



High Level Design Document

Author	Shubham Gaikwad
Company	iNeuron.ai
Project Name	Amazon Sales Data Analysis

Contents

1.Introduction	
2.Problem Statement	
3.Business Requirements	
4.Architecture	
5.WorkFlow	
6.Technologies	
7.Data Source	
8.EDA	
9.Conclusion	

Introduction:

The purpose of this High-Level Design Document is to provide an overview of the Business Intelligence (BI) solution for analysing Amazon Sales data. Using Exploratory Data Analysis (EDA) approaches, the data was evaluated and processed in Jupyter Notebook. The BI solution intends to provide actionable insights and support data-driven decision-making for sales analysis, KPIs, items, sales profitability, and trends over time.

Problem Statement:

Sales management has gained importance to meet increasing competition and the need for improved methods of distribution to reduce cost and to increase profits. Sales management today is the most important function in a commercial and business enterprise.

ETL: Extract-Transform-Load some Amazon dataset and find sales trend month wise, year wise, yearly month wise, Sales Comparison, 10 items most and least sold, Sales and Profitability Analysis

Business Requirements:

The business intelligence solution must meet the following business needs and give a comprehensive monthly and annual sales analysis to identify sales trends and seasonality.

- Analyse sales data and items in depth to identify top-performing products and key performance indicators (KPIs).
- Analyse overall business success by analysing sales and profitability.

Architecture:

The BI Solution architecture is made up of several parts that work together to process and analyse Amazon sales data. The following components were used to create the BI solution:

A. Data Source: - Extracts data from a data source and downloads it onto a local system.

B. BI Tools: Business intelligence tools (Jupyter Notebook). Jupyter Notebook is a helpful tool. It provides a platform for combining code, text, interactive data analysis, and visualisations in a single document, as well as ad-hoc reporting.

C. Data Governance and Security: Implement data governance practises to assure data accuracy, consistency, and privacy. Implement role-based access control to limit data access depending on user roles.

Work Flow:

1. Dataset: The first stage in the workflow is to collect and obtain the necessary dataset for the sales and profit analysis. Sales transactions, product details, revenue, costs, timestamps, and other relevant metrics may be included in the dataset.

2. Tool (Jupyter Notebook): For exploratory data analysis (EDA) and data manipulation, Jupyter Notebook will be used. It gives an interactive environment to execute code, displaying data, and presenting results.

3. ETL (Extract, Transform, Load): The dataset will be extracted from the source (e.g., Drive, Excel file) and loaded into Jupyter Notebook for further analysis during the ETL phase. This phase will also include transformation operations such as data cleaning, filtering, and handling missing values.

4. Data Pre-processing: During the data pre-processing step, the dataset is prepared for sales and profit analysis. Among the responsibilities are: - Aggregating sales data by relevant time intervals (e.g., month, year).

- Determining profitability measures such as Gross Profit Margin and Net Profit.

- Developing derived variables like Average Order Value (AOV).

5. Analysis of Sales and Profits: The sales and profit analysis will be performed using the pre-processed data to reveal important insights. This includes the following:

- Determining the best-selling products and categories.

- Identifying seasonality patterns and analysing sales trends over time.

- Customer segmentation based on purchasing behaviour.

- Conducting individual product profitability analyses

6. Data visualisation: Data visualisation is critical for communicating findings in an understandable and effective manner. Various visualisation approaches, such as:

- line charts and cat plots for sales trend analysis used.

- A bar chart that compares sales indicators and item sales.

- Pie chart plots for visualising sales-profitability correlations.

Technologies:

1. Microsoft Excel: Microsoft Excel is a popular spreadsheet tool that allows for data processing, visualisation, and analysis. Excel is used to clean basic data, organise it, and do simple computations. It has an easy-to-use user interface, making it suitable for quick exploratory analysis by a wide range of users.

2. Jupyter Notebook: Jupyter Notebook is an interactive computing environment that enables data analysts to create and share documents that integrate live code, visualisations, and explanatory prose. Jupyter Notebook is a data visualisation and exploratory data analysis (EDA) tool. It supports a variety of computer languages, including Python, allowing for easy interaction with other data analysis frameworks.

3. Python: Python is a high-level, adaptable programming language that is well-known for its readability and ease of use. Because of its huge libraries and ecology, it is often used in data analysis. Python offers a strong basis for data processing, numerical computation, and statistical analysis.

4. NumPy: NumPy is a key Python module for numerical computing. It supports arrays and matrices, allowing for the efficient handling of big datasets and mathematical calculations. NumPy accelerates numerical computations, making it an essential component in data analysis workflows.

5. Pandas: pandas are a strong NumPy-based toolkit that provides data structures and functions for rapid data manipulation and analysis. Data Frame and Series objects are provided by pandas to help with data organisation, indexing, and filtering. It is crucial in data pre-processing and transformation operations.

6. Matplotlib: Matplotlib is a popular Python data visualisation toolkit. It enables users to produce static, interactive, and publication-quality visualisations such as line plots, bar charts, and scatter plots, among other things. Matplotlib enables data researchers to efficiently communicate insights through visually appealing representations.

7. Seaborn: Seaborn is a Matplotlib-based statistical data visualisation library. It offers a higher-level interface for producing visually appealing and informative visualisations. Seaborn streamlines the process of constructing complicated graphs and allows for the examination of data linkages.

Data Source:

The Amazon Sales dataset, which includes raw sales data such as sales transactions, products, revenue, costs, and timestamps, is the major data source for this BI solution. The information retrieve can from a source like xlsx, CSV, Drive or database etc.

EDA:

The following tasks were done in Jupyter Notebook during the Exploratory Data Analysis (EDA) phase:

- ETL: Extract, Transform, Load is a method for extracting data from a variety of sources, changing it into an appropriate format, and then loading it into a destination.
- Data Cleaning: Handled missing values, duplication, and data integrity problems.

- Data Transformation: The process of converting data kinds and standardising variables for analysis.
- Data Aggregation: To aid in further analysis, Aggregation, Statistics, and Grouping were conducted on data by month and year.
- Data Visualisation: Various plots and charts were used to display sales patterns and variable relationships.

A. Month and Year Sales Analysis: The BI solution will deliver interactive graphs and visuals for month and year sales analysis. It will include the following:

- Line and bar charts showing sales patterns over time.
- Year to year and month to month comparisons.

B. Sales Metrics and Item Analysis: The BI solution will provide insights into sales metrics and items such as:

- Top-selling items and quantities of them
- Sales Comparison - Sales Metrics and Item Bar Graph

C. Sales and Profitability Analysis: The BI solution could be used for sales and profitability analysis, such as gross, profit, loss, average, and discount analysis etc.

- A comparison of sales and profitability
- Pie and Donut charts for highlighting Sales Metrics and Items

Conclusion:

The designed Business Intelligence solution for Amazon Sales data will allow stakeholders to make data-driven choices, find sales patterns, and improve business performance. Customers would be able to acquire a better understanding of sales trends and take strategic adjustments if EDA insights from Jupyter Notebook were integrated into an appropriate BI platform.

The Amazon logo is centered at the bottom of the page. It features the word "amazon" in a white, lowercase, sans-serif font. A yellow curved arrow, representing the "smile" logo, is positioned below the text, starting under the 'a' and ending under the 'n'.