

UNIVERSIDADE FEDERAL DE SANTA CATARINA
DEPARTAMENTO DE INFORMÁTICA E ESTATÍSTICA

Cinthy Weigel

**O USO DE BIG DATA PARA ANÁLISE DE OFERTA DE IMÓVEIS VIA AIRBNB
EM DESTINOS TURÍSTICOS**

Florianópolis

2019

Cintha Weigel

**O USO DE BIG DATA PARA ANÁLISE DE OFERTA DE IMÓVEIS VIA AIRBNB
EM DESTINOS TURÍSTICOS**

Trabalho Conclusão do Curso submetido ao curso
de Sistemas de Informação para a obtenção do
Grau de Bacharel em Sistemas de Informação.
Orientador: Prof. José Leomar Todesco

Florianópolis

2019

AGRADECIMENTOS

À minha mãe, Salete, pelo amor, incentivo, e apoio incondicional durante esta caminhada. À minha prima Kelly Cristina Weigel Pianna da Rocha, por todo apoio e mensagens de carinho e suporte. E a toda minha família, que sempre me incentivou.

Ao meu orientador, José Leomar Todesco, pelo suporte, paciência e auxílio durante o desenvolvimento do trabalho.

Aos meus amigos, em especial a Richard Junior Bertozzo, Júlia Nakayama Silveira, Sabrina Schutz, Luiza Cardoso, Leon Daros, Sarah Pfaffenzeller e Matheus Gentil por todo apoio e compreensão e à Laís Ferrigo Perazzolo, por toda ajuda que dedicou à mim durante o desenvolvimento do projeto.

E a todos que direta ou indiretamente fizeram parte da minha formação, os meus mais sinceros muito obrigada.

RESUMO

Sensores, dispositivos móveis, computadores e aplicações fazem parte da nossa vida diária. Nesses dispositivos circulam diariamente uma enorme quantidade de dados que, se utilizados corretamente e com segurança podem fornecer indicadores e *insights* sobre melhorias em diversos setores. Por meio desses dispositivos a forma como armazenamos e obtemos dados também mudou, acarretando no que chamamos hoje de *Big Data*. Este termo denomina grandes bases de dados com grande volume, variedade e que crescem em grande velocidade. Contudo, esta grande variedade de dispositivos trouxe complexidade quando o assunto é a obtenção desses dados, sendo necessário criar ferramentas de mineração capazes de navegar por arquivos web, buscando e tratando informações que possam ser úteis para a formação de conhecimento. A busca pode ser feita através de agentes inteligentes, como *Web Crawlers*. Estes são *scripts* que simulam a atuação de humanos em domínios da internet, procurando por dados que podem ser úteis para algum fim específico. Neste sentido o presente estudo utiliza técnicas de mineração de conteúdo através de *Web Crawlers* para a coleta de dados sobre habitações disponíveis em plataforma de aluguel de temporada, com o intuito de aplicar procedimentos de engenharia do conhecimento e *Data Warehouse* para criação de indicadores, por meio de um modelo dimensional e realização de consultas a fim de gerar gráficos informativos e mapas de calor, criando um ambiente de tomada de decisão para as partes interessadas no setor turístico da Ilha de Florianópolis. O ambiente de análise criado proporcionou a visualização dos indicadores referentes a habitações e ocupação, permitindo inferir em quais regiões da Ilha estão concentradas as habitações e a pouca variação de procura entre os meses de março, abril e maio, sendo abril o mês com maior demanda.

Palavras-chave: *Big Data*, *Web Crawler*, engenharia do conhecimento, *Data Warehouse*, Turismo, Florianópolis.

ABSTRACT

Sensors, mobile devices, computers and applications are part of our daily lives. These devices circulate a huge amount of data on a daily basis, which, if used correctly and safely, can provide indicators and insights into improvements across industries. Through these devices the way we store and obtain has also changed, resulting in what we now call Big Data. This term refers to large databases with large volume, variety and growing at great speed. However, this wide variety of devices has brought complexity when it comes to obtaining this data, and it is necessary to create data mining tools that can browse web files, seeking and treating information that may be useful for knowledge formation. Search can be done through intelligent agents such as Web Crawlers. These are scripts that simulate humans acting in Internet domains, looking for data that may be useful for some specific purpose. In this sense the present study uses content mining techniques through Web Crawlers to collect data on housing available in season rental platform, in order to apply knowledge engineering procedures and Data Warehouse to create indicators, through a dimensional model and consultation to generate informative graphs and heat maps, creating a decision-making environment for stakeholders in the tourism sector of Florianópolis Island. The analysis environment created provided the answers of the indicators related to housing and occupations, allowing to infer in which regions of the island are concentrated the dwellings and with little variation of demand between the months of March, April and May, and April or month with the highest demand.

Keywords: Big Data. Web Crawler. Knowledge Engineering. Data Warehouse. Tourism. Florianópolis.

LISTA DE FIGURAS

Figura 1 – Os 3V's do Big Data.....	22
Figura 2 – Categorias do Big Data.....	23
Figura 3 – Taxonomia da Mineração de Dados da Web.....	25
Figura 4 – Movimento Turístico Estimado em Santa Catarina.....	32
Figura 5 – Modelo Dimensional.....	39
Figura 6 – Total Máximo de Habitações Oferecidas por Mês	43
Figura 7 – Total de Habitações Ocupadas por Dia	44
Figura 8 – Demanda de Habitações por Mês	44
Figura 9 – Média de Demanda de Habitações por Mês	45
Figura 10 – Média de Dias Ocupados por Mês.....	46
Figura 11 – Máximo de Hóspedes em Habitações por Mês	47
Figura 12 – Máximo de Hóspedes em Habitações por Dia	48
Figura 13 – Habitações Disponíveis em Março, Abril e Maio.....	53
Figura 14 – Habitações Alugadas em Março.....	54
Figura 15 – Habitações Alugadas em Abril.....	55
Figura 16 – Habitações Alugadas em Maio.....	56

LISTA DE TABELAS

Tabela 1 – 10 Habitações mais Ocupadas em Março.....	49
Tabela 2 – 10 Habitações mais Ocupadas em Abril.....	50
Tabela 3 – 10 Habitações mais Ocupadas em Maio.....	50
Tabela 4 – Top 10 Habitações com maior Arrecadação em Março.....	51
Tabela 5 – Top 10 Habitações com maior Arrecadação em Abril.....	51
Tabela 6 – Top 10 Habitações com maior Arrecadação em Maio.....	52

LISTA DE ABREVIATURAS E SIGLAS

CSV Comma Separated Values

DW Data Warehouse

ETL Extraction, Transformation, Loading

JSON JavaScript Object Notation

RSS Rich Site Summary

SQL Structured Query Language

HTML Hypertext Markup Language

GLOSSÁRIO

Dias ocupados - dias nos quais determinada habitação estava sendo utilizada para acomodar hóspedes.

Habitação - Uma casa ou apartamento, podendo ser integral ou não integral, que está disponível para ser alugado.

Hóspede - Indivíduo que se acomoda, por tempo determinado, em uma determinada habitação.

Ocupação - Denomina que a habitação está ocupada.

Ocupação média - O número de dias que determinada habitação estava ocupada no mês, dividido pelo número de dias do mês.

Valor da diária - Valor por dia, pelo qual o hóspede paga, para se acomodar em determinada habitação.

SUMÁRIO

1 INTRODUÇÃO	15
1.1 OBJETIVOS.....	16
1.1.1 Objetivo Geral	16
1.1.2 Objetivos Específicos	16
1.2 JUSTIFICATIVA.....	17
1.3 RESULTADOS ESPERADOS.....	17
1.4 ORGANIZAÇÃO DOS CAPÍTULOS.....	18
2 REFERENCIAL TEÓRICO	19
2.1 REVOLUÇÃO DOS DADOS.....	19
2.2 BIG DATA.....	20
2.2.1 Big Data Analytics	23
2.3 MINERAÇÃO DE DADOS DA WEB.....	24
2.3.1 Extração de Dados da Web	27
2.3.2 Web Crawler	27
2.4 TURISMO.....	28
2.4.1 Turismo Inteligente	29
2.4.2 Turismo em Florianópolis	31
2.4.3 Airbnb: Aluguel por Temporada	33
3 DESENVOLVIMENTO	35
3.1 PROCEDIMENTOS METODOLÓGICOS.....	35
3.1.1 Definição da Localização	35
3.1.2 Desenvolvimento e Coleta de Dados	36
3.1.3 Organização dos Dados	37
3.1.4 Modelagem Dimensional	37
3.1.5 Carga de Dados no Modelo Dimensional	39
3.1.6 Frontend	40

4	RESULTADOS E ANÁLISES.....	42
5	CONCLUSÕES E TRABALHOS FUTUROS.....	58
5.1	CONCLUSÃO.....	58
5.2	TRABALHOS FUTUROS.....	59
	REFERÊNCIAS.....	60
	APÊNDICE A – Consulta: Oferta Máxima de Habitações Por Mês.....	66
	APÊNDICE B – Consulta: Demanda de Habitações Por Dia.....	67
	APÊNDICE C – Consulta: Demanda de Habitações por Mês.....	68
	APÊNDICE D – Consulta: Média Mensal de Demanda de Habitações....	69
	APÊNDICE E – Consulta: Média de Dias Ocupados por Mês.....	70
	APÊNDICE F – Consulta: Máximo de Hóspedes em Habitações por Mês	71
	APÊNDICE G – Consulta: Máximo de Hóspedes em Habitações por Dia	72
	APÊNDICE H – Script do Mapa de Calor.....	73
	APÊNDICE I – Consulta: Top 10 Habitações mais Ocupadas.....	75
	APÊNDICE J – Consulta: Top 10 Habitações com Maior Arrecadação..	76
	APÊNDICE K – Artigo.....	77
	APÊNDICE L – Código Fonte.....	97

1 INTRODUÇÃO

Dados sempre foram produzidos e consumidos, ao longo da história. No entanto, nem sempre foram de fácil acesso ou facilmente interpretáveis, como hoje. A produção de dados tem aumentado constantemente, e isso se deve a grande abundância de dispositivos digitais presentes em casas, ambientes corporativos e espaços públicos, das mídias sociais e da internet das coisas (KITCHIN, 2014).

O resultado do uso desses dispositivos e de mídias sociais são os dados digitais, que contemplam informações como roteiros de viagens, trabalho, comunicação, consumo e interesses de seus usuários. (RIAHI, 2018). A utilização de dispositivos móveis e o constante aumento do fluxo de dados estão modificando a forma como os dados são armazenados, distribuídos e interpretados, produzindo o que chamamos hoje de *Big Data* (KITCHIN, 2014).

O termo *Big Data* é bastante complexo e pode ser caracterizado por dados com três principais atributos: volume, variedade e velocidade. O volume, representa a grande quantidade de dado gerado. A variedade se refere aos tipos de dados (imagem, texto, entre outros) que são gerenciados por um sistema de informação e a velocidade se refere a frequência na qual os dados são gerados, compartilhados e armazenados (RIAHI, 2018). Dessa forma, o *Big Data*, representa uma grande mudança no escopo de conhecimento sobre determinados eventos (SCHROEDER, 2018).

Através dos dados obtidos por meio de páginas web e armazenados em bases de dados *Big Data* podemos criar novos conhecimentos para auxiliar na tomada de decisão (MCAFEE; BRYNJOLFSSON, 2012). Dados fazem parte da nossa vida cotidiana e possuem grande impacto em negócios e na sociedade como um todo. Para o setor turístico este fenômeno modificou a forma tradicional de viajar, desde a busca por roteiros em cidades turísticas, até a busca por acomodações (HUNTER et al., 2015).

A partir desse cenário, aplicações como o *Airbnb* começaram a surgir com o intuito de facilitar a busca por hospedagens. Considerado como uma inovação disruptiva, o aplicativo é simples e intuitivo, e busca diminuir a distância entre hóspede e anfitrião, bem como trazer uma experiência de acomodação

transformada, diferente das oferecidas por métodos tradicionais (GUTTENTAG, 2013; GUTTENTAG; SMITH, 2017).

O ambiente proposto pela plataforma, no âmbito econômico, poderia reduzir o desequilíbrio entre a oferta e a procura de acomodações durante períodos de maior movimentação turística devido a lista imensa de imóveis disponíveis para locação, explicando seu sucesso através da conexão entre pessoas e informação por meio da plataforma digital, preços abaixo do mercado tradicional e a cultura de apresentação de recomendações de outros usuários sobre determinada acomodação. (TAVOLARI, 2017)

Neste contexto, o presente estudo busca implementar uma aplicação prática que atenda aos interesses de órgãos responsáveis pelo turismo na região de Florianópolis e aos usuários de plataformas de aluguel por temporada. De modo específico, o objetivo deste trabalho é explorar técnicas de mineração de dados oriundos de páginas *web*, *data warehouse* e engenharia do conhecimento para agregar novas descobertas que auxiliem no processo de tomada de decisão do setor turístico da região de Florianópolis. Considerando desde a coleta de informações sobre os imóveis disponíveis para locação até a implementação de um esquema de visualização das informações, que possa contribuir para o acompanhamento do movimento turístico na região (MOODY; KORTINK, 2000).

1.1 OBJETIVOS

1.1.1 Objetivo Geral

O objetivo geral do presente estudo é propor um ambiente de análise dos dados sobre oferta de imóveis em Florianópolis disponibilizados no Airbnb aplicando técnicas de extração de conhecimento.

1.1.2 Objetivos Específicos

1. Coletar dados de site de aluguel por temporada (Airbnb), por meio de *Web Crawler*, disponibilizado no *GitHub* por Tom Slee.

2. Realizar o pré-processamento dos dados obtidos.
3. Criar um modelo dimensional para organização dos dados obtidos.
4. Disponibilizar um conjunto de indicadores para apoio a tomada de decisão sobre turismo.

1.2 JUSTIFICATIVA

Dados provenientes da Internet são criados de forma digital, por isso são fáceis de armazenar e gerenciar. São produzidos em grande escala e utilizam das mais novas tecnologias para sua manipulação, além de serem aplicados em diversos problemas que enfrentamos atualmente. Os dados presentes na Internet podem ser considerados um universo de respostas a questões as quais nem sequer pensamos. (ASKITAS; ZIMMERMANN, 2015). Deste modo, podemos dizer que os dados afetam diretamente a compreensão de fenômenos presentes em nosso dia a dia (SCHROEDER, 2018).

Neste sentido, o presente estudo justifica-se pela falta de informações referentes ao turismo em Florianópolis, principalmente quanto à ocupação e disponibilização de imóveis para a temporada, buscando explorar técnicas de mineração de conteúdo da web para buscar dados sobre estes imóveis, sendo capaz de apoiar agentes públicos e não públicos interessados em informações sobre oferta e ocupação, no caso, do *Airbnb*.

1.3 RESULTADOS ESPERADOS

- Aquisição de uma base de dados que forneça conhecimento sobre a oferta e a ocupação de imóveis disponíveis para aluguel de temporada na região de Florianópolis.
- Proporcionar um ambiente para discussão sobre o setor turístico na região de Florianópolis.

1.4 ORGANIZAÇÃO DOS CAPÍTULOS

O presente estudo foi organizado da seguinte forma: no capítulo 2 são contextualizados os referenciais teóricos que embasaram o desenvolvimento do trabalho iniciando com a Revolução dos Dados, seguido por *Big Data*, aprofundando em *Big Data Analytics*, nas seções 2.1, 2.2. e 2.2.1, respectivamente. Na seção 2.3., 2.3.1 e 2.3.2 são descritos os temas Mineração de Dados, Extração de Dados e *Web Crawler*. Nas seção 2.4 e suas subseções são apresentados um histórico sobre Turismo, Turismo Inteligente e Turismo em Florianópolis. Em seguida, no capítulo 3 são descritos os passos para o desenvolvimento do estudo. No capítulo 4, são apresentados os resultados e análises realizados. No capítulo 5, são apontadas as conclusões e possíveis trabalhos futuros

2 REFERENCIAL TEÓRICO

Neste capítulo serão apresentados temas que serviram de base para o desenvolvimento do presente estudo.

2.1 REVOLUÇÃO DOS DADOS

A revolução dos dados teve início entre os anos 50 e 70, com o advento da criação de computadores digitais e outras tecnologias e desde então vêm gerando grande impacto na sociedade em que vivemos (ASKITAS; ZIMMERMANN, 2015). Para (RIAHI, 2018) estamos nos tornando uma sociedade fundamentada pelo conhecimento, onde a informação atua como protagonista em nossa economia, cultura e política.

Essa revolução se deu através da grande "onda" de tecnologias de comunicação e informação (KITCHIN, 2014). Devido ao número expressivo de usuários em redes sociais e de tecnologias onipresentes sempre conectados a Internet, ou seja, o número de dados gerados aumenta a cada instante (JOHNSON et al., 2017).

Para (ČIŽIK, 2018), o advento da internet trouxe muitas mudanças para a forma como nós nos relacionamos e nos comunicamos com o mundo. Através do crescimento das tecnologias digitais, computadores e *smartphones*, boa parte de nosso comportamento na internet passou a ser armazenado, desde compras online, pesquisas em buscadores virtuais, até comentários que fazemos e vídeos que assistimos. A diferença dos dados armazenados atualmente, para os dados armazenados a 20 ou 30 anos atrás, no entanto, são significantes e estão listadas a seguir: disponibilidade em tempo real, maior quantidade e variedade (EINAV; LEVIN, 2013). Dessa forma, podemos atribuir o conceito de revolução digital ao processo que levou a sociedade a substituir tecnologias analógicas, por tecnologias digitais. Assim, promovendo mudanças importantes em nosso âmbito social, político e econômico. (SCHOENHERR, 2004)

Para (JOHNSON et al., 2017) a ciência está sendo revolucionada pela era da digitalização. Grandes quantidades de dados estão sendo coletadas por meio de

inúmeras fontes, sendo elas: dispositivos móveis, sensores, páginas *web*, entre outros. Estes dados são chamados de dados digitais e estão presentes em grande parte de nosso cotidiano. (RIAHI, 2018).

São considerados, por muitos, matéria prima para abstração de problemas, em diversas áreas, em categorias, medidas, gráficos e outras formas representativas e apresentam grande utilidade quando aplicados para descoberta do conhecimento, pois possibilitam a criação e inovação (KITCHIN, 2014). Para, (ALMEIDA, 2018) a evolução dos sistemas de informação e a digitalização de informações resultou em uma explosão de dados de interesse de grandes empresas. Tornando a informação essencial como diferencial competitivo, pois através dela, como observa (KITCHIN, 2014), o processo de tomada de decisão pode se tornar menos complexo e custoso.

Podemos definir a Revolução dos Dados como resultado da revolução digital, pois através dos desenvolvimentos tecnológicos dos últimos anos, vêm tornando os dados cada vez mais abertos e acessíveis e suportados por grandes infraestruturas. Além disso, reformulando o desenvolvimento de conhecimento, levantando questões importantes como segurança e privacidade. (KITCHIN, 2014).

Neste contexto, temos o surgimento do termo *Big Data*, que se refere a grande quantidade de dados digitais que podem ser gerados, armazenados e compartilhados através de redes conectadas em um curto espaço de tempo (KITCHIN, 2014). Desta maneira, sendo visto como uma alteração no escopo de conhecimento sobre determinados assuntos, significando a capacidade de agregar valor e fornecer à quem o está utilizando novos *insights* para resolução de problemas pois, reformula a maneira de construirmos conhecimento e de nos relacionarmos com a informação (MCAFEE; BRYNJOLFSSON, 2012, BOYD; CRAWFORD, 2012). O resultado dessa revolução se apresenta na forma de uma grande "boom" de aplicações construídas através de dados digitais e da automação de tarefas diárias, tornando nosso dia a dia mais produtivo e eficiente (JOHNSON et al., 2017).

2.2 BIG DATA

Desde sempre, empresas de diversos tamanhos e de todos os tipos têm se beneficiado com o gerenciamento e a análise de dados. Contudo, este tipo de processo ficou cada vez mais difícil pois, dados podem ser obtidos de diversas maneiras, podendo ser estruturados ou não estruturados, vindos de bancos de dados tradicionais ou de documentos, vídeos e até mesmo imagens. Além disso, também deve-se considerar dados oriundos das mais diversas tecnologias, assim como sensores e *smartphones* (HURWITZ; NUGENT; HALPER, 2013).

E apesar de todas essas informações formarem sentidos separadamente, a vantagem competitiva existe nas organizações quando elas conseguem reunir esses dados, de diferentes fontes e de diferentes tipos, em uma informação que agregará valor em sua tomada de decisão (HURWITZ; NUGENT; HALPER, 2013).

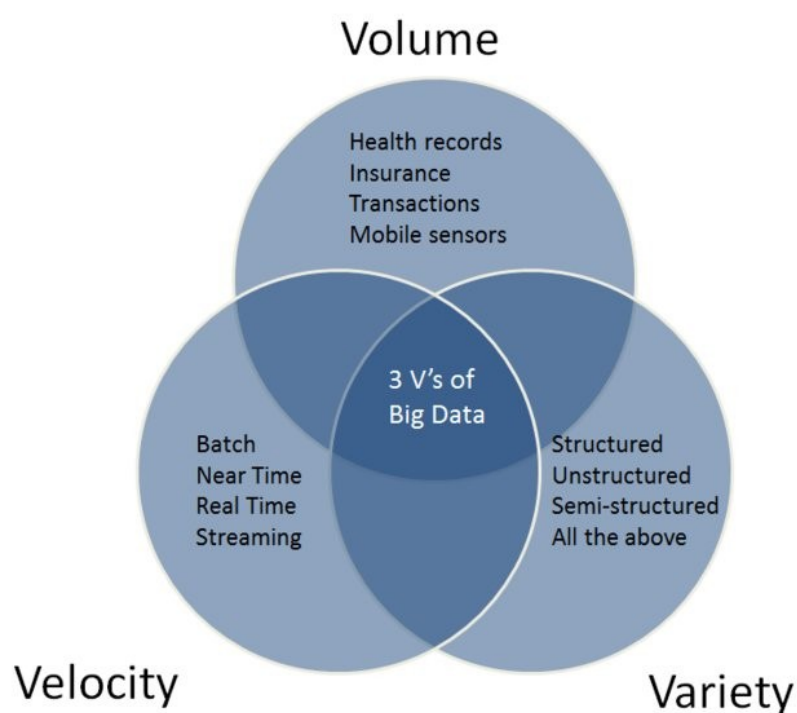
Existem diversas definições para o termo *Big Data*. Para (BOYD; CRAWFORD, 2012), pode ser definido como um fenômeno cultural, acadêmico e tecnológico que baseia-se na interação entre tecnologia, análise e mitologia, pois potencializa os algoritmos para a realização de análises, auxiliando na identificação de padrões, construindo conhecimentos de forma superior. Referindo-se a processos que não podem ser feitos em pequena escala, para extrair novas descobertas ou criar novas formas de valorizar produtos e negócios de forma que transforme também, o mercado, as organizações e as relações entre cidadãos e governo (ONBERGER; CUKIER, 2013).

Assim definimos *Big Data* como a capacidade de gerenciar volumes imensos de dados, variáveis, com velocidade e no tempo certo, para que possibilite uma análise em tempo real, permitindo uma reação também em tempo real. (HURWITZ; NUGENT; HALPER, 2013) É um objeto complexo, que se apresenta como uma solução projetada para fornecer acesso, em tempo real, à bancos de dados gigantes, contemplando também uma categoria de tecnologias e técnicas. Volume, variedade e velocidade são suas três principais características. Uma vez que falamos de imensas quantidades de dados, armazenados em *datasets* gigantes, que incluem dados estruturados, semiestruturados e não estruturados obtidos cada vez mais rápido (RIAHI, 2018).

Como mostra a figura 1, o grande volume de dados pode ser derivado de registros de saúde, transações financeiras, sensores móveis e seguros. Por meio

dessas diferentes fontes, dados, podem ser apresentados com as mais diversas características, sendo estruturados e não estruturados, em formato de texto, imagem ou som. Afetando também a velocidade de como esses dados são obtidos podendo ser adquiridos em tempo real, em lote ou através de *streaming*, tecnologia que utiliza a internet para realizar a transferência de dados (AUSTIN; KUSUMOTO, 2016, SIGNIFICADOS, 2017).

Figura 1 – Os 3 V's do Big Data



Fonte: (AUSTIN; KUSUMOTO, 2016)

No entanto, como observa (ALMEIDA, 2018), algumas novas características foram contempladas ao longo do tempo, sendo agrupadas em cinco categorias, são elas: variabilidade, veracidade, valor, visualização e solidez dos dados coletados.

Figura 2 – Categorias do Big Data.

Categories	Big Data Characteristics
Collecting data	Veracity, Variety
Processing data	Velocity, Volume
Integrity data	Validity, Variability, Volatility
Visualization data	Visualization
Worth of data	Value

Fonte: (ALMEIDA, 2018)

As tecnologias *Big Data* realizam conexões entre pontos de dados distintos, possibilitando a identificação de padrões que antes poderiam passar despercebidos. Sem responder o "por quê" ou "como", essas ligações podem responder perguntas que ainda nem foram pensadas. O principal objetivo do Big Data é destacar o uso de grandes *datasets* para descoberta de conhecimento, auxiliando o processo de tomada de decisão e tornando-o mais eficiente (AUSTIN; KUSUMOTO, 2016). Para as grandes organizações, o Big Data se torna um elemento crucial, pois permite a coleta, armazenamento e manipulação de grande quantidade de dados, com rapidez e no tempo certo para obtenção das melhores vantagens (HURWITZ; NUGENT; HALPER, 2013).

2.2.1 Big Data Analytics

O processo de análise de dados do tipo "Big Data" é complexo e formas tradicionais de programação e modelos simples de organização de dados não são suficientes para sua realização, principalmente pela quantidade de dados não estruturados presentes nestes *datasets*. Por isso, são necessárias ferramentas e técnicas específicas para realização de estudos com estas bases de dados (ALMEIDA, 2018).

Para (RUSSOM, 2011), *Big Data Analytics* pode ser definido pela aplicação de técnicas avançadas de análise ou técnicas avançadas de descoberta, como *data mining*, análises estatísticas e preditivas aplicadas sobre bases de dado Big Data, que possuem dados diversos e granulares. Ainda sobre sua definição, (RIAHI, 2018) estabelece *Big Data Analytics* como a coleta organização e estudo de *datasets* para identificação de padrões e outras informações.

Este processo, consiste em determinar, automaticamente, correlações, entre uma variedade imensa de dados que mudam constantemente, auxiliando a exploração dos mesmos. E para auxiliar neste método de descoberta diversas ferramentas, de análise são empregadas como, *data mining*, análises estatísticas, *queries* em SQL, visualização de dados e inteligência artificial (RUSSOM, 2011).

Neste contexto, grandes benefícios podem ser obtidos através da aplicação de técnicas de análise de dados nestes *datasets*. Através de descoberta de comportamentos de compra de seus clientes, organizações podem influenciar a compra de produtos similares ou obter uma vantagem competitiva para com seus concorrentes (RUSSOM, 2011).

Conseqüentemente, o principal objetivo do *Big Data Analytics* é identificar informações importantes em um curto espaço de tempo, auxiliando diversas áreas e principalmente empresas, no seu processo estratégico, tático e operacional (ALMEIDA, 2018).

2.3 MINERAÇÃO DE DADOS DA WEB

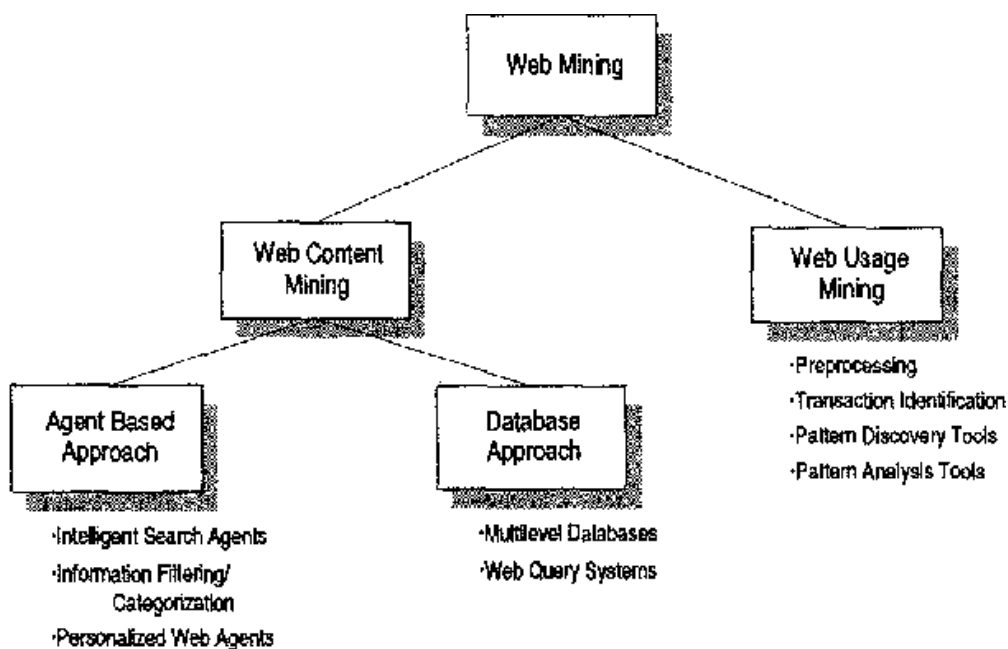
Com a grande explosão de dispositivos de comunicação conectados através da Internet, tornou-se indispensável a busca por ferramentas automáticas que capturem dados relevantes da Web para identificação de preferências e análise de padrões de seus usuários. Devido a estes fatores, a criação de sistemas inteligentes, que busquem por informações relevantes, "*server-side*" e "*client-side*" tornaram-se muito necessárias (COOLEY; MOBASHER; SRIVASTAVA, 1997).

A mineração de dados da *World Wide Web* tem estado em destaque pois reúne em seu meio, dados complexos e simples, além de possuir um volume

enorme de informações e para (SRIVASTAVA et al., 2000) pode ser descrita como o processo de aplicação de técnicas para descoberta de padrões de uso de determinados usuários em páginas específicas. Neste contexto, a mineração de dados pode ser conceituada como a descoberta e análise de informações úteis disponíveis pela Internet, utilizando técnicas como a Mineração de Conteúdo (COOLEY; MOBASHER; SRIVASTAVA, 1997), composta por documentos da *Web* e ferramentas presentes na Taxonomia da Mineração da Web (MARINHO; GIRARDI, 2003).

Como observado na figura 3, a mineração da web pode ser classificada em mineração de conteúdo e mineração de uso. Por sua vez, a mineração de conteúdo pode ser dividida em duas abordagens: baseada em agentes e a baseada em banco de dados. A abordagem baseada em agentes da mineração de conteúdo, é separada em três categorias distintas: Agentes Inteligentes; Filtragem/Categorização; e Agentes Personalizados. Agentes inteligentes buscam por dados relevantes através de particularidades do domínio e perfis de usuário. Para a filtragem/categorização de informação utiliza-se de técnicas automáticas de recuperação da informação, em páginas da web, para filtrar e categorizar a informação. Agentes personalizados estudam as preferências do usuário e buscam outras fontes de informação que sejam similares. Diferente da abordagem fundamentada em agentes, a baseada em banco de dados tem o intuito de organizar os dados semiestruturados obtidos da web em conjuntos mais estruturados de dados, fazendo o uso de formatos padrão de queries em bancos de dados e *data mining* para realizar análises (COOLEY; MOBASHER; SRIVASTAVA, 1997).

Figura 3 – Taxonomia da Mineração de Dados da Web



Fonte: (COOLEY; MOBASHER; SRIVASTAVA, 1997).

Uma das principais ações para descoberta do conhecimento se encaixa na criação de um *dataset* para as questões de Mineração de Dados. No entanto, além dos dados serem coletados de inúmeras fontes, também podem ser de tipos variáveis. Dessa forma, (SRIVASTAVA et al., 2000) classifica os dados nas seguintes categorias:

- Conteúdo: Geralmente composto por textos e gráficos. Se refere ao conteúdo de páginas web.
- Estrutural: Dados que se referem à organização da página web (HTML e XML).
- Dados de Uso: Se referem a hábitos de uso da internet do usuário, considerando endereços IP e referências a páginas específicas.
- Perfil do Usuário: Consiste em informações sobre hábitos de consumo e perfil do usuário de determinada página

Assim, podemos compreender que a Web possui um amplo conjunto de dados diversificados e dinâmicos e por causa dessas características, a web se

estabelece como "solo fértil" para aplicação de pesquisas em diversas áreas (PAL; TALWAR; MITRA, 2002).

2.3.1 Extração de Dados da Web

Usualmente, usuários buscam por informações na web através de pesquisas por palavras específicas ou apenas navegando entre as páginas. Essas são maneiras intuitivas de coletar dados, contudo não são indicadas para buscas específicas de determinados itens, pois são uma maneira fácil de "se perder". Para maior eficiência na pesquisa e coleta de dados, são utilizados programas especializados, chamados de *Crawlers*, que identificam os dados de interesse e os mapeiam para um formato possível de analisar (LAENDER et al., 2002).

Estes *Web Crawlers*, buscam simular a interação humana com o site alvo da pesquisa, e estão relacionados à Indexação da Web, que é uma forma de indexar informações presentes em documentos HTML na Internet através de robôs (VARGIU; URRU, 2013). Portanto, esses sistemas automatizados, desenvolvidos para coleta das informações, tem por objetivo transformar os dados não estruturados e salvá-los em bases de dados estruturadas (VARGIU; URRU, 2013).

2.3.2 Web Crawler

Chamamos de *Web Crawler*, o programa de software utilizado para extrair informações de páginas web, simulando a exploração humana da internet. A extração de informações, está intimamente ligada a indexação da web, técnica de recuperação de informações através de um robô, focando na transformação de dados não estruturados que podem ser transformados e analisados em bancos de dados ou planilhas (VARGIU; URRU, 2013).

A extração de dados, é realizada através do desenvolvimento de um programa automatizado que conectado a um servidor web solicita dados para extração das informações necessárias (MITCHEL, 2015). Segundo (CASTILLO, 2005), a arquitetura utilizada para criação de um *web crawler* geralmente se

apresenta em forma de cascata, na qual um *crawler* cria uma coleção de dados que serão indexados e pesquisados.

Ainda para (CASTILLO, 2005), a pesquisa de páginas web possui duas principais partes, uma *off-line* e outra *on-line*. A primeira é responsável por realizar o download de subconjuntos da web e é executada periodicamente e a segunda é executada toda vez que uma *query* é feita pelo usuário.

Hoje, páginas web são apresentadas aos usuários nos mais diversos formatos, assim, sua indexação é iniciada através de uma visão lógica de seu conteúdo. E segundo (CASTILLO, 2005) o método mais utilizado para esta fase é chamado de “*bag of words*”. Este método trata o conteúdo de cada página como um conjunto de palavras desordenadas e não estruturadas. Sendo assim, a arquitetura utilizada para seu desenvolvimento deve ser organizada levando em consideração que é parte essencial do projeto.

Contudo, apesar de tornar o processo de obtenção de dados através de páginas web mais simples e ágil, os *Web Crawlers* ainda possuem alguns desafios. O grande volume de informações, e a maneira rápida como páginas mudam e são removidas, tornam este processo mais complexo (CASTILLO, 2005).

2.4 TURISMO

Turismo, pode ser definido como o ato de passar algum tempo longe de casa, buscando divertimento, descanso e contentamento, enquanto há o uso de serviços comerciais oferecidos pelo destino alvo da viagem (WALTON, 2018). Visto como um fenômeno global, por sua infraestrutura imensa, sua importância se manifesta claramente através da sociedade, da política e principalmente na economia (GYR, 2010). Iniciando na Europa Ocidental, no século XVII e com origens na Antiguidade Clássica, teve seu princípio muito antes da palavra turismo surgir. Na Grécia Antiga e Roma, por exemplo, viagens com infraestrutura de suporte, e destinos essenciais com foco na experiência já existiam (WALTON, 2018).

“Desde a sua criação, o turismo polarizou: revela inúmeras visões, desde a aprovação total de seu potencial para enriquecer a

autorrealização combinada à recreação e à rejeição crítica devido à crença de que causa danos através da redução sistemática do entretenimento e da destruição ambiental evitável.” (GYR, 2010)

Neste sentido, podemos perceber o turismo moderno como um conjunto de atividades intensivas, organizadas de modo comercial e voltadas à negócios, impactando diretamente a economia e a sociedade como um todo. Que desde seus primórdios, para o início do século XXI, até os dias atuais se tornou uma das atividades econômicas essenciais em todo o mundo. Através da rápida evolução nos transportes de massa e no início do século 20, através das companhias aéreas, o turismo tornou-se um grande negócio global, criando oportunidades para viagens à passeio e/ou a trabalho (WALTON, 2018).

Hoje, a indústria do turismo recebe grande impacto proveniente da Web 2.0, segundo (SIGALA, 2007) a criação de ferramentas como o RSS, que permite a subscrição de usuários para recebimento de notícias, fez com que empresas do setor buscassem incorporá-la em seus próprios domínios, procurando manter seus clientes informados por meio de notícias, como também otimizando suas próprias ferramentas de busca.

Essa evolução dos meios de comunicação trouxe grande impacto à indústria turística. Segundo (HUNTER et al., 2015), a ligação entre as tecnologias da informação e o turismo fez com que a troca de informações utilizando dispositivos móveis conectados a internet, não fosse manipulada apenas por empresas agentes de viagem, tornando-se pública, por meio de comentários em redes sociais e *posts* com críticas pessoais acerca do destino alvo do turismo. Deste modo, a evolução da Web 2.0, criou uma espécie de coparticipação de usuários da internet na geração de experiências turísticas, impactando o modelo de negócio do setor (SIGALA, 2007).

2.4.1 Turismo Inteligente

Para (GRETZEL et al., 2015), no âmbito do turismo, inteligente é a palavra utilizada para descrever o processo de conexão e troca de informações de diversas fontes, oriundas de tecnologias Big Data, sensores e *smartphones* que, quando

aplicadas em conjunto são capazes de realizar a otimização de recursos, o desenvolvimento sustentável e a comodidade. Neste contexto, (GRETZEL et al., 2015) define inteligente como “palavra popular” para caracterizar o processo de desenvolvimento social, tecnológico e econômico que acontece por meio dessas tecnologias.

Com a grande explosão do uso de tecnologias da Web 2.0, grandes quantidades de informação circulam diariamente e quando unidas às experiências turísticas em diversas cidades do mundo, podem gerar grandes propostas de valor. Ao evento social e econômico surgido através do encontro de tecnologias da informação e de experiências turísticas, damos o nome de Turismo Inteligente (HUNTER et al., 2015).

Para (HWANG; PARK; HUNTER, 2015), a competição entre cidades que recebem o rótulo de cidades turísticas tem sido intensa durante os últimos anos. Sendo assim, estas cidades tendem a querer se desenvolver para adaptar-se ao gosto de um determinado público e mercado, empenhando-se para construir sua imagem para com seu mercado alvo. Essa construção, no entanto, acaba tornando-se complexa. Por causa, das tecnologias da informação utilizadas hoje, é preciso entender como a cidade é vista on-line e também off-line, desta forma podendo alcançar seu propósito de melhoria (HWANG; PARK; HUNTER, 2015).

A tecnologia da informação tem um papel essencial para a experiência turística atual, criando, desenvolvendo e aplicando sistemas técnicos de suporte ao progresso de negócios relacionados ao turismo por meio da internet, validando como este processo afeta a relação consumidor x comercial (HWANG; PARK; HUNTER, 2015).

Para aplicação de sistemas técnicos de apoio a decisão, destinos turísticos podem ser equipados com sensores, câmeras e outros dispositivos capazes de obter informação e armazená-las em bancos de dados através de redes wi-fi, para serem posteriormente analisados e utilizados de forma eficiente pelas partes interessadas (KAUR; KAUR, 2016).

A capacidade do turismo inteligente, de gerar conhecimentos ricos e profundos, tanto da visão do turista quanto da visão de negócio fez com que o método ganhasse rápido reconhecimento. Diferentemente de métodos tradicionais

de turismo, que utilizam principalmente hipóteses para testar e criar modelos, o turismo inteligente utiliza fontes de dados reais, de onde podem ser extraídas informações de qualidade (HWANG; PARK; HUNTER, 2015).

Toda essa construção, desde o estudo da visão de uma cidade turística até a aplicação de sistemas inteligentes capazes de auxiliar as tomadas de decisão, transforma um destino turístico em um destino turístico inteligente. Um destino turístico inteligente, é um local que possui uma boa infraestrutura de tecnologias da informação, otimizadas e empregadas para a construção de experiências turísticas cada vez melhores (GRETZEL et al., 2015).

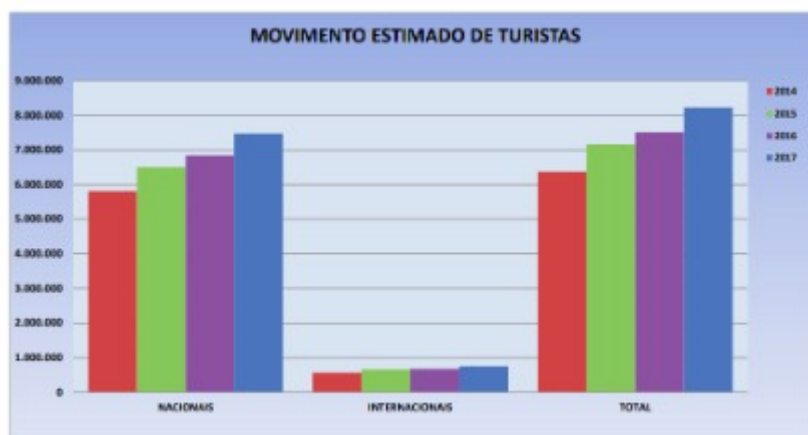
Com a tecnologia como principal chave para seu desenvolvimento, o turismo inteligente trouxe uma mudança de paradigma, levando o turismo tradicional para o futuro e tem sido adotada em diversos locais no mundo todo. A adesão, se bem estruturada, pode trazer inúmeros benefícios às partes interessadas no negócios e também para os turistas, que usufruirão de melhores experiências de viagem (KAUR; KAUR, 2016).

2.4.2 Turismo em Florianópolis

Para (LINS, 2000) o turismo em Florianópolis despontou há algumas décadas atrás, após a descoberta do litoral pelos Argentinos. Deste ponto em diante, o município cresceu em termos de fluxo turístico e de serviços disponibilizados para o turista, como por exemplo, os hotéis.

Como apresentado na pesquisa de demanda turística, disponibilizada no página do setor em Santa Catarina, em 2017 estimou-se o fluxo de 7.471.779 turistas nacionais e 751.921 turistas internacionais.

Figura 4 – Gráfico de Movimento Turístico Estimado em Santa Catarina



Fonte: (SANTUR, 2018)

Essa evolução se dá pelo fato de que a cidade possui atributos exuberantes, ligados à natureza, como praias, cachoeiras e paisagens muito belas. Além disso, também contempla a raiz açoriana, oriunda da imigração séculos antes. Esse conjunto de fatores, atrai turistas de todas as partes do Brasil e do mundo, impulsionando o crescimento do setor turístico na região (LINS, 2000).

No entanto, o fluxo turístico observado na região é sazonal e concentrado no verão, devido as inúmeras praias. Dessa forma, podemos compreender a grande demanda do setor turístico, principalmente do setor hoteleiro para atender uma grande quantidade de pessoas em um curto período de tempo e a sua subutilização no restante do ano (LINS, 2000).

Para (LINS, 2007), porém, cabe destacar a quantidade de serviços fornecidos à turistas na cidade, distribuídas nas mais variadas localidades. Além das opções pernoite, que se caracterizam por hotéis, albergues, *hostels*, entre outros. Somando os agentes comentados, com as instituições públicas do setor turístico percebe-se que Florianópolis possui o turismo como centro de sua atividade econômica.

Neste cenário, está claro que tal movimento turístico não despontaria caso a região não tivesse grande potencial natural, paisagístico, histórico e cultural. Portanto, é necessário que haja manutenção dessas riquezas para que Florianópolis continue a se desenvolver no setor turístico (LINS, 2007).

2.4.3 Airbnb: Aluguel por Temporada

A plataforma *Airbnb* foi fundada em 2008, descrevendo sua missão como “criar um mundo onde qualquer pessoa possa se sentir em casa em viagens autênticas, diversificadas, inclusivas e sustentáveis.” No site, pessoas de todo o mundo podem alugar mais de 7 bilhões de acomodações, entre casas inteiras e quartos, únicos listados em mais de 100,000 cidades em 191 países e regiões (AIRBNB, 2019).

Desde seu lançamento, o modelo de negócio inovador e baseado na internet, trouxe grandes mudanças para o modelo tradicional de acomodação turística. Em seu domínio de mercado, é a empresa que possui mais destaque e faz parte do grupo de companhias que permitem à pessoas comuns oferecerem suas casas ou apartamentos como acomodação, ou seja, representa um produto moderno que mudou a percepção de hospitalidade para a indústria (GUTTENTAG, 2013; GUTTENTAG; SMITH, 2017).

Comparado com meios tradicionais de estadia em uma visita turística, no entanto, habitações listadas podem faltar com algumas características de qualidade de serviço que muitos turistas procuram quando buscam onde ficar. Contudo, por ser inovador, geralmente tem preços abaixo do mercado e outros benefícios, entre eles a experiência de viajar acomodando-se em um “lar” e as informações sobre a região da viagem fornecidas por um anfitrião local, o que pode atrair inúmeros consumidores (GUTTENTAG, 2013).

O fato de possuir um modelo de negócio baseado na Internet e um cuidado único para com os turistas faz com que o *Airbnb* possa ser visto como uma “inovação disruptiva”. (GUTTENTAG; SMITH, 2017) Uma inovação disruptiva “é uma inovação que ajuda a criar um novo mercado e uma rede de valor, e eventualmente acaba impactando em um mercado e uma rede de valor existentes.” (CHRISTENSEN, 2015).

A ascensão do *Airbnb* é imensa e transformou a forma como turistas reservam suas acomodações, para uma plataforma de reservas em larga escala. Por meio da internet, o *Airbnb*, conseguiu ultrapassar obstáculos que, antes, dificultavam aos turistas a escolha da acomodação para sua viagem. Hoje, a

tecnologia da Web 2.0 facilita o trabalho dos anfitriões em promover sua casa, apartamento ou quarto, atingindo um número muito maior de pessoas. Além disso, usuários também podem descrever suas experiências, concedendo comentários e notas aos lugares que ficaram, diminuindo a distância entre hóspede e anfitrião (GUTTENTAG, 2013).

3 DESENVOLVIMENTO

A forma como adquirimos informações na *web* pode ser feita de maneira síncrona ou assíncrona. Na forma síncrona, temos a necessidade do envolvimento direto do pesquisador e se encaixam bem para pesquisas de âmbito qualitativo. De outro modo, as formas assíncronas de aquisição de dados não exigem envolvimento direto do pesquisador para cumprir seu papel, pois utilizam as mais novas tecnologias para obtê-las (FOINA, 2003).

Neste sentido, o presente estudo utiliza de técnicas assíncronas para coleta de dados sobre a disponibilidade de imóveis presentes em um site de aluguel por temporada em Florianópolis, nos meses de março, abril e maio de 2019. Para o levantamento de informações, os dados são coletados por mineração de conteúdo na web por meio de um *Web Crawler*, também conhecido como *Web Spider*, um programa focado para *World Wide Web* utilizado para obter informações específicas de páginas web (MACHADO et al., 2015).

A seção seguinte apresenta os procedimentos metodológicos realizados para alcançar os resultados esperados, descritos na seção 1.3. Nas subseções são descritos, os passos de definição da cidade alvo do estudo, o desenvolvimento do *web crawler* e a coleta de dados, a organização dos dados, processos de modelagem dimensional e carga de dados no modelo e implementação do *frontend*.

3.1 PROCEDIMENTOS METODOLÓGICOS

Nesta seção serão apresentados os passos realizados para o desenvolvimento deste trabalho, desde a definição da localização até a implementação do ambiente de visualização dos dados.

3.1.1 Definição da Localização

Florianópolis é situada no litoral de Santa Catarina e é também a capital do estado, possuindo cerca de 42 praias. Como observado na seção 2.4.2, a cidade

possui um fluxo turístico sazonal, concentrado principalmente nos meses de verão, entre dezembro e março. O fato de Florianópolis ser considerada uma cidade turística e a falta de informações sobre os imóveis disponíveis para aluguel por temporada do local, dificultando a pesquisa e o desenvolvimento de ambientes de tomada de decisão, que auxiliem as partes interessadas no setor a realizarem melhorias para recebimento destes turistas, levou à escolha da cidade para o desenvolvimento deste estudo.

3.1.2 Desenvolvimento e Coleta de Dados

Para obtenção dos dados sobre as habitações da região de Florianópolis, necessários para o estudo, foi utilizado o *Web Crawler* disponibilizado no *GitHub* por Tom Slee. Este *Web Crawler* é desenvolvido em *python*, com integrações de bibliotecas como *Requests* e *BeautifulSoup*. Os dados coletados via *Script*, utilizaram apenas o site do *Airbnb* como fonte e os dados ficam armazenadas em um banco de dados *Postgresql* contemplando os seguintes atributos dos imóveis:

- *room_id*: número único identificando um dos itens listados no site Airbnb. Este id possui uma *url* no site http://airbnb.com.br/rooms/room_id;
- *host_id*: número único que identifica o anfitrião;
- *room_type*: atributo que identifica o tipo de habitação, podendo ser “Apartamento/Casa inteira”, “Quarto Privativo” ou “Quarto Compartilhado”;
- *Borough*: atributo que identifica o município;
- *Neighborhood*: atributo que identifica o bairro;
- *Reviews*: número de *reviews* que a habitação recebeu;
- *Overall_satisfaction*: média de notas que a habitação recebeu (de 0 até 5);
- *Accommodates*: número de hóspedes que uma habitação pode acomodar;
- *Bedrooms*: número de quartos oferecidos na habitação;
- *Price*: o valor em dólares recebido por uma noite no dia da coleta;
- *Minstay*: a estadia mínima em uma visitada, escolhida pelo anfitrião;
- Latitude e Longitude.

(SLEE, 2017).

Para contemplação dos dados referentes às disponibilidades de cada imóvel na cidade, foi necessário a implementação de um novo *Web Scraper*. No desenvolvimento do mesmo, foram utilizadas as mesmas tecnologias presentes no primeiro *script* de retorno de dados sobre as acomodações, *python* e *requests*, simulando a ação de acesso a uma habitação individual. Os dados adquiridos por esse *script* foram armazenados em um arquivo em formato JSON. Deste, adicionamos os dados de dia e disponibilidade do imóvel. Os dados coletados representam o período de noventa dias, entre Março de 2019 e Maio de 2019.

3.1.3 Organização dos Dados

Ao final da coleta de dados, os arquivos JSON com os dados sobre disponibilidade do imóvel foram transformados em arquivos CSV, para facilitar o trabalho de limpeza e carga de dados no modelo dimensional, feitos posteriormente. Para transformação dos arquivos foi utilizado um serviço de conversão online. Este processo foi somente realizado nos dados obtidos no *script* de disponibilidade.

Finalizado o processo de conversão, os dados foram limpos e organizados, tratando valores nulos e normalizando atributos como *availability (true/false)* e *room_type (entire home/apt, shared room, private room)*. As colunas escolhidas para integrarem o projeto e serem adicionadas ao modelo foram: *dia_semana*, *dia_mes*, *mes*, *avaibility*, *num_quartos*, *num_banheiros*, *num_hospedes*, *latitude*, *longitude*, *id_habitacao*, *valor_aluguel*, *descricao*, *habitacao_integral*.

3.1.4 Modelagem Dimensional

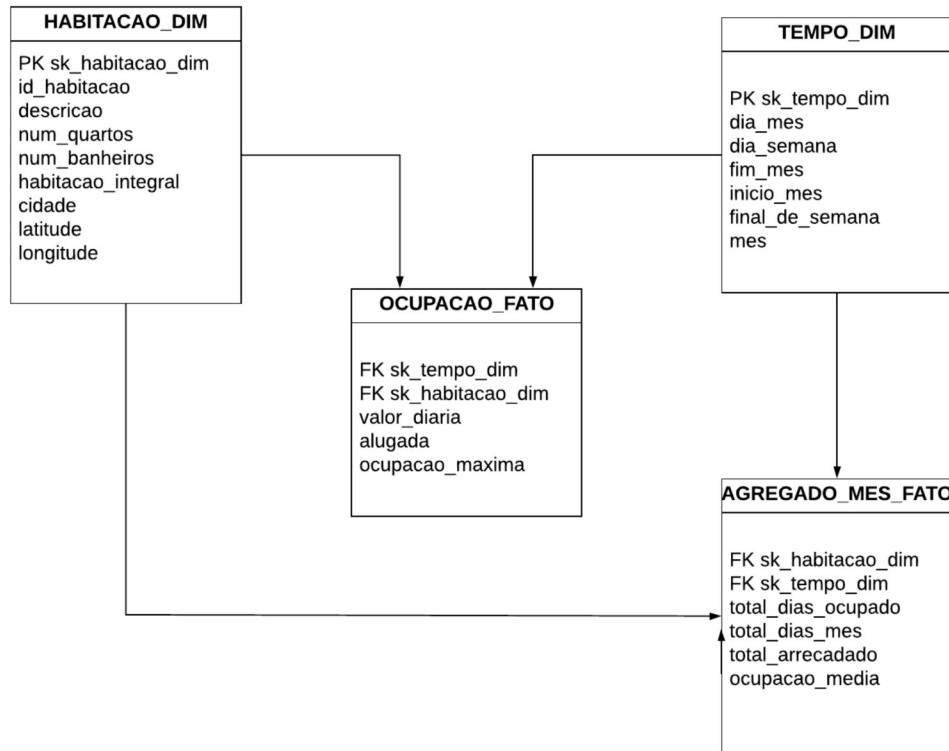
Iniciando a fase de análise do projeto, foi criado um *Data Mart*, que é um subconjunto lógico do *Data Warehouse* (KIMBALL et al., 1998). Este, diz respeito ao conjunto de tecnologias utilizadas para tomar decisões mais acertadas e de forma mais rápida, pois diferentemente de bancos de dados comuns, eles são orientados ao assunto, integrados e com variação de tempo (EL-SAPPAGH; HENDAWI; BASTAWISSY, 2011).

Portanto, essa diferença existente comparada à banco de dados operacionais, faz com que seja necessária a construção de uma modelagem dimensional que, tem por objetivo, facilitar a escrita e leitura de “*queries*” pelo usuário final. A principal forma de construir uma modelagem dimensional é o esquema estrela (MOODY; KORTINK, 2000).

Um esquema estrela consiste em inúmeras tabelas pequenas, chamadas de dimensões e uma ou mais tabelas maiores, chamadas de fato. As tabelas fato, contém todos os tipos de medidas relacionadas ao negócio, representando relações “muitos-para-muitos” e contendo uma ou mais chaves estrangeira que as ligam às suas tabelas dimensão. Por sua vez as tabelas dimensão, são definidas por suas chaves primárias, que serão referenciadas em suas respectivas tabelas fato, contendo as informações necessárias para criação de medidas e *group by* nas *queries* do *Data Mart* (MOODY; KORTINK, 2000; KIMBALL et al., 1998).

Com essas definições, a construção do modelo dimensional utilizado neste estudo, agrupou as variáveis de descrição da habitação, na tabela dimensão Habitação e os dados de descrição de tempo na tabela dimensão Tempo. Os dados utilizados para criação de medidas no modelo foram enviados para a tabela fato, Ocupação, que contém as chaves estrangeiras ligando-a às tabelas Habitação e Tempo. Para criação de medidas mais específicas a tabela fato Agregado Mês, que agrupa os dias totais ocupados por cada habitação, bem como sua ocupação média. As tabelas e seus respectivos atributos podem ser observados na figura 5.

Figura 5 – Modelo Dimensional



Fonte: Elaborada pelo autor (2019)

A grande vantagem da criação do esquema estrela para realizar a análise de dados é que reduz o número de tabelas, bem como o número de joins necessários para realizar consultas ao banco de dados, pois os relacionamentos também ficam reduzidos (MOODY; KORTINK, 2000).

3.1.5 Carga de Dados no Modelo Dimensional

Nesta etapa, após a organização dos dados em planilhas e da criação de um banco de dados que suportasse o modelo dimensional criado, foi realizado o processo de Extração, Transformação e Carregamento dos dados (ETL, do inglês *Extraction, Transformation e Loading*). A ferramenta de Extração busca ou recebe os dados de um ambiente operacional e pode ocorrer de diversas formas. Sendo um receptor passivo ou ativo. A transformação, é definida pelo processo de modificar

os dados, para que eles estejam consistentes com o modelo para o qual serão enviados. A etapa de carregamento representa a carga das informações extraídas e transformadas para o modelo (INMON, 2002).

A fim de realizar a carga dos dados no modelo dimensional, foi utilizado a ferramenta de *Data Integration*, Kettle da suite Pentaho. Para o processo de ETL, foram realizadas quatro cargas, para preencher as quatro tabelas presentes no modelo, dimensão Habitação, dimensão Tempo, fato Ocupação e fato Agregado Mês.

Para a carga da dimensão Habitação, foram utilizadas as planilhas CSV com os dados das habitações, construindo o processo para a não inclusão de valores repetidos nas tabelas, visto que, habitações podem estar disponíveis em um mês e não no mês seguinte. A carga da tabela dimensão Tempo, foi realizada também com um arquivo CSV, que continha as datas de 01 de março de 2019 a 31 de maio de 2019.

O processo de carga da tabela fato Ocupação, utilizou-se do *Lookup* para as tabelas já existentes, dimensão Tempo e dimensão Habitação, com a intenção de realizar o link entre o id da habitação, o id do tempo e a disponibilidade. Em seguida, para a última carga, da tabela fato Agregado Mês foi utilizada a tabela Ocupação para realizar o *group by* necessário para o preenchimentos das informações de total de dias ocupados por habitação, total arrecadado e ocupação média.

Ao final de cada um dos processos de carga, os dados foram carregados em um banco de dados *postgresql* que corresponde ao modelo criado e apresentado na seção 3.1.4, na figura 5, do presente estudo.

3.1.6 Frontend

Para diminuir a complexidade e tornar a informação do *Data Warehouse* o mais acessível possível, é necessário que se crie uma camada entre os dados e usuários. Essa camada tem por objetivo ajudar usuários finais a encontrar o que estão procurando, por esse motivo o *Frontend* pode ser definido como “a face pública do *Data Warehouse*” (KIMBALL et al., 1998). A camada de *frontend* pode ser construída através de serviços específicos, criados para este fim. Estes serviços

abstraem grande parte das atividades, inicialmente necessárias para o desenvolvimento do DW, como por exemplo: gerenciamento de buscas e *queries*, acesso e segurança e relatórios padrão (KIMBALL et al., 1998). Preenchendo a distância entre o usuário e os dados, aplicações desenvolvidas para o usuário final apresentam importância, tanto para os usuários técnicos quanto aos usuários do negócio. Os usuários técnicos se beneficiam da aplicação por meio dos relatórios, que podem ser utilizados como método de aprendizagem para alcançar certos resultados, além de monitorar a qualidade das *queries*. Mudando o panorama para os usuários de negócio, a aplicação representa um modo fácil e barato para que consigam os dados que procuram (KIMBALL et al., 1998).

Buscando construir relatórios com os dados organizados e carregados no modelo, conforme descrição na seção 3.1.5, a ferramenta escolhida foi o Metabase, uma aplicação *open source* com a capacidade de conectar-se ao banco de dados *postgresql*, onde estão armazenados os dados do modelo. Após conectar o banco de dados à ferramenta, foi possível criar os gráficos de oferta de habitações por mês, demanda de habitações por mês, habitações ocupadas por dia, média de habitações ocupadas por mês, máximo de ocupantes em todas as habitações por mês, máximo de ocupantes em todas as habitações por dia, top 10 habitações que passaram mais dias alugadas e top 10 habitações que mais arrecadaram. Estes relatórios serão apresentados no decorrer do capítulo 4.

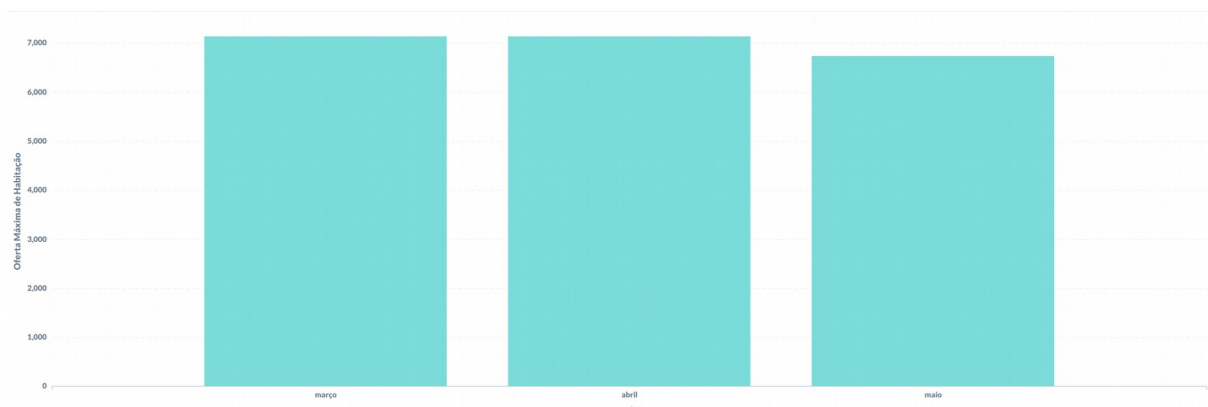
Para o desenvolvimento dos mapas de calor, que utilizaram as latitudes e longitudes das habitações, foi implementado um *script* na linguagem *python*, utilizando a biblioteca *pandas* e a biblioteca *gmpplot*. *Pandas* é uma biblioteca *open source*, que provê facilidades e alta performance para utilização de estrutura de dados e realização de análise e *gmpplot* fornece a facilidade de “plotar” dados no Google Maps. Os gráficos de mapa de calor também serão apresentados ao final do capítulo 4.

4 RESULTADOS E ANÁLISES

Após as etapas de desenvolvimento descritas no capítulo 3, foi possível a construção de gráficos que propiciaram *insights* e indicadores sobre os dados. Indicadores são medidas de performance calculadas através de um conjunto de métricas, deste modo são mais complexos e portanto podem servir como base para tomada de decisão (GOMES, 2017). Os dados coletados acerca da ilha de Florianópolis forneceram os indicadores para habitação, oferta, demanda e ocupação nos meses de março, abril e maio, fora da alta temporada turística na região, que acontece nos meses de verão, e estão apresentados nos gráficos a seguir.

O primeiro gráfico construído, com o auxílio da ferramenta de consulta Metabase, foi o de oferta máxima de habitações na região. A consulta necessária para isso está disponível no **apêndice A**. Esta conta é feita por meio da seleção do valor máximo de habitações disponíveis, da contagem de habitações distintas presentes em cada dia, agrupadas por dia e mês, e por fim agrupando o valor máximo por mês. O resultado da consulta está no gráfico da figura 6. Nele podemos observar que o mês onde há maior oferta de habitações é Março com 7.133. Seguido por Abril com 7.131 e Maio com 6.731. O número de habitações ofertadas corresponde ao final de temporada na Ilha de Florianópolis, diminuindo o número conforme a passagem dos meses.

Figura 6 – Total Máximo de Habitações Oferecidas por Mês



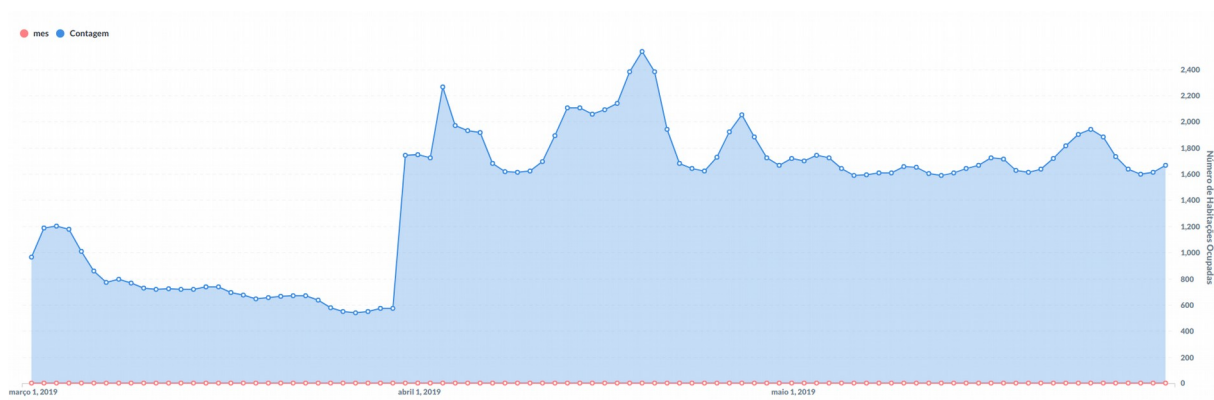
Fonte: Elaborado pelo Autor (2019)

Estabelecida a oferta de Habitações na região durante o período de três meses, também foi estabelecida a demanda, ou seja quantas dessas habitações disponíveis foram ocupadas.

Na figura 7 são apresentados os totais de habitação que estavam alugadas por dia. A consulta que gerou este gráfico está disponível no **apêndice B**. Nela foi selecionado o atributo dia do mês e a contagem distinta de habitações da tabela fato Ocupação, agrupando-as por dia. Na imagem podemos observar que o mês de março possui valores menores, comparando-o com os meses seguintes. Isso se deve à um problema técnico, onde o *web crawler* foi bloqueado pelo site, o que impossibilitou a obtenção de dados de algumas das habitações do mês de março.

Nos meses de abril e maio, podemos perceber que o gráfico segue um padrão de picos e vales, com as maiores contagens presentes nos finais de semana. Neste gráfico, também podemos perceber que a maior contagem encontrada foi no mês de abril, durante o final de semana do feriado de Páscoa.

Figura 7 – Total de Habitações Ocupadas por Dia

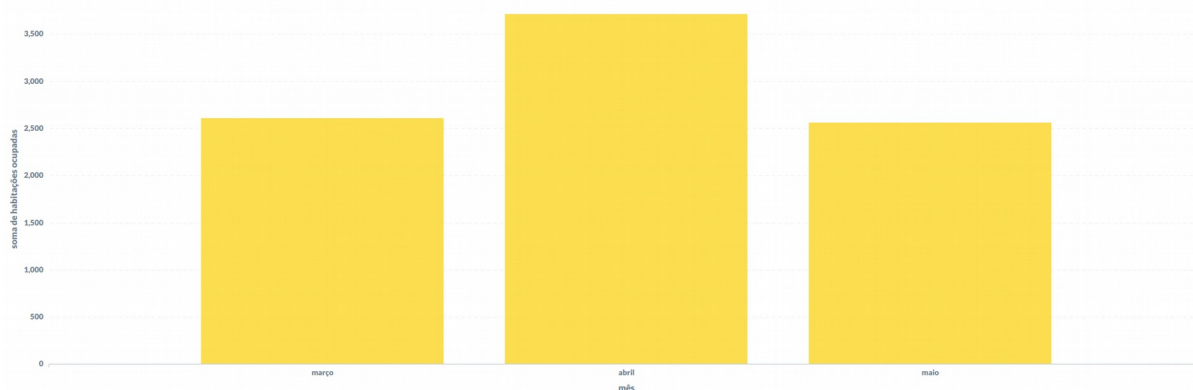


Fonte: Elaborado pelo Autor (2019)

O gráfico da figura 8 apresenta a contagem máxima de procura por habitações em cada mês. A consulta que gerou este gráfico está exposta no **apêndice C**. Sobre a tabela fato Ocupação foi selecionado o valor máximo de habitações em cada mês de uma sub consulta, que por si conta as habitações distintas de cada dia que estão ocupadas.

Considerando o número ofertado de habitações como 100%, em cada mês, como apresentadas no gráfico da figura 6, no mês de março temos 2,605 habitações ocupadas das 7.133 oferecidas, ou seja 36,5% das habitações foram ocupadas neste período. Já no mês seguinte o número de habitações ocupadas passou a ser 3,712 isto é 52,05% das 7.131 habitações oferecidas no mês de abril. Seguindo para o mês de maio temos 2,562 habitações oferecidas, sendo 39,39% das ofertadas.

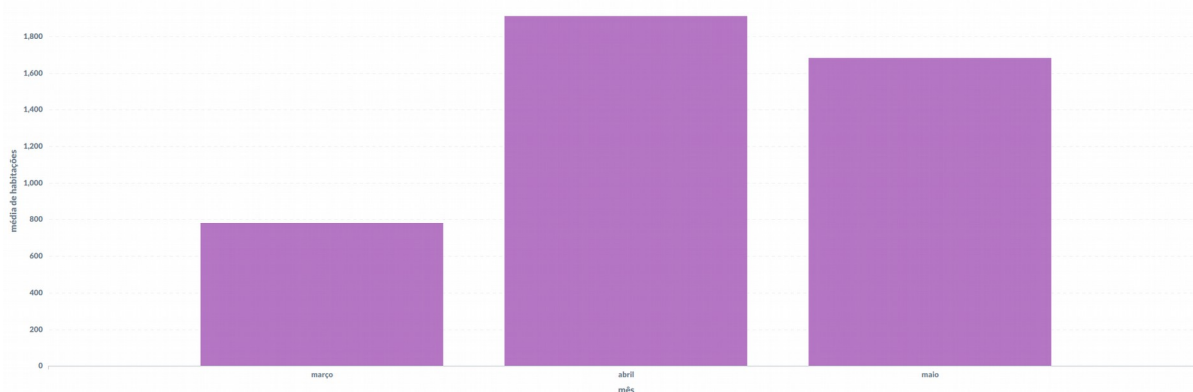
Figura 8 – Demanda de Habitações por Mês



Fonte: Elaborada pelo Autor (2019)

Deste modo temos a média de habitações ocupadas ou procuradas por mês, que é exibida no gráfico da figura 9. A consulta que gerou o gráfico está disponível no **apêndice D**. Primeiramente foram somados o número de habitações alugadas em cada dia em uma sub consulta. Na consulta principal, utilizando a função *average* foi gerado a média de demanda por habitações agrupadas por mês. A maior média foi para o mês de abril com 1.910 habitações. Seguidos pelo mês de maio com 1.680 habitações e março com 780 habitações

Figura 9 – Média de Demanda de Habitações por Mês



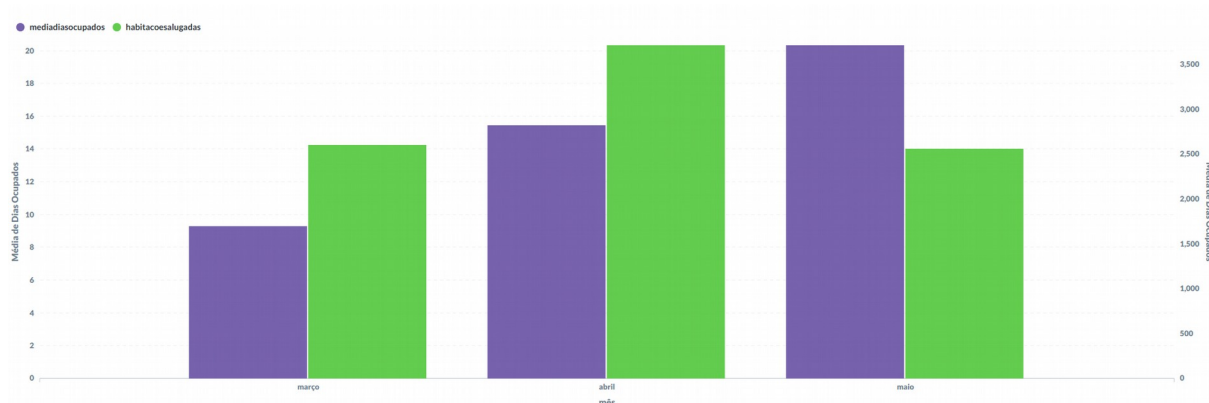
Fonte: Elaborada pelo Autor (2019)

Por meio das figuras 8 e 9 vemos que o padrão de demanda por habitações segue alguns critérios, no mês de março possivelmente pelo problema de bloqueio encontrado no *web crawler* a demanda é menor, contudo no mês de abril e maio a demanda volta a aumentar com alguns feriados de páscoa e Tiradentes, bem como o dia das mães no mês seguinte.

Em cada mês habitações podem estar ocupadas durante todo o período ou apenas durante alguns dias. Neste sentido, a figura 10 apresenta o gráfico com a média de dias ocupados por mês em habitações. A consulta que resultou no gráfico está no **apêndice D**. O cálculo de dias ocupados é feito realizando a contagem de

dias alugados em todas as habitações dividido pela contagem de habitações alugadas agrupadas mensalmente. Resultando no gráfico a seguir, onde a barra verde representa as habitações alugadas e a barra roxa representa a média de dias alugados. O resultado do gráfico mostra que o mês de maio tem a maior média de dias ocupados, sendo 20.33 dias, seguido por abril com 15.44 dias e março com 9.28 dias. Assim podemos perceber que apesar de possuir menos habitações em oferta e ter a segunda menor demanda, o mês de maio tem a maior média de quantidade dias alugados.

Figura 10 – Média de Dias Ocupados no Mês



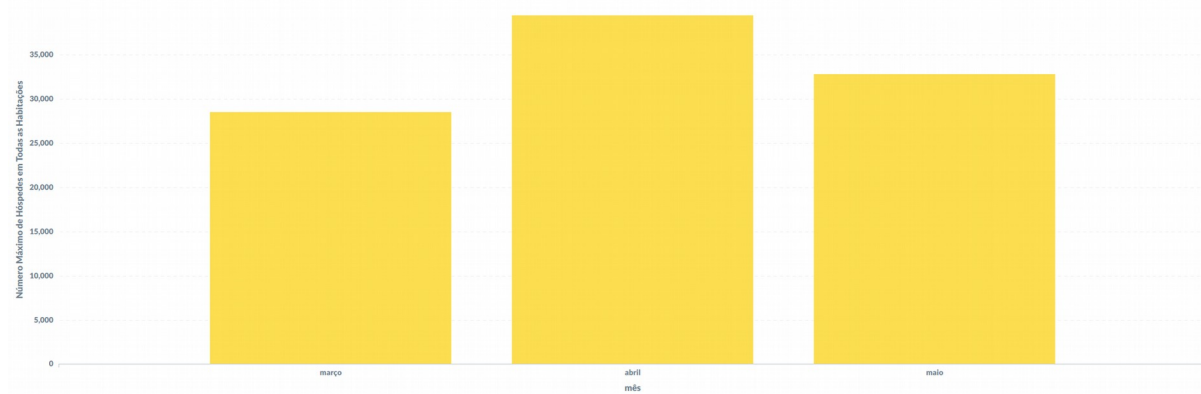
Fonte: Elaborado pelo Autor (2019)

Os dois próximos gráficos consideram o atributo ocupação máxima presente no modelo dimensional. Para cada habitação existe um número máximo de ocupantes que pode hospedar. Este número é definido pelo próprio anfitrião ao anunciar a sua casa ou apartamento no site de aluguel por temporada *Airbnb* e pode sofrer alterações enquanto a habitação estiver disponível. Os gráficos seguintes consideram que o número máximo de hóspedes esteve em cada uma das habitações ocupadas.

