**Project Based Learning Report**

on

**Build an application for customer segments with machine learning (k means clustering**

Submitted in the partial fulfillment of the requirements

For the Project based learning in (**Essential of Data Science**)

in

Electronics & Communication Engineering

By

**20141111027 Aashish Anand**

**2014111038 Kartik Devarde**

**2014111032 Krishna Bella**

Under the guidance of Course In-charge

Prof. **Arundhati A Shinde**

Department of Electronics & Communication Engineering

Bharati Vidyapeeth

(Deemed to be University)

College of Engineering,

Pune – 4110043

**Academic Year: 2021-22**

**Bharati Vidyapeeth**

**(Deemed to be University)**

**College of Engineering,**

**Pune – 411043**

**DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING**

**CERTIFICATE**

Certified that the Project Based Learning report entitled, **“Build an application for customer segments with machine learning (k means clustering )”** is work done by

**20141111027 Aashish Anand**

**2014111038 Kartik Devarde**

**2014111032 Krishna Bella**

in partial fulfillment of the requirements for the award of credits for Project Based Learning (PBL) in **Essential of Data Science** of Bachelor of Technology Semester IV, in Branch name.

**Prof. Dnyndeo L Dr.Tanuja Dhope**

**Course In-charge PBLCo-Ordinator**

**Dr. Arundhati A Shinde**

**Professor & Head**

**ELECTRONICS & COMMUNICATION ENGINEERING**

|  |  |
| --- | --- |
| **INDEX: -** | |
| **Page No.** | **Contents** |
| 1-2 | Description of |
| 3-10 | Results with Analysis |
| 10-10 | Outcome & Conclusion |

**Description of K-mean Clustering:**

**1**

# **K-Means Clustering Algorithm**

K-Means Clustering is an unsupervised learning algorithm that is used to solve the clustering problems in machine learning or data science. In this topic, we will learn what is K-means clustering algorithm, how the algorithm works, along with the Python implementation of k-means clustering.

What is K-Means Algorithm?

K-Means Clustering is an  [Unsupervised Learning algorithm](https://www.javatpoint.com/unsupervised-machine-learning) which groups the unlabeled dataset into different clusters. Here K defines the number of pre-defined clusters that need to be created in the process, as if K=2, there will be two clusters, and for K=3, there will be three clusters, and so on.

It allows us to cluster the data into different groups and a convenient way to discover the categories of groups in the unlabeled dataset on its own without the need for any training.

It is a centroid-based algorithm, where each cluster is associated with a centroid. The main aim of this algorithm is to minimize the sum of distances between the data point and their corresponding clusters.

The algorithm takes the unlabeled dataset as input, divides the dataset into k-number of clusters, and repeats the process until it does not find the best clusters. The value of k should be predetermined in this algorithm.

The k-means [clustering](https://www.javatpoint.com/clustering-in-machine-learning)

The algorithm mainly performs two tasks:

* Determines the best value for K center points or centroids by an iterative process.
* Assigns each data point to its closest k-center. Those data points which are near to the particular k-center, create a cluster.

Hence each cluster has data points with some commonalities, and it is away from other clusters.

The below diagram explains the working of the K-means Clustering Algorithm:



## **How does the K-Means Algorithm Work?**

The working of the K-Means algorithm is explained in the below steps:

**Step-1:** Select the number K to decide the number of clusters.

**Step-2:** Select random K points or centroids. (It can be other from the input dataset).

**Step-3:** Assign each data point to their closest centroid, which will form the predefined K clusters

**Step-4:** Calculate the variance and place a new centroid of each cluster.

**Step-5:** Repeat the third steps, which means reassign each datapoint to the new closest centroid of each cluster.

**Step-6:** If any reassignment occurs, then go to step-4 else go to FINISH.

**Step-7**: The model is ready.

**Results with Analysis:**

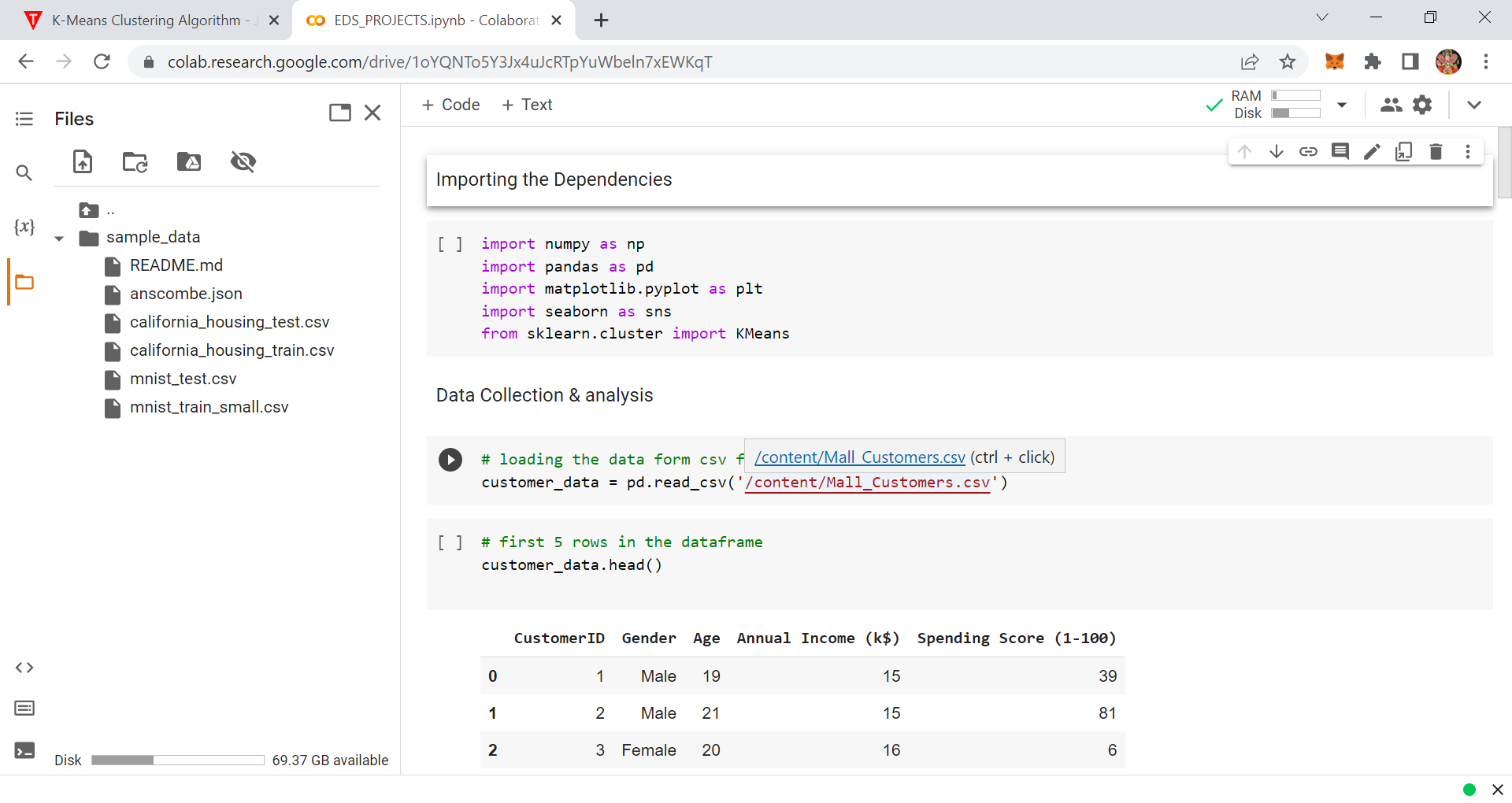
Environment and Tools:

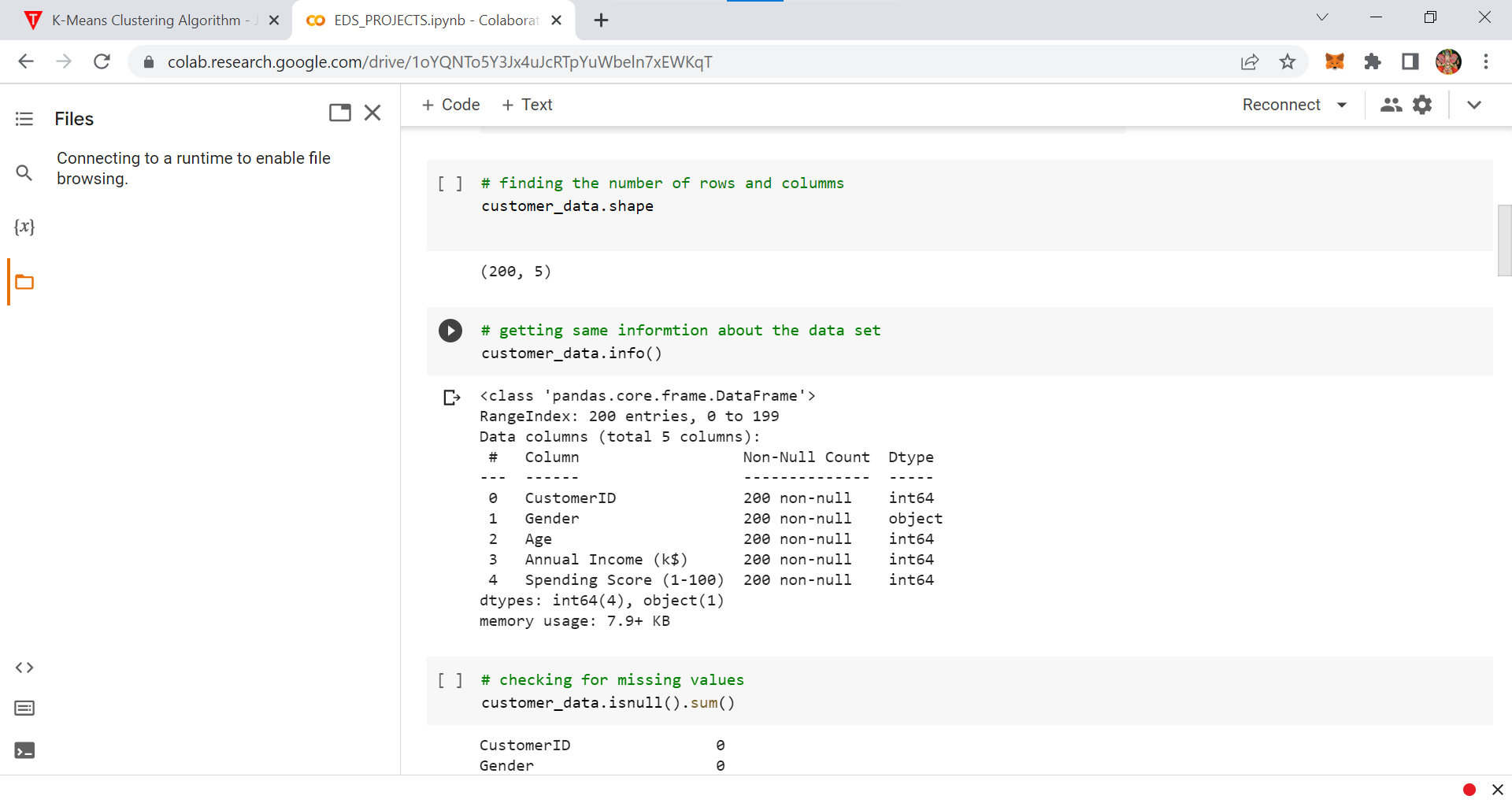
1.Colab

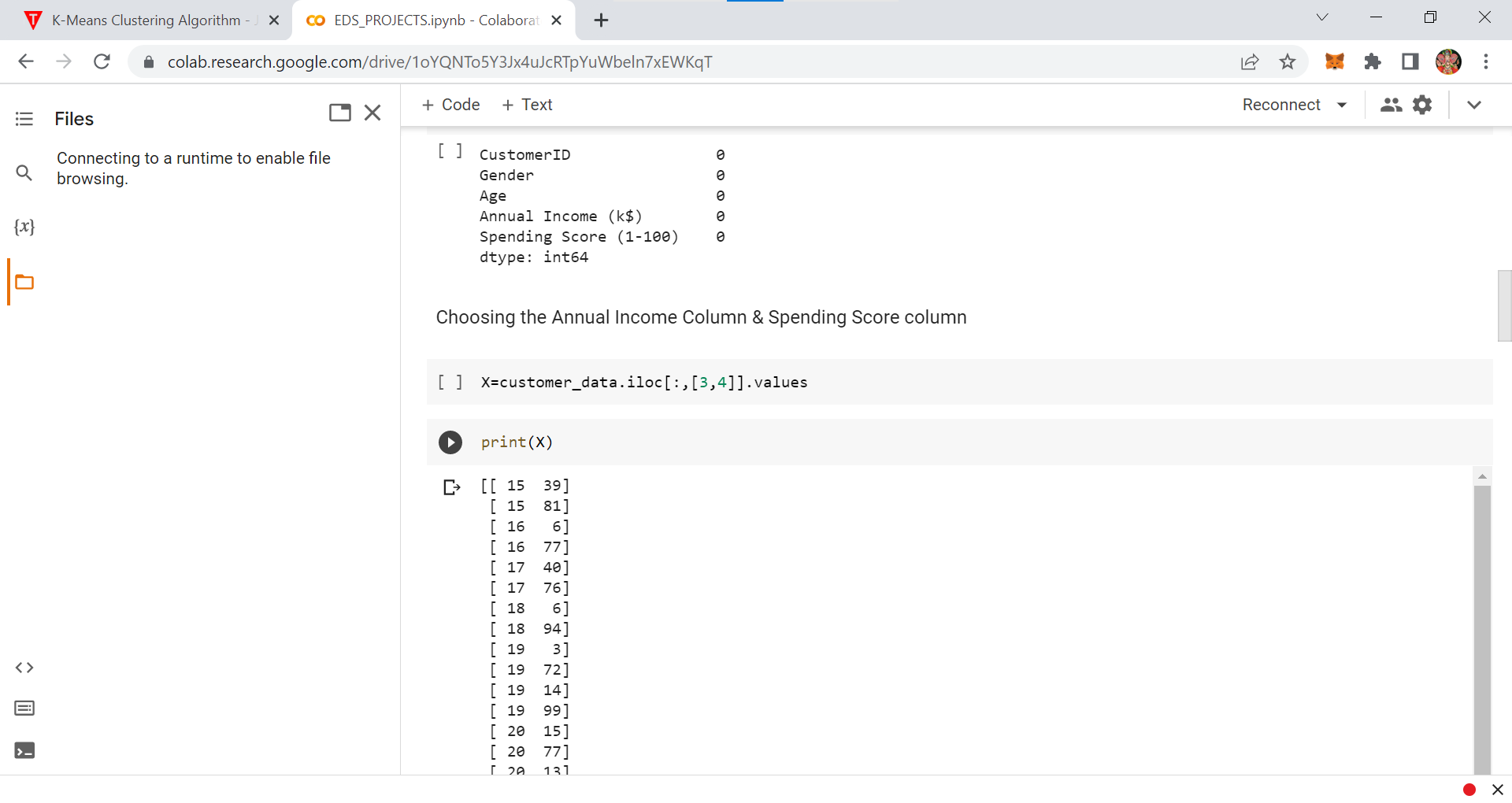
2.numpy

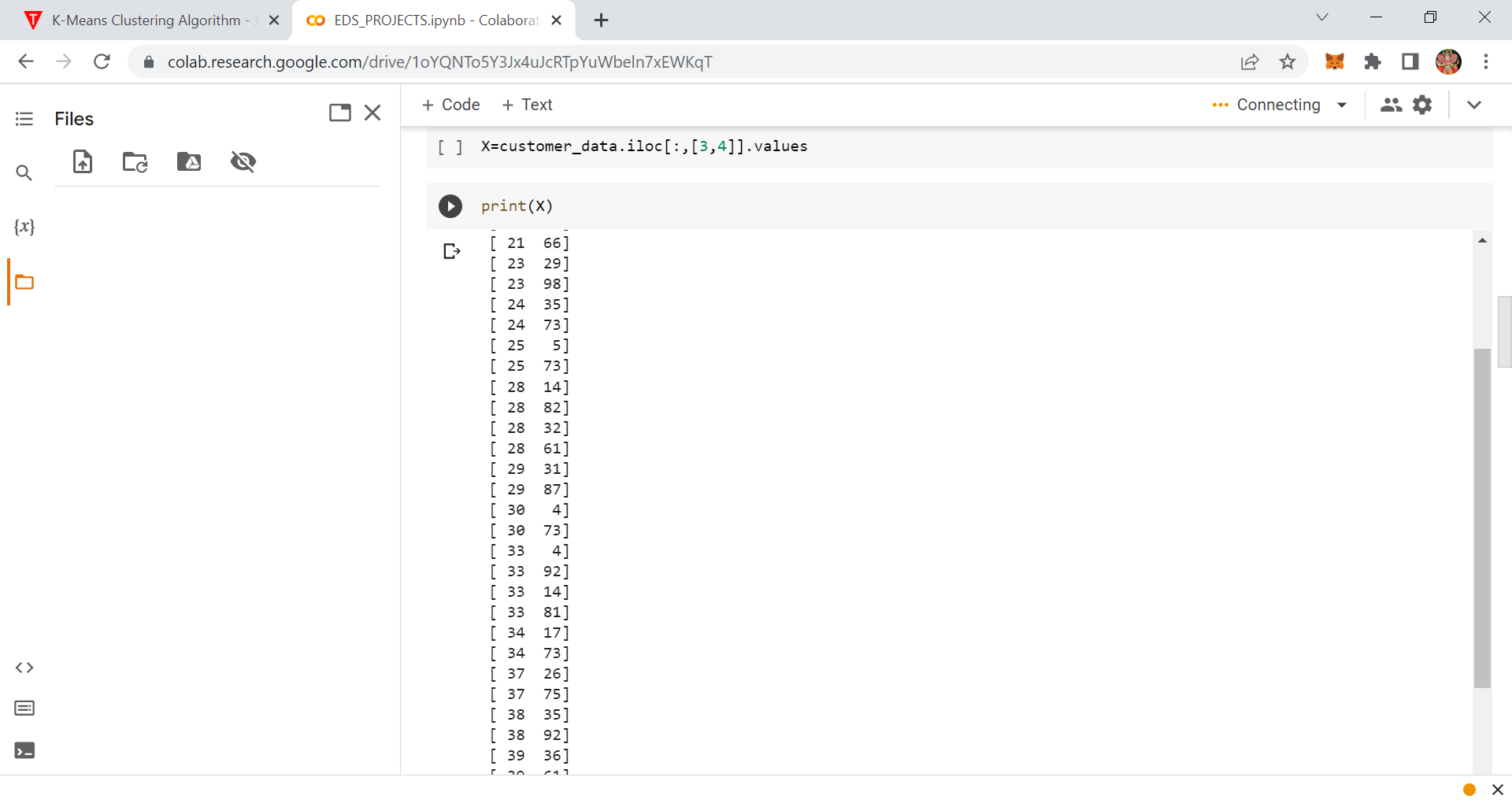
3.keggly data sets

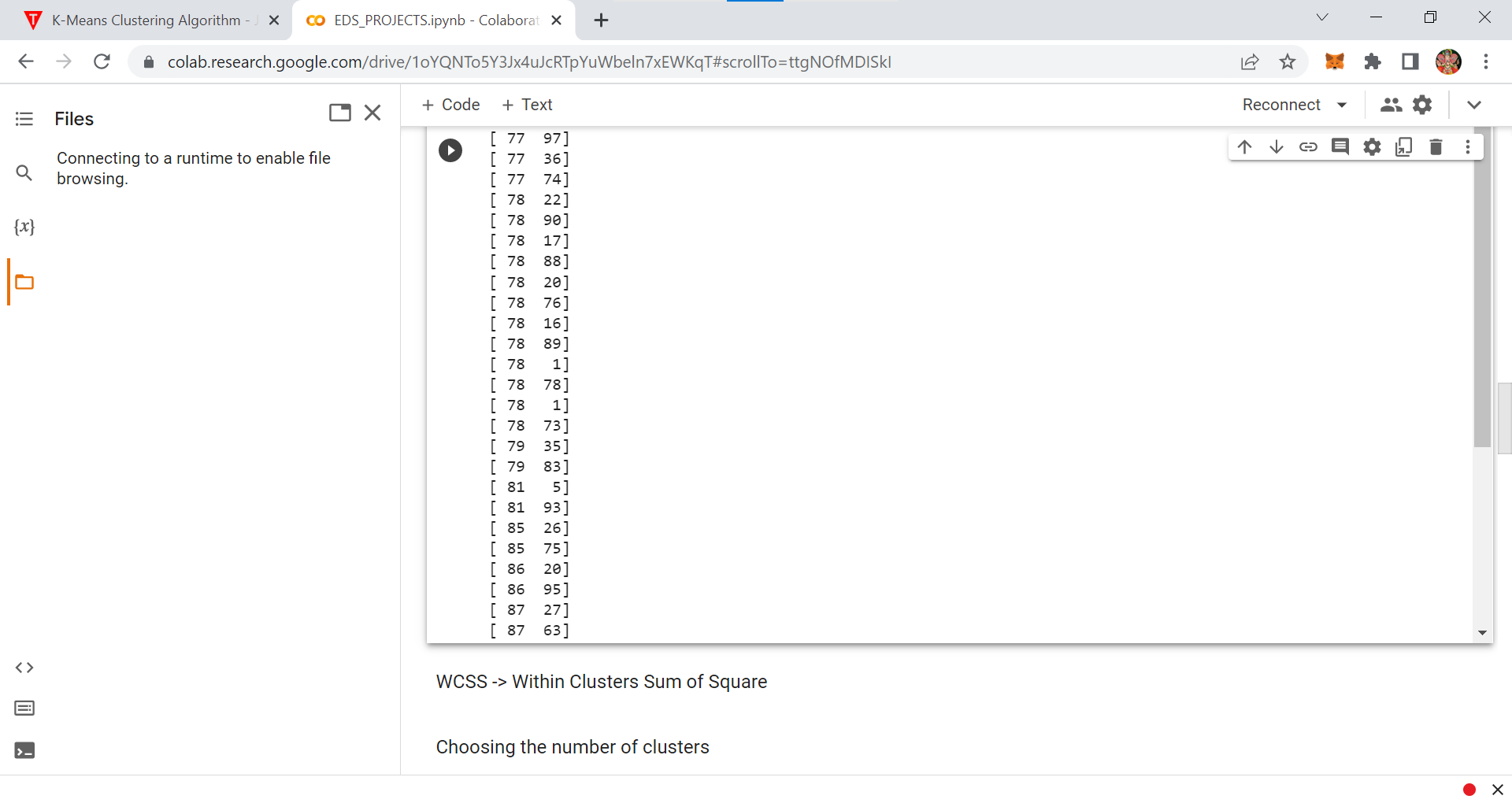
4.python

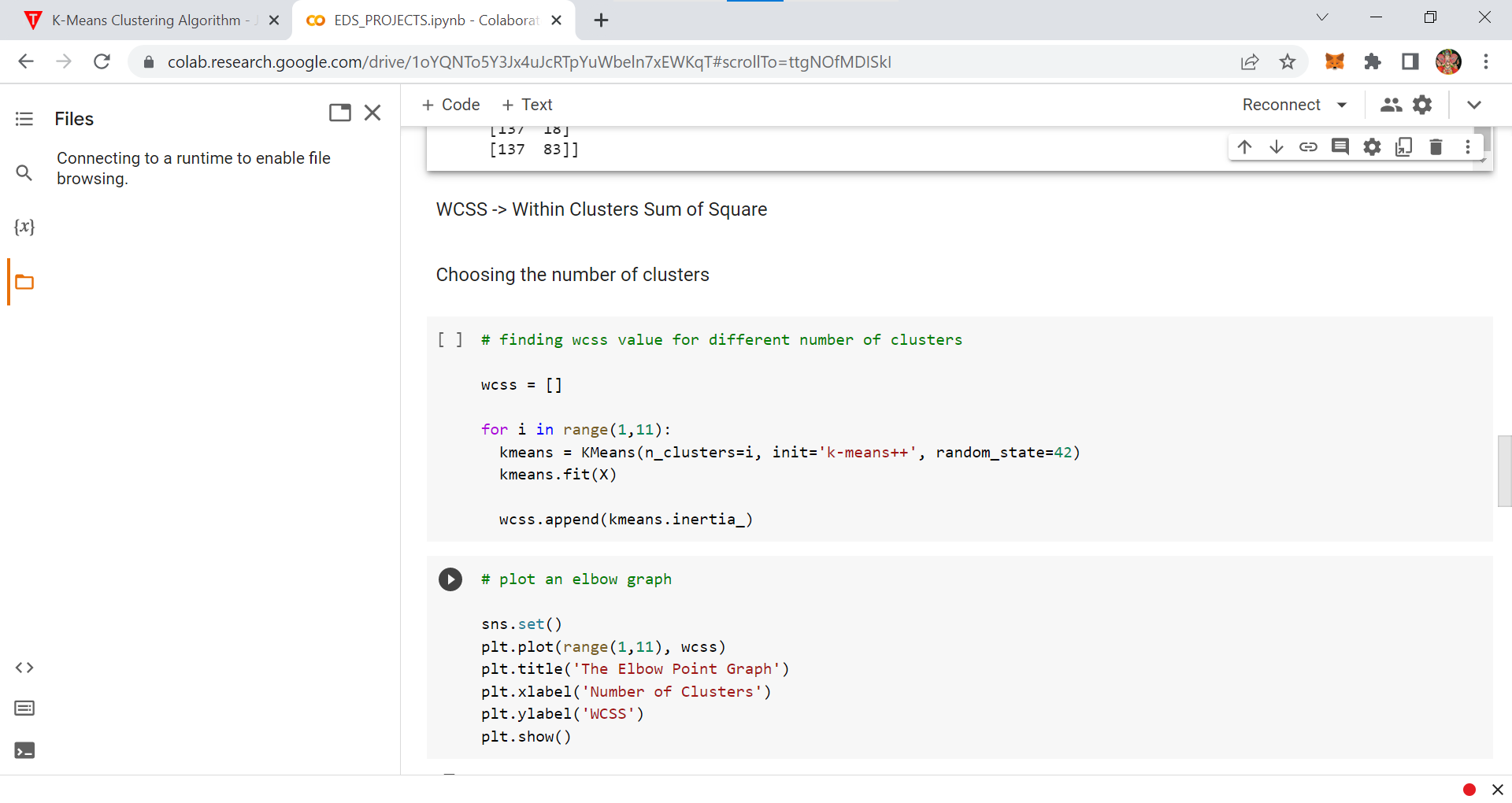
****

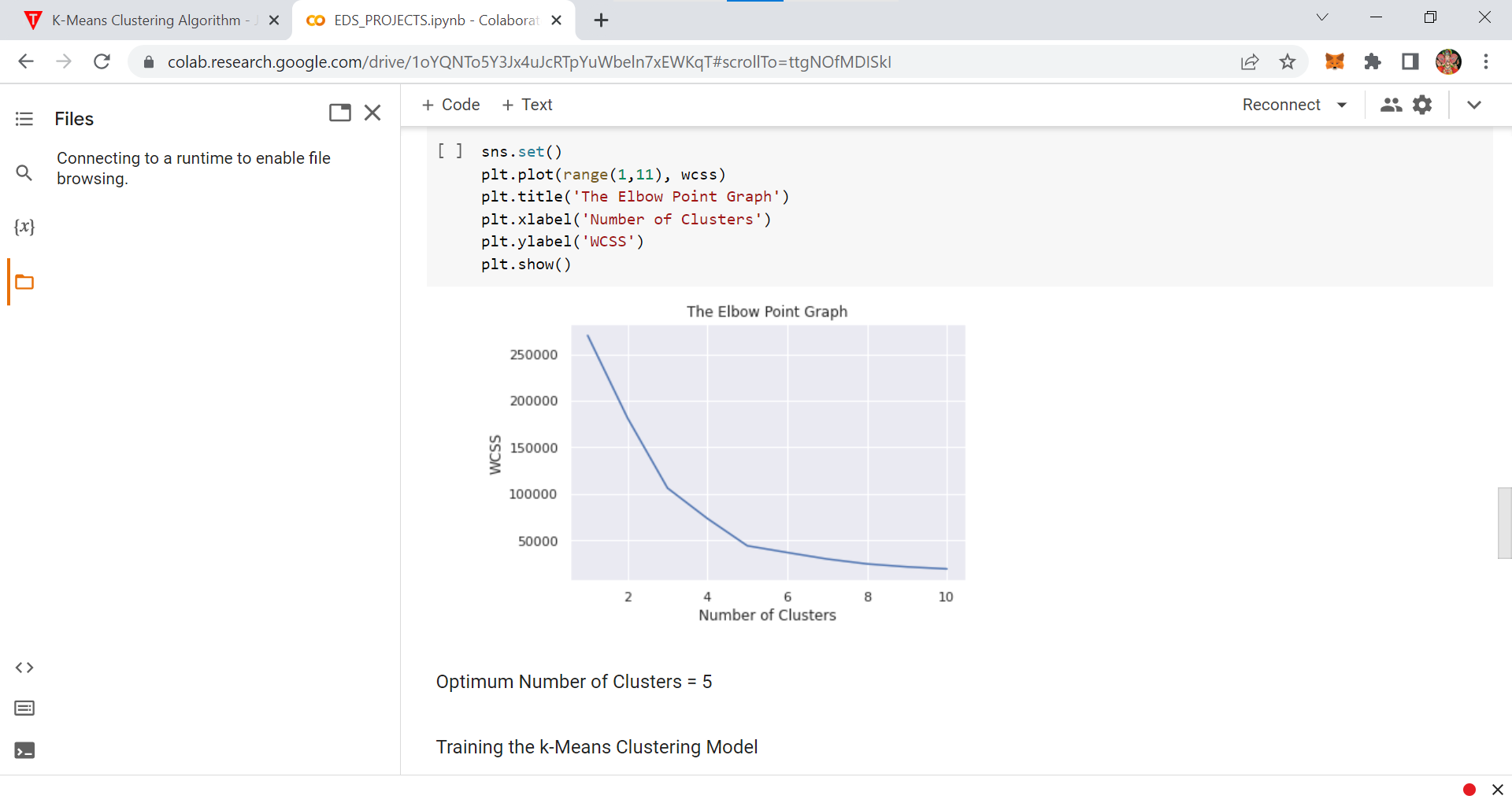
****

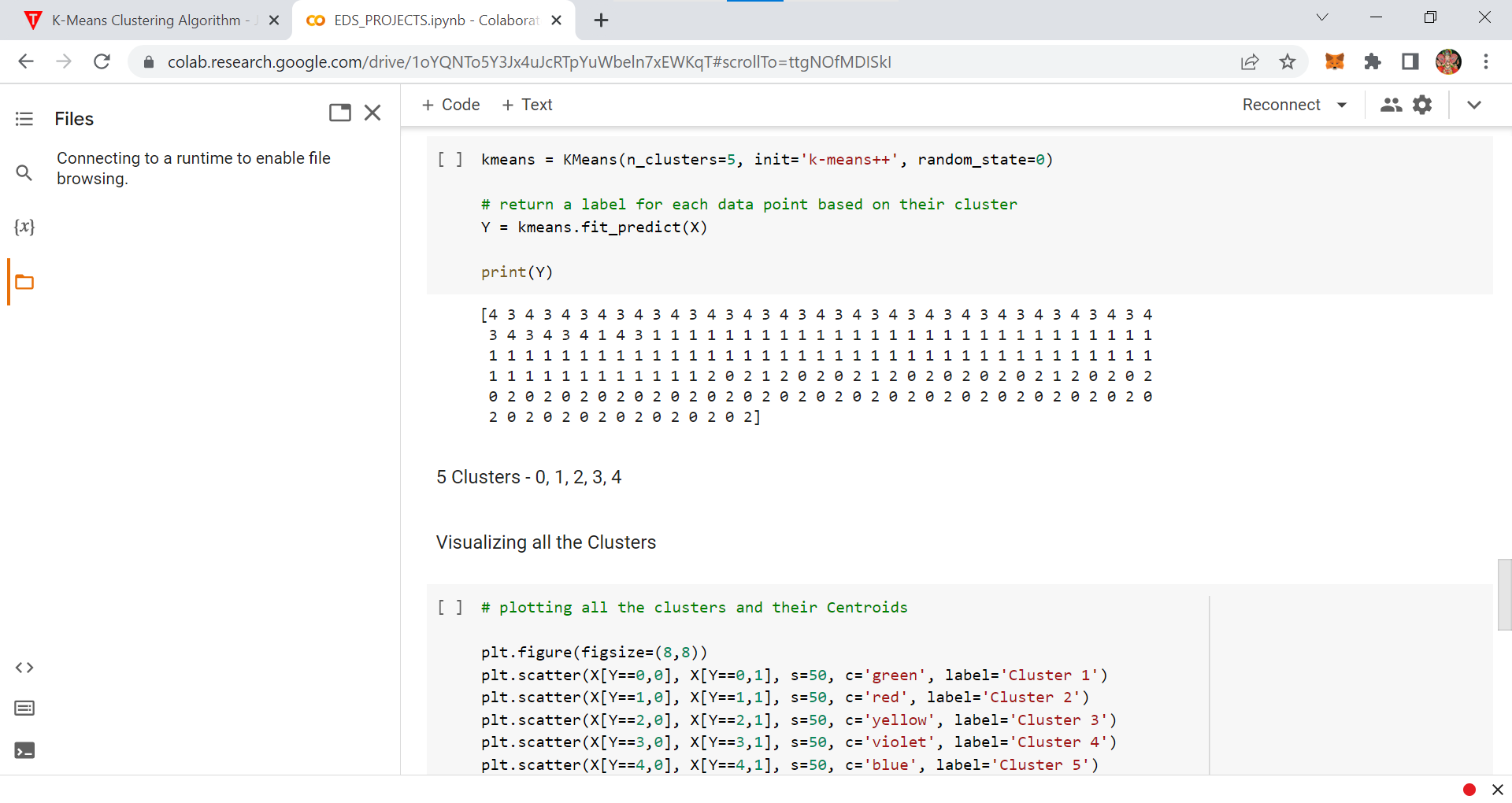
****

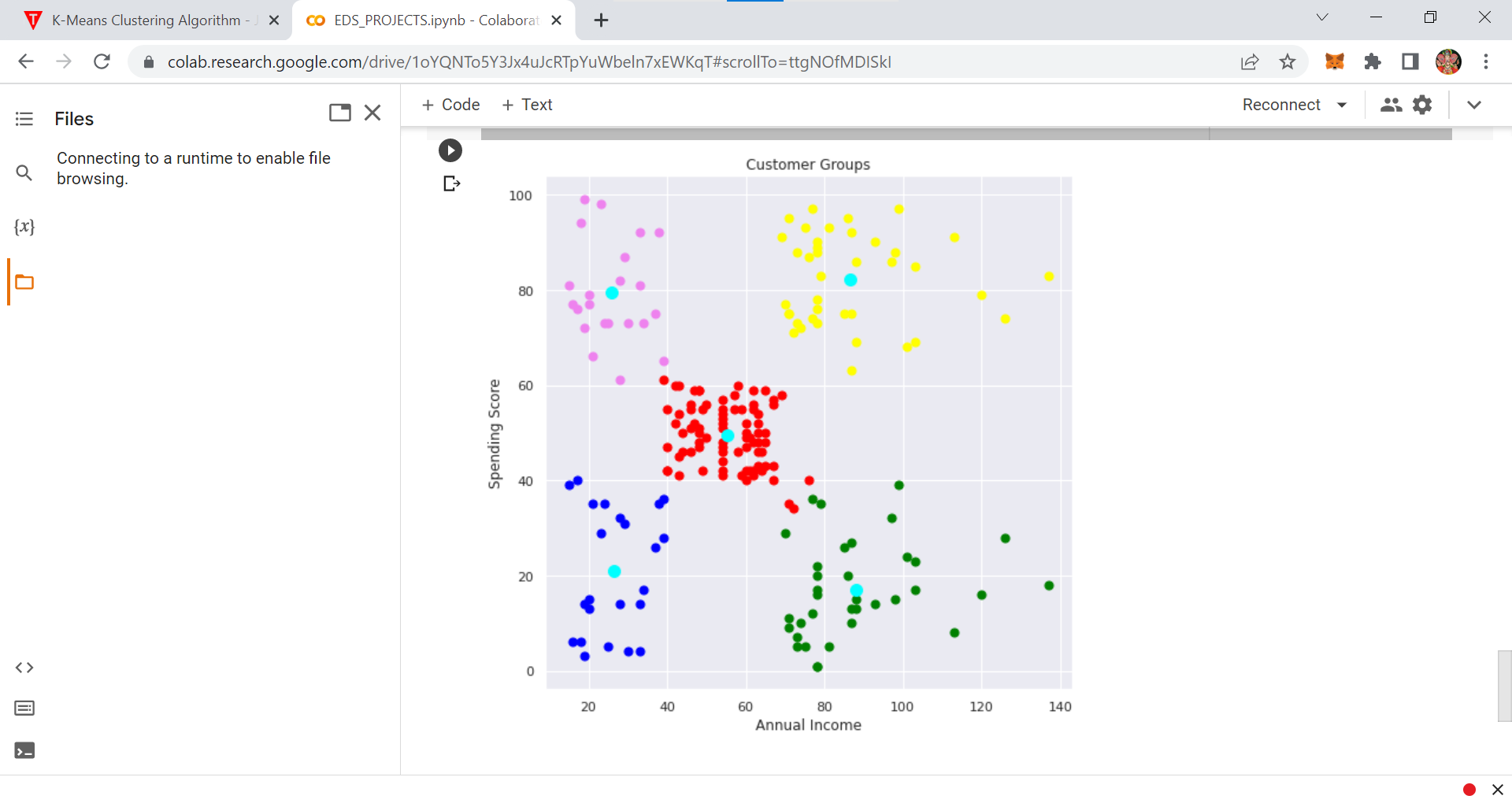
****

****

****

****

****

****

**Project Conclusion: -**

From this project, we gained the knowledge of one of the most popular clustring algorithms and usually the first thing we apply when solving clustring tasks to get an ideal data structure and data set.the goal of k means is to group data points into distinct non overlapping subgroups.

**10**