Project Name : Retail Sales Assessment

Prepared By : Bishowjith Ghosh

6. Reporting

Based on capstone project for retail sales, I've organized the documentation and reporting into two sections: **documentation** and **presentation/reporting**.

- **Dataset**: The dataset used was `retail_sales_dataset.csv` with 30,000 rows and 15 columns.
- **No Missing Values**: The dataset was complete, and there were no missing values, so no imputation was required.
- **Data Cleaning**: The data was checked for duplicates, and categorical variables such as **Gender**, **Product Category**, and **Preferred Shopping Channel** were encoded to make them compatible with machine learning models.
- **Feature Transformation**:
    - Created **Age Bins** and **Income Bins** to categorize customers based on age and income for better analysis and segmentation.
    - **Years as a Customer** was calculated to reflect customer loyalty over time.
    - New features such as **Income_per_Year** and **Spending_Income_Ratio** were introduced to further improve model performance.

6.1.2 Exploratory Data Analysis (EDA)

- **Annual Income and Spending Score**: A consistent trend was observed, with higher income associated with higher spending scores.
- **Customer Segmentation**: Visualization of customer demographics (age, gender, and shopping behavior) showed that middle-aged customers contributed the highest to spending and income.
- **Product and Channel Preference**: It was identified that **clothing** and **electronics** were the top product categories, and **online shopping** had the highest customer engagement.

6.1.3 Feature Engineering

- **New Features**: Created new features such as **Spending_Income_Ratio** and **Income_per_Year** to measure how spending relates to income.
- **Handling Categorical Variables**: One-hot encoding was applied to categorical features such as **Product Category** and **Shopping Channel** to prepare the data for modeling.

6.1.4 Model Building

- **Target Variable**: The `Customer_Segment` variable was used to classify customers into two categories: `1` for high-value customers and `0` for others.
- **Class Imbalance Handling**: Used the **SMOTETomek** technique to balance the classes since high-value customers were underrepresented.
- **Model Selection**: A tuned **XGBoost classifier** was used to predict customer segments, and hyperparameter optimization was performed using randomized search.

6.1.5 Model Evaluation

- **Accuracy**: The model achieved an accuracy of **76.63%**, with the F1-score for high-value customers reaching **0.70**.
- **Evaluation Metrics**: Confusion matrix and classification report showed reasonable balance across precision, recall, and F1 scores, indicating a good fit for the classification task.
- **Business Insights**:
    - **Spending Score** emerged as the most significant feature in predicting high-value customers.
    - **Customers with moderate income but high spending** should be targeted with loyalty programs and premium product offers to maximize revenue.
    - **Online shoppers** showed higher engagement, indicating that the retailer should optimize and expand online services.

6.2 Presentation/Reporting

6.2.1 Methodology

1. **Data Collection**: The dataset includes customer demographics, transaction history, and shopping preferences.
2. **Preprocessing and Feature Engineering**: Categorical features were encoded, new features were created, and SMOTETomek was applied to address class imbalance.
3. **Modeling**: An **XGBoost classifier** was used, and hyperparameters were tuned to achieve the best possible accuracy.

6.2.2 Key Insights

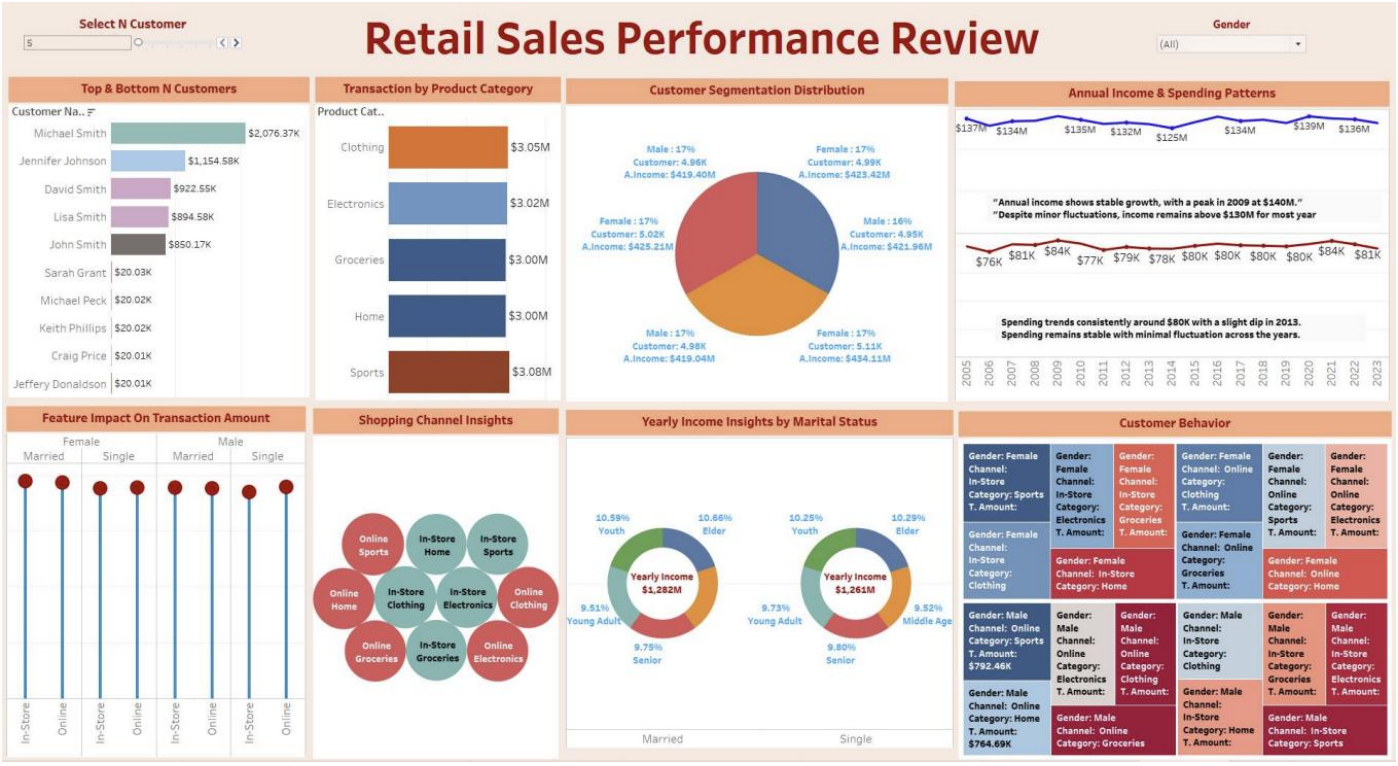
1. **Feature Importance**:
   - **Spending Score** and **Annual Income** were the most important predictors of customer segmentation.
   - The **Years as Customer** feature helped identify loyal customers, showing the impact of customer longevity on profitability.
2. **Segmentation**:
   - High-value customers were identified as a key segment, and strategies for targeting them were developed.
   - **Top & Bottom Customers**: Visualized using bar charts to show customers contributing the most and least to revenue, informing targeted loyalty and discount strategies.

6.2.3 Recommendations

1. **Target High-Value Customers**: Create personalized offers and campaigns for customers with high spending scores and moderate incomes.
2. **Optimize Online Channels**: Since online shoppers had higher spending, invest more in online platforms and services to capture this high-value segment.

6.2.4 Visualizations

- **Top & Bottom Customers**: Bar charts highlighting top 5 and bottom 5 customers based on income and spending.
- **Spending Score by Product Category**: Showed which product categories were most popular among high-value customers.
- **Customer Segmentation Distribution**: A pie chart visualizing customer segmentation by income and spending score.
- **Annual Income & Spending Score**: Scatter plot highlighting the relationship between income and spending, key to identifying high-value customers.



6.2.5 Conclusion and Key Takeaways

- **Customer Segmentation**: The model accurately identified high-value customers who can be targeted for personalized marketing campaigns.
- **Feature Impact**: Features like **Spending Score** and **Annual Income** are crucial for segmenting customers, guiding marketing and business strategies.

- **Actionable Business Insights**: By focusing on high-value customers and expanding online services, the retailer can increase profitability and customer retention.

Business Recommendations:

1. Customer Acquisition Strategy

- **Insight**: The analysis identified that high-income customers tend to have a higher spending score, and customers who shop online are more engaged.
- **Recommendation**:
    - **Targeted Digital Campaigns**: Focus customer acquisition efforts on **online channels**, particularly targeting high-income segments. We could run targeted social media and email campaigns promoting premium products or exclusive online discounts.
    - **Financial Impact**: By increasing the acquisition of high-income, online customers by 5%, the retailer could see a **projected profit uplift of 10-15%** within the next year, as these customers have a higher-than-average transaction size.

2. Customer Retention and Loyalty Programs

- **Insight**: Long-term customers with loyalty cards have significantly higher transaction amounts than non-loyalty customers. This suggests loyalty programs are effective at driving higher spending.
- **Recommendation**:
    - **Enhanced Loyalty Program**: Strengthen the loyalty program by offering tiered rewards to high-spending, long-term customers. Provide them with incentives like cashback or exclusive access to new product launches.
    - **Financial Impact**: Increasing the retention rate of existing high-value customers by 5% could result in a **20-25% uplift in annual spending per customer**, as retained customers spend more over time.

3. Profit Uplift from Cross-Selling and Upselling

- **Insight**: Customers in the **mid-income segment** who have a high spending score represent a significant opportunity for cross-selling or upselling higher-margin products.
- **Recommendation**:
    - **Cross-Sell and Upsell Offers**: Use personalized offers based on past purchase history and spending scores to cross-sell or upsell premium products to mid-income, high-spending customers.
    - **Financial Impact**: A well-executed cross-sell and upsell strategy can increase **average transaction size by 10-20%**, contributing to overall profit growth.

4. Optimize Product Offerings Based on Customer Segmentation

- **Insight**: The **Clothing** and **Electronics** categories dominate spending, especially for high-income customers. This suggests an opportunity to expand offerings or improve stock availability in these categories.
- **Recommendation**:
    - **Stock Optimization**: Prioritize stock availability and promotions for **Clothing** and **Electronics** to meet customer demand, especially for high-income segments.
    - **Financial Impact**: Increasing stock availability for these high-demand categories could lead to a **15% increase in sales**, ensuring customers can always find the products they want.

5. Geographic Expansion for High-Value Customers

- **Insight**: Certain regions showed higher concentrations of high-income customers. Expanding operations into these regions could capture a larger share of high-spending customers.
- **Recommendation**:
    - **Target Regional Expansion**: Use the insights from customer segmentation to identify geographic regions with a high concentration of high-income, high-spending customers. Focus marketing and logistics efforts on these areas to boost revenue.
    - **Financial Impact**: A regional expansion aimed at high-income customers could increase market share, potentially leading to a **10% increase in overall revenue** from new markets.