

Assessing Travellers Perception Levels towards Select Online Travel Agents (TA) in India using Netnography and Text Mining

Dashrath Mane, Dr. Prateek Srivastava, Dr. Amit Jain, Dr. D S Chouhan

Department of CSE, School of Engineering, Sir Padampat Singhanian University, INDIA

Abstract - This paper explores the overall effect of the use of online services by travelers and their feelings resulting from a series of feedback. Customer reviews are obtained from various sources and the satisfactions of travelers using different online travel related service providers is analyzed. A new term is coined and referred to as "Netno-Mining," which integrates two words Text mining and Netnography. Different results obtained by processing netnographic data and text based mining techniques are used to interpret satisfaction, perception of tourist/ travellers using online services. This paper covers perception derived by analysis of review rating values and sentiment-based levels.

Keywords: OTA, Netnography, Text mining, perceptions, traveller

1. Introduction

In India's economy, the tourism industry plays an important role (bhayani & rana , 2016) . India's foreign exchange earnings and GDP increased during 2007-2008. Hotels are a major service industry that in terms of foreign exchange earnings, contributes most to tourism and also helps to increase India's GDP. Tourism also plays a vital role in this growing trend in GDP & foreign exchange based earnings in India. Due to the development of greater jobs opportunities, the tourism industry has invented more space. Even then it must be very seriously considered to conquer this hotel industry, which faces many problems. The current situation shows that India has a promising future because of the phenomenal growth in the tourism industry. (bhayani & rana , 2016)

2. OTA Market in India

The scale of the online booking industry is expected to be approximately 4 billion in India by 2020. The supplier environment is expanding with new startups such as Oyo, which combine hotel chain delivery with the agility of online travel agencies. ("Top Online Travel Agencies (OTAs) in India - THE 'PERFECT' HOTEL," n.d.)

In the year 2016, Ibibo & MakeMyTrip , the two biggest online travel agency(OTA) groups in India, merged to create MMT and Goibibo. Paytm again entered the OTA market with a bang after their initial venture in 2015 did not take off. For eg, Expedia & Booking.com Trail Goibibo-Make My Trip are international giants in the domestic market. In this list below, we have only included the domestic Indian OTAs and omitted the other four major international OTAs: Airbnb, Booking.com, Agoda, and Expedia which function efficiently in the Indian online travel industry.

3. Analysis of Average sentiment to rating

This research uses the reviews collected from various online platforms including mouthshut.com and other rating and referral sites. all the reviews collected for online travel service providers which includes Goibibo ,makemytrip, cleartrip ,yatra and red bus . all the reviews are analysed using netnographic analysis and sentiment analysis .each of the reviews having the rating values associated with ranging from 1 to 5. To understand the perception of travellers various sentiment analysis procedures associated with rating values for each of the reviews collected, results obtained are presented in this section .

Figure 1 represents the line of code used to perform the average sentiment to review ratings.

```
rating_sentiments = reviews_df.groupby('rating').mean()["compound"]

ratings=[1,2,3,4,5]
sentiment=[rating_sentiments[i] for i in range(1,6) ]
xpos = np.arange(1,len(ratings)+1)
plt.bar(xpos,sentiment, label="sentiments")

plt.xlabel("Rating")
plt.ylabel("Average sentiment")
plt.savefig("Overall Average sentiment to review rating", dpi = 200)
plt.legend()
```

Figure 1 Average sentiment to review rating -LOC

4. Review Analysis Based on Rating & Visualization

Figure 2 illustrates different steps in the method used for research depending on the ranking of the study. This includes information about online travel service providers and unique values for different scores. There will be removal of the empty rows and cleaning and percentage values for rating are determined and filtering is performed based on rating and OTA values. Using different graphs, the results obtained from these measures are graphed.

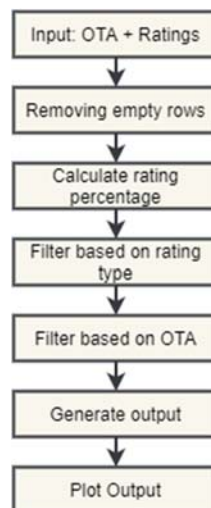


Figure 2 Process used for review analysis based on rating values

Table 1 represents statistical values for various online travel agents used in the studies and the level-wise classification of reviews and percentage values for the same. Based on the value for each of the level from level 1 to level 5 which are mapped to greeting values ranging from 125 respectively, this helps us to understand the volume of reviews belongs to each of the levels and hence satisfaction and dissatisfaction can be measured in percentage for each of the online travel service providers.

Table 1 OTA wise Frequency of Rating

Frequency of rating											
OTA	Level 1	%	Level 2	%	Level 3	%	Level 4	%	Level 5	%	Total
ClearTrip	236	67.2	13	3.7	12	9.4	33	9.4	57	16.2	351
Goibibo	321	61.8	33	6.4	36	10.4	54	10.4	75	14.5	432
MMT	281	52.9	18	3.4	48	16.9	90	16.9	94	17.7	531
Redbus	292	62.3	38	8.1	45	9.8	46	9.8	48	10.2	469
Yatra	168	64.9	16	6.4	15	15.1	38	15.1	21	7.9	258
	1298	63.59	85	4.16	156	7.64	207	10.14	295	14.45	2041

Various rating values ranging from 1 to 5 are shown in Figure 3-4, where 1 is low and 5 is high for clear-trip. It is quite clear from the line chart shown that a large volume of reviews belongs to level 1 with rating value 1 which reflects the frustration of travellers using different services provided by clear ride, whereas the significantly low volume of reviews are identified for level 2 and 3 and moderate for 4 and 5 while the volume of reviews are identified that are closer to the satisfaction of users.

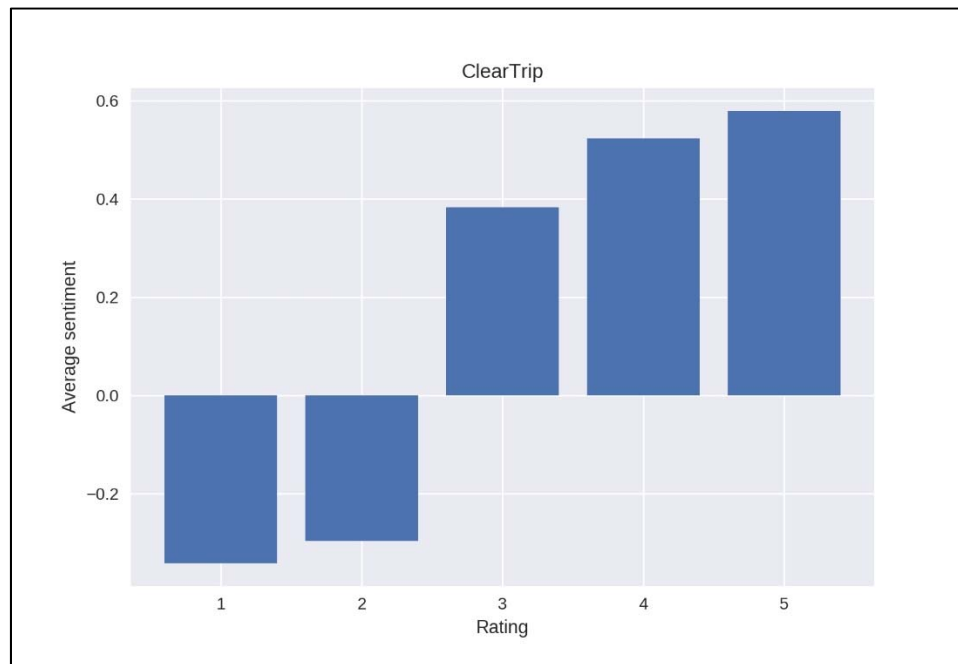


Figure 3 Cleartrip average sentiment to review rating

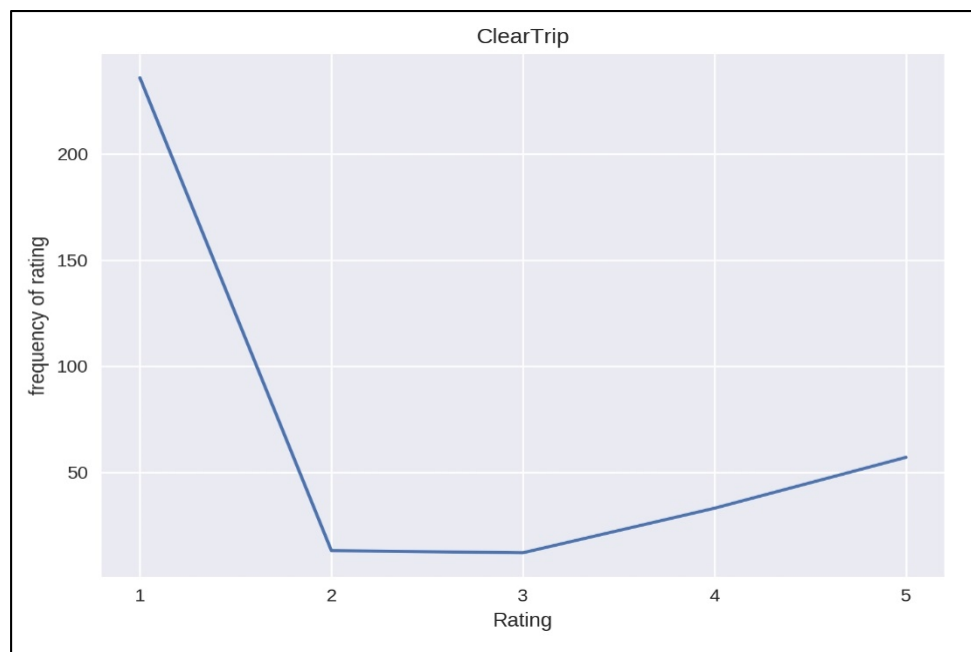


Figure 4 Cleartrip review frequency to review rating

The figure5-6 indicates various rating values from 1 to 5, with 1 being low and 5 being high. From the line chart shown, it is also very likely that a large volume of reviews belongs to level 1 with a rating value of 1, which suggests disappointment of travellers using different services offered by goibibo, while a significantly small volume of reviews is identified for level 2 and 3 and a modest volume of reviews that are significantly closer to that is identified for levels 4 and 5. The volume of Level 5 here reflects the potential number of extremely happy travellers.

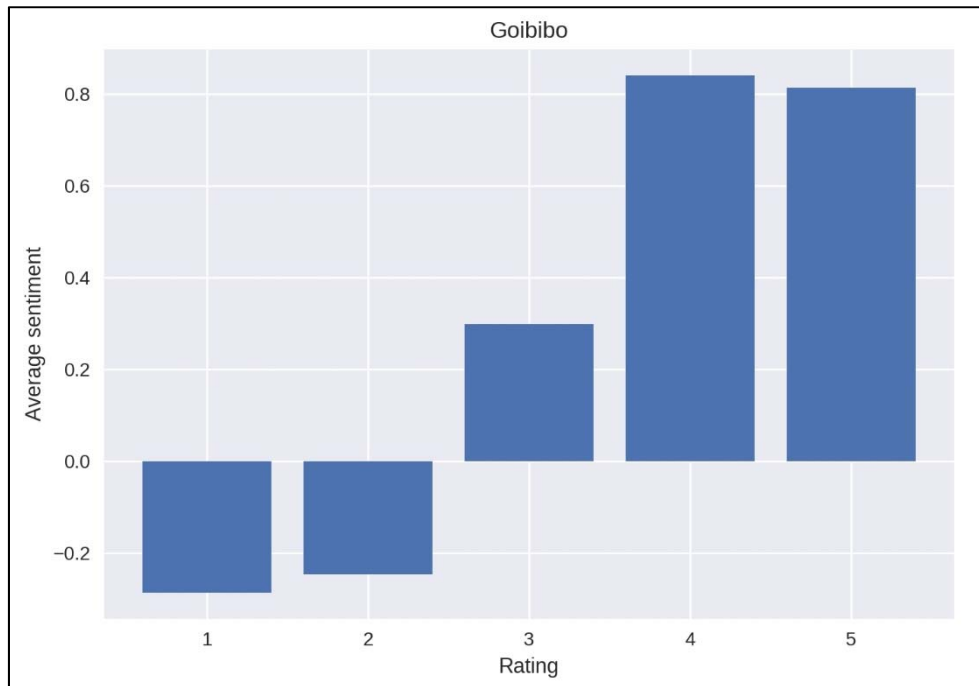


Figure 5 Goibibo average sentiment to review rating

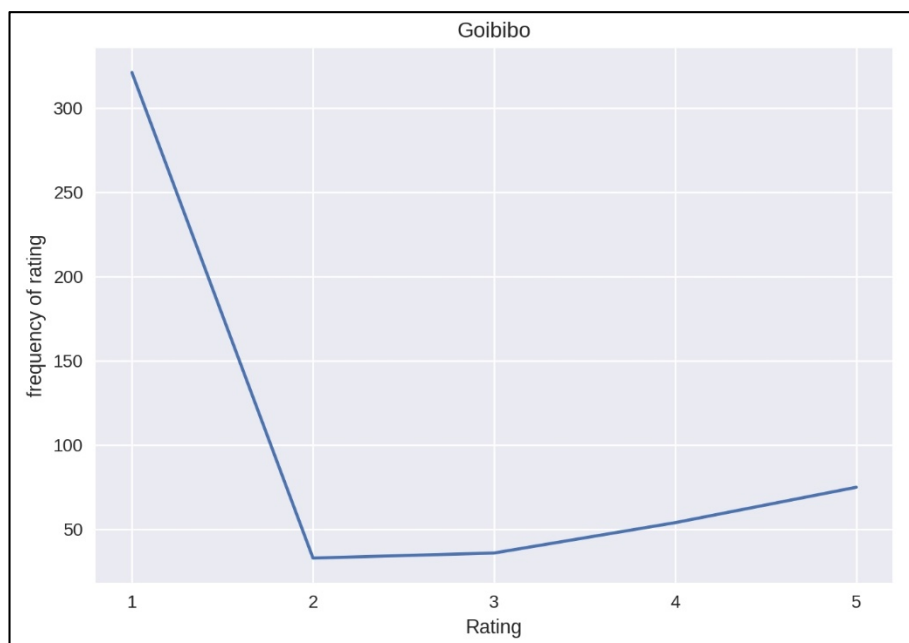


Figure 6 Goibibo review frequency to review rating

The figure 7-8 shows various rating values ranging from 1 to 5, with 1 being low and 5 for MMT being high. It is very clear from the line graph seen that a large volume of reviews belongs to Level 1 with a rating value of 1, which means that travellers are dissatisfied with the various services provided by MMT, while a considerably small volume of reviews is identified for Level 2 and Level 3 and a modest volume of reviews is identified for Level 4 and Level 5, which are far closer to that level.

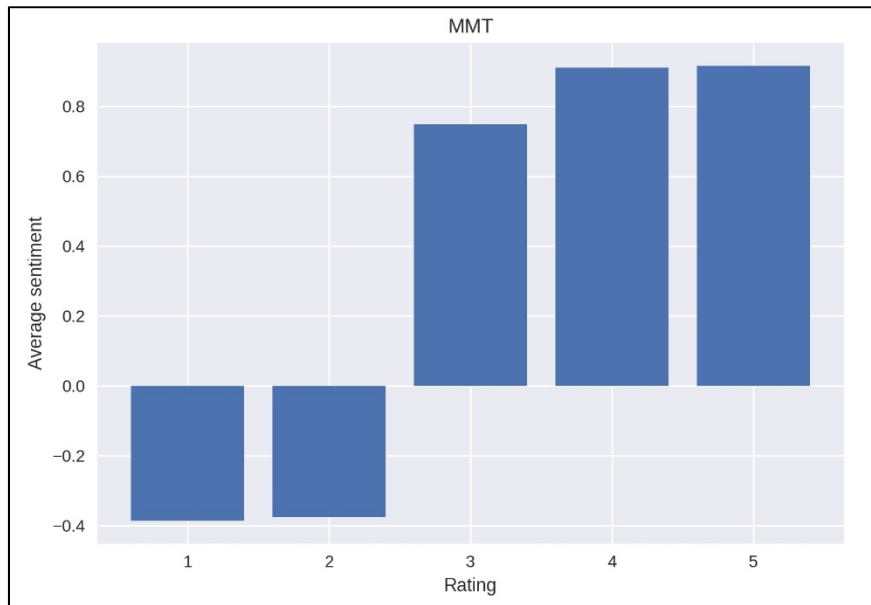


Figure 7 MMT average sentiment to review rating

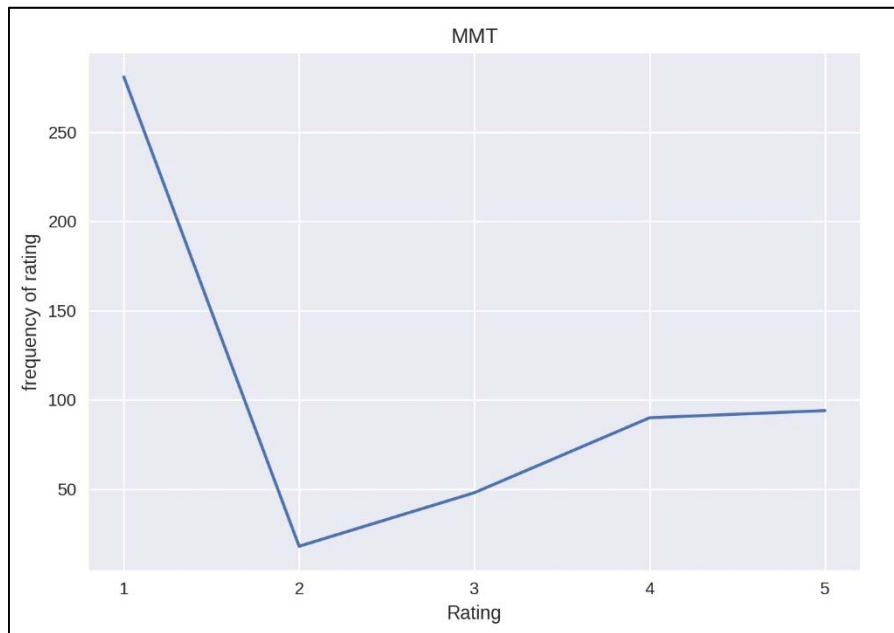


Figure 8 MMT review frequency to review rating

Figure 9-10 indicates various rating values, ranging from 1 to 5, 1 low and 5 high. It is very clear from the line graph seen that a significant number of reviews belongs to level 1 with a rating value of 1, which means that of travellers are dissatisfied with the different services offered by red bus, while level 2 has a moderate volume, while level 3,4,5 has a comparatively small volume of reviews. This trend differs very little from other OTAs. For various services rendered by this OTA, low volumes in 5 states, low satisfaction counts and high volumes in level 1 reflect dissatisfaction.

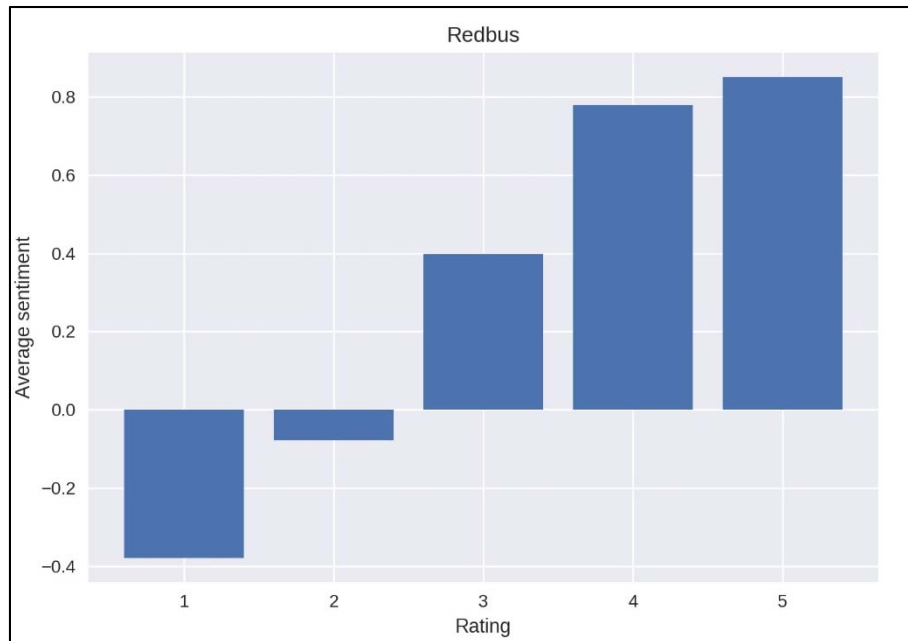


Figure 9 Redbus average sentiment to review rating

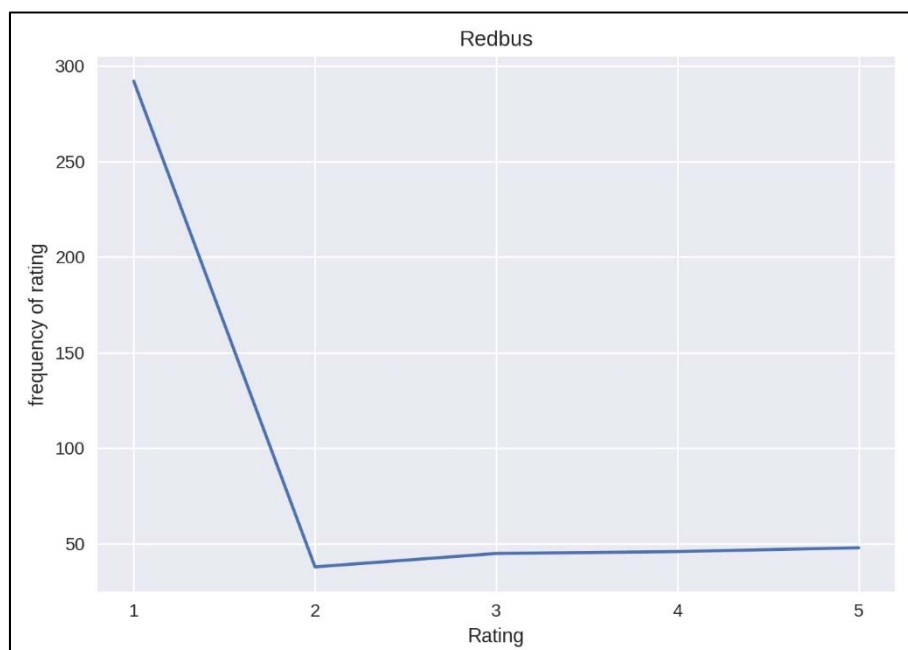


Figure 10 Redbus review frequency to review rating

The figure 11-12 indicates various rating values, ranging from 1 to 5, 1 low and 5 high. It is very clear from the line graph shown that a large volume of reviews belongs to level 1 with a rating value of 1, which means that travellers using different yatra services are dissatisfied, whereas level 4 has a moderate volume, whereas level 2.3 has a considerably small volume of reviews. For various services rendered by this OTA, low volumes in 5 states and high volumes in 1 state are low and unhappy with different services.

