

UNIVERSIDADE DE BRASÍLIA  
DEPARTAMENTO DE ADMINISTRAÇÃO  
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE (FACE).



**UnB**

**FERNANDA SANTOS AMORIM**

**Modelo de Auxílio à Tomada de Decisão em Turismo:  
Geomarketing e Hotelaria utilizando Aprendizado de  
Máquina.**

Brasília, DF  
2017

**FERNANDA SANTOS AMORIM**

**Modelo de Auxílio à Tomada de Decisão em Turismo:  
Geomarketing e Hotelaria utilizando Aprendizado de  
Máquina.**

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

Brasília, DF  
2017

**FERNANDA SANTOS AMORIM**

**Modelo de Auxílio à Tomada de Decisão em Turismo:  
Geomarketing e Hotelaria utilizando Aprendizado de  
Máquina.**

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

Área de concentração:  
Marketing

Supervisor:  
Prof. Dr. Pedro Henrique Melo Albu-  
querque

Brasília, DF  
2017

## FICHA

### CATALOGRÁFICA

Amorim, Fernanda Santos

Modelo de Auxílio à Tomada de Decisão em Turismo: Geomarketing e Hotelaria utilizando Aprendizado de Máquina / . – Brasília, DF, 2017. 88 p.

Monografia - Bacharelado — Universidade de Brasília - Departamento de Administração.

1. Geomarketing. 2. Turismo .3. Hotelaria. 4. Aprendizado de Máquina. I. Amorim, Fernanda Santos II. Universidade de Brasília.

## DEDICATÓRIA

Dedico esse trabalho a Deus, que sempre guiou os vários aspectos da minha vida. Á minha família, que sempre apoiou todas as minhas decisões e foram suporte nas horas de dificuldade. Ao meu orientador, Pedro Albuquerque, que aguentou as minhas mensagens desesperadas até a última palavra escrita desse trabalho.

## AGRADECIMENTOS

Agradeço a Deus, a quem eu dou glória e honra por cada detalhe que acontece em minha vida. Ao meu orientador, Pedro Albuquerque, que viu algum potencial em mim desde o dia que eu entrei na sala pela primeira vez, obrigada por todo o aprendizado, conselhos e orientações acadêmicas. À minha família, Mônica, Daniel, Daniel Júnior e Arthur, vocês são o melhor suporte que eu poderia pedir a Deus, obrigada por aguentar o meu estresse durante esse ano de trabalho.

Agradeço aos meus amigos, em especial, Carolina Lacerda, Edevaldo Henrique, Jonas Ramos, Ana Carolina Silveira, por serem suporte em oração e ajuda nas horas de angústia e desespero. Agradeço aos meus amigos do Laboratório de Aprendizado de Máquina em Finanças e Organizações, são muitos nomes e eu não citarei todos, mas vocês são parte crucial do meu crescimento, muito obrigada. Às minhas amigas e companheiras de UnB, Talita Lima e Gabriela Nobre, obrigada por aguentar as minhas reclamações ao longo do ano e por cada conselho, a UnB foi melhor por causa de vocês.

## RESUMO

Este estudo procura relacionar as teorias de Geomarketing, Turismo e Aprendizado de Máquina para criar um modelo quantitativo que possa auxiliar na tomada de decisão gerencial para alocação de um hotel. De acordo com Latour e Floc'h (2001), Geomarketing consiste na junção dos estudos do marketing, da economia e da geografia para avaliar o comportamento do consumidor no espaço e elencar informações estratégicas para os gestores na tomada de decisão. O foco deste estudo está na aplicação do Geomarketing dentro de um dos elos da cadeia da indústria do Turismo, o mercado hoteleiro. O Aprendizado de Máquina é uma das linhas da Inteligência Artificial que tem por objetivo a criação de algoritmos que conseguem aprender de forma automatizada baseados características já aprendidas, ou seja, são algoritmos que conseguem generalizar uma situação quando apresentados aos dados que a descrevem. Nesta pesquisa, serão utilizados os estudos de Aprendizado de Máquina derivados de Drucker et al. (1997) e Smola e Schölkopf (2004) que geraram o *Support Vector Regression* (SVR), um método consequente do *Support Vector Machine* (SVM) de Cortes e Vapnik (1995). O modelo de previsão foi elaborado relacionando o preço das diárias, a localização dos hotéis e as amenidades oferecidas de hotéis de 11 capitais brasileiras com potencial turístico, são elas: Brasília, Belo Horizonte, Curitiba, Florianópolis, Fortaleza, Maceió, Manaus, Natal, Porto Alegre, Recife e Salvador.

Palavras Chaves: Geomarketing, Turismo, Hotelaria, Aprendizado de Máquina, *Support Vector Regression*

## LISTA DE FIGURAS

2.1	Aneis de Thünen Fonte: Aranha e Figoli (2001) . . . . .	19
2.2	Hierarquia de Localidades Conforme Teoria do Lugar Central Fonte: Aranha e Figoli (2001) . . . . .	21
2.3	Base de Dados Geocodificados Fonte: Elaboração da Autora . . . . .	23
2.4	Esferas do Geomarketing de acordo com Cliquet (2013) Fonte: Elaboração da Autora . . . . .	24
2.5	Hierarquia do Aprendizado Indutivo Fonte: Monard e Baranauskas () . . . . .	28
2.6	Indução do classificador em aprendizado supervisionado Fonte: Lorena e Carvalho (2007) . . . . .	28
2.7	Três diferentes hipóteses de classificador para um conjunto de exemplos de treinamento Fonte: Lorena e Carvalho (2007) . . . . .	29
2.8	Separador de Máxima Margem Fonte: Soman, Loganathan e Ajay (2009) . . . . .	30
2.9	Espaço de Entrada Fonte: Soman, Loganathan e Ajay (2009) . . . . .	31
2.10	Espaço de Característica Fonte: Soman, Loganathan e Ajay (2009) . . . . .	31
2.11	Curva de Separação Fonte: Soman, Loganathan e Ajay (2009) . . . . .	32
3.1	Representação gráfica da função de perda Fonte: Albuquerque (2014) . . . . .	38
3.2	Representação Gráfica do SVR na forma linear Fonte: Albuquerque (2014) . . . . .	39
3.3	Representação Gráfica do SVR na forma não linear Fonte: Albuquerque (2014) . . . . .	40
3.4	<i>Kernel Mexican Hat</i> Fonte: Ding, Wu e Shi (2014) . . . . .	42
3.5	<i>Kernel Mexican Hat</i> Fonte: Ding, Wu e Shi (2014) . . . . .	43
4.1	Mapa de Preços - Brasília - <i>Kernel Mexican Hat</i> . . . . .	49
4.2	Mapa de Preços - Brasília - <i>Kernel Polynomial</i> . . . . .	50
4.3	Mapa de Preços - Belo Horizonte - <i>Kernel Mexican Hat</i> . . . . .	52

4.4	Mapa de Preços - Belo Horizonte - <i>Kernel Polynomial</i> . . . . .	53
4.5	Mapa de Preços - Curitiba - <i>Kernel Mexican Hat</i> . . . . .	55
4.6	Mapa de Preços - Curitiba - <i>Kernel Mexican Hat</i> . . . . .	56
4.7	Mapa de Preços - Florianópolis - <i>Kernel Mexican Hat</i> . . . . .	58
4.8	Mapa de Preços - Florianópolis - <i>Kernel Polymonial</i> . . . . .	59
4.9	Mapa de Preços - Fortaleza - <i>Kernel Mexican Hat</i> . . . . .	61
4.10	Mapa de Preços - Fortaleza - <i>Kernel Polymonial</i> . . . . .	62
4.11	Mapa de Preços - Maceió - <i>Kernel Mexican Hat</i> . . . . .	64
4.12	Mapa de Preços - Maceió - <i>Kernel Polymonial</i> . . . . .	65
4.13	Mapa de Preços - Manaus - <i>Kernel Mexican Hat</i> . . . . .	67
4.14	Mapa de Preços - Manaus - <i>Kernel Polymonial</i> . . . . .	68
4.15	Mapa de Preços - Natal - <i>Kernel Mexican Hat</i> . . . . .	70
4.16	Mapa de Preços - Natal - <i>Kernel Polymonial</i> . . . . .	71
4.17	Mapa de Preços - Porto Alegre - <i>Kernel Mexican Hat</i> . . . . .	73
4.18	Mapa de Preços - Porto Alegre - <i>Kernel Polymonial</i> . . . . .	74
4.19	Mapa de Preços - Recife - <i>Kernel Mexican Hat</i> . . . . .	76
4.20	Mapa de Preços - Recife - <i>Kernel Polymonial</i> . . . . .	77
4.21	Mapa de Preços - Salvador - <i>Kernel Mexican Hat</i> . . . . .	79
4.22	Mapa de Preços - Salvador - <i>Kernel Polymonial</i> . . . . .	80

## LISTA DE TABELAS

2.1	Histórico do Geomarketing . . . . .	25
3.1	Tabela de Parâmetros - <i>Kernel Mexican Hat</i> . . . . .	46
3.2	Tabela de Parâmetros - <i>Kernel Polynomial</i> . . . . .	46

# SUMÁRIO

<b>1</b>	<b>Introdução.</b>	<b>12</b>
1.1	Problema da Pesquisa . . . . .	13
1.2	Objetivo Geral . . . . .	15
1.3	Objetivos Específicos . . . . .	15
1.4	Motivação/Justificativa . . . . .	15
<b>2</b>	<b>Referencial Teórico</b>	<b>18</b>
2.1	Histórico do Geomarketing . . . . .	18
2.2	Sistema de Informação Geográfica (SIG) . . . . .	25
2.3	Máquina de Suporte Vetorial . . . . .	27
2.4	Máquina de Suporte Vetorial em Marketing . . . . .	32
2.5	Máquina de Suporte Vetorial em Turismo . . . . .	35
2.6	Turismo e Hotelaria . . . . .	35
<b>3</b>	<b>Metodologia</b>	<b>37</b>
3.1	<i>Support Vector Regression - SVR</i> . . . . .	37
3.2	<i>Kernel Mexican Hat</i> . . . . .	41
3.3	<i>Kernel Polynomial</i> . . . . .	43
3.4	Modelo Conceitual - Geomarketing e Turismo . . . . .	43
3.5	Dados do Modelo . . . . .	45
<b>4</b>	<b>Resultados</b>	<b>47</b>
4.1	Análise do Modelo . . . . .	47

4.1.1	Brasília - DF . . . . .	47
4.1.2	Belo Horizonte - MG . . . . .	51
4.1.3	Curitiba - PR . . . . .	54
4.1.4	Florianópolis - SC . . . . .	57
4.1.5	Fortaleza - CE . . . . .	60
4.1.6	Maceió - AL . . . . .	63
4.1.7	Manaus - AM . . . . .	66
4.1.8	Natal - RN . . . . .	69
4.1.9	Porto Alegre - RS . . . . .	72
4.1.10	Recife - PE . . . . .	75
4.1.11	Salvador - BA . . . . .	78
<b>5</b>	<b>Conclusão</b>	<b>81</b>
	<b>Referências Bibliográficas</b>	<b>84</b>

# 1 INTRODUÇÃO.

O setor do Turismo movimenta muitas das economias mundiais. Segundo o *World Economic Forum*, este setor tem impulsionado o crescimento econômico mundial nas últimas décadas, colaborando em 10,2% no PIB mundial e correspondendo a 292 milhões de empregos, isto é, 1 em cada 10 empregos no mundo (FORUM, 2017). Mesmo com a instabilidade econômica e política em alguns países, a previsão do *World Economic Forum* é de desenvolvimento para o setor nos próximos anos, com expectativas de crescimento de 3,9% ao ano para os próximos 10 anos.

A cada dois anos, o *World Economic Forum* publica, em seu *The Travel & Tourism Competitiveness Report* (Relatório de Competitividade de Viagens e Turismo em português), as análises referentes ao Índice de Competitividade de Viagens e Turismo (*Travel & Tourism Competitiveness Index – TTCI*) (FORUM, 2017). No ano de 2017, as pesquisas relatam quatro resultados principais:

1. Melhora na competitividade da indústria de *Travel&Tourism* (T&T) principalmente em países em desenvolvimento, o que impulsiona as economias emergentes;
2. Esta indústria continua construindo pontes entre as pessoas com um número maior de indivíduos atravessando as fronteiras entre os países.
3. Com advento da Quarta Revolução Industrial, a indústria tem procurado aprimorar a conectividade dos processos e os países têm buscado o desenvolvimento das estruturas tecnológicas.
4. Por fim, a indústria de T&T, no momento, precisa reforçar o desenvolvimento de políticas mais sustentáveis devido à crescente conscientização da importância do meio ambiente.

O primeiro resultado, derivado das análises da *Travel & Tourism Competitiveness Report* (FORUM, 2017), mostra que os países emergentes têm se tornado mais competitivos na indústria. O crescimento do Turismo e a maior proporção de estrangeiros visitando os países

em desenvolvimento são motivos para impulsionar as economias emergentes e proporcionar mais investimento na infraestrutura dos países e desenvolvimento socioeconômico (AKAMA; KI-ETI, 2007).

Mesmo diante de um contexto, por vezes protecionista, o intenso fluxo de pessoas entre os países gerado pela indústria de T&T ajuda a construir relações entre as nações, provendo tendências menos restritivas. O segundo resultado da pesquisa mostra que essas tendências de diminuição das barreiras facilitam o comércio global, incentivam a criação de empresas e possibilitam um crescimento econômico (FORUM, 2017).

O terceiro resultado da pesquisa mostra que esta indústria tem buscado usufruir da tecnologia nos seus processos, devido às influências da globalização, da internet e dos avanços tecnológicos. A tecnologia tem mudado a forma como as pessoas se relacionam e as empresas precisam estar preparadas para as mudanças que isso acarretará. Na indústria do Turismo, as aplicações tecnológicas e comerciais têm sido incentivo para a criação de novos modelos de negócio e desenvolvimento de diferentes mercados (WANG; YU; FESENMAIER, 2002).

A indústria do Turismo é formada por vários serviços, como hotelaria, alimentação, transporte, entretenimento e vários outros serviços correlatos (THEOBALD, 2001). Neste trabalho, será estudado o ramo de hotelaria com foco principal em como a localização geográfica influencia na movimentação do mercado, aplicando as teorias de Geomarketing neste setor.

O setor de hotelaria corresponde a um dos elos da cadeia de serviços turísticos. Este setor representa um serviço de alto contato com o cliente, logo tem grande influência na qualidade percebida pelos usuários (BRANCO; RIBEIRO; TINOCO, 2010). No Brasil, este ramo tem grande influência na economia e possui uma pesquisa de mercado específica no IBGE, a Pesquisa de Serviços de Hospedagem (PSH) realizada desde 2011.

## **1.1 Problema da Pesquisa**

De acordo com Yrigoyen (2003), o sucesso empresarial tornou-se cada vez mais difícil de ser alcançado devido às novas tecnologias, às mudanças nos hábitos de consumo e ao aparecimento de novas formas de negócio. Entretanto, paralelamente ao surgimento dos fatores citados, o desenvolvimento de novas perspectivas dentro do marketing tem possibilitado a inclusão de novos conceitos e ferramentas, fornecendo aos gestores uma melhor compreensão do ambiente comercial e um maior controle das mudanças advindas da tecnologia (YRIGOYEN, 2003). O presente trabalho tem como objetivo estudar fatores que auxiliam na análise de mercado utilizando informações comportamentais combinado com informações espaciais. Com ligação

destes dados, é possível fazer uma análise profunda de mercado utilizando Geomarketing. O Geomarketing vem da junção dos estudos espaciais e dos estudos de mercado, unindo assim fenômenos socioeconômicos e geográficos. Os aspectos espaciais e a informação geográfica que os determinam, possibilitam ir a fundo em estudos indispensáveis para as definições estratégicas (CLIQUET, 2013; GIOVANARDI; LUCARELLI, 2018). Como já abordado, a indústria escolhida para a análise é a indústria do Turismo, com foco no setor de hotelaria. Além de analisar os dados espaciais e socioeconômicos, será utilizado Aprendizado de Máquina como método para auxiliar na decisão locacional.

A crescente competitividade organizacional força as empresas a buscarem inovações e, por consequência, vantagens competitivas para os negócios. Segundo Bose e Mahapatra (2001), para fidelizar e alcançar novos segmentos de mercado é necessário o uso de novas tecnologias para estudo do comportamento de compra dos clientes. Para estudar os consumidores de maneira mais profunda, as empresas têm utilizado conhecimento de tecnologia da informação, construindo grandes bases de dados com informações dos clientes (LIM; CHEN; CHEN, 2013; SHAW et al., 2001; NGAI; XIU; CHAU, 2009). A análise destas bases com milhares de dados é um trabalho minucioso que pode ser aprimorado como o uso de aprendizado de máquina para automatizar os processos (BOSE; MAHAPATRA, 2001).

O aprendizado de máquina é uma vertente da área de Inteligência Artificial que estuda métodos computacionais capazes de automatizar o processo de aprendizagem por meio do conhecimento de exemplo, isto é, por meio de uma base de dados de interesse (BOSE; MAHAPATRA, 2001; LANGLEY; SIMON, 1995; MONARD; BARANAUSKAS, ). Na ótica organizacional, o uso deste método já é presente, em algumas empresas, nas seguintes áreas:

1. Marketing, com os estudos de *Customer Relationship Management* (CRM) e uso de *data mining* e aplicações em *big data* (LIM; CHEN; CHEN, 2013);
2. Finanças, principalmente em análise de risco de crédito e previsão de falência em bancos (LANGLEY; SIMON, 1995; LIN; HU; TSAI, 2012);
3. Logística, com previsão de demanda em *supply-chain* (CARBONNEAU; LAFRAMBOISE; VAHIDOV, 2008)

Além das aplicações citadas acima, as implicações do aprendizado de máquina também estão presentes nas áreas da saúde, da biologia, da química e da mecânica (LANGLEY; SIMON, 1995). Com o advento da globalização e das intensas mudanças que as tecnologias têm causado no relacionamento consumidor-organização, as funções do aprendizado de máquina podem ser a chave para o sucesso empresarial.

Neste estudo, o objetivo é aplicar o aprendizado de máquina na Indústria do Turismo, mais especificamente no mercado hoteleiro. Pretende-se criar um mecanismo de auxílio na decisão de se alocar um hotel de acordo com os serviços oferecidos por este e a perspectiva de preço. No Brasil, o mercado possui grande influência na economia e o bom desempenho empresarial da área pode fomentar ainda mais o crescimento desse setor.

Dessa forma, o problema de pesquisa que este trabalho busca averiguar pode ser delimitado por esta pergunta de pesquisa: "De que maneira o Geomarketing aplicado ao Aprendizado de Máquina pode auxiliar na Tomada de Decisão Gerencial relacionado à localização no mercado hoteleiro?"

## **1.2 Objetivo Geral**

Por meio deste trabalho, espera-se desenvolver um novo método para a Tomada de Decisão Gerencial que auxilie a gerência na decisão locacional e na identificação de potenciais oportunidades de negócio. Dessa forma, pretende-se: construir uma Máquina de Suporte Vetorial de Regressão (*Support Vector Regression*) que auxilie na decisão gerencial de localização de hotéis de 11 capitais turísticas, são elas: Brasília, Belo Horizonte, Curitiba, Florianópolis, Fortaleza, Maceió, Manaus, Natal, Porto Alegre, Recife e Salvador.

## **1.3 Objetivos Específicos**

1. Analisar o mercado de hotéis de 11 capitais com potencial turístico utilizando Máquina de Suporte Vetorial;
2. Prever os possíveis locais para alocação de hotéis observando os preços dos hotéis vizinhos e as amenidades oferecidas;
3. Analisar o comportamento dos preços nos bairros das cidades analisadas.

## **1.4 Motivação/Justificativa**

A falta de planejamento formal e a falta e/ou má gestão dos recursos das áreas funcionais são uns dos fatores citados e discutidos na literatura como contribuintes para o insucesso empresarial (ALBUQUERQUE; FILHO, 2011). Para que os negócios das empresas sobrevivam, é preciso observar minuciosamente os mercados em questão, estudando o comportamento da

concorrência e do consumidor, analisando as particularidades do negócio e prevendo a demanda do produto ou serviço. Para contribuir com a análise de mercado, é preciso uso de bases com dados dos consumidores e concorrentes e cerca de 70% dos dados que auxiliam no processo decisório são de caráter espacial (MACHADO; FRANCISCO; RIBEIRO, 2016). O Geomarketing estuda a junção de dados socioeconômicos e espaciais, potencializando assim os estudos de segmentação do mercado e possíveis consumidores.

Com clientes cada vez mais exigentes, as empresas precisam compreender de maneira específica o mercado. Segundo Cliquet (2013), a compreensão do mercado produz uma maior segmentação, forçando as organizações a se adaptar às necessidades de cada segmento de clientes. A junção do estudo da segmentação de mercado relacionado com a localização geográfica traz benefícios para gerência que consegue conhecer de maneira mais completa os anseios de cada segmento de mercado e o perfil dos consumidores locais. Cliquet (2013) argumenta que por muitos anos os pesquisadores se preocuparam em focar no comportamento do consumidor, entretanto, com o desenvolvimento das técnicas de Geomarketing, as empresas, principalmente do setor varejista, têm alargado dos horizontes e, com isso, os aspectos espaciais voltaram a ser prioridades nas pesquisas.

Mesmo com o avanço dos estudos na área, ainda há poucos estudos relacionando *Support Vector Machine*(SVM) e Marketing. Entretanto, segundo Cui e Curry (2005), muitos dos problemas desta área da Administração exigem previsão, o que poderia ser resolvido aplicando SVM nos estudos. Os autores provaram em seus estudos que o SVM é uma ferramenta promissora para previsão de escolhas do consumidor em ambiente de previsão pura, trazendo resultados significantes podendo ser aplicado em várias áreas dentro do Marketing (CUI; CURRY, 2005). Portanto, além da justificativa prática vista no auxílio da análise mercado; há, também, uma justificativa teórica para o trabalho vista na contribuição científica, devido à pouca literatura.

Além disso, a economia mundial tem voltado a sua atenção para a economia de serviços. Este setor tem representado cerca de 70% do PIB (Produto Interno Bruto) nos países desenvolvidos e tem gerado uma grande quantidade de empregos. No Brasil, segundo a última Pesquisa Anual de Serviços - PAS em 2015 (Departamento de Comércio e Serviços IBGE, 2015) realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE), o setor de serviços tem empregado 12,7 milhões de pessoas no país, movimentando cerca de R\$ 1, 4 trilhões em Receita Operacional Líquida.

O Turismo é um setor que tem crescido nos últimos anos e está inserido nos serviços que têm fomentado a economia brasileira. De acordo com a Pesquisa de Serviços em Hospedagem - PSH (Coordenação de Serviços e Comércio IBGE, 2016), realizada em 2016 pelo IBGE, o ingresso

de turistas no Brasil aumentou de 5,4 milhões para 6,6 milhões entre os anos de 2011 e 2016. O crescimento do setor no Brasil atinge todos os elos da cadeia de serviços turísticos, entre eles o serviço de hotelaria/hospedagem que é alvo deste presente estudo.

## 2 REFERENCIAL TEÓRICO

### 2.1 Histórico do Geomarketing

De acordo com Aranha e Figoli (2001), o Geomarketing consiste em muito mais que um simples estudo de pontos comerciais. A união do estudo de dados espaciais com dados socioeconômicos também avalia impacto que o novo ponto comercial causará à rede já consolidada, a contribuição deste novo ponto para a rede em si, viabiliza o estudo da abertura de novos pontos para expansão do mercado, possibilita identificar áreas de influência do mercado e o perfil de cada unidade de negócio.

Os estudos sobre Geomarketing se iniciam com Von Thünen (1966) e a “A Teoria do Estado Isolado”. Thünen introduziu em 1966 um modelo que busca analisar as relações entre mercados, produção e distância geográfica (RODRIGUE; COMTOIS; SLACK, 2013; THUNEN; HALL, 1966). A Teoria do Estado Isolado diz que o uso da terra ao redor das cidades era estipulado pelos custos relativos aos transportes de distintos produtos para o mercado central, ou seja, as atividades com mais produção disputariam pelas terras mais próximas ao mercado central e vice-versa, além disso, as terras com maior proximidade dos mercados centrais fornecem uma maior renda em relação às terras mais distantes. Os estudos de Thünen propuseram que os custos de transporte de produtos agrícolas, indivíduos e conhecimentos entre os ambientes, é melhor maneira de explicar o desenvolvimento dos espaços geográficos (FURLAN, 2011).

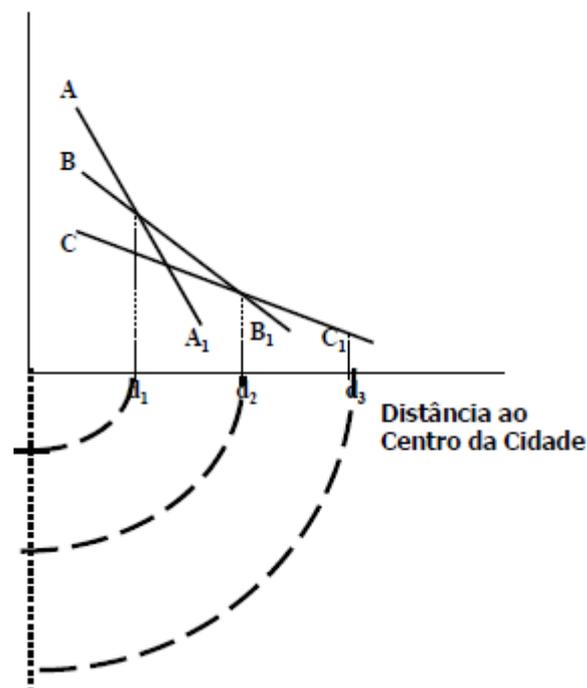
Os estudos resultam nos “Anéis de Thünen”; que, embora muito criticado pelos teóricos da área pela falta de realidade nas suposições, possui importância nos estudos primordiais de Economia de Localização. Para o modelo são feitas as seguintes suposições:

1. Livre concorrência: os agentes utilizam os recursos que estão ao alcance para desenvolvimento comercial, assim o mercado é competitivo e equilibrado;
2. Agentes possuem comportamento racional e procuram maximizar sua utilidade ou lucro: a comunidade é autossuficiente e não possui influências externas;

3. Condições naturais homogêneas: a terra ao redor do mercado é plana e a fertilidade é homogênea;
4. Custos de transporte uniforme: só há um tipo de transporte e este é feito por meio terrestre.

