

# Level 2: Restaurant Data Analytics | Data Analysis

- Task 1: Restaurant Ratings
- Task 2: Cuisine Combination
- Task 3: Geographic Analysis
- Task 4: Restaurant Chains

## Step 1: Import necessary Python libraries.

```
In [3]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from tabulate import tabulate
import itertools
import plotly.express as px
from sklearn.cluster import KMeans
```

## Step 2: Load the dataset into a DataFrame.

```
In [5]: resto_df = pd.read_csv(r"Dataset .csv")
resto_df
```

Out[5]:

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose
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..
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..
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..
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..
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..
...	...	...	...	...	...	...	..
9546	5915730	Naml Gurme	208	İstanbul	Kemankeş Karamustafa Paşası Mahallesi, Rıhtım ...	Karaköy	Karaköy, İstanbul
9547	5908749	Ceviz Aca	208	İstanbul	Koşuyolu Mahallesi, Muhittin Köstendağ Cadd...	Koşuyolu	Koşuyolu, İstanbul
9548	5915807	Huqqa	208	İstanbul	Kuruçeşme Mahallesi, Muallim Naci Caddesi, N...	Kuruçeşme	Kuruçeşme, İstanbul
9549	5916112	Ak Kahve	208	İstanbul	Kuruçeşme Mahallesi, Muallim Naci Caddesi, N...	Kuruçeşme	Kuruçeşme, İstanbul
9550	5927402	Walter's Coffee Roastery	208	İstanbul	Cafea Mahallesi, Bademaltı Sokak, No 21/B, ...	Moda	Moda, İstanbul

9551 rows × 21 columns

## Step 3: Basic Inspection on given dataset

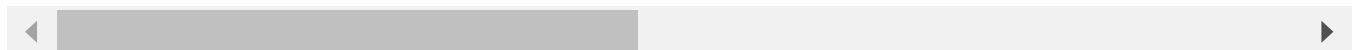
- Top 5 rows - using head

```
In [8]: resto_df.head()
```

```
Out[8]:
```

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitude
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
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
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
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
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

5 rows × 21 columns



- bottom 5 rows using tail

```
In [10]: resto_df.tail()
```

Out[10]:

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose
9546	5915730	Namlı Gurme	208	İstanbul	Kemankeş Karamustafa Paşa Mahallesi, Rıhtım ...	Karaköy	Karaköy, İstanbul
9547	5908749	Ceviz Aca	208	İstanbul	Koşuyolu Mahallesi, Muhittin İsmet Paşa Caddesi	Koşuyolu	Koşuyolu, İstanbul
9548	5915807	Huqqa	208	İstanbul	Kuruçeşme Mahallesi, Muallim Naci Caddesi, N...	Kuruçeşme	Kuruçeşme, İstanbul
9549	5916112	Ak Kahve	208	İstanbul	Kuruçeşme Mahallesi, Muallim Naci Caddesi, N...	Kuruçeşme	Kuruçeşme, İstanbul
9550	5927402	Walter's Coffee Roastery	208	İstanbul	Cafea Mahallesi, Bademaltı Sokak, No 21/B, ...	Moda	Moda, İstanbul

5 rows × 21 columns

• Inspecting Column Names and Data Types

```
In [12]: resto_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
```

- **Checking for Missing Values**

```
In [14]: resto_df.isnull().sum()
```

```
Out[14]: 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
```

```
In [24]: cuisines = resto_df['Cuisines'].dropna().str.split(", ").explode()
```

- **Basic Statistical Summary**

```
In [27]: resto_df.describe()
```

Out[27]:

	Restaurant ID	Country Code	Longitude	Latitude	Average Cost for two	Price range	Aggregate rating
count	9.551000e+03	9551.000000	9551.000000	9551.000000	9551.000000	9551.000000	9551.000000
mean	9.051128e+06	18.365616	64.126574	25.854381	1199.210763	1.804837	2.666370
std	8.791521e+06	56.750546	41.467058	11.007935	16121.183073	0.905609	1.516378
min	5.300000e+01	1.000000	-157.948486	-41.330428	0.000000	1.000000	0.000000
25%	3.019625e+05	1.000000	77.081343	28.478713	250.000000	1.000000	2.500000
50%	6.004089e+06	1.000000	77.191964	28.570469	400.000000	2.000000	3.200000
75%	1.835229e+07	1.000000	77.282006	28.642758	700.000000	2.000000	3.700000
max	1.850065e+07	216.000000	174.832089	55.976980	800000.000000	4.000000	4.900000

- Checking Unique Values

In [30]:

```
resto_df.nunique()
```

Out[30]:

Restaurant ID	9551
Restaurant Name	7446
Country Code	15
City	141
Address	8918
Locality	1208
Locality Verbose	1265
Longitude	8120
Latitude	8677
Cuisines	1825
Average Cost for two	140
Currency	12
Has Table booking	2
Has Online delivery	2
Is delivering now	2
Switch to order menu	1
Price range	4
Aggregate rating	33
Rating color	6
Rating text	6
Votes	1012
dtype:	int64

- Checking shape

In [33]:

```
resto_df.shape
```

Out[33]:

```
(9551, 21)
```

# Task 1: Restaurant Ratings

- \*\*Analyze the distribution of aggregate

ratings and determine the most common rating range.\*\*

```
In [36]: agg_val_count = resto_df["Aggregate rating"].value_counts()
agg_val_count
```

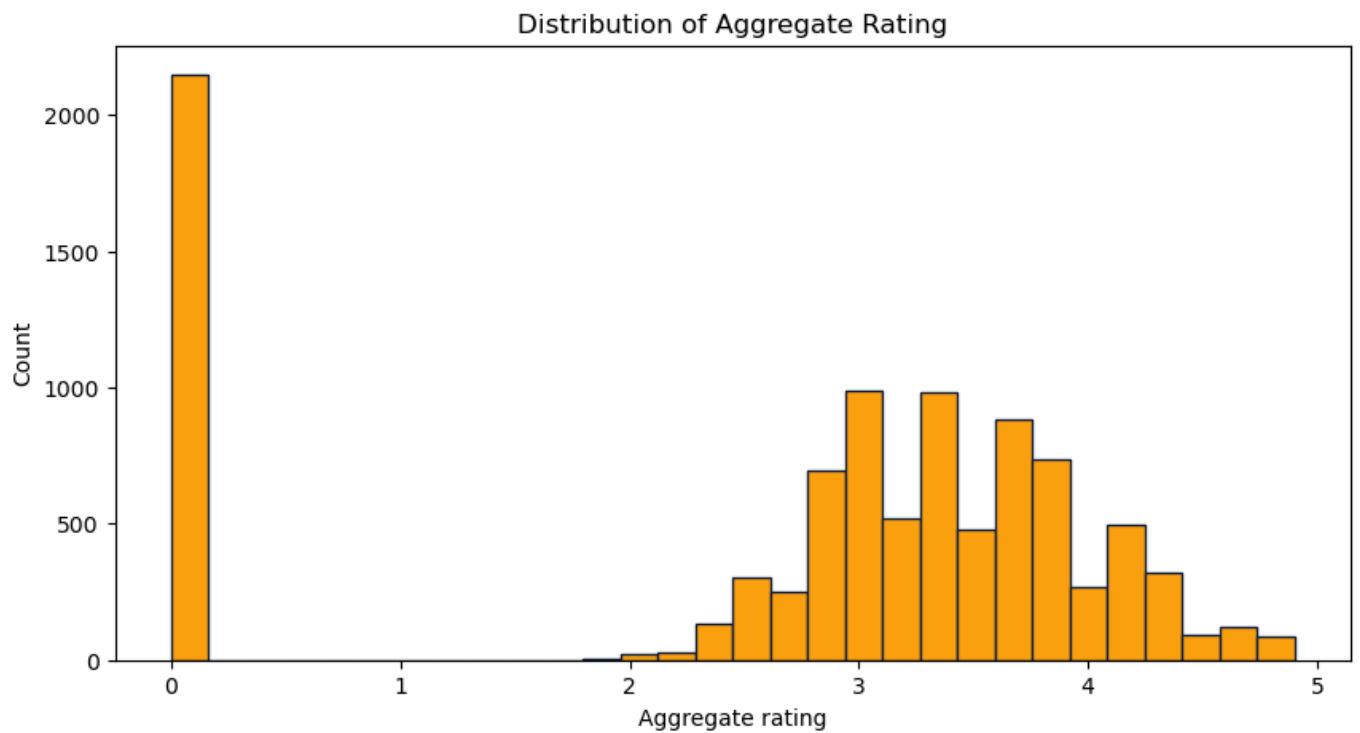
```
Out[36]: Aggregate rating
0.0      2148
3.2       522
3.1       519
3.4       498
3.3       483
3.5       480
3.0       468
3.6       458
3.7       427
3.8       400
2.9       381
3.9       335
2.8       315
4.1       274
4.0       266
2.7       250
4.2       221
2.6       191
4.3       174
4.4       144
2.5       110
4.5        95
2.4        87
4.6        78
4.9        61
2.3        47
4.7        42
2.2        27
4.8        25
2.1        15
2.0         7
1.9         2
1.8         1
Name: count, dtype: int64
```

```
In [38]: rating_most_common = agg_val_count.idxmax()
print(f'The Most Common rating range is: {rating_most_common}')
```

The Most Common rating range is: 0.0

```
In [40]: plt.figure(figsize=(10,5))
plt.hist(resto_df["Aggregate rating"], bins=30, color='#fca311', edgecolor='#14213d')
plt.xlabel('Aggregate rating')
plt.ylabel('Count')
plt.title('Distribution of Aggregate Rating')
plt.show
```

```
Out[40]: <function matplotlib.pyplot.show(close=None, block=None)>
```



- **\*\*Calculate the average number of votes**

received by restaurants.\*\*

```
In [43]: avg_vote = round(resto_df['Votes'].mean(), 2)
print(f'The Average number of Votes received by restaurants : {avg_vote}')
```

The Average number of Votes received by restaurants : 156.91

## Task 2: Cuisine Combination

- **\*\*Identify the most common combinations of**

cuisines in the dataset.\*\*

```
In [58]: common_cuisine_combinations = resto_df.groupby('Cuisines')['Aggregate rating'].mean().sort_values(ascending=False)
top_10_combinations = common_cuisine_combinations.head(10)
print(f'The Top 10 most common combinations are : {top_10_combinations}')
```

The Top 10 most common combinations are : Cuisines

Italian, Deli	4.9
Hawaiian, Seafood	4.9
American, Sandwich, Tea	4.9
Continental, Indian	4.9
European, Asian, Indian	4.9
European, Contemporary	4.9
European, German	4.9
BBQ, Breakfast, Southern	4.9
American, Coffee and Tea	4.9
Sunda, Indonesian	4.9

Name: Aggregate rating, dtype: float64

- **\*\*Determine if certain cuisine combinations**

tend to have higher ratings.\*\*



```
In [65]: max_rating = common_cuisine_combinations.iloc[0]
print(f'The Max Rating is: {max_rating}')
```

The Max Rating is: 4.9

```
In [67]: max Rated_rest = resto_df.loc[resto_df['Aggregate rating'] == max_rating]
print('Restorents having the Maximum Ratings: ')
max Rated_rest['Restaurant Name']
```

Restorents having the Maximum Ratings:

```
Out[67]: 3 Ooma
8 Spiral - Sofitel Philippine Plaza Manila
10 Silantro Fil-Mex
39 Coco Bambu
48 Braseiro da Govea
...
9484 Restaurant Mosaic @ The Orient
9514 Ministry of Crab
9524 Gaga Manjero
9538 Starbucks
9540 Draft Gastro Pub
Name: Restaurant Name, Length: 61, dtype: object
```

## Task 3: Geographic Analysis

- \*\*Plot the locations of restaurants on a

map using longitude and latitude coordinates.\*\*

```
In [70]: resto_df[['Latitude', 'Longitude']]
```

```
Out[70]:
```

	Latitude	Longitude
0	14.565443	121.027535
1	14.553708	121.014101
2	14.581404	121.056831
3	14.585318	121.056475
4	14.584450	121.057508
...	...	...
9546	41.022793	28.977392
9547	41.009847	29.041297
9548	41.055817	29.034640
9549	41.057979	29.036019
9550	40.984776	29.026016

9551 rows × 2 columns

```
In [72]: print(resto_df['Latitude'].isnull().sum())
print(resto_df['Longitude'].isnull().sum())
```

0  
0

```
In [157... fig = px.scatter_mapbox(resto_df,lat='Latitude', lon='Longitude',hover_name='Restaurant Name')
fig.update_layout(mapbox_style="open-street-map",)
```

- **Identify any patterns or clusters of restaurants in specific areas.**

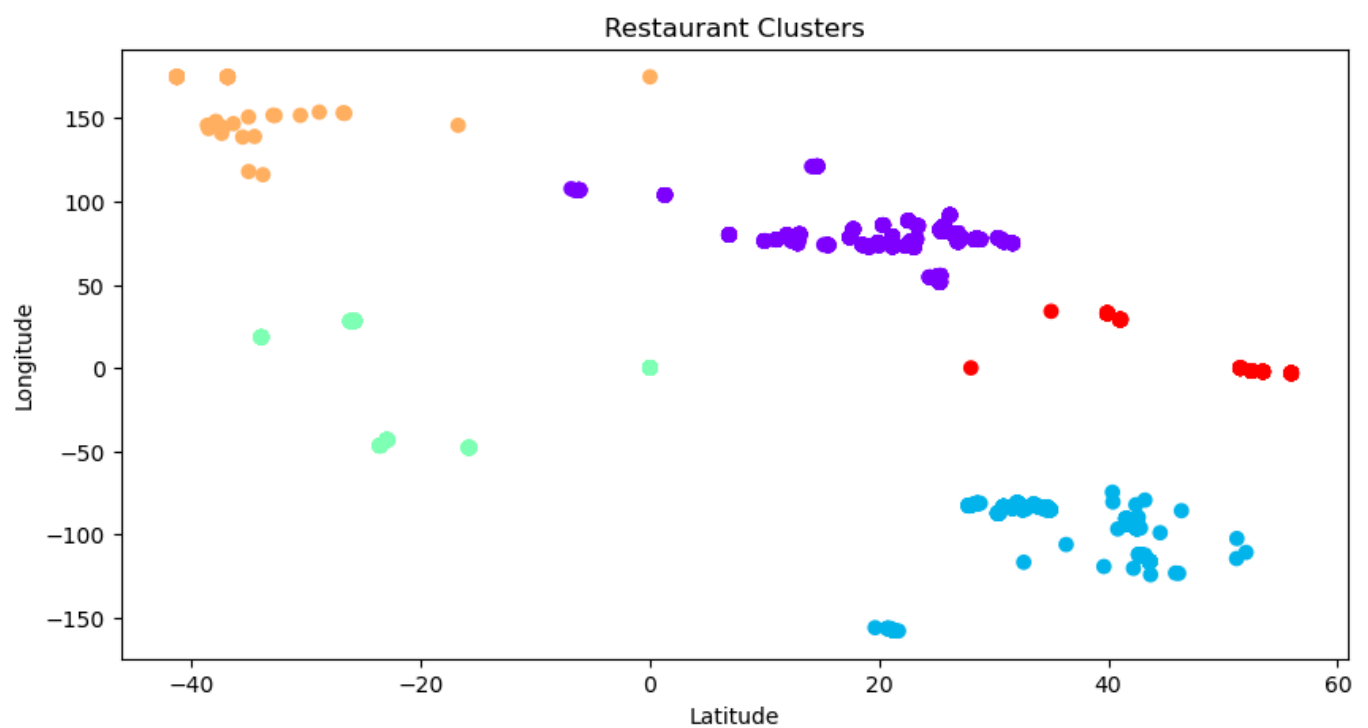
```
In [102... X = resto_df[['Latitude','Longitude']]
num_cluster = 5
kmeans=KMeans(n_clusters=num_cluster,n_init=10,random_state=42)
resto_df['Cluster']=kmeans.fit_predict(X)
```

```
In [119... fig=px.scatter_mapbox(resto_df,lat='Latitude',lon='Longitude',hover_name='Restaurant Name',ho
color_continuous_scale='reds',zoom=2,)

fig.update_layout(mapbox_style="open-street-map",)
```

In [131...

```
plt.figure(figsize=(10,5))
plt.scatter(resto_df['Latitude'],resto_df['Longitude'], c=resto_df['Cluster'], cmap='rainbow')
plt.title('Restaurant Clusters')
plt.xlabel('Latitude')
plt.ylabel('Longitude')
plt.show()
```



# Task 4: Restaurant Chains

- Identify if there are any restaurant chains present in the dataset

In [134...

```
resto_df.head(2)
```

Out[134...

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitude	Latitude	C
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	Jap D
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	Ja

2 rows × 22 columns



In [142...

```
res_count = resto_df['Restaurant Name'].value_counts()
potential_chains=res_count[res_count > 10].index
print("Potential restaurant chain:")
for chain in potential_chains:
    print(f"--{chain}")
```

Potential restaurant chain:

- Cafe Coffee Day
- Domino's Pizza
- Subway
- Green Chick Chop
- McDonald's
- Keventers
- Pizza Hut
- Giani
- Baskin Robbins
- Barbeque Nation
- Giani's
- Barista
- Dunkin' Donuts
- Costa Coffee
- Pind Balluchi
- Wah Ji Wah
- Twenty Four Seven
- Pizza Hut Delivery
- Sagar Ratna
- Republic of Chicken
- KFC
- Starbucks
- Chaayos
- Burger King
- Haldiram's
- Shree Rathnam
- Frontier
- Moti Mahal Delux
- Bikanervala
- Aggarwal Sweets
- Behrouz Biryani
- Karim's
- Bikaner Sweets
- Chicago Pizza
- Apni Rasoi
- 34, Chowringhee Lane
- Wow! Momo
- Madras Cafe
- Burger Point

In [148...

```
restaurant_chain_stats=resto_df.groupby('Restaurant Name').agg({'Aggregate rating':'mean','Vo
restaurant_chain_stats.columns=['Restaurant Name','Average rating','Total Votes']
restaurant_chain_stats=restaurant_chain_stats.sort_values(by='Total Votes',ascending=False)
print("Restaurant Chain Rating and Popularity Analysis (Sorted by Total Votes):")
print(restaurant_chain_stats.head(20))
```

# Restaurant Chain Rating and Popularity Analysis (Sorted by Total Votes):

	Restaurant Name	Average rating	Total Votes
663	Barbeque Nation	4.353846	28142
101	AB's - Absolute Barbecues	4.825000	13400
6943	Toit	4.800000	10934
785	Big Chill	4.475000	10853
2297	Farzi Cafe	4.366667	10098
6988	Truffles	3.950000	9682
1510	Chili's	4.580000	8156
2879	Hauz Khas Social	4.300000	7931
3261	Joey's Pizza	4.250000	7807
4902	Peter Cat	4.300000	7574
796	Big Yellow Door	4.266667	7511
5571	Saravana Bhavan	4.133333	7238
6080	Starbucks	3.805556	7139
4941	Pirates of Grill	4.025000	7091
3405	Karim's	3.030769	6878
2098	Domino's Pizza	2.740506	6643
6106	Subway	2.907937	6124
2145	Dunkin' Donuts	3.136364	5974
783	Big Brewsky	4.500000	5705
4924	Pind Balluchi	2.630000	5582

## Observations

- Restaurant Chain Rating and Popularity Analysis (Sorted by Total Votes)
  - Barbeque Nation
  - AB's - Absolute Barbecues
  - Toit
  - Big Chill
  - Farzi Cafe