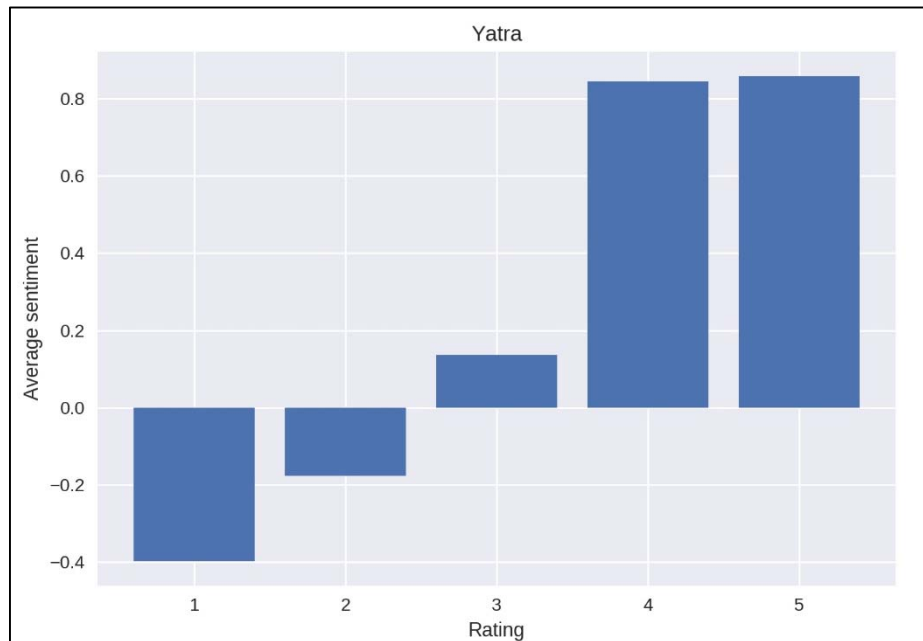


Figure 11 yatra average sentiment to review rating

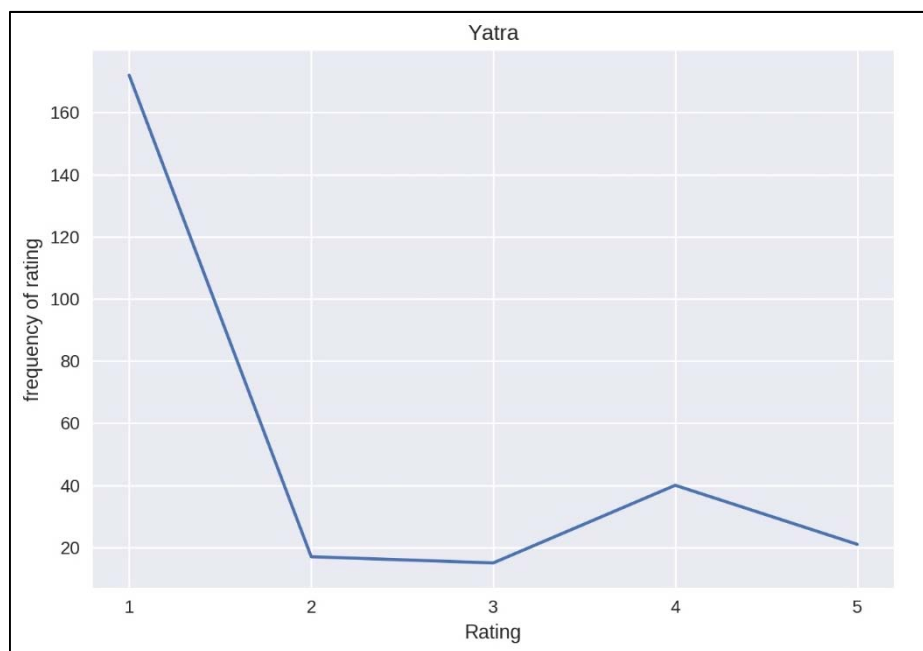


Figure 12 yatra review frequency to review rating

Figure 13 below represents line of code from Python used for the purpose of deriving positive negative ratio from reviews.

```
posm = reviews_df[(reviews_df["compound"] >= 0) & (reviews_df["gender"] == "male")]
posm.groupby('OTA')['OTA'].apply(list)
posf = reviews_df[(reviews_df["compound"] >= 0) & (reviews_df["gender"] == "female")]
posf.groupby('OTA')['OTA'].apply(list)
negm = reviews_df[(reviews_df["compound"] < 0) & (reviews_df["gender"] == "male")]
negm.groupby('OTA')['OTA'].apply(list)
negf = reviews_df[(reviews_df["compound"] < 0) & (reviews_df["gender"] == "female")]
negf.groupby('OTA')['OTA'].apply(list)
xpos = np.arange(len(OTA))
plt.bar(xpos-0.4, posm, width=0.2, label="Positive Male")
plt.bar(xpos-0.2, posf, width=0.2, label="Positive Female")
plt.bar(xpos, negm, width=0.2, label="Negative Male")
plt.bar(xpos+0.2, negf, width=0.2, label="Negative Female")
plt.xticks(xpos, OTA)
plt.ylabel("Number of reviews")
plt.xlabel("OTA")
plt.title('Sentiment of reviews')
plt.legend()
plt.savefig("Sentiment of reviews with gender.png", dpi = 300)
plt.show()
```

Figure 13 Review positive negative ratio -LOC

Table 2 represents these statistical values for online travel agents and corresponding positive negative review volume. This basically helps us to understand the ratio off positive and negative sentiments for each of the online travel service providers used in this study.

Table 2 OTA wise Sentiment ratio count

OTA	Positive	Negative	Total
ClearTrip	178	208	386
Goibibo	308	286	594
MMT	301	231	532
Redbus	225	244	469
Yatra	122	144	266

Figure 14 represents the sentiment of reviews as positive and negative for all 5 online travel service providers that make my trip, Goibibo, cleartrip, yatra and red bus .from the graphical representation it is very much clear that the volume of reviews in positive and negative category is almost balanced. It means there is no scope for biased kind of interpretation of results

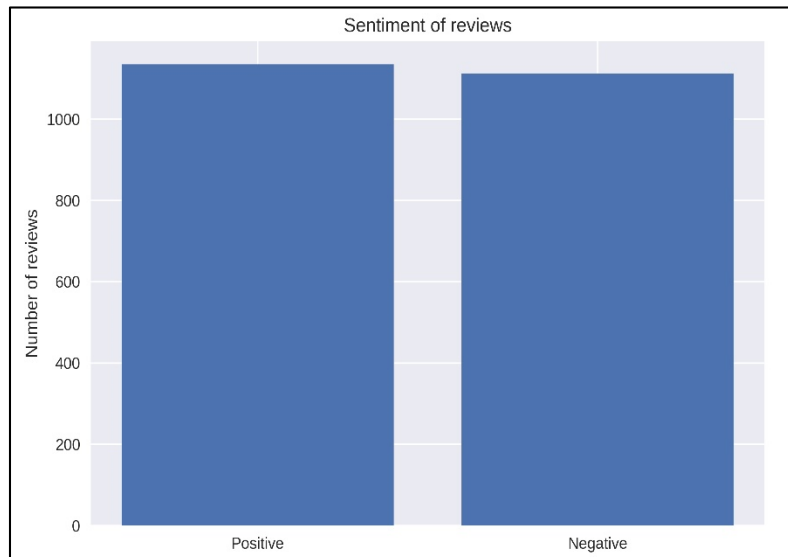


Figure 14. Sentiment Ratio

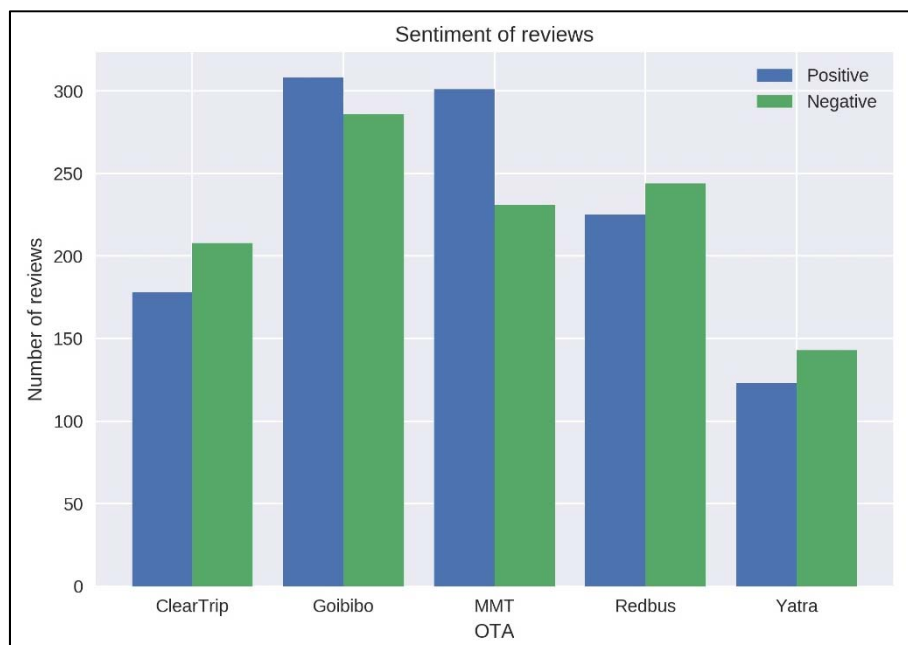


Figure 15 OTA specific sentiment ratios of reviews

Figure 15 graphically represents online travel agent specific sentiment ratios of reviews, from the representation shown we can check out on the ratio of positive and negative reviews for each of the travel agent and it is very much clear that for make my trip and goibibo the overall positive sentiment count is more than negative sentiment count whereas, for cleartrip, redbus and yatra the negative sentiment values or count is more than positive sentiment. when positive sentiment count is more than negative sentiment count, we can conclude that travellers satisfaction is more there then rest of that travel service providers.

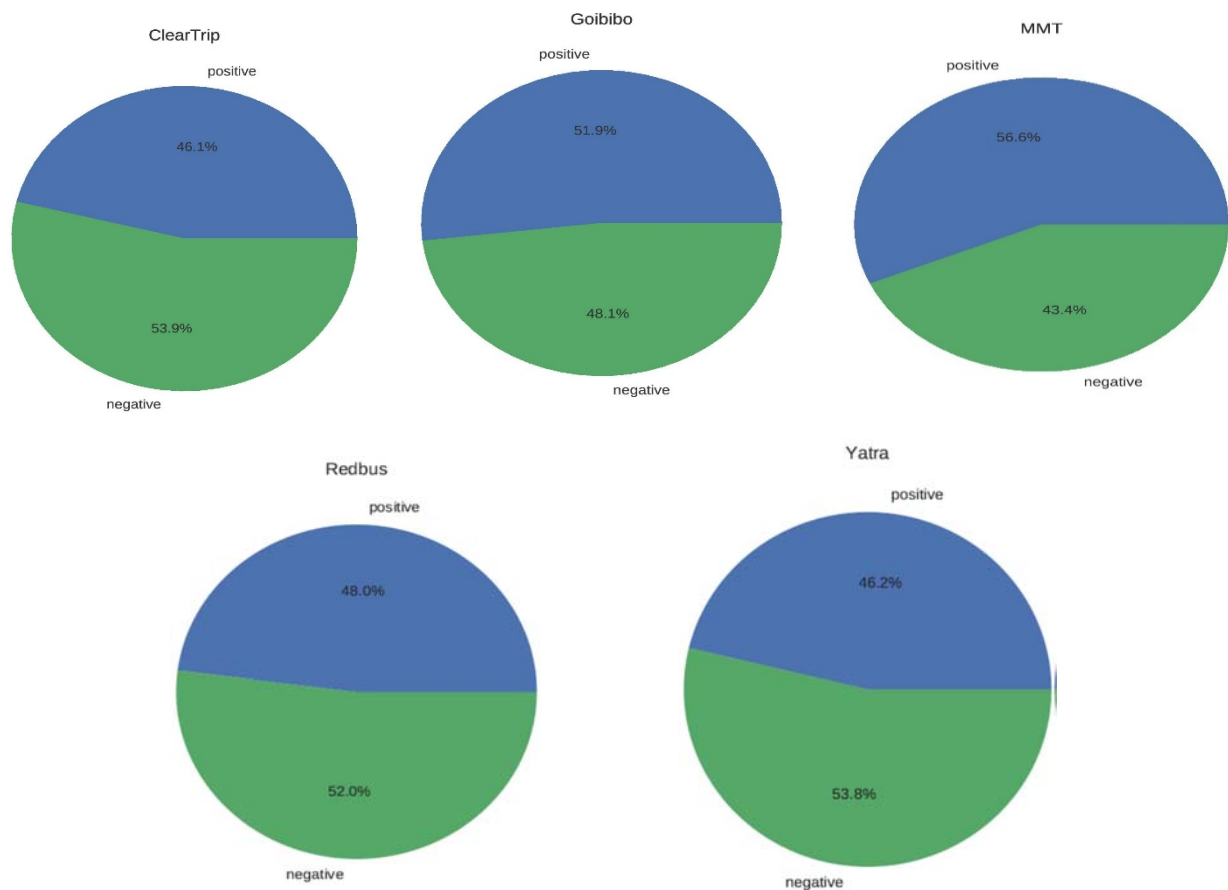


Figure 16 Positive negative sentiments for OTA

Figure 16 represents OTA wise positive negative sentiments in the form of the pie chart. From the representation we can interpret that for cleartrip positive sentiment percentage is 46.1%, for Goibibo it is 51.9%, for MMT it is 56.6%, similarly for red bus it is 48% and for yatra it is 46.2%. Hence make my trip is carrying the highest positive sentiment percentage, then comes Goibibo then comes red bus then comes yatras and lastly cleartrip.

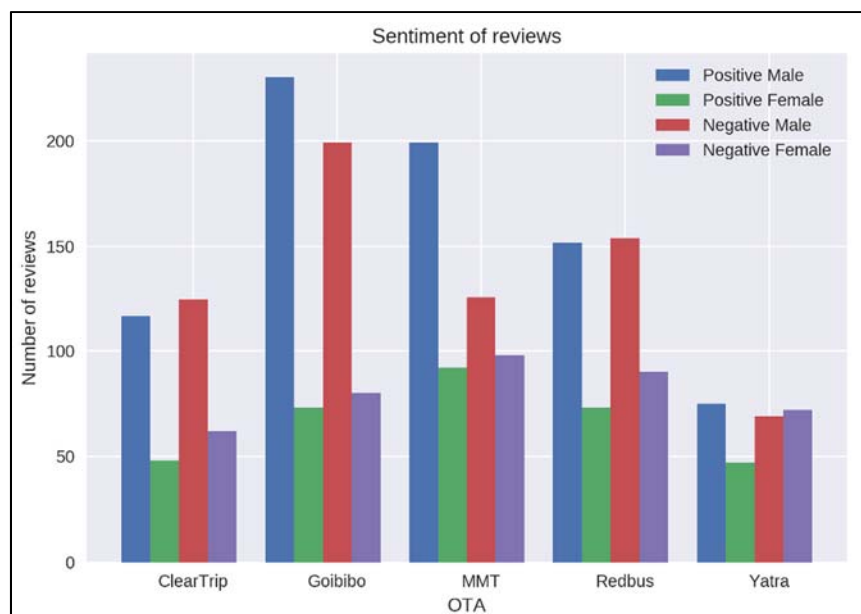


Figure 17 Gender Specific sentiment of reviews

Figure 17, Overall from the results obtained males are more satisfied with the various services offered by online travel service providers than female. Almost for all the online travel service providers, the positive percentage values for Male are higher than female.

5. Perception analysis based on Sentiment Levels

This section covers, various results obtained by analysing sentiment of reviews and it's classification into various sentiment levels. For each of the reviews collected a corresponding review rating and but identifying this sentiment score for each of the review and mapping that into the various levels called as sentiment levels all the reviews based on the sentiment score are being classified into level one to level 5 the statistical values obtained as a part of this are shown in Table 3

Table 3 Classification of sentiment levels

Sentiment Levels		level 1	level 2	level 3	level 4	level 5
OTA	ClearTrip	120	45	42	50	74
	Goibibo	164	57	44	56	199
	MMT	154	30	27	18	254
	Redbus	172	31	23	38	147
	Yatra	96	21	14	20	79

figure 18 below represents various line of code used from Python to derive sentiment scores from the review and classification of reviews into various levels based on sentiment values, this analysis is done for all the five online travel service providers collectively and also individually.

```

leveln=[ "level"+str(i) for i in range(1,6)]
OTAs = ['ClearTrip', 'Goibibo', 'MMT', 'Redbus', 'Yatra']
for OTA in OTAs:
    levels = []
    reviews_OTA_df = reviews_df[reviews_df['OTA'] == OTA]
    for i in range(1,6):
        levels.append(len(reviews_OTA_df[reviews_OTA_df["level"] == "level"+str(i) ]))

    xpos = np. arange(len(levels))
    plt. bar(leveln,levels)

    plt. xticks(xpos,leveln)
    plt. ylabel("Number of reviews")
    plt. xlabel("Levels based on sentiment of each review")

    plt.title(OTA+' levels')
    plt.savefig(OTA+" levels", dpi = 400 )
    plt.show()

```

Figure 18 Analysis of the reviews based on sentiment values -LOC

Figure 19, systematically representing various steps carried out as a part of the process used for measuring perception based on sentiment levels. This process takes online travel service provider related information and reviews collected as an input.

All the reviews collected are cleaned as a part of textual descriptions and sentiment analysis has been carried out for all the reviews collected based on the sentiment scores for each of the review various levels are being identified and all these reviews are map are classified into level 1 to level 5 based on the values of sentiment, for each of the levels and each of the online travel service provider filters can be used to get statistical and graphical results.

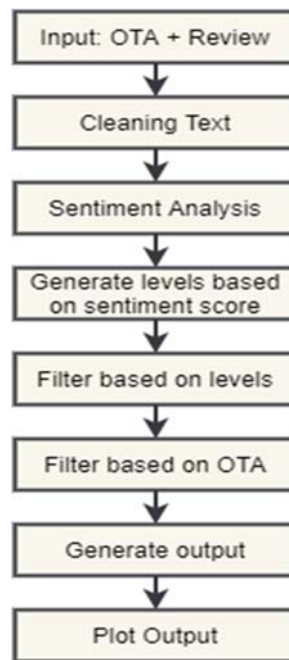


Figure 19 Process used for measuring perception using sentiment levels

Table 4 covers summary of level-specific and OTA specific volume and percentage of reviews belongs to each of the level of sentiment.

Table 4 Summary of Levels Based on Sentiment values

Sentiment Levels		Levels										Total
		1	%	2	%	3	%	4	%	5	%	
O T A	ClearTrip	120	36.25	45	13.60	42	12.69	50	15.11	74	22.36	331
	Goibibo	164	31.54	57	10.96	44	8.46	56	10.77	199	38.27	520
	MMT	154	31.88	30	6.21	27	5.59	18	3.73	254	52.59	483
	Redbus	172	41.85	31	7.54	23	5.60	38	9.25	147	35.77	411
	Yatra	96	41.74	21	9.13	14	6.09	20	8.70	79	34.35	230
Overall		706	35.75	184	9.32	150	7.59	182	9.22	753	38.13	1975

Figure 20 is a graphical representation of results opted as part of analysis carried out based on sentiment value and classification of reviews into levels of sentiments. Level 1 corresponds to extremely dissatisfied travellers, level 2 corresponds to dissatisfied travellers, and Level 3 corresponds to neutral, level 4 and 5 responds to satisfied and extremely satisfied travellers respectively.

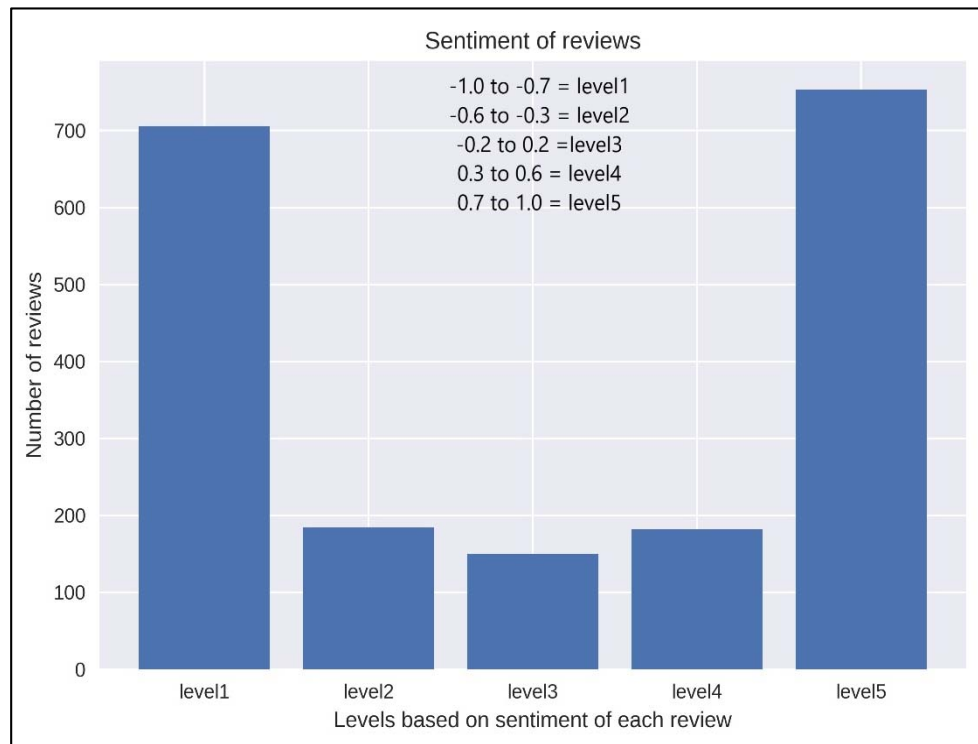


Figure 20 Analysis of levels of Sentiments for reviews

Figure 21 represents cleartrip level of sentiments for reviews, from the representation we can make out that good volume of reviews belongs to level 1 whereas level 4 and 5 are also having or moderate volume of reviews which corresponds to satisfied and extremely satisfied travellers using various services offered by ClearTrip.

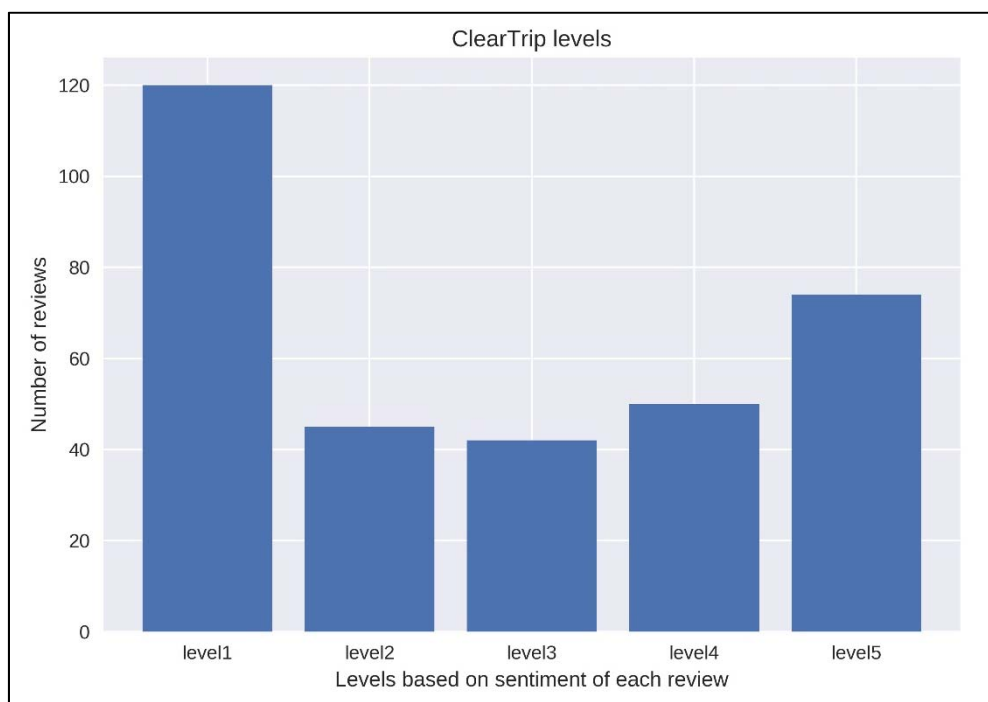


Figure 21 Cleartrip levels of Sentiments for reviews

Figure 22 represents graphically the various levels based on the sentiment of each review for Goibibo, from the representation again it is very much clear that level 5 is having extremely good count of reviews it represents the number of travellers using Goibibo services are extremely happy with the kind of services offered by Goibibo

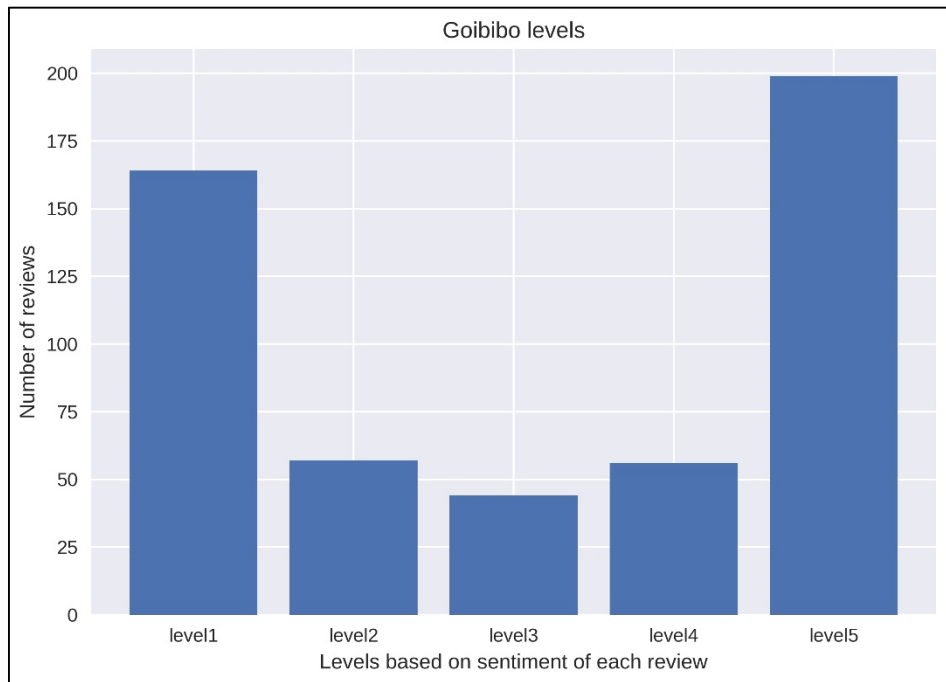


Figure 22 Goibibo levels of Sentiments for reviews

Figure 23 represents MMT levels of sentiment for reviews, from the representation we can see that a large number of reviews belongs to level 5 which represents extremely satisfied travellers using various services offered by make my trip.

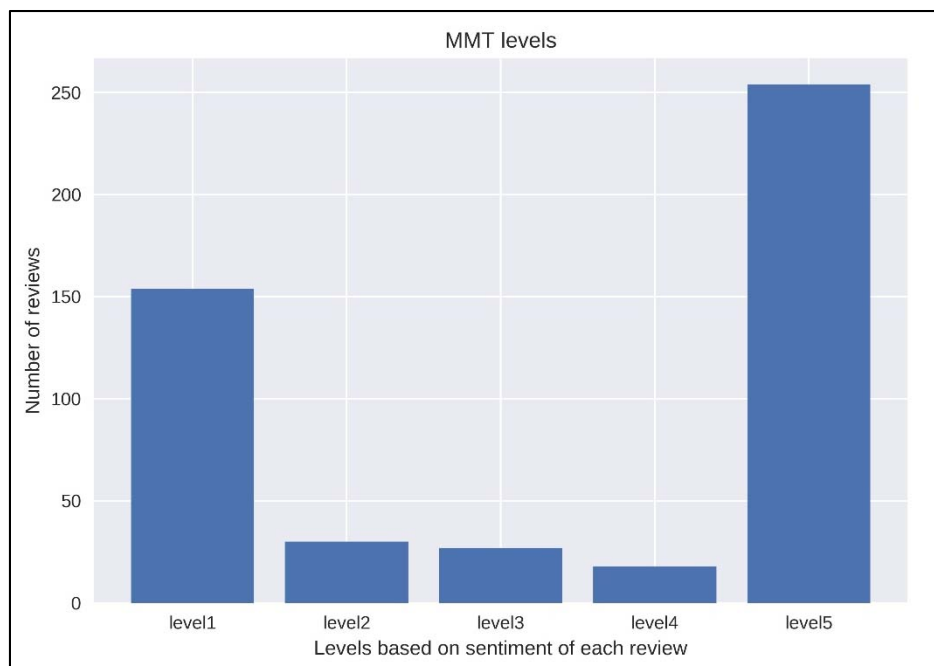


Figure 23 MMT levels of Sentiments for reviews

Figure 24 represents graphically various levels of sentiment for reviews in Redbus, From the graphical representation we can see that good number of reviews belongs to level one which means dissatisfied or extremely dissatisfied number of customers which belongs to redbus at the same time level 5 the volume of reviews which belongs to this level also represents a good amount of extremely satisfied count of customers are travellers availing services offered by red bus.

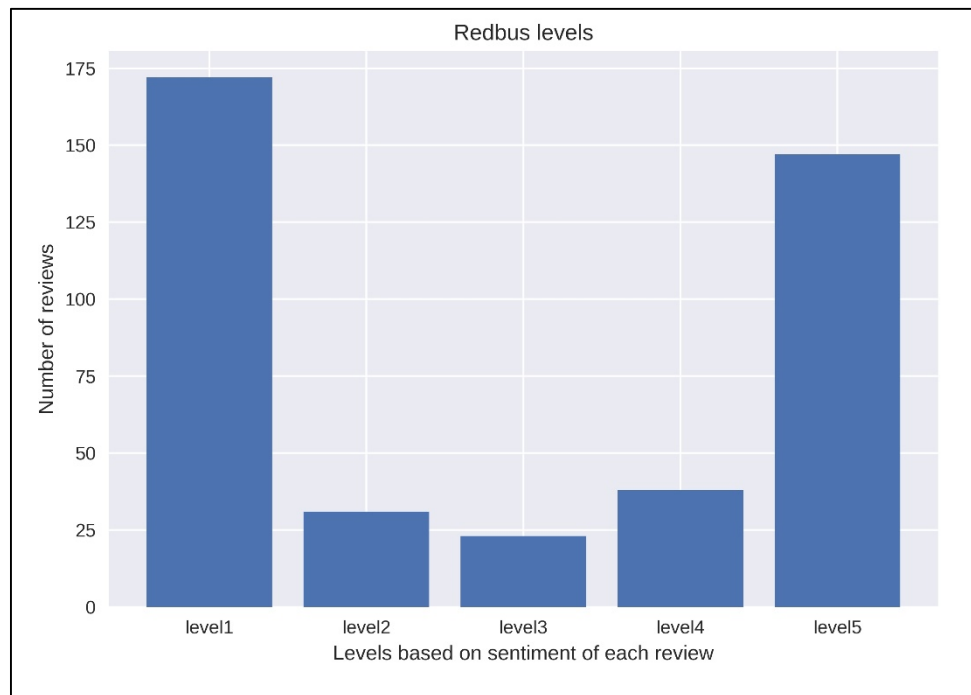


Figure 24 Redbus levels of Sentiments for reviews

Figure 25 represents yatra level of sentiments for reviews, level one is having the highest volume of reviews it means the count of travellers belongs to the extremely dissatisfied category is more, level 2,3, 4 are having Low count of reviews whereas level phase having a moderate count of reviews which represents the moderate number of travellers belongs to extremely satisfied category using services offered by Yatra

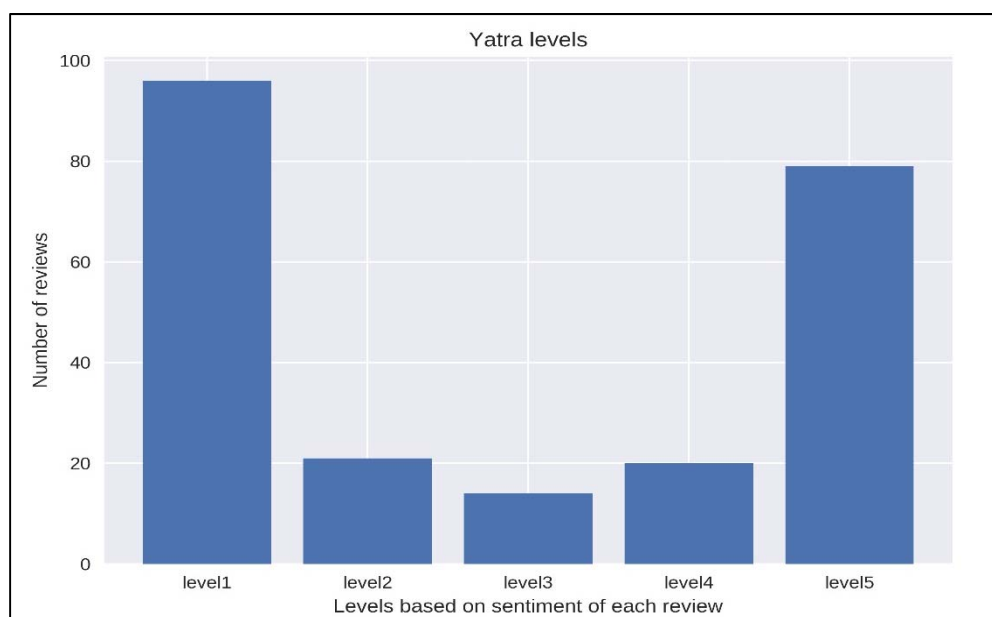


Figure 25 Yatra levels of Sentiments for reviews

6. Statistical Analysis

This section interprets or covers the statistical data obtained from the Chi-square test. The research is carried out in this scenario to assess if the variables are kind of independent or not independent. The mathematical analysis of Chi-square is used for this. Chi-square statistical hypothesis test is applied to assess if the online travel agent has any correlation with the level of sentiment.

