

```
In [1]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
```

```
In [2]: data = pd.read_excel('data.xlsx')
data.head()
```

```
Out[2]:
```

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitude	Latitude
0	7402935	Skye	94	Jakarta	Menara BCA, Lantai 56, Jl. MH. Thamrin, Thamri...	Grand Indonesia Mall, Thamrin	Grand Indonesia Mall, Thamrin, Jakarta	106.821999	-6.19677
1	7410290	Satoo - Hotel Shangri-La	94	Jakarta	Hotel Shangri-La, Jl. Jend. Sudirman	Hotel Shangri-La, Sudirman	Hotel Shangri-La, Sudirman, Jakarta	106.818961	-6.20329
2	7420899	Sushi Masa	94	Jakarta	Jl. Tuna Raya No. 5, Penjaringan	Penjaringan	Penjaringan, Jakarta	106.800144	-6.10129
3	7421967	3 Wise Monkeys	94	Jakarta	Jl. Suryo No. 26, Senopati, Jakarta	Senopati	Senopati, Jakarta	106.813400	-6.23524
4	7422489	Avec Moi Restaurant and Bar	94	Jakarta	Gedung PIC, Jl. Teluk Betung 43, Thamrin, Jakarta	Thamrin	Thamrin, Jakarta	106.821023	-6.19627

```
In [3]: cc = pd.read_excel('Country-Code.xlsx')
```

```
In [4]: df_rest = pd.merge(data, cc, on='Country Code', how='left')
```

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

Out[5]:

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitude	Latitude
0	7402935	Skye	94	Jakarta	Menara BCA, Lantai 56, Jl. MH. Thamrin, Thamri...	Grand Indonesia Mall, Thamrin	Grand Indonesia Mall, Thamrin, Jakarta	106.821999	-6.19677
1	7410290	Satoo - Hotel Shangri-La	94	Jakarta	Hotel Shangri-La, Jl. Jend. Sudirman	Hotel Shangri-La, Sudirman	Hotel Shangri-La, Sudirman, Jakarta	106.818961	-6.20329
2	7420899	Sushi Masa	94	Jakarta	Jl. Tuna Raya No. 5, Penjaringan	Penjaringan	Penjaringan, Jakarta	106.800144	-6.10129
3	7421967	3 Wise Monkeys	94	Jakarta	Jl. Suryo No. 26, Senopati, Jakarta	Senopati	Senopati, Jakarta	106.813400	-6.23524
4	7422489	Avec Moi Restaurant and Bar	94	Jakarta	Gedung PIC, Jl. Teluk Betung 43, Thamrin, Jakarta	Thamrin	Thamrin, Jakarta	106.821023	-6.19627

```
In [6]: df_rest.columns = df_rest.columns.str.replace(' ', '_')
df_rest.columns
```

```
Out[6]: 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', 'Price_range', 'Aggregate_rating',
'Rating_color', 'Rating_text', 'Votes', 'Country'],
dtype='object')
```

```
In [7]: df_rest.info()
```

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 9551 entries, 0 to 9550
Data columns (total 20 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   Restaurant_ID         9551 non-null   int64
 1   Restaurant_Name       9550 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   Price_range         9551 non-null   int64
15   Aggregate_rating     9551 non-null   float64
16   Rating_color        9551 non-null   object
17   Rating_text         9551 non-null   object
18   Votes              9551 non-null   int64
19   Country             9551 non-null   object
dtypes: float64(3), int64(5), object(12)
memory usage: 1.5+ MB

```

```
In [8]: df_rest.isnull().sum()
```

```

Out[8]: Restaurant_ID      0
Restaurant_Name      1
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
Price_range        0
Aggregate_rating    0
Rating_color        0
Rating_text         0
Votes              0
Country            0
dtype: int64

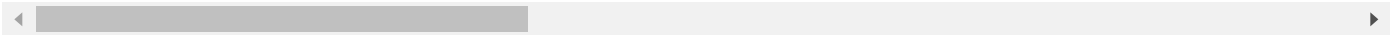
```

```
In [9]: df_rest[df_rest['Restaurant_Name'].isnull()]
```

Out[9]:

	Restaurant_ID	Restaurant_Name	Country_Code		City	Address	Locality	Locality_Ve
--	---------------	-----------------	--------------	--	------	---------	----------	-------------

	1646	113702	NaN	1	Ahmedabad	Opposite Sindhu Bhawan, Bodakdev, Ahmedabad	Bodakdev	Boda Ahme
--	------	--------	-----	---	-----------	---	----------	--------------



```
In [10]: ##Since the restaurant name is missing, we dropped the record and reset the index.
df_rest.dropna(axis=0,subset=['Restaurant_Name'],inplace=True)
df_rest.reset_index(drop=True,inplace=True)
df_rest[df_rest['Cuisines'].isnull()]
```

Out[10]:

	Restaurant_ID	Restaurant_Name	Country_Code	City	Address	Locality	Locality_Verl
9082	17374552	Corkscrew Cafe	216	Gainesville	51 W Main St, Dahlonega, GA 30533	Dahlonega	Dahlonega, Ga
9085	17501439	Dovetail	216	Macon	543 Cherry St, Macon, GA 31201	Macon	Macon, Ma
9093	17059060	Hillstone	216	Orlando	215 South Orlando Avenue, Winter Park, FL 32789	Winter Park	Winter Park, Orla
9405	17284158	Jimmie's Hot Dogs	216	Albany	204 S Jackson St, Albany, GA 31701	Albany	Albany, All
9493	17142698	Leonard's Bakery	216	Rest of Hawaii	933 Kapahulu Ave, Honolulu, HI 96816	Kaimuki	Kaimuki, Re H
9503	17616465	Tybee Island Social Club	216	Savannah	1311 Butler Ave, Tybee Island, GA 31328	Tybee Island	Tybee Isl Savar
9532	17284105	Cookie Shoppe	216	Albany	115 N Jackson St, Albany, GA 31701	Albany	Albany, All
9534	17284211	Pearly's Famous Country Cookng	216	Albany	814 N Slappey Blvd, Albany, GA 31701	Albany	Albany, All
9538	17606621	HI Lite Bar & Lounge	216	Miller	109 N Broadway Ave, Miller, SD 57362	Miller	Miller, N

In [11]: `#Since there were only 9 records without cuisines, we have replace the null values with
df_rest['Cuisines'].fillna('Others',inplace=True)`

In [12]: `df_rest.isnull().sum()
df_rest.info()`

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9550 entries, 0 to 9549
Data columns (total 20 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Restaurant_ID          9550 non-null   int64
1   Restaurant_Name        9550 non-null   object
2   Country_Code           9550 non-null   int64
3   City                   9550 non-null   object
4   Address                9550 non-null   object
5   Locality               9550 non-null   object
6   Locality_Verbose       9550 non-null   object
7   Longitude              9550 non-null   float64
8   Latitude               9550 non-null   float64
9   Cuisines               9550 non-null   object
10  Average_Cost_for_two   9550 non-null   int64
11  Currency               9550 non-null   object
12  Has_Table_booking      9550 non-null   object
13  Has_Online_delivery    9550 non-null   object
14  Price_range            9550 non-null   int64
15  Aggregate_rating       9550 non-null   float64
16  Rating_color           9550 non-null   object
17  Rating_text            9550 non-null   object
18  Votes                  9550 non-null   int64
19  Country                9550 non-null   object
dtypes: float64(3), int64(5), object(12)
memory usage: 1.5+ MB

```

EDA-1

Explore the geographical distribution of the restaurants.

