Zomato_Analysis(Data_Analysis_Project)

May 21, 2024

[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

[2]:
```
df = pd.read_csv('Zomatodataset/zomato.csv',encoding='latin-1')
df.head()
```

[2]:

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitude	Latitude	Cuisines	…	Currency	Has Table booking	\
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	SM Megamall, Ortigas, Mandaluyong City, Mandal…	121.056475	14.585318					
4	6314302	Sambo Kojin	162	Mandaluyong City	Third Floor, Mega Atrium, SM Megamall, Ortigas…	SM Megamall, Ortigas, Mandaluyong City	SM Megamall, Ortigas, Mandaluyong City, Mandal…	121.057508	14.584450					

```
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]
```

[3]:
```
df.columns
```

[3]:
```
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')
```

[4]:
```
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
```

[5]:
```
df.describe()
```

[5]:

	Restaurant ID	Country Code	Longitude	Latitude	\
count	9.551000e+03	9551.000000	9551.000000	9551.000000	
mean	9.051128e+06	18.365616	64.126574	25.854381	
std	8.791521e+06	56.750546	41.467058	11.007935	
min	5.300000e+01	1.000000	-157.948486	-41.330428	
25%	3.019625e+05	1.000000	77.081343	28.478713	
50%	6.004089e+06	1.000000	77.191964	28.570469	
75%	1.835229e+07	1.000000	77.282006	28.642758	
max	1.850065e+07	216.000000	174.832089	55.976980	

	Average Cost for two	Price range	Aggregate rating	Votes
count	9551.000000	9551.000000	9551.000000	9551.000000
mean	1199.210763	1.804837	2.666370	156.909748
std	16121.183073	0.905609	1.516378	430.169145
min	0.000000	1.000000	0.000000	0.000000
25%	250.000000	1.000000	2.500000	5.000000
50%	400.000000	2.000000	3.200000	31.000000
75%	700.000000	2.000000	3.700000	131.000000
max	800000.000000	4.000000	4.900000	10934.000000

[6]:
```
df.shape
```

[6]:
```
(9551, 21)
```

[7]:
```
df.isnull().sum()
```

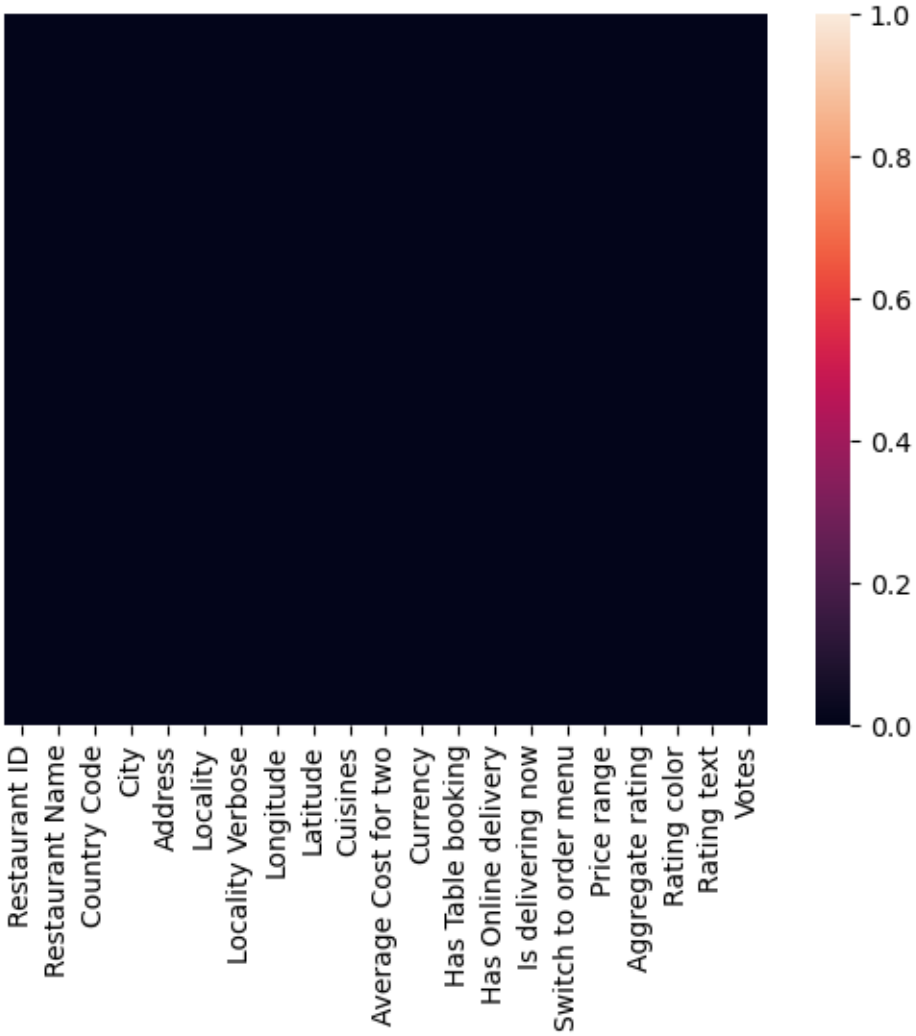
[7]:
```
Restaurant ID         0
Restaurant Name       0
Country Code          0
City                  0
Address               0
```

```
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
```

[8]:
```
[features for features in df.columns if df[features].isnull().sum()>0]
```

[8]:
```
['Cuisines']
```

[9]:
```
sns.heatmap(df.isnull(),yticklabels=False)
plt.show()
```



[10]:
```
df_country = pd.read_excel("Zomatodataset/Country-Code.xlsx")
df_country.head()
```

[10]:

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

[11]:
```
final_df=pd.merge(df,df_country,on='Country Code',how='left')
```

[12]:
```
final_df.head()
```

[12]:

	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]
```

[13]:
```
final_df.columns
```

[13]:
```
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')
```

[14]:
```
final_df.Country.value_counts()
```

[14]:
```
India             8652
United States      434
United Kingdom      80
Brazil              60
UAE                 60
South Africa        60
New Zealand         40
Turkey              34
Australia           24
Phillipines         22
Indonesia           21
Singapore           20
Qatar               20
Sri Lanka           20
Canada               4
Name: Country, dtype: int64
```

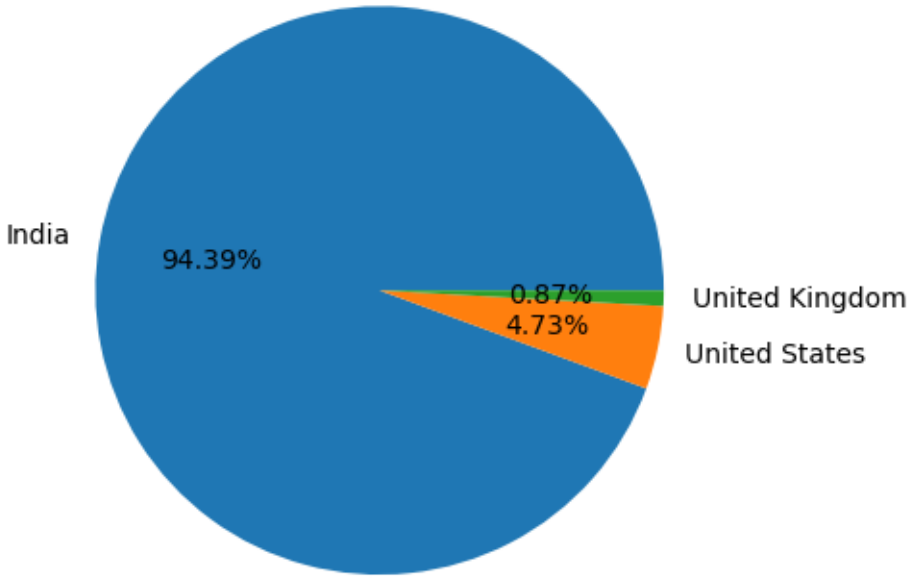
[15]:
```
country_names=final_df.Country.value_counts().index
country_names
```

[15]:
```
Index(['India', 'United States', 'United Kingdom', 'Brazil', 'UAE',
       'South Africa', 'New Zealand', 'Turkey', 'Australia', 'Phillipines',
       'Indonesia', 'Singapore', 'Qatar', 'Sri Lanka', 'Canada'],
      dtype='object')
```

[16]:
```
country_val=final_df.Country.value_counts().values
country_val
```

[16]:
```
array([8652,  434,   80,   60,   60,   60,   40,   34,   24,   22,   21,
         20,   20,   20,    4], dtype=int64)
```

[17]:
