

CHAPTER -4

DATA ANALYSIS AND INTERPRETATION

4.1 ZOMATO EXPLORATORY DATA ANALYSIS

4.1.1 Introduction

This Python code is used for analyzing the Zomato restaurant data of Bangalore city. It consists of various statistical tests, data cleaning, and data visualization techniques to extract useful insights from the dataset.

4.1.2 Libraries Used

The following libraries are used in the code:

- pandas: For data manipulation and analysis
- numpy: For numerical computations
- seaborn and matplotlib.pyplot: For data visualization
- pingouin: For statistical analysis

4.1.3 Data

The data used in the code is present in a CSV file named "BangaloreZomatoData-2022.csv". The data contains information about various restaurants in Bangalore, including their name, location, type of cuisine, ratings, and reviews.

4.1.4 Code Explanation

Reading Data

The first step of the code is to read the CSV file containing the restaurant data using pandas' read_csv method. The data is stored in a variable named "zomatodata".

Test of Reliability

In this section, Cronbach's alpha test is performed to determine the reliability of the data. The "pg" library is used for calculating Cronbach's alpha. The columns that contain the items or measures of interest are selected, and Cronbach's alpha is calculated using the `pg.cronbach_alpha()` method. The result is stored in the "alpha" variable, which is then printed to the console.

Test of Normality

This section performs the Kolmogorov-Smirnov test to check the normality of the data. The `scipy.stats` library's `kstest` method is used to perform the test. The columns with numerical data are selected, and their values are flattened into a single array. The `kstest` method is then applied to this array, and the p-value is stored in the "p" variable. The alpha level is set to 0.05, and the test result is printed to the console.

Descriptive Statistics

This section displays the descriptive statistics of the entire dataset using the `describe` method of the `pandas` library.

Table 4.1

	IsHomeDelivery	isTakeaway	isIndoorSeating	isVegOnly	Dinner Reviews	Delivery Reviews	AverageCost
count	8923.000000	8923.000000	8923.000000	8923.000000	8923.000000	8923.000000	8923.000000
mean	0.997871	0.660092	0.442676	0.072285	157.106018	2014.709403	340.225261
std	0.046098	0.473704	0.496731	0.258974	731.834227	5524.430542	308.338943
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	50.000000
25%	1.000000	0.000000	0.000000	0.000000	0.000000	42.000000	150.000000
50%	1.000000	1.000000	0.000000	0.000000	0.000000	279.000000	250.000000
75%	1.000000	1.000000	1.000000	0.000000	43.000000	1493.500000	400.000000
max	1.000000	1.000000	1.000000	1.000000	26500.000000	99600.000000	4200.000000

Table of Descriptive Statistics

Understanding the Data

Several methods are used to gain a better understanding of the data, such as:

1) Head

The head method displays the first 5 rows of the dataset.

2) Tail

The tail method displays the last 5 rows of the dataset.

3) Datatype of Columns

The dtypes method displays the datatype of each column in the dataset.

4) Dataset Shape

The shape method displays the number of rows and columns in the dataset.

5) Column Names

The columns method displays the names of all the columns in the dataset.

6) Unique Values in Columns

The nunique method displays the number of unique values present in each column of the dataset.

7) Cleaning Data

The isnull method is used to check if any columns contain null values. If the "Timing" column is present, it is dropped from the dataset. The isnull method is used again to confirm that the null values have been removed.

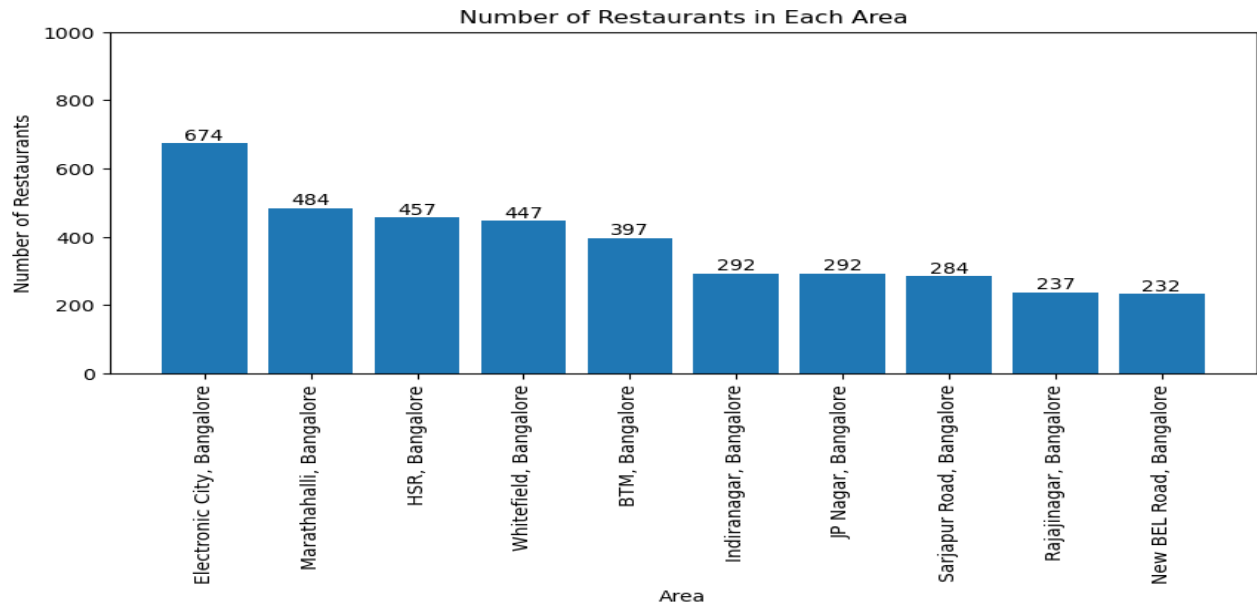
4.2 Data Visualization

Various data visualization techniques are used to gain insights into the dataset. The following methods are used:

8) Bar Chart to Show the Number of Restaurants in Each Location

A bar chart is plotted to show the number of restaurants in each area of Bangalore. The top 10 areas with the highest number of restaurants are selected and plotted. The `plt.bar()` method is used to plot the bar chart.

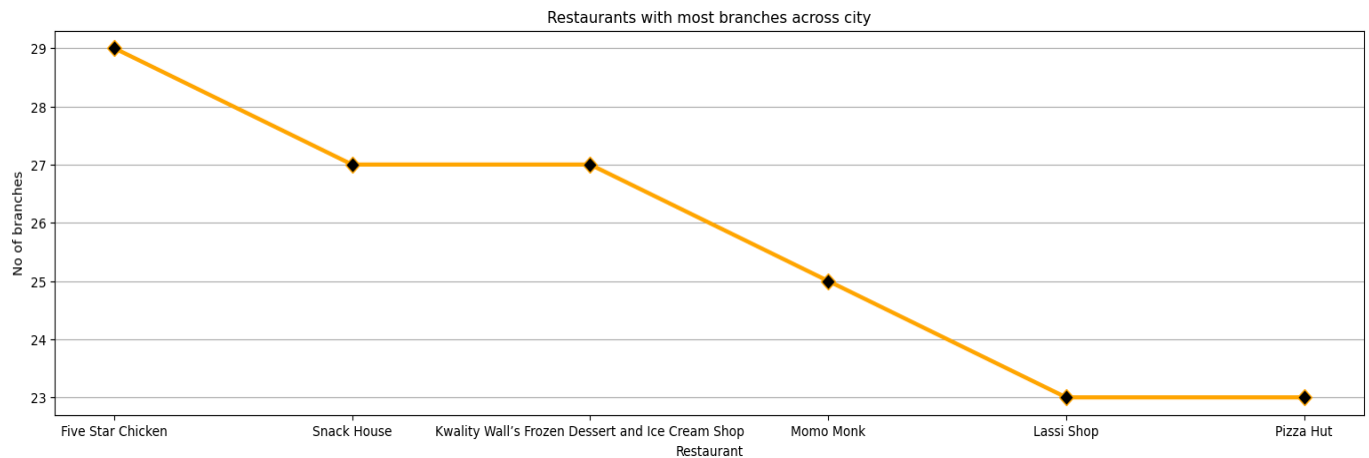
Chart 4.1



9) Count of the Number of Times a Restaurant's Name has been Repeated

The number of times each restaurant's name appears in the dataset is counted. The `group by` method is used to group the dataset by the "Name" column, and the count of each group is stored.

Chart 4.2



10)Here I filtered the restaurants by top rating and cheap price

The next section filters the dataset by the top rating and cheap price. It replaces the missing values with NaN using the. replace() function, calculates the total rating by adding the Dinner and Delivery Ratings, and selects the top 6 values. It groups the dataset by the 'Name' column and calculates the mean of the total rating, sorts it in descending order, and selects the top 5 values.

Table 4.2

	IsHomeDelivery	isTakeaway	isIndoorSeating	isVegOnly	Dinner Reviews	Delivery Reviews	AverageCost	count	total_rating
Name									
Brahmin's Coffee Bar	1.0	1.0	0.0	1.0	2976.0	3923.000000	100.0	1.0	9.4
CTR Shri Sagar	1.0	1.0	1.0	1.0	4869.0	27600.000000	150.0	1.0	9.3
Burma Burma	1.0	1.0	1.0	0.0	2899.0	1123.000000	1500.0	1.0	9.3
Bologna	1.0	1.0	1.0	0.0	1542.0	344.000000	1600.0	1.0	9.0
The Blue Wagon - Coffee & Kitchen	1.0	1.0	1.0	0.0	370.0	244.000000	400.0	1.0	9.0
...
Zomoz - The Momo Company	1.0	0.0	0.0	0.0	0.0	497.727273	200.0	1.0	NaN
Zoya Kabab Centre	1.0	1.0	0.0	0.0	0.0	5.000000	200.0	1.0	NaN
Zwama	1.0	1.0	0.0	0.0	2.0	67.000000	300.0	1.0	NaN
energii - Bowls, Salads & Smoothies	1.0	0.0	0.0	0.0	0.0	63.000000	250.0	1.0	NaN
foodle	1.0	0.0	0.0	0.0	0.0	2.000000	150.0	1.0	NaN

Table showing the list Restaurants based on price

11) Took the name of Restaurant for Bar Chard

The next section creates a bar chart using the matplotlib library to show the name and rating of the top 5 restaurants. It creates two series 'price' and 'rating' containing the average cost and total rating of the top 5 restaurants and plots them using the .bar() function. It also adds labels, a title, and limits to the chart.

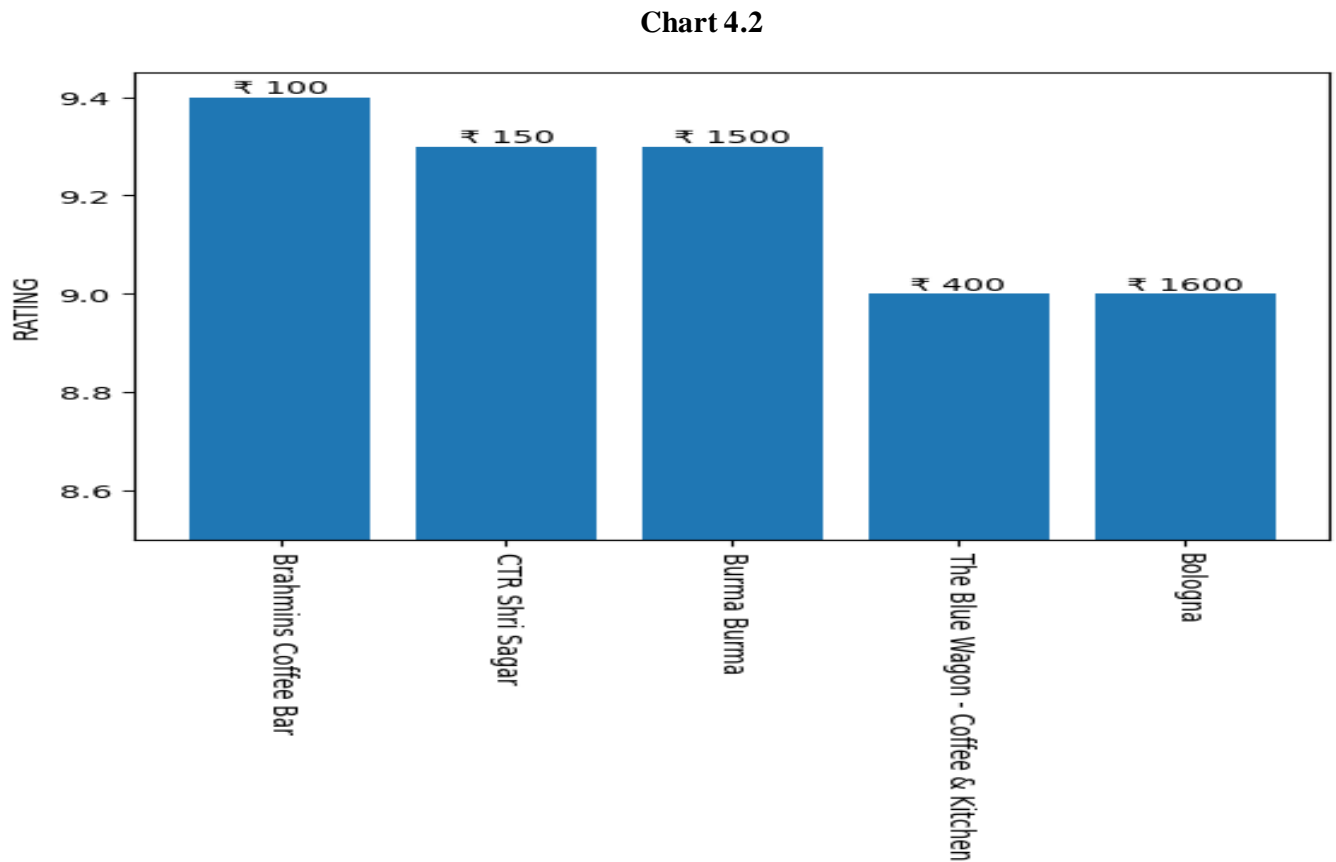


Table shows Top 5 Restaurants and their price