Level 2 Data Analysis Report
Cognifyz Technologies Internship Program
Date: October 2025

Intern Name: Shazia Afraz

Organization: Cognifyz Technologies Internship

Duration: 28 September 2025 – 28 October 2025

Program: Data Analysis Internship With Python

Introduction

The Level 2 analysis investigates key patterns influencing the restaurant market by examining customer ratings, cuisine diversity, location clusters, and the performance of restaurant chains. This stage offers valuable insights into customer engagement, popular culinary trends, geographic concentration, and the operational reach and reputation of major brands. These insights help stakeholders make informed decisions regarding growth, differentiation, and enhancing the customer experience.

Task 1: Restaurant Ratings

Objective:

To analyze the distribution of aggregate ratings for restaurants, determine the most common rating range, and assess engagement by calculating the average number of votes received (avg votes = 156.99). This insight aims to understand customer satisfaction trends and public engagement with restaurant services.

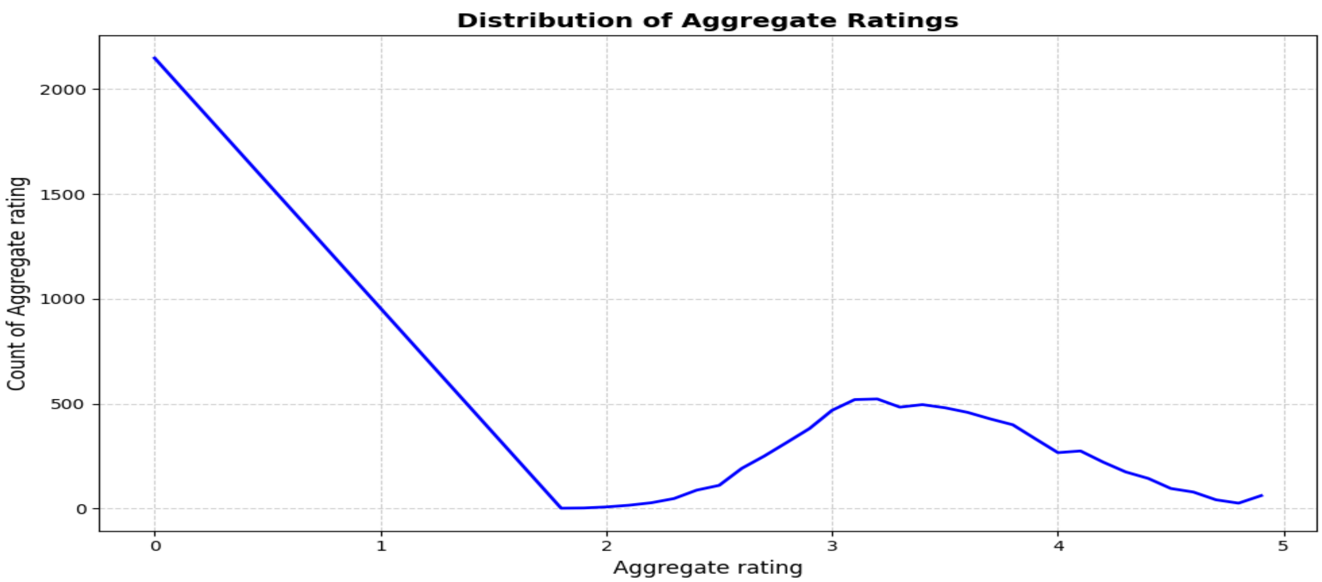
Code Explanation:

- The code calculates the frequency of each aggregate rating by counting how often each rating appears in the 'Aggregate rating' column and sorting these counts by rating value.
- It visualizes the counted ratings as a blue line chart, with aggregate ratings on the x-axis and their corresponding counts on the y-axis, customizing the plot’s style, color, and line width.
- The code then adds a descriptive title, axis labels, grid lines, and adjusts the plot layout for clarity before displaying the final chart.

```python
# Assuming the column is named 'Aggregate rating'
rating_counts = df['Aggregate rating'].value_counts().sort_index()

# Plot the line chart
plt.figure(figsize=(10, 6))
plt.plot(rating_counts.index, rating_counts.values, color='blue', linewidth=2)

# Add labels and title
plt.title('Distribution of Aggregate Ratings', fontsize=14, fontweight='bold')
plt.xlabel('Aggregate rating', fontsize=12)
plt.ylabel('Count of Aggregate rating', fontsize=12)

# Add grid and style
plt.grid(True, linestyle='--', alpha=0.6)
plt.tight_layout()

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

Results:

- The distribution of aggregate ratings is highly right-skewed with the largest number of restaurants (2,146) receiving a rating of 0.
- Rating frequency drops sharply after 0, with a trough around the 2.0–2.5 mark, and then shows a moderate increase for ratings between 3.0 and 4.0, where the majority of rated restaurants cluster.
- The most common rating value is 0 (unrated or no feedback), followed by moderate frequencies in the mid-rating ranges.
- The average number of votes per restaurant is 156.99, indicating a moderate level of customer engagement and feedback collection across establishments.



Visualization:

The line graph titled "Distribution of Aggregate Ratings" illustrates:

- A peak at the 0-rating mark, indicating many restaurants are either unrated or have not received customer reviews.

- A steep decline moving toward ratings of 1 and 2, with counts rising again moderately in the 3–4 range, before tapering off toward the maximum rating of 5.

- The y-axis represents the number of restaurants at each aggregate rating value, and the x-axis shows the rating scale from 0–5.

Business Implications:

- Customer Engagement Gap: The prevalence of restaurants with an aggregate rating of 0 exposes a significant gap in customer feedback, signaling a need for proactive review and engagement strategies.

- Reputational Opportunity: Restaurants in the 3–4 rating range are performing well and should leverage their reputation to attract more customers, whereas those unrated risk being overlooked.

- Action on Reviews: Upskilling staff, encouraging customer reviews, and soliciting feedback through loyalty programs or digital means can improve rating counts and credibility.

- Competitive Benchmarking: With an average vote count of 156.99, establishments can gauge their engagement level against market norms, aiming to exceed this benchmark for stronger online presence and trust.

- Establishments with no ratings should prioritize strategies to collect feedback, as a credible rating boosts customer confidence and positively influences dining choices.

Task 2 – Cuisine Combination

Objective:

To analyze the restaurant distribution by cuisine type and identify which cuisines are most commonly served, helping inform market trends, diversification strategies, and consumer preferences across the market.

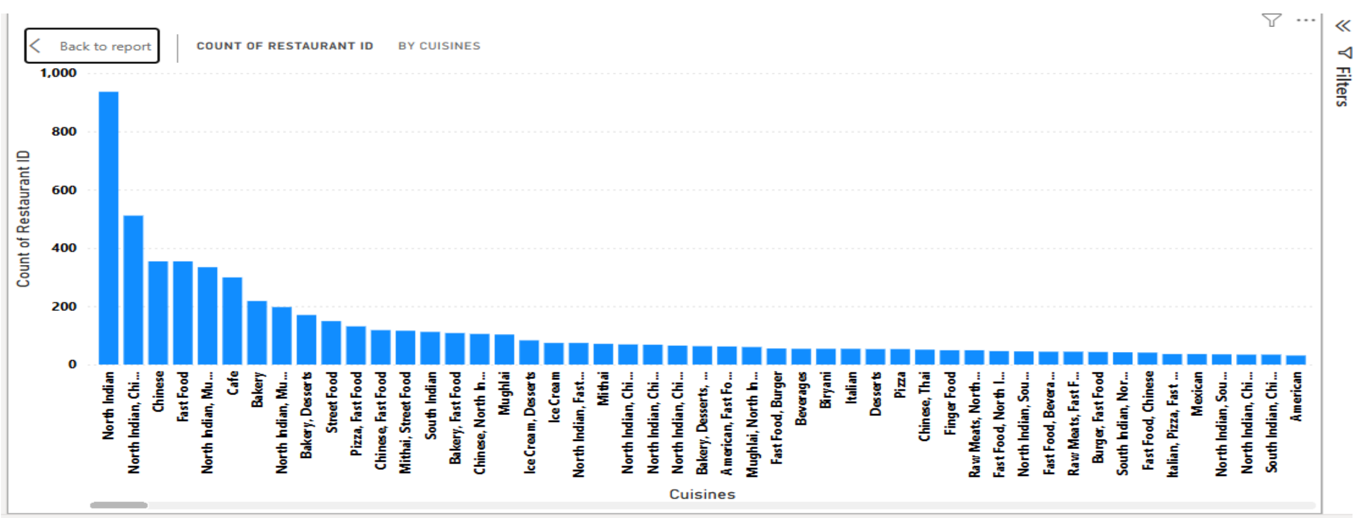
Code Explanation:

- The code groups the dataframe by 'Cuisines' and counts the number of unique restaurant IDs for each cuisine, then sorts these counts in descending order to identify cuisines with the most restaurants.
- It creates a bar chart with the cuisines on the x-axis and the count of unique restaurant IDs on the y-axis, using a blue color scheme and an extra-wide figure size for better readability.
- The plot includes clear axis labels, a descriptive title, and uses tight layout adjustment to prevent label overlap before displaying the visualized result.

```python
# Group by cuisine and count unique restaurant IDs
cuisine_counts = df_exploded.groupby('Cuisines')['Restaurant ID'].nunique().sort_values(ascending=False)

# Plot the bar chart
plt.figure(figsize=(18, 6))
cuisine_counts.plot(kind='bar', color='dodgerblue')
plt.xlabel('Cuisines')
plt.ylabel('Count of Restaurant ID')
plt.title('Count of Restaurant ID by Cuisines')
plt.tight_layout()
plt.show()
```

Results:

- North Indian cuisine is the clear leader, with the highest restaurant count by a significant margin.
- Following this are North Indian, Chinese fusion, Chinese, Fast Food, and Cafe, all with substantial representation.
- There is a long tail of less common cuisines—typified by more specialized or international offerings—showing some degree of menu diversification but much lower occurrence.
- The distribution is highly right-skewed; only a handful of cuisines dominate the restaurant landscape, while most others are niche options.



Visualization:

The bar chart "Count of Restaurant ID by Cuisines":

- Shows counts for each cuisine (x-axis: Cuisines; y-axis: Number of Restaurants).
- Bars for North Indian, North Indian & Chinese, and Chinese are much taller than others.
- Provides an immediate comparative overview, with declining counts moving toward specialty cuisines at the chart's right side.

Business Implications:

- Menu Planning: Restaurants should prioritize offering North Indian, Chinese, and fusion menus to align with leading consumer demand and increase competitive viability.
- Differentiation: Potential exists for niche cuisines (Italian, Continental, Mexican, etc.) to stand out in markets oversaturated with mainstream cuisine types.
- Expansion: Businesses seeking rapid growth should leverage popular cuisine types, while innovators and specialty brands can capture loyal segments by exploring lesser-served cuisines.
- Marketing Focus: Campaigns centered around the most common cuisines have the widest reach, while focused storytelling around niche options can cultivate dedicated customer bases and premium positioning

Task 3 – Geographic Analysis

Objective:

To plot restaurant locations on a geographic map using longitude and latitude data and analyze spatial distribution patterns, identifying clusters or areas of concentration.

Code Explanation:

- The dataset groups restaurant data by latitude, longitude, and name, calculating total votes for each unique restaurant.
- Using Plotly’s scatter_geo, this data is visualized on an interactive world map with bubbles representing vote counts; hovering reveals the restaurant name.
- The map uses the 'natural earth' projection for realism, titled “Sum of Votes and First Restaurant Name by Latitude and Longitude,” allowing users to explore restaurant clusters by engagement and location.

```python
# Group by latitude, longitude, and restaurant name, summing votes
map_data = df.groupby(['Latitude', 'Longitude', 'Restaurant Name'], as_index=False)['Votes'].sum()

# Plot using Plotly for interactive bubbles
fig = px.scatter_geo(
    map_data,
    lat='Latitude',
    lon='Longitude',
    size='Votes',
    hover_name='Restaurant Name',
    projection='natural earth',
    title='Sum of Votes and First Restaurant Name by Latitude and Longitude'
)

fig.show()
```

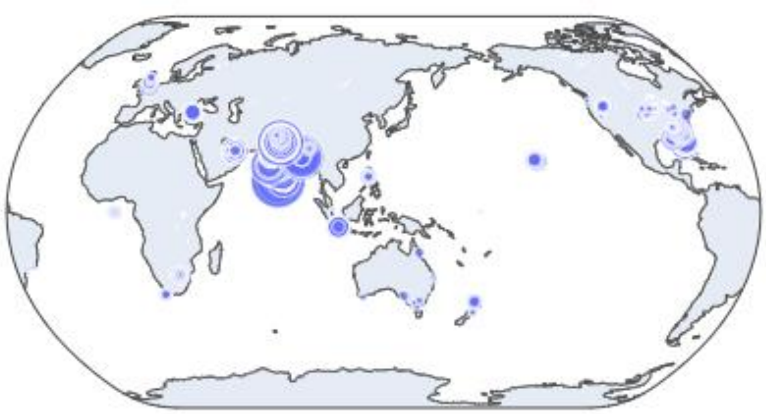
Results:

- The plotted map reveals distinct clusters of restaurant locations, with the largest gatherings in high-population or culinary destination regions—especially in South Asia, Southeast Asia, and major cities in Europe, North America, and Australia.

- Several intense clusters are visible in India and surrounding regions, indicating local hotspots for dining activity.

- Additional clusters appear in Europe and North America, while other continents show sparser distributions of restaurants.

- The bubble sizes indicate the density or engagement (such as number of restaurants or customer votes) at each location, with denser regions usually receiving more feedback.



Visualization:

- The chart displays restaurant locations as bubbles on a world map.
- Larger bubbles correspond to higher counts (restaurant density or collective engagement).
- The geographic spread emphasizes the global reach of the dataset, while visually highlighting urban centers and popular regions.

Business Implications:

- Expansion Planning: Identified clusters suggest strategic areas for new restaurant or outlet launches, leveraging existing demand and familiarity.
- Regional Marketing: Densely populated clusters benefit from tailored local promotions to stand out in competitive markets.
- Operational Efficiency: Cluster analysis enables more efficient inventory delivery, staffing, and logistics due to proximity among outlets.
- Opportunity Identification: Sparse regions present potential for new ventures aiming for early market entry and leadership.

Task 4 – Restaurant Chains

Objective:

- To evaluate the performance of top restaurant chains by comparing the number of operational outlets and their corresponding average customer ratings, thus assessing the relationship between brand presence and service quality.

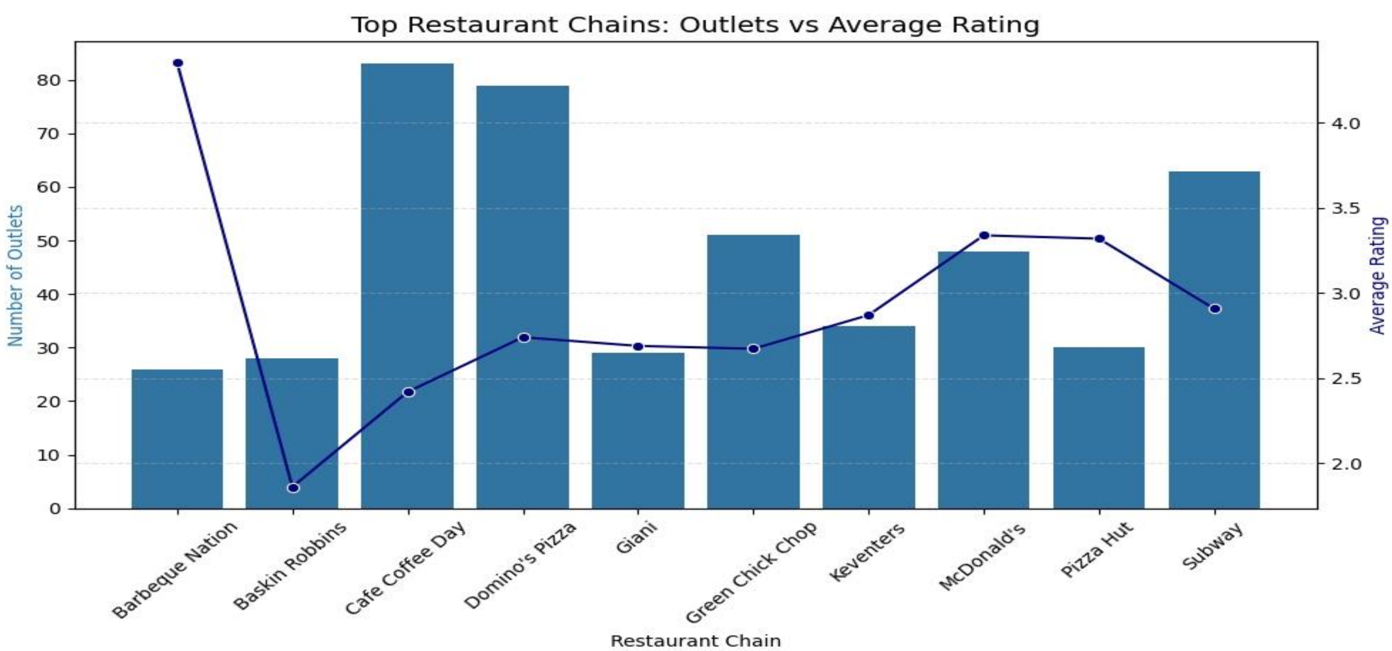
Code Explanation :

- Counts each restaurant name and selects chains with more than one outlet, focusing on the top 10.
- Calculates the average rating for each chain and merges outlet counts with ratings for a summary table.
- Plots a dual-axis chart: outlet count as bars and average rating as a line for clear comparison.

```python
# Visualization: Dual-axis Bar + Line Chart (Blue)
fig, ax1 = plt.subplots(figsize=(10,6))

# Bars: Number of outlets
sns.barplot(x='Restaurant Name', y='Outlet Count', data=chain_df, color='#1f77b4', ax=ax1)
ax1.set_ylabel('Number of Outlets', color='#1f77b4')
ax1.set_xlabel('Restaurant Chain')
ax1.tick_params(axis='x', rotation=45)

# Line: Average rating
ax2 = ax1.twinx()
sns.lineplot(x='Restaurant Name', y='Aggregate rating', data=chain_df, color='navy', marker='o', ax=ax2)
ax2.set_ylabel('Average Rating', color='navy')

plt.title("Top Restaurant Chains: Outlets vs Average Rating", fontsize=14)
plt.grid(axis='y', linestyle='--', alpha=0.4)
plt.tight_layout()
plt.show()
```

Results:

- Cafe Coffee Day and Domino's Pizza lead in the number of outlets, each operating more than 80 locations.
- Subway, McDonald's, and Green Chick Chop also maintain a strong multi-outlet presence.
- Barbeque Nation has a smaller outlet footprint but stands out due to its exceptional average rating (above 4).
- Most other chains exhibit average ratings between 2.0 and 3.5, with some (e.g., Baskin Robbins) notably lower despite a moderate presence.



Visualization:

- The dual-axis chart "Top Restaurant Chains: Outlets vs Average Rating":
- The blue bars represent the number of outlets for each chain (left y-axis), highlighting market reach.

- The navy line shows average customer ratings (right y-axis), offering quick visibility into perceived quality.
- This combination reveals chains that balance expansion with consistent customer satisfaction and those with varying ratings despite high outlet counts.

Business Implications:

- Scalability vs. Quality: Chains with many outlets (like Cafe Coffee Day or Domino's) must focus on maintaining rating standards across branches, as wide reach alone does not guarantee high ratings.
- Premium Brands: Restaurants like Barbeque Nation can leverage their strong reputation for further expansion or premium positioning, especially in cities where reputation drives customer choice.
- Strategic Benchmarking: Chains with high ratings and broad presence provide models for service consistency, operational excellence, and brand equity.
- Improvement Focus: Underperforming chains—even with many outlets—should invest in service training, menu quality, and customer engagement to lift ratings and overall brand.

Conclusion:

The Level 2 Data Analysis Report delivers powerful insights into how customer ratings, cuisine trends, regional clusters, and chain performance shape the restaurant market. Results highlight the dominance of mainstream cuisines like North Indian and Chinese, significant geographic clustering in high-population and culinary destination areas, and a clear gap in customer engagement for many unrated establishments. Analysis of leading chains shows that while some brands combine scale and quality, others struggle to maintain high ratings as they expand. These findings empower businesses to refine their menu strategies, target high-potential geographic zones, prioritize feedback collection, and balance growth with consistent service quality for long-term competitive advantage.