SCHOOL OF COMPUTER ENGINEERING

Fast Track Semester -2021

PROJECT REPORT

CSE3020 – DATA VISUALIZATION

Visualizing Data of Unicorn Food Delivery Startup
"Zomato"

Team No. 5

BY:

ARPIT SHARMA 19BDS0027

ADITYA CHAUDHARY 19BCE0544

VIRAJ GUPTA 19BCE2151

Submitted to

Prof. N NALINI

1. ABSTRACT

Zomato is an online website which provides restaurant search and discovery service. It provides its customers a platform to evaluate choices for great places to eat. The purpose of this project is to do an analysis of the way Zomato has scaled up its operations, expanded its business into various countries. The data is retrieved from Kaggle which was extracted by a user from Zomato API. It is about Zomato registered restaurants across several countries. The data has 9552 records and 21 columns. The primary columns used for aggregating are Votes, Aggregate Rating and Price Range. Currently, the average Urban millennials spend a chunk of their income on food. This has opened new opportunities for investing in restaurants across cities in India. We are going to study what factors contribute to customer attraction towards restaurants. The analysis shows how location, budget, quality of food, types of cuisines available and services offered by restaurant, affect average ratings for restaurants.We are also going to develop a restaurant recommendation algorithm using Machine Learning models.Our motive is to find similar restaurants to a given target restaurant.

2. INTRODUCTION

The importance of Data Visualization arises because our human brain required processed data to understand and use. Data is enormously present everywhere but in the raw and assorted format, it has to be transformed into a meaningful, processed, and organized way to bring out value and predictions from it. We need data in the visualized form to identify and examine current trends for better decision-making. Visualized data gives a better understanding for businesses and every corner truly relies on it. User-friendly and beautiful visuals are conveyed more accurately and yet less time-consuming. Time is a very important aspect and with the data visualization concept, every data becomes quick and fast for decision making and interpretation.

In the food industry, the data visualization concept, gives true meaning to the restaurant managers and their customers by providing clear insight from the processed data.

A data-driven program gives value to all the categories of people, including managers, restaurant owners, travelers, businesses all highly rely on the importance of meaningful data. Data visualization gives better control of prediction, analysis, increases security and satisfaction to people. Imagine if we could get the visual representation of the past, present, and future data then it will be very easy to predict any mishappening or downfall that will occur in the future with the help of data visualization. Based on the history, present scenarios the future prediction can be most appropriately generated. We can find out the best other possible alternatives and if not then correct it. The possibilities are endless with the data visualization. It keeps you ahead with the time.

To break down the Zomato dataset for a better visualization of data, which can be used for further statistical analysis by many analyst, business firms and startups. To expand the knowledge of programming in R by using various libraries. To explore the possible opportunities after visualizing the complex dataset. Desire to assist fellow food lovers who are habitual in trying new restaurants and cuisines in finding the most cost effective and good quality restaurants across India. Aid travelers who want to try the best regional food in the tourist region. Help restaurant owners to understand how to attract more customers and Investors to invest in profitable restaurants. It can be used to expand the business through analysis and statistical observation, and they can achieve their goals such as financial strategy, i.e., fund and sales, marketing strategy to increase their clients, globalization strategy, i.e., spreading around the globe.

Food Recommendation System has proven critical by empowering users to overcome information overload, help decision-making, and change user behavior. The suggestion of food products is one subject that has gotten significantly less attention in the past, especially when contrasted with sectors connected to leisure and entertainment. People often are confused about what to eat and what to choose after logging into a website and seeing so many delicious food items. Many people, on average, take 30 minutes before ordering a particular food. Here are a few fundamental questions we asked people in the survey to know their opinion regarding the food choice and the cuisine before selecting and ordering it. On average, more than 80% of people prefer to eat outside food. This implies that they want to try something new in terms of choices and do not want to have the same type of food again and again frequently. It is evident that though people want to eat outside food, around 55% of people find it challenging to choose a restaurant to eat at. The main reason could be the lack of clarity of what to eat and when to eat, maybe it could be based on mood, or they cannot find a restaurant based on the cuisine they want. Maybe this is why people take so much time to find an excellent restaurant. Only around 1/4th of the users can find a restaurant based on their preferred cuisine. The main reason could be that whenever the user goes online to order a portion of food, the user gets diverted by seeing other restaurants, and due to a large number of items available in front of him at that time, he is not able to find the cuisine which he wants to have as per the planned earlier based on mood. More than 80% of people have food based on their mood. The moods could be either sad, happy, lonely, bored, comfort, stressful, frustrated or angry, and many more.

Nevertheless, one can conclude that mood influences the choice of food and is directly proportional to what we eat and order online. However, only about 20% of people can find the restaurant and eat the food based on their mood. This might be because of the lack of clarity in finding the right food at the right time and too many options available in front of the screen, which diverts and tempts one away. So, we can conclude that people are facing problems finding the right food based on their mood at the right time.

3. LITERATURE SURVEY

<u>1)*"A sentiment-enhanced personalized location recommendation system."*</u>

By implementing advanced techniques in data mining like machine learning, it would enhance the scope of this system. Depending on your patterns the system will guide end user breakfast, brunch, lunch and dinner. Every human has a different taste when it comes to food. It can relate patterns of the system to the taste of an individual. Recommendation system provides the best alternative hence individuals have a list of food that can be made with the items available from the inventory. Injecting the approach in the market gains the maximum profit guarantee. Basic idea behind the Mining process is to extract patterns from massive databases. Recognition of specific patterns from databases is a challenging task. By avoiding the redundancy from the patterns specific patterns can be made. Massive datasets have similarities in patterns. Proposed system neglects the similarity between the patterns resulting in best solutions possible. Databases containing massive data might have noisy datasets in it. This will result in a pattern anomaly. Retrieving the specific pattern process extremely beneficial in knowledge and discoveries from processed data to benefit the industries. In comparison to using the Naive Bayes Machine Learning Algorithm, we preferred to use natural language processing which gives us a variety of options in choosing food based on various moods of the person.We have taken into consideration 8 moods.The dataset is worked over by creating a Made comfort_food_reasons as a single string removing all punctuations and performed lemmatization on this string for every row.Based on the user mood input and if it exists in the lemmatized string, we append the corresponding food choice and return it to the user.

2) *"Machine learning based food recipe recommendation system."*

In this paper, they presented two approaches to recommend recipes based on preferences of the user given in the form of ratings and compared them to identify which approach suits the dataset better. Two approaches were used namely, item based approach and user based approach to recommend recipes. For an item based approach Tanimoto Coefficient Similarity and Log Likelihood Similarity would be used to compute similarities between different recipes. For user based approach EuclideanDistance and Pearson Correlation are used. We use similarity techniques of user based approach and introduce fixed size neighborhood and threshold-based neighborhood to the same. The performance of the user based approach is found to be better than the item based approach. The performance for the All Recipe data set is found to be better than the simulated dataset since there are more interactions between users and items. We have worked with 2 main datasets.The first dataset is for recommending food based on mood and the second dataset is for recommending restaurants based on cuisine decided by taking a visited restaurant within Bangalore. Apart from giving options which are concerned with suggesting similar types of food, we have primarily emphasized on giving recommendations and suggestions based on the mood of the person and the type of food the person wants to eat in that particular mood. Apart from this, our focus is also to suggest a cuisine based on the past experience of a person using cosine similarities.This involves using the concept of TF IDF which is Term Frequency-Inverse Document Frequency. A word/term frequency is counted and multiplied with a global component inverse document frequency. In the next step TF IDF model normalizes the result to the unit length. And by using a linear kernel on this TFIDF obtained matrix, we obtain the cosine similarities matrix.