```
plt.pie(country_val[:3],labels=country_names[:3],autopct='%1.2f%%')  ## Top 3
↪countries that uses zomato
plt.show()
```



[18]:
```
final_df.columns
```

[18]:
```
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')
```

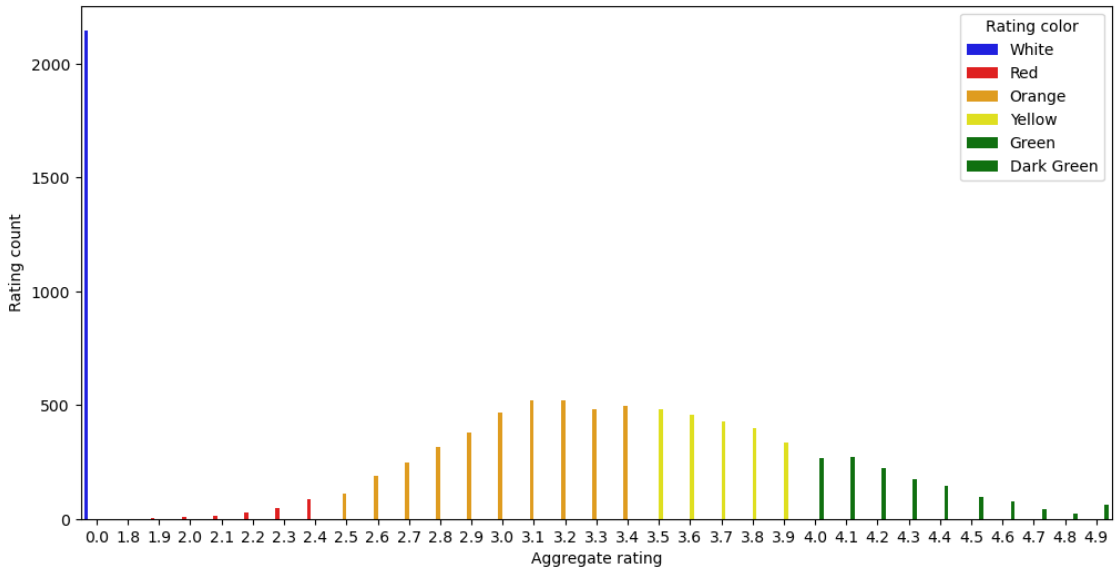
[19]:
```
ratings=final_df.groupby(['Aggregate rating', 'Rating color', 'Rating text']).
↪size().reset_index().rename(columns={0:'Rating count'})
ratings
```

[19]:

	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

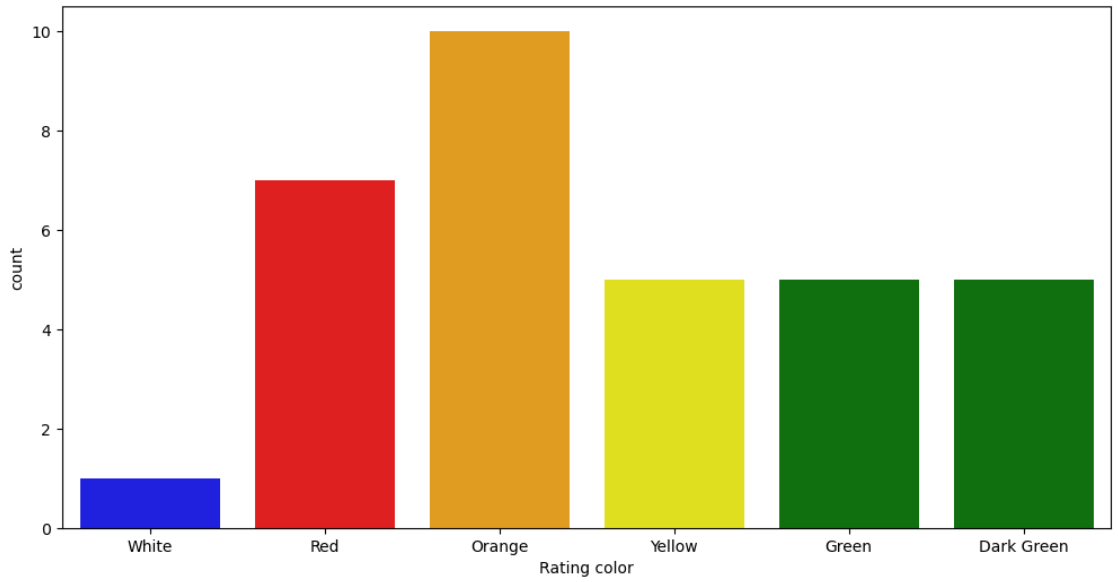
4	2.1	Red	Poor	15
5	2.2	Red	Poor	27
6	2.3	Red	Poor	47
7	2.4	Red	Poor	87
8	2.5	Orange	Average	110
9	2.6	Orange	Average	191
10	2.7	Orange	Average	250
11	2.8	Orange	Average	315
12	2.9	Orange	Average	381
13	3.0	Orange	Average	468
14	3.1	Orange	Average	519
15	3.2	Orange	Average	522
16	3.3	Orange	Average	483
17	3.4	Orange	Average	498
18	3.5	Yellow	Good	480
19	3.6	Yellow	Good	458
20	3.7	Yellow	Good	427
21	3.8	Yellow	Good	400
22	3.9	Yellow	Good	335
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	144
28	4.5	Dark Green	Excellent	95
29	4.6	Dark Green	Excellent	78
30	4.7	Dark Green	Excellent	42
31	4.8	Dark Green	Excellent	25
32	4.9	Dark Green	Excellent	61

[20]:
```
import matplotlib
matplotlib.rcParams['figure.figsize'] = (12, 6)
sns.barplot(x="Aggregate rating",y="Rating count",hue='Rating␣
 ↪color',data=ratings,palette=['blue','red','orange','yellow','green','green'])
plt.show()
```



[21]:
```
## Count plot
sns.countplot(x="Rating␣
↪color",data=ratings,palette=['blue','red','orange','yellow','green','green'])
```

[21]: <Axes: xlabel='Rating color', ylabel='count'>



[22]:
```
final_df[final_df['Rating color']=='White'].groupby(['Country']).size().
↪reset_index()
```

[22]:
```
           Country     0
0           Brazil     5
1            India  2139
2   United Kingdom     1
3    United States     3
```

[23]:
```
final_df.groupby(['Currency','Country']).size().reset_index()
```

[23]:
```
                   Currency         Country     0
0          Botswana Pula(P)      Phillipines    22
1        Brazilian Real(R$)           Brazil    60
2                 Dollar($)        Australia    24
3                 Dollar($)           Canada     4
4                 Dollar($)        Singapore    20
5                 Dollar($)    United States   434
6        Emirati Diram(AED)              UAE    60
7        Indian Rupees(Rs.)            India  8652
8    Indonesian Rupiah(IDR)        Indonesia    21
9             NewZealand($)      New Zealand    40
10               Pounds( £)   United Kingdom    80
11          Qatari Rial(QR)            Qatar    20
12                Rand(R)     South Africa    60
13   Sri Lankan Rupee(LKR)        Sri Lanka    20
14         Turkish Lira(TL)           Turkey    34
```

[24]:
```
final_df.columns
```

[24]:
```
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')
```

[25]:
```
final_df[final_df['Has Online delivery']=='Yes'].groupby(['Country']).size().
↪reset_index()
```

[25]:
```
  Country     0
0   India  2423
1     UAE    28
```

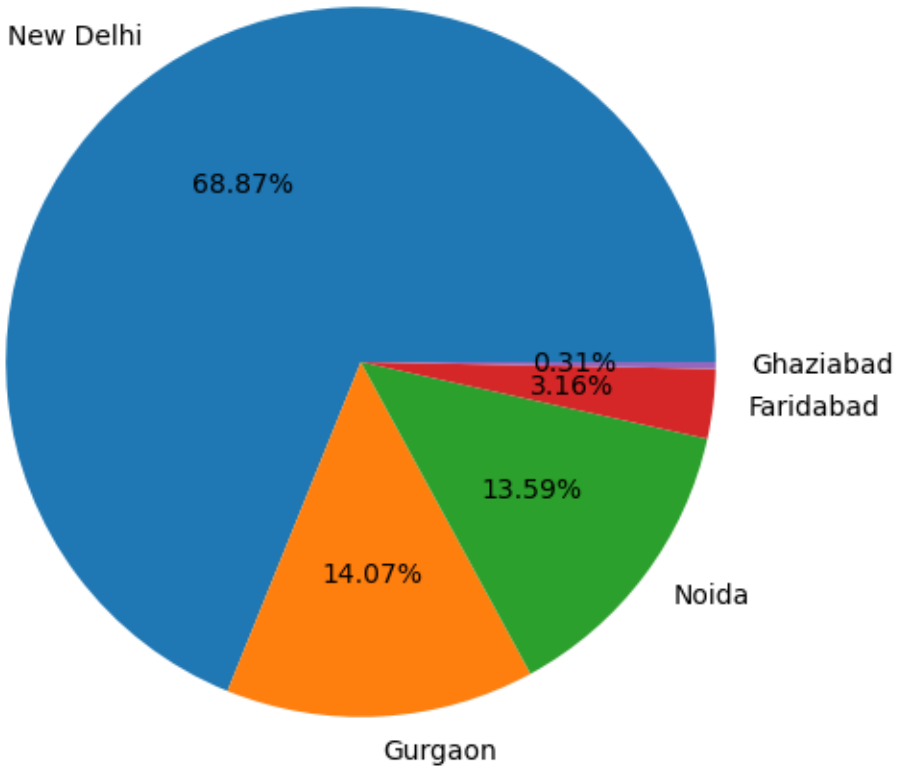
[26]:
```
final_df[final_df['Has Online delivery']=='No'].groupby(['Country']).size().
↪reset_index()
```

[26]:
```
           Country     0
0        Australia    24
1           Brazil    60
```

```
2           Canada     4
3            India  6229
4        Indonesia    21
5      New Zealand    40
6      Phillipines    22
7            Qatar    20
8        Singapore    20
9     South Africa    60
10       Sri Lanka    20
11          Turkey    34
12             UAE    32
13  United Kingdom    80
14   United States   434
```

[27]:
```
city_count=final_df.City.value_counts().index
city_val=final_df.City.value_counts().values
```

[28]:
```
plt.pie(city_val[:5],labels=city_count[:5],autopct='%1.2f%%')
plt.show()
```



[35]:
```
final_df.Cuisines.value_counts()[:10]
```

[35]:
```
North Indian                          936
North Indian, Chinese                 511
Chinese                               354
Fast Food                             354
North Indian, Mughlai                 334
Cafe                                  299
Bakery                                218
North Indian, Mughlai, Chinese        197
Bakery, Desserts                      170
Street Food                           149
Name: Cuisines, dtype: int64
```

[47]:
```
sd=final_df.City.value_counts().reset_index()
display(sd)
```

```
                index  City
0           New Delhi  5473
1             Gurgaon  1118
2               Noida  1080
3           Faridabad   251
4           Ghaziabad    25
..                ...   ...
136         Panchkula     1
137         Mc Millan     1
138          Mayfield     1
139           Macedon     1
140  Vineland Station     1

[141 rows x 2 columns]
```

[ ]: