Analysis of Zomato Dataset

Zomato is an Indian restaurant aggregator and food delivery start-up founded in 2008. Zomato provides information, menus and user-reviews of restaurants as well as food delivery options from partner restaurants in selected cities.

# Dataset

The dataset used for this project was found on Kaggle. Zomato API Analysis is one of the most useful analyses for foodies who want to taste the best cuisines of every part of the world which lies in their budget. This analysis is also for those who want to find the value for money restaurants in various parts of the country for the cuisines. Additionally, this analysis caters the needs of people who are striving to get the best cuisine of the country and which locality of that country serves those cuisines with maximum number of restaurants.

Tools & Libraries

• Python • Jupyter Notebook • Pandas • Numpy • Seaborn • Matplotlib • Plotly & Cufflinks

Data Storage

The collected data has been stored in the Comma Separated Value file Zomato.csv. Each restaurant in the dataset is uniquely identified by its Restaurant Id. Every Restaurant contains the following variables:

• Restaurant Id: Unique id of every restaurant across various cities world

• Restaurant Name: Name of the restaurant

• Country Code: Country in which restaurant is located

• City: City in which restaurant is located

• Address: Address of the restaurant

• Locality: Location in the city

• Locality Verbose: Detailed description of the locality

• Latitude: Latitude coordinate of the restaurant's location

• Cuisines: Cuisines offered by the restaurant

• Average Cost for two: Cost for two people in different currencies

• Currency: Currency of the country

• Has Table booking: yes/no

• Has Online delivery: yes/ no

• Is delivering: yes/ no

• Switch to order menu: yes/no

• Price range: range of price of food

• Aggregate Rating: Average rating out of 5

• Rating colour: depending upon the average rating colour

• Rating text: text on the basis of rating of rating

• Votes: Number of ratings casted by people

## Data Cleaning

I made the following changes and created the following variables:

• Deleted the columns URL, address and phone as they were not important for analysis

• Cleaned the rate column and converted it into the numeric data type

• Heatmap to see missing values

EDA

I looked at the different-different trends of the data and below is a few highlights of the

analysis.

• Which country has maximum records

• Zomato Ratings

• which country of customers has maximum numbers of zero ratings

• which country has online delivery options

• cities distribution

• top 10 Cuisines