Project Report: Online Retail Segmentation

Name: Sher Ali Shahid

Roll No: sk23057

1. Introduction

The objective of this project is to analyze an online retail dataset to gain insights into customer behavior, segment customers based on their purchase frequency, calculate the average order value by country, identify potential churned customers, analyze product affinities, and explore time-based trends. By performing these analyses, we aim to provide actionable insights to improve customer engagement, tailor marketing strategies, and optimize business performance.

2. Dataset Overview

The dataset used in this project contains information about online retail transactions. Key columns include InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, and Country. This dataset serves as the foundation for conducting various analyses to understand customer behavior and business patterns.

3. Analysis and Findings

3.1 Customer Segmentation by Purchase Frequency

We segmented customers into three groups: High, Medium, and Low frequency customers based on their purchase frequency. This segmentation helps in identifying loyal customers and those who might need more attention. Using SQL queries, we calculated the number of purchases made by each customer and assigned them to appropriate segments. This information can be utilized for targeted marketing campaigns and personalized customer interactions.

3.2 Average Order Value by Country

We calculated the average order value for each country to identify regions with the highest order values. This analysis helps in identifying valuable customer bases and potential market expansion opportunities. By grouping transactions by country and performing calculations using SQL, we were able to uncover insights into the distribution of customer spending across different countries.

3.3 Customer Churn Analysis

Customer churn analysis helps in identifying customers who haven't made a purchase in a specific period. By defining a churn threshold (e.g., 6 months), we identified potential churned customers who might need re-engagement efforts. Utilizing SQL, we filtered customers based on their last purchase dates, providing a basis for targeted retention strategies.

3.4 Product Affinity Analysis

We conducted a product affinity analysis to identify products frequently purchased together. By calculating the correlation between product purchases using SQL queries, we discovered products that are often bought as a bundle. This information can guide product bundling strategies and enhance recommendations for cross-selling.

3.5 Time-based Analysis

Exploring trends over time is crucial for understanding customer behavior patterns. By grouping transactions by time intervals (e.g., months or quarters) and calculating total sales, we identified monthly or quarterly sales patterns. This analysis aids in seasonality predictions, inventory management, and campaign planning.

4. Conclusion and Insights

Through our analyses, we gained valuable insights into customer behavior and business dynamics. The segmentation by purchase frequency allows us to target customers more effectively, the average order value analysis highlights promising markets, churn analysis aids in customer retention efforts, product affinity analysis enhances cross-selling opportunities, and time-based analysis uncovers sales trends. These insights collectively provide a foundation for data-driven decision-making and strategic planning.

5. Future Directions

While our project covers a range of analyses, there are additional directions that can be explored. Future enhancements may include more advanced machine learning techniques for customer segmentation, predictive models for churn analysis, and deeper product association analysis. Additionally, incorporating external data sources for demographic insights could provide a more comprehensive view of customer behavior.

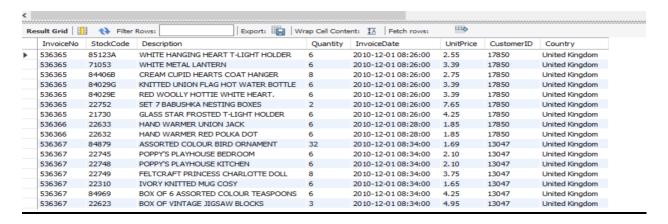
6. Acknowledgments

We acknowledge the dataset source [provide dataset source] for providing the data necessary for our analyses. We also appreciate the support of our mentors, peers, and resources that contributed to the successful completion of this project.

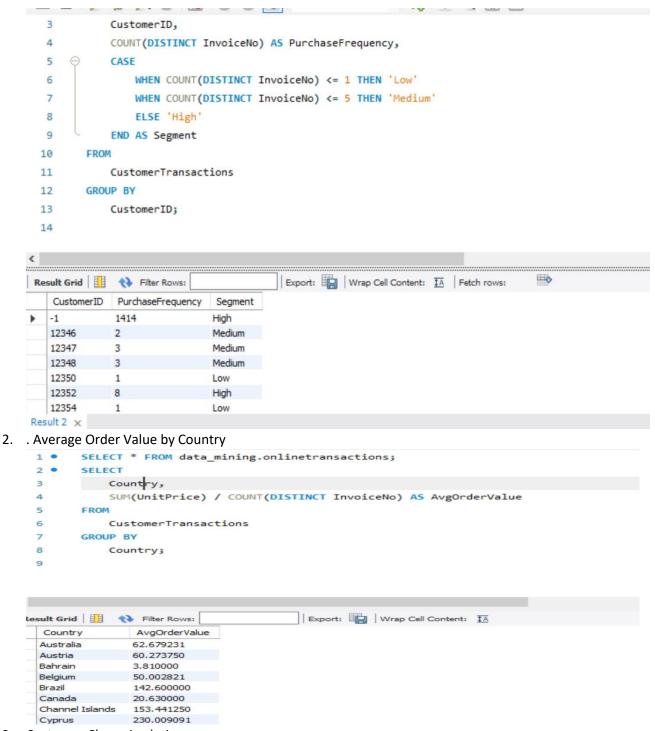
SQL Queries

Dataset

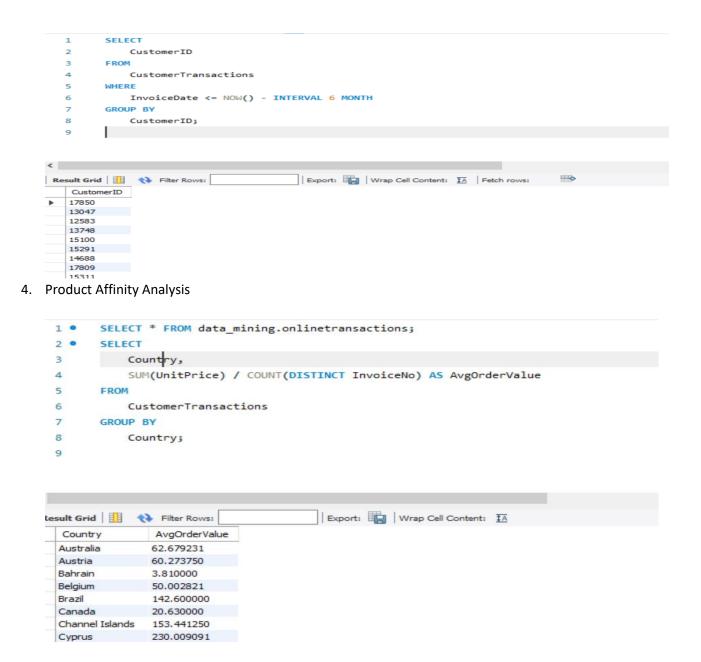
1 • SELECT * FROM data_mining.customertransactions;



1. Customer Segmentation by Purchase Frequency



3. Customer Churn Analysis



5. . Time-based Analysis

```
SELECT
1
      YEAR(InvoiceDate) AS Year,
2
      MONTH(InvoiceDate) AS Month,
3
4
        SUM(UnitPrice) AS TotalSales
5
    FROM
6
       CustomerTransactions
    GROUP BY
8
        Year, Month;
9
```

Export: Wrap Cell Content: IA

