📄 Cognifyz Internship Level 1&2 – Data Analysis Report

Title: Restaurant Data Insights Using Power BI & Jupyter Notebook

Prepared by: Challagundla Pavani Prudhvila

Tools Used: Power BI, Python (Jupyter Notebook), Pandas, Matplotlib, Seaborn

🧩 Overview

This report explores restaurant data to uncover insights about customer ratings, cuisine combinations, geographic patterns, and restaurant chains. Two tools were used:

📊 Power BI Dashboard

✅ Level 1 – Basic Restaurant Analysis Report

🔹 Task 1: Cuisine Analysis

📌 Objective

Identify the top 3 most common cuisines and calculate their share as a percentage of the total restaurants.

🧪 Methodology

Used value\_counts() to count occurrences in the Cuisines column.

Selected top 3 cuisines , nlargest(3).

Calculated percentage share of each using simple division.

Visualized the results using a pie chart.

📈 Visualization

A pie chart was used to display the top 3 cuisines and their proportions, styled with labels and optional emojis for clarity.

✅ Insights

Most common cuisine: North Indian

Followed by: Chinese and South Indian

These 3 account for a majority of restaurant listings, showing customer demand trends.

🔹 Task 2: City-wise Restaurant Analysis

📌 Objective

Identify which cities have the most restaurants and which cities offer the highest average restaurant ratings.

🧪 Methodology

Counted number of restaurants by City using value\_counts()

Grouped data by city and calculated average rating using groupby('City')['Aggregate rating'].mean()

Visualized city-wise average ratings with a bar chart of the top 10 cities.

📈 Visualization

Bar chart with cities on the X-axis and average ratings on the Y-axis. Gridlines and titles added for clarity.

✅ Insights

Highest average rating: Jammu (example)

Popular cities don’t always mean higher-rated ones — smaller cities can outperform on quality

🔹 Task 3: Price Range Distribution

📌 Objective

Understand how restaurants are distributed across different pricing categories.

🧪 Methodology

Used value\_counts() on the Price range column to count restaurants by price level.

Calculated percentage for each range.

Visualized using a bar chart and optional donut chart.

📈 Visualization

Bar chart: Price range vs number of restaurants

Donut chart: Price range vs percentage distribution

✅ Insights

Most restaurants fall into Price Range 2 (moderately priced)

Very few offer luxury pricing (Range 4), indicating affordability dominance in the dataset

📦 Summary of Level 1

Task Focus Area Key Output

1 Cuisine Top 3 cuisines, pie chart

2 City analysis Most/Best cities, bar chart

3 Price distribution % per price range, bar/donut chart

📊 Power BI Dashboard

✅ Task 1: Restaurant Ratings

Goal: Analyze rating distribution, find most common range, calculate average votes

Visuals:

Bar Chart showing count of restaurants by Aggregate Rating

Card showing average number of Votes

Card displaying most common rating value

DAX Used:

Average Votes = AVERAGE('food'[Votes])

Rating Range =

SWITCH(TRUE(),

'food'[Aggregate rating] < 2, "Below 2.0",

'food'[Aggregate rating] < 3, "2.0 - 2.9",

'food'[Aggregate rating] < 4, "3.0 - 3.9",

'food'[Aggregate rating] <= 5, "4.0 - 5.0",

"Unknown"

)

✅ Task 2: Cuisine Combination

Goal: Identify popular cuisine combos and check if they get higher ratings

Visuals:

Bar Chart showing top cuisine combinations

Table showing cuisine vs average rating

Insights:

North Indian + Chinese is one of the most frequent combinations

Some fusion cuisines show above-average ratings

✅ Task 3: Geographic Patterns (Without Maps)

Goal: Understand city-level trends without using maps

Visuals:

Bar Chart: Top 15 cities by number of restaurants

Matrix/Table: City-wise average ratings and total votes

Slicer added for filtering by city

🧪 Jupyter Notebook Analysis (Python)

📌 Steps Performed:

1. Import libraries: pandas, matplotlib, seaborn

2. Read dataset and check for nulls

3. Exploratory Analysis

Value counts of Cuisines, City, Restaurant Name

Grouped bar charts (Cuisine vs Rating, City vs Votes)

4. Cuisine Combination Analysis

df['Cuisines'].value\_counts().head(10)

df.groupby('Cuisines')['Aggregate rating'].mean().sort\_values(ascending=False)