

CERTIFICATE

This is to certify that the Project Report titled “Uncovering trends in Taxi Aggregator Industry Using Twitter Sentiment Analysis”, is a bonafide work carried out by Mr. Gaurang Manchanda of MBA 2014-16 and submitted to Delhi School of Management, Delhi Technological University, Bawana Road, Delhi-42 in partial fulfillment of the requirement for the award of the Degree of Masters of Business Administration.

Signature of Guide

(Dr. Rajan Yadav)

Signature of Head (DSM)

Signature of Guide

(Mr. Anuraag Tiruwa)

Place: New Delhi

Date:

DECLARATION

I, Gaurang Manchanda, student of MBA 2014-16 of Delhi School of Management, Delhi Technological University, Bawana Road, Delhi-42 declare that project dissertation report on “Uncovering trends in Taxi Aggregator Industry Using Twitter Sentiment Analysis” submitted in partial fulfillment of Degree of Masters of Business Administration is the original work conducted by me.

The information and data given in the report is authentic to the best of my knowledge.

This report is not being submitted to any other University for award of any other Degree, Diploma and Fellowship.

Place: New Delhi

Gaurang Manchanda

Date:

(2K14/MBA/22)

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to Dr. Rajan Yadav, associate professor, Delhi School of Management, Delhi Technological University, for encouraging me to work on this project and helping me understand it. I am extremely grateful to him for his kind support and guidance throughout the project.

The completion of this project would not have been possible without the valuable contribution of Mr. Anuraag Tiruwa, PHD scholar, DSM, DTU. I am sincerely thankful for his guidance, invaluable constructive criticism and friendly advice during the project work.

I would also like to thank my college, Delhi School of Management, Delhi Technological University for providing me with the facilities and making me capable enough to complete this project successfully for future development.

EXECUTIVE SUMMARY

Micro-blogging websites have evolved to become a source of all kinds of information. People post all kinds of real time messages on micro-blogs including their experience of a service they use, opinions on a variety of topics and current issues, complains and positive sentiments about the product they use.

Twitter offers a unique dataset in the world of brand sentiment. Brands receive sentiment messages directly from their customers in real time on twitter. These brands have the opportunity to analyze these messages to determine the consumer sentiment. Taxi aggregator industry being a high volume service industry receives hundreds of comments on their social media pages daily from their customers regarding their experiences, complaints and opinions on the services provided.

The aim of this study is to analyze the sentiments of a corpus of tweets posted with hashtags and twitter handles of major taxi aggregator players in India, including Ola cabs and Uber Cabs. The study aims to classify the tweet sentiments as positive, negative and neutral. The purpose of this study is to identify key service areas of these companies which require further improvements and the areas which provide positive experience to the customers. The study further discovers trends from the data which may generate actionable insights.

TABLE OF CONTENTS

1.	Introduction	1
2.	Social Media Trends In India	3
	2.1 Measuring Success of Social Media Platform	4
	2.2 Measuring Brand Sentiment	5
	2.3 Using Twitter in Sentiment Analysis	8
3.	Industry Profile	11
	3.1 Taxi Aggregators In India	11
	3.2 Evolution of the Organized Taxi Sector in India	12
	3.3 Taxi Aggregators: Facts and Figures	14
4.	Methodology	19
	4.1 Significance of the project	19
	4.2 Scope of the study	19
	4.3 Research Methodology	19
	4.4 Sample Size	19
	4.5 Data Extraction	19
	4.6 Data Cleaning and Pre-Processing	21
	4.7 Tools Used	22
	4.8 Steps involved in sentiment analysis and category creation	25
5.	Analysis, Inferences and Recommendations	28
	5.1 Tweet Analysis Of Uber India	28

5.1.1 Inferences and Recommendations for Uber India	33
5.2 Tweet Analysis Of Ola Cabs	34
5.2.1 Inferences and Recommendations for Ola Cabs	39
6. Limitations of the study	41
7. Conclusion	42
References	43
Annexures	44
Adherence Sheet	50

LIST OF FIGURES

Figure	Title	Page No
Figure 2.1	Primary objectives for being present on Social Media	4
Figure 2.2	Companies measuring Brand Sentiment using Social Listening Tools	5
Figure 2.3	Impact on Brand Sentiment after using Social Listening Tools	6
Figure 2.4	Frequency of Updates on Social Media	7
Figure 2.5	Average Response Time on Social Media	7
Figure 2.6	Tweets addressed to Uber and Ola Handles	9
Figure 3.1	Comparison of different taxi models	12
Figure 3.2	Facts about major taxi aggregators	14
Figure 3.3	Twitter Statistics of Ola Cabs	15
Figure 3.4	Twitter followers of Ola Cabs trend for April 2016	15
Figure 3.5	Twitter Statistics of Uber Cabs	16
Figure 3.6	Twitter followers of Uber Cabs trend for April 2016	17
Figure 3.7	Twitter Statistics of Meru Cabs	17
Figure 3.8	Twitter followers of Meru Cabs trend for April 2016	18
Figure 4.1	Twitter object ids extracted for Ola cabs	20
Figure 4.2	JSON response (right pane) for statuses/lookup query on Twitter API console	21
Figure 4.3	SPSS Text Analytics for Surveys Libraries	23
Figure 4.4	SPSS Negative Words Library	24
Figure 4.5	SPSS Positive Words Library	24
Figure 4.6	Flowchart for Sentiment Analysis and Category Creation	27
Figure 5.1	Sentiment Analysis and Category creation for Uber India data in SPSS TAS	28
Figure 5.2	Category Web for Uber India data	29

Figure 5.3	Distribution of positive and negative tweets across various categories for Uber Cabs	30
Figure 5.4	Word cloud for high frequency themes extracted from Uber India data	30
Figure 5.5	Word Cloud of High Frequency Positive and Negative Words for Uber Cabs	31
Figure 5.6	Trend analysis of sentiment over the month for Uber cabs	31
Figure 5.7	Trend analysis of tweets over different time slots for Uber Cabs	32
Figure 5.8	Trend analysis of tweets over different days for Uber Cabs	32
Figure 5.9	Follower count of users who tweeted vs. sentiment of posts for Uber Cabs	33
Figure 5.10	Sentiment Analysis and Category creation for Ola Cabs data in SPSS TAS.	34
Figure 5.11	Detailed Category Web for Ola Cabs data	35
Figure 5.12	Simplified Category Web for Ola Cabs data	35
Figure 5.13	Distribution of positive and negative tweets across various categories for Ola Cabs	36
Figure 5.14	Word cloud for high frequency themes extracted from Ola cabs data	37
Figure 5.15	Word Cloud of High Frequency Positive and Negative Words for Ola Cabs	37
Figure 5.16	Trend analysis of sentiment over the month for Ola Cabs	38
Figure 5.17	Trend analysis of tweets over different days for Ola Cabs	38
Figure 5.18	Trend analysis of tweets over different time slots for Ola cabs	39
Figure 5.19	Follower count of users vs. sentiment of posts for Ola cabs	39

LIST OF TABLES

Tables	Title	Page no
Table 4.1	Category wise negative and positive responses for Uber India	29
Table 4.2	Category wise positive and negative responses for Ola Cabs	36

INTRODUCTION

Social media has off-late changed the way companies interact with their customers. It has enabled them to conduct business in an entirely different manner wherein they get to interact with their customer base via social networking sites. As millions of people share and post on social media, the rate at which this data is increasing is unfathomable. (Reza Zafarani, Huan Liu et al, Cambridge Univeristy Press, 2014). Industry- consumer interaction happens in real time with the advent of these forums. Blogs, customer review websites, social media networking sites like Facebook and Twitter have emerged out to be the greatest sources of data for the companies to analyze and develop marketing intelligence solutions.

Social media data mining is the art and science of extracting, analyzing and presenting patterns from social media data in order to fetch meaningful insights. A study indicates that out of 98 percent of customers who raise issues about customer service about 30 percent do so via online social networking sites (Chandra, International Journal of Business Research, 2011). To the marketer, this mined data can provide insights into behavior patterns of the customer and help him understand customer feelings and sentiments in real time. (Shintaro Okazaki et al, 2014). However, along with the benefits come the challenges in exploiting this huge dataset. User generated content on social media is often fragmented and disorganized. Special techniques are needed to analyze and mine patterns out of it. One such technique is sentiment analysis.

Sentiment analysis is a data mining technique that uses machine learning algorithms to analyze and infer the sentiment of a piece of text. In the field of marketing, it acts as a research method to effectively evaluate the consumer opinions in real time. (SearchBusinessAnalytics, 2016). It allows data extraction and analysis from a very large corpus without any time delays. Marketers, with the help of sentiment analysis are able to gain information on attitudes and opinions of consumers as they occur, without having to invest in lengthy and costly market research activities.

In this study, we apply sentiment analysis techniques to gauge the sentiments of the customers regarding the services provided by taxi aggregator companies. The objective of this study is to analyze the sentiments of a corpus of tweets posted with hashtags and twitter handles of major taxi aggregator players in India, including Ola cabs and Uber Cabs. The study aims to classify the tweet sentiments as positive, negative and neutral. Further, the purpose of this study is to identify key service areas which require further improvements and the areas which provide positive experience to the customers.

SOCIAL MEDIA TRENDS IN INDIA

Social media usage in India increased manifolds, as the number of internet users in India reached 375 million users. (Statista.com, 2016). India is one the key markets for social media giants as India has mobile social media penetration of about 9% (Statista.com, 2016). The active social media users in India has grown to around 106 million and India is among the top three countries in number of users for Facebook and twitter, the numbers being 100m+ for Facebook and over 33 million for Twitter. (Ernst Young, 2015). About 84% of Facebook's 100 million users in India access it using their mobile devices (Ernst Young, 2015).

Almost 81% of the brands surveyed consider Facebook to be the most important platform for customer engagement, while about 48% of surveyed brands consider Twitter as the second-most important channel after Facebook, followed by YouTube (with 43% surveyed brands considering it to be the third-most important channel). (Ernst Young, 2015)

75% of India's online population is digital consumers i.e. those who use digital media for purchasing goods and services (Ernst Young, 2015). Trust in a brand and company is no longer dependent only on company-controlled, traditional mass media channels, but rather on peers and communities through social media and other digital channels. Therefore, brands need to allocate a significant proportion of marketing budget to social media marketing and digital in order to successfully market to these customers.

The top 3 objectives to be present on social media for the brands include Building Brand Awareness, Customer Engagement and Building a Community.

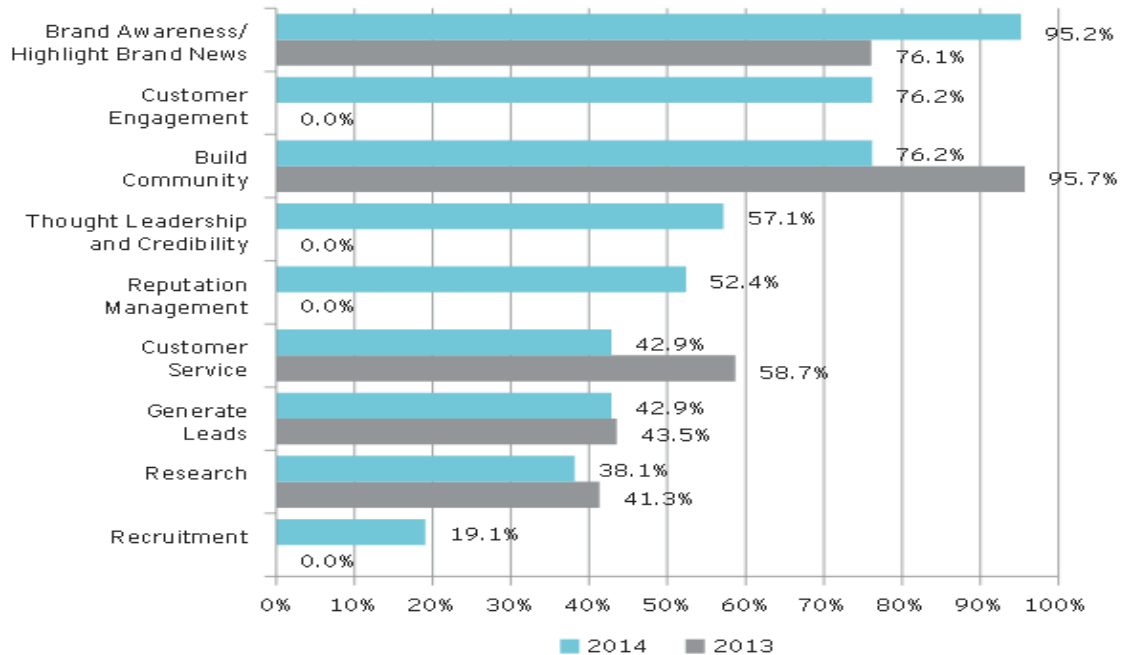


Fig 2.1: Primary objectives for being present on Social Media
 (Source: Social Media Marketing-India Trends Study, EY, 2015)

2.1 Measuring Success of Social Media Platform

For measuring the success at social media platform, marketers use various metrics like

Brand sentiment analysis

It refers to analyzing the emotion behind a social media post. Sentiment analysis is a technique which can help measure the tone of a conversation and adds important context to social conversations.

Customer engagement

It refers to how often and how much people interact with a brand and its content in social media. When a visitor or customer likes, re-tweets, shares or comments on something a brand has posted, they're actively engaging with the brand.

Brand mentions

This is the total number of times a brand is mentioned on a webpage or social media network over a particular period of time.

Social reach

It refers to the total number of people a brand is able to reach via its various social media networks. It is used to measure the influence of a brand and usually takes into account the number of shares, retweets, repins, click-through rates, the number of visitors etc.

2.2 Measuring Brand Sentiment

Brand sentiment can be used to evaluate the performance of campaigns and other initiatives for competitive research and also evaluate brand health.

Over the time, as a company's online footprint increases manifolds, it becomes almost impossible to track the sentiment manually. There are a host of tools that can help in tracking and evaluating the brand sentiment.

A majority of organizations monitor their brand sentiment through automated social media tools. Tools used for social listening purpose include Meltwater , SAS text analytics, Radian6, Simplify 360, Iristrack, Social Mention, Hootsuite and Netbase.

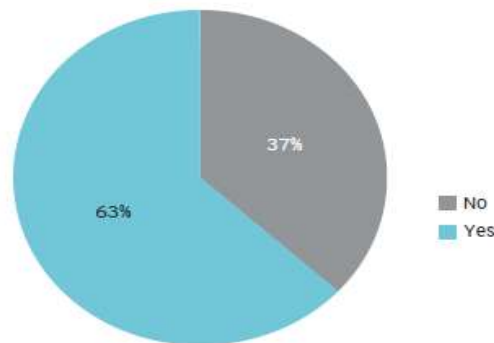


Fig 2.2: Companies measuring Brand Sentiment using Social Listening Tools
(Source: Social Media Marketing-India Trends Study, EY, 2015)

Marketers believe that sentiment analysis has reaped benefits for their companies. Some of the major benefits that organizations have realized by via social listening tools include effective management of customer queries, better understanding of the perception of the brand and effective resolution on requests.

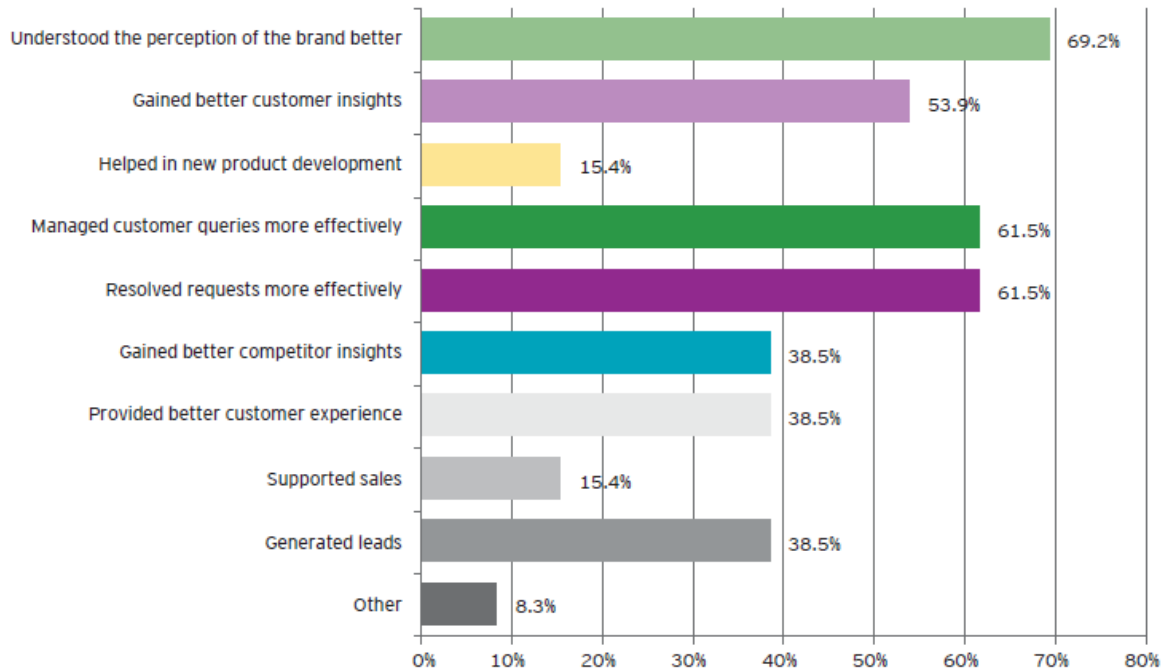


Fig 2.3: Impact on Brand Sentiment after using Social Listening Tools
 (Source: Social Media Marketing-India Trends Study, EY, 2015)

The time taken to respond via social media channels is another critical factor that determines the performance of a company’s social media presence success. Companies now a days do much more than marketing on social media, they also indulge in customer service and crisis management. From a marketing point of view, well stated responses prove to of more value than a quicker response.

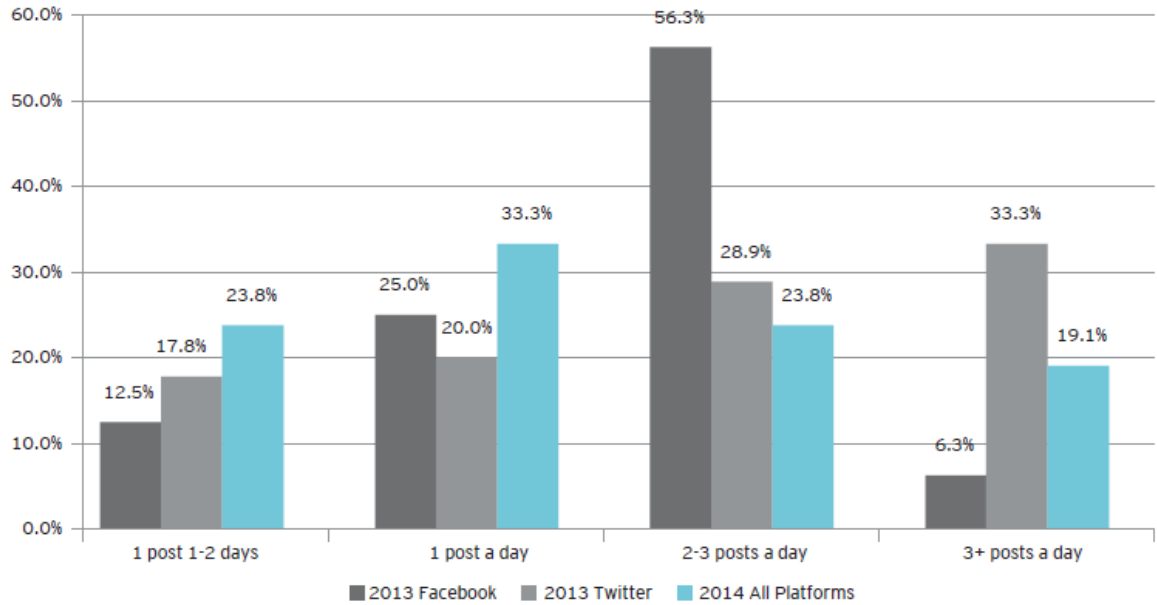


Fig 2.4: Frequency of Updates on Social Media

(Source: Social Media Marketing-India Trends Study, EY, 2015)

In 2014, the average response time for a company to respond was 30 min or less for 38% of brands (compared to average response times stated by 25% brands for Facebook and 28% of brands in 2013). (Ernst Young, 2015).

Average Response Time

What is your average response time in which you attempt to reply to users on a social networking platform?

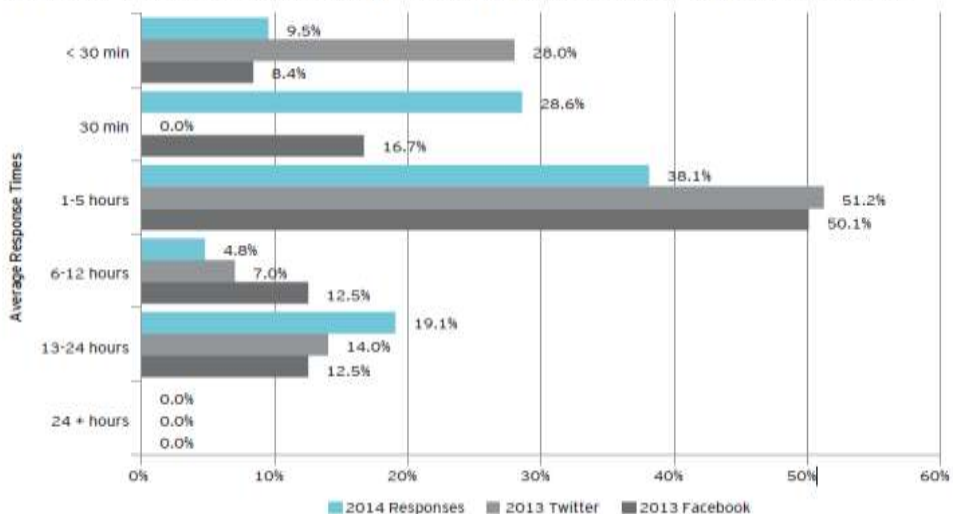


Fig 2.5: Average Response Time on Social Media

(Source: Social Media Marketing-India Trends Study, EY, 2015)

2.3 Using Twitter in Sentiment Analysis

Micro-blogging site Twitter is among the world's top ten social media channels in terms of daily number of visitors and registered users (Garg et al., Journal of Management Information Systems, 2011). It is one of the most popular social media channels for receiving real-time updates with around 160 million registered users, receiving around 55 million tweets per day and 600 million search queries. (Savage, 2011; Thelwall et al., Journal of American Society for Information Science and Technology, 2011). Twitter has proved to be one of the favorite social networking sites owing to its concise format which allows only 140 characters. (Geho et al.,2010; Gayo-Avello, 2011; O'Leary, Communications of the ACM, 2011). Companies are constantly pushing content to reach their customer base via twitter. They are interacting with them in real time to solve any issue the customers face while using their products and services. Sentiment analysis can be used to judge the sentiment of the tweets and discover patterns in them in order to understand the perception of the customer and analyze the effect of distinct events. Further, customer behavior can be predicted using this technique. (Fotis Misopoulos et al, 2014).

When a user responds to an event by tweeting, she/he demonstrates an "information behavior". This contains his judgments on the event and therefore reveals important information about users' sentiments and on the tweeted topic. Savage (2011) observed that while individual tweets might not be of great value, but a corpus of tweets when taken together can uncover information about the customer's opinions and moods which can help marketers in understanding their customers better.

The use of twitter as an information source is increasing over the time with the advent of new tools and techniques. Researchers are experimenting with various textual information analysis techniques for analyzing Twitter feeds to get actionable insights. (Thelwall et al., Journal of American Society for Information Science and Technology 2011). Based on the aforementioned, the text that follows presents a study conducted within the framework of sentiment analysis and provides an example of how twitter information can be used to assess customer experiences in the taxi aggregator industry.

Consumers of taxi aggregator services actively post their experiences in the form of opinions, complaints, suggestions etc. on a daily basis.

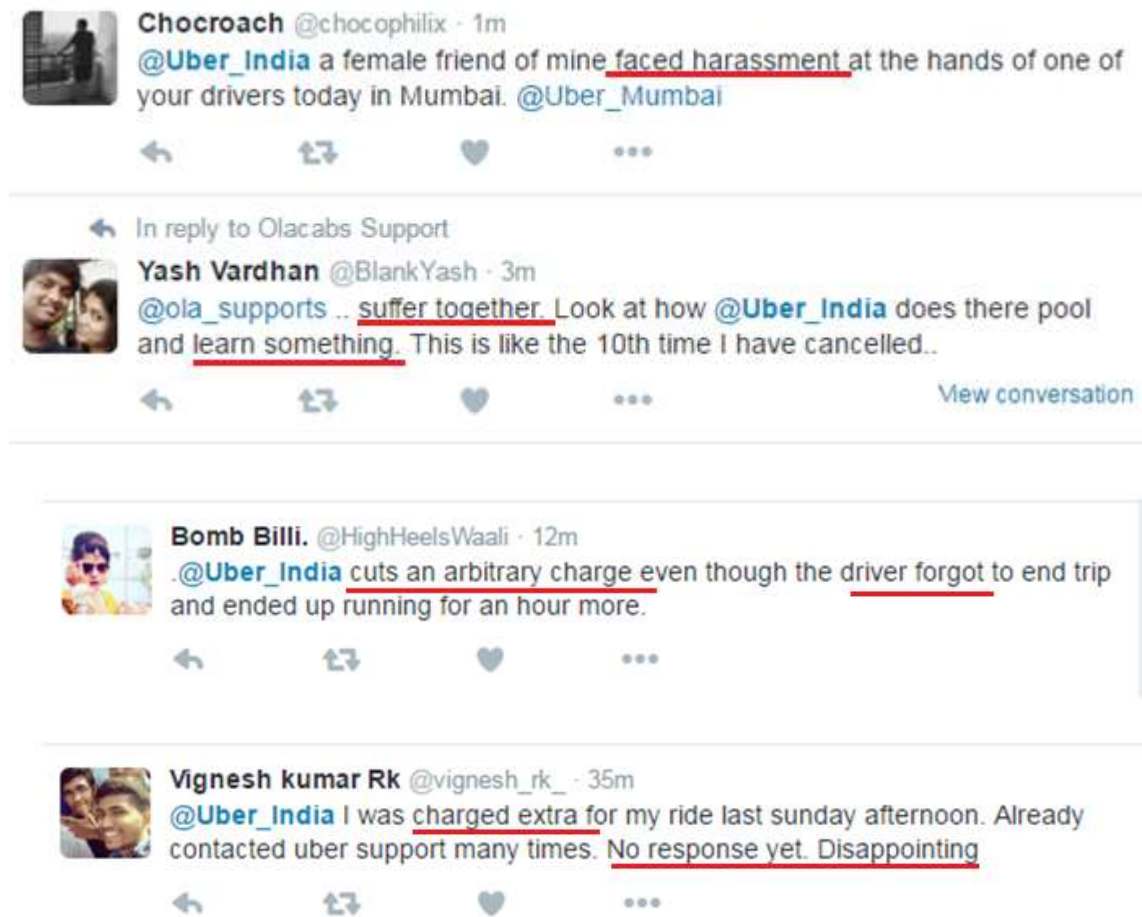


Fig 2.6: Tweets addressed to Uber and Ola Handles

(Source: Twitter.com, 2016)

These aggregators can mine the customer comments and analyze them to understand the general customer sentiment. They can further analyze areas which particularly require immediate attention for improvement and areas which delight the customer.