Na figura 11 são apresentados os resultados da consulta do **apêndice F**. Essa consulta retorna o número máximo de ocupantes que estiveram na região de

Florianópolis agrupando-os mensalmente. Este número é obtido através de uma sub consulta que soma o atributo ocupação máxima da tabela fato Ocupação, adicionando qualquer habitação que foi alugada por pelo menos um dia, realizando um agrupamento por dia e mês. Comparando aos gráficos de oferta mensal de habitação, este gráfico segue um padrão visualmente semelhante. O mês de março apresentou o resultado de 28,481 ocupantes máximos em habitações na ilha. O mês de abril apresentou o maior resultado, considerando o número máximo de ocupantes em cada habitação com 39.419 e o mês de maio apresentou o resultado de 32.766.

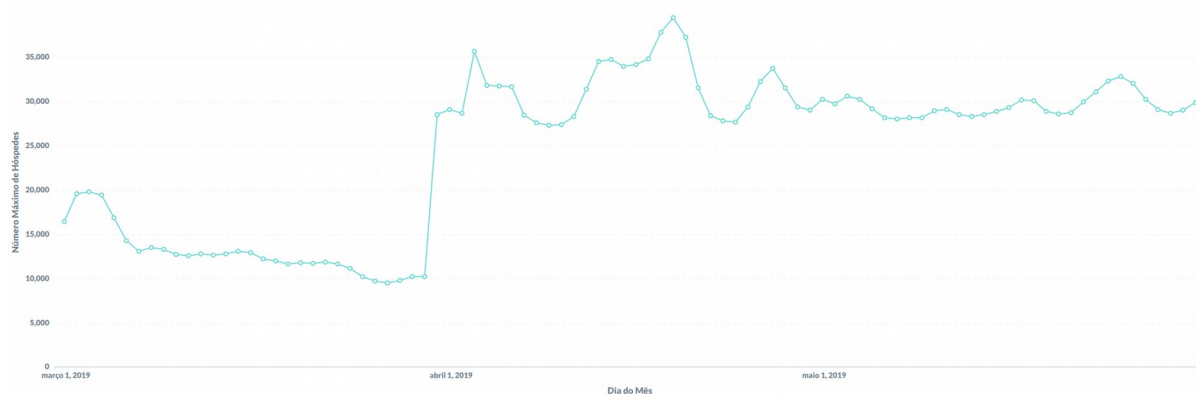
Figura 11 – Máximo de Hóspedes em Habitações por Mês



Fonte: Elaborado pelo Autor (2019)

O seguinte gráfico apresenta o número máximo de ocupantes agrupados por dia em todas as habitações marcadas como alugadas. Para esta consulta, através da função *sum*, foram somados os ocupantes máximos de cada habitação, agrupando-os por dia. Os dados foram selecionados da tabela fato Ocupação com os dados adicionais de dia selecionados através de um join com a tabela dimensão Tempo, a consulta está presente no **Apêndice G**. O gráfico se torna bem semelhante com o gráfico da figura 7, que por sua vez conta o total de habitações ocupadas por dia durante o período. Dessa forma o maior pico se encontra no dia 19 de abril, final de semana de páscoa e possuindo os maiores números entre os finais de semana de cada mês.

Figura 12 – Máximo de Hóspedes em Habitações por Dia



Fonte: Elaborado pelo Autor (2019)

Todos os cálculos foram feitos considerando a quantidade máxima de hóspedes que cada habitação pode hospedar, sem considerar a mudança de hóspedes em cada uma delas, por exemplo: um habitação pode estar ocupada durante todo o mês e haver um rotação de hóspedes a cada três ou cinco dias, bem como os mesmos hóspedes podem permanecer durante todo mês. Sendo assim o número obtido serve de base para análise de quantidade de fluxo turístico da região.

Considerando o atributo número de dias ocupados da tabela fato Agregado Mês foram levantadas as 10 habitações que mais estiveram ocupadas durante cada mês, contanto o número de dias total em que estiveram indisponíveis. A consulta foi feita selecionando o atributo total de dias ocupados da tabela fato Agregado Mês, ordenando descendente e buscando pelos primeiros dez resultados, a consulta está disponível no **Apêndice I**. No mês de março, as dez selecionadas passaram todos os dias do mês indisponíveis, duas delas são habitações não integrais, ou seja, habitações de aluguel para apenas um quarto ou quarto compartilhado. O valor da diária, que foi obtido através da primeira coleta referente as habitações, variou entre

65 e 511 reais. Outro fator interessante é que todas as habitações são de regiões próximas a praias, como Ingleses, Canasvieiras e Campeche.

Tabela 1 – 10 Habitações mais Ocupadas em Março

id_habitacao	descricao	num_quartos	num_banheiros	num_hospedes	cidade	latitude	longitude	habitacao_integral	valoraluguel	total_dias_ocupados
18089352	Excelente Padrão em Área Nobre - Portaria 24 horas	1	1	4	Florianopolis	-27,589,162	-48,552,581	habitacao integral	148	31
13688682	Quitinete (Studio) a 50m do mar Prala dos Ingleses	1	1	2	Florianopolis	-27,441,947	-48,386,572	habitacao integral	80	31
29117667	Cabanas da Lua	1	1	2	Florianopolis	-27,428,776	-48,407,442	habitacao não integral	244	31
21045807	Praia dos Ingleses! 100 metros do mar!	2	2	5	Florianopolis	-27,425,305	-48,402,274	habitacao integral	511	31
14593755	Apto. confortável a 50m da praia em Canasvieiras	1	1	4	Florianopolis	-27,429,148	-48,456,632	habitacao integral	110	31
20047607	Casa 2qtos (c/ Ar) p/ Família a 8mín a pé do Mar	2	2	4	Florianopolis	-27,430,600	-48,402,866	habitacao integral	120	31
15342452	Casa na praia, aproveite a natureza próximo ao mar	2	1	5	Florianopolis	-27,458,084	-48,384,239	habitacao integral	252	31
16883601	Apto para temporada nos Ingleses	1	1	4	Florianopolis	-27,451,536	-48,405,566	habitacao integral	125	31
15359185	Clean ,aconchegante e no Novo Campeche	1	1	4	Florianopolis	-27,662,722	-48,480,749	habitacao integral	156	31
9530891	Quarto confortável em região central na ilha	1	1	2	Florianopolis	-27,580,631	-48,510,719	habitacao não integral	65	31

Fonte: Elaborado pelo Autor (2019)

No mês de abril das 10 habitações selecionadas todas permaneceram o mês inteiro indisponíveis, com apenas uma habitação sendo habitação não integral, de quarto compartilhado ou quarto individual. Semelhante ao mês de março às habitações se encontram em regiões de praia da Ilha, Jurerê, Canasvieiras e Ingleses. O valor da diária foi de 65 a 601 reais e o número de hóspedes máximo pesquisado foi de 2 à 10.

Tabela 2 – 10 Habitações mais Ocupadas em Abril

id_habitacao	descricao	num_quartos	num_banheiros	num_hospedes	cidade	latitude	longitude	habitacao_integral	valoraluguel	total_dias_ocupados
9743168	Apartamento nos Ingleses Norte, Florianópolis - SC	2	2	6	Florianopolis	-27,430,920	-48,401,800	habitacao integral	260	30
2104066	Kitnete linda em Floripa	1	1	2	Florianopolis	-27,583,660	-48,526,350	habitacao integral	88	30
21472614	Quarto na lagoa 2 Ju	1	1	2	Florianopolis	-27,602,270	-48,466,170	habitacao não integral	150	30
16257786	Casa do Cauê	2	2	4	Florianopolis	-27,504,180	-48,418,000	habitacao integral	260	30
23553799	Floripa Cobertura com piscina Gaivotas Floripa	3	3	8	Florianopolis	-27,423,397	-48,404,220	habitacao integral	300	30
9547413	Casa muito aconchegante em Jurerê	2	1	6	Florianopolis	-27,441,220	-48,482,060	habitacao integral	249	30
19513624	Mountain & Bamboo fields View Apartment	1	1	2	Florianopolis	-27,576,584	-48,422,669	habitacao integral	65	30
30105969	Apartamento praia dos Ingleses-Floripa (211) SI	3	2	7	Florianopolis	-27,422,930	-48,402,930	habitacao integral	376	30
22272952	Casa no Cacupé para temporada	1	1	5	Florianopolis	-27,520,100	-48,517,590	habitacao integral	161	30
27427016	Casa de temporada Rio Vermelho	3	3	10	Florianopolis	-27,483,798	-48,414,587	habitacao integral	601	30

Fonte: Elaborado pelo Autor (2019)

O resultado da consulta para o mês de maio resultou na tabela 3. Todas as habitações selecionadas permaneceram indisponíveis durante os 31 dias do mês, todas sendo habitações integrais e com os valores de diária mais esparsos, de 65 até 4.999 reais. As habitações se encontram em região de praia da cidade, sendo elas Cachoeira do Bom Jesus, Jurerê Internacional, Ingleses e Lagoa da Conceição. O número de hóspedes máximo variou de 2 até 10.

Tabela 3 – 10 Habitações mais Ocupadas em Maio

id_habitacao	descricao	num_quartos	num_banheiros	num_hospedes	cidade	latitude	longitude	habitacao_integral	valoraluguel	total_dias_ocupados
23126358	Cobertura Top 100m da praia	4	4	10	Florianopolis	-27,426,968	-48,401,955	habitacao integral	899	31
21603142	Apartamento/Cachoeira Bom Jesus/quadra mar	2	2	5	Florianopolis	-27,423,063	-48,436,474	habitacao integral	501	31
10448366	Porto Del Mar - Vista Interna e Piscina	2	2	6	Florianopolis	-27,429,167	-48,465,054	habitacao integral	256	31
19513624	Mountain & Bamboo fields View Apartment	1	1	2	Florianopolis	-27,576,584	-48,422,669	habitacao integral	65	31
16065832	Casa entre mar e lagoa	2	1	6	Florianopolis	-27,606,630	-48,438,430	habitacao integral	201	31
2201377	Jurerê a 50m da praia	1	1	3	Florianopolis	-27,439,380	-48,487,320	habitacao integral	384	31
28005599	Apto "o melhor no Villa Giardino" Ingleses Fpolis	2	2	5	Florianopolis	-27,416,878	-48,404,413	habitacao integral	451	31
30004834	Mansão na Búzios perto do Café de lá Music!	5	5	10	Florianopolis	-27,440,680	-48,498,170	habitacao integral	4,999	31
10574578	Centrinho da Lagoa	2	2	4	Florianopolis	-27,606,420	-48,466,840	habitacao integral	279	31
23878735	Apartamento em Ingleses	1	1	3	Florianopolis	-27,449,620	-48,396,735	habitacao integral	68	31

Fonte: Elaborado pelo Autor (2019)

Da mesma forma em que há habitações que ficaram ocupadas durante boa parte do período analisado, algumas habitações arrecadaram mais diárias durante o tempo em que ficaram alugadas. As consultas que geraram as tabelas abaixo estão disponíveis no **apêndice J**. A consulta busca pelo atributo total arrecadado da tabela fato Agregado Mês e ordena-os do maior para o menor, buscando pelos primeiros 10 resultados.

No mês de março podemos observar que as habitações que mais arrecadaram possuíam os valores de diárias mais caros, entre 958 reais e 5.948. Com a diária de valor mais alto, o número de hóspedes ficou entre duas e dezesseis pessoas. Como observado na tabela 4 a seguir, as habitações com maior

arrecadação, são de alto padrão e estão nas praias conhecidas por serem mais sofisticadas da Ilha de Florianópolis.

Tabela 4 - Top 10 Habitações com Maior Arrecadação em Março

id_habitacao	descricao	num_quartos	num_banheiros	num_hospedes	cidade	latitude	longitude	habitacao_integral	valoraluguel	total_arrecadado
21986720	Costão do Santinho Resort - Vilas Portuguesas - Ap. Superio (Website hidden by Airbnb) Dias - 4 adultos e 2 crianças	2	2	6	Florianopolis	-27,466,276	-48,380,509	habitacao integral	2,704	83,824
27898194	Suite Master Costão do Santinho	1	1	3	Florianopolis	-27,444,379	-48,377,919	habitacao não integral	1,731	53,661
28956907	Casa temporaria Ingleses	3	5	7	Florianopolis	-27,432,622	-48,403,978	habitacao não integral	1,499	44,970
16602339	Férias dos Sonhos	2	2	5	Florianopolis	-27,462,674	-48,377,943	habitacao integral	1,299	40,269
10032944	Cobertura na Praia dos Ingleses - Florianópolis	3	2	6	Florianopolis	-27,424,220	-48,401,840	habitacao integral	1,201	37,231
7642148	Prédio inteiro - 44 pessoas !!!	1,000	8	16	Florianopolis	-27,441,930	-48,385,500	habitacao integral	5,498	32,988
19907262	Cobertura em Condomínio a 50 Metros da Praia	5	3	10	Florianopolis	-27,430,608	-48,397,162	habitacao integral	1,299	31,176
20520426	Cobertura área nobre dos Ingleses	2	2	4	Florianopolis	-27,423,730	-48,403,280	habitacao integral	999	30,969
999011	Vista espetacular Cobertura frente ao mar Ingleses	4	4	8	Florianopolis	-27,418,280	-48,402,500	habitacao integral	967	29,977
21986719	Costão do Santinho Resort - Hotel Vilas Portuguesas - Apto Standard - 7 Dias - 2 Adultos + 1	1	1	2	Florianopolis	-27,464,877	-48,380,066	habitacao integral	958	29,698

Fonte: Elaborada pelo Autor (2019)

No mês de abril as habitações com maior arrecadação seguem os mesmos padrões do mês anterior, estando localizadas em região de alto padrão. Essas habitações hospedam um número máximo de hóspedes entre quatro e dez pessoas e são em maioria habitações integrais.

Tabela 5 - Top 10 Habitações com Maior Arrecadação em Abril

id_habitacao	descricao	num_quartos	num_banheiros	num_hospedes	cidade	latitude	longitude	habitacao_integral	valoraluguel	total_arrecadado
28249878	COBERTURA 2 SUITES SPA PISCINA EM JURERE	2	3	6	Florianopolis	-27,444,193	-48,489,252	habitacao integral	24,999	749,970
22451529	Apto 203 - Boulevard Ponta das Canas	1	1	4	Florianopolis	-27,403,262	-48,426,904	habitacao integral	10,001	300,030
83547	Best place in Jurerê Internacional, Florianópolis	4	5	8	Florianopolis	-27,437,320	-48,512,950	habitacao integral	5,000	150,000
14551534	Ótima casa bem próxima ao mar	5	5	10	Florianopolis	-27,435,960	-48,501,950	habitacao integral	3,250	97,500
22231063	Jurerê Internacional Brazil Beach House	5	5	10	Florianopolis	-27,439,700	-48,498,630	habitacao integral	3,001	90,030
30004834	Mansão na Búzios perto do Café de lá Music!	5	5	10	Florianopolis	-27,440,680	-48,498,170	habitacao integral	4,999	84,983
21924933	Beautiful home in Jurerê Internacional Brazil	7	5	12	Florianopolis	-27,440,480	-48,506,570	habitacao integral	2,500	67,500
1302206	Casa magnifica de frente para o mar	5	7	10	Florianopolis	-27,415,950	-48,429,320	habitacao integral	2,202	66,060
3944422	Casa com piscina e vista da lagoa e mar	4	4	8	Florianopolis	-27,607,440	-48,485,140	habitacao integral	1,949	58,470
27692835	Elite Praia Mole Villa	4	400	10	Florianopolis	-27,600,904	-48,436,043	habitacao integral	1,874	56,220

Fonte: Elaborada pelo Autor (2019)

Em maio o padrão de habitações se repete, sendo que as habitações com maior arrecadação se encontram em regiões de alto padrão e próximas ao mar. O número de hóspedes máximo que pode ser recebido também permanece maior, entre quatro e quatorze.

Tabela 6 - Top 10 Habitações com Maior Arrecadação em Maio

id_habitacao	descricao	num_quartos	num_banheiros	num_hospedes	cidade	latitude	longitude	habitacao_integral	valoraluguel	total_arrecadado
28249878	COBERTURA 2 SUÍTES SPA PISCINA EM JURERÊ	2	3	6	Florianópolis	-27,444,193	-48,489,252	habitacao integral	24,999	774,969
22451529	Apto 203 - Boulevard Ponta das Canas	1	1	4	Florianópolis	-27,403,262	-48,426,904	habitacao integral	10,001	310,031
83547	Best place in Jurerê Internacional, Florianópolis	4	5	8	Florianópolis	-27,437,320	-48,512,950	habitacao integral	5,000	155,000
30004834	Mansão na Búzios perto do Café de lá Music!	5	5	10	Florianópolis	-27,440,680	-48,498,170	habitacao integral	4,999	154,969
27902175	Casa completa para 14 hóspedes com piscina	7	5	14	Florianópolis	-27,439,030	-48,504,680	habitacao integral	4,117	111,159
14551534	Ótima casa bem próxima ao mar	5	5	10	Florianópolis	-27,435,960	-48,501,950	habitacao integral	3,250	100,750
8652841	FLORIANOPOLIS EN LA MEJOR PLAYA A 30 METROS DEL MAR	4	3	8	Florianópolis	-27,424,050	-48,438,810	habitacao integral	2,228	69,068
1302206	Casa magnífica de frente para o mar	5	7	10	Florianópolis	-27,415,950	-48,429,320	habitacao integral	2,202	68,262
3944422	Casa com piscina e vista da lagoa e mar	4	4	8	Florianópolis	-27,607,440	-48,485,140	habitacao integral	1,949	60,419
27692835	Elite Praia Mole Villa	4	400	10	Florianópolis	-27,600,904	-48,436,043	habitacao integral	1,874	58,094

Fonte: Elaborada pelo Autor (2019)

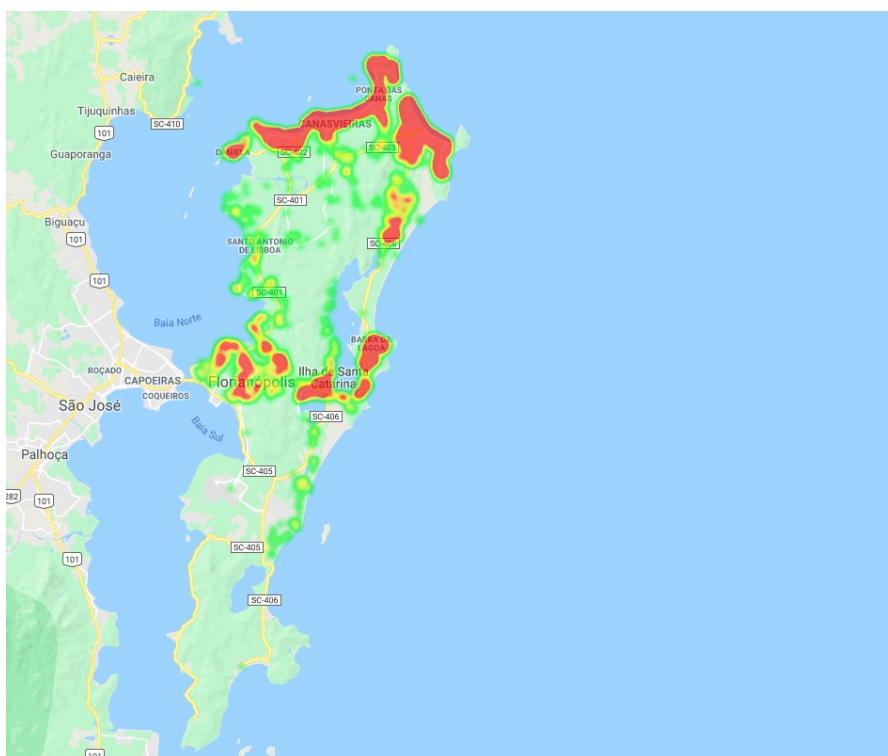
Para os mapas de calor das figuras 13, 14, 15 e 16 foram utilizadas as informações sobre latitude e longitude de cada habitação presentes na tabela dimensão Habitação. Para a implementação do mapa foi desenvolvido um *script* na ferramenta *Jupyter Notebook* que está disponível no **apêndice G**. O programa utiliza as bibliotecas *pandas* e *matplotlib* para desenhar o mapa de calor com as habitações disponíveis e alugadas. Primeiramente é realizada a consulta que seleciona as habitações distintas e disponíveis durante o intervalo do dia 01 de março de 2019 ao dia 31 de maio de 2019, selecionando também a latitude e longitude de cada uma delas. Essa consulta é feita na plataforma *Metabase* que disponibiliza um arquivo *CSV* para download dos dados retornados pela consulta. Feito este processo, é possível gerar o mapa de calor com o *script* anteriormente criado. A biblioteca *pandas* faz a leitura do *CSV* e com a biblioteca *matplotlib* é possível conectar as informações ao *Google Maps* gerando um arquivo *HTML* com o mapa de calor. No mapa as regiões em vermelho mostram a maior concentração em número e as regiões em verde as menores concentrações.

A figura 13 apresenta o gráfico de calor obtido para todas as habitações disponíveis nos meses de março à maio. Nele podemos observar que há grande concentração de habitações no norte da Ilha de Florianópolis, em Canasvieiras, Jurerê, Costão do Santinho, Daniela e Santo Antônio de Lisboa. Outras regiões com

alta concentração de habitações disponíveis são Lagoa da Conceição, Barra da Lagoa, Itacorubi, Carvoeira e Centro.

Nas regiões do Sul da Ilha também há alguma concentração maior, representada pela região em amarelo, no Campeche. Muitas das regiões próximas à praias também apresentam alguma concentração de imóveis disponíveis, apesar de ser menor que nos bairros de praia.

Figura 13 – Habitações Disponíveis em Março, Abril e Maio



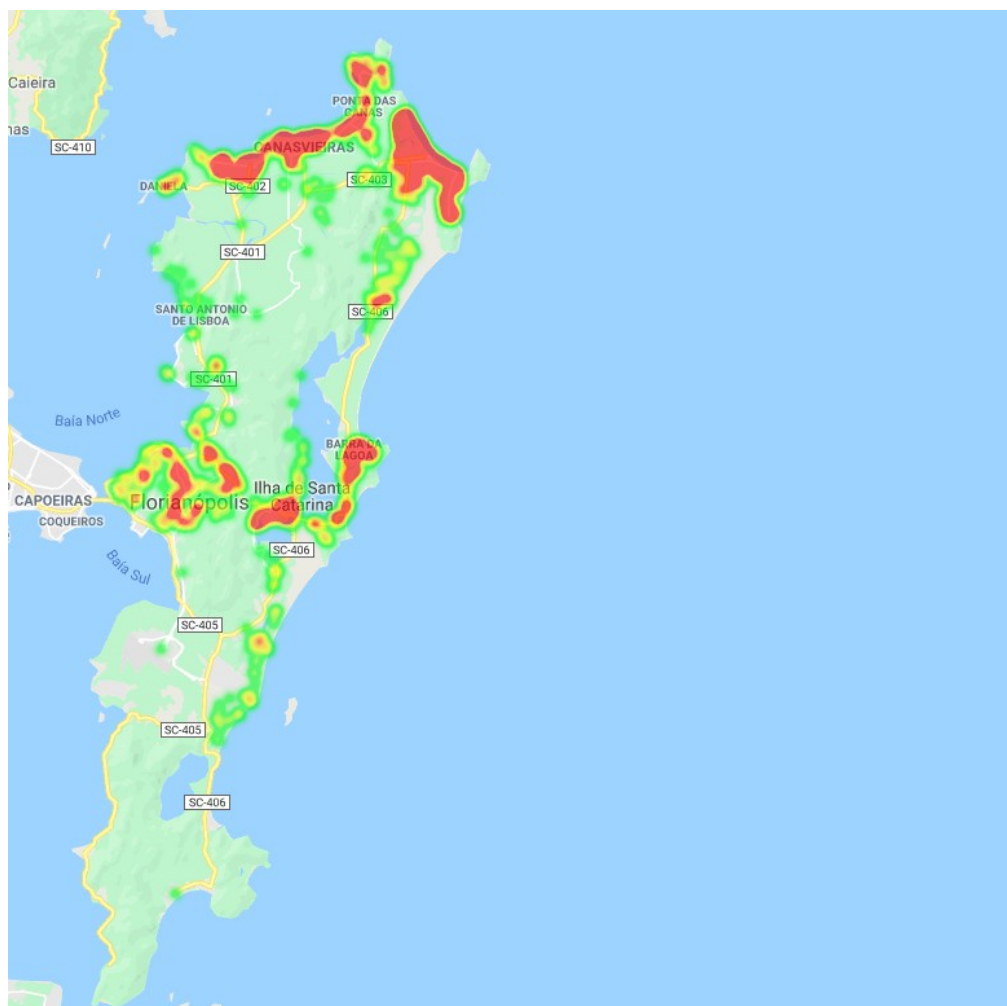
Fonte: Elaborado pelo Autor (2019)

A seguir os mapas de calor apresentados referem-se apenas às habitações que estiveram ocupadas, ou seja marcadas como alugadas na tabela fato Ocupação. Para a implementação de cada um deles foi realizado o mesmo processo

descrito para a figura 12, selecionando as habitações distintas da tabela, junto com suas latitudes e longitudes.

No mês de março de 2019 a concentração de habitações permanece nas regiões próximas à praia e está semelhante ao gráfico de habitações totais disponíveis durante o período. Nas regiões do norte da Ilha há uma menor concentração, observado pela diminuição da região em vermelho em Jurerê, Daniela, Costão do Santinho e Ingleses. Na região do Rio Vermelho podemos observar que há menos habitações ocupadas do que disponíveis. Conclui-se a partir do mapa da figura 14 que as habitações alugadas permanecem em maioria nas regiões de praia e centro.

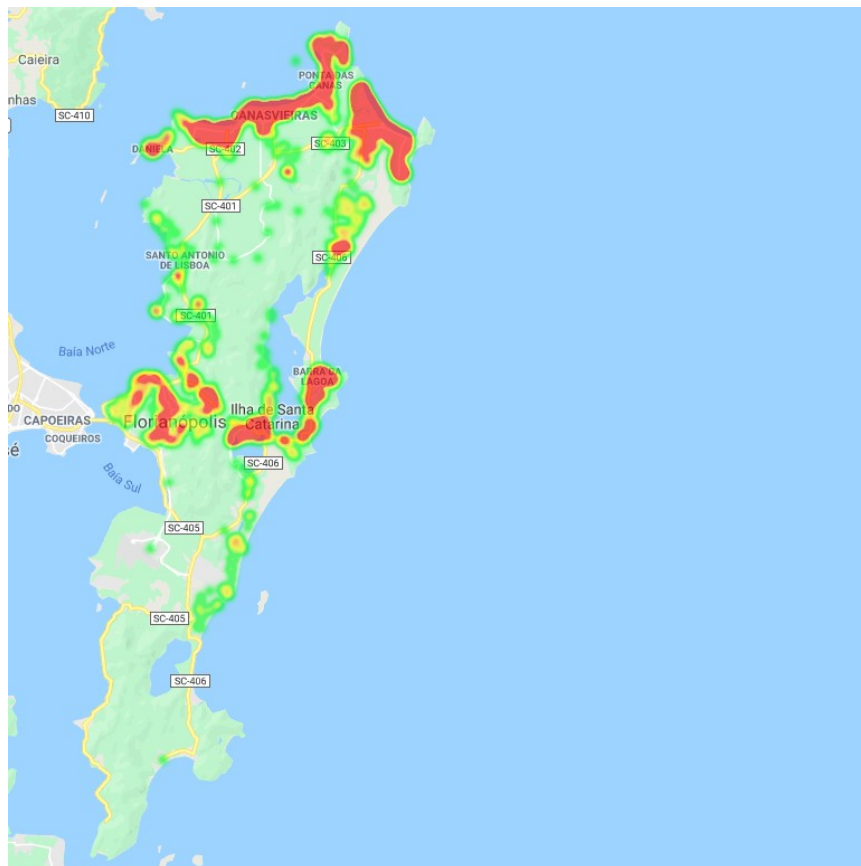
Figura 14 – Habitações Alugadas em Março



Fonte: Elaborada pelo Autor(2019)

No mês de abril a concentração de habitações ocupadas aumenta nas regiões próximas as praias de Jurerê Internacional, Daniela, Canasvieiras, Costão do Santinho, Barra da Lagoa e Lagoa da Conceição e regiões próximas como Rio Vermelho. As regiões dos bairros Itacorubi e Carvoeira também apresentam concentração significativa de imóveis ocupados. Ainda podemos ver algum acúmulo de habitações na região de Campeche. Em comparação ao mês anterior, podemos perceber que há pontos que cresceram em concentração de habitações na região de Santo Antônio de Lisboa.

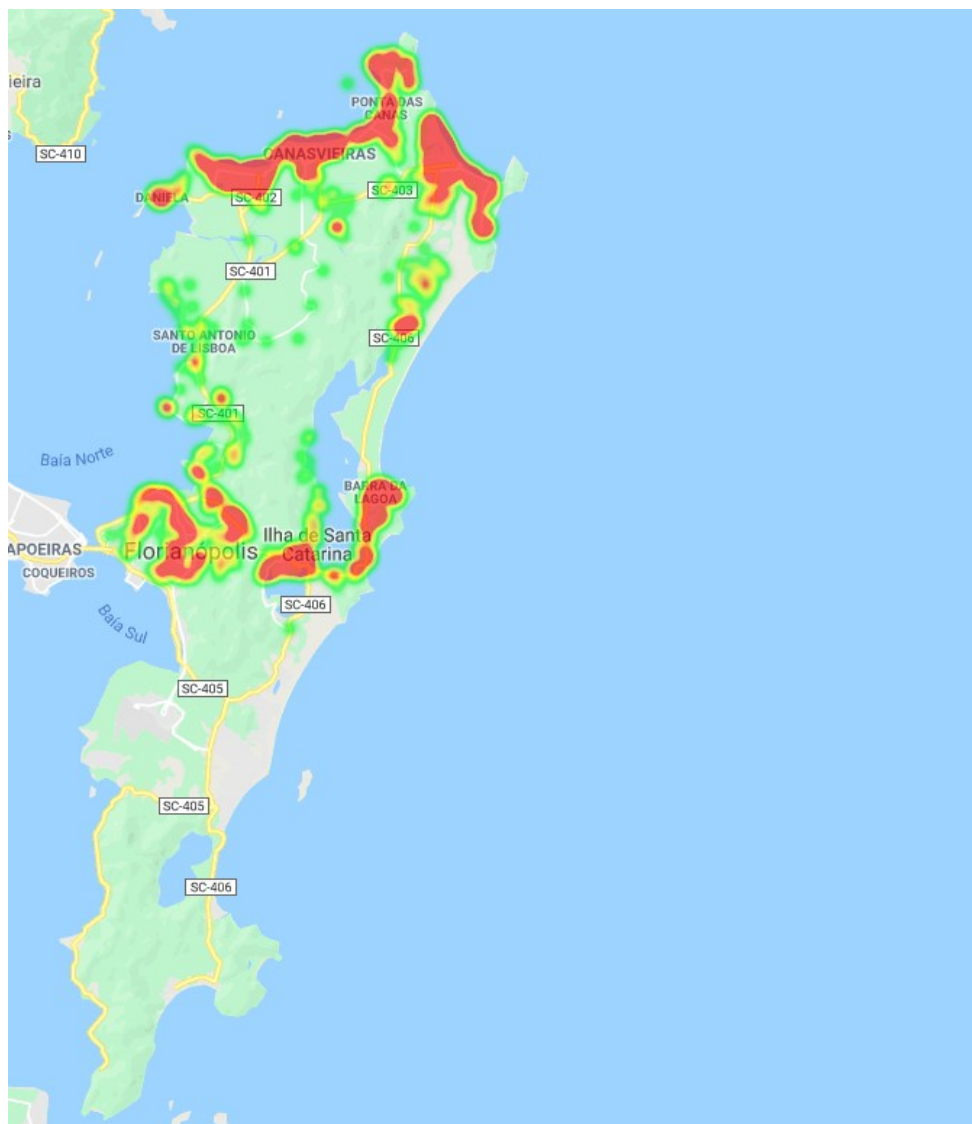
Figura 15 – Habitações Alugadas em Abril



Fonte: Elaborada pelo Autor (2019)

No mês de maio, a concentração de habitações permanece no norte da Ilha, e não ocorre no Sul, pois a região do Campeche não é apresentada no mapa a seguir. No entanto, vemos que há um acúmulo intenso de habitações alugadas nas regiões próximas a praia, apesar de Maio ser um mês com baixa temporada de turistas. Neste mês a concentração de habitações alugadas também cresce nos bairros próximos ao centro da cidade.

Figura 16 – Habitações Alugadas em Maio



Fonte: Elaborada pelo Autor (2019)

Através dos mapas e indicadores gerados no desenvolvimento deste estudo

podemos perceber que a procura pelos imóveis disponibilizados via Airbnb na Ilha de Florianópolis, fora da alta temporada, tende a crescer durante os finais de semana e feriados, bem como o número de hóspedes. Este último, sendo um número máximo estimado através do atributo ocupação máxima de cada habitação.

As tabelas com as dez habitações com maior arrecadação e maior ocupação, nos forneceram algumas características sobre a procura. As habitações com maior ocupação tendem a possuir um menor valor de diária e hospedar menos pessoas, diferente das habitações com maior arrecadação que tendem a ser lugares de alto nível e estarem em regiões de alto padrão da ilha, bem como hospedar um maior número de pessoas.

Por meio dos mapas de calor, foi possível observar onde se concentram o maior número de habitações que estão disponíveis na região pelo Airbnb, concluindo que se encontram em maioria nos bairros próximos da praia como Jurerê e Canasvieiras. Contudo também é possível observar uma grande oferta nos bairros próximos ao centro, levantando a hipótese de que há turistas buscando lazer, mas também há turistas em viagens de negócios.

Dessa forma, pode-se concluir que os indicadores para habitação, número de hóspedes e ocupação propiciaram novos conhecimentos sobre a oferta, a procura e ocupação de imóveis via plataforma de aluguel por temporada, podendo servir de apoio a tomada de decisão para as partes interessadas do setor na região.

5 CONCLUSÕES E TRABALHOS FUTUROS

O processo de revolução que levou a sociedade a trocar tecnologias analógicas por tecnologias digitais, trouxe muitos benefícios mas também muitos desafios. O grande volume e variedade de dados que circulam diariamente, trouxe a evolução de ferramentas de captura e análise de dados para criação de conhecimentos sobre negócios e sobre a própria sociedade. No setor turístico esse crescimento modificou a forma como agentes e empresas do setor se organizam, levando a atenção das partes interessadas às experiências de turistas, buscando utilizar dados digitais para analisar fluxos turísticos e propor melhorias em cidades consideradas como turísticas. Neste ambiente, diversas inovações surgiram e trouxeram mudanças nos meios de acomodação tradicionais encontrados em viagens. O Airbnb, plataforma de aluguel por temporada, propôs uma maneira simples e mais barata de hospedagem, nos mais diversos lugares do mundo, levando ao turista a experiência de se acomodar em um lugar aconchegante e familiar.

5.1 CONCLUSÃO

Este trabalho teve como intuito a criação de um ambiente de análise dos dados sobre oferta de imóveis em Florianópolis disponibilizados no Airbnb aplicando técnicas de extração de conhecimento, através das análises e do ambiente projetado na ferramenta Metabase, concluindo assim que o trabalho cumpriu com seu objetivo geral.

Os objetivos específicos também foram atingidos. Por meio da coleta de dados e da projeção do ambiente de análise foi possível inferir alguns conhecimentos sobre a oferta e a ocupação dos imóveis disponibilizados para aluguel por temporada no período de outono na Ilha com os indicadores gerados. Por meio do atributo ocupação máxima foi possível inferir uma quantidade máxima de turistas que a região pode ter recebido durante o período de março a maio de 2019, construindo gráficos de linha que nos mostraram uma maior ocupação durante períodos de feriado e finais de semana, como pode ser observado na figura 11 do

capítulo 4. Além disso, foi possível verificar o número de habitações oferecidas, bem como a demanda. Os mapas de calor nos proporcionaram uma visão de onde se localizam a maior parte das habitações oferecidas, que são regiões de praia, como Canasvieiras, Jurerê, Costão do Santinho, Barra da Lagoa e Lagoa da Conceição e regiões mais próximas ao centro da cidade.

Proporcionando um ambiente para discussão sobre o setor turístico da cidade de Florianópolis, percebemos que dados podem ser utilizados para melhorar a experiência de turistas que visitam a região durante o ano todo. Neste contexto, podemos concluir que o advento do Big Data e a Revolução dos Dados tem suma importância na transformação de dados em conhecimento.

5.2 TRABALHOS FUTUROS

Assim através do desenvolvimento deste estudo pode-se perceber a oportunidade para o desenvolvimento de alguns trabalhos futuros que aprimorem o que aqui foi descrito, como por exemplo:

- A automatização e melhoria do *script* de coleta de dados para habitações individuais, buscando construir uma base de dados maior;
- A adição de mais dados ao modelo a fim de criar outras métricas e indicadores;
- A análise e coleta de dados da alta temporada na região de Florianópolis.

REFERÊNCIAS

AIRBNB. **About Us**, 2019. <<https://news.airbnb.com/about-us/>> Acesso em: 28 out. 2019.

ALMEIDA, F. **Big data: Concept, potentialities and vulnerabilities**. Emerging Science Journal, v. 2, 03 2018.

ASKITAS, N.; ZIMMERMANN, K. **The internet as a data source for advancement in social sciences**. International Journal of Manpower, v. 36, p. 2–12, 04 2015.

AUSTIN, C.; KUSUMOTO, F. **The application of big data in medicine: current implications and future directions**. Journal of interventional cardiac electrophysiology : an international journal of arrhythmias and pacing, v. 47, 01 2016.

BOYD, D.; CRAWFORD, K. **Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon**. Information, Communication Society, v. 15, p. 662–679, 01 2012.

CASTILLO, C. **Effective web crawling**. SIGIR Forum, v. 39, p. 55–56, 06 2005.

CHRISTENSEN, CLAYTON M. Disruptive Innovation. **The Encyclopedia of Human-Computer Interaction, 2nd Ed.**, [S. l.], p. -, 15 mar. 2015. Disponível em: <https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/disruptive-innovation>. Acesso em: 3 dez. 2019.

COOLEY, R.; MOBASHER, B.; SRIVASTAVA, J. **Web mining: information and pattern discovery on the world wide web**. In: Proceedings Ninth IEEE International Conference on Tools with Artificial Intelligence. [S.l.: s.n.], 1997. p. 558–567.

EINAV, L.; LEVIN, J. **The data revolution and economic analysis**. Innovation Policy and the Economy, v. 14, 05 2013.

EL-SAPPAGH, S. H. A.; HENDAWI, A. M. A.; BASTAWISSY, A. H. E. **A proposed model for data warehouse etl processes**. Journal of King Saud University - Computer and Information Sciences, v. 23, n. 2, p. 91 – 104, 2011. ISSN 1319-1578. <<https://www.sciencedirect.com/science/article/pii/S131915781100019X>> Acesso em: 28 out. 2019.

FOINA, A. **Métodos de aquisição de dados quantitativos na internet: o uso da rede como fonte de dados empíricos**. Ciência e Trópico, v. 30, p. 283–296, 07 2003.

GOMES, P. C. T. **WHAT ARE THE MAIN DIFFERENCES BETWEEN METRICS AND INDICATORS?** Apr 2017. <<https://www.opservices.com/differences-between-metrics-and-indicators/>> Acesso em: 29 out. 2019.

GRETZEL, U. et al. **Smart tourism: foundations and developments**. Electronic Markets, v. 25, 08 2015.

GUTTENTAG, D. **Airbnb: Disruptive innovation and the rise of an informal tourism accommodation sector**. Current Issues in Tourism, p. 1–26, 12 2013.

GUTTENTAG, D. A.; SMITH, S. L. **Assessing airbnb as a disruptive innovation relative to hotels: Substitution and comparative performance expectations**. International Journal of Hospitality Management, v. 64, p. 1 – 10, 2017. ISSN 0278-4319.

GYR, U. **The history of tourism: Structures on the path to modernity**. European History Online, 12 2010.

HUNTER, W. C. et al. **Constructivist research in smart tourism**. Asia Pacific Journal of Information Systems, v. 25, p. 105–120, 03 2015.

HURWITZ, J.; NUGENT, A.; HALPER, F. **Big Data for Dummies**. [S.l.]: Wiley Brand, 2013.

HWANG, J.; PARK, H.-Y.; HUNTER, W. C. **Constructivism in smart tourism research: Seoul destination image**. Asia Pacific Journal of Information Systems, v. 25, p. 163–178, 03 2015.

JOHNSON, J. et al. **Big data: big data, digitization, and social change**. Ubiquity, v. 2017, p. 1–8, 12 2017.

KAUR, K.; KAUR, R. **Internet of things to promote tourism: An insight into smart tourism**. International Journal of Recent Trends in Engineering Research, v. 2, p. 357 – 361, 04 2016.

KIMBALL, R. et al. **The Data Warehouse Lifecycle Toolkit**. [S.l.]: Wiley, 1998.

KITCHIN, R. **The data revolution: big data, open data, data infrastructures and their consequences**. [S.l.]: SAGE, 2014.

LAENDER, A. H. F. et al. **A brief survey of web data extraction tools**. ACM SIGMOD Record Homepage archive, v. 31, p. 84–93, 06 2002.

LINS, H. **Florianópolis, cluster turístico?** Revista Turismo em Análise, v. 11, n. 2, p. 55–70, 11 2000.

LINS, H. **Interações, aprendizagem e desenvolvimento:ensaio sobre o turismo em florianópolis**. Revista Turismo em Análise, v. 9, n. 1, p. 107–121, 01 2007

MACHADO, C. C. et al. **Um web crawler para projeções e análise de vulnerabilidades de segurança e consistência estrutural de páginas web**. Revista de Empreendedorismo, Inovação e Tecnologia, v. 2, p. 3–12, 2015.

MARINHO, L; GIRARDI, R. **Mineração na Web**. Maranhão: UFMA, 2003. Disponível em: https://www.researchgate.net/publication/228369738_Mineracao_na_Web.

Acesso em: 2 dez. 2019.

MCAFEE, A.; BRYNJOLFSSON, E. **Big data: The management revolution**. *Harvard business review*, v. 90, p. 60–6, 68, 128, 10 2012.

MITCHEL, R. **Web Scraping with Python: Collecting Data from the Modern Web**. [S.l.]: O'Reilly, 2015.

MOODY, D.; KORTINK, M. **From enterprise models to dimensional models: a methodology for data warehouse and data mart design**. In: . [S.l.: s.n.], 2000. p. 5.

ONBERGER, V. M.-S.; CUKIER, K. **Big Data: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK**. [S.l.]: HMH Books, 2013.

PAL, S. K.; TALWAR, V.; MITRA, P. **Web mining in soft computing framework: relevance, state of the art and future directions**. *IEEE Transactions on Neural Networks archive*, v. 13, p. 1163–1177, 05 2002.

RIAHI, Y. **Big data and big data analytics: Concepts, types and technologies**. v. 5, p. 524–528, 11 2018.

RUSSOM, P. **Big data analytics. Transforming Data With Inteligence**, 2011.

SCHROEDER, R. **Big data: Shaping knowledge, shaping everyday life**. *MATRIZES*, v. 12, p. 135–136, 08 2018.

SIGALA, M. **Web 2.0 in the tourism industry: a new tourism generation and new e-business models**. 01 2007.

SRIVASTAVA, J. et al. **Web usage mining: Discovery and applications of usage patterns from web data**. SIGKDD Explorations, v. 1, n. 2, p. 12–23, 2000.

VARGIU, E.; URRU, M. **Exploiting web scraping in a collaborative filtering-based approach to web advertising**. Artificial Intelligence Research, v. 2, 01 2013.

WALTON, J. K. **Tourism**. Encyclopædia Britannica, inc., 11 2018. <<https://www.britannica.com/topic/tourism>>. Acesso em: 28 out. 2019.

ČIŽIK, T. **The information revolution and the european security architecture**. 07 2018.

SANTUR, **Estatísticas e Indicadores Turísticos: Pesquisa de Demanda Turística - Alta Temporada Anual - 2014-2017**. [S. l.], 2018. Disponível em: <http://turismo.sc.gov.br/institucional/index.php/pt-br/informacoes/estatisticas-e-indicadores-turisticos/category/51-pesquisa-de-demanda-turistica-alta-temporada-anual-2014-2017>. Acesso em: 29 out. 2019.

METABASE: is the easy, open source way for everyone in your company to ask questions and learn from data.. Versão 0.33.4. [S. l.], 19 nov. 2007. Disponível em: <https://www.metabase.com/>. Acesso em: 14 out. 2019.

KETTLE: Data Integration. Versão 8.2. [S. l.], 2004. Disponível em: <https://community.hitachivantara.com/s/article/data-integration-kettle>. Acesso em: 9 set. 2019.

INMON, William H. **Building the Data Warehouse**. 3. ed. [S. l.]: Wiley, 2002. 412 p.

SIGNIFICADOS. *In: O que é Streaming*. [S. l.], 20 out. 2017. Disponível em: <https://www.significados.com.br/streaming/>. Acesso em: 31 out. 2019.

SLEE, Tom. **Airbnb web site scraper**. [S. l.], 02 2018. Disponível em:

<https://github.com/tomslee/airbnb-data-collection/blob/master/README.md>. Acesso em: 31 out. 2019.

SLEE, Tom. **Airbnb Data Collection: Get the Data**. [S. l.], 7 ago. 2017. Disponível em: <http://tomslee.net/airbnb-data-collection-get-the-data>. Acesso em: 31 out. 2019.

SCHOENHERR, Steven E. **The Digital Revolution**. 10. mai. 2004. Disponível em: <https://web.archive.org/web/20070103161029/http://history.sandiego.edu/gen/recording/digitalrev.html>. Acesso em: 3 dez. 2019.

TAVOLARI, Bianca. **AirBnB e os impasses regulatórios para o compartilhamento de moradia: notas para uma agenda de pesquisa em direito**. Nov. 2017. Disponível em: https://www.researchgate.net/publication/321156274_AirBnB_e_os_impasses_regulatorios_para_o_compartilhamento_de_moradia_notas_para_uma_agenda_de_pesquisa_em_direito. Acesso em: 6 dez. 2019.

APÊNDICE A – Consulta: Oferta Máxima de Habitações Por Mês

```
select mes, titulo_mes, max(cnt) as ofertaMaxima
from (select td.mes, td.titulo_mes, count(distinct oc.sk_habitacao_dim) as cnt
      from ocupacao_fato oc
      left join tempo_dim td
      on oc.sk_tempo_dim = td.sk_tempo_dim
      left join habitacao_dim hd
      on oc.sk_habitacao_dim = hd.sk_habitacao_dim
      where cidade = 'Florianopolis' and dia_mes > '2019-02-28' and dia_mes
         < '2019-06-01'
      group by mes, dia_mes, titulo_mes
) p
group by p.mes, p.titulo_mes
```


APÊNDICE B – Consulta: Demanda de Habitações Por Dia

```
select td.dia_mes, td.mes, count(distinct of.sk_habitacao_dim)
from ocupacao_fato of
left join tempo_dim td
on of.sk_tempo_dim = td.sk_tempo_dim
left join habitacao_dim hd
on of.sk_habitacao_dim = hd.sk_habitacao_dim
where cidade = 'Florianopolis' and alugada = 'alugada' and dia_mes > '2019-02-28'
and dia_mes < '2019-06-01'
group by dia_mes, mês
```

APÊNDICE C – Consulta: Demanda de Habitações por Mês

```
select mes, titulo_mes, sum(cnt) as habitacoesOcupadasPorMes
from (select td.mes, td.titulo_mes, count(distinct oc.sk_habitacao_dim) as cnt
      from ocupacao_fato oc
      left join tempo_dim td
      on oc.sk_tempo_dim = td.sk_tempo_dim
      left join habitacao_dim hd
      on oc.sk_habitacao_dim = hd.sk_habitacao_dim
      where cidade = 'Florianopolis' and alugada = 'alugada' and dia_mes > '2019-02-
28' and dia_mes < '2019-06-01'
      group by mes, titulo_mes
      ) p
group by p.mes, p.titulo_mes
```

APÊNDICE D – Consulta: Média Mensal de Demanda de Habitações

```
select x.mes, avg(x.countHabitacoesDia), x.titulo_mes
from ( select td.mes, td.titulo_mes, td.dia_mes, count(distinct of.sk_habitacao_dim)
as countHabitacoesDia
  from ocupacao_fato of
  left join tempo_dim td
  on of.sk_tempo_dim = td.sk_tempo_dim
  left join habitacao_dim hd
  on of.sk_habitacao_dim = hd.sk_habitacao_dim
  where cidade = 'Florianopolis' and alugada = 'alugada' and dia_mes > '2019-02-28'
and dia_mes < '2019-06-01'
group by dia_mes, td.mes, td.titulo_mes
) x
group by x.mes, x.titulo_mes
```

APÊNDICE E – Consulta: Média de Dias Ocupados por Mês

```
select  td.mes,      td.titulo_mes,      count(distinct  oc.sk_habitacao_dim)  as
habitacoesAlugadas,      count(alugada)      as      diasAlugados,
count(alugada)::decimal/count(distinct  oc.sk_habitacao_dim)::decimal  as
mediaDiasOcupados
from ocupacao_fato oc
left join tempo_dim td
on oc.sk_tempo_dim = td.sk_tempo_dim
left join habitacao_dim hd
on oc.sk_habitacao_dim = hd.sk_habitacao_dim
where cidade = 'Florianopolis' and dia_mes > '2019-02-28' and dia_mes < '2019-06-
01' and alugada='alugada'
group by td.mes, td.titulo_mes
```

APÊNDICE F – Consulta: Máximo de Hóspedes em Habitações por Mês

```
select mes, titulo_mes, max(ocupantesDia)
from (
  select t.mes, t.titulo_mes, t.dia_mes, sum(ocupacao_maxima) as ocupantesDia
  from agregado_mes_fato a
  left join ocupacao_fato o
  on a.sk_habitacao_dim = o.sk_habitacao_dim
  left join tempo_dim t
  on o.sk_tempo_dim = t.sk_tempo_dim
  left join habitacao_dim h
  on a.sk_habitacao_dim = h.sk_habitacao_dim
  where cidade = 'Florianopolis' and alugada = 'alugada' and dia_mes > '2019-02-28'
  and dia_mes < '2019-06-01'
  group by t.dia_mes, t.mes, t.titulo_mes
) x
group by x.mes, x.titulo_mes
```

APÊNDICE G – Consulta: Máximo de Hóspedes em Habitações por Dia

```
select t.mes, t.dia_mes, sum(ocupacao_maxima) as ocupantesDia
from agregado_mes_fato a
left join ocupacao_fato o
on a.sk_habitacao_dim = o.sk_habitacao_dim
left join tempo_dim t
on o.sk_tempo_dim = t.sk_tempo_dim
left join habitacao_dim h
on a.sk_habitacao_dim = h.sk_habitacao_dim
where cidade = 'Florianopolis' and alugada = 'alugada' and dia_mes > '2019-02-28'
and dia_mes < '2019-06-01'
group by t.mes, t.dia_mes
```

APÊNDICE H – Script do Mapa de Calor

```
import pandas as pd
import gmplot
from IPython.display import display

latitudes = data["latitude"]
longitudes = data["longitude"]

gmap = gmplot.GoogleMapPlotter(-27.354800, -48.325700, 10)

gmap.heatmap(latitudes, longitudes)

gmap.draw("habitacoesDisponiveisEm3Meses.html")

hab = pd.read_csv("alugadasmes3.csv")

display(hab.head(n=5))
latitudes2 = hab["latitude"]
longitudes2 = hab["longitude"]

gmap = gmplot.GoogleMapPlotter(-27.354800, -48.325700, 10)

gmap.heatmap(latitudes2, longitudes2)

gmap.draw("habitacoesAlugadasEmMarco.html")

hab = pd.read_csv("alugadasmes4.csv")

display(hab.head(n=5))

latitudes2 = hab["latitude"]
longitudes2 = hab["longitude"]

gmap = gmplot.GoogleMapPlotter(-27.354800, -48.325700, 10)
```

```
gmap.heatmap(latitudes2, longitudes2)

gmap.draw("habitacoesAlugadasEmAbril.html")

hab = pd.read_csv("alugadasmes5.csv")

display(hab.head(n=5))

latitudes2 = hab["latitude"]
longitudes2 = hab["longitude"]

gmap = gmapplot.GoogleMapPlotter(-27.354800, -48.325700, 10)

gmap.heatmap(latitudes2, longitudes2)

gmap.draw("habitacoesAlugadasEmMaio.html")
```


APÊNDICE I – Consulta: Top 10 Habitações mais Ocupadas

```
select hd.*, total_dias_ocupados, ag.sk_habitacao_dim, mes
from agregado_mes_fato ag
left join habitacao_dim hd
on ag.sk_habitacao_dim = hd.sk_habitacao_dim
where cidade = 'Florianopolis' and mes = 3
group by ag.sk_habitacao_dim, mes, total_dias_ocupados, hd.sk_habitacao_dim
order by total_dias_ocupados desc
fetch first 10 rows only
```

APÊNDICE J – Consulta: Top 10 Habitações com Maior Arrecadação

```
select h.*, count(alugada), h.valoraluguel, count(alugada) * h.valoraluguel as
total_arrecadado
from ocupacao_fato oc
left join habitacao_dim h
on oc.sk_habitacao_dim = h.sk_habitacao_dim
left join tempo_dim t
on oc.sk_tempo_dim = t.sk_tempo_dim
where h.cidade = 'Florianopolis' and t.mes = 5 and alugada = 'alugada'
group by oc.sk_habitacao_dim, h.sk_habitacao_dim
order by total_arrecadado desc
fetch first 10 rows only
```

APÊNDICE K – Artigo

O uso de Big Data para análise de imóveis via Airbnb em destinos turísticos

Cinthya Weigel¹

¹Departamento de Informática e Estatística – Universidade Federal de Santa Catarina (UFSC)

weigelcinthya@gmail.com

Abstract. *This work uses content mining techniques through web crawlers to collect data on housing available on the season rental platform, in order to apply knowledge engineering and data warehouse procedures for indicator creation through a dimensional model. , informative graphics and heat maps, creating an environment to support decision making in the tourism sector of Florianópolis Island. The created environment provides the answers of the indicators on the occupation and occupation, allowing to observe where the dwellings are concentrated and with little variation in the demand of entries from March to May 2019, being April or month with the highest demand.*

Resumo. *Este trabalho utiliza técnicas de mineração de conteúdo através de Web Crawlers para a coleta de dados sobre habitações disponíveis em plataforma de aluguel por temporada, com o intuito de aplicar procedimentos de engenharia do conhecimento e Data Warehouse para criação de indicadores, por meio de um modelo dimensional, gráficos informativos e mapas de calor, criando um ambiente para apoio a tomada de decisão ao setor turístico da Ilha de Florianópolis. O ambiente criado proporcionou a visualização de indicadores sobre habitação e ocupação, permitindo observar onde estão concentradas as habitações e a pouca variação de procura entres os meses de março a maio de 2019, sendo abril o mês com a maior demanda.*

1. Introdução

A produção de dados tem aumentado constantemente e isso deve a grande abundância de dispositivos digitais presentes em casas, ambientes corporativos e espaços públicos, das mídias sociais e da internet das coisas. (Kitchin, 2014) O produto do uso desses dispositivos são os dados digitais, que contemplam informações como roteiros de viagens, trabalho, comunicação, consumo e interesses de seus usuários. (Riahi, 2018) Deste modo a forma como os dados são armazenados, distribuídos e interpretados, produzindo o que chamamos hoje de Big Data. (Kitchin, 2014)

Big Data pode ser caracterizado por grandes bases de dados com três principais atributos: volume, variedade e velocidade. (Riahi, 2018) Por meio de dados obtidos em páginas web e armazenados em bases de dados Big Data por criar novos conhecimentos para auxiliar a tomada de decisão. (McAfee; Brynjolfsson, 2012) Para o setor turístico este fenômeno modificou a forma tradicional de viajar, desde a busca por roteiros em cidades turísticas até a busca por acomodações.

Neste cenário, aplicações como o Airbnb começaram a surgir com o intuito de facilitar a busca por hospedagens. O aplicativo se apresenta de forma simples e intuitiva, buscando diminuir a distância entre hóspede e anfitrião. (Gutteng, 2013; Gutteng; Smith, 2017) No ambiente econômico, o ambiente proposto pela plataforma reduziria o desequilíbrio entre oferta e procura pois possui uma grande quantidade de imóveis disponíveis em diversas cidades, explicando seu sucesso através da conexão entre informação e pessoas, preços abaixo do mercado e a cultura de recomendações. (Tavolari, 2017)

No contexto apresentado, o objetivo do presente estudo é explorar técnicas de mineração de dados oriundos de páginas web, data warehouse e engenharia do conhecimento para agregar novas descobertas que auxiliem no processo de tomada de decisão do setor turístico da região de Florianópolis. Considerando desde a coleta de dados sobre os imóveis disponíveis para locação até a implementação de um esquema de visualização das informações, que possa contribuir para o acompanhamento do movimento turístico na região. (Moody; Kortink, 2000).

2. Fundamentação Teórica

2.1. Revolução dos Dados

A revolução dos dados teve início entre os anos de 1950 e 1970, com o advento da criação de computadores digitais e outras tecnologias, e desde então vêm gerando grande impacto na sociedade em que vivemos. (Askitas; Zimmermann, 2015). Com o número expressivo de usuários em mídias sociais e de tecnologias onipresentes sempre conectados a Internet o número de dados gerados aumenta a cada instante. (Johnson et al., 2017).

O crescimento de tecnologias digitais, computadores e smartphones fez com que boa parte do nosso comportamento na Internet passasse a ser armazenado, desde compras online, pesquisas em buscadores virtuais, comentários e vídeos que assistimos. Contudo dados armazenados hoje se diferem dos armazenados a 20 ou 30 anos atrás pela disponibilidade em tempo real, a maior quantidade e a maior variedade. (Einav; Levin, 2013). Assim, podemos atribuir o conceito de revolução digital ao processo que levou a sociedade a substituir tecnologias analógicas por tecnologias digitais, promovendo mudanças importantes em nosso âmbito social, político e econômico. (Schorenherr, 2004).

Grandes quantidades de dados estão sendo coletadas por meio de inúmeras fontes: dispositivos móveis, sensores, página web, entre outros. Estes dados são chamados de dados digitais e para muitos são considerados matéria prima para abstração de problemas em diversas áreas, apresentando grande utilidade quando aplicados para descoberta de conhecimento. (Riahi, 2018; Kitchin, 2014). Definimos a Revolução dos Dados como um produto da revolução digital, pois através do desenvolvimento tecnológico vêm tornando os dados cada vez mais abertos, acessíveis e suportados por grandes infraestruturas. Além disso, reformulando o desenvolvimento de conhecimento, levantando questões importantes como privacidade e segurança. (Kitchin, 2014).

Desta maneira, sendo vista como uma alteração no escopo de conhecimento sobre determinados assuntos e modificando a maneira de construirmos conhecimento e de nos relacionarmos com a informação, o resultado dessa revolução se apresenta na forma de um “boom” de aplicações construídas através de dados digitais e da automação de tarefas diárias tornando nosso dia a dia mais produtivo e eficiente. (Boyd; Crawford, 2012; Johnson et al. 2017)

2.2 Big Data

Desde sempre, empresas de diversos tamanhos e de todos os tipos têm se beneficiado com o gerenciamento e análise de dados. Entretanto, este tipo de processo ficou mais difícil pois, dados podem ser obtidos de diversas maneiras, podendo ser estruturados ou não, vindos de bancos de dados tradicionais ou da própria web. (Hurwitz; Nugent; Halper, 2013).

Big Data pode ser definido como a capacidade de gerenciar volumes imensos de dados, variáveis, com velocidade e no tempo certo, para que possibilite uma análise ágil, permitindo uma resposta também rápida (Hurwitz; Nugent; Halper, 2013). É um objeto complexo que se apresenta como uma solução projetada para fornecer acesso, em tempo real, à banco de dados gigantes, contemplando também uma categoria de tecnologias e técnicas. Volume, variedade e velocidade são suas principais características (Riahi, 2018).

As tecnologias *Big Data* realizam conexões entre pontos de dados distintos, possibilitando a identificação de padrões que antes poderiam passar despercebidos. Sem responder o "por quê" ou "como", essas ligações podem responder perguntas que ainda nem foram pensadas. O principal objetivo do Big Data é destacar o uso de grandes *datasets* para descoberta de conhecimento, auxiliando o processo de tomada de decisão e tornando-o mais eficiente (Austin; Kusumoto, 2016). Para as grandes organizações, o Big Data se torna um elemento crucial, pois permite a coleta, armazenamento e manipulação de grande quantidade de dados, com rapidez e no tempo certo para obtenção das melhores vantagens (Hurwitz; Nugent; Halper, 2013).

2.3 Mineração de Dados da Web

O aumento da utilização de dispositivos conectados à Internet fez com que a busca por ferramentas automáticas que capturem dados relevantes da Web se tornasse indispensável. A mineração de dados da Internet, tem estado em destaque pois reúne em seu meio dados simples e complexos, além de possuir um volume imenso de informações. Deste modo, podemos definir a mineração de dados como a descoberta e análise de informações úteis disponíveis pela Internet, utilizando técnicas como por exemplo, a Mineração de Conteúdo (Cooley; Mobasher; Srivastava, 1997), composta por documentos e ferramentas presentes na taxonomia da mineração da web (Marinho; Girardi, 2003).

A taxonomia da mineração da web à classifica em mineração de conteúdo e mineração de uso. A mineração de conteúdo por sua vez é dividida em duas abordagens: baseada em agentes e baseada em bancos de dados. A primeira é dividida em: agentes inteligentes; Filtragem/Categorização; E agentes personalizados. A segunda tem como intuito organizar os dados semi estruturados obtidos da web em conjuntos mais estruturados fazendo o uso de formatos padrão de queries em bancos de dados e *data mining* para realizar análises (Cooley; Mobasher; Srivastava, 1997).

Para coleta de informações geralmente são utilizados programas especializados chamados de Web Crawlers, que identificam os dados de interesse e os mapeiam para um

formato possível de analisar. Estes, buscam simular a interação humana com o site alvo da pesquisa e estão relacionados à indexação da Web, que é uma forma de indexar informações presentes em documentos HTML na Internet através de robôs, que tem por objetivo transformar os dados não estruturados coletados e salvá-los em bases de dados estruturadas (Laender et al, 2002; Vargiu; Urru, 2013).

A extração de dados, é realizada através do desenvolvimento de um programa automatizado que conectado a um servidor web solicita dados para extração das informações necessárias (Mitchel, 2015). Segundo (Castillo, 2005), a arquitetura utilizada para criação de um *web crawler* geralmente se apresenta em forma de cascata, na qual um *crawler* cria uma coleção de dados que serão indexados e pesquisados.

Hoje, páginas web são apresentadas aos usuários nos mais diversos formatos, assim, sua indexação é iniciada através de uma visão lógica de seu conteúdo. E segundo (Castillo, 2005) o método mais utilizado para esta fase é chamado de “*bag of words*”. Contudo, apesar de tornar o processo de obtenção de dados através de páginas web mais simples e ágil, os *Web Crawlers* ainda possuem alguns desafios. O grande volume de informações, e a maneira rápida como páginas mudam e são removidas, tornam este processo mais complexo (Castillo, 2005).

2.4 Turismo Inteligente

A evolução dos meios de comunicação trouxe grande impacto à indústria turística. Segundo (Hunter et al., 2015), a ligação entre as tecnologias da informação e o turismo fez com que a troca de informações utilizando dispositivos móveis conectados a internet, não fosse manipulada apenas por empresas agentes de viagem, tornando-se pública, por meio de comentários em redes sociais e *posts* com críticas pessoais acerca do destino alvo do turismo.

Para (Gretzel et al., 2015), no âmbito do turismo, inteligente é a palavra utilizada para descrever o processo de conexão e troca de informações de diversas fontes, oriundas de tecnologias Big Data, sensores e *smartphones* que, quando aplicadas em conjunto são capazes de realizar a otimização de recursos, o desenvolvimento sustentável e a comodidade. Com a grande explosão do uso de tecnologias da Web 2.0, grandes quantidades de informação circulam diariamente e quando unidas às experiências turísticas em diversas cidades do mundo, podem gerar grandes propostas de valor. Ao evento social e econômico surgido através do encontro de tecnologias da informação e de experiências turísticas, damos o nome de Turismo Inteligente (Hunter et al., 2015).

A tecnologia da informação tem um papel essencial para a experiência turística atual, criando, desenvolvendo e aplicando sistemas técnicos de suporte ao progresso de negócios relacionados ao turismo por meio da internet, validando como este processo afeta a relação consumidor x comercial (Hwang; Park; Hunter, 2015). Com a tecnologia como principal chave para seu desenvolvimento, o turismo inteligente trouxe uma mudança de paradigma, levando o turismo tradicional para o futuro e tem sido adotada em diversos locais no mundo todo. A adesão, se bem estruturada, pode trazer inúmeros benefícios às partes interessadas no negócios e também para os turistas, que usufruirão de melhores experiências de viagem (Kaur; Kaur, 2016).

2.4.1 Turismo em Florianópolis

Para (Lins, 2000) o turismo em Florianópolis despontou há algumas décadas atrás, após a descoberta do litoral pelos Argentinos. Deste ponto em diante, o município cresceu em termos de fluxo turístico e de serviços disponibilizados para o turista, como por exemplo, os hotéis. Como apresentado na pesquisa de demanda turística, disponibilizada no página do setor em Santa Catarina, em 2017 estimou-se o fluxo de 7.471.779 turistas nacionais e 751.921 turistas internacionais. (Santur, 2018)

No entanto, o fluxo turístico observado na região é sazonal e concentrado no verão, devido as inúmeras praias. Dessa forma, podemos compreender a grande demanda do setor turístico, principalmente do setor hoteleiro para atender uma grande quantidade de pessoas em um curto período de tempo e a sua subutilização no restante do ano (Lins, 2000).

Neste cenário, está claro que tal movimento turístico não despontaria caso a região não tivesse grande potencial natural, paisagístico, histórico e cultural. Portanto, é necessário que haja manutenção dessas riquezas para que Florianópolis continue a se desenvolver no setor turístico (Lins, 2007).

2.4.2 Airbnb: Aluguel por Temporada

A plataforma *Airbnb* foi fundada em 2008, descrevendo sua missão como “criar um mundo onde qualquer pessoa possa se sentir em casa em viagens autênticas, diversificadas, inclusivas e sustentáveis.” No site, pessoas de todo o mundo podem alugar mais de 7 bilhões de acomodações, entre casas inteiras e quartos, únicos listados em mais de 100,000 cidades em 191 países e regiões (Airbnb, 2019).

Desde seu lançamento, o modelo de negócio inovador e baseado na internet, trouxe grandes mudanças para o modelo tradicional de acomodação turística. Em seu domínio de mercado, é a empresa que possui mais destaque e faz parte do grupo de companhias que permitem à pessoas comuns oferecerem suas casas ou apartamentos como acomodação, ou seja, representa um produto moderno que mudou a percepção de hospitalidade para a indústria (Guttentag, 2013; Guttentag; Smith, 2017).

A ascensão do *Airbnb* é imensa e transformou a forma como turistas reservam suas acomodações, para uma plataforma de reservas em larga escala. Por meio da internet, o *Airbnb*, conseguiu ultrapassar obstáculos que, antes, dificultavam aos turistas a escolha da acomodação para sua viagem. Hoje, a tecnologia da Web 2.0 facilita o trabalho dos anfitriões em promover sua casa, apartamento ou quarto, atingindo um número muito maior de pessoas. Além disso, usuários também podem descrever suas experiências, concedendo comentários e notas aos lugares que ficaram, diminuindo a distância entre hóspede e anfitrião (Guttentag, 2013).

3. Procedimentos metodológicos

3.1 Definição da Localização

Situada no litoral de Santa Catarina e também capital do estado, Florianópolis possui cerca de 42 praias e tem seu maior movimento turístico concentrados nos meses de verão, entre dezembro e março. A falta de informação sobre os imóveis disponíveis para aluguel por temporada no local, dificultando a pesquisa e o desenvolvimento de ambientes de tomada de decisão que auxiliem as partes interessadas, levou a escolha da cidade para o desenvolvimento do estudo.

3.2 Coleta de Dados

Para obtenção dos dados sobre as habitações da região de Florianópolis foi utilizado o Web Crawler disponibilizado no GitHub por Tom Slee. Este Web Crawler é desenvolvido em python, com integrações de bibliotecas como Requests e BeautifulSoup. Os dados coletados via script, utilizaram apenas o site do Airbnb como fonte e armazenou os dados em banco de dados PostgreSQL, contemplando os seguintes atributos dos imóveis: *room_id*; *Host_id*; *Room_type*; *Borough*; *Neighborhood*; *Reviews*; *Overall_satisfaction*; *Accommodates*; *Bedrooms*; *Price*; *Minstay*; *Latitude*; E *longitude* (Slee, 2017).

As informações sobre a disponibilidade de cada imóvel foram obtidas através do desenvolvimento de um segundo Web Crawler que utilizou as informações de id, do primeiro script. Este programa foi escrito também na linguagem python, em conjunto com a biblioteca requests e salvou os dados em um arquivo JSON. Desta forma foram adicionadas os dados de dia e disponibilidade de cada imóvel, representando o período de 1 de março de 2019 até 31 de maio de 2019.

3.3 Organização dos Dados

Ao final da coleta de dados, os arquivos JSON com os dados sobre disponibilidade foram transformados em arquivos CSV, com o auxílio da ferramenta SQLfy, para facilitar o trabalho de limpeza e carga de dados no modelo dimensional. Finalizado este processo os dados foram limpos e organizados, tratando valores nulos e normalizando atributos como *availability (true/false)* e *room_type (entire home/apt, shared room, private room)*. As colunas escolhidas para integrarem o projeto e serem adicionadas ao modelo foram: *dia_semana*, *dia_mes*, *mes*, *availability*, *num_quartos*, *num_banheiros*, *num_hospedes*, *latitude*, *longitude*, *id_habitacao*, *valor_aluguel*, *descricao*, *habitacao_integral*.

3.4 Modelagem Dimensional

Para a etapa de modelagem dimensional foi desenvolvido um esquema estrela que se apresenta na figura abaixo. Um esquema estrela consiste em inúmeras tabelas pequenas, chamadas de dimensões e uma ou mais tabelas maiores, chamadas de fato. As tabelas fato, contém todos os tipos de medidas relacionadas ao negócio, representando relações “muitos-para-muitos” e contendo uma ou mais chaves estrangeira que as ligam às suas tabelas dimensão. Por sua vez as tabelas dimensão, são definidas por suas chaves primárias, que serão referenciadas em suas respectivas tabelas fato, contendo as informações necessárias para criação de medidas e *group by* nas *queries* do *Data Mart* (Moody; Kortink, 2000; Kimball et al., 1998).

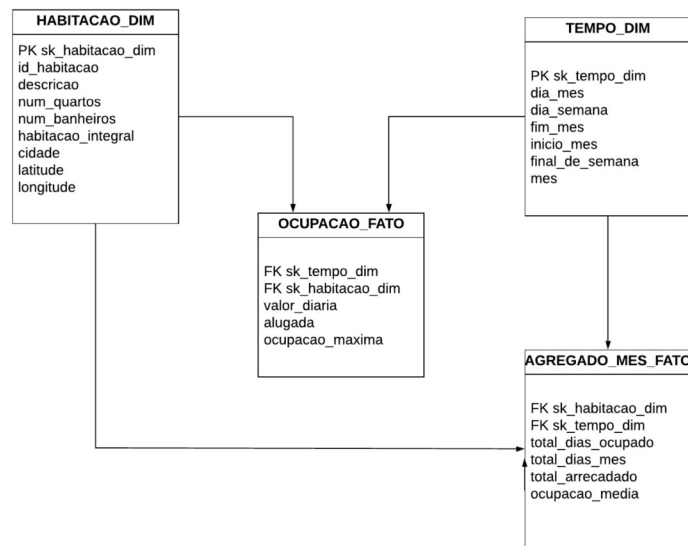


Figura 1. Modelo Dimensional

Com essas definições, a construção do modelo dimensional utilizado neste estudo, agrupou as variáveis de descrição da habitação, na tabela dimensão Habitação e os dados de descrição de tempo na tabela dimensão Tempo. Os dados utilizados para criação de medidas no modelo foram enviados para a tabela fato, Ocupação, que contém as chaves estrangeiras ligando-a às tabelas Habitação e Tempo. Para criação de medidas mais específicas a tabela fato Agregado Mês, que agrupa os dias totais ocupados por cada habitação, bem como sua ocupação média.

3.5 Carga de Dados

Nesta etapa foi realizado o processo de ETL para o modelo dimensional criado na etapa anterior, utilizando a ferramenta Data Integration, Kettle da suite Pentaho. Para o processo de ETL, foram realizadas quatro cargas, para preencher as quatro tabelas presentes no modelo, dimensão Habitação, dimensão Tempo, fato Ocupação e fato Agregado Mês. Ao final de cada um dos processos de carga, os dados foram carregados em um banco de dados *postgresql* que corresponde ao modelo criado e apresentado na seção anterior, figura 1.

3.6 Frontend

Para diminuir a complexidade e tornar a informação do *Data Warehouse* o mais acessível possível, é necessário que se crie uma camada entre os dados e usuários. A camada de *frontend* pode ser construída através de serviços específicos, criados para este fim. Estes serviços abstraem grande parte das atividades, inicialmente necessárias para o desenvolvimento do DW, como por exemplo: gerenciamento de buscas e *queries*, acesso e segurança e relatórios padrão (Kimball et al, 1998)

Neste sentido a visualização dos dados foi feita através do Metabase, uma aplicação open source com a capacidade de conectar-se ao banco de dados *postgresql*, onde estão armazenados os dados do modelo. Após a conexão foi possível criar os gráficos de oferta e demanda de habitações, máximo de ocupantes por dia e por mês e os top 10 de habitações que mais estiveram ocupadas e as que mais arrecadaram. Para o desenvolvimento dos mapas de calor, que utilizam as latitudes e longitudes presentes no modelo, foi desenvolvido um

script em python que utilizou as bibliotecas pandas e gmpplot para renderizar as informações em um mapa.

4. Resultados e Análises

Os dados coletados acerca da ilha de Florianópolis forneceram os indicadores para habitação, oferta, demanda e ocupação nos meses de março, abril e maio, fora da alta temporada turística na região, que acontece nos meses de verão, e estão apresentados nos gráficos a seguir.

O primeiro gráfico construído, com o auxílio da ferramenta de consulta Metabase, foi o de oferta máxima de habitações na região. Nele podemos observar que o mês onde há maior oferta de habitações é Março com 7.133. Seguido por Abril com 7.131 e Maio com 6.731. O número de habitações ofertadas corresponde ao final de temporada na Ilha de Florianópolis, diminuindo o número conforme a passagem dos meses.

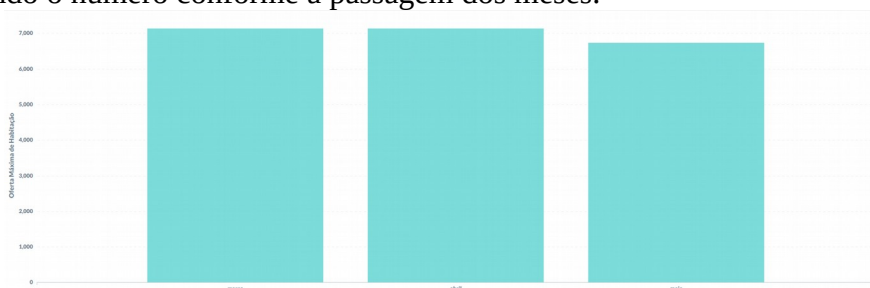


Figura 2. Oferta de Habitações por mês

Estabelecida a oferta de Habitações na região durante o período de três meses, também foi estabelecida a demanda, ou seja quantas dessas habitações disponíveis foram ocupadas. Na imagem podemos observar que o mês de março possui valores menores, comparando-o com os meses seguintes. Isso se deve à um problema técnico, onde o *web crawler* foi bloqueado pelo site, o que impossibilitou a obtenção de dados de algumas das habitações do mês de março. Nos meses de abril e maio, podemos perceber que o gráfico segue um padrão de picos e vales, com as maiores contagens presentes nos finais de semana. Neste gráfico, também podemos perceber que a maior contagem encontrada foi no mês de abril, durante o final de semana do feriado de Páscoa.

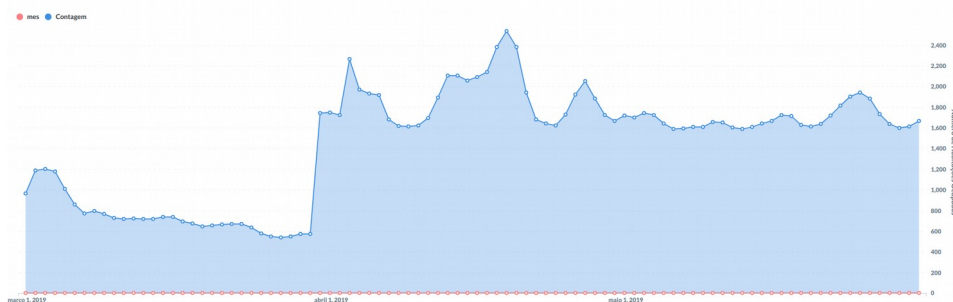


Figura 3. Demanda Diária de Habitações

O gráfico seguinte apresenta a contagem máxima de procura por habitações por mês. Considerando o número ofertado de habitações como 100%, em cada mês, como apresentadas no gráfico da figura 6, no mês de março temos 2,605 habitações ocupadas das 7.133 oferecidas, ou seja 36,5% das habitações foram ocupadas neste período. Já no mês seguinte o

número de habitações ocupadas passou a ser 3,712 isto é 52,05% das 7.131 habitações oferecidas no mês de abril. Seguindo para o mês de maio temos 2,562 habitações oferecidas, sendo 39,39% das ofertadas.

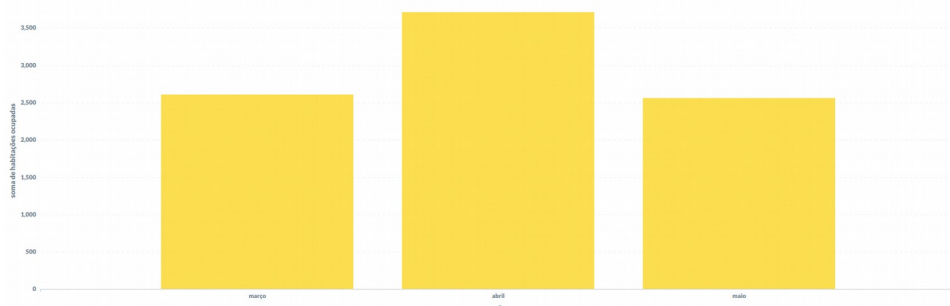


Figura 4. Demanda de Habitações Mensal

O gráfico a seguir apresenta o número máximo de ocupantes que estiveram na região de Florianópolis agrupando-os mensalmente. Este número é obtido através de uma sub consulta que soma o atributo ocupação máxima da tabela fato Ocupação, adicionando qualquer habitação que foi alugada por pelo menos um dia, realizando um agrupamento por dia e mês. Comparando aos gráficos de oferta mensal de habitação, este gráfico segue um padrão visualmente semelhante. O mês de março apresentou o resultado de 28,481 ocupantes máximos em habitações na ilha. O mês de abril apresentou o maior resultado, considerando o número máximo de ocupantes em cada habitação com 39.419 e o mês de maio apresentou o resultado de 32.766.

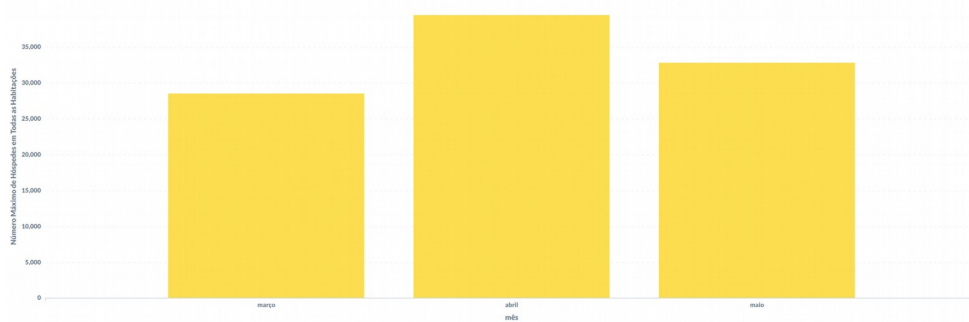


Figura 5. Máximo de Hóspedes em Habitações por mês

O seguinte gráfico apresenta o número máximo de ocupantes agrupados por dia em todas as habitações marcadas como alugadas. Para esta consulta, através da função *sum*, foram somados os ocupantes máximos de cada habitação, agrupando-os por dia. Os dados foram selecionados da tabela fato Ocupação com os dados adicionais de dia selecionados através de um join com a tabela dimensão Tempo. O gráfico se torna bem semelhante com o gráfico da figura 7, que por sua vez conta o total de habitações ocupadas por dia durante o período. Dessa forma o maior pico se encontra no dia 19 de abril, final de semana de páscoa e possuindo os maiores números entre os finais de semana de cada mês.

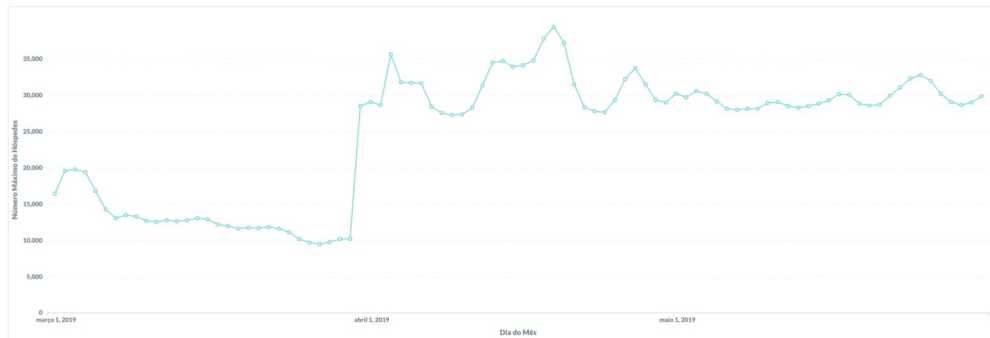


Figura 6. Máximo de Hóspedes em Habitações por dia

Considerando o atributo número de dias ocupados da tabela fato Agregado Mês foram levantadas as 10 habitações que mais estiveram ocupadas durante cada mês, contando o número de dias total em que estiveram indisponíveis. No mês de março, as dez selecionadas passaram todos os dias do mês indisponíveis, duas delas são habitações não integrais, ou seja, habitações de aluguel para apenas um quarto ou quarto compartilhado. O valor da diária, que foi obtido através da primeira coleta referente as habitações, variou entre 65 e 511 reais. Outro fator interessante é que todas as habitações são de regiões próximas a praias, como Ingleses, Canasvieiras e Campeche.

id_habitacao	descricao	num_quartos	num_banheiros	num_hospedes	cidade	latitude	longitude	habitacao_integral	valoraluguel	total_dias_ocupados
18089352	Excelente Padrão em Área Nobre - Portaria 24 horas	1	1	4	Florianopolis	-27,589,162	-48,552,581	habitacao integral	148	31
13688682	Quintete (Studio) a 50m do mar Praia dos Ingleses	1	1	2	Florianopolis	-27,441,947	-48,386,572	habitacao integral	80	31
29117667	Cabanas da Lua	1	1	2	Florianopolis	-27,428,776	-48,407,442	habitacao não integral	244	31
21045807	Praia dos Ingleses! 100 metros do mar!	2	2	5	Florianopolis	-27,425,305	-48,402,274	habitacao integral	511	31
14593755	Apto. confortável a 50m da praia em Canasvieiras	1	1	4	Florianopolis	-27,429,148	-48,456,632	habitacao integral	110	31
20047607	Casa 2qtos (c/ Ar) p/ Família a 8min a pé do Mar	2	2	4	Florianopolis	-27,430,600	-48,402,866	habitacao integral	120	31
15342452	Casa na praia, aproveite a natureza próximo ao mar	2	1	5	Florianopolis	-27,458,084	-48,384,239	habitacao integral	252	31
16883601	Apto para temporada nos Ingleses	1	1	4	Florianopolis	-27,451,536	-48,405,566	habitacao integral	125	31
15359185	Clean ,aconchegante e no Novo Campeche	1	1	4	Florianopolis	-27,662,722	-48,480,749	habitacao integral	156	31
9530891	Quarto confortável em região central na ilha	1	1	2	Florianopolis	-27,580,631	-48,510,719	habitacao não integral	65	31

Figura 7. Top 10 Habitações mais ocupadas em Março

No mês de abril das 10 habitações selecionadas todas permaneceram o mês inteiro indisponíveis, com apenas uma habitação sendo habitação não integral, de quarto compartilhado ou quarto individual. Semelhante ao mês de março às habitações se encontram em regiões de praia da Ilha, Jurerê, Canasvieiras e Ingleses. O valor da diária foi de 65 a 601 reais e o número de hóspedes máximo pesquisado foi de 2 à 10.

id_habitacao	descricao	num_quartos	num_banheiros	num_hospedes	cidade	latitude	longitude	habitacao_integral	valoraluguel	total_dias_ocupados
9743168	Apartamento nos Ingleses Norte, Florianopolis - SC	2	2	6	Florianopolis	-27,430,920	-48,401,800	habitacao integral	260	30
2104066	Kitnete linda em Floripa	1	1	2	Florianopolis	-27,583,660	-48,526,350	habitacao integral	88	30
21472614	Quarto na lagoa 2 Ju	1	1	2	Florianopolis	-27,602,270	-48,466,170	habitacao não integral	150	30
16257786	Casa do Caue	2	2	4	Florianopolis	-27,504,180	-48,418,000	habitacao integral	260	30
23553799	Floripa Cobertura com piscina Galvotas Floripa	3	3	8	Florianopolis	-27,423,397	-48,404,220	habitacao integral	300	30
9547413	Casa muito aconchegante em Jurerê	2	1	6	Florianopolis	-27,441,220	-48,482,060	habitacao integral	249	30
19513624	Mountain & Bamboo fields View Apartment	1	1	2	Florianopolis	-27,576,584	-48,422,669	habitacao integral	65	30
30105969	Apartamento praia dos Ingleses-Floripa [211] SI	3	2	7	Florianopolis	-27,422,930	-48,402,930	habitacao integral	376	30
22272952	Casa no Caçupé para temporada	1	1	5	Florianopolis	-27,520,100	-48,517,590	habitacao integral	161	30
27427016	Casa de temporada Rio Vermelho	3	3	10	Florianopolis	-27,483,798	-48,414,587	habitacao integral	601	30

Figura 8. Top 10 Habitações mais ocupadas em Abril

O resultado da consulta para o mês de maio resultou na tabela 3. Todas as habitações selecionadas permaneceram indisponíveis durante os 31 dias do mês, todas sendo habitações integrais e com os valores de diária mais esparsos, de 65 até 4.999 reais. As habitações se encontram em região de praia da cidade, sendo elas Cachoeira do Bom Jesus, Jurerê Internacional, Ingleses e Lagoa da Conceição. O número de hóspedes máximo variou de 2 até 10.

id_habitacao	descricao	num_quartos	num_banheiros	num_hospedes	cidade	latitude	longitude	habitacao_integral	valoraluguel	total_dias_ocupados
23126358	Cobertura Top 100m da praia	4	4	10	Florianopolis	-27,426,968	-48,401,955	habitacao integral	899	31
21603142	Apartamento/Cachoeira Bom Jesus/quadra mar	2	2	5	Florianopolis	-27,423,063	-48,436,474	habitacao integral	501	31
10448366	Porto Del Mar - Vista Interna e Piscina	2	2	6	Florianopolis	-27,429,167	-48,465,054	habitacao integral	256	31
19513624	Mountain & Bamboo fields View Apartment	1	1	2	Florianopolis	-27,576,584	-48,422,669	habitacao integral	65	31
16065832	Casa entre mar e lagoa	2	1	6	Florianopolis	-27,606,630	-48,438,430	habitacao integral	201	31
2201377	Jurerê a 50m da praia	1	1	3	Florianopolis	-27,439,380	-48,487,320	habitacao integral	384	31
28005599	Apto "o melhor no Villa Giardino" Ingleses Fpolis	2	2	5	Florianopolis	-27,416,878	-48,404,413	habitacao integral	451	31
30004834	Mansão na Búzios perto do Café de lá Music!	5	5	10	Florianopolis	-27,440,680	-48,498,170	habitacao integral	4,999	31
10574578	Centrinho da Lagoa	2	2	4	Florianopolis	-27,606,420	-48,466,840	habitacao integral	279	31
23878735	Apartamento em Ingleses	1	1	3	Florianopolis	-27,449,620	-48,396,735	habitacao integral	68	31

Figura 9. Top 10 Habitações mais ocupadas em Maio

As tabelas a seguir apresentam as habitações que arrecadaram o maior valor durante o período em que ficaram alugadas. A consulta busca pelo atributo total arrecadado da tabela fato Agregado Mês. No mês de março podemos observar que as habitações que mais arrecadaram possuíam os valores de diárias mais caros, entre 958 reais e 5.948. Com a diária de valor mais alto, o número de hóspedes ficou entre duas e dezesseis pessoas. Como observado na tabela 4 a seguir, as habitações com maior arrecadação, são de alto padrão e estão nas praias conhecidas por serem mais sofisticadas da Ilha de Florianópolis.

id_habitacao	descricao	num_quartos	num_banheiros	num_hospedes	cidade	latitude	longitude	habitacao_integral	valoraluguel	total_arrecadado
21986720	Costão do Santinho Resort - Vilas Portuguesas - Ap. Superio (Website hidden by Airbnb) Dias - 4 adultos e 2 crianças	2	2	6	Florianopolis	-27,464,276	-48,380,509	habitacao integral	2.704	83.824
2798194	Suite Master Costão do Santinho	1	1	3	Florianopolis	-27,444,379	-48,377,919	habitacao não integral	1.731	53.641
28954907	Casa temporaria Ingleses	3	5	7	Florianopolis	-27,432,622	-48,403,978	habitacao não integral	1.499	44.970
16602339	Férias dos Sonhos	2	2	5	Florianopolis	-27,462,674	-48,377,943	habitacao integral	1.299	40.269
1002944	Cobertura na Praia dos Ingleses - Florianopolis	3	2	6	Florianopolis	-27,424,220	-48,401,840	habitacao integral	1.201	37.231
7462148	Prédio Inteiro - 44 pessoas !!!	1.000	8	16	Florianopolis	-27,441,930	-48,385,500	habitacao integral	5.498	32.988
19907262	Cobertura em Condomínio a 50 Metros da Praia	5	3	10	Florianopolis	-27,430,608	-48,397,162	habitacao integral	1.299	31.176
20520426	Cobertura área nobre dos Ingleses	2	2	4	Florianopolis	-27,423,730	-48,403,280	habitacao integral	999	30.969
999011	Vista espetacular Cobertura frente ao mar Ingleses	4	4	8	Florianopolis	-27,418,280	-48,402,500	habitacao integral	967	29.977
21986719	Costão do Santinho Resort- Hotel Vilas Portuguesas - Apto Standard - 7 Dias - 2 Adultos + 1	1	1	2	Florianopolis	-27,464,877	-48,380,066	habitacao integral	958	29.698

Figura 10. Habitações com Maior Arrecadação em Março

No mês de abril as habitações com maior arrecadação seguem os mesmos padrões do mês anterior, estando localizadas em região de alto padrão. Essas habitações hospedam um número máximo de hóspedes entre quatro e dez pessoas e são em maioria habitações integrais.

id_habitacao	descricao	num_quartos	num_banheiros	num_hospedes	cidade	latitude	longitude	habitacao_integral	valoraluguel	total_arrecadado
28249878	COBERTURA 2 SUITES SPA PISCINA EM JURERÊ	2	3	6	Florianopolis	-27,444,193	-48,489,252	habitacao integral	24.999	749.970
22451529	Apto 203 - Boulevard Ponta das Canas	1	1	4	Florianopolis	-27,403,262	-48,426,904	habitacao integral	10.001	300.030
83547	Best place in Jurerê Internacional, Florianopolis	4	5	8	Florianopolis	-27,437,320	-48,512,950	habitacao integral	5.000	150.000
14551534	Ótima casa bem próxima ao mar	5	5	10	Florianopolis	-27,435,960	-48,501,950	habitacao integral	3.250	97.500
22231063	Jurerê Internacional Brazil Beach House	5	5	10	Florianopolis	-27,439,700	-48,498,630	habitacao integral	3.001	90.030
30004834	Mansão na Búzios perto do Café de lá Music!	5	5	10	Florianopolis	-27,440,680	-48,498,170	habitacao integral	4.999	84.983
21924933	Beautiful home in Jurerê Internacional Brazil	7	5	12	Florianopolis	-27,440,480	-48,506,570	habitacao integral	2.500	67.500
1302206	Casa magnifica de frente para o mar	5	7	10	Florianopolis	-27,415,950	-48,429,320	habitacao integral	2.202	66.060
3944422	Casa com piscina e vista da lagoa e mar	4	4	8	Florianopolis	-27,607,440	-48,485,140	habitacao integral	1.949	58.470
27692835	Elite Praia-Mole-Villa	4	400	10	Florianopolis	-27,600,904	-48,436,043	habitacao integral	1.874	56.220

Figura 11. Top 10 Habitações com Maior Arrecadação em Abril

Em maio o padrão de habitações se repete, sendo que as habitações com maior arrecadação se encontram em regiões de alto padrão e próximas ao mar. O número de hóspedes máximo que pode ser recebido também permanece maior, entre quatro e quatorze.

id_habitacao	descricao	num_quartos	num_banheiros	num_hospedes	cidade	latitude	longitude	habitacao_integral	valor4lug	total_arrecadado
28249878	COBERTURA 2 SUITES SPA PISCINA EM JURERÊ	2	3	6	Florianópolis	-27,444,193	-48,489,252	habitacao integral	24,999	774,969
22451529	Apto 203 - Boulevard Ponta das Canas	1	1	4	Florianópolis	-27,403,262	-48,426,904	habitacao integral	10,001	310,021
83547	Best place in Jurerê Internacional, Florianópolis	4	5	8	Florianópolis	-27,437,320	-48,512,950	habitacao integral	5,000	155,000
30004834	Mansão na Búzios perto do Café de lá Music!	5	5	10	Florianópolis	-27,440,680	-48,498,170	habitacao integral	4,999	154,969
27902175	Casa completa para 14 hóspedes com piscina	7	5	14	Florianópolis	-27,439,030	-48,504,680	habitacao integral	4,117	111,159
14551534	Ótima casa bem próxima ao mar	5	5	10	Florianópolis	-27,435,960	-48,501,950	habitacao integral	3,250	100,750
8652841	FLORIANOPOLIS EN LA MEJOR PLAYA A 30 METROS DEL MAR	4	3	8	Florianópolis	-27,424,050	-48,438,810	habitacao integral	2,228	69,068
1302206	Casa magnífica de frente para o mar	5	7	10	Florianópolis	-27,415,950	-48,429,320	habitacao integral	2,202	68,262
3944422	Casa com piscina e vista da lagoa e mar	4	4	8	Florianópolis	-27,607,440	-48,485,140	habitacao integral	1,949	60,419
27692835	Elite Praia-Hotel Villa	4	400	10	Florianópolis	-27,600,904	-48,436,043	habitacao integral	1,874	58,094

Figura 12. Habitações com Maior Arrecadação em Maio

Para os mapas de calor a seguir foram utilizadas as informações sobre latitude e longitude de cada habitação presentes na tabela dimensão Habitação. A figura a seguir apresenta o gráfico de calor obtido para todas as habitações disponíveis nos meses de março à maio. Nele podemos observar que há grande concentração de habitações no norte da Ilha de Florianópolis, em Canasvieiras, Jurerê, Costão do Santinho, Daniela e Santo Antônio de Lisboa. Outras regiões com alta concentração de habitações disponíveis são Lagoa da Conceição, Barra da Lagoa, Itacorubi, Carvoeira e Centro. Nas regiões do Sul da Ilha também há alguma concentração maior, representada pela região em amarelo, no Campeche. Muitas das regiões próximas à praias também apresentam alguma concentração de imóveis disponíveis, apesar de ser menor que nos bairros de praia.

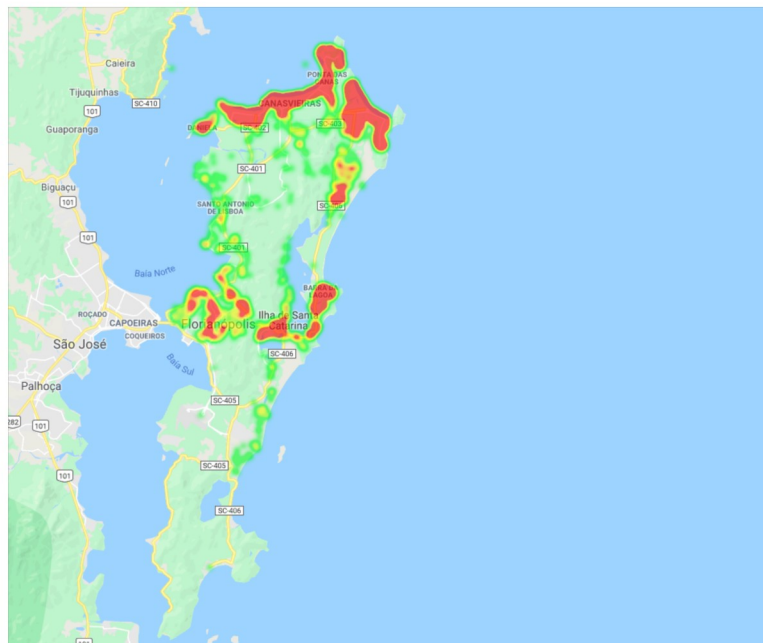


Figura 13. Habitações disponíveis em Março, Abril e Maio

A seguir os mapas apresentados se referem apenas às habitações que estiverem indisponíveis no período. No mês de março de 2019 a concentração de habitações permanece nas regiões próximas à praia e está semelhante ao gráfico de habitações totais disponíveis durante o período. Nas regiões do norte da Ilha há uma menor concentração, observado pela

diminuição da região em vermelho em Jurerê, Daniela, Costão do Santinho e Ingleses. Na região do Rio Vermelho podemos observar que há menos habitações ocupadas do que disponíveis. Conclui-se a partir do mapa da figura 14 que as habitações alugadas permanecem em maioria nas regiões de praia e centro.

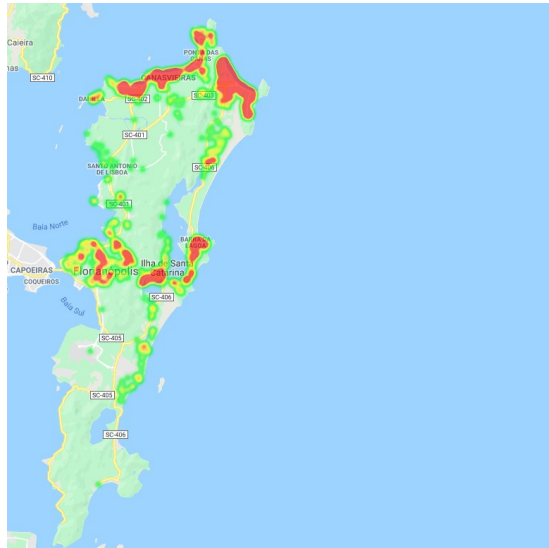


Figura 14. Habitações Alugadas em Março

No mês de abril a concentração de habitações ocupadas aumenta nas regiões próximas as praias de Jurerê Internacional, Daniela, Canasvieiras, Costão do Santinho, Barra da Lagoa e Lagoa da Conceição e regiões próximas como Rio Vermelho. As regiões dos bairros Itacorubi e Carvoeira também apresentam concentração significativa de imóveis ocupados. Ainda podemos ver algum acúmulo de habitações na região de Campeche. Em comparação ao mês anterior, podemos perceber que há pontos que cresceram em concentração de habitações na região de Santo Antônio de Lisboa.

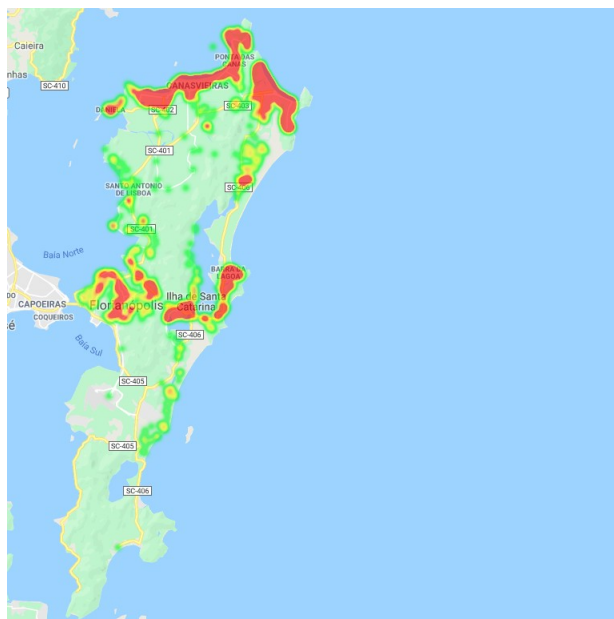


Figura 15. Habitações Alugadas em Abril

No mês de maio, a concentração de habitações permanece no norte da Ilha, e não ocorre no Sul, pois a região do Campeche não é apresentada no mapa a seguir. No entanto, vemos que há um acúmulo intenso de habitações alugadas nas regiões próximas a praia, apesar de Maio ser um mês com baixa temporada de turistas. Neste mês a concentração de habitações alugadas também cresce nos bairros próximos ao centro da cidade.

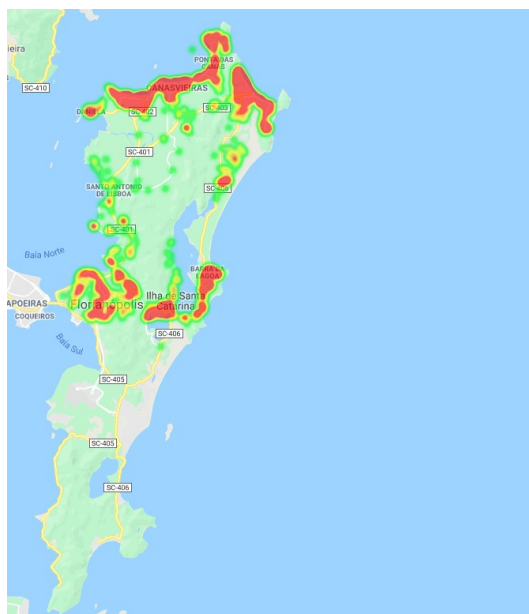


Figura 16. Habitações Alugadas em Maio

Através dos mapas e indicadores gerados no desenvolvimento deste estudo podemos perceber que a procura pelos imóveis disponibilizados via Airbnb na Ilha de Florianópolis, fora da alta temporada, tende a crescer durante os finais de semana e feriados, bem como o número de hóspedes. Este último, sendo um número máximo estimado através do atributo ocupação máxima de cada habitação.

As tabelas com as dez habitações com maior arrecadação e maior ocupação, nos forneceram algumas características sobre a procura. As habitações com maior ocupação tendem a possuir um menor valor de diária e hospedar menos pessoas, diferente das habitações com maior arrecadação que tendem a ser lugares de alto nível e estarem em regiões de alto padrão da ilha, bem como hospedar um maior número de pessoas.

Por meio dos mapas de calor, foi possível observar onde se concentram o maior número de habitações que estão disponíveis na região pelo Airbnb, concluindo que se encontram em maioria nos bairros próximos da praia como Jurerê e Canasvieiras. Contudo também é possível observar uma grande oferta nos bairros próximos ao centro, levantando a hipótese de que há turistas buscando lazer, mas também há turistas em viagens de negócios.

Dessa forma, pode-se concluir que os indicadores para habitação, número de hóspedes e ocupação propiciaram novos conhecimentos sobre a oferta, a procura e ocupação de imóveis via plataforma de aluguel por temporada, podendo servir de apoio a tomada de decisão para as partes interessadas do setor na região

5. Conclusão

Este trabalho teve como intuito a criação de um ambiente de análise dos dados sobre oferta de imóveis em Florianópolis disponibilizados no Airbnb aplicando técnicas de extração de conhecimento, através das análises e do ambiente projetado na ferramenta Metabase, concluindo assim que o trabalho cumpriu com seu objetivo geral.

Os objetivos específicos também foram atingidos. Por meio da coleta de dados e da projeção do ambiente de análise foi possível inferir alguns conhecimentos sobre a oferta e a ocupação dos imóveis disponibilizados para aluguel por temporada no período de outono na Ilha com os indicadores gerados. Por meio do atributo ocupação máxima foi possível inferir uma quantidade máxima de turistas que a região pode ter recebido durante o período de março a maio de 2019, construindo gráficos de linha que nos mostraram uma maior ocupação durante períodos de feriado e finais de semana, como pode ser observado na figura 11 do capítulo 4. Além disso, foi possível verificar o número de habitações oferecidas, bem como a demanda. Os mapas de calor nos proporcionaram uma visão de onde se localizam a maior parte das habitações oferecidas, que são regiões de praia, como Canasvieiras, Jurerê, Costão do Santinho, Barra da Lagoa e Lagoa da Conceição e regiões mais próximas ao centro da cidade.

Proporcionando um ambiente para discussão sobre o setor turístico da cidade de Florianópolis, percebemos que dados podem ser utilizados para melhorar a experiência de turistas que visitam a região durante o ano todo. Neste contexto, podemos concluir que o advento do Big Data e a Revolução dos Dados tem suma importância na transformação de dados em conhecimento.

6. Trabalhos Futuros

Através do desenvolvimento deste estudo pode-se perceber a oportunidade para o desenvolvimento de alguns trabalhos futuros que aprimorem o que aqui foi descrito, como por exemplo: a automatização melhora do script de coleta de dados para habitações individuais buscando construir uma base de dados maior; A adição de mais dados ao modelo afim de criar outras métricas e indicadores; E a análise e coleta de dados da alta temporada na região de Florianópolis.

Referências

Airbnb. About Us, 2019. <<https://news.airbnb.com/about-us/>> Acesso em: 28 out. 2019.

Almeida, F. Big data: Concept, potentialities and vulnerabilities. *Emerging Science Journal*, v. 2, 03 2018.

Askitas, N.; Zimmermann, K. The internet as a data source for advancement in social sciences. *International Journal of Manpower*, v. 36, p. 2–12, 04 2015.

Austin, C.; Kusumoto, F. The application of big data in medicine: current implications and future directions. *Journal of interventional cardiac electrophysiology : an international journal of arrhythmias and pacing*, v. 47, 01 2016.

Boyd, D.; Crawford, K. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication Society*, v. 15, p. 662–679, 01 2012.

Castillo, C. Effective web crawling. *SIGIR Forum*, v. 39, p. 55–56, 06 2005.

Christesen, Clayton M. Disruptive Innovation. *The Encyclopedia of Human-Computer Interaction*, 2nd Ed., [S. l.], p. -, 15 mar. 2015. Disponível em: <https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/disruptive-innovation>. Acesso em: 3 dez. 2019.

Cooley, R.; Mobasher, B.; Srivastava, J. Web mining: information and pattern discovery on the world wide web. In: *Proceedings Ninth IEEE International Conference on Tools with Artificial Intelligence*. [S.l.: s.n.], 1997. p. 558–567.

Einav, L.; Levin, J. The data revolution and economic analysis. *Innovation Policy and the Economy*, v. 14, 05 2013.

El-Sappagh, S. H. A.; Hendawi, A. M. A.; Bastawissy, A. H. E. A proposed model for data warehouse etl processes. *Journal of King Saud University - Computer and Information Sciences*, v. 23, n. 2, p. 91 – 104, 2011. ISSN 1319-1578.

<<https://www.sciencedirect.com/science/article/pii/S131915781100019X>> Acesso em: 28 out. 2019.

Foina, A. Métodos de aquisição de dados quantitativos na internet: o uso da rede como fonte de dados empíricos. *Ciência e Trópico*, v. 30, p. 283–296, 07 2003.

Gomes, P. C. T. WHAT ARE THE MAIN DIFFERENCES BETWEEN METRICS AND INDICATORS? Apr 2017. <<https://www.opservices.com/differences-between-metrics-and-indicators/>> Acesso em: 29 out. 2019.

Gretzel, U. et al. Smart tourism: foundations and developments. *Electronic Markets*, v. 25, 08 2015.

Gutteng, D. Airbnb: Disruptive innovation and the rise of an informal tourism accommodation sector. *Current Issues in Tourism*, p. 1–26, 12 2013.

Gutteng, D. A.; Smith, S. L. Assessing airbnb as a disruptive innovation relative to hotels: Substitution and comparative performance expectations. *International Journal of Hospitality Management*, v. 64, p. 1 – 10, 2017. ISSN 0278-4319.

Gyr, U. The history of tourism: Structures on the path to modernity. *European History Online*, 12 2010.

Hunter, W. C. et al. Constructivist research in smart tourism. *Asia Pacific Journal of Information Systems*, v. 25, p. 105–120, 03 2015.

Hurwitz, J.; Nugent, A.; Halper, F. *Big Data for Dummies*. [S.l.]: Wiley Brand, 2013.

Hwang, J.; Park, H.-Y.; Hunter, W. C. Constructivism in smart tourism research: Seoul destination image. *Asia Pacific Journal of Information Systems*, v. 25, p. 163–178, 03 2015.

Johnson, J. et al. Big data: big data, digitization, and social change. *Ubiquity*, v. 2017, p. 1–8, 12 2017.

Kaur, K.; Kaur, R. Internet of things to promote tourism: An insight into smart tourism. *International Journal of Recent Trends in Engineering Research*, v. 2, p. 357 – 361, 04 2016.

Kimball, R. et al. *The Data Warehouse Lifecycle Toolkit*. [S.l.]: Wiley, 1998.

Kitchin, R. *The data revolution: big data, open data, data infrastructures and their consequences*. [S.l.]: SAGE, 2014.

Laender, A. H. F. et al. A brief survey of web data extraction tools. *ACM SIGMOD Record Homepage archive*, v. 31, p. 84–93, 06 2002.

Lins, H. Florianópolis, cluster turístico? Revista Turismo em Análise, v. 11, n. 2, p. 55–70, 11 2000.

Lins, H. Interações, aprendizagem e desenvolvimento:ensaio sobre o turismo em florianópolis. Revista Turismo em Análise, v. 9, n. 1, p. 107–121, 01 2007

Machado, C. C. et al. Um web crawler para projeções e análise de vulnerabilidades de segurança e consistência estrutural de páginas web. Revista de Empreendedorismo, Inovação e Tecnologia, v. 2, p. 3–12, 2015.

Marinho, L; Girardi, R. Mineração na Web. Maranhão: UFMA, 2003. Disponível em: https://www.researchgate.net/publication/228369738_Mineracao_na_Web. Acesso em: 2 dez. 2019.

McAfee, A.; Brynjolfsson, E. Big data: The management revolution. Harvard business review, v. 90, p. 60–6, 68, 128, 10 2012.

Mitchel, R. Web Scraping with Python: Collecting Data from the Modern Web. [S.l.]: O'Reilly, 2015.

Moody, D.; Kortink, M. From enterprise models to dimensional models: a methodology for data warehouse and data mart design. In: . [S.l.: s.n.], 2000. p. 5.

Onberger, V. M.-S.; Cukier, K. Big Data: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK. [S.l.]: HMH Books, 2013.

Riahi, Y. Big data and big data analytics: Concepts, types and technologies. v. 5, p. 524–528, 11 2018.

Russom, P. Big data analytics. Transforming Data With Inteligence, 2011.

Schoeder, R. Big data: Shaping knowledge, shaping everyday life. *MATRIZES*, v. 12, p. 135–136, 08 2018.

Sigala, M. Web 2.0 in the tourism industry: a new tourism generation and new e-business models. 01 2007.

Srivastava, J. et al. Web usage mining: Discovery and applications of usage patterns from web data. *SIGKDD Explorations*, v. 1, n. 2, p. 12–23, 2000.

Vargiu, E.; Urru, M. Exploiting web scraping in a collaborative filtering-based approach to web advertising. *Artificial Intelligence Research*, v. 2, 01 2013.

Walton, J. K. Tourism. *Encyclopædia Britannica*, inc., 11 2018. <<https://www.britannica.com/topic/tourism>>. Acesso em: 28 out. 2019.

ČIŽIK, T. The information revolution and the european security architecture. 07 2018.

Santur, Estatísticas e Indicadores Turísticos: Pesquisa de Demanda Turística - Alta Temporada Anual - 2014-2017. [S. l.], 2018. Disponível em: <http://turismo.sc.gov.br/institucional/index.php/pt-br/informacoes/estatisticas-e-indicadores-turisticos/category/51-pesquisa-de-demanda-turistica-alta-temporada-anual-2014-2017>. Acesso em: 29 out. 2019.

Metabase: is the easy, open source way for everyone in your company to ask questions and learn from data.. Versão 0.33.4. [S. l.], 19 nov. 2007. Disponível em: <https://www.metabase.com/>. Acesso em: 14 out. 2019.

Kettle: Data Integration. Versão 8.2. [S. l.], 2004. Disponível em: <https://community.hitachivantara.com/s/article/data-integration-kettle>. Acesso em: 9 set. 2019.

Inmon, William H. *Building the Data Warehouse*. 3. ed. [S. l.]: Wiley, 2002. 412 p.

Significados. *In*: O que é Streaming. [S. l.], 20 out. 2017. Disponível em: <https://www.significados.com.br/streaming/>. Acesso em: 31 out. 2019.

Slee, Tom. Airbnb web site scraper. [S. l.], 02 2018. Disponível em: <https://github.com/tomslee/airbnb-data-collection/blob/master/README.md>. Acesso em: 31 out. 2019.

Slee, Tom. Airbnb Data Collection: Get the Data. [S. l.], 7 ago. 2017. Disponível em: <http://tomslee.net/airbnb-data-collection-get-the-data>. Acesso em: 31 out. 2019.

Schoenherr, Steven E. The Digital Revolution. 10. mai. 2004. Disponível em: <https://web.archive.org/web/20070103161029/http://history.sandiego.edu/gen/recording/digitalrev.html>. Acesso em: 3 dez. 2019.

Tavolari, Bianca. AirBnB e os impasses regulatórios para o compartilhamento de moradia: notas para uma agenda de pesquisa em direito. Nov. 2017. Disponível em: https://www.researchgate.net/publication/321156274_AirBnB_e_os_impasse_regulatorios_para_o_compartilhamento_de_moradia_notas_para_uma_agenda_de_pesquisa_em_direito. Acesso em: 6 dez. 2019.

APÊNDICE L – Código Fonte

```

'''
Requirements: pip install fake_useragent
               pip install bs4
               pip install requests

Cmd line execution: python crawler.py <arquivo_com_cnpj>
'''

import requests
import json
import random
import time
import logging
import logging.handlers as handlers
import pandas as pd
from urllib.request import Request, urlopen
from bs4 import BeautifulSoup
from fake_useragent import UserAgent
from urllib.error import URLError, HTTPError
from requests.exceptions import ProxyError

start = time.time()

#code related with random proxies
ua = UserAgent() # From here we generate a random user agent
proxies = [] # Will contain proxies [ip, port]

timestamp = time.strftime('%Y%m%d_%H%M%S')

# Logging setup
logger = logging.getLogger('crawler')
logger.setLevel(logging.INFO)

# Set the Log formatter
formatter = logging.Formatter('%(asctime)s - %(name)s - %(levelname)s - %
(message)s')

#time rotating log -> when: s (second) / m (minute) / h (hour) / d (day) / w0-w6

```

```

(weekday) / midnight
logHandler = handlers.TimedRotatingFileHandler('normal_'+timestamp+'.log',
when='d', interval=1, backupCount=500)
logHandler.setLevel(logging.DEBUG)
logHandler.setFormatter(formatter)

errorLogHandler = handlers.RotatingFileHandler('error_'+timestamp+'.log',
maxBytes=500000, backupCount=500)
errorLogHandler.setLevel(logging.ERROR)

logger.addHandler(logHandler)
logger.addHandler(errorLogHandler)

logger.info('Starting program...')

'''Get the latest proxies'''
def getLatestProxies():

    try:
        logger.debug('[getLatestProxies] Grabbing latest proxies... ')

        # Retrieve latest proxies
        proxies_req = Request('http://givemeproxy.com/')
        proxies_req.add_header('User-Agent', ua.random)
        proxies_doc = urlopen(proxies_req, timeout=10).read().decode('utf8')

        soup = BeautifulSoup(proxies_doc, 'html.parser')
        proxies_table = soup.find(id='tablepress-2')

        # Save proxies in the array
        for row in proxies_table.tbody.find_all('tr'):
            if row.find_all('td')[3].string == 'YES':
                proxies.append({
                    'ip': row.find_all('td')[0].string,
                    'port': row.find_all('td')[1].string
                })

```



```

        i=len(proxies)
        logger.debug('[getLatestProxies] proxies len: %s' % i)
        logger.debug('[getLatestProxies] %s' % proxies)
    except Exception as ex:
        logger.error('[getLatestProxies] Erro ao tentar pegar os proxies.
Exception: %s' % ex)

```

```

''' Retrieve a random index proxy (we need the index to delete it if not
working)'''

```

```

def random_proxy():
    return random.randint(0, len(proxies) - 1)

```

```

''' Test if a proxy is working '''

```

```

def proxyIsAlive(s, proxy):
    try:
        req = Request('http://icanhazip.com')
        req.set_proxy(proxy['ip'] + ':' + proxy['port'], 'http')
        req.add_header('User-Agent', ua.random)
        urlopen(req, timeout=2)
        my_ip = urlopen(req).read().decode('utf8')
        logger.debug('[proxyIsAlive] IP: %s' % my_ip)
        return True
    except Exception:
        logger.warning("[proxyIsAlive] Proxy não funcionou.")
        return False

```

```

'''Get the data from url'''

```

```

def getUrl(url):
    try:
        s = requests.Session()
        a = requests.adapters.HTTPAdapter(max_retries=0)
        s.mount('https://', a)

        # Choose a random proxy
        proxy_index = random_proxy()
        proxy = proxies[proxy_index]

```

```

for n in range(100):
    isOpen = proxyIsAlive(s, proxy)
    if isOpen:
        break
    else:
        if len(proxies) == 5:
            getLatestProxies()
        else:
            del proxies[proxy_index]
            logger.warning('[getUrl] Proxy %s:%s deleted.' % (proxy['ip'],
proxy['port']))

            proxy_index = random_proxy()
            proxy = proxies[proxy_index]

    # call url
    https_proxy=proxy['ip'] + ':' + proxy['port']
    prox = {"https": https_proxy}
    logger.info('[getUrl] %s' % url)

    response = s.get(url, proxies=prox, headers={'user-agent':
ua.random}, timeout=3)
    print(response.content)

    if response == None:
        logger.warning("[getUrl] Response is None. Trying again...")
        getUrl(url)
    else:
        statuscode = response.status_code
        if statuscode != 200:
            logger.warning("[getUrl] Status code is: %s" % statuscode)
            if statuscode == 429:
                # logger.warning("[getUrl] Too many requests...trying
again...")

                # getUrl(url)
                raise ValueError("Muitas requisições")
            if statuscode == 403:
                return response
            else:
                logger.warning("[getUrl] Trying again...")

```

```

        getUrl(url)
    else:
        return response
except ProxyError as e:
    logger.warning("[getUrl] Erro: %s " % e)
    logger.warning("[getUrl] Trying again...")
    if len(proxies) == 5:
        getLatestProxies()
    else:
        del proxies[proxy_index]
        logger.warning('[getUrl] Proxy %s:%s deleted.' % (proxy['ip'],
proxy['port']))
        getUrl(url)

```

```
'''Main method: reads cnpjs, get json and record'''
```

```
def main():
```

```
    room = pd.read_csv('arquivos/room.csv', sep=',')
```

```
    room_ids = room["room_id"].tolist()
```

```
    lista_ids = room_ids
```

```
    getLatestProxies()
```

```
    refresh_start = time.time()
```

```
    with open('availability_june.txt', 'a+', encoding='utf-8') as f:
```

```
        #iterate over id list (this id list is retrivied using Tom Slee Crawler for
Rooms in Airbnbn
```

```
        for id in lista_ids:
```

```

logger.info('-----')
-----')

```

```
        logger.info('[main] Processing')
```

```

        # https://www.airbnb.com.br/api/v2/ calendar_months?
        _format=with_conditions&count=2&currency=BRL&key=d306zoyjsyarp7ifhu67rjxn52tv0t20&l
        isting_id=21976565&locale=pt&month=4&year=2019
        # url = 'https://www.airbnb.com.br/api/v2/calendar_months?
        _format=with_conditions&count=2&currency=BRL&key=d306zoyjsyarp7ifhu67rjxn52tv0t20&l
        isting_id=%s&locale=pt&month=4&year=2019' % id
        url = '' % id
        print(url)

    try:
        response = getUrl(url)
        if response:
            statuscode = response.status_code
            if statuscode == 200:

                #load json data
                logger.debug(['main] Getting json...')
                dados = dict(json.loads(response.content, encoding='utf-
8'))

                #handle the error that comes from api
                if 'message' in dados:
                    logger.error("[main] ##### Erro para url: %s" %
url)

                    errorMsg = dados['message']
                    logger.error("[main] Não pegou os dados da URL: %s
Motivo: %s " % (url, errorMsg))
                    try:
                        logger.error("erro")
                    except Exception as ex:
                        logger.error("[main] ### Erro ao escrever a id %s
no arquivo " % id)

                        logger.error("[main] ### Erro %s " % ex)
                    pass

                continue

        lastmodified = time.strftime('%Y-%m-%d %H:%M:%S')

```

```

dados['insert_date']= lastmodified
dados['last_update']= lastmodified
dados['id']= id

logger.debug(dados)

try:
    json.dump(dados,f, ensure_ascii=False)
    f.write('\n')
except Exception as e:
    logger.error('[main] Erro ao escrever o json no
arquivo. %s' % e)

else:
    logger.error("[main] ##### Erro para url: %s" % url)
    logger.error("[main] Status code: %s " % statuscode)

else:
    logger.error('[main] Response: %s' % response)
    logger.error("[main] ##### Erro para id: %s" % id)

refresh_end = time.time()
execution_time = (refresh_end - refresh_start)

if(execution_time >= 600):
    refresh_start = time.time()
    logger.info('----- Total de id processados
-----')
    getLatestProxies()

except HTTPError as e:
    logger.error("[main] ### HttpError ##### Erro para id: %s" % id)
    logger.error('[main] The server couldn\'t fulfill the request.
Error code: %s' % e.code)
    continue

except URLError as e:
    logger.error("[main] ### URLError ##### Erro para url: %s" % url)
    if hasattr(e, 'reason'):

```

```
        logger.error('[main] We failed to reach a server. Reason: %s' %
e.reason)

    elif hasattr(e, 'code'):
        logger.error('[main] Error code: %s' % e.reason)
    continue

except Exception as e:
    logger.error("[main] ### Exception ##### Erro para url: %s" % url)
    logger.error("[main] Exception: %s" % e)
    continue

f.close()

logger.info('-----
-----')

#-----end main -----

if __name__ == '__main__':
    main()
```