CLUSTERING

1.Segmentation:

- The data has been segmented into five distinct groups (clusters), each represented by different colors (red, orange, green, blue, purple). These clusters indicate groups of customers with similar characteristics or behaviors.

2.Separation:

- The clusters appear relatively well-separated, meaning the clustering algorithm effectively differentiated groups. This suggests that the underlying data has distinct patterns.

3.Cluster Characteristics:

- Each cluster likely corresponds to a unique customer behavior or demographic profile (e.g., high spenders, frequent shoppers, discount-seekers). Additional data analysis would reveal the specific features defining each group.

4.Decision-Making:

- This clustering could be used for decision-making, such as targeted marketing strategies. For example:
    - Customers in Cluster 0 (red) might need a different marketing approach than those in Cluster 3 (blue).
    - Clusters with overlapping areas may indicate customer behaviors with shared characteristics.

5.Outliers:

- Some data points (e.g., those at the extreme edges) may represent outliers or unique customer behaviors. These could require special attention or further investigation.

6.Insights You Can Derive:

<u>**Cluster Characteristics**</u>:

- Each cluster can now be described by its mean values across the numerical features. These values can highlight distinct characteristics for each cluster. For example:
  - **Cluster 0** might have high values for Annual Income but low values for Age.
  - **Cluster 1** This group has the lowest **Purchase Frequency**, with a median around 2–3.The narrow range suggests low purchasing frequency with minimal variation.
  - **Cluster 2** represents customers with moderate **Purchase Frequency** (median around 5–6).The variability is low, with most customers behaving similarly.
  - **Cluster 3** might show high values for Spending Score, indicating more frequent or larger purchases.
  - **Cluster 4**: Cluster 4 shows high **Purchase Frequency** (median around 8–9), second only to Cluster 0.The wider range suggests more variability in purchasing behavior.

<u>**Comparing Clusters**</u>:

- By comparing the mean values across clusters, you can identify trends or patterns:

    - Which cluster represents high-value customers (e.g., high income, high spending)?
    - Which cluster represents budget-conscious customers (e.g., low income, low spending)?

Actionable Segments:

- These profiles help in creating tailored strategies for different clusters:
    - **High-value clusters**: Focus on loyalty programs or premium services.
    - **Low-value clusters**: Implement targeted promotions or educational campaigns to encourage engagement.

Validation of Clustering:

- By looking at the mean values, you can validate if the clusters make sense:
    - Do the features within each cluster align with expected patterns?
    - Are the clusters sufficiently distinct?

Feature Importance:

- The variation in mean values across clusters highlights which features (columns) contribute most to differentiating clusters. Features with little variation may have less impact on the clustering.

7.Actionable Insights

<u>High-Value Customers (Clusters 0 and 4)</u>:

- Customers in these clusters are frequent purchasers and represent high-value segments.
- You could prioritize retention strategies like loyalty programs, exclusive discounts, or personalized recommendations.

<u>Low-Value Customers (Clusters 1 and 3)</u>:

- These clusters indicate infrequent purchasers.
- Strategies to engage these customers might include promotional campaigns, reminders, or incentives to encourage more frequent buying.

<u>Moderate Segment (Cluster 2)</u>:

- Cluster 2 represents a middle-ground group.
- Focus on converting them into more frequent buyers through targeted upselling or bundling offers.

<u>Outliers</u>:

- Identify and investigate outliers in Cluster 3 or others to understand anomalies. These might indicate special cases or data quality issues.