Advanced Explorations Using Python

Problem Statement

Problem Statement In a fast-evolving food culture, diners often struggle to identify quality restaurants that offer the best value for money. This project aims to help food lovers explore and analyze real-world restaurant data to discover the best-rated, most cost-effective, and cuisine-specific dining options in their locality. Using exploratory data analysis (EDA), we will uncover patterns in restaurant offerings, pricing, ratings, service availability, and geographical distribution. The end goal is to empower users with insights that can drive smarter food decisions based on factual data rather than personal recommendations.

Tasks to be completed

1. Import Libraries
2. Set Options
3. Read Data
4. Understand and Prepare the Data
5. Understand the variables
6. Check for Missing Values
7. Study Correlation
8. Detect Outliers
9. Create a new variable 'region'

Learning Objectives

Upon completing this project, learners will be able to:

- Understand the fundamentals of structured EDA using pandas, matplotlib, and seaborn
- Identify and handle missing or inconsistent data
- Perform feature engineering (e.g., creating a custom region column)
- Explore correlations between variables like cost, rating, and popularity
- Use Python to generate visual insights from real-world data

Dataset Description

The dataset represents information about dining outlets, including geolocation, cost metrics, cuisines, services, and user feedback.

Variable Name	Description	Type
name	Official name of the food outlet	Categorical (object)
establishment	Category of the eatery (e.g., Café, Fine Dining)	Categorical (object)
city	City in which the outlet operates	Categorical (object)
locality	Neighborhood within the city	Categorical (object)
latitude	Geographical latitude of the outlet	Numerical (float64)
longitude	Geographical longitude of the outlet	Numerical (float64)
cuisines	List of cuisine types offered	Categorical (object)
average_cost_for_two	Estimated cost for two people	Numerical (int64)
price_range	Tiered cost bracket (1 = low, 4 = premium)	Categorical (object) (May appear as int)
highlights	Notable features (e.g., Live Music, Outdoor Seating)	Categorical (object)
aggregate_rating	Overall rating based on user reviews	Numerical (float64)

rating_text	Textual label (e.g., Excellent, Good)	Categorical (object)
votes	Total user feedback entries	Numerical (int64)
photo_count	Number of images uploaded	Numerical (int64)
delivery	Indicates if delivery is offered	Categorical (object) (May appear as int)