

**O uso da *Data Science* na análise dos indicadores do Turismo:
uma revisão das pesquisas no Brasil (2012-2022)**

Gabriela Costa Gonçalves

Bacharel em Turismo pela Universidade de Brasília, Brasil.
E-mail: gabrielacg.tur@gmail.com

Fagno Tavares de Oliveira

Doutor em Ciências Florestais - Conservação da Natureza - pela Universidade de Brasília, Brasil.
Professor Adjunto do Centro de Excelência em Turismo da Universidade de Brasília, Brasil.
E-mail: fagno@unb.br

Resumo

Este artigo apresenta a *Data Science* como discute a interseção entre turismo e ciência de dados, particularmente o uso da *Data Science* para analisar os dados gerados pelo turismo. O aumento exponencial do volume de dados gerados pelo turismo requer novas abordagens analíticas, e a análise estatística orientada por tecnologias computacionais dos dados torna-se essencial para compreender tendências, preferências e comportamentos dos consumidores e oferecer serviços mais personalizados e eficientes. O estudo verifica a aplicabilidade da *Data Science* na análise dos indicadores do turismo, com base na análise, identificação e avaliação dos indicadores comumente utilizados no desenvolvimento do turismo. Para isto, foi realizado um levantamento na base de dados *Scielo* e *Redalyc* nos últimos 10 anos (2012-2022) com auxílio das técnicas *skimming* e *scanning* foi realizada a análise de conteúdo. Em síntese, foram identificadas 180 produções científicas, porém é insipiente estudos que relacionam *Data Science* e Turismo no Brasil. O uso da *Data Science* no turismo possui enorme potencial para contribuir na construção de políticas e diretrizes mais eficazes e sustentáveis no setor.

Palavras-chave: Turismo. Ciência de dados. Indicadores do turismo. Análise de dados.

1 INTRODUÇÃO

A era dos dados abrange uma variedade de setores da sociedade, incluindo o turismo. A crescente popularidade das agências de viagens on-line, redes sociais e sites de reserva tem levado a uma redução de serviços turísticos físicos. As avaliações dos usuários nessas plataformas produzem uma grande quantidade de informações sobre o comportamento dos viajantes. Todavia, grande parte das informações geradas são no formato digital e coletada por meio de computadores, provenientes de vários usuários, é possível organizá-los e extrair *insights* relevantes apenas com o auxílio de recursos computacionais (PUSCHMANN; BURGESS, 2014).

O uso da *Data Science* na análise de dados tem se mostrado cada vez mais relevantes para o turismo, especialmente nas esferas econômica, social e ambiental que compõem esta atividade. O tema *Data Science* surge como uma abordagem inesperada que pode ser explorada pelo turismo com a finalidade de desenvolver melhorias para o setor.

Para Fuchs, Höpken e Lexhagen (2014), a aplicação de um modelo de análise de dados multidimensional pode elevar a competitividade e potencial de um destino turístico moderno, contribuindo para a criação, implementação e exploração do conhecimento. Li e Law. (2020), destacam a importância da *Data Science* para avaliar o crescimento do turismo e mostram como ele melhora a precisão e a eficiência na análise de dados turísticos. Davenport e Harris (2017) apontam que, ao usar técnicas de análise de dados, as empresas podem obter *insights* valiosos sobre seus clientes, produtos e processos internos, o que pode levar a melhorias significativas de eficiência e rentabilidade.

Ao considerar o turismo como um fenômeno que pode ser estudado sob diferentes perspectivas, surgem oportunidades analíticas desafiadoras que merecem a atenção. A complexidade e a diversidade do turismo exigem abordagens multidisciplinares e metodologias inovadoras para entender e explicar os diferentes aspectos e dimensões envolvidos. Nesse sentido, a análise de dados se torna fundamental para investigar as dinâmicas do turismo e, assim contribuir para a elaboração de políticas e diretrizes mais eficazes, na definição e avaliação de indicadores no setor.

Os indicadores do turismo podem ser divididos em diversas categorias como, social, cultural, ambiental, econômica, infraestrutura e gestão. Cada categoria abrange indicadores específicos que auxiliam na análise do desempenho e impacto do turismo em determinada região ou localidade. Para que “*esses indicadores possam ser utilizados por um público amplo e não especializado, é necessário encontrar um equilíbrio entre facilidade de comunicação e personalização às necessidades dos usuários, de um lado, e precisão estatística, solidez analítica e coerência científica, de outro*” (UNWTO; Programa de Conta Satélite do Turismo, 2015, p. 07).

A utilização da *Data Science* na construção dos indicadores do turismo se apresenta como uma alternativa vantajosa. Tendo em vista que o uso da mesma permite a adoção de métodos mais rigorosos e abrangentes utilizando equipamentos tecnológicos de ponta.

O turismo está intimamente ligado ao uso das novas tecnologias de comunicação e informação, que tem impulsionado seu crescimento e transformação. Isso tem levado a mudanças no comportamento dos turistas e conseqüentemente, no fenômeno turismo, tornando essencial a busca por ferramentas que permitam compreender as novas tendências.

Este estudo teve como questão central “como a *Data Science* pode auxiliar na análise dos indicadores do turismo?”. Para responder a esta questão norteamos a pesquisa

objetivando verificar a aplicabilidade da *Data Science* na análise dos indicadores do Turismo, tendo como suporte a análise, a identificação do uso da *Data Science* na área do turismo, o levantamento dos indicadores comumente utilizados no turismo e por último, a avaliação do uso da *Data Science* na análise desses indicadores.

2 METODOLOGIA

2.1 Métodos de Pesquisa

A metodologia adotada nesta pesquisa foi a análise bibliométrica e de conteúdo. A primeira se refere a contagem de publicações ou citações encontradas nas bases de publicações científicas e acadêmicas (COATES *et al.*, 2001). Desta forma buscou-se identificar, analisar e sintetizar as principais informações sobre o tema proposto *Data Science*, Indicadores e Turismo.

A análise bibliométrica, de acordo com Porter (2007), enfatiza além das contagens, as conexões entre os temas rastreados, ou entre autores ou entre instituições, pode dar indicações sobre a evolução dos desenvolvimentos e inovações em estágios iniciais. Sendo o caso desta pesquisa sobre uso da *data science* no turismo.

A segunda se constitui de várias técnicas onde se busca descrever o conteúdo emitido no processo de comunicação, seja ele por meio de falas ou textos (BARDIN, 2007), no caso desta pesquisa, de textos sobre *data science*, indicadores e turismo. A escolha deste método de análise permite, de forma sistemática, a descrição das mensagens, e das atitudes atreladas ao contexto na enunciação, bem como as inferências sobre os dados coletados (CAVALCANTE; CALISTO; PINHEIRO, 2014).

Nesta pesquisa foi realizado o levantamento nas bases de dados *Scielo* e *Redalyc*. A escolha das bases de dados justifica-se pelo número elevado de periódicos indexados do Brasil, possibilitando um levantamento maior de artigos sobre o tema abordado. O levantamento foi realizado longitudinalmente (2012 a 2022), visando avaliar a quantidade de publicações neste período no Brasil. Foram utilizados descritores-chave na consulta, como, “*data science*”, “ciência de dados”, “análise de dados”, “indicadores” e “turismo”, uma vez que a temática “uso da *Data Science* no turismo” é recente no país, obrigando a divisão dos termos para ampliar as possibilidades de resultados. Na busca foram encontrados 180 artigos publicados nas bases de dados que foram avaliados por meio das técnicas de leitura *skimming* e *scanning*.

As técnicas *skimming* e *scanning* foram utilizadas para auxiliar na análise de conteúdo e estabelecer critérios de leitura na investigação, objetivando remover trabalhos que não possuem relação com a temática. A técnica *skimming* consiste na leitura rápida para identificar se o texto é relevante ou não para as suas necessidades (HARMER, 2007). Por sua vez, a técnica *scanning* compreende uma leitura mais específica e objetiva, onde o leitor procura informações específicas em um texto. Nesta técnica, o leitor se concentra nas palavras-chave e frases relevantes para a sua busca, sem se preocupar com detalhes secundários (BROWN; LALAND, 2001).

Para a realização deste estudo adotamos as fases propostas por Bardin (2011): Pré-análise; Exploração do material; e Tratamentos dos resultados. A pré-análise consiste na organização do material útil à pesquisa, como: a coleta de dados mediante a pesquisa de artigos na base de dados *Scielo* e *Redalyc*. Na exploração do material: a identificação das variáveis de interesse para o estudo e a organização dos indicadores relevantes; Por fim, o tratamento dos resultados que consiste na análise reflexiva e crítica do tema em estudo.

3 REFERENCIAL TEÓRICO

3.1 Ciência de Dados ou *Data Science*: um breve contexto

A ciência de dados ou *data science* é um campo interdisciplinar que fica na fronteira da estatística e da ciência da computação e que utiliza o método científico e é responsável pela extração de informações valiosas a partir de grandes conjuntos de dados que são complexos, heterogêneos e distribuídos. (RAUDEMBERG; CARMO, 2019).

Esta disciplina envolve a aplicação de técnicas de estatística, aprendizado de máquina e análise exploratória, por conseguinte, está baseada em diversas matérias tais como, a matemática, estatística, programação, computação, negócios, mineração e visualização de dados. O interesse desta área de estudos se baseia em descobrir padrões, relações e tendências nos dados utilizando, linguagens de programação, modelos estatísticos e métodos.

De acordo com Hayashi (1998), em seu texto “*What is Data Science? Fundamental Concepts and a Heuristic Example*” a Ciência de Dados tem como objetivo analisar e compreender fenômenos reais por meio da utilização de dados. Além de buscar revelar as características e estruturas ocultas de fenômenos naturais, humanos e sociais complexos a partir de uma perspectiva diferente daquela estabelecida pelos métodos tradicionais. Esse novo ponto de vista implica em abordagens de pensamento multidimensionais, dinâmicas e flexíveis.

Adicionalmente, a *Data Science* utiliza quatro tipos de análises distintas em sua abordagem: 1) descritiva, que descreve eventos ou situações passadas; 2) análise diagnóstica, que identifica as causas e efeitos dos eventos baseado em dados; 3) preditiva, esta análise usa algoritmos e aprendizado de máquina para prever eventos futuros; e 4) análise preditiva fornece recomendações sobre as ações que devem ser tomadas para atingir os resultados, com base na análise de dados e modelos preditivos. As análises mencionadas têm importância em diversos contextos e são aplicáveis para solucionar uma ampla variedade de problemas em diferentes setores.

A fim de compreender e analisar os temas abordados neste estudo, é essencial distinguir os conceitos de *Data Science* e *Big Data*, que embora relacionados, são distintos no campo da tecnologia e da análise de dados. Enquanto, a *Data Science* visa extrair *insights* significativos do estudo dos dados, por sua vez, o *Big Data* é um termo que se refere ao grande volume de dados em si, e às tecnologias utilizadas para lidar com eles.

Para Arrigoni (2013), o conceito de *Big Data* surgiu em meados de 1990 pela Nasa, como uma forma de se referir a conjuntos de dados complexos e volumosos que desafiam os limites computacionais tradicionais em termos de captura, processamento, análise e armazenamento.

De acordo com Oliveira (2016), em sua pesquisa “*Extração de dados do site TripAdvisor como suporte na elaboração de indicadores do turismo de Minas Gerais: Uma iniciativa Big Data*” afirma que, a organização de dados se torna um elemento essencial para análise de informações caracterizadas como *Big Data*, a fim de lidar com a variedade de informações provenientes de diferentes dinâmicas e formatos.

Biz, De Souza e Caumo (2020) argumentam que o *Big Data*, antes mesmo de ter um conceito formalizado, este processo foi reconhecido como uma inovação disruptiva, desafiando o status quo ao questionar os métodos tradicionais de produção, gestão,

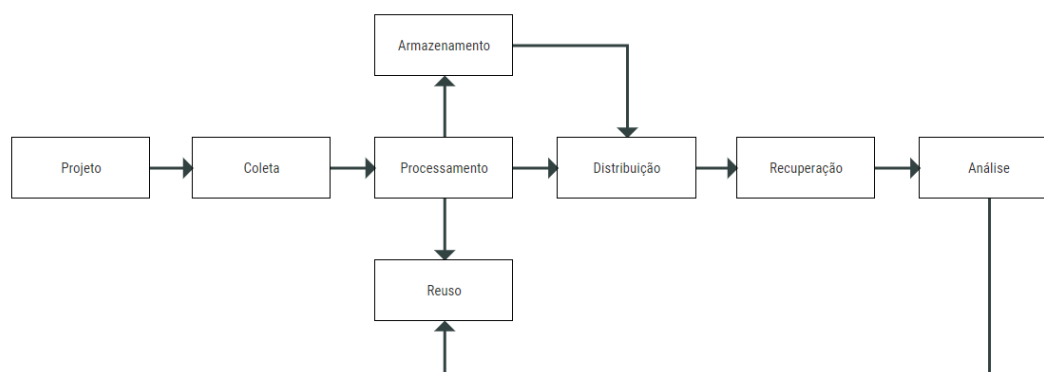
análise, armazenamento e utilização de dados. Esta transformação afetou profundamente a forma como o conhecimento é gerado, os negócios são conduzidos e a governança é estabelecida.

No que diz respeito, à análise e processamento de dados as etapas principais incluem captura de informações, transformação e posteriormente, apresentação dos resultados relevantes da amostra. Para utilizar a *Data Science* de forma efetiva, é essencial compreender o ciclo de vida de um projeto de *Data Science* e estabelecer uma estrutura que guie o desenvolvimento de soluções desde a concepção dos dados até a sua implementação.

Provost e Fawcett (2013) em seu livro, “*Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*”, mostram que o ciclo de vida de um projeto de *Data Science* é dividido em sete etapas tais como: a definição do problema, coleta e preparação de dados, análise exploratória, modelagem, avaliação e implementação. Complementarmente, o livro *Data Science from Scratch First Principles* de Joel Grus (2015), entende que as etapas básicas para um projeto de *Data Science* contêm a definição de problemas, coleta de dados, análise exploratória, modelagem, avaliação, implementação e monitoramento contínuo.

Um dos modelos mais notáveis no que se refere ao ciclo de vida de um projeto de *Data Science* (figura 1) é o modelo proposto pela Iniciativa de Documentação de Dados (*Data Documentation Initiative - DDI*), que segue um fluxo de funcionamento com entrada, processamento, saída e feedback de informações.

Figura 1. Exemplo de DDI



Fonte: Adaptado de DDI (2004).

3.2 Tecnologia da Informação e o Turismo

A tecnologia da informação e comunicação (TIC) pode ser entendida como um conjunto de instrumentos, técnicas e soluções que são realizadas com auxílio de recursos computacionais. Consiste em uma aplicação prática do conhecimento científico em diversas áreas de pesquisa e tem como objetivo e finalidade a obtenção de informações, armazenamento, proteção, processamento, acesso, gerenciamento e uso das informações e dados de uma pessoa, seja ela física ou jurídica. É importante ressaltar que a informação consiste em dados organizados de forma lógica e compreensível. Especialmente, porque a informação é a “matéria-prima de todas as áreas do conhecimento” (MIRANDA; SIMEÃO, 2003).

A incorporação da tecnologia da informação no cotidiano das pessoas “é um dos principais indicadores do desenvolvimento da sociedade da informação e a inserção das tecnologias de informação no cotidiano das pessoas e no desempenho e na transformação da sociedade como um todo” (GUIMARÃES, 2007). A ciência da informação, por sua vez, estuda as propriedades da informação, o uso e a transmissão e o processamento da informação para armazenamento e recuperação.

No turismo, a informação é essencial e pode ser considerada a matéria-prima do sistema turístico, por meio de sistemas de informação, o turismo reúne dados internos e externos que podem ser transformados em informação útil para o setor. Com isso, é possível armazenar, gerir e apresentar informações que possibilitem realizar análises a partir da tecnologia.

Para Hassan,

o setor do turismo está intrinsecamente ligado à tecnologia em seus diversos segmentos, sendo que seu crescimento depende da capacidade de inovação e do aproveitamento dessa tecnologia para aprimorar a gestão e desenvolver novos produtos, aperfeiçoar a comunicação, melhorar as experiências de viagem e personalizar o atendimento (HASSAN, 2011, p. 15).

Dessa forma, no decorrer dos anos o processo tecnológico e o turismo têm caminhado lado a lado. Outro ponto importante, de acordo com Buhalis e Law (2008), as TICs transformaram o setor de turismo, tanto em termos de estratégias de negócios quanto de estrutura industrial. A interação entre viagens e tecnologia transformou radicalmente o fenômeno turístico nos últimos anos. Tanto que, as tendências da web e uso de sites levaram a um crescimento exponencial na forma de fazer turismo, desde transações on-line até a descoberta de novos destinos. Além disso, o turismo e a Internet são os parceiros ideais (WTO, 2001) as aplicações disponíveis na web possibilitam o turista a ter acesso a informações relevantes sobre os destinos e os produtos turísticos que desejam consumir.

A utilização das TICs podem proporcionar novas oportunidades e desafios para a sustentabilidade do turismo (MARTINS; FIATES; PINTO, 2016). Baseado nisso, destaca-se o crescimento da interdisciplinaridade em diversas áreas do conhecimento, pois constitui-se, por si só, um desafio no desenvolvimento do turismo, tornando-o mais sustentável, inclusivo, acessível e coeso em que o desenvolvimento de tecnologias da informação e aplicações permite a diversificação, difusão e personificação de produtos turísticos. Por fim, ao serem incorporadas no setor de turismo auxiliam o turista, a experiência, e principalmente o planejamento e gestão dos destinos e locais turísticos.

3.3 O papel dos indicadores para análise do desenvolvimento do turismo

Os indicadores são utilizados para identificar, medir e ilustrar aspectos de um determinado conceito ou fenômeno e traduzir, de forma mensurável e simples. O termo “indicador” tem sua origem no latim “indicare” que significa “descobrir, apontar, anunciar, estimar”. Ele é usado para comunicar informações sobre o progresso em direção a uma meta específica e é usado como um recurso para tornar uma tendência ou fenômeno não imediatamente visível a partir de dados isolados (BELLEN, 2005).

De acordo com Ott (1978), em sua obra “*Environmental Indices: Theory and Practice*”, os indicadores são um meio encontrado para reduzir uma ampla quantidade de dados a uma forma mais simples de informação, retendo o significado essencial do que está sendo perguntado.

Netto (2020), explica que a importância dos indicadores é percebida quando se verifica a possibilidade de monitoramento da atividade turística, contribuindo no aumento da responsabilidade dos atores envolvidos, na ampliação da consciência ambiental e turística, bem como no suporte ao planejamento e às ações envolvidas.

Considerando que um indicador é um valor numérico que representa o status de algo relevante para a tomada de decisão de entidades governamentais e empresas, é fundamental compreender sua criação que envolve a definição de métodos de coleta e análise de dados, incluindo o uso de técnicas de ciência de dados, isto é, *Data Science*. Ao dissertar sobre o assunto Camargo (2000), destaca que os indicadores devem “transmitir credibilidade, e, para isso, devem ser produzidos de maneira criteriosa, garantindo a disponibilidade de dados e resultados relevantes no menor tempo e custo possível”.

No âmbito do Turismo a *United Nations World Tourism Organization* (UNWTO), considera que os indicadores são ferramentas essenciais para os processos de planejamento, gestão e monitoramento do turismo fornecendo informação acurada para os processos de tomada de decisão (UNWTO, 2007). Eles são especialmente necessários para a elaboração de estratégias de marketing, fortalecimento das relações entre as instituições, avaliação da eficiência e da efetividade das decisões gerenciais e a mensuração do turismo no âmbito da economia nacional (UN, 2010).

A definição de um sistema de indicadores e a criação de instrumentos e técnicas para medir esses indicadores se tornaram uma prioridade e um componente fundamental nos processos de planejamento e gestão de destinos turísticos. Isso requer uma aplicação prática de modelos propostos (HANAI, 2009).

Nessa perspectiva, a UNWTO (2007), publicou um guia que lista vários indicadores relacionados às áreas ambientais, sociais, econômicas e culturais. Esses indicadores são uma importante fonte de pesquisa e um banco de dados que pode ser usado em todas as fases do planejamento e desenvolvimento do turismo.

O quadro 1 a seguir contém os principais indicadores utilizados no turismo pela UNWTO (2018).

Quadro 1. Indicadores utilizados no turismo.

Tipos de Indicadores	O que são?
Receita Turística	é uma medida de receita gerada pelo turismo, como gastos com hospedagem, transporte, alimentação e compras.
Número de chegadas de turistas	é uma medida do número de pessoas que visitam um destino, região ou país com fins turísticos.
Tempo médio de estadia	é uma média de tempo que os turistas passam em um destino, região ou país.
Gasto por turista	é a média de gastos dos turistas por dia ou por viagem.
Taxa de Ocupação dos Hotéis	é uma medida de porcentagem de quartos de hotel que estão ocupados em uma determinada época do ano.
Indicador de Sustentabilidade do Turismo	É uma ferramenta utilizada para mensurar o grau de sustentabilidade de uma região.

Fonte: Adaptado de UNWTO (2018).

Todavia, ao considerar a capacidade dos indicadores de fornecer informações específicas para a mensuração de uma realidade estudada, pode-se imaginar bom uso destas ferramentas nas questões que envolvem o desenvolvimento das diversas atividades, dentre elas as preocupações advindas do turismo.

4. RESULTADOS E DISCUSSÃO

4.1 Uso e avaliação da *Data Science* na análise de indicadores do desenvolvimento do turismo

A partir da análise de 180 produções científicas relacionadas sobre o tema, levantadas nas bases de dados *Redalyc* e *SciELO* nos últimos 10 anos (2012-2022). Dentre as publicações, 06 foram consideradas de alta relevância por apresentarem pesquisas que relacionam a *Data Science* e o Turismo, porém apenas um estudo foi desenvolvido no Brasil. Para auxiliar na pesquisa e compreensão dos temas abordados, o escopo das palavras-chave foi expandido e uma nuvem de palavras foi elaborada (figura 2).

Figura 2. Nuvem de palavras



Elaboração própria (2023).

Os artigos que abordam o uso da *Data Science* estão apresentados a seguir (quadro 2), com título/autores, palavras-chaves e seu objeto de estudo.

Quadro 2. Estudos relacionados com a temática *Data Science*

Título do artigo / Autores	Palavras-chave	Objeto de estudo
Modelos preditivos de cancelamento de reservas de hotéis: uma análise semiautomática da literatura. ANTÓNIO, Nuno; ALMEIDA, Ana; NUNES, Luis. <i>Tourism & Management Studies</i> , 15(1), 7-21. 2019.	<i>Data Science</i> , Gestão de receitas, previsão, revisão de literatura, processamento de linguagem natural.	Este trabalho aproveita os métodos avançados de ciência de dados para sintetizar as descobertas atuais da pesquisa sobre o desenvolvimento de modelos de previsão/predição para cancelamento de reservas em setores relacionados a viagens e turismo e para identificar os principais tópicos cobertos pela pesquisa de cancelamento de reservas.
Avaliação do impacto das Tecnologias da Informação e Comunicação no setor hoteleiro português: uma análise exploratória com <i>Data Envelopment Analysis</i> . PAÇO, Cidália Maria Leal; PÉREZ, Juan Manuel Cepeda. <i>Tourism & Management Studies</i> , 11, 35-43. 2015	<i>Data envelopment analysis</i> , tecnologias da informação e comunicação, produtividade, hotéis.	Este estudo proporcionou valor adicional ao setor hoteleiro, turístico, acadêmico, profissional e outros setores em geral. Concentrou-se no desenvolvimento e implementação de uma metodologia apropriada, destacando o potencial das Tecnologias de Informação e Comunicação (TIC) para melhorar a eficiência. Além de abordar deficiências metodológicas em estudos anteriores, busca superar inconsistências e resultados contraditórios. O estudo também destaca a importância de considerar questões específicas relacionadas às TIC para aproveitar plenamente seus benefícios. Identificar aplicações específicas das TIC e seus benefícios é fundamental para que as empresas desenvolvam um conjunto adequado de ferramentas alinhadas com sua estratégia.
Desempenho das capitais brasileiras como destinos turísticos. SANTOS, Erick da Silva; FORTES, Mauri. <i>Revista Brasileira de Pesquisa em Turismo</i> , 9(1), 22-39. 2015	<i>Efficiency. Data Envelopment Analysis (DEA). Factor Analysis. Tourism Demand. Brazilian Tourism Destinations.</i>	O principal objetivo deste estudo foi fornecer uma análise de <i>benchmarking</i> para contribuir com a literatura de turismo, oferecendo novos <i>insights</i> sobre a avaliação de desempenho de destinos turísticos no Brasil, bem como aplicar o modelo DEA (Análise Envoltória de Dados) para atingir tal objetivo.
<i>Determinantes del tipo impositivo efectivo en el sector turístico: un modelo dinámico con datos de panel.</i> MORENO-ROJAS, José; GONZÁLEZ-RODRÍGUEZ, Rosário; MARTÍN-SAMPER, Rosário Carmem. <i>Tourism & Management Studies</i> , 13(3), 31-38. 2017.	<i>Effective tax rate, asset composition, firm size, financial structure, equity, dynamic panel data, ROA.</i>	O estudo usa dados de painel para analisar a importância de certas variáveis como tamanho da empresa, estrutura financeira, rentabilidade, ativos fixos e tipo de empresa (hotéis ou agências de viagens).

<p><i>Evaluación de Destinos Turísticos mediante la Tecnología de Ciencia de Datos.</i></p> <p>MOLINAR, Carlos Mario Amaya; ESPINOZA, Pedro Magan; LLAMAS, Ileana Ochoa. <i>Estudios y Perspectivas en Turismo</i>, 26(2), 286-305. 2017.</p>	<p>Destinos turísticos, <i>Big Data</i>, <i>competitividad</i>, imagen, experiencia de visita.</p>	<p>O artigo apresenta resultados de um estudo exploratório aplicando uma metodologia de <i>Big Data</i> em dois destinos turísticos mexicanos, utilizando o método de análise de sentimentos e opiniões em avaliações publicadas online sobre quatro tipos de serviços turísticos: atrações turísticas, hospedagem, alimentação e bebidas, e transporte.</p>
<p><i>Similarities and correlation between resident tourist overnights and Google Trends information in Portugal and its tourism regions.</i></p> <p>DINIS, Gorete; COSTA, Carlos; PACHECO, Osvaldo. <i>Tourism & Management Studies</i>, 13(3), 15-22. 2017</p>	<p><i>Google Trends</i>, turismo doméstico, dados de pesquisa, Portugal, <i>big data</i>.</p>	<p>Este trabalho de pesquisa tem como objetivo destacar a capacidade das ferramentas <i>Google Trends</i> e <i>Google Analytics</i> em fornecer indicadores quase em tempo real sobre os interesses e comportamentos coletivos online. Iniciou-se com uma revisão da literatura e análise de dados secundários, permitindo compreender as políticas, indicadores e estatísticas oficiais do turismo, a utilização da Internet e o potencial das ferramentas <i>Google Trends</i> e <i>Google Analytics</i> para os agentes do setor.</p>

Fonte: Elaboração própria, 2023.

Verifica-se a ausência de estudos na área do turismo sobre o uso da *Data Science*, o que indica uma lacuna na produção do conhecimento. Tal resultado pode estar relacionado a vários fatores, como a falta de interesse na área; recursos para realizar pesquisas; desenvolvimento tecnológico para abordar essas questões. Todavia, a ausência de estudos também pode ser uma oportunidade para o desenvolvimento de futuras pesquisas. A realização de estudos pode fornecer novos *insights* e conhecimentos valiosos para a comunidade científica, para a sociedade e em especial para o turismo.

Apesar do número insuficiente de artigos encontrados nas bases de dados, *Scielo* e *Redalyc* (Quadro 2) e principalmente de estudos no Brasil, vale destacar que estas pesquisas desempenham um papel significativo na análise do setor de turismo. Eles utilizam métodos estatísticos, bem como exploram o potencial da ciência de dados e tecnologia da informação para alcançar esse objetivo.

No estudo "Modelos preditivos de cancelamento de reservas de hotéis: uma análise semiautomática da literatura", os pesquisadores realizaram uma investigação minuciosa para identificar e examinar modelos preditivos de cancelamento de reservas em hotéis. O foco principal do estudo foi ressaltar a importância da previsão e predição de cancelamentos para a indústria do turismo, particularmente no setor de hospitalidade. Ao revisar a literatura existente, os autores identificaram uma lacuna na pesquisa atual relacionada ao desenvolvimento e teste de modelos de cancelamento, enfatizando a relevância de modelos de previsão e provisão para compreender os fatores que influenciam os cancelamentos.

Essa análise aprofundada contribui para o campo do turismo, fornecendo perspectivas valiosas sobre estratégias de previsão e predição de cancelamentos de reservas, permitindo que as empresas tomem decisões fundamentadas para aprimorar a gestão de reservas e a utilização eficiente de recursos. Por meio da aplicação de técnicas

e recursos avançados da *Data Science*, colaboram para a estimação, previsão, coleta de dados, otimização e controle, desempenhando papel fundamental na gestão de receitas.

O artigo "*Evaluación de destinos turísticos mediante la tecnología de la ciencia de datos*" aborda a aplicação de técnicas de ciência de dados para avaliar a competitividade de destinos turísticos com base em comentários e opiniões de turistas. Destaca-se nesta pesquisa a relevância da mineração de opiniões como uma técnica complementar à análise de texto tradicional, permitindo a identificação da polaridade dos comentários (positiva, negativa ou neutra) e sua relação com os serviços e agentes envolvidos. Ao apresentar uma abordagem prática e inovadora para avaliar a competitividade dos destinos turísticos por meio da mineração de opiniões dos turistas.

Baseado na análise de mineração de dados, os autores da pesquisa "Similaridades e correlação entre as dormidas dos residentes e a informação do *Google Trends* para Portugal e suas regiões de turismo" sugerem que o uso do *Google Trends* (GT) pelas organizações de turismo, especialmente as organizações de gestão de destinos, por permitir monitorar os interesses dos turistas em uma região ou país específico em tempo quase real, identificando inclusive picos de interesse. Além disso, os tomadores de decisão no setor turístico podem utilizar os dados do GT para estimar a demanda turística em determinados territórios e desenvolver estratégias de marketing adequadas, adaptando mensagens aos diferentes canais, tipos de consumidores e estágios da viagem.

A análise de dados no turismo fornece informações essenciais para os gestores de destinos turísticos e outras partes interessadas, permitindo a identificação de áreas de melhoria e a formulação de estratégias para aprimorar a qualidade dos serviços. Além disso, ao identificar os serviços que impactam negativamente a imagem e competitividade do destino, podem ser implementadas medidas corretivas para melhorar a experiência dos visitantes e impulsionar o desempenho do setor turístico. Sua importância para o desenvolvimento dos indicadores do turismo reside na capacidade de fornecer informações valiosas para melhorar a qualidade dos serviços, identificar áreas de melhoria e impulsionar a competitividade do setor turístico como um todo.

As abordagens utilizadas no estudo "Avaliação do impacto das Tecnologias da Informação e Comunicação no setor hoteleiro português: uma análise exploratória com *Data Envelopment Analysis*" e "Desempenho das capitais brasileiras como destinos turísticos", utilizam o modelo DEA (Análise Envoltória de Dados) que consiste numa metodologia de avaliação de desempenho e eficiência utilizada para comparar a eficiência relativa de unidades homogêneas, como empresas, organizações ou unidades produtivas. A DEA permite a medição da eficiência considerando múltiplos inputs e outputs simultaneamente, o que a torna uma abordagem poderosa para avaliar a produtividade e identificar as melhores práticas. A DEA constrói uma fronteira eficiente que representa o desempenho máximo alcançável pelos diferentes participantes em um conjunto de dados.

Na análise realizada pelos autores Paço e Pérez (2015), é possível observar que utilizando o método DEA como a Tecnologia da Informação e Comunicação (TIC) têm o potencial de gerar um impacto positivo no setor hoteleiro. As TIC permitem que as empresas alcancem um grande número de pessoas sem limitações geográficas, criando valor agregado para os clientes. Além de, possibilitar lançar rapidamente no mercado novos produtos e serviços, incluindo catálogos online com descrições e preços, proporcionando uma maior disponibilidade para os clientes finais. Ademais, a adoção das TIC também pode resultar em um aumento na receita proveniente de serviços adicionais, devido à abertura de novos canais de distribuição e vendas.

O estudo de Paço e Pérez (2015) contribui significativamente para a análise do turismo e seus indicadores por várias razões. Ao identificar a eficiência dos hotéis no

setor hoteleiro português, oferece uma visão clara de como a implementação das TIC pode impactar a eficiência operacional das empresas hoteleiras. Para determinar a fronteira eficiente dos hotéis e identificar aqueles que são mais e menos eficientes, fornecendo um ponto de referência para avaliar o desempenho dos hotéis em termos de geração de estadias. Favorece a entender sua posição em relação à concorrência e quais áreas precisam ser aprimoradas para melhorar a eficiência e, conseqüentemente, aumentar o retorno sobre o investimento.

Na segunda pesquisa que também utiliza o modelo DEA (Análise Envoltória de Dados) para avaliar o desempenho dos destinos turísticos no Brasil. O objetivo principal foi construir fronteiras de produção e medir a eficiência em relação a essas fronteiras, empregando uma abordagem de programação matemática. Santos e Fortes (2015), evidenciam a importância crucial dos dados disponíveis para realizar essa análise. Apesar de reconhecer as limitações decorrentes da falta de dados em alguns setores e destinos específicos, os autores enfatizam a necessidade de obter informações abrangentes e atualizadas para alcançar conclusões mais precisas. Isso ressalta a relevância de coletar e utilizar dados de forma eficaz no setor do turismo, visando obter *insights* valiosos para melhorar a gestão e o desempenho dos destinos.

No estudo de Moreno-Rojas *et al.* (2017), "*Determinantes del tipo impositivo efectivo en el sector turístico: un modelo dinámico con datos de panel*", apresentam uma abordagem ao estudo dos determinantes da carga tributária por dois tipos de empresas muito relevantes no setor do turismo, como hotéis e agências de viagens. Utilizando dados em painel para analisar a importância de certas variáveis, como tamanho da empresa, estrutura financeira, rentabilidade, ativos fixos e tipo de empresa (hotéis ou agências de viagens), como determinantes da Taxa Efetiva de Imposto (*Effective Tax Rate - ETR*) ao longo do período de 2008-2013 para empresas espanholas. Uma das principais contribuições do estudo foi a identificação dos fatores que afetam a carga tributária dessas empresas turísticas. Ao analisar variáveis como tamanho da empresa, estrutura financeira, rentabilidade e ativos fixos, o estudo destaca quais desses fatores têm maior impacto na ETR. Isso pode auxiliar no desenvolvimento de indicadores mais precisos e abrangentes para avaliar a saúde financeira e a competitividade das empresas do setor de turismo.

Os artigos analisados desempenham um papel fundamental no avanço do conhecimento do fenômeno do turismo, na análise de dados turísticos e no desenvolvimento e aprimoramento dos indicadores existentes. As informações levantadas e analisadas por meio da *Data Science* possibilitam realizar a previsão e predição de cancelamentos de reservas, avaliação da competitividade de destinos turísticos, o monitoramento de interesses dos turistas em tempo real e a utilização das TIC para impulsionar o setor hoteleiro. Essas contribuições têm um impacto significativo na gestão de destinos, na tomada de decisões informadas e no aprimoramento da qualidade dos serviços turísticos, impulsionando a competitividade do setor como um todo.

Compreender como o uso da *Data Science* pode proporcionar uma análise mais ampla e objetiva do desenvolvimento do turismo é necessária. Sendo, inicialmente preciso abordar questões relacionadas ao uso da estatística em pesquisas turísticas, isto é, melhor organização dos dados e métodos de análise para que seja possível atingir as metas apresentadas no Plano Nacional de Turismo.

O Plano Nacional de Turismo 2018-2022 destaca que existe uma carência de pesquisas mais específicas sobre como aplicar de forma eficaz os conhecimentos fornecidos pela estatística no setor do turismo, ou seja, como dados sobre o fluxo de turistas em uma região podem ser usados no seu desenvolvimento e planejamento (PNT 2018-2020).

Nesse sentido, o Ministério do Turismo (2013) já ressaltava a importância de realizar estudos que subsidiem as tomadas de decisão, incluindo pesquisas sobre a oferta e demanda turística, a satisfação dos visitantes, oportunidades de investimentos e financiamento e impactos gerados pelo turismo; bem como a identificação de atuais e potenciais produtos e roteiros turísticos. Desta forma, o levantamento desses dados é fundamental para a definição de estratégias, o enfrentamento de crises e a análise de cenários do turismo, sendo crucial o papel do Ministério do Turismo em fomentar e incentivar tais pesquisas que é identificado como carente, mas importante.

Apesar dos avanços do uso de tecnologias digitais no Brasil, ainda há escassez de estudos e pesquisas sobre o emprego da *Data Science* no turismo. A insuficiência de dados precisos e confiáveis, juntamente com a falta de conhecimento sobre os benefícios do uso da *Data Science* é um fator limitante para o desenvolvimento de estudos e pesquisas na área. É fundamental que o setor invista em soluções de coleta e gerenciamento de dados para melhorar a quantidade e qualidade de informações disponíveis para o planejamento e gerenciamento do turismo.

CONSIDERAÇÕES FINAIS

O presente artigo discutiu o uso da *Data Science* como ferramenta na análise de indicadores do turismo. A *Data Science* apresenta-se como uma ferramenta para identificar, tendências, padrões e correlações que poderiam passar despercebidos em análises convencionais. Esse conhecimento pode ser usado para desenvolver políticas públicas e contribuir para a tomada de decisões mais precisas no turismo.

