In [17]:
```python
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
```

In [9]:
```python
path=r"C:\Users\Sruth\Downloads\cognifyz dataset.csv"
restaurant_df=pd.read_csv(path)
restaurant_df
```

Out[9]:

	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, Rhtm ...	Karak�_y	Karak�_y, ��stanbul	28.977392	41.022793	Turkish	...
9547	5908749	Ceviz A��acl	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

In [10]:
```python
# handling missing values
#For a categorical variable, determine the most frequent value, known as the mode.
cuisine_mode = resturant_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 confirmatio
restaurant_df.isnull().sum()
```

```
North Indian
```

Out[10]:
```
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

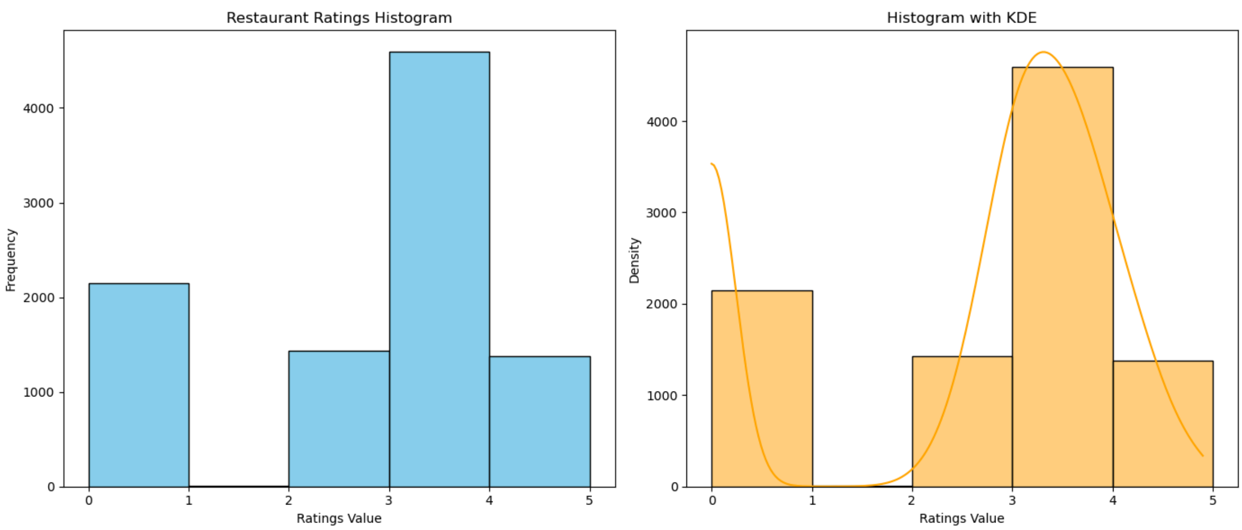
Analyzethe distribution of aggreate ratings and determine the most common rating range

In [11]:
```
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 [12]:
```
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 [19]:
```
import warnings
warnings.filterwarnings('ignore')
bins = [x for x in range(0,6,1)]

distribution_rating("Aggregate rating",bins)
```



In [20]:
```
# Calculate the average number of votes recieved by resturants
# 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[20]:

```
156.91
```

TASK-2 :CUISINE COMBINATION

In [21]:

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

In [22]:

```
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())
```

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

In [23]:

```
# Determine if certain cuisine combinations tend to have higher ratings
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[23]:

```
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
```

TASK-3 : Geographicla analysis

Plot the locations of restutants on a map using latitude coordinate

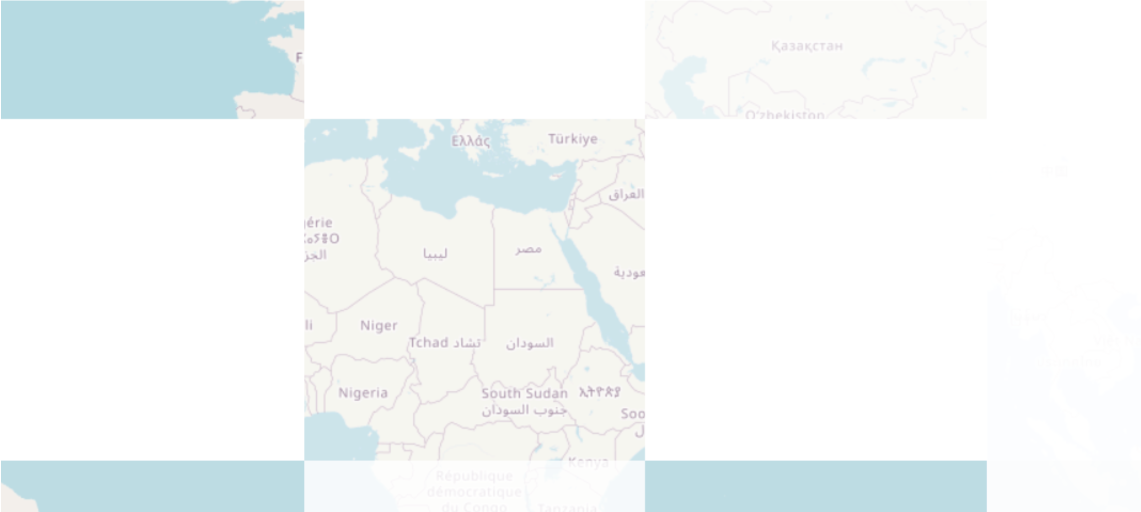
In [24]:

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

```
0
0
```

In [25]:

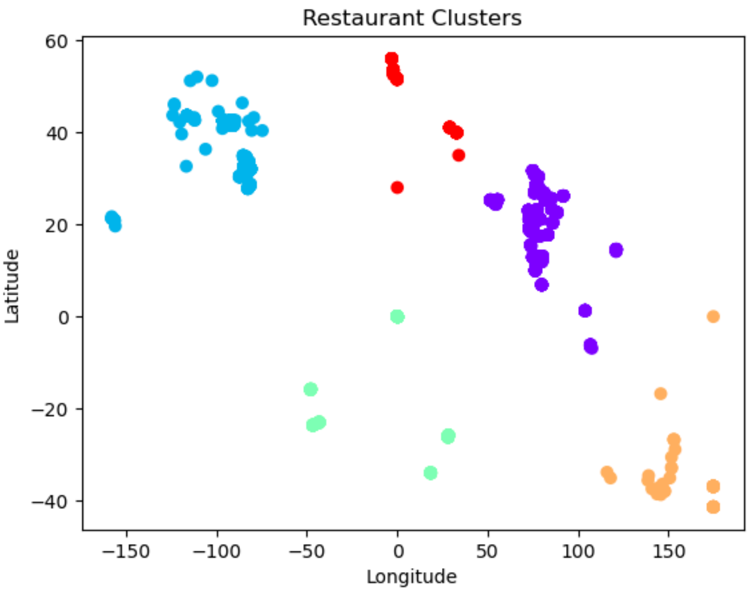
```
# plot the restaurents on the map
fig = px.scatter_mapbox(restaurant_df,lat='Latitude',  lon='Longitude',
     hover_name='Restaurant Name',       color_discrete_sequence=['red'],
     zoom=2,
)
fig.update_layout(
    mapbox_style="open-street-map",
)
```



Identify any patterns or clusters of reaturants in specific areas

In [26]:
```
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 [28]:
```
# 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()
```



TASK-4:Resturant chains

Identify if there is any resturant chains avaliable in dataset

In [29]:
```
restaurant_df.head(2)
```

Out[29]:

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitude	Latitude	Cuisines	...	Has Table booking	Has Online delivery	deliverir no
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 [30]:
```
res_count=restaurant_df['Restaurant Name'].value_counts()
potential_chains=res_count[res_count > 10].index
print("Potential restaurant chains:")
for chain in potential_chains:
    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
```

Analysis the ratings and popularity of different restutrants chains

In [31]:
```
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
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
```

In [ ]: