**Mini Project 2 Report**

1. **The data and its source**

* Dataset name: Swiggy Restaurants Dataset.
* Source: Kaggle.
* Link: [Swiggy Restaurants Dataset | Kaggle](https://www.kaggle.com/datasets/ashishjangra27/swiggy-restaurants-dataset)

1. **Description of data exploration and data cleaning steps**

|  |  |
| --- | --- |
| **Column name** | **Data type** |
| id | Object |
| Restaurant | Float64 |
| City | Object |
| SubRegion | Object |
| cost | Object |
| cuisine | Object |
| rating | Float64 |
| rating\_count | Object |
| lic\_no | Object |
| address | Object |
| menu | Object |
| link | Object |
| type | Object |
| price | Object |

* The dataset in total has 14 columns and 160621 rows.
* Type and price columns have more than 90% null values hence both the columns are dropped and rows with null values are dropped.
* After cleaning the dataset we have 12 columns and 148290 rows.
* Restaurant ‘id’ should always be unique hence dropped the duplicate rows.

1. **Unit of analysis**
2. **Rating Analysis**
3. Restaurant with Maximum rating throughout the dataset

* Getting the name of the restaurant with maximum rating.
* There are 209 Restaurant with 5.0 rating.

ii. Restaurant Counts w.r.t rating

* Getting to know which rating has maximum number of restaurants and which rating has a minimum number of restaurants.
* Plotting it on bar graph with the help of Seaborn package.

1. **City Analysis**
2. Number of cities with least number of restaurants listed

* Getting to know the cities with least number of restaurants using min () function.
* There are 5 cities with least number of restaurants to know it.

ii. Least Popular Cities as per Restaurants Counts

* Getting to know the least popular cities with number of restaurants in that city.
* Plotting it on bar graph with the help of Seaborn package.

1. **Cuisines Analysis**
2. Most popular Cuisines served throughout the dataset for the unique cuisine pair.
   * Getting to know which unique cuisine pair is most popularly served.
   * Plotting it on bar graph with the help of Seaborn package.
3. **Description of the program**
4. Importing the packages and accessing the JSON file.

**Graphical user interface, text, application

Description automatically generated**

1. Knowing the number of rows and columns, checking for duplicates and dropping the duplicate rows.

A picture containing scatter chart

Description automatically generated

1. Checking for null values and dropping the columns with more null values and rows too.

Graphical user interface, text, application

Description automatically generated

A picture containing text

Description automatically generated

1. Generating the list of restaurants with maximum rating.

Graphical user interface, application

Description automatically generated

1. Plotting a bar graph of restaurant counts with respect to rating.

Text

Description automatically generated

Chart, histogram

Description automatically generated

1. Generating a list of cities with least number of restaurants.

Graphical user interface, text, application

Description automatically generated

1. Plotting a bar graph of least popular city as per restaurants count.

Text

Description automatically generated

Chart, bar chart, histogram

Description automatically generated

1. Plotting a bar graph for the most popular Cuisines served throughout the dataset [In terms of frequency]

Graphical user interface, text

Description automatically generated

Chart, bar chart

Description automatically generated

1. **Description of the output files.**

1. Output file name: City.csv

The csv file has the city name and the count of restaurants.