For example: Multiple posts on a particular day about the booking being cancelled by drivers can make the company look into its operational issues and rectify them to deliver better services and ensure customer satisfaction.

Customer sentiment can be measured over a long duration to analyze the impact of a newly adopted strategy. These companies can mine for trends in their feed to understand the customer better and thereby deliver a better service.

INDUSTRY PROFILE

3.1 Taxi Aggregators in India

The taxi business in the country is growing at 20 - 25 per cent per year as per the reports of Radio Taxi Association of India. Currently the organized taxi sector is just around four to five percent of the industry and has a valuation of around \$800 million. It is expected to grow to \$7 billion by 2020 (businesstoday.in, 2015). The taxi aggregator services have become popular in the last few years in India. The companies which manage these services call themselves as technology companies. This exempts them from the transport regulations that a typical transport operator has to adhere to.

Ola is the overall market leader in terms of market share. According to SoftBank Corp, based on the data of registered vehicles, the company has a 65 per cent share. TaxiForSure, a company Ola acquired, has 14 per cent share.(businesstoday.in, 2015) Meru cabs had the second-largest share at 16 per cent while Uber cabs just has five percent of the market. (businesstoday.in, 2015) The combined market share of Ola and TaxiForSure is around 80 per cent and all the other taxi companies together have the rest. While Ola got an earlier start on the market by launching its service before Uber arrived in August 2013, Ola's lead may also have resulted due to the breadth of services that it offers. Last week, Ola revealed that its 'Micro' vehicle service alone is bigger than Uber, covering 75 cities and over a million daily rides. The acquisition of TaxiForSure for \$200 million last year will also have helped.

One of the major reasons for the success of Ola and Uber, the two major taxi aggregator services in India, is their ability to digitally match supply and demand by successful deployment of technology. Technology has enabled the right information to be available to the right persons at the right time. The business model of taxi aggregators has provided intelligent solutions to address the gaps in the present call taxi system – namely lack of focus on performance, driver behavior, difficulty in matching capacity with demand, uncertainty of demand, increase in prices etc. Using technology

comprising of software algorithms enabling accurate matching of demand and supply, the idle time is reduced for the drivers and for the customers, the wait time is reduced. The other benefit for the consumer is that travel using taxi aggregator services is at an affordable cost due to the economies of scale. This has created a win-win-win situation for all – the driver gets assurance of demand, the taxi aggregator gets his commission and the consumer has to wait less and pay reasonable fares for availing the taxi services.

	Unorganized Market	Radio Cabs	Affiliator	Aggregator
Type of Operator	<ul style="list-style-type: none"> Individual operators or cabs running in different part of country 88-92% of the market is still unorganized. 	<ul style="list-style-type: none"> The companies registered under Radio Cab Scheme. It just constitutes 4-5% of the market. 	<ul style="list-style-type: none"> The company are affiliating with Taxi owners. Follows Garage-to-Garage Model 	<ul style="list-style-type: none"> The new model which is aggregating taxi under single brand. It constitutes less than 0.5% .
Key Players	<ul style="list-style-type: none"> Individual Owners Small Operators Taxi Stands Pre-paid Taxi 	<ul style="list-style-type: none"> Meru Cabs Easy Cabs Mega Cabs TAB Cab 	<ul style="list-style-type: none"> TaxiGuide Savaari Cab24X7 TaxiPixi 	<ul style="list-style-type: none"> Uber Ola Cabs Taxi for Sure Bookmycab

Fig 3.1: Comparison of different taxi models

(Source: Evolution of Indian Taxi Market, INC42.com, 2015)

These taxis capitalize on their availability quotient and comfort quotient to attract people. Use of GPRS (General Packet Radio Service based technology) and Global Positioning System (GPS) can help operators monitor the cabs on real time basis thereby making efficient use of cabs in terms of a high utilization ratio. This results in a good financial performance.

3.2 Evolution of the Organized Taxi Sector in India

The organized taxi sector has its inception in 2001 with Fast Track Taxi and Mega Cabs. But it was only after 2006 that the market saw traction when players like Easy Cabs, Meru Cabs and Savaari came up. There are now several competitors in the organized taxi business.

Phase 1 – Fully Owned Fleets

Initially the companies owned the complete fleet of cars. The drivers were salaried employees. This led to high capital costs to the company including car loan EMI's, high maintenance costs etc. Although this model helped in rapid expansion, however it also came at a huge cost. Bookings were mainly done via telephone calls. High stress levels of drivers and driver strikes were common.

Phase 2 – Fleet Aggregation Model

This model was started by companies such as TaxiForSure and Ola. In this, car owners or small fleet owners can get registered with the company to deploy their cars as taxis. Cars could take up non-company rides as well, however for every company-initiated ride; they had to pay certain percentage as commission to the company. This model had lower maintenance costs and low capital expenditure.

Booking was done via websites as well as through telephone calls. While cash was still the dominant mode of payment, in-cab POS terminal for credit / debit cards was also used in this phase.

Phase 3 – The Hybrid Model (Current Phase)

In this model, part of the fleet is from an aggregation model and part of the fleet is owned by the company and hence providing the best of both worlds i.e. while keeping costs low, they offered better control on service quality and cab availability.

Booking are done via website, telephone as well as mobile apps and the payments via cash, card and wallets. Major players include Ola Cabs, Uber and Meru Cabs.

3.3 Taxi Aggregators: Facts and Figures

Ola was founded in Mumbai by IIT Bombay graduates Bhavish Aggarwal and Ankit Bhati in 2010. They shifted headquarters to Bangalore in 2012. It currently serves 100 cities in India and it follows a 100% aggregated model.



Fig 3.2: Facts about major taxi aggregators
 (Source: Evolution of Indian Taxi Market, INC42.com, 2015)

According to a report by www.techinasia.com :

- Average number of Ola rides per day: 1 million
- Number of cabs on Ola's platform: 350,000
- Reported Number of Daily trips: 200,000.
- Number of cities Ola cabs operates in : 75

The twitter handle of Olacabs has around 59k followers and 13k tweets.



Fig 3.3: Twitter Statistics of Ola Cabs
(Source: Socialbakers.com, 2016)

A steady increase in the number of followers from 56k to 59k during the month of April was observed.



Fig 3.4: Twitter followers of Ola Cabs trend for April 2016
(Source: Socialbakers.com, 2016)

Uber, on the other hand, founded in 2009 entered India a little later than Ola cabs. Similar to Ola, it also follows a 100% aggregated model. Recently, Uber's Alexander said in his interview that Uber's market share increased from a mere 5% in January last year to nearly 50%. Ola Micro (priced at Rs 6/km) is the closest competitor to UberGo (priced at Rs 7/km) in terms of pricing.

According to a report on www.techinasia.com:

- Average number of Uber rides per day: 2,00,000 (in last august)
- Number of cities Uber operates in India: 27

The twitter handle of Uber India has around 13.5k followers and 7.2k tweets.



Fig 3.5: Twitter Statistics of Uber Cabs
(Source: Socialbakers.com, 2016)

Uber India also saw a steady rise in the number of followers in the month of April.

Uber India Twitter Followers



Fig 3.6: Twitter followers of Uber Cabs trend for April 2016

(Source: Socialbakers.com, 2016)

Founded in 2007 and active in 23 cities, Meru is the only player amongst the three to have a hybrid model. According to CEO Siddhartha Pahwa, 70 per cent of its cars are today aggregated. The rest are owned by Meru and are given under a subscription model to drivers.

The twitter following of Meru is 10.8K users and has around 7.5k tweets.



Fig 3.7: Twitter Statistics of Meru Cabs

(Source: Socialbakers.com, 2016)

Meru saw a gradual rise in the number of followers in the month of April with 400 new followers added up during the month.



Fig 3.8: Twitter followers of Meru Cabs trend for April 2016
(Source: Socialbakers.com, 2016)

METHODOLOGY

4.1 Significance of the study

The study is aimed at analyzing the customer sentiment of taxi aggregator companies (Ola and Uber cabs) using their twitter handles and associated hashtags. The goal of the study was to harness data on social media platform Twitter for monitoring and analyzing customer experiences for the purpose of optimizing service performance. The study further aimed at uncovering trends in the tweets of the customers. The study was conducted under the framework of sentiment analysis

4.2 Scope of the study

This study is focused on the taxi aggregators including Ola cabs and Uber cabs, their customers and their management executives.

4.3 Research Methodology

Descriptive research techniques were employed for the study. Tweets addressed to @ola_support and @uber_india were extracted and preprocessed to remove hashtags, URLs, @ symbols etc. These tweets were then analyzed using SPSS Text Analytics for Surveys and results were plotted in the same.

4.4 Sample Size

A corpus of 2096 tweets of Uber India and 1430 tweets of Ola cabs posted between the dates 1-April-2016 to 30-April-2016 was analyzed to determine their sentiment polarity.

4.5 Data Extraction:

A number of free GUI tools like Facepager etc. are available for data extraction which use the Twitter API in the background. However, a limitation of using such tools is that for a particular search term, tweets of past 7-8 days or a maximum of 1500 tweets can be extracted. This study involves analysis of a data collected over a period of month, so

a workaround had to be employed in order to extract more tweets. A semi-automatic approach was employed to extract the tweets (Tillkeyling.com, 2016).

For extraction, the following steps were taken:

1. Using twitter advanced search, tweets addressed to various taxi aggregators were searched for using their handle (For Example: @Uber_India) between the dates 1st April 2016 to 30th April 2016. The search results were scrolled to the bottom. A script was used to extract the objects ids out of the results page.

```
723661200920350720, 723660298041843713, 723660160946847744, 723653890223144961, 7
23653324352819200, 723653056441651200, 723652886912065537, 723649257790533632, 72
3648682281832449, 723647841936105474, 723641591035006977, 723641498781229056, 723
639501432623104, 723631761112989696, 723625873300017153, 723625691707625472, 7236
25046397227008, 723624603935866880, 723620405102043136, 723608117586452482, 72360
7823892897792, 723607648604700673, 723606600666107904, 723605329863385089, 723605
147511742464, 723604872206135296, 723604482836172800, 723603952449773568, 7236034
59270959104, 723603012392886275, 723602952322187264, 723602182763876353, 72360182
3630811136, 723600977539043328, 723599510694334464, 723599208218017793, 723598773
394485248, 723598016410689536, 723597417015316480, 723597196885561344, 7235971657
38643457, 723596583032524800, 723596425926303744, 723595772164366336, 72359561161
9004416, 723595249273065473, 723594699966029824, 723594413939650560, 723593261139
415040, 723593138359525376, 723593100770172928, 723592759605489664, 7235910862543
50336, 723590690228211712, 723590551518343168, 723589492209233920, 72358839906457
1908, 723587821529927682, 723587687639195648, 723587686116655104, 723587078945169
409, 723587032405147653, 723585803935473666, 723585604232069125, 7235848032835788
81, 723583997184380928, 723583726295228416, 723583654304190464, 72358360086049587
2, 723583347654422529, 723582676284878848, 723582402908422144, 723581931007307776
, 723580892183183360, 723580147970920448, 723579471723433984, 723579426512863232,
723579395445682176, 723579142818652160, 723579063957250048, 723579058957770753, 7
23578685522104320, 723578547558866944, 723578305920655360, 723578217089654785, 72
3578147581513732, 723577930928918529, 723577851577044992, 723577733112979456, 723
577439876579328, 723577140063543296, 723576911457357824, 723576569009041408, 7235
76235415080961, 723576203274149889, 723576142402322432, 723575791204851713, 72357
5747219083264, 723575725698256896, 723575633134030849
```

Fig 4.1: Twitter object ids extracted for Ola cabs

2. The twitter API 1.1 has a statuses/lookup function which allows fetching data by giving ids as input. However, it allows fetching data of only 100 ids in one go. The ids extracted in the previous step were automatically chunked in groups of 100 comma separated records by the script. Chunks of 100 ids were fed to the Twitter API console to fetch 100 records. The Twitter API returns a JSON script as response.

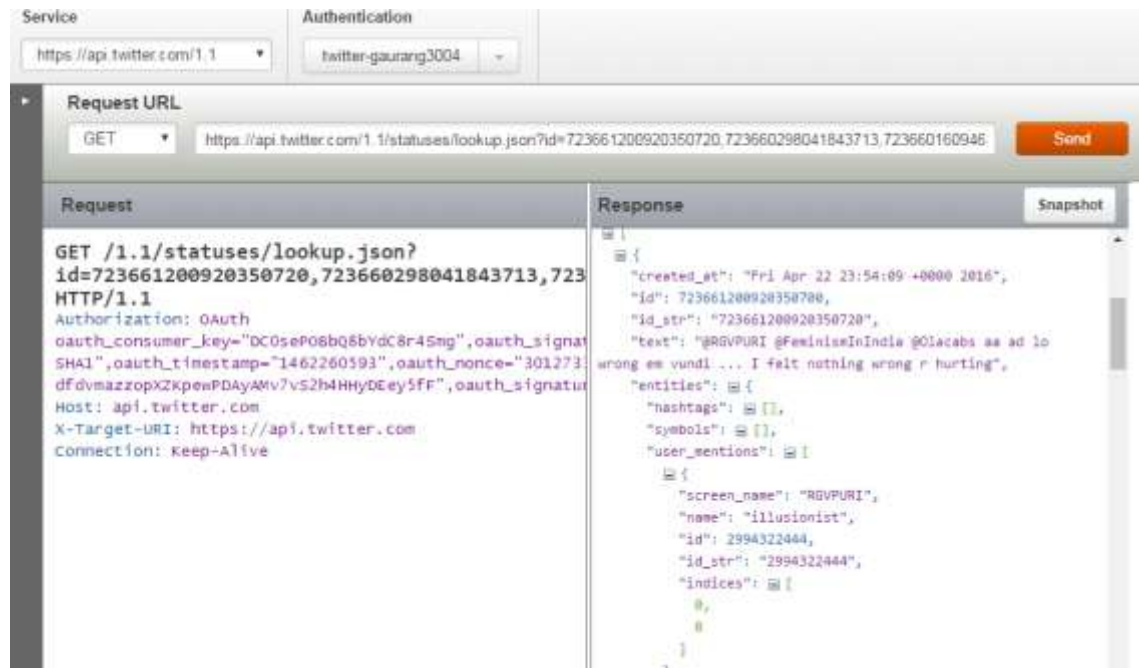


Fig 4.2: JSON response (right pane) for statuses/lookup query on Twitter API console.
(Source: Twitter API Console Tool, 2016)

3. The JSON response was validated and converted to Excel format.

4.6 Data Cleaning and Pre-Processing

Before analyzing the records, they need to be pre-processed to remove different symbols and URLs which are irrelevant for sentiment analysis. Following steps were taken in data pre-processing.

1. Convert the tweets to lower case.
2. URLs –The URLs in the tweets don't fetch any information about the sentiment of the tweet, so eliminate all of these URLs and replace with generic word URL
3. @username - Replace "@username" with generic word AT_username via regex matching.
4. #hashtag - hash tags can give us some useful information, so it is useful to replace them with the exact same word without the hash. (RavikantRajBlog, 2015) E.g. #olacabs was replaced with 'olacabs'

4.7 TOOLS USED

1. IBM SPSS Text Analytics for Surveys 4.0.1

SPSS Text Analytics for Surveys uses linguistic techniques to extract and classify key concepts from open-ended survey responses. It has reliable category building algorithms which enables the users to categorize the responses of the open ended text. The categories produced can also be reused to provide consistent results across the same or similar studies.

It extracts key concepts and terms using linguistics-based text analysis which offers speed and accuracy. Linguistics-based text analysis is based on the field of study known as natural language processing,

For a thorough analysis, SPSS Text Analytics for Surveys provides a list of lexicon libraries including:

1. Customer Satisfaction Library
2. Product Satisfaction Library
3. Opinions Library
4. Slang Library
5. Information Library
6. Emoticon Library

These dictionaries provide an effective way of classifying text as expressing positive or negative sentiment (Apoorv Agarwal, 2011). In the figure 4.3, the left pane shows the different libraries present in the text analysis software and the types, the right pane shows exclude list (stop words), bottom pane shows synonyms and optional elements.

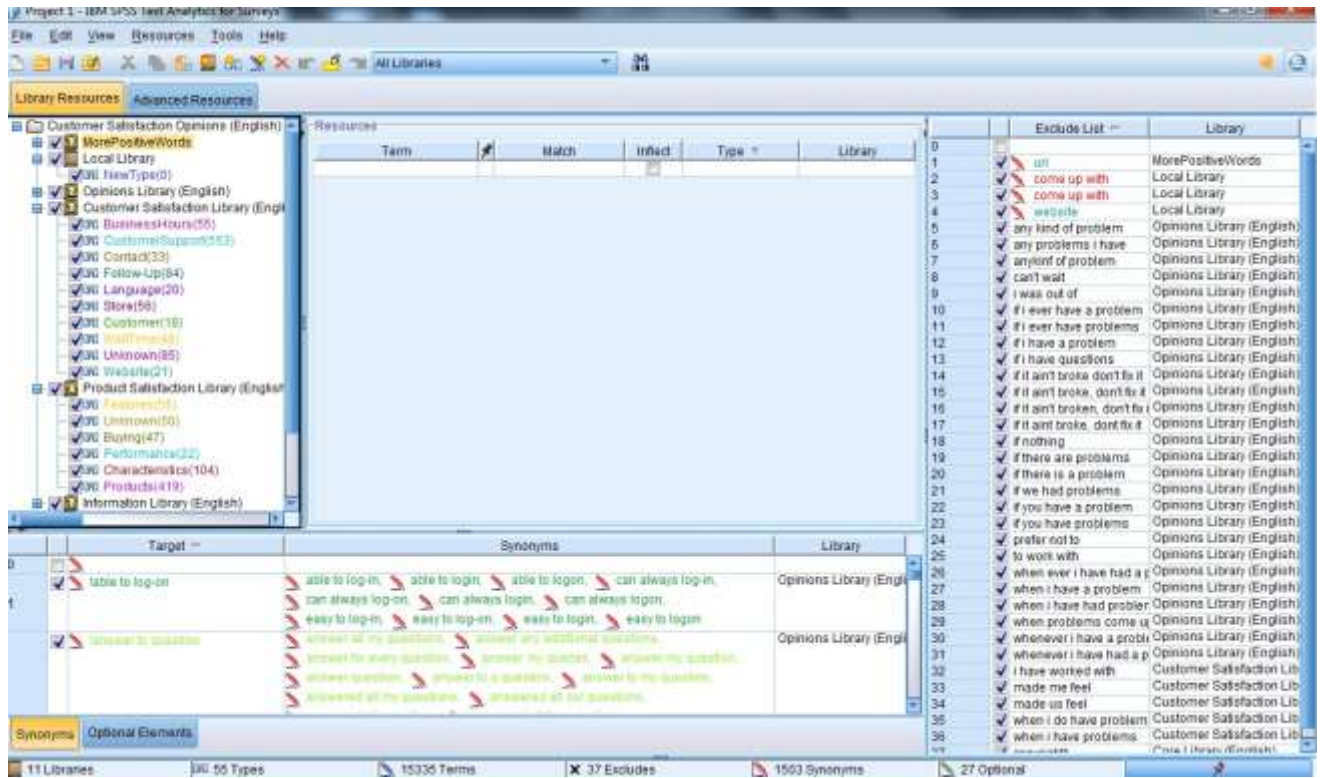


Fig 4.3: SPSS Text Analytics for Surveys Libraries

Each library has further certain types defined in it. For example: The Opinions library has Positive, Positive Attitude, Positive Budget, Positive Competence, Positive Feeling, Positive Functioning, Negative, Uncertain, Negative Attitude, Negative Budget, Negative Competence, Negative Feeling, Negative Functioning types defined under it.

Each type has a number of words defined under it. In figure 4.4, the center pane shows words belonging to 'Negative' Type in Opinions Library. Further, in figure 4.5, the center pane shows words belonging to 'Positive Attitude' Type in Opinions Library.

The screenshot shows the SPSS interface for the Negative Words Library. The left pane displays a hierarchical tree view of resources, including 'Opinions Library (English)' and 'Customer Satisfaction Library (English)'. The right pane shows a table of terms with columns for Term, Match, Inflect, Type, and Library.

Term	Match	Inflect	Type	Library
a bit less than expected	Entire (no compounds)	<input type="checkbox"/>	Negative	Opinions Library (Engl)
a little doubt	Entire (no compounds)	<input type="checkbox"/>	Negative	Opinions Library (Engl)
a little tired of	Entire (no compounds)	<input type="checkbox"/>	Negative	Opinions Library (Engl)
abashed	Entire (no compounds)	<input type="checkbox"/>	Negative	Opinions Library (Engl)
abhor	Entire (no compounds)	<input type="checkbox"/>	Negative	Opinions Library (Engl)
abnormal	Entire (no compounds)	<input type="checkbox"/>	Negative	Opinions Library (Engl)
abominable	Entire (no compounds)	<input type="checkbox"/>	Negative	Opinions Library (Engl)
abrasive	Entire (no compounds)	<input type="checkbox"/>	Negative	Opinions Library (Engl)
abrupt	Entire (no compounds)	<input type="checkbox"/>	Negative	Opinions Library (Engl)
absolute nightmare	Entire (no compounds)	<input type="checkbox"/>	Negative	Opinions Library (Engl)
absurd	Entire (no compounds)	<input type="checkbox"/>	Negative	Opinions Library (Engl)
abused	Entire (no compounds)	<input type="checkbox"/>	Negative	Opinions Library (Engl)
abysmal	Entire (no compounds)	<input type="checkbox"/>	Negative	Opinions Library (Engl)
achilles heel	Entire (no compounds)	<input type="checkbox"/>	Negative	Opinions Library (Engl)
add more	Entire (no compounds)	<input type="checkbox"/>	Negative	Opinions Library (Engl)
adverse	Entire (no compounds)	<input type="checkbox"/>	Negative	Opinions Library (Engl)
adverse reaction	Entire (no compounds)	<input type="checkbox"/>	Negative	Opinions Library (Engl)
afraid	Entire (no compounds)	<input type="checkbox"/>	Negative	Opinions Library (Engl)
aggravating	Entire (no compounds)	<input type="checkbox"/>	Negative	Opinions Library (Engl)
agrivating	Entire (no compounds)	<input type="checkbox"/>	Negative	Opinions Library (Engl)
aimless	Entire (no compounds)	<input type="checkbox"/>	Negative	Opinions Library (Engl)
alarming	Entire (no compounds)	<input type="checkbox"/>	Negative	Opinions Library (Engl)
allergic	Entire (no compounds)	<input type="checkbox"/>	Negative	Opinions Library (Engl)
along time	Entire (no compounds)	<input type="checkbox"/>	Negative	Opinions Library (Engl)
alongtime	Entire (no compounds)	<input type="checkbox"/>	Negative	Opinions Library (Engl)
always late	Entire (no compounds)	<input type="checkbox"/>	Negative	Opinions Library (Engl)
always the same	Entire (no compounds)	<input type="checkbox"/>	Negative	Opinions Library (Engl)
ambig	Entire (no compounds)	<input type="checkbox"/>	Negative	Opinions Library (Engl)

Fig 4.4: SPSS Negative Words Library

The screenshot shows the SPSS interface for the Positive Words Library. The left pane displays a hierarchical tree view of resources, including 'Opinions Library (English)' and 'Customer Satisfaction Library (English)'. The right pane shows a table of terms with columns for Term, Match, Inflect, Type, and Library.

Term	Match	Inflect	Type	Library
100 % commitment	Entire (no compounds)	<input type="checkbox"/>	PositiveAttitude	Opinions Library (Engl)
100% commitment	Entire (no compounds)	<input type="checkbox"/>	PositiveAttitude	Opinions Library (Engl)
accomadating	Entire (no compounds)	<input type="checkbox"/>	PositiveAttitude	Opinions Library (Engl)
accomidating	Entire (no compounds)	<input type="checkbox"/>	PositiveAttitude	Opinions Library (Engl)
accomindating	Entire (no compounds)	<input type="checkbox"/>	PositiveAttitude	Opinions Library (Engl)
accommodating	Entire (no compounds)	<input type="checkbox"/>	PositiveAttitude	Opinions Library (Engl)
accomodating	Entire (no compounds)	<input type="checkbox"/>	PositiveAttitude	Opinions Library (Engl)
adorable	Entire (no compounds)	<input type="checkbox"/>	PositiveAttitude	Opinions Library (Engl)
aim to please	Entire (no compounds)	<input type="checkbox"/>	PositiveAttitude	Opinions Library (Engl)
always ready to help	Entire (no compounds)	<input type="checkbox"/>	PositiveAttitude	Opinions Library (Engl)
always smile	Entire (no compounds)	<input type="checkbox"/>	PositiveAttitude	Opinions Library (Engl)
always taking care	Entire (no compounds)	<input type="checkbox"/>	PositiveAttitude	Opinions Library (Engl)
always there to help	Entire (no compounds)	<input type="checkbox"/>	PositiveAttitude	Opinions Library (Engl)
ambitious	Entire (no compounds)	<input type="checkbox"/>	PositiveAttitude	Opinions Library (Engl)
amiable	Entire (no compounds)	<input type="checkbox"/>	PositiveAttitude	Opinions Library (Engl)
amicable	Entire (no compounds)	<input type="checkbox"/>	PositiveAttitude	Opinions Library (Engl)
approachable	Entire (no compounds)	<input type="checkbox"/>	PositiveAttitude	Opinions Library (Engl)
approachable	Entire (no compounds)	<input type="checkbox"/>	PositiveAttitude	Opinions Library (Engl)
attended	Entire (no compounds)	<input type="checkbox"/>	PositiveAttitude	Opinions Library (Engl)
attentative	Entire (no compounds)	<input type="checkbox"/>	PositiveAttitude	Opinions Library (Engl)
attention to detail	Entire (no compounds)	<input type="checkbox"/>	PositiveAttitude	Opinions Library (Engl)
attentive	Entire (no compounds)	<input type="checkbox"/>	PositiveAttitude	Opinions Library (Engl)
bend backwards to help	Entire (no compounds)	<input type="checkbox"/>	PositiveAttitude	Opinions Library (Engl)
big smile	Entire (no compounds)	<input type="checkbox"/>	PositiveAttitude	Opinions Library (Engl)
care about	Entire (no compounds)	<input type="checkbox"/>	PositiveAttitude	Opinions Library (Engl)
care about my issues	Entire (no compounds)	<input type="checkbox"/>	PositiveAttitude	Opinions Library (Engl)
care about my problems	Entire (no compounds)	<input type="checkbox"/>	PositiveAttitude	Opinions Library (Engl)
care deeply about	Entire (no compounds)	<input type="checkbox"/>	PositiveAttitude	Opinions Library (Engl)

Fig 4.5: SPSS Positive Words Library

2. **Microsoft Excel** -Microsoft Excel is a spreadsheet developed by Microsoft for Windows, Mac OS X, Android and iOS. It features calculation, pivot tables, graphing tools. The twitter data is cleaned and analyzed using Microsoft Excel once sentiment analysis and categorization is done.
3. **Tableau** - Tableau, developed by Tableau Software, is software used for business intelligence and analytics. It produces family of interactive data visualization based on data.