3)*"Survey on recommendation system using data mining and clustering techniques."*

In this we need to create a numerical model to anticipate the accomplishment of the restaurants i.e. to know whether the restaurant is famous or not good at taste. For this we have to make a reasonable technique. In this system we ought to have authentic information of each factor: famous restaurant , on-famous restaurant, music, five star hotels.these variables are given equivalent weightage in the technique to know whether a hotel is good to go or to miss it. At that point dependent on the expressive measurements of the gathered information of every segment the hotel is named famous or not. Administrator is the individual who encourages the information in the hotel . He will include information about specific hotels. This framework will assist us with knowing whether a motion picture is hit or not. We have worked with 2 main datasets.The first dataset is for recommending food based on mood and the second dataset is for recommending restaurants based on cuisine decided by taking a visited restaurant within Bangalore. Apart from giving options which are concerned with suggesting similar types of food, we have primarily emphasized on giving recommendations and suggestions based on the mood of the person and the type of food the person wants to eat in that particular mood. Apart from this, our focus is also to suggest a cuisine based on the past experience of a person using cosine similarities.This involves using the concept of TF IDF which is Term Frequency-Inverse Document Frequency. A word/term frequency is counted and multiplied with a global component inverse document frequency. In the next step TF IDF model normalizes the result to the unit length. And by using a linear kernel on this TFIDF obtained matrix, we obtain the cosine similarities matrix.

and humanic –explain a customer's overall perception of a dining experience that lead to satisfaction or dissatisfaction. A focus on food, menu offerings, ambiance, and service create buzz in social media. Furthermore, several implications inform customer researchers about effective ways to investigate and serve the restaurant industry**.**

4) "*ZOMATO DATA WITH EDA, GEOSPATIAL AND SENTIMENT ANALYSIS Priyadharshini R Department Of Computing, Coimbatore Institute Of Technology, Coimbatore, Tamil Nadu, India."*

Zomato is food delivery company and multinational restaurant aggregator. Zomato Food Delivery service is available in 24 countries and more than 10,000 cities as of 2019. The major objective is to perform Exploratory Data Analysis, Geospatial Analysis and Sentiment analysis using Zomato dataset of particular City "Bangalore". Bangalore is Largest City of South Indian State of Karnataka and also Capital City. Bangalore is also known as "Bengaluru" is 27th Largest city all over the world and 3rd Largest city in India with a population over 15 million. Bangalore also referred as “Silicon Valley of India” or “IT Capital of India”. Bangalore has unique food culture, Different types of Restaurants from all over the world can be found in Bangalore with various kinds of cuisines. Main goal of this project is to find insights, Interesting Facts and figures and drawing valid conclusion with different visualization techniques.

5) “Predicting Restaurant Rating using Machine Learning and comparison of Regression Models , 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)”

The restaurant industry is one of the prevailing competitive sectors. People enjoy cherishing communal dining for centuries, hence the demand for restaurants increasing day by day. Bangalore is a heaven for foodies with a range of cuisines from different parts of the world. In this paper, the data set for restaurants for a specific location is identified and the Data Visualization tools are applied to understand the trends and patterns of the food culture. This paper proposes a model to understand the factors affecting the rating of restaurants. Machine learning and predictive analytics with wide spread range of tools and techniques aids to predict the rating of restaurants. In this paper model is built using various regression algorithms and the most efficient algorithm is considered. The result of this paper helps new restaurants in deciding their menu, cuisine, theme, cost, demographic location etc. thereby increasing the business.

6) "Anil Bilgihan, Soobin Seo & Jihee Choi (2018) Identifying restaurant satisfiers and dissatisfiers: Suggestions from online reviews, Journal of Hospitality Marketing & Management"

Analyzing data from user-generated websites with data mining practices can suggest meaningful insights into services performances. Such techniques may help identifying the precursors of service satisfiers and dis-satisfiers of restaurants. The aim of this research is to analyze online customer reviews for restaurants and underpin the restaurant satisfier and nature of online restaurant reviews. After retrieving 2214 usable customer comments on Yelp.com, using a coded web spider, a series of both qualitative (e.g., tag clouds, word networks, and word trees were created using data visualization techniques) and quantitative (e.g., MANOVA) data analysis techniques were deployed to detect the satisfiers and dis-satisfiers of different types of restaurants. Three distinct types of clues – functional, mechanic,

4. PROPOSED WORK

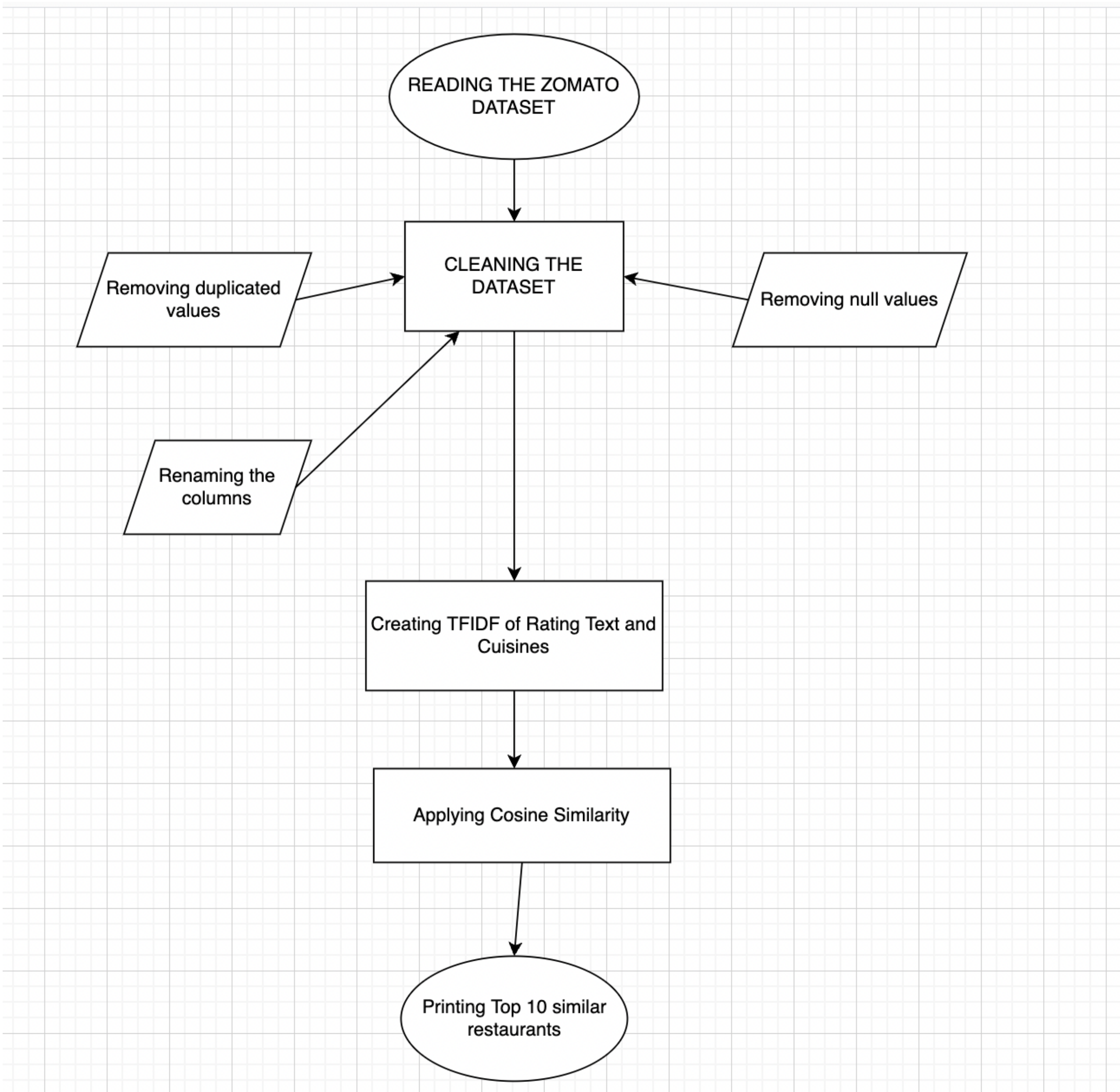
1.)<u>Restaurant Recommender</u>

In this system, user gives a restaurant name as an input and it provides top 10 restaurants similar to that restaurant from the dataset.This recommender uses the concept of cosine similarity.

Cosine similarity-

We use the concept of cosine similarity to identify how close each restaurant is to other restaurants based on the reviews given by users in the data. Cosine similarity is a value between 0 and 1, and helps identify how close one entity is to another. This is calculated using the concept of TF IDF which is Term Frequency-Inverse, Document Frequency. A word/term frequency is counted and multiplied with a global component inverse document frequency. In the next step TF IDF model normalizes the result to the unit length. And by using a linear kernel on this TFIDF obtained matrix, we obtain the cosine similarities matrix.

Architecture Diagram:



2.) Zomato Data Visualization

a. Location of All Restaurants

The first module "Location of all restaurants" shows the availability of Zomato services around the world. This module contains two plots. First is a Geo Map of all the Restaurant's Location which is created using rworldmap library. Second is a Bar Plot which contains the distribution of Zomato Restaurants across Countries.

b. Aggregate Rating

The second module "Aggregate Rating" shows the aggregate rating as a Histogram with 100 break points and also shows the scatter plot alongside showing the number of Ratings.

c. Price Range

The third module "Price Range" shows the distribution of price range of the restaurants in Scatter, Bar and Box Plots.

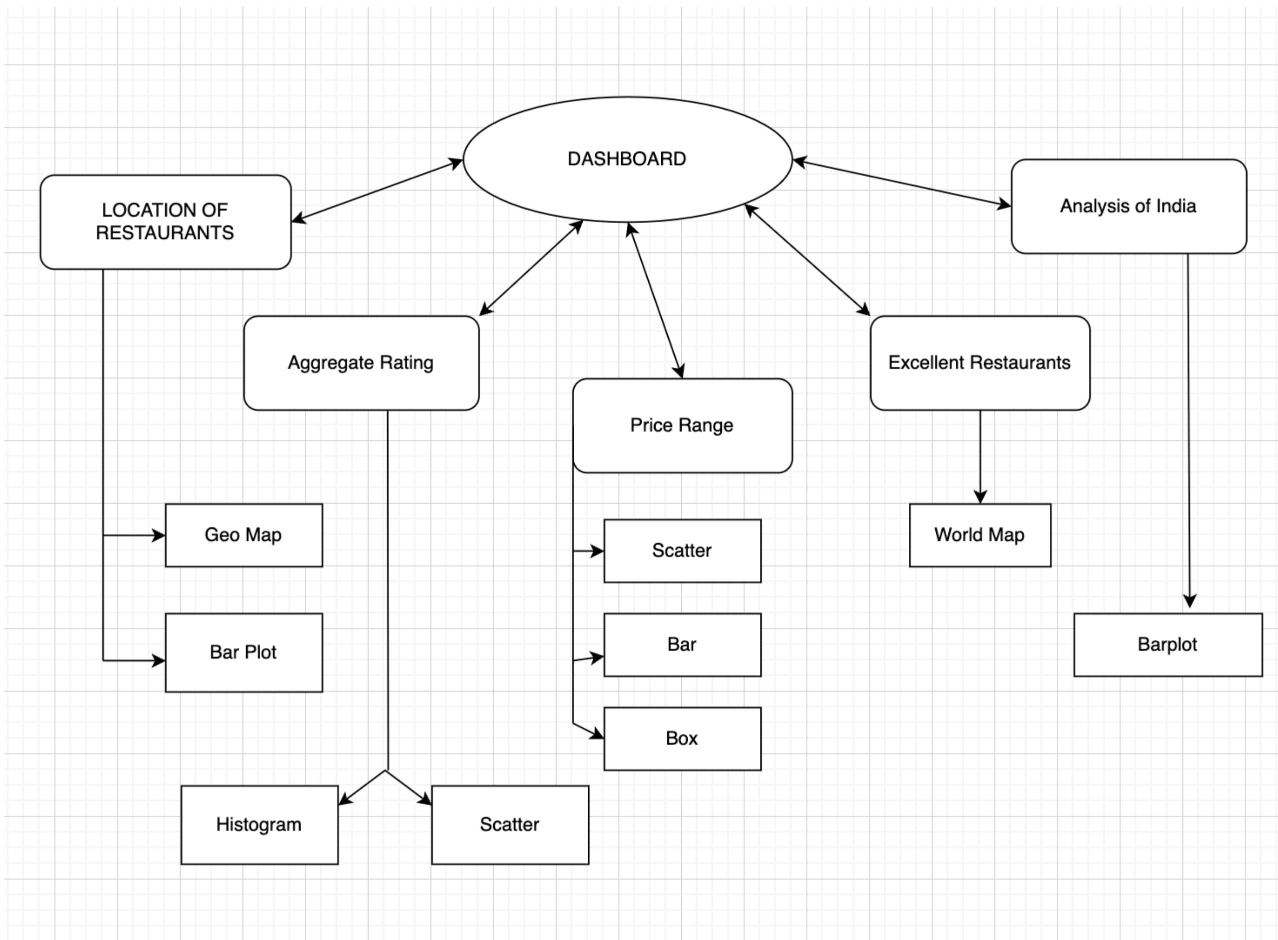
d. Excellent Ratings

The fourth module "Location of excellent restaurants" show exclusively 5 star rated restaurant locations on the world map.

e. Analysis of India

The fifth module "Analysis of India" shows the restaurant ratings, average cost of two, price range distribution, services offered by restaurants, and type of cuisines available in the Restaurants that are located in India.

Architecture Diagram:



5. RESULTS

(i) Restaurant Recommender System:

a. Top 10 Restaurants similar to Ooma.

```
[ ] recommend('Ooma')
```

TOP 10 RESTAURANTS LIKE Ooma WITH SIMILAR REVIEWS:

	Restaurant Name	Mean Rating	cost
9299	Milse	4.9	50.0
9303	Miann	4.9	25.0
304	Atlanta Highway Seafood Market	4.9	25.0
9296	Talaga Sampireun	4.9	200000.0
0	Le Petit Souffle	4.8	1100.0
2400	Spice Kraft	4.8	1200.0
9301	Depot Eatery and Oyster Bar	4.8	90.0
1839	Prankster	4.8	1500.0
352	Earl of Sandwich	4.7	35.0
9315	Eight - The Langham Hotel	4.7	190.0

b. Top 10 Restaurants similar to Vikings

```
[ ] recommend('Vikings')
```

TOP 10 RESTAURANTS LIKE Vikings WITH SIMILAR REVIEWS:

	Restaurant Name	Mean Rating	cost
6618	Dukes Pastry Shop	4.200000	300.0
6620	London Street Kitchen	4.200000	600.0
6617	Dilli Treat	4.200000	500.0
6619	HotMess Bakes	4.100000	750.0
4956	QD's Restaurant	4.033333	800.0
9550	Walter's Coffee Roastery	4.000000	55.0
2609	Biryani Binge	3.900000	600.0
2610	City of Joy	3.900000	800.0
2629	Lemon Drops	3.900000	500.0
8179	Jungle Jamboree	3.825000	1200.0

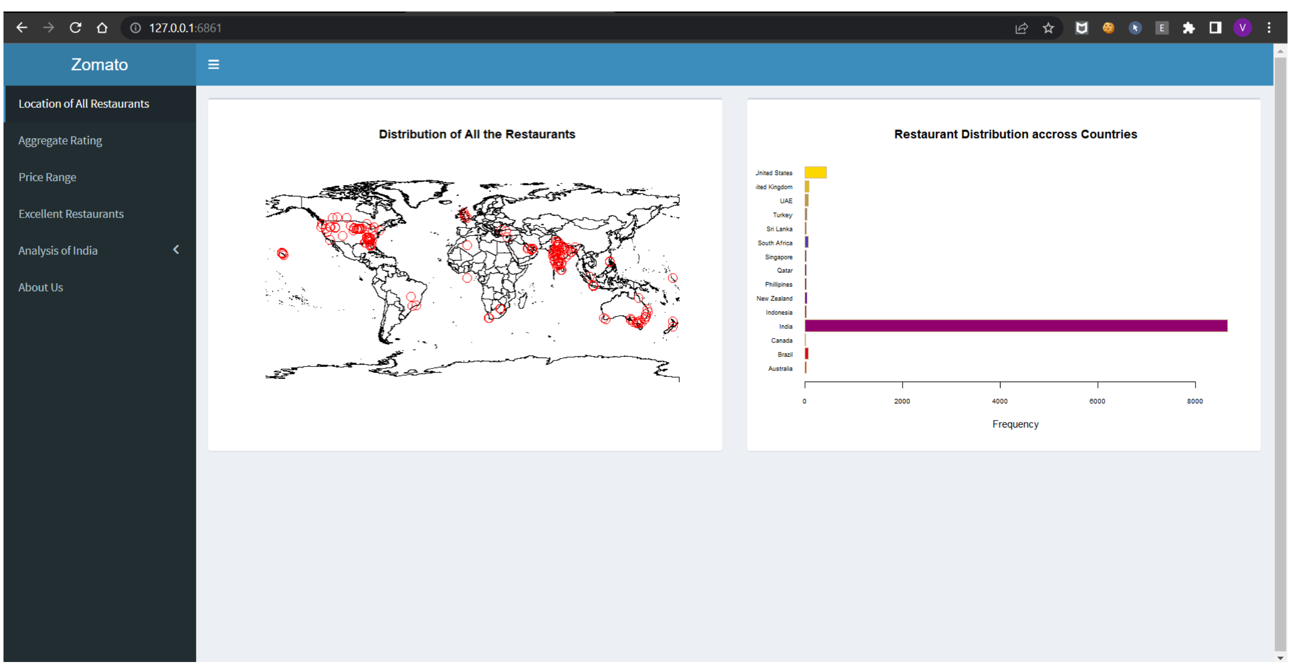
c. Top 10 Restaurants similar to Earl of Sandwich

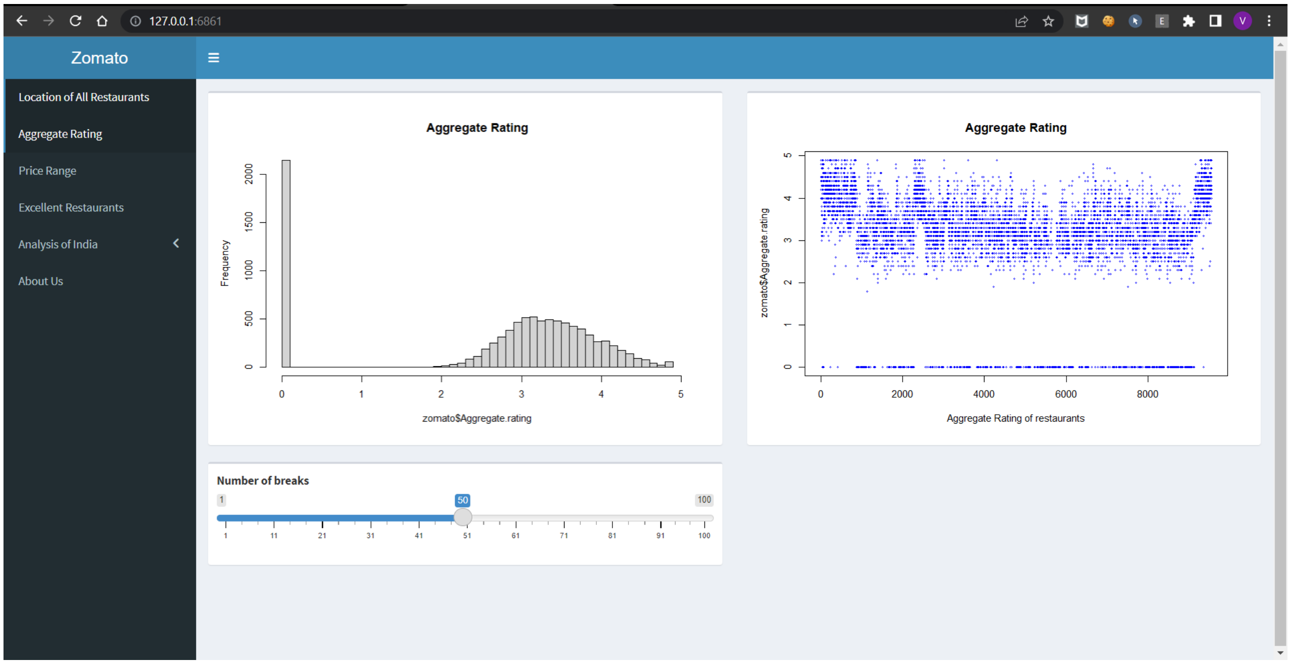
```
[ ] recommend('Earl of Sandwich')
```

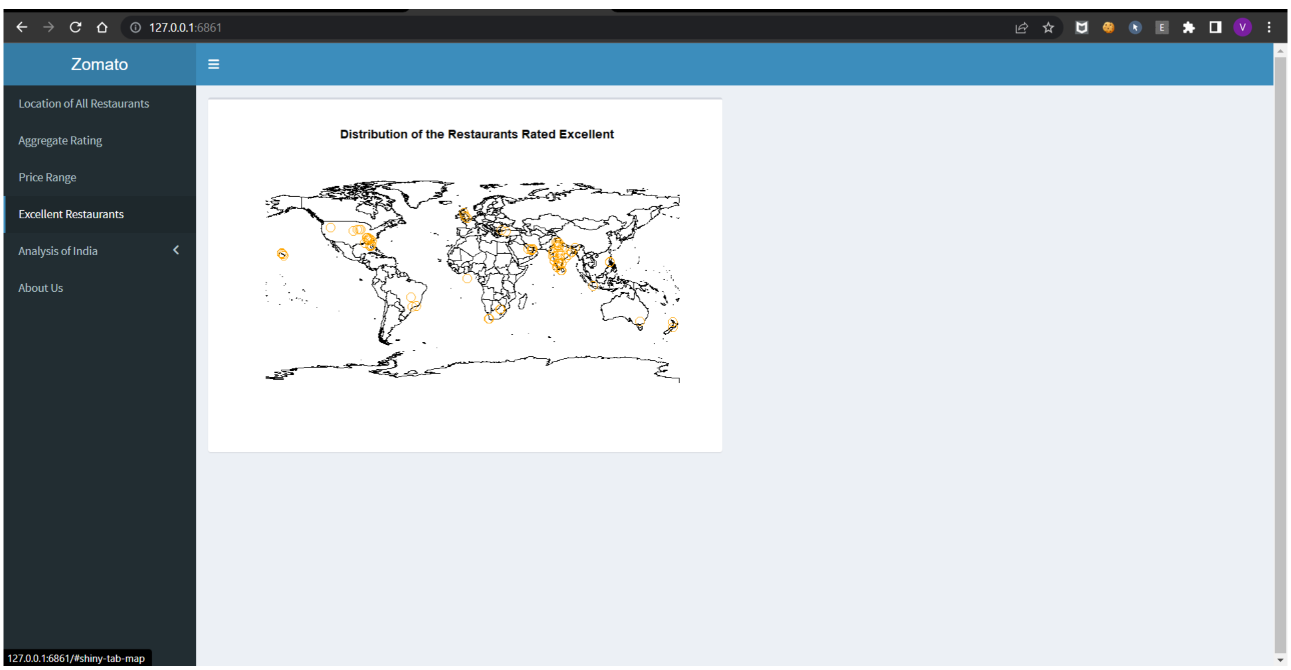
```
TOP 10 RESTAURANTS LIKE Earl of Sandwich WITH SIMILAR REVIEWS:
```

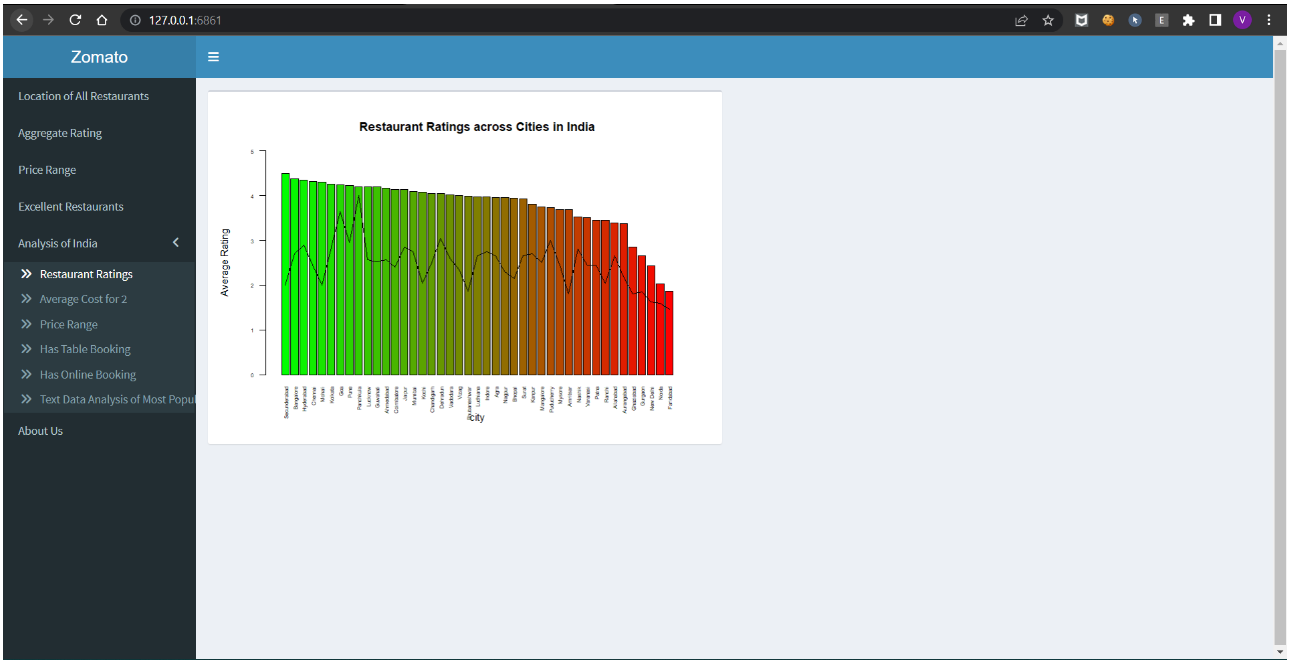
	Restaurant Name	Mean Rating	cost
9299	Milse	4.9	50.0
9303	Miann	4.9	25.0
304	Atlanta Highway Seafood Market	4.9	25.0
9296	Talaga Sampireun	4.9	200000.0
0	Le Petit Souffle	4.8	1100.0
2400	Spice Kraft	4.8	1200.0
9301	Depot Eatery and Oyster Bar	4.8	90.0
1839	Prankster	4.8	1500.0
352	Earl of Sandwich	4.7	35.0
9315	Eight - The Langham Hotel	4.7	190.0

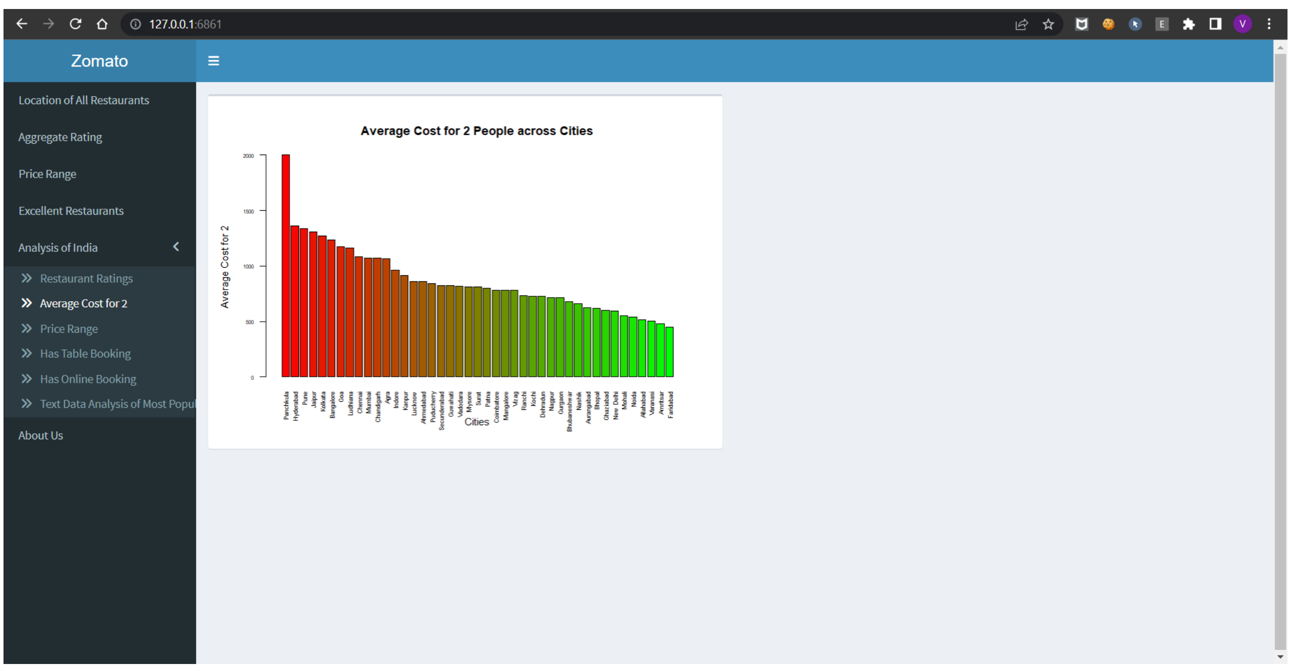
(ii) Dashboard:

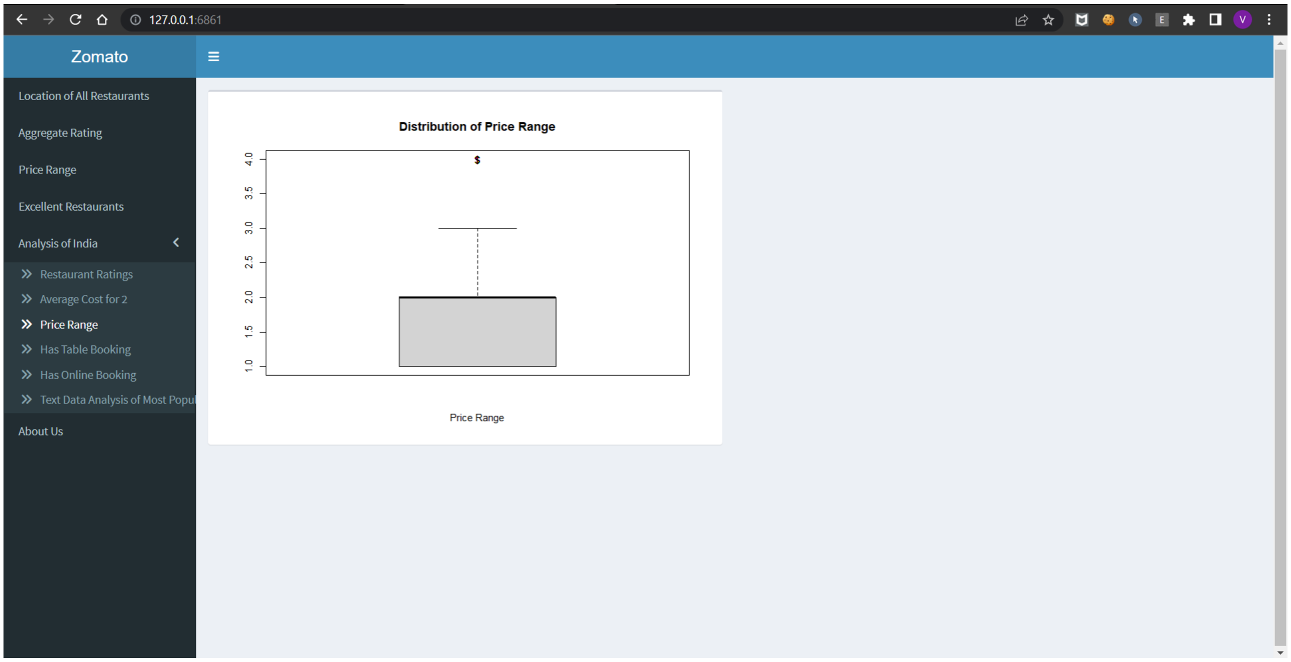



127.0.0.1:6861
Zomato
Location of All Restaurants
Aggregate Rating
Price Range
Excellent Restaurants
Analysis of India
About Us
Aggregate Rating
Frequency
zomato$Aggregate.rating
Aggregate Rating
zomato$Aggregate.rating
Aggregate Rating of restaurants
Number of breaks

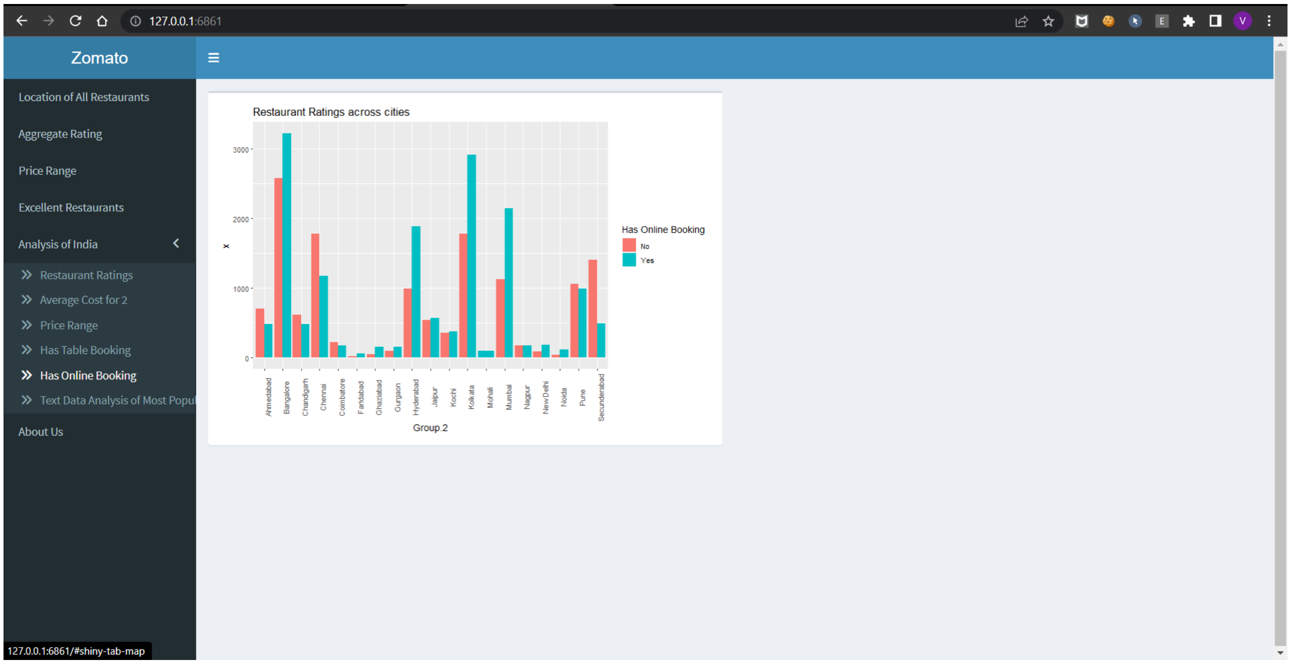

127.0.0.1:6861
Zomato
Location of All Restaurants
Aggregate Rating
Price Range
Excellent Restaurants
Analysis of India
About Us
Distribution of the Restaurants Rated Excellent
127.0.0.1:6861/#shiny-tab-map

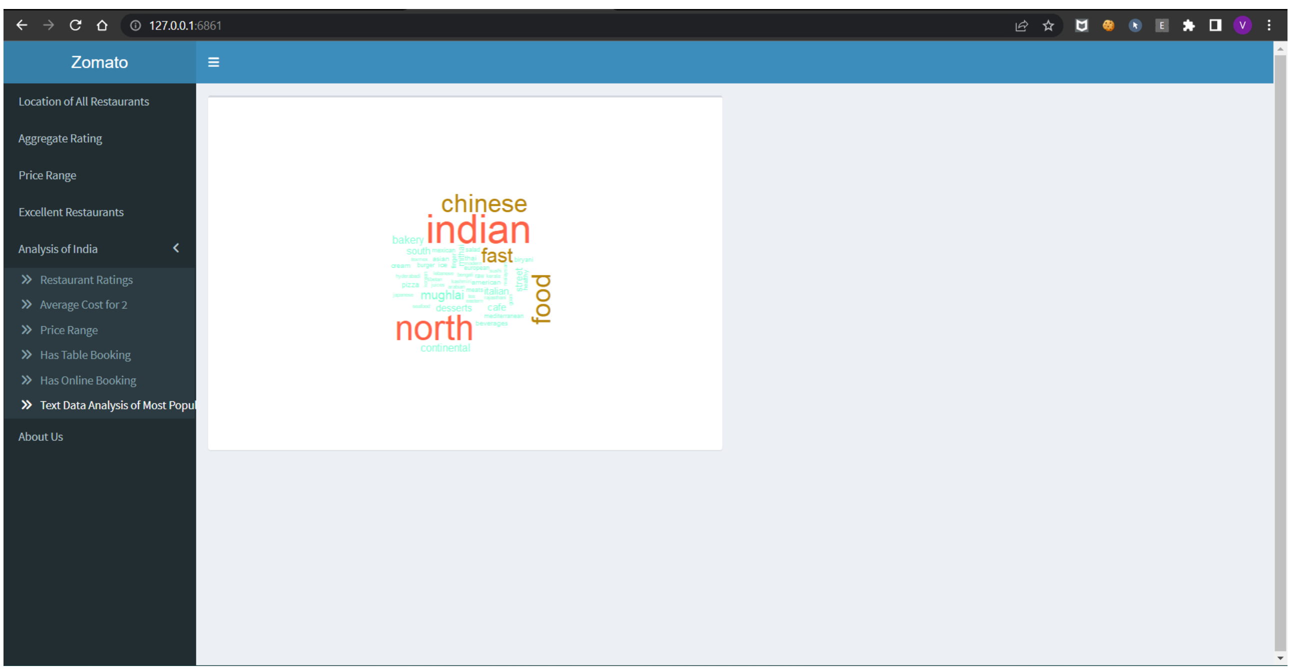

127.0.0.1:6861
Zoomato
Location of All Restaurants
Aggregate Rating
Price Range
Excellent Restaurants
Analysis of India
Restaurant Ratings
Average Cost for 2
Price Range
Has Table Booking
Has Online Booking
Text Data Analysis of Most Popul
About Us
Restaurant Ratings across Cities in India
Average Rating
City


127.0.0.1:6861
Zoomato
Location of All Restaurants
Aggregate Rating
Price Range
Excellent Restaurants
Analysis of India
Restaurant Ratings
Average Cost for 2
Price Range
Has Table Booking
Has Online Booking
Text Data Analysis of Most Popul
About Us
Average Cost for 2 People across Cities
Average Cost for 2
Cities


127.0.0.1:6861
Zomato
Location of All Restaurants
Aggregate Rating
Price Range
Excellent Restaurants
Analysis of India
Restaurant Ratings
Average Cost for 2
Price Range
Has Table Booking
Has Online Booking
Text Data Analysis of Most Popul
About Us
Distribution of Price Range
Price Range


127.0.0.1:6861
Zomato
Location of All Restaurants
Aggregate Rating
Price Range
Excellent Restaurants
Analysis of India
Restaurant Ratings
Average Cost for 2
Price Range
Has Table Booking
Has Online Booking
Text Data Analysis of Most Popul
About Us
Restaurant Ratings across cities
Has Table Booking
No
Yes
Bangalore
Chennai
Faridabad
Ghaziabad
Gurgaon
Hyderabad
Kolkata
Mumbai
New Delhi
Noida
Pune
Secunderabad
Group.2


127.0.0.1:6861
Zomato
Location of All Restaurants
Aggregate Rating
Price Range
Excellent Restaurants
Analysis of India
Restaurant Ratings
Average Cost for 2
Price Range
Has Table Booking
Has Online Booking
Text Data Analysis of Most Popul
About Us
Restaurant Ratings across cities
3000
2000
1000
0
Ahmedabad
Bangalore
Chandigarh
Chennai
Coimbatore
Faridabad
Ghaziabad
Gurgaon
Hyderabad
Jaipur
Kochi
Kolkata
Mohali
Mumbai
Nagpur
New Delhi
Noida
Pune
Secunderabad
Group 2
Has Online Booking
No
Yes
127.0.0.1:6861/#shiny-tab-map


127.0.0.1:6861
Zomato
Location of All Restaurants
Aggregate Rating
Price Range
Excellent Restaurants
Analysis of India
Restaurant Ratings
Average Cost for 2
Price Range
Has Table Booking
Has Online Booking
Text Data Analysis of Most Popul
About Us
chinese
indian
bakery
south
fast
food
pizza
mughlai
italian
desserts
cafe
north
continental

6. CONCLUSION:

Following our completion of this work, we can conclude that clustering cosine similarity is a commonly used machine learning algorithm. It can be used in various data sets and grouping types. As a user interface, we used a Shiny app and grouped the Zomato data set. We used various Plots to analyze our Zomato Data Set. We used a Geo map to define the restaurant's location. We found in our final result that we are getting a good representation of clusters and a better model that delivers better results. It will also help other startup companies to refer to the stats and the statistical spread of a global company like Zomato.

7. References:

1) *Improving Service Processes Based on Visualization of Human-Behavior and POS Data: A Case Study in a Japanese Restaurant* *Tomohiro Fukuhara, Ryuhei Tenmoku, Takashi Okuma, Ryoko Ueoka, Masanori Takehara & Takeshi Kurata,First Online: 15 April 2014*
2) Anil Bilgihan, Soobin Seo & Jihee Choi (2018) Identifying restaurant satisfiers and dissatisfiers: Suggestions from online reviews, Journal of Hospitality Marketing & Management, 27:5, 601-625, DOI: 10.1080/19368623.2018.1396275
3) R Development Core Team. (2008) "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing, Vienna
4) Predicting Restaurant Rating using Machine Learning and comparison of Regression Models, 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)
5) Microsoft Corporation. (2014) "Power BI: Interactive Data Visualization BI Tools." Microsoft Corporation, Albuquerque.
6) Visualization of Fast Food Restaurant Location using Geographical Information System ,D L Widaningrum1, I Surjandari1 and A M Arymurthy2 Published under licence by IOP Publishing Ltd
7) Birkin M, Clarke G, Clarke M and Culf R 2004 *Applied GIS and Spatial Analysis* ed J Stillwell and G Clarke (England: John Wiley & Sons, Ltd.) Using Spatial Models to Solve Difficult Retail Location Problem 35-5
8) Spatial Distribution and Characteristics of Restaurant Inspection Results in Toronto, Ontario, 2017–2018 DLT Ng, MA Harris, TH Oiamo, I Young - Food Protection Trends, 2020
9) Fernández de Córdova G, Fernández-Maldonado A M and del Pozo J M 2016 *Habitat International* 54 28-39
10) Restaurant review system with data visualization and text search using TF-IDF algorithm,S Abd Hamid - 2020 - ir.uitm.edu.my
11) Visualization of restaurant information on web maps,YJ Ah, KH Jeong, SJ Woo, LS Ho - The 5th International …, 2011
12) What affects the online ratings of **restaurant** consumers: a research perspective on text-mining big **data** analysis,J Liu, Y Yu, F Mehraliyev, S Hu, J Chen - International Journal of …, 2022

Appendix

CODE:

1. Dashboard

```
library(shiny)
library(shinydashboard)
library(rworldmap)
library(dplyr)
library(xlsx) #Read xlsx

zomato = read.csv("zomato.csv")

cn <- read.xlsx2("Country-Code.xlsx", sheetIndex = 1, header
= TRUE, stringsAsFactors = FALSE)


for (dfindex in 1:nrow(zomato)) {
  i <- 1
  while(zomato$Country.Code[dfindex]!= cn$Country.Code[i]){
    i = i+1 }
```

```
  zomato$country.Name[dfindex] <-cn$Country[i]

}




ui <- dashboardPage(dashboardHeader(title="Zomato"),

                    dashboardSidebar(

                      sidebarMenu(menuItem("Location of All
Restaurants",tabName = "map"), menuItem("Aggregate Rating",
tabName = "dash1"), menuItem("Price Range", tabName =
"dash2"), menuItem("Excellent Restaurants", tabName = "exe"),


                                  menuItem("Analysis of
India", menuSubItem("Restaurant Ratings", tabName =
"mapind"), menuSubItem("Average Cost for 2", tabName =
"avg_c"), menuSubItem("Price Range", tabName = "price_ind"),
menuSubItem("Has Table Booking", tabName = "tb"),
menuSubItem("Has Online Booking", tabName = "ob"),
menuSubItem("Text Data Analysis of Most Popular Cuisines",
tabName = "cuis")),


                                  menuItem("About Us",
tabName = "abt")


                      )),




                    dashboardBody(
```

```
                        tabItems(

                          tabItem(tabName = "map",
fluidRow(box(plotOutput("mapp")),
box(plotOutput("locbar")))),

                          tabItem(tabName = "dash1",
fluidRow(box(plotOutput("histogram")),
box(plotOutput("bar"))),
fluidRow(box(sliderInput("bins","Number of
breaks",1,100,50)))),

                          tabItem(tabName = "dash2",
fluidRow(box(plotOutput("line")), box(plotOutput("bars")),
box(plotOutput("boxp")))),

                          tabItem(tabName = "exe",
fluidRow(box(plotOutput("exew")))),

                          tabItem(tabName = "abt",
fluidRow(textOutput("text"))),

                          tabItem(tabName = "mapind",
fluidRow(box(plotOutput("maprevind")))),

                          tabItem(tabName = "invest",
fluidRow(box(plotOutput("inv")))),
```

```
                        tabItem(tabName = "tb",
fluidRow(box(plotOutput("table")))),

                        tabItem(tabName = "ob",
fluidRow(box(plotOutput("online")))),

                        tabItem(tabName = "avg_c",
fluidRow(box(plotOutput("averagecost")))),

                        tabItem(tabName = "price_ind",
fluidRow(box(plotOutput("priceindia")))),

                        tabItem(tabName =
"cuis",fluidRow(box(plotOutput("cuisine")))))))




server <- function(input, output) {


  library('rworldxtra')

  output$mapp <- renderPlot({newmap <- getMap(resolution =
"high")

  plot(newmap, xlim = range(zomato$Longitude),ylim =
range(zomato$Latitude), asp = 1, main="Distribution of All
the Restaurants")

  points(zomato$Longitude, zomato$Latitude, col = "red", cex
= 2)})
```

```
  output$locbar <- renderPlot({
    countryfreq<- table(zomato$country.Name)
    colFunc <- colorRampPalette(c("red","blue", "gold"))
    my.cols <- colFunc(length(countryfreq))
    barplot(countryfreq, col=my.cols,
            las = 1,
            horiz = TRUE,
            border = "tan",
            cex.axis = 0.7,
            cex.names = 0.6,
            xlab = "Frequency",
            xlim = c(0, max(countryfreq)),
            main="Restaurant Distribution accross Countries"
    )})


  output$histogram <-
renderPlot({hist(zomato$Aggregate.rating, main="Aggregate
Rating", breaks = input$bins)})


  output$bar <- renderPlot({plot(zomato$Aggregate.rating,
col="blue", main="Aggregate Rating", xlab="Aggregate Rating
of restaurants", cex=.4)})
```

```
  output$line <-
renderPlot({plot(zomato$Restaurant.ID,zomato$Price.range,col=
"red", main="Price Range", ylab="Price Range of
Restaurants")})

  output$boxp <- renderPlot({
    boxplot(zomato$Price.range,
            xlab = "Price Range",
            main = "Distribution of Price Range",
            pch="$")
    abline(h = median(zomato$Price.range), lty = 2, col =
"red")
  })



  output$bars <-
renderPlot({price_tab<-table(zomato$Price.range)
  barplot(price_tab, main = "Price Range of the Restaurants",
ylab="No of Restaurants", xlab="Price Range")})



  output$exew <- renderPlot({zomatoE <-
filter(zomato,zomato$Rating.text == "Excellent")
  newmap <- getMap(resolution = "low")
  plot(newmap, xlim = range(zomatoE$Longitude),ylim =
range(zomatoE$Latitude), asp = 1,
       main = "Distribution of the Restaurants Rated
Excellent")
```

```
  points(zomatoE$Longitude, zomatoE$Latitude, col = "orange",
cex = 2)})



  output$maprevind <- renderPlot({z2<-
filter(zomato,zomato$country.Name=="India")
  library(RColorBrewer)
  pal2 <- colorRampPalette(c("green","red"))
  colg <- pal2(43)

  rating<-aggregate(z2$Aggregate.rating, list(z2$City),mean)
  colnames(rating) <- c("city","rating")
  pr<- aggregate(z2$Price.range,list(z2$City),mean)
  colnames(pr)<- c("city","price.range")
  library(plyr)
  ratprt<- merge(pr,rating, by.x = "city", by.y = "city")

  ratprt <- ratprt[order(ratprt[,3], decreasing = TRUE),]
  x<- barplot(ratprt[,3],names.arg=ratprt[,1], las =2,
              ylab = "Average Rating",
              ylim=c(0,5),
              xlab = "city",
              cex.axis = 0.5,
              cex.names = 0.5,
              main = "Restaurant Ratings across Cities in
India",
```

```
              bty="n",
              col=colg,)
  lines(x=x, y=ratprt[,2])})

  output$inv <- renderPlot({
    ## Adding Population ###

    fnamep <- paste0("C:/Users/aksha/Documents/R
Dataset/population.csv")

    pop <- read.csv(fnamep, stringsAsFactors = FALSE, header
= TRUE)

    z2<- zomato[zomato$country.Name=="India",]
    z3 <-merge(pop,z2, by.x= "City", by.y = "City")

    ## Calculating top 5 cities to invest using weights

z3<-z3[,c("City","Population","Votes","Average.Cost.for.two",
"Aggregate.rating")]

z3[,c("Population","Votes","Average.Cost.for.two","Aggregate.
rating")]<-
scale(z3[,c("Population","Votes","Average.Cost.for.two","Aggr
egate.rating")])
    z3$scores<-
round(z3$Population*0.2+z3$Votes*0.2+z3$Average.Cost.for.two*
0.4+z3$Aggregate.rating*0.4)
```

```
    scores<- tapply(z3$scores, list(z3$City), mean)
    scores2<- sort(scores, decreasing = TRUE)

    library(RColorBrewer)
    pal2 <- colorRampPalette(c("green","red"))
    colg <- pal2(43)

    barplot(scores2[1:5],
            las=2,
            ylab = "Scores",
            xlab = "city",
            ylim = c(0,max(scores2)),
            cex.axis = 0.5,
            cex.names = 0.7,
            main = "Top 5 Cities to Invest",
            bty="n",
            col = colg)})



  output$table <- renderPlot({

    z2<- zomato[zomato$country.Name=="India",]
    d1 <- aggregate(z2$Votes, list(z2$Has.Table.booking,
z2$City), mean)
    names1 <- d1$Group.2[d1$Group.1=="Yes"]
```

```
    z<- d1[d1$Group.2 %in% names1,]

    library(ggplot2)

    ggplot(z, aes(fill=Group.1, y=x, x=Group.2))  +

      geom_bar(position="dodge", stat="identity") +

      theme(axis.text.x = element_text(angle = 90)) +

      ggtitle("Restaurant Ratings across cities") +

      scale_fill_discrete(name = "Has Table Booking")})


  output$online <- renderPlot({


    z2<- zomato[zomato$country.Name=="India",]

    d2 <- aggregate(z2$Votes, list(z2$Has.Online.delivery,
z2$City), mean)

    names2 <- d2$Group.2[d2$Group.1=="Yes"]

    z<- d2[d2$Group.2 %in% names2,]

    library(ggplot2)

    ggplot(z, aes(fill=Group.1, y=x, x=Group.2))  +

      geom_bar(position="dodge", stat="identity") +

      theme(axis.text.x = element_text(angle = 90)) +

      ggtitle("Restaurant Ratings across cities") +

      scale_fill_discrete(name = "Has Online Booking")})


  output$averagecost <- renderPlot({


    z2<- zomato[zomato$country.Name=="India",]
```

```
    e <- tapply(z2$Average.Cost.for.two, list(z2$City), mean)
    cost<- sort(e,decreasing = TRUE)
    library(RColorBrewer)
    pal <- colorRampPalette(c("red", "green"))
    colred <- pal(43)

    barplot(cost,
            las=2,
            ylab = "Average Cost for 2",
            xlab = "Cities",
            cex.axis = 0.5,
            cex.names = 0.6,
            main = "Average Cost for 2 People across Cities",
            bty="n",
            col = colred)})

  output$priceindia <- renderPlot({

    z2<- zomato[zomato$country.Name=="India",]
    boxplot(z2$Price.range,
            xlab = "Price Range",
            main = "Distribution of Price Range",
            pch="$")})

  output$cuisine <- renderPlot({
```

```r
library('tm')
library('textstem')
library('qdap')
z2<- zomato[zomato$country.Name=="India",]
cF1<- Corpus(VectorSource(z2$Cuisines))
cF1<- tm_map(cF1,tolower)
cF1 <- tm_map(cF1, removePunctuation)
cF1 <- tm_map(cF1, removeWords, stopwords("english"))
cF1 <- tm_map(cF1, removeNumbers)
stem_word1 <- lemmatize_words(cF1, dictionary =
lexicon::hash_lemmas)
freq_terms(stem_word1, 20)
tdm<- TermDocumentMatrix(stem_word1)
tdmatrix<-as.matrix(tdm)
colT<- apply(tdmatrix, 2, sum)
tdmclean<- tdmatrix[,colT>0]
tdfreq<- rowSums(tdmclean)
tdfreq<- sort(tdfreq,decreasing = T)
tdfreq[1:20]
wordfreq<- data.frame(term=names(tdfreq),num=tdfreq)
library('wordcloud')
wordcloud(wordfreq$term, wordfreq$num, max.words=50,
colors=c("aquamarine","darkgoldenrod","tomato"))})
```

```
  output$text<-renderText("\t\t\t\n\n\n Data visualisation
Project by group : --> \n\t  Viraj Gupta : 19BCE2151 \n\t  ,
Aditya Chaudhary : 19BCE0544 \n\t  ,    Arpit Sharma :
19BDS0027")


}



shinyApp(ui, server)
```

2. Restaurant Recommendation System

```
import numpy as np

import pandas as pd

import os

from google.colab import drive

drive.mount('/content/drive',force_remount=True)


import os
```

```
os.chdir("/content/drive/MyDrive/Colab Notebooks/dv")


data=pd.read_csv("zomato.csv",encoding = "ISO-8859-1")

print(data)


data.duplicated().sum()

data.drop_duplicates(inplace=True)


data.dropna(how='any',inplace=True)


data.rename(columns={'Average Cost for two':'cost'},inplace=True)


data['cost']=data['cost'].astype(str)

data['cost']=data['cost'].apply(lambda x: x.replace(',',''))

data['cost']=data['cost'].astype(float)


restaurants = list(data['Restaurant Name'].unique())

data['Mean Rating'] = 0

for i in range(len(restaurants)):

    data['Mean Rating'][data['Restaurant Name'] == restaurants[i]] = data['Aggregate 
rating'][data['Restaurant Name'] == restaurants[i]].mean()


from sklearn.feature_extraction.text import CountVectorizer

from sklearn.feature_extraction.text import TfidfVectorizer
```

```
tfidf = TfidfVectorizer(analyzer='word', ngram_range=(1, 2), min_df=0, stop_words='english')
tfidf_matrix = tfidf.fit_transform(data['Rating text'])

from sklearn.metrics.pairwise import linear_kernel
cosine_similarities = linear_kernel(tfidf_matrix, tfidf_matrix)

indices = pd.Series(data['Restaurant Name'])

def recommend(name, cosine_similarities = cosine_similarities):

    recommend_restaurant = []

    idx = indices[indices == name].index[0]

    score_series = pd.Series(cosine_similarities[idx]).sort_values(ascending=False)

    top30_indexes = list(score_series.iloc[0:31].index)

    for each in top30_indexes:
        recommend_restaurant.append(list(data.index)[each])

    df_new = pd.DataFrame(columns=['Restaurant Name','Mean Rating', 'cost'])
```

```
    for each in recommend_restaurant:

        df_new = df_new.append(pd.DataFrame(data[['Restaurant Name','Mean
Rating','cost']][data.index == each].sample()))


    df_new = df_new.drop_duplicates(subset=['Restaurant Name','Mean Rating', 'cost'], keep=False)

    df_new = df_new.sort_values(by='Mean Rating', ascending=False).head(10)


    print('TOP %s RESTAURANTS LIKE %s WITH SIMILAR REVIEWS: ' % (str(len(df_new)),
name)

    return df_new


recommend('Vikings')

recommend('Ooma')

recommend('Earl of Sandwich')
```

*************THANK YOU*****************