

UNIVERSIDADE FEDERAL DE MINAS GERAIS  
ESCOLA DE CIÊNCIA DA INFORMAÇÃO

RAFAEL ALMEIDA DE OLIVEIRA

EXTRAÇÃO DE DADOS WEB COMO SUPORTE NA ELABORAÇÃO DE  
INDICADORES DO TURISMO DE MINAS GERAIS: UMA INICIATIVA EM BIG DATA

BELO HORIZONTE

2017

RAFAEL ALMEIDA DE OLIVEIRA

EXTRAÇÃO DE DADOS WEB COMO SUPORTE NA ELABORAÇÃO DE  
INDICADORES DO TURISMO DE MINAS GERAIS: UMA INICIATIVA EM BIG DATA

Dissertação apresentada ao Programa de Pós-Graduação em Gestão e Organização do Conhecimento da Escola de Ciência da Informação da Universidade Federal de Minas Gerais para obtenção do grau de Mestre em Ciência da Informação.

Linha de Pesquisa: Gestão e Tecnologia

Orientadora: Renata Maria Arantes Baracho Porto

BELO HORIZONTE

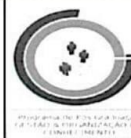
2017

O48e Oliveira, Rafael Almeida de.  
Extração de dados web como suporte na elaboração de indicadores do turismo de Minas Gerais [manuscrito] : uma iniciativa em big data / Rafael Almeida de Oliveira. – 2017.  
88 f., enc. : il.

Orientadora: Renata Maria Arantes Baracho Porto.  
Dissertação (Mestrado) – Universidade Federal de Minas Gerais, Escola de Ciência da Informação.  
Referências: f. 84-88.

1. Ciência da informação – Teses. 2. Turismo e Estado – Teses. 3. Sites da Web – Indicadores – Teses. 4. Gerenciamento da informação – Teses. I. Título. II. Porto, Renata Maria Arantes Baracho. III. Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

CDU: 004: 379.85




## FOLHA DE APROVAÇÃO

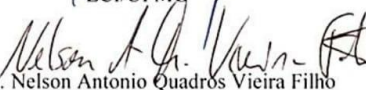
### EXTRAÇÃO DE DADOS WEB COMO SUPORTE NA ELABORAÇÃO DE INDICADORES DO TURISMO DE MINAS GERAIS: UMA INICIATIVA EM BIG DATA

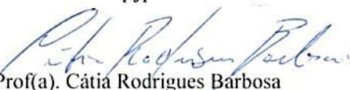
#### RAFAEL ALMEIDA DE OLIVEIRA

Dissertação submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, como requisito para obtenção do grau de Mestre em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, área de concentração REPRESENTAÇÃO DO CONHECIMENTO, linha de pesquisa Gestão e Tecnologia.

Aprovada em 06 de março de 2017, pela banca constituída pelos membros:

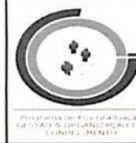
  
Prof(a). Renata Maria Abrantes Baracho Porto - Orientadora  
ECI/UFMG

  
Prof(a). Nelson Antonio Quadros Vieira Filho  
FJP

  
Prof(a). Cátia Rodrigues Barbosa  
ECI/UFMG

  
Prof(a). Max Cirino de Mattos  
UNA

Belo Horizonte, 6 de março de 2017.



## ATA DA DEFESA DA DISSERTAÇÃO DO ALUNO RAFAEL ALMEIDA DE OLIVEIRA

Realizou-se, no dia 06 de março de 2017, às 14:00 horas, sala 1000 - ECI/UFMG, da Universidade Federal de Minas Gerais, a defesa de dissertação, intitulada *EXTRAÇÃO DE DADOS WEB COMO SUPORTE NA ELABORAÇÃO DE INDICADORES DO TURISMO DE MINAS GERAIS: UMA INICIATIVA EM BIG DATA*, apresentada por RAFAEL ALMEIDA DE OLIVEIRA, número de registro 2015665859, graduado no curso de TURISMO, como requisito parcial para a obtenção do grau de Mestre em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, à seguinte Comissão Examinadora: Prof(a). Renata Maria Abrantes Baracho Porto - Orientadora (ECI/UFMG), Prof(a). Nelson Antonio Quadros Vieira Filho (FJP), Prof(a). Cátia Rodrigues Barbosa (ECI/UFMG), Prof(a). Max Cirino de Mattos (UNA).

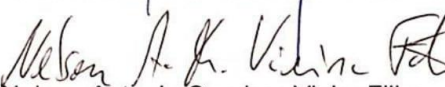
A Comissão considerou a dissertação:

Aprovada

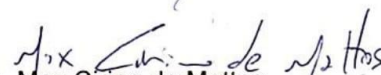
Reprovada

Finalizados os trabalhos, lavrei a presente ata que, lida e aprovada, vai assinada por mim e pelos membros da Comissão.  
Belo Horizonte, 06 de março de 2017.

  
Prof(a). Renata Maria Abrantes Baracho Porto

  
Prof(a). Nelson Antonio Quadros Vieira Filho

  
Prof(a). Cátia Rodrigues Barbosa

  
Prof(a). Max Cirino de Mattos

## **AGRADECIMENTOS**

Agradeço aos meus pais e meus familiares por serem meus guias durante toda a caminhada;

À minha esposa pela paciência, dedicação, apoio e amor incondicional que me fortalece dia-a-dia;

À minha orientadora pelos ensinamentos e por ter apostado nesse desafio;

Aos colegas e demais professores da ECI pela caminhada, troca de conhecimento e aprendizado constante;

A equipe do Import.io pelas orientações que foram fundamentais para esse trabalho;

Aos amigos e colegas da SETUR-MG por auxiliarem e acreditarem no potencial dessa pesquisa e pela amizade de longa data;

A todos aqueles que durante a trajetória me inspiraram a acreditar que é possível criar políticas públicas com criatividade e inovação.

A todos aqueles que acreditam que o conhecimento não tem preço.

*“Seja curioso. Leia de tudo. Tente coisas novas. O que as pessoas chamam de inteligência se resume a curiosidade”.*  
*(Aaron Swartz. 1983 - 2013)*

## RESUMO

A pesquisa teve como objetivo estudar o fenômeno *Big Data* e a possibilidade de utilização de ferramentas de extração de dados em ambiente *web* para auxiliar na elaboração de indicadores referente ao turismo em Minas Gerais. Primeiramente, realizou-se uma fundamentação teórica sobre conceitos fundamentais da ciência da informação para a contextualização do tema, além de pontuar o papel das ferramentas de extração de informações via *web*. Após essa etapa, foi realizado um estudo de caso com uma ferramenta de extração de dados denominada *Import.io* para coletar dados do site de viagens *TripAdvisor*, buscando as principais informações dos atrativos turísticos de Minas Gerais transformando-as em um banco de dados estruturado. Assim, foi possível analisar informações como a oferta de atrativos por categorias no estado e por município, o número de avaliações, o perfil dos visitantes, o nível de satisfação e o período de maior visitação de cada um dos atrativos. Para comprovar a utilidade das informações capturadas foi realizado um acompanhamento dos dados referentes ao Conjunto Arquitetônico da Pampulha com o objetivo de avaliar um possível impacto do seu reconhecimento como patrimônio mundial da humanidade na percepção dos visitantes. Os resultados mostraram que é possível utilizar a extração de dados disponibilizados em espaço *web* para a criação de indicadores de monitoramento do turismo que auxiliem o poder público na tomada de decisão. Constatou-se também a necessidade de uma maior discussão sobre a forma de disponibilização dos dados por parte das plataformas *online* para órgãos governamentais, visando dar utilidade à informação que possam beneficiar o público final. A metodologia apresentada poderá auxiliar as esferas públicas em extrair informações estratégicas que estão disponibilizadas na *web*, com baixo custo, otimizando ações e garantindo uma melhoria no uso de recursos.

**Palavras-chave:** *big data*, extração de dados, turismo, políticas públicas, recuperação da informação

## ABSTRACT

The research aims to study the phenomenon called “Big Data” and the possibility of using web data extraction tools (web scrapers) to help the development of indicators about tourism in Minas Gerais State (Brazil). For that, it was carried out a bibliographical review of authors related to information science to contextualize the subject, as well as to emphasize the role of web information extraction tools. After this step, we used a case study with a web scraper tool to collect data from TripAdvisor, searching for key information about Minas Gerais’ tourist attractions and turning them into a structured database. Thus, it was possible to analyse information such as the division of tourist attractions by categories from the state and municipalities, the number of evaluations, visitors’ profiles, satisfaction levels, and the period of most visits at each of the attractions. To prove the use of the information captured it was carried out a follow-up of the data concerning the Pampulha Architectural Complex with the objective of evaluating a possible impact of its recognition as a world heritage site in the visitors’ perception. The results showed that it is possible to use data from the platform to monitor actions and create indicators that aim to assist public decision-making. However, there is still need for further discussion about the availability of data delivered by online companies to the final public, which could be used by government agencies. We expect this methodology to assist the state authorities and municipalities to extract strategic information that is already available on the web at low costs, improving actions and ensuring an improvement in the use of public resources in tourism policies.

**Keywords:** big data, web scraping, tourism, information retrieval, public policies

## LISTA DE FIGURAS

|   |    |
|---|----|
| FIGURA 1 – Principais etapas da pesquisa.....   | 20 |
| FIGURA 2 – Subcategorias no cadastramento de atrativos.....   | 36 |
| FIGURA 3 – Tela inicial do <i>import.io</i> .....   | 45 |
| FIGURA 4 – Seleção automática de informações pelo <i>import.io</i> .....                              | 46 |
| FIGURA 5 – Tabela em branco para criação de colunas.....  | 46 |
| FIGURA 6 – Informações selecionadas para extração no <i>TripAdvisor</i> .....                         | 48 |
| FIGURA 7 – Visualização das informações selecionadas na aba “ <i>data view</i> ”.....                 | 49 |
| FIGURA 8 – Tela de trabalho ( <i>dashboard</i> ) – <i>import.io</i> .....                             | 50 |
| FIGURA 9 – Página “o que fazer” do município de Belo Horizonte .....                                  | 52 |
| FIGURA 10 – Página “o que fazer” em Minas Gerais e <i>link</i> “mais” .....                           | 53 |
| FIGURA 11 – Tela de trabalho ( <i>dashboard</i> ) do extrator 2 com as <i>URLs</i> dos destinos ..... | 54 |

## LISTA DE TABELAS

|  |    |
|--|----|
| TABELA 1 – Percentual de cadastros de atrativos de Minas Gerais no site <i>TripAdvisor</i> por categoria .....   | 59 |
| TABELA 2 – Distribuição dos atrativos de Ouro Preto cadastrados no <i>TripAdvisor</i> por categorias .....   | 60 |
| TABELA 3 – Percentual de avaliações e satisfação dos atrativos do município de Mariana por categorias.....   | 61 |
| TABELA 4 – Percentual de avaliações e nota média da categoria “museus especializados” por município.....   | 62 |
| TABELA 5 – Municípios com maior número percentual de avaliações por perfil de visitante.....   | 64 |
| TABELA 6 – Principais categorias visitadas pelos tipos de visitantes nos atrativos de Minas Gerais cadastrados no <i>TripAdvisor</i> .....   | 65 |
| TABELA 7 – Percentual de avaliações realizadas por período do ano no município de Carrancas e no distrito de Monte Verde.....  | 67 |
| TABELA 8 – Percentual de avaliações realizadas nas categorias “museus especializados” e “obras arquitetônicas” por período do ano em Belo Horizonte.....                             | 68 |
| TABELA 9 – Número de avaliações absolutas e a variação percentual dos principais atrativos de Belo Horizonte entre os meses de julho e agosto de 2016 no <i>TripAdvisor</i> .....    | 70 |
| TABELA 10 – Número de avaliações absolutas e a variação percentual dos principais atrativos de Belo Horizonte entre os meses de julho e dezembro de 2016 no <i>TripAdvisor</i> ..... | 70 |
| TABELA 11 – Nota média das avaliações e a variação percentual dos principais atrativos de Belo Horizonte entre os meses de julho e agosto de 2016 .....                              | 71 |
| TABELA 12 – Nota média das avaliações e a variação percentual dos principais atrativos de Belo Horizonte entre os meses de julho e dezembro de 2016 .....                            | 72 |

## LISTA DE ABREVIATURAS

|           |   |   |
|-----------|---|---|
| API       | - | <i>Application Programming Interface</i>                                |
| CSV       | - | <i>Comma Separated Values</i>   |
| GPS       | - | <i>Global Positioning System</i>  |
| IP        | - | <i>Internet Protocol</i>  |
| MTUR      | - | Ministério do Turismo   |
| OMT       | - | Organização Mundial de Turismo  |
| PMDI      | - | Plano Mineiro de Desenvolvimento Integrado                              |
| PPAG      | - | Plano Plurianual de Ação Governamental                                  |
| SEPLAG-MG | - | Secretaria de Estado de Planejamento e Gestão de Minas Gerais           |
| SETUR-MG  | - | Secretaria de Estado de Turismo de Minas Gerais                         |
| URL       | - | <i>Uniform Resource Locator</i>   |
| UNESCO    | - | <i>United Nations Educational, Scientific and Cultural Organization</i> |
| WTTC      | - | <i>World and Travel Tourism Council</i>                                 |

## SUMÁRIO

|   |           |
|---|-----------|
| <b>1 INTRODUÇÃO .....</b>   | <b>15</b> |
| <b>2 FUNDAMENTAÇÃO TEÓRICA .....</b>  | <b>21</b> |
| 2.1 BIG DATA.....   | 21        |
| 2.1.1 Definição de <i>Big Data</i> .....  | 22        |
| 2.1.2 Big Data e a sociedade .....  | 24        |
| 2.1.3 <i>Big Data</i> nas organizações.....   | 26        |
| 2.1.4 Extração de dados <i>web</i> .....  | 29        |
| 2.2 TURISMO.....  | 31        |
| 2.2.1 O turismo na era digital .....  | 32        |
| 2.2.2 Avaliação de viagens em ambiente <i>web</i> : o exemplo do <i>site TripAdvisor</i> .....    | 33        |
| 2.2.3 O uso de indicadores no turismo.....  | 38        |
| <b>3 METODOLOGIA .....</b>  | <b>41</b> |
| 3.1 COLETA DE DADOS .....   | 42        |
| 3.1.1 Seleção de informações para captura: extrator 1 .....                                       | 44        |
| 3.1.2 Coleta de <i>URLs</i> dos atrativos de Minas Gerais: extrator 2.....                        | 50        |
| 3.2 ORGANIZAÇÃO DO BANCO DE DADOS .....   | 55        |
| <b>4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS .....</b>  | <b>58</b> |
| 4.1 OFERTA DE ATRATIVOS.....  | 58        |
| 4.2 AVALIAÇÕES.....   | 61        |
| 4.3 PERFIL DOS VISITANTES .....   | 63        |
| 4.4 ÉPOCA DE VISITAÇÃO.....   | 66        |
| <b>5 PATRIMÔNIO CULTURAL DA HUMANIDADE: O CASO DO CONJUNTO<br/>ARQUITETÔNICO DA PAMPULHA.....</b> | <b>69</b> |
| <b>6 CRIAÇÃO DE INDICADORES.....</b>  | <b>74</b> |
| <b>7 CONSIDERAÇÕES FUTURAS.....</b>   | <b>82</b> |

|                          |           |
|--------------------------|-----------|
| <b>8 CONCLUSÃO .....</b> | <b>84</b> |
| <b>REFERÊNCIAS.....</b>  | <b>87</b> |

## 1 Introdução

Presenciamos na sociedade moderna, um aumento do volume de dados por cada pessoa ou organização no seu dia-a-dia. Com o avanço da tecnologia exemplificado pelas altas conexões de internet e utilização de dispositivos móveis, o acesso à informação se tornou mais fácil, gerando um volume de dados bem maior do que nas décadas anteriores. De acordo com levantamento realizado pela empresa de tecnologia Cisco (2015) o tráfego de *IPs*<sup>1</sup> na internet mundial ultrapassará 1 *zettabyte* ( $10^{21}$ ) no final do ano de 2016, sendo esse valor duplicado até o ano de 2019. Em média o tráfego de *IPs* irá crescer em média, 23% ao ano.

Desde as décadas passadas, esses números eram observados (em escala menor), porém já impactavam diretamente na economia. Cohen (2002) afirmou que vivíamos na era da “economia da informação” onde a forma de utilizar a informação por empresas e governos era decisiva para obter bons resultados econômicos. Essa lógica também é percebida nos dias de hoje, onde se pode afirmar que qualquer atividade organizacional é dependente de informação (BASTOS, 2006), criando uma nova dinâmica nos ambientes organizacionais internos e externos. Zdepsky (2004) diz que a busca incessante de informações para a sobrevivência das organizações faz com que seus gestores busquem, cada vez mais, metodologias que facilitem a filtragem e integração desses dados.

Para alguns autores, uma das formas de auxiliar na interpretação e utilização dos dados está relacionada ao fenômeno informacional definido como “*Big Data*”, ou seja, são conjuntos de dados gigantescos (que podem chegar entre *terabytes* e *exabytes*) e muito complexos (tais como informações contidas desde aplicativos móveis a redes sociais) que necessitam de servidores, formas de gerenciamento, análises e tecnologias de visualização robustos e diferenciados (CHEN, CHIANG e STOREY, 2012), que não podem ser analisados por programas ou ferramentas de uso comum da informação (BARBIERI, 2013; DAVENPORT, 2014).

As discussões sobre *Big Data* vêm ganhando a cada dia, maior importância no campo da Ciência da Informação. Ribeiro (2014; p.102) cita que “a Ciência da Informação nasceu e foi gestada com o objetivo maior de apresentar

---

<sup>1</sup> Conhecido como “*internet protocol*”, o IP pode ser definido como um protocolo para entrega universal de dados na rede de computadores, sendo que cada protocolo possui um endereço específico correspondente a uma máquina específica (FARREL, 2005).

solução para problemas ligados ao uso de dados e informação e como tal, tem um importante papel nos estudos que envolvem o tema *Big Data*". Para Boyd e Crawford (2012), estudos de *Big Data* são multidisciplinares, visto sua potencialidade de aplicação para diferentes áreas.

Dentre essas áreas, o turismo caracteriza-se por ser uma atividade que lida com diferentes setores da economia, agregando-os e fazendo-os trabalhar de forma integrada e em rede. Assim, é necessário intermediar não só os interesses, mas principalmente a necessidade de informações das dimensões que compõem o turismo, tanto no âmbito da prestação de serviços (hotéis, agências, restaurantes, atrativos turísticos, etc.), mas também referente ao meio ambiente, sociedade civil, iniciativa privada, entidades de classe e poder público, Ramos (2010, p.108) afirma que, no turismo, a informação é o "combustível da economia".

Segundo dados da Organização Mundial do Turismo (OMT)<sup>2</sup>, a atividade turística será responsável por movimentar, diretamente, 2 trilhões de dólares no ano de 2020 com um número total de pessoas circulando em 1,6 bilhões, fazendo com que o turismo se torne o principal difusor de atividades econômicas no mundo. De acordo com números do *World Travel and Tourism Council* (2016), em 2015 o turismo foi responsável por 9,8% do Produto Interno Bruto mundial, estimando um crescimento de 3,3% para 2016. Além disso, o turismo foi responsável pelo emprego direto de 107 milhões de pessoas em 2015, e estima-se um crescimento de 1,9% para 2016, se tornando uma das atividades de maior impacto financeiro em âmbito global.

Mesmo tendo essa grande importância econômica, o setor de turismo ainda carece de melhorias na gestão da informação para tomada de decisões estratégicas, principalmente em âmbito nacional. Em pesquisa realizada pelo Ministério do Turismo (MTUR) que avalia o índice de competitividade dos 65 destinos turísticos mais importantes do Brasil, mostrou que em 2014, das 13 dimensões avaliadas para a composição do índice<sup>3</sup>, o monitoramento de informações realizado

---

<sup>2</sup> The Globalization of Tourism: [http://www.unesco.org/courier/1999\\_08/uk/dossier/txt13.htm](http://www.unesco.org/courier/1999_08/uk/dossier/txt13.htm)  
Acesso em 03/10/2014.

<sup>3</sup> O MTUR definiu no Plano Nacional de Turismo 2007-2010, 65 destinos que representassem a diversidade da oferta turística nacional, denominando-os "destinos indutores", visando monitorar a competitividade desses territórios em relação ao mercado turístico ao longo do tempo através da elaboração de um índice com 13 dimensões, sendo: infraestrutura geral, serviços e equipamentos turísticos, acesso, atrativos turísticos, marketing e promoção do destino, políticas públicas, cooperação regional, monitoramento, economia local, capacidade empresarial, aspectos sociais, aspectos ambientais e por fim, aspectos culturais (BRASIL, 2014).

por esses destinos foi a dimensão com a menor nota (36,2) em uma escala de 01 a 100, sendo que a média de todas as dimensões foi de 59,5.

Sabe-se que grande parte das informações características do fenômeno *Big Data* é disponibilizada de forma dispersa na internet, dificultando sua captura, tratamento e análise para tomada de decisões por gestores de organizações e empresas. Muitas dessas informações são disponibilizadas por diferentes usuários em redes sociais a partir de avaliações e comentários sobre produtos e serviços com o intuito de auxiliar demais usuários da *web* na solução de problemas.

Um dos exemplos pode ser caracterizado no setor de turismo pelo site *TripAdvisor*. A plataforma se caracteriza como a maior rede social para troca de informações sobre destinos turísticos no mundo e é alimentada pelos próprios usuários a partir da avaliação quantitativa e qualitativa de produtos e serviços turísticos, facilitando a escolha dos destinos e o planejamento de viagem dos turistas. De acordo com informações do próprio site ([www.tripadvisor.com.br](http://www.tripadvisor.com.br)), as páginas de viagens do *TripAdvisor* atingiram em fevereiro de 2016, 340 milhões de visitantes únicos por mês, com 350 milhões de avaliações e opiniões cadastradas de mais de 6,5 milhões de meios de hospedagem, restaurantes e atrativos de 136 mil destinos, tornando-se um grande banco de dados com informações relevantes para elaboração de estratégias de desenvolvimento do setor, principalmente do setor público.

Para a realização de pesquisas e levantamento de informações no setor público, são necessários recursos que na maioria das vezes, encontram-se escassos. De acordo com o Portal da Transparência do Governo de Minas Gerais, no ano de 2015, a Secretaria de Estado de Turismo de Minas Gerais (SETUR-MG) foi responsável por apenas 0,02% das despesas do Estado (R\$ 16 milhões ante os R\$ 80,79 bilhões), no qual mais da metade dessas despesas foram destinadas a custos relacionados à própria administração da Secretaria (tais como pagamento de vencimento dos servidores, diárias, materiais de consumo), sendo o recurso destinado à execução de projetos insuficiente para a elaboração de pesquisas de grande porte.

