**Low-Level Document (LLD)**

**Amazon Sales Data**

**Analysis**

Revision Number: 0.1

Deepraj Arya

**DOCUMENT CONTROL**

**Change Record:**

|  |  |  |  |
| --- | --- | --- | --- |
| **VERSION** | **DATE** | **AUTHOR** | **COMMENTS** |
| 0.1 | 12- Oct -  2023 | Deepraj Arya | Introduction and architecture defined |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Reviews:**

|  |  |  |  |
| --- | --- | --- | --- |
| **VERSION** | **DATE** | **REVIEWER** | **COMMENTS** |
| 0.1 |  |  |  |

**Approval Status:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **VERSION** | **REVIEW**  **DATE** | **REVIEWED BY** | **APPROVED BY** | **COMMENTS** |
|  |  |  |  |  |

**Contents**

# Project Overview

# Data Loading and Preprocessing

# Data Loading

# Data Preprocessing

# Exploratory Data Analysis

# Time Series Analysis

# Feature Engineering

1. **Project Overview**

The Amazon Sales Data Analysis project aims to analyze sales data for Amazon products, gaining insights into sales trends, profitability, and key performance metrics. This LLD will provide a detailed breakdown of the code and technical aspects involved in the project.

1. **Data Loading and Preprocessing**
   1. **Data Loading**

* The project begins with data loading from a CSV file named "Amazon\_Sales\_Records.csv" using the Pandas library.
* The data is stored in a Data Frame named "df."
  1. **Data Preprocessing**

Data Type Conversion

* The data type of the 'Order Date' and 'Ship Date' columns is converted to datetime using the pd.to\_datetime function.

Missing Values

* Check for missing values in the dataset. Fortunately, there are no missing values in any of the columns.

Data Description

* Generate a summary of the dataset using the describe method to get statistics on numerical columns.

Copy of Data

* Create a copy of the original data as "sales\_data" to maintain the integrity of the original dataset.

Feature Engineering

* Create new columns such as 'Order\_Year,' 'Order\_Month,' 'Order\_Quarter,' and 'Order\_Day' from the 'Order Date' column to facilitate time-based analysis.

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

**Visualizing Sales Trends**

* Visualize sales trends over time using line plots for year, month, and day.
* Calculate and analyze correlations between 'Order Year' and 'Units Sold,' 'Order Month' and 'Units Sold,' and 'Order Day' and 'Units Sold.'

**Grouping and Aggregation**

* Group the data by 'Item Type' and calculate the sum of 'Units Sold' for each item type.
* Group the data by 'Item Type' and 'Sales Channel' and calculate the sum of 'Units Sold' for each combination.

**Data Distribution**

* Select numerical columns for distribution analysis and create histograms to visualize the distribution of each numerical feature.
* Identify and handle outliers in the 'Unit Price,' 'Unit Cost,' 'Total Revenue,' 'Total Cost,' and 'Total Profit' columns using the Interquartile Range (IQR) method.

**Correlation Analysis**

* Calculate the correlation matrix for numerical columns.
* Create a heatmap to visualize the correlations between features.

1. **Time Series Analysis**

**Time Series Transformation**

* Convert the 'Order Date' column to a datetime format.
* Set 'Order Date' as the index for time series analysis.

**Profitability Analysis**

* Visualize the distribution of 'Total Profit.'
* Analyze 'Total Profit' by sales channel.
* Identify and handle extreme outliers in the 'Online' sales channel.

1. **Feature Engineering**

**Additional Features**

* Calculate 'Profit Margin (%)' as (profit/revenue) \* 100.
* Calculate 'Sales Growth Rate (%)' as the percentage change in total revenue compared to the previous month.
* Calculate 'Cost-to-Revenue Ratio (%)' as (cost/revenue) \* 100.

**Visualization of Newly Created Features**

* Visualize 'Sales Growth Rate,' 'Profit Margin,' and 'Cost-to-Revenue Ratio' over time.

**Month-Wise Analysis**

* Analyze and visualize metrics on a month-to-month basis.
* Calculate and visualize average 'Sales Growth Rate,' 'Total Profit,' 'Total Revenue,' and 'Total Cost' for each month.

**Year-Wise Analysis**

* Analyse and visualize metrics on a year-to-year basis.
* Calculate and visualize the average 'Sales Growth Rate,' 'Total Profit,' 'Total Revenue,' and 'Total Cost' for each year.