4.8 Steps involved in sentiment analysis and category creation

1. **Extraction of Concepts and Patterns:** The extraction engine identifies candidate key terms. These key terms are grouped under a main concept. Concepts are then grouped into types, which are a collection of similar words such as positive opinion words, words related to customer support etc.

For example: Words and patterns such as overcharge, cheap fare, surge pricing, inexpensive can all be grouped under the concept 'fare'.

Depending on the dictionary the word matches, it is categorized into different types.

For Example: Overcharge, Surge Pricing come under 'NegativeBudget' type in Opinions Library.

Cheap Fare, Inexpensive come under 'PositiveBudget' type in Opinions Library.

2. **Refine and fine tune extraction results:** The automatic extraction results depend on the existing linguistic resources that come bundled with the software. The results need to be fine-tuned for more accurate results. Fine tuning can be done in the following ways:
 - a. Adding new libraries, types and words specific to the domain of words being analyzed. For example: Words related to taxi industry can be added to the linguistic resources for generating more accurate 'concepts'.
 - b. Adding new synonyms to the existing words from the corpus being analyzed.
 - c. Excluding irrelevant words and patterns from further extractions.

- d. Changing the ‘type’ of current words in the context of the text being analyzed. For example: The word ‘cool’ can be generally associated as a positive feature when associated with humans, however if the comments being analyzed are from a feedback to a restaurant, ‘cool’ would be a negative word to associate with food.
3. **Build Categories:** The software uses linguistic techniques to build categories based on the concepts and types extracted.

For example: <Driver> + <Negative> may be a category with all the terms related to type <driver> and <negative> together.

4. **Refine Categories:** The categories built automatically need further refinement for more accurate results. Fine tuning of categories can be done in the following ways:

- a. **Define category rules:** Rules can be defined to make categories. One can combine the different ‘types’ and ‘concepts’ extracted in previous steps to make categories. Boolean operators AND, OR and NOT can be used to combine types.

For example:

Category ‘Negative Service Quality’ can be defined rules such as:

[waiting time] & <cab> & (<Negative> | <NegativeFeeling Emoticon>)

- b. Manually forcing comments to a particular category.
- c. Combining similar categories.

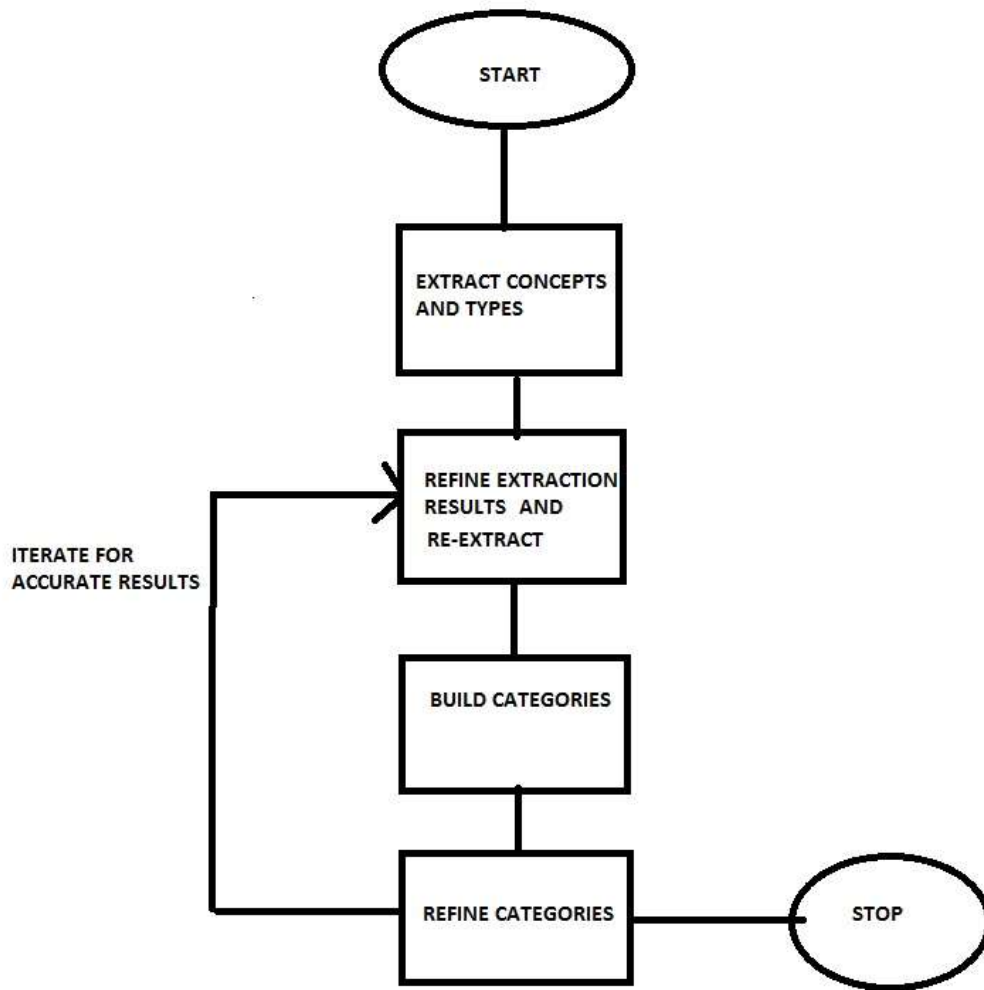


Fig 4.6: Flowchart for Sentiment Analysis and Category Creation

ANALYSIS, INFERENCES AND RECOMMENDATIONS

5.1 Tweet Analysis for Uber India

A corpus of 2096 tweets addressed to @uber_india was analyzed to identify customer sentiment polarity. The records were further classified into categories such as Customer Support, App, Driver Related Issues, Service Quality, Contextual to identify areas which attracted maximum positive and negative comments.

970	at_uber_india plus i called him and told him where are you! when i told him it's not me. he dint even bother cancelling the trip.	Service Quality Negative
1642	at_uber_india uber udsipur doesn't have a twitter id?	Service Quality Negative
1644	at_uber_india why is there an error in adding a payment method while registering a new id? i've been trying since morning!! frustrated	Service Quality Negative
1688	at_uber_india one of the uber driver doesnt knowv how to cancel a trip in tere end after accepting the booking :0.	Service Quality Negative
2033	at_uber_india pls giv attention on issue 95% of ur chennai drivers are cancelling the trip if it's above 15km reason is incentive on 15 trips	Service Quality Negative
5	at_uber_india i know surge price works. it's just since the last month there is a constant surge in pricing	Fare Negative
247	at_uber_india seems to have surge everywhere every night in mumbai! can't get a cab even on a tuesday night without surge price. at_uber_mumbai	Fare Negative
364	at_uber_india at_sirjadeja at_suhelseth shame on uber ..inspite of numerous cabs on road..they are charging 3times more than normal fare..cheat!	Fare Negative
413	at_uber_india i don't c any logic behind u chargin surge price whn thr r more thn 5 cabs avlbl at d same location. surge pricing z extortion.	Fare Negative
513	at_uber_india i don't c any logic behind u chargin surge price whn thr r more thn 5 cabs avlbl at d same location. surge pricing z extortion.	Fare Negative

Fig 5.1: Sentiment Analysis and Category creation for Uber India data in SPSS TAS.

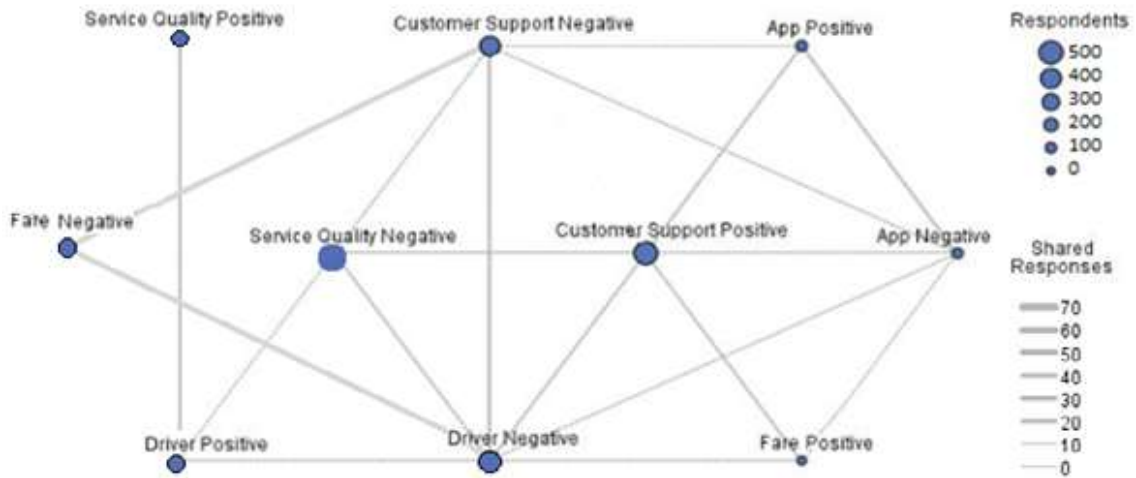


Fig 5.2: Category Web for Uber India data

Service Quality attracted the maximum number of comments, followed by customer support, driver experience, fare related issues, app and contextual.

Out of the total of 2096 tweets, 1314 tweets corresponded to negative sentiment, 441 to positive sentiment and rest 341 were neutral/contextual.

	Positive	Negative	Total
Fare	22	201	223
Service Quality	82	404	486
Customer Support	195	226	421
Driver Experience	105	311	416
App	37	172	209
Contextual			341

Table 5.1: Category wise negative and positive responses for Uber India

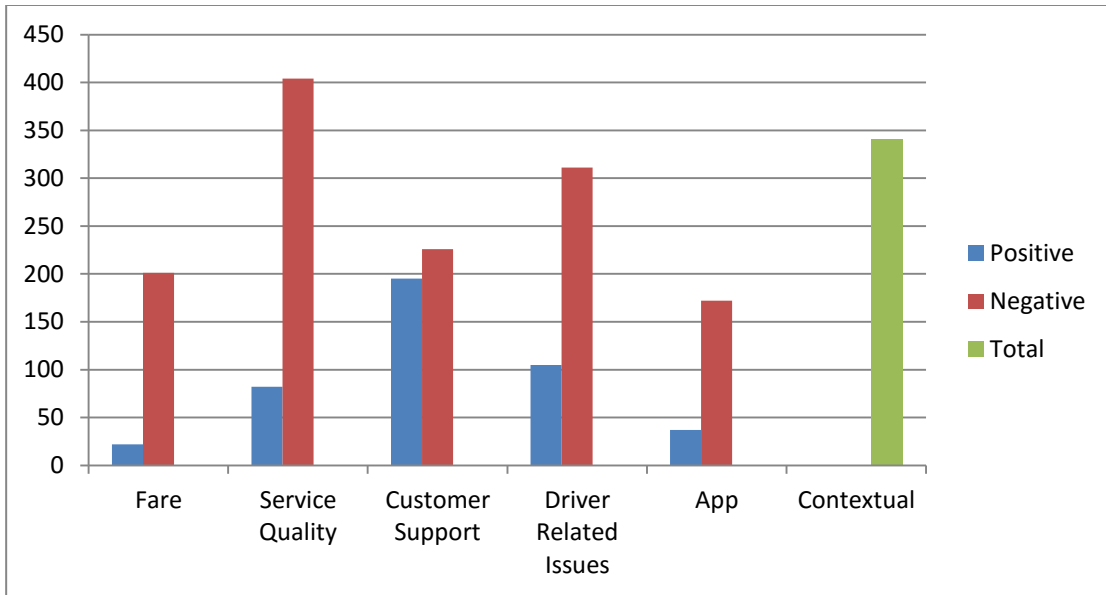


Fig 5.3: Distribution of positive and negative tweets across various categories for Uber Cabs



Fig 5.4: Word cloud for high frequency themes extracted from Uber India data

Main themes which extracted with high frequency include surge pricing, pathetic service, cancelled trip, prompt reply, quick reply, quick response, support team, pool ride, totally frustrated etc.



Figure 5.5: Word Cloud of High Frequency Positive and Negative Words for Uber Cabs

As seen from the above figure, high frequency negative words such as cancel, problem, bad, complaint, declined, pathetic, wrong, unable, frustrating, expensive, overcharge, not fair, loot etc. were extracted. Amongst the positive words, thanks, fast, excellent, available, problem resolved, answered properly, resolved were common which mainly reflect effective customer support system of Uber.

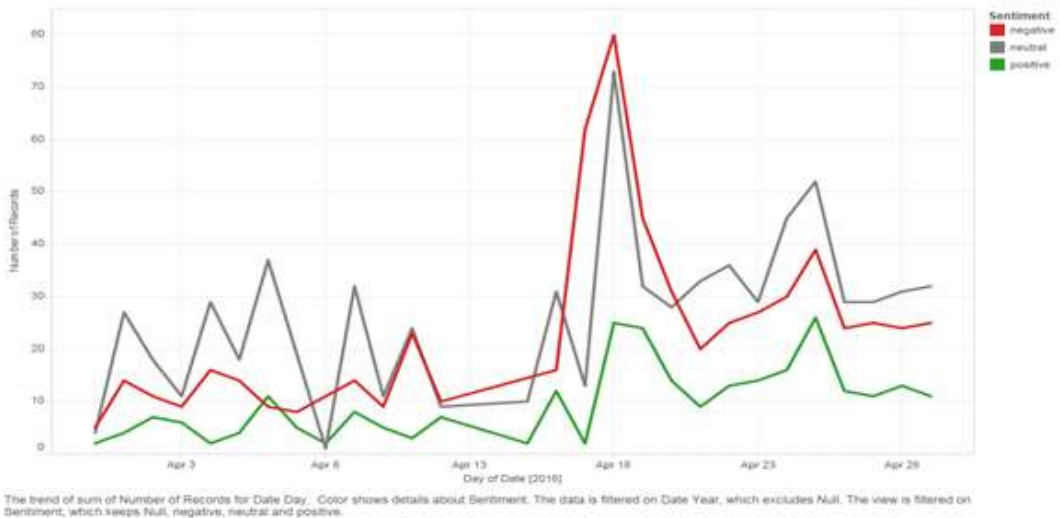


Fig 5.6: Trend analysis of sentiment over the month for Uber cabs

The above graph shows the line plot for sentiments throughout the month. A sharp spike in the number of tweets particularly negative and neutral tweets was observed around 18th April – 25th April 2016.

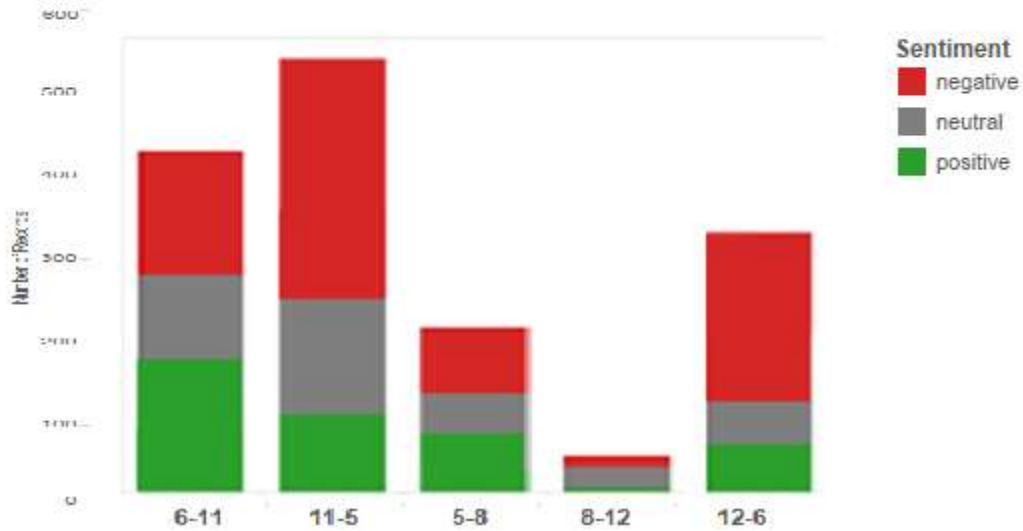


Fig 5.7: Trend analysis of tweets over different time slots for Uber Cabs

From the above graph, it is observed that maximum number of tweets were captured in 11 am-5 pm slot, followed by 6 am-11am slot and then by 12 am – 6 am slot.

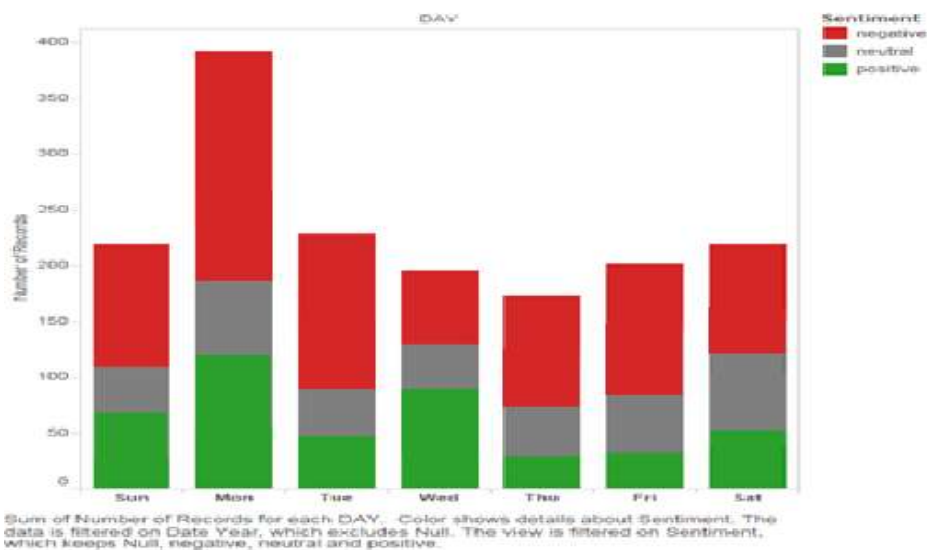
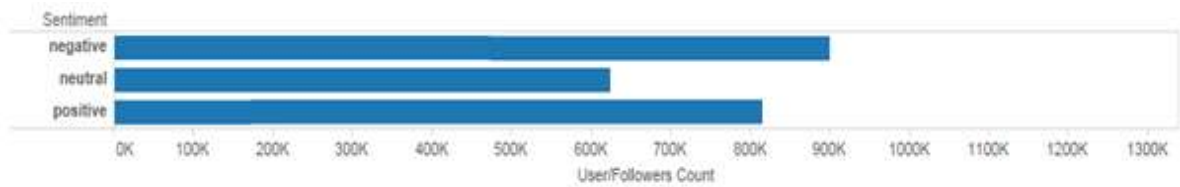


Fig 5.8: Trend analysis of tweets over different days for Uber Cabs

From the above graph, it is observed that Monday received the maximum number of tweets with tweet numbers gradually decreasing over the week.



Sum of User/Followers Count for each Sentiment. The view is filtered on Sentiment, which keeps negative, neutral and positive.

Fig 5.9: Follower count of users who tweeted vs. sentiment of posts for Uber Cabs

The above graph shows the negative posts reach the maximum number of users, followed by positive and neutral posts. This post audience analysis can help a company measure the damage to its reputation by the negative posts and measure positive marketing by the positive posts.

5.1.1 Inferences and Recommendations for Uber India:

1. Based on the above data, service quality has attracted maximum share of negative comments and the company needs to focus in this area. The customers also faced issues related to surge prices, overbilling, long waiting times, and booking getting cancelled, unavailability of cabs and drivers being rude and unprofessional.
2. Customer support attracted more positive than negative comments, positive being related to problem being resolved, helpful; answering queries properly, knowledgeable executives.
3. “Surge Pricing” was the main captured theme from the analysis which was mainly captured during the ‘odd-even scheme’ period in Delhi.
4. Issues with apps were also common.
5. A sharp spike in the trend analysis plot of sentiment over the month was observed around 18th April 2016 (refer figure 5.5) which was the first Monday after the odd-even scheme in Delhi started. A sudden rise in the number of negative tweets was observed. This could be correlated to the consumers expressing their negative sentiments regarding unavailability of cabs, service issues, surge pricing etc.
6. The number of tweets was highest in the 11 am – 5 pm slot followed by 6 am – 11am slot and then by 12am to 6 am slot which shows that maximum tweets

were posted during office hours, followed by the tweets during the time of travel, followed by post-midnight tweets. Monday observed the maximum number of tweets with the number gradually dipping throughout the week. Customer support can be done managed better using this data to handle heavy volume of tweets addressed to customer care for providing quick and effective response.

7. There is a small difference in the reach of positive and negative posts which is good for the health of the brand.

5.2 Tweet Analysis for Ola Cabs

A corpus of 1430 tweets addressed to @ola_support was analyzed to identify customer sentiment polarity. The records were further classified into categories such as Customer Support, Fare Related, App, Driver Related Issues, Service Quality, Contextual to identify areas which attracted maximum positive and negative comments.

Out of the total of 1430 tweets, 834 tweets corresponded to negative sentiment, 345 to positive sentiment and rest 251 were neutral/contextual.

1	709681410026508030	olacabs dont ever book your ride,'ride now' is better as driver can deny his duty end moment and u can miss your train or flight	book/ride/ride+<Contextual>
2	724893439024881020	at_ola_supports i have forwarded the correct booking reference over dm. pl check . the entire confirmation msg is forwarded	book/booking+<Positive>
3	714095792038936060	at_ola_supports olacabs 1) i did 2nd immediate booking as well which got cancelled from yr side even after providing details	book/booking+<Positive>
4	714058479267409020	olacabs worst service in ahmedabad . just before 5 min from scheduled time they cancelled the booking . i booked 3 hrs in advance.	book/booking+<Negative> service/service+<Negative>
5	704479313249218050	pathetic service by olacabs . at_olacabs . cancelled my 5:45 am booking at 5:40 am. cancelled the booking at will of driver .	book/booking+<Negative> service/service+<Negative>
6	712780811792162050	olacabs prebooking not trustworthy . you may end up receiving a msg that your booking has been cancelled just at the pickup scheduled time	book/booking+<Negative> olacabs/olacabs+<Negative>
7	724950055094905960	at_ola_supports not able to submit the rating and it throws an error as a result i am not able to book cabs . check no of attempts to submit	book/booking+<Negative>
8	724953142366994050	at_ola_supports i missed my train . even after booking the cab much in advance , your driver s cancel the booking just 2 minutes bfr arriving	book/booking+<Negative>

Fig 5.10: Sentiment Analysis and Category creation for Ola Cabs data in SPSS TAS.

	Positive	Negative	Total
Fare	72	191	263
Service Quality	63	223	286
Customer Support	95	195	290
Driver Related Issues	71	166	237
App	64	39	103
Contextual			251

Table 5.2: Category Wise Positive and Negative Responses for Ola Cabs

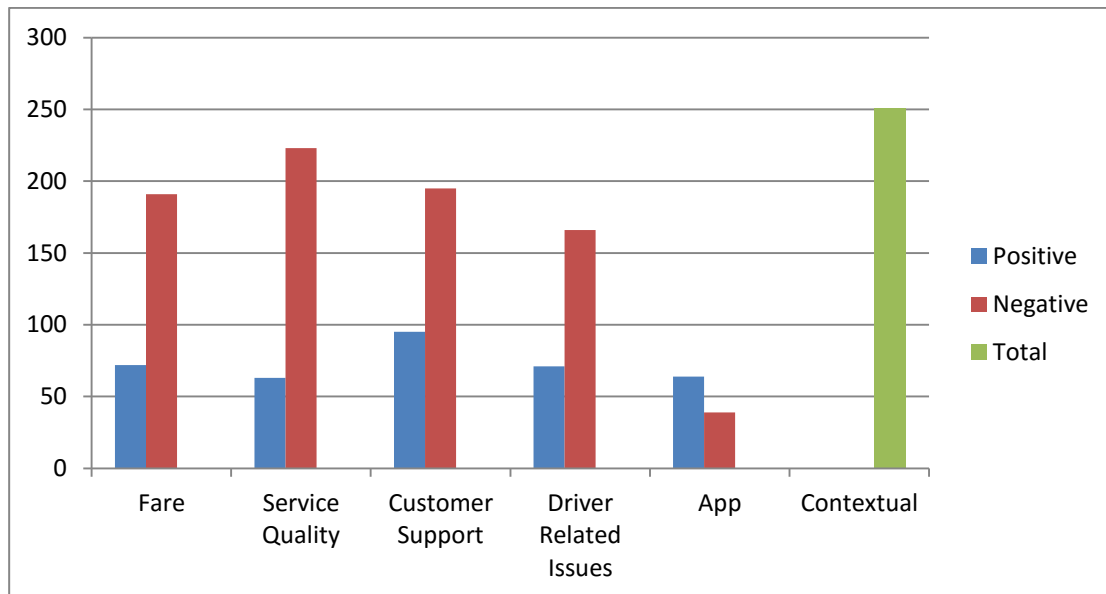


Fig 5.13: Distribution of positive and negative tweets across various categories for Ola Cabs

Customer support attracted maximum number of comments followed by service quality, fare related issues, driver experience, app and contextual.



Fig 5.14: Word cloud for high frequency themes extracted from Ola cabs data

Surge pricing, deal of the day, booming taxi market, pathetic service, legal recourse, fare estimates, bad customer service, pathetic Ola cabs etc. were the main themes extracted from the data.



Figure 5.15: Word Cloud of High Frequency Positive and Negative Words for Ola Cabs

High frequency positive words like better, updated, 😊 (happy smiley), great, hope, quick, best etc. were extracted from the data. Negative words include unprofessional, pathetic, problem, worst, ridiculous, missed etc.

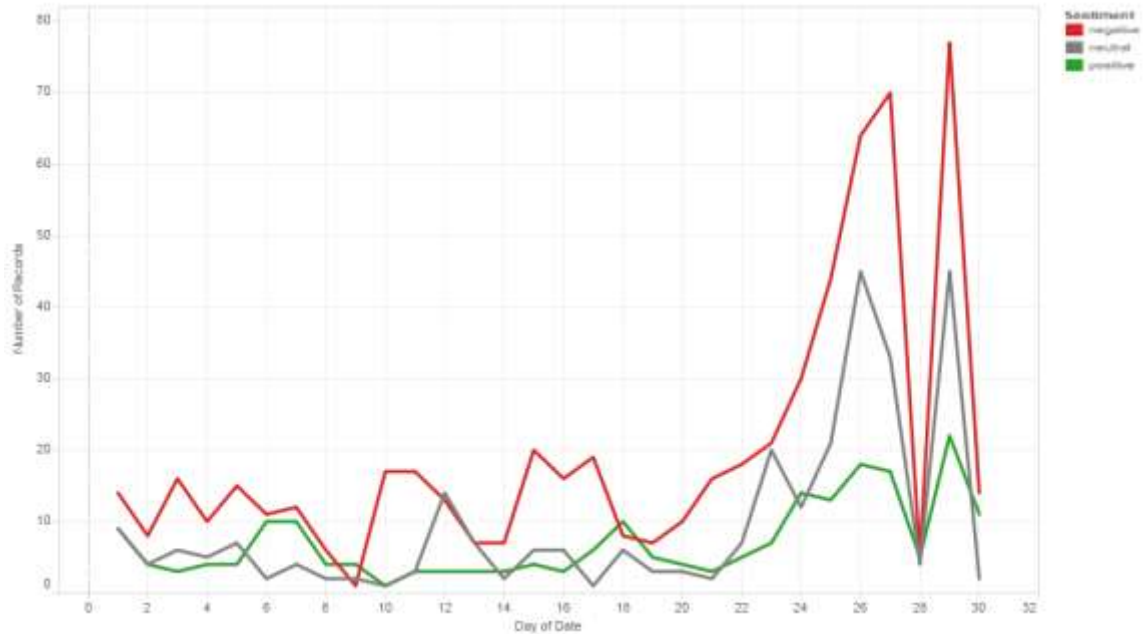


Fig 5.16: Trend analysis of sentiment over the month for Ola Cabs

The above graph shows the line plot for sentiments throughout the month. A sharp spike in the no of tweets particularly negative and neutral tweets was observed around 24th April to 30th April 2016.

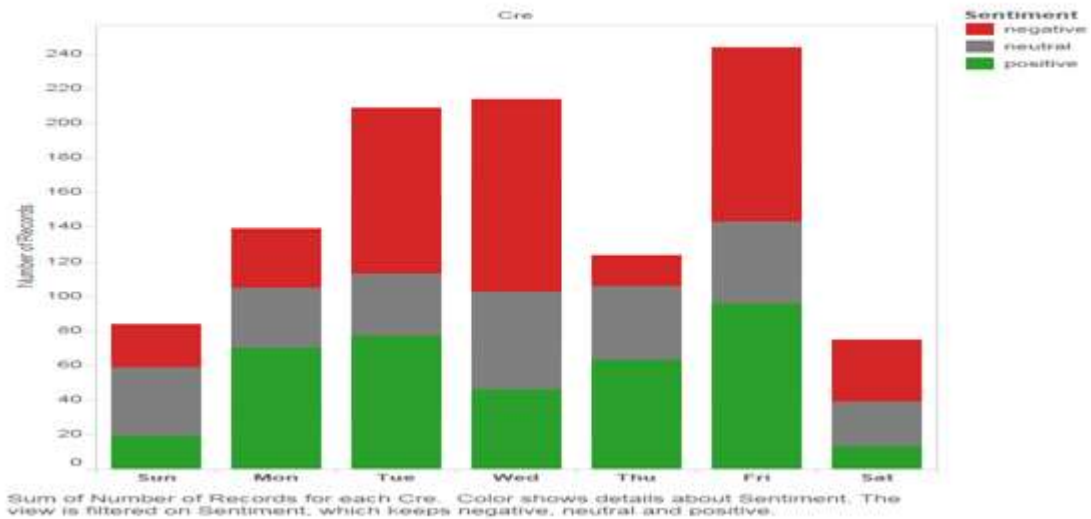


Fig 5.17: Trend analysis of tweets over different days for Ola Cabs

From the above graph, it is observed that Friday received the maximum number of tweets followed by Tuesday and Wednesday.

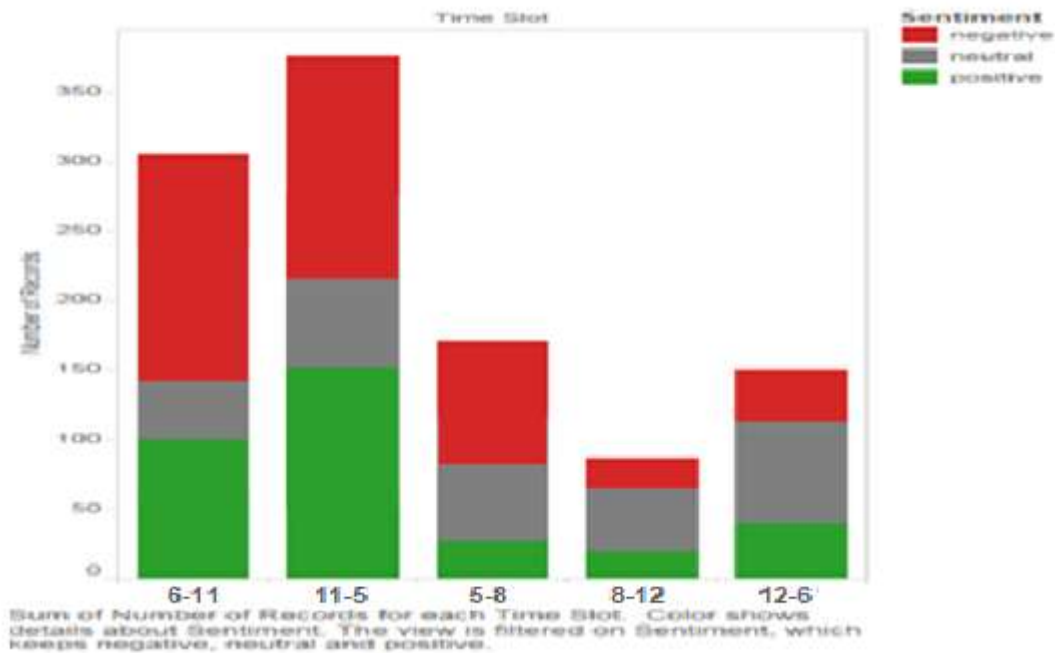


Fig 5.18: Trend analysis of tweets over different time slots for Ola cabs

From the above graph, it is observed that maximum number of tweets were captured in 11 am - 5 pm slot, followed by 6 am-11 am slot and then by 12 am – 6 am slot.

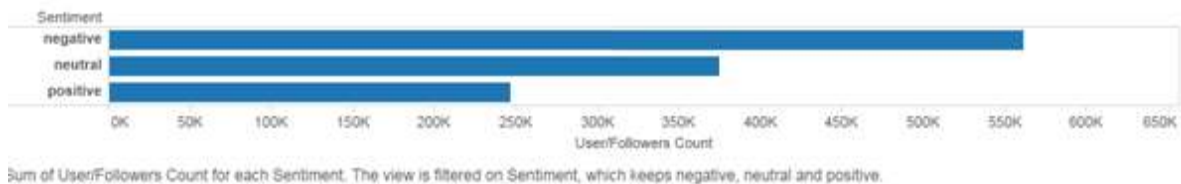


Fig 5.19: Follower count of users vs. sentiment of posts for Ola cabs

The above graph shows the negative posts reach the maximum number of users, followed by neutral and positive posts. This can help a company measure the damage to its reputation

5.2.1 Inferences and Recommendations for Ola Cabs:

1. Service quality issues, fare related issues and customer care received similar share of negative comments with service quality issues leading slightly than others. Bad customer service, surge pricing, pathetic service, legal recourse, fare estimates, pathetic Ola cabs were the main themes extracted which reflect the same. The company needs to focus on these areas.

2. Surge pricing was again the main theme in a lot of tweets which may be correlated to odd-even rule in Delhi as Delhi has a huge customer base for Ola.
3. Words such as unprofessional, pathetic, problem, worst, ridiculous, missed etc. dominated the list of negative words which again reflect on service quality issues and driver behavior.
4. Issues with apps were in less proportion as compared to Uber cabs. Apps is the only category which received more positive than negative comments.
5. A sudden spike in the number of tweets particularly negative and neutral tweets was observed around 24th April – 30th April 2016 which can be correlated to the operational issues including non-availability of cabs, surge pricing etc. during the odd-even scheme period.
6. The number of tweets was highest in the 11 am – 5 pm slot followed by 6 am – 11am slot and then by 12am to 6 am slot which shows that maximum tweets were posted during office hours, followed by the tweets during the time of travel, followed by post-midnight tweets. Friday observed the maximum number of tweets followed by Tuesday and Wednesday. Customer support can be done managed using this data to handle high volume of tweets for providing quick and effective response.
7. There is a huge difference between the reach of positive and negative posts which is not a good sign for the health of the brand.

LIMITATIONS OF THE STUDY

1. Sarcasm can't be detected by automated sentiment analysis tools as it involves use of positive words to express a negative meaning or vice versa.
2. The precision of sentiment analysis tool is not 100% and the tool can report false positives and false negatives.
3. The recall of sentiment analysis tool is not 100% and the tool can possibly report less number of positives and negative comments than the actual number.
4. All the tweets associated with the taxi aggregators during the period for which the analysis is done could not be captured as users can use any expression in hashtags to express their opinions. For example: The tweets addressed to ola cabs can be addressed with @ola_cabs, @ola_support, #olacabs, #ola_cabs, #OlaSucks, #OlaIsAwesome etc. There is no standard convention of addressing companies in tweets and hence it is impossible to consider all the cases.
5. Only one month's data was used to conduct the study which could be a limiting factor in determining the trends.

CONCLUSION

Tweets addressed to twitter handles of Ola cabs and Uber cabs for the month of April were extracted and analyzed. The tweets were analyzed for their sentiment polarity in terms of positive, negative and neutral polarities. Further, the study categorized the tweets into various service KPIs including Service Quality, Fare related issues, App Performance, Driver Issues, Customer Support etc and found areas which required improvements and areas of customer delight. The study also extracted various entities to understand the general trends in the tweets.

Monthly trend analysis, time slot trend analysis and weekday-volume of tweets trend analysis was done to uncover insights in the tweet patterns. Tweet audience reach analysis graphs were plotted to determine positive and negative effect of tweets to the brand reputation. Recommendations based on the data and its analysis were made to various departments of the aggregators.

The study concludes that the social media data can indeed be a rich source of information which, if harnessed by the marketers, can lend organisations an upper edge over its competitors. Analysis of this huge chunk of unstructured data can lead to actionable insights and help marketers in better understanding of customer's behaviour, perceptions and feelings. It can help organisations in improving their products and services by listening to their customers in real time. Organisations should build suitable talent and invest in social media mining to reap the benefits of this data.

REFERENCES

Shintaro Okazaki, Ana M. Diaz Martin, Hector David et al (2014). Using Twitter to engage with customers: a data mining approach. Retrieved from: www.emeraldinsight.com/1066-2243.htm.

Fotis Misopoulos, Miljana Mitic, Alexandros Kapoulas et al (2014). Uncovering customer service experiences with Twitter: the case of airline industry. Retrieved from: www.emeraldinsight.com/0025-1747.htm

Apoorv Agarwal, Boyi Xie, Ilia Vovsha et al (2011), SENTIMENT ANALYSIS OF TWITTER DATA. Retrieved from: <http://www.cs.columbia.edu/~julia/papers/Agarwaletal11.pdf>.

SocialBakers.com. (2016). Social Media Marketing, Statistics and Monitoring Tools. Retrieved on 30th April 2016, from <http://www.socialbakers.com/>.

Ernst Young (2015), Social Media Marketing India Trends Study 2nd Edition. Retrieved on 30th April 2016, from [http://www.ey.com/Publication/vwLUAssets/EY-social-media-marketing-india-trends-study-2014/\\$FILE/EY-social-media-marketing-india-trends-study-2014.pdf](http://www.ey.com/Publication/vwLUAssets/EY-social-media-marketing-india-trends-study-2014/$FILE/EY-social-media-marketing-india-trends-study-2014.pdf)

Ravikant Raj Blog (2015), How to build a twitter sentiment analyzer? Retrieved on 30th Jan 2016, from: <http://ravikiranj.net/posts/2012/code/how-build-twitter-sentiment-analyzer/>

Tillkeyling.com (2016), A workaround for Twitter's Search-API limitations: Using the Twitter Websearch and Facepager. Retrieved on 30th Jan 2016, from: <http://tillkeyling.com/a-workaround-for-twitters-search-api-limitations-using-the-twitter-websearch-and-facepager.html>

Inc42.com (2016). The Evolution Of The Indian Taxi Market - Comparing The Biggies Retrieved on 30th Jan 2016, from: <https://inc42.com/features/evolution-indian-taxi-market-comparison/>

Statista.com(2016). The statistics Portal | Statistics and facts on internet usage in India. Retrieved on 30th Jan 2016, from: <http://www.statista.com/topics/2157/internet-usage-in-india/>

What is opinion mining (sentiment mining)? - Definition from WhatIs.com. (2016). SearchBusinessAnalytics(2016). Retrieved 6 May 2016, from <http://searchbusinessanalytics.techtarget.com/definition/opinion-mining-sentiment-mining>

Businesstoday.in (2016). Taxi wars |How Ola and Uber are making other taxi companies irrelevant in India. Retrieved 6 May 2016, from <http://www.businesstoday.in/magazine/cover-story/india-taxi-market-war-heats-up-ola-cabs-uber-strategy-leaders/story/222542.html>

Sentiment analysis. (2016). Wikipedia. Retrieved 6 May 2016, from https://en.wikipedia.org/wiki/Sentiment_analysis

Opinion Mining. (2016). swshar. Retrieved from <https://swshar.wordpress.com/2016/04/14/56/>

Techinasia.com(2016), Ola hints cheapest cabs get 1 million rides a day, beats Uber in India. Retrieved 6 May 2016, from <https://www.techinasia.com/ola-micro-cabs-expansion-india>

ANNEXURE A

Screenshot of JSON response from Twitter API.

```
[{
  "created_at": "Thu Apr 21 04:14:39 +0000 2016",
  "id": 723001982437326800,
  "id_str": "723001982437326848",
  "text": "@MeruCabs @ArvindKejriwal and how about quality of service? There is no comparison about
the same",
  "entities": {
    "hashtags": [],
    "symbols": [],
    "user_mentions": [{
      "screen_name": "MeruCabs",
      "name": "Meru Cabs",
      "id": 1021937828,
      "id_str": "1021937828",
      "indices": [
        0,
        9
      ]
    }
  ],
  {
    "screen_name": "ArvindKejriwal",
    "name": "Arvind Kejriwal",
    "id": 405427035,
    "id_str": "405427035",
    "indices": [

```

ANNEXURE B

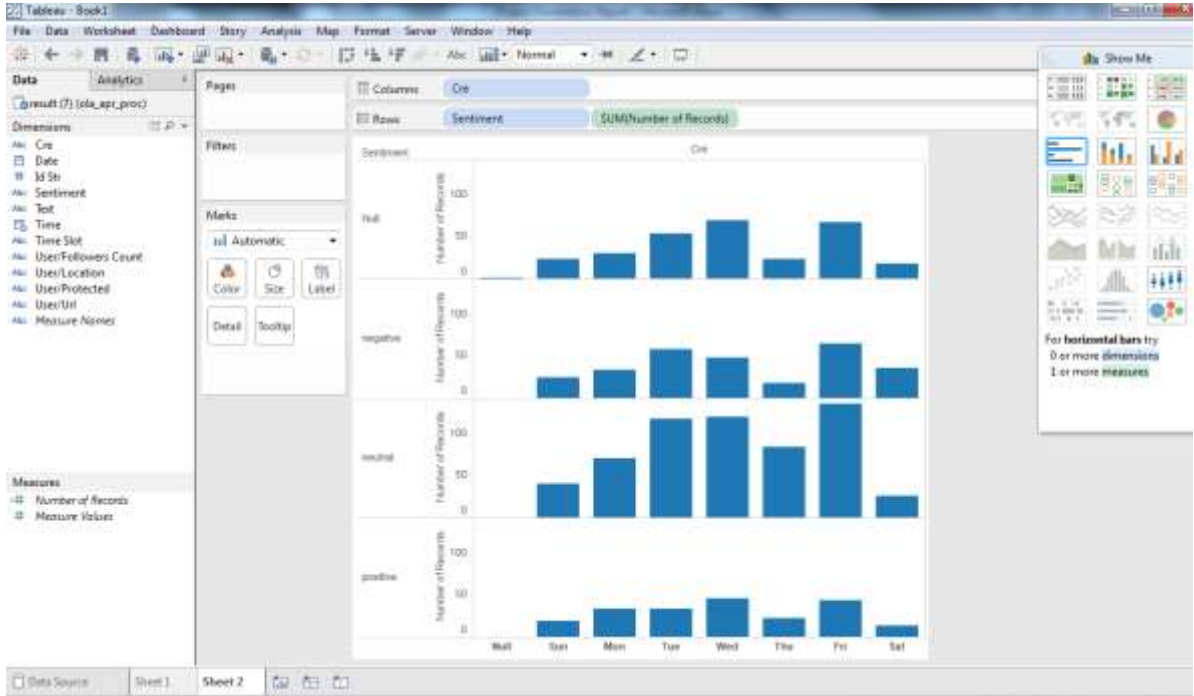
Screenshot of Sentiment Analysis by SPSS Text Analytics for Surveys

The screenshot displays the SPSS Text Analytics for Surveys interface. On the left, a 'Categories' pane shows a hierarchical tree of sentiment categories with counts for '#Decide' and '#Response'. Below it, the 'Unused Extractions' pane lists various extraction rules. The main window shows a list of 11 survey responses, each with a unique ID, the original text, and the assigned sentiment category.

ID	Response	Categories	Ratio
1	AT_Online - Do more? Another of your 1st April surveys? OnDate, OnRooms, AprilFools	Negative	-
2	On auto-connect to VAFI a your On Price is just a little less. AT_Online AT_Online, OnRooms, OnDate, OnRooms	PositiveSentiment	-
3	Yes... Now I can take a booking. OnDate is nice to try OnRooms, "	PositiveSentiment	-
4	Great! On rooms, the On are interesting. OnRooms https://booking.onrooms.com	PositiveSentiment	-
5	OnDate to make VAFI Facility. OnRooms Across Categories. Book URL.	PositiveSentiment OnRooms/OnDate/Positive	-
6	OnDate will open. OnRooms, OnRooms to some extent at segments including VAFI. Most even automatically. On? Much needed in Singapore.	-	-
7	OnDate Special Discount Super: Use 100% Off as code when booking. On?	-	-
8	AT_Online all these or others about OnRooms, many of them are really dumb. onrooms onrooms onrooms	Negative	-
9	AT_Online searching. OnRooms on 1st April why not to smart. OnRooms?	Negative OnRooms/Negative	-
10	An OnDate relative to AT_Online. OnRooms, & reminds me of OnDates from last year. OnDate.	PositiveSentiment	-
11	Is it April Fools day relative?? AprilFools. OnRooms. OnDate.	-	-

ANNEXURE C

Screenshot of Tableau Worksheet



ANNEXURE D

Screenshot of output file for Uber cabs.

	A	B	C	D	E	F	G	H	I
1	DAY	DATE	TIME	TIME SLOT	id	text	sentiment	Category	user/location
32	Fri	29-Apr	6:28:11 AM	6-11	31	at_uber_india at_saila88shama wah ra	negative	Driver Experience Negative	
40	Fri	29-Apr	8:05:27 AM	6-11	39	at_uber_india i have been a loyal custom	negative	Driver Experience Negative	Bangalore , India
45	Fri	29-Apr	2:37:35 PM	11-5	44	at_uber_indiahow can you make right the positive		Driver Experience Negative	New Delhi, Delhi
48	Fri	29-Apr	2:29:02 PM	11-5	47	at_uber_india rocks. at_ola drivers are	ic negative	Driver Experience Negative	
51	Fri	29-Apr	4:32:12 PM	11-5	50	at_uber_india but my driver is saying you	neutral	Contextual	Anywhere
65	Fri	29-Apr	11:51:24 AM	11-5	65	at_uber_india so dis driver has a problem	negative	Driver Experience Negative	Mumbai
70	Fri	29-Apr	8:48:20 AM	6-11	70	at_uber_india how can i get sms of my tri	neutral	Contextual	INDIA
81	Fri	29-Apr	4:21:05 AM	12-6	81	at_uber_india : inate .ur drivers keep th	negative	Fare Negative	New Delhi
83	Fri	29-Apr	2:19:42 PM	11-5	83	at_uber_india at_uber_blr why should i	neutral	Fare Negative	Bengalore
89	Thu	28-Apr	7:09:17 PM	5-8	89	at_uber_india driver behaving very rudely	negative	Driver Experience Negative	Mumbai, Maharashtra
96	Fri	29-Apr	4:14:22 PM	11-5	96	at_uber_india one of your drivers actual	neutral	Contextual	Haryana, India
97	Fri	29-Apr	5:04:02 AM	12-6	97	at_uber_india at_tarinivifisigh and if n	neutral	Contextual	
112	Thu	28-Apr	4:22:22 AM	12-6	112	at_uber_india my driver ended my trip by	negative	Driver Experience Negative	
113	Wed	27-Apr	5:35:10 PM	5-8	113	at_uber_india why am i charged? - i had t	negative	Driver Experience Negative	earth
117	Wed	27-Apr	5:28:32 PM	5-8	117	at_uber_india why do your driver's alway	negative	Driver Experience Negative	Mumbai
124	Thu	28-Apr	2:36:45 PM	11-5	124	at_uber_india i had to take 1.4 surge now	positive	Fare Negative	
129	Thu	28-Apr	1:43:46 PM	11-5	129	at_uber_india jst booked a cab driver wz	neutral	Driver Experience Negative	
139	Thu	28-Apr	2:06:54 PM	11-5	139	at_uber_india done, had a terrifying expe	negative	Driver Experience Negative	Bangalore
141	Wed	27-Apr	4:08:35 PM	11-5	141	at_uber_india this is the 3rd time your dr	negative	Service Quality Negative	
145	Thu	28-Apr	2:05:39 AM	12-6	145	at_uber_india how much of the surged pr	neutral	Fare Negative	Berlin, Germany
159	Wed	27-Apr	2:38:13 PM	11-5	159	at_uber_india what do we do about drive	neutral	Driver Experience Negative	Noida

ANNEXURE E

Screenshot of output file for Ola cabs

day	date	time	time slot	id	str	text	user/location	sentiment
Tue	26-Apr	8:45:29 AM	6-11	1	at_ola_supports	>> please go and listen to the call records where i asked your customer care to book me a cab and he told me th		neutral
Tue	26-Apr	12:36:50 PM	11-5	2	at_ola_supports	at_olacabs worst service even though it say cab available it never gets booked. url	Hyderabad	negative
Tue	26-Apr	8:48:08 AM	6-11	3	at_ola_supports	>> consumer court and show this as proof how you are customer for a ride. thanks for showing true colour at_bi		neutral
Tue	26-Apr	12:27:34 PM	11-5	4	at_ola_supports	after being assured by pr team dat i can collect it anytime i was sent back as the 'relevant' was not functional on		negative
Tue	26-Apr	1:15:36 PM	11-5	5	at_ola_supports	not able to submit the rating and it throws an error as a result i am not able to book cabs. check no of attempts t		negative
Tue	26-Apr	12:25:49 PM	11-5	6	at_ola_supports	any update?	Back 2 chennai	#N/A
Tue	26-Apr	1:15:01 PM	11-5	7	at_ola_supports	y alt num? i m v much available on my no. at_olacabs. these silly reasons unprofessional at_4ni	Chennai	negative
Tue	26-Apr	10:01:23 AM	6-11	8	at_ola_supports	yesterday's mumbai cheerday		neutral
Tue	26-Apr	8:35:14 AM	6-11	9	at_ola_supports	your ceo needs to train you on reading at_bhash. if you check, it is there in post at_nikesharora	Delhi, India	neutral
Tue	26-Apr	12:46:28 PM	11-5	10	at_ola_supports	yes and it shows 2mins and now it doesn't booked before it was no surge these guys don't take i	Hyderabad	neutral
Tue	26-Apr	12:33:00 PM	11-5	11	at_ola_supports	and that too during peak hours team. this was the most ridiculous experience i ever had with at	Chandigarh, In	negative
Tue	26-Apr	9:33:02 AM	6-11	12	at_ola_supports	i work in hsr layout. is there a office nearby of urs?	Mumbai	neutral
Tue	26-Apr	10:16:02 AM	6-11	13	at_ola_supports	thx.	Bhopal & Pune	#N/A
Tue	26-Apr	9:51:59 AM	6-11	14	at_ola_supports	receipt for booking id tfs-pp-c66898026		#N/A
Tue	26-Apr	10:23:20 AM	6-11	15	at_ola_supports	karnataka govt has advised not to collect peak time charges but u guys are still cheating the publ	Bangalore	negative
Tue	26-Apr	8:37:33 AM	6-11	16	at_ola_supports	or are you saying that you will refund me money when i take my next ride		neutral
Tue	26-Apr	2:19:00 PM	11-5	17	at_ola_supports	why u are taking this much time to credit the money?this is really disappointing.		negative
Tue	26-Apr	9:34:29 AM	6-11	18	at_ola_supports	thanks for addressing the issue.	New Delhi	#N/A

ADHERENCE SHEET

	Date	Signature
Proposal Submission	5 th April 2016	
Data Collection and Analysis	12 th April 2016	
First Draft	19 th April 2016	
Final Report Submission	26 th April 2016	