Uma das pesquisas de maior relevância para traçar o perfil do visitante em Minas Gerais, assim como o seu grau de satisfação com os produtos e serviços turísticos do estado é realizada pela Secretaria de Estado de Turismo de Minas Gerais (SETUR-MG). Porém, por se tratar de uma pesquisa aplicada a partir da contratação de pesquisadores que devem visitar diversos municípios *in loco*, a qualidade e o

número de questionários aplicados a cada ano dependem diretamente de recursos orçamentários que, quando não disponíveis, acabam acarretando na descontinuidade da série histórica e na dificuldade de mensuração de indicadores de desempenho do turismo no estado (MORAIS, OLIVEIRA, PAIM. 2015).

A utilização de resultados de pesquisas possibilita avaliar os resultados de programas e projetos do setor público, aprimorando-os, com objetivo de responder às necessidades dos beneficiários dessas ações (SERAPIONI, 2016). Esse monitoramento deve ser feito de forma regular, tempestiva e sistêmica pelo Estado, e uma das formas de realizar esse monitoramento é a partir da construção de indicadores que são “úteis ao gerenciamento tempestivo das ações de governo” (CARDOSO JUNIOR, 2015, p.12), facilitando a tomada de decisões.

Assim, buscar alternativas de baixo custo para coletar dados que auxiliem monitorar o impacto de ações realizadas no setor de turismo se torna fundamental.

Dentro desse quadro, surge a seguinte questão de pesquisa: utilizar informações extraídas a partir de sites de avaliação de viagens pode auxiliar na elaboração de indicadores de desempenho do turismo em Minas Gerais?

Esse trabalho checkou a viabilidade de elaborar indicadores que possam ser utilizados no monitoramento de informações do turismo em Minas Gerais, a partir da extração de dados em plataformas *web*. Para tanto, utilizou-se uma ferramenta de extração de dados que possibilitou a análise de informações de forma clara, visando atingir os objetivos propostos e respondendo assim, a pergunta em questão.

Essa pesquisa tem como objetivos específicos:

- Facilitar o acesso a técnicas de extração de dados a partir da utilização de uma ferramenta gratuita e acessível que não exige conhecimentos de programação;
- Gerar indicadores de monitoramento para o turismo a partir da análise dos pontos fortes e fracos de cada atrativo do estado oriundos dos dados extraídos;

- Contribuir para melhoria da tomada de decisão dos gestores públicos e dos responsáveis pelos atrativos privados pesquisados;

O monitoramento e a avaliação do comportamento dos usuários em plataformas *web*, a partir da realização de um estudo de caso instrumental (exemplificado pelo site *TripAdvisor*), possa auxiliar na elaboração de indicadores que contribuam para a criação de estratégias de melhoria do atendimento e da satisfação dos visitantes em atrativos e destinos turísticos de Minas Gerais.

De forma aplicada, esse estudo servirá de base para que gestores do turismo tratem dados característicos de *Big Data* gerando insumos que possibilitem a criação de indicadores para o monitoramento da atividade turística. Os indicadores contribuirão para a criação de estratégias mais efetivas que visem à melhoria do atendimento e da satisfação dos visitantes em atrativos e destinos turísticos de Minas Gerais. A utilização de uma ferramenta para extração de dados e posterior utilização das informações contribui na elaboração de novas metodologias de monitoramento e novas fontes de informação para o turismo.

Essa dissertação está dividida em 8 capítulos. Além do primeiro capítulo já apresentado e focado na contextualização da pesquisa, o restante do trabalho será dividido em mais sete capítulos.

O segundo capítulo apresenta a fundamentação teórica onde são abordados os temas que serão explorados de forma a elucidar a discussão proposta, focados nas temáticas referentes ao fenômeno conhecido como *Big Data*, desde sua definição até a ilustração de ferramentas de extração de informações. Discute-se adiante uma contextualização do setor de turismo e aborda-se uma discussão sobre o turista na era digital e a importância do compartilhamento de informações realizados através de plataformas online, exemplificada pelo site denominado *TripAdvisor*. Por fim, mostra a importância da criação de indicadores para o setor de turismo.

O terceiro capítulo foca na apresentação da metodologia do trabalho abordando o passo-a-passo para a realização da pesquisa, desde a escolha e aplicação da ferramenta de extração de dados a organização do banco de dados.

O quarto capítulo é reservado para a apresentação e a análise crítica das informações coletadas tais como a oferta de atrativos, avaliações dos usuários, perfil e época de visitação.

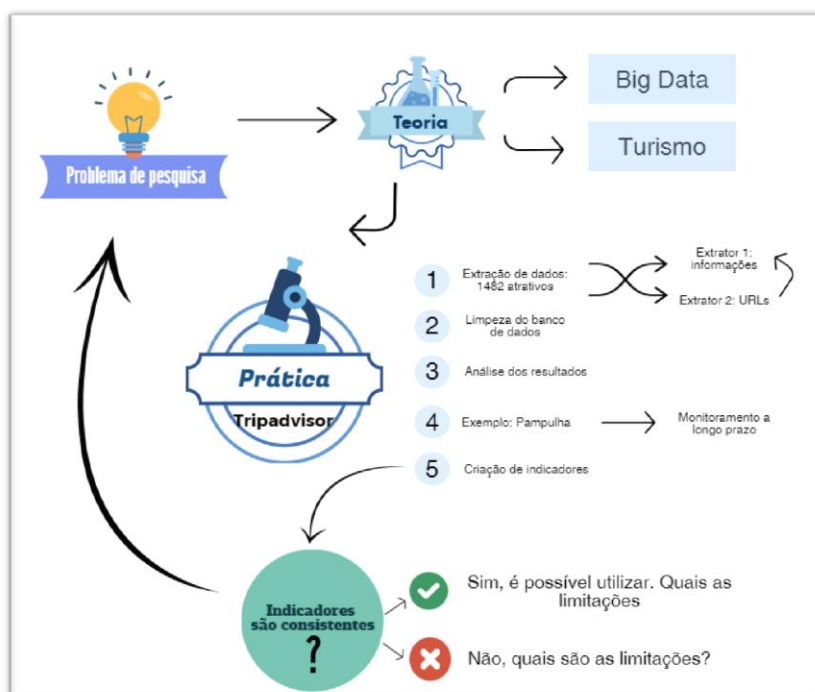
O quinto capítulo apresenta, de forma particular, os resultados obtidos pelo atrativo “Conjunto Arquitetônico da Pampulha” que foi reconhecido como patrimônio mundial da humanidade durante a realização da pesquisa. O objetivo é mostrar o impacto causado por esse reconhecimento no nível de satisfação dos visitantes ao longo do tempo.

Após a análise, propõe-se no sexto capítulo a criação de 4 indicadores para o monitoramento dos dados levantados a partir da metodologia apresentada.

O sétimo capítulo destina-se às considerações futuras observadas durante a realização da pesquisa, visando ampliar a discussão sobre o acesso à informação contida no meio digital para o setor público.

O último capítulo apresenta as principais conclusões e possíveis propostas de continuidade de pesquisa, esperando contribuir para o debate de temas correlatos. A figura 1 objetiva visualizar a metodologia de forma mais clara.

Figura 1: Principais etapas da pesquisa



Fonte: Elaboração do autor

## 2 Fundamentação teórica

Para enriquecer as discussões apresentadas, é essencial compreender alguns conceitos básicos e conteúdos chave que norteiam esse trabalho.

Pode-se dizer que a informação é um insumo presente em qualquer área do conhecimento, devido a sua ampla definição e utilização para transmissão de conhecimento. Mas, é a Ciência da Informação que se preocupa de forma primária, “em estudar os processos de representação do conhecimento e de seus registros associados – a informação” (SOUZA; ALMEIDA; BARACHO, p.161, 2015).

Considera-se que a Ciência da Informação é um campo de conhecimento interdisciplinar, principalmente focado na organização da informação e na utilização de sistemas computacionais que possibilita interagir com diferentes áreas (TARAPANOFF, 2006; SARACEVIC, 2009), visando resolver problemas informacionais de diferentes naturezas. Essa inter-relação contribuiu diretamente para a evolução da atual sociedade da informação, caracterizada pela grande explosão informacional a partir do final da 2ª Guerra Mundial até os dias de hoje, fazendo com que a informação esteja arquivada em formatos muito distintos em âmbito universal, principalmente após o advento da *internet* (SARACEVIC, 2009), ditando novos comportamentos sociais.

Estudar o fenômeno das novas formas de captura, análise e disseminação das informações, principalmente por meios digitais, torna-se de grande importância para a Ciência da Informação, pois seu impacto na área acadêmica é enorme (SOUZA; ALMEIDA; BARACHO, 2015), visto que os novos meios de comunicação ampliaram sistemicamente o volume, a variedade e a velocidade das informações, gerando no fenômeno conhecido como *Big Data*.

### 2.1 Big Data

A seguir, serão apresentados os principais conceitos referentes ao termo *Big Data* que serão fundamentais para o entendimento da contextualização da pesquisa e visa ilustrar as discussões mais relevantes sobre o tema.

### 2.1.1 Definição de *Big Data*

A utilização do termo “*Big Data*” foi observada aproximadamente no início da década de 1990, descrita pela *Association for Computing Machinery* simplesmente como um conjunto de dados muito grande que não poderia ser analisado por apenas um computador e que apenas nos últimos cinco anos, começou a ganhar popularidade (CRAWFORD; MITNER; GRAY, 2014), ampliando o seu conceito além do volume de informações, defendido de forma ampla por diversos autores.

De acordo com Kitchin (2014), um conceito baseado apenas no grande volume de informações é limitado, visto que esse fenômeno já ocorria em décadas anteriores como, por exemplo, os dados censitários nacionais. Concordando com esse pensamento, Boyd e Crawford (2012) afirmam que:

De fato, alguns dados englobados pela *Big Data* (ex: todas as mensagens de *Twitter* referentes a um tema específico) não são maiores do que bancos de dados anteriores que nunca foram considerados como *Big Data* (ex: CENSUS). *Big Data* refere-se menos sobre grande volume de dados e mais sobre a capacidade de buscar, agregar e combinar grandes bancos de dados (BOYD; CRAWFORD, 2012, p.663, tradução nossa).

McAfee e Brynjolfsson (2012) elencam três propriedades que são observadas apenas no conjunto de informações denominado como *Big Data*, denominadas como os “3 Vs”:

- Volume: Tem-se hoje, maior fluxo de informação na internet por segundo do que era armazenado em toda a web há vinte anos. Isso dá a oportunidade de empresas trabalharem com uma enorme quantidade de dados em apenas um ambiente;
- Velocidade: em muitos casos, a velocidade de processamento dos dados é mais relevante que a quantidade de dados disponíveis. Assim, *Big Data* possui a característica de trabalhar com as informações em curto espaço de tempo e principalmente em tempo real, garantindo competitividade para as empresas que a utilizam;

- Variedade: a *Big Data* assume diferentes formatos de dados, desde imagens, sinais de GPS, mensagens em redes sociais, dentre outros, que são processadas de forma conjunta.

Alguns autores ainda acrescentam outra propriedade (ou um quarto “V”) para *Big Data*. Han e Lu (2014) definem o “4º V” com a “veracidade” que reflete se os dados utilizados na análise comparativa estão de acordo com as características encontradas nos dados brutos. Já para a empresa Oracle (2015), o “4º V” seria dedicado ao “valor”, ou seja, é necessária uma análise dos dados de forma coerente para que a informação seja repassada de acordo com as expectativas e de forma a auxiliar na resolução de problemas.

Para lidar com todas essas informações de diferentes dinâmicas e formatos, a organização de dados se torna essencial para análise de informações caracterizadas como *Big Data*. De acordo com Pimenta (2013), para controlar os “Vs” é necessário escolher quais as informações devem ser descartadas e para isso é fundamental a gestão da informação e do conhecimento.

Grande parte das informações produzidas hoje é em formato digital, alimentada por diversos usuários, coletada e armazenada por computadores, sendo possível de ser organizada e extraída apenas por ferramentas computacionais (PUSCHMANN; BURGESS, 2014). Elas acabam dando significado aos bancos de dados que não conseguiriam ser compreendidos por um indivíduo ou um grupo de indivíduos. Portanto, criam-se ferramentas de “análise de *Big Data*”. De acordo com Andrejevic (2014, p.3, tradução nossa) “*Big Data* denota o momento onde formas automatizadas de reconhecimento de padrões conhecidas como “análise de dados” se encontram com formas automatizadas de coleta e armazenamento de dados”. O autor, conclui que a partir de então, os dados podem ser coletados e correlacionados como nunca haviam sido antes, destacando padrões de comportamento “muito além do que a mente humana é capaz de detectar ou até explicar” (ANDREJEVIC, 2014, p.4, tradução nossa).

Percebe-se que as conceituações de *Big Data* defendida pelos autores apresentados permeiam não só os campos da informação e da computação, mas também abrem caminho para discussões sobre novos paradigmas na sociedade. Uma das definições que consegue consolidar esses pensamentos é destacada a seguir:

Nós definimos *Big Data* como um fenômeno cultural, tecnológico e acadêmico que repousa sobre a interação de: (1) Tecnologia: maximizando o poder computacional e a eficiência algoritma para reunir, analisar, relacionar e comparar bancos de dados volumosos. (2) Análise: lidar com grandes bancos de dados para identificar padrões com vistas a realizar reivindicações sociais, técnicas e legais. (3) Mitologia: a crença generalizada que conjuntos de vários dados oferecem uma alta forma de inteligência e conhecimento que podem gerar “insights” que anteriormente eram impossíveis, com a aura da verdade, objetividade e assertividade (BOYD; CRAWFORD, 2012, p.663, tradução nossa).

Essa definição apresenta-se como a mais indicada para a compressão do conceito e que norteará as demais discussões durante esse trabalho. Compreender *Big Data* não só como um problema computacional, mas entender que suas consequências podem de fato criar novos paradigmas para a sociedade é um ponto crucial que deve ser debatido com maior aprofundamento por diversas áreas do conhecimento.

### **2.1.2 Big Data e a sociedade**

Entende-se que a partir do conceito apresentado, a noção do termo *Big Data* discutida por diversos especialistas na área acadêmica tende a gerar grandes consequências em diversas áreas. Essa ideia pode ser apresentada da seguinte forma:

*Big Data* fez emergir um sistema de conhecimento que (...) tem o poder de informar como nós entendemos as redes humanas e sociais. (...) *Big Data* reformula questões-chave sobre a constituição do conhecimento, o processo de pesquisa, como devemos nos engajar com a informação, e a natureza e a categorização da realidade (BOYD, CRAWFORD, 2012, p.665, tradução nossa).

A partir desse princípio, *Big Data* se torna uma visão de mundo, enraizado em múltiplos campos nos setores público e privado, necessitando de críticas mais profundas e engajadas, tornando-se não só uma ferramenta de análise, mas sim como uma emergente teoria do conhecimento (CRAWFORD; MILTNER; GRAY, 2014).

Ao mesmo tempo em que o fenômeno *Big Data* pode auxiliar em inúmeras questões da sociedade em diversas áreas tais como saúde, segurança ou meio ambiente (BOYD; CRAWFORD, 2012), alguns dilemas relacionados a ela devem ser considerados durante sua utilização.

Para alguns autores (BOYD; CRAWFORD, 2012; ANDREJEVIC, 2014), quem tiver acesso à extração de dados de *Big Data* poderá ter mais informação e poder de decisão do que as pessoas que não possuem esse acesso, podendo gerar padrões de comportamento e opiniões para a classe menos favorecida. Esse ecossistema gerado pela *Big Data* poderá criar então uma “divisão digital”, conforme defendido por Boyd e Crawford (2012, p.665). Por isso, é necessário que as ferramentas de análise de dados em ambiente *Big Data* possam ser utilizadas de forma mais ampla e mais acessível a um maior número de pessoas e com vistas a gerar mais transparência, principalmente nas políticas públicas como direito de todos os cidadãos e auxiliando as entidades privadas a serem mais sustentáveis (ANDREJEVIC, 2014).

Outra discussão trazida pelo tema baseia-se na ideia de que uma coleta de grande conjunto de dados e a análise dos mesmos possa trazer uma preocupação com a privacidade das informações e indivíduos (TENE; POLONETSKY, 2012), o que pode acarretar em perdas de liberdades e aumentar o controle do estado ou de corporações (BOYD; CRAWFORD, 2012; PIMENTA, 2013).

Grande parte dessas informações é emitida de forma “passiva” por usuários que, de fato, estão transmitindo essas informações através de geolocalização, análise de preferências pela internet, dentre outros, sem o conhecimento explícito dos mesmos. Para Pimenta (2013, p.8) “tornamo-nos “sensores humanos” graças ao uso de aparelhos inteligentes que, conectados à “nuvem”, nos tornam emissores de dados em tempo real com uma velocidade inequívoca”.

Esse contexto faz com que para o “*projeto Big Data*” obter sucesso, os cidadãos devem ter a consciência da importância das informações que eles emitem pois essas informações poderão auxiliar toda a sociedade. “Contribuir com dados individuais para o conjunto é contribuir com o avanço da ciência, inovação e aprendizado” (CRAWFORD; MILTNER; GRAY, 2014, p.4, tradução nossa). Corroborando com esse pensamento, Pimenta (2013) defende que:

(...) compreender os dados, saber usá-los e controlá-los é o maior desafio do *Big Data*. Para vencer esse desafio é necessário ter acesso ao mundo dos dados e para ter acesso, é necessário permissão. Os dados que emitidos por cada momento em meios tecnológicos influenciam diretamente na identidade individual e

consequentemente coletiva da sociedade, criando novas tendências e comportamentos. A memória coletiva, então, não nos pertence completamente (PIMENTA, 2013, p.4, tradução nossa).

Além disso, Andrejevic (2014) afirma que a utilização dos dados a partir do fenômeno *Big Data* deve ser feita de forma coletiva e não individual, através da análise de padrões de comportamento de uma sociedade, a partir do contexto nos quais os indivíduos se inserem. A partir dessa lógica, a individualidade não ficará ao todo comprometida.

No caso de estudos acadêmicos, deve-se verificar com cuidado como as ferramentas de *Big Data* serão utilizadas, principalmente na extração e análise dessas informações, garantindo a ética dos estudos apresentados (CRAWFORD; MILTNER; GRAY, 2014).

Todo esse contexto faz com que a análise de *Big Data* seja fundamental para compreendermos a sociedade na qual vivemos; suas tendências e comportamentos. Para se adaptarem a essa realidade, as organizações veem a coleta e análise de dados de *Big Data* como um elemento essencial para a melhoria dos produtos e serviços, garantindo competitividade. Para o setor público, possibilita monitorar ações que possam beneficiar a vida de cada indivíduo.

### **2.1.3 *Big Data* nas organizações**

O fenômeno *Big Data* possibilitou aos gerentes e tomadores de decisões conhecerem melhor seus ambientes de negócios, além de transformar o conhecimento adquirido em aumento de desempenho e estratégias mais precisas (MCAFEE; BRYNJOLFSSON, 2012). Complementando esse pensamento, “*Big Data* é capaz de trazer *insights* importantíssimos para a empresa, mas a decisão e visão do executivo sempre será essencial para o negócio, porém agora, muito mais embasada em dados”. (NOVO; NEVES, 2013, p.34).

No mundo dos negócios, o termo *Big Data* foi uma forma de se opor ao tradicional modo de levantamento de dados, sendo relacionado praticamente a uma “marca comercial”. *Big Data* marcou uma mudança significativa entre os sistemas de gerenciamento de bancos de dados relacionais para plataformas que ofereciam vantagens de desempenho em longo prazo, ao contrário das soluções que eram encontradas no mercado (PUSCHMANN; BRUGESS, 2014).

A importância da utilização de *Big Data* nas organizações pode ser ilustrada por uma pesquisa realizada por McAfee e Brynjolfsson (2012), na qual concluíram que as empresas que estão entre as três principais de seu setor produtivo que utilizam informações de banco de dados para auxiliar na tomada de decisões eram, em média, 5% mais produtivas e 6% mais lucrativas do que os concorrentes. Assim, os autores afirmam que os grandes líderes perceberão em pouco tempo que a *Big Data* pode ser considerada como uma “revolução gerencial”. Grande parte dessa revolução pode ser explicada pelo custo-benefício de se possuir ferramentas de análises de dados que auxiliem de forma precisa a elaborar estratégias para atingir públicos-alvo ou nichos de mercado (BOYD; CRAWFORD, 2012). Algumas ferramentas podem ser encontradas gratuitamente na internet, garantindo maior acessibilidade às informações e utilização por parte das organizações.

Porém, os dados levantados por essas soluções não conseguem interpretar resultados por si só. É fundamental o trabalho de um analista de dados ou especialista na área de conhecimento do problema para que os resultados possam ser compreendidos a partir de um contexto (KITCHIN, 2014). Concordando com essa visão Boyd e Crawford (2012) afirmam que os bancos de dados não são entidades neutras ou livres de teorias esperando para divulgar seus segredos. Para isso, é fundamental a interpretação de pesquisadores que também possuem seus próprios pontos de vista. Mesmo com essa necessidade, as organizações ainda possuem um déficit de profissionais hábeis para analisar e dar sentido aos dados levantados. (KITCHIN, 2014; LAROSE; LAROSE, 2015).

Não só os resultados devem ser analisados de forma criteriosa, mas é também essencial fazer as perguntas certas das quais as organizações e seus gestores querem extrair as informações. McAfee e Brynjolfsson (2012), afirmam que os tomadores de decisões das empresas atualmente se espelham muito na experiência pessoal e na intuição para direcionar as estratégias a serem seguidas, mas com as possibilidades geradas pela *Big Data*, só a experiência não é suficiente: é necessário moldar as decisões a partir do levantamento sistemático de dados, fundamentados em questões elaboradas de forma criteriosa.

Além de elaborar perguntas e analisar os dados de forma coerente, McAfee e Brynjolfsson (2012), elencam cinco desafios que devem ser enfrentados pelas organizações que pretender lidar com as informações oriundas da *Big Data*:

- Liderança: para obter sucesso, as organizações e o público devem contar com grupos de lideranças que saibam exatamente quais os objetivos e metas a serem seguidos, além de fazerem as perguntas corretas. O poder da *Big Data* é indissociável aos *insights* humanos.
- Gerenciamento de talentos: mais importante do que contar com equipes de estatística e análise de dados, é necessário investir em profissionais que possuam habilidades de limpar e organizar banco de dados, criar interfaces de visualização e que consigam aliar as linguagens de negócios à análise de dados.
- Tecnologia: grande parte das tecnologias que lidam com *Big Data* possuem códigos abertos e são de fácil customização e implantação. Porém, ainda são tecnologias novas e que necessitam de bom conhecimento das áreas de tecnologia de informação para conseguir integrar as fontes de dados internas e externas da organização.
- Tomada de decisão: deve-se criar organizações flexíveis que possam maximizar a cooperação entre os especialistas de funções diferentes. As pessoas que entendem os problemas devem andar junto com os profissionais que conseguem trabalhar a favor da tecnologia de extração de dados, auxiliando na tomada de decisão.
- Cultura organizacional: as organizações devem se perguntar menos “o que pensamos?” e perguntar mais “o que nós sabemos?”, fazendo quebrar uma cultura de atuar apenas pelo instinto e utilizando as informações de fato para gerar estratégias e não apenas para justificar aquilo que já se sabe.

Em resumo, para as organizações “o grande desafio de lidar com *Big Data* é essencialmente humano, que é fazer as perguntas corretas, reconhecer padrões, fazer suposições e prever o comportamento baseados em dados” (ORACLE, 2015, p.4, tradução nossa). Portanto, cabe às organizações utilizarem a experiência das equipes e a cooperação entre áreas do conhecimento amparadas em dispositivos tecnológicos para solucionar problemas e buscar oportunidades de desenvolvimento.

Dentre as soluções tecnológicas disponíveis para auxiliar as organizações, *Big Data* popularizou duas ferramentas fundamentais para trabalhar com as informações sendo a primeira ligada ao armazenamento dos dados denominada de *Apache Hadoop* e a segunda ligada à tecnologia de processamento de dados denominada de *NoSQL database* (ORACLE, 2015). Para realizar essas duas tarefas de forma eficiente, é essencial utilizar técnicas de extração de dados, principalmente em ambiente *web*, com vistas a facilitar a organização das informações que elucidem a tomada de decisão.

#### **2.1.4 Extração de dados web**

Em tempos de *Big Data*, a extração e a organização de dados torna-se crucial para a sobrevivência das organizações. Essa relevância fez com que surgisse a “ciência de dados” (ou apenas *Data Science*) como campo de pesquisa. *Data Science* consiste em um conjunto de princípios fundamentais que dão suporte e orientam o princípio da extração de informações e conhecimento a partir de dados:

*Data Science* envolve princípios, processo e técnicas para compreender fenômenos através da análise (automatizada) de dados. (...) O objetivo final da *Data Science* é melhorar a tomada de decisões e portanto, sendo de interesse primordial para as empresas (PROVOST; FAWCETT, p.53, 2013)

Para auxiliar na tomada de decisões, a *Data Science* deve lidar diretamente com os processos de extrair, preparar, analisar, visualizar, gerenciar e arquivar grandes volumes de dados (STANTON, 2013). Então, para que todo o processo se torne bem executado, a extração das informações deve ser realizada da melhor maneira possível, pois afetará todas as demais etapas do processo, e conseqüentemente na tomada final de decisão.

Uma forma de coleta de dados é extrair informações constantes na *web*, principalmente a partir do acúmulo de informações divulgadas em redes sociais e conteúdo pela internet (CHEN; CHIANG; STOREY, 2012) que podem auxiliar os gestores na tomada de decisão. Para Devika e Surendran (2013) muitas aplicações no meio dos negócios dependem da *web* para coleta de informações que são fundamentais durante os processos de tomada de decisão. A extração de dados em formato digital, principalmente de *websites*, se tornou uma das técnicas mais utilizadas em *Big Data* (MARRES; WELTEVREDE, 2012).

*Websites* são criados, na maioria das vezes, para auxiliar na visualização de informações e não para exposição de dados de forma estruturada. Extrair essas informações de forma manual pode consumir muito tempo e ser suscetível a erros de extração. Vargiu e Urru (2013) afirmam que o método manual mais efetivo para extrair informações da *web* ainda é o conhecido “copiar-colar”, na qual o próprio usuário deve copiar os dados de uma determinada página, um a um, e depois inseri-los em um banco de dados único. Porém, os autores afirmam que apesar de esse método ainda ser a única saída para exportar informações de páginas na *web*, na prática não se torna factível, especialmente para grandes organizações e projetos, sendo seu custo muito alto. Nesse contexto, a extração das informações de forma automatizada possui um papel fundamental para a análise dos dados expostos. (DEVIKA; SURENDRAN, 2013).

As páginas de um mesmo *website* geralmente possuem *templates* (ou *layouts*) semelhantes entre elas e que podem expor dados de uma forma padronizada, o que facilita a utilização de ferramentas de extração de dados *web*. De acordo com Devika e Surendran (2013, p.278, tradução nossa): “A construção de uma página é o processo em transformar dados em *templates*. A extração de dados *web* é o processo inverso da criação da página”. Ou seja, tenta-se criar um banco de dados a partir das informações contidas em cada uma das páginas, sendo que o primeiro passo para a alimentação das informações dessas páginas foi criar um banco de dados que geralmente o usuário comum não tem acesso.

Uma das possibilidades de extrair os dados dos *websites* é utilizar ferramentas denominadas *Web Scrapers* (ou apenas *Scraping*). De acordo com Vargiu e Urru (2013, p.44, tradução nossa) um *web scraper* “foca em transformar dados não-estruturados da *web*, tipicamente em formato HTML, em dados estruturados que possam ser arquivados e analisados em bancos de dados locais (...)”. Já para Marres e Weltevrede (2012), além de ser uma técnica para coletar informações relevantes de forma automatizada de dados *online*, tem a característica marcante de possibilitar estudos em tempo real de diferentes situações sociais, o que aumentou muito a sua utilização para áreas do conhecimento que inicialmente não estavam ligadas à programação, como o caso das ciências sociais. O processo de *scraping* é composto por:

(...) uma série de passos nos quais dados estruturados são extraídos de uma desordem informacional. Para isso, é necessário criar um

processo para transformar a informação uniforme de dados *online* em informações estruturadas e dentro desse processo, os dados textuais brutos são progressivamente despojados de seus elementos sem relevância e formatados para produzir de forma bem ordenada um conjunto de dados utilizáveis (MARRES; WELTEVREDE, 2012, p.8, tradução nossa).

Para as mesmas autoras, o processo de *scraping* tornou-se essencial para a elaboração de pesquisas relacionadas ao comportamento social e cada vez mais são utilizadas no campo das ciências sociais. Complementando esse pensamento, elas afirmam que:

(...) *scraping* faz nada menos do que destravar o “potencial sociológico” da internet: ele torna viável para as pesquisas sociais trabalhar com grandes quantidades de dados gerados pelo próprio usuário, que ultimamente estão se acumulando em plataformas online como por exemplo o *Facebook*, *Twitter*, *Wikipedia*, além de outros (MARRES; WELTEVREDE, 2012, p.10, tradução nossa).

A extração de dados *web* possibilita então, analisar as informações em tempo real produzidas e alimentadas em plataformas digitais pelos próprios usuários e, compreender o comportamento desses usuários pode auxiliar diretamente a responder diversos questionamentos e problemas enfrentados diariamente por pesquisadores e organizações de diferentes setores, assim como o turismo.

## **2.2 Turismo**

Esse tópico é destinado a elucidar algumas questões relacionadas sobre o conceito de “turismo”, seu contexto em relação à era digital a partir do exemplo do site *TripAdvisor* e a importância do levantamento de indicadores para a atividade.

### **2.2.1 O turismo na era digital**

A definição de maior consenso por estudiosos para o termo “turismo” foi formulada pela Organização Mundial de Turismo (OMT). Segundo a entidade, o turismo é definido como:

(...) as atividades que as pessoas realizam durante suas viagens e permanência em lugares distintos dos que vivem, por um período de

tempo inferior a um ano consecutivo, com fins de lazer, negócios e outros (OMT, 1995, p.1, tradução nossa).

Deve-se lembrar que o turismo sempre foi visto como uma atividade de realização social. Assim como existiam os cartões-postais e cartas para troca de informações sobre viagens, atualmente as mídias digitais e as redes sociais contribuem para transformar as formas de compartilhamento de informações entre os viajantes (MUNAR; JACOBSEN, 2013). De acordo com Xiang, Magnini e Fesenmaier (2015) desde os anos 90, o setor de viagens adotou a *internet* como um novo canal de distribuição de serviços e de propaganda e com o passar do tempo, a relação entre os prestadores de serviço e seus clientes foi se aperfeiçoando até as redes sociais. Essas redes possibilitaram não só a interação entre esses atores, mas também a troca de opiniões *online* entre indivíduos, despertando atenção por parte das empresas e organizações.

A troca de informações entre usuários entre redes digitais possibilitou a criação do termo “eWOW” (*eletronic word of mouth*), ou “boca-a-boca digital”, na qual a informação é repassada para cada indivíduo sem intermediação, na qual as recomendações de pessoas conhecidas ou não acabam influenciando na escolha dos destinos de viagem dos usuários dessas redes (GRETZEL; YOO, 2008). As autoras ainda afirmam que “como o uso da *internet* para o planejamento de viagem se tornou mais relevante, espera-se que os processos de decisão sejam cada vez mais influenciados pelo eWOW” (GRETZEL; YOO, p.36, 2008, tradução nossa).

Esse contexto fez com que a busca de informações para o planejamento de uma viagem se tornasse uma parte importante da experiência do turista, na qual ele necessita do maior número de informações possíveis para auxiliar na tomada de decisão (XIANG; MAGNINI; FESENMAIER, 2015), fazendo com que sites de avaliação de viagens se tornem uma importante ferramenta de planejamento. Esses sites “provêm ideias, facilitam as decisões, acrescentam diversão no processo de planejamento e aumentam a confiança ao reduzir riscos, fazendo com que seja mais fácil imaginar como o destino deve ser” (GRETZEL; YOO, p.44, 2008, tradução nossa).

Nos sites de avaliação os consumidores passam a avaliar os perfis dos destinos e das empresas a partir da distribuição de notas e comentários, fazendo com que as organizações incentivem os seus clientes a avaliá-los nas plataformas *online*.

Com isso, o eWOW ocorre de forma mais efetiva, gerando mais clientes para os estabelecimentos (GRETZEL; YOO, 2008), aumentando a competitividade dos destinos turísticos na disputa pelas indicações de viagem.

### **2.2.2 Avaliação de viagens em ambiente web: o exemplo do site *TripAdvisor***

A informação é um insumo essencial para qualquer atividade, inclusive o turismo. Sem ela, o turista não teria como se localizar ou saber o que fazer dentro de um determinado destino, por exemplo. De acordo com Cacho e De Azevedo (2010) a evolução nos sistemas de informação possibilitou diretamente não só a sustentabilidade do turismo, mas também sua expansão. Os mesmos autores afirmam que a tecnologia possibilitou o aumento das fontes de informação para o turismo:

A crescente demanda turística vem impulsionando a ampliação das fontes de informação às quais o potencial turista possa vir a ter acesso. Neste contexto, torna-se cada vez mais urgente o uso das novas fontes de informação como veículo difusor dos serviços e informações turísticas (CACHO; DE AZEVEDO, p.43, 2010).

Esses fatos fizeram com que vários sites de informações para o turista surgissem, dentre eles o site *TripAdvisor*. Considerado o maior site de compartilhamento de informações de viagens do mundo, o *TripAdvisor* foi fundado no ano 2000 e desde então, se tornou uma corporação que através de suas subsidiárias, administra e opera sites sob domínio de outras 24 empresas de viagens online, empregando 3100 pessoas em março de 2016<sup>4</sup>.

De acordo com informações do próprio site ([www.tripadvisor.com.br](http://www.tripadvisor.com.br)), as páginas de viagens do *TripAdvisor* atingiram em fevereiro de 2016, 340 milhões de visitantes únicos por mês, com 350 milhões de avaliações e opiniões cadastradas de mais de 6,5 milhões de meios de hospedagem, restaurantes e atrativos de 136 mil destinos. O site atua em 48 países e está disponível em 28 idiomas.

Para Munar e Jacobsen (2013), o *TripAdvisor* e os demais sites de avaliações de viagens se beneficiam dos conteúdos gerados pelos turistas a partir da necessidade dos mesmos em divulgar informações para uma comunidade, explorando

---

<sup>4</sup> Fonte: *TripAdvisor* – [www.tripadvisor.com.br](http://www.tripadvisor.com.br) . Acesso em 9 de mai. 2016.

esses conteúdos de forma comercial, sendo que esse tipo de mídia é considerado como a plataforma de rede social mais confiável perante os usuários na internet.

Dentre os recursos oferecidos pelo site, os usuários podem comparar preços e reservar hotéis, casas de temporada, voos e passeios de forma online, receber recomendações de destinos e meios de hospedagem a partir do perfil do usuário, publicar fotos de viagens, interagir através de fórum de perguntas e dúvidas para cada destino, buscar locais e atrações a partir de interação com mapas virtuais e principalmente avaliar os produtos e destinos turísticos a partir da distribuição de notas e comentários abertos.

Para Torres, Morales e Jiménez (2013), os comentários e opiniões dos usuários do *TripAdvisor* possuem não só componentes quantitativos (posições em rankings, avaliações, número de comentários...) mas também contam com informações qualitativas, possibilitando a um proprietário de um estabelecimento, por exemplo, identificar o que os seus clientes estão dizendo sobre seu estabelecimento e em caso negativo, poder corrigir falhas de forma rápida e precisa. O autor também afirma que:

Processar as informações e opiniões dos clientes de forma correta possibilitará as empresas a melhorar os processos de produção, o que é uma grande valia para as empresas que sabem que ter todas as informações dos consumidores na ponta dos dedos e de forma gratuita é vantajoso (TORRES; MORALES; JIMÉNEZ, 2013, p. 23, tradução nossa).

Limberger (et al, 2014, p.62), afirma que é difícil categorizar o *TripAdvisor* como uma rede social, uma comunidade virtual ou *blog*. Porém, fica claro que o objetivo do site é “coletar e disseminar conteúdos elaborados pelos usuários sobre viagens (...), sendo que suas características mais marcantes são os comentários e avaliações”. A interação desses usuários dentro da plataforma cria um grande banco de dados sobre os atrativos e destinos turísticos, fazendo com que o *site* se torne referência em informações na *web*.

Com o intuito de construção de conteúdo colaborativo, os próprios usuários do site cadastram os atrativos turísticos de cada destino, dividindo-os em categorias pré-determinadas. Essas categorias facilitam a busca de informações através de filtros pelos interessados em conhecer um específico tipo de atrativo como, por exemplo, um museu.

O site *TripAdvisor* possibilita o cadastramento de estabelecimentos comerciais (restaurantes, hotéis, casas para aluguel de temporada e companhias aéreas) a partir da criação do perfil da empresa prestadora do serviços turístico na plataforma pelo proprietário ou responsável pelo local, assim como possibilita a inclusão por qualquer usuário da plataforma de atrações turísticas permanentes (ou seja, aquelas que não ocorrem eventualmente como é o caso de *shows* ou espetáculos teatrais de curta temporada, dentre outros) ou excursões que possam ser realizadas em um único dia através da categoria “o que fazer”.

O usuário, ao tentar realizar a avaliação de um atrativo que não está cadastrado no *site* (a partir do campo de busca ou da aba “avaliação”) visualiza uma mensagem, convidando-o a cadastrar o perfil da atração na plataforma.

Ao acessar o *link*, o usuário é levado para uma página com um formulário de preenchimento de informações sobre o local. No caso de um usuário escolher um atrativo relacionado à categoria “o que fazer”, ele deve seguir as diretrizes de cadastramento para essa categoria conforme divulgadas no *site*<sup>5</sup>. As diretrizes destacam que as atrações permanentes cadastradas devem: ser de interesse dos turistas, possibilitar o acesso para todas as idades, estar de acordo com a política de animais do *TripAdvisor* - não incentivar a promoção de locais de exploração animal, mal tratos ou que possibilitam contato direto entre turistas e animais ameaçados de extinção, a não ser de forma gratuita – ter nome oficial e endereço permanente, estar abertas e disponíveis à visitação pelo público geral de forma regular, devem estar abertas por pelo menos 12 semanas consecutivas e estar em pleno funcionamento ou recebendo reservas para uma data de abertura futura.

A partir de então, o sistema permite classificar o atrativo em, no máximo, 3 subcategorias das 212 disponíveis inicialmente na tela de seleção que, ao serem selecionadas, são abertas novas opções de subcategorias que se relacionam com as escolhidas, ampliando ainda a gama de opções de categorização, conforme figura 2:

---

<sup>5</sup> Disponível em <https://www.tripadvisor.com/pt-br/articles/200614767>. Acesso em 09 jan. 2017.

Figura 2: Subcategorias no cadastramento de atrativos

### Editar subcategorias (máx. de 3)

Busque subcategorias para incluir neste lugar

|   |  |   |  |
|---|--|---|--|
| <input type="checkbox"/> Academias de ginástica                       | <input type="checkbox"/> Corrida por pontos turísticos         | <input type="checkbox"/> Lojas outlet                       | <input type="checkbox"/> Passeios particulares             |
| <input type="checkbox"/> Acampamentos esportivos e colônias de férias | <input type="checkbox"/> Cânions                               | <input type="checkbox"/> Lounges de aeroporto               | <input type="checkbox"/> Passeios por bares, clubes e pubs |
| <input type="checkbox"/> Adegas e vinícolas                           | <input type="checkbox"/> Deques e torres de observação         | <input type="checkbox"/> Luaus                              | <input type="checkbox"/> Passeios por fábricas             |
| <input type="checkbox"/> Aluguel de barcos                            | <input type="checkbox"/> Desertos                              | <input type="checkbox"/> Marinas                            | <input type="checkbox"/> Passeios sem guias e alugueis     |
| <input type="checkbox"/> Aluguel de equipamentos                      | <input type="checkbox"/> Destilarias                           | <input type="checkbox"/> Mercados de pulgas e feiras livres | <input type="checkbox"/> Passeios turísticos               |
| <input type="checkbox"/> Antiquários                                  | <input type="checkbox"/> Esportes aquáticos                    | <input type="checkbox"/> Mergulho com e sem snorkel         | <input type="checkbox"/> Passeios à noite                  |
| <input type="checkbox"/> Apresentações                                | <input type="checkbox"/> Esqui aquático e jet ski              | <input type="checkbox"/> Mergulho com tubarões              | <input type="checkbox"/> Perto d'água                      |
| <input type="checkbox"/> Apresentações do Cirque du Soleil            | <input type="checkbox"/> Estúdios de pintura e cerâmica        | <input type="checkbox"/> Minas                              | <input type="checkbox"/> Planos-bar                        |
| <input type="checkbox"/> Aquários                                     | <input type="checkbox"/> Excursões arqueológicas               | <input type="checkbox"/> Minigolfe                          | <input type="checkbox"/> Pistas de boliche                 |
| <input type="checkbox"/> Arenas e estádios                            | <input type="checkbox"/> Excursões com caminhada e acampamento | <input type="checkbox"/> Mirantes                           | <input type="checkbox"/> Pistas de corrida automobilística |
| <input type="checkbox"/> Atividades e parques da Disney               | <input type="checkbox"/> Excursões culturais                   | <input type="checkbox"/> Missões                            | <input type="checkbox"/> Pistas e trilhas para corrida     |
| <input type="checkbox"/> Aulas de culinária                           | <input type="checkbox"/> Excursões de café e chá               | <input type="checkbox"/> Montanhas                          | <input type="checkbox"/> Playgrounds                       |
|   | <input type="checkbox"/> Excursões de fantasmas e              | <input type="checkbox"/> Monumentos e estátuas              |  |

Concluído
Cancelar

Fonte: Tripadvisor

Além da categorização, o usuário deve preencher de forma obrigatória o nome do local, país, cidade e estado e endereço. Já as informações opcionais se relacionam com o preenchimento de maiores informações sobre o endereço, CEP/código postal, telefone, *site*, horários de funcionamento e duração recomendada da visita. Por último, o usuário pode responder também de forma opcional uma lista de 36 perguntas a partir das opções “sim”, “não” e “não tenho certeza” que auxiliam o *site* a categorizar e divulgar a atração na plataforma a partir do interesse do usuário de forma mais precisa<sup>6</sup>.

Ao finalizar o cadastro, o usuário visualiza na tela uma mensagem do *site*, informando que o cadastro será analisado em até 5 dias, sendo que o usuário receberá por *email* a confirmação ou não da aceitação do cadastro da atração.

Caso o cadastro esteja de acordo com as diretrizes propostas pelo site, o atrativo é disponibilizado na plataforma para receber avaliações dos demais usuários.

<sup>6</sup> Tem-se como exemplo perguntas como “Esta atração envolve animais?”, “Esta atração é boa para casais?”, “Esta atração oferece locais para comer?” ou “Esta atração é gratuita para o público?”.

Para avaliar uma atração, o usuário deve entrar com o seu *login* (evitando o anonimato) e, de forma obrigatória, dar uma nota de satisfação para a atração numa escala de 1 e 5 (sendo 1 = horrível, 2 = ruim, 3 = razoável, 4 = muito bom e 5 = excelente), escrever um comentário com, no mínimo, 100 caracteres além do título do texto, responder a pergunta sobre que tipo de visita ele realizou (casal, família, amigos, negócios ou sozinho) e qual foi o mês da visita (sendo possível escolher 1 entre os últimos 12 meses do ano corrente).

De forma opcional, o usuário pode preencher a duração recomendada da visita, além de responder 6 perguntas a partir das opções “sim”, “não” e “não tenho certeza” (da mesma forma que ocorre no cadastramento de um atrativo) e adicionar fotos.

Ao finalizar a avaliação, o usuário deve selecionar de forma obrigatória a opção que certifica que a avaliação é baseada na sua própria experiência e feita de forma sincera, sem possuir nenhuma relação pessoal ou comercial com o local e não tendo recebido incentivo ou pagamento algum para escrevê-la. Além disso, deve concordar que está ciente de que o *site* não possui tolerância com avaliações falsas. Dessa forma, espera-se que não haja intervenção dos próprios estabelecimentos para contratarem pessoas que possam realizar diversos comentários positivos na plataforma para aumentar a nota de avaliação do lugar, reforçando a importância de realizar comentários feitos pela própria vivência dos visitantes, sem manipulação de terceiros.

Assim como o cadastro de atrativos, a avaliação também deve ser aprovada pelo *site* e o usuário recebe a confirmação em até 5 dias por *email*.

É interessante observar que cada avaliação aprovada é convertida em pontos para o perfil do usuário<sup>7</sup>. Quanto maior o número de pontos, maior é o nível do usuário dentro da plataforma<sup>8</sup>, ou seja, maior será sua relevância ou reconhecimento como colaborador dentro da comunidade de usuários. Além dos pontos, o usuário também pode receber selos que são exibidos no seu perfil, caso ele cumpra alguns

---

<sup>7</sup> Pode-se conseguir pontos não só pela avaliação (100 pontos), mas também pela inserção de foto ou vídeo (30 pontos), publicação no fórum de usuários (20 pontos) ou uma avaliação simples de um estabelecimento ou atrativo numa escala de 01 a 05 que aparece de forma aleatória no site (05 pontos).

<sup>8</sup> Para atingir o nível 1 é necessário = 300 pontos, o nível 2 = 500 pontos, nível 3 = 1000 pontos, nível 4 = 2.500 pontos, nível 5 = 5.000 pontos e nível 6 = 10.000 pontos.

critérios de avaliações como, por exemplo, publicar 3 avaliações numa mesma categoria (expert em hotéis, expert em restaurantes ou expert em atrações).

Essa estratégia tem como objetivo incentivar a participação ativa dos usuários, gerando cada vez mais comentários e avaliações fidedignas dentro da plataforma, fortalecendo o *TripAdvisor* a ser um *site* de referência para busca de informações por qualquer viajante que queira opiniões e dicas de estabelecimentos e atrativos antes de programar a sua viagem.