Finding out the cities with maximum / minimum number of restaurants

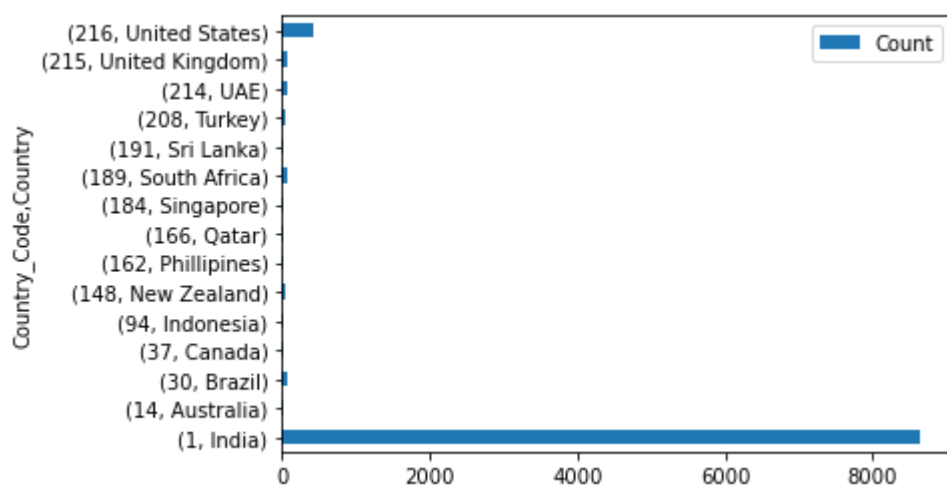
```

In [13]: cntry_dist = df_rest.groupby(['Country_Code', 'Country']).agg( Count = ('Restaurant_ID'
cntry_dist.sort_values(by='Count', ascending=False)
#We observe that India has then highest number of restaurants with 8651 restaurants ar

```

Out[13]:

		Count
Country_Code	Country	
1	India	8651
216	United States	434
215	United Kingdom	80
30	Brazil	60
189	South Africa	60
214	UAE	60
148	New Zealand	40
208	Turkey	34
14	Australia	24
162	Phillipines	22
94	Indonesia	21
166	Qatar	20
184	Singapore	20
191	Sri Lanka	20
37	Canada	4

In [14]: `cntry_dist.plot(kind='barh')`Out[14]: `<AxesSubplot:ylabel='Country_Code,Country'>`

```
In [15]: city_dist = df_rest.groupby(['Country', 'City']).agg(Count=('Restaurant_ID', 'count'))
city_dist.describe()
#city with max restaurant has count = 5473
#city with min restaurant has count = 1
```

Out[15]:

Count	
count	141.000000
mean	67.730496
std	476.723952
min	1.000000
25%	1.000000
50%	20.000000
75%	20.000000
max	5473.000000

```
In [16]: city_dist.sort_values(by='Count',ascending=False)
# we see that new Delhi has the maximum restaurant with 5473
# we observe that multiple cities have only one restaurant.
```

Out[16]:

		Count
Country	City	
India	New Delhi	5473
	Gurgaon	1118
	Noida	1080
	Faridabad	251
	Ghaziabad	25

	Panchkula	1
Australia	Balingup	1
Indonesia	Bandung	1
Phillipines	Quezon City	1
United States	Winchester Bay	1

141 rows × 1 columns

```
In [17]: min_cnt_rest = city_dist[city_dist['Count']==1]
min_cnt_rest.info()
min_cnt_rest
#There are 46 cities in 7 different countries with 1 restaurants
```



```
<class 'pandas.core.frame.DataFrame'>
MultiIndex: 46 entries, ('Australia', 'Armidale') to ('United States', 'Winchester Ba
y')
Data columns (total 1 columns):
#   Column  Non-Null Count  Dtype
---  ---
0   Count   46 non-null      int64
dtypes: int64(1)
memory usage: 1.8+ KB
```

Out[17]:

		Count
Country	City	
Australia	Armidale	1
	Balingup	1
	Beechworth	1
	Dicky Beach	1
	East Ballina	1
	Flaxton	1
	Forrest	1
	Huskisson	1
	Inverloch	1
	Lakes Entrance	1
	Lorn	1
	Macedon	1
	Mayfield	1
	Middleton Beach	1
	Montville	1
	Palm Cove	1
	Paynesville	1
	Penola	1
	Phillip Island	1
	Tanunda	1
Canada	Trentham East	1
	Victor Harbor	1
	Chatham-Kent	1
	Consort	1
India	Vineland Station	1
	Yorkton	1
	Mohali	1
Indonesia	Panchkula	1
	Bandung	1
Phillipines	Quezon City	1
	Tagaytay City	1
South Africa	Randburg	1

		Count
Country	City	
United States	Clatskanie	1
	Cochrane	1
	Fernley	1
	Lakeview	1
	Lincoln	1
	Mc Millan	1
	Miller	1
	Monroe	1
	Ojo Caliente	1
	Potrero	1
	Princeton	1
	Vernonia	1
	Weirton	1
	Winchester Bay	1

EDA-II

Explore how ratings are distributed overall

Evaluate the Highest Rated and Lowest Rated Restaurant of the City in all the countries.

```
In [18]: max_rate = df_rest.sort_values(by='Aggregate_rating',ascending=False).groupby(['Country', 'City'])
#highest rating restaurants

min_rate = df_rest.sort_values(by='Aggregate_rating',ascending=False).groupby(['Country', 'City'])
#lowest rating restaurants

df_max=max_rate[['Country', 'City', 'Restaurant_Name', 'Aggregate_rating']] #new dataframe
df_min=min_rate[['Country', 'City', 'Restaurant_Name', 'Aggregate_rating']] #new dataframe

rating_rest=df_max.merge(df_min,left_on='City',right_on='City',how='inner') #merge into one dataframe
```

```
In [19]: rating_rest
```

Out[19]:

	Country_x	City	Restaurant_Name_x	Aggregate_rating_x	Country_y	Restaurant_Name_y
0	Australia	Armidale	Whitebull Hotel	3.5	Australia	Whitebull Hotel
1	Australia	Balingup	Taste of Balingup	3.2	Australia	Taste of Balingup
2	Australia	Beechworth	Bridge Road Brewers	4.6	Australia	Bridge Road Brewers
3	Australia	Dicky Beach	The Giggling Goat	3.6	Australia	The Giggling Goat
4	Australia	East Ballina	The Belle General	4.1	Australia	The Belle General
...
136	United States	Valdosta	Smok'n Pig B-B-Q	4.1	United States	El Tereo Mexican Restaurant
137	United States	Vernonia	Blue House Cafe	4.3	United States	Blue House Cafe
138	United States	Waterloo	Four Queens Dairy Cream	3.9	United States	Masala Grill & Coffee House
139	United States	Weirton	Theo Yianni's Authentic Greek Restaurant	3.9	United States	Theo Yianni's Authentic Greek Restaurant
140	United States	Winchester Bay	Fishpatrick's Crabby Cafe	3.2	United States	Fishpatrick's Crabby Cafe

141 rows × 7 columns



```
In [20]: rating_rest.drop(columns='Country_y',axis=1,inplace=True)
rating_rest.columns = ['Country','City','Highest Rated Restaurant','Rating Max','Lowest
rating_rest
```

Out[20]:

	Country	City	Highest Rated Restaurant	Rating Max	Lowest Rated Restaurant	Rating Min
0	Australia	Armidale	Whitebull Hotel	3.5	Whitebull Hotel	3.5
1	Australia	Balingup	Taste of Balingup	3.2	Taste of Balingup	3.2
2	Australia	Beechworth	Bridge Road Brewers	4.6	Bridge Road Brewers	4.6
3	Australia	Dicky Beach	The Giggling Goat	3.6	The Giggling Goat	3.6
4	Australia	East Ballina	The Belle General	4.1	The Belle General	4.1
...
136	United States	Valdosta	Smok'n Pig B-B-Q	4.1	El Tereo Mexican Restaurant	3.1
137	United States	Vernonia	Blue House Cafe	4.3	Blue House Cafe	4.3
138	United States	Waterloo	Four Queens Dairy Cream	3.9	Masala Grill & Coffee House	3.2
139	United States	Weirton	Theo Yianni's Authentic Greek Restaurant	3.9	Theo Yianni's Authentic Greek Restaurant	3.9
140	United States	Winchester Bay	Fishpatrick's Crabby Cafe	3.2	Fishpatrick's Crabby Cafe	3.2

141 rows × 6 columns

```
In [21]: df_rest1 = df_rest.copy()
df_rest1.columns
```

```
Out[21]: 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', 'Price_range', 'Aggregate_rating',
'Rating_color', 'Rating_text', 'Votes', 'Country'],
dtype='object')
```

```
In [22]: dummy = ['Has_Table_booking', 'Has_Online_delivery']
df_rest1 = pd.get_dummies(df_rest1, columns=dummy, drop_first=True)
df_rest1.head()
# 0 indicates 'NO'
# 1 indicates 'YES'
```

Out[22]:

	Restaurant_ID	Restaurant_Name	Country_Code	City	Address	Locality	Locality_Verbose
0	7402935	Skye	94	Jakarta	Menara BCA, Lantai 56, Jl. MH. Thamrin, Thamri...	Grand Indonesia Mall, Thamrin	Grand Indonesia Mall, Thamrin, Jakarta
1	7410290	Satoo - Hotel Shangri-La	94	Jakarta	Hotel Shangri-La, Jl. Jend. Sudirman	Hotel Shangri-La, Sudirman	Hotel Shangri-La, Sudirman, Jakarta
2	7420899	Sushi Masa	94	Jakarta	Jl. Tuna Raya No. 5, Penjaringan	Penjaringan	Penjaringan, Jakarta
3	7421967	3 Wise Monkeys	94	Jakarta	Jl. Suryo No. 26, Senopati, Jakarta	Senopati	Senopati, Jakarta
4	7422489	Avec Moi Restaurant and Bar	94	Jakarta	Gedung PIC, Jl. Teluk Betung 43, Thamrin, Jakarta	Thamrin	Thamrin, Jakarta

EDA - III

Ratio between restaurants that allow table booking vs that do not allow table booking.

Percentage of restaurants providing online delivery.

Difference in no. of votes for the restaurants that deliver and the restaurant that don't.

```
In [23]: #Ration between restaurants allowing table booking and those which dont
table_booking = df_rest1[df_rest1['Has_Table_booking_Yes']==1]['Restaurant_ID'].count()
table_nbooking =df_rest1[df_rest1['Has_Table_booking_Yes']==0]['Restaurant_ID'].count()
print('Ratio between restaurants that allow table booking vs. those that do not allow
      round((table_booking/table_nbooking),2))
```

Ratio between restaurants that allow table booking vs. those that do not allow table booking: 0.14

```
In [24]: print(table_booking,table_nbooking)
```

1158 8392

```
In [25]: #Percentage of restaurant that has online delivery
rest_od = df_rest1[df_rest1['Has_Online_delivery_Yes'] == 1]['Restaurant_ID'].count()
rest_nod = df_rest1[df_rest1['Has_Online_delivery_Yes'] == 0]['Restaurant_ID'].count()
print('Percentage of restaurants providing online delivery : {} %'.format((round(rest_
```

Percentage of restaurants providing online delivery : 25.7 %

```
In [26]: rest_deliver = df_rest1[df_rest1['Has_Table_booking_Yes'] == 1]['Votes'].sum()
rest_ndeliver = df_rest1[df_rest1['Has_Table_booking_Yes'] == 0]['Votes'].sum()
```

```
print('Difference in number of votes for restaurants that deliver and dont deliver: ',
```

Difference in number of votes for restaurants that deliver and dont deliver: 680082

EDA -IV

What are the top 10 cuisines served across cities?

What is the maximum and minimum no. of cuisines that a restaurant serves?

```
In [27]: df_rest.columns
cuisines = df_rest['Cuisines'].apply(lambda x: pd.Series(x.split(',')))
```

```
In [28]: cuisines.columns = ['Cuisine_1','Cuisine_2','Cuisine_3','Cuisine_4','Cuisine_5','Cuisi
cuisines.tail()
```

```
Out[28]:
```

	Cuisine_1	Cuisine_2	Cuisine_3	Cuisine_4	Cuisine_5	Cuisine_6	Cuisine_7	Cuisine_8
9545	Chinese	North Indian	Fast Food	NaN	NaN	NaN	NaN	NaN
9546	Indian	Chinese	Continental	NaN	NaN	NaN	NaN	NaN
9547	Cafe	Continental	Desserts	Ice Cream	Italian	Beverages	NaN	NaN
9548	Street Food	NaN	NaN	NaN	NaN	NaN	NaN	NaN
9549	Chinese	North Indian	NaN	NaN	NaN	NaN	NaN	NaN

```
In [29]: df_cuisines = pd.concat([df_rest,cuisines],axis=1)
df_cuisines.head()
```

Out[29]:

	Restaurant_ID	Restaurant_Name	Country_Code	City	Address	Locality	Locality_Verbose
0	7402935	Skye	94	Jakarta	Menara BCA, Lantai 56, Jl. MH. Thamrin, Thamri...	Grand Indonesia Mall, Thamrin	Grand Indonesia Mall, Thamrin, Jakarta
1	7410290	Satoo - Hotel Shangri-La	94	Jakarta	Hotel Shangri-La, Jl. Jend. Sudirman	Hotel Shangri-La, Sudirman	Hotel Shangri-La, Sudirman, Jakarta
2	7420899	Sushi Masa	94	Jakarta	Jl. Tuna Raya No. 5, Penjaringan	Penjaringan	Penjaringan, Jakarta
3	7421967	3 Wise Monkeys	94	Jakarta	Jl. Suryo No. 26, Senopati, Jakarta	Senopati	Senopati, Jakarta
4	7422489	Avec Moi Restaurant and Bar	94	Jakarta	Gedung PIC, Jl. Teluk Betung 43, Thamrin, Jakarta	Thamrin	Thamrin, Jakarta

5 rows × 28 columns

```
In [30]: cuisine_loc = pd.DataFrame(df_cuisines[['Country','City','Locality_Verbose','Cuisine_1',
                                                'Cuisine_4','Cuisine_5','Cuisine_6','Cuisine_7']])
```

```
In [31]: cuisine_loc_stack=pd.DataFrame(cuisine_loc.stack()) #stacking the columns
cuisine_loc.head()
```

```
Out[31]:
```

	Country	City	Locality_Verbose	Cuisine_1	Cuisine_2	Cuisine_3	Cuisine_4	Cuisine_5	Cuisine_6
0	Indonesia	Jakarta	Grand Indonesia Mall, Thamrin, Jakarta	Italian	Continental	NaN	NaN	NaN	NaN
1	Indonesia	Jakarta	Hotel Shangri-La, Sudirman, Jakarta	Asian	Indonesian	Western	NaN	NaN	NaN
2	Indonesia	Jakarta	Penjaringan, Jakarta	Sushi	Japanese	NaN	NaN	NaN	NaN
3	Indonesia	Jakarta	Senopati, Jakarta	Japanese	NaN	NaN	NaN	NaN	NaN
4	Indonesia	Jakarta	Thamrin, Jakarta	French	Western	NaN	NaN	NaN	NaN

```
In [33]: df_rest.to_csv("Identifying_recommending_restaurant.csv", index=False)
```

In []: