- *Restaurant Data Analysis Project*

- *Internship Project @ Cognifyz Technologies*
- *Tool Used : Python, Excel, MySQL*

- *Duration : 17/03/2025 to 17/04/2025*

SHYAM DHUDIYA
From : Dhoraji, Gujarat

- ## *Introduction*

- *This project is a part of my internship at Cognifz technologyies, where i was assigned a real-word dataset containing information about various restaurants.*

- *The main goal of this project is to extract meaningful insights from the data using tools such as Python, Excel, MySQL.*

- *Dataset Overview*

- ***The dataset contains detailed information about restaurants such as their names, country code, city, cuisines, locations, delivery, booking table, currency, rating color, rating text, votes***

- ***It consists of more than 9550 rows and 21 columns. This data was analyzed using Python to uncover useful business insights.***

- *Python Analysis*

- *Python was used for in-depth data analysis and visualization, Libraries such as Pandas, Numpy, Matplotlib, and Seaborn were used to clean, analyze, and visualize key trends in the dataset.*

- *Imported and cleaned daset using Pandas*
- *Handling missing values and removed outliers*
- *Used groupby() and filtering for insights*
- *Created visualizations with Matplotlib & Seaborn*

[1]:
```
# import necessary libraries
import pandas as pd
import matplotlib.pyplot as plt
import folium
import nltk
import string
import seaborn as sns
from nltk.tokenize import word_tokenize
from collections import Counter
from nltk.corpus import stopwords
from nltk.sentiment import SentimentIntensityAnalyzer
from wordcloud import WordCloud
```

[2]:
```
# Load data
file_path = "C:/Users/dhudi/Downloads/Restaurants.xlsx"
df = pd.read_excel(file_path)
```

