

Assignment 1

Data Analysis and Visualization

Preprocessing Steps:

1. Combining Datasets:

- Merged df1, df2, and df3 into one dataset using `pd.concat()`.
- **Reason:** To put all data together for easier analysis.

2. Removing Unnecessary Columns:

- Deleted columns Unnamed: 21 to Unnamed: 25.
- **Reason:** These columns didn't have any useful information and were empty.

3. Filling Missing Values:

- Fill missing values in numeric columns with the average (mean) of each column.
- **Reason:** To avoid losing data while filling in gaps.

4. Detecting and Removing Outliers:

- Found outliers using the Z-score method with a threshold of 2.5 and removed them.
- **Reason:** To remove values that were too different and could affect the results.

5. Normalization:

- Scaled numeric columns between 0 and 1 using Min-Max scaling.
- **Reason:** To make sure the values are in the same range, which helps with analysis and modeling.

6. Label Encoding:

- Converted text categories into numbers using `LabelEncoder`.
- **Reason:** Most models only work with numbers, so text data needs to be converted.

7. Creating New Features:

- Added new columns: `Total_Sales`, `Discount_Effectiveness`, and `Sales_per_Customer`.
- **Reason:** These new features add more useful information for analysis.

Insights and Narrative:

Key Insights

1. Customer Segmentation and CLV:

- High CLV segments contribute significantly to total sales, indicating that a small percentage of customers generate the majority of revenue. This highlights the importance of focusing on high-value customers.

2. Price Sensitivity:

- Changes in unit price have a noticeable impact on CLV. Increasing prices generally boosts average CLV, suggesting that pricing strategies directly affect profitability.

3. Discount Effectiveness:

- The effectiveness of discounts varies, with some discounts significantly boosting sales. Optimizing discount strategies is crucial for balancing revenue and profitability.

4. Dimensionality Reduction via PCA:

- PCA results show that the first two principal components capture most of the data variance, simplifying the analysis while retaining essential information about customer behavior.

5. CLV Distribution:

- The distribution of CLV reveals a concentration of high-value customers. Strategies should focus on retaining these customers to maximize revenue.

Narrative

The retail industry is experiencing a major shift as e-commerce continues to expand. Traditional retail faces significant challenges from the convenience and reach of online shopping. Our analysis and visualizations demonstrate that high-value customers drive the majority of sales, highlighting the need for targeted retention strategies. Pricing changes impact CLV, suggesting that strategic pricing can enhance profitability. PCA analysis simplifies the complex data, focusing on the most influential features, while the distribution of CLV emphasizes the importance of high-value segments. These insights paint a picture of a retail landscape increasingly shaped by e-commerce, where customer value and strategic pricing are key to success.

Recommendations:

For Traditional Retailers:

1. Adopt Omnichannel Strategies:

- Integrate online and offline experiences to provide a seamless shopping journey and leverage data from both channels for personalized customer engagement.

2. Personalize Marketing Efforts:

- Use data analytics to create tailored marketing campaigns aimed at retaining high-value customers and enhancing their shopping experience.

3. Enhance Inventory Management:

- Implement predictive analytics to optimize inventory levels, reducing costs and improving customer satisfaction.

For E-Commerce Platforms:

1. Improve Customer Experience:

- Continuously enhance website usability and customer service to retain existing customers and attract new ones.

2. Leverage Advanced Analytics:

- Use data insights to refine pricing strategies, target marketing efforts, and improve overall business performance.

3. Expand Product Offerings:

- Diversify product lines to cater to a broader audience and maintain competitive growth.

For Policymakers:

1. Invest in Digital Infrastructure:

- Support the development of robust digital infrastructure and cybersecurity to facilitate secure and efficient e-commerce operations.

2. Ensure Fair Competition:

- Implement regulations to promote fair competition and prevent monopolistic practices, supporting both traditional and online retailers.

3. Support Workforce Transition:

- Develop policies to assist workers displaced by the shift to e-commerce, including retraining programs and job transition support.