Zomato Data Analysis

Introduction

Zomato API Analysis is one of the most useful analysis for foodies who want to taste the best cuisines of every part of the world which lies in their budget. This analysis is also for those who want to find the value for money restaurants in various parts of the country for the cuisines. Additionally, this analysis caters the needs of people who are striving to get the best cuisine of the country and which locality of that country serves that cuisines with maximum number of restaurants.

Requirement analysis

Dataset Details

- Restaurant Id: Unique id of every restaurant across various cities of the world
- Restaurant Name: Name of the restaurant
- Country Code: Country in which restaurant is located
- City: City in which restaurant is located
- Address: Address of the restaurant
- Locality: Location in the city

- Locality Verbose: Detailed description of the locality -_Longitude: Longitude coordinate of the restaurant’s location
- Latitude: Latitude coordinate of the restaurant’s location
- Cuisines: Cuisines offered by the restaurant
- Average Cost for two: Cost for two people in different currencies
- Currency: Currency of the country
- Has Table booking: yes/no
- Has Online delivery: yes/ no
- Is delivering: yes/ no
- Switch to order menu: yes/no
- Price range: range of price of food
- Aggregate Rating: Average rating out of 5
- Rating color: depending upon the average rating color
- Rating text: text on the basis of rating of rating
- Votes: Number of ratings casted by people

Software requirement specification

1. Importing Dataset
2. Exploratory Analysis and Visualization
3. List of countries the survey is spread across
4. Understanding the Rating aggregate, color and text
5. Country and Currency
6. Online delivery distribution
7. From which Locality maximum hotels are listed in Zomato
8. What kind of Cuisine these highly rates restaurants offer
9. How many of such restaurants accept online delivery
10. Understanding the Restaurants Rating localities.
11. Rating VS Cost of dinning
12. Location of Highly rated restaurants across New Delhi
13. Common Eateries

Analysis of Data

The dataset can be downloaded from Kaggle. I have build this project on google Colab. The dataset can be downloaded as follows

```
import pandas as pd

df = pd.read_csv('zomato.csv', encoding='ISO-8859-1')
```

```
df.head(2)
```

Reading the data using pandas

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitude	Latitude	Cuisines	...	Currency	Has Table booking	Has Online delivery
0	6317637	Le Petit Souffle	162	Makati City	Third Floor, Century City Mall, Kalayaan Avenu...	Century City Mall, Poblacion, Makati City	Century City Mall, Poblacion, Makati City, Mak...	121.027535	14.565443	French, Japanese, Desserts	...	Botswana Pula(P)	Yes	No
1	6304287	Izakaya Kikufuji	162	Makati City	Little Tokyo, 2277 Chino Roces Avenue, Legaspi...	Little Tokyo, Legaspi Village, Makati City	Little Tokyo, Legaspi Village, Makati City, Ma...	121.014101	14.553708	Japanese	...	Botswana Pula(P)	Yes	No

Checking if dataset contains any null

```
## Checking if dataset contains any null

nan_values = df.isna()
nan_columns = nan_values.any()
columns_with_nan = df.columns[nan_columns].tolist()
print(columns_with_nan)
```

Cuisines seems to contain null values. Hence any further analysis involving Cuisines the NaN values has to be considered.

There is an other file which is also available along with this dataset

```
df1 = pd.read_excel('/content/zomato-restaurants-data/Country-Code.xlsx')
df1.head()
```

	Country Code	Country
0	1	India
1	14	Australia
2	30	Brazil
3	37	Canada
4	94	Indonesia

Let us merge both the datasets. This will help us to understand the dataset country wise.

```
df2 = pd.merge(df,df1,on='Country Code',how='left')
df2.head(2)
```

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitude	Latitude	Cuisines	...	Currency	Has Table booking	Has Online delivery
0	6317637	Le Petit Souffle	162	Makati City	Third Floor, Century City Mall, Kalayaan Avenu...	Century City Mall, Poblacion, Makati City	Century City Mall, Poblacion, Makati City, Mak...	121.027535	14.565443	French, Japanese, Desserts	...	Botswana Pula(P)	Yes	No
1	6304287	Izakaya Kikufuji	162	Makati City	Little Tokyo, 2277 Chino Roces Avenue, Legaspi...	Little Tokyo, Legaspi Village, Makati City	Little Tokyo, Legaspi Village, Makati City, Ma...	121.014101	14.553708	Japanese	...	Botswana Pula(P)	Yes	No

Implementation

print('List of counteris the survey is spread accross - ')
for x in pd.unique(df2.Country): print(x)
print()
print('Total number to country', len(pd.unique(df2.Country)))

```
List of counteris the survey is spread accross -
Phillipines
Brazil
United States
Australia
Canada
Singapore
UAE
India
Indonesia
New Zealand
United Kingdom
Qatar
South Africa
Sri Lanka
Turkey

Total number to country 15
```

Understanding the Rating aggregate, color and text

df3 = df2.groupby(['Aggregate rating','Rating color', 'Rating text']).size().reset_index().rename(columns={0:'Rating Count'})
df3

	Aggregate rating	Rating color	Rating text	Rating Count
0	0.0	White	Not rated	2148
1	1.8	Red	Poor	1
2	1.9	Red	Poor	2
3	2.0	Red	Poor	7
4	2.1	Red	Poor	15
5	2.2	Red	Poor	27
6	2.3	Red	Poor	47
7	2.4	Red	Poor	87
8	2.5	Orange	Average	110
9	2.6	Orange	Average	191
10	2.7	Orange	Average	250
11	2.8	Orange	Average	315
12	2.9	Orange	Average	381
13	3.0	Orange	Average	468
14	3.1	Orange	Average	519
15	3.2	Orange	Average	522
16	3.3	Orange	Average	483
17	3.4	Orange	Average	498
18	3.5	Yellow	Good	480
19	3.6	Yellow	Good	458
20	3.7	Yellow	Good	427

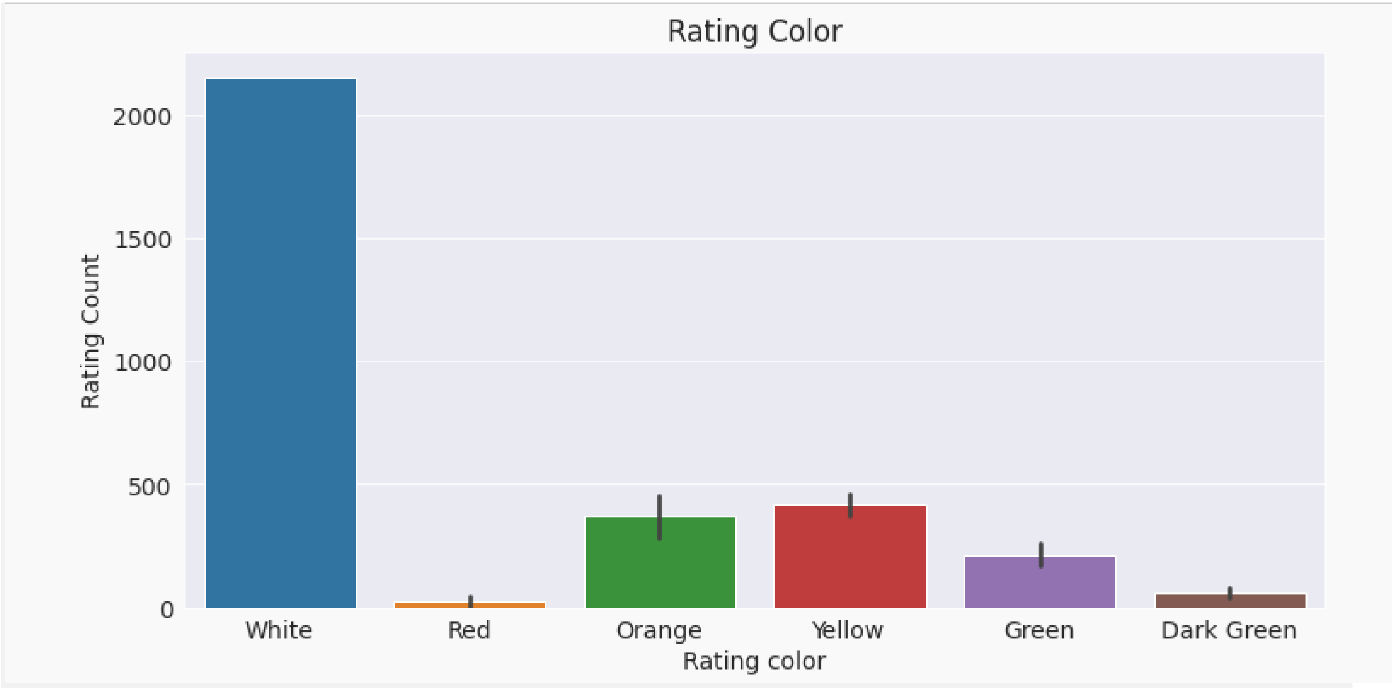
The above information helps us to understand the realation between Aggregate rating, color and text. We conclude the following color assigned to the ratings:

- Rating 0 — White — Not rated
- Rating 1.8 to 2.4 — Red — Poor

- Rating 2.5 to 3.4 — Orange — Average
- Rating 3.5 to 3.9 — Yellow — Good
- Rating 4.0 to 4.4 — Green — Very Good
- Rating 4.5 to 4.9 — Dark Green — Excellent

Let us try to understand the spread of rating across restaurants

```
import seaborn as sns
import matplotlib
import matplotlib.pyplot as plt
%matplotlib inline

sns.set_style('darkgrid')
matplotlib.rcParams['font.size'] = 14
matplotlib.rcParams['figure.figsize'] = (9, 5)
matplotlib.rcParams['figure.facecolor'] = '#00000000'

plt.figure(figsize=(12,6))
# plt.xticks(rotation=75)
plt.title('Rating Color')


sns.barplot(x=df3['Rating color'], y=df3['Rating Count']);
```



Interesting, Maximum restaurants seems to have gone No ratings. Let us check if these restaurants belong to some specific country.

```
No_rating = df2[df2['Rating color']=='White'].groupby('Country').size().reset_index().rename(columns={0:'Rating Count'})
No_rating
```

	Country	Rating Count
0	Brazil	5
1	India	2139
2	United Kingdom	1
3	United States	3

India seems to have maximum unrated restaurants. In India the culture of ordering online food is still gaining momentum hence most of the restaurants are still unrated on Zomato as people might be preferring to visiting the restaurant for a meal.

Country and Currency

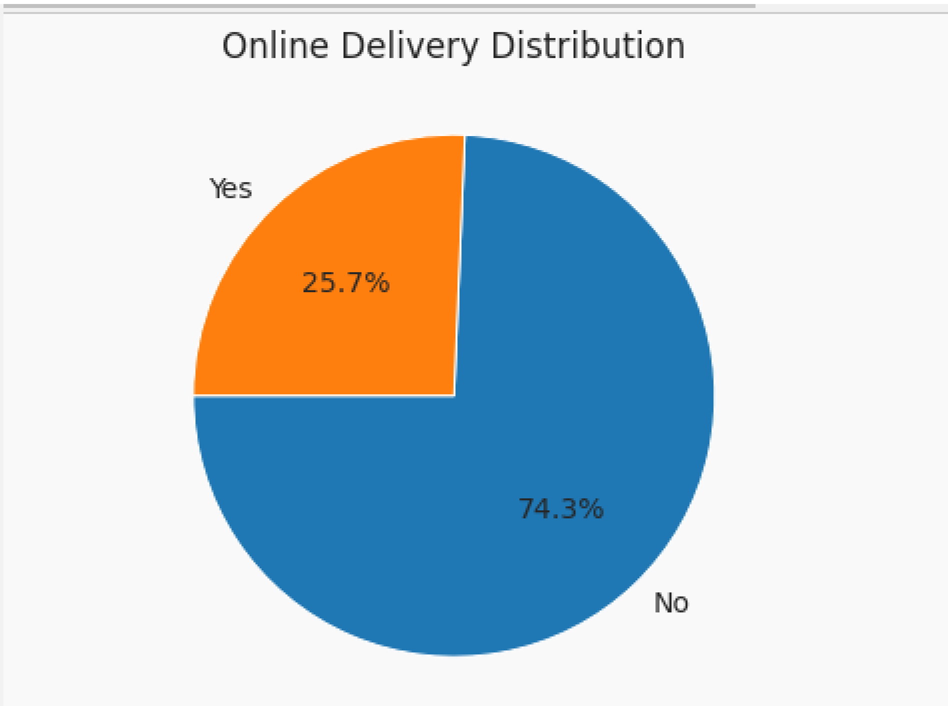
```
country_currency =
df2[['Country','Currency']].groupby(['Country','Currency']).size().reset_index(name='count').drop(
'count', axis=1, inplace=False)
country_currency.sort_values('Currency').reset_index(drop=True)
```

	Country	Currency
0	Phillipines	Botswana Pula(P)
1	Brazil	Brazilian Real(R$)
2	Australia	Dollar($)
3	Canada	Dollar($)
4	Singapore	Dollar($)
5	United States	Dollar($)
6	UAE	Emirati Diram(AED)
7	India	Indian Rupees(Rs.)
8	Indonesia	Indonesian Rupiah(IDR)
9	New Zealand	NewZealand($)
10	United Kingdom	Pounds(��)
11	Qatar	Qatari Rial(QR)
12	South Africa	Rand(R)
13	Sri Lanka	Sri Lankan Rupee(LKR)
14	Turkey	Turkish Lira(TL)

Above table display country and the currency they accept. Interestingly four countries seems to be accepting currency in dollars.

Online delivery distribution

```
plt.figure(figsize=(12,6))
plt.title('Online Delivery Distribution')


plt.pie(df2['Has Online delivery'].value_counts()/9551*100, labels=df2['Has Online
delivery'].value_counts().index, autopct='%1.1f%%', startangle=180);
```



Only 25% of restaurants accepts online delivery. This data might be biased as we have maximum restaurants listed here are from India. Maybe analysis over city wise would be more helpful.

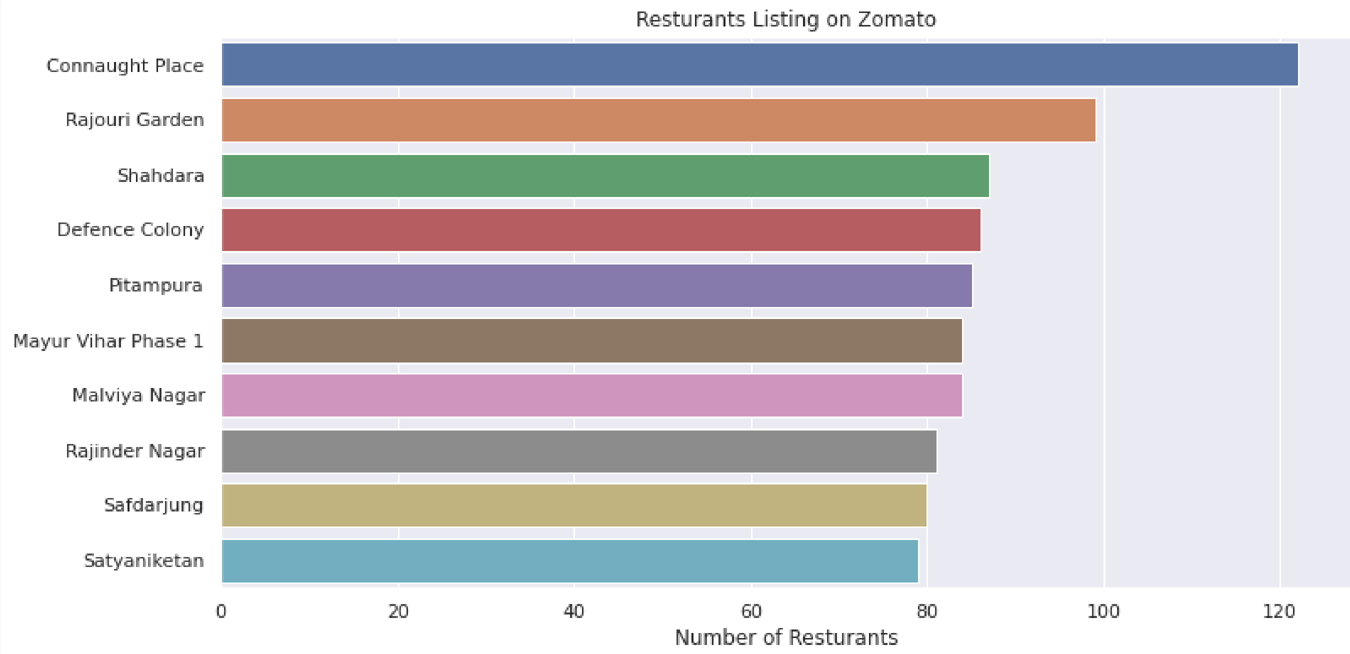
Let us try to understand the coverage of city

The data seems to be skewed towards New Delhi, Gurgaon and Noida. I see minimal data for other cities. Hence I would do my analysis predominantly on New Delhi.

From which Locality maximum hotels are listed in Zomato

```
Delhi = df2[(df2.City == 'New Delhi')]
plt.figure(figsize=(12,6))
sns.barplot(x=Delhi.Locality.value_counts().head(10),
y=Delhi.Locality.value_counts().head(10).index)
```

```
plt.ylabel(None);
plt.xlabel('Number of Resturants')
plt.title('Resturants Listing on Zomato');
```



Connaught place seems to have high no of restaurants registered with Zomato, Let us understand the cuisines the top rated restaurants has to offer

What kind of Cuisine these highly rates restaurants offer

```
# I achieve this by the following steps

## Fetching the resturants having 'Excellent' and 'Very Good' rating
ConnaughtPlace = Delhi[(Delhi.Locality.isin(['Connaught Place'])) & (Delhi['Rating text'].isin(['Excellent','Very Good']))]

ConnaughtPlace = ConnaughtPlace.Cuisines.value_counts().reset_index()

## Extracing all the cuisens in a single list
cuisien = []
for x in ConnaughtPlace['index']:
  cuisien.append(x)
```

cuisien = '[%s]'%', '.join(map(str, cuisien))

Cuisine

```
'North Indian, Italian, Asian, American',
'Continental, North Indian, Chinese, Mediterranean',
'Chinese',
'Continental, American, Asian, North Indian',
'North Indian, Continental',
'Continental, Mediterranean, Italian, North Indian',
'North Indian, European, Asian, Mediterranean',
'Japanese',
'Bakery, Desserts, Fast Food',
'North Indian, Afghani, Mughlai',
'Biryani, North Indian, Hyderabadi',
'Ice Cream',
'Continental, Mexican, Burger, American, Pizza, Tex-Mex',
'North Indian, Chinese',
'North Indian, Mediterranean, Asian, Fast Food',
'North Indian, European',
'Healthy Food, Continental, Italian',
'Continental, North Indian, Italian, Asian',
'Continental, Italian, Asian, Indian',
'Biryani, Hyderabadi',
'Bakery, Fast Food, Desserts',
'Italian, Mexican, Continental, North Indian, Finger Food',
'Cafe',
'Asian, North Indian',
'South Indian',
'Modern Indian',
'North Indian, Chinese, Continental, Italian',
'North Indian, Chinese, Italian, American, Middle Eastern',
'Fast Food, American, Burger']
```

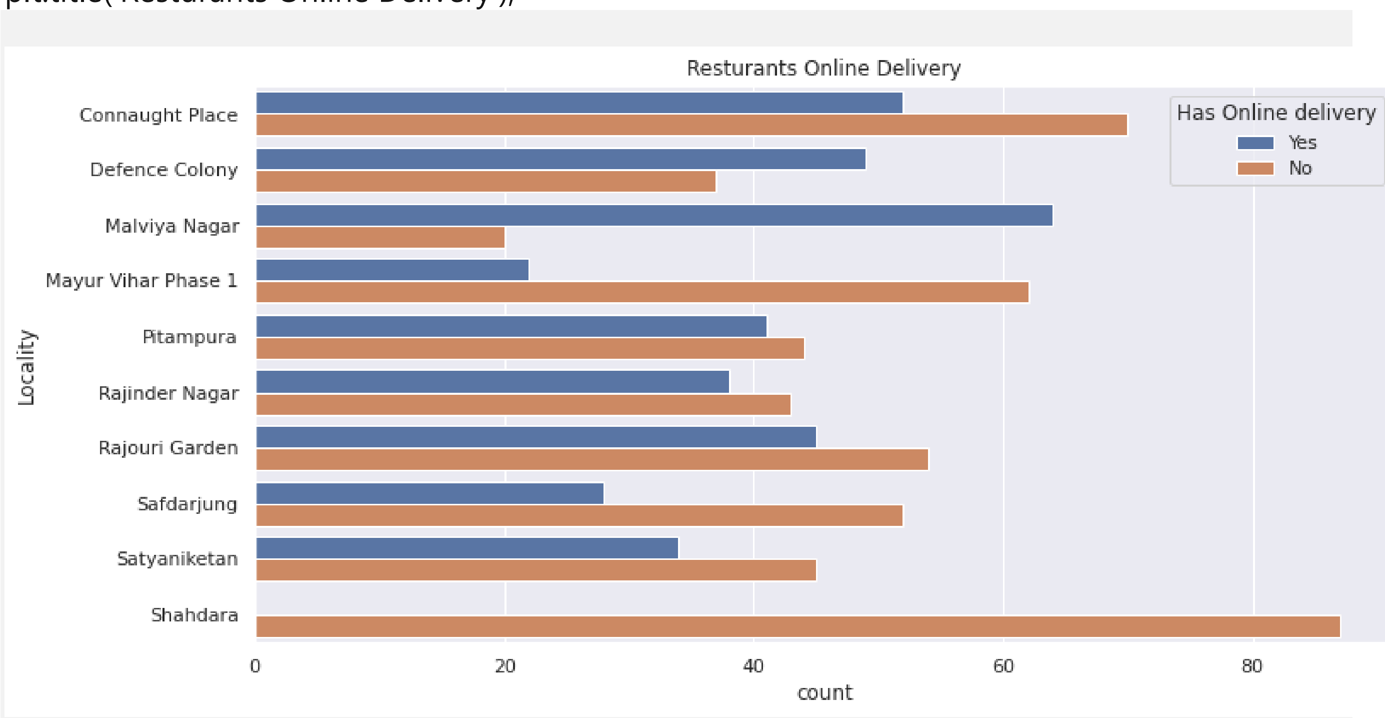
Top rated restaurants seems to be doing well in the following cuisine

- North Indian

- Chinese
- Italian
- American

How many of such restaurants accept online delivery

```
top_locality = Delhi.Locality.value_counts().head(10)
sns.set_theme(style="darkgrid")
plt.figure(figsize=(12,6))
ax = sns.countplot(y= "Locality", hue="Has Online delivery", data=Delhi[Delhi.Locality.isin(top_locality.index)])


plt.title('Resturants Online Delivery');
```



Apart from Shahdara locality, restaurants in other locality accepts online delivery.

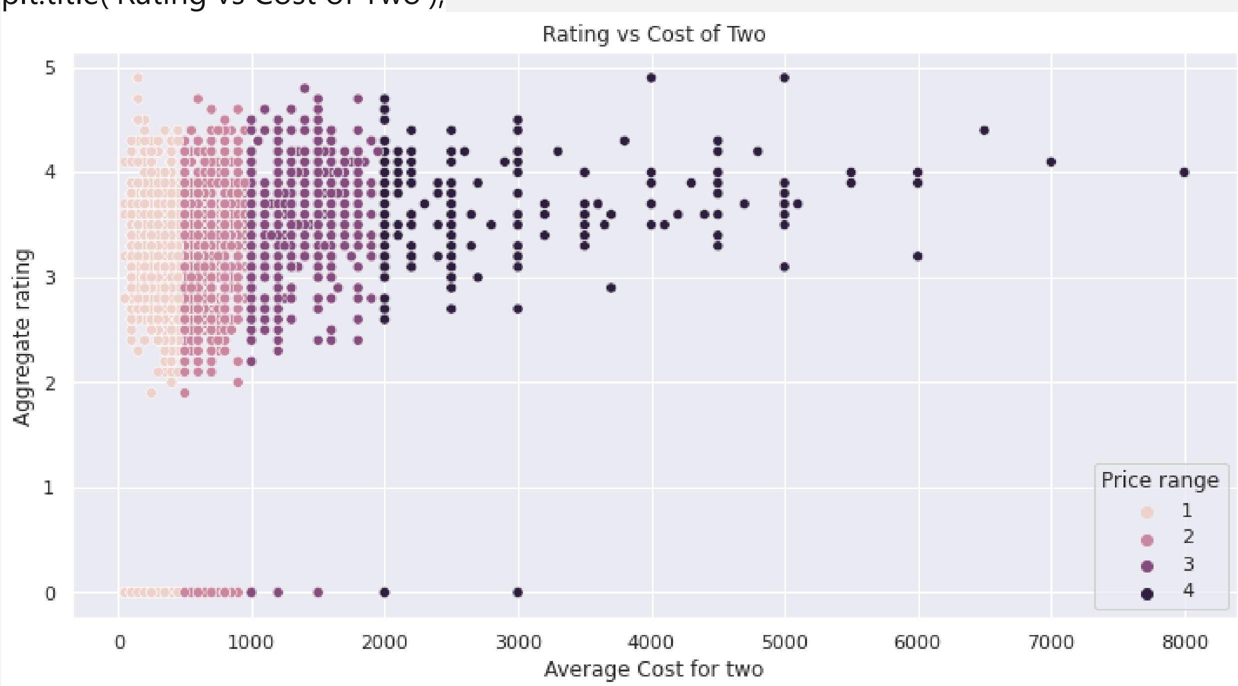
Online Delivery seems to be on higher side in Defence colony and Malviya Nagar

