

SENTIMENTOS DE TURISTAS EM DESTINO EMERGENTE: ANÁLISE LEXICOMÉTRICA E DE SENTIMENTOS DE AVALIAÇÕES NO TRIPADVISOR SOBRE O LITORAL DO PIAUÍ - BRASIL

TOURIST SENTIMENTS IN AN EMERGING DESTINATION: LEXICOMETRIC AND SENTIMENT ANALYSIS OF TRIPADVISOR REVIEWS ON THE PIAUÍ COAST - BRAZIL

SENTIMIENTOS TURÍSTICOS EN UN DESTINO EMERGENTE: ANÁLISIS LEXICOMÉTRICO Y DE SENTIMIENTOS DE RESEÑAS EN TRIPADVISOR SOBRE EL LITORAL DE PIAUÍ - BRASIL

André Riani Costa Perinotto^{1 2} (perinotto@ufdpar.edu.br) 
Maria Clara Silva da Rocha¹ (claraaslv@outlook.com) 
Priscila Cembranel² (priscila_cembranel@yahoo.com.br) 

¹Universidade Federal do Delta do Parnaíba, Piauí, Brasil

²Universidade Federal do Paraná, Paraná, Brasil

RESUMO:

Objetivo: Compreender como os sentimentos e os padrões discursivos expressos no conteúdo gerado por usuários (CGU) configuram a imagem digital do litoral do Piauí como destino turístico emergente.

Desenho/metodologia/abordagem: Adotou-se um desenho de métodos mistos convergente, de caráter exploratório e descritivo. Foram analisados 2.400 comentários sobre 16 atrativos do TripAdvisor, publicados entre 2011 e 2023. A dimensão quantitativa combinou Classificação Hierárquica Descendente (CHD), Análise de Similitude e Análise Fatorial de Correspondência (AFC) no IRAMUTEQ com classificação independente de polaridade pelo DistilBERT multilíngue; a dimensão qualitativa compreendeu a interpretação temática das classes lexicais. A classificação automatizada foi submetida à validação manual por dois pesquisadores.

Resultados: Identificaram-se quatro categorias temáticas: natureza e ecoturismo; lazer e hospitalidade; cultura e patrimônio histórico; e infraestrutura e serviços. A AFC explicou 62,82% da inércia total. A classificação automatizada registrou 42,8% de sentimentos positivos, 33,8% neutros e 14,4% negativos; para 9,0% dos comentários, a classificação original não pôde ser recuperada nos registros preservados. O conjunto evidencia uma imagem globalmente favorável, porém heterogênea.

Limitações/implicações da pesquisa: O *corpus* restringe-se a uma plataforma e agrega avaliações de doze anos, o que introduz heterogeneidade temporal; os registros preservados não permitem recuperar o tamanho da subamostra de validação nem um coeficiente formal de concordância.

Implicações práticas: Os achados subsidiam gestores públicos, organizações do turismo e agentes culturais na priorização de infraestrutura e na valorização dos atributos naturais e culturais.

Originalidade/valor: O estudo combina técnicas lexicométricas e classificação automatizada de sentimentos em um destino emergente brasileiro pouco representado na literatura, oferecendo um protocolo analítico transferível a outros contextos.

Palavras-chave: análise de sentimentos; conteúdo gerado por usuários; imagem de destino; TripAdvisor; litoral do Piauí; destinos emergentes.

Informações Editoriais:

Double Blind Review

Submissão: 23/04/2026

Avaliação: 27/05/2026

Aceite: 21/09/2026

Editor:

Luiz Carlos da Silva Flores

Assistente Editorial:

Alexandre Ferraz

Disponibilidade dos dados:

Os dados estão disponíveis no corpo do artigo. Os dados podem ser solicitados aos autores.

ABSTRACT:

Purpose: To understand how sentiments and discursive patterns expressed in user-generated content (UGC) configure the digital image of the Piauí coast as an emerging tourism destination.

Design/methodology/approach: A convergent mixed-methods design with an exploratory and descriptive scope was adopted. The study analyzed 2,400 TripAdvisor reviews of 16 attractions published between 2011 and 2023. The quantitative dimension combined Descending Hierarchical Classification (DHC), Similarity Analysis and Correspondence Factor Analysis (CFA) in IRAMUTEQ with independent polarity classification using multilingual DistilBERT; the qualitative dimension involved thematic interpretation of the lexical classes. The automated classification was manually validated by two researchers.

Findings: Four thematic categories were identified: nature and ecotourism; leisure and hospitality; culture and historical heritage; and infrastructure and services. The CFA explained 62.82% of total inertia. Automated classification recorded 42.8% positive, 33.8% neutral, and 14.4% negative sentiments; for 9.0% of the reviews, the original classification could not be recovered from the preserved records. Overall, the findings indicate a broadly favorable yet heterogeneous image.

Research limitations/implications: The corpus is restricted to one platform and aggregates reviews over a twelve-year period, introducing temporal heterogeneity; the preserved records do not allow recovery of the validation subsample size or a formal inter-rater agreement coefficient.

Practical implications: The findings support public managers, tourism organizations and cultural agents in prioritizing infrastructure and leveraging the destination's natural and cultural attributes.

Originality/value: The study combines lexicometric techniques and automated sentiment classification in an underrepresented Brazilian emerging destination, offering an analytical protocol transferable to other contexts.

Keywords: sentiment analysis; user-generated content; destination image; TripAdvisor; Piauí coast; emerging destinations.

RESUMEN:

Objetivo: Comprender cómo los sentimientos y los patrones discursivos expresados en el contenido generado por el usuario (CGU) configuran la imagen digital del litoral de Piauí como destino turístico emergente.

Diseño/metodología/enfoque: Se adoptó un diseño convergente de métodos mixtos, de carácter exploratorio y descriptivo. Se analizaron 2.400 reseñas de TripAdvisor sobre 16 atractivos, publicadas entre 2011 y 2023. La dimensión cuantitativa combinó Clasificación Jerárquica Descendente (CJD), Análisis de Similitud y Análisis Factorial de Correspondencia (AFC) en IRAMUTEQ con clasificación independiente de polaridad mediante DistilBERT multilingüe; la dimensión cualitativa comprendió la interpretación temática de las clases léxicas. La clasificación automatizada fue sometida a validación manual por dos investigadores.

Hallazgos: Se identificaron cuatro categorías temáticas: naturaleza y ecoturismo; ocio y hospitalidad; cultura y patrimonio histórico; e infraestructura y servicios. El AFC explicó el 62,82% de la inercia total. La clasificación automatizada registró 42,8% de sentimientos positivos, 33,8% neutros y 14,4% negativos; para el 9,0% de las reseñas, la clasificación original no pudo recuperarse de los registros conservados. En conjunto, los hallazgos evidencian una imagen globalmente favorable, aunque heterogénea.

Limitaciones/implicaciones de la investigación: El corpus se restringe a una plataforma y agrega reseñas de doce años, lo que introduce heterogeneidad temporal; los registros conservados no permiten recuperar el tamaño de la submuestra de validación ni un coeficiente formal de concordancia entre evaluadores.

Implicaciones prácticas: Los hallazgos apoyan a gestores públicos, organizaciones turísticas y agentes culturales en la priorización de infraestructura y en la valorización de los atributos naturales y culturales.

Originalidad/valor: El estudio combina técnicas lexicométricas y clasificación automatizada de sentimientos en un destino emergente brasileño poco representado en la literatura, ofreciendo un protocolo analítico transferible a otros contextos.

Palabras clave: análisis de sentimientos; contenido generado por el usuario; imagen de destino; TripAdvisor; litoral de Piauí; destinos emergentes.

INTRODUÇÃO

A transformação digital tem reconfigurado profundamente as práticas turísticas contemporâneas. Planejar viagens, escolher destinos e avaliar experiências tornou-se uma atividade essencialmente mediada por plataformas digitais e pelo conteúdo que os próprios usuários produzem e compartilham *online* (Otabek, 2024; Sivarethinamohan, 2023). Nesse cenário, o conteúdo gerado por usuários (CGU), avaliações, comentários, fotos e vídeos publicados voluntariamente em plataformas especializadas, emerge como uma das fontes de informação mais influentes na formação da imagem dos destinos e na tomada de decisão de viagem (Liu, 2010; Kaplan & Haenlein, 2010). Contudo, apesar do crescente interesse acadêmico pelo tema, persistem lacunas relevantes no entendimento sobre como o CGU opera especificamente em destinos emergentes brasileiros, onde a escassez de estudos reduz a capacidade de gestores e formuladores de política turística de interpretar as percepções espontâneas dos visitantes.

O litoral do Piauí destaca-se como um dos destinos emergentes mais promissores do Brasil, reconhecido por atrativos como o Delta do Parnaíba, a Praia de Atalaia, a Lagoa do Portinho e o patrimônio histórico-cultural de Parnaíba. Com a expansão do acesso à internet e das plataformas de avaliação digital, turistas passaram a registrar e compartilhar suas experiências de forma espontânea, criando um acervo de percepções que reflete, em tempo real, os atributos valorizados e as fragilidades percebidas no destino (Dutra & Gosling, 2021; Pérez et al., 2022). Esse tipo de dado, oriundo de comentários publicados em plataformas como o TripAdvisor, possibilita análises sistemáticas por meio de técnicas de mineração de texto e análise de sentimentos, capazes de revelar padrões discursivos e emocionais que métodos tradicionais dificilmente captariam (Pang & Lee, 2008; Cambria et al., 2013).

Estudos anteriores demonstram que as avaliações online exercem impacto direto sobre a atratividade dos destinos e a intenção de visita (Fotis et al., 2012; Santos et al., 2024), funcionando como um mecanismo de comunicação que, em muitos casos, supera o *marketing* institucional em credibilidade e alcance (Litvin et al., 2008). No contexto dos destinos emergentes, essa dinâmica é especialmente significativa: a imagem construída digitalmente pode antecipar ou substituir a experiência presencial na percepção de potenciais visitantes, tornando o monitoramento e a interpretação do CGU uma ferramenta estratégica indispensável para o planejamento turístico (Xiang & Gretzel, 2010).

Nesse contexto, o TripAdvisor destaca-se como uma das plataformas que reúnem um vasto acervo de avaliações e comentários de turistas do mundo inteiro sobre atrações, hospedagens e serviços. Essas informações impactam diretamente a percepção e a tomada de decisão dos futuros viajantes, tornando-se uma ferramenta estratégica de marketing e posicionamento de destinos turísticos (Pérez et al., 2022).

As opiniões compartilhadas por turistas nas redes sociais e em sites de avaliação exercem forte influência sobre a decisão de viagem, podendo valorizar ou desestimular determinados destinos. Essas avaliações refletem vivências pessoais, trazendo à tona tanto elogios quanto críticas relacionadas à infraestrutura, atendimento e conservação dos locais visitados (Freire et al., 2022; Pérez et al., 2022). Além disso, essas manifestações espontâneas permitem acessar sentimentos expressos textualmente pelos usuários e revelar percepções que dificilmente seriam obtidas por métodos tradicionais (Dutra & Gosling, 2021).

Com a ascensão do conteúdo gerado por usuários, a *internet* transformou-se em um espaço dinâmico de troca de experiências. Fóruns, *blogs*, grupos e plataformas especializadas reúnem avaliações que influenciam diretamente a imagem dos destinos turísticos, promovendo melhorias nos serviços e fortalecendo o relacionamento com o visitante (Liu, 2010).

Os comentários disponíveis no TripAdvisor representam uma fonte rica para essa investigação, revelando impressões que afetam diretamente a atratividade e a escolha do destino. Técnicas de análise de sentimentos têm se mostrado eficazes na interpretação dessas opiniões, sendo amplamente utilizadas em áreas como *marketing*, finanças e políticas públicas (Dutra & Gosling, 2021). Dentre as possibilidades, o TripAdvisor permite uma análise de sentimentos capaz de fornecer um retrato mais fiel do destino do que métodos tradicionais (Vetitnev et al., 2021), além de favorecer a compreensão de opiniões e avaliações dos turistas (Gao et al., 2015), de suas motivações e dos atributos elogiados (Ramos & George, 2024).

Diante desse cenário, este estudo parte da seguinte questão de pesquisa: como os sentimentos expressos pelos turistas no TripAdvisor contribuem para a construção da imagem do litoral do Piauí enquanto destino emergente? Com vistas a responder a essa questão, o objetivo geral é analisar como o conteúdo gerado por usuários (CGU) na plataforma TripAdvisor configura, no plano discursivo e sentimental, a imagem do litoral piauiense enquanto destino turístico emergente, a partir da identificação de padrões sentimentais, temáticos e discursivos nos comentários dos turistas. Como objetivos específicos, busca-se: (a) identificar a proporção e a natureza dos sentimentos positivos, neutros e negativos examinados nas avaliações, considerando também a parcela cuja classificação original não foi passível de recuperação nos registros disponíveis; (b) mapear os aspectos mais mencionados pelos visitantes, incluindo infraestrutura, hospitalidade, beleza natural e patrimônio cultural; e (c) categorizar as principais temáticas dos comentários, relacionando-as às discussões sobre formação da imagem de destinos emergentes e às implicações práticas para a gestão do turismo local.

Do ponto de vista do *gap* de pesquisa, identificam-se duas lacunas que justificam e orientam este estudo. A primeira é teórico-empírica: embora a análise de sentimentos aplicada ao turismo seja um campo em expansão (Vetitnev et al., 2021; Ramos & George, 2024), os destinos emergentes brasileiros permanecem sub-representados na literatura nacional e internacional, especialmente aqueles localizados fora do eixo Sul-Sudeste. A segunda é metodológica: parte expressiva dos estudos brasileiros sobre CGU e imagem de destino recorre a análises de conteúdo tradicionais, sem combinar lexicometria, estatística textual e classificação automatizada de polaridade. Este estudo posiciona-se nessa interseção, oferecendo evidência empírica sobre um destino emergente e um protocolo metodológico integrado para a pesquisa em turismo digital no Brasil.

REVISÃO TEÓRICA

A transformação digital provocou mudanças na forma como os turistas pesquisam, escolhem e avaliam destinos. O TripAdvisor desponta como uma das plataformas de compartilhamento de experiências turísticas, permitindo que os viajantes publiquem avaliações, comentários e recomendações sobre hospedagens, atrações e restaurantes (Miguéns et al., 2008; Dutra & Gosling, 2021).

A ascensão das plataformas digitais ampliou o alcance e a credibilidade do conteúdo gerado por usuários (CGU), que compreende textos, fotos, vídeos e demais formatos publicados voluntariamente por consumidores na *internet* (Liu, 2010). No turismo, esse tipo de conteúdo substitui gradualmente as fontes institucionais tradicionais, sendo percebido como mais autêntico, confiável e influente nas decisões de viagem (Kaplan & Haenlein, 2010; Xiang & Gretzel, 2010).

Sabe-se que as avaliações online exercem impacto direto sobre a atratividade dos destinos, influenciando não apenas a imagem, mas também a intenção de visita (Pérez et al., 2022; Fotis et al., 2012). Essa influência ocorre porque os turistas enxergam, nas experiências compartilhadas por outros visitantes, elementos de identificação, validação e comparação. Sigala et al. (2012) e Gretzel et al. (2007) reforçam que o CGU contribui ativamente para a construção coletiva da reputação dos destinos, tornando-se uma ferramenta estratégica tanto para consumidores quanto para gestores.

Além disso, o TripAdvisor permite interações entre usuários e disponibiliza recursos como *rankings*, fóruns, filtros de pesquisa, notas e mapas, consolidando-se como uma fonte de dados rica e diversificada (Dutra & Gosling, 2021). Essas informações oferecem insumos relevantes para análise sistemática da experiência turística, principalmente em destinos emergentes ainda pouco estudados.

Litvin et al. (2008) argumentam que o CGU, em muitos casos, supera o *marketing* tradicional, justamente por ser espontâneo e emocionalmente carregado. Tal característica torna as avaliações *online* especialmente valiosas para estudos sobre percepção turística, pois revelam não apenas dados objetivos, mas também sentimentos, expectativas e frustrações.

A imagem de um destino turístico é resultado da combinação entre atributos objetivos (como infraestrutura, preço e clima) e percepções subjetivas, influenciadas pelas emoções e experiências dos visitantes (Echtner & Ritchie, 1993; Gallarza et al., 2002). Em um cenário cada vez mais digital, essa percepção passa a ser formada e comunicada também por meio de comentários e avaliações publicadas *online*.

Para evitar o uso intercambiável de termos afins, adota-se a seguinte distinção conceitual. Atributo corresponde a uma característica identificável do destino, como infraestrutura, preço ou paisagem (Echtner & Ritchie, 1993). Percepção refere-se ao processo cognitivo de interpretação desses atributos, enquanto avaliação corresponde ao julgamento favorável ou desfavorável resultante dessa percepção (Beerli & Martín, 2004). Emoção designa um estado afetivo associado a episódios da experiência turística (Hosany, Martin, & Woodside, 2021). Sentimento, tal como operacionalizado neste estudo, corresponde à polaridade linguística positiva, neutra ou negativa inferida automaticamente do texto, constituindo uma aproximação computacional simplificada do componente afetivo (Pang & Lee, 2008; Liu, 2012). A imagem de destino constitui um construto de ordem superior que integra percepções, avaliações e componentes afetivos associados a múltiplos atributos (Gallarza et al., 2002; Lai & Li, 2016; Yang et al., 2022). Assim, esta pesquisa mensura polaridade textual e percepção discursiva, sem pretender medir diretamente as emoções vividas pelos turistas.

