Architecture Design

House Price Prediction

Written By	Gouthamram Puranik
Document Version	0.1
Last Revised Date	25/06/2023

DOCUMENT CONTROL

Change Record:

VERSION	DATE	AUTHOR	COMMENTS
0.1	25- June - 2023	Gouthamram Puranik	Complete architecture

Approval Status:

VERSION	REVIEW DATE	REVIEWED BY		APPROVED BY	COMMENTS

Contents

1. Introduction

1.1 What is Architecture design document?

Any software needs the architectural design to represents the design of software. IEEE defines architectural design as “the process of defining a collection of hardware and software components and their interfaces to establish the framework for the development of a computer system.” The software that is built for computer-based systems can exhibit one of these many architectures.

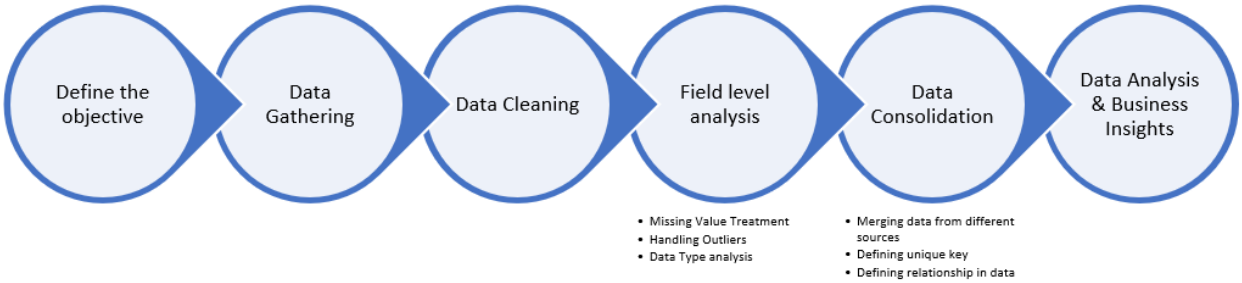
Each style will describe a system category that consists of :

- A set of components (eg: a database, computational modules) that will perform a function required by the system.
- The set of connectors will help in coordination, communication, and cooperation between the components.
- Conditions that how components can be integrated to form the system.
- Semantic models that help the designer to understand the overall properties of the system.

1.2 Scope

Architecture Design Document (ADD) is an architecture design process that follows a step-by-step refinement process. The process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the design principles may be defined during requirement analysis and then refined during architectural design work.

2. Architecture



2.1. Data description

	Shop_Name	Cuisine	Location	Rating	Cost_for_Two
count	118	118	118	118	118
unique	115	79	65	13	30
top	La Pino'z Pizza	North Indian	BTM, BTM	4.1	₹ 300
freq	2	12	13	30	16

In the Swiggy Bangalore Outlet dataset, we have observed a total of 118 records, each consisting of 5 distinct features. These features can be categorized into two types: continuous and categorical. Specifically, we have 2 continuous features and 3 categorical features. The dataset is provided in the widely used Comma Separated Value (.csv) format .

2.2. Data Import

We can import the data using Pandas read.csv() function as the dataset is in the form of Comma separated (.csv) format.

2.3. Extract, Transform, Load (ETL)

- ETL techniques are employed to gain a deeper understanding of the dataset being analyzed.
- The goal of ETL includes extracting important variables, identifying outliers, missing values, or human errors, and comprehending relationships between variables.
- ETL helps maximize insights derived from the dataset while minimizing potential errors that may arise later in the analysis process.
- In essence, ETL provides a better understanding of variables and their relationships within the dataset.

2.4. Exploratory data analysis(EDA)

Exploratory data analysis (EDA) involves thoroughly analyzing the data to extract meaningful insights, facilitating a comprehensive understanding of the dataset and providing answers to key questions. It entails a systematic exploration of the data through various statistical techniques, visualizations, and descriptive summaries. By delving into the data's characteristics, patterns, and relationships, EDA enables the identification of important trends, anomalies, and correlations. Through this process, we gain valuable insights that contribute to our understanding of the data and aid in addressing specific questions or objectives.

2.5. Data Visualization

Data visualization is the process of converting complex datasets and metrics into visual representations, such as charts, graphs, and diagrams. It encompasses various types of visualizations, including bar plots, pie charts, box plots, scatter plots, and more. These visual representations serve the purpose of simplifying and conveying information in a visually appealing and intuitive manner. By using visuals, data visualization aids in the identification and communication of insights, patterns, and relationships present in the data, enabling effective decision-making and knowledge sharing.