Low Level Design

Amazon Sales Analysis

Written By	Shubham tembhurne
Document Version	0.3
Last Revised Date	13/05/23

DOCUMENT CONTROL

Change Record:

VERSION	DATE	AUTHOR	COMMENTS
0.1	13- May - 2023	Shubham temburne	Introduction and architecture defined
0.2	13 - May - 2023	Shubham tembhurne	Architecture & Architecture description appended and updated.

Contents

1. Introduction

1.1 What is Low-Level design document?

The goal of the LDD or Low-level design document (LLDD) is to give the internal logic design of the actual program code for the House Price Prediction dashboard. LDD describes the class diagrams with the methods and relations between classes and programs specs. It describes the modules so that the programmer can directly code the program from the document.

1.2 Scope

Low-level design (LLD) is a component-level 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 data organization may be defined during requirement analysis and then refined during data design work.

1.3 Project Introduction

The analysis aims to gain insights and extract meaningful information from the data to understand various aspects of the sales performance on Amazon. By conducting a comprehensive analysis of the given Amazon sales dataset, this project aims to provide valuable insights that can guide decision-making and optimize sales strategies for improved performance on the platform.

2. Problem Statement

This dataset consists more than 1000 of real products with their identification number listed in the Amazon marketplace specifically from the region India. I noticed the region due to the currency used in the dataset is Rupee India. My objective is to clean and prepare the data due to the raw data being very unorganized. I will then move on to finding insights about the data and try to elaborate in the form of visualization.

3. Dataset Information

PRODUCT_ID: A product ID is a unique identifier assigned to a specific product or item for the purpose of tracking and managing sales transactions. It helps to distinguish one product from another and enables efficient inventory management, order processing, and sales analysis.

PRODUCT_NAME: A product name refers to the specific name or designation given to a product or item being sold. It is used to identify and describe the product to customers, sales representatives, and other stakeholders involved in the sales process. The product name plays a crucial role in marketing, branding, and communicating the features and benefits of the product to potential buyers

CATEGORY: A category refers to a grouping or classification system used to categorize products based on shared characteristics, attributes, or functions. Categories help organize and manage inventory, facilitate browsing and search for customers, and provide insights into sales performance by product type. Common examples of sales categories include electronics, clothing, home appliances, beauty products, and automotive. Categorizing products allows for better organization, targeted marketing, and more effective sales analysis.

DISCOUNTED_PRICE: A discounted price refers to a reduced price offered to customers as an incentive or promotion. It is typically lower than the original or regular price of a product. Discounts can be in the form of a percentage off, a fixed amount off, or a combination of both. The discounted price is

communicated to customers to encourage them to make a purchase and create a sense of value or savings. Discounted prices can be temporary, such as during a sale or promotion, or they can be available for specific customer segments, such as loyal customers or bulk buyers.

ACTUAL_PRICE: the actual price refers to the original or regular price of a product before any discounts or promotional offers are applied. It represents the standard price at which the product is typically sold. The actual price is often used as a reference point to calculate discounts or determine the savings a customer can receive through promotional pricing. It is important for customers to know the actual price to understand the value they are receiving when comparing it to discounted prices or competing offers.

DISCOUNT_PERCENTAGE: The discount percentage in a sale refers to the percentage by which the original price of a product is reduced to calculate the discounted price.

RATING: Ratings refer to the numerical or qualitative assessments provided by customers or experts to evaluate the quality, performance, or satisfaction level of a product or service.

RATING_COUNT: Rating count in sales refers to the total number of ratings or reviews received for a product or service

DIFFERENCE_PRICE: The difference price refers to the variance or discrepancy between two prices.

1. Architecture

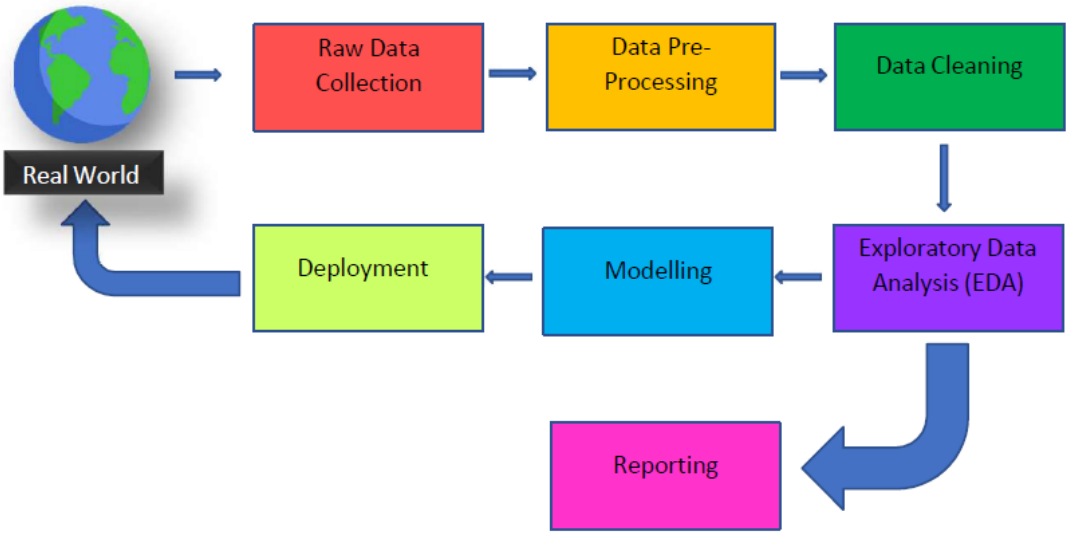


Tableau Server Architecture

Tableau has a highly scalable, n-tier client-server architecture that serves mobile clients, web clients and desktop-installed software. Tableau Server architecture supports fast and flexible deployments.

The following diagram shows Tableau Server's architecture:

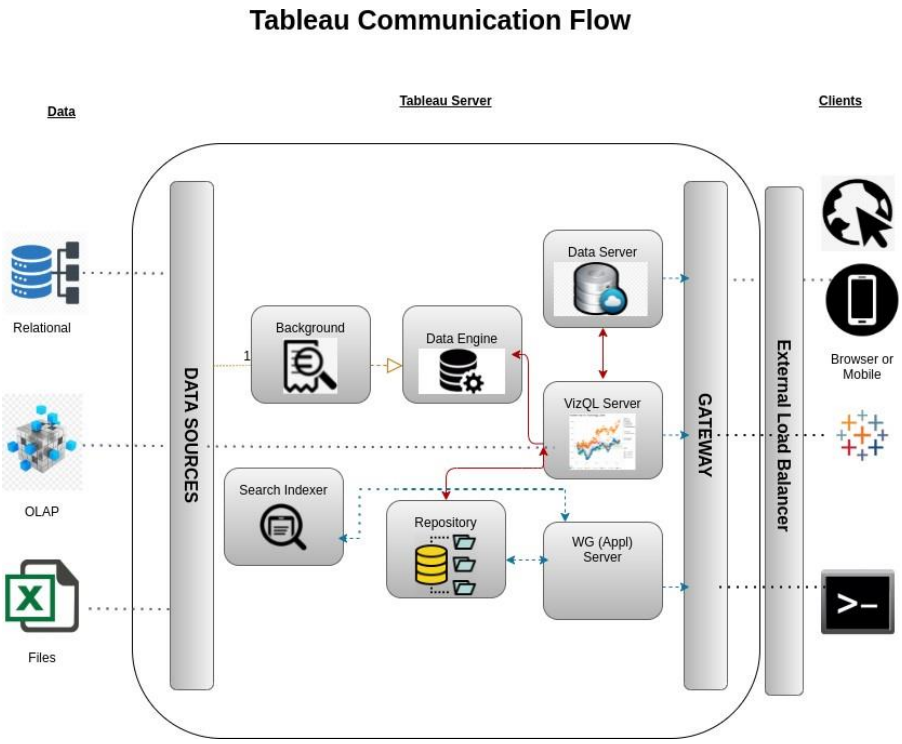


Tableau Server is internally managed by the multiple server processes.

1). Gateway/Load Balancer

It acts as an Entry gate to the Tableau Server and also balances the load to the Server if multipleProcesses are configured.

2) Application Server:-

Application Server processes (wgserver.exe) handle browsing and permissions for the Tableau Server web and mobile interfaces. When a user opens a view in a client device, that user starts a session on Tableau Server. This means that an Application Server thread starts and checks the permissions for that user and that view.

3) Repository:-

Tableau Server Repository is a PostgreSQL database that stores server data. This data includes information about Tableau Server users, groups and group assignments, permissions, projects, data sources, and extract metadata and refresh information.

4) VIZQL Server:-

Once a view is opened, the client sends a request to the VizQL process (vizqlserver.exe). The VizQL process then sends queries directly to the data source, returning a result set that is rendered as images and presented to the user. Each VizQL Server has its own cache that can be shared across multiple users

5) Data Engine: -

It Stores data extracts and answers queries.

6) Backgrounder: -

The backgrounder Executes server tasks which includes refreshes scheduled extracts, tasks initiated from tabcmd and manages other background tasks.

7) Data Server: -

Data Server Manages connections to Tableau Server data sources

It also maintains metadata from Tableau Desktop, such as calculations, definitions, and groups.

3. Architecture Description

1. Raw Data Collection

The Dataset was taken from Kaggle Provided.

2. Data Pre-Processing

Before building any model, it is crucial to perform data pre-processing to feed the correct data to the model to learn and predict. Model performance depends on the quality of data fed to the model to train.
This Process includes)
Handling Null/Missing Values
b) Handling Skewed Data
c) Outliers Detection and Removal

3. Data Cleaning

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.
a) Remove duplicate or irrelevant observations
b) Filter unwanted outliers
c) Renaming required attributes

4. Reporting

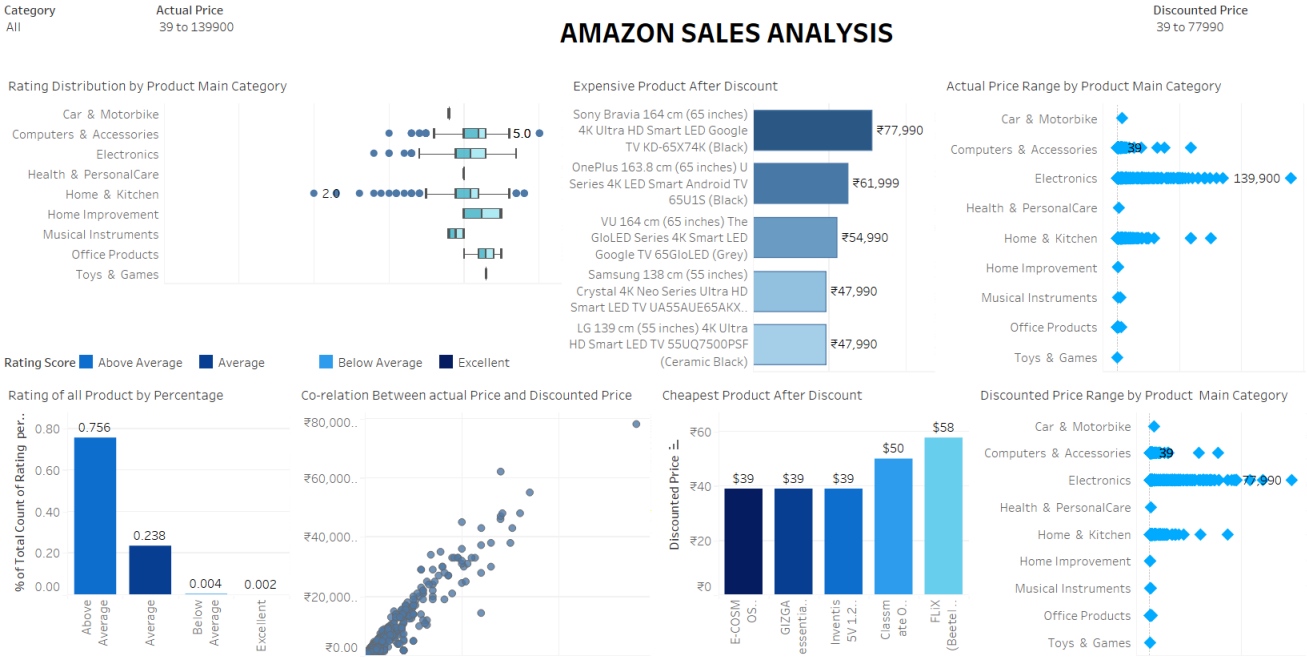
Reporting is a most important and underrated skill of a data analytics field. Because being a Data Analyst you should be good in the easy and self-explanatory report because your model will be used by many stakeholders who are not from a technical background. a) High-Level Design Document (HLD) b) Low-Level Design Document (LLD) c) Architecture d) Wireframe e) Detailed Project Report f) PowerPoint Presentation

5. Modelling Data

Modelling is the process of analyzing the data objects and their relationship to the other objects. It is used to analyze the data requirements that are required for the business processes. The data models are created for the data to be stored in a database. The Data Model's main focus is on what data is needed and how we have to organize data rather than what operations we have to perform.

6. Deployment

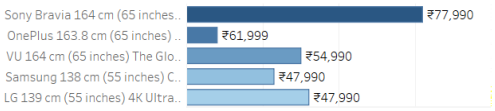
We created a Tableau Dashboard



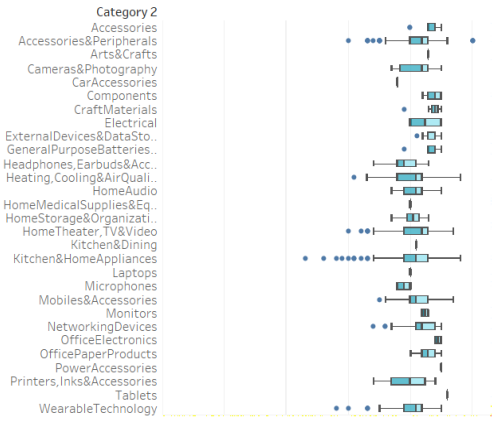
AMAZON SALES ANALYSIS

Rating 2.04 to 5

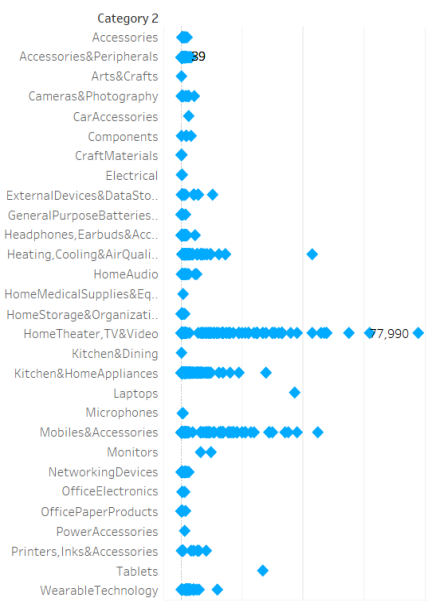
Largest Price difference due to discount in Product



Rating Distribution by Product Sub Category



Discounted Price Range by Product Sub Category



Actual Price Range by Product Sub Category