**Customer Segmentation with K-Means Clustering**

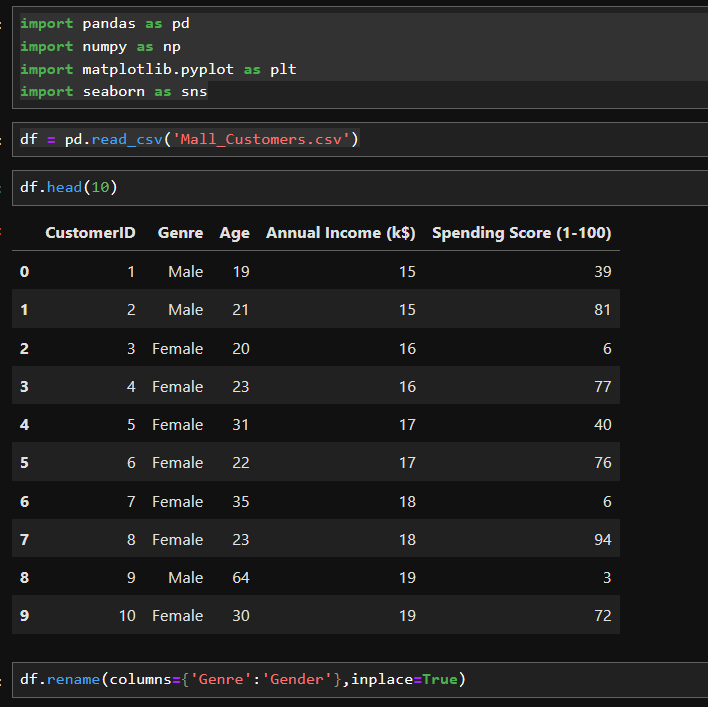
**Aim:** Apply K-Means clustering to segment customers based on their purchase behaviour.

**Description:** Use a customer purchase dataset to identify distinct segments using the K-Means clustering algorithm.

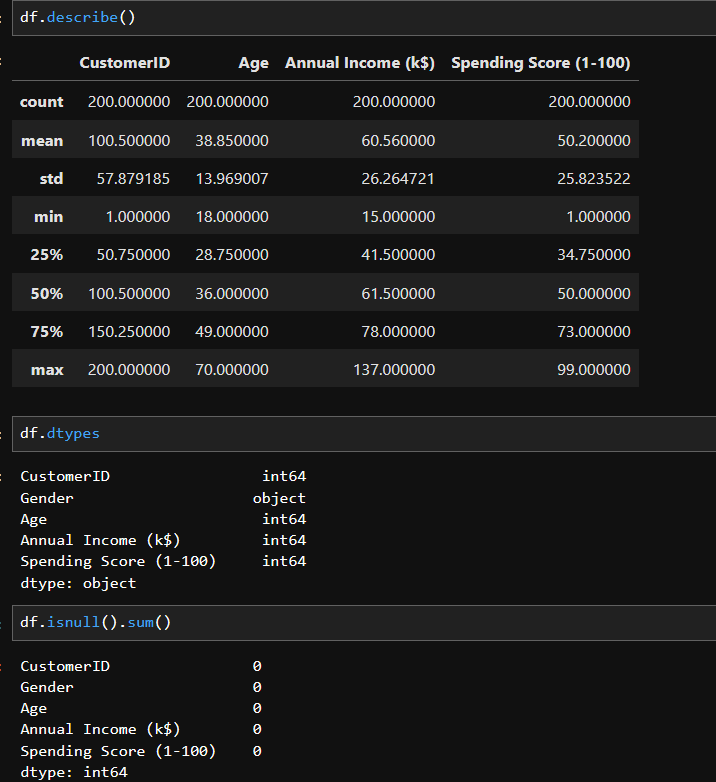
**Technologies:** Python, Pandas, Scikit-learn.

**Analysis Process:**

**1.Importing and understanding the data:**

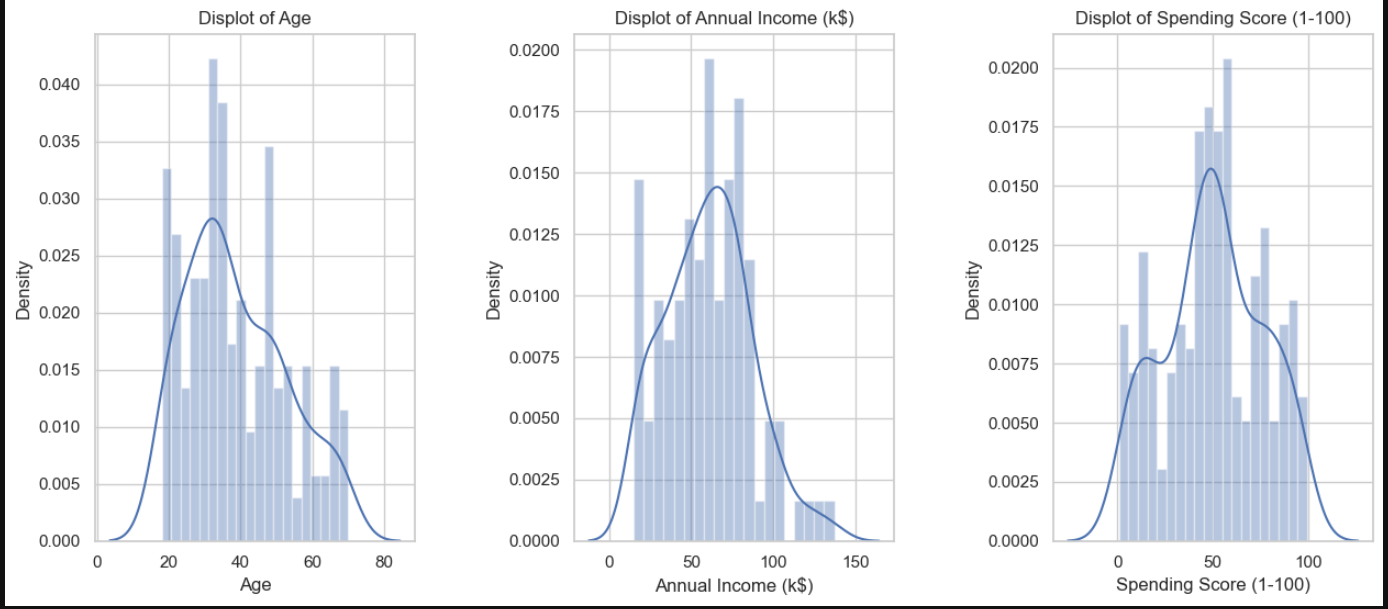
****

* Displaying first 10 records of the dataset

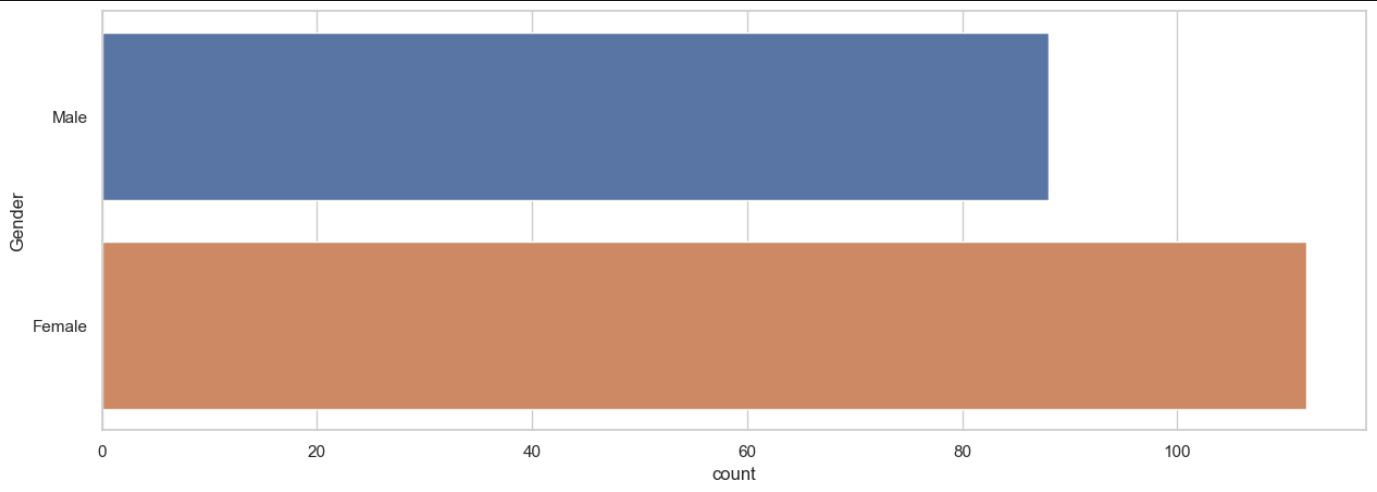
****

* Analysing the statistics of the data i.e., mean, count, maximum, minimum, etc.
* Analysing the type of the variable and to count number of null values present in each column.

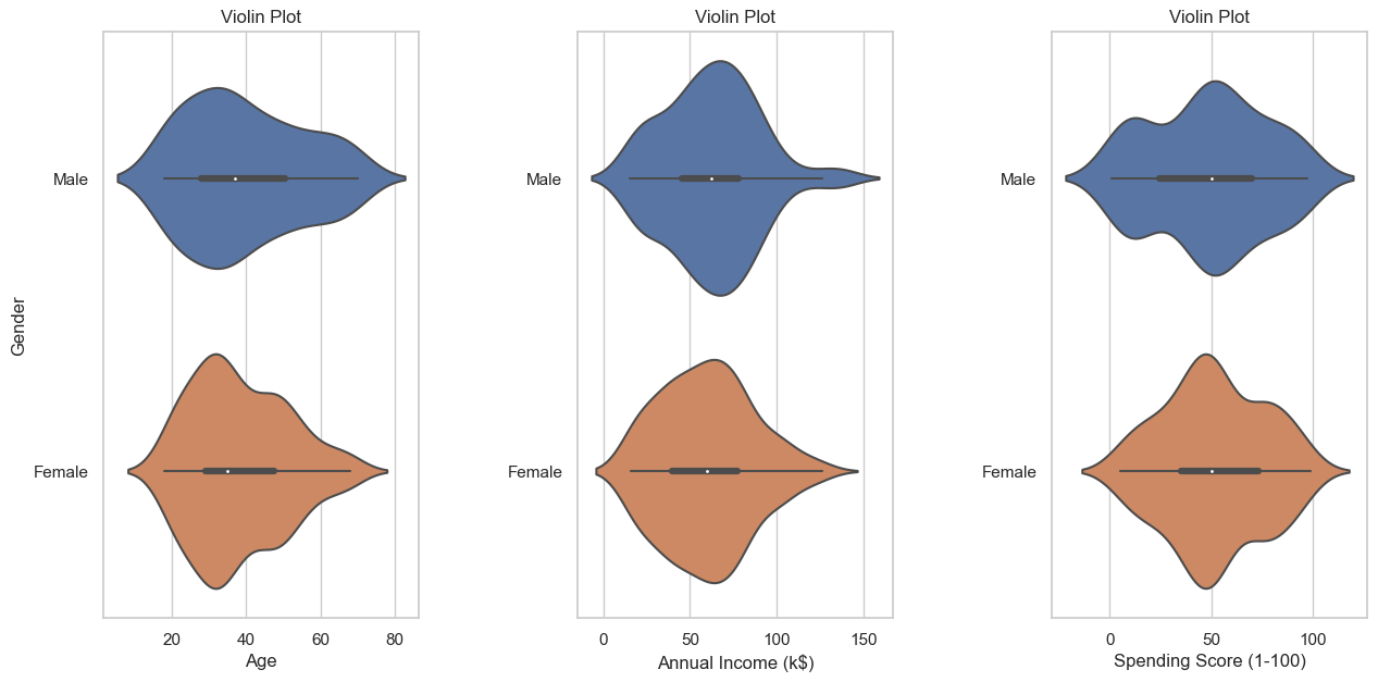
**2. Visualising the data:**

****

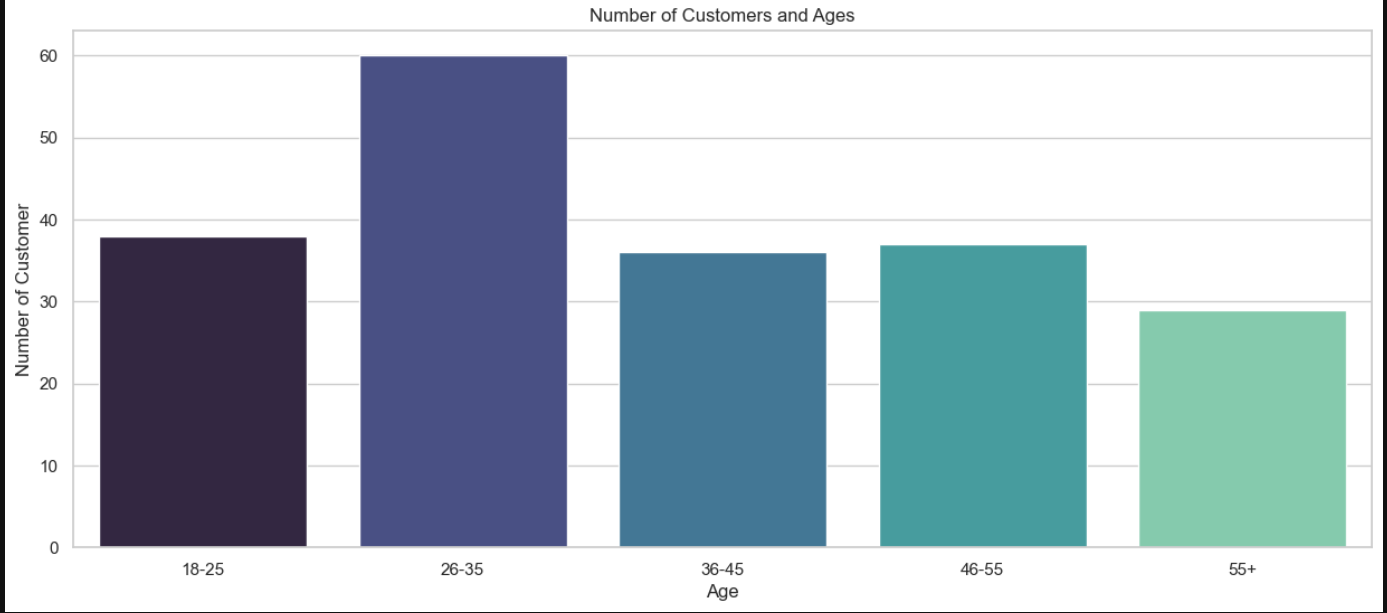
* We create three univariate graphs using displot of Age, Annual income, Spending Score (1-100) respectively.
* We can see that people of age around 30 years has the highest density, highest density of annual income is around 60-65 (k$) and highest density in spending score is around 60.

****

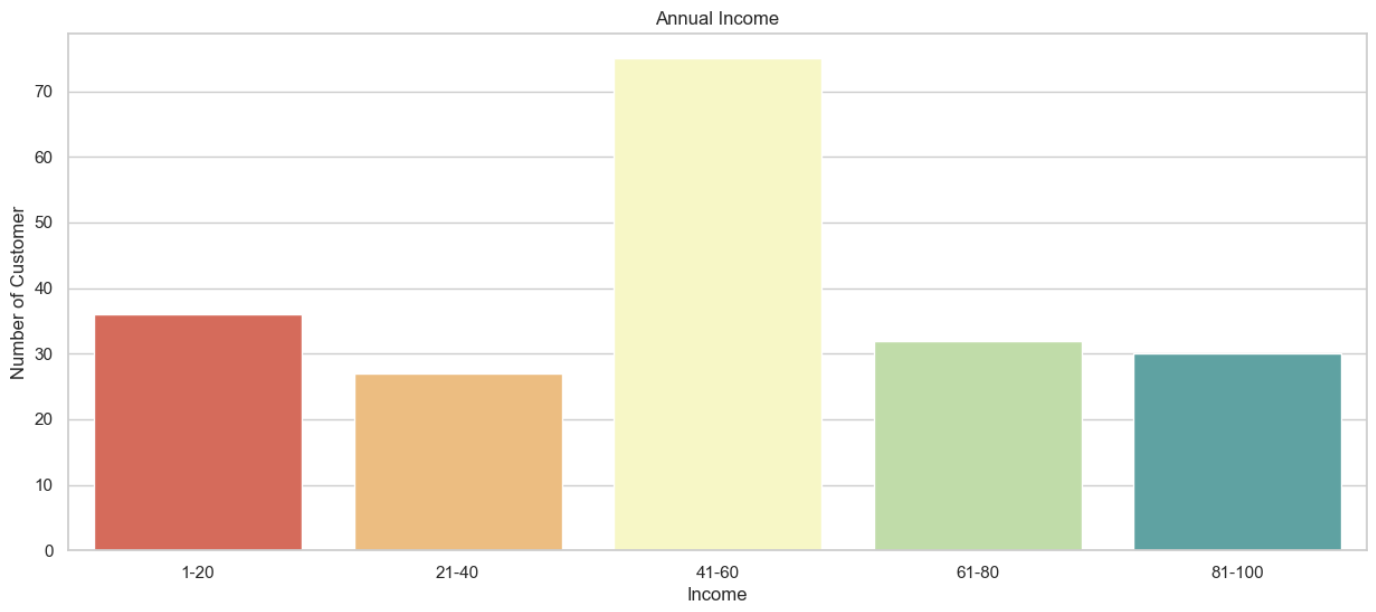
* We compare the count of female and male customers. There are more number of female customers than male customers.

****

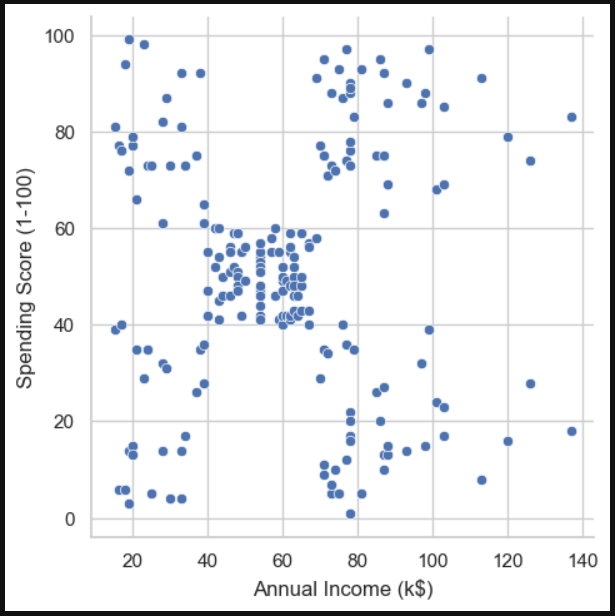
* We create three violin plots for bivariate analysis for both category of gender.

****

****

****

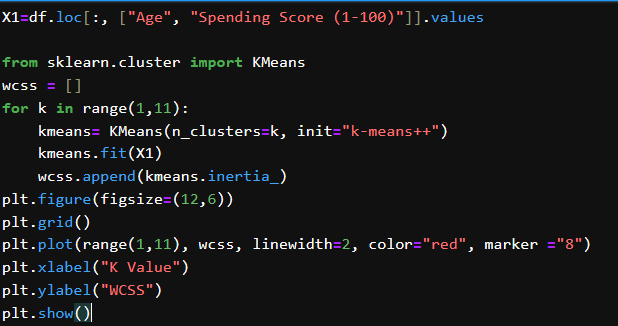
* We create three bar plots to visualise each variable’s maximum and minimum values.

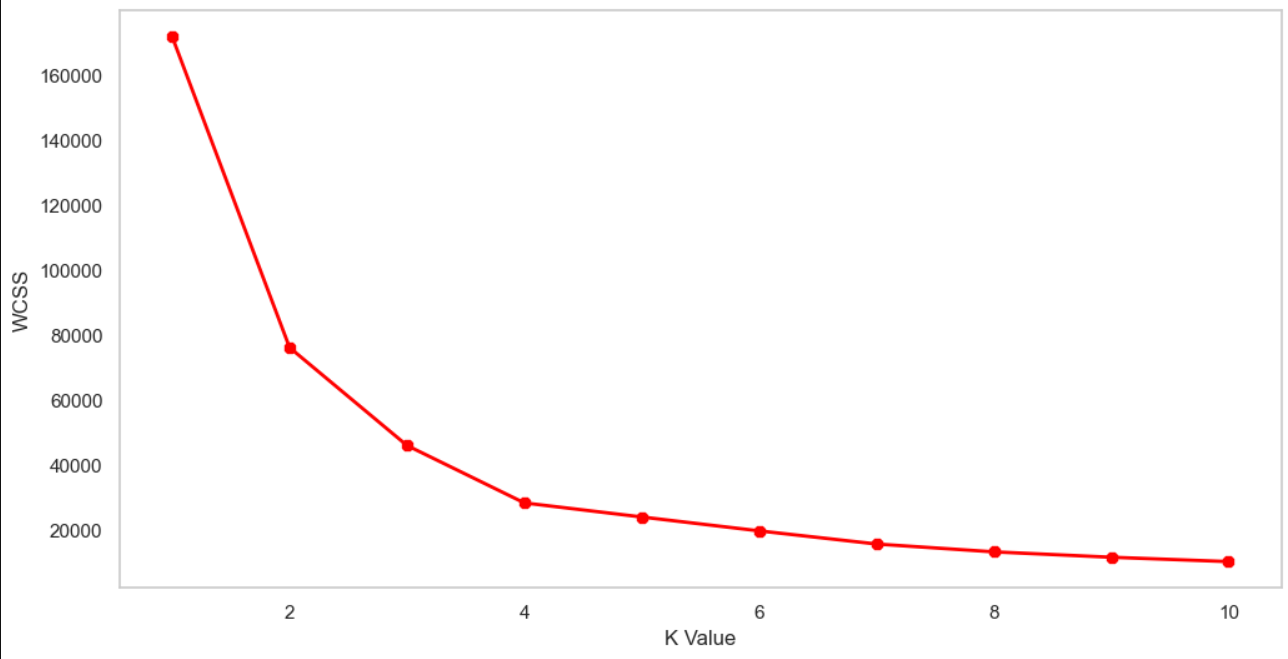
****

* In this graph, we plot a relation graph between annual income and spending score. We see there is no clear relation between these two variables but there is a similar trend in the range 60-80 annual income(k$) has spending score of 40-60.

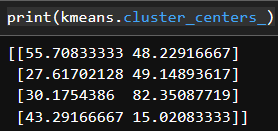
**3. K-Means Clustering:**

* Grouping age and spending score to find the number of clusters required.

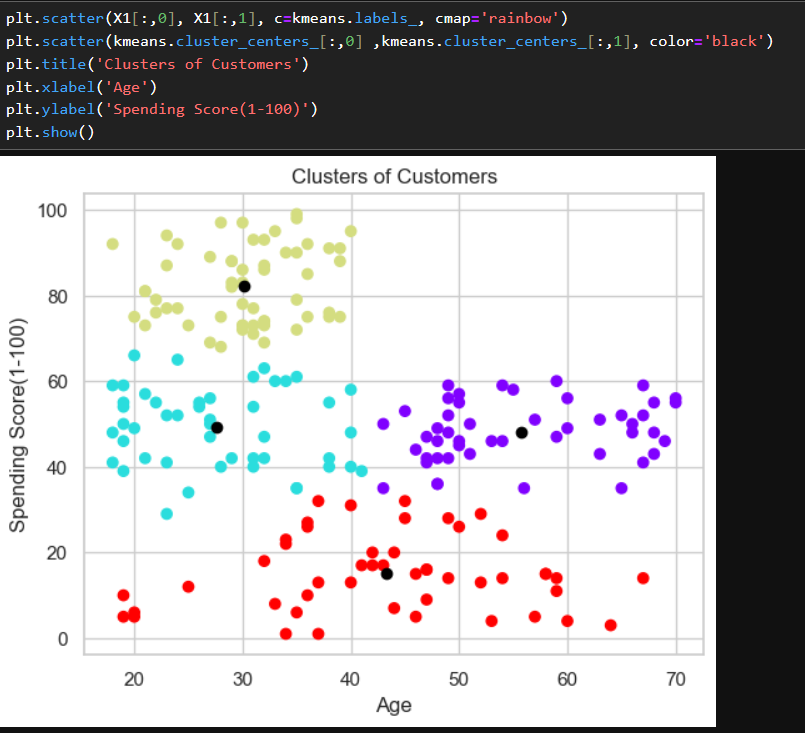
****

****

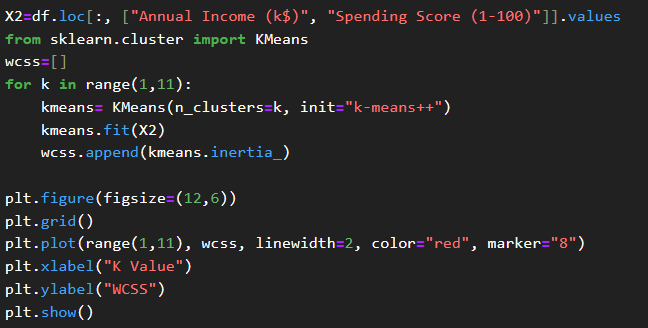
* We can see there is bend at 4 after which the graph’s slope is constant. So, we will consider 4 as number of clusters.
* We can calculate the centroids for X1:

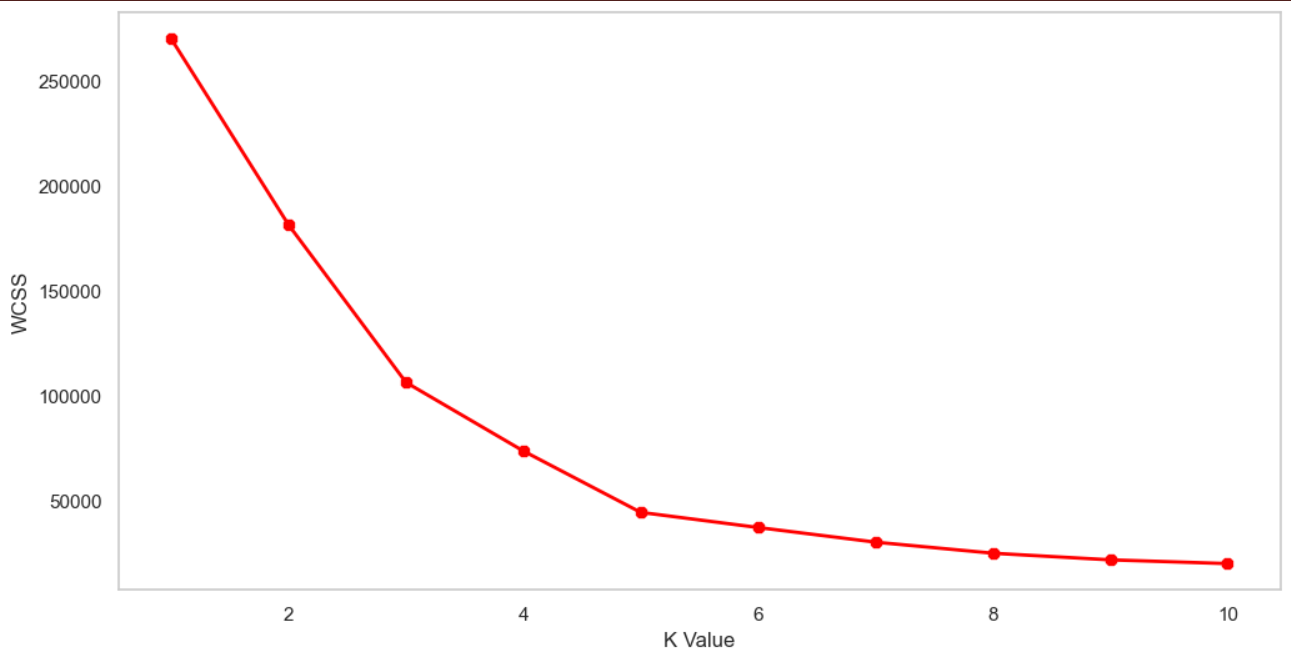
****

* We will represent the clusters in a scatter plot where, centroids will be represented in black colour.

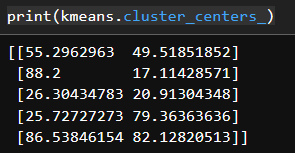
****

* Now, grouping annual income and spending score to find the number of clusters required.

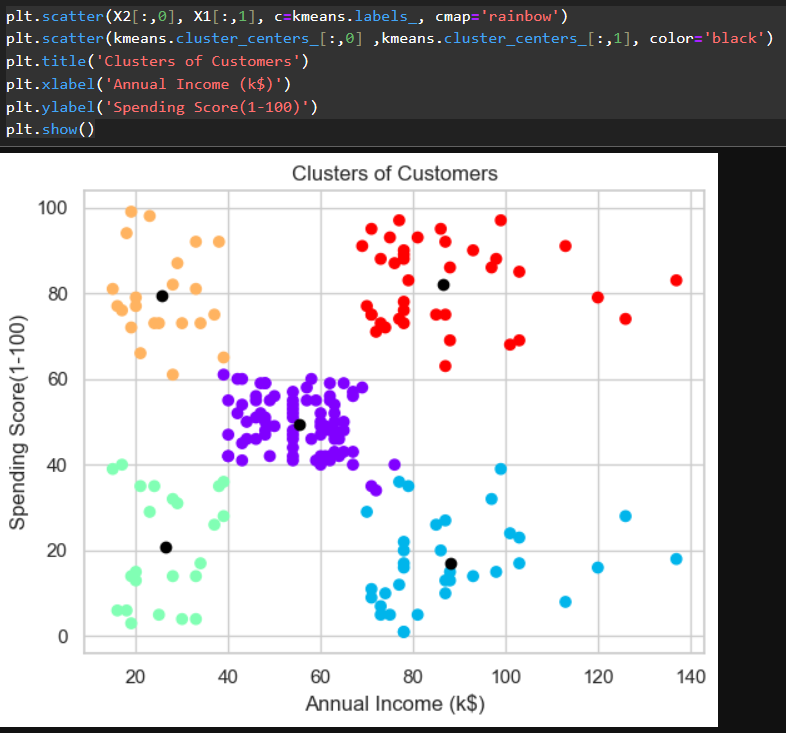
****

****

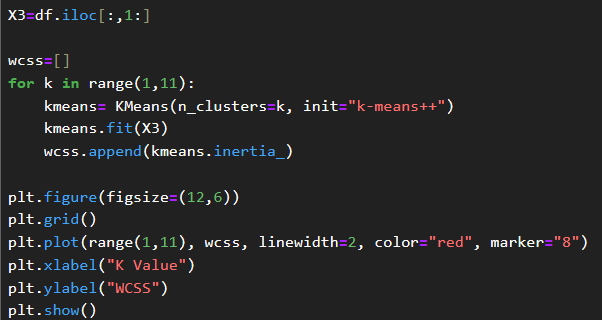
* We can see there is bend at 5 after which the graph’s slope is constant. So, we will consider 5 as number of clusters.
* We can calculate the centroids for X2:

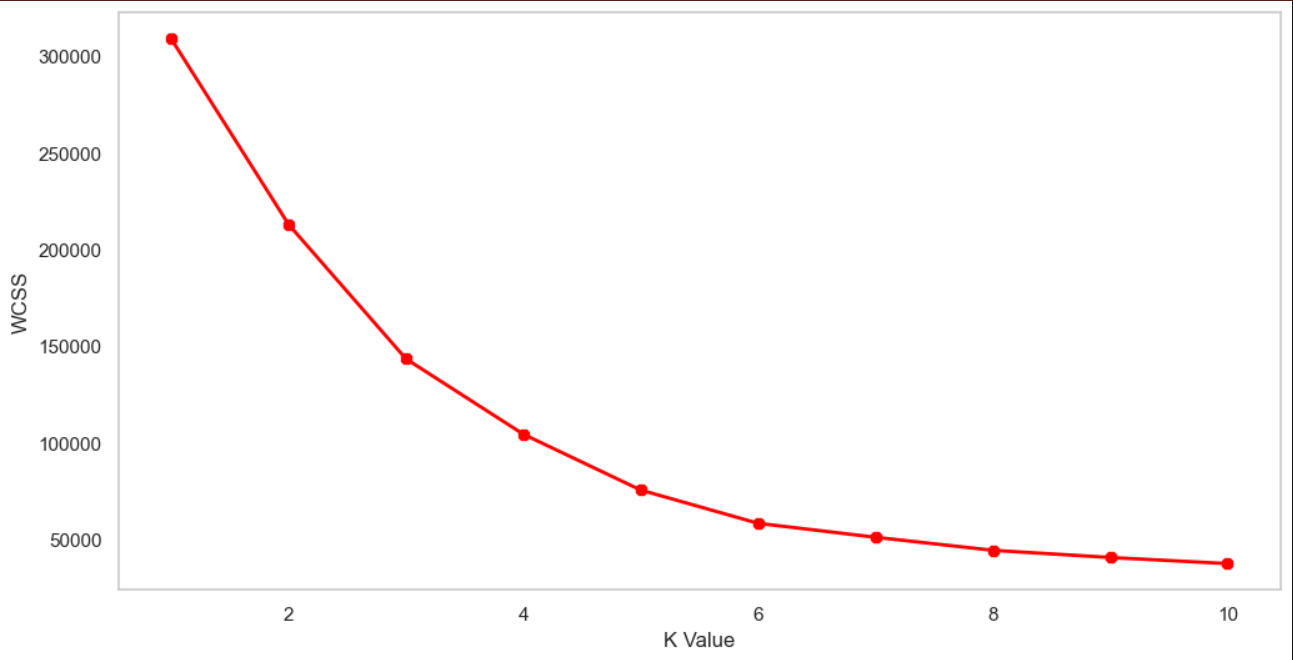
****

* We will represent the clusters in a scatter plot where, centroids will be represented in black colour.

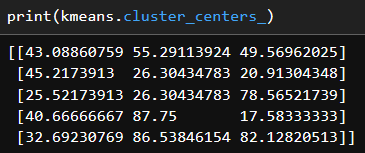
****

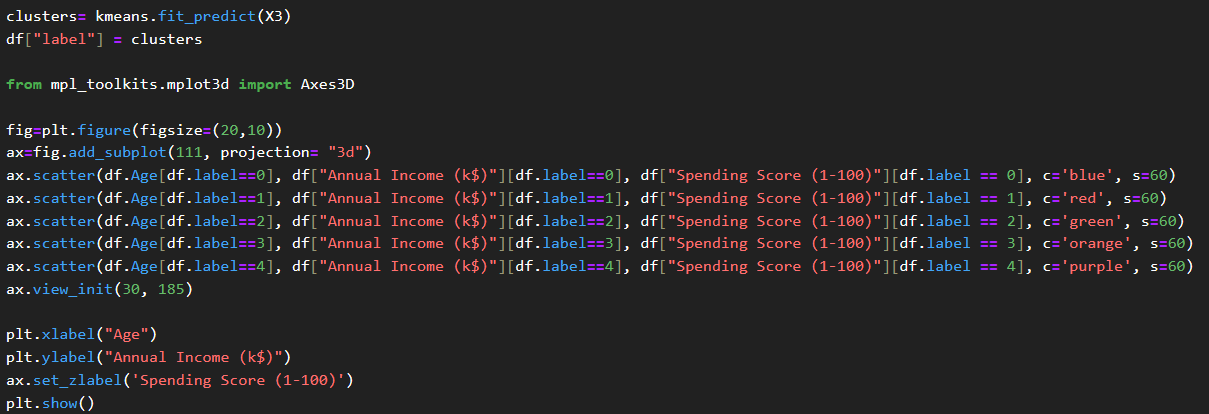
* Now, we create a 3D representation of clusters and group them in similar way.

****

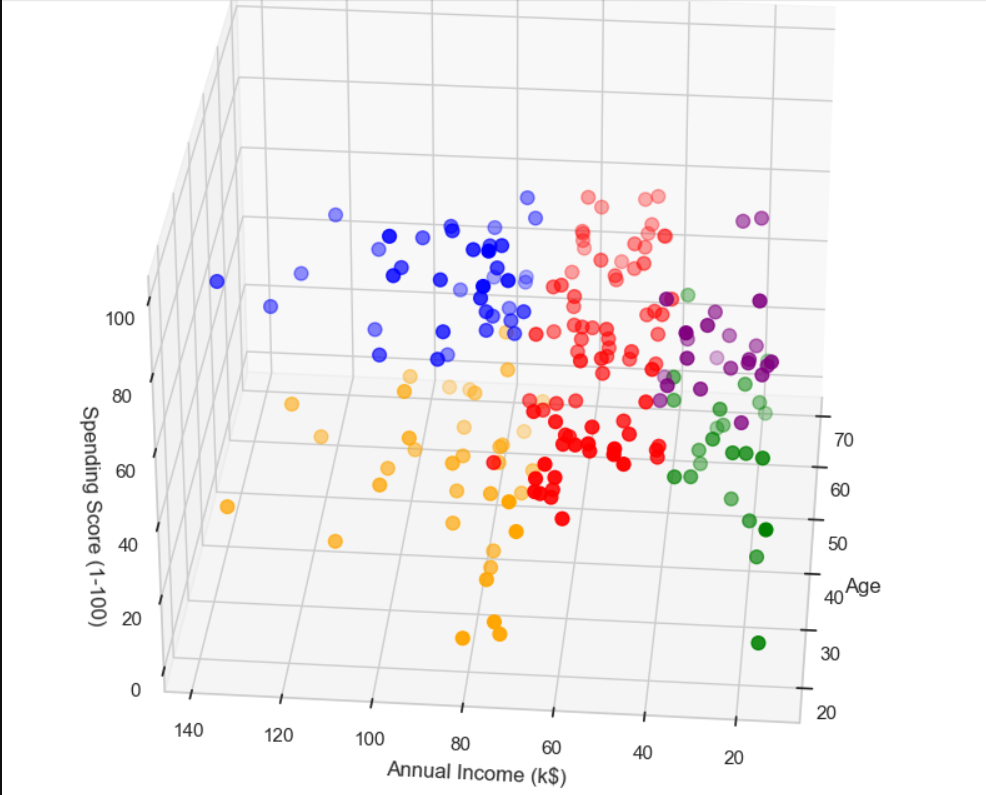
****

* We can see there is bend at 5 after which the graph’s slope is constant. So, we will consider 5 as number of clusters.
* We can calculate the centroids for X3:

****

****

**4. Result:**

****

***-Debopriya Chatterjee***