

Analyzing Amazon Sales data

High-Level Design

Version: 1.0

Revision Date: 21-11-24

Riya Singh

Document Version Control Table

version	Date	Author	Changes
1.0	21-11-2024	Riya Singh	Tools Used
1.1	22-11-2024	Riya Singh	KPI

Contents

1. Introduction.....	5
1.1 Objective.....	5
1.2 Scope.....	5
1.3 Audience.....	6
2. Project Overview.....	7
2.1 Background.....	7
2.2 Goals.....	7
3. System Architecture.....	9
3.1 ETL Process.....	9
3.2 Architecture Diagram.....	10
Diagram.....	10
4. Dataset Description.....	11
4.1 Overview.....	11
4.2 Key Attributes.....	11
4.3 Data Quality.....	12
5. Technical Design.....	13
5.1 Data Modeling.....	13
5.4 Tools Used.....	13
6. KPIs (Key Performance Indicators).....	14
6.1 Introduction to KPIs.....	14
6.2 KPI Visualization Details.....	14
6.3 Interpretation of KPIs.....	14
6.4 Enhancements for Future.....	14
7. Deployment.....	15

Abstract

The Amazon Sales Analysis project focuses on understanding sales trends and key metrics to enhance business intelligence and improve profitability. Utilizing an ETL process, data was extracted, transformed for analysis, and loaded into Power BI. The project identifies sales trends across monthly, yearly, and yearly-monthly dimensions while uncovering meaningful relationships between attributes such as region, item type, and sales channel. Key metrics like total revenue, profit, and cost were calculated to provide actionable insights. A blue-themed, interactive Power BI dashboard was designed to visualize these findings. The analysis highlights seasonal trends, top-performing regions, and high-profit product categories, supporting data-driven decision-making. This project demonstrates how effective sales management can reduce costs, increase profits, and refine distribution strategies. It provides a scalable foundation for future enhancements, including predictive analytics and integration of additional datasets.

1. Introduction

The introduction chapter of the High-Level Documentation (HLD) provides a clear and detailed context for the project, outlining its purpose, scope, and intended audience. This chapter serves as the foundation for understanding the objectives and expected outcomes of the Amazon Sales Analysis project.

1.1 Objective

The primary goal of this project is to analyze Amazon sales data to uncover trends, identify key metrics, and provide actionable insights for enhancing business intelligence in the e-commerce domain.

- **Key Objectives Include:**
 - Identifying **sales trends** on a monthly, yearly, and yearly-month basis to understand seasonal and long-term patterns.
 - Calculating critical metrics such as total revenue, profit, and cost to evaluate financial performance.
 - Exploring the relationships between attributes like region, product type, and sales channel to discover meaningful correlations and opportunities.
 - Providing visual insights through a comprehensive Power BI dashboard to facilitate data-driven decision-making.

This analysis supports the development of strategies to optimize sales management, reduce costs, and improve profitability in an increasingly competitive market.

1.2 Scope

The scope of the project encompasses all activities required to transform raw sales data into valuable insights using modern analytical tools and techniques.

- **Dataset:**

The project uses a clean Amazon dataset containing 101 rows and 18 columns, including attributes like region, country, item type, sales channel, order priority, order date, units sold, total revenue, and total profit.
- **Analytical Dimensions:**
 - **Time-Based Analysis:** Monthly, yearly, and year-month sales trends.
 - **Performance Metrics:** Total revenue, profit margins, and cost-effectiveness across regions, channels, and product types.

- **Attribute Relationships:** Correlations and patterns between sales attributes (e.g., product type vs. profitability).
- **Tools and Techniques:**
Power BI is utilized for ETL (Extract-Transform-Load), analysis, and interactive visualization. Additional tools may include DAX formulas and SQL queries.
- **Outcomes:**
Deliver a visually striking Power BI dashboard, enabling stakeholders to interpret insights intuitively and support informed decision-making.

1.3 Audience

This document is intended for a range of stakeholders involved in sales management, business analysis, and decision-making:

- **Business Analysts:** To gain insights into sales performance and identify areas for improvement.
- **Data Teams:** For technical implementation, understanding the analysis workflow, and supporting data integration tasks.
- **Management:** To make strategic decisions based on the insights and trends highlighted in the analysis.
- **Developers and Engineers:** To understand the ETL and dashboard creation processes, enabling future scalability.

2. Project Overview

The **Project Overview** chapter provides a deeper understanding of the business context, outlining the problem being addressed and the goals the project aims to achieve. It sets the stage for the analytical and technical processes discussed later in the document.

2.1 Background

Sales management plays a critical role in the success of commercial and business enterprises, especially in the competitive e-commerce market. With the rapid growth of online retail, businesses face increasing pressure to optimize their operations and maximize profits. Key challenges include:

- Identifying and responding to seasonal and long-term sales trends.
- Managing distribution effectively to reduce operational costs.
- Understanding regional and product-level sales performance to focus resources on high-impact areas.

In this context, Amazon, as a leading global e-commerce platform, provides an ideal case for analyzing sales data to derive actionable insights. This project addresses the need for improved sales management by leveraging data analysis to uncover patterns, identify performance metrics, and recommend strategies for increasing efficiency and profitability.

2.2 Goals

The project aims to achieve the following key objectives:

1. **Analyze Sales Trends:**
 - Identify trends on **monthly**, **yearly**, and **yearly-month** dimensions to gain insights into seasonality and long-term growth patterns.
2. **Evaluate Key Metrics:**
 - Calculate and analyze metrics such as total revenue, total profit, total cost, and units sold to assess financial performance.
 - Identify top-performing regions, products, and sales channels to inform strategic decision-making.
3. **Uncover Meaningful Relationships:**
 - Explore correlations between attributes like region, product type, sales channel, and profitability.
 - Highlight areas with potential for cost reduction and revenue optimization.

4. Develop a Comprehensive Dashboard:

- Create an interactive Power BI dashboard with visually appealing, blue-themed designs.
- Provide stakeholders with intuitive tools to explore insights and trends for decision-making.

5. Support Data-Driven Strategies:

- Equip business leaders with actionable recommendations based on the analysis, enabling them to improve sales strategies, optimize resource allocation, and increase profitability.

3. System Architecture

The **System Architecture** chapter describes the technical approach used to process and analyze the data. It outlines the ETL (Extract-Transform-Load) process, highlighting the tools and techniques employed, and includes a high-level architecture diagram to visualize the workflow.

3.1 ETL Process

The ETL process involves three key stages: **Extract**, **Transform**, and **Load**, designed to prepare the dataset for analysis and visualization.

1. Extract

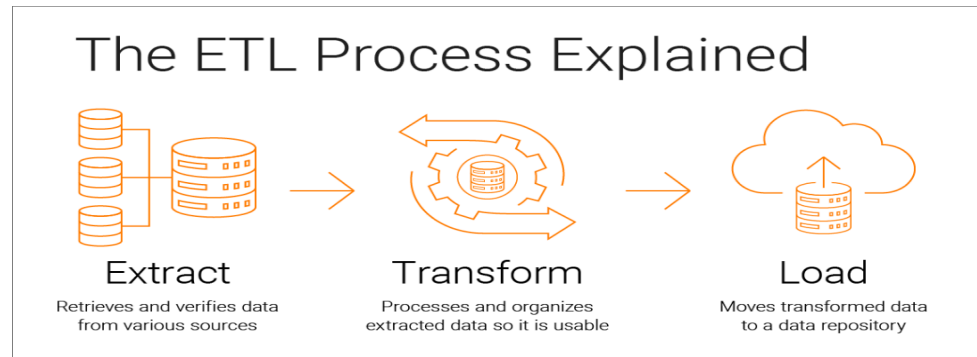
- The dataset, containing 101 rows and 18 columns, was sourced as a clean, structured CSV file.
- Key attributes include region, country, item type, sales channel, order priority, order date, units sold, unit price, unit cost, total revenue, total cost, and total profit.
- The data was imported into **Power BI** for processing.

2. Transform

- **Data Cleaning**: Ensured all data entries were consistent, removing duplicates and handling missing values (if any).
- **DateTable Creation**: A separate DateTable was created in Power BI for time-based analysis, including attributes such as year, month, and month-year combinations.
- **Derived Metrics**: Calculated additional columns like monthly revenue, yearly profit, and yearly-month sales trends using DAX formulas.
- **Attribute Relationships**: Established relationships between DateTable and the main dataset (cleaned_amazon_sales_data) to enable seamless filtering and analysis.

3. Load

- The transformed data was loaded into **Power BI** for creating an interactive, visually appealing dashboard.
- Visualizations such as bar charts, line graphs, heatmaps, and KPI cards were implemented to present insights clearly.

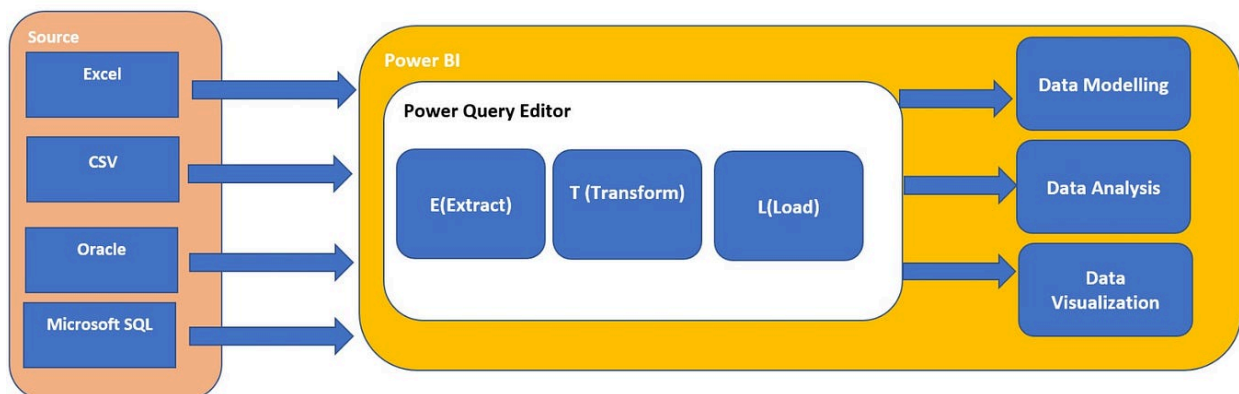


3.2 Architecture Diagram

Below is a description of the architecture diagram. You can create this using tools like PowerPoint, Canva, or Lucidchart. Here's how it flows:

1. **Data Source:** Amazon sales dataset (CSV file).
2. **ETL Process:**
 - **Extract:** Data imported into Power BI.
 - **Transform:** Cleaning, DataTable creation, and calculations using DAX.
 - **Load:** Transformed data prepared for visualization.
3. **Data Analysis and Visualization:**
 - Power BI Dashboard displaying trends, metrics, and relationships.
4. **Output:** Actionable insights shared with stakeholders for decision-making.

Diagram



You can use a tool to visually enhance this with icons and arrows for a professional look.

4. Dataset Description

The **Dataset Description** chapter provides an overview of the data used in the project, including its structure, key attributes, and preparation for analysis.

4.1 Overview

The dataset used for this project contains sales data from Amazon, structured in a tabular format. It includes a variety of attributes that capture essential details about sales transactions, allowing for comprehensive analysis and visualization.

- **Total Rows:** 101
- **Total Columns:** 18

4.2 Key Attributes

Attribute	Description
Region	Geographic region where the sales occurred.
Country	Specific country within the region where the sales occurred.
Item Type	Category of the item sold (e.g., electronics, furniture, etc.).
Sales Channel	Mode of sales (e.g., online or offline).
Order Priority	Priority level of the order (e.g., High, Medium, Low).
Order date	Date when the order was placed.
Ship date	Date when the order was shipped.
Units sold	Number of units sold for each transaction.
Unit price	Price of a single unit of the item.
Unit cost	Cost of a single unit of the item.
Total Revenue	Total revenue generated from the sale of units (calculated as Units Sold \times Unit Price).
Total Cost	Total cost incurred for the sale of units (calculated as Units Sold \times Unit Cost).
Total Profit	Profit generated from the transaction (calculated as Total Revenue - Total Cost).

Year	Extracted year of the sales order for time-based analysis.
Month	Extracted month of the sales order for time-based analysis.
Year_ Month	Concatenated year and month for trend analysis (e.g., "2024-11").
Total Sales	Aggregated sales value for specific time periods or attributes.

4.3 Data Quality

The dataset is clean and pre-processed, ensuring accuracy and consistency:

- **No Missing Values:** All required data points are present.
- **No Duplicates:** Redundant entries were removed during the cleaning process.
- **Date Formatting:** Dates are properly formatted for time-based analysis.
- **Derived Columns:** Additional columns, such as **Year**, **Month**, and **Year_Month**, were created to facilitate trend analysis.

5. Technical Design

The **Technical Design** chapter outlines the approach taken to structure, calculate, and visualize data to achieve the project objectives. It provides details on data modeling, key metrics, visualizations, and the tools used.

5.1 Data Modeling

To facilitate efficient analysis and meaningful insights, data modeling was performed using **Power BI**:

- **Primary Tables:**
 - **cleaned_amazon_sales_data**: The main dataset containing sales details.
 - **DateTable**: A custom table created for time-based analysis.
- **Relationships:**
 - A **one-to-one relationship** was established between the **Date** column in **DateTable** and the **Order Date** column in **cleaned_amazon_sales_data**. This relationship ensures seamless filtering and aggregation of data by time periods.
 - The relationship enabled advanced time intelligence calculations using DAX formulas for trend analysis (e.g., year-over-year comparisons).

5.4 Tools Used

1. **Python (Google Colab):**
 - Data cleaning and preprocessing were performed using Python libraries like **pandas** and **numpy**.
 - Ensured the dataset was free of duplicates and inconsistencies.
2. **Power BI:**
 - Used for data modeling, analysis, and creating an interactive dashboard.
3. **DAX Formulas:**
 - Dynamic measures and calculated columns were created using Data Analysis Expressions (DAX) in Power BI.
4. **SQL:**
 - Used for querying and validating dataset attributes during the cleaning process.



6. KPIs (Key Performance Indicators)

6.1 Introduction to KPIs

Key Performance Indicators (KPIs) are essential metrics that provide a quick snapshot of business performance. In the Amazon Sales Analysis dashboard, KPIs focus on sales trends, product performance, and market analysis to help stakeholders make informed decisions.

6.2 KPI Visualization Details

- **Placement:**
 - Positioned at the top of the dashboard for immediate visibility.
- **Design:**
 - Each KPI uses **cards** with large font sizes and visually appealing color themes (blue theme with white text).
- **Dynamic Updates:**
 - KPIs dynamically adjust based on user filters, such as year, region, or sales channel.

6.3 Interpretation of KPIs

- **Product Units Sold** helps assess overall market demand.
- **Sales Projection** aids in forecasting future revenue.
- **KPI Rating** combines several performance indicators into a single metric, summarizing overall performance.

6.4 Enhancements for Future

- Include real-time data updates to track KPIs continuously.
- Add additional KPIs like Customer Retention Rate or Average Order Value for deeper insights.

7. Deployment

The deployment process of the Amazon Sales Analysis dashboard ensures that stakeholders have secure and seamless access to insights for effective decision-making. The finalized Power BI dashboard is published on the **Power BI Service**, enabling cloud-based access and sharing through secure links or embedding into organizational intranet portals. Role-based access control is implemented to manage user permissions, ensuring sensitive data is accessible only to authorized personnel such as managers and analysts. The dashboard is integrated into existing business intelligence workflows, allowing for consistent and centralized reporting. Thorough testing is conducted to validate the accuracy of visualizations, KPIs, and filtering functionalities before deployment. Additionally, the dashboard is designed for scalability, allowing the integration of additional datasets or modules in the future. To ensure smooth operation post-deployment, a maintenance plan is established to update data periodically, address user feedback, and optimize performance. Tools like **Power BI Service**, along with platforms like **Azure** or **SharePoint**, can be used for secure hosting and access. This approach ensures that the dashboard remains a reliable and actionable resource for driving business growth and strategic decisions.

