



DADOS DO TRABALHO

CÓDIGO DE IDENTIFICAÇÃO: 135

TÍTULO

Utilização de dados extraídos do site TripAdvisor para elaboração de indicadores do turismo em Minas Gerais

ÓRGÃO/ENTIDADE EXECUTOR(A)

Secretaria de Estado de Turismo

CATEGORIA

Ideias Inovadoras Implementáveis

MODALIDADE

Inovação em Processos Organizacionais

ÁREA TEMÁTICA

AVALIAÇÃO E MONITORAMENTO DE POLÍTICAS PÚBLICAS

PÚBLICO ALVO

O projeto tem como público alvo os visitantes de atrativos e espaços turísticos de Minas Gerais; gestores do poder público estadual e municipal; gestores de espaços públicos e privados voltados à visitação turística.

RESUMO



A iniciativa tem como objetivo avaliar a possibilidade de construir indicadores para monitorar o desempenho do turismo em espaços turísticos de Minas Gerais a partir de dados extraídos da maior plataforma web de avaliação de destinos turísticos mundial denominada TripAdvisor. Para tanto, foi utilizada uma ferramenta de extração de dados gratuita para coletar dados do site TripAdvisor, buscando as principais informações dos atrativos turísticos de Minas Gerais e transformando-as em um banco de dados estruturado. Os resultados preliminares possibilitaram 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. Assim, acredita-se que essas informações possam ser trabalhadas para se tornarem indicadores de desempenho da atividade turística no estado, considerando que as informações são preenchidas pelos próprios visitantes.

PALAVRAS-CHAVE

monitoramento; tripadvisor; turismo; indicadores

PROBLEMA ENFRENTADO OU OPORTUNIDADE PERCEBIDA

Segundo dados da Organização Mundial do Turismo (OMT), 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 com uma 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, o monitoramento de informações realizado 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 também que para a realização de pesquisas e levantamento de informações no setor público, são necessários recursos que na grande maioria das vezes, encontram-se limitados, dificultando a continuidade para a elaboração de pesquisas de grande porte, tal como a pesquisa de demanda turística. Essa pesquisa é considerada como um dos estudos realizados pela SETUR-MG 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. 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. Assim, buscar alternativas de coleta de dados que possibilitem auxiliar com menor custo a monitorar o impacto da atividade e conseqüentemente, auxiliar na elaboração de políticas, se torna fundamental.



Sabe-se também que, atualmente, o turista durante a sua viagem, compartilha sua opinião em tempo real sobre os produtos e serviços consumidos nas redes sociais, se tornando uma grande fonte de informação para outros usuários das redes que planejam viajar para os mesmos destinos e também para os gestores públicos que acabam ganhando uma fonte de informação de dados atualizada sobre a qualidade e a satisfação desses turistas nos destinos mineiros, mas que muitas vezes não se encontra disponível em um banco de dados estruturado.

JUSTIFICATIVA

O grande volume de dados gerados por diferentes meios de comunicação em alta velocidade e de diferentes formatos caracteriza-se o fenômeno conhecido atualmente como "Big Data", gerando novos questionamentos sobre a utilização de informações de forma estratégica pelas organizações. Esse contexto faz com que, conforme Bastos (2006), a otimização de esforços para o gerenciamento de processos de busca e uso de informação nas organizações seja caracterizado como um tema atual de grande interesse para os gestores e tomadores de decisões. Complementando esse pensamento, Beuren (1998, p. 43) afirma que "o desafio maior da informação é o de habilitar os gestores a alcançar os objetivos propostos para a organização, por meio do uso eficiente dos recursos disponíveis".

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. Considerado o maior site de compartilhamento de informações de viagens do mundo, é alimentado 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), 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.

Assim, acredita-se que o monitoramento e a avaliação do comportamento dos usuários nessa rede possam auxiliar na elaboração de indicadores, com possibilidade de contribuir para a criação de estratégias que visem à melhoria do atendimento e da satisfação dos visitantes em atrativos e destinos turísticos de Minas Gerais. 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).

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 (...)?.

Para Marres e Weltevrede (2012, p.10, tradução nossa) "(...) 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".

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.

Um dos aplicativos gratuitos que possibilitam a extração de dados é o "Import.io". De acordo com informações retiradas do site da empresa (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 da Big Data. Além disso, não é necessário conhecimentos de programação para o uso da ferramenta, possibilitando sua utilização por qualquer servidor do Estado.

Portanto, acredita-se que o uso da ferramenta para a extração de informações dos usuários do TripAdvisor, possibilitará analisar de forma quantitativa e qualitativa a opinião consolidada de todos os atrativos turísticos do estado existentes na plataforma, o que não pode ser feito por métodos de pesquisas survey, in loco, devido a dificuldade de se alocar um pesquisador para cada atrativo do estado durante todos os dias do ano para coletar opiniões dos visitantes. Assim, a alternativa de buscar as informações que já estão sendo alimentadas pelo próprio cidadão ou visitante em uma plataforma online de reconhecimento mundial pode servir como complemento de pesquisas já realizadas pela SETUR-MG, diminuindo seu custo ou até possibilitando a elaboração de indicadores de desempenho da atividade.

REFERÊNCIAS:

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.

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.

TORRES, J.; MORALEZ, A.; JIMENEZ, S. Quantitative analysis os hotels in TripAdvisor: spanish and portuguese destinations. Málaga, Espanha, 2013.

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.

OBJETIVO



- Ampliar a nota do índice de monitoramento de destinos turísticos mineiros do MTUR;
- Auxiliar na elaboração de indicadores que possam ser utilizados no desenvolvimento de políticas de turismo em Minas Gerais, a partir da extração de dados em redes sociais;
- Aprimorar as técnicas de extração de dados a partir da utilização de uma ferramenta gratuita e acessível que não necessite de conhecimentos de programação;
- Contribuir para melhoria da tomada de decisão dos gestores públicos e dos atrativos pesquisados a partir da utilização de uma ferramenta de extração de dados;
- Servir como modelo de metodologia de extração de dados para demais Secretarias.

RESULTADOS ESPERADOS PARA O GOVERNO E/OU SOCIEDADE

Espera-se que essa iniciativa possa contribuir na criação de indicadores de desempenho do turismo no estado a partir da opinião dos cidadãos e turistas no estado de Minas Gerais. Esse monitoramento poderá facilitar a realização de políticas de visem melhorias para o aumento do nível de satisfação dos visitantes, além de aumentar as avaliações positivas dos atrativos do estado na plataforma TripAdvisor, e consequentemente, gerando maior número de pessoas interessadas em visitar Minas Gerais a partir dos comentários positivos, movimentando a economia do turismo.

DIFICULDADES QUE PODEM SER ENFRENTADAS DURANTE E/OU APÓS SUA IMPLANTAÇÃO

- Por ser utilizada uma ferramenta gratuita para a extração de dados, pode ocorrer alguma mudança na forma de utilização da mesma que impossibilite o acompanhamento ou atualização dos dados, sendo necessária a busca por uma outra solução que atenda os mesmos requisitos.
- Falta de entendimento da gestão pública sobre a importância de monitoramento de redes sociais para tomada de decisão e validação das informações.

ESTUDOS PRELIMINARES (SE HOVER)

Os estudos preliminares apontaram que foi possível coletar todas as 146.371 avaliações de cada um dos 1.384 atrativos de Minas Gerais cadastrados no site TripAdvisor e trabalhá-las de forma eficiente a partir de uma planilha do excel denominada "tabela dinâmica". Deve-se levar em consideração que grande parte dos atrativos possuem poucos comentários o que pode prejudicar a análise crítica desses atrativos, e para fins de análise dos resultados, optou-se por considerar apenas os atrativos que possuam no mínimo, 106 avaliações de usuários, sendo esse valor considerado como a média de avaliações de cada atrativo no estado. Assim, as análises foram feitas a partir dos 235 atrativos com maior número de avaliações.

Foi possível, dentre várias opções, identificar:

- a oferta de atrativos por município e por categoria (museus, igreja, shopping...);
- os atrativos, municípios e categorias com maior número de avaliações absolutas, além de avaliações por segmento (excelente, muito bom...);



- o perfil dos visitantes para cada município, atrativo e categoria (família, sozinho, amigos, romance, a negócios...)
- a época de visitação (por trimestre) mais pertinente de cada município, atrativo e categoria.

Como exemplo, foi possível identificar quais são os principais atrativos turísticos com avaliações no município de Mariana (tabela 01 - anexo), divididos em categorias. Pelos dados apresentados, percebe-se que a categoria "minas" representa 34% do total de avaliações dos atrativos apresentados. Já a categoria "igrejas e catedrais" representa 32% 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.

É possível observar também que de todos os atrativos do município, a maior nota de satisfação foi alcançada pela Praça Minas Gerais (4,5) e a menos pelo museu histórico Casa de Câmara e Candeia (3,9).

Em relação ao perfil dos visitantes, observa-se que Mariana possui uma distribuição próxima nos perfis de pessoas que viajam a casal (35%), com a família (32%) e com amigos (23%), o que acaba sendo observado também em cada um dos atrativos, possuindo perfis de visitação semelhante entre eles.

A tabela 2 (anexo) mostra todos os museus de Minas Gerais caracterizados como "Museus Especializados" que possuem o maior número de avaliações no site TripAdvisor, sendo que o Inhotim representa 35% de todas as avaliações realizadas.

Em relação a satisfação dos usuários, percebe-se que o Museu Dona Beja em Araxá 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. Ao identificar problemas de satisfação no museu Dona Beja, foi realizada uma análise qualitativa dos comentários no site e percebeu-se que o museu ficou mais de um ano fechado para reforma, o que gerou insatisfação dos usuários. No caso do Museu do Diamante em Diamantina, as maiores reclamações foram focadas na falta de infraestrutura e o baixo acervo. Assim, é possível criar políticas que auxiliem na melhoria das condições dos museus de forma pontual, visando a melhoria da satisfação dos usuários ao longo do tempo. Caso seja realizada uma reforma ou melhoria nas condições de visitação, o indicador poderá refletir numa melhoria das notas, possibilitando analisar de forma quantitativa o impacto de uma determinada ação de melhoria. A tabela apresentada possibilita também visualizar os períodos com o maior número de comentários dos viajantes, que se acredita ser em grande parte condizentes com as datas de visitação desses museus. Assim, é possível checar a possibilidade de criar roteiros integrados de visitação em museus que possuem uma taxa de avaliações baixas, como por exemplo em Dezembro a Fevereiro, criando iniciativas como programações gratuitas, maior divulgação dos espaços, dentre outras ações, como é o caso do Museu de Mineralogia (20%), Memorial Minas Gerais Vale (21%), Palácio da Liberdade (23%) e Museu de Artes e Ofícios (21%), todos eles localizados em Belo Horizonte.

Já a tabela 3 (anexo) mostra que os destinos com maior proporção de viagens com amigos são característicos de natureza (cachoeiras, lagos, serras...) tais como Andradas, São Thomé das Letras, Conceição do Ibitipoca e Serra do Cipó. Já os destinos característicos de casais (romântico) foram destacados em locais de pousadas, ligados a gastronomia e cultura, condizente com o que é divulgado para esse público como nos casos de Gonçalves, Lavras Novas e Monte Verde, cidades já conhecidas por serem atrativas para visitantes em lua de mel. Essa informações possibilitam elaborar propostas de roteiros ou divulgações desses destinos para um público específico, garantindo uma maior eficiência nas promoções realizadas pelo poder público.

A partir dessas informações, espera-se que os dados possam auxiliar posteriormente na elaboração de indicadores de desempenho (tais como taxa de variação do número de avaliações para os atrativos ou taxa de variação de visitação em baixa temporada) monitorando mensalmente as informações e validando-as como indicadores de acompanhamento da SETUR.



Também foram realizados testes em outros sites como Booking.com (site de reserva de hospedagem) e AirBNB (site de reserva de hospedagem alternativa), possibilitando buscar informações como a oferta de hospedagem em determinado município ou o preço médio da diária praticada, mas que deve ser aprofundado posteriormente.

GRAU DE NOVIDADE

Novo para a sociedade

CUSTO DE IMPLEMENTAÇÃO/MANUTENÇÃO

A ideia não possui custos de implementação ou manutenção, visto que utiliza ferramentas gratuitas para a extração das informações. Caso haja a possibilidade de utilização dessa metodologia para substituir indicadores coletados pela Pesquisa de Demanda Turística realizada por survey em apenas 03 momentos do ano (alta, média e baixa temporada), estima-se uma economia de, no mínimo, R\$ 90 mil, considerando ainda que o levantamento online pode ser realizado mensalmente, o que pode gerar uma economia ainda maior caso essa metodologia fosse feita por survey mensalmente.

PRAZO DE EXECUÇÃO (EM MESES)

3

DESCREVA AS ETAPAS DE IMPLANTAÇÃO

DESCRIÇÃO	INICIO	TÉRMINO
1º Teste da captura de informações	01/06/2016	08/06/2016
2º Teste da captura de informações	01/08/2016	05/08/2016
3º Teste da captura de informações	03/10/2016	07/10/2016
Análise das informações dos 3 testes	10/10/2016	21/10/2016
Elaboração de proposta de indicadores	24/10/2016	04/11/2016
Monitoramento dos indicadores de desempenho	31/01/2017	01/01/2018

ENVOLVE MAIS DE UM ÓRGÃO/ENTIDADE NA IMPLEMENTAÇÃO E EXECUÇÃO?

Não

ALGUMA OUTRA INSTITUIÇÃO JÁ EXECUTA ESTA INICIATIVA?



Não

Data de entrega no Sistema

13/07/2016 14:52:51