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;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850	United Kingdom
536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850	United Kingdom
536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850	United Kingdom
536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850	United Kingdom
536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850	United Kingdom
536365	22752	SET 7 BABUSHKA NESTING BOXES	2	2010-12-01 08:26:00	7.65	17850	United Kingdom
536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	2010-12-01 08:26:00	4.25	17850	United Kingdom
536366	22633	HAND WARMER UNION JACK	6	2010-12-01 08:28:00	1.85	17850	United Kingdom
536366	22632	HAND WARMER RED POLKA DOT	6	2010-12-01 08:28:00	1.85	17850	United Kingdom
536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	2010-12-01 08:34:00	1.69	13047	United Kingdom
536367	22745	POPPY'S PLAYHOUSE BEDROOM	6	2010-12-01 08:34:00	2.10	13047	United Kingdom
536367	22748	POPPY'S PLAYHOUSE KITCHEN	6	2010-12-01 08:34:00	2.10	13047	United Kingdom
536367	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8	2010-12-01 08:34:00	3.75	13047	United Kingdom
536367	22310	IVORY KNITTED MUG COSY	6	2010-12-01 08:34:00	1.65	13047	United Kingdom
536367	84969	BOX OF 6 ASSORTED COLOUR TEASPOONS	6	2010-12-01 08:34:00	4.25	13047	United Kingdom
536367	22623	BOX OF VINTAGE JIGSAW BLOCKS	3	2010-12-01 08:34:00	4.95	13047	United Kingdom

1. . Customer Segmentation by Purchase Frequency

```
    CustomerID,
    COUNT(DISTINCT InvoiceNo) AS PurchaseFrequency,
    CASE
        WHEN COUNT(DISTINCT InvoiceNo) <= 1 THEN 'Low'
        WHEN COUNT(DISTINCT InvoiceNo) <= 5 THEN 'Medium'
        ELSE 'High'
    END AS Segment
FROM
    CustomerTransactions
GROUP BY
    CustomerID;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

CustomerID	PurchaseFrequency	Segment
-1	1414	High
12346	2	Medium
12347	3	Medium
12348	3	Medium
12350	1	Low
12352	8	High
12354	1	Low

Result 2

2. . Average Order Value by Country

```
SELECT * FROM data_mining.onlinetransactions;
SELECT
    Country,
    SUM(UnitPrice) / COUNT(DISTINCT InvoiceNo) AS AvgOrderValue
FROM
    CustomerTransactions
GROUP BY
    Country;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Country	AvgOrderValue
Australia	62.679231
Austria	60.273750
Bahrain	3.810000
Belgium	50.002821
Brazil	142.600000
Canada	20.630000
Channel Islands	153.441250
Cyprus	230.009091

3. Customer Churn Analysis

```sql
SELECT
    CustomerID
FROM
    CustomerTransactions
WHERE
    InvoiceDate <= NOW() - INTERVAL 6 MONTH
GROUP BY
    CustomerID;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

CustomerID
17850
13047
12583
13748
15100
15291
14688
17809
15311

4. Product Affinity Analysis

```sql
SELECT * FROM data_mining.onlinetransactions;
SELECT
    Country,
    SUM(UnitPrice) / COUNT(DISTINCT InvoiceNo) AS AvgOrderValue
FROM
    CustomerTransactions
GROUP BY
    Country;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Country	AvgOrderValue
Australia	62.679231
Austria	60.273750
Bahrain	3.810000
Belgium	50.002821
Brazil	142.600000
Canada	20.630000
Channel Islands	153.441250
Cyprus	230.009091

5. . Time-based Analysis

```sql
SELECT
    YEAR(InvoiceDate) AS Year,
    MONTH(InvoiceDate) AS Month,
    SUM(UnitPrice) AS TotalSales
FROM
    CustomerTransactions
GROUP BY
    Year, Month;

```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Year	Month	TotalSales
2010	12	332188.80
2011	1	172752.80
2011	2	127448.77
2011	3	171486.51
2011	4	116747.96