National Institute Of Technology Delhi

A

Project Report

On

“Customer Segmentation”

For The Course

Data Science

SUBMITTED BY

Ruttala Gowthami Priya

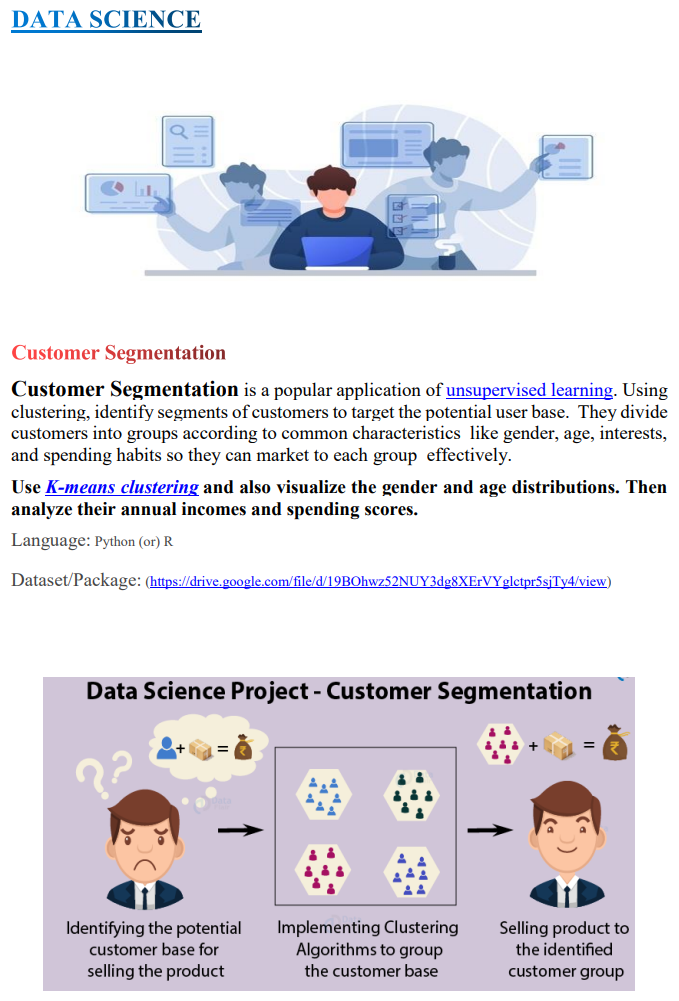
SUBMITTED TO

Exposys Data Labs

For Academic Year 2022-23

CONTENTS

* Introduction
* Implementation of Mall Customers Segmentation using K-Means Clustering
* Importing Libraries
* Working with Data set
* Visualize The data points
* Find K value using the Elbow method
* Training the K-means Algorithm on the training dataset
* Visualize the clusters formed
* References
* Conclusion



INTRODUCTION

The main objective of this project is to analyze the given data. Customer Segmentation is a popular application of unsupervised learning. Using clustering, identify segments of customers to target the potential user base. They divide customers into groups according to common characteristics like gender, age, interests, and spending habits so they can market to each group effectively.

Customer segmentation is the practice of dividing a company’s customers into groups that reflect similarity among customers in each group. The goal of segmenting customers is to decide how to relate to customers in each segment in order to maximize the value of each customer to the business.

Implementation of Mall Customers Segmentation using K-Means Clustering

What is Clustering?

Clustering is a set of techniques used to partition data into groups or clusters.

What is K-Means Algorithm?

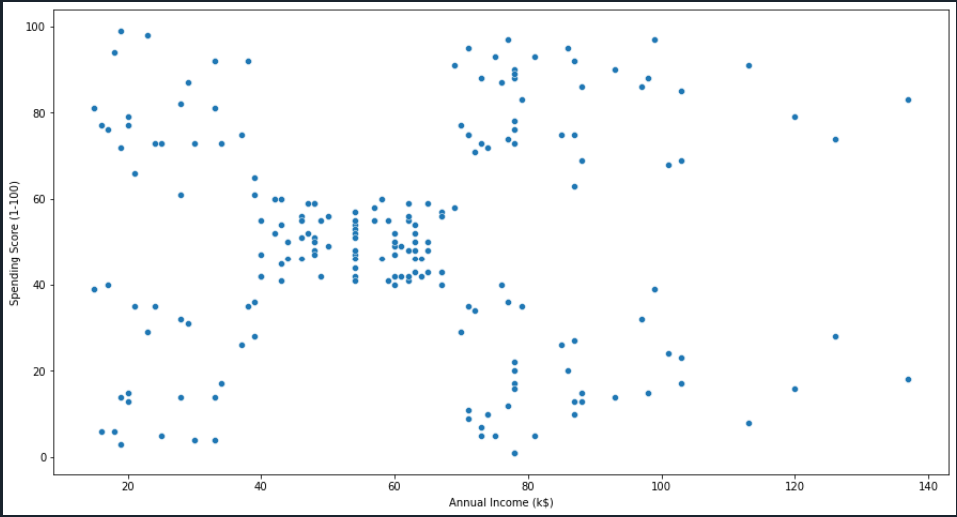
K-Means clustering is an centroid-based Unsupervised Learning Algorithm , which groups the unbalanced data set into different clusters.

Here , K defines the number of pre-defined clusters or groups that need to be created in that process , as if K=5 , there will be five clusters and for k=10 , there will be ten clusters and soon.

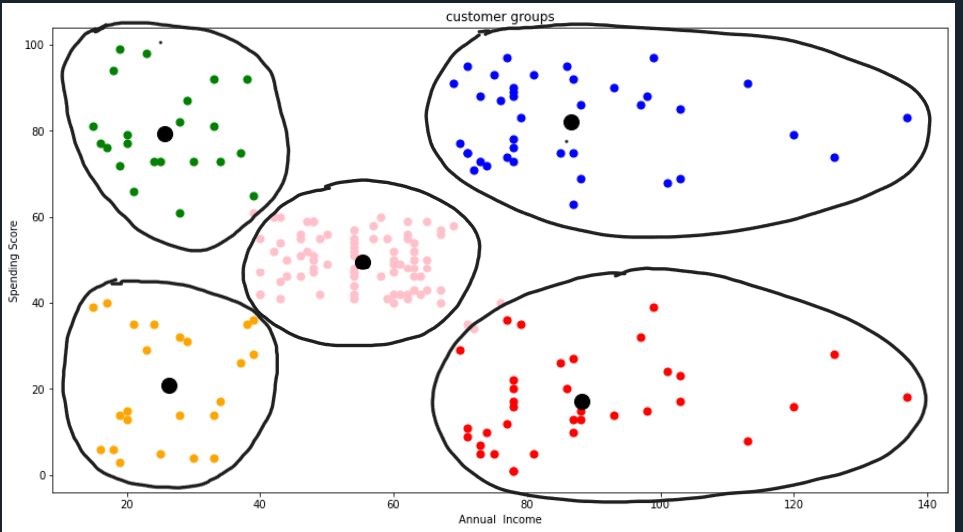
The K-Means clustering algorithm mainly performs two tasks :

1. Determines the best value for K center points.
2. Assigns each data points to its closest K-center.

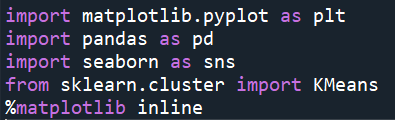
Before K-Means



After K-Means

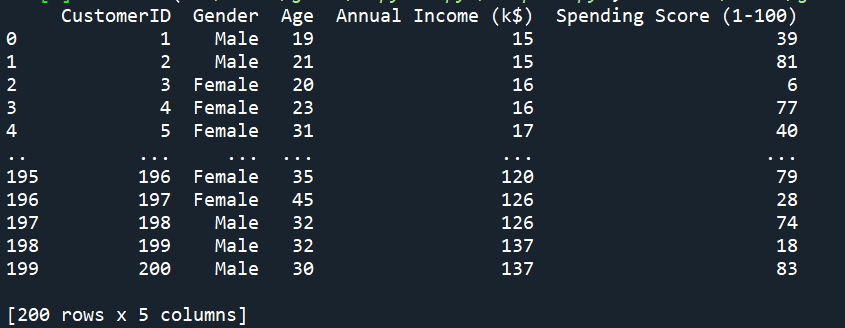


Importing Libraries

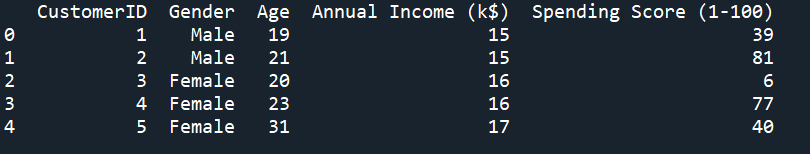


Working with Data set

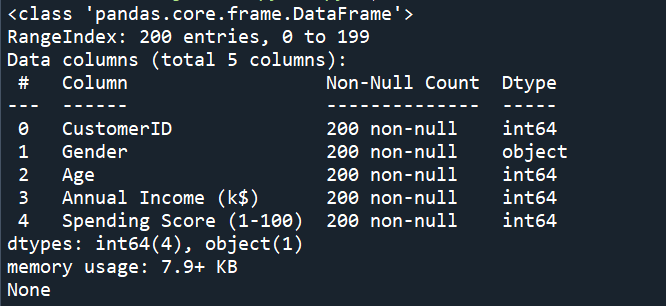




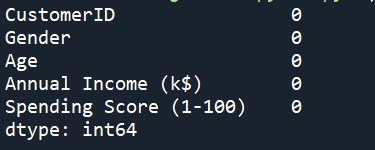








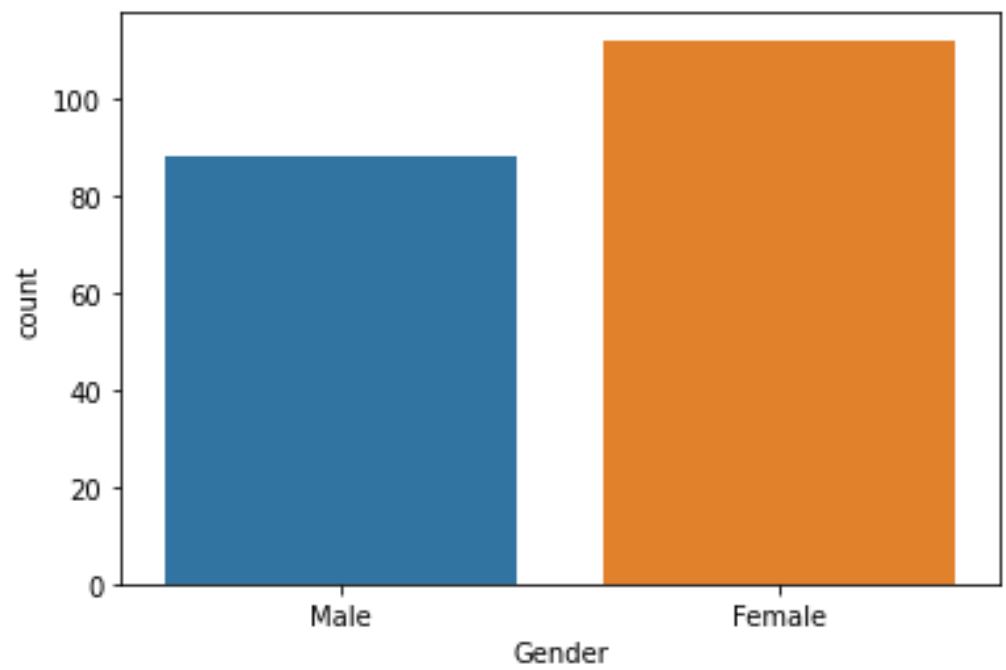




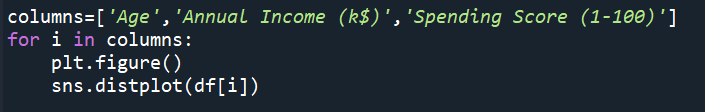
Visualize The data points

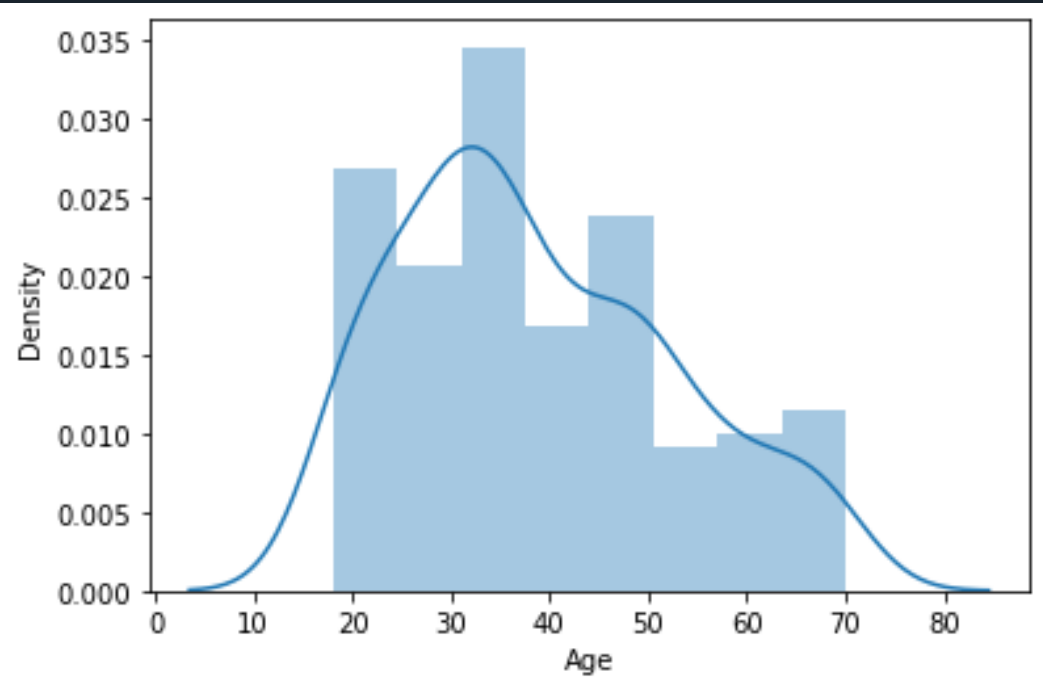
Gender vs no. of customers

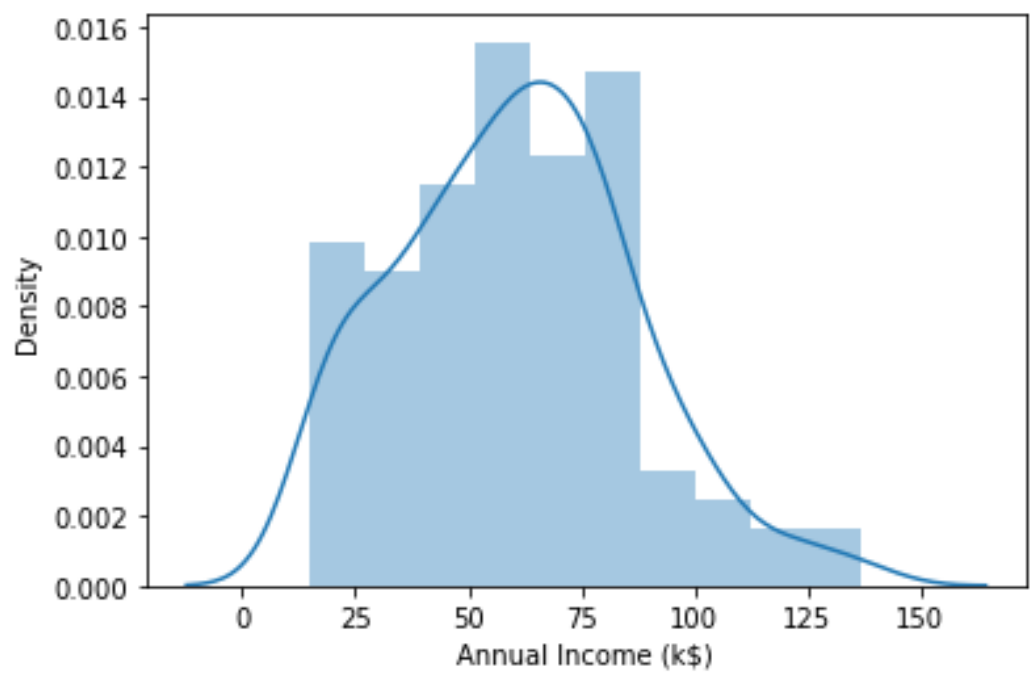


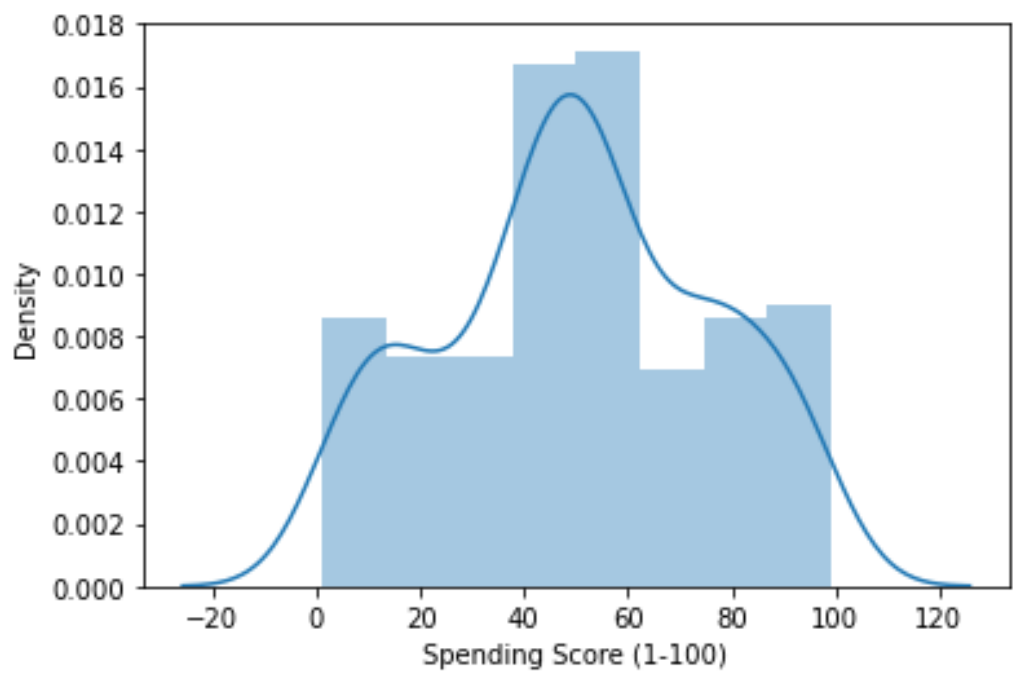


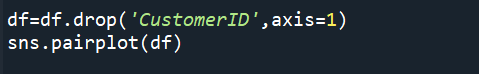
Age , Annual Income , Spending Score

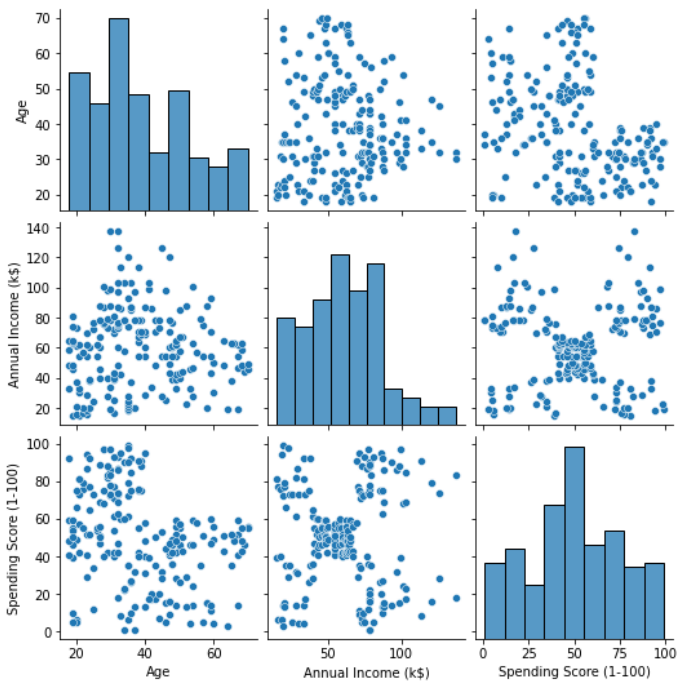


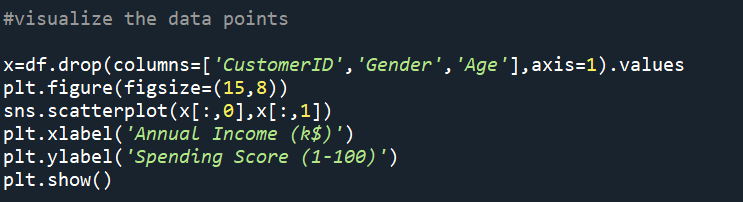


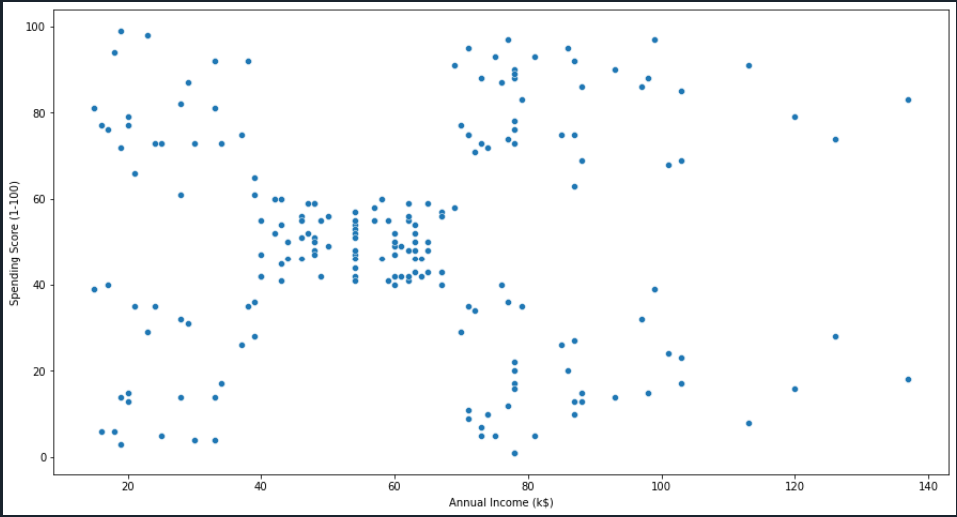




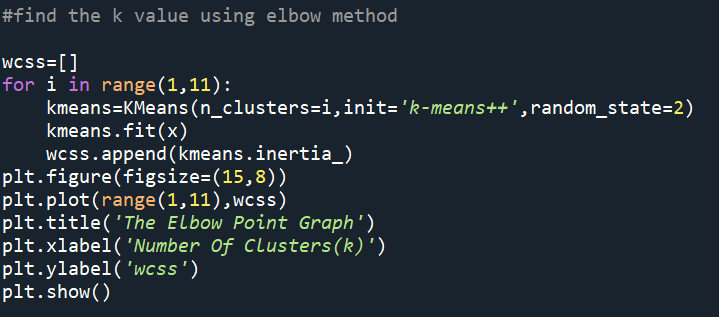


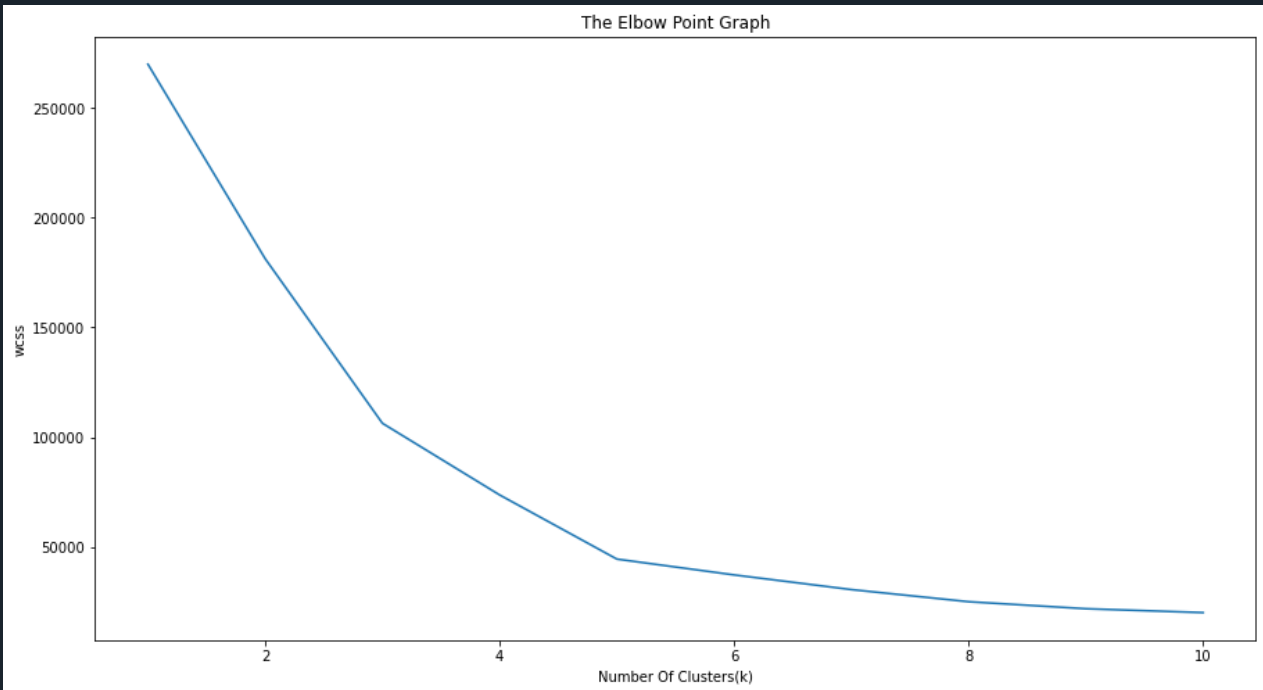






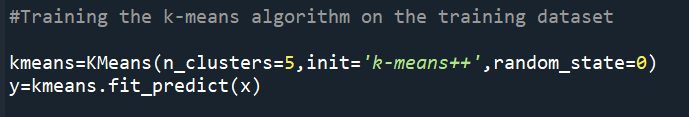
Find K value using the Elbow method



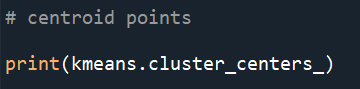


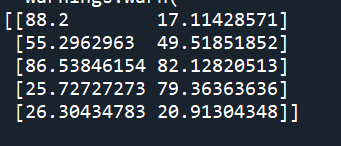
Wcss doesn’t reduce much after k=5 so, we can choose 5 as the perfect K value or clusters.

Training the K-means Algorithm on the training dataset

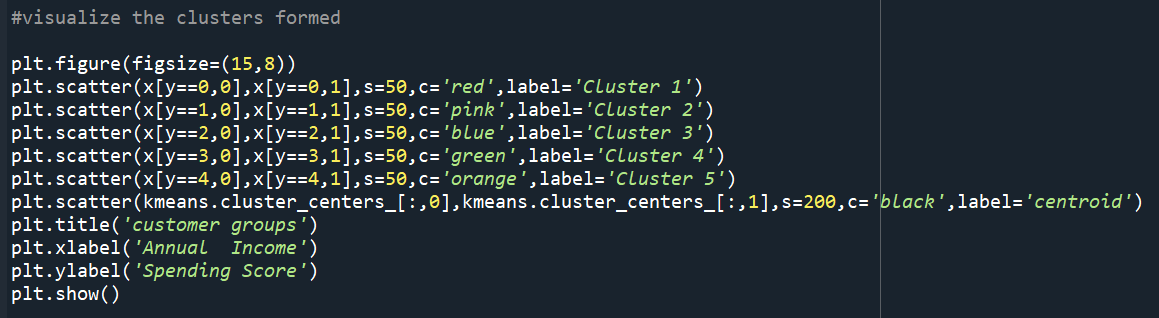


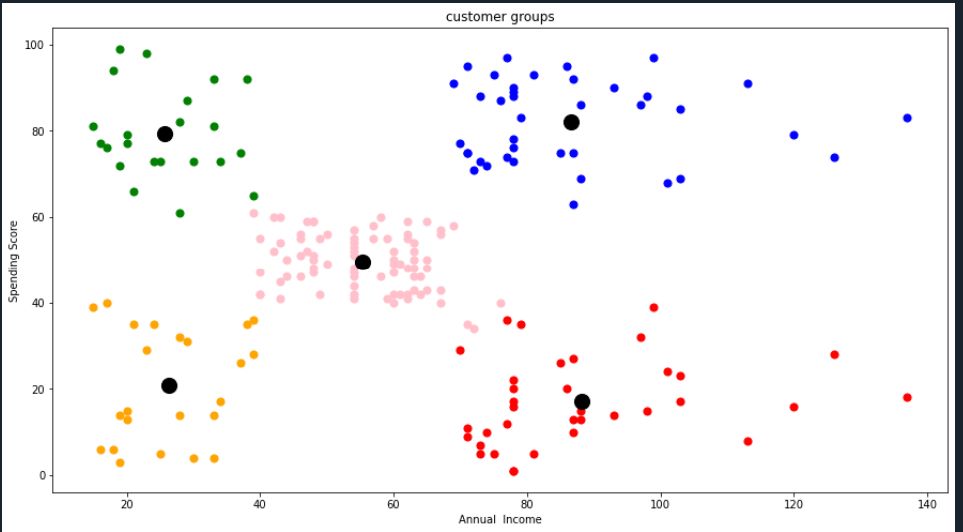
Centroid points



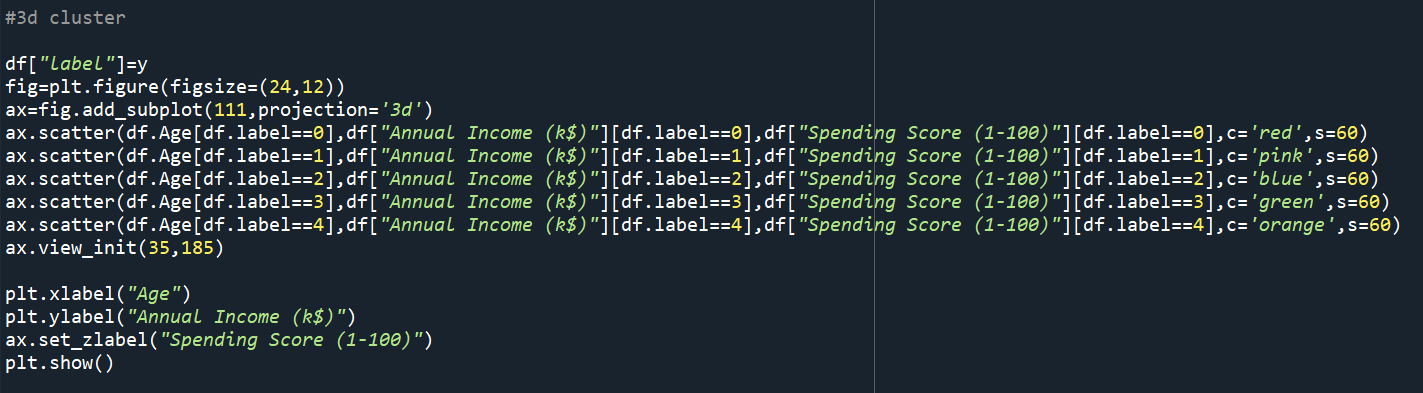


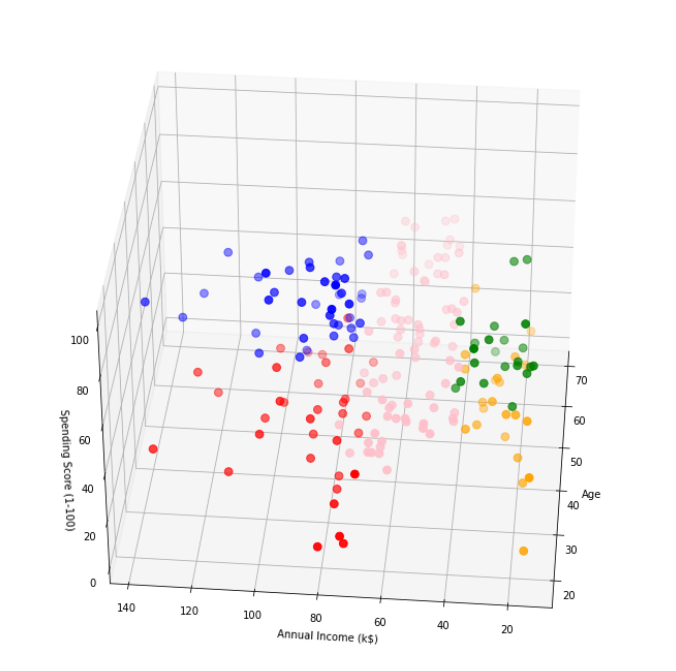
Visualize the clusters formed





3D cluster





References :-

* <https://en.wikipedia.org/wiki/Market_segmentation>
* Python Data Science Handbook Essential Tools for Working with Data by Jake Vander Plas
* <https://www.youtube.com/>

Conclusion :-

Hence , we learned one of the most famous clustering algorithms : K-Means

We can clearly see that five clusters are formed.

The black dots represents the centroid for each cluster.

THANKYOU!!!