É importante ressaltar que, a utilização da *Data Science* na análise dos indicadores do turismo requer uma infraestrutura tecnológica e de dados sólidos. Porém, é igualmente necessário considerar a proteção da privacidade dos dados pessoais dos turistas e demais agentes envolvidos.

Em decorrência da evolução dos mecanismos e técnicas de pesquisa, em particular na área da Ciência de Dados, estão surgindo novas possibilidades para a pesquisa e desenvolvimento no campo do turismo. No entanto, ainda há desafios a serem superados em relação aos métodos científicos e técnicas para analisar grandes volumes de dados e solucionar problemas relacionados à pesquisa acadêmica.

Apesar da insuficiência de estudos que correlacionam *Data Science* e Turismo ser incipiente no Brasil, é necessário que novos estudos sejam realizados a fim de elucidar questões ainda não exploradas. Desta forma é fundamental o desenvolvimento de pesquisas com *Data Science*, uma vez que o próprio Órgão Gestor do Turismo no país reconhece a carência e a importância da análise de dados estatísticos, sendo essencial que o mesmo fomente e incentive tais pesquisas para o desenvolvimento do turismo.

A *Data Science* surge como um meio para melhor compreender o turismo e aprimorar suas práticas, além de favorecer a análise do desenvolvimento e a avaliação das políticas pensadas para a atividade. À medida que as aplicações e instrumentos de pesquisa continuarem a evoluir, o uso da ciência de dados continuará a se expandir, contribuindo cada vez mais para o desenvolvimento sustentável e sucesso do turismo.

REFERÊNCIAS

- ANTÓNIO, Nuno; ALMEIDA, Ana; NUNES, Luis. Modelos preditivos de cancelamento de reservas de hotéis: uma análise semiautomática da literatura. **Tourism & Management Studies**, v. 15, n. 1, p. 7-21, 2019. Disponível em: < <https://doi.org/10.18089/tms.2019.15011> >. Acesso em: 10. jun. 2023.
- ARRIGONI, Rodrigo. **Big data: Conceitos, tecnologias e aplicativos**. São Paulo: Novatec Editora, 2013.
- BARDIN, Laurence. *Análise de conteúdo*. Edições 70. Lisboa, 1977, rev. 2007.
- BARDIN, Laurence. *Análise de conteúdo*. São Paulo: Edições 70, 2011.
- BELLEN, Hans Michael van. *Indicadores de sustentabilidade: uma análise comparativa*. Editora: SENAC, São Paulo, 2005.
- BIZ, Alexandre Augusto.; DE SOUZA, João Arthur de Souza; CAUMO, Rafael Bassegio. *Big Data para a geração de indicadores de turismo: fontes de dados e aplicações*. **Revista E-Tech: Tecnologias para Competitividade Industrial**, 13(1), 53-70, 2020.
- BROWN, Culum., & LALAND, Kevin. N. *Social learning in animals: The roots of culture*. *Academic Press*.pp. 42-45. 2001.
- BUHALIS, Dimitrios; LAW, Rob. *Progress in information technology and tourism management: 20 years on and 10 years after the Internet—The state of eTourism research*. *Tourism Management*, 29(4), 609-623.2008. 2008. Disponível em: < <https://doi.org/10.1016/j.tourman.2008.01.005> >. Acesso em: 10.fev.2023.
- NETTO, Joviniano Pereira da Silva. *Indicadores de Sustentabilidade como Suporte ao Planejamento do Turismo: Aspectos Conceituais e Metodológicos*. **Rosa dos Ventos**, vol. 13, núm. 1, pp. 260-277, 2021.
- CAMARGO, Hélio Luís. *Indicadores: conceitos e aplicações*. São Paulo: Editora Atlas, 2000.
- CAVALCANTE, Ricardo Bezerra; CALIXTO, Pedro; PINHEIRO, Marta Macedo Kerr. *Content analysis: General considerations, relationship with the research question, possibilities, and limitations of the method*. *Inf. & Soc.:Est.*, João Pessoa, v.24, n.1, p. 13-18, Jan./Apr. 2014.
- PAÇO, Cidália Maria Leal; PÉREZ, Juan Manuel Cepeda. *Avaliação do impacto das Tecnologias da Informação e Comunicação no setor hoteleiro português: uma análise exploratória com Data Envelopment Analysis*. **Tourism & Management Studies**, 11, 35-43, 2015.

COATES, Vary; FAROOQUE, Mahmud; KLAVANS, Richard; LAPID, Koty; LINSTONE, Harold; A, Linstone; PISTORIUS, Carl; PORTER, Alan. *On the Future of Technological Forecasting*. North-Holland: Elsevier Science, 2001.

DAVENPORT, Thomas; HARRIS, Jeanne. *Competing on analytics: The new science of winning*. Cambridge, USA: Harvard Business Press. 2017.

DDI - *Data Documentation Initiative: Structural Reform Group*. DDI Version 3.0 *Conceptual Model*. DDI Alliance. 2004. Disponível em: <<http://libraries.mit.edu/guides/subjects/datamanagement/cycle.html>>. Acessado em: <10 mai. 2023>.

DINIS, Gorete; COSTA, Carlos; PACHECO, Osvaldo. *Similarities and correlation between resident tourist overnights and Google Trends information in Portugal and its tourism regions*. *Tourism & Management Studies*, 13(3), 15-22. 2017. Disponível em: <<https://doi.org/10.18089/tms.2017.13302>>. Acesso em 09. jun. 2023.

FUCHS, Matthias; ABADZHIEV, Andrey; SVENSSON, Bo; HÖPKEN, Wolfram; LEXHAGEN, Maria. *A Knowledge Destination Framework for Tourism Sustainability – a Business Intelligence Application from Sweden*, *Tourism - An Interdisciplinary Journal*, 61(2), 121-148. 2014. Disponível em: <<https://miun.diva-portal.org/smash/record.jsf?pid=diva2%3A668442&dswid=434>>. Acesso em 08. jun. 2023.

GRUS, Joel. *Data Science from Scratch First Principles: With Python*. O'Reilly Media, Inc. 2015.

GUIMARÃES, Sueli Édir Rufini. Educação Matemática e Transformação da Sociedade., Docência em Educação Matemática: Temas e Debates (pp. 65-83). Campinas, SP: Papirus, 2007.

HARMER, Jeremy. *The practice of English language teaching (4th ed.)*. Pearson Longman. p. 57-58, 2007.

HANAI, Frederico Yuri. *Gestão de destinos turísticos: aplicação prática*. São Paulo: Editora Senac São Paulo, 2009.

HASSAN, Hussein Ali - *Tecnologias de informação e turismo : e-tourism*. Coimbra: [s.n.], 2011.

HAYASHI, Chikio. *What is Data Science ? Fundamental Concepts and a Heuristic Example*. In: Hayashi, C., Yajima, K., Bock, HH., Ohsumi, N., Tanaka, Y., Baba, Y. (eds) *Data Science, Classification, and Related Methods. Studies in Classification, Data Analysis, and Knowledge Organization*. Springer, Tokyo. 1998. Disponível em: <https://doi.org/10.1007/978-4-431-65950-1_3>. Acesso em: 10.fev.2023.

LI, Xin; LAW, Rob. (2020). Network analysis of big data research in tourism. *Tourism Management Perspectives*, 33, 100608. 2020 .Disponível em: <<https://doi.org/10.1016/j.tmp.2019.100608>> Acesso em: 10. mai. 2023

MORENO-ROJAS, José; GONZÁLEZ-RODRÍGUEZ, Rosário; MARTÍN-SAMPER, Rosário Carmem. *Determinants of the effective tax rate in the tourism sector: a dynamic panel data model*. *Tourism & Management Studies*, vol. 13, núm. 3, pp. 31-38. 2017. Disponível em: <<https://www.redalyc.org/articulo.oa?id=388753560003>>. Acesso em: 12 jun 2023.

MARTINS, Cristina; FIATES, Gabriela; PINTO, Adilson Luiz. (2016). A relação entre os clusters de turismo e tecnologia e seus impactos para o desenvolvimento local: Um estudo bibliométrico da produção científica. *Revista Brasileira de Pesquisa em Turismo*, 10(1), 65–88. 2016.

MIRANDA, Antônio; SIMEÃO, Elmira Luzia Melo Soares. Conceituação de massa documental e a interação da tecnologia com o conhecimento registrado. *Revista Ciência da Informação*. v.3, n4. 2003. 43-49 Disponível em: <http://www.datagramazero.org.br/ago02/F_I_art.htm>. Acesso em: 26 mai. 2023.

Ministério do Turismo. mais pelo Brasil 2013 2016. Ministério do Turismo. Plano Nacional do Turismo: o turismo fazendo muito –. Brasília/DF, 2013. Disponível em: <<https://www.gov.br/turismo/pt-br/centrais-de-conteudo/plano-nacional-2013-pdf>>. Acesso em 26 mai 2023.

MTUR – Ministério do Turismo. Plano Nacional de Turismo 2018 – 2022. Brasília, 2018. Disponível em: http://www.turismo.gov.br/images/pdf/PNT_2018-2022.pdf. Acesso em: 10 fev. 2023.

MOLINAR, Carlos Mario Amaya; ESPINOZA, Pedro Magan; LLAMAS, Ileana Ochoa. Evaluación de destinos turísticos mediante la tecnología de la ciencia de datos. **Estudios y Perspectivas en Turismo**, vol. 26, n. 2, pp. 286-305. 2017. Disponível em: <<https://www.redalyc.org/articulo.oa?id=180750377004>>. 08. jun. 2023.

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. 2017.

OTT, W. *Environment indices: theory and practice*. Michigan: Arnn Arbor, 1978.

PORTER, Alan L. *How tech mining can enhance R&D management*. *Research Technology Management*, 50 (2), 15. 2007. Disponível em: <<https://doi.org/10.1080/08956308.2007.11657425>>. Acesso: 26. mai. 2023.

PROVOST, Foster; FAWCETT, Tom. *Data science for business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media, Inc. 2013.

PUSCHMANN, Cornelius.; BURGESS, Jean. *Metaphors of Big Data*. *International Journal of Communication*, v. 8, n. 1, p. 1690-1709, 2014.

RAUDEMBERG, Sandro; CARMO, Paulo Ricardo. Ciência de dados e seu papel na transformação digital: Uma revisão integrativa. *Perspectivas em Ciência da Informação*,

24(4), 70-87. 2019. Disponível em: < <https://doi.org/10.1590/1981-5344/3907> >. Acesso em: 10. fev. 2023.

SANTOS, Erick da Silva; FORTES, Mauri. Desempenho das capitais brasileiras como destinos turísticos. *Revista Brasileira De Pesquisa Em Turismo*, 9(1), 22–39. 2015. Disponível em: <<https://doi.org/10.7784/rbtur.v9i1.697>> Acesso: 10. fev. 2023.

UNITED NATIONS. *System of National Accounts 2008*. New York: United Nations, 2010.

UNWTO. *Indicators of Sustainable Development for Tourism Destinations: A Guidebook*. Madrid: World Tourism Organization, 2007.

WORLD TOURISM ORGANIZATION – Unwto & Tourism Satellite Account Programme. *A statistical project to support mainstreaming tourism in sustainable development. Project Proposal*. 2015.

WORLD TOURISM ORGANIZATION. *Indicators of sustainable tourism: Development and applications*. UNWTO. 2018. Recuperado de <https://www.e-unwto.org/doi/book/10.18111/9789284420295>.

WORLD TOURISM ORGANIZATION. *Tourism and the internet: The ideal partners*. Madrid, Spain: World Tourism Organization. 2001

***The use of Data Science in the analysis of Tourism indicators:
a review of research in Brazil (2012-2022)***

Abstract

This article presents Data Science as it discusses the intersection between tourism and data science, particularly the use of Data Science to analyze the data generated by tourism. The exponential increase in the volume of data generated by tourism requires new analytical approaches, and statistically driven analysis using computational technologies becomes essential to understand trends, preferences, and consumer behavior, and to offer more personalized and efficient services. The study examines the applicability of Data Science in the analysis of tourism indicators, based on the analysis, identification, and evaluation of commonly used indicators in tourism development. To do so, a survey was conducted in the Scielo and Redalyc databases over the last 10 years (2012-2022), and content analysis was performed using skimming and scanning techniques. In summary, 180 scientific productions were identified, but there is a lack of studies that relate Data Science and Tourism in Brazil. The use of Data Science in tourism has enormous potential to contribute to the development of more effective and sustainable policies and guidelines in the sector.

Keywords: *Tourism. Data science. Tourism indicators. Data analysis.*