As suposições de Thünen formam o chamado "Estado Isolado". A livre concorrência, o comportamento racional dos agentes e a procura por maximizar a utilidade ou o lucro, no caso dos comerciantes, compõem um mercado isolado de influências externas de outros mercados, o que já impossibilita a prática da teoria, de certa forma, porque os mercados atuais sofrem influências externas devido à intensa globalização. As condições naturais homogêneas estabelecem que os produtores agrícolas possuem as mesmas condições de produção de fertilidade e estrutura do solo e os custos de transporte uniforme garantem que os transportes ocorrem da mesma maneira para todos os produtores, neste caso, o custo de transporte dependeria somente da mercadoria transportada (RODRIGUE; COMTOIS; SLACK, 2013).

As categorias de produção se distribuem ao longo do território formando o que pode ser visto na Figura 2.1



**Figura 2.1** – Anéis de Thünen  
Fonte: Aranha e Figoli (2001)

No anel central, há predominância da categoria comercial. No anel seguinte ao comércio, há predominância da categoria industrial. E, por fim, a curva mais distante do centro predomina

a categoria residencial. O modelo de Thünen teve impacto nas teorias de geomarketing, entretanto a falta de realidade nas suposições faz com que não seja possível aplicar a teoria em sua forma completa (FURLAN, 2011).

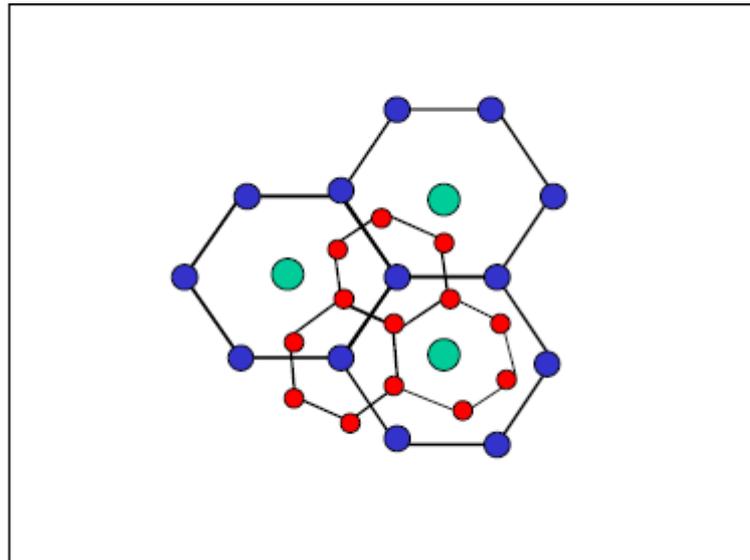
Outro estudo que contribuiu para o histórico do Geomarketing foi a Teoria do Lugar Central de Christaller (1966). Esta Teoria estuda o número, o tamanho, o distanciamento e composição funcional de centros comerciais, inseridos em livre concorrência (ARANHA; FIGOLI, 2001). O autor ainda faz algumas suposições com relação a três fatores:

**Consumidores:** igualmente afluentes, perfeitamente informados, racionais e tomam suas decisões procurando maximizar a sua utilidade. Dessa forma, supõe-se que um consumidor isolado não tenha poder de mudar o mercado.

**Fornecedores:** racionais, tomam decisões procurando maximizar o lucro, operam com custos de produção equivalentes dos competidores e num ambiente onde não há barreira de entrada. Assim como os consumidores, não possuem poder de mudar o mercado.

**Geografia:** supõe-se que o espaço seja homogêneo, ou seja, os custos de transportes são os mesmos para todas as direções e os consumidores e fornecedores possuem distribuição uniforme no espaço. Para o comportamento geográfico dos consumidores, pressupõe-se que estes procuram comprar produtos nos locais mais próximos de si.

De acordo com os pressupostos citados, Christaller (1966) chegou em algumas conclusões. Primeiramente, o aumento nos custos de transporte faz com que a demanda por certos produtos diminua de acordo com o aumento da distância entre os consumidores e os fornecedores, isso significa que a partir de uma estipulada distância, a demanda tende a ir a zero. A máxima distância que os consumidores se propõe a percorrer à procura de um produto chama-se “área de influência” ou “área de amplitude” (CHRISTALLER, 1966). Com a construção destes níveis de influência geográfica, foi criada uma hierarquia das localidades dos centros de comércio conforme a Figura 2.2:



**Figura 2.2** – Hierarquia de Localidades Conforme Teoria do Lugar Central  
Fonte: Aranha e Figoli (2001)

Na Figura 2.2 é possível observar as fronteiras delimitando as áreas de influência de cada centro comercial. A Teoria do Lugar Central possui várias suposições para que ocorra, entretanto não há observações empíricas. Por mais que a teoria contribua para os estudos geográficos e socioeconômicos, a aplicação desta ainda precisa de adaptações à realidade das atividades comerciais devido às várias suposições feitas pelo autor. Entretanto, o modelo criou os fundamentos do que hoje é chamado de Marketing Geográfico (FURLAN, 2011)

Outra teoria que influenciou as vertentes do Marketing Geográfico é a Teoria da Localização Industrial de Weber (1909). Para Weber, a localização industrial deve ser determinada pela minimização dos custos de matéria prima e transporte. O autor estudou a influência, de forma individual, dos custos de transporte, do fator de mão-de-obra e das chamadas forças aglomerativas – forças que são capazes de explicar o movimento de concentração e dispersão de uma indústria em determinada região. O modelo ainda possui um conjunto de simplificações para ser aplicado: não há variações nos custos de transporte no espaço dado, o mercado possui concorrência perfeita – empresas e consumidores possuem conhecimento das condições de mercado -, não há influências externas no mercado e os mercados se localizam em um número certo de centros urbanos (WEBER; FRIEDRICH, 1962; WEBER; FRIEDRICH, 1929; WEBER, 1909)

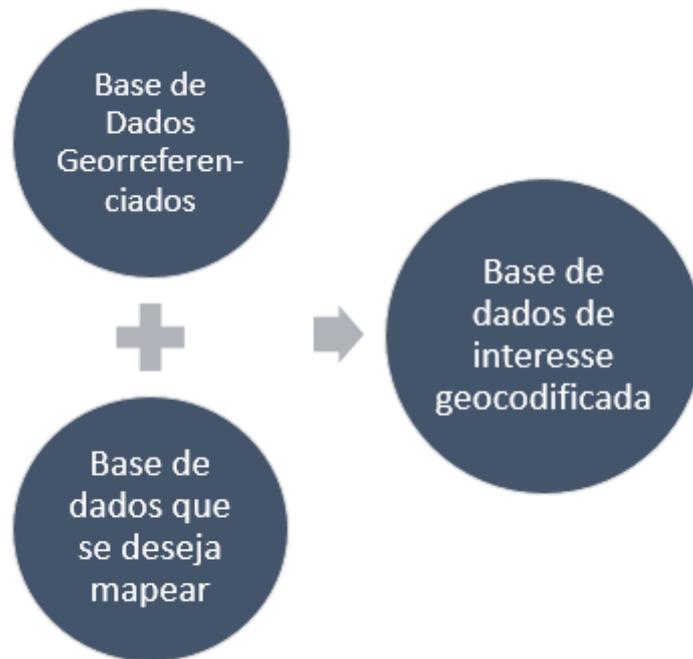
A relação dos três fatores de influência resulta no Triângulo Locacional de Weber. No modelo, procura-se obter a localização que proporcione o menor custo de transporte, ajustando com a necessidade de matéria prima da produção.

O histórico do Geomarketing ainda possui outros autores que colaboraram com teorias e modelos como Hotelling (1929), Reilly (1931) e os Modelos Gravitacionais, Huff (1962) e Losch

et al. (1954) que influenciou os trabalhos de Christaller (1966) na Teoria do Lugar Central. Os estudos destes autores foram muito importantes para a construção e no desenvolvimento do conceito de Geomarketing. As teorias da Economia e da Geografia se mostram aplicáveis até na realidade atual das empresas em relação aos custos de transporte, localização e preços.

Em termos práticos, o Geomarketing é definido como estudo das relações existentes entre as estratégias e ações de marketing e o território onde a empresa, os consumidores e os pontos comerciais se localizam (DAVIES, 1976) . A teoria auxilia as empresas tomar decisões referentes à localização e à relação dos consumidores com espaço geográfico das organizações. As principais aplicações técnicas do Geomarketing são: percepção de nichos de mercado e pontos comerciais, definição de áreas de vendas baseados na demanda e atratividade do consumo dos produtos (MACHADO; FRANCISCO; RIBEIRO, 2016).

De acordo com Machado, Francisco e Ribeiro (2016), para que seja feita uma análise em Geomarketing, primeiramente, é preciso se atentar para duas etapas: existência de uma base de dados georreferenciados de elementos sistemáticos – latitude e longitude – e uma base de dados de objetos da qual se deseja mapear. Com estas entrada de dados, é possível criar uma base que possua as informações de interesse geocodificadas e, assim, dar início às análises propostas. Na Figura 2.3, há exemplificado as etapas para a formação da base de dados geocodificados.



**Figura 2.3** – Base de Dados Geocodificados

Fonte: Elaboração da Autora

Segundo Cliquet (2013) para introduzir os estudos espaciais no marketing, é preciso levar em consideração três esferas principais: comportamento do consumidor, local de vendas, gestão de marketing. Entre essas três dimensões, os pesquisadores de marketing consideram o comportamento do consumidor como a mais importante, cerca de 50% das pesquisas de marketing são focadas nesta dimensão (CLIQUET, 2013). Entretanto, a globalização e a consequente maior mobilidade dos consumidores tem aberto novas perspectivas para os pesquisadores considerarem também nas pesquisas de marketing o comportamento espacial dos clientes, o que proporciona uma maior abertura para os estudos de Geomarketing.



**Figura 2.4** – Esferas do Geomarketing de acordo com Cliquet (2013)  
 Fonte: Elaboração da Autora

No marketing geográfico, procura-se assimilar a forma como análise dos dados geográficos pode ser empregada nas estratégias de marketing das empresas. Segundo Yrigoyen (2003), o Geomarketing consiste na “análise geográfica da realidade socioeconômica através de instrumentos cartográficos e ferramentas de estatística espacial”. A junção do Marketing e da Geografia auxilia na resposta de algumas questões relevantes para o mercado como quem compra os produtos, quais são as características dos consumidores onde são comprados os produtos, qual a fatia de mercado esperada, se é necessário expandir a marca, onde deve ser localizada a marca, como os produtos devem ser promovidos e porque são comprados (CARDOSO, 2011). Latour e Floc’h (2001) descreveram o Geomarketing como sistema integrado de dados, métodos estatísticos e gráficos designado para elaborar informações que sejam úteis para a tomada de decisão dos gestores. Na Tabela 2.1 é possível observar os autores e as respectivas contribuições à definição do Marketing Geográfico.

**Tabela 2.1 – Histórico do Geomarketing**

<b>Autor</b>	<b>Teoria, Modelo ou Definição</b>
Von Thünen (1966)	Teoria do Estado Isolado: esta teoria busca analisar as relações entre mercados, produção e distância geográfica
Christaller (1966)	Teoria do Lugar Central: esta Teoria estuda o número, o tamanho, o distanciamento e composição funcional de centros comerciais, inseridos em livre concorrência
Weber (1909)	Teoria da Localização Industrial: para Weber, a localização industrial deve ser determinada pela minimização dos custos de matéria prima e transporte
Hotelling (1929) Reilly e Huff (1929)	<i>Stability of Competition</i> Modelos Gravitacionais
Davis (1997)	Definiu Geomarketing como estudo das relações existentes entre as estratégias e ações de marketing e o território onde a empresa, os consumidores e os pontos comerciais se localizam
Yrigoyen (2003)	Definiu que o Geomarketing consiste na "análise geográfica da realidade socioeconômica através de instrumentos cartográficos e ferramentas de estatística espacial"
Latour e Floc'h (2001)	Definem o Geomarketing como a junção de dados, de sistemas de processamento, de métodos estatísticos designados para elaborar informações que sejam úteis para a tomada de decisão relacionada com localização

## 2.2 Sistema de Informação Geográfica (SIG)

Observando as várias definições apresentadas de Geomarketing, percebe-se que há uma necessidade de uso de uma grande quantidade de dados e diversos modelos estatísticos para fazer as análises necessárias. As bases de dados utilizadas para estas pesquisas demanda de inúmeras informações dos consumidores, concorrentes e fornecedores; e, por isso, é preciso o uso de softwares cartográficos e de tecnologia da informação o chamado Sistema de Informação Geográfica (SIG).

O SIG consiste em um conjunto de aplicativos computacionais desenvolvidos para lidar com informações relativas à superfície da Terra, além disso, os sistemas têm função de arquivar, adquirir, verificar, recuperar, integrar, consultar, analisar e associar as informações georreferenciadas com fim de gerar conhecimento para tomada de decisão gerencial (FURLAN, 2011). Esse sistema consegue abranger vários tipos de dados e aplicações nas mais diversas áreas do conhecimento, como otimização de tráfego, controle cadastral, gerenciamento de serviços de utilidade pública, demografia, cartografia, administração de recursos naturais, monitoramento

costeiro, controle de epidemias, planejamento urbano e entre outros (CÂMARA, 1996) .

De acordo com Davis e Câmara (2001), existem três formas principais de se utilizar um Sistema de Informação Geográfica: como ferramenta para confecção de mapas, como suporte para análise espacial e como um banco de dados geográficos. O desenvolvimento deste tipo de software possibilita novas aplicações do uso de dados geográficos, resolvendo análises cada vez mais complexas.

O SIG possui uma estrutura geral em que os componentes funcionam de uma forma hierárquica e os componentes são: interface com o usuário, entrada e integração de dados, funções de consulta e análise espacial, visualização e plotagem e armazenamento e recuperação de dados (DAVIS; CÂMARA, 2001). O geoprocessamento de dados, processo anterior ao uso de sistemas computacionais para dados georreferenciados como SIG que consistia em simplesmente tratamento das informações referentes acontecimentos associados à posição na superfície, mostrou-se mais dinâmico nos processos de representação terrestres. (MAGUIRE; GOODCHILD; RHIND, 1991) propõe que as aplicações do SIG devem ser classificadas em: socioeconômicas, ambientais e gerencias.

De acordo com Câmara (1996) as aplicações socioeconômicas ocorrem tanto como análise preliminar como para análise posterior e se distinguem em três grupos: uso do solo, ocupação humana e atividades econômicas. As análises ambientais geralmente ocorrem antes dos sistemas computacionais e se distinguem em dois grupos: meio ambiente e uso de recursos naturais. Para implicações no meio ambiente, pode-se aplicar o SIG para analisar modelagem climática e ambiental, como monitoração de desflorestamento e da emissão e ação de poluentes. No caso do uso de recursos naturais, é possível aplicar na identificação e mapeamento mineral e petrolífero, planejamento e supervisão de redes hidroelétricas, gerenciamento costeiro e marítimo e sistema de informação de recursos hidrológicos. As aplicações de gerenciamento concentram-se principalmente no auxílio à tomada de decisão tanto como forma de determinar novas políticas de planejamento quanto na avaliação das decisões já tomadas. Este tipo de aplicação tem sido muito utilizado para planejamento urbano ou ambiental (CÂMARA, 1996).

O geomarketing se enquadra no Sistema de Informação Geográfica aplicado às análises socioeconômicas mais especificamente no grupo das atividades econômicas que incluem estudos de marketing e indústria. Como já foi explicitado, os estudos locacionais são muito importantes para solucionar várias perguntas sobre o marketing das empresas.

Segundo a visão baseada em recursos de Wernerfelt (1984), o desempenho estratégico das empresas depende da forma como estas lidam com seus recursos, sejam estes tangíveis ou intangíveis. A localização geográfica, para a estratégia empresarial, é considerada um recurso

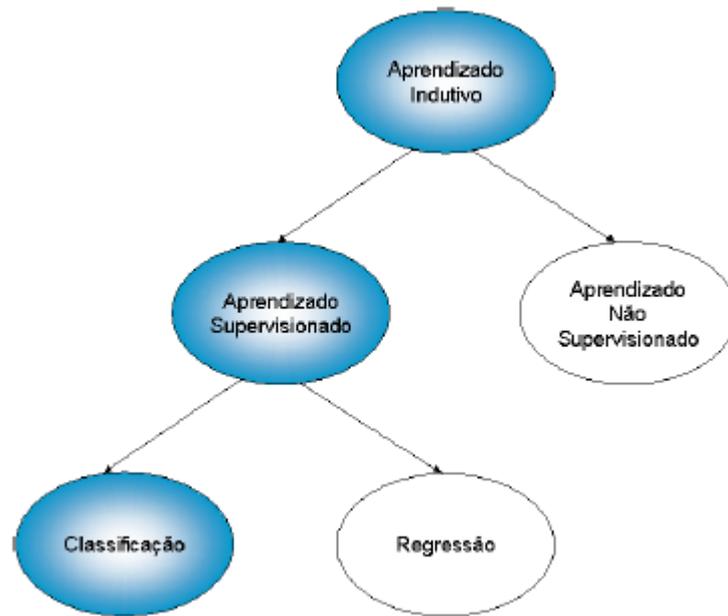
físico das empresas e recurso é tudo aquilo que pode gerar algum tipo de vantagem competitiva para as organizações, ou seja, estratégia que crie valor para a firma e não seja duplicada por concorrentes ou potenciais concorrentes (BARNEY, 1991). Isso significa que, tomada as devidas proporções, as decisões relacionadas à localização geográfica possuem considerável relevância para a sobrevivência das empresas no mercado e, por isso, os estudos de geomarketing são de grande utilidade para as organizações.

## 2.3 Máquina de Suporte Vetorial

Em junção com o SIG, neste trabalho será utilizado Máquina de Suporte Vetorial, em inglês, *Support Vector Machine* (SVM). SVM é uma técnica de aprendizado de máquina que utiliza Teoria do Aprendizado Estatístico, teoria baseado em estudos de Cortes e Vapnik (1995), para procurar um padrão que possa classificar dados em classes distintas ou generalizar o comportamento de uma série de dados. O aprendizado de máquina (AM) é uma área da Inteligência Artificial que desenvolve técnicas computacionais para construção de sistema com capacidade de adquirir conhecimento de forma automática, ou seja, um programa de computador que toma as decisões de acordo com experiências de resolução de problemas anteriores (MONARD; BARANAUSKAS, ; MITCHELL, 1997; CUI; WONG; LUI, 2006).

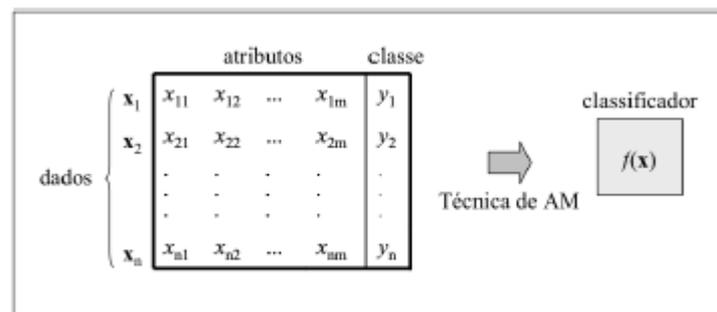
O AM utiliza indução para predizer os eventos futuros, a indução generaliza um raciocínio que se origina de um conceito específico e os conceitos são aprendidos por meio de inferência indutiva, um dos principais métodos para redirecionar conhecimento novo e fazer previsões (MONARD; BARANAUSKAS, ). Para generalizar um conjunto de elementos específicos, o aprendizado de máquina utiliza o Princípio da Inferência Indutiva em que o fenômeno é estudado, cria-se um modelo de explicação deste e, após isso, é postulada uma hipótese geral sobre o fenômeno. O aprendizado indutivo é dividido em três vertentes: supervisionado, por reforço e não-supervisionado. No aprendizado supervisionado, há a necessidade de uso de um indutor (ou professor), ou seja, um conjunto de exemplos de treino para os processos de aprendizagem o que não ocorre no caso do aprendizado não-supervisionado (HAYKIN, 1999). O indutor do aprendizado supervisionado, tem como objetivo construir um classificador que determine a classe de novos exemplos não rotulados no conjunto de treinamento, quando os rótulos de classe são discretos, há um problema de classificação, quando rótulos de classe são contínuos, há um problema de regressão (MONARD; BARANAUSKAS, ). No aprendizado por reforço, não há indutor, dessa forma, o modelo aprende o fenômeno em cada tentativa de solucioná-lo. No aprendizado não-supervisionado, não há presença de um indutor e, com isso, não há conjunto de treinamento prévio. Neste caso, o algoritmo de aprendizado agrupa os insumos recebidos formando

*clusters* (MONARD; BARANAUSKAS, ). Na Figura 2.5, é possível ver as hierarquias do aprendizado indutivo para aprendizado supervisionado e não supervisionado .



**Figura 2.5** – Hierarquia do Aprendizado Indutivo  
Fonte: Monard e Baranauskas ( )

Para este trabalho, o tipo de aprendizado abordado será o supervisionado, logo, o conjunto de exemplos utilizados pelo indutor devem gerar um classificador que terá a capacidade de prever a classe dos novos dados introduzidos. Este primeiro processo em que o indutor gera o classificador de acordo com os dados fornecidos é chamado de treinamento. Na Figura 2.6, tem-se a representação de como é gerado o classificador de forma simplificada. O conjunto de dados com  $n$  observações, cada observação  $x_i$  possui  $m$  atributos e as variáveis  $y_i$  representam as classes. De acordo com os dados e as respectivas classes, o algoritmo de aprendizado de máquina deduz o classificador (LORENA; CARVALHO, 2007).

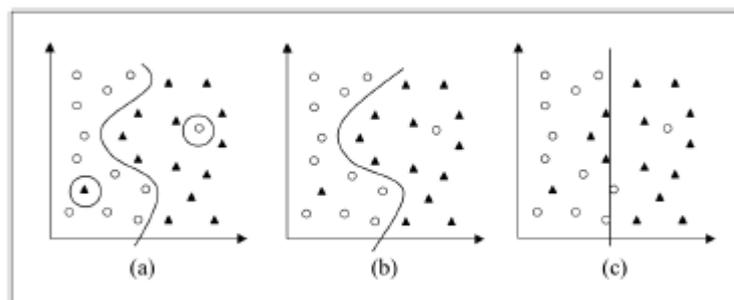


**Figura 2.6** – Indução do classificador em aprendizado supervisionado  
Fonte: Lorena e Carvalho (2007)

Posteriormente à geração do classificador, se faz necessário a validação deste para avaliar

a precisão dos processos e de acordo com essa análise são feitas alterações nos princípios do problema ou na triagem dos dados afim de aprimorar os resultados do classificador. Na validação, são usados outro conjunto de dados, chamados dados de teste. Isso significa que no conjunto de dados de exemplo existem dois subconjuntos de dados: de treinamento e de teste. Os dados de treinamento são utilizados para gerar o classificador e os dados de teste são utilizados para avaliar a efetividade do aprendizado.

A Teoria do Aprendizado Estatístico (CORTES; VAPNIK, 1995) que propõe as condições matemáticas para a escolha dos classificadores a partir do conjunto de dados de treinamento, esta teoria é a principal base do *Support Vector Machine*. Para escolher os classificadores, a teoria analisa o desempenho do classificador nos dados de treinamento e a complexidade para averiguar como o classificador irá manipular novos dados. O conjunto de dados usados por um sistema no processo de aprendizado pode gerar vários classificadores de acordo com os dados de treinamento fornecidos, entretanto o classificador escolhido deve ser aquele que apresenta uma boa capacidade de generalização para que consiga manter o desempenho para dados novos. Isso mostra a importância do uso de um conjunto de exemplos de treinamento representativos em relação ao total de dados, para que não ocorra de situações de *overfitting* – hipótese se ajusta em excesso ao conjunto de treinamento – ou *underfitting* - hipótese se ajusta pouco ao conjunto de treinamento. Como exemplo da importância de escolha do classificador, considere o conjunto de exemplos de treinamento da Figura 2.7. Supõe-se que o sistema de aprendizado procura distinguir os dados em duas classes: círculo e triângulo.



**Figura 2.7** – Três diferentes hipóteses de classificador para um conjunto de exemplos de treinamento

Fonte: Lorena e Carvalho (2007)

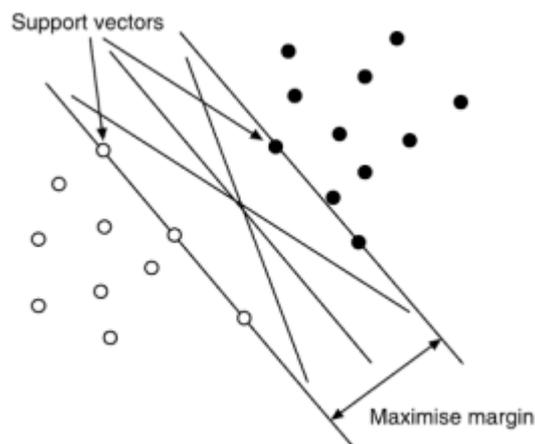
No gráfico 2.7(a), há uma hipótese de classificador que possui grande especificidade de classificação, isto é, tem capacidade de classificar os dados de maneira correta e incluindo os *outliers*, também chamado de ruídos. No caso desta hipótese, a alta especificidade da hipótese pode impedir que o classificador acerte as classes na análise de novos dados, representando a ocorrência de *overfitting*.

No gráfico 2.7(c), é possível é o outro extremo da situação de *overfitting*. Neste caso, a hipótese não se ajusta adequadamente aos dados e, por isso, não classifica os exemplos corretamente, representando a ocorrência de *underfitting*.

Estas duas situações caracterizam o dilema viés-variância. Este dilema demonstra sobre dualidade presente entre precisão dos dados, isto é, o viés das estimativas, e a magnitude do intervalo de confiança, ou seja, a variância entre os dados. Para que a classificação seja precisa, é necessário dosar entre quantidade de dados da amostra e a complexidade do modelo.

Como melhor hipótese para classificar os dados, temos o gráfico 2.7 (b). Nesta hipótese, o classificador consegue separar a maioria dos dados nas classes de maneira correta, sem especificar excessivamente possibilitando o ajuste para dados novos.

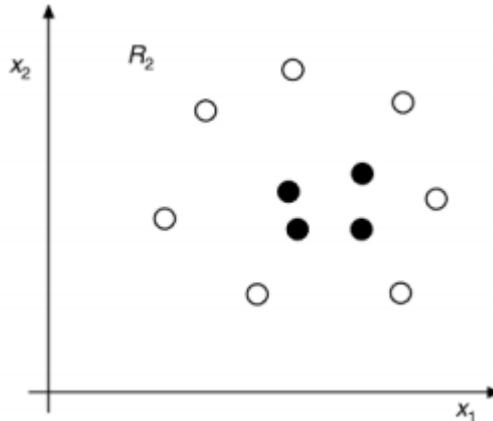
As Máquinas de Suporte Vetorial buscam generalizar uma função utilizando as chamadas função kernel, ou seja, o SVM procura aprender como os modelos ocorrem e generalizar os resultados de dentro da amostra de treinamento para todo o conjunto de dados. O SVM mais simples de ser aplicado é o que busca obter a melhor margem para separação de dois grupos de dados, como é possível ver abaixo, graficamente na Figura 2.8:



**Figura 2.8** – Separador de Máxima Margem  
Fonte: Soman, Loganathan e Ajay (2009)

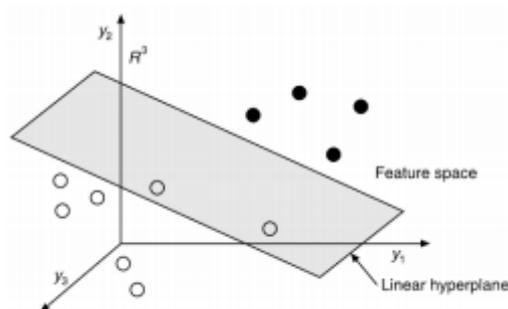
Os pontos que estão na fronteira de separação são chamados de Suportes Vetoriais (*Support Vectors*) que compõem o limite das margens de separação entre os dados. Este tipo de SVM tem por objetivo separar um conjunto de variáveis, na Figura 2.8 os pontos pretos e os pontos brancos tem características distintas, logo é possível fazer a separação de maneira linear com uma reta que esteja equidistante das duas margens. Para os casos em que os dados não podem ser separados linearmente, a generalização da função é feita com uso de funções *kernel*. Neste caso, é feita uma transformação linear nos dados do espaço de entrada para uma

dimensão superior, chamada espaço de característica, isto é,  $x \rightarrow \Phi(x)$  em que  $\mathbb{R}^p \rightarrow \mathbb{R}^q$  tal que  $q > p$ . Nesta dimensão superior  $\mathbb{R}^q$ , os dados são separados por hiper-planos e ao voltar para a dimensão do espaço de entrada  $\mathbb{R}^p$ , os hiper-planos geram uma curva de separação dos dados.



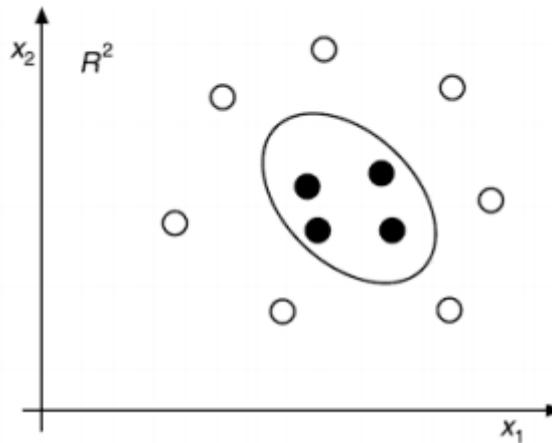
**Figura 2.9** – Espaço de Entrada  
Fonte: Soman, Loganathan e Ajay (2009)

Na Figura 2.9, há o exemplo de espaço de entrada com um conjunto de dados em que não é possível fazer a separação de forma linear. Aplicando a função *kernel* para aumentar a dimensão da função de  $\mathbb{R}^2 \rightarrow \mathbb{R}^3$ .



**Figura 2.10** – Espaço de Característica  
Fonte: Soman, Loganathan e Ajay (2009)

Com a dimensão maior no espaço de característica, é possível enxergar o hiper-plano que separa os dados. Voltando para a dimensão original da função  $\mathbb{R}^3 \rightarrow \mathbb{R}^2$ , o hiper-plano gera uma curva de separação como pode ser visto na Figura 2.11.



**Figura 2.11 – Curva de Separação**  
 Fonte: Soman, Loganathan e Ajay (2009)

O *Support Vector Machine* possui diversas aplicações nas mais variadas áreas do conhecimento. Como o objetivo do SVM é prever questões baseadas em dados conhecidos, este modelo tem grande aplicabilidade nas áreas de Marketing porque muitos dos problemas desta área estão relacionados precisam usar previsões para ser solucionados (CUI; CURRY, 2005). Nos últimos anos, muitos dos estudos de previsão têm se voltado para área de comportamento do consumidor, entretanto estes estudos também podem auxiliar diversos outros fenômenos do marketing para atingir objetivos mais profundos nos modelos mercadológicos e assim gerar resultados mais precisos. Entre os exemplos de fenômenos temos: previsão de demanda, propagandas mais direcionadas e efetivas e questões relacionadas à navegação em websites (CUI; CURRY, 2005).

## 2.4 Máquina de Suporte Vetorial em Marketing

A grande maioria das decisões que precisam ser tomadas na área de marketing necessitam de modelos preditivos e para se obter resultados que sejam úteis para a organização, é preciso que o problema a ser resolvido seja definido da forma mais apurada possível. Cui e Curry (2005) classificam em quatro contextos:

1. Previsão Pura: ocorre em casos onde o conhecimento estrutural do fenômeno não é viável ou necessária;
2. Previsão Robusta: neste contexto, a previsão é fator chave na tomada de decisão e proporcionam uma compreensão intuitiva dos dados para os gestores;

3. Previsão Analítica: contexto em que os modelos são construídos ou adaptados para solucionar um ou mais problemas;
4. Previsão e Análise de *Gaps* Estruturais: neste contexto, os modelos de previsão devem ser altamente precisos para aumentar o conhecimento estrutural do processo organizacional.

As análises e modelos preditivos serão cada vez mais necessários para a tomada de decisão em marketing e as máquinas de suporte vetorial conseguem prever de maneira precisa os ambientes propostos. A globalização e as facilidades atuais de acesso a informação fazem com que os consumidores se tornem cada vez mais exigentes e específicos e, com isso, o mercado de consumidores se encontra cada vez mais segmentado. A segmentação é o processo de divisão de um grande mercado em pequenos grupos de consumidores (HUANG; TZENG; ONG, 2007; CLIQUET, 2013).

Para atender as necessidades dos consumidores, os gestores precisam estar mais atentos no planejamento e nas estratégias de marketing que estejam adequadas com as necessidades dos consumidores e que, ainda assim, consigam alcançar os mais diversos perfis de consumo. O uso de modelos e ferramentas de inteligência artificial pode ser a chave para atingir esses objetivos e solucionar problemas. Huang, Tzeng e Ong (2007) propõe o uso *support vector clustering* (SVC) para a segmentação de mercado, estudando os dados de uma companhia de bebidas. Os estudos compararam os resultados do SVC com outros métodos e comprovaram que o SVC possui mais vantagens de uso para segmentação, como manipular o problema de *outliers* e lidar com diferentes tipos de atributos.

No marketing direto tem crescido a necessidade de análise de dados, devido ao aumento de banco de dados de informações sobre clientes das empresas e os estudos em big data. Esta área do Marketing procura identificar potenciais consumidores de produtos ou serviços e estimular a compra destes por meio de canais (SHIN; CHO, 2006). Utilizando as informações dos consumidores é possível identificar grupos de clientes que com maior potencial de resposta que outros, a partir disso, Shin e Cho (2006) aplicaram as teorias do SVM para construir um modelo-resposta buscando prever a possibilidade de resposta dos consumidores.

Este modelo-resposta proporciona uma redução nos custos globais de marketing. A precisão do modelo é pautada na quantidade de consumidores que responderão de forma positiva à promoção do produto ou serviço, quando maior o número de respondentes, maior a precisão do modelo (SHIN; CHO, 2006).

Dessa forma, o SVM buscar por classificar os consumidores entre respondentes e não-

respondentes de modo a produzir a maior margem entre as classes. Por mais que a metodologia do SVM seja muito útil e ofereça ótimos resultados para o estudo, Shin e Cho (2006) evidenciam que método possui algumas complicações na aplicação devido à grande quantidade de dados de treinamento, ao desequilíbrio de classe e ao *output* binário do SVM.

A grande quantidade de dados de treinamento, na maioria das vezes não tratados, pode influenciar nos resultados da máquina, tornando o método computacionalmente ineficiente (SHIN; CHO, 2006). O desequilíbrio de classe é uma dificuldade recorrente no SVM em *big data*, essa limitação faz com que o SVM não apresente resultados eficientes. Desequilíbrio de classe ocorre em casos que o número da maioria dos resultados supera significativamente a minoria dos resultados causando um viés na classificação das classes (KIM; CHAE; OLSON, 2013; SHIN; CHO, 2006; LING; LI, 1998). No modelo-resposta o resultado do SVM, diferentemente de outros métodos de aprendizado de máquina, fornece um valor binário, isto é -1 ou 1, para solucionar esta dificuldade e não interferir no objetivo de descobrir a propensão de compra do consumidor, Shin e Cho (2006) estimaram a probabilidade com base na distância entre o padrão e o limite de decisão, ou seja, o padrão que está mais distante do limite de decisão possui maior probabilidade de pertencer à classe representada, e assim é possível estimar as probabilidades de compra dos clientes.

Ainda no Marketing Direto, é possível explorar o SVM em conjunto com sistemas especializados para fornecer recomendações de produtos personalizados para cada cliente utilizando os dados das preferências dos próprios consumidores (CHEUNG et al., 2003). Para avaliar as preferências dos consumidores, Cheung et al. (2003) utilizam o *Content-Based System* e *Collaborative System*.

No *Content-Based System*, os produtos são descritos com atributos extraídos das descrições disponíveis e as preferências dos clientes são previstas por meio do estudo da relação entre as classificações do produto dadas pelo consumidor e os atributos correspondentes, neste sistema é possível avaliar os dados usando SVM (CHEUNG et al., 2003). O *Collaborative System* avalia as preferências fazendo recomendações por meio da combinação das classificações de preferências com as recomendações dos chamados "clientes de opinião", para esse sistema Cheung et al. (2003) utilizam um modelo chamado *Latent Class Model*.

Utilizando dados de recomendação de filmes do IMDb e EachMovie, Cheung et al. (2003) analisaram a performance do SVM em comparação com outros modelos e constataram que o SVM conseguiu efetuar um desempenho superior aos outros métodos comumente utilizados como Naive Bayes e Árvore de Decisão. Com a complexidade de espaço e tempo bastante baixas em comparação a outros modelos, o SVM se mostrou muito eficiente na análise ad

preferências dos consumidores (CHEUNG et al., 2003).

## 2.5 Máquina de Suporte Vetorial em Turismo

No Turismo, a principal aplicação do SVM é na modelagem de previsão de demanda turística, neste caso utilizando Máquina de Suporte Vetorial para Regressão (*Support Vector Regression* - SVR). Dentro do Aprendizado de Máquina, muitos são métodos, além do SVM, que se aplicam para prever demanda, entre eles: Algoritmo Genético e Redes Neurais Artificiais (SONG; LI, 2008).

De acordo com Song e Li (2008), a vantagem de se utilizar métodos de Inteligência Artificial (IA), como Redes Neurais, *Fuzzy Logic*, Algoritmo Genético e Máquina de Suporte Vetorial, é que não há necessidade de informações adicionais a respeito dos dados, como distribuição ou probabilidade e os altos graus de precisão gerados pelos métodos. Por mais que as técnicas de IA possuem limitações teóricas de interpretação da demanda turística sob perspectiva econômica, os modelos têm grande potencial de uso nas análises de previsão (SONG; LI, 2008).

A acurácia em previsão de demanda de Turismo é importante para a tomada de decisão de pesquisadores, investidores e gestores, que podem estabelecer estratégias específicas para o setor (HONG et al., 2011; SHELDON; VAR, 1985; GOH; LAW, 2011). Além disso, uma previsão de demanda precisa auxilia gestores empresariais e governamentais a tomar decisões nos níveis operacional, tático e estratégico, baseados nas necessidades da região estudada (HONG et al., 2011). Para Sheldon e Var (1985), a previsão de turismo influencia, principalmente, em três classes de organização: agências governamentais interessadas, organizações de turismo nacionais e regionais e fornecedores de instalações turísticas e viagens.

As técnicas comumente utilizadas para fazer previsão de demanda turística, de acordo com Song e Li (2008) e Goh e Law (2011), são Séries de Tempo e Métodos Econométricos. Entretanto, os mesmos autores evidenciaram o potencial dos métodos de Inteligência Artificial para prever demanda, mostrando que estes modelos têm desempenhos tão bons, e por vezes até melhores, que as outras técnicas na literatura acadêmica.

## 2.6 Turismo e Hotelaria

O setor hoteleiro abrange diversas atividades dentro dos serviços turísticos, como alimentação, entretenimento e hospedagem. O segmento emprega um número muito grande de fun-

cionários devido às várias atividades envolvidas no serviço. Uma dificuldade inerente, não só a este setor, mas como em todo o Mercado de Turismo é a sazonalidade da demanda que depende das altas temporadas (BRANCO; RIBEIRO; TINOCO, 2010).

Para que o setor turístico se desenvolva e cresça dentro de suas perspectivas, é necessário se atentar para a hotelaria. Dependendo da qualidade e da extensão da hotelaria presente na cidade, isto vai representar o tipo de viajantes locais e possibilita melhores análises estratégicas para o setor (COOPER et al., 2000). Há vários tipos de hospedagem possível, como hotéis, *flats*, pousadas, hostels, alugueis por temporada, *bed & breakfast* e entre outros. A escolha do tipo de acomodação depende das preferências do consumidor diante das opções apresentada e dos serviços oferecidos.

A escolha do acomodação interfere na estrutura turística antes do cliente chegar ao destino, logo, é um serviço que requer investimentos para proporcionar desenvolvimento do destino turístico (COOPER et al., 2000). Desde da década de 90, a indústria hoteleira do Brasil tem sido fomentada com investimentos, principalmente com a chegada de grande grupos hoteleiros ao país, o que tem tornado o mercado cada vez mais competitivo (DONAIRE; SILVA; GASPAR, 2009).

Como este setor possui influência em várias economias, a brasileira inclusa, há a necessidade de um constante fomento à inovação no mercado para que os serviços tragam vantagem competitiva para o negócio. Para isso, os gestores precisam estar atentos às novas tecnologias que possibilitem a adoção de estratégias que tragam vantagem competitiva para o empreendimento (CRUZ; GÂNDARA, 2003; LICHROU; O'MALLEY; PATTERSON, 2008).

## 3 METODOLOGIA

Nesta sessão será trabalhado o método da pesquisa e a natureza dos dados utilizados para análise. Entre os métodos de Aprendizado de Máquina apresentados, para ser aplicado no contexto desde pesquisa foi escolhido a Máquina de Suporte Vetorial para Regressão (*Support Vector Regression - SVR*). Além disso, será construído um Modelo Conceitual para explicar a relação entre o Geomarketing e o Mercado de Turismo.

### 3.1 *Support Vector Regression - SVR*

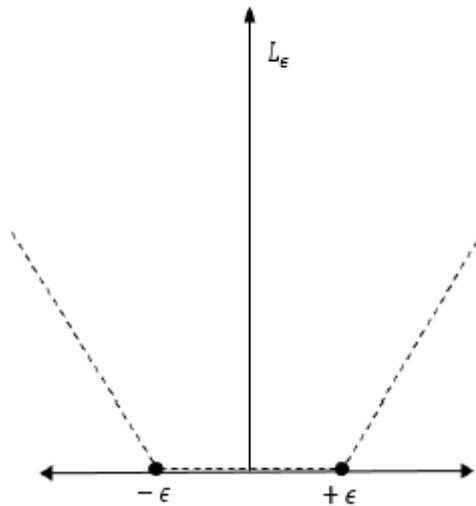
O SVR é uma aplicação do *Support Vector Machine (SVM)*, a diferença entre as duas metodologias está no *output* do SVR e do SVM. No SVM, o algoritmo de aprendizado supervisionado tem por objetivo discriminar os dados em classes de maneira binária, ou seja, +1 e -1 ou 0 e 1, também há casos de aprendizado não-supervisionado, em que o objetivo do algoritmo também é separar em classes, não de forma binária, mas de acordo com o comportamento dos dados. No SVR, o *output* do algoritmo de aprendizado é uma estimativa numérica referente às variáveis estudadas, ou seja, pelo uso de modelos de regressão é possível fazer a classificação de infinitos valores reais. Nesta aplicação do SVM, o objetivo é encontrar uma função  $f(x_i)$  em que os resultados da previsão, isto é,  $y_i$ , não desviem demasiadamente da tolerância desejada,  $\epsilon$  (BASAK; PAL; PATRANABIS, 2007; SMOLA; SCHÖLKOPF, 2004). O SVR pode ser utilizado como metodologia de previsão em diversas áreas, entre elas, previsão de ativos financeiros (YEH; HUANG; LEE, 2011) e, também, previsão de demanda turística (HONG et al., 2011).

Para fazer as previsões nos algoritmos de SVM é necessária uma função de decisão para classificação com objetivo de maximizar a margem entre as classes. O algoritmo do SVR utilizará uma matriz de dados  $\mathbf{X}$  de dimensão  $l \times p$  fazer previsão de uma variável  $y_i \in \mathbb{R}^q$  de forma que  $q > p$ . Como no SVR, não há separação de classes, é preciso selecionar uma função de perda em casos que a previsão foge do "banda de tolerância", isto é, do parâmetro escolhido como limite do erro. Esta tolerância  $\epsilon$  seria uma região de permissibilidade da função. Caso algum dado ultrapasse essa região, a função é penalizada por uma constante de custo.

A função de perda característica dos modelos de SVR é dada pela seguinte estrutura:

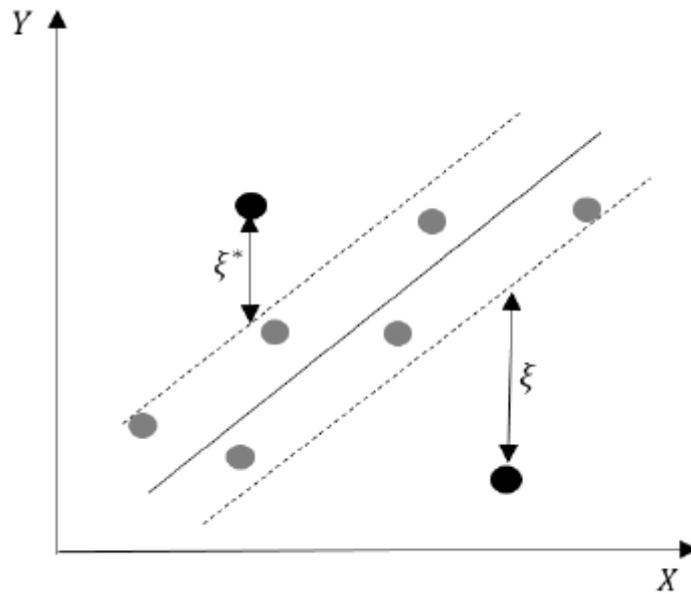
$$L_{\epsilon}(y_i, f(\mathbf{x}_i)) = \begin{cases} 0 & , \text{ se } |y_i - f(\mathbf{x}_i)| \leq \epsilon \\ |y_i - f(\mathbf{x}_i)| - \epsilon & , \text{ caso contrário} \end{cases} \quad (3.1)$$

Graficamente, a expressão se comporta da seguinte forma:



**Figura 3.1** – Representação gráfica da função de perda  
Fonte: Albuquerque (2014)

A função de perda é representada por  $\xi$ , ou seja,  $\xi = y_i - f(\mathbf{x}_i)$ . Esta função é uma forma de penalizar as variáveis que estão fora do limite do erro estabelecido previamente no modelo. O  $\xi$  pode ser ocorrer como um erro para "mais" ( $\xi^*$ ) ou como um erro para "menos" ( $\xi$ ) em ambos os casos, os erros são penalizados com a constante de custo. Determinar o grau de perda, isto é, o limite do erro é importante para que o SVR encontre uma função de decisão que se adeque melhor ao conjunto de dados. A figura 3.2 mostra uma representação gráfica do SVR na forma linear em que é possível ver o comportamento de  $\xi$  na função.



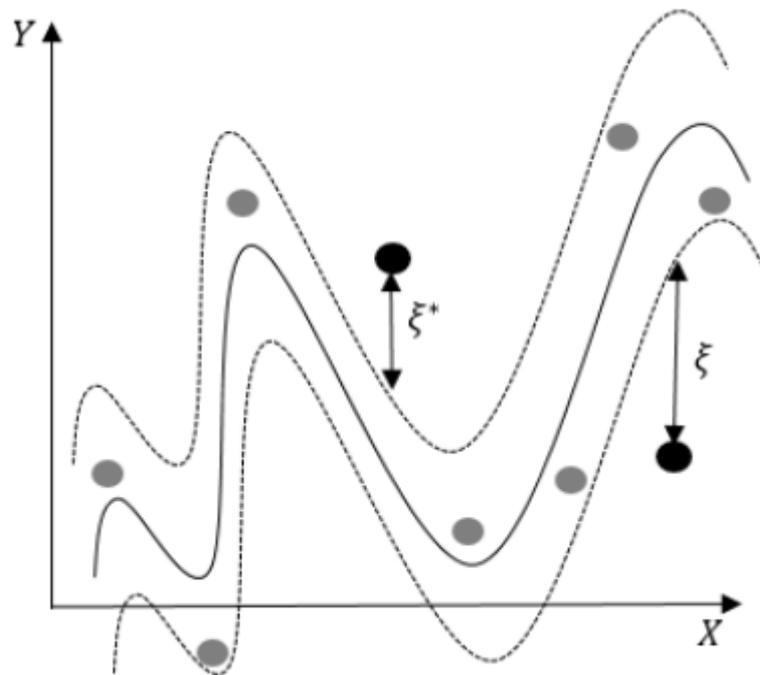
**Figura 3.2** – Representação Gráfica do SVR na forma linear  
Fonte: Albuquerque (2014)

Para o caso não-linear, presente em várias situações nas distribuições de dados reais, o SVR se comporta da seguinte maneira representada na Figura 3.3.

A função de decisão característica de *Support Vector Regression*, conhecida como forma primal do problema de otimização, é descrita abaixo:

$$\begin{aligned}
 &\text{Minimizar} && \frac{1}{2}\omega' \omega + C\mathbf{1}'\xi_i + C\mathbf{1}'\xi_i^* \\
 &\text{Sujeito a} && y_i - [\phi(\mathbf{x}_i)' \omega - \gamma] \leq \varepsilon + \xi_i^*, \text{ para } i = 1, \dots, l \\
 &&& [\phi(\mathbf{x}_i)' \omega - \gamma] - y_i \leq \varepsilon + \xi_i, \text{ para } i = 1, \dots, l. \\
 &\text{Com} && \xi_i \geq 0; \xi_i^* \geq 0
 \end{aligned} \tag{3.2}$$

Na Equação 3.2,  $\frac{1}{2}\omega' \omega$  representa o recíproco da margem, isto é, o contrário da margem. Quando se minimiza o recíproco, maximiza-se a margem, no caso do SVR a banda de tolerância, para que os vetores de suporte das fronteiras alcancem o máximo de dados possível sem prejudicar o modelo. A constante de custo  $C$  é uma variável de folga para penalizar os desvios da função, ou seja, para os dados que se encontram fora da região de permissibilidade  $\varepsilon$  será computado uma penalização à função de decisão. Esta constante é o *trade-off* entre a complexidade da função e a escolha da tolerância do erro, representada pela diferença entre o  $y_i$  que está fora de  $\varepsilon$  e o limite da função  $f(\mathbf{x}_i)$  (SMOLA; SCHÖLKOPF, 2004). A constante de custo influencia na complexidade da função em vista do dilema viés-variância citado na seção anterior.



**Figura 3.3** – Representação Gráfica do SVR na forma não linear  
Fonte: Albuquerque (2014)

Nas restrições do problema de otimização,  $y_i$  representa os dados previstos pelo modelo.  $\phi(\mathbf{x}_i)^t \omega$  e  $\gamma$  representam a função  $f(\cdot) = \omega^t \phi(\cdot) - \gamma$  que mapeia os dados  $\mathbf{x} \in \mathbb{R}^l$  para uma outra dimensão superior conhecida ou representada por um espaço de Hilbert por meio de um kernel.

A seguir, temos os problema primal do SVR escrito na forma matricial:

$$\begin{aligned}
 &\text{Minimizar} && \frac{1}{2} \omega^t \omega + C \mathbf{1}^t \xi_i + C \mathbf{1}^t \xi_i^* \\
 &\text{Sujeito a} && -\mathbf{y} - [\phi^t \omega - \gamma \mathbf{1}] + \epsilon \mathbf{1} + \xi^* \geq \mathbf{0} \\
 &&& [\phi^t \omega - \gamma \mathbf{1}] - \mathbf{y} + \epsilon \mathbf{1} + \xi \leq \mathbf{0} \\
 &\text{Com} && \xi_i \geq \mathbf{0}; \xi_i^* \geq \mathbf{0}
 \end{aligned} \tag{3.3}$$

Para facilitar a resolução do problema de otimização descrito nas Equações 3.2 e 3.3, Smola e Schölkopf (2004) sugerem que seja construído a Forma Dual de Wolfe (1961) do problema e esta pode ser formulada seguindo alguns passos. Primeiro, é preciso aplicar os Multiplicadores de Lagrange no problema primal; depois, deve-se encontrar as condições de Karush Kuhn Tucker e, por fim, substituir os resultados das condições na função. O resultado é a Forma Dual de

Wolfe (1961) explicitada abaixo.

$$\mathcal{L} = -\frac{1}{2}(\lambda^{*t} - \lambda^t)\phi^t\phi(\lambda^* - \lambda) + \begin{pmatrix} \lambda^{*t} & \lambda \end{pmatrix} \begin{pmatrix} +\mathbf{y} - \mathbf{1}\epsilon \\ -\mathbf{y} - \mathbf{1}\epsilon \end{pmatrix} \quad (3.4)$$

Na Equação 3.4, os multiplicadores de Lagrange tem restrições quanto à constante de custo  $C$ ,  $0 \leq \lambda_i$ ,  $\lambda_i^* \leq C$  e  $\sum_{i=1}^l (\lambda_i^* - \lambda_i) = 0$ .  $\phi^t\phi$  representam a matriz associada a função *kernel*. As funções *kernel* são utilizadas no SVM quando os dados não permitem ser classificados de maneira linear, o que acontece frequentemente em distribuições reais. Neste caso, é feito um mapeamento da função para uma dimensão superior em que os dados podem ser separados por um hiperplano linear (SOMAN; LOGANATHAN; AJAY, 2009). Para o SVR, a função *kernel* é utilizada para solucionar os problemas de regressão linear dentro no espaço de dimensão superior (CHEN; WANG, 2007).

Para este estudo, serão utilizadas duas funções *kernel*: *Mexican Hat* e *Polynomial*.

## 3.2 Kernel Mexican Hat

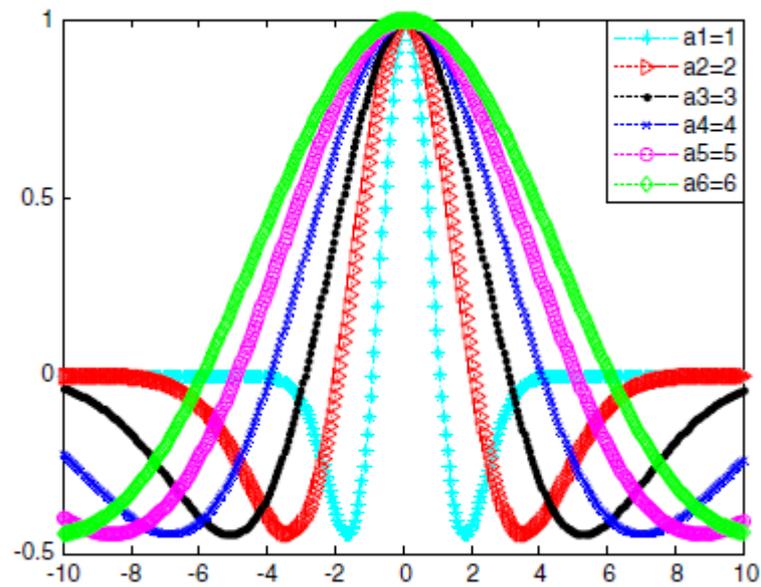
A função *Mexican Hat* faz parte da família das chamadas funções de Ondaletas (*Wavelets*). A *Mexican Hat* possui a seguinte estrutura na Equação 3.5, de acordo com Ding, Wu e Shi (2014):

$$\Psi(x) = (1 - x^2)\exp\left(-\frac{1}{2}x^2\right) \quad (3.5)$$

A função *Kernel* do *Mexican Hat* que satisfaz as Condições de Mercer e a expressão  $K(x_i, x_j) = \phi^t(x_i)\phi(x_j)$  está representada abaixo na Equação 3.6:

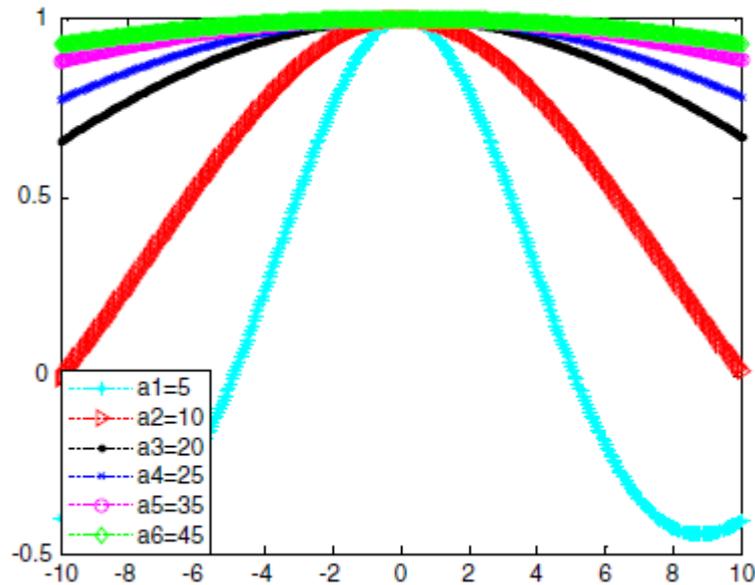
$$K(x, z) = \prod_{i=1}^d \left(1 - \left(\frac{x_i - z_i}{a_i}\right)^2\right) \exp\left(-\frac{1}{2}\left(\frac{x_i - z_i}{a_i}\right)^2\right) \quad (3.6)$$

Nas figuras 3.4 e 3.5 abaixo, é possível ver como o *Kernel Mexican Hat* atua para diferentes valores de  $a$ , onde  $d = 1$  e o *test point* é igual a 0, 1.



**Figura 3.4** – *Kernel Mexican Hat*  
 Fonte: Ding, Wu e Shi (2014)

Com a Figura 3.4, fica nítida a observação de que o *Kernel Mexican Hat* possui uma capacidade de aprendizado superior em casos em que os valores de  $a$  são menores. Entretanto, a Figura 3.5 também mostra a boa capacidade de generalização do *Kernel* em situações em que os valores de  $a$  são maiores. Isto mostra a importância de se selecionar parâmetros que realmente representem os dados reais (DING; WU; SHI, 2014). Segundo Ding, Wu e Shi (2014), as funções da Família *Wavelets* possuem capacidade de aprendizado e eficiência na generalização melhores em casos em que os parâmetros escolhidos para o modelo são apropriados para o conjunto de dados.



**Figura 3.5 – Kernel Mexican Hat**  
 Fonte: Ding, Wu e Shi (2014)

### 3.3 Kernel Polynomial

O *Kernel Polynomial* é muito utilizado para soluções em *Support Vector Machine*, sendo um dos *Kernels* mais comuns nas pesquisas da área em que os modelos não são lineares (GUNN et al., 1998). Este *kernel* faz parte da classe dos *kernels* não estacionários, ou seja, dos *kernels* que dependem de dois exemplos (GENTON, 2001). A estrutura da Função está representada abaixo na Equação 3.7:

$$K(x_i, x_j) = (ax_i^t x_j + c)^d \quad (3.7)$$

Em que,  $x_i$  e  $x_j$  são os vetores do espaço de entrada e  $a, c, d$  são parâmetros do *kernel*. Na maioria dos casos do SVM que procura soluções para problemas binários, os valores de  $a$  e  $c$  assumem o valor de 1 resultando na expressão  $K(x_i, x_j) = (x_i^t x_j + 1)^d$ .

### 3.4 Modelo Conceitual - Geomarketing e Turismo

Para fazer as análises necessárias, neste estudo é proposto um Modelo Conceitual para relacionar as variáveis de Geomarketing e Turismo. Este trabalho procura propor um método quantitativo para auxílio da decisão gerencial de localização, ou seja, que consiga prever um lugar ótimo para a alocação do negócio, neste caso de um hotel. Como variáveis a ser estudadas

neste modelo, foram escolhidas: a localização (latitude e longitude), as amenidades do hotel (serviços oferecidos pelo estabelecimento) e o preço.

A localização será utilizada como variável independente do modelo, pelo qual será inserido os estudos de Geomarketing no contexto turístico. O local onde o hotel se encontra influencia na compra do consumidor dependendo das preferências deste. No modelo, a localização será agregada aos serviços prestados no hotel (amenidades) para determinar se o cliente escolherá ou não o hotel em questão.

As amenidades, como já citado acima, são os serviços que o hotel oferece a seus hóspedes como *Wi-Fi* e Café da manhã incluso. De acordo com Chan e Wong (2006), a qualidade dos serviços prestados pelo hotel tem mais influência na preferência do cliente até que a redução dos preços. Por esses clientes, entende-se que são viajantes independentes que tomam decisões individualmente, sem o uso de agência de viagens ou grupos de turismo. Estes clientes viajam a trabalho ou a lazer e isso influencia as suas preferências de escolha e na percepção dos atributos do hotel (CHAN; WONG, 2006).

Para os dois tipos de viajantes, os atributos de escolha mais importantes são a localização, a atitude dos funcionários e os quartos (CHAN; WONG, 2006). O preço será utilizado como variável dependente do modelo, esta variável atua na percepção de valor do cliente em relação aos serviços, ou seja, atua na expectativa dos consumidores. Quando o cliente paga um valor alto por um serviço, este espera que este também a qualidade elevada, ou seja, há uma alta expectativa. Entretanto quando o cliente não paga um valor tão alto, a expectativa do serviço não é tão elevada (CHAN; WONG, 2006).

Para este modelo de previsão, pretende-se prever o preço do hotel de acordo com a localização, as amenidades oferecidas pelo hotel em questão, as amenidades dos hotéis vizinhos e o preço dos hotéis vizinhos.

A Equação 3.8 abaixo, representa a função que será aplicada ao SVR com as variáveis de interesse:

$$y = f(\mathbf{W}\mathbf{y}, \mathbf{X}, \mathbf{W}\mathbf{X}) \quad (3.8)$$

A Equação 3.8 está em forma matricial, em que  $\mathbf{W}\mathbf{y}$  representa a média de preço dos hotéis vizinhos, calculada pela multiplicação da matriz de preços  $\mathbf{y}$  e da matriz  $\mathbf{W}$  com a distância entre os hotéis da cidade, dessa forma é possível descobrir os preços ponderados pela localização. O  $\mathbf{X}$  representa as amenidades dos hotel escolhido. O  $\mathbf{W}\mathbf{X}$  representa as amenidades dos hotéis vizinhos relacionadas com a distância entre eles, calculada pela multiplicação das matrizes  $\mathbf{x}$

que consta as amenidades dos hotéis próximos e  $w$  com a localização dos hotéis próximos  $y$  representa o preço do hotel.

### 3.5 Dados do Modelo

O objetivo deste estudo é propor um método quantitativo que auxilie na tomada de decisão locacional de gestores de Hotéis. Para tal, o modelo necessita de uma base de dados georreferenciada do mercado hoteleiro da cidade estudada e a malha territorial da cidade para que seja gerado os mapas.

Para esta pesquisa, foram utilizados dados do *site TripAdvisor* coletados entre os meses de Agosto, Setembro e Outubro. Foram coletados dados de preços da diária de cada hotel, das amenidades do estabelecimento e da localização - latitude e longitude - de hotéis em cidades turísticas brasileiras. Além disso, para elaborar os mapas foram usadas as malhas territoriais do censo de 2010 e para fazer as análises do modelo conceitual e aplicar o SVR aos resultados, foi utilizado o *software* livre R.

Para a analisar a realidade turística do Brasil, foram escolhidas as capitais que possuíam mais de 50 hotéis dentro do município. Neste estudo, foram analisadas as seguintes cidades: Brasília, Belo Horizonte, Curitiba, Florianópolis, Fortaleza, Maceió, Manaus, Natal, Porto Alegre, Recife e Salvador

O parâmetro  $C$  do *Support Vector Regression* é definido pelo usuário do modelo. Este parâmetro está relacionado com os erros de previsão dos dados, isto é, ele determina a influência dos erros dentro do modelo, se em alto ou baixo grau, por este motivo é chamado de parâmetro de regularização. O parâmetro selecionado para compor o modelo é aquele que obtiver o menor Erro Quadrático Médio - EQM (*Mean Square Error* - MSE) para que a máquina obtenha valores mais próximos da realidade. O MSE é o somatório dos quadrados da diferença entre os valores de previsão ( $x_i$ ) e dos valores realmente observados ( $\hat{x}_i$ ), dividido pelo número total de observações ( $n$ ).

$$MSE = \sum_i^n \frac{(x_i - \hat{x}_i)^2}{n} \quad (3.9)$$

Como foi abordado na seção anterior, os *kernels* selecionados para análise são o *Mexican Hat* e o *Polynomial*. Na Tabela 3.1 a seguir, estão os parâmetros e o MSE utilizados para o *Kernel Mexican Hat* em cada uma das cidades analisadas.

**Tabela 3.1** – Tabela de Parâmetros - *Kernel Mexican Hat*

<b>Cidade</b>	<b>Parâmetro</b>	<b>MSE</b>
Brasília	0,55	70583,13685
Belo Horizonte	5	152914,3996
Curitiba	5	322217,271
Florianópolis	0,45	621291,1204
Fortaleza	0,5	283791,9072
Maceió	0,45	257876,5337
Manaus	5	67465,78031
Natal	5	246689,7601
Porto Alegre	0,45	133653,3733
Recife	5	42156,56169
Salvador	0,6	629992,0156

Para o *Kernel Polynomial*, são utilizados dois parâmetros. Na Tabela 3.2, encontra-se os parâmetros encontrados para cada cidade de acordo com MSE correspondente.

**Tabela 3.2** – Tabela de Parâmetros - *Kernel Polynomial*

<b>Cidade</b>	<b>Parâmetro 1</b>	<b>Parâmetro 2</b>	<b>MSE</b>
Brasília	5	1	54071,14167
Belo Horizonte	2,9	1	100184,2214
Curitiba	1,25	1	230692,4728
Florianópolis	0,1	1	568366,4849
Fortaleza	4,6	1	273544,9026
Maceió	0,15	2	160285,8994
Manaus	4,35	1	35807,53683
Natal	3,55	1	112434,3941
Porto Alegre	1,75	1	57252,20317
Recife	3,55	1	30081,78961
Salvador	5	1	524001,6852

Para fazer a análise de cada cidade, escolheu-se um perfil hipotético de cliente do mercado hoteleiro para selecionar as amenidades que mais lhe interessa. Para este trabalho, o perfil de cliente desejado selecionou as seguintes amenidades para um hotel: *Free Wi-fi*, Estacionamento, Café da Manhã incluso, Ar Condicionado, Piscina, Restaurante, Serviço de Quarto, Portaria, Lavanderia, Quarto para Família, Funcionários Multilíngue, Translado para Aeroportos e, para o caso de cidades no litoral, a Proximidade da Praia. Pela seleção das amenidades, é possível inferir que o objetivo da viagem do turista é lazer.

## 4 RESULTADOS

### 4.1 Análise do Modelo

O modelo analisa como o preço dos hotéis vizinhos, as amenidades oferecidas pelo hotel e as amenidades oferecidas pelos hotéis vizinhos influenciam no preço do hotel estudado. O *Support Vector Regression* atua na previsão das localizações mais propícias mais o estabelecimento de acordo com as características dos bairros dentro da cidade. Dependendo do preço e das amenidades dos hotéis da região o algoritmo consegue prever que tipos de hotéis que se adequam à região de acordo com o preço da diária. Como o resultado da previsão é gerado um mapa de preços da cidade estudada, mostrando os lugares onde se deve alocar novos hotéis de acordo com a região e o preço mais propício para aquela região.

#### 4.1.1 Brasília - DF

A primeira cidade analisada é Brasília - DF, a seguir podem ser visto os Mapas gerados pelo modelo nas Figuras 4.1 e 4.2.

O Mapa da Figura 4.1 foi gerado aplicando o *Kernel Mexican Hat*, a faixa de preço prevista é discriminada por cores, os preços assinalados com cores mais avermelhadas são os maiores preços do intervalo e os preços assinalados com cores azuladas são os menores preços do intervalo. Para o Mapa do *Kernel Mexican Hat*, o resultado da previsão gerou um intervalo de preço entre R\$ 192,50 e R\$ 182,50. Em Brasília, as regiões com maiores preços - mais avermelhadas - situam-se nos bairros mais nobres da cidade, como Lago Sul, Lago Norte e Sudoeste. As regiões com menores preços - mais azuladas - situam-se em bairros no sudoeste do Distrito Federal, como Taguatinga, Gama e Valparaíso.

O Mapa da Figura 4.2 foi gerado pelo *Kernel Polynomial*. A faixa de preço prevista por este *Kernel* tem uma maior amplitude que a previsão do *Kernel Mexican Hat*. Neste caso, os preços variam de R\$ 2000,00 a R\$ 500,00. Segundo a previsão, as regiões periféricas ao centro de Brasília apresentaram os preços mais baixos e no centro da capital estariam concentrados as

diárias mais caras, logo, mais propícia para hotéis de luxo. Por mais que os preços previstos, não estejam tão de acordo com a realidade da cidade, o algoritmo conseguiu generalizar o comportamento socioeconômico da cidade, prevendo o preço da diária dos hotéis de acordo com as características socioeconômicas de cada região.

Para a cidade de Brasília - DF, a generalização do modelo mostrou que as regiões mais distantes do centro da cidade possuem os preços mais baixos e as regiões mais próximas do centro e dos pontos turísticos possuem os preços mais altos.

### Hotel prices – Mexican Hat

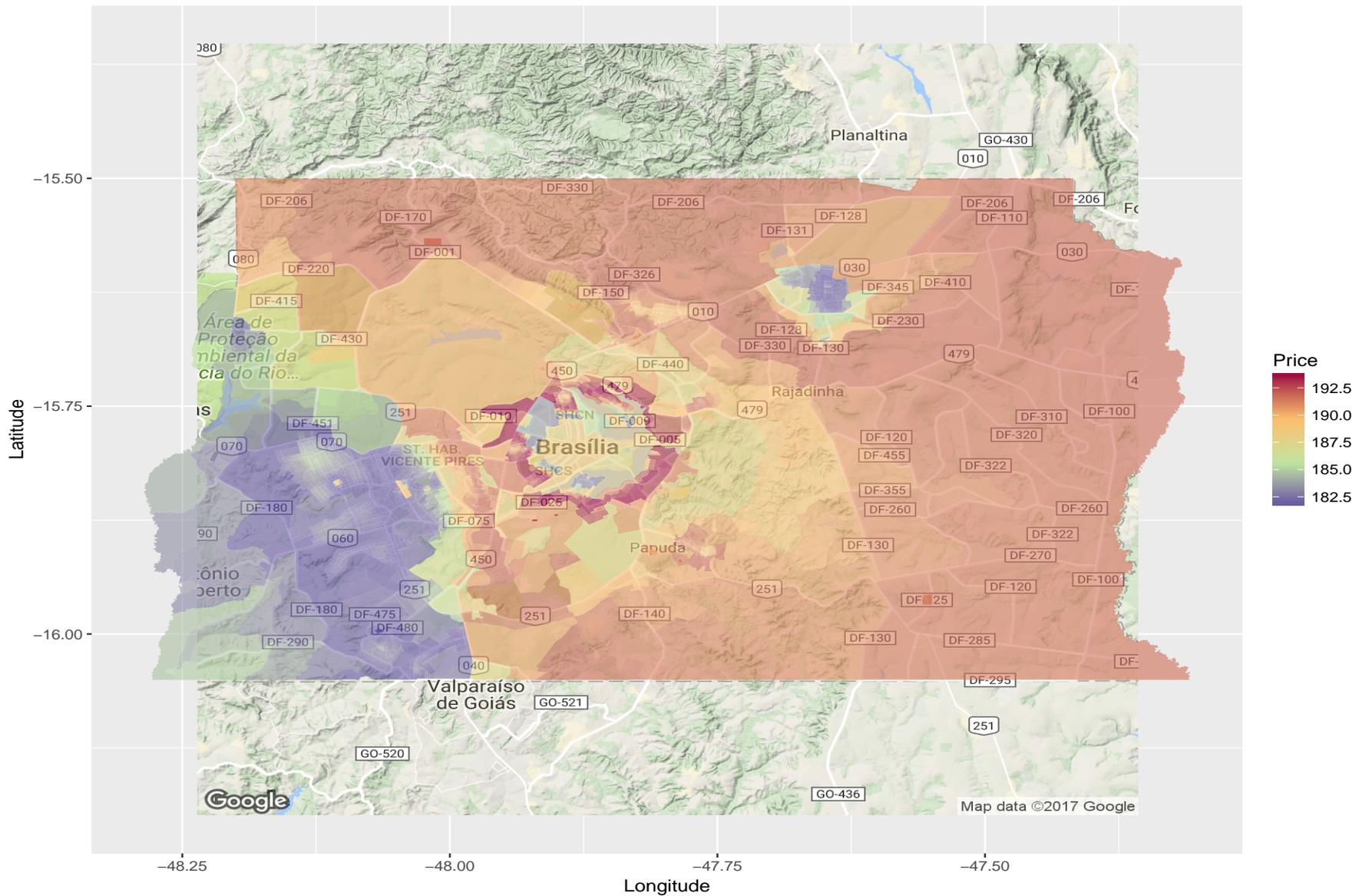


Figura 4.1 – Mapa de Preços - Brasília - Kernel Mexican Hat

### Hotel prices – Polynomial

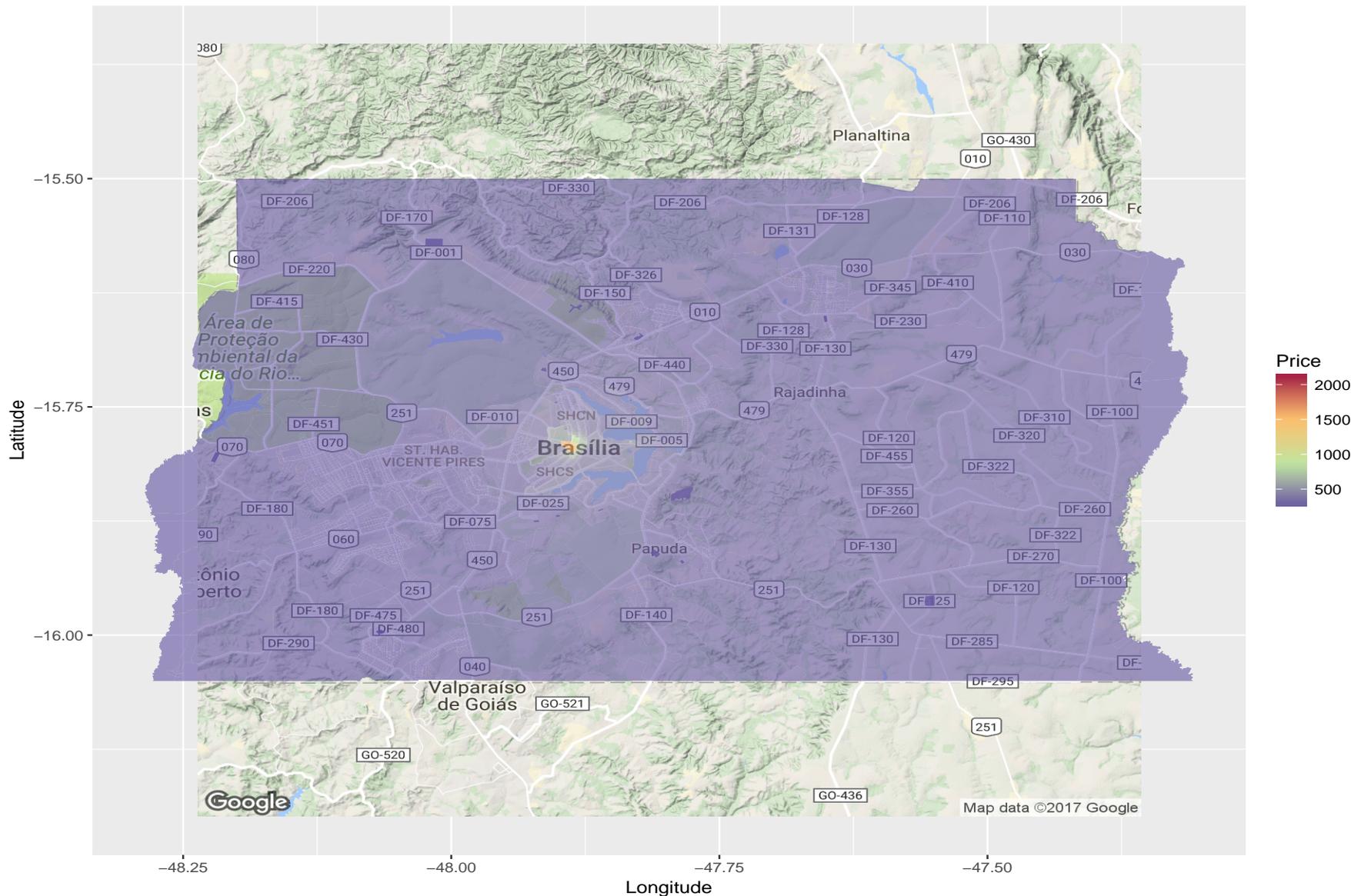


Figura 4.2 – Mapa de Preços - Brasília - Kernel Polynomial

### 4.1.2 Belo Horizonte - MG

Para a cidade de Belo Horizonte - MG, os Mapas gerados pelo modelo podem ser vistos a seguir nas Figuras 4.3 e 4.4

O Mapa da Figura 4.3 foi gerado pelo *Kernel Mexican Hat*. Para esta cidade, a faixa de preço prevista variou entre R\$ 149,00 e R\$ 144,00, com uma baixa amplitude de preços. Os bairros com maior incidência de variação de preços são os bairros mais próximos do Centro, isto é, Santa Efigênia, Funcionários, Savassi e entre outros. Nestes bairros, a discrepância de preços é evidente no Mapa, aos arredores do Centro, há o indicativo de maiores preços - assinalados em vermelho - iguais ou superiores a R\$ 149,00. Entretanto, na região central, os preços se mantiveram mais baixos - assinalados em azul e verde - variando entre R\$ 146,00 e R\$ 144,00. Outras dois bairros foram evidenciadas no Mapa por preços mais altos: Pampulha e União. Os bairros restantes permaneceram com a mesma média de preço, assinalados de laranja.

O Mapa da Figura 4.4 foi gerado pelo *Kernel Polynomial*. A previsão feita por esse *kernel* produziu uma faixa de preços de maior amplitude e com maiores preços que a faixa prevista pelo *Kernel Mexican Hat*. No Mapa, as regiões com maiores preços também se encontram nos arredores do Centro de Belo Horizonte, com a maior preço para uma pequena região no bairro de Funcionários, próximo ao Praça Diogo de Vasconcelos. Além dos bairros da região central de Belo Horizonte, Pampulha e União também apresentaram maiores preços. Para os bairros restantes, os preços se mantiveram com valores semelhantes.

Para Belo Horizonte, o modelo mostrou que os bairros próximos à região central da cidade e os bairros Pampulha e União possuem os preços mais destoantes das outras regiões da cidade.

Hotel prices – Mexican Hat

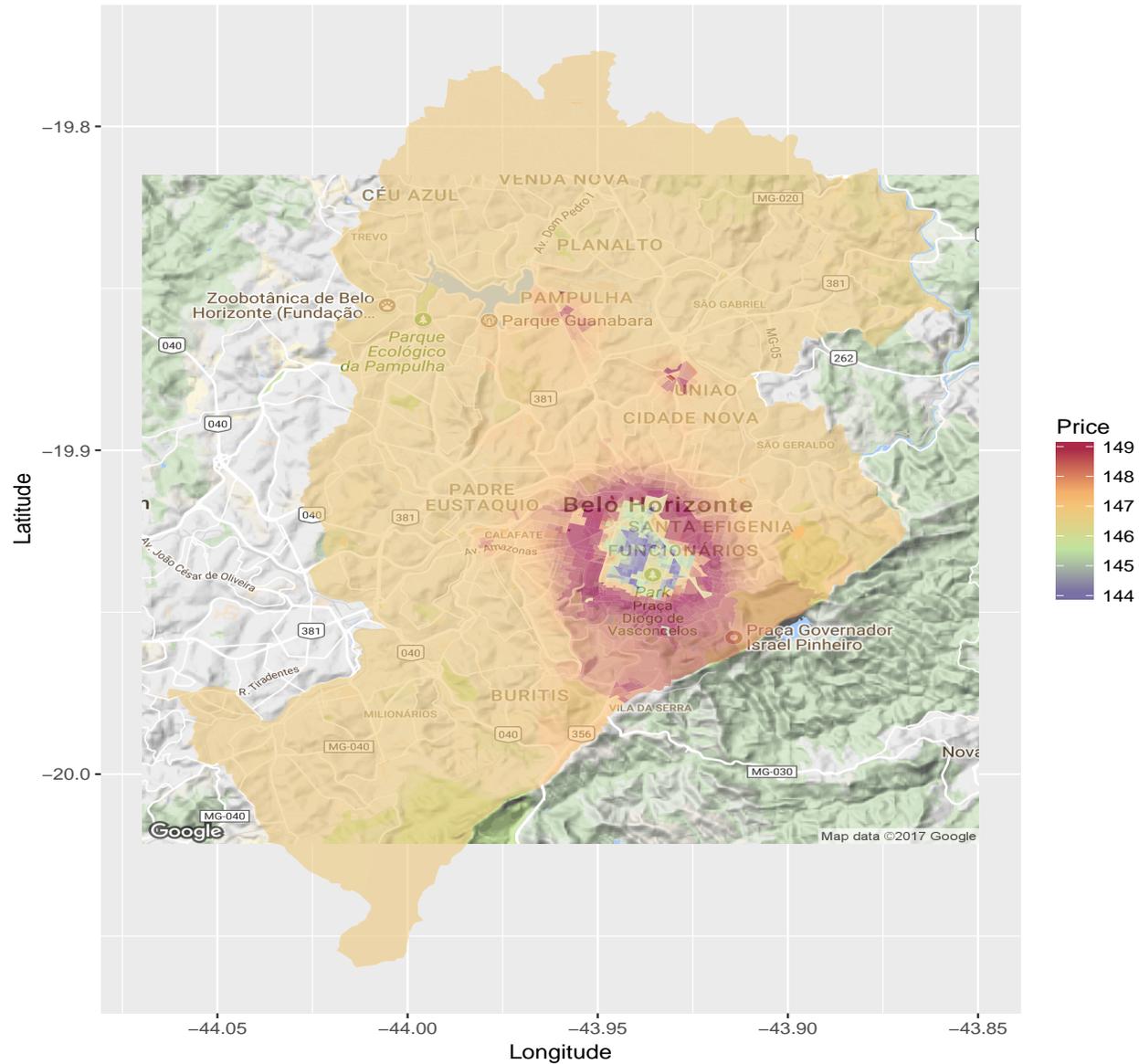


Figura 4.3 – Mapa de Preços - Belo Horizonte - Kernel Mexican Hat

Hotel prices – Polynomial

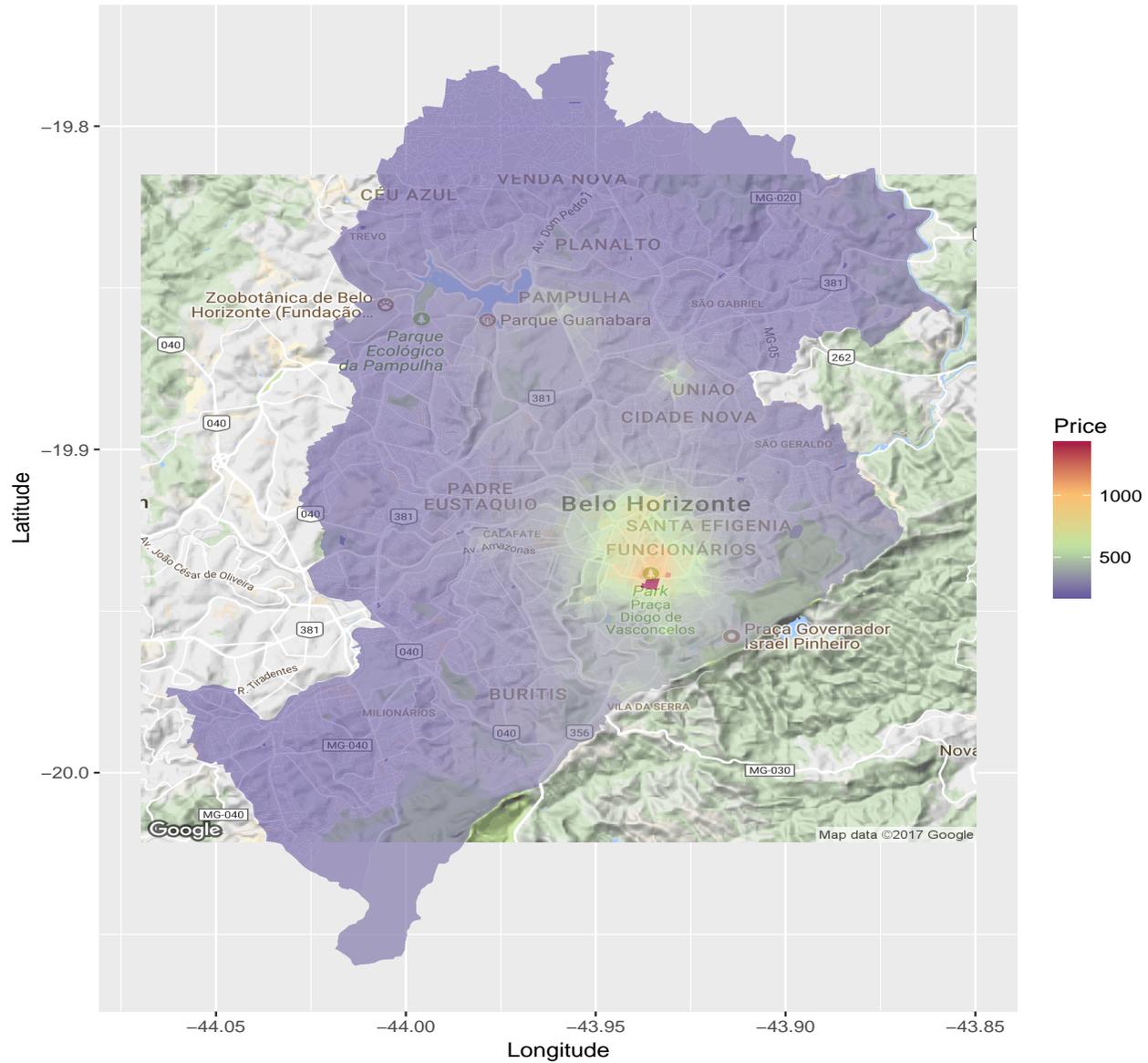


Figura 4.4 – Mapa de Preços - Belo Horizonte - Kernel Polynomial

### 4.1.3 Curitiba - PR

Para a cidade Curitiba - PR, os Mapas gerados pelo modelo podem ser vistos nas Figuras 4.5 e 4.6.

O Mapa da Figura 4.5 foi gerado pelo *Kernel Mexican Hat*. A faixa de preço prevista pelo modelo possui baixa amplitude, variando entre R\$ 180,00 E R\$ 170,00. Os bairros com maiores indício de previsão para preços altos - acima de R\$ 180,00 - estão na região Sul da cidade, evidenciado no mapa com os bairros Novo Mundo, Pinheirinho e Umbará. As regiões com menores preços, assinaladas em verde e azul, estão situadas nos bairros de próximos ao Centro de Curitiba, como Água Verde, Ahu e Mercês. A mancha azul, próxima ao bairro de Água Verde, revela a região com a menor previsão de preço da cidade.

O Mapa da Figura 4.6 foi gerado pelo *Kernel Polynomial*. Neste caso, o modelo não se ajustou aos dados da cidade e, por isso, não obteve um valor predito para análise porque os preços previstos são negativos. Existem várias razões que podem causar a não adequação do modelo, entre eles estão a falta de dados que expliquem o modelo ou a escolha do *kernel*.

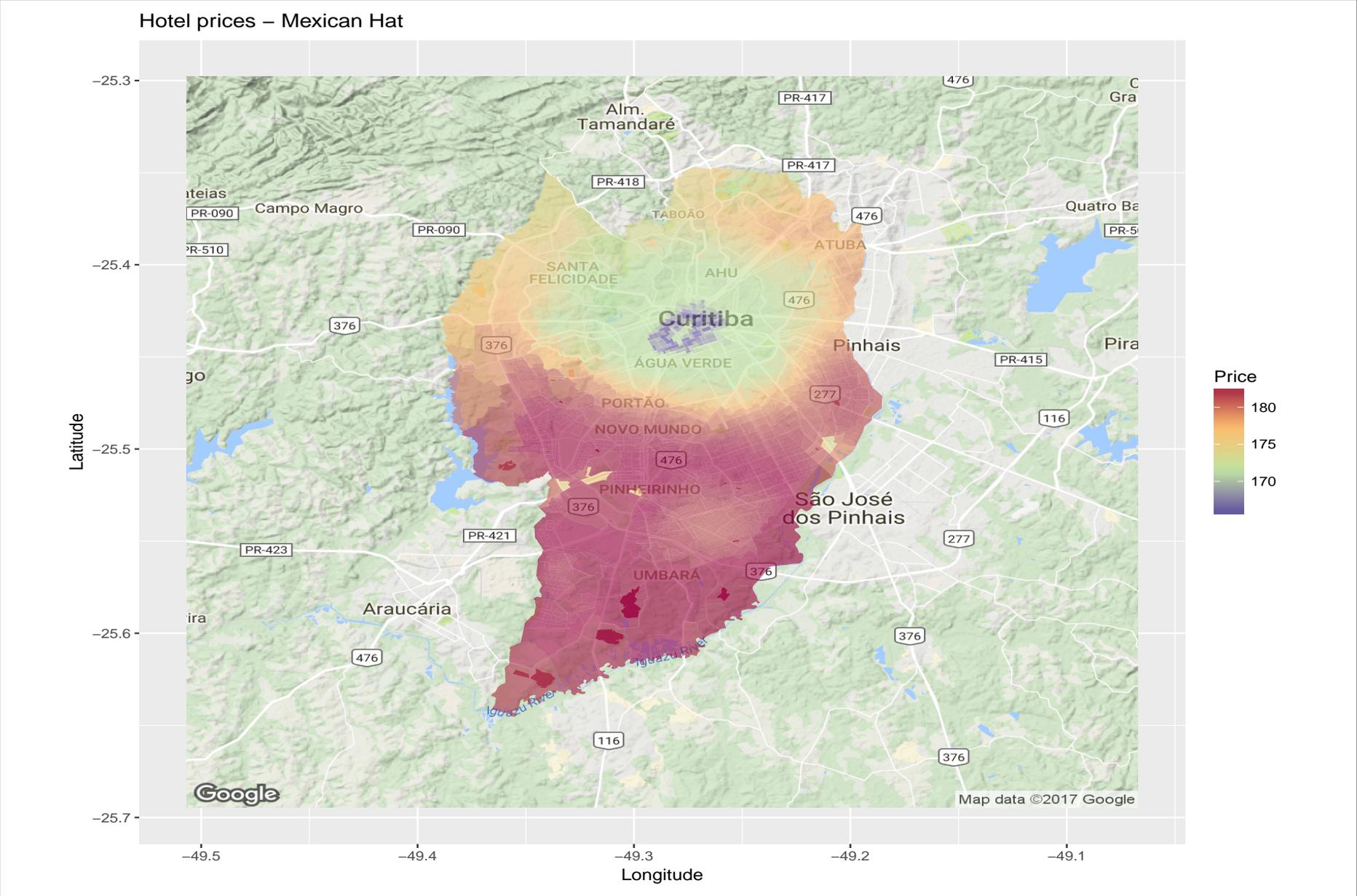


Figura 4.5 – Mapa de Preços - Curitiba - Kernel Mexican Hat

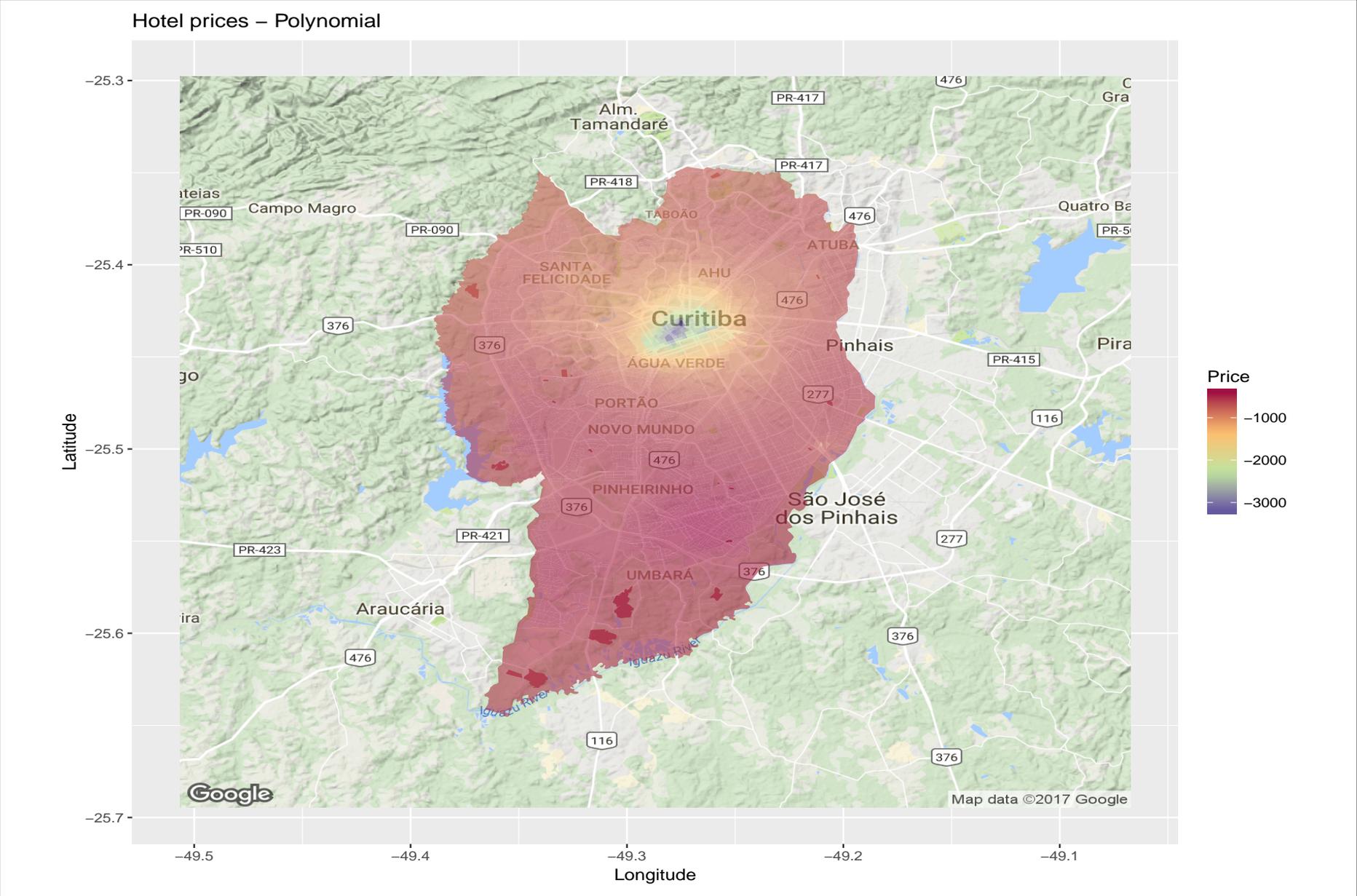


Figura 4.6 – Mapa de Preços - Curitiba - Kernel Mexican Hat

#### 4.1.4 Florianópolis - SC

Para a cidade de Florianópolis - SC, os Mapas gerados pelo modelo podem ser vistos nas Figuras 4.7 e 4.8.

O Mapa da Figura 4.7 foi gerado pelo *Kernel Mexican Hat*. É possível verificar que o modelo não se adequou para esta cidade, porque há preços negativos na faixa de preço prevista. Da mesma maneira que ocorreu com a aplicação do *Kernel Polynomial* para Curitiba - PR, o modelo não se ajusta aos dados pela falta de dados que expliquem o modelo ou pela escolha do *kernel*.

O Mapa da Figura 4.8 foi gerado pelo *Kernel Polynomial*. Para este *kernel*, o modelo gerou os maiores preços para as regiões próximas a Barra da Lagoa, Canasvieiras e Daniela, principalmente nas próximas às principais praias desses bairros. Na região azulada, assinalada no mapa, encontram-se os menores preços previstos pelo modelo. Pode-se inferir que os hotéis mais próximos ao litoral norte da cidade e próximos à Baía da Lagoa, tem tendência a ter os maiores preços que outras regiões.

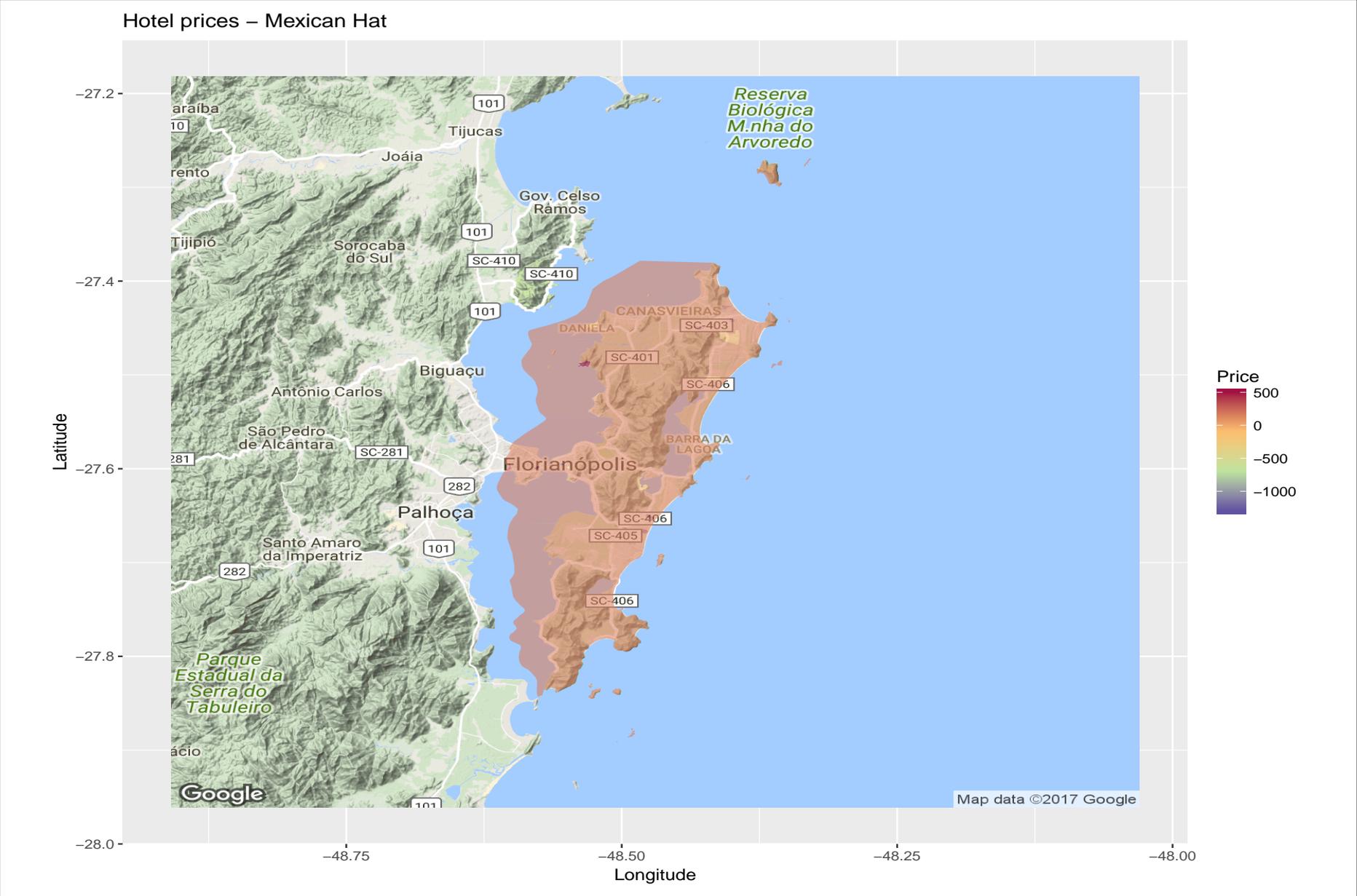
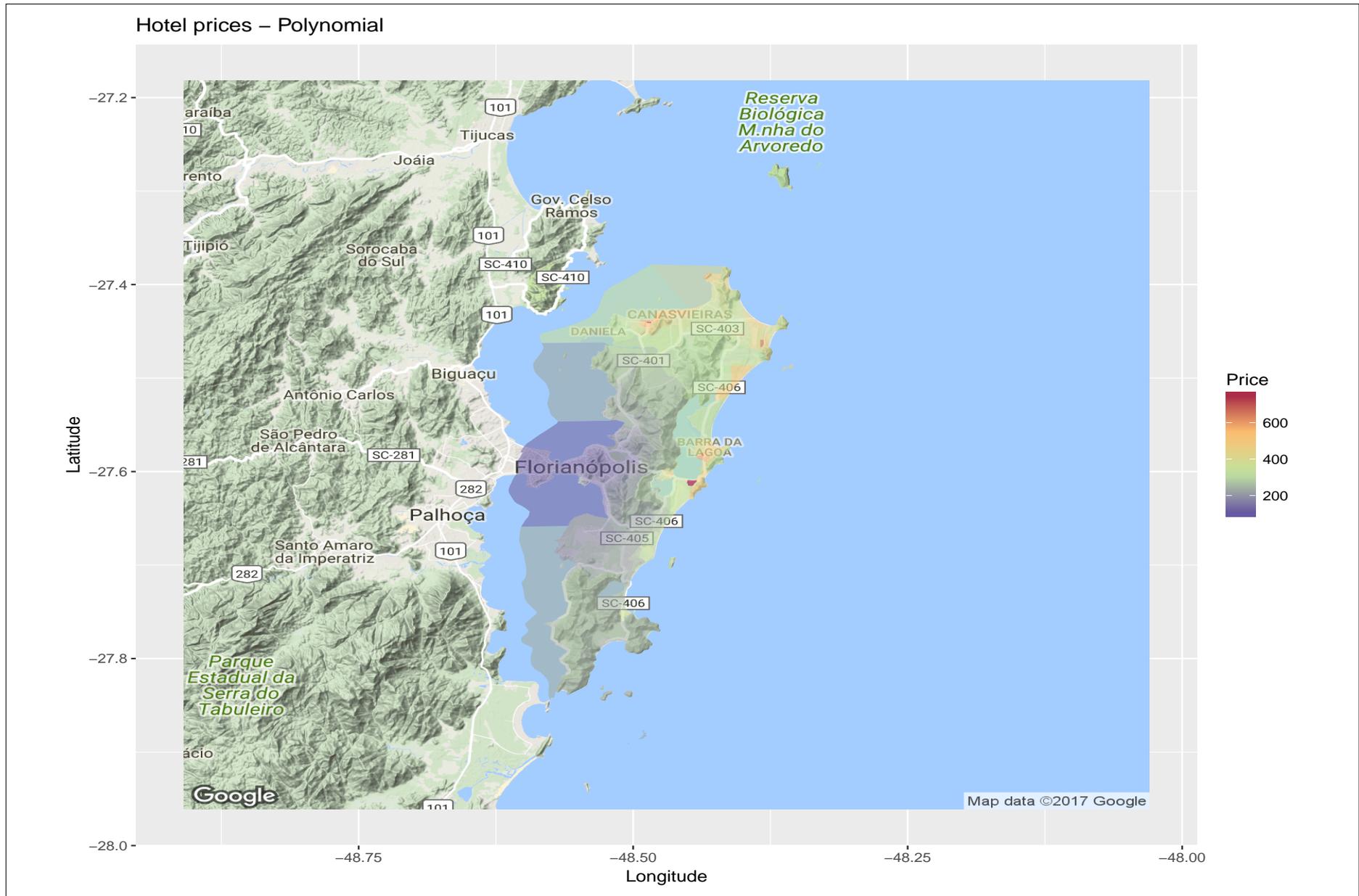


Figura 4.7 – Mapa de Preços - Florianópolis - Kernel Mexican Hat



**Figura 4.8 – Mapa de Preços - Florianópolis - Kernel Polynomial**

#### 4.1.5 Fortaleza - CE

Para da cidade de Fortaleza - CE, os mapas gerados pelo modelo podem ser vistos nas Figuras 4.9 e 4.10

O Mapa da Figura 4.9 foi gerado pelo *Kernel Mexican Hat*. Para Fortaleza, o modelo mostrou que os maiores preços, dentro do intervalo previsto, situam-se em uma pequena faixa no litoral, porém se estendem para regiões mais distantes da praia chegando ao Bairro de Passaré. Diferentemente do esperado, o modelo não previu preços mais altos para a região mais próxima às praias, o preço previsto das diárias dessa região ficou na média de R\$ 184,00. A amplitude dos preços previstos foi baixa, os menores preços previstos estão abaixo de R\$ 180,00 e estão situados na região sudoeste da cidade, mais distante dos litoral.

O Mapa da Figura 4.10 foi gerado pelo *Kernel Polynomial*. Diferentemente do resultado gerado pelo *Kernel Mexican Hat*, para o *Kernel Polynomial* os maiores preços se mantiveram nas regiões mais próximas às praias, principalmente próximo ao Bairro Dionísio Torres onde os preços variaram entre R\$ 400,00 e R\$ 800,00. Para restante dos Bairros da cidade, o modelo previu preços mais baixos, abaixo de R\$ 200,00.

Os dois *kernels* não apresentaram resultados tão semelhantes para se fazer uma análise conjunta e generalizar o comportamento do preço na cidade. Entretanto, é possível inferir há uma maior probabilidade de preços maiores para os bairros próximos ao litoral e na região próxima ao Bairro Passaré.

Hotel prices – Mexican Hat

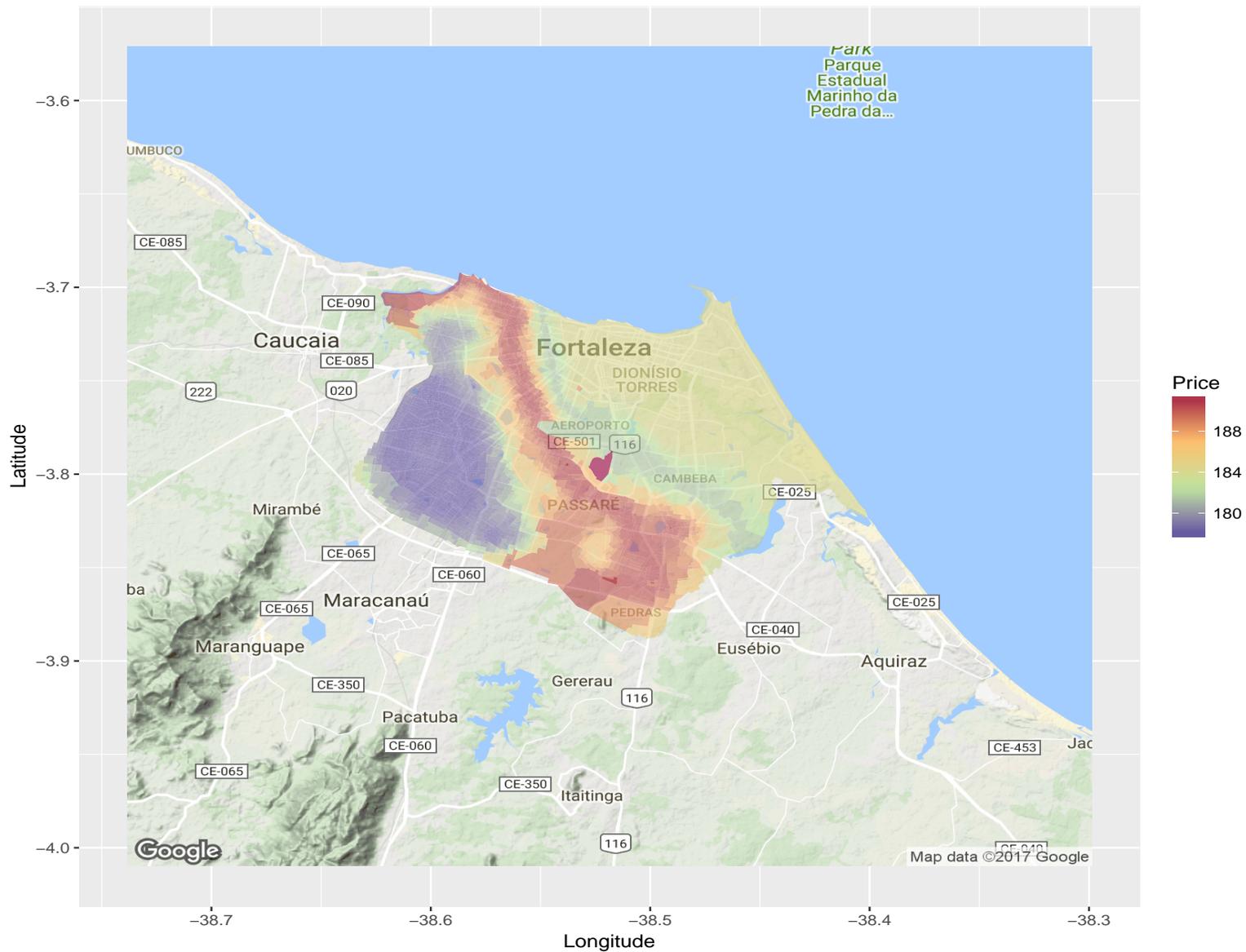


Figura 4.9 – Mapa de Preços - Fortaleza - Kernel Mexican Hat

Hotel prices – Polynomial

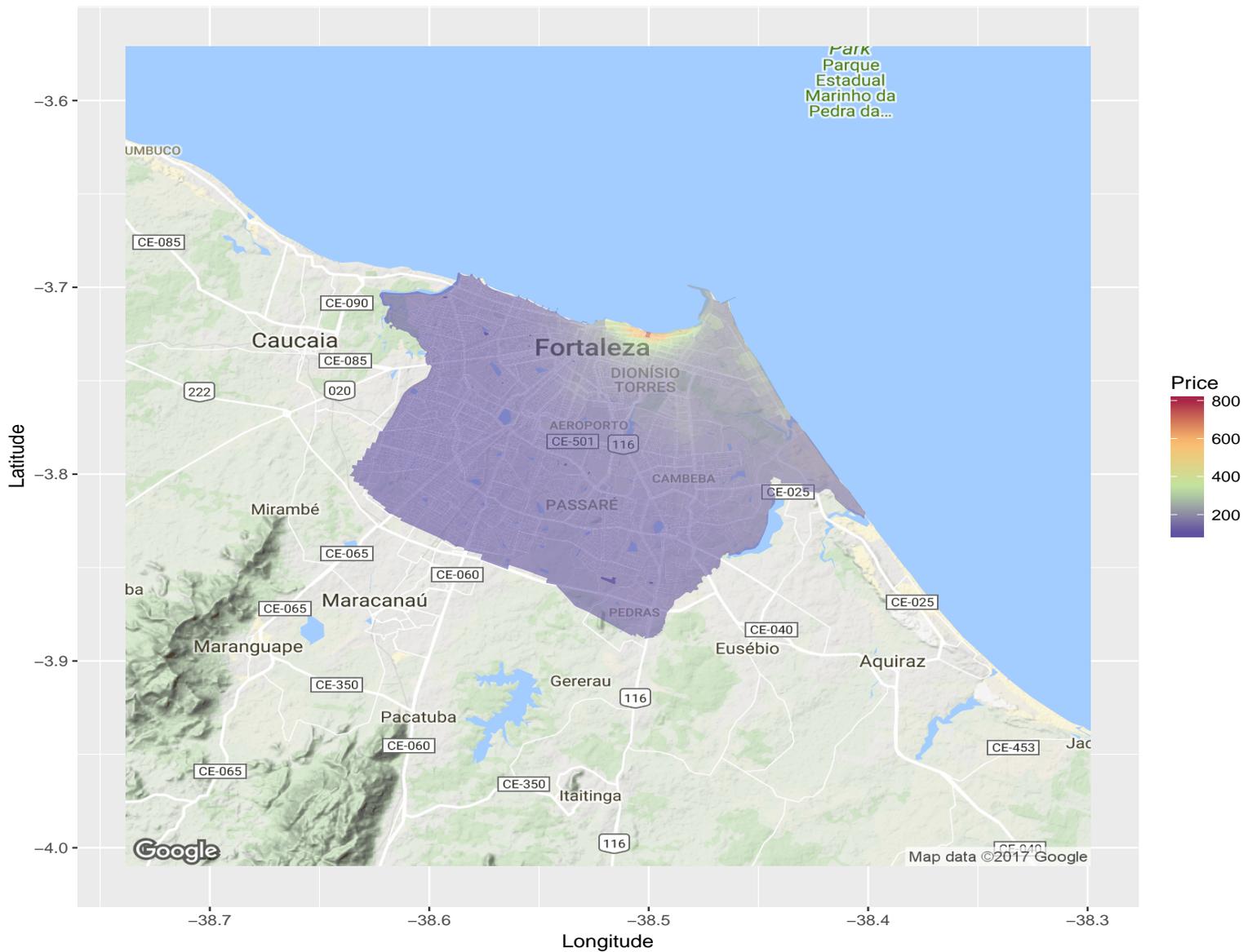


Figura 4.10 – Mapa de Preços - Fortaleza - Kernel Polynomial

#### 4.1.6 Maceió - AL

Para a cidade de Maceió - AL, os mapas gerados pelo modelo podem ser vistos nas Figuras 4.11 e 4.12.

O Mapa da Figura 4.11 foi gerado pelo *Kernel Mexican Hat*. Para este *kernel*, não houve uma grande variabilidade de preços entre os bairros da cidade, a grande maioria das regiões permaneceu com preço similar igual ou menos que R\$ 200,00. Duas pequenas regiões no litoral norte e próxima ao litoral sul apresentaram previsão de preços mais altos que outras regiões.

O Mapa da Figura 4.12 foi gerado pelo *Kernel Polynomial*. Assim como o *Kernel Mexican Hat*, também não houve grande variabilidade de preços na previsão. A grande maioria dos bairros da cidade foi prevista com a menor faixa dos preços previstos igual ou abaixo de R\$ 200,00. Apenas uma região, no litoral sul, apresentou preços mais elevados, no Bairro Ponta Verde.

Os resultados apresentados pelos *kernels* mostram de maneira similar o comportamento do preço na cidade, o que se pode inferir é que o litoral sul, próximo ao Bairro Ponta Verde, e as regiões do litoral norte tem tendência a preços mais elevados.

### Hotel prices – Mexican Hat

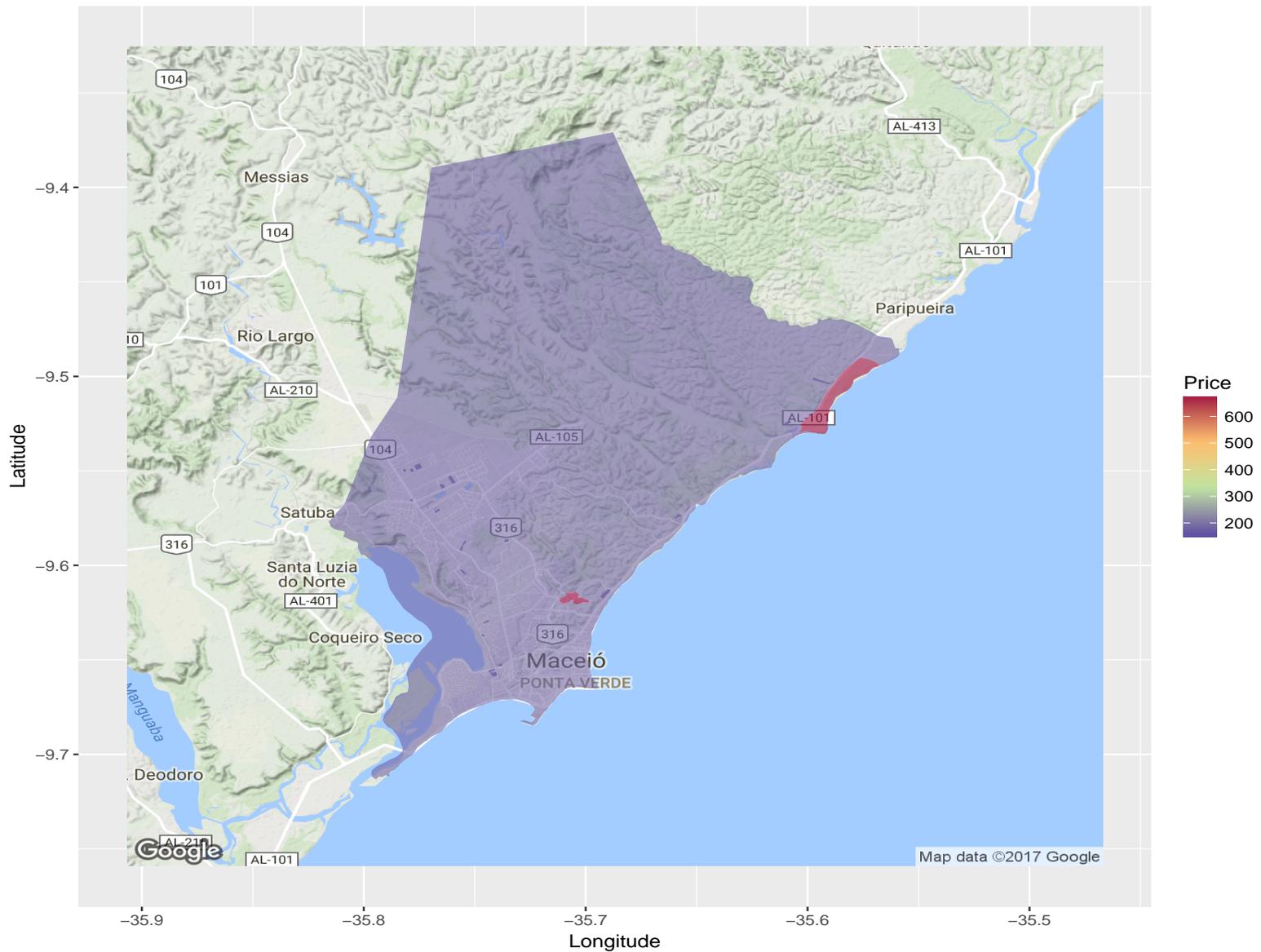


Figura 4.11 – Mapa de Preços - Maceió - Kernel Mexican Hat

Hotel prices – Polynomial

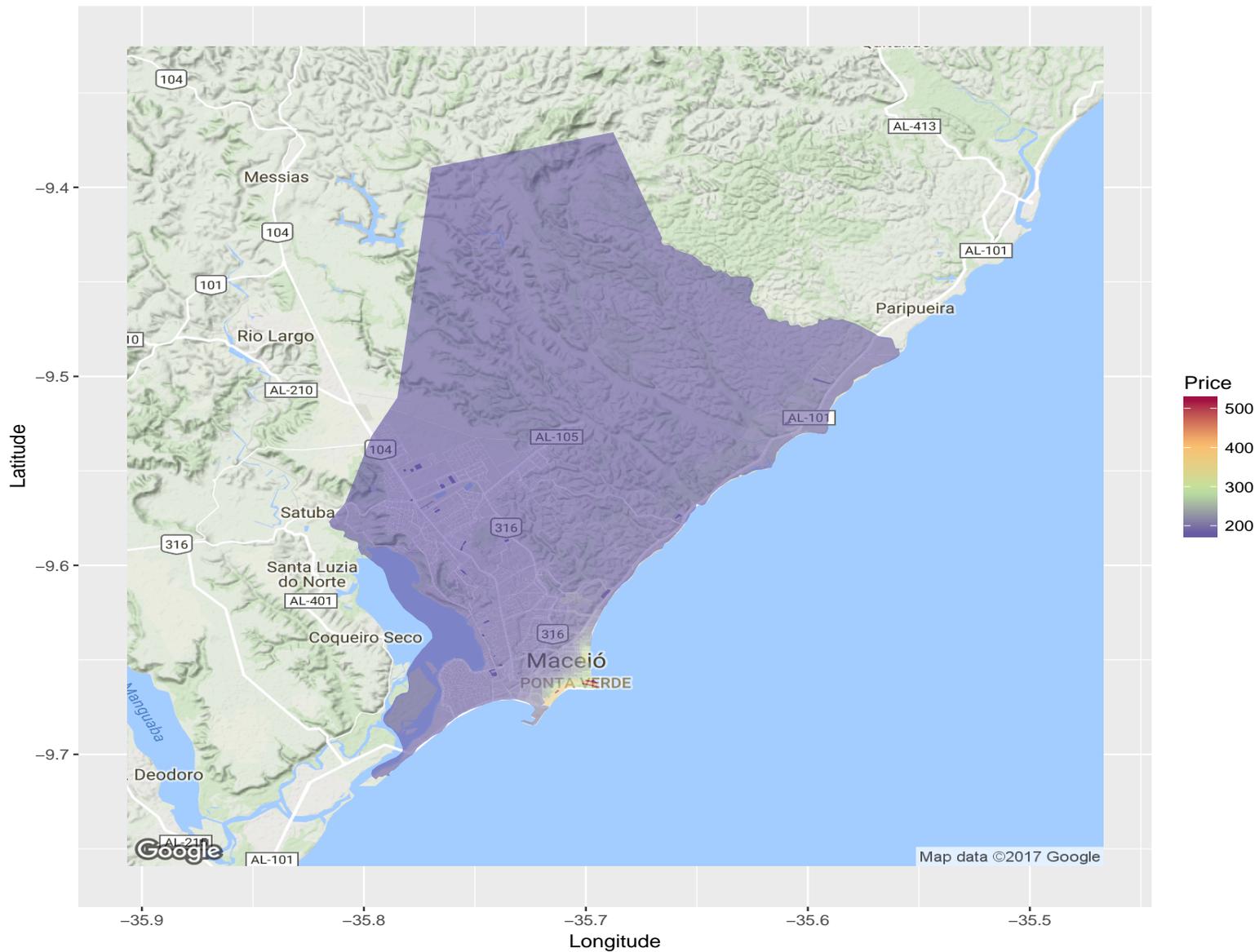


Figura 4.12 – Mapa de Preços - Maceió - Kernel Polynomial

#### 4.1.7 Manaus - AM

Para a cidade de Manaus - AM, os mapas gerados pelo modelo podem ser vistos a seguir nas Figuras 4.13 e 4.14.

O Mapa da Figura 4.13 foi gerado pelo *Kernel Mexican Hat*. Assim como o município de Maceió, para Manaus, a grande maioria dos bairros não possuiu grande variabilidade nos preços previstos. A região gerada no Mapa consiste em toda extensão do município, entretanto a parte mais populosa da cidade se encontra no sul do município onde é possível verificar uma maior variabilidade dos preços previstos. Observando apenas esta região, verifica-se que os maiores preços situam-se nos bairros mais próximos do centro da cidade e dos pontos turísticos. A previsão de preços para os outros bairros permaneceu similar entre R\$ 135,00 e R\$ 132,50.

O Mapa da Figura 4.14 foi gerado pelo *Kernel Polynomial*. O mapa gerado por este *kernel* é bastante semelhante ao mapa do *kernel* anterior. Para este *kernel*, somente uma pequena parte das regiões centrais da cidade apresentaram os maiores preços da previsão. O restante do município apresenta o mesmo preço de previsão. Generalizando os resultados do modelo, é possível inferir que apenas a região central de Manaus apresentou preços maiores.

### Hotel prices – Mexican Hat

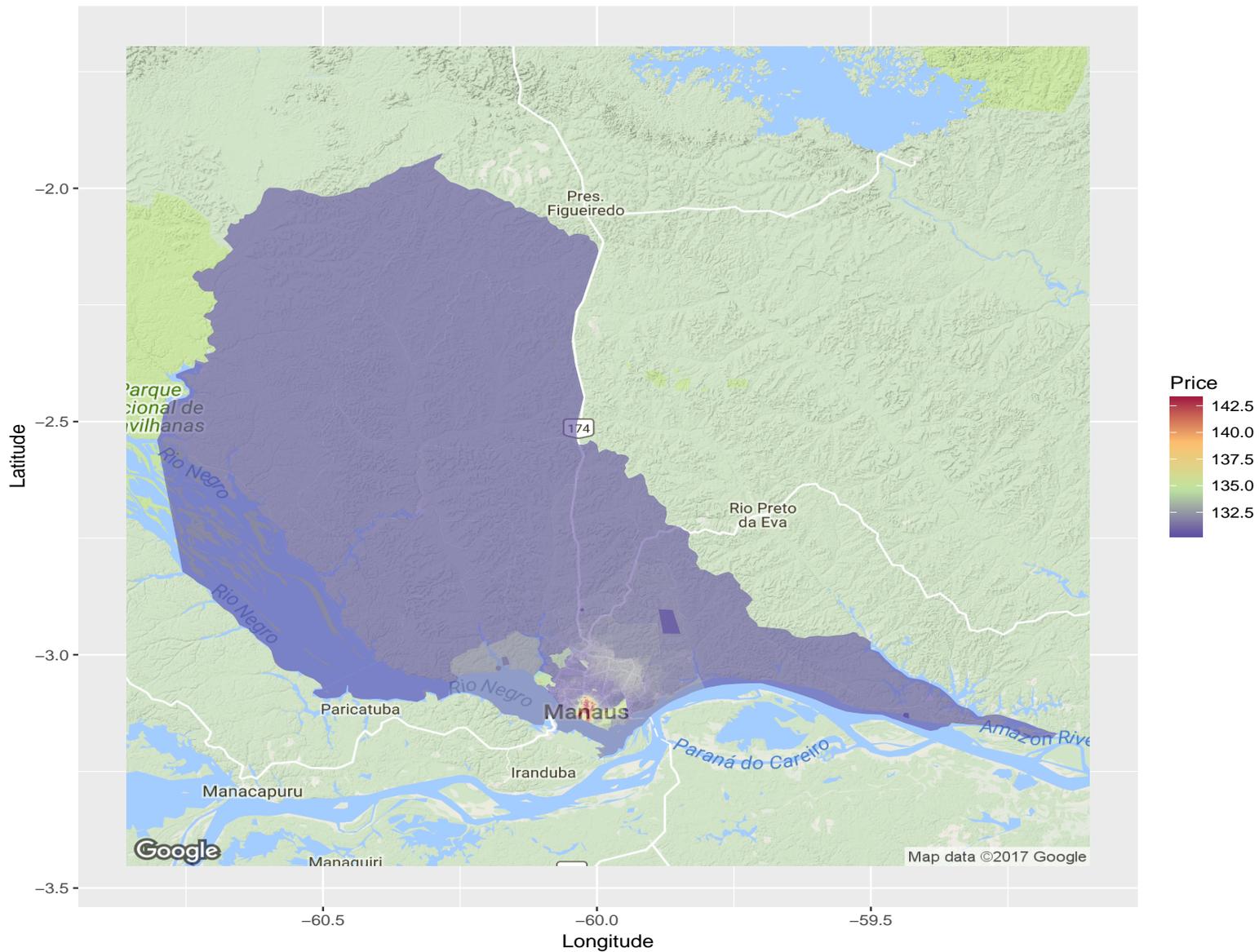


Figura 4.13 – Mapa de Preços - Manaus - *Kernel Mexican Hat*

### Hotel prices – Polynomial

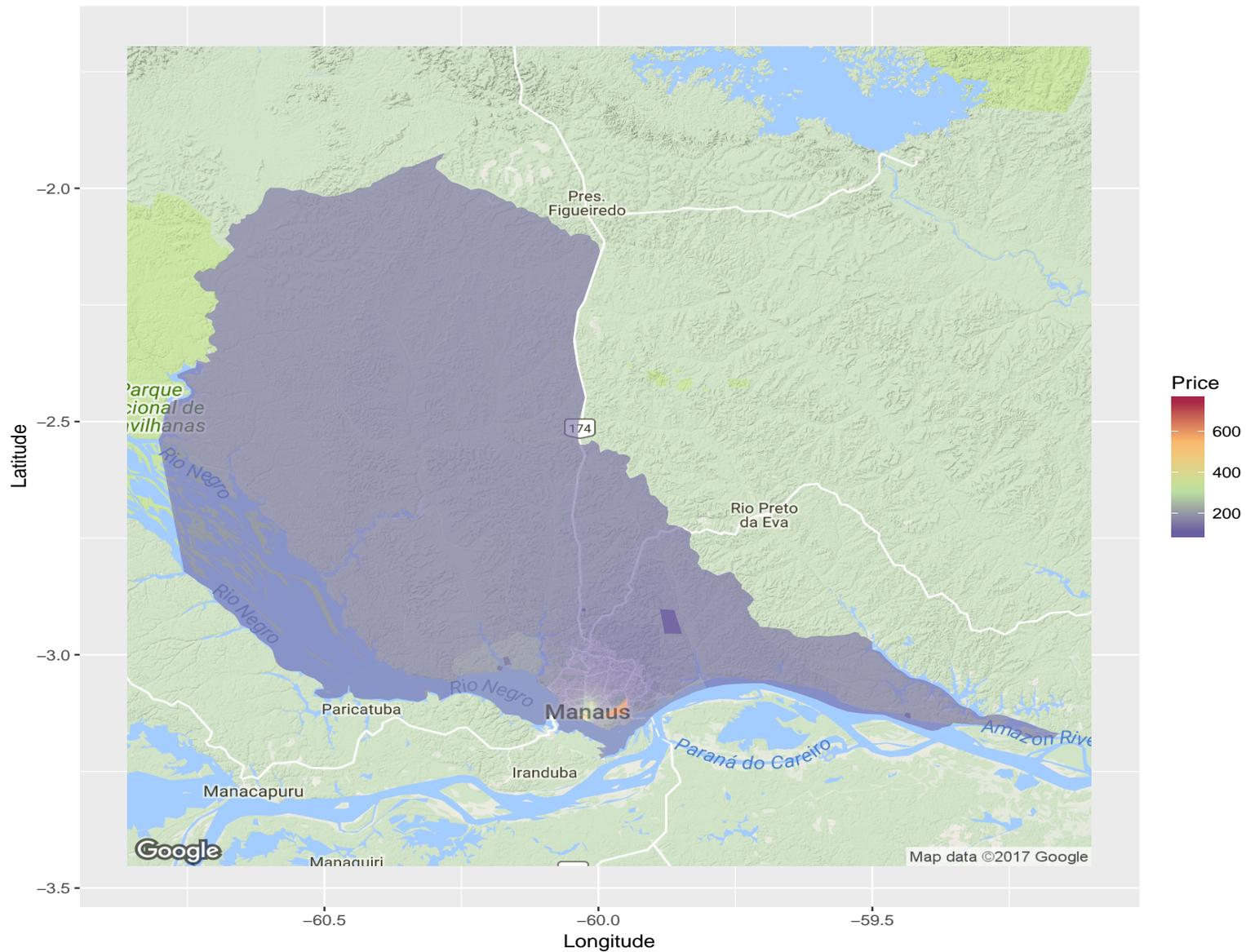


Figura 4.14 – Mapa de Preos - Manaus - Kernel Polynomial

#### 4.1.8 Natal - RN

Para a cidade de Natal - RN, os mapas gerados pelo modelo podem ser vistos nas Figuras 4.15 e 4.16.

O Mapa da Figura 4.15 foi gerado pelo *Kernel Mexican Hat*. Para este *kernel*, a previsão apresentou uma grande variabilidade de preços entre os bairros. A região com preços altos - assinalada em vermelho - situa-se mais ao norte da cidade, nos bairros de Potengi, Pajuçara, Iguapó e as suas proximidades. Os bairros da região próxima ao litoral sul também apresentou previsão de preço elevado, aproximadamente R\$ 178,00 a diária. Nesta região se encontra algumas das praias famosas da cidade, a Ponta Negra, e o Parque da Dunas, ponto turístico da cidade. A região que apresentou os preços mais baixos encontra-se nos bairros de Alecrim e Lagoa Nova e se estende até os bairros da parte oeste de Natal.

O Mapa da Figura 4.16 foi gerado pelo *Kernel Polynomial*. Diferentemente do *Kernel Mexican Hat*, o resultado gerado pelo *Kernel Polynomial* apresentou uma grande amplitude no intervalo de preços. A região que mostrou previsão de preço mais elevada se encontra no litoral no Bairro Ponta Negra e Neópolis, com preços variando de R\$ 500,00 a R\$ 1500,00. O restante dos bairros permaneceram com os mesmos preços de previsão, abaixo de R\$ 500,00.

Para Natal - RN, o modelo propôs resultados relativamente semelhantes. Pode-se inferir que o litoral sul tem tendência a preços mais elevados que as outras regiões, devido à proximidade com as praias mais frequentadas e os principais pontos turísticos.

### Hotel prices – Mexican Hat

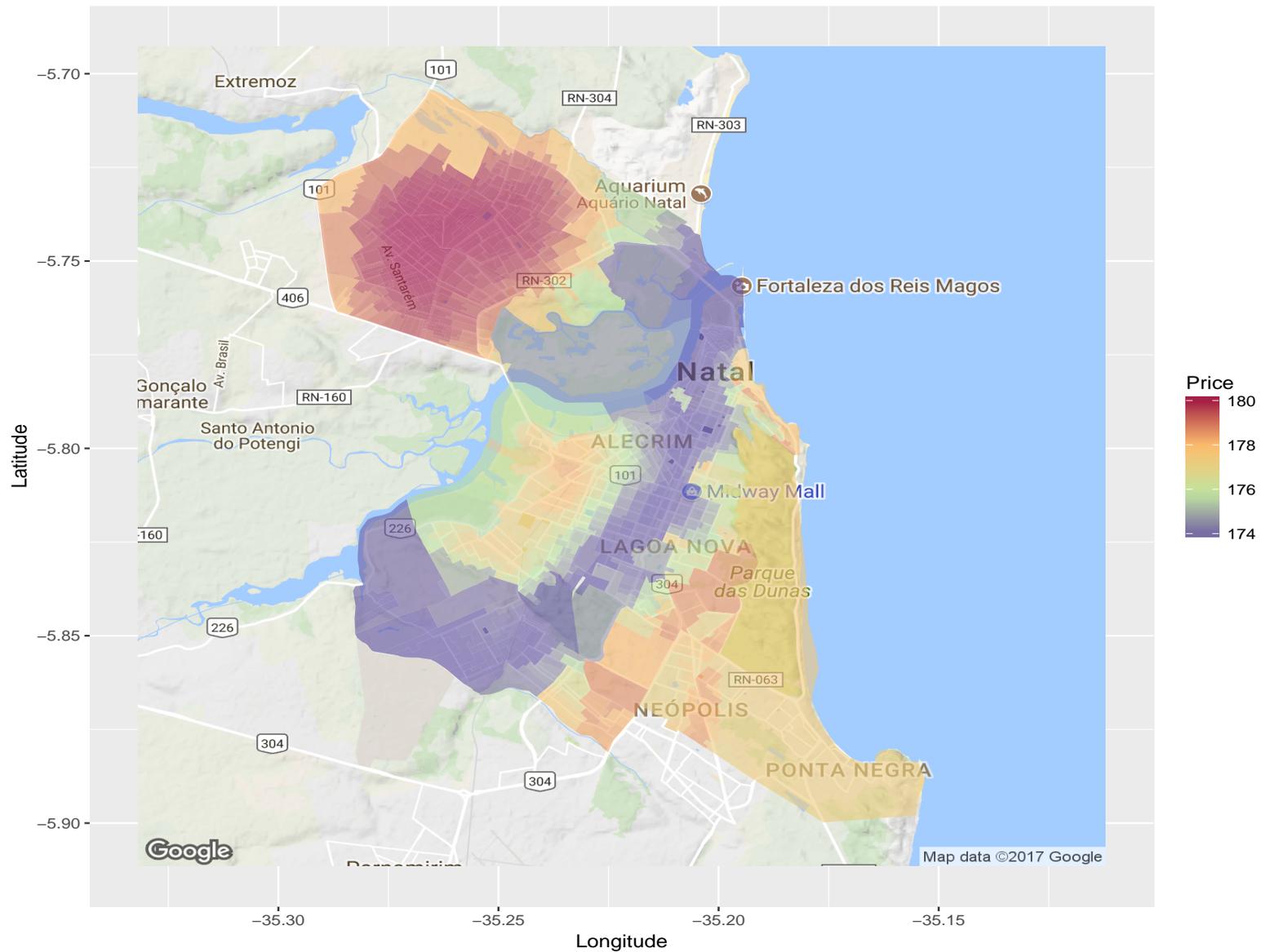


Figura 4.15 – Mapa de Preços - Natal - Kernel Mexican Hat

Hotel prices – Polynomial

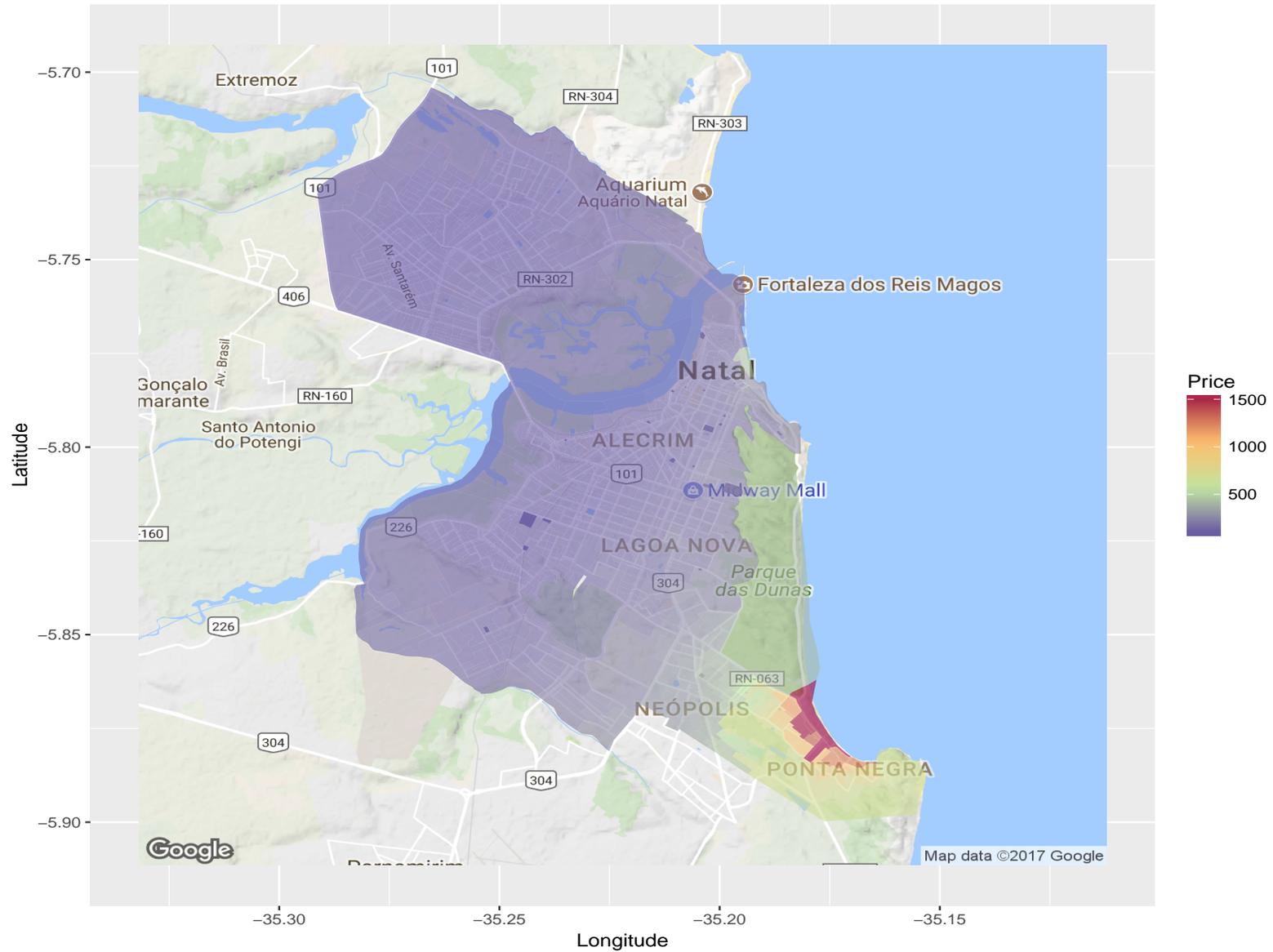


Figura 4.16 – Mapa de Preços - Natal - Kernel Polynomial

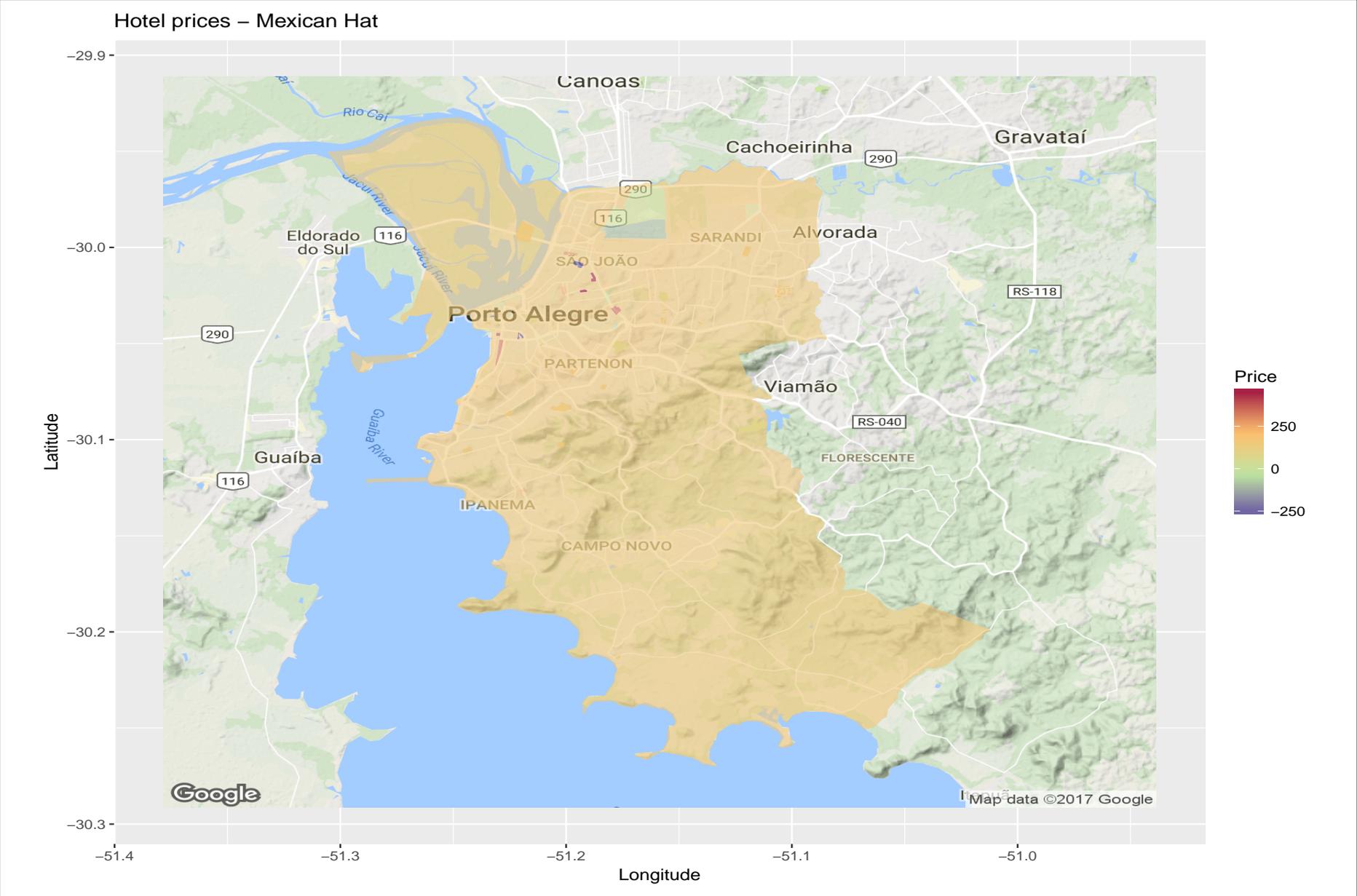
#### 4.1.9 Porto Alegre - RS

Para a cidade de Porto Alegre - RS, os mapas gerados pelo modelo podem ser vistos nas Figuras 4.17 e 4.18.

O Mapa da Figura 4.17 foi gerado pelo *Kernel Mexican Hat*. O resultado apresentado pelo *kernel* não indicou grande variabilidade na previsão dos preços. A quase totalidade dos bairros do município permaneceram com o mesmo preços previsto, com média de R\$ 250,00. Com exceção apenas para algumas pequenas regiões no norte do município, próxima ao Bairro São João que apresentaram preços diferentes do restante das regiões do município.

O Mapa da Figura 4.18 foi gerado pelo *Kernel Polynomial*. Para este *kernel*, a previsão apresentou preços mais elevados nos bairros da região norte do município, com foco aos Bairros Cidade Baixa e Bom Fim com os preços mais altos. O restante dos bairros de Porto Alegre permaneceu com o mesmo preço de previsão, abaixo de R\$ 200,00.

É possível inferir por meio dos mapas, que a região próximas aos Bairros São João, Cidade Baixa e Bom Fim possuem tendência de preços mais elevados para as diárias, diferenciando dos outros bairros da cidade.



**Figura 4.17 – Mapa de Preços - Porto Alegre - Kernel Mexican Hat**

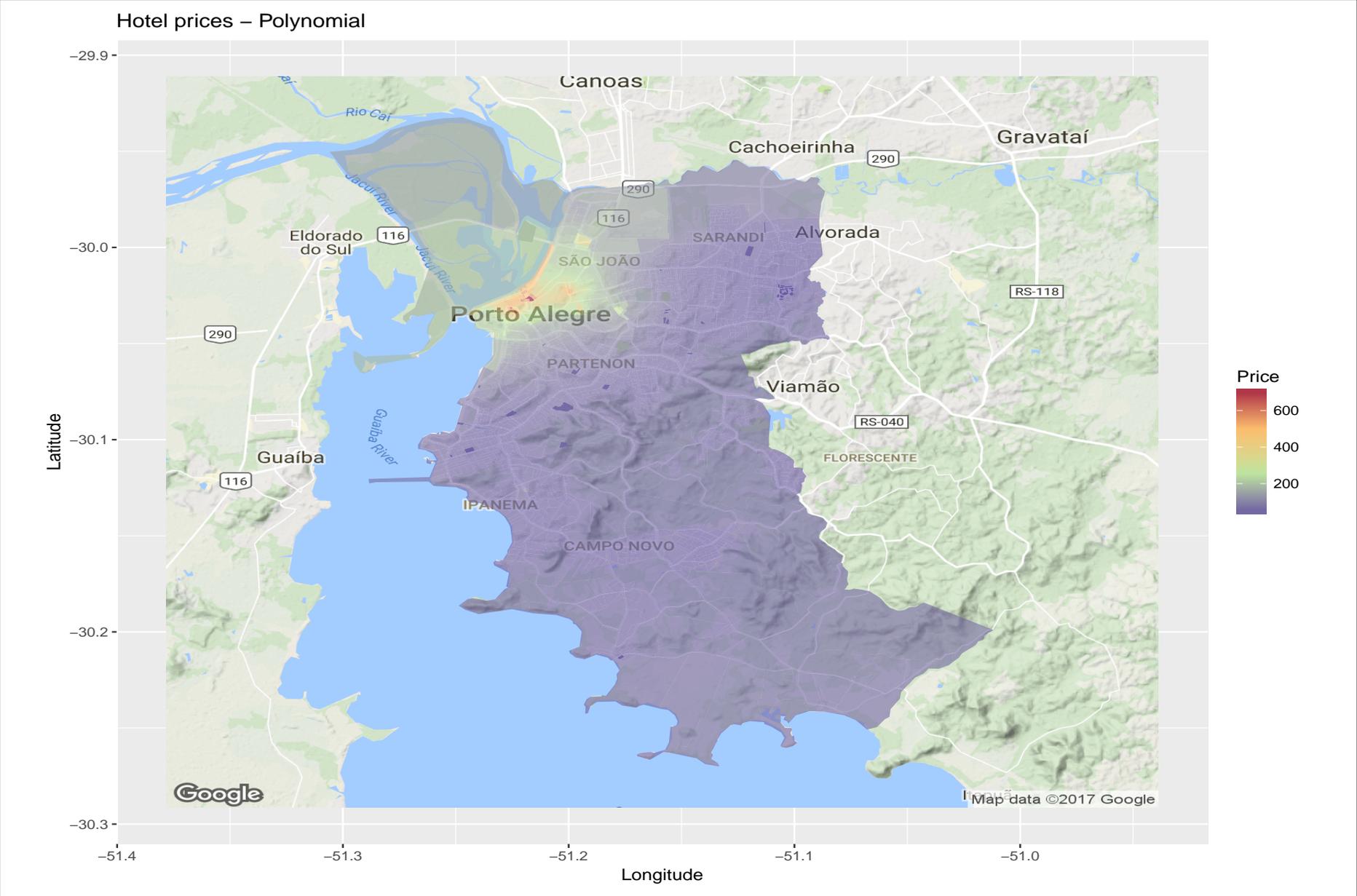


Figura 4.18 – Mapa de Preços - Porto Alegre - Kernel Polynomial

#### 4.1.10 Recife - PE

Para a cidade de Recife - PE, os mapas gerados pelo modelo podem ser vistos nas Figuras 4.19 e 4.20.

o Mapa da Figura 4.19 foi gerado pelo *Kernel Mexican Hat*. Diferentemente do esperado, as maiores previsões de preço, para este *kernel*, não foram para os bairros próximos ao litoral. Os maiores preços da previsão situam-se nos bairros Casa Amarela, Encruzilhada e os arredores, regiões relativamente distantes do litoral. Os Bairros que apresentaram os menores preços da previsão são Imbiribeira, Pina e Boa Viagem, no litoral sul de Recife.

O Mapa da Figura 4.20 foi gerado pelo *Kernel Polynomial*. O resultado apresentado por este *kernel* mostrou que as maiores previsões de preço situam-se nos Bairros do litoral sul, principalmente Pina e Boa Viagem. O restante dos bairros permaneceu com a mesma média de preço prevista pelo modelo, entre R\$ 240,00 e R\$ 220,00.

A diferença nos resultados dos dois *kernels* é explicada pela escolha dos parâmetros de cada *kernel*, que podem adaptar melhor o modelo aos dados estudados. No caso, o *Kernel Polynomial* se adaptou melhor aos dados e conseguiu gerar resultados mais plausíveis para a realidade da cidade litorânea.

### Hotel prices – Mexican Hat

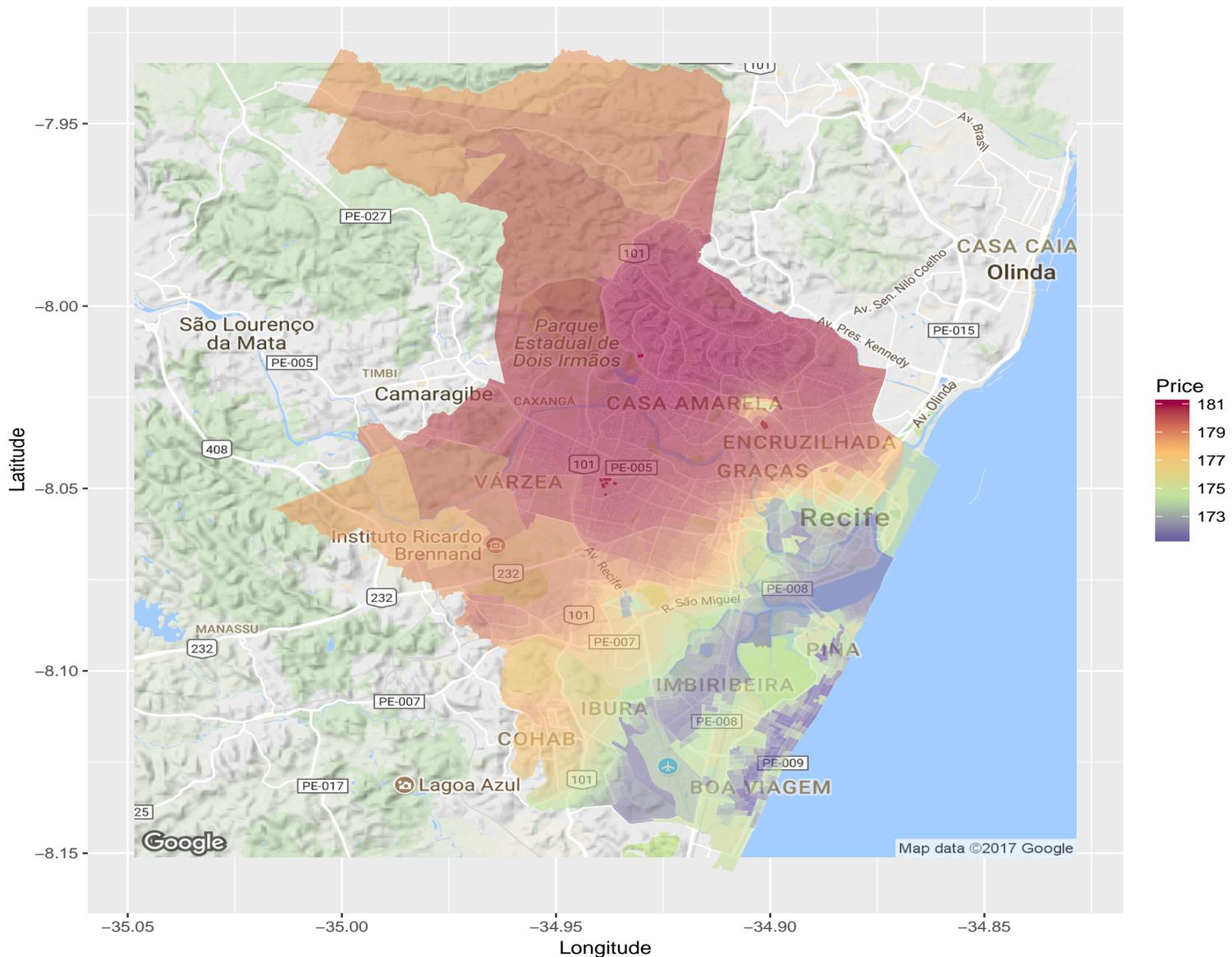


Figura 4.19 – Mapa de Preços - Recife - Kernel Mexican Hat

Hotel prices – Polynomial

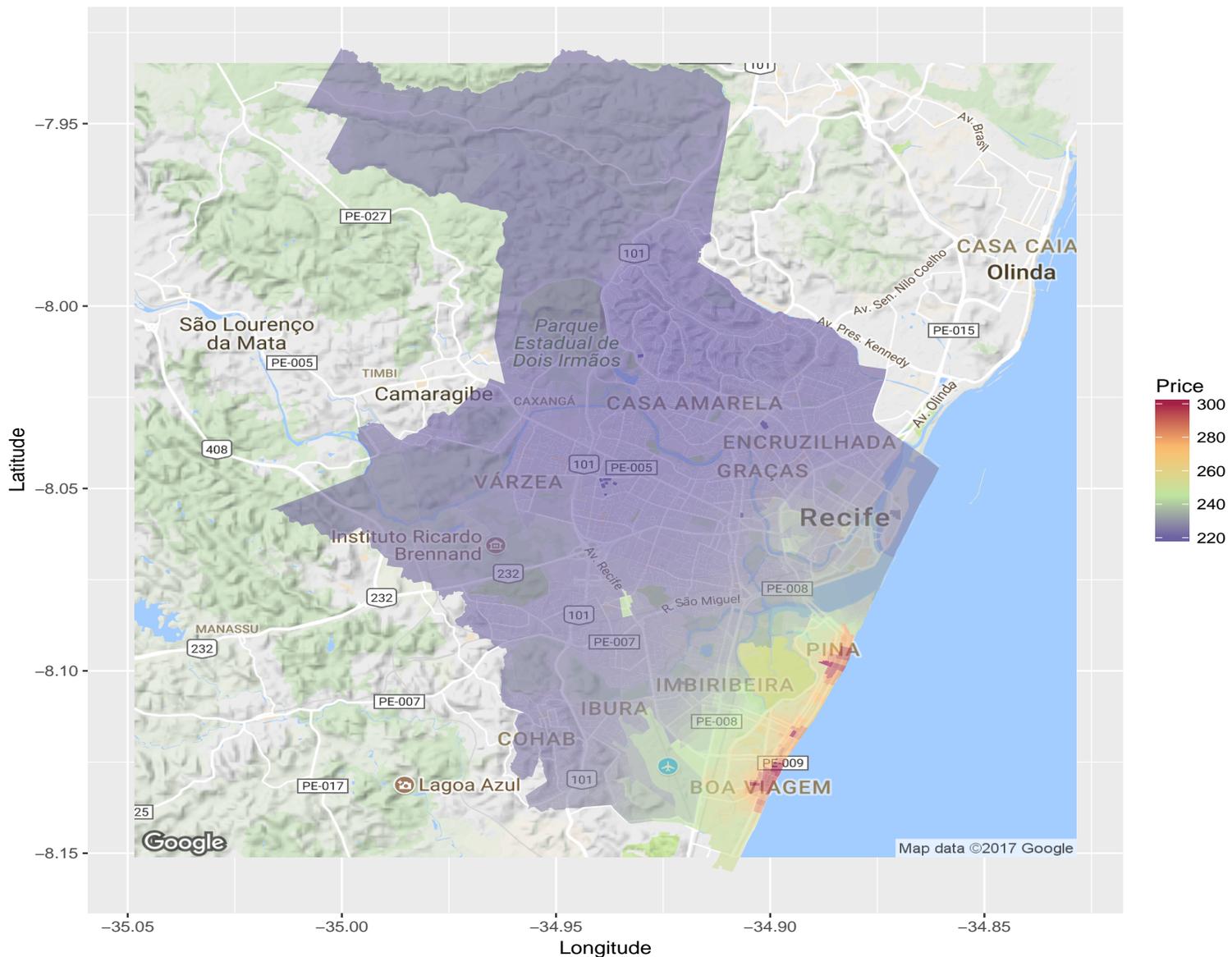


Figura 4.20 – Mapa de Preços - Recife - Kernel Polynomial

#### 4.1.11 Salvador - BA

Para a cidade de Salvador - BA, os mapas gerados pelo modelo podem ser vistos nas Figuras 4.21 e 4.22.

O Mapa da Figura 4.21 foi gerado pelo *Kernel Mexican Hat*. O resultado deste *kernel* previu uma grande variabilidade de preço entre os bairros de Salvador. As regiões com maior previsão de preço situam-se nos Bairros de Paripe, Plataforma, São Francisco e proximidades ao norte do município, Liberdade, Imbuí, região próxima ao litoral sul nos Bairros de Brotas e Rio Vermelho e algumas pequenas regiões no litoral norte. Grande parte do município permaneceu com a mesma previsão de preço na faixa de de R\$ 155,00. As regiões com os menores preços de previsão situam-se no litoral norte do município, nas proximidades dos Bairros São Cristóvão e Piata e, além disso, em uma pequena região no litoral sul próxima aos bairros com previsão de preços mais altos.

O Mapa da Figura 4.22 foi gerado pelo *Kernel Polynomial*. A faixa de preço prevista pelo *kernel* apresentou preços altos e, por isso, grande parte dos bairros do município permaneceu com na mesma faixa de previsão, abaixo de R\$ 1000,00. Apenas duas regiões apresentaram preços maiores em relação aos outros Bairros, situadas no litoral sul, próxima ao Bairro Brotas, e em uma pequena região no litoral norte.

A partir dos mapas, pode-se inferir que as regiões próximas aos Bairros Brotas e Rio Vermelho, no litoral sul, e aos Bairros São Cristóvão e Piata, o litoral norte, são mais propícias para hotéis mais caros, principalmente pela proximidade das praias mais famosas do município.

Hotel prices – Mexican Hat

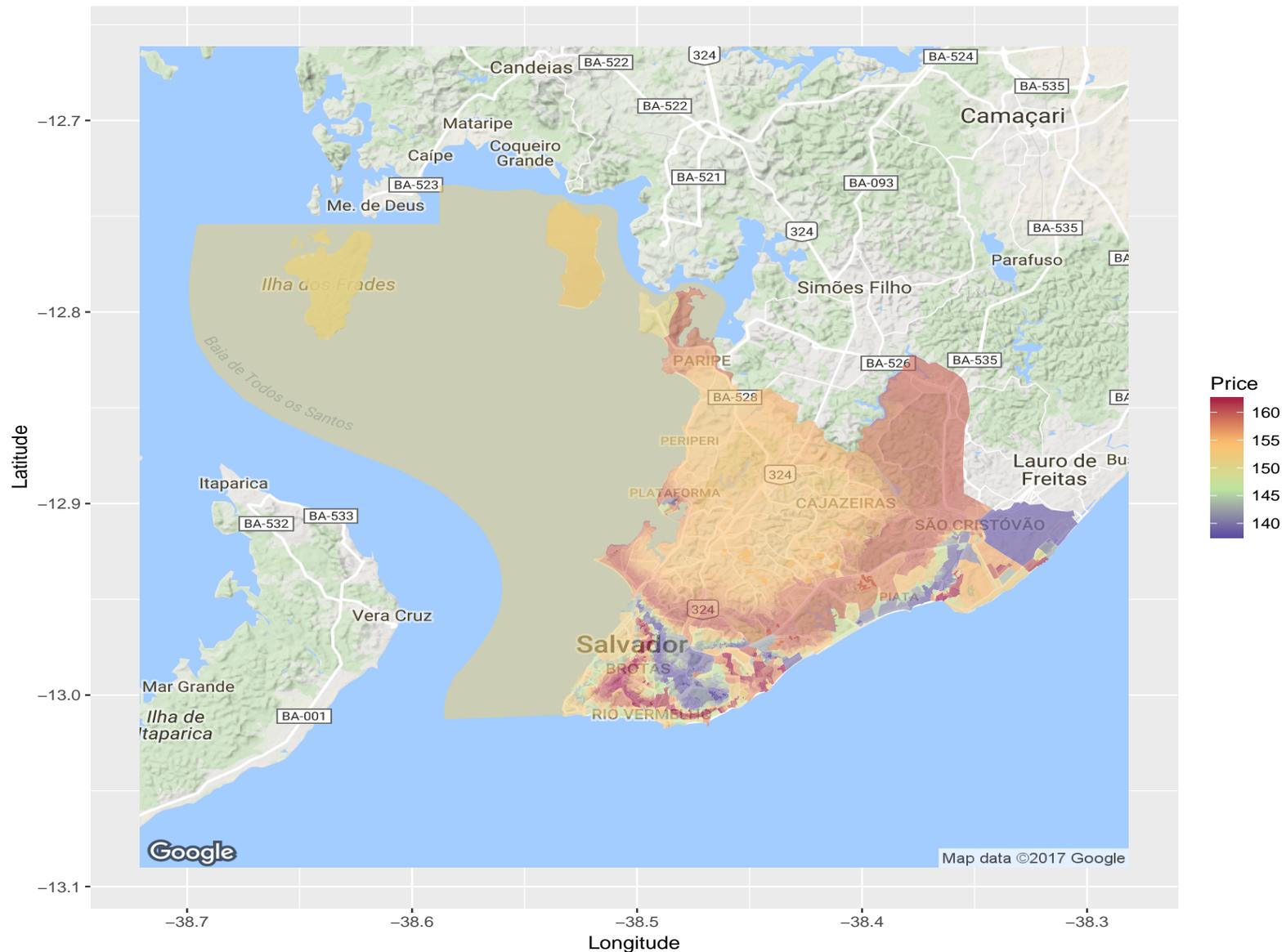


Figura 4.21 – Mapa de Preços - Salvador - Kernel Mexican Hat

Hotel prices – Polynomial

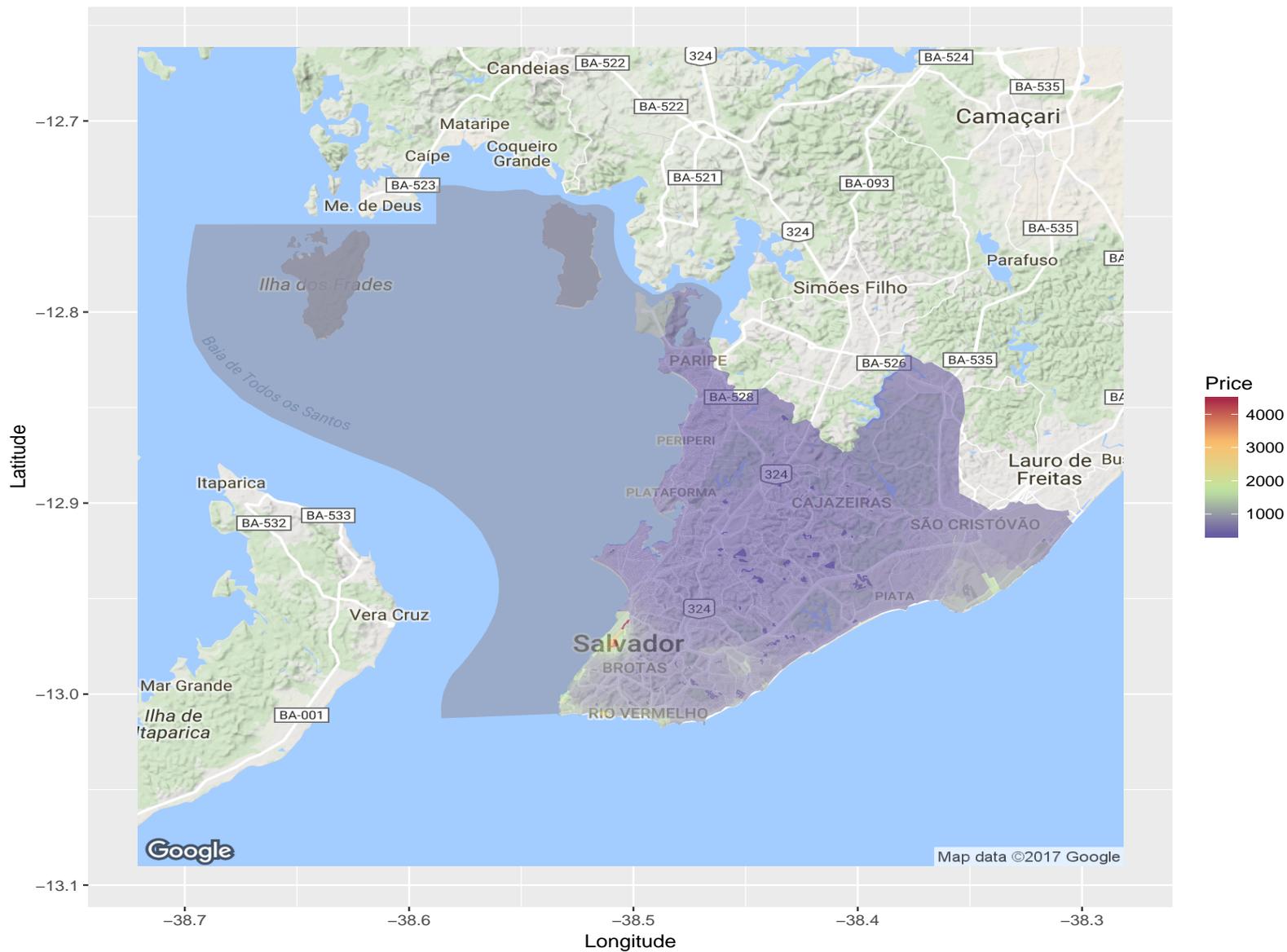


Figura 4.22 – Mapa de Preços - Salvador - Kernel Polynomial

## 5 CONCLUSÃO

Com o passar dos anos e o avanço da tecnologia, o sucesso empresarial se torna cada vez mais dependente do uso de novos meios tecnológicos para a tomada de decisão. Focando nisso, este estudo teve como objetivo criar um método quantitativo utilizando Aprendizado de Máquina para auxílio na tomada de decisão gerencial de alocação de um hotel. Foi proposto o método quantitativo com auxílio das teorias de Aprendizado de Máquina e Geomarketing e com foco para aplicação de um modelo de *Support Vector Regression (SVR)* para previsão dos preços de acordo com a localização e os serviços oferecidos pelo hotel. Dessa forma, o modelo gerou mapas de cada uma das cidades analisadas caracterizando os bairros por preço para apoiar na indicação de seria a localização mais propícia de um hotel de acordo com os serviços prestados e das variáveis . Aplicando o Sistema de Informação Geográfica, foram criadas bases geocodificadas - com latitude e longitude - para relacionar com os dados do preços e das amenidades dos hotéis e, dessa forma, aplicar as teorias de Geomarketing ao mercado hoteleiro.

Com os dados geocodificados, os dados de preço das diárias e os dados das amenidades dos hotéis, foi aplicado o *Support Vector Regression* e o Modelo Conceitual para cada cidade fixando os serviços desejados por um perfil hipotético de turista, gerando um mapa de previsão de preços da cidade estudada. Para o *Support Vector Regression*, foram utilizados dois *kernels*: *Mexican Hat* e *Polynomial*. A partir dos resultados, foi possível perceber que a escolha dos *kernels* é de suma importância para o desempenho do modelo, porque afeta diretamente o *output*.

A partir dos mapas gerados pelos *kernels*, os gestores podem verificar com clareza o comportamento dos preços dentro de cada bairro da cidade e, com isso, decidir qual é o melhor local para alocar um hotel de acordo com as características da região e os serviços oferecidos pelo estabelecimento. Neste estudo, foram analisadas 11 capitais turísticas do Brasil. Para cada cidade, foram gerados dois mapas, um de cada *kernel*, e foi feita a análise descritiva do comportamento dos preços na cidade para cada mapa apresentado. Para cada município, o modelo buscou se adequar aos dados e generalizar a estrutura, entretanto em alguns cidades

isto não foi possível, como foi o caso de Curitiba, Florianópolis e Porto Alegre para o *Kernel Mexican Hat*.

Observando os mapas das capitais distantes do litoral, isto é, Brasília, Belo Horizonte, Curitiba, Porto Alegre e Manaus, a grande maioria dos resultados mostrou que os maiores preços dos hotéis situam-se principalmente, nos bairros próximos às regiões centrais dos municípios ou próximos às regiões mais nobres. Isso mostra que modelo consegue distinguir a realidade socioeconômica de cada cidade, para aplicar ao comportamento dos preços dos hotéis. Apenas uma cidade não seguiu o padrão citado acima: Curitiba. Em Curitiba, diferentemente do esperado, as regiões centrais apresentaram os menores preços da previsão.

Para os mapas das cidades litorâneas, ou seja, Florianópolis, Fortaleza, Maceió, Recife e Salvador, esperava-se que as regiões com previsão de altos preços se encontrassem próximos ao litoral. Em algumas cidades, os *kernels* apresentaram resultados divergentes em si quanto ao esperado, como é o exemplo de Fortaleza e Recife que apresentaram padrões de previsão muito diferentes entre os dois mapas. O mapa gerado pelo *Kernel Mexican Hat* para Fortaleza, as regiões com preços mais altos não ficaram apenas no litoral da cidade, entretanto se estenderam para Bairros distante da praia e próximos ao bairro de Passaré, e para o *Kernel Polynomial* as regiões com preços mais altos se limitou apenas às praias do litoral centro. Para Recife, o *Kernel Mexican Hat* também gerou resultados diferentes do esperado, prevendo os menores preços para o litoral do município, já o *Kernel Polynomial* divergindo da previsão do outro *kernel*, previu os maiores preços para o litoral. Contudo, o comportamento dos preços nas outras cidades permaneceu como o esperado previamente, tendo em vista os serviços selecionados para o modelo deste estudo.

Com os mapas apresentados e as análises descritivas propostas, o estudo consegue atender ao Objetivo Geral do trabalho de "desenvolver um novo método para a Tomada de Decisão Gerencial que auxilie a gerência na decisão locacional e na identificação de potenciais oportunidades de negócio". Além disso, os mapas do modelo mostraram uma visão geral do como o mercado hoteleiro se comporta em cada uma das cidades analisadas, salientando que as decisões estratégicas deste mercado podem ser aprimoradas com uso de tecnologia e, dessa forma, atendendo aos Objetivos Específicos do estudo.

Para estudos futuros, espera-se a aplicação de outros *kernels* no modelo para ampliar as análises e aplicar para outras cidades com objetivo de se obter novas perspectivas acerca desta pesquisa. A implementação de novos *kernels* ao modelo trará novas interpretações possíveis para os mapas já apresentados e também poderá se adaptar melhor aos dados para os casos em que o *kernel* não se adequou ao conjunto de informações da cidade. Além disso, espera-

se que com a mudança da escolha das amenidades e, por consequência, do perfil turístico do cliente, o modelo identifique novos comportamentos de preço diante dos serviços fixados.

## REFERÊNCIAS BIBLIOGRÁFICAS

- AKAMA, J. S.; KIETI, D. Tourism and socio-economic development in developing countries: A case study of mombasa resort in kenya. *Journal of sustainable tourism*, Taylor & Francis, v. 15, n. 6, p. 735–748, 2007.
- ALBUQUERQUE, A. F.; FILHO, E. E. Fatores de mortalidade de pequenas empresas: uma análise da produção acadêmica no período 2000-2010. *3Es-ENCONTRO DE ESTUDOS EM ESTRATÉGIA-ANPAD*, v. 5, p. 1–15, 2011.
- ALBUQUERQUE, P. H. *Previsão de séries temporais financeiras por meio de máquinas de suporte vetorial e ondaletas*. [S.l.], 2014.
- ARANHA, F.; FIGOLI, S. Geomarketing: memórias de viagem. *São Paulo, GeoBrasil*, p. 1–73, 2001.
- BARNEY, J. Firm resources and sustained competitive advantage. *Journal of management*, Sage Publications, v. 17, n. 1, p. 99–120, 1991.
- BASAK, D.; PAL, S.; PATRANABIS, D. C. Support vector regression. *Neural Information Processing-Letters and Reviews*, v. 11, n. 10, p. 203–224, 2007.
- BOSE, I.; MAHAPATRA, R. K. Business data mining—a machine learning perspective. *Information & management*, Elsevier, v. 39, n. 3, p. 211–225, 2001.
- BRANCO, G. M.; RIBEIRO, J. L. D.; TINOCO, M. A. C. Determinantes da satisfação e atributos da qualidade em serviços de hotelaria. *Produção. São Paulo, SP. Vol. 20, n. 4 (out./dez. 2010), p. 576-588*, SciELO Brasil, 2010.
- CÂMARA, G. *Anatomia de sistemas de informação geográfica*. [S.l.]: UNICAMP-Instituto de Computação, 1996.
- CARBONNEAU, R.; LAFRAMBOISE, K.; VAHIDOV, R. Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research*, Elsevier, v. 184, n. 3, p. 1140–1154, 2008.
- CARDOSO, C. E. P. *Geomarketing como Suporte de Decisão em Gestão do Território*. Tese (Doutorado) — Faculdade de Ciências Sociais e Humanas, Universidade Nova de Lisboa, 2011.
- CHAN, E. S.; WONG, S. C. Hotel selection: When price is not the issue. *Journal of Vacation Marketing*, Sage Publications London, Thousand Oaks, CA, and New Delhi, v. 12, n. 2, p. 142–159, 2006.
- CHEN, K.-Y.; WANG, C.-H. Support vector regression with genetic algorithms in forecasting tourism demand. *Tourism Management*, Elsevier, v. 28, n. 1, p. 215–226, 2007.

- CHEUNG, K.-W.; KWOK, J. T.; LAW, M. H.; TSUI, K.-C. Mining customer product ratings for personalized marketing. *Decision Support Systems*, Elsevier, v. 35, n. 2, p. 231–243, 2003.
- CHRISTALLER, W. *Central places in southern Germany*. [S.l.]: Prentice-Hall, 1966.
- CLIQUET, G. *Geomarketing: Methods and strategies in spatial marketing*. [S.l.]: John Wiley & Sons, 2013.
- COOPER, C.; FLETCHER, J.; FYALL, A.; GILBERT, D.; WANHILL, S. *Turismo: princípios e prática*. [S.l.]: Bookman, 2000.
- Coordenação de Serviços e Comércio IBGE. *Pesquisa de Serviços de Hospedagem*. 2016. Disponível em: <<https://biblioteca.ibge.gov.br/visualizacao/livros/liv100623.pdf>>.
- CORTES, C.; VAPNIK, V. Support-vector networks. *Machine learning*, Springer, v. 20, n. 3, p. 273–297, 1995.
- CRUZ, G. da; GÂNDARA, J. M. O turismo, a hotelaria e as tecnologias digitais. *Turismo-Visão e Ação*, v. 5, n. 2, p. 107, 2003.
- CUI, D.; CURRY, D. Prediction in marketing using the support vector machine. *Marketing Science*, INFORMS, v. 24, n. 4, p. 595–615, 2005.
- CUI, G.; WONG, M. L.; LUI, H.-K. Machine learning for direct marketing response models: Bayesian networks with evolutionary programming. *Management Science*, INFORMS, v. 52, n. 4, p. 597–612, 2006.
- DAVIES, R. *Marketing geography: with special reference to retailing* (methuen, london). 1976.
- DAVIS, C.; CÂMARA, G. *Arquitetura de sistemas de informação geográfica. Introdução à ciência da geoinformação*. São José dos Campos: INPE, 2001.
- Departamento de Comércio e Serviços IBGE. *Pesquisa Anual de Serviços*. 2015. Disponível em: <<https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=7150>>.
- DING, S.; WU, F.; SHI, Z. Wavelet twin support vector machine. *Neural Computing and Applications*, Springer, v. 25, n. 6, p. 1241–1247, 2014.
- DONAIRE, D.; SILVA, M. P. da; GASPAR, M. A. A rede de negócios do turismo: um estudo sobre suas características e implicações estratégicas. *Turismo-Visão e Ação*, v. 11, n. 1, p. 112–134, 2009.
- DRUCKER, H.; BURGESS, C. J.; KAUFMAN, L.; SMOLA, A. J.; VAPNIK, V. Support vector regression machines. **Advances in neural information processing systems**. [S.l.: s.n.], 1997. p. 155–161.
- FORUM, W. E. *The Travel & Tourism Competitiveness Report 2017*. 2017. Disponível em: <<https://www.weforum.org/reports/the-travel-tourism-competitiveness-report-2017>>.
- FURLAN, A. A. Geoprocessamento: estudos de geomarketing e as possibilidades de sua aplicação no planejamento do desenvolvimento socioeconômico. *GEOUSP: Espaço e Tempo (Online)*, n. 29, p. 97–105, 2011.

- GENTON, M. G. Classes of kernels for machine learning: a statistics perspective. *Journal of machine learning research*, v. 2, n. Dec, p. 299–312, 2001.
- GIOVANARDI, M.; LUCARELLI, A. Sailing through marketing: A critical assessment of spatiality in marketing literature. *Journal of Business Research*, Elsevier, v. 82, p. 149–159, 2018.
- GOH, C.; LAW, R. The methodological progress of tourism demand forecasting: a review of related literature. *Journal of Travel & Tourism Marketing*, Taylor & Francis, v. 28, n. 3, p. 296–317, 2011.
- GUNN, S. R. et al. Support vector machines for classification and regression. *ISIS technical report*, v. 14, p. 85–86, 1998.
- HAYKIN, S. *Neural networks a comprehensive introduction*. [S.l.]: Prentice Hall, New Jersey, 1999.
- HONG, W.-C.; DONG, Y.; CHEN, L.-Y.; WEI, S.-Y. Svr with hybrid chaotic genetic algorithms for tourism demand forecasting. *Applied Soft Computing*, Elsevier, v. 11, n. 2, p. 1881–1890, 2011.
- HOTELLING, H. Stability in competition. *The Economic Journal*, JSTOR, v. 39, n. 153, p. 41, mar 1929.
- HUANG, J.-J.; TZENG, G.-H.; ONG, C.-S. Marketing segmentation using support vector clustering. *Expert systems with applications*, Elsevier, v. 32, n. 2, p. 313–317, 2007.
- HUFF, D. L. A note on the limitations of intraurban gravity models. *Land Economics*, JSTOR, v. 38, n. 1, p. 64–66, 1962.
- KIM, G.; CHAE, B. K.; OLSON, D. L. A support vector machine (svm) approach to imbalanced datasets of customer responses: comparison with other customer response models. *Service Business*, Springer, v. 7, n. 1, p. 167–182, 2013.
- LANGLEY, P.; SIMON, H. A. Applications of machine learning and rule induction. *Communications of the ACM*, ACM, v. 38, n. 11, p. 54–64, 1995.
- LATOURE, P.; FLOU'CH, J. L. *Géomarketing: Principes, méthodes et applications*. [S.l.]: Ed. d'Organisation, 2001.
- LICHROU, M.; O'MALLEY, L.; PATTERSON, M. Place-product or place narrative (s)? perspectives in the marketing of tourism destinations. *Journal of Strategic Marketing*, Taylor & Francis, v. 16, n. 1, p. 27–39, 2008.
- LIM, E.-P.; CHEN, H.; CHEN, G. Business intelligence and analytics: Research directions. *ACM Transactions on Management Information Systems (TMIS)*, ACM, v. 3, n. 4, p. 17, 2013.
- LIN, W.-Y.; HU, Y.-H.; TSAI, C.-F. Machine learning in financial crisis prediction: a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, IEEE, v. 42, n. 4, p. 421–436, 2012.
- LING, C. X.; LI, C. Data mining for direct marketing: Problems and solutions. **KDD**. [S.l.: s.n.], 1998. v. 98, p. 73–79.
- LORENA, A. C.; CARVALHO, A. C. de. Uma introdução às support vector machines. *Revista de Informática Teórica e Aplicada*, v. 14, n. 2, p. 43–67, 2007.

- LOSCH, A. et al. *Economics of location*. Yale University Press, 1954.
- MACHADO, C. C.; FRANCISCO, E. de R.; RIBEIRO, J. G. da P. A geodemografia e o geomarketing na identificação de mercados potenciais. *Séries Demográficas*, v. 3, p. 77–95, 2016.
- MAGUIRE, D. J.; GOODCHILD, M. F.; RHIND, D. W. *Geographical information systems: principles and applications*. [S.l.]: Longman Scientific & Technical London, 1991.
- MITCHELL, T. M. Machine learning. 1997. *Burr Ridge, IL: McGraw Hill*, v. 45, n. 37, p. 870–877, 1997.
- MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina.
- NGAI, E. W.; XIU, L.; CHAU, D. C. Application of data mining techniques in customer relationship management: A literature review and classification. *Expert systems with applications*, Elsevier, v. 36, n. 2, p. 2592–2602, 2009.
- REILLY, W. J. *The law of retail gravitation*. [S.l.]: WJ Reilly, 1931.
- RODRIGUE, J.-P.; COMTOIS, C.; SLACK, B. *The geography of transport systems*. [S.l.]: Routledge, 2013.
- SHAW, M. J.; SUBRAMANIAM, C.; TAN, G. W.; WELGE, M. E. Knowledge management and data mining for marketing. *Decision support systems*, Elsevier, v. 31, n. 1, p. 127–137, 2001.
- SHELDON, P. J.; VAR, T. Tourism forecasting: a review of empirical research. *Journal of Forecasting*, Wiley Online Library, v. 4, n. 2, p. 183–195, 1985.
- SHIN, H.; CHO, S. Response modeling with support vector machines. *Expert Systems with applications*, Elsevier, v. 30, n. 4, p. 746–760, 2006.
- SMOLA, A. J.; SCHÖLKOPF, B. A tutorial on support vector regression. *Statistics and computing*, Springer, v. 14, n. 3, p. 199–222, 2004.
- SOMAN, K.; LOGANATHAN, R.; AJAY, V. *Machine learning with SVM and other kernel methods*. [S.l.]: PHI Learning Pvt. Ltd., 2009.
- SONG, H.; LI, G. Tourism demand modelling and forecasting—a review of recent research. *Tourism management*, Elsevier, v. 29, n. 2, p. 203–220, 2008.
- THEOBALD, W. *Turismo global*. [S.l.]: Senac, 2001.
- THUNEN, J. H. V.; HALL, P. G. *Isolated state*. [S.l.]: Pergamon, 1966.
- WANG, Y.; YU, Q.; FESENMAIER, D. R. Defining the virtual tourist community: implications for tourism marketing. *Tourism management*, Elsevier, v. 23, n. 4, p. 407–417, 2002.
- WEBER, A. *Ueber den Standort der Industrien*. [S.l.: s.n.], 1909.
- WEBER, A.; FRIEDRICH, C. J. Alfred weber's theory of the location of industries. University of Chicago, 1929.
- WEBER, A.; FRIEDRICH, C. J. *Theory of the Location of Industries*. [S.l.]: University of Chicago Press Chicago, Ill, USA, 1962.

WERNERFELT, B. A resource-based view of the firm. *Strategic management journal*, Wiley Online Library, v. 5, n. 2, p. 171–180, 1984.

WOLFE, P. A duality theorem for non-linear programming. *Quarterly of applied mathematics*, v. 19, n. 3, p. 239–244, 1961.

YEH, C.-Y.; HUANG, C.-W.; LEE, S.-J. A multiple-kernel support vector regression approach for stock market price forecasting. *Expert Systems with Applications*, Elsevier, v. 38, n. 3, p. 2177–2186, 2011.

YRIGOYEN, C. C. El geomarketing y la distribución comercial. *Investigación y marketing*, Aedemo, n. 79, p. 6–14, 2003.