High Level Design (HLD)

Analyzing Swiggy : Bangalore delivery outlet data

Revision Number: 1.0

Last date of revision: 10/08/2024

Saurabh Kumar Kushwaha

**Document Version Control**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date Issued** | **Version** | **Description** | **Author** |
| **10th Aug 2024** | 1.0 | First Version of Complete HLD | Saurabh Kumar Kushwaha |

**Contents Page No**

Document Version Control……………………………………………………

Abstract………………………………………………………………………..

1 Introduction………………………………………………………………

1.1 Why this Low-Level Design Document?..........................................

1.2 Scope…………………………………………………………………….

1.3 Constraints……………………………………………………………….

2 Technical Specifications………………………………………………………………………

2.1 Swiggy Banglore Outlet Dataset………………………………………………

2.1.1 Swiggy Bangalore Outlet Dataset Overview………………………………

2.1.2 Overview of Variables……………………………………………….

3 Architecture…………………………………………………………………………….

3.1 Architecture Description……………………………………………………..

3.1.1 Data Description

3.1.2 Define the Use Cases

3.1.3 Import the Dataset

3.1.4 Exploratory Data Analysis (EDA)

3.1.5 Data Pre-processing, Data Cleaning & Imputation (Handling the Categorical & Numerical Variables)

3.1.6 Analyse the Data

3.1.7 Visualize & Share Meaningful Insights

4 Technology Stack

Abstract

## The "Analyze Swiggy" project is focused on exploring the Swiggy Bangalore Outlet Details dataset through an ETL (Extract, Transform, Load) and analysis approach. The main goal is to gain valuable insights from the data.

## To start, the data is extracted from a CSV file, and then transformed using techniques such as converting columns into suitable data types, creating new columns, and handling missing data. The transformed data is then saved into a new CSV file for further analysis.

## The project analyzes the transformed data using a range of visualizations and metrics. The key areas of interest include:

## 1. Distribution of restaurants based on cost categories.

## 2. Average cost per person at different locations.

## 3. Number of restaurants in each location.

## 4. Average rating for each cost category.

## 5. Distribution of ratings across restaurants.

## 6. Average cost for each cuisine.

## 7. Average rating for each cuisine.

## To achieve these insights, the project employs various tools, including pandas for data manipulation, seaborn, and matplotlib for data visualization.

## The functional architecture of the project comprises three main stages:

## 1. Extract: This stage involves fetching the data from the CSV file.

## 2. Transform: The data undergoes transformations using pandas functions to prepare it for analysis.

## 3. Load: The transformed data is then saved into a new CSV file and further analyzed using visualizations and metrics.

## A key focus of the project is to optimize performance. This is done by using appropriate data types and handling missing data to avoid errors during analysis.

## Throughout the project, Key Performance Indicators (KPIs) are used to measure and evaluate the effectiveness of the analysis. However, the specific KPIs used in this project are not mentioned in the provided text.

## Introduction

## Over the years, the online food ordering market has experienced remarkable growth, primarily due to the rising popularity of food delivery platforms on the internet. As technology continues to advance, more and more people are turning to their smartphones and mobile apps to conveniently order food online. As a consequence, the online food ordering industry has become intensely competitive, with big players like Swiggy, Zomato, Uber Eats, and others competing fiercely to gain a larger share of the market.

## In this case, the Swiggy Bangalore Outlet Details dataset offers valuable insights into the online food ordering market in Bangalore, India. It provides a wealth of information about various Swiggy outlets in the city, including their names, locations, types of cuisine served, average cost for two people, and ratings. By studying this dataset, we can uncover significant findings about the online food ordering scene in Bangalore. This analysis allows us to identify important metrics and factors that influence the market and also helps us understand how different attributes are related to each other.

## Purpose

The purpose of this project is to gain insights into the online food ordering market in Bangalore, India, using the Swiggy Bangalore Outlet Details dataset. By analyzing this dataset, we can identify key metrics and factors that are relevant to the market and understand the relationships between these factors. The insights gained from this analysis can be used by stakeholders such as online food ordering platforms, restaurants, and customers to make informed decisions and improve their operations.

## Scope

## The goal of this project is to analyze the Swiggy Bangalore Outlet dataset using data mining techniques. By extracting, transforming, and loading the data, we aim to uncover essential insights relevant to the online food ordering market in Bangalore. Our focus will be on finding answers to the following questions:

## 1. What is the average cost and average rating for each cuisine?

## 2. How does the cost category of an outlet influence its rating? Additionally, what is the distribution of restaurants across different cost categories?

## 3. What is the distribution of ratings among the restaurants in the dataset?

## 4. Can we determine the number of restaurants present in each location?

## Through this analysis, we aim to identify key metrics and factors that contribute to the success of restaurants in the online food ordering market in Bangalore.

General Description

**Product Perspective :**

The main objective of the Analyze Swiggy project is to provide valuable insights and analysis about the popular food delivery platform, Swiggy. With the increasing popularity of online food delivery services, it has become essential to understand how these platforms work and perform. Analyze Swiggy has been carefully designed to help different stakeholders, including restaurant owners, delivery partners, and customers, to better grasp Swiggy's business dynamics. By doing so, it empowers them to make informed decisions with confidence.

## Problem Statement:

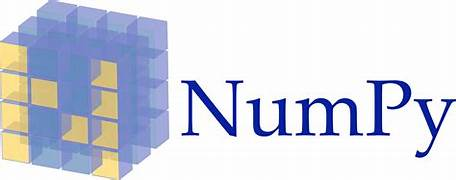
The online food ordering market encompasses a variety of food options, including meals prepared by restaurants, independent chefs, and groceries that can be ordered online and either picked up or delivered. The first-ever online food ordering service, World Wide Waiter (now known as Waiter.com), was established in 1995. Online food ordering involves placing orders through websites or other applications. The food offered can either be ready-to-eat meals or ingredients for cooking that haven't been specifically prepared for immediate consumption.

Perform ETL (Extract-Transform-Load) on the dataset to gather relevant information from this extensive data. This process is a form of data mining.

By mining this data, we can extract several types of information and key metrics. Analyzing the relationships between different attributes can reveal meaningful insights. For the research, it would be essential to investigate and uncover on own findings.

**Tools used**

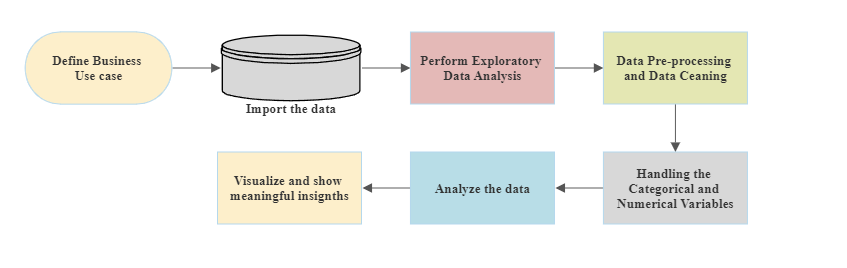
The entire framework is constructed using tools and libraries like Python 3, Jupyter Notebook, Numpy, Pandas, Matplotlib, and Seaborn.



**Design Details**

**3.1 Functional Architecture**

****

**Error Handling:**

The project has been designed to ensure error-free execution by thoroughly testing and running the entire script multiple times throughout the process flow.

**Conclusion**:

The examination of the Swiggy Bangalore Outlet Details dataset has yielded numerous fascinating findings regarding the restaurants and their attributes in Bangalore.

The dataset encompasses details about different restaurants in Bangalore, including cuisine type, location, rating, and cost for two people. Initially, the dataset was extracted and processed, converting the cost for two and rating columns to numerical data, removing rows with missing information, creating new columns for average cost per person and cost category, and standardizing column names by eliminating special characters and spaces.

After processing, the dataset was loaded and analyzed using various visualizations and statistical techniques. A basic overview of the data revealed 1448 restaurants, representing 21 distinct cuisines across 92 different locations. The first five rows of the data displayed essential information such as the restaurant's name, location, cuisine, cost for two, rating, and other attributes.

The exploration of unique values for each attribute indicated 21 diverse cuisines, with North Indian cuisine being the most prevalent in the dataset. The cost for two ranged from 50 to 6000, with an average cost of 682.48. The rating ranged from 2.0 to 4.9, with an average rating of 3.5.

Analyzing the average cost for each cuisine, it was found that Chinese cuisine had the highest average cost for two, followed by Thai and Italian cuisines. Conversely, Street Food had the lowest average cost for two. In terms of average ratings, Mediterranean cuisine received the highest average rating, followed by European and Chinese cuisines, while South Indian cuisine had the lowest average rating.

A heatmap illustrating the average cost and rating for each cuisine showed that most cuisines had higher ratings for restaurants with higher costs. However, there were exceptions, like Street Food, which had higher ratings for low-cost restaurants. A pie chart depicting the distribution of restaurants by cost category revealed that the majority fell under the medium-cost category, with 56.4% of restaurants having an average cost per person between 200 and 500.

The rating distribution revealed that most restaurants received ratings between 3.5 and 4.0. The distribution of restaurants by location demonstrated that Koramangala 5th Block had the highest number of restaurants, followed by BTM and Indiranagar. The average cost per person by location revealed that Jayanagar had the highest average cost per person, followed by Indiranagar and MG Road.

Lastly, the analysis revealed that high-cost restaurants generally had higher average ratings compared to low-cost restaurants, with the median rating for high-cost restaurants being 3.8 and the median rating for low-cost restaurants being 3.5.

Overall, the examination of the Swiggy Bangalore Outlet Details dataset provides valuable insights into the diverse restaurant landscape of Bangalore. The data highlights the prevalence of North Indian cuisine and the predominance of restaurants falling under the medium-cost category, with average costs per person ranging between 200 and 500. Additionally, the analysis establishes a correlation between restaurant costs and ratings, with most cuisines receiving higher ratings for higher-cost establishments.

**References:**

1. <https://en.wikipedia.org/wiki/Swiggy>
2. <https://medium.com/@triveni.mishra1921/an-interactive-analysis-of-swiggy-bangalore-data-d9c292947a4c>
3. https://techstory.in/everything-you-need-to-know-about-swiggy/