IBM Data Science Professional Course

COURSERA CAPSTONE COURSE

The Battle of Neighborhoods

Starting a New Indian Restaurant in Delhi, India

By- Mukta Prafulla Sapre

July 2020

INTRODUCTION

- Delhi : A union territory in India containing New Delhi, a capital of India.
- Census 2011 :-

City Population : 11 Million which is second-highest in India

Urban Area Population : Over 26 Million which is world's second-largest urban area

Type equation here.

- Euromonitor International :-

Ranked as 28th-most visited city in the world and first in India by foreign visitors in 2015

- There are already many Restaurants, Hotels, Coffee Shops etc.
- Increasing population and tourism are increasing the demand for good quality Restaurants.

- The tourists mostly from foreign countries always want to taste Indian Regional food.
- Regularly eating continental food is not hygienic.
- It also creates so many economic problems.
- Due to Covid-19 pandemic, people are realising the importance of eating home-made or local food.
- Preferring to eat Indian food is the best solution to make the country self-dependant.
- Finding a place for such Restaurant will decide whether the Restaurant will run successfully or not.

BUSINESS PROBLEM

- Aim of Project :

  To analyze and find area in Delhi, India where a new Indian Regional Food Restaurant can be started and run successfully

- Business Problem :

  'Where to start new Indian Regional Food Restaurant in Delhi, India?'

INTEREST

- This project is useful to the businessmen, investors and developers who want to start a new Indian Restaurant in Delhi, India.
- Starting an Indian Restaurant will increase the demand of raw food materials that produce in remote areas of country. Indian farmers will get their productions for sell in market. It will definitely help to solve a problem of unemployment.
- Therefore, it will automatically raise the Indian economy.

DATA

- The list of suburbs (neighborhoods) in Delhi, India. Website: https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Delhi
- The latitude and longitude coordinates of each neighborhood using **Geocoder**.
- The venues data particularly related to Indian Restaurants from **Foursquare**.

METHODOLOGY

- I used the Python Jupyter Notebook in IBM Watson Studio to extract all the data and analyse.
- The data was extracted and transformed into dataframe with column 'Neighborhood' using BeautifulSoup.
- The latitude and longitude coordinates for each neighbourhood was found using Geocoder.
- Then using these values, the map of Delhi was drawn with all the neighborhoods superimposed on top with coloured circles with the help of Folium.
- Using Foursquare API, top 100 venues in 2000 meters radius was found.

- I extract the data into dataframe with columns neighbourhood, venue-name, venue-category, latitude and longitude.
- Checking how many venues are there for each neighbourhood, I also found that there are 70 unique categories of venues.
- The rows were grouped by neighborhood and by taking the mean of frequency of occurrence of each category.
- The dataframe containing venue category 'Indian Restaurant' was separated.
- K-Means clustering and some other analysis was done further.

K-MEANS CLUSTERING ALGORITHM

Before explaining Exploratory Analysis, I would like to describe about K-Means Clustering Algorithm.

K-Means Clustering Algorithm :

K-means clustering algorithm is one of the simplest unsupervised learning algorithms that solve all the well-known clustering problems like the problem of this project.

The algorithm classifies the data into K centres which are as small as possible.

EXPLORATORY ANALYSIS

K-Means Clustering :

- I transformed the data into 3 clusters based on the frequency derived.
- The result allows determining the area where the number of 'Indian Restaurant' is to be increased i.e. which area is convenient and suitable to start a new Indian Restaurant.
- Then I had to determine which cluster has less number of Indian restaurants, which cluster has moderate number of Indian restaurants and which cluster has very high number of Indian restaurants.
- Visualizing these clusters with the help of map is really very simple way to understand the results.

Other Analysis :

- I separated the data with venue categories that are related to food restaurants i.e. Indian Restaurant, Pizza Place, Fast Food Restaurant, American Restaurant, Italian Restaurant etc.etc.
- Calculating total number of Restaurants in each category, dataframe was found from which Bar Graph and Pie Chart were drawn.
- After this, the above categories were classified into three types that are Indian, Foreign and Both.

Indian : Restaurants that serve only Indian Food from different parts of India.

Foreign : Restaurants that serve only foreign food from different countries in the world.

Both : Restaurants that serve Indian food as well as foreign food.

- Calculating the number of Restaurants in each type and also its percentage, the Bar Graph and Pie Charts were drawn.

- The dataframes found are as follows

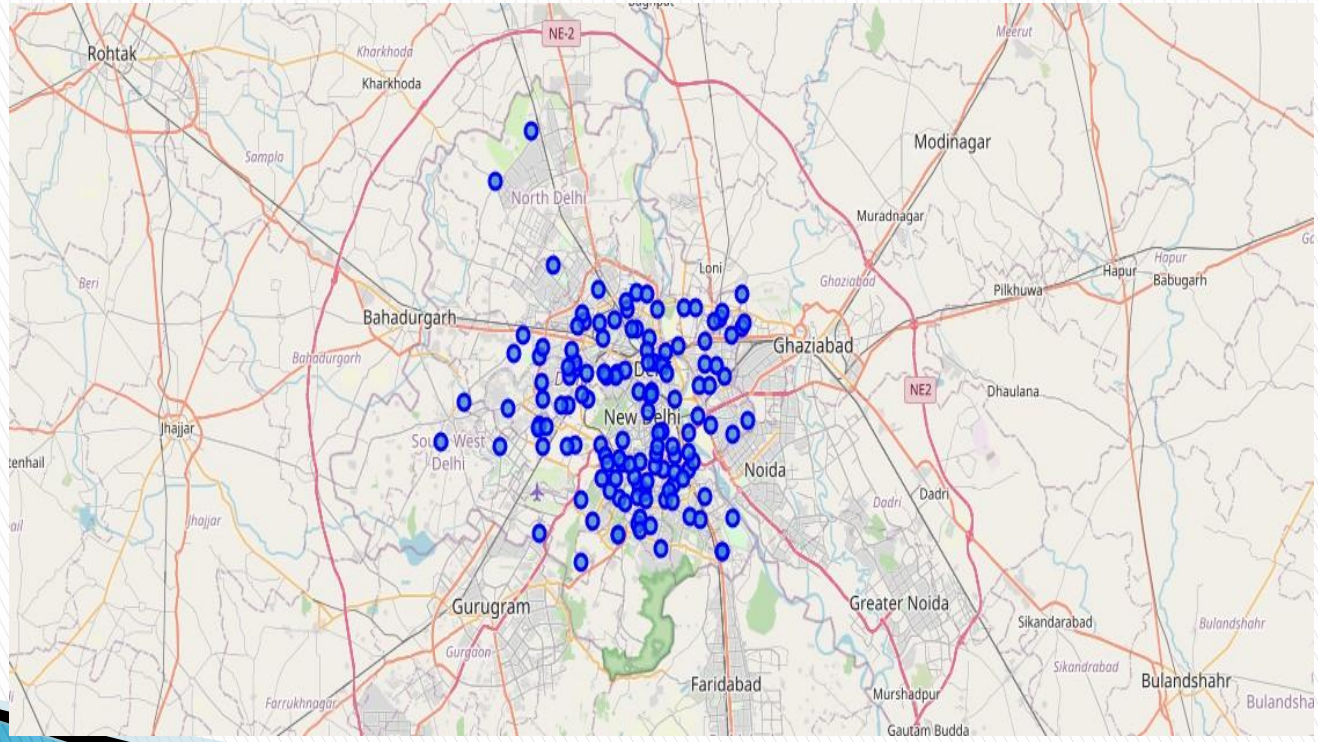
Category	Total
Indian Restaurant	145
Pizza Place	182
Fast Food Restaurant	94
American Restaurant	1
Middle Eastern Restaurant	27
Chinese Restaurant	41
Sandwich Place	42
Hotel	121
Diner	27
Food Truck	26
Thai Restaurant	27
Italian Restaurant	27
Salad Place	17
Vegetarian / Vegan Restaurant	18
Food Court	13

Types of Food	Number of Restaurants
Indian	145
Foreign	458
Both	205

Types of Food	Percentage
Indian	17.94554455
Foreign	56.68316832
Both	25.37128713

RESULTS

Following picture shows the map of Delhi with all the neighborhoods superimposed on top.



- Classifying the data into clusters, I found 3 clusters as follows:

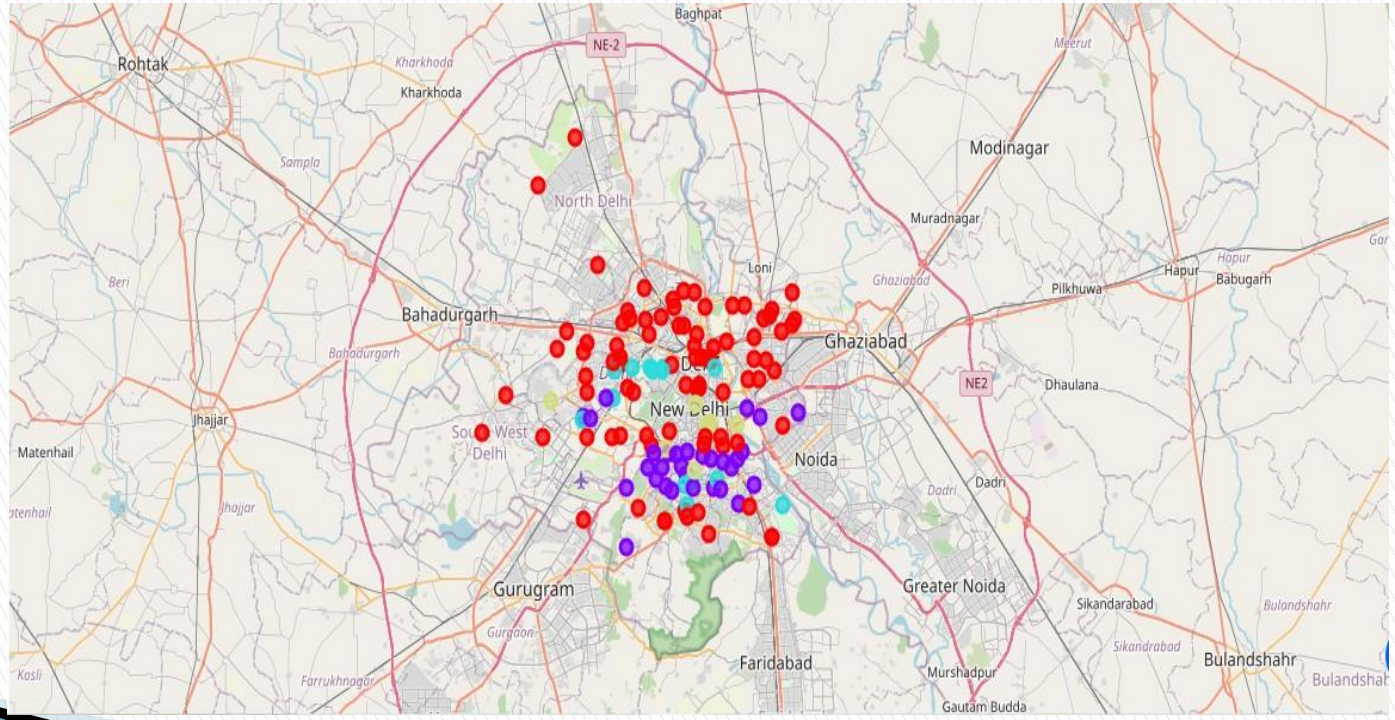
  Cluster 0: Neighborhoods with less number of Indian Restaurants.

  Cluster 1: Neighborhoods with large number of Indian Restaurants.

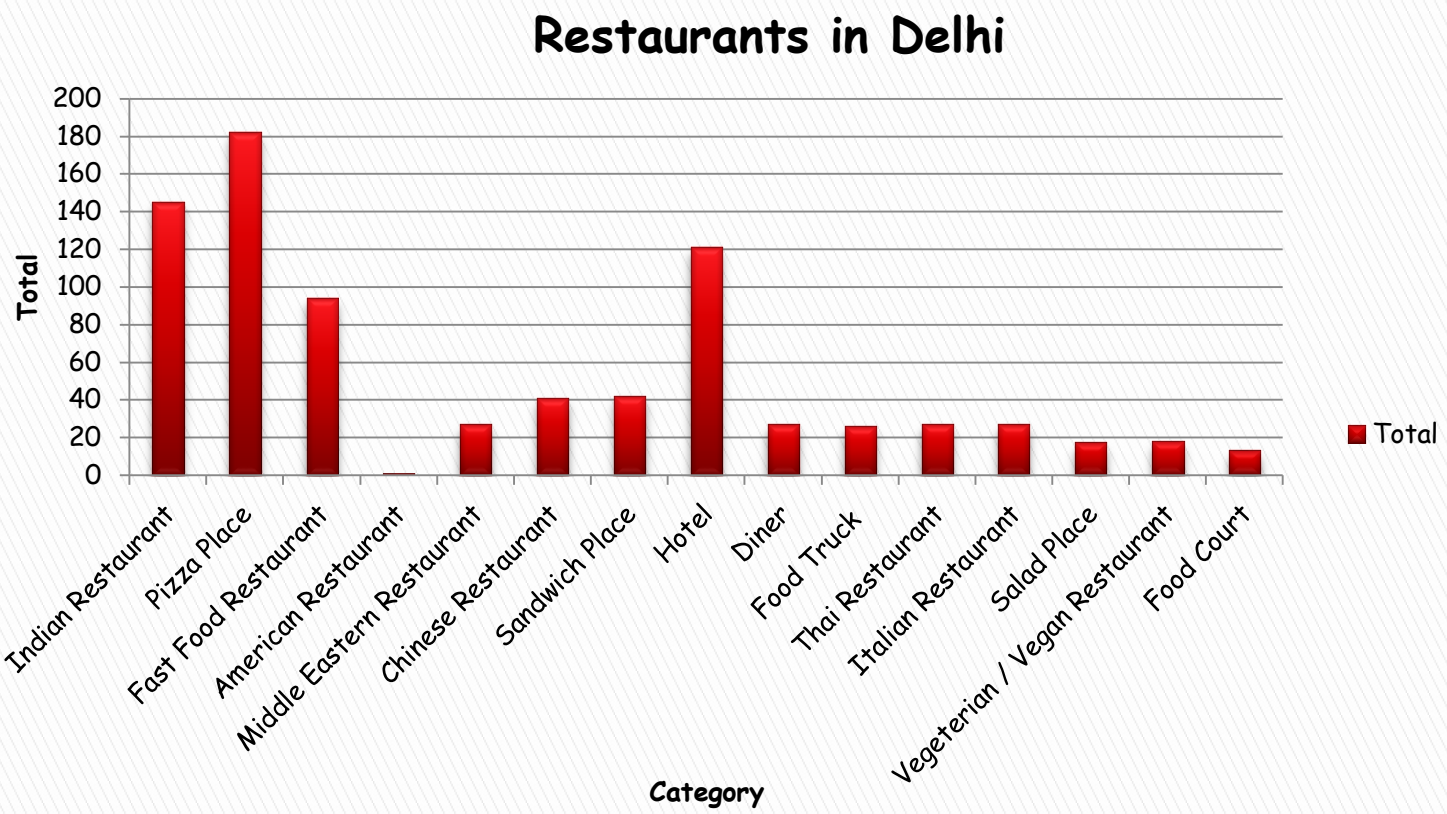
  Cluster 2: Neighborhoods with moderate number of Indian Restaurants.

