**High – Level Design (HLD)**

**ANALYZING AMAZON SALES DATA**

**GAUTAM VORA**

**Document Version Control**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date Issued** | **Version** | **Description** | **Author** |
| 23/05/2023 | 1.0 | Abstract,  Introduction,  General Description | Gautam Vora |
| 24/05/2023 | 1.1 | Design Detail, KPI, Deployment | Gautam Vora |
| 26/05/2023 | 1.2 | Final Revision | Gautam Vora |

**Contents**

Document Version Control .................................................................................................. 2

Abstract ............................................................................................................................... 3

1. Introduction ......................................................................................................................4

1.1 Why this High-Level Design Document? ....................................................................4

1.2 Objective of HLD .........................................................................................................4

1.3 Scope of HLD .............................................................................................................4

1. General Description .........................................................................................................5
   1. Product Perspective & Problem Statement ...........................................................5
   2. Data Requirements .................................................................................................5
   3. Tools Used ..............................................................................................................6
2. Design Details
   1. Functional Architecture ...........................................................................................8
   2. Error Handling ……………………………………….…………….………………………………………………9
   3. Optimization ………………………………………..……………………………………………………………..
3. KPI’s
   1. KPI’s (Key Performance Indicators) ................................................................10
4. Deployment ...............................................................................................................11

**Abstract**

Amazon data analytics refers to the process of analyzing consumer buying and preference patterns, sales and many other data points, which helps to make well-informed and data-driven decisions. By analyzing specific metrics, you can drastically increase Amazon sales and also improve your overall Amazon sales strategy. Amazon Sales data refers to sales, high-performing sellers, and several other data points. There are millions of Amazon sellers around the world. Nearly half of them are self-employed and live off their ecommerce/retail businesses (47%), and 22% earn income from their Amazon businesses alone. Amazon sales data Analysis focuses on the process of analyzing consumer behavior, sales, and several other attributes in order to make improved, data-driven decisions. It is key to successfully sustaining their businesses and earning profits and for this purpose, they analyze different metrics like sales, Sales Quantity, Discount rate, Sales over years etc. Analysis of the sales data the main factor that contributes to sellers improving their business and increasing their revenue. They can better understand the market trends and customers’ buying behaviors and help them cater to what the customers really want.

Data analysis can help them to understand their business in a quiet different manner and helps to improve the quality of the service by identifying the weak areas of the business. This study demonstrates the how different analysis help to make better business decisions and help analyze customer trends and satisfaction, which can lead to new and better products and services. Different analysis performed to get the key insights from this data based on which business decisions will be taken.

**1.Introduction**

This document will be used for documenting High-level designs of project.

* 1. **Purpose of the Document**

* Describe different design approaches.
* helps to make well-informed and data-driven decisions.
* Describe different analysis approaches based on variety of Use Cases.
* Describe third party components/tools required for the system.
* Present complete Process Flow followed for this project.

**1.2 Objective of HLD**

* 1. The objective of this paper is to carefully study the sales made by Amazon and present that analysis in the form of different Data. Seeing ways of doing things
  2. To provide an overview of the entire system.
  3. To provide an introduction of the Problem’s way of seeing things / sensible view of what is and is not important & Statement, Data Needed things, Tools used, and many more.
  4. To provide a module-wise breakup of the whole system.

**1.3 Scope of HLD**

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**

Sales management has gained importance to meet increasing competition and the need for improved methods of distribution to reduce costs and increase profits. Sales management today is the most important function in a commercial and business enterprise.

The goal of the project is to Carefully study Amazon Sales data to get big data that will help in bringing changes in a business in the future. It will help to show/tells about flaws in the business model or in the way that one is going about conducting business. Sellers will be able to clearly see where they're losing money, and what the problem is, and reduce their losses in the same way/in that way. It helps come up with (related to a plan to reach a goal) solutions to problems. This project aims to provide a visual understanding of the data using Microsoft Power BI.

## **2.2 Data Requirements**

Data Requirement completely depend on our problem.

* In this project, to perform analysis, we are using datasets that is provided by iNeuron Intelligence Pvt. Ltd.
* We make a use of those datasets as per the requirement and the problem statement.
* The features which are taken into consideration are:

Some of the important features are:

|  |  |
| --- | --- |
| Name | Description |
| Region | Region Name from which order was placed |
| Country | Country Name from which order was placed |
| Item Type | Product Category from which Items were ordered |
| Sale Channel | Mode of a sale |
| Order Priority | Delivery mode opted by the customer |
| Order Date | Date on which order was placed |
| Order ID | A unique order ID pertaining to each order placed |
| Ship Date | Date on which order expected to be delivered |
| Units Sold | Quantity of units sold |
| Unit Price | Selling Price per unit of a product |
| Unit Cost | Purchased cost per unit of a product |
| Total Revenue | Total Revenue/Sales made |
| Total Cost | Total cost which has invested |
| Total Profit | Total Profit which is made after deducting the Total Cost from Total Revenue |

**2.3 Tools Used**

* Pandas, Microsoft Excel were used for Data Manipulation & Pre-processing.
* GitHub is used as version control system.
* These are all the tools and libraries were used to build the whole framework.
* Jupyter Notebook is used as IDE.
* Python is the Programming Language used.
* EDA is done using Numpy& Pandas.
* Visualizations were done using Matplotlib.

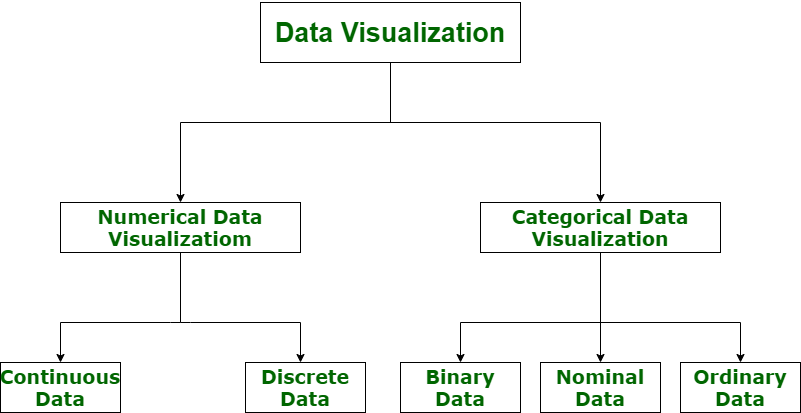




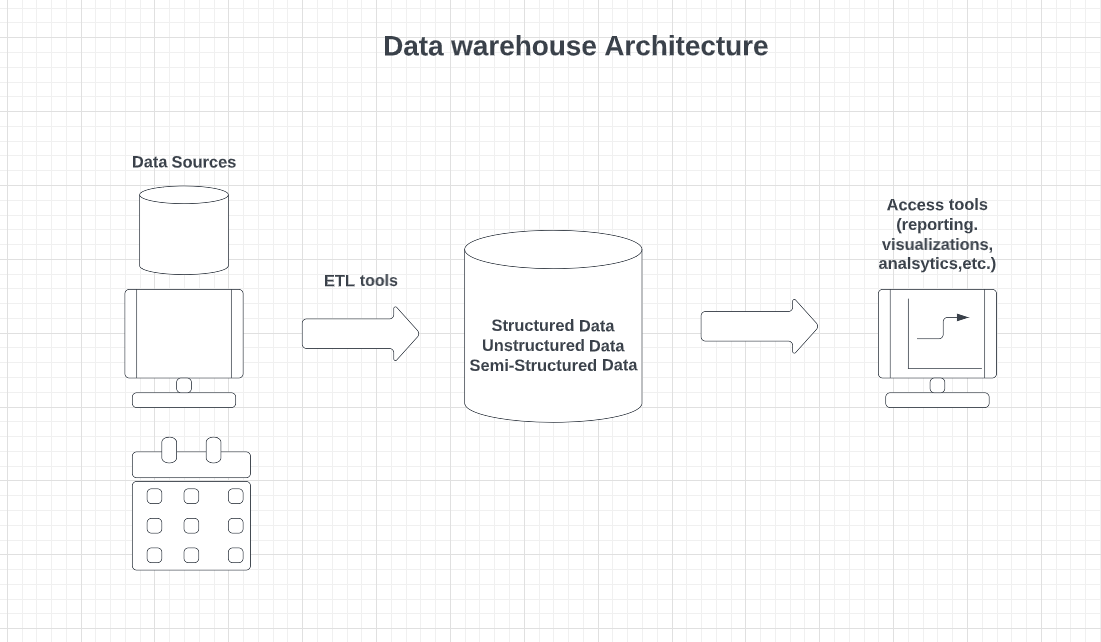
  

1. **Design Details**

**Data Visualization**



**3.1 Functional Architecture**



## **3.2 Optimization**

**1. 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

##### **2. 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.

##### **3. 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 including 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-ofdate filters) can take advantage of the indexing properties in your database and are faster than discrete data 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)

**4. 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 KPI**

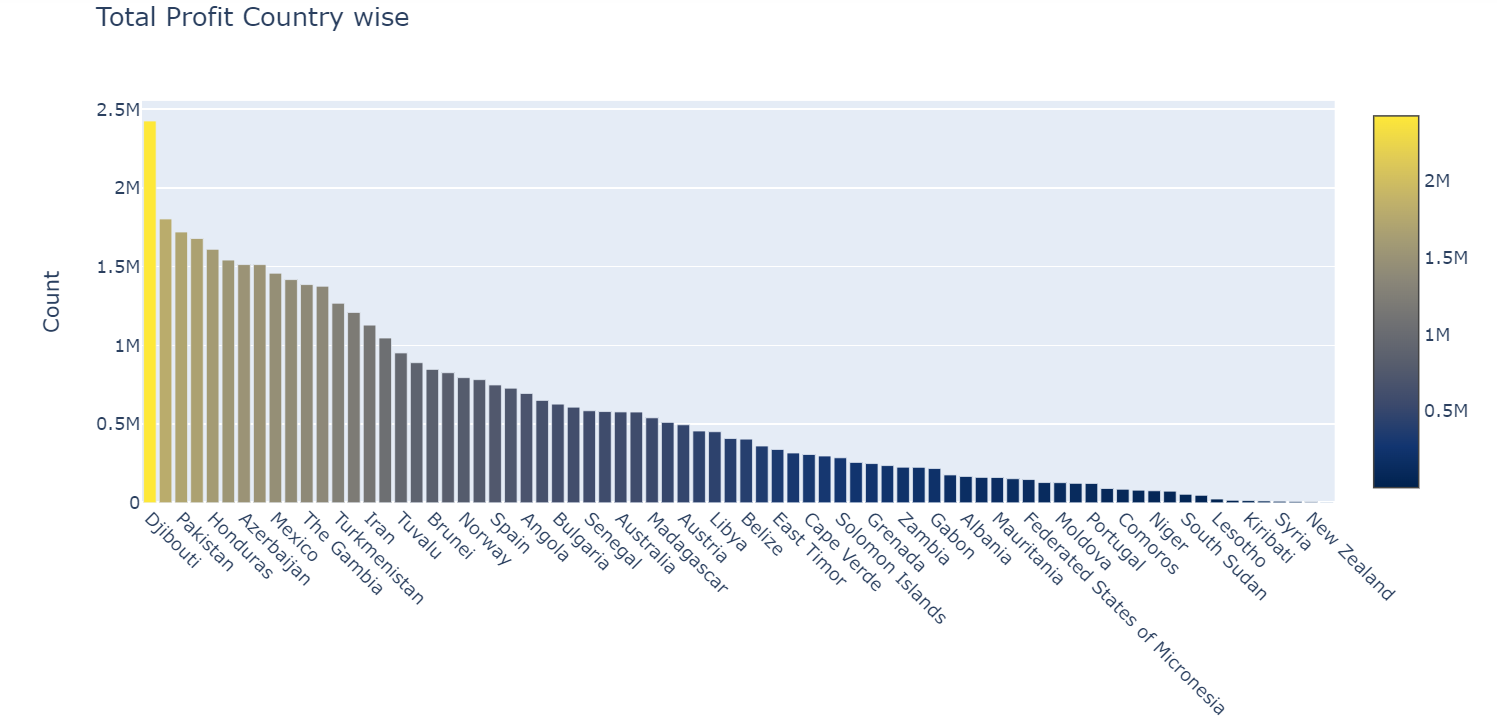
Dashboards will be implemented to display and indicate certain KPIs and relevant indicators for the disease.

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 Sales Data and its relationships with different metrics.

1. Yearly, Quarterly, Monthly Ups and Downs in Sales & Profits.
2. Items That Generated Highest Sales, Profit etc.
3. Top 5 Items that generated highest Sales and Top 5 Items by Quantity.
4. Bottom 5 Items that generated Lowest Sales and Bottom 5 Items by Quantity.
5. Forecasting.

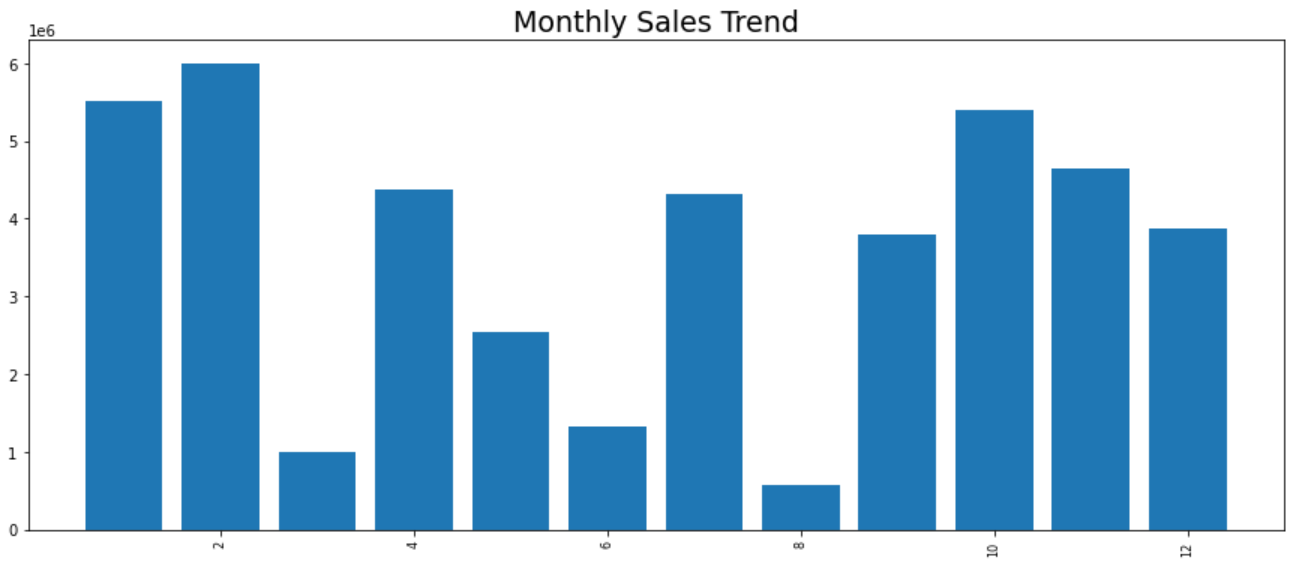


**5 Deployment**

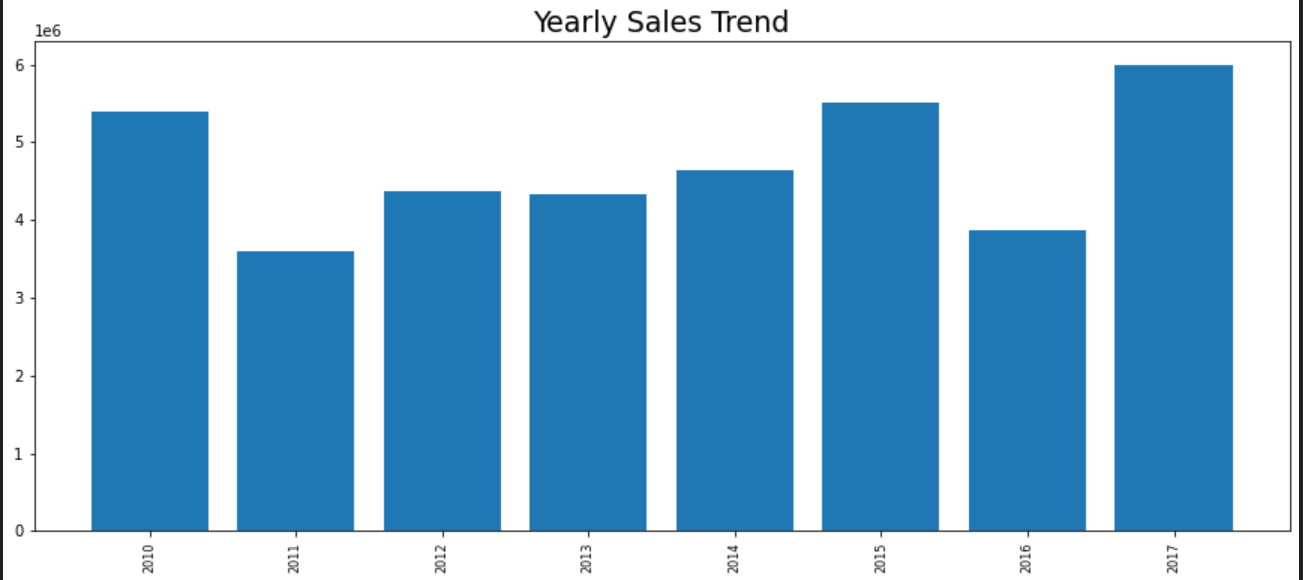
Prioritizing data and analytics couldn’t come at a better time. Your company, no matter what size, is already collecting data and most likely Analysing just a portion of it to solve business problems, gain competitive advantages, and drive enterprise transformation. With the explosive growth of enterprise 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.

Power BI prioritizes choice in flexibility to fit, rather than dictate, your enterprise architecture. Power BI Desktop and Power BI Service leverage your existing technology investments and integrate them into your IT infrastructure to provide a self-service, modern analytics platform for your users. With on-premises, cloud, and hosted options,

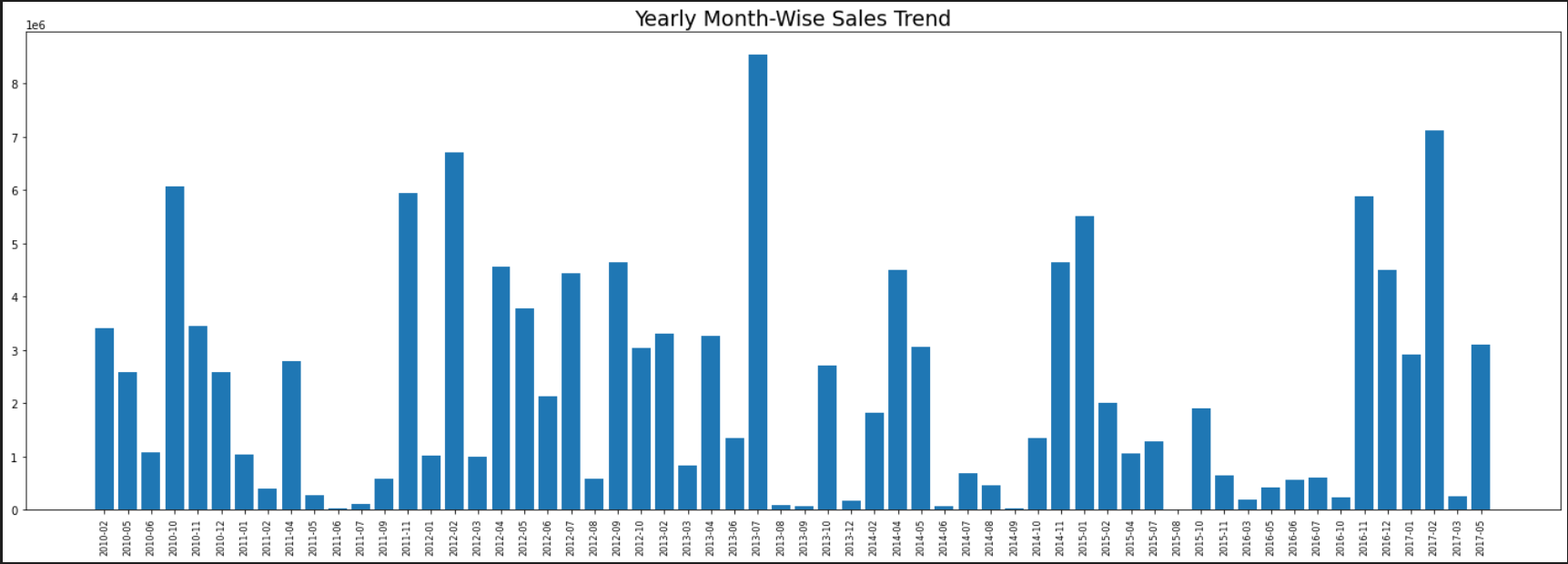
1. **Monthly Sales Trends :-**



1. **Yearly Sales Trend**



1. **Yearly Month-Wise Sales Trend**



**2.Total Revenue**