### **2.2.3 O uso de indicadores no turismo**

A criação de indicadores possui um papel fundamental no monitoramento de ações de um determinado programa ou projeto. Dispor de informações periódicas e específicas acerca da implementação e do alcance dos resultados durante o acompanhamento de programas e projetos é uma grande dificuldade para os gestores. (JANNUZZI, 2005).

Os indicadores possibilitam “estabelecer um padrão normativo, por meio do qual é possível construir um diagnóstico para subsidiar a formulação e a avaliação de políticas públicas” (PEREIRA; PINTO, p.365, 2012). Para os autores, é cada vez mais necessário que órgãos de pesquisa estaduais e federais invistam na construção de indicadores elaborados, incentivando os gestores a utilizarem os mesmos como ferramenta de gestão.

Assim como o avanço de sistemas de informação e tecnologias moldaram toda uma nova sociedade da informação, estes também contribuíram diretamente na disseminação do uso de indicadores a partir do acesso mais amplo a banco de dados.

Dados cadastrais antes esquecidos em armários e fichários passam a transitar pela Internet, transformando-se em informação estruturada para análise e tomada de decisão. Dados estatísticos antes inacessíveis em enormes arquivos digitais passam a ser “customizados” na forma de tabelas, mapas e modelos quantitativos construídos por usuários não especializados (JANNUZZI, p.138, 2005).

Por ser uma atividade chave para o desenvolvimento de diversos países e possibilitar aumento de arrecadação, empregos e dinamização de diversos setores,

mensurar os impactos do turismo também é desafiador (DUPEYRAS; MACCALLUM, 2013). Assim, elaborar estatísticas e indicadores para a atividade se torna fundamental.

Viegas (1997) defende a importância das estatísticas para o desenvolvimento do turismo:

A existência de estatísticas exatas sobre o turismo é de vital importância para determinar os seus impactos diretos ou indiretos na economia, para auxiliar no planejamento e desenvolvimento de novos produtos turísticos, para determinar as características do turista, para permitir a formulação de estratégias promocionais e de marketing e para detectar mudanças nas preferências e características do turista (VIEGAS, 1997, p.2).

A partir da evolução dos meios digitais, a captura de informação para monitoramento da atividade turística se tornou possível e relevante para compreender o comportamento dos consumidores e os níveis de satisfação. O conteúdo digital encontra-se amplamente disponível, transformando o que anteriormente era considerado como uma experiência individual em conhecimento compartilhado em diversas bases de dados ao redor do mundo. Agora, essas informações podem ser monitoradas e analisadas por empresas e organizações tanto públicas quanto privadas (MUNAR; JACOBSEN, 2013).

Monitorar indicadores no turismo se torna crucial para gestores públicos e para os profissionais que necessitam de evidências para auxiliar os tomadores de decisões nas organizações. A “economia global e tendências do turismo, incluindo mudanças nos mercados e no comportamento dos turistas, o papel das mídias sociais e novas formas de demanda (...) aumentam a importância da questão” (DUPEYRAS; MACCALLUM, p.6, 2013, tradução nossa).

De forma geral, as redes que circundam os viajantes durante o planejamento das viagens capturam e geram um enorme volume de informações, característico de *Big Data*, criando oportunidades e desafios para o turismo. O levantamento de informações para os destinos turísticos:

(...) permitem aos profissionais (...) entender melhor onde e como vivem os visitantes reais e potenciais, a natureza da informação utilizada para planejar uma viagem, bem como com quem os

viajantes compartilham suas experiências antes, durante e depois da viagem (XIANG; MAGNINI; FESENMAIER, p. 524, 2015).

Para o turismo, a elaboração de indicadores, a partir dos atuais contextos de inovação tecnológica e relacionamento entre as pessoas em plataformas online, podem servir como meio de acompanhamento de políticas públicas, facilitando a tomada de decisão e direcionando esforços para a melhoria de projetos destinados ao desenvolvimento do setor. Para tanto, é essencial que os indicadores propostos possam ser elaborados de forma clara e transparente, aumentando sua legitimidade como ferramenta de gestão social (JANNUZZI, 2005).

### 3 Metodologia

A atividade de pesquisa visa descobrir novos conhecimentos e, portanto, é necessário modelar o trajeto a ser seguido, uma vez que cada caminho pode levar o pesquisador a alcançar diferentes resultados, devendo avaliar os limites da pesquisa e as oportunidades a partir do contexto que se pretende trabalhar (TEIXEIRA, 2003). Portanto, descrever a metodologia científica de forma clara é fundamental para o acompanhamento e compreensão do objeto estudado.

A pesquisa apresentada neste trabalho pode ser classificada como exploratória, na qual tem como objetivo familiarizar com um problema em questão, buscando torna-lo mais explícito. Além disso, visa aprimorar ideias ou descoberta de intuições (GIL, 2002). No caso apresentado, pretende-se explorar as possibilidades de análise e elaboração de indicadores a partir da utilização de dados disponibilizados em meio digital. Conforme destacado por Gil (2002), as pesquisas exploratórias geralmente são elaboradas de forma bibliográfica ou de estudo de caso.

Para aprofundamento das discussões da pesquisa exploratória, utilizou-se um estudo de caso instrumental. O estudo de caso geralmente consiste no estudo profundo de um ou poucos objetos, de maneira que permita ampliar e detalhar seu conhecimento (GIL, 2002). Já o estudo de caso instrumental é:

(..) aquele que é desenvolvido com o propósito de auxiliar no conhecimento ou redefinição de determinado problema. O pesquisador não tem interesse específico no caso, mas reconhece que pode ser útil para alcançar determinados objetivos (GIL, p.139, 2002).

De forma prática, foi utilizado como universo os comentários e avaliações dos usuários postadas nas páginas de atrativos turísticos de Minas Gerais contidas no *site TripAdvisor*. É importante destacar que as avaliações nos atrativos e destinos podem ser feitas tanto por visitantes quanto por moradores que visitaram os espaços, mas sem descaracterizar a importância do *site* para o segmento do turismo, conforme discussão apresentada no capítulo anterior.

O objetivo do estudo de caso é analisar a possibilidade de criar indicadores para políticas públicas a partir de dados digitais e que possa ser replicado em outras plataformas.

Além disso, a pesquisa utilizou métodos quantitativos e qualitativos para realizar seu processo investigativo. O método quantitativo se baseia na utilização de instrumentos padronizados e neutros para a coleta de informações, assegurando generalizações de forma precisa e objetiva. No caso desse estudo, a extração das informações, assim como o tratamento dos dados para a elaboração de indicadores foram tratados quantitativamente, buscando de forma censitária trabalhar com as informações dos atrativos e destinos de Minas Gerais cadastrados no site *TripAdvisor*. Isso possibilitou a realização de cálculos numéricos, a partir da utilização de uma ferramenta de extração de dados *web* denominada *import.io*. Para Teixeira (2003), na pesquisa de caráter quantitativo, dados são analisados de forma estatística, geralmente com suporte de aplicações de computadores, complementando as qualidades de interpretação do pesquisador.

Já o método qualitativo é mais indicado para investigações críticas e interpretativas, onde se deseja estudar fenômenos que envolvem diversas relações sociais entre indivíduos ou grupos de indivíduos, onde o investigador é considerado como instrumento primário e essencial na coleta e análise de dados (TEIXEIRA, 2003). Portanto, a análise das informações e sua interpretação se deram de forma qualitativa. Para a conclusão, foi necessário contextualizar as informações coletadas com a realidade vivida e experimentada pelo autor. Isso possibilitou que o uso da estatística descritiva pudesse apoiar uma interpretação considerada como subjetiva (TEIXEIRA, 2003).

Assim, considera-se que essa pesquisa possa ser caracterizada com elementos quantitativos e qualitativos, sendo denominada de híbrida (CRESWELL, 2003).

A partir dessa metodologia científica, serão detalhados a seguir, os passos para a coleta dos dados pela ferramenta *import.io*, a limpeza e organização do banco de dados e a análise dos resultados que possibilitaram a construção de indicadores para o monitoramento do turismo em Minas Gerais.

### 3.1 Coleta de dados

O aplicativo denominado *import.io* é um produto da *startup* de tecnologia<sup>9</sup> de mesmo nome, fundada no dia 21 de junho de 2012 por três investidores. Atualmente, a empresa é sediada na cidade de Londres (Inglaterra) e com uma base de operação em São Francisco (Estados Unidos). Desde a sua fundação até o mês de janeiro de 2016, a organização já havia conseguido um aporte financeiro de U\$ 17,3 milhões de 13 investidores<sup>10</sup> para desenvolvimento de ferramentas de extração de dados de sites na internet.

De acordo com informações retiradas do site da empresa ([www.import.io](http://www.import.io)), o *import.io* é uma plataforma que possibilita de forma rápida extrair dados de páginas da *web* e transformá-los em bancos de dados estruturados, facilitando a organização da informação por usuários que necessitam analisar dados que estão disponíveis na rede mundial de computadores, mas que estão distribuídos de forma dispersa ou em grande volume em um site, por exemplo. Assim, ela se denomina como uma solução tecnológica para coleta de dados com características de *Big Data*.

A escolha de utilização dessa aplicação para esse estudo foi realizada a partir de testes com diversas ferramentas de extração de dados disponíveis a partir de buscas pela *internet*. Das ferramentas encontradas (tais como o *Kimono*, *Scraper.io* e *Rapid Miner*), o *import.io* além de ser uma ferramenta de extração de dados disponível de forma gratuita na *web* e principalmente por não necessitar de conhecimentos de linguagem de programação para sua utilização, permitiu de forma mais didática compreender seu funcionamento. Portanto foi considerada a ferramenta mais acessível para os propósitos da pesquisa. Ressalta-se que nos momentos de dúvida de utilização da ferramenta, a equipe de apoio do *import.io* foi acionada por e-mail e respondeu rapidamente e de forma efetiva os questionamentos, possibilitando extrair todas as informações necessárias para este trabalho.

---

<sup>9</sup> De acordo com Paternoster et al. (2014) uma *startup* de tecnologia é considerada como uma companhia recém-criada sem histórico operacional e com produção rápida de tecnologia de ponta.

<sup>10</sup> Fonte: <https://www.crunchbase.com/organization/importio#/entity> Acesso em: 6 de abr. de 2016

Para a extração das informações, foi utilizada a versão *web* da ferramenta que não necessita de instalação de *software* em um computador para seu funcionamento<sup>11</sup>. Essa versão possibilitou a extração gratuita de até 10.000 páginas (ou *URLs*) individuais por mês, sendo suficiente para essa pesquisa. Ou seja, seria possível extrair as informações de até 10.000 atrativos cadastrados em Minas Gerais na plataforma *TripAdvisor* e monitorá-los mês a mês.

De forma geral, a ferramenta utilizada possibilita extrair informações de uma ou mais páginas que se conheça a *URL* de extração (e que possuem a mesma padronização de layout e organização das informações) selecionadas pelo próprio usuário. O extrator possibilita ao usuário “ensinar” o aplicativo do *import.io* a extrair apenas aquelas informações que constam em uma determinada página (*URL*), sendo ela uma informação de texto, imagem, *link* ou número, registrando-as em um arquivo compatível com a utilização em banco de dados.

Para coletar todas as informações selecionadas dos atrativos turísticos de Minas Gerais, foi realizado um trabalho de duas etapas detalhadas a seguir, sendo a primeira para selecionar as informações que seriam coletadas a partir de uma página modelo de um atrativo de Minas Gerais e a segunda etapa para coletar todas as *URLs* dos demais atrativos do estado cadastrados no *TripAdvisor* com vistas a replicar a extração das informações da página modelo para as demais páginas, possibilitando a criação do banco de dados.

### **3.1.1 Seleção de informações para captura: extrator 1**

Para realizar o levantamento dos dados dos atrativos turísticos do estado cadastrados no site *TripAdvisor*, primeiramente foi necessário selecionar as principais informações que seriam extraídas de uma página modelo de um atrativo no *TripAdvisor*, “ensinando” a ferramenta do *import.io* a extraí-las para um banco de

---

<sup>11</sup> A partir do dia 2 de abril de 2016, a ferramenta *import.io* passou por uma grande atualização, sendo criado um novo extrator de forma gratuita via *web* substituindo uma versão mais completa ofertada pela instalação de um aplicativo *desktop*, com um limite de extrações e possibilidade de pacotes pagos com maiores funcionalidades de extração de dados. Porém, para os usuários que já estavam cadastrados na plataforma antes dessa data (no caso, o autor do trabalho), a versão *desktop* ainda continua disponível com todas as funcionalidades. Vale ressaltar que mesmo com a versão paga, é possível conseguir a versão completa de forma gratuita para fins de estudo e trabalhos acadêmicos, entrando em contato diretamente com a empresa. Mas para fins desse trabalho e visto que a grande maioria das pessoas terá acesso apenas a versão limitada, optou-se em utilizar a versão *web*.

dados. A tela principal do *import.io* (figura 3), nos dá primeiramente a opção de uma extração de informações selecionadas automaticamente pela própria ferramenta a partir de um *URL* qualquer, inserido no campo destacado.

No caso desse trabalho, foi inserida a *URL* referente ao atrativo “Praça da Liberdade”<sup>12</sup> em Belo Horizonte, e em seguida ativado o botão de extração “*Try it out*”. A partir desse comando, a ferramenta de extração analisa de forma automática a estrutura da página e seleciona, também de forma automática, as informações nas quais ela considera como relevante, colocando-as em uma planilha, conforme a figura 4.

Figura 3 – Tela inicial do *import.io*



Fonte: Import.io

<sup>12</sup> A *URL* referente à Praça da Liberdade utilizada foi: [https://www.tripadvisor.com.br/Attraction\\_Review-g303374-d1534424-Reviews-Praça\\_da\\_Liberdade-Belo\\_Horizonte\\_State\\_of\\_Minus\\_Gerais.html](https://www.tripadvisor.com.br/Attraction_Review-g303374-d1534424-Reviews-Praça_da_Liberdade-Belo_Horizonte_State_of_Minus_Gerais.html). Acesso em: 12 de jun. de 2016.

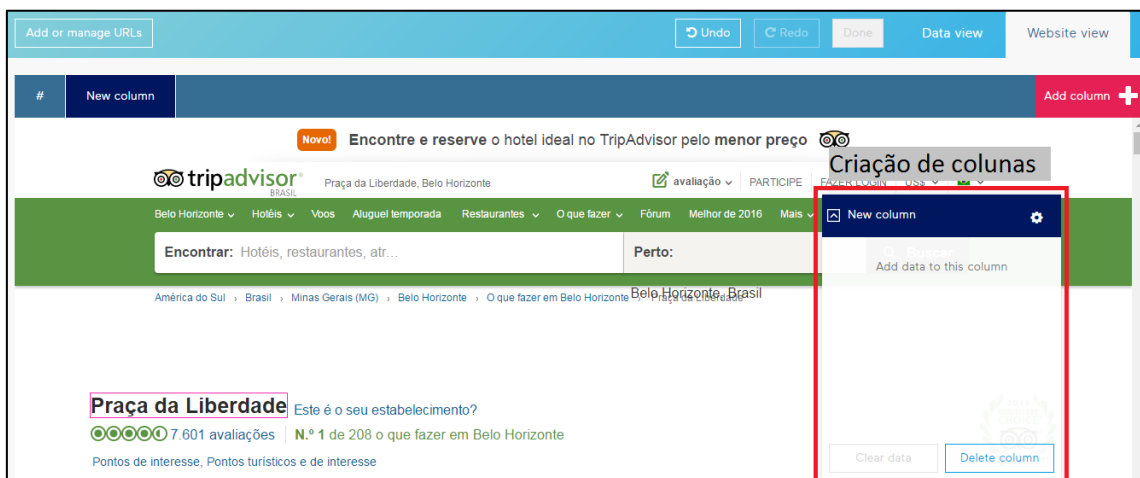
Figura 4: Seleção automática de informações pelo *import.io*



Fonte: Import.io

Porém, as informações capturadas de forma automática não foram consideradas ideais para o trabalho e assim, conforme destacado em vermelho na figura 5, optou-se por criar uma tabela em branco, possibilitando a escolha de quais informações seriam extraídas no site. Ao acionar o botão, a ferramenta mostra a opção de criar uma tabela totalmente em branco (*start with a blank page*) ou ainda possibilita pré-definir o cabeçalho das colunas a partir do tipo de página na qual as informações serão extraídas: produto (*product*), negócios (*business*) ou notícias (*news*). Optou-se então, pela criação de uma tabela totalmente em branco.

Figura 5: Tabela em branco para criação de colunas



Fonte: Import.io

Com a página em branco, foi necessário então, selecionar cada uma das informações da página que seriam inseridas nas colunas da tabela em branco, considerando que cada informação diferente ocupasse uma coluna diferente da planilha.

Finalmente, as informações selecionadas para a captura e para a criação do título de cada uma das colunas foram destacadas conforme numeração da figura 6, e listadas como: nome do atrativo (1), município (2), categoria do atrativo (3), número de avaliações por faixas de satisfação (excelente (4), muito bom (5), razoável (6), ruim (7), horrível (8)), motivação para a visita ao atrativo (família (9), romântico (10), sozinho (11), negócios (12) ou amigos (13)) além da temporada de viagem (março a maio (14), junho a agosto (15), setembro a novembro (16) e dezembro a fevereiro (17)) e endereço (18).

É importante destacar que na figura 6, no item “3”, o *TripAdvisor* possibilita a visualização de mais de uma opção de categorização do atrativo (no exemplo, a Praça da Liberdade possui a categorização de “pontos de interesse” e “pontos turísticos e de interesse”). Porém, percebeu-se em alguns testes realizados antes da extração final deste trabalho que a primeira classificação mostrada na página (e na qual foi extraída a informação de “categoria do atrativo” para esse trabalho) foi considerada a mais significativa para classificar o atrativo, sendo que a segunda categoria se apresentava mais ampla do que a primeira, a terceira mais ampla do que a segunda e assim sucessivamente. Desta forma, preferiu-se extrair para a análise dos resultados apenas uma categoria que foi entendida como a classificação mais segmentada para os atrativos (no caso do exemplo foi considerada apenas a classificação “pontos de interesse”).

Outro ponto destacado na figura 6 refere-se às informações numéricas demonstradas no intervalo dos números 4 a 17. Percebe-se que os valores apresentados no site são relacionados ao filtro de comentários realizados na língua utilizada pelo visitante do site, que no caso, foi a língua portuguesa (conforme destacado no item 19). Caso a pesquisa fosse realizada no site *TripAdvisor* em inglês ([www.tripadvisor.com](http://www.tripadvisor.com) ou [www.tripadvisor.co.uk](http://www.tripadvisor.co.uk)), o filtro automaticamente mudaria para a língua inglesa, e os valores mostrados para cada uma das avaliações seria limitados aos comentários em inglês, reduzindo muito os resultados apresentados para os atrativos de Minas Gerais.

Figura 6: Informações selecionadas para extração no *TripAdvisor*

Belo Horizonte ▾ Hotéis ▾ Voos Aluguel temporada Restaurantes ▾ O que fazer ▾ Fórum Melhor de 2016 Mais ▾

Encontrar: Hotéis, restaurantes, atrações Perto: Belo Horizonte, Brasil

América do Sul ▾ Brasil ▾ Minas Gerais (MG) ▾ **Belo Horizonte** ▾ **2** O que fazer em Belo Horizonte ▾ Praça da Liberdade

**1** **Praça da Liberdade** Este é o seu estabelecimento?

7.602 avaliações | N.º 1 de 208 o que fazer em Belo Horizonte

**3** Pontos de interesse Pontos turísticos e de interesse

SOMENTE NO SITE OFICIAL RESERVE JÁ

7.602 avaliações feitas pela comunidade do TripAdvisor 5 pontos

Deivison Henrique, Débora Carmo, Fernanda Brito de Pinho e mais 31 amigos estiveram aqui. Veja as pontuações deles...

Veja a opinião dos viajantes:

| Pontuação dos viajantes                             | Tipo de viajante                                     | Época do ano                                       | Idioma   |
|---|--|--|--|
| <input type="checkbox"/> Excelente (4.311) <b>4</b> | <input type="checkbox"/> Famílias (2.026) <b>9</b>   | <input type="checkbox"/> Mar-Mai (1.903) <b>14</b> | <input type="radio"/> Todos os idiomas             |
| <input type="checkbox"/> Muito bom (2.526) <b>5</b> | <input type="checkbox"/> Romântico (1.267) <b>10</b> | <input type="checkbox"/> Jun-Ago (2.324) <b>15</b> | <input checked="" type="radio"/> Português (7.353) |
| <input type="checkbox"/> Razoável (477) <b>6</b>    | <input type="checkbox"/> Sozinho (949) <b>11</b>     | <input type="checkbox"/> Set-Nov (1.535) <b>16</b> | <input type="radio"/> Inglês (153)                 |
| <input type="checkbox"/> Ruim (30) <b>7</b>         | <input type="checkbox"/> Negócios (496) <b>12</b>    | <input type="checkbox"/> Dez-Fev (1.591) <b>17</b> | <input type="radio"/> Espanhol (60)                |
| <input type="checkbox"/> Horrível (9) <b>8</b>      | <input type="checkbox"/> Amigos (1.836) <b>13</b>    |  | <a href="#">Mais</a>                               |

Filtrado por: Português **Caixa de filtro** **19** [Limpar tudo](#)

Reserve com antecedência

Excursão particular: descubra a cidade de Belo Horizonte A partir de **R\$281\*** [Mais informações](#)

**Por que reservar na Viator?**

- Destaques da excursão e itinerário completo
- Reserva online fácil
- Menor preço garantido

[Mais informações >](#)

Esta atração é gratuita para o público?  Sim  Não  Incerteza

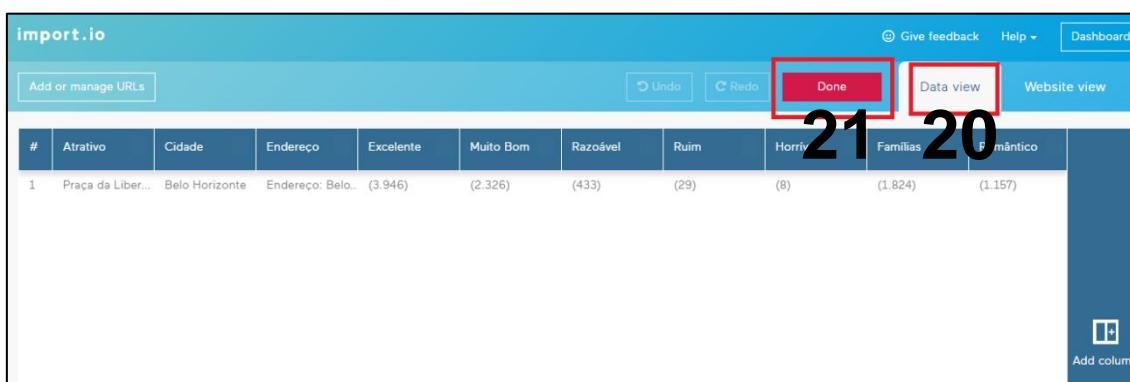
**18** **Endereço:** Belo Horizonte, Minas Gerais, Brasil  
Número de telefone: 30140-050-31-3217-9538  
[Site](#)

Fonte: Tripadvisor

Ainda nesse aspecto, não considerou interessante extrair as informações destacadas no campo “idioma”, pois há que se considerar que alguns usuários da plataforma podem morar no Brasil e escrever o comentário em outra língua (como inglês ou espanhol), afetando os resultados para esse quesito e dificultando uma análise real do idioma de cada comentário. Além disso, de acordo com a pesquisa de demanda turística realizada pela SETUR-MG em 2014, apenas 1% dos visitantes do estado de Minas Gerais eram estrangeiros, o que reforça que esse público ainda não é expressivo em relação ao total de visitantes (que pode se refletir também no *TripAdvisor*), não prejudicando os resultados da análise de comentários apenas em português.

Logo após a seleção das informações, pode-se observar a relação de informações coletadas para cada coluna ao selecionar a aba “*data view*”, conforme item 20 da figura 7:

Figura 7: Visualização das informações selecionadas na aba “*data view*”

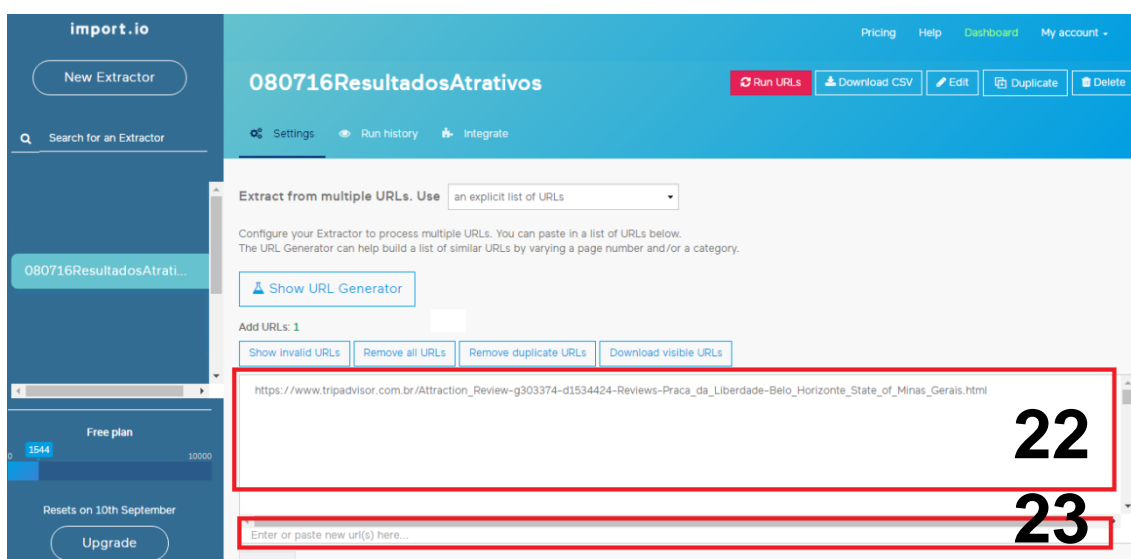


| # | Atrativo          | Cidade         | Endereço          | Excelente | Muito Bom | Razoável | Ruim | Horrrível | Famílias | Romântico |
|---|-------------------|----------------|-------------------|-----------|-----------|----------|------|-----------|----------|-----------|
| 1 | Praça da Liber... | Belo Horizonte | Endereço: Belo... | (3.946)   | (2.326)   | (433)    | (29) | (8)       | (1.824)  | (1.157)   |

Fonte: Import.io

Após a checagem das informações dispostas na planilha de forma satisfatória, selecionou-se o botão “*done*” (item 21, figura 7), finalizando a etapa de seleção das informações a serem capturadas pelo extrator, abrindo uma tela de trabalho para início da extração, denominada “*dashboard*”, conforme figura 8.

Figura 8: Tela de trabalho (*dashboard*) – *import.io*



Fonte: Import.io

O item 22 da figura 8, destaca as *URLs* que serão extraídas as informações selecionadas, sendo possível inserir novas *URLs* no campo 23. Como a seleção das informações foi realizada apenas na página modelo da Praça da Liberdade, o extrator (denominado a partir de agora como “extrator 1”) apresentou apenas a *URL* relacionada à Praça da Liberdade. Então, para conseguir utilizar o extrator 1 para todas as páginas de atrativos de Minas Gerais, foi necessário criar outro extrator apenas para coletar as *URLs* dos atrativos para serem inseridas no campo 23 e assim, extrair as informações dos demais atrativos. Esse passo será destacado na etapa a seguir.

### 3.1.2 Coleta de *URLs* dos atrativos de Minas Gerais: extrator 2

A próxima etapa da extração foi coletar todas as *URLs* dos atrativos turísticos de Minas Gerais para posterior inserção no extrator 1, fazendo-o coletar as mesmas informações da página modelo, porém de todos os demais atrativos e transformando-as no banco de dados.

Para tanto, foi necessário criar um novo extrator, denominado de “extrator 2”, apenas para a coleta de *URLs* dos atrativos para cada página de municípios do estado. A figura 9, mostra que cada *URL* de um determinado município no site *TripAdvisor* possui a listagem de informações básicas de todos os atrativos

daquele município, sendo que a informação do nome do atrativo, pode ser extraída na forma de um *link*, que direciona para a página do atrativo.

No caso da figura 9, tem-se a página principal dos atrativos de Belo Horizonte, com um destaque para as informações que ao serem selecionadas pelo usuário do site *TripAdvisor*, direciona-o para a página do atrativo. Ou seja, ao selecionar o título “Praça da Liberdade”, o usuário irá diretamente para a página da figura 9 com as informações completas da “Praça da Liberdade” (a mesma página da que serviu de modelo para extrator 1).

Nesse caso, foi necessário criar um extrator a partir da página modelo de Belo Horizonte que extraísse apenas as *URLs* dos campos dos títulos dos atrativos, criando uma tabela de apenas uma coluna, com o nome “*URL*”. Nesse caso, as informações dos títulos foram transformadas em *URLs* conforme exemplo abaixo, sendo que cada página do *TripAdvisor* lista, no máximo, 30 atrativos:

- Praça da Liberdade =  
[https://www.tripadvisor.com.br/Attraction\\_Review-g303374-d1534424-Reviews-Praca da Liberdade-Belo Horizonte State of Minas Gerais.html](https://www.tripadvisor.com.br/Attraction_Review-g303374-d1534424-Reviews-Praca_da_Liberdade-Belo_Horizonte_State_of_Minas_Gerais.html)
- Mercado Central de Belo Horizonte =  
[https://www.tripadvisor.com.br/Attraction\\_Review-g303374-d1534420-Reviews-Mercado Central de Belo Horizonte-Belo Horizonte State of Minas Gerais.html](https://www.tripadvisor.com.br/Attraction_Review-g303374-d1534420-Reviews-Mercado_Central_de_Belo_Horizonte-Belo_Horizonte_State_of_Minas_Gerais.html)
- Memorial Minas Gerais Vale =  
[https://www.tripadvisor.com.br/Attraction\\_Review-g303374-d2360286-Reviews-Memorial Minas Gerais Vale-Belo Horizonte State of Minas Gerais.html](https://www.tripadvisor.com.br/Attraction_Review-g303374-d2360286-Reviews-Memorial_Minas_Gerais_Vale-Belo_Horizonte_State_of_Minas_Gerais.html)
- ...Atrativo N30 = *URL\_atrativo\_N30*

Figura 9: Página “o que fazer” do município de Belo Horizonte

América do Sul > Brasil > Minas Gerais (MG) > Belo Horizonte > O que fazer em Belo Horizonte

## O que fazer em Belo Horizonte, MG

Ordenar por: **Classificação** Reserve online

Ver mapa

TIPOS DE ATRAÇÃO:

- Pontos turísticos e de interesse (53)
- Compras (36)
- Museus (54)
- Passeios gastronômicos (6)
- Natureza e parques (12)
- Diversão e jogos (13)
- Concertos e shows (50)

**Praça da Liberdade**  
Nº 1 de 208 o que fazer em Belo Horizonte  
7.611 avaliações  
"espetacular" 25/08/2016  
"Museus" 25/08/2016  
Pontos de interesse  
Vencedor do Travellers' Choice™ 2016

**Mercado Central de Belo Horizonte**  
Nº 2 de 208 o que fazer em Belo Horizonte  
7.668 avaliações  
"Pura sinestesia!!!" 25/08/2016  
"Muito grande" 25/08/2016  
Feiras de agricultores  
Certificado de Excelência

**Memorial Minas Gerais Vale**  
Nº 3 de 208 o que fazer em Belo Horizonte  
1.855 avaliações  
"Local calmo, fresco e com..."  
"Um pouco da história e d..."  
Museus especializados  
Certificado de Excelência

URL de acesso para a página do atrativo

URL de acesso para a página do atrativo

URL de acesso para a página do atrativo

Fonte: Tripadvisor

As demais URLs dos municípios de Minas Gerais foram encontradas na página denominada “O que fazer em Minas Gerais” e também no *link* “mais” da mesma página, que ao ser selecionado, abre uma listagem com as URLs dos demais destinos, conforme figura 10.

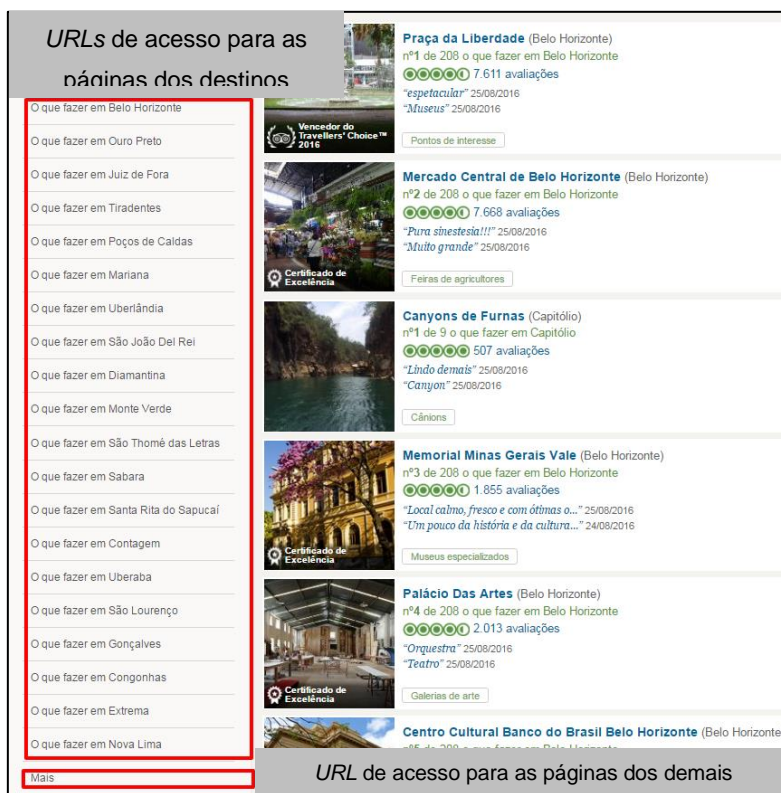
Com a localização das páginas com a listagem das *URLs* dos municípios, criou-se um extrator preliminar apenas para coleta dessas *URLs*. Apenas para alguns municípios que possuíam mais de uma página de atrativos foi necessário acessar cada um dessas páginas para coletar o endereço das *URLs* visto que a *URL* é diferente para cada uma delas. Belo Horizonte, por exemplo, possuía 7 páginas de listagem de atrativos, e foi necessário copiar para um arquivo de texto as 7 *URLs*, sendo uma respectiva de cada página. Porém, a maioria dos destinos possuía apenas uma página de atrativos, que foi coletada de forma automática.

Foram coletados no total 267 URLs que foram inseridas no campo 24 do extrator 2 conforme figura 11.

O próximo passo do processo foi acionar o botão de extração (figura 11), possibilitando o *download* de um arquivo em formato de editor de planilhas, com apenas uma coluna com os endereços de 1482 URLs dos atrativos turísticos de Minas Gerais cadastrados no site *TripAdvisor*.

Em seguida, as 1482 URLs extraídas do extrator 2 foram copiadas para o item 23 (figura 8) do extrator 01, possibilitando a extração das informações selecionadas na primeira etapa para todos os atrativos turísticos de Minas Gerais, após o acionamento do botão de extração.

Figura 10: Página “o que fazer” em Minas Gerais e link “mais”



Fonte: TripAdvisor

Figura 11: Tela de trabalho (*dashboard*) do extrator 2 com as *URLs* dos destinos

Fonte: Import.io

## 3.2 Organização do banco de dados

Após a extração das informações, foi realizado o *download* do banco de dados para o *excel*, no qual foi necessário organizar as informações para posterior análise. No total, foram coletadas as informações de 1.482 *URLs*, sendo que 98 (6,6%) foram desconsideradas, pois se tratavam de *URLs* de grupos de atrativos ou serviços (e não de um único atrativo ou serviço), que são divulgadas nas páginas de alguns destinos e que redirecionam para outros atrativos. Além disso, caso as informações desses grupos de atrativos fossem extraídas, haveria uma grande repetição de informações de atrativos, portanto, para fins dessa pesquisa, optou-se por não extrair essas informações e retirá-las do banco de dados.

Dessa forma, foi considerado para a formação do banco de dados um total de 1.384 atrativos de 253 municípios cadastrados. Porém, percebeu-se que parte dos atrativos extraídos possuía um número muito pequeno de avaliações, o que dificultaria a análise de informações quantitativas como, por exemplo, o percentual de pessoas que avaliaram o atrativo como excelente, muito bom, regular, ruim ou péssimo. Sendo assim, decidiu-se por fazer um recorte no número de atrativos a partir do número total de avaliações de cada um (somatório das avaliações de excelente, muito bom, regular, ruim ou péssimo).

Para a realização desse recorte, tirou-se a média do número de avaliações dos atrativos. O resultado mostrou que, em média, cada atrativo possuía 106 avaliações e portanto, foram considerados para análise dos resultados apenas os atrativos que possuíam número igual ou acima dessa média.

Após esse recorte foram considerados para a análise dos resultados 235 atrativos de 46 municípios, ou seja, 17% do número de atrativos e 18% dos municípios extraídos inicialmente.

Deve-se ressaltar que o banco de dados extraído pelo *import.io* foi realizado no formato *CSV*<sup>13</sup>, sendo necessária para a análise das informações a separação dos valores em colunas distintas. Para isso, foi selecionada toda a 1ª

---

<sup>13</sup> A abreviação *CSV* significa em inglês *comma-separated values*, podendo ser traduzido para “valores separados por vírgulas”, ou seja, todos os resultados de um banco de dados são exibidos em apenas uma coluna, separados por vírgulas, sendo necessário o desmembramento de cada valor entre as vírgulas para colunas distintas, facilitando a análise das informações para o usuário que pode trabalhá-las manualmente.

coluna da planilha em *excel* com as informações captadas e posteriormente utilizado o comando “texto para colunas”, separando os valores com a seleção da opção “vírgula”. Após essa transformação, cada um dos valores foi transferido para colunas distintas, possibilitando a análise das informações.

O banco de dados foi transformado nas seguintes colunas: *URL*, atrativo, cidade, categoria, endereço, excelente, muito bom, razoável, ruim, horrível, famílias, romântico, sozinho, negócios, amigos, mar-mai, jun-ago, set-nov, dez-fev, sendo que as cinco primeiras colunas possuíam dados no formato de texto e as demais em números. Porém, a informação numérica disponibilizada pelo *TripAdvisor* (conforme itens 4 a 17 da figura 6), é representada em parênteses, o que o banco de dados, ao extrair as informações, considerou como um sinal negativo “-”, ou seja, todos os dados foram representados no banco com o sinal negativo. Assim, todos os valores numéricos da planilha tiveram que ser multiplicados por -1, fazendo com que os valores se tornassem positivos.

Após essa etapa, foram criadas três colunas denominadas “avaliações”, com o somatório dos valores absolutos da satisfação dos visitantes (nos campos “excelente”, “muito bom”, “razoável”, “ruim”, “horrível”), do tipo de viagem (“famílias”, “romântico”, “sozinho”, “negócios”, “amigos”) e da época de visitaç o (“mar-mai”, “jun-ago”, “set-nov”, “dez-fev”). Isso foi necess rio para checar se todos os usu rios avaliaram os tr s quesitos obrigatoriamente no site *TripAdvisor* e percebeu-se que o n mero de avalia es para a satisfa o do visitante e  poca de visita o coincidiram, diferentemente do tipo de viagem, que se apresentava com valores menores. Ou seja, o usu rio do *TripAdvisor* ao avaliar um destino ou atrativo obrigatoriamente preenche os campos de satisfa o e  poca de visita o, mas nem sempre preenche o tipo de viagem. Como os somat rios de cada um dos tr s quesitos, foi poss vel calcular o valor percentual de cada coluna em rela o ao total de seu quesito, facilitando a compreens o das informa es.

Outra coluna criada no banco de dados fez refer ncia   nota m dia obtida por cada destino em rela o   satisfa o dos visitantes. Para isso, para cada avalia o no campo “excelente” foi determinada uma nota 5, o campo “muito bom” uma nota 4, o campo “razo vel” uma nota 3, o campo “ruim” uma nota 2 e o campo “horr vel” como nota 1. Em seguida, somou-se todos os valores, dividindo o resultado

pelo total das avaliações, chegando ao valor médio da avaliação para cada atrativo ou destino, sempre no intervalo de 1 a 5, sendo 1 o valor mínimo e 5 o valor máximo.

A extração de dados deste estudo foi realizada no dia 12 de julho de 2016 na qual se baseou a análise descritiva dos dados a seguir.

## 4 Apresentação e análise dos resultados

A metodologia apresentada possibilitou coletar as informações de cada um dos atrativos selecionados e trabalhá-las de forma eficiente a partir de uma planilha do *excel* denominada “tabela dinâmica”. Foi possível identificar:

- a oferta de atrativos por município e por categoria de atrativo (museus, igreja, shopping...);
- os atrativos, municípios e categorias com maior número de avaliações absolutas, média de nota, além do percentual de avaliações por faixa de satisfação (excelente, muito bom...);
- o perfil dos visitantes para cada município, atrativo e categoria (família, sozinho, a negócios...)
- a época de visitação mais pertinente de cada município, atrativo e categoria.

A seguir, serão detalhados os resultados mais relevantes para cada um dos itens elencados a partir da extração de informações realizada no dia 12 de julho de 2016.

### 4.1 Oferta de atrativos

Dentre o recorte de atrativos realizado, os resultados apontaram que os principais atrativos e destinos de Minas Gerais cadastrados no *TripAdvisor* estavam divididos em 53 categorias, sendo que uma categoria foi considerada como “sem identificação”, pois o atrativo “Minascentro” não apresentou no *site* uma categorização. As categorias com o maior número de cadastros podem ser conferidas na tabela 1.

Por permitir que os próprios usuários categorizem os atrativos e pela extração ter sido feita apenas com a primeira categoria destacada no *site*, percebeu-se que para uma análise estadual alguns atrativos poderiam ser categorizados de forma distinta, como é o caso do Parque Estadual do Itacolomi. Na extração, o atrativo foi categorizado na categoria “áreas com vida selvagem e natureza”, sendo que havia a categoria “parques” ou ainda “parques estaduais”. Ao conferir no *site*, percebeu-se que além da categoria informada na extração, o parque estava categorizado apenas em “natureza e parques”. Já na categoria “parques estaduais” há apenas o cadastro do

Monumento Natural Estadual Gruta do Rei do Mato, sendo que os demais parques estaduais estavam categorizados na categoria “parques”, o que pode prejudicar uma análise consolidada das informações por categorias. Esse fator pode inclusive dificultar ao usuário do *TripAdvisor* na busca de informações pelo site a partir das categorias, deixando de informar sobre atrativos relevantes que foram categorizados de forma incoerente.

Tabela 1 – Percentual de cadastros de atrativos de Minas Gerais no site *TripAdvisor* por categoria

| <b>Categoria</b>                   | <b>Cadastros</b> |
|------------------------------------|------------------|
| Pontos de interesse                | 13,62%           |
| Igrejas e catedrais                | 13,19%           |
| Museus especializados              | 10,64%           |
| Obras arquitetônicas               | 6,38%            |
| Outras atividades ao ar livre      | 4,68%            |
| Parques                            | 4,26%            |
| Lojas especializadas e de souvenir | 3,40%            |
| Áreas com vida selvagem e natureza | 2,55%            |
| Cachoeiras                         | 2,55%            |
| Shopping centers                   | 2,55%            |
| Teatros                            | 2,55%            |
| Trilhas para caminhada             | 2,55%            |
| Cavernas e grutas                  | 2,13%            |
| Museus de história                 | 2,13%            |
| Perto d'Água                       | 2,13%            |
| Outros                             | 24,68%           |

Fonte: *TripAdvisor*  
 Elaboração: próprio autor

Posto essa consideração, sugere-se que as análises realizadas a partir de filtros de categorias sejam avaliadas de forma cuidadosa ao se tratar de informações completas para o estado. Já no caso de municípios, a categorização se apresentou de forma mais aceitável, apesar de serem encontradas algumas incoerências em alguns casos. No exemplo da tabela 2, temos as categorias dos atrativos de Ouro Preto.

Tabela 2 – Distribuição dos atrativos de Ouro Preto cadastrados no *TripAdvisor* por categorias

**Ouro Preto**

---

**Areas com vida selvagem e natureza**

Parque Estadual de Itacolomi

**Formacoes geologicas**

Mina du Veloso

**Igrejas e catedrais**

Basilica de Nossa Senhora do Pilar

Igreja de Sao Francisco de Assis

Igreja Matriz N S da Conceicao

Igreja N S do Rosario dos Pretos

Igreja Nossa Senhora do Carmo

Igreja Nossa Senhora do Rosario

Igreja Sao Francisco de Paula

Igreja Sta Efigenia dos Pretos

**Mercados de pulgas e feiras livres**

Feira do Largo de Coimbra (Pedra Sabao)

**Minas**

Minas de Ouro

**Museus de historia**

Casa dos Contos

Museu da Inconfidencia

**Museus especializados**

Estacao Ferroviaria de Ouro Preto

Museu Casa dos Inconfidentes

Museu da Mineralogia

Museu de Ciencia e Tecnica da Escola de Minas/UFOP

Museu do Aleijadinho

Museu do Oratorio

**Passeios de trem com paisagem**

Passeio de Maria Fumaca ate Mariana

**Pontos de interesse**

Centro Historico de Ouro Preto

Praca Tiradentes

Rua Conde de Bobadela

**Teatros**

Teatro Municipal

---

Fonte: *TripAdvisor*

Elaboração: próprio autor

Apesar de a maioria estar categorizada de forma coerente, há classificações distintas para a “Mina du Veloso” (formações geológicas) e “Minas de Ouro” (minas) que não se invalidam, mas que poderiam ser agregadas em apenas

uma categoria para uma análise mais precisa da oferta de atrativos do município. De qualquer forma, a divisão possibilitou compreender o tipo de oferta turística contida no município.

## 4.2 Avaliações

Os dados extraídos possibilitaram analisar as informações sobre o percentual de avaliações por municípios e por faixas de satisfação, resumidas em uma nota média. A tabela 3 exemplifica os resultados para o município de Mariana:

Tabela 3: Percentual de avaliações e satisfação dos atrativos do município de Mariana por categorias

| <b>Categorias e atrativos</b>        | <b>Avaliações (%)</b> | <b>Excelente (%)</b> | <b>Muito Bom (%)</b> | <b>Razoável (%)</b> | <b>Ruim (%)</b> | <b>Horrível (%)</b> | <b>Nota</b> |
|--------------------------------------|-----------------------|----------------------|----------------------|---------------------|-----------------|---------------------|-------------|
| <b>Igrejas e catedrais</b>           | <b>31,98</b>          | <b>43,01</b>         | <b>45,03</b>         | <b>10,04</b>        | <b>0,98</b>     | <b>0,94</b>         | <b>4,28</b> |
| Basilica de Sao Pedro dos Clerigos   | 23,83                 | 42,70                | 45,51                | 9,55                | 1,12            | 1,12                | 4,28        |
| Catedral Basilica da Se              | 33,07                 | 43,72                | 45,34                | 8,91                | 1,21            | 0,81                | 4,30        |
| Centro Historico de Mariana          | 26,37                 | 41,62                | 43,65                | 13,71               | 0,00            | 1,02                | 4,25        |
| Igreja N S do Carmo - Mariana        | 16,73                 | 44,00                | 45,60                | 8,00                | 1,60            | 0,80                | 4,30        |
| <b>Minas</b>                         | <b>34,16</b>          | <b>43,61</b>         | <b>39,85</b>         | <b>13,16</b>        | <b>2,76</b>     | <b>0,63</b>         | <b>4,23</b> |
| Minas da Passagem                    | 100,00                | 43,61                | 39,85                | 13,16               | 2,76            | 0,63                | 4,23        |
| <b>Museus de historia</b>            | <b>7,06</b>           | <b>22,42</b>         | <b>49,70</b>         | <b>26,06</b>        | <b>1,82</b>     | <b>0,00</b>         | <b>3,93</b> |
| Casa de Camara e Cadeia              | 100,00                | 22,42                | 49,70                | 26,06               | 1,82            | 0,00                | 3,93        |
| <b>Passeios de trem com paisagem</b> | <b>10,19</b>          | <b>42,86</b>         | <b>37,39</b>         | <b>14,29</b>        | <b>3,36</b>     | <b>2,10</b>         | <b>4,16</b> |
| Maria Fumaca ate Ouro Preto          | 100,00                | 42,86                | 37,39                | 14,29               | 3,36            | 2,10                | 4,16        |
| <b>Pontos de interesse</b>           | <b>16,61</b>          | <b>45,62</b>         | <b>42,08</b>         | <b>11,25</b>        | <b>0,61</b>     | <b>0,43</b>         | <b>4,32</b> |
| Praca Gomes Freire                   | 29,90                 | 37,93                | 44,83                | 15,52               | 0,86            | 0,86                | 4,18        |
| Praca Minas Gerais                   | 70,10                 | 53,31                | 39,34                | 6,99                | 0,37            | 0,00                | 4,46        |
| <b>Mariana</b>                       | <b>100,00</b>         | <b>41,35</b>         | <b>43,47</b>         | <b>12,91</b>        | <b>1,46</b>     | <b>0,82</b>         | <b>4,23</b> |

Fonte: *TripAdvisor*

Elaboração: próprio autor

Pelos dados apresentados, percebe-se que a categoria “minas” representa 34,16% do total de avaliações dos atrativos apresentados. Já a categoria “igrejas e catedrais” representa 31,98% de todas as avaliações. Dentro dessa categoria, é possível observar que a Catedral Basílica da Sé é a que possui maior número de avaliações em comparação com as demais igrejas (33,07%).

É possível observar também que de todos os atrativos do município, a maior nota de satisfação foi alcançada pela Praça Minas Gerais (4,46) e a menor pelo museu histórico Casa de Câmara e Candeia (3,93).

A tabela 4 apresenta os resultados de avaliações e notas para a categoria “museus especializados” de Minas Gerais por município, sendo que o Inhotim representa 35,72% de todas as avaliações realizadas.

Tabela 4 – Percentual de avaliações e nota média da categoria “museus especializados” por município

| <b>Museus e cidades</b>                            | <b>Avaliações</b> | <b>Nota</b> |
|--|-------------------|-------------|
| <b>Belo Horizonte</b>                              | <b>32,53%</b>     | <b>4,42</b> |
| Casa Kubitschek                                    | 2,50%             | 4,42        |
| Memorial Minas Gerais Vale                         | 37,02%            | 4,66        |
| MM Gerdau - Museu das Minas e do Metal             | 23,70%            | 4,54        |
| Museu de Arte da Pampulha                          | 4,83%             | 3,87        |
| Museu De Artes & Ofícios                           | 11,94%            | 4,55        |
| Palacio da Liberdade                               | 20,00%            | 4,48        |
| <b>Brumadinho</b>                                  | <b>35,72%</b>     | <b>4,82</b> |
| Inhotim  | 100,00%           | 4,82        |
| <b>Cordisburgo</b>                                 | <b>0,99%</b>      | <b>4,38</b> |
| Museu Casa Guimaraes Rosa                          | 100,00%           | 4,38        |
| <b>Diamantina</b>                                  | <b>1,02%</b>      | <b>3,72</b> |
| Museu do Diamante                                  | 100,00%           | 3,72        |
| <b>Juiz de Fora</b>                                | <b>2,69%</b>      | <b>4,09</b> |
| Museu Mariano Procopio                             | 100,00%           | 4,09        |
| <b>Ouro Preto</b>                                  | <b>12,00%</b>     | <b>4,42</b> |
| Estacao Ferroviaria de Ouro Preto                  | 9,31%             | 4,19        |
| Museu Casa dos Inconfidentes                       | 8,41%             | 4,56        |
| Museu da Mineralogia                               | 30,21%            | 4,49        |
| Museu de Ciencia e Tecnica da Escola de Minas/UFOP | 14,65%            | 4,51        |
| Museu do Aleijadinho                               | 11,71%            | 4,24        |
| Museu do Oratorio                                  | 25,71%            | 4,53        |
| <b>Pocos de Caldas</b>                             | <b>1,10%</b>      | <b>4,19</b> |
| Museu Historico e Geografico                       | 100,00%           | 4,19        |
| <b>Sao Joao Del Rei</b>                            | <b>1,84%</b>      | <b>4,29</b> |
| Museu Ferroviario                                  | 100,00%           | 4,29        |
| <b>Tiradentes</b>                                  | <b>10,91%</b>     | <b>4,32</b> |
| Museu da Liturgia                                  | 28,14%            | 4,53        |
| Museu de Sant'Ana                                  | 39,50%            | 4,54        |
| Museu do Automovel                                 | 32,36%            | 4,11        |
| <b>Uberaba</b>                                     | <b>1,19%</b>      | <b>4,59</b> |
| Casa da Memoria de Chico Xavier                    | 100,00%           | 4,59        |
| <b>Total Geral</b>                                 | <b>100,00%</b>    | <b>4,35</b> |

Fonte: *TripAdvisor*

Elaboração: próprio autor

Em relação à satisfação dos usuários, percebe-se que o Museu do Diamante em Diamantina alcançou a menor nota. Esse resultado auxilia diretamente o setor público e os gestores dos museus a identificarem onde se encontram os pontos de maior atenção para elaborar propostas de melhoria. Para identificar problemas de satisfação no museu, foi realizada uma análise qualitativa dos comentários no site e percebeu-se que o museu conta com falta de informações, falta de relação do acervo com a extração dos diamantes e problemas de infraestrutura o que gerou insatisfação dos usuários. Já o Museu de Arte da Pampulha foi elogiado pela arquitetura, porém questionado pelos usuários pela falta de exposições e acervo dentro do espaço.

Os resultados possibilitam criar ações que auxiliem na melhoria das condições dos museus de forma pontual, visando melhorar a satisfação dos usuários ao longo do tempo.

Além disso, é possível criar estratégias de promoção visando aumentar o número de avaliações dos museus que possuem uma boa nota de satisfação, mas que possuem baixo percentual de avaliação como é o caso do Museu dos Inconfidentes em Ouro Preto que possui uma nota média de satisfação de 4,56 e contempla apenas 8,34% das avaliações dos museus de Ouro Preto ou o Museu Casa de Guimarães Rosa, em Cordisburgo (0,99% de avaliações entre todos os museus do estado e 4,38 de satisfação).

### **4.3 Perfil dos visitantes**

Os dados permitiram também analisar os municípios mais avaliados por perfil de visitante, conforme exemplo da tabela 5.

Os resultados mostram que os destinos com maior proporção de viagens com famílias são locais relacionados a estâncias termais como o caso de Baependi, São Lourenço, Passa Quatro e Lambari. No caso de Betim, as avaliações foram centralizadas no atrativo “Vale Verde”, caracterizado por ser um parque ecológico e com contato direto com animais, o que favorece o público familiar.

Tabela 5 – Municípios com maior número percentual de avaliações por perfil de visitante

| <b>Município</b>                 | <b>Famílias (%)</b>  |
|----------------------------------|----------------------|
| Baependi                         | 52,83                |
| São Lourenço                     | 52,60                |
| Betim                            | 51,76                |
| Passa Quatro                     | 50,71                |
| Lambari                          | 50,47                |
| <b>Município</b>                 | <b>Romântico (%)</b> |
| Goncalves                        | 68,75                |
| Lavras Novas                     | 64,69                |
| Monte Verde                      | 59,50                |
| Cardeal Mota                     | 44,13                |
| Carrancas                        | 43,31                |
| <b>Município</b>                 | <b>Amigos (%)</b>    |
| Andradas                         | 49,46                |
| São Thomé das Letras             | 43,79                |
| Conceição da Ibitipoca           | 43,74                |
| Parque Nacional da Serra do Cipó | 43,51                |
| Lima Duarte                      | 40,58                |
| <b>Município</b>                 | <b>Negócios (%)</b>  |
| Uberlândia                       | 13,61                |
| Montes Claros                    | 13,29                |
| Uberaba                          | 11,78                |
| Nova Lima                        | 10,81                |
| Araxá                            | 10,09                |
| <b>Município</b>                 | <b>Sozinho (%)</b>   |
| Montes Claros                    | 16,46                |
| Sabará                           | 16,42                |
| Belo Horizonte                   | 15,21                |
| Uberaba                          | 13,23                |
| Uberlândia                       | 13,08                |

Fonte: *TripAdvisor*

Elaboração: próprio autor

Os destinos mais avaliados por amigos são característicos de natureza (cachoeiras, lagos, serras...) tais como Andradas, São Thomé das Letras, Conceição do Ibitipoca, Lima Duarte (Serra do Ibitipoca) e Serra do Cipó. Já os destinos característicos de casais foram destacados em locais de pousadas, ligados à gastronomia e cultura, condizente com o que é divulgado para esse público como nos casos de Gonçalves, Lavras Novas, Cardeal Mota (Serra do Cipó), Monte Verde e Carrancas - cidades já conhecidas por serem atrativas para visitantes em lua de mel.

No caso dos perfis “sozinho” e “negócios” os resultados se assemelham, destacando municípios com vocação para o turismo de negócios e eventos, tal como Uberlândia, Montes Claros, Uberaba e Belo Horizonte.

Aliado a isso, conforme apresentado na tabela 6, percebe-se uma coerência entre as categorias mais visitadas para cada perfil.

Tabela 6 – Principais categorias visitadas pelos tipos de visitantes nos atrativos de Minas Gerais cadastrados no *TripAdvisor*

| <b>Categorias</b>                        | <b>Famílias (%)</b>  |
|--|----------------------|
| Parques temáticos                        | 68,7                 |
| Jardins                                  | 64,5                 |
| Centros de entretenimento e jogos        | 63,1                 |
| Aquários                                 | 57,8                 |
| Museus de ciência                        | 54,2                 |
| <b>Categorias</b>                        | <b>Romântico (%)</b> |
| Excursões de off-road, quadriciclo e 4WD | 66,8                 |
| Canyoning e rapel                        | 63,8                 |
| Barragens                                | 63,3                 |
| Trilhas para caminhada                   | 57,2                 |
| Excursões pela natureza e vida selvagem  | 48,4                 |
| <b>Categorias</b>                        | <b>Amigos (%)</b>    |
| Arenas e estádios                        | 46,8                 |
| Feiras de agricultores                   | 42,0                 |
| Mirantes                                 | 41,0                 |
| Atividades ao ar livre                   | 36,4                 |
| Parques nacionais                        | 34,5                 |
| <b>Categorias</b>                        | <b>Sozinho (%)</b>   |
| Bibliotecas                              | 28,3                 |
| Feiras de agricultores                   | 17,6                 |
| Observatórios e planetários              | 17,0                 |
| Shopping centers                         | 16,7                 |
| Locais religiosos e sagrados             | 15,6                 |
| <b>Categorias</b>                        | <b>Negócios (%)</b>  |
| Centros de conveniências e conferencias  | 27,8                 |
| Sem identificação (Minascentro)          | 25,7                 |
| Teatros                                  | 11,4                 |
| Locais religiosos e sagrados             | 10,4                 |
| Shopping centers                         | 10,1                 |

Fonte: *TripAdvisor*  
Elaboração: próprio autor

No caso de famílias, as categorias com maior número de avaliações foram destinadas para locais de grande fluxo de pessoas, ligados à diversão ou ao conhecimento tais como os casos de parques temáticos, centro de entretenimento e jogos, aquários e museus de ciência. Já as categorias mais avaliadas para perfis românticos ficaram caracterizadas por atrativos na natureza, e esportes de aventura que geralmente são ofertados em municípios com vocação para lua-de-mel, tal como Monte Verde ou Lavras Novas. Na categoria amigos, os destaques foram para atrativos ao ar livre ou focado em eventos, como no caso de estádios ou parques nacionais que possuem um perfil de áreas de camping com um público jovem. No caso de pessoas sozinhas, os destaques foram para biblioteca ou locais de compras ou reflexão, enquanto que para pessoas motivadas a negócios, as categorias destacadas foram relacionadas a locais de eventos e exposições, ou programações possíveis de lazer que podem ser realizadas em um curto espaço de tempo tais como teatro, locais religiosos ou *shopping*.

Essas informações possibilitam elaborar propostas de roteiros ou divulgações desses destinos para públicos específicos, garantindo uma maior eficiência nas promoções realizadas pelo estado.

#### **4.4 Época de visitação**

Os dados possibilitaram visualizar os períodos com o maior número de comentários dos viajantes por intervalo de meses. A tabela 7 apresenta como exemplo dados comparativos de dois destinos de Minas Gerais que possuem vocações turísticas distintas. Carrancas é um município conhecido por ter vocação para turismo de natureza, pela quantidade de pousadas, cachoeiras e atrativos naturais que atraem um público mais jovem. Os dados da tabela 7 apresentam um grande número de avaliações no período entre dezembro e fevereiro que condiz com o período de férias escolares, que pode sugerir uma relação com o público visitante do local. Já o distrito de Monte Verde é vendido no mercado como um destino de inverno por seu clima frio e a quantidade de pousadas focadas no perfil de romance, tendo como pico de avaliações exatamente o período entre os meses de junho e agosto, condizente com os dados apresentados.

Tabela 7 – Percentual de avaliações realizadas por período do ano no município de Carrancas e no distrito de Monte Verde

| Municípios e atrativos                         | Mar-Mai (%)  | Jun-Ago (%)  | Set-Nov (%)  | Dez-Fev (%)  |
|--|--------------|--------------|--------------|--------------|
| <b>Carrancas</b>                               | <b>34,32</b> | <b>13,80</b> | <b>16,76</b> | <b>35,11</b> |
| Cachoeira da Esmeralda                         | 54,62        | 9,23         | 0,00         | 36,15        |
| Cachoeira Da Fumaca                            | 35,77        | 19,51        | 17,07        | 27,64        |
| Carrancas Eco Adventure                        | 35,85        | 13,21        | 16,98        | 33,96        |
| Centro Historico de Carrancas                  | 32,57        | 17,14        | 22,29        | 28,00        |
| Complexo da Zilda                              | 24,06        | 16,04        | 20,28        | 39,62        |
| Poço do Coração                                | 23,08        | 7,69         | 23,93        | 45,30        |
| <b>Monte Verde</b>                             | <b>25,85</b> | <b>32,01</b> | <b>24,12</b> | <b>18,01</b> |
| Casa dos Beija Flores                          | 24,79        | 30,58        | 28,93        | 15,70        |
| Chapeu do Bispo                                | 26,28        | 34,95        | 23,21        | 15,56        |
| Igreja de Sao Francisco de Assis - Monte Verde | 23,97        | 26,71        | 25,34        | 23,97        |
| Pedra Partida                                  | 25,52        | 35,86        | 23,22        | 15,40        |
| Pedra Redonda                                  | 28,90        | 32,14        | 19,69        | 19,27        |
| Pico Selado                                    | 27,04        | 31,60        | 24,76        | 16,61        |
| Trilha do Plato                                | 25,66        | 33,63        | 25,66        | 15,04        |
| Unger's Pottery House                          | 24,65        | 30,63        | 22,18        | 22,54        |
| <b>Total Geral</b>                             | <b>29,48</b> | <b>24,21</b> | <b>20,97</b> | <b>25,34</b> |

Fonte: *TripAdvisor*

Elaboração: próprio autor

Outro exemplo de análise foi feita para as categorias “museus especializados” e “obras arquitetônicas” do município de Belo Horizonte, conforme apresentado na tabela 8.

Pelas informações apresentadas, pode-se sugerir a elaboração de ações promocionais tais como um passeio integrado com descontos ou incentivo de eventos para os atrativos da região da Pampulha (Casa Kubitschek, Museu de Arte da Pampulha, Casa do Baile e Igreja da Pampulha), visando aumento de visitação nos meses entre setembro e novembro. Já no caso dos atrativos da Praça da Liberdade (Memorial Minas Gerais Vale, Museu das Minas e do Metal, Palácio da Liberdade), as ações poderiam ser feitas entre os meses de dezembro a fevereiro, diminuindo a sazonalidade desses atrativos.

Tabela 8 – Percentual de avaliações realizadas nas categorias “museus especializados” e “obras arquitetônicas” por período do ano em Belo Horizonte

| <b>Categorias e atrativos</b>         | <b>Mar-Mai (%)</b> | <b>Jun-Ago (%)</b> | <b>Set-Nov (%)</b> | <b>Dez-Fev (%)</b> |
|---------------------------------------|--------------------|--------------------|--------------------|--------------------|
| <b>Museus especializados</b>          | <b>26,85</b>       | <b>27,77</b>       | <b>22,85</b>       | <b>22,52</b>       |
| Casa Kubitschek                       | 33,63              | 27,43              | 16,81              | 22,12              |
| Memorial Minas Gerais Vale            | 25,85              | 29,38              | 24,18              | 20,59              |
| MM Gerda - Museu das Minas e do Metal | 24,86              | 27,57              | 23,46              | 24,11              |
| Museu de Arte da Pampulha             | 27,98              | 26,61              | 20,64              | 24,77              |
| Museu De Artes & Ofícios              | 23,75              | 30,61              | 24,68              | 20,96              |
| Palácio da Liberdade                  | 25,03              | 25,03              | 27,35              | 22,59              |
| <b>Obras arquitetônicas</b>           | <b>29,46</b>       | <b>25,06</b>       | <b>22,18</b>       | <b>23,31</b>       |
| Casa do Baile                         | 32,47              | 22,51              | 20,35              | 24,68              |
| Cidade Administrativa                 | 27,83              | 22,63              | 23,85              | 25,69              |
| Conjunto Arquitetônico da Pampulha    | 26,38              | 27,12              | 23,25              | 23,25              |
| Edifício Maletta                      | 36,41              | 25,17              | 16,20              | 22,22              |
| Edifício Niemeyer                     | 26,37              | 25,55              | 26,10              | 21,98              |
| Igreja São Francisco De Assis         | 27,27              | 27,37              | 23,33              | 22,02              |
| <b>Total</b>                          | <b>28,15</b>       | <b>26,42</b>       | <b>22,52</b>       | <b>22,92</b>       |

Fonte: *TripAdvisor*

Elaboração: próprio autor

Para uma análise mais completa dos resultados para todo o território, sugere-se o acompanhamento das informações pelo período de um ano, comparando os resultados de um determinado mês pelos resultados do mesmo mês no ano anterior, pois grande parte dos atrativos sofre com períodos sazonais de alta e baixa visitação. Porém, um levantamento mensal da informação pode trazer resultados imediatos de uma intervenção física num atrativo ou ação promocional específica, como foi observado durante a extração de informações do Conjunto Arquitetônico da Pampulha. Durante a realização da pesquisa, o atrativo foi reconhecido como patrimônio mundial da humanidade, possibilitando um acompanhamento desse reconhecimento por um período de 5 meses e que será analisado de forma mais precisa no próximo capítulo por sua particularidade.

## 5 Patrimônio Cultural da Humanidade: o caso do Conjunto Arquitetônico da Pampulha

O Conjunto Arquitetônico da Pampulha foi criado a partir da década de 1940, em Belo Horizonte, com intuito de fomentar o desenvolvimento da capital mineira a partir da criação de diversas edificações e jardins voltados ao lazer e a cultura. Além da lagoa artificial da Pampulha, o complexo possui o antigo Casino (transformado no atual Museu de Arte da Pampulha), o late Clube, a Casa do Baile, a Casa Kubitscheck e a Igreja São Francisco de Assis, todas assinadas pelo arquiteto Oscar Niemeyer (em colaboração com diversos outros artistas tais como Portinari e Burle Marx), sendo considerado um marco da arquitetura moderna no país e que serviu de inspiração para diversas obras ao redor do mundo, a partir da utilização das formas arrojadas e do concreto armado nas suas construções. Sua importância como patrimônio cultural da humanidade foi reconhecida pela *United Nations Educational, Scientific and Cultural Organization* (UNESCO) no dia 17 de julho de 2016<sup>14</sup>.

Essa data favoreceu analisar um possível impacto do desempenho dos atrativos turísticos da região da Pampulha pela metodologia apresentada neste trabalho, pois a primeira extração foi realizada 05 dias antes do anúncio pela UNESCO uma segunda extração foi realizada aproximadamente 20 dias depois. Por fim, realizou-se uma nova extração no dia 14 de dezembro de 2016, ou seja, cinco meses depois do anúncio. O comparativo dos dados entre os três períodos possibilitou avaliar a variação no nível de satisfação e no número de avaliações no *TripAdvisor* e um possível impacto do anúncio nos resultados alcançados

Ao comparar o número de avaliações de todos os 50 atrativos cadastrados em Belo Horizonte, percebeu-se que, em média, cada atrativo aumentou o valor entre julho e agosto em 5,95%. O atrativo “Lagoa da Pampulha” foi exatamente o que obteve a maior variação entre todos, atingindo um aumento de 14,36% no número de avaliações. Considera-se também que três atrativos da região estiveram entre os 11 com maior taxa de variação para o período (Parque Guanabara, Igreja São Francisco de Assis e Conjunto Arquitetônico da Pampulha), conforme tabela 9.

