



Big Data Course Project

Phase 2: Proposal

Team 17

Team members:-

Name	Sec	BN
Michael Aziz Faheem	2	10
Mostafa Magdy Abdelrahman	2	25
Mostafa Ahmed Sobhy	2	24
Mostafa Mahmoud Kamal	2	27

Brief problem description

- **Zomato Bangalore Restaurants**

Bengaluru being one such city has more than 12,000 restaurants with restaurants serving dishes from all over the world. And we have rate and votes also some knowledge of the restaurants.

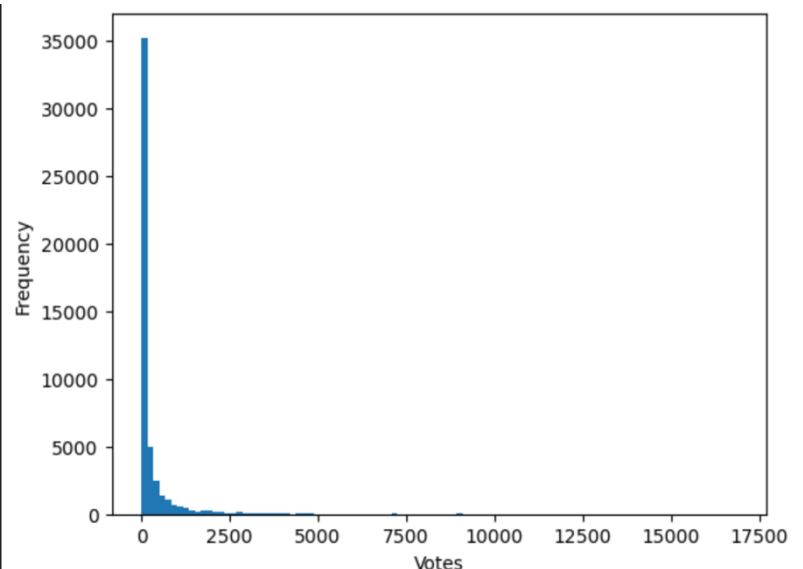
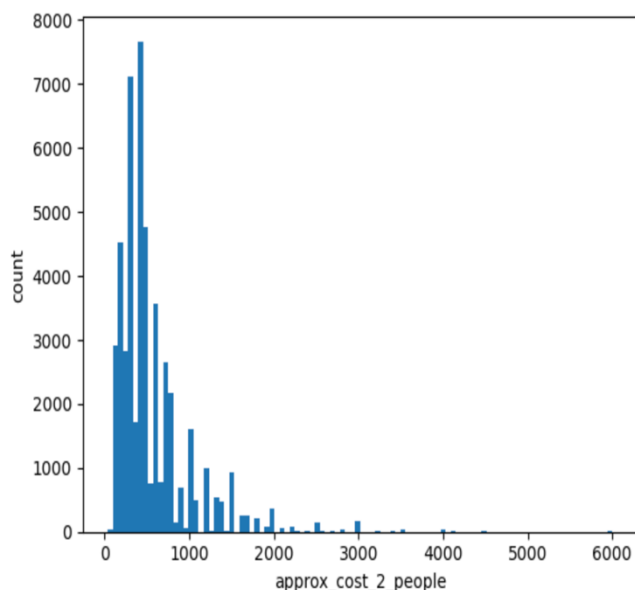
The data is actually just for EDA but we used some suitable features to predict if the rate if restaurant more or less than 3.7.

Project pipeline

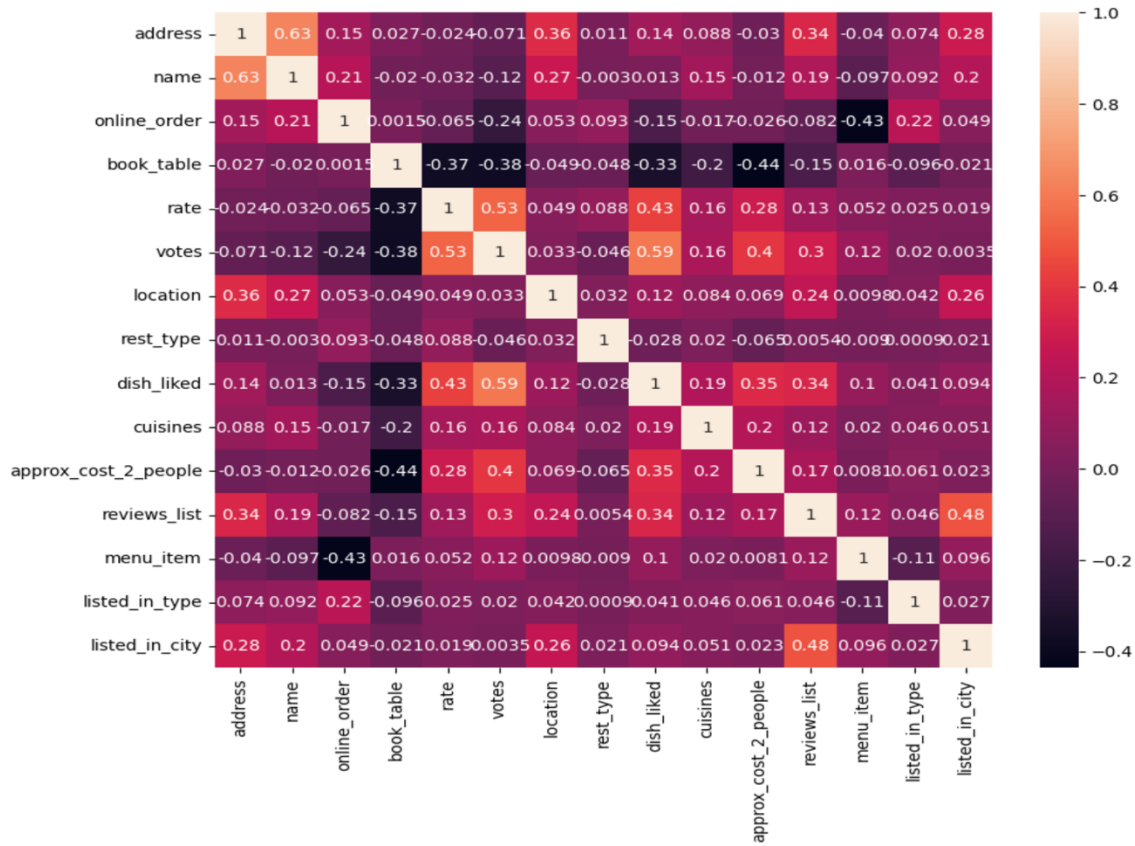
- Reading data.
- Clean and preprocess the data.
- Data visualization and extract info from the data and correlation
- Model naïve bayes without map-reduce
- Model naïve bayes with map-reduce

Analysis and solution of the problem

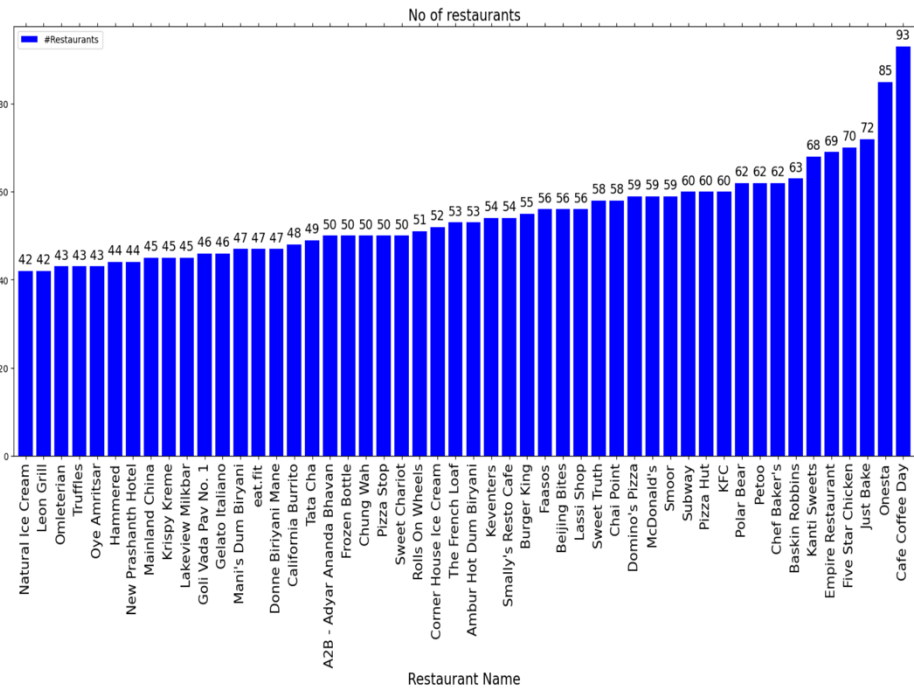
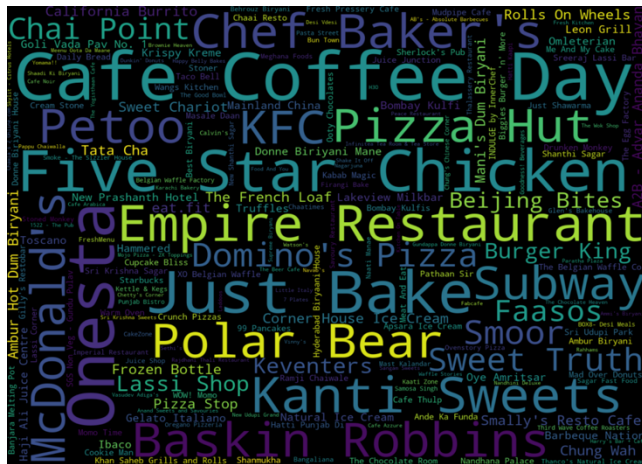
1. Data preprocessing:
 - a. We convert rate from string that have scale like ("3.1/5") to float that we can deal with like 3.1
 - b. Drop all columns expect important ones (votes , approximate cost for 2 persons)
 - c. Rename columns to be able to deal with
 - d. Remove "nan" and "-" from features
 - e. We convert rate that if we have rate more than 3.7 output 1 and if less than 3.7 output 0
2. Data visualization and Extracting insights from data (by coding) :
 - a. See the distribution of features to detect the type of naïve bayes that will use.



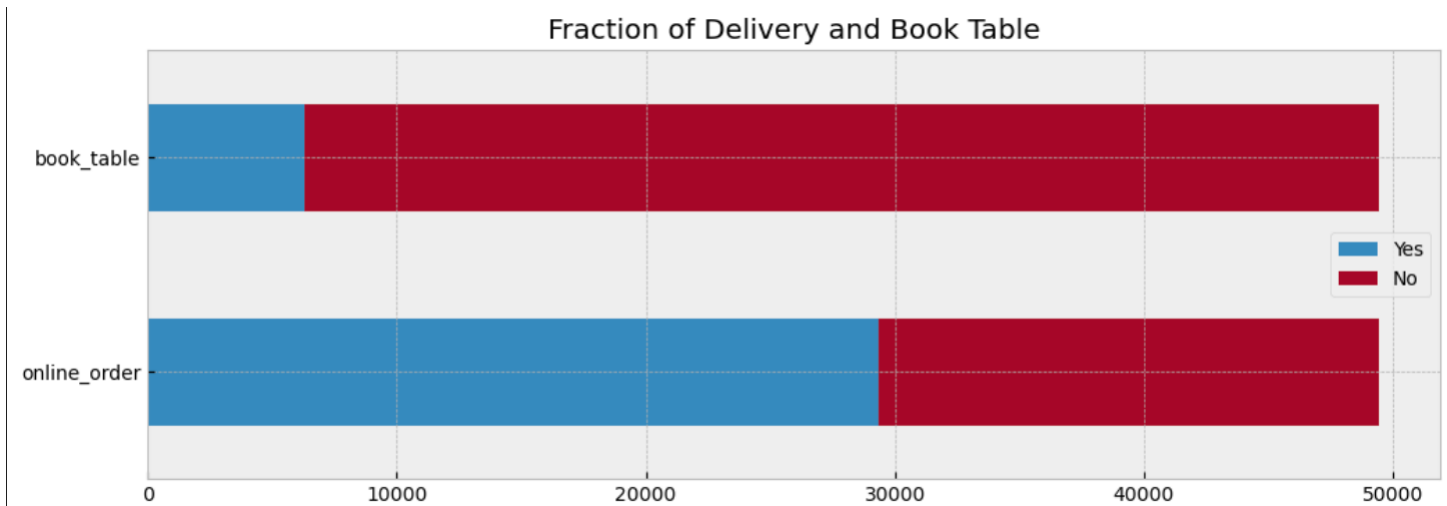
b. See the correlation of features to get the most important features that affect the target



c. Show the most restaurant repeated in the data



d. Show that if book table in the restaurant is provided in the most restaurant



3. Model/Classifier training :

- We trained naïve bayes with map-reduce we built a class for NB and create model from it then use it map function to count the class that we have from numbers then at reduce function we collect and add the frequency then predict In the same class
- We try naïve bayes without map reduce ready made from sklearn

Result and Evaluation

Without map reduce we get 83% accuracy

```
Model without map-reduce accuracy: 0.8297596285167987
```

With map reduce we get 84% accuracy

```
Model with map-reduce accuracy = 0.8474251497005988
```

Unsuccessful trials that were not included in the final solution

Firstly we tried multinomial naïve bayes without Laplace smoothing but getting low accuracy

Also we tried to percentile the values of the votes and cost columns but get a very low accuracy

Any Enhancements and future work

Clean and preprocess some other features and use them in the prediction.