Low level document
Analysing Amazon Sales data

Revision Number: 1.0

Last date of revision: 30/10/2023

Document Version control

Date issued	Version	Description	Author
30/10/2023	1.0	First Version of LLD	Deepak V

CONTENT

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.

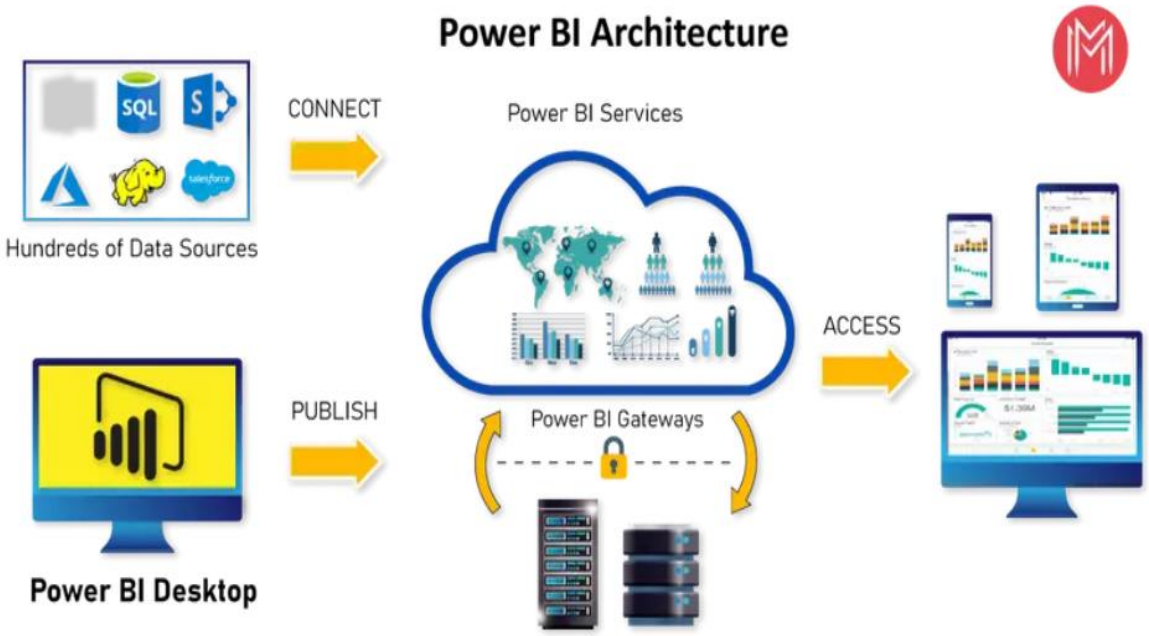
1.2 Scope

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.

2. Architecture

Power BI is a business solution that combines multiple technologies to work together as a system. Microsoft's Power BI technology consists of a suite of components that deliver exceptional business intelligence solutions, such as:

- Power Query (for data mash-up and transformation)
- Power BI Desktop (a companion development tool)
- Power BI Mobile (for Android, iOS, Windows phones)
- Power Pivot (for in-memory tabular data modelling)
- Power View (for viewing data visualizations)
- Power Map (for visualizing 3D geo-spatial data)
- Power Q&A (for natural language Q&A).



2.1. Data Sources

Power BI's extensive data source support is a key feature. You can import data from local files, cloud-based online sources, or connect directly to live data sources. There is a 1 GB limit on imports from on-premises or online services. Some common Power BI data sources include:
a) Excel
b) Text/CSV
c) XML
d) JSON
e) Oracle Database
f) IBM DB2 Database
g) MySQL Database
h) Postgres SQL Database
i) Sybase Database
j) Teradata Database
k) SAP HANA Database
l) SAP Business Warehouse server
m) Amazon Redshift
n) Impala o) Google Big Query (Beta)
p) Azure SQL Database
q) Salesforce Reports
r) Google Analytics
s) Facebook
t) GitHub.

2.2 Power BI desktop

Power BI Desktop allows users to connect to a wide variety of data sources, both on-premises and in the cloud. Once connected, users can clean and transform their data, and create reports and dashboards that visualize their data in a meaningful way. Power BI Desktop also includes a number of advanced features, such as the ability to create custom visualizations, develop DAX expressions, and publish reports to the Power BI Service.

2.3 Power BI service

Power BI Service is a cloud-based platform for storing, sharing, and collaborating on Power BI content. It is the central hub for accessing Power BI content from anywhere, and it provides a number of features for managing and distributing Power BI content, such as

workspaces, apps, datasets, reports, and dashboards. This is a powerful tool for sharing and collaborating on BI content, and it plays a vital role in the data ingestion, transformation, modelling, visualization, and publishing process.

2.4 Power BI report server

Power BI Report Server is a self-hosted solution for publishing and managing Power BI reports on-premises. It is suitable for organizations that need to control their own data and environment, or that have compliance requirements that prevent them from using the cloud-based Power BI Service. It also provides a number of features that make it easy to manage and distribute Power BI reports, including centralized management and scalability. The only difference between these two is that Power BI Report Server is an on-premise platform.

2.5 Power BI Gateway

Power BI Gateways are used in organizations to securely extract on-premises data to Power BI for analysis and reporting. This component is used to connect and access on-premise data in secured networks. Power BI Gateways are generally used in organizations where data is kept in security and watch.

2.6 Power BI Mobile

Power BI Mobile is a native Power BI app for iOS, Android, and Windows mobile devices. It allows users to view reports and dashboards.

3. Architecture Design

3.1 Data Description

Amazon Sales Data Features

1. Cust key – It is a Unique Id used to define a customer.
2. Date key – It is the date on which transaction took place.
3. Discount amount – It is the difference between Sales amount based on list price and Sales amount.
4. Invoice Date – It is the date on Which the Ordered delivered and invoice created.
5. Invoice Number – It is a Unique number generated by the system after making of invoice
6. Item Class – It is the class of the Item.

7. Item Number – It is a Unique number used to define an item.
8. Item – It is the name of the item for which transaction took place.
9. Line Number – It is the number of line from which it is ordered.
10. List Price – It is the price quoted by the manufacturer.
11. Order Number – It is the Unique Number for the particular order.
12. Promised delivery date – It is the date provided on which delivery is expected.
13. Sales Amount – It is the Product of Sales Price and Quantity.
14. Sales amount based on List Price – It is the product of List price and Quantity.
15. Sales Cost amount – It is the amount caused for making sales of the item.
16. Sales Margin amount - It is the difference between Sales amount and Sales cost amount.
17. Sales Price – It is the price at which Item is Sold.
18. Sales Quantity – It is the quantity of the ordered item.
19. Sales Rep. – It is the unique number or Id of the sales representative.
20. U/M – It Is the Unit of measurement for particular item.

3.2 Raw Data Collection:

The Dataset was taken from iNeuron's Provided Project Description Document.

3.3 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 includesa) Handling Null/Missing Values b) Handling Skewed Data c) Outliers Detection and Removal

3.4 Data Cleaning

Data cleaning is the process of removing incorrect, incorrectly formatted, duplicate, or incomplete data within a dataset.

a) Remove duplicate
b) Unwanted Observations
c) Renaming required attributes.

3.5 Exploratory Data Analysis (EDA)

Exploratory Data Analysis refers to the critical process of performing initial investigations on data to discover patterns, spot anomalies, test hypothesis and check assumptions with the help of summary statistics and graphical representations.

3.6. Modelling

Data Modelling is the process of analysing the data objects and their relationship to the other objects. It is used to analyse the data requirements that are required for the business processes. The data models are created to store the data 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.

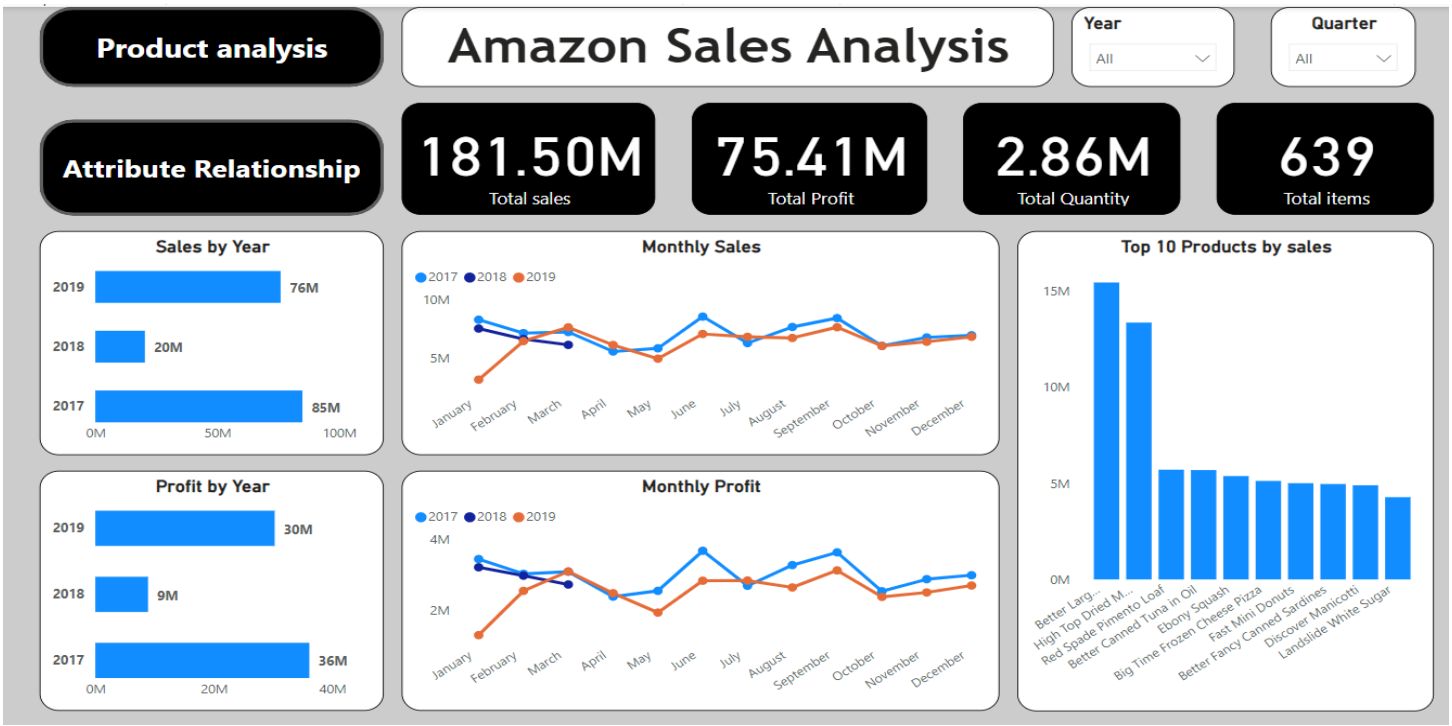
3.7 Reporting

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

3.8 Deployment

Creating of Reports in Power BI.



4. Unit Test Case

Test	Description
Slicer of year and quarter	When clicked on the slicer, a dropdown should occur which shows different years and quarters.
Charts	All the charts are working and has no errors.
Page Buttons	All page buttons are working and has no errors.