Zomato Data Analysis

June 18, 2022

1 Import Basic Libraries

[41]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

[42]:
```
#Load the dataset
df=pd.read_csv("zomato.csv",encoding='latin-1')
```

[43]:
```
df.head()
```

[43]:

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitude	Latitude
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
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
2	6300002	Heat - Edsa Shangri-La	162	Mandaluyong City	Edsa Shangri-La, 1 Garden Way, Ortigas, Mandal…	Edsa Shangri-La, Ortigas, Mandaluyong City	Edsa Shangri-La, Ortigas, Mandaluyong City, Ma…	121.056831	14.581404
3	6318506	Ooma	162	Mandaluyong City	Third Floor, Mega Fashion Hall, SM Megamall, O…	SM Megamall, Ortigas, Mandaluyong City			
4	6314302	Sambo Kojin	162	Mandaluyong City	Third Floor, Mega Atrium, SM Megamall, Ortigas…	SM Megamall, Ortigas, Mandaluyong City			

```
3  SM Megamall, Ortigas, Mandaluyong City, Mandal…  121.056475  14.585318
4  SM Megamall, Ortigas, Mandaluyong City, Mandal…  121.057508  14.584450

                          Cuisines  …          Currency Has Table booking  \
0       French, Japanese, Desserts  …  Botswana Pula(P)               Yes
1                         Japanese  …  Botswana Pula(P)               Yes
2  Seafood, Asian, Filipino, Indian  …  Botswana Pula(P)               Yes
3                  Japanese, Sushi  …  Botswana Pula(P)                No
4                 Japanese, Korean  …  Botswana Pula(P)               Yes

  Has Online delivery Is delivering now Switch to order menu Price range  \
0                  No                No                    No           3
1                  No                No                    No           3
2                  No                No                    No           4
3                  No                No                    No           4
4                  No                No                    No           4

   Aggregate rating  Rating color Rating text Votes
0               4.8    Dark Green   Excellent   314
1               4.5    Dark Green   Excellent   591
2               4.4         Green   Very Good   270
3               4.9    Dark Green   Excellent   365
4               4.8    Dark Green   Excellent   229

[5 rows x 21 columns]
```

2 Exploratery Data Analysis

[44]:
```
#Data Cleaning
df.isnull().sum()
```

[44]:
```
Restaurant ID         0
Restaurant Name       0
Country Code          0
City                  0
Address               0
Locality              0
Locality Verbose      0
Longitude             0
Latitude              0
Cuisines              9
Average Cost for two  0
Currency              0
Has Table booking     0
Has Online delivery   0
Is delivering now     0
```

```
Switch to order menu    0
Price range             0
Aggregate rating        0
Rating color            0
Rating text             0
Votes                   0
dtype: int64
```

[45]:
```
df.columns
```

[45]:
```
Index(['Restaurant ID', 'Restaurant Name', 'Country Code', 'City', 'Address',
       'Locality', 'Locality Verbose', 'Longitude', 'Latitude', 'Cuisines',
       'Average Cost for two', 'Currency', 'Has Table booking',
       'Has Online delivery', 'Is delivering now', 'Switch to order menu',
       'Price range', 'Aggregate rating', 'Rating color', 'Rating text',
       'Votes'],
      dtype='object')
```

