**Customer Segmentation and Sales Analysis Documentation**

**1. Introduction**

This document provides a comprehensive analysis of the Online Retail dataset, focusing on customer segmentation using RFM (Recency, Frequency, Monetary) analysis and K-Means clustering. It also includes insights into top-selling and least-selling products, sales trends, and business recommendations based on data-driven insights.

**2. Data Cleaning & Preparation**

**Data Source**

* **Dataset**: Online Retail Dataset
* **Source**: [UCI Machine Learning Repository](https://archive.ics.uci.edu/ml/datasets/Online+Retail)

**Data Cleaning Steps:**

1. **Removed missing Customer IDs** to ensure accurate customer-level analysis.
2. **Filtered out transactions with negative or zero quantity and unit prices** to remove invalid sales records.
3. **Created a 'Total Price' column**: TotalPrice=Quantity×UnitPriceTotalPrice = Quantity × UnitPrice
4. **Saved cleaned data as** OnlineRetail\_Cleaned.csv for further processing in Power BI and analysis.

**3. Customer Segmentation: RFM & K-Means Clustering**

**RFM Analysis:**

* **Recency (R)**: Number of days since the last purchase.
* **Frequency (F)**: Total number of transactions.
* **Monetary (M)**: Total spending amount.
* Segmented customers into three categories based on their RFM scores:
  + **High-Value Customers**: 0.3%
  + **Medium-Value Customers**: 74.83%
  + **Low-Value Customers**: 24.87%

**K-Means Clustering for Enhanced Segmentation:**

* Applied **K-Means clustering** to segment customers based on purchasing patterns.
* Enhanced the RFM model by fine-tuning customer segmentation.

**4. Sales Insights**

**Top 5 Best-Selling Products (Based on Quantity Sold)**

| **Product** | **Units Sold** |
| --- | --- |
| JUMBO BAG RED RETROSPOT | 46,181 |
| MEDIUM CERAMIC TOP STORAGE JAR | 77,916 |
| PAPER CRAFT, LITTLE BIRDIE | 80,995 |
| WHITE HANGING HEART T-LIGHT HOLDER | 36,725 |
| WORLD WAR 2 GLIDERS ASSTD DESIGNS | 54,415 |

**Visualization:** Bar chart showcasing the top 5 best-selling products.

**Bottom 5 Least-Selling Products**

| **Product** | **Units Sold** |
| --- | --- |
| AMBER BERTIE GLASS BEAD BAG CHARM | 1 |
| BLACK VINT ART DEC CRYSTAL BRACELET | 1 |
| FLOWER SHOP DESIGN MUG | 1 |
| HEN HOUSE W CHICK IN NEST | 1 |
| SET 36 COLOURING PENCILS DOILEY | 1 |

**Visualization:** Bar chart for least-selling products.

**5. Sales Trends Analysis**

**Sales by Country (Top 5)**

* Percentage of total sales contribution per country.
* **Bar Chart** displaying sales distribution.

**Yearly Sales Trend**

* **Line Chart** illustrating total sales per year.

**6. Business Recommendations**

Based on the analysis, the following strategic recommendations are proposed:

1. **Increase High-Value Customers**
   * Personalized marketing campaigns for high spenders.
   * Exclusive deals and early access promotions.
2. **Convert Medium-Value Customers into High-Value Customers**
   * Introduce bundle deals and loyalty rewards.
   * Provide targeted discount offers based on purchase history.
3. **Optimize Low-Performing Products**
   * Reduce stock for least-selling products.
   * Offer clearance sales to improve inventory turnover.
4. **Expand International Market Reach**
   * Focus on high-performing countries to maximize revenue.
   * Localized marketing campaigns to drive growth in key markets.

**7. Conclusion**

This analysis provides actionable insights to improve sales performance, enhance customer retention, and drive business growth. The findings will be integrated into a **Power BI Dashboard** for real-time monitoring and decision-making.

**Visuals**

**Top Five Country Based Upon Percentage**

A white background with black dots

AI-generated content may be incorrect.

**Yearly Wise Sales**

A white background with black text

AI-generated content may be incorrect.

The Reason for Use of Kmeans Algorithm



Thank You

Syed Asad