High Level Design (HLD) Amazon Sales Data Analysis

**Siri S**

# Document Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| **Date Issued** | **Version** | **Description** | **Author** |
| 22 June, 2024 | 1.0 | First Version of Complete HLD | Siri S |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Content**

# Document Version Control…………………………………………………………………………………3

1. [Introduction](#_tyjcwt) 4
   1. Project objectives and scope4
   2. Importance of sales management in e-commerce4
2. Dataset Overview4
   1. Description of the provided dataset4
   2. Relevance to the project objectives4
3. Technologies and Tools5
   1. Overview of the programming languages and tools used5
   2. Justification for the selection of specific technologies5
4. Data Extraction and Preprocessing6
   1. Steps involved in data extraction from the dataset6
   2. Data cleaning and preprocessing techniques applied6
5. Exploratory Data Analysis (EDA)6

5.1 Analysis of sales trends6

5.2 Distribution analysis of key variables6

5.3 Calculation and interpretation of key metrics6

6 Sales Trend Analysis7

6.1 Monthly sales trend analysis7

6.2 Yearly sales trend analysis7

6.3 Yearly-monthly sales trend analysis7

7 Key Metrics and Factors7

7.1 Average resolution time7

7.2 Average requester wait time7

7.3 Analysis of satisfaction scores7

8 Visualization8

8.1 Visual representations of sales trends and distributions8

9 Conclusion and Recommendations9

9.1 Summary of findings from the analysis9

9.2 Actionable recommendations for sales management9

10 Future Work9

10.1 Potential enhancements or extensions to the project9

10.2 Areas for further research and analysis9

# 

# 

# Introduction

# 1.1 Project Objectives and Scope:

* The primary objective of this project is to analyze Amazon sales data comprehensively to extract meaningful insights that can inform business decisions and strategies.
* Scope includes examining sales trends, customer satisfaction levels, and operational efficiency metrics derived from the dataset.

# 1.2 Importance of Sales Management in E-commerce:

* Sales management is vital in e-commerce due to increasing competition and the need for cost reduction and profit maximization.
* Effective sales management involves optimizing various aspects such as inventory management, pricing strategies, and customer support processes.

**2. Dataset Overview**

**2.1 Description of the Provided Dataset:**

* The dataset comprises Amazon sales data, including transaction details, customer feedback, and support query resolution times.
* It includes attributes such as sales transaction IDs, product details, customer information, timestamps, satisfaction scores, and resolution times.

**2.2 Relevance to the Project Objectives:**

* The dataset is highly relevant as it provides crucial insights into sales performance, customer behavior, and service quality, aligning with the project's objectives of analyzing Amazon sales data.

Amazon Sales Data Analysis

# 3. ****Technologies and Tools****

**3.1 Overview of Programming Languages and Tools Used:**

Python is chosen as the primary programming language for its versatility and extensive libraries for data analysis.

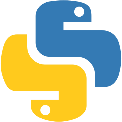
Libraries such as pandas, NumPy, and scikit-learn are utilized for data manipulation, analysis, and machine learning tasks.

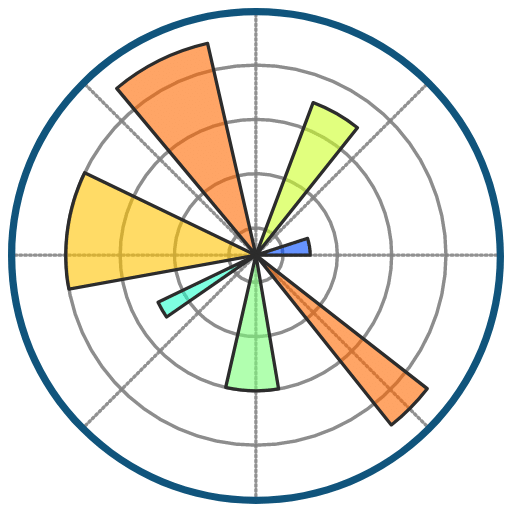
Matplotlib, seaborn are employed for data visualization.

## Tools used

Business intelligence tools like NumPy, Pandas, Matpoltlib, Seaborn used to visualize.







* 1. **Justification for Technology Selection:**
* Python is widely used in data science and offers powerful tools for data manipulation, analysis, and visualization.
* Pandas provides efficient data structures and functions for data manipulation, making it ideal for working with structured data like the Amazon sales dataset.
* Matplotlib and seaborn offer a wide range of visualization options for exploring and presenting data effectively.

# 4 Data Extraction and Preprocessing

# 4.1 Steps Involved in Data Extraction:

* Loading the dataset using pandas' read\_excel() function to create a DataFrame.
* Extracting relevant columns and rows based on the project objectives and requirements.

**4.2 Data Cleaning and Preprocessing Techniques Applied:**

* Handling missing values by either imputing them or removing rows/columns with missing values.
* Removing duplicates to ensure data integrity.
* Converting data types to appropriate formats (e.g., converting date strings to datetime objects).
* Checking for and handling outliers that may affect the analysis results.
* Encoding categorical variables if necessary for analysis or modeling purposes.

**5. Exploratory Data Analysis (EDA)**

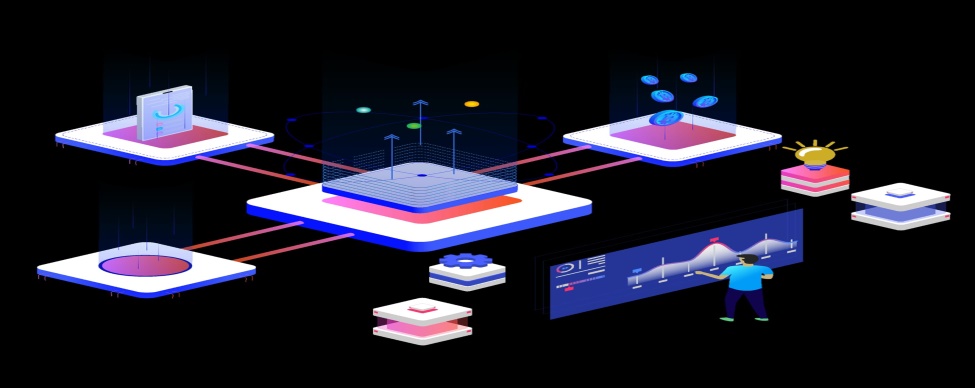
**5.1 Analysis of Sales Trends:**

* Analyzing sales trends on a monthly, yearly, and yearly-monthly basis to identify patterns, seasonality, and trends.
* Visualizing sales trends using line plots, bar plots, or heatmaps to understand variations over time.

**5.2 Distribution Analysis of Key Variables:**

* Analyzing the distribution of key variables such as sales volumes, resolution times, and satisfaction scores to identify central tendencies and variability.
* Visualizing distributions using histograms, box plots, or kernel density estimates.

**5.3 Calculation and Interpretation of Key Metrics:**

* Calculating key metrics such as average resolution time, average requester wait time, and satisfaction scores.
* Interpreting key metrics to assess operational efficiency, customer satisfaction levels, and overall business performance.

**6. Sales Trend Analysis**

**6.1 Monthly Sales Trend Analysis:**

* Analyzing sales performance on a month-by-month basis to identify seasonal patterns, peak months, and trends over time.
* Visualizing monthly sales trends using line plots or bar plots to highlight fluctuations and trends.

**6.2 Yearly Sales Trend Analysis:**

* Analyzing sales performance on a yearly basis to identify long-term growth patterns and trends.
* Visualizing yearly sales trends using line plots or bar plots to assess overall performance over time.

**6.3 Yearly-Monthly Sales Trend Analysis:**

* Analyzing sales trends for each month across different years to identify any specific trends within months.
* Visualizing yearly-monthly sales trends using heatmaps or line plots to observe seasonal variations or changes over time.

**7. Key Metrics and Factors**

**7.1 Average Resolution Time:**

* Calculating the average resolution time to assess the efficiency of customer support processes.
* Analyzing trends and variations in resolution times to identify potential areas for improvement.

**7.2 Average Requester Wait Time:**

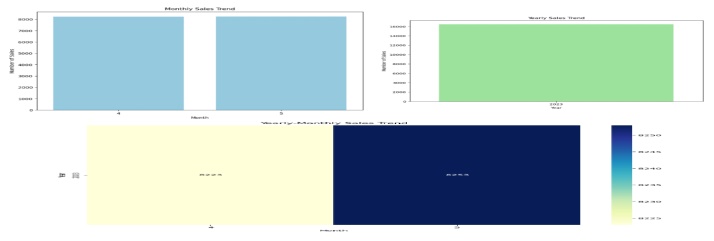
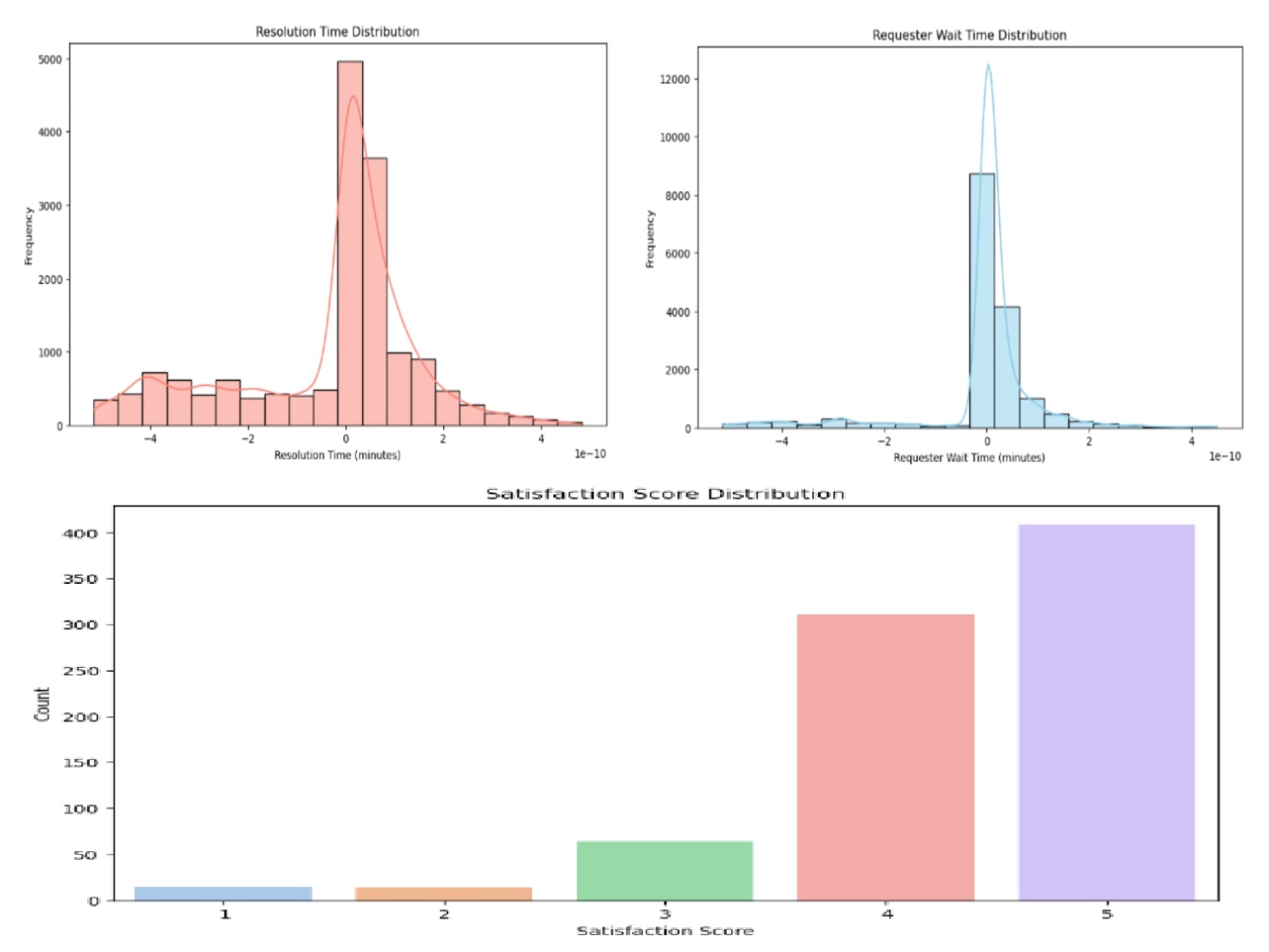
* Calculating the average requester wait time to evaluate the responsiveness of support services.
* Assessing trends and variations in requester wait times to optimize service delivery.

**7.3 Analysis of Satisfaction Scores:**

* Analyzing customer satisfaction scores to gauge customer feedback and satisfaction levels.
* Identifying factors influencing satisfaction scores and addressing areas for improvement.

**8. Visualization**

**8.1 Visual Representations of Sales Trends and Distributions:**

* Creating visualizations such as line plots, bar plots, histograms, or heat maps to visualize sales trends and distributions effectively.
* Utilizing interactive visualization tools to explore data dynamically and identify insights.

**9. Conclusion and Recommendations**

**9.1 Summary of Findings from the Analysis:**

* Summarizing key findings and insights derived from the analysis of Amazon sales data, including sales trends, key metrics, and factors influencing performance.

**9.2Actionable Recommendations for Sales Management:**

* Providing actionable recommendations for improving sales management processes, optimizing inventory management, and enhancing customer satisfaction based on the analysis.

**10. Future Work**

**10.1 Potential Enhancements or Extensions to the Project:**

* Identifying potential enhancements or extensions to the project, such as incorporating additional datasets or conducting further analysis on specific aspects of sales management.
* Exploring advanced analytical techniques or machine learning models for sales forecasting or customer segmentation.

**10.2 Areas for Further Research and Analysis:**

* Identifying areas for further research and analysis, such as exploring customer behavior patterns, conducting sentiment analysis on customer feedback data, or investigating the impact of external factors on sales performance.