Task 6.1 Bangalore Zomato Data Set Analysis

DATA

- **Source of Data:** The data is available on one of open source platform – kaggle.com and downloaded from following link: https://www.kaggle.com/datasets/himanshupoddar/zomato-bangalore-restaurants?resource=download

- **Data Collection:** The data was scraped from Zomato website and collected in two phases. In Phase I, the URLs, names, and addresses of restaurants visible on the front page of Zomato were extracted and recorded in a CSV file, simplifying the later individual data extraction process. In Phase II, data for each restaurant and category was scraped individually.

- **Data Content**: obtaining 15 variables such as online order, book_table, rate, votes, phone, location, rest_type, dish_liked, cuisines, approx_cost (for two people), reviews_list, and menu_item.

- **Data Profile:**

1. The dataset contains 51717 rows and 17 columns; after data Quality check, it has dataset contains 49853 rows and 10 columns
2. The missing Value have been replaced with or calculating avg of the column.
3. Deleted some null values.

- **Data Understanding:**

Column Name	Description	Data Type
url	contains the url of the restaurant in the Zomato website	
address	contains the address of the restaurant in Bengaluru	object
name	contains the name of the restaurant	object
online_order	whether online ordering is available in the restaurant or not	object
book_table	table book option available or not	object
rate	the overall rating of the restaurant out of 5	object
votes	total number of ratings for the restaurant	int
phone	the phone number of the restaurant	object
location	contains the neighborhood in which the restaurant is located	object
rest_type	restaurant type	
dish_liked	dishes people liked in the restaurant	object
cuisines	food styles, separated by comma	object
approx_cost (for 2 people)	contains the approximate cost for meal for two people	object

reviews_list	list of tuples containing reviews for the restaurant, each tuple consists of two values, rating and review by the customer	object
menu_item	contains list of menus available in the restaurant	object
listed_in(type)	type of meal	object
listed_in(city)	contains the neighborhood in which the restaurant is listed	object

An explanation for why you've chosen this data set

The basic idea of analyzing the Zomato dataset is to get a fair idea about the factors affecting the establishment of different types of restaurants at different places in Bengaluru, aggregate rating of each restaurant.

Bengaluru being one such city has more than 12,000 restaurants with restaurants serving dishes from all over the world. With each day new restaurants opening the industry hasn't been saturated yet and the demand is increasing day by day.

In spite of increasing demand, it however has become difficult for new restaurants to compete with established restaurants. Most of them serving the same food. Bengaluru being an IT capital of India. Most of the people here are dependent mainly on the restaurant food as they don't have time to cook for themselves. With such an overwhelming demand of restaurants it has therefore become important to study the demography of a location.

Limitation and ethics

1. Limitation: The contains a lot of missing values and also contains unnecessary information. The Data contains many errors in typing.

2. Ethical issue: The data contains some personal Information so getting rid of those columns for security purposes.

Define questions to explore. In a third section of your project document, define a list of questions to explore with your analysis.

- What kind of a food is more popular in a locality
- What are the popular restaurants by location
- Cheapest, highest rated and largely voted restaurants
- Is there a relation between cuisine, location and the cost?
- Expensive, Highest rated and largely voted.
- Which locality of that city serves those cuisines with maximum number of restaurants
- The needs of people who are striving to get the best cuisine of the neighborhood
- Is a particular neighborhood famous for its own kind of food.