**DAV ASSIGNMENT REPORT**

**22I-1914**

**PRE-PROCESSING STEPS:**

1. Handling Missing Values

* Filled Missing Values: Used 'Unknown' for categorical columns and median values for numerical ones to preserve data integrity.
* Purchase Date: Replaced missing dates with '1900-01-01' to ensure all records have a valid date.
* Binary Columns: Imputed Returns and Churn with the most frequent value to maintain distribution.
* Dropped Rows: Removed rows with missing Customer ID, as it's essential for grouping.

2. Removing Outliers

* Z-Scores: Calculated Z-scores for numerical columns and identified outliers as those with Z-scores above 3 or below -3.
* Removed Outliers: Filtered out rows with extreme Z-scores to ensure data quality.

3. Standardizing Numerical Variables

* StandardScaler: Standardized numerical columns to have a mean of 0 and a standard deviation of 1, ensuring fair comparisons.

4. Encoding Categorical Variables

* One-Hot Encoding: Applied to Product Category, Payment Method, and Gender to convert them into numerical format.
* Binary Variables: Returns and Churn were cast to integers as they are already binary.

5. Creating New Features

* Total Purchase Value Per Customer: Aggregated total spending by each customer to gauge overall value.
* Number of Purchases: Counted transactions per customer to understand engagement.
* Average Purchase Frequency: Calculated as the number of purchases divided by total customers to measure purchase activity.

These steps ensure that my data is clean, standardized, and ready for analysis

**c) Discussing the Impact of Dimensionality Reduction**

1. Impact of PCA:
   * Loss of Information: PCA reduces dimensionality, which may lead to loss of some information. However, choosing the number of components that capture a high percentage of variance minimizes this loss.
   * Understanding Data Structure: PCA helps in understanding the structure of data by reducing complexity and revealing patterns or clusters that may not be obvious in high-dimensional space.
   * Visualization: Visualizing the first two principal components allows for easier identification of patterns or groupings in the data. For example, clusters or separations among different categories (like Churn) can become more apparent.

Overall, PCA simplifies the dataset while retaining the most important features, aiding in better interpretation and visualization of the data structure.

**Key Insights:**

1. Higher CLV for Non-Churn Customers: The analysis reveals that the Customer Lifetime Value (CLV) for customers who do not churn is approximately $68,455.79, compared to $68,862.35 for those who do churn. This finding indicates that retaining customers who are more likely to churn can still yield substantial value, but it is essential to focus on retention efforts to maximize CLV across all segments.
2. Minimal Difference Between Segments: The small difference in CLV between customers who churn and those who do not suggests that the value generated from both groups is relatively similar. This emphasizes the need for effective strategies to reduce churn, as retaining customers could lead to slightly higher CLV.
3. Impact of Purchase Value on CLV: The "What-if" analysis demonstrates that increasing the purchase value positively affects the average CLV. For instance, a 5% increase in purchase value leads to a noticeable rise in CLV, reinforcing the importance of strategies that encourage higher spending per transaction.
4. Retention Rate's Role in CLV: The retention rate of 70% used in the CLV calculation highlights the significance of customer retention. Although the CLV for customers who churn is slightly higher, improving retention can enhance overall profitability and customer value, indicating that investing in retention strategies is crucial.
5. Consistent Value Across Segments: The relatively close CLV values across churn segments suggest that customers who churn and those who do not have comparable potential for generating revenue. This indicates that effective engagement and retention strategies could help in balancing the CLV across different customer segments.

**Narrative on the Retail Industry Transformation:**

The retail industry is experiencing a profound shift due to the rise of e-commerce. Traditional retailers, once reliant on physical stores, are now facing increasing competition from online platforms that offer convenience, variety, and accessibility. This transformation is reshaping how consumers shop and how businesses engage with their customers.

E-commerce platforms have leveraged technology to provide personalized shopping experiences and seamless transactions, leading to a decline in foot traffic for brick-and-mortar stores. Our analysis of Customer Lifetime Value (CLV) illustrates the impact of this shift. The CLV for customers who churn is slightly higher than for those who do not, highlighting the potential value of retaining customers and the effectiveness of tailored retention strategies.

The "What-if" analysis underscores the importance of increasing purchase values to boost CLV. As e-commerce continues to grow, understanding customer behavior and implementing strategies that enhance customer value are crucial for both traditional retailers and e-commerce platforms. The close CLV values between churn and non-churn segments suggest that optimizing customer retention efforts can lead to significant improvements in profitability.

The PCA analysis further supports the need for targeted strategies by revealing distinct customer behavior patterns. The ongoing evolution of the retail industry emphasizes the importance of adapting to digital trends and leveraging data-driven insights to remain competitive.

**Actionable Recommendations:**

**For Traditional Retailers:**

1. Adopt Omnichannel Strategies: Integrate online and offline channels to offer a seamless shopping experience. Invest in e-commerce capabilities and digital marketing to enhance customer reach and engagement.
2. Focus on Retention Strategies: Implement targeted loyalty programs and personalized marketing efforts to retain customers. Improving retention can balance CLV across different segments and maximize revenue.
3. Enhance Customer Engagement: Offer unique in-store experiences and exclusive promotions to attract and retain customers who value physical shopping. This can complement online efforts and drive higher CLV.

**For E-Commerce Platforms:**

1. Optimize Pricing and Promotions: Regularly adjust pricing strategies and offer promotions to encourage higher purchase values. This can directly impact CLV and drive profitability.
2. Invest in Personalization: Use customer data to provide tailored recommendations and offers. Enhancing the shopping experience through personalization can increase CLV and reduce churn.
3. Improve Customer Support: Expand support channels and ensure prompt resolution of customer issues. Effective customer service can reduce churn and enhance overall customer satisfaction.

**For Policymakers:**

1. Support Small and Traditional Retailers: Provide financial assistance and resources to help small and traditional retailers transition to e-commerce. This can foster a more competitive and diverse market landscape.
2. Ensure Data Privacy and Security: Implement regulations to protect customer data and build trust in both online and offline transactions. Data privacy is essential for sustaining e-commerce growth.
3. Invest in Digital Infrastructure: Enhance digital infrastructure and technology to support e-commerce growth and enable traditional retailers to integrate online capabilities effectively. This investment can drive overall industry advancement.