

Identifying and Recommending Best Restaurants

Andrei Enescu

11/13/2020

DESCRIPTION

A restaurant consolidator is looking to revamp the B2C portal using intelligent automation tech. This requires a different matrix to identify the star restaurants and generate recommendations. To make sure an effective model can be achieved, it is important to understand the behavior of the data in hand.

There are 2 datasets, data and Country-Code. Dataset data has 19 attributes and Country-Code has two attributes.

You are required to determine certain matrices to identify the star restaurants and generate recommendations.

Step to perform:

Importing, Understanding, and Inspecting Data :

Perform preliminary data inspection and report the findings as the structure of the data, missing values, duplicates, etc.

Based on the findings from the previous questions, identify duplicates and remove them

Performing EDA:

1. Explore the geographical distribution of the restaurants and identify the cities with the maximum and minimum number of restaurants
2. Restaurant franchising is a thriving venture. So, it is very important to explore the franchise with most national presence
3. Find out the ratio between restaurants that allow table booking vs. those that do not allow table booking
4. Find out the percentage of restaurants providing online delivery
5. Calculate the difference in number of votes for the restaurants that deliver and the restaurants that do not deliver
6. What are the top 10 cuisines served across cities?

7. What is the maximum and minimum number of cuisines that a restaurant serves? Also, which is the most served cuisine across the restaurant for each city?
8. What is the distribution cost across the restaurants?
9. How ratings are distributed among the various factors?
10. Explain the factors in the data that may have an effect on ratings. For example, number of cuisines, cost, delivery option, etc.

Dashboarding:

Visualize the variables using Tableau to help user explore the data and create a better understanding of the restaurants to identify the “star” restaurant

Demonstrate the variables associated with each other and factors to build a dashboard

These are the libraries I used:

```
library(rio)
library(tmaptools)
library(plyr)
library(dplyr)
library(janitor)
library(ggplot2)
library(forcats)
library(plotly)
library(wordcloud)
library(tm)
library(tidyr)
library(stringr)
```

Importing, Understanding, and Inspecting Data :

Perform preliminary data inspection and report the findings as the structure of the data, missing values, duplicates, etc.

Based on the findings from the previous questions, identify duplicates and remove them

```
Cntry_code <- import("Country-Code.xlsx")
RestDF <- import("data.xlsx")

head(Cntry_code)
```

```
##   Country Code   Country
## 1           1      India
```

```
## 2      14  Australia
## 3      30    Brazil
## 4      37    Canada
## 5      94  Indonesia
## 6     148 New Zealand
```

```
str(Cntry_code)
```

```
## 'data.frame':  15 obs. of  2 variables:
## $ Country Code: num  1 14 30 37 94 148 162 166 184 189 ...
## $ Country      : chr  "India" "Australia" "Brazil" "Canada" ...
```

```
colnames(Cntry_code)
```

```
## [1] "Country Code" "Country"
```

```
summary(Cntry_code)
```

```
## Country Code      Country
## Min.   : 1.0      Length:15
## 1st Qu.: 65.5     Class :character
## Median :166.0     Mode  :character
## Mean   :137.9
## 3rd Qu.:199.5
## Max.   :216.0
```

```
which(is.na(Cntry_code))
```

```
## integer(0)
```

```
sum(duplicated(Cntry_code))
```

```
## [1] 0
```

```
head(RestDF)
```

```
## Restaurant ID      Restaurant Name Country Code    City
## 1      7402935                Skye          94 Jakarta
## 2      7410290    Satoo - Hotel Shangri-La      94 Jakarta
## 3      7420899                Sushi Masa      94 Jakarta
## 4      7421967            3 Wise Monkeys      94 Jakarta
## 5      7422489    Avec Moi Restaurant and Bar  94 Jakarta
## 6     18352452    Lucky Cat Coffee & Kitchen  94 Jakarta
##
##                                Address
## 1      Menara BCA, Lantai 56, Jl. MH. Thamrin, Thamrin, Jakarta
## 2                                Hotel Shangri-La, Jl. Jend. Sudirman
## 3                                Jl. Tuna Raya No. 5, Penjaringan
## 4                                Jl. Suryo No. 26, Senopati, Jakarta
## 5      Gedung PIC, Jl. Teluk Betung 43, Thamrin, Jakarta
```

```

## 6 Plaza Festival, South Parking, Jl. HR Rasuna Said, Kuningan, Jakarta
##          Locality                      Locality Verbose
## 1 Grand Indonesia Mall, Thamrin Grand Indonesia Mall, Thamrin, Jakarta
## 2   Hotel Shangri-La, Sudirman   Hotel Shangri-La, Sudirman, Jakarta
## 3             Penjaringan             Penjaringan, Jakarta
## 4             Senopati             Senopati, Jakarta
## 5             Thamrin             Thamrin, Jakarta
## 6   Plaza Festival, Kuningan   Plaza Festival, Kuningan, Jakarta
## Longitude Latitude          Cuisines Average Cost for two
## 1  106.8220 -6.196778      Italian, Continental      800000
## 2  106.8190 -6.203292 Asian, Indonesian, Western      800000
## 3  106.8001 -6.101298          Sushi, Japanese      500000
## 4  106.8134 -6.235241          Japanese      450000
## 5  106.8210 -6.196270      French, Western      350000
## 6  106.8317 -6.218932      Cafe, Western      300000
##          Currency Has Table booking Has Online delivery Price range
## 1 Indonesian Rupiah(IDR)          No          No          3
## 2 Indonesian Rupiah(IDR)          No          No          3
## 3 Indonesian Rupiah(IDR)          No          No          3
## 4 Indonesian Rupiah(IDR)          No          No          3
## 5 Indonesian Rupiah(IDR)          No          No          3
## 6 Indonesian Rupiah(IDR)          No          No          3
## Aggregate rating Rating color Rating text Votes
## 1          4.1      Green   Very Good  1498
## 2          4.6  Dark Green   Excellent   873
## 3          4.9  Dark Green   Excellent   605
## 4          4.2      Green   Very Good   395
## 5          4.3      Green   Very Good   243
## 6          4.3      Green   Very Good   458

```

```
str(RestDF)
```

```

## 'data.frame':   9551 obs. of  19 variables:
## $ Restaurant ID      : num  7402935 7410290 7420899 7421967 7422489 ...
## $ Restaurant Name    : chr   "Skye" "Satoo - Hotel Shangri-La" "Sushi Masa" "3 Wise Monkeys" ...
## $ Country Code       : num   94 94 94 94 94 94 94 94 94 94 ...
## $ City               : chr   "Jakarta" "Jakarta" "Jakarta" "Jakarta" ...
## $ Address            : chr   "Menara BCA, Lantai 56, Jl. MH. Thamrin, Thamrin, Jakarta" "Hotel Shangri-La, Sudirman" "Penjaringan, Jakarta" ...
## $ Locality           : chr   "Grand Indonesia Mall, Thamrin" "Hotel Shangri-La, Sudirman" "Penjaringan, Jakarta" ...
## $ Locality Verbose   : chr   "Grand Indonesia Mall, Thamrin, Jakarta" "Hotel Shangri-La, Sudirman, Jakarta" ...
## $ Longitude          : num   107 107 107 107 107 ...
## $ Latitude           : num  -6.2 -6.2 -6.1 -6.24 -6.2 ...
## $ Cuisines            : chr   "Italian, Continental" "Asian, Indonesian, Western" "Sushi, Japanese" ...
## $ Average Cost for two: num   800000 800000 500000 450000 350000 300000 300000 250000 250000 200000 ...
## $ Currency           : chr   "Indonesian Rupiah(IDR)" "Indonesian Rupiah(IDR)" "Indonesian Rupiah(IDR)" ...
## $ Has Table booking   : chr   "No" "No" "No" "No" ...
## $ Has Online delivery : chr   "No" "No" "No" "No" ...
## $ Price range         : num    3 3 3 3 3 3 3 3 3 3 ...
## $ Aggregate rating    : num    4.1 4.6 4.9 4.2 4.3 4.3 3.7 4 4.2 4.9 ...
## $ Rating color        : chr   "Green" "Dark Green" "Dark Green" "Green" ...
## $ Rating text         : chr   "Very Good" "Excellent" "Excellent" "Very Good" ...
## $ Votes               : num   1498 873 605 395 243 ...

```

```
colnames(RestDF)
```

```
## [1] "Restaurant ID"      "Restaurant Name"    "Country Code"
## [4] "City"               "Address"            "Locality"
## [7] "Locality Verbose"   "Longitude"          "Latitude"
## [10] "Cuisines"           "Average Cost for two" "Currency"
## [13] "Has Table booking"  "Has Online delivery" "Price range"
## [16] "Aggregate rating"   "Rating color"       "Rating text"
## [19] "Votes"
```

```
summary(RestDF)
```

```
## Restaurant ID      Restaurant Name      Country Code      City
## Min. :      53      Length:9551      Min. :  1.00      Length:9551
## 1st Qu.: 301962      Class :character  1st Qu.:  1.00      Class :character
## Median : 6004089      Mode  :character  Median :  1.00      Mode  :character
## Mean   : 9051128                      Mean   : 18.37
## 3rd Qu.:18352292                      3rd Qu.:  1.00
## Max.   :18500652                      Max.   :216.00
## Address            Locality            Locality Verbose    Longitude
## Length:9551         Length:9551         Length:9551         Min.   : -157.95
## Class :character    Class :character    Class :character    1st Qu.:  77.08
## Mode  :character    Mode  :character    Mode  :character    Median :  77.19
##                                     Mean   :  64.13
##                                     3rd Qu.:  77.28
##                                     Max.   : 174.83
## Latitude           Cuisines            Average Cost for two  Currency
## Min.   : -41.33      Length:9551         Min.   :    0      Length:9551
## 1st Qu.: 28.48      Class :character    1st Qu.:  250      Class :character
## Median : 28.57      Mode  :character    Median :   400      Mode  :character
## Mean   : 25.85                      Mean   : 1199
## 3rd Qu.: 28.64                      3rd Qu.:  700
## Max.   : 55.98                      Max.   :800000
## Has Table booking  Has Online delivery  Price range          Aggregate rating
## Length:9551         Length:9551         Min.   :1.000      Min.   :0.000
## Class :character    Class :character    1st Qu.:1.000      1st Qu.:2.500
## Mode  :character    Mode  :character    Median :2.000      Median :3.200
##                                     Mean   :1.805      Mean   :2.666
##                                     3rd Qu.:2.000      3rd Qu.:3.700
##                                     Max.   :4.000      Max.   :4.900
## Rating color        Rating text          Votes
## Length:9551         Length:9551         Min.   :    0.0
## Class :character    Class :character    1st Qu.:    5.0
## Mode  :character    Mode  :character    Median :   31.0
##                                     Mean   : 156.9
##                                     3rd Qu.: 131.0
##                                     Max.   :10934.0
```

```
colnames(RestDF)[colSums(is.na(RestDF)) > 0]
```

```
## [1] "Restaurant Name" "Cuisines"
```

```
which (is.na(RestDF$`Restaurant Name`))
```

```
## [1] 1647
```

```
which (is.na(RestDF$Cuisines))
```

```
## [1] 9084 9087 9095 9407 9495 9505 9534 9536 9540
```

There is one restaurant name missing and 9 records without cuisine.

Omitting the row without restaurant name and replacing the cuisine NA values with “Other”.

```
RestDF$Cuisines[is.na(RestDF$Cuisines)] <- "Other"  
RestDF <- na.omit(RestDF)
```

```
sum(duplicated(RestDF))
```

```
## [1] 0
```

```
RestDF %>% get_dupes()
```

```
## No variable names specified - using all columns.
```

```
## No duplicate combinations found of: Restaurant ID, Restaurant Name, Country Code, City, Address, Locality
```

```
## [1] Restaurant ID      Restaurant Name      Country Code  
## [4] City                Address             Locality  
## [7] Locality Verbose     Longitude           Latitude  
## [10] Cuisines              Average Cost for two Currency  
## [13] Has Table booking    Has Online delivery Price range  
## [16] Aggregate rating     Rating color        Rating text  
## [19] Votes                dupe_count  
## <0 rows> (or 0-length row.names)
```

There are no duplicates

Geocoding the missing locations

```
Rest_temp <- subset(RestDF, subset = Longitude == 0)  
Rest_temp1 <- geocode_OSM(Rest_temp$City, as.sf = T)  
Rest_temp1 <- as.data.frame(Rest_temp1)  
Rest_temp1 <- Rest_temp1[,1:3]  
colnames(Rest_temp1) <- c("City", "Latitude", "Longitude")  
Rest_temp1 <- Rest_temp1[c(1,3,2)]  
Rest_temp2 <- cbind(Rest_temp, Rest_temp1, by = "City")  
Rest_temp2 <- Rest_temp2[c(1:7,21,22, 8:20, 23)]  
Rest_temp2 <- Rest_temp2[, -c(10,11,22,23)]
```

```
RestDF <- RestDF[-which(RestDF$Longitude == 0),]  
RestDF <- rbind(RestDF, Rest_temp2)
```

Performing EDA:

1. Explore the geographical distribution of the restaurants and identify the cities with the maximum and minimum number of restaurants

Merging the two data sets

```
RestDF1 <- merge(RestDF, Cntry_code, by = "Country Code")
head(RestDF1)
```

```
##      Country Code Restaurant ID      Restaurant Name
## 1          1         2701      Orient Express - Taj Palace Hotel
## 2          1       309548      Tian - Asian Cuisine Studio - ITC Maurya
## 3          1         7066              Berco's
## 4          1       301523      Nostalgia at 1911 Brasserie - The Imperial
## 5          1         2724              1911 - The Imperial
## 6          1         2742      Bukhara - ITC Maurya
##      City      Address
## 1 New Delhi Taj Palace Hotel, Diplomatic Enclave, Chanakyapuri, New Delhi
## 2 New Delhi      ITC Maurya, Diplomatic Enclave, Chanakyapuri, New Delhi
## 3 New Delhi      Food Court, Moments Mall, Kirti Nagar, New Delhi
## 4 New Delhi      The Imperial, Janpath, New Delhi
## 5 New Delhi      The Imperial, Janpath, New Delhi
## 6 New Delhi      ITC Maurya, Chanakyapuri, New Delhi
##      Locality
## 1 The Taj Palace Hotel, Chanakyapuri
## 2      ITC Maurya, Chanakyapuri
## 3      Moments Mall, Kirti Nagar
## 4      The Imperial, Janpath
## 5      The Imperial, Janpath
## 6      ITC Maurya, Chanakyapuri
##      Locality Verbose Longitude Latitude
## 1 The Taj Palace Hotel, Chanakyapuri, New Delhi 77.17009 28.59501
## 2      ITC Maurya, Chanakyapuri, New Delhi 77.17345 28.59735
## 3      Moments Mall, Kirti Nagar, New Delhi 77.14673 28.65686
## 4      The Imperial, Janpath, New Delhi 77.21819 28.62544
## 5      The Imperial, Janpath, New Delhi 77.21819 28.62544
## 6      ITC Maurya, Chanakyapuri, New Delhi 77.17372 28.59747
##      Cuisines Average Cost for two
## 1          European              8000
## 2      Asian, Japanese, Korean, Thai, Chinese 7000
## 3          Chinese, Thai              1100
## 4          European, Continental        6000
## 5      North Indian, Chinese, South Indian, Italian 6000
## 6          North Indian              6500
##      Currency Has Table booking Has Online delivery Price range
## 1 Indian Rupees(Rs.)      Yes      No      4
## 2 Indian Rupees(Rs.)      No      No      4
## 3 Indian Rupees(Rs.)      No      Yes      3
## 4 Indian Rupees(Rs.)      Yes      No      4
## 5 Indian Rupees(Rs.)      Yes      No      4
## 6 Indian Rupees(Rs.)      No      No      4
```

	Aggregate rating	Rating	color	Rating text	Votes	Country
## 1	4.0		Green	Very Good	145	India
## 2	4.1		Green	Very Good	188	India
## 3	3.5		Yellow	Good	120	India
## 4	3.2		Orange	Average	12	India
## 5	3.9		Yellow	Good	272	India
## 6	4.4		Green	Very Good	2826	India

```
count(RestDF1, Country, sort = T)
```

	Country	n
## 1	India	8651
## 2	United States	434
## 3	United Kingdom	80
## 4	Brazil	60
## 5	South Africa	60
## 6	UAE	60
## 7	New Zealand	40
## 8	Turkey	34
## 9	Australia	24
## 10	Phillipines	22
## 11	Indonesia	21
## 12	Qatar	20
## 13	Singapore	20
## 14	Sri Lanka	20
## 15	Canada	4

```
ggplot(RestDF1,
  aes(x = Country)) +
  geom_bar(aes(x = fct_infreq(Country), fill = Country),
    position = "dodge",
    stat = "count") +
  labs(title = "No of restaurants in each country",
    x = "Countries",
    y = "No. of restaurants") +
  theme(axis.text.x = element_text(angle = 75, size = 8),
    plot.title = element_text(hjust = 0.5))
```




India has the most number of restaurants - 8651, followed by USA - 434 The least number of restaurants are in Canada - 4

```
count(RestDF1, Country, City, sort = T) %>% top_n(5, wt = n)
```

```
##   Country      City    n
## 1   India New Delhi 5473
## 2   India Gurgaon 1118
## 3   India Noida 1080
## 4   India Faridabad 251
## 5   India Ghaziabad 25
```

New Delhi has the most restaurants - 5473

```
count(RestDF1, Country, City, sort = T) %>% top_n(-46, wt = n)
```

```
##           Country      City n
## 1   Australia Armidale 1
## 2   Australia Balingup 1
## 3   Australia Beechworth 1
## 4   Australia Dicky Beach 1
## 5   Australia East Ballina 1
## 6   Australia Flaxton 1
## 7   Australia Forrest 1
## 8   Australia Huskisson 1
```

```
## 9      Australia      Inverloch 1
## 10     Australia    Lakes Entrance 1
## 11     Australia           Lorn 1
## 12     Australia      Macedon 1
## 13     Australia      Mayfield 1
## 14     Australia    Middleton Beach 1
## 15     Australia      Montville 1
## 16     Australia      Palm Cove 1
## 17     Australia      Paynesville 1
## 18     Australia      Penola 1
## 19     Australia    Phillip Island 1
## 20     Australia      Tanunda 1
## 21     Australia    Trentham East 1
## 22     Australia    Victor Harbor 1
## 23           Canada    Chatham-Kent 1
## 24           Canada      Consort 1
## 25           Canada Vineland Station 1
## 26           Canada      Yorkton 1
## 27           India      Mohali 1
## 28           India    Panchkula 1
## 29     Indonesia      Bandung 1
## 30    Phillipines    Quezon City 1
## 31    Phillipines    Tagaytay City 1
## 32    South Africa    Randburg 1
## 33    United States    Clatskanie 1
## 34    United States    Cochrane 1
## 35    United States    Fernley 1
## 36    United States    Lakeview 1
## 37    United States    Lincoln 1
## 38    United States    Mc Millan 1
## 39    United States    Miller 1
## 40    United States    Monroe 1
## 41    United States    Ojo Caliente 1
## 42    United States    Potrero 1
## 43    United States    Princeton 1
## 44    United States    Vernonia 1
## 45    United States    Weirton 1
## 46    United States    Winchester Bay 1
```

There are 46 cities in 7 countries that have only 1 restaurant

2. Restaurant franchising is a thriving venture. So, it is very important to explore the franchise with most national presence

```
count(RestDF1, Country, 'Restaurant Name', sort = T) %>% top_n(10, wt = n)
```

```
##      Country Restaurant Name  n
## 1      India  Cafe Coffee Day 83
## 2      India  Domino's Pizza 79
## 3      India      Subway 63
```

```
## 4    India Green Chick Chop 51
## 5    India      McDonald's 48
## 6    India      Keventers 34
## 7    India      Giani 29
## 8    India      Pizza Hut 29
## 9    India      Baskin Robbins 28
## 10   India      Barbeque Nation 25
```

