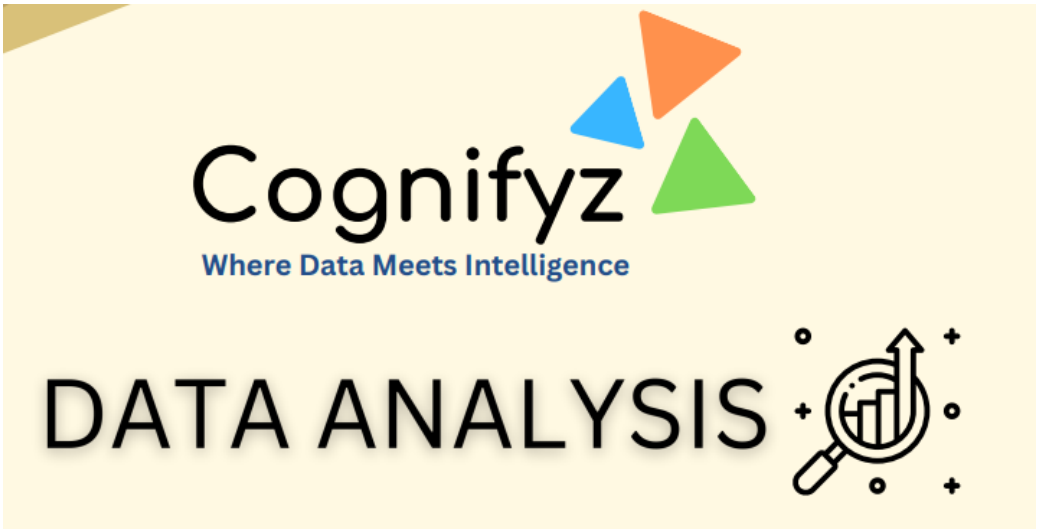


Task - Objective(Level 2)

- Expertise in Python programming and Data Manipulation
- Extract valuable insights from large datasets and drive informed decision-making.
- Data cleaning and preprocessing data, performing statistical analysis, or creating data visualizations,
- Proficiency in Python will play a crucial role in delivering meaningful results.

1.Load Python Modules

In [1]:
```
# Import all the packages - required for all the tasks
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from tabulate import tabulate
import itertools
import plotly.express as px
from sklearn.cluster import KMeans
```

2. Read the Dataset CSV - using Pandas

In [2]:
```
# read the data from csv file using  pandas
restaurant_df=pd.read_csv("Dataset.csv")
restaurant_df
```

Out[2]:

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitude	Latitude	Cuisines	...
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	French, Japanese, Desserts	...
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	Japanese	...
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	Seafood, Asian, Filipino, Indian	...
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	Japanese, Sushi	...
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	Japanese, Korean	...
...	...	...	...	...	...	...	...	...	...	...	...
9546	5915730	Naml\ Gurme	208	��stanbul	Kemanke�� Karamustafa Pa��a Mahallesi, R\ht\m ...	Karak�_y	Karak�_y, ��stanbul	28.977392	41.022793	Turkish	...
9547	5908749	Ceviz A��ac\	208	��stanbul	Ko��uyolu Mahallesi, Muhittin ��st�_nda�� Cadd...	Ko��uyolu	Ko��uyolu, ��stanbul	29.041297	41.009847	World Cuisine, Patisserie, Cafe	...
9548	5915807	Huqqa	208	��stanbul	Kuru�_e��me Mahallesi, Muallim Naci Caddesi, N...	Kuru�_e��me	Kuru�_e��me, ��stanbul	29.034640	41.055817	Italian, World Cuisine	...
9549	5916112	A���k Kahve	208	��stanbul	Kuru�_e��me Mahallesi, Muallim Naci Caddesi, N...	Kuru�_e��me	Kuru�_e��me, ��stanbul	29.036019	41.057979	Restaurant Cafe	...
9550	5927402	Walter's Coffee Roastery	208	��stanbul	Cafea��a Mahallesi, Bademalt\ Sokak, No 21/B, ...	Moda	Moda, ��stanbul	29.026016	40.984776	Cafe	...

9551 rows × 21 columns

3. Basic Inspection - dataset

In [3]:

```
def basic_inspection_dataset(table):

    print("top 5 rows - using head")
    print(table.head())
    print()

    print("bottom 5 rows using tail")
    print(table.tail())
    print()

    print("numbers of samples and columns")
    print(table.shape)
    print()

    print("numbers of samples ")
    print(len(table))
    print()

    print("numbers of entries in the data frame")
    print(table.size)
    print()

    print("Columns Names")
    print(table.columns)
    print()

    print("Columns dtypes")
    print(table.dtypes)
    print()

    print("Dataframe info")
    print(table.info())
    print()

    print()
    print("check the missing value in each column")
    print(table.isnull().sum())

    print()
    print("check the missing value in each column")
    print(table.isna().sum())

basic_inspection_dataset(restaurant_df)
```

```
top 5 rows - using head
   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  ...          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]

bottom 5 rows using tail
      Restaurant ID           Restaurant Name  Country Code       City  \
9546        5915730               Naml‘ Gurme           208  ��stanbul
9547        5908749              Ceviz A��ac‘           208  ��stanbul
9548        5915807                    Huqqa           208  ��stanbul
9549        5916112                A���k Kahve           208  ��stanbul
9550        5927402  Walter's Coffee Roastery           208  ��stanbul

                                             Address     Locality  \
9546  Kemanke�� Karamustafa Pa��a Mahallesi, R‘ht‘m ...     Karak�_y
9547  Ko��uyolu Mahallesi, Muhittin ��st�_nda�� Cadd...    Ko��uyolu
9548  Kuru�_e��me Mahallesi, Muallim Naci Caddesi, N...  Kuru�_e��me
9549  Kuru�_e��me Mahallesi, Muallim Naci Caddesi, N...  Kuru�_e��me
9550  Cafea��a Mahallesi, Bademalt‘ Sokak, No 21/B, ...         Moda

           Locality Verbose  Longitude   Latitude  \
9546    Karak�_y, ��stanbul  28.977392  41.022793
9547   Ko��uyolu, ��stanbul  29.041297  41.009847
9548  Kuru�_e��me, ��stanbul  29.034640  41.055817
9549  Kuru�_e��me, ��stanbul  29.036019  41.057979
9550        Moda, ��stanbul  29.026016  40.984776

                           Cuisines  ...          Currency  \
9546                         Turkish  ...  Turkish Lira(TL)
9547  World Cuisine, Patisserie, Cafe  ...  Turkish Lira(TL)
9548          Italian, World Cuisine  ...  Turkish Lira(TL)
9549                Restaurant Cafe  ...  Turkish Lira(TL)
9550                           Cafe  ...  Turkish Lira(TL)

     Has Table booking Has Online delivery Is delivering now  \
9546               No                 No                No
9547               No                 No                No
9548               No                 No                No
```

```
9549               No                  No                No
9550               No                  No                No

     Switch to order menu Price range  Aggregate rating  Rating color  \
9546                   No           3               4.1         Green
9547                   No           3               4.2         Green
9548                   No           4               3.7        Yellow
9549                   No           4               4.0         Green
9550                   No           2               4.0         Green

     Rating text Votes
9546   Very Good   788
9547   Very Good  1034
9548        Good   661
9549   Very Good   901
9550   Very Good   591

[5 rows x 21 columns]

numbers of samples and columns
(9551, 21)

numbers of samples
9551

numbers of entries in the data frame
200571

Columns Names
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')

Columns dtypes
Restaurant ID             int64
Restaurant Name          object
Country Code              int64
City                     object
Address                  object
Locality                 object
Locality Verbose         object
Longitude               float64
Latitude                float64
Cuisines                 object
Average Cost for two      int64
Currency                 object
Has Table booking        object
Has Online delivery      object
Is delivering now        object
Switch to order menu     object
Price range               int64
Aggregate rating        float64
Rating color             object
Rating text              object
Votes                     int64
dtype: object

Dataframe 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
None


check the missing value in each column
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

check the missing value in each column
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
```

4. Handling Missing Values

In [4]:
```
#For a categorical variable, determine the most frequent value, known as the mode.
cuisine_mode = restaurant_df['Cuisines'].mode()[0]
print(cuisine_mode)

# fill the missing value with mode
restaurant_df['Cuisines'].fillna(cuisine_mode,inplace=True)

# check for missing values - for confirmation
restaurant_df.isnull().sum()
```

```
North Indian
```

Out[4]:
```
Restaurant ID           0
Restaurant Name         0
Country Code            0
City                    0
Address                 0
Locality                0
Locality Verbose        0
Longitude               0
Latitude                0
Cuisines                0
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
```

Level 2, Task 1: Restaurant Ratings

2.1.1 Analyze the distribution of aggregate ratings and determine the most common rating range.

In [5]:
```
def distribution_rating(rating,bins):
    # Create a figure and axes object
    fig, axes = plt.subplots(1, 2, figsize=(14, 6))

    # Plot histogram without KDE on the left
    axes[0].hist(restaurant_df[rating], bins=bins, color='skyblue', edgecolor='black')
    axes[0].set_xlabel('Ratings Value')
    axes[0].set_ylabel('Frequency')
    axes[0].set_title('Restaurant Ratings Histogram')

    # Plot histogram with KDE on the right
    sns.histplot(data=restaurant_df, x=rating, bins=bins, kde=True, color='orange', edgecolor='black',
ax=axes[1])
    axes[1].set_xlabel('Ratings Value')
    axes[1].set_ylabel('Density')
    axes[1].set_title('Histogram with KDE')

    # Adjust layout
    plt.tight_layout()
    plt.show()
```

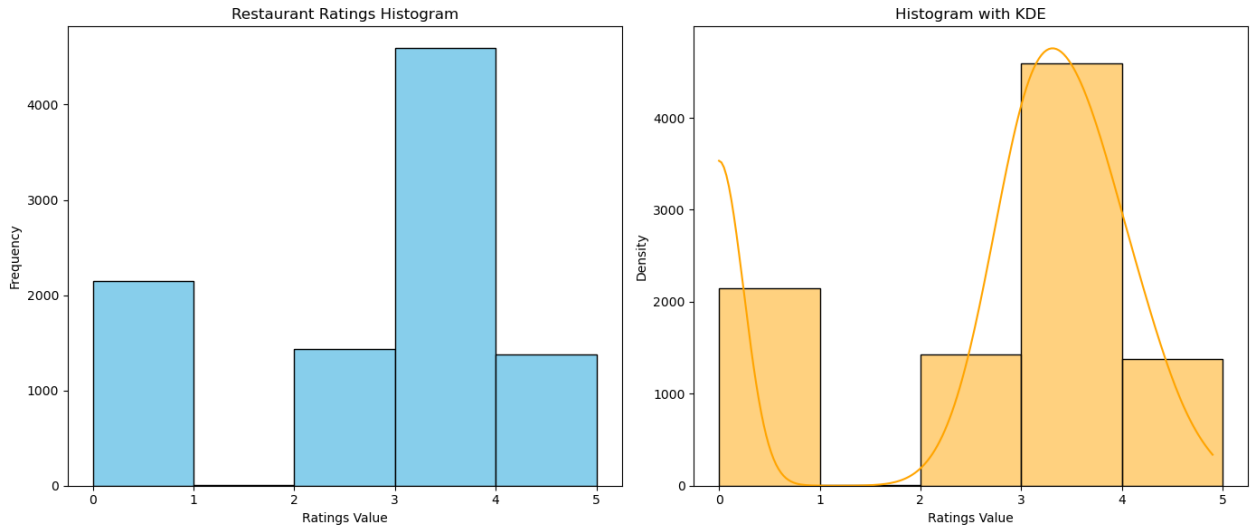
In [6]:
```
print("Rating Max Count -", restaurant_df["Aggregate rating"].max())
print("Rating Min Count - ", restaurant_df["Aggregate rating"].min())
```

```
Rating Max Count - 4.9
Rating Min Count -  0.0
```

In [7]:
```
bins = [x for x in range(0,6,1)]
print(bins)
```

```
[0, 1, 2, 3, 4, 5]
```

In [8]:
```
distribution_rating("Aggregate rating",bins)
```



Observations

- Ratings Range
  1. 0-1
  2. 1-2 - Min
  3. 2-3
  4. 3-4 - Max
  5. 4-5

2.1.2 Calculate the average number of votes received by restaurants.

In [9]:
```
# Average votes received by the restaurent
avg_votes=restaurant_df['Votes'].mean()
print("Average votes received by the restaurent")
round(avg_votes,2)
```

```
Average votes received by the restaurent
```

Out[9]:
```
156.91
```

Observations

- Average votes received by the restaurent is 156.91

Level 2, Task 2: Cuisine Combination

2.2.1 Identify the most common combinations of cuisines in the dataset

In [10]:
```
restaurant_df['Cuisines'] = restaurant_df['Cuisines'].str.split(',')
```

In [11]:
```
comninations_list = []
for i in restaurant_df['Cuisines']:
    comninations_list.extend(set(c) for c in itertools.combinations(i, 2))

combination_counts = pd.Series(comninations_list).value_counts()
print(combination_counts.head())
```

```
{ Chinese, North Indian}      1314
{ Mughlai, North Indian}       689
{ Chinese,  Mughlai}           323
{ Fast Food, North Indian}     296
{ Chinese,  North Indian}      268
Name: count, dtype: int64
```

Observations

- Cuisine Combinations
  1. Chinese, North Indian is in top position
  2. Mughlai, North Indian is in second position

3. Chinese, Mughlai is in third position

2.2.2 Determine if certain cuisine combinations tend to have higher ratings.

In [12]:
```
# Assuming 'idf' is your DataFrame
restaurant_df['Cuisines'] = restaurant_df['Cuisines'].apply(lambda x: ', '.join(x) if isinstance(x, list) else x)

# Display the updated DataFrame
print(restaurant_df['Cuisines'])

avg_rating=restaurant_df.groupby('Cuisines')['Aggregate rating'].mean()

# Average rating in descending order
avg_rating=avg_rating.sort_values(ascending=False)
print('The Cuisines Combination that have higher ratings:')
avg_rating.head()
```

```
0             French,  Japanese,  Desserts
1                              Japanese
2       Seafood,  Asian,  Filipino,  Indian
3                       Japanese,  Sushi
4                      Japanese,  Korean
                    ...
9546                           Turkish
9547      World Cuisine,  Patisserie,  Cafe
9548               Italian,  World Cuisine
9549                     Restaurant Cafe
9550                               Cafe
Name: Cuisines, Length: 9551, dtype: object
The Cuisines Combination that have higher ratings:
```

Out[12]:
```
Cuisines
Italian,  Deli              4.9
Hawaiian,  Seafood          4.9
American,  Sandwich,  Tea   4.9
Continental,  Indian        4.9
European,  Asian,  Indian   4.9
Name: Aggregate rating, dtype: float64
```

Observations

- Certain cuisine combinations tend to have higher ratings.
  1. Italian, Deli
  2. Hawaiian, Seafood
  3. American, Sandwich, Tea
  4. Continental, Indian
  5. European, Asian, Indian

Level 2, Task 3: Geographic Analysis

2.3.1 Plot the locations of restaurants on a map using longitude and latitude coordinates

In [13]:
```
print(restaurant_df["Longitude"].isnull().sum())
print(restaurant_df["Latitude"].isnull().sum())
```

```
0
0
```

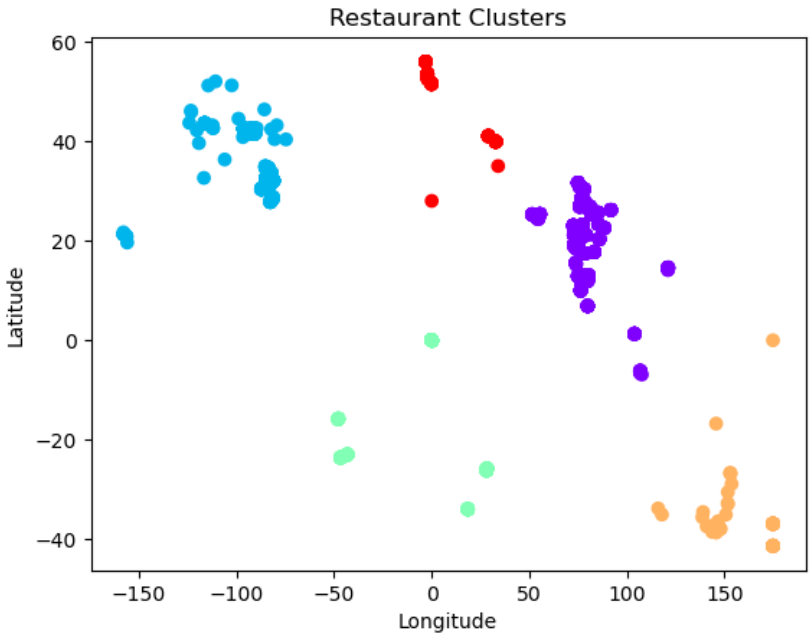
In [14]:
```
# plot the restaurents on the map
fig = px.scatter_mapbox(restaurant_df,lat='Latitude',  lon='Longitude',
     hover_name='Restaurant Name',       color_discrete_sequence=['red'],
     zoom=2,
)
fig.update_layout(
    mapbox_style="open-street-map",
)
```

2.3.2 Identify any patterns or clusters of restaurants in specific areas.

In [15]:
```
X=restaurant_df[['Latitude','Longitude']]
num_cluster=5
# k mean clustering
kmeans=KMeans(n_clusters=num_cluster,n_init=10,random_state=42)
restaurant_df['cluster']=kmeans.fit_predict(X)
```

In [16]:
```
# plot on the map
fig=px.scatter_mapbox(  restaurant_df,    lat='Latitude',    lon='Longitude',
    hover_name='Restaurant Name',    hover_data=['Cuisines','Country Code'],
    color='cluster',    color_continuous_scale='reds',
    zoom=2,
)
fig.update_layout(
    mapbox_style="open-street-map",
)
```

In [17]:
```
# Plotting the clusters
plt.scatter(restaurant_df['Longitude'], restaurant_df['Latitude'], c=restaurant_df['cluster'],
cmap='rainbow')
plt.title('Restaurant Clusters')
plt.xlabel('Longitude')
plt.ylabel('Latitude')
plt.show()
```



Level 2, Task 4: Restaurant Chains

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

In [18]:
```
1 restaurant_df.head(2)
```

Out[18]:

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitude	Latitude	Cuisines	...	Has Table booking	Has Online delivery	I deliverin nov
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	French, Japanese, Desserts	...	Yes	No	N
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	Japanese	...	Yes	No	N

2 rows × 22 columns

In [19]:
```
1 res_count=restaurant_df['Restaurant Name'].value_counts()
2 potential_chains=res_count[res_count > 10].index
3 print("Potential restaurant chains:")
4 for chain in potential_chains:
5     print(f"--{chain}")
```

```
Potential restaurant chains:
--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
```

2.4.2 Analyze the ratings and popularity of different restaurant chains.

In [20]:
```
restaurant_chain_stats=restaurant_df.groupby('Restaurant Name').agg({
    'Aggregate rating':'mean',
    'Votes':'sum',
}).reset_index()

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
662             Barbeque Nation        4.353846        28142
100   AB's - Absolute Barbecues        4.825000        13400
6943                      Toit        4.800000        10934
784                   Big Chill        4.475000        10853
2296                 Farzi Cafe        4.366667        10098
6988                   Truffles        3.950000         9682
1509                    Chili's        4.580000         8156
2878           Hauz Khas Social        4.300000         7931
3260               Joey's Pizza        4.250000         7807
4902                  Peter Cat        4.300000         7574
795             Big Yellow Door        4.266667         7511
5571            Saravana Bhavan        4.133333         7238
6080                  Starbucks        3.805556         7139
4941           Pirates of Grill        4.025000         7091
3404                    Karim's        3.030769         6878
2097             Domino's Pizza        2.740506         6643
6106                     Subway        2.907937         6124
2144             Dunkin' Donuts        3.136364         5974
782                 Big Brewsky        4.500000         5705
4924              Pind Balluchi        2.630000         5582
```

Observations

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