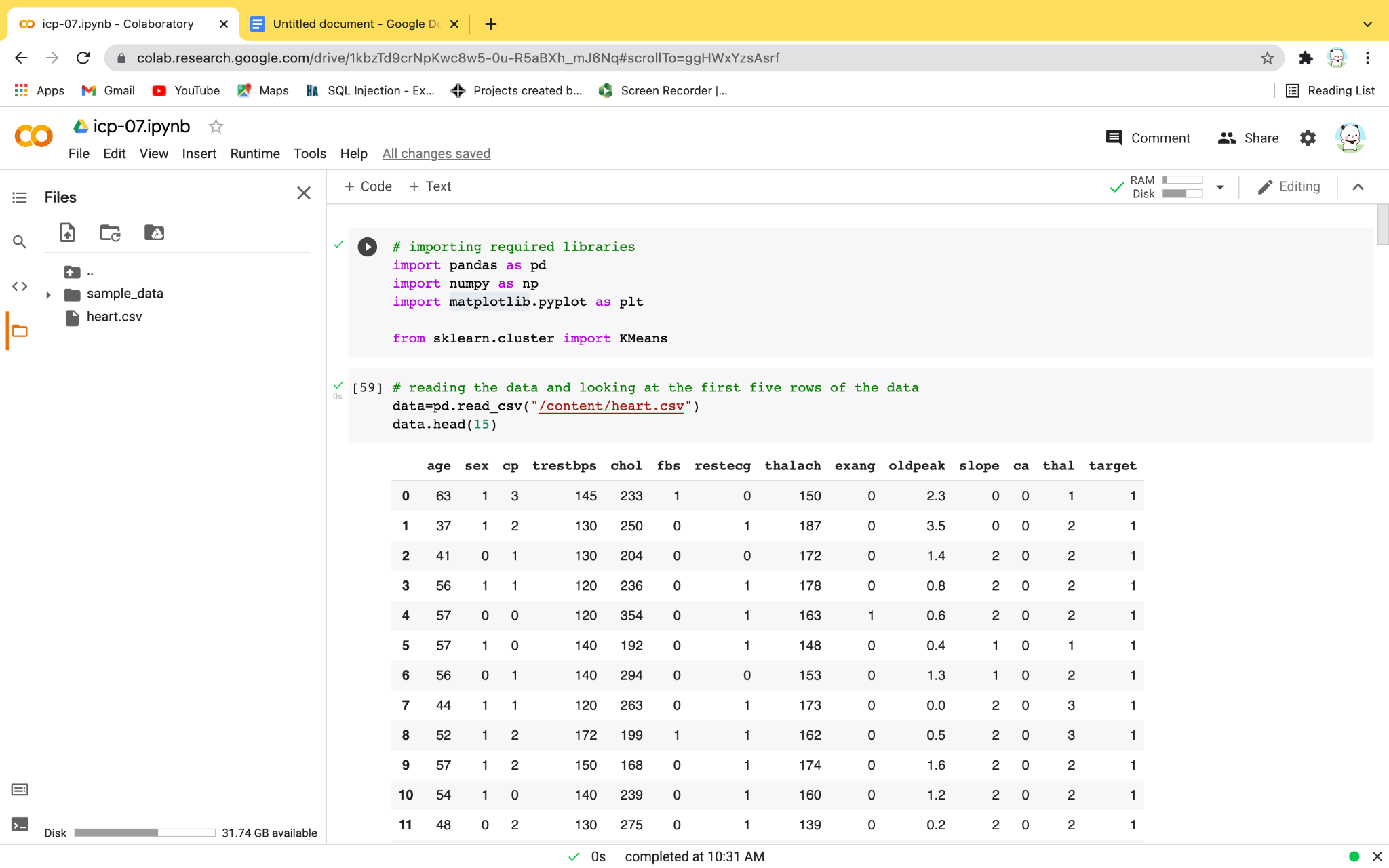
**Wiki Report**

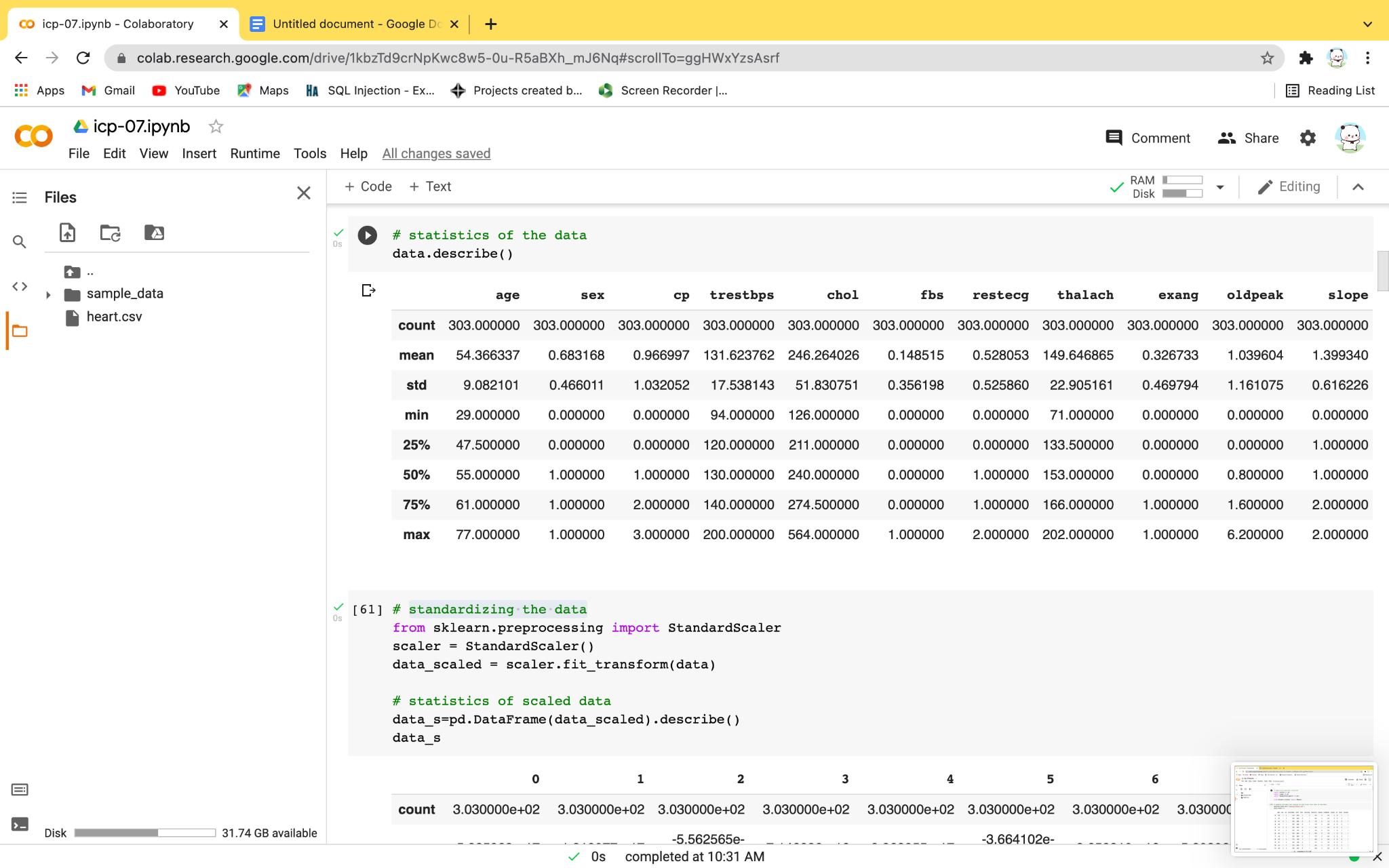
**A)**In the icp i have learned how to perform k-means clustering with different k values on a given dataset**.**

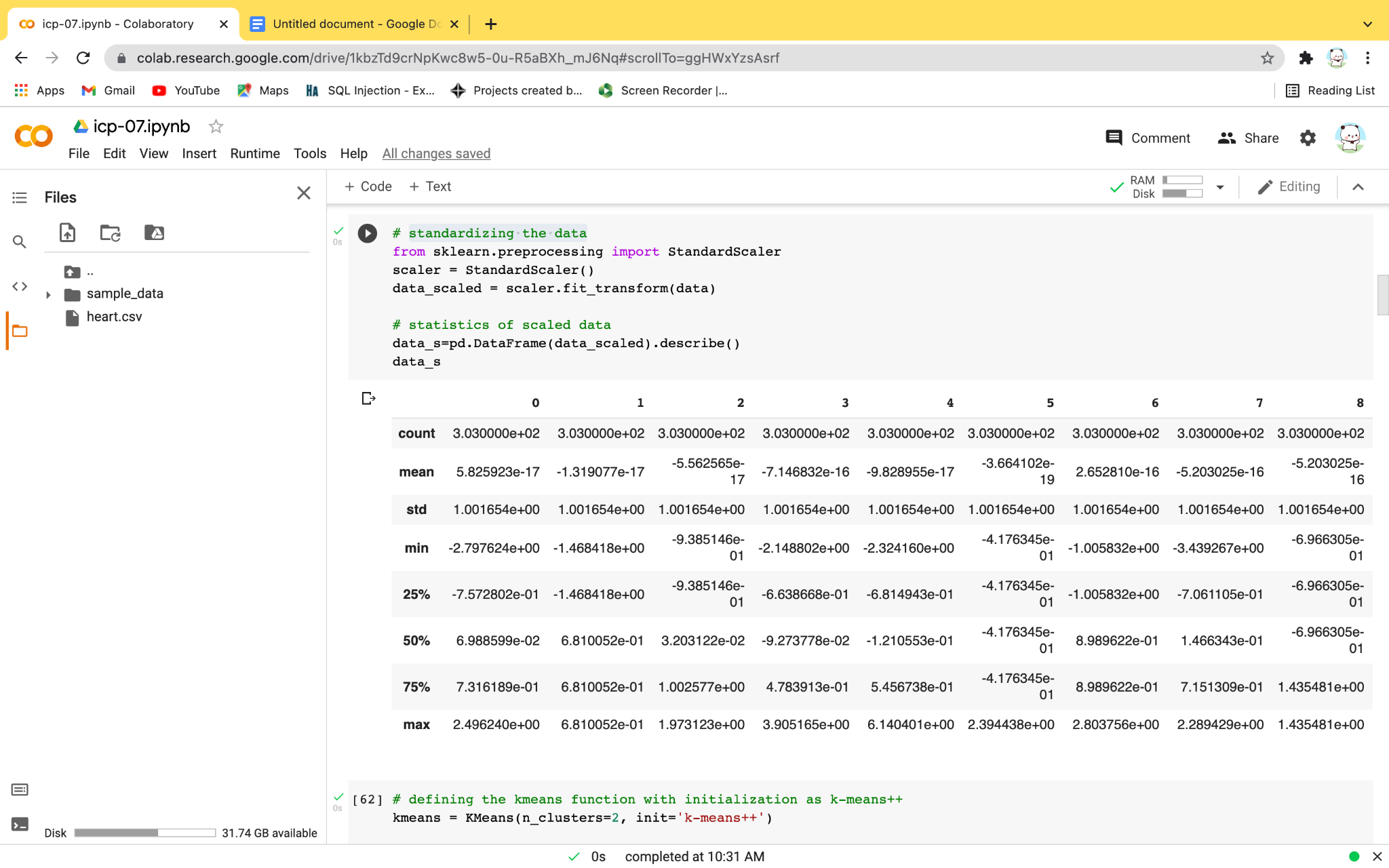
**B)**imported pandas,numpy,matplotlib,k-means libraries to perform data analysis, mathematical, plotting graphs and performed clustering using those libraries. And read the csv file which refers to the presence of heart disease in the patient. Using describe function printed the statistical values because k-means is sensitive to outliers ie.. there can be some errors that can potentially harm the data so to take care of that i have used describe function. Then standardizing the data using standard scale function to scale data such that the mean becomes 0 ie..it scales down to number in a reasonable way .Then defined the k-means function and given the initial k value then fitting the k-means algorithm on scaled data.Then found the inertia on a fitted data.Then using the empty list stored the multiple k values using multiple k-means algorithms then converting the results into data-frame and plotted them on a graph. So, an output we get is an elbow graph where we can predict the correct k value ie..ideal value of cluster, how many clusters can be done. As per the icp i have performed k-means algorithm on a new dataset and on different k values. By observing the elbow curve the five values that I have used are 4,5,6,7,8. When I used k=4 ie.. it had to be categorized in 4 groups. In each group it grouped all similar values. In this data using clustering we will get to know or predict the chances of getting heart diseases at what age or with their cholesterol levels and all. So the k-means algorithm groups the similar data into a group. By that we can predict that certain age group people or people with certain cholesterol levels are very much likely to get heart diseases. By observing all the 5 clusters i feel cluster 5 fits better because its values are segregated properly.

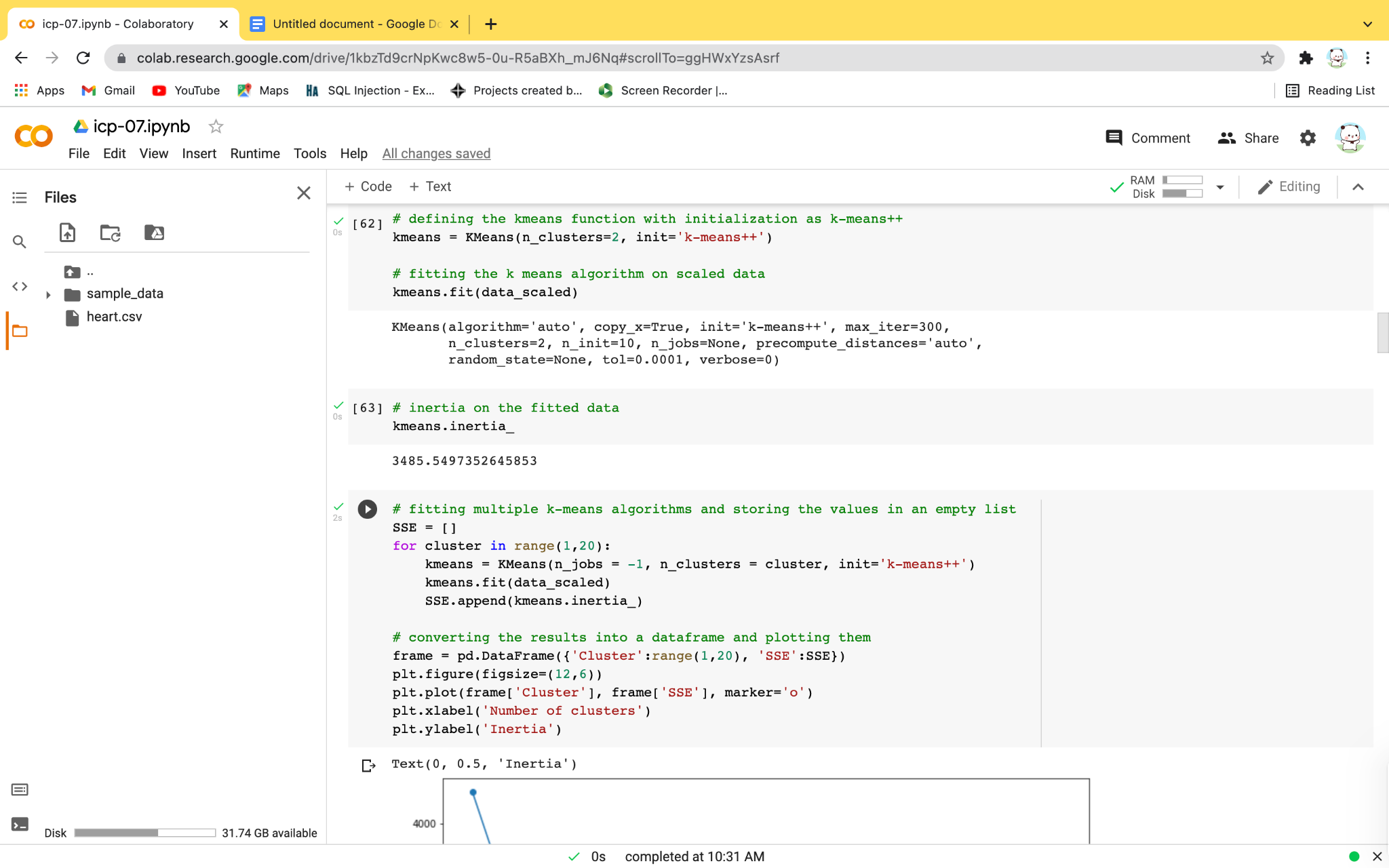
**C)**The challenges that I have faced is that it's hard to predict the ideal k value.

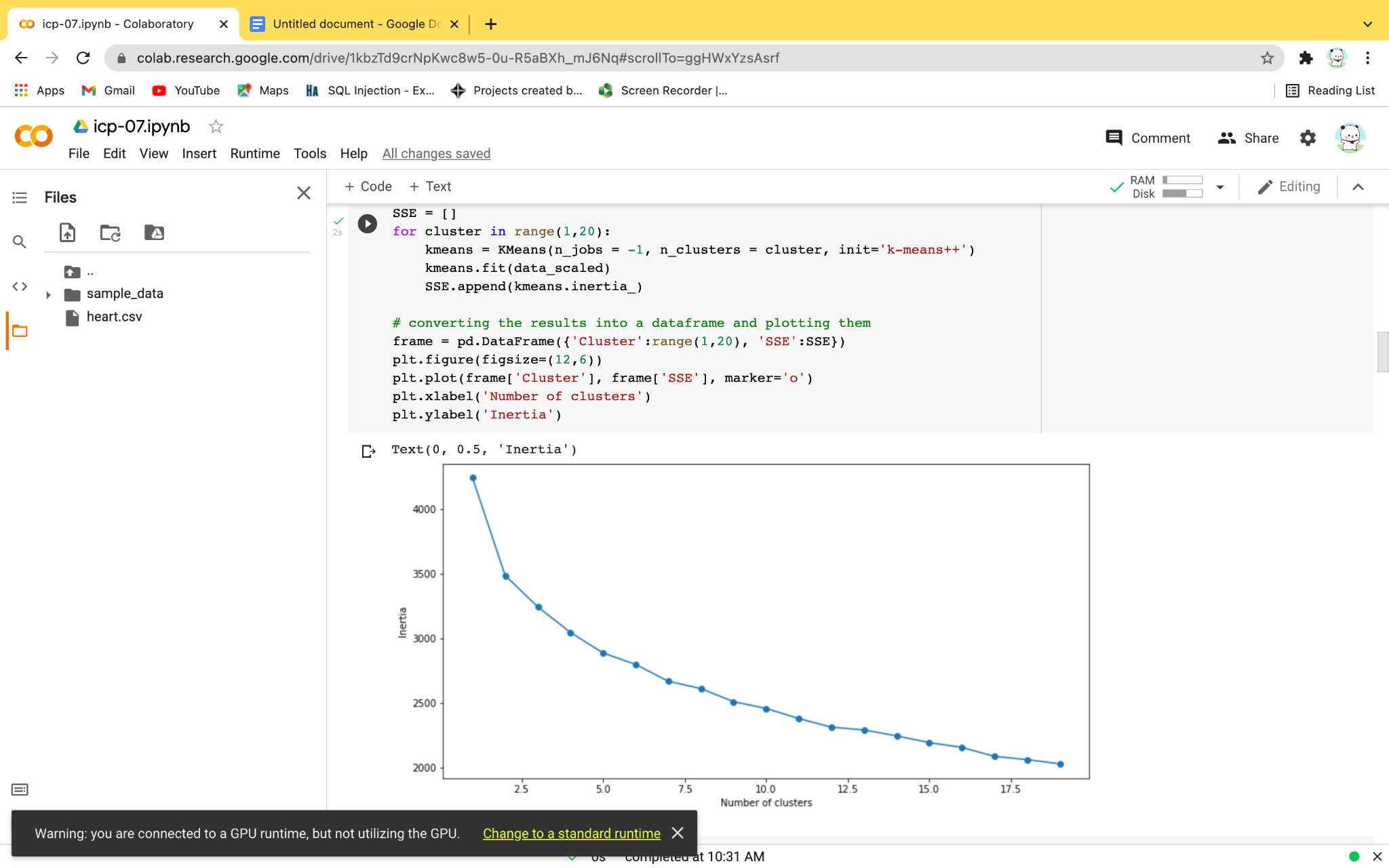
**D)SCREENSHOTS:**

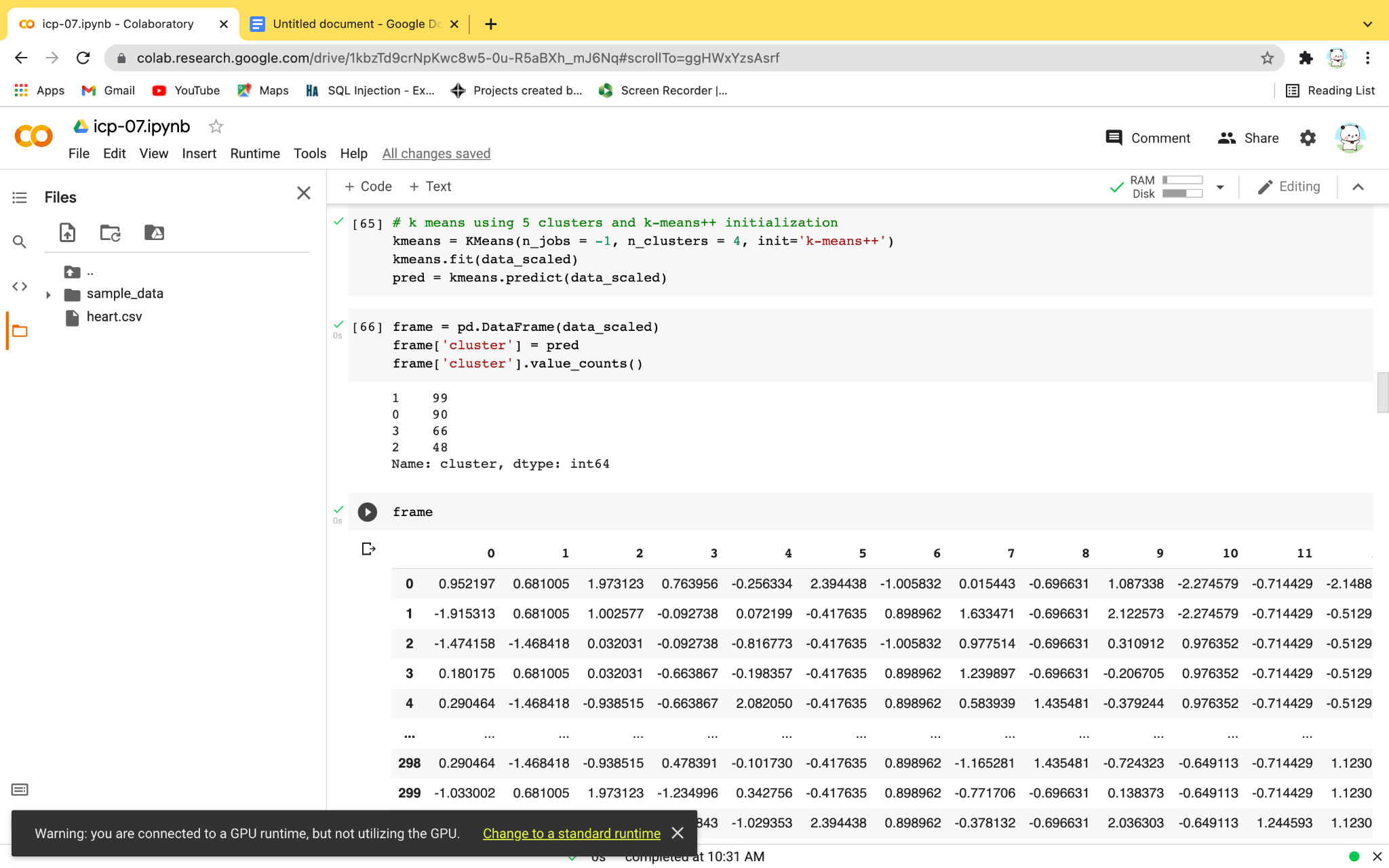
****

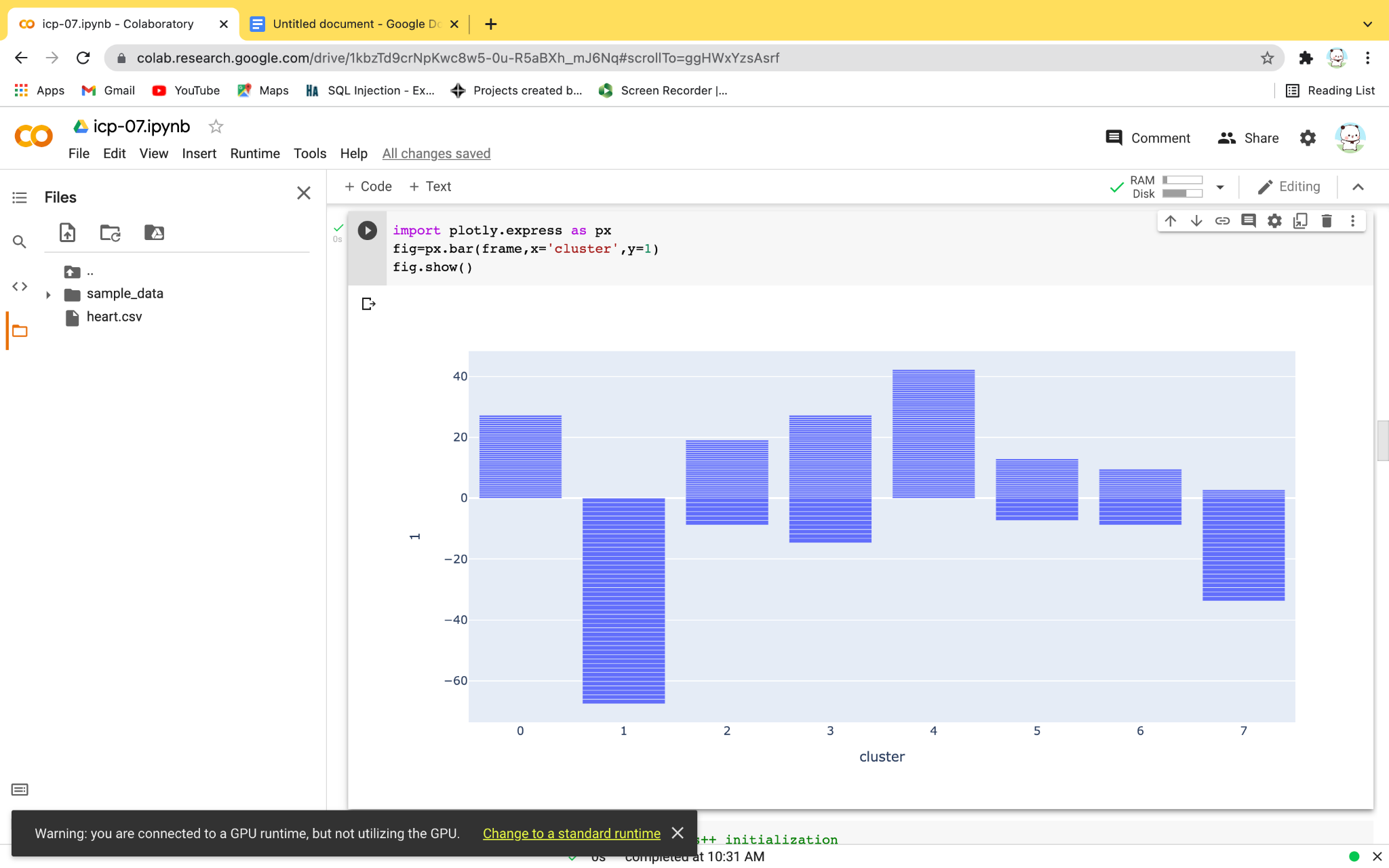
****

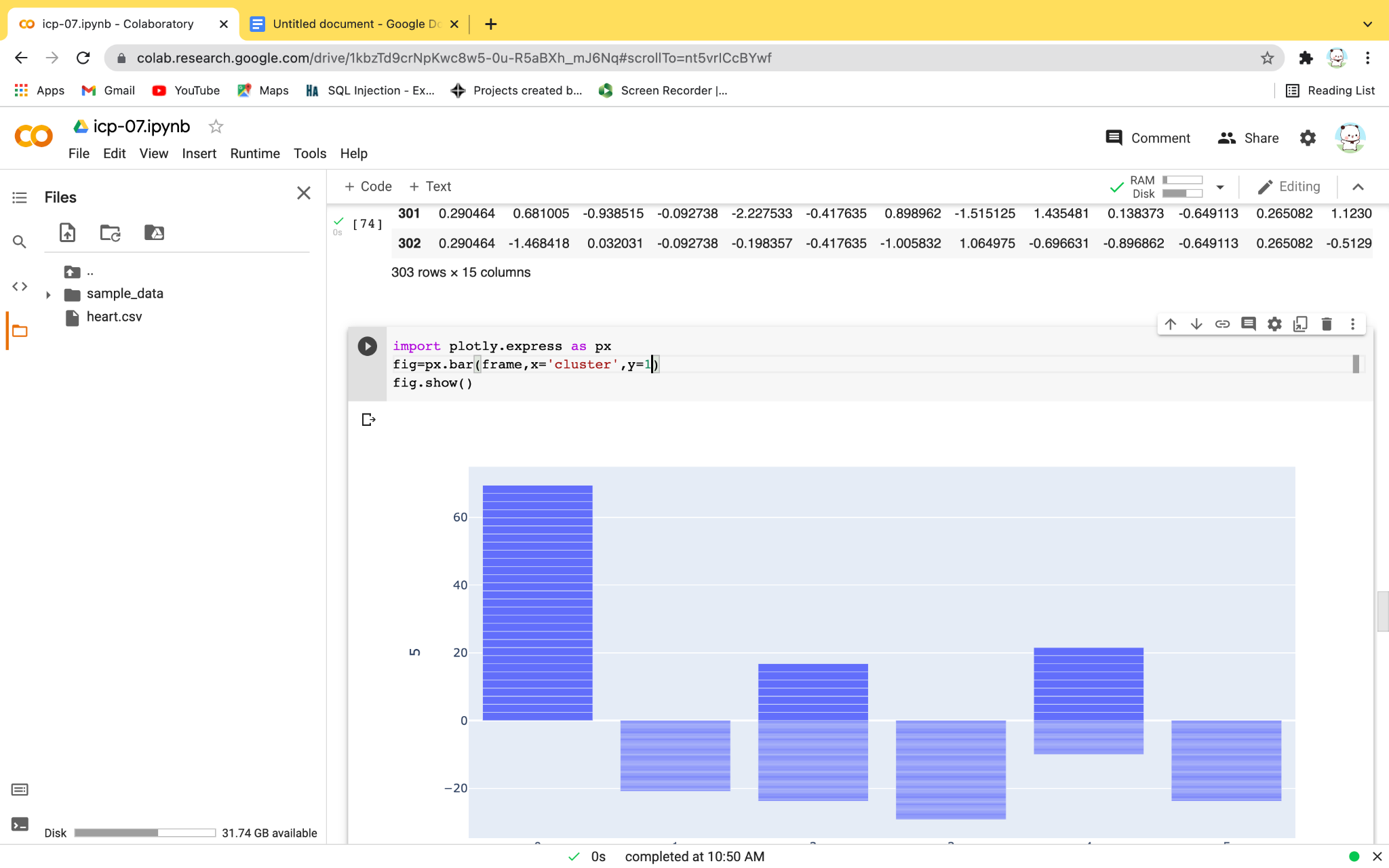
****

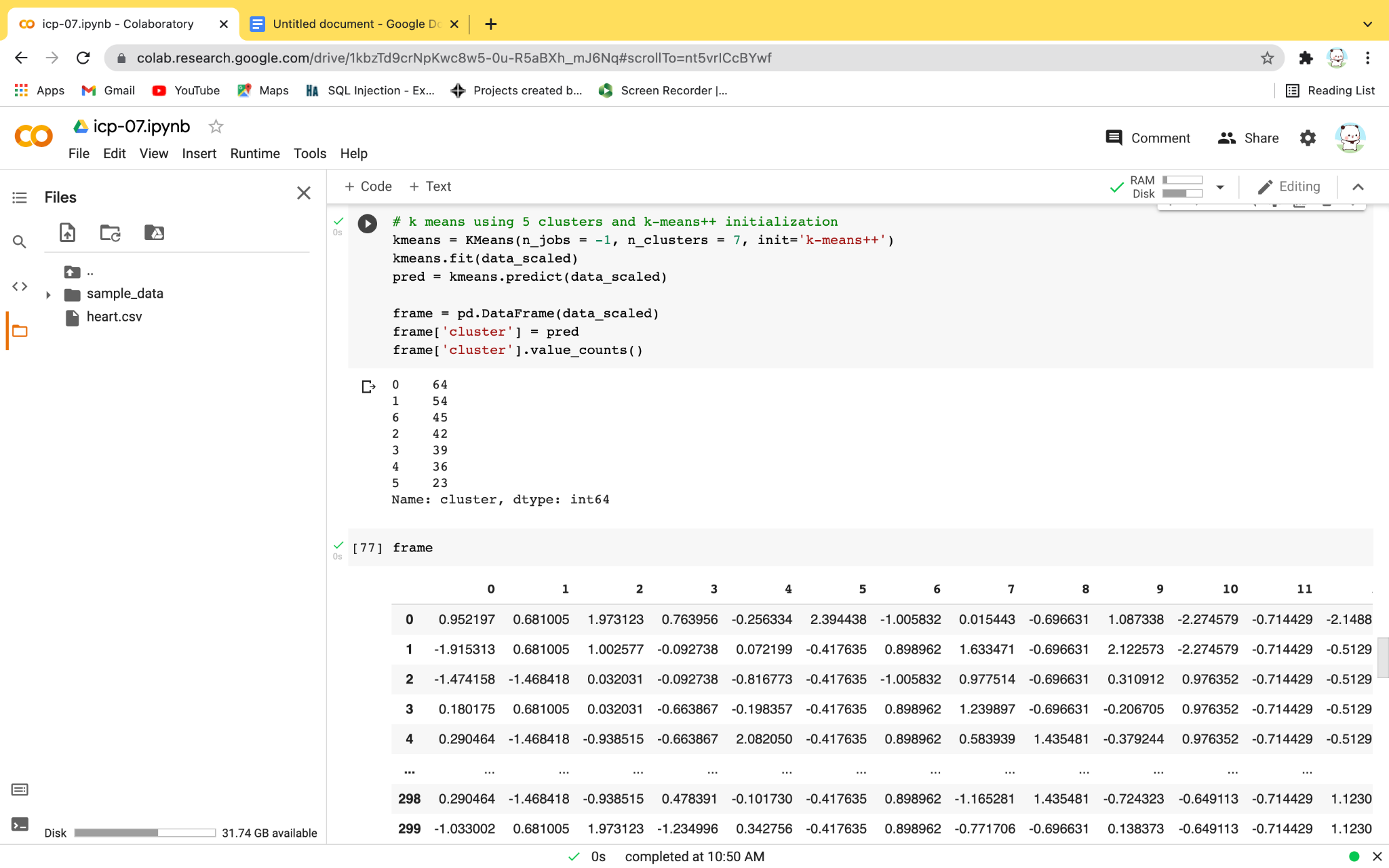
****

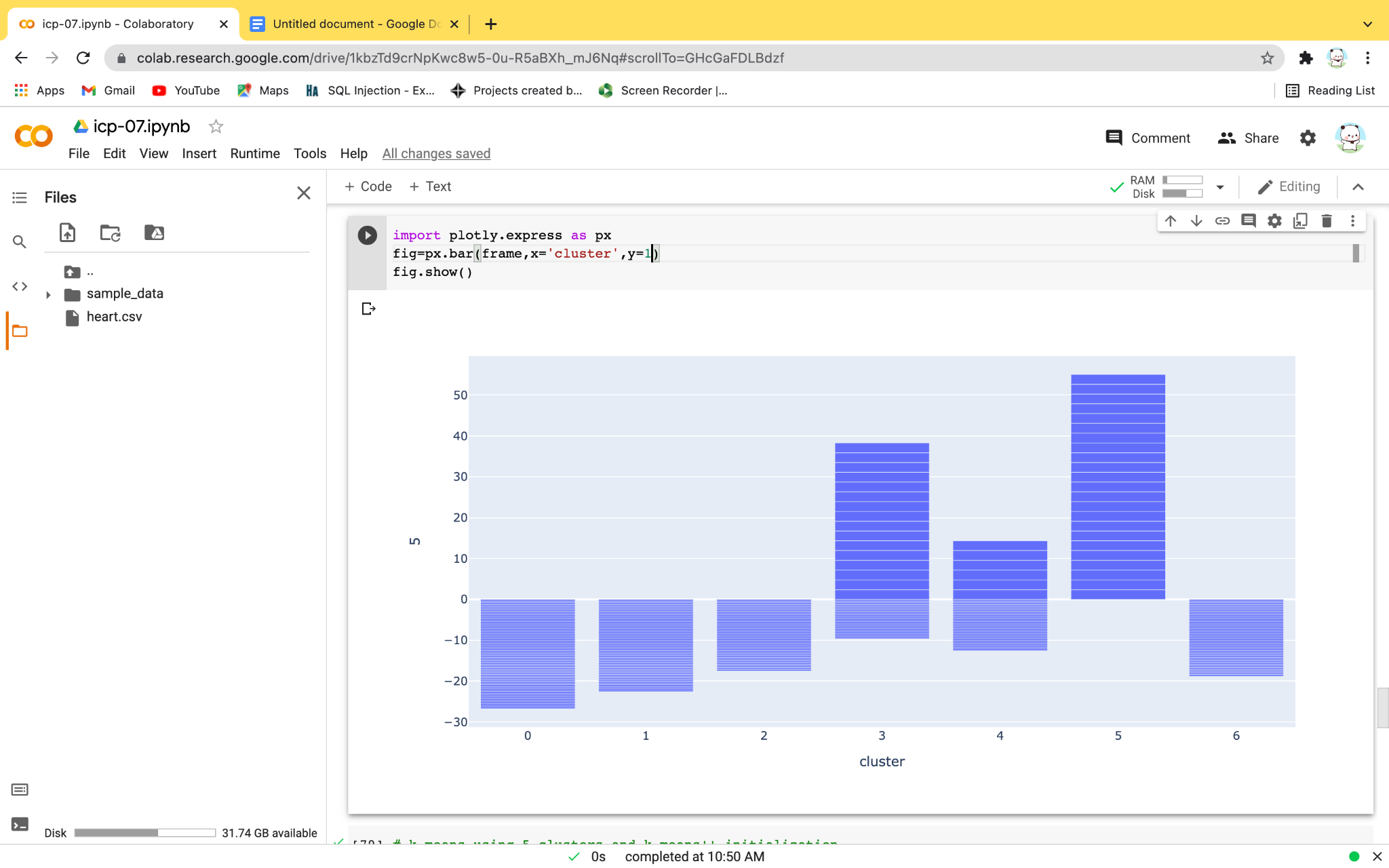
****

