Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Aplicação de técnicas de mineração em um programa de concessão de benefícios ao consumidor: o caso do Programa Nota Legal do Distrito Federal

Mário Henrique Paes Vieira

Dissertação apresentada como requisito parcial
para conclusão do Programa de Pós-Graduação em Computação Aplicada

Orientador
Prof. Dr. Marcelo Ladeira

Brasília
2014

Ficha catalográfica elaborada pela Biblioteca Central da Universidade de Brasília. Acervo 1018084.

Vieira, Mário Henrique Paes.
V658a Aplicação de técnicas de mineração em um programa de concessão de benefícios ao consumidor : o caso do Programa Nota Legal do Distrito Federal / Mário Henrique Paes Vieira. -- 2014.
xiv, 117 f. : il. ; 30 cm.

Dissertação (mestrado) - Universidade de Brasília, Instituto de Ciências Exatas, Departamento de Ciência da Computação, 2014.
Orientação: Marcelo Ladeira.
Inclui bibliografia.

1. Banco de dados - Gerência. 2. Sistemas de recuperação da informação - Administração. 3. Notas fiscais. 4. Mineração de dados (Computação). I. Ladeira, Marcelo. II. Título.

CDU 004.658

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Aplicação de técnicas de mineração em um programa de concessão de benefícios ao consumidor: o caso do Programa Nota Legal do Distrito Federal

Mário Henrique Paes Vieira

Dissertação apresentada como requisito parcial para conclusão do
Mestrado Profissional em Computação Aplicada

Prof. Dr. Marcelo Ladeira (Orientador)
CIC/UnB

Prof. Dr. Rommel Novaes Carvalho
CIC/UnB e CGU

Prof. Dr. Remis Balaniuk
UCB e TCU

Prof. Dr. Marcelo Ladeira
Coordenador do Programa de Pós-graduação em Computação Aplicada

Brasília, 02 de julho de 2014

Dedicatória

A todos servidores públicos que buscam aprimoramento em suas atividades e ao seu trabalho, mesmo relegando temporariamente tempo com a família, que possam dar retorno merecido à sociedade. A todos que tenham êxito ao superar os caminhos mais difíceis. A nós.

"Há homens que lutam um dia e são bons. Há outros que lutam um ano e são melhores. Há aqueles que lutam muitos anos e são muito bons. Mas há os que lutam por toda a vida, esses são os imprescindíveis". Bertold Brecth

Agradecimentos

À minha esposa e filhos (João Vitor e Amanda) pela compreensão durante o tempo que me dediquei ao mestrado. João Vitor, por sempre me apoiar com um sorriso, não importando as dificuldades sofridas no último ano. Amanda, mesmo estando à sua espera, a expectativa de sua chegada me deu forças para completar este trabalho.

Aos professores do MPCA que incentivaram a proposta do mestrado profissional.

À Secretaria de Estado de Fazenda do Distrito Federal pelo apoio, recursos e acesso aos dados utilizados na pesquisa.

Agradeço a Edson Pinheiro Alvarista por ajuda nos procedimentos do INSS para correção da base de dados e aos demais colegas pelas trocas de informações ao longo de todo o curso.

Resumo

O Programa Nota Legal (PNL) da Secretaria de Estado da Fazenda do Distrito Federal (SEF) é um programa de concessão de benefícios fiscais que permite que consumidores pessoa física e empresas optantes pelo Simples Nacional tenham benefícios fiscais sobre aquisição de bens e na prestação de serviços. Ao exigir o documento fiscal, os consumidores podem recuperar até 30% do Imposto sobre Circulação de Mercadorias e Prestação de Serviços (ICMS), em aquisição de bens, e do Imposto sobre Serviços de Qualquer Natureza (ISS), na prestação de serviços. Os consumidores podem usar os benefícios fiscais no Imposto sobre a Propriedade Predial e Territorial Urbana (IPTU) e no Imposto sobre Propriedade de Veículos Automotores (IPVA), ou ainda receber como crédito em sua conta bancária caso não possuam bens. No período de vigência do Programa Nota Legal, entre 2008 a 2013, no banco de dados se encontram cadastradas aproximadamente: 95.000 empresas, 830.000 consumidores e 157.000.000 de documentos fiscais processados.

O objetivo desta pesquisa é analisar, via técnicas de mineração de dados, dois perfis que visam melhoria da gestão do programa: a fidelidade das pessoas físicas ao programa e a obtenção de créditos de consumo pelos beneficiários. Enquanto o perfil de fidelidade leva em conta a variação do tempo sobre as pessoas físicas participantes e propõe indicadores para avaliação do programa, o perfil de créditos analisa a distribuição dos créditos entre os consumidores. A metodologia CRISP-DM é utilizada e ao longo de suas fases é realizada a integração do banco de dados do PNL com outras bases de dados existentes na SEF. Para melhoria da qualidade da informação, são removidos ruídos, *missing values* e *outliers*. Estatística é utilizada para extração do conhecimento gerado sobre os perfis desejados.

Dados foram obtidos para melhoria da gestão desse programa com os perfis e indicadores apurados, ao propiciar uma linha de base para comparação do PNL com outros programas de concessão de benefícios. Foram também obtidos meios para permitir maior transparência aos cidadãos com painéis de informações sobre benefícios e beneficiários do PNL.

Palavras-chave: Programa Nota Legal, CRISP-DM, Documento fiscal, Mineração de Dados, ICMS, ISS

Abstract

The Programa Nota Legal (PNL, in Portuguese) from the Department of Treasury of the Federal District (SEF) is a program for granting tax benefits that allows individual consumers and companies opting for "Simples Nacional" to have tax benefits on purchase of goods and provision of services. By requiring the tax document, consumers can recover up to 30% on Tax on the Circulation of Goods (ICMS, in Portuguese), when purchasing goods, and Services Tax (ISS, in Portuguese), when provisioning services. Consumers can use the tax benefits in the Tax on Land Property (IPTU, in Portuguese) and Tax on Motor Vehicles (IPVA, in Portuguese), or receive the credit in their bank account in case they neither own a land nor a vehicle. From 2008 to 2013, in the database are stored approximately: 95,000 companies, 830,000 consumers, and 157,000,000 processed tax documents.

The objective of this research is to analyze, via data mining techniques, two profiles aimed at improving the program management: fidelity of physical people to the program and obtaining consumer credit by beneficiaries. While the profile of fidelity takes into account the time variation of the participating individuals and proposes indicators for evaluating the program, the credit profile analyzes the distribution of credits among consumers. The CRISP-DM methodology is used and throughout its phases is performed the PNL database integration with other existing databases in SEF. To improve the quality of information, noises, missing values and outliers are removed. Statistics is used for knowledge extraction about the desired profiles.

Data were obtained for management improval of this program with calculated profiles and indicators, providing a baseline for comparison from PNL to other programs of benefit payments. Were also obtained means to allow greater transparency for citizens with information panels about benefits and beneficiaries of PNL.

Keywords: Programa Nota Legal, CRISP-DM, Taxation document, Data Mining, ICMS, ISS

Sumário

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

Este capítulo apresenta a especificação geral do problema abordado, sua relevância e as áreas de pesquisa relacionadas. A Seção 1.1 introduz o problema a ser abordado. A Seção 1.2 apresenta o principal objetivo desta pesquisa. A Seção 1.3 ilustra alguns trabalhos relacionados ao tema.

1.1 Justificativa do Tema

Em 13 de junho de 2008, foi sancionada a Lei nº 4.159 que dispõe sobre a criação do programa de concessão de créditos para adquirentes de mercadorias ou bens e tomadores de serviços [14], mais conhecido como Programa Nota Legal.

O Programa Nota Legal (PNL) da Secretaria de Estado da Fazenda do Distrito Federal (SEF) permite que consumidores pessoa física e empresas optantes pelo Simples Nacional possam recuperar até 30% do Imposto sobre Circulação de Mercadorias e Prestação de Serviços (ICMS) e do Imposto sobre Serviços de Qualquer Natureza (ISS) efetivamente recolhido pelos estabelecimentos fornecedores ou prestadores de serviço.

Ao mesmo tempo em que se pretende recompensar o cidadão que exerce seus direitos exigindo o documento fiscal, o PNL também busca reduzir o mercado informal e propiciar o incremento da arrecadação tributária, visando suprir o Distrito Federal de recursos financeiros necessários para o cumprimento de sua função social. A sociedade ganha também com a redução da concorrência desleal, coibindo a sonegação fiscal.

Embora o PNL já fosse vislumbrado desde 2007, apenas com a regulamentação da lei de criação do PNL, a partir de [13], que houve esforços para sua operacionalização. Mesmo com dificuldade na obtenção de recursos e estrutura organizacional dentro da SEF, foi implementado um sistema para cadastramento de pessoas físicas e jurídicas, processamento de documentos fiscais, cálculo de benefícios e verificação de reclamações dos usuários.

Nos anos seguintes houve um melhor aparelhamento da SEF, em que foi destinada uma área exclusiva do Fisco para cuidar do PNL. Além disso, foram destinados mais recursos de TI para aquisições em infraestrutura e desenvolvimento de software. Mesmo com esta nova estrutura, o crescimento do PNL foi além do esperado, enquanto a área do Fisco se especializou no tratamento aos consumidores, contadores e empresas; houve atenção da TI para melhorias no site disponibilizado.

Uma vez que a SEF tem a competência institucional de promover a gestão tributária e financeira distrital, a motivação da presente pesquisa reside na necessidade da SEF estar munida de ferramental tecnológico para melhorar a tomada de decisões e fazer frente ao desafio de financiamento do setor público e uso dos recursos necessários para investimentos na sociedade.

1.2 Objetivo

É de interesse da SEF o conhecimento de perfis de créditos de consumo pelos beneficiários e de perfis de fidelidade das pessoas físicas ao PNL no intuito de melhorar a gestão desse programa, ou ainda garantir maior transparência das informações aos cidadãos.

1.3 Áreas de Pesquisa Relacionadas

Foram pesquisados trabalhos que envolvessem programas de concessão de benefícios ao consumidor de todo o país. Foram encontradas pesquisas de áreas multidisciplinares como:

- Direito - que em [33] afere a sensibilidade da cidadania fiscal entre contribuintes do PNL, em [43] discute de maneira empírica estratégias regulatórias para um programa de concessão de benefícios e em [34] analisa o uso de habilidades de negociação junto ao programa Nota Paulista;
- Economia: que em [36] e avalia o impacto do programa Nota Paulista sobre a arrecadação de São Paulo e em [48] avalia o impacto do programa de concessão de benefícios sobre a arrecadação de São Paulo e Alagoas;
- Ciências contábeis: que em [49] através de métodos contábeis relaciona redução da sonegação fiscal com o crescimento da receita tributária no DF e em [46] avalia a repercussão do programa Nota Fiscal Paulista;
- Engenharia: que em [24] propõe um novo equipamento emissor de nota fiscal para melhoria tecnológica do programa Nota Paulista

- Administração: que em [39] analisa elementos de governo eletrônico junto ao Nota Paulista

Foram também pesquisados trabalhos que utilizaram técnicas de mineração de dados para problemas da área tributária. Cabe citar:

- Em [21] a tese intitulada "Um Modelo para Gerenciamento, Avaliação e Planejamento da Arrecadação de Impostos Estaduais" propõe um modelo visando combater a sonegação fiscal e aumentar a receita estadual sem elevação da carga tributária. Para coibir a evasão dos tributos devidos, este trabalho se utiliza de *datawarehouse* e *datamining* na Secretaria de Estado de Fazenda do Tocantins.

- Em [23] mineração de dados é aplicada ao problema da sonegação do ICMS na Secretaria de Estado de Fazenda do Ceará com a utilização de Redes Neurais Artificiais. O modelo disponibilizado serve como uma ferramenta flexível de controle fiscal ao ser treinada para novas tendências ou padrões de sonegação.

- [2] utiliza rede neural e regressão linear no auxílio na identificação de indícios de infração à legislação tributária na Receita Federal do Brasil. O modelo obtido se mostrou eficiente para predizer o valor da receita bruta dos contribuintes e identificar omissões de receitas.

- O artigo em [26] propõe que o planejamento é o fator crítico de sucesso do processo de mineração de dados. Como estudo de caso, ele se utiliza do modelo de descoberta do conhecimento (KDD) usando árvores de decisão, direcionado à detecção de fraudes fiscais. Como existem conflitos entre maximizar o quanto deve ser auditado e minimizar o custo da auditoria, a utilização do KDD apresentou diminuição da complexidade do estudo ao dividir a mineração de dados em fases e realizar o planejamento específico para cada fase.

Mesmo não se restringindo apenas a estas referências, não foram encontrados trabalhos que fizessem análises semelhantes a esta pesquisa ao relacionar obtenção de perfis com programas de concessão de benefícios ao consumidor.

1.4 Organização desta Dissertação

Esta dissertação está organizada da seguinte forma: no Capítulo 2 são apresentados os fundamentos teóricos necessários para desenvolvimento dos capítulos seguintes. No Capítulo 3 é apresentada a organização atual do PNL na SEF e suas peculiaridades. No Capítulo 4 é apresentado o modelo de referência CRISP-DM sendo detalhadas as fases que serão abordadas nesta pesquisa. No Capítulo 5 é apresentada a fase de entendimento

do negócio. No Capítulo 6 é apresentada a fase de compreensão dos dados. No Capítulo 7 é apresentada a fase de preparação dos dados. No Capítulo 8 temos a apresentação e análise dos resultados. No Capítulo 9 temos as conclusões e trabalhos futuros.

Capítulo 2

Fundamentação Teórica

Este capítulo apresenta fundamentos teóricos necessários para desenvolvimento dos capítulos seguintes. A Seção 2.1 conceitua mineração de dados. A Seção 2.2 introduz conceitos de estatística. A Seção 2.3 conceitua *Outliers* e seu uso. A Seção 2.4 apresenta o WEKA, um dos principais *frameworks* de Mineração de Dados.

2.1 Mineração de Dados

Diversas definições de mineração de dados podem ser encontradas na literatura. Entre as diversas definições podem ser destacadas as seguintes:

- Conforme [29], "Mineração de Dados é o processo de descoberta de padrões de interesse e de conhecimento em grandes quantidades de dados".
- Conforme [47], "Mineração de Dados é o processo de proposição de várias consultas e extração de informações úteis, padrões e tendências, frequentemente desconhecidos, a partir de grande quantidade de dados armazenada em bancos de dados".
- Em [31], a definição é dada de uma perspectiva estatística: "Mineração de Dados é a análise de grandes conjuntos de dados a fim de encontrar relacionamentos inesperados e de resumir os dados de uma forma que eles sejam tanto úteis quanto compreensíveis ao dono dos dados".
- Conforme [41], "Mineração de Dados é um campo interdisciplinar que junta técnicas de máquinas de conhecimentos, reconhecimento de padrões, estatísticas, banco de dados e visualização, para conseguir extrair informações de grandes bases de dados".

Para esta pesquisa considera-se mineração de dados como exploração de grandes bancos de dados visando extração de conhecimento através do uso de técnicas que envolvem

relacionamento entre variáveis, de forma que o conhecimento gerado seja útil e compreensível.

Mineração de dados é parte de um processo maior denominado Busca de Conhecimento em Bancos de Dados (Knowledge Discovery in Database - KDD), o qual possui uma metodologia própria para preparação e exploração dos dados, interpretação de seus resultados e assimilação de conhecimento. No entanto, mineração de dados se tornou mais conhecida do que o próprio processo de KDD em função de ser a etapa onde são aplicadas as técnicas de busca de conhecimento.

2.1.1 Tarefas de mineração de dados

As tarefas correspondem aos problemas que podem ser tratados por mineração de dados. Diferentes tipos de métodos e técnicas são necessários para encontrar diferentes padrões de dados resultando em tarefas distintas. Seguem algumas definições para tarefas usadas na literatura:

- Em [37] - classificação/predição, segmentação, associação, *clustering*, visualização e otimização;
- Em [51] - classificação, estimação, segmentação e descrição;
- Em [50] - predição, detecção de desvios, segmentação, *clustering*, associação, sumarização, visualização e mineração de textos;
- Em [53] - sumarização, classificação, associação, *clustering* e análise de tendências.

Uma breve descrição é apresentada para cada uma das tarefas citadas:

Descrição - descreve os padrões e tendências revelados pelos dados. A descrição geralmente oferece uma possível interpretação para os resultados obtidos. A tarefa de descrição é utilizada em conjunto com técnicas de análise exploratória de dados, para comprovar a influência de certas variáveis no resultado obtido.

Classificação - visa identificar a qual classe um determinado registro pertence. Nesta tarefa, o modelo analisa o conjunto de registros fornecidos, com cada registro já contendo a indicação à qual classe pertence, a fim de "aprender"como classificar um novo registro (aprendizado supervisionado).

Estimação - similar à classificação, porém é usada quando o registro é identificado por um valor numérico e não nominal. Assim, pode-se estimar o valor de uma determinada variável analisando-se os valores das demais.

Predição - similar às tarefas de classificação e estimação, porém ela visa descobrir o valor futuro de uma determinada variável.

Clustering - visa identificar e aproximar registros similares. Um agrupamento (ou *cluster*) é uma coleção de registros similares entre si, porém diferentes dos outros registros nos demais agrupamentos. Esta tarefa difere da classificação pois não necessita que os registros sejam previamente categorizados (aprendizado não-supervisionado). Além disso, ela não tem a pretensão de classificar, estimar ou predizer o valor de uma variável, ela apenas identifica os grupos de dados similares.

Associação - consiste em identificar quais variáveis estão relacionadas. Apresentam a forma: SE variável X ENTÃO variável Y. Esta tarefa é conhecida principalmente em análises de "Cestas de Compras"(*Market Basket*), onde são identificados quais produtos comprados em conjunto pelos consumidores.

Segmentação - subdivide-se os dados em conjuntos menores com comportamento similar pelas variáveis de segmentação. Com estes subconjuntos, pode-se determinar novos agrupamentos ou fazer algum tipo de prognóstico. Difere de *clustering* pois a formação de grupos é conduzida pelo usuário e não determinada pelo método ou técnica usada.

Visualização - utilizada, principalmente, quando não há conhecimento prévio da distribuição dos dados e para encontrar disparidades nos dados. Após esta análise, é possível identificar como segmentar dados ou selecionar variáveis no uso de *clustering*.

Otimização - visa otimizar recursos limitados como: tempo, espaço, custo. Busca-se maximizar variáveis de resultado como venda, lucro, economia, ... Esta tarefa é similar a área de pesquisa operacional, a qual trata problemas de otimização, sempre sujeito a restrições.

Detecção de desvios - similar à análise de *outliers* que é detalhada em 2.3.

Sumarização - é a abstração ou generalização dos dados resultando em um conjunto de dados menor que fornece uma visão geral dos dados com informações agregadas.

Mineração de textos - trabalha com dados armazenados como texto. Deseja-se extrair informações com base em técnicas de tratamento e exploração de textos.

Análise de tendências - procura encontrar padrões e regularidades na evolução dos dados ao longo do tempo.

2.2 Conceitos de Estatística Descritiva

Esta Seção apresenta alguns conceitos de estatística que foram utilizados nesta pesquisa conforme [1] e [25].

Estatística Descritiva é a parte da Estatística que utiliza técnicas para descrever e sumarizar dados representativos do comportamento de uma variável, onde se utilizam tabelas, gráficos e medidas que resumem a distribuição desta variável.

As variáveis podem ser classificadas em qualitativas e quantitativas. As variáveis qualitativas podem ser nominais quando permitem identificar categorias; ou ordinais quando permitem ordenar categorias. Nas variáveis quantitativas suas realizações são números resultantes de contagem ou mensuração que podem ser discretas quando assumem somente valores inteiros; ou contínuas quando assumem valores fracionários.

Entre as medidas que resumem as distribuição de variáveis temos as medidas de posição e as medidas de dispersão.

Medidas de posição são valores que representam o conjunto de dados observados ou então promovem uma partição sobre este conjunto. Entre as medidas de posição destacam-se as medidas de tendência central e as separatrizes.

Medidas de tendência central representam os fenômenos pelos seus valores médios em torno dos quais tendem a se concentrar os dados. São parâmetros que permitem que se tenha uma primeira ideia, um resumo, de como se distribuem os dados de um experimento.

Existem três medidas principais que refletem a tendência central de uma distribuição de frequências: média, moda e mediana.

Média é a soma de todos os resultados dividida pelo número total de casos.

Moda é o valor mais frequente do conjunto de dados observados. É o evento ou categoria de eventos que ocorre com maior frequência indicando o valor ou categoria mais provável. A moda não é necessariamente única, ao contrário da média ou da mediana.

Mediana é o valor numérico que separa a metade superior de uma variável, da metade inferior. Se o número de observações for impar, a mediana será o valor central da distribuição; se o número for par, a mediana será a média dos dois valores centrais.

Medidas separatrizes dividem uma base de dados em "n" partes iguais. Em especial, temos que: mediana - divide o conjunto em 2 partes iguais; quartil - divide o conjunto em 4 partes iguais; decil - divide o conjunto em 10 partes iguais; e percentil - divide o conjunto em 100 partes iguais.

Portanto o 1º percentil determina o 1% menor dos dados; o 98º percentil determina o 98% menor dos dados; o 25º percentil é o primeiro quartil; e o 50º percentil é a mediana. De igual forma, o 10º percentil é o primeiro decil e o 80º percentil é o oitavo decil.

Medidas de dispersão complementam as informações fornecidas pelas medidas de posição. Descrevem a variabilidade que ocorre na base de dados analisada. Permitem identificar até que ponto os resultados se concentram ou não ao redor da tendência central de um conjunto de observações.

Existem várias medidas para avaliar a dispersão. As principais são: amplitude, variância, desvio padrão e coeficiente de variação.

Amplitude é a diferença entre o maior e o menor valor que foi observado para a variável, servindo para caracterizar a abrangência.

Variância é a soma dos quadrados dos desvios de cada ponto em torno da média. Caracteriza a dispersão dos pontos de uma amostra potencializando as diferenças.

Desvio padrão é a raiz quadrada da variância. Um baixo desvio padrão indica que os dados tendem a estar próximos da média. Um desvio padrão alto indica que os dados estão espalhados por uma gama de valores.

Coeficiente de variação é o resultado do desvio padrão dividido pela média, transformado em percentual sendo independente das unidades adotadas. Por essa razão, é vantajosa para a comparação de distribuições cujas unidades podem ser diferentes. O coeficiente de variação indica a homogeneidade da distribuição. Considera-se nesta pesquisa como baixa dispersão valores menores ou iguais a 15%, média dispersão valores entre 15% a 30% e alta dispersão para valores maiores ou iguais a 30%.

Esta pesquisa ainda se utiliza do conceito de distribuição acumulada. A distribuição acumulada, dada por P[X<=x], descreve a distribuição da probabilidade de uma variável aleatória X assumir valores menores ou iguais a x.

2.3 Análise de *Outliers*

De acordo com [30], um *outlier* é um objeto de dados que desvia significativamente dos outros objetos, como se fosse gerado por um mecanismo diferente dos demais.

A análise de *outliers* visa identificar e tratar objetos que possuam características significativamente diferentes dos demais registros do conjunto de dados. Conforme [31], em muitas situações o objetivo da mineração de dados é detectar anomalias. Na detecção de fraudes ou de falhas estes objetos que diferem da maioria é o que está sendo procurado. Por outro lado, se o objetivo de mineração de dados é construção de modelos, *outliers* podem simplesmente obscurecer os pontos principais do modelo. Neste caso, deve-se identificar e removê-los antes de construir os modelos.

Um *outlier* pode ocorrer devido a uma variabilidade na medida ou pode indicar um erro experimental, sendo que neste caso devem ser excluídos da base de dados. Certos métodos estatísticos e algoritmos de mineração de dados são sensíveis à presença de *outliers*, sendo que seus resultados podem ser comprometidos.

Em [27], o estudo de *outliers*, independentemente da(s) sua(s) causa(s), pode ser realizado em várias fases:

- A fase inicial é a da identificação das observações que são potencialmente aberrantes. A identificação de textitoutliers consiste na detecção, com métodos subjetivos, das observações surpreendentes. A identificação é feita, geralmente, por análise gráfica ou, no caso de um número de dados ser pequeno, por observação direta dos mesmos.

São assim identificadas as observações que têm fortes possibilidades de virem a ser designadas por *outliers*.

- Na segunda fase tem-se como objetivo a eliminação da subjetividade inerente à fase anterior. Pretende-se saber se as observações identificadas como *outliers* potenciais o são, efetivamente. São realizados testes junto às observações "preocupantes". Devem ser escolhidos os testes mais adequados para a situação em estudo. As observações suspeitas são testadas quanto à sua discordância. Se for aceita a hipótese de algumas observações serem *outliers*, elas podem ser designadas como discordantes. Uma observação diz-se discordante se puder considerar-se inconsistente com os restantes valores depois da aplicação de um critério estatístico objetivo. Muitas vezes o termo discordante é usado como sinônimo de *outlier*.

- Na última fase é necessário decidir o que fazer com as observações discordantes. A maneira mais simples de lidar com estas observações é eliminá-las. Esta abordagem, apesar de ser muito utilizada, não é aconselhável. Ela só se justifica no caso de os *outliers* serem devidos a erros cuja correção é inviável. Caso contrário, as observações consideradas como *outliers* devem ser tratadas cuidadosamente pois contêm informação relevante sobre características subjacentes aos dados e poderão ser decisivas no conhecimento da população à qual pertence a amostra em estudo.

2.4 *Framework* WEKA

A *framework* WEKA começou a ser idealizada em 1993 por um grupo de pesquisadores da universidade de Waikato, localizada na Nova Zelândia. Ao longo dos anos se consolidou como uma das ferramentas de mineração mais utilizadas. Dentre seus recursos, contém uma coleção de algoritmos de aprendizado de máquina para as atividades de Mineração de Dados que podem ser aplicados diretamente a uma base de dados. Além disso, o WEKA possui facilidades para o desenvolvimento de novos algoritmos.

O WEKA é um software livre dentro das especificações *General Public License* (GPL). Possui implementados recursos para pré-processamento, classificação, regressão, *clustering*, associação e visualização. [52]

2.4.1 Arquivo ARFF

O método preferido do WEKA para carga de dados é o Formato de Arquivo de Atributo-Relação (ARFF), onde é possível definir os tipos de dados serão carregados e seus valores.

O arquivo ARFF é formado por duas seções: cabeçalho e dados.

O cabeçalho contêm um nome para a base de dados, uma lista das variáveis e seus tipos de dados.

Cada variável declarada no cabeçalho indica a ordem em que seus valores serão declarados na seção de dados. Há quatro tipo de dados que o WEKA dá suporte: nominal, numérico, String (valores de texto arbitrário) e data.

Na seção de dados, cada instância da base de dados representa uma linha do arquivo. Os valores das variáveis é separado por vírgulas. *Missing values* são representados pelo caractere "?".

Capítulo 3

Contextualização e Perspectivas

Este capítulo contextualiza o Programa Nota Legal junto à Secretaria de Estado de Fazenda do DF e apresenta os desafios e possibilidades de melhoria. A Seção 3.1 contextualiza a SEF, apresenta suas atribuições e termos usados em suas atividades. A Seção 3.2 detalha algumas características do ICMS usadas no PNL. A Seção 3.3 apresenta um resumo sobre o funcionamento do PNL. A Seção 3.4 apresenta algumas características sobre o envio de Livro Fiscal Eletrônico (LFE) à SEF. A Seção 3.5 apresenta limitações desta pesquisa.

3.1 SEF

A Secretaria de Estado de Fazenda do Distrito Federal, diferente dos outros Estados da Federação, acumula a arrecadação tributária dos tributos de competência estadual e municipal. Além disso, tem atuação e competência nas seguintes áreas: arrecadação de tributos; política tributária e fiscal; gestão financeira e contabilidade pública; e operações de crédito e dívida pública.

Estas atividades finalísticas têm suporte na Tecnologia da Informação da SEF que mantem os maiores sistemas governamentais do DF: SIGEST - Sistema Integrado de Gestão Tributária, onde são realizados os cadastros e pautas de valores necessários aos sistemas de lançamento de tributos; SITAF - Sistema Integrado de Tributação e Administração Fiscal, onde é processada a arrecadação tributária; SIGGO - Sistema Integrado de Gestão Governamental, onde são realizadas: contabilidade, planejamento, orçamento público e as despesas governamentais; e SISGEPAT - Sistema Geral de Patrimônio, destinado à execução das atividades de administração e controle dos bens patrimoniais móveis e semoventes de propriedade do Distrito Federal.

Para cumprimento de sua missão e suporte aos Sistemas, a SEF possui convênios de troca de informações com a Receita Federal do Brasil (RFB), responsável pela arredação

da União; e outros órgãos do DF como: TJDFT, Câmara Legislativa e outras Secretarias de Governo. Para execução de atividades financeiras há também troca de informações com o Banco do Brasil e o Banco de Brasília (BRB). Em especial, a SEF possui contrato com os Correios para comunicação com os cidadãos e empresas no intuito de realizar cobranças e outros tipos de comunicados.

O Programa Nota Legal se utiliza da infraestrutura existente principalmente do SIGEST e do SITAF para obtenção de informações tributárias e, tendo suporte nos convênios existentes, permite troca de informações com os órgãos externos.

Para melhor contextualização às atribuições fazendárias, segue abaixo terminologia usada no PNL pela SEF.

Beneficiário, no contexto do PNL, é a pessoa física ou jurídica, esta última desde que esteja cadastrada no Simples Nacional, adquirente de mercadorias ou bens e tomadores de serviços que será beneficiada com créditos para desconto em tributos: IPVA e/ou IPTU.

Contribuinte, no contexto do PNL, é a pessoa que recolhe o ISS ou o ICMS.

CNAE apresenta dados com codificação e descrição das atividades econômicas dos estabelecimentos. Acrônimo para Classificação Nacional de Atividades Econômicas.

Documentos fiscais são documentos, em papel ou meio eletrônico, utilizados pelos contribuintes (pessoas físicas ou jurídicas) para demonstrar e registrar sua conformidade com as obrigações tributárias. São regulados pelos órgãos fazendários federal, estadual e municipal. Cupom fiscal e nota fiscal são espécies de documentos fiscais.

Cupom fiscal é um documento fiscal, de emissão por meio de um equipamento obrigatório nas vendas por empresas, onde o vendedor está obrigado a emiti-lo, sendo que o mesmo deve transitar junto com os produtos, pois é o que garante o trânsito nacional, sem que haja a apreensão dos produtos pelos órgãos de fiscalização dos Estados e Municípios, ou mesmo da União.

ICMS, sigla para Imposto sobre Circulação de Mercadorias e Prestação de Serviços, tem cinco hipóteses de incidência: operações de circulação de mercadorias; prestação de serviços de transporte interestadual ou intermunicipal; serviços de comunicação; energia elétrica; e importação. Pela legislação do PNL, apenas operações de circulação de mercadorias podem gerar benefícios aos consumidores. Maior detalhamento é apresentado na Seção 3.2.

IPTU, sigla para Imposto sobre a Propriedade Predial e Territorial Urbana. O fato gerador do IPTU ocorre na aquisição de imóvel. A base de cálculo do IPTU é o valor venal do imóvel, que é definido como o valor de venda do imóvel em condições normais de mercado. O valor venal é determinado pela SEF por meio de avaliação realizada, na qual são considerados alguns fatores que interferem na composição do valor do imóvel, como, por exemplo: a área do terreno, a destinação ou a natureza da utilização do terreno,

a área construída, o valor unitário do metro quadrado, os serviços públicos existentes, a valorização do logradouro, e outros fatores aferidos no mercado imobiliário, conforme consta em [12].

IPVA, sigla para Imposto sobre Propriedade de Veículos Automotores. O fato gerador do IPVA ocorre na aquisição de veículo automotor. A base de cálculo é o valor venal do veículo, conforme consta em [16].

ISS, sigla para Imposto sobre Serviços de Qualquer Natureza. Tem como fato gerador a prestação de serviços constantes da lista anexa à Lei Complementar 116/2003, disponível em [6], ainda que esses não se constituam como atividade preponderante do prestador.

Simples Nacional é o sistema simplificado de recolhimento de tributos e contribuições. Abrange a participação de todos os entes federados (União, Estados, Distrito Federal e Municípios). Para o ingresso de empresas no Simples Nacional é necessário o cumprimento das seguintes condições: enquadrar-se na definição de microempresa ou de empresa de pequeno porte; cumprir os requisitos previstos na legislação; e formalizar a opção pelo Simples Nacional.

LFE, sigla para Livro Fiscal Eletrônico. É composto por documentos fiscais e contêm informações para tributação. Enviado mensalmente à SEF como obrigação tributária acessória.

3.2 O ICMS

O Imposto sobre Circulação de Mercadorias e Prestação de Serviços é um tributo de competência estadual. Como a legislação do ICMS não é unificada, já que cada um dos 26 estados e o Distrito Federal têm competência para legislar nesta matéria, a legislação desse tributo é frequentemente alterada e atualizada com novos procedimentos tributários. Sendo assim, contribuintes que atuem em diversos estados precisam ter conhecimento de uma grande quantidade de normas, dentre elas protocolos, convênios, leis estaduais e decretos envolvendo o ICMS.

O ICMS é um imposto não cumulativo, ou seja, significa dizer que o cálculo do valor a ser recolhido pelos estabelecimentos recolhedores do imposto deve ser calculado como uma conta corrente em que os débitos são o imposto devido nas vendas e os créditos são os impostos recolhidos nas operações anteriores.

Por exemplo, admitindo-se que a alíquota do ICMS seja 17,00 %, uma empresa adquire mercadoria no valor de R$ 100,00 com R$ 17,00 de ICMS já embutido no valor da compra. Se a empresa vender a mercadoria por R$ 150,00, terá que embutir R$ 25,50 de ICMS. O ICMS a recolher é a diferença entre o valor do imposto incluso na venda (débito) e o valor incluso na compra (crédito), ou seja, R$ 8,50 (R$ 25,50 – R$ 17,00).

Esse cálculo deve ser processado por meio de uma apuração periódica com a diferença entre o somatório de todos os débitos e o somatório de todos os créditos.

O ICMS é um imposto lançado pelo próprio contribuinte que tem a obrigação de apurar o que é devido, com base na legislação vigente, e de informar todos os elementos do cálculo para a administração tributária.

3.3 Procedimentos do PNL

O Programa Nota Legal é um programa de estimulo à cidadania fiscal no Distrito Federal, que tem por objetivo estimular os consumidores a exigirem a entrega do documento fiscal na hora da compra. Como compensação ao contribuinte, o PNL gera créditos aos consumidores.

Para obtenção do benefício, é necessário que o consumidor exija o registro do seu CPF ou CNPJ no documento fiscal emitido. A empresa participante, por sua vez, para a concretização do benefício deve encaminhar mensalmente os documentos fiscais emitidos com a identificação do CPF/CNPJ do consumidor, bem como efetuar o pagamento dos impostos devidos (ICMS/ISS).

O cadastramento dos beneficiários no Programa Nota Legal dá-se de forma automática na data do primeiro registro pela empresa participante com a indicação do CPF/CNPJ do consumidor. Contudo, para fins de consulta, acompanhamento, utilização de créditos e registro de reclamação, o beneficiário deve registrar suas informações cadastrais por meio do portal do programa[1]. Os créditos, após a liberação da SEF, ficam disponíveis por 2 anos para utilização.

O cálculo de créditos é efetuado sobre o valor recolhido pelos estabelecimentos comerciais, o que é feito por período e não no momento de cada operação. Além disso, o cálculo dos créditos é realizado após o vencimento dos prazos de registro e de retificação dos documentos fiscais. O estabelecimento comercial que deixar de emitir ou de entregar ao consumidor documento hábil ou não efetuar o registro eletrônico no prazo estabelecido, fica sujeito a multa, sem prejuízo de penalidades tributárias.

Os créditos recebidos pelo consumidor variam de acordo com diversas circunstâncias, como o valor da nota fiscal de compra, o quanto o estabelecimento recolheu de imposto no mês e se o estabelecimento tem créditos de ICMS. O documento fiscal pode ter qualquer valor, entretanto o consumidor terá direito aos créditos proporcionais ao valor de suas compras. Após os cálculos dos créditos, o consumidor aguarda a liberação para utilização na compensação dos impostos devidos: IPVA ou IPTU. Após 2013 houve alteração na

[1] http://www.notalegal.df.gov.br/

lei permitindo que contribuintes que não possuam imóveis ou veículos possam receber os créditos em suas contas bancárias.

Para utilização dos créditos no abatimento do valor do IPTU e do IPVA, se consumidor pessoa física, não se exige vínculo entre o detentor do crédito e os imóveis ou veículos. Contudo, não pode haver débito pendente de pagamento em nome do titular dos créditos, para os imóveis e os veículos indicados e seus proprietários/arrendatários.

A empresa participante deve transmitir à SEF os dados dos documentos fiscais e do consumidor até o final do mês subsequente da compra ou aquisição feita no estabelecimento. Encerrado este prazo, caso o documento não conste em consulta disponibilizada no site do PNL, ou conste com divergência de dados, o consumidor poderá registrar reclamação no segundo mês subsequente, exclusivamente pelo site do programa, guardando o original do documento para apresentação à SEF, no caso de ser notificado pela não regularização efetuada pelo contribuinte.

O início do prazo para apresentação do documento fiscal pelo consumidor pode ser suspenso pelo período necessário ao processamento dos documentos fiscais transmitidos pelas empresas participantes do Nota Legal.

Os seguintes produtos e serviços não dão direito a créditos do PNL:

- combustíveis líquidos ou gasosos e lubrificantes, derivados ou não de petróleo;
- serviços de comunicação (exemplos: conta de telefone, TV a cabo, internet);
- operações não sujeitas à tributação (exemplos: livros, revistas e produtos hortifrutigranjeiros isentos);
- operações de fornecimento de energia elétrica;
- prestação de serviços bancários ou financeiros;
- serviços prestados por profissionais autônomos ou sociedades uniprofissionais;
- operações realizadas por feirante, ambulante ou produtor rural;
- operações ou prestações de microempresa optante do Simples Nacional cuja receita bruta seja, no ano calendário anterior, igual ou inferior a R$ 36.000,00;
- se o adquirente for contribuinte do ICMS ou do ISS, não optante do Simples Nacional;
- se o adquirente ou o tomador for órgão ou entidade da administração pública direta ou indireta;

- e na hipótese de documento: a) inidôneo; b) não hábil para acobertar a operação ou prestação; c) que não identifique corretamente o adquirente ou tomador; d) emitido mediante fraude, dolo ou simulação.

3.3.1 Participação das empresas no PNL

Cada pessoa jurídica legalmente cadastrada possui um CNAE associado. A participação da empresa no PNL ocorre de acordo com a CNAE do estabelecimento, caso seja uma das relacionadas no Anexo Único da Portaria SEF nº 323/2008 [15], observado a data de ingresso em caráter obrigatório ou facultativo.

A adesão, prevista na legislação como facultativa (a critério da empresa) ocorre mediante a identificação do consumidor no documento fiscal, sujeitando-se, a partir de então, à legislação do PNL.

Caso a pessoa jurídica emita documentos fiscais não relacionados ao seu CNAE, ou seja, à sua atividade econômica preponderante, o contribuinte não deverá informar a identificação do consumidor (CPF/CNPJ) no envio do Livro Fiscal Eletrônico.

3.3.2 Cálculo do crédito do PNL

O crédito do PNL é calculado após uma consolidação do sistema, observando-se:

I) cada aquisição possui uma fração para fins de atribuição do crédito, sendo o numerador o valor do documento fiscal emitido e o denominador o total de vendas do estabelecimento dos bens sobre os quais incidiram aquele imposto (ICMS ou ISS) no respectivo mês, considerando as operações e prestações incluídas no programa;

II) o valor limite para rateio entre os consumidores com identificação no documento fiscal é de 30% do valor recolhido pelo contribuinte para o mês da emissão.

O valor do crédito é obtido mediante a multiplicação da fração descrita no item I pelo valor limite para rateio descrito no item II.

O cálculo do crédito considera, ainda:

a) o valor limite de crédito por documento fiscal é de 7,5%, se tributado pelo ICMS, e de 1,5%, se pelo ISS;

b) o total de vendas descrito no item I e os recolhimentos citados no item II são para o CNPJ raiz do estabelecimento emitente;

c) a fração descrita no item I é para os documentos fiscais emitidos com a identificação do consumidor (CPF/CNPJ) que forem corretamente declarados nos documentos fiscais até a consolidação;

d) o valor do crédito não poderá ultrapassar 30% do imposto incidente para a operação, quando declarado pelo contribuinte.

O percentual de recolhimento de ICMS/ISS utilizado no cálculo do crédito de documento fiscal emitido a partir de dezembro de 2012 observa o Fator de Multiplicação para o Cálculo do Crédito (FMCC), de acordo com a atividade econômica preponderante (CNAE) do contribuinte, na forma estabelecida em [19].

Empresas optantes pelo Simples Nacional terão o FMCC igual a 1 para o cálculo do crédito relativo às suas operações e prestações, ou seja, observará o teto de 30% do ICMS/ISS recolhido até a consolidação.

Se o documento fiscal emitido com identificação do consumidor para o período consolidado for corretamente declarado pela empresa após a data de fechamento do cálculo, somente poderá gerar crédito se houver reclamação do consumidor em relação ao mesmo, estando pendente de conclusão.

Nessa hipótese, bem como para as reclamações protocolizadas pelo consumidor para análise da SEF e consideradas como procedente pelo Fisco, o crédito será atribuído mediante o uso do Índice Médio de Crédito (IMC) do imposto incidente sobre a operação realizada, condicionado ao recolhimento das receitas consideradas no cálculo para o mês de emissão do documento fiscal.

O IMC é calculado por tributo (ICMS e ISS), após a data de consolidação, correspondendo à razão entre os totais de créditos disponibilizados para determinado mês e os valores dos documentos fiscais para os quais foram realizados os cálculos:

$$IMC(In) = \frac{TC(In)}{TD(In)}$$

Sendo que:

- IMC (In) = Índice Médio de Créditos referente ao imposto (ICMS ou ISS), no mês de referência;
- TC (In) = valor total de créditos calculados referente ao imposto (ICMS ou ISS), de todos os contribuintes participantes, no mês de referência;
- TD (In) = valor total dos documentos fiscais de ICMS ou de ISS de todos os contribuintes participantes para o mês de referência, com identificação de CPF ou CNPJ do beneficiário, considerando as operações e prestações incluídas no PNL.

O índice é utilizado na concessão de crédito para documento fiscal objeto de reclamação do beneficiário, após a consolidação do cálculo, condicionado ao recolhimento do tributo pela empresa para o mês de sua emissão, nas seguintes situações: reclamação concluída pela SEF/DF com análise pela sua procedência; documento fiscal regularizado pela empresa antes da conclusão pelo Fisco.

O cálculo do crédito, nessas hipóteses, é obtido pela multiplicação do valor do documento fiscal pelo IMC do mês de sua emissão, observado o imposto correspondente.

Os créditos de reclamações analisadas como procedentes pelo Fisco, de período consolidado e em que haja o recolhimento do tributo pelo contribuinte, terão o crédito pelo IMC anteriormente à lavratura do auto de infração, permanecendo em uma situação transitória até a finalização do procedimento de fiscalização com a situação "Concluída pelo Fisco".

Caso ocorra a anterior regularização do documento fiscal pelo contribuinte, a reclamação será tramitada para a situação "Concluída pelo Sistema".

3.4 Livro Fiscal Eletrônico

A escrituração dos Livros Fiscais previstos no regulamento do ICMS e do ISS foi substituída em 2006 pela escrituração eletrônica do Livro Fiscal Eletrônico, conforme leiaute previsto em [7]. A escrituração no LFE foi determinada em [10] e regulamentada em [11].

Todo contribuinte do ICMS e/ou do ISS no DF está obrigado a escrituração do Livro Fiscal Eletrônico, salvo contribuintes enquadrados no Simples Nacional com faturamento anual inferior ao limite estabelecido para os Microempreendedores Individuais (para 2012, o valor é de R$ 60.000,00).

As informações devem ser geradas em arquivo no formato TXT, por aplicativo de responsabilidade do contribuinte. Existem diversos aplicativos de escrituração fiscal que geram LFE no mercado. Cabe ao contribuinte escolher o que melhor se adapta às suas necessidades. Além de não disponibilizar nenhum aplicativo para escrituração, a SEF também não se responsabiliza pela homologação dos aplicativos existentes no mercado.

3.5 Sigilo Fiscal e Funcional

Os dados utilizados nesta pesquisa foram obtidos da extração do banco de dados de produção do PNL da SEF e representam informações protegidas pelo sigilo fiscal.

O sigilo fiscal tem como objetivo garantir a privacidade do cidadão, impedindo que as informações confiadas ao Fisco, isto é, órgãos da Administração Pública encarregados da arrecadação de tributos e da fiscalização dos contribuintes, sejam divulgadas e possam causar prejuízos ao cidadão.

O sigilo fiscal é um direito fundamental do cidadão. O Art. 5º da Constituição Federal, em [4], garante aos brasileiros e estrangeiros residentes no país o direito à inviolabilidade da intimidade, da vida privada, da honra e da imagem das pessoas. Nestes casos é assegurado o direito à indenização pelo dano material ou moral decorrente de sua violação.

Neste sentido, o Código Tributário Nacional (CTN) [3], em seu Artigo 198, estabelece que qualquer informação obtida em razão do ofício sobre a situação econômica ou financeira de um contribuinte não pode ser divulgada.

Para garantir o sigilo fiscal nesta pesquisa, os contribuintes e beneficiários não serão identificados e as informações declaradas não serão identificadas.

O sigilo funcional visa a garantir que as informações e processos sigilosos da organização sejam preservados. Neste sentido, as informações sobre o trâmite interno da SEF não serão divulgadas.

O próximo capítulo apresenta uma metodologia para mineração de dados que é referência de mercado, o CRISP-DM. A partir da customização de suas fases foi desenvolvida esta pesquisa.

Capítulo 4

Metodologia de Mineração de Dados

Esta pesquisa apresenta uma iniciativa da SEF pioneira ao analisar os dados existentes nas bases do PNL usando mineração de dados. Conforme apresentado na Seção 1.3, o uso de mineração de dados tem sido promissor no estudo de problemas de mesma natureza apresentada. A Seção 4.1 apresenta a metodologia de referência em mineração de dados, o CRISP-DM [40].

4.1 Modelo de Referência CRISP-DM

Para guiar o trabalho de mineração de dados utilizou-se a metodologia CRISP-DM. O CRISP-DM (*CRoss Industry Standard Process for Data Mining*) destaca-se como um padrão de fato em mineração de dados, principalmente por não ser proprietário e pretender ser independente do setor e das aplicações em que é utilizado.

De acordo com [35], entre 2006 e 2008, um grupo de interesse especial para a versão 2.0 foi formado para discussões e atualizações do modelo CRISP-DM. Não houve conclusão deste trabalho até o momento. Desta forma, será utilizada a versão 1.0 que está consolidada desde o ano 2000. O CRISP-DM envolve um ciclo composto por seis fases para um projeto de mineração de dados. São elas:

- Entendimento do negócio: esta fase parte do entendimento dos objetivos do projeto e requisitos de uma perspectiva negocial, para então converter este conhecimento em uma definição de projeto de mineração de dados e um plano preliminar para alcançar os objetivos;
- Compreensão dos dados: a fase de entendimento dos dados começa com o obtenção de um conjunto de dados inicial e procede com atividades para familiarização com os dados, identificar problemas de qualidade dos dados e ter noções de como o conjunto de dados se compõe;

- Preparação dos dados: a fase de preparação dos dados cobre todas as atividades para construir o conjunto de dados a partir dos dados iniciais coletados, com foco em seleção de atributos, limpeza, construção, integração e formatação dos dados de entrada;
- Modelagem: várias técnicas de modelagem são selecionadas e aplicadas. Seus parâmetros são calibrados para obtenção de valores ótimos;
- Avaliação: os modelos obtidos são avaliados e os passos executados para construir os modelos são revisados para garantir que alcancem os objetivos negociais;
- Aplicação: o conhecimento obtido com os modelos gerados deve ser aplicado na organização e este conhecimento deve ser disseminado e apresentado para os usuários de uma forma que eles possam usá-lo.

A Figura 4.1, adaptada de [40], apresenta como as fases se relacionam ao longo do tempo. Apesar de ter fases bem definidas, a sequência de passos entre as fases do CRISP-DM não é linear, apresentando ciclos e retornos, o que o torna mais flexível. É esperado que haja interação entre as fases conforme seja necessário corrigir ou ajustar o processo. A seta de saída de uma fase indica o início da próxima. As setas internas indicam as dependências mais importantes e frequentes entre as fases. O círculo de setas externo simboliza a natureza cíclica do processo de mineração de dados. Lições aprendidas durante o processo de mineração de dados e da aplicação entregue podem disparar novas questões negociais a serem avaliadas.

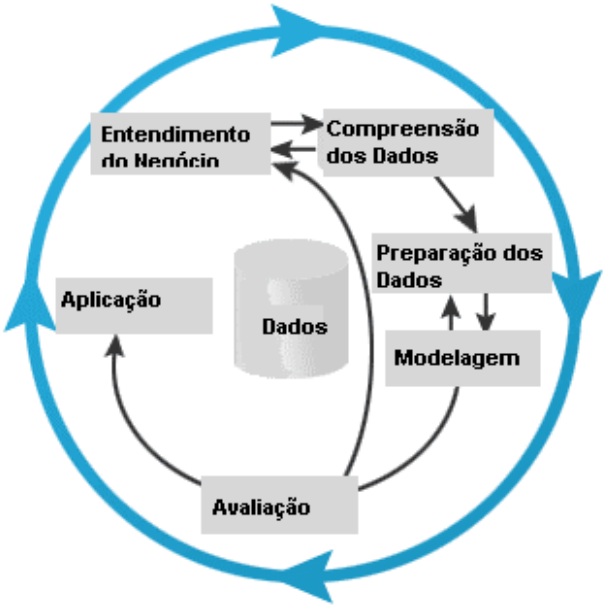

Figura 4.1: Fases do CRISP-DM

Capítulo 5

Entendimento do Negócio

Parte do entendimento do negócio foi abordado no Capítulo 3 ao apresentar características do funcionamento da SEF e do PNL. A Seção 5.1 apresenta outros aspectos relevantes para o entendimento do negócio. A Seção 5.2 apresenta o planejamento para obtenção dos objetivos da pesquisa.

5.1 Definições

Uma vez definido o objetivo do uso de mineração de dados para extração dos perfis dos participantes do PNL, o próximo passo baseou-se em dois aspectos de perfis para estudo: obtenção de créditos de consumo pelos beneficiários e fidelidade das pessoas físicas ao PNL.

A análise de fidelidade para pessoas jurídicas não foi contemplada por esta pesquisa, uma vez que, na prática, houve pouca adesão dos beneficiários que são pessoas jurídica ao PNL, cerca de 1,5% do total de documentos fiscais. Como o uso dos créditos é limitado à compensação do IPVA e IPTU de bens de titularidade das empresas, há apenas uma pequena fração de empresas do Distrito Federal que são optantes do Simples Nacional e ainda possuem a titularidade de bens elegíveis ao PNL.

Os dois perfis selecionados para estudo constituem dificuldades relatadas pelos especialistas da SEF. Em momentos anteriores já houve questionamentos de órgãos de controle envolvendo questões sobre os perfis e comportamento dos beneficiários. Informações pontuais foram relatadas à época mas um panorama geral envolvendo os perfis não tiveram progresso posterior.

Nos dois casos, os perfis de estudo tem como base o documento fiscal, sendo este conhecido pela população em geral como o cupom fiscal emitido em aquisições de mercadorias e a nota fiscal que é solicitada geralmente em aquisições mais onerosas e na obtenção de serviços.

Apesar de possuir um banco de dados de mais de 150.000.000 de documentos fiscais com dados de empresas e de contribuintes, este é utilizado primariamente para funções como: cálculo dos benefícios do programa, envio de e-mails, iteração com a aplicação web. Desta forma, o potencial de análise de sua base de dados ainda é pouco explorado.

A origem dos dados do PNL envolvendo os documentos fiscais baseia-se em outros sistemas da SEF. Estes sistemas recebem o Livro Fiscal Eletrônico (LFE), apresentado na Seção 3.4, gerado pelos contadores das empresas mensalmente. Os documentos fiscais que são identificados com o CPF/CNPJ do consumidor, são processados, têm seus tipos de dados validados e suas principais informações inseridas no banco de dados do PNL.

Informações que podem relacionar o beneficiário (pessoa física ou jurídica) ou o contribuinte do PNL com o documento fiscal precisam ser obtidas de outras fontes. Em conjunto com os especialistas da SEF foi acordado o uso de dados obtidos da Receita Federal para o beneficiário pessoa física; e dados de extração do cadastro fiscal de empresas, pertencente à SEF, para o beneficiário pessoa jurídica e contribuintes.

Conforme será detalhado nos Capítulos 6 e 7, as informações relacionadas à endereços postais nos bancos de dados se apresentaram deficientes. Para suprir esta carência recomendou-se o uso do banco de dados dos Correios. Embora estas informações não estivessem disponíveis em banco de dados da SEF, mediante convênio existente, as informações foram solicitadas e processadas.

Com estas definições, foi disponibilizado acesso irrestrito aos dados de produção do PNL e servidor de alta performance com 80 núcleos e 1 TB de memória RAM para processamento das informações e testes. Devido ao sigilo fiscal, dados sensíveis de beneficiários e contribuintes foram mascarados e respeitados; e os dados do PNL não foram retirados da SEF.

5.2 Planejamento de Mineração

Junto aos especialistas da SEF, para alcance do levantamento dos perfis desejados, objetivou-se, a partir das informações dos documentos fiscais, a inclusão de dados que descrevessem pessoas físicas e jurídicas. Realizando o cruzamento entre os bancos de dados, poderiam ser obtidas informações que pudessem descrever de forma clara padrões e comportamentos.

No contexto de mineração de dados, verificou-se que determinadas tarefas seriam adequadas para determinação dos perfis. Por conta da limitação do tempo para conclusão da pesquisa e para obtenção de resultados concretos, o escopo deste projeto foi dividido em duas fases:

- Na fase 1, que se trata da pesquisa atual, seguindo o CRISP-DM, a extração dos bancos de dados deveriam ser selecionados, processados e transformados em bases de dados prontas para mineração de dados. Para modelagem poderiam ser utilizadas técnicas de sumarização e de visualização nas bases de dados. A avaliação dos padrões explicitados poderia ser realizada junto aos especialistas da SEF.
- Na fase 2, em uma próxima pesquisa, pretende-se utilizar as bases de dados da fase 1, usando técnicas de associação e de *clustering* com o *framework* WEKA.

No que tange a modelagem na fase 1, para sumarização poderia ser utilizada análise estatística e para a visualização seria possível a utilização do software Qlikview[1] para obtenção de painéis de informações em suporte aos dados sumarizados.

Para execução da fase 1, decidiu-se extrair os documentos fiscais do banco de dados do PNL já processados e validados desde seu início até o final de 2013. Ou seja, documentos fiscais inválidos ou não processados, com créditos ainda não calculados, deveriam ser descartados.

Seriam geradas duas extrações de dados. A primeira que, através de comandos de Linguagem de Consulta Estruturada, tradução para *Structured Query Language* (SQL), faria a extração das informações de documentos fiscais em conjunto com informações de pessoas físicas. A outra extração de dados seria feita com os dados disponíveis de pessoas jurídicas. Evitou-se uma única extração de dados devido à grande quantidade de informações e dificuldade posterior de processamento para eliminar a redundância dos dados.

De posse dos dados, deveriam ser removidos campos que não agregassem informações à pesquisa. Objetivou-se a inclusão de variáveis que descrevessem idade das pessoas físicas, as atividades econômicas existentes no PNL para as pessoas jurídicas e informações sobre endereços postais para pessoas físicas e jurídicas.

O próximo passo avaliaria a qualidade das informações existentes, com correção de dados onde aplicável, e remoção dos dados em casos sem solução. Possíveis casos de *outliers* já eram presumidos, uma vez que a validação do LFE se restringe ao tipo de dados sem realizar críticas aos valores das variáveis.

Com estas informações validadas seria possível a junção das duas bases de dados e extração de estatísticas descritivas sobre cada uma das variáveis a serem analisadas.

Para avaliação do perfil de créditos de consumo, a quantidade de créditos obtida por documento fiscal poderia ser agrupada em faixas de crédito que descrevessem comportamentos similares das pessoas. Variáveis escolhidas que descrevessem a obtenção de créditos de consumo deveriam ser tabuladas duas a duas, em tabelas com sua probabili-

[1] (www.qlik.com)

dade de ocorrência, ao longo das faixas de crédito. Para cada tabela gerada deveria ser criado seu gráfico correspondente. Com tabelas e gráficos sobre as variáveis, seria possível extrair informações sobre os padrões observados. Estas faixas poderiam ser analisadas por *clustering* na próxima fase.

Para avaliação do perfil de fidelidade, deveria ser levado em conta a variação do tempo sobre as pessoas físicas que participaram do PNL. Desta forma poderiam ser analisados comportamentos em períodos de tempo. Da mesma forma que no perfil de créditos de consumo, variáveis escolhidas que descrevessem a fidelidade dos consumidores deveriam ser tabuladas duas a duas, em tabelas com sua probabilidade de ocorrência, ao longo de faixas de fidelidade que fossem encontradas. Para cada tabela gerada deveria ser criado seu gráfico correspondente. Com tabelas e gráficos sobre as variáveis seria possível extrair informações sobre os padrões observados. Ainda sobre o perfil de fidelidade, deveriam ser propostos indicadores para avaliação do PNL. Uma vez que estes indicadores estivessem definidos, poderiam ser extraídas informações de seu desempenho ao longo das faixas de fidelidade. Estas faixas poderiam ser analisadas por associação na próxima fase.

Capítulo 6

Compreensão dos Dados

Com o objetivo de determinar os padrões de consumo e de fidelidade dos beneficiários, foi feito o levantamento das informações no banco de dados do PNL, no cadastro fiscal de empresas e nos dados de pessoas físicas disponibilizados pela RFB. A Seção 6.1 apresenta como os dados foram originalmente extraídos e as descrições de seus campos. A Seção 6.2 apresenta remoção de campos que não estão alinhados ao objetivo da pesquisa. A Seção 6.3 analisa problemas de qualidade e suas soluções. A Seção 6.4 analisa, via estatística descritiva, os campos selecionados.

A sequência dos passos desta fase não foram rigidamente seguidos conforme o CRISP-DM. Os passos de Seleção de Dados e de Limpeza de Dados da fase de Preparação dos Dados foram realizadas em conjunto à fase de Compreensão dos Dados.

Para maior clareza textual, neste Capítulo e nos seguintes, consumidores equivalem aos beneficiários e empresas equivalem aos contribuintes do PNL.

6.1 Coleta de dados inicial

Os documentos fiscais emitidos pelos contribuintes do PNL são armazenados na tabela Documento_Fiscal.

A Tabela 6.1 apresenta as descrições dos campos presentes na tabela Documento_Fiscal e seu tipo de dado.

As tabelas Mensagem e Livro_Processado são vinculadas à Documento_Fiscal. Enquanto Mensagem não acrescenta informações que descrevem o consumo dos usuários do PNL; através da tabela Livro_Processado é possível extrair o CNPJ da empresa vinculada ao documento fiscal emitido. As outras informações existentes na tabela Livro_Processado são referentes a controles de processamento dos dados enviados pelos contadores das empresas à SEF, não descrevendo o consumo dos usuários do PNL.

Tabela 6.1: Campos Tabela Documento_Fiscal

Tipo de Dado	**Campo**	**Descrição**
Numérico	Seq_doc_fiscal	Chave primária tabela
Numérico	Seq_livro_processado	Chave estrangeira tabela Livro_Processado
Numérico	Num_doc_fiscal	Número do documento fiscal
Data	Dta_emissao_doc_fiscal	Data de emissão
Data	Dta_registro_doc_fiscal	Data de registro no sistema
Numérico	Val_doc_fiscal	Valor emitido
Nominal	Tipo_registro_documento_fiscal	Tipo do documento fiscal
Nominal	Cnpj_destinatario	CNPJ
Nominal	Cod_mod_doc_fiscal	Modelo
Nominal	Cpf_destinatario	CPF
Nominal	Serie_doc_fiscal	Série
Numérico	Val_bc_iss	Valor da base de cálculo do ISS
Numérico	Val_bc_icms	Valor da base de cálculo do ICMS
Numérico	Val_iss	Valor declarado de ISS
Numérico	Val_icms	Valor declarado de ICMS
Data	Data_hora_incl	Auditoria - data/hora de inclusão do registro
Nominal	Usuario_altr	Auditoria - usuário de inclusão
Nominal	Usuario_incl	Auditoria - usuário de alteração
Data	Data_hora_altr	Auditoria - data/hora de alteração
Numérico	Val_credito	Valor do crédito do PNL calculado
Nominal	Sit_calculo	Situação do cálculo do crédito
Numérico	Seq_mensagem	Chave estrangeira tabela Mensagem
Numérico	Val_cal_credito	Não usado
Numérico	Val_st	Não usado
Numérico	Cfop	Código fiscal de operações e prestações, conforme [8]
Numérico	Cfps	Código fiscal de prestação de serviço, conforme [8]

A Tabela 6.2 apresenta as descrições dos campos presentes no cadastro de pessoas físicas e seu tipo de dado. Tratam-se das informações repassadas pela RFB à SEF conforme convênio existente.

A Tabela 6.3 apresenta as descrições das informações referentes ao cadastro fiscal de empresas da SEF. Embora o cadastro fiscal da SEF seja mais amplo com informações detalhadas sobre faturamento e composição societária das empresas, decidiu-se que estas informações não seriam abrangidas pelo escopo da pesquisa e por serem sigilosas não serão detalhadas.

Com base nestas informações, foram extraídos dois conjuntos de dados. O primeiro deu origem a um arquivo texto contendo extração conjunta de Documento_Fiscal e de Pessoa_Fisica relacionados pelo campo CPF que é comum às duas tabelas. Foi incluído, a esta primeira base de dados, o campo nominal CNPJ extraído do relacionamento do campo seq_livro_processado com a tabela Livro_Processado. Este campo apresenta o CNPJ do contribuinte que declarou o documento fiscal. A partir dele seria possível relacionar as informações com a tabela Cadastro_Fiscal. Foram extraídos os dados plenamente processados pelo PNL no período de 16/09/2008 a 31/12/2013.

O segundo conjunto de dados armazenou as informações referentes ao cadastro fiscal de empresas em um segundo arquivo texto.

Estes arquivos geraram em torno de 30 GB de informações.

6.2 Descrição dos dados

Já em uma primeira análise das informações extraídas, foi possível perceber que alguns campos não seriam de ajuda para descrever os perfis desejados. Desta forma, foram removidas as seguintes informações dos dados extraídos:

- Dados de auditoria: Data_hora_incl, Usuario_altr, Usuario_incl, Data_hora_altr, usuario_incl, usuario_altr, data_hora_incl, data_hora_altr;
- *Constraints* do banco de dados: Seq_doc_fiscal, Seq_livro_processado, Seq_ mensagem;
- Campos derivados/calculados: Val_bc_iss, Val_bc_icms. Em breve explanação podemos dizer que base de cálculo é a grandeza econômica (valor) sobre a qual se aplica a alíquota para calcular o imposto a pagar. O Decreto 25.508/2005, disponível em [9], define no Art. 27, que base de cálculo do imposto é o preço do serviço. Como os valores calculados do ISS e do ICMS no documento fiscal já estão presentes na tabela, as bases de cálculo não agregam informações. Estes campos são utilizados pela SEF para conferência dos cálculos e auditoria;

Tabela 6.2: Campos Tabela Pessoa_Fisica

Tipo de Dado	**Campo**	**Descrição**
Nominal	Cpf_dados	CPF
Nominal	Nom_cpf_dados	Nome completo
Numérico	Sit_cadastral_dados	Indicador de situação
Numérico	Ind_residente_ext_dados	Indicador residente no exterior
Numérico	Cod_pais_ext_dados	Código país exterior
Nominal	Nom_pais_ext_dados	Nome país exterior
Nominal	Nom_mae_cpf_dados	Nome da mãe
Numérico	Dat_nasc_cpf_dados	Data de nascimento
Numérico	Sex_cpf_dados	Sexo
Numérico	Nat_ocupacao_dados	Natureza da ocupação
Numérico	Cod_ocupacao_princ_dados	Código ocupação
Numérico	Ano_exerc_ocupacao_dados	Ano da ocupação
Nominal	End_tip_logradouro_dados	Endereço - Tipo de logradouro
Nominal	End_logradouro_dados	Endereço - Logradouro
Nominal	End_num_logradouro_dados	Endereço - Número logradouro
Nominal	End_complemento_dados	Endereço - Complemento
Nominal	End_bairro_dados	Endereço - Bairro
Numérico	End_cep_dados	Endereço - Código de Endereçamento Postal - CEP
Nominal	End_uf_dados	Endereço - UF
Nominal	End_cod_municipio_dados	Endereço - Código do município
Numérico	Tel_ddd_dados	Telefone - DDD
Numérico	Tel_dados	Telefone
Numérico	Cod_unidade_adm_dados	Unidade Organizacional
Numérico	Ano_obito_dados	Ano de óbito
Numérico	Ind_estrangeiro	Indicador estrangeiro
Numérico	Num_titulo_eleitor_dados	Título de eleitor
Nominal	Usuario_incl	Auditoria - usuário de inclusão
Nominal	Usuario_altr	Auditoria - usuário de alteração
Data	Data_hora_incl	Auditoria - data/hora de inclusão
Data	Data_hora_altr	Auditoria - data/hora de alteração
Nominal	Nom_municipio_dados	Município
Nominal	Virtual_cpf_digito	Dígito verificador CPF
Nominal	Virtual_cpf_por_estado	Não usado

Tabela 6.3: Campos Tabela Cadastro_Fiscal

Tipo de Dado	**Campo**	**Descrição**
Numérico	inscricao	Inscrição da empresa perante a SEF
Numérico	CNPJ	CNPJ empresa
Nominal	razao_social	Razão social
Nominal	nome_fantasia	Nome fantasia
Nominal	descricao	Endereço - Descrição
Nominal	Bairro	Endereço - Bairro
Nominal	Cidade	Endereço - Cidade
Nominal	UF	Endereço - UF
Numérico	cep	Endereço - Código de Endereçamento Postal

- Identificadores únicos da tabela de cadastro fiscal das empresas: inscricao, razao_social, nome_fantasia. Removidos uma vez que o CNPJ é o identificador principal dos dados;
- Identificadores únicos da tabela de documentos fiscais: Num_doc_fiscal, Dta_ registro _doc_fiscal. Removidos uma vez que o CNPJ é o identificador principal dos dados;
- Identificadores únicos da tabela de pessoa física: num_titulo_eleitor_dados. Removido uma vez que o CPF é o identificador principal dos dados;
- Campos com uma única informação em todos os registros: sit_cadastral_dados, Sit_calculo, virtual_cpf_digito, virtual_cpf_ por_estado. Desta forma, os valores de cada um destes campos são unimodais com a moda representando 100% da base de dados;
- Campos com valores nulos em todos os registros: Cfop, Cfps, Val_cal_credito, Val_st;
- Informações técnicas: Tip_registro_documento_fiscal, Cod_mod_doc_fiscal, Serie_doc_fiscal. Informações utilizadas para identificar o tipo, modelo e formato do documento fiscal. Estas informações não descrevem os perfis desejados;
- Dados não relacionados ao perfis desejados: ind_residente_ext_dados, cod_pais _ext_dados, nom_pais_ext_dados, nom_mae_cpf_dados, nat_ocupacao _dados, cod_ocupacao_princ_dados, ano_exerc_ocupacao_dados, tel _ddd _dados, tel_dados, cod_unidade_adm_dados, ind_estrangeiro. Junto a um especialista da SEF, são se identificou possível contribuição destes campos para definição dos perfis desejados.

6.3 Verificação da qualidade dos dados

Embora os dados de documentos fiscais já tivessem sido previamente processados e validados, os dados das outras tabelas não foram. Desta forma, foram verificados os aspectos de ruído, *missing values* e formato de dados durante a determinação da qualidade dos dados.

6.3.1 Ruído

Foram verificados problemas de consistência nos dados que envolvem endereço postal, tanto no cadastro fiscal quanto nos dados da Receita Federal. Como exemplo, o fato de existirem 121 cidades do DF cadastradas na tabela de cadastro fiscal e o fato de terem sido encontrados 17842 bairros no DF na tabela da RFB.

Os campos de bairro, cidade e UF apresentaram informações inconsistentes entre si. Foram encontradas cidades de um Estado cadastrados para outro Estado e bairros de uma cidade cadastradas para cidades distintas.

A Figura 6.1 ilustra algumas situações de ruído. Na lista da esquerda, referente aos dados de bairro coletados junto ao Cadastro Fiscal, são apresentadas as diversas variações para o bairro "Samambaia". Na lista do centro, referente aos dados de cidades coletadas junto ao Cadastro Fiscal, são apresentadas as diversas variações para a cidade "Brasília". Na lista da direita, referente aos dados de bairros coletados junto aos dados da Receita Federal, são apresentadas as diversas variações para o bairro "Lago Sul".

Como os dados de endereço apresentaram dificuldades para sua utilização, optou-se em utilizar os campos de CEP para levantamento dos endereços junto aos Correios uma vez que os campos de CEP foram os que apresentaram menos problemas de qualidade de dados e a SEF possui convênio com os Correios para utilização de sua base de dados.

Desta forma, foram removidos os seguintes campos: End_tip_logradouro_dados, End_logradouro_dados, End_num_logradouro_dados, End_complemento_dados, End_bairro_dados, End_cep_dados, End_uf_dados, End_cod_municipio_dados.

6.3.2 *Missing Values*

No campo Sex_cpf_dados foram detectados em torno de 28.000 pessoas físicas sem definição de sexo.

Para tratamento deste campo utilizou-se procedimentos da Previdência Social (INSS) para determinação de sexo. Baseando-se em uma base de dados com 1.747.422 registros que relaciona unicamente primeiros nomes de pessoas e o sexo correspondente, via programação, foi extraído do campo Nom_cpf_dados o primeiro nome de cada pessoa sem

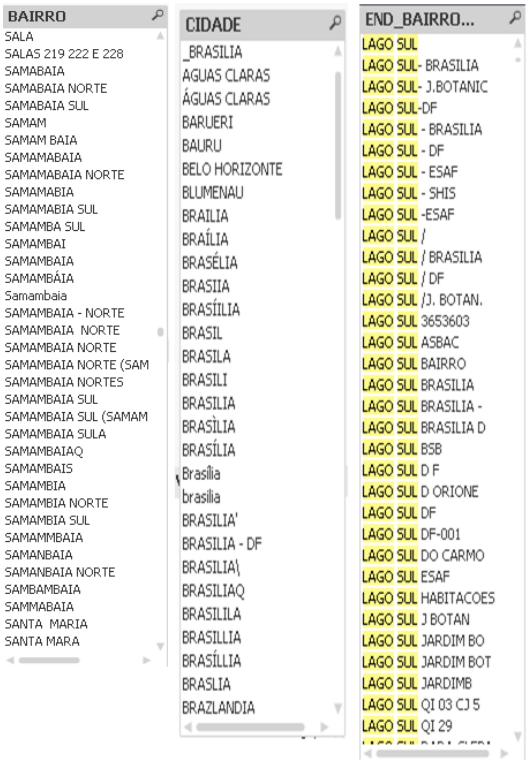

Figura 6.1: Ruído nos Dados

definição de sexo de Pessoa_Fisica e realizado comparação com os nomes existentes. Para cada caso em que o nome foi encontrado, o sexo existente foi atualizado. Após a comparação de todos os nomes, apenas 11 pessoas ainda ficaram sem definição. Os registros associados a estes nomes foram removidos da base de dados.

Com a resolução dos *missing values*, o campo Nom_cpf_dados foi descartado, já que o nome de pessoa física trata de informação sigilosa.

6.3.3 Formato dos Dados

O campo Dat_nasc_cpf_dados inicialmente foi definido como numérico na base de dados com formato AAAAMMDD, embora seja uma informação de data. Os campos que continham números flutuantes (Val_doc_fiscal, Val_iss, Val_icms e Val_credito) durante a extração apresentaram padrão com os milhares separados com o caractere ',' e o caractere '.' para separação fracionária.

Nos dois casos o problema foi resolvido via programação. Houve correção para o tipo de dados data no primeiro caso e remoção dos caracteres ',' nos campos de número.

6.4 Exploração dos dados

O objetivo da exploração dos dados é extrair os primeiros conhecimentos e ajudar a conferir se os dados são consistentes e se não houve erro de carga. Além disso, a estatística descritiva pode ajudar a entender o comportamento dos dados.

Para cada um dos campos das Tabelas 6.4 e 6.5 são apresentadas estatísticas de quantidade de valores distintos e moda para campos nominais e data. Mínimo, máximo, média, desvio padrão e moda para campos numéricos. A informação sobre moda apresenta o valor da moda encontrado, quantidade de repetições da moda e a porcentagem que estas repetições representam do total da base de dados.

A Tabela 6.4 apresenta informações sobre os campos que envolvem o arquivo texto de contribuintes.

O valor da moda do CNPJ é uma informação sigilosa e não foi apresentada propositalmente. No campo CEP, o valor de maior incidência para a moda é um *missing value*. Não foram constatados meios para resolução deste problema de qualidade de dados.

Tabela 6.4: Análise Base de Dados de Contribuintes

Campo	**Descrição**
CNPJ	Distintos = 94.945, Moda com ocorrência de 2.773.740 documentos fiscais para a moda, representando 2% da base de dados
cep	Distintos = 15.309, Moda = " ", com ocorrência de 4.915 documentos fiscais para a moda, representando 4,4% da base de dados

Ao analisar a Tabela 6.5, em que foram carregados 157.192.269 registros dos campos do arquivo texto de beneficiários, as informações que se seguem foram objetos de análise.

Embora a extração tenha sido feita até 31/12/2013, conforme planejamento relatado na Seção 5.2, a quantidade de documentos fiscais processados do PNL só estavam disponíveis até 31/10/2013.

Os valores de máximo e de desvio padrão em Val_doc_fiscal causaram suspeita de erro ao carregar as informações. Mas houve conferência junto ao banco de dados e os valores deste campo estavam cadastrados corretamente. Verificou-se ser necessária análise de *outliers* em um momento posterior.

Os valores Val_doc_fiscal, Val_icms e Val_iss são declarados pelas empresas ao enviarem o LFE para SEF mensalmente. Embora tenham seus valores validados pelo LFE, o cálculo do imposto referente ao valor do documento fiscal é realizado pela empresa que o declarou e não pela SEF. Sendo assim, o valor de Val_credito é calculado conforme Seção 3.3.2 durante a carga de dados do LFE para o PNL. Uma vez que os outros parâmetros

Tabela 6.5: Análise da Base de Dados de Beneficiários

Campo	**Descrição**
Dta_emissao_doc_fiscal	Datas entre 16/09/2008 a 31/10/2013 Distintos = 1.871 Moda = 10/08/2013, com ocorrência de 328.010 documentos fiscais para a moda, representando 0,2% da base de dados
Val_doc_fiscal	Mínimo = 0,01 Máximo= 9.003.600.097,03 Média= 336,68 Desvio padrão = 1.247.483,03 Moda = 5, com ocorrência de 646.195 documentos fiscais para a moda, representando 0,4% da base de dados
Cpf_destinatario	Distintos = 830.156 Moda com ocorrência de 246.022 documentos fiscais, representando 0,2% da base de dados
Cnpj_destinatario	Distintos = 37.305 Moda com ocorrência de 6.034 documentos fiscais, representando menos de 0,001% da base de dados
Val_iss	Mínimo=0 Máximo=50.000 Média= 0,38 Desvio padrão = 12,28 Moda= 0, com ocorrência de 147.566.777 documentos fiscais para a moda, representando 93,9% da base de dados
Val_icms	Mínimo=0 Máximo= 701.880,00 Média= 5.455,49 Desvio padrão= 647.917,81 Moda =0, com ocorrência de 36.585.707 documentos fiscais para a moda, representando 23,3% da base de dados
Val_credito	Mínimo= 0 Máximo= 144.237,12 Média= 2,21 Desvio padrão = 22,06 Moda = 0, com ocorrência de 16.956.211 documentos fiscais para a moda, representando 10,8% da base de dados
Sex_cpf_dados	Distintos = 2, Moda = 2-feminino, com ocorrência de 80.433.006 documentos fiscais para a moda, representando 51,2% da base de dados
Dat_nasc_cpf_dados	Distintos = 26.183, Moda = "13/09/1982", com ocorrência de 262.263 documentos fiscais para a moda, representando 0,2% da base de dados
CNPJ	Descrição na Tabela 6.4
End_cep_dados	Distintos = 66.608, Moda = "70000000", com ocorrência de 5.159.500 documentos fiscais para a moda, representando 1,8% da base de dados

para o cálculo dos créditos não estavam presentes no banco de dados do PNL, se assumiu que os valores calculados encontram-se corretos.

O valor da moda do Cpf_destinatario e Cnpj_destinatario são informações sigilosas e não foram apresentadas propositalmente.

A quantidade de documentos fiscais associados à moda do Cpf_destinatario causou suspeita de erro ao carregar as informações. Isso implicaria dizer que uma pessoa física emitiu em média mais de 100 documentos fiscais por dia durante a vigência do PNL. Mas houve conferência junto ao banco de dados e os valores deste campo estavam cadastrados corretamente. Verificou-se ser necessária análise de *outliers* em momento posterior.

Ao analisar as informações sobre Cpf_destinatario e Cnpj_destinatario constatou-se que abrangem populações diferentes. Pessoas físicas e pessoas jurídicas tem tratamento diferenciado na lei que regulamenta o PNL. Então o comportamento esperado para estes dois perfis deve ser pesquisado separadamente.

Ao verificar a grande quantidade de valores 0 para a moda de Val_iss e o Val_icms também constatou-se que abrangem populações diferentes. A tributação para a aquisição de bens é realizada via ICMS, enquanto a tributação de serviços é realizada via ISS. Quando há valor declarado para um imposto, o campo do outro imposto é preenchido com 0.

Há equilíbrio entre os dois sexos, sendo que há predominância do sexo feminino. Enquanto o sexo feminino é responsável por 51,2% dos documentos fiscais, o sexo masculino obteve 48,8% dos documentos fiscais restantes. A Figura 6.2 apresenta a distribuição dos sexos.

Justifica-se o valor para a moda de End_cep_dados, uma vez que este é o CEP padrão de Brasília e a base de dados revela esta realidade. Outros CEPs genéricos analisados não apresentaram tanta frequência quanto este.

Desta forma, foram apuradas 12 variáveis a serem pesquisadas; sendo 6 nominais, 4 numéricas e 2 de data; em torno de 157.000.000,00 registros; 95.000 empresas; e 830.000 pessoas físicas.

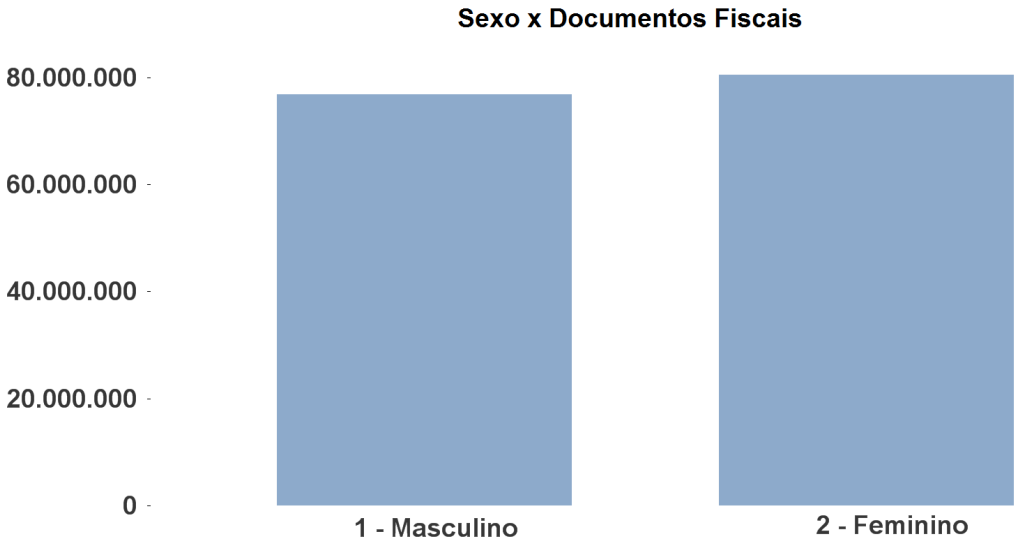

Figura 6.2: Distribuição Sexo

Capítulo 7

Preparação dos Dados

Nesta fase do CRISP-DM objetivou-se a adequação das bases de dados à lógica negocial do PNL. Para o adequado estudo dos perfis, foi proposta inclusão de informações de idade, atividade econômica e regiões administrativas (RA) que poderiam ser derivadas dos dados atuais. A Seção 7.1 apresenta inclusão/remoção de variáveis e suas razões. A Seção 7.2 apresenta uma tentativa de resolução para o problema de qualidade dos endereços. A Seção 7.3 inclui à base de dados as variáveis solicitadas pelos especialistas da SEF. A Seção 7.4 apresenta a junção das bases de dados existentes e estatística descritiva sobre as novas variáveis. A Seção 7.5 ajusta a base de dados aos objetivos da pesquisa.

7.1 Seleção de Dados

Parte do trabalho de seleção de dados foi apresentado no capítulo anterior nas Seções 6.2 e 6.3.

Em [15] foi possível verificar que no início do PNL, em 2008, poucas atividades econômicas estavam elegíveis para o Programa Nota Legal. Em especial, o cadastro obrigatório das empresas que recolhem ICMS só começou em 01/11/2008. E ao pesquisar na base de dados, foi verificado que apenas 3.402 documentos fiscais estavam cadastrados em 2008. Além disso, a população ainda não tinha sido conscientizada sobre a existência do PNL. Apenas em 2010 que houve o primeiro período de indicação para uso dos créditos no IPVA e IPTU.

Desta forma, decidiu-se por eliminar os dados de 2008 para obtenção dos períodos coerentes à utilização do PNL pela população.

7.2 Limpeza de Dados

Para tentar resolver o problema de *missing values* para os CEP de pessoas jurídicas, foi indicado por especialista da SEF o uso da tabela de Beneficiario do banco de dados do PNL. Para cada pessoa física cadastrada no programa há informações de endereço, autenticação e outros critérios de acesso da SEF. Por conta do sigilo fiscal esta tabela não terá seus campos detalhados.

Ao extrair seus dados, os mesmos problemas de qualidade se fizeram presentes. Mesmo analisando apenas o campo CEP da tabela, foi possível verificar que o tipo de dados do campo era nominal, e os dados cadastrados para este campo tinham problemas de qualidade piores que o anterior ao aceitar sequências de caracteres ao invés de apenas números. Desta forma, não foi possível a utilização destas informações.

7.3 Construção de Dados

A partir das bases de dados existentes decidiu-se incluir informações de idade, atividade econômica e regiões administrativas que poderiam ser derivadas dos dados atuais.

Por idade entende-se a idade da pessoa física no momento em que foi realizada a emissão do documento fiscal.

Por atividade econômica entende-se um agrupamento no fornecimento de bens e serviços similares, prestadas pelas pessoas jurídicas.

Como o DF tem a característica de ter apenas um município, a divisão em RAs se mostrou mais adequada para a pesquisa que outras proposições, como divisão em cidades ou bairros. Nas regiões administrativas pretendeu-se agrupar as pessoas ao local que moram ou tem sua sede. Para tanto inclui-se: a subdivisão do Distrito Federal nas regiões administrativas; os municípios fora do DF que formam a Região Integrada de Desenvolvimento do Distrito Federal e Entorno (RIDE); e uma região que abrangesse casos dos outros Estados.

7.3.1 RIDE e RA

A Região Integrada de Desenvolvimento do Distrito Federal e Entorno (RIDE) é uma região integrada de desenvolvimento econômico, criada em [5], e regulamentada em [17], para efeitos de articulação da ação administrativa da União, dos Estados de Goiás, Minas Gerais e do Distrito Federal.

A RIDE tem como objetivo articular e harmonizar as ações administrativas da União, dos estados e dos municípios para a promoção de projetos que visem à dinamização econômica e provisão de infraestruturas necessárias ao desenvolvimento em escala regional. Há

prioridade no recebimento de recursos públicos destinados a investimentos que estejam de acordo com os interesses consensuados entre os entes. Esses recursos devem contemplar demandas por equipamentos e serviços públicos, fomentar arranjos produtivos locais, propiciar o ordenamento territorial e assim promover o seu desenvolvimento integrado.

Consideram-se de interesse da RIDE os serviços públicos comuns ao Distrito Federal, Estados de Goiás, Minas Gerais e aos Municípios que a integram, relacionados com as seguintes áreas: infraestrutura, geração de empregos e capacitação profissional, saneamento básico, em especial o abastecimento de água, a coleta e o tratamento de esgoto e o serviço de limpeza pública, uso, parcelamento e ocupação do solo, transportes e sistema viário, proteção ao meio ambiente e controle da poluição ambiental, aproveitamento de recursos hídricos e minerais, saúde e assistência social, educação e cultura, produção agropecuária e abastecimento alimentar, habitação popular, serviços de telecomunicação, turismo, e segurança pública.

A área de abrangência da RIDE compreende as Regiões Administrativas do DF e municípios de Goiás e de Minas Gerais situados no entorno do DF.

As Regiões Administrativas do Distrito Federal são: RA I - Brasília, RA II - Gama, RA III - Taguatinga, RA IV - Brazlândia, RA V - Sobradinho, RA VI - Planaltina, RA VII - Paranoá, RA VIII - Núcleo Bandeirante, RA IX - Ceilândia, RA X - Guará, RA XI - Cruzeiro, RA XII - Samambaia, RA XIII - Santa Maria, RA XIV - São Sebastião, RA XV - Recanto das Emas, RA XVI - Lago Sul, RA XVII - Riacho Fundo, RA XVIII - Lago Norte, RA XIX - Candangolândia, RA XX - Águas Claras, RA XXI - Riacho Fundo II, RA XXII - Sudoeste/Octogonal, RA XXIII - Varjão, RA XXIV - Park Way, RA XXV - SCIA - Setor Complementar de Indústria e Abastecimento (Cidade Estrutural e Cidade do Automóvel), RA XXVI - Sobradinho II, RA XXVII - Jardim Botânico, RA XXVIII - Itapoã, RA XXIX - SIA - Setor de Indústria e Abastecimento, RA XXX - Vicente Pires, RA XXXI - Fercal.

Os municípios do Estado de Goiás são: Abadiânia, Água Fria de Goiás, Águas Lindas de Goiás, Alexânia, Cabeceiras, Cidade Ocidental, Cocalzinho de Goiás, Corumbá de Goiás, Cristalina, Formosa, Luziânia, Mimoso de Goiás, Novo Gama, Padre Bernardo, Pirenópolis, Planaltina, Santo Antônio do Descoberto, Valparaíso de Goiás e Vila Boa.

Os municípios do Estado de Minas Gerais são: Buritis, Cabeceira Grande e Unaí.

A Figura 7.1 permite uma visualização dos municípios de Goiás e de Minas Gerais junto ao Distrito Federal[1]. Os municípios em destaque formam a área conhecida por "Entorno"do Distrito Federal.

[1]Figura disponível em http://g1.globo.com/distrito-federal/noticia/2011/08/entorno-do-df-concentra-quase-40-dos-assassinatos-de-goias.html. Acesso em 02/06/2014

Figura 7.1: Região Integrada de Desenvolvimento do Distrito Federal e Entorno

7.3.2 Inclusão da variável idade

A variável idade pôde ser calculado a partir da diferença entre Dta_emissao_doc_fiscal e Dat_nasc_cpf_dados.

Uma vez que esta variável foi calculada, a variável Dat_nasc_cpf_dados foi descartada já que não agregaria mais informações.

7.3.3 Inclusão da variável atividade_economica

No site do Programa Nota Legal as atividades econômicas estão denominadas como ramos comerciais. Estas atividades foram definidas previamente no início do PNL pelos especialistas da SEF. Tratam-se de agrupamentos de CNAE.

A tabela Lista_Contribuintes do banco de dados do Programa Nota Legal possui o mapeamento entre as atividades econômicas cadastradas e seus CNAE correspondentes. A partir do cadastro fiscal de empresas da SEF foi possível obter o CNAE correspondente ao CNPJ das empresas utilizadas até este momento. Através de consultas SQL no banco de dados foi possível obter a atividade econômica associada a cada CNPJ. O resultado desta consulta ao banco de dados foi incorporado às bases de dados.

As variáveis foram incluídas nas bases de dados como atividade_origem para as empresas beneficiárias e atividade_destino para as empresas contribuintes.

7.3.4 Inclusão da variável RA

Como as informações relacionadas a endereços apresentavam problemas de qualidade de dados, tomou-se a decisão de utilizar a base de dados dos Correios para padronização de endereços.

A mídia fornecida pelos Correios apresentava arquivos de cargas de dados com formato de texto delimitado, documentação para processamento destes arquivos e uma sugestão de modelagem de dados para carregamento das informações.

Foi constatado na documentação que o banco de dados não oferecia a informação sobre regiões administrativas. Além disso, o banco de dados dos Correios cobria várias informações que não eram pertinentes a esta pesquisa.

Decidiu-se por customizar o modelo de dados fornecido para, a partir dos CEP existentes, obter as informações sobre as regiões administrativas, e remover aquelas que não eram de interesse desta pesquisa.

Seguindo a modelagem criada, foi implementado o banco de dados, e feito cargas nas tabelas. O preenchimento das RA foi feito manualmente.

A Figura 7.2 apresenta parte da modelagem realizada para obtenção das regiões administrativas. As outras tabelas e seus relacionamentos foram omitidos por objetividade e clareza textual. Em resumo, os Correios dividem o endereçamento postal em localidades e conforme seu tamanho são mapeados um CEP ou uma faixa de CEP para cada localidade. A tabela Localidades guarda informações sobre as localidades, sendo que o campo cep_localidade guarda o valor de CEP e o campo ra guarda a informação de região administrativa da localidade. A tabela Faixas_Bairro apresenta a faixa de CEP para uma determinada localidade, sendo que o campo cep_ini_fx_bai guarda o valor do início da faixa de CEP, o campo cep_fim_fx_bai guarda o valor do final da faixa de CEP e o campo ra guarda a informação de região administrativa da faixa.

Para cada registro da tabela Faixas_Bairro foi preenchido a RA correspondente às 281 faixas do Distrito Federal encontradas. Foi preenchido a RA "Entorno"para as 374 faixas referentes aos municípios de Goiás e Minas Gerais que fazem parte da RIDE. Para cada registro da tabela Localidades foi preenchido o valor "Entorno"aos municípios de Goiás e Minas Gerais que fazem parte da RIDE. Todos os outros casos, fora destas regiões, foram preenchidas com o valor "OUTROS".

Para extração destas informações, através de consultas SQL foi informado o CEP e obtida a RA correspondente. Se convencionou que consultas em que o CEP era inválido ou que não retornassem dados seriam preenchidas com o valor "erro". As regiões administrativas obtidas para cada CEP foram incluídas nas bases de dados. Apenas 38.916 documentos fiscais, dos 157.188.867 existentes, foram preenchidos como "erro".

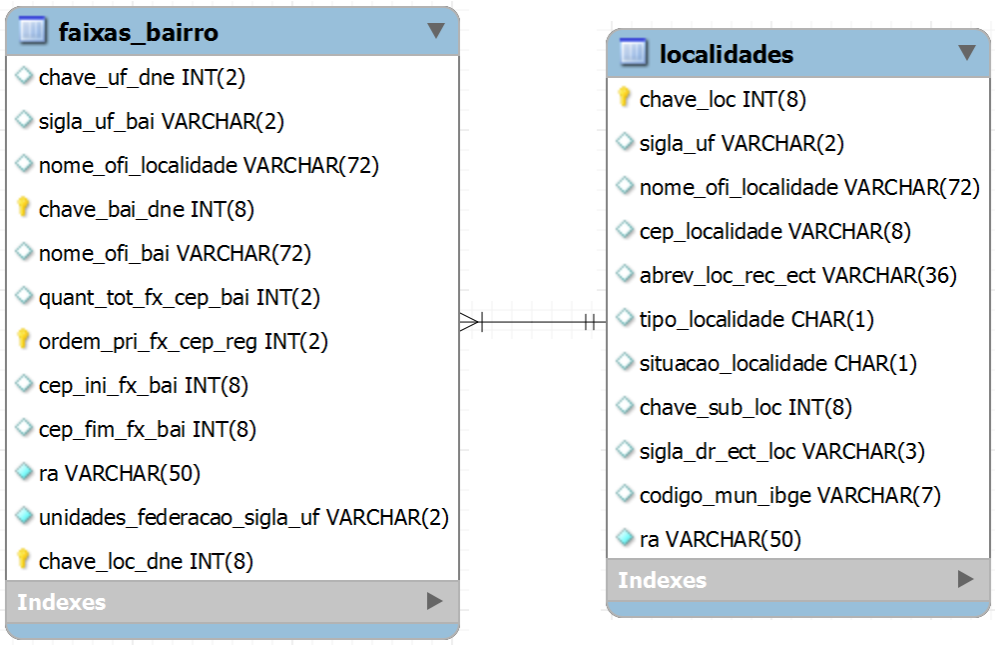

Figura 7.2: Modelo de Dados para busca de RA

Uma vez que as informações foram incluídas nas bases de dados como ra_origem para as empresas beneficiárias e ra_destino para as empresas contribuintes, as variáveis End_cep_dados e cep foram descartadas.

7.4 Integração de Dados

Uma vez que a adição dos novas variáveis foi concretizada não havia mais motivos para manter duas bases de dados. Então, a partir da variável CNPJ que é comum às duas bases, foi feito a junção das informações.

A Tabela 7.1 apresenta o resultado alcançado, e da mesma forma que na Seção 6.4 do capítulo anterior, foi utilizada estatística descritiva para ajudar a entender o comportamento dos dados, em que foram carregados 157.188.867 registros. As informações que se seguem foram objetos de análise.

A variável CNPJ diminuiu sua quantidade de dados distintos devido à junção das duas bases de dados em que só permaneceram os dados em comum.

Os valores em idade de máximo igual a 121,3 anos e de mínimo igual a 0,01 ano causaram suspeita de erro no cálculo das informações. Mas houve conferência junto ao banco de dados e os valores desta variável estavam cadastrados corretamente. Verificou-se ser necessária análise de *outliers* em um momento posterior.

Tabela 7.1: Análise da Base de Dados envolvendo a Preparação dos Dados

Campo	**Descrição**
Dta_emissao_doc_fiscal	Datas entre 01/01/2009 a 31/10/2013. Distintos = 1.765 Moda= 10/08/2013, com ocorrência de 328.010 documentos fiscais para a moda, representando 0,2% da base de dados
Val_doc_fiscal	Mínimo= 0,01 Máximo= 9.003.600.097,03 Média= 336,68 Desvio padrão= 1.247.496,53 Moda= 5, com ocorrência de 646.195 documentos fiscais para a moda representando 0,4% da base de dados
Cpf_destinatario	Distintos= 804.683 Moda com ocorrência de 246.022 documentos fiscais, representando 0,2% da base de dados
Cnpj_destinatario	Distintos= 37.302 Moda com ocorrência de 6.034 documentos fiscais, representando menos de 0,001% da base de dados
Val_iss	Mínimo= 0 Máximo= 50.000 Média= 0,37 Desvio padrão = 12,28 Moda = 0, com ocorrência de 147.566.777 documentos fiscais para a moda, representando 93,9% da base de dados
Val_icms	Mínimo= 0 Máximo= 701.880,00 Média= 5.454,89 Desvio padrão= 647.924,50 Moda= 0, com ocorrência de 36.585.707 documentos fiscais para a moda, representando 23,3% da base de dados
Val_credito	Mínimo= 0 Máximo= 144.237,12 Média= 2,21 Desvio padrão= 22,06 Moda= 0, com ocorrência de 16.956.211 documentos fiscais para a moda, representando 10,8% da base de dados
Sex_cpf_dados	Distintos= 2, Moda= 2-feminino, com ocorrência de 80.431.049 documentos fiscais para a moda, representando 51,2% da base de dados
CNPJ	Distintos = 30.265, Moda com ocorrência de 3.123.557 documentos fiscais, representando 2% da base de dados
idade	Mínimo= 0,01 Máximo= 121,3 Média= 43,86 Desvio padrão= 13,32 Moda= 32,02, com ocorrência de 60.886 documentos fiscais para a moda, representando menos de 0,001% da base de dados
atividade_economica_origem	Distintos= 46, Moda= "Outras_atividades", com ocorrência de 4.303 documentos fiscais para a moda, representando menos de 0,001% da base de dados
atividade_economica_destino	Distintos= 46, Moda= "Hipermercados, supermercados, padarias e confeitarias", com ocorrência de 60.237.806 documentos fiscais para a moda, representando 38,32% da base de dados
ra_origem	Distintos= 34, Moda= "Brasília", com ocorrência de 44.427.933 documentos fiscais para a moda, representando 28,26% da base de dados
ra_destino	Distintos= 33, Moda= "Brasília", com ocorrência de 67.596.748 documentos fiscais para a moda, representando 43% da base de dados

Nota-se que a moda em atividade_economica_destino, ra_origem e ra_destino dominam grande parte da base de dados. Este comportamento era esperado pelo que os especialistas observaram na vigência do PNL.

Em complemento às observações feitas, as figuras 7.3, 7.4, 7.5 e 7.6 ajudam a compreender o comportamento das variáveis adicionadas.

A Figura 7.3 apresenta as atividades econômicas de maior incidência na base de dados. O valor "Outras atividades"agrupa os outros valores existentes.

A Figura 7.4 apresenta o comportamento da variável idade ao longo do tempo.

As Figuras 7.5 e 7.6 apresentam as regiões administrativas de maior incidência na base de dados. O valor "Outras RA"agrupa os outros valores existentes.

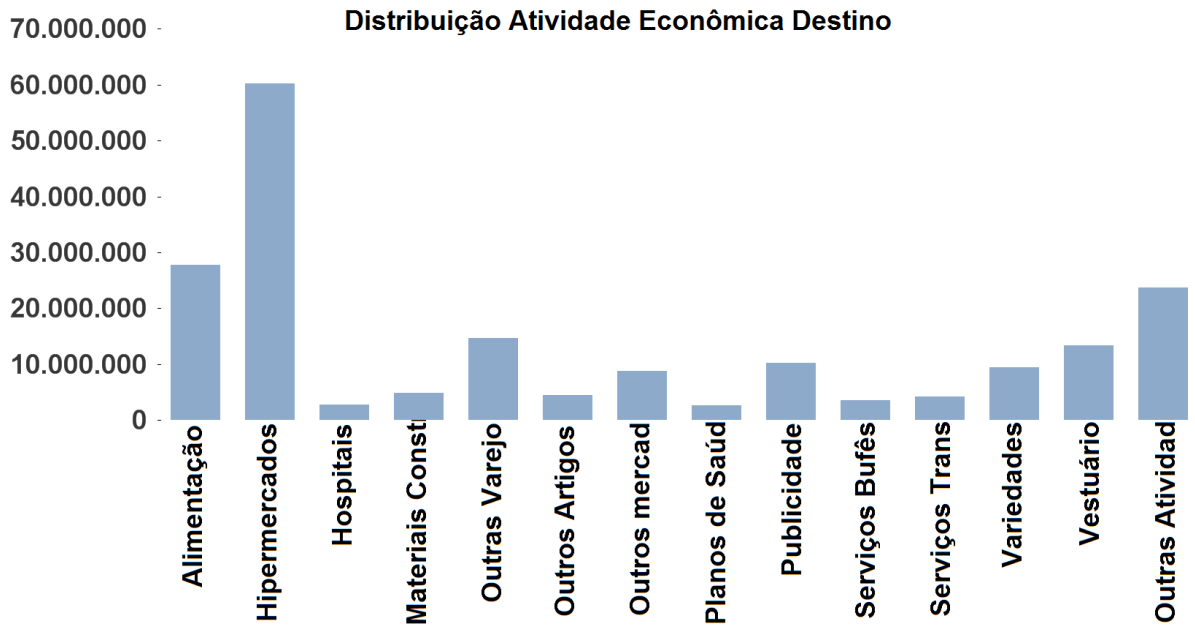

Figura 7.3: Distribuição Atividade Destino

7.5 Formatação de Dados

Como passo final à preparação dos dados analisou-se sua adequação visando facilitar a extração dos perfis desejados. Criou-se então várias visões dos mesmos dados adequados à diferentes análises. Desta forma, foram gerados 5 arquivos: icms_pj.txt, icms_pf.txt, iss_pj.txt e iss_pf.txt para o perfil de créditos de consumo e pf_temp.txt para o perfil de fidelidade.

Para o perfil de créditos de consumo, conforme detalhado na Seção 6.4, análises tanto de ICMS e ISS quanto de pessoa física e pessoa jurídica deveriam ser processadas separadamente. Para o arquivo icms_pf.txt foram removidas as variáveis Cpf_destinatario,

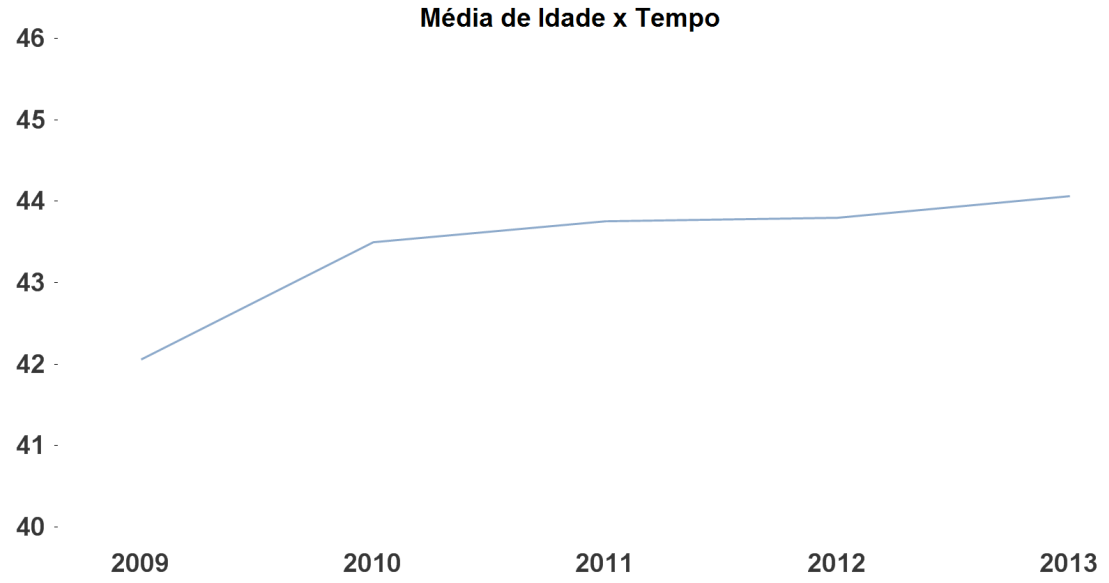

Figura 7.4: Média de Idade x Tempo

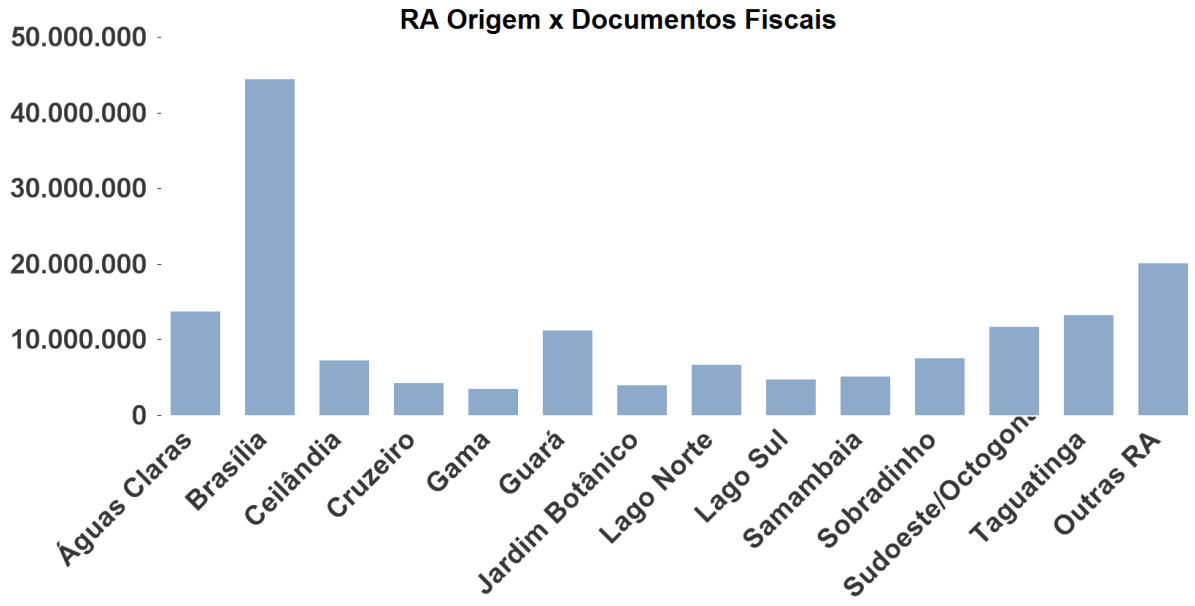

Figura 7.5: Distribuição RA Origem

Cnpj_destinatario, Sex_cpf_dados, idade e Val_iss. Para o arquivo icms_pj.txt foram removidas as variáveis Cpf_destinatario, Cnpj_destinatario, atividade_economica_origem, ra_origem e Val_iss. Para o arquivo iss_pf.txt foram removidas as variáveis Cpf_destinatario, Cnpj_destinatario, Sex_cpf_ dados, idade e Val_icms. Para o arquivo iss_pj.txt foram removidas as variáveis Cpf _destinatario, Cnpj_destinatario, atividade_economica_origem, ra_origem e Val_icms.

Como o perfil de fidelidade abrange apenas o comportamento das pessoas físicas, para o

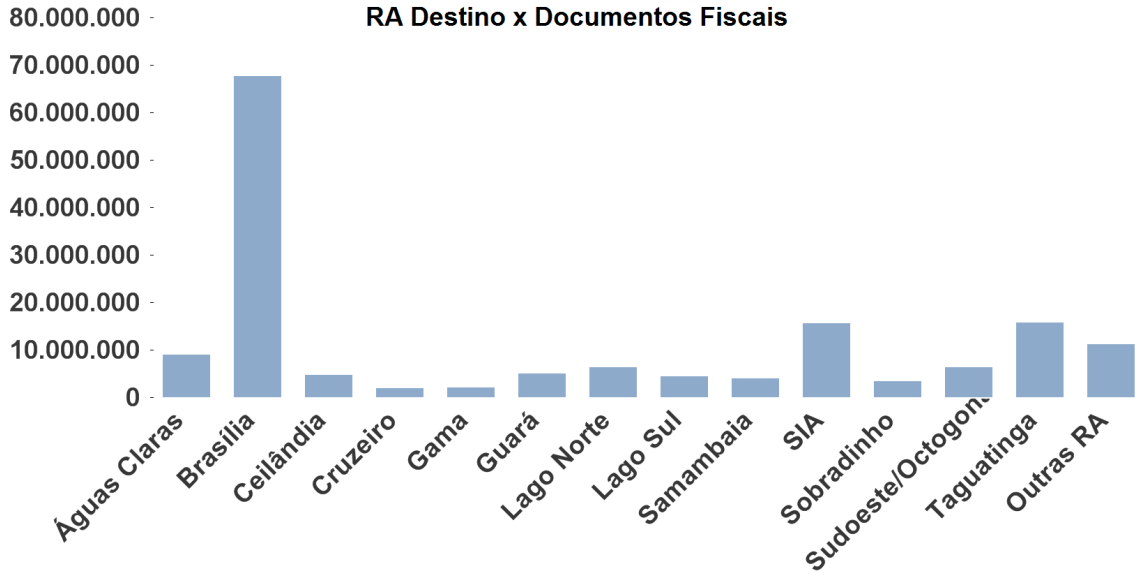

Figura 7.6: Distribuição RA Destino

arquivo pf_temp.txt foram removidas as variáveis: Cpf_destinatario, Cnpj_destinatario e atividade_economica_ origem. O comportamento do cidadão ao solicitar o documento fiscal não faz distinção entre aquisições de bens ou serviços, desta forma para esta base de dados os valores de ICMS e ISS foram mantidos juntos.

As bases de dados obtidas somadas registraram em torno de 8GB de tamanho.

Capítulo 8

Apresentação e Análise de Resultados

Neste capítulo os dados obtidos serão analisados e extraídos os perfis desejados. A Seção 8.1 apresenta análises e refinamentos sobre os dados. A Seção 8.2 analisa situações passíveis de *outliers*. A Seção 8.3 analisa os dados sobre a perspectiva do perfil sobre fidelidade. A Seção 8.4 analisa os dados sobre a perspectiva do perfil sobre créditos.

Este Capítulo cobre as fases de modelagem e avaliação do CRISP-DM.

8.1 Análises Gerais

Ao inserir as variáveis RA e atividade_economica percebe-se grande quantidade de valores distintos associados a estas variáveis. É proposto um refinamento para agrupar os valores destas variáveis.

8.1.1 Agrupamento de RA

Conforme apresentado nos capítulos anteriores, as informações de endereço do Distrito Federal foram divididos em regiões administrativas. Embora seja uma alternativa pertinente para o problema existente de qualidade dos dados, inferir dados sobre estas 34 localidades torna-se uma tarefa árdua.

Conforme apresentado nas figuras 7.5 e 7.6, a emissão de documentos fiscais é desbalanceada entre as RA. Como a atribuição dos CEP às RA foi feita de forma manual, esta atribuição pode incorrer em erros em regiões limítrofes entre si. Pela característica do Distrito Federal de organização das residências das pessoas físicas em condomínios, há dificuldade em determinar a qual RA pertencem certos condomínios. Como exemplo, é possível citar a região entre Sobradinho e Sobradinho II e a região entre entre Jardim Botânico e São Sebastião. Além disso, parte do comércio se encontra nas rodovias que dividem estas RA. Nos perfis desejados é esperado que o consumo resultante em benefícios

ao PNL seja realizado próximo da residência das pessoas. Classificar erroneamente as RA poderia sugerir padrões inexistentes.

Para resolver estas questões, tomou-se a decisão de realizar o agrupamento da RA em 9 grupos, visando a proximidade das regiões em que as pessoas participantes do PNL habitam.

Um grupo foi criado para o valor "Entorno"e outro para o valor "Outros". Como são regiões externas ao DF, na verdade, já se tratam de agrupamentos previamente criados. O grupo "Entorno"engloba a região da RIDE enquanto o grupo "Outros"engloba outras regiões do país que não sejam do Distrito Federal e da RIDE.

Como os dados com valor “erro” não poderiam ajudar na inferência de dados, estas informações foram removidas das bases de dados ao realizar o agrupamento.

Para cada um dos grupos foi criada uma palavra mnemônica para facilitar a localização do grupo.

A Tabela 8.1 apresenta o agrupamento realizado.

Tabela 8.1: Agrupamento de Regiões Administrativas

Grupo	**Regiões Administrativas**
0 - Norte	Planaltina, Sobradinho, Sobradinho II, Fercal, Paranoá, Itapoã, Varjão, Brazlândia
1 - Entorno	Entorno
2 - Outros	OUTROS
3 - Oeste	Águas Claras, Taguatinga, Vicente Pires, Ceilândia, Recanto das Emas, Riacho Fundo, Riacho Fundo II, Samambaia
4 - Sul	Santa Maria, Gama
5 - Abastecimento	SCIA, SIA
6 - Sudeste	São Sebastião, Jardim Botânico
7 - Sudoeste	Candangolândia, Núcleo Bandeirante, Park Way, Guará, Cruzeiro
8 - Centro	Brasília, Sudoeste/Octogonal, Lago Norte, Lago Sul

A Figura 8.1 permite melhor visualização do agrupamento das 31 RA existentes do Distrito Federal nos grupos de RA.

As Figuras 8.2 e 8.3 apresentam a distribuição de documentos fiscais por agrupamentos de RA criados.

Nas bases de dados, as informações existentes sobre RA foram substituídas pelo índice do grupo, visando melhoria de performance e memória ao processar estas informações.

8.1.2 Agrupamento de Atividades Econômicas

Ao utilizar as atividades econômicas definidas pelo PNL, pôde ser notado, conforme apresentado na Seção 7.1, que existe uma grande quantidade de atividades econômicas

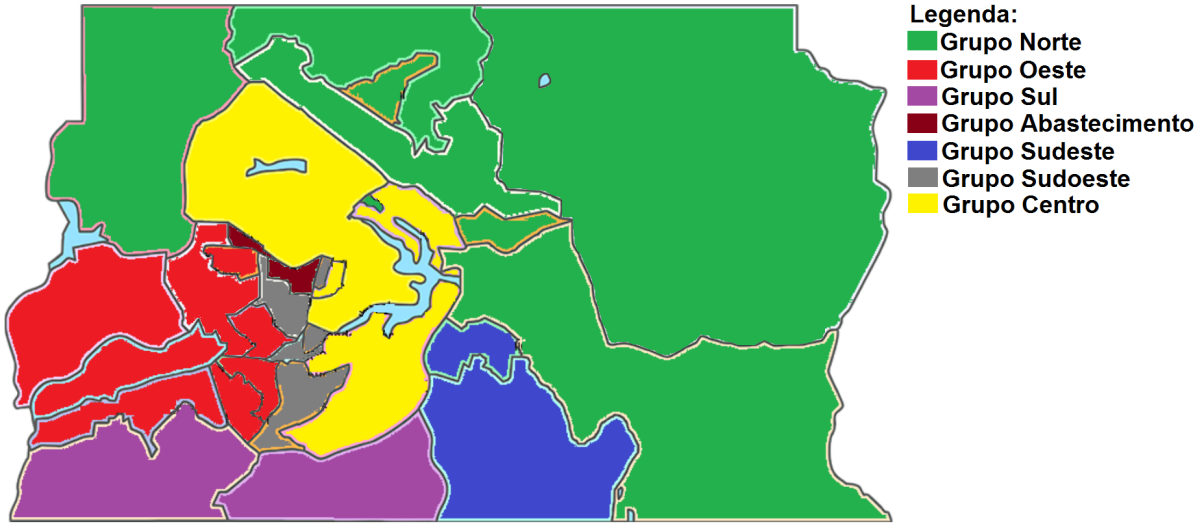

Figura 8.1: Agrupamentos de RAs no DF

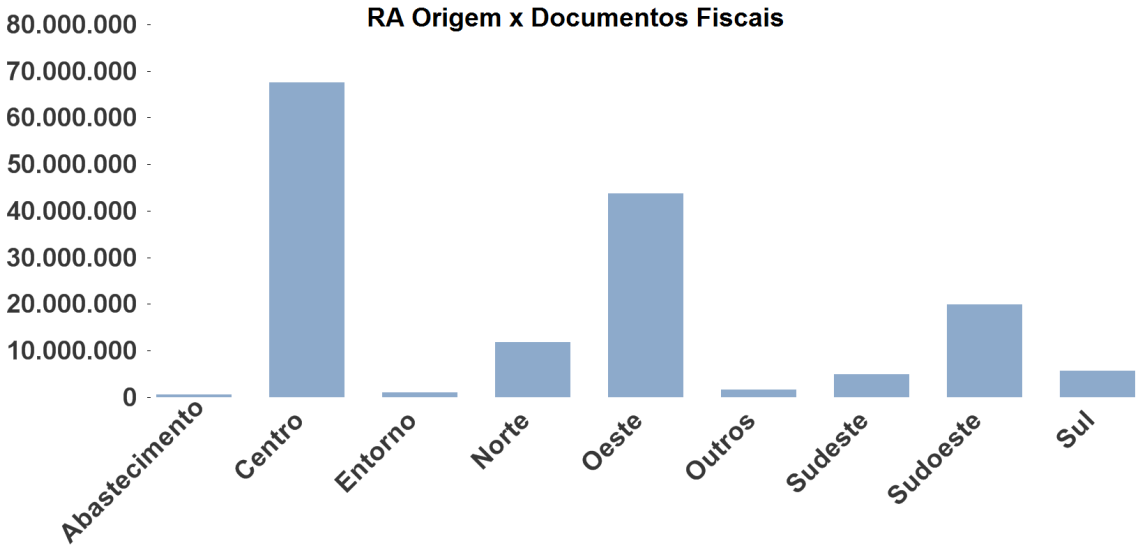

Figura 8.2: Agrupamentos RA Origem x Documentos Fiscais

com incidências diversas no PNL, conforme apresentado pela Figura 7.3. Pode-se perceber que certas atividades econômicas guardam grande semelhança entre si, como exemplo, "Educação", "Ensino superior e atividades de apoio à educação" e "Outras atividades de ensino". Em seu dia a dia, para os beneficiários do PNL, há pouca diferenciação em atividades econômicas tão similares, já que todas se referem a aquisições e serviços realizados em Educação de modo geral.

Visando mitigar a quantidade de atividades econômicas, foi realizado um agrupamento de atividades econômicas que guardam semelhança entre si, resultando em 7 grupos.

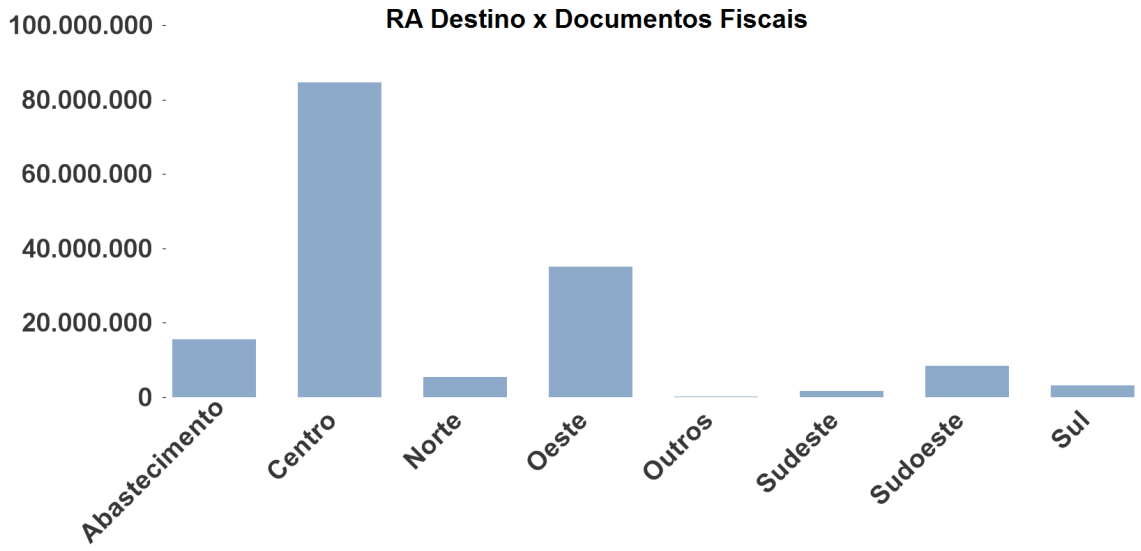

Figura 8.3: Agrupamentos RA Destino x Documentos Fiscais

Para cada um dos grupos foi criada uma palavra mnemônica em referência às atividades econômicas exercidas pelo grupo.

A Tabela 8.2 apresenta o agrupamento realizado.

A Figura 8.4 permite visualizar a distribuição de atividades econômicas nos agrupamentos criados. Apesar de ainda haver uma grande dispersão das frequências das atividades econômicas, a quantidade de valores a serem analisados é menor e mais intuitivo aos beneficiários do PNL.

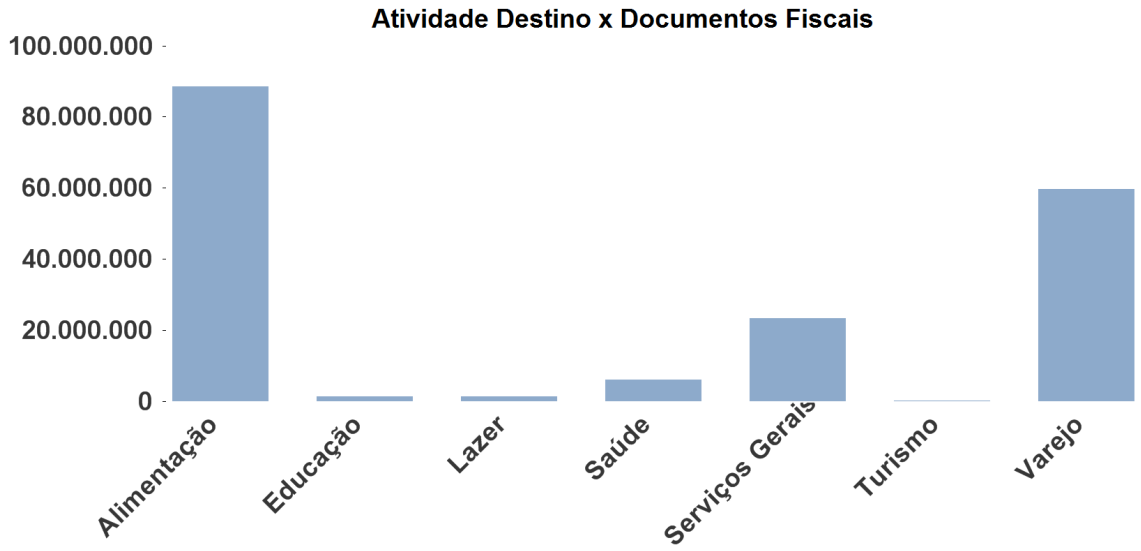

Figura 8.4: Agrupamentos Atividade Destino x Documentos Fiscais

Tabela 8.2: Agrupamento de Atividades Econômicas

Grupo	**Atividades econômicas**
0 - Alimentação	Alimentação
	Hipermercados, supermercados, padarias e confeitarias
	Serviços de bufês e comidas preparadas
1 - Varejo	Brinquedos e artigos recreativos
	Material para construção e para comunicação
	Óticas, relojoarias, joalherias e bijuterias
	Outras atividades de varejo
	Outros artigos de uso doméstico
	Outros mercados varejistas
	Variedades, informática, papelaria e eletrodomésticos
	Cabeleireiros, tratamento de beleza e estética
	Vestuário, calçados e artigos de viagem
	Veículos e motocicletas usados
	Engenharia, arquitetura e urbanismo
	Manutenção de máquinas e equipamentos
	Instalações elétricas
	Serviços em veículos automotores
	Serviços para casa, decoração e reparação
	Vigilância, segurança privada e monitoramento de sistemas de segurança
2 - Educação	Educação
	Ensino superior e atividades de apoio à educação
	Outras atividades de ensino
3 - Lazer	Academias e outras atividades esportivas
	Cinemas, parques, discotecas, boliches, sinuca e jogos eletrônicos
	Campings, pensões e outros alojamentos
4 - Serviços Gerais	Atividades de contabilidade e auditoria
	Atividades funerárias
	Atividades imobiliárias
	Atividades veterinárias
	Estacionamento de veículos
	Lavanderia, tinturaria e toalheiro
	Outros serviços
	Publicidade, fotografia, filmagem de festas e eventos e microfilmagem
	Representação comercial
	Serviços advocatícios
	Serviços de gráfica e editoração
	Serviços de informática
	Serviços de limpeza
	Serviços de seleção e agenciamento de mão-de-obra
	Serviços de transporte, guarda e entrega
5 - Turismo	Agências de viagem e turismo
	Hotéis e similares
6 - Saúde	Hospitais, odontologia, vacinação, laboratórios e exames
	Medicamentos homeopáticos, veterinários e artigos médicos
	Planos de saúde, seguros e previdência complementar

Nas bases de dados, as informações existentes foram substituídas pelo índice do grupo, visando melhoria de performance e memória ao processar estas informações.

8.2 Análise de *Outliers*

Nos capítulos anteriores foram destacados três atributos para avaliação de *outliers*: idade das pessoas físicas; quantidade de documentos fiscais emitidos por pessoas físicas; e valores dos documentos fiscais.

8.2.1 Pessoas físicas com idades não produtivas

No capítulo anterior, a Figura 7.4 apresenta a variação da idade média encontrada na base de dados ao longo do tempo. Pela Tabela 7.1, embora a idade média calculada de 43,86 anos esteja de acordo com as expectativas, os valores em idade de máximo igual a 121,3 anos e de mínimo igual a 0,01 ano dos usuários do PNL são atípicos.

Como beneficiários do PNL precisam fornecer o CPF no momento de alguma aquisição ou serviço para que no ano seguinte possam ter descontos nos impostos distritais, não era esperado interesse no PNL por crianças devido a suas limitações financeiras e nem por idosos com idade acima de 90 anos devido a suas limitações físicas.

No primeiro caso, foram encontrados 36.244 pessoas físicas (vinculadas a 0,3% dos documentos fiscais) com idade até 20 anos. Como é possível a emissão do CPF desde o momento do nascimento e há casos em que crianças já nascem herdeiras de bens, não há motivos objetivos para exclusão destas pessoas.

Para o caso dos idosos, foram encontrados 230 pessoas físicas (vinculadas a 0,02% dos documentos fiscais) com idade acima de 90 anos. Neste caso também não há motivos objetivos para exclusão destas pessoas.

Não deve ser descartado que no momento da aquisição ou da prestação de um serviço, por erro da pessoa física ou da empresa ao emitir o documento fiscal, seja informado o CPF de outra pessoa. Tal pessoa poderia estar presente nos casos relatados ou ainda ser alguém já falecido. No processamento da SEF, por ser um documento fiscal válido, ele é processado e inserido no banco de dados.

Estes casos motivaram uma pesquisa paralela. Todos os CPF foram pesquisados para verificação de óbito na tabela de pessoa física da RFB. Foram encontrados 198 pessoas, independente de idade, com status de falecimento no período do PNL entre 2008 a 2013.

Os casos relatados para crianças, idosos e com status de falecimento foram encaminhados à SEF para avaliação.

Como não há restrições na Lei que instituiu o PNL, estes casos foram mantidos na base de dados. No entanto, para diferenciar os extremos, as análises de idade para os

perfis foi dividida em 9 faixas com o objetivo de segmentar as pessoas físicas. São elas: 0-20 anos; 20-30 anos; 30-40 anos; 40-50 anos; 50-60 anos; 60-70 anos; 70-80 anos; 80-90 anos; e acima de 90 anos.

8.2.2 Quantidade de documentos fiscais emitidos

Para avaliação da quantidade de documentos fiscais emitidos por pessoa física, foram geradas as seguintes estatísticas para cada valor encontrado de documento fiscal (doc fiscais): quantidade de pessoas físicas vinculadas ao documento fiscal (quant), probabilidade de ocorrência (prob), probabilidade acumulada de ocorrência (prob acum), média (media), desvio padrão (dp) e coeficiente de variação (cv). Essas probabilidades foram estimadas com base nas frequências de ocorrência na base de dados. A Tabela 8.3 apresenta os valores e as estatísticas calculadas.

Valores de documentos fiscais que apresentavam estatísticas similares foram agrupados em faixas de valores até o limite de 2.000 documentos fiscais. Dois mil documentos fiscais por pessoa física é o valor limite esperado pela SEF para emissão de documentos fiscais no período de 2009 a 2013.

Como base de comparação, a média de documentos fiscais emitidos por pessoa física, no período de 2009 a 2013, foi de 195 documentos fiscais. Então o valor de 2.000 documentos fiscais representa um valor limite pelo menos 10 vezes superior à média.

Constata-se que 367 pessoas físicas apresentaram mais de 2.000 documentos fiscais na vigência pesquisada e, em especial para este caso, o coeficiente de variação apresenta alta dispersão dos dados.

Desta forma, decidiu-se remover da base de dados estas 367 pessoas físicas com seus documentos fiscais vinculados.

Esta avaliação levou em conta toda a vigência do PNL. Na Seção 8.3.1 é feita a avaliação da quantidade de documentos por períodos da vigência, implicando na remoção de 8341 pessoas físicas.

Embora tenham sido removidas 9208 (367 + 8341) pessoas físicas das bases de dados, isso não implica que estes dados sejam irrelevantes e deve-se analisar o porquê deste comportamento.

Presume-se que poderiam ser erros na declaração do LFE, intencionais ou não, por parte do contribuinte. Tais erros podem se enquadrar em tentativa de fraude dos contribuintes.

Como citado em 5.2, não há críticas aos valores processados pelo LFE ao obter os documentos fiscais com CPF/CNPJ declarados. Neste caso, a partir da reclamação dos beneficiários originários é possível correção e mitigação das ocorrências.

Tabela 8.3: Análise da Quantidade de Documentos Fiscais

doc fiscais	quant	prob	prob acum	media	dp	cv (em %)
1	6246	0,776	0,776	1,000	0,000	0,000
2	6256	0,777	1,553	2,000	0,000	0,000
3	6306	0,784	2,337	3,000	0,000	0,000
4	6052	0,752	3,089	4,000	0,000	0,000
5	6232	0,774	3,863	5,000	0,000	0,000
6	6136	0,763	4,626	6,000	0,000	0,000
7	6155	0,765	5,391	7,000	0,000	0,000
8	6038	0,75	6,141	8,000	0,000	0,000
9	6026	0,749	6,89	9,000	0,000	0,000
10	6020	0,748	7,638	10,000	0,000	0,000
11	5906	0,734	8,372	11,000	0,000	0,000
12	5811	0,722	9,094	12,000	0,000	0,000
13	5859	0,728	9,822	13,000	0,000	0,000
14	5852	0,727	10,549	14,000	0,000	0,000
15	5841	0,726	11,275	15,000	0,000	0,000
16	5581	0,694	11,969	16,000	0,000	0,000
17	5551	0,69	12,659	17,000	0,000	0,000
18	5636	0,7	13,359	18,000	0,000	0,000
19	5450	0,677	14,036	19,000	0,000	0,000
20	5349	0,665	14,701	20,000	0,000	0,000
21-30	50179	6,236	20,937	25,411	2,867	11,282
31-40	44196	5,492	26,429	35,412	2,868	8,098
41-50	38236	4,752	31,181	45,399	2,871	6,323
51-60	33914	4,215	35,396	55,411	2,866	5,172
61-70	30197	3,753	39,149	65,432	2,856	4,364
71-80	27169	3,376	42,525	75,427	2,865	3,798
81-90	24377	3,029	45,554	85,407	2,873	3,363
91-100	22156	2,753	48,307	95,458	2,862	2,998
101-200	152766	18,985	67,292	145,167	28,781	19,826
201-300	87824	10,914	78,206	246,606	28,757	11,661
301-400	57663	7,166	85,372	346,932	28,775	8,294
401-500	38633	4,801	90,173	447,225	28,763	6,431
501-750	49877	6,198	96,371	604,952	70,112	11,589
751-1000	18287	2,273	98,644	855,014	70,704	8,269
1001-1250	6640	0,825	99,469	1105,828	69,967	6,327
1251-1500	2499	0,311	99,78	1356,036	71,35	5,261
1501-1750	979	0,122	99,902	1610,419	70,744	4,392
1751-2000	421	0,052	99,954	1860,014	73,729	3,963
>2000	367	0,046	100	3495,954	12922,04	369,628

Para os casos do beneficiário não ter solicitado inclusão do seu CPF/CNPJ no documento fiscal, a Portaria [18], estabelece que, a partir de dezembro de 2011, apenas 5 documentos fiscais por pessoa física ou jurídica são viáveis para emissão diariamente em um mesmo estabelecimento. Antes desta data, 10 documentos fiscais eram permitidos. Caso estes limites sejam ultrapassados o valor do crédito é bloqueado. Mesmo assim, a ocorrência repetida dos limites de documentos fiscais de beneficiários ao longo do tempo poderia caracterizar tentativa de fraude.

Em qualquer das hipóteses levantadas, durante a fase de indicação de créditos do PNL, são feitas verificações de bloqueio do beneficiário e de seus créditos vigentes antes que eles sejam disponibilizados para uso.

Por serem extremamente atípicos, os dados removidos não poderiam colaborar para determinação do comportamento padrão de obtenção de créditos e de fidelidade, ou seja, os perfis desejados. Cabe ressaltar que estes casos foram encaminhados à SEF para verificação.

8.2.3 Valores dos documentos fiscais

Após a remoção das pessoas físicas, conforme Seções 8.2.2 e 8.3.1, foi realizada a análise dos valores para a variável Val_doc_fiscal. Como Val_iss, Val_icms e Val_credito são calculados a partir da variável Val_doc_fiscal, estas variáveis não foram analisados novamente.

Foram obtidos os valores Mínimo= 0,01 Máximo= 9.003.600.097,03 Média= 348.87, Desvio padrão= 1.293.361,52 e Moda= 5, com ocorrência de 583.007 documentos fiscais para a moda, representando 0,4% da base de dados.

À primeira vista não houve grande variação para esta variável ao comparar com as estatísticas apresentadas na Tabela 7.1. Ao ordenar os valores desta variável, foi possível perceber que apenas 7 documentos fiscais tinham valores acima de R$ 10.000.000,00, sendo que 4 deles apresentaram valores acima de R$ 1.000.000.000,00. Para os outros 3 documentos fiscais foi verificado a atividade econômica vinculada e contatou-se o grupo 0 relacionado à alimentação. Decidiu-se então por eliminar estes 7 documentos fiscais da base de dados já que são valores irreais para aquisições e serviços por pessoas físicas.

Estes valores provavelmente foram inseridos no banco de dados da SEF por erro das empresas ao enviar o LFE, sendo que a criação do LFE é feito por aplicativos terceiros à SEF, conforme apresentado na Seção 3.4.

As estatísticas da variável Val_doc_fiscal após a remoção destes dados são Mínimo= 0,01, Máximo= 8.902.366,70, Média= 152,24, Desvio padrão= 3.950,90 e Moda= 5, com ocorrência de 583.007 documentos fiscais para a moda, representando 0,4% da base de dados.

O novo valor encontrado para a média da variável Val_doc_fiscal está de acordo com as expectativas da SEF para o PNL.

8.3 Estudo de caso: Perfil fidelidade

A pesquisa da fidelidade dos usuários ao PNL envolve analisar o tempo em que os beneficiários se encontram na base de dados e sua participação efetiva.

Para tanto foi proposto a divisão do período de vigência em cinco faixas de fidelidade: 1-11 meses, 12-23 meses, 24-35 meses, 36-47 meses, e 48-58 meses [1]. Por faixa de fidelidade entende-se o tempo de participação efetivo no PNL, ou seja, a quantidade de meses em que a pessoa física emitiu documentos fiscais.

São analisadas, em maiores detalhes, as iterações entre as variáveis ra_origem, atividade_economica_destino, idade e Sex_cpf_dados. Para facilitar o entendimento das tabelas e gráficos, as variáveis serão apresentadas respectivamente como RA, Atividade, Idade e Sexo. Entende-se que estas variáveis são fatores determinantes para o comportamento das pessoas físicas no PNL.

8.3.1 Indicadores para avaliação da fidelidade

Foram propostos dois indicadores para avaliação da fidelidade das pessoas físicas:

- Indicador "intensidade" definido como a quantidade de documentos fiscais emitidos por mês de participação da pessoa física no PNL (#doc/mp). Por meses de participação (mp) entende-se a quantidade de meses em que houve emissão de documentos fiscais.
- Indicador "perseverança" definido como a quantidade de meses de participação da pessoa física pela quantidade de meses de cadastro (mp/mc). Por meses de cadastro (mc) entende-se a quantidade de meses que se passaram desde o cadastro da pessoa física no PNL até a data final da vigência pesquisada (31/10/2013).

Enquanto o primeiro indicador avalia o desempenho real obtido, o segundo indicador avalia a perseverança do beneficiário em participar do PNL.

A Tabela 8.4 apresenta os valores: faixa de fidelidade (meses), quantidade de pessoas físicas (quant), porcentagem de pessoas físicas (freq), média (media), desvio padrão (dp) e coeficiente de variação (cv) para os indicadores criados.

Pela tabela, para o indicador intensidade, quanto maior o tempo de participação no PNL, maior a média de documentos fiscais emitidos e maior o desvio padrão. Mesmo com

[1] Como a vigência pesquisada foi de 01/01/2009 até a data de 31/10/2013, o tempo máximo de participação no PNL é de 58 meses.

Tabela 8.4: Estatísticas para os indicadores de fidelidade

Meses	quant	freq	Perseverança			Intensidade		
			media	dp	cv	media	dp	cv
1-11	143344	17,821	0,273	0,188	0,688	2,725	3,594	1,318
12-23	213154	26,501	0,475	0,155	0,326	3,877	3,483	0,898
24-35	223473	27,784	0,665	0,112	0,169	5,988	4,363	0,728
36-47	193494	24,056	0,832	0,070	0,085	9,391	5,385	0,573
48-58	30856	3,836	0,940	0,024	0,026	14,703	6,466	0,439

o coeficiente de variação diminuindo com o passar do tempo, este indicador apresentou alta dispersão de dados. Para o indicador perseverança, o coeficiente de variação teve comportamento cada vez mais homogêneo nas faixas de fidelidade.

Para avaliação da tendência central dos dados, foi realizada a distribuição dos valores de média encontrados nos indicadores, em percentis. As Tabelas 8.5 e 8.6 apresentam os valores encontrados. As Figuras 8.5 e 8.6, respectivamente, apresentam os gráficos das tabelas. Analisando a Tabela 8.5 e a Figura 8.5, percebe-se que a variação entre os percentis de 99% para 100% nas faixas foi muito alta. Embora tenha sido atenuada nas últimas faixas de fidelidade, esta variação representa valores atípicos do comportamento dos beneficiários do PNL.

Tabela 8.5: Percentis calculados para o indicador intensidade

Meses	25%	50%	75%	85%	95%	96%	97%	98%	99%	100%
1-11	1,250	1,700	2,727	3,888	8,000	9,000	10,600	13,000	17,400	347,500
12-23	1,947	2,761	4,400	5,904	10,133	11,150	12,529	14,615	18,368	114,562
24-35	3,133	4,676	7,346	9,461	14,303	15,357	16,742	18,750	22,393	75,875
36-47	5,526	8,128	11,842	14,365	19,736	20,837	22,181	24,138	27,652	52,631
48-58	9,941	13,479	18,145	21,220	27,166	28,312	29,686	31,530	33,960	41,666

Tabela 8.6: Percentis calculados para o indicador perseverança

Meses	25%	50%	75%	85%	95%	96%	97%	98%	99%	100%
1-11	0,148	0,224	0,333	0,444	0,714	0,750	0,800	0,833	0,888	0,916
12-23	0,367	0,444	0,547	0,633	0,814	0,846	0,880	0,909	0,941	0,958
24-35	0,583	0,653	0,723	0,781	0,894	0,914	0,928	0,941	0,960	0,972
36-47	0,775	0,833	0,886	0,918	0,940	0,953	0,959	0,959	0,959	0,979
48-58	0,924	0,942	0,960	0,962	0,979	0,979	0,980	0,980	0,980	0,983

Desta forma, decidiu-se eliminar as pessoas físicas encontradas para o centil de 100% do indicador intensidade, ou seja, foram eliminados da base de dados 8.341 pessoas físicas com seus documentos fiscais vinculados.

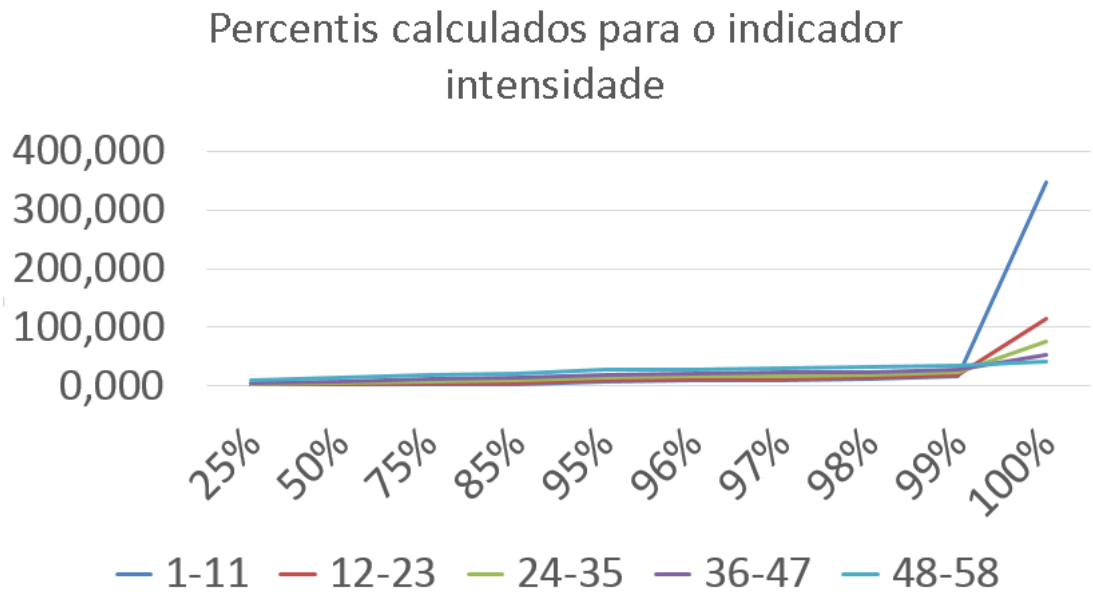

Figura 8.5: Gráfico de percentis para o indicador intensidade

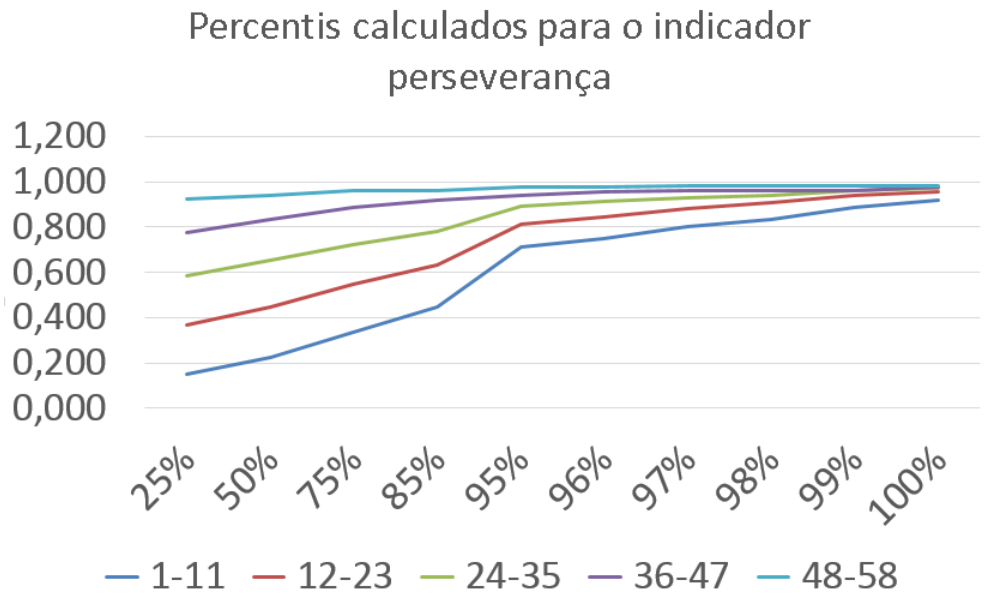

Figura 8.6: Gráfico de percentis para o indicador perseverança

As Tabelas 8.7, 8.8 e 8.9 apresentam os valores recalculados após a remoção dessas pessoas físicas.

Houve melhoria no comportamento do coeficiente de variação do indicador intensidade, mas os dados continuaram apresentando alta dispersão de dados. O indicador mp/mc foi minimamente afetado pela remoção dos dados.

Independente das faixas de fidelidade, para a vigência do PNL, o indicador intensidade calculado foi 3,9 documentos fiscais / mês e o indicador perseverança calculado foi 63,6%.

Tabela 8.7: Estatísticas recalculadas para os indicadores intensidade e perseverança

Meses	quant	freq	Perseverança			Intensidade		
			media	dp	cv	media	dp	cv
1-11	141876	17,824	0,269	0,183	0,681	2,490	2,275	0,913
12-23	210971	26,504	0,473	0,152	0,322	3,663	2,689	0,734
24-35	221169	27,785	0,664	0,112	0,168	5,755	3,678	0,639
36-47	191470	24,054	0,831	0,070	0,085	9,146	4,833	0,528
48-58	30494	3,831	0,940	0,024	0,026	14,444	6,048	0,418

Tabela 8.8: Percentis recalculados para o indicador intensidade

Meses	25%	50%	75%	85%	95%	96%	97%	98%	99%	100%
1-11	1,250	1,667	2,667	3,727	7,000	7,909	9,000	10,500	13,000	17,400
12-23	1,944	2,739	4,333	5,727	9,333	10,136	11,150	12,500	14,562	18,368
24-35	3,121	4,636	7,235	9,226	13,500	14,312	15,357	16,735	18,706	22,394
36-47	5,500	8,070	11,692	14,098	18,872	19,745	20,830	22,156	24,087	27,652
48-58	9,898	13,380	17,896	20,792	26,061	26,961	27,980	29,347	30,940	33,961

Tabela 8.9: Percentis recalculados para o indicador perseverança

Meses	25%	50%	75%	85%	95%	96%	97%	98%	99%	100%
1-11	0,148	0,222	0,333	0,435	0,688	0,750	0,786	0,833	0,875	0,917
12-23	0,366	0,444	0,545	0,629	0,800	0,833	0,867	0,905	0,933	0,958
24-35	0,581	0,653	0,723	0,780	0,892	0,912	0,923	0,941	0,946	0,972
36-47	0,776	0,833	0,887	0,917	0,940	0,952	0,958	0,959	0,959	0,979
48-58	0,925	0,942	0,961	0,962	0,980	0,980	0,980	0,980	0,981	0,983

8.3.2 Análise das variáveis selecionadas

Tabela 8.10: Perfil fidelidade: Sexo x Idade

1-11	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	>90	Total
masc	3,042	18,577	14,399	7,993	4,117	1,943	0,758	0,167	0,010	51,005
fem	3,808	17,467	13,395	7,294	4,260	1,956	0,642	0,157	0,016	48,995
Total	6,851	36,044	27,793	15,287	8,376	3,899	1,399	0,324	0,026	100,000
12-23	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	>90	Total
masc	0,792	16,173	16,089	9,614	5,445	2,665	1,082	0,253	0,011	52,125
fem	1,009	15,145	14,457	8,370	5,211	2,588	0,892	0,188	0,013	47,875
Total	1,801	31,318	30,547	17,984	10,657	5,253	1,974	0,441	0,025	100,000
24-35	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	>90	Total
masc	0,156	10,595	15,985	11,399	7,294	3,785	1,490	0,313	0,013	51,031
fem	0,164	10,841	15,333	10,409	7,163	3,634	1,163	0,249	0,014	48,969
Total	0,320	21,437	31,318	21,809	14,457	7,419	2,653	0,562	0,027	100,000
36-47	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	>90	Total
masc	0,009	4,389	12,969	11,551	8,470	5,050	2,165	0,433	0,018	45,055
fem	0,016	5,394	15,716	14,176	11,492	6,039	1,814	0,289	0,009	54,945
Total	0,025	9,783	28,686	25,727	19,962	11,089	3,979	0,722	0,027	100,000
48-58	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	>90	Total
masc	0,000	3,301	16,474	14,813	8,798	4,251	1,604	0,229	0,005	49,475
fem	0,006	2,957	14,672	14,848	11,708	5,091	1,107	0,130	0,006	50,525
Total	0,007	6,259	31,146	29,660	20,505	9,342	2,711	0,359	0,011	100,000

Foram calculadas tabelas com a probabilidade de ocorrência dos valores da base de dados de fidelidade (pf_temp.txt) com tabulações para as variáveis atividade, idade, RA e sexo, tabuladas duas a duas, ao longo das cinco faixas de fidelidade. Os gráficos plotam estas tabulações ao longo das cinco faixas de fidelidade. As informações dos gráficos e tabelas são todas em porcentagem.

A Tabela 8.10 apresenta a análise da probabilidade de ocorrência das variáveis sexo e idade.

A Tabela 8.11 apresenta a análise da probabilidade de ocorrência das variáveis sexo e atividade.

A Tabela 8.12 apresenta a análise da probabilidade de ocorrência das variáveis RA e sexo.

A Tabela 8.13 apresenta a análise da probabilidade de ocorrência das variáveis RA e idade.

Tabela 8.11: Perfil fidelidade: Sexo x Atividade

1-11	masc	fem	Total
Alimentação	22,773	19,953	42,725
Varejo	20,547	21,872	42,420
Educação	0,348	0,376	0,724
Lazer	0,326	0,240	0,567
Serviços	4,981	4,113	9,094
Turismo	0,146	0,063	0,209
Saúde	1,975	2,286	4,262
Total	51,096	48,904	100,000

12-23	masc	fem	Total
Alimentação	25,641	20,831	46,472
Varejo	19,365	20,715	40,081
Educação	0,251	0,291	0,542
Lazer	0,280	0,188	0,469
Serviços	4,880	3,876	8,756
Turismo	0,062	0,037	0,099
Saúde	1,703	1,878	3,581
Total	52,183	47,817	100,000

24-35	masc	fem	Total
Alimentação	26,398	22,239	48,637
Varejo	17,396	20,096	37,492
Educação	0,239	0,268	0,507
Lazer	0,272	0,194	0,466
Serviços	5,124	4,281	9,405
Turismo	0,048	0,033	0,082
Saúde	1,596	1,815	3,411
Total	51,072	48,928	100,000

36-47	masc	fem	Total
Alimentação	23,859	25,182	49,040
Varejo	14,046	21,629	35,674
Educação	0,270	0,367	0,637
Lazer	0,248	0,234	0,482
Serviços	5,278	5,488	10,766
Turismo	0,032	0,034	0,066
Saúde	1,383	1,952	3,334
Total	45,114	54,886	100,000

48-58	masc	fem	Total
Alimentação	28,506	24,828	53,334
Varejo	13,324	18,073	31,397
Educação	0,377	0,428	0,804
Lazer	0,338	0,265	0,603
Serviços	5,709	5,337	11,046
Turismo	0,028	0,029	0,057
Saúde	1,231	1,527	2,757
Total	49,513	50,487	100,000

Tabela 8.12: Perfil fidelidade: RA x Sexo

1-11	Norte	Entorno	Outros	Oeste	Sul	Abast	Sudeste	Sudoeste	Centro	Total
masc	6,885	2,090	2,746	18,413	4,477	0,510	1,301	4,138	10,442	51,000
fem	7,063	1,758	1,881	18,951	4,847	0,456	1,374	3,938	8,733	49,000
Total	13,947	3,848	4,626	37,364	9,324	0,966	2,674	8,076	19,175	100,000

12-23	Norte	Entorno	Outros	Oeste	Sul	Abast	Sudeste	Sudoeste	Centro	Total
masc	6,483	1,242	1,400	20,412	4,174	0,374	1,460	4,899	11,678	52,123
fem	6,215	0,992	0,973	19,268	3,909	0,284	1,343	4,919	9,973	47,877
Total	12,699	2,234	2,373	39,680	8,083	0,659	2,803	9,818	21,651	100,000

24-35	Norte	Entorno	Outros	Oeste	Sul	Abast	Sudeste	Sudoeste	Centro	Total
masc	5,318	0,534	0,718	18,698	2,859	0,224	1,538	6,073	15,066	51,028
fem	5,202	0,513	0,581	18,050	2,709	0,184	1,467	6,318	13,949	48,972
Total	10,520	1,047	1,298	36,748	5,568	0,408	3,005	12,391	29,015	100,000

36-47	Norte	Entorno	Outros	Oeste	Sul	Abast	Sudeste	Sudoeste	Centro	Total
masc	3,083	0,148	0,337	12,080	1,266	0,158	1,506	6,248	20,230	45,057
fem	3,509	0,174	0,400	12,648	1,304	0,140	1,875	7,812	27,081	54,943
Total	6,592	0,322	0,737	24,728	2,570	0,298	3,381	14,061	47,311	100,000

48-58	Norte	Entorno	Outros	Oeste	Sul	Abast	Sudeste	Sudoeste	Centro	Total
masc	1,912	0,049	0,257	9,652	0,475	0,203	1,710	6,133	29,088	49,480
fem	1,617	0,055	0,224	6,554	0,356	0,125	1,446	6,324	33,820	50,520
Total	3,529	0,104	0,482	16,206	0,830	0,328	3,156	12,458	62,908	100,000

A Tabela 8.14 apresenta a análise da probabilidade de ocorrência das variáveis RA e atividade.

A Tabela 8.15 apresenta a análise da probabilidade de ocorrência das variáveis atividade e idade.

Os gráficos referentes às tabelas analisadas nessa seção estão apresentados no Anexo C.

A Tabela 8.16 apresenta a quantidade de pessoas físicas (quant PF), porcentagem de pessoas físicas (% PF), quantidade de documentos fiscais (quant doc) e porcentagem de documentos fiscais (% doc) por faixa de fidelidade.

Com base nas observações sobre as tabelas e gráficos foi possível extrair as informações que se seguem sobre as faixas de fidelidade.

A variável sexo teve distribuição de probabilidade equilibrada entre os valores masculino e feminino. Na faixa de fidelidade de 12-23 cabe ressaltar que na idade 0-20, o sexo feminino chega ser 20% maior que o masculino, no entanto para a Alimentação, o sexo masculino chega a ser 20% maior que o feminino. Na faixa de 36-47 em diante há inversão no sexo majoritário para a população feminina. É relevante notar que na faixa de fidelidade 36-47 a população feminina no grupo Centro concentra mais de 25% dos documentos fiscais e na faixa de fidelidade 48-58 mais de um terço dos documentos fiscais.

Tabela 8.13: Perfil fidelidade: RA x Idade

1-11	Norte	Entorno	Outros	Oeste	Sul	Abast	Sudeste	Sudoeste	Centro	Total
0-20	1,218	0,207	0,144	2,723	0,652	0,067	0,279	0,563	1,000	6,852
20-30	5,164	1,524	1,522	13,589	3,308	0,360	1,031	2,992	6,558	36,047
30-40	3,490	1,147	1,560	10,268	2,500	0,303	0,716	2,047	5,760	27,790
40-50	2,114	0,602	0,813	5,345	1,520	0,152	0,373	1,234	3,137	15,289
50-60	1,186	0,266	0,391	3,200	0,838	0,061	0,189	0,663	1,580	8,374
60-70	0,534	0,084	0,148	1,661	0,367	0,015	0,067	0,345	0,678	3,900
70-80	0,198	0,016	0,038	0,498	0,112	0,007	0,018	0,181	0,329	1,398
80-90	0,042	0,003	0,007	0,077	0,026	0,000	0,002	0,049	0,118	0,324
>90	0,003	0,000	0,004	0,002	0,001	0,000	0,000	0,002	0,015	0,026
Total	13,947	3,848	4,626	37,364	9,324	0,966	2,674	8,076	19,175	100,000

12-23	Norte	Entorno	Outros	Oeste	Sul	Abast	Sudeste	Sudoeste	Centro	Total
0-20	0,260	0,041	0,034	0,705	0,121	0,012	0,080	0,188	0,361	1,801
20-30	4,041	0,760	0,720	12,251	2,372	0,227	1,009	3,213	6,727	31,319
30-40	3,844	0,781	0,801	12,429	2,368	0,242	0,823	2,726	6,533	30,547
40-50	2,370	0,399	0,455	6,946	1,707	0,110	0,491	1,652	3,853	17,983
50-60	1,359	0,195	0,254	4,324	0,938	0,051	0,272	1,068	2,195	10,657
60-70	0,592	0,052	0,075	2,222	0,406	0,012	0,104	0,611	1,183	5,254
70-80	0,195	0,006	0,023	0,691	0,151	0,004	0,021	0,303	0,580	1,974
80-90	0,035	0,001	0,012	0,106	0,020	0,000	0,004	0,055	0,209	0,441
>90	0,003	0,000	0,000	0,007	0,001	0,000	0,000	0,003	0,012	0,025
Total	12,699	2,234	2,373	39,680	8,083	0,659	2,803	9,818	21,651	100,000

24-35	Norte	Entorno	Outros	Oeste	Sul	Abast	Sudeste	Sudoeste	Centro	Total
0-20	0,034	0,005	0,005	0,097	0,014	0,003	0,016	0,040	0,107	0,320
20-30	2,292	0,262	0,276	7,590	1,131	0,101	0,636	2,780	6,368	21,436
30-40	3,261	0,395	0,458	12,386	1,719	0,159	0,886	3,595	8,461	31,319
40-50	2,503	0,233	0,270	8,277	1,431	0,092	0,722	2,515	5,764	21,807
50-60	1,541	0,116	0,192	5,178	0,836	0,034	0,481	1,922	4,158	14,459
60-70	0,651	0,030	0,065	2,513	0,331	0,016	0,189	1,037	2,586	7,419
70-80	0,216	0,006	0,025	0,623	0,095	0,002	0,068	0,415	1,202	2,652
80-90	0,022	0,000	0,007	0,083	0,010	0,000	0,008	0,081	0,351	0,562
>90	0,000	0,000	0,001	0,001	0,000	0,000	0,000	0,005	0,020	0,027
Total	10,520	1,047	1,298	36,748	5,568	0,408	3,005	12,391	29,015	100,000

36-47	Norte	Entorno	Outros	Oeste	Sul	Abast	Sudeste	Sudoeste	Centro	Total
0-20	0,003	0,000	0,000	0,006	0,001	0,000	0,001	0,004	0,011	0,025
20-30	0,686	0,046	0,086	2,805	0,322	0,036	0,267	1,438	4,097	9,783
30-40	1,882	0,134	0,216	8,694	0,809	0,109	0,934	3,772	12,138	28,688
40-50	2,026	0,077	0,195	6,619	0,805	0,084	1,006	3,670	11,243	25,725
50-60	1,305	0,047	0,134	4,211	0,434	0,042	0,766	3,051	9,974	19,963
60-70	0,530	0,016	0,071	1,913	0,157	0,021	0,310	1,530	6,538	11,087
70-80	0,145	0,001	0,032	0,440	0,040	0,004	0,085	0,537	2,695	3,979
80-90	0,015	0,000	0,003	0,040	0,003	0,002	0,011	0,057	0,591	0,723
>90	0,000	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,025	0,027
Total	6,592	0,322	0,737	24,728	2,570	0,298	3,381	14,061	47,311	100,000

48-58	Norte	Entorno	Outros	Oeste	Sul	Abast	Sudeste	Sudoeste	Centro	Total
0-20	0,000	0,000	0,000	0,005	0,000	0,000	0,000	0,000	0,001	0,007
20-30	0,278	0,015	0,041	1,363	0,099	0,030	0,132	0,981	3,317	6,256
30-40	1,132	0,047	0,138	7,303	0,289	0,164	0,931	4,022	17,124	31,151
40-50	1,232	0,018	0,139	4,698	0,315	0,081	1,109	3,841	18,227	29,661
50-60	0,631	0,023	0,089	2,039	0,103	0,040	0,656	2,484	14,447	20,511
60-70	0,218	0,000	0,055	0,653	0,023	0,013	0,266	0,946	7,160	9,334
70-80	0,036	0,000	0,017	0,124	0,001	0,001	0,054	0,178	2,300	2,711
80-90	0,002	0,000	0,002	0,021	0,000	0,000	0,007	0,006	0,320	0,359
>90	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,011	0,011
Total	3,529	0,104	0,482	16,206	0,830	0,328	3,156	12,458	62,908	100,000

Tabela 8.14: Perfil fidelidade: RA x Atividade

1-11	Norte	Entorno	Outros	Oeste	Sul	Abast	Sudeste	Sudoeste	Centro	Total
Aliment	5,175	1,389	2,200	16,306	4,354	0,323	0,993	3,245	8,739	42,725
Varejo	6,355	1,765	1,734	16,525	3,626	0,394	1,362	3,351	7,311	42,421
Educação	0,094	0,021	0,025	0,324	0,038	0,004	0,031	0,053	0,136	0,724
Lazer	0,085	0,020	0,023	0,129	0,074	0,004	0,007	0,079	0,146	0,567
Serviços	1,576	0,424	0,417	2,523	0,828	0,154	0,180	0,947	2,044	9,093
Turismo	0,016	0,006	0,059	0,044	0,006	0,002	0,003	0,012	0,061	0,209
Saúde	0,715	0,229	0,177	1,333	0,374	0,087	0,115	0,437	0,793	4,261
Total	14,016	3,854	4,636	37,184	9,298	0,968	2,691	8,123	19,230	100,000

12-23	Norte	Entorno	Outros	Oeste	Sul	Abast	Sudeste	Sudoeste	Centro	Total
Aliment	5,139	0,921	1,139	19,196	3,933	0,233	1,136	4,283	10,494	46,473
Varejo	5,418	0,913	0,881	16,521	3,037	0,252	1,357	3,844	7,859	40,081
Educação	0,060	0,012	0,010	0,238	0,026	0,003	0,020	0,040	0,133	0,542
Lazer	0,067	0,013	0,012	0,112	0,050	0,004	0,010	0,068	0,133	0,469
Serviços	1,468	0,251	0,234	2,338	0,717	0,115	0,201	1,127	2,304	8,756
Turismo	0,009	0,003	0,007	0,031	0,005	0,001	0,001	0,010	0,031	0,099
Saúde	0,575	0,127	0,091	1,122	0,302	0,051	0,087	0,483	0,744	3,581
Total	12,735	2,239	2,374	39,558	8,070	0,658	2,812	9,856	21,698	100,000

24-35	Norte	Entorno	Outros	Oeste	Sul	Abast	Sudeste	Sudoeste	Centro	Total
Aliment	4,433	0,468	0,649	18,929	2,646	0,162	1,307	5,660	14,382	48,637
Varejo	4,132	0,385	0,458	14,313	2,030	0,149	1,350	4,499	10,176	37,492
Educação	0,052	0,005	0,005	0,187	0,016	0,002	0,021	0,050	0,168	0,507
Lazer	0,052	0,006	0,006	0,111	0,034	0,002	0,010	0,074	0,171	0,466
Serviços	1,360	0,127	0,134	2,111	0,606	0,068	0,238	1,569	3,192	9,405
Turismo	0,007	0,001	0,002	0,027	0,004	0,000	0,002	0,011	0,027	0,082
Saúde	0,499	0,057	0,044	0,988	0,227	0,025	0,081	0,558	0,931	3,411
Total	10,536	1,048	1,298	36,667	5,564	0,409	3,010	12,422	29,047	100,000

36-47	Norte	Entorno	Outros	Oeste	Sul	Abast	Sudeste	Sudoeste	Centro	Total
Aliment	2,775	0,145	0,369	13,110	1,138	0,127	1,546	6,526	23,305	49,041
Varejo	2,371	0,116	0,254	9,102	0,918	0,106	1,419	4,773	16,616	35,673
Educação	0,045	0,001	0,004	0,151	0,009	0,002	0,025	0,077	0,323	0,637
Lazer	0,033	0,002	0,003	0,085	0,015	0,002	0,015	0,074	0,253	0,482
Serviços	1,039	0,044	0,080	1,602	0,381	0,043	0,286	2,021	5,270	10,766
Turismo	0,004	0,000	0,001	0,017	0,002	0,000	0,002	0,009	0,031	0,066
Saúde	0,328	0,014	0,027	0,638	0,110	0,018	0,093	0,608	1,499	3,335
Total	6,595	0,322	0,737	24,706	2,572	0,298	3,385	14,089	47,296	100,000

48-58	Norte	Entorno	Outros	Oeste	Sul	Abast	Sudeste	Sudoeste	Centro	Total
Aliment	1,655	0,055	0,257	9,487	0,388	0,174	1,640	6,476	33,202	53,335
Varejo	1,072	0,031	0,153	5,004	0,267	0,090	1,115	3,704	19,962	31,398
Educação	0,036	0,001	0,003	0,129	0,008	0,002	0,025	0,098	0,503	0,804
Lazer	0,023	0,002	0,003	0,083	0,005	0,003	0,018	0,080	0,386	0,603
Serviços	0,583	0,012	0,051	1,167	0,139	0,037	0,281	1,713	7,063	11,045
Turismo	0,002	0,000	0,000	0,009	0,001	0,000	0,001	0,010	0,033	0,057
Saúde	0,153	0,003	0,014	0,320	0,024	0,022	0,078	0,400	1,743	2,757
Total	3,525	0,104	0,481	16,199	0,831	0,328	3,158	12,481	62,892	100,000

Tabela 8.15: Perfil fidelidade: Atividade x Idade

1-11	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	>90	Total
Aliment	2,914	15,717	11,999	6,371	3,467	1,596	0,534	0,122	0,006	42,725
Varejo	2,943	15,086	11,720	6,574	3,636	1,697	0,619	0,131	0,014	42,420
Educação	0,126	0,339	0,164	0,063	0,023	0,007	0,001	0,000	0,000	0,724
Lazer	0,062	0,280	0,136	0,051	0,024	0,009	0,004	0,001	0,000	0,567
Serviços	0,550	3,259	2,588	1,476	0,750	0,319	0,118	0,030	0,003	9,094
Turismo	0,005	0,056	0,074	0,041	0,023	0,009	0,002	0,001	0,000	0,209
Saúde	0,278	1,330	1,120	0,689	0,428	0,257	0,116	0,039	0,004	4,262
Total	6,878	36,066	27,801	15,264	8,350	3,894	1,395	0,325	0,027	100,000

12-23	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	>90	Total
Aliment	0,816	15,139	14,142	8,020	4,890	2,384	0,881	0,192	0,009	46,472
Varejo	0,731	12,066	12,310	7,377	4,351	2,203	0,842	0,189	0,012	40,081
Educação	0,026	0,235	0,159	0,086	0,027	0,007	0,001	0,000	0,000	0,542
Lazer	0,016	0,225	0,132	0,059	0,023	0,009	0,003	0,001	0,000	0,469
Serviços	0,149	2,709	2,701	1,692	0,905	0,408	0,152	0,038	0,002	8,756
Turismo	0,001	0,027	0,031	0,020	0,012	0,006	0,001	0,000	0,000	0,099
Saúde	0,061	0,927	1,073	0,722	0,443	0,236	0,096	0,023	0,002	3,581
Total	1,800	31,329	30,548	17,976	10,650	5,251	1,977	0,443	0,025	100,000

24-35	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	>90	Total
Aliment	0,159	11,081	15,378	10,250	6,775	3,488	1,224	0,267	0,014	48,637
Varejo	0,111	7,479	11,693	8,377	5,620	2,929	1,054	0,219	0,010	37,492
Educação	0,003	0,116	0,176	0,146	0,050	0,012	0,003	0,001	0,000	0,507
Lazer	0,004	0,164	0,163	0,079	0,035	0,015	0,005	0,001	0,000	0,466
Serviços	0,033	2,019	2,887	2,105	1,389	0,666	0,254	0,050	0,002	9,405
Turismo	0,000	0,015	0,026	0,019	0,013	0,006	0,002	0,000	0,000	0,082
Saúde	0,009	0,548	0,989	0,836	0,579	0,308	0,116	0,025	0,001	3,411
Total	0,319	21,423	31,312	21,811	14,461	7,425	2,658	0,563	0,027	100,000

36-47	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	>90	Total
Aliment	0,012	5,223	14,437	12,208	9,489	5,277	1,988	0,393	0,015	49,040
Varejo	0,009	3,152	10,123	9,366	7,271	4,101	1,402	0,241	0,009	35,674
Educação	0,000	0,050	0,195	0,266	0,095	0,024	0,006	0,001	0,000	0,637
Lazer	0,000	0,076	0,180	0,123	0,066	0,026	0,008	0,002	0,000	0,482
Serviços	0,003	1,018	2,854	2,824	2,309	1,259	0,434	0,062	0,003	10,766
Turismo	0,000	0,007	0,018	0,017	0,013	0,008	0,002	0,000	0,000	0,066
Saúde	0,001	0,230	0,862	0,928	0,731	0,410	0,147	0,026	0,001	3,334
Total	0,025	9,755	28,669	25,732	19,975	11,105	3,986	0,725	0,028	100,000

48-58	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	>90	Total
Aliment	0,003	3,801	17,295	15,465	10,469	4,698	1,403	0,194	0,006	53,334
Varejo	0,003	1,576	9,370	9,411	6,763	3,238	0,914	0,119	0,004	31,397
Educação	0,000	0,029	0,235	0,387	0,125	0,024	0,004	0,001	0,000	0,804
Lazer	0,000	0,050	0,226	0,192	0,091	0,035	0,008	0,001	0,000	0,603
Serviços	0,001	0,639	3,173	3,344	2,459	1,082	0,310	0,037	0,001	11,046
Turismo	0,000	0,005	0,018	0,015	0,012	0,005	0,001	0,000	0,000	0,057
Saúde	0,000	0,117	0,801	0,857	0,612	0,280	0,081	0,010	0,000	2,757
Total	0,007	6,218	31,118	29,671	20,530	9,362	2,722	0,361	0,011	100,000

Tabela 8.16: Quantitativo das faixas de fidelidade

faixa	quant PF	% PF	quant doc	% doc
1-11	141876	17,824	2.486.704	1,666
12-23	210971	26,504	13.974.911	9,365
24-35	221169	27,785	38.225.386	25,615
36-47	191470	24,054	72.792.393	48,779
48-58	30494	3,831	21.749.534	14,575

Na variável idade a média de idade tende a aumentar ao longo da fidelidade. Nas idades mais avançadas as porcentagens encontradas previamente diminuem na faixa 48-58.

Na variável atividade os grupos de Alimentação e Varejo são os maiores responsáveis pela fidelidade dos consumidores do PNL. Em uma intensidade menor os Serviços Gerais também tem um peso significativo. Há crescimento da Alimentação ao longo do tempo de fidelidade chegando a responder por mais de 50% dos documentos fiscais. Ao longo das faixas de fidelidade é significativo notar que um terço das emissões de documentos fiscais do PNL ocorrem para a atividade de Alimentação no Plano Piloto. Educação, Lazer e Turismo são pouco representativos não chegando a representar 1% da fidelidade ao longo do tempo.

A probabilidade baixa de fidelidade para Educação, Lazer e Turismo não é sinônimo de sonegação já que as tabelas refletem a quantidade de documentos fiscais solicitadas pela população. Esta probabilidade representa que nestas atividades econômicas, aquisições e serviços são realizados com menos frequência em comparação com as outras atividades econômicas. Trata-se de um dado de interesse para a gestão do PNL, em que estímulos poderiam ser realizados para que fossem solicitados mais documentos fiscais pela população nestas atividades econômicas.

Na variável RA pode-se constatar que a participação de pessoas físicas residentes no grupo Centro (Brasília, Lagos e Sudoeste/Octogonal) cresce ao longo do tempo de fidelidade em detrimento dos residentes no grupo Oeste (Águas Claras, Taguatinga, Vicente Pires, Ceilândia, Recanto das Emas, Riacho Fundo, Riacho Fundo II e Samambaia). As percentagens de pessoas cadastradas no PNL de outros Estados ou residentes no Entorno decrescem ao longo das faixas de fidelidade. A participação de pessoas de outros Estados é sempre maior do que as do Entorno. Como a população destes locais tem seu cotidiano no Distrito Federal esperava-se maior fidelidade destes locais.

É uma contradição que as pessoas de outros Estados peçam mais documentos fiscais que a população do Entorno. Para tentar entender porque esta situação ocorre, foi feita uma comparação entre as RA obtidas com os dados da RFB e com os dados da tabela Beneficiario para cada pessoa física. Eliminado os casos de erros, *missing values* e *outliers*

relatados previamente, 9345 das 18434 pessoas físicas cadastradas em outros Estados pela RFB, se encontram residentes no Distrito Federal pela tabela Beneficiario. Conclui-se que pode haver distorção nas informações fornecidas à SEF, o que agrava ainda mais os problemas de endereçamento postal.

8.4 Estudo de caso: Perfil crédito

A pesquisa sobre obtenção de créditos de consumo pelos beneficiários do PNL envolve analisar condições em que os beneficiários receberam créditos ao participar do PNL.

Da mesma forma que na Seção 8.3, são analisadas, em maiores detalhes, as interações entre as variáveis ra_origem, atividade_economica_destino, idade e Sex_cpf_dados. Para facilitar o entendimento das tabelas e gráficos, as variáveis serão apresentados respectivamente como RA, Atividade, Idade e Sexo. Entende-se que estas variáveis são fatores determinantes para o comportamento dos contribuintes no PNL.

Conforme apresentado na Seção 7.5 análises sobre os créditos devem ser realizados separadamente para ICMS, ISS, pessoas físicas e pessoas jurídicas.

8.4.1 Divisão em faixas de crédito

Como primeiro passo para avaliação da quantidade de créditos obtidos pelos beneficiários, foram geradas as seguintes estatísticas para cada valor encontrado de documento fiscal: quantidade de documentos com o mesmo valor de crédito (freq), probabilidade de ocorrência (prob), probabilidade acumulada de ocorrência (prob acum), média (media), desvio padrão (dp) e coeficiente de variação (cv).

Ao separar os valores entre pessoas físicas e jurídicas e entre ICMS e ISS para avaliação, em cerca de 14.000.000 de documentos fiscais não foi possível distinguir se o crédito 0 era referente a um documento fiscal de ISS ou de ICMS. A empresa declarou à SEF o documento fiscal com valor maior que 0, no entanto, não houve declaração sobre recolhimento do imposto para este valor. Logo não houve crédito do PNL a ser concedido para o beneficiário. Como não foi possível distinguir entre o ISS e o ICMS, não foi possível categorizar entre as quatro tabelas para estes casos. Desta forma, eles foram excluídos das bases de dados de crédito de consumo. Por este motivo e pelas razões apresentadas em 8.4.2 decidiu-se analisar separadamente o caso de crédito 0.

Valores de documentos fiscais que apresentavam estatísticas similares foram agrupados em faixas de valores até o limite de R$ 1.000 em créditos. R$ 1.000 em crédito por documento fiscal é o teto limite esperado pela SEF no cálculo de benefícios. Documentos fiscais com créditos acima de R$ 1.000 precisam ser comprovados pelo beneficiário conforme Inciso I do Artigo 13 da Lei nº 4.159, em [14].

Tabela 8.17: Estatísticas crédito com ISS de pessoa física

PF cred	freq	prob %	prob acum %	media	dp	cv
0	736586	7,239	7,239	0,000	0,000	
0,01-1	5693744	55,959	63,198	0,273	0,273	100,000
1,01-2	1502470	14,766	77,965	1,454	0,283	19,464
2,01-3	707625	6,955	84,919	2,464	0,303	12,297
3,01-4	344899	3,390	88,309	3,496	0,281	8,038
4,01-5	238914	2,348	90,657	4,469	0,287	6,422
5,01-6	219453	2,157	92,814	5,539	0,314	5,669
6,01-7	140257	1,378	94,192	6,467	0,285	4,407
7,01-8	111008	1,091	95,283	7,479	0,275	3,677
8,01-9	105916	1,041	96,324	8,522	0,335	3,931
9,01-10	47508	0,467	96,791	9,503	0,287	3,020
10,01-11	44167	0,434	97,225	10,487	0,281	2,680
11,01-12	40553	0,399	97,624	11,546	0,328	2,841
12,01-13	27030	0,266	97,889	12,503	0,273	2,183
13,01-14	23716	0,233	98,122	13,500	0,276	2,044
14,01-15	24602	0,242	98,364	14,587	0,335	2,297
15,01-16	14286	0,140	98,505	15,527	0,287	1,848
16,01-17	14406	0,142	98,646	16,475	0,277	1,681
17,01-18	14634	0,144	98,790	17,599	0,347	1,972
18,01-19	10391	0,102	98,892	18,574	0,283	1,524
19,01-20	9111	0,090	98,982	19,487	0,290	1,488
20,01-30	54439	0,535	99,517	24,321	2,985	12,273
30,01-40	17941	0,176	99,693	34,522	2,842	8,232
40,01-50	10614	0,104	99,797	44,591	2,900	6,504
50,01-60	6612	0,065	99,862	54,994	3,195	5,810
60,01-70	3138	0,031	99,893	64,846	2,894	4,463
70,01-80	2676	0,026	99,920	74,489	2,789	3,744
80,01-90	1877	0,018	99,938	85,504	3,263	3,816
90,01-100	1134	0,011	99,949	94,623	2,798	2,957
100,01-200	3727	0,037	99,986	133,416	26,408	19,794
200,01-300	781	0,008	99,993	241,947	28,984	11,979
300,01-400	302	0,003	99,996	344,478	30,392	8,823
400,01-500	129	0,001	99,998	437,261	27,962	6,395
500,01-750	144	0,001	99,999	595,346	70,343	11,815
750,01-1000	30	0,000	99,999	859,688	91,280	10,618
>1000	60	0,001	100,000	2202,171	2531,218	114,942

Tabela 8.18: Estatísticas crédito com ICMS de pessoa física

PF cred	freq	prob %	prob acum %	media	dp	cv
0	15015660	12,072	12,072	0,000	0,000	
0,01-1	62670374	50,383	62,455	0,268	0,295	110,075
1,01-2	16138672	12,974	75,429	1,571	0,326	20,751
2,01-3	8054399	6,475	81,904	2,384	0,306	12,836
3,01-4	4873060	3,918	85,822	3,466	0,292	8,425
4,01-5	3349039	2,692	88,514	4,466	0,292	6,538
5,01-6	2488426	2,001	90,515	5,469	0,301	5,504
6,01-7	1807462	1,453	91,968	6,479	0,292	4,507
7,01-8	1472140	1,184	93,152	7,474	0,287	3,840
8,01-9	1169784	0,940	94,092	8,487	0,302	3,558
9,01-10	908515	0,730	94,822	9,482	0,289	3,048
10,01-11	786360	0,632	95,455	10,468	0,290	2,770
11,01-12	662749	0,533	95,987	11,491	0,301	2,619
12,01-13	535515	0,431	96,418	12,494	0,288	2,305
13,01-14	465932	0,375	96,792	13,489	0,288	2,135
14,01-15	427760	0,344	97,136	14,515	0,308	2,122
15,01-16	341468	0,275	97,411	15,492	0,288	1,859
16,01-17	299503	0,241	97,652	16,494	0,287	1,740
17,01-18	267110	0,215	97,866	17,509	0,304	1,736
18,01-19	232064	0,187	98,053	18,513	0,292	1,577
19,01-20	195526	0,157	98,210	19,491	0,288	1,478
20,01-30	1166500	0,938	99,148	24,154	2,912	12,056
30,01-40	459340	0,369	99,517	34,342	2,896	8,433
40,01-50	223757	0,180	99,697	44,423	2,926	6,587
50,01-60	122503	0,098	99,796	54,557	3,018	5,532
60,01-70	70778	0,057	99,853	64,589	2,931	4,538
70,01-80	46157	0,037	99,890	74,613	2,917	3,910
80,01-90	32474	0,026	99,916	84,976	3,181	3,743
90,01-100	21984	0,018	99,933	94,752	2,951	3,114
100,01-200	64305	0,052	99,985	132,045	26,966	20,422
200,01-300	10118	0,008	99,993	239,168	28,862	12,068
300,01-400	3476	0,003	99,996	342,772	28,507	8,317
400,01-500	1622	0,001	99,997	445,100	28,653	6,437
500,01-750	1777	0,001	99,999	602,099	71,885	11,939
750,01-1000	627	0,001	99,999	847,421	73,446	8,667
>1000	923	0,001	100,000	2736,140	8062,117	294,653

Tabela 8.19: Estatísticas crédito com ISS de pessoa jurídica

PJ cred	freq	prob %	prob acum %	media	dp	cv
0	93705	18,457	18,457	0,000	0,000	
0,01-1	119114	23,462	41,919	0,462	0,269	58,225
1,01-2	87545	17,244	59,163	1,495	0,283	18,930
2,01-3	60800	11,976	71,139	2,461	0,295	11,987
3,01-4	25087	4,941	76,080	3,507	0,273	7,784
4,01-5	22014	4,336	80,417	4,479	0,267	5,961
5,01-6	14054	2,768	83,185	5,520	0,312	5,652
6,01-7	9183	1,809	84,994	6,566	0,298	4,539
7,01-8	7739	1,524	86,518	7,506	0,254	3,384
8,01-9	9324	1,837	88,354	8,523	0,333	3,907
9,01-10	6720	1,324	89,678	9,488	0,258	2,719
10,01-11	5301	1,044	90,722	10,442	0,279	2,672
11,01-12	4394	0,865	91,588	11,546	0,325	2,815
12,01-13	3247	0,640	92,227	12,515	0,262	2,093
13,01-14	3693	0,727	92,955	13,477	0,312	2,315
14,01-15	2917	0,575	93,529	14,579	0,317	2,174
15,01-16	2271	0,447	93,977	15,474	0,278	1,797
16,01-17	1959	0,386	94,362	16,488	0,259	1,571
17,01-18	2179	0,429	94,792	17,615	0,325	1,845
18,01-19	2150	0,423	95,215	18,607	0,252	1,354
19,01-20	1328	0,262	95,477	19,463	0,273	1,403
20,01-30	9150	1,802	97,279	24,927	3,051	12,240
30,01-40	6334	1,248	98,527	34,174	2,704	7,912
40,01-50	2248	0,443	98,969	44,202	2,816	6,371
50,01-60	1032	0,203	99,173	55,143	3,218	5,836
60,01-70	807	0,159	99,332	65,505	2,869	4,380
70,01-80	554	0,109	99,441	74,622	2,640	3,538
80,01-90	419	0,083	99,523	86,058	3,201	3,720
90,01-100	268	0,053	99,576	95,172	2,870	3,016
100,01-200	1256	0,247	99,824	136,710	26,779	19,588
200,01-300	420	0,083	99,906	244,208	31,123	12,744
300,01-400	172	0,034	99,940	342,536	28,363	8,280
400,01-500	76	0,015	99,955	451,744	29,906	6,620
500,01-750	124	0,024	99,980	602,228	75,799	12,586
750,01-100	48	0,009	99,989	864,821	66,472	7,686
>1000	56	0,011	100,000	1627,745	659,204	40,498

Tabela 8.20: Estatísticas crédito com ICMS de pessoa jurídica

PJ cred	freq	prob %	prob acum %	media	dp	cv
0	521213	34,462	34,462	0,000	0,000	
0,01-1	299122	19,777	54,239	0,406	0,294	72,414
1,01-2	147215	9,734	63,973	1,464	0,288	19,672
2,01-3	97954	6,477	70,449	2,477	0,289	11,667
3,01-4	70359	4,652	75,101	3,482	0,287	8,242
4,01-5	54062	3,574	78,676	4,480	0,290	6,473
5,01-6	42104	2,784	81,459	5,485	0,291	5,305
6,01-7	34075	2,253	83,712	6,487	0,287	4,424
7,01-8	27314	1,806	85,518	7,486	0,285	3,807
8,01-9	23071	1,525	87,044	8,496	0,292	3,437
9,01-10	18932	1,252	88,295	9,488	0,286	3,014
10,01-11	16457	1,088	89,384	10,488	0,288	2,746
11,01-12	13821	0,914	90,297	11,494	0,293	2,549
12,01-13	11478	0,759	91,056	12,492	0,288	2,305
13,01-14	9995	0,661	91,717	13,493	0,289	2,142
14,01-15	9042	0,598	92,315	14,504	0,293	2,020
15,01-16	7773	0,514	92,829	15,493	0,287	1,852
16,01-17	7010	0,463	93,292	16,490	0,285	1,728
17,01-18	6083	0,402	93,695	17,504	0,297	1,697
18,01-19	5703	0,377	94,072	18,517	0,284	1,534
19,01-20	4731	0,313	94,384	19,500	0,291	1,492
20,01-30	30291	2,003	96,387	24,327	2,874	11,814
30,01-40	14972	0,990	97,377	34,501	2,853	8,269
40,01-50	8671	0,573	97,950	44,481	2,853	6,414
50,01-60	5640	0,373	98,323	54,717	2,941	5,375
60,01-70	3995	0,264	98,588	64,756	2,850	4,401
70,01-80	3018	0,200	98,787	74,925	2,878	3,841
80,01-90	2578	0,170	98,958	85,239	3,078	3,611
90,01-100	1808	0,120	99,077	94,825	2,918	3,077
100,01-200	8025	0,531	99,608	139,049	28,144	20,240
200,01-300	2530	0,167	99,775	244,181	28,367	11,617
300,01-400	1276	0,084	99,859	344,632	28,055	8,141
400,01-500	651	0,043	99,902	445,548	28,558	6,410
500,01-750	811	0,054	99,956	604,568	70,552	11,670
750,01-1000	323	0,021	99,977	854,746	69,475	8,128
>1000	343	0,023	100,000	1859,160	1590,982	85,575

A Tabela 8.17 apresenta os valores calculados para a quantidade de créditos obtidos por pessoas físicas no ISS.

A Tabela 8.18 apresenta os valores calculados para a quantidade de créditos obtidos por pessoas físicas no ICMS.

A Tabela 8.19 apresenta os valores calculados para a quantidade de créditos obtidos por pessoas jurídicas no ISS.

A Tabela 8.20 apresenta os valores calculados para a quantidade de créditos obtidos por pessoas jurídicas no ICMS.

Ao analisar as Tabelas 8.17, 8.18, 8.19 e 8.20, as informações que se seguem foram objetos de análise.

O coeficiente de variação para valores de créditos acima de R$ 1.000 é atípico com relação aos outros coeficientes obtidos, pois apresentam alta dispersão de dados. Por este motivo e pelas razões apresentadas em 8.4.2 decidiu-se analisar separadamente este caso.

É possível perceber que a distribuição acumulada alcança 99% da distribuição para créditos até R$ 20 de pessoas físicas e alcança 95% da distribuição para créditos até R$ 20 de pessoas jurídicas. Como o crescimento dos créditos acontece de forma diferenciada até este limite, decidiu-se por categorizar separadamente os casos até R$ 20 e acima de R$ 20.

Mesmo com coeficiente de variação indicando alta dispersão de dados para valores encontrados entre R$ 0,01 e R$ 1 e cerca de 20% da base de dados neste faixa, decidiu-se por analisá-la até o valor de R$ 20. Levando-se em conta o cálculo do crédito do PNL, como apresentado em 3.3.2, foi verificado na base de dados que mesmo para documentos fiscais de mesmo valor, suponha-se R$ 30, dependendo da atividade econômica da empresa em que foi feito a aquisição do bem ou do serviço, o valor do crédito calculado poderia ter valores que variam entre R$ 0,01 a R$ 2. Há também a hipótese apresentada em 8.4.2 que reduz o crédito do beneficiário. Desta forma, não houve alterações para este caso.

Observa-se que o número de documentos fiscais associados ao ISS são menos que 10% dos dados de pessoas físicas. Enquanto que para pessoas jurídicas o número de documentos fiscais associados ao ISS representam cerca de 25% dos dados.

Cabe ainda verificar que a quantidade de documentos fiscais de pessoas jurídicas representa apenas 1,5% do total de documentos fiscais da base de dados.

Desta forma, para as análises seguintes será considerada a divisão dos créditos em quatro faixas de crédito: 0, onde não há créditos ao beneficiários; 0-20, valor entre R$0,01 a R$20 de créditos; 20-1000, valor de R$20,01 a R$1.000 de créditos; e >1000 para valores acima de R$1.000. Por faixa de crédito entende-se a quantidade de crédito de consumo obtida nos documentos fiscais para grupos de beneficiários.

8.4.2 Análises de crédito zero ou superior a um mil reais

A respeito da quantidade de documentos fiscais encontrados na base de dados do PNL com crédito zero, sua justificativa se deve, em essência, aos seguintes fatores: inadimplência das empresas, possibilidade de estar em curso uma fiscalização tributária e existência de créditos tributários do ICMS para a empresa emissora do documento fiscal.

Deve-se levar em conta os princípios constitucionais quanto à despesa pública em que não é válido criar despesa sem que seja criado uma receita associada.

Os fatores de inadimplência das empresas participantes e dos créditos tributários são consequências direta do Artigo 3º da Lei nº 4.159[2], na íntegra em [14]; ou seja, somente são concedidos créditos quando as empresas participantes recolherem o imposto devido. No caso de inadimplência das empresas participantes, há recusa pela empresa no recolhimento de suas obrigações tributárias. Para o caso de existência de créditos tributários, devido à legislação do ICMS, e conforme apresentado em 3.2, é possível a uma empresa, em um determinado período, não recolher imposto à SEF caso já tenha pago um montante superior ao devido em momento anterior. Ainda existem casos, em que as empresas podem gozar de isenções ou substituições tributárias, reduzindo parcialmente ou totalmente a quantidade de créditos disponibilizados aos contribuintes.

Outra hipótese é a abertura para realização da fiscalização tributária, conforme Artigo 13 da Portaria nº 4[3], na íntegra em [20]. Caso a empresa esteja sob investigação da SEF, não serão concedidos créditos no período da investigação da contabilidade da empresa.

8.4.3 Análise das variáveis selecionadas para pessoas físicas

Foram calculadas tabelas com a probabilidade de ocorrência dos valores da base de dados de crédito (icms_pf.txt e iss_pf.txt) com tabulações para as variáveis atividade, idade, RA e sexo, tabuladas duas a duas, ao longo das quatro faixas de crédito. Os gráficos plotam estas tabulações ao longo das quatro faixas de crédito. As informações dos gráficos e tabelas são todas em porcentagem.

[2]1. "LEI Nº 4.159, DE 13 DE JUNHO DE 2008. ... Art. 3º O beneficiário do programa, adquirente ou tomador, fará jus ao valor de até 30% (trinta por cento) do ICMS ou do ISS efetivamente recolhido pelo estabelecimento fornecedor ou prestador. ... § 2º Os créditos previstos neste artigo não serão concedidos: IX - na hipótese de documento: a) inidôneo; d) emitido mediante fraude, dolo ou simulação."

[3]2. "PORTARIA Nº 4, DE 4 DE JANEIRO DE 2012. ... Art. 13. A SEF poderá efetuar o bloqueio de créditos consolidados nas seguintes hipóteses: I - de valor superior a R$ 1.000,00 (um mil reais), provenientes de um único documento fiscal; II - provenientes de elevado número de registros de documentos fiscais emitidos por um determinado contribuinte do Nota Legal que identifique um mesmo adquirente; III - de forma preventiva, quando houver indício de irregularidade ou fraude. § 1º Para fins de desbloqueio do crédito a que se refere o caput, o adquirente deverá apresentar o original ou cópia autenticada, em qualquer Agência de Atendimento da Receita, em até 10 (dez) dias antes de expirado o prazo para indicar os veículos e (ou) imóveis sobre os quais deverá ser efetuado o abatimento do IPTU e (ou) do IPVA, observado o prazo de prescrição do crédito."

Tabela 8.21: Perfil crédito ICMS PF: Sexo x Atividade

0	Alimentação	Varejo	Educação	Lazer	Serviços	Turismo	Saúde	Total
masc	31,650	13,397	0,011	0,086	6,195	0,012	0,411	51,762
fem	26,587	15,564	0,012	0,072	5,635	0,006	0,362	48,238
Total	58,237	28,961	0,023	0,158	11,83	0,018	0,773	100,000
0-20	Alimentação	Varejo	Educação	Lazer	Serviços	Turismo	Saúde	Total
masc	28,554	13,534	0,021	0,117	4,217	0,008	0,802	47,253
fem	27,734	20,088	0,019	0,087	3,983	0,005	0,831	52,747
Total	56,288	33,622	0,04	0,204	8,2	0,013	1,633	100,000
20-1000	Alimentação	Varejo	Educação	Lazer	Serviços	Turismo	Saúde	Total
masc	8,673	25,904	0,008	0,151	11,643	0,002	1,131	47,512
fem	7,587	32,273	0,005	0,116	11,496	0,001	1,01	52,488
Total	16,26	58,177	0,013	0,267	23,139	0,003	2,141	100,000
>1000	Alimentação	Varejo	Educação	Lazer	Serviços	Turismo	Saúde	Total
masc	10,831	29,723	0,000	0,000	7,822	0,000	6,378	54,753
fem	8,544	30,806	0,000	0,000	1,805	0,000	4,091	45,246
Total	19,375	60,529	0,000	0,000	9,627	0,000	10,469	100,000

Tabela 8.22: Perfil crédito ICMS PF: RA x Sexo

0	Norte	Entorno	Outros	Oeste	Sul	Abastec	Sudeste	Sudoeste	Centro	Total
masc	5,538	0,369	0,585	15,397	3,503	0,203	1,275	7,079	17,723	51,672
fem	4,991	0,292	0,445	13,270	2,998	0,134	1,172	6,909	18,117	48,326
Total	10,529	0,661	1,030	28,667	6,501	0,337	2,447	13,988	35,840	100,000
0-20	Norte	Entorno	Outros	Oeste	Sul	Abastec	Sudeste	Sudoeste	Centro	Total
masc	3,636	0,353	0,557	14,150	1,687	0,206	1,565	5,838	19,212	47,204
fem	3,913	0,354	0,511	14,143	1,750	0,173	1,732	6,925	23,295	52,797
Total	7,549	0,707	1,068	28,293	3,437	0,379	3,297	12,763	42,507	100,000
20-1000	Norte	Entorno	Outros	Oeste	Sul	Abastec	Sudeste	Sudoeste	Centro	Total
masc	4,649	0,452	0,551	14,259	1,997	0,246	1,545	5,592	18,059	47,350
fem	4,675	0,402	0,493	13,800	2,012	0,195	1,696	6,489	22,888	52,650
Total	9,324	0,854	1,044	28,059	4,009	0,441	3,241	12,081	40,947	100,000
>1000	Norte	Entorno	Outros	Oeste	Sul	Abastec	Sudeste	Sudoeste	Centro	Total
masc	2,925	0,758	0,217	12,135	0,867	0,108	2,059	5,092	30,011	54,171
fem	2,275	0,650	0,217	5,850	0,325	0,108	1,733	4,442	30,228	45,828
Total	5,200	1,408	0,434	17,985	1,192	0,216	3,792	9,534	60,239	100,000

Tabela 8.23: Perfil crédito ICMS PF: RA x Idade

0	Norte	Entorno	Outros	Oeste	Sul	Abast	Sudeste	Sudoeste	Centro	Total
0-20	0,051	0,004	0,007	0,108	0,030	0,002	0,009	0,029	0,053	0,293
20-30	1,625	0,159	0,192	4,645	1,082	0,062	0,311	1,886	4,061	14,023
30-40	3,019	0,246	0,334	9,725	1,924	0,128	0,744	3,793	9,942	29,851
40-50	2,878	0,148	0,239	6,908	1,824	0,083	0,689	3,422	8,487	24,678
50-60	1,871	0,078	0,151	4,502	1,061	0,040	0,460	2,773	6,977	17,913
60-70	0,807	0,023	0,077	2,189	0,434	0,018	0,183	1,502	4,263	9,496
70-80	0,251	0,002	0,026	0,525	0,130	0,003	0,046	0,523	1,706	3,212
80-90	0,025	0,000	0,004	0,064	0,015	0,000	0,005	0,058	0,339	0,510
>90	0,001	0,000	0,000	0,002	0,000	0,000	0,000	0,001	0,016	0,020
Total	10,528	0,660	1,030	28,668	6,500	0,336	2,447	13,987	35,844	100,000

0-20	Norte	Entorno	Outros	Oeste	Sul	Abast	Sudeste	Sudoeste	Centro	Total
0-20	0,056	0,009	0,007	0,142	0,026	0,003	0,018	0,040	0,086	0,387
20-30	1,427	0,192	0,216	4,966	0,710	0,078	0,444	1,913	4,997	14,943
30-40	2,275	0,264	0,348	9,944	1,069	0,148	0,932	3,626	11,451	30,057
40-50	1,967	0,144	0,240	6,793	0,904	0,088	0,918	3,128	10,106	24,286
50-60	1,194	0,076	0,158	4,102	0,489	0,040	0,644	2,413	8,339	17,455
60-70	0,477	0,021	0,067	1,853	0,183	0,017	0,259	1,177	5,054	9,108
70-80	0,136	0,003	0,027	0,443	0,051	0,003	0,072	0,409	2,013	3,157
80-90	0,015	0,000	0,005	0,051	0,005	0,001	0,009	0,054	0,443	0,583
>90	0,000	0,000	0,000	0,001	0,000	0,000	0,000	0,002	0,019	0,022
Total	7,547	0,709	1,068	28,295	3,437	0,378	3,296	12,762	42,508	100,000

20-1000	Norte	Entorno	Outros	Oeste	Sul	Abast	Sudeste	Sudoeste	Centro	Total
0-20	0,051	0,007	0,004	0,136	0,028	0,003	0,010	0,029	0,060	0,328
20-30	1,441	0,198	0,195	4,703	0,714	0,077	0,362	1,559	3,980	13,229
30-40	2,717	0,306	0,320	9,335	1,198	0,170	0,957	3,199	10,565	28,767
40-50	2,712	0,202	0,269	7,315	1,165	0,109	0,993	3,409	11,068	27,241
50-60	1,615	0,107	0,165	4,346	0,600	0,056	0,640	2,484	8,845	18,858
60-70	0,606	0,029	0,064	1,797	0,236	0,023	0,221	1,031	4,511	8,518
70-80	0,165	0,004	0,023	0,390	0,062	0,004	0,052	0,331	1,579	2,610
80-90	0,016	0,000	0,004	0,037	0,007	0,000	0,005	0,038	0,323	0,430
>90	0,000	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,017	0,018
Total	9,323	0,853	1,044	28,060	4,010	0,442	3,240	12,080	40,948	100,000

>1000	Norte	Entorno	Outros	Oeste	Sul	Abast	Sudeste	Sudoeste	Centro	Total
0-20	0,108	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,108
20-30	0,433	0,433	0,000	2,059	0,217	0,000	0,000	0,433	2,600	6,175
30-40	1,083	0,108	0,108	7,259	0,433	0,000	1,192	2,925	12,788	25,892
40-50	2,817	0,217	0,000	3,684	0,325	0,217	1,300	2,600	17,551	28,711
50-60	0,325	0,542	0,325	2,817	0,217	0,000	0,650	1,950	12,784	19,610
60-70	0,325	0,000	0,000	1,408	0,000	0,000	0,325	0,867	11,701	14,626
70-80	0,108	0,108	0,000	0,650	0,000	0,000	0,325	0,542	2,384	4,117
80-90	0,000	0,000	0,000	0,108	0,000	0,000	0,000	0,217	0,325	0,650
>90	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,108	0,108
Total	5,199	1,408	0,433	17,985	1,192	0,217	3,792	9,534	60,240	100,000

Tabela 8.24: Perfil crédito ICMS PF: RA x Atividade

0	Norte	Entorno	Outros	Oeste	Sul	Abast	Sudeste	Sudoeste	Centro	Total
Aliment	5,928	0,393	0,632	17,920	3,704	0,186	1,058	8,352	20,065	58,238
Varejo	2,574	0,185	0,274	8,412	0,996	0,104	1,168	3,595	11,651	28,959
Educação	0,001	0,000	0,000	0,003	0,000	0,000	0,001	0,002	0,018	0,025
Lazer	0,008	0,001	0,001	0,029	0,003	0,000	0,005	0,017	0,095	0,159
Serviços	1,975	0,072	0,111	2,198	1,775	0,036	0,192	1,903	3,568	11,830
Turismo	0,001	0,000	0,002	0,003	0,000	0,000	0,000	0,001	0,010	0,017
Saúde	0,120	0,012	0,010	0,095	0,041	0,011	0,015	0,194	0,274	0,772
Total	10,607	0,663	1,030	28,660	6,519	0,337	2,439	14,064	35,681	100,000

0-20	Norte	Entorno	Outros	Oeste	Sul	Abast	Sudeste	Sudoeste	Centro	Total
Aliment	3,513	0,344	0,598	16,689	1,780	0,177	1,741	6,722	24,725	56,289
Varejo	2,828	0,260	0,352	10,068	1,285	0,124	1,313	4,123	13,268	33,620
Educação	0,003	0,001	0,000	0,005	0,000	0,000	0,001	0,003	0,027	0,040
Lazer	0,024	0,002	0,002	0,035	0,012	0,001	0,007	0,022	0,099	0,204
Serviços	0,963	0,077	0,093	1,285	0,279	0,056	0,202	1,525	3,720	8,200
Turismo	0,001	0,000	0,001	0,002	0,000	0,000	0,000	0,002	0,006	0,012
Saúde	0,222	0,023	0,021	0,200	0,078	0,021	0,036	0,389	0,644	1,634
Total	7,554	0,707	1,067	28,284	3,434	0,379	3,300	12,786	42,489	100,000

20-1000	Norte	Entorno	Outros	Oeste	Sul	Abast	Sudeste	Sudoeste	Centro	Total
Aliment	1,022	0,112	0,155	6,414	0,381	0,057	0,514	1,484	6,121	16,260
Varejo	4,763	0,519	0,621	16,172	2,267	0,266	1,984	6,702	24,883	58,177
Educação	0,000	0,000	0,000	0,001	0,000	0,000	0,000	0,001	0,010	0,012
Lazer	0,098	0,004	0,003	0,024	0,082	0,001	0,004	0,008	0,042	0,266
Serviços	3,020	0,186	0,232	4,658	0,990	0,107	0,721	3,534	9,692	23,136
Turismo	0,000	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,002	0,003
Saúde	0,367	0,028	0,024	0,197	0,111	0,006	0,063	0,528	0,818	2,142
Total	9,270	0,849	1,035	27,467	3,831	0,437	3,286	12,257	41,568	100,000

>1000	Norte	Entorno	Outros	Oeste	Sul	Abast	Sudeste	Sudoeste	Centro	Total
Aliment	0,481	0,000	0,241	6,619	0,361	0,120	1,444	0,481	9,627	19,374
Varejo	3,369	1,203	0,000	6,017	0,481	0,000	1,564	4,452	43,442	60,528
Educação	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Lazer	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Serviços	0,481	0,241	0,000	2,046	0,120	0,000	0,361	0,963	5,415	9,627
Turismo	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Saúde	0,963	0,000	0,241	1,083	0,241	0,000	0,241	3,129	4,573	10,471
Total	5,294	1,444	0,482	15,765	1,203	0,120	3,610	9,025	63,057	100,000

Tabela 8.25: Perfil crédito ICMS PF: Atividade x Idade

0	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	>90	Total
Aliment	0,171	8,823	18,085	14,069	9,992	5,108	1,721	0,259	0,009	58,237
Varejo	0,087	3,662	8,530	7,353	5,377	2,885	0,911	0,149	0,006	28,960
Educação	0,000	0,005	0,010	0,004	0,003	0,001	0,000	0,000	0,000	0,023
Lazer	0,000	0,036	0,066	0,027	0,019	0,008	0,003	0,000	0,000	0,159
Serviços	0,032	1,393	2,932	3,022	2,371	1,420	0,555	0,099	0,006	11,830
Turismo	0,000	0,003	0,006	0,005	0,003	0,001	0,000	0,000	0,000	0,018
Saúde	0,002	0,090	0,199	0,195	0,159	0,088	0,034	0,006	0,000	0,773
Total	0,292	14,012	29,828	24,675	17,924	9,511	3,224	0,513	0,021	100,000

0-20	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	>90	Total
Aliment	0,193	8,639	17,142	13,367	9,642	5,080	1,841	0,370	0,015	56,287
Varejo	0,158	4,943	10,171	8,362	5,848	3,007	0,964	0,161	0,006	33,620
Educação	0,000	0,005	0,013	0,011	0,007	0,003	0,001	0,000	0,000	0,040
Lazer	0,001	0,038	0,071	0,048	0,028	0,013	0,005	0,001	0,000	0,205
Serviços	0,030	1,099	2,202	2,079	1,607	0,845	0,292	0,044	0,002	8,200
Turismo	0,000	0,003	0,005	0,002	0,002	0,001	0,000	0,000	0,000	0,013
Total	0,004	0,193	0,439	0,426	0,328	0,173	0,061	0,009	0,000	1,633
	0,386	14,920	30,043	24,295	17,462	9,122	3,164	0,585	0,023	100,000

20-1000	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	>90	Total
Aliment	0,033	1,755	4,678	4,571	3,103	1,476	0,521	0,116	0,007	16,260
Varejo	0,198	8,157	17,302	15,465	10,634	4,821	1,386	0,210	0,008	58,176
Educação	0,000	0,001	0,004	0,003	0,002	0,002	0,000	0,000	0,000	0,012
Lazer	0,007	0,051	0,071	0,067	0,042	0,020	0,007	0,001	0,000	0,266
Serviços	0,075	2,854	6,208	6,621	4,670	1,997	0,617	0,094	0,003	23,139
Turismo	0,000	0,000	0,001	0,000	0,001	0,000	0,000	0,000	0,000	0,002
Saúde	0,003	0,177	0,508	0,646	0,474	0,232	0,087	0,013	0,000	2,140
Total	0,316	12,995	28,772	27,373	18,926	8,548	2,618	0,434	0,018	100,000

>1000	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	>90	Total
Aliment	0,000	2,046	4,573	4,452	3,249	3,249	1,444	0,361	0,000	19,374
Varejo	0,000	2,647	15,042	19,374	12,635	9,025	1,444	0,241	0,120	60,528
Educação	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Lazer	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Serviços	0,000	1,083	2,529	2,888	1,203	1,685	0,241	0,000	0,000	9,627
Turismo	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Saúde	0,120	0,722	1,805	2,166	2,768	1,805	0,963	0,120	0,000	10,469
Total	0,120	6,498	23,949	28,880	19,855	15,764	4,092	0,722	0,120	100,000

Tabela 8.26: Perfil crédito ICMS PF: Sexo x Idade

0	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	>90	Total
masc	0,138	7,291	15,614	12,752	8,746	4,855	1,948	0,316	0,013	51,673
fem	0,155	6,734	14,235	11,927	9,167	4,643	1,265	0,194	0,007	48,326
Total	0,293	14,025	29,849	24,679	17,913	9,498	3,213	0,510	0,020	100,000
0-20	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	>90	Total
masc	0,169	7,180	14,503	11,512	7,619	4,159	1,708	0,339	0,014	47,203
fem	0,217	7,765	15,551	12,774	9,834	4,952	1,449	0,245	0,010	52,798
Total	0,386	14,945	30,054	24,286	17,453	9,111	3,157	0,584	0,024	100,000
20-1000	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	>90	Total
masc	0,159	6,597	13,972	12,752	8,314	3,934	1,374	0,240	0,007	47,349
fem	0,169	6,630	14,792	14,491	10,541	4,590	1,237	0,190	0,011	52,651
Total	0,328	13,227	28,764	27,243	18,855	8,524	2,611	0,430	0,018	100,000
>1000	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	>90	Total
masc	0,000	4,225	14,301	14,086	10,184	8,559	2,492	0,325	0,000	54,171
fem	0,108	1,950	11,593	14,626	9,426	6,067	1,625	0,325	0,108	45,828
Total	0,108	6,175	25,894	28,712	19,610	14,626	4,117	0,650	0,108	100,000

Tabela 8.27: Perfil crédito ISS PF: Sexo x Idade

0	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	>90	Total
masc	0,246	6,812	14,164	11,981	7,687	4,018	1,443	0,295	0,010	46,655
fem	0,215	7,448	16,838	13,609	9,529	4,364	1,161	0,171	0,009	53,344
Total	0,461	14,260	31,002	25,590	17,216	8,382	2,604	0,466	0,019	100,000
0-20	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	>90	Total
masc	0,194	6,915	14,072	12,252	7,647	3,766	1,433	0,268	0,014	46,561
fem	0,221	7,546	16,711	14,083	9,408	4,089	1,161	0,210	0,010	53,439
Total	0,415	14,461	30,783	26,335	17,055	7,855	2,594	0,478	0,024	100,000
20-1000	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	>90	Total
masc	0,092	5,401	12,480	12,780	9,389	5,751	2,662	0,568	0,050	49,173
fem	0,095	5,684	14,505	12,728	9,750	5,600	2,009	0,435	0,021	50,829
Total	0,187	11,085	26,985	25,508	19,139	11,351	4,671	1,003	0,071	100,000
>1000	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	>90	Total
masc	0,000	3,333	5,000	30,000	6,667	15,000	6,667	5,000	0,000	71,667
fem	0,000	3,333	5,000	5,000	5,000	8,333	1,667	0,000	0,000	28,333
Total	0,000	6,666	10,000	35,000	11,667	23,333	8,334	5,000	0,000	100,000

Tabela 8.28: Perfil crédito ISS PF: Sexo x Atividade

0	Alimentação	Varejo	Educação	Lazer	Serviços	Turismo	Saúde	Total
masc	0,402	13,936	4,281	2,054	15,678	0,242	10,129	46,722
fem	0,341	15,178	5,678	1,432	16,497	0,178	13,974	53,279
Total	0,743	29,114	9,959	3,486	32,175	0,420	24,103	100,000
0-20	Alimentação	Varejo	Educação	Lazer	Serviços	Turismo	Saúde	Total
masc	0,656	9,950	3,621	2,433	18,667	0,491	10,725	46,543
fem	0,518	10,701	4,593	2,162	18,723	0,452	16,308	53,457
Total	1,174	20,651	8,214	4,595	37,390	0,943	27,033	100,000
20-1000	Alimentação	Varejo	Educação	Lazer	Serviços	Turismo	Saúde	Total
masc	0,994	14,020	6,504	1,955	8,188	0,958	16,654	49,274
fem	1,000	10,657	6,921	2,456	6,480	0,662	22,551	50,727
Total	1,994	24,677	13,425	4,411	14,668	1,620	39,205	100,000
>1000	Alimentação	Varejo	Educação	Lazer	Serviços	Turismo	Saúde	Total
masc	1,667	10,000	5,000	0,000	33,333	0,000	21,667	71,667
fem	0,000	5,000	3,333	0,000	13,333	0,000	6,667	28,333
Total	1,667	15,000	8,333	0,000	46,666	0,000	28,334	100,000

Tabela 8.29: Perfil crédito ISS PF: RA x Sexo

0	Norte	Entorno	Outros	Oeste	Sul	Abastec	Sudeste	Sudoeste	Centro	Total
masc	3,972	0,398	0,501	12,605	1,318	0,216	1,325	5,420	20,899	46,654
fem	3,976	0,332	0,453	12,561	1,287	0,192	1,482	6,170	26,893	53,345
Total	7,948	0,730	0,954	25,166	2,605	0,408	2,807	11,590	47,792	100,000
0-20	Norte	Entorno	Outros	Oeste	Sul	Abastec	Sudeste	Sudoeste	Centro	Total
masc	3,498	0,396	0,517	11,871	1,492	0,200	1,475	6,197	20,915	46,560
fem	3,708	0,357	0,497	11,721	1,427	0,170	1,683	7,187	26,689	53,439
Total	7,206	0,753	1,014	23,592	2,919	0,370	3,158	13,384	47,604	100,000
20-1000	Norte	Entorno	Outros	Oeste	Sul	Abastec	Sudeste	Sudoeste	Centro	Total
masc	3,123	0,504	0,683	11,987	1,423	0,243	1,768	5,565	23,869	49,164
fem	3,128	0,351	0,504	11,247	1,422	0,162	1,983	5,785	26,253	50,835
Total	6,251	0,855	1,187	23,234	2,845	0,405	3,751	11,350	50,122	100,000
>1000	Norte	Entorno	Outros	Oeste	Sul	Abastec	Sudeste	Sudoeste	Centro	Total
masc	0,000	0,000	0,000	11,667	0,000	1,667	0,000	3,333	55,000	71,667
fem	0,000	0,000	0,000	3,333	0,000	0,000	1,667	3,333	20,000	28,333
Total	0,000	0,000	0,000	15,000	0,000	1,667	1,667	6,666	75,000	100,000

Tabela 8.30: Perfil crédito ISS PF: RA x Idade

0	Norte	Entorno	Outros	Oeste	Sul	Abastec	Sudeste	Sudoeste	Centro	Total
0-20	0,055	0,005	0,005	0,218	0,039	0,005	0,014	0,037	0,082	0,460
20-30	1,438	0,181	0,174	4,493	0,580	0,068	0,377	1,643	5,307	14,261
30-40	2,409	0,277	0,299	9,300	0,828	0,159	0,858	3,507	13,367	31,005
40-50	2,178	0,146	0,245	6,084	0,697	0,114	0,775	3,109	12,239	25,587
50-60	1,272	0,095	0,145	3,351	0,301	0,030	0,547	2,090	9,387	17,218
60-70	0,467	0,020	0,059	1,369	0,127	0,030	0,188	0,911	5,211	8,382
70-80	0,113	0,003	0,026	0,312	0,031	0,002	0,042	0,236	1,839	2,604
80-90	0,016	0,002	0,002	0,040	0,002	0,000	0,008	0,054	0,342	0,466
>90	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,002	0,016	0,018
Total	7,948	0,729	0,955	25,167	2,605	0,408	2,809	11,589	47,790	100,000

0-20	Norte	Entorno	Outros	Oeste	Sul	Abastec	Sudeste	Sudoeste	Centro	Total
0-20	0,048	0,008	0,006	0,148	0,025	0,003	0,017	0,055	0,104	0,414
20-30	1,249	0,188	0,191	4,276	0,641	0,068	0,400	2,163	5,284	14,460
30-40	2,140	0,277	0,343	8,588	0,907	0,146	0,959	4,026	13,400	30,786
40-50	2,051	0,157	0,238	5,975	0,791	0,096	0,947	3,505	12,572	26,334
50-60	1,131	0,096	0,150	3,047	0,364	0,039	0,575	2,370	9,287	17,059
60-70	0,449	0,022	0,060	1,235	0,145	0,017	0,205	0,935	4,786	7,854
70-80	0,122	0,004	0,022	0,291	0,043	0,002	0,049	0,295	1,766	2,594
80-90	0,015	0,001	0,004	0,032	0,004	0,000	0,006	0,034	0,382	0,478
>90	0,000	0,000	0,000	0,001	0,000	0,000	0,000	0,001	0,021	0,023
Total	7,205	0,753	1,014	23,593	2,920	0,371	3,158	13,384	47,602	100,000

20-1000	Norte	Entorno	Outros	Oeste	Sul	Abastec	Sudeste	Sudoeste	Centro	Total
0-20	0,014	0,005	0,001	0,059	0,004	0,001	0,006	0,035	0,063	0,188
20-30	0,793	0,128	0,173	3,569	0,400	0,048	0,292	1,351	4,330	11,084
30-40	1,679	0,248	0,350	7,283	0,783	0,145	1,221	2,771	12,508	26,993
40-50	1,809	0,199	0,251	5,773	0,820	0,128	1,118	2,790	12,613	25,501
50-60	1,133	0,203	0,220	3,989	0,487	0,048	0,687	2,326	10,048	19,141
60-70	0,593	0,061	0,128	1,947	0,263	0,031	0,318	1,434	6,575	11,350
70-80	0,214	0,009	0,058	0,545	0,078	0,003	0,090	0,553	3,123	4,673
80-90	0,016	0,003	0,006	0,064	0,011	0,001	0,020	0,090	0,792	1,003
>90	0,000	0,000	0,000	0,001	0,000	0,000	0,000	0,001	0,070	0,072
Total	6,251	0,856	1,187	23,230	2,846	0,405	3,752	11,351	50,122	100,000

>1000	Norte	Entorno	Outros	Oeste	Sul	Abastec	Sudeste	Sudoeste	Centro	Total
0-20	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
20-30	0,000	0,000	0,000	1,667	0,000	0,000	0,000	0,000	5,000	6,667
30-40	0,000	0,000	0,000	1,667	0,000	0,000	0,000	3,333	5,000	10,000
40-50	0,000	0,000	0,000	10,000	0,000	0,000	0,000	1,667	23,333	35,000
50-60	0,000	0,000	0,000	1,667	0,000	0,000	1,667	0,000	8,333	11,667
60-70	0,000	0,000	0,000	0,000	0,000	1,667	0,000	1,667	20,000	23,334
70-80	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	8,333	8,333
80-90	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	5,000	5,000
>90	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Total	0,000	0,000	0,000	15,000	0,000	1,667	1,667	6,667	74,999	100,000

Tabela 8.31: Perfil crédito ISS PF: RA x Atividade

0	Norte	Entorno	Outros	Oeste	Sul	Abast	Sudeste	Sudoeste	Centro	Total
Aliment	0,113	0,017	0,011	0,139	0,034	0,002	0,015	0,053	0,358	0,742
Varejo	1,906	0,239	0,279	6,348	0,635	0,143	0,890	3,172	15,501	29,113
Educação	1,416	0,082	0,107	3,364	0,373	0,024	0,234	1,188	3,171	9,959
Lazer	0,220	0,016	0,028	0,838	0,138	0,011	0,140	0,428	1,666	3,485
Serviços	1,342	0,157	0,282	6,178	0,419	0,151	0,938	3,732	18,979	32,177
Turismo	0,039	0,002	0,014	0,087	0,013	0,001	0,012	0,044	0,209	0,421
Saúde	2,954	0,222	0,229	8,227	1,014	0,065	0,573	2,960	7,858	24,102
Total	7,990	0,735	0,950	25,181	2,626	0,397	2,802	11,577	47,741	100,000

0-20	Norte	Entorno	Outros	Oeste	Sul	Abast	Sudeste	Sudoeste	Centro	Total
Aliment	0,068	0,009	0,027	0,212	0,028	0,006	0,027	0,416	0,383	1,176
Varejo	1,356	0,171	0,211	4,432	0,556	0,087	0,678	2,575	10,583	20,649
Educação	0,597	0,037	0,061	2,251	0,152	0,026	0,314	0,930	3,847	8,215
Lazer	0,296	0,035	0,044	0,910	0,173	0,022	0,104	0,845	2,168	4,597
Serviços	2,253	0,245	0,359	6,922	0,797	0,134	1,219	5,537	19,919	37,390
Turismo	0,070	0,012	0,024	0,279	0,036	0,003	0,024	0,128	0,368	0,944
Saúde	2,627	0,247	0,288	8,089	1,201	0,095	0,817	3,050	10,620	27,034
Total	7,267	0,756	1,014	23,095	2,943	0,373	3,183	13,481	47,893	100,000

20-1000	Norte	Entorno	Outros	Oeste	Sul	Abast	Sudeste	Sudoeste	Centro	Total
Aliment	0,111	0,102	0,106	0,326	0,041	0,004	0,063	0,155	1,086	1,994
Varejo	1,646	0,190	0,278	5,453	0,578	0,101	0,952	2,698	12,784	24,677
Educação	0,874	0,047	0,102	3,013	0,381	0,043	1,000	1,444	6,512	13,416
Lazer	0,089	0,005	0,036	0,451	0,031	0,020	0,112	0,310	3,356	4,410
Serviços	0,912	0,179	0,200	3,923	0,506	0,090	0,380	1,688	6,792	14,670
Turismo	0,108	0,028	0,065	0,438	0,048	0,004	0,058	0,176	0,696	1,621
Saúde	2,497	0,300	0,369	9,691	1,259	0,133	1,202	4,935	18,823	39,209
Total	6,237	0,851	1,156	23,295	2,844	0,395	3,767	11,406	50,046	100,000

>1000	Norte	Entorno	Outros	Oeste	Sul	Abast	Sudeste	Sudoeste	Centro	Total
Aliment	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1,667	0,000	1,667
Varejo	0,000	0,000	0,000	0,000	0,000	0,000	1,667	1,667	11,666	15,001
Educação	0,000	0,000	0,000	5,000	0,000	0,000	0,000	0,000	3,333	8,333
Lazer	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Serviços	0,000	0,000	0,000	8,333	0,000	0,000	0,000	1,667	36,666	46,667
Turismo	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Saúde	0,000	0,000	0,000	1,667	0,000	1,667	0,000	1,667	23,333	28,334
Total	0,000	0,000	0,000	15,000	0,000	1,667	1,667	6,668	75,000	100,000

Tabela 8.32: Perfil crédito ISS PF: Atividade x Idade

0	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	>90	Total
Aliment	0,010	0,228	0,278	0,124	0,070	0,025	0,007	0,001	0,000	0,743
Varejo	0,123	4,504	8,954	7,126	5,095	2,381	0,771	0,151	0,007	29,114
Educação	0,154	2,418	3,774	2,493	0,818	0,237	0,052	0,012	0,000	9,958
Lazer	0,026	0,884	1,262	0,728	0,400	0,140	0,040	0,006	0,001	3,487
Serviços	0,037	3,437	9,676	8,362	6,203	3,260	1,024	0,169	0,008	32,176
Turismo	0,002	0,091	0,124	0,086	0,067	0,035	0,014	0,003	0,000	0,422
Saúde	0,109	2,717	6,936	6,681	4,541	2,297	0,691	0,126	0,004	24,102
Total	0,461	14,279	31,006	25,600	17,194	8,375	2,599	0,468	0,020	100,000

0-20	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	>90	Total
Aliment	0,014	0,313	0,384	0,297	0,122	0,032	0,011	0,001	0,000	1,174
Varejo	0,054	2,825	6,037	5,145	3,905	1,939	0,631	0,109	0,005	20,650
Educação	0,069	1,051	2,435	3,273	1,090	0,231	0,057	0,008	0,000	8,214
Lazer	0,042	1,144	1,648	1,072	0,474	0,165	0,043	0,007	0,000	4,595
Serviços	0,092	5,923	12,445	9,140	6,225	2,624	0,806	0,129	0,005	37,389
Turismo	0,002	0,131	0,272	0,239	0,172	0,098	0,026	0,004	0,000	0,944
Saúde	0,143	3,072	7,540	7,205	5,082	2,757	1,003	0,218	0,014	27,034
Total	0,416	14,459	30,761	26,371	17,070	7,846	2,577	0,476	0,024	100,000

20-1000	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	>90	Total
Aliment	0,000	0,209	0,853	0,564	0,264	0,073	0,027	0,005	0,000	1,995
Varejo	0,055	3,209	6,457	6,393	5,032	2,493	0,873	0,157	0,008	24,677
Educação	0,062	1,815	5,446	4,503	1,247	0,290	0,057	0,005	0,000	13,425
Lazer	0,011	0,648	1,708	1,171	0,591	0,216	0,058	0,008	0,000	4,411
Serviços	0,021	2,515	4,240	3,325	2,560	1,349	0,546	0,105	0,009	14,670
Turismo	0,000	0,097	0,342	0,477	0,391	0,214	0,081	0,018	0,000	1,620
Saúde	0,035	2,522	7,939	9,110	9,041	6,751	3,041	0,708	0,055	39,205
Total	0,184	11,015	26,985	25,546	19,126	11,386	4,683	1,006	0,072	100,000

>1000	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	>90	Total
Aliment	0,000	0,000	1,667	0,000	0,000	0,000	0,000	0,000	0,000	1,667
Varejo	0,000	1,667	1,667	6,667	3,333	1,667	0,000	0,000	0,000	15,001
Educação	0,000	1,667	1,667	1,667	3,333	0,000	0,000	0,000	0,000	8,334
Lazer	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Serviços	0,000	1,667	3,333	16,666	3,333	8,333	8,333	5,000	0,000	46,666
Turismo	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Saúde	0,000	1,667	1,667	10,000	1,667	13,333	0,000	0,000	0,000	28,334
Total	0,000	6,668	10,000	35,000	11,666	23,333	8,333	5,000	0,000	100,000

A Tabela 8.26 apresenta a análise da probabilidade de ocorrência das variáveis sexo e idade para o ICMS de pessoas físicas.

A Tabela 8.21 apresenta a análise da probabilidade de ocorrência das variáveis sexo e atividade para o ICMS de pessoas físicas.

A Tabela 8.22 apresenta a análise da probabilidade de ocorrência das variáveis RA e sexo para o ICMS de pessoas físicas.

A Tabela 8.23 apresenta a análise da probabilidade de ocorrência das variáveis RA e idade para o ICMS de pessoas físicas.

A Tabela 8.24 apresenta a análise da probabilidade de ocorrência das variáveis RA e atividade para o ICMS de pessoas físicas.

A Tabela 8.25 apresenta a análise da probabilidade de ocorrência das variáveis atividade e idade para o ICMS de pessoas físicas.

A Tabela 8.27 apresenta a análise da probabilidade de ocorrência das variáveis sexo e idade para o ISS de pessoas físicas.

A Tabela 8.28 apresenta a análise da probabilidade de ocorrência das variáveis sexo e atividade para o ISS de pessoas físicas.

A Tabela 8.29 apresenta a análise da probabilidade de ocorrência das variáveis RA e sexo para o ISS de pessoas físicas.

A Tabela 8.30 apresenta a análise da probabilidade de ocorrência das variáveis RA e idade para o ISS de pessoas físicas.

A Tabela 8.31 apresenta a análise da probabilidade de ocorrência das variáveis RA e atividade para o ISS de pessoas físicas.

A Tabela 8.32 apresenta a análise da probabilidade de ocorrência das variáveis atividade e idade para o ISS de pessoas físicas.

Os gráficos referentes às tabelas analisadas nessa Seção estão apresentados nos Anexos A e B.

Com base nas observações sobre as tabelas e gráficos foi possível extrair as informações que se seguem sobre as faixas de crédito.

De maneira geral as tendências observadas para as faixas de crédito não se aplicam na faixa de crédito >1000. Há inversões bruscas de valores e tendências para todos as variáveis analisadas.

Para a variável sexo, no ICMS, embora o sexo masculino seja predominante na faixa 0, há inversão na tendência inicial dos créditos do sexo masculino para o sexo feminino na faixa de 20-1000 e nas seguintes.

Para o ISS, o sexo feminino é predominante até a faixa de 20-1000. Na faixa >1000 há inversão da tendência e o sexo masculino é majoritário com mais de 70% dos casos. Ainda nesta faixa de crédito, os valores encontrados previamente de cerca de 12% dos

documentos fiscais para ambos os sexos na idade 40-50, variam para valores de 30% para o sexo masculino e 5% para o sexo feminino na idade 40-50. Considerando o Centro, temos que cerca de 55% dos documentos fiscais do sexo masculino se encontram nesta faixa.

Na variável idade, foi verificado no ISS e no ICMS, que há uma leve tendência que quanto maior a idade mais créditos sejam obtidos. Presume-se que pessoas mais velhas tenham despesas maiores e maior renda explicando esta tendência, pois, em geral, quanto maior o gasto, mais créditos são concedidos. No ISS para >1000 verificou-se peso relevante para a idade de 60 a 70 anos, diferente das outras faixas.

Para a variável atividade, no ICMS, da mesma forma que no perfil de fidelidade são predominantes Alimentação, Varejo e Serviços Gerais. Quanto maior o crédito, menor as porcentagens da Alimentação e maiores as porcentagens de Varejo e de Serviços Gerais. É relevante notar que para a faixa >1000 a atividade Varejo no grupo Centro abrange mais de 40% dos documentos fiscais desta faixa de crédito.

Para o ISS são predominantes Varejo, Serviços Gerais, Saúde e Educação. Para o valor inicial dos créditos há uma distribuição de porcentagens similar entre estes serviços, no entanto, à medida que o valor dos créditos aumenta, diminuem as porcentagens dos Serviços Gerais e aumentam principalmente as porcentagens da Saúde até a faixa de crédito 20-1000. Presume-se que os consumidores do PNL solicitam documentos fiscais com valores baixos para um conjunto de atividades econômicas e documentos fiscais de valores mais elevados para outro conjunto de atividades econômicas. Embora Turismo, Educação e Lazer não representem nem 1% dos valores para o ICMS, para o ISS estas atividades possuem porcentagens mais significantes, principalmente no caso de Educação. Para a faixa de crédito >1000, a atividade Serviços Gerais para a população masculina e na região Centro abrange um terço dos documentos fiscais desta faixa, enquanto este comportamento não se apresentou pelas outras faixas de crédito.

Considera-se que serviços para Turismo, Educação e Lazer são mais comuns que aquisições para estas atividades econômicas, o que explicaria os valores encontrados. De forma geral, para a gestão do PNL, seria interessante revisar os agrupamentos de CNAE para evitar o desbalanceamento entre as atividades econômicas e entre o ISS e o ICMS.

Na variável RA, observou-se, para o ISS e o ICMS, pesos estáveis ao longo das faixas de crédito. Há destaque na origem dos créditos respectivamente para as regiões Centro, Oeste, Sudoeste e Norte. O Centro responde por mais de 40% das emissões de crédito. A expectativa inicial era que houvesse uma distribuição de créditos similar a que ocorre com a população destas regiões e esta expectativa não se confirmou. As porcentagens encontradas para o Entorno e outros Estados não chegou a 2% para as faixas de crédito tanto para o ISS quanto para o ICMS.

8.4.4 Análise das variáveis selecionadas para pessoas jurídicas

Tabela 8.33: ICMS RA x Atividade

0	Norte	Entorno	Outros	Oeste	Sul	Abastec	Sudeste	Sudoeste	Centro	Total
Aliment	1,162	0,000	0,000	16,705	0,499	0,551	0,122	1,395	4,394	24,827
Varejo	3,306	0,000	0,018	17,707	2,500	5,352	0,873	3,979	15,328	49,063
Educação	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,002	0,003
Lazer	0,000	0,000	0,000	0,002	0,000	0,003	0,000	0,000	0,013	0,017
Serviços	2,427	0,000	0,000	2,064	0,183	1,266	0,409	0,713	18,890	25,951
Turismo	0,000	0,000	0,000	0,001	0,000	0,005	0,000	0,000	0,020	0,026
Saúde	0,024	0,000	0,000	0,020	0,001	0,007	0,001	0,007	0,053	0,112
Total	6,920	0,000	0,018	36,497	3,183	7,183	1,405	6,094	38,700	100,000

0-20	Norte	Entorno	Outros	Oeste	Sul	Abastec	Sudeste	Sudoeste	Centro	Total
Aliment	0,502	0,000	0,001	15,522	0,179	1,055	0,289	1,448	9,574	28,571
Varejo	2,643	0,000	0,009	16,738	1,718	5,115	0,960	3,998	17,754	48,936
Educação	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,003	0,003
Lazer	0,002	0,000	0,000	0,005	0,000	0,001	0,000	0,001	0,008	0,017
Serviços	1,887	0,000	0,001	1,679	0,107	1,274	0,299	0,593	16,326	22,166
Turismo	0,000	0,000	0,000	0,002	0,001	0,009	0,000	0,000	0,035	0,047
Saúde	0,013	0,000	0,000	0,029	0,007	0,028	0,005	0,015	0,163	0,259
Total	5,047	0,000	0,010	33,975	2,012	7,483	1,555	6,055	43,863	100,000

20-1000	Norte	Entorno	Outros	Oeste	Sul	Abastec	Sudeste	Sudoeste	Centro	Total
Aliment	0,728	0,000	0,000	6,240	0,130	0,739	0,163	0,802	7,725	16,528
Varejo	5,516	0,000	0,001	27,522	3,508	7,234	1,999	6,347	20,579	72,706
Educação	0,000	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,001	0,003
Lazer	0,001	0,000	0,000	0,006	0,000	0,001	0,000	0,000	0,022	0,030
Serviços	1,153	0,000	0,000	2,055	0,164	0,648	0,210	0,598	5,469	10,299
Turismo	0,000	0,000	0,000	0,010	0,000	0,017	0,000	0,000	0,174	0,200
Saúde	0,015	0,000	0,000	0,065	0,011	0,017	0,008	0,014	0,104	0,234
Total	7,414	0,000	0,001	35,899	3,813	8,657	2,380	7,761	34,075	100,000

>1000	Norte	Entorno	Outros	Oeste	Sul	Abastec	Sudeste	Sudoeste	Centro	Total
Aliment	0,000	0,000	0,000	0,778	0,000	0,778	0,000	0,389	7,393	9,339
Varejo	15,953	0,000	0,000	31,518	4,280	10,895	1,556	5,447	14,397	84,047
Educação	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,389	0,389
Lazer	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Serviços	0,778	0,000	0,000	1,167	0,000	0,389	0,000	0,778	1,556	4,669
Turismo	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Saúde	0,000	0,000	0,000	0,389	0,000	0,000	0,000	0,000	1,167	1,556
Total	16,732	0,000	0,000	33,852	4,280	12,062	1,556	6,615	24,903	100,000

Foram calculadas tabelas com a probabilidade de ocorrência dos valores da base de dados de crédito de pessoas jurídicas (icms_pj.txt e iss_pj.txt) com tabulações para as variáveis atividade e RA ao longo das quatro faixas de crédito.

A Tabela 8.33 apresenta a análise da probabilidade de ocorrência das variáveis atividade e RA para o ICMS.

A Tabela 8.34 apresenta a análise da probabilidade de ocorrência das variáveis atividade e RA para o ISS.

Os gráficos referentes às tabelas analisadas nessa seção estão apresentados nos Anexos D e E. Os gráficos plotam as tabulações ao longo das quatro faixas de crédito.

Tabela 8.34: ISS RA x Atividade

0	Norte	Entorno	Outros	Oeste	Sul	Abastec	Sudeste	Sudoeste	Centro	Total
Aliment	0,009	0,000	0,000	0,065	0,005	0,080	0,001	0,023	0,214	0,397
Varejo	1,373	0,000	0,018	5,833	0,814	2,729	0,333	1,851	10,211	23,162
Educação	0,008	0,000	0,000	0,080	0,001	0,060	0,001	0,028	0,211	0,389
Lazer	0,000	0,000	0,000	0,001	0,000	0,003	0,000	0,001	0,011	0,016
Serviços	1,948	0,000	0,001	10,882	0,837	3,520	0,410	3,396	21,157	42,149
Turismo	0,006	0,000	0,000	0,088	0,005	0,022	0,000	0,017	2,189	2,327
Saúde	2,038	0,000	0,017	9,582	1,138	2,709	0,606	2,753	12,715	31,558
Total	5,382	0,000	0,036	26,529	2,800	9,123	1,351	8,069	46,708	100,000

0-20	Norte	Entorno	Outros	Oeste	Sul	Abastec	Sudeste	Sudoeste	Centro	Total
Aliment	0,004	0,000	0,000	0,183	0,002	0,185	0,007	0,043	3,390	3,814
Varejo	1,715	0,000	0,004	8,194	0,801	3,689	0,397	2,302	11,054	28,156
Educação	0,013	0,000	0,000	0,076	0,004	0,034	0,001	0,018	0,249	0,395
Lazer	0,001	0,000	0,000	0,018	0,000	0,004	0,000	0,003	0,025	0,051
Serviços	3,872	0,000	0,018	17,185	2,531	3,878	1,354	4,404	23,198	56,440
Turismo	0,021	0,000	0,000	0,243	0,006	0,344	0,002	0,073	5,790	6,479
Saúde	0,247	0,000	0,000	1,018	0,128	0,435	0,032	0,496	2,309	4,665
Total	5,873	0,000	0,022	26,917	3,472	8,569	1,793	7,339	46,015	100,000

20-1000	Norte	Entorno	Outros	Oeste	Sul	Abastec	Sudeste	Sudoeste	Centro	Total
Aliment	0,051	0,000	0,000	0,727	0,034	0,938	0,006	0,200	9,085	11,041
Varejo	0,984	0,000	0,000	5,778	0,687	3,341	0,063	1,270	7,786	19,909
Educação	0,034	0,000	0,000	0,332	0,006	0,051	0,023	0,149	0,961	1,556
Lazer	0,000	0,000	0,000	0,029	0,000	0,011	0,000	0,000	0,029	0,069
Serviços	3,049	0,000	0,006	15,010	1,751	4,531	0,389	3,278	24,046	52,057
Turismo	0,063	0,000	0,000	0,366	0,017	0,646	0,000	0,057	11,219	12,368
Saúde	0,246	0,000	0,000	0,509	0,114	0,429	0,006	0,057	1,636	2,997
Total	4,427	0,000	0,006	22,748	2,609	9,947	0,487	5,011	54,762	100,000

>1000	Norte	Entorno	Outros	Oeste	Sul	Abastec	Sudeste	Sudoeste	Centro	Total
Aliment	0,000	0,000	0,000	3,922	0,000	5,882	0,000	0,000	47,059	56,863
Varejo	0,000	0,000	0,000	0,000	0,000	1,961	0,000	0,000	5,882	7,843
Educação	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Lazer	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Serviços	0,000	0,000	0,000	0,000	0,000	1,961	0,000	0,000	19,608	21,569
Turismo	0,000	0,000	0,000	0,000	0,000	5,882	0,000	0,000	7,843	13,725
Saúde	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Total	0,000	0,000	0,000	3,922	0,000	15,686	0,000	0,000	80,392	100,000

De maneira geral, talvez por dispor de poucas observações de ICMS e principalmente de ISS, as tendências observadas tanto para RA quanto para atividade variam bruscamente entre as faixas de crédito.

Com base nas observações sobre as tabelas e gráficos foi possível extrair as informações que se seguem sobre as faixas de crédito.

Foram encontrados poucos documentos fiscais de outros Estados e não foi encontrado nenhum documento fiscal de empresas do Entorno. Estes valores eram esperados uma vez que pelas regras do PNL, o uso dos créditos só pode ser realizado para imóveis e veículos de propriedade da empresa e registrados no Distrito Federal.

Na variável RA são predominantes para o ICMS, respectivamente, Centro, Oeste, Norte e Abastecimento. Para as faixas de crédito iniciais a RA Abastecimento tem peso maior que a RA Norte. Nas maiores faixas de crédito a situação se inverte.

Para o ISS também são predominantes respectivamente, Centro, Oeste, Norte e Abastecimento. O peso do Centro aumenta bruscamente na faixa de crédito >1000 em detrimento das outras RA.

Por se tratar da área industrial do DF, faz sentido que na análise para pessoa jurídica haja participação da RA de Abastecimento que foi registrada com pouco peso nas análises para pessoas físicas.

Na variável atividade, para o ICMS, há predominância do Varejo e um crescimento expressivo dele ao longo das faixas de crédito, seguido de Alimentação e Serviços Gerais. Para a faixa de crédito >1000, o Varejo chega a obter 84% dos documentos fiscais emitidos.

Para o ISS, para os valores iniciais de crédito, há predominância de Varejo, Serviços Gerais e Saúde, respectivamente. No entanto, para a faixa de crédito >1000 os pesos variam muito com relação às outras faixas, sendo que são predominantes Alimentação, Serviços Gerais e Turismo. Quanto a Turismo, em nenhuma outra análise houve peso relevante desta atividade.

Capítulo 9

Conclusões e Trabalhos Futuros

Esse capítulo descreve as conclusões, resultados obtidos e as linhas de trabalho futuro.

9.1 Conclusões

Cabe ressaltar que, ao invés de uma amostra, esta pesquisa utilizou todas as informações dos bancos de dados disponíveis. A manipulação de grandes quantidades de dados se apresentou difícil e demorada para processamento. O uso de dados reais utilizando quatro banco de dados diferentes apresentou um desafio para corrigir e integrar as informações. Ao seguir o CRISP-DM, atingir a completude da preparação dos dados consumiu mais da metade do tempo para realizar esta pesquisa.

Contudo, o objetivo desta pesquisa foi alcançado ao explicitar informações sobre os perfis dos beneficiários. Essa informação pode auxiliar a gestão do Programa Nota Legal da SEF.

Quanto ao perfil de créditos de consumo as seguintes conclusões podem ser extraídas sobre as faixas de crédito.

Faixa 0: Ao se somar os documentos fiscais com crédito 0, inclusive os excluídos na Seção 8.4.1, chega-se a 20% do total dos documentos fiscais. Há predominância do sexo masculino.

Faixa 0-20: Houve surpresa ao constatar que em torno de 65% dos documentos fiscais geraram créditos abaixo de R$ 1. Para os beneficiários pode se tornar frustante a solicitação dos documentos fiscais junto às empresas e, em momento posterior, receber uma quantidade pequena de crédito. Ao considerar toda a faixa 0-20, 78,5% dos documentos fiscais se encontram nesta faixa de crédito. Há predominância do sexo feminino.

Faixa 20-1000: Embora representem apenas 1,5% dos documentos fiscais emitidos, as atividades econômicas de maior relevâncias são distintas das encontradas nas faixas 0 e 0-20. Há predominância do sexo feminino.

Faixa >1000: Menos de 0,001% dos documentos fiscais se encontram nesta faixa. Há tendências diferentes em relação às outras faixas nas atividades econômicas, idade e em sexo.

Conforme apresentado na Seção 8.4.1, embora o PNL tenha tido grande aceitação pelas pessoas físicas, ele não se popularizou ou não foram encontrados incentivos para seu uso pelas pessoas jurídicas.

De forma geral, as maiores porcentagens para a variável idade se mantiveram estáveis de 30 a 50 anos ao longo das faixas de crédito. Nas avaliações sobre as Regiões Administrativas, havia expectativa de maior participação do Entorno já que o consumo de grande parte da população destas áreas ocorre no Distrito Federal. Diferente de outros programa de concessão de benefícios, há pouca participação de outros Estados. As cidades satélites principalmente do grupo Oeste, representam participação significativa para o PNL. No entanto, o grupo Centro é responsável pela maioria de emissão de documentos fiscais encontrados hoje. A atividade econômica mais relevante para a emissão de documentos fiscais é o grupo de Alimentação.

Quanto ao perfil de fidelidade as seguintes conclusões podem ser extraídas sobre as faixas de fidelidade.

Faixa 1-11: A minoria dos documentos fiscais se encontram nesta faixa. As idades de maior peso para esta variável variam de 20 a 40 anos. Há predominância do sexo masculino. A RA mais relevante é a Oeste.

Faixa 12-23: As idades de maior peso para esta variável variam de 20 a 40 anos. Há predominância do sexo masculino. A RA mais relevante é a Oeste.

Faixa 24-35: esta faixa concentra a maioria dos consumidores. As idades de maior peso para esta variável variam de 30 a 50 anos. Há predominância do sexo masculino. A RA mais relevante é a Oeste.

Faixa 36-47: a maioria dos documentos fiscais emitidos se encontra nesta faixa chegando a quase 50% do total. As idades de maior peso para esta variável variam de 30 a 50 anos. Há inversão na tendência da variável sexo com predominância do sexo feminino. Também há mudança da tendência de RA, onde a RA mais relevante é o Centro.

Faixa 48-58: faixa que abrange ainda uma pequena proporção dos consumidores. O grupo Centro tem maior participação nesta faixa. Presume-se que por alcançar maior quantidade de créditos concedidos, por maior poder aquisitivo, haja tendência dos consumidores em continuar no programa. As idades de maior peso para esta variável variam de 30 a 50 anos. Há leve tendência de quanto maior a idade, maior seja a fidelidade dos consumidores. Há predominância do sexo feminino. Presume-se que embora homens sejam mais entusiastas com o PNL, as mulheres que tendem a ser mais fiéis ao longo do tempo.

De forma geral, pela distribuição da quantidade de pessoas físicas pela faixas de fidelidade, uma vez cadastradas, há participação regular no programa. Alimentação e Varejo respondem pela grande maioria das atividades econômicas, no entanto, ao longo das faixas de fidelidade, o peso da Alimentação cresce enquanto do Varejo diminui. Pode-se constatar que o peso das cidades satélites cai e a região central de Brasília aumenta significativamente ao longo do tempo de fidelidade. Apesar de pessoas de outros Estados estarem cadastradas no PNL, sua fidelidade não se mantem ao longo do tempo. O mesmo se observa para a população do Entorno em que seu peso chega a ser menor que da pessoas de outros Estados. Como a população destes locais tem seu cotidiano no Distrito Federal esperava-se maior fidelidade destes locais.

No perfil de fidelidade foram apresentados dois indicadores visando melhoria do PNL. Foi constatado que a intensidade e a perseverança dos consumidores vem crescendo ao longo das faixas de fidelidade. Mesmo com a quantidade de pessoas se estabilizando, pode-se trabalhar na melhoria da gestão do PNL através destes indicadores. Conforme apresentado na Subseção 8.3.1, ao calcular os indicadores pela vigência do PNL, temos que o indicador perseverança calculado foi de 63,6% e o indicador intensidade calculado foi de 3,9 documentos fiscais / mês. Cabe à SEF definir políticas para que o indicador perseverança possa se aproximar de 100%. Da mesma forma, o indicador intensidade pode ser melhorado com incentivos da SEF, pois há beneficiários que chegam a emitir cerca de 30 documentos fiscais/mês em média.

Por fim, a Tabela 9.1 apresenta a evolução dos documentos fiscais e quantidade de participantes efetivos do PNL nos anos apurados. Conforme [45], o número de pessoas ocupadas da população economicamente ativa do Distrito Federal é de 1.269.000. Desta forma, o número de pessoas no PNL já equivale a 61% da população economicamente ativa do Distrito Federal. Embora ainda exista espaço para crescimento, a gestão do PNL deve garantir novos estímulos para conseguir novas adesões e evitar desistências do programa.

Tabela 9.1: Crescimento PNL

Ano	**Qtd Doc Fiscais**	**Qtd pessoas**
2009	1543116	357269
2010	8238436	591191
2011	29116856	703053
2012	57327881	759027
2013	60962578	773963

Pode-se afirmar que o PNL teve sucesso em sua adesão pela população. Mesmo com crescimento menor apurado nos últimos anos, o PNL ganha cada vez mais adeptos e mais documentos fiscais válidos com CPF/CNPJ são registrados pela SEF. A gestão do

PNL agora pode se valer desta pesquisa para avaliar políticas para os próximos anos do programa.

9.2 Resultados obtidos

Como a administração fazendária tem uma série de atribuições e necessidades, diversas ações foram postas em prática visando dotá-la de condições e instrumentos capazes de propiciar uma política fiscal transparente, priorizando o aprimoramento da gestão:

- Ao evidenciar problemas de qualidade nos dados e, em especial para endereço postal, há iniciativa para uso sistematizado do banco de dados dos Correios como padronização de endereços no PNL. Outros sistemas da SEF poderiam se beneficiar deste banco de dados, inclusive para comunicação aos contribuintes.

- Pela detecção de *outliers*, as pessoas físicas identificadas foram encaminhadas à SEF para verificação se houve uso dos créditos durante os períodos de indicação. Para os casos de óbitos e idades discrepantes houve alteração no sistema de concessão de créditos para geração de alertas nos casos encontrados e bloqueio preventivo para utilização.

- Como subsídio a este trabalho foram montados painéis de informações com o software Qlikview[1]. Este software tem sido utilizado na SEF como solução para problemas de *Business Discovery* [2]. Em consonância com os conceitos apresentados em [32], uma vez que a pesquisa produzida possui semelhanças com a modelagem dimensional - ao estabelecer o documento fiscal como fato e as características de pessoas, tempo e local como dimensões - as cargas de dados realizadas e os gráficos produzidos foram disponibilizadas à gestão do PNL. Além de contribuir para melhoria da gestão, há o objetivo de disponibilização das informações pelo site do PNL como transparência ao cidadão.

Para a sociedade, os perfis pesquisados apresentam fonte de informação para os benefícios apurados e como esclarecimentos gerais sobre os beneficiários existentes.

9.3 Trabalhos futuros

Este trabalho abre caminho para uma série de análises usando outras tarefas e técnicas de mineração de dados. Conforme mencionado em 5.2, em uma segunda fase, planeja-se

[1] www.qlik.com

[2] Conforme [44], método que visa ajudar as empresas a tomar decisões inteligentes mediante informações recolhidas por diversas fontes de informação

utilizar a base de dados do perfil de fidelidade na tarefa de associação e as bases de dados do perfil de crédito para a tarefa de *clustering*. Além disso, utilizando o banco de dados do PNL há ainda várias oportunidades de pesquisa: avaliação de como se dá o uso dos créditos pelos beneficiários, análise das reclamações feitas pelos beneficiários e concluídas pela SEF, levantamento de indicadores sobre os e-mails enviados, avaliação dos autos de infração à empresa, entre outros.

É viável o uso desta pesquisa como auxílio à auditoria e fiscalização tributária ao identificar padrões atípicos nos perfis.

Há possibilidade de uso desta pesquisa para cálculo da Curva de Laffer conforme trabalhos apresentados em [38] e [42]. A inclinação desta curva pode fornecer informações sobre como se comporta a arrecadação em função das alíquotas dos impostos. Se a inclinação for positiva, indica que o aumento da alíquota induz ao aumento da arrecadação. Se a inclinação for zero indica que a arrecadação não respondeu às variações de alíquotas. Se a inclinação for negativa, é sinal de que o aumento nas alíquotas levou a redução da arrecadação. O cálculo da Curva de Laffer permitiria ter indicadores sobre a carga tributária praticada no Distrito Federal.

Em testes preliminares usando as bases de dados obtidas, houve dificuldades na execução do WEKA e impossibilidade de uso em máquinas *desktop* tradicionais pela limitação de memória RAM. Os arquivos ARFF prontos para processamento chegaram a ocupar 50GB de RAM com um tempo de carga de cerca de 30 min. A paralelização na execução dos algoritmos de mineração de dados em *clusters* de máquinas poderia viabilizar novas aplicações, conforme trabalhos apresentados em [28] e [22].

Vários estados já adotam programas similares ao PNL. No momento desta pesquisa já existem iniciativas em: Alagoas, Rondônia, São Paulo, Minas Gerais, Paraná, Rio Grande do Sul, Rio de Janeiro, Pernambuco e Pará. Seria de grande valia a comparação de perfis e indicadores dos outros programas afim de avaliar padrões de comportamento visando melhores comparações e conclusões mais robustas. Desta forma, cabe ressaltar o impacto desta pesquisa envolvendo DF e entorno que pode ser expandida para todo o país ao permitir a comparação entre diferentes programas de concessão de benefícios ao consumidor.

Referências

[1] *Introduction to the Theory of Statistics*. McGraw-Hill, third edition, 1974. 7

[2] Cláudio Vasconcelos Braga. Rede neural e regressão linear: comparativo entre as técnicas aplicadas a um caso prático na Receita Federal. Master's thesis, Faculdade de Economia e Finanças IBMEC, 2010. 3

[3] Brasil. Lei nº 5.172, de 25 de outubro de 1966, denominado Código Tributário Nacional. Dispõe sobre o Sistema Tributário Nacional e institui normas gerais de direito tributário aplicáveis à União, Estados e Municípios. Câmara dos Deputados, 1966. Disponível em http://www.planalto.gov.br/ccivil_03/leis/l5172.htm. Acesso em 06/07/2014. 20

[4] Brasil. Constituição da República Federativa do Brasil, de 05 de outubro de 1988. Câmara dos Deputados, 1988. Disponível em http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em 06/07/2014. 19

[5] Brasil. Lei Complementar nº 94, de 19 de fevereiro de 1998. Autoriza o Poder Executivo a criar a Região Integrada de Desenvolvimento do Distrito Federal e Entorno - RIDE e instituir o Programa Especial de Desenvolvimento do Entorno do Distrito Federal, e dá outras providências. Câmara dos Deputados, 1998. Disponível em http://www.planalto.gov.br/ccivil_03/leis/lcp/Lcp94.htm. Acesso em 06/07/2014. 39

[6] Brasil. Lei Complementar nº 116, de 31 de julho de 2003. Dispõe sobre o Imposto Sobre Serviços de Qualquer Natureza, de competência dos Municípios e do Distrito Federal, e dá outras providências. Câmara dos Deputados, 2003. Disponível em http://www.planalto.gov.br/ccivil_03/leis/lcp/lcp116.htm. Acesso em 06/07/2014. 14

[7] Brasil. Ato COTEPE nº 35, de 05 de julho de 2005. Dispõe sobre as especificações técnicas para a geração, o armazenamento e o envio de arquivos em meio digital relativos aos registros de documentos fiscais, livros fiscais, lançamentos contábeis, demonstrações contábeis, documentos de informação econômico-fiscais e outras informações de interesse do fisco. Câmara dos Deputados, 2005. Disponível em http://www1.fazenda.gov.br/confaz/confaz/atos/atos_cotepe/2007/..%5C2005%5CAC035_05.htm. Acesso em 06/07/2014. 19

[8] Brasil. Ato COTEPE nº 70, de 02 de dezembro de 2005. Dá nova redação ao Manual de Orientação do Leiaute Fiscal de Processamento

de Dados, instituído pelo Anexo Único do Ato Cotepe nº 35/05, de 5 de julho de 2005. Câmara dos Deputados, 2005. Disponível em http://www1.fazenda.gov.br/confaz/confaz/atos/atos_cotepe/2007/..%5C2005%5CAC070_05.htm. Acesso em 13/07/2014. 28

[9] Brasil. Decreto nº 25.508, de 19 de janeiro de 2005. Regulamenta o Imposto Sobre Serviços de Qualquer Natureza - ISS. Secretaria de Estado de Fazenda do Distrito Federal, 2005. Disponível em http://www.fazenda.df.gov.br/aplicacoes/legislacao/legislacao/TelaSaidaDocumento.cfm?txtNumero=25508&txtAno=2005&txtTipo=6&txtParte=A)%20TEXTO%20ORIGINAL. Acesso em 09/07/2014. 29

[10] Brasil. Decreto nº 26.529, de 13 de janeiro de 2006. Institui o Livro Fiscal Eletrônico que substitui os livros fiscais relacionados no Decreto nº 18.955, de 22 de dezembro de 1997, e no Decreto nº 25.508, de 19 de janeiro de 2005. Secretaria de Estado de Fazenda do Distrito Federal, 2006. Disponível em http://www.fazenda.df.gov.br/aplicacoes/legislacao/legislacao/TelaSaidaDocumento.cfm?txtNumero=26529&txtAno=2006&txtTipo=6&txtParte=. Acesso em 06/07/2014. 19

[11] Brasil. Portaria nº 210, de 14 de julho de 2006. Estabelece normas para fins de aplicação do Decreto nº 26.529, de 13 de janeiro de 2006, que instituiu o Livro Fiscal Eletrônico que substitui os livros fiscais relacionados no Decreto nº 18.955, de 22 de dezembro de 1997, e no Decreto nº 25.508, de 19 de janeiro de 2005. Câmara dos Deputados, 2006. Disponível em http://www.fazenda.df.gov.br/aplicacoes/legislacao/legislacao/TelaSaidaDocumento.cfm?txtNumero=210&txtAno=2006&txtTipo=7&txtParte=. Acesso em 06/07/2014. 19

[12] Brasil. Decreto nº 28.445, de 20 de novembro de 2007. Consolida a legislação que institui e regulamenta o Imposto sobre a Propriedade Predial e Territorial Urbana - IPTU. Secretaria de Estado de Fazenda do Distrito Federal, 2007. Disponível em http://www.fazenda.df.gov.br/aplicacoes/legislacao/legislacao/TelaSaidaDocumento.cfm?txtNumero=28445&txtAno=2007&txtTipo=6&txtParte=. Acesso em 06/07/2014. 14

[13] Brasil. Decreto nº 29.396, de 13 de agosto de 2008. Regulamenta a Lei nº 4.159, de 13 de junho de 2008, que dispõe sobre a criação do programa de concessão de créditos para adquirentes de mercadorias ou bens e tomadores de serviços, nos termos que especifica, e dá outras providências. Secretaria de Estado de Fazenda do Distrito Federal, 2008. Disponível em http://www.fazenda.df.gov.br/aplicacoes/legislacao/legislacao/TelaSaidaDocumento.cfm?txtNumero=29396&txtAno=2008&txtTipo=6&txtParte=. Acesso em 06/07/2014. 1

[14] Brasil. Lei nº 4.159, de 13 de junho de 2008. Dispõe sobre a criação do programa de concessão de créditos para adquirentes de mercadorias ou bens e tomadores de serviços, nos termos que especifica. Câmara Legislativa do Distrito Federal, 2008. Disponível em

http://www.fazenda.df.gov.br//aplicacoes/legislacao/legislacao/TelaSaidaDocumento.cfm?txtNumero=4159&txtAno=2008&txtTipo=5&txtParte=COMPILADO. Acesso em 06/07/2014. 1, 68, 74

[15] Brasil. Portaria nº 323, de 13 de agosto de 2008. Estabelece cronograma de implantação do programa de que trata a lei nº 4.159, de 13 de junho de 2008, e dá outras providências. Câmara Legislativa do Distrito Federal, 2008. Disponível em http://www.fazenda.df.gov.br//aplicacoes/legislacao/legislacao/TelaSaidaDocumento.cfm?txtNumero=323&txtAno=2008&txtTipo=7&txtParte=TEXTO%20COMPILADO. Acesso em 06/07/2014. 17, 38

[16] Brasil. Anexo Único da Lei nº 4.722, de 27 de dezembro de 2011. Estabelece a pauta de valores venais dos veículos automotores registrados e licenciados no Distrito Federal para efeito de lançamento do Imposto sobre a Propriedade de Veículos Automotores – IPVA para o exercício de 2012 e dá outras providências. Câmara dos Deputados, 2011. Publicada no Diário Oficial do Distrito Federal (DODF) nº 248, de 28 de dezembro de 2011. 14

[17] Brasil. Decreto nº 7.469, de 04 de maio de 2011. Regulamenta a Lei Complementar no 94, de 19 de fevereiro de 1998, que autoriza o Poder Executivo a criar a Região Integrada de Desenvolvimento do Distrito Federal e Entorno - RIDE e instituir o Programa Especial de Desenvolvimento do Entorno do Distrito Federal. Câmara dos Deputados, 2011. Disponível em http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2011/Decreto/D7469.htm. Acesso em 06/07/2014. 39

[18] Brasil. Portaria nº 102, de 10 de julho de 2012. Altera a Portaria nº 4, de 4 de janeiro de 2012, que estabelece procedimentos relativos à concessão, à consolidação e à utilização de créditos no âmbito do programa instituído pela Lei nº. 4.159, de 13 de junho de 2008; e dá outras providências. Câmara Legislativa do Distrito Federal, 2012. Disponível em http://www.fazenda.df.gov.br/aplicacoes/legislacao/legislacao/TelaSaidaDocumento.cfm?txtNumero=102&txtAno=2012&txtTipo=7&txtParte=. Acesso em 30/07/2014. 56

[19] Brasil. Portaria nº 187, de 22 de novembro de 2012. Altera a Portaria nº 323, de 13 de agosto de 2008, e a Portaria nº 4, de 4 de janeiro de 2012, que estabelecem procedimentos relativos ao cronograma de implantação de atividades e à concessão, à consolidação e à utilização de créditos do Programa Nota Legal. Câmara Legislativa do Distrito Federal, 2012. Disponível em http://www.fazenda.df.gov.br/aplicacoes/legislacao/legislacao/TelaSaidaDocumento.cfm?txtNumero=187&txtAno=2012&txtTipo=7&txtParte=. Acesso em 06/07/2014. 18

[20] Brasil. Portaria nº 4, de 4 de janeiro de 2012. Estabelece procedimentos relativos à concessão, à consolidação e à utilização de créditos no âmbito do programa instituído pela Lei nº 4.159, de 13 de junho de 2008, e dá outras providências. Câmara Legislativa do Distrito Federal, 2012. Disponível em http://www.fazenda.df.gov.br/aplicacoes/legislacao/legislacao/TelaSaidaDocument

o.cfm?txtNumero=4&txtAno=2012&txtTipo=7&txtParte=. Acesso em 06/07/2014. 74

[21] Eugênio Rubens Cardoso Braz. *Um modelo para gerenciamento, avaliação e planejamento da arrecadação de tributos estaduais*. PhD thesis, Universidade Federal de Santa Catarina, 2001. 3

[22] Cheng Chu, Sang Kyun Kim, Yi-An Lin, YuanYuan Yu, Gary Bradski, Andrew Y Ng, and Kunle Olukotun. Map-reduce for machine learning on multicore. *Advances in neural information processing systems*, 19:281, 2007. 93

[23] Helder da Silva Andrade. Um processo de mineração de dados aplicado ao combate á sonegação fiscal do ICMS. Master's thesis, Universidade Estadual do Ceará, 2009. 3

[24] Newton Oller De Mello, Marcelo Luiz Alves Fernandez, Vidal Augusto Zapparoli Castro Melo, Eduardo Mario Dias, and Caio Fernando Fontana. New technologies for nota fiscal paulista (sao paulo tax invoice): automation of the tax documents issue process in the retail of the state of sao paulo-brazil. In *Proceedings of the 8th WSEAS international conference on System science and simulation in engineering*, pages 251–258. World Scientific and Engineering Academy and Society (WSEAS), 2009. 2

[25] Douglas Downing and Jeffrey Clark. *Business Statistics*, chapter 2. Fifth edition. 7

[26] G. Mainetto F. Bonchi, F. Giannotti and D. Pedreschi. Using data mining techniques in fiscal fraud detection. In *Lecture Notes in Computer Science*, volume 1676, pages 369–376, 1999. 3

[27] Maria Manuela Caria Figueira. Identificação de outliers. 1998. 9

[28] Dan Gillick, Arlo Faria, and John DeNero. Mapreduce: Distributed computing for machine learning. *Berkley, Dec*, 18, 2006. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary10.1.1.111.9204. 93

[29] Jiawei Han and Micheline Kamber. *Data Mining - Concepts and Techniques*, chapter 1 pg 8. Third edition. 5

[30] Jiawei Han and Micheline Kamber. *Data Mining - Concepts and Techniques*, chapter 12 pg 544. Third edition. 9

[31] David J Hand, Heikki Mannila, and Padhraic Smyth. *Principles of data mining*, chapter 1. MIT press, 2001. 5, 9

[32] Ralph Kimball and Margy Ross. *The data warehouse toolkit: the complete guide to dimensional modeling*. John Wiley & Sons, 2011. 92

[33] Arthur Luis Pinho de Lima. Cidadania fiscal e o programa nota legal. 2011. Universidade de Brasília. 2

[34] Carissa Marcondes Macéa. Applying negotiation skills in the design of public policies: Analysis of the city of são paulo's invoice program. *Direito GV Research Paper Series*, (98), 2014. 2

[35] Óscar Marbán, Gonzalo Mariscal, and Javier Segovia. A data mining & knowledge discovery process model. *Data Mining and Knowledge Discovery in Real Life Applications*, pages 1–17, 2009. 21

[36] Enlinson Mattos, Fabiana Rocha, and Patrícia Toporcov. Programas de incentivos fiscais são eficazes?: evidência a partir da avaliação do impacto do programa nota fiscal paulista sobre a arrecadação de icms. *Revista Brasileira de Economia*, 67(1):97–120, 2013. 2

[37] Jesus Mena. *Data mining your website*. Digital Press, 1999. 6

[38] Alfredo Meneghetti Neto. O aumento do ICMS e a Curva de Laffer. *Análise (PUCRS), Porto Alegre*, 3(1):59–72, 1992. 93

[39] Gustavo Hermínio Salati Marcondes de Moraes. Adoção de governo eletrônico no brasil: a perspectiva do usuário do programa nota fiscal paulista. 2013. 3

[40] P. Chapman (NCR), J. Clinton (SPSS), R. Kerber (NCR), T. Khabaza (SPSS), T. Reinartz (DaimlerChrysler), C. Shearer (SPSS), and R. Wirth (DaimlerChrysler). Crisp 1.0. process and user guide, Acesso em 31/05/2014. Disponível em CRISP 1.0. Process and User Guide. 21, 22

[41] R. Stadler J. Verhess A. Zanasi P. Cabena, P. Hadjinian. *Discovering data mining: from concept to implementation*. Prentice Hall, 1997. 5

[42] Nelson Leitão Paes. A Curva de Laffer e o imposto sobre produtos industrializados - evidências setoriais. In *Cadernos de Finanças Públicas*, number 10, pages 5–22. Escola de Administração Fazendária - ESAF, Brasília, Dezembro 2010. ISSN 1806-8944. 93

[43] Bruno Vinicius Luchi Paschoal. Punição, recompensa, persuasão e ajuda: estratégias regulatórias a partir do caso nota fiscal paulista. 2012. Fundação Getúlio Vargas. 2

[44] QlikView. Business Discovery: Powerful, User-Driven BI. Technical report, Qlik, 2011. A QlikView White Paper. 92

[45] Secretaria de Estado de Trabalho. Pesquisa de Emprego e Desemprego no Distrito Federal - PED-DF. Technical report, CODEPLAN, Resultado Anual 2012. Disponível em http://www.codeplan.df.gov.br/areas-tematicas/pesquisas-socioeconomicas/258-ped .html. Acesso em 21/07/2014. 91

[46] Andreia Diniz AH Siqueira, Celene da Silva Oliveira, and Raquel Prediger Anjos. Nota fiscal paulista: uma estratégia para reduzir a sonegação fiscal. *Revista Interatividade*, pages 89–97, 2014. 2

[47] Bhavani Thuraisingham. *Data Mining: Technologies, Techniques, Tools, and Trends*, chapter 1 pg 1. CRC press, 1999. 5

[48] Patrícia Ferreira Toporcov. Evidências empíricas do efeito da nota fiscal paulista e alagoana sobre a arrecadação estadual. Master's thesis, Fundação Getúlio Vargas, 2009. 2

[49] Hermano Cláudio Vieira. Relação existente entre a diminuição da sonegação fiscal ea devolução dos tributos por meio de crédito aos contribuintes pelo governo do distrito federal. 2012. UniCEUB. 2

[50] Sholom M Weiss. *Predictive data mining: a practical guide*. Morgan Kaufmann, 1998. 6

[51] Christopher Westphal and Teresa Blaxton. *Data mining solutions: methods and tools for solving real-world problems*. Wiley, 1998. 6

[52] Ian H Witten, Eibe Frank, and Mark A Hall. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, third edition edition, 2011. ISBN 978-0-12-374856-0. 10

[53] Fu Yongjiao. Data mining: Tasks, techniques and applications. *IEEE Potentials*, 16(4):18–20, 1997. 6

Anexo A

Gráficos de Probabilidade do Perfil Crédito: ICMS Pessoa Física

A Figura A.1 apresenta o gráfico de superfície das informações tabeladas em 8.26 das variáveis sexo e idade para o ICMS de pessoas físicas.

A Figura A.2 apresenta o gráfico de superfície das informações tabeladas em 8.21 das variáveis sexo e atividade para o ICMS de pessoas físicas.

A Figura A.3 apresenta o gráfico de superfície das informações tabeladas em 8.22 das variáveis ra e sexo para o ICMS de pessoas físicas.

A Figura A.4 apresenta o gráfico de superfície das informações tabeladas em 8.23 das variáveis ra e idade para o ICMS de pessoas físicas.

A Figura A.5 apresenta o gráfico de superfície das informações tabeladas em 8.24 das variáveis ra e atividade para o ICMS de pessoas físicas.

A Figura A.6 apresenta o gráfico de superfície das informações tabeladas em 8.25 das variáveis atividade e idade para o ICMS de pessoas físicas.

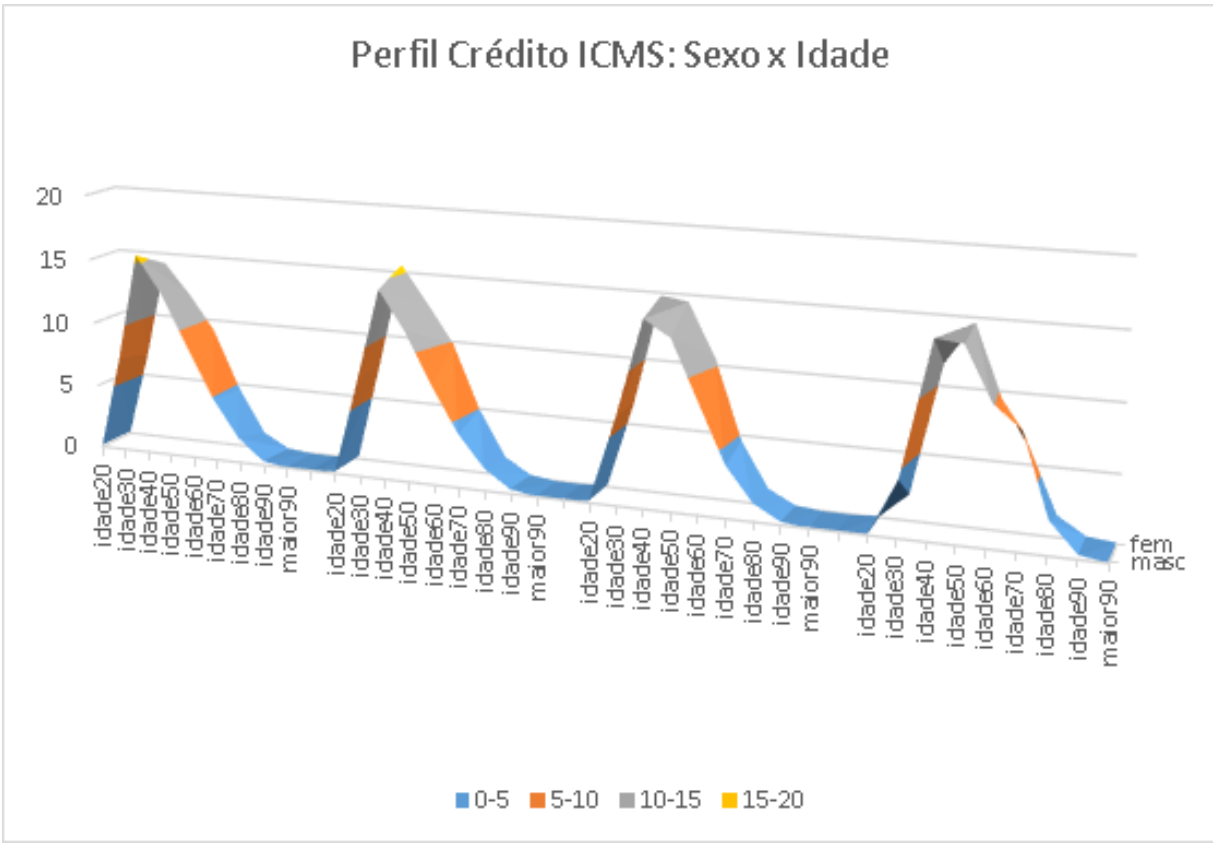

Figura A.1: Perfil Crédito ICMS: Sexo x Idade

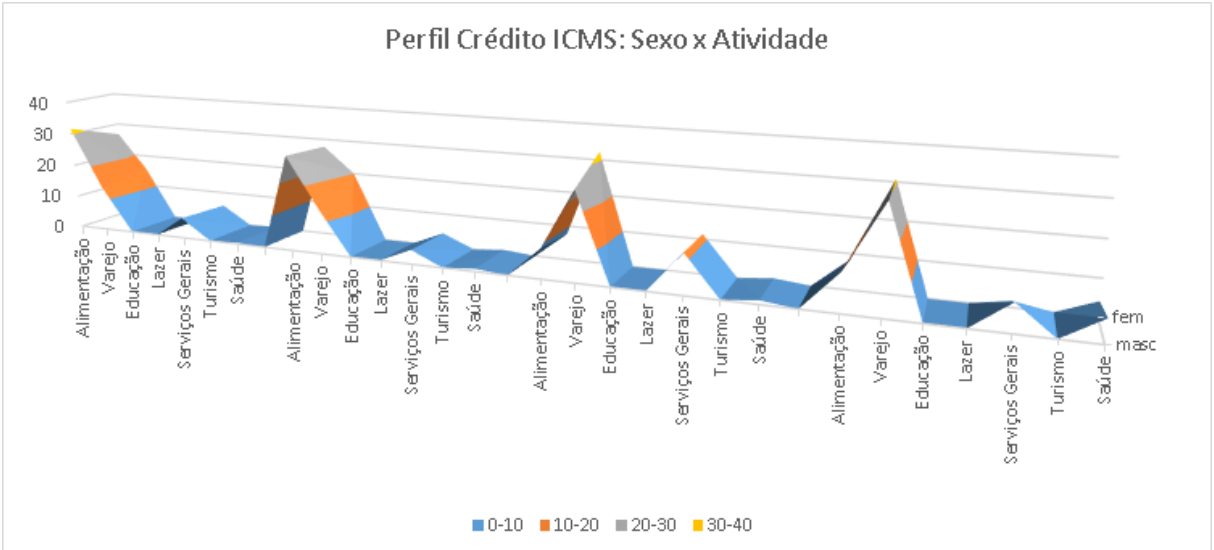

Figura A.2: Perfil Crédito ICMS: Sexo x Atividade

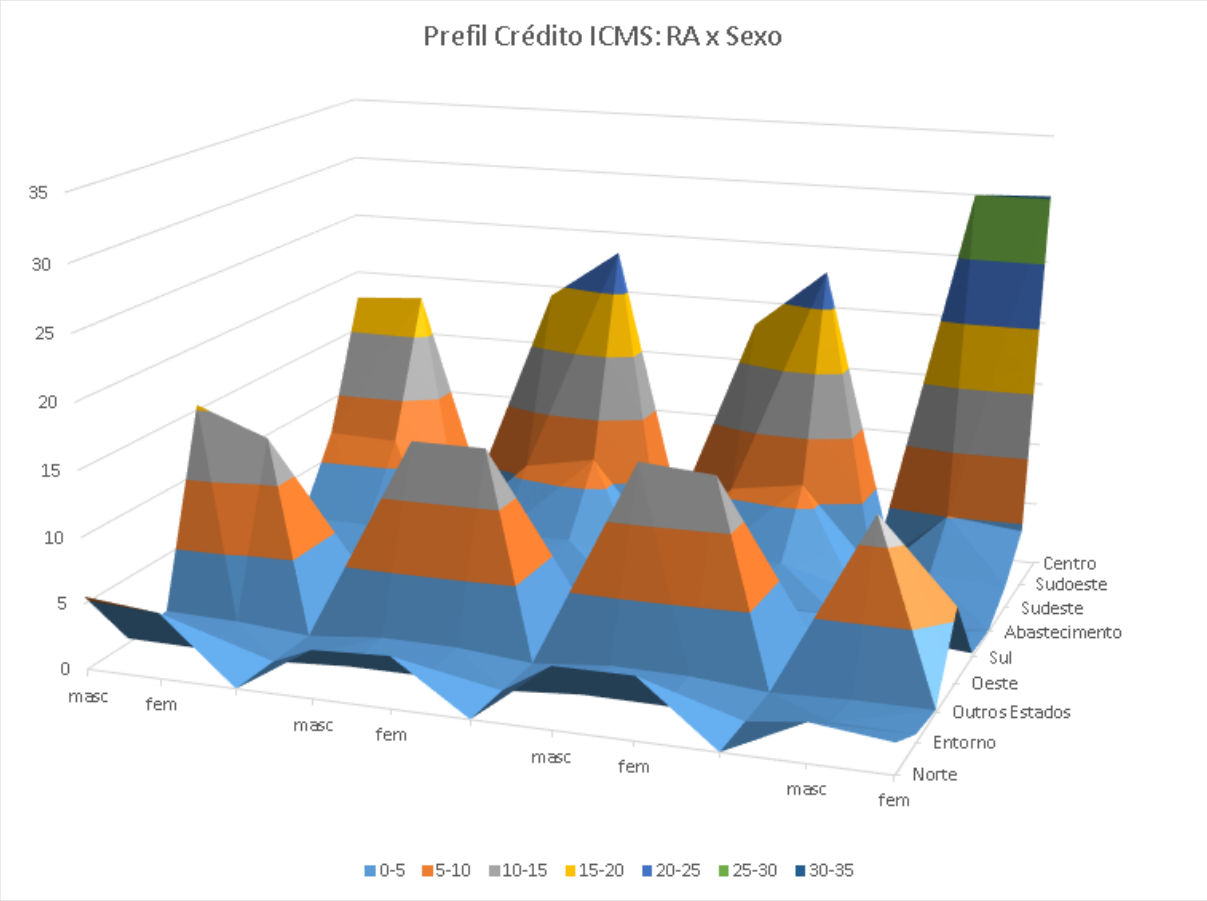

Figura A.3: Perfil Crédito ICMS: RA x Sexo

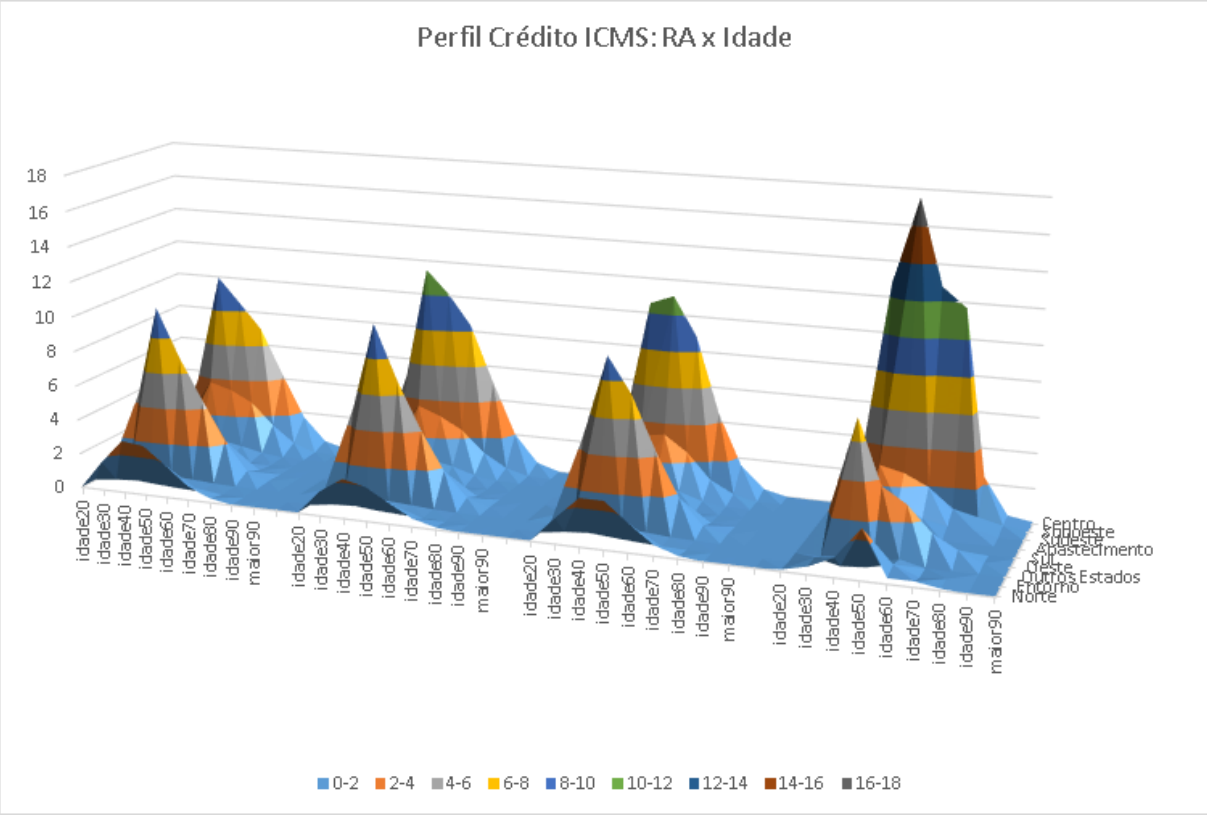

Figura A.4: Perfil Crédito ICMS: RA x Idade

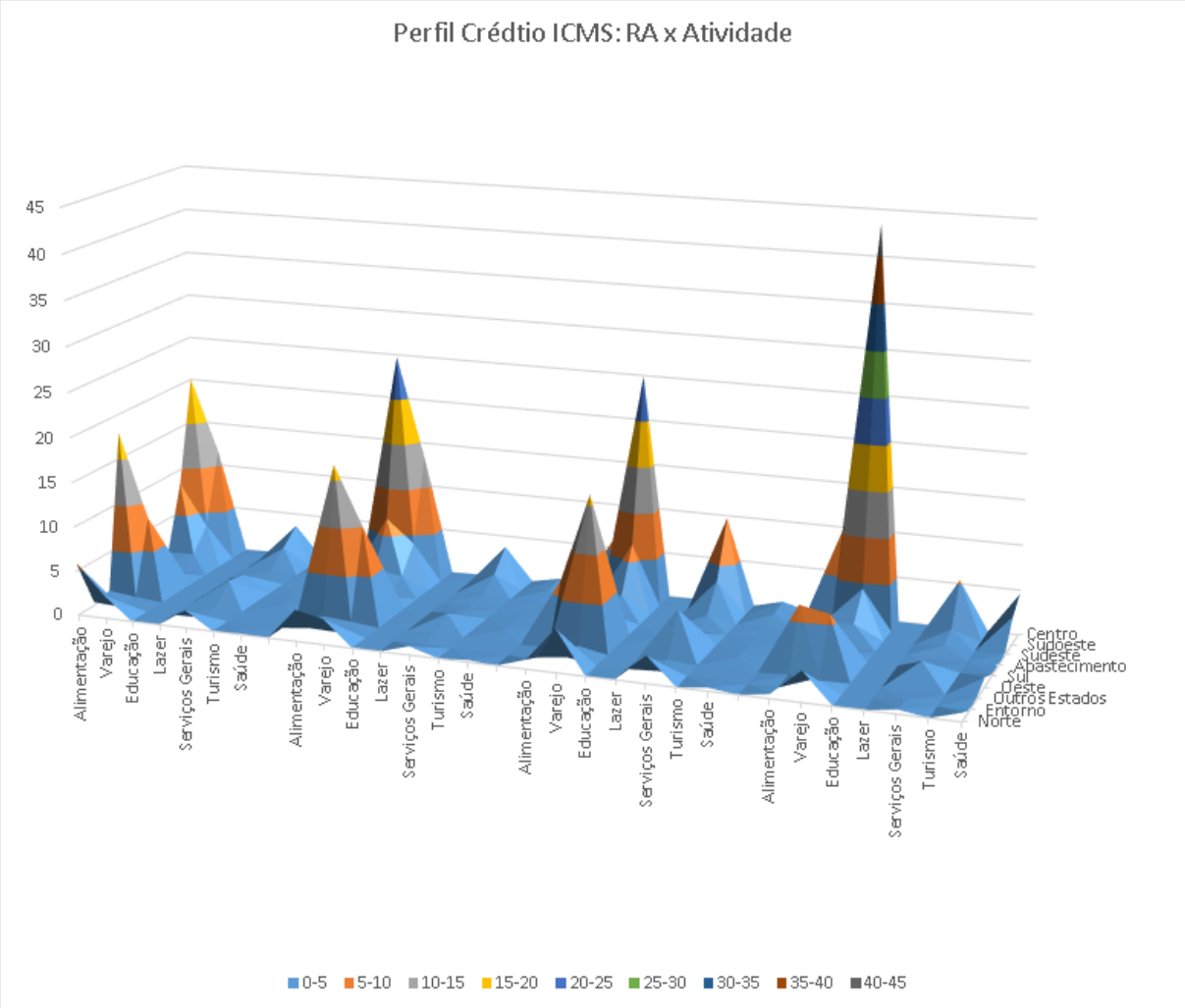

Figura A.5: Perfil Crédito ICMS: RA x Atividade

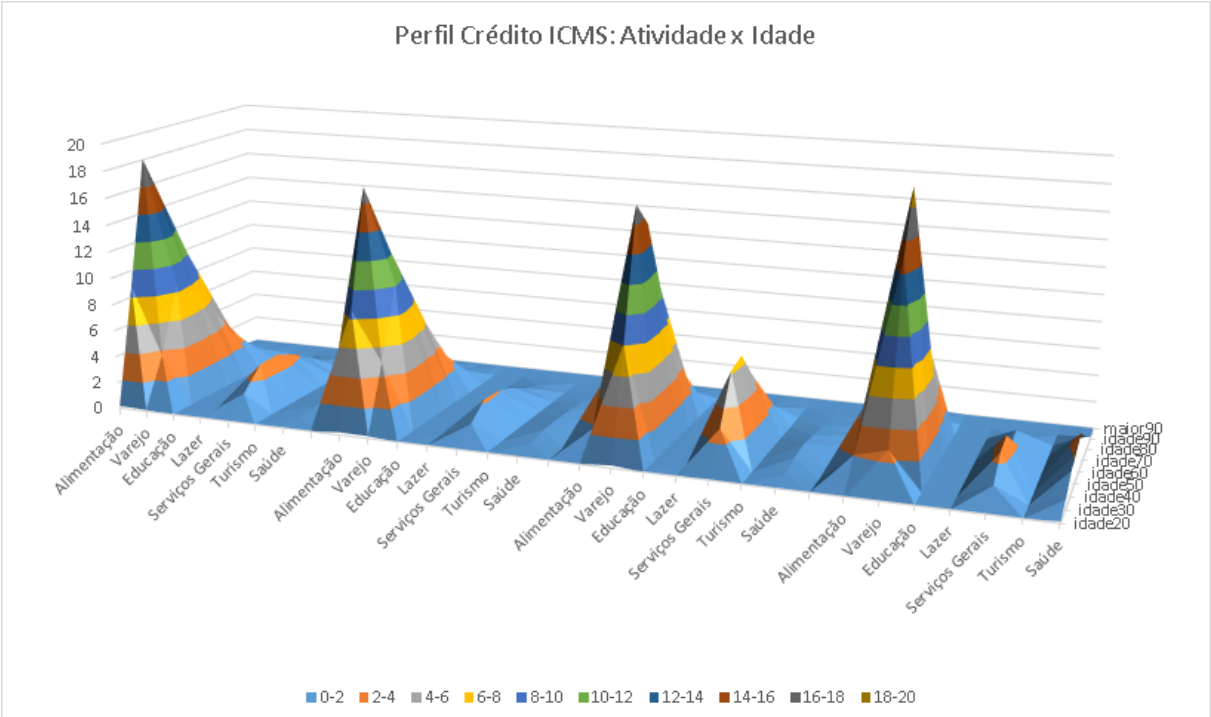

Figura A.6: Perfil Crédito ICMS: Atividade x Idade

Anexo B

Gráficos de Probabilidade do Perfil Crédito: ISS Pessoa Física

A Figura B.1 apresenta o gráfico de superfície das informações tabeladas em 8.27 das variáveis sexo e idade para o ISS de pessoas físicas.

A Figura B.2 apresenta o gráfico de superfície das informações tabeladas em 8.28 das variáveis sexo e atividade para o ISS de pessoas físicas.

A Figura B.3 apresenta o gráfico de superfície das informações tabeladas em 8.29 das variáveis RA e sexo para o ISS de pessoas físicas.

A Figura B.4 apresenta o gráfico de superfície das informações tabeladas em 8.30 das variáveis RA e idade para o ISS de pessoas físicas.

A Figura B.5 apresenta o gráfico de superfície das informações tabeladas em 8.31 das variáveis RA e atividade para o ISS de pessoas físicas.

A Figura B.6 apresenta o gráfico de superfície das informações tabeladas em 8.32 das variáveis atividade e idade para o ISS de pessoas físicas.

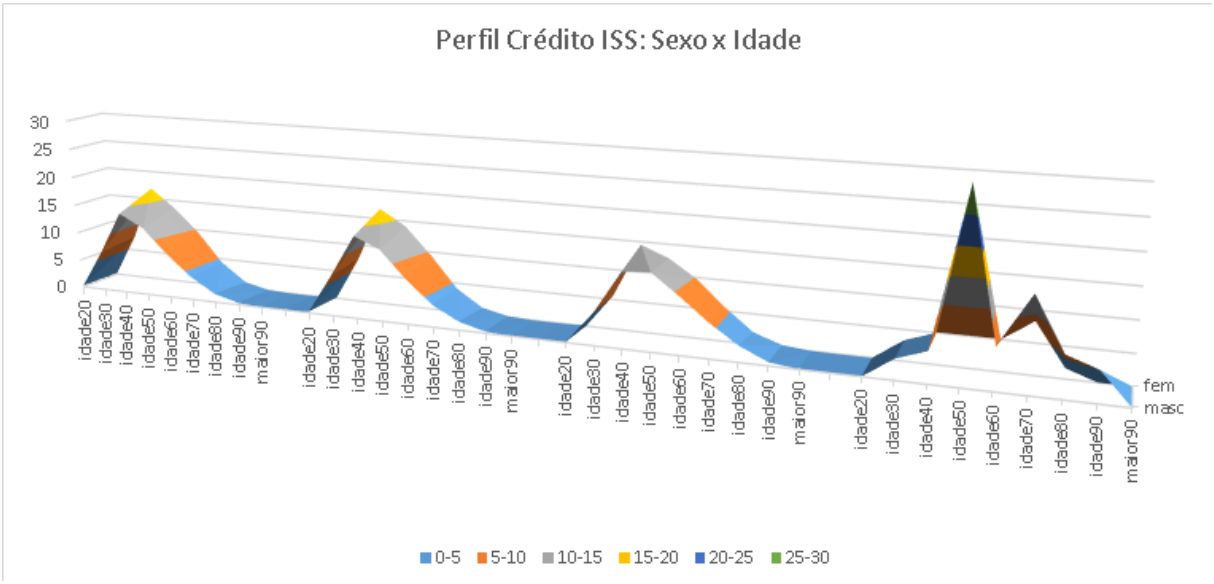

Figura B.1: Perfil Crédito ISS: Sexo x Idade

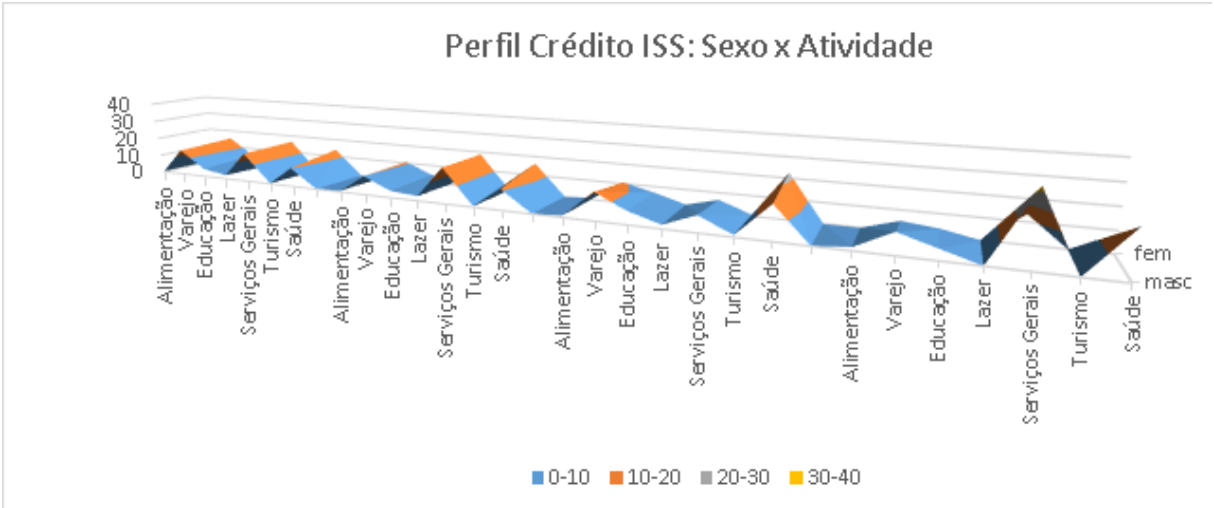

Figura B.2: Perfil Crédito ISS: Sexo x Atividade

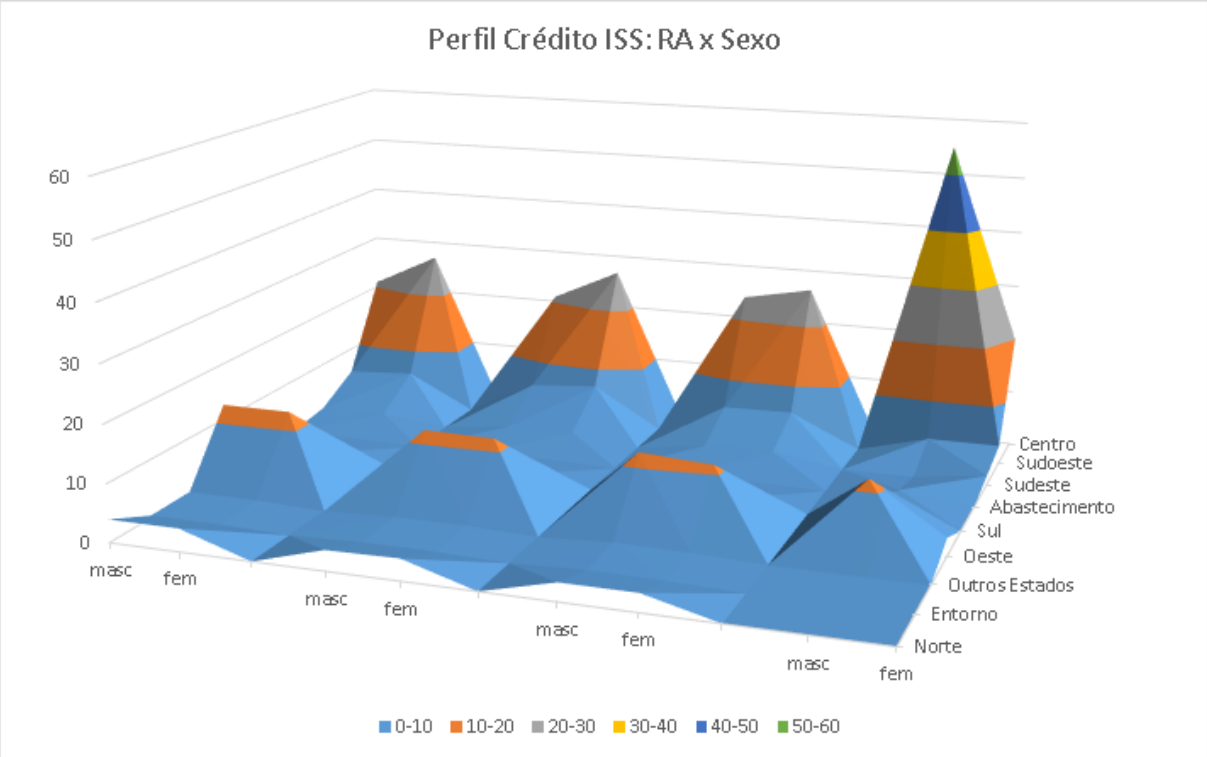

Figura B.3: Perfil Crédito ISS: RA x Sexo

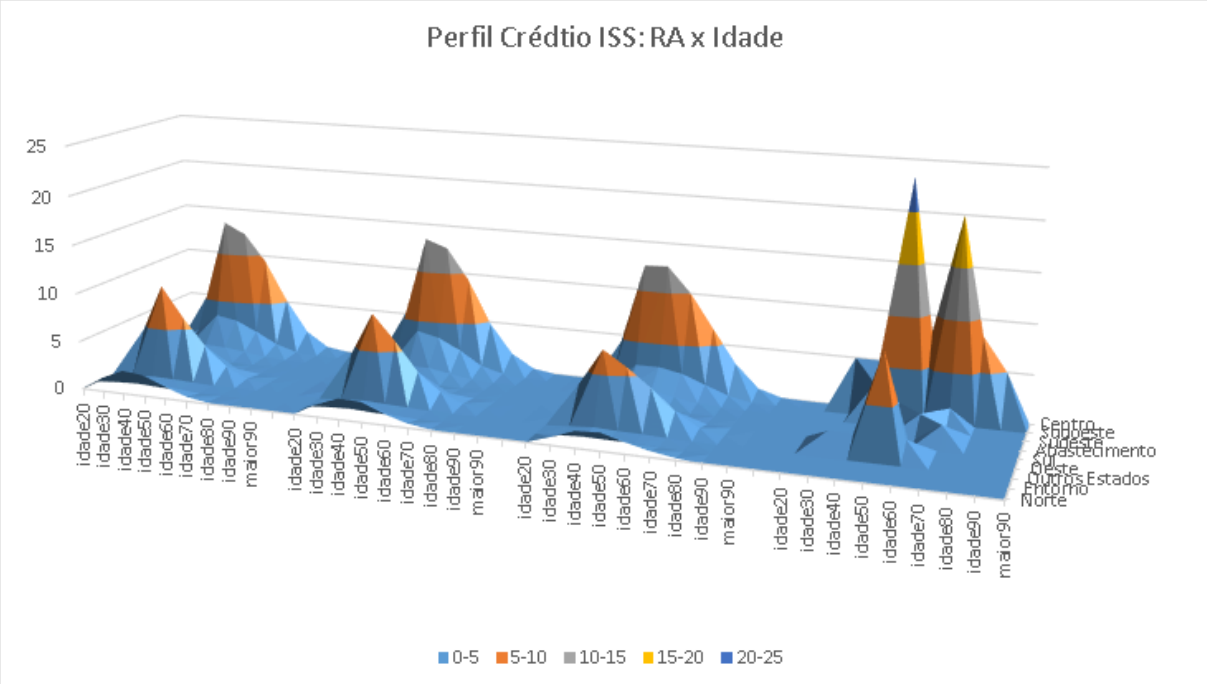

Figura B.4: Perfil Crédito ISS: RA x Idade

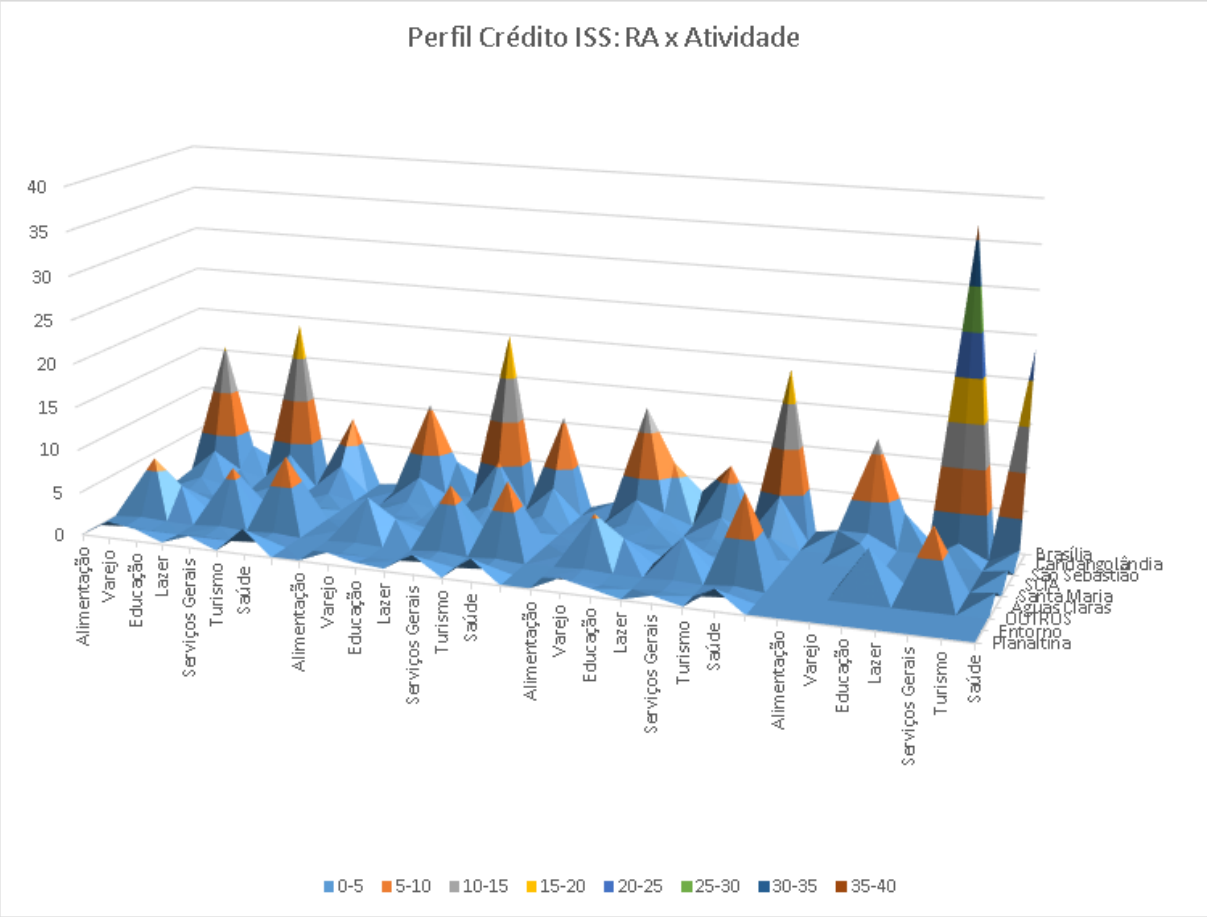

Figura B.5: Perfil Crédito ISS: RA x Atividade

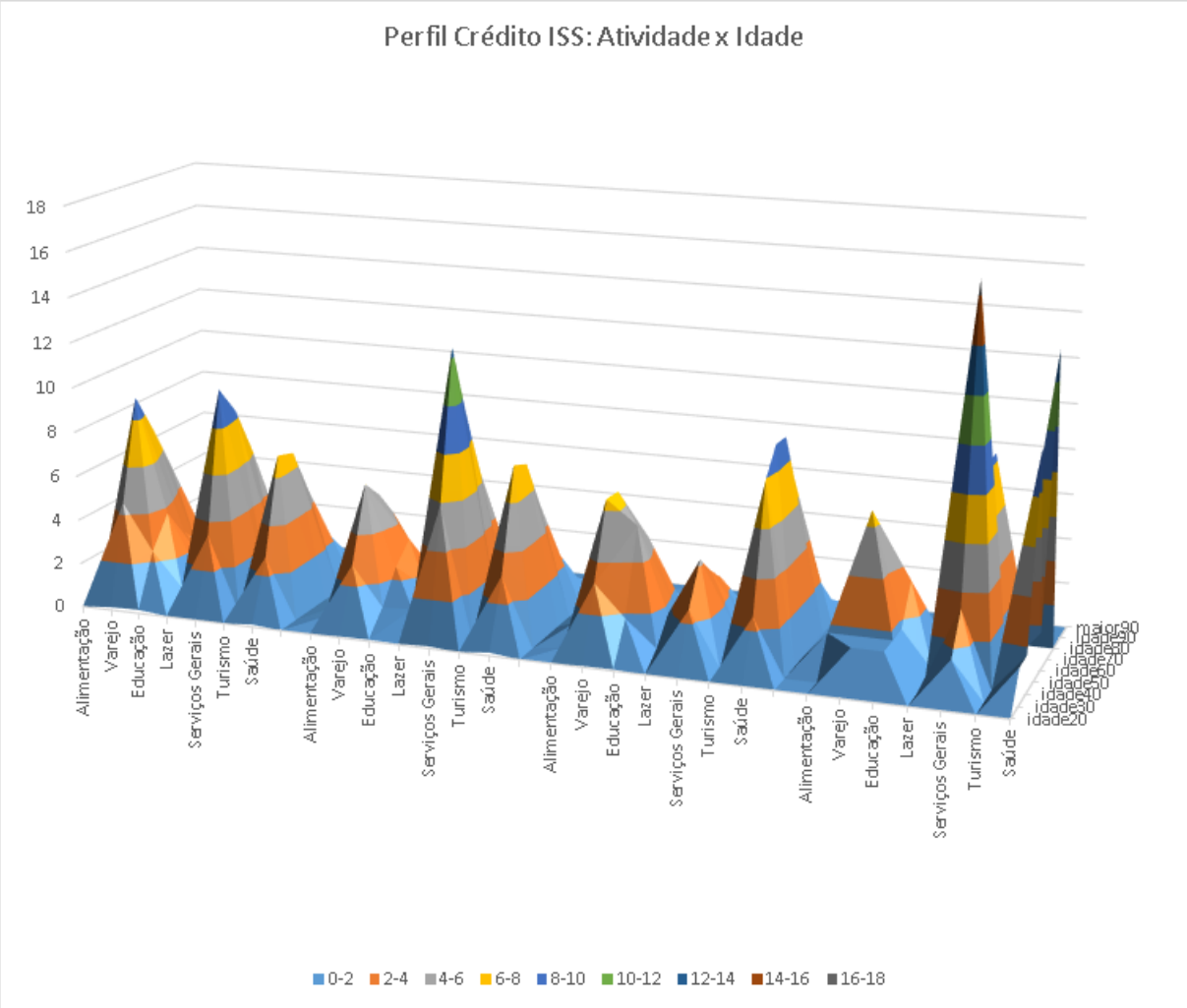

Figura B.6: Perfil Crédito ISS: Atividade x Idade

Anexo C

Gráficos de Probabilidade do Perfil Fidelidade

A Figura C.1 apresenta o gráfico de superfície das informações tabeladas em 8.10 das variáveis sexo e idade para o perfil fidelidade.

A Figura C.2 apresenta o gráfico de superfície das informações tabeladas em 8.11 das variáveis sexo e atividade para o perfil fidelidade.

A Figura C.3 apresenta o gráfico de superfície das informações tabeladas em 8.12 das variáveis ra e sexo para o perfil fidelidade.

A Figura C.4 apresenta o gráfico de superfície das informações tabeladas em 8.13 das variáveis ra e idade para o perfil fidelidade.

A Figura C.5 apresenta o gráfico de superfície das informações tabeladas em 8.14 das variáveis ra e atividade para o perfil fidelidade.

A Figura C.6 apresenta o gráfico de superfície das informações tabeladas em 8.15 das variáveis atividade e idade para o perfil fidelidade.

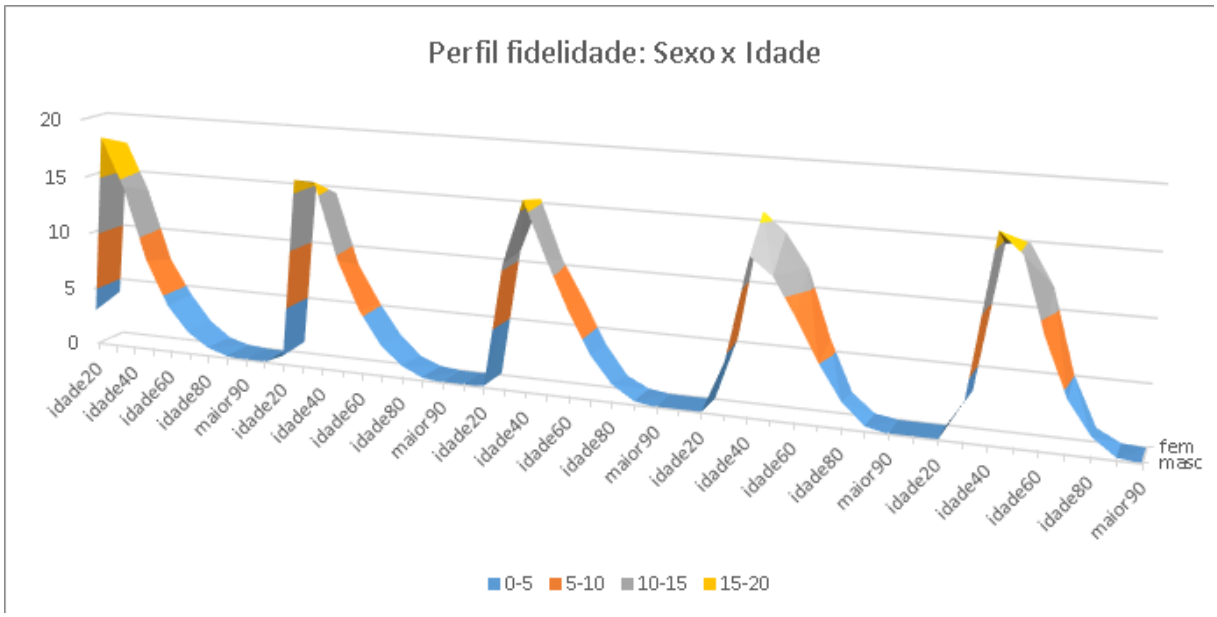

Figura C.1: Perfil fidelidade: Sexo x Idade

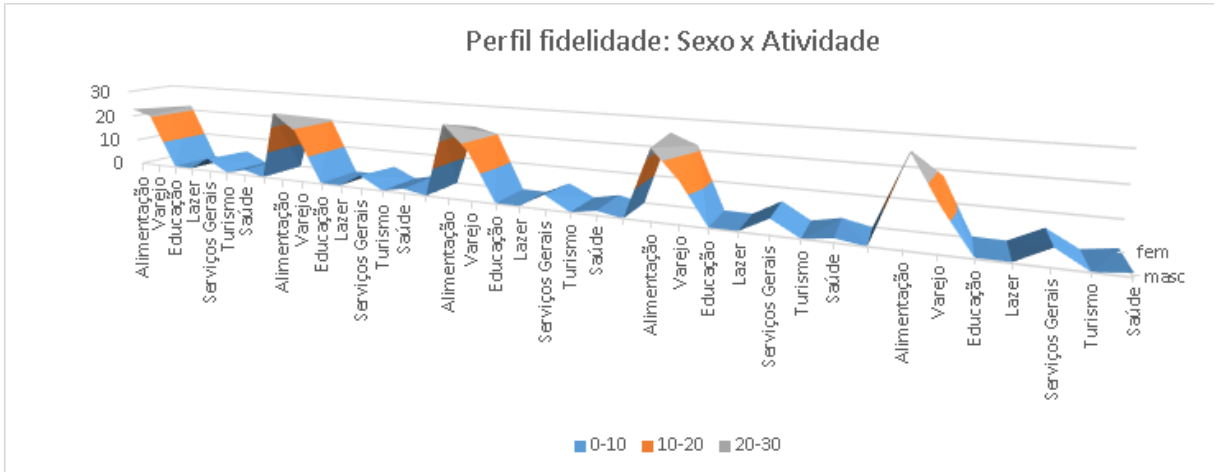

Figura C.2: Perfil fidelidade: Sexo x Atividade

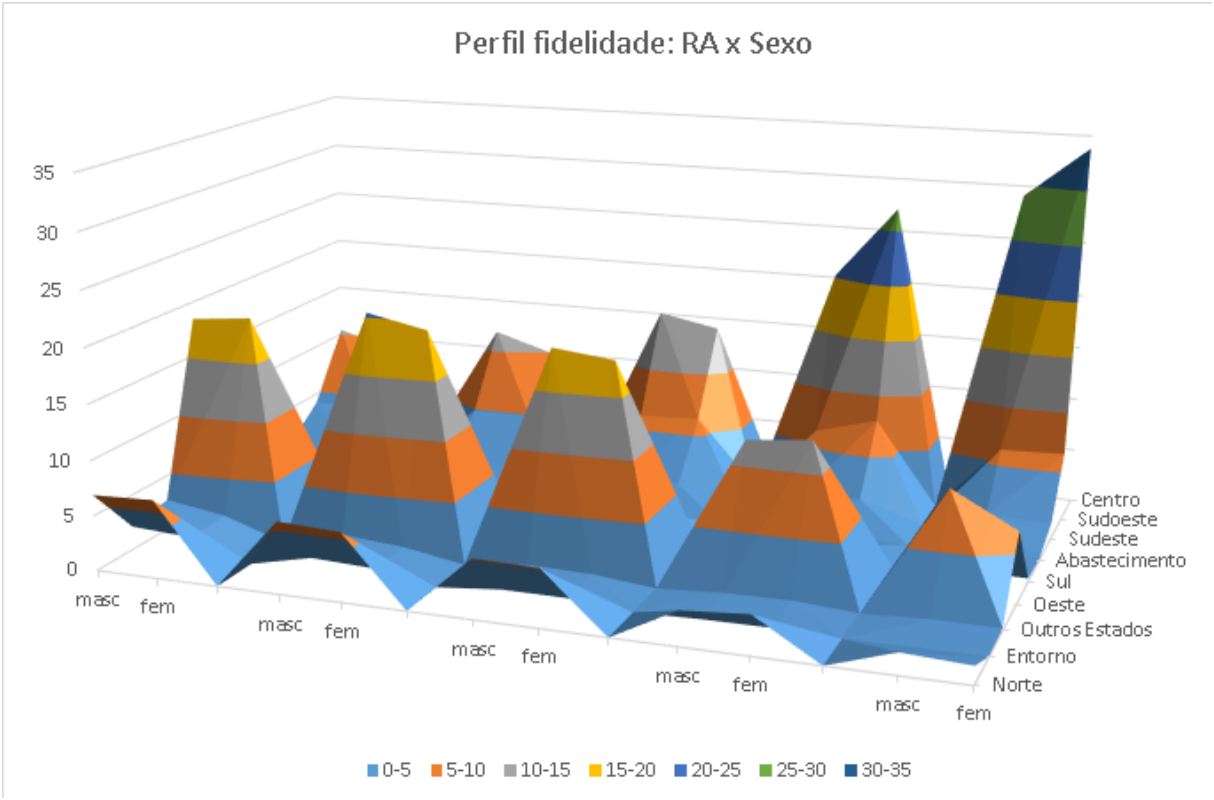

Figura C.3: Perfil fidelidade: RA x Sexo

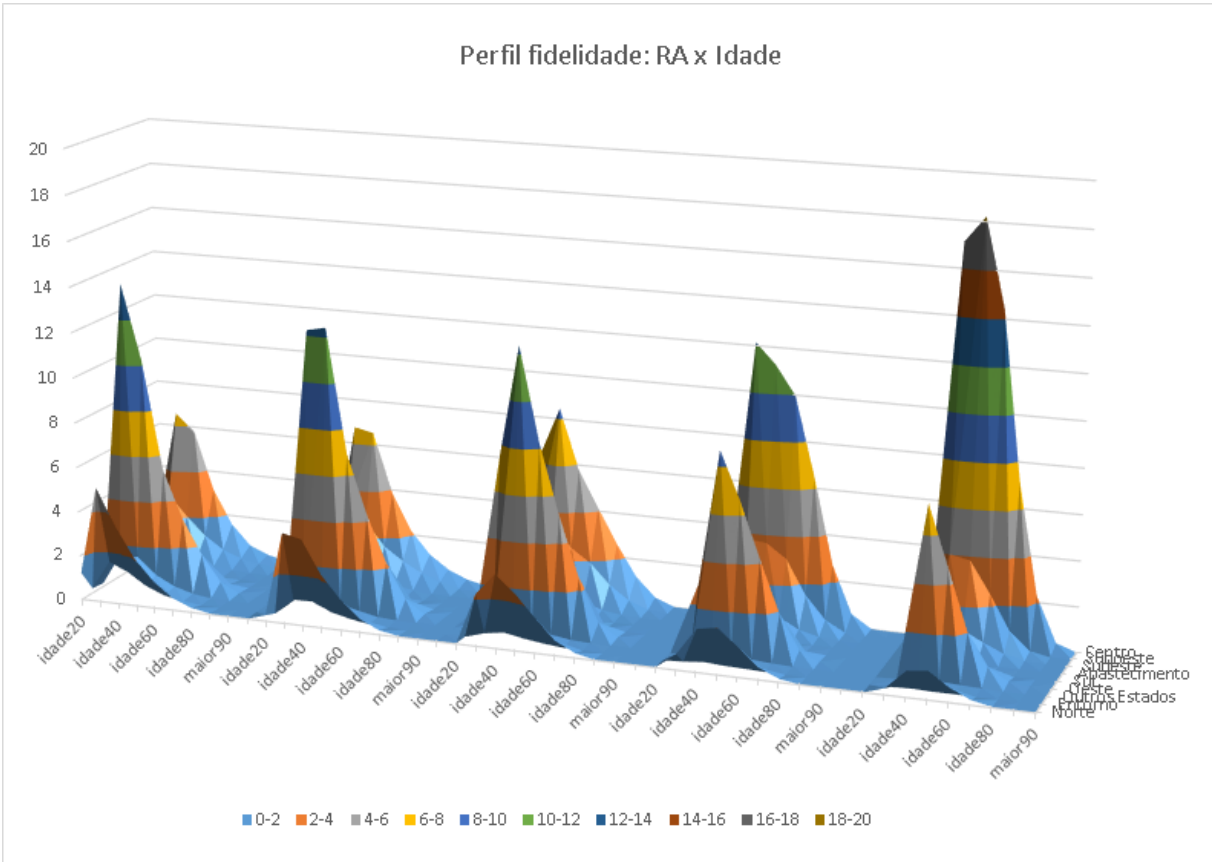

Figura C.4: Perfil fidelidade: RA x Idade

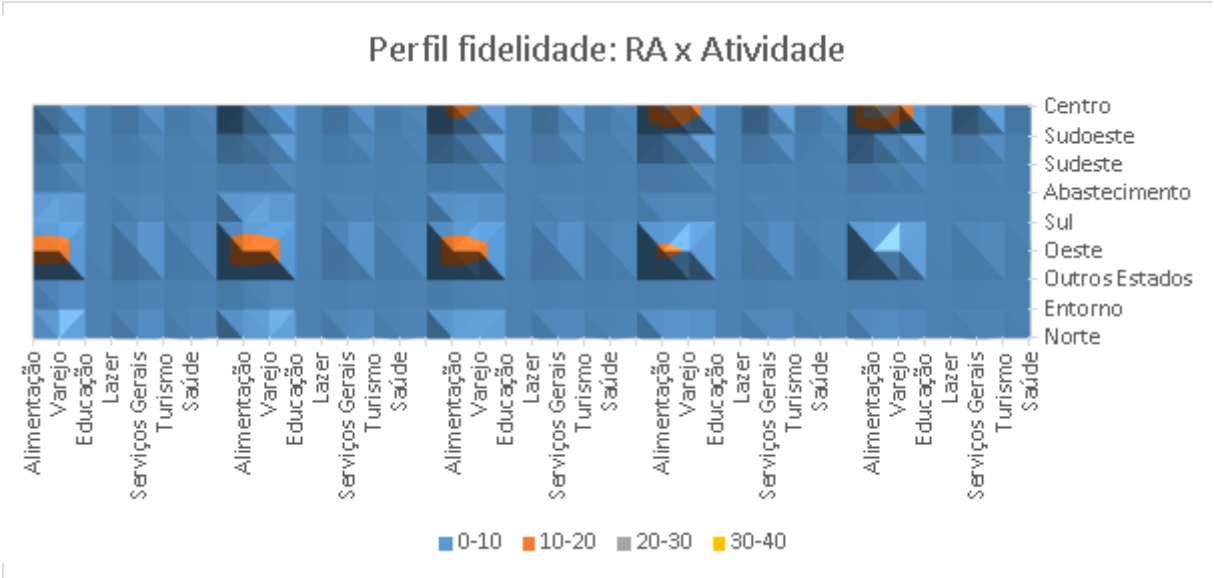

Figura C.5: Perfil fidelidade: RA x Atividade

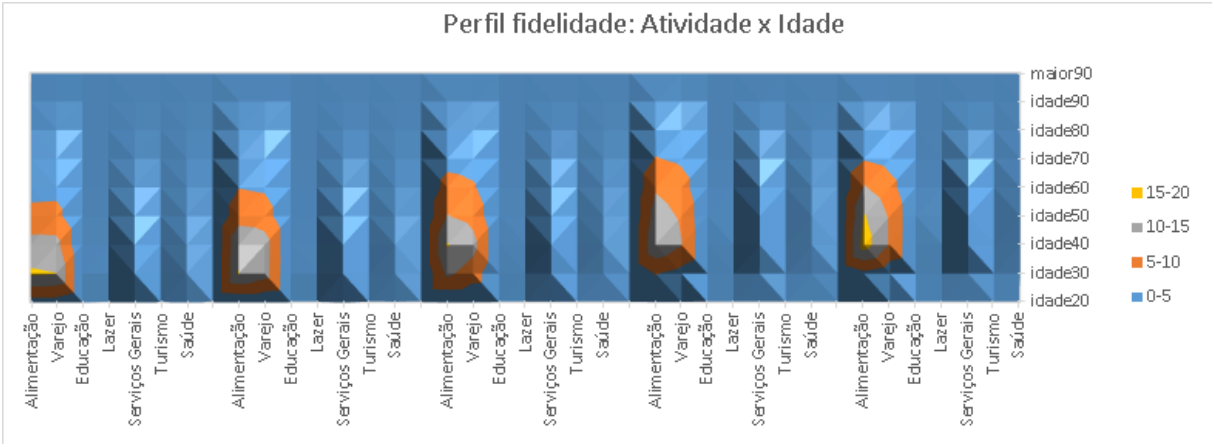

Figura C.6: Perfil fidelidade: Atividade x Idade

Anexo D

Gráfico do Perfil Crédito: ICMS Pessoa Jurídica

A Figura D.1 apresenta o gráfico de superfície das informações tabeladas em 8.33 das variáveis RA e atividade para o ICMS de pessoas jurídicas.

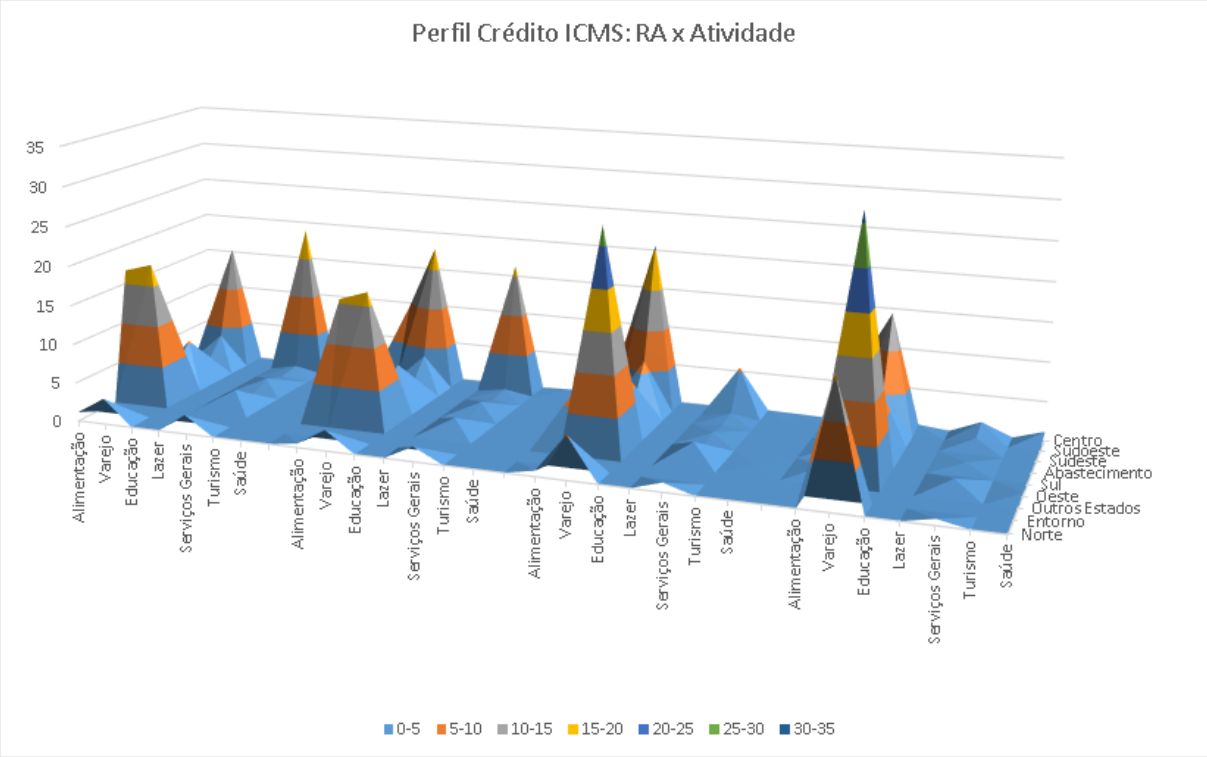

Figura D.1: Perfil Empresa Crédito ICMS: RA x Atividade

Anexo E

Gráfico do Perfil Crédito: ISS Pessoa Jurídica

A Figura E.1 apresenta o gráfico de superfície das informações tabeladas em 8.34 das variáveis RA e atividade para o ISS de pessoas jurídicas.

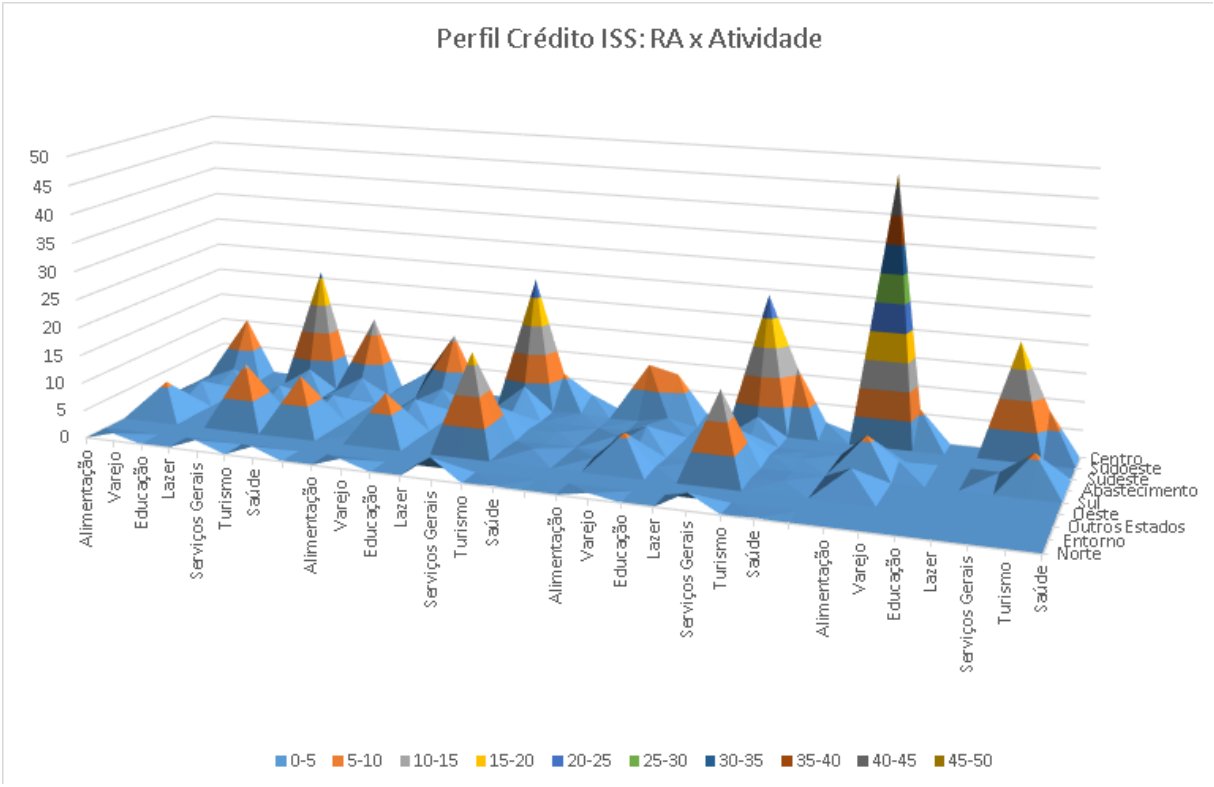

Figura E.1: Perfil Empresa Crédito ISS: RA x Atividade