

# **📊 Zomato Bangalore Restaurant Trends**

## **🎯 Objective**

The aim of this hackathon is to perform **data-driven analysis of restaurant trends** in **Bangalore** using Zomato’s real-world dataset. Participants are expected to:

* Clean and preprocess the data
* Perform exploratory data analysis (EDA)
* Merge location data for mapping
* Extract actionable insights
* Answer MCQ-style reasoning questions based on analysis

## **🧑‍💻 Environment Setup**

Participants must use **Jupyter Notebook** for this hackathon. Ensure the following Python libraries are installed:

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from geopy.geocoders import Nominatim

from geopy.exc import GeocoderTimedOut

from IPython.core.display import display, HTML

import time

Additional libraries such as folium, plotly, or geopandas may be used for mapping if preferred.



## **📦 Datasets Provided**

### **1. data1.csv –** [**Zomato Restaurant Data**](https://drive.google.com/file/d/1IfdLKZJR_ETJUa3BQYM11j0GSEVt6YcT/view?usp=sharing)

| **Column Name** | **Description** |
| --- | --- |
| online\_order | Indicates if online ordering is available (Yes/No) |
| book\_table | Indicates if table booking is available (Yes/No) |
| rate | Average rating (e.g., 4.1/5) |
| votes | Number of votes received by the restaurant |
| rest\_type | Type of restaurant (e.g., Casual Dining, Quick Bites) |
| dish\_liked | Popular dishes liked by customers |
| cuisines | Cuisines offered by the restaurant |
| approx\_costfor\_two\_people | Approximate cost for two people in INR |
| listed\_intype | Restaurant listing category (e.g., Buffet, Delivery) |
| listed\_incity | Local area in Bangalore where the restaurant is located |

### **2. data2.csv –** [**Geographical Coordinates**](https://drive.google.com/file/d/1xlJ--IT_vMXlUXcvBvHkvTuyBSEJ8mhN/view?usp=sharing)

| **Column Name** | **Description** |
| --- | --- |
| listed\_incity | Local area name (used for merging) |
| Latitude | Latitude of the location |
| Longitude | Longitude of the location |



## **Data Cleaning & Preprocessing**

Participants are required to clean and preprocess the data using the following guidelines:

### **Step 1: Rating Column (rate)**

* Replace '-' values with NaN
* Remove /5 and retain only the numeric value
* Convert to float using pd.to\_numeric()
* Fill missing values using the **median rating**

### **Step 2: Cost Column (approx\_costfor\_two\_people)**

* Remove commas from numeric strings ('1,000' → '1000')
* Convert to **numeric**
* Fill missing values using the **median cost**

### **Step 3: Categorical Columns**

* dish\_liked → Replace NaN with "Not Available"
* cuisines → Replace NaN with "Other"
* rest\_type → Replace NaN with "Unknown"

### 



### **Step 4: Votes Column**

* Fill missing values in votes with **median**

### **Step 5: Binary Encoding**

Convert the following binary fields:

| **Column** | **Mapping** |
| --- | --- |
| online\_order | Yes → 1, No → 0 |
| book\_table | Yes → 1, No → 0 |

### **Step 6: Data Type Conversion**

Ensure the following conversions:

* rate → float
* votes → integer
* approx\_costfor\_two\_people → integer

### **📝 Note**

## Ensure that after every transformation, you verify the changes using functions like .info(), .isnull().sum(), and .describe() to confirm successful cleaning.



## **🔗 Dataset Merging for Mapping**

### **1. data1.csv –** [**Zomato Restaurant Data**](https://drive.google.com/file/d/1IfdLKZJR_ETJUa3BQYM11j0GSEVt6YcT/view?usp=sharing)

### **2. data2.csv –** [**Geographical Coordinates**](https://drive.google.com/file/d/1xlJ--IT_vMXlUXcvBvHkvTuyBSEJ8mhN/view?usp=sharing)

To incorporate location-based analysis, merge the datasets using the listed\_incity column:

*merged\_df = pd.merge(data1\_cleaned, data2, on='listed\_incity', how='left')*

This will provide Latitude and Longitude values for each restaurant, enabling geographical visualizations.

## **🗺️ Geospatial Visualization Task**

## **Objective**

The **Geospatial Visualization Task** aims to help participants perform location-based analysis using a restaurant dataset. This task focuses on visualizing restaurant density and creating cuisine-specific maps using **Folium**—a Python library for generating interactive maps. The goal is to provide insights into the spatial distribution of restaurants across **Bangalore** and identify patterns based on **restaurant type**, **ratings**, and **location**.

## **Task Overview**

### **Restaurant Density Map**

Participants will create an interactive map to visualize the density of restaurants across different areas of Bangalore. This map will highlight the geographic distribution of restaurants and identify regions with higher concentrations of eateries.

### 

### **Cuisine-Specific Map (Italian Restaurants)**

In this task, participants will filter the dataset to focus on **Italian cuisine** and create a map that visualizes the locations of these restaurants across the city. This task allows participants to explore the popularity of Italian restaurants in various neighborhoods.

## 

## **Submission Guidelines**

Each team must submit:

1. 📓 A **Jupyter Notebook (.ipynb)** that includes:  
   * All data cleaning steps
   * Merged dataset
   * Visualizations and EDA findings
2. 📄 A set of answers to **MCQ-based questions** (provided separately or generated from the notebook insights)

### **Conclusion**

This hackathon provides an engaging opportunity for participants to analyze and visualize restaurant data through **geospatial techniques**. By creating interactive maps, participants will gain valuable insights into the **restaurant industry** in Bangalore, with a focus on **restaurant density** and **cuisine-specific trends**.