****

**Amazon Sales Data Analysis**

\*\*Last Revised Date\*\*: [24-08-2023]

**Problem Statement:**

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

DOCUMENT CONTROL

|  |
| --- |
| **| VERSION | DATE | AUTHOR | COMMENTS |** |
| **| 0.1 | [ 19-08-2023] | Iqra | Initial architecture and objectives |** |
| **| 0.2 | [20-08-2023] | Iqra | Data preparation and cleaning |** |
| **| 0.3 | [21-08-2023] | Iqra | Exploratory data analysis and metrics |** |
| **| 0.4 | [22-08-2023] | Iqra | Customer segmentation and conclusions |** |

**Table of Contents**

1. **Introduction**
   * Background and Problem Statement
   * Objectives
2. **Data Preparation**
   * Loading Libraries
   * Loading Dataset
   * Data Cleaning and Preprocessing
     + Handling Null Values
     + Data Type Conversion
     + Data Quality Checks
3. **Exploratory Data Analysis (EDA)**
   * Understanding Data Attributes
     + Summary Statistics
     + Top Selling Items by Sales
     + Top Mean Discount by Item
     + Items with High Profit
     + Items with Discounts and Without Discounts
     + Yearly Sales Price Trends
     + Discount Spending Trends
     + Monthly Discount and Sales Patterns
4. **Optimizing Pricing and Discounts**
   * Key Metrics
     + Discount Effectiveness
     + Sales Margin
     + Sales Quantity and Price
   * Correlation Analysis
   * Visualization
   * Hypothesis Testing
5. **Inventory Management**
   * Monthly Sales Trend
   * Monthly Discount Trend
   * Seasonal Decomposition of Sales
   * Recency, Frequency, and Monetary (RFM) Analysis
     + Data Preparation
     + RFM Clustering
6. **Conclusion**
   * Summary of Findings
   * Business Insights
   * Future Steps

**1. Introduction**

**Background and Problem Statement**

The analysis focuses on Amazon's sales data to optimize pricing and discounts for improved sales management. The goal is to reduce costs and increase profits through strategic data-driven decisions. Key aspects include optimizing pricing and discounts, effective inventory management, customer segmentation, sales forecasting, promotion analysis, process efficiency, sales performance evaluation, and cross-selling/upselling.

**Objectives**

The primary objectives of the analysis are as follows:

* Understand the sales data and its attributes.
* Identify trends, patterns, and insights to optimize pricing and discount strategies.
* Evaluate the impact of discounts on sales and margins.
* Conduct hypothesis testing to validate findings.
* Provide recommendations for inventory management and customer segmentation.

**2. Data Preparation**

**Loading Libraries**

The necessary Python libraries are imported, including pandas, numpy, matplotlib, seaborn, and statsmodels.

**Loading Dataset**

The Amazon sales dataset is loaded using pandas from an Excel file.

**Data Cleaning and Preprocessing**

The dataset undergoes thorough cleaning and preprocessing, including handling null values, data type conversion, and data quality checks. Null values are dropped, and data types are standardized. Columns with low variances are dropped, and categorical columns are converted to the category data type.

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

EDA is performed to gain insights into the dataset.

**Understanding Data Attributes**

Summary statistics are calculated, and visualizations are created for various attributes such as top selling items by sales, top mean discount by item, items with high profit, items with and without discounts, yearly sales price trends, discount spending trends, and monthly discount and sales patterns.

**4. Optimizing Pricing and Discounts**

This section focuses on optimizing pricing and discount strategies.

**Key Metrics**

Key metrics related to discount effectiveness, sales margin, and sales quantity and price are calculated.

**Correlation Analysis**

Correlation analysis is conducted to understand relationships between attributes such as discount amount, sales amount, and sales margin.

**Visualization**

Visualizations are created to visualize the relationships between discount amount, sales amount, and sales margin amount.

**Hypothesis Testing**

Hypothesis testing is performed to validate findings related to the impact of changes in pricing and discount strategies on sales and margins.

**5. Inventory Management**

This section explores inventory management using time-based analysis.

**Monthly Sales Trend**

Monthly sales trends are plotted to visualize sales patterns over time.

**Monthly Discount Trend**

The trend of monthly discounts given is plotted to observe patterns in discount distribution.

**Seasonal Decomposition of Sales**

Seasonal decomposition is performed to separate the data into its underlying components: observed, trend, seasonal, and residual.

**Recency, Frequency, and Monetary (RFM) Analysis**

RFM analysis is conducted to segment customers based on their recency of purchase, frequency of purchase, and monetary value of purchases. Clustering techniques are applied to group customers into segments.

**6. Conclusion**

The analysis concludes with a summary of findings, key business insights, and recommendations for Amazon's sales management strategies based on the insights gained from the data analysis. Future steps for further exploration and improvement are also discussed.