High-Level Design (HLD)

Project Name: Amazon Sales Data Analysis
Revision Number: 1.0
Last Revised: 08/07/2025

Author: Sourav Patra

Date	Version	Description	Author
08 July 2025	1.0	Initial version of HLD	Sourav Patra

Abstract

Sales data analysis plays a critical role in understanding and improving business performance. In the e-commerce sector, efficient sales management helps track key trends, customer behaviour, and product performance. This project focuses on analysing Amazon's sales dataset to derive insights on revenue, profit, sales trends, and geographical performance using Power BI. The goal is to empower decision-making through visual storytelling of data.

1. Introduction

1.1 Why this High-Level Design Document?

This High-Level Design (HLD) document outlines the architectural overview and design principles of the Amazon Sales Data Analytics project. It helps clarify the approach, structure, and tools used before actual development. The HLD ensures that the project is scalable, maintainable, and aligns with business intelligence goals.

1.2 Scope

The scope of this document is to:

- Describe the system architecture and data flow
- Identify key design components and BI tools used
- Define KPIs and dashboards to be delivered
- Outline optimization strategies and deployment details

---

2. General Description

2.1 Product Perspective & Problem Statement

The problem involves analysing a dataset of Amazon sales records to identify:

- Month-wise, year-wise, and year-month-wise sales trends
- Key business metrics (KPIs) such as revenue, profit, and unit sales
- Relationships between region, product types, channels, and performance

The output includes interactive dashboards that present sales trends, product performance, regional revenue, and more.

2.2 Tools Used

- **Microsoft Power BI** – For dashboards, KPIs, and visual storytelling
- **Microsoft Excel** – For data cleaning using Power Query
- **Power Query** – To extract and transform the raw dataset
- **DAX** – For calculated columns and measures
- **CSV Dataset** – “Amazon Sales Records”

3. Design Details

3.1 Functional Architecture

ETL Flow:

1. **Extract**: Load the raw dataset (CSV) into Excel and Power BI
2. **Transform**: Clean data, convert date formats, create new time-based columns (Year, Month, YearMonth)
3. **Load**: Load the cleaned dataset into Power BI

Visualization Layer:

- Dashboards: Monthly Revenue Trend, Profit by Region, Units Sold by Product
- Filters: Region, Year, Item Type

Data Layer:

- Sales Table (Amazon data)
- Date Table (created using DAX for time intelligence)

3.2 Optimization

- Used DateTable for optimized time filtering
- Minimized columns and unused data fields during ETL
- Applied slicers instead of heavy filter panes
- Used numeric-based KPIs (faster to compute than text fields)
- Created summary views for overview analysis, with drilldowns for details

4. KPIs

KPI Name	Description
Total Revenue	Sum of revenue across all orders
Total Profit	Total revenue minus total cost
Total Units Sold	Aggregated unit sales
Average Unit Price	Revenue ÷ Units Sold
Monthly Revenue	Revenue grouped by Year-Month
Profit by Region	Total profit grouped by geographical region

5. Deployment

The dashboard is developed using **Microsoft Power BI Desktop** and can be shared as:

- .pbix file
- Embedded in Power BI service (web)
- Exported to PDF or PowerPoint for presentation

Data can be updated through Power Query or connected to a live data source in enterprise settings.