Data Warehouse and Integration

Project report on Zomato

DHRUV GOEL (BS19BDS031)

VEDANT KABRA (BS19BDS020)

TUSHAR SONATHALIA (BS19BDS022)

PRADEEP SANDILYA (BS19BDS023)

Table of Contents

BACKGROUND INFORMATION

The objective of the project is to provide insights, recommendations, and solutions to the CEO of Zomato to help them with their business. This is being done as a part of the Data Warehouse and Integration course. The report focuses on analysing restaurants that fall under major TIER 1 cities of India that use Zomato as a medium to reach out to their customers.

Back in 2008, all it took was an idea to enable digital access to thousands of restaurant menus. Three passionate foodies who hated waiting in lines, drove around Delhi to collect menus from restaurants, scan them and put them online. Their idea has now grown into the vision that drives the team of 5000+ people every day.[1]

Zomato is an Indian restaurant aggregator and food delivery start-up founded by Deepinder Goyal and Pankaj Chaddah in 2008. Zomato provides information, menus, and user-reviews of restaurants as well as food delivery options from partner restaurants in select cities. Zomato also began grocery delivery amid the Covid-19 outbreak. As of 2019, the service is available in 24 countries and in more than 10,000 cities.

Zomato was founded as Foodiebay in 2008 and was renamed Zomato on 18th January 2010 as Zomato Media Pvt. Ltd. In 2011, Zomato expanded across India to Delhi NCR, Mumbai, Bangalore, Chennai, Pune, and Kolkata. In 2012, the company expanded operations internationally in several countries, including the United Arab Emirates, Sri Lanka, Qatar, the United Kingdom, the Philippines, and South Africa. In 2013, Zomato was launched in New Zealand, Turkey, Brazil, and Indonesia, with its website and apps available in Turkish, Portuguese, Indonesian and English languages. In April 2014, Zomato launched its services in Portugal, which was followed by launches in Canada, Lebanon, and Ireland in 2015.

In 2019, Zomato acquired Seattle-based food portal Urbanspoon, which led to the firm's entry into the United States and Australia. This U.S.-expansion brought Zomato into direct competition with similar models such as Yelp and Foursquare.

---

[1] https://www.zomato.com/who-we-are

With the introduction of .xxx domains in 2011, Zomato also launched zomato.xxx, a site dedicated to food porn. In May 2012, it launched a print version of the website named "Citibank Zomato Restaurant Guide," in collaboration with Citibank, but it has since been discontinued.

Zomato had also made a name for itself for its prowess in digital marketing. Coming from the online food industry it has made a revenue of $210 million.

Info Edge (27.6%), Ant Financial (23%) and Uber (9.9%) are the parent companies of Zomato. [2]

Zomato provides various services like Restaurant Search & Discovery, Online Ordering, Table Reservations & Management, POS Systems, Subscription Services etc. [3]

Zomato has also partnered with Paytm, PayPal and Freecharge to enable online payments.

DATASET INFORMATION:

The data for this project has been acquired from Kaggle. Kaggle is a subsidiary of Google LLC and is an online community of data scientists and machine learning practitioners. [4]

The dataset[5] consists of 56664 restaurants spread across six Tier-1 cities of India, namely, Hyderabad, Delhi, Mumbai, Kolkata, Chennai, and Bengaluru.

ATTRIBUTE INFORMATION:

- NAME
- PRICE (approximate price for two people)
- CUISINE CATEGORY
- CITY
- REGION (under which suburb of the city does the restaurant fall in)
- URL (website of the restaurant)
- CUISINE TYPE (type of restaurant)

---

[2] https://en.wikipedia.org/wiki/Zomato#cite_note-2
[3] https://www.financialexpress.com/industry/sme/zomato-swiggy-launch-alcohol-delivery-service-in-odisha/1971433/
[4] https://en.wikipedia.org/wiki/Kaggle
[5] https://www.kaggle.com/ronidas39/zomato-india-data-set

- TIMING
- RATING (out of 5)
- VOTES (number of people who have rated the restaurant)

RESEARCH

DATA PRE-PROCESSING:

To clean the dataset Python was used as ETL tool. Python is a high-level general purpose programming language.[6]

The dataset retrieved from Kaggle was broken down into multiple files that contained Zomato data of each city in India. This was further broken down to multiple CSV files that contained data of each sub-locality of the cities of India.

