In [1]:
```
import pandas as pd
```

In [2]:
```
df=pd.read_csv("dataset restrarant new.csv" )
```

In [3]:
```
df
```

Out[3]:

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitude	Latitude	Cuisines	Averag Cost fo tw
0	51040	Truffles	1	Bangalore	28, 4th 'B' Cross, Koramangala 5th Block, Bang...	Koramangala 5th Block	Koramangala 5th Block, Bangalore	77.614293	12.933298	American, Burger, Cafe	80
1	51705	Toit	1	Bangalore	298, Namma Metro Pillar 62, 100 Feet Road, Ind...	Indiranagar	Indiranagar, Bangalore	77.640709	12.979166	Italian, American, Pizza	200
2	308322	Hauz Khas Social	1	New Delhi	9A & 12, Hauz Khas Village, New Delhi	Hauz Khas Village	Hauz Khas Village, New Delhi	77.194471	28.554285	Continental, American, Asian, North Indian	160
3	20404	Peter Cat	1	Kolkata	18A, Park Street, Park Street Area, Kolkata	Park Street Area	Park Street Area, Kolkata	88.352885	22.552672	Continental, North Indian	100
4	56618	AB's Absolute Barbecues	1	Bangalore	90/4, 3rd Floor, Outer Ring Road, Munnekollaly...	Marathahalli	Marathahalli, Bangalore	77.699386	12.949934	European, Mediterranean, North Indian	140
...	...	...	...	...	...	...	...	...	...	...	
9543	18486858	6 Packs Momos	1	Noida	Spice World Mall, Sector 25, Noida	Spice World Mall, Sector 25	Spice World Mall, Sector 25, Noida	77.340602	28.586000	Chinese	30
9544	18431152	Cafe' Wow	1	Noida	Food Court, 3rd Floor, The Great India Palace ...	The Great India Place, Sector 38	The Great India Place, Sector 38, Noida	77.325600	28.567514	Fast Food	20
9545	18439721	Chef's Basket Pop Up Cafe	1	Noida	Inside Big Bazaar, The Great India Place, Sect...	The Great India Place, Sector 38	The Great India Place, Sector 38, Noida	0.000000	0.000000	Italian, Chinese	20

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitude	Latitude	Cuisines	Averag Cost f tw
9546	18428504	The HangoutDeli	1	Noida	320A, 3rd Floor, The Great India Place Mall, S...	The Great India Place, Sector 38	The Great India Place, Sector 38, Noida	77.323213	28.567751	Continental, Lebanese, Mexican	10(
9547	18254559	Platters	1	Noida	Shop 1, Tulip Mall, Near HDFC Bank, Sector 48,...	Tulip Mall, Sector 48, Noida	Tulip Mall, Sector 48, Noida, Noida	77.367322	28.557930	North Indian, Chinese	5(

9548 rows × 20 columns

level 1

TASK 1: TOP CUISINES

In [4]:
```
top_cuisines = df['Cuisines'].value_counts().head(3)
print(top_cuisines)
```

```
North Indian            936
North Indian, Chinese   511
Chinese                 354
Name: Cuisines, dtype: int64
```

In [5]:
```
Cuisines_counts = df['Cuisines'].value_counts()
Cuisines_percentages = (Cuisines_counts / len(df)) * 100
top_Cuisines = Cuisines_percentages.head(3)
print(top_Cuisines)
```

```
North Indian            9.803100
North Indian, Chinese   5.351906
Chinese                 3.707583
Name: Cuisines, dtype: float64
```

TASK 2: City Analysis

In [6]:
```
top_City = df['City'].value_counts().idxmax()

print("City with the highest number of restaurants:", top_City)
```

```
City with the highest number of restaurants: New Delhi
```

In [7]:
```
df["Aggregate rating"] = pd.to_numeric(df["Aggregate rating"], errors="coerce")
average_ratings = df.groupby('City')['Aggregate rating'].mean()
print(average_ratings)
```

```
City
Abu Dhabi          4.300000
Agra               3.965000
Ahmedabad          4.161905
Albany             3.555000
Allahabad          3.395000
                       ...
Weirton            3.900000
Wellington City    4.250000
Winchester Bay     3.200000
Yorkton            3.300000
estanbul           4.292857
Name: Aggregate rating, Length: 141, dtype: float64
```

In [8]:
```
df["Aggregate rating"] = pd.to_numeric(df["Aggregate rating"], errors="coerce")

average_ratings = df.groupby("City")["Aggregate rating"].mean()

top_city = average_ratings.idxmax()

print("City with the highest average rating:", top_city)
```

```
City with the highest average rating: Inner City
```

Task 3: Price Range Distribution

In [9]:
```
import matplotlib.pyplot as plt
```

In [10]:
```
price_range_counts = df["Price range"].value_counts().sort_index()

# Plot the bar chart
plt.figure(figsize=(10, 5))
plt.bar(price_range_counts.index, price_range_counts.values, color="purple")
plt.xlabel("Price Range")
plt.ylabel("Number of Restaurants")
plt.title("Distribution of Price Ranges Among Restaurants")
plt.xticks(price_range_counts.index)
plt.grid(axis="y", linestyle="--", alpha=0.9)

# Show the plot
plt.show()
```

In [11]:
```
# Count occurrences of each price range and calculate percentages
price_range_percentages = (df["Price range"].value_counts(normalize=True) * 100).sort_index()

# Display the result
print(price_range_percentages)
```

```
1    46.533305
2    32.593213
3    14.736070
4     6.137411
Name: Price range, dtype: float64
```

Task 4: ONLINE DELIVERY

In [12]:
```
online_delivery_count = df["Has Online delivery"].value_counts(normalize=True) * 100

print(online_delivery_count)
```

```
False    74.329703
True     25.670297
Name: Has Online delivery, dtype: float64
```

In [13]:
```
df["Aggregate rating"] = pd.to_numeric(df["Aggregate rating"], errors="coerce")

average_ratings = df.groupby("Has Online delivery")["Aggregate rating"].mean()
print(average_ratings)
```

```
Has Online delivery
False    3.467453
True     3.381274
Name: Aggregate rating, dtype: float64
```

LEVEL 2

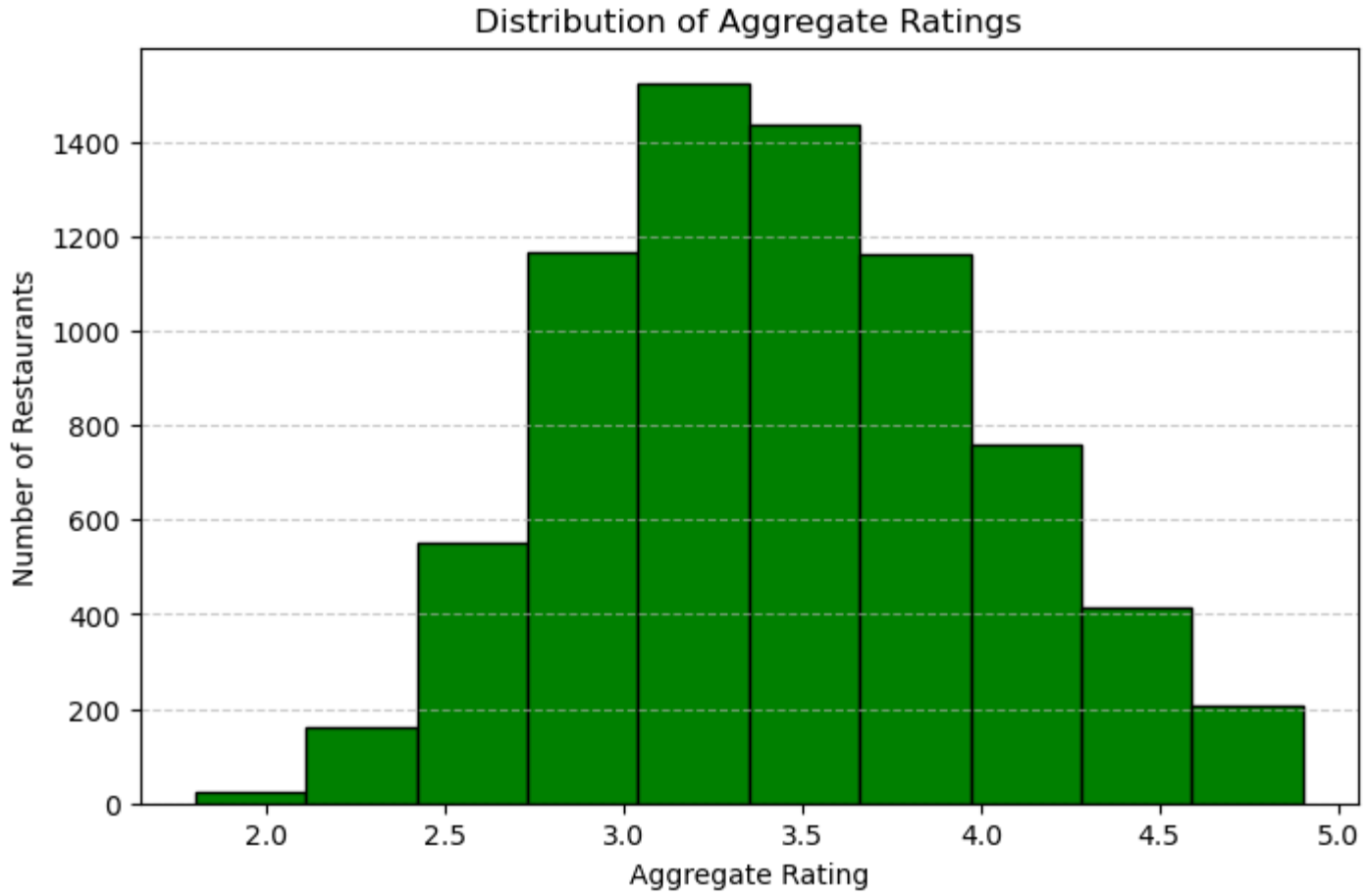
Task 1: Restrarant rating

In [14]:
```
df["Aggregate rating"] = pd.to_numeric(df["Aggregate rating"], errors="coerce")

plt.figure(figsize=(8, 5))
plt.hist(df["Aggregate rating"].dropna(), bins=10, color="Green", edgecolor="black")
plt.xlabel("Aggregate Rating")
plt.ylabel("Number of Restaurants")
plt.title("Distribution of Aggregate Ratings")
plt.grid(axis="y", linestyle="--", alpha=0.7)

# Show the plot
plt.show()

most_common_rating_range = df["Aggregate rating"].round(1).value_counts().idxmax()
print("Most common rating:", most_common_rating_range)
```




```
Most common rating: 3.2
```

In [15]:
```
df["Votes"] = pd.to_numeric(df["Votes"], errors="coerce")

average_votes = df["Votes"].mean()

average_votes = round(df["Votes"].mean(), 2)

print("Average number of votes received by restaurants:", average_votes)
```

```
Average number of votes received by restaurants: 156.78
```

Task 2: Cuisines Combination

In [16]:
```
cuisine_combinations = df["Cuisines"].value_counts()
print("Most common cuisine combinations:")
print(cuisine_combinations.head(5))  # Show top 10
```

```
Most common cuisine combinations:
North Indian              936
North Indian, Chinese     511
Chinese                   354
Fast Food                 353
North Indian, Mughlai     333
Name: Cuisines, dtype: int64
```

In [17]:
```
df["Aggregate rating"] = pd.to_numeric(df["Aggregate rating"], errors="coerce")

cuisine_ratings = df.groupby("Cuisines")["Aggregate rating"].mean()

top_cuisine_combinations = cuisine_ratings.sort_values(ascending=False)
print("Cuisine combinations with highest ratings:")
print(top_cuisine_combinations.head(5))
```

```
Cuisine combinations with highest ratings:
Cuisines
Burger, Bar Food, Steak         4.9
American, BBQ, Sandwich         4.9
Sunda, Indonesian               4.9
European, Asian, Indian         4.9
Italian, Bakery, Continental    4.9
Name: Aggregate rating, dtype: float64
```

Task 3: Geographic Analysis

In [18]:
```
df["Longitude"] = pd.to_numeric(df["Longitude"], errors="coerce")
df["Latitude"] = pd.to_numeric(df["Latitude"], errors="coerce")

# Plot the locations
plt.figure(figsize=(10, 6))
plt.scatter(df["Longitude"], df["Latitude"], alpha=0.5, c="skyblue", edgecolors="black")
plt.xlabel("Longitude")
plt.ylabel("Latitude")
plt.title("Restaurant Locations on Map")
plt.grid(True)

# Show the plot
plt.show()
```

In [19]:
```
from sklearn.cluster import KMeans
```

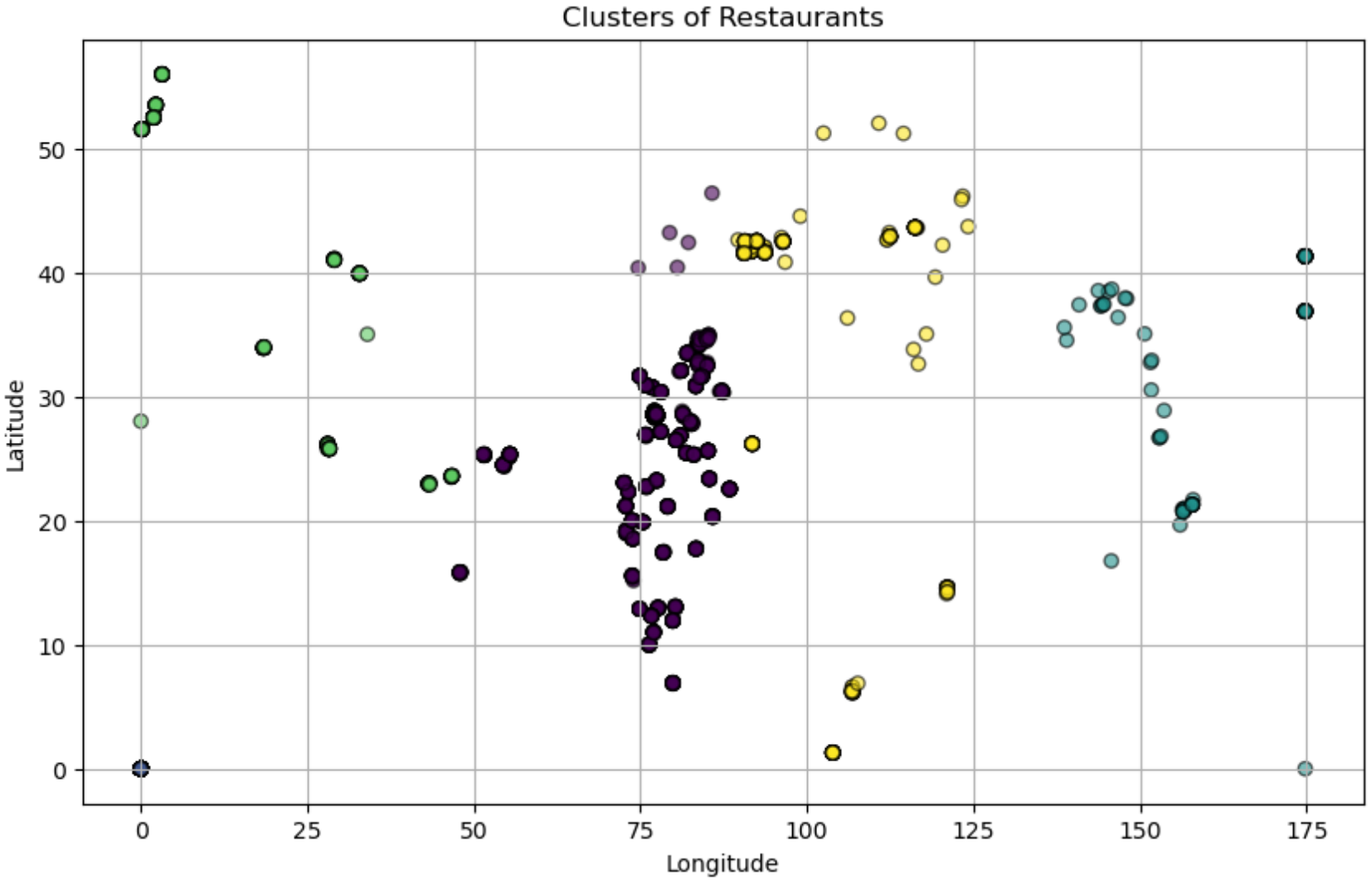
In [20]:
```
df["Longitude"] = pd.to_numeric(df["Longitude"], errors="coerce")
df["Latitude"] = pd.to_numeric(df["Latitude"], errors="coerce")


df = df.dropna(subset=["Longitude", "Latitude"])


kmeans = KMeans(n_clusters=5, random_state=42, n_init=10)
df["Cluster"] = kmeans.fit_predict(df[["Longitude", "Latitude"]])

#Plot the clusters
plt.figure(figsize=(10, 6))
plt.scatter(df["Longitude"], df["Latitude"], c=df["Cluster"], cmap="viridis", alpha=0.6, edgecolors="black")
plt.xlabel("Longitude")
plt.ylabel("Latitude")
plt.title("Clusters of Restaurants")
plt.grid(True)

# Show the plot
plt.show()
```



