Architecture

# (ANALYZING SWIGGY)

|  |  |
| --- | --- |
| Written By / Author | 1.Krishnachaitanya Pabbu  2. Gangireddy Madhava Reddy |
| Document Version | V1.0 |
| Last Revised Date | 23/02/2023 |

**Document Version Control:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Version** | **Author** | **Comments** |
| **23/02/2023** | V1.0 | Krishnachaitanya Pabbu | First Draft |
|  |  |  |  |

**Approval Status:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Version** | **Review Data** | **Reviewed By** | **Approved By** | **Comments** |
| V1.0 |  |  |  |  |

**Contents**

## Document Version Control 2

1. **Introduction** 4
   1. Why this Architecture design document? 4
   2. Scope 4
2. **Architecture** 5
   1. Architecture Description 5

2.1.1 Data Description 5

2.1.2 Define the Use Cases 5

2.1.3 Extraction of Dataset 5

2.1.4 Extract Transform Load (ETL) 6

2.1.5 Data Pre-processing, Data Cleaning & Imputation (Handling the Categorical &

Numerical Variables) 6

## 2.1.6 Analyse the Data 7

## 2.1.7 Visualize & Share Meaningful Insights 7

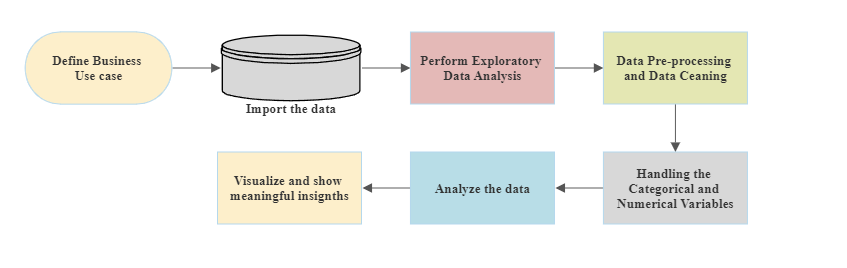
1. **Introduction**
   1. Why this Architecture design document?

This document aims to present a comprehensive architectural blueprint for the Analyzing Swiggy Project, delving into every aspect of our design. It will delve into the project's origins and crucial functional demands from an architectural perspective. The primary goal of this paper is to guide the development group in establishing the overarching structure of the system.

* 1. Scope

The Architecture Design Document (ADD) outlines a systematic approach to architecture design, progressing through a structured series of steps. This methodology is applicable for shaping data frameworks, essential software architecture, coding elements, and ultimately, optimizing algorithms for performance. In essence, the guiding principles of design are initially established during requirement analysis and subsequently honed throughout the process of architectural development.

1. **Architecture**



* 1. Architecture Description –

## 2.1.1 Data Description –

## In the previous instance, within our dataset pertaining to the Swiggy Bangalore Outlet, we observe a collection of approximately 118 entries, each comprising 5 distinct attributes. These attributes are classified into 2 continuous characteristics and 3 categorical attributes. The data is presented in the format of Comma Separated Values (.csv).

## 2.1.2 Define the Use Cases –

## In this phase, utilizing the provided dataset and identified business challenges, we have outlined multiple scenarios for conducting analyses. These scenarios aim to extract crucial observations from the data, which will serve as the foundation for making informed business choices. Additionally, this process facilitates not just the comprehension of significant attribute interconnections, but also empowers us to independently explore and unveil our own discoveries.

## 2.1.3 Extraction of Dataset –

Given that we've been provided with the dataset in the structure of a CSV (Comma Separated Value) file, we have the option to utilize the Pandas read\_csv( ) function to bring in the data.



## 2.1.4 Extract Transform Load (ETL) –

• "Extract Transform Load" (ETL) constitutes a crucial phase known as "Data Exploration" within the Data Analysis Process. During this stage, various methodologies are applied to gain a deeper comprehension of the employed dataset.

• Grasping the nature of the dataset encompasses a range of aspects, which include but are not confined to the following:

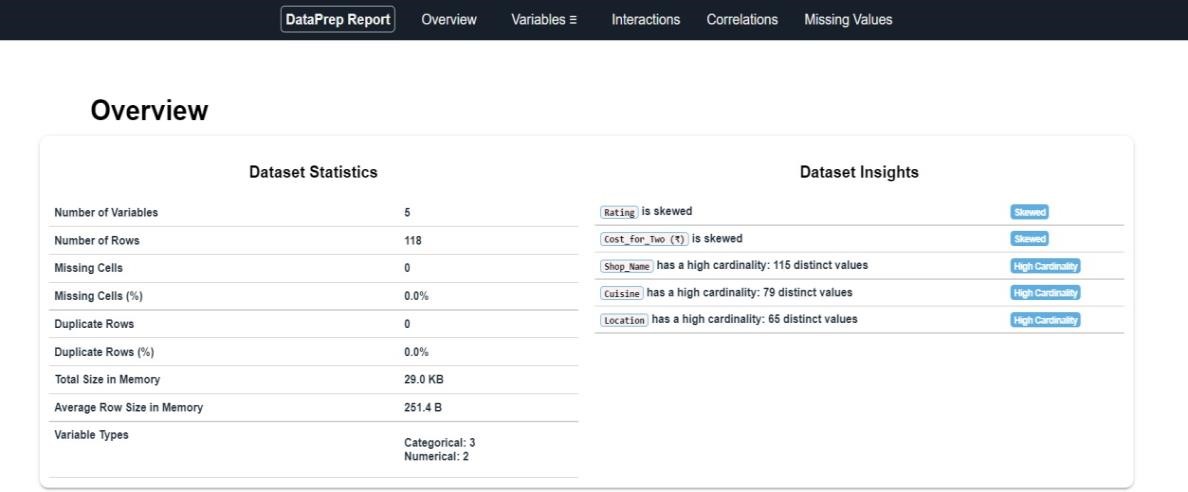
* Extracting significant "Variables."
* Detecting "Outliers," "Missing Values," or instances of "Human Error."
* Uncovering the interplay between different variables.
* Ultimately, enhancing insights into the dataset while minimizing the potential for subsequent errors.

• In simpler terms, this process facilitates a more profound grasp of both the "Variables" and the "Relationships" they share.

• To automate our Exploratory Data Analysis (EDA), we leverage the dataprep module.

• The module provides the ensuing information:

* Overview: Identifying column types within a Data Frame.
* Variables: Detailing variable attributes, unique values, distinct count, and instances of missing values.
* Quartile statistics, encompassing minimum and maximum values, Q1, median, Q3, range, and interquartile range.
* Descriptive statistics, such as mean, mode, standard deviation, sum, median absolute deviation, coefficient of variation, kurtosis, and skewness.
* Correlations: Signaling highly correlated variables, along with Spearman, Pearson, and Kendall matrices.
* Handling Missing Values: Visual representations through Bar Charts, Heatmaps, and a spectrum of missing value occurrences.



## 2.1.5 Data Pre-processing, Data Cleaning & Imputation (Handling the Categorical & Numerical Variables) –

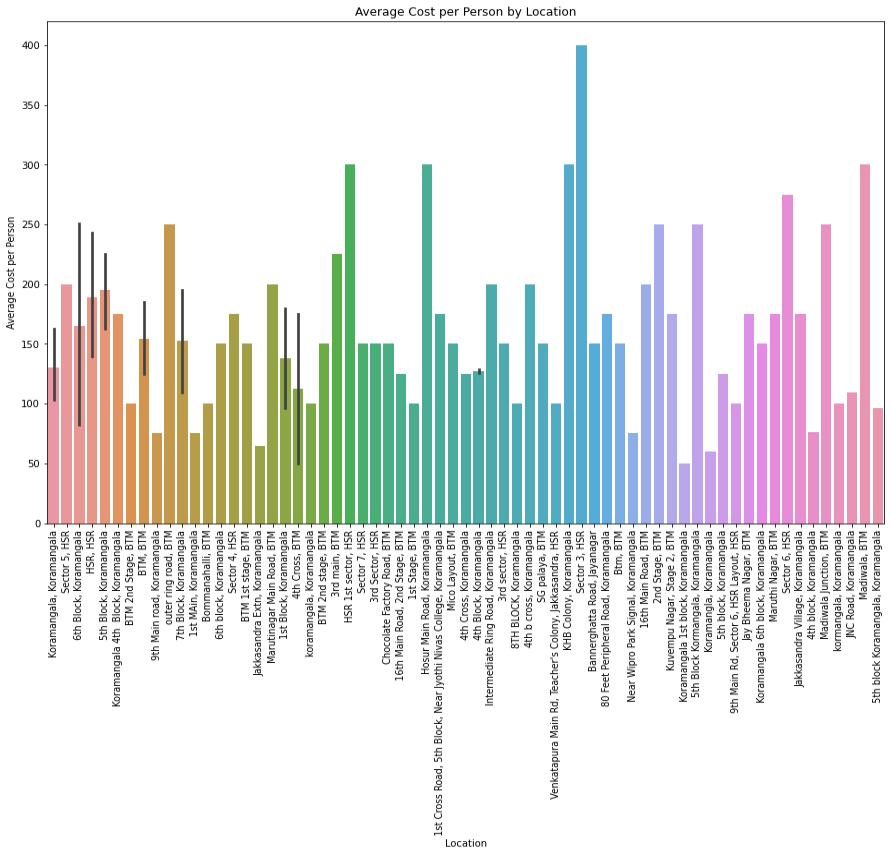
## Data preprocessing involves the transformation of raw data into a usable format for analysis. This entails extensive data cleansing and addressing missing values through suitable imputation methods, considering the categorical or numerical nature of variables. In the context of this undertaking, we performed replacement or imputation of absent data points using mean, median, or mode, based on variable characteristics. Furthermore, we excluded columns that do not contribute to our analytical objectives.

## 2.1.6 Analyze the Data –

## After completing the initial pre-processing phase, we are ready to proceed with our substantive examination. Here, we craft code segments and establish logical procedures to format our data in alignment with the specified intended applications.

## 2.1.7 Visualize & Share Meaningful Insights –

At last, the moment has arrived to transform our data into a visual portrayal of information. In essence, data visualization involves the conversion of extensive data sets and measurements into various visual forms like Bar Plots, Pie Charts, Heat maps, Box Plots, Scatter Plots, and numerous others. This resulting visual depiction of data simplifies the process of recognizing and disseminating valuable observations regarding the data's content. Behold, here is a captivating preview of one of our visual creations –



All those different analyses help to make better business decisions and help analyse customer trends and satisfaction, which can lead to new and better products and services.