---

<sup>14</sup> Fonte: <http://whc.unesco.org/en/news/1528/> Acesso em: 06 de ago. de 2016.

Tabela 9 – Número de avaliações absolutas e a variação percentual dos principais atrativos de Belo Horizonte entre os meses de julho e agosto de 2016 no *TripAdvisor*

| Atrativo                                  | Avaliações Jul. | Avaliações Ago. | Variação (%) |
|---|-----------------|-----------------|--------------|
| <b>Lagoa da Pampulha</b>                  | <b>926</b>      | <b>1059</b>     | <b>14,36</b> |
| Teatro Topázio                            | 151             | 167             | 10,60        |
| Edifício Maletta                          | 747             | 821             | 9,91         |
| <b>Parque Guanabara</b>                   | <b>534</b>      | <b>585</b>      | <b>9,55</b>  |
| Edifício Niemeyer                         | 364             | 398             | 9,34         |
| Praça Israel Pinheiro                     | 357             | 390             | 9,24         |
| Museu de Ciências Naturais                | 138             | 150             | 8,70         |
| Mercado Central de Belo Horizonte         | 6632            | 7163            | 8,01         |
| Lagoa Santa                               | 208             | 223             | 7,21         |
| <b>Igreja Sao Francisco De Assis</b>      | <b>2057</b>     | <b>2198</b>     | <b>6,85</b>  |
| <b>Conjunto Arquitetônico da Pampulha</b> | <b>2168</b>     | <b>2316</b>     | <b>6,83</b>  |
| <b>Belo Horizonte</b>                     | <b>48155</b>    | <b>51018</b>    | <b>5,95</b>  |

Fonte: *TripAdvisor*

Elaboração: próprio autor

Ao compararmos os dados entre julho e dezembro (tabela 10), percebe-se que a Lagoa da Pampulha também foi o atrativo que obteve a maior variação dentre os 50 pesquisados, crescendo em 57,67% no número de avaliações, acima do dobro da média de crescimento dos demais atrativos da cidade (22,04%). Percebe-se também que outros dois atrativos que estão incluídos no complexo reconhecido apareceram no 2º. e 9º. lugar (Casa Kubitschek e Casa do Baile).

Tabela 10 – Número de avaliações absolutas e a variação percentual dos principais atrativos de Belo Horizonte entre os meses de julho e dezembro de 2016 no *TripAdvisor*

| Atrativo                          | Avaliações Jul. | Avaliações Dez. | Variação (%) |
|-----------------------------------|-----------------|-----------------|--------------|
| <b>Lagoa da Pampulha</b>          | <b>926</b>      | <b>1460</b>     | <b>57,67</b> |
| <b>Casa Kubitschek</b>            | <b>113</b>      | <b>161</b>      | <b>42,48</b> |
| Teatro Topázio                    | 151             | 209             | 38,41        |
| Edifício Maletta                  | 747             | 990             | 32,53        |
| Museu de Ciências Naturais        | 138             | 181             | 31,16        |
| Shopping Diamond Mall             | 382             | 499             | 30,63        |
| Minascentro                       | 771             | 1003            | 30,09        |
| Mercado Central de Belo Horizonte | 6632            | 8619            | 29,96        |
| <b>Casa do Baile</b>              | <b>231</b>      | <b>297</b>      | <b>28,57</b> |
| Parque Guanabara                  | 534             | 685             | 28,28        |
| Praça Israel Pinheiro             | 357             | 451             | 26,33        |
| <b>Belo Horizonte</b>             | <b>48155</b>    | <b>58770</b>    | <b>22,04</b> |

Fonte: *TripAdvisor*

Elaboração: próprio autor

É interessante ressaltar que os atrativos da Pampulha destacados entre julho e agosto se localizam próximos entre eles, o que poderia facilitar uma visita nos atrativos pelo turista que não tinha conhecimento dos demais atrativos da região (que foram contemplados no reconhecimento). Já nos atrativos da Pampulha destacados entre julho e dezembro, percebe-se um crescimento maior nos atrativos ligados à proposta de visita do Complexo da Pampulha, que podem ter sido mais reconhecidos pelos visitantes a partir do anúncio, ou seja, ganharam maior relevância entre os demais atrativos da própria região.

Ao analisarmos o valor da nota média das avaliações, outros resultados também se destacaram para a região da Pampulha. Conforme demonstrado na tabela 11, em média, os 50 atrativos turísticos de Belo Horizonte tiveram um desempenho de 0,04% melhor no mês de agosto em relação ao mês de julho de 2016.

Tabela 11 – Nota média das avaliações e a variação percentual dos principais atrativos de Belo Horizonte entre os meses de julho e agosto de 2016

| Atrativos                                 | Nota média<br>Jul. | Nota média<br>Ago. | Variação (%) |
|---|--------------------|--------------------|--------------|
| <b>Casa do Baile</b>                      | <b>4,00</b>        | <b>4,03</b>        | <b>0,60</b>  |
| <b>Parque Guanabara</b>                   | <b>3,96</b>        | <b>3,98</b>        | <b>0,46</b>  |
| Lagoa Santa                               | 4,18               | 4,20               | 0,46         |
| Praça Raul Soares                         | 3,22               | 3,23               | 0,44         |
| <b>Conjunto Arquitetônico da Pampulha</b> | <b>4,38</b>        | <b>4,40</b>        | <b>0,37</b>  |
| <b>Lagoa da Pampulha</b>                  | <b>4,26</b>        | <b>4,28</b>        | <b>0,37</b>  |
| Teatro Topázio                            | 4,46               | 4,47               | 0,35         |
| <b>Museu de Arte da Pampulha</b>          | <b>3,87</b>        | <b>3,88</b>        | <b>0,35</b>  |
| Minascentro                               | 4,25               | 4,26               | 0,26         |
| <b>Igreja São Francisco De Assis</b>      | <b>4,44</b>        | <b>4,45</b>        | <b>0,23</b>  |
| <b>Belo Horizonte</b>                     | <b>4,26</b>        | <b>4,26</b>        | <b>0,04</b>  |

Fonte: *TripAdvisor*

Elaboração: próprio autor

Se analisarmos os 10 atrativos com as maiores taxas de crescimento, 06 deles possuem ligação direta com a região da Pampulha (destacados em negrito). É interessante ressaltar que no espaço de um mês após o anúncio pela UNESCO, não houve investimento de infraestrutura ou melhoras significativas na região que fizesse com que a variação se tornasse positivamente acima da média. Pode-se sugerir que apenas o anúncio fez com que a valorização dos atrativos da região por parte dos visitantes tenha aumentado. Ao analisar de forma individual os comentários dos

atrativos no site *TripAdvisor*, notou-se nitidamente referências à conquista do reconhecimento pelo Conjunto Arquitetônico da Pampulha.

O mesmo pode ser observado na comparação dos dados entre julho e dezembro conforme tabela 12. Dos 10 primeiros atrativos com maior variação positiva na nota média de satisfação, 06 deles são atrativos da região da Pampulha, reforçando a ideia de que o reconhecimento da UNESCO pode ter influenciado diretamente nos resultados, mesmo em um período mais longo de análise.

Tabela 12 – Nota média das avaliações e a variação percentual dos principais atrativos de Belo Horizonte entre os meses de julho e dezembro de 2016

| Atrativos                                      | Nota média Jul. | Nota média Ago. | Variação (%) |
|--|-----------------|-----------------|--------------|
| <b>Casa do Baile</b>                           | <b>4,01</b>     | <b>4,07</b>     | <b>1,46</b>  |
| Lagoa Santa                                    | 4,18            | 4,24            | 1,40         |
| Cidade Administrativa                          | 3,89            | 3,93            | 1,13         |
| <b>Lagoa da Pampulha</b>                       | <b>4,27</b>     | <b>4,30</b>     | <b>0,90</b>  |
| <b>Conjunto Arquitetônico da Pampulha</b>      | <b>4,39</b>     | <b>4,42</b>     | <b>0,74</b>  |
| <b>Parque Guanabara</b>                        | <b>3,97</b>     | <b>4,00</b>     | <b>0,73</b>  |
| Centro Cultural Banco do Brasil Belo Horizonte | 4,58            | 4,61            | 0,62         |
| <b>Museu de Arte da Pampulha</b>               | <b>3,87</b>     | <b>3,90</b>     | <b>0,61</b>  |
| <b>Casa Kubitschek</b>                         | <b>4,42</b>     | <b>4,44</b>     | <b>0,57</b>  |
| Edifício Maletta                               | 4,05            | 4,07            | 0,49         |
| <b>Belo Horizonte</b>                          | <b>4,26</b>     | <b>4,27</b>     | <b>0,20</b>  |

Fonte: *TripAdvisor*

Elaboração: próprio autor

A metodologia possibilita analisar de forma mais ampla os resultados, mês a mês, para aprofundar a discussão em trabalhos futuros e checar se de fato, há um impacto direto do reconhecimento da UNESCO no patrimônio em questão, principalmente ao compararmos com demais atrativos do município. Preliminarmente é interessante ressaltar que há uma sinalização positiva dos visitantes em relação a uma maior valorização do patrimônio, mesmo sem alterações significativas na estrutura turística do local. Espera-se que em longo prazo, haja maiores intervenções do poder público no espaço, com fins a manter o reconhecimento e atrair mais visitantes para a cidade e o acompanhamento da evolução dos dados poderá servir como uma ferramenta para avaliar a eficiência de ações de promoção ou melhoria estrutural.

Após uma análise descritiva das informações extraídas e das possibilidades de cruzamento de dados, checkou-se a viabilidade da criação de

indicadores para auxiliar os gestores públicos na tomada de decisão, o que será abordado no próximo tópico.

## 6 Criação de indicadores

Após a análise dos dados, é possível estabelecer uma proposta de indicadores de desempenho que possam auxiliar os gestores das políticas de turismo de Minas Gerais a monitorar a evolução dos atrativos turísticos do estado na plataforma *TripAdvisor*.

A construção das fichas técnicas de cada indicador foi adaptada da metodologia utilizada pelo Governo de Minas Gerais na construção do “Caderno de Indicadores” que teve sua publicação impressa feita no ano de 2011 e foi transformado num site interativo de monitoramento de dados em 2013<sup>15</sup> para consulta customizada dos principais indicadores monitorados. A publicação tinha como objetivo “apresentar os indicadores de resultados das políticas públicas estaduais” (MINAS GERAIS, 2011), no qual apresentava 108 indicadores de diferentes áreas que eram descritos em 80 fichas técnicas. Após 2013, ainda não houve uma publicação consolidada pelo Governo de Minas que enfatize de forma direta o monitoramento de indicadores, sendo os mesmos definidos de forma apenas física e financeira constantes no Plano Plurianual de Ação Governamental (PPAG)<sup>16</sup>.

Acredita-se conforme os objetivos dessa pesquisa, a apresentação de uma proposta de indicadores baseada nas fichas técnicas do Caderno de Indicadores se torna suficiente para o monitoramento, visto seu maior nível de detalhes em relação à forma de apresentação dos indicadores no PPAG. Os atributos utilizados para a ficha técnica de cada indicador são (MINAS GERAIS, 2011):

- Título: nome do indicador;
- Descrição: tradução simples do indicador, descrevendo ou conceituando todas as variáveis que o compõem;
- Fórmula de cálculo: fórmula matemática, com a devida tradução de cada um dos termos utilizados;

---

<sup>15</sup> O site denominado “Minas em Números” pode ser acessado pelo seguinte endereço: [www.numeros.mg.gov.br](http://www.numeros.mg.gov.br)

<sup>16</sup> De acordo com o site oficial da Secretaria de Planejamento e Gestão de Minas Gerais (SEPLAG), o PPAG: “O PPAG define qual será o escopo de atuação do Estado para um período de quatro anos, ou seja, define os programas e ações de governo, com suas respectivas metas físicas e orçamentárias, que serão executados durante esse período”. Disponível em: <http://www.planejamento.mg.gov.br/planejamento-e-orcamento/plano-plurianual-de-acao-governamental>. Acesso em: 26 de jan. 2017.

- Polaridade: direção desejada para a medida em questão. Para medidas que se deseja aumentar, “maior melhor” e para medidas que se deseja reduzir, “menor melhor”;
- Aplicação: justificativa para a escolha do indicador, o que ele mensura e a forma como é utilizado, assim como sua especificidade.
- Limites e limitações: referem-se ao uso do indicador e às restrições à sua aplicação.

É importante destacar que alguns elementos constantes na ficha técnica do Caderno de Indicadores não foram utilizados neste trabalho, destacados a seguir:

- Fonte de informação: todas os indicadores sugeridos possuem a mesma fonte de dados, o *site TripAdvisor*;
- Periodicidade: visto que a sazonalidade é um item de grande relevância para a atividade turística, espera-se que os indicadores propostos sejam analisados em um intervalo de um ano;
- Defasagem: pode ocorrer uma defasagem de tempo entre a captura de informações da ferramenta e a elaboração dos indicadores, porém, como a premissa é utilizar de forma quase imediata os dados coletados, acredita-se que mesmo pequena, possa ser de difícil mensuração para uma ficha técnica.

A proposta de monitoramento se limitará a descrever possíveis indicadores e a forma de coleta e cálculo dos mesmos, visando um futuro uso após um ano de monitoramento para elaboração de metas de curto, médio e longo prazo. Como o turismo é realizado de forma sazonal, acredita-se que esse intervalo é o mais indicado para o monitoramento dos atrativos para Minas Gerais como um todo. Espera-se que os primeiros resultados extraídos em Julho de 2016 possam servir como linha de base para o acompanhamento em Julho de 2017, 2018 e assim em diante.

Para a apresentação da proposta, foram levados em consideração indicadores que atendessem os atuais objetivos e estratégias para o turismo do

estado, contidos no Plano Mineiro de Desenvolvimento Integrado (PMDI) – 2016 a 2027. O PMDI “tem como propósito fundamental fixar as estratégias para a promoção do desenvolvimento econômico e social de longo prazo de Minas Gerais” (MINAS GERAIS, 2016), determinado pelo artigo 231 da Constituição Estadual. Os indicadores propostos visam auxiliar no monitoramento de dados para alcance dos seguintes objetivos e estratégias definidas para o turismo no PMDI:

- Objetivos estratégicos:
  - Aumentar a competitividade das regiões turísticas do Estado;
  - Consolidar a identidade turística do Estado de Minas Gerais nos mercados nacional e internacional;
- Estratégias prioritárias:
  - Definir e consolidar a identidade turística de Minas Gerais para o posicionamento nos mercados nacional e internacional;
  - Melhorar a infraestrutura dos atrativos e destinos turísticos.
- Estratégias complementares:
  - Promover a excelência na prestação de serviços turísticos;
  - Promover os destinos turísticos prioritários do Estado junto aos principais mercados emissores, nacional e internacional;
  - Qualificar, consolidar e divulgar as fontes de dados do turismo para subsidiar as decisões estratégicas tanto de mercado quanto de governo.

A partir desses fatos, serão apresentados os indicadores baseados na extração de informações do *TripAdvisor*.

### **1 – Número de avaliações nos atrativos de Minas Gerais**

**Descrição:** o indicador tem como objetivo mensurar a variação no número de avaliações absolutas em todos os atrativos de Minas Gerais dentro da plataforma *TripAdvisor*.

**Fórmula de cálculo:**  $AvaAtrat = AvaTotal$

Em que  $AvaAtrat$  = número de avaliações nos atrativos do estado e  $AvaTotal$  = número absoluto de todas as avaliações dos atrativos do estado.

**Valor de referência:** 122.076

**Polaridade:** Maior melhor

**Aplicação:** o indicador visa mensurar possíveis ações de promoção dos atrativos turísticos de Minas Gerais que incentivem a utilização do *TripAdvisor* como ferramenta de avaliação pelos visitantes, com o intuito de aumentar a quantidade do número de avaliações do estado na plataforma, e conseqüentemente a relevância dos destinos mineiros em comparação aos demais destinos nacionais no *site*.

**Limites e limitações:** O indicador não visa monitorar o aumento da qualidade ou satisfação dos visitantes nos atrativos. São considerados na análise apenas os atrativos extraídos a partir da linha de base e do mês e ano de referência da primeira extração, ou seja, se um novo atrativo surgir com alto número de avaliações durante um novo período de análise, ele não será considerado para fins de análise da série histórica. Como o indicador representa a totalidade do estado, seu valor não suprime a necessidade de avaliar o mesmo indicador para cada município, categoria ou atrativo, caso seja necessária uma apuração mais precisa de um possível impacto positivo ou negativo no indicador geral. Se um atrativo constante da linha de base perder o cadastro na plataforma, o valor da linha de base pode ser revisto para fins de comparação dos resultados. Há que se considerar de forma hipotética que o *TripAdvisor* pode perder sua base de usuários futuramente por qualquer motivo (tal como fechamento do *site*, migração de usuários para outras plataformas), o que pode refletir no aumento do número de avaliações absolutas.

## **2 – Média de satisfação nos destinos turísticos de Minas Gerais**

**Descrição:** o indicador pretende transformar as avaliações realizadas pelos usuários do *TripAdvisor* em uma nota única, variando na escala de 1 a 5 pontos, sendo 1 a nota mais baixa e 5 a nota mais alta, facilitando o seu acompanhamento. Desta forma, cada avaliação

realizada como “excelente” é pontuada com uma nota 5, “muito bom” é avaliada como nota 4, “razoável” como nota 3, “ruim” como nota 2 e “péssimo” como nota 1. A partir de então, é tirada a média das notas de cada município a partir das médias das notas de cada atrativo do município. Em seguida, o valor geral de Minas Gerais será a média das notas de todos os municípios.

**Fórmula de cálculo:** 
$$\text{Satisf} = \frac{(\text{not}_{m1} + \text{not}_{m2} + \dots + \text{not}_{mn})}{m\text{Total}}$$

Em que Satisf = média da satisfação nos destinos turísticos de Minas Gerais;  $\text{not}_{m1}$  = nota média do município 1 qualquer;  $\text{not}_{m2}$  = nota média do município 2 qualquer;  $\text{not}_{mn}$  = nota média do município n qualquer e  $m\text{Total}$  = número total de municípios

**Valor de referência:** 4,29

**Polaridade:** Maior melhor

**Aplicação:** o indicador possibilita checar o nível de satisfação geral dos usuários do *TripAdvisor* nos atrativos de Minas Gerais. Possibilita monitorar se uma política de estruturação de atrativos, promoção ou melhoria de infraestrutura está impactando de forma direta na qualidade da oferta de atrativos dos destinos mineiros.

**Limites e limitações:** Como o indicador representa a totalidade do estado, seu valor não suprime a necessidade de avaliar o mesmo indicador para cada município, categoria ou atrativo, caso seja necessária uma apuração mais precisa de um possível impacto positivo ou negativo no indicador geral. Se um atrativo constante da linha de base perder o cadastro na plataforma, o valor da linha de base pode ser revisto para fins de comparação dos resultados.

### **3 – Percentual de avaliações nos destinos menos avaliados**

**Descrição:** o indicador tende a mensurar o percentual de avaliações nos destinos com menor número de avaliações do *TripAdvisor* em

relação aos destinos com maior número de avaliações. Percebe-se que o número médio de avaliações para cada destino na plataforma é de 2.654, porém apenas 7 destinos estão acima desse valor, concentrando 77,83% de todas as avaliações. Os demais 39 municípios estão abaixo da média, concentrando apenas 22,17% das avaliações.

**Fórmula de cálculo:** 
$$\text{DestMenos} = \frac{\text{AvaMenos} \times 100}{\text{AvaTotal}}$$

Em que DestMenos = Percentual de avaliações nos destinos menos avaliados; AvaMenos = total de avaliações absolutas dos destinos menos avaliados e AvaTotal = total de avaliações absolutas de todos os destinos.

**Valor de referência:** 22,17%

**Polaridade:** Maior melhor

**Aplicação:** esse indicador tem como objetivo mensurar um possível impacto de ações para descentralização e diversificação da oferta de produtos e serviços para o maior número de municípios do estado. É interessante que o benefício da atividade turística não fique apenas concentrado em determinados municípios, englobando o maior número de destinos possíveis.

**Limites e limitações:** são considerados como destinos mais avaliados: Belo Horizonte, Ouro Preto, Poços de Caldas, Tiradentes, Brumadinho, Monte Verde, São Lourenço. Os municípios menos avaliados são: Uberlândia, Mariana, São Thomé das Letras, São João del-Rei, Juiz de Fora, Diamantina, Capitólio, Lavras Novas, Sete Lagoas, Carrancas, Caxambu, Parque Nacional da Serra do Cipó, Uberaba, Ipatinga, Conceição da Ibitipoca, Lima Duarte, Lagoa Santa, Cordisburgo, Betim, Santa Bárbara, Passa Quatro, Governador Valadares, Monte Sião, Caeté, Cardeal Mota, Montes Claros, Contagem, Alto Caparaó, Gonçalves, Sabará, Pouso Alegre, Extrema, Lambari, Nova Lima, Baependi, Divinópolis, Araxá, Andradas, Carmo do Rio Claro. Se um atrativo constante da linha de base perder o cadastro na plataforma, o

valor da linha de base pode ser revisto para fins de comparação dos resultados.

#### **4 – Percentual de avaliações durante a baixa temporada em Minas Gerais (setembro a fevereiro)**

**Descrição:** o indicador permite identificar a variação do número de avaliações durante os meses de baixa temporada em Minas Gerais na plataforma *TripAdvisor*. Os dados extraídos para a linha de base mostram que 55,27% das avaliações são feitas no período entre março e agosto, sendo considerado como alta temporada e 44,73% entre os meses de setembro a fevereiro, sendo considerado como baixa temporada. É interessante para Minas Gerais, que os valores das duas temporadas se aproximem, garantindo uma maior distribuição dos visitantes durante todo o ano, diminuindo o impacto da sazonalidade no estado.

**Fórmula de cálculo:**  $BaixaTemp = \frac{AvaBaixa \times 100}{AvaTotal}$

Em que  $BaixaTemp$  = Percentual de avaliações durante a baixa temporada em Minas Gerais;  $AvaBaixa$  = total de avaliações absolutas durante os meses de setembro a fevereiro e  $AvaTotal$  = total de avaliações absolutas de todos os destinos.

**Valor de referência:** 44,73%

**Polaridade:** Maior melhor (até o teto de 50%)

**Aplicação:** o indicador possibilita criar estratégias para aumentar o fluxo de visitantes no estado durante os períodos de menor visitação, através da criação de roteiros integrados, passaportes de descontos em atrativos de um mesmo município, investimento em campanhas de divulgação com vistas à diminuir o impacto da sazonalidade. Quanto mais igual a distribuição dos visitantes durante todos os períodos do ano, mais estável é o mercado de prestação de serviços turísticos,

garantindo lucro e renda para a população de forma uniforme no ano pesquisado.

