METHODOLOGY

Name: Zorawar Singh Mahal

1) Web Scraping:

I have used the beautifulsoup library of python for scraping the data out of the following three restaurants on zomato:

a) https://www.zomato.com/bangalore/andhra-gunpowder-brigade-road-bangalore/order

b) https://www.zomato.com/bangalore/ubq-by-barbeque-nation-1-indiranagar-bangalore/order

c) https://www.zomato.com/bangalore/box8-desi-meals-indiranagar-bangalore/order

I extracted the following things: Restaurant name, Restaurant area, Item name, Item price, Item rating, Number of ratings, Item description and Bestseller tags; appended them in lists and then made a dataframe. Then exported the dataframe in csv format.

2) Data Cleaning and Analysis:

I applied the following steps:

a) Merged the three datasets together in a dataframe.

b) Checked for null values. Filled the null values of 'Item description' column with their respective 'Item names'. Replaced the nan values in the Bestseller tag with 0 and 'Must try' tags with 1.

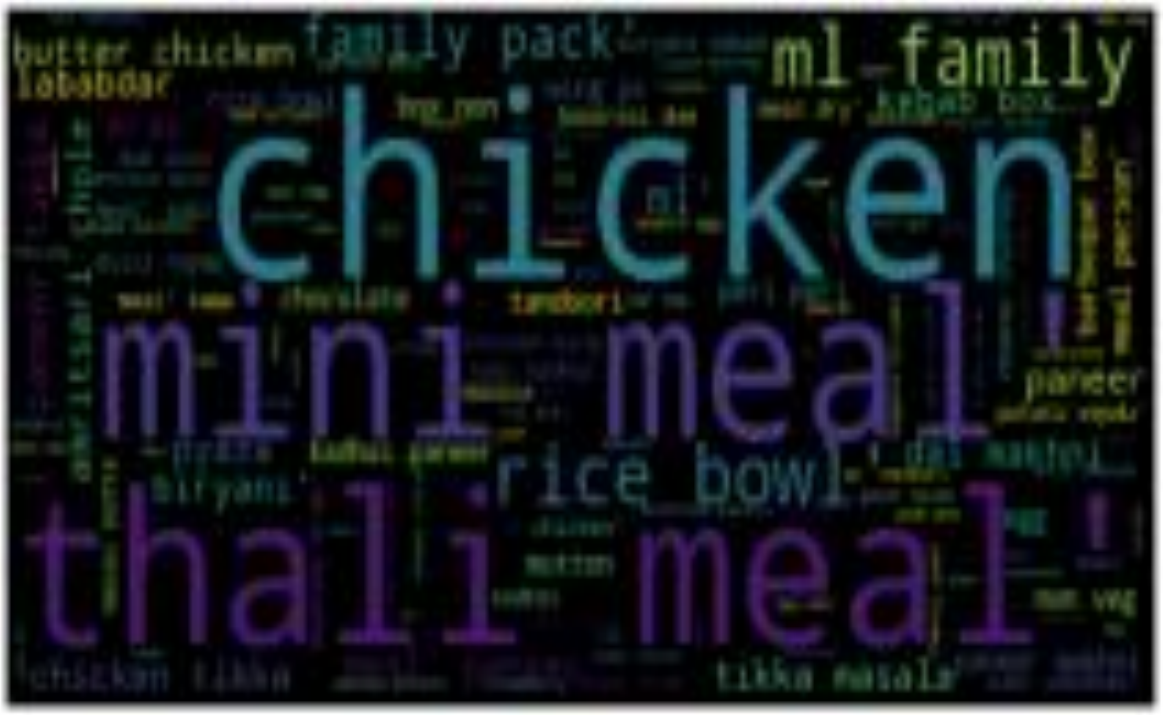
c) Replaced the 'Rating' values with 0 (for nan values), 1 (for ratings lesser than 3.5) and 2 (for ratings greater than 3.5). Replaced the null values in the 'No of Ratings' column with 0.

d) Removed the unnecessary symbols and words from the 'Item price' and 'No of Ratings' columns. Then convert them into float values.

e) Preprocessed the Item name and Item description columns using NLP. Removed the unwanted symbols and digits, made everything into lower cases, removed the stopwords.

f) Used the wordcloud to find out the frequently used words in each column. For eg,

this wordcloud is for the Item names column.



3) Visualization using Power BI:

Created a power bi dashboard for the restaurants.