Minor Project

Chennai Zomato EDA

Prepared by

A9920121001111(el)
Syed Manzoor Ahmed
MBA - Final Year (2021 - 2023)
Data Science

Project Report

Abstract :

To start a restaurant in Chennai we are using data to analyse on which location will be the best place to open. We are also analysing what type of restaurant are located in Chennai, area wise types of restaurants, what are their price range, area wise price range, rating of the restaurants, cuisine types whether its North Indian, South Indian, Chinese and so on.

Introduction of the Study :

Today the Internet is booming in India. We are using the internet for many purposes such as ordering foods, buying electronics, for education etc. So In this project, using various data we are going to discuss people's behaviour and restaurant categories, and will discuss area wise restaurants in Chennai. So first we need data to analyse these things. From the zomato API we have got a huge dataset. (These dates are available openly on the internet for education purposes). After collecting the raw data we have to first analyse how many records and attributes are present in the dataset. How many features are available from that dataset. Then based on the requirement we have to clean the raw data for analysing process.

Objective(s) of the Study :

To find the best place to open a restaurant in Chennai. Understanding of restaurant space in Chennai. What kind of foods are preferable by the people of Chennai. Which cuisine is mostly chosen by people on area wise. Finding the price range. Also analyzing the restaurant type based on loaction wise. Restaurant basis most preferred dishes.

Background Study :

So In this analysis we are using python which is open source. we are performing this analysis in kaggle platform. This dataset is collected from zomato API. In python we can delete, visualize, sort and do many things. We are also using some famous libraries called matplot and seaborn.

As dataset file is high so we can't able to attach. So for dataset reference kindly use the below.

https://www.kaggle.com/datasets/rabhar/zomato-restaurants-in-india?select=zomato_restaurants_in_India.csv

For python documentation kindly use the below link

https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.drop.html

To see the project workbook use the below link_

https://www.kaggle.com/code/syedmanzoor/chennai-zomato-restaurants

Research Methodology :

Data Collection :

First we have to collect data from the source. For this topic we have got dataset readily from internet. So this collected data from the source is called as raw data. Raw data contains many duplicates, many unwanted attributes which will look like a messy data.

Data loading :

Since we are working this in python. we have to load the csv file in kaggle platform once the data is loaded, firstly the data has to be cleaned. So the raw data consists of 211944 records and 26 attributes. Data is loaded through pandas library using the command,

"zomato = pd.read_csv('/kaggle/input/zomato-restaurants-in-india/Zomato_restaurants_in_india.csv')

zomato.head()"

Data Cleaning :

Based on our objective are cleaning the data. So we are dropping some unwanted attributes (columns) from the dataset using the command,

zomato.drop(['res_id', 'address', 'latitude', 'longitude', 'locality_verbose',

'timings', 'rating_text', 'delivery', 'takeaway'],axis=1,inplace=True)

Then in data cleaning process we are removing the duplicate data. So many duplicate records will be presented in dataset. This duplicate is removed using the drop command,

```
print('No.of duplicate entries:',zomato[zomato.duplicated()].count()[0])
zomato.drop_duplicates(inplace=True)
print('After removal:\nNo. of features:',zomato.shape[1],'\nNo. of resturants:',zomato.shape[0])
```

After this we found that, 156014 are duplicate records. so remaining 55930 records are validated for analyzing process. Then we also found some blank fields present in the dataset. so we are replacing that blank fields by 'NA'. Then finally we are filtering only Chennai city data. So total zomato restaurants located in Chennai city is 1891

Data Distribution :

Once our data is cleaned, then we can able to analyze it. Data distribution is done for better understanding of the data in the related manner. In this we are grouping the data based on locality, restaurant type, based on rating and price range.

Visualization :

Visualization is process showing the insights (data) in visual format for a good understand of what actually happened. Visualization process, we can also tell this as a story telling process. using the data we are telling the story or showing in visual format. For visualization in python we are using some libraries such as matplot and seaborn.

Data Analysis & Interpretation :

we have to dataset in csv file, we have to load it in kaggle using the below command

```
zomato = pd.read_csv('/kaggle/input/zomato-restaurants-in-india/Zomato_restaurants_in_india.csv')

zomato.head()
```

Then in data cleaning process we are removing the duplicate data. So many duplicate records will be presented in dataset. This duplicate is removed using the drop command,

```
print('No.of duplicate entries:',zomato[zomato.duplicated()].count()[0])
zomato.drop_duplicates(inplace=True)
print('After removal:\nNo. of features:',zomato.shape[1],'\nNo. of resturants:',zomato.shape[0])
```

Ttal recNo. of features: 11
No. of resturants: 211944
No.of duplicate entries: 156014
After removal:
No. of features: 11
No. of resturants: 55930

As dataset file is high so we can't able to attach. So for dataset reference kindly use the below.

https://www.kaggle.com/datasets/rabhar/zomato-restaurants-in-india?

select=zomato_restaurants_in_India.csv

For python documentation kindly use the below link

https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.drop.html

To see the project workbook use the below link_

https://www.kaggle.com/code/syedmanzoor/chennai-zomato-restaurants

Results & Discussions :

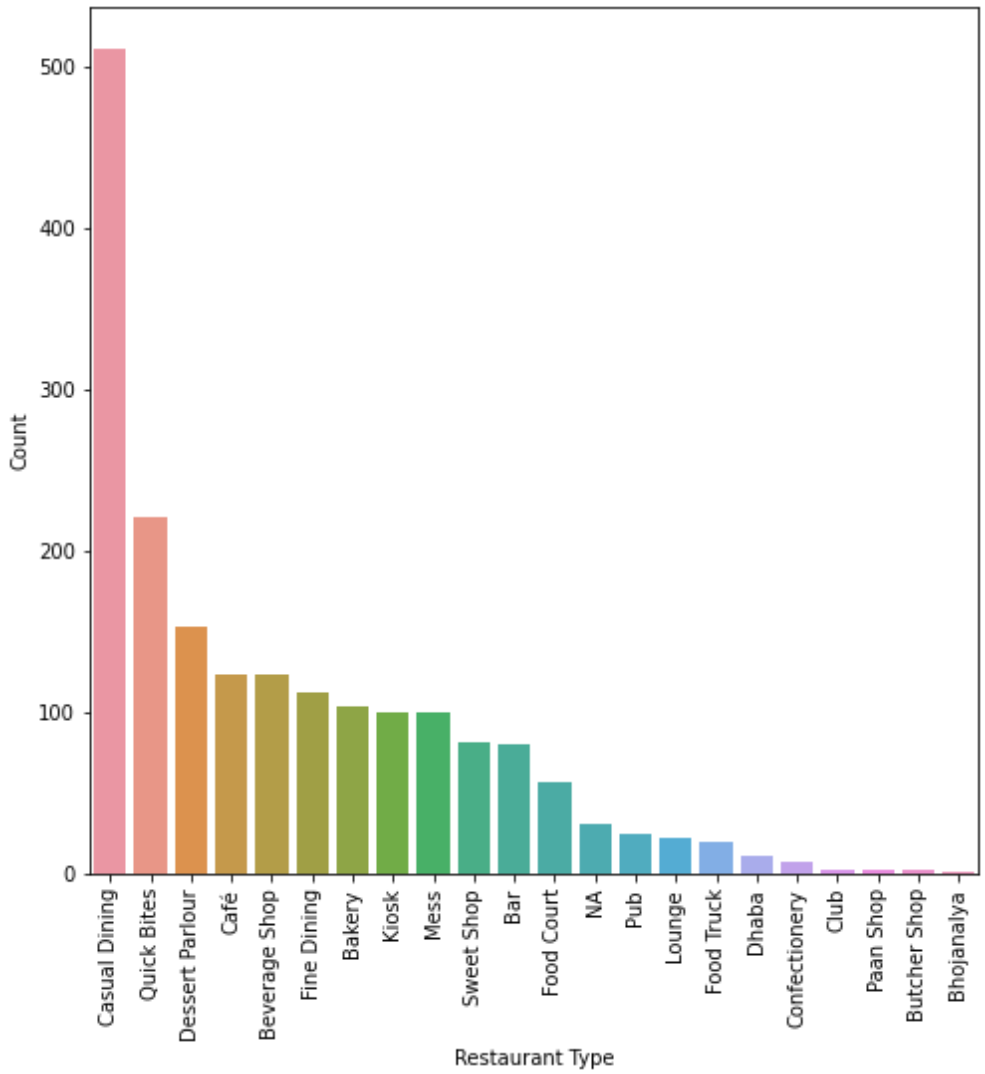
From the dataset we can able to see that, total 26 attributes (columns) are presented. Such as restaurant id, name of the restaurant, establishment, restaurant address, city, locality, latitude, longitude, locality verbose, cuisine restaurant timing, average cost for two, price range, highlights, aggregate ratings, rating in text format, votes, photo count, delivery, takeaway. So for our analyses we don't need this much of data. We have to validate the data and need to analyse. So here we're going to remove some columns using the command 'drop' For our analyses we need restaurant name, establishment, city, locality, cuisine, average cost for two, price range, highlights, aggregate rating, votes, photo count.