**Limites e limitações:** Como o indicador representa a totalidade do estado, seu valor não suprime a necessidade de avaliar o mesmo indicador para cada município, categoria ou atrativo, caso seja necessária uma apuração mais precisa de um possível impacto positivo ou negativo no indicador geral. Se um atrativo constante da linha de base perder o cadastro na plataforma, o valor da linha de base pode ser revisto para fins de comparação dos resultados.

Deve-se destacar que a proposta de indicadores pode ser ampliada a partir da necessidade de acompanhamento dos dados e da elaboração de novos projetos. Pode-se, por exemplo, criar indicadores que comparem o desempenho dos patrimônios mineiros reconhecidos pela UNESCO em relação aos demais patrimônios nacionais ou com patrimônios internacionais que possuam características semelhantes aos mineiros. Além disso, podem-se propor indicadores que monitorem a evolução de determinados municípios ou atrativos de Minas Gerais em relação a municípios ou atrativos brasileiros, tornando-se uma importante ferramenta para análise do mercado concorrente.

## 7 Considerações futuras

Conforme destacado no primeiro capítulo, há uma discussão enfrentada pela sociedade sobre a utilização dos dados relacionados à *Big Data* e sobre a disponibilização das informações das empresas privadas de tecnologia para o público interessado. Para instituições públicas, mensurar o impacto do turismo num determinado destino se torna fundamental e uma das formas de diagnosticar a oferta presente no *site* é pela extração de informações, organizando-as num banco de dados possível de análise por pesquisadores.

Tem-se como exemplo o principal *site* de reserva de hospedagem alternativa no mundo, conhecido como “AirBNB”, no qual diversos usuários podem alugar seus próprios quartos para outros visitantes, impactando diretamente no mercado de hotéis de diversos destinos. Isso acontece, porque o *site* pode aumentar a oferta de quartos em um destino quase que instantaneamente e a custo zero, sendo que o custo de ampliação do número de quartos em um hotel regular é muito superior e demorado, criando uma concorrência diferenciada para o setor hoteleiro (ZERVAS; PROSEPIO; BYERS, 2014).

A partir de uma busca de trabalhos correlatos na *internet* não é raro encontrar pesquisas que utilizam o *scraping* em sites de turismo como forma de extração de informações para fins de pesquisa. Schmücker, Sonntag e Wagner (2016, p.2, tradução nossa), utilizaram o *scraping* como metodologia para avaliar o impacto do *AirBNB* em alguns destinos da Europa e afirmaram que “as informações foram extraídas do *site AirBNB* sem o consentimento da empresa”, pois “há evidências de que as informações providas pela própria plataforma de hospedagem podem ser manipuladas (tradução nossa)”, não gerando dados confiáveis. Essa mesma pesquisa foi apresentada em comunicação oral no dia 25 de novembro durante o *14th Global Forum on Tourism Statistics*, realizado na cidade de Veneza<sup>17</sup>, e foi relatado por Sonntag (2016) que por diversas vezes os autores tentaram entrar em contato com a equipe do *AirBNB*, sendo impossível obter qualquer informação sobre a disponibilização de dados. Afirmou-se também que a realização de parcerias entre *sites* e órgãos públicos para obtenção de informações no setor de turismo era muito difícil pela necessidade das empresas de tentar vender as informações de seus bancos de dados.

---

<sup>17</sup> O autor deste trabalho esteve presente no evento, no qual teve acesso direto à informação.

No mesmo evento, foram apresentados artigos afirmando que os serviços dessas plataformas influenciam diretamente na composição da cadeia produtiva do turismo e em comportamentos sociais como acarretar na diminuição de contatos entre vizinhos, aumento de aluguéis em áreas turísticas e migração de moradores para as periferias no caso de *sites* de hospedagem alternativa e a diminuição do número de usuários em viagens de trens pela Europa a partir de *sites* de carona (WEISS, 2016; DEL AMO, HURTADO, 2016; LO MONACO, 2016).

A necessidade ou não de autorização para a realização de pesquisas com utilização de *scraping* ainda é uma incógnita. Para Munzert et al (2015, p. 278), mesmo para juristas que são dedicados às causas sobre internet matérias sobre *web scraping* são difíceis de lidar. Os autores afirmam que em todos os casos de litígio jurídicos sobre a questão envolviam uso comercial da informação, mas mesmo assim “não existe clareza sobre quais são as ações sobre a égide do *web scraping* que podem ser consideradas como legais ou ilegais (MUNZERT et al. 2015, p. 279)”.

Ainda é necessário um maior entendimento sobre a melhor forma de disponibilização de dados entre as plataformas digitais e as instituições interessadas nesses dados, possibilitando compreender melhor os impactos da economia digital e compartilhada no setor de turismo. Os serviços dessas plataformas influenciam diretamente na composição da cadeia produtiva do turismo, tornando essenciais pesquisas que possibilitem ao poder público entender esses fenômenos.

Espera-se que os resultados desse trabalho possam, inclusive, auxiliar nesse debate.

## 8 Conclusão

O fenômeno *Big Data* alterou de diversas formas a relação entre os indivíduos e as fontes de informação, principalmente em meios digitais. O grande volume, velocidade e variedade de dados disponíveis fizeram com que pesquisas relacionadas a esse fenômeno ganhassem interesse em várias áreas destacando a acadêmica, principalmente na Ciência da Informação.

A partir de então, a busca por ferramentas de gestão de informação torna-se essencial para qualquer entidade, seja privada ou pública, garantindo elaboração de estratégias que visem atingir públicos-alvo, além de monitorar o sucesso ou fracasso de produtos e serviços.

No setor de turismo, existem vários *sites* que consolidam informações sobre destinos e atrativos turísticos, baseados na experiência dos próprios usuários e trabalhar esses dados se torna um desafio para os gestores do setor.

O estudo apresentado demonstra a viabilidade da extração de informações disponibilizadas em formato digital para a elaboração de indicadores que podem ser utilizados pela administração pública, com vistas a monitorar programas e ações a partir da satisfação dos usuários, respondendo positivamente ao problema de pesquisa apresentado. É importante ressaltar que a metodologia apresentada não tem como objetivo substituir integralmente as formas tradicionais de levantamento de dados já utilizadas, mas servir como uma pesquisa complementar para auxiliar na elaboração de indicadores e facilitar tomada de decisão de gestores públicos e privados. Sugere-se também a combinação dos resultados apresentados com demais fontes de informações sobre o turismo já disponíveis, buscando a criação de análises e indicadores mais integrados, trazendo de forma mais precisa a realidade da visitação nos espaços turísticos.

Ilustrou-se (por um estudo de caso instrumental), como as informações extraídas no *site TripAdvisor* podem servir para compreender o comportamento dos usuários da plataforma em atrativos turísticos do estado. Os dados demonstram que as informações extraídas são compatíveis com o que é observado no mercado de turismo em Minas Gerais, corroborando que as informações disponíveis no *site* se aproximam da realidade vista nos atrativos.

Com uma análise mais profunda realizada sobre o Conjunto Arquitetônico da Pampulha, observou-se que a metodologia possibilitou comparar os resultados antes e depois do reconhecimento do conjunto como patrimônio mundial da humanidade e que há indícios de que esse reconhecimento aumentou instantaneamente a percepção de satisfação dos visitantes dentro do *TripAdvisor*, mesmo sem grandes investimentos do setor público para a melhoria do quadro.

As extrações permitiram analisar dados referentes à oferta de atrativos, satisfação e perfil dos visitantes e também a época de visitação mais pertinente para cada local em Minas Gerais. Para facilitar o monitoramento em longo prazo, propõem-se quatro indicadores que consolidam essas informações: número de avaliações nos atrativos de Minas Gerais, média de satisfação nos destinos turísticos de Minas Gerais, percentual de avaliações nos destinos menos avaliados e o percentual de avaliações durante a baixa temporada em Minas Gerais (setembro a fevereiro). Deve-se enfatizar que a visitação nos atrativos e destinos turísticos ocorre de forma sazonal, e para uma melhor comparação das informações, sugere-se continuar as extrações no prazo de um ano, pois assim será possível comparar os resultados em um universo considerando todas as temporadas (tais como alta, média e baixa) que podem variar de local para local.

Espera-se que a pesquisa seja utilizada por municípios, possibilitando com que eles possam fazer seu próprio monitoramento de informações em um espaço de tempo mais curto (mensal, trimestral, semestral), facilitando identificar os gargalos apontados pelos usuários do *TripAdvisor* e criar estratégias para mitigação desses apontamentos de forma mais ágil.

Já em relação ao setor privado, os resultados podem servir como ferramenta de subsídio para cobrar melhorias de infraestrutura aos órgãos públicos (tais como acesso, iluminação pública) e para que possam criar ações diversas (tais como promoções, melhoria nos serviços prestados, estrutura de apoio, divulgação dos estabelecimentos) para aumento de visitantes e na qualidade da experiência dos usuários nos espaços. A coleta de informações a partir da metodologia apresentada pode ser expandida para páginas de conteúdo e avaliação de bares, restaurantes, hotéis, agências de viagens, dentre outros segmentos do turismo.

Percebe-se também que ainda há um embate sobre a disponibilização de informações contidas em plataformas digitais para utilização de órgãos públicos,

visto que esses serviços estão impactando cada dia mais na estrutura social de destinos turísticos. Entender esses impactos se torna uma peça-chave para auxiliar os gestores na tomada de decisão com vistas a elaborar estratégias de desenvolvimento do turismo. Sugere-se que uma das propostas de pesquisas futuras baseadas no estudo deste trabalho que pode auxiliar nessa questão seja a criação de *sites* de avaliação de destinos e serviços turísticos pelo próprio poder público ou organizações não-governamentais, desvinculando a dependência de coleta de informações apenas por *sites* privados, possibilitando ainda o uso de informações por qualquer usuário interessado nos dados de forma transparente.

Acredita-se que essa pesquisa contribui para ampliar esse debate e espera-se que incentive ou sirva para incentivar demais pesquisadores na busca por métodos alternativos de captura, análise e gestão da informação para fins estratégicos, beneficiando não só o setor de turismo, mas demais áreas que ainda necessitam criar formas inovadoras para a solução de problemas informacionais. A pesquisa permite replicar a metodologia em diversos sites, caso haja necessidade, tais como nas áreas de saúde, segurança ou educação (inclusive no setor acadêmico como plataformas de currículos e grupos de pesquisa) expandindo as possibilidades de utilização da ferramenta para o setor público.

Por fim, espera-se que discussões na Ciência da Informação possam elucidar maneiras de enfrentar esses problemas, não só a partir da criação de soluções tecnológicas, mas principalmente no fomento ao debate sobre o acesso à informação digital num mundo caracterizado cada vez mais pela dinâmica da informação no dia-a-dia dos indivíduos e sua relação com a sociedade.

## Referências

ANDREJEVIC, Mark. The Big Data divide. **International Journal of Communication**, nº 8, 2014.

ARAÚJO; Vânia Maria Rodrigues Hermes de. Sistemas de informação: nova abordagem teórico-conceitual. **Ciência da Informação**, Brasília, v.24, n. 1, p. 54-76, 1995.

BARACHO, R. M. A. **Sistema de recuperação de informação visual em desenhos técnicos de engenharia e arquitetura**: modelo conceitual, esquema de classificação e protótipo. Tese de Doutorado. Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2007.

BARBIERI, Carlos. **Uma visão sintética e comentada do Data Management Body of Knowledge (DMBOK)**. Fumsoft - Belo Horizonte, 2013.

BASTOS, Jaime S. Y. Utilização de fontes de informação por executivos do setor de tecnologia da informação no Brasil. **Revista Fonte (PRODEMGE)**, Belo Horizonte, ano 3, n. 5, p.113-121, 2006.

BEUREN, Ilse Maria. **Gerenciamento da informação**: um recurso estratégico no processo de gestão empresarial. Editora Atlas. São Paulo, 1998.

BOYD, Danah; CRAWFORD Kate. Critical questions for Big Data. **Information, Communication & Society**, v.15, nº5, p.662-679, 2012.

BRASIL. Ministério do Turismo. **Índice de competitividade do turismo nacional**: destinos indutores do desenvolvimento turístico regional: relatório Brasil 2014. Brasília, 2014.

CACHO, Andréa do Nascimento Barbosa; DE AZEVEDO, Francisco Fransualdo. O turismo no contexto da sociedade informacional. **Revista Brasileira de Pesquisa em Turismo**, v. 4, n. 2, 2010.

CARDOSO JÚNIOR, José Celso. **Monitoramento estratégico de políticas públicas**: requisitos tecnopolíticos, proposta metodológica e implicações práticas para a alta administração pública brasileira. Texto para discussão. Instituto de Pesquisa Econômica Aplicada (IPEA). Brasília: Rio de Janeiro, 2015.

CHEN, Hsinchun; CHIANG, Roger HL; STOREY, Veda C. **Business intelligence and analytics**: from Big Data to Big Impact. **MIS quarterly**, v. 36, n. 4, p. 1165-1188, 2012.

CISCO. The Zettabyte Era: **Trends and Analysis**. Estados Unidos. 2015. Disponível em: [http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/VNI\\_Hyperconnectivity\\_WP.pdf](http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/VNI_Hyperconnectivity_WP.pdf). Acesso em: 4 abr. 2016.

COHEN, Max F. Alguns aspectos do uso da informação na economia da informação. Brasília: **Ciência da Informação**, v.31, nº 03, p.26-36, 2002.

CRAWFORD, Kate; MILTNER, Kate; GRAY, Mary L. Critiquing Big Data: Politics, Ethics, Epistemology. **International Journal of Communication**, nº 8, 2014.

CRESWELL, J.W. **Research design**: qualitative, quantitative, and mixed methods approaches. 2. ed. Thousand Oaks: Sage Publications, 2003. 245 p.

DAVENPORT, Thomas. H. **Big Data no trabalho**. Elsevier Brasil, 2014.

DEL AMO, O.P; HURTADO, Eva. **Measuring the impact of short term rentals in Spain: decline of a myth**. 14th Global Forum on Tourism Statistics. Itália, Veneza, 2016.

DEVIKA, K.; SURENDRAN, Subu. An overview of web data extraction techniques. **International Journal of Scientific Engineering and Technology**, v. 2, n. 4, 2013.

DUPEYRAS, A.; MACCALLUM, N. Indicators for measuring competitiveness in tourism: a guidance document. **OECD Tourism Papers**, 2013/02. OECD, Paris, 2013.

FARREL, Adrian. **A Internet e seus Protocolos**. Elsevier Brasil, 2005.

GIL, A.C. **Como elaborar projetos de pesquisa**. São Paulo: Atlas, 2002. 231 p.

GOLDMAN, Alfredo et al. Apache Hadoop: Conceitos teóricos e práticos, evolução e novas possibilidades. **XXXI Jornadas de atualizações em informática**, 2012.

GÜNTHER, Hartmut. Pesquisa qualitativa versus pesquisa quantitativa: Esta é a questão? **Revista Psicologia: Teoria e Pesquisa**, v. 22, n. 2, p. 201-210, maio/ago. 2006. Disponível em: <http://www.scielo.br/pdf/ptp/v22n2/a10v22n2>. Acesso em: 07 mai. 2016.

JANNUZZI, Paulo de Martino. Indicadores para diagnóstico, monitoramento e avaliação de programas sociais no Brasil. **Revista do Serviço Público**, v. 56, n. 2, p. 137-160, 2015.

KITCHIN, Rob. Big Data, new epistemologies and paradigm shifts. **Big Data & Society**, abr. jun., p. 1-12, 2014.

LO MONACO, Jean-Christophe. **Long-distance ridesharing has had an impact on domestic tourism travel in France since 2010**. 14th Global Forum on Tourism Statistics. Itália, Veneza, 2016.

LIMBERGER, Pablo Flôres et al. Satisfaction in hospitality on TripAdvisor.com: An analysis of the correlation between evaluation criteria and overall satisfaction. **Tourism & Management Studies**, v. 10, n. 1, p. 59-65, 2014.

MACULAN, B. C. M. S. **Manual de normalização**: padronização de documentos acadêmicos do NITEG/UFMG e do PPGCI/UFMG. 2. ed. atual. e rev. Belo Horizonte: UFMG, 2011.

MARRES, Noortje; WELTEVREDE, Esther. Scraping the social? Issues in real-time social research. **Journal of Culture Economy** (subm), p. 1-52. Goldsmiths Research online, 2012. Disponível em <http://eprints.gold.ac.uk/6768/>. Acesso em: 3 mai. 2016.

MINAS GERAIS. Escritório de Prioridades Estratégicas. **Caderno de Indicadores**. Escritório de Prioridades Estratégicas, Belo Horizonte, 2011.

MINAS GERAIS. **Plano Mineiro de Desenvolvimento Integrado**: Desenvolvimento econômico e social sustentável de Minas Gerais – 2016 a 2017. V. I, Belo Horizonte, 2016.

MINAS GERAIS. **Plano Plurianual de Ação Governamental**: 2016 – 2019. V. III, Belo Horizonte, 2016b.

MORAIS, Raul S.; OLIVEIRA, Rafael A.; PAIM, Alessandra H. C. A. As Pesquisas de Demanda Turística em Minas Gerais: evolução histórica, principais desafios e perspectivas de futuro. **Revista Turismo e Análise**. V. 26, n.1, p. 21-37. Universidade de São Paulo (USP), 2015.

MUNAR, Ana Maria; JACOBSEN, Jens Kr. (2013) Trust and Involvement in Tourism Social Media and Web-Based Travel Information Sources. **Scandinavian Journal of Hospitality and Tourism**, p. 1-19, 2013.

MUNZERT, Simon et al. **Automated data collection with R: A practical guide to web scraping and text mining**. John Wiley & Sons, 2015.

NOVO, Rafael; NEVES, J. M. Souza. Inovação na inteligência analítica por meio do Big Data: características de diferenciação da abordagem tradicional. **VIII Workshop de pós-graduação e pesquisa do Centro Paula Souza**: Sistemas produtivos: da inovação à sustentabilidade. São Paulo, 9 e 10 de outubro, 2013.

ORACLE. **An Enterprise Architect's Guide to Big Data** - Reference Architecture Overview. – Oracle Enterprise Architecture White Paper. Estados Unidos, 2015.

Disponível em: <http://www.oracle.com/technetwork/topics/entarch/articles/oea-big-data-guide-1522052.pdf> Acesso em: 19 abr. 2016.

ORGANIZAÇÃO MUNDIAL DO TURISMO. **Compilación de las estadísticas del gasto turístico**. (Manual Técnico No. 2), OMT, 1995.

PATERNOSTER, Nicolò et al. Software development in startup companies: A systematic mapping study. **Information and Software Technology**, v. 56, n. 10, p. 1200-1218, 2014.

PEREIRA, D. Ramos de Miranda; PINTO, Marcelo Rezende. A importância do entendimento dos indicadores na tomada de decisão de gestores públicos. **Revista do Serviço Público Brasília** v. 63, n.3. p.363-380 jul/set 2012.

PIMENTA, Ricardo M. Big Data e controle da informação na era digital: tecnogênese de uma memória a serviço do mercado e do Estado. **Revista Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 6, n. 2, jul./dez. 2013.

PROVOST, Foster; FAWCETT, Tom. Data science and its relationship to big data and data-driven decision making. **Big Data**, v. 1, n. 1, p. 51-59, 2013.

PUSCHMANN, Cornelius; BURGESS, Jean. Metaphors of Big Data. **International Journal of Communication**, nº 8, 2014.

RAMOS, Célia. **Sistemas de informação para a gestão turística**. Revista Encontros Científicos. Faro, n. 6, p.107-116, 2010.

RIBEIRO, Cláudio José Silva. Big Data: os novos desafios para o profissional da informação. **Revista Informação & Tecnologia (ITEC): João Pessoa/Marília**, 1(1): 96-105, jan./jun., 2014.

SANDVIG, Christian et al. Auditing algorithms: Research methods for detecting discrimination on internet platforms. **Data and discrimination: converting critical concerns into productive inquiry**, 2014.

SARACEVIC, T. Information science. In: Marcia J. Bates and Mary Niles Maack (Eds.) **Encyclopedia of Library and Information Science**. New York: Taylor & Francis. pp. 2570-2586, 2009.

SCHMUCKER, Dirk; SONNTAG, Ulf; WAGNER, Philipp. **Assessing the impact of “shared accommodation” for city tourism**. 14th Global Forum on Tourism Statistics. Itália, Veneza, 2016.

SERAPIONI, Mauro. Conceitos e métodos para a avaliação de programas sociais e políticas públicas. **Sociologia, Revista da Faculdade de Letras da Universidade do Porto**, Vol. XXXI, 2016, p. 59-80.

SONNTAG, Ulf. **Assessing the impact of “shared accommodation” for city tourism**. 14th Global Forum on Tourism Statistics. Itália, Veneza, 2016. (Comunicação oral).

SOUZA, Renato Rocha; ALMEIDA, Maurício Barcellos; BARACHO, Renata Maria Abrantes. Ciência da informação em transformação: Big Data, nuvens, redes sociais e Web Semântica. **Ciência da Informação**, v. 42, n. 2, 2015.

STANTON, Jeffrey M. **Introduction to data science**. 2013.

TARAPANOFF, Kira. Informação, conhecimento e inteligência em corporações: relações e complementaridade. In: TARAPANOFF, Kira (Org.) **Inteligência, informação e conhecimento**. Brasília: UNESCO/IBICT, pag. 19-36, 2006.

TEIXEIRA, Enise Barth. A análise de dados na pesquisa científica: importância e desafios em estudos organizacionais. **Desenvolvimento em questão**, v. 1, n. 2, p. 177-201, 2003.

VARGIU, Eloisa; URRU, Mirko. Exploiting web scraping in a collaborative filtering-based approach to web advertising. **Artificial Intelligence Research**, v. 2, n. 1, p. p44, 2012.

VIEGAS, Maria Margarida Viçoso de Arrais. **As estatísticas do turismo e a uniformização de conceitos**. Lisboa: Instituto Nacional de Estatística, 1997.

WEISS, Juergen. **The impact of shared accommodation for the overall accommodation industry**. 14th Global Forum on Tourism Statistics. Itália, Veneza, 2016.

WORLD TRAVEL AND TOURISM COUNCIL. Travel & Tourism: **Economic impact 2016 – World**. Londres, 2016.

XIANG, Zheng; MAGNINI, Vincent P.; FESENMAIER, Daniel R. Information technology and consumer behavior in travel and tourism: Insights from travel planning using the internet. **Journal of Retailing and Consumer Services**, v. 22, p. 244-249, 2015.

ZERVAS, Georgios; PROSERPIO, Davide; BYERS, John W. The rise of the sharing economy: Estimating the impact of Airbnb on the hotel industry. **Journal of Marketing Research**, 2014.

ZDEPSKI, Fabíola Bevervanço. **Proposta de metodologia para definição de requisitos de um sistema de informação de gestão estratégica do turismo municipal**: O caso do município de Ponta Grossa - Paraná. Dissertação de mestrado. Curitiba: Universidade Católica do Paraná, 2004.