Most franchises are in India. The biggest franchise in term of location is Cafe Coffee Day with 83 restaurants
 Checking for franchises in the rest of the countries from the data set

```
RestDF1 %>%
  group_by(Country) %>%
  filter(any(Country == "Australia")) %>%
  count('Restaurant Name', sort = T) %>%
  top_n(10, wt = n)
```

```
## # A tibble: 24 x 3
## # Groups:   Country [1]
##   Country 'Restaurant Name'      n
##   <chr>    <chr>              <int>
## 1 Australia 1918 Bistro & Grill      1
## 2 Australia 5 Little Pigs           1
## 3 Australia Anchorage Cafe Restaurant Wine Bar 1
## 4 Australia Beach Box Cafe         1
## 5 Australia Bespoke Harvest        1
## 6 Australia Blue Bean Love Cafe     1
## 7 Australia Bridge Road Brewers     1
## 8 Australia DiVine                  1
## 9 Australia Flaxton Gardens         1
## 10 Australia Funkey Monkey          1
## # ... with 14 more rows
```

```
RestDF1 %>%
  group_by(Country) %>%
  filter(any(Country == "Brazil")) %>%
  count('Restaurant Name', sort = T) %>%
  top_n(10, wt = n)
```

```
## # A tibble: 58 x 3
## # Groups:   Country [1]
##   Country 'Restaurant Name'      n
##   <chr>    <chr>              <int>
## 1 Brazil  Coco Bambu           2
## 2 Brazil  Garota de Ipanema      2
## 3 Brazil  A Figueira Rubaiyat     1
## 4 Brazil  Aconchego Carioca       1
## 5 Brazil  Amir                   1
## 6 Brazil  Aprazível               1
## 7 Brazil  Balada Mix               1
## 8 Brazil  Beirute                  1
## 9 Brazil  Bibi                     1
```

```
## 10 Brazil Braseiro da GiÁvea      1
## # ... with 48 more rows
```

```
RestDF1 %>%
  group_by(Country) %>%
  filter(any(Country == "Canada")) %>%
  count('Restaurant Name', sort = T) %>%
  top_n(10, wt = n)
```

```
## # A tibble: 4 x 3
## # Groups:   Country [1]
##   Country 'Restaurant Name'      n
##   <chr>   <chr>                <int>
## 1 Canada Arigato Sushi          1
## 2 Canada Consort Restaurant      1
## 3 Canada Lake House Restaurant    1
## 4 Canada Tokyo Sushi              1
```

```
RestDF1 %>%
  group_by(Country) %>%
  filter(any(Country == "Indonesia")) %>%
  count('Restaurant Name', sort = T) %>%
  top_n(10, wt = n)
```

```
## # A tibble: 18 x 3
## # Groups:   Country [1]
##   Country 'Restaurant Name'      n
##   <chr>   <chr>                <int>
## 1 Indonesia Talaga Sampireun      3
## 2 Indonesia Fish Streat           2
## 3 Indonesia 3 Wise Monkeys         1
## 4 Indonesia Avec Moi Restaurant and Bar 1
## 5 Indonesia Flip Burger           1
## 6 Indonesia Lemongrass             1
## 7 Indonesia Lucky Cat Coffee & Kitchen 1
## 8 Indonesia Momo Milk              1
## 9 Indonesia MONKS                  1
## 10 Indonesia Noah's Barn Coffeenery    1
## 11 Indonesia OJJU                  1
## 12 Indonesia Onokabe                1
## 13 Indonesia Satoo - Hotel Shangri-La 1
## 14 Indonesia Skye                   1
## 15 Indonesia Sushi Masa             1
## 16 Indonesia Toodz House            1
## 17 Indonesia Union Deli             1
## 18 Indonesia Zenbu                  1
```

```
RestDF1 %>%
  group_by(Country) %>%
  filter(any(Country == "New Zealand")) %>%
  count('Restaurant Name', sort = T) %>%
  top_n(10, wt = n)
```

```
## # A tibble: 40 x 3
## # Groups:   Country [1]
##   Country      'Restaurant Name'      n
##   <chr>        <chr>                <int>
## 1 New Zealand Baduzzi                  1
## 2 New Zealand Big Fish Eatery          1
## 3 New Zealand Burger Liquor            1
## 4 New Zealand Caffè L'affare            1
## 5 New Zealand Charley Noble Eatery & Bar 1
## 6 New Zealand Chinoiserie              1
## 7 New Zealand De Fontein Belgian Beer Cafe 1
## 8 New Zealand Depot Eatery and Oyster Bar 1
## 9 New Zealand Dragonfly                 1
## 10 New Zealand Eden Noodles Cafe Ė_Ėàù_ŮĖ_ŮĖ_ùċ 1
## # ... with 30 more rows
```

```
RestDF1 %>%
  group_by(Country) %>%
  filter(any(Country == "Phillipines")) %>%
  count('Restaurant Name', sort = T) %>%
  top_n(10, wt = n)
```

```
## # A tibble: 21 x 3
## # Groups:   Country [1]
##   Country      'Restaurant Name'      n
##   <chr>        <chr>                <int>
## 1 Phillipines Silantro Fil-Mex          2
## 2 Phillipines Balay Dako                1
## 3 Phillipines Buffet 101                1
## 4 Phillipines Cafe Arabelle              1
## 5 Phillipines Din Tai Fung               1
## 6 Phillipines Guevarra's                1
## 7 Phillipines Heat - Edsa Shangri-La     1
## 8 Phillipines Hobing Korean Dessert Cafe 1
## 9 Phillipines Izakaya Kikufuji           1
## 10 Phillipines Le Petit Souffle           1
## # ... with 11 more rows
```

```
RestDF1 %>%
  group_by(Country) %>%
  filter(any(Country == "Qatar")) %>%
  count('Restaurant Name', sort = T) %>%
  top_n(10, wt = n)
```

```
## # A tibble: 20 x 3
## # Groups:   Country [1]
##   Country      'Restaurant Name'      n
##   <chr>        <chr>                <int>
## 1 Qatar       7st by Mumbai Spices          1
## 2 Qatar       Aalishan                          1
## 3 Qatar       Applebee's                          1
## 4 Qatar       Coral - InterContinental Doha         1
## 5 Qatar       Eatopia                             1
```

```
## 6 Qatar Gokul Gujarati Restaurant 1
## 7 Qatar Gymkhana 1
## 8 Qatar Indian Coffee House 1
## 9 Qatar Mainland China Restaurant 1
## 10 Qatar MRA Bakery Sweets & Restaurant 1
## 11 Qatar Paper Moon 1
## 12 Qatar Ponderosa 1
## 13 Qatar Punjab Restaurant 1
## 14 Qatar Roti & Boti 1
## 15 Qatar Sabai Thai - The Westin Doha Hotel & Spa 1
## 16 Qatar Texas Roadhouse 1
## 17 Qatar The Manhattan FISH MARKET 1
## 18 Qatar Vine - The St. Regis 1
## 19 Qatar Zaffran Dining Experience 1
## 20 Qatar Zaoq 1
```

```
RestDF1 %>%
  group_by(Country) %>%
  filter(any(Country == "Singapore")) %>%
  count('Restaurant Name', sort = T) %>%
  top_n(10, wt = n)
```

```
## # A tibble: 20 x 3
## # Groups:   Country [1]
##   Country   'Restaurant Name'      n
##   <chr>      <chr>                <int>
## 1 Singapore Al'frank Cookies      1
## 2 Singapore Artichoke Cafe        1
## 3 Singapore Artistry              1
## 4 Singapore Bitters & Love        1
## 5 Singapore Boufe Boutique Cafe   1
## 6 Singapore Chye Seng Huat Hardware 1
## 7 Singapore Colony                1
## 8 Singapore Cut By Wolfgang Puck   1
## 9 Singapore Fratini La Trattoria   1
## 10 Singapore I Am                 1
## 11 Singapore Jaan                 1
## 12 Singapore Makansutra Gluttons Bay 1
## 13 Singapore Potato Head Folk      1
## 14 Singapore Restaurant Andre       1
## 15 Singapore Rhubarb Le Restaurant 1
## 16 Singapore Sky On 57             1
## 17 Singapore Summer Pavilion       1
## 18 Singapore Super Loco             1
## 19 Singapore The Lokal             1
## 20 Singapore The Refinery Singapore 1
```

```
RestDF1 %>%
  group_by(Country) %>%
  filter(any(Country == "South Africa")) %>%
  count('Restaurant Name', sort = T) %>%
  top_n(10, wt = n)
```

```
## # A tibble: 60 x 3
## # Groups:   Country [1]
##   Country      'Restaurant Name'      n
##   <chr>        <chr>                <int>
## 1 South Africa 23 On Hazelwood          1
## 2 South Africa Active Sushi            1
## 3 South Africa Baobab Cafe & Grill      1
## 4 South Africa Beluga                  1
## 5 South Africa Bloss Cafe               1
## 6 South Africa Brooklyn Brothers        1
## 7 South Africa Cafe Del Sol Botanico     1
## 8 South Africa Capital Craft Beer Academy 1
## 9 South Africa Carbon Bistro            1
## 10 South Africa Coco Safar              1
## # ... with 50 more rows
```

```
RestDF1 %>%
  group_by(Country) %>%
  filter(any(Country == "Sri Lanka")) %>%
  count('Restaurant Name', sort = T) %>%
  top_n(10, wt = n)
```

```
## # A tibble: 20 x 3
## # Groups:   Country [1]
##   Country      'Restaurant Name'      n
##   <chr>        <chr>                <int>
## 1 Sri Lanka Arabian Knights             1
## 2 Sri Lanka Burger's King              1
## 3 Sri Lanka Butter Boutique            1
## 4 Sri Lanka Cafe Beverly               1
## 5 Sri Lanka Cafe Shaze                 1
## 6 Sri Lanka Carnival Ice Cream         1
## 7 Sri Lanka Chinese Dragon Cafe        1
## 8 Sri Lanka CIOCONAT Lounge            1
## 9 Sri Lanka Cricket Club Cafe          1
## 10 Sri Lanka Elite Indian Restaurant    1
## 11 Sri Lanka Malay Restaurant           1
## 12 Sri Lanka Ministry of Crab           1
## 13 Sri Lanka Queen's Cafe              1
## 14 Sri Lanka Simply Strawberries By Jagro 1
## 15 Sri Lanka T.G.I. Friday's           1
## 16 Sri Lanka The Commons                1
## 17 Sri Lanka The Manhattan Fish Market  1
## 18 Sri Lanka The Paddington             1
## 19 Sri Lanka The Sizzle                 1
## 20 Sri Lanka Upali's                   1
```

```
RestDF1 %>%
  group_by(Country) %>%
  filter(any(Country == "Turkey")) %>%
  count('Restaurant Name', sort = T) %>%
  top_n(10, wt = n)
```

```
## # A tibble: 32 x 3
## # Groups:   Country [1]
##   Country 'Restaurant Name'      n
##   <chr>   <chr>                 <int>
## 1 Turkey Dî_vero€Ülu             2
## 2 Turkey Timboo Cafe             2
## 3 Turkey AÂÛÂÛk Kahve            1
## 4 Turkey Baltazar                1
## 5 Turkey Cafemiz                 1
## 6 Turkey Ceviz A€Üac€±          1
## 7 Turkey Dem Karakî_y            1
## 8 Turkey Draft Gastro Pub        1
## 9 Turkey Emirgan Sî_tîÂÛ        1
## 10 Turkey Gaga Manjero           1
## # ... with 22 more rows
```

```
RestDF1 %>%
  group_by(Country) %>%
  filter(any(Country == "UAE")) %>%
  count('Restaurant Name', sort = T) %>%
  top_n(10, wt = n)
```

```
## # A tibble: 55 x 3
## # Groups:   Country [1]
##   Country 'Restaurant Name'      n
##   <chr>   <chr>                 <int>
## 1 UAE     Applebee's                 3
## 2 UAE     AB's Absolute Barbecues      2
## 3 UAE     Gazebo                       2
## 4 UAE     The Cheesecake Factory        2
## 5 UAE     Al Mukhtar Bakery            1
## 6 UAE     Aroos Damascus               1
## 7 UAE     Bait El Khetyar              1
## 8 UAE     Barbeque Nation              1
## 9 UAE     Carnival By Tresind           1
## 10 UAE     Cho Gao - Crowne Plaza Abu Dhabi 1
## # ... with 45 more rows
```

```
RestDF1 %>%
  group_by(Country) %>%
  filter(any(Country == "United Kingdom")) %>%
  count('Restaurant Name', sort = T) %>%
  top_n(10, wt = n)
```

```
## # A tibble: 74 x 3
## # Groups:   Country [1]
##   Country      'Restaurant Name'      n
##   <chr>        <chr>                 <int>
## 1 United Kingdom Chaophraya           3
## 2 United Kingdom Jamie's Italian       3
## 3 United Kingdom Dishoom                2
## 4 United Kingdom San Carlo             2
## 5 United Kingdom 10 To 10 In Delhi       1
```



```
## 6 United Kingdom Akbars 1
## 7 United Kingdom Almost Famous Burgers 1
## 8 United Kingdom Archies 1
## 9 United Kingdom Bank 1
## 10 United Kingdom Bao 1
## # ... with 64 more rows
```

```
RestDF1 %>%
  group_by(Country) %>%
  filter(any(Country == "United States")) %>%
  count('Restaurant Name', sort = T) %>%
  top_n(10, wt = n)
```

```
## # A tibble: 423 x 3
## # Groups:   Country [1]
##   Country      'Restaurant Name'      n
##   <chr>         <chr>                <int>
## 1 United States Texas Roadhouse      4
## 2 United States HuHot Mongolian Grill 3
## 3 United States Ceviche Tapas Bar & Restaurant 2
## 4 United States Chick-fil-A          2
## 5 United States Granite City Food & Brewery 2
## 6 United States Los Agaves           2
## 7 United States Mellow Mushroom      2
## 8 United States Rhinehart's Oyster Bar 2
## 9 United States 'Ohana                1
## 10 United States 2 Dog                1
## # ... with 413 more rows
```

After India, the US has the most number of franchises - 4, followed by Indonesia, UAE and UK with 3 franchises each

3. Find out the ratio between restaurants that allow table booking vs. those that do not allow table booking

```
RestDF1$'Has Table booking' <-
  as.factor(RestDF1$'Has Table booking')

Tbl_bkg <- ddply(
  RestDF1,
  ~ 'Has Table booking',
  summarize,
  No = sum('Has Table booking' == "No"),
  Yes = sum('Has Table booking' == "Yes")
)
Tbl_bkg$'Table booking' <- Tbl_bkg$No + Tbl_bkg$Yes
Tbl_bkg
```

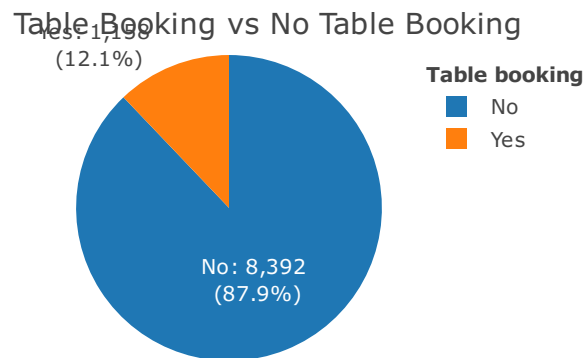
```
##   Has Table booking   No   Yes Table booking
## 1                No 8392    0      8392
## 2                Yes   0 1158      1158
```

```
(1158 / 9550) * 100
```

```
## [1] 12.12565
```

Only 12.1 % of the restaurants have the option to book a table

```
plot_ly(  
  Tbl_bkg,  
  labels = ~ 'Has Table booking',  
  values = ~ 'Table booking',  
  type = 'pie',  
  texttemplate = "%{label}: %{value:,.} <br>(%{percent})"  
) %>%  
  layout(  
    title = "Table Booking vs No Table Booking",  
    legend = list(title = list(text = '<b> Table booking </b>')),  
    xaxis = list(  
      showgrid = FALSE,  
      zeroline = FALSE,  
      showticklabels = FALSE  
    ),  
    yaxis = list(  
      showgrid = FALSE,  
      zeroline = FALSE,  
      showticklabels = FALSE  
    )  
  )  
)
```



We can observe in the above pie chart that 87.9% of restaurants don't have a table booking option

4. Find out the percentage of restaurants providing online delivery

```
RestDF1$`Has Online delivery` <-
  as.factor(RestDF1$`Has Online delivery`)

Online_bkg <- ddply(
  RestDF1,
  ~ `Has Online delivery`,
  summarize,
  No = sum(`Has Online delivery` == "No"),
  Yes = sum(`Has Online delivery` == "Yes")
)
Online_bkg$Online <- Online_bkg$No + Online_bkg$Yes
Online_bkg
```

```
##   Has Online delivery   No  Yes Online
## 1                No 7099    0   7099
## 2                Yes    0 2451   2451
```

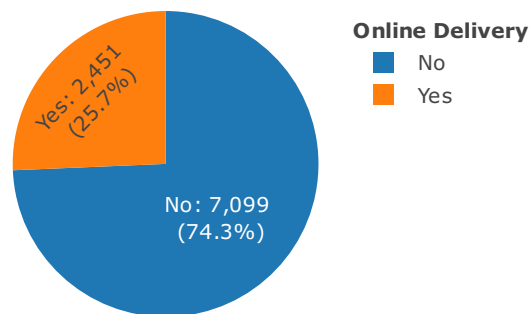
```
(2451 / 9550) * 100
```

```
## [1] 25.66492
```

25.7 % of the restaurants provide online delivery

```
plot_ly(  
  Online_bkg,  
  labels = ~ 'Has Online delivery',  
  values = ~ Online ,  
  type = 'pie',  
  texttemplate = "%{label}: %{value:,.} <br>(<math>\%{percent}</math>)"  
) %>%  
  layout(  
    title = "Online Delivery vs No Online Delivery",  
    legend = list(title = list(text = '<b> Online Delivery </b>')),  
    xaxis = list(  
      showgrid = FALSE,  
      zeroline = FALSE,  
      showticklabels = FALSE  
    ),  
    yaxis = list(  
      showgrid = FALSE,  
      zeroline = FALSE,  
      showticklabels = FALSE  
    )  
  )  
)
```

Online Delivery vs No Online Delivery



We can observe in the above pie chart that 74.3% of restaurants don't have an online delivery service

5. Calculate the difference in number of votes for the restaurants that deliver and the restaurants that do not deliver

```
Vote <- RestDF1 %>%  
  group_by('Has Online delivery') %>%  
  summarise(Votes = sum(Votes))
```

```
## 'summarise()' ungrouping output (override with '.groups' argument)
```

```
Vote
```

```
## # A tibble: 2 x 2  
##   'Has Online delivery'  Votes  
##   <fct>                <dbl>  
## 1 No                  979962  
## 2 Yes                 517914
```

```
sum(RestDF1$Votes)
```

```
## [1] 1497876
```

```
979962 - 517915
```

```
## [1] 462047
```

Restaurants with Online delivery have 517915 Votes

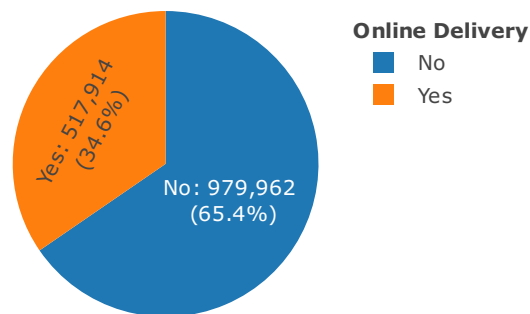
Restaurants without Online delivery have 979962 votes

The difference in number of votes is 462047 in favor for the restaurants without online delivery

```
Vote <- as.data.frame(Vote)
```

```
plot_ly(  
  Vote,  
  labels = ~ 'Has Online delivery',  
  values = ~ Votes ,  
  type = 'pie',  
  texttemplate = "%{label}: %{value:,.} <br>(%{percent})"  
) %>%  
  layout(  
    title = "Votes - Online Delivery vs No Online Delivery",  
    legend = list(title = list(text = '<b> Online Delivery </b>')),  
    xaxis = list(  
      showgrid = FALSE,  
      zeroline = FALSE,  
      showticklabels = FALSE  
    ),  
    yaxis = list(  
      showgrid = FALSE,  
      zeroline = FALSE,  
      showticklabels = FALSE  
    )  
  )  
)
```

Votes - Online Delivery vs No Online Delivery



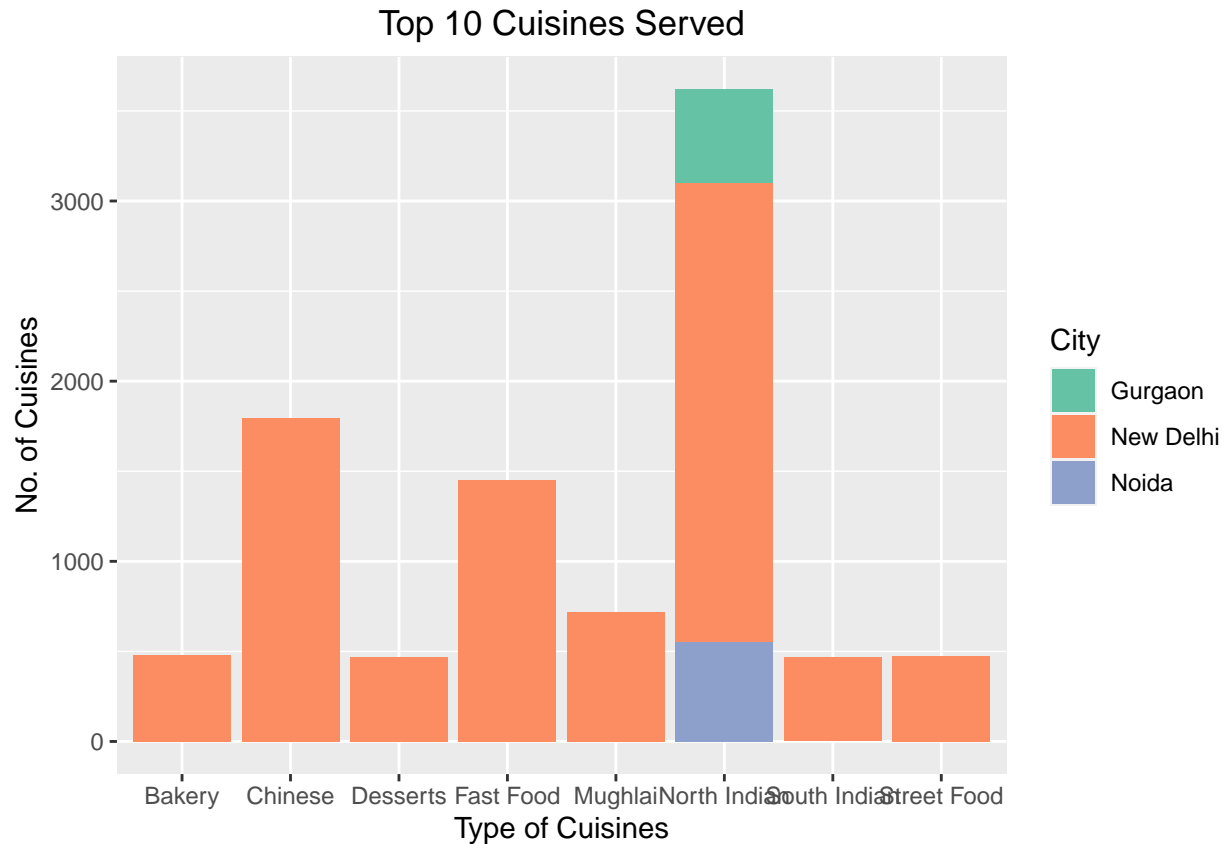
Restaurants without delivery have 65.4% out of the total votes and restaurants that have delivery have only 34.6% of total votes.

we conclude that other factors are more important in according higher votes.

6. What are the top 10 cuisines served across cities?

```
Cuisine <-  
  select(RestDF1, Country, City, 'Restaurant Name', Cuisines)  
  
Ungrouped <- Cuisine %>%  
  select(Cuisines, 'Restaurant Name') %>%  
  separate_rows(Cuisines, sep = ", ") %>%  
  count('Restaurant Name', Cuisines)  
  
Cuisine1 <- merge(Cuisine, Ungrouped, by = "Restaurant Name")  
  
Top10_City <- count(Cuisine1, City, Cuisines.y, sort = T) %>%  
  top_n(10, wt = n)  
  
ggplot(Top10_City,  
  aes(x = Cuisines.y, y = n)) +  
  geom_bar(stat = "identity", aes(fill = as.factor(City))) +
```

```
labs(title = "Top 10 Cuisines Served",
     x = "Type of Cuisines",
     y = "No. of Cuisines",
     fill = "City") +
scale_fill_brewer(palette = "Set2") +
theme(plot.title = element_text(hjust = 0.5))
```



As we can observe from the above output in the top 10 cuisines served we have only 3 cities: New Delhi, Noida and Gurgaon.

We can also see that “North Indian” cuisine is the most served.

Here are the top 10 cuisines in all of the countries

```
set.seed(1)
wordcloud(
  words = RestDF1$Cuisines,
  min.freq = 1,
  max.words = 10,
  random.order = FALSE,
  rot.per = 0.35,
  colors = brewer.pal(8, "Set2")
)
```




Below are the top 10 cuisine per country

Australia

```
Australia <- RestDF1 %>%  
  group_by(Country) %>%  
  filter(any(Country == "Australia")) %>%  
  count(Cuisines, sort = T)  
set.seed(2)  
wordcloud(  
  words = Australia$Cuisines,  
  min.freq = 1,  
  max.words = 10,  
  random.order = FALSE,  
  rot.per = 0.35,  
  colors = brewer.pal(8, "Set2")  
)
```



Brazil

```
Brazil <- RestDF1 %>%  
  group_by(Country) %>%  
  filter(any(Country == "Brazil")) %>%  
  count(Cuisines, sort = T)  
set.seed(3)  
wordcloud(  
  words = Brazil$Cuisines,  
  min.freq = 1,  
  max.words = 10,  
  random.order = FALSE,  
  rot.per = 0.35,  
  colors = brewer.pal(8, "Set2")  
)
```



Canada

```
Canada <- RestDF1 %>%  
  group_by(Country) %>%  
  filter(any(Country == "Canada")) %>%  
  count(Cuisines, sort = T)  
set.seed(4)  
wordcloud(  
  words = Canada$Cuisines,  
  min.freq = 1,  
  max.words = 10,  
  random.order = FALSE,  
  rot.per = 0.35,  
  colors = brewer.pal(8, "Set2")  
)
```

sushi
japanese
italian
asian
canadian
pizza
chinese

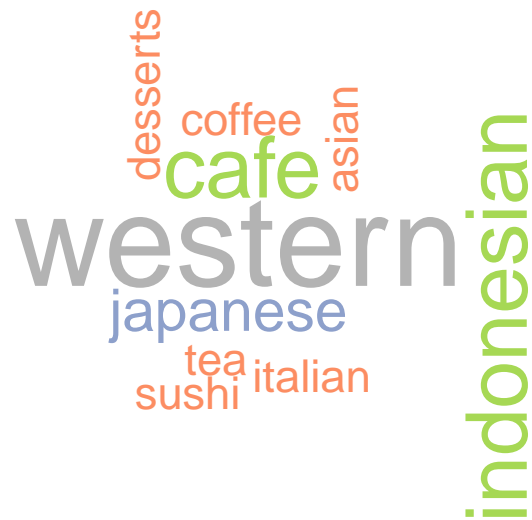
India

```
India <- RestDF1 %>%  
  group_by(Country) %>%  
  filter(any(Country == "India")) %>%  
  count(Cuisines, sort = T)  
set.seed(5)  
wordcloud(  
  words = India$Cuisines,  
  min.freq = 1,  
  max.words = 10,  
  random.order = FALSE,  
  rot.per = 0.35,  
  colors = brewer.pal(8, "Set2")  
)
```



Indonesia

```
Indonesia <- RestDF1 %>%  
  group_by(Country) %>%  
  filter(any(Country == "Indonesia")) %>%  
  count(Cuisines, sort = T)  
set.seed(6)  
wordcloud(  
  words = Indonesia$Cuisines,  
  min.freq = 1,  
  max.words = 10,  
  random.order = FALSE,  
  rot.per = 0.35,  
  colors = brewer.pal(8, "Set2")  
)
```



New Zealand

```
New_Zealand <- RestDF1 %>%  
  group_by(Country) %>%  
  filter(any(Country == "New Zealand")) %>%  
  count(Cuisines, sort = T)  
set.seed(7)  
wordcloud(  
  words = New_Zealand$Cuisines,  
  min.freq = 1,  
  max.words = 10,  
  random.order = FALSE,  
  rot.per = 0.35,  
  colors = brewer.pal(8, "Set2")  
)
```



Phillipines

```
Phillipines <- RestDF1 %>%  
  group_by(Country) %>%  
  filter(any(Country == "Phillipines")) %>%  
  count(Cuisines, sort = T)  
set.seed(8)  
wordcloud(  
  words = Phillipines$Cuisines,  
  min.freq = 1,  
  max.words = 10,  
  random.order = FALSE,  
  rot.per = 0.35,  
  colors = brewer.pal(8, "Set2")  
)
```



Qatar

```
Qatar <- RestDF1 %>%  
  group_by(Country) %>%  
  filter(any(Country == "Qatar")) %>%  
  count(Cuisines, sort = T)  
set.seed(9)  
wordcloud(  
  words = Qatar$Cuisines,  
  min.freq = 1,  
  max.words = 10,  
  random.order = FALSE,  
  rot.per = 0.35,  
  colors = brewer.pal(8, "Set2")  
)
```




Singapore

```
Singapore <- RestDF1 %>%  
  group_by(Country) %>%  
  filter(any(Country == "Singapore")) %>%  
  count(Cuisines, sort = T)  
set.seed(10)  
wordcloud(  
  words = Singapore$Cuisines,  
  min.freq = 1,  
  max.words = 10,  
  random.order = FALSE,  
  rot.per = 0.35,  
  colors = brewer.pal(8, "Set2")  
)
```



South Africa

```
South_Africa <- RestDF1 %>%  
  group_by(Country) %>%  
  filter(any(Country == "South Africa")) %>%  
  count(Cuisines, sort = T)  
set.seed(11)  
wordcloud(  
  words = South_Africa$Cuisines,  
  min.freq = 1,  
  max.words = 10,  
  random.order = FALSE,  
  rot.per = 0.35,  
  colors = brewer.pal(8, "Set2")  
)
```



Sri Lanka

```
Sri_Lanka <- RestDF1 %>%  
  group_by(Country) %>%  
  filter(any(Country == "Sri Lanka")) %>%  
  count(Cuisines, sort = T)  
set.seed(12)  
wordcloud(  
  words = Sri_Lanka$Cuisines,  
  min.freq = 1,  
  max.words = 10,  
  random.order = FALSE,  
  rot.per = 0.35,  
  colors = brewer.pal(8, "Set2")  
)
```



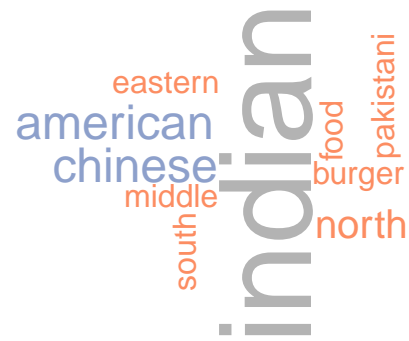
Turkey

```
Turkey <- RestDF1 %>%  
  group_by(Country) %>%  
  filter(any(Country == "Turkey")) %>%  
  count(Cuisines, sort = T)  
set.seed(13)  
wordcloud(  
  words = Turkey$Cuisines,  
  min.freq = 1,  
  max.words = 10,  
  random.order = FALSE,  
  rot.per = 0.35,  
  colors = brewer.pal(8, "Set2")  
)
```



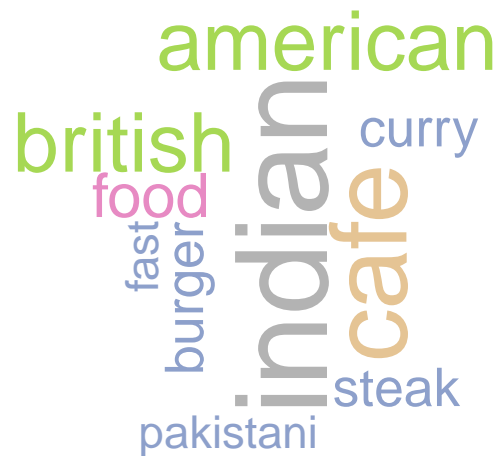
UAE

```
UAE <- RestDF1 %>%  
  group_by(Country) %>%  
  filter(any(Country == "UAE")) %>%  
  count(Cuisines, sort = T)  
set.seed(14)  
wordcloud(  
  words = UAE$Cuisines,  
  min.freq = 1,  
  max.words = 10,  
  random.order = FALSE,  
  rot.per = 0.35,  
  colors = brewer.pal(8, "Set2")  
)
```



UK

```
UK <- RestDF1 %>%
  group_by(Country) %>%
  filter(any(Country == "United Kingdom")) %>%
  count(Cuisines, sort = T)
set.seed(15)
wordcloud(
  words = UK$Cuisines,
  min.freq = 1,
  max.words = 10,
  random.order = FALSE,
  rot.per = 0.35,
  colors = brewer.pal(8, "Set2")
)
```



United States

```
US <- RestDF1 %>%
  group_by(Country) %>%
  filter(any(Country == "United States")) %>%
  count(Cuisines, sort = T)
set.seed(16)
wordcloud(
  words = US$Cuisines,
  min.freq = 1,
  max.words = 10,
  random.order = FALSE,
  rot.per = 0.35,
  colors = brewer.pal(8, "Set2")
)
```



7. What is the maximum and minimum number of cuisines that a restaurant serves?

Also, which is the most served cuisine across the restaurant for each city?

```
count(Cuisine1, 'Restaurant Name', Cuisines.y, sort = T) %>%
  top_n(10, wt = n)
```

```
##      Restaurant Name  Cuisines.y  n
## 1  Cafe Coffee Day      Cafe 83
## 2  Domino's Pizza     Fast Food 79
## 3  Domino's Pizza      Pizza 79
## 4      Subway        American 63
## 5      Subway        Fast Food 63
## 6      Subway Healthy Food 63
## 7      Subway          Salad 63
## 8 Green Chick Chop     Fast Food 51
## 9 Green Chick Chop North Indian 51
## 10 Green Chick Chop    Raw Meats 51
```

```
max(Cuisine1$n)
```

```
## [1] 83
```


The maximum number of cuisines served by a restaurant is 83: Cuisine is Cafe and the franchise name is Cafe Coffee Day

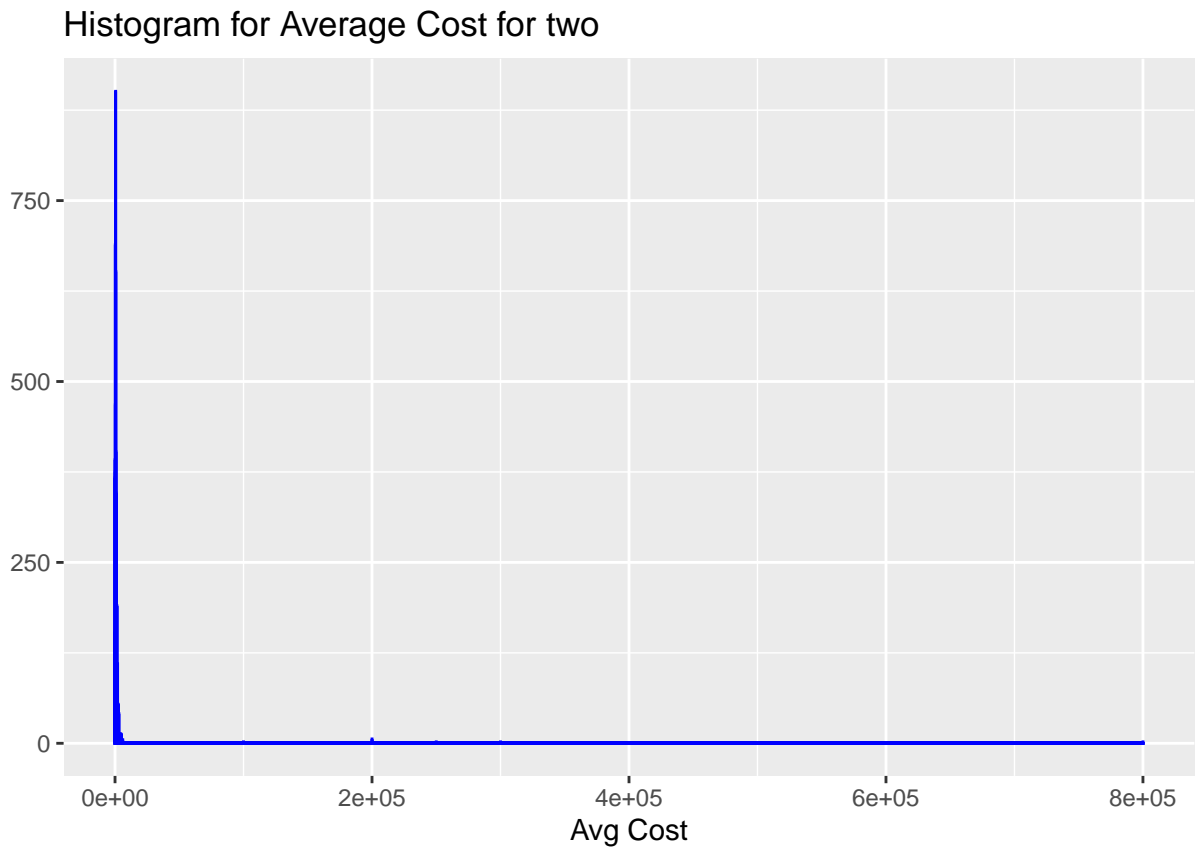
```
min(Cuisine1$n)
```

```
## [1] 1
```

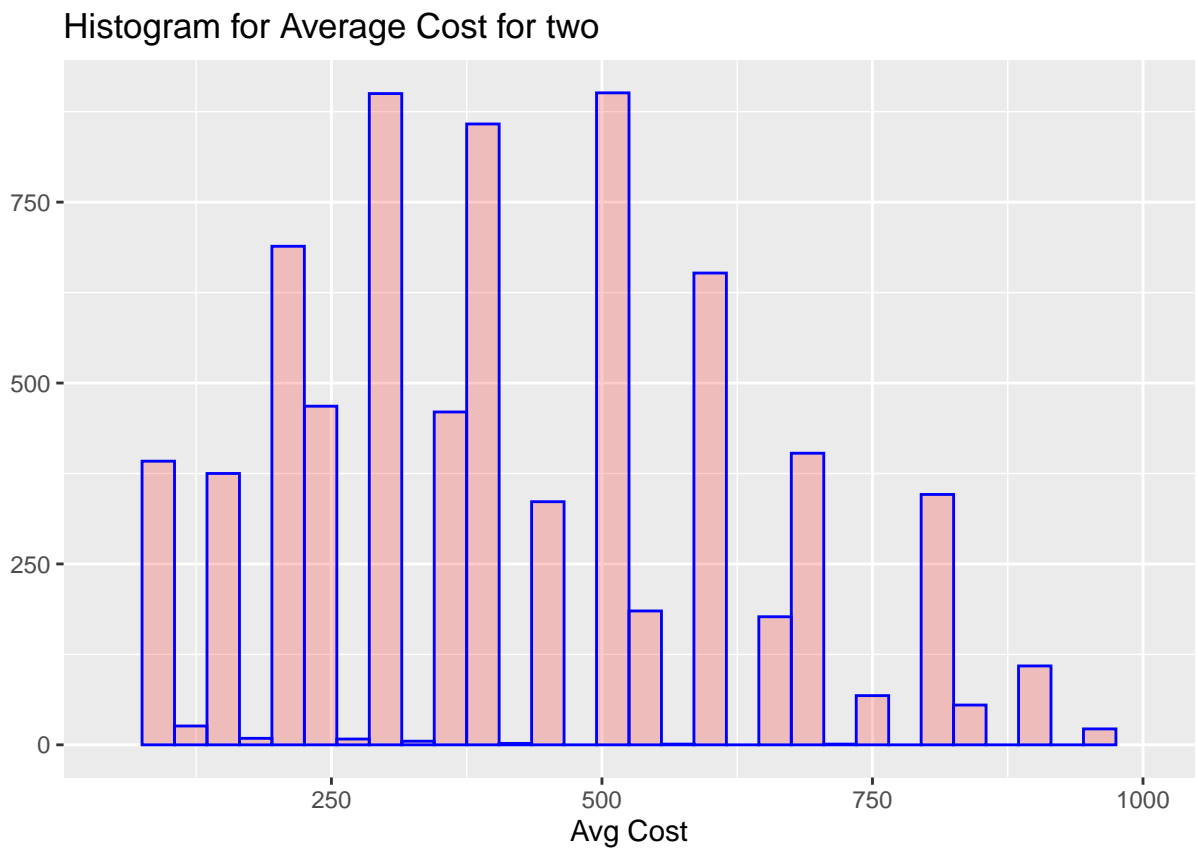
There are multiple restaurants that serve only one cuisine.

8. What is the distribution cost across the restaurants?

```
qplot(  
  RestDF1$`Average Cost for two`,  
  geom = "histogram",  
  binwidth = 30,  
  main = "Histogram for Average Cost for two",  
  xlab = "Avg Cost",  
  fill = I("red"),  
  col = I("blue"),  
  alpha = I(.2)  
)
```



```
qplot(
  RestDF1$`Average Cost for two`,
  geom = "histogram",
  binwidth = 30,
  main = "Histogram for Average Cost for two",
  xlab = "Avg Cost",
  fill = I("red"),
  col = I("blue"),
  alpha = I(.2),
  xlim = c(50, 1000)
)
```



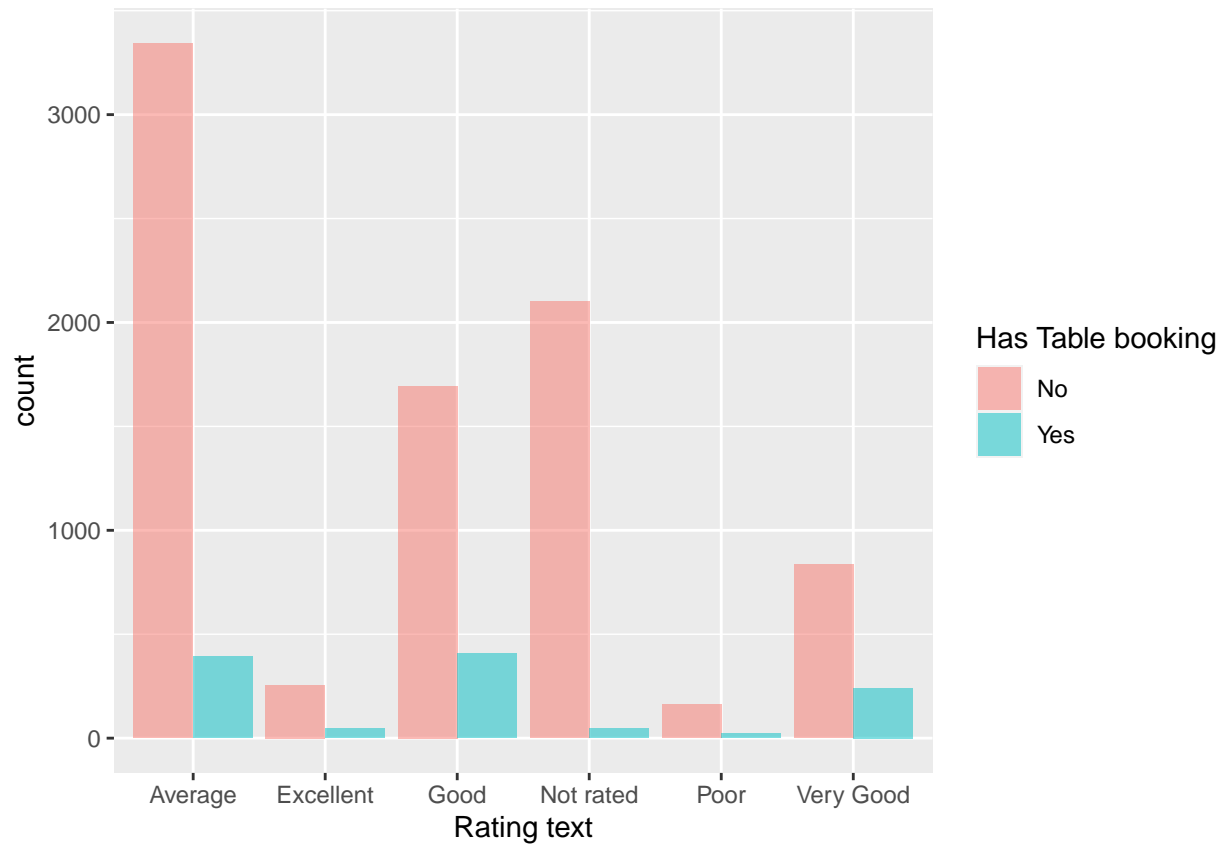
From the above histograms we can observe that mostly the costs are between 250 and 550

9. How ratings are distributed among the various factors?

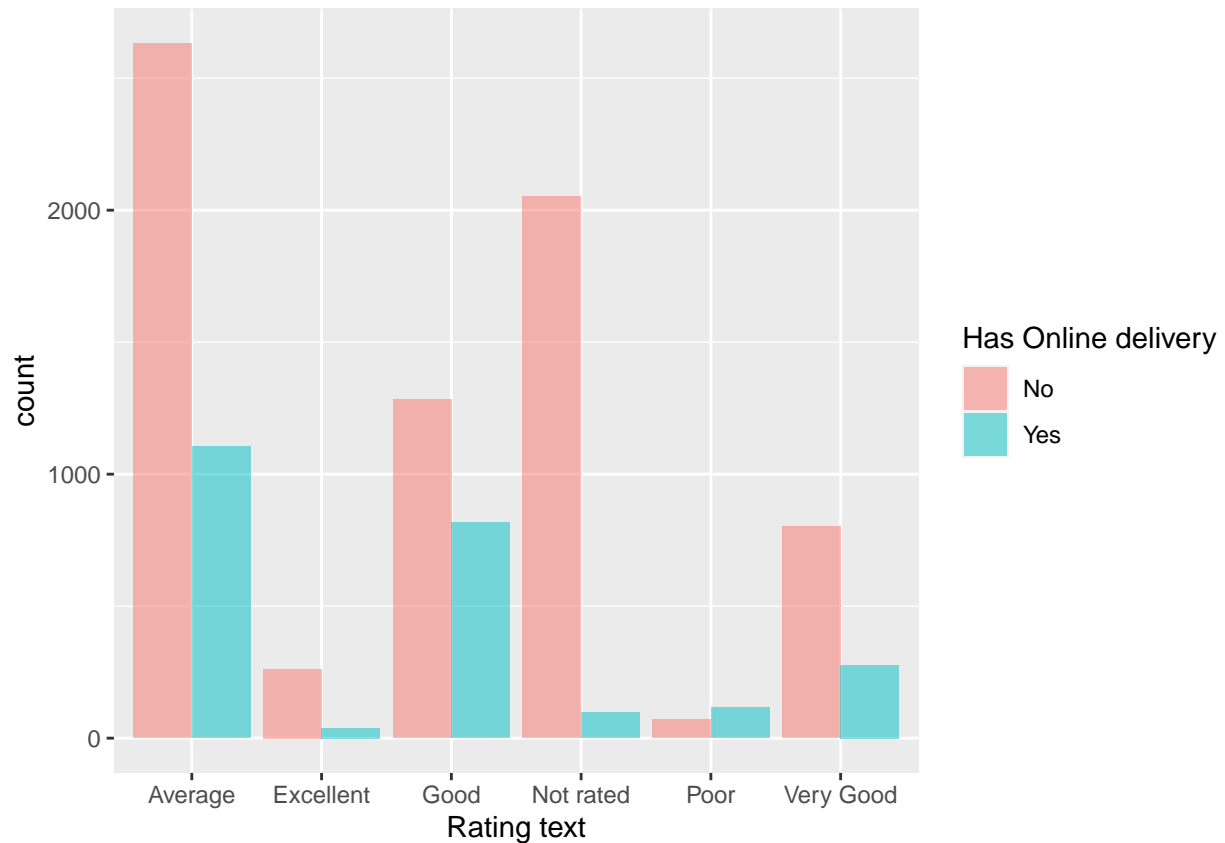
10.Explain the factors in the data that may have an effect on ratings.

For example, number of cuisines, cost, delivery option, etc.

```
ggplot(RestDF1, aes(x = `Rating text`, fill = `Has Table booking`)) +
  geom_histogram(alpha = 0.5,
    position = "dodge",
    stat = "count")
```



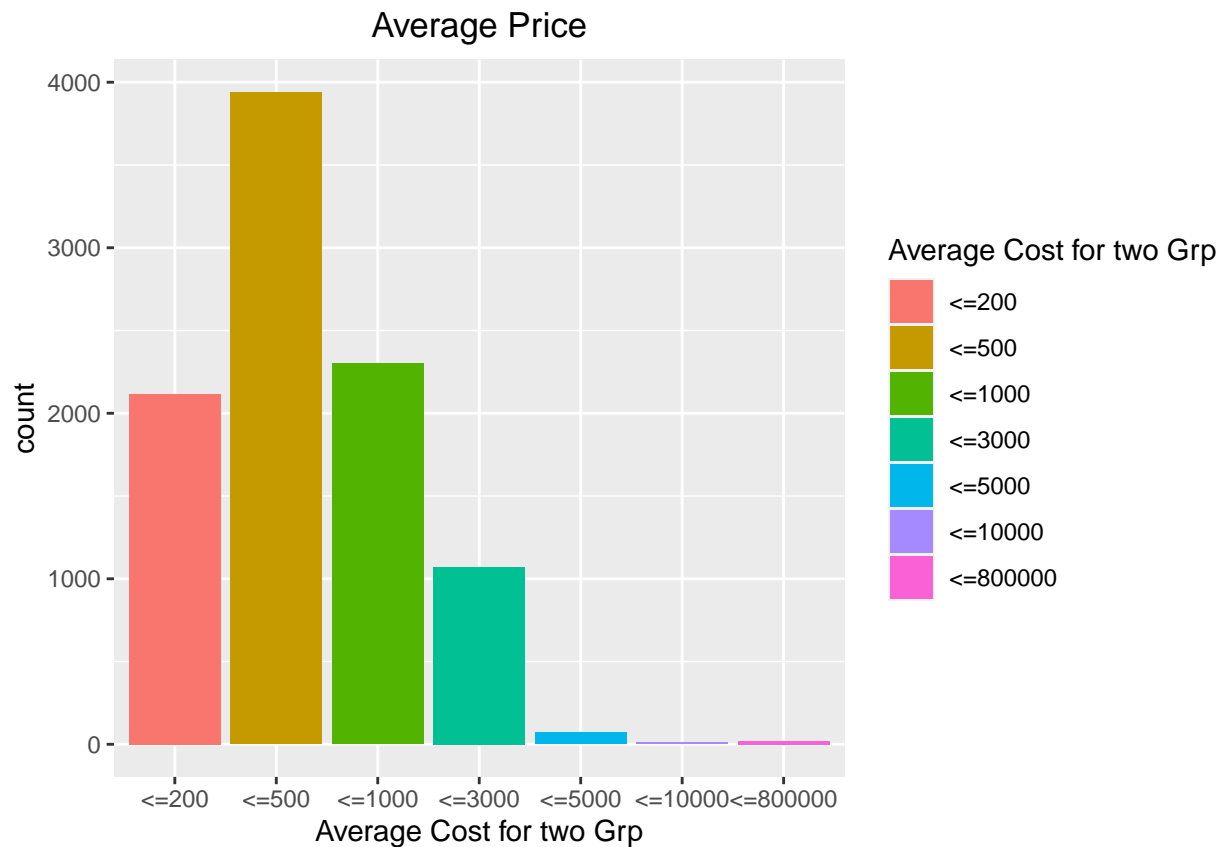
```
ggplot(RestDF1, aes(x = 'Rating text', fill = 'Has Online delivery')) +  
  geom_histogram(alpha = 0.5,  
    position = "dodge",  
    stat = "count")
```



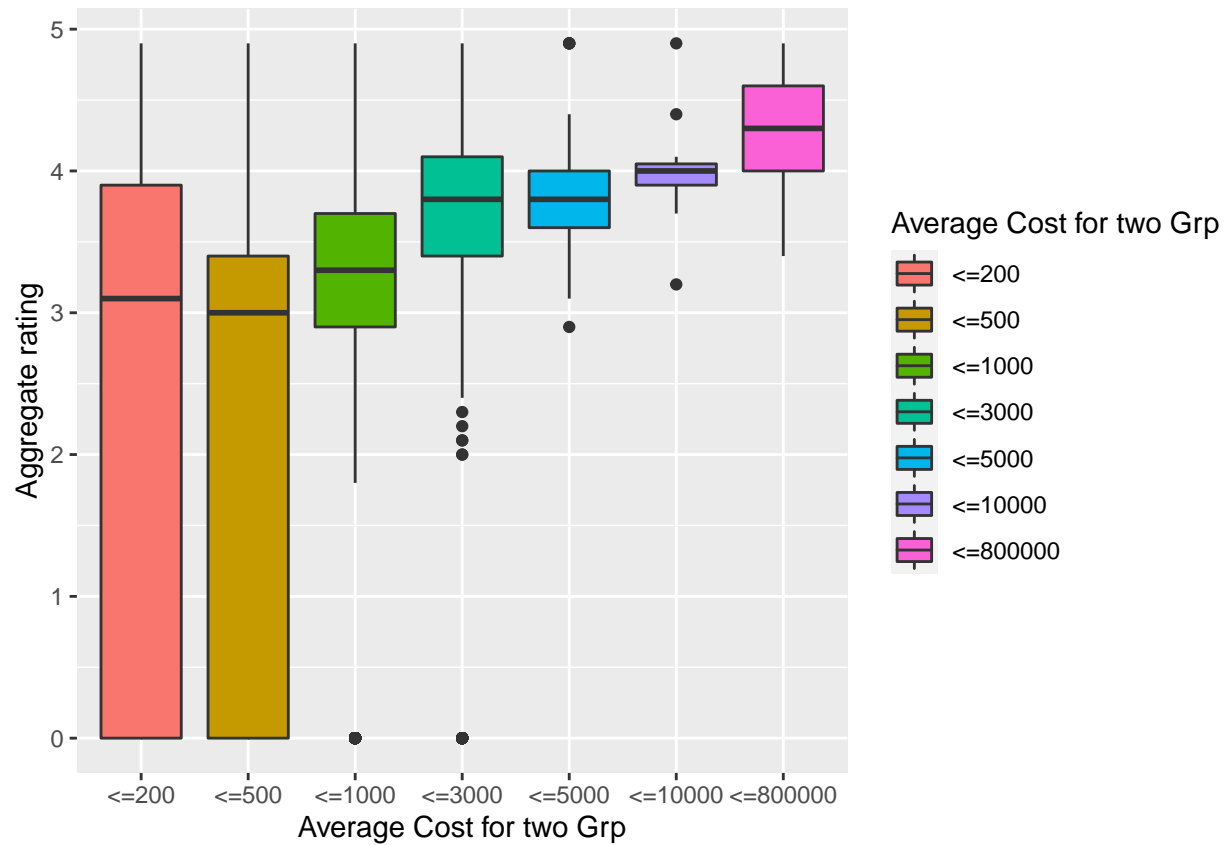
Price Count & Rating Distribution

```
RestDF1$`Average Cost for two Grp` <-
  cut(
    RestDF1$`Average Cost for two`,
    c(0, 200, 500, 1000, 3000, 5000, 10000, 800000),
    labels = c(
      "<=200",
      "<=500",
      "<=1000",
      "<=3000",
      "<=5000",
      "<=10000",
      "<=800000"
    )
  )

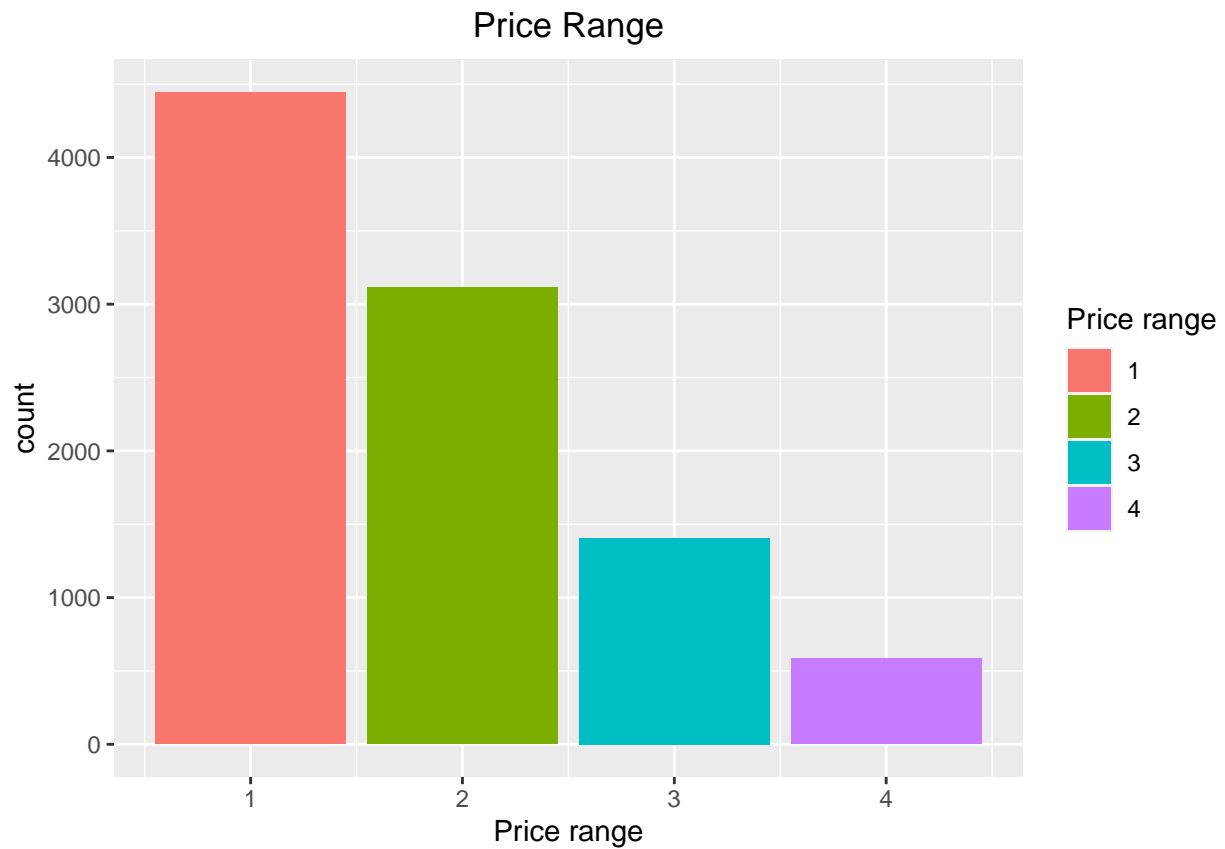
RestDF1 %>%
  filter(!is.na(`Average Cost for two Grp`)) %>%
  ggplot() +
  geom_bar(aes(x = `Average Cost for two Grp`, fill = `Average Cost for two Grp`),
    stat = "count") +
  labs(title = "Average Price") +
  theme(plot.title = element_text(hjust = 0.5))
```



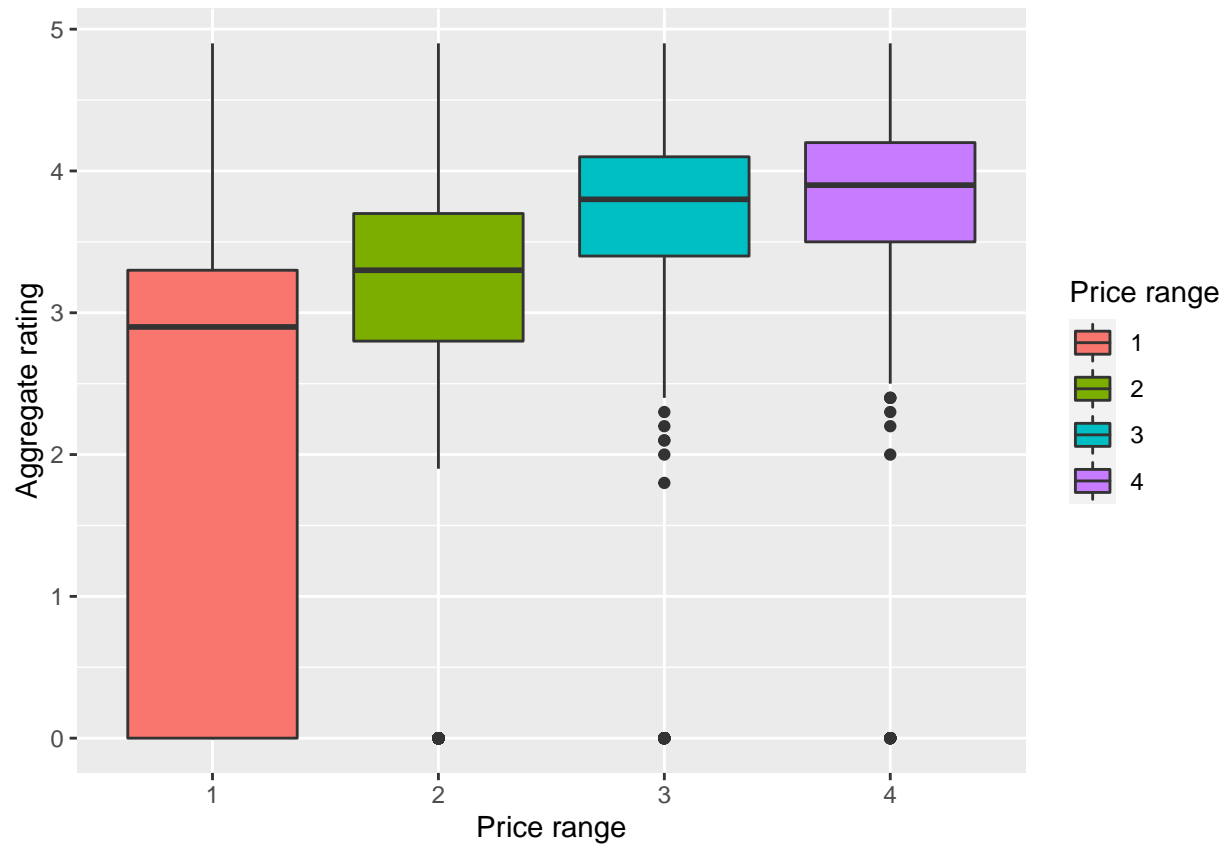
```
RestDF1 %>%
  filter(!is.na('Average Cost for two Grp')) %>%
  ggplot() +
  geom_boxplot(aes(x = 'Average Cost for two Grp', y = 'Aggregate rating',
                  fill = 'Average Cost for two Grp')) +
  theme(plot.title = element_text(hjust = 0.5))
```



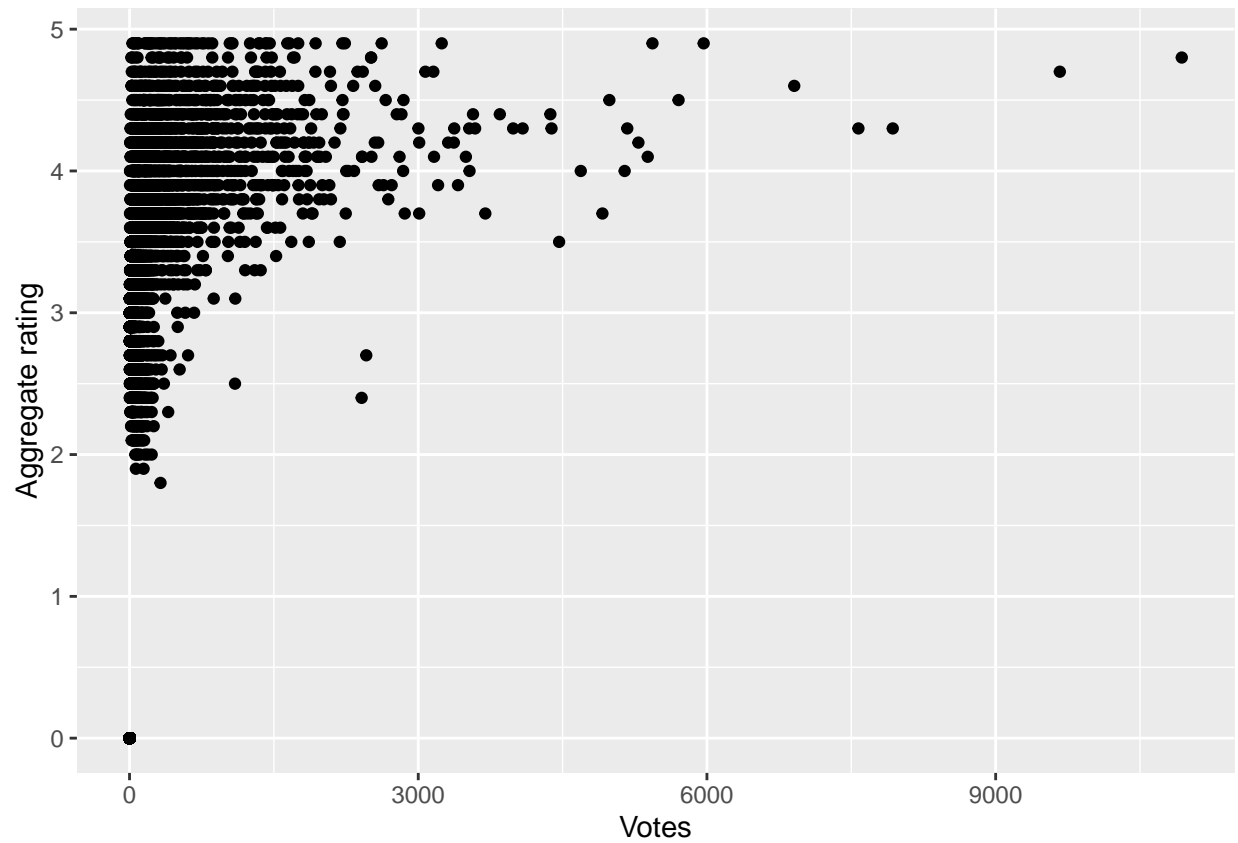
```
RestDF1 %>%
  ggplot() +
  geom_bar(aes(x = 'Price range', fill = as.factor('Price range')),
    stat = "count") +
  labs(title = "Price Range") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_fill_discrete(name = "Price range")
```



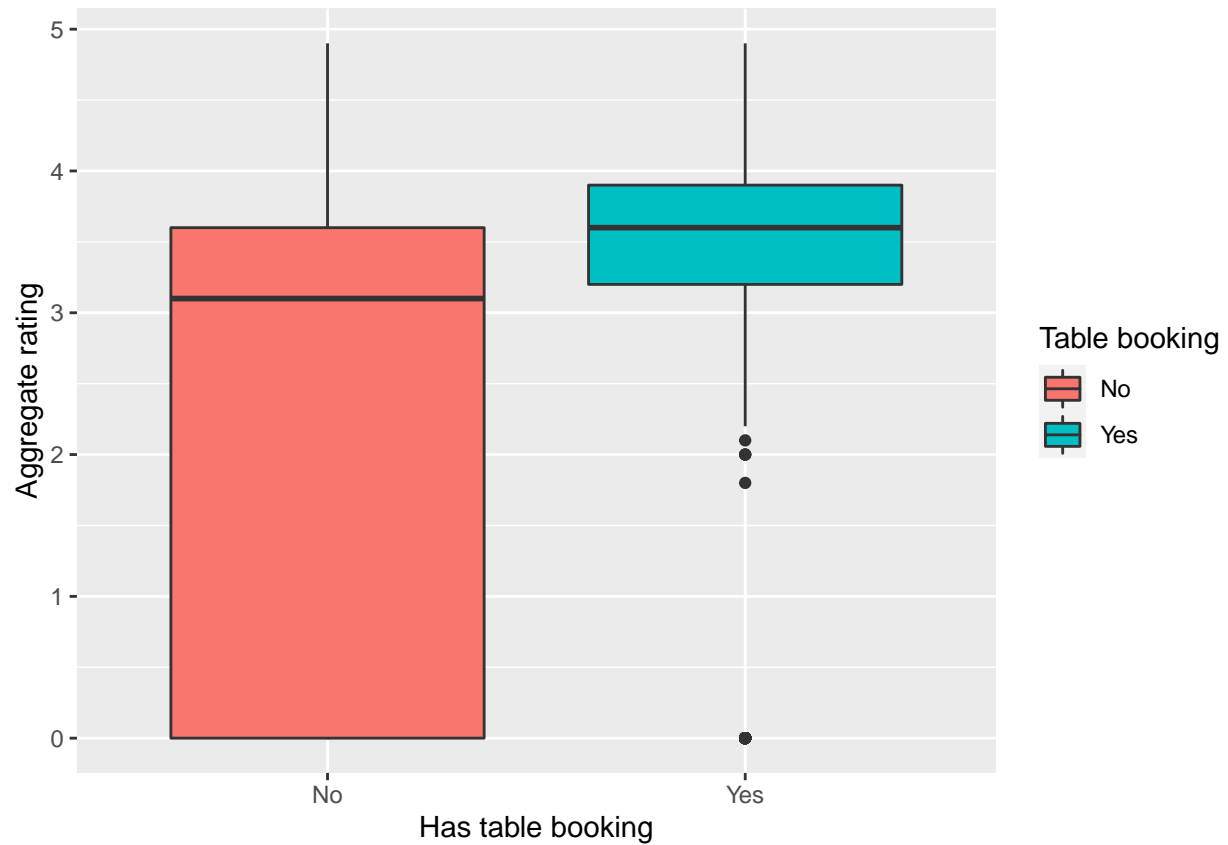
```
RestDF1 %>%  
  ggplot() +  
  geom_boxplot(aes(  
    x = as.factor('Price range'),  
    y = 'Aggregate rating',  
    fill = as.factor('Price range')  
  )) +  
  theme(plot.title = element_text(hjust = 0.5)) +  
  scale_fill_discrete(name = "Price range") +  
  xlab("Price range")
```



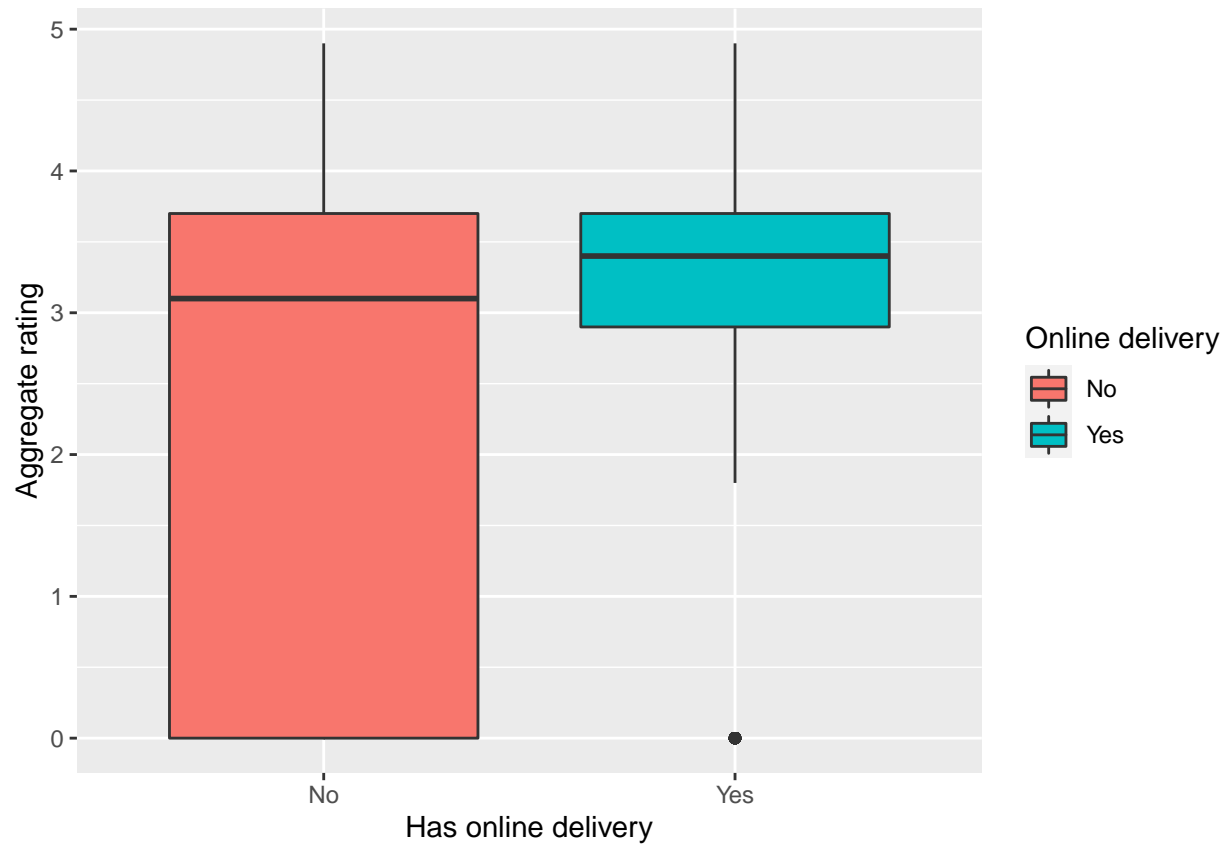
```
RestDF1 %>%  
  ggplot() +  
  geom_point(aes(x = Votes, y = 'Aggregate rating'))
```

```
RestDF1 %>%
  ggplot() +
  geom_boxplot(aes(
    x = 'Has Table booking',
    y = 'Aggregate rating',
    fill = as.factor('Has Table booking')
  )) +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_fill_discrete(name = "Table booking") +
  xlab("Has table booking")
```



```
RestDF1 %>%  
  ggplot() +  
  geom_boxplot(aes(  
    x = 'Has Online delivery',  
    y = 'Aggregate rating',  
    fill = as.factor('Has Online delivery')  
  )) +  
  theme(plot.title = element_text(hjust = 0.5)) +  
  scale_fill_discrete(name = "Online delivery") +  
  xlab("Has online delivery")
```



We can observe that there is no single variable that affects the rating in a strong way however average cost for two, price range, number of votes, table booking and online delivery are affecting in part the rating of a restaurant.