After removing these columns, still many duplicate records (rows) are present in our dataset. To remove duplicates we are using 'drop_duplicates' Then many empty cells are replaced by 'NA'

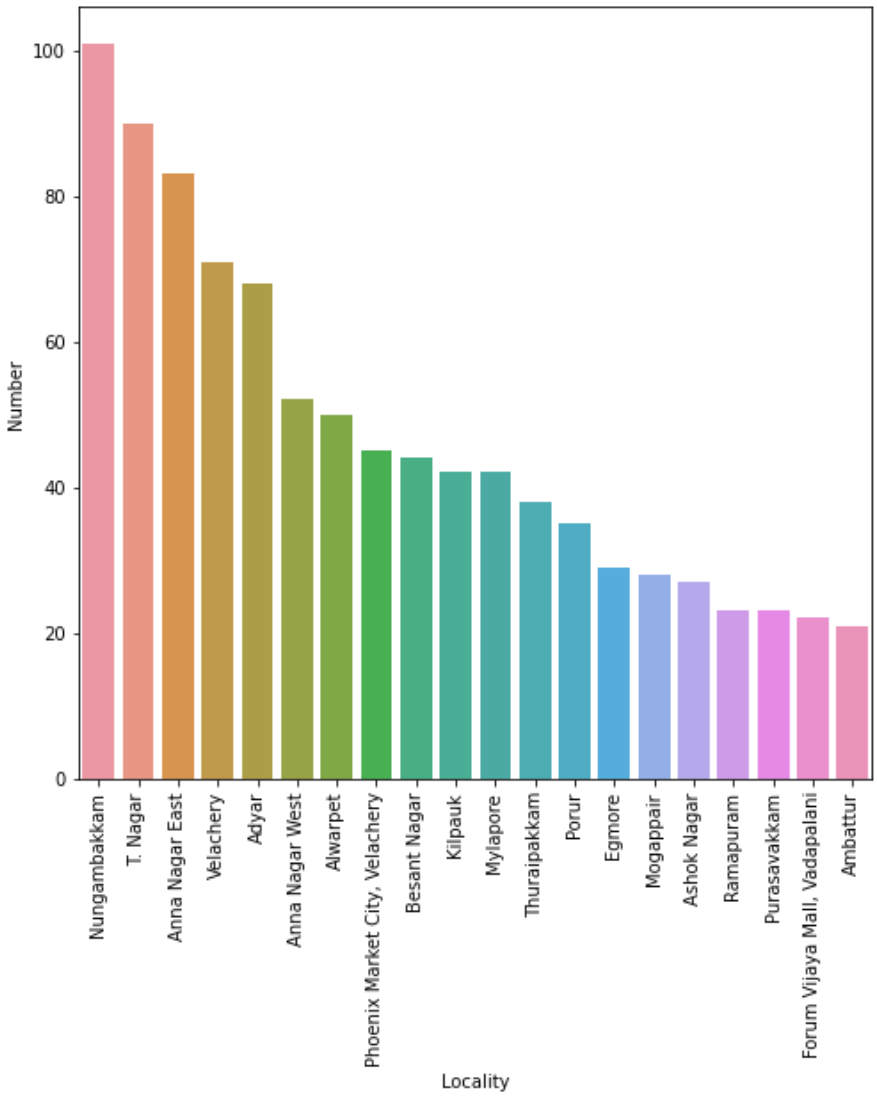
After the cleaning process we got 1891 restaurants located in Chennai city.

Now data distribution will be done for deep understanding of the restaurants. The data is grouped by its type, locality, price range and ratings then its been plotted for visualization.

From this visual we can see that the city serves different cuisines such as north indian, south indian, arabian, chinese, dessert, fastfood, continental, pizza, ice cream etc. Among these cuisines north indian, south indian and chinese are mostly preferred in Chennai.



This bar chart represents the type of restaurants in Chennai. As we can see there are ma y types of restaurants presents in Chennai such as casual dining, quick bites, dessert parlour, cafe, beverage shops, fine dining, bakery, kiosk, mess, sweet shop, bar, food court, pub, lounge, food truck, dhaba, club etc. So from the above bar chart we can have analysed most of the restaurants are labelled as casual dining followed by quick bites. Restaurants such as food court, bar, sweet shop, mess kiosk, bakery, fine dining, beverage shop, cafe, dessert parlour are equal in number in city.



This bar chart represents the location wise restaurants in Chennai. So here we can able to see many locations such as besant nagar, t nagar, anna nagar, nungambakkam, velachery, adyar, alwarpet, phoenix market city, vadapalani, kilpauk, mylapore, thuraipakkam, porur, mogappair, pursaiwakkam, ambattur and so on. So this bar chart shows top 20 localition in Chennai. From this we have analysednungambakkam occupies the first place in the list followed by t nagr , velachery, besant nagar, kilpauk, mylapore, thuraipakkam lies in similar number of restaurants.

Recommendations & Conclusion :

From the exploratory data analysis EDA, we can recommend some ideas to start the restaurant.

In location wise we can recommend to go for nungambakkam, t nagar, anna nagar east, velachery & adyar. In restaurant type type we can highly recommend casual dinings. because from the data analyse we came to know that casual dining are more preferred by the people followed by quick bites, dessert parlour, cafe, beverage shops.

In cuisine, we would recommend to go with north indian, south indian, and chinese. These cuisines are highly preferred by the peoples.

Bibliography & References :

As dataset file is high so we can't able to attach. So for dataset reference kindly use the below.

https://www.kaggle.com/datasets/rabhar/zomato-restaurants-in-india?select=zomato_restaurants_in_India.csv

For python documentation kindly use the below link

https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.drop.html