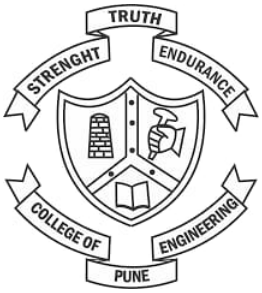
College Of Engineering, Pune



Sentiment Analysis on Reliance Jio Reviews

Guided By
Dr. Y. V. Haribhakta
Ms. Rugwedi Mam

Report By
-Vivek Bhand (111903129)
-Vrushabh Patil (111903130)
-Yadnyadeep Khadke(111903131)

04 May 2022

Contents

SENTIMENT ANALYSIS ON RELIANCE JIO REVIEWS

111903129 Vivek V. Bhand[1*]
111903130 Vrushabh P. Patil[2*]
111903131 Yadnyadeep N. Khadke[3*]

*Corresponding E-mails :
[1] bhandvv19.comp@coep.ac.in
[2] patilvp19.comp@coep.ac.in
[3] khadkeyn19.comp@coep.ac.in

PROBLEM STATEMENT.
To classify the reviews using Sentiment Analyzer and studying the analysis of customers Satisfaction over period of time and over different regions.
To build Sentiment Analyzer using Logistic And Vader Algorithms .

ABSTRACT. The purpose of this project is to analysis the satisfaction of Jio Network Customers . This project will be helpful to inculcate changes in the service , as per the Customer needs.

KEYWORDS: Sentiment Analysis, Sentiment Polarity Categorization , nltk , Customers Reviews

1 Introduction

For the Data Science project, we have decided to develop a web scrapper to scrap review website for the reviews available on the Reliance Jio network and internet service. Our aim is to analyse the scrapped data to classify into Satisfied, Unsatisfied, and Neutral based on the sentiment of reviewer. Our aim to find the general overview of the application user's sentiment toward the services provided by the Reliance Jio, as are they satisfied with the product. Also, we aim to analyse their satisfaction level towards features such as.

- Internet Speed
- Network coverage
- Quality of voice call
- Customer support

- Flexible data plan
- Affordable rates on data and talk-time

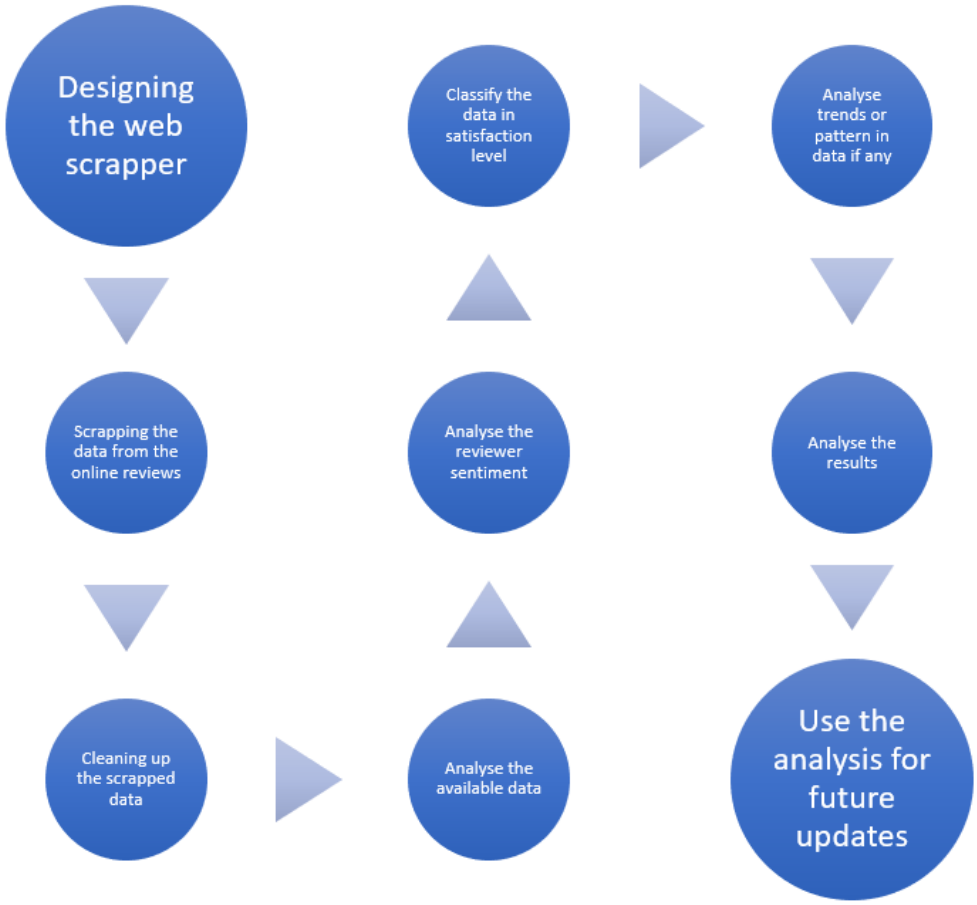
As the insight to the user's approach towards the product s crucible for the future development of the network company, our aim to find when the user are most satisfied with the feature so it may be further developed or when to limit error when there is unsatisfaction in the user-base. For the process of the data scrapping, we went to the Reliance Jio review web page on the `https://www.mouthshut.com/mobile-operators/Reliance-Jio-reviews-925812061` Mouthshut website (A renowned website for users to give reviews) to get access to the reviews. The tools we used for the Web scrapping are as follows:

- Octo-parse
- Web scrapper extension for Chrome
- Custom Web Scrapper

2 Project Objectives

- Classify the sentiments into Satisfied / Unsatisfied / Neutral
- Classify the data region wise and year wise
- Identifying the area for improvement and monitoring the variation of satisfiability over years
- Building Sentiment Analyzer to classify the network related comments

3 Design Model



4 Data Preprocessing

- Adding latitude longitude to attributes using geopy library
- Removing NaN values
- Removing extra white spaces, special characters
- Converting number to words in reviews
- Converting date of reviews into number of days ago
- Normalization of data
- Removing unnecessary attributes

- Removing outliers

4.1 Lemmatization

In the lemmatization, we have used nltk open source modules. The number of processes we have included for this are as follows:

- Tagging reviews as Nouns, verbs, adjectives and None.
- Removing irrevelent Stop words
- Lemmatization to convert any word to corresponding root words

5 Exploratory Data Analysis

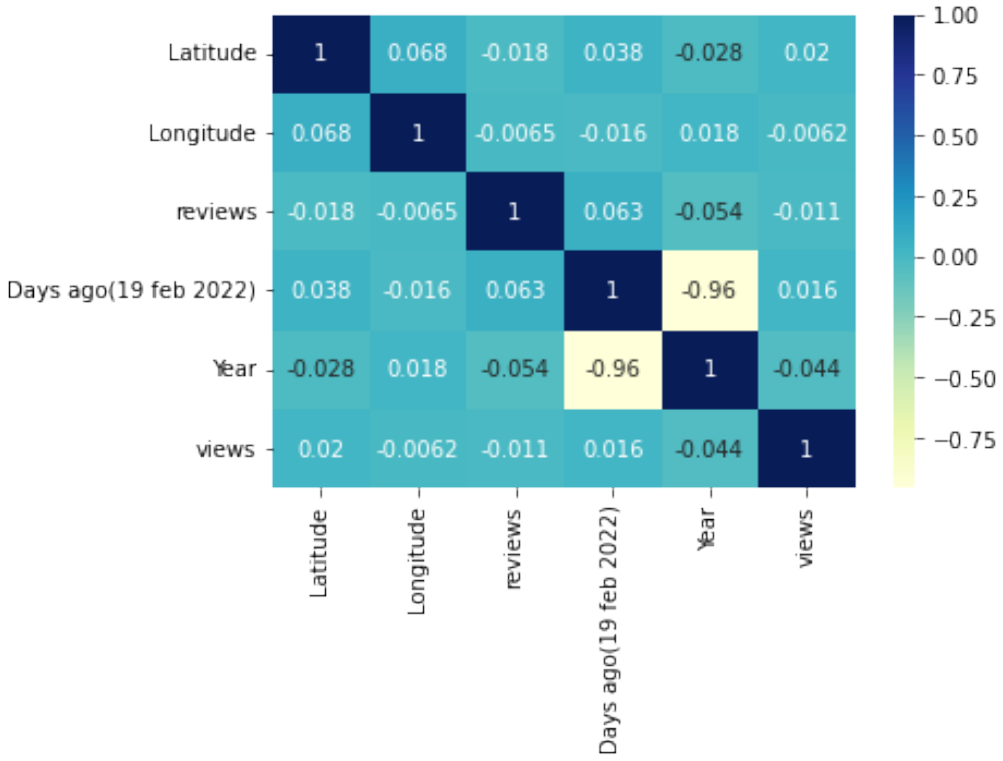
CORRELATION BETWEEN ATTRIBUTES : HEATMAP, PAIRPLOTS

GEOSPATIAL PLOT BASED ON LOCATION OF REVIEWER

REGION-WISE DISTRIBUTION OF REVIEWS

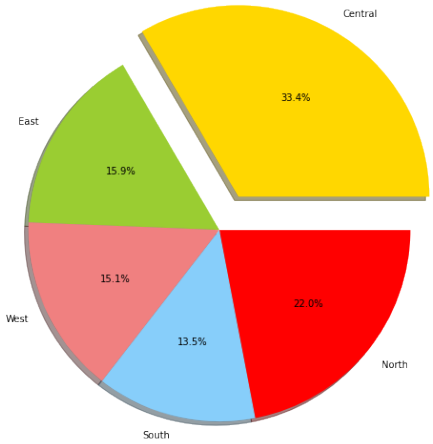
5.0.1 HeatMap

To get how various attributes are inter-linked with each other, we took heatmap to see to which extent they are related.



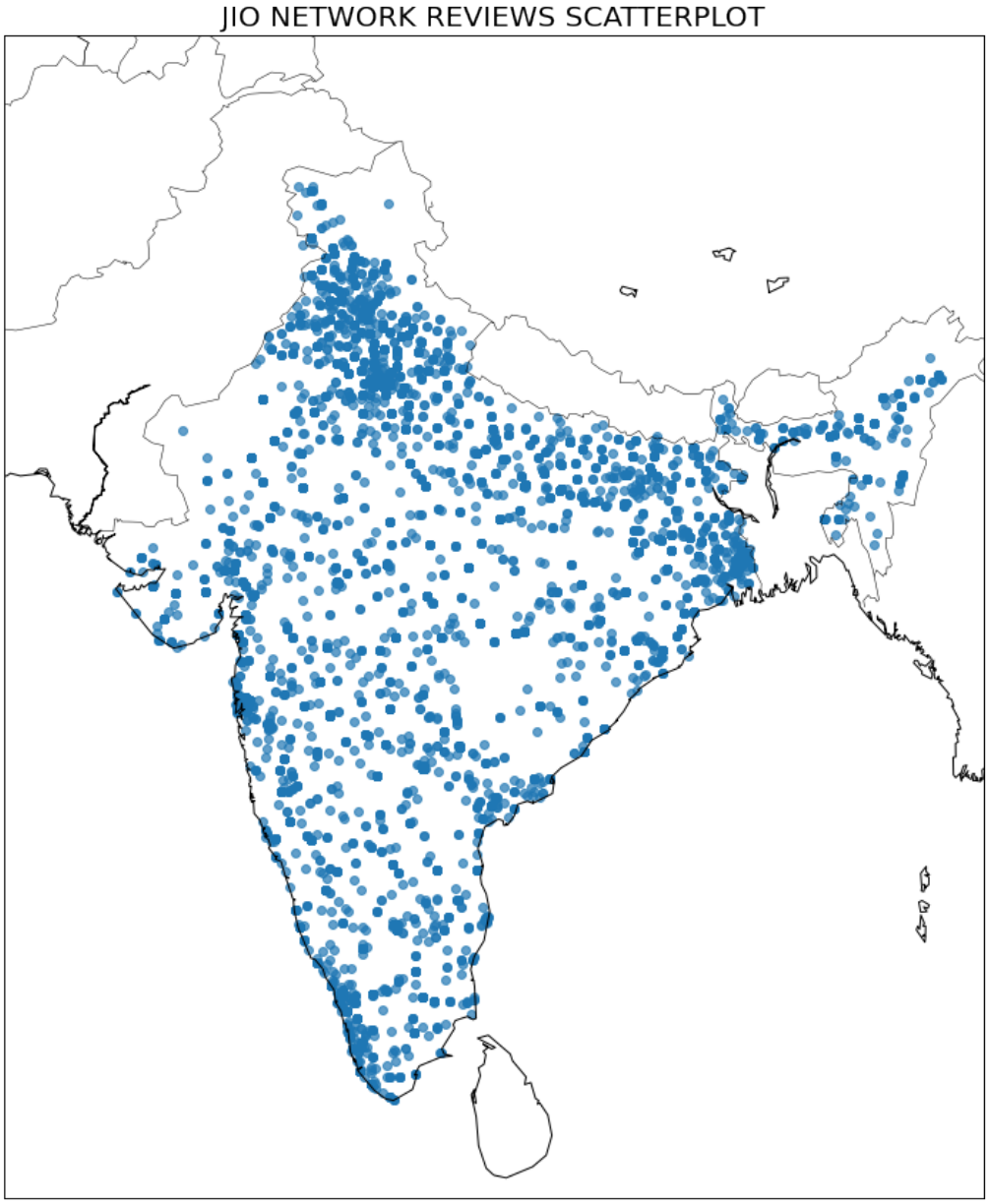
5.0.2 Region wise Count

We classified the scrapped data into States and their respective count. We divided into sections As Central, North, South, East, West sector.



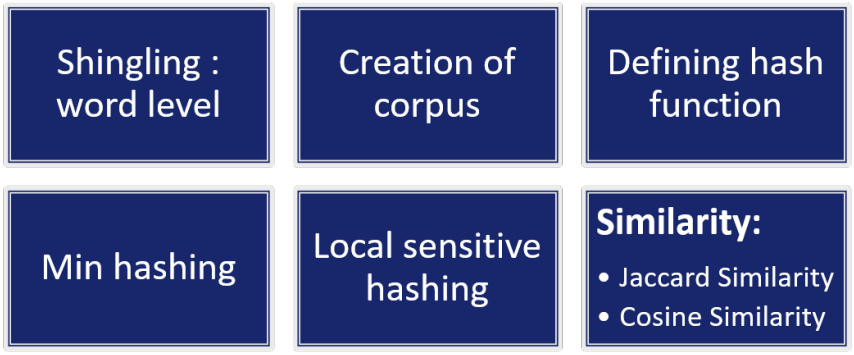
5.0.3 India BaseMap

Plotting all locations of the reviewers in the Map on India.
The Reviews is observed to clustered in the areas of Developed Cities such as Mumbai, Delhi, Kolkata and Kerala.
The possible behind this is high population density in the developed cities, more literacy rate and more number of users.

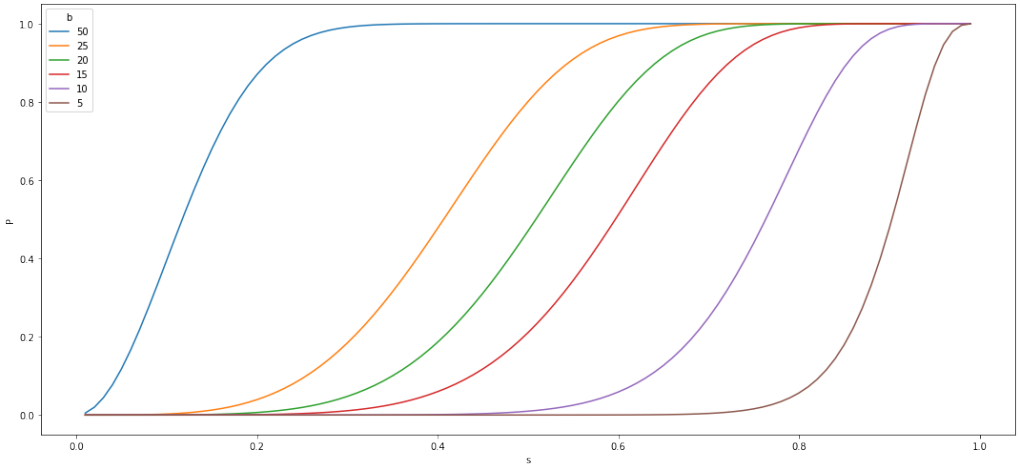

JIO NETWORK REVIEWS SCATTERPLOT

6 Hashing(Word Level)

6.1 Steps Followed :

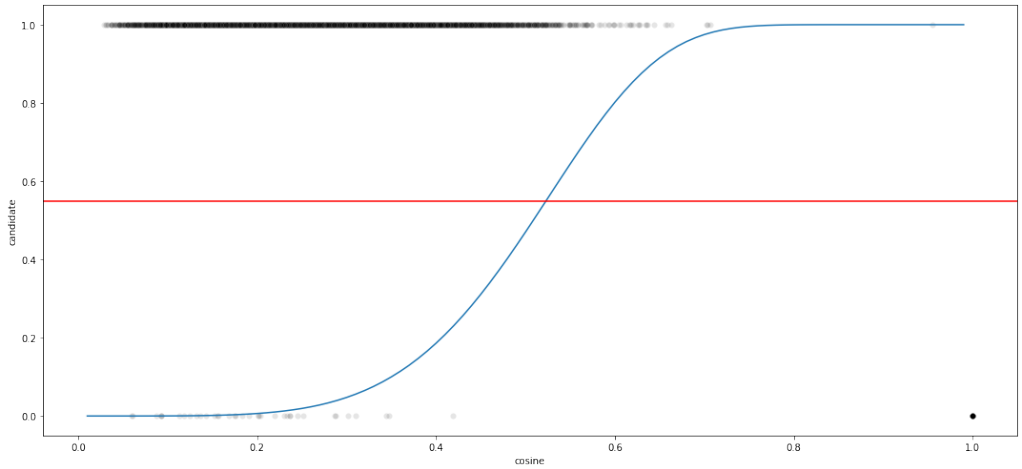


6.2 LSH Graph for Different Band Size:

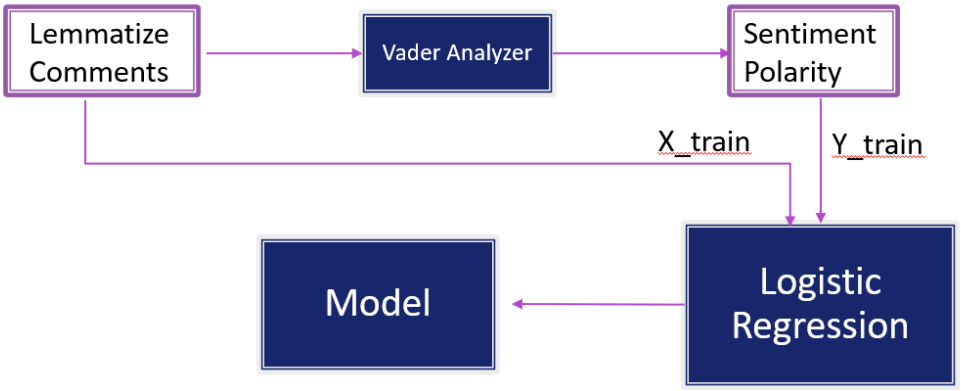


6.3 LSH and Threshold (for Band size = 20)

As threshold is 0.54 Optimal Band Size is 20.

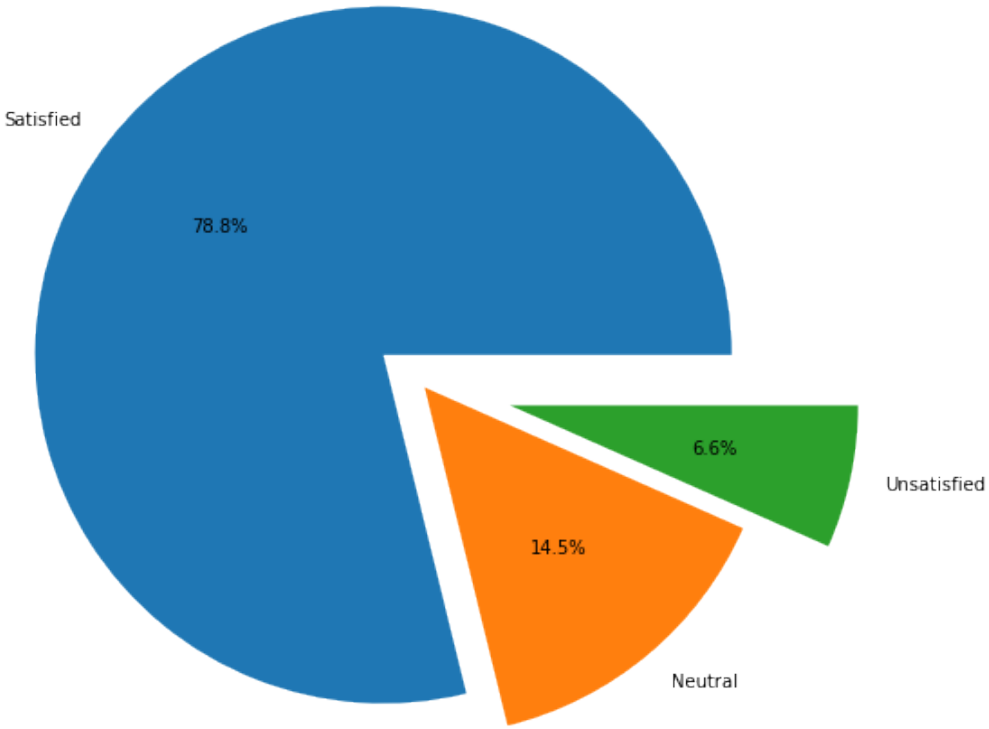


7 Build Model



7.1 VADER

VADER module, Valence Aware Dictionary for Sentiment Reasoning, is used for describing a sentiment score in the inputted string, It calculates compound score based on the positive, negative and neutral words to encountered in the review. We analysed the observed results to classify the reviews into Satisfied, Unsatisfied and Neutral.



8 Analysis :

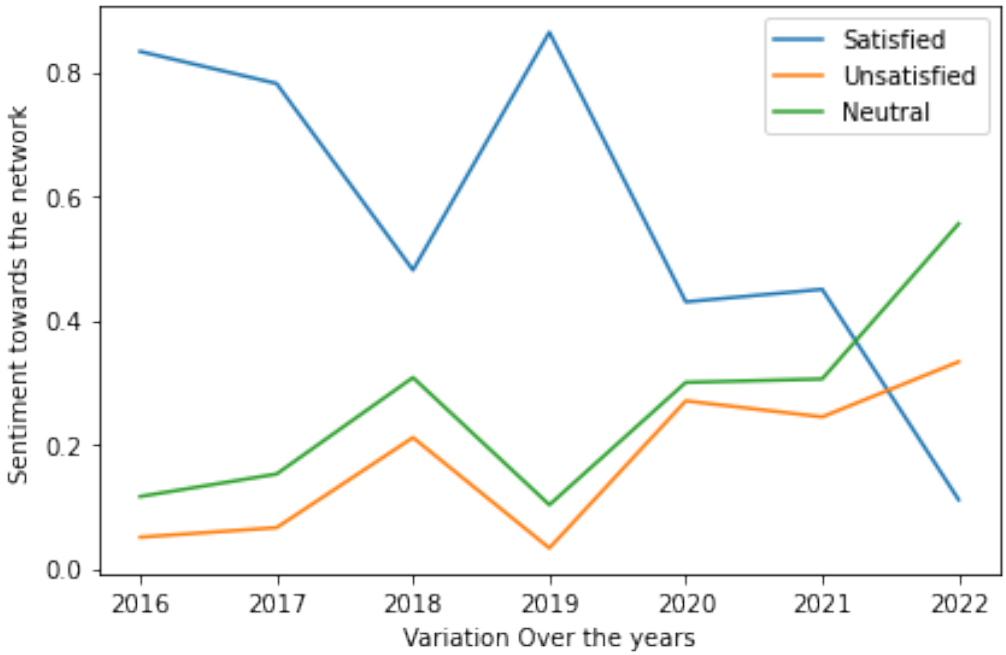
Based on the analysis, we have observed interesting trends in the sentiments based on the location of the user and time of the review.

8.1 Year-wise Analysis :

We have observed that the satisfaction level of the user are decreasing in the years from 2016-2022. The speculated reasons behind the decrease in the satisfaction level may be of

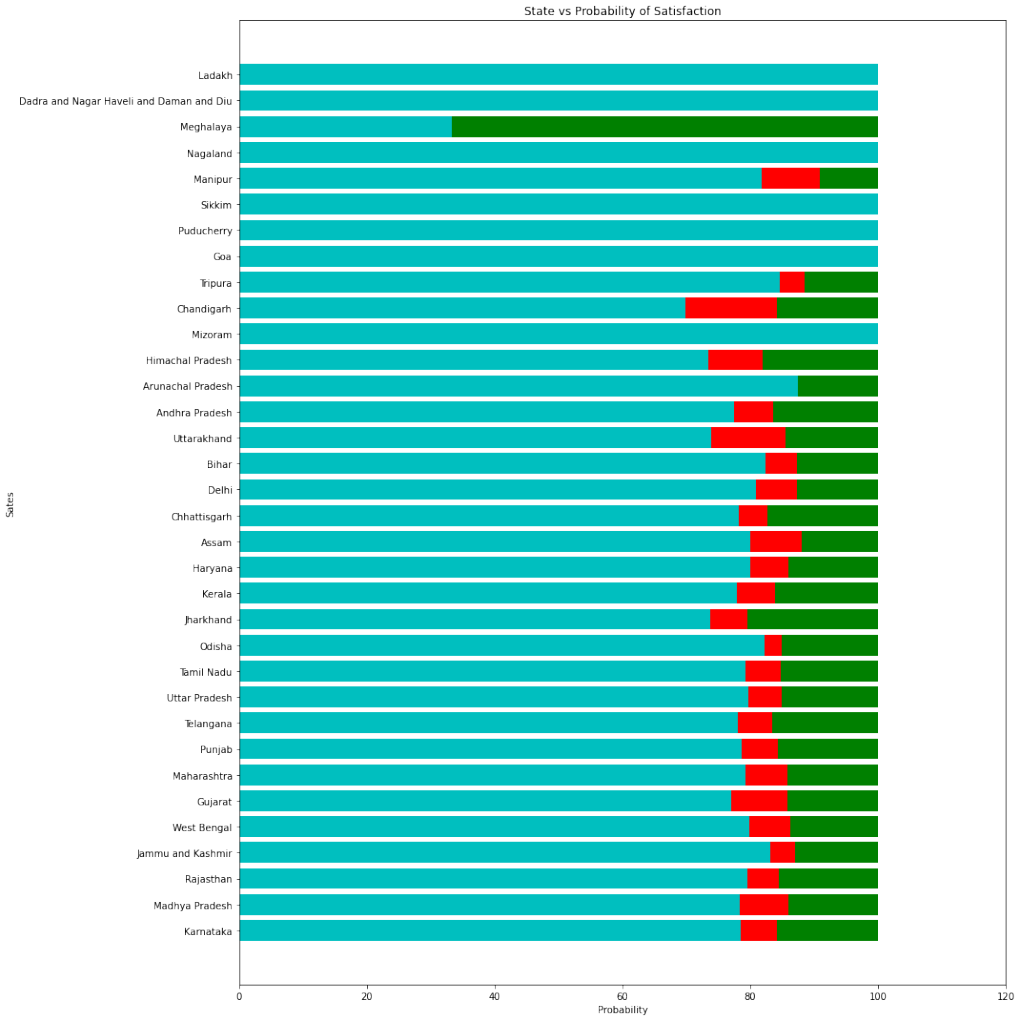
- Inefficient scaling of a good network over increased user data base.
- Number of user have increased exponentially the number of grieving customers was

also observed to be increased.



8.2 State-wise Analysis :

Plotting State-wise distribution of the reviews into satisfaction levels as Satisfied, Unsatisfied and Neutral.



9 Conclusion

In this Data science project, we learned how to extract suitable information, how to clean data to remove noises, outliers and any irrelevant attributes, how to pre-process the data to be suitable for usage, and creating a model for detect the sentiment of user from the reviews.

10 Link for the Project

Google Drive : https://drive.google.com/drive/folders/1WJvYR2JOuJ4FD4KAc4e3DJjPYZUanOV5?usp=sha