After downloading data of the six cities on the local machine, the data files were loaded on a jupyter notebook and following is the code that helped in cleaning the dataset to perform further analysis:

```python
import pandas as pd
import numpy as np

globbed_files = glob.glob("C:\\Users\\kabra\\Downloads\\archive (1)\\Hyderabad\\*.csv") #creates a list of all csv files

data = [] # pd.concat takes a list of dataframes as an agrument
for csv in globbed_files:
    frame = pd.read_csv(csv,sep = "|")
    data.append(frame)

Hyd = pd.concat(data, ignore_index=True) #dont want pandas to try an align row indexes
#bigframe.to_csv("Pandas_output2.csv")

globbed_files = glob.glob("C:\\Users\\kabra\\Downloads\\archive (1)\\Mumbai\\*.csv") #creates a list of all csv files

data = [] # pd.concat takes a list of dataframes as an agrument
for csv in globbed_files:
    frame = pd.read_csv(csv,sep = "|")
    data.append(frame)

Mumbai = pd.concat(data, ignore_index=True) #dont want pandas to try an align row indexes
#bigframe.to_csv("Pandas_output2.csv")

globbed_files = glob.glob("C:\\Users\\kabra\\Downloads\\archive (1)\\Chennai\\*.csv") #creates a list of all csv files

data = [] # pd.concat takes a list of dataframes as an agrument
for csv in globbed_files:
    frame = pd.read_csv(csv,sep = "|")
    data.append(frame)

Chennai = pd.concat(data, ignore_index=True) #dont want pandas to try an align row indexes
#bigframe.to_csv("Pandas_output2.csv")

globbed_files = glob.glob("C:\\Users\\kabra\\Downloads\\archive (1)\\Bengaluru\\*.csv") #creates a list of all csv files

data = [] # pd.concat takes a list of dataframes as an agrument
for csv in globbed_files:
    frame = pd.read_csv(csv,sep = "|")
    data.append(frame)

blore = pd.concat(data, ignore_index=True) #dont want pandas to try an align row indexes
#bigframe.to_csv("Pandas_output2.csv")

globbed_files = glob.glob("C:\\Users\\kabra\\Downloads\\archive (1)\\Delhi NCR\\*.csv") #creates a list of all csv files
```

Data Pre-processing in python

---

[6] https://en.wikipedia.org/wiki/Python_(programming_language)

```
data = [] # pd.concat takes a list of dataframes as an agrument
for csv in globbed_files:
    frame = pd.read_csv(csv,sep = "|")
    data.append(frame)

delhi = pd.concat(data, ignore_index=True) #dont want pandas to try an align row indexes
#bigframe.to_csv("Pandas_output2.csv")

globbed_files = glob.glob("C:\\Users\\kabra\\Downloads\\archive (1)\\Kolkata\\*.csv") #creates a list of all csv files

data = [] # pd.concat takes a list of dataframes as an agrument
for csv in globbed_files:
    frame = pd.read_csv(csv,sep = "|")
    data.append(frame)

kol = pd.concat(data, ignore_index=True) #dont want pandas to try an align row indexes
#bigframe.to_csv("Pandas_output2.csv")

res = [Hyd,Mumbai,Chennai,blore,delhi,kol]

result = pd.concat(res)
result.info()

result.to_csv("C:\\Users\\kabra\\OneDrive - SP JAIN SCHOOL OF GLOBAL MANAGEMENT\\Desktop\\result.csv")
```

Data Pre-processing in python

REASONS TO USE TABLEAU:

- Tableau, being a data analytics tool, is a very effective tool to create interactive data visualizations. It is also a drag-and-drop software which makes the user interface very easy to use.
- This interactivity provided by the software enabled us to prototype different graphs much rapidly and understand the intricacies of the data.
- Tableau also easily handles large amounts of data and has readily available plots that help in understanding the data.
- It also takes almost any type of data file as an input and converts the raw data into easily understandable data.

REASONS TO USE MATPLOTLIB:

Matplotlib is a plotting library of python. It provides an object-oriented API for embedding plots into applications. It is a comprehensive library for creating static, animated and interactive visualizations.

Although Tableau is a great tool to visualize data, it does have certain restrictions and boundaries. For this very reason, we found Matplotlib to be able to overcome those restrictions and continue with our analysis.

Along with the use of Matplotlib, we had access to pandas, another library available in python which enabled us to carry out complicated filtering and transformations on the data.
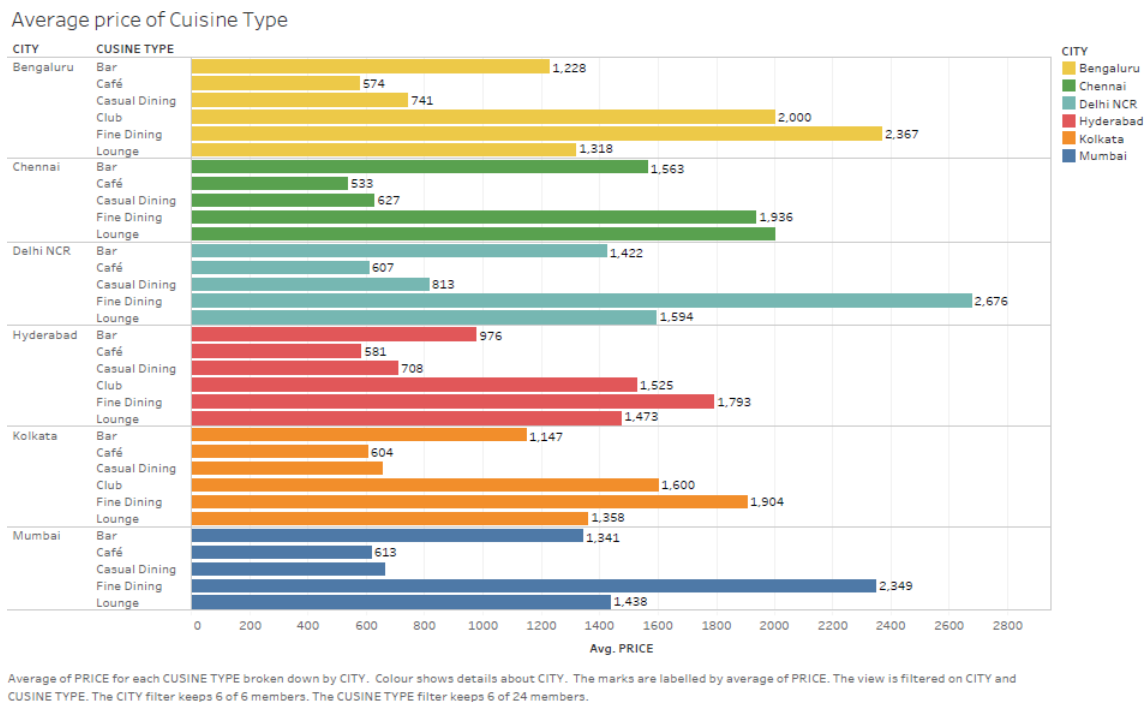
ANALYSIS OF DATA:

Average price of Cuisine Type

Average of PRICE for each CUSINE TYPE broken down by CITY. Colour shows details about CITY. The marks are labelled by average of PRICE. The view is filtered on CITY and CUSINE TYPE. The CITY filter keeps 6 of 6 members. The CUSINE TYPE filter keeps 6 of 24 members.

Figure 1 - TABLEAU

Tableau was used as the Business Intelligence tool for finding the average price for each cuisine type broken down by the cities. This provided insight on which city was the most expensive and least expensive for various types of cuisines served. For instance, the average price to be paid by two people in a bar in Bengaluru would be Rs 1228 but in Hyderabad would only be Rs 976.

Following are our observations for each cuisine Type:
1. Bar: The bar's in Chennai are the most expensive from the cities that have been analysed by us costing over Rs 1500 approximately for two people but Hyderabad on the other hand is very reasonable costing under Rs 1000 approximately for two people.
2. Café: As per our analysis, the cafés of all six cities cost between Rs 500 – Rs 600 approximately for two people. Even though the difference between each city is minute, the café's in Mumbai are moderately more expensive than the cafés of the other cities.
3. Casual Dining: The average price for a meal for two at a casual dining restaurant cost between Rs 600 – Rs 800 approximately. Delhi is the most expensive vis à vis Chennai.
4. Club: The clubs of only Hyderabad, Bengaluru and Kolkata have been listed in our data. Clubbing at Bengaluru is much more expensive than

clubbing at Hyderabad or Kolkata as they charge Rs 2000 approximately for two people.

5. Fine Dining: Delhi again is the most expensive place amongst the six cities charging an average of Rs 2600 for a fine dining meal for two approximately.

6. Lounge: The lounges of Chennai are the most expensive from the lot of six cities costing and average of Rs 2000 approximately for a meal for two.
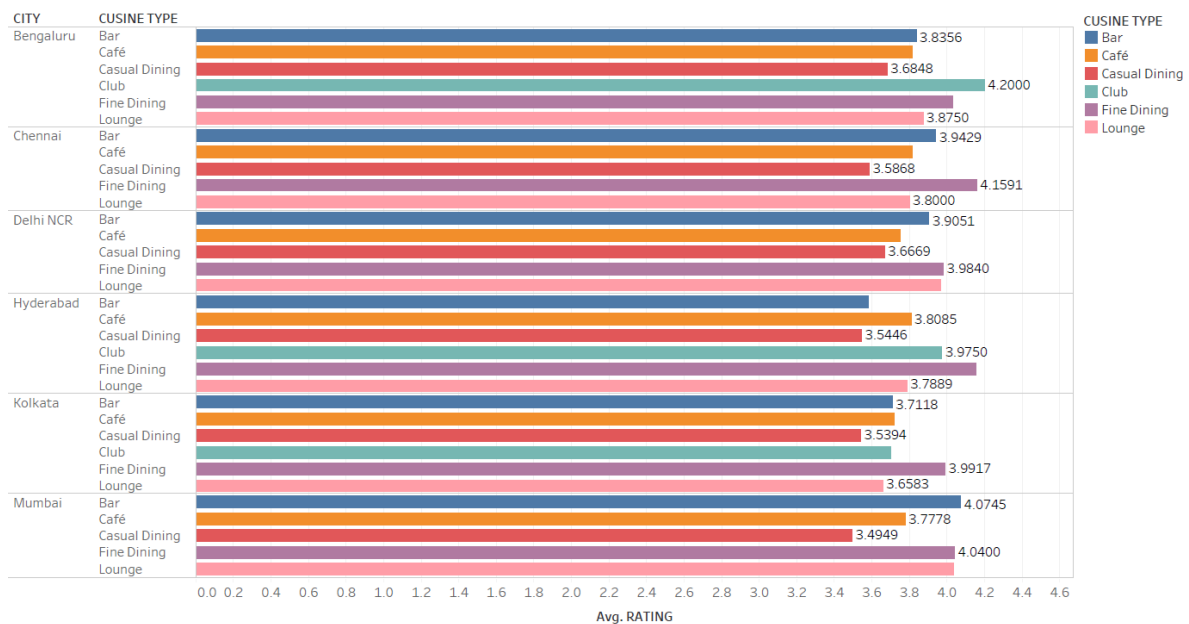
Average of RATING for each CUSINE TYPE broken down by CITY. Colour shows details about CUSINE TYPE. The marks are labelled by average of RATING. The view is filtered on CUSINE TYPE, which keeps 6 of 24 members.
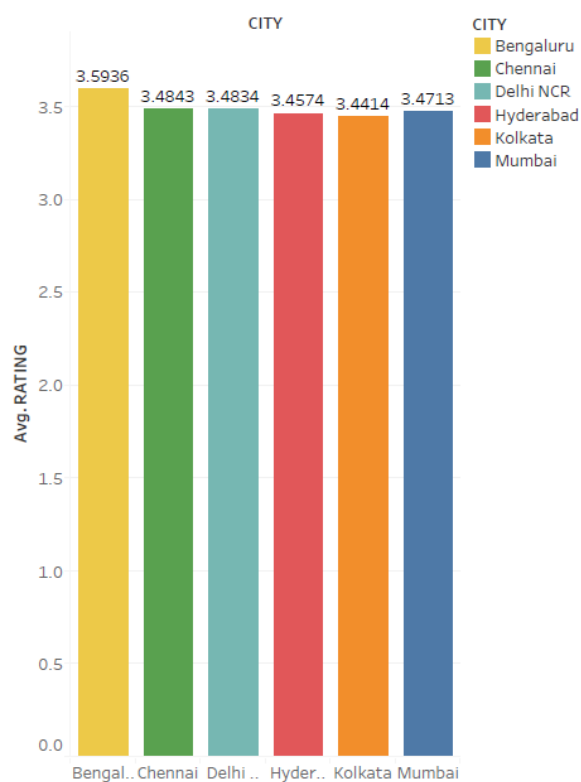
Figure 2 - TABLEAU

