Level 2: Task

Restaurant_ratings

```
import pandas as pd
import numpy as np

data = pd.read_csv('Datasetc.csv')
data
data.head()
data.columns
data.isnull().sum()
data.describe()

average_votes = data['Votes'].mean()
average_votes
rating_distribution = data['Aggregate rating'].value_counts().sort_index()
rating_distribution
most_common_range = rating_distribution.idxmax()
most_common_range
```

Cuisine_combination

```
import pandas as pd
from itertools import combinations
from collections import Counter

data = pd.read_csv('Datasetc.csv')
data

data.head()
data.columns
```

```
data.isnull().sum()


def Cuisines(Cuisines_str):
  Cuisines = [c.strip() for c in Cuisines_str.split(',')]
  return list(combinations(sorted(Cuisines), 2))
data['Cuisine_combinations'] = data['Cuisines'].apply(Cuisines)


all_combinations = [combo for sublist in data['Cuisine_combinations'] for combo in sublist]
combo_counts = Counter(all_combinations)
combo_counts_df = pd.DataFrame(combo_counts.items(), columns=['Combinations', 'Count'])
combo_counts_df = combo_counts_df.sort_values(by='Count', ascending=False)
print("Most Common Cuisine Combinations:")
print(combo_counts_df)


expanded_combinations = data.explode('Cuisine_combinations','Aggregate_rating')
print(expanded_combinations.head())


rating_by_combo =expanded_combinations.groupby('Cuisine_combinations')['Aggregate
rating'].mean().reset_index()


print(rating_by_combo.head())
rating_by_combo = rating_by_combo.sort_values(by='Aggregate rating', ascending=False)
print(rating_by_combo)
print("Cuisine Combinations with Higher Ratings:")
print(rating_by_combo.head())
```

Geographic Analysis

```
import pandas as pd
import plotly.express as px
df = pd.read_csv('Datasetc.csv')
fig = px.scatter_geo(df,
                     lat='Latitude',
```

```
                lon='Longitude',
                title='Restaurant Locations')
fig.show()
```

Restaurant Chains

```
import numpy as np
import pandas as pd
import plotly.express as px

df = pd.read_csv('Datasetc.csv')
df

chain_counts = df['Restaurant Name'].value_counts()
print(chain_counts)

chains = chain_counts[chain_counts > 50]
print(chains)

total_chars = df['Rating text'].apply(len).sum()
print(f'Total number of characters in the column: {total_chars}')
chain_analysis = df.groupby('Restaurant Name').agg({
   'Rating text': 'count',
   'Votes': 'sum',
   'Aggregate rating': 'sum'
})
print(chain_analysis)

chain_analysis = chain_analysis.loc[chains.index]
print(chain_analysis)
fig_rating = px.bar(chain_analysis, x=chain_analysis.index, y='Rating text', title='Average Ratings by Restaurant Chain')
```

```
fig_rating.show()


fig_popularity = px.bar(chain_analysis, x=chain_analysis.index, y='Votes', title='Total Votes by Restaurant Chain')

fig_popularity.show()
```