



Universidade de Brasília

Faculdade de Economia, Administração e Contabilidade

Departamento de Administração

CAMILA DE ARAÚJO E SILVA

**MERCADO DE COMIDA JAPONESA NO DISTRITO
FEDERAL: Análise das Oportunidades de Negócio por
meio de Geomarketing e Máquinas de Suporte Vetorial**

Brasília – DF
2014



Universidade de Brasília
Faculdade de Economia, Administração e Contabilidade
Departamento de Administração

CAMILA DE ARAÚJO E SILVA

**MERCADO DE COMIDA JAPONESA NO DISTRITO
FEDERAL: Análise das Oportunidades de Negócio por
meio de Geomarketing e Máquinas de Suporte Vetorial**

Monografia apresentada ao
Departamento de Administração como
requisito parcial à obtenção do título de
Bacharel em Administração.

Professor Orientador: Doutor, Pedro
Henrique Melo Albuquerque.

Brasília – DF
2014

Silva, Camila de Araújo e.

Mercado de Comida Japonesa no Distrito Federal: Análise das Oportunidades de Negócio por meio de Geomarketing e Máquinas de Suporte Vetorial / Camila de Araújo e Silva. – Brasília, 2014.

143 f. : il.

Monografia (bacharelado) – Universidade de Brasília, Departamento de Administração, 2014.

Orientador: Prof. Dr. Pedro Henrique Melo Albuquerque, Departamento de Administração.

1. Geomarketing. 2. Máquinas de Suporte Vetorial. 3. Comida Japonesa. 4. Oportunidades de Negócio. 5. Distrito Federal.

CAMILA DE ARAÚJO E SILVA

**MERCADO DE COMIDA JAPONESA NO DISTRITO
FEDERAL: Análise das Oportunidades de Negócio por
meio de Geomarketing e Máquinas de Suporte Vetorial**

A Comissão Examinadora, abaixo identificada, aprova o Trabalho de
Conclusão do Curso de Administração da Universidade de Brasília da
aluna

Camila de Araújo e Silva

Doutor, Pedro Henrique Melo Albuquerque
Professor-Orientador

Doutor, Rafael Barreiros Porto
Professor-Examinador

Doutora, Marina Figueiredo Moreira
Professora-Examinadora

Brasília, 20 de novembro de 2014

“Our existence in time is determined for us, but we are largely free to select our location. This is influenced, though not dictated, by our place of origin. Finding the right location is essential to successful life, but it is essential also to a successful enterprise, to the establishment of a lasting settlement – in short, to group survival. In addition, a suitable location must be a location for the rights events.[...]”

Lösch, 1954, p. 3

RESUMO

Este estudo tem por objetivo identificar oportunidades de negócio no mercado de comida japonesa do Distrito Federal, definidas como sendo aquelas regiões onde existe, simultaneamente, uma alta demanda por comida japonesa e uma baixa oferta desses produtos, considerando-se a amostra de franquias utilizada. Para uma melhor compreensão técnica e teórica do trabalho, foram apresentados inicialmente os principais conceitos, histórico e noções introdutórias sobre os campos de estudo de Geomarketing e de Máquinas de Suporte Vetorial, que são os principais campos de estudo que embasam o desenvolvimento deste trabalho, bem como sobre a Pesquisa de Orçamentos Familiares, estudo realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), que constitui a principal fonte de dados deste trabalho. Nos procedimentos metodológicos, foram selecionados os produtos que compõe a cesta de produtos japoneses e, posteriormente foram descritas as variáveis que poderiam influenciar o consumo de comida japonesa. Procedeu-se para o treinamento da máquina e posterior aplicação do método no mercado de comida japonesa do Distrito Federal. Por meio da previsão de valores gerados pela máquina foram obtidos os dados referentes ao gasto padronizado, que fornece uma estimativa da demanda por comida japonesa para cada setor censitário do Distrito Federal. Para a identificação de oportunidades de negócio, foram cruzados os dados referentes à demanda estimada por setor censitário e ao mapeamento das 39 unidades de franquias utilizadas como amostra e que já atuam na região. Os resultados apontaram 23 possíveis oportunidades de negócio para o mercado de comida japonesa no Distrito Federal, localizadas tanto na região central do Plano Piloto, quanto em cidades satélites de Brasília.

Palavras-chave: 1. Geomarketing. 2. Máquinas de Suporte Vetorial. 3. Comida Japonesa. 4. Oportunidades de Negócio. 5. Distrito Federal.

LISTA DE FIGURAS

Figura 1 – Composição do Geocódigo	20
Figura 2 – Anéis de Thünen	25
Figura 3 – Triângulo Locacional de Weber.....	29
Figura 4 – Teoria do Lugar Central	30
Figura 5 – Sistemas de Informação Geográfica	34
Figura 6 – Marketing-Mix.....	37
Figura 7 – Aprendizado de Máquina Supervisionado.....	43
Figura 8 – Aprendizado de Máquina Não Supervisionado	44
Figura 9 – A hierarquia do Aprendizado	46
Figura 10 – Classificadores no Processo de Aprendizado	48
Figura 11 – Procedimentos Metodológicos	70
Figura 12 – <i>PrintScreen</i> Google Maps	77

LISTA DE GRÁFICOS

Gráfico 1 – Consumo <i>Per Capita</i> de Pescados	14
Gráfico 2 – Produtos Japoneses Consumidos na POF 2008-2009	65

LISTA DE MAPAS

Mapa 1 – Franquias no Distrito Federal	112
Mapa 2 – Franquias no Plano Piloto	113
Mapa 3 – Franquias nas Cidades Satélites	114
Mapa 4 – Franquias e Gasto no Distrito Federal.....	115
Mapa 5 – Oportunidades de Negócio em Brasília	116
Mapa 6 – Oportunidades de Negócio em Brazlândia	117
Mapa 7 – Oportunidades de Negócio na Candangolândia.....	118
Mapa 8 – Oportunidades de Negócio na Ceilândia	119
Mapa 9 – Oportunidades de Negócio no Cruzeiro	120
Mapa 10 – Oportunidades de Negócio no Gama	121
Mapa 11 – Oportunidades de Negócio no Guará	122
Mapa 12 – Oportunidades de Negócio no Lago Norte	123
Mapa 13 – Oportunidades de Negócio no Lago Sul.....	124
Mapa 14 – Oportunidades de Negócio no Núcleo Bandeirante	125
Mapa 15 – Oportunidades de Negócio no Paranoá	126
Mapa 16 – Oportunidades de Negócio em Planaltina	127
Mapa 17 – Oportunidades de Negócio no Recanto das Emas.....	128
Mapa 18 – Oportunidades de Negócio no Riacho Fundo.....	129
Mapa 19 – Oportunidades de Negócio em Samambaia.....	130
Mapa 20 – Oportunidades de Negócio em Santa Maria.....	131
Mapa 21 – Oportunidades de Negócio em São Sebastião.....	132
Mapa 22 – Oportunidades de Negócio em Sobradinho.....	133
Mapa 23 – Oportunidades de Negócio em Taguatinga	134
Mapa 24 – Oportunidades de Negócio no Distrito Federal.....	135

LISTA DE QUADROS

Quadro 1 – Aplicações da POF	23
Quadro 2 – Fatores de Segmentação de Mercado	36
Quadro 3 – Cesta de Produtos Japoneses	59
Quadro 4 – Classes das Variáveis	72
Quadro 5 – Classificação do Gasto Padronizado.....	74
Quadro 6 – Coordenadas Geográficas da Amostra	78
Quadro 7 – Quadro Resumo dos Objetivos Específicos	91

LISTA DE SIGLAS

- ABCJ – Associação Brasileira de Culinária Japonesa
- ABIA – Associação Brasileira das Indústrias de Alimentação
- AM – Aprendizado de Máquina
- CEP – Código de Endereçamento Postal
- CPJ – Cesta de Produtos Japoneses
- CSV – *Comma Separated Value*
- EQM – Erro Quadrático Médio
- GIS – *Geographical Information Systems*
- IBGE – Instituto Brasileiro de Geografia e Estatística
- INPC – Índice Nacional de Preços ao Consumidor
- IPCA – Índice Nacional de Preços ao Consumidor Amplo
- MPA – Ministério da Pesca e Aquicultura
- POF – Pesquisa de Orçamentos Familiares
- SVC – *Support Vector Clustering*
- SVM – *Support Vector Machines*
- TAE – Teoria do Aprendizado Estatístico

SUMÁRIO

1. INTRODUÇÃO.....	13
1.1 Contextualização.....	13
1.2 Problema de Pesquisa	15
1.3 Justificativas.....	15
1.3.1 Justificativa Prática	15
1.3.2 Justificativa Teórica	16
1.4 Objetivos	17
1.4.1 Objetivo Geral.....	17
1.4.2 Objetivos Específicos	17
2 REFERENCIAL TEÓRICO	18
2.1 A Pesquisa de Orçamentos Familiares	18
2.2 Geomarketing.....	23
2.3 Geomarketing e Segmentação de Mercado.....	35
2.4 Máquinas de Suporte Vetorial	42
3 MÉTODO.....	56
3.1 Elaboração da Cesta.....	58
3.2 Seleção das Variáveis.....	59
3.3 Treinamento da Máquina	66
3.4 Aplicação do Método para a amostra do Distrito Federal.....	71
4 RESULTADOS	76
4.1 Identificação de Oportunidades de Negócio.....	76
5 CONCLUSÕES E RECOMENDAÇÕES	90
5.1 Principais Conclusões do Estudo	90
5.2 Limitações do Estudo e Sugestões para Trabalhos Futuros	92
REFERÊNCIAS.....	94
APÊNDICES.....	99
Apêndice A – Programação utilizada no RStudio	99
Apêndice B – Mapas formados no QGIS	111
ANEXOS	136
Anexo A – Bloco de Consumo Alimentar Pessoal.....	136

1. INTRODUÇÃO

1.1 Contextualização

Diversos acontecimentos contribuíram para a chegada da cultura japonesa no Brasil. Um dos primeiros fatos aconteceu em 1908, com a chegada dos primeiros imigrantes japoneses ao porto de Santos, em São Paulo. Segundo o IBGE (2014), o número de japoneses residentes no Brasil cresceu cerca de 82,3% entre 1920 e 1970. Desde então, a cultura japonesa vem se tornando cada vez mais presente no Brasil, por meio, principalmente de sua gastronomia, de seus esportes de combate e artes marciais, entre outras expressões culturais.

A gastronomia japonesa, contudo, é a que melhor representa a nação nipônica no território brasileiro. No total, os japoneses trouxeram mais de 50 tipos de alimentos ao Brasil. Essa influência se reflete ainda mais nos dias de hoje, por meio principalmente dos restaurantes de comida japonesa. Segundo a Associação Brasileira de Culinária Japonesa – ABCJ (2014) o número de restaurantes japoneses cresceu cerca de 700% nos últimos 10 anos.

Deve-se considerar também que os gastos com alimentação fora do domicílio vêm aumentando cada vez mais. Segundo o IBGE (2010) o percentual das despesas com alimentação fora de casa na região Centro-Oeste cresceu de 24,1% para 31,1% de acordo com as Pesquisas de Orçamentos Familiares dos anos de 2002-2003 e 2008-2009, respectivamente.

Além disso, o setor encontra-se em uma fase de crescimento e aceitação do público brasileiro em geral. Segundo o Boletim Estatístico de Pesca e Aquicultura de 2011, elaborado pelo Ministério da Pesca e Aquicultura (MPA), a média do consumo de pescados por habitante/ano no País alcançou 11,17 quilos em 2011, o que representa um aumento de 14,5% com relação ao ano anterior. Em dois anos (2010 e 2011) o crescimento da demanda por peixes e frutos do mar aumentou cerca de 23,7%, em média, conforme ilustra o Gráfico 1.

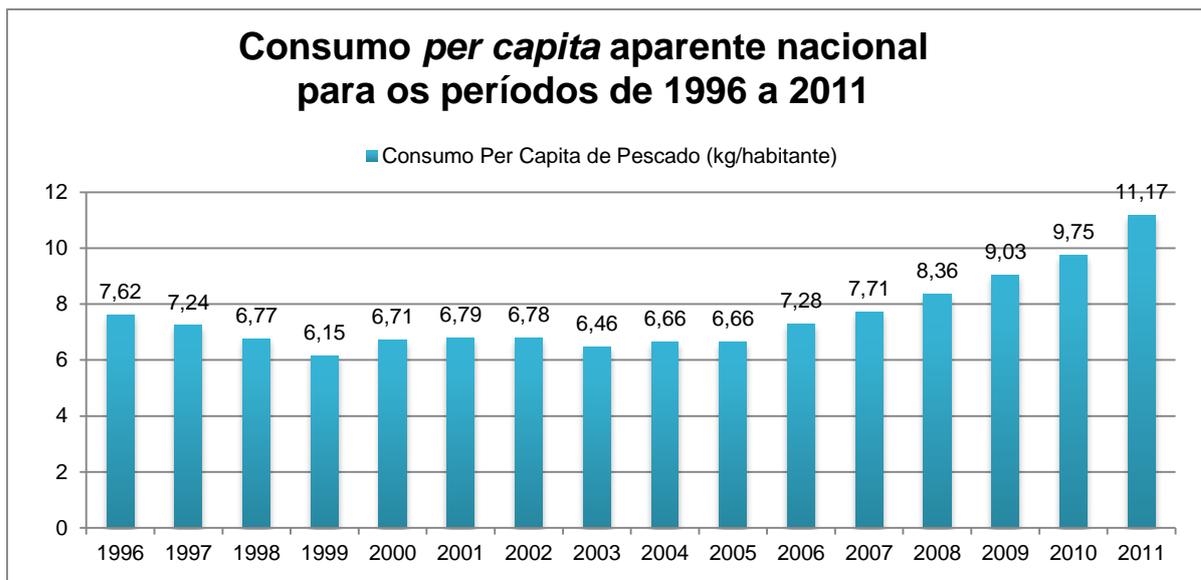


Gráfico 1 – Consumo *Per Capita* de Pescados

Fonte: Boletim Estatístico de Pesca e Aquicultura de 2011.

Dessa forma, o aumento dos gastos com alimentação fora do domicílio, somado ao crescimento do consumo de pescados refletem, conseqüentemente, um possível aumento na escolha pelos consumidores de estabelecimentos de alimentação que possuem peixes e frutos do mar como produto principal, como é o caso da culinária japonesa.

Não obstante, os consumidores desse tipo de culinária tendem a associar a comida japonesa com o título de um alimento mais saudável. Segundo a Associação Brasileira das Indústrias de Alimentação (ABIA), a difusão de informações nutricionais e o relacionamento direto entre alimentação e bem-estar fizeram com que a preocupação com a saúde e boa forma ganhasse cada vez mais espaço.

Assim, a opção por comida japonesa, com seus pratos leves e nutritivos tais como *sushi* e *sashimi*, cresceu na decisão do consumidor. A popularização do preço da comida japonesa decorrente do aumento da concorrência também é um fator fundamental para a decisão de consumo.

Dessa forma, a expansão do mercado de comida japonesa aliada à crescente aceitação por parte do público brasileiro evidencia que a abertura de um negócio nesse ramo mostra-se uma estratégia promissora do ponto de vista econômico. O gestor deve, portanto, avaliar a melhor forma de abertura para o novo negócio, de acordo com as metas e orçamento estabelecidos, a fim de se aproveitar ao máximo os benefícios do cenário atual.

Nesse contexto, o conhecimento do perfil local dos consumidores no que diz respeito à definição de um local potencialmente propenso ao consumo de alimentos típicos da culinária japonesa se mostra um fator crucial para o desenvolvimento de uma estratégia de sucesso.

1.2 Problema de Pesquisa

Segundo Masano et al. (2006), uma boa localização do empreendimento pode representar uma grande variação no volume de negócios, além de ser um fator crucial e determinante para o sucesso ou fracasso de um empreendimento. Tomando como base essa definição, fica evidente que a determinação do local ideal para o negócio pode ser um fator tão importante quanto um cenário econômico favorável ou um público receptivo.

Nesse contexto, o problema de pesquisa o qual este trabalho busca investigar pode ser definido na seguinte questão: “É possível construir uma ferramenta quantitativa que auxilie a Tomada de Decisão do Gestor quanto ao local de abertura de seu negócio utilizando Máquinas de Suporte Vetorial e dados secundários?”.

1.3 Justificativas

1.3.1 Justificativa Prática

A necessidade de um conhecimento e compreensão mais específicos do mercado, que se manifesta a partir do aumento de segmentações mais precisas, constitui, segundo Cliquet (2006), um dos pilares de desenvolvimento do Geomarketing. Assim, Cliquet (2006) explica que essa segmentação de mercado é baseada tanto no aumento de fragmentação da população observada ao longo dos últimos anos, quanto na necessidade de definição de uma estratégia de diferenciação das empresas em um contexto cada vez mais competitivo.

Dessa forma, o estudo de segmentação baseado em localização geográfica apresenta vantagens para o gestor. Além de ser capaz de determinar o potencial de consumo de uma região, o Geomarketing se mostra como uma ferramenta essencial para que as empresas definam seu público em potencial antes de abrir um novo ponto de venda ou de lançar um novo produto no mercado.

Nesse contexto, o Geomarketing desempenha o papel de cruzar dados tais como a renda média dos consumidores da região e o perfil de consumo local, definindo a localização ideal que represente o máximo possível o público-alvo que a empresa deseja atingir.

Assim, a principal justificativa prática deste trabalho consiste em propor uma abordagem capaz de auxiliar os gestores na tomada de decisão no que diz respeito à localização de seu negócio. A delimitação do público-alvo a ser atingido partir de técnicas de Aprendizado de Máquina combinada ao estudo e análise dos dados sociodemográficos da região através de técnicas de Geomarketing, caracterizam ações estrategicamente orientadas para a escolha adequada do local a ser implementado o negócio.

Pode-se citar algumas vantagens advindas a partir da adoção dessas ações como estratégias. Primeiramente a minimização dos custos provenientes da má alocação dos recursos disponíveis. Além disso, há uma nítida redução nos riscos do negócio devido ao estudo prévio das características do mercado local. Sem a localização adequada, a empresa pode deixar de identificar oportunidades de aumento de cobertura, de expansão de mercado, ter prejuízos ou até mesmo ir à falência por não possuir um público local suficiente para sustentar o negócio.

1.3.2 Justificativa Teórica

Cui e Curry (2005) afirmam que até o presente momento não existem aplicações de Máquinas de Suporte Vetorial (ou simplesmente SVM, do inglês *Support Vector Machines*) reportadas na literatura de Marketing. Apesar de serem relativamente pouco estudadas na área de Marketing, os autores ressaltam que as SVMs têm demonstrado sua utilidade em uma grande variedade de outras disciplinas, tais como estatística, ciência da computação, agricultura e engenharia.

Dessa forma, a principal justificativa teórica deste trabalho consiste na contribuição científica fornecida a um campo de estudo pouco explorado, especialmente no Brasil, tendo em vista que não foram encontrados nenhum trabalho científico de autoria brasileira que envolvesse a aplicação de Máquinas de Suporte Vetorial em Marketing.

Não obstante, as técnicas e procedimentos aqui demonstrados também apresentam possibilidade de aplicação no processo decisório de abertura de negócios em quaisquer outros ramos de atuação do mercado, o que evidencia o grande valor agregado ao trabalho, não apenas para os gestores, como também para a academia como um todo.

1.4 Objetivos

1.4.1 Objetivo Geral

Propor um método que auxilie os gestores a tomarem decisão no que diz respeito à localização de seu negócio a partir da utilização de dados secundários, técnicas de Geomarketing e Máquinas de Suporte Vetorial.

1.4.2 Objetivos Específicos

- I. elaborar uma cesta com os produtos relacionados à culinária japonesa identificados na Pesquisa de Orçamentos Familiares 2008-2009;
- II. selecionar, a partir da Pesquisa de Orçamentos Familiares 2008-2009, as variáveis que estariam relacionadas com o consumo de comida japonesa;
- III. programar a leitura dos dados da Pesquisa de Orçamentos Familiares 2008-2009 e a construção de Máquinas de Suporte Vetorial a ser utilizada para amostra do Distrito Federal;
- IV. aplicar o método proposto ao mercado de comida japonesa do Distrito Federal para a identificação das regiões mais propensas ao consumo desses produtos;
- V. estimar a demanda por comida japonesa para cada setor censitário do Distrito Federal por meio da previsão de valores de gasto padronizado gerada pelo aprendizado de máquina;
- VI. identificar os locais com oportunidades de negócio no Distrito Federal a partir da análise da localização das franquias utilizadas como amostra, dos resultados obtidos pela Máquina de Suporte Vetorial e dos dados da Pesquisa de Orçamentos Familiares 2008-2009.

2 REFERENCIAL TEÓRICO

2.1 A Pesquisa de Orçamentos Familiares

A Pesquisa de Orçamentos Familiares (POF) é um estudo realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE) cujo principal objetivo consiste em obter, em um determinado período, informações referentes à composição orçamentária das famílias brasileiras.

Nesse contexto, os dados obtidos por meio da POF possibilitam o conhecimento de informações dessas famílias em três principais esferas: econômica, social e alimentar. Assim, torna-se possível traçar um perfil de consumo dessas famílias.

No que diz respeito à esfera econômica, a POF permite conhecer a situação financeira das famílias a partir do fornecimento de informações referentes tanto ao rendimento familiar (receita), quanto à destinação desses recursos (despesa). A POF revela também informações referentes à composição dos gastos e quais foram os bens ou produtos consumidos e serviços utilizados nesse período.

Na esfera social, a POF fornece informações sobre as condições de vida da população, bem como uma abordagem acerca da qualidade de vida percebida por essa população. Essas informações são levadas em consideração sob a ótica dos orçamentos familiares.

Por fim, na esfera alimentar, a POF possibilita a geração de bases de dados e a realização de estudos referentes ao perfil nutricional da população a partir da análise dos dados relacionados às despesas com saúde, às características antropométricas, bem como dados sobre o consumo alimentar dessas famílias, obtidos ao longo da pesquisa.

2.1.1 Histórico da POF

O Estudo Nacional da Despesa Familiar, realizado pelo IBGE no período entre 1974 e 1975, foi à primeira pesquisa a contemplar em seu escopo a estrutura orçamentária doméstica brasileira.

Os resultados obtidos nessa pesquisa possibilitaram a criação dos índices de inflação medidos pelo IBGE: o Índice Nacional de Preços ao Consumidor (INPC) e outros índices derivados, como por exemplo, o Índice Nacional de Preços ao Consumidor Amplo (IPCA).

Anteriormente, as pesquisas eram limitadas a apenas grandes cidades, o que restringia, em maior ou menor grau, a generalização dos dados obtidos. Com o passar dos anos, a pesquisa passou a contemplar todos os estados e teve sua metodologia aperfeiçoada para obtenção de dados cada vez mais precisos e relevantes para a sociedade.

Até o momento, já foram realizadas cinco pesquisas de orçamentos familiares no Brasil. A primeira, como já citado, ocorreu entre 1974 e 1975, e a segunda, entre 1987 e 1988. Já a terceira pesquisa, foi realizada em 1995 e 1996, enquanto que a quarta foi realizada em 2002 e 2003. Por fim, a última pesquisa de orçamentos familiares realizada, e que será a utilizada nesse trabalho, está compreendida entre o período de 2008 e 2009.

2.1.2 Amostra da POF

A Pesquisa de Orçamentos Familiares 2008-2009 foi feita por meio da utilização de uma técnica estatística denominada amostragem aleatória simples. Por motivos operacionais, durante a realização dos Censos Demográficos os municípios do Brasil são divididos em várias regiões distintas.

A primeira e maior região diz respeito às Unidades da Federação. As Unidades da Federação são compostas por delimitações menores, denominadas Municípios. Já os Municípios são subdivididos em Distritos e estes, por sua vez, são subdivididos em Subdistritos. Por fim, tem-se que os Subdistritos são então subdivididos em regiões ainda menores, denominadas setores censitários, que constituem a menor unidade territorial da pesquisa. A combinação numérica dos códigos da Unidade da Federação, do Município, do Distrito, do Subdistrito e do Setor Censitário compõe o chamado Geocódigo, conforme ilustra a Figura 1.

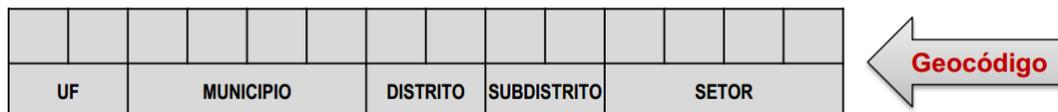


Figura 1 – Composição do Geocódigo

Fonte: IBGE (2011)

Primeiramente, foram selecionados aleatoriamente setores censitários para cada uma das Unidades da Federação para compor a amostra. Esses setores escolhidos constituem, por sua vez, as denominadas subamostras de setores que serão utilizadas na POF.

A seleção desses setores por meio da técnica de amostragem se deu de tal modo que a probabilidade de escolha esteja proporcionalmente relacionada à quantidade de domicílios que existem no referido setor, dentro do conjunto de setores censitários que compõem a amostra mestra.

Posteriormente, foram selecionados de forma aleatória e sem reposição domicílios tanto de áreas urbanas, quanto de áreas rurais, dentro de cada um dos setores selecionados. A soma das amostras de todas as Unidades da Federação do Brasil totaliza mais de 68.000 domicílios (IBGE, 2010).

2.1.3 Conceitos da POF

Para uma melhor compreensão metodológica da Pesquisa de Orçamentos Familiares, faz-se necessário a introdução de alguns conceitos e definições que serão utilizados ao longo deste trabalho.

O termo Domicílio refere-se à unidade amostral da pesquisa. É uma das principais unidades de investigação e análise dessa pesquisa, pois é a partir do domicílio que serão obtidas informações relevantes a respeito das condições de moradia das famílias (IBGE, 2010).

Já o termo Unidade de Consumo refere-se à unidade básica de investigação e análise dos orçamentos familiares. No contexto da Pesquisa de Orçamentos Familiares, o termo unidade de consumo é tido como sinônimo do termo família. Dessa forma, para a POF, os termos “família” e “unidade de consumo” são equivalentes do ponto de vista conceitual. Além disso, tem-se que o tamanho da unidade de consumo

corresponde ao número de pessoas que compõe aquela unidade de consumo, ou família (IBGE, 2010).

O termo Pessoas se refere aos indivíduos que residem em determinado domicílio, sendo considerados moradores presentes aqueles que se encontravam na residência durante a realização da pesquisa, e moradores ausentes aqueles que não estavam presentes no momento da pesquisa, mas que habitualmente residem no domicílio. Assim, o número total de pessoas do domicílio corresponde à soma dos moradores ausentes e presentes durante a realização da pesquisa (IBGE, 2010).

Já a Unidade de Orçamento corresponde àquele morador que participou de alguma forma no orçamento daquela unidade de consumo, durante o período de pesquisa. Nesse contexto, a pessoa de referência da unidade de consumo, considerada como sendo o “chefe de família”, é aquela responsável por despesas relacionadas à habitação (aluguel, prestação de imóvel, condomínio, imposto, etc). No que diz respeito às características dessas pessoas, serão investigadas em cada uma das unidades de consumo o gênero dos moradores, nível de escolaridade, a raça, a idade, dentre outras variáveis (IBGE, 2010).

Com relação às despesas, foram investigados todos os tipos de despesas, monetárias e não monetárias, referentes ao consumo de bens e serviços de qualquer natureza, despesas correntes, despesas de consumo relacionadas à alimentação, habitação, transporte, higiene pessoal, vestuário, educação, entretenimento, saúde, entre outros. Para o presente trabalho, serão utilizados os dados referentes às despesas com alimentação (IBGE, 2010).

Outras duas informações investigadas pela POF são: a renda familiar, considerando todos os rendimentos obtidos pelos moradores integrantes da unidade de consumo em estudo; e a avaliação subjetiva a respeito das condições de vida. Este último não será utilizado neste trabalho (IBGE, 2010).

2.1.4 Metodologia da POF

A Pesquisa de Orçamentos Familiares possui uma duração de 12 meses. Existem dois principais motivos capazes de justificar esse período de duração. O primeiro, diz respeito à dificuldade de operacionalização da pesquisa, tendo em vista

sua abrangência territorial. Outro motivo é que a POF considera também em suas investigações as alterações de consumo, rendimentos e despesas que ocorrem ao longo de um ano (IBGE, 2010).

Durante esse período total de 12 meses, diversos pesquisadores visitam cada um dos domicílios selecionados durante um período menor de nove dias consecutivos, com o intuito de coletar os dados da POF (IBGE, 2010).

Nesse contexto, o principal método para se obter os dados dos orçamentos das famílias foi a aplicação de diversos questionários por meio de entrevista presencial com os moradores. Os pesquisadores aplicavam questionários com temas específicos, dentro do período de nove dias estabelecido para a visita ao domicílio (IBGE, 2010).

A forma de registro das informações sobre as aquisições feitas durante esse período variava de acordo com o valor e a frequência do bem ou serviço adquirido. Ou seja, para as aquisições que ocorriam com frequência, como por exemplo, transporte e alimentação, o registro era feito diariamente (IBGE, 2010).

Devido às características desse tipo de levantamento de informações, bem como devido ao grau de complexidade e detalhamento de algumas variáveis investigadas na pesquisa, optou-se pela utilização de computadores portáteis para o registro dos dados coletados (IBGE, 2010).

2.1.5 Aplicações da POF

Tendo em vista a abrangência temática envolvida na Pesquisa de Orçamentos Familiares, tem-se que os dados coletados nessa pesquisa podem ser utilizados nas mais diversas áreas de estudo. Alguns exemplos das principais aplicações a partir dos resultados obtidos na POF podem ser descritos, conforme descrito no Quadro 1.

ÁREA DE ESTUDO	APLICAÇÃO
Políticas Públicas	I. Subsidia a criação políticas públicas nas mais diversas áreas, tais como: <ul style="list-style-type: none"> a) Educação; b) Saúde; c) Moradia; d) Social; e) Nutricional; II. Possibilita o desenvolvimento e a gestão de políticas públicas para melhorias nas

	<p>condições e qualidade de vida da população;</p> <p>III. Permite a construção de indicadores baseados em dados socioeconômicos, capazes de mensurar e analisar a pobreza e a desigualdade social.</p>
Economia	<p>I. Auxilia na definição de estratégias de investimentos;</p> <p>II. Atualiza e pondera medidas de inflação;</p> <p>III. Estima o consumo nos níveis nacional e regional;</p>
Nutrição	<p>I. Possibilita a visualização do estado antropométrico da população;</p> <p>II. Viabiliza ações voltadas à orientação e segurança alimentar da população.</p>
Análise de Mercado	<p>I. Permite o conhecimento do perfil do consumidor e da demanda por bens e serviços.</p>

Quadro 1 – Aplicações da POF

Fonte: Elaborado pela autora.

Dessa forma, torna-se nítido o potencial que a Pesquisa de Orçamentos Familiares tem, tanto para o Setor Público, quanto para o Privado. A POF fornece uma rica base de dados que pode ser utilizada em diversas aplicações capazes de beneficiar, direta ou indiretamente, toda a sociedade.

2.2 Geomarketing

O processo de globalização proporciona a integração social, política, econômica e cultural em um nível internacional, trazendo uma nova configuração para o cenário mundial atual. Diante dessas mudanças, a necessidade de conhecimento do ambiente geográfico no qual estamos inseridos adquire uma nova dimensão e valor para o processo decisório nas organizações.

As constantes mudanças no ambiente externo, como por exemplo, as atividades de outras organizações, ações dos concorrentes, mudanças nas necessidades dos clientes e mudanças no mercado, fazem com que o entendimento geográfico se torne uma ferramenta de extrema relevância no processo decisório, por tornar viável a integração entre os objetivos da organização e os recursos e tempo disponíveis (TAKETA, 1993).

Entretanto, a adoção de um plano rígido e de longo prazo em um ambiente tão mutável pode determinar o fracasso de uma organização, pois dificulta a criação de novas abordagens e a identificação de ameaças e oportunidades. Assim, devido à

constante necessidade de adaptação e resolução de problemas em prazos cada vez mais curtos, os gestores vêm substituindo o planejamento estratégico pelo pensamento estratégico (TAKETA, 1993). Nesse contexto, pensamento estratégico é definido como sendo um processamento contínuo de informações internas e externas, de forma a possibilitar o ajuste às mais diversas situações. Segundo o mesmo autor, o entendimento geográfico compõe esse pensamento estratégico e pode ajudar a enfrentar esses problemas organizacionais.

Contudo, o conhecimento espacial implica no conhecimento dos agentes que atuam em nosso sistema. De um lado temos a oferta dos estabelecimentos, que configuram a concorrência, e de outro temos a demanda, representada pelos clientes e consumidores. Todos esses agentes podem ser localizados de maneira estática no espaço, seja pela localização dos estabelecimentos, seja pela própria residência dos clientes (TAKETA, 1993).

Atualmente, com o surgimento das geotecnologias, o Geomarketing adquiriu uma nova concepção. O advento de tecnologias de georreferenciamento, tais como o GPS, os satélites, e os softwares de geoprocessamento com modelos de mercado que geram dados georreferenciados, promoveram a integração entre a geografia e o marketing, fazendo surgir o Geomarketing tal como o conhecemos hoje.

Nesse contexto, os Sistemas de Informações Geográficas (GIS, do inglês *Geographical Information Systems*) utilizados para a prática do Geomarketing, disponibilizam ferramentas que permitem a localização desses agentes no espaço a partir da utilização de diversas bases de dados, fornecendo assim informações de extrema relevância e utilidade para a tomada de decisão.

2.2.1 Histórico do Geomarketing

Inicialmente conhecido como geografia de mercado, o Geomarketing tem suas origens a partir da união entre a Economia e a Geografia. Já no século XIX, Von Thünen (1966) introduziu em 1826 o conceito de área de influência em sua “Teoria do Estado Isolado”, também conhecida como “Teoria da Localização Agrícola”. Essa teoria trata da relação existente entre a renda gerada por uma terra e sua localização com relação ao mercado central.

Dessa forma, segundo Thünen (1966), as terras mais próximas do centro consumidor geram uma renda maior em comparação àquelas mais distantes. Assim, os produtos se distribuíam regularmente em torno do mercado, formando o que ficou conhecido como os “Anéis de Thünen”, conforme ilustra a Figura 2:

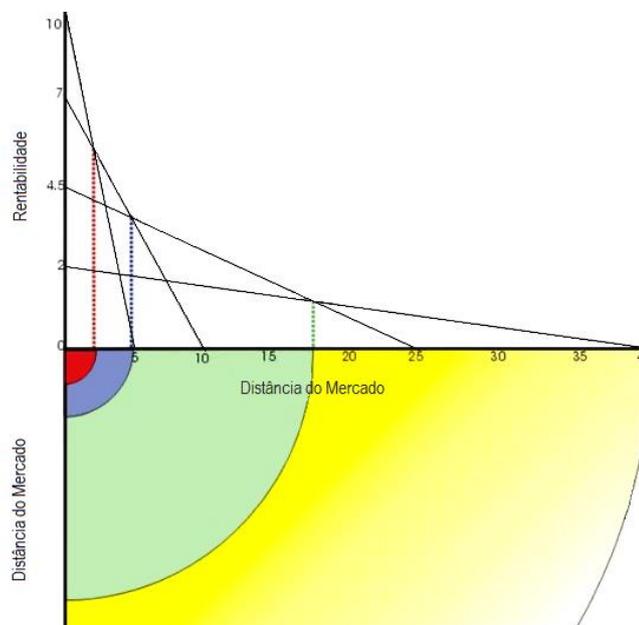


Figura 2 – Anéis de Thünen

Fonte: Adaptado de Center for Spatially Integrated Social Science

Parte dessa variação na rentabilidade pode ser justificada com base na ideia de que, quanto mais afastado o produto estiver com relação ao mercado ou centro consumidor, maiores serão os gastos com transporte. Assim, Thünen (1966) definiu que a distância até o mercado determinava não apenas o excedente do produtor, como também a seleção de culturas.

Nesse contexto, segundo essa teoria, a região em vermelho seria o centro consumidor, composto pelas residências e pelo mercado urbano. Já a região em azul, por estar mais próxima ao centro consumidor, seria ideal para o cultivo de produtos mais perecíveis, como por exemplo horticultura e laticínios. A região em verde serviria para o cultivo de grãos em geral, enquanto que a região em amarelo seria ideal para pastagem.

É importante ressaltar os princípios adotados por Thünen (1966) no desenvolvimento de sua teoria, contextualizando os critérios por ele adotados. Tais princípios podem ser sintetizados, conforme listagem a seguir:

- I. o mercado central está localizado dentro do referido “Estado Isolado”, sugerindo uma comunidade autossuficiente e sem influências externas;
- II. o “Estado Isolado” está cercado por terras desocupadas e sem uso e todos dispõem da mesma tecnologia e recursos (mesmos custos de produção);
- III. as condições naturais são uniformes em todas as terras do Estado, ou seja, não há rios, montanhas ou outros obstáculos. O solo, o clima e todos os outros fatores na agricultura são os mesmos;
- IV. só há um tipo de transporte: o terrestre. Os agricultores do Estado devem transportar seus próprios produtos diretamente até o mercado central;
- V. os agricultores do Estado fazem tudo o que for preciso para ganhar o maior lucro no mercado (concorrência perfeita).

No contexto atual de Marketing, esses princípios se mostram parcialmente presentes, ainda que em aspectos diferentes. A ideia de inexistência de influências externas, trazida no primeiro princípio, evidencia um fato improvável, dado o cenário do mercado atual. A globalização trouxe uma integração na qual a interdependência entre as empresas representa um fator de sobrevivência e adaptabilidade no mercado.

No *outsourcing*, por exemplo, parte dos produtos e serviços que são utilizados por uma empresa em sua cadeia produtiva é adquirida a partir de uma empresa externa, de forma que a empresa-cliente possa se concentrar em sua competência principal, definida como *core competence* (ARNOLD, 2000). Práticas como essa estão se tornando cada vez mais presentes na realidade das empresas e representam um relacionamento de interdependência mútua e de uma estreita colaboração, evidenciando que, atualmente, o fator “influências externas” adquiriu outro valor para as empresas.

Assim, o segundo princípio se mostra particularmente inaplicável no contexto de mercado atual. As empresas, mesmo que de mesmo porte e ramo de atuação, possuem diversificadas fontes de informações e dispõem da utilização de ferramentas diferenciadas, possibilitando assim que algumas empresas se destaquem com relação às outras no mercado.

Neste cenário, o terceiro princípio aplicar-se-ia não à uniformidade das condições naturais da terra, mas sim, de forma equivalente, à igualdade de condições

necessárias à permanência no mercado. Impostos governamentais, sazonalidades na produção, bem como variações de necessidades dos clientes ou consumidores são, por exemplo, algumas das circunstâncias às quais, em geral, as empresas estão igualmente sujeitas.

Considerando-se as alterações trazidas pelos avanços tecnológicos, o quarto princípio se mostra ainda verdadeiro, no sentido de que as empresas dispõem das mesmas opções de transporte. Assim, os fatores que definirão o tipo de transporte a ser utilizado serão, dentre outros, as restrições orçamentárias da empresa, as características do produto a ser transportado, bem como o prazo a ser cumprido.

Por fim, tem-se que o quinto e último princípio apresentado ainda hoje se faz presente no contexto de mercado atual. As empresas, ainda que não estejam inseridas no formato de concorrência perfeita, buscam os meios necessários para que possam obter o maior lucro no mercado.

Todos esses princípios refletem a forma atual com que as empresas buscam se inserir no mercado. No caso de restaurantes japoneses, bem como no de outros ramos de negócio, os gestores procuram instalar seus estabelecimentos de forma a satisfazer essas condições, buscando sempre atingir o melhor desempenho dentro de seu mercado. Estar localizado próximo a seu fornecedor, por exemplo, pode ser uma boa estratégia de minimização de custo de transporte e desperdício, tendo em vista principalmente a alta perecibilidade dos produtos em questão.

Além de Thüner, outros grandes autores desenvolveram, posteriormente, teorias que se remetem de alguma forma à essência teórica do Geomarketing.

Diferentemente de Von Thünen, que procura responder quais atividades deverão se localizar em uma determinada região, Alfred Weber (1957) busca predizer em sua “Teoria da Localização Industrial” onde se localiza uma dada atividade industrial.

Segundo essa teoria, uma indústria está localizada onde os custos de transporte com matérias-primas e produto final são mínimos. Essa teoria envolve uma série de simplificações, especificamente:

- I. a localização é definida em uma região isolada, sem influências externas e composta de apenas um mercado;
- II. o espaço é isotrópico, ou seja, sem variações nos custos de transporte;
- III. os mercados estão localizados em um número específico de centros urbanos e;
- IV. o mercado possui concorrência perfeita, ou seja, existe um alto número de empresas e de clientes, onde ambos detêm perfeito conhecimento das condições do mercado.

Weber (1957) define então, três fatores que influenciam na localização de uma indústria: (1) custos de transporte, (2) custos de mão-de-obra e (3) economias de aglomeração. O primeiro fator é considerado o elemento mais importante desse modelo, tendo em vista que os outros fatores são considerados apenas para efeitos de ajuste. Deve-se, portanto, encontrar a localização que forneça o menor custo de transporte para, então, ajustar essa localização de forma a considerar os custos de mão-de-obra e economias de aglomeração.

Rodrigue (2013) ilustra a questão de minimização dos custos de transporte a partir da “Localização Triangular”, utilizada por Weber (1957) com o intuito de determinar a localização industrial ótima, conforme apresentado na Figura 3.

Nessa figura, considera-se um produto de $w(M)$ toneladas a ser vendido no mercado M , onde são necessárias $w(S1)$ e $w(S2)$ toneladas de material provenientes de $S1$ e $S2$, respectivamente. O problema consiste em encontrar a localização industrial ótima P situada nas respectivas distâncias $d(M)$, $d(S1)$ e $d(S2)$.

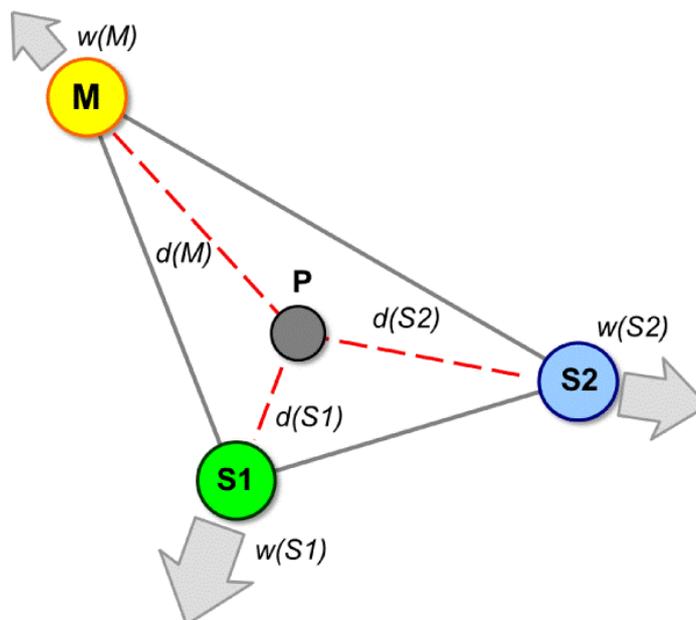


Figura 3 – Triângulo Locacional de Weber
 Fonte: Rodrigue (2013)

Weber (1957) ressalta também alguns fatores que podem influenciar os custos de transporte, tais como o tipo e a extensão do sistema de transporte utilizado, as condições naturais da região e os tipos de estrada, a própria natureza dos bens, ou seja, as qualidades que, além do peso, determinam a facilidade de transporte.

Por fim, outro autor que merece destaque no que diz respeito à sua contribuição para o surgimento das ideias essenciais que compõe o Geomarketing é Christaller (1966) e sua “Teoria do Lugar Central”. Essa teoria é uma tentativa de explicar e identificar padrões no arranjo espacial, tamanho e número de centros comerciais em um sistema urbano. Para o desenvolvimento de sua teoria, Christaller (1966) fez algumas considerações:

Consumidores: assume-se que todos os consumidores possuem igual influência e tamanho, de forma que apenas um único consumidor sozinho não seja suficiente para causar grandes impactos no mercado. Além disso, todos os consumidores são racionais, possuem o mesmo nível de informação e buscam sempre a maximização de sua utilidade;

Fornecedores: assim como os consumidores, assume-se que os fornecedores são igualmente racionais, afluentes, possuem o mesmo nível de informação e buscam sempre a maximização de seu lucro.

Adicionalmente, considera-se também que todos os fornecedores possuem os mesmos custos de produção e que não existem barreiras de entrada no mercado;

Geografia: assume-se que todo o espaço seja homogêneo, onde os custos de transporte não variem, independentemente da direção escolhida. Além disso, pressupõe-se a uniforme distribuição no mercado de todos os consumidores e fornecedores. Já no que diz respeito ao comportamento geográfico dos consumidores, considera-se que todos eles adquiram um único produto no local mais próximo.

Tendo como bases as condições simplificadoras citadas, é possível fazer algumas conclusões com relação à “Teoria do Lugar Central” de Christaller (1966). Os constantes aumentos nos custos de transporte fazem com que a demanda por certo tipo de produto reduza à medida que a distância entre o mercado consumidor e o fornecedor aumente.

A partir de determinada distância, a demanda por este produto será igual a zero. Essa distância é conhecida como “área de influência” ou “amplitude” do bem, ou seja, é a distância máxima que o consumidor se dispõe a percorrer para obter determinado produto. Cria-se então uma hierarquia de localizações, conforme ilustra a Figura 4.

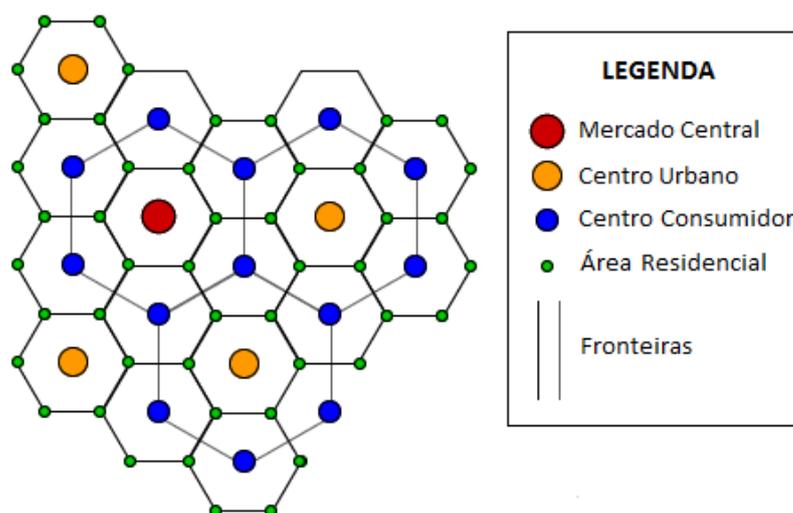


Figura 4 – Teoria do Lugar Central
Fonte: Adaptado de The Wolf at the Door

Na Figura 4 apresentada, as fronteiras demarcam o limite das áreas de influência. Nas áreas residenciais (sub-centros), onde a maioria das pessoas vive, existem apenas serviços básicos, tais como padaria, lojas de maquinários e serviços de carpintaria. Nos centros consumidores (pequenos centros), que são cercados de áreas residenciais, estariam concentrados outros serviços tais como médicos, dentistas, veterinários, fabricantes de roupas, vidraceiros, comerciantes especializados e assim por diante. Já os centros urbanos (centros intermediários) mantêm as fábricas, indústrias, como mineração de carvão e siderurgia, universidades e centros de lazer. Por fim, no topo da hierarquia, está o mercado central (grande centro) que oferece a maior variedade de bens possível.

A abordagem desses autores com suas teorias econômico-geográficas representa uma contribuição de fundamental importância para o surgimento e desenvolvimento do que hoje é conhecido como Geomarketing. Ainda hoje, o preço de áreas urbanas e os custos de transporte estão inversamente relacionados à distância até o centro.

Assim, pode-se observar que com o desenvolvimento e integração dos mercados, o processo de escolha da localização geográfica foi adquirindo cada vez mais importância com o passar do tempo. Feita a análise das origens do Geomarketing, passemos agora para sua definição.

2.2.2 Conceituação do Geomarketing

Diferentemente do conceito apresentado anteriormente no histórico do Geomarketing, que define seu surgimento a partir da união entre a economia e a geografia, Yrigoyen (2003) defende que a geografia introduziu no campo do marketing a dimensão espacial dos fenômenos socioeconômicos do mercado em que analisa. Por isso, segundo a mesma autora, o termo Geomarketing nasce da confluência entre geografia e marketing.

Tal definição não invalida as ideias anteriormente apresentadas, tendo em vista que, segundo Chauvel (2002), os primeiros estudos de marketing surgiram somente no pós-guerra, por volta dos anos 40, ou seja, quase 100 anos depois dos primeiros estudos relacionando localização e rentabilidade.

Entretanto, a introdução da dimensão espacial às decisões de marketing envolve, segundo Cliquet (2006), pelo menos três domínios do marketing: (1) comportamento do consumidor, (2) local de varejo e (3) gestão de marketing. O primeiro domínio está mais relacionado com o comportamento espacial ou geográfico do consumidor o que, segundo o mesmo autor, advém essencialmente da geografia e da economia.

Levando em consideração esses domínios, Cliquet (2006) define Geomarketing como sendo um “conjunto de técnicas que permite a manipulação de dados geocodificados”. Porém, Cliquet (2006) ressalta que essas técnicas somente adquirem o entendimento de marketing espacial quando as informações geográficas obtidas são aplicadas em áreas do marketing suscetíveis aos efeitos desses tipos de dados geocodificados, tornando-se assim uma ferramenta útil para desenvolvimento de estratégias e tomada de decisão.

No contexto do presente trabalho, esses domínios se mostram presentes em diferentes fases. O domínio de “comportamento do consumidor” seria uma análise prévia que orientaria o segundo e terceiro domínios: “local de varejo” e “gestão de marketing”. O primeiro domínio contempla não apenas as características comportamentais do consumidor, mas principalmente suas características sociodemográficas, como gênero, idade, renda, etc.

Com base nas informações obtidas a partir do primeiro domínio, nortear-se-á o desenvolvimento de estratégias de marketing específicas para o público identificado (gestão de marketing). Definir-se-á também a melhor localização para a instalação do estabelecimento, tendo em vista as características observadas no primeiro domínio (local de varejo).

Dessa forma, os estudos de marketing espacial buscam compreender de que maneira a análise geoespacial pode ser utilizada para melhorar os aspectos operacionais e estratégicos do marketing (SMITH, 2007). Nesse contexto, o Geomarketing surge com intuito principal de fornecer aos gestores ferramentas que permitam compreender como funciona a organização dos mercados em termos espaciais.

Partindo desse princípio, Yrigoyen (2003) define Geomarketing como sendo um conjunto de técnicas que permite analisar a realidade socioeconômica sob um ponto de vista geográfico, através de instrumentos cartográficos e ferramentas de estatística espacial.

Outro conceito trazido por Latour e Le Floch (2001) define o termo Geomarketing como um sistema integrado por programas de processamento de dados, métodos estatísticos e representações gráficas destinadas a produzir informações úteis para a tomada de decisões, através de instrumentos que combinam cartografia digital, gráficos e tabelas.

Analisando o conceito sob um ponto de vista mais sociológico, Alcaide et al. (2012) argumentam que Geomarketing é a área do Marketing que tem como objetivo o conhecimento global do cliente, bem como as necessidades e comportamentos do cliente dentro de uma determinada área geográfica. Toda essa informação ajuda a empresa a ter uma visão mais completa de seus clientes e a identificar suas necessidades.

Por fim, Baviera-Puig et al. (2013) sugerem que o Geomarketing pode ser definido como a disciplina que se utiliza de Sistemas de Informação Geográfica como uma ferramenta para análise e tomada de decisão em marketing, visando conciliar as necessidades e desejos do cliente de uma maneira lucrativa para a empresa.

Dessa forma, pode-se dizer que o Geomarketing envolve basicamente os seguintes elementos: base de dados, informações cartográficas e Sistemas de Informação Geográfica, ou simplesmente SIG, que é responsável por relacionar os dois primeiros elementos.

Tomando como base todos os aspectos suscitados nas definições apresentadas, considerar-se-á neste trabalho, para efeitos de conceituação, a seguinte definição de Geomarketing: *é um conjunto de técnicas que permite a visualização e análise de informações de mercado relevantes sob uma ótica geográfica, através da utilização de ferramentas específicas para o processamento de dados geocodificados, auxiliando os gestores na elaboração de estratégias, na leitura espacial de mercado e, conseqüentemente, na tomada de decisões.*

É importante ressaltar como a introdução dessa dimensão espacial na análise social e econômica de um mercado permite a uma empresa um melhor conhecimento do mercado em que atua, além de prover informações de extrema relevância para o Marketing, tais como: quem compra, quando compra, com que frequência se dá esse consumo e, o foco maior deste trabalho, onde compra.

A visualização dessas informações sob a forma de mapas geográficos apresenta uma vantagem considerável no que diz respeito à interpretação das informações obtidas. A análise da distribuição e da relação entre variáveis se torna mais fácil devido à possibilidade de se trabalhar com várias “camadas de informação” simultaneamente, além de permitir a agregação espacial de dados que geralmente são dispostos em tabelas, onde as relações entre essas variáveis e o espaço no qual estão inseridas nem sempre são percebidas. A Figura 5 ilustra bem essa ideia:

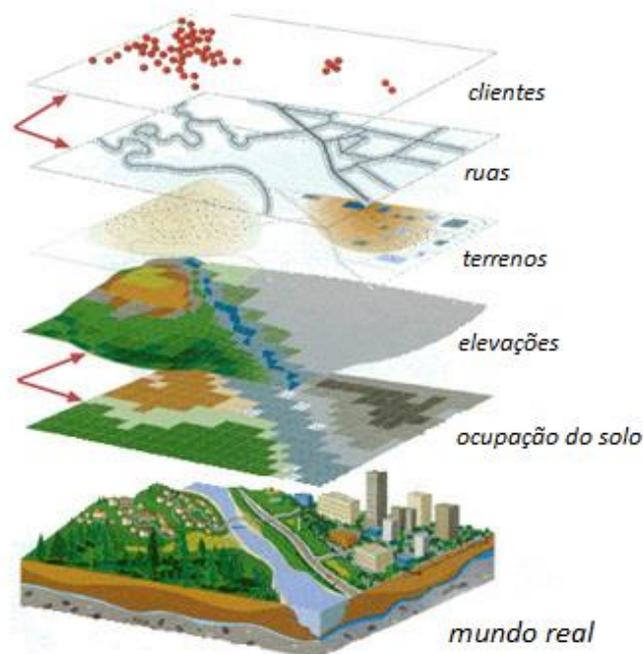


Figura 5 – Sistemas de Informação Geográfica
Fonte: Adaptado de The University of Chicago

O desenvolvimento de mapas que forneçam a visualização sobreposta dessas “camadas de informação” permite que empresa obtenha uma melhor compreensão sobre a realidade do mercado. Assim, é possível verificar a influência que uma determinada localização exerce tanto sobre seus clientes, quanto sobre seus concorrentes de mercado.

Nesse contexto, o Geomarketing é capaz de prover um melhor desempenho operacional por meio da adaptação do marketing-mix da empresa de acordo com segmento de mercado a ser atingido, permitindo assim a identificação dos locais que possuam maior potencial de consumo para determinado produto ou serviço, conforme veremos mais adiante.

2.3 Geomarketing e Segmentação de Mercado

Desde o surgimento das primeiras organizações as empresas sentem a necessidade de separar seus mercados por áreas distintas, reconhecendo as características específicas de cada região. Ortigoza (1997) ressalta como o avanço no processo de globalização reforçou ainda mais a importância da escolha do local ideal para a abertura de um negócio.

Nota-se, portanto, que a prática do Geomarketing está muitas vezes implícita na estratégia da organização, embora nem sempre de uma maneira formal e sistematizada. Atualmente, muitas empresas ainda utilizam como base apenas as fronteiras físicas delimitadas entre regiões.

Entretanto, esse tipo de segmentação de mercado baseada apenas em definições geográficas pré-determinadas, sem a devida fundamentação metodológica, pode provocar uma leitura errônea do mercado. O reconhecimento dessa falha mostra a necessidade de uma abordagem mais precisa no processo de segmentação de mercado.

Segundo Richers e Lima (1991), além do fator geográfico, existem diversos outros aspectos que influenciam a segmentação de mercado, tais como fatores demográficos, socioeconômicos, padrões de consumo, benefícios procurados, estilos de vida, personalidade e caracterização econômica do setor, conforme apresentado no Quadro 2:

FATORES	CRITÉRIOS
Geográfico	Extensão do mercado potencial, concentração geográfica, transporte e acesso, polarização, bairros e ruas, tráfego, centros de compras.
Demográfico	Idade, gênero, domicílio, família, ciclo de vida (jovem, adulto, idoso).
Socioeconômico	Classe de renda, instrução, ocupação, status, migração, mobilidade social.

Padrões de Consumo	Frequência de compra, local de compra, lealdade a marcas, Heavy & Light Users, curva ABC.
Benefícios Procurados	Satisfação sensual, prestígio social, emulação-preço favorável, qualidade/durabilidade, redução de custos, atendimento/serviços.
Estilos de Vida	Expectativas de vida, uso do tempo, interesses predominantes, participação em eventos e agrupamentos sociais, uso do dinheiro, amizades e relações pessoais.
Personalidade	Bases culturais, atitudes e valores, liderança, agentes de mudanças.
Caracterização Econômica	Setor de atividade, tamanho das empresas, atuação dos concorrentes, acessibilidade, usos e aplicações, unidade de decisão de compra, demanda derivada.

Quadro 2 – Fatores de Segmentação de Mercado

Fonte: Richers e Lima (1991).

Dentre os fatores apresentados no Quadro 2, os principais fatores a serem considerados no processo de segmentação de mercado deste trabalho serão os fatores geográfico, demográfico, socioeconômico e caracterização econômica do setor. Assim, realizou-se por meio da utilização do Geomarketing, a identificação e delimitação geográfica dos segmentos de mercado, de acordo com critérios previamente estabelecidos.

De acordo com Douard (2006), os objetivos do Geomarketing podem ser sintetizados em quatro pontos fundamentais: (1) gerenciamento de informações múltiplas; (2) domínio mercadológico de territórios; (3) adaptação aos clientes e; (4) predição de evoluções.

Já Yrigoyen (2003) define que a missão do Geomarketing consiste em abordar os quatro elementos do marketing-mix (produto, comunicação, distribuição e preço) sob a perspectiva espacial subjacente a todos eles. De certa forma, a prática do Geomarketing envolve, direta ou indiretamente, a consideração desses fatores no processo de segmentação de mercado.

Isso se deve ao fato de que a dimensão espacial é evidente em todo esse processo. O **produto**, seja um bem ou serviço, deve ser **comunicado** ao público-alvo através de meios promocionais, como por exemplo um *outdoor*, bem como **distribuído** em um meio de transporte para sua comercialização em um ponto de venda, com um **preço** muitas vezes condicionado a questões locais (um bairro, um centro comercial, uma cidade, etc), conforme ilustra a Figura 6.



Figura 6 – Marketing-Mix
 Fonte: Elaborado pela autora.

Dessa forma, a segmentação de mercado descreve a divisão de um mercado heterogêneo em grupos homogêneos, os quais responderão de forma diferente aos estímulos do mercado, tais como promoções, comunicações, propagandas, preço entre outras variáveis do marketing-mix.

Cada grupo, ou segmento, está sujeito aos efeitos de um marketing-mix diferente, tendo em vista que os segmentos são criados justamente para minimizar as diferenças inerentes aos indivíduos dentro de cada segmento, bem como maximizar as diferenças entre um segmento e outro. Uma vez identificado o segmento de mercado em questão, qualquer solução personalizada pode ser, conseqüentemente, implementada e testada de forma a prover a uma melhor utilização do serviço proposto.

De fato, o Geomarketing tem progressivamente se tornado o novo elemento do marketing-mix, permitindo a melhora do conhecimento e compreensão dos mercados (DOUARD, 2006). Sobre Geomarketing, o mesmo autor afirma ainda que:

“O Geomarketing fornece as chaves para decodificar o significado que o território contém. De agora em diante, no marketing-mix, as

empresas devem progressivamente integrar esse, ainda pequeno, componente geográfico. Uma melhor compreensão das conexões entre os clientes e o espaço fornece a possibilidade de abordagens de marketing inovadoras, localmente adaptadas e provendo, portanto, vantagem competitiva". (DOUARD, 2006, p. 100)

Nesse contexto, é notável a contribuição do Geomarketing para um melhor aproveitamento das informações de mercado obtidas. A incorporação de informações geográficas ao marketing-mix permite a obtenção de informações mais ricas do ponto de vista espacial, possibilitando a formulação de estratégias de ação mais precisas.

Do ponto de vista do gestor, essas informações georreferenciadas permite o desenvolvimento de estratégias de marketing, produtos ou serviços específicos, orientados às necessidades dos consumidores. A longo prazo, essas informações podem trazer grandes economias, na medida em que canaliza a utilização dos insumos disponíveis, evitando assim os desperdícios proveniente de uma má alocação de recursos.

Entretanto, Aranha e Figoli (2001) defendem que, devido à evolução da informática e ao barateamento dos bancos de dados, a segmentação tradicional está se tornando obsoleta. Atualmente, o conhecimento das características e comportamento de compra de cada indivíduo está mais viável do ponto de vista econômico, permitindo assim uma segmentação de mercado mais detalhada do que era feito anteriormente.

Assim, os Sistemas Geográficos de Informação (SIG) estão, segundo Aranha e Figoli (2001), possibilitando uma nova abordagem do mercado, substituindo as amplas estratégias de segmentação de mercado atuais. Essa nova abordagem geográfica de mercado é definida pelos mesmos autores como "*micromarketing*" ou "*marketing de vizinhança*".

A justificativa dessa nomenclatura se deve ao fato de que as pessoas que residem no mesmo local (ou vizinhança) tendem a ser, de alguma forma, parecidas, seja sobre aspectos culturais, econômicos ou históricos. Dessa forma, Aranha e Figoli (2001) explicam que o micromarketing é capaz de tornar esses dados em informações mais precisas, permitindo, por meio da utilização de ferramentas estatísticas e bancos de dados, a fragmentação do público em grupos menores e uniformes, também chamados de *clusters*.

Partindo-se do princípio de que a localização da residência de um consumidor possui parcial responsabilidade sobre seu comportamento de consumo, Cardoso (2011) afirma que o perfil do consumidor ideal pode ser obtido a partir da sobreposição das vendas de determinada área às informações sobre os consumidores que residem nessa região.

Dessa forma, torna-se possível a adaptação da oferta local de acordo com o perfil identificado de seus consumidores, permitindo assim a visualização do desempenho de determinados ramos de negócio que atuam nessa região.

Outro agente do marketing-mix positivamente afetado pelas leituras de mercado que o Geomarketing oferece é o fornecedor, que poderá analisar o melhor local para seu ponto de distribuição, baseando-se no perfil do público local e no número de vendas de determinados produtos, conforme enfatiza Cardoso (2011).

Analisando sob o ponto de vista da concorrência, o Geomarketing fornece meios para a identificação do posicionamento dos concorrentes no mapa, fazendo com que a empresa tenha uma visão mais ampla do mercado local, evitando assim, por exemplo, a fixação em locais já saturados por outras empresas do mesmo ramo.

A grande vantagem advinda da utilização dessa abordagem de micro segmentação consiste na identificação de padrões de consumo específicos que, quando conhecidos, podem orientar o emprego dos recursos disponíveis, potencializando os resultados estratégicos da empresa. Reforçando esta ideia, Cardoso (2011) afirma que:

“Uma das bases estruturais para o desenvolvimento do Geomarketing consiste na procura de um conhecimento e compreensão precisos sobre os mercados, que se manifestam através de segmentações mais precisas baseadas na análise comportamental. Esta mesma segmentação de mercado, é baseada na cada vez maior fragmentação da população e na necessidade de definir a diferenciação estratégica para as empresas num contexto cada vez mais competitivo”. (CARDOSO, 2011, p. 28)

Dessa forma, o cruzamento entre as informações sobre o consumidor (tais como idade, renda, gênero, despesas) e os padrões de consumo identificados, tornará possível a segmentação da base de dados, permitindo elaborar estratégias

específicas e, conseqüentemente, obter resultados mais efetivos para a compreensão do fenômeno de consumo dos consumidores.

Embora a relevância dessa compreensão seja notável, Cardoso (2011) ressalta que existem diferentes formas para se avaliar o potencial econômico de um determinado território. O estilo de vida adotada pelo consumidor, a mobilidade e as características geográficas da região, são variáveis que vêm se tornando cada vez mais presentes como dimensões do Marketing e por isso devem ser levadas em consideração no momento de definir a localização comercial de um negócio.

O crescimento da mobilidade entre os consumidores é um fato inegável, seja devido à maior viabilidade econômica de acesso aos mais variados tipos de transporte, seja pela diminuição das barreiras políticas, culturais ou sociais que antes segregavam os espaços. Atualmente, segundo Cliquet (2006), as pessoas percorrem uma distância 10 vezes maior do que era percorrido a cerca de 30 anos atrás. Esse crescimento na mobilidade dos consumidores revela uma nova realidade que deve ser considerada pelas empresas ao segmentarem um mercado.

Nesse contexto, Douard (2006) ressalta que o desenvolvimento de bases de dados de comportamento, oferecendo múltiplos dados sobre o comportamento de consumo, permite, graças ao Geomarketing, a realização de uma nova abordagem de clientes. Esse fenômeno é duplamente facilitado por dois fatores: (1) desenvolvimento de *softwares* de cartografia mais completos e de mais fácil utilização que se tornaram possíveis devido ao (2) desenvolvimento de tecnologias baseadas em Sistemas de Informação Geográfica (SIG).

Tal fato faz com que, segundo o mesmo autor, as empresas passem a integrar cada vez mais as ferramentas do Geomarketing a seus departamentos de pesquisa. Assim, dentre as mais frequentes aplicações do Geomarketing, Douard (2006) frisa as seguintes:

- I. adaptar a oferta comercial do estabelecimento, otimizando seu desempenho;
- II. trabalhar em micro mercados;
- III. reestruturar a rede;
- IV. determinar uma nova localização;

- V. fixar objetivos de forma relacional ao potencial;
- VI. identificar os clientes ideais por meio de campanhas específicas;
- VII. otimizar investimentos em marketing direto;
- VIII. assistir às vendas;
- IX. analisar a competitividade.

Baviera-Puig et al. (2013) trazem outra aplicação do Geomarketing a partir da utilização de Sistemas de Informação Geográfica. Para definir as áreas de consumo em potencial, os autores se utilizaram de dois conceitos. O primeiro diz respeito à geodemanda (*geodemand*), definida como sendo a localização em mapas digitais dos consumidores de um produto ou serviço em um mercado específico.

Já o segundo conceito trata sobre à geocompetição (*geocompetition*), definida como a localização em mapas digitais dos concorrentes de mercado, bem como o delineamento de suas áreas de alcance (*trade area*) em um mercado específico.

Dessa forma, a partir da intersecção desses dois mapas, o autor define as áreas de consumo em potencial como sendo aquelas áreas onde a geodemanda não se encontra com a geocompetição, ou seja, são as áreas onde existem simultaneamente um alto índice de demanda e um baixo índice de oferta.

É notável, portanto, que todas as aplicações e utilizações do Geomarketing até aqui apresentadas têm em comum a busca das respostas às mesmas questões: onde estão os clientes, onde vivem, quais são suas características, quando e com que frequência compram, onde compram, onde estão meus concorrentes, onde posso encontrar novos mercado, dentre outros questionamentos pertinentes à dimensão espacial do consumo.

Assim, torna-se visível a importância do papel do Geomarketing no que diz respeito ao refinamento do marketing-mix de forma direcionada para os diferentes segmentos de mercado geograficamente delimitados.

Essa delimitação geográfica em mapas, cruzada às informações de comportamento e compra dos consumidores gera, por sua vez, segmentos significativamente mais específicos do que a simples segmentação feita com base em meras divisões políticas ou geográficas.

Constata-se, portanto, a expressiva importância do Geomarketing como uma ferramenta de auxílio tanto na definição de produto, preço e promoções a serem aplicadas, quanto na percepção da localização geográfica ideal. Tais informações servirão de base para definir onde determinado produto possuirá maior aceitação do mercado, constituindo um fator de enorme vantagem competitiva para a empresa que o adotar.

2.4 Máquinas de Suporte Vetorial

A implementação do Geomarketing implica, conforme visto anteriormente, na seleção e processamento de dados sobre o consumidor, fornecendo assim uma segmentação de mercado mais precisa. Uma ferramenta que pode ser útil nesse processo de classificação de dados são as Máquinas de Suporte Vetorial, ou simplesmente SVM, do inglês *Support Vector Machines*.

As Máquinas de Suporte Vetorial, cujo objetivo principal consiste na separação de dados pertencentes a classes distintas, constituem uma técnica de aprendizado de máquina que tem como principal fundamentação teórica a “Teoria do Aprendizado Estatístico” desenvolvida por Vladimir Vapnik (1995).

Para efeitos de delimitação operacional, conceituou-se de forma sistemática cada um dos termos anteriormente apresentados, fornecendo assim um suporte teórico para um melhor entendimento e compreensão sobre a aplicação dessas técnicas no âmbito do Geomarketing.

2.4.1 Aprendizado de Máquina

Segundo Mitchell (1997), uma perspectiva de grande utilidade em aprendizado de máquina envolve a procura de um amplo espaço de possíveis hipóteses, a fim de determinar aquela que melhor se adapta aos dados observados ou a qualquer outro critério tido como prioritário pelo usuário.

As técnicas de aprendizado de máquina utilizam o **princípio da inferência indutiva**, definido por Angluin e Smith (1983) como sendo o processo de hipotetizar uma regra geral para um conjunto específico de exemplos ou dados. Assim, o

aprendizado em inferência indutiva é dividido em dois paradigmas: o aprendizado supervisionado e o não supervisionado.

O **supervisionado** envolve a participação de um *professor* no processo de aprendizado. Segundo Haykin (1999) esse *professor* fornecerá o conhecimento sobre o ambiente, por meio de um conjunto de entrada e saída de exemplos (dados) representados nesse ambiente.

A ideia é que o Aprendizado de Máquina (AM) crie, utilizando esses exemplos, uma representação que produza saídas corretas para cada nova entrada de dados que não estavam incluídos nos exemplos fornecidos pelo *professor*, conforme ilustra a Figura 7:



Figura 7 – Aprendizado de Máquina Supervisionado
Fonte: Elaborado pela autora.

Já no aprendizado **não supervisionado** ou auto-organizado não existe a figura externa de um *professor* para intervir no processo de aprendizado. A previsão é feita de forma independente, de maneira que o AM aprenda a identificar e classificar uma nova entrada de dados, com base nos parâmetros utilizados na representação obtida (HAYKIN, 1999), conforme ilustra a Figura 8.

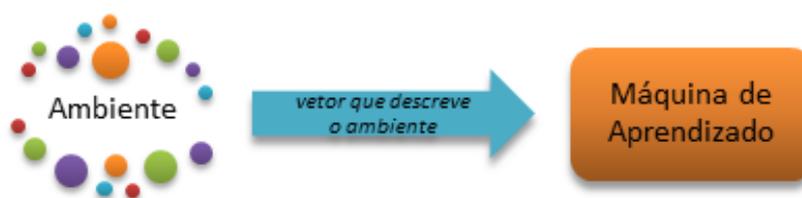


Figura 8 – Aprendizado de Máquina Não Supervisionado

Fonte: Elaborado pela autora.

Uma vez que a rede tenha se adaptado às regularidades estatísticas da entrada de dados, ela desenvolve a habilidade de formar representações internas para codificar as características da entrada de dados, criando assim novas classes de forma automática (BECKER, 1991 *apud*. HAYKIN, 1999).

Essas técnicas, segundo Lorena e Carvalho (2007) são mais utilizadas quando o objetivo principal for a identificação de padrões ou tendências que possam orientar na compreensão dos dados.

Neste trabalho, considerou-se a aprendizagem **supervisionada**, ou seja, há um *professor* como agente conhecedor do ambiente. Dessa forma, suponha um conjunto de dados onde x_i represente um vetor dado e y_i represente a classe a qual esse dado pertence, usualmente codificadas como +1 e -1.

Deve-se, portanto, produzir um modelo classificador (ou hipótese) que seja capaz de prever a classes a qual os novos dados pertencerão. Esse processo de estimar uma classificação com base em um conjunto específico de dados é conhecido como **treinamento** (KE-LIN e SWAMY, 2014).

Além do conjunto de dados de treinamento, que fará com que a máquina aprenda o algoritmo, existem outros dois conjuntos. O conjunto de **validação** será utilizado tanto para a verificação da capacidade de generalização do algoritmo quanto para o ajuste dos parâmetros. Após o treinamento e os devidos ajustes do algoritmo, o conjunto de **teste** é aplicado com o intuito de avaliar o desempenho de generalização da máquina de aprendizado.

Nesse contexto, segundo Baranauskas et al. (2001), as classes desses dados podem ser consideradas como a descrição propriamente dita do fenômeno sobre o qual se tem interesse. Em outras palavras, as classes representam o evento de

interesse sobre o qual se deseja que a máquina aprenda (treine) para que seja capaz de realizar previsões.

No contexto deste trabalho, o fenômeno de interesse seria justamente a propensão ao consumo de comida japonesa. Assim, as classes representadas seriam (1) indivíduo propenso ao consumo de comida japonesa e (2) indivíduo não propenso ao consumo de comida japonesa. Com o aprendizado, a máquina deverá ser capaz de prever, para cada entrada de um novo dado, a qual dessas duas classes esse novo dado (indivíduo) pertencerá.

Dessa forma, segundo Mitchell (1997), as classes podem assumir dois tipos de valores: discretos (1, 2, 3, ... , n) ou contínuos. Quando as classes apresentam valores discretos, trata-se de um problema de **classificação**. Entretanto, se elas apresentarem valores contínuos, o problema em questão será denominado como sendo de **regressão**.

Com relação aos problemas de classificação, eles podem ser caracterizados de duas formas: se as classes assumirem apenas dois valores, ou seja, $n = 2$, ele é denominado como um problema de classificação **binário**. Se ele possuir mais de duas classes, ou seja, $n > 2$, ele será considerado como um problema de classificação **multiclasse** (VAPNIK, 1998).

Cada um desses exemplos, também denominados de casos, registros ou dados, possui determinadas características ou aspectos que o representam dentro do conjunto. Baranauskas e Monard (2003) denominam essas características descritivas como um **atributo** do dado.

Segundo os mesmos autores, existem ainda dois tipos de atributos: o **nominal** e o **contínuo**. O primeiro tipo é caracterizado pela não existência de ordem entre os valores obtidos, como por exemplo, gênero feminino e masculino. Já o segundo tipo caracteriza-se justamente pela existência de ordem entre os valores apresentados, como por exemplo, idade $\in \mathbb{R}$. A Figura 9 sintetiza os conceitos até então apresentados:

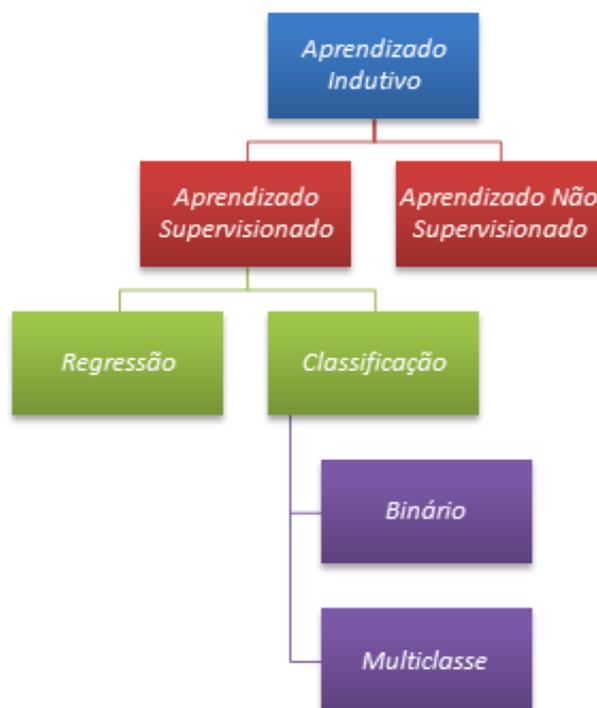


Figura 9 – A hierarquia do Aprendizado
 Fonte: Adaptado de Baranauskas e Monard (2003)

Dessa forma, dispõe-se de um embasamento teórico suficiente para realizar a classificação do problema de pesquisa do presente trabalho, de acordo com os conceitos sobre aprendizado de máquina anteriormente apresentados.

Primeiramente, tem-se que o atual problema de pesquisa trata-se de um aprendizado **indutivo**, pois objetiva-se a identificação de padrões, bem como o fornecimento de previsões de comportamento para cada novo dado entrante, que nesse caso são os indivíduos consumidores de comida japonesa.

Posteriormente, pode-se definir o problema como sendo de aprendizado **supervisionado**, devido ao fato da existência de uma figura externa, denominada como *professor*, que fornece as representações de conhecimento sobre o ambiente estudado, com exemplos de previsões corretas.

Com relação à análise do problema de pesquisa, no que diz respeito à natureza de suas classes, pode-se defini-lo como sendo um problema de **classificação**, tendo em vista a existência de apenas duas classes de valores discretos possíveis: indivíduo propenso ao consumo de comida japonesa ou indivíduo não propenso ao consumo de comida japonesa.

Por fim, tem-se que o presente problema de pesquisa apresenta um formato de classes **binário**, por possuir, como dito anteriormente, somente duas classes possíveis: indivíduo propenso ou não propenso ao consumo de comida japonesa, ou seja, $n = 2$.

No que diz respeito aos **atributos** observados nos dados (aqui considerados como sendo os consumidores), tem-se uma variedade de características descritivas, tais como gênero, idade, renda, despesas, local de residência, gastos com alimentação fora do domicílio, dentre outros.

Assim, após as análises apresentadas, tem-se que a classificação final do problema pode ser definida como sendo de aprendizado indutivo, com um problema de classificação binário.

2.4.2 Teoria do Aprendizado Estatístico

A Teoria do Aprendizado Estatístico é a principal fundamentação das Máquinas de Suporte Vetorial. Segundo Vapnik (1998), essa teoria explora as diversas maneiras possíveis de estimar a dependência funcional de um determinado conjunto de dados.

Vapnik (1998) afirma ainda que a Teoria do Aprendizado Estatístico foi desenvolvida para pequenas amostras de dados (finitas), que não dependem de informações prévias para a resolução dos problemas, diferentemente da estatística clássica, que foi desenvolvida para amplas amostras e baseada no uso de diversos tipos de informações prévias.

Em vez disso, segundo o mesmo autor, a Teoria do Aprendizado Estatístico (TAE) considera a estrutura do conjunto de funções implementado pela máquina de aprendizado, onde as medidas de capacidade do subconjunto estão especificamente definidas (aprendizado preditivo).

Dessa forma, o problema de treinamento supervisionado consiste em encontrar um classificador f , dentro de um conjunto F de todos os possíveis classificadores que o aprendizado de máquina pode gerar, que durante o processo de aprendizado utilize o conjunto T (onde estão contidos todos os dados de treinamento) para fornecer um classificador específico.

O conjunto de dados de treinamento T é composto de n dados, onde x_i é o conjunto de dados propriamente dito e y_i é a classe a qual esses dados pertencem. Assim, o conjunto T deverá ser utilizado para gerar esse classificador específico f , de maneira que $f \in F$.

Tomando como base os conceitos apresentados, considere o conjunto de dados de treinamento T representados na Figura 10:

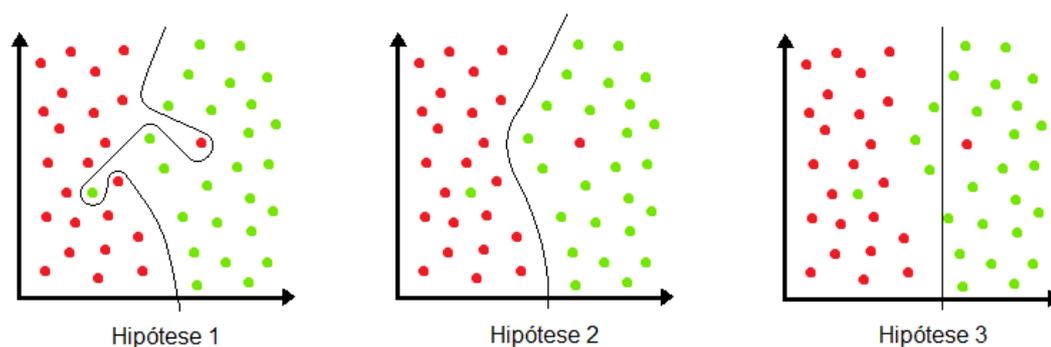


Figura 10 – Classificadores no Processo de Aprendizado

Fonte: Adaptado de Stanford University

Nesse caso, o principal objetivo do processo de aprendizado é, a partir da utilização do conjunto T que contém todos os dados da amostra, fornecer um classificador que seja capaz de separar esses dados, dividindo-os entre as classes “vermelho” e “verde”.

Na primeira hipótese, observa-se um classificador de excelente desempenho, capaz de separar corretamente todos os dados do conjunto T , inclusive os *outliers*, também conhecidos como ruídos. Entretanto, por apresentarem tamanha especificidade para o conjunto de dados em questão, esse classificador pode apresentar erros quando inseridos novos dados. Esse fenômeno é chamado de *overfitting* e representa um super ajustamento do classificador aos dados do conjunto de treinamento T (EVERITT e SKRONDAL, 2002).

Em contrapartida, a terceira hipótese evidencia um classificador que por sua vez desconsidera alguns dados que pertencem a classes divergentes, por estarem muito próximos uns dos outros. Portanto, é notável como a hipótese 3 está suscetível a muitos erros, até mesmo com os próprios dados do conjunto de treinamento. Esse

fenômeno é conhecido como *underfitting* e representa o sub-ajustamento do classificador aos dados do conjunto T .

Por fim, temos a segunda hipótese. O classificador apresentado nessa hipótese se mostra como um intermediário entre as hipóteses anteriores. Assim, esta função f apresenta um bom nível de generalização, classificando de forma correta a maioria dos dados da amostra. Este seria, portanto, o classificador adequado para o conjunto T apresentado.

Reforçando essa ideia de variações nos processos de classificação, Semolini (2002) afirma que a viabilidade do treinamento supervisionado está intimamente relacionada com o conjunto T , tendo em vista que o treinamento depende da quantidade de informação contida na amostra. Caso essa quantidade não seja suficiente, o classificador não será capaz de fornecer uma generalização confiável.

Portanto, pode-se constatar que a Teoria do Aprendizado Estatístico tem como objetivo auxiliar no processo de escolha de um classificador específico f a partir de um conjunto de treinamento T , visando o estabelecimento de condições matemáticas que levam em consideração tanto o desempenho do classificador nesse conjunto, quanto sua complexidade, a fim de se obter um desempenho de igual capacidade também na classificação de novos dados entrantes (LORENA e CARVALHO, 2007).

Após exposição dos tópicos essenciais sobre aprendizado de máquinas, necessários à compreensão teórica da técnica de aprendizado em Máquinas de Suporte Vetorial, serão apresentados a seguir os principais conceitos, aplicações e ideias centrais que envolvem o uso dessa técnica, bem como sua relação com o Geomarketing.

2.4.3 Máquinas de Suporte Vetorial em Marketing

Muitos problemas de marketing exigem precisão ao se prever resultados ou estados futuros. Nas últimas duas décadas, previsões sobre o comportamento do consumidor tem atraído muita atenção. Entretanto, a previsão de outros fenômenos de mercado desempenha um papel fundamental na prática do marketing moderno, além de ser essencial para atingir os objetivos mais profundos na ciência do marketing (CUI e CURRY, 2005).

Cui e Curry (2005) ressaltam ainda que uma previsão precisa, embora essencial, é muitas vezes dificultada pelas complexas relações entre o agente previsor, as variáveis-alvo e uma falta de embasamento teórico para orientar a identificação do modelo.

Nesse contexto, as Máquinas de Suporte Vetorial podem prever de forma precisa tais ambientes. Assim, as principais características que tornam a utilização de Máquinas de Suporte Vetorial tão atrativa são, segundo Smola et al. (1999 *apud* LORENA e CARVALHO, 2007), as seguintes:

- I. Boa capacidade de generalização;
- II. Robustez em grandes dimensões;
- III. Convexidade da função objetivo;
- IV. Fundamentação teórica bem definida;

A capacidade de generalização de um classificador pode ser mensurada por meio da análise de sua eficácia ao se classificar outros dados cujos conjuntos de origem sejam diferentes do conjunto de dados de treinamento utilizado no processo de aprendizado. Os classificadores gerados por Máquinas de Suporte Vetorial apresentam baixo índice de *overfitting*, fazendo com que esses classificadores apresentem um bom desempenho de generalização (VAPNIK, 1995).

Com relação à robustez, os classificadores gerados por Máquinas de Suporte Vetorial apresentam um melhor desempenho diante de um conjunto de dados de grandes dimensões do que outros métodos de aprendizado de máquina, os quais frequentemente apresentam incidência de *overfitting* quando confrontados com dados dessa natureza (SEMOLINI, 2002).

Outra vantagem advinda da utilização de Máquinas de Suporte Vetorial consiste na existência de somente um mínimo global, pois se trata da otimização de uma função quadrática, diferentemente do que ocorre com outros métodos inteligentes, como as Redes Neurais Artificiais, que podem apresentar mais de um valor (CHEN e ODOBEZ, 2002).

As Máquinas de Suporte Vetorial apresentam também uma fundamentação teórica muito bem estabelecida, tanto no campo da matemática quanto no da

estatística. Segundo Hearst (1998), o algoritmo de suporte vetorial pode ser considerado como o ponto de intersecção entre o aprendizado prático e teórico.

Ainda, segundo a mesma autora, o algoritmo de suporte vetorial alcança tanto os fatores que precisam ser levados em consideração para um aprendizado de sucesso, precisamente definidos pela teoria de aprendizado estatístico; quanto os complexos modelos e algoritmo usualmente aplicados no mundo real.

Com as variadas demandas e o ambiente dinâmico, encontrados no contexto de marketing, a segmentação de mercado tem se tornado um conceito central tanto na teoria do marketing, quanto em sua prática (HUANG et al., 2007). Segundo os mesmos autores, a segmentação de mercado pode ser descrita como sendo o processo de dividir um amplo mercado em pequenos grupos de consumidores que possuem desejos e necessidades semelhantes, chamados de *clusters*.

As semelhanças entre cada segmento indicam comportamentos de compra semelhantes. As informações sobre os segmentos são úteis para a que os tomadores de decisão possam alcançar todos os consumidores de forma efetiva com um marketing-mix básico (ANDERSON e VINCZE, 2000).

Huang et al. (2007) destacam alguns benefícios que podem ser obtidos a partir da implementação de uma estratégia de segmentação de mercado. Um dos principais benefícios diz respeito ao fato de que os tomadores de decisão podem utilizar um marketing-mix específico para classificar pequenos mercados com uma melhor precisão, permitindo aos gestores aplicarem seus recursos de forma mais efetiva e eficiente.

Outro benefício advindo dessa prática é que a segmentação de mercado provê uma relação mais próxima e íntima entre os consumidores e a empresa. Além disso, os resultados de uma segmentação de mercado podem ser utilizados para que os tomadores de decisões determinem estratégias competitivas específicas (AAKER, 2001).

Assim, tendo em vista o alto valor estratégico agregado ao processo de segmentação de mercado, torna-se notável a necessidade de utilizar ferramentas que possuam um bom desempenho nessa atividade. Huang et al. (2007) propõem a

utilização de uma técnica de análise de *cluster* denominada *Support Vector Clustering* (SVC).

Essa técnica é utilizada para particionar um conjunto de dados em k grupos, de forma que cada grupo seja homogêneo em certos atributos, baseado em um critério específico previamente estabelecido. Essa proposta fez com que, baseado em resultados numéricos, a técnica SVC se tornasse uma ferramenta apropriada para a segmentação de mercado e até mesmo superasse outros métodos de segmentação (HUANG et al., 2007).

Entretanto, SVC nada mais é do que uma técnica de agrupamento de conjunto de dados baseado na teoria de *Support Vector Machines* (SVM) desenvolvida por Vapnik (1995, 1998). SVM utiliza o princípio da Minimização do Risco Estrutural para obter um melhor desempenho de generalização do que os métodos de aprendizado de máquina convencionais, que utilizam o princípio de Minimização do Risco Empírico. No caso do SVC, os autores estenderam o entendimento de SVM com o intuito de considerar também o problema de segmentação, denominado *clustering*.

A fim de demonstrarem a proposta desse método comparada a outras, Huang et al. (2007) desenvolveram em seu trabalho um estudo de caso envolvendo 40 consumidores em potencial de uma empresa de bebidas. O conjunto de dados contém quatro atributos de estilo de vida: (1) nível de socialização, (2) tempo de lazer, (3) conhecimento obtido e (4) realização.

Para comparar o desempenho dos três diferentes métodos os autores utilizaram como parâmetro o valor dos atributos e o erro-padrão gerado por cada método, representando-os em uma tabela. Concluiu-se que o SVC pôde separar bem cada segmento (*cluster*) com base na diferença de média entre os atributos. No que diz respeito ao erro-padrão, o SVC superou os outros métodos, pois agrupou os *clusters* de uma forma mais homogênea.

Shin e Cho (2006) apresentam uma abordagem semelhante. Parte-se do princípio de que o marketing está preocupado em identificar consumidores em potencial para determinados produtos ou serviços, promovendo-os aos potenciais clientes por meio de vários canais (SHIN e CHO, 2006). Dentre esses canais,

podemos citar como exemplo os componentes do marketing-mix, conforme visto anteriormente.

Nesse contexto, Shin e Cho (2006) propõem o conceito de modelo-resposta, cujo objetivo consiste em prever a probabilidade desse consumidor responder às promoções e ofertas. A partir desse modelo, pode-se identificar, dentro do conjunto de dados, um subconjunto de consumidores mais propensos à resposta do que outros consumidores.

Um modelo-resposta mais preciso obterá um maior número de respondentes e um baixo número de não-respondentes, ou seja, um menor número de pessoas que não responderam aos estímulos promocionais desenvolvidos pela empresa. A implementação deste método pode reduzir significativamente o custo global de marketing, sem sacrificar as oportunidades (SHIN e CHO, 2006).

Os mesmos autores ressaltam ainda que, vários métodos de aprendizado de máquina foram propostos como um modelo-resposta para esses problemas de marketing. Entretanto, as *Support Vector Machines* (SVM) se destacaram na comunidade de aprendizado de máquina por possuir uma teoria sólida e um desempenho prático (SHIN e CHO, 2006).

Assim, as SVM buscam posicionar uma fronteira de decisão de forma a maximizar a margem entre as duas classes. No caso do presente trabalho, essa fronteira será estabelecida entre as classes “indivíduos propensos” e “indivíduos não propensos” ao consumo de comida japonesa, provendo assim um modelo que auxilie na tomada de decisão do gestor.

O adequado posicionamento dessa fronteira é crucial para um bom desempenho do método. Métodos convencionais, tais como as Redes Neurais, utilizam o princípio de Minimização do Risco Empírico e tendem a se super ajustar (*overfit*) ao conjunto de dados de treinamento, resultando em uma má generalização, como visto anteriormente.

O fato das SVM se utilizarem de outro princípio, denominado princípio de Minimização do Risco Estrutural, permite que as SVM minimizem o erro no conjunto de dados de treinamento e, conseqüentemente, apresentem um melhor desempenho

de generalização, que é o principal objetivo em aprendizado estatístico (SHIN e CHO, 2006).

Entretanto, a aplicação de SVM como uma forma de modelo-resposta do método apresenta, segundo Shin e Cho (2006), algumas dificuldades. Primeiro, o treinamento por SVM pode se tornar computacionalmente intratável. Geralmente os gestores mantêm grandes quantidades de dados sobre seus clientes, adicionando registros de novos clientes sem o devido tratamento.

Segundo, o modelo-resposta tende a apresentar um grave problema de desequilíbrio de classes, tendo em vista que a taxa de resposta dos consumidores é geralmente muito baixa. Dessa forma, a maioria dos consumidores pertenceriam a classes dos “não-respondentes”, enquanto que apenas alguns consumidores estariam na classe de “respondentes”, fazendo com que o modelo não se comporte bem (SHIN e CHO, 2006).

A terceira dificuldade diz respeito às probabilidades. O modelo deve encontrar alguma maneira de estimar *scores* ou probabilidades a partir do SVM (SHIN e CHO, 2006). Dado que os recursos de marketing são limitados, os gestores de marketing visam maximizar o retorno ou receita total. Para tanto, faz-se necessário saber, dentre o total de indivíduos considerados, quais são aqueles mais prováveis a consumir determinado produto ou serviço.

Nesse contexto, o modelo-resposta computa, de cada indivíduo, sua propensão ou probabilidade de responder a estímulos de consumo específicos de um produto ou serviço. Esses valores são, então, usados para classificar os consumidores em uma ordem decrescente (SHIN e CHO, 2006).

Assim, os gestores de marketing podem direcionar seus recursos apenas aos indivíduos que possuem um maior *score* de consumo, ou seja, uma maior probabilidade consumir aquele produto. Porém, os classificadores de SVM retornam saídas (*outputs*) binárias, e não contínuas.

Para solucionar esses problemas, Shin e Cho (2006) apresentam em seu trabalho algumas soluções. Para o problema de intratabilidade computacional, eles apresentaram um algoritmo de seleção padrão que reduz o conjunto de dados de

tratamento sem perda de precisão. O algoritmo seleciona apenas os padrões próximos à fronteira de decisão, com base nas propriedades da vizinhança.

Para o problema de desequilíbrio de classes, os autores mostram como incorporar os diferentes custos de erros de classificação na função objetivo. Por fim, para o problema de probabilidades, os autores demonstram como a distância entre um padrão de consumo observado e o hiperplano de decisão pode ser utilizada como um *score* de consumo.

Cada indivíduo possui desejos e necessidades específicos e para que seja possível agrupá-las, é necessário dividir o mercado em partes menores e significativas, denominadas segmentos. Os trabalhos aqui apresentados evidenciam o crescimento da utilização de ferramentas de inteligência artificial para lidar com o problema de segmentação, bem como a importância da utilização de um método adequado e eficaz para se obter uma segmentação de mercado mais precisa. Como resultado, os gestores estarão mais aptos a desenvolver planejamentos e estratégias de marketing e de posicionamento de seus produtos adequadas ao perfil de consumo de seus clientes e consumidores.

3 MÉTODO

Como principal fonte de dados, foram utilizadas a Pesquisa de Orçamentos Familiares (POF) 2008-2009 e o Censo Demográfico de 2010, ambos realizados pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Como visto, a POF tem por objetivo fornecer informações sobre a composição dos orçamentos domésticos das famílias brasileiras, a partir da investigação dos hábitos de consumo, da alocação de gastos e da distribuição dos rendimentos, segundo as características dos domicílios e das pessoas.

A fim de determinar a localização mais vantajosa para a abertura do negócio, este trabalho se utilizou de duas áreas distintas para trabalhar os dados: Geomarketing e Aprendizado de Máquina, conforme visto anteriormente no Referencial Teórico deste trabalho.

Latour e Le Floch (2001) definem o termo Geomarketing como sendo um sistema integrado por programas de processamento de dados, métodos estatísticos e representações gráficas destinadas a produzir informações úteis para a tomada de decisões, através de instrumentos que combinam cartografia digital, gráficos e tabelas.

Nessa perspectiva, as técnicas utilizadas pelo Geomarketing têm o propósito inicial de avaliar a atividade econômica de uma determinada região e analisar seus respectivos dados sociodemográficos, traduzindo-os em ações estratégicas que podem ser mensuradas, planejadas e então implementadas pela empresa. Assim, é possível definir as áreas real e potencialmente propensas ao consumo de determinado produto ou serviço.

Por outro lado, faz-se necessário também uma seleção precisa do público alvo a ser atingido pela empresa. Do contrário, de nada valerá todo o esforço para o mapeamento de áreas em potencial.

No presente trabalho, tem-se que a **população** (universo) são todos os restaurantes de comida japonesa localizados no Distrito Federal, bem como todos os indivíduos que consomem comida japonesa e residem no Distrito Federal.

Para a **amostra**, foram selecionadas 39 unidades de franquias de *fast-food* de comida japonesas que atuam no Distrito Federal e apenas os indivíduos consumidores de comida japonesas que foram identificados na POF 2008-2009.

É necessária também a correta segmentação do mercado. Para tanto, este trabalho se utilizou de técnicas específicas da área de Aprendizado de Máquina, mais especificamente da técnica conhecida como Máquinas de Suporte Vetorial (SVM, do inglês *Support Vector Machine*).

Conforme visto anteriormente, as Máquinas de Suporte Vetorial (SVM) constituem uma técnica de aprendizado de máquina onde, a partir de um conjunto particular de exemplos, são obtidas conclusões genéricas. A Teoria de Aprendizado Estatístico desenvolvido por Vapnik (1995) é o fundamento principal que sustenta a técnica de aprendizado baseada em suporte vetorial. Essa teoria estabelece uma série de técnicas de aprendizado de máquina que devem ser seguidas para se possa obter classificadores que sejam capazes de prever corretamente a classe de novos dados do mesmo domínio, ou seja, maximizar a capacidade de generalização.

Considerando então os dados sociodemográficos obtidos a partir da Pesquisa de Orçamentos Familiares (POF) dos anos de 2008 a 2009, o SVM teria o intuito de reconhecer padrões no conjunto de dados, que nesse caso seriam os consumidores de comida japonesa, e poder assim prever para cada nova entrada de dados, a classe à qual os novos dados pertencerão.

Assim, as técnicas de suporte vetorial atuam como um classificador de uma amostra de dados, que nesse caso seria o conjunto de consumidores. O classificador obtido pode ser interpretado como uma função f , a qual recebe um conjunto de dados x e fornece uma predição y . No contexto desse trabalho, o conjunto de dados x representa o conjunto de consumidores e a predição y diz respeito à previsão de consumo ou não consumo de comida japonesa por parte do conjunto de consumidores analisado.

Nessa sessão, prosseguiu-se com a apresentação dos principais conceitos e definições necessários à compreensão das técnicas de coleta e análise de dados a serem executadas neste trabalho.

3.1 Elaboração da Cesta

Para a elaboração da Cesta de Produtos a ser utilizadas nessa pesquisa, foram selecionados, dentre todos os itens listados no Cadastro de Produtos da Pesquisa de Orçamentos Familiares (POF) 2008-2009, apenas os produtos específicos da culinária japonesa.

Com intuito de compor uma cesta mais precisa, não foram incluídos na Cesta de Produtos Japoneses desta pesquisa os itens considerados apenas ingredientes, e não específicos da culinária japonesa.

Assim, foram retirados da cesta todo e qualquer produto que, mesmo sendo utilizado como insumo para o preparo de pratos tipicamente japoneses, não podem ser considerados como um produto específico apenas desta culinária, como por exemplo, legumes, frutos do mar, verduras e frutas.

Dessa forma, após essa seleção de produtos a cesta final gerada, aqui denominada Cesta de Produtos Japoneses (CPJ), apresentou em sua composição um total de 25 itens, relacionados no Quadro 3.

QUADRO	GRUPO DE DESPESA	CODÍGO DO ITEM	PRODUTO
CADERNETA DE DESPESA	CEREAIS E LEGUMINOSAS	1802	ARROZ JAPONES ESPECIAL
		1801	ARROZ ESPECIAL JAPONES
	HORTALIÇAS FOLHOSAS, FRUTOSAS E OUTRAS	6001	NIRA
		3102	MOYASHI
	SAIS E CONDIMENTOS	3604	SHOYO
		3607	MOLHO SHOYU
		3605	TSUQUEMONO
		3601	MOLHO DE SOJA
		3603	MOLHO JAPONES
		3606	SAKURO
		6801	MISSO
		10503	MOLHO JAPONES LIGHT
		5502	KATSUOBUSHI (PEIXE EM PO)
		10505	TSUQUEMONO LIGHT
		10506	SAKURO LIGHT
	10504	SHOYO LIGHT	
	5901	KANI KAMA	

	ENLATADOS E CONSERVAS	1406	YAKISSOBA (SOPA DE LEGUMES DESIDRATADA)
		4402	ALGA EM CONSERVA
		6106	YAKISSOBA (SOPA DE LEGUMES DESIDRATADA) LIGHT
		4401	ALGAS EM CONSERVA
	LATICÍNIOS	3402	TOFU
	BEBIDAS ALCOÓLICAS	2302	SAQUE
	ALIMENTOS PREPARADOS OU SEMI PREPARADOS	9201	YAKISSOBA PARA VIAGEM
		4201	SUSHI PARA VIAGEM
		4202	SUSHI PRONTO PARA VIAGEM
		9401	FAMILIA FELIZ PARA VIAGEM (COMIDA JAPONESA)

Quadro 3 – Cesta de Produtos Japoneses

Fonte: Elaborado pela autora a partir da Pesquisa de Orçamentos Familiares 2008-2009 (IBGE,2010).

Durante a realização da POF 2008-2009, alguns desses itens foram descritos de diferentes formas pelos pesquisados. Entretanto, esses itens se referem ao mesmo produto, conforme veremos mais adiante.

Vale ressaltar que a seleção dos produtos que compõe essa cesta não envolveu a utilização de técnicas aprofundadas. Os produtos aqui listados foram selecionados basicamente por meio de pesquisas a respeito da origem cultural e gastronômica desses alimentos, bem como pela análise qualitativa dos produtos já cadastrados na POF.

3.2 Seleção das Variáveis

Após a elaboração da Cesta de Produtos Japoneses, prosseguiu-se com a seleção das variáveis que possam explicar ou influenciar o consumo dos alimentos anteriormente apresentados.

Na revisão de literatura realizada, não foram encontrados estudos que indiquem relação entre as características sociodemográficas dos indivíduos e o consumo de comida japonesa.

Os poucos estudos encontrados relacionavam a alimentação de uma forma geral com comportamentos psicológicos e antropológicos dos indivíduos, foco este que não seria de muita utilidade para o presente trabalho.

Dessa forma, com intuito de identificar quais variáveis poderiam explicar o consumo de comida japonesa, optou-se por prosseguir com a análise dos dados constantes na Pesquisa de Orçamentos Familiares.

Primeiramente, realizou-se uma análise qualitativa de todas as bases de consumo fornecidas pela POF 2008-2009, a fim de se identificar qual dessas bases contém os produtos que compõe a Cesta de Produtos Japoneses (CPJ). Após essa análise, foi identificado que a base de consumo CADERNETA DESPESA apresentava todos os produtos da CPJ.

Definida a base de consumo a ser investigada, prosseguiu-se com a leitura e análise dos dados. Para tanto, foram utilizados dois *softwares* distintos para se trabalhar os dados: o RStudio e o Microsoft Excel 2010.

O primeiro software, RStudio, foi utilizado principalmente para selecionar, dentro da base de dados CADERNETA DESPESA, apenas aqueles indivíduos que consumiram os produtos que compõe a CPJ, de acordo com critérios de seleção previamente estabelecidos.

Programou-se a leitura da referida base, de forma que o *software* gerasse outra base, denominada CONSUMO OBSERVADO, com os dados de todos os indivíduos que durante o período de realização da pesquisa, consumiram no Brasil produtos relacionados à culinária japonesa, aqui representados pela CPJ.

Com a base de dados CONSUMO OBSERVADO, deu-se início à utilização do segundo *software*, o Microsoft Excel. Foram feitas análises qualitativas, em busca de padrões de consumo nos dados observados. É importante frisar que nem todas as variáveis sociodemográficas aqui apresentadas foram utilizadas na etapa de previsão da máquina de aprendizado.

O objetivo principal a partir da análise dessas variáveis consiste em obter uma melhor compreensão do fenômeno de consumo de comida japonesa, identificando possíveis características que possam vir a influenciar o consumo desse tipo de alimento. As informações obtidas serão descritas a seguir, de acordo com as variáveis selecionadas.

3.2.1 Unidade da Federação

No que diz respeito à Unidade da Federação, foi observado que as maiores concentrações de consumidores de comida japonesa encontram-se no Paraná e em São Paulo, com 13,37% e 12,54% do total de respondentes de todo o Brasil, respectivamente.

Os outros estados classificados de forma subsequente aos anteriores foram, nessa ordem, Mato Grosso do Sul, Minas Gerais, Pará, Ceará, Pernambuco, Santa Catarina, Espírito Santo e Mato Grosso.

Com relação ao Distrito Federal, foram observados apenas 19 indivíduos que consumiram algum dos produtos que compõem a CPJ, o que representa cerca de apenas 3,17% do total de respondentes da amostra selecionada.

As diferenças observadas nesse percentual se devem, principalmente, à quantidade de pessoas pesquisadas em cada Unidade da Federação. Sabendo-se que esse número deve ser proporcional ao número de setores censitários que compõem cada Estado, é natural que os estados mais populosos apresentem os maiores percentuais.

Portanto, considera-se que não existe, necessariamente, um padrão de consumo associado à região na qual o indivíduo se encontra. Sendo assim, assumiu-se nesse trabalho que o padrão de consumo para comida japonesa observado em território nacional será o mesmo em qualquer estado do País, inclusive no Distrito Federal.

3.2.2 Idade

Levando-se em consideração a variável idade, pode-se observar um consumo de comida japonesa significativamente maior na faixa etária entre os 10 e 29 anos de idade. Juntos totalizam 221 indivíduos, o que representa cerca de 36,95% do total de consumidores identificados.

A segunda faixa etária mais consumidora desses produtos se concentra nos indivíduos que possuem entre 30 e 49 anos de idade, representado 29,93% dos entrevistados. Já com relação ao público idoso, indivíduos com mais de 70 anos de

idade representam apenas 5,18% dos consumidores de comida japonesa para a amostra selecionada.

Dessa forma, é possível visualizar uma maior predominância do público jovem no consumo de comida japonesa. Tal constatação fornece insumos para que os gestores desenvolvam uma estratégia de marketing mais direcionada, utilizando-se, por exemplo, de meios de comunicação que sejam mais presentes no cotidiano do público identificado, tais como a web e redes sociais.

3.2.3 Gênero

Com relação à variável gênero, foi observada uma ligeira predominância de mulheres no consumo de comida japonesa, que juntas representam cerca de 55,18% dos total de consumidores dessa amostra.

Sendo assim, pode-se afirmar com base nos dados analisado, que não existe predominância de gênero no que diz respeito ao consumo de comida japonesa. Do ponto de vista do gestor, tem-se que, considerando o contexto analisado, essa não seria uma variável tão relevante para o desenvolvimento de estratégias de marketing específicas para o público masculino ou feminino.

3.2.4 Escolaridade

No quesito grau de escolaridade, quase 40% dos indivíduos da amostra selecionada não informaram seu nível instrucional. Entretanto, dentre os indivíduos que optaram por responder, os que possuem Ensino Médio representam 25,54% desse grupo, enquanto que o Nível Superior (Graduação) representa cerca de 23,90% do total de respondentes.

Considerando a alta porcentagem de indivíduos não respondentes sobre o grau de escolaridade, não é possível afirmar que exista alguma predominância relevante para esse quesito no que diz respeito ao consumo de comida japonesa. Dessa forma, para que seja possível desenvolver estratégias de mercado mais confiáveis no âmbito do nível de escolaridade, deve-se investigar outras fontes que ofereçam dados mais precisos.

3.2.5 Raça

A variável raça apresentou entre os consumidores de comida japonesa uma nítida prevalência de indivíduos que se auto declaram brancos, representando 56,52% do total da amostra. Na sequência, seguem os indivíduos que se declaram Pardos, com 32,10% de representação.

Por fim, os indivíduos da raça amarela ocupam a terceira posição nesse cenário, representando cerca de 8,86% da amostra. Os indivíduos declarados como negros, indígenas ou que declararam não saber sua cor ou raça, juntos somam apenas 2,50% do total de consumidores da amostra.

Do ponto de vista gerencial, pode-se afirmar que a variável raça possui uma nítida influência no consumo de comida japonesa. Conseqüentemente, identificar as regiões nas quais os indivíduos que se auto declaram brancos se concentram pode representar um fator diferencial no negócio.

3.2.6 Cartão de Crédito

Possuir ou não possuir cartão de crédito se mostrou ser uma variável de pouca influência no consumo de comida japonesa. Do total de consumidores observados na amostra, 41,47% afirmaram não ter cartão de crédito, 35,78% declarou possuir cartão de crédito e os outros 22,74% restantes não forneceram essa informação na pesquisa.

Apesar dos resultados obtidos na análise, é importante ressaltar que, do ponto de vista do gestor, possuir cartão de crédito como forma de pagamento pode ser um diferencial. Segundo Ikeda e Wang (2005) o mercado de cartões de crédito é um dos que obtiveram maior lucro e expansão no Brasil nos últimos dez anos.

3.2.7 Renda

O consumo de comida japonesa mostrou-se ser influenciável pela variável renda. Um total de 76,58% dos consumidores da amostra possui uma renda *per capita* de até R\$ 3000,00. Dentre esse grupo, cerca de 44,97% possui uma renda *per capita* de até R\$ 1000. Os indivíduos que possuem uma renda *per capita* acima de R\$ 6000 somam apenas 1,003% da amostra.

A predominância dessa faixa de renda pode trazer informações relevantes para o gestor. De acordo com Rodrigues et al. (2014) essa é justamente a faixa de renda que compõe a denominada “nova classe média” que, segundo os mesmos autores, corresponde a mais da metade da população brasileira. Reforçando a ideia do crescimento do uso de cartão de crédito apresentada anteriormente, Rodrigues et al. (2014) cita ainda que o 62% dos cartões de créditos atualmente disponíveis no Brasil se concentra justamente nessa classe social.

3.2.8 Produtos Consumidos

O produto japonês mais consumido pelos brasileiros no referido período foi o Molho Shoyu, representando 72,40% dos produtos consumidos. Tal produto foi descrito de quatro diferentes formas pelos pesquisados: (1) Molho de Soja, (2) Molho Shoyu, (3) Shoyo e (4) Molho Japonês. Apesar dessas definições, todos esses itens se referem ao mesmo produto, aqui denominado apenas de Molho Shoyu. O Gráfico 2 a seguir ilustra a representação percentual de cada um dos produtos na Cesta de Produtos Japoneses.

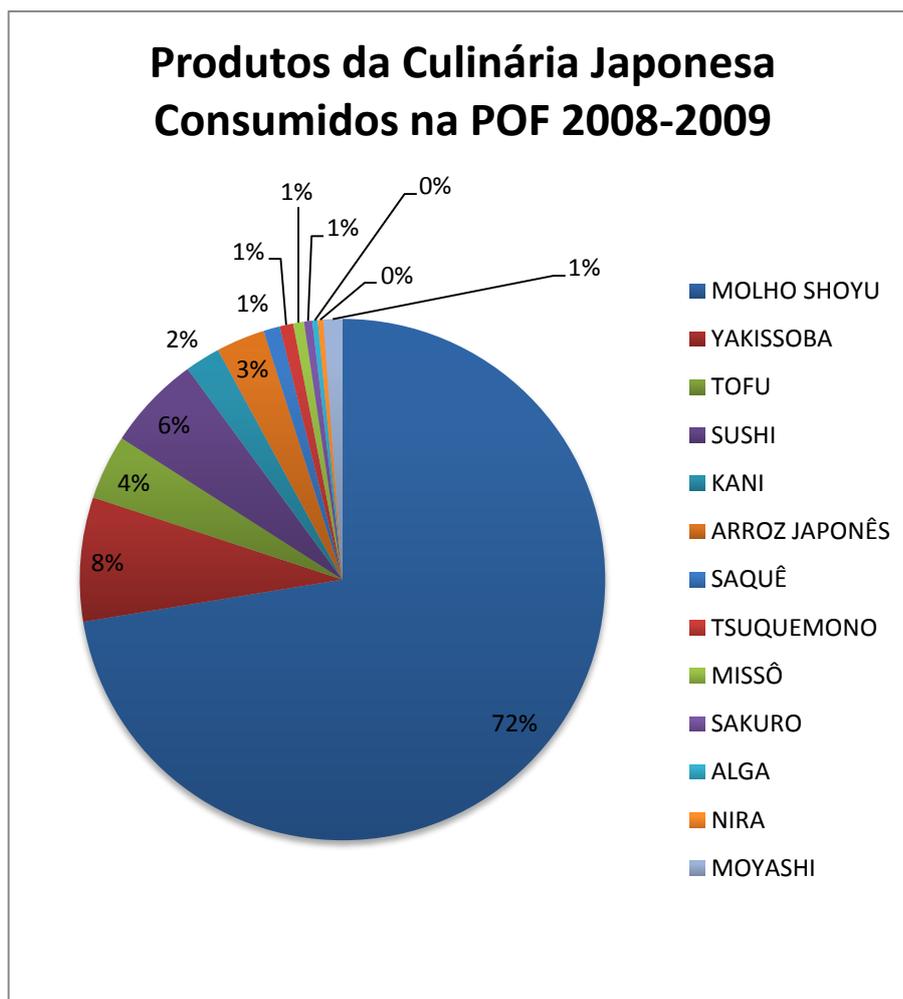


Gráfico 2 – Produtos Japoneses Consumidos na POF 2008-2009

Fonte: Elaborado pela autora a partir da Pesquisa de Orçamentos Familiares 2008-2009 (IBGE, 2010).

Depois do Molho, o segundo produto mais consumido foi o Yakissoba, com 7,69% de representação. Os itens “YAKISSOBA (SOPA DE LEGUMES DESIDRATADA)” e “YAKISSOBA PARA VIAGEM” também foram considerados como sendo apenas Yakissoba.

O mesmo aconteceu para os itens “ARROZ ESPECIAL JAPONÊS”, “ARROZ JAPONÊS ESPECIAL” e “SUSHI PARA VIAGEM”, “SUSHI PRONTO PARA VIAGEM”, “FAMILIA FELIZ PARA VIAGEM (COMIDA JAPONESA)” que foram considerados como sendo apenas Arroz Japonês e Sushi, respectivamente.

3.2.9 Valor Gasto

Com relação ao valor dispendido no consumo desses produtos, cerca de 92,15% dos consumidores da amostra gastaram até R\$ 10,00. Enquanto isso, apenas 0,34% dos pesquisados gastaram mais de R\$ 30,00.

O principal motivo que pode justificar a prevalência do baixo valor dispendido no consumo desses produtos diz respeito ao fato de que, como visto anteriormente, o produto japonês mais consumido pelos brasileiros foi o Molho Shoyu, que possui um baixo valor monetário. O preço médio do frasco de 150ml da marca mais consumida deste produto, gira em torno de R\$ 3,00 a unidade. Considerando que o Molho Shoyu representou cerca de 72% dos produtos japoneses consumidos, é natural que o valor dispendido seja baixo, podendo não representar fielmente a realidade de consumo atual.

3.3 Treinamento da Máquina

Após a seleção das variáveis sociodemográficas que podem influenciar o consumo de comida japonesa, prosseguiu-se com a aplicação de Máquinas de Suporte Vetorial, com o intuito de realizar uma previsão de consumo dos indivíduos identificados.

Para essa etapa foi utilizado o Censo Demográfico de 2010, também realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), a fim de determinar a localização geográfica na qual os indivíduos com alta propensão de consumo se encontram, mais especificamente o setor censitário ao qual esses indivíduos pertencem. Segundo o IBGE (2010), um setor censitário pode ser definido como sendo:

[...] “a unidade de controle cadastral formada por área contínua, de característica totalmente urbana ou totalmente rural, com dimensão e número de domicílios, ou de estabelecimentos, que permitam o levantamento das informações por um único agente” e [...] “a menor unidade territorial, com limites físicos identificáveis em campo, com dimensão adequada à operação de pesquisas e cujo conjunto esgota a totalidade do Território Nacional, o que permite assegurar a completa cobertura do País”. (IBGE, 2010, P. 34)

Ainda sobre os setores censitários, o IBGE (2010) afirma que:

[...] “seus limites devem respeitar os limites territoriais legalmente definidos e os estabelecidos pelo IBGE para fins estatísticos. São definidos, preferencialmente, por pontos de referência estáveis e de fácil identificação em campo, de modo a evitar que um agente omita a coleta na área sob sua responsabilidade, ou invada a unidade territorial de responsabilidade de outro agente”. (IBGE, 2010, p. 34)

É importante citar também que cada setor censitário, dependendo do seu tipo, urbano ou rural, pode possuir de 250 a 350 domicílios; ou de 150 a 250 domicílios ou de 100 a 200 estabelecimentos agropecuários, respectivamente.

Como já mencionado anteriormente, a base de dados utilizada para esta etapa do trabalho foi a CADERNETA DESPESA da POF, por conter todos os produtos da CPJ. Dessa forma, prosseguiu-se com a importação dessa base para o *software* RStudio. A partir dessa base, foram selecionados apenas os produtos definidos na CPJ, cujos números quadro variavam de 63 a 69. Dentro desses quadros, cada produto possuía um número de código item individual, conforme visto anteriormente no Quadro 3.

Após a seleção desses produtos, foi atribuído a cada indivíduo (observação) um valor referente ao gasto com consumo de comida japonesa. Dessa forma, quem consumiu algum dos produtos da CPJ recebe o valor consumido e quem não consumiu nenhum desses produtos recebe o valor zero.

Obtidos os valores de gasto com comida japonesa, procedeu-se para a aplicação do peso no valor gasto, multiplicando-se o gasto pelo fator de expansão da POF com o intuito de se obter um valor mais aproximado da realidade. Após a ponderação, calculou-se o gasto total com comida japonesa por pessoa. Dessa forma foi possível visualizar as observações de consumo no Distrito Federal por setor censitário, que totalizaram 4.690 setores.

Para a próxima etapa foi utilizada outra base de dados da POF, denominada MORADOR. Essa base contém todas as informações sociodemográficas de cada um dos indivíduos identificados na POF. Para esta etapa foram consideradas como variáveis de característica a idade, o gênero, a renda, a alfabetização e a raça dos indivíduos. Para as variáveis numéricas (idade e renda) foram criados intervalos de valor específicos. Todas as variáveis foram ponderadas com o fator de expansão.

Após a criação e ponderação das variáveis acima citadas, os resultados obtidos foram agregados aos setores censitários. Assim, procedeu-se com a união entre as bases de gasto com comida japonesa e dos setores censitários agregados às variáveis descritas. O resultado dessa união foi uma nova base de dados onde consta o consumo de comida japonesa por setor censitário, agregado às variáveis sociodemográficas. Com essa nova base de dados iniciou-se a análise por meio de Máquinas de Suporte Vetorial.

Como visto anteriormente, a máquina de aprendizado funciona como uma ferramenta com a capacidade de realizar uma análise dos dados provenientes do mundo real e classifica-los. É justamente essa classificação correta dos dados que a máquina tenta aprender. Assim, cada possível classificação que a máquina prevê para os dados é denominada uma hipótese.

Nesse contexto, o principal objetivo que se busca com a aplicação da máquina de aprendizado é que ela seja capaz de gerar a hipótese que mais se aproxima da realidade dos dados. Para tanto, é necessário que a máquina utilize em seu treinamento exemplos de dados que já foram classificados da forma correta.

Assim, a etapa de treinamento tem o intuito de fazer com que a máquina defina o padrão a ser utilizado para classificar os dados. Após o treinamento da máquina é necessário executar a validação do classificador, utilizando-se para tal um novo conjunto de dados que, apesar de ter a mesma origem, é constituído de dados diferente dos utilizados na amostra de treinamento. A etapa de validação visa ajustar o classificador, aperfeiçoando-o. Por fim, a etapa de teste é aplicada com o intuito de avaliar o desempenho de generalização da máquina de aprendizado, por meio da classificação de novos dados.

Por meio da criação de uma variável binária, a base de dados utilizada nesta etapa do trabalho foi separada nas duas partes: treinamento e validação. A parte de treinamento corresponde à 70% dos dados da base, enquanto que a parte de validação corresponde aos 30% restantes dos dados da base.

De forma sucinta, o método pode ser descrito da seguinte maneira: considere um conjunto de dados de treinamento $T = \{(x_k, y_k)\}_{k=1}^N$, de tal forma que o x_k é o dado entrante do k -ésimo padrão de classificação e y_k é o “rótulo” da classe pertencente

ao k -ésimo padrão de classificação. Assim, o objetivo do método é encontrar o classificador que separe os dados do conjunto de treinamento T com o menor erro possível, de forma a maximizar a capacidade de generalização da máquina, ou seja, maximizar a margem que separa as classes dos dados, conforme visto anteriormente na Figura 10. Criadas as bases de treinamento e validação, procedeu-se para a criação da lista de parâmetros a ser utilizada para o aprendizado da máquina. São eles os parâmetros C e Sigma (σ).

Ambos os parâmetros são definidos pelo usuário do método. O parâmetro C , também conhecido como parâmetro de regularização, determina, em maior ou menor grau, a importância dos erros de classificação gerados pelo algoritmo aprendido pela máquina. Já o parâmetro σ está relacionado com a qualidade da previsão feita pela máquina. Dessa forma, é possível perceber a existência de um *trade-off* no processo de seleção dos parâmetros no que diz respeito à complexidade do método. Se apresentarmos um baixo valor do parâmetro de qualidade σ e um alto valor do parâmetro de erro C , o classificador se super ajustará aos dados do conjunto de treinamento, causando, como já visto anteriormente, o fenômeno denominado *overfitting*. Em contrapartida, se o valor do parâmetro σ for alto e o do parâmetro C for baixo, o classificador se sub ajustará aos dados do conjunto de treinamento, tornando-se suscetível a muitos erros, causando o chamado *underfitting*. Dessa forma, utilizou-se para cada par de parâmetros, aquele cujo Erro Quadrático Médio (EQM) apresentasse o menor valor.

O EQM consiste no somatório do quadrado diferenças entre o valor que foi previsto (y_i) pela máquina e o valor real observado dos dados (\hat{y}_i), ponderados pelo número de termos observados (n), conforme descrito na equação a seguir.

$$EQM = \sum_i \frac{(y_i - \hat{y}_i)^2}{n}$$

A obtenção dos parâmetros de menor EQM visa diminuir a diferença entre o estimador gerado pelo aprendizado da máquina e o verdadeiro valor da quantidade estimada, fazendo com que a máquina seja capaz de realizar uma previsão mais próxima possível da realidade.

De uma forma mais objetiva, a aplicação do método e a descrição dos procedimentos metodológicos utilizados neste trabalho podem ser apresentadas de maneira sintética, conforme ilustra a Figura 11.

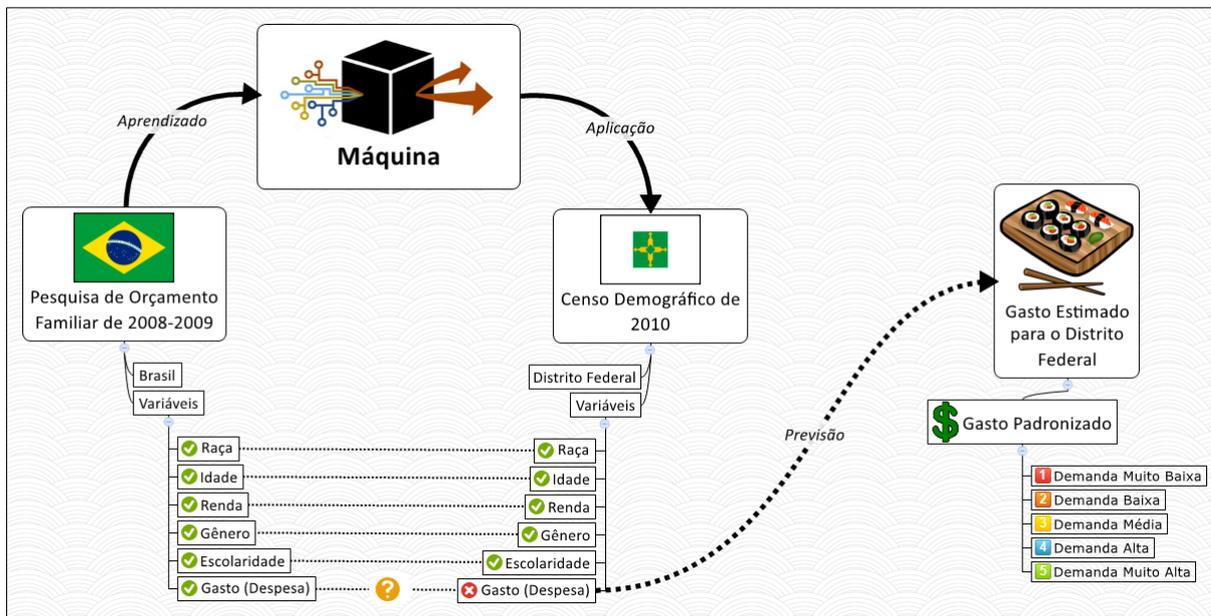


Figura 11 – Procedimentos Metodológicos

Fonte: Elaborado pela autora.

Primeiramente, é possível verificar na Figura 11 que a máquina se utilizou de duas bases de dados distintas para os processos de aprendizado e de aplicação do método, conforme já visto anteriormente. Para a etapa de aprendizado da máquina, utilizou-se como fonte de dados a Pesquisa de Orçamentos Familiares de 2008-2009, em um nível de abrangência nacional, com o intuito de identificar padrões no consumo de comida japonesa no Brasil. Como dito anteriormente, 70% da base de dados da POF foi destinada ao treinamento da máquina, enquanto que os 30% restantes foram destinados à validação do método. Já para a aplicação do padrão aprendido pela máquina, utilizou-se outra base de dados, proveniente do Censo Demográfico de 2010. É importante ressaltar que ambas as pesquisas possuem abrangência nacional. No entanto, na etapa de aplicação, este trabalho se utilizou apenas dos dados do Censo Demográfico de 2010 para o Distrito Federal, que é a região alvo de estudo deste trabalho. Com base na ilustração da Figura 11, é possível notar que ambas as bases de dados possuem variáveis em comum, tais como raça, idade, renda, gênero, escolaridade, entre outras.

Existem, porém, algumas variáveis que são específicas a apenas uma das bases, como é o caso da variável gasto, que está presente apenas na Pesquisa de Orçamentos Familiares. O gasto, também conhecido como despesa, diz respeito à destinação que um indivíduo dá a seus recursos, ou seja, o que é consumido por este indivíduo. No contexto deste trabalho, para cada indivíduo identificado, atribuiu-se na variável gasto o valor em reais (R\$) dispendido pelo indivíduo com comida japonesa neste período e o valor zero caso o indivíduo não tenha consumido nenhum dos produtos da Cesta de Produtos Japoneses. Dessa forma, foram considerados para o processo de treinamento da máquina, apenas aqueles indivíduos que consumiram pelo menos um dos produtos que compõe a cesta de produtos japoneses.

Assim, a etapa de aprendizado consiste em fazer com que a máquina identifique padrões no consumo de comida japonesa, com bases nos indivíduos consumidores destes produtos identificados no Brasil, ou seja, indivíduos que apresentaram na variável gasto um valor diferente de zero. Dessa forma, na etapa de aplicação do método, o principal objetivo consiste em fazer com que a máquina seja capaz de prever para o Distrito Federal, por meio do padrão de consumo nacional identificado, aquela variável que não é comum às bases, ou seja, o gasto com comida japonesa. O resultado dessa previsão fornece um gasto estimado em valores padronizados, tornando possível a identificação da demanda por comida japonesa para cada um dos 4.690 setores censitários que compõe o Distrito Federal, conforme veremos mais adiante.

3.4 Aplicação do Método para a amostra do Distrito Federal

Após o aprendizado da máquina, a próxima etapa envolveu a criação de uma nova base de dados. Essa nova base será utilizada para que a máquina possa prever, por meio de valores padronizados, o gasto com comida japonesa em cada um dos setores censitários do Distrito Federal.

Conforme dito anteriormente, considerou-se que o padrão de consumo para comida japonesa observado em território nacional será o mesmo em qualquer estado do País, inclusive no Distrito Federal. Dessa forma, a máquina predirá, com base no padrão de consumo nacional observado, a demanda por comida japonesa nos setores

censitários do Distrito Federal, utilizando-se para tal o Censo Demográfico de 2010, também elaborado pelo IBGE.

Essa nova base foi constituída por meio da união das classes criadas para cada uma das variáveis citadas anteriormente, a saber – alfabetização, idade, gênero, renda e raça. As variáveis foram divididas em várias classes, conforme ilustra o Quadro 4.

VARIÁVEL	CLASSE
Alfabetização	Pessoas Alfabetizadas
Idade	Pessoas de 0 a 9 anos de idade
	Pessoas de 10 a 19 anos de idade
	Pessoas de 20 a 29 anos de idade
	Pessoas de 30 a 39 anos de idade
	Pessoas de 40 a 49 anos de idade
	Pessoas de 50 a 59 anos de idade
	Pessoas de 60 a 69 anos de idade
	Pessoas de 70 anos de idade ou mais
Gênero	Homens
	Mulheres
Renda	Renda 1: Domicílios particulares com rendimento nominal mensal domiciliar <i>per capita</i> de até 1/8 salário mínimo (R\$ 63,75)
	Renda 2: Domicílios particulares com rendimento nominal mensal domiciliar <i>per capita</i> de mais de 1/8 a 1/4 salário mínimo (R\$ 63,76 até R\$ 127,50)
	Renda 3: Domicílios particulares com rendimento nominal mensal domiciliar <i>per capita</i> de mais de 1/4 a 1/2 salário mínimo (R\$ 127,51 até R\$ 255,00)
	Renda 4: Domicílios particulares com rendimento nominal mensal domiciliar <i>per capita</i> de mais de 1/2 a 1 salário mínimo (R\$ 255,01 até R\$ 510,00)
	Renda 5: Domicílios particulares com rendimento nominal mensal domiciliar <i>per capita</i> de mais de 1 a 2 salários mínimos (R\$ 510,01 até R\$ 1.020,00)
	Renda 6: Domicílios particulares com rendimento nominal mensal domiciliar <i>per capita</i> de mais de 2 a 3 salários mínimos (R\$ 1.020,01 até R\$ 1.530,00)
	Renda 7: Domicílios particulares com rendimento nominal mensal domiciliar <i>per capita</i> de mais de 3 a 5 salários mínimos (R\$ 1.530,01 até R\$ 2.550,00)
	Renda 8: Domicílios particulares com rendimento nominal mensal domiciliar <i>per capita</i> de mais de 5 a 10 salários mínimos (R\$ 2.550,00 até R\$ 5.100,00)
	Renda 9: Domicílios particulares com rendimento nominal mensal domiciliar <i>per capita</i> de mais de 10 salários mínimos (acima de R\$ 5.100,00)
Raça	Branca
	Preta
	Amarela
	Parda
	Indígena

Quadro 4 – Classes das Variáveis

Fonte: Elaborado pela autora a partir do Censo Demográfico de 2010 (IBGE, 2011).

Vale ressaltar que, diferentemente da Pesquisa de Orçamentos Familiares, o Censo Demográfico fornece apenas o número de pessoas alfabetizadas por setor censitário, e não fornece o número de pessoas analfabetas. Por consequência, o número de pessoas analfabetas é o igual ao total de pessoas identificadas naquele setor, subtraídas as pessoas alfabetizadas.

Outra ressalva diz respeito à variável renda. As classes de renda do Censo Demográfico são baseadas no salário mínimo do ano no qual a pesquisa é realizada. De acordo com a Lei nº 12.255, de 15 de junho de 2010 o salário mínimo estabelecido no ano foi no valor de R\$ 510,00. Este foi o parâmetro utilizado para classificar as faixas de renda.

Os dados foram agrupados de forma a identificar a quantidade de indivíduos com determinada característica, como por exemplo indivíduos entre 10 e 19 anos de idade, dentro de cada setor censitário. Assim, todas as variáveis e suas respectivas classes possuem em comum a coluna com o geocódigo do setor.

Dessa forma, tem-se que a base de dados do Distrito Federal é constituída das 26 classes de variáveis identificadas acima, a saber – Código do Setor Censitário, Alfabetizados, Idade 0 a 9, Idade 10 a 19, Idade 20 a 29, Idade 30 a 39, Idade 40 a 49, Idade 50 a 59, Idade 60 a 69, Idade 70 acima, Homem, Mulher, Renda 1, Renda 2, Renda 3, Renda 4, Renda 5, Renda 6, Renda 7, Renda 8, Renda 9, Branca, Preta, Amarela, Parda e Indígena – para cada um dos 4.690 setores censitários do Distrito Federal.

Com a base de dados do Distrito Federal pronta, é possível prever o gasto com comida japonesa no DF por meio da máquina de aprendizado constituída na seção anterior. Para tanto, serão utilizadas a base de dados do Distrito Federal e a base de validação do método.

Ambas as bases foram programadas para possuírem as mesmas 26 classes de variáveis citadas anteriormente. A base de dados do Distrito Federal apresenta as observações identificadas em cada setor censitário do Distrito Federal por meio do Censo Demográfico de 2010, enquanto que a base de validação apresenta os dados de validação referentes ao padrão de consumo observado no Brasil por meio da Pesquisa de Orçamentos Familiares 2008-2009.

Dessa forma, base de dados final gerada nesta etapa apresenta 27 variáveis: as 26 classes já citadas anteriormente e mais 1 variável, denominada aqui de Gasto Padronizado (GP). Esta variável é obtida por meio da diferença entre o Gasto registrado no referido setor censitário (G) e o Gasto Médio registrado no Distrito Federal (\bar{G}), ponderada pelo desvio-padrão (σ), fornecendo valores padronizados que

representam a demanda por comida japonesa para cada um dos 4.690 setores censitários do Distrito Federal, conforme descrito na equação:

$$GP = \frac{(G - \bar{G})}{\sigma}$$

A ideia de estabelecer uma escala de valores padronizados é a de que, quanto mais próximo de zero o valor do gasto padronizado for, mais próximo ele estará da média de gasto observada no Distrito Federal.

Para uma melhor análise espacial dos gastos padronizados obtidos nesta etapa do trabalho, optou-se por selecionar intervalos de valores específicos, que possuem suas respectivas classificações e cores a serem exibidas nos mapas, conforme demonstra as informações apresentadas no Quadro 5.

Nº	COR	INTERVALO DE VALORES	CLASSIFICAÇÃO
1		-0.7418 até -0.0796	Demanda Muito Baixa
2		-0.0796 até 0.1202	Demanda Baixa
3		0.1202 até 0.3722	Demanda Média
4		0.3722 até 0.8787	Demanda Alta
5		0.8787 até 1.9196	Demanda Muito Alta

Quadro 5 – Classificação do Gasto Padronizado

Fonte: Elaborado pela autora.

Um valor de gasto padronizado que esteja contido no primeiro intervalo de valores, representa um setor censitário onde o consumo de comida japonesa foi muito abaixo da média de consumo observada no Distrito Federal. Já um gasto padronizado cujo valor esteja contido no segundo intervalo, diz respeito a um setor censitário onde o consumo observado foi um pouco abaixo da média do Distrito Federal. Um setor censitário cujo valor do gasto padronizado esteja contido no terceiro intervalo, representa uma demanda média, ou seja, o consumo identificado neste setor foi próximo do consumo médio observado no Distrito Federal. Já um valor de gasto padronizado que esteja contido no quarto intervalo de valores, representa um setor onde a demanda pode ser considerada como sendo alta, pois apresenta um gasto acima do observado no Distrito Federal. Por fim, um gasto padronizado cujo valor esteja contido no quinto e último intervalo, representa um setor censitário onde o gasto padronizado observado foi muito acima da média do Distrito Federal, ou seja, um setor censitário que apresenta uma alta demanda por comida japonesa.

Nesse contexto, serão considerados como sendo oportunidade de negócio, todas as regiões em que for observada uma baixa oferta de comida japonesa, levando-se em consideração a amostra de 39 unidades de franquias utilizada neste trabalho, mas que apresentem uma alta concentração de indivíduos com gasto padronizado elevado para o consumo de comida japonesa, ou seja, indivíduos considerados como sendo consumidores em potencial. Na próxima seção foram definidos, de maneira mais específica, todos os critérios que devem ser cumpridos para que um local seja considerado uma oportunidade de negócio.

4 RESULTADOS

4.1 Identificação de Oportunidades de Negócio

Após a obtenção do gasto padronizado para cada um dos setores censitários do Distrito Federal, procedeu-se para o georreferenciamento dos dados, com intuito de identificar quais são as regiões no Distrito Federal mais propensas ao consumo de comida japonesa, ou seja, regiões que possuem um gasto padronizado alto, mas que não possuem restaurantes de comida japonesa próximos.

De maneira geral, foi estabelecido para o presente trabalho que uma região será considerada como sendo uma oportunidade de negócio caso cumpra, concomitantemente, os seguintes critérios:

- I. Estar localizado em um setor censitário do tipo urbano;
- II. Não estar localizado em uma região residencial;
- III. Não estar próximo a nenhuma das franquias da amostra em um raio de pelo menos 1 km;
- IV. Estar dentro ou próximo de centros comerciais;
- V. Possuir fácil acesso por meio de transporte público ou particular;
- VI. Estar dentro, ou a no máximo 500 metros de distância, de um setor censitário cujo gasto padronizado possua um valor acima de 0.3722;
- VII. O subdistrito analisado deve possuir uma quantidade mínima de três setores censitários que cumpram o critério anterior.

Outra forma de interpretar o sexto critério pode ser descrita da seguinte maneira: o valor do gasto padronizado deve estar contido no quarto ou quinto intervalos, que representam, respectivamente, uma demanda alta e uma demanda muito alta, conforme visto no Quadro 5.

Para esta etapa do trabalho, foi utilizado outro *software* livre chamado Quantum GIS (QGIS) 2.4, que é um Sistema de Informação Geográfica (SIG). No QGIS é possível inserir, visualizar, gerenciar, georreferenciar, editar e analisar os mais diversos dados, bem como compor mapas.

Primeiramente, foi feito o mapeamento das 39 franquias de *fast-food* de comida japonesa utilizadas como amostra neste trabalho. Para tanto, foram obtidas as

coordenadas geográficas de cada franquia por meio de seus respectivos CEPs (Código de Endereçamento Postal), disponíveis para consulta nos próprios sites das franquias.

Para obter as coordenadas geográficas das franquias, inseriu-se os CEPs no Google Maps e posteriormente utilizou-se a função “O que há aqui?”, clicando com o botão direito do mouse sobre o ponto indicado no mapa. Essa função fornece as coordenadas geográficas no formato de grau decimal, conforme ilustra a Figura 12.

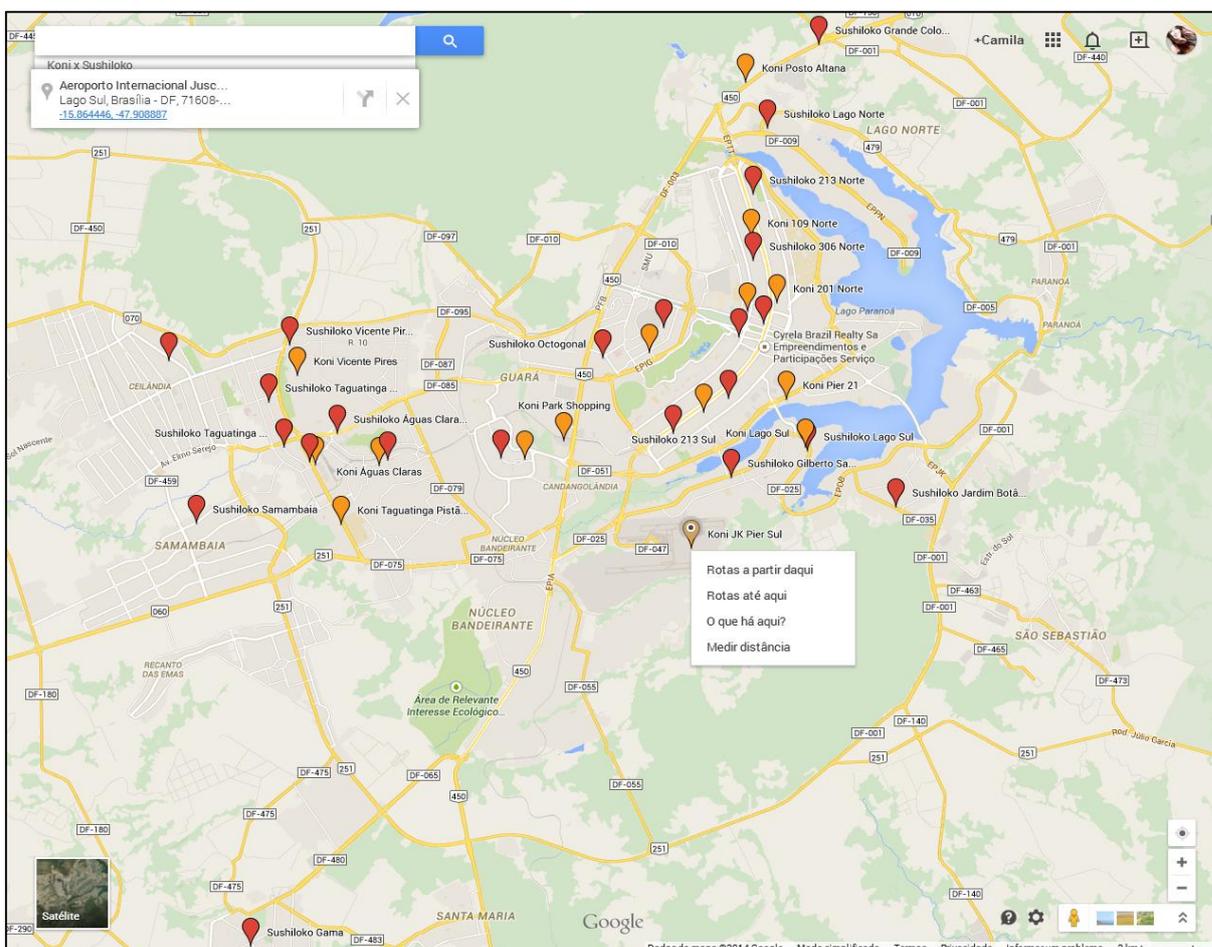


Figura 12 – PrintScreen Google Maps

Fonte: Google Maps.

Após a obtenção das 39 coordenadas geográficas, criou-se um arquivo no Excel no formato CSV (*Comma Separated Value*), formato este que permite a decodificação e georreferenciamento pelo QGIS. O arquivo criado com todas as coordenadas é ilustrado no Quadro 6.

Nº	FRANQUIA	LATITUDE	LONGITUDE
1	Sushiloko Águas Claras Av. Araucárias	-15.839.574	-48.018930
2	Sushiloko Águas Claras Av. Castanheiras	-15.830300	-48.037122
3	Sushiloko 213 Norte	-15.744461	-47.886885
4	Sushiloko 306 Norte	-15.768456	-47.887164
5	Sushiloko Conjunto Nacional	-15.791155	-47.883085
6	Sushiloko 213 Sul	-15.830479	-47.915906
7	Sushiloko 406 Sul	-15.818113	-47.895875
8	Sushiloko Pátio Brasil	-15.795733	-47.892281
9	Sushiloko Gama	-16.014693	-48.068384
10	Sushiloko Guará II	-15.839377	-47.977919
11	Sushiloko Jardim Botânico	-15.856995	-47.835516
12	Sushiloko Lago Norte	-15.720955	-47.881831
13	Sushiloko Lago Sul	-15.836972	-47.867309
14	Sushiloko Gilberto Salomão	-15.846405	-47.894953
15	Sushiloko Grande Colorado	-15.690321	-47.863538
16	Sushiloko Octogonal	-15.803598	-47.941206
17	Sushiloko Sobradinho	-15.654941	-47.811472
18	Sushiloko Sudoeste	-15.792522	-47.919221
19	Sushiloko Taguatinga Norte	-15.819228	-48.061872
20	Sushiloko Taguatinga Sul	-15.840853	-48.046873
21	Sushiloko Vicente Pires	-15.798902	-48.054287
22	Sushiloko Samambaia	-15.862337	-48.088220
23	Sushiloko Taguatinga Centro	-15.835511	-48.056402
24	Sushiloko Ceilândia JK Shopping	-15.803874	-48.098005
25	Koni Águas Claras	-15.842125	-48.022089
26	Koni Guará II	-15.839704	-47.969468
27	Koni Lago Sul	-15.835506	-47.868070
28	Koni Posto Altana	-15.704002	-47.889698
29	Koni Sudoeste	-15.801420	-47.924294
30	Koni Taguatinga Shopping	-15.841827	-48.044788
31	Koni 201 Norte	-15.783458	-47.878566
32	Koni 109 Norte	-15.759752	-47.887750
33	Koni Park Shopping	-15.833174	-47.955470
34	Koni 209 Sul	-15.822521	-47.904916
35	Koni Brasília Shopping	-15.786702	-47.888989
36	Koni JK Pier Sul	-15.871724	-47.909286
37	Koni Pier 21	-15.818136	-47.874954
38	Koni Taguatinga Pistão Sul	-15.862973	-48.035887
39	Koni Vicente Pires	-15.809465	-48.051680

Quadro 6 – Coordenadas Geográficas da Amostra

Fonte: Elaborado pela autora.

Vale ressaltar que foram consideradas nessa amostra inclusive as unidades que ainda não foram entregues, mas que já estão em construção e possuem CEP disponível no site, que totalizaram duas unidades, a saber – Sushiloko Águas Claras Av. Castanheiras e Sushiloko Ceilândia JK Shopping.

Tal consideração se deve ao fato de que, independentemente da data de entrega do empreendimento, já é certo que haverá um ponto de venda de comida japonesa naquela região. Dessa forma, invalida-se a possibilidade de que essa localidade venha a se tornar uma oportunidade de negócio, conforme os critérios aqui estabelecidos.

Obtidas as coordenadas geográficas de cada franquía da amostra, procedeu-se para o mapeamento dessas franquías no mapa do Distrito Federal. Optou-se por posicionar todos os mapas na seção dos Apêndices, mais especificamente no apêndice intitulado como “Apêndice B – Mapas formados no QGIS”, em orientação no formato paisagem, permitindo-se assim uma visualização mais detalhada das informações contidas nos mapas. Assim, tem-se como resultado o primeiro mapa construído no *software* QGIS, onde é possível observar a distribuição geográfica das franquías da amostra pelo Distrito Federal, conforme representado pelo Mapa 1.

Observa-se uma grande concentração principalmente nas regiões próximas ao centro de Brasília, como é o caso do Plano Piloto, conforme ilustrado de maneira mais detalhada pelo Mapa 2.

Entretanto, é notável também uma presença dessas franquías em cidades satélites, tais como Guará, Taguatinga, Águas Claras e Samambaia, conforme ilustrado de maneira mais detalhada pelo Mapa 3.

Feito o mapeamento das franquías da amostra no QGIS, procedeu-se para a importação da base de dados gerada na seção anterior deste trabalho, onde constam as estimativas de gasto com comida japonesa para cada setor censitário do Distrito Federal. Essa base de dados foi inserida como uma camada vetorial no mapa do Distrito Federal, mapa este disponível no site do IBGE no formato de arquivo específico para georreferenciamento, o *shapefile*.

Por meio do mapeamento das franquías já realizado neste trabalho, juntamente com a inserção e união da base de dados referente ao gasto padronizado, foi possível

constituir um novo mapa com o cruzamento desses dados, capaz de fornecer a localização das franquias no espaço geográfico e as respectivas demandas identificadas em cada setor censitário do DF, conforme ilustra o Mapa 4.

Assim, serão considerados como sendo oportunidade de negócio, todas as regiões em que for observada uma baixa oferta de comida japonesa, mas que apresentem uma alta concentração de indivíduos com gasto padronizado elevado para o consumo de comida japonesa, ou seja, indivíduos considerados como sendo consumidores em potencial.

Para a identificação das oportunidades de negócios, foi realizada uma análise qualitativa, de acordo com os critérios previamente estabelecidos. Para fins de adequação aos critérios, a demanda da região analisada deve ser acima da demanda média do Distrito Federal. Sendo assim, a faixa de gasto padronizado a ser considerada para a análise das oportunidades deverá estar acima de 0.3722, ou seja, estar inclusa em um dos dois últimos intervalos do gasto padronizado.

Naturalmente, um maior gasto padronizado será identificado em regiões residenciais, devido ao fato de que são essas as regiões nas quais os consumidores em potencial residem. Dessa forma, serão selecionadas as regiões mais próximas possíveis das residências do público-alvo.

A análise de oportunidades de negócio foi feita individualmente, por meio de uma análise qualitativa, com os 19 subdistritos que compõe o Distrito Federal identificados no Censo Demográfico de 2010.

Vale ressaltar que os 19 subdistritos em conjunto cobrem toda a extensão geográfica do Distrito Federal. Entretanto, nem todas as Regiões Administrativas do Distrito Federal são descritas individualmente. Isso se deve ao fato de que o mesmo subdistrito pode abranger mais de uma Região Administrativa.

Optou-se por trabalhar com o modelo de subdistritos por se tratar de uma definição geográfica, diferentemente do modelo de Regiões Administrativas, é uma definição política e pode sofrer alterações no decorrer do tempo. Seguem os resultados obtidos para cada subdistrito.

4.1.1 Subdistrito Brasília

Considerando-se os critérios estabelecidos para que uma região seja considerada como uma oportunidade de negócio, tem-se que, para o subdistrito de Brasília, foram identificadas cinco possíveis oportunidades de negócio, conforme ilustra o Mapa 5.

Apesar de Brasília ser um dos subdistritos que mais possuem franquias a saber, onze franquias, ele é também um dos subdistritos onde foram identificadas uma das maiores demandas por comida japonesa. Além disso, o subdistrito de Brasília está privilegiadamente localizado no centro econômico e político do Distrito Federal, o que contribui para um maior fluxo de pessoas na região. Não obstante, Brasília possui uma das maiores extensões de áreas comerciais do Distrito Federal, o que também contribui para a identificação de oportunidades de negócio.

Como dito anteriormente, foram identificadas cinco oportunidades de negócio. Iniciando-se a contagem no sentido Sul-Norte do mapa, a primeira está localizada na altura da Quadra comercial 202 Sul, abaixo do Eixo L Sul. A segunda está localizada na altura da Quadra comercial 303 Sul, próxima a Via W1 Sul. A terceira oportunidade identificada se encontra no Comércio Local Norte (CLN), na altura da Quadra 405 Norte. As últimas duas oportunidades de negócio identificadas em Brasília se encontram no final da Asa Norte, uma na altura da Quadra 513 Norte, próximo à Via W3 Norte, e a outra no Comércio Local do Setor De Habitações Coletivas e Geminadas Norte (SHCGN), na altura da Quadra 715 Norte.

Todas as cinco oportunidades identificadas no subdistrito de Brasília se encontram em áreas comerciais, estão a mais de 1km de distância das franquias identificadas no subdistrito, estão dentro ou próximas a centros comerciais, com estacionamento e fácil acesso ao transporte público, além de estarem dentro ou a no máximo 500 metros de distância de setores censitários cujo valor do gasto padronizado tenha sido superior a 0.3722.

4.1.2 Subdistrito de Brazlândia

Para o subdistrito de Brazlândia, foi identificada apenas uma oportunidade de negócio, no comércio local da Quadra 04 Conjunto C, próximo ao centro da cidade. A

oportunidade identificada se encontra próxima diversos pontos de ônibus e possui disponibilidade de estacionamento. O Mapa 6 ilustra o subdistrito de Brazlândia, onde apesar de ser observada a existência de apenas quatro setores censitários que possuem um gasto padronizado acima de 0.3722, não possui nenhuma franquias localizada neste subdistrito.

4.1.3 Subdistrito Candangolândia

Para o subdistrito de Candangolândia, apenas uma localidade foi identificada como sendo uma possível oportunidade de negócio. Essa oportunidade se mostra viável, principalmente devido ao fato de que não foram identificadas franquias concorrentes nessa região. Além disso, a oportunidade identificada está localizada no Setor de Postos de Motéis Sul (SPMS) Conjunto E, bem próxima às áreas residenciais e também próxima a uma das vias mais movimentadas que cruzam o Distrito Federal, a Via DF-003, também conhecida como EPIA (Estrada Parque Indústria e Abastecimento). O Mapa 7 ilustra as informações apresentadas.

4.1.4 Subdistrito Ceilândia

Para o subdistrito de Ceilândia, duas localidades foram identificadas como sendo uma oportunidade de negócio, conforme ilustra o Mapa 8. Iniciando-se a contagem no sentido Norte-Sul do mapa, a primeira oportunidade está localizada no Setor O, entre as Quadras 15 e 13 e a segunda se encontra no Setor M, entre as Quadras 12 e 18. Apesar de não haver franquias concorrentes identificadas nesse subdistrito propriamente dito, é possível notar a presença delas nas proximidades do subdistrito (vide Mapa 4). Além disso, grande parte dos setores censitários desse subdistrito que possuem a demanda acima da média são áreas residenciais, inviabilizando a possibilidade de identificação de uma oportunidade de negócio na maior parte da extensão geográfica do subdistrito. Ambas as oportunidades identificadas se encontram em áreas comerciais, estão dentro ou próximas a centros comerciais, possuem estacionamento e fácil acesso ao transporte público, especialmente às estações de metrô, além de estarem dentro ou a no máximo 500 metros de distância de setores censitários cuja demanda tenha sido acima da média identificada no Distrito Federal. O Mapa 8 ilustra essas informações.

4.1.5 Subdistrito Cruzeiro

Apesar de o subdistrito Cruzeiro apresentar três unidades de franquias concorrentes, essa região possui um gasto padronizado considerado alto, o que facilita a inserção de novos estabelecimentos de comida japonesa na região. Dessa forma, foram identificadas duas oportunidades de negócio, conforme os critérios anteriormente estabelecidos. Iniciando-se a contagem no sentido Oeste-Leste, a primeira oportunidade se encontra no Comércio Local do Cruzeiro Velho 2, de frente a uma das principais vias de acesso à cidade, além de estar cercada de regiões residenciais, o que facilita o acesso tanto de pessoas que estão passando por aquela região, quanto de pessoas que já residem nas proximidades. O fluxo nessa região também tende a ser alto por estar próxima a uma das vias mais movimentadas do Distrito Federal, a via DF-003, também conhecida como EPIA (Estrada Parque Indústria e Abastecimento).

Além disso, a região também está localizada próximo a um dos maiores centros comerciais do Distrito Federal, o Setor de Indústria e Abastecimento, conhecido como SIA, onde se encontra um grande fluxo de pessoas devido ao forte comércio local, que oferece os mais diversos tipos de produtos e serviços, e abriga também uma das feiras mais movimentadas do Distrito Federal, conhecida como Feira dos Importados. A segunda oportunidade identificada está localizada no Comércio Local próximo à Avenida das Jaqueiras, uma das principais avenidas da XXII Região Administrativa do Distrito Federal, conhecida como Sudoeste, que também está incluída no subdistrito do Cruzeiro. Além disso, é um dos subdistritos mais próximos do centro do Distrito Federal e possui fácil acesso por transporte público. O Mapa 9 demonstra as informações apresentadas.

4.1.6 Subdistrito Gama

O subdistrito Gama já possui uma franquia em seu principal centro comercial. Entretanto, devido ao alto gasto padronizado identificado na região do subdistrito, duas novas localidades foram identificadas como sendo uma oportunidade de negócio, conforme ilustra o Mapa 10. Iniciando-se a contagem no sentido Norte-Sul, a primeira oportunidade está localizada entre as Quadras comerciais 6 e 7, enquanto que a segunda está localizada entre as Quadras 41 e 42 do Gama.

Ambas as oportunidades identificadas se encontram em áreas comerciais, estão dentro ou próximas a centros comerciais, com estacionamento e fácil acesso ao transporte público, especialmente pontos de ônibus, além de estarem dentro ou a no máximo 500 metros de distância de setores censitários cujo valor do gasto padronizado identificado tenha sido superior a 0.3722.

4.1.7 Subdistrito Guar

J existem, nesse subdistrito, trs franquias concorrentes. Uma delas est localizada no maior shopping do Distrito Federal, conhecido como Park Shopping. Entretanto, todas se concentram no mesmo lado da regio. Dessa forma, foram identificadas duas possveis oportunidades de negcio nas outras regies do Guar, conforme representado pelo Mapa 11.

Iniciando-se a contagem no sentido Norte-Sul, a primeira oportunidade de negcio se encontra no Setor Residencial, Indstria e Abastecimento (Sria) I Quadra Externa (QE) 7 do Guar I, no principal centro comercial da regio. Essa rea dispe de estacionamento amplo e fcil acesso ao transporte pblico, especialmente nibus. A segunda oportunidade est localizada no Sria II Quadra Externa (QE) 13, na sada do Guar II, prximo a duas estaes do metr e possui estacionamento pblico disponvel, alm de estar exatamente na rea comercial que faz fronteira com a rea residencial da regio.

4.1.8 Subdistrito Lago Norte

Para o subdistrito Lago Norte, foi identificada uma oportunidade de negcio, localizada no Setor de Habitaes Individuais Norte (SHIN) QI 9, ao lado da Estrada Parque do Lago Norte (EPPN). Esta oportunidade possui uma localizao comercial privilegiada, pois est cercada de quadras residenciais. Alm disso, o local dispe de um amplo estacionamento pblico e possui um posto policial a poucos metros de distncia. Esta oportunidade est localizada em setor censitrio do tipo urbano, em uma rea comercial distante a mais de 1 km das franquias identificadas neste subdistrito e possui um gasto padronizado acima da mdia do Distrito Federal, conforme ilustra o Mapa 12.

4.1.9 Subdistrito Lago Sul

Para o subdistrito Lago Sul, não foram identificadas localidades oportunas para o negócio em nenhum setor censitário. Como pode ser visto no Mapa 13, já existem cinco franquias na região, e todos os outros setores censitários restantes que apresentaram um valor de gasto padronizado acima de 0.3722 e que se encontram a pelo menos 1 km de distância das cinco franquias, estão ou localizados em áreas residenciais, ou o comércio mais próximo está a mais de 500 metros de distância. Além disso, esses setores censitários restantes não possuem fácil acesso por meio de transporte público.

4.1.10 Subdistrito Núcleo Bandeirante

Esse é o subdistrito que se mostra mais promissor para a inserção de novos estabelecimentos de comida japonesa. Como pode ser observado no Mapa 14, não existe nenhuma franquia concorrente na região. Além disso, esse é o subdistrito de maior gasto padronizado com comida japonesa do Distrito Federal. Dessa forma, foi identificada uma possível oportunidade de negócio no centro comercial do Núcleo Bandeirante, mais especificamente na Avenida Central. Esta oportunidade está localizada próxima à principal via de acesso à cidade e também uma das mais movimentadas do Distrito Federal, a Via DF-075, também conhecida como EPNB (Estrada Parque Núcleo Bandeirante). A não identificação de outras oportunidades de negócio ao longo da região se deve ao fato de que a maior extensão geográfica desse subdistrito abriga áreas residenciais, região conhecida como Park Way.

4.1.11 Subdistrito Paranoá

A maior parte da extensão geográfica desse subdistrito abriga zonas rurais, como pode ser visto no Mapa 15. Dessa forma, foi identificado apenas um setor censitário que possui um gasto padronizado acima de 0.3722, ou acima da demanda média do Distrito Federal. Levando-se em consideração os critérios previamente estabelecidos neste trabalho, especialmente o último critério, a saber, o subdistrito analisado deve possuir uma quantidade mínima de três setores censitários que possuam um gasto padronizado acima da média, tem-se que o subdistrito Paranoá não possui nenhuma oportunidade de negócio.

4.1.12 Subdistrito Planaltina

Assim como o Paranoá, o subdistrito de Planaltina apresenta uma vasta extensão territorial em zonas rurais. Observando-se de maneira mais próxima, é possível identificar uma oportunidade de negócio no centro da cidade, localizada na Via NS 1 Quadra 1 Conjunto 4, conforme ilustra o Mapa 16. O local dessa oportunidade dispõe de um amplo estacionamento público e possui um grande fluxo de pessoas por estar localizado a poucos metros da Feira de Planaltina, um dos principais centros comerciais da região. Além disso, dispõe também de fácil acesso por transporte público, possuindo diversas paradas de ônibus nas proximidades.

4.1.13 Subdistrito Recanto das Emas

O subdistrito Recanto das Emas apresentou o gasto padronizado mínimo estabelecido nos critérios de seleção deste trabalho em apenas um setor censitário do subdistrito, conforme representado pelo Mapa 17. Dessa forma, tem-se que o empreendimento de um estabelecimento de comida japonesa nesse subdistrito é contra indicado, devido ao fato da região por possuir uma baixa demanda desse tipo de produto. Portanto, de acordo com o sétimo e último critério de seleção, não foram identificadas oportunidades de negócios nessa região.

4.1.14 Subdistrito Riacho Fundo

Da mesma forma do subdistrito anterior, o subdistrito Riacho Fundo não apresentou oportunidades de negócio possíveis na região, considerando-se os critérios de seleção de oportunidades previamente estabelecidos neste trabalho. Foram observados apenas dois setores censitários onde o gasto padronizado observado foi acima da média registrada no Distrito Federal, como pode ser observado no Mapa 18.

4.1.15 Subdistrito Samambaia

Assim como os subdistritos Planaltina e Paranoá, o subdistrito de Samambaia possui uma vasta extensão territorial de zonas rurais e já possui uma franquias localizada próximo ao principal centro comercial da cidade, conforme ilustra o Mapa

19. Entretanto, foi identificada uma oportunidade de negócio na região que cumprem os critérios previamente estabelecidos. Esta oportunidade está localizada próximo ao subdistrito de Taguatinga, mais especificamente entre o Setor L Norte Quadra 616 e a Avenida Leste Quadra 416. Está localizado também próximo ao Hospital de Samambaia e possui estacionamento público disponível, além do fácil acesso ao transporte público, dispendo de várias paradas de ônibus ao seu redor.

4.1.16 Subdistrito Santa Maria

Assim como os subdistritos Paranoá, Recanto das Emas e Riacho Fundo, o subdistrito de Santa Maria só apresentou o gasto padronizado mínimo estabelecido nos critérios deste trabalho em apenas dois setores censitários, conforme representado pelo Mapa 20. Dessa forma, não foram identificadas oportunidades de negócio nessa região.

4.1.17 Subdistrito São Sebastião

A oportunidade de negócio encontrada nesse subdistrito se encontra próximo ao centro comercial da cidade, mais especificamente na Avenida Comercial, e dispõe de diversos pontos de ônibus ao longo de toda sua extensão. Apesar do subdistrito de São Sebastião possuir setores censitários que apresentam um valor de gasto padronizado correspondente a uma demanda considerada muito alta, a identificação de apenas uma oportunidade de negócio se deve ao fato de que a maior extensão territorial desse subdistrito corresponde a setores censitários do tipo rural, composto por diversos núcleos rurais, ou está contido em áreas residenciais que não apresentam comércio nas proximidades, como é o caso do condomínio residencial Jardins Mangueiral, em São Sebastião, conforme ilustra o Mapa 21.

4.1.18 Subdistrito Sobradinho

No subdistrito de Sobradinho, onde está localizada apenas uma franquia, foi identificada uma oportunidade de negócio próxima ao centro da cidade, como pode ser observado no Mapa 22. Ela está localizada entre as Quadras 10 e 11 de Sobradinho e está próxima à biblioteca pública de Sobradinho e a diversas quadras

esportivas. Além disso, a região dispõe de estacionamento público e fácil acesso por meio de transporte público, especificamente ônibus.

4.1.19 Subdistrito Taguatinga

O subdistrito de Taguatinga abrange em sua extensão as cidades de Águas Claras e Vicente Pires. Nessa região, foi registrada a presença onze franquias no total, o que representa pouco mais de 25% da amostra utilizada neste trabalho, conforme pode ser observado no Mapa 23. Apesar da alta concorrência, a região apresenta um elevado gasto padronizado e algumas localidades ainda não atendidas pelas franquias atuais, o que levou à identificação de duas possíveis oportunidades de negócio.

Iniciando-se a contagem no sentido Oeste-Leste, a primeira oportunidade está localizada no Setor J Norte de Taguatinga, no comércio local entre as Quadras 3 e 9 e dispõe de estacionamento público. A segunda oportunidade está localizada no comércio local da Rua número 4, na cidade de Vicente Pires. O local dispõe de um amplo estacionamento público e possui um posto policial a poucos metros de distância.

Assim, considerando os critérios de seleção previamente estabelecidos, tem-se que para toda a extensão geográfica do Distrito Federal, foram identificadas 23 oportunidades de negócio, representadas em conjunto no Mapa 24.

Com a conclusão da etapa de identificação de oportunidades, foi possível constatar a aplicação do conceito de Geomarketing estabelecido anteriormente no Referencial Teórico deste trabalho, onde Geomarketing foi definido como sendo um conjunto de técnicas que permite a visualização e análise de informações de mercado relevantes sob uma ótica geográfica, através da utilização de ferramentas específicas para o processamento de dados geocodificados, auxiliando os gestores na elaboração de estratégias, na leitura espacial de mercado e, conseqüentemente, na tomada de decisões.

Analisando de forma mais detalhada este conceito no contexto deste trabalho, temos que (1) a visualização e análise de informações de mercado relevantes sob uma ótica geográfica se deu justamente por meio da obtenção do gasto padronizado

observado para cada setor censitário do Distrito Federal, (2) através da utilização de ferramentas específicas para o processamento de dados geocodificados, que no contexto deste trabalho foi o *software* QGIS 2.4, (3) auxiliando os gestores na elaboração de estratégias, na leitura espacial de mercado e, conseqüentemente, na tomada de decisões por meio principalmente das oportunidades de negócio que foram identificadas neste trabalho.

Baviera-Puig et al. (2013) apresentaram em seus estudos, conforme visto anteriormente, os conceitos de *geodemand* e *geocompetition*, ou, em tradução livre, geodemanda e geocompetição, respectivamente. A geodemanda (*geodemand*) é definida como sendo a localização em mapas digitais dos consumidores de um produto ou serviço em um mercado específico. Já a geocompetição (*geocompetition*) é definida como sendo a localização em mapas digitais dos concorrentes de mercado, bem como o delineamento de suas áreas de alcance (*trade area*) em um mercado específico.

De maneira análoga, o presente trabalho apresenta como sua geodemanda justamente o gasto padronizado identificado em cada setor censitário do Distrito Federal, enquanto que a geocompetição, por sua vez, refere-se ao mapeamento das unidades de franquias utilizadas como amostra. A área de alcance considerada neste trabalho foi de 1 km de distância, conforme estabelecido nos critérios para a seleção de uma oportunidade de negócio. Assim, foram identificadas oportunidades de negócio justamente nas regiões que apresentavam, simultaneamente, um alto nível de geodemanda e um baixo nível de geocompetição.

Dessa forma, é possível constatar que este estudo apresenta consonância com a literatura revisada. A principal diferença entre o presente trabalho e os demais estudos analisados diz respeito, especialmente, ao método empregado para a obtenção dos dados georreferenciados referentes ao perfil de consumo local dos indivíduos. O presente trabalho se baseou na proposição de uma nova abordagem que pudesse viabilizar a obtenção de dados mais precisos, por meio da utilização de Máquinas de Suporte Vetorial aplicadas ao Geomarketing.

5 CONCLUSÕES E RECOMENDAÇÕES

5.1 Principais Conclusões do Estudo

Como principal conclusão deste trabalho, tem-se a confirmação de que é possível construir uma ferramenta quantitativa que possa auxiliar a Tomada de Decisão do Gestor quanto ao local de abertura de seu negócio utilizando Máquinas de Suporte Vetorial e dados secundários. Para se atingir tal conclusão, fez-se necessário o cumprimento de etapas específicas que comprovem a possibilidade de construção dessa ferramenta.

A primeira etapa envolveu a escolha do ramo de mercado a ser utilizado como objeto de estudo. O presente trabalho utilizou para a aplicação do método o mercado de comida japonesa, por ser um dos mercados com um maior crescimento no segmento da alimentação nos últimos anos no Brasil, conforme visto anteriormente na introdução deste trabalho.

Posteriormente, para uma melhor compreensão técnica e teórica do trabalho, foram apresentados os principais conceitos, histórico e noções introdutórias sobre os campos de estudo do Geomarketing e de Aprendizado de Máquina, que são os campos pilares para o desenvolvimento do trabalho. Foram apresentadas também as principais definições e metodologia da Pesquisa de Orçamentos Familiares, realizada pelo IBGE, cujos resultados constituíram-se como principal fonte de dados deste trabalho.

Nos procedimentos metodológicos, selecionaram-se os produtos que compuseram a Cesta de Produtos Japoneses e, posteriormente foram selecionadas as variáveis que poderiam influenciar o consumo de comida japonesa. Feito isso, procedeu-se para o treinamento da máquina e posterior aplicação do método no mercado de comida japonesa do Distrito Federal. Após o cruzamento dos dados obtidos referentes ao gasto padronizado por setor censitário, bem como o mapeamento das franquias que já atuam na região, foram identificadas 23 possíveis oportunidades de negócio para o mercado de comida japonesa no Distrito Federal. O Quadro 7 resume a metodologia utilizada e as informações obtidas, de acordo com os objetivos específicos propostos neste trabalho.

OBJETIVO ESPECÍFICO	MÉTODO	RESULTADO
Elaborar uma cesta com os produtos relacionados à culinária japonesa identificados na Pesquisa de Orçamentos Familiares 2008-2009.	Seleção dos principais produtos utilizados na culinária japonesa que foram identificados na base de dados CADERNETA DESPESA, que compõe parte dos resultados da Pesquisa de Orçamentos Familiares 2008-2009	Cesta de Produtos onde constam apenas produtos específicos da culinária japonesa que foram consumidos pelos brasileiros no período compreendido entre 2008 e 2009.
Selecionar, a partir da Pesquisa de Orçamentos Familiares 2008-2009, as variáveis que estariam relacionadas com o consumo de comida japonesa.	Análise das características sociodemográficas dos indivíduos que consumiram os alimentos que compõe a Cesta de Produto Japoneses.	Relação de variáveis sociodemográficas que influenciam ou estão relacionadas ao consumo de comida japonesa no Brasil.
Programar a leitura dos dados da Pesquisa de Orçamentos Familiares 2008-2009 e a construção de Máquinas de Suporte Vetorial a ser utilizada para amostra do Distrito Federal.	Programação, utilizando o <i>software</i> RStudio, da leitura dos dados da Pesquisa de Orçamentos Familiares 2008-2009 e da construção de Máquinas de Suporte Vetorial.	Constituição do método de aprendizado de máquina capaz de prever, com base na Pesquisa de Orçamentos Familiares 2008-2009, a demanda por comida japonesa no Distrito Federal.
Aplicar o método proposto ao mercado de comida japonesa do Distrito Federal para a identificação das regiões mais propensas ao consumo desses produtos.	Aplicação do método constituído, no mercado de comida japonesa do Distrito Federal para prever, com base nos dados do Censo Demográfico de 2010, o gasto padronizado em cada setor censitário.	Previsão dos gastos padronizados que refletem a demanda por comida japonesa em cada um dos setores censitários do Distrito Federal.
Estimar a demanda por comida japonesa para cada setor censitário do Distrito Federal por meio da previsão de valores de gasto padronizado gerada pelo aprendizado de máquina;	Utilização dos gastos padronizados obtidos na etapa anterior para classificar a demanda por comida japonesa em cada setor censitário do Distrito Federal, atribuindo-se um nível de demanda de acordo com intervalos de valores específicos de gasto padronizado.	Visualização espacial dos níveis de demanda por comida japonesa no mapa do Distrito Federal em cada setor censitário, considerando-se o gasto padronizado médio observado no Distrito Federal.
Identificar os locais com oportunidades de negócio no Distrito Federal a partir da análise da localização das franquias utilizadas como amostra, dos resultados obtidos pela Máquina de Suporte Vetorial e dos dados da Pesquisa de Orçamentos Familiares 2008-2009	Mapeamento, por meio de coordenadas geográficas, das franquias utilizadas na amostra deste trabalho e aplicação dos gastos padronizados previstos para cada setor censitário no mapa do Distrito Federal, utilizando para ambas as ações o <i>software</i> QGIS.	Cruzamento das informações obtidas a partir da localização das franquias e do gasto padronizado previsto, identificando regiões que não possuam franquias próximas, mas que possuam uma alta demanda por comida japonesa, ou seja, regiões caracterizadas neste trabalho como sendo oportunidades de negócio.

Quadro 7 – Quadro Resumo dos Objetivos Específicos

Fonte: Elaborado pela autora.

Dessa forma, tem-se que a justificativa teórica proposta inicialmente para o desenvolvimento deste trabalho, a saber – contribuir cientificamente para um campo de estudo pouco explorado, especialmente no Brasil – foi alcançada por meio,

principalmente, da revisão de literatura e dos procedimentos metodológicos realizados neste trabalho. A justificativa prática proposta, a saber – propor uma abordagem capaz de auxiliar os gestores na tomada de decisão no que diz respeito à localização de seu negócio – também foi contemplada, na medida em que o gestor dispõe das informações necessárias à execução dessa abordagem.

Primeiramente, o gestor deve dispor de dados sociodemográficos referentes à região na qual pretende atuar e investigar quais são as variáveis que podem justificar o consumo do produto ou serviço que pretende ofertar. No caso deste trabalho, os dados sociodemográficos utilizados referentes ao Distrito Federal foram provenientes do Censo Demográfico de 2010, realizado pelo IBGE. Além disso, é necessário também possuir uma base de dados referente ao consumo do produto ou serviço o qual o gestor pretende ofertar. No caso deste trabalho, a base de dados utilizada para obter as informações referentes ao consumo de comida japonesa no Brasil e, posteriormente, no Distrito Federal, foi Pesquisa de Orçamentos Familiares de 2008-2009, também realizada pelo IBGE. Além disso, é importante também obter o máximo de informações possíveis sobre os concorrentes diretos do produto ou serviço a ser ofertado. No contexto deste trabalho foram obtidas informações referentes aos fatores demográficos, socioeconômicos, padrões de consumo, caracterização econômica do setor e localizações geográficas das franquias concorrentes. Entretanto, outras informações tais como benefícios procurados, estilos de vida, personalidade, estratégias de divulgação, público-alvo, promoções e *feedback* de clientes, também são alguns fatores relevantes que podem colaborar para a elaboração de uma pesquisa de mercado mais completa. Uma vez que o gestor dispõe desses dados, é possível, por meio das técnicas aqui apresentadas, adaptá-las ao seu contexto e construir uma ferramenta que o auxilie na tomada de decisão no que diz respeito à localização de seu negócio.

5.2 Limitações do Estudo e Sugestões para Trabalhos Futuros

Uma das principais limitações desse estudo diz respeito ao fato de que a amostra de franquias considerada é limitada. Apesar de serem as duas maiores franquias de *fast-food* de comida japonesa atuantes no Distrito Federal, existe a possibilidade de que haja outros restaurantes de comida japonesa que não foram considerados na amostra desse estudo, o que poderia comprometer os resultados.

Dessa forma, para que os resultados obtidos tornem-se mais confiáveis, sugere-se que os trabalhos futuros nessa área levem em consideração não apenas as franquias de comida japonesa, mas também os restaurantes de comida japonesa que atuam no Distrito Federal.

Outra limitação do estudo diz respeito à Cesta de Produtos Japoneses, que se apresentou muito restrita para o Distrito Federal na POF de 2008-2009. Tal restrição pode ser justificada pelo fato de que durante os anos de realização da POF 2008-2009, uma das franquias utilizada na amostra era recém-chegada à Brasília, enquanto que a outra franquia, que é brasiliense, havia sido criada no ano anterior, em 2007.

De uma forma geral, durante os anos de 2008 e 2009 o mercado de comida japonesa ainda estava conquistando seu espaço no mercado brasileiro de alimentação. Dessa forma, os dados da POF 2008-2009 podem não refletir de forma fidedigna a realidade atual de consumo de comida japonesa.

Vale ressaltar que até o presente momento, esta é a Pesquisa de Orçamentos Familiares mais recente disponibilizada pelo IBGE. Uma Pesquisa de Orçamentos Familiares mais recente foi realizada nos anos de 2012 e 2013, mas sem previsão de divulgação até o momento.

Dessa forma, outra sugestão para trabalhos futuros diz respeito justamente à utilização de uma base de dados mais atualizada, que possa representar de uma maneira mais fiel à realidade de consumo atual no mercado de comida japonesa.

Outra limitação importante a ser considerada envolve as oportunidades de negócio apresentadas nos resultados. Não foram avaliados neste estudo as questões relacionadas à viabilidade econômica ou física de implementação de um novo estabelecimento de comida japonesa nas regiões sugeridas. Como a análise envolvida nesta etapa foi qualitativa, com base nos dados fornecidos pelo *Google Maps* e pelo *Google Satellite*, é possível que a região indicada como uma oportunidade de negócio esteja indisponível. Portanto, deve-se considerar nesses casos a possibilidade de implementação em outra região que esteja disponível e mais próxima possível do local indicado.

REFERÊNCIAS

- AAKER, D. A. **Strategic market management**. New York: John Wiley & Sons, 2001.
- ABCJ - ASSOCIAÇÃO BRASILEIRA DE CULINÁRIA JAPONESA. Disponível em: <<http://www.baressp.com.br>>. Acesso em: 17/03/2014.
- ABIA – ASSOCIAÇÃO BRASILEIRA DAS INDÚSTRIAS DA ALIMENTAÇÃO. Disponível em: <<http://www.abia.org.br/>>. Acesso em: 17/03/2014.
- ALCAIDE, J. C., CALERO, R., HERNÁNDEZ, R. **Geomarketing. Marketing territorial para vender y fidelizar más**. Madrid: ESIC, 2012.
- ANDERSON, C., VINCZE, J. W. **Strategic marketing management**. New York: Houghton Mifflin, 2000.
- ANGLUIN, D., SMITH, C. H. Inductive inference: Theory and methods. **ACM Computing Surveys (CSUR)**, v. 15, n. 3, p. 237-269, 1983.
- ARANHA, F., FIGOLI, S. Geomarketing: memórias de viagem. **São Paulo**, p. 1-73, 2001.
- ARNOLD, U. New dimensions of outsourcing: a combination of transaction cost economics and the core competencies concept. **European Journal of Purchasing & Supply Management**, Alemanha, v. 6, n. 1, p. 23-29, 2000.
- BRASIL. Lei nº 12.255, de 15 de junho de 2010. Dispõe sobre o salário mínimo a partir de 1º de janeiro de 2010, estabelece diretrizes para a política de valorização do salário mínimo entre 2012 e 2023 e revoga a Lei no 11.944, de 28 de maio de 2009. **Diário Oficial da República Federativa do Brasil**, Brasília, DF, 16 jun. 2010. Seção I, p.1.
- CARDOSO, C. E. P. **Geomarketing como suporte de Decisão em Gestão do Território**. 84 f. Dissertação (Especialização em Detecção Remota e Sistemas de Informação Geográfica) – Faculdade de Ciências Sociais e Humanas, Universidade Nova Lisboa, Lisboa, 2011.
- CENTER FOR SPATIALLY INTEGRATED SOCIAL SCIENCE. Disponível em: <<http://www.csiss.org/>>. Acesso em: 16/05/2013.

CHAUVEL, M. A. **Consumidores insatisfeitos: uma oportunidade para as empresas**. Rio de Janeiro: Mauad, 2000.

CHEN, D., ODOBEZ, J. M. Comparison of support vector machine and neural network for text texture verification. **IDIAP**, Suíssa, 2002.

CHRISTALLER, Walter. **Central places in Southern Germany**. New Jersey: Prentice-Hall, 1966.

CLIQUET, G. **Geomarketing: methods and strategies in spatial marketing**. Londres: ISTE, 2006.

CUI, D., CURRY, D. Prediction in marketing using the support vector machine. **Marketing Science**, v. 24, n. 4, p. 595-615, 2005.

DOUARD. J. P. Geomarketing and Consumer Behavior. In: CLIQUET, G. **Geomarketing: Methods and Strategies in Spatial Marketing**. Londres: ISTE, 2006. p. 91-113.

DU, Ke-Lin; SWAMY, M. N. S. Fundamentals of Machine Learning. **Neural Networks and Statistical Learning**, Londres, p. 15-65, 2014.

EVERITT, B., SKRONDAL, A. **The Cambridge dictionary of statistics**. Cambridge: Cambridge University Press, 2002.

HAYKIN, S. **Neural Networks – A Comprehensive Foundation**. 2. Ed. Ontario: Pearson Prentice Hall, 1999.

HEARST, M. Support vector machines. **Intelligent Systems and their Applications, IEEE**, v. 13, n. 4, p. 18-28, 1998.

HUANG, Jih-Jeng; TZENG, Gwo-Hshiung; ONG, Chornng-Shyong. Marketing segmentation using support vector clustering. **Expert systems with applications**, v. 32, n. 2, p. 313-317, 2007.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. **Brasil 500 anos. Território Brasileiro e Povoamento – Japoneses – Razões da Emigração Japonesa**. Disponível em: <http://cod.ibge.gov.br/210N6>. Acesso em 18 de Março de 2014.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. **Censo Demográfico de 2010: características da população e dos domicílios**. Rio de Janeiro: IBGE, 2011.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. **Pesquisa de Orçamentos Familiares 2002-2003: primeiros resultados**. Rio de Janeiro: IBGE, 2004.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. **Pesquisa de Orçamentos Familiares 2008-2009: despesas, rendimentos e condições de vida**. Rio de Janeiro: IBGE, 2010.

LATOURE, P.; LE FLOC'H, J. **Géomarketing: principes, méthodes et applications**. França: Editions d'Organisation, 2001.

LORENA, A. C., DE CARVALHO, A. C. P. L. F. Uma introdução às support vector machines. **Revista de Informática Teórica e Aplicada**, São Paulo, v. 14, n. 2, p. 43-67, 2007.

LÖSCH, A. **Economics of location**. New Haven e Londres: Yale University Press, 1954.

MASANO, T. F. et al. **Gestão de marketing**. São Paulo: Saraiva, 2006.

MINISTÉRIO DA PESCA E AQUICULTURA. **Boletim Estatístico da Pesca e Aquicultura – Brasil - 2010**. Brasília. Disponível em: <<http://www.mpa.gov.br>>. Acesso em: 20/04/2013.

MINISTÉRIO DA PESCA E AQUICULTURA. **Boletim Estatístico da Pesca e Aquicultura – Brasil - 2011**. Brasília. Disponível em: <<http://www.mpa.gov.br>>. Acesso em: 20/04/2013.

MITCHELL, T. M. **Machine Learning**. Portland: McGraw-Hill, 1997.

MONARD, M. C., BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. **Sistemas Inteligentes-Fundamentos e Aplicações**, p. 89-114, 2003.

ORTIGOZA, S. A. O fast food e a mundialização do gosto. **Cadernos de Debate**, São Paulo, v. 5, p. 21-45, 1997.

PRATI, R. C., BARANAUSKAS, J. A.; MONARD, M. C. Extração de informações padronizadas para a avaliação de regras induzidas por algoritmos de aprendizado de máquina simbólico. **Relatório Técnico do ICMC**, São Carlos, v. 145, 2001.

RICHERS, R., LIMA, C. P. **Segmentação**. São Paulo: Nobel, 1991.

RODRIGUE, J. P., COMTOIS, C., SLACK, B. **The geography of transport systems**. New York: Routledge, 2013.

RODRIGUES, A. L.; CALGARO, C.; PEREIRA, A. O. K. Conquistando Direitos: Ascensão Da Nova Classe Média No Brasil E Aumento Do Consumo. In: **Congresso de Pesquisa e Extensão da Faculdade da Serra Gaúcha**. 2014. p. 98-109.

ROIG-TIERNO, N., BAVIERA-PUIG, A., BUITRAGO-VERA, J. Business opportunities analysis using GIS: the retail distribution sector. **Global Business Perspectives**, Valência, v. 1, n. 3, p. 226-238, 2013.

SEMOLINI, R., ZUBEN, F. J. **Support vector machines, inferência transdutiva eo problema de classificação**. 142 f. Dissertação (Mestrado em Engenharia Elétrica) - Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, São Paulo, 2002.

SHIN, H. J., CHO, S. Response modeling with support vector machines. **Expert Systems with Applications**, v. 30, n. 4, p. 746-760, 2006.

SMITH, C. ESRI ArcGIS Business Analyst: An Assessment of its value for Spatial Marketing Research. **Fall**, 2007.

STANFORD UNIVERSITY. **Machine Learning**. Disponível em: <<http://openclassroom.stanford.edu/>>. Acesso em: 24/05/2014.

TAKETA, R. Management and the geographer: The relevance of geography in strategic thinking. **The Professional Geographer**, Texas, v. 45, n. 4, p. 465-470, 1993.

THE UNIVERSITY OF CHICAGO. Disponível em: <<http://www.uchicago.edu/>>. Acesso em: 16/05/2014.

THE WOLF AT THE DOOR. Disponível em: <<http://www.wolfatthedoor.org.uk>>.

Acesso em: 16/05/2013.

VAPNIK, V. **Statistical learning theory**. New York: John Wiley & Sons, 1998.

VAPNIK, V. **The Nature of Statistical Learning**. New York: Springer, 1995.

VON THÜNEN, J. H. **The Isolated State**. Oxford: Pergamon Press, 1966.

WANG, H. Y. F.; IKEDA, A. A. Análise do mercado de cartão de crédito brasileiro. **VII SEMEAD**, São Paulo, Brasil, 2005.

WEBER, A. **Theory of Location of Industries**. Chicago: University of Chicago Press, 1957.

YRIGOYEN, C. C. El geomarketing y la distribución comercial. **Investigación y Marketing**, Madrid, n. 79, p. 6-13, 2003.

APÊNDICES

Apêndice A – Programação utilizada no RStudio

```

#Limpa o workspace
rm(list=ls())

#Local dos arquivos
setwd("C:\\Users\\Camila\\Dados")

#Importa os dados
consumo.df<-read.csv("T_CADERNETA_DESPESA_S.csv")

#Primeiras observações
head(consumo.df)

#Guarda a base na memoria
attach(consumo.df)

#Confere o quadro
table(NUM_QUADRO)

#Lista de produtos de comida japonesa
japonesa<-paste(PROD_NUM_QUADRO_GRUPO_PRO,COD_ITEM, sep = "")

#Quem consumiu ganha o valor, quem não consumiu ganha zero
gastocomidajaponesa<-
ifelse(japonesa%in%c("631802","859201","703604","703607","8323
02","703605","703601","793402","703603","775901","854201","859
401","771406","854202","703606","706801","631801","774402","67
6001","673102","705502","776106","7010506","7010503","7010504"
,"7010505","774401"),VAL_DESPESA_CORRIGIDO,0 )

#Estatísticas descritivas do gasto de comida japonesa no
Brasil
summary(gastocomidajaponesa[which(gastocomidajaponesa>0)],weig
ht=FATOR_EXPANSAO1)
hist(gastocomidajaponesa[which(gastocomidajaponesa>0)])
boxplot(gastocomidajaponesa[which(gastocomidajaponesa>0)])

#Aplica o peso no gasto
gastocomidajaponesaponderado<-
gastocomidajaponesa*FATOR_EXPANSAO1

#Calcula o gasto total com comida japonesa por pessoa

```

```

japonesa.df<-
aggregate(gastocomidajaponesaponderado,by=list(COD_UF,NUM_SEQ,
NUM_DV),sum)

#Remove a base da memoria
detach(consumo.df)

#Dá nome as colunas
colnames(japonesa.df)<-c("COD_UF","NUM_SEQ","NUM_DV","GASTO")

#Primeiras observações da base por Setor Censitário (4690
setores)
head(japonesa.df)

#Cria as variáveis de característica do Setor Censitário com
base no arquivo de morador

#Importa os dados
morador.df<-read.csv("T_MORADOR_S.csv")

#Primeiras observações
head(morador.df)

#Guarda a base na memoria
attach(morador.df)

#Cria as variáveis
idade0a9<-ifelse(IDADE_ANOS>=0 &
IDADE_ANOS<9,1,0)*FATOR_EXPANSAO1
idade10a19<-ifelse(IDADE_ANOS>=10 &
IDADE_ANOS<19,1,0)*FATOR_EXPANSAO1
idade20a29<-ifelse(IDADE_ANOS>=20 &
IDADE_ANOS<29,1,0)*FATOR_EXPANSAO1
idade30a39<-ifelse(IDADE_ANOS>=30 &
IDADE_ANOS<39,1,0)*FATOR_EXPANSAO1
idade40a49<-ifelse(IDADE_ANOS>=40 &
IDADE_ANOS<49,1,0)*FATOR_EXPANSAO1
idade50a59<-ifelse(IDADE_ANOS>=50 &
IDADE_ANOS<59,1,0)*FATOR_EXPANSAO1
idade60a69<-ifelse(IDADE_ANOS>=60 &
IDADE_ANOS<69,1,0)*FATOR_EXPANSAO1
idade70acima<-ifelse(IDADE_ANOS>=70 &
IDADE_ANOS<99,1,0)*FATOR_EXPANSAO1
homem<-ifelse(COD_SEXO==1,1,0)*FATOR_EXPANSAO1
mulher<-ifelse(COD_SEXO==2,1,0)*FATOR_EXPANSAO1
RENDA_PERCAPITA<-as.numeric(RENDA_PERCAPITA)
renda1<-ifelse(RENDA_PERCAPITA>=63.75,1,0)*FATOR_EXPANSAO1
renda2<-ifelse(RENDA_PERCAPITA<63.75 &
RENDA_PERCAPITA>=127.5,1,0)*FATOR_EXPANSAO1
renda3<-ifelse(RENDA_PERCAPITA<127.5 &
RENDA_PERCAPITA>=255,1,0)*FATOR_EXPANSAO1

```

```

renda4<-ifelse(RENDA_PERCAPITA<255 &
RENDA_PERCAPITA>=510,1,0)*FATOR_EXPANSAO1
renda5<-ifelse(RENDA_PERCAPITA<510 &
RENDA_PERCAPITA>=1020,1,0)*FATOR_EXPANSAO1
renda6<-ifelse(RENDA_PERCAPITA<1020 &
RENDA_PERCAPITA>=1530,1,0)*FATOR_EXPANSAO1
renda7<-ifelse(RENDA_PERCAPITA<1530 &
RENDA_PERCAPITA>=2550,1,0)*FATOR_EXPANSAO1
renda8<-ifelse(RENDA_PERCAPITA<2550 &
RENDA_PERCAPITA>=5100,1,0)*FATOR_EXPANSAO1
renda9<-ifelse(RENDA_PERCAPITA<5100,1,0)*FATOR_EXPANSAO1
alfabetizado<-ifelse(COD_SABE_LER==1,1,0)*FATOR_EXPANSAO1
analfabeto<-ifelse(COD_SABE_LER==2,1,0)*FATOR_EXPANSAO1
branca<-ifelse(COD_COR_RACA==1,1,0)*FATOR_EXPANSAO1
preta<-ifelse(COD_COR_RACA==2,1,0)*FATOR_EXPANSAO1
amarela<-ifelse(COD_COR_RACA==3,1,0)*FATOR_EXPANSAO1
parda<-ifelse(COD_COR_RACA==4,1,0)*FATOR_EXPANSAO1
indigena<-ifelse(COD_COR_RACA==5,1,0)*FATOR_EXPANSAO1
naosaberaca<-ifelse(COD_COR_RACA==9,1,0)*FATOR_EXPANSAO1

#Remove a base da memoria
detach(morador.df)

#Agrega os resultados para as variáveis criadas
setores.df<-
aggregate(cbind(idade0a9,idade10a19,idade20a29,idade30a39,idad
e40a49,idade50a59,idade60a69,idade70acima,homem,mulher,RENDA_P
ERCAPITA,renda1,renda2,renda3,renda4,renda5,renda6,renda7,rend
a8,renda9,alfabetizado,analfabeto,branca,preta,amarela,parda,i
ndigena,naosaberaca),by=list(morador.df$COD_UF,morador.df$NUM_
SEQ,morador.df$NUM_DV),sum)

#Dá nome as colunas
colnames(setores.df)<-
c("COD_UF","NUM_SEQ","NUM_DV","Idade0a9","Idade10a19","Idade20
a29","Idade30a39","Idade40a49","Idade50a59","Idade60a69","Idad
e70a99","Homem","Mulher","RENDA_PERCAPITA","Renda1","Renda2","
Renda3","Renda4","Renda5","Renda6","Renda7","Renda8","Renda9",
"alfabetizado","analfabeto","branca","preta","amarela","parda"
,"indigena","naosaberaca")

#Une as bases
tudo.df<-merge(japonesa.df,setores.df,by=c
("COD_UF","NUM_SEQ","NUM_DV"))

#Retira quem naosaberaca, RENDA_PERCAPITA e analfabeto
tudo.df<-tudo.df[,-32]
tudo.df<-tudo.df[,-15]
tudo.df<-tudo.df[,-25]

head(tudo.df)

```

```

### Comeca a análise SVM ###

#habilita o pacote kernlab
library(kernlab)

### Separa a base em duas partes: Treinamento (70%) e
Validação (30%) ###

#Fixa a semente
set.seed(23238)

#Cria a dummy para separar a base
sorteio<-rbinom(nrow(tudo.df),1,0.7)

#Retira as variáveis de controle
tudoLimpa.df<-tudo.df[,-c(1:3)]

#Base de treinamento
treina.df<-tudoLimpa.df[which(sorteio==1),]
treina.df<-scale(treina.df)

#Retira as Rendas de 2 a 8
treina.df<-treina.df[,-c(13:19)]

#Base de Validação
valida.df<-tudoLimpa.df[which(sorteio!=1),]
valida.df<-as.data.frame(scale(valida.df))

#Retira as Rendas de 2 a 8
valida.df<-valida.df[,-c(13:19)]

#Cria a lista de parâmetros para a busca
listaC<-seq(0.01, 10, length.out = 50)
listaSigma2<- c(seq(0.01, 10, length.out = 25),seq(0.01, 10,
length.out = 25))
parms <- expand.grid(C = listaC, sigma =listaSigma2)

#str(valida.df)

#Para cada par de parâmetros pegue aquele com o menor EQM
apply_pb <- function(X, MARGIN, FUN, ...)
{
  env <- environment()
  pb_Total <- sum(dim(X)[MARGIN])
  counter <- 0
  pb <- txtProgressBar(min = 0, max = pb_Total,
                      style = 3)

  wrapper <- function(...)
  {
    curVal <- get("counter", envir = env)

```

```

    assign("counter", curVal +1 ,envir= env)
    setTxtProgressBar(get("pb", envir= env),
                      curVal +1)
  FUN(...)
}
res <- apply(X, MARGIN, wrapper, ...)
close(pb)
res
}

EQM<-apply_pb(parms,1,function(x){ svm<-
ksvm(GASTO~.,data=treina.df, kernel="rbfdot",type="eps-svr",
kpar=list(sigma=as.numeric(x[2])),C=as.numeric(x[1]),scaled
=FALSE); mean((valida.df$GASTO-predict(svm,valida.df)^2))})

#Junta todos os resultados
parms$EQM<-EQM

#Obtém os parâmetros de menor valor
summary(parms$EQM)
hist(parms$EQM)
iMin<-which(parms$EQM==min(parms$EQM))

#Resultado final
parms[iMin[1],]
C<-as.numeric(parms[iMin[1],][1])
sigma<-as.numeric(parms[iMin[1],][2])
#sigma<-ceiling(sigma)

#C<-1
#sigma<-0.01
library(kernlab)
svm<-ksvm(GASTO~.,data=treina.df, kernel="rbfdot",type="eps-
svr", kpar=list(sigma=sigma),C=C,scaled =FALSE)

#Cria a matriz de pontos
library(reshape2)
z<-dcast(parms,parms[,1]~parms[,2],value.var="EQM")
z<-as.matrix(z[,-1])

#Faz o gráfico
library(plot3D)
library(grDevices)
M <- mesh(listaC,listaSigma2)
surf3D(M$x, M$y, z, colvar = z, facets = FALSE,
      bty="b2", main="Erro Quadrático Médio",phi = 30, theta
= -30,
      xlab="Parâmetro C",ylab= "Parâmetro
sigma2",zlab="EQM",
      colkey = list(length = 0.2, width = 0.4, shift = 0.05,

```

```

        cex.axis = 0.8, cex.clab = 0.85),
    clab = c("Valor do EQM"))

bmp(filename = "Grafico1.bmp",width = 480, height = 480, units
= "px", pointsize = 12,bg = "white")

#Faz os gráficos
par(mfrow = c(3, 3), mar = c(0, 0, 0, 0))
surf3D(M$x, M$y, z, colvar = z,bty="b2", facets = FALSE,phi =
30, theta = -30,
      xlab="Parametro C",ylab= "Parametro
sigma2",zlab="EQM")
surf3D(M$x, M$y, z, colvar = z,bty="b2", facets = FALSE,phi =
20, theta = -20,
      xlab="Parametro C",ylab= "Parametro
sigma2",zlab="EQM")
surf3D(M$x, M$y, z, colvar = z,bty="b2", facets = FALSE,phi =
10, theta = -10,
      xlab="Parametro C",ylab= "Parametro
sigma2",zlab="EQM")
surf3D(M$x, M$y, z, colvar = z,bty="b2", facets = FALSE,phi =
60, theta = -60,
      xlab="Parametro C",ylab= "Parametro
sigma2",zlab="EQM")
surf3D(M$x, M$y, z, colvar = z,bty="b2", facets = FALSE,phi =
40, theta = -40,
      xlab="Parametro C",ylab= "Parametro
sigma2",zlab="EQM")
surf3D(M$x, M$y, z, colvar = z,bty="b2", facets = FALSE,phi =
50, theta = -50,
      xlab="Parametro C",ylab= "Parametro
sigma2",zlab="EQM")
surf3D(M$x, M$y, z, colvar = z,bty="b2", facets = FALSE,phi =
60, theta = -60,
      xlab="Parametro C",ylab= "Parametro
sigma2",zlab="EQM")
surf3D(M$x, M$y, z, colvar = z,bty="b2", facets = FALSE,phi =
40, theta = -40,
      xlab="Parametro C",ylab= "Parametro
sigma2",zlab="EQM")
surf3D(M$x, M$y, z, colvar = z,bty="b2", facets = FALSE,phi =
50, theta = -50,
      xlab="Parametro C",ylab= "Parametro
sigma2",zlab="EQM")
bmp(filename = "Grafico2.bmp",width = 480, height = 480, units
= "px", pointsize = 12,bg = "white")
dev.off()

#Remove o que não interessa
#rm(list= ls()[!(ls() %in%
c('parms','tudoLimpa.df','treina.df','valida.df'))])

```

```

#Salva os resultados
save.image("Resultados10092014.RData")

#CENSO
setwd("C:\\Users\\Camila\\CENSO 2010\\Base informações
setores2010 universo DF\\CSV")

#Cria a base para previsão.
censo.df<-data.frame()

#Verifica quantas variáveis tem na base
scan("Pessoa01_DF.csv",sep=';', what="character" , nlines=1 )

#Cria a classe cada variável
class<-rep("character",88)

#Importa os dados
alfabetizado<-read.csv2("Pessoa01_DF.csv",colClasses=class )
alfabetizado<-alfabetizado[,c(1,3:79)]

#Código do Setor Censitário
SETOR<-alfabetizado[,1]

#Função para conversão de character para numeric
asNumeric <- function(x) as.numeric(as.character(x))
charNumeric <- function(d) modifyList(d, lapply(d[, sapply(d,
is.character)], asNumeric))
factorsNumeric <- function(d) modifyList(d, lapply(d[,
sapply(d, is.factor)], asNumeric))

#Alfabetização
alfabetizado<-rowSums(charNumeric(alfabetizado[,2:78]),na.rm =
T)

#Junta
censo.df<-as.data.frame(SETOR)
censo.df$alfabetizado<-alfabetizado
str(censo.df)

#Verifica quantas variáveis tem na base
scan("Pessoa13_DF.csv",sep=';', what="character" , nlines=1 )

#Cria a classe cada variável
class<-rep("character",137)

#Importa os dados
idade<-read.csv2("Pessoa13_DF.csv",colClasses=class )

#Código do Setor Censitário
SETOR<-idade[,1]

```

```

#Idade
idade0a9<-rowSums(charNumeric(idade[,c(24:45)]),na.rm = T)
idade10a19<-rowSums(charNumeric(idade[,c(46:55)]),na.rm = T)
idade20a29<-rowSums(charNumeric(idade[,c(56:65)]),na.rm = T)
idade30a39<-rowSums(charNumeric(idade[,c(66:75)]),na.rm = T)
idade40a49<-rowSums(charNumeric(idade[,c(76:85)]),na.rm = T)
idade50a59<-rowSums(charNumeric(idade[,c(86:95)]),na.rm = T)
idade60a69<-rowSums(charNumeric(idade[,c(96:105)]),na.rm = T)
idade70acima<-rowSums(charNumeric(idade[,c(105:136)]),na.rm =
T)

temp<-as.data.frame(SETOR)
temp$idade0a9<-idade0a9
temp$idade10a19<-idade10a19
temp$idade20a29<-idade20a29
temp$idade30a39<-idade30a39
temp$idade40a49<-idade40a49
temp$idade50a59<-idade50a59
temp$idade60a69<-idade60a69
temp$idade70acima<-idade70acima
str(temp)

#Junta com a base censo.df
censo.df<-merge(censo.df,temp,by="SETOR")

#### Homem
#Verifica quantas variaveis tem na base
scan("Pessoa11_DF.csv",sep=';', what="character" , nlines=1 )

#Cria a classe cada variável
class<-rep("character",137)

#Importa os dados
homem<-read.csv2("Pessoa11_DF.csv",colClasses=class )

#Código do Setor Censitário
SETOR<-homem[,1]

#Idade
homem<-rowSums(charNumeric(homem[,24:136]),na.rm = T)

tempH<-as.data.frame(SETOR)
tempH$homem<-homem

#Junta com a base censo.df
censo.df<-merge(censo.df,tempH,by="SETOR")

```

```

#### Mulher
#Verifica quantas variaveis tem na base
scan("Pessoa12_DF.csv",sep=';', what="character" , nlines=1 )

#Cria a classe cada variável
class<-rep("character",137)

#Importa os dados
mulher<-read.csv2("Pessoa12_DF.csv",colClasses=class )

#Código do Setor Censitário
SETOR<-mulher[,1]

#Idade
mulher<-rowSums(charNumeric(mulher[,24:136]),na.rm = T)
tempm<-as.data.frame(SETOR)
tempm$mulher<-mulher

#Junta com a base censo.df
censo.df<-merge(censo.df,tempm,by="SETOR")

#### Renda

#Verifica quantas variáveis tem na base
scan("DomicilioRenda_DF.csv",sep=';', what="character" ,
nlines=1 )

#Cria a classe cada variável
class<-rep("character",17)

#Importa os dados
renda<-read.csv2("DomicilioRenda_DF.csv",colClasses=class )

#Código do Setor Censitário
SETOR<-renda[,1]
temp<-as.data.frame(SETOR)

renda1<-renda[, 7]
renda2<-renda[, 8]
renda3<-renda[, 9]
renda4<-renda[,10]
renda5<-renda[,11]
renda6<-renda[,12]
renda7<-renda[,13]
renda8<-renda[,14]
renda9<-renda[,15]

temp$renda1<-renda1
#temp$renda2<-renda2
#temp$renda3<-renda3
#temp$renda4<-renda4

```

```

#temp$renda5<-renda5
#temp$renda6<-renda6
#temp$renda7<-renda7
#temp$renda8<-renda8
temp$renda9<-renda9

#Junta com a base censo.df
censo.df<-merge(censo.df,temp,by="SETOR")

### Raça

#Verifica quantas variáveis tem na base
scan("Pessoa03_DF.csv",sep=';', what="character" , nlines=1 )

#Cria a classe cada variável
class<-rep("character",254)

#Importa os dados
raca<-read.csv2("Pessoa03_DF.csv",colClasses=class )
branca<-raca[,4]
preta<-raca[,5]
amarela<-raca[,6]
parda<-raca[,7]
indigena<-raca[,8]

#Código do Setor Censitário
SETOR<-raca[,1]
temp<-as.data.frame(SETOR)

temp$branca<-branca
temp$preta<-preta
temp$amarela<-amarela
temp$parda<-parda
temp$indigena<-indigena

#Junta com a base censo.df
censo.df<-merge(censo.df,temp,by="SETOR")
#Nomes nas bases iguais
names(valida.df)
names(censo.df)

names(censo.df)<-
c("SETOR","Idade0a9","Idade10a19","Idade20a29","Idade30a39","I
dade40a49",

"Idade50a59","Idade60a69","Idade70a99","Homem","Mulher","Renda
1","Renda9","alfabetizado","branca","preta","amarela",
"parda", "indigena")

```

```

#Converte as variaveis para numeric
censo.df$Renda1<-as.numeric(censo.df$Renda1)
#censo.df$Renda2<-as.numeric(censo.df$Renda2)
#censo.df$Renda3<-as.numeric(censo.df$Renda3)
#censo.df$Renda4<-as.numeric(censo.df$Renda4)
#censo.df$Renda5<-as.numeric(censo.df$Renda5)
#censo.df$Renda6<-as.numeric(censo.df$Renda6)
#censo.df$Renda7<-as.numeric(censo.df$Renda7)
#censo.df$Renda8<-as.numeric(censo.df$Renda8)
censo.df$Renda9<-as.numeric(censo.df$Renda9)
censo.df$alfabetizado<-as.numeric(censo.df$alfabetizado)
censo.df$branca<-as.numeric(censo.df$branca)
censo.df$preta<-as.numeric(censo.df$preta)
censo.df$amarela<-as.numeric(censo.df$amarela)
censo.df$parda<-as.numeric(censo.df$parda)
censo.df$indigena<-as.numeric(censo.df$indigena)
censo.df[is.na(censo.df)] <- 0

#Faz a previsão do gasto com comida japonesa
str(censo.df)
temp<-scale(censo.df[,-1])
censoNorm.df<-data.frame(censo.df$SETOR,temp)

censoNorm.df$GASTO<-predict(svm,censoNorm.df)
summary(censoNorm.df$GASTO)

#censo.df$GASTO<-predict(svm,censo.df)

##### ANALISE ESPACIAL #####

setwd("C:\\Users\\Camila\\Dados")

#Faz o mapa
library(maptools)
library(rgdal)

#Lê as malhas
bsb <- readOGR(".", "BSB")
setor <- readOGR(".", "SETOR")

#Faz o merge
setor.spdf <-
merge(setor,censoNorm.df,by.x="CD_GEOCODI",by.y="censo.df.SETO
R")
names(setor.spdf)

#Pinta o mapa
library(RColorBrewer)

#Pinta o MAPA
gasto <- setor.spdf$GASTO

```

```

gasto[is.na(gasto)] <- 0

#Definindo o número de classes:
nClasses <- 5

#Definindo as cores do mapa
tCores = brewer.pal(nClasses,"Spectral")

#Gera os intervalos
library(classInt)
classe<-classIntervals(gasto,nClasses,style="quantile")

#Atribuindo cada cor a sua respectiva classe
colcode <- findColours(classe, tCores)

#Pinta o mapa
plot(setor.spdf, col=colcode, border ="NA", axes = TRUE)
plot(bsb, col=NA, border=grey(.5), lwd=1, add=TRUE)
title(main="Gasto com Comida
Japonesa",xlab="Longitude",ylab="Latitude")

#Define a rosa dos ventos
source("Legenda e Rosa dos Ventos.r")
northarrow(c(2,3),0.5)

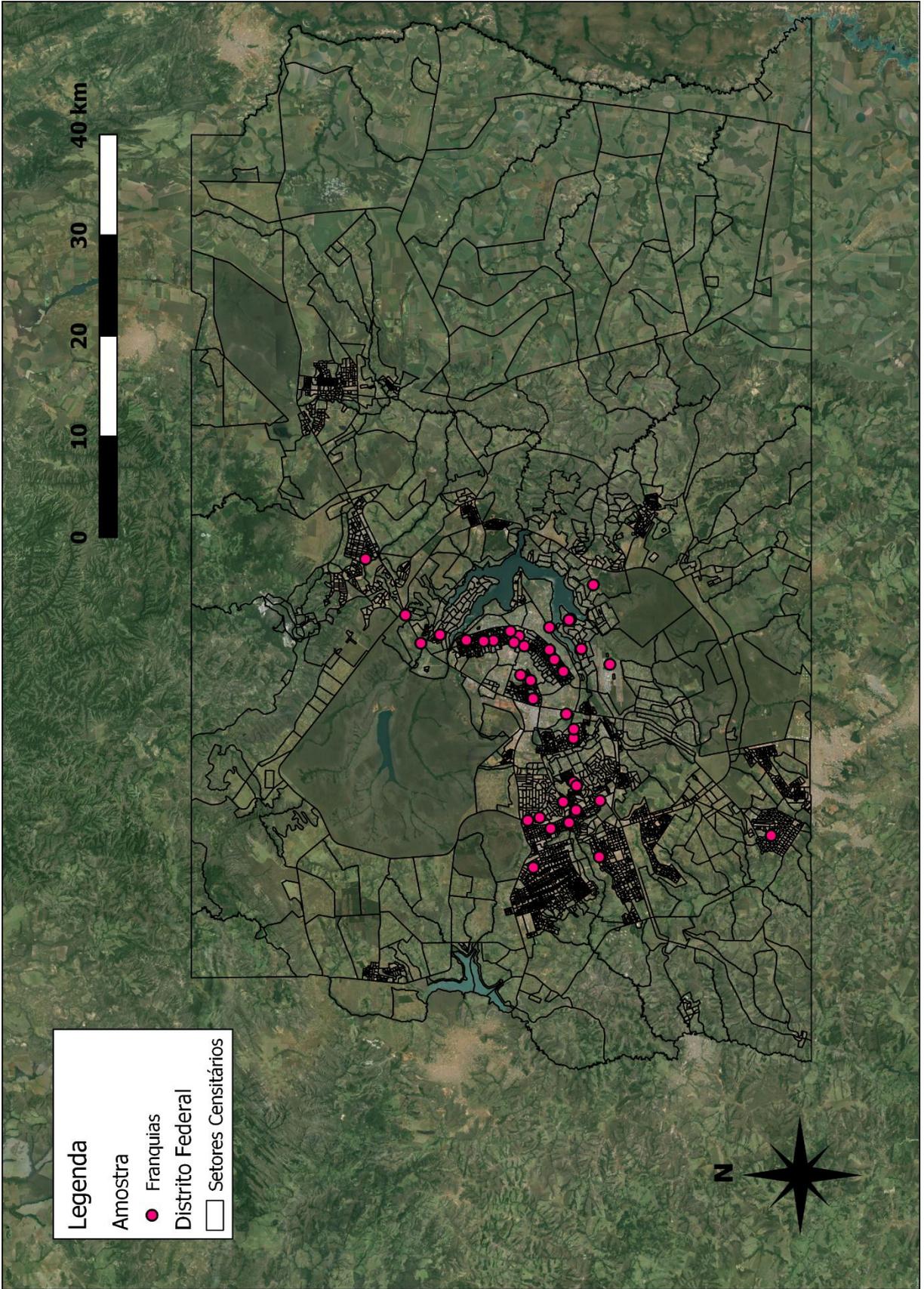
#Legenda
legend(10,15, legend=names(attr(colcode,
"table")),fill=attr(colcode, "palette"), cex=0.6,bty="n",
title="Quantil do Gasto")

#Salvando o mapa
windows()
savePlot(filename = "MapaGasto",type = c("bmp"), device
=dev.cur(),restoreConsole = TRUE)
library(shapefiles)
write.shp(setor.spdf, 'MapaGasto.shp')
#write.csv(setor.spdf, file = 'MapaGasto.csv')

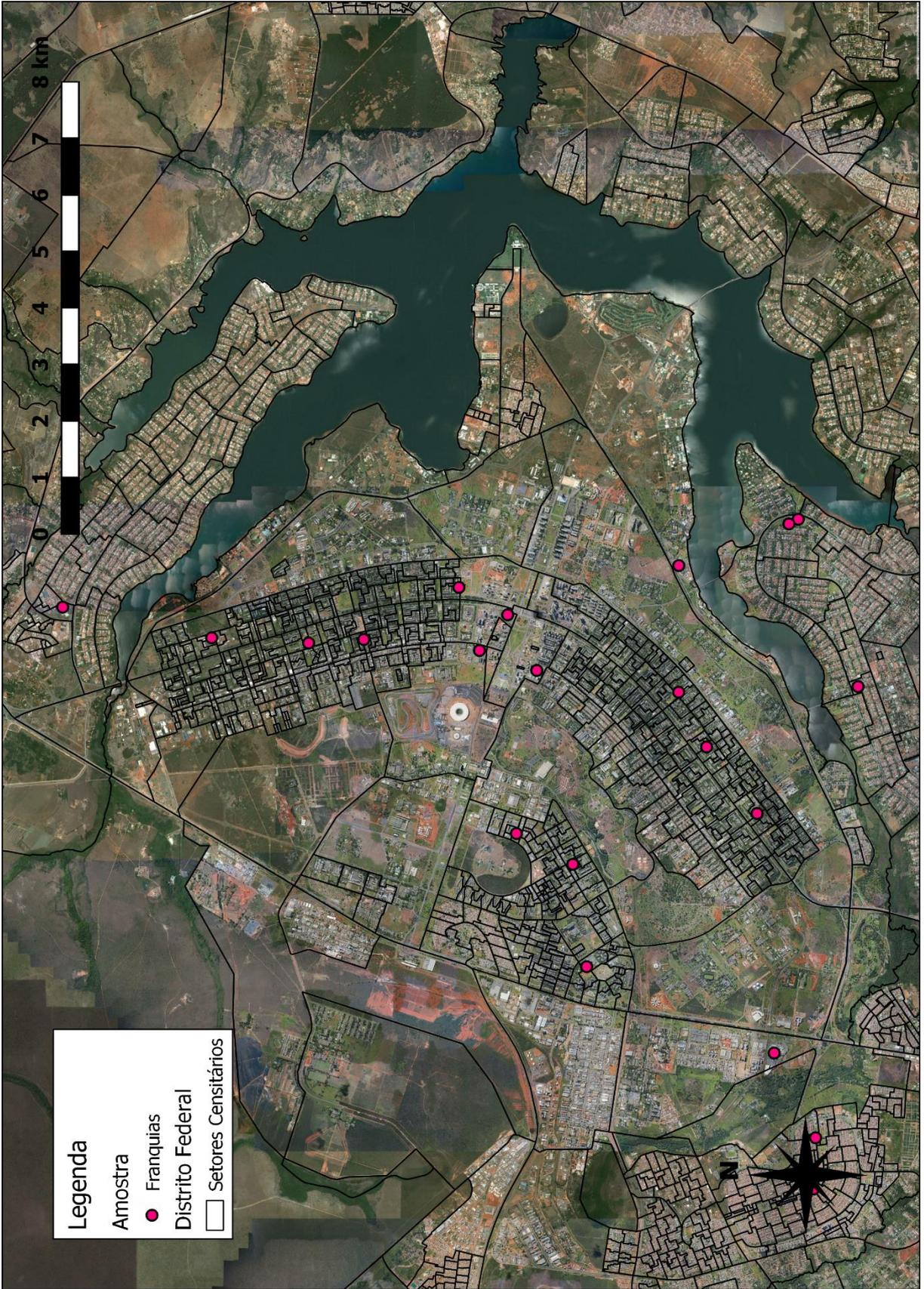
```

Apêndice B – Mapas formados no QGIS

Mapa 1 – Franquias no Distrito Federal



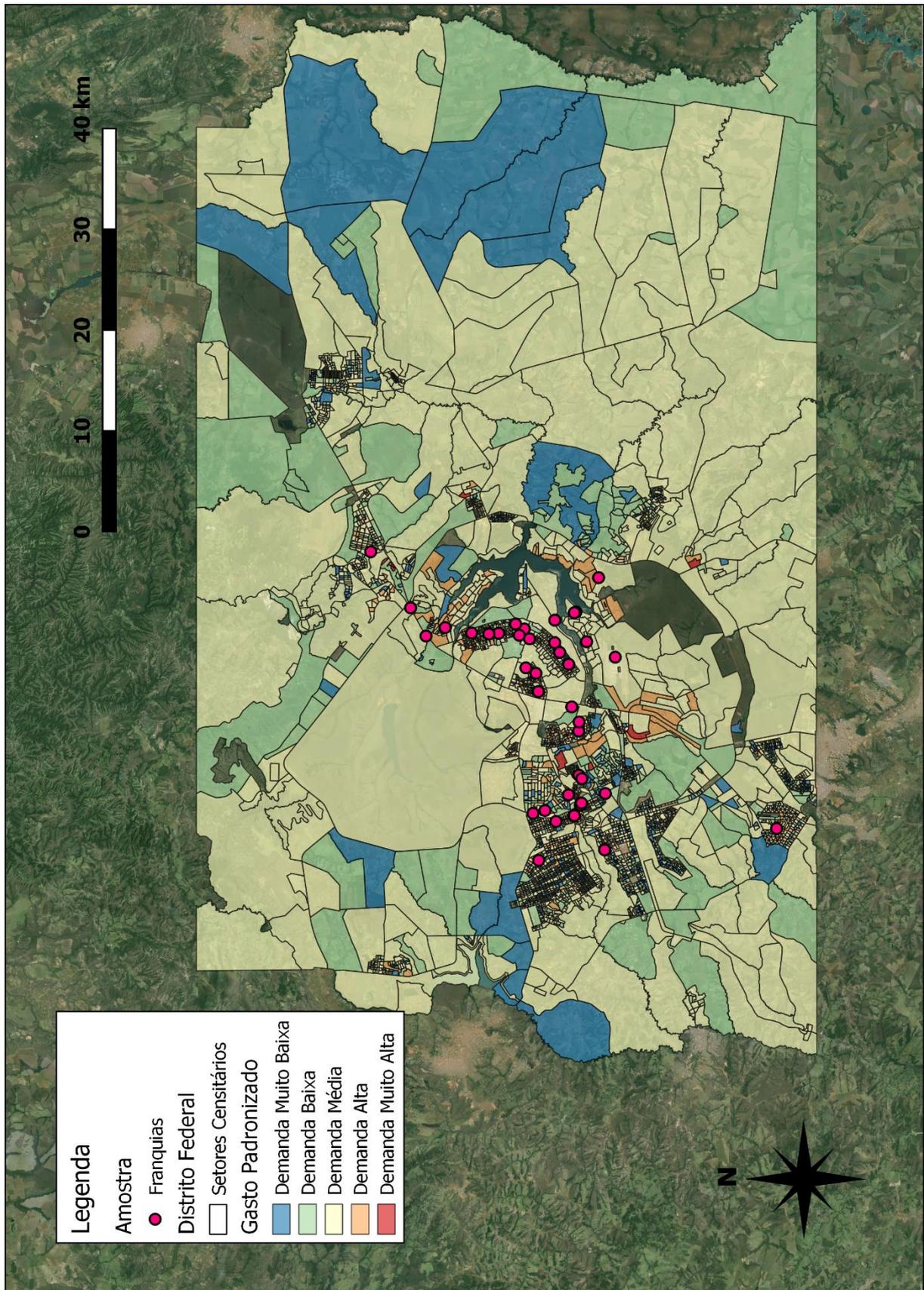
Mapa 2 – Franquias no Plano Piloto



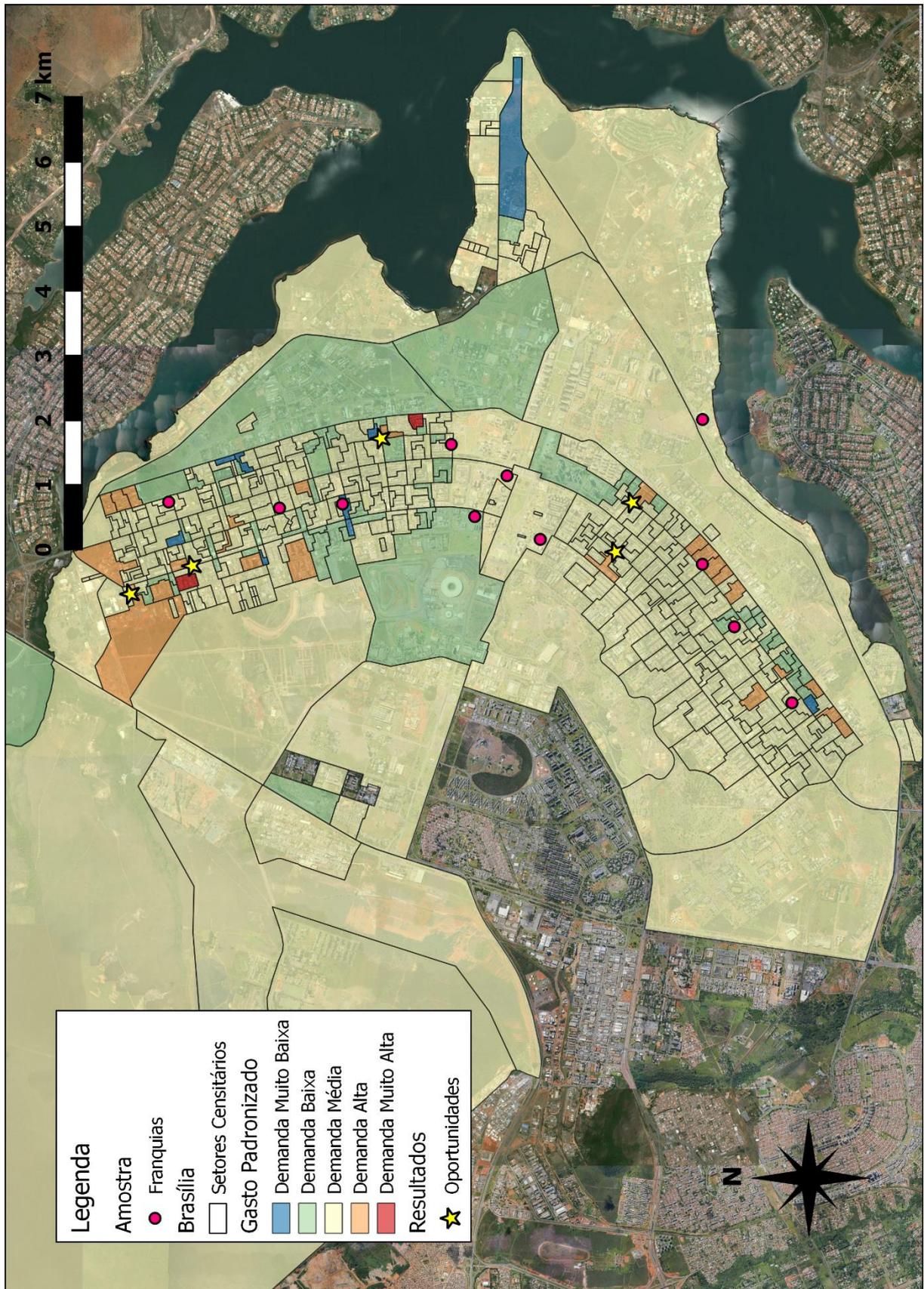
Mapa 3 – Franquias nas Cidades Satélites



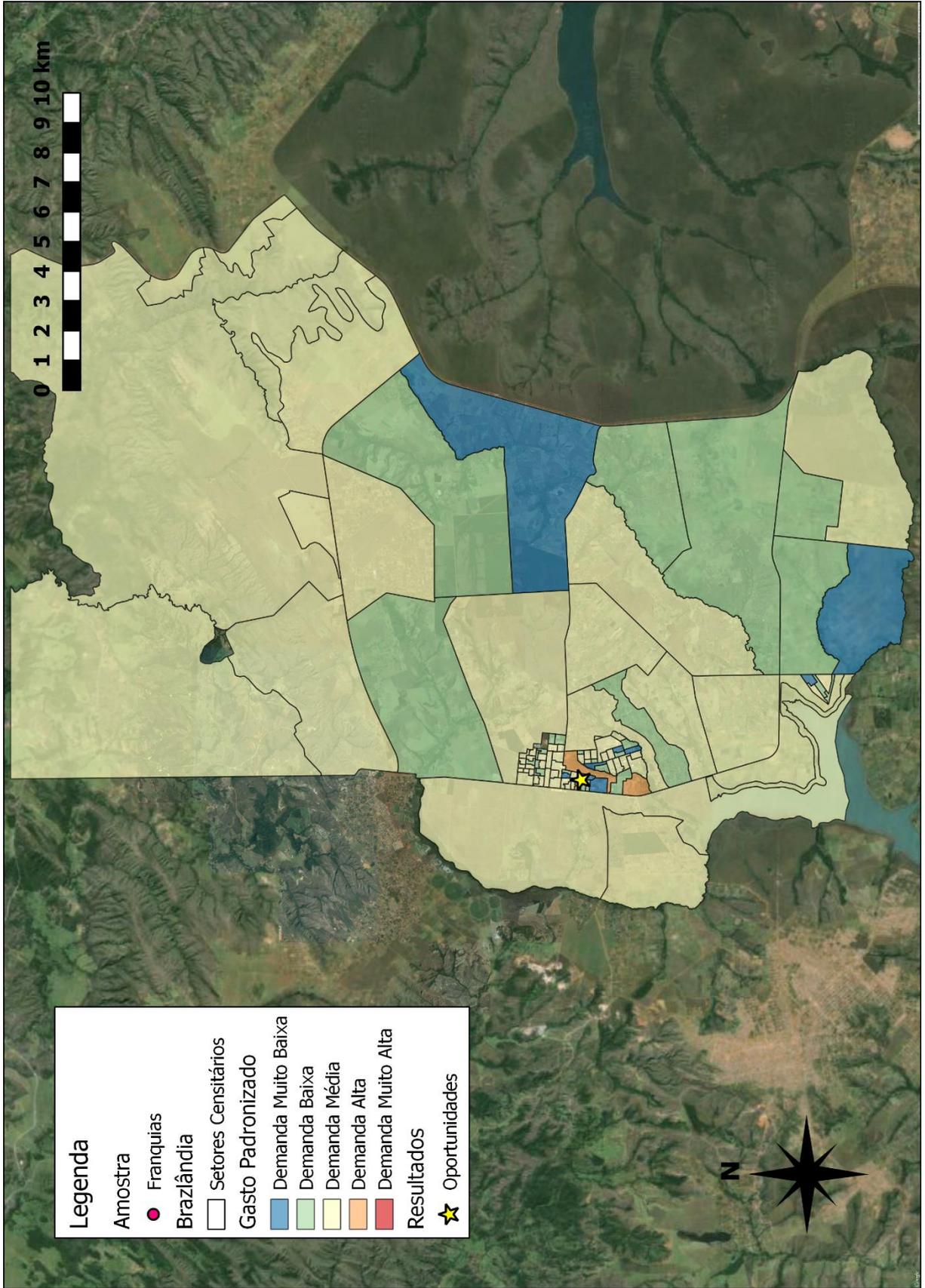
Mapa 4 – Franquias e Gasto no Distrito Federal



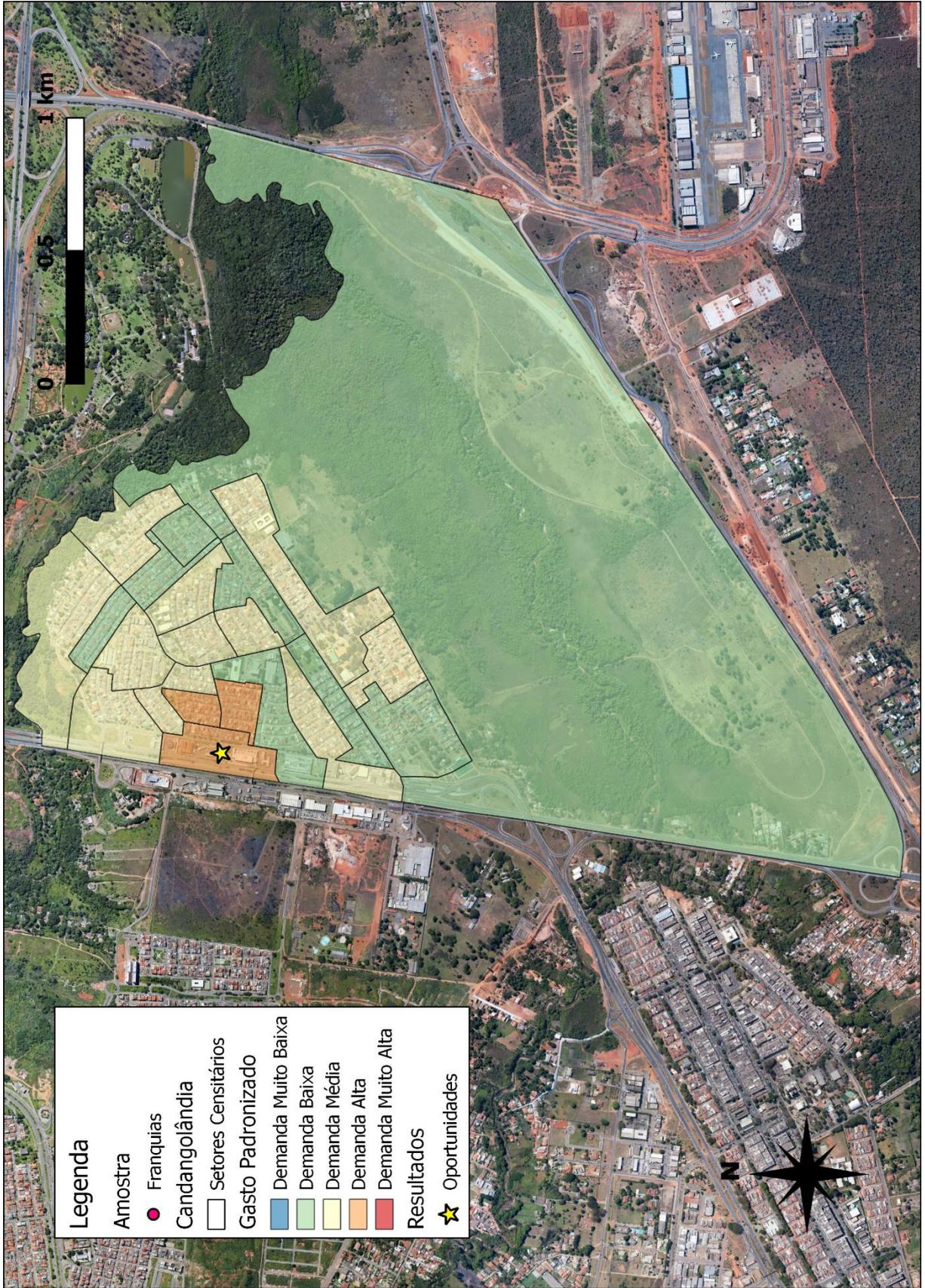
Mapa 5 – Oportunidades de Negócio em Brasília



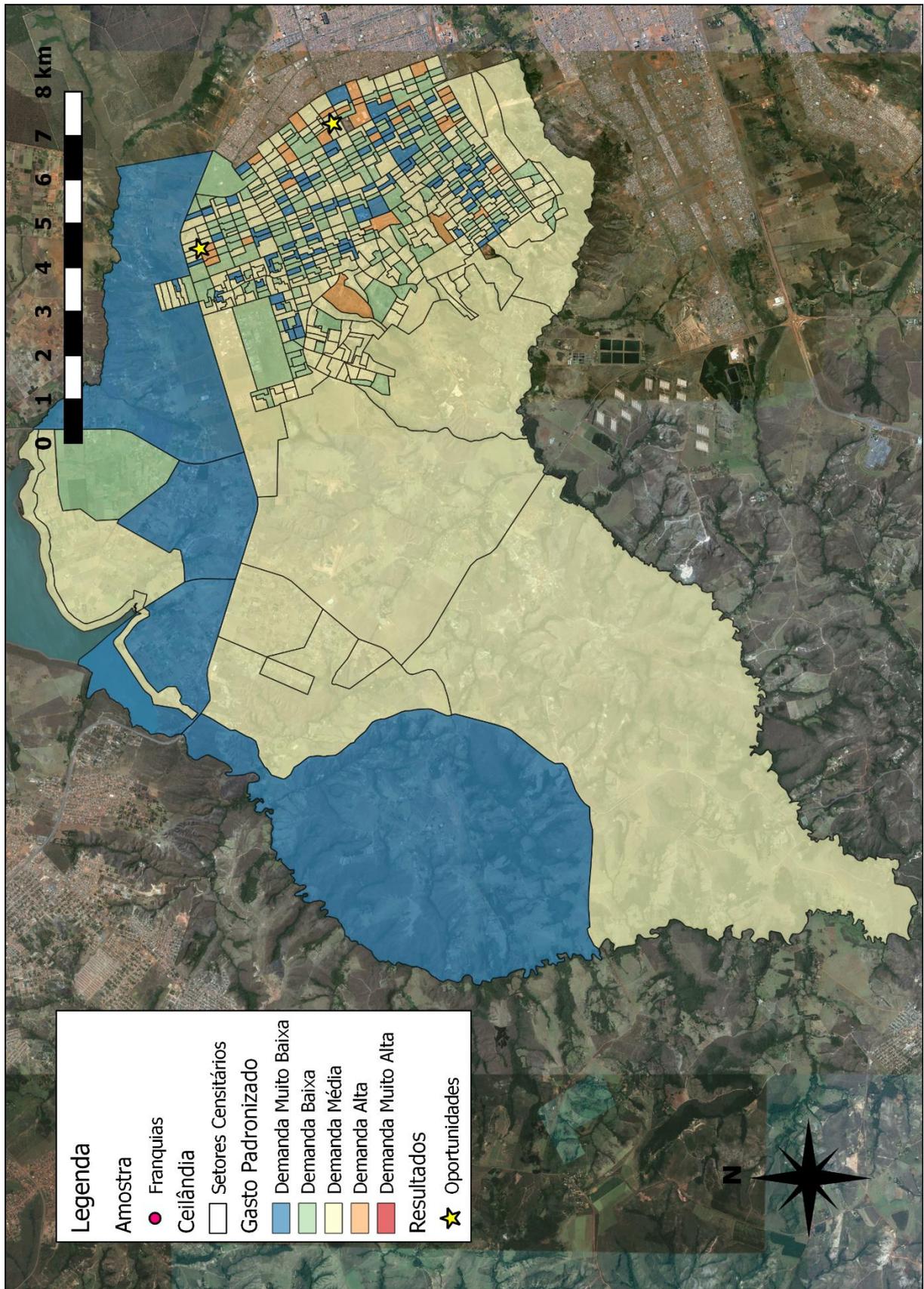
Mapa 6 – Oportunidades de Negócio em Brazlândia



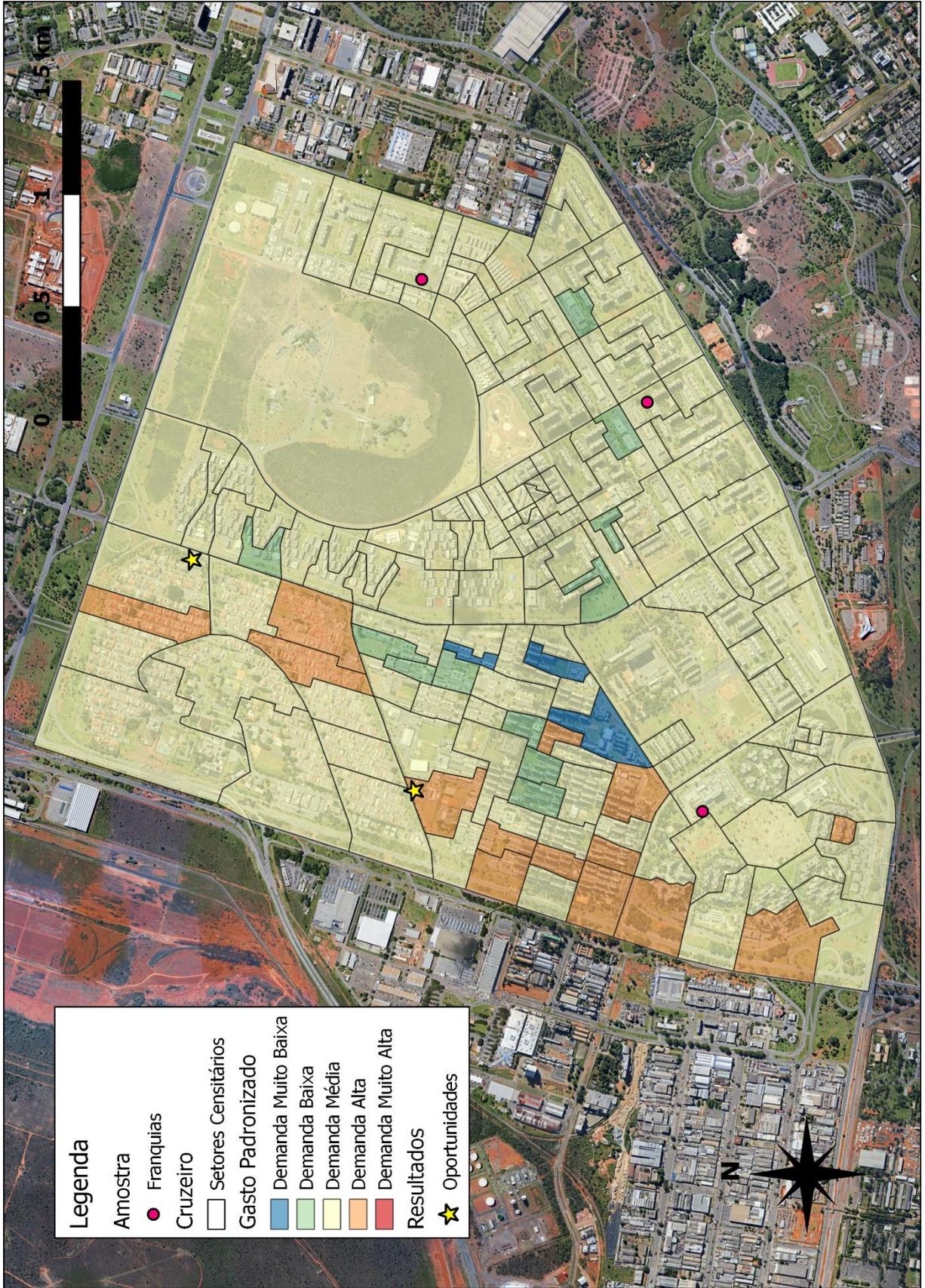
Mapa 7 – Oportunidades de Negócio na Candangolândia



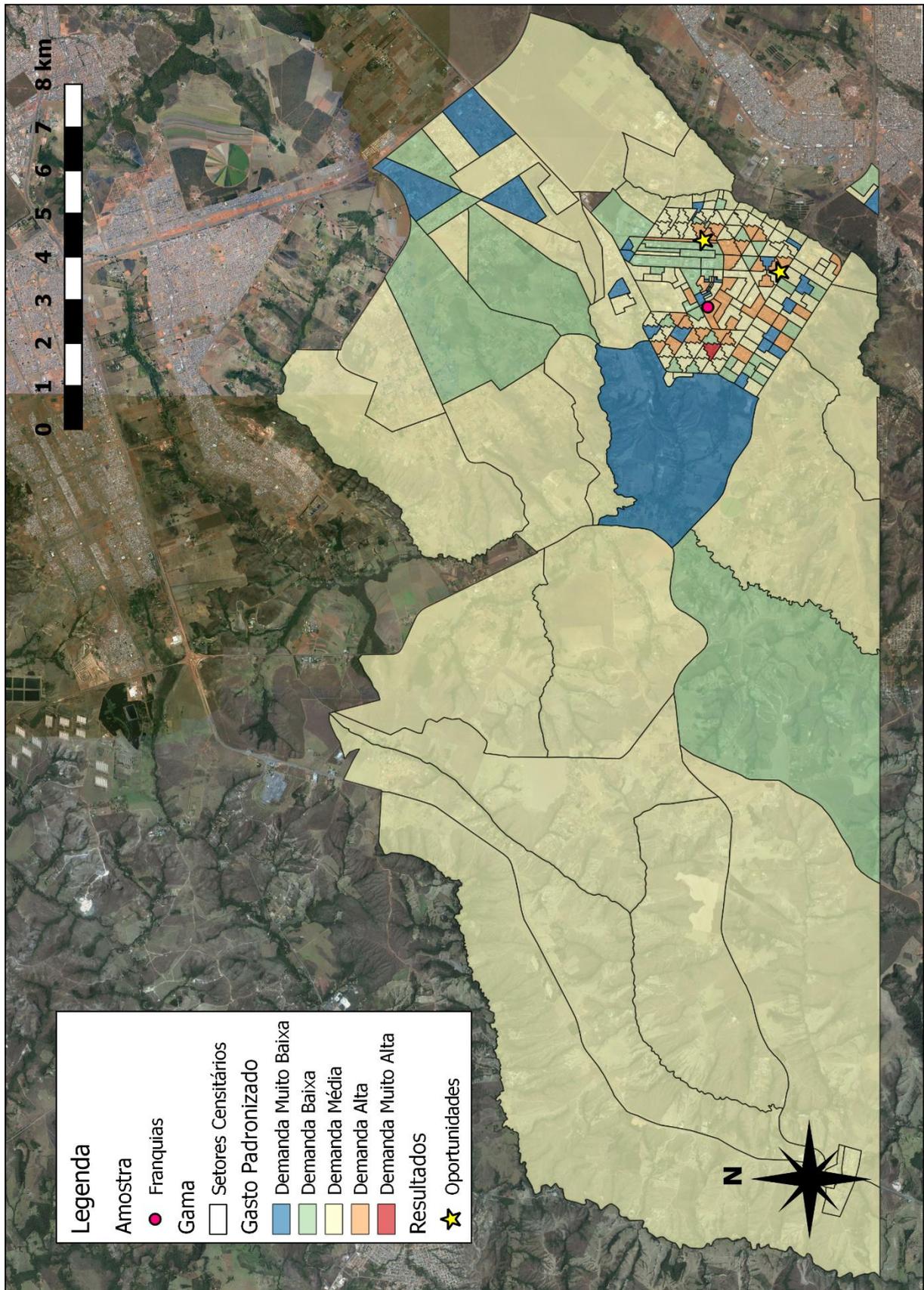
Mapa 8 – Oportunidades de Negócio na Ceilândia



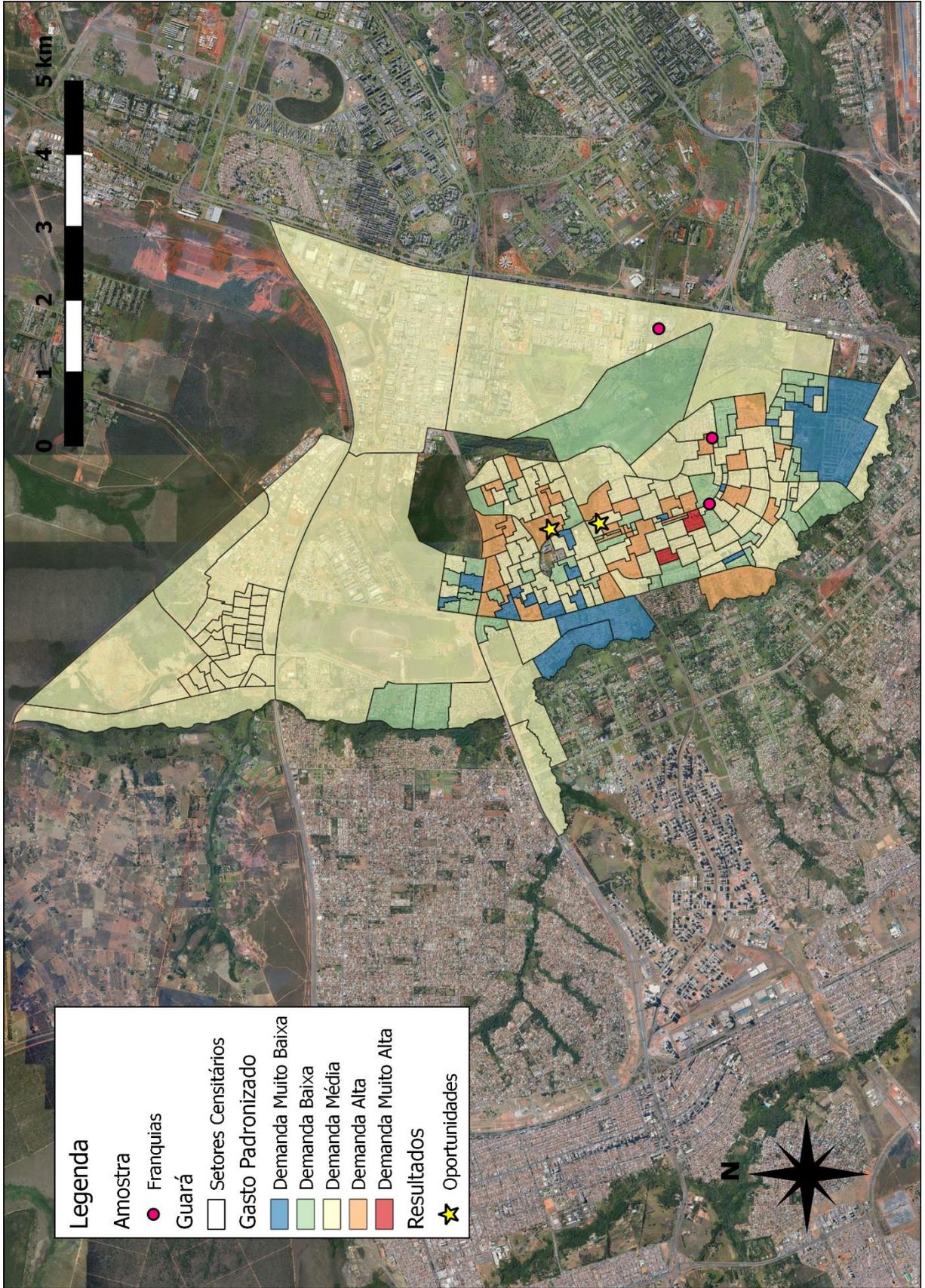
Mapa 9 – Oportunidades de Negócio no Cruzeiro



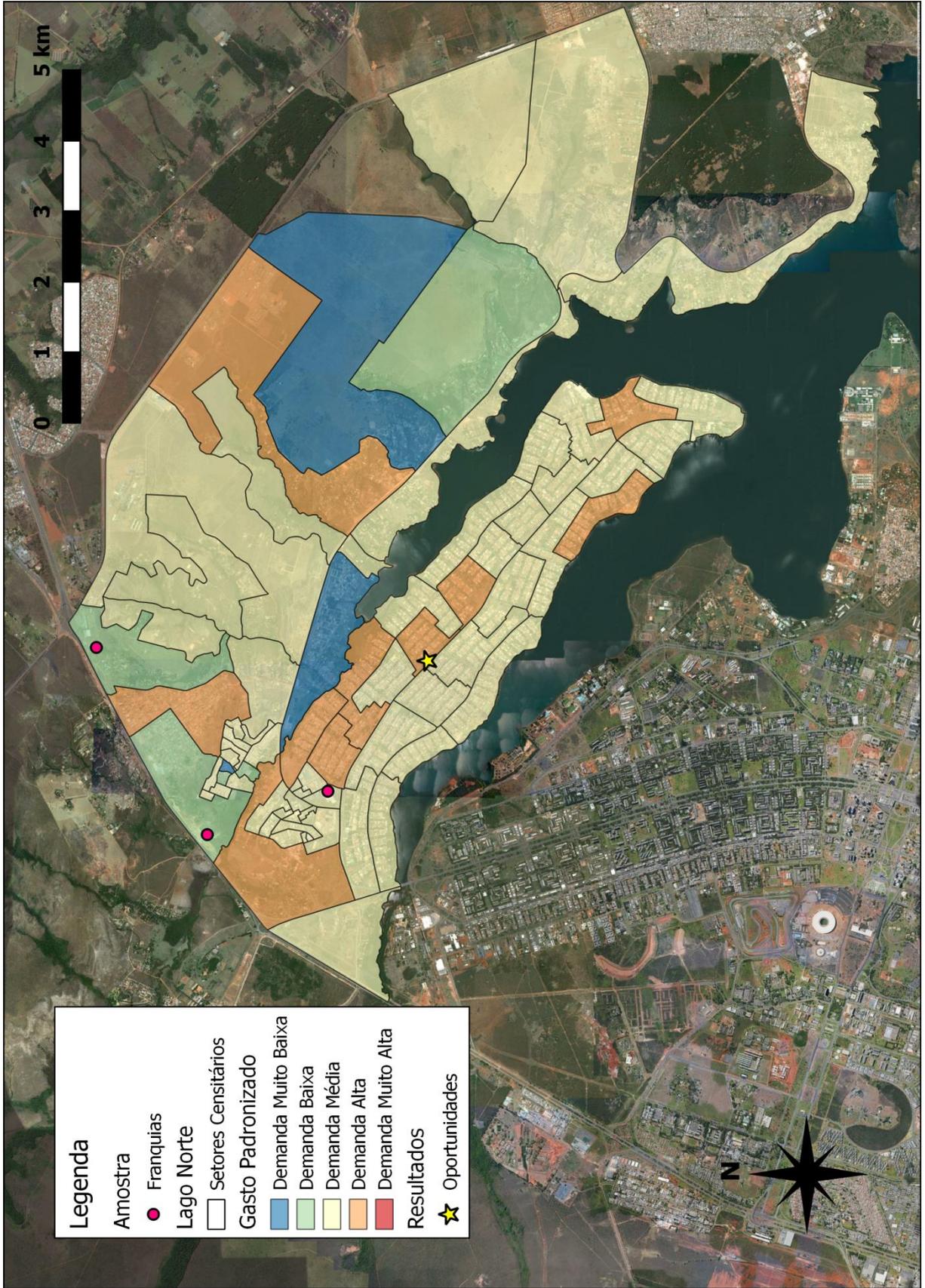
Mapa 10 – Oportunidades de Negócio no Gama



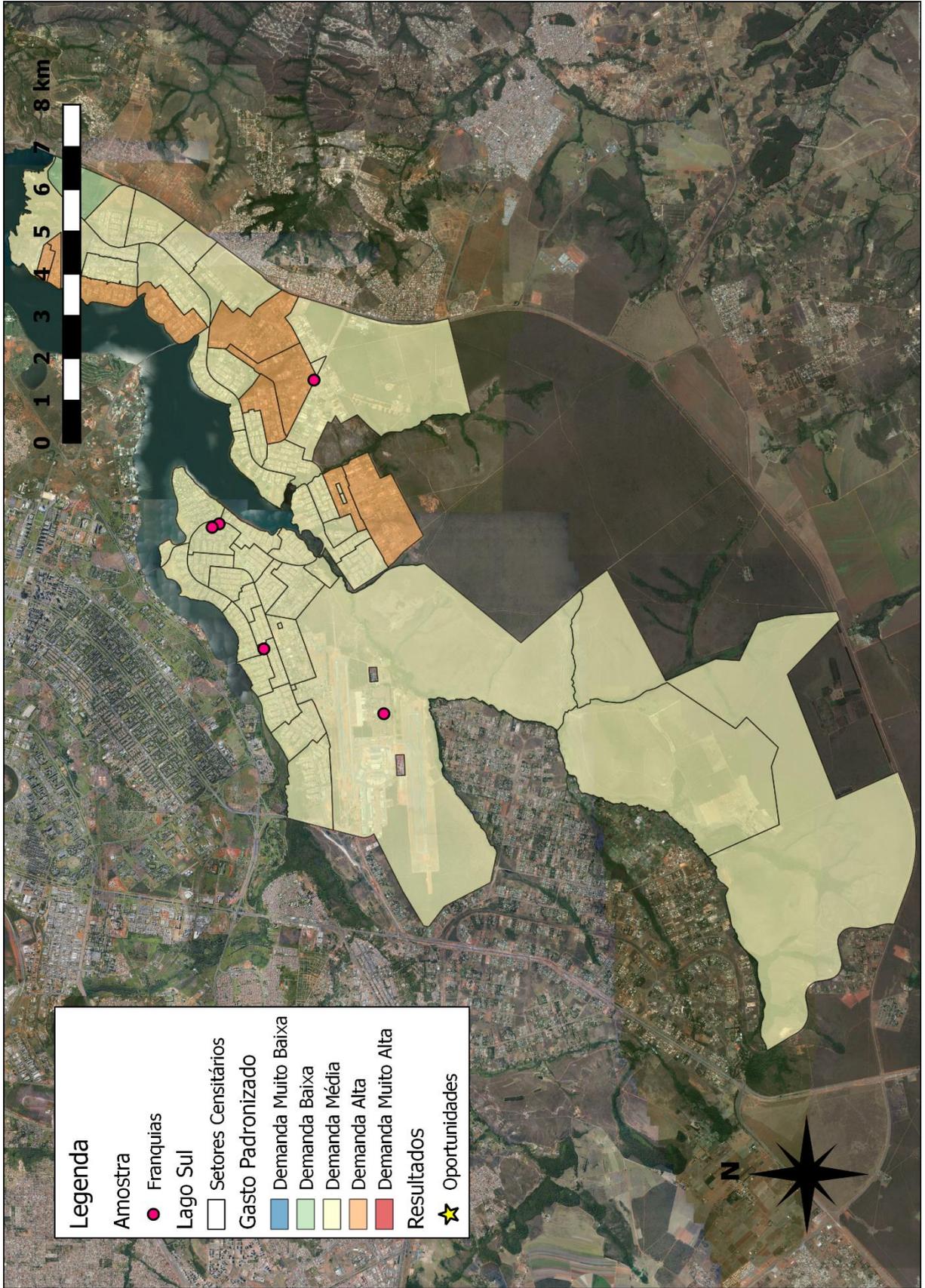
Mapa 11 – Oportunidades de Negócio no Guará



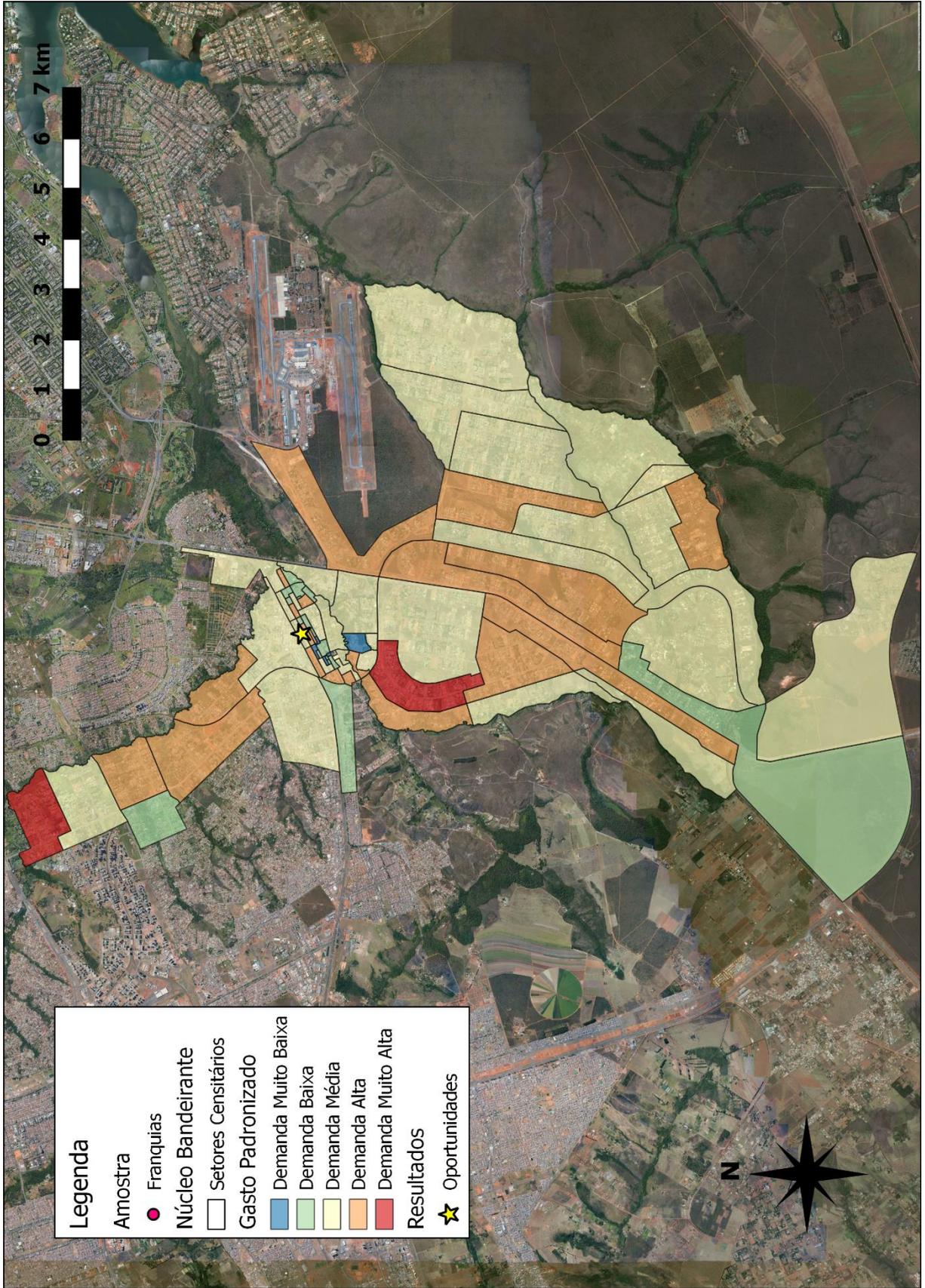
Mapa 12 – Oportunidades de Negócio no Lago Norte



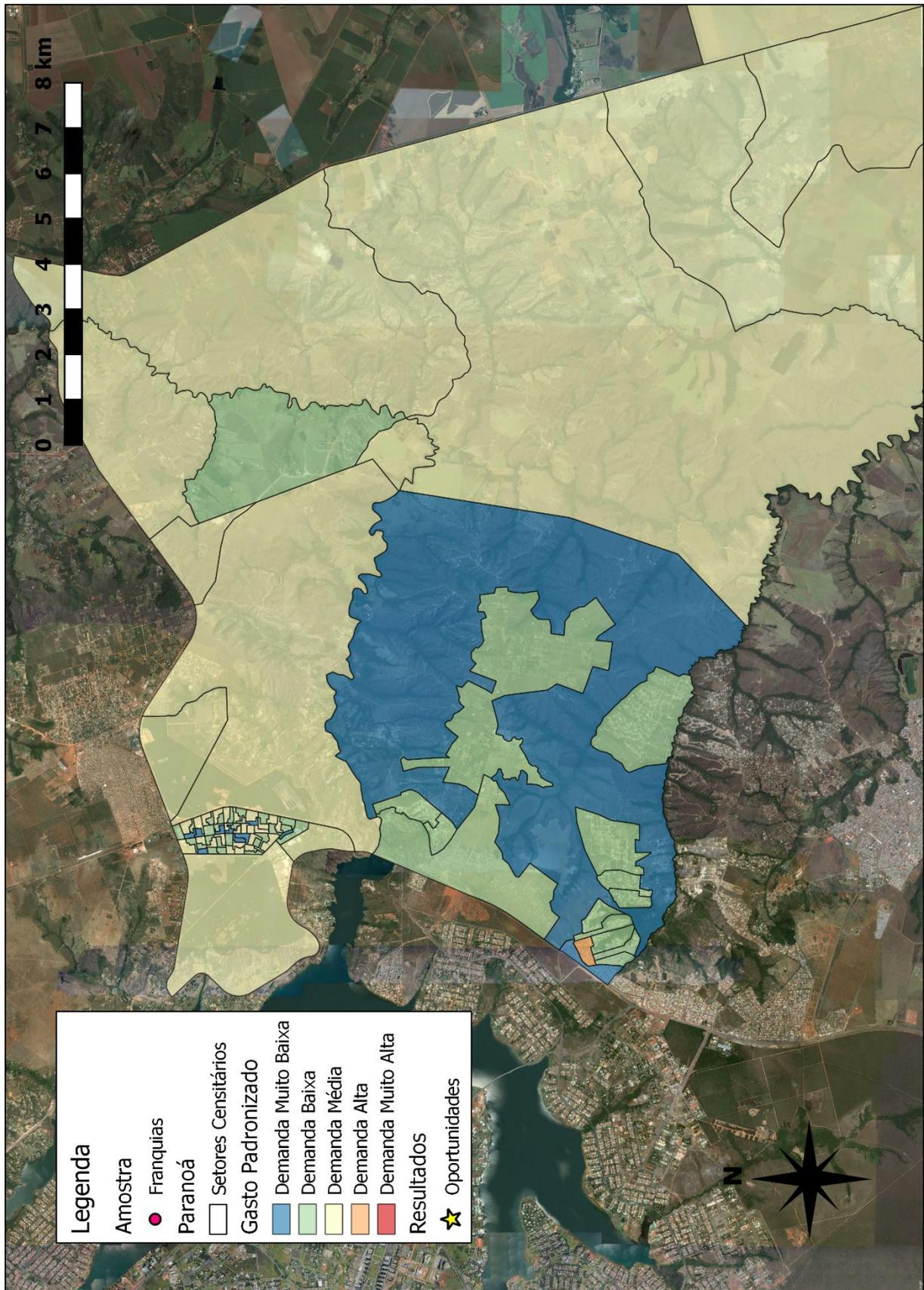
Mapa 13 – Oportunidades de Negócio no Lago Sul



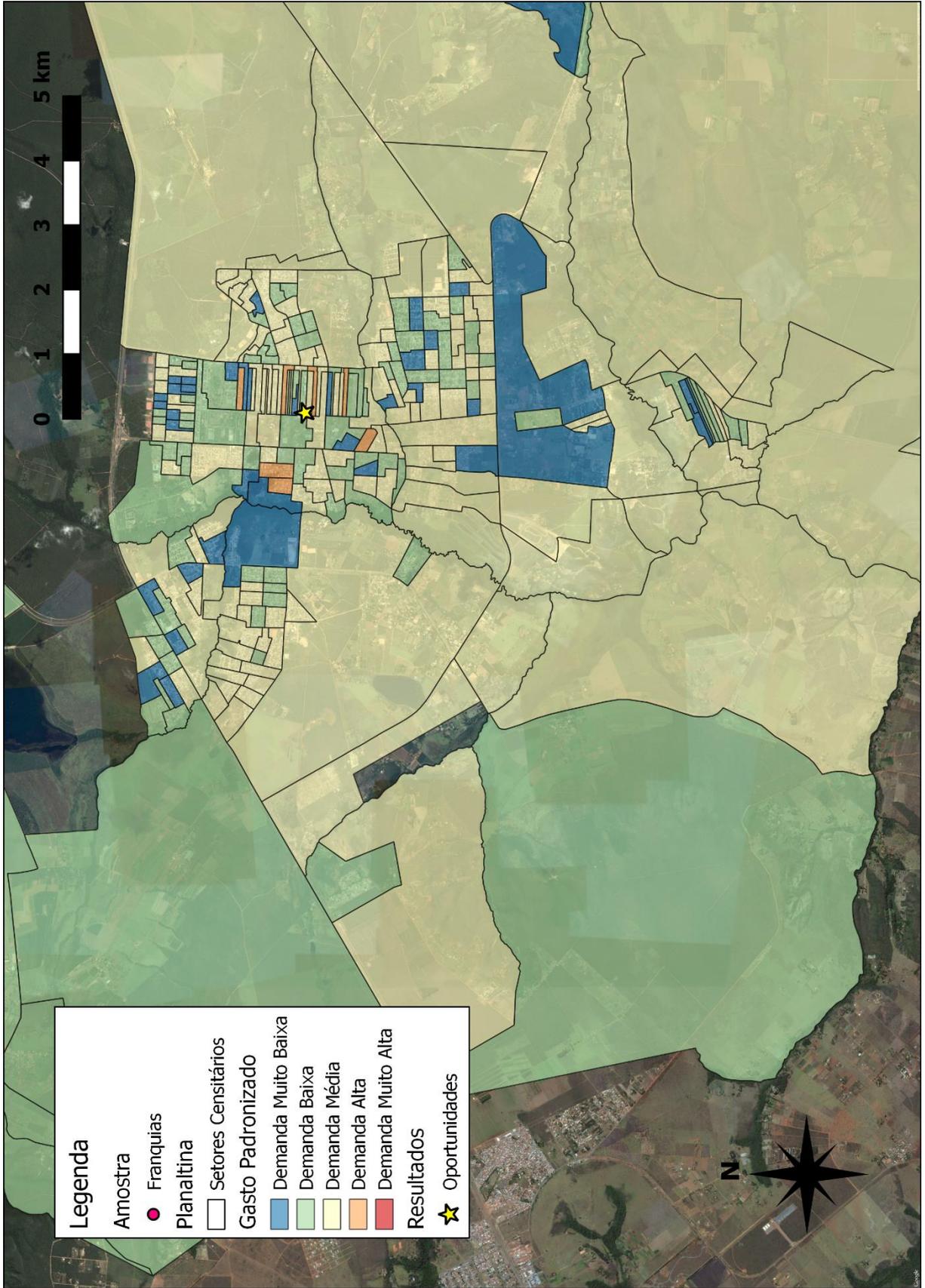
Mapa 14 – Oportunidades de Negócio no Núcleo Bandeirante



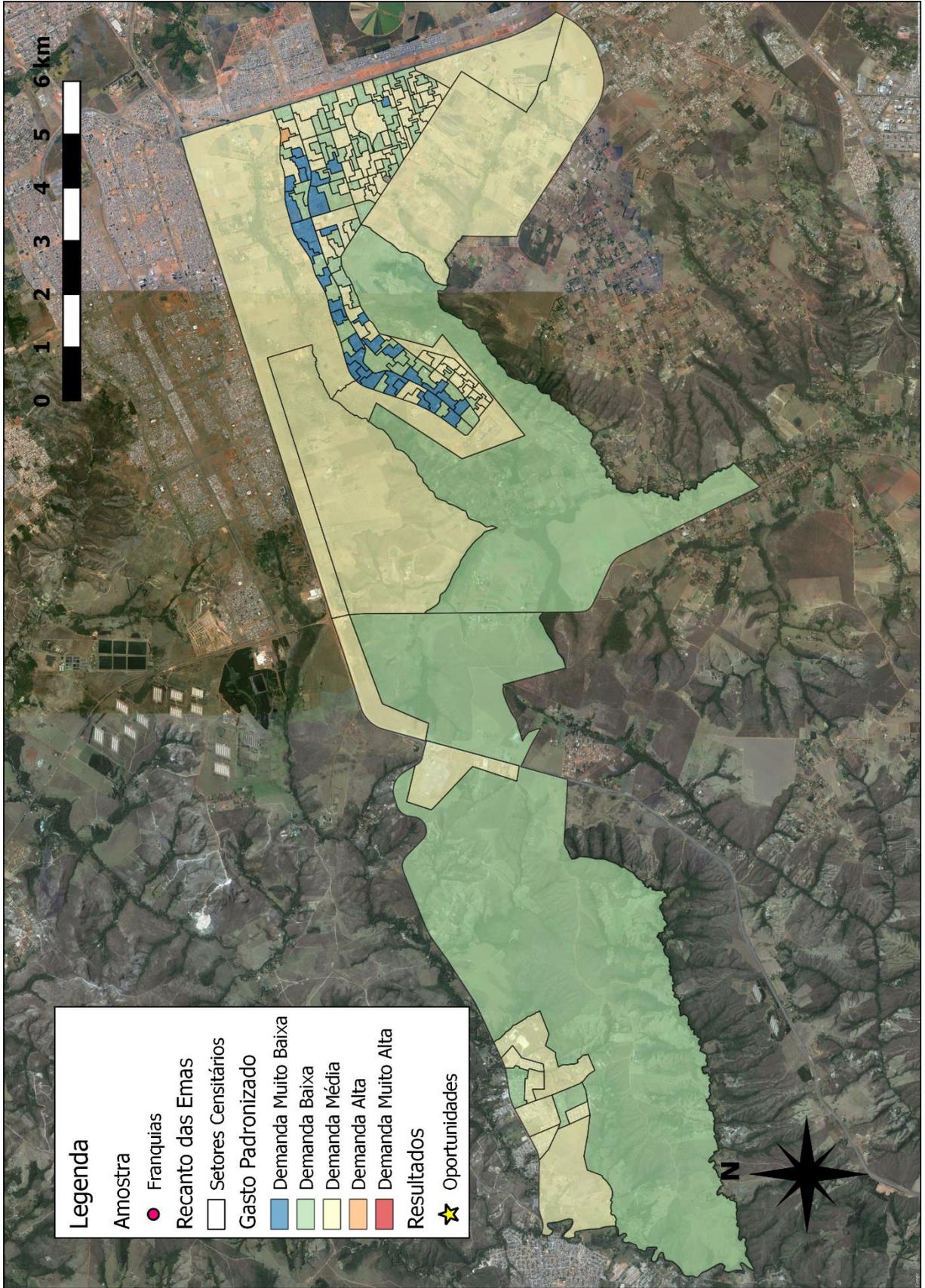
Mapa 15 – Oportunidades de Negócio no Paranoá



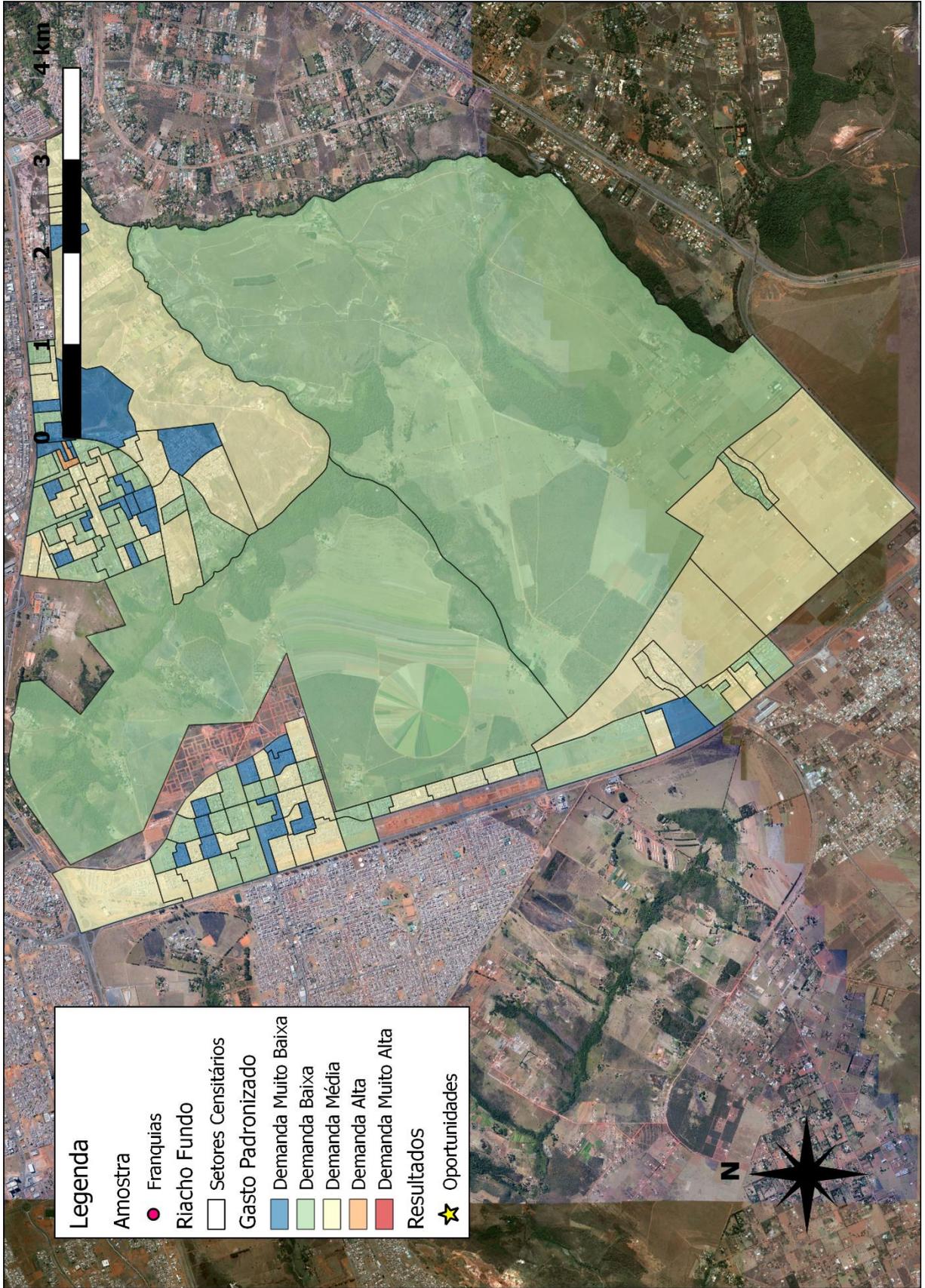
Mapa 16 – Oportunidades de Negócio em Planaltina



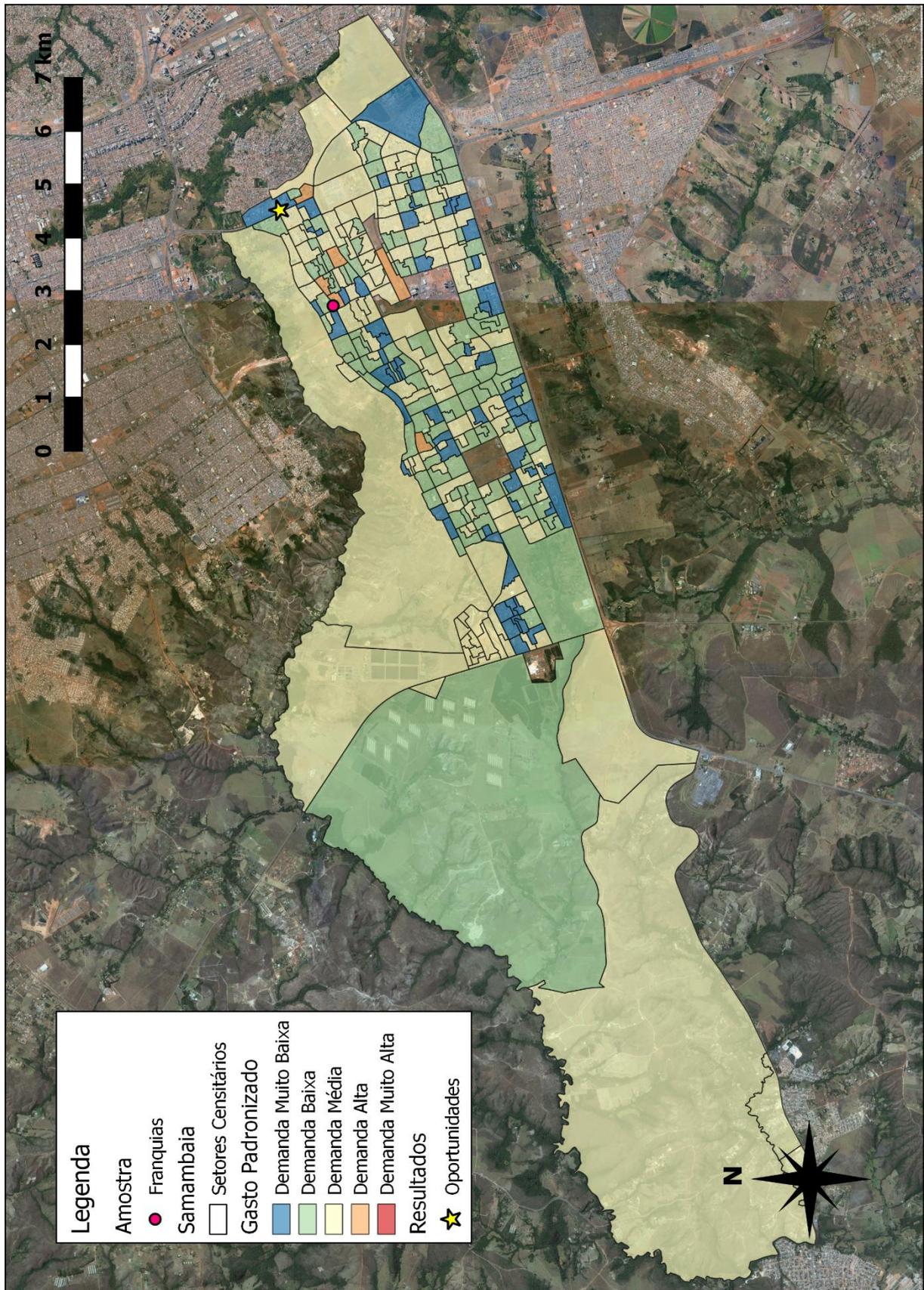
Mapa 17 – Oportunidades de Negócio no Recanto das Emas



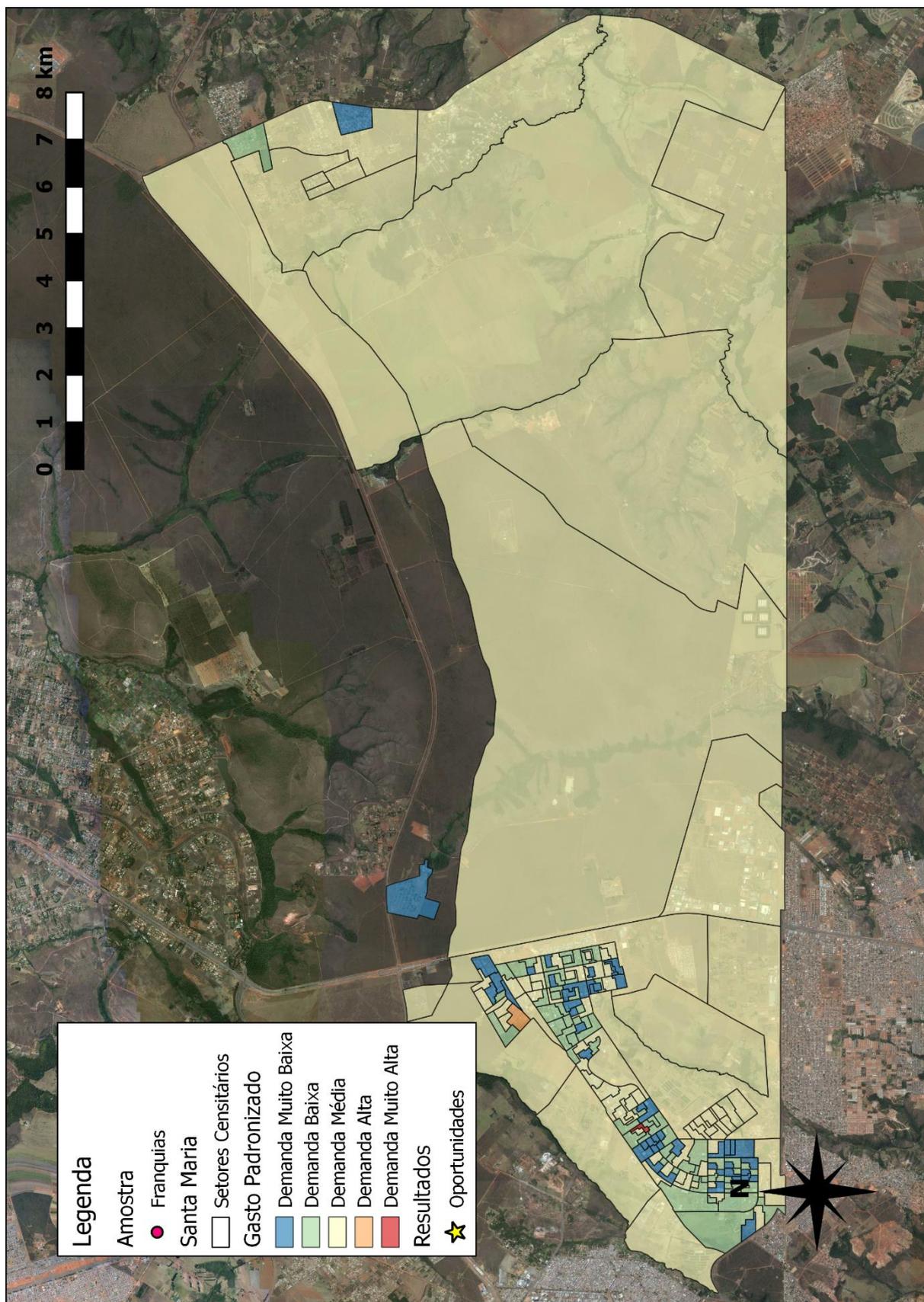
Mapa 18 – Oportunidades de Negócio no Riacho Fundo



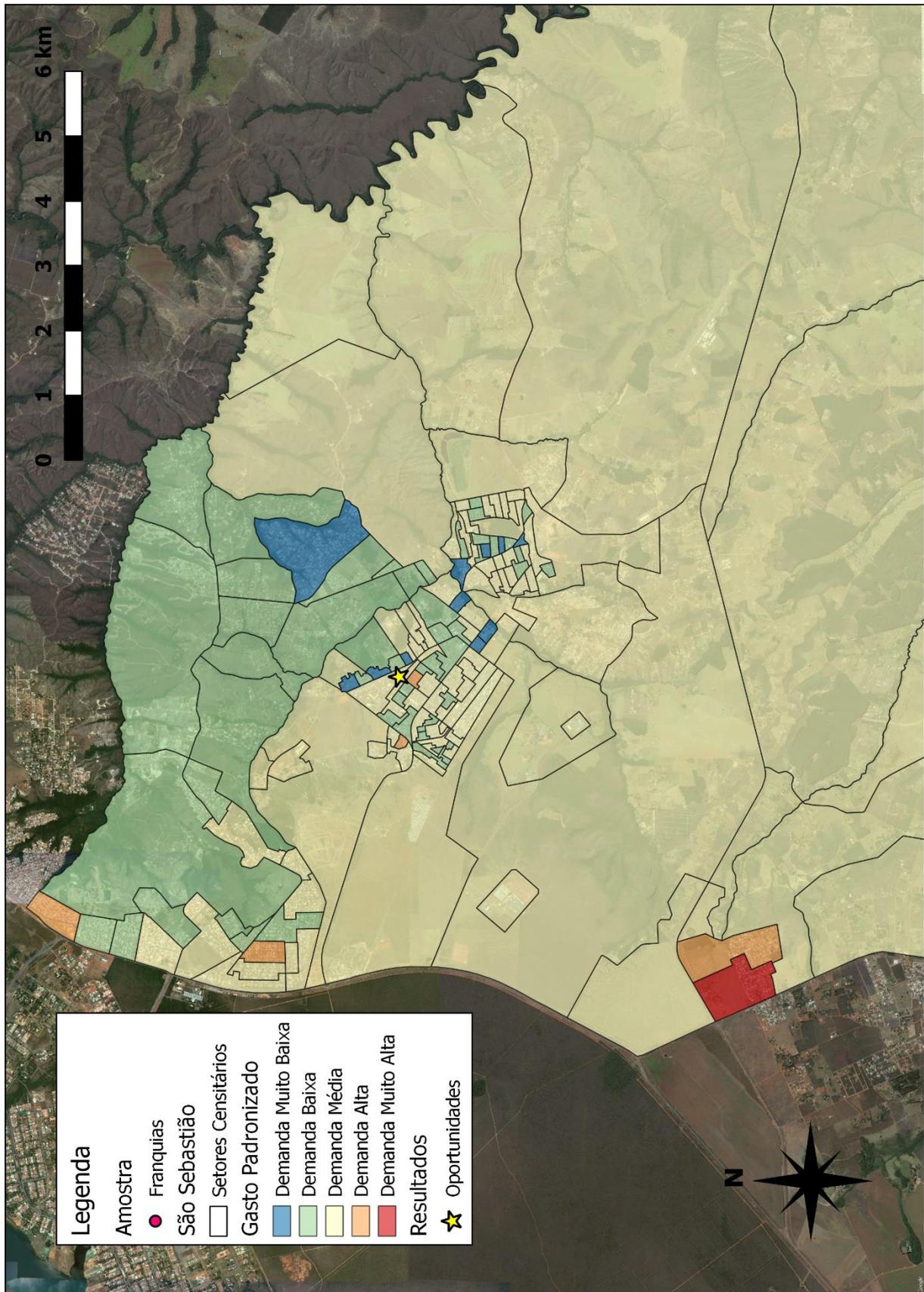
Mapa 19 – Oportunidades de Negócio em Samambaia



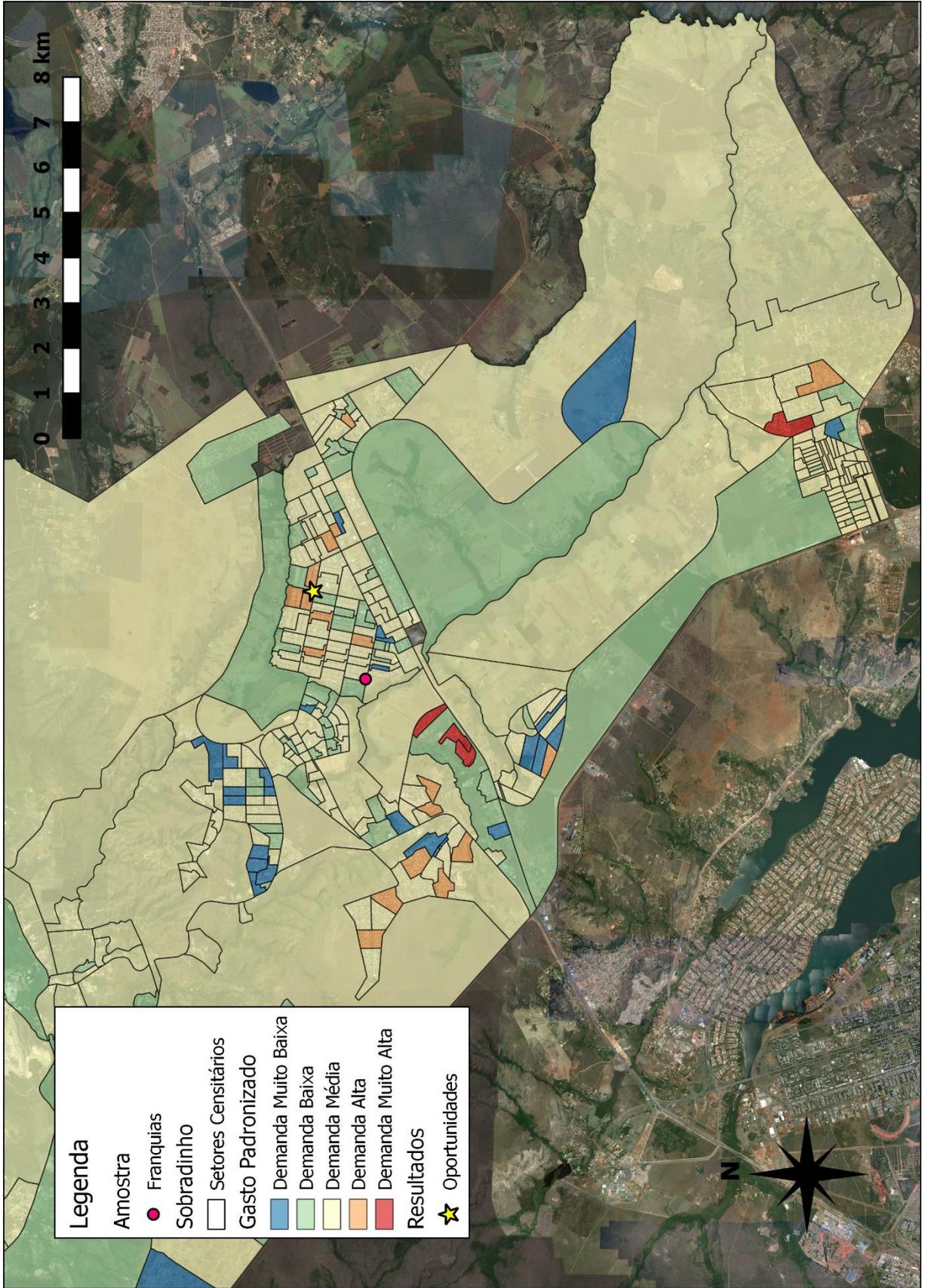
Mapa 20 – Oportunidades de Negócio em Santa Maria



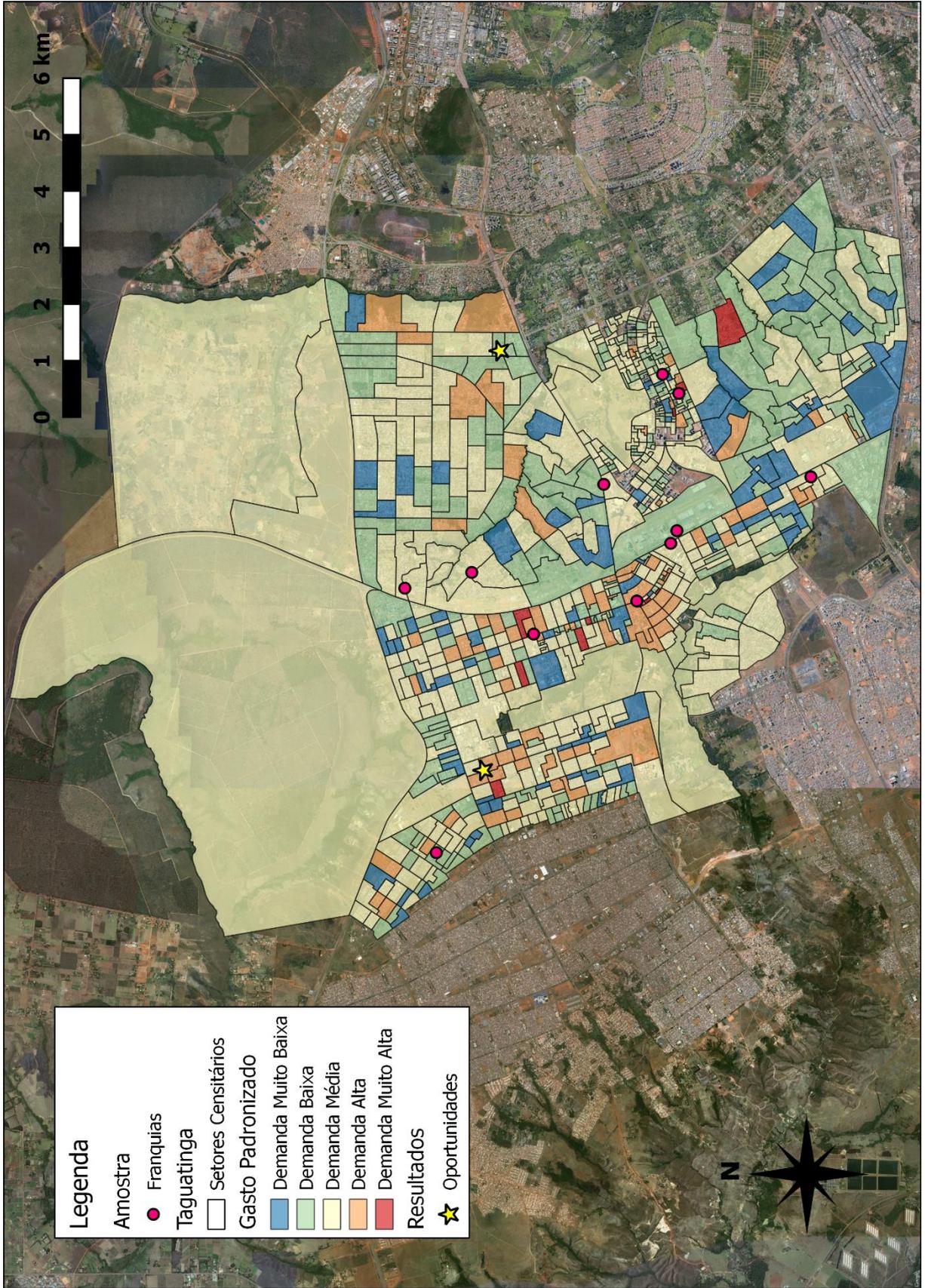
Mapa 21 – Oportunidades de Negócio em São Sebastião



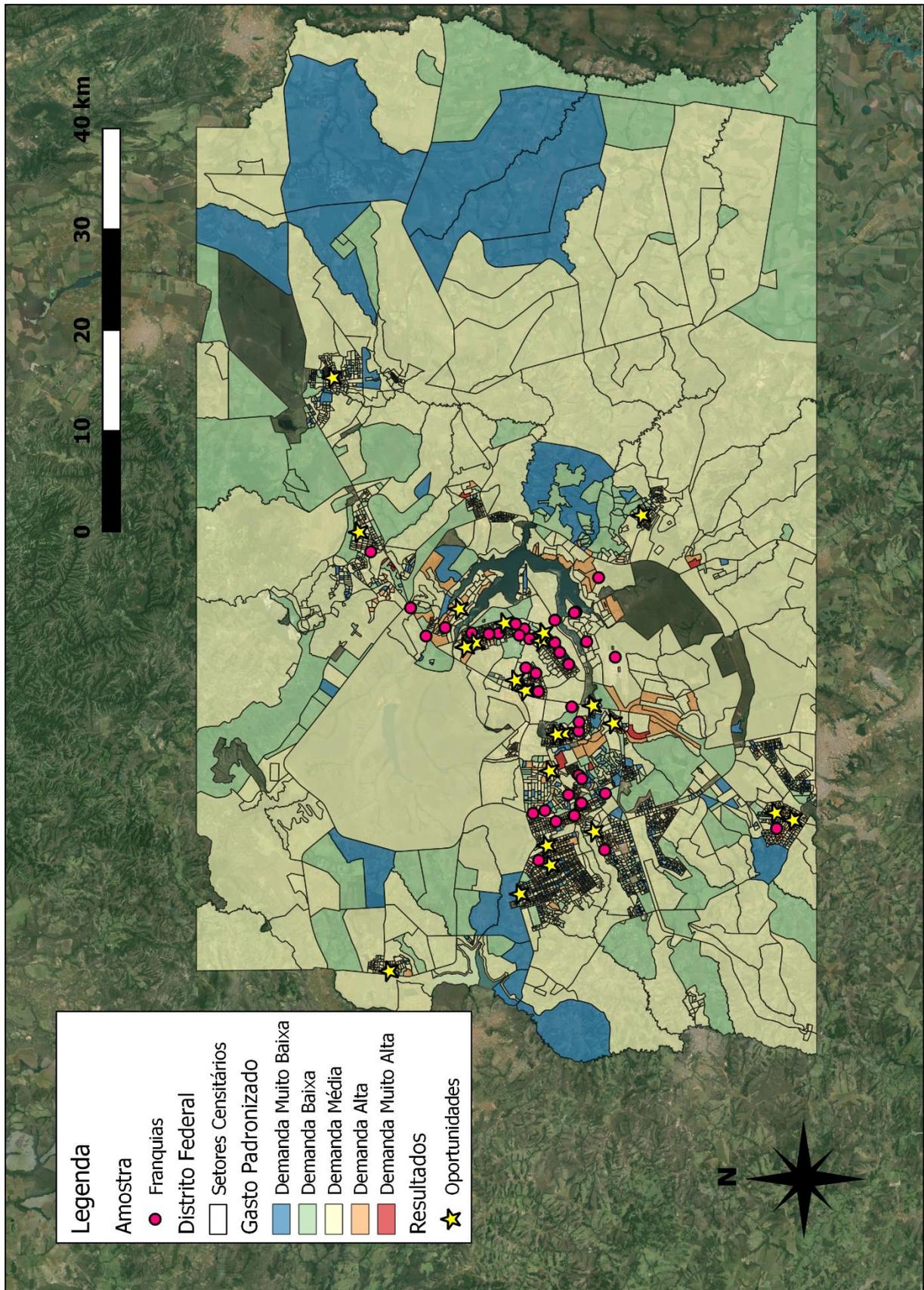
Mapa 22 – Oportunidades de Negócio em Sobradinho



Mapa 23 – Oportunidades de Negócio em Taguatinga



Mapa 24 – Oportunidades de Negócio no Distrito Federal



ANEXOS

Anexo A – Bloco de Consumo Alimentar Pessoal

Ministério do Planejamento, Orçamento e Gestão  Instituto Brasileiro de Geografia e Estatística Diretoria de Pesquisas Coordenação de Trabalho e Rendimento Gerência da Pesquisa de Orçamentos Familiares								
Pesquisa de Orçamentos Familiares 2008 - 2009								
POF 7 - Bloco de Consumo Alimentar Pessoal								
70	IDENTIFICAÇÃO E CONTROLE DO QUESTIONÁRIO							
01	IDENTIFICAÇÃO GERAL							
	UF	MUNICÍPIO	SUBDIS-TRITO	SETOR	Nº DE ORDEM NA LISTAGEM			
	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>			
02	IDENTIFICAÇÃO POF							
	UF	SEQUENCIAL	DV	CÓDIGO DO DOMICÍLIO	PERÍODO TEÓRICO	PERÍODO REAL	Nº DA UC	Nº DO INFORMANTE
	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
03	NOME DO INFORMANTE <input type="text"/>							
04	TOTAL DE DIAS PESQUISADOS <input type="text"/>							
05	REGISTROS FEITOS PELO PRÓPRIO INFORMANTE? 1 <input type="checkbox"/> SIM 3 <input type="checkbox"/> NÃO							
<p style="text-align: center;">Prezado(a) senhor(a), A sua colaboração no preenchimento deste bloco representa uma efetiva contribuição para o sucesso da Pesquisa de Orçamentos Familiares. Recordamos que as informações prestadas serão usadas exclusivamente para fins estatísticos e serão mantidas em sigilo, conforme estabelecido na lei 5.534 de 14/11/1968. Muito obrigado por sua colaboração.</p>								
<p style="text-align: center;">O modo de viver e consumir da família brasileira.</p> 								

71	1	DATA: 17 / 06 / 2008	DIA DO POF 3: 1	DIA DA SEMANA: Terça-feira	1º DIA (CONTINUA)
2 SITUAÇÃO DO QUADRO					
1 <input checked="" type="checkbox"/> PESQUISADO COM REGISTRO		3 <input type="checkbox"/> PESQUISADO SEM REGISTRO		5 <input type="checkbox"/> NÃO-PESQUISADO	
INSTRUÇÕES PARA PREENCHIMENTO DO QUADRO					
<p>1 - Registre a quantidade de todos os alimentos consumidos por você, no domicílio ou fora, conforme exemplo de preenchimento desta página. D = Domicílio: o alimento consumido é proveniente do seu domicílio. F = Fora: o alimento é adquirido e também consumido fora do seu domicílio.</p> <p>2 - Procure descrever separadamente todos os itens consumidos. Se não for possível, registre junto, conforme o seguinte exemplo: 1 pão francês com manteiga.</p> <p>3 - Escolha dois dias da semana, com um intervalo de pelo menos um dia entre eles, para registrar seu consumo.</p>					
_____ utiliza com frequência: <input type="checkbox"/> Açúcar <input checked="" type="checkbox"/> Adoçante <input type="checkbox"/> Açúcar e Adoçante <input type="checkbox"/> Não utiliza					
FORNE DO ALIMENTO	HORÁRIO	DESCRIÇÃO DO ALIMENTO CONSUMIDO			
(3)	(4)	(5)			
D	7h	1 copo médio de café			
D	7h	2 ovos de galinha fritos			
D	7h	3 pontas de faca de manteiga			
D	7h	1 copo grande de leite com sabor			
D	7h	1 pão francês			
F	7h	2 balas			
F	9h	1 laranja			
F	10h	2 bifes de alcatra fritos			
F	13h	1 porção de batata-inglesa cozida			
F	13h	0,5 copo médio de café			
F	13h	1 taça de salada de frutas			
F	13h	3 escumadeiras de arroz			
F	13h	1 copo grande de refrigerante de guaraná			
F	13h	2 fatias de queijo prato			
F	16h	2 rodélas de abacaxi			
D	18h	3 conchas de sopa de legumes			
D	20h	1 colher de chá de azeite de oliva			
D	20h	3 colheres de sopa de doce de abóbora			
F	23h	1 lata de 350 ml de cerveja			

PESQUISA DE ORÇAMENTOS FAMILIARES 2008-2009			
Anexo auxiliar do POF 7 - Bloco de Consumo Alimentar Pessoal			
Para facilitar o preenchimento deste instrumento de coleta, utilize os exemplos abaixo.			
MEDIDAS			
Asa	Copo de cafezinho	Gomo	Prato de sobremesa
Bago	Copo de requeijão	Gramma	Prato fundo
Banda	Copo grande	Lata de ___ ml	Prato raso
Barra	Copo médio	Litro	Quilo
Bife	Copo tulipa	Maço	Ramo
Bisnaga	Costela	Metade	Rodela
Bola	Coxa	Mililitros	Sachê
Cacho	Cumbuca	Pacote	Saco
Caneca	Dose	Pedaço	Sobrecoxa
Caneco	Escumadeira	Pegador	Tablete
Casquinha	Espetinho	Peito	Taça
Colher de arroz / servir	Espeto	Pescoço	Tigela
Colher de café	Espiga	Pires	Unidade
Colher de chá	Fatia	Ponta de faca	Unidade pequena
Colher de sobremesa	Filé	Porção	Xícara de café
Colher de sopa	Folha	Punhado	Xícara de chá
Concha	Garfada	Posta	
Copo americano	Garrafa de ___ ml	Pote	
PREPARAÇÕES			
Cru(a)	Empanado(a)/à milanesa	Com manteiga/óleo	
Cozido(a)	Refogado(a)	Ao vinagrete	
Grelhado(a)/brasa/churrasco	Molho vermelho	Ensopado(a)	
Assado(a)	Molho branco	Mingau	
Frito(a)	Ao alho e óleo	Sopa	

OBSERVAÇÕES