Research paper / Оригинальная статья
https://doi.org/10.51176/1997-9967-2024-3-60-72
МРНТИ 06.71.57
JEL: L83, Z31, Z32



# Analysis of Online User Reviews for Popular Tourist Attractions: Almaty Case

**Alper Kürşat Uysal[a]*, Murat Alper Başaran[a], Kemal Kantarcı[a]**

[a]*Alanya Alaaddin Keykubat University, Kestel district, 80 University Str., 07425, Alanya, Antalya, Turkiye*

**For citation:** Uysal, A. K., Başaran, M.A. & Kantarcı, K. (2024). Analysis of Online User Reviews for Popular Tourist Attractions: Almaty Case. Economy: strategy and practice, 19(3), 60-72, https://doi.org/10.51176/1997-9967-2024-3-60-72

## ABSTRACT

Attractions in the tourism industry are one of the components that motivate tourists to visit destinations, such as entertainment, natural, cultural, and historical richness. For such reasons, people decide to visit unique destinations and spend time there. Almaty, the largest city of Kazakhstan, is one of the significant attraction centers of the Central Asia region, offering tourists unique and pleasant features with several tourist attractions. This study aims to analyze online user reviews of tourist attractions in Almaty, Kazakhstan, using machine learning and text mining methods. The primary focus is on identifying the main thematic clusters of reviews and their sentiment and comparing these themes with the types of attractions: historical, natural, and man-made. A total of 7,515 reviews were collected from the TripAdvisor website. The data was processed using sentiment analysis, topic modeling, and hierarchical clustering methods. The analysis revealed that 38% of the reviews were related to natural attractions, 34% to man-made, and 28% to historical ones. The most positive reviews were associated with natural attractions, while historical and man-made attractions received 79.38% and 81.40% positive reviews, respectively. In addition, the items that make up these attractions are identified, and their sentiment levels are pointed out. In addition to this situation, visitors have the most positive expressions for natural attractions, especially landscapes and lakes. The findings emphasize the importance of considering review themes to improve the quality of tourist services and to enhance the positive image of Almaty as a tourist destination.

**KEYWORDS:** Economic Impact, Attraction, Marketing, Travel Services, Destination Image, Text Mining, Tourism, Strategic Tourism Management, Almaty, Kazakhstan

**CONFLICT OF INTEREST:** the authors declare that there is no conflict of interest

**FINANCIAL SUPPORT:** This research was funded by grant number AP19679687 between 2023 and 2025.

**Article history:**
Received 28 April 2024
Accepted 03 June 2024
Published 30 September 2024

*** Corresponding author: Alper Kürşat Uysal -** Associate Professor, , Alanya Alaaddin Keykubat University, Kestel district, 80 University Str., 07425, Alanya, Antalya, Türkiye, +90 535 977 7370, email: alper.uysal@alanya.edu.tr, ORCID ID: https://orcid.org/0000-0002-4057-934X

# Анализ онлайн-отзывов пользователей о популярных туристских достопримечательностях: пример г. Алматы

**Уйсал А.К.[a]*, Башаран М.А.[a], Кантарджы К.[a]**

[a] *Алания Алааддин Кейкубат Университет, район Кестел, ул. Университет 80, 07425, Алания, Анталия, Турция*

**Для цитирования:** Уйсал А.К., Башаран М.А., Кантарджы К. (2024). Анализ онлайн-отзывов пользователей о популярных туристских достопримечательностях: пример г. Алматы. Экономика: стратегия и практика, 19(3), 60-72, https://doi.org/10.51176/1997-9967-2024-3-60-72

**Аннотация**

Туристические достопримечательности являются одним из важных компонентов, которые мотивируют туристов посещать различные дестинации. По этим причинам люди решают посещать уникальные места и проводить там время. Алматы, крупнейший город Казахстана, является одним из основных центров притяжения туристов в Центральной Азии, предлагая уникальные и приятные особенности с множеством туристических достопримечательностей. Целью данного исследования является анализ онлайн-отзывов пользователей о туристических достопримечательностях Алматы с использованием методов машинного обучения и текстового майнинга. Основное внимание уделяется выявлению основных тематических кластеров отзывов, их эмоциональной окраски и сопоставлению этих тем с типами достопримечательностей: историческими, природными и рукотворными. В рамках исследования было собрано 7 515 отзывов с сайта TripAdvisor. Данные были обработаны с использованием методов анализа тональности, тематического моделирования и иерархической кластеризации. Анализ показал, что 38% отзывов относятся к природным достопримечательностям, 34% — к рукотворным и 28% — к историческим. Наиболее положительные отзывы были связаны с природными достопримечательностями, тогда как исторические и рукотворные объекты получили 79,38% и 81,40% положительных отзывов соответственно. Кроме того, были определены элементы, составляющие эти достопримечательности, и указаны уровни их эмоциональной оценки. Посетители оставляют наиболее положительные отзывы о природных достопримечательностях, особенно пейзажах и озёрах. Полученные результаты подчеркивают важность учета тематики отзывов для улучшения качества туристических услуг и укрепления положительного имиджа Алматы как туристического направления.

**Ключевые слова:** экономическое влияние, достопримечательность, маркетинг, туристические услуги, имидж дестинации, туризм, стратегическое управление туризмом, Алматы, Казахстан

**Конфликт интересов:** авторы заявляют об отсутствии конфликта интересов.

**Финансирование.** Данное исследование финансировалось по гранту AP19679687 на период 2023-2025.

**История статьи:**
Получено 28 April 2024
Принята 03 June 2024
Опубликовано 30 Сентября 2024

***Корреспондирующий автор:** **Уйсал А.К.** – PhD, Алания Алааддин Кейкубат Университет, район Кестел, ул. Университет 80, 07425, Алания, Анталия, Турция, +90 534 310 54 74, email: egemen.tukenmez@alanya.edu.tr

## INTRODUCTION

The attractions of destinations are one of the main factors in the formation of tourist demand (Nasir et al., 2020). Historically significant places, natural beauties, and man-made elements within the scope of these attractions motivate tourists to visit destinations and make significant economic and social contributions to destinations (Pike & Page, 2014). For this reason, businesses and destination management organizations want to attract tourists to relevant areas in quantity and quality by effectively managing these attraction elements and developing efficient strategies in marketing activities (Go & Govers, 2000; Gato et al., 2022; Blain et al., 2005). This tourist attraction is an important indicator that increases the service quality, brand value, and image of the countries in the tourism sector (Kazmi et al., 2020).

The situation and quality of attractions in a destination have an essential role in the travel satisfaction level of tourists (Mariani et al., 2014). This is an important fact that naturally affects the image of the destination and the country where the attractions are located (Nadeau et al., 2008). With the increasing accessibility of transportation and technological advancements, there has been a growing interest in tourism among people over time (Hącia, 2019; Pencarelli, 2020). Alongside developments in tourism, individuals have been inclined towards traditional tourism types and destinations while paving the way for visits to unique cities and regions (Cimbaljević et al., 2019; Kim et al., 2007; Kim & Brown, 2012). One of these unique places is the Central Asian geography. The cultural, local, and geographical attractions of this region have increased people's travels to this region, and in this context, tourists are provided with different elements in their travels by diversifying touristic products (Mukhambetov & Ottenbacher, 2021; Panzabekova, 2018).

Kazakhstan is one of the most important attractions in Central Asia and has an essential position in its region due to its geopolitical importance (Wang & Liu, 2020). In addition, it has an essential position in tourism in its region with its cultural, natural, and man-made attractions (Kantarci et al., 2017; Kadyrbekova et al., 2023). Within the scope of these attractions, Kazakhstan ranked 33rd in the world in natural resource attractiveness and 63rd in cultural resources in the report prepared by the Travel and Tourism Development Index for 2021 (World Economic Forum, 2021). For this reason, using these resources efficiently, protecting them within the scope of sustainable development, carrying out restoration works, and using publicity and promotion activities effectively all over the world can increase the number of tourists in terms of quality and quantity (Brokaj, 2014; Larson & Poudyal, 2012). This country has significant attractions from post-Soviet times and nomadic culture (Tiberghien et al., 2018). This place also has enormous nature-based attractions, but these resources cannot be sufficiently adapted to the tourism industry for competitive advantage because of lack of professionalism and investments (Tleuberdinova et al., 2022; Shayakhmetova et al., 2020). For this reason, natural resources in this region should be protected efficiently for future generations (Igaliyeva et al., 2020).

With Web 2.0 applications, how consumers convey their satisfaction by expressing their opinions on digital platforms has accelerated (Garner & Kim, 2022). The fact that this interaction occurs in the field of tourism is significant for components in the tourism industry, which are part of the service sector (Herrero et al., 2015). For this reason, examining the reviews written in these areas and measuring their impact on consumers' experiences is essential in revealing satisfaction with the attractions in destinations (Lu & Stepchenkova, 2015). In big data analysis, developments in machine learning, text mining, and natural language processing have paved the way for the practical analysis of large amounts of data presented on these platforms (Ghavami, 2019; Marine-Roig, 2021). Attractions in destinations have complex structures; therefore, researchers apply machine learning models to reveal hidden features from text data, such as online reviews (Taecharungroj & Mathayomchan, 2019).

Within the development in this field, online user-generated reviews have been studied for attractions in a destination management perspective with machine learning techniques in growing text mining and natural language processing methods like topic modeling, information retrieval, text clustering, text classification, and sentiment analysis. In a study conducted on 40 cities, Latent Dirichlet Allocation, one of the topic modeling methods, was used to reveal the topics in online reviews in TripAdvisor's "Things to Do" section. According to the results, castle, religion, ancient, sports, and theater topics are the most mentioned topics in the reviews of these 40 cities frequently visited by tourists (McKenzie & Adams, 2018). A relevant study conducted for Marrakech, one of the famous destinations in Morocco, via a topic modeling approach showed that tourists mainly mention the atmosphere, shopping, behaviors of citizens, and general experiences in their reviews (Ali et al., 2021). Location-provided services and activities are also important features for tourists visiting popular Indian hotspots (Singh et al., 2021).

In this study, online user reviews on the TripAdvisor website, one of the world's largest online travel platforms, were examined regarding the cultural, natural, and man-made tourist attractions of Almaty, the largest city in Kazakhstan. In this way, the content of the reviews of people who visited the relevant attractions and the effect of these contents on satisfaction were revealed. Few related studies about attraction types in Almaty have online user-generated reviews. To reveal and understand the contents of reviews, machine learning, and text mining techniques were used together to gather the related information from the obtained dataset. Furthermore, it is aimed to reveal which topics have the most positive and negative sentiments for each attraction type and how the subjects within these topics are used together.

## MATERIALS AND METHODS

Attractions with 5 or more comments on the attractions of Almaty city on TripAdvisor were included in the research in this context, and a total of 76 attractions were included in the research (TripAdvisor, 2024). When looking at these attractions, it was seen that most of the reviews were written in English and Russian. In addition, attractions are generally divided into three groups: natural beauties such as lakes, mountains, and snow; historical places such as history, culture, and architectural structures; and later, human structures such as subways, shopping areas, and streets. In this context, 7515 reviews were obtained from the relevant site and analyzed with the Orange data mining program (Demšar et al., 2004). The data range covers between 2011 and 2024. After all data were scraped from TripAdvisor, preprocessing was performed on the raw text data. At this stage, the numerical data in the comments were removed, and all capital letters were converted to lowercase letters. After this stage, the VADER sentiment analysis method was applied to every review. The non-negative matrix factorization method was performed to obtain meaningful and representative word clusters for topic modeling. After this phase, topic distribution and sentiment scores of every review were identified, and hierarchical clustering analysis was performed on this dataset to observe co-occurrences between topics.

## RESULTS

In the analysis of tourist attractions, Figure 1 illustrates the distribution of review frequencies across different types of attractions. The data reveals that nature-based attractions received the highest reviews, followed closely by man-made attractions, with historical attractions garnering the least attention. Specifically, when expressed in percentage terms, nature-based attractions accounted for 38% of the total reviews, man-made attractions represented 34%, and historical attractions comprised 28%. This distribution highlights tourists' predominant interest in natural landscapes and environments, although artificial and historically significant sites also maintain substantial appeal.

Figure 1 shows the attraction types and their review frequencies.

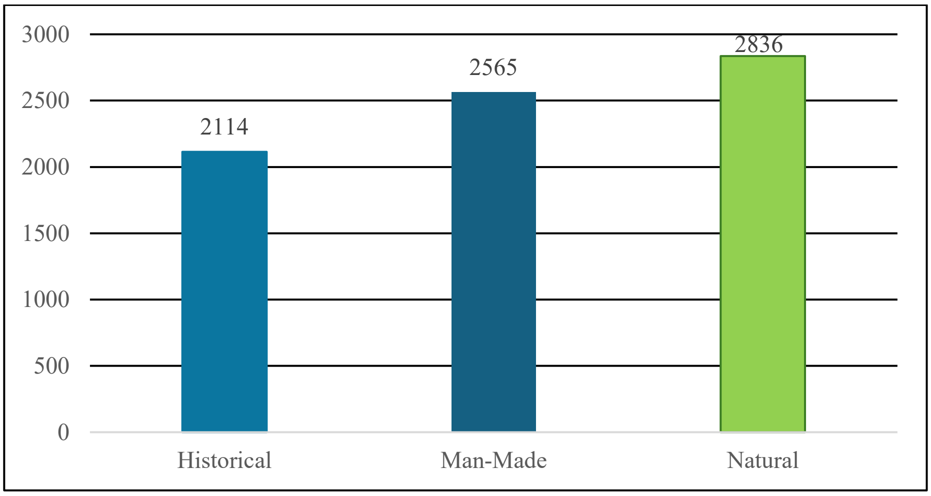


**Figure 1.** Attraction types and frequency distributions obtained from the data set

After the relevant reviews were filtered under three different attraction types, sentiment analysis, topic modeling, and text classification were performed for each group. Sentiment analysis is an analysis method used to express a sentence's emotional state and intensity (Hutto & Gilbert, 2014). In the relevant research, the VADER method was used, and the emotion type of each sentence was classified according to the resulting compound result.

Figure 2 shows the emotion types of all data. Accordingly, 6,160 positive emotions,

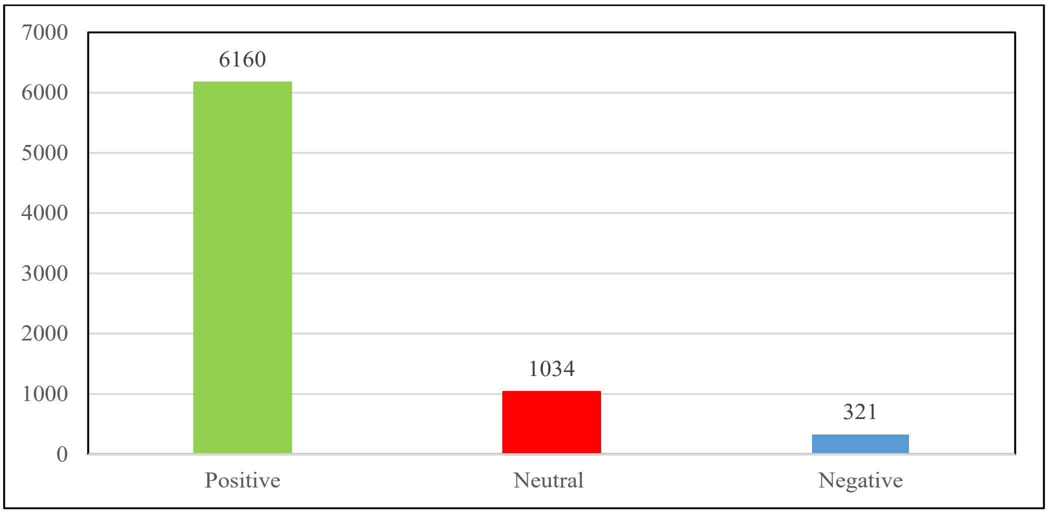


**Figure 2.** Sentiment analysis results of all reviews in the data set

In the classification of emotional responses, 1,034 instances were identified as neutral emotions, while 321 instances were classified as negative emotions. The data highlights a significant predominance of positive emotions, constituting 81.97% of the total emotional responses. This indicates that most of the emotions expressed were positive, reflecting a generally favorable sentiment among the subjects. Neutral emotions, which account for 13.76% of all emotions, suggest some responses where individuals neither exhibited strong positive nor negative feelings. This could indicate a more reserved or indifferent reaction to the analyzed stimuli or context. Negative emotions, making up only 4.27% of the total, are the least represented among the classified emotions.

Figure 3 shows the distribution of the sentiments of all three attraction types.

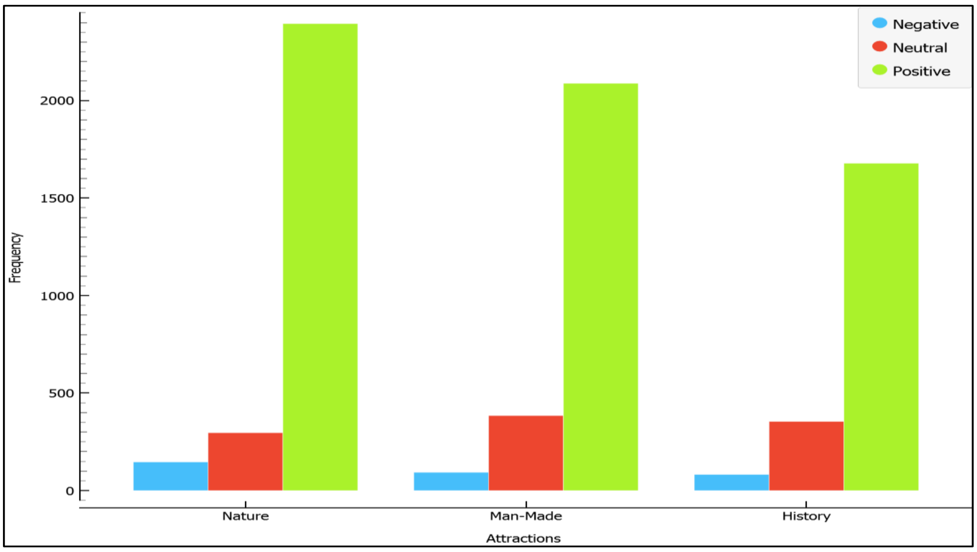


**Figure 3.** Sentiment Types of All Attractions

According to the sentiments of all attractions, the most positively commented attraction type is nature-based. In this type, 84.41% of all nature-based attraction reviews are positive, 10.44% are neutral, and %5.15 are negative, respectively. In addition, positive sentiments make up 79.38% of all reviews, neutral 16.75%, and negative emotions 3.88% in history-based attractions. In man-made attractions, positive sentiments have a percentage of 81.40% of all reviews, neutral has a percentage of 14.97%, and negative one has a percentage of 3.63%. According to findings, although nature-based attractions have more negative sentiments in percentage, the highest percentage difference between negative and neutral sentiments of attraction types is in history-based ones. Man-made and nature-based attractions follow this situation. Furthermore, the most significant difference between neutral and positive sentiment belongs to nature-based attractions. This proportional size difference is followed by man-made and history-based attractions seriatim.

After this process, the topic distributions of each attraction were analyzed using the topic modeling algorithm called Non-Negative Matrix Factorization. It is a dimension-reduction technique used in machine learning and text mining areas such as topic modeling, document similarity, and feature extraction from text documents (Zurada et al., 2013). As a result of the analysis, the value with the highest probability in the topic distribution of each review was classified as the topic of that review. Then, after the sentiment and topic classes of the classified reviews were determined, the topics and sentiment states of each of the three attraction types were analyzed. Considering the topic integrity within the scope of topic modeling, a total of 4 topics were presented for each attraction type, and the distribution of the words in the topics that constitute the attractiveness in the sentences of the relevant attractiveness type was shown with marginal topic probability at the top of each column in Tables 1, 2, and 3 respectively.

Table 1 presents the topic modeling results related to the history-based attraction type, providing a detailed overview in this category.

**Table 1.** Topic modeling results of history-based attraction type

| Architectural | Distribution | Social Memory | Distribution | Museum | Distribution | Art | Distribution |
|---|---|---|---|---|---|---|---|
| Marginal Topic Probability | 0.360218 | Marginal Topic Probability | 0.329586 | Marginal Topic Probability | 0.103076 | Marginal Topic Probability | 0.199693 |
| Cathedral | 0.0553142 | Monument | 0.0528893 | Museum | 0.0861758 | Art | 0.0851309 |
| Church | 0.0507704 | Park | 0.0468671 | Hour | 0.0510682 | Instrument | 0.0507274 |
| Beautiful | 0.049173 | Memorial | 0.0431495 | Exhibit | 0.0425852 | Music | 0.0475057 |
| Orthodox | 0.0353729 | Tree | 0.0368214 | Spend | 0.0425006 | Folk | 0.0433161 |
| Wooden | 0.0293969 | War | 0.0325079 | Collection | 0.0412122 | Display | 0.0385741 |
| Mosque | 0.0252959 | Panfilov | 0.0320651 | English | 0.04051 | Pleasant | 0.0361879 |
| Building | 0.0229772 | Moscow | 0.0278459 | Room | 0.0330272 | Guide | 0.0353366 |
| Love | 0.0226024 | Victory | 0.0277006 | History | 0.0298839 | Interesting | 0.0337542 |
| Russian | 0.0210765 | Walk | 0.0257135 | Modern | 0.0281714 | Local | 0.0285703 |
| Architecture | 0.0208298 | Impressive | 0.025166 | Painting | 0.0266259 | Exhibition | 0.0259649 |

Note: compiled by authors

Table 1 shows that in the history-based attraction type, words cluster about architecture, social memory, museums, and art topics. In other words, related words in these clusters are written more in the reviews of this type of attraction. According to each topic cluster's marginal topic probability scores, the words in the architectural element are used the most, and the words in the museum element are used the least.

Table 2 shows the topic modeling results of the nature-based attraction type.

**Table 2.** Topic modeling results of nature-based attraction type

| Panoramic View | Distribution | Mountain | Distribution | Animals | Distribution | Lake | Distribution |
|---|---|---|---|---|---|---|---|
| Marginal Topic Probability | 0,36747 | Marginal Topic Probability | 0,205227 | Marginal Topic Probability | 0,157897 | Marginal Topic Probability | 0,260507 |
| View | 0,031936 | Ski | 0,06425 | Animal | 0,0654574 | Lake | 0,075785 |
| City | 0,029488 | Resort | 0,038887 | Garden | 0,0463533 | Road | 0,028548 |
| Kok | 0,026131 | Shymbulak | 0,029663 | Zoo | 0,0448032 | Water | 0,025099 |
| Cable | 0,025422 | Slope | 0,025616 | Bird | 0,0280845 | Mountain | 0,022887 |
| Tobe | 0,025059 | Lift | 0,023345 | Child | 0,0236987 | Beautiful | 0,021268 |
| Hill | 0,024558 | Hotel | 0,01904 | Cage | 0,0207323 | Bus | 0,019236 |
| Sunset | 0,022619 | Winter | 0,017821 | Entrance | 0,0202134 | Blue | 0,016984 |
| Top | 0,022411 | Snow | 0,016478 | Walk | 0,0195675 | Driver | 0,015903 |
| Nice | 0,021837 | Cable | 0,015889 | Park | 0,0188086 | Taxi | 0,015541 |
| Child | 0,019084 | Station | 0,01581 | Enclosure | 0,0184302 | Nature | 0,015472 |

Note: compiled by authors

When we look at the words that make up the nature-based attraction type in Table 2, we see that they relate to panoramic views, mountains, animals, and lakes. The most common subject here is the panoramic view topic, and the least common is the animal topic.

Further, Table 3 shows the topic modeling results of the man-made-based attraction type.

**Table 3.** Topic modeling results of man-made-based attraction type.

| Winter Tourism | Distribution | Parks | Distribution | Shopping | Distribution | Transportation | Distribution |
|---|---|---|---|---|---|---|---|
| Marginal Topic Probability | 0,352633 | Marginal Topic Probability | 0,277102 | Marginal Topic Probability | 0,137912 | Marginal Topic Probability | 0,223863 |
| Skate | 0,0536991 | Park | 0,086159 | Market | 0,05202 | Metro | 0,055952 |
| Rink | 0,0348971 | Child | 0,032666 | Fruit | 0,044675 | Station | 0,052272 |
| Mountain | 0,0337025 | Theater | 0,027746 | Bazaar | 0,04057 | Mall | 0,031133 |
| View | 0,0272242 | Walk | 0,027437 | Street | 0,033274 | Line | 0,027259 |
| Car | 0,0265342 | Photo | 0,023492 | Shop | 0,033193 | Train | 0,025237 |
| Ice | 0,0257745 | Place | 0,023432 | Green | 0,032052 | Clean | 0,025037 |
| Medeo | 0,0254322 | Nice | 0,023185 | Meat | 0,03201 | City | 0,023031 |
| Cable | 0,0224777 | Love | 0,022828 | Building | 0,03193 | Shop | 0,020814 |
| Ski | 0,0192443 | Beatles | 0,021869 | Food | 0,030681 | Staff | 0,019853 |
| Medeu | 0,0177825 | Attraction | 0,019395 | Price | 0,02927 | Experience | 0,01968 |

Note: compiled by authors

Table 3 shows the topics of man-made attractions, and words related to winter tourism, parks, shopping, and transportation come together. Winter tourism is the most common topic among these, and shopping is the least common.

Figures 4, 5, and 6 show the sentiment distribution of the attraction types in percentage according to the topics.

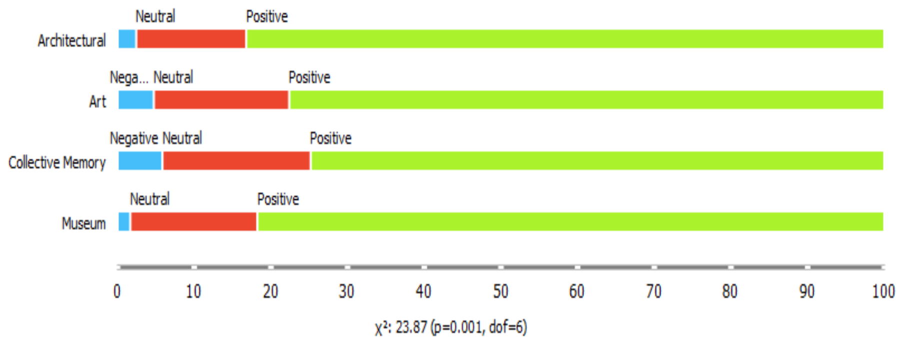


**Figure 4.** Topics and sentiment distribution of the history-based attraction type

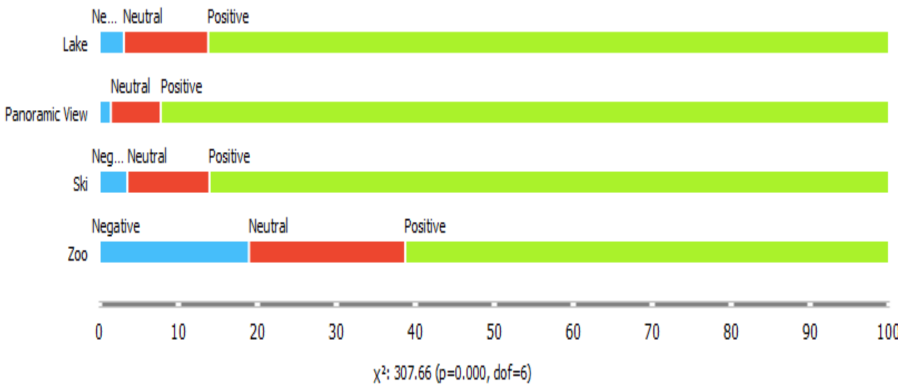


**Figure 5.** Topics and sentiment distribution of the nature-based attraction type

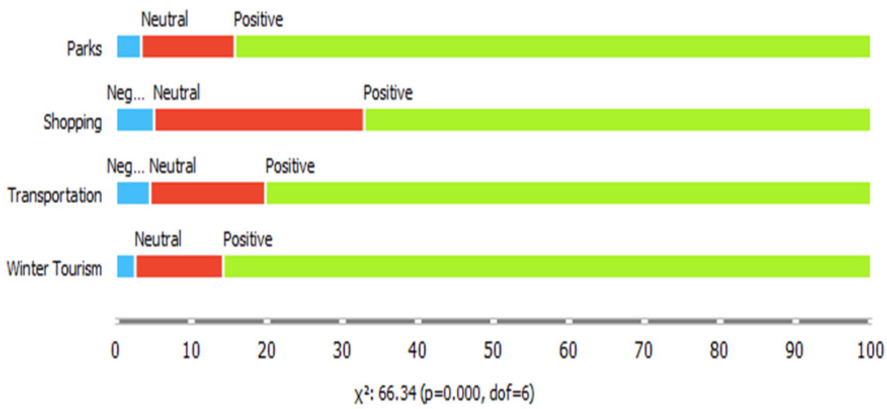


**Figure 6.** Topics and sentiment distribution of the man-made attraction type

Figure 4 illustrates that the element expressed in the most positive context is the subject of architecture. Museum, art, and social memory elements follow this situation. Within this subject, words about architectural buildings, structures, and museums are essential for tourist satisfaction when visiting these related history-based attractions. Figure 5 shows that, while the panoramic view has the highest distribution of positive opinions in nature attraction, the zoo topic emerged as the element with the highest distribution of negative emotions. Figure 6 presents the issues of man-made attractions, and among these issues, it is seen that the elements related to winter tourism are considered the most positive. In contrast, the shopping element stands out the most as unfavorable. In addition, the Chi-Square analysis also showed that all topics reached the significance level.

Furthermore, hierarchical cluster analysis is performed to understand which topics tourists use together in the same review. Cosine similarity and ward linkage parameters are used to perform a dendrogram. Figures 7,8 and 9 show the co-occurrence of topics in history, nature, and man-made attraction types, respectively.

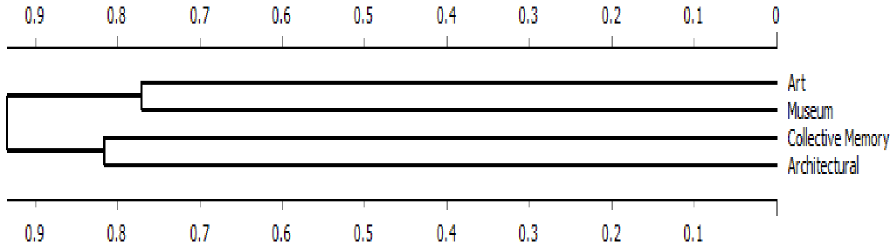


**Figure 7**. Hierarchical Cluster Analysis of Reviews in History-Based Attractions

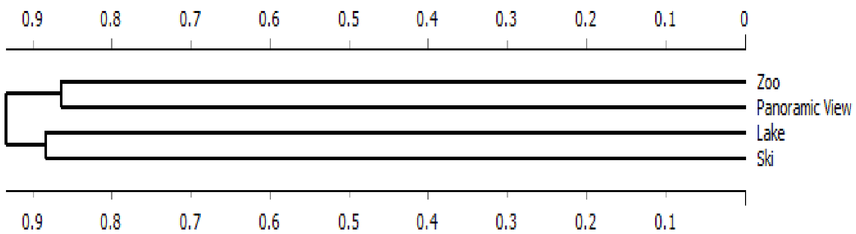


**Figure 8.** Hierarchical Cluster Analysis of Reviews in Nature-Based Attractions

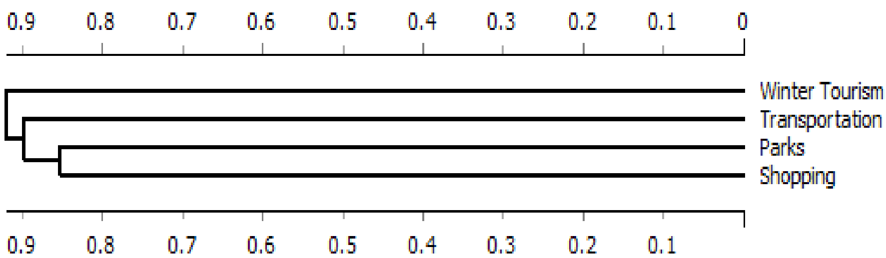


**Figure 9.** Hierarchical Cluster Analysis of Reviews in Man-Made-Based Attractions

Figure 7 shows that visitors mainly refer to art and museum-based words in the same review. These topics are mentioned together more than other topics in this section. Moreover, it is seen that collective memory and architecture-based words also tended to be written in reviews by tourists together. Figure 8 shows the nature-based hierarchical clustering analysis. According to the findings in this topic, lake and ski, on the other hand, panoramic view and zoo tended to be written together. Figure 9 shows that parks and shopping topics are used together. Subsequently, the transportation topic is added to this group. Winter tourism and its related words appear to form a separate cluster of man-made attractions.

## DISCUSSION AND CONCLUSION

In this study, the attractions in Almaty were examined, the reviews of the attractions in the relevant region were taken from the TripAdvisor website, and content analysis of the reviews was conducted using text mining and machine learning methods. At this stage, the topics and sentiment states of the reviews made on the relevant attraction types were examined using sentiment analysis, topic modeling text classification, and unsupervised learning hierarchical clustering methods. In the research, it was seen that attractions are classified by history, nature, and man-made based in Almaty. In attractions related to history, people generally commented positively on the historical elements, and it is shown that people comment on architectural structures, collective memory, arts, and museums. In this part, architecture and museums have the most positive sentiments. Art and collective memory also have positive emotions higher than negative. However, when comparing the percentage of sentiments to other topics in history-based attractions, the negative sentiment percentage of these two topics is higher than others. In this context, monuments, structures, and events that express the social memory of Kazakhstan and Kazakh citizens, as well as art topics that show Kazakh culture, can be improved to enhance visitor satisfaction for history-based attractions in Almaty in the future. Furthermore, carrying out promotional materials and restoration works for the relevant areas will significantly increase visitor loyalty and revisit actions and eWOM in the relevant areas. When considering this situation, our analysis shows that people tend to write their expression about history-based attraction considering the co-occurrence topics. Museum and art topics are considered together by visitors in their reviews. On the other hand, collective memory and architecture tend to be mentioned together. From a destination management perspective, handling and taking over these topics together for tourism planning and policy in Almaty is better. In restoration and protection strategies and promotional actions, these topics should be considered together, and these strategies should support each other.

In reviews about nature-based attractions, visitors tend to make the highest positive comments about the landscape, including activities and attractions such as panoramic views, mountains, lakes, and skiing. Although the Kazakh and Central Asia region has important culture-based attractions, nature-based ones take the most satisfied reviews by tourists, and this type of attraction has an enormous role in attracting tourists from all over the world. However, in this context, the comments about the zoo have a higher negative share in attractiveness than others. For this reason, strategies such as improving the service quality in natural areas where animals live, wildlife parks, and zoos, as well as improving the living spaces of animals, may enable people who visit these areas to make more positive comments. When we look at the man-made attractions that humans mainly constructed in recent years, the reviews about winter tourism and parks have positive sentiments7. However, the reviews given in the shopping area were seen to be more damaging than others. In this context, improvements, infrastructure, superstructure works, and pricing strategies regarding shopping areas, shopping malls, bazaars, and similar places may increase the number of reviews regarding this area to a positive level in the future.

One main limitation encountered in this study is that most reviews are in English and Russian. For this reason, it seems to be challenging to reveal the content of reviews written outside these two languages and to make comparisons between them. Furthermore, it wasn't easy to classify three types of attractions according to our findings, which were classified based on current attractions and their review contents in related travel platforms. In this context, tourist attractions in Almaty should be developed and diversified, and the existing attractions should be better marketed to encourage people from different countries to visit these attractions. In future research, researchers can analyze the reviews on other travel and tourism platforms and use different digital sources to reveal different topics and approaches. In addition, different text mining, natural language processing, and machine learning models can be used to reveal different topics and contexts from scraped data.

## AUTHOR CONTRIBUTIONS

Conceptualization and theory: AKU, MAB and KK; research design: AKU, MAB and KK; data collection: AKU, MAB and KK; analysis and interpretation: AKU and MAB; writing draft: AKU and MAB; supervision: AKU, MAB and KK; correction of article: AKU, MAB and KK; proofread and final approval of article AKU, MAB and KK. All authors have read and agreed to the published version of the manuscript.

## REFERENCES


Ali, T., Marc, B., Omar, B., Soulaimane, K., & Larbi, S. (2021). Exploring destination's negative e-reputation using aspect-based sentiment analysis approach: Case of Marrakech destination on TripAdvisor. *Tourism Management Perspectives, 40*, 100892. https://doi.org/10.1016/j.tmp.2021.100892

Blain, C., Levy, S. E., & Ritchie, J. B. (2005). Destination branding: Insights and practices from destination management organizations. *Journal of Travel Research, 43*(4), 328-338. https://doi.org/10.1177/0047287505274646

Brokaj, R. (2014). Local governments role in the sustainable tourism development of a destination. *European Scientific Journal, 10*(31), 103-117.

Cimbaljević, M., Stankov, U., & Pavluković, V. (2019). Going beyond the traditional destination competitiveness–reflections on a smart destination in the current research. *Current Issues in Tourism, 22*(20), 2472-2477. https://doi.org/10.1080/13683500.2018.1529149

Demšar, J., Zupan, B., Leban, G., & Curk, T. (2004). Orange: From experimental machine learning to interactive data mining. In *Knowledge discovery in databases: PKDD 2004: 8th European conference on principles and practice of knowledge discovery in databases, Pisa, Italy, September 20-24, 2004. Proceedings 8* (pp. 537-539). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-30116-5_58

Garner, B., & Kim, D. (2022). Analyzing user-generated content to improve customer satisfaction at local wine tourism destinations: An analysis of Yelp and TripAdvisor reviews. *Consumer Behavior in Tourism and Hospitality, 17*(4), 413-435. https://doi.org/10.1108/CBTH-03-2022-0077

Gato, M., Dias, Á., Pereira, L., da Costa, R. L., & Gonçalves, R. (2022). Marketing communication and creative tourism: An analysis of the local destination management organization. *Journal of Open Innovation: Technology, Market, and Complexity, 8*(1), 40. https://doi.org/10.3390/joitmc8010040

Ghavami, P. (2019). *Big data analytics methods: Analytics techniques in data mining, deep learning and natural language processing* (2nd ed.). Walter de Gruyter GmbH & Co KG. https://doi.org/10.1515/9781547401567

Go, F. M., & Govers, R. (2000). Integrated quality management for tourist destinations: A European perspective on achieving competitiveness. *Tourism Management, 21*(1), 79-88. https://doi.org/10.1016/S0261-5177(99)00098-9

Hącia, E. (2019). The role of tourism in the development of the city. *Transportation Research Procedia, 39*, 104-111. https://doi.org/10.1016/j.trpro.2019.06.012

Herrero, Á., San Martín, H., & Hernández, J. M. (2015). How online search behavior is influenced by user-generated content on review websites and hotel interactive websites. *International Journal of Contemporary Hospitality Management, 27*(7), 1573-1597. https://doi.org/10.1108/IJCHM-05-2014-0255

Hutto, C., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8, No. 1, pp. 216-225). https://doi.org/10.1609/icwsm.v8i1.14550

Igaliyeva, L., Yegemberdiyeva, S., Utepkaliyeva, K., & Bakirbekova, A. (2020). Development of economic mechanism for ensuring ecological security in Kazakhstan. *International Journal of Energy Economics and Policy, 10*(4), 240-250. https://doi.org/10.32479/ijeep.9634

Kadyrbekova, D., Kassenali, A., & Yevloyeva, A. (2023). A comprehensive study of Kazakhstan's cultural heritage and its impact on domestic tourism. *ECONOMIC Series of the Bulletin of the LN Gumilyov ENU, 4*, 339-352. https://doi.org/10.32523/2789-4320-2023-4-339-352

Kantarci, K., Uysal, M., Magnini, V., & Basaran, M. A. (2017). *Tourism in Central Asia.* In Hall, M., & Page, S. (Eds.), Handbook of Tourism in Asia (pp. 275-286). Routledge.

Kazmi, S. H. A., Raza, M., & Ahmed, J. (2020). Impact of destination service quality on revisit intention in tourism. *Journal of Organisational Studies & Innovation, 7*(3), 26-45. http://dx.doi.org/10.13140/RG.2.2.23418.31680

Kim, A. K., & Brown, G. (2012). Understanding the relationships between perceived travel experiences, overall satisfaction, and destination loyalty. *Anatolia, 23*(3), 328-347. https://doi.org/10.1080/13032917.2012.696272

Kim, H., Cheng, C. K., & O'Leary, J. T. (2007). Understanding participation patterns and trends in tourism cultural attractions. *Tourism Management, 28*(5), 1366-1371. https://doi.org/10.1016/j.tourman.2006.09.023

Larson, L. R., & Poudyal, N. C. (2012). Developing sustainable tourism through adaptive resource management: A case study of Machu Picchu, Peru. *Journal of Sustainable Tourism, 20*(7), 917-938. https://doi.org/10.1080/09669582.2012.667217

Lu, W., & Stepchenkova, S. (2015). User-generated content as a research mode in tourism and hospitality applications: Topics, methods, and software. *Journal of Hospitality Marketing & Management, 24*(2), 119-154. https://doi.org/10.1080/19368623.2014.907758

Mariani, M. M., Buhalis, D., Longhi, C., & Vitouladiti, O. (2014). Managing change in tourism destinations: Key issues and current trends. *Journal of Destination

*Marketing & Management, 2*(4), 269-272. https://doi.org/10.1016/j.jdmm.2013.11.003

Marine-Roig, E. (2021). Measuring online destination image, satisfaction, and loyalty: Evidence from Barcelona districts. *Tourism and Hospitality, 2*, 62-78. https://doi.org/10.3390/tourhosp2010004

McKenzie, G., & Adams, B. (2018). A data-driven approach to exploring similarities of tourist attractions through online reviews. *Journal of Location Based Services, 12*(2), 94-118. https://doi.org/10.1080/17489725.2018.1493548

Mukhambetov, T., & Ottenbacher, M. (2021). Cluster approach in cultural heritage tourism: Case of the Central Asian section of Silk Road. *Farabi Journal of Social Sciences, 7*(1), 49-70. https://doi.org/10.26577/CAJSH.2021.v7.i1.06

Nadeau, J., Heslop, L., O'Reilly, N., & Luk, P. (2008). Destination in a country image context. *Annals of Tourism Research, 35*(1), 84-106. https://doi.org/10.1016/j.annals.2007.06.012

Nasir, M., Mohamad, M., Ghani, N., & Afthanorhan, A. (2020). Testing mediation roles of place attachment and tourist satisfaction on destination attractiveness and destination loyalty relationship using phantom approach. *Management Science Letters, 10*(2), 443-454. https://doi.org/10.5267/j.msl.2019.8.026

Panzabekova, A. Z. (2018). Diversification of tourism and economic development of Kazakhstan. *R-Economy, 4*(3), 82-87. https://doi.org/10.15826/recon.2018.4.3.012

Pencarelli, T. (2020). The digital revolution in the travel and tourism industry. *Information Technology & Tourism, 22*(3), 455-476. https://doi.org/10.1007/s40558-019-00160-3

Pike, S., & Page, S. J. (2014). Destination Marketing Organizations and destination marketing: A narrative analysis of the literature. *Tourism Management, 41*, 202-227. https://doi.org/10.1016/j.tourman.2013.09.009

Shayakhmetova, L., Maidyrova, A., & Moldazhanov, M. (2020). State regulation of the tourism industry for attracting international investment. *Journal of Environmental Management and Tourism, 11*(6), 1489-1495. https://doi.org/10.14505/jemt.11.6(46).19

Singh, S., Chauhan, T., Wahi, V., & Meel, P. (2021). Mining tourists' opinions on popular Indian tourism hotspots using sentiment analysis and topic modeling. In *5th International Conference on Computing Methodologies and Communication* (pp. 1306-1313). https://doi.org/10.1109/ICCMC51019.2021.9418341

Taecharungroj, V., & Mathayomchan, B. (2019). Analyzing TripAdvisor reviews of tourist attractions in Phuket, Thailand. *Tourism Management, 75*, 550-568. https://doi.org/10.1016/j.tourman.2019.06.020

Tiberghien, G., Bremner, H., & Milne, S. (2018). Authenticating eco-cultural tourism in Kazakhstan: A supply side perspective. *Journal of Ecotourism, 17*(3), 306-319. https://doi.org/10.1080/14724049.2018.1502507

Tleuberdinova, A., Salauatova, D., & Pratt, S. (2022). Assessing tourism destination competitiveness: The case of Kazakhstan. *Journal of Policy Research in Tourism, Leisure and Events, 16*(2), 265-283. https://doi.org/10.1080/19407963.2022.2027954

TripAdvisor (2024). *TripAdvisor attractions in Almaty*. [cited April 18, 2024]. Available: https://www.tripadvisor.com/Attractions-g298251-Activities-oa0-Almaty.html

Wang, Y., & Liu, Y. (2020). Central Asian geo-relation networks: Evolution and driving forces. *Journal of Geographical Sciences, 30*(11), 1739-1760. https://doi.org/10.1007/s11442-020-1810-z

World Economic Forum. (2021). *Travel and tourism development index 2021: Explore the data*. [cited April 18, 2024]. Available: https://www.weforum.org/publications/travel-and-tourism-development-index-2021/explore-the-data/

Zurada, J. M., Ensari, T., Asl, E. H., & Chorowski, J. (2013). Nonnegative matrix factorization and its application to pattern analysis and text mining. In *Federated conference on computer science and information systems* (pp. 11-16). Krakow, Poland.

**Information about the authors**

*__Alper Kürşat Uysal__ – PhD, Associate Professor, Department of Computer Engineering, Alanya Alaaddin Keykubat University, Antalya, Türkiye, email: alper.uysal@alanya.edu.tr, ORCID ID: https://orcid.org/0000-0002-4057-934X

**Murat Alper Başaran -** PhD, Professor, Department of Industrial Engineering, Alanya Alaaddin Keykubat University, 07425, Alanya, Antalya, Türkiye, email: murat.basaran@alanya.edu.tr, ORCID ID: https://orcid.org/0000-0001-9887-5531

**Kemal Kantarcı** – PhD, Professor, Department of Tourism Management, Alanya Alaaddin Keykubat University, Antalya, Türkiye, email: kemal.kantarci@alanya.edu.tr, ORCID ID: https://orcid.org/0000-0002-9572-2959

**Авторлар туралы мәліметтер**

*__Алпер Күршат Уйсал__ – PhD, қауымдастырылған профессор, компьютерлік инженерия кафедрасы, Алания Алааддин Кейкубат Университеті, Анталия, Түркия, email: alper.uysal@alanya.edu.tr, ORCID ID: https://orcid.org/0000-0002-4057-934X

**Мурат Алпер Башаран -** PhD, профессор, Өндірістік инжиниринг кафедрасы, Алания Алааддин Кейкубат Университеті, 07425, Алания, Анталия, Түркия, email: murat.basaran@alanya.edu.tr, ORCID ID: https://orcid.org/0000-0001-9887-5531
**Кемал Кантаржы** – PhD, профессор, туризм менеджменті кафедрасы, Алания Алааддин Кейкубат Университеті, Анталия, Түркия, email: kemal.kantarci@alanya.edu.tr, ORCID ID: https://orcid.org/0000-0002-9572-2959

**Сведения об авторах**

***Альпер Куршат Уйсал** – PhD, доцент, кафедра компьютерной инженерии, Алания Университет Алааддина Кейкубата, Анталья, Турция, email: alper.uysal@alanya.edu.tr, ORCID ID: https://orcid.org/0000-0002-4057-934X
**Мурат Альпер Башаран -** PhD, профессор, кафедра промышленного инжиниринга, Алания Университет Алааддина Кейкубата, 07425, Алания, Анталья, Турция, email: murat.basaran@alanya.edu.tr, идентификатор ORCID: https://orcid.org/0000-0001-9887-5531
**Кемаль Кантарджи** – PhD, профессор, кафедра менеджмента туризма, Алания Университет Алааддина Кейкубата, Анталья, Турция, email: kemal.kantarci@alanya.edu.tr, ORCID ID: https://orcid.org/0000-0002-9572-2959