Jupyter Notebook

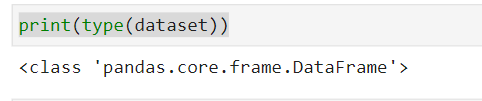
To read data from csv file we use pandas library,

Import pandas as pd

Store it in a variable

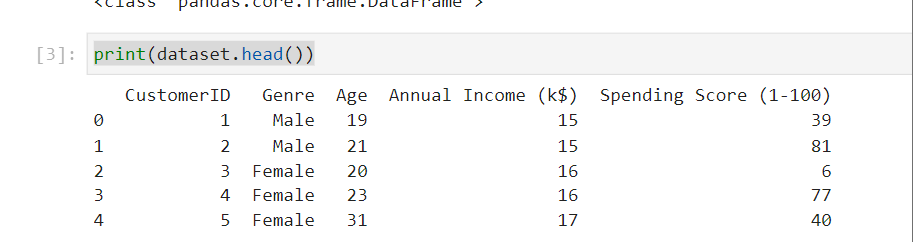
Var\_name = pd.read\_csv(‘data\_mall.csv’)



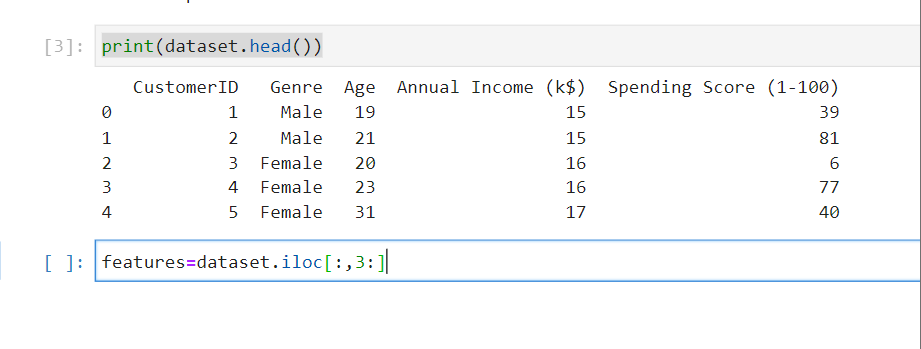


Dataset will be of type dataframe

To print top 5 data in dataset

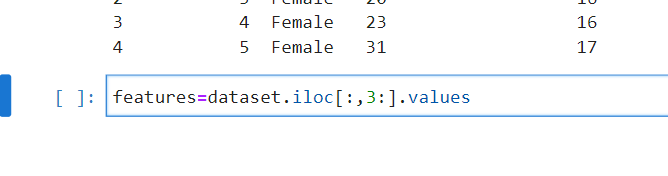


Here we want all last two columns for analysis

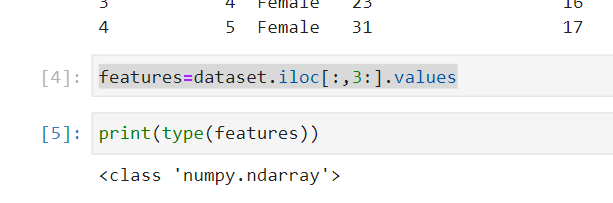


The : indicates all the rows,

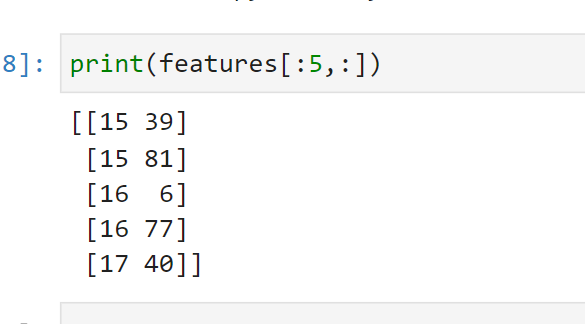
3: indicates from index 3 till end



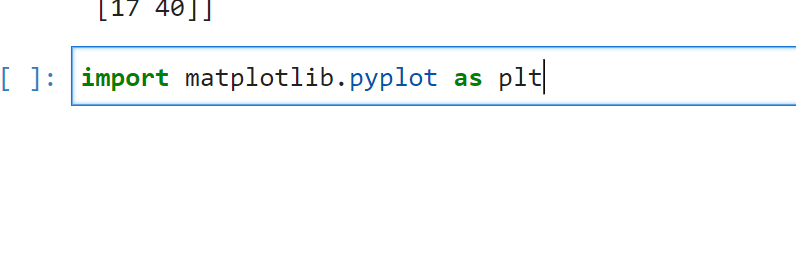
If we don’t put .values it will in dataframe, to make it ndarray we use values

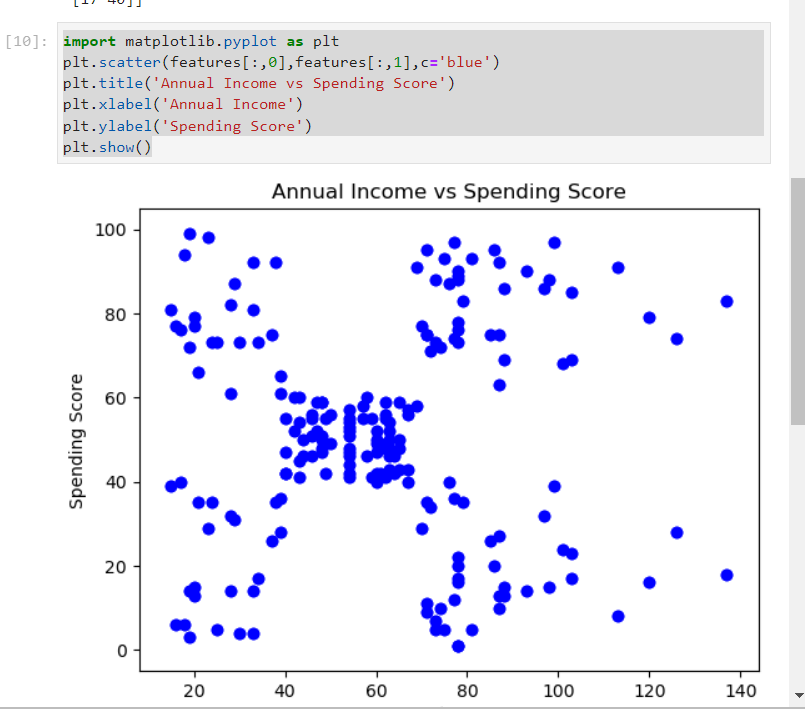


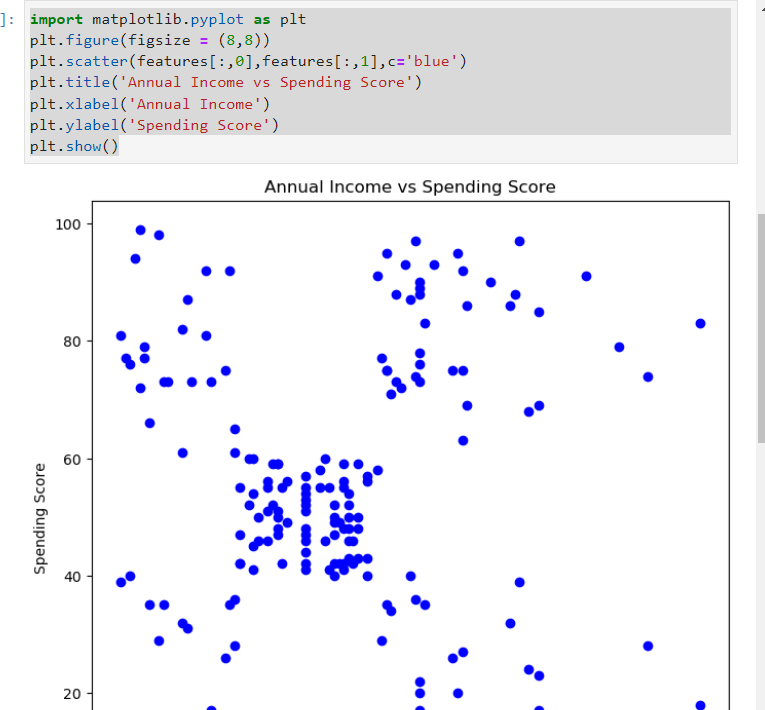
To print first 5 in features variable



For visulaisation we have to import matplotlib library

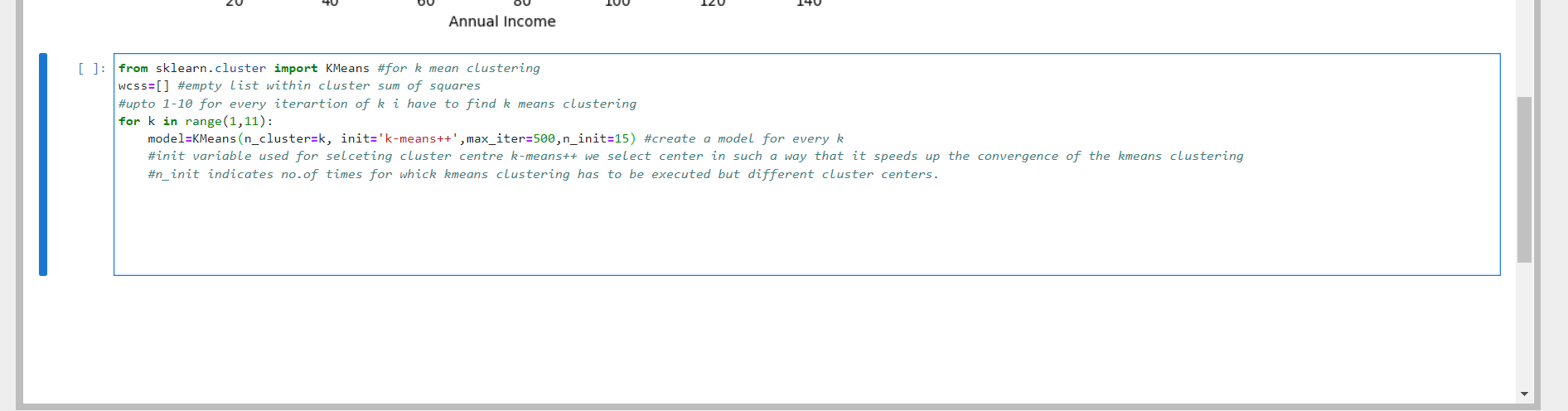




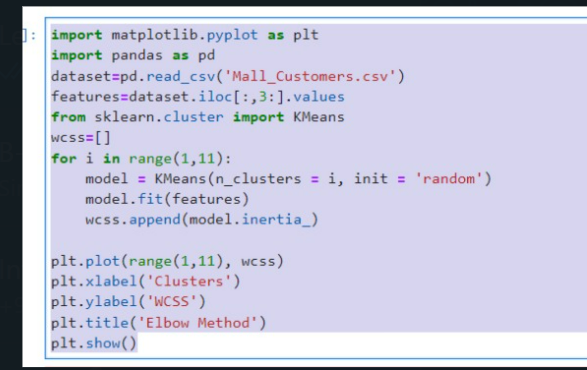


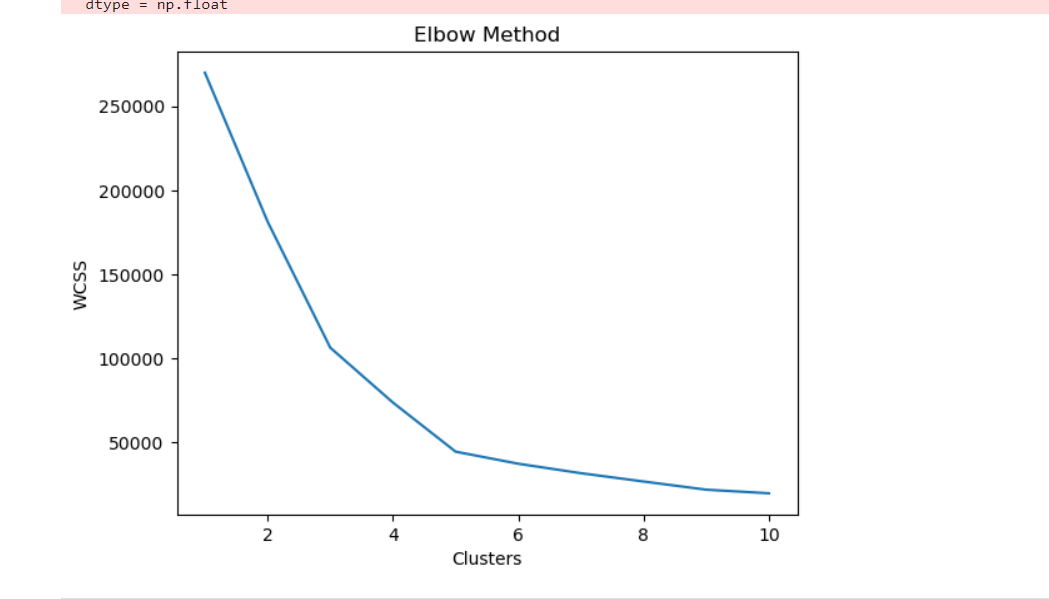
Plt.figure(figsize=(8,8)) to make the size bigger

We have to make use of elbow method to find the number of clusters(k) to divide our dataset

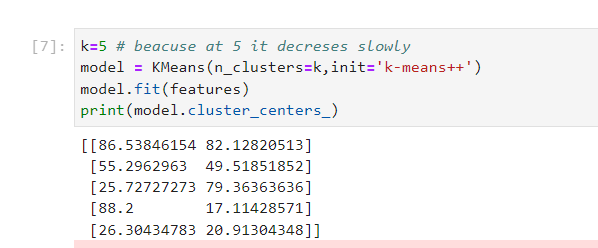


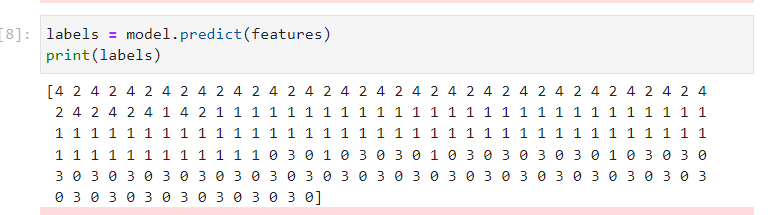
Last two parameters not required.

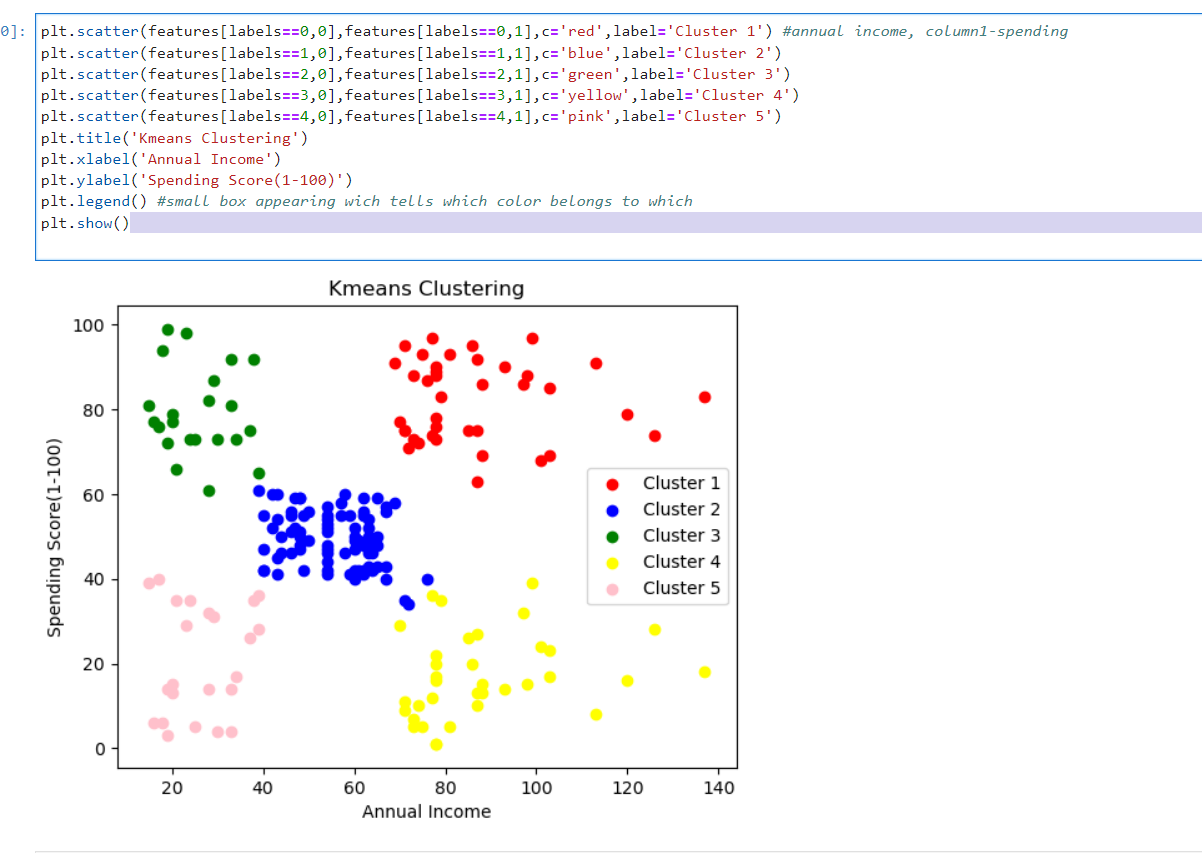


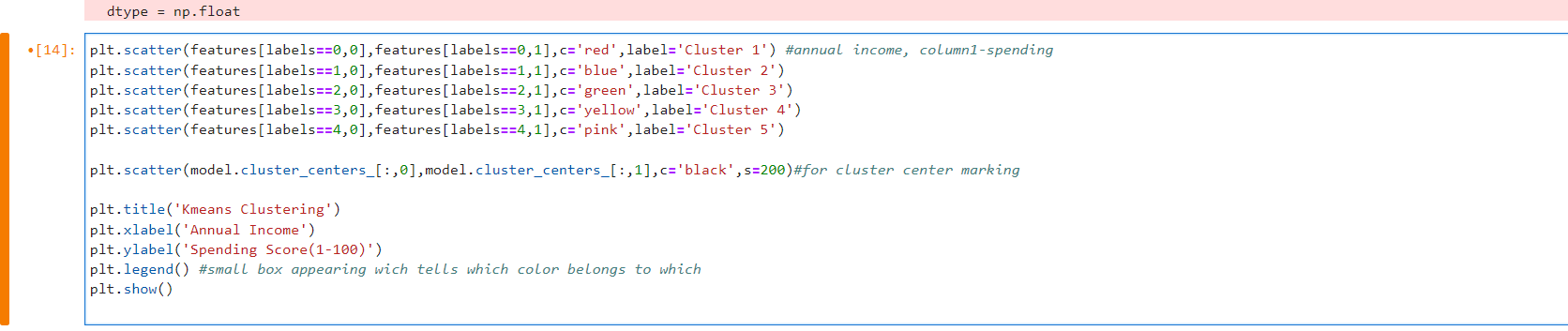


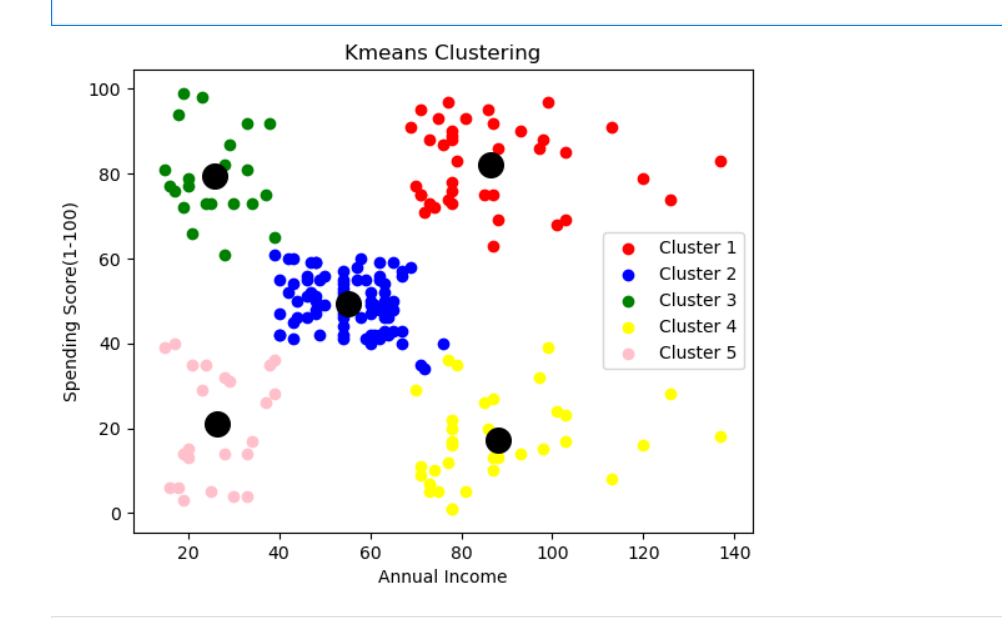
Wcss: within cluster sum of squares











Extracting meaning from the graph

The pink cluster belongs to people having low income and low spending scores.

The green cluster belongs to customer having low income and high spending scores

The red cluster belongs to customer having high income and high spending scores

The blue cluster belongs to customer having medim income and high medim scores

The yellow cluster belongs to customer having high income and low spending scores

Bigger size plot:

