

# Architecture Design

## House Price Prediction

Written By	Sahil & Siddhi
Document Version	
Last Revised Date	

## DOCUMENT CONTROL

### Change Record:

VERSION	DATE	AUTHOR	COMMENTS
0.1	29-10-2022	Sahil	Introduction and architecture defined
0.2	30-10-2022	Siddhi	Architecture & Architecture description appended and updated.

### Approval Status:

VERSION	REVIEW DATE	REVIEWED BY		APPROVED BY	COMMENTS

# Contents

1.

Introduction.....

..... 2

1.1 What is Architecture Design Document?

..... 2

1.2 Scope

.....

..... 3

2. Architecture

.....

..... 3

2.1 Architecture Description

.....4

## 4 ARCHITECTURE DESIGN

HOUSE PRICE PREDICTION 3



## 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

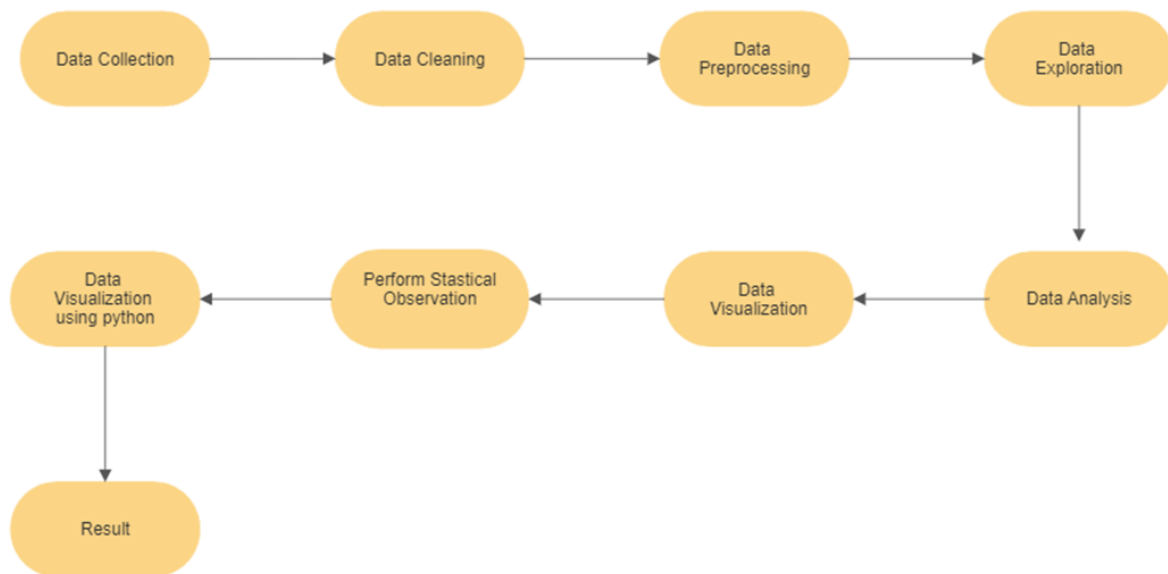
architectures.

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.



## 2.1 Architecture Description

### Data Collection -

As we have seen earlier, in our Dataset preview, we have around hundreds of records with 14 different features. Features are distributed as Continuous features, Categorical features and Geographical feature in our Sales Dataset. The "100 Sales Records.csv" dataset was given in the form of Comma Separated Value (.csv) format and we load this dataset in Python.

### Data Cleaning -

At this stage, based on the given knowledge unit and business problems we have formed the several Use examples to act the observations on and this will definitely help out get the key knowledge from this facts based on which business decisions will be taken. in addition, It helps in not only getting through knowledge the purposeful relationships between given properties, but it also lets us to do our own operation of making observations and come-up with our decisions in law. happily, on condition that knowledge unit has no missing values, for this reason we proceeded farther.

### Import the Dataset -

In this process, we changed the data into highly understandable format by applying Python's Pandas and Numpy libraries. We eliminated the unnecessary rows and columns. We changed categorical data into number-based dummy numbers that change/things that change in order to make it more good for related to studying numbers/ number-based analysis which makes/gives meaning to the data up to higher extent. We drew attention to upon removing unnecessary thing in data and made it good for doing exploratory data analysis.

## **Exploratory Data Analysis (EDA)-**

It is a "Data Exploration" step in the Data Analysis Process, where several techniques are used to better understand the dataset being used. Understanding the Dataset can refer to a number of things including but not limited to

Extracting Important "Variables

Identifying "Outliers

Missing Values", or "Human Error".

Understanding the Relationships between variables. Ultimately, maximizing our insights of a dataset and minimizing potential "Error" that may occur later in the process.

In other words, it will give you a better Understanding of the "Variables" and the "Relationships" between them. Here, we make use of dataprep module to automate our EDA process. It provides the following information:

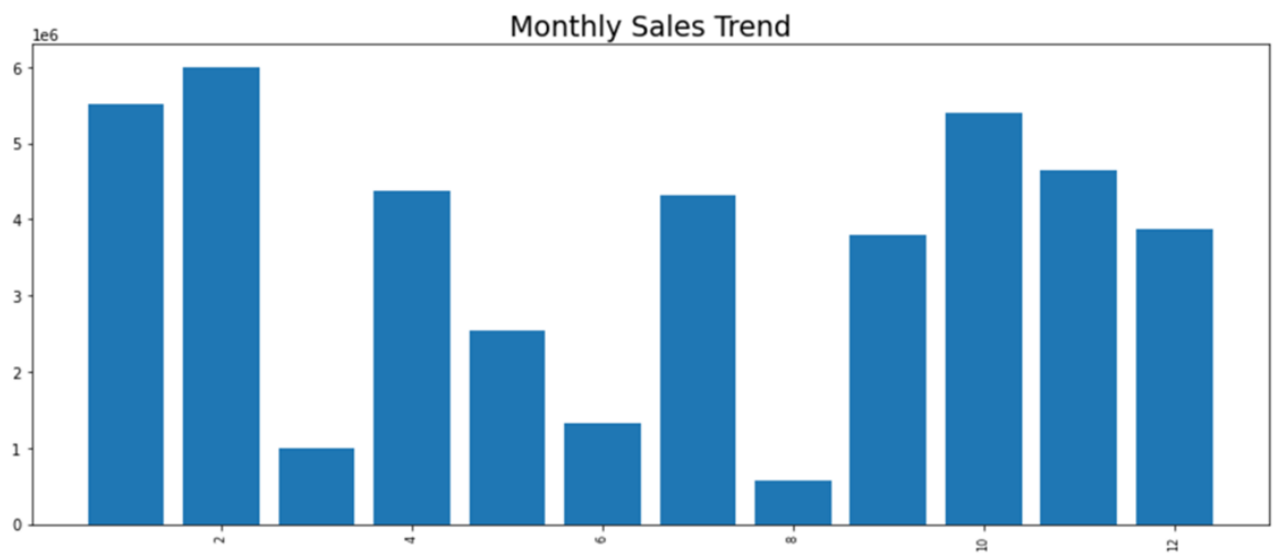
Overview: detect the types of columns in a DataFrame. Variables: variable type, unique values, distinct count, missing values Quartile statistics like minimum value, Q1, median, Q3, maximum, range, interquartile range Descriptive statistics like mean, mode, standard deviation, sum, median absolute deviation, coefficient of variation, kurtosis, skewness. Correlations: highlighting of highly correlated variables, Spearman, Pearson and Kendall matrices Missing Values: Bar Chart, Heatmap and spectrum of missing values.

## **Data Visualization Using Python-**

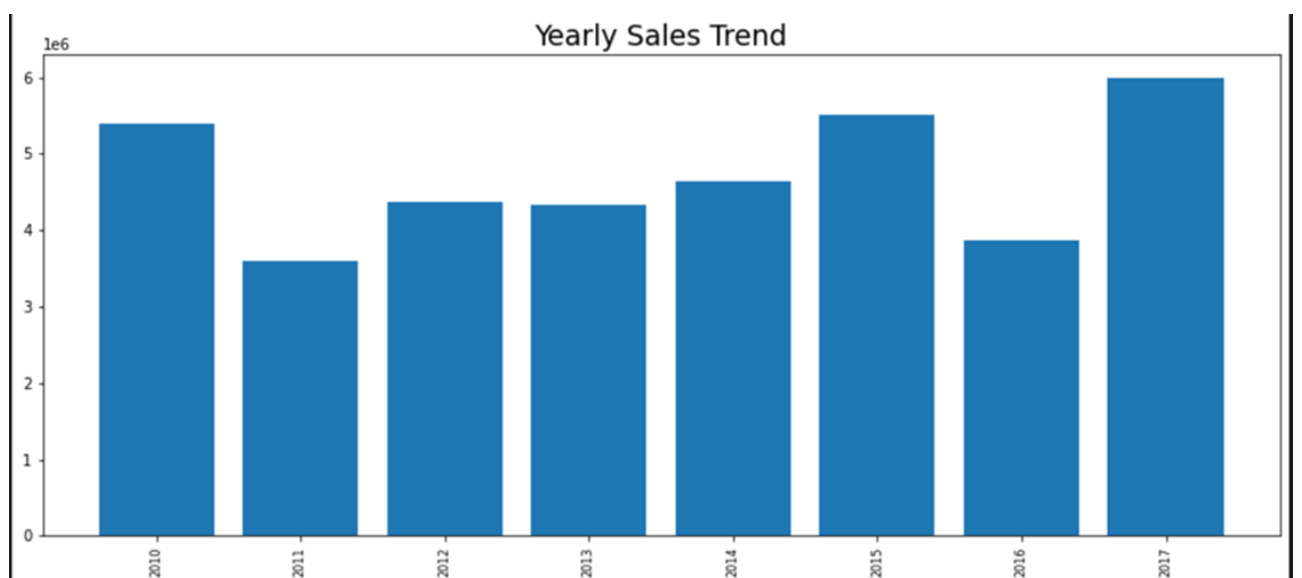
Initially, we extracted data from the datasets into the Python. We prepared various charts and plots based on meaningful data. We established several mathematical relations between the numerical attributes present in the data. We prepared various visuals for different sections of the data and filtered the charts according to various parameters to make it more user interactive and user friendly.. The resulting visual representation of data makes it easier to identify and share insights about the information represented in the data.

## **Result:**

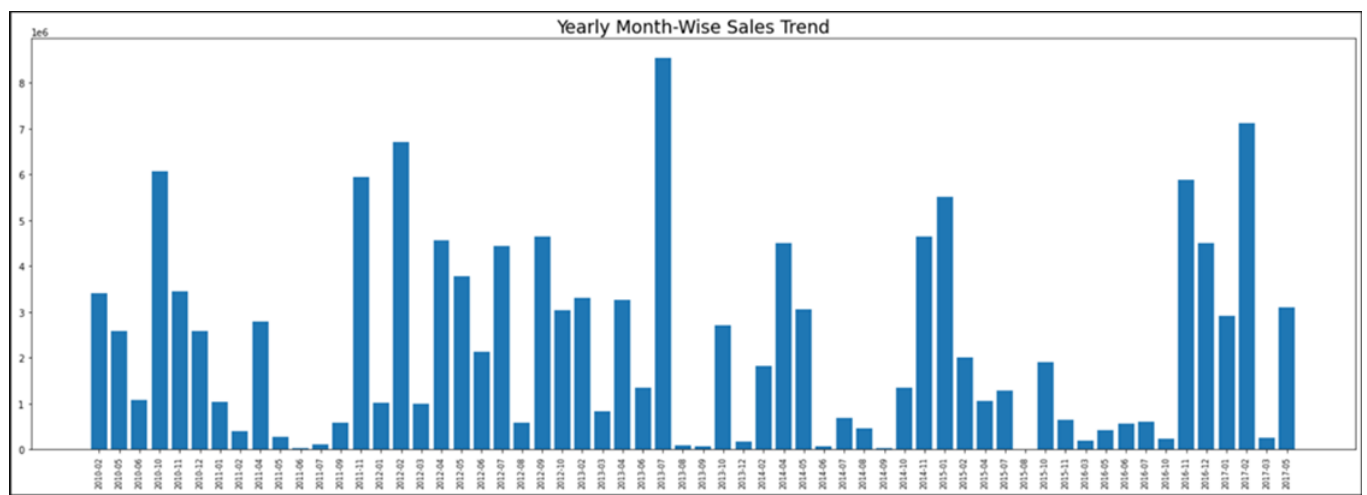
### **1. Monthly Sales Trends :-**



## 2. Yearly Sales Trend



## 3. Yearly Month-Wise Sales Trend



#### 4. Total Revenue

Total_Revenue	
Item_Type	
Cosmetics	36601509.60
Office Supplies	30585380.07
Household	29889712.29
Baby Food	10350327.60
Clothes	7787292.80
Cereal	5322898.90
Meat	4503675.75
Personal Care	3980904.84
Vegetables	3089057.06
Beverages	2690794.60