H0 (Null hypothesis): There is no relation between levels of sentiment and individuals using Online Travel Agents

H1(Hypothesis): Correlation between levels of sentiment and individuals using Online Travel Agents

Table 5 Test Statistics for OTA and Sentiment Levels-Observed

Levels	level 1	Level 1 %	level 2	level 2 %	level 3	Level 3%	level 4	level 4 %	level 5	level 5 %
ClearTrip	120	17	45	24.46	42	28	50	27.47	74	9.83
Goibibo	164	23.23	57	30.98	44	29.33	56	30.77	199	26.43
MMT	154	21.81	30	16.3	27	18	18	9.89	254	33.73
Redbus	172	24.36	31	16.85	23	15.33	38	20.88	147	19.52
Yatra	96	13.6	21	11.41	14	9.33	20	10.99	79	10.49
	706	100	184	100	150	100	182	100	753	100

Table 6 Test Statistics for OTA and Sentiment Levels-Expected

Levels	level 1	level 2	level 3	level 4	level 5
ClearTrip	118.32	30.84	25.14	30.50	126.20
Goibibo	185.88	48.45	39.49	47.92	198.26
MMT	172.66	45.00	36.68	44.51	184.15
Redbus	146.92	38.29	31.22	37.87	156.70
Yatra	82.22	21.43	17.47	21.19	87.69

The P value resulted is 0.00. With the DOF(degrees of freedom)16, critical value or tabular value of 26.30 and at a 5 per cent significance level. The calculated value of chi-square(122.08) is greater or more than the value or tabular value (26.30). There is ample statistical proof or evidence to perform rejection on the H0 null hypothesis and support the alternative one, considering the reality of association or relationship between different OTA and sentiment values levels. $122.08 > 26.30$ or our P value < 0.05 is acknowledged as the alternative hypothesis. There is also a link between the level sentiment evaluation and OTA users. Figure 26 reflects the statistical relationship for all five travel agents between the different review sentiment levels and the online travel agent; the predicted values and the observed value is reflected. We can see in this figure that for goibibo, redbus, MMT with more relative values for rating levels from 1 to 5. Their relationship is thus reflected.

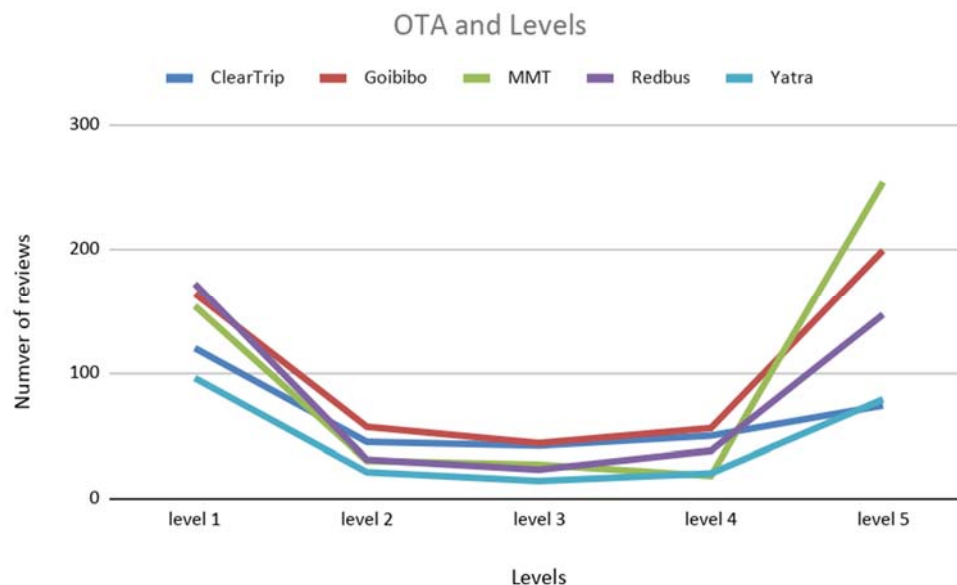


Figure 26 Statistical Relationship between Sentiment levels and OTA

7. Conclusion

Netnographic and text based mining techniques are used to analyze and process feedback and comments obtained from different online platforms. For flight booking, the satisfaction of the traveler with relevant services is greater with MMT, Goibibo and Redbus. Compared to others, the satisfaction of travelers using services offered by MMT is considerably high for hotel booking and related services. Based on the results obtained from the report, online travel based service provider in india can use this for taking appropriate decisions.

This paper helps us understand the perceptions of travellers/tourist using online services offered by MMT, Goibibo, cleartrip, redbus, yatra. The tourism sector in India is growing speedily. Advancements in the social platforms, the large volume of data generated online are available for the processing of results. Online travel agents can refer to results obtained to understand their customers' expectations.

References

- [1] Chen W, Xu Z, Zheng X, Yu Q and Luo Y. 2020. Research on sentiment classification of online travel review text. *Applied Sciences* (Switzerland), 10(15). <https://doi.org/10.3390/AP10155275>
- [2] Dashrath Mane and Bhatnagar D. 2020 Integrating Netnographic Analysis and Text Mining for Assessing Satisfaction of Travellers Visiting to India - A Review of Literature. *Learning and Analytics in Intelligent Systems*, vol 3. Springer, Cham DOI: https://doi.org/10.1007/978-3-030-24322-7_67, pp.565-572
- [3] Dashrath Mane and Prateek Srivastava. 2020. Netnography and Text Mining to Understand Perceptions of Indian Travellers using Online Travel Services. *Intl. Journal of Advanced Computer Science and Applications (IJACSA)*, 11(9):559-569.
- [4] Dashrath Mane and Srivastava P. 2020 Text Mining to Understand Major Keywords Explaining Sentiments of Travelers Using Travel Related Online Services in India. *Learning and Analytics in Intelligent Systems*, vol 16. Springer, Cham DOI: https://doi.org/10.1007/978-3-030-46943-6_32, pp.287-295.
- [5] Dashrath Mane, Dr. Prateek Srivastava, Dr. Amit Jain and Dr. D S Chouhan. 2020. Netno-Mining: Integrating Text Mining with Netnographic Analysis to Assess the Perception of Travelers using Select OTA Services in India. *International Journal of Engineering Trends and Technology* 68(9):121-129. DOI 10.14445/22315381/IJETT-V68I9P217
- [6] Dr. Sanjay J Bhayani, Sheetalba N Rana (2016). A Study on Tourist Satisfaction towards Hotel Related Services in Gujarat Tourism. 4(6), 105–110.
- [7] <http://dx.doi.org/10.14569/IJACSA.2020.0110967>
- [8] Kozinets R. 2010. Click to connect. *Netnography and tribal advertising*, *J. Advertising Research*. 46: 279-288.
- [9] Kozinets R. 2010a. Netnography: the marketer's secret ingredient, *MIT Technology Review*. Online at <http://www.technologyreview.com/business/26434/>.
- [10] Kozinets R. 2010b. Netnography: doing ethnographic research online. Sage Publication, London
- [11] Mkono M. 2012. Netnographic tourist research: The Internet as a virtual fieldwork site, *Tourism Analysis*. 17:553-555.
- [12] Top Online Travel Agencies (OTAs) in India - THE "PERFECT" HOTEL. (n.d.). Retrieved July 28, 2020, from <http://whiteskyhospitality.co.uk/top-online-travel-agencies-otas-in-india/>
- [13] Tourism Economic Impact 2017 India. (n.d). World Travel and Tourism Council. Retrieved from <https://www.wttc.org/-/media/files/reports/economic-impact-research/-countries-2017/india2017.pdf>
- [14] Tsujii K, Tsuda K and Takahashi M. 2015. Towards Extracting the Hotel Evaluations from the Comments by the Foreign Tourists with Text Mining. *Proc, IIAI 4th International Congress on Advanced Applied Informatics, Okayama, 2015*, pp.46-49