

Sentiment analysis as a tool for assessing the negative impact of tourism on a destination

*Aleksandr Vetitnev*¹, *Dmitriy Chigarev*^{1*}, and *Vladlena Enushevskaya*²

¹Faculty of Economics and Management Processes, Sochi State University, 354000 Sochi, Russian Federation

²Environmental Engineering Faculty, Sochi State University, 354000 Sochi, Russian Federation

Abstract. This article presents the deliverables of the analysis of the experiences of the visitors to the Sochi resort, which is one of the criteria for assessment of the risk of problems caused by excessive tourism pressure. The authors highlight the influence of perception and experiences of the visitors and their informal web communication on the creation of the area's image. The visitors' reviews of the resort on TripAdvisor were studied using sentiment analysis. The article outlines the essence of the sentiment analysis method, the reasons for its choice to study the visitors' attitude towards the Sochi resort, and the relevance of determining the experiences of its visitors. The authors describe the process of studying 33,475 city visitors reviews in 2011-2021 of 58 Sochi tourist attractions and makes conclusions about the visitors' attitude towards the city, the assessment of its attractions, areas, the level of satisfaction of the tourists from various regions with their vacation. The authors underline predominance of positive reviews and confirms broader possibilities of the sentiment analysis method in comparison with traditional methods of studying public opinion and the possibility of obtaining an objective view of the area image with this method.

1 Introduction

The effectiveness of the activities of regional governments accurately determines the satisfaction of the population with the quality of life.

In the last decade, the experts' attention has been attracted by the problem of the deterioration of the quality of life of the local population of the most demanded destinations, caused by the excessive tourism pressure, which in 2017-2018 even spawned a wave of protests against tourists in Europe [3, 11]. For destination management organizations, especially those that “suffer from their attractiveness”, planning issues - scientifically based foresight and forecasting of tourism development are of paramount importance.

Many academic publications have been devoted to the study of overtourism; in the topic of congestion of the areas demanded by visitors, which has been discussed in the tourism discourse for more than 50 years, the emphasis is shifting from a quantitative approach to a

*Corresponding author: dvchigarev@gmail.com

psychological component, perceived pressure, moreover, perceived not only by the locals but as well by the tourists [6].

Perception as a subjective psychological aspect of the experience is the commencement of the creation of the area's image in the eyes of visitors, its assessment, which can be used, *inter alia*, when forecasting the development of this specific destination, taking into account its tourist overload. Informal travelers communication, their comments, reviews and messages on the Internet play an important role in shaping the destination image.

Originated in the early 2000s, the analysis of the text sentiment (sentiment analysis), which belongs to the group of content analysis methods in computational linguistics, is one of the major methods of identifying the emotional attitude of the text author towards the object under study. It is a method for quantification of qualitative data; its purpose is the automated extraction of emotional content from the text and determination of the polarity of emotional assessments in the text: positive, negative, and neutral [2]. And, since the experience is a sensual emotional reflection of reality, it is practicable to study the tourists reviews using sentiment analysis of these reviews, which allows to better understand the opinions of visitors about leisure facilities.

In Russia, domestic tourism is becoming increasingly sought after. The leading positions of the Sochi resort, the increasing number of tourist arrivals at this destination, their sharp growth after the lifting of quarantine restrictions in 2020 which is observed already now, in 2021 may cause a number of problems in the city that would adversely affect the quality of life of the locals and the quality of the experience of visitors [21, 22]. Therefore, we believe that the issue of studying the Sochi guests experience through sentiment analysis of their reviews on the Internet, which allows to understand the emotional attitude of the authors of messages towards the resort city, is currently quite relevant.

The purpose of this work is to assess the experiences of the visitors to the Sochi resort destination; its task is to test the sentiment analysis method in determining their attitude towards the city.

2 A Literature Review

In the model proposed by Therckelsen & Halkier [20], the perception of people and areas is based, firstly, on the dichotomy of "we" and "they, others" and, secondly, on the fact that the attitude towards the area can be both positive and negative. Based on this, four types of an area's image are possible (which in one of his definitions Philip L. Pearce [12] considers the perception and experience of a site), namely – "friendly", "exotic", "boring", and "hostile". The first two types shape a positive attitude towards the area, while the second create a negative one.

We believe that the Sochi resort belongs to the areas with an exotic image, full of allure and appeal; the attitude towards such sites is most likely positive. In order to confirm or deny the positive attitude of tourists towards Sochi, we have studied their experiences, which they share in their web reviews of the resort.

In the academic community in recent years, the analysis of reviews on the Internet is becoming increasingly more popular. It helps determine travelers satisfaction and a destination image [5, 9, 10, 19], segmentation of the tourist market [1, 8, 13], the assessment of the quality of tourist services [4, 18, 24], the evaluation of the attitude of the locals and guests of host destinations towards mega-events [7, 16].

TripAdvisor, the world's largest travel website, available in 49 countries in 28 languages, was launched in 2000 and immediately became the most popular social network. This platform is visited monthly by 463 million current and potential travelers. The service contains over 884 million reviews and opinions about tourist attractions ("Things to Do"),

which number has the largest share in comparison with other websites. Therefore, this platform was selected to study the tourists reviews of Sochi.

As already noted, the method of sentiment analysis of the text allows us to reveal the emotional attitude of the author of the text to the object in question. The Sochi tourist attractions have become such an object in our study. According to Sazykin & Glushko [15], a tourist attraction is "a tourist and recreational resource localized in space, having a structure, attractiveness, hierarchy, informational support, involved in tourist production activities", the main quality of which is attractiveness – the ability to attract the attention of tourists.

3 Materials and Methods

The sentiment analysis of the reviews of visitors to the Sochi resort destination was carried out using the Python programming language. The data was saved in the .csv format. 58 CSV files were prepared for analysis; then, they were encoded for subsequent statistical processing in SPSS (IBM SPSS Statistics v.21). After all the reviews had been collected, a sentiment score was calculated for each review.

Reviews can have negative, neutral and positive score. The summary index ("compound") is an integrated assessment of the first three scores, which varies from -1 to 1. In order to determine the sentiment of each review, we set the threshold of the compound score as ± 0.2 . If the compound score of the review is more than 0.2, then the review is positive; if less than 0.2 - negative; if from -0.2 to 0.2 - neutral [23].

4 Results

This study covered 33,475 reviews of visitors to the Sochi resort destination in 2011–2021 published on TripAdvisor. The perception of the whole always depends on the perception of the parts; therefore, the empirical basis of this study was the tourists' reviews of 58 Sochi sights (attractions). Of them, two sights received from 90 to 100% of the total number of positive reviews, 35 – from 80 to 90%, 15 – from 70 to 80%, 5 – from 60 to 70%, and only one – 42%. In total, there were 80% positive reviews, 10% negative, 10% neutral, and 0.576 compound ones.

The attractions under consideration were grouped by category and city districts; the reviewers were segmented by their place of residence. The number of reviews in each group and their share of the total number of reviews are presented in Table 1:

Table 1. Segmentation of 58 Sochi attractions by category and district. Segmentation of reviewers by their place of residence

Segmentation of attractions by category and district	No. of reviews	%
By category		
Alpine ski resorts	5,990	17.9
Nature and man-made attractions	6,405	19.1
Theme amusement parks	7,340	21.9
Culture venues, embankments	4,215	12.6
Sports facilities, infrastructure	5,310	15.9
Retail facilities, shopping malls, casinos	1,395	4.2
Water parks, dolphinariums, aquariums	2,560	7.6
Other	260	0.8

Table 1. Continued

By district		
Krasnaya Polyana	10,445	31.2
Adler	13,285	39.7
Khosta	5,260	15.7
Central	3,480	10.4
Lazarevsky	1,005	3.0
Segmentation of reviewers by their place of residence	No. of reviews	%
Central Federal District	3,135	11.8
Northwestern Federal District	2,562	9.7
Southern Federal District	4,711	17.8
North Caucasian Federal District	362	1.4
Volga Federal District	2,190	8.3
Ural federal district	1,363	5.2
Siberian Federal District	1,140	4.3
Far Eastern Federal District	649	2.5
Foreign state	1283	4.8
Moscow	9068	34.3

Based on the number of the tourists' reviews, grouped by thematic criterion, the leading attractions are theme amusement parks – 21.9%, followed (in popularity) by natural and man-made attractions – 19.1% and ski resorts – 17.9%. Visitors to the city contributed the largest number of reviews of the attractions of the Adler district – 39.7%, Krasnaya Polyana – 31.2%, and Khosta district – 15.7%. One third of all comments was made by Moscow residents (34.3%), one fifth – by the residents of the Southern Federal District (17.8%), and one eighth – by the residents of the Central Federal District (11.8%).

Based on this segmentation of reviews, we've analyzed the experiences of the visitors to the Sochi resort destination by two indicators: the overall index (compound) and the negative connotation (Tables 2, 3, and 4).

Table 2. Deliverables of the sentiment analysis of reviews of 58 Sochi attractions, grouped by category

Category	Compound score			Negative connotation		
	No.	Average	Standard deviation	No.	Average	Standard deviation
Alpine ski resorts	5,824	0.637*	0.402	2,211	0.075	0.050
Nature and man-made attractions	6,233	0.548	0.465	3,169	0.082	0.055
Theme amusement parks	7,148	0.593	0.429	3,250	0.075	0.048
Culture venues, embankments	4,063	0.602	0.407	1,571	0.077	0.049
Sports facilities, infrastructure	5,104	0.599	0.415	1,915	0.081	0.054
Retail facilities, shopping malls, casinos	1,353	0.504	0.483	632	0.084	0.060
Water parks, dolphinariums, aquariums	2,482	0.406*	0.566	1,329	0.093	0.061
Other	248	0.531	0.467	118	0.076	0.047
TOTAL	32,455	0.576	0.445	14,195	0.080	0.053

Notes: One-way ANOVA test: $F = 83.184$, $p = 0.000$; * - $p < 0.05$ using Games-Howell test

The data in the table show that the highest summary index (compound score – 0.637) and the lowest negative connotation (0.075) were received by alpine ski resorts. The group,

which included water parks, dolphinariums and aquariums, received the lowest compound score, 0.406, and the largest negative connotation, 0.093.

The deliverables of the sentiment analysis of the reviews by the city districts showed (Table 3) that the reviewers were most satisfied with Krasnaya Polyana (compound score – 0.608) and Adler district (compound score – 0.575); least of all were they satisfied with the Lazarevsky district (compound score – 0.372), which also demonstrated stronger negative connotation compared to Krasnaya Polyana and Adler district (0.106 and 0.078, respectively).

Table 3. Deliverables of sentiment analysis of reviews of 58 Sochi attractions, grouped by the city districts

Group	Compound score			Negative connotation		
	No.	Average	Standard deviation	No.	Average	Standard deviation
Krasnaya Polyana	10,154	0.608*	0.427	4,027	0.078	0.051
Adler	12,834	0.575	0.437	5,637	0.078	0.051
Khosta	5,139	0.562	0.453	2,532	0.078	0.052
Central	3,353	0.563	0.447	1,464	0.086	0.055
Lazarevsky	975	0.372*	0.601	535	0.106	0.067
TOTAL	32,455	0.576	0.445	14,195	0.080	0.053

Notes: One-way ANOVA test: $F = 66,952$, $p = 0.000$; * - $p < 0.05$ using Games-Howell test

The greatest satisfaction with the Sochi resort (Table 4) was noted by residents of the Southern Federal District (compound score – 0.612), the least – by residents of the North Caucasian Federal District (compound score – 0.548); the latter showed the highest negative connotation (0.082). The reviews of tourists from the Ural Federal District contain the least negative assessments (0.073).

Table 4. Deliverables of the analysis of the satisfaction of visitors to the Sochi resort, residing in different regions of Russia and the world

Reviewer's place of residence	Compound score			Negative connotation		
	No.	Average	Standard deviation	No.	Average	Standard deviation
Central Federal District	3,035	0.591	0.430	1,312	0.076	0.049
Northwestern Federal District	2,474	0.556	0.453	1,118	0.080	0.052
Southern Federal District	4,582	0.612*	0.426	1,871	0.079	0.052
North Caucasian Federal District	353	0.548*	0.460	158	0.082	0.058
Volga Federal District	2,116	0.582	0.427	919	0.078	0.050
Ural federal district	1,325	0.575	0.444	590	0.073	0.046
Siberian Federal District	1,093	0.590	0.412	449	0.079	0.052
Far Eastern Federal District	623	0.596	0.423	272	0.076	0.044
Foreign state	1,246	0.568	0.444	554	0.081	0.055
Moscow	8,787	0.580	0.438	3,835	0.079	0.052
TOTAL	25,634	0.584	0.435	11,078	0.078	0.051

Notes: One-way ANOVA test: $F = 66,952$, $p = 0.000$; * - $p < 0.05$ using Games-Howell test

After the 2014 Winter Olympic Games, Sochi has become very popular among Russian travelers. The number of TripAdvisor reviews of the resort has skyrocketed with the creation of the pages dedicated to the Olympic venues. This is also confirmed in Figure 1 (the Y-axis indicates the number of the tourists' reviews). The figure shows that the number of reviews in 2015–2017 has increased significantly compared to 2012–2014, which positively correlates with the growth rates of tourist arrivals during this period: in 2014, the

city was visited by 4.6 million people, in 2015 – 5 million, in 2016 – 5,2 million, and in 2017 – 6 million [17].

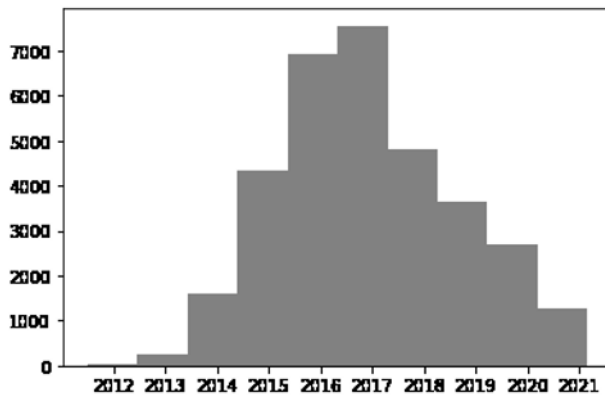


Fig. 1. Distribution of the reviews from 2012 to 2021

In order to visualize the most common words in 33,475 tourist reviews, we've created a cloud using the wordcloud package (Fig. 2). The almost complete absence of negative words confirms the positive attitude of the reviews authors.



Fig. 2. Word Cloud

5 Discussion

Sentiment analysis of the reviews of visitors to Sochi showed their positive attitude towards the city. The tourists rated alpine ski resorts, as well as cultural venues and embankments better than all other attractions (compound score 0.637 and 0.602, respectively). Resort guests were less satisfied with water parks, dolphinariums, and aquariums (compound score – 0.406).

Although the Lazarevsky district is the largest in the city (79.59 km²) and, for example, it exceeds the area of the Adler district together with Krasnaya Polyana (35.25 km²) more than twice, visitors contributed the least number of reviews of this district (1,005 and 23,730, respectively) which proves lesser popularity of the district, which is confirmed by its lower compound score (0.372) compared to the Adler district (0.575) and Krasnaya Polyana (0.608) and higher negative connotation in the reviews of the attractions of the Lazarevsky region (0.106) vs. 0.078 for the Adler district and Krasnaya Polyana.

The XXII Olympic Winter Games boosted interest in the Sochi resort, which caused an increase in the number of tourist arrivals (Vetitnev et al., 2021). Figure 1 shows a skyrocketed number of tourists' reviews of Sochi after 2014 compared to, for example, pre-Olympic 2012. A joint study by the social media reputation agency Digital Guru and Brand Analytics [14] also showed that the number of positive messages about winter vacations in Krasnaya Polyana increased significantly.

In the 33,475 reviews of Sochi sights studied, words with positive connotation prevailed (Figure 2). The tourists' reviews of the resort city, which have positive connotation, encourage potential travelers to visit the Sochi resort destination.

6 Conclusion

For tourism researchers, the Internet, as a global network that allows its users to exchange information, serves as the most important and objective source of information.

Our study confirmed the possibility to get a general idea of the Sochi resort image using sentiment analysis of the visitors' reviews: the positive image of the city is characterized by the prevalence of positive reviews and comments of the tourists over negative ones. The attractiveness of Sochi as an area with a positive image encourages excessive tourism, which may cause certain negative consequences for the Sochi resort destination. This conclusion, drawn, in turn, on the base of sentiment analysis of the visitors reviews can be considered the practical relevance of this study, and the method of sentiment analysis is a tool for assessing the negative impact of tourism on the destination.

The practical significance of this work also lies in the fact that it has shown the efficiency of sentiment analysis of the texts in comparison with sociological surveys and other traditional methods of studying public opinion. Besides considerable saving in time, money, and labor resources, the advantage of this method was manifested in the speed of the study, the possibility of processing a large number of reviews, and, most importantly, in determining their emotional substance.

However, the following limitations should be noted: the general sentiment of the review could be influenced by certain factors that distort the meaning of the text, such as, for example, incorrect translation of some words, the impossibility of recognizing ironic and sarcastic statements, the presence of many negatives in the text, etc. All this could change the statement's sentiment to the opposite. The use of just one website, TripAdvisor, has significantly reduced the scope of the study which could have been expanded if using other travel websites and social networks.

In the future, in order to determine the main reasons for the attractiveness of tourist areas, it is recommended to apply other methods, for example, topic modeling. It is also possible to expand the empirical base of the study with such TripAdvisor items as hotels, restaurants, and entertainment venues.

Therefore, sentiment analysis of tourist reviews could become both an alternative and a supplement to the study of the experiences of tourist destinations visitors.

References

1. A. Ahani, M. Nilashi, O. Ibrahim, L. Sanzogni, S. Weaven, *International Journal of Hospitality Management*, **80**, 52-77 (2019).
2. C. Bhadane, H. Dalal, H. Doshi, *Procedia Computer Science*, **45**, 808-814 (2015).
3. I. Egresi, *Proceedings of the TAW2018 International Scientific Conference*, Tiranë, Albania, from 20th to 23rd September 2018 (Albania: POLIS University, 2018), 701-714.

4. M. Geetha, P. Singha, S. Sinha, *Tourism Management*, **61**, 43-54 (2017).
5. M. González-Rodríguez, R. Martínez-Torres, S.L. Toral, *International Journal of Contemporary Hospitality Management*, **28**, 2609-2627 (2016).
6. V. Gonzalez, L. Coromina, N. Galí, *Tourism Review*, **73**, 277-296 (2018).
7. A. P. Kirilenko, S. O. Stepchenkova, *Tourism Management*, **63**, 54-65 (2017).
8. A. P. Kirilenko, S. O. Stepchenkova, J. M. Hernandez, *Tourism Management*, **72**, 400-410 (2019).
9. S. Kladou, E. Mavragani, *Journal of Destination Marketing & Management* (2015).
10. E. Marine-Roig, *Sustainability*, **9**, 1425 (2017).
11. M. Martins, *Journal of Tourism, Heritage & Services Marketing*, **4(1)**, 3-7 (2018).
12. P. L. Pearce, Bristol, Blue Ridge Summit: Channel View Publications (2005).
13. J. A. Pesonen, A. Tuohino, *Journal of Vacation Marketing*, **23(2)**, 145–158 (2017).
14. RBC, Nice, but ... What people like and dislike about Krasnaya Polyana (2018) <https://kuban.rbc.ru/>
15. A. M. Sazykin, A. A. Glushko, *Service and Tourism: Current Challenges*, **11(2)**, 7–16 (2017).
16. A. Seilsepour, R. Ravanmehr, H. Sima, *Journal of Advances in Computer Engineering and Technology*, **5(3)**, 143-160 (2019).
17. Sochi Municipal Administration, Socio-economic development of the city of Sochi (2021) <https://sochi.ru/>
18. Y. Su, W. Teng, *Tourism Management*, **69**, 214-222 (2018).
19. L. G. Tamajón, G. C. Valiente, *Current Issues in Tourism* (2015).
20. A. Therkelsen, H. Halkier, *Umbrella Place Branding. A Study of Friendly Exoticism and Exotic Friendliness in Coordinated National Tourism and Investment Promotion*, SPIRIT (2004).
21. A. M. Vetitnev, D. V. Chigarev, *Journal of Environmental Management and Tourism*, **12(1)**, 218-232 (2021).
22. A. M. Vetitnev, D. V. Chigarev, O. E. Afanasiev, *European Journal of Tourism Research*, **28**, 2812 (2021).
23. J. Wu, *Sentiment Analysis for Hotel Reviews* (2019) <https://www.octoparse.com/>
24. Y. Yang, S. Park, X. Hu, *Tourism Management*, **67**, 248-260 (2018).