- These clusters are visualized with the help of map.

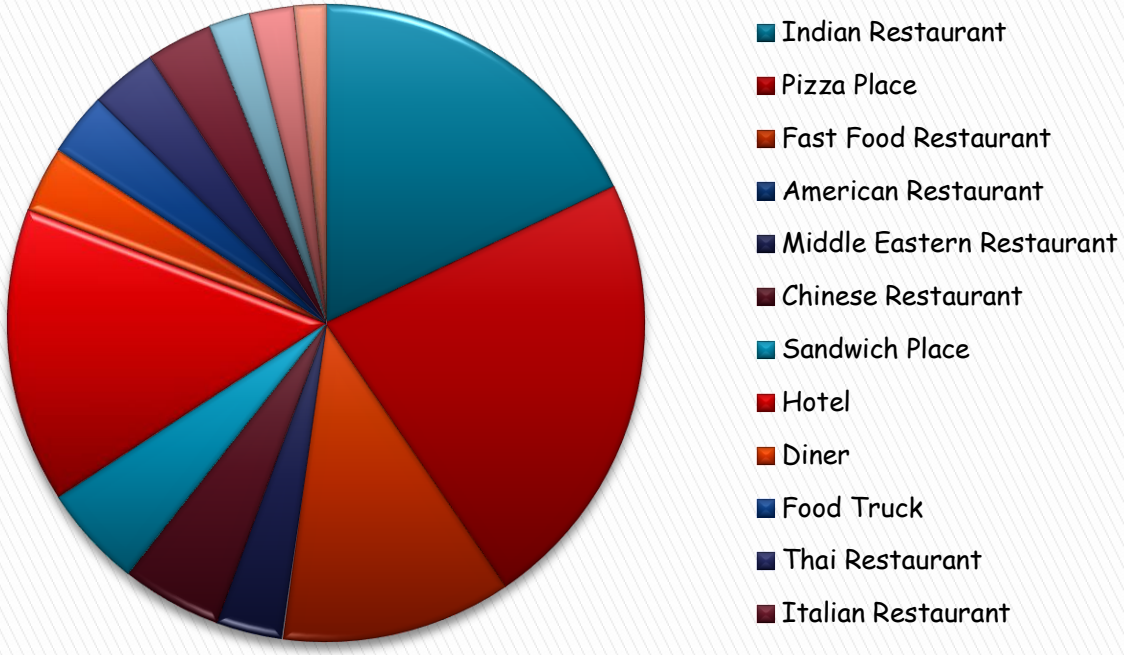
- In the following map, cluster 0 is represented by Blue colour, Cluster 1 by Red colour and Cluster 2 by Purple colour.

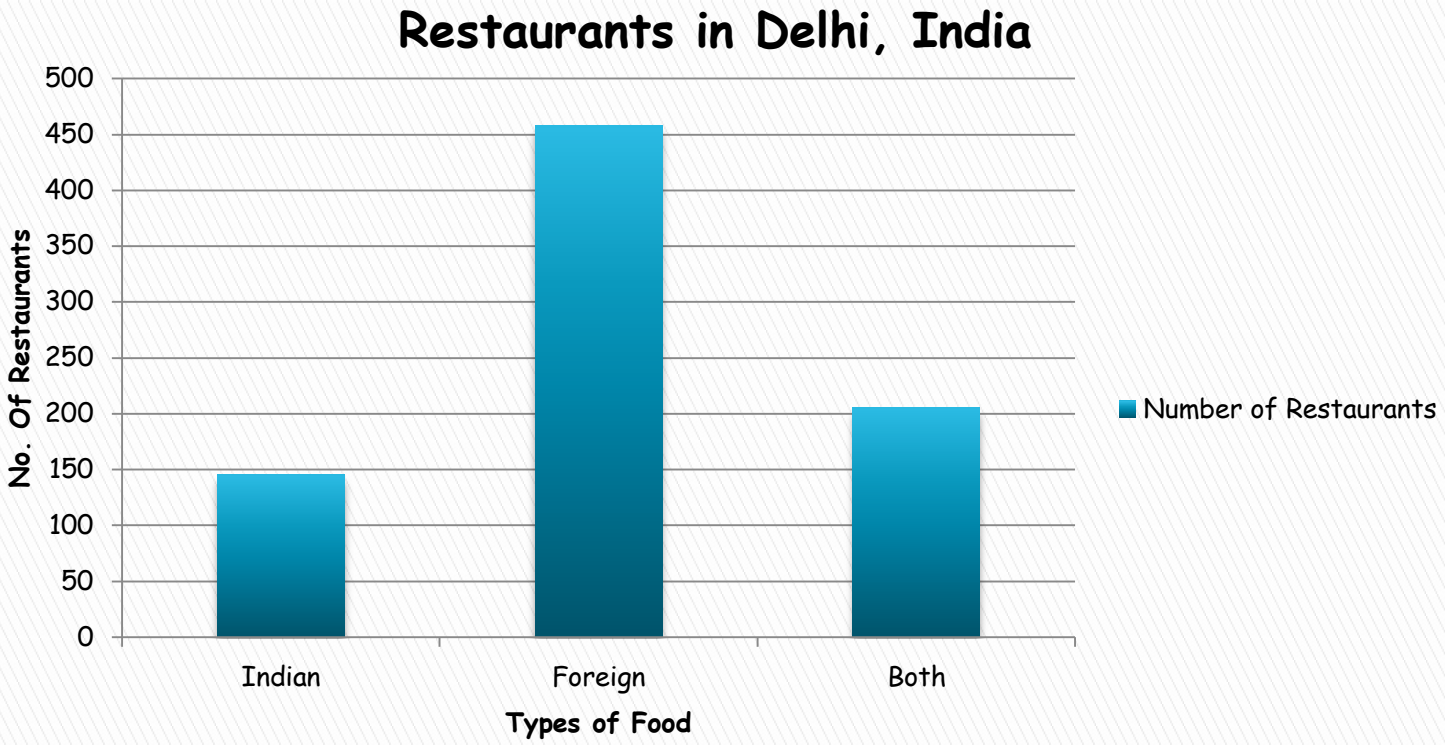


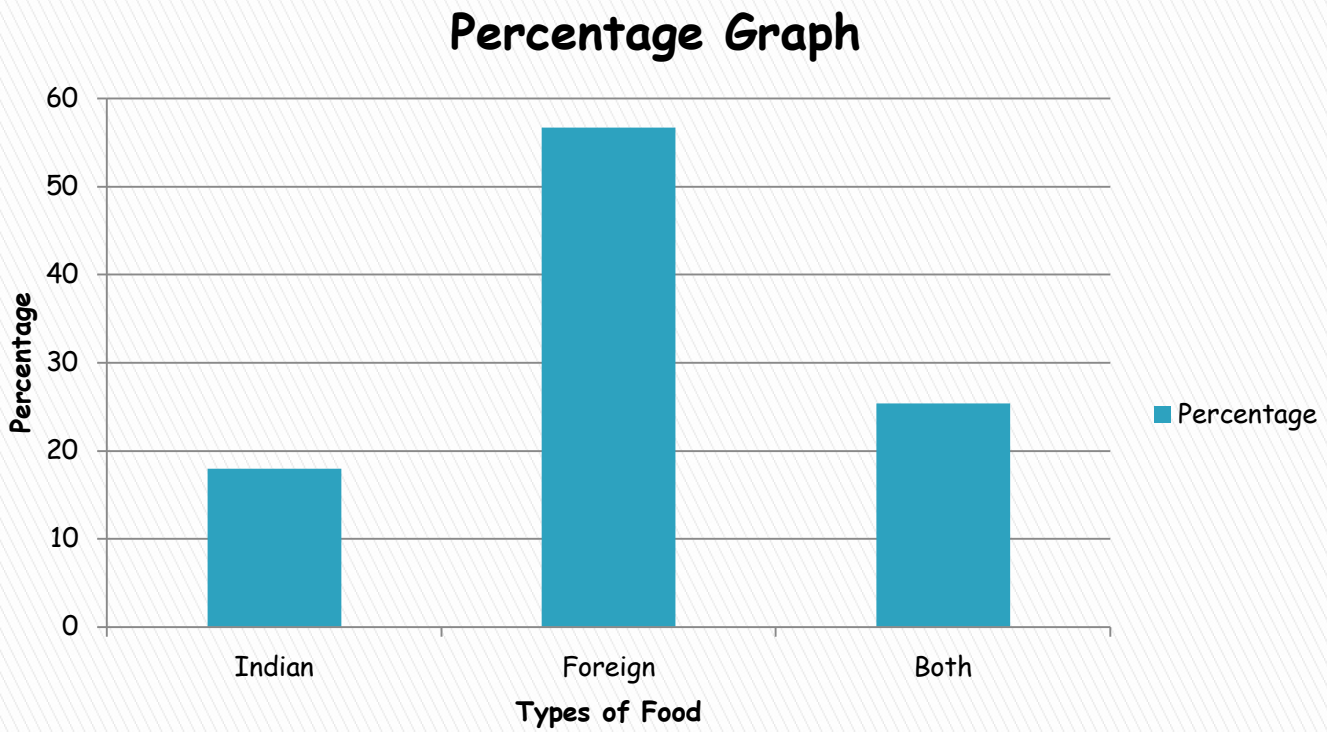
BAR GRAPHS AND PIE CHARTS

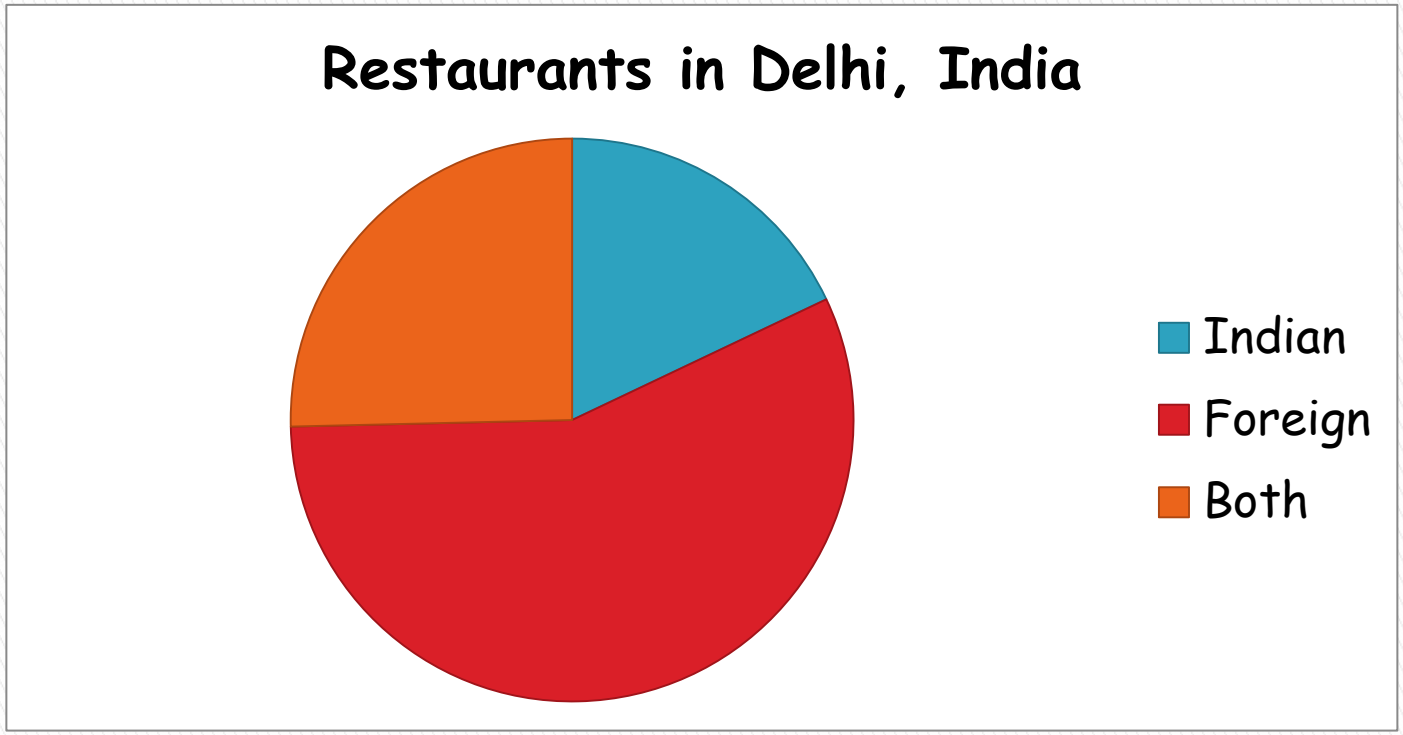


Restaurants in Delhi, India




Restaurants in Delhi, India
500
450
400
350
300
250
200
150
100
50
0
No. Of Restaurants
Indian
Foreign
Both
Types of Food
Number of Restaurants


Percentage Graph
60
50
40
30
20
10
0
Percentage
Indian
Foreign
Both
Types of Food
Percentage


Restaurants in Delhi, India
Indian
Foreign
Both

DISCUSSION

- The results of K-Means clustering show that there is highest concentration of Indian restaurants in northern part of Delhi. The number of Indian restaurants is moderate in southern part. The central part of Delhi has very few numbers of Indian Restaurants.
- Since, there are large numbers of Indian restaurants in northern part; there will be large competition for a new restaurant. Therefore, it will not be convenient to open a new restaurant there.
- So, central part of Delhi will be more suitable and potential to start a new Indian Restaurant.

- From another analysis of types of restaurants in Delhi, it is shown that the overall availability of Indian food is less in comparison with the foreign food.
- My project suggests the businessmen, investors, developers and people interested in hoteling sector to start new Indian restaurants in the Central part of Delhi.
- The number of Indian restaurants has to be increased in Delhi.
- I recommend all these people to increase their business in Indian Food to help the country to be self-dependent.

CONCLUSION

- The new Indian Restaurant can be started in Central areas of Delhi.
- The number of only Indian Food Restaurants is very less in Delhi in comparison with Foreign Restaurants.
- The number of restaurants only serving Indian food should be increased in entire Delhi. Especially, it has to be increased in areas in Central part of Delhi.

LIMITATIONS AND FUTURE DIRECTIONS

- My project is based on the data of only number of restaurants in Delhi. But, there are so many factors that affect the hoteling sector in that area such as population, location, class of living population in respective areas, choice of majority people etc. etc…
- It is practically complicated and time consuming to collect all these data and analyse.
- Now a days, the percentage of online food deliveries is rapidly increasing. Someone can consider this factor in this project while doing further analysis.

I hope my project will help maximum people to make good career in hoteling sector. Anyone can give me more suggestions on this project so that I can improve it. Anyone can point out my mistakes and suggest more ideas.

Thank You!!!