# Identifying and Recommending Best Restaurants

# Software Used:

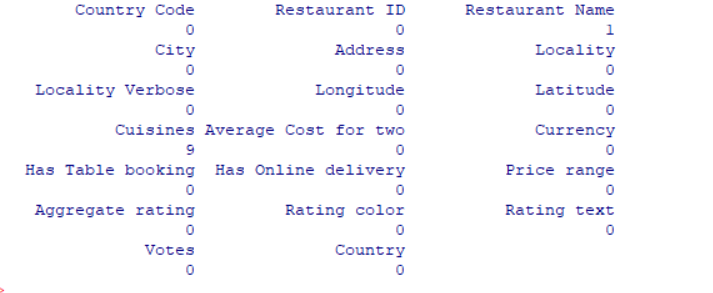
**R Studio**

## Data Preliminary analysis:

1. Dealing with **missing data:**

Only 1 missing value in the Restaurant Name column.

9 missing values in the Cuisines column.



Before Removing row count: 

After Removing row count:

1. Dealing with **duplicates**:

No duplicates Found in the ‘data’ dataset.



1. Replace cities with ambiguous names: 44) 
2. Remove Columns names spacing to prevent Error in R:



1. Unify **Currencies** into **Indian Rupel** to make it more efficient to visualize big numbers:

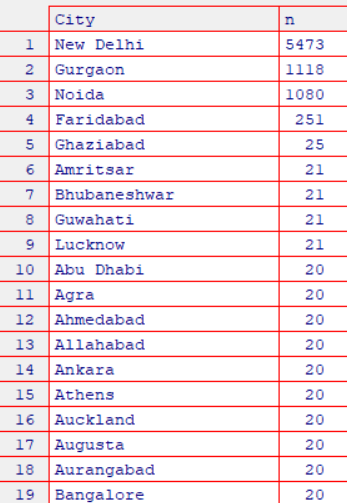


I downloaded necessary packages and imported them to use them in R:

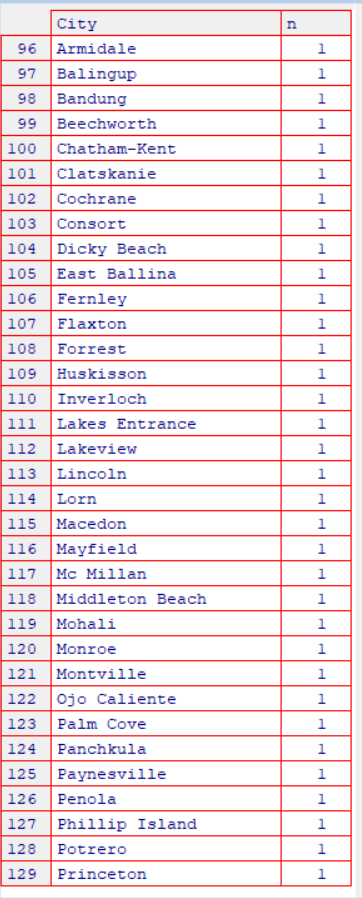
**dpylr** for data manipulation, **ggplot2** for visualization, **readxl** to read xlsx files in R.

1. **Maximum / minimum** number of restaurants by **Cities**.

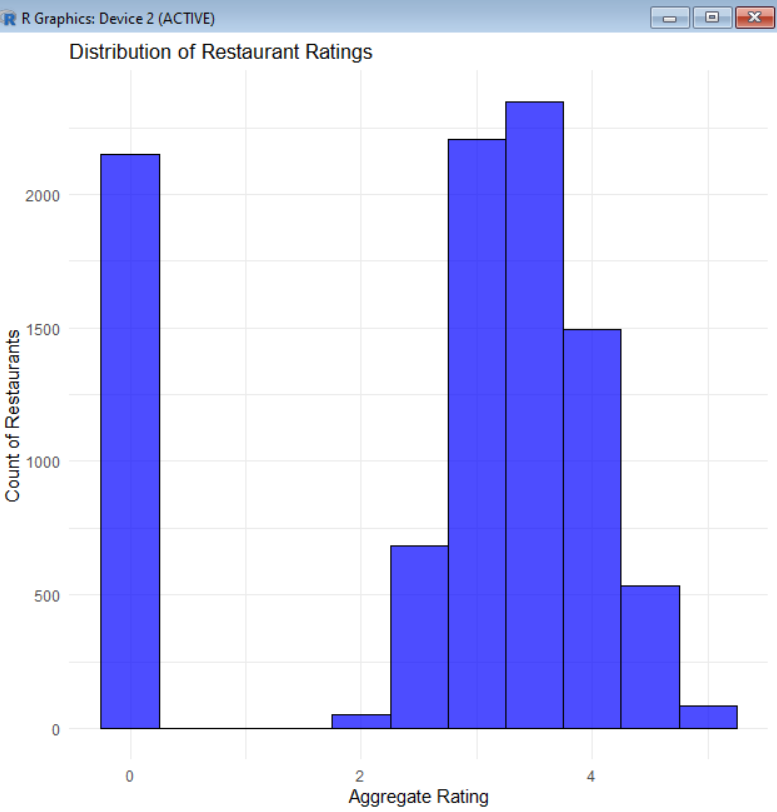
Max number of restuarants:



