High Level Design Amazon Food Sales Data Analysis

Revision Number: 1.0
Last date of revision: 1/03/2023

Amit Patil

iNeuron

Document Version Control

Date Issued	Version	Description	Author
1 March 2023	1.0	First Version of Complete HLD	Amit Patil

Contents

Abstract

This High-Level Design (HLD) document outlines a project for analyzing Amazon sales data. Theobjective of the project is to use data visualization techniques to gain insights into Amazon sales data and identify trends and patterns.libraries like Numpy and Pandas, will be used to build the framework. The document presents the functional architecture of the project and op-timization strategies for maximizing performance. This HLD document serves as a reference manual for how the modules interact at a high level and provides a detailed description of the design features and architecture of the project.

1. Introduction

1.1 Why this High-Level Design Document?

The purpose of this High-Level Design (HLD) Document is to add the necessary detail to the current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to coding, and can be used as a reference manual for how the modules interact at a high level.

The HLD will:

- Present all of the design aspects and define them in detail
- Describe the user interface being implemented
- Describe the hardware and software interfaces
- Describe the performance requirements
- Include design features and the architecture of the project
- List and describe the non-functional attributes like:
    - Security
    - Reliability
    - Maintainability
    - Portability
    - Re usability
    - Application compatibility
    - Resource utilization
    - Serviceability

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.

2. General Description

2.1 Product Perspective & Problem Statement

The objective of analyzing Amazon sales data is to gain insights into the company's salesperformance. By analyzing sales data, it is possible to identify patterns, trends, and op- portunities for growth. However, analyzing sales data can be a complex process due to the large amount of data involved.

2.2 Tools used

3. Design Details

3.1 Functional Architecture

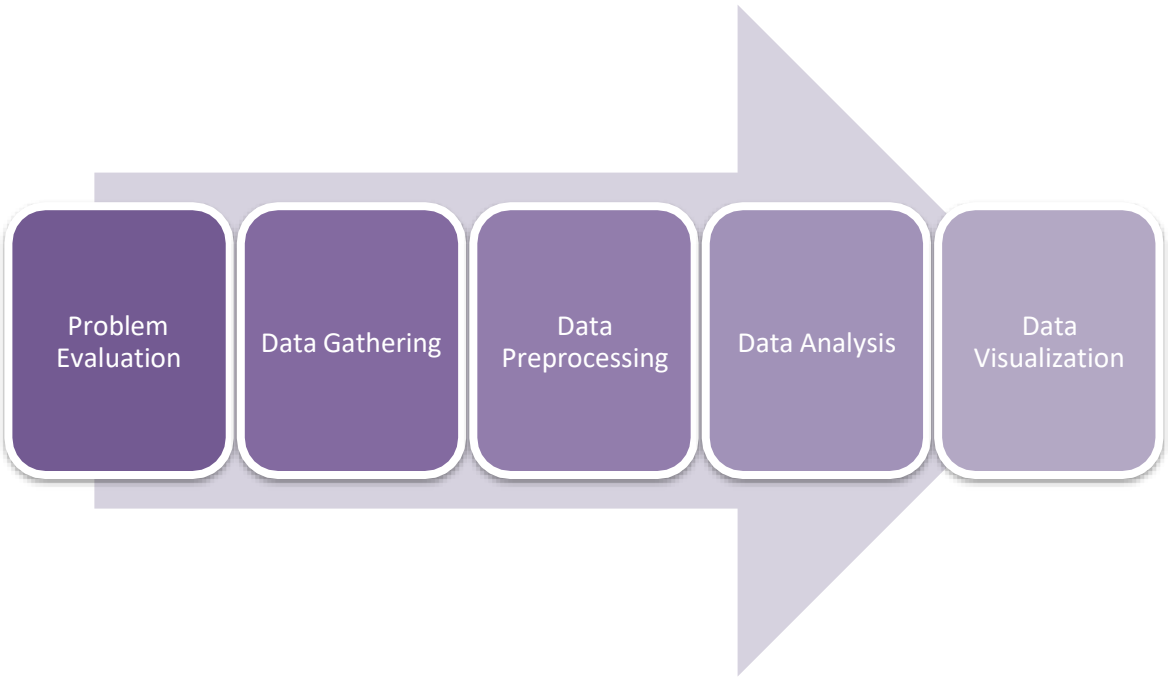

Figure 1: Functional Architecture

3.2 Optimization

Your data strategy drives performance

• Minimize the number of fields
• Minimize the number of records
• Optimize extracts to speed up future queries by materializing calculations, removing columns and the use of accelerated views

Reduce the marks (data points) in your view

• Practice guided analytics. There's no need to fit everything you plan to show in a single view. Compile related views and connect them with action filters to travel from overview to highly-granular views at the speed of thought.
• Remove unneeded dimensions from the detail shelf.
• Explore. Try displaying your data in different types of views.

Limit your filters by number and type

• Reduce the number of filters in use. Excessive filters on a view will create a more complex query, which takes longer to return results. Double-check your filters and remove any that aren't necessary.
• Use an include filter. Exclude filters load the entire domain of a dimension, while include filters do not. An include filter runs much faster than an exclude filter, especially for dimensions with many members.
• Use a continuous date filter. Continuous date filters (relative and range-of-date filters) can take advantage of the indexing properties in your database and are faster than discrete date filters.
• Use Boolean or numeric filters. Computers process integers and Booleans (t/f) much faster than strings.
• Use parameters and action filters. These reduce the query load (and work across data sources).

Optimize and materialize your calculations

• Perform calculations in the database
• Reduce the number of nested calculations.
• Reduce the granularity of LOD or table calculations in the view. The more granular the calculation, the longer it takes.
o LODs - Look at the number of unique dimension members in the calculation.
o Table Calculations - the more marks in the view, the longer it will take to calculate.
• Where possible, use MIN or MAX instead of AVG. AVG requires more processing than MIN or MAX. Often rows will be duplicated and display the same result with MIN, MAX, or AVG.
• Make groups with calculations. Like include filters, calculated groups load only named members of the domain, whereas Tableau's group function loads the entire domain.
• Use Booleans or numeric calculations instead of string calculations. Computers can process integers and Booleans (t/f) much faster than strings.
Boolean>Int>Float>Date>Date Time>String

4. KPIs & Charts

Dashboards will be implemented to display and indicate certain KPIs and relevant indicators forthe food sales, the dashboards will be included to display charts over time with progress on various indicators or factors

4.1 KPIs (Key Performance Indicators)

The KPIs (key performance indicators) for this code are the metrics that are being an- alyzed, which are:

1.Sales revenue, cost, and profit
2. Sales by region, country, item type, sales channel, and order priority
3. 3.Sales trends over time
4.Top-selling products and regions 5.Customer demographics and behavior

By analyzing these metrics, we can gain insights into the performance of the busi- ness and identify areas where improvements can be made. For example, if sales are low in a particular region or for a particular product type, the business may need to adjust its marketing or pricing strategies. Similarly, if customer behavior shows that a particular order priority is more profitable than others, the business may want to focus on promoting that priority to increase overall profitability.

4.2 Charts

Charts displaying an understanding of Amazon Food Sales Data

1. Bar chart for sales revenue, cost, and profit
2. Stacked bar chart for sales by region
3. Pie chart for sales by sales channel
4. Line chart for sales trends over time
5. Horizontal bar chart for top-selling products and reagions
6. Bubble chart for customer demographics and behavior

5. Deployment

Prioritizing data and analytics couldn't come at a better time. Your company, no matter What size, is already collecting data and most likely analyzing just a portion of it to solve business problems, gain competitive advantages, and drive enterprise transformation. With the explosive growth of enterprise data, database technologies, and the high demand for analytical skills, today's most effective IT organizations have shifted their focus to enabling self-service by deploying and operating Power BI at scale, as well as organizing, orchestrating, and unifying disparate sources of data for business users and experts alike to author and consume content.

Amazon Food Sales data is been cleaned using ETL tool Power Query and analyzed on Power BI which will provide a better key insight for the data and tell you a better story of the raw data. The Power BI report is published on workspace where you can play with the data for necessary insights.