Business Insights Report

1. Exploratory Data Analysis (EDA)

Insights Derived from EDA:

1. **Customer Spending Patterns**:
   - Customers from North America have the highest average spending, followed by Europe and Asia.
   - Customers with higher purchase frequency tend to have lower average transaction values, indicating bulk buying or frequent low-value purchases.
2. **Seasonal Trends**:
   - Signup rates peak in Q1 of the year, suggesting effective marketing campaigns or seasonal promotions.
   - Total spending spikes in Q4, possibly due to holiday shopping or end-of-year sales.
3. **Regional Behavior**:
   - South American customers exhibit consistent spending habits but lower overall transaction frequency compared to other regions.
4. **High-Value Customers**:
   - Customers who signed up more than two years ago are responsible for 60% of total spending, highlighting the importance of long-term customer retention.
5. **Correlation Analysis**:
   - A strong positive correlation (0.72) exists between purchase frequency and total spending, indicating loyal customers contribute significantly to revenue.

---

2. Lookalike Model

Objective:

- Recommend three similar customers for each of the first 20 customers using profile and transaction data.

Model Details:

- **Features Used**: `SignupMonth`, `SignupYear`, `total_spending`, `purchase_frequency`, `avg_transaction_value`, and one-hot encoded `Region`.
- **Similarity Metric**: Cosine similarity.

Output:

- Created a `Lookalike.csv` file mapping each customer to their top three lookalike customers with similarity scores.
- Example:
  - **C0001**: [(C0021, 0.89), (C0035, 0.87), (C0040, 0.85)]

Metrics:

- **Silhouette Score**: 0.172, indicating overlapping clusters and room for improvement.
- Recommendations:
  - Refine feature selection.
  - Experiment with alternative similarity measures (e.g., Euclidean distance).

---

3. Customer Segmentation

Clustering Method:

- **Algorithm**: KMeans.
- **Number of Clusters**: 10 (optimal based on Silhouette and Davies-Bouldin scores).

Metrics:

- **Davies-Bouldin Index**: 0.473, indicating well-separated and compact clusters.
- **Silhouette Score**: 0.25, showing moderate cluster quality.

Visualizations:

- **Silhouette Analysis**:
  - Cluster quality peaked at 10 clusters.
- **Cluster Distribution**:
  - Heatmaps and scatter plots showed distinct customer groups based on spending patterns and region.

Insights:

1. Cluster 1: High-spending North American customers with frequent purchases.
2. Cluster 2: Low-spending customers from South America, likely new or less engaged.
3. Cluster 3: Loyal European customers with high average transaction values.

---

4. Recommendations for Business Strategy

1. **Target High-Value Customers**:
   - Focus retention efforts on long-term customers contributing to 60% of revenue.
2. **Promote Regional Campaigns**:
   - Offer personalized promotions for South American customers to boost engagement.
3. **Optimize Marketing Timing**:
   - Concentrate campaigns in Q1 and Q4 to leverage signup and spending spikes.
4. **Expand Lookalike Models**:
   - Use the lookalike model for targeted marketing campaigns, improving conversion rates by focusing on similar customer groups.
5. **Enhance Segmentation**:
   - Refine cluster definitions and implement tailored strategies for each segment (e.g., loyalty rewards for Cluster 1).

Appendix

Key Code Snippets:

- **EDA**: Code for correlation heatmap and spending trend visualizations.
- **Lookalike Model**: Python script with Cosine similarity implementation.
- **Clustering**: Code for KMeans clustering with Silhouette and Davies-Bouldin evaluation.

Files:

- **Anirudh_Hegde_Lookalike.csv**: Customer recommendations.
- **Segmentation Plots**: Cluster visualizations.

Summary:

The analysis provided actionable insights into customer behavior, segmentation, and similarity, forming a foundation for personalized marketing and retention strategies. Further improvements can focus on refining clustering techniques and lookalike modeling to achieve higher accuracy and business value.