Least Numbers of restaurants by cities:

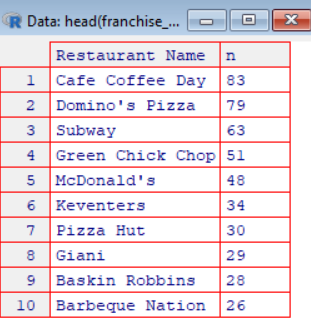


1. Distribution of **ratings:**



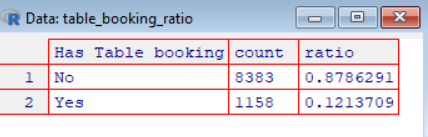
Most restaurants have ratings **around 3.2**, meaning the average experience is good. The **slightly lower average** suggests some restaurants have **very low ratings**. Ratings vary widely, ranging from **0 to 4.9**, showing differences in customer satisfaction.

1. **Franchise** with most national presence:



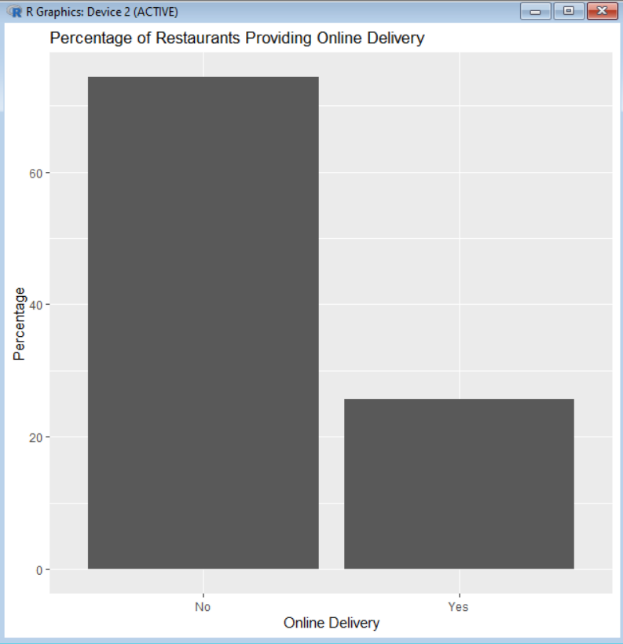
Top 10 Franchise

1. **Ratio** for restaurants with **booking vs not having booking:**



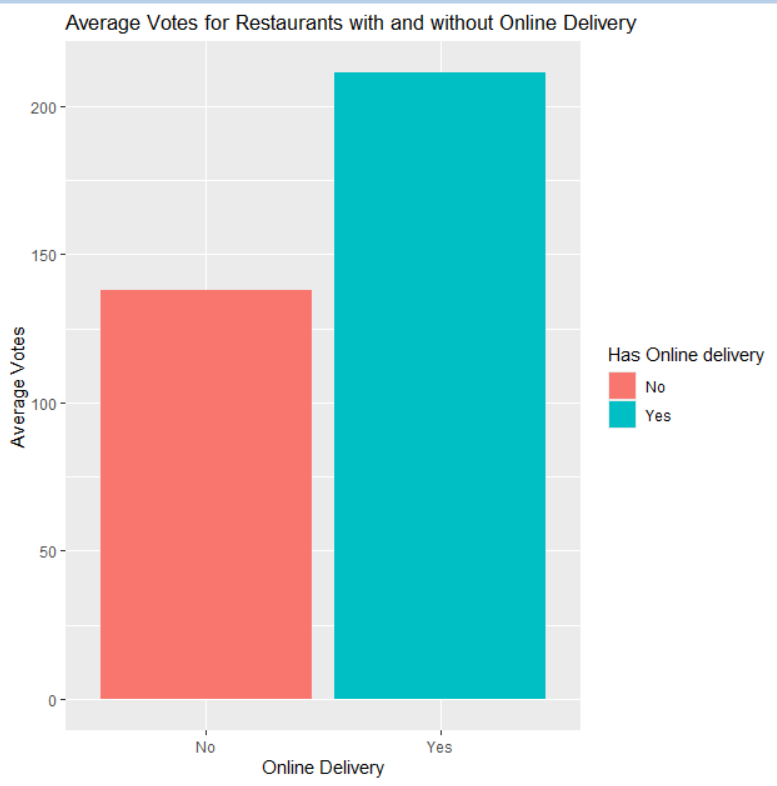
Ratio of restaurants that has no booking is 87% and 12% for restaurants that has booking.

1. **Percentage** of restaurants providing **online delivery:**



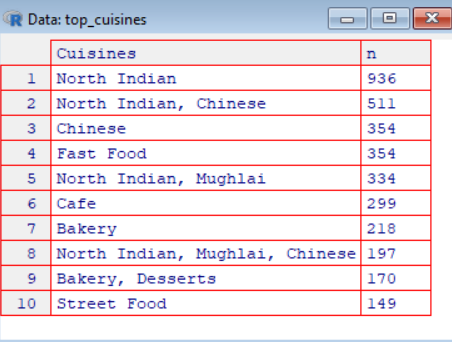
Percentage of restaurants with delivery is **less** than restaurants without online delivery.

1. **Average votes** based on service online delivery:



Average votes are higher in customers with online delivery which shows that online delivery has positive effects on votes.

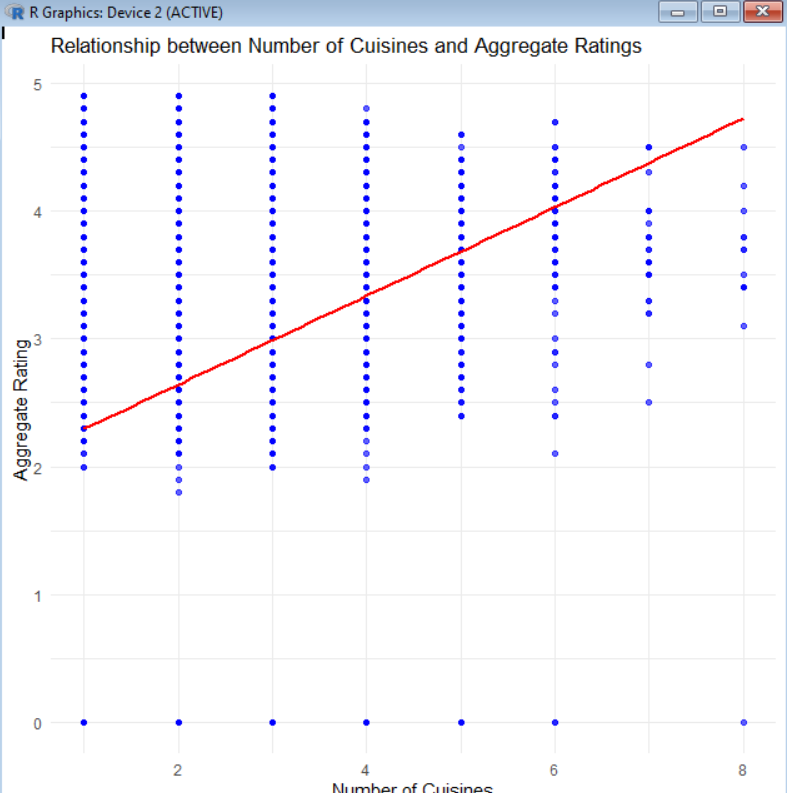
1. **Top 10 cuisines** served across cities:



1. **Maximum and Minimum no.** of cuisines that a restaurant serves:

Maximum number of cuisines served by a restaurant: 8

Minimum number of cuisines served by a restaurant: 1



Correlation between Number of Cuisines and Aggregate Rating: 0.2501937 **a weak positive relationship.**

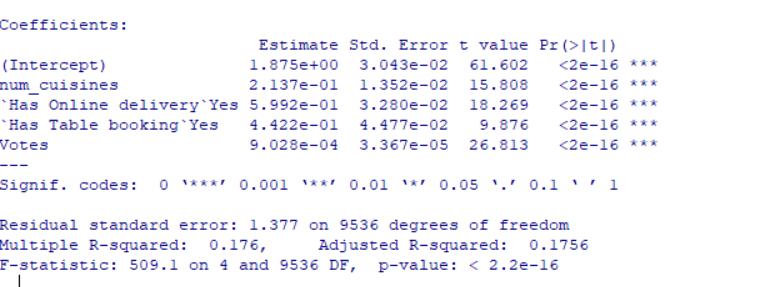
1. **Cost vs other variables:**

Correlation between Cost and Rating: 0.05186896 **the cost has little to no effect on the ratings.**

Correlation between Cost and Number of Cuisines: 0.01328691 **This is an even weaker positive correlation.**

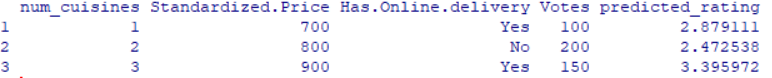
1. **Effect on Ratings:**

The number of cuisines (num\_cuisines), online delivery availability (Has Online delivery), table booking (Has Table booking), and the number of votes are significant columns of aggregate ratings, with p-values well below 0.05.

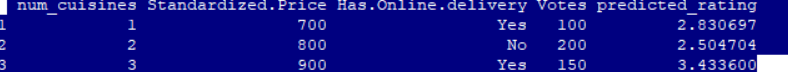


**F-statistic:** 509.1, with a p-value < 2.2e-16, suggests that the overall model is highly significant.

1. **Predicting rating:**



1. Prediction based on Polynomial Regression



1. **Export data into xlsx file cleaned\_data.**xlsx to work on Tableau.