

TOURIST VOICES ON TWITTER: SENTIMENT AND TOPIC ANALYSIS FOR SUSTAINABLE TOURISM IN PENANG

Kun Zhu

School of Computer Science, Universiti Sains Malaysia, Penang 11800, Malaysia

Nor Hasliza Md Saad*

School of Management, Universiti Sains Malaysia, Penang 11800, Malaysia

Ghada ElSayad

Business Information Systems Department, College of Management and Technology, Arab Academy for Science, Technology and Maritime Transport (AASTMT), Heliopolis, Cairo, Egypt

Muhamad Mu'az Mohd Ali

School of Management, Universiti Sains Malaysia, Penang 11800, Malaysia

Jiahui Qian

School of Management, Universiti Sains Malaysia, Penang 11800, Malaysia

ABSTRACT

This research aims to analyse the opinions expressed in tweets about a tourism destination in Malaysia and elucidate the dominant topics discussed on Twitter. This research examines tourist sentiment on a Malaysian destination from a sustainability perspective by analysing 18,018 tweets, collected from November 9, 2021, to January 9, 2022, using Lexicon-based sentiment analysis for polarity detection and Latent Dirichlet Allocation (LDA) for topic modelling. Results revealed that 46% of sentiments were positive, 39% were neutral, and 15% were negative. Tourists primarily discussed food, tourist spots, events, hotels, traffic, and driver attitudes. Core tourism services received positive feedback, whereas additional fees were noted as less favourable. The integration of sentiment analysis and LDA enables a nuanced understanding of tourist experiences, providing actionable insights for policymakers and businesses to enhance sustainable tourism practices and optimize resource management in Penang.

Keywords: Covid-19, Sentiment analysis, Sustainable tourism, Sustainability, Topic modelling

Received: 5th December 2024

Accepted: 2nd March 2025

<https://doi.org/10.33736/ijbs.9568.2025>

* Corresponding author: Associate Professor, School of Management, Universiti Sains Malaysia, Penang 11800, Malaysia. Tel: +6(04) 653 6012 Email: norhasliza@usm.my

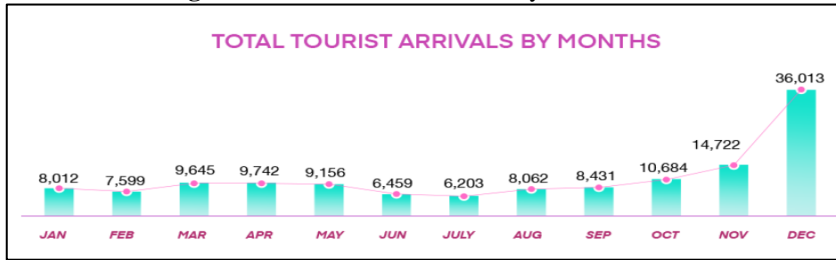
1. INTRODUCTION

Tourism has become one of Malaysia's most influential businesses and one of the most critical contributors to Gross Domestic Product (GDP): in 2019, the tourist industry contributed 15.9% of Malaysia's GDP (RM220.4 billion), making it the third-largest contributor (Department of Statistics Malaysia, 2020). The tourist industries provided RM86.1 billion to the national economy in 2019, representing a 2.3% rise over the previous year's figure of RM84.1 billion. Tourist arrivals in 2018 were down by 0.5% from the previous year, when they reached 25.95 million, but were expected to expand at an average rate of 4% per year by 2020, reaching 36 million. Tourism earnings were expected to increase by 13.6% per year, reaching RM168 billion. As a result, it was predicted that the tourist sector would provide 2.34 million jobs by 2020.

The Covid-19 pandemic has drastically affected the global economy, with tourism experiencing profound disruptions due to government-imposed measures like social distancing, travel restrictions, and quarantine mandates (BBC News, 2020). These actions severely curtailed visitor mobility, intensified perceptions of travel risk, and led to a steep decline in global tourism, with major destinations seeing up to a 90% drop in visitors (UNWTO, 2020a). By early 2020, nearly all international destinations enforced travel restrictions (UNWTO, 2020b). In Malaysia, the pandemic caused an 83.4% decrease in inbound arrivals, reducing numbers from 26.1 million in 2019 to 4.3 million in 2020, with tourism revenue falling by 85.3% (Tourism Malaysia, 2022). Consequently, the tourism sector faced economic challenges, leading to hotel closures and substantial financial losses, estimated at over RM6.53 billion (The EdgeProp, 2021). Despite these setbacks, Penang, a renowned tourist destination, displayed remarkable recovery following the lifting of the Movement Control Order (MCO). Penang saw a surge in tourism, with hotel occupancy rates reaching 70%, and beach hotels reaching 90% occupancy, attracting more than 50,000 visitors on the first weekend after restrictions were lifted (Dermawan, 2021), as shown in Figure 1. These indicators underscore the importance of examining tourist perceptions in the post-pandemic recovery phase.

Given the heightened uncertainty introduced by the pandemic, an in-depth analysis of tourists' sentiments and risk perceptions on platforms such as Twitter is crucial for practitioners and policymakers. Such insights can aid in assessing tourists' confidence, understanding perceived travel risks, and recognizing shifts in preferences and behaviours. The findings of this research will be instrumental for the Malaysian government in developing strategies that not only restore tourist confidence but also promote sustainable tourism practices. Moreover, this research is timely, as it directly addresses the challenges faced by governments in managing potential new waves of the virus while simultaneously revitalizing the tourism sector. By integrating sustainable development principles into recovery efforts, Penang can foster a resilient tourism sector that balances economic growth with environmental protection and community well-being. This approach will help ensure that the benefits of tourism are distributed equitably and that the natural and cultural resources that attract visitors are preserved for future generations.

Figure 1: Total Tourist Arrivals by Month, 2021



Source: Tourism Malaysia (2022), <https://www.tourism.gov.my/statistics>

2. LITERATURE REVIEW

This section reviews the existing literature on topic modelling, sentiment analysis, and the intersection of social media with tourism experiences, highlighting the methodologies and findings relevant to this research.

2.1. Penang Tourism Related Research

Penang, located on the northwest coast of Peninsular Malaysia, is renowned for its diverse attractions that combine cultural heritage, natural beauty, and culinary delights. One of the most significant draws for tourists is its cultural heritage, especially in George Town, a UNESCO World Heritage site. The city is home to numerous historical buildings, temples, and museums, which showcase its multicultural history. The vibrant street art scene adds a modern twist to the historical backdrop, offering an engaging and educational experience for visitors (Gumelar, 2020). Furthermore, community participation plays a crucial role in sustaining cultural heritage tourism. Local involvement in the tourism development process can enhance the visitor experience while preserving local culture and benefiting the community (Huibin & Marzuki, 2012).

In addition to cultural attractions, Penang offers stunning natural sites such as Penang Hill, known for its breathtaking views and cooler climate, making it a popular retreat from the bustling city below (Pokok Kelapa, n.d.). The island also caters to adventure enthusiasts with activities like hiking, cycling, and water sports, particularly in Penang National Park, which provides diverse trails for nature exploration and promotes environmental awareness (Md Saad & Yaacob, 2021). Penang's culinary scene is another major attraction, with the state being dubbed the food capital of Malaysia. The diverse blend of Malay, Chinese, and Indian cuisines, especially street food, draws food tourists from around the world. The use of spices in local dishes highlights the region's cultural diversity, making it a central part of Penang's tourism appeal (Sibila Lebe, 2015).

Furthermore, sentiment analysis of tourist experiences via platforms like Twitter reveals important insights into tourist perceptions of Penang, emphasizing the significance of service infrastructure and the destination environment in shaping tourist satisfaction (Md Saad & Yaacob, 2021). Overall, Penang's tourism landscape offers a unique combination of cultural, natural, and culinary attractions, with community involvement being vital for sustainable development.

2.2. Penang Tourism Related Research

"Topic model" refers to a statistical framework enabling the identification of prevalent subjects within specific document collections. One prominent approach is the Latent Dirichlet Allocation (LDA) model, a generative probabilistic method widely used in social media research to uncover hidden topics in large-scale text datasets (Cao et al., 2023). As an unsupervised machine learning technique, LDA does not require predefined topics. Its primary objectives are to identify latent themes within textual data by clustering semantically similar words and categorizing content according to the detected subjects. In the tourism sector, topic modelling provides valuable insights from public opinion, feedback, and reviews regarding travellers' experiences shared on platforms such as Twitter (Egger & Yu, 2022). By aggregating related words into thematic groups, topic modelling enhances categorization and reveals significant social media trends (Lamba & Madhusdhan, 2018). LDA's relevance to this research lies in its ability to efficiently extract recurring themes from large volumes of unstructured text, enabling a deeper understanding of public sentiment regarding tourism in Penang.

Topic modelling through LDA is an essential tool for extracting hidden patterns and thematic structures from large-scale textual data. It allows researchers to move beyond simple keyword analysis, revealing deeper insights into the underlying themes and sentiments within public discourse. In the context of this research, the application of LDA to Twitter data provides a systematic method for capturing diverse tourist experiences, thus offering a robust framework for exploring public perceptions of Penang as a travel destination.

2.3. Sentiment Analysis

Sentiment analysis, a branch of natural language processing (NLP), involves assessing and detecting the sentiment expressed in written content, which can originate from diverse sources like social media posts and product reviews (Bing, 2012). This analytical approach aids businesses in understanding social sentiments linked to their brand, products, or services while monitoring online discussions (Rambocas & Pacheco, 2018). It captures individuals' sentiments, opinions, attitudes, emotions, and evaluations regarding services, products, organizations, and events (Alessia et al., 2015). By analysing textual data, sentiment analysis is useful for interpreting brief communications across various mediums.

Over the past decade, sentiment analysis or opinion mining using new technologies and algorithms has become a significant analytical tool for various industries, including tourism. By leveraging large-scale sentiment analysis of social media data, researchers and practitioners can understand market trends, public reactions, and perceptions of specific tourism destinations (Prananda & Thalib, 2020), in boosting business performance (Ahmed et al., 2022) and for many other business-related purposes (Mehraliyev et al., 2022). In the context of tourism, sentiment analysis not only gauges tourist satisfaction but also identifies key areas where destinations like Penang can improve, aligning with sustainable tourism development goals.

Sentiment analysis offers critical insights into public opinions, which are especially valuable in tourism research. By assessing the sentiments expressed by tourists on social media, this technique will provide a comprehensive view of tourist satisfaction and highlight areas for improvement in Penang's tourism offerings.

2.4. Tourism Experiences and Social Media

Research into tourist behaviour and perceptions has typically relied on quantitative data analysed through advanced statistical techniques using questionnaires (Mirzaalian & Halpenny, 2019). However, the emergence of text mining allows for a more nuanced exploration of tourist behaviour and perceptions through social media content (Mehraliyev et al., 2022). Advanced text mining techniques facilitate the examination of large volumes of data generated on social media platforms (Cui et al., 2023). Social media platforms like Facebook, Instagram, and Twitter are extensively utilized for sharing thoughts and opinions on service quality and product experiences (Trinh Ngo et al., 2022). Among social media platforms, Twitter is particularly suited for this research due to its real-time nature, large user base, and the prevalence of short, opinion-driven content related to tourism. This makes it an ideal source for extracting public sentiment and perceptions. Many individuals now prefer to seek travel information online, as it offers convenient access to information, opportunities for opinion sharing, and travel booking (Li et al., 2021).

Moreover, travellers often share photos, videos, and reviews to document their experiences and perceptions of destinations (Martínez-Torres et al., 2015; Trinh Ngo et al., 2022). This shift from traditional sources like brochures to dynamic, real-time online reviews and posts underscores the value of social media in capturing authentic tourist sentiments. An experiential perspective on tourist behaviour is essential for understanding the quality of travel experiences, encompassing both positive emotions like enjoyment and negative feelings such as dissatisfaction. This perspective redefines the roles of tourism professionals and tourists, emphasizing the importance of realistic and immersive experiences that highlight the senses, emotions, and enjoyment derived from travel (Liu et al., 2021). Through social media analysis, researchers and stakeholders can identify the underlying drivers of satisfaction and dissatisfaction, fostering more responsive and adaptive tourism strategies. The integration of social media data, particularly from Twitter, offers a valuable opportunity to understand the nuanced experiences and perceptions of tourists. This approach enhances the ability to identify both the strengths and areas for improvement in tourism destinations, thereby supporting sustainable tourism development.

This research aims to mine Twitter data to explore tourist perceptions and behaviour regarding Penang, serving as an alternative to traditional data sources. By employing sentiment analysis and topic modelling, this research seeks to identify positive tourist experiences and uncover negative perceptions needing enhancement. Furthermore, it aims to utilize topic modelling to discover recurring themes within the context of Penang tourism as reflected in the collected Twitter data. This approach not only enhances understanding of tourist sentiment but also contributes to sustainable tourism development by identifying areas for improvement and fostering positive experiences for visitors.

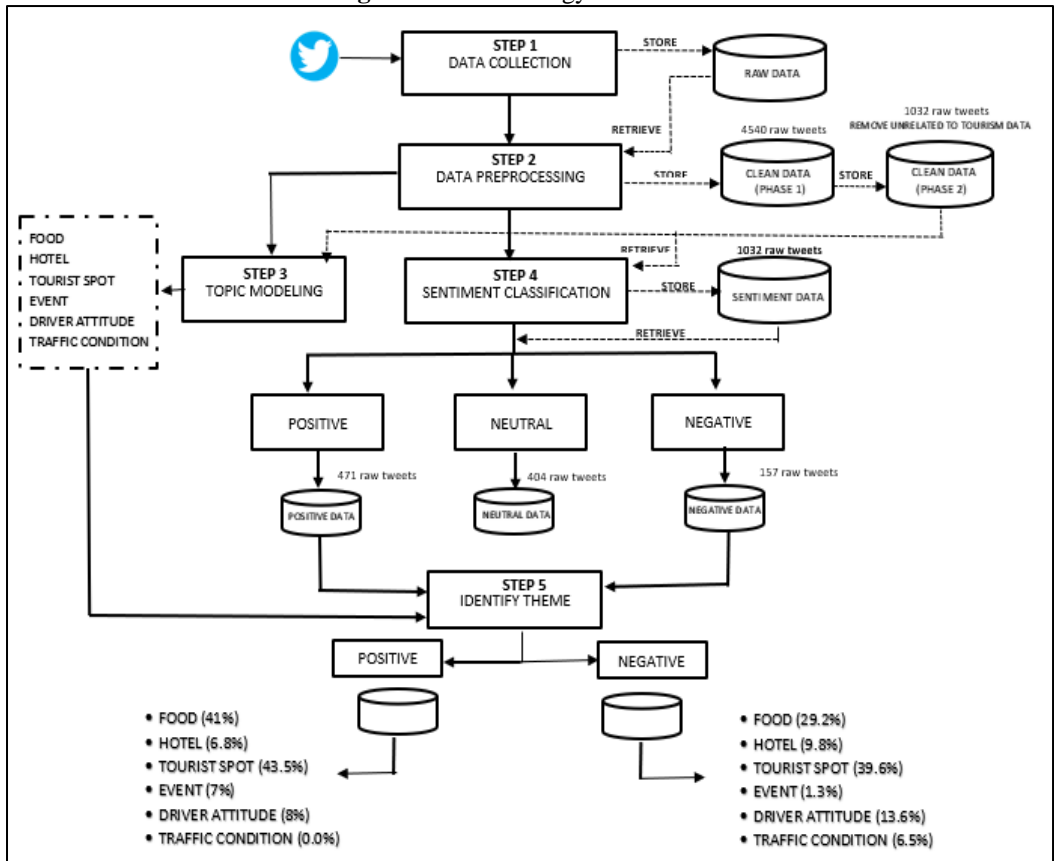
3. METHODOLOGY

This section outlines the methodological framework employed in this research, detailing the systematic approach taken to analyse tourist sentiment and perceptions related to Penang tourism through social media data. The research process consists of five key stages: data collection, data pre-processing, topic modelling, sentiment analysis, and data visualization, as illustrated in Figure 2.

3.1. Data Collection

For the Twitter data collection, this research employed RapidMiner's Search Twitter feature in conjunction with the Twitter API to retrieve tweets containing specific keywords and their associated metadata. The research was conducted using a Twitter user account, with the keywords "Penang" and "Georgetown" selected for analysis in English. The data mining process spanned two months, from November 9, 2021, to January 9, 2022. This period was strategically chosen as it followed Malaysia's initial implementation of 'bubble travel' on September 16, 2021, and the subsequent reopening of domestic travel on October 11, 2021. These developments marked a resurgence in domestic tourism in Malaysia. A total of 18,018 raw tweets were collected during this period and stored for further analysis. The time frame provides critical context, as the dataset captures tourist sentiment during Malaysia's tourism recovery phase.

Figure 2: Methodology Framework



3.2. Data Pre-Processing

The data cleaning process involved a series of steps to ensure the collected data was accurately prepared and consolidated for analysis, thereby improving dataset quality and reliability. Analytical tools facilitated streamlined processing and report generation.

In the pre-processing phase, unnecessary data attributes were removed by removing irrelevant characters, trimming white spaces, discarding duplicates, and addressing missing values. These steps ensured consistency, minimized potential biases, and enhanced analysis accuracy. Extraneous characters and spaces were removed using a subprocess operator to ensure data consistency, while duplicates were eliminated, and missing values were replaced to avoid skewed results.

After the initial cleaning in Phase 1 with RapidMiner, a secondary manual review was conducted in Excel to enhance dataset relevance. Specifically, tweets containing "Penang" and "Georgetown" were examined, and those unrelated to tourism—such as those focused on politics, marketing, or personal content—were excluded. This process resulted in a refined dataset that concentrated on tourism-specific content, including user experiences, perceptions of locations, and recommendations. The exclusion criteria ensured that only tweets relevant to the research's objectives were included, resulting in a final dataset optimized for tourism analysis.

3.3. Topic Modelling

Following data cleaning and pre-processing, topic modelling was conducted using the Latent Dirichlet Allocation (LDA) algorithm to uncover themes within the Twitter dataset. LDA, a generative probabilistic model, identifies topics as mixtures of words, making it effective for extracting patterns in large text corpora. In this research, the optimal number of topics was determined through iterative testing, balancing model complexity and interpretability, with coherence scores guiding topic relevance. The LDA model then highlighted significant terms, allowing each topic to be labelled according to its primary keywords. Ultimately, seven topics were identified and grouped into five overarching themes capturing key concerns in tourist sentiment on Penang tourism.

These themes reflect major areas of tourist interest, including local attractions, cuisine, events, accommodation quality, and transportation issues. The identification of these themes is significant, as they provide actionable insights for tourism stakeholders, including policymakers, business owners, and marketers. These insights enable the development of strategies aimed at improving visitor experiences, enhancing tourism services, and fostering sustainable tourism practices in Penang.

3.4. Sentiment Analysis

The tweet data collected from Twitter was analysed using sentiment analysis techniques to categorize and cluster the sentiments expressed in the tweets as positive, negative, or neutral. This research employed the Valence Aware Dictionary and sEntiment Reasoning (VADER) framework, a widely recognized tool for sentiment analysis, particularly suited for social media content (Hossain et al., 2023). VADER is effective because it not only assesses the polarity of sentiment (positive or negative) but also evaluates the intensity of these emotions, enabling a nuanced

understanding of user sentiments. VADER operates on a lexicon-based approach that links words to specific sentiment scores, which reflect both their emotional intensity and polarity. This lexicon includes various lexical attributes and is particularly adept at analysing informal language typical in social media, including slang, emoticons, punctuation, and capitalized words. It also incorporates rules that enable it to recognize context and modifiers that can amplify or diminish sentiment intensity (Qi & Shabrina, 2023). To compute the sentiment score for a given tweet, VADER aggregates the intensity scores of all individual words within the text. Each tweet is then classified based on its overall sentiment score: tweets with a positive score are classified as expressing positive sentiment, those with a negative score as negative sentiment, and those with a score close to zero as neutral.

The sentiment analysis process plays a pivotal role in identifying the emotional landscape of tourist experiences. It provides insights into common themes and emotions expressed by visitors, offering stakeholders a data-driven basis for improving services and addressing key concerns. This analysis also helps monitor shifts in public sentiment over time, particularly relevant in understanding the impacts of the Covid-19 pandemic on tourism recovery.

4. RESULTS

A total of 18,018 raw data items were initially collected, and following a cleaning process, 1,029 clean data items were retained. These data were subsequently classified into three main categories: positive, negative, and neutral. For the purpose of this paper, we will focus on the positive and negative sentiment groups, while the neutral group will not be discussed further, as it does not contribute significantly to the understanding of positive or negative sentiment.

4.1. Topic Modelling

The outcomes of the topic modelling analysis, following the text annotation process, are presented in Table 1. Each topic is defined by its associated keywords, identified based on frequency and relevance within the discussions on Twitter. The analysis highlights key themes frequently addressed by users in tweets about Penang, which include tourist attractions, culinary experiences, accommodation, events, driver behaviour, and traffic conditions. A higher confidence score corresponds with greater relevance, suggesting that the keywords within each identified topic accurately reflect users' experiences.

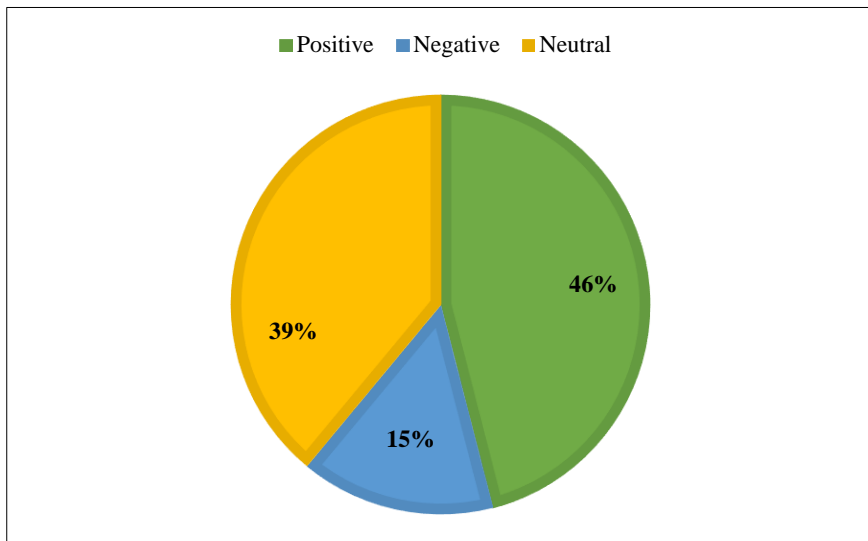
In Table 1, the identified topics are categorized and assigned descriptive labels that encapsulate the essence of the keywords within them. This approach not only reveals the dominant themes in the discourse surrounding Penang tourism but also provides insights that can be utilized by stakeholders in the tourism sector to enhance the visitor experience.

Table 1: Results of the topic modelling analysis

Topic No	Top 5 word	Topic Description
0	food, time, try, see, think	Food
1	come, curry, got, street, went	Food
2	hill, town, george, beautiful, know	Tourist spot
3	love, time, weekend, visit, work	Events
4	island, drive, laksa, eat, bridge	Driver attitude and traffic
5	hotel, town, get, photo, post	Hotel
6	café, night, restaurant, road, want	Food

4.2. Sentiment Analysis

Text visualization technique was used to provide a general picture of tourists' overall sentiment towards Penang Island as a travel destination from the 1,032 cleaned data items analysed. The results reveal that the majority of sentiments expressed were positive (46%; n=471), followed by neutral sentiments (39%; n=404), and negative sentiments (15%; n=157), as depicted in Figure 3.

Figure 3: Sentiment Overview

4.2.1 Positive Sentiment

Among the positive sentiments, 471 data items were categorized, as detailed in Table 2. The "Tourist Spot" category emerged as the most frequently discussed, representing 43.5% of the analysed tweets. This dominance emphasizes the significant role of tourist spots in shaping visitors' experiences and perceptions of Penang.

Table 2: Positive Sentiment Categories

Category	Tourist Spot	Food	Event	Hotel	Driver Attitude & Traffic	TOTAL
Frequency	205	193	33	32	8	471
Percentage (%)	43.5%	41%	7.0%	6.8%	1.7%	100

Table 3 provides examples of tweets that reflect this sentiment. One notable tweet, scoring 3.1, stated: "Blessed weekend enjoying Penang beaches ... fall in love with this island ... the hill down to the shoreline, everything is so beautiful," with high sentiment scores attributed to words such as 'blessed' (0.74) and 'beautiful' (0.74). Other highly rated tweets praised popular tourist spots such as Penang Hill, Penang National Park, and the Penang Escape Theme Park, emphasizing users' enjoyment of nature and outdoor activities.

Food emerged as the second most mentioned theme, comprising 41% of the total tweets. This highlights the central role of culinary experiences in Penang's tourism appeal. Table 4 presents tweets reflecting positive sentiments towards dining experiences on the island. One tweet stating, "The best fine dining restaurant in Batu Feringghi. The food was superb," achieved a score of 3.8, with words like 'best' (0.82) and 'superb' (0.79) contributing to its high positive score. Mentions of local dishes such as Nasi Kandar and Penang Laksa further emphasize the importance of food in enriching cultural identity and enhancing visitor satisfaction.

Table 3: Positive Example Tweets from Tourist Spot Category

Tweet	Score	Scoring String	Negativity	Positive
Blessed weekend enjoying Penang beaches. I've totally fallen in love with this Island. From the top of the hill down to the shoreline, everything is so beautiful ??!	3.1	blessed (0.74), enjoying (0.62), love (0.82), top (0.21), beautiful (0.74)	0.0	3.1
Good morning, Penang! How lucky to be able to wake up to this! I'm not knocking those that don't, because every place has its own particular beauty, I'm just celebrating this. I grew up in Malaysia and I'm loving living here now.	3.1	good (0.49), lucky, (0.46), beauty (0.72), celebrating (0.69), loving (0.74)	0.0	3.1

Table 4: Positive Example Tweets from Food Category

Tweet	Score	Scoring String	Negative	Positive
The best fine dining restaurant in Batu Feringghi. The food was superb. Impressed with lovely and romantic atmosphere. They have fresh flowers on every table! Ferringhigardenrestaurant ★★★★★	3.8	best (0.82), fine (0.21), superb (0.79), impressed (0.54), lovely (0.72), romantic (0.44), fresh (0.33)	0.0	3.8
Yes ... nasi kandar always the best in Penang. My children also love nasi kandar. I will try to find some other best place for pasembor. It's because I love pasembor ??	3.7	yes (0.44), best (0.82), love (0.82), best (0.82), love (0.82)	0.0	3.7

The hotel category accounted for 32% of the total tweets analysed, underscores the significant role of accommodations in the overall tourist experience. Table 5 provides examples of tweets expressing positive sentiments regarding hotels in Penang, with the highest score in this category being 3.6 for a tweet about the Eastern & Oriental Hotel. The combination of terms like 'friendly' (0.56), 'amazing' (0.72), and 'recommend' (0.38) reflects the influence of exceptional service and hospitality on enhancing the visitor experience.

Table 5: Positive Example Tweets from Hotel Category

Tweet	Score	Scoring String	Negative	Positive
Wow 5 nights! Yes, staff were very friendly. If you love a staycation and hotel breakfast, next time try Eastern & Oriental Hotel Penang. The hotel is just amazing 🌟 100% would recommend.	3.6	wow (0.72), yes (0.44), friendly (0.56), love (0.82), amazing (0.72), recommend (0.38)	0.0	3.6
I've been to Mauritius, the hotel was fantastic but apart from a beautiful river trip the rest of the Island didn't do it for me. Where I did really enjoy was Penang, the Island and the people were lovely.	2.7	fantastic (0.67), beautiful (0.74), enjoy (0.56), lovely (0.72)	0.0	2.7

The combination of these positive descriptors highlights the importance of exceptional service and hospitality in enhancing the overall visitor experience. Additionally, users often shared sentiments related to their dining experiences at hotels, particularly those at Batu Feringghi, where meals are enjoyed alongside picturesque beach views. Such experiences contribute to a holistic vacation, emphasizing the intersection of accommodation, culinary offerings, and natural beauty. The inclusion of hotels like Hard Rock Hotel, Lexis Hotel, and E&O Hotel in user discussions suggests that these establishments play a crucial role in the tourism landscape of Penang, providing not only lodging but also enriching dining experiences.

Tweets regarding events constituted 7% of the analysed dataset, as summarized in Table 6. Most tweets cantered on user engagement with national celebrations, such as Deepavali and Christmas, as well as major concerts featuring artists like Arashi and TWICE. These events promote community spirit and significantly contribute to tourism in Penang. A notable tweet encapsulated positive sentiments from a concert experience: "For the third screening, I went to Penang... I feel loved," scoring 2.8 due to positive words like 'kind' (0.62) and 'welcoming' (0.49).

Table 6: Positive Example Tweets from Event Category

Tweet	Score	Scoring String	Negativity	Positive
For the third screening, I went to Penang. It's true the phrase I read before that 'because of Arashi, I met so many kind people in my life'???? Arashians just so nice and welcoming. I feel loved. Arashi – MY ARASHI	2.8	wow (0.72), yes (0.44), friendly (0.56), love (0.82), amazing (0.72), recommend (0.38)	0.0	2.8
It was such a wonderful time immersing ourselves in 'The Senses' rainforest concert! Our sincere thanks go to Penang Hill Corporation and The Habitat Penang Hill as well as all the artists for making this happen! We shall see you again!	1.9	wonderful (0.69), sincere (0.44), thanks (0.49), well (0.28)	0.0	1.9

In contrast, the Driver Attitude and Traffic Conditions categories received the lowest positive scores, indicating concerns that may diminish visitor satisfaction. The Driver Attitude category, illustrated in Table 7, included mixed sentiments, with one tweet remarking that "any driver could become an expert driver in Penang." However, the Traffic Conditions category predominantly reflected frustration, exemplified by tweets criticizing local driving experiences, highlighting the need for improved road safety and traffic management.

Table 7: Positive Example Tweets from Driver Attitude Category

Tweet	Score	Scoring String	Negative	Positive
Driving in Penang without a P number plate really feels like a whole Get Out situation	0.5	number (0.08), like (0.38)	0.0	0.5
Myvi driver in Penang just like 'pencilok' expert but using a car not motorcycle ????	0.4	like (0.38)	0.0	0.4

4.2.2 Negative Sentiment

Similar to positive sentiments, negative sentiments were categorized into themes relating to user experiences. A total of 154 data items were classed as negative and allocated to categories as shown in Table 8.

Table 8: Negative Sentiment Categories

Tourist Spot	Food	Driver Attitude & Traffic	Hotel	Event	TOTAL
61	45	31	15	2	154
39.6%	29.2%	20.1%	9.8%	1.3%	100

The "Tourist Spot" category, with 39.6% of the negative sentiment tweets, highlights significant dissatisfaction with certain tourist attractions in Penang. As shown in Table 9, these tweets reflect a range of grievances, often highlighting specific shortcomings encountered during visits. For example, one tweet states, "Still kinda piss that they stop ferry service getting in and out of Penang. I live for that shit since I was a kid." This comment conveys a strong sense of frustration, featuring words with notably negative sentiment scores: 'piss' (-0.44), 'stop' (-0.31), and 'shit' (-0.67). Such expressions not only indicate dissatisfaction but also reflect a deeper emotional connection to the ferry service, which has been a significant part of the user's experience in Penang. Other tweets with high negative scores focus on locations such as the Penang War Museum, Ghost Museum, Penang Hill, and Penang Ferry. Common criticisms include perceptions that these sites are overrated, as portrayed in online advertisements, and concerns about poor service and inadequate facilities. Notably, the discontinuation of the iconic Penang ferry service has provoked strong reactions from users, who lament the loss of a traditional and convenient means of transportation.

Table 9: Negative Example Tweets from Tourist Spot Category

Tweet	Score	Scoring String	Negativity	Positive
A visit to the Penang War Museum today was a sober reminder of the horror of armed conflict. Lest we forget...	-2.0	war (-0.74), horror (-0.69), conflict (-0.33), forget (-0.23)	2.0	0
Still kinda piss that they stop ferry service getting in and out of penang. I live for that shit since I was a kid.	-1.4	piss (-0.44), stop (-0.31), shit (-0.67)	1.4	0

In the Food category, examples of tweets are shown in Table 10. The highest negativity score of -2.1 was attributed to a tweet stating, "we used to have it as a family meal... I absolutely hate it...", with contributing words including "hate" (-0.69), "no" (-0.31) used twice, and "problem" (-0.44). Another tweet received a score of -0.8, expressing frustration with the comment, "damn Penang is so crazy, literally can't dine in fast; all queue like crazy," where the repetition of "damn" (-0.44) and "crazy" (-0.36) influenced the negative sentiment. Although the Food category was anticipated to have the highest frequency of occurrences, it ranked second with 45 instances, compared to 47 in the Tourist Spot category. Users predominantly expressed dissatisfaction with overcrowded restaurants, particularly popular ones known for their exceptional cuisine, citing issues such as long queues, lack of hot food, high prices, and instances of food poisoning.

Table 10: Negative Example Tweets from Food Category

Tweet	Score	Scoring String	Negative	Positive
We used to have it as a family meal throughout the day and it was always eaten hot. I don't know whyyyyy Penang serves this with ice. I absolutely hate it. So I won't eat Bubur Chacha here at all. Melaka, Singapore, even Hua Hin is ok, no problem. Penang Bubur Chacha no. Hate.	-2.1	hate (-0.69), ok (0.31), no (-0.31), problem (-0.44), no (-0.31), hate (-0.69)	2.4	0.3
If I tell you I'm from Penang and your first comment is 'I've visited Penang! And got food poisoning (((('what am I supposed to say. It's not my fault you're born with a weak stomach	-1.6	poisoning (-0.72), fault (-0.44), weak (-0.49)	1.6	0

Conversely, tweets addressing driving experiences yielded more negative sentiment scores than those related to food; 12.3% of the 154 negative sentiment data items pertained to driver attitudes in Penang. The Driver Attitude category highlights unpleasant experiences with local drivers, as detailed in Table 11. Dangerous driving and traffic violations are prevalent concerns, with users expressing frustration over drivers who fail to use indicators or adhere to traffic signals, potentially leading to undesirable outcomes for tourists.

Table 11: Negative Example Tweets from Driver Attitude Category

Tweet	Score	Scoring String	Negativity	Positive
Yes, when driving in Penang, beware of idiots riding motorcycles. They won't stop at traffic light or else, also need to beware of idiotic drivers that make a u-turn at the wrong place.	-1.7	yes (0.44), idiotic (-0.67), stop (-0.31), idiotic (-0.67), wrong (-0.54)	2.2	0.4
Are Penang's drivers originally rude and that's why people call them rude, or are they rude because people call them rude anyway so they might as well	-1.8	rude (-0.51), rude (-0.51), rude (-0.51), rude (-0.51), well (0.28)	2.1	0.3

Some users have reported negative experiences in certain hotels in Penang, as shown in Table 12. Most tweets do not specify hotel names, except for one of the oldest hotels in Georgetown. One tweet combined the words "bad" (-0.64) and "disappointed" (-0.54), resulting in the highest sentiment score in this category, citing issues such as a slow check-in process and a lack of water supply.

Additionally, 6.5% of negative tweets addressed driver attitudes and traffic, with users expressing frustration over being stuck in traffic and experiences of getting lost in Penang. Table 13 presents example tweets expressing negative sentiment regarding the traffic conditions in Penang. The first tweet conveys frustration over the poor traffic situation, receiving a sentiment score of -0.6, with the negative term "bad" (-0.64) dominating the sentiment analysis. The second tweet, while expressing some irritation, points out the inconsistency of complaints about traffic, especially when it is caused by the drivers themselves, leading to a slightly less negative score of -0.4, with "complain" (-0.38) influencing the sentiment.

Table 12: Negative Example Tweets from Hotel Category

Tweet	Score	Scoring String	Negativity	Positive
This is Batu Ferringhi's only luxury boutique hotel by the beach! ??? It's Lone Pine, tucked away in an idyllic spot...	-0.3	lone (-0.28)	0.3	0
Unfortunately, yes. There have been a lot of tweets complaining about the slow check ins from different hotels, due to staff shortages mostly. Royale Chulan Penang is one of them ??	-0.4	Unfortunately (-0.36), yes (0.44), complaining (-0.21), shortage (-0.26)	0.8	0.4

Table 13: Negative Example Tweets from Traffic Condition Category

Tweet	Score	Scoring String	Negativity	Positive
Why traffic so bad this morning in Penang??	-0.6	bad (-0.64)	0.6	0
Eh hello, don't come to Penang and complain about the traffic. YOU'RE the traffic	-0.4	complain (-0.38)	0.4	0

Table 14 showcases example tweets expressing fatigue or disappointment in relation to events. The first tweet illustrates exhaustion from a packed schedule of travel and social events, reflected in the sentiment score of -0.2, with the term "tired" (-0.49) contributing the most to the negative sentiment. The second tweet expresses a mild frustration with the inability to hold offline events due to the COVID-19 pandemic, yielding a sentiment score of -0.1, with the term "offline" (-0.13) affecting the overall sentiment.

Table 14: Negative Example Tweets from Event Category

Tweet	Score	Scoring String	Negativity	Positive
Omg I'm so tired. Landed from Langkawi yesterday, going off to Penang today till Wednesday for work, family staycation Thursday to Saturday, then 2 weddings this weekend. Gosh gais I'm out of energy.	-0.2	tired (-0.49, energy (0.28)	0.5	0.3
[?? MYMYs] Due to Malaysia Covid-19, We are not able to make any offline events for Aespa as we are mostly based in...	-0.1	offline (-0.13)	0.1	0

5. DISCUSSION

This research highlights the significant relationship between cultural events and sustainable tourism in Penang. The positive sentiment expressed in event-related tweets indicates that engaging tourists through community celebrations and cultural activities, such as the Penang International Food Festival or George Town Festival, enhances visitor satisfaction, promotes repeat visits, and encourages positive word-of-mouth. These events not only enrich the cultural experience but also contribute significantly to the local economy. For instance, the George Town Festival, known for its diverse art performances and exhibitions, attracts both local and international visitors, fostering cultural exchange and boosting tourism revenues.

However, negative sentiments regarding driver attitudes and traffic conditions present challenges that can detract from the tourism experience. Addressing these logistical issues is crucial for sustainable tourism. Improvements in infrastructure, such as the expansion of public transportation options, better traffic management systems, and the implementation of public awareness campaigns promoting responsible driving practices, could greatly enhance the visitor experience. For example, initiatives like dedicated tourist bus routes or incentives for the use of electric vehicles could reduce congestion and improve air quality, creating a more welcoming environment for tourists.

The analysis reveals that 46% of tweets about Penang expressed positive sentiments, while only 15% were negative. This favourable perception reflects Penang's status as a premier tourist destination, but it also signals the need for ongoing evaluations of service infrastructure and the tourism environment (Nadeau et al., 2022). Regular assessments will help align local services with evolving visitor expectations, thus contributing to long-term sustainability.

Food quality emerged as a critical theme in the sentiment analysis. While many visitors praised local cuisine, a notable number expressed dissatisfaction, indicating the necessity for consistent

culinary standards, which are vital to travel motivations (Rousta & Jamshidi, 2020; Su, 2018). Penang's food, often regarded as a primary attraction, must meet high expectations to maintain its culinary reputation. This suggests that initiatives aimed at ensuring quality standards across all food service establishments could enhance the overall tourism experience.

The prominence of hotel-related tweets underscores the importance of hospitality providers effectively responding to customer feedback to enhance service quality and foster brand loyalty (Ho et al., 2020; Puh & Bađić Babac, 2023). By leveraging guest reviews and feedback, hotels can improve their services, contributing to a more personalized and positive experience for visitors. Moreover, effective mobility is crucial for a positive tourism experience. Enhancing transportation options will facilitate smoother interactions between tourists and their environment, further contributing to sustainable tourism development (La Rocca, 2015; Le-Klähn et al., 2015; Ali, 2020). This research underscores the interconnectedness of cultural events, food quality, hospitality, and transportation in shaping sustainable tourism in Penang. By addressing identified challenges and enhancing both cultural and logistical aspects, stakeholders can enrich visitor experiences and ensure the long-term sustainability of the region's tourism sector.

6. CONCLUSION

This research offers important insights into tourists' sentiments regarding Penang, emphasizing both its strengths and areas for improvement. Analysing tweets reveals that while many visitors express positive feelings about cultural events, such as the George Town Festival and Penang International Food Festival, and local cuisine, there are notable concerns about transportation and service quality.

Cultural events significantly enhance visitor satisfaction, contributing to sustainable tourism by attracting both domestic and international tourists. For example, the Penang International Food Festival is a major event that not only celebrates local cuisine but also attracts tourists globally, boosting both tourism and the local economy. However, negative sentiments surrounding driver attitudes and traffic conditions highlight the need for better transportation infrastructure and road safety to improve the overall tourism experience. Addressing these logistical challenges, such as implementing effective traffic management systems and promoting public awareness on road safety, would alleviate some of these concerns and create a more enjoyable experience for visitors. Additionally, the dual nature of food experiences in Penang is noteworthy. Local cuisine serves as a primary attraction, yet some visitors express dissatisfaction. This underscores the importance of ensuring consistent quality across food services to meet diverse expectations and reinforce Penang's reputation as a culinary destination. It may be beneficial for the tourism stakeholders to establish food quality certification programs or encourage establishments to adopt standardized practices to maintain high culinary standards.

The research also illustrates the necessity of understanding tourists' emotions and perceptions to improve their travel experiences. By addressing the identified challenges—traffic, service quality, and food consistency—while leveraging the strengths of its tourism offerings, stakeholders can foster a more sustainable tourism environment that benefits both visitors and the local community. In conclusion, this research contributes to the literature on sustainable tourism by utilizing innovative methods to analyse social media content. The findings emphasize the importance of

ongoing evaluation and adaptation in the tourism sector, ensuring that Penang continues to be a vibrant and appealing destination for future travellers. Policymakers and tourism stakeholders should take specific actions based on these findings, including improving transportation infrastructure, ensuring consistent food quality, and supporting cultural events that enhance the overall tourism experience.

This research has several limitations. Firstly, the reliance on Twitter data introduces certain constraints, as it predominantly reflects the opinions of a specific subset of tourists who are active on social media platforms. This may not be representative of the broader tourist population, particularly those who do not engage with such platforms. Furthermore, sentiment analysis is inherently prone to biases, including the misinterpretation of nuanced expressions such as sarcasm or irony, which could potentially distort the accuracy of sentiment classification. In terms of future research, it would be valuable to mitigate these limitations by incorporating data from a wider range of social media platforms, such as Instagram or Facebook, which may offer additional perspectives on tourist sentiment. Moreover, survey-based research could complement the social media analysis, providing direct insights from a more diverse and representative sample of tourists. Further studies could also investigate the differential impact of specific cultural events on various tourist segments, exploring whether domestic and international visitors exhibit divergent attitudes and behaviours toward these events.

ACKNOWLEDGEMENT

We would like to express our sincere thanks to the reviewers for their valuable suggestions to help improve the paper.

REFERENCES

- 12th Malaysian Plan. (2021). A prosperous, inclusive, sustainable Malaysia. 12th Malaysian Plan.
- Ahmed, A. A. A., Agarwal, S., Kurniawan, I. G. A., Anantadjaya, S. P., & Krishnan, C. (2022). Business boosting through sentiment analysis using Artificial Intelligence approach. *International Journal of System Assurance Engineering and Management*, 13(Suppl 1), 699-709.
- Alessia, D., Ferri, F., Grifoni, P., & Guzzo, T. (2015). Approaches, tools and applications for sentiment analysis implementation. *International Journal of Computer Applications*, 125(3), 26–33.
- Ali, B. (2020). Tourism mobility in ecotourism cities: A case study of low urban density. *Journal of Tourism and Hospitality Management*, 8(2), 1-9.
- BBC News. (2020, March 17). *Coronavirus: How are Lockdowns and Other Measures Being Enforced?* <https://www.bbc.com/news/world-51911340>
- Bing, L. (2012). Sentiment analysis and opinion mining (synthesis lectures on human language technologies). University of Illinois: Chicago, IL, USA.
- Cao, Q., Cheng, X., & Liao, S. (2023). A comparison study of topic modeling based literature analysis by using full texts and abstracts of scientific articles: A case of COVID-19 research. *Library Hi Tech*, 41(2), 543-569.
- Cui, Y., Kim, S., & Feng, S. (2023). Exploring success factors of tourism performing arts by analyses of online reviews. *Journal of Hospitality and Tourism Technology*, 14(1), 37-52.

- Department of Statistics Malaysia. (2020). Tourism satellite account. https://www.dosm.gov.my/v1/index.php?r=column/cthemByCat&cat=111&bul_id=dEZ6N0dYUDJEWkVxMzdOalY3UUJSdz09&menu_id=TE5CRUZCblh4ZTZMODZlBmk2aWRRQT09
- Dermawan, A. (2021, October 15). Penang's tourism industry off to a good start. *New Straits Times*. <https://www.nst.com.my/news/nation/2021/10/736779/penangs-tourism-industry-good-start>
- Egger, R., & Yu, J. (2022). Identifying hidden semantic structures in Instagram data: A topic modeling comparison. *Tourism Review*, 77(4), 1234-1246.
- Gumelar, G. K. (2020). Permanent and temporary aspects of urban tourism (Case study: George Town, Malaysia). *International Journal of Tourism and Hospitality Studies*, 10(2), 45-56.
- Hossain, M. S., Begum, H., Rouf, M. A., & Sabuj, M. M. I. (2023). Investigation and prediction of users' sentiment toward food delivery apps applying machine learning approaches. *Journal of Contemporary Marketing Science*, 6(2), 109-127.
- Ho, R. C., Withanage, M. S., & Khong, K. W. (2020). Sentiment drivers of hotel customers: A hybrid approach using unstructured data from online reviews. *Asia-Pacific Journal of Business Administration*, 12(3/4), 237-250.
- Huibin, X., & Marzuki, A. (2012). Community participation of cultural heritage tourism from innovation system perspective. *International Journal of Services Technology and Management*, 18(3-4), 105-127..
- La Rocca, R. A. (2015). Tourism and mobility: Best practices and conditions to improve urban livability. *TeMA-Journal of Land Use, Mobility and Environment*, 8(3), 311-330.
- Lamba, M., & Madhusudhan, M. (2022). Topic modeling, text mining for information professionals: An uncharted territory. In L. P. Webster (Ed.), *Text mining for information professionals* (pp. 105-137). Springer.
- Lebe, S. S. (2015). Spices and Tourism: Destinations, Attractions, and Cuisines. *Annals of Tourism Research*, 54, 1-16.
- Le-Klähn, D. T., Roosen, J., Gerike, R., & Hall, C. M. (2015). Factors affecting tourists' public transport use and areas visited at destinations. *Tourism Geographies*, 17(5), 738-757.
- Li, X., Law, R., Xie, G., & Wang, S. (2021). Review of tourism forecasting research with internet data. *Tourism Management*, 83, 104245.
- Liu, M. W., Zhu, Q., & Wang, X. (2021). Building consumer connection with new brands through rituals: The role of mindfulness. *Marketing Letters*, 32(2), 271-284.
- Martínez-Torres, M. del, Rodríguez-Piñero, F., & Toral, S. L. (2015). Customer preferences versus managerial decision-making in open innovation communities: The case of Starbucks. *Technology Analysis & Strategic Management*, 27(10), 1226-1238.
- Md Saad, N. H., & Yaacob, Z. (2021). Tourism Destination Product Characteristics Based on Twitter Sentiment Analysis: A Case Study of Penang, Malaysia. *Tourism Management and Sustainable Development*, 6(3), 98-112.
- Mehraliyev, F., Chan, I. C. C., & Kirilenko, A. P. (2022). Sentiment analysis in hospitality and tourism: A thematic and methodological review. *International Journal of Contemporary Hospitality Management*, 34(1), 46-77.
- Mirzaalian, F., & Halpenny, E. (2019). Social media analytics in hospitality and tourism: A systematic literature review and future trends. *Journal of Hospitality and Tourism Technology*, 10(4), 764-790.

- Nadeau, J., Wardley, L. J., & Rajabi, E. (2022). Tourism destination image resiliency during a pandemic as portrayed through emotions on Twitter. *Tourism and Hospitality Research*, 22(1), 60-70.
- New Straits Times. (2022, January 3). Penang in CNN Travel's list of 'Where to Travel 2022: The best destinations to go'. <https://www.nst.com.my/news/nation/2022/01/759810/penang-cnn-travels-list-where-travel-2022-best-destinations-go>
- Pamukçu, H., Saraç, Ö., Aytuğar, S., & Sandıkçı, M. (2021). The effects of local food and local products with geographical indication on the development of tourism gastronomy. *Sustainability*, 13(12), 6692.
- Pokok Kelapa. (2016). *Penang Hill Hike via Moon Gate*. <https://pokokkelapa.wordpress.com/2016/09/06/penang-hill-hike-via-moon-gate/>
- Prananda, A. R., & Thalib, I. (2020). Sentiment analysis for customer review: Case study of GO-JEK expansion. *Journal of Information Systems Engineering and Business Intelligence*, 6(1), 1-8.
- Puh, K., & Bagić Babac, M. (2023). Predicting sentiment and rating of tourist reviews using machine learning. *Journal of Hospitality and Tourism Insights*, 6(3), 1188-1204.
- Qi, Y., & Shabrina, Z. (2023). Sentiment analysis using Twitter data: a comparative application of lexicon-and machine-learning-based approach. *Social network analysis and mining*, 13(1), 31.
- Rambocas, M., & Pacheco, B. G. (2018). Online sentiment analysis in marketing research: A review. *Journal of Research in Interactive Marketing*, 12(2), 146-163.
- Rousta, A., & Jamshidi, D. (2020). Food tourism value: Investigating the factors that influence tourists to revisit. *Journal of Vacation Marketing*, 26(1), 73-95.
- Spennemann, D. H., & Whitsed, R. (2023). The impact of COVID-19 on the Australian outdoor recreation industry from the perspective of practitioners. *Journal of Outdoor Recreation and Tourism*, 41, 100445.
- Su, N. D. (2018). *Travel Motivation and Behavior Toward Visiting a Food Tourism Destination* [Unpublished doctoral thesis]. Swinburne University of Technology.
- The EdgeProp. (2021, January 22). *Malaysia hotel industry suffers RM6.53b losses last year – MAH*. <https://www.edgeprop.my/content/1850913/malaysia-hotel-industry-suffers-rm653b-losses-last-year-mah>
- The Star. (2017, January 9). *Penang among must-visit destinations this year in CNN list*. <https://www.thestar.com.my/news/nation/2017/01/09/cnn-recognises-penang-as-among-17-must-visit-destinations-this-year/>
- The Star. (2019, June 4). *CNN Travel: Penang among best destinations for ultimate Asia experience*. <https://www.thestar.com.my/news/nation/2019/06/04/cnn-travel-penang-among-best-destinations-for-ultimate-asia-experience/>
- Trinh Ngo, T.T., Phuong Bui, L.A., Dung Pham, T.H., Tram Mai, N. and Luan Bui, K. (2022), “Social media research trends in tourism city context”, *International Journal of Tourism Cities*, Vol. 8 No. 3, pp. 779-798.
- Tourism Malaysia. (2022). *Tourist arrivals and receipts in Malaysia by year 2022*. <https://www.tourism.gov.my/statistics>
- United Nations World Tourism Organization (UNWTO). (2020a). *UNWTO Briefing Note – Tourism and COVID-19, understanding domestic tourism and seizing its opportunities*. <https://www.eunwto.org/doi/pdf/10.18111/9789284422111>

United Nations World Tourism Organization (UNWTO). (2020b). *100% of global destinations now have COVID-19 related travel restrictions*. <https://www.unwto.org/news/100-global-destinations-now-have-covid-19-related-travel-restrictions>