Mall\_Customers

To continue building the customer segmentation model, we will proceed with feature engineering, applying clustering algorithms, visualization, and interpretation. Let's start with feature engineering:

Feature engineering

1. RFM Analysis:

* RFM (Recency, Frequency, Monetary) analysis is a widely used technique in customer segmentation. It involves creating features based on the recency of customer purchases, the frequency of purchases, and the monetary value of those purchases. These features can provide valuable insights into customer behavior and help in creating segments based on their buying patterns.

1. Customer Lifetime Value (CLV):

* Calculating the CLV of each customer and using it as a feature can help in understanding the long-term value that each customer brings to the business. This can be a useful metric for segmentation, as it can help identify high-value and low-value customers.

1. Time-based Features:

* Creating features based on the time of day, week, month, or year when customers make purchases can help identify seasonal trends and patterns in customer behavior. These features can be particularly useful for businesses with seasonal products or services.

1. Demographic Features:

* Incorporating demographic information such as age, gender, location, and income can provide insights into customer preferences and behavior based on different demographic segments. This can help in creating more targeted and personalized marketing strategies.

1. Behavioral Features:

* Analyzing customer behavior on the company's website or app, such as click-through rates, browsing history, and time spent on different pages, can provide valuable insights into customer interests and preferences. Creating features based on this information can help in understanding customer engagement and interaction with the brand.

1. Social Media Interactions:

* If available, incorporating data on customers' social media interactions with the brand, such as likes, shares, and comments, can provide insights into customer engagement and brand perception. These features can help in understanding the impact of social media on customer segmentation.

1. Segmentation-Specific Features:

* Depending on the specific segmentation approach being used, creating features that are tailored to the segmentation criteria can be beneficial. For instance, if using a behavioral segmentation approach, features related to specific behavioral patterns can be created.

Applying clustering algorithms

1. K-means Clustering:

* Choose the number of clusters (K) based on domain knowledge or by using techniques like the elbow method or silhouette analysis.
* Apply the K-means algorithm to the preprocessed dataset.
* Use the cluster centers and assignments to interpret the segments and their characteristics.
* Evaluate the quality of the clusters using metrics like silhouette score and within-cluster sum of squares (WSS).

#### kmeans **=** KMeans(n\_clusters**=**5)

#### label **=** kmeans.fit\_predict(X3)

#### print(label)

#### print(kmeans.cluster\_centers\_)

1. Hierarchical Clustering:

* Decide on the type of hierarchical clustering, such as agglomerative or divisive clustering.
* Choose a suitable linkage method (e.g., complete, single, or average linkage) based on the nature of the data.
* Construct a dendrogram to visualize the clustering process and decide on the number of clusters.
* Cut the dendrogram at the appropriate height to form clusters.
* Evaluate the clusters using metrics such as cophenetic correlation or silhouette score.

### print(kmeans.cluster\_centers\_)

1. DBSCAN (Density-Based Spatial Clustering of Applications with Noise):

* Determine the appropriate values for the minimum number of points and the neighborhood distance (epsilon) based on the data's density and distribution.
* Apply the DBSCAN algorithm to the data.
* Identify core points, border points, and noise points.
* Analyze the resulting clusters and their characteristics.
* Evaluate the quality of the clusters using metrics like silhouette score or Davies-Bouldin index.

### **dbscan = DBSCAN(eps=0.5, min\_samples=5)**

### **y\_dbscan = dbscan.fit\_predict(X\_scaled)**

1. Gaussian Mixture Models (GMM):
   * Estimate the parameters of the Gaussian mixture models, including the number of components (clusters) and the means and covariances of each component.
   * Use the Expectation-Maximization (EM) algorithm to fit the GMM to the data.
   * Assign data points to different clusters based on the highest probability of belonging to each component.
   * Analyze the resulting clusters and their characteristics.
   * Evaluate the quality of the clusters using metrics such as the Akaike information criterion (AIC) or Bayesian information criterion (BIC).

cluster **=** kmeans.fit\_predict(X3)

df["label"] **=** cluster

​

​

**from** mpl\_toolkits.mplot3d **import** Axes3D

​

​

fig **=** plt.figure(figsize**=**(20,10))

ax **=** fig.add\_subplot(111,projection **=** '3d')

​

plt.xlabel("Age")

plt.ylabel("Annual Income (K$)")

ax.set\_zlabel('Spending Score(1-100)')

​

plt.show()

Visualization

* Certainly, visualization is crucial for understanding the patterns and insights derived from the customer segmentation model. Here's an example of how you can visualize the customer segments using Python libraries like Matplotlib and Seaborn.
* Visualizations help in better understanding the underlying patterns and can facilitate decision-making processes for targeted marketing strategies and customer engagement initiatives.

from mpl\_toolkits.mplot3d import Axes3D

plt.figure(1,figsize=(15,6))

n = 0

for cols in ['Age','Annual Income (k$)','Spending Score (1-100)']:

n +=1

plt.subplot(1,3,n)

sns.set(style="whitegrid")

plt.subplots\_adjust(hspace=0.5,wspace=0.5)

sns.violinplot(x = cols,y = 'Gender',data=df)

plt.ylabel('Gender' if n== 1 else '')

plt.title('Violin Plot')

plt.show()

Interpretation:

Segment Characteristics Analysis:

* Examine the descriptive statistics for each segment, including mean, median, and mode for various features.
* Identify the distinguishing characteristics that define each segment, such as purchasing behavior, engagement patterns, and demographic attributes.

Tailored Strategy Development:

* + Develop customized marketing, sales, and product strategies for each segment based on their unique characteristics and needs.
  + Analyze the characteristics of each cluster, such as average spending, age group, geographic location, and purchase frequency.
  + Identify key features that distinguish one cluster from another.
  + Explore the differences in customer behavior and preferences among clusters.
  + Use the insights to tailor marketing strategies, improve customer experience, and personalize product offerings for each segment

Conclusion :

* Ensure to validate the clustering results with domain experts and conduct further analysis to refine the segmentation model.