Task 4:Restrarant Chains

In [21]:
```
restaurant_counts = df["Restaurant Name"].value_counts()

restaurant_chains = restaurant_counts[restaurant_counts > 1]

 #Display the result
print("Restaurant chains found in the dataset:")
print(restaurant_chains)
```

```
Restaurant chains found in the dataset:
Cafe Coffee Day     83
Domino's Pizza      79
Subway              63
Green Chick Chop    51
McDonald's          48
                    ..
Chaap Point          2
Punjabi Chicken      2
Shawarma King's      2
Italiano             2
Wood Box Cafe        2
Name: Restaurant Name, Length: 736, dtype: int64
```

In [22]:
```
df["Aggregate rating"] = pd.to_numeric(df["Aggregate rating"], errors="coerce")
df["Votes"] = pd.to_numeric(df["Votes"], errors="coerce")

restaurant_counts = df["Restaurant Name"].value_counts()
restaurant_chains = restaurant_counts[restaurant_counts > 1].index

chain_data = df[df["Restaurant Name"].isin(restaurant_chains)]
chain_analysis = chain_data.groupby("Restaurant Name").agg(
    Average_Rating=("Aggregate rating", "mean"),
    Total_Votes=("Votes", "sum")
).sort_values(by="Average_Rating", ascending=False)

# Display the result
print("Ratings and Popularity of Restaurant Chains:")
print(chain_analysis.head(10))  # Show top 10 chains
```

```
Ratings and Popularity of Restaurant Chains:
                         Average_Rating  Total_Votes
Restaurant Name
Talaga Sampireun                  4.900         5514
Silantro FilMex                   4.850         1364
AB's Absolute Barbecues           4.850         3151
AB's  Absolute Barbecues          4.825        13400
Naturals Ice Cream                4.800         3094
Gymkhana                          4.700          328
The Cheesecake Factory            4.650         3010
Garota de Ipanema                 4.600           59
Dishoom                           4.600         1269
Chili's                           4.580         8156
```

In [ ]: