



Cal State LA University

Zomato Restaurants Dataset analysis

By: Bitasadat Faraji Tabrizi & Yangyang Jia

A. Dataset URL's:

<https://www.kaggle.com/shrutihehta/zomato-restaurants-data/data>

This URL's contain the entire data regarding the Zomato, It is an online website which provides restaurant search and discovery service. It provide its customer a platform to evaluate choices for great place to eat We are curious to study and analyse this huge dataset to understand the buyers' point of view, preferences and their ability to spend. The dataset has total 21 columns and contains the information like, Restaurant ID, Restaurant Name, Country Code, City, Address, Locality, Locality Verbose, Longitude, Latitude, Cuisines, Average Cost for two, Currency, Has Table booking, Has Online delivery, Is delivering now, Switch to order menu, Price range, Aggregate rating, Rating colour, Rating text, Votes. The study will give us the in detail idea about Market and Consumer Analysis – Zomato. It is important to study this Zomato market penetrating into sectors which introducing different [Cuisines, country code, locality longitude latitude, with various currency] Now to study the impact on about the market which the TOP 10 RESTAURANT'S, the average cost of two for the aggregate rating, customer vote, their price range and currency, customer votes and aggregate rating.

Before knowing the reason for the choosing this topic, lets know about: - Zomato. It is an online restaurant guide; it is one of the fastest ways to search places to eat.

Searching with autonomous reviews and rating where “FoodieBay” it’s the initial name of Zomato started in July 2008 by IIT graduates Mr. Deepinder Goyal (Founder) and Pankaj Chaddah (Co-Founder). Zomato has its headquarters in Delhi. “Info edge” it’s the parent company of Zomato where it was categorised in “mobile applications” and “food and restaurant guide” is its sector. It was renamed as Zomato in November 2010. It expand to Pune and Bangalore and then to Chennai,

Ahmadabad and Hyderabad in 2011. In a very short span of time, Zomato went global and started providing its service in Dubai in September 2012. As the company is going more and more global, it rebranded its logo that so that it could transcend cultures, languages and geographic boundaries.

Mission statement of the Zomato states that “Our mission is to ensure nobody has a bad meal”. Vision statement of the Zomato states that “Zomato is used by millions every day to decide where to eat in over 10,000 cities across 23 countries. In a few years, we should be able help point you to a great place to eat no matter what part of the world you're in”.

To differentiate themselves from their competitors. Zomato concentrated on adding approx. 18,000 new places to eat from. Along with they also decorated many special features such as pointed to particular dishes or opening times.

To be the largest resource in food supply market, Zomato bought urban spoon, a leading restaurant service providing portal for \$52 million to enter US, Canada and Australia to leverage local insights and experiences and to expand their business in overseas seeing the future goal and objectives.

Finally, the reason in choosing this topic is to have a better insight on how the Zomato's service will affect on the customer in different parts of the world. Also, how customer review will help Zomato improve their services. The analysis will help categorizing restaurants, their locations, quality, etc. and will help to understand user requirement and increase customer satisfaction. As an objective we also need to know about the customers likes and dislikes with respect to places and locality to improve the quality of services accordingly.

Justifying Columns to Analyze:

Restaurant ID	Identification Number	Numeric	
Restaurant Name	Name Of the Restaurant	String	
Country Code	Country Code of the Restaurant	Numeric	
City	City Name of the Restaurant	String	
Address	Address Of the Restaurant	String	
Locality	Shot Address Of the Restaurant	String	✓
Locality Verbose	Long Address of the Restaurant	String	
Longitude	Longitude	Numeric	
Latitude	Latitude	Numeric	
Cuisines	Types Of Cuisines Served	String	
Average Cost for two	Average Cost if two people visit the Restaurant	Numeric	
Currency	Type of Currency paid in the Restaurant	String	

Currency	Type of Currency paid in the Restaurant	String
Has Table booking	Can we book tables in Restaurant? Yes/No	String
Has Online delivery	Can we have online delivery ? Yes/No	String
Is delivering now	Is the Restaurant delivering food now? Yes/No	String
Switch to order menu	Switch to order menu ? Yes/ No	String
Price range	Categorized price between 1 -4	Numeric
Aggregate rating	Categorizing ratings between 1-5	Numeric
Rating color	Different colors representing Customer Rating	String
Rating text	Different Rating like Excellent, Very Good ,Good, Avg., Poor, Not Rated	String
Votes	No.Of Votes received by restaurant from customers.	Numeric

B. Data Cleaning:

1. Deleting/Removing irrelevant columns

The dataset has two similar columns with the column names as “Locality” and “Locality.Verbose”. We removed one unwanted column (“Locality” which had less detailed data) and cleaned the data.

Before:

The screenshot shows the RStudio interface with the following details:

- Environment Pane:** Shows a data frame 'a' with 9551 observations and 21 variables.
- Data Viewer:** Displays the first 12 columns of the dataset: Address, Locality, Locality.Verbose, Longitude, and Latitude. The data is shown in a table format with rows representing individual observations.
- Console:** Contains the following R code:

```
> setwd("C:/Users/Dear User/Desktop")
> a<-read.csv("zomato.csv",header=T,sep=",")
> View(a)
> |
```

After:

The screenshot shows the RStudio interface after removing the 'Locality' column. The details are as follows:

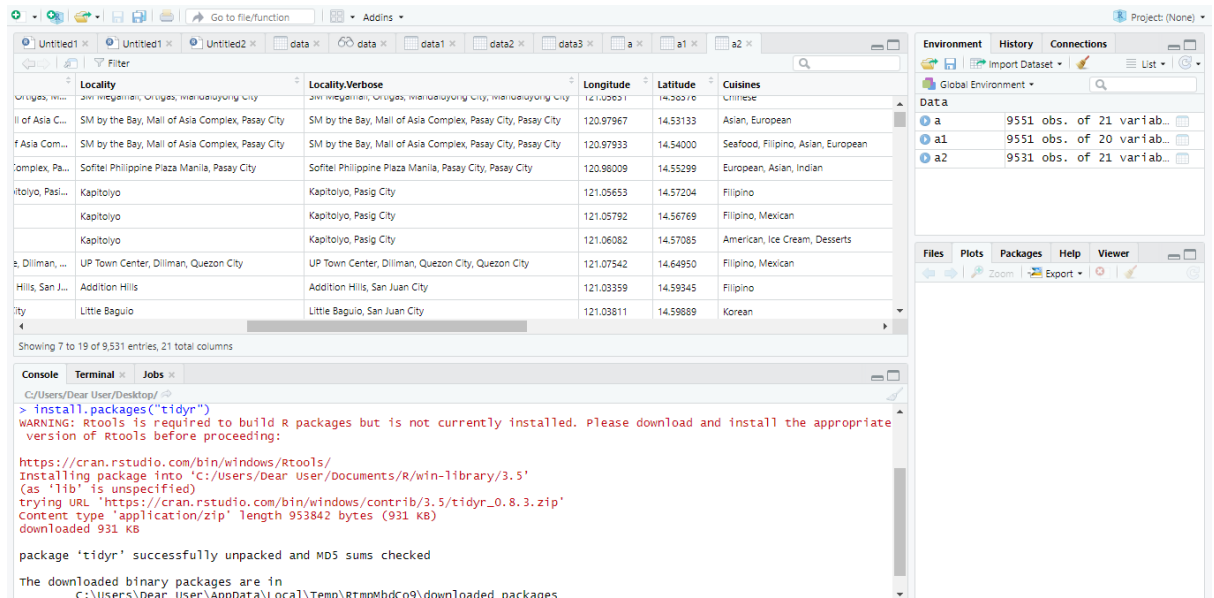
- Environment Pane:** Shows a data frame 'a1' with 9551 observations and 20 variables.
- Data Viewer:** Displays the first 12 columns of the dataset: Address, Locality.Verbose, Longitude, Latitude, and Cuisines. The 'Locality' column has been removed.
- Console:** Contains the following R code:

```
> setwd("C:/Users/Dear User/Desktop")
> a<-read.csv("zomato.csv",header=T,sep=",")
> View(a)
> a1<-a[,-6]
> View(a1)
> |
```

2. Combining Columns

Combining latitude and longitude to a single column named location using unite function and installing “tidyr” package.

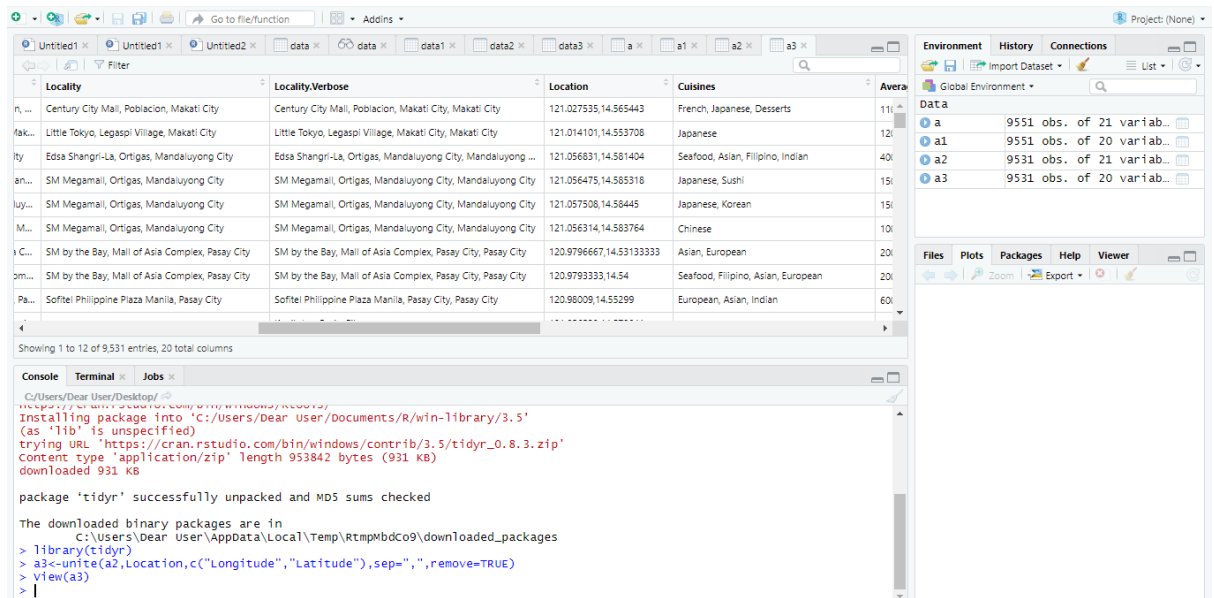
Before:



The screenshot shows the RStudio interface before the transformation. The data frame has columns: Locality, Locality.Verbose, Longitude, Latitude, and Cuisines. The console shows the installation of the 'tidyr' package.

```
install.packages("tidyr")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:
https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/Dear User/Documents/R/win-library/3.5'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.5/tidyr_0.8.3.zip'
Content type 'application/zip' length 953842 bytes (931 KB)
downloaded 931 KB
package 'tidyr' successfully unpacked and MD5 sums checked
The downloaded binary packages are in
C:/Users/Dear User/AppData/Local/Temp/RtmpMbdCo9/downloaded_packages
```

After:



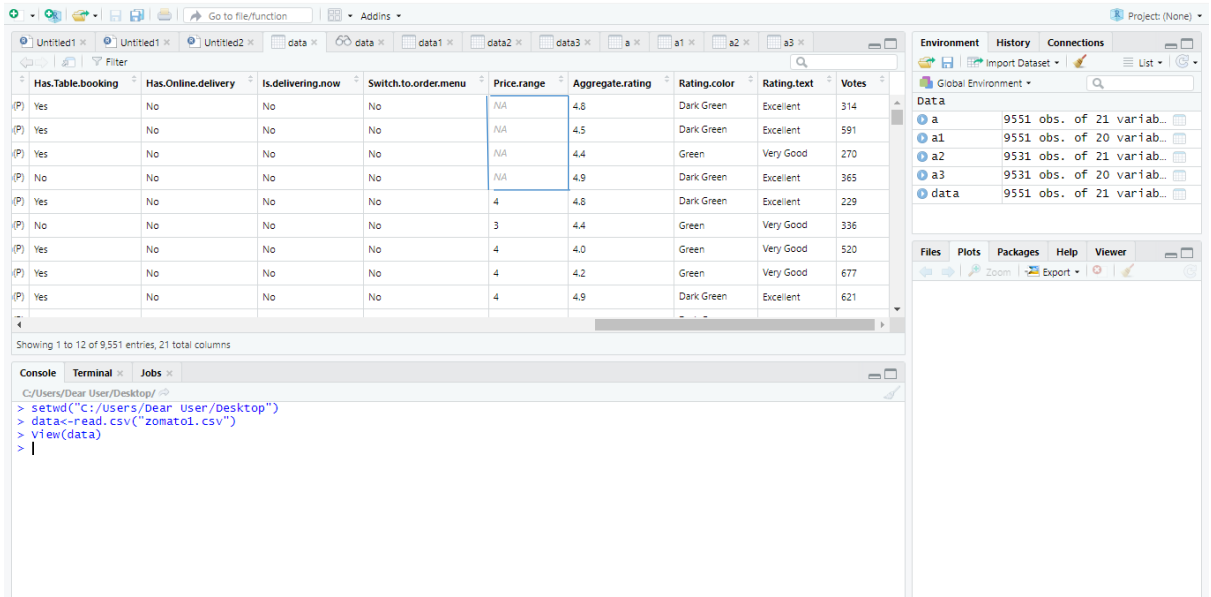
The screenshot shows the RStudio interface after the transformation. The data frame now has columns: Locality, Locality.Verbose, Location, Cuisines, and Average. The console shows the execution of the unite function.

```
library(tidyr)
a3<-unite(a2,Location,c("Longitude","Latitude"),sep=" ",remove=TRUE)
view(a3)
```

3. Removing Null Values :

Removing Null Values and replacing them with 0 in price range column.

Before:

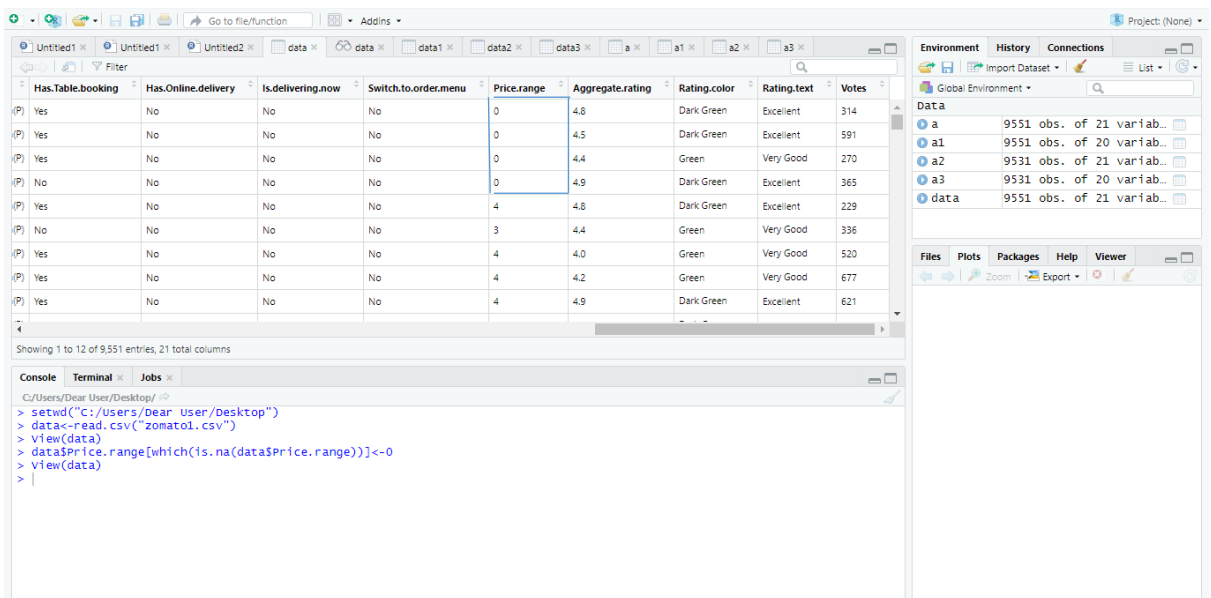


The screenshot shows the RStudio interface with a data frame loaded from 'zomato1.csv'. The 'Price.range' column contains several NA values. The console shows the following commands:

```
C:/Users/Dear User/Desktop/ >
> setwd("C:/Users/Dear User/Desktop")
> data<-read.csv("zomato1.csv")
> view(data)
> |
```

	Has.Table.booking	Has.Online.delivery	Is.delivering.now	Switch.to.order.menu	Price.range	Aggregate.rating	Rating.color	Rating.text	Votes
(P)	Yes	No	No	No	NA	4.8	Dark Green	Excellent	314
(P)	Yes	No	No	No	NA	4.5	Dark Green	Excellent	591
(P)	Yes	No	No	No	NA	4.4	Green	Very Good	270
(P)	No	No	No	No	NA	4.9	Dark Green	Excellent	365
(P)	Yes	No	No	No	4	4.8	Dark Green	Excellent	229
(P)	No	No	No	No	3	4.4	Green	Very Good	336
(P)	Yes	No	No	No	4	4.0	Green	Very Good	520
(P)	Yes	No	No	No	4	4.2	Green	Very Good	677
(P)	Yes	No	No	No	4	4.9	Dark Green	Excellent	621

After:



The screenshot shows the RStudio interface after running the following commands in the console:

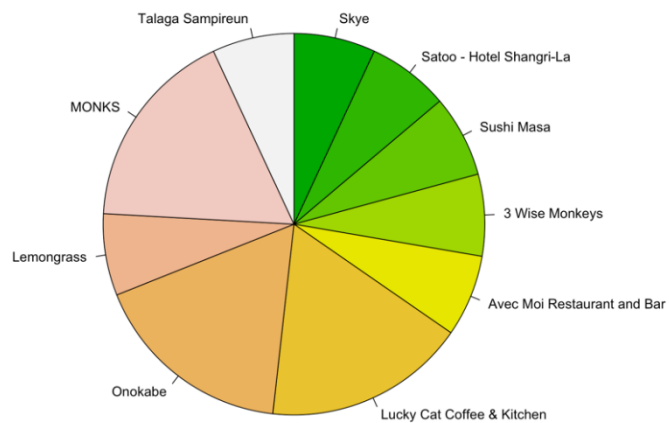
```
C:/Users/Dear User/Desktop/ >
> setwd("C:/Users/Dear User/Desktop")
> data<-read.csv("zomato1.csv")
> view(data)
> data$Price.range[which(!is.na(data$Price.range))]<-0
> view(data)
> |
```

	Has.Table.booking	Has.Online.delivery	Is.delivering.now	Switch.to.order.menu	Price.range	Aggregate.rating	Rating.color	Rating.text	Votes
(P)	Yes	No	No	No	0	4.8	Dark Green	Excellent	314
(P)	Yes	No	No	No	0	4.5	Dark Green	Excellent	591
(P)	Yes	No	No	No	0	4.4	Green	Very Good	270
(P)	No	No	No	No	0	4.9	Dark Green	Excellent	365
(P)	Yes	No	No	No	4	4.8	Dark Green	Excellent	229
(P)	No	No	No	No	3	4.4	Green	Very Good	336
(P)	Yes	No	No	No	4	4.0	Green	Very Good	520
(P)	Yes	No	No	No	4	4.2	Green	Very Good	677
(P)	Yes	No	No	No	4	4.9	Dark Green	Excellent	621

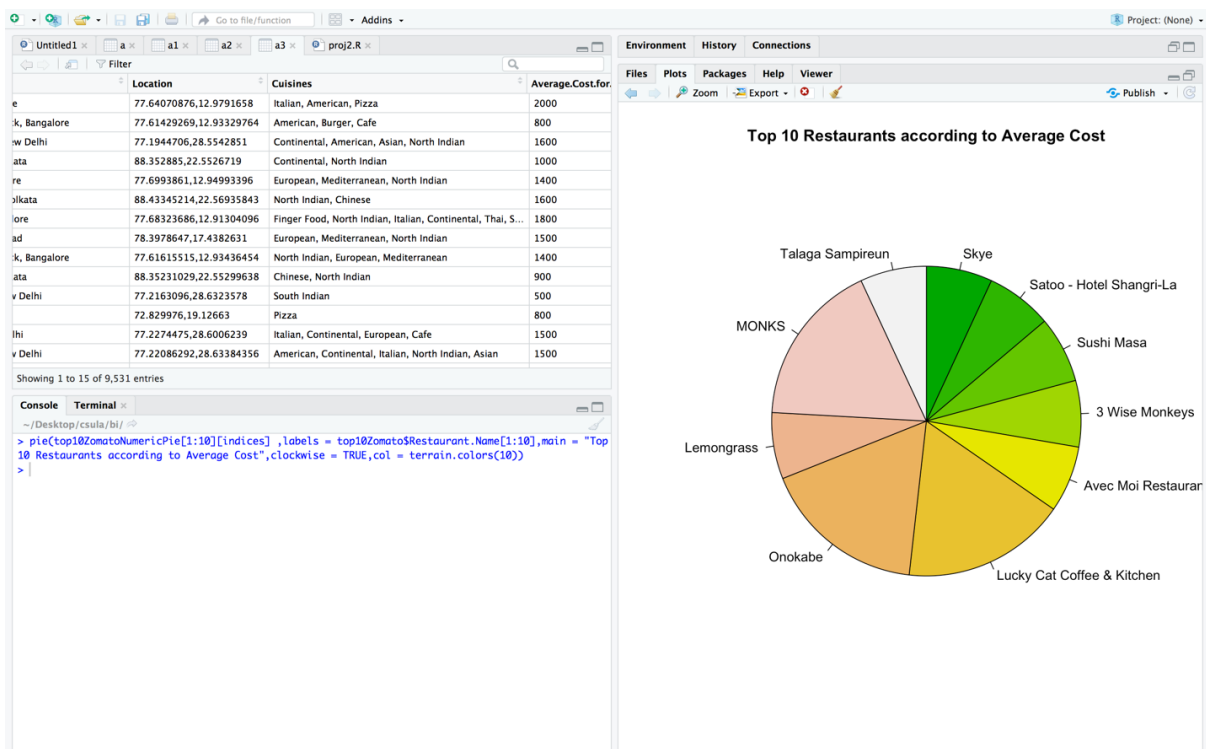
C. Data Visualizations:

1. Finding the Average cost of top 10 restaurants?

Top 10 Restaurants according to Average Cost

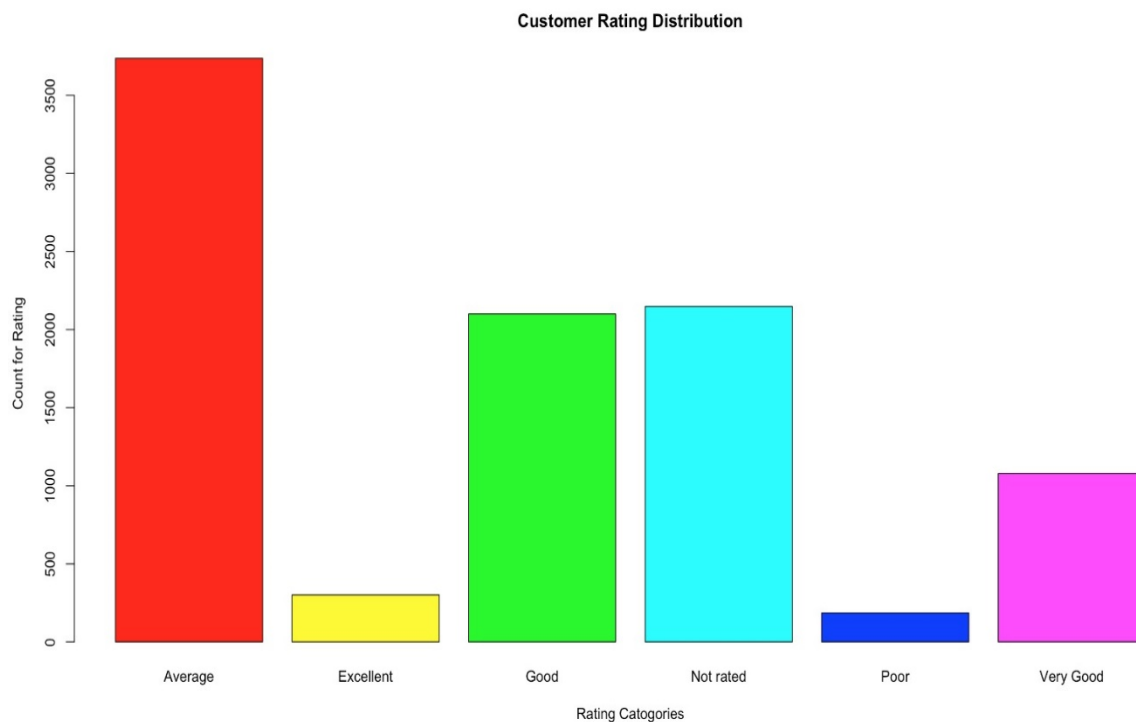


(Highlights from RScript – plotrix, plyr, pie chart)

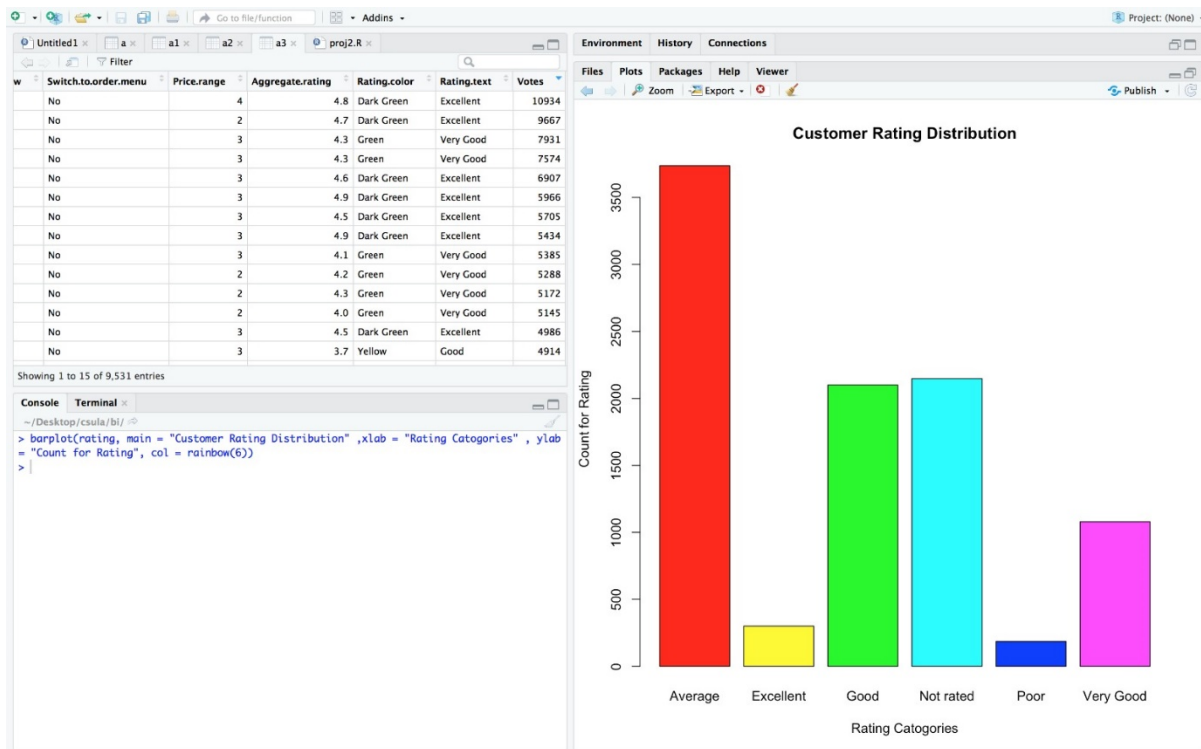


With the use of **R-Script** we have analysed the facts about the restaurant wise average cost proportion in above pie chart. It shows the average cost for two people of top ten restaurants. We can see the names of restaurants for which the average cost of two persons for top ten restaurants. Maximum is Lucky cat coffee & kitchen followed by Monks, Onokabe, Lemongrass, 3 wise monkeys, Sushi masa, Skye, Avec moi restaurant, Satoo – hotel Shangri la. By looking at this statistics, we can analyze the top 10 restaurants according to average cost.

2. Different types of Customer ratings of all the restaurants?



(Highlights from RScript –plotrix, plyr, Bar chart)

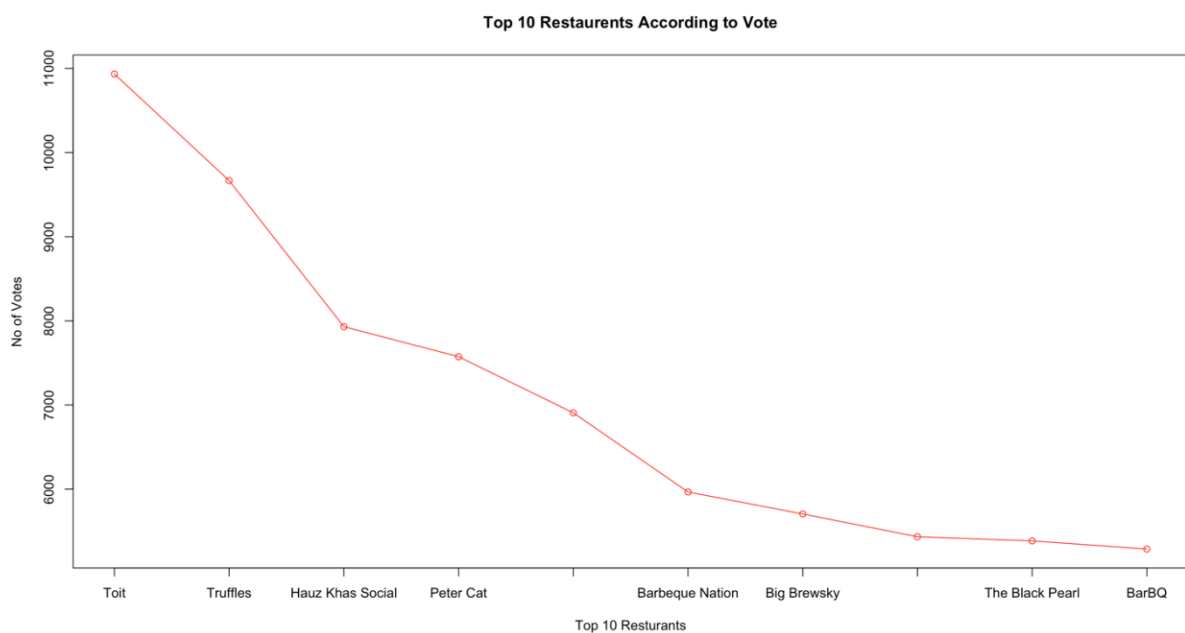


In the above Bar chart, we have tried to analyse the customer rating of all the restaurants. We can analyse that, here the rating categorised in various dimensions that is Average, Excellent, Good, Not Rated, Poor, Very Good. Maximum restaurants are rated “Average” by the customers with the count for rating greater than 3500 followed by Not Rated, Good, Very Good, Excellent, Poor has been rated by customers.

Rating Categories	Aggregate rating	Rating Colour
Excellent	= 4.5>	Dark Green
Very good	4 to 4.5	Green
Good	3.5 to 3.9	Yellow

Average	2.5 to 3.4	Orange
Poor	1 to 2.5	Red
Not Rated	Nil	White

3. Finding how much Customers voted of top ten restaurants?

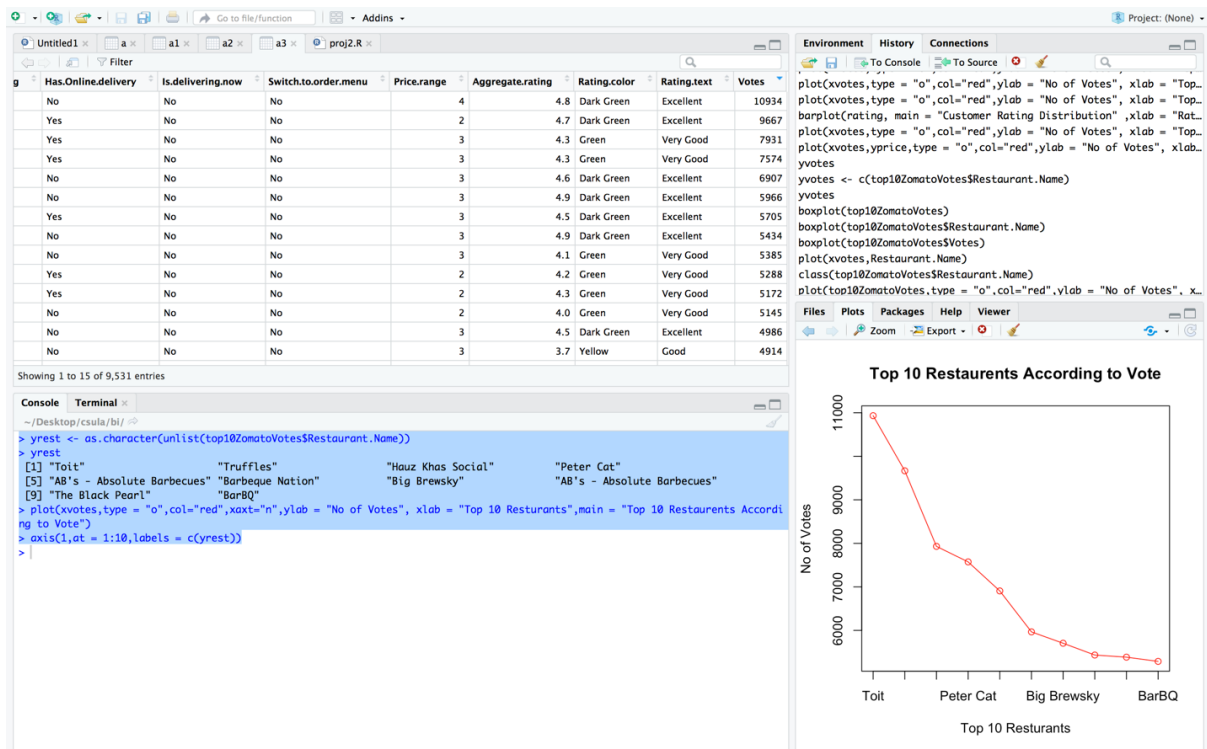


(Highlights from RScript –plotrix, plyr, Line chart)

```

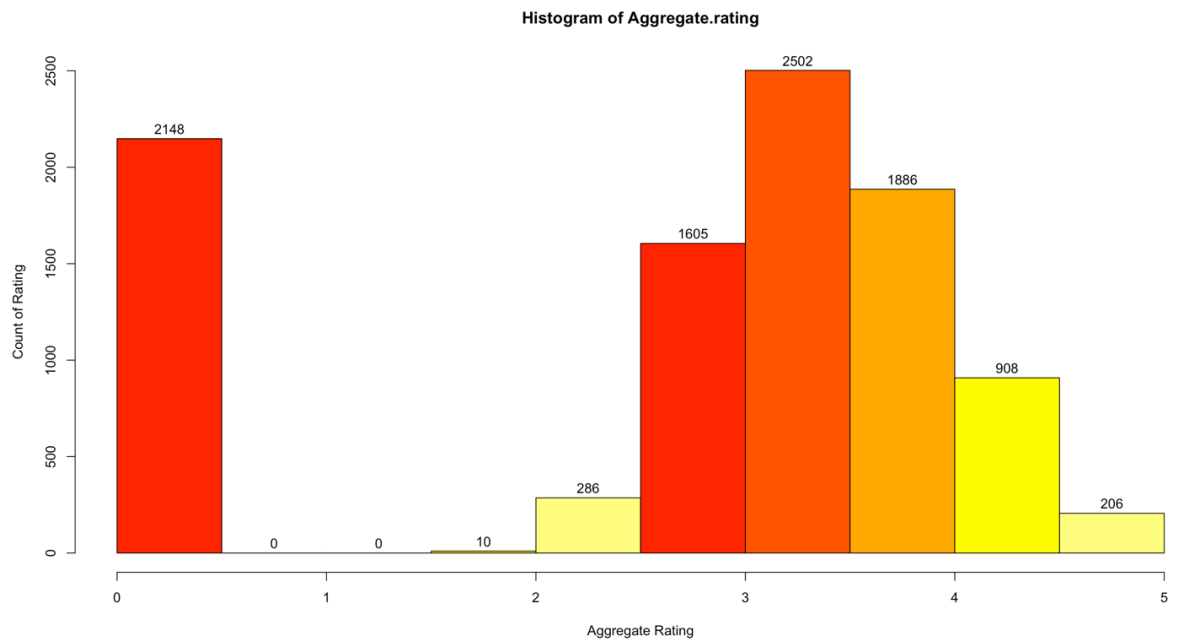
Console Terminal x
~/Desktop/csula/bi/
> zomatoRestNameAndAverage <- data.frame(Restaurant.ID, Restaurant.Name, Average.Cost.for.two)
> zomatoRestNameAndAverage$Restaurant.ID <- a$Restaurant.ID
> zomatoRestNameAndAverage$Restaurant.Name <- a$Restaurant.Name
> zomatoRestNameAndAverage$Average.Cost.for.two <- a$Average.Cost.for.two
> top10Zomato <- zomatoRestNameAndAverage[order(-zomatoRestNameAndAverage$Average.Cost.for.two) , ][1:10,]
> print(top10Zomato)
  Restaurant.ID Restaurant.Name Average.Cost.for.two
9284      7402935          Skye          800000
9285      7410290 Satoo - Hotel Shangri-La          800000
9288      7420899          Sushi Masa          500000
9290      7421967      3 Wise Monkeys          450000
9295      7422489 Avec Moi Restaurant and Bar          350000
9289      18352452 Lucky Cat Coffee & Kitchen          300000
9296      18386856          Onokabe          300000
9279      7423482          Lemongrass          250000
9286      18391256          MONKS          250000
9280      7422633      Talaga Sampireun          200000

```



In the above line chart, we have tried to analyse the customer votes of all the restaurants. We can analyse that, here the Maximum vote 10934; Minimum vote 0. The Customer votes of top 10 restaurants; where, for TOIT there are votes **10934**. By looking at this statistics, we can analysis the top 10 restaurants according to customer votes.

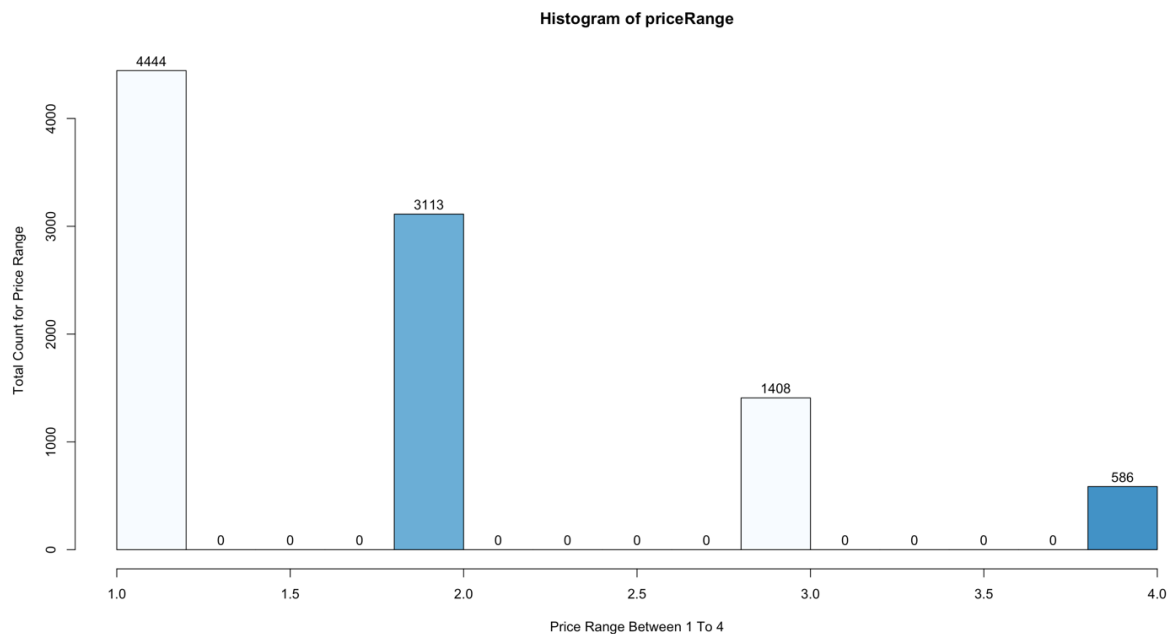
4. Finding Overall Customer ratings for all the restaurants?



In the above histogram, we have tried to analyze the overall customer ratings of all the restaurants where horizontal axis indicates aggregate ratings and vertical axis indicates count of rating. Rating is ranged between 0 to 5.

By analysis of the above data, we found that the highest count of rating is 2502 which is indicated between 3-4 in restaurants aggregate rating

5. Finding Customer preferable price range on all the restaurants?



In the above graph, we have tried to compare between price range and the count of price range. So that we can analyze the price range which has been focused by the customer.

Horizontal axis indicates the price range which ranges from 0 to 5000 and vertical axis indicates the total count of price which ranges from 1.0 to 4.0. (1= Lowest, 4= Highest)

From the above graph, we found that 4444 restaurants falls between 1 – 1.5 ratings and 586 restaurants falls between 3.5 – 4.

D. Statistical Summary and Functions

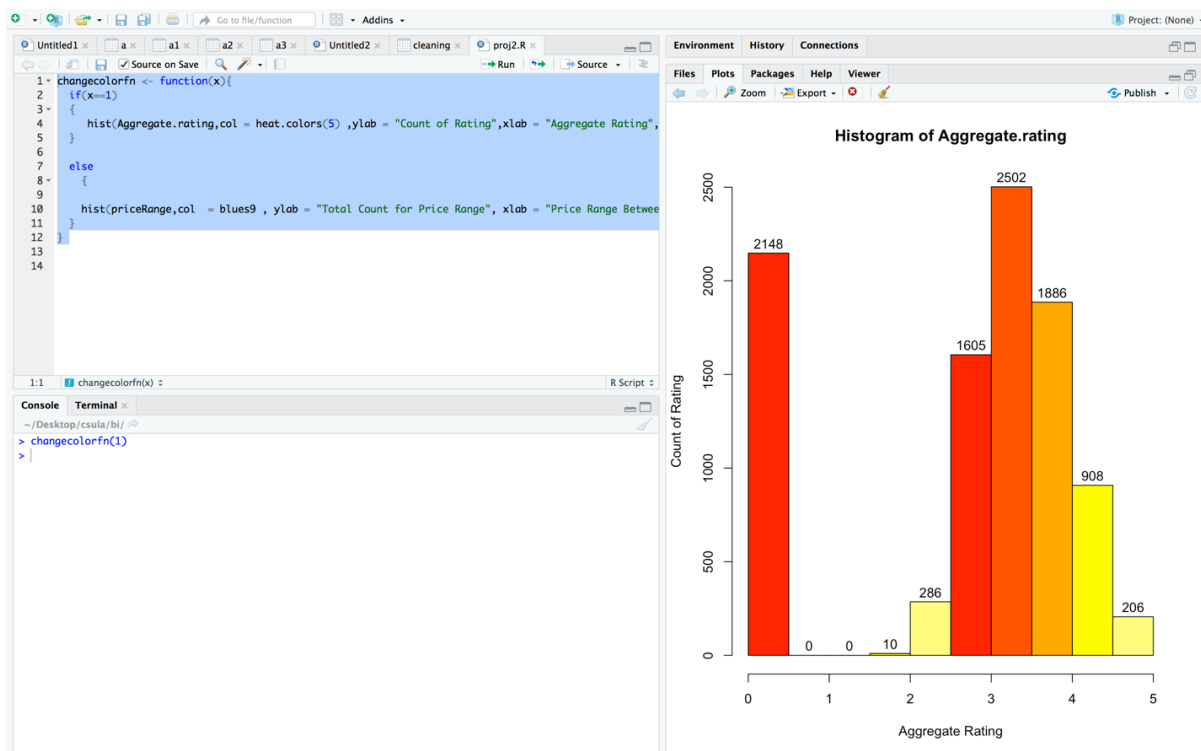
Function:

User Defined Function:

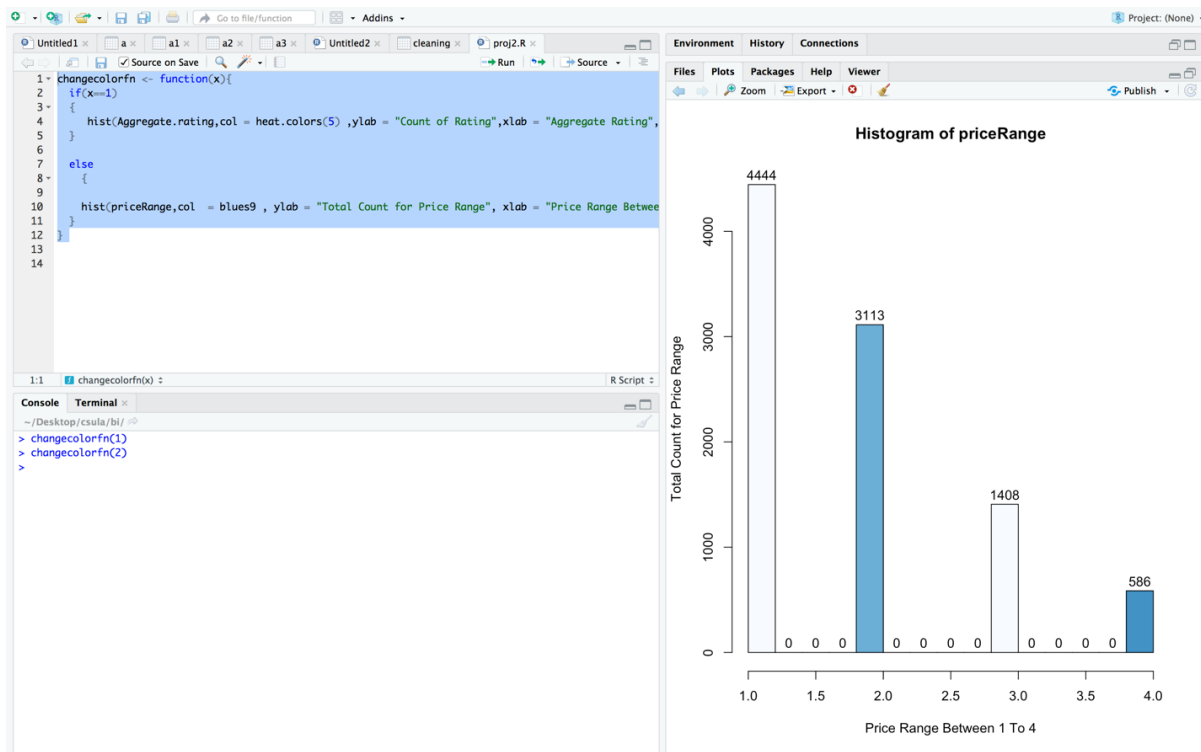
This function is used to switch histogram graph for price range and aggregate rating, means if we call the function `changecolorfn()` with value if `x` as 1(i.e.

`changecolorfn(1))` then it will give a histogram for aggregate ratings where as is the value of `x` is any number other than 1 then it will give a histogram for price range also with different colours.

Output when `changecolorfn(1)` :



Output when changecolorfn(2) or any number :



Summary :

```
> summary(Aggregate.rating)  
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
 0.000  2.500   3.200   2.666  3.700   4.900   
> summary(Average.Cost.for.two)  
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
    0    250    400   1199    700 800000   
> summary(Votes)  
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
  0.0    5.0    31.0   156.9   131.0 10934.0   
>
```

R Code for all the Analysis and Visualizations:

DATA - CLEANING :

```
>setwd("C:/Users/Dear User/Desktop")
```

```

>getwd()

> a <- read.csv("zomato.csv", header = T, sep = ",")

>View(a)

> a1 <- a[-6]

>View(a1)

>install.packages("tidyr")

>library(tidyr)

> a3 <- unite(a2, Location, c("Longitude","Latitude"), sep = ",", remove = TRUE)

>View(a3)

> data<- read.csv("zomato1.csv")

> View(data)

> data$Price.range[which(is.na(data$Price.range))]<-0

> View(data)

```

PIE - CHART: “Average cost for two of top 10 restaurants”

```

> top10ZomatoNumeric <- data.frame(Restaurant.ID,Restaurant.Name, Average.Cost
.for.two)

> top10ZomatoNumeric$Average.Cost.for.two <- a$Average.Cost.for.two

> top10ZomatoNumeric$Restaurant.ID <- a$Restaurant.ID

> top10ZomatoNumeric$Restaurant.Name <- a$Restaurant.Name

> top10ZomatoNumericPie <- top10ZomatoNumeric[order(-top10ZomatoNumeric$A
verage.Cost.for.two) , ][1:10,]

> print(top10ZomatoNumericPie)

Convert to numeric :

> class(top10ZomatoNumericPie)

[1] "data.frame"

```

```
> top10ZomatoNumericPie <- as.numeric(as.character(unlist(top10ZomatoNumericPie
e)))

> class(top10ZomatoNumericPie)

[1] "numeric"

> pie(top10ZomatoNumericPie[1:10][indices] ,labels = top10Zomato$Restaurant.Name[1:10],main = "Top 10 Restaurants according to Average Cost",clockwise = TRUE
,col = terrain.colors(10))
```

BAR - CHART: “Customer ratings of all the restaurants”

```
> head(a$Rating.text)

[1] Excellent Excellent Very Good Excellent Excellent Very Good

Levels: Average Excellent Good Not rated Poor Very Good

> rating <- table(a$Rating.text)

> barplot(rating, main = "Customer Rating Distribution" ,xlab = "Rating Catogories" ,
ylab = "Count for Rating", col = rainbow(6))
```

LINE-CHART: “Customer votes of top 10 restaurant”

```
> yrest <- as.character(unlist(top10ZomatoVotes$Restaurant.Name))

> yrest

[1] "Toit"          "Truffles"       "Hauz Khas Social"  "Peter Cat"

[5] "AB's - Absolute Barbecues" "Barbeque Nation"  "Big Brewsky"

"AB's - Absolute Barbecues"

[9] "The Black Pearl"  "BarBQ"

> plot(xvotes,type = "o",col="red",xaxt="n",ylab = "No of Votes", xlab = "Top 10
Resturants",main = "Top 10 Restaurents According to Vote")
```

```
> axis(1,at = 1:10,labels = c(yrest))
```

HISTOGRAM: “Overall Customer ratings on all the restaurants”

```
changecolorfn(1)
```

HISTOGRAM: “Customer preferable price range on all the restaurants”

```
changecolorfn(2) OR changecolorfn(3) OR changecolorfn(...n except 1)
```

USER DEFINED FUNCTION : “changecolorfn(x)”

HISTOGRAM: “Overall Customer ratings on all the restaurants”

```
changecolorfn <- function(x){
```

```
  if(x==1)
```

```
  {
```

```
    hist(Aggregate.rating,col = heat.colors(5) ,ylab = "Count of Rating",xlab = "Aggregate Rating",labels = TRUE)
```

```
  }
```

HISTOGRAM: “Customer preferable price range on all the restaurants”

```
  else
```

```
  {
```

```
    hist(priceRange,col = blues9 , ylab = "Total Count for Price Range", xlab = "Price Range Between 1 To 4",labels = TRUE)
```

```
  }
```

```
}
```

Statistical Averages Of Data : “Summary Of Aggregate Customer Rating, Average Cost to two person, Customer Votes for Restaurant”

```
> summary(Aggregate.rating)
```

```
Min. 1st Qu. Median Mean 3rd Qu. Max.
0.000 2.500 3.200 2.666 3.700 4.900
```

```
> summary(Average.Cost.for.two)
```

```
Min. 1st Qu. Median Mean 3rd Qu. Max.
0 250 400 1199 700 800000
```

```
> summary(Votes)
```

```
Min. 1st Qu. Median Mean 3rd Qu. Max.
0.0 5.0 31.0 156.9 131.0 10934.0
```

An example of Pie Chart Script in R(Click On the Image Below And Download the R

File): https://drive.google.com/open?id=1F304RkF0zu6DIeHZOfkYoXYNpItq-0_w



PieChart.R

References:

1. Patodi, Shaneel. (Nov 2016). "Presentation on zomato". Retrieved from "<https://www.slideshare.net/mobile/ShaneelPatodi/zomato-presentation>".
2. "CASE STUDY- Zomato: An Indian Startup acquiring the world". Retrieved from "<http://dsim.in/blog/2015/04/14/zomato-an-indian-startup-acquiring-the-world>".
3. Ravish , Vishal. (Aug 2017). "Case Study on Zomato : A Global Indian Company". Retrieved from "<https://googleweblight.com/i?u=https://www.whizsky.com/2017/08/case-study-zomato-global-indian-company/&hl=en-IN>".
4. Mathur, Sneha. (July 2015). "Zomato: An Indian startup acquiring the world". Retrieved from "<https://googleweblight.com/i?u=https://www.digitalvidya.com/blog/zomato-an-indian-startup-acquiring-the-world/&hl=en-IN>".

