iNeuron

High Level Design (HLD)

Amazon Sales Data Analysis

Revision Number: 1.0
Last date of revision: 03/07/2024

Ruchika Doley
Adithya Sai
Document Version
Control

Date Issued	Version	Description	Author
03rd July 2024	1.0	First Version of Complete HLD	Ruchika Doley

Contents

Housing prices are an important reflection of the economy, and housing price ranges are of great interest for both buyers and sellers. In this project, house prices will be predicted given explanatory variables that cover many aspects of residential houses**Error! Bookmark not defined.**

Abstract

Analyzing Amazon sales data can provide valuable insights into consumer behavior, market trends, and product performance, which are crucial for sellers and the platform itself. Amazon's vast and diverse product range makes it a rich source of data for such analysis. This project aims to delve into Amazon sales data to identify key factors influencing sales, such as product category, price, customer reviews, and seasonal trends. By leveraging data mining and machine learning techniques, we will uncover patterns and correlations that can help predict sales performance and optimize inventory management.

This analysis not only assists sellers in strategizing their offerings but also enhances the overall shopping experience for customers by ensuring the availability of popular products. The ultimate goal is to provide actionable insights that drive better business decisions and improve sales outcomes on the Amazon marketplace.

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
  - Reusability
  - 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

Amazon sales data is a crucial indicator of market dynamics, consumer preferences, and product performance. Analyzing this data is of significant interest to both sellers and Amazon as a platform. In this project, sales data will be analyzed using explanatory variables that cover various aspects of products and their sales performance. This analysis will provide insights into consumer behavior, market trends, and product success factors, which can inform strategic decision-making for sellers and improve the overall marketplace experience.

The objective of this project is to perform data visualization techniques to uncover insights from Amazon sales data. By applying various Business Intelligence tools such as Tableau or Power BI, the project aims to achieve a visual understanding of the data. This will help identify key factors influencing sales, detect patterns and trends, and provide actionable insights for optimizing sales strategies and inventory management.

2.2 Tools used

Business Intelligence tools and libraries works such as Numpy, Pandas, Excel, R, Tableau, Power BI are used to build the whole framework.
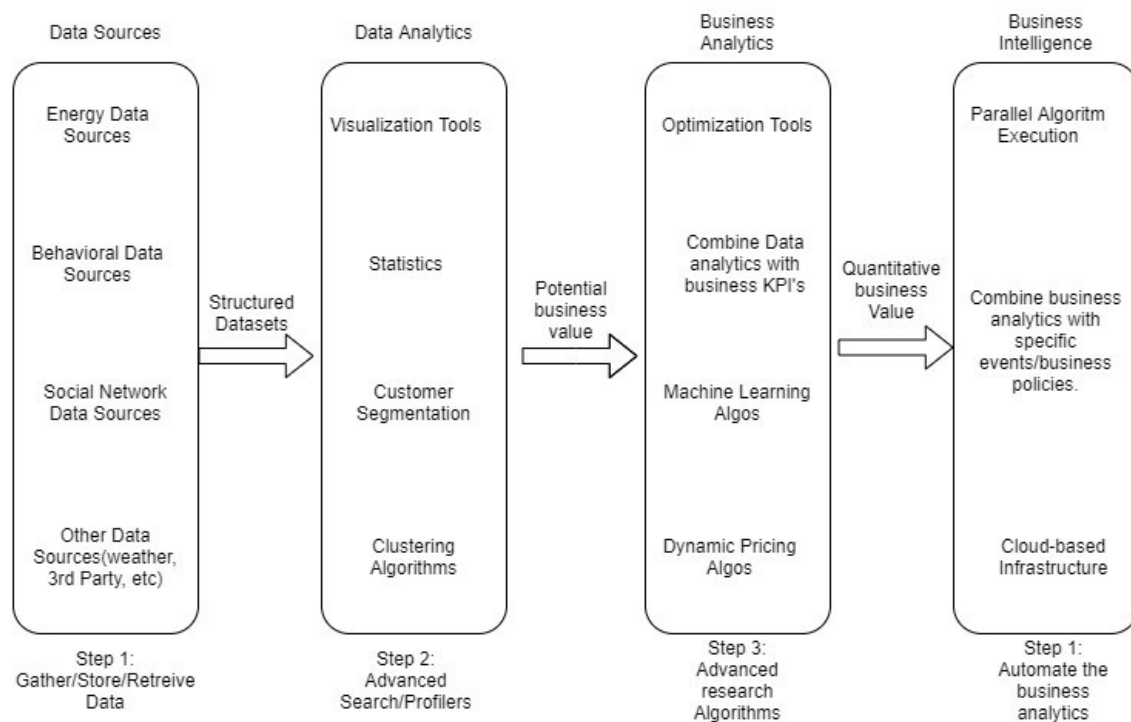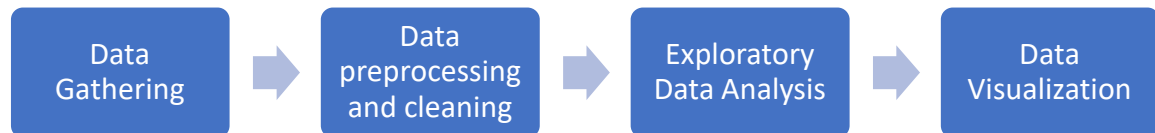
3 Design Details
3.1 Functional Architecture

Data Gathering → Data preprocessing and cleaning → Exploratory Data Analysis → Data Visualization

Figure 1: Functional Architecture of Business Intelligence

How BI Really Works

Organizational Memory	Information Integration	Insight Creation	Presentation
• Data Warehouse • ERP • Knowledge Repository • CMS • DMS	• Business Analytics Tool • Data Mining • Real-time Decision	• Text mining tools • Web mining tools • Environmental Scanning • RFID	• OLAP Tools • Visualization tools • Digital Dashboards • Score Card

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.
  - LODs - Look at the number of unique dimension members in the calculation.
  - 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>DateTime>String

4 KPIs

Dashboards will be implemented to display and indicate certain KPIs and relevant indicators for the disease.

As and when, the system starts to capture the historical/periodic data for a user, the dashboards will be included to display charts over time with progress on various indicators or factors

4.1 KPIs (Key Performance Indicators)

Key indicators displaying a summary of the Housing Price and its relationship with different metrics

1. Yearly, Quarterly and Monthly changes in sales amounts, sales quantity and number of orders.
2. The best performing sales reps in terms of their contribution to total sales margin amount.
3. The items which are purchased in largest quantity.
4. Variation in Sales amount, Sales quantity and Sales margin.
5. Percentage of the total sales margin brought in by the top selling items in each year, quarter and month.
6. Average amount spent per order, average number of orders placed per month and average spending per month of customers.

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 Tableau at scale, as well as organizing, orchestrating, and unifying disparate sources of data for business users and experts alike to author and consume content.

Tableau prioritizes choice in flexibility to fit, rather than dictate, your enterprise architecture. Tableau Server and Tableau Online leverage your existing technology investments and integrate into your IT infrastructure to provide a self-service, modern analytics platform for your users. With on-premises, cloud, and hosted options, there is a version of Tableau to match your requirements. Below is a comparison of the three types:

TYPE PROS CONS

Tableau Server - On Premises

- Full control of hardware and software
- Infrastructure and data remain behind your firewall
- Need dedicated administrators to manage hardware and software
- Additional infrastructure needed to access off-network (mobile, external)

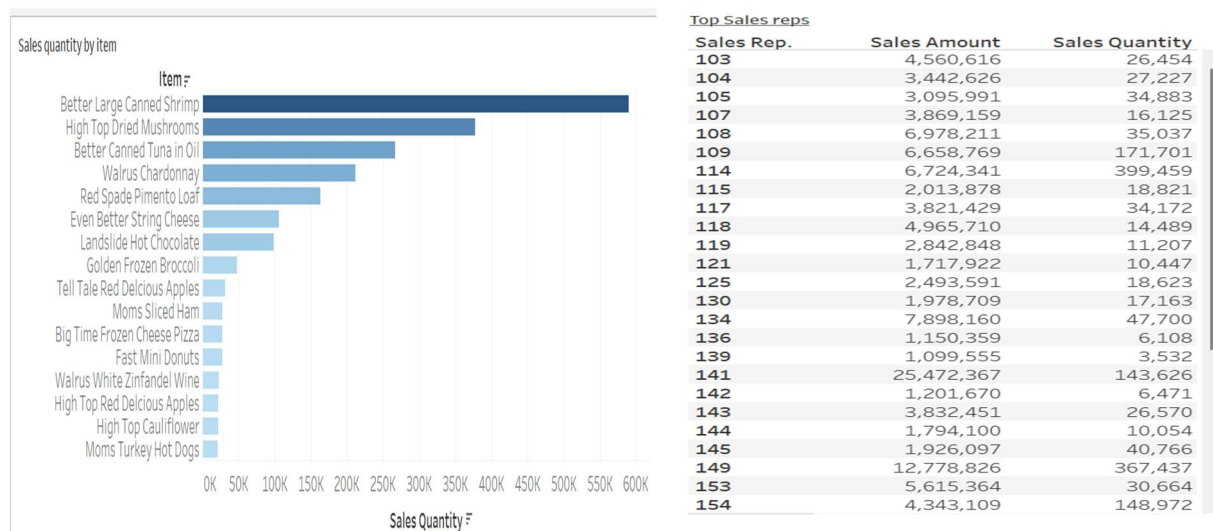**Tableau Server - Public Cloud (IaaS)**

- Full control of software on managed hardware
- Puts infrastructure in same place as data (for migration to cloud)
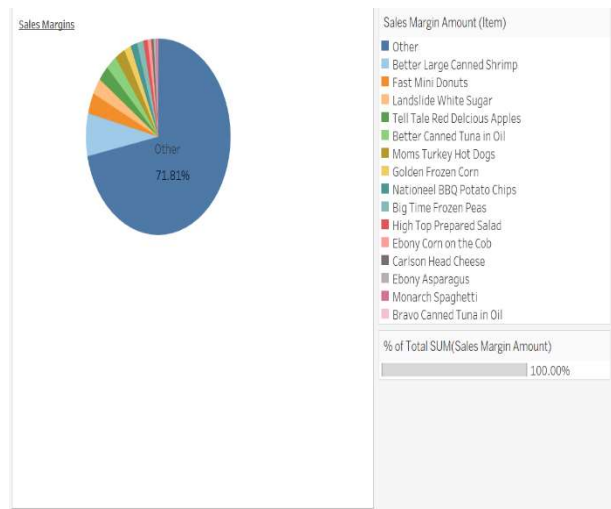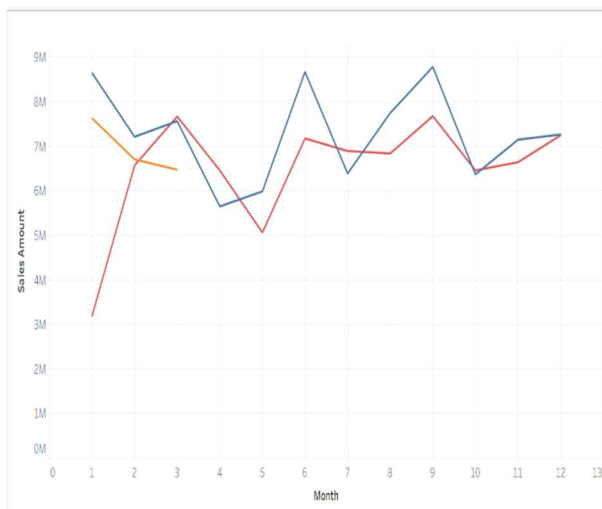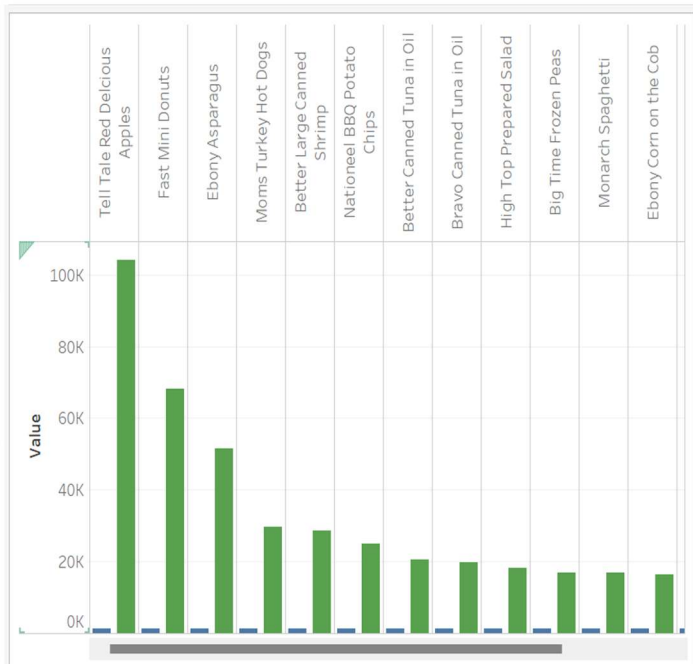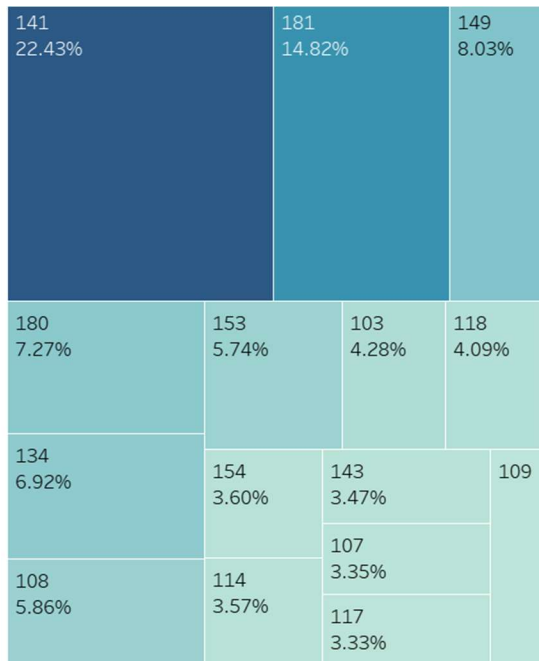- Flexibility to spin up/down hardware as needed

- Need dedicated administrators to manage software
- Additional infrastructure needed to access off-network (mobile, external)

Tableau Online (SaaS)

- Fully hosted solution (hardware, software upgrades)
- Fast to deploy
- Easy for external audience to access
- Single-site in multi-tenant environment
- Cubes are not supported
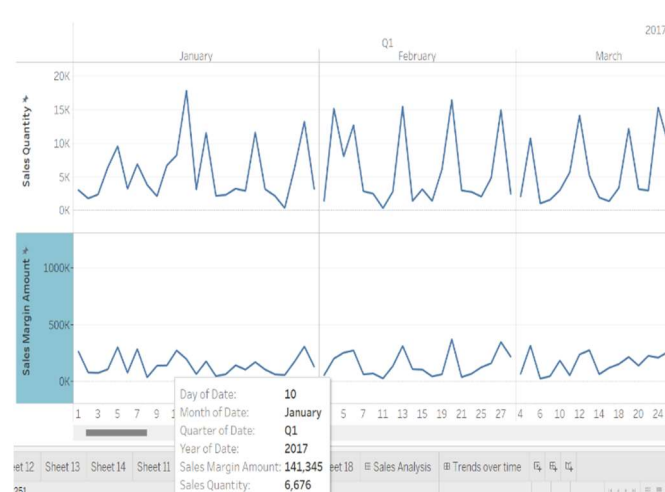- No guest account access

Depending on your organizational roles and responsibilities, Tableau Server should be installed by a systems administrator and the designated Tableau Server Administrator in coordination with the appropriate IT roles. For Tableau Online, you will integrate with your existing technology and configure the site settings. The Data & Analytics Survey, completed by business teams, identifies and prioritizes data use cases, audience size, and users. You will use the information collected in both surveys to plan your deployment strategy, including sizing, installation, and configuration of your Tableau Server or integration and configuration of Tableau Online. In addition to installing Tableau Server or configuring Tableau Online, administrators will also need to plan for the client software installation of Tableau Prep Builder, Tableau Desktop, Tableau Mobile, and Tableau Bridge for Tableau Online where applicable.

Top Sales reps

| Sales Rep.	Sales Amount	Sales Quantity
103	4,560,616	26,454
104	3,442,626	27,227
105	3,095,991	34,883
107	3,869,159	16,125
108	6,978,211	35,037
109	6,658,769	171,701
114	6,724,341	399,459
115	2,013,878	18,821
117	3,821,429	34,172
118	4,965,710	14,489
119	2,842,848	11,207
121	1,717,922	10,447
125	2,493,591	18,623
130	1,978,709	17,163
134	7,898,160	47,700
136	1,150,359	6,108
139	1,099,555	3,532
141	25,472,367	143,626
142	1,201,670	6,471
143	3,832,451	26,570
144	1,794,100	10,054
145	1,926,097	40,766
149	12,778,826	367,437
153	5,615,364	30,664
154	4,343,109	148,972

Sheet 16

Bravo Large Canned Shrimp
High Top Dried Mushrooms
over Manicotti Ebony Prepared Salad
Better Fancy Canned Sardines
Red Spade Foot-Long Hot Dogs
Cutting Edge Foot-Long Hot Dogs
Big Time Frozen Cheese Pizza
Imagine Popsicles Moms Sliced Turkey

Day of Date:	10
Month of Date:	January
Quarter of Date:	Q1
Year of Date:	2017
Sales Margin Amount:	141,345
Sales Quantity:	6,676

Sheet 12 Sheet 13 Sheet 14 Sheet 11 eet 18 ⊞ Sales Analysis ⊞ Trends over time

Custkey	Year of 1stOrder	Avg. Amt/Order	Orders/Month	Avg Spending/Month
10025919	2017	7,099	38	270,951
10019194	2017	6,888	38	261,754
10026081	2017	5,097	23	114,681
10018871	2017	5,354	22	119,578
10026606	2017	11,428	22	249,506
10021986	2017	919	22	20,069
10008638	2017	796	20	15,780
10017638	2017	3,746	20	73,361
10020184	2017	13,142	19	249,691
10023524	2017	4,265	18	78,193
10025552	2017	4,541	18	80,226
10023793	2017	1,618	18	28,586
10013080	2017	1,347	17	23,007
10002506	2017	3,574	14	50,035
10017036	2017	6,542	14	88,864
10021485	2017	74,982	13	949,767
10015793	2017	6,443	12	77,318
10025025	2017			80,822
10013238	2017			74,021
10012226	2017			153,514

Custkey:	10015793
Year of 1stOrder:	2017
Avg. Amt/Order:	6,443

Sales Analysis Dashboard:
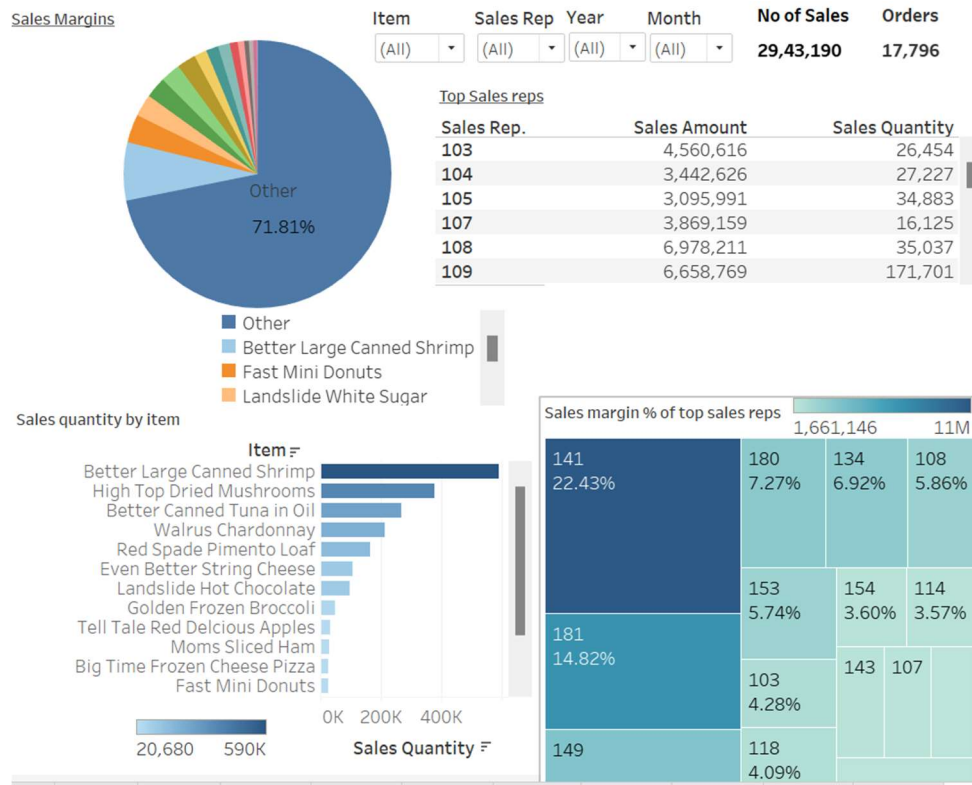
Time series analysis dashboard: