Business Insights Report

I have derived the following key business insights:

1. Customer Distribution by Region

The majority of the customers are located in South America and Asia, accounting for approximately 60% and 35% of the customer base respectively. This distribution highlights potential regional hotspots for targeted marketing campaigns, much like focusing on areas where your favorite coffee shop sees the most foot traffic.

2. Monthly Sales Trend

Analyzing the monthly sales trend from January 2024 to November 2024, I observed a fluctuating pattern with sales peaking in July 2024 and experiencing a significant decline in January 2024.

- **Seasonal Fluctuations:** Sales are subject to seasonal changes, with higher sales during certain months (e.g., July) and lower during others (e.g., January).
- **Peak Sales Period:** July 2024 saw the highest sales, potentially due to summer sales, holidays, or special events.
- **Lowest Sales Period:** January 2024 had the lowest sales, possibly due to post-holiday season slump, winter weather, or reduced consumer spending.
- **Sales Trend:** Overall, the sales trend is relatively stable with some ups and downs, indicating a consistent customer base that maintains sales levels despite seasonal variations.
- **Opportunities for Improvement:**
    - Increasing sales in January through targeted marketing campaigns or promotions.
    - Maximizing revenue during July by offering special deals or discounts.
    - Analyzing factors contributing to sales fluctuations to develop strategies for mitigating their impact.

Recommendation: By understanding these sales patterns, what we can better is plan out marketing strategies, ensuring we boost sales during slower months and capitalize on peak periods effectively.

3. Product Category Performance

The 'Electronics' and 'Books' categories are the top performers, contributing over 50% of the total sales value. This dominance underscores the importance of maintaining diverse stock and focusing on high-demand categories to sustain sales growth.

4. Customer Tenure and Spending Behavior

Customers with longer tenure tend to have higher total spending and purchase frequency. This correlation emphasizes the value of customer retention strategies in boosting overall sales, much like nurturing long-term relationships with loyal clients.

5. Transaction Validation

I identified 88 invalid transactions where the total value did not match the product of quantity and price, pointing to potential data quality issues. Addressing these discrepancies is essential to ensure accurate financial reporting and inventory management.

6. Customer Segmentation Insights

Using clustering techniques and analyzing the Davies-Bouldin Index (DBI), I determined that **4 clusters** are optimal for segmenting the customers.

- **Optimal Number of Clusters:** The DBI plot showed a significant dip around 4 clusters, indicating well-separated and cohesive groups.
- **Diminishing Returns Beyond 4 Clusters:** Increasing the number of clusters beyond 4 did not improve segmentation quality and sometimes worsened it.
- **Strategic Sweet Spot:** Focusing on 4 distinct customer segments strikes a balance between detailed understanding and practical management.

Business Strategy Implications:

- **Tailored Marketing:** Develop targeted campaigns for each of the 4 segments to enhance engagement and conversion.
- **Product Development:** Design products that cater specifically to the needs and preferences of each segment.
- **Customer Service:** Provide differentiated service experiences based on segment characteristics.
- **Resource Allocation:** Allocate resources efficiently by focusing on the most valuable segments.

Recommendation: By segmenting the customers into 4 distinct groups, we can easily tailor out strategies more effectively, leading to improved customer satisfaction and increased revenue.

7. Opportunities for Continuous Improvement

- **Diversify Product Offerings:** Cater to different customer segments to reduce dependency on seasonal fluctuations.
- **Targeted Marketing:** Increase sales during low periods and capitalize on peak periods with specialized campaigns.
- **Inventory Management:** Ensure adequate stock during peak times to meet customer demand.
- **Customer Retention:** Implement strategies to maintain a consistent customer base and mitigate the impact of seasonal sales variations.
- **Data Analysis:** Continuously monitor and analyze sales and customer data to identify trends and adjust strategies accordingly.

Lookalike Model Development

In developing the Lookalike Model, my goal was to identify and recommend three similar customers for each target customer based on their profiles and transaction history. Think of it like finding friends who share similar interests and habits as you do.

Data Preprocessing: I began by merging the Customers, Products, and Transactions datasets to create a comprehensive dataset. Key features such as Total Spend, Purchase Frequency, and Average Order Value were calculated to capture each customer's purchasing behavior. Additionally, I computed customer tenure (days since signup) and recency (days since last transaction) to reflect customer engagement over time. Categorical variables like customer regions were one-hot encoded to facilitate numerical analysis.

Feature Scaling and Similarity Computation: To ensure all features contributed equally to the similarity measurement, I applied the StandardScaler to normalize the feature values. Using cosine similarity, I computed a similarity matrix that quantifies the similarity between each pair of customers based on their scaled feature vectors. This approach effectively captures the directionality of the feature vectors, making it suitable for assessing customer similarities irrespective of their absolute spending amounts.

Generating Lookalike Recommendations: For each target customer (CustomerID: C0001 - C0020), I extracted their similarity scores with all other customers from the similarity matrix. By sorting these scores in descending order and excluding the target customer themselves, I identified the top three most similar customers along with their corresponding similarity scores. These recommendations are compiled into a "Lookalike.csv" file, providing a clear mapping between each target customer and their lookalikes.

Final Output: The resulting Lookalike Model offers valuable insights into customer similarities, enabling personalized marketing strategies, targeted promotions, and enhanced customer engagement initiatives. By leveraging this model, the business can effectively identify and reach out to potential high-value customers who share similar profiles with existing loyal customers.

My Approach to the Data Science Assignment

In tackling the eCommerce Data Science Assignment, I focused on three main tasks as mentioned in pdf: Exploratory Data Analysis (EDA) with business insights, building a Lookalike Model, and customer segmentation through clustering.

Task 1: Exploratory Data Analysis and Business Insights

I started by importing the necessary datasets: Customers, Products, and Transactions. Using pandas, I loaded the data and checked for any missing values or duplicates, ensuring the data was clean for analysis. I then explored the data to understand customer demographics, product categories, and transaction patterns. By visualizing sales trends over time and analyzing customer tenure and spending behavior, I was able to derive meaningful business insights that could help improve marketing strategies and operational decisions.

Task 2: Lookalike Model

For the Lookalike Model, my goal was to find similar customers based on their profiles and purchasing history. I first merged the customer and transaction data to get a comprehensive view of each customer’s activity. I calculated features like total spending, purchase frequency, and average order value to represent each customer. After scaling these features, I used cosine similarity to measure how alike each customer was to others. By ranking the similarity scores, I recommended the top three similar customers for each of the first 20 customers in the dataset. This model helps in targeting customers with similar behaviors, enhancing personalized marketing efforts.

Task 3: Customer Segmentation / Clustering

For customer segmentation, I aimed to group customers based on their purchasing habits and profile information. I combined features such as total spend, purchase frequency, recency of the last purchase, and tenure. After handling any missing values and scaling the data, I applied the KMeans clustering algorithm to identify distinct customer segments. I experimented with different numbers of clusters and evaluated them using the Davies-Bouldin Index to choose the optimal number of segments. Finally, I visualized the clusters using PCA to reduce the dimensions, which helped in understanding the distinct groups and tailoring strategies for each segment.