Defence colony seems to have high no of highly rated restaurants but Malviya Nagar seems to done better in terms of Good and Average restaurants.

As restaurants with ‘Poor’ and ‘Not Rated’ is far lesser that ‘Good’, ‘Very Good’ and ‘Excellent’ restaurants. Hence people in these localities prefer online ordering

Rating VS Cost of dinning

```
plt.figure(figsize=(12,6))
sns.scatterplot(x="Average Cost for two", y="Aggregate rating", hue='Price range', data=Delhi)
plt.xlabel("Average Cost for two")
plt.ylabel("Aggregate rating")
plt.title('Rating vs Cost of Two');
```



I observe there is no linear relation between price and rating. For instance, Restaurants with good rating (like 4–5) have restaurants with all the price range and spread across the entire X axis

Location of Highly rated restaurants across New Delhi

```
Delhi['Rating text'].value_counts()
```

```
Average      2495
Not rated    1425
Good         1128
Very Good     300
Poor           97
Excellent      28
Name: Rating text, dtype: int64
```

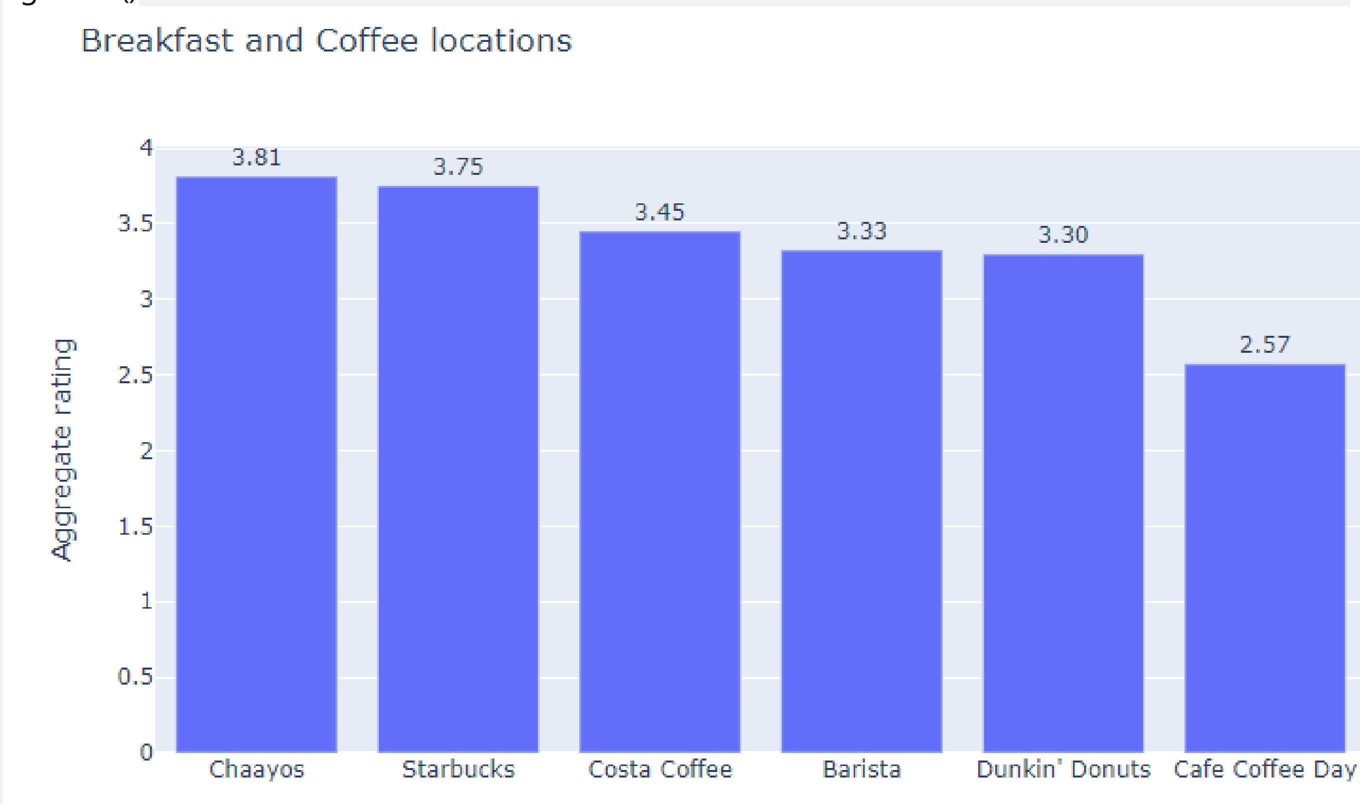
The aforementioned four cities represent nearly 65% of the total data available in the dataset. Apart from the higly rated local restaurants, it'd be intersting to know where the known-eateries that are commonplace. The verticles across which these can be located are -

- Breakfast
- American Fast Food
- Ice Creams, Shakes & Desserts

Common Eateries

Breakfast and Coffee locations

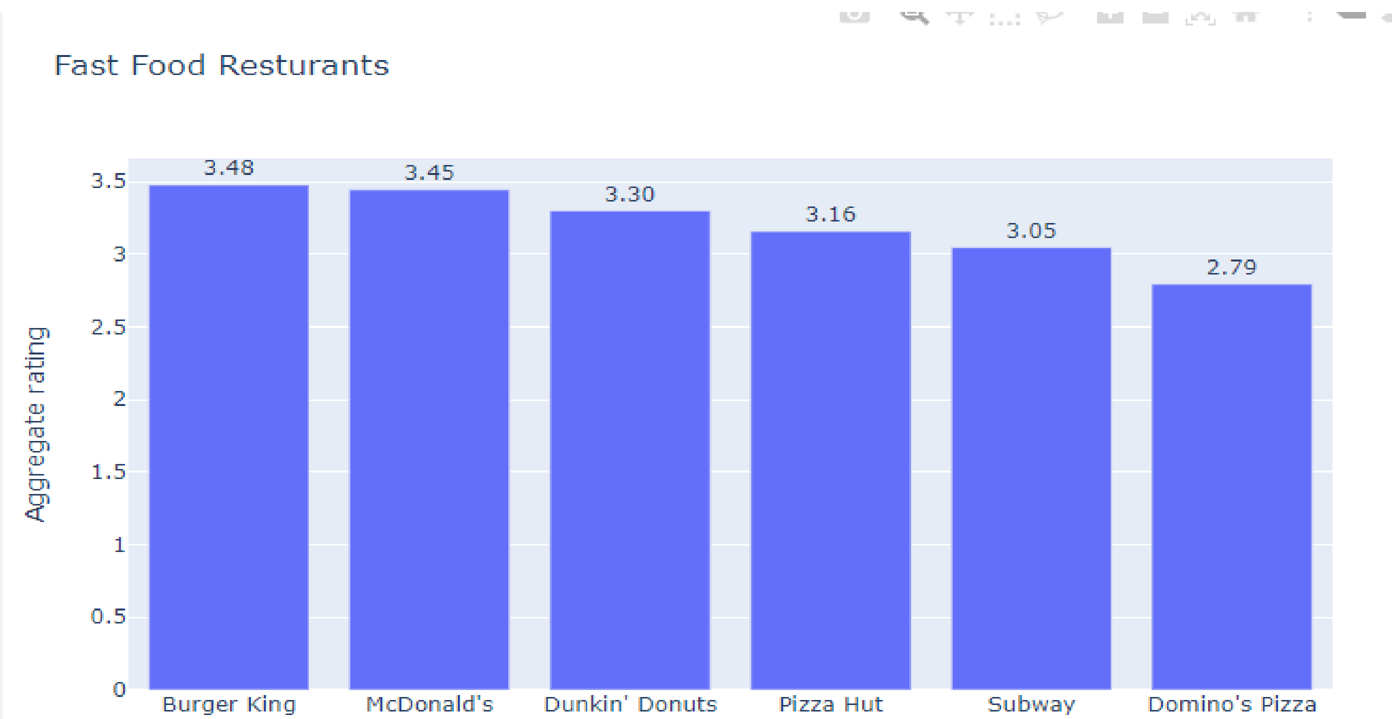
```
import plotly.express as px

df= breakfast
fig = px.bar(df, y='Aggregate rating', x='Restaurant Name', text='Aggregate rating',
title="Breakfast and Coffee locations")
fig.update_traces(texttemplate='%{text:.3s}', textposition='outside')
fig.update_layout(
    autosize=False,
    width=800,
    height=500,)
fig.show()
```



Chaayos outlets are doing better. We need more of those in Delhi. Café coffee day seems to be performing poorly in avg rating. They are required to improve their services.

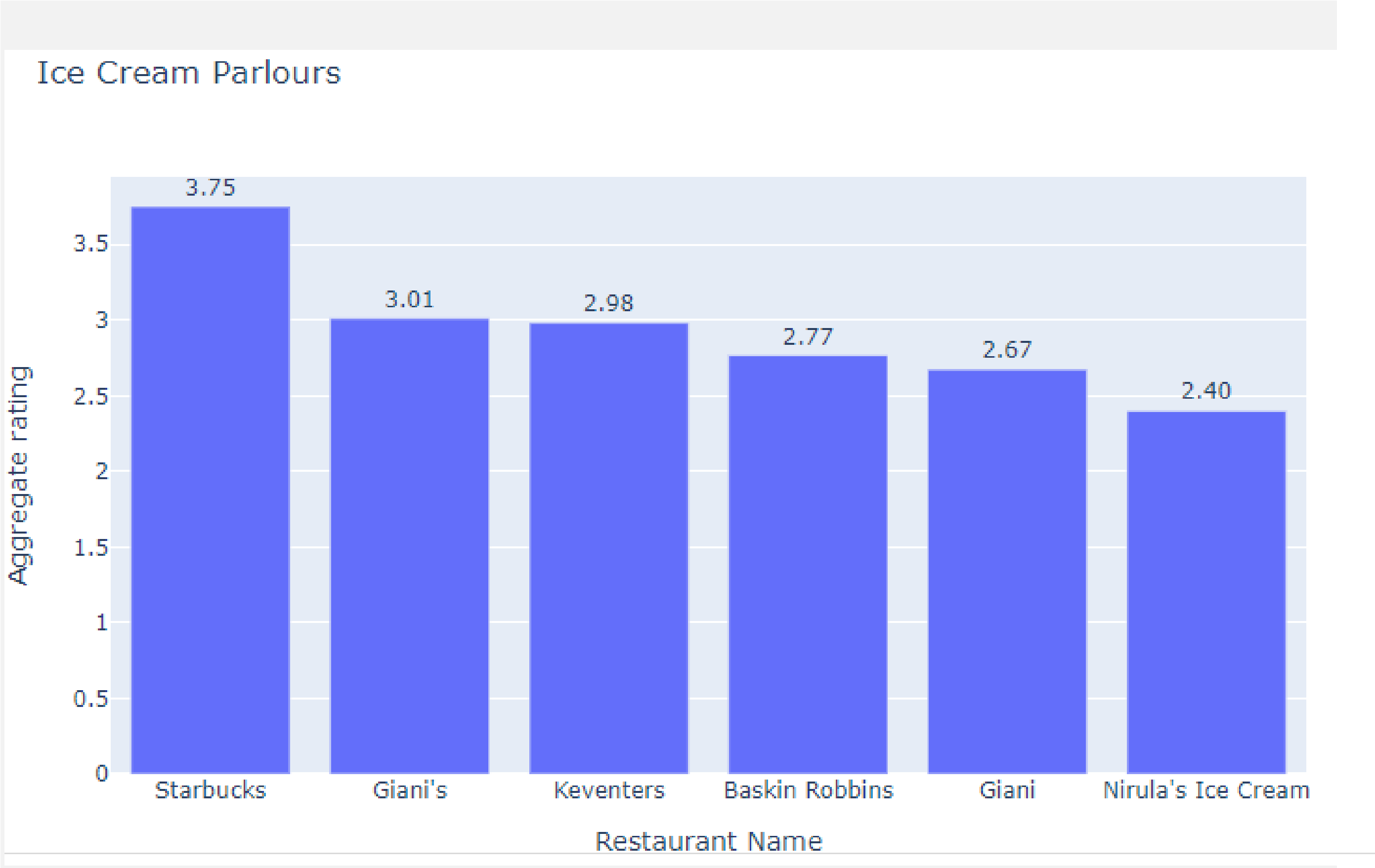
Fast Food Restaurants

```
import plotly.express as px

df= american
fig = px.bar(df, y='Aggregate rating', x='Restaurant Name', text='Aggregate rating',
title="Fast Food Resturants")
fig.update_traces(texttemplate='%{text:.3s}', textposition='outside')
fig.update_layout(
    autosize=False,
    width=800,
    height=500,)

fig.show()
```



Ice Cream Parlors

```
import plotly.express as px

df= ice_cream
fig = px.bar(df, y='Aggregate rating', x='Restaurant Name', text='Aggregate rating',
title="Ice Cream Parlours")
fig.update_traces(texttemplate='%{text:.3s}', textposition='outside')
fig.update_layout(
    autosize=False,
    width=800,
    height=500,)
fig.show()
```



Foreign brands seems to be doing better than the local brands

Inferences and Conclusions

We've drawn many inferences from the survey. Here's a summary of a few of them:

- The dataset is skewed towards India and doesn't represent the complete data of restaurants worldwide.
- Restaurants rating is categorized in six categories

1. Not Rated

2. Average
3. Good
4. Very Good
5. Excellent

- Connaught Palace have maximum restaurants listed on Zomato but in terms of online delivery acceptance Defence colony and Malviya nagar seems to be doing better.
- The top rated restaurants seems to be getting better rating on the following cuisine

1. North Indian
2. Chinese
3. American
4. Italian

- There is no relation between cost and rating. Some of the best rated restaurants are low on cost and vice versa.
- On common Eateries, For Breakfast and Coffee location Indian restaurants seems to be better rated but for Fast food chain and Ice cream parlors American restaurants seems to be doing better.

References and Future Work

Check out the following resources to learn more about the dataset and tools used in this notebook:

- Zomato Restaurants Data: https://www.kaggle.com/shrutimehta/zomato-restaurants-data
- Pandas user guide: https://pandas.pydata.org/docs/user_guide/index.html
- Matplotlib user guide: https://matplotlib.org/3.3.1/users/index.html
- Seaborn user guide & tutorial: https://seaborn.pydata.org/tutorial.html
- opendatasets Python library: https://github.com/JovianML/opendatasets