Amazon Sales Data Analysis

High Level Design (HLD)

Shubham Ghosh & Vipin Bhagat

Revision Number: 1.0
Last date of revision: 11/09/2023

Document Version Control

Date Issued	Version	Description	Author
12-09-2023	**1.0**	**First version of Complete HLD**	**Shubham Ghosh & Vipin Bhagat**

Contents

Abstract

Analyzing Amazon sales data is crucial for sellers to make informed decisions and improve their performance in the competitive e-commerce market. This data-driven approach allows sellers to understand consumer behavior, market trends, and customer preferences. Here's a breakdown of key points and benefits associated with Amazon sales data analysis:

1. **Metrics for Analysis**

- Total Sales: The overall revenue generated from sales.

- Sales Quantity: The total number of products sold.

- Total Profit: The net profit earned after deducting costs.

- Last Year Sales: A comparison with previous year's sales to identify growth or decline trends.

- Other Metrics: Depending on the dataset, various additional metrics can be considered, such as customer reviews, return rates, advertising spend, and more.

2. **Performance Improvement**

- Data analysis helps sellers identify areas for improvement in their business operations, pricing strategies, and product offerings.

- It enables sellers to optimize advertising campaigns, inventory management, and pricing strategies based on historical sales data.

3. **Market Trends and Customer Behavior**

- By analyzing sales data, sellers can gain insights into market trends and changing customer behaviors.

- This information can be used to adapt to market shifts, identify emerging product categories, and anticipate consumer demands.

4. **Customer Preferences**:

- Sales data analysis can reveal customer preferences and buying patterns, allowing sellers to tailor their product listings and marketing efforts accordingly.

- Understanding customer preferences can lead to the development of products that better meet market demands.

5. **Data Analytics in E-commerce**

- In the e-commerce industry, data analytics plays a critical role in understanding business performance and customer satisfaction.

- Sellers can use data analytics to identify weak areas in their operations and take corrective actions.

6. **Business Decision Making**

- Data analysis informs strategic decisions, such as expanding product lines, entering new markets, or adjusting pricing strategies.

- It provides a factual basis for decision-making rather than relying solely on intuition.

7. **Sales Trend Analysis**

- Analyzing sales data month-wise, year-wise, and quarter-wise helps identify seasonal trends and fluctuations.

- Sellers can adjust their marketing and inventory strategies to capitalize on peak sales periods.

8. **Regional and Country Insights**

- Data analysis can also provide valuable insights into regional and country-specific sales patterns.

- Sellers can tailor their marketing and inventory strategies to specific geographic regions.

9. **Key Indicators and Metrics**

- Identifying key performance indicators (KPIs) and metrics that influence customer choices is essential.

- These indicators can include conversion rates, customer reviews, and product ratings.

1 Introduction

1.1 Why this High-Level Design Document?

A High-Level Design (HLD) document is a crucial part of the software development process, as it provides a blueprint for how the system will be designed and built. Here's an overview of what the HLD document typically includes:

1. Design Aspects:
- The document should thoroughly present and define all design aspects of the project. This includes architectural decisions, data structures, algorithms, and design patterns that will be used.

2. User Interface:
- Describe the user interface (UI) being implemented. This includes the layout, navigation, and user interactions. It may include wireframes or mock-ups to illustrate the UI design.

3. Hardware and Software Interfaces:
- Detail the hardware components, software libraries, and third-party services that the system will interact with. Explain how these interfaces will be utilized in the project.

4. Performance Requirements:
- Specify the performance expectations and requirements for the system. This may include response times, throughput, and scalability considerations.

5. Design Features and Architecture:
- Explain the overall architecture of the project, including the high-level components/modules, their responsibilities, and how they interact with each other.
- Highlight any innovative design features or unique approaches being employed.

6. Non-Functional Attributes:
- Discuss non-functional attributes that are critical to the success of the project. Common attributes include:
  - **Security:** Describe how data and system security will be ensured.
  - **Reliability:** Explain measures to ensure system uptime and fault tolerance.
  - **Maintainability:** Detail plans for code maintainability, updates, and enhancements.
  - **Portability:** Discuss how the system will be made compatible with different platforms.
  - **Reusability**: Explain how components/modules can be reused in other projects.
  - **Application Compatibility**: Address compatibility with other software or systems.
  - **Resource Utilization:** Discuss how system resources (CPU, memory, etc.) will be managed.
  - **Serviceability**: Describe how maintenance, debugging, and troubleshooting will be handled.

7. Dependencies:
- Identify any external dependencies, such as third-party APIs or external databases, and how they will be managed.

8. Testing and Quality Assurance:
- Outline the testing strategies and quality assurance processes that will be used to ensure the reliability and correctness of the software.

9. Deployment Plan:
- Provide an overview of the deployment process, including any specific deployment tools or methodologies.

10. Risk Assessment:
- Identify potential risks that could impact the project and propose mitigation strategies.

11. Appendices:

- Include any supplementary materials, diagrams, flowcharts, or reference materials that support the design.

12. References:

- If applicable, include references to external documentation, standards, or research used in the design process.

The HLD document serves as a reference guide for developers, testers, and other stakeholders involved in the project. It ensures that the project's design is well-defined, consistent, and aligned with the project's goals and requirements. Additionally, it helps detect and resolve design contradictions or issues before coding begins, saving time and resources in the long run.

1.2 Scope

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

1.3 Definitions

- ➢ ETL – Extract, Transform, Load
- ➢ EDA – Exploratory Data Analysis
- ➢ CSV file – Comma-separated values file, opened in MS Excel
- ➢ Power Query – With Power Query, you can connect to many different data sources and transform the data into the shape you want ➢ Power BI - An interactive data visualization software company focused on business intelligence.

2 General Description

2.1 Product Perspective & Problem Statement

Sales management is of paramount importance in today's competitive business landscape, serving as a vital function for commercial enterprises seeking to optimize distribution, reduce costs, and enhance profits. The primary objective of this project is to conduct a thorough analysis of Amazon sales data, generating valuable insights that can drive future business decisions. By scrutinizing this data, sellers can pinpoint weaknesses in their business models and operational processes, allowing for cost reduction and improved profitability. Furthermore, this initiative enables the formulation of strategic solutions to address identified issues. The project's ultimate goal is to harness Microsoft Power BI for creating visual representations of the data, facilitating a clearer and more actionable understanding of critical trends and patterns.

2.2 Tools used

Microsoft Excel

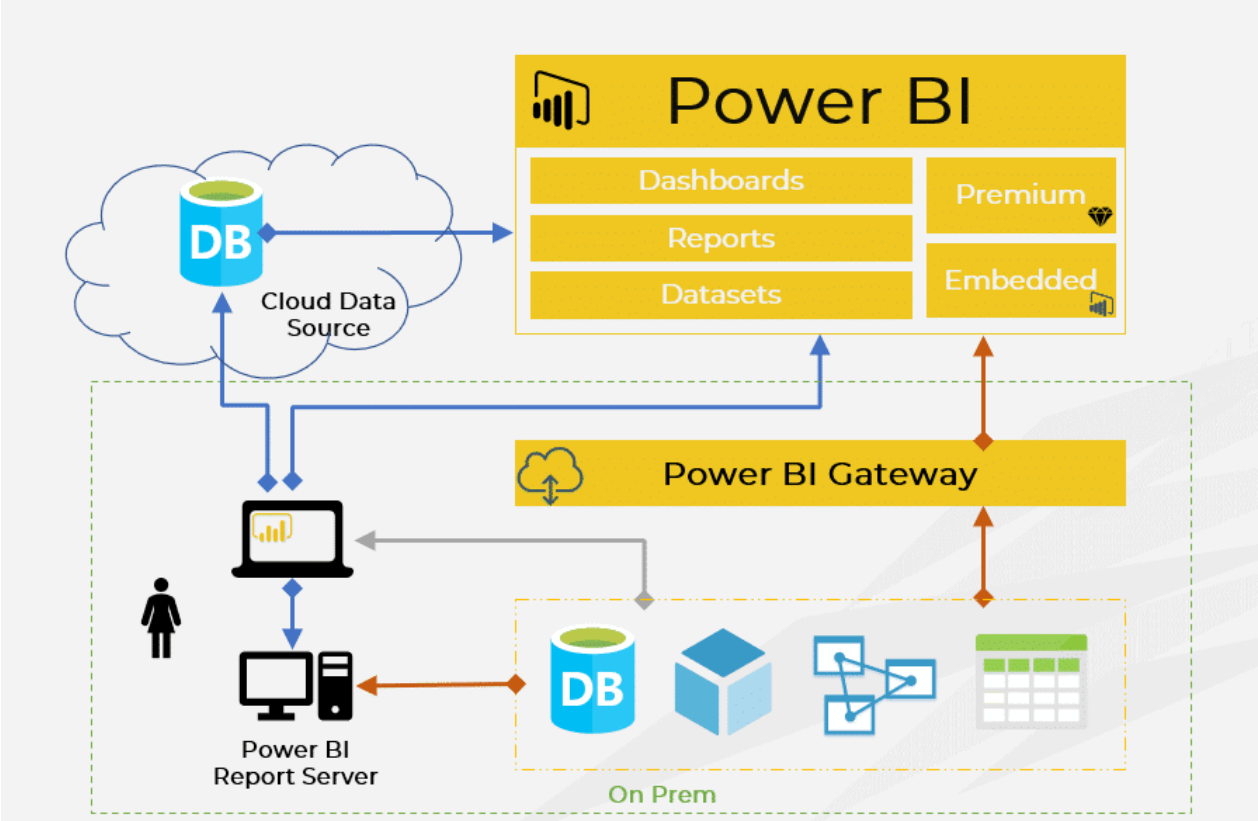
Microsoft Excel, part of Microsoft Office, aids data loading and basic cleaning in CSV format. It's integrated into Power BI for further analysis.

Microsoft Power BI

Power BI, by Microsoft, is a business intelligence tool that transforms diverse data sources into interactive visual insights. It reads data from various formats, including databases and spreadsheets, facilitating comprehensive analysis and visualization.

3 Design Details

3.1 Functional Architecture



Power BI

Power BI is a business analytics solution that lets you visualize your data and share insights across your organization, or embed them in your app or website. Connect to hundreds of data sources and bring your data to life with live dashboards and reports. It provides interactive visualizations with self-service business intelligence capabilities, where end users can create reports and dashboards by themselves, without having to depend on any information technology staff or database administrator.

Benefit of Business Intelligence

The benefits of Business intelligence are as follows:

- ❖ Business intelligence is faster more accurate process of reporting critical information.
- ❖ Business intelligence facilitates better and efficient decision-making process.
- ❖ Business intelligence provides timely information for better customer relationship management.
- ❖ Business intelligence improves profitability of the company.
- ❖ Business intelligence provides a facility of assessing organization's readiness in meeting new business challenges.
- ❖ Business intelligence supports usage of best practices and identifies every hidden cost

3.2 Optimization

Efficient data strategy:

- Minimize fields and records.
- Optimize extracts by materializing calculations, removing columns, and using accelerated views.
- Reduce data points in views through guided analytics.
- Remove unnecessary dimensions from the detail shelf.
- Explore different data views.
- Limit and streamline filters to speed up queries.
- Prefer include filters over exclude filters, especially for dimensions with many members.
- Use continuous date filters for faster processing.
- Utilize Boolean or numeric filters for quicker computations.
- Implement parameters and action filters to reduce query load across data sources.

To optimize and materialize calculations:

- Perform calculations directly in the database.
- Minimize nested calculations to enhance performance.
- Reduce the granularity of Level of Detail (LOD) or table calculations in the view, as highly granular calculations take longer.
- For LODs, consider the number of unique dimension members in the calculation.
- For table calculations, be mindful that more data points in the view lead to longer processing times. When possible, use MIN or MAX instead of AVG, as AVG requires more processing and can lead to duplicated results with MIN and MAX.

4.KPIs

Dashboards will be implemented to display and indicate certain KPIs and relevant indicators about the sales of the products in various years.

Sales representative according to the number of the sales.

4.1 KPIs (Key Performance Indicator)

Key indicators for Sales Data analysis and their relationships with different metrics:

1. **Total Profit Distribution Yearly, Quarterly, and Monthly:** Analyze profit trends at different time intervals to identify seasonal patterns and opportunities for growth.

2. **Items That Generated Highest Revenue and Profit:** Identify the top-performing products in terms of both revenue and profit to focus on their success factors.

3. **Top 5 Items that generated highest Profit in terms of percentage:** Determine which products yield the highest profit margins, allowing for strategic pricing and marketing decisions.

4. **Total Profit Across Various Regions and Countries**: Understand the geographic distribution of profits to allocate resources effectively and target high-profit regions.

5. **Total Revenue Contribution per Item**: Break down revenue by individual items to assess their impact on overall sales and prioritize product development or marketing efforts.

6. **Total Revenue and Total Cost by Item type**: Compare revenue and costs by item type to evaluate profitability within different product categories.

7. **Revenue generated Year over Year**: Monitor revenue growth or decline over consecutive years to identify trends and opportunities for improvement.

8. **Last two years Revenue distribution with profit percentage**: Visualize revenue distribution over the past two years alongside profit percentages to assess recent performance.

9. **Total Unit sold Across all the years**: Determine the total volume of products sold over all years to gauge overall market demand and sales consistency.

These indicators provide valuable insights into sales performance, profitability, and market dynamics, helping businesses make data-driven decisions and optimize their strategies.

5 Deployment

Prioritizing data and analytics is crucial in the current business landscape. Companies of all sizes are gathering vast amounts of data, but often only analyzing a fraction of it. By harnessing data analytics, businesses can address challenges, gain a competitive edge, and undergo transformative changes. With the proliferation of enterprise data, advancements in database technologies, and the increasing demand for analytical expertise, successful IT organizations are now emphasizing self-service analytics. This involves deploying and scaling tools like Tableau, as well as harmonizing diverse data sources to empower both business users and experts to create and utilize content effectively..