[46]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9551 entries, 0 to 9550
Data columns (total 21 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Restaurant ID         9551 non-null   int64
 1   Restaurant Name       9551 non-null   object
 2   Country Code          9551 non-null   int64
 3   City                  9551 non-null   object
 4   Address               9551 non-null   object
 5   Locality              9551 non-null   object
 6   Locality Verbose      9551 non-null   object
 7   Longitude             9551 non-null   float64
 8   Latitude              9551 non-null   float64
 9   Cuisines              9542 non-null   object
 10  Average Cost for two  9551 non-null   int64
 11  Currency              9551 non-null   object
 12  Has Table booking     9551 non-null   object
 13  Has Online delivery   9551 non-null   object
 14  Is delivering now     9551 non-null   object
 15  Switch to order menu  9551 non-null   object
 16  Price range           9551 non-null   int64
 17  Aggregate rating      9551 non-null   float64
 18  Rating color          9551 non-null   object
 19  Rating text           9551 non-null   object
 20  Votes                 9551 non-null   int64
dtypes: float64(3), int64(5), object(13)
memory usage: 1.5+ MB
```

[47]:
```
df.describe().T
```

[47]:

	count	mean	std	min	\
Restaurant ID	9551.0	9.051128e+06	8.791521e+06	53.000000	
Country Code	9551.0	1.836562e+01	5.675055e+01	1.000000	
Longitude	9551.0	6.412657e+01	4.146706e+01	-157.948486	
Latitude	9551.0	2.585438e+01	1.100794e+01	-41.330428	
Average Cost for two	9551.0	1.199211e+03	1.612118e+04	0.000000	
Price range	9551.0	1.804837e+00	9.056088e-01	1.000000	
Aggregate rating	9551.0	2.666370e+00	1.516378e+00	0.000000	
Votes	9551.0	1.569097e+02	4.301691e+02	0.000000	

	25%	50%	75%	max
Restaurant ID	301962.500000	6.004089e+06	1.835229e+07	1.850065e+07
Country Code	1.000000	1.000000e+00	1.000000e+00	2.160000e+02
Longitude	77.081343	7.719196e+01	7.728201e+01	1.748321e+02
Latitude	28.478713	2.857047e+01	2.864276e+01	5.597698e+01
Average Cost for two	250.000000	4.000000e+02	7.000000e+02	8.000000e+05
Price range	1.000000	2.000000e+00	2.000000e+00	4.000000e+00
Aggregate rating	2.500000	3.200000e+00	3.700000e+00	4.900000e+00
Votes	5.000000	3.100000e+01	1.310000e+02	1.093400e+04

[48]:
```
df.isnull().sum()
```

[48]:

Restaurant ID	0
Restaurant Name	0
Country Code	0
City	0
Address	0
Locality	0
Locality Verbose	0
Longitude	0
Latitude	0
Cuisines	9
Average Cost for two	0
Currency	0
Has Table booking	0
Has Online delivery	0
Is delivering now	0
Switch to order menu	0
Price range	0
Aggregate rating	0
Rating color	0
Rating text	0
Votes	0

dtype: int64

[49]:
```
#Drop the null values
df.dropna(subset=['Cuisines'],inplace=True)
```

[50]:
```
#Load another Dataset
df_country=pd.read_csv("Country-Code.csv")
```

[51]:
```
df_country.head()
```

[51]:

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

[52]:
```
#Rename the CountryCode to Country Code ,Tomerge both the dataset
df.rename(columns={"CountryCode":"Country Code"},inplace=True)
```

[53]:
```
#Merging datasets
final_df=pd.merge(df,df_country,on="Country Code",how='left')
```

[54]:
```
final_df.head()
```

[54]:

	Restaurant ID	Restaurant Name	Country Code	City	\
0	6317637	Le Petit Souffle	162	Makati City	
1	6304287	Izakaya Kikufuji	162	Makati City	
2	6300002	Heat - Edsa Shangri-La	162	Mandaluyong City	
3	6318506	Ooma	162	Mandaluyong City	
4	6314302	Sambo Kojin	162	Mandaluyong City	

	Address	\
0	Third Floor, Century City Mall, Kalayaan Avenu…	
1	Little Tokyo, 2277 Chino Roces Avenue, Legaspi…	
2	Edsa Shangri-La, 1 Garden Way, Ortigas, Mandal…	
3	Third Floor, Mega Fashion Hall, SM Megamall, O…	
4	Third Floor, Mega Atrium, SM Megamall, Ortigas…	

	Locality	\
0	Century City Mall, Poblacion, Makati City	
1	Little Tokyo, Legaspi Village, Makati City	
2	Edsa Shangri-La, Ortigas, Mandaluyong City	
3	SM Megamall, Ortigas, Mandaluyong City	
4	SM Megamall, Ortigas, Mandaluyong City	

	Locality Verbose	Longitude	Latitude	\
0	Century City Mall, Poblacion, Makati City, Mak…	121.027535	14.565443	
1	Little Tokyo, Legaspi Village, Makati City, Ma…	121.014101	14.553708	

```
2  Edsa Shangri-La, Ortigas, Mandaluyong City, Ma…  121.056831  14.581404
3  SM Megamall, Ortigas, Mandaluyong City, Mandal…  121.056475  14.585318
4  SM Megamall, Ortigas, Mandaluyong City, Mandal…  121.057508  14.584450

                          Cuisines  …  Has Table booking  \
0       French, Japanese, Desserts  …                Yes
1                         Japanese  …                Yes
2  Seafood, Asian, Filipino, Indian  …                Yes
3                  Japanese, Sushi  …                 No
4                 Japanese, Korean  …                Yes

  Has Online delivery Is delivering now Switch to order menu Price range  \
0                  No                No                    No           3
1                  No                No                    No           3
2                  No                No                    No           4
3                  No                No                    No           4
4                  No                No                    No           4

  Aggregate rating  Rating color  Rating text Votes      Country
0              4.8    Dark Green    Excellent   314  Phillipines
1              4.5    Dark Green    Excellent   591  Phillipines
2              4.4         Green    Very Good   270  Phillipines
3              4.9    Dark Green    Excellent   365  Phillipines
4              4.8    Dark Green    Excellent   229  Phillipines

[5 rows x 22 columns]
```

[55]:
```
final_df.columns
```

[55]:
```
Index(['Restaurant ID', 'Restaurant Name', 'Country Code', 'City', 'Address',
       'Locality', 'Locality Verbose', 'Longitude', 'Latitude', 'Cuisines',
       'Average Cost for two', 'Currency', 'Has Table booking',
       'Has Online delivery', 'Is delivering now', 'Switch to order menu',
       'Price range', 'Aggregate rating', 'Rating color', 'Rating text',
       'Votes', 'Country'],
      dtype='object')
```

3 Which are the Top 3 coutries using Zomato

[56]:
```
country_name=final_df.Country.value_counts().index
```

[57]:
```
country_name
```

[57]:
```
Index(['India', 'United States', 'United Kingdom', 'Brazil', 'South Africa',
       'UAE', 'New Zealand', 'Turkey', 'Australia', 'Phillipines', 'Indonesia',
       'Qatar', 'Singapore', 'Sri Lanka', 'Canada'],
```

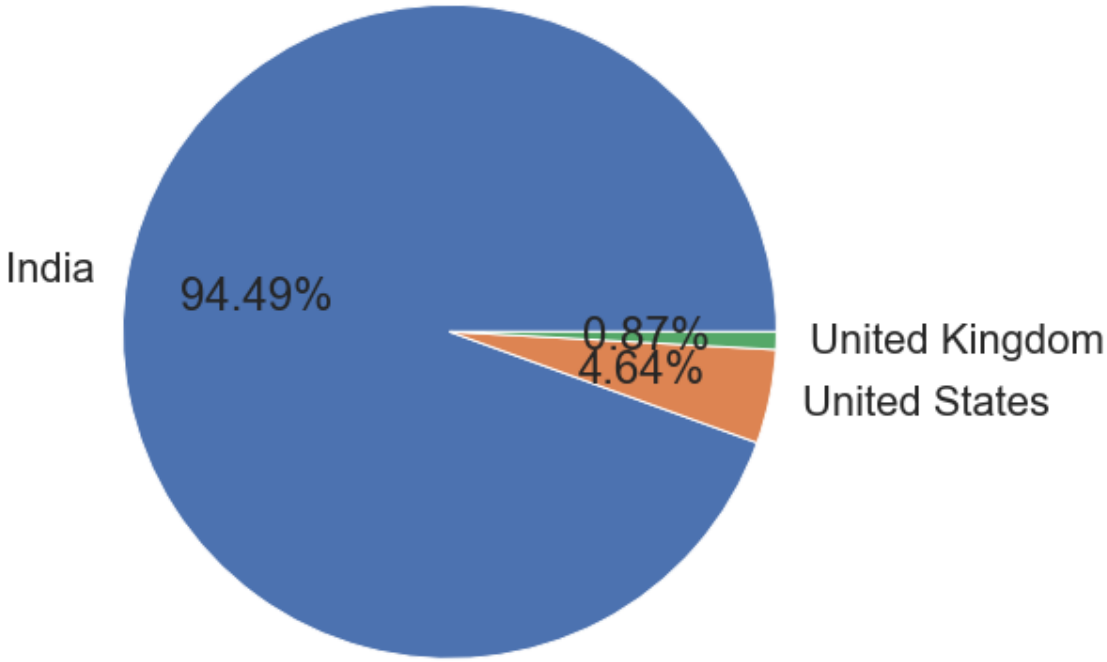
```
      dtype='object')
```

[58]:
```
country_val=final_df.Country.value_counts().values
```

[59]:
```
country_val
```

[59]:
```
array([8652,  425,   80,   60,   60,   60,   40,   34,   24,   22,   21,
         20,   20,   20,    4], dtype=int64)
```

[60]:
```
#top 3 countries uses zomato
plt.figure(figsize=(8,8))
sns.set(font_scale=2)
plt.pie(country_val[:3],labels=country_name[:3],autopct="%1.2f%%")
plt.show()
```



zomato maximum records or transaction are from india

4 How the Aggregate rating is associated with Rating Color and Rating text

[61]:
```
final_df.groupby(["Aggregate rating","Rating color","Rating text"]).size()
```

[61]:
```
Aggregate rating  Rating color  Rating text
0.0               White         Not rated      2148
1.8               Red           Poor              1
1.9               Red           Poor              2
2.0               Red           Poor              7
2.1               Red           Poor             15
2.2               Red           Poor             27
2.3               Red           Poor             47
2.4               Red           Poor             87
2.5               Orange        Average         110
2.6               Orange        Average         191
2.7               Orange        Average         250
2.8               Orange        Average         315
2.9               Orange        Average         381
3.0               Orange        Average         468
3.1               Orange        Average         519
3.2               Orange        Average         522
3.3               Orange        Average         483
3.4               Orange        Average         495
3.5               Yellow        Good            480
3.6               Yellow        Good            458
3.7               Yellow        Good            427
3.8               Yellow        Good            399
3.9               Yellow        Good            332
4.0               Green         Very Good       266
4.1               Green         Very Good       274
4.2               Green         Very Good       221
4.3               Green         Very Good       174
4.4               Green         Very Good       143
4.5               Dark Green    Excellent        95
4.6               Dark Green    Excellent        78
4.7               Dark Green    Excellent        41
4.8               Dark Green    Excellent        25
4.9               Dark Green    Excellent        61
dtype: int64
```

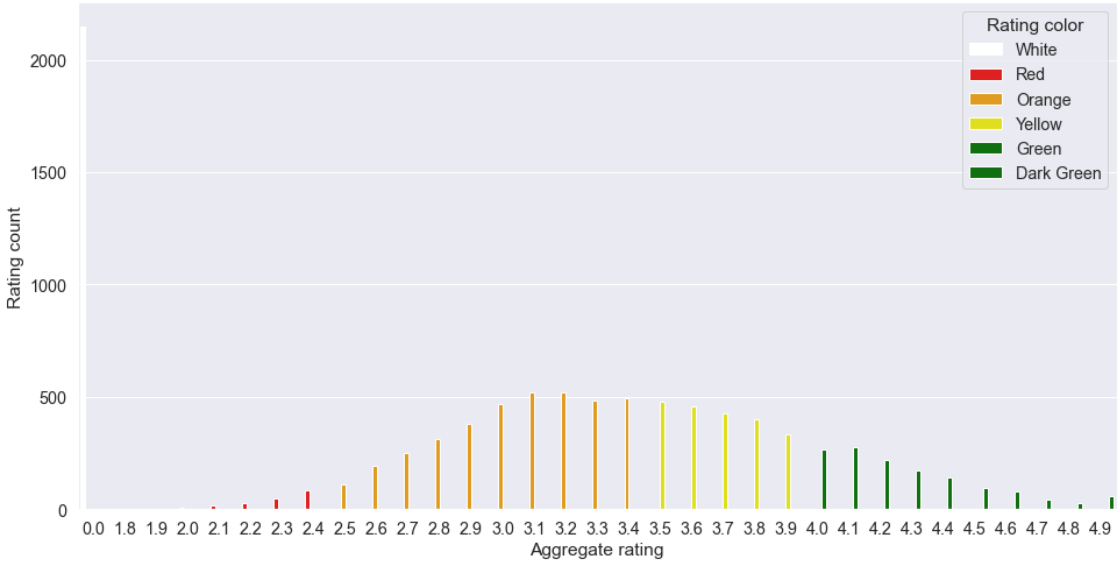
[62]:
```
final_df.groupby(["Aggregate rating","Rating color","Rating text"]).size().
↪reset_index()
```

[62]:
```
    Aggregate rating Rating color Rating text     0
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   495
18                3.5       Yellow        Good   480
19                3.6       Yellow        Good   458
20                3.7       Yellow        Good   427
21                3.8       Yellow        Good   399
22                3.9       Yellow        Good   332
23                4.0        Green   Very Good   266
24                4.1        Green   Very Good   274
25                4.2        Green   Very Good   221
26                4.3        Green   Very Good   174
27                4.4        Green   Very Good   143
28                4.5   Dark Green   Excellent    95
29                4.6   Dark Green   Excellent    78
30                4.7   Dark Green   Excellent    41
31                4.8   Dark Green   Excellent    25
32                4.9   Dark Green   Excellent    61
```

[63]:
```
ratings=final_df.groupby(["Aggregate rating","Rating color","Rating text"]).
↪size().reset_index().rename(columns={0:"Rating count"})
```

[64]:
```
plt.figure(figsize=(16,8))
sns.set_style('dark')
sns.set(font_scale=1.3)
sns.barplot(x="Aggregate rating",y="Rating count",hue="Rating␣
↪color",palette=["white","red","orange","yellow","green","green"],data=ratings)
plt.show()
```

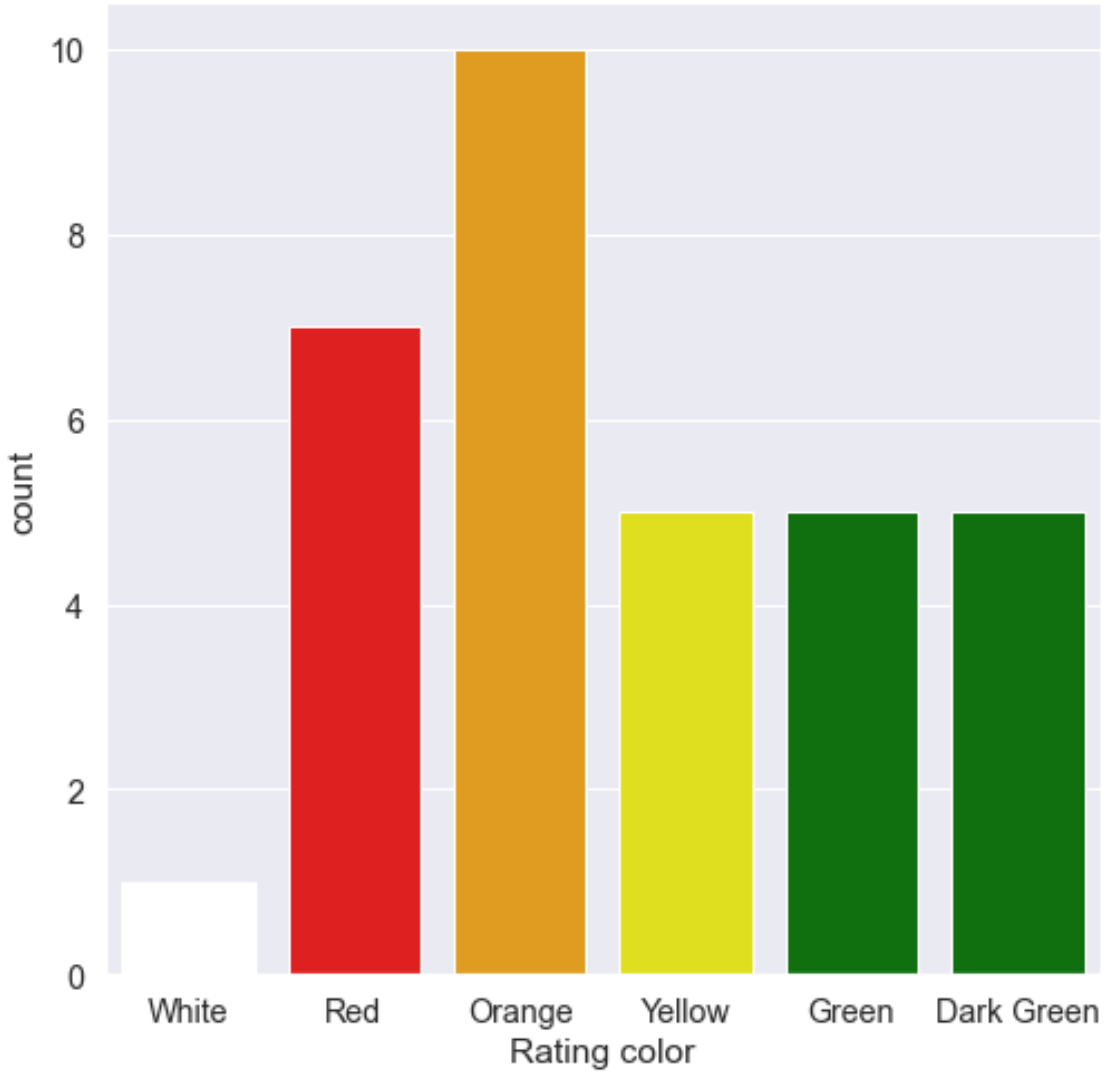


Most of the people did not rated

5 Which rating color is mostly used in this dataset

[65]:
```
plt.figure(figsize=(8,8))
sns.countplot(x="Rating␣
 ↪color",data=ratings,palette=["white","red","orange","yellow","green","green"])
```

[65]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1c326be0e80>
```



Orange colour is mostly used in Rating color

Most of the people rated 'Average' or between 2.9 to 3.4

6 Which Country gave maximum 0 ratings

[66]:
```
final_df[final_df["Rating color"]=="White"].groupby("Country").size().
↪reset_index()
```

[66]:

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

maximun no. of 0 rating is from india

[67]:
```
final_df.head(2)
```

[67]:
```
   Restaurant ID   Restaurant Name  Country Code         City  \
0        6317637  Le Petit Souffle           162  Makati City
1        6304287  Izakaya Kikufuji           162  Makati City

                                              Address  \
0  Third Floor, Century City Mall, Kalayaan Avenu…
1  Little Tokyo, 2277 Chino Roces Avenue, Legaspi…

                                     Locality  \
0   Century City Mall, Poblacion, Makati City
1  Little Tokyo, Legaspi Village, Makati City

                                    Locality Verbose   Longitude   Latitude  \
0  Century City Mall, Poblacion, Makati City, Mak…  121.027535  14.565443
1  Little Tokyo, Legaspi Village, Makati City, Ma…  121.014101  14.553708

                     Cuisines  …  Has Table booking Has Online delivery  \
0  French, Japanese, Desserts  …                Yes                  No
1                    Japanese  …                Yes                  No

  Is delivering now Switch to order menu Price range Aggregate rating  \
0                No                   No           3              4.8
1                No                   No           3              4.5

   Rating color  Rating text Votes      Country
0    Dark Green    Excellent   314  Phillipines
1    Dark Green    Excellent   591  Phillipines

[2 rows x 22 columns]
```

[68]:
```
final_df.columns
```

[68]:
```
Index(['Restaurant ID', 'Restaurant Name', 'Country Code', 'City', 'Address',
       'Locality', 'Locality Verbose', 'Longitude', 'Latitude', 'Cuisines',
       'Average Cost for two', 'Currency', 'Has Table booking',
       'Has Online delivery', 'Is delivering now', 'Switch to order menu',
       'Price range', 'Aggregate rating', 'Rating color', 'Rating text',
       'Votes', 'Country'],
      dtype='object')
```

7 Show the Currencies owned by each Countrty

[69]:
```
final_df.groupby(["Country",'Currency']).size().reset_index()
```

[69]:

	Country	Currency	0
0	Australia	Dollar($)	24
1	Brazil	Brazilian Real(R$)	60
2	Canada	Dollar($)	4
3	India	Indian Rupees(Rs.)	8652
4	Indonesia	Indonesian Rupiah(IDR)	21
5	New Zealand	NewZealand($)	40
6	Phillipines	Botswana Pula(P)	22
7	Qatar	Qatari Rial(QR)	20
8	Singapore	Dollar($)	20
9	South Africa	Rand(R)	60
10	Sri Lanka	Sri Lankan Rupee(LKR)	20
11	Turkey	Turkish Lira(TL)	34
12	UAE	Emirati Diram(AED)	60
13	United Kingdom	Pounds(£)	80
14	United States	Dollar($)	425

8 Which country has online delivery

[70]:
```
final_df[final_df['Has Online delivery']=="Yes"].Country.value_counts()
```

[70]:
```
India    2423
UAE        28
Name: Country, dtype: int64
```

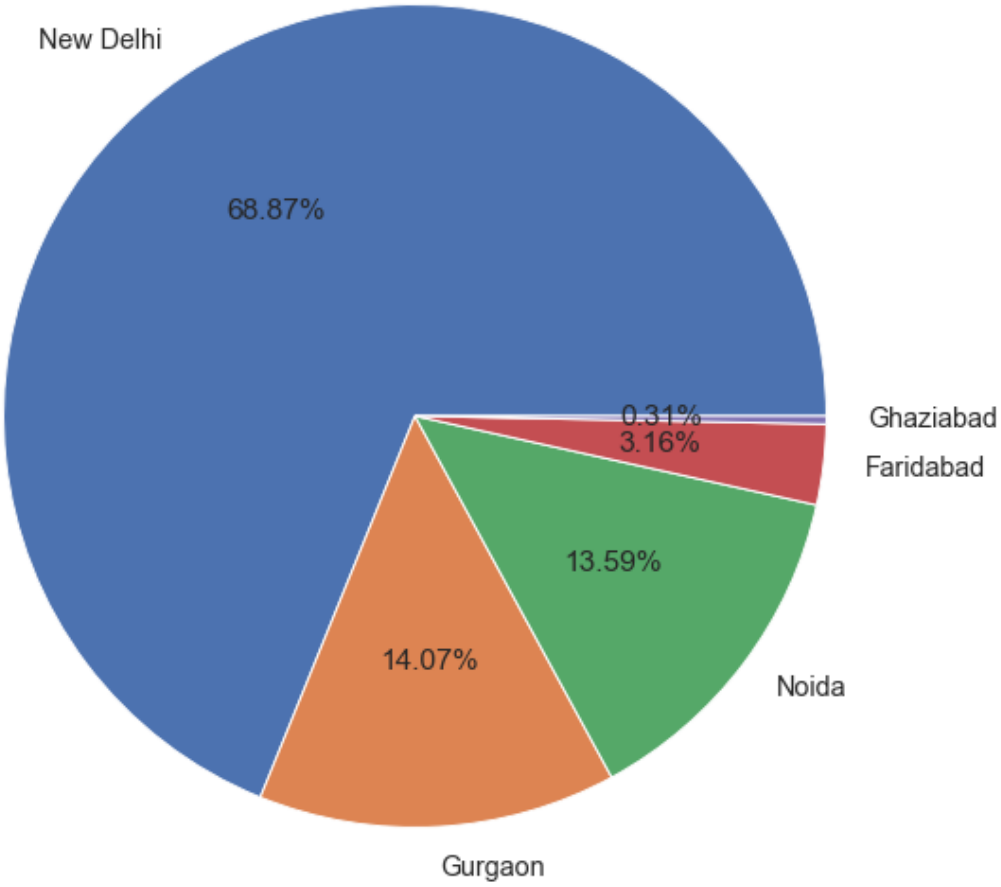
9 Which City has maximum restaurants in India

[71]:
```
final_df.City.value_counts().index
```

[71]:
```
Index(['New Delhi', 'Gurgaon', 'Noida', 'Faridabad', 'Ghaziabad', 'Lucknow',
       'Bhubaneshwar', 'Ahmedabad', 'Amritsar', 'Guwahati',
       …
       'Potrero', 'Fernley', 'Penola', 'Inverloch', 'Balingup', 'Panchkula',
       'Palm Cove', 'Dicky Beach', 'Winchester Bay', 'Bandung'],
      dtype='object', length=140)
```

[72]:
```
city_values=final_df.City.value_counts().values
city_labels=final_df.City.value_counts().index
```

[73]:
```
plt.figure(figsize=(10,10))
plt.pie(city_values[:5],labels=city_labels[:5],autopct='%1.2f%%')
plt.show()
```



New Delhi has maximum Restaurants

[74]:
```
df.corr()
```

[74]:

	Restaurant ID	Country Code	Longitude	Latitude	\
Restaurant ID	1.000000	0.146270	-0.224362	-0.052626	
Country Code	0.146270	1.000000	-0.694629	0.018049	
Longitude	-0.224362	-0.694629	1.000000	0.045415	
Latitude	-0.052626	0.018049	0.045415	1.000000	
Average Cost for two	-0.001629	0.043717	0.045948	-0.111080	
Price range	-0.134528	0.245363	-0.080257	-0.166735	
Aggregate rating	-0.327160	0.281295	-0.114733	0.000197	
Votes	-0.147434	0.154361	-0.084371	-0.022914	

	Average Cost for two	Price range	Aggregate rating	\

```
Restaurant ID                -0.001629    -0.134528         -0.327160
Country Code                  0.043717     0.245363          0.281295
Longitude                     0.045948    -0.080257         -0.114733
Latitude                     -0.111080    -0.166735          0.000197
Average Cost for two          1.000000     0.075111          0.051864
Price range                   0.075111     1.000000          0.438356
Aggregate rating              0.051864     0.438356          1.000000
Votes                         0.067833     0.309474          0.313474
```

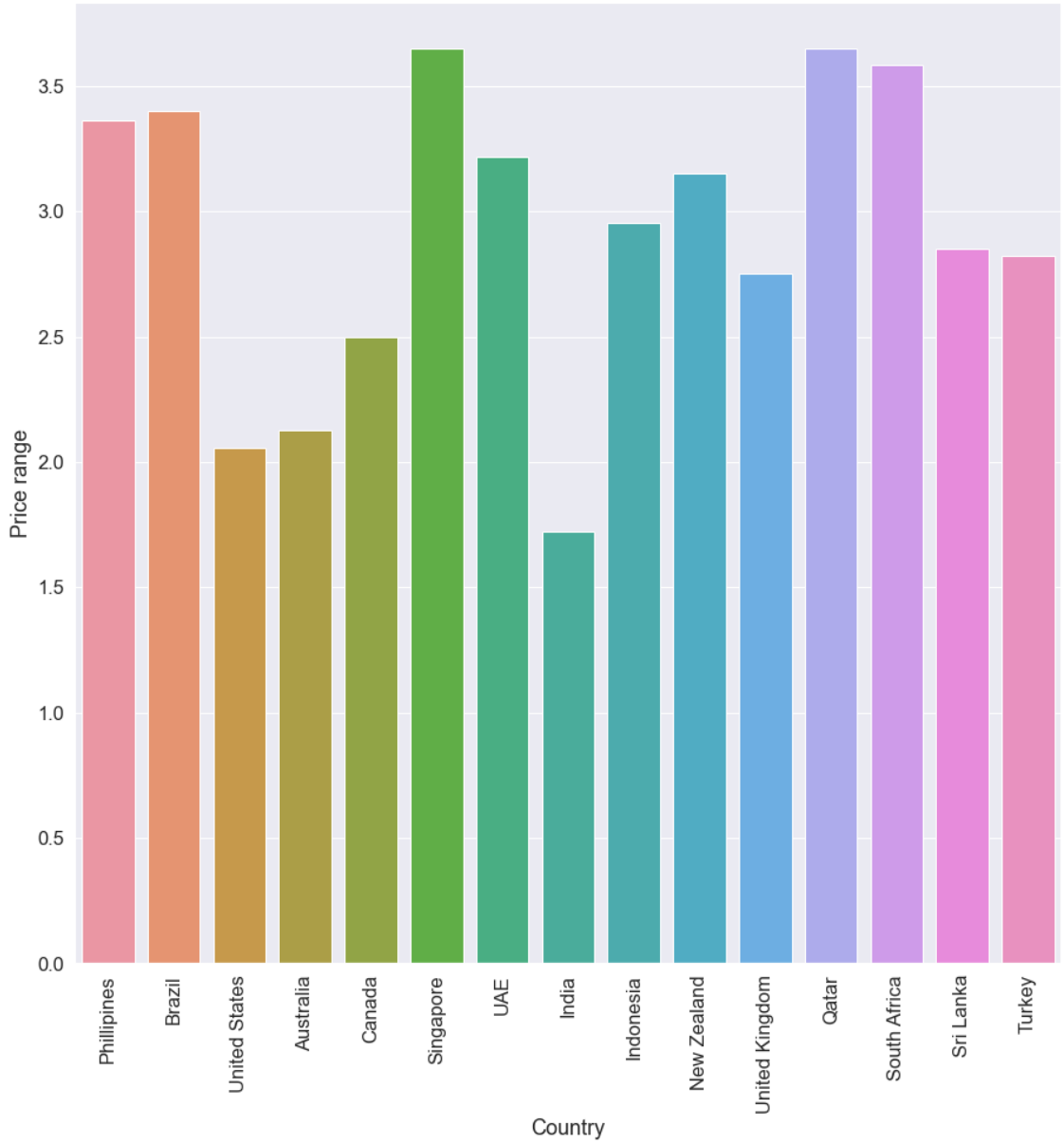
```
                          Votes
Restaurant ID        -0.147434
Country Code          0.154361
Longitude            -0.084371
Latitude             -0.022914
Average Cost for two  0.067833
Price range           0.309474
Aggregate rating      0.313474
Votes                 1.000000
```

[75]:
```
final_df['Locality'].unique()
```

[75]:
```
array(['Century City Mall, Poblacion, Makati City',
       'Little Tokyo, Legaspi Village, Makati City',
       'Edsa Shangri-La, Ortigas, Mandaluyong City', …, 'Ko\x81ôuyolu',
       'Kuruí_e\x81ôme', 'Moda'], dtype=object)
```

10 Which country has cheapest price of foods

[76]:
```
plt.figure(figsize=(15,15))
sns.set(font_scale=1.5)
sns.barplot(x='Country',y='Price range',data=final_df,ci=None)
plt.xticks(rotation=90)
plt.show()
```



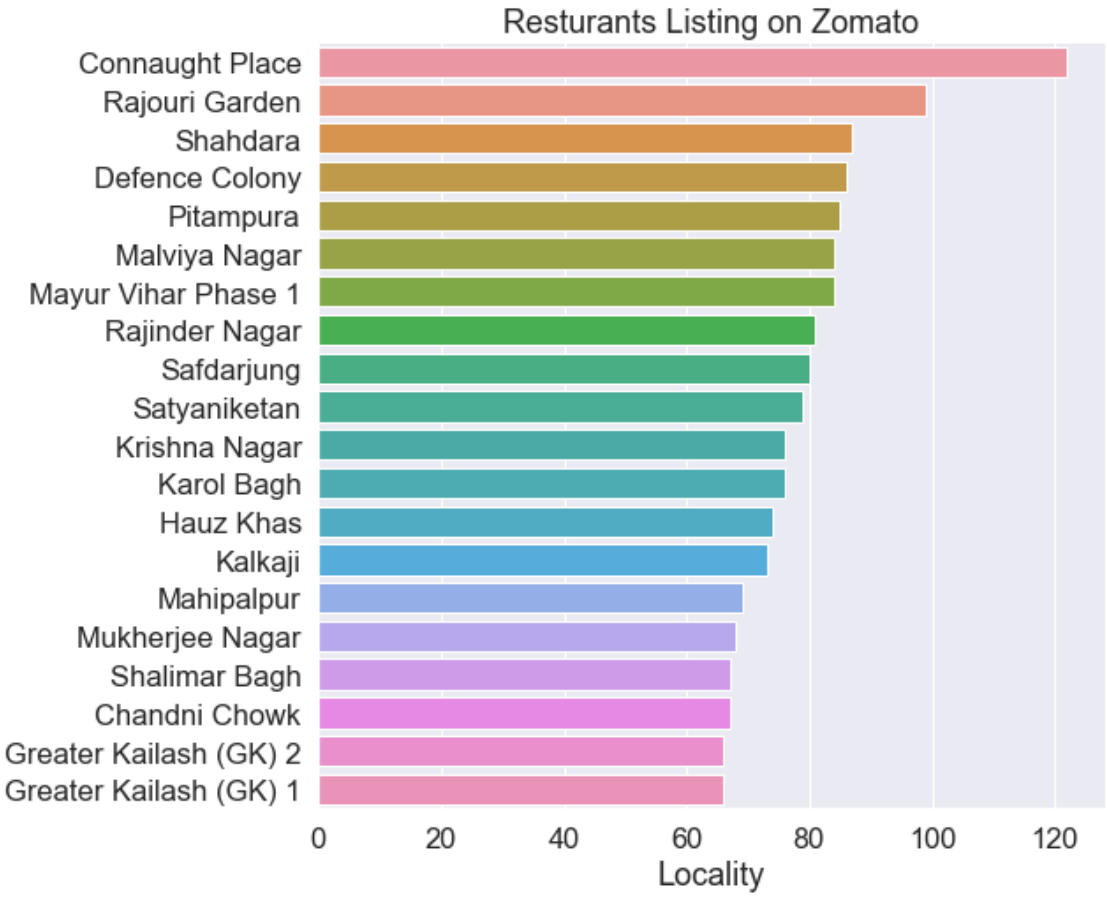
India has cheapest price of food

[77]:
```
#which locality has highest price range
```

11 From which Locality in Delhi maximum hotels are listed in Zomato

[78]:
```
Delhi=final_df[final_df["City"]=="New Delhi"]
```

[79]:
```
plt.figure(figsize=(8,8))
sns.barplot(x=Delhi.Locality.value_counts().head(20), y=Delhi.Locality.
 ↪value_counts().head(20).index)
plt.title('Resturants Listing on Zomato')
plt.show()
```



Connaught Place has maximum no. of Restaurants in Delhi

12 Which Kind of cuisiens are famous in Connaught Place

[80]:
```
## Fetching the resturants having 'Excellent' rating
ConnaughtPlace = Delhi[(Delhi.Locality.isin(['Connaught Place'])) &␣
 ↪(Delhi['Rating text'].isin(['Excellent']))]
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
```

[80]: ['Ice Cream', 'Continental, Italian, Asian, Indian', 'North Indian']

13 Result

Advice to other restaurants

Include 'Continental, Italian, Asian, Indian', 'North Indian', 'Ice Cream' In the Cuisines to increase their Sale

14 Thank You