i) Find and download the customer dataset as provided at: https://

gist.github.com/pravalliyaram/5c05f43d2351249927b8a3f3cc3e5ecf

**ii) How many records are there? [1]**

**Total records is 200. We have 200 rows and 5 columns.**

**iii) What features is it composed of ?[2]**

We have five features (columns)

CustomerID, Gender, Age, Annual Income(k$), Spending Score

**iv) Does it contain missing values?[2]**

This dataset does not have any missing values

**v) How will you deal with the missing values, if they are present?[3]**

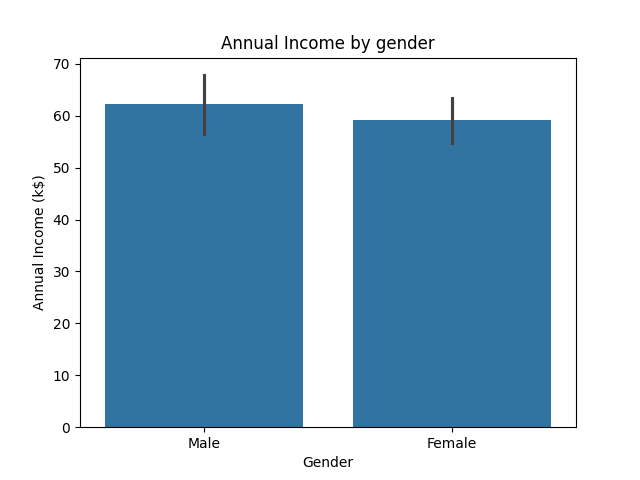
1.Removing the columns or rows with missing values

2.Imputation by filling missing values with a specific value example mean

3.Forward or backward fill - filling the missing values with the value from the previous or next observation.

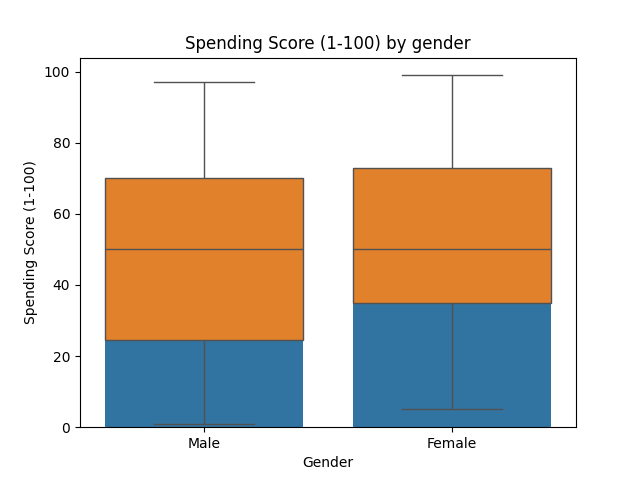
**vii) How does the 'Annual income', 'Age' and 'Spending Score' vary with**

**'Gender'? Use graphs as part of your description.[9]**

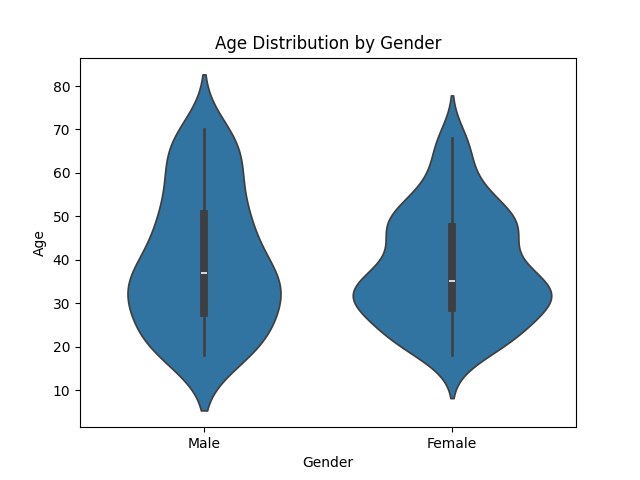


The bar chart above shows that Males Annual Income is higher than the Females.

The Males annual income in (k$) 62.6 while the Females annual income in (k$) is 59.7.



Female Spending score(1-100) is higher than males. Females spend 73.5 which is 2.8 higher than males whose spending score is 70.4



The shape of the violin shows that Females had a higher probability density than men.

The Males median was 37.4 while the Females median was 35.3.

viii) Use the KMeans method to place the data in groups (clusters) based on

customer 'Age' and 'Annual Income'.

Show your work using groups (value of k).[13]

Steps

1.Normalizing the testing and training data.

2. Choosing the number of clusters and adjusting based on my analysis, 2 is my number of

clusters.

3.Creating a KMeans model with fixed random state for reproducibility.

4.Fitting the model to the normalized training data.

5.Calculating the silhouette score to assess how well data points are separated within clusters.

My Silhouette score: 0.6479480688309227.

\* Visualize the cluster distribution.

