

# METHODOLOGY

**Name: Sonal Singh**

## 1) Web Scrapping:

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.