Tableau was used as the Business Intelligence tool for finding the average rating of each cuisine type broken down by the cities.

Following are our observation for each cuisine type:
1. Bar: It is observed that the rating for all the cities is similar. None of the cities have a very low or high rating. The city which has the highest rated bar is Mumbai i.e., 4.07 (out of 5) and Kolkata has the lowest rating i.e., 3.71.

2. Café: After looking at the above graph we can say that Delhi NCR has highest rated café.

3. Casual Dining: The rating ranges between 3.4 – 3.6 with Mumbai having a rating of 3.49 and Bengaluru having a rating of 3.68. Rest of the cities having an average rating of 3.5.

4. Club: As mentioned above, the clubs of Bengaluru, Hyderabad and Chennai have been only listed in our data. Among the clubs of these cities, Bengaluru holds the clubs with the highest rating whereas, Clubs of Kolkata again do not hold a great club rating.
5. Fine Dining: As per our analysis, Chennai has been rated the highest (4.15) among the cities, whereas, Delhi NCR has the lowest rating for a Fine Dining restaurant.
6. Lounge: Well, Lounges of Mumbai have the highest rating. Whereas Kolkata again holds a poor rating.

Average rating of all restaurants

Average of RATING for each CITY. Colour shows details about CITY. The marks are labelled by average of RATING.

Figure 3 - TABLEAU

Tableau was used as the Business Intelligence tool in Figure 3 for finding the average rating of all restaurants broken by city.

Figure 3 shows the overall rating of all the restaurants for each city. Based on the above graph, we can say that the Zomato affiliated restaurants of Bengaluru have the highest rating i.e., 3.59 (out of 5) and Mumbai has the lowest rating i.e., 3.47 (out of 5). It can be said that all the cities roughly fall under the $3.4 - 3.6$ rating belt.
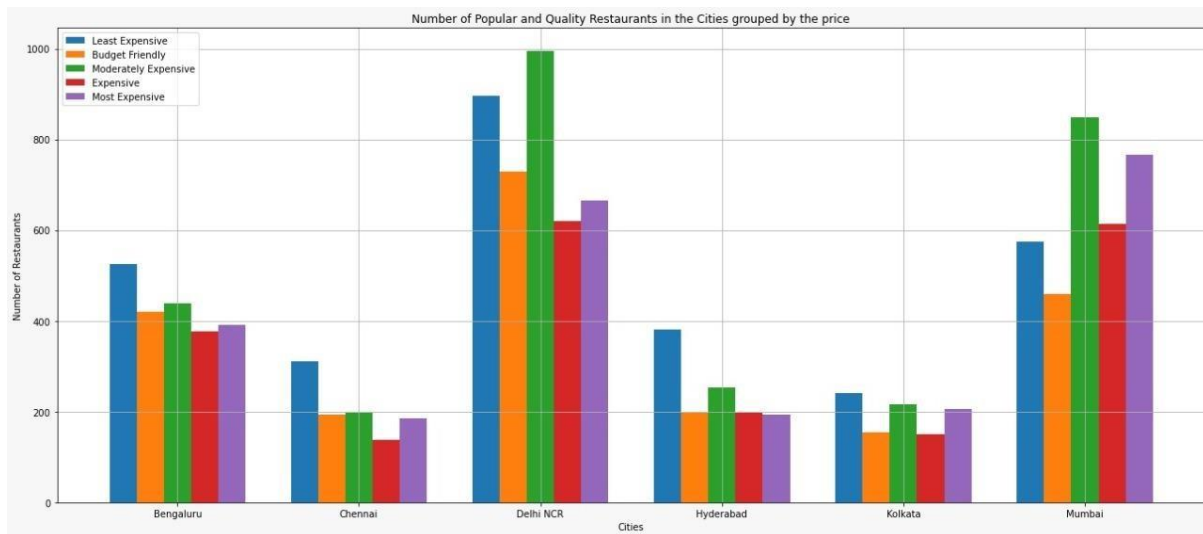
Figure 4 - MATPLOTLIB

Matplotlib (a library of python) was used in Figure 4 to plot the number of popular and Quality restaurants broken down by city and grouped by price.

The votes & rating were first filtered based on their median. The price was then divided into 5 quantiles and were named Least expensive, Budget Friendly, Moderately expensive, Expensive, and most Expensive.

As per Figure 4:
1. Bengaluru showed that major number of popular restaurants fell in the least expensive to moderately expensive section of the graph.
2. Chennai on the other hand majorly had popular restaurants that catered to the least expensive side of the graph.
3. Delhi showed that moderately expensive restaurants were more popular vis-à-vis least expensive or the budget friendly ones.
4. Hyderabad, like Chennai showed that least expensive restaurants were more popular as to the other categories.
5. Kolkata on the other hand saw almost equal number of restaurants catering to the least expensive, moderately expensive, and most expensive categories.
6. Mumbai showed that the restaurants which belonged to the moderately expensive belt to the most expensive belt were more in number.
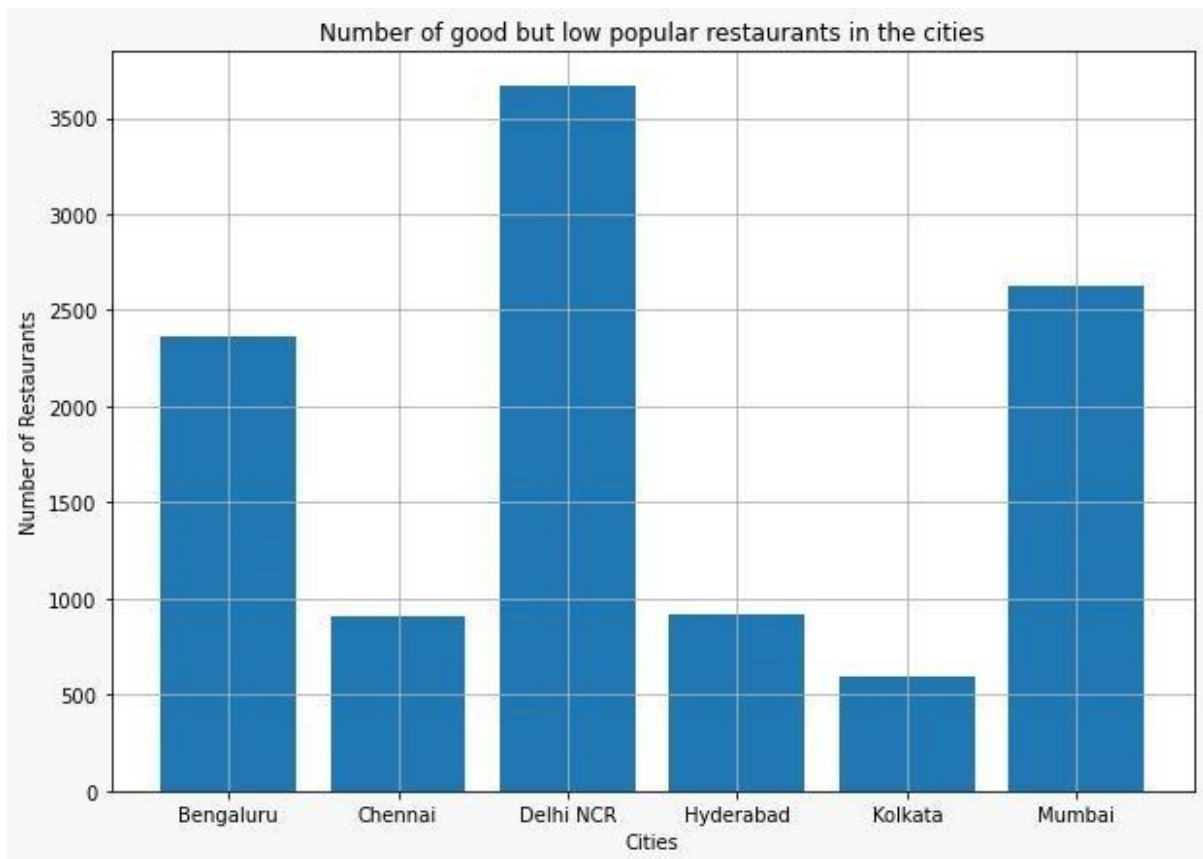
Figure 5 - MATPLOTLIB

Matplotlib (a library of python) was used in Figure 5 to plot the number of good but not as popular restaurants broken down by city.

As per Figure 5, Delhi NCR had the most restaurants that were good by rating but not as popular by votes. Mumbai and Bengaluru had above 2000 restaurants that were good but not as popular.
Chennai, Hyderabad, and Kolkata on the other hand did not have as many restaurants which belonged to such a category.

RECOMMENDATIONS

The CEO of Zomato would be pleased to know that we have found some insightful insights with the analysis shown in the previous section. Even though we focussed majorly on the TIER 1 cities, the side motive was to also implement this in various countries and cities to improve marketing, sales, cost etc.

Our analysis focussed on understanding the average price of various cuisine types, average rating of the restaurants for various cuisine types, average rating of all restaurants, number of popular and quality restaurants and number of unpopular but quality restaurants.

Average price and average rating of various cuisines types of each city helped us in understanding the type of income categories that were able to afford food in such restaurants. Even among Tier 1 cities, there were drastic differences in some of the prices between cities for the same cuisine type. Even though this does not speak enough on its own, when number of quality restaurants were assessed, every city had at least 100 restaurants that were popular and the rating given by the users spoke good about the quality of food.

When we looked at the unpopular restaurants but rating that assured quality, we saw a great spike in such number of restaurants. This spoke poorly on the marketing done by Zomato to help upcoming businesses that are not as popular. Restaurants that are popular usually have a fixed client base and would only need Zomato as a medium to deliver the food. But a lot of clients that would want to explore new restaurants would end up not finding the hidden potential of the upcoming restaurants due to poor marketing.

We feel that Zomato could also make more money if they looked at promoting the unpopular restaurants. A popular restaurant already has a brand image in the market whereas an unpopular restaurant is looking for its big break. To do so, they would be willing to find every opportunity to expand its customer base.

As per our analysis, we found that the number of restaurants that fall in the unpopular category are much more than the popular ones and hence if Zomato strongly recommended its users to order or explore the unpopular restaurants,

they would be able to make much more money. The not as popular restaurants would also be willing to provide more offers and schemes than an established brand making the marketing easier.

More importantly, the understanding of quality and popular restaurants in the cities enable Zomato to build a better recommendation engine. For example, based on a user's search history in their city, they can be associated with the class of restaurants they prefer, and this association can be used by recommending the similar restaurants to users who have the same preferences. Furthermore, this recommendation engine can be used to suggest restaurants to the user in other cities, should they ever find themselves travelling to other destinations.

By using the aforementioned strategies and model we strongly believe that Zomato would be able to expand and grow.

REFERENCES

The codes were all written using Python. These have been uploaded on Github. You can access them using https://github.com/tusharsonthalia/Data-Warehousing-Project.