Segundo Beerli e Martín (2004), fatores como paisagens naturais, hospitalidade, segurança e serviços influenciam a imagem do destino. O componente afetivo, por sua vez, participa da avaliação global da experiência. Freire et al. (2022) mostram que estados afetivos como alegria, frustração e encantamento se relacionam às expectativas construídas antes da viagem e às experiências vividas durante a estadia.

Nesse contexto, técnicas de análise de sentimentos emergem como ferramentas eficazes para identificar a polaridade positiva, negativa ou neutra em relatos publicados em plataformas digitais (Pang & Lee, 2008; Cambria et al., 2013). Essa abordagem integra a área de mineração de opinião e utiliza métodos baseados em léxicos ou algoritmos de *machine learning* para extrair e interpretar avaliações expressas em textos (Liu, 2010; Liu, 2012).

Liu (2012) destaca que a análise de sentimentos cresceu exponencialmente no campo do Processamento de Linguagem Natural (PLN), sendo adotada por setores como *marketing*, política, saúde e turismo. No setor turístico, sua aplicação vem sendo testada com êxito na interpretação de grandes volumes de avaliações, proporcionando insights valiosos para a gestão de destinos (Ricardo et al., 2022).

A análise de sentimentos aplicada aos comentários do TripAdvisor tem se mostrado uma ferramenta eficaz para captar percepções sobre a experiência dos visitantes (Vetitnev et al., 2021). Além disso, essa abordagem permite mapear atributos mais elogiados ou criticados com base nas avaliações textuais dos turistas (Ramos & George, 2024). Comparações entre serviços de análise de sentimentos também demonstram sua utilidade para interpretar percepções subjetivas em contextos turísticos (Gao et al., 2015).

Outros estudos também demonstram como o uso combinado de inteligência artificial e mineração de texto pode identificar temas recorrentes, avaliar padrões afetivos e orientar decisões estratégicas baseadas em evidências (Aulia et al., 2025; Susanto et al., 2024).

De forma convergente, os estudos revisados reconhecem o CGU como fator central de influência sobre a imagem de destino e a intenção de visita, destacando o papel das avaliações digitais na construção coletiva da reputação turística (Liu, 2010; Gretzel et al., 2007; Pérez et al., 2022). Contudo, divergências importantes emergem quanto à confiabilidade desse conteúdo: enquanto parte da literatura enfatiza a autenticidade e espontaneidade das avaliações como vantagem sobre o *marketing* tradicional (Litvin et al., 2008), outro conjunto de estudos alerta para riscos de vieses sistemáticos, *fake reviews* e assimetrias informacionais que podem distorcer a imagem percebida dos destinos (Cambria et al., 2013; Pang & Lee, 2008). Essa tensão é especialmente relevante em destinos emergentes, nos quais o volume de avaliações ainda é relativamente baixo, amplificando o peso de comentários individuais na formação da reputação digital. Ao mesmo tempo, a articulação teórica entre CGU, análise de sentimentos e imagem de destino ainda carece de maior densidade: poucos estudos examinam simultaneamente a polaridade emocional dos comentários e suas implicações para a construção simbólica da imagem de um destino específico (Ye et al., 2011; Guo et al., 2017). É precisamente nessa articulação que o presente estudo busca contribuir.

METODOLOGIA

Este estudo adota um desenho de métodos mistos do tipo convergente, de caráter exploratório e descritivo (Molina-Azorín & Font, 2016). As dimensões quantitativa e qualitativa são conduzidas sobre o mesmo *corpus* e integradas na interpretação. A dimensão quantitativa compreende a Classificação Hierárquica Descendente (CHD), a Análise de Similitude e a Análise Fatorial de Correspondência (AFC), processadas pelo IRAMUTEQ, além da classificação automática e independente de polaridade pelo DistilBERT. A dimensão qualitativa corresponde à leitura interpretativa dos segmentos de texto e dos eixos semânticos, utilizada para nomear e atribuir sentido substantivo às classes lexicais. A integração ocorre na interpretação das classes quantitativamente extraídas e na comparação entre a estrutura temática da CHD e a distribuição de polaridade do DistilBERT, configurando triangulação metodológica (Camargo & Justo, 2013). Os dados foram coletados por *web scraping* no TripAdvisor, abrangendo comentários sobre atrativos do litoral do Piauí.

Como os critérios originais de inclusão por atrativo não estão documentados no material disponível, não se atribui retrospectivamente um limiar numérico. Para fins de transparência, o conjunto analisado é não probabilístico e compreende atrativos do litoral do Piauí presentes no TripAdvisor, distribuídos entre praias, áreas naturais e patrimônio histórico-cultural. Para cada atrativo selecionado, a coleta abrangeu a totalidade dos comentários disponíveis até 4 de fevereiro de 2023. Foram considerados os seguintes locais: a praia de Barra Grande e a Praia do Sardin, em Barra Grande; a praia de Macapá; Cajueiro da Praia; o Delta do Parnaíba; a Árvore Penteada, a Praia do Itaqui e a Praia de Atalaia, em Luís Correia; a Lagoa do Portinho, a Praia dos Poldros, a Praia Ponta das Canárias, o Porto das Barcas, o Espaço Cultural Porto das Barcas, o Museu do Mar do Delta do Parnaíba, a Catedral Nossa Senhora das Graças e o Sesc Caixeiral, em Parnaíba.

Quadro 1 – Delineamento metodológico

Etapa	Descrição	Ação	Ferramentas/Referências
Definição do corpus	Seleção dos atrativos turísticos do litoral do Piauí a serem analisados (praias, patrimônio histórico, áreas naturais).	Delimitar o universo da pesquisa.	Planejamento da coleta de dados; revisão de literatura sobre o litoral do Piauí.
Coleta de dados	Extração automatizada de comentários publicados no TripAdvisor	Obter dados autênticos sobre as percepções dos turistas.	Técnica de <i>Web Scraping</i> (Glez-Peña et al., 2013).
Tratamento dos dados	Organização em planilha Excel, remoção de registros duplicados e exclusão de textos com menos de 20 caracteres.	Garantir a consistência e a qualidade do <i>corpus</i> textual.	Excel (Microsoft Office).
Análise lexical	Processamento dos textos para identificar frequência de termos e coocorrências semânticas.	Explorar padrões de linguagem e temas recorrentes.	IRAMUTEQ (Lahlou, 1994; Camargo & Justo, 2013).
Classificação Hierárquica Descendente (CHD)	Agrupamento dos comentários em classes lexicais com base na proximidade das palavras.	Identificar categorias temáticas predominantes.	IRAMUTEQ (Marchand & Ratinaud, 2012).
Análise de Similitude	Construção da rede de coocorrência entre palavras.	Mapear conexões semânticas e estruturas discursivas.	IRAMUTEQ.
Análise Fatorial de Correspondência (AFC)	Cruzamento do vocabulário com as classes lexicais.	Visualizar proximidades e oposições entre percepções.	IRAMUTEQ.
Análises complementares	Estatísticas descritivas e geração de gráficos.	Aprofundar a interpretação quantitativa dos dados.	Software R (R Core Team, 2022).
Análise de sentimentos por modelo de linguagem	Aplicação do DistilBERT multilíngue ao comentário integral para classificação independente de polaridade (positivo/neutro/negativo), seguida de validação manual por dois pesquisadores.	Obter a polaridade automatizada, confrontá-la com a leitura humana e triangular seus resultados com a estrutura temática da CHD, sem derivar uma técnica da outra.	HuggingFace; DistilBERT (Sanh et al., 2019; Yuan, 2023); validação manual por dois pesquisadores.
Interpretação dos resultados	Discussão dos achados à luz da literatura sobre turismo, CGU e imagem de destino.	Relacionar resultados empíricos com teoria.	Estudos de Silva e Rocha (2014), Braga e Guzzi (2021), Braga et al. (2022), Camargo e Justo (2013), Coalsante et al., (2022), Silva e Melo (2012).

Fonte: autoria própria (2025).

A análise foi conduzida por meio de técnicas de análise textual e mineração de dados. A coleta seguiu a estrutura apresentada no Quadro 1 e utilizou *web scraping* na plataforma TripAdvisor, conforme o protocolo de extração sistemática descrito por Glez-Peña et al. (2013), com acesso ao site, extração do conteúdo e organização em banco textual (Borges et al., 2024; Perinotto et al., 2024). O procedimento foi realizado em 20 de fevereiro de 2023 e contemplou comentários publicados entre 8 de dezembro de 2011 e 4 de fevereiro de 2023. A documentação metodológica preservada não especifica a linguagem ou biblioteca utilizada na raspagem, razão pela qual esse parâmetro não é reconstruído retrospectivamente. Essa ausência limita a replicabilidade computacional estrita da etapa de extração; contudo, permanecem documentados a plataforma de origem, o intervalo temporal, a abrangência da coleta, os critérios de tratamento, o tamanho final do corpus e o fluxo analítico subsequente. Como o *corpus* agrega avaliações de aproximadamente doze anos, os resultados representam uma imagem digital cumulativa e sujeita à heterogeneidade temporal, decorrente de mudanças na oferta turística, nas práticas de uso da plataforma e do período da pandemia de covid-19. A dimensão temporal pode alterar de forma assimétrica o conteúdo e a valência de avaliações *online* (Xu et al., 2024); portanto, os achados não devem ser interpretados como retrato homogêneo de um único momento.

No que se refere à composição do *corpus*, foram coletados todos os comentários disponíveis nos 16 atrativos selecionados até 4 de fevereiro de 2023. O tratamento documentado incluiu organização em planilha Excel, remoção de registros duplicados e exclusão de textos com menos de 20 caracteres. O *corpus* final reuniu 2.400 comentários. Após o processamento no IRAMUTEQ, obtiveram-se 3.093 segmentos de texto (ST). A CHD, fundamentada no método de Reinert (1990), identifica classes lexicais temáticas a partir da coocorrência e da proximidade estatística entre palavras, sem inferir automaticamente a valência emocional do conteúdo (Camargo & Justo, 2013; Marchand & Ratinaud, 2012). A nomeação e a caracterização das classes resultam da interpretação qualitativa do vocabulário dominante. A polaridade positiva, neutra ou negativa, por sua vez, foi estimada de forma independente pelo DistilBERT, aplicado aos comentários integrais e não aos ST. Assim, CHD e DistilBERT constituem procedimentos distintos e complementares.

Na análise qualitativa, adotou-se uma abordagem baseada na análise lexical e na técnica de clusters. A análise lexical busca superar a tradicional separação entre abordagens quantitativas e qualitativas, pois permite quantificar e aplicar cálculos estatísticos a informações que, por natureza, são qualitativas, como os textos (Lahlou, 1994). Já a técnica de *clusters* visa a simplificar um conjunto de dados, concentrando as informações em seu ponto central (Linde, 2009). A análise dos comentários foi conduzida com auxílio do *software* IRAMUTEQ, que permitiu a identificação de padrões linguísticos e

a estruturação das relações semânticas entre os termos. Além disso, o software R foi utilizado para análises estatísticas complementares e para geração de gráficos.

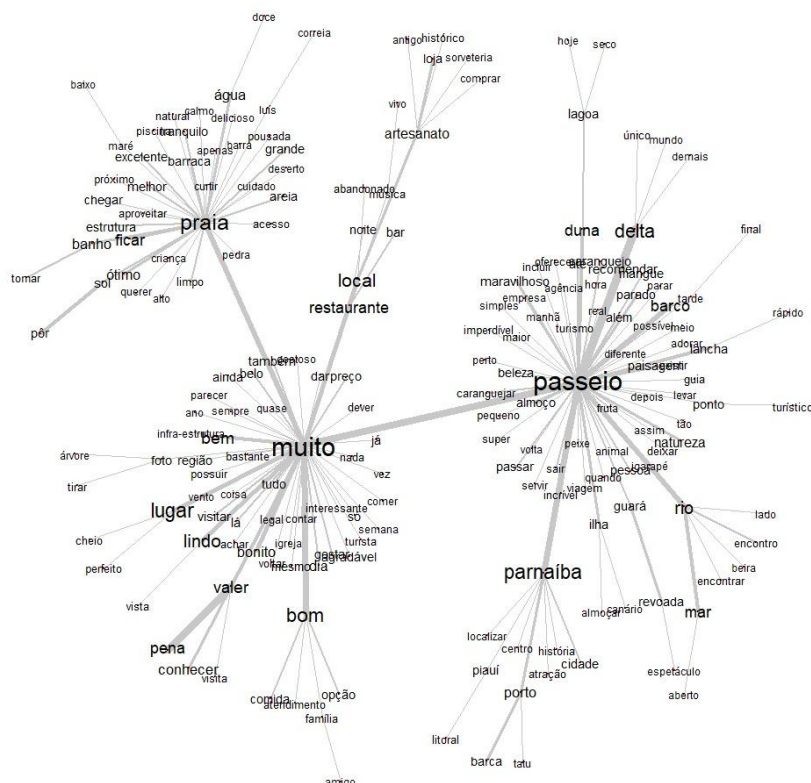
A análise de similaridade foi escolhida como principal técnica de investigação devido à sua capacidade de mapear as co-ocorrências entre palavras, revelando suas conexões dentro de um texto, ajudando a compreender a estrutura do *corpus*, destacando tanto os elementos comuns quanto as particularidades associadas às variáveis descritivas identificadas na análise (Marchand & Ratinaud, 2012).

Adicionalmente, a polaridade dos sentimentos foi verificada com o modelo distilbert-base-multilingual-cased-sentiments-student (Yuan, 2023), disponibilizado na plataforma HuggingFace e baseado na arquitetura DistilBERT (Sanh et al., 2019). Conforme a documentação do modelo, seu treinamento por destilação utiliza um modelo professor mDeBERTa-v3-base-mnli-xnli e um *template* de hipótese aplicado a um *corpus* multilíngue de sentimentos. O modelo retorna probabilidades para três rótulos mutuamente exclusivos, *positive*, *neutral* e *negative*. A unidade de análise foi o comentário individual em sua forma integral, mantendo essa etapa independente da segmentação realizada pelo IRAMUTEQ. Nos registros cuja saída classificatória foi preservada, adotou-se a categoria de maior probabilidade como classificação final. A classificação automatizada foi submetida à validação manual por dois pesquisadores, que confrontaram os rótulos atribuídos pelo modelo com a leitura humana dos comentários. A documentação metodológica remanescente, entretanto, não preserva o tamanho da subamostra utilizada nessa validação nem um coeficiente formal de concordância entre avaliadores, razão pela qual esses parâmetros não são reconstruídos retrospectivamente. Além disso, a documentação disponível não permite recuperar a atribuição original de 9,0% do *corpus*. A comparação com a CHD foi utilizada como triangulação entre uma estrutura temática e uma medida independente de polaridade, e não como validação de uma técnica pela outra.

RESULTADOS E DISCUSSÕES

A Figura 1 apresenta a árvore de similitude gerada pelo software IRAMUTEQ. Esta evidencia os principais termos extraídos do corpus textual e suas conexões semânticas. Essa técnica permite identificar as palavras mais frequentes e os agrupamentos lexicais que estruturam o discurso dos turistas em comentários publicados no TripAdvisor sobre os atrativos do litoral piauiense. Por meio dessa análise, observa-se uma narrativa turística que valoriza atributos específicos da região e reforça sua imagem enquanto destino natural e cultural.

Figura 1 – Árvore de similitude dos comentários sobre os atrativos turísticos do litoral piauiense no TripAdvisor



Fonte: dados do TripAdvisor, processados por meio do software IRAMUTEQ.

O termo “passeio” ocupa posição central na estrutura lexical, sendo amplamente conectado a palavras como “delta”, “barco”, “rio”, “paisagem”, “duna” e “natureza”. Essa configuração indica que o passeio de barco no Delta do Parnaíba representa uma das experiências mais marcantes para os visitantes, destacando a valorização de elementos do ecoturismo e da contemplação da paisagem natural. Tal percepção está em consonância com Silva e Rocha (2014), que ressaltam a integração do Delta à Rota das Emoções e seu potencial para o turismo sustentável e de aventura.

O termo “praia” também se destaca com forte densidade relacional, associado a adjetivos como “ótimo”, “limpo”, “tranquilo” e “lindo”, o que reforça o caráter estético e recreativo do litoral piauiense. Palavras como “água”, “areia”, “banho” e “ficar” ampliam essa dimensão experiencial, revelando o grau de satisfação dos turistas com os ambientes costeiros.

Além disso, o termo “Parnaíba” aparece conectado a palavras como “porto”, “cidade”, “centro histórico” e “atração”, indicando que a dimensão urbana e cultural da cidade também compõe o imaginário turístico da região. Segundo Braga e Guzzi (2021) e Braga et al. (2022), o Porto das Barcas, situado no centro histórico de Parnaíba, constitui um ponto turístico relevante, com patrimônio arquitetônico dos períodos colonial e imperial que atrai o interesse de visitantes.

Outros agrupamentos significativos envolvem termos como “lugar”, “lindo”, “valer”, “bom”, e “muito”, que funcionam como indicadores da avaliação subjetiva dos turistas, reforçando uma percepção majoritariamente positiva. Por outro lado, temas como infraestrutura, restaurantes e acesso aparecem em menor densidade, o que sugere que, embora mencionados, não centralizam as experiências relatadas.

Dessa forma, a análise da árvore de similitude revela que os comentários enfatizam, sobretudo, os atributos naturais, paisagísticos e emocionais do litoral piauiense, o que contribui para a consolidação de uma imagem turística ancorada na beleza cênica, na tranquilidade das praias e nas experiências de contato com a natureza.

Figura 2 – Nuvem de palavras com os termos mais frequentes nos comentários sobre os atrativos turísticos do litoral piauiense no TripAdvisor



Fonte: dados do TripAdvisor, processados por meio do software IRAMUTEQ.

A Figura 2 apresenta a nuvem de palavras gerada com base nos comentários dos turistas sobre o litoral piauiense. Essa visualização, produzida pelo software IRAMUTEQ, destaca graficamente os termos mais frequentes do corpus textual, sendo o tamanho das palavras proporcional à sua ocorrência. A ferramenta permite uma análise visual direta e acessível, facilitando a identificação dos temas mais recorrentes e das impressões predominantes nos relatos dos visitantes (Carmargo & Justo, 2013).

A visualização reforça a importância de termos como “passeio”, “praia”, “muito”, “Parnaíba”, “local” e “bom”, que novamente aparecem em destaque, corroborando os achados da árvore de similitude. Esses termos indicam que as experiências turísticas relatadas estão centradas em atividades de lazer, contemplação e apreciação dos elementos naturais da região.

Quanto ao campo emocional, observa-se um vocabulário amplamente positivo, com alta incidência de adjetivos como “lindo”, “bonito”, “maravilhoso”, “ótimo” e “excelente”, o que revela uma percepção geral de encantamento e satisfação com o destino. A baixa frequência de termos com carga negativa sugere que os comentários disponíveis no TripAdvisor reforçam uma imagem favorável do litoral do Piauí, destacando seus aspectos estéticos, naturais e afetivos.

Nesse sentido, a nuvem de palavras funciona como um retrato sintético e espontâneo da experiência turística, evidenciando os elementos mais valorizados pelos visitantes e contribuindo para a compreensão do conteúdo gerado pelo usuário (CGU) como fonte legítima na construção da imagem do destino.

Por meio do método de Reinert (Reinert, 1990), a Classificação Hierárquica Descendente (CHD) produziu originalmente cinco classes lexicais, organizadas segundo coocorrência e proximidade estatística entre palavras. Para fins de síntese interpretativa, duas dessas classes foram posteriormente apresentadas em uma mesma categoria temática, resultando em quatro categorias analíticas discutidas no texto. Essa consolidação não altera a solução estatística original do IRAMUTEQ, preservada no dendrograma da Figura 3.

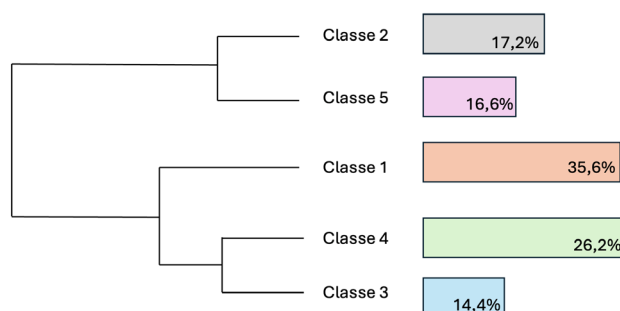
A primeira categoria temática está associada às atividades náuticas e de ecoturismo, reunindo termos como passeio, delta, rio, barco, lancha, mar, duna, catamarã, mangue, revoada e caranguejo. Esses elementos apontam para experiências no Delta do Parnaíba, com destaque para passeios de barco e contato com a fauna local, como a observação dos guarás. O vocabulário de valência predominantemente positiva é compatível com avaliações favoráveis das paisagens e atividades ligadas à natureza. Segundo Colasante et al. (2022), esses elementos reforçam o potencial do ecoturismo na região e contribuem para a valorização do destino.

A segunda classe está relacionada ao lazer e descanso em praias e lagoas, com destaque para palavras como praia, lagoa, barraca, árvore, relaxar, tranquilidade, natureza, pôr do sol, maravilhoso e delicioso. As menções indicam a valorização do ambiente natural e da infraestrutura de apoio, como barracas e restaurantes, além da busca por momentos de tranquilidade. O vocabulário fortemente positivo reflete a apreciação dos visitantes por esse tipo de experiência contemplativa.

A terceira classe remete ao turismo cultural e ao patrimônio histórico, incluindo termos como artesanato, loja, histórico, porto, cidade, igreja, centro, bar e sorveteria. Essa classe representa percepções sobre o centro histórico de Parnaíba, combinando atrações culturais, comércio local e vida noturna. Embora sejam feitas menções elogiosas ao patrimônio arquitetônico e à produção artesanal, também surgem apontamentos quanto à necessidade de melhorias na conservação e infraestrutura desses espaços. Como observam Braga e Guzzi (2021), o Porto das Barcas e o centro histórico configuram-se como um marco histórico e cultural relevante, reunindo edificações do período colonial e contribuindo para o valor simbólico do destino (Braga et al., 2022).

Por fim, a quarta classe agrega os relatos mais críticos, com destaque para termos como abandono, falta, precisar, devido, estrada, limpeza, cuidado e interditar. Embora seja menos frequente no *corpus* analisado, essa classe concentra avaliações negativas, geralmente associadas à infraestrutura precária, falta de manutenção e acessibilidade limitada. Do ponto de vista analítico, esses termos não se limitam a descrever deficiências físicas: eles sinalizam uma ruptura entre a expectativa construída pelo *marketing* do destino e a experiência vivida *in loco*, um hiato expectativa-experiência reconhecido como um dos principais vetores de formação de imagem negativa e de redução da intenção de revisita (Freire et al., 2022; Beerli & Martín, 2004). Em destinos emergentes como o litoral piauiense, onde o volume total de avaliações ainda é relativamente baixo, comentários negativos tendem a exercer influência desproporcional sobre a reputação digital, fenômeno que amplifica a urgência das intervenções infraestruturais identificadas e que deve orientar prioritariamente a agenda dos gestores públicos e privados do turismo local. Esses relatos funcionam, portanto, como alertas estratégicos que, se incorporados ao planejamento turístico, podem reverter fragilidades estruturais em oportunidades de diferenciação competitiva do destino.

Esses agrupamentos temáticos oferecem uma visão sobre a forma como os turistas percebem e avaliam o destino, e constituem subsídios valiosos para o aprimoramento das estratégias de gestão turística local.

Figura 3 – Dendrograma da Classificação Hierárquica Descendente (CHD) dos comentários sobre os atrativos turísticos do litoral piauiense no TripAdvisor

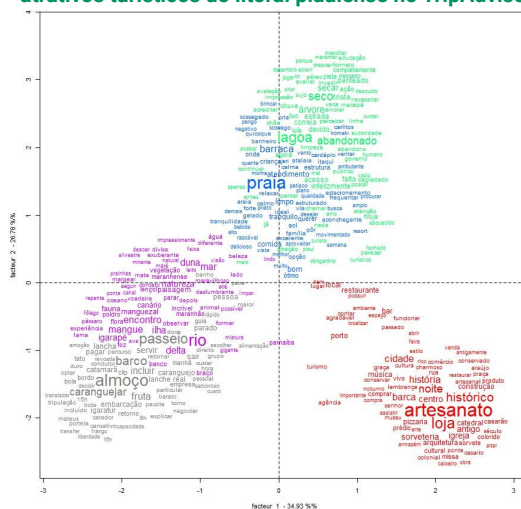
Fonte: Dados do TripAdvisor, processados por meio do software IRAMUTEQ.

O processamento pelo IRAMUTEQ resultou originalmente em cinco classes lexicais, conforme o dendrograma da Figura 3. Na etapa interpretativa, as Classes 2 e 5 foram apresentadas conjuntamente sob a categoria temática Natureza e Ecoturismo por compartilharem vocabulário e eixos semânticos convergentes (Camargo & Justo, 2013). Trata-se de uma consolidação temática posterior, e não de uma fusão estatística das classes: os percentuais e a estrutura das cinco classes originais permanecem preservados na Figura 3. Essa distinção explica a presença de cinco classes no dendrograma e de quatro categorias temáticas no Quadro 2 sem atribuir à consolidação um critério estatístico que não foi aplicado.

A Análise Fatorial de Correspondência (AFC) ofereceu uma representação complementar da organização lexical do corpus. O Fator 1 explicou 36,25% da inércia total e o Fator 2 explicou 26,57%, totalizando 62,82% de inércia acumulada, conforme o output do IRAMUTEQ. Os dois primeiros fatores evidenciam a oposição entre vocabulário de valência predominantemente positiva, associado à natureza, ao lazer e à hospitalidade, e um conjunto menor de referências críticas ligadas à infraestrutura e aos serviços. Essa configuração indica que a imagem discursiva do litoral piauiense é majoritariamente favorável, embora coexistam fragilidades estruturais que demandam atenção da gestão pública e privada (Silva & Melo, 2012).

De forma geral, as categorias temáticas 1 e 2 apresentam vocabulário avaliativo predominantemente positivo, reforçando a imagem favorável do litoral piauiense. A categoria cultural evidencia que a experiência turística também incorpora atrativos urbanos e históricos, enquanto a categoria crítica concentra referências à manutenção e infraestrutura. Essa leitura qualitativa das classes não substitui a classificação automatizada de polaridade realizada de forma independente pelo DistilBERT. O potencial turístico relaciona-se, portanto, à articulação entre infraestrutura turística, infraestrutura básica, serviços e atrativos oferecidos (Silva & Melo, 2012).

Na Figura 4, revelam-se quatro eixos centrais das experiências turísticas no litoral piauiense: (a) Atividades náuticas e ecoturismo – centradas no Delta do Parnaíba e associadas a paisagens naturais e fauna local; (b) Lazer, praias e hospitalidade – destacando o descanso, a tranquilidade e a infraestrutura de apoio à beira-mar; (c) Turismo cultural e patrimônio histórico – voltado ao centro histórico e à economia criativa da cidade de Parnaíba; e (d) Críticas à infraestrutura – indicando fragilidades estruturais percebidas pelos turistas durante a visita.

Figura 4 – Análise Fatorial de Correspondência (AFC) com destaque das classes lexicais identificadas nos comentários sobre os atrativos turísticos do litoral piauiense no TripAdvisor

Fonte: dados do TripAdvisor, processados por meio do software IRAMUTEQ

Como análise complementar, o modelo multilíngue DistilBERT (Sanh et al., 2019; Yuan, 2023) foi aplicado ao corpus de 2.400 comentários. As proporções registradas nos resultados preservados correspondem a 42,8% de sentimentos positivos, 33,8% neutros e 14,4% negativos, totalizando 91,0% do corpus. Para os 9,0% restantes, aproximadamente 216 comentários, a atribuição original de polaridade não pôde ser recuperada de modo verificável. Essa parcela é, portanto, reportada como classificação não recuperável nos registros preservados e não como uma quarta classe produzida pelo DistilBERT. A predominância de polaridade positiva converge, em termos gerais, com o vocabulário avaliativo favorável observado nas categorias temáticas de natureza e lazer, sem implicar relação de derivação entre CHD e DistilBERT.

O Quadro 2 evidencia que o litoral do Piauí é valorizado, sobretudo, por seus recursos naturais e pela autenticidade cultural, enquanto críticas à infraestrutura urbana e à manutenção de atrativos constituem um contraponto à imagem favorável. De forma integrada, os achados indicam uma imagem digital agregada globalmente positiva, porém heterogênea, construída ao longo do período analisado por avaliações que combinam belezas naturais, experiências de lazer, patrimônio cultural e fragilidades estruturais. Essa dualidade deve orientar gestores públicos e privados na formulação de estratégias de desenvolvimento turístico mais sustentáveis e competitivas.

Quadro 2 – Síntese das análises por classes temáticas

Classe Temática	Principais termos associados	Percepções predominantes	Valência qualitativa do vocabulário	Implicações para o turismo
Natureza e Ecoturismo	Passeio, delta, rio, barco, paisagem, natureza, dunas, mangue, guarás, caranguejo.	Forte valorização das belezas naturais e atividades ligadas ao ecoturismo, como passeios de barco e observação da fauna.	Vocabulário de valência predominantemente positiva, associado à admiração, ao encantamento e à satisfação.	Reforça a imagem do litoral piauiense como destino de turismo de natureza e ecoturismo, demandando políticas de preservação ambiental.
Lazer e Hospitalidade	Praia, lagoa, barraca, árvore, pôr do sol, relaxar, tranquilidade, maravilhoso, delicioso.	Experiências de lazer, descanso e hospitalidade; destaque para infraestrutura de apoio (barracas, restaurantes).	Vocabulário de valência predominantemente positiva, associado a prazer, conforto e relaxamento.	Necessidade de qualificação contínua da hospitalidade e dos serviços de apoio para manter altos níveis de satisfação.
Cultura e Patrimônio Histórico	Artesanato, loja, histórico, porto, cidade, igreja, noite, centro, bar, gastronomia.	Valorização do patrimônio histórico e da cultura local, especialmente em Parnaíba e no Porto das Barcas.	Vocabulário majoritariamente positivo, com ocorrências de valência negativa relacionadas à conservação.	Potencial fortalecimento do turismo cultural e da economia criativa; demanda por políticas de preservação e revitalização urbana.
Infraestrutura e Serviços	Abandono, falta, estrada, limpeza, cuidado, interditar.	Críticas pontuais à manutenção, limpeza e acessibilidade em alguns atrativos.	Vocabulário de valência predominantemente negativa, associado a frustração e insatisfação.	Necessidade de investimentos em infraestrutura básica e turística para reduzir pontos de insatisfação e elevar a competitividade do destino.

Fonte: autoria própria (2025).

Os resultados demonstram que o CGU permite compreender simultaneamente atributos valorizados e fragilidades percebidas na construção discursiva da imagem do litoral piauiense. A centralidade dos recursos naturais e das experiências recreativas converge com a literatura sobre fatores de imagem de destino (Beerli & Martín, 2004; Freire et al., 2022), mas a distribuição temática observada não é equilibrada. Diferentemente de destinos mais consolidados, nos quais as avaliações tendem a distribuir maior atenção entre natureza, infraestrutura e serviços (Vetitnev et al., 2021; Ramos & George, 2024), o corpus analisado concentra-se de forma marcante nos atributos naturais e experienciais, enquanto a infraestrutura aparece sobretudo como contraponto crítico.

Esse padrão sugere que, em um destino emergente e de menor visibilidade, a atratividade paisagística pode assumir peso discursivo superior ao dos atributos estruturais, compensando parcialmente fragilidades de infraestrutura na avaliação global. Ao mesmo tempo, as críticas à manutenção, acesso e serviços mostram que essa compensação não elimina limitações capazes de afetar a reputação digital. O resultado particulariza a natureza multidimensional da imagem de destino proposta por Beerli e Martín (2004) e Gallarza et al. (2002), indicando que o peso relativo de cada dimensão varia conforme o contexto de desenvolvimento turístico. A distinção entre imagem percebida e projetada também reforça a necessidade de interpretar o CGU como expressão situada da experiência do visitante, e não como representação total da imagem do destino (Sun et al., 2021).

CONSIDERAÇÕES FINAIS

Retomando explicitamente a questão de pesquisa que orientou este estudo – como os sentimentos expressos pelos turistas no TripAdvisor contribuem para a construção da imagem do litoral do Piauí enquanto destino emergente? – os resultados indicam que esses sentimentos operam simultaneamente como reforço e alerta. A predominância de polaridade positiva, associada aos recursos naturais, às experiências de lazer e ao patrimônio histórico-cultural, sustenta uma imagem digital favorável e diferenciada. Em sentido complementar, avaliações negativas relacionadas à infraestrutura, manutenção e acessibilidade sinalizam fragilidades que tensionam essa imagem. Assim, o CGU revela uma imagem agregada favorável, porém heterogênea, na qual atributos experienciais coexistem com limitações estruturais.

No âmbito prático, os resultados oferecem subsídios diferenciados aos atores da governança turística local. Para gestores públicos municipais e estaduais, as críticas apontam prioridades em manutenção viária, sinalização, limpeza e acessibilidade. Para organizações do trade, guias e operadoras, a forte valorização dos atributos naturais e experienciais sustenta estratégias de comunicação digital centradas no Delta do Parnaíba, nas praias e nas paisagens costeiras. Para gestores do patrimônio histórico-cultural de Parnaíba, os comentários reforçam a necessidade de conservação e revitalização dos espaços valorizados pelos visitantes. Para pesquisadores e demais atores do ecossistema turístico regional, o protocolo combinado oferece uma base para monitoramento periódico da reputação digital do destino.

Do ponto de vista metodológico e empírico, a pesquisa contribui ao demonstrar a aplicação combinada de mineração lexicométrica, por meio de CHD, Similitude e AFC, e classificação automatizada de sentimentos por DistilBERT na caracterização da imagem digital de um destino emergente ainda pouco estudado. A contribuição teórica é mais delimitada: os achados evidenciam, nesse contexto, um padrão em que a valorização de atributos naturais e experienciais coexiste com fragilidades estruturais, refinando empiricamente a proposição de Beerli e Martín (2004) e Gallarza et al. (2002) sobre a natureza multidimensional e heterogênea da imagem de destino.

Em termos sociais, o estudo destaca a importância da escuta ativa das percepções dos turistas como mecanismo de fortalecimento da competitividade regional, além de evidenciar como o turismo pode se consolidar como vetor de desenvolvimento sustentável, capaz de gerar benefícios econômicos, culturais e ambientais para as comunidades locais.

Entretanto, esta pesquisa apresenta limitações. O corpus reúne apenas comentários do TripAdvisor e agrega avaliações publicadas entre 2011 e 2023, o que introduz heterogeneidade temporal e impede interpretar os resultados como retrato de um único momento. A classificação automatizada está sujeita a ambiguidades linguísticas, ironia e especificidades culturais. Embora tenha sido submetida à validação manual por dois pesquisadores, a documentação preservada não permite recuperar o tamanho da subamostra validada nem um coeficiente formal de concordância entre avaliadores. Além disso, para 9,0% dos comentários, aproximadamente 216 registros, a atribuição original de polaridade não pôde ser recuperada na documentação preservada. Por fim, a documentação metodológica disponível não registra a linguagem ou biblioteca utilizada no web scraping, o que limita a replicabilidade computacional estrita da etapa de coleta, ainda que estejam documentados o universo analisado, o intervalo temporal, os critérios de tratamento e o fluxo analítico.

Diante dessas limitações, pesquisas futuras podem ampliar a coleta para Google Reviews, Booking e redes sociais, segmentar o corpus por subperíodos, inclusive pré e pós-pandemia, e replicar a validação manual com amostra estratificada, registro do tamanho amostral e coeficientes formais de concordância entre avaliadores e entre avaliação humana e classificação automatizada. Também se recomenda que futuras coletas registrem e, quando possível, disponibilizem a linguagem, a biblioteca e o script empregados na web scraping, fortalecendo a replicabilidade computacional. Estudos comparativos com destinos consolidados também permitiriam testar se a predominância discursiva dos atributos naturais se mantém em diferentes estágios de desenvolvimento turístico. Do ponto de vista metodológico, modelos treinados especificamente para o português e para o domínio turístico podem ser comparados ao DistilBERT. A relação entre polaridade e indicadores de comportamento, como fluxo de visitantes, reservas ou posição em rankings, pode, ainda, ampliar o desenho de exploratório para explicativo.

AGRADECIMENTOS

Agradecimentos ao Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPQ, Brasil.

REFERÊNCIAS

- Antunes, N., Ramos, C., & Sousa, C. (2018). Who Are The Tourists Sharing Content On Social Media? Behaviour And Characteristics. *Journal of Tourism, Sustainability and Well-being*, Cinturs – Research Centre for Tourism, Sustainability and Well-being, University of Algarve, 6(3), 237-257.
- Aulia, M. A., Solihah, B., & Zuhdi, A. (2025). Sentiment analysis and topic modeling of tourist reviews on Bali Island attractions on Tripadvisor using lexicon-based method and Latent Dirichlet Allocation (LDA). *Intelmatika*, 5(1), 1-7. <https://doi.org/10.25105/v5i1.17619>
- Beerli, A., & Martín, J. D. (2004). Factors influencing destination image. *Annals of Tourism Research*, 31(3), 657-681. <https://doi.org/10.1016/j.annals.2004.01.010>
- Borges, J. L. C., Perinotto, A. R. C., & Braga, S. S. (2024). O uso de técnicas de crowdsourcing, big data e análise de redes aplicadas à demanda turística. *Marketing & Tourism Review*, 9(1), 1-39.
- Braga, S. S., & Guzzi, A. (2021). Organização espacial da atividade turística no litoral piauiense. *Mercator*, 20, 1-22. <https://doi.org/10.4215/rm2021.e20012>
- Braga, S. S., Guzzi, A., Perinotto, A. R. C., & Malta, G. A. P. (2022). Análise da atratividade turística do litoral piauiense: Atualização da avaliação dos atrativos turísticos, entre 2010 e 2020. *Revista Turismo em Análise*, 33(1), 29-49. <https://doi.org/10.11606/issn.1984-4867.v33i1p29-49>
- Camargo, B. V., & Justo, A. M. (2013). IRAMUTEQ: Um software gratuito para análise de dados textuais. *Temas em Psicologia*, 21(2), 513-518. <https://doi.org/10.9788/TP2013.2-16>
- Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, 28(2), 15-21. <https://doi.org/10.1109/MIS.2013.30>
- Chu, M., Chen, Y., Yang, L., & Wang, J. (2022). Language interpretation in travel guidance platform: Text mining and sentiment analysis of TripAdvisor reviews. *Frontiers in psychology*, 13, 1029945. <https://doi.org/10.3389/fpsyg.2022.1029945>
- Colasante, T., Ferreira, L., & Almeida, P. (2022). A revoada dos guarás como atrativo turístico: Estudo de caso no litoral do Brasil. *Revista Brasileira de Ecoturismo*, 15(3), 369-395. <https://doi.org/10.34024/rbecotur.2022.v15.13159>
- Dutra, L. F., & Gosling, M. S. (2021). Analisando sentimentos no TripAdvisor: A voz do consumidor como parâmetro decisório na medição da qualidade da informação em museus. *Informação & Sociedade*, 31, 1-23. <https://doi.org/10.22478/ufrpb.1809-4783.2021v31n1>
- Echtner, C. M., & Ritchie, J. R. B. (1993). The measurement of destination image: An empirical assessment. *Journal of Travel Research*, 31(4), 3-13. <https://doi.org/10.1177/004728759303100402>
- Fotis, J., Buhalis, D., & Rossides, N. (2012). Social media impact on holiday travel planning: The case of the Russian and the FSU markets. *International Journal of Online Marketing*, 2(1), 1-19. <https://doi.org/10.4018/ijom.2012010101>
- Freire, R. M. L., Moreira, S., & Dias, C. (2022). Análise da imagem afetiva dos turistas no destino Vale dos Vinhedos – RS. *Turismo: Visão e Ação*, 25(1), 114-133. <https://doi.org/10.14210/rtva.v25n1.p114-133>
- Gallarza, M. G., Saura, I. G., & García, H. C. (2002). Destination image: Towards a conceptual framework. *Annals of Tourism Research*, 29(1), 56-78. [https://doi.org/10.1016/S0160-7383\(01\)00031-7](https://doi.org/10.1016/S0160-7383(01)00031-7)
- Gao, S., Hao, J., & Fu, Y. (2015). The application and comparison of web services for sentiment analysis in tourism," In: 12th International Conference on Service Systems and Service Management (ICSSSM), Guangzhou, pp. 1-6. IEEE. <https://doi.org/10.1109/ICSSSM.2015.7170341>
- Glez-Peña, D., Lourenço, A., López-Fernández, H., Reboiro-Jato, M., & Fdez-Riverola, F. (2013). Web scraping technologies in an API world. *Briefings in Bioinformatics*, 15(5), 788-797. <https://doi.org/10.1093/bib/bbt026>
- Gretzel, U., Yoo, K. H., & Purifoy, M. (2007). Online travel review study: Role & impact of TripAdvisor. *Laboratory for Intelligent Systems in Tourism*, Texas A&M University.
- Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59, 467-483. <https://doi.org/10.1016/j.tourman.2016.09.009>
- Hosany, S., Martin, D., & Woodside, A. G. (2021). Emotions in tourism: Theoretical designs, measurements, analytics, and interpretations. *Journal of Travel Research*, 60(4), 773-787. <https://doi.org/10.1177/0047287520937079>
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, 53(1), 59-68. <https://doi.org/10.1016/j.bushor.2009.09.003>
- Lahlou, S. (1994). L'analyse lexicale. *Variances*, 2, 13-24.
- Lai, K., & Li, X. (2016). Tourism destination image: Conceptual problems and definitional solutions. *Journal of Travel Research*, 55(8), 1065-1080. <https://doi.org/10.1177/0047287515619693>
- Linden, R. (2009). Técnicas de agrupamento. *Revista de Sistemas de Informação da FSMA*, 4, 18-36.
- Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, 29(3), 458-468. <https://doi.org/10.1016/j.tourman.2007.05.011>
- Liu, B. (2010). Sentiment analysis and subjectivity. In: N. Indurkha & F. J. Damerau (Eds.), *Handbook of natural language processing* (pp. 627-666). CRC Press.
- Liu, B. (2012). *Sentiment analysis and opinion mining*. Morgan & Claypool.
- Marchand, P., & Ratinaud, P. (2012). L'analyse de similitude appliquée aux corpus textuels: Les primaires socialistes pour l'élection présidentielle française (septembre-octobre 2011). In *Actes des 11e Journées internationales d'Analyse statistique des Données Textuelles (JADT 2012)* (pp. 687-699). Recuperado de: <http://lexicometrica.univ-paris3.fr/jadt/jadt2012/Communications/Marchand,%20Pascal%20et%20al.%20-%20L%27analyse%20de%20similitude%20appliquee%20aux%20corpus%20textuels.pdf>

- Marrese-Taylor, E., Velásquez, J. D., Bravo-Marquez, F., & Matsuo, Y. (2013). Identifying customer preferences about tourism products using an aspect-based opinion mining approach. *Procedia Computer Science*, 22, 182-191. <https://doi.org/10.1016/j.procs.2013.09.094>
- Miguéns, J., Baggio, R., & Costa, C. (2008). Social media and tourism destinations: TripAdvisor case study. *Advances in Tourism Research*, 26, 1-6.
- Molina-Azorín, J. F., & Font, X. (2016). Mixed methods in sustainable tourism research: An analysis of prevalence, designs and application in JOST (2005–2014). *Journal of Sustainable Tourism*, 24(4), 549-573. <https://doi.org/10.1080/09669582.2015.1073739>
- Otabek, J. (2024). Establishing The Foundations Of Sustainable Tourism And The Use Of Digital Technologies To Reduce Negative Impacts On The Environment And Public Health: Global Experience. *American Journal of Economics and Business Management*, 8(1), 1-8. <https://doi.org/10.31150/ajeblm.v8i1.3184>
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135. <https://doi.org/10.1561/1500000011>
- Pérez R., Elizabeth C., Sánchez, E. B., & Labrada, J. R. M. (2022). Exploring consumer behavior through user-generated content on TripAdvisor. The case of Holguin destination. *Revista interamericana de ambiente y turismo*, 18(2), 184-194. <https://dx.doi.org/10.4067/S0718-235X2022000200184>
- Perinotto, A. R. C., Borges, J. L. C., Braga, S. S., & Cembranel, P. (2024). Caracterização de segmentos e atrativos turísticos por meio da mineração de dados e análise de redes no TripAdvisor. *Revista Brasileira de Pesquisa em Turismo*, 18, 2950. <https://doi.org/10.7784/rbtur.v18.2950>
- R Core Team. (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Ramos, C. M. Q., & George, O. A. (2024). Sentiment analysis applied to tourism: exploring tourist-generated content in the case of a wellness tourism destination. *International Journal of Spa and Wellness*, 7(2), 139-161. <https://doi.org/10.1080/24721735.2024.2352979>
- Reinert, M. (1990). Alceste, une méthodologie d'analyse des données textuelles et une application: Aurelia de Gérard de Nerval. *Bulletin de Méthodologie Sociologique*, 26(1), 24-54. <https://doi.org/10.1177/075910639002600103>
- Ricardo, E. C. P., Sanchez, E. B., & Labrada, J. R. M. (2022). Exploring consumer behavior through user-generated content on TripAdvisor: The case of Holguin destination. *Revista Interamericana de Ambiente y Turismo*, 18(2), 184-194. <https://doi.org/10.4067/S0718-235X2022000200184>
- Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*. <https://arxiv.org/abs/1910.01108>
- Santos, S., Ferreira, S., & Vasconcelos, M. (2024). User-generated content in tourism: Could it impact brand equity and intention to visit? In: *Proceedings of the 11th European Conference on Social Media*. Recuperado de: <https://pdfs.semanticscholar.org/111d/4e752d6c4d3b9a5124c0df7eff1f5d328d6e.pdf>
- Sigala, M., Christou, E., & Gretzel, U. (2012). Social media in travel, tourism and hospitality: Theory, practice and cases. Ashgate.
- Silva, A. M. S., & Rocha, R. R. N. (2014). Uma reflexão sobre turismo e desenvolvimento sustentável na APA do Delta do Parnaíba/PI – Rota das Emoções. *Turismo: Estudos & Práticas*, 3(1), 79-101.
- Silva, F. G. S. S., & Melo, R. S. (2012). A contribuição da sinalização turística para o desenvolvimento turístico da cidade de Parnaíba (PI, Brasil). *Revista Brasileira de Pesquisa em Turismo*, 6(2), 129-146. <https://doi.org/10.7784/rbtur.v6i2.488>
- Sivarethnamohan, R. (2023). Exploring the Transformation of Digital Tourism: Trends, Impacts, and Future Prospects. In: *International Conference on Digital Applications, Transformation & Economy (ICDATE)*, Miri, Sarawak, Malaysia, pp. 260-266. <https://doi.org/10.1109/ICDATE58146.2023.10248691>.
- Sun, W., Tang, S., & Liu, F. (2021). Examining perceived and projected destination image: A social media content analysis. *Sustainability*, 13(6), 3354. <https://doi.org/10.3390/su13063354>
- Susanto, P. C., Gunawan, P. W., & Widarnandana, I. G. D. (2024). Artificial intelligence assisted insights into Bali's destination image: Sentiment and thematic analyses of TripAdvisor reviews. *International Journal of Advances in Applied Sciences*, 13(1), 63-72. <https://doi.org/10.11591/ijaas.v13.i1.pp63-72>
- Vetitnev, A., Chigarev, D., & Enushevskaya, V. (2021). Sentiment analysis as a tool for assessing the negative impact of tourism on a destination. In: *International Conference Ecological Paradigms of Sustainable Development: Political, Economic and Technological Dimension of Biosphere Problems" (EPSD 2021)*, Saint Petersburg, Russia, Edited by Ereemeev, S.G.; E3S Web of Conferences, Volume 311. Recuperado de: <https://ui.adsabs.harvard.edu/abs/2021E3SWC.31106008V/abstract>
- Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism Management*, 31(2), 179-188. <https://doi.org/10.1016/j.tourman.2009.02.016>
- Xu, X., Wang, Y., Zhu, Q., & Zhuang, Y. (2024). Time matters: Investigating the asymmetric reflection of online reviews on customer satisfaction and recommendation across temporal lenses. *International Journal of Information Management*, 75, 102733. <https://doi.org/10.1016/j.ijinfomgt.2023.102733>
- Yang, S., Isa, S. M., Yao, Y., Xia, J., & Liu, D. (2022). Cognitive image, affective image, cultural dimensions, and conative image: A new conceptual framework. *Frontiers in Psychology*, 13, 935814. <https://doi.org/10.3389/fpsyg.2022.935814>
- Ye, Q., Law, R., Gu, B., & Chen, W. (2011). The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings. *Computers in Human Behavior*, 27(2), 634-639. <https://doi.org/10.1016/j.chb.2010.04.014>
- Yuan, L.-X. (2023). distilbert-base-multilingual-cased-sentiments-student [Modelo de aprendizado de máquina]. HuggingFace. Recuperado de: <https://huggingface.co/lxyuan/distilbert-base-multilingual-cased-sentiments-student>

CONTRIBUIÇÃO DOS AUTORES

André Riani Costa Perinotto: Conceitualização; Curadoria de dados; Análise de dados; Recebimento de financiamento; Pesquisa; Metodologia; Administração do projeto; Disponibilização de ferramentas; Desenvolvimento, implementação e teste de software; Supervisão; Validação de dados e experimentos; Design da apresentação de dados; Redação do manuscrito original; Redação – revisão e edição.

Maria Clara Silva da Rocha: Conceitualização; Curadoria de dados; Análise de dados; Recebimento de financiamento; Pesquisa; Metodologia; Disponibilização de ferramentas; Desenvolvimento, implementação e teste de software; Validação de dados e experimentos; Design da apresentação de dados; Redação do manuscrito original.

Priscila Cembranel: Curadoria de dados; Análise de dados; Disponibilização de ferramentas; Validação de dados e experimentos; Design da apresentação de dados; Redação do manuscrito original; Redação – revisão e edição.