Foundation University Journal of Engineering and Applied Sciences

FUJEAS
Vol. 4, Issue 1, 2023.
DOI:10.33897/fujeas.v4i1.785

Research Article

Benchmarking Travelling Reviews using Opinion Mining

Bushra Kanwal [a], Muhammad Rizwan Rashid Rana [a, *], Asif Nawaz [a], Anum Nawaz Kiani [b]

[a] University Institute of Information Technology, PMAS Arid Agriculture University, Rawalpindi, Pakistan

[b] Department of Computer Science, Rawalpindi Women University, Rawalpindi, Pakistan

* **Corresponding author**: rizwanrana315@gmail.com

Article Citation:
Kanwal et al. (2023). "Benchmarking Travelling Reviews using Opinion Mining". *Foundation University Journal of Engineering and Applied Sciences*
DOI:10.33897/fujeas.v4i1.785

Published by Foundation University Islamabad.
Web: https://fui.edu.pk/

Abstract:

Online travel reviews offer valuable data, yet it remains uncertain if those most influenced by these reviews actually read them. This research aims to uncover consistent patterns and explain variations in online travel ratings, comments, and reviews. To accomplish this, millions of reviews were collected from Pakistan's top online travel companies, Uber and Careem. Utilizing semantic affiliation analysis, subject terms were extracted, forming a semantic affiliation structure. The findings highlight significant differences among channels concerning topical vocabulary, subject distribution, structural traits, and community links. The network visualization results are particularly noteworthy, as they illustrate connections between key concepts and words within each topic, making them easily understandable. With the proposed logical method, we can better understand the strategic snafus in the travel sector and gain fresh insights into how to dig up popular assessments to better serve tourists, lodging establishments, and trade groups.

Keywords: Decision Opinion, Travel Reviews, Visualization, Social Network Association, Travel Industry.

1. Introduction

Whether a traveler is in the midst of a trip, just beginning to plan one, or has already returned home from their adventure, the Internet is a valuable resource that may help them in a number of ways. [1]. User-generated content (UGC) is a relatively new sort of travel database. You may monitor tour-goers' comments, profiles, and photos from their trips right here [2]. User-generated content (UGC) facilitates a traveler's ability to share his or her thoughts and impressions on a variety of social media sites, including some that aren't specifically geared toward that purpose. This material features a traveler's comments about his or her impressions of a certain location or travel-related service. When planning trips to uncharted or hazardous destinations, numerous tourists rely on the abundance of online user-generated content (UGC) to seek advice and insights shared by fellow travelers [3].

The rise of new Internet technologies has led to the widespread use of user reviews, which are particularly valuable in the travel and hotel sectors. These reviews offer essential insights and guidance for travelers, aiding them in making informed decisions. They have become a crucial resource for individuals seeking authentic and reliable information about destinations and accommodations. Online travel surveys in the form of dynamic client-shared recordings of explorers' interactions, such as those seen on Trip Advisor [4], have been widespread for some time. The decision to use a website has shifted from read-only to write-only [5]. In recent years, online testimonials have risen in prominence. Companies place a premium on customer reviews and ratings because they are used as a reference by potential buyers [6]. When culled and studied in light of one's specific problems, the wealth of knowledge contained in online reviews and comments can be invaluable [7].

The advent of social media and user-generated content has revolutionized communication and knowledge-sharing among individuals. Electronic word of mouth (eWoM) has become a key factor in shaping contemporary advertising strategies, allowing consumers to influence each other's decisions. In the travel and hospitality industries, the term "traveler-generated content" (TGC) emphasizes the significance of user-contributed information in guiding potential travelers. Internet reviews, ratings, travel blogs, and vlogs serve as valuable resources for tourists seeking authentic and reliable insights about destinations. TGC empowers travelers with first-hand experiences, enabling them to make informed choices and plan their trips more effectively. This shift towards peer-to-peer information exchange has transformed how people approach travel decision-making and has further bridged the gap between consumers and businesses.

For the goal of gauging customer satisfaction, several travel-related websites gather the opinions and feedback of past customers on a variety of products, activities, attractions, locations, and services. In the future, travelers will benefit from the wealth of information available online to enhance their trip planning [8]. Travel websites actively promote user-generated content, including trip stories, reviews, photos, and more, providing modern vacationers with easy access to valuable insights. However, the sheer volume of data on these platforms can overwhelm customers, leading to a growing need for improved data organization and user-friendly interfaces [9].

Researchers have delved into the impact of website design on user rating behaviors and review management, with a focus on popular platforms like the Chinese travel website qunar.com [10]. The user text reviews involve two critical factors: a) Reviewer-specific information, such as the public availability of reviewer names, their expertise, and the company's reputation in the community, and b) Feedback and reviews, encompassing anticipated satisfaction and review adaptation, which provides both quantitative (average rating and response count) and qualitative (reviewer thoughtfulness) data [11]. A significant challenge in this context lies in the analysis of the tone of online reviews, as they reflect the emotions and opinions of individuals, often ranging from positive to negative or neutral [12]. With the advent of natural language processing techniques, researchers are continually seeking innovative ways to interpret and extract valuable insights from this vast trove of textual data. By better understanding and harnessing user-generated content, the travel industry can offer improved services, while travelers can make more informed decisions, ensuring memorable and satisfying travel experiences.

Tourists are swayed by the opinions of others before making a purchase, and research shows that they value hearing about the experiences of others [4, 13–15]. Over 77% of prospective tourists wait until they read online reviews before making a booking [16]; this is according to survey data gathered by businesses. Exploring knowledge nodes (such as company travel websites or online travel evaluations) might help tourists reduce their sense of uncertainty and gain useful indirect knowledge for making purchases, leading to a more positive psychosomatic experience [17], [18]. While internet consumer reviews are helpful and often more relevant in the information search, they come at a greater intellectual cost. Consumers can get confused and disoriented when confronted with a deluge of online reviews, leading to a loss of sense of agency and an increase in theoretical tension [19]. The tourism organization can affect both the quality of online reviews and client attitudes toward review information [13].

The key variables that determine the excellence of information collection are size, accuracy community qualities, and verbal eloquence of evaluations [19]. Items were primarily highlighted through the use of words, though certain physical components were also included. Difficult-to-understand wording in evaluations is ignored by most customers [20], which can hurt a business's ability to compete in the tourism industry. (Such as popularity and earnings) Customer perceptions of the usefulness of online review information in making travel decisions are affected by the accuracy of such reviews [14], [18], [21]. According to [20], if consumers find online evaluations difficult to understand, they are less likely to make a reservation, which has an impact on the competitiveness of tourism enterprises (e.g., in for example, repute and profits).

Therefore, it has become a serious issue in both the realm of practice and the realm of academia to efficiently extract high-quality facts from a mountain of data culled from internet evaluations. Surveys [21], Computer simulations [22], and a quantitative analysis [21], [23]. Previous studies on online vacation reviews have made use of grounded hypotheses [20], [24]. Furthermore, textual online reviews often incorporate extensive data sources exhibiting logical capabilities akin to cultural econometric and statistical methodologies.[25]. Finding statistically significant variations between groups in the review data is challenging [26], [27].

Words that are central to a topic can be quickly extracted from a large corpus of texts via semantic association exploration. OTAs, or online travel agencies, are increasing in popularity in today's technologically advanced world [24]. Before these technological advances, people used to be able to swiftly and easily get ready thanks to them. Business success for online travel agencies requires semantic analysis of associations, and this can be accomplished by making use of the plentiful concept tree structure and semantic data provided. [26], [27], [26].

By using a semantic association network, the tourism association can even out the dispersion in the distribution of thematic word contents, organizational plots, and social relationships [28]. The primary goal is to determine the most frequently used terms within the dataset [29]. In order to do this, we used the Jieba toolbox to pick out the overarching themes and cull the dataset of duplicate or unhelpful evaluations. A Python toolbox called Jieba was released specifically for the task of segmenting Chinese words [28]. This technique saw extensive use for manual mining. The NLTK was also utilized to determine how often particular theme-related words were used as shown in Eq. 1.

$$Fi = Ri\ X\ \left(\frac{Li}{Lt}\right) \tag{1}$$

For any given thematic word I let Fi stand for its occurrence in the datasets, Ri for the total number of reviews, Li for the size of I, and Lt for the total size of review words.

Semantic linking of two archives in online travel reviews is typically determined by insistent emphasis as well as the significance rationale behind the issue. Applying a combination of semantic strategies to the study of online trip evaluations has been shown to reduce the test predisposition of topic-identifiable evidence to the word quantity. Thus, it can greatly improve the efficiency and quality of analysis of online travel reviews by selectively omitting the thematic words and association words. The value of the data can be increased when it is transformed into information, and finally, insight [29].

1.1. Research Contributions

Findings from this study will be used by online travel agencies to improve the service they provide to riders by incorporating feedback from customers into their operations.

Conceptual contributions could involve such things as:

a) Advised an efficient means of tracking down travel-related problems.
b) Using feedback from customers, the suggested model will pinpoint travel-related issues and guide improvements in infrastructure design.
c) The proposed methodology will aid in pinpointing the source of the issue and examining how much time and effort will be needed to rectify the situation.
d) The goal is to create a model that specifies how to tackle what will have the most impact and then to evaluate the results of those adjustments.

The remainder of this study is structured as follows: Section 2 presents a foundational understanding of research work linked to online reviews, and extraction of expert opinion from online reviews. In Section 3, we detail our planned study approach, including the procedures we will follow to gather information and ensure its accuracy. Results and analysis are presented in Section 4. Discussion, findings, limitations, and suggested directions for the future are presented in Section 5.

2. Related Work

This section presents the literature on the topic of opinion mining and online travel reviews.

2.1. Online Travel Reviews

Consumers who have done their research and are ready to make a purchase will often submit reviews and ratings online [30]. Internet reviews are both a means of gathering information upon which to base decisions and the driving force behind the vacation diligence decisions that are ultimately made. Judges' expressed feelings on matters of detail or procedure can influence the behavior of other visitors [31]. Features of internet reviews including completeness [32], professionalism [33], quality [34], and character [35] significantly impact product sales [36] for businesses and merchants.

2.2. Introduction for Online Travel Reviews Used for Opinion Mining

Opinion mining is a technology that automatically categorizes online comment data via text-based analysis, including encoding and regular language processing. It delves into people's examinations, thoughts, and emotions about groups, substances, people, problems, pursuits, and characteristics [37]. The method uses an ordered strategy to dissect web-based audits in order to extract textual data from item reviews (for example, the sack of words). When applied to the context of online vacation surveys, the proposed assessment mining method has proven to be both accurate and efficient at handling unstructured audit texts, both of which are invaluable. [4], [39] used Inactive Dirichlet Analysis (IDA) opinion extraction methods to sift through millions of reviews taken from travel sites in order to identify the top concerns of riders and reveal the importance of well-regulated scopes for travel companies to effectively manage their interactions with passengers. Sentiment analysis, an application of natural language processing, has seen a dramatic uptick in popularity over the past decade [40]. Sentiment analysis is referred to by a variety of different names in the academic literature i.e. opinion mining.

2.3. Semantic Association Analysis and SNA

The investigation of the brain's reaction to anticipated arguments confirmed that this is a significant analytic mechanism [41]. Technology that processes natural language is one popular design trend, as are content prototypes. Bigram co-occurrence may aid in preventing data distortion and damage during verbal assessment data aggregation and calculation [43]. When using external data and a semantic relation to build a model of point arguments, semantic analysis allows for a more effective textbook bracket than plant-based models. This strategy has been well-established in the context of business and earthly performance intelligence.

3. Method

In this section, the proposed methodology has been discussed.

3.1. Research Design

As shown in Figure 1, which provides a detailed account of the procedure followed in this study, a semantic association analysis approach was supplied as a means of classifying riders' potential requests based on internet travel reviews and expanding riders' delight.

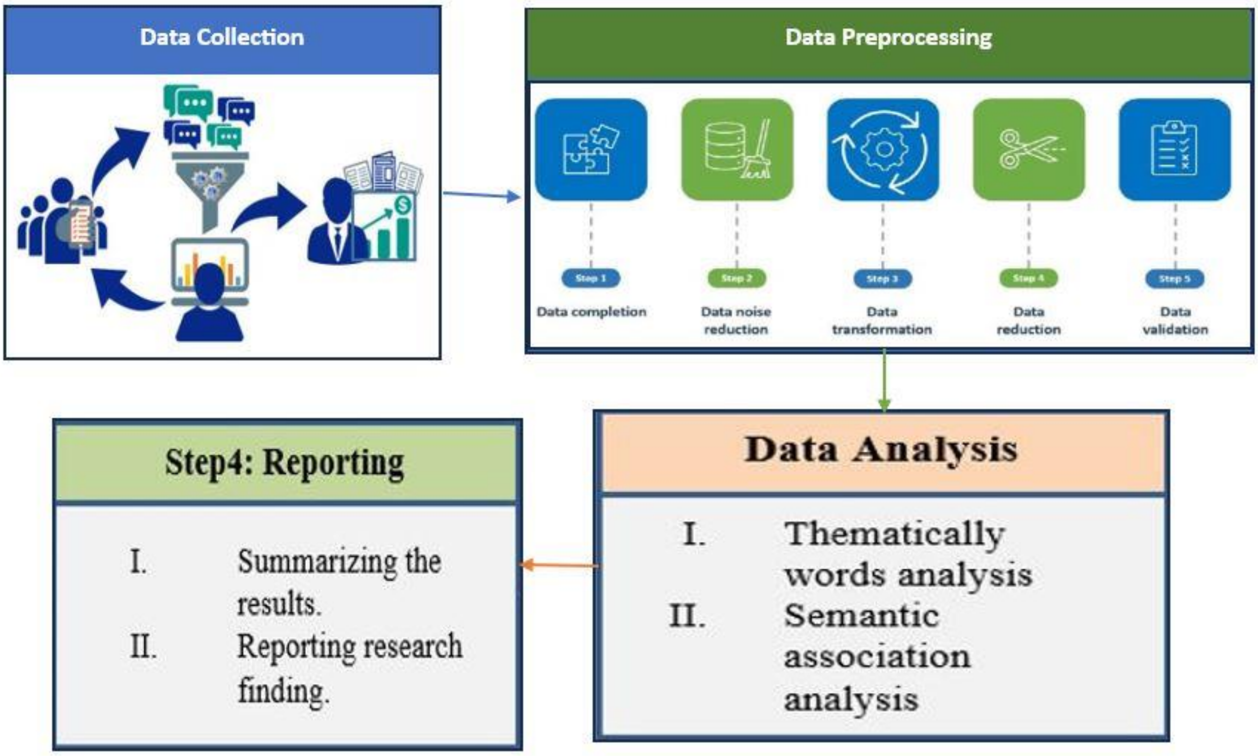

Figure 1: A semantic-based framework for reviews

3.2. Data Collection

Prior studies on online travel reviews have primarily concentrated on specific domains. In contrast, this research centers on two prominent platforms, Uber and Careem, as the primary data sources. Pakistan's top two businesses in terms of client growth in the ridesharing industry are Careem and Uber. Two benchmark datasets are used in this research, one containing the reviews of Careem and the second containing the reviews of Uber. The details are shown in Table 1. The snapshot of these reviews is shown in Figure 2.

Table 1: Dataset details

#	Dataset	URL
1	Cream	https://www.kaggle.com/code/mahmoudeldesuky/eda-for-careem
2	Uber	https://www.kaggle.com/code/hershyandrew/uber-reviews-text-analysis

We also use the reviewers' rating data. Analysis of tourist spots in 24 broad classifications, as displayed in Figure 3. Resorts, beaches, parks, museums, churches, shopping centers, zoos, restaurants, art galleries, swimming pools, clubs, bakeries, salons, cafes, viewpoints, gardens, etc. all have 1–5 star ratings, and the average user rating is determined for each category.

3.3. Data Preprocessing

The reviews obtained from the platforms of two nomadic firms, Uber and Careem, underwent three main processes: data cleaning, tokenization, and the removal of stop words and duplicates. Data cleaning is essential to ensure that only relevant and valuable information related to riders is retained. This process involves identifying and eliminating errors or unnecessary data from the dataset [4]. Common examples of such information include misspelled words and jargon. Stop words are removed

	comment	rating
0	Mohammad harun, he is an awesome guy very info...	5
1	amazing guy gaurav was, so patience and kind. ...	5
2	Gaurav was very knowledgeable and very helpful...	5
3	I called them regarding my flight cancellation...	5
4	VERY GOOD SERVICE BY GAURAV LOHAT SERVED AS SO...	4

Figure 2: User review

user	churches	resorts	beaches	parks	theatres	museums	malls	zoo	restaurants	pubs/bars	...	art galleries	dance clubs	swimming pools	gyms	bakeries	beauty & spas	cafes	view points	monuments	gardens
User 1	0.0	0.0	3.63	3.65	5.0	2.92	5.0	2.35	2.33	2.64	...	1.74	0.59	0.5	0.0	0.5	0.0	0.0	0.0	0.0	0.0
User 2	0.0	0.0	3.63	3.65	5.0	2.92	5.0	2.64	2.33	2.65	...	1.74	0.59	0.5	0.0	0.5	0.0	0.0	0.0	0.0	0.0
User 3	0.0	0.0	3.63	3.63	5.0	2.92	5.0	2.64	2.33	2.64	...	1.74	0.59	0.5	0.0	0.5	0.0	0.0	0.0	0.0	0.0
User 4	0.0	0.5	3.63	3.63	5.0	2.92	5.0	2.35	2.33	2.64	...	1.74	0.59	0.5	0.0	0.5	0.0	0.0	0.0	0.0	0.0
User 5	0.0	0.0	3.63	3.63	5.0	2.92	5.0	2.64	2.33	2.64	...	1.74	0.59	0.5	0.0	0.5	0.0	0.0	0.0	0.0	0.0

Figure 3: User rating

using the Python Natural Language Toolkit and the panda's library is used to identify and remove duplicate rows from the review's dataset. In this study, we discarded any rider reviews that were less than 10 words long before proceeding. The minimal text size of reviews can have an impact on how they are perceived [45].

4. Analysis of Outcomes

The section delves into a detailed examination of outcome analysis. It explores and scrutinizes the results and findings of the study.

4.1. Statistical Analysis of Thematical Words

This study established the definition of manual content analysis. At the outset, we merged the top 50 theme words from two systems and established their order. Three e-commerce experts came to a consensus on how to classify the designated thematic words, agreeing that words with similar connotations should be grouped together in specific classes based on an examination of their relationships and logical hierarchy.

Table 2 reveals a striking similarity in the four most common bigram co-occurrence terms between Uber and Careem, such as "horrible accident," "ride explanation," and "ride satisfaction." These findings suggest that passengers highly prioritize well-prepared trips and dependable transportation services.

Figure 4 presents a comparison of four bigram co-occurrence phrases on the ridesharing platforms, Careem and Uber. These are "Horrible-Accident," "Ride-Explanation", "Ride-Satisfaction", and "Guide-Happy". The X-axis displays the bigram phrases, while the Y-axis illustrates their usage frequency. This

Table 2: The Bigram co-occurrence phrases on the Cream and Uber

Bigram Co-occurrence Phrases	Uber	Careem
horrible-accident	25	35
ride–explanation	995	853
ride–satisfaction	565	610
Guide-happy	570	590

visualization allows for a clear understanding of the relative prevalence and differences in the utilization of these phrases on both platforms.

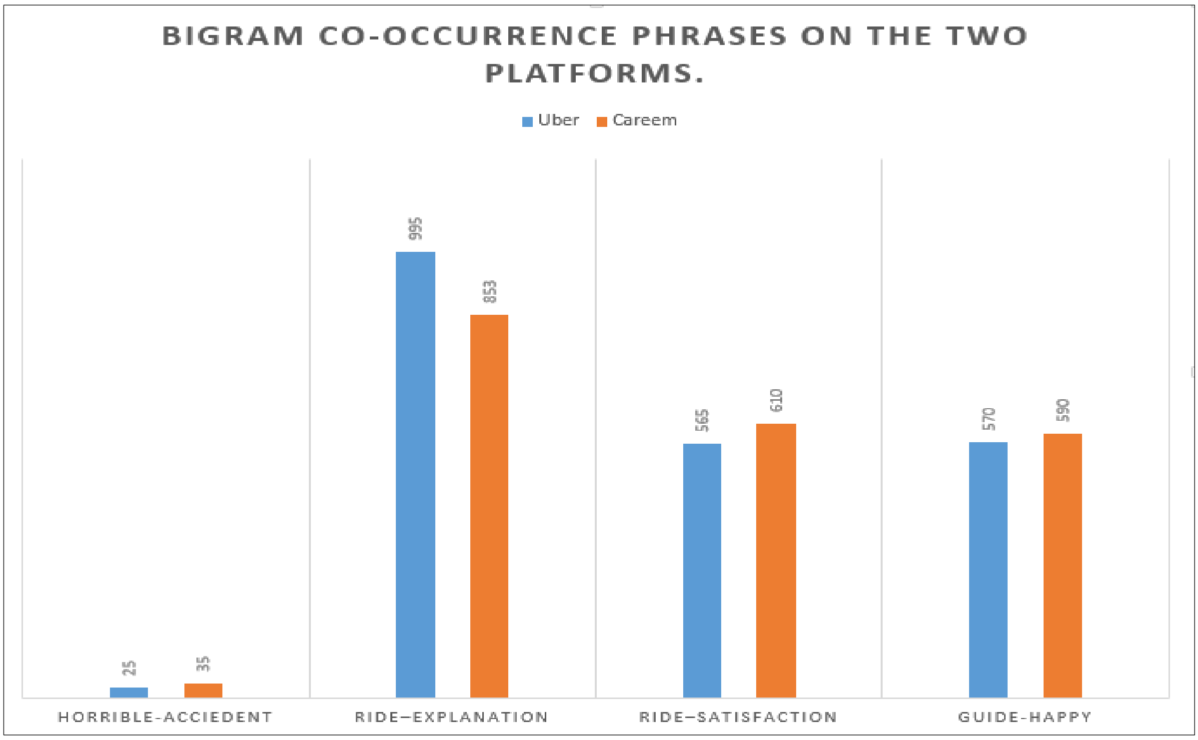

Figure 4: The bigram co-occurrence phrases

4.2. Analysis of Rating Reviews

The average star rating is between 0.5 and 5. There is a large variation in reviews for some venues (such as restaurants and bars). The most popular place that all tourists should go to. On the other hand, places like gyms, bakeries, and swimming pools tend to get bad marks from visitors. Our main goal was to divide users into several groups according to their preferences. K-Means clustering offers advantages over other algorithms by providing simplicity, efficiency, and interpretability, allowing businesses to quickly identify clusters of similar reviews and extract actionable insights to improve products or services with minimal computational complexity. We use the K-Means clustering algorithm on four different case studies and evaluate the outcomes.

a) Cluster analysis using K-means on the raw data (24 features).
b) K-Means Clustering with the original data that has been scaled using the Standard Scalar (24 features with scaled).

c) Clustering using K-means on a principal component analysis
d) Clustering using K-means on the principal component analysis scaled data.

Figure 5 shows the K-Means clustering result in 4 clusters (segments) of users as follows:

a) Users who enjoy cultural activities and the great outdoors make up Cluster #0 (Green).
b) Cluster 1 (Orange): Artistic souls who enjoy fast food, who are aware of the hotel and juice bar scene, and who appreciate design.
c) Cluster 2 (Light Blue): The Activity Seeker. People who fall into this category are always on the lookout for new and exciting things to do, and they frequently use services like those provided by the zoo, shopping centers, restaurants, and bars in their area.
d) The third cluster (pink) shows members are not evenly dispersed; they may or may not have similar interests.

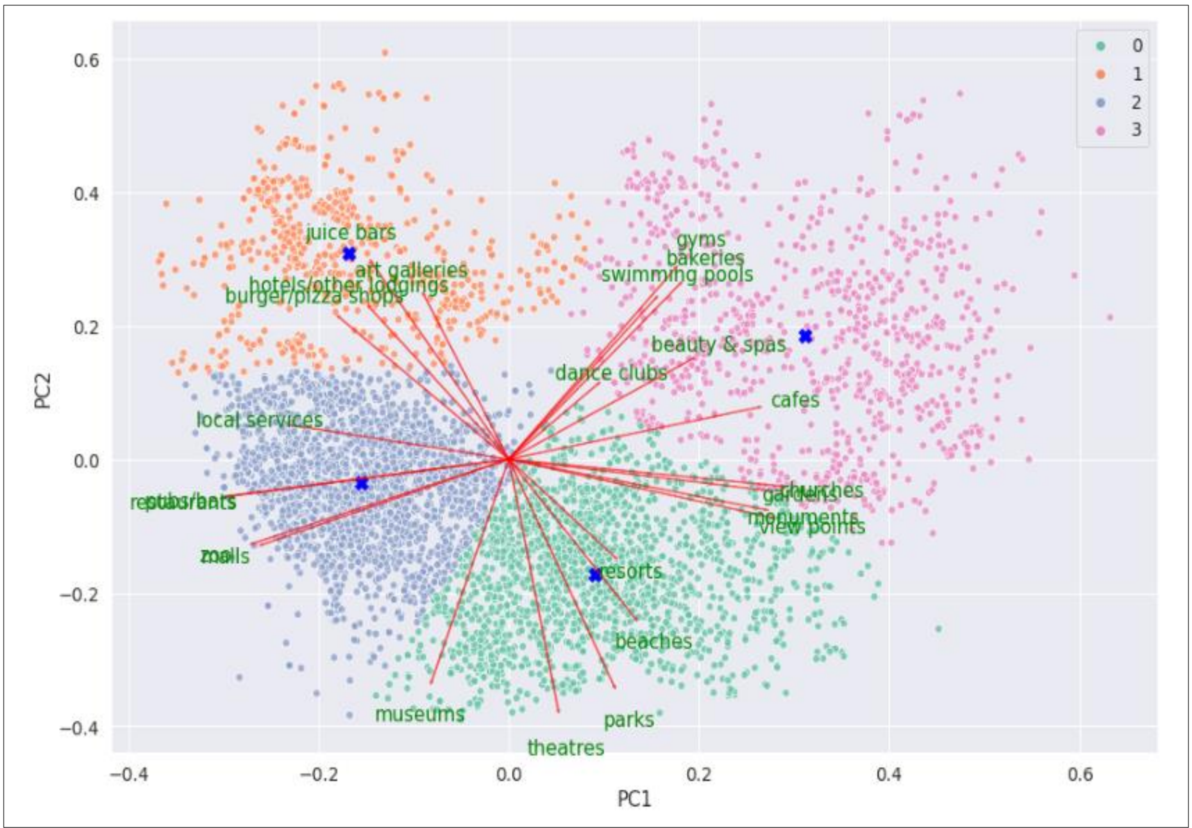

Figure 5: Clustering result

Figure 6 is a bar chart depicting the median user rating within each group. For the reasons stated above:

a) Cluster #0: those who spend the most time at the beach, in the park, at the theatre, and at museums
b) Group #1 is the most enthusiastic about fast-food restaurants, hotels, juice bars, and museums.
c) Cluster 2: Favorites include shopping centers, amusement parks, eateries, watering holes, and neighborhood businesses.
d) Users in the third cluster gave an average rating of two stars across the board, although they seemed to place a greater emphasis on nightclubs, swimming pools, gyms, bakeries, beauty and spas, cafes, viewpoints, monuments, and gardens than those in the first three groups.

Therefore, as depicted in Figure 6, Cluster #3 can be summed up as an individual who values both physical health and travel.

Figure 6: Median user rating

5. Discussion and Conclusion

5.1. Major Finding

This research brings a fresh and vital dimension to the study of online vacation reviews, aiming to explore the extent to which internet tourist reviews can be effectively acknowledged. Understanding the impact and credibility of these reviews is of utmost importance in influencing decision-making within the tourism sector. As the digital landscape continues to shape travelers' choices, online reviews have become a crucial source of information for potential tourists.

The study delves into the value and reliability of internet tourist reviews, shedding light on their influence on consumer decision-making. By analyzing the content, sentiment, and patterns within these reviews, researchers can uncover valuable insights that can aid businesses in understanding customer preferences and expectations. Moreover, this research has implications for improving the overall quality of services and products in the tourism industry.

Acknowledging the significance of Internet reviews empowers businesses to address customer concerns and enhance their offerings, ultimately leading to higher levels of customer satisfaction. This can positively impact the reputation and competitiveness of tourism establishments. Additionally, understanding the dynamics of online reviews can guide tourism organizations in developing effective marketing and communication strategies to engage with their audience more effectively.

This article's primary goal is to identify common themes and provide examples of how these themes are contrasted in online travel evaluations and comments. Topical words were extracted using a semantic association analysis from millions of reviews collected from two major online travel agencies in Pakistan. By incorporating concepts from social network theory into our study, we are able to expand our scope beyond the bounds of previous information. Our study successfully classifies online travel review topics and the social connections generated by those topics. Our research suggests that there are two distinct types of employees for ridesharing services like Uber and Careem: guides and riders.

5.2. Implication for Research

This research introduces a novel theoretical framework that explores user-generated content in the context of online travel review websites, incorporating the influence of social networks to expand on previous studies. The findings demonstrate that internet travel reviews serve as a reflection of the social connections between reviewers and travelers, revealing a form of social attachment. The study highlights how the interpersonal relationships between reviewers and riders impact the tone and subject matter of these reviews. The term "social network" encompasses various types of connections, be it formal or informal, local or global, direct or indirect. By analyzing theme words from online travel reviews, this research traces semantic connections between phrases, shedding light on the underlying dynamics of user interactions and the formation of opinions within the travel community. This framework offers valuable insights for understanding the intricacies of user-generated content and its role in shaping perceptions and decision-making in the travel industry.

5.3. Limitations and Future Work

The present study, despite its limitations, holds significant value for researchers in the realm of semantic association analysis. By utilizing data from the prominent ridesharing platforms Uber and Careem in Pakistan, lays a foundation for future investigations in this area. To enhance its practicality, researchers should expand the study's scope to include non-local factors like travel destinations and cultural norms. Additionally, extending the research timeframe beyond 2015 would provide a more comprehensive understanding and further validate the study's conclusions. While the semantic network structure was based solely on online traveler reviews, future research could benefit from considering the practicality and framework of tourism products, which would bolster its credibility and generalizability, making it a valuable resource for the broader travel industry.

6. References

[1] Z. Liu and S. Park, "What makes a useful online review? implication for travel product websites," *Tourism Management*, vol. 47, pp. 140-151, 2015.

[2] P. Blackshaw and M. Nazzaro, "Consumer-generated media (CGM) 101 word-of-mouth in the age of the Web-fortified consumer," New York: BuzzMetrics, 2004. http://www. nielsen-online. com/downloads/us/buzz/nbzm_wp_CGM101. pdf

[3] C. Cox, S. Burgess, C. Sellitto and J. Buultjens, "The role of user-generated content in tourists' travel planning behavior," *Journal of Hospitality Marketing & Management*, vol. 18(8), pp. 743-764, 2009.

[4] Y. Guo, S. J. Barnes and Q. Jia, "Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation," *Tourism Management*, vol. 59, pp. 467–483, 2017.

[5] M. R. R. Rana, S. U. Rehman, A. Nawaz, T. Ali and M. Ahmed, "A Conceptual Model for Decision Support Systems Using Aspect Based Sentiment Analysis," In *Proceedings of The Romanian Academy Series A-Mathematics Physics Technical Sciences Information Science*, Editura Academiei Romane, 2021, vol. 22(4), pp. 381-390.

[6] M. R. R. Rana, S. U. Rehman, A. Nawaz, T. Ali, A. Imran, A. Alzahrani and A. Almuhaimeed, "Aspect-Based Sentiment Analysis for Social Multimedia: A Hybrid Computational Framework,". *Computer Systems Science & Engineering*, vol. 46(2), pp. 2415-2428, 2023.

[7] H. Jangid, S. Singhal, R. R. Shah and R. Zimmermann, "Aspect-based financial sentiment analysis using deep learning," In *Companion Proceedings of the The Web Conference,* April 2018, pp. 1961-1966.

[8] E. Marine-Roig, "Content analysis of online travel reviews," In *Handbook of e-Tourism*, Cham: Springer International Publishing, 2022, pp. 1-26.

[9] B. Pan and D. R. Fesenmaier, "Online information search: vacation planning process," *Annals of Tourism Research*, vol. 33(3), pp. 809-832, 2006.

[10] Z. Zhang, Z. Zhang and Y. Yang, "The power of expert identity: How website-recognized expert reviews influence travelers' online rating behavior," *Tourism Management*, vol. 55, pp. 15-24, 2016.

[11] Z. Liu and S. Park, "What makes a useful online review? Implication for travel product websites," *Tourism Management*, vol. 47, pp. 140-151, 2015.

[12] R. Kumar, H. S. Pannu and A. K. Malhi, "Aspect-based sentiment analysis using deep networks and stochastic optimization," *Neural Computing and Applications,* vol. 32, pp. 3221–3235, 2020.

[13] Y. H. Hu, Y. L. Chen and H. L. Chou, "Opinion mining from online hotel reviews – a text summarization approach". *Information Processing & Management*, vol. 53(2), pp. 436–449 ,2017.

[14] M. Schuckert, X. Liu and R. Law, "Hospitality and tourism online reviews: Recent trends and future directions," *Journal of Travel & Tourism Marketing*, vol. 32(5), pp. 608–621, 2015.

[15] A. García-Pablos, M. Cuadros and M. T. Linaza, "Automatic analysis of textual hotel reviews," *Information Technology & Tourism*, vol. 16(1), pp. 45–69, 2016.

[16] Q. Ye, R. Law and B. Gu, "The impact of online user reviews on hotel room sales," *International Journal of Hospitality Management*, vol. 28(1), pp. 180–182, 2009.

[17] T. Lian and C. Yu, "Representation of online image of tourist destination: A content analysis of huangshan," *Asia Pacific Journal of Tourism Research*, vol. 22(10), pp. 1063-1082, 2017.

[18] J. Li, L. Xu, L. Tang, S. Wang and L. Li, "Big data in tourism research: A literature review," *Tourism Management*, vol. 68, pp. 301–323, 2018.

[19] S. K. Lakshmanaprabu, K. Shankar, D. Gupta, A. Khanna, J. J. Rodrigues, P. R. Pinheiro and V. H. C. de Albuquerque, "Ranking analysis for online customer reviews of products using opinion mining with clustering," *Complexity*, Article ID 3569351, pp. 1–9, 2018.

[20] A. Papathanassis and F. Knolle, "Exploring the adoption and processing of online holiday reviews: A grounded theory approach," *Tourism Management*, vol. 32(2), 215–224, 2011.

[21] M. Schuckert, X. Liu and R. Law, "A segmentation of online reviews by language groups: How English and non-English speakers rate hotels differently," *International Journal of Hospitality Management*, vol. 48, pp. 143–149, 2015.

[22] Z. Xiang, Z. Schwartz, J. H. Gerdes Jr. and M. Uysal, "What can big data and text analytics tell us about hotel guest experience and satisfaction?," *International Journal of Hospitality Management*, vol. 44, pp. 120–130, 2015.

[23] P. Racherla and W. Friske, "Perceived 'usefulness' of online consumer reviews: An exploratory investigation across three services categories," *Electronic Commerce Research and Applications*, vol. 11(6), pp. 548–559, 2012.

[24] R. Annisa and I. Surjandari, "Opinion Mining on Mandalika Hotel Reviews Using Latent Dirichlet Allocation," *Procedia Computer Science*, vol. 161, pp. 739–746, 2019.

[25] R. Filieri, "What makes an online consumer review trustworthy?," *Annals of Tourism Research*, vol. 58, pp. 46–64, 2016.

[26] S. Boo and J. A Busser, "Meeting planners' online reviews of destination hotels: A twofold content analysis approach," *Tourism Management*, vol. 66, pp. 287–301, 2018.

[27] Z. Xiang, Q. Du, Y. Ma and W. Fan, "A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism," *Tourism Management*, vol. 58, pp. 51–65, 2017.

[28] Z. Hou, F. Cui, Y. Meng, T. Lian and C. Yu, "Opinion mining from online travel reviews: A comparative analysis of Chinese major OTAs using semantic association analysis," *Tourism Management*, vol. 74, pp. 276-289, 2019.

[29] J. Li, L. Xu, L. Tang, S. Wang and L. Li, "Big data in tourism research: A literature review," *Tourism Management*, vol. 68, pp. 301–323, 2018.

[30] S. W. Litvin, R. E. Goldsmith and B. Pan, "Electronic word-of-mouth in hospitality and tourism management," *Tourism Management*, vol. 29(3), pp. 458–468, 2008.

[31] W. Wang and L. Wang, "Merger case analysis of tongcheng and yilong, two online travel agents," *Business & Economy*, vol. 13(5), pp. 93–113, 2018.

[32] R. Chen, C. Y. Liang, W. C. Hong and D. X. Gu, "Forecasting holiday daily tourist flow based on seasonal support vector regression with adaptive genetic algorithm," *Applied Soft Computing Journal*, vol. 26, pp. 435–443, 2015.

[33] R. Ladhari and M. Michaud, "eWOM effects on hotel booking intentions, attitudes, trust, and website perceptions," *International Journal of Hospitality Management*, vol. 46(3), pp. 36–45, 2015.

[34] T. Wenchin, H. Mingtsang, S. Liwen and T. M. Y. Lin, "Compliance with eWOM: The influence of hotel reviews on booking intention from the perspective of consumer conformity," *International Journal of Hospitality Management*, vol. 46, pp. 99–111, 2015.

[35] L. V. Casaló, C. Flavián, M. Guinalíu and Y. Ekinci, "Do online hotel rating schemes influence booking behaviors?," *International Journal of Hospitality Management*, vol. 49, pp. 28–36, 2015.

[36] M. R. R. Rana, A. Nawaz and J. Iqbal, "A survey on sentiment classification algorithms, challenges and applications," *Acta Universitatis Sapientiae Informatica*, vol. 10(1), pp. 58-72, 2018.

[37] F. Hemmatian and M. K. Sohrabi, "A survey on classification techniques for opinion mining and sentiment analysis," *Artificial Intelligence Review*, vol. 52(3), pp. 1495-1545, 2019.

[38] H. Raisi, R. Baggio, L. Barratt-Pugh and G. Willson, "Hyperlink network analysis of a tourism destination," *Journal of Travel Research*, vol. 57(5), pp. 671-686, 2018.

[39] J. Yuan and Z. T. Wu, "Research on effect of convergence between creative industry park and tourism based on network text: A case study of shenzhen dafen oil painting village," *Resource Development & Market*, vol. 32(5), pp. 631–636, 2016.

[40] C. Messaoudi, Z. Guessoum and L. Ben Romdhane, "Opinion mining in online social media: a survey," *Social Network Analysis and Mining*, vol. 12(1), pp. 1-18, 2022.

[41] P. Xiang, K. Tian and Q. Huang, "Semantic association identification and knowledge discovery for business applications," In *The International Conference Information Computing and Automation*, 2007, pp. 480–483.

[42] M. Kutas, S. A. Hillyard, "Brain potentials during reading reflect word expectancy and semantic association," *Nature*, vol. 307(5947), pp. 161–163, 1984.

[43] F. Herrera and L. Martinez, "A 2-tuple fuzzy linguistic representation model for computing with words," *IEEE Transactions on Fuzzy Systems*, vol. 8(6), pp. 746–752, 2000.

[44] X. Cheng, S. Fu, J. Sun, A. Bilgihan and F. Okumus, "An investigation on online reviews in sharing economy driven hospitality platforms: A viewpoint of trust," *Tourism Management*, vol. 71, pp. 366–377, 2019.

[45] P. Racherla and W. Friske, "Perceived 'usefulness' of online consumer reviews: An exploratory investigation across three services categories," *Electronic Commerce Research and Applications*, vol. 11(6), pp. 548–559, 2012.