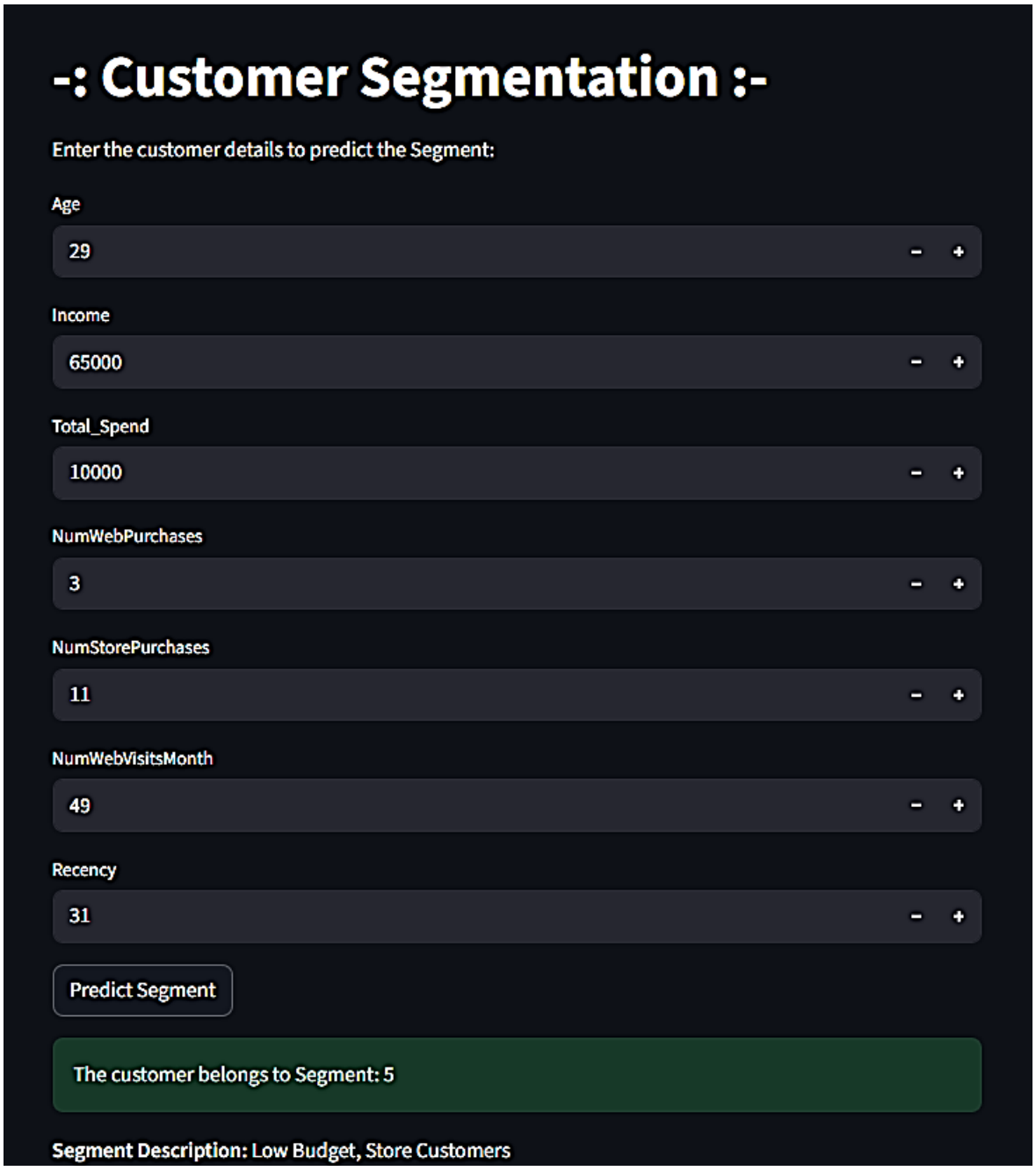
Customer Segmentation Using Data Analysis

Abstract

Customer segmentation is a key strategy for understanding customer behavior and tailoring marketing efforts. This project focuses on analyzing customer data to uncover patterns using Exploratory Data Analysis (EDA) and preprocessing techniques. By examining various attributes such as demographics, purchasing behavior, and engagement metrics, this analysis helps in grouping customers into meaningful segments for better decision-making.

Streamlit application: http://192.168.29.22:8501



Introduction

Understanding customer behavior is critical for businesses to maintain competitiveness. Customer segmentation involves dividing a customer base into distinct groups that share similar characteristics. This segmentation allows businesses to target specific audiences more effectively, resulting in optimized marketing strategies and improved customer satisfaction.

In this project, we utilize a customer dataset to perform segmentation using EDA, preprocessing, and feature engineering. Our main objective is to derive useful insights that can aid in identifying customer groups based on age, income, spending behavior, family size, and customer loyalty.

Methodology

1. **Data Loading**: The dataset (customer_segmentation.csv) was loaded using pandas.
2. **Data Preprocessing**: Null values were handled, columns were cleaned, and necessary features were derived.
3. **Feature Engineering**: New features like Age, Total number of Children, Total Spend, and Customer Since were created.
4. **EDA & Visualization**: Univariate and bivariate analysis were conducted to understand distributions and relationships.
5. **Segmentation Goals**: Identifying high-value customers, loyal customers, and low-spending customers based on income, spend, and tenure.

Data Preprocessing

- Removed null values using df.dropna().
- Converted Dt_Customer to datetime format for tenure calculation.
- Derived Age using current year and Year_Birth.
- Computed Total number of Children as the sum of Kidhome and Teenhome.
- Created Total_Spend by summing all product-related spend columns.
- Calculated Customer Since in days from the joining date to the current date.

These transformations helped prepare the data for meaningful analysis.

Exploratory Data Analysis (EDA)

Key observations from EDA include:

- **Age Distribution**: Customers range from young adults to elderly. Most are between 30–60 years.
- **Income Distribution**: Income is skewed right, with a few high-income outliers.
- **Spending Behavior**:
    - Highest spending is on wines and meat products.
    - Very low spending on gold and fruits.
- **Marital Status & Education**:
    - Majority are married and highly educated.
- **Children**: Most customers have 0–2 children.
- **Tenure**: Some customers have been associated with the business for over 2000 days.

Data Visualization

Some of the key visualizations used:

- **Histograms** for:
    - Income
    - Total Spend

    - Age
- **Boxplots** for:
    - Income vs Total Spend
- **Heatmap** of Correlation:
    - Helps identify how variables like income, age, and spending are related.
- **Bar plots** for:
    - Marital Status
    - Education Distribution

These plots enabled better understanding of customer clusters and behavior patterns.

Result

From the analysis, the following insights were obtained:

- Customers with higher income tend to spend more, especially on wines and meat products.
- Customers with fewer children generally spend more than those with larger families.
- Senior customers (age > 50) spend less on luxury items like gold.
- Long-term customers (higher tenure) tend to spend more consistently.

These insights can help create customer personas and recommend targeted offers or marketing campaigns.

Conclusion

The project successfully segmented customers using various features like age, income, spending patterns, and tenure. The analysis uncovered key traits that define high-value and loyal customers, which can be instrumental in customer relationship management and marketing strategies. Further work could include applying clustering algorithms like K-Means or Hierarchical Clustering to define the segments more explicitly.