What Do Guests Really Think? A Multi-Scenario ABSA Framework in Brazilian Amazon Tourism

Victor I. S. Silveira [1], **Cecília A. Silva** [3], **Omar A. C. Cortés** [1,2], **Antonio F. L. J. Júnior** [1], **Fábio M. F. Lobato** [1,3,4]

[1]Universidade Estadual do Maranhão (UEMA)
São Luís – MA – Brasil

[2]Instituto Federal do Maranhão (IFMA)
São Luís – MA – Brasil

[3]Universidade Federal do Oeste do Pará (UFOPA)
Santarém – PA – Brasil

[4]Universidade de São Paulo (USP)
São Carlos – SP – Brasil

omar@ifma.edu.br, fabio.lobato@icmc.usp.br

Abstract. *Tourism is a significant socioeconomic activity, contributing $11.6 trillion to global GDP in 2025, and online platforms such as Booking have become rich sources of user-generated content that reflect tourists' perceptions on a large scale; however, a significant gap remains in studies focused on tourism in the Brazilian Amazon. Thus, this study collects, processes, and analyzes Booking reviews from three destinations of varying sizes, namely Belém, Santarém, and Alter do Chão (Pará), combining exploratory analysis with Aspect-Based Sentiment Analysis in Portuguese. Approximately 103,000 reviews were analyzed using three variations of BERTimbau-Large fine-tuned on the ABSAPT-2024 corpus. The results show a predominance of positive sentiment, seasonal peaks in the second half of the year, and contrasts among the cities in their most prominent aspects, providing insights for accommodation managers and public tourism policies in the Amazon.*

Resumo. *O turismo é uma atividade socioeconômica relevante, contribuindo com 11,6 trilhões de dólares para o PIB global em 2025, e plataformas online como o Booking tornaram-se ricas fontes de conteúdo gerado pelo usuário, refletindo a percepção de turistas em larga escala. Contudo, há uma lacuna significativa de estudos sobre o turismo na Amazônia brasileira. Dessa forma, este trabalho realiza a coleta, o processamento e a análise de avaliações do Booking de três destinos de dimensões distintas, Belém, Santarém e Alter do Chão (Pará), combinando análise exploratória e análise de sentimentos baseada em aspectos em português. Foram analisadas cerca de 103 mil avaliações, com três variações do BERTimbau-Large ajustadas ao corpus ABSAPT-2024. Os resultados evidenciam predominância de sentimento positivo, picos sazonais no segundo semestre e contrastes entre as cidades nos aspectos mais notáveis, fornecendo subsídios a gestores de hospedagem e a políticas públicas de turismo na Amazônia.*

1. Introdução

O turismo é uma atividade socioeconômica relevante para a geração de renda, emprego e desenvolvimento territorial [Thommandru et al. 2023]. Em 2025, o setor de viagens e turismo representou aproximadamente 9,8% da economia mundial[1]. No Brasil, o turismo internacional apresentou crescimento expressivo, com aumento de 45% nas chegadas de visitantes estrangeiros e receitas superiores a US$ 6,6 bilhões nos primeiros nove meses de 2024. Nesse cenário, destinos amazônicos do Pará possuem relevância estratégica: Santarém e Alter do Chão destacam-se pelo ecoturismo e pelas praias fluviais do Rio Tapajós[2], enquanto Belém se destaca pelo turismo urbano e gastronômico[3]. Assim, o crescimento das hospedagens é refletido nas plataformas de avaliação, como TripAdvisor e Booking, que são fontes fundamentais de Conteúdo Gerado pelo Usuário (CGU), refletindo percepções e preferências de turistas em larga escala [Chu et al. 2022]. Além de influenciarem decisões de viagem, essas plataformas fornecem informações relevantes para estratégias de *Customer Relationship Management* (CRM) e para o monitoramento da experiência do consumidor [Buhalis and Volchek 2022], sendo necessária a utilização de técnicas de Processamento de Linguagem Natural (PLN) para mapear e caracterizar as necessidades e opiniões dos turistas [Egger and Gokce 2022].

Apesar da crescente disponibilidade de avaliações online, ainda há uma lacuna de estudos que aplicam técnicas de PLN ao turismo na Amazônia brasileira, especialmente a Análise de Sentimentos Baseada em Aspectos, do inglês *Aspect-Based Sentiment Analysis* (ABSA), em destinos paraenses analisados a partir de plataformas de hospedagem. A ABSA permite identificar não apenas a polaridade geral de uma avaliação, mas também os aspectos específicos mencionados pelos turistas, como atendimento, localização, limpeza e café da manhã [Cheng et al. 2025]. A literatura aponta que modelos de linguagem baseados em arquiteturas *Transformer*, como o *Bidirectional Encoder Representations from Transformers* (BERT), permanecem eficazes em tarefas de ABSA devido à sua capacidade de capturar dependências contextuais relevantes para a identificação da polaridade em nível de aspecto [Tripathy and Sharaff 2026].

Embora abordagens léxicas, redes recorrentes com CRF e, mais recentemente, *Large Language Models* (LLMs) também sejam empregadas em ABSA [Tripathy and Sharaff 2026], a opção pela família BERT neste trabalho justifica-se por três fatores: (i) modelos *encoder-only* são naturalmente adequados à formulação de extração de aspectos como rotulagem de *tokens* e de sentimento do aspecto como classificação de pares de sentenças, sem exigir estratégias de decodificação ou engenharia de *prompts*; (ii) seu custo computacional de inferência é substancialmente menor que o de LLMs, viabilizando a aplicação em escala às cerca de 50 mil avaliações do corpus consolidado; e (iii) a disponibilidade do BERTimbau, pré-treinado especificamente em português brasileiro [Souza et al. 2020], e sua adoção pelos trabalhos de referência do

[1]Travel & Tourism Economic Impact Research (EIR). Wttc.org. Disponível em: https://wttc.org/research/economic-impact. Acesso em: 05 de junho de 2026

[2]Ministério do Turismo. Disponível em: https://www.gov.br/turismo/pt-br/assuntos/noticias/caribe-amazonico-alter-do-chao-surpreende-com-praias-de-areia-branca-e-aguas-mornas-no-coracao-da-floresta. Acesso em: 05 jun. 2026

[3]Agência Pará de Notícias. Disponível em: https://www.agenciapara.com.br/noticia/76376/belem-disputa-premio-nacional-de-melhor-experiencia-em-turismo-gastronomico. Acesso em: 05 de junho de 2026

ABSAPT-2024 [Thurow Bender et al. 2024, Pereira et al. 2026] permitem comparabilidade direta com os *benchmarks* estabelecidos.

Diante disso, este trabalho tem como objetivo coletar, processar e analisar avaliações da plataforma Booking referentes a Belém, Santarém e Alter do Chão, combinando Análise Exploratória de Dados (AED) e ABSA em português. A AED busca identificar padrões temporais e geográficos, enquanto a ABSA é empregada para extrair aspectos da experiência de hospedagem e suas respectivas polaridades. Assim, o estudo busca reduzir a lacuna de trabalhos sobre PLN aplicados ao turismo amazônico e gerar conhecimento acionável para gestores de hospedagem e para políticas públicas de turismo.

O restante deste artigo está organizado da seguinte forma: a Seção 2 apresenta trabalhos relacionados que abordam ABSA e turismo amazônico; a Seção 3 descreve a metodologia adotada, incluindo procedimentos de coleta, pré-processamento e modelagem; a Seção 4 detalha os resultados experimentais e a discussão; e a Seção 5 conclui o estudo e apresenta sugestões para trabalhos futuros.

2. Trabalhos Relacionados

A análise do CGU em plataformas de turismo é uma área ativa de pesquisa, envolvendo mineração de texto, análise de sentimentos e modelagem de tópicos a partir de avaliações online. Nesta Seção, destacam-se estudos relacionados a duas frentes principais: (i) análise do turismo regional amazônico e (ii) análise de sentimentos baseada em aspectos de dados de plataformas de avaliação online. No contexto amazônico, [de Souza et al. 2020] investigaram, por meio de questionários em escala Likert aplicados a 94 hóspedes de 13 hotéis e de análise fatorial, os atributos que turistas de negócios consideram na escolha de hospedagem em Belém, destacando "presteza" e "higiene e segurança" como fatores centrais. [Junior et al. 2020] analisaram 3.881 avaliações de restaurantes de Santarém no TripAdvisor com *Polyglot*, *Non-Negative Matrix Factorization* (NMF) e *Term Frequency-Inverse Document Frequency* (TF-IDF), identificando predominância de comentários positivos (62,5%) e tópicos associados a atendimento, localização e cardápio. [Aleixo et al. 2023] compararam o CGU do setor hoteleiro em Alter do Chão nas plataformas Booking e TripAdvisor, também com *Polyglot*, *Latent Dirichlet Allocation* (LDA) e TF-IDF, e encontraram predominância de sentimentos positivos e tópicos recorrentes sobre localização, atendimento, conforto e limpeza.

Na frente de ABSA em português, [Thurow Bender et al. 2024] apresentaram a tarefa ABSAPT-2024, voltada ao desenvolvimento de métodos para Extração de Termo de Aspectos (do inglês *Aspect Term Extraction*, ATE) e Classificação de Sentimento por Aspecto (do inglês *Aspect Sentiment Classification*, ASC) no domínio hoteleiro, utilizando modelos baseados em BERT e esquema de anotação BIO Tagging (B - *Begin*; I - *Inside*; O - *Out*); os melhores resultados reportados foram F1-score=0,63 para ATE e F1-score=0,65 para ASC. Posteriormente, [Pereira et al. 2026] investigaram abordagens com *transfer learning* usando BERTimbau-Large no corpus ABSAPT-2024, comparando um pipeline sequencial ATE + ASC e uma abordagem fim-a-fim (E2E) com BIO expandido; os resultados indicaram equilíbrio entre as estratégias, com maior Recall no pipeline, maior Precisão na E2E e F1-score=0,72 em ambas. Ademais, [Pereira et al. 2024] exploraram *instruction tuning* em modelos de linguagem de pequeno porte para ABSA E2E em português, com 5.000 sentenças de avaliações de restaurantes rotuladas com GPT-

3.5, comparando PTT5 Base, FLAN-T5 Base e mT0 Small, em que o PTT5 Base obteve o melhor desempenho, com F1-score=0,60, Precisão=0,61 e Recall=0,59, reforçando a importância do pré-treinamento específico para o português. A Tabela 1 sintetiza as principais características dos estudos discutidos nesta Seção, destacando o domínio e a região de aplicação, a fonte de dados, a técnica empregada e se o estudo aplica ABSA em cidades amazônicas.

Tabela 1. Comparação entre os trabalhos relacionados e este trabalho.

Estudo	Domínio / Região	Dados	Técnica
[de Souza et al. 2020]	Hospedagem (negócios) em Belém/PA	Questionário (94 hóspedes)	Escala Likert + análise fatorial
[Junior et al. 2020]	Restaurantes em Santarém/PA	TripAdvisor (3.881 av.)	*Polyglot* + NMF + TF-IDF (sentimento e tópicos)
[Aleixo et al. 2023]	Hospedagem em Alter do Chão/PA	Booking + TripAdvisor	*Polyglot* + LDA + *Bag-of-Words*/TF-IDF
[Thurow Bender et al. 2024]	Hospedagem em diversas cidades (BR)	TripAdvisor (ABSAPT-2024)	BERT + BIO *tagging* (benchmark ATE e ASC)
[Pereira et al. 2026]	Hospedagem em diversas cidades (BR)	TripAdvisor (ABSAPT-2024)	BERTimbau-Large, *pipeline* ATE + ASC e E2E
[Pereira et al. 2024]	Restaurantes (BR)	Avaliações rotuladas via GPT-3.5	*Instruction tuning* em LLMs pequenos (PTT5, FLAN-T5, mT0)
Este trabalho	**Hospedagem em Belém, Santarém e Alter do Chão/PA**	**Booking (103.848 av.)**	**BERTimbau-Large (ATE, ASC, E2E) + PyABSA, comparando cenários e cidades**

Observa-se que, embora [Thurow Bender et al. 2024] e [Pereira et al. 2026] apliquem ABSA com modelos BERT ao domínio hoteleiro em português, nenhum dos dois é voltado a destinos amazônicos, enquanto os estudos regionais sobre o Pará [de Souza et al. 2020, Junior et al. 2020, Aleixo et al. 2023] não empregam ABSA, restringindo-se a análises de sentimento geral, modelagem de tópicos ou métodos estatísticos tradicionais. Identifica-se, portanto, uma lacuna tanto na aplicação do ABSA no setor hoteleiro em cidades amazônicas quanto nos estudos comparativos entre cidades de diferentes dimensões. O presente estudo aplica ABSA ao domínio de hospedagem em destinos amazônicos de diferentes portes, com auxílio de um corpus anotado para a tarefa em português, comparando diferentes cenários e abordagens de inferência para avançar da avaliação técnica de modelos à geração de conhecimento turístico aplicável.

3. Metodologia

O presente trabalho foi desenvolvido com base no método *Data Science Trajectories* (DST), proposto por [Martínez-Plumed et al. 2019]. A escolha justifica-se por sua natureza exploratória e flexível. A metodologia foi organizada em seis fases: Entendimento do Negócio, Aquisição dos Dados, Entendimento dos Dados, Preparação dos Dados, Modelagem, Avaliação e Exploração dos Resultados. Na fase do **Entendimento do Negócio**, foi realizada uma revisão da literatura do tipo *ad hoc* [Ralph and Baltes 2022], investigando trabalhos sobre turismo em regiões amazônicas, PLN aplicado a avaliações turísticas e pipelines com modelos BERT para ABSA, discutidos nas Seções 1 e 2. Na etapa de **Aquisição dos Dados**, os dados foram coletados do Booking via *crawler*[4]. As coletas de dados das três cidades ocorreram em outubro de 2025 e em abril de 2026, gerando dois *datasets* complementares, armazenados no formato BSON no MongoDB.

[4]Utilizando Python (3.12) e as seguintes bibliotecas: Selenium (4.33), Flask (3.1.1) e pymongo (4.13.1).

Depois, para o ajuste fino dos modelos de ABSA em português, foi solicitado o *corpus* ABSAPT-2024 aos autores de [Thurow Bender et al. 2024] via formulário[5].

Em **Entendimento dos Dados**, os dados do Booking incluem informações do estabelecimento (nome, localização, pontuação, comodidades) e de cada avaliação (nacionalidade, tipo de quarto e de viagem, duração da estadia, comentários positivos/negativos e resposta do estabelecimento). O ABSAPT-2024 é um *corpus* anotado de ABSA em português brasileiro no domínio hoteleiro, composto por treino com 4.828 registros, teste de ATE (Tarefa 1) com 283 instâncias e teste de ASC (Tarefa 2) com 1.176 instâncias, com 82 aspectos únicos e distribuição assimétrica: *hotel* (902), *quarto* (662), *localização* (494) e *café da manhã* (346) concentram aproximadamente metade dos registros. Na fase de **Preparação dos Dados**, os dois *datasets* complementares do Booking foram concatenados e deduplicados, garantindo a presença dos estabelecimentos em ambos os períodos de coleta. A deduplicação de estabelecimentos considerou URL, localização e nome; a de avaliações considerou título, textos positivos e negativos, pontuação, data, hóspede, tipo de viagem e de quarto, com verificação manual da qualidade da concatenação. A Tabela 2 apresenta o quantitativo do processo, com *Ant./Nov.*: estabelecimentos antigos/novos; *Conc.*: uniões realizadas; *Est. final*: estabelecimentos consolidados; *Rev. br./Dup. rem./Rev. final*: avaliações brutas, duplicadas removidas e finais.

Tabela 2. Resultados da consolidação por cidade.

Cidade	*Ant.*	*Nov.*	*Conc.*	*Est. final*	*Rev. br.*	*Dup. rem.*	*Rev. final*
Alter do Chão	131	56	48	139	11.002	2.881	8.121
Santarém	76	78	48	106	20.086	8.233	11.853
Belém	222	407	115	514	129.822	45.948	83.874
Total	**429**	**541**	**211**	**759**	**160.910**	**57.062**	**103.848**

Adicionalmente, gerou-se um *dataframe* de avaliações (`df_reviews`), mantendo apenas avaliações de hóspedes brasileiros e removendo textos compostos exclusivamente por padrões (*e. g.*, "Não há comentários disponíveis para esta avaliação.") ou por *placeholders*, como *nenhum*, *nada* e *n/a*. Os campos título, comentário positivo e comentário negativo foram concatenados, quando válidos, em uma nova coluna do *dataframe*. No corpus ABSAPT-2024, realizaram-se o recálculo dos índices de início e de fim dos aspectos no texto normalizado e a conversão das classes originais de polaridade de sentimento $\{1, -1, 0\}$ para $\{\texttt{Positive}, \texttt{Negative}, \texttt{Neutral}\}$. Na etapa de **Modelagem**, realizou-se o *fine-tuning* do BERTimbau Large [Souza et al. 2020], seguindo a metodologia de [Pereira et al. 2026], com três modelos especializados em ABSA: ATE, ASC e E2E. A tarefa de ATE foi tratada como rotulagem de sequências (*Sequence Labeling*) [Jakob and Gurevych 2010], com o esquema BIO [Ramshaw and Marcus 1995] convertendo as anotações para os rótulos `B-ASP`, `I-ASP` e `O`, consolidados pelos benchmarks do SemEval-2014 [Pontiki et al. 2014].

O modelo ATE foi formulado como uma classificação de *tokens* com esquema BIO e comprimento máximo de 512 *tokens*, seguindo a arquitetura baseada em BERT [Devlin et al. 2019]. O modelo ASC foi formulado como uma classificação de sequência, recebendo o par $[\,\text{texto} \mid \text{aspecto}\,]$ separado por `[SEP]`, com três classes de saída (`Pos` - Positivo; `Neu` - Neutro; `Neg` - Negativo) e comprimento máximo de 256 *tokens*, con-

[5]ABSAPT 2024. Google Sites. Disponível em: `https://sites.google.com/inf.ufpel.edu.br/absapt-2024/`. Acesso em: 08 de dezembro de 2025

forme a metodologia de construção de quase-frases proposta por [Sun et al. 2019]. Por fim, o E2E realiza a extração e a classificação simultaneamente, mapeando o problema para uma tarefa de rotulagem com sete classes (`O`, `B-POS`, `I-POS`, `B-NEG`, `I-NEG`, `B-NEU`, `I-NEU`) e comprimento máximo de 512 *tokens*, com base no esquema unificado de [Li et al. 2019]. O treinamento seguiu os hiperparâmetros de [Pereira et al. 2026], com a adição do hiperparâmetro *Patience_E2E*=3, pois a tarefa E2E apresentou maior instabilidade nas primeiras épocas (F1-score próximo de zero), o que exigiu maior paciência para evitar interrupções prematuras.

Antes do treinamento final, os três modelos (ATE, ASC e E2E) foram submetidos a uma validação cruzada k-fold ($k = 5$) sobre o conjunto de treino do ABSAPT-2024 (4.828 registros), com partições geradas por `KFold` (*scikit-learn*), embaralhamento habilitado e *seed* fixa (42), em uma única execução. O agrupamento das partições foi realizado em nível de texto de avaliação único, evitando que aspectos de uma mesma avaliação fossem divididos entre treino e validação no mesmo *fold*. Em cada *fold*, um modelo foi treinado do zero e avaliado com as métricas do conjunto de teste oficial (Precisão, *Recall* e F1-score via *seqeval* para ATE e E2E; *Balanced accuracy*, F1-macro, Precisão-macro e Recall-macro para ASC). A validação cruzada estimou a estabilidade de cada modelo (média e desvio padrão entre os cinco folds, reportados na Seção 4) e definiu o número de épocas do modelo final, calculado pela média das épocas ótimas (*early stopping*) observadas nos folds. A inferência sobre o `df_reviews` seguiu um desenho fatorial de dois cenários de entrada $\times$ três abordagens, conforme a Tabela 3.

Tabela 3. Cenários de entrada e abordagens avaliadas na inferência

Cenários de entrada	
Cenário A (*fulltext*)	Concatena o título, o texto positivo e o texto negativo em uma única entrada.
Cenário B (*separate*)	Processa o título, o texto positivo e o texto negativo separadamente, consolidando os aspectos ao final.
Abordagens avaliadas	
Abordagem 1 (ATE + ASC)	Pipeline em dois estágios com os modelos ajustados neste trabalho: ATE identifica os *spans*; ASC classifica o sentimento a partir do par [texto, aspecto].
Abordagem 2 (E2E)	Modelo E2E ajustado neste trabalho, que extrai aspectos e classifica polaridade em uma única passagem, mapeando as sete classes para três polaridades.
Abordagem 3 (PyABSA)	Linha de base multilíngue e não ajustada (*off-the-shelf*) da biblioteca PyABSA (v. 2.3.4) [Yang et al. 2023], checkpoint público `multilingual` do módulo *Aspect Term Extraction and Polarity Classification* (ATEPC), que realiza extração de aspectos e classificação de polaridade conjuntamente, correspondendo, em escopo de tarefa, à Abordagem 2 (E2E), com saída normalizada no mesmo formato das demais abordagens.

Na fase de **avaliação**, o desempenho dos modelos foi mensurado nos conjuntos de teste oficiais do ABSAPT-2024, com as métricas de [Thurow Bender et al. 2024, Pereira et al. 2026]: para ATE, Precisão, Recall e F1-score em nível de *span* via *seqeval* [chakki-works 2024]; para ASC, *Balanced accuracy* (Bacc) e F1-macro nas três classes de polaridade; o E2E foi avaliado conjuntamente, considerando a localização do *span*[6] e a polaridade associada. Após a inferência no Booking, realizou-se uma análise de erros por combinação cenário $\times$ abordagem e uma verificação de sanidade da polaridade, esperando-se predominância positiva em campos positivos e negativa em campos negativos, usada como indicador qualitativo de coerência das predições.

Por fim, na **exploração dos resultados**, os resultados brutos foram normalizados

[6]Segmento de texto contínuo, contendo uma ou mais palavras, que identifica o alvo específico (aspecto) de um texto.

antes da análise final, consolidando variantes morfológicas em formas canônicas por meio de um dicionário manual (*e. g.*, *quartos* → *quarto*; *café* → *café da manhã*). Os aspectos corrigidos foram então enriquecidos com metadados dos hotéis, incluindo nome, cidade, coordenadas geográficas e datas das avaliações.

4. Resultados e discussão

Esta Seção apresenta os principais resultados: AED do dataset do Booking; ajuste fino das três variações do BERTimbau (ATE, ASC e E2E) no ABSAPT-2024, comparado a [Thurow Bender et al. 2024, Pereira et al. 2026]; inferência dos modelos no `df_reviews`; e análise temporal e regional dos resultados. Os materiais complementares estão disponíveis no GitHub[7]. Após a consolidação dos dois períodos de coleta, o corpus reuniu 103.848 avaliações das três localidades, sendo 49.885 em português, apresentadas na Tabela 4. Belém concentra 80,8% das avaliações, enquanto Santarém e Alter do Chão representam 11,4% e 7,8%, respectivamente.

Tabela 4. Volume de avaliações por município no corpus consolidado.

Município	Avaliações (%)	Avaliações PT-BR (%)
Belém	83.874 (80,8%)	39.125 (78,4%)
Santarém	11.853 (11,4%)	5.706 (11,4%)
Alter do Chão	8.121 (7,8%)	5.054 (10,1%)
Total	**103.848 (100,0%)**	**49.885 (100,0%)**

A análise do perfil revelou predominância de avaliadores brasileiros, reforçando a pertinência de modelos especializados em português. Em Belém e Santarém predomina a viagem individual (respectivamente, 32,5% e 37%), seguida da viagem em casal, enquanto Alter do Chão apresenta predominância de viagens em casal (42,9%). Belém e Santarém concentram estadias de uma diária (47% e 51,7%), ao passo que Alter do Chão apresenta maior equilíbrio entre duas e três diárias. Quanto à sazonalidade, Belém apresenta distribuição mais constante, enquanto Santarém e Alter do Chão registram aumento de aproximadamente 50% nos meses de alta (ago–out) em relação ao período chuvoso (fev–mai) [8].

Na validação cruzada, ATE e E2E apresentaram maior estabilidade, enquanto ASC apresentou maior variabilidade (Bacc $0{,}689 \pm 0{,}178$), associada principalmente à classe `Neutra`. Nos testes oficiais, ATE alcançou F1 de 0,651 (pipeline) e 0,633 (E2E), próximos ao ABCD da ABSAPT-2024, mas inferiores a [Pereira et al. 2026], diferença atribuída à ausência de normalização lexical dos aspectos na Tarefa 1. Na Tarefa 2, ASC obteve os melhores resultados entre os sistemas comparados (Bacc=0,770; F1-macro=0,772), superando os participantes oficiais e apresentando desempenho equivalente a [Pereira et al. 2026]. A classe `Neutra` permaneceu como principal desafio. Na avaliação composta, pipeline e E2E apresentaram desempenho praticamente idêntico (F1=0,557 e 0,558), indicando equilíbrio entre as estratégias.

Na inferência sobre o `df_reviews`, o Cenário A apresentou cobertura superior a 82% nas três abordagens, com o pipeline atingindo 86,43% de avaliações com pelo menos um aspecto. PyABSA, entretanto, identificou 5.738 aspectos únicos, contra 851

[7]Disponível em: `https://github.com/vitinhx/absa_eniac_2026`

[8]Instituto Nacional de Meteorologia (INMET). Disponível em: `https://portal.inmet.gov.br/informativos`. Acesso em: 19 de junho de 2026

do pipeline e 916 do E2E, evidenciando menor consolidação lexical. No Cenário B, os textos positivos concentraram polaridades positivas (81–90%), enquanto os negativos apresentaram predominantemente polaridades negativas (68–90%), fornecendo validação qualitativa das predições. Os títulos apresentaram baixa cobertura em todas as abordagens, resultado que pode decorrer tanto da menor riqueza descritiva desse campo quanto da dificuldade de extração em textos curtos; essa distinção requer validação adicional e é discutida na Seção 5.

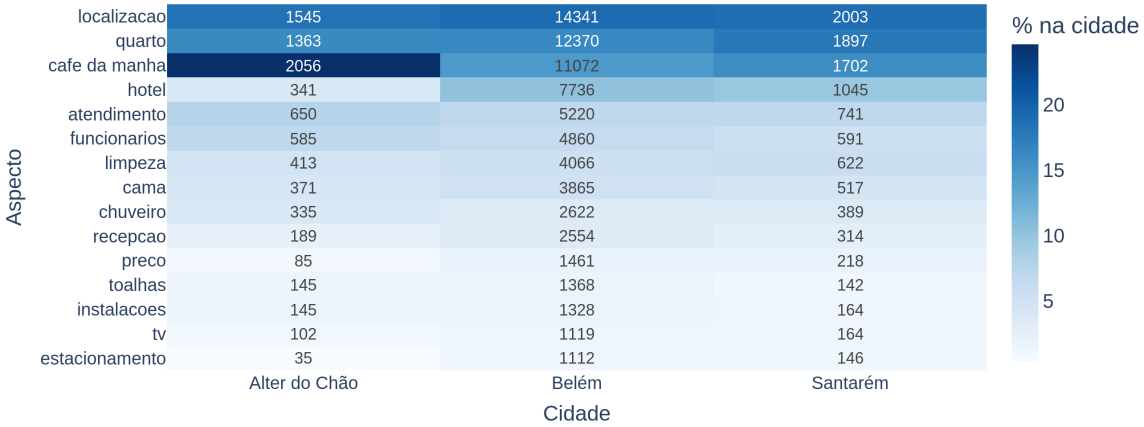


Figura 1. Frequência dos 15 aspectos mais frequentes por cidade (% normalizada)

Na Figura 1, em Alter do Chão, predomina o aspecto ‘Café da manhã’, com 24,6%. Por outro lado, em Santarém e Belém destacam-se ‘Localização’ e ‘Hotel’, em comparação com Alter. Na análise de sentimento dos quinze aspectos mais frequentes, na Figura 2, a cor de cada célula representa o Sentimento Médio (SM) do aspecto naquela cidade, calculado pela média das polaridades individuais das ocorrências do aspecto (mapeadas para $\{-1, 0, +1\}$) e segmentado por cidade; a escala varia de vermelho (SM próximo de -1, sentimento negativo) a azul (SM próximo de $+1$, sentimento positivo), com branco representando neutralidade. O número exibido em cada célula corresponde à quantidade e ao percentual de ocorrências do aspecto em relação ao total dos quinze aspectos exibidos naquela cidade, normalização que segue a mesma base da Figura 1 e mitiga o efeito do desbalanceamento geográfico do corpus (Tabela 4); a cor (SM), por se tratar de uma média e não de uma contagem, já não é diretamente afetada por esse desbalanceamento. Observa-se que Alter do Chão apresentou as maiores pontuações médias para a maior parte dos aspectos, com exceção do ‘Preço’. A pontuação média de Alter do Chão é a maior, com SM de 0,63 (12.292 aspectos), seguida de Santarém com 0,31 (13.721) e Belém com 0,28 (98.182), sugerindo que destinos menores apresentaram avaliações mais favoráveis. Ainda na Figura 2, os aspectos ‘TV’, ‘Chuveiro’ e ‘Toalhas’ apresentaram as menores pontuações em todas as três cidades, indicando que problemas de infraestrutura e manutenção tendem a ser recorrentes independentemente do destino. Em contrapartida, ‘Localização’, ‘Atendimento’ e ‘Funcionários’ estão entre os mais bem avaliados, evidenciando a importância dos elementos ligados à experiência e ao serviço prestado.

Na Figura 3, apresenta-se a coocorrência dos aspectos em avaliações, utilizando como referência o Cenário A e a abordagem *pipeline*. Observa-se que

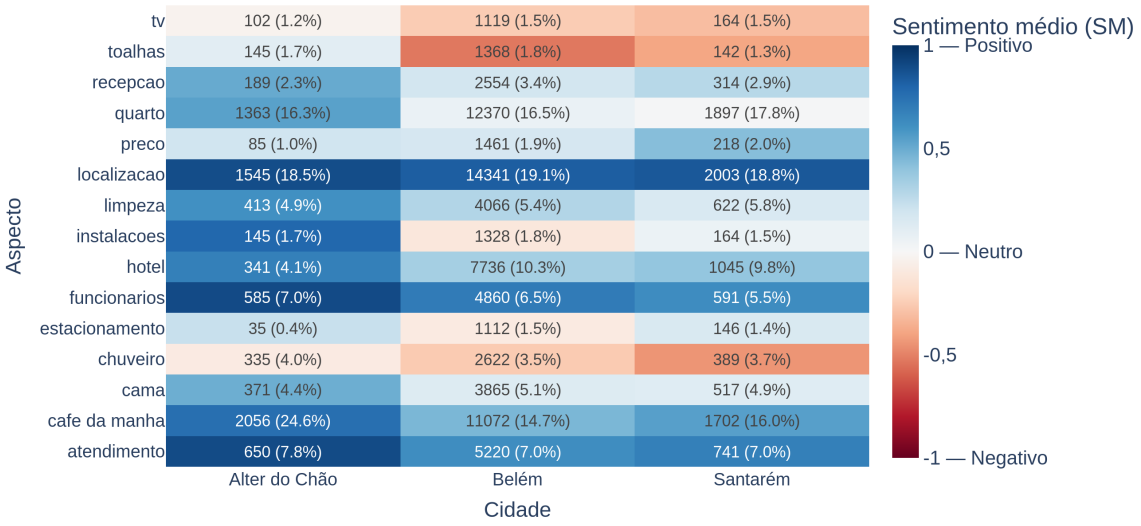


Figura 2. Distribuição (quantidade e percentual) e sentimento médio por Aspecto e Cidade.

'Atendimento' coocorre tanto com aspectos do mesmo contexto ('Hotel', 1.324 coocorrências; 'Funcionários', 1.123) quanto com dimensões distintas da experiência, como 'Localização' (2.390), em que o par `Atendimento–Localização` é mais frequente do que `Atendimento–Hotel` ou `Atendimento–Funcionários`. Destacam-se ainda pares sinérgicos de alta frequência, como `Café da Manhã–Quarto` (6.108) e `Quarto–Hotel` (3.718), e pares complementares, como `Café da manhã–Localização` (6.239) e `Funcionários–Localização` (2.475).

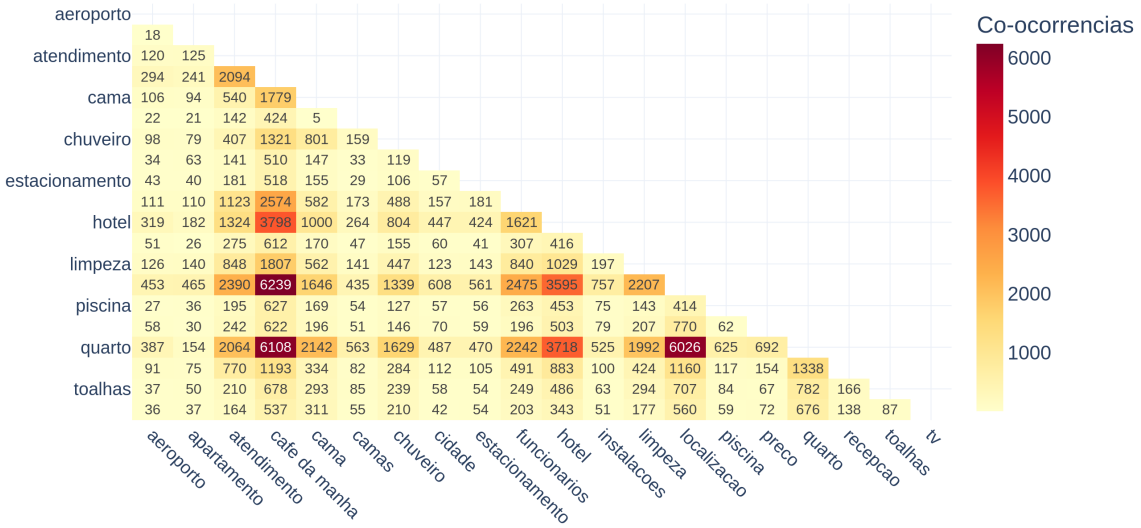


Figura 3. Coocorrências dos 20 aspectos mais frequentes (Cenário A - Pipeline)

Esses resultados sugerem grupos de aspectos relacionados entre si, seja por pertencerem à mesma dimensão do serviço, como 'Quarto' e 'Hotel', seja por representarem dimensões complementares da experiência, como 'Localização' e 'Café da Manhã'. A gestão da experiência do hóspede deve, portanto, considerar como os aspectos são percebidos conjuntamente pelos turistas, e não de forma isolada, o que pode subsidiar futuras pesquisas sobre agrupamento e normalização semântica de aspectos no domínio hoteleiro.

5. Considerações finais

Este trabalho investigou a extração de conhecimento acionável, em nível de aspecto, sobre a percepção dos hóspedes em três destinos turísticos do Pará a partir de avaliações do Booking, demonstrando que a combinação de análise exploratória com análise de sentimentos baseada em aspectos em português é viável para transformar avaliações textuais em uma análise estruturada da experiência de hospedagem, revelando padrões temporais, diferenças geográficas e sentimentos associados a aspectos específicos. Os principais achados evidenciam forte concentração das avaliações em Belém, predominância de avaliadores brasileiros, notas elevadas e picos sazonais no segundo semestre. Os modelos ajustados apresentaram estabilidade na validação cruzada, com desempenho equilibrado entre as abordagens *pipeline* e E2E; na inferência sobre os dados do Booking, observou-se predominância de aspectos positivos, coerência entre a origem do texto e a polaridade predita, e contrastes entre destinos: Belém e Santarém destacaram localização, hotel e quarto, enquanto Alter do Chão evidenciou a predominância de café da manhã. Assim, o estudo contribui para reduzir a lacuna de trabalhos de processamento de linguagem natural aplicados ao turismo na Amazônia brasileira, empregando modelos *Transformer* ajustados ao português e comparando diferentes cenários e abordagens de inferência para apoiar a gestão hoteleira e a formulação de políticas públicas.

Como limitações, destacam-se: a ABSA adotada prioriza aspectos e opiniões explícitos; a classe Neutra permaneceu como o principal gargalo de desempenho; e o limite máximo de *tokens* gerou truncamentos nos resultados brutos, ainda que corrigidos posteriormente. Além disso, os dados restringem-se a uma única plataforma e região, com desbalanceamento em favor de Belém; a baixa cobertura observada no campo título não permite distinguir, com os dados atuais, entre ausência real de informação descritiva e falha de extração dos modelos; e a coerência das predições sobre o corpus do Booking foi avaliada apenas qualitativamente (verificação de sanidade da polaridade), devido à ausência de um *gold standard* anotado para essa base. Como trabalhos futuros, propõe-se ampliar a coleta para outras plataformas e municípios amazônicos, realizar monitoramento longitudinal, tratar aspectos implícitos e comparativos, integrar a normalização de aspectos ao protocolo de avaliação, conduzir uma validação manual de uma amostra do corpus do Booking (ou validação auxiliar com LLM) para mensurar diretamente a confiabilidade dos aspectos e polaridades extraídos, e desenvolver um painel interativo de apoio à decisão para gestores de hospedagem e para políticas públicas de turismo.

Agradecimentos e Uso de IA generativa

Este trabalho foi apoiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), por meio das bolsas DT-303031/2023-9, DT-302850/2023-6, DT-315052/2026-0 e PIBIC - 179796/2025-9; pelo Tribunal de Justiça do Maranhão, por meio do Acordo de Cooperação Técnica nº 02/2021 (processo nº 38328/2020-TJ/MA); e pela Financiadora de Estudos e Projetos (FINEP) - Pró-Amazônia - 2.373/24 - CTCCA-II.

Declara-se que os modelos de IA generativa foram utilizados como ferramenta de apoio. O *ChatGPT* foi utilizado para a melhoria dos termos de busca de artigos, a revisão gramatical e a elaboração de tabelas, enquanto o *Claude* foi utilizado para a revisão e a orquestração do pipeline desenvolvido. O Grammarly Pro foi utilizado para a revisão gramatical. A autoria e a responsabilidade integral pelo conteúdo final, incluindo a verificação de plágio e as correções, são da primeira pessoa autora.

Referências

Aleixo, M. J. A., de Sousa Araujo, G., do Carmo, F. A., Junior, A. F. L. J., and Lobato, F. M. F. (2023). Análise comparativa de plataformas de mídias sociais turísticas: um estudo de caso de booking e tripadvisor para o município de santarém, pará. *Anais do Computer on the Beach*, 14:070–077.

Buhalis, D. and Volchek, K. (2022). Bridging marketing theory and big data analytics: The taxonomy of marketing attribution. *International Journal of Information Management*, 56:102253.

chakki-works (2024). seqeval: A python framework for sequence labeling evaluation. GitHub. Acesso em: 05 de junho de 2026.

Cheng, L.-C., Huang, H.-Y., and Huang, Y.-W. (2025). Multi-task chinese aspect-based sentiment analysis framework for service improvement: a case study on bnb reviews. *Electronic Commerce Research*, 25(6):4765–4798.

Chu, S.-C., Kim, J., and Chen, H. (2022). The power of social media in destination choice: Exploring the roles of online reviews and user-generated content in tourism. *Journal of Travel & Tourism Marketing*, 39(2):214–234.

de Souza, A. C., Batista, K. R. M., and Gomes, S. C. (2020). Hospedagens para turistas a negócios: quais atributos importam? *RITUR-Revista Iberoamericana de Turismo*, 10(1):202–220.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 4171–4186.

Egger, R. and Gokce, E. (2022). Natural language processing (nlp): An introduction: making sense of textual data. In *Applied data science in tourism: Interdisciplinary approaches, methodologies, and applications*, pages 307–334. Springer.

Jakob, N. and Gurevych, I. (2010). Extracting opinion targets in a single and cross-domain setting with conditional random fields. In *Proceedings of the 2010 conference on empirical methods in natural language processing*, pages 1035–1045.

Junior, L. F., Junior, J. S., and Lobato, F. (2020). Um olhar sobre turismo gastronômico: Um caso no tripadvisor. In *Encontro Nacional de Inteligência Artificial e Computacional (ENIAC)*, pages 519–530. SBC.

Li, X., Bing, L., Li, P., and Lam, W. (2019). A unified model for opinion target extraction and target sentiment prediction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 6714–6721.

Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernández-Orallo, J., Kull, M., Lachiche, N., Ramírez-Quintana, M. J., and Flach, P. (2019). Crisp-dm twenty years later: From data mining processes to data science trajectories. *IEEE transactions on knowledge and data engineering*, 33(8):3048–3061.

Pereira, F. N., de Freitas, L. A., and Corrêa, U. B. (2026). Análise de sentimento baseada em aspectos no domínio de acomodações utilizando o modelo bertimbau. In *Procee-*

dings of the 17th International Conference on Computational Processing of Portuguese (PROPOR 2026)-Vol. 1, pages 830–837.

Pereira, G., Barbosa, L., Moreira, J., Melo, T., and Silva, A. (2024). Enhancing aspect-based sentiment analysis for portuguese using instruction tuning. In *Encontro Nacional de Inteligência Artificial e Computacional (ENIAC)*, pages 990–1001. SBC.

Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., and Manandhar, S. (2014). SemEval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, Dublin, Ireland. Association for Computational Linguistics.

Ralph, P. and Baltes, S. (2022). Paving the way for mature secondary research: the seven types of literature review. In *Proceedings of the 30th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, pages 1632–1636.

Ramshaw, L. and Marcus, M. (1995). Text chunking using transformation-based learning. In *Third workshop on very large corpora*.

Souza, F., Nogueira, R., and Lotufo, R. (2020). BERTimbau: pretrained BERT models for Brazilian Portuguese. In *9th Brazilian Conference on Intelligent Systems, BRACIS, Rio Grande do Sul, Brazil, October 20-23 (to appear)*.

Sun, C., Huang, L., and Qiu, X. (2019). Utilizing bert for aspect-based sentiment analysis via constructing auxiliary sentence. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 380–385.

Thommandru, A., Espinoza-Maguiña, M., Ramirez-Asis, E., Ray, S., Naved, M., and Guzman-Avalos, M. (2023). Role of tourism and hospitality business in economic development. *Materials Today: Proceedings*, 80:2901–2904.

Thurow Bender, A., Gomes, G. A., Lopes, E. P., Araujo, R. M., de Freitas, L. A., and Corrêa, U. B. (2024). Overview of absapt at iberlef 2024: Overview of the task on aspect-based sentiment analysis in portuguese. *Procesamiento del lenguaje natural*, 73:315–322.

Tripathy, G. and Sharaff, A. (2026). Traversing the landscape of aspect-based sentiment analysis: Delving deeper into techniques, trends, and future directions. *Computer Science Review*, 60:100885.

Yang, H., Zhang, C., and Li, K. (2023). Pyabsa: A modularized framework for reproducible aspect-based sentiment analysis. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, CIKM 2023*, pages 5117–5122. ACM.