Task: Descriptive Analysis

Importing Libraries

In [1]:
```
import pandas as pd
import statistics as ST
import matplotlib.pyplot as plt
import seaborn as sns
```

In [3]:
```
df=pd.read_csv('N:\Dataset .csv')
```

In [4]:
```
df.columns
```

Out[4]:
```
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')
```

Data characteristics

In [5]:
```
df.head()
```

Out[5]:

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitu
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...	121.0275
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...	121.0141
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...	121.0568
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...	121.0564
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...	121.0575

5 rows × 21 columns

In [6]:
```
df.tail()
```

Out[6]:

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Loc Ver
9546	5915730	Naml\ Gurme	208	��stanbul	Kemanke�� Karamustafa Pa��a Mahallesi, R\ht\m ...	Karak�_y	Karak ��st
9547	5908749	Ceviz A��ac\	208	��stanbul	Ko��uyolu Mahallesi, Muhittin ��st�_nda�� Cadd...	Ko��uyolu	Ko��u ��st
9548	5915807	Huqqa	208	��stanbul	Kuru�_e��me Mahallesi, Muallim Naci Caddesi, N...	Kuru�_e��me	Kuru�_e�(��st
9549	5916112	A���k Kahve	208	��stanbul	Kuru�_e��me Mahallesi, Muallim Naci Caddesi, N...	Kuru�_e��me	Kuru�_e�(��st
9550	5927402	Walter's Coffee Roastery	208	��stanbul	Cafea��a Mahallesi, Bademalt\ Sokak, No 21/B, ...	Moda	M ��st

5 rows × 21 columns

In [7]:
```
df.describe()
```

Out[7]:

	Restaurant ID	Country Code	Longitude	Latitude	Average Cost for two	Price range	Aggr
count	9.551000e+03	9551.000000	9551.000000	9551.000000	9551.000000	9551.000000	9551.0
mean	9.051128e+06	18.365616	64.126574	25.854381	1199.210763	1.804837	2.6
std	8.791521e+06	56.750546	41.467058	11.007935	16121.183073	0.905609	1.5
min	5.300000e+01	1.000000	-157.948486	-41.330428	0.000000	1.000000	0.0
25%	3.019625e+05	1.000000	77.081343	28.478713	250.000000	1.000000	2.5
50%	6.004089e+06	1.000000	77.191964	28.570469	400.000000	2.000000	3.2
75%	1.835229e+07	1.000000	77.282006	28.642758	700.000000	2.000000	3.7
max	1.850065e+07	216.000000	174.832089	55.976980	800000.000000	4.000000	4.9

Calculate basic statistical measures (mean,median, standard deviation, etc.) for

numerical columns

In [8]:
```
print(df['Country Code'].mean())
print(df['Country Code'].median())
print(ST.stdev(df['Country Code']))
```

```
18.365616165846507
1.0
56.75054560094657
```

In [9]:
```
print(df['Longitude'].mean())
print(df['Longitude'].median())
print(ST.stdev(df['Longitude']))
```

```
64.12657446168706
77.1919642
41.46705784761728
```

In [10]:
```
print(df['Latitude'].mean())
print(df['Latitude'].median())
print(ST.stdev(df['Latitude']))
```

```
25.854380700074756
28.57046888
11.007935124784668
```

In [11]:
```
print(df['Average Cost for two'].mean())
print(df['Average Cost for two'].median())
print(ST.stdev(df['Average Cost for two']))
```

```
1199.2107632708617
400.0
16121.183073499646
```

In [12]:
```
print(df['Price range'].mean())
print(df['Price range'].median())
print(ST.stdev(df['Price range']))
```

```
1.804837189823055
2.0
0.905608847397614
```

In [13]:
```
print(df['Aggregate rating'].mean())
print(df['Aggregate rating'].median())
print(ST.stdev(df['Aggregate rating']))
```

```
2.66637001361114
3.2
1.5163775396521326
```

In [14]:
```
print(df['Votes'].mean())
print(df['Votes'].median())
print(ST.stdev(df['Votes']))
```

```
156.909747670401
31.0
430.1691453762912
```

Explore the distribution of categorical variables

s

In [15]:
```
top_cuis=df['Cuisines'].value_counts()
print(top_cuis.head())
```

```
Cuisines
North Indian              936
North Indian, Chinese     511
Chinese                   354
Fast Food                 354
North Indian, Mughlai     334
Name: count, dtype: int64
```

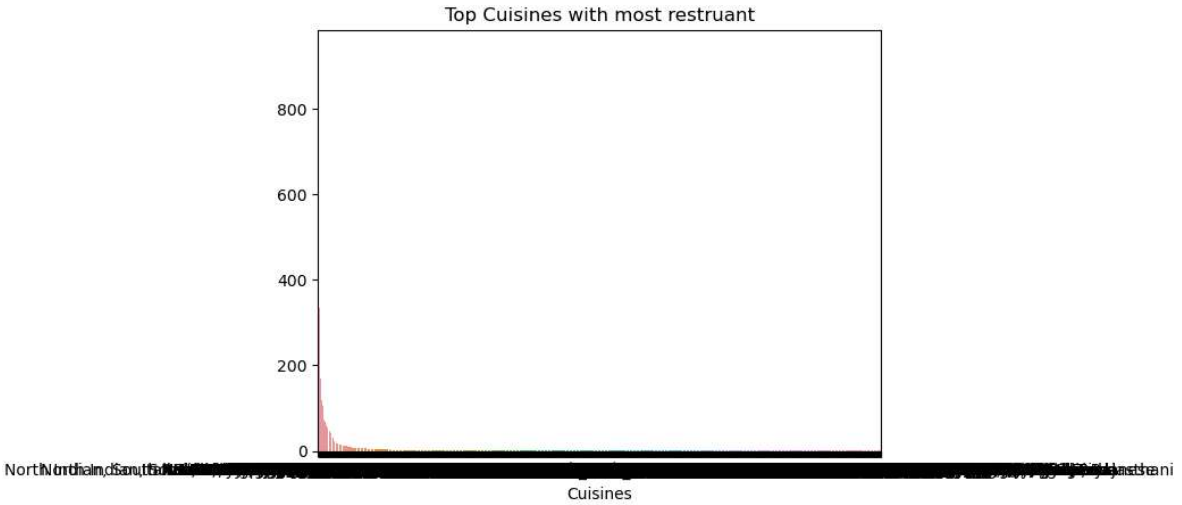
In [16]:
```
top_city=df['City'].value_counts()
print(top_city.head())
```

```
City
New Delhi    5473
Gurgaon      1118
Noida        1080
Faridabad     251
Ghaziabad      25
Name: count, dtype: int64
```

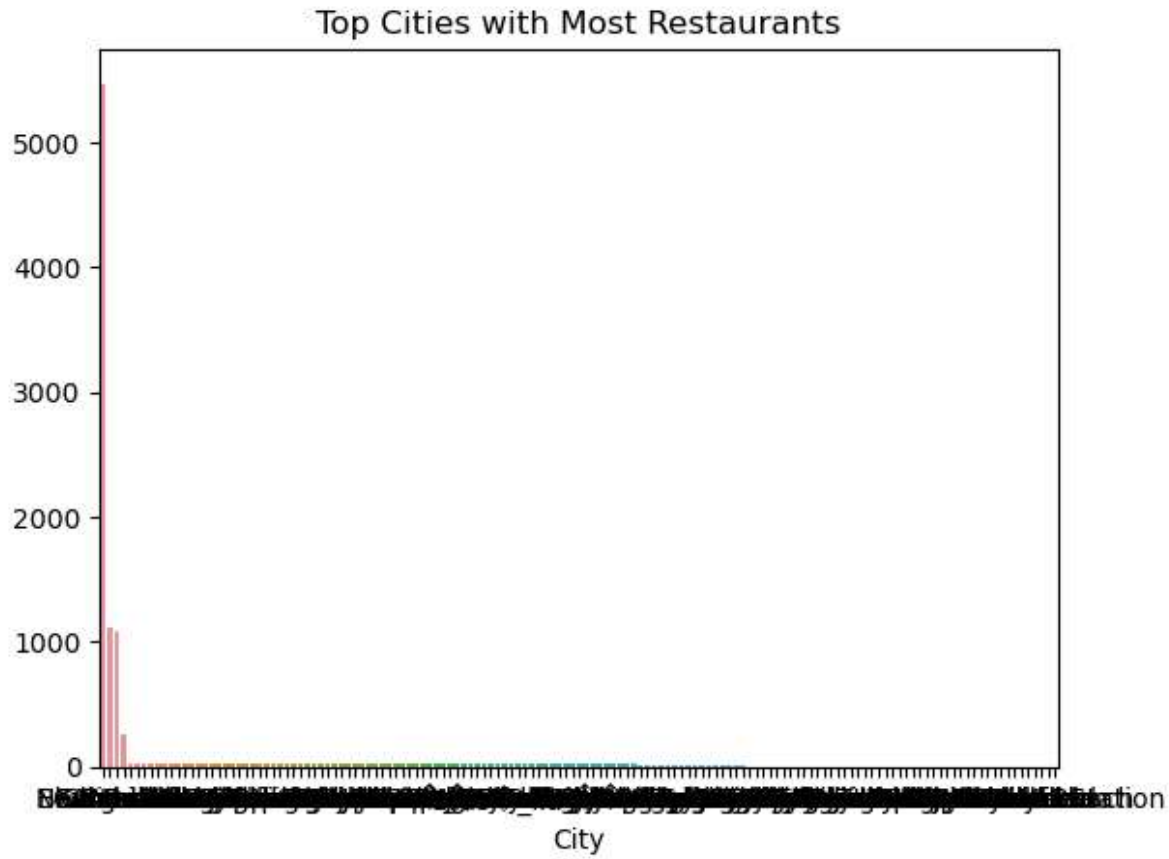
In [17]:
```
top_restu=df['Restaurant Name'].value_counts()
print(top_restu.head())
```

```
Restaurant Name
Cafe Coffee Day     83
Domino's Pizza      79
Subway              63
Green Chick Chop    51
McDonald's          48
Name: count, dtype: int64
```

In [21]:
```
sns.barplot(x=top_cuis.index ,y=top_cuis.values)
plt.title('Top Cuisines with most restruant')
plt.show()
```



In [19]:
```
sns.barplot(x=top_city.index, y=top_city.values)
plt.title('Top Cities with Most Restaurants')
plt.show()
```



In [ ]: