# SUMMER INTERNSHIP PROJECT REPORT

## *"SWIGGY DATA ANALYSIS"*

*Submitted in partial fulfilment of the requirements for the award of the degree of*

**Master of Technology**
**In**
**Data Science**
**By**

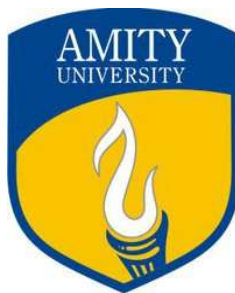**Gathram Kesava Murthy**

**Enrollment No: - A50568423003**

*Under the guidance of*

**Dr. Riya Sapra**

**Associate Professor**

**CSE department**



**Department of Computer Science &Engineering**

**Amity School of Engineering & Technology**

**Amity University Haryana**

**Gurgaon, India**

**November 2024**

**Department of Computer Science and Engineering**

**Amity School of Engineering and Technology**

## DECLARATION

I, **Gathram Kesava Murthy** ,student of Master of Technology in Data Science from the Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University Haryana, hereby declare that I am fully responsible for the information and results provided in this project report titled "**SWIGGY DATA ANALYSIS**" submitted in Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University Haryana, Gurgaon for the partial fulfilment of the requirement for the award of the degree of *Master of Technology in Data Science*. I have taken care in all respects to honor the intellectual property rights and have acknowledged the contributions of the others for using them. I further declare that in the case of any violation of intellectual property rights or copyrights, I as a candidate will be fully responsible for the same, Head of department and the Institute should not be held for full or partial violation of copyrights if found at any stage of my degree.

**Gathram Kesava Murthy**

**A50568423003**

**Department of Computer Science and Engineering**

**Amity School of Engineering and Technology**

## CERTIFICATE

This to certify that the work in the project report entitled "**Swiggy Data Analysis**" by **Gathram Kesava Murthy** bearing Enrollment No. **A50568423003** is a Bonafede report of project work carried out by him under my supervision and guidance in partial fulfilment of the requirement for the award of the degree of **Master of Technology in Data Science** in the Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University Haryana, Gurgaon. Neither this project nor any part of it has been submitted for any degree or academic award elsewhere. The work was satisfactory. They have shown complete dedication and devotion to the given project work.

**Signature of Supervisor(s)**

**Dr. Riya Sapra**
**Associate Professor,**
**Computer Science & Engineering,**
**ASET, Amity University Haryana**

**Dr. Shivani Sharma**
**Associate Professor,**
**Computer Science & Engineering,**
**ASET, Amity University Haryana**

# CERTIFICATE OF INTERNSHIP



**Issued Date:** *01-09-2024*

CIN No- U85500HR2023PTC115118

## CERTIFICATE
### OF INTERNSHIP

**UNIFIED MENTOR**
YOUR SKILL, SUCCESS & JOURNEY

*Gathram Kesava Murthy*

For successfully completing three months internship as Data Science Intern at Unified Mentor Pvt Ltd. Dated from 01-06-2024 to 01-09-2024
During the internship we found him/her consistent & hard-working. We wish them all the best for their future endeavors.

Paras Grover
**Paras Grover**
Director

ISO
9001:2015
CERTIFIED COMPANY

Sanket Patil
**Sanket Patil**
Awarded By

Verify at:

AN ISO **9001:2015** Certified Company

# LETTER OF RECOMMENDATION OF INTERNSHIP

**UNIFIED MENTOR**
YOUR SKILL, SUCCESS & JOURNEY

**Address**
SCO 17-18, STREET 31C, S BLOCK
SECTOR 24, GURUGRAM, INDIA
PIN - 122010

Unified Mentor Pvt. Ltd.
Tel:+91 6283 800330
www.unifiedmentor.com
CIN: U85500HR2023PTC115118

Date: 10-10-2024

## To whom it may concern

Dear **Gathram Kesava Murthy**,

I am delighted to recommend **Gathram Kesava Murthy** for the **Data Science Internship** at **Unified Mentor Pvt Ltd**. Throughout his tenure, he exhibited a strong passion for data analysis and a solid understanding of data science principles.

Demonstrating exceptional analytical skills and a keen attention to detail, he effectively utilized tools such as Python, SQL, and other relevant tools to extract insights and derive meaningful conclusions from data.

One of his notable strengths is his ability to tackle complex problems and find innovative solutions.

He also demonstrated strong communication skills and the ability to collaborate effectively with team members. His positive attitude and eagerness to learn made him a valuable addition to our team.

Based on his performance and dedication, I am confident that he will continue to excel in the field of data science.

**Regards,**

*Paras Grover*

**Paras Grover**
Director/Founder

**Phone**
+91 6283 800330

**Email**
info@unifiedmentor.com

UNIFIED MENTOR
INDIA

**Department of Computer Science and Engineering**

**Amity School of Engineering and Technology**

# ACKNOWLEDGEMENT

This summer internship project has been a wonderful experience which consumed hard work and dedication. The successful completion of the summer internship project and compilation of report would not have been possible if I did not have the support of some very important people who have been a constant source of guidance and supervision. Therefore, I would like to extend my sincere gratitude to each one of them.

It's my radiant sentiment to place on record my best regards and express my sincere thanks to director of Amity School of Engineering and Technology, **Prof. (Dr.) Shalini Bhaskar Bajaj**, my supervisor **Dr. Riya Sapra** for providing me with this opportunity to gain practical field knowledge through this summer internship project.

I would like to thank the organization **UNIFIED MENTOR PRIVATE LIMITED** and **Mr. Paras Grover** to give me the opportunity to provide training in the field of Data Science and analytics using python, sql and tableau. I had a great experience and learnt a lot from this internship.

I would also like to thank **Mr. Paras Grover** for giving me this auspicious opportunity.

**Gathram Kesava Murthy**
**Enrollment No. – A50568423003**

# ABSTRACT

This analysis dives into Swiggy's restaurant data to uncover key trends that shape customer preferences and guide business strategies. As one of India's leading food delivery platforms, Swiggy depends on data-driven insights to improve services and strengthen customer satisfaction. The dataset used in this study includes essential details about restaurants, such as ratings, costs, cuisines, and city-wise distribution, providing a rich foundation for exploration.

The goal of this analysis was to understand how restaurant ratings, pricing, and cuisine popularity vary across different cities. To achieve this, Python libraries like pandas and numpy were used for data manipulation and analysis. The data was carefully cleaned and preprocessed to address missing values, standardize formats, and ensure reliability. Clear and engaging visualizations were also created to make the findings accessible and actionable.

The results revealed fascinating trends. Some cities have a dense concentration of restaurants, while others offer fewer but often higher-rated establishments. Popular cuisines, especially those associated with high ratings, were identified, helping to understand what customers prefer. Cost analysis shed light on affordability trends, showing where premium or budget-friendly options are more common. Customer reviews offered valuable insights into what drives positive feedback and repeat orders.

This study emphasizes the power of data in decision-making. The findings can help Swiggy tailor its services to meet customer demands, improve partnerships with restaurants, and refine marketing strategies. Restaurant owners, in turn, can use this information to better align with customer expectations and boost their performance. By acting on these insights, Swiggy can continue to innovate and deliver a superior experience to its users, maintaining its position as a leader in the competitive food delivery market.

# LIST OF FIGURES

# Contents

# CHAPTER 1
# INTRODUCTION

The food delivery industry has revolutionized how people enjoy their meals, bringing convenience and variety to customers' fingertips. Swiggy, one of India's leading food delivery platforms, has played a significant role in shaping this transformation. With thousands of restaurants listed on its platform, Swiggy offers a diverse range of cuisines, price points, and services to cater to the preferences of millions of users. However, staying ahead in a competitive market requires more than just a wide selection of restaurants; it demands a deep understanding of customer preferences, restaurant performance, and pricing trends.

Data-driven insights are vital for Swiggy to enhance its offerings, improve user satisfaction, and optimize its operations. These insights also benefit restaurant owners by helping them align their services with customer expectations. This study aims to analyze a dataset containing restaurant information from Swiggy's platform, focusing on key factors such as ratings, costs, cuisines, and city-specific performance. By exploring and understanding these aspects, the study seeks to uncover patterns and trends that can drive strategic decisions for both Swiggy and its partner restaurants.

## 1.1    Aim

The primary aim of this study is to perform a comprehensive analysis of Swiggy's restaurant data to identify meaningful trends and insights. The analysis focuses on factors such as ratings, pricing, and cuisine preferences, with the ultimate goal of improving customer experiences and helping Swiggy optimize its services. Through data-driven exploration, this study also aims to support restaurant owners in refining their offerings to meet market demands.

## 1.2    Objective

 To achieve the overarching aim, this study is guided by several key objectives:

### 1.2.1 Analyze Restaurant Ratings

Restaurant ratings are a critical indicator of customer satisfaction. This study examines the distribution of ratings across different restaurants to identify patterns that contribute to high performance. It also explores the relationship between ratings and other factors, such as cuisine type and cost, to understand what drives positive feedback.

### 1.2.2 Understand Cost Patterns

The pricing of meals plays a significant role in attracting customers. By analyzing the cost distribution across restaurants, this study identifies locations with premium offerings and those known for affordable meals. Understanding these patterns can help Swiggy tailor its platform to accommodate diverse customer segments.

### 1.2.3 Explore Cuisine Popularity

Cuisines are a major factor in customer decision-making. This study investigates the most popular cuisines, particularly those associated with high ratings or frequent customer orders. Insights into cuisine preferences can guide Swiggy and restaurant owners in designing menus that resonate with customers.

### 1.2.4 City-wise Performance Analysis

Swiggy operates across multiple cities, each with unique customer preferences and restaurant dynamics. This objective focuses on evaluating city-level restaurant distributions, identifying high-performing locations, and understanding regional trends in ratings and pricing.

### 1.2.5 Generate Insights for Business Strategy

Ultimately, the findings of this analysis aim to provide actionable insights that can inform Swiggy's business strategy. These insights will help Swiggy enhance customer engagement, improve restaurant partnerships, and optimize its operational efficiency.

## 1.3   Proposed Work

To achieve the objectives outlined above, the study adopts a systematic approach that combines data exploration, preprocessing, analysis, and visualization. This structured methodology ensures that the insights generated are reliable, actionable, and relevant to Swiggy and its stakeholders.

### 1.3.1 Data Exploration

The first step involves exploring the dataset to understand its structure and content. The dataset includes information such as restaurant ratings, pricing, cuisines, and city details. Initial exploration involves summarizing the data using basic statistical methods to identify trends,

patterns, and potential issues such as missing or inconsistent values. The exploration phase also provides an overview of key metrics, such as the number of restaurants in the dataset, the range of ratings, and the distribution of cost values. This preliminary analysis helps frame the focus areas for deeper investigation.

### 1.3.2 Data Preprocessing

Before diving into detailed analysis, the dataset undergoes preprocessing to address any inconsistencies and prepare it for exploration. This step includes:

- **Handling Missing Values**: Missing data is managed by either removing incomplete entries or imputing missing values with suitable defaults, such as the median rating or average cost.
- **Standardizing Formats**: Features like "cost," which may include non-numeric characters (e.g., currency symbols), are converted to numeric formats for seamless analysis.
- **Encoding Categorical Features**: Text-based features, such as cuisine types, are encoded into numerical representations where necessary to facilitate analysis.

By ensuring a clean and consistent dataset, this phase lays a strong foundation for generating accurate and meaningful insights.

### 1.3.3 Exploratory Data Analysis (EDA)

EDA is the heart of this study, aiming to uncover patterns and trends within the data. Key areas of focus include:

- **Rating Analysis**: This involves examining the distribution of ratings across restaurants, identifying the characteristics of highly-rated establishments, and understanding how ratings vary by cuisine type or city.
- **Cost Analysis**: Cost distribution is analyzed to identify price trends and outliers. Restaurants are categorized into different pricing tiers, such as budget-friendly, mid-range, and premium, to understand their appeal to different customer segments.
- **Cuisine Analysis**: Popular cuisines are identified using frequency counts and rating thresholds. This analysis also explores which cuisines are most common among high-rated restaurants.

- **City-wise Insights**: City-level comparisons are conducted to evaluate restaurant density, average ratings, and cost trends. Cities with exceptional performance are highlighted as benchmarks for success.

### 1.3.4 Visualization

Data visualizations play a crucial role in communicating the findings effectively. Using charts, graphs, and heatmaps, the study represents trends and patterns in a clear and engaging manner. Visualizations help stakeholders quickly grasp the insights and their implications, making it easier to act on the findings.

## 1.4 Benefit of Study

The insights generated from this study are expected to have a significant impact on Swiggy's operations and restaurant partnerships. By understanding customer preferences and market trends, Swiggy can:

- Enhance its platform's user experience by promoting restaurants and cuisines that align with customer demands.
- Optimize marketing strategies to attract and retain more customers.
- Strengthen relationships with restaurant owners by providing them with data-driven guidance to improve their performance.

Restaurant owners can also leverage these insights to:

- Align their pricing and menus with customer preferences.
- Focus on cuisines that are most popular in their target regions.
- Improve their service quality to achieve higher ratings and customer satisfaction.

# CHAPTER 2

# BACKGROUND

To truly understand the scope of this analysis, it's essential to explore the broader context in which it operates. Swiggy, one of India's most prominent food delivery platforms, has revolutionized the way people enjoy their meals. By bringing a diverse range of restaurants to customers' doorsteps, Swiggy has become synonymous with convenience and choice. However, its continued success depends on its ability to adapt and evolve, guided by data-driven insights. This section delves into the company's history, its impact on the food delivery landscape, and the dataset that serves as the foundation of this study.

## 2.1 About Swiggy



Figure 2.1 Swiggy Logo

Swiggy, founded in 2014 in Bangalore, began as a solution to a simple problem: connecting people with great food efficiently. Over the years, it has grown into one of India's most beloved food delivery platforms, known for its seamless experience and wide range of offerings.

### 2.1.1 Swiggy's Core Strengths

Swiggy stands out for its focus on user convenience and its commitment to providing value for both customers and restaurant partners. Some of its standout features include:

- **Intuitive User Interface**: Swiggy's app makes it easy for users to browse menus, compare options, and place orders quickly.
- **Diverse Restaurant Network**: From small local eateries to luxury dining, Swiggy ensures that users can find something for every occasion.
- **Reliable Delivery System**: Known for its quick deliveries, Swiggy's delivery network has set a benchmark in the industry.

- **Customer-Centric Approach**: By offering frequent discounts, personalized recommendations, and loyalty programs, Swiggy keeps its customers engaged and satisfied.

### 2.1.2 Swiggy's Market Presence

Today, Swiggy operates in more than 500 cities across India, serving millions of customers daily. It has also expanded its services to include grocery delivery through Swiggy Instamart and personalized deliveries through Swiggy Genie. The platform's adaptability and innovation have made it a trusted name in urban and suburban areas alike.

### 2.1.3 Why Swiggy Relies on Data

In a competitive market with other key players like Zomato and Dunzo, data has become Swiggy's secret weapon. By analyzing customer behavior, restaurant performance, and delivery trends, Swiggy can stay ahead of the curve. For instance, identifying which cuisines are trending in certain cities or understanding why some restaurants perform better than others can lead to more strategic decisions.

## 2.2 About Dataset

The dataset used in this analysis provides a window into Swiggy's ecosystem, showcasing the variety and performance of restaurants listed on the platform. It offers valuable details that help paint a picture of customer preferences, restaurant trends, and market dynamics.

### 2.2.1 What is in dataset?

The dataset includes comprehensive information about Swiggy's restaurant network, covering key attributes like:

- **Restaurant Name and ID**: Unique identifiers and names of restaurants help track individual performance.
- **City**: Each restaurant's location, giving insights into regional preferences.
- **Ratings**: Average customer ratings (on a scale from 0 to 5) reflect the quality of food, service, and overall experience.
- **Rating Counts**: The number of reviews contributes to understanding the reliability of the ratings.

- **Cost**: The average cost for two people indicates a restaurant's affordability.
- **Cuisine Types**: Cuisines offered by the restaurant, such as Indian, Chinese, or Italian.

**2.2.2 Challenges in dataset**

- **Missing Data**: Some fields, such as ratings or costs, have gaps that need to be filled or accounted for during analysis.
- **Inconsistent Formats**: For instance, cost data may include currency symbols that must be cleaned before analysis.
- **Categorical Complexity**: Features like "cuisines" often include multiple entries per restaurant, requiring careful processing to extract useful insights.

# Chapter 3
# TECHNOLOGIES USED

In the rapidly evolving world of data analysis, the tools and technologies employed are crucial to efficiently process, analyze, and interpret large datasets. In this study, a combination of cutting-edge technologies and well-established libraries have been utilized to extract actionable insights from Swiggy's restaurant data. These technologies include programming languages, data manipulation libraries, visualization tools, and statistical models. Each of these technologies plays an integral role in transforming raw data into meaningful conclusions that can inform decision-making for both Swiggy and restaurant owners.

## 3.1 Python: The Core Programming Language

Python is a popular high-level programming language that is widely used in various fields, including data science, web development, scientific computing, artificial intelligence, and more. It was first released in 1991 by Guido van Rossum and has since become one of the most widely used programming languages in the world. One of the key features of Python is its simplicity and readability. It has a clear and concise syntax that makes it easy to learn and use, even for beginners. Python is also known for its versatility, as it can be used for a wide range of tasks, from simple scripting to complex software development. Python has a vast collection of libraries and frameworks that make it easy to perform various tasks, including data analysis, machine learning, and web development. For example, NumPy and Pandas are popular libraries for data manipulation and analysis, while Scikit-learn and TensorFlow are widely used for machine learning tasks. Figure 3.1 shows the logo of python. Python is an open-source language, which means that its source code is freely available and can be modified and distributed by anyone. This has led to a large and active community of developers who contribute to the development of the language and its associated libraries and frameworks.



Figure 3.1 Python

## 3.2 Data Manipulation and Analysis Libraries

The heart of any data analysis lies in its ability to manipulate and process raw data. For this study, several Python libraries are used to perform tasks such as data cleaning, handling

missing values, feature engineering, and preparing the data for visualization and statistical analysis.

**3. 2.1 Pandas: The Powerhouse of Data Analysis**

Pandas is the cornerstone of data manipulation in Python. It provides powerful and flexible data structures that allow for efficient handling of structured data. Pandas' key data structure, the **DataFrame**, is a two-dimensional size-mutable, potentially heterogeneous tabular data structure with labeled axes (rows and columns), which makes it ideal for analyzing datasets with mixed data types. Pandas facilitates the following tasks:

- **Data Cleaning**: One of the most important aspects of data analysis is cleaning the data. Pandas provides robust methods to identify and handle missing data, duplicate records, and outliers.
- **Data Transformation**: Using methods like groupby(), merge(), and pivot_table(), Pandas makes it easy to aggregate, join, and reshape data. For instance, when dealing with restaurant ratings and costs, Pandas allows you to group data by different criteria, such as city or cuisine, and then aggregate or summarize it.
- **Data Filtering**: Pandas makes it easy to filter the data based on conditions, which allows analysts to focus on specific subsets of data relevant to the analysis. For instance, filtering out restaurants with fewer than 10 reviews could help ensure that ratings are statistically significant.



Figure 3.2.1 Pandas

**3.2.2 NumPy: Numerical Operations at Scale**

While **Pandas** handles most of the data manipulation tasks, **NumPy** is used when complex numerical computations are needed. NumPy provides support for working with large, multi-dimensional arrays and matrices, and it integrates well with Pandas for numerical operations. In this study, NumPy supports operations such as:

- **Mathematical Operations**: Many of the analyses require simple or complex mathematical operations like calculating the mean, median, variance, and standard deviation of ratings and costs. NumPy enables these operations efficiently.
- **Handling Arrays**: For tasks such as handling large datasets or performing element-wise operations, NumPy's n-dimensional arrays are invaluable. These arrays allow for efficient computations without the overhead of traditional loops.



Figure 3.2.2 Numpy

## 3.3 Data Visualization Tools

Data visualization plays a key role in conveying insights clearly and concisely. The ability to visualize complex data and trends helps stakeholders (such as Swiggy or restaurant owners) understand key findings quickly and make informed decisions. In this study, two powerful Python libraries are used for visualization: **Matplotlib.**

### 3.3.1 Matplotlib: The Foundation of Data Visualization

**Matplotlib** is one of the most widely used libraries for generating static plots in Python. It provides a low-level interface for creating a wide range of visualizations, from basic line charts to complex scatter plots and heatmaps. Some of the key uses of Matplotlib in this study include:

- **Line and Bar Graphs**: These are commonly used to show trends over time or compare different categories, such as restaurant performance across cities or cost distributions across cuisines.
- **Pie Charts**: To visualize the distribution of different cuisines or restaurant types across the dataset.
- **Histograms**: To illustrate the distribution of restaurant ratings, cost categories, and other continuous variables.



Figure 3.3.1 Matplotlib

10

# Chapter 4

## DESIGN OF PROJECT

## 4.1 Hardware Configuration

**RAM**: 16 GB.

**Storage**: 1 TB SSD.

**CPU**: 12th Gen Intel(R) Core(TM) i7-12650HX @ 2.60GHz.

## 4.2 Software Configuration

**Operating System**: Windows 11

**Code editor**: VS Code

**Python Version:** Python 3.9.10

**Libraries Used:** Pandas, Numpy, Matplotlib

## 4.3 Data Flow Diagram

### 4.3.1 Level 0 DFD

This represents the system at a high level:

- Input: Raw restaurant data (ID, Name, Rating, Cost, Cuisine, etc.).
- Process: Preprocess data, analyze trends, and generate insights.
- Output: Analytical insights on ratings, costs, cuisines, and city-specific performance.
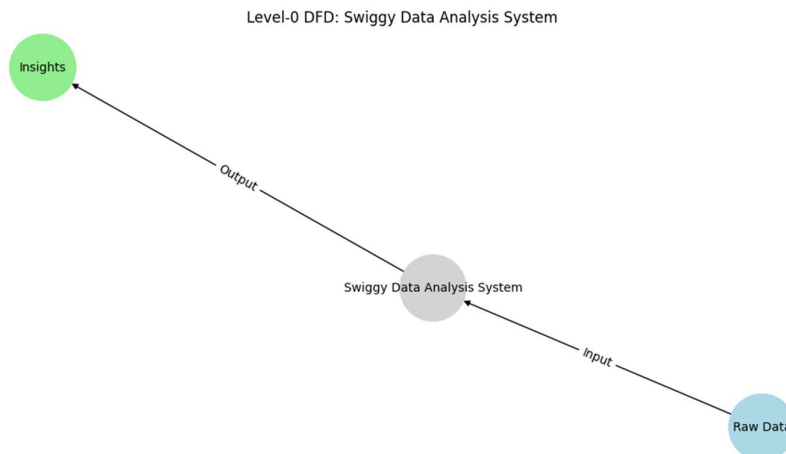


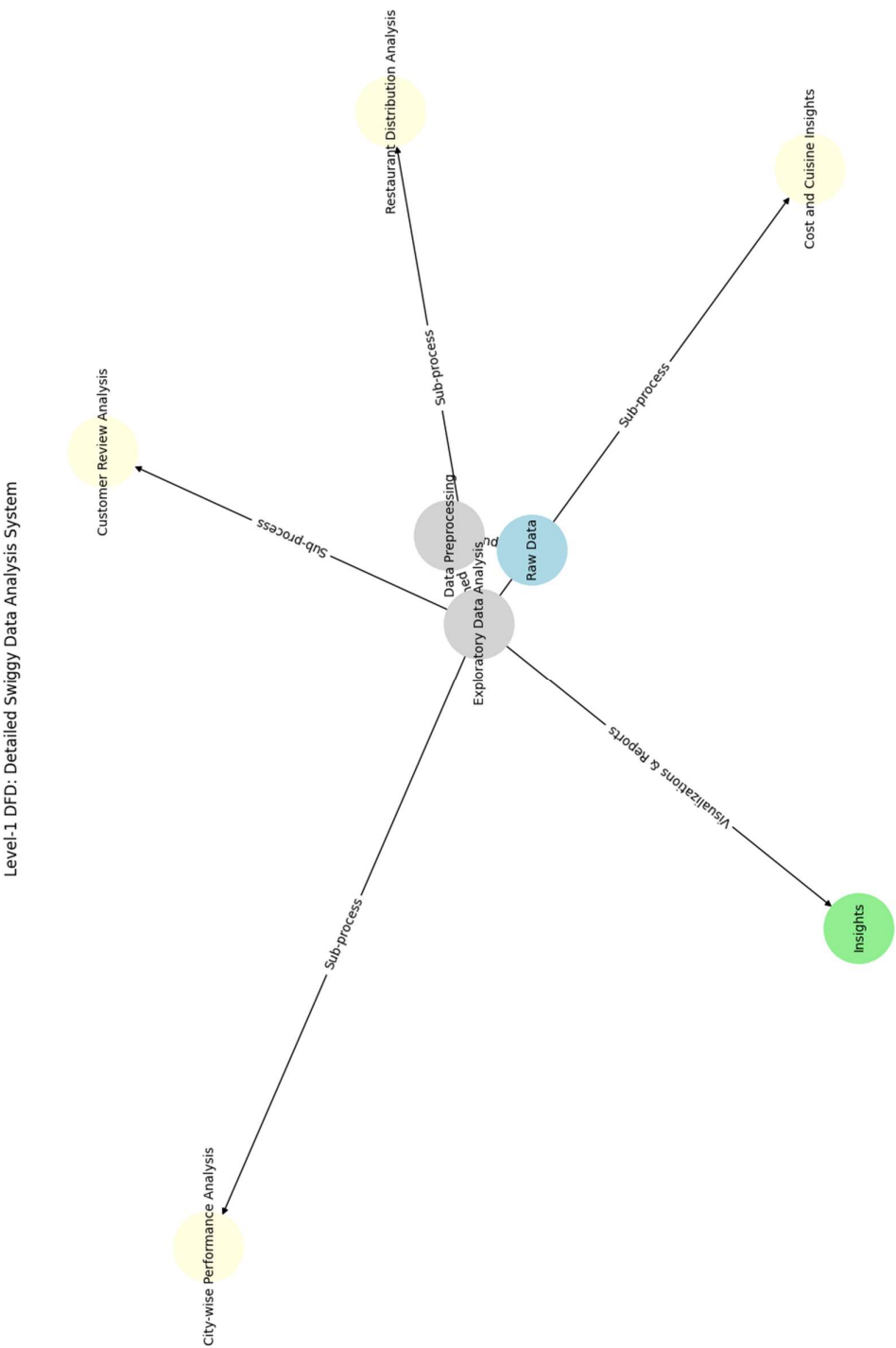Figure 4.3.1: Data Flow Diagram level 0

**4.3.2 LEVEL 1 DFD**



Figure 4.3.2: Data Flow Diagram level 1

This breaks down the processes further:

**1.Data Preprocessing**

- Input: Raw dataset

- Activities: Handling missing values, data type conversion, encoding.

- Output: Cleaned dataset

**2. Exploratory Data Analysis (EDA)**

- Input: Cleaned dataset

- Sub-processes:

    o Restaurant Distribution Analysis

    o Customer Review Analysis

    o City-wise Performance Analysis

    o Cost and Cuisine Insights

- Output: Visualizations, key trends, and detailed reports.

# Chapter 5

# IMPLEMENTATION

## 5.1 Data Preprocessing

The data preprocessing phase ensures the dataset is clean and ready for analysis. Key steps include:

**Handling Missing Values**

- **Null Value Analysis**: Missing values are identified and addressed.
- **Imputation Strategy**: Missing ratings are filled with -999.0 to ensure data integrity without removing rows. This prevents losing potentially useful data while clearly marking invalid entries.

**Data Type Conversion**

- The "cost" column is converted from a string format (e.g., "₹500") to numeric values. This involves removing symbols and spaces to enable numerical computations.

**Categorical Encoding**

- Features like "rating count," which have categorical distributions, are optionally encoded for advanced analyses.

**Cleaning Text Data**

- Cuisines, addresses, and other text-based features are cleaned using string operations to standardize formats and remove inconsistencies.

## 5.2 Exploratory Data Analysis (EDA)

EDA focuses on uncovering trends and insights through statistical and visual techniques. The steps involved are as follows:

**a. Restaurant Distribution**

- **City Analysis**:
  - Count the number of restaurants in each city to identify Swiggy's most significant markets.
  - Visualize restaurant density across cities using bar charts.
- **Popular Chains**:
  - Determine the most frequently occurring restaurant names to highlight popular chains.

**b. Customer Reviews**

- **Rating Analysis**:
    - Explore the distribution of ratings across all restaurants.
    - Categorize restaurants into high-rated, average-rated, and low-rated based on thresholds (e.g., >3.5 is high-rated).
- **Review Frequency**:
    - Examine the number of ratings per restaurant to understand customer engagement levels.

**c. City Performance**

- Identify cities with the highest average ratings.
- Compare ratings and review counts across cities to identify patterns.

**d. Cost Analysis**

- Analyze the distribution of costs among restaurants using histograms or box plots.
- Highlight the most expensive and affordable restaurants.

**e. Cuisine Popularity**

- Use string manipulation techniques to split and analyze cuisine types.
- Identify cuisines frequently associated with high-rated restaurants.

## 5.3 Implementation Details

### a. Python Libraries

The following libraries are used:

- **Pandas**: For data manipulation and analysis.
- **Numpy**: For numerical operations.
- **Matplotlib and Seaborn**: For data visualization.

### b. Code Workflow

The code workflow includes:

1. **Loading Data**:
    - Load the CSV dataset into a Pandas DataFrame.
    - Display the first few rows using .head() for an overview.
2. **Data Cleaning**:
    - Handle null values and convert data types.
    - Standardize the formats of columns like cost and cuisine.
3. **Analysis**:
    - Group data by city, rating, and cuisine to answer specific questions.

4. **Visualization**:
   - o Create bar plots, pie charts, histograms, and scatter plots to visualize trends.
5. **Insights Extraction**:
   - o Use summary statistics and visual findings to derive actionable insights.

## 5.4 Key Findings

**Restaurant Distribution**

- Cities like Bangalore and Delhi have the highest number of restaurants listed.
- Several chains, such as Domino's and McDonald's, are dominant across multiple cities.

**Customer Reviews**

- Restaurants with higher ratings often have more reviews, indicating a correlation between popularity and perceived quality.
- Cities with the highest average ratings include Hyderabad and Pune.

**Cost Analysis**

- Most restaurants fall within a moderate price range (₹300-₹800 for two).
- A few high-end restaurants charge upwards of ₹2,500, catering to premium customers.

**Cuisine Popularity**

- North Indian and Chinese cuisines dominate the high-rated restaurant category.
- Emerging cuisines like Mexican and Thai have gained popularity in metro cities.

## 5.5 Challenges and Solutions

**Challenge: Handling Missing Data**

- Missing values in critical features like "rating" could skew the analysis.
- **Solution**: Assign placeholder values (-999.0) or compute medians for imputation.

**Challenge: Cleaning Cost Data**

- Costs in the dataset included symbols and inconsistent formats.
- **Solution**: Use regular expressions to extract numeric values.

**Challenge: Analyzing Cuisines**

- Restaurants often serve multiple cuisines, stored as comma-separated values.
- **Solution**: Use string manipulation and list comprehension for parsing.

## 5.6 Visualizations

Key visualizations generated include:

1. **Bar Chart**: Number of restaurants in each city.
2. **Histogram**: Distribution of restaurant ratings.
3. **Box Plot**: Cost distribution across cities.
4. **Pie Chart**: Popular cuisines among high-rated restaurants.
5. **Scatter Plot**: Correlation between cost and ratings.

## Chapter 6

## RESULTS

In the Swiggy Data Analysis project, various visualizations were used to explore and present insights into the dataset effectively. Below is a detailed explanation of each visualization and its significance in understanding the results:

## 6. 1 Distribution of Restaurant Ratings:

- **Purpose:** This line of code creates a bar graph to visualize the distribution of restaurant ratings.

- **Data:** It uses the ratings_df DataFrame, which contains the rating values and their corresponding percentages.

- **Visualization:** It plots the rating values on the x-axis and the percentages on the y-axis. alpha=0.5 makes the bars slightly transparent. This graph likely shows the distribution of ratings, such as 4.0, 4.5, 5.0, etc. and the percentage of restaurants having each of these ratings.
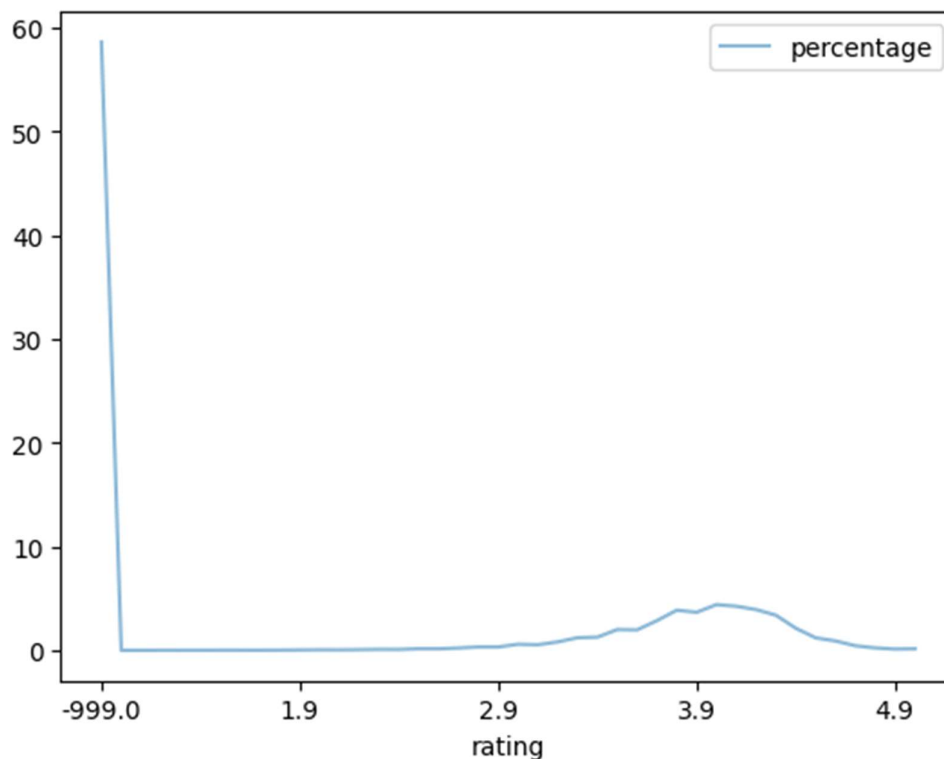


Figure 6.1: Distribution of Restaurant Ratings

## 6.2 Distribution of Restaurant Ratings (Excluding Missing Data):

- **Purpose:** This plot is similar to the previous one, but it excludes the "-999.0" rating, possibly representing restaurants with missing rating data.

- **Data:** It uses a filtered version of ratings_df where rows with rating '-999.0' are removed.

- **Visualization:** Same as the first graph, it shows a bar plot of rating distributions but focuses on the ratings available in the original dataset.
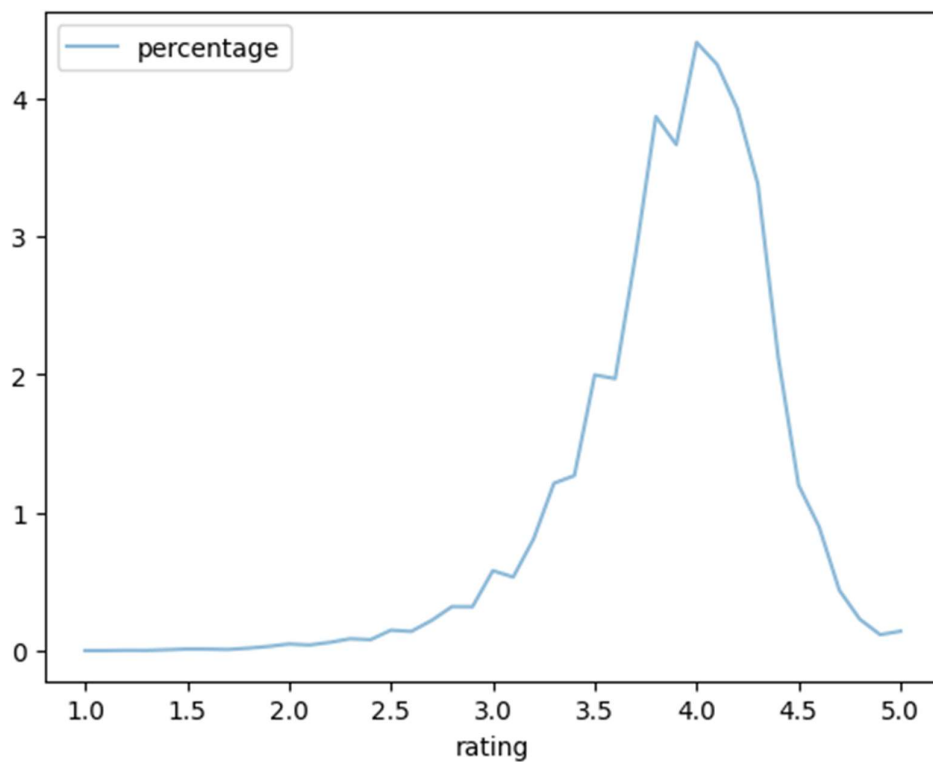


Figure 6.2: Distribution of Restaurant Ratings
(Excluding Missing Data)

## 6.3 Variation in Meal Costs Across Restaurants:

- **Purpose:** This line creates a line graph to visualize the distribution of the 'cost' feature (cost for two people).
- **Data:** It uses the 'cost' column of the df DataFrame.
- **Visualization:** It plots the index of the data points on the x-axis and the corresponding 'cost' values on the y-axis, providing a visual representation of the cost distribution, potentially showing outliers or general price trends.
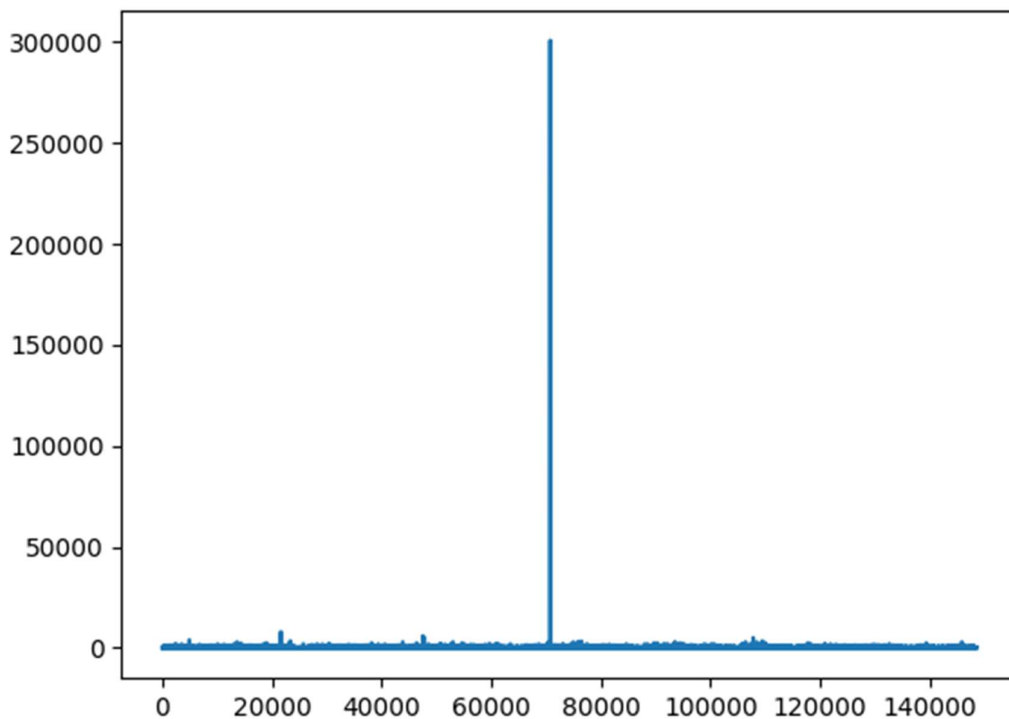


Figure 6.3: Variation in Meal Costs Across Restaurants

## 6.4 Statistical Distribution of Meal Cost for Two:

- **Purpose:** This line generates a box plot of the 'cost' feature.
- **Data:** It also uses the 'cost' column of the df DataFrame.
- **Visualization:** A box plot summarizes the distribution of the 'cost' variable by showing quartiles, median, and outliers. It helps in understanding the central tendency, spread, and skewness of the data for cost for two people.
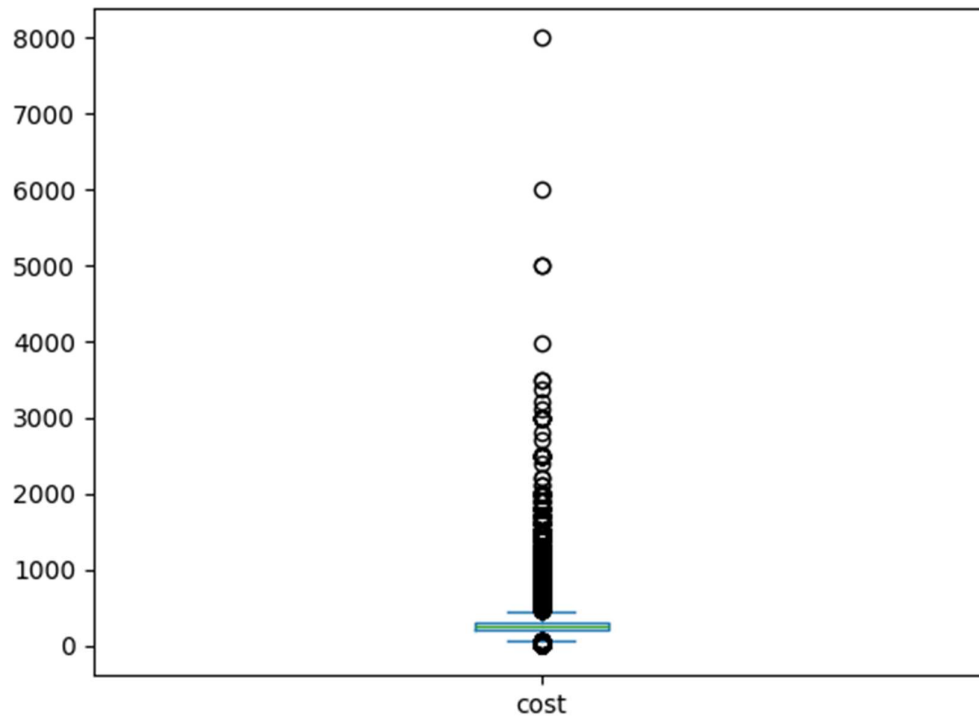


Figure 6.4: Statistical Distribution of Meal Cost for Two

# Chapter 7

# CONCLUSION

The analysis of Swiggy's restaurant data provides valuable insights into how the platform is performing and where opportunities lie. Major cities like Bangalore, Delhi, and Mumbai are hotspots for restaurant listings, while smaller cities like Pune and Chandigarh show promise for further growth. Most restaurants have ratings between 3.0 and 4.5, which indicates that customers are generally satisfied, though the few low-rated restaurants suggest areas where quality can be improved. When it comes to cost, mid-range restaurants priced between ₹300 and ₹800 are the most popular, offering a good balance of quality and affordability, while high-end restaurants with prices above ₹2,500 appeal to a smaller but highly satisfied audience. North Indian and Chinese cuisines remain customer favorites, but global cuisines like Italian and Mexican are becoming increasingly popular in metro cities. These insights suggest that Swiggy can grow by onboarding more restaurants in smaller cities, promoting top-rated eateries, addressing quality concerns, and catering to evolving customer tastes. By focusing on these strategies, Swiggy can continue to enhance the dining experience for its customers and strengthen its presence in the competitive food delivery market.

# Chapter 8

## REFERENCES

[1]      Swiggy Resturant Dataset (Kaggle)

[2]      https://pandas.pydata.org/

[3]      http://www.numpy.org

[4]      http://www.matplotlib.org

[5]      https://www.swiggy.com/