[3]:
```
df.shape
```

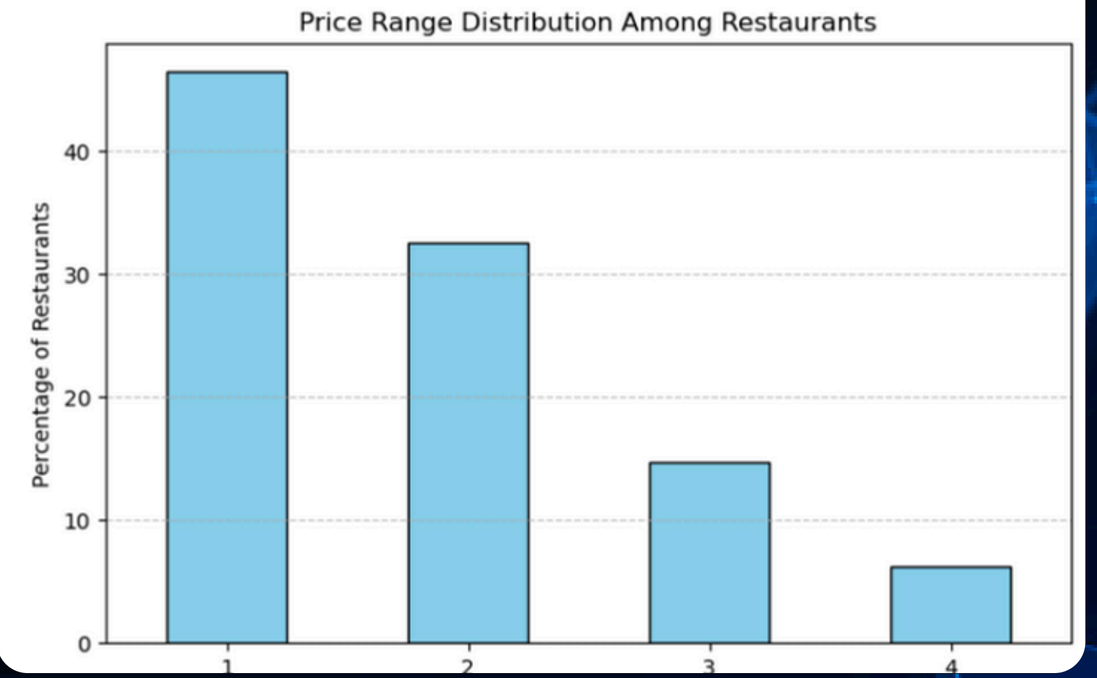
[3]:
```
(9551, 21)
```

[4]:
```
df.columns
```

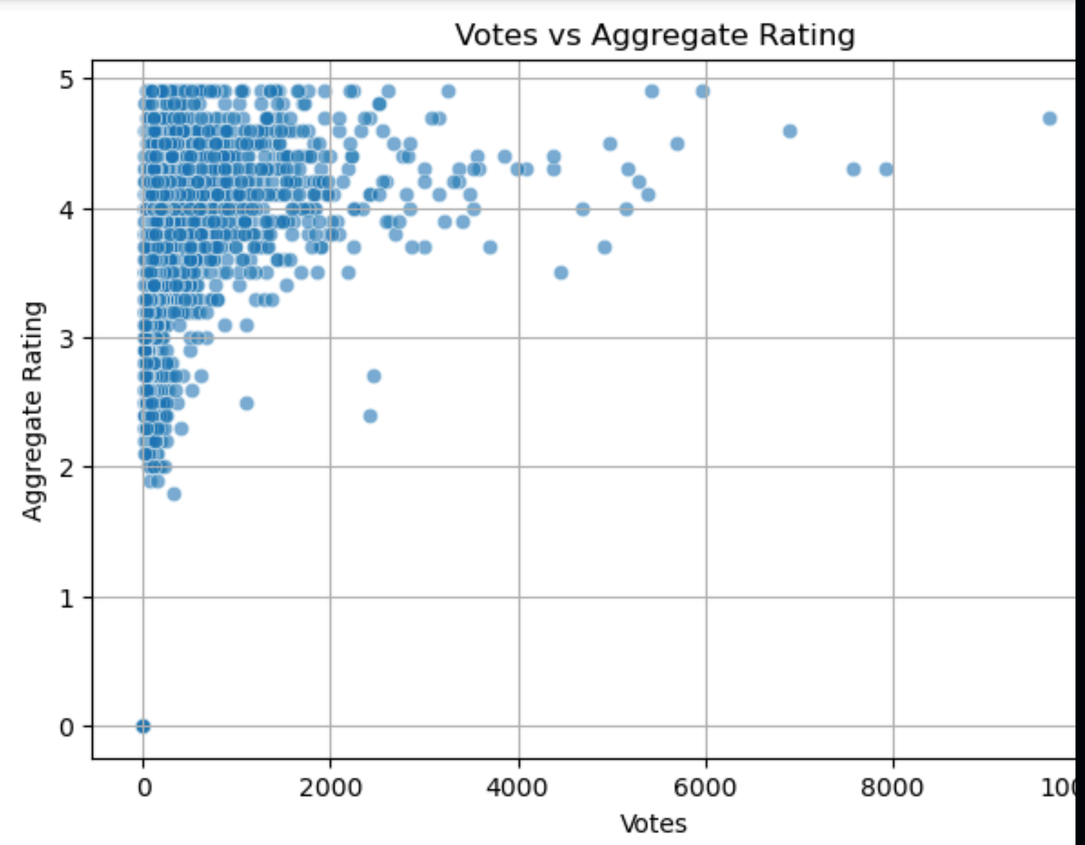
[4]:
```
Index(['Restaurant ID', 'Restaurant Name', 'Country Code', 'City', 'A
       'Locality', 'Locality Verbose', 'Longitude', 'Latitude', 'Cuis
       'Average Cost for two', 'Currency', 'Has Table booking',
       'Has Online delivery', 'Is delivering now', 'Switch to order m
       'Price range', 'Aggregate rating', 'Rating color', 'Rating tex
       'Votes'],
      dtype='object')
```

Task 3 : Price Range Distribution

```
price_range_counts = df["Price range"].value_counts(normalize=True) * 100

plt.figure(figsize=(8,5))
price_range_counts.sort_index().plot(kind='bar', color='skyblue', edgecolor='black')

plt.xlabel("Price Range")
plt.ylabel("Percentage of Restaurants")
plt.title("Price Range Distribution Among Restaurants")
plt.xticks(rotation=0)
plt.grid(axis='y', linestyle='--', alpha=0.7)

plt.show()

price_range_counts
```



```
plt.figure(figsize=(8, 5))
sns.scatterplot(data=df_clean, x='Votes', y='Aggregate rating', alpha=0.6)
plt.title('Votes vs Aggregate Rating')
plt.xlabel('Votes')
plt.ylabel('Aggregate Rating')
plt.grid(True)
plt.show()
```

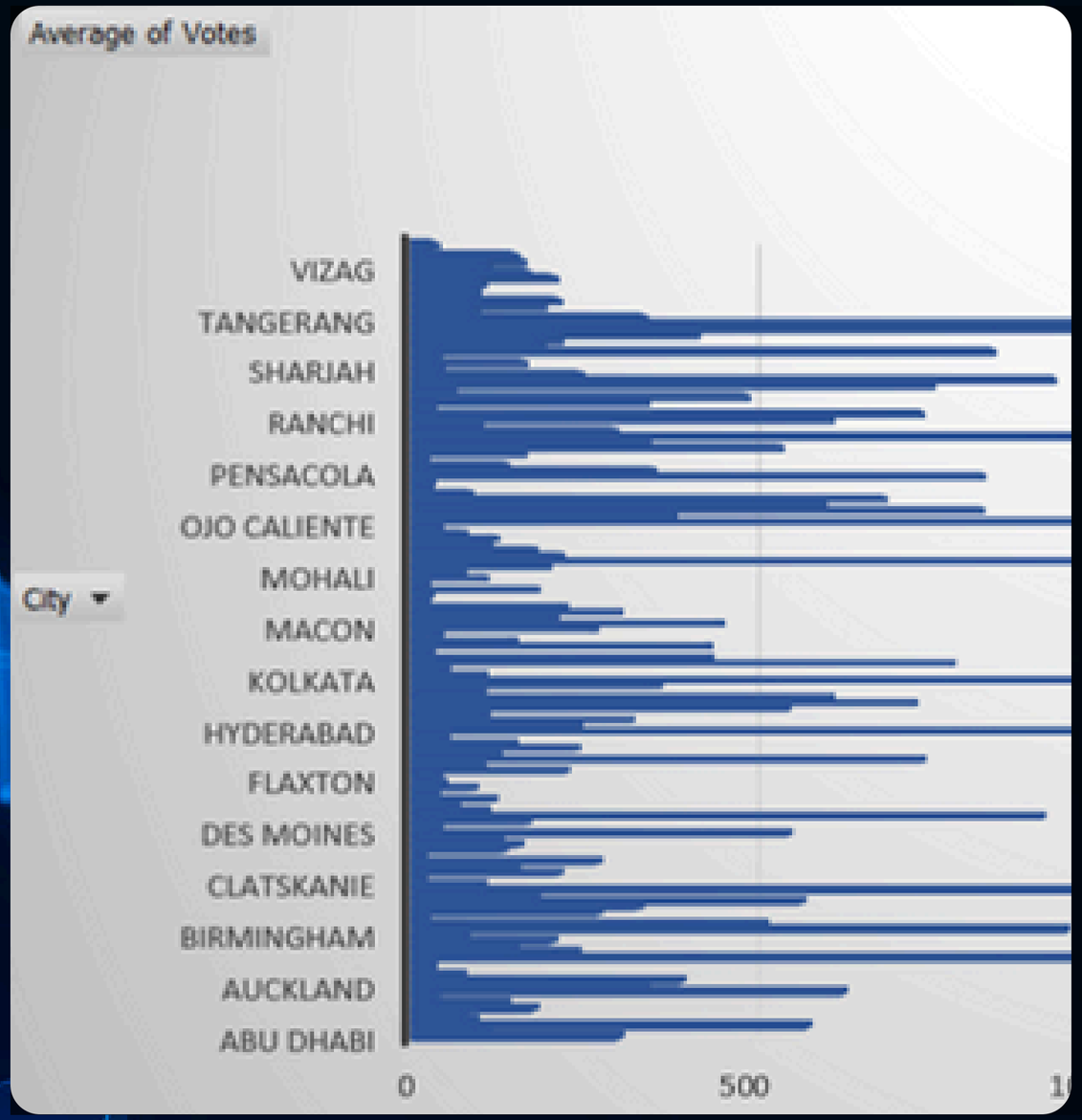


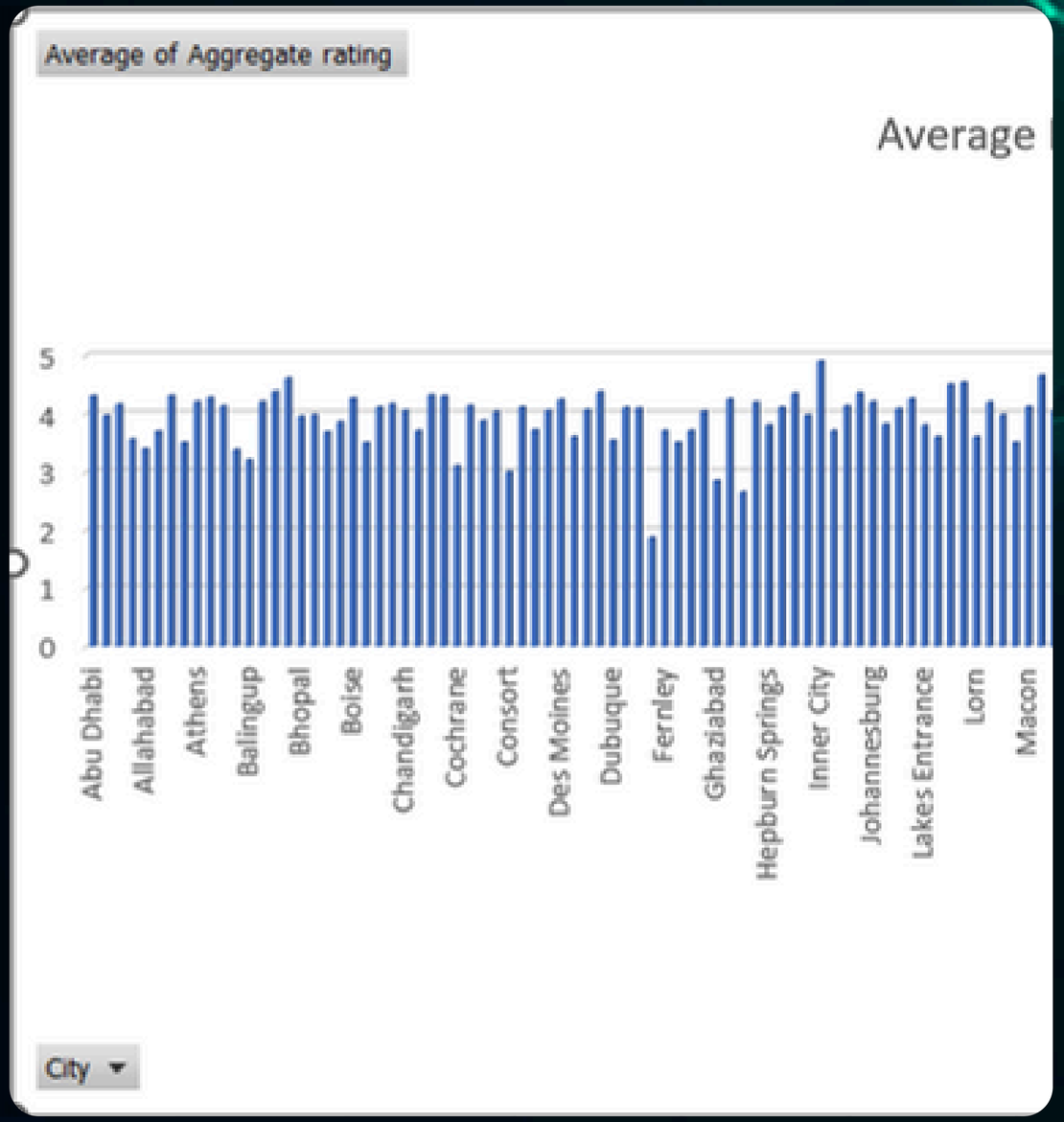

```
# Table booking
booking_vs_price.plot(kind='bar', stacked=True, colormap='Pastel1')
plt.title('Table Booking Availability by Price Range')
plt.ylabel('Percentage')
plt.xlabel('Price Range')
plt.legend(title='Table Booking')
plt.grid(True)
plt.show()
```

- ***Excel Analysis***

- *The initial phase of analysis was done using Microsoft Excel. The dataset was cleaned, filtered, and visualized to understand patterns in restaurants rating, votes, cuisines.*

- *Excel allowed for quick exploration and visual representation of the data through pivot tables and charts*
- *Cleaned and removed duplicates*
- *Created Pivot Table to Summarize*
- *Created Charts*
- *Applied Conditional Formatting to Highlight*


Average of Votes
VIZAG
TANGERANG
SHARJAH
RANCHI
PENSACOLA
OJO CALIENTE
MOHALI
MACON
KOLKATA
HYDERABAD
FLAXTON
DES MOINES
CLATSKANIE
BIRMINGHAM
AUCKLAND
ABU DHABI
City
0
500


Average of Aggregate rating
Average
5
4
3
2
1
0
Abu Dhabi
Allahabad
Athens
Balingup
Bhopal
Boise
Chandigarh
Cochrane
Consort
Des Moines
Dubuque
Fernley
Ghaziabad
Hepburn Springs
Inner City
Johannesburg
Lakes Entrance
Lorn
Macon
City


Cuisines Distribution by Restaurants


Count of Aggregate rating
PRICE RANG
4444
3113
1
2
Price range

- *MySQL Analysis*

- *In this phase, the cleaned restaurant dataset was uploaded into MySQL for structured querying and deeper analysis. Using SQL queries, I extracted valuable insights such as:*
- *Top 10 Cities with Most Restaurants*
- *Average Price Range by Cuisine*
- *Count Restaurants Based on Rating Text*

[1]:
```
# import necessary libraries
import pandas as pd
import matplotlib.pyplot as plt
import folium
import nltk
import string
import seaborn as sns
from nltk.tokenize import word_tokenize
from collections import Counter
from nltk.corpus import stopwords
from nltk.sentiment import SentimentIntensityAnalyzer
from wordcloud import WordCloud
```

[2]:
```
# Load data
file_path = "C:/Users/dhudi/Downloads/Restaurants.xlsx"
df = pd.read_excel(file_path)
```

[3]:
```
df.shape
```

[3]:
```
(9551, 21)
```

[4]:
```
df.columns
```

[4]:
```
Index(['Restaurant ID', 'Restaurant Name', 'Country Code', 'City', 'A
       'Locality', 'Locality Verbose', 'Longitude', 'Latitude', 'Cuis
       'Average Cost for two', 'Currency', 'Has Table booking',
       'Has Online delivery', 'Is delivering now', 'Switch to order m
       'Price range', 'Aggregate rating', 'Rating color', 'Rating tex
       'Votes'],
      dtype='object')
```

```
9       -- 5. Count Restaurants Based on
0 •     SELECT Rating_text, COUNT(*) AS c
1       FROM restaurant
2       GROUP BY Rating_text;
3
```

sult Grid | Filter Rows: | E

Rating_text	count
Excellent	301
Very Good	1079
Good	2100
Average	3737
Not rated	2148
Poor	186

```
8       -- 3. Average Price Range by Cuisine
9 •     SELECT Cuisines, AVG(Price_range) AS avg_price
0       FROM restaurant
1       GROUP BY Cuisines
2       ORDER BY avg_price DESC;
3
4       -- 4. Restaurants in a Specific City
```

sult Grid | Filter Rows: | Export: | Wrap

Cuisines	avg_price
Seafood, Asian, Filipino, Indian	4.0000
Seafood, Filipino, Asian, European	4.0000
European, Asian, Indian	4.0000
Cafe, Bakery, American, Italian	4.0000
Seafood, American, Mediterranean, Japanese	4.0000
American, Asian, Italian, Seafood	4.0000
Mexican, Grill	4.0000
Peruvian, Latin American	4.0000
American, Grill	4.0000
Seafood, Bar Food, Brazilian	4.0000

```sql
-- 1. Top 10 Cities with Most Restaurants
SELECT City, COUNT(*) AS total_restaurants
FROM restaurant
GROUP BY City
ORDER BY total_restaurants DESC
LIMIT 10;

-- 2. Top 10 Highest Rated Restaurants
```

Result Grid | Filter Rows: | Export: | Wrap C

City	total_restaurants
New Delhi	5473
Gurgaon	1118
Noida	1080
Faridabad	251
Ghaziabad	25
Guwahati	21
Ahmedabad	21
Lucknow	21
Amritsar	21
Bhubaneshwar	21

- ***Key Insights & Recommendations***

- *After analyzing the restaurants dataset using Python, Excel, and SQL several valuable insights were identified. These insights can help improve customers satisfaction, business growth, and operational efficiency.*

- *Conclusion*

  - *This internship project provided hands-on experience in analyzing real-word restaurants data using various tools such as Python, Excel, MySQL*

  - *Strengthened skill in data cleaning, EDA, and visualization*
  - *Learned to device actionable insights using Python, SQL, Excel*

THANK YOU!

FOR YOUR ATTENTION

linkedin.com/in/shyam-dhudiya

https://github.com/Shyamdhudiya

dhudiyashyam84@gmail.com

91049 41922