Customer Segmentation Using Python

**Importing Libraries**

In [1]:

**import** pandas **as** pd

**import** matplotlib.pyplot **as** plt

**import** seaborn **as** sns

**from** sklearn.cluster **import** KMeans

**import** warnings

warnings**.**filterwarnings("ignore")

**Reading the Data**

**Data is available in CSV file**

In [2]:

data **=** pd**.**read\_csv('Mall\_Customer.csv')

In [3]:

data**.**columns

Out[3]:

Index(['CustomerID', 'Gender', 'Age', 'Annual Income (k$)',

'Spending Score (1-100)'],

dtype='object')

In [4]:

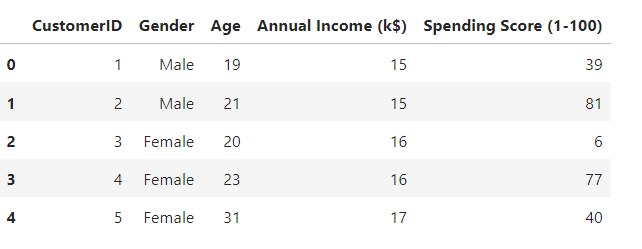
*# Total rows and columns*

data**.**shape

Out[4]:

(200, 5)

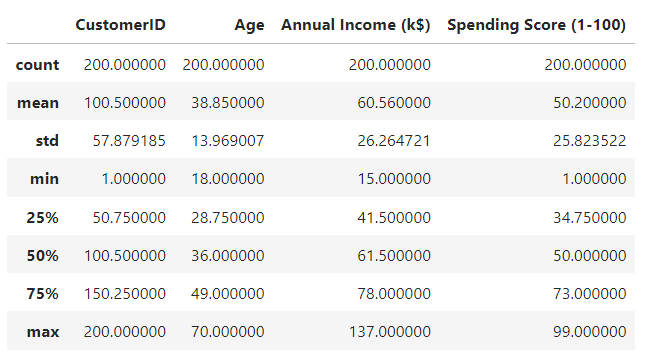
data**.**head()



**Variable Identification**

In [6]:

data**.**describe()



data**.**info()

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 200 entries, 0 to 199

Data columns (total 5 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 CustomerID 200 non-null int64

1 Gender 200 non-null object

2 Age 200 non-null int64

3 Annual Income (k$) 200 non-null int64

4 Spending Score (1-100) 200 non-null int64

dtypes: int64(4), object(1)

memory usage: 7.9+ KB

In [8]:

*# Finding Null Values*

print(data**.**isnull()**.**sum())

CustomerID 0

Gender 0

Age 0

Annual Income (k$) 0

Spending Score (1-100) 0

dtype: int64

In [9]:

**from** sklearn.cluster **import** KMeans

*# Select the relevant features for clustering*

X **=** data[['Age', 'Annual Income (k$)', 'Spending Score (1-100)']]

*# Set the number of clusters*

n\_clusters **=** 5

*# Fit the K-means model to the data*

kmeans **=** KMeans(n\_clusters**=**n\_clusters, random\_state**=**42)

kmeans**.**fit(X)

*# Obtain the cluster labels*

labels **=** kmeans**.**labels\_

In [10]:

*# Plot the clusters*

plt**.**scatter(X['Annual Income (k$)'], X['Spending Score (1-100)'], c**=**labels)

plt**.**xlabel('Annual Income (k$)')

plt**.**ylabel('Spending Score (1-100)')

plt**.**title('Customer Segmentation')

plt**.**show()



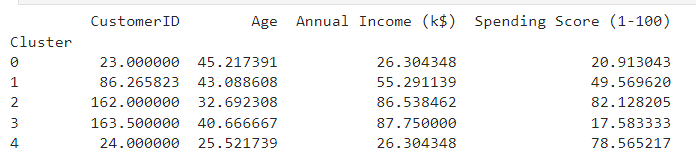
*# Add the cluster labels to the dataset*

data['Cluster'] **=** labels

*# Analyze the segments*

segment\_data **=** data**.**groupby('Cluster')**.**mean()

print(segment\_data)



*# Plot the clusters with centroids*

plt**.**scatter(X['Annual Income (k$)'], X['Spending Score (1-100)'], c**=**labels, cmap**=**'viridis')

plt**.**scatter(kmeans**.**cluster\_centers\_[:, 0], kmeans**.**cluster\_centers\_[:, 1], s**=**300, c**=**'red', label**=**'Centroids')

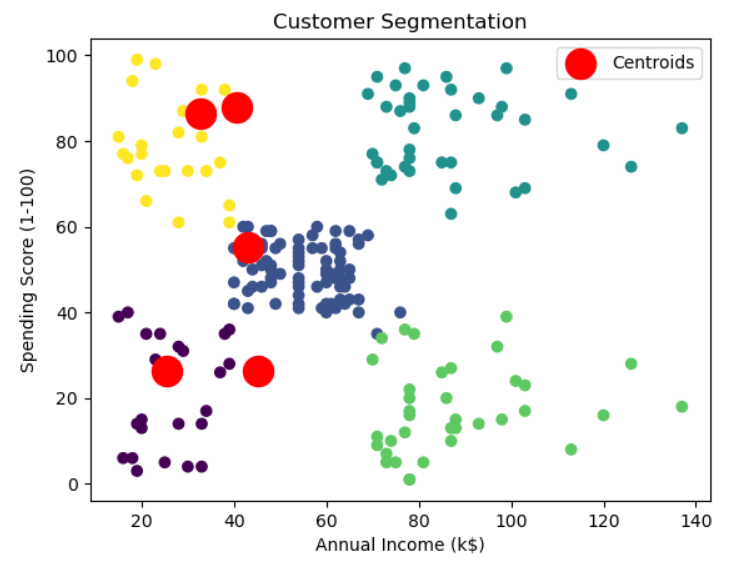
plt**.**xlabel('Annual Income (k$)')

plt**.**ylabel('Spending Score (1-100)')

plt**.**title('Customer Segmentation')

plt**.**legend()

plt**.**show()



**from** mpl\_toolkits **import** mplot3d

*# Create a 3D scatter plot*

fig **=** plt**.**figure()

ax **=** plt**.**axes(projection**=**'3d')

ax**.**scatter3D(X['Age'], X['Annual Income (k$)'], X['Spending Score (1-100)'], c**=**labels, cmap**=**'viridis')

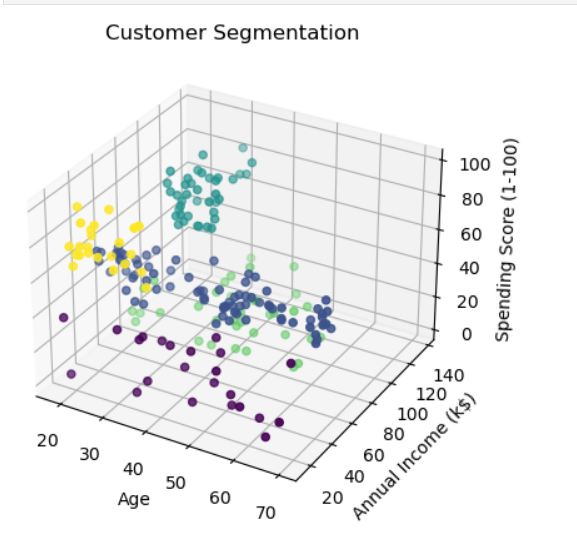
ax**.**set\_xlabel('Age')

ax**.**set\_ylabel('Annual Income (k$)')

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

ax**.**set\_title('Customer Segmentation')

plt**.**show()



**from** sklearn.preprocessing **import** StandardScaler

**import** scipy.cluster.hierarchy **as** sch

In [15]:

*# Select the relevant features for clustering*

features **=** data[['Age', 'Annual Income (k$)', 'Spending Score (1-100)']]

*# Scale the features*

scaler **=** StandardScaler()

scaled\_features **=** scaler**.**fit\_transform(features)

*# Perform hierarchical clustering with the 'ward' method (Euclidean distance metric)*

dendrogram **=** sch**.**dendrogram(sch**.**linkage(scaled\_features, method**=**'ward'))

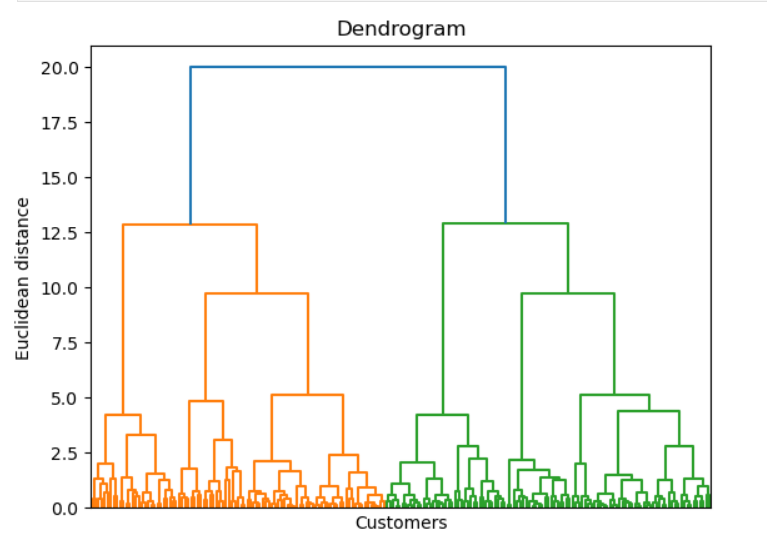
plt**.**title("Dendrogram")

plt**.**xlabel("Customers")

plt**.**ylabel("Euclidean distance")

plt**.**tick\_params(axis**=**'x',labelbottom**=False**);

plt**.**show()



**from** sklearn.decomposition **import** PCA

In [17]:

*# Preprocess the data*

features **=** ['Age', 'Annual Income (k$)', 'Spending Score (1-100)']

X **=** data[features]

scaler **=** StandardScaler()

X\_scaled **=** scaler**.**fit\_transform(X)

In [18]:

*# Apply PCA for dimensionality reduction*

pca **=** PCA(n\_components**=**2)

X\_pca **=** pca**.**fit\_transform(X\_scaled)

In [19]:

*# Add the PCA components to the dataset*

data['PC1'] **=** X\_pca[:, 0]

data['PC2'] **=** X\_pca[:, 1]

*# Visualize the segments*

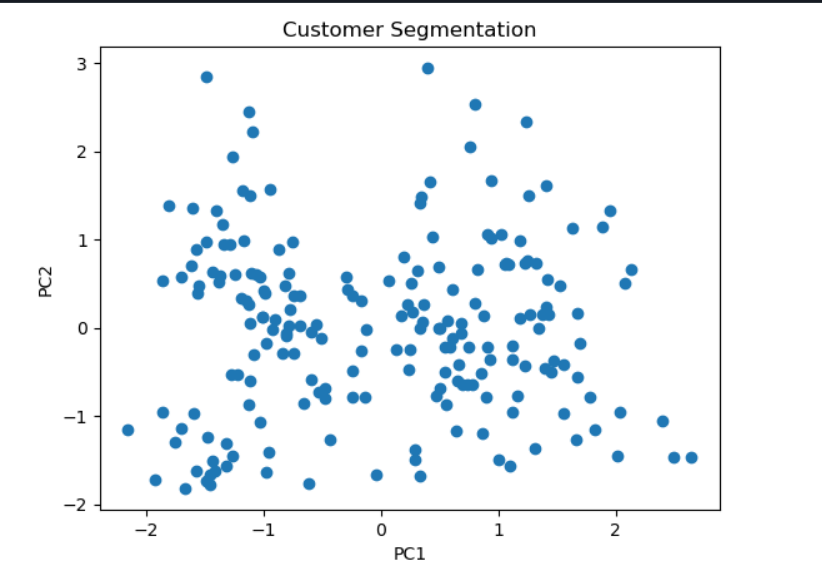
plt**.**scatter(data['PC1'], data['PC2'])

plt**.**xlabel('PC1')

plt**.**ylabel('PC2')

plt**.**title('Customer Segmentation')

plt**.**show()



Conclusion

**The customer segmentation analysis project using Python revealed valuable insights. Here is a summary of the findings:**

**K-means Clustering: The K-means algorithm was applied to cluster customers based on their age, annual income, and spending score. By setting the number of clusters to 5, distinct segments of customers were identified. The scatter plot visualization helped understand the distribution of customers within each cluster.**

**Hierarchical Clustering: The hierarchical clustering algorithm, with the ward method and Euclidean distance metric, was utilized. The dendrogram provided insights into the hierarchical structure of the data, showcasing the linkage between different clusters and the Euclidean distances between them.**

**Principal Component Analysis (PCA): PCA was performed to reduce the dimensionality of the data and visualize customer segments in a two-dimensional space. By representing customers based on the first two principal components, patterns and relationships among the segments were observed.**

**The obtained customer segments enable businesses to gain a deeper understanding of their customer base. This understanding allows for personalized marketing strategies, improved customer satisfaction, and optimized business strategies. By tailoring their offerings to specific segments, businesses can effectively target customers and enhance their overall performance.**

**It is important to note that these conclusions are based on the specific dataset and context of the project. Adjustments and further analysis may be required based on the unique characteristics and goals of the business.**