# Capstone project 4 Unsupervised

Individual project by Sharath S

#### About the dataset

- This dataset contains information about different restaurants under Zomato.
- There are 2 datasets, Review and metadata.

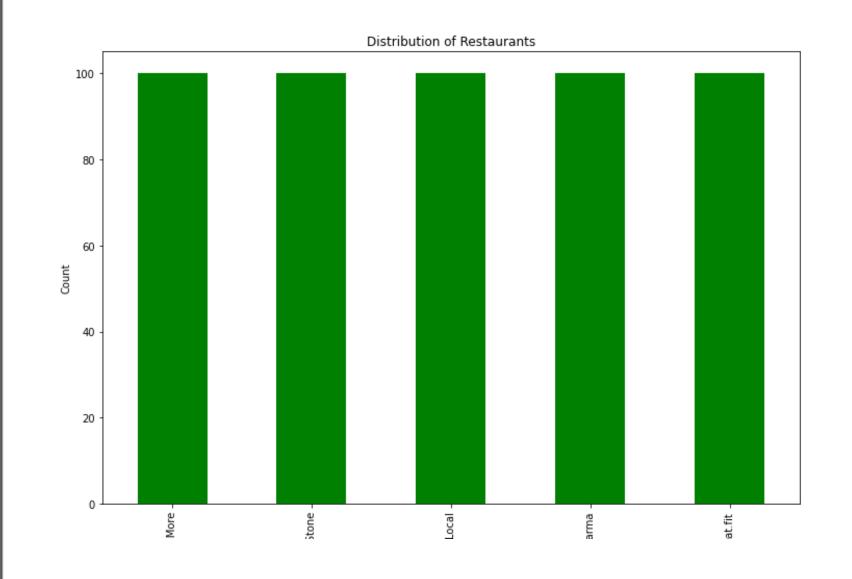




#### Problem Statement

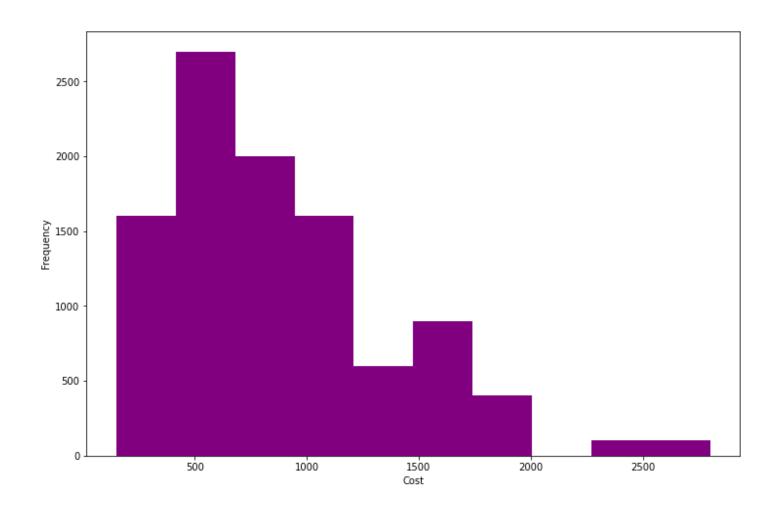
- The objective is to use data about reviews of different restaurants for sentiment analysis and use data about the different types of cuisines and form clusters based on similar attributes to help the business.
- Both sentiment analysis and clustering can help the business.

# Exploratory data analysis

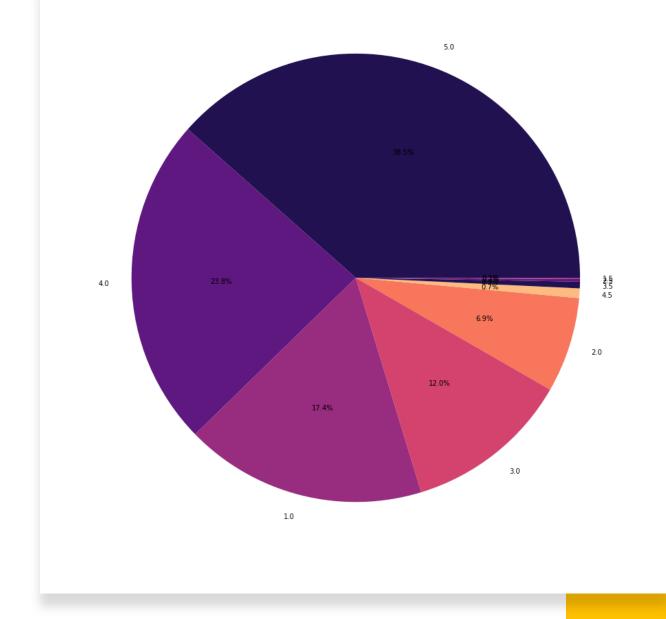


Bar Plot of distribution of restaurants

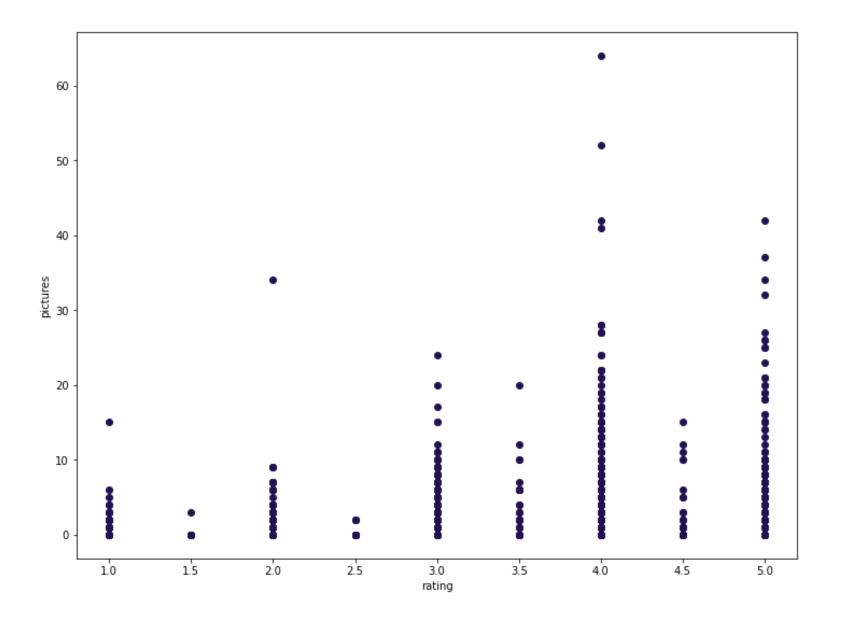
# Histogram of cost



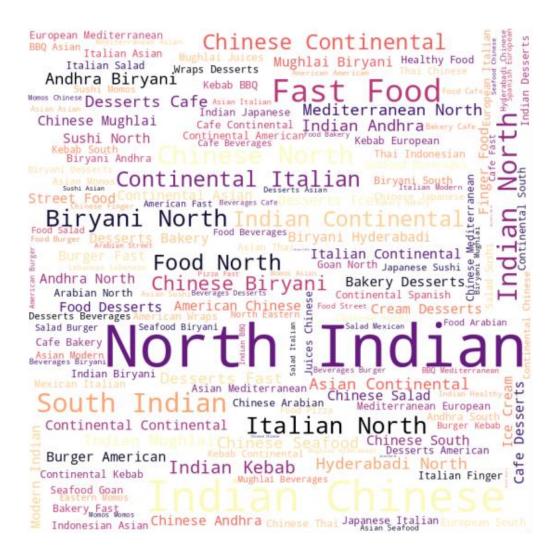
# Pie chart of rating percentage



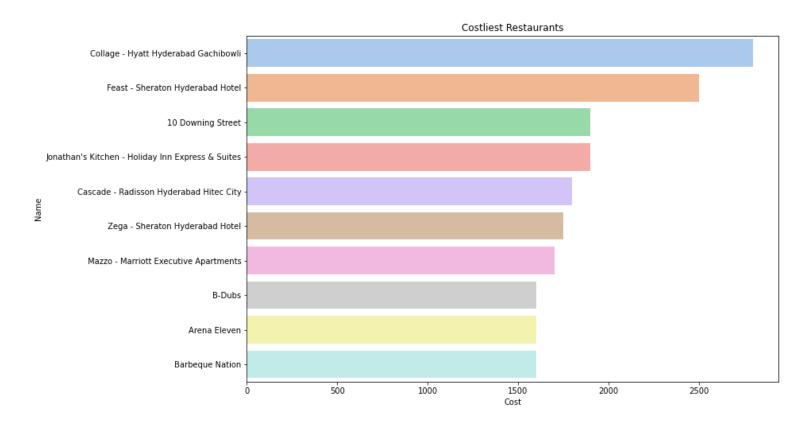
Scatterplot of rating vs pictures



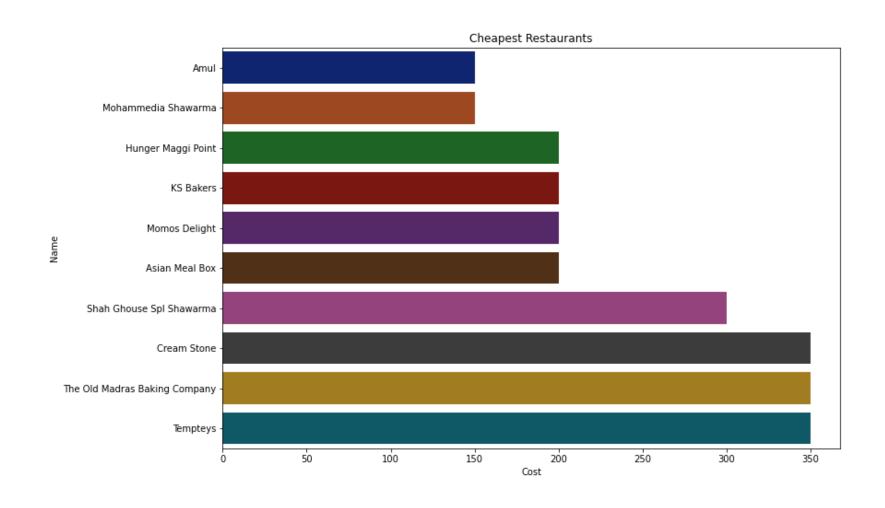




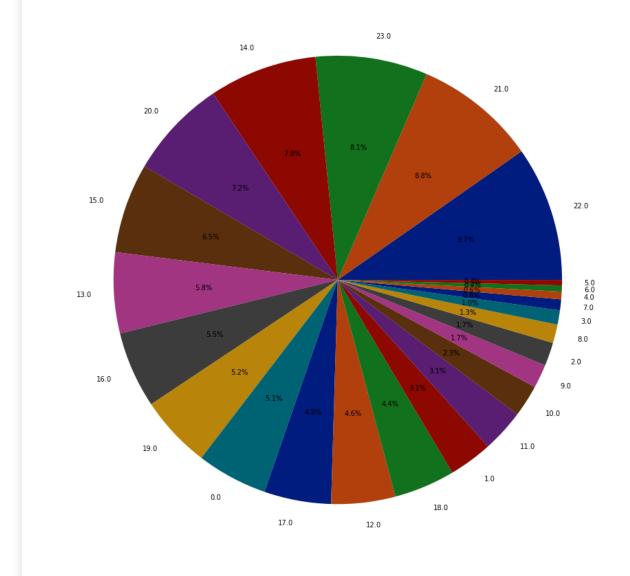




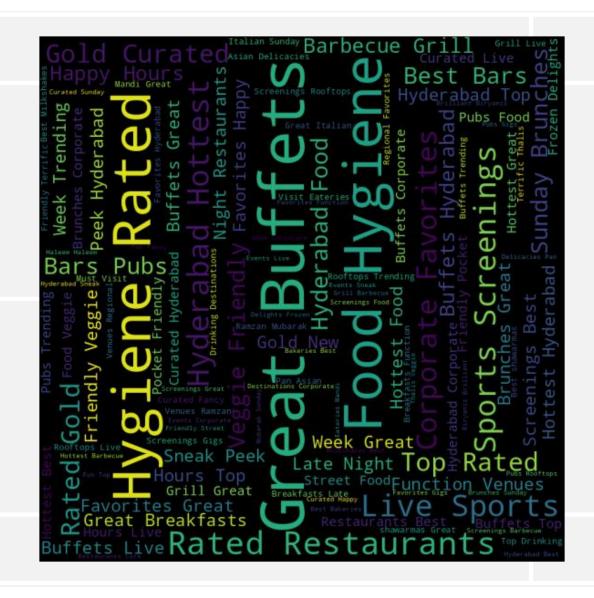
#### Cheapest restaurants

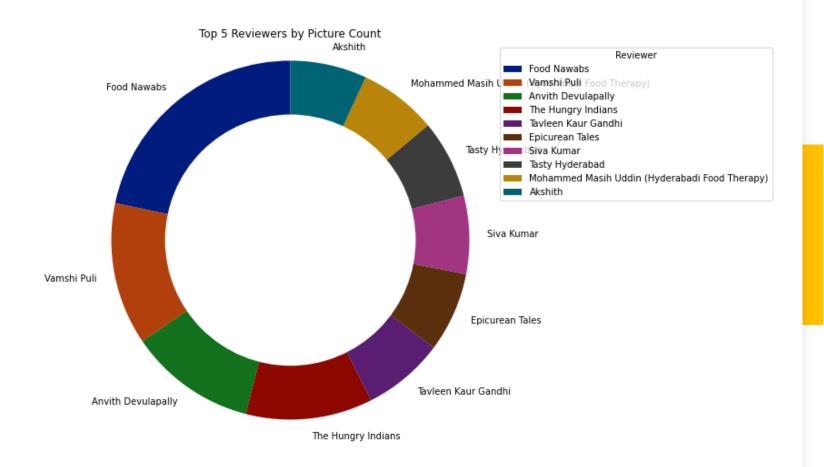


# Pie chart of hour of review



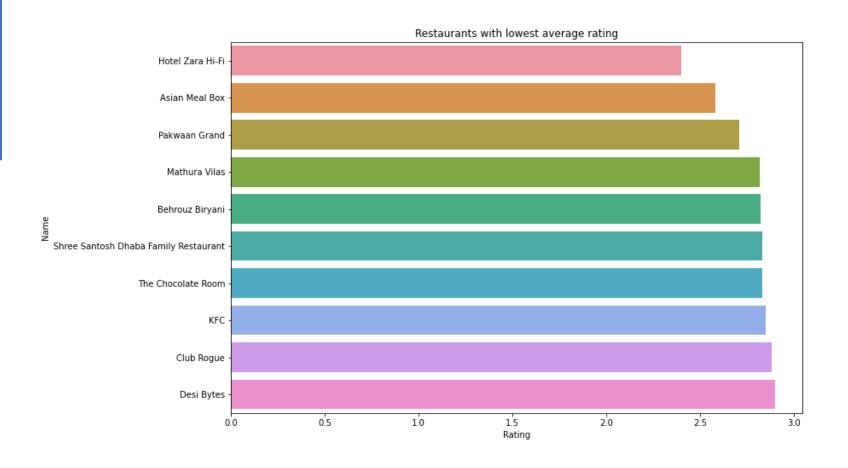
# Word cloud of collections (tags)

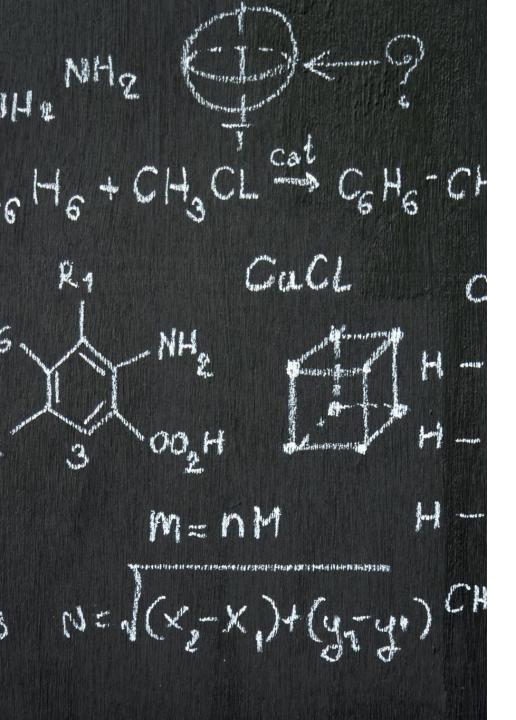




Donut chart of top 5 reviewers by picture count.

## Lowest rated restaurants

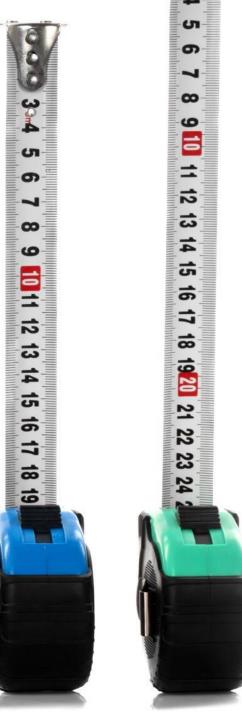




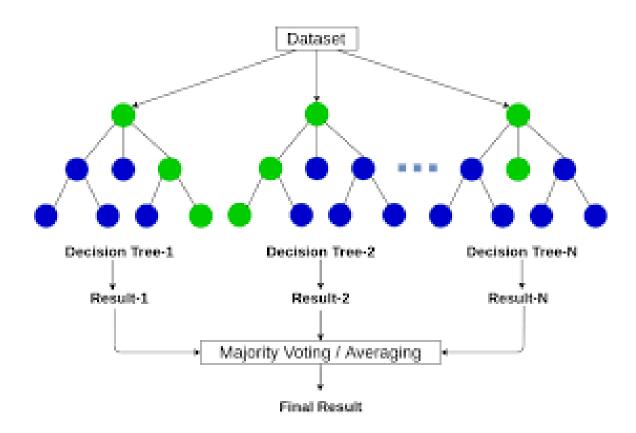
### Model 1: Naïve Bayes ( Sentiment Analysis)

- Naive Bayes is a probabilistic machine learning algorithm used for classification problems. It is based on Bayes' theorem.
- Naive Bayes is considered "naive" because it assumes that all features are independent of each other.









#### Random Forest

It is an ensemble learning method which builds multiple decision trees on different random subsets of the training data and features.

### Classification report

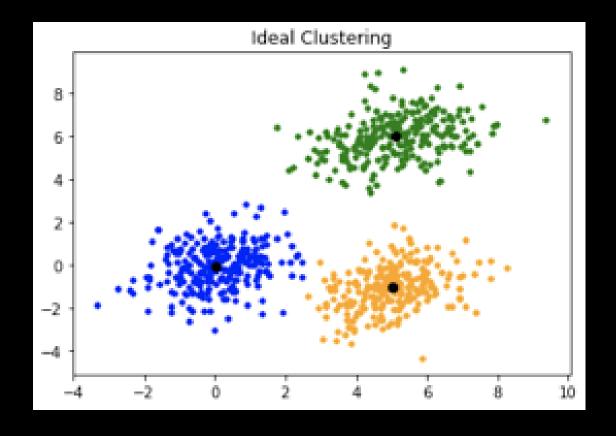
```
• precision recall f1-score support
```

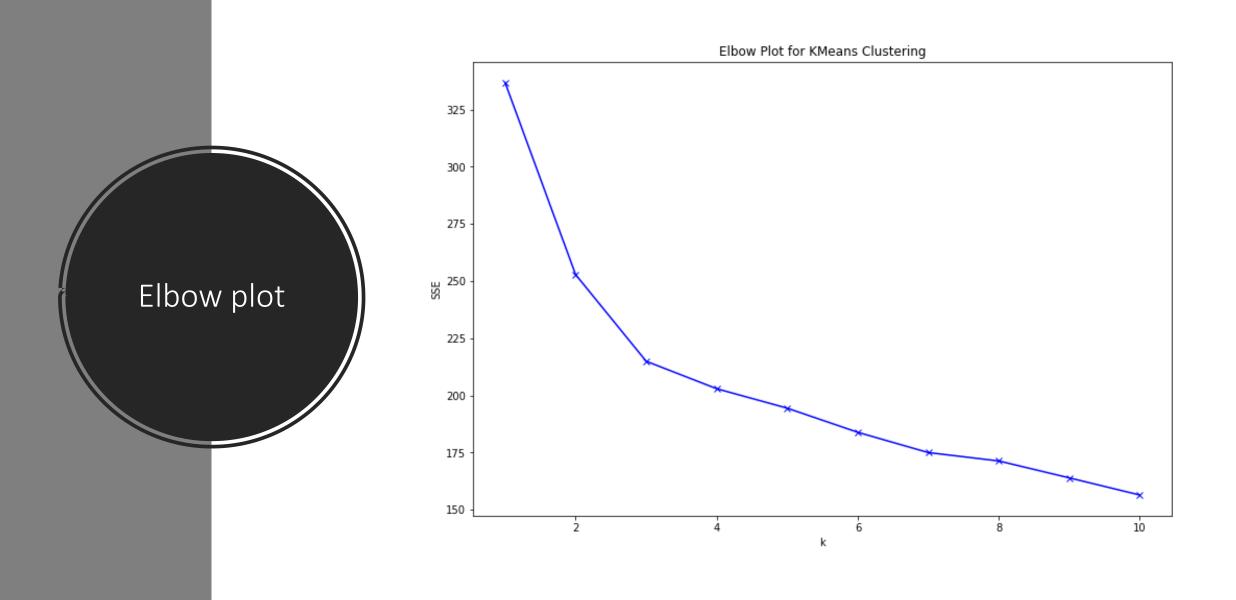
```
• 0 0.88 0.69 0.77 725
```

- 1 0.84 0.95 0.89 1266
- accuracy 0.85 1991
- macro avg 0.86 0.82 0.83 1991
- weighted avg 0.86 0.85 0.85 1991

# Model 2 : K means clustering

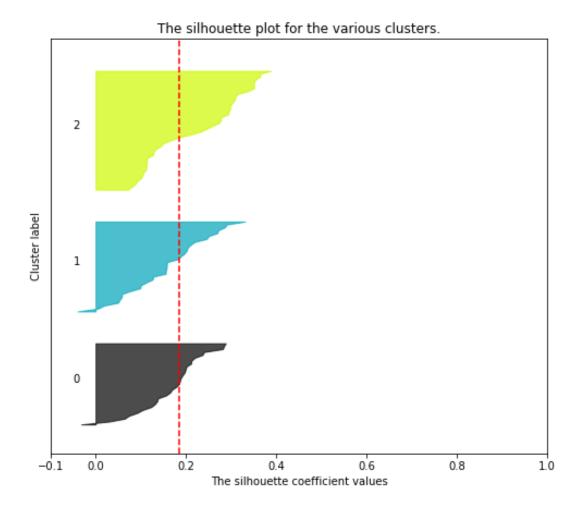
 It is an unsupervised machine learning algorithm used for grouping similar data points together based on their features. The algorithm works by dividing a set of data points into a predetermined number of clusters (k).





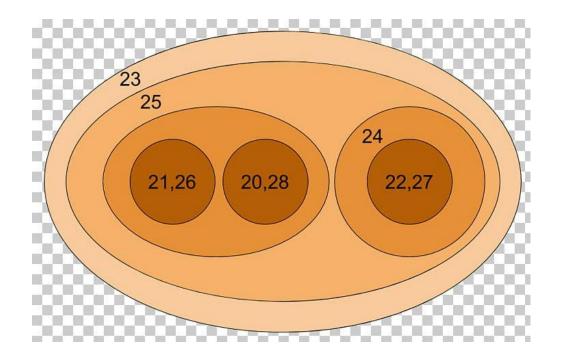
#### Silhouette analysis for KMeans clustering

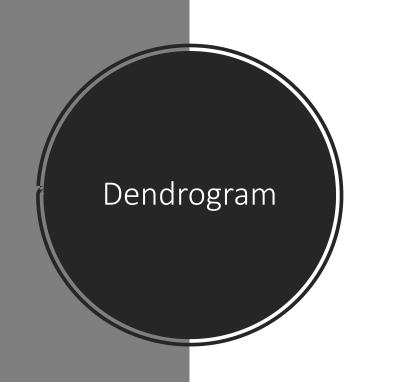


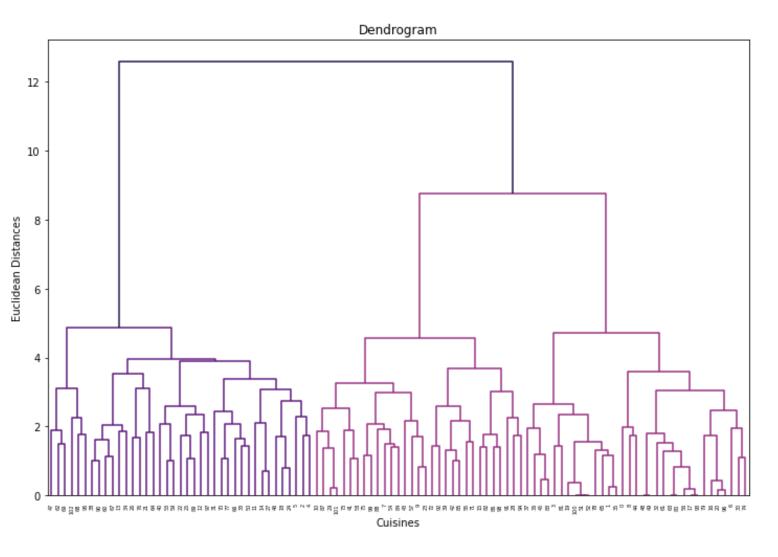


## Model 3: Heirarchical clustering.

 In agglomerative hierarchical clustering, each data point starts as a separate cluster, and the algorithm iteratively merges the most similar clusters until all data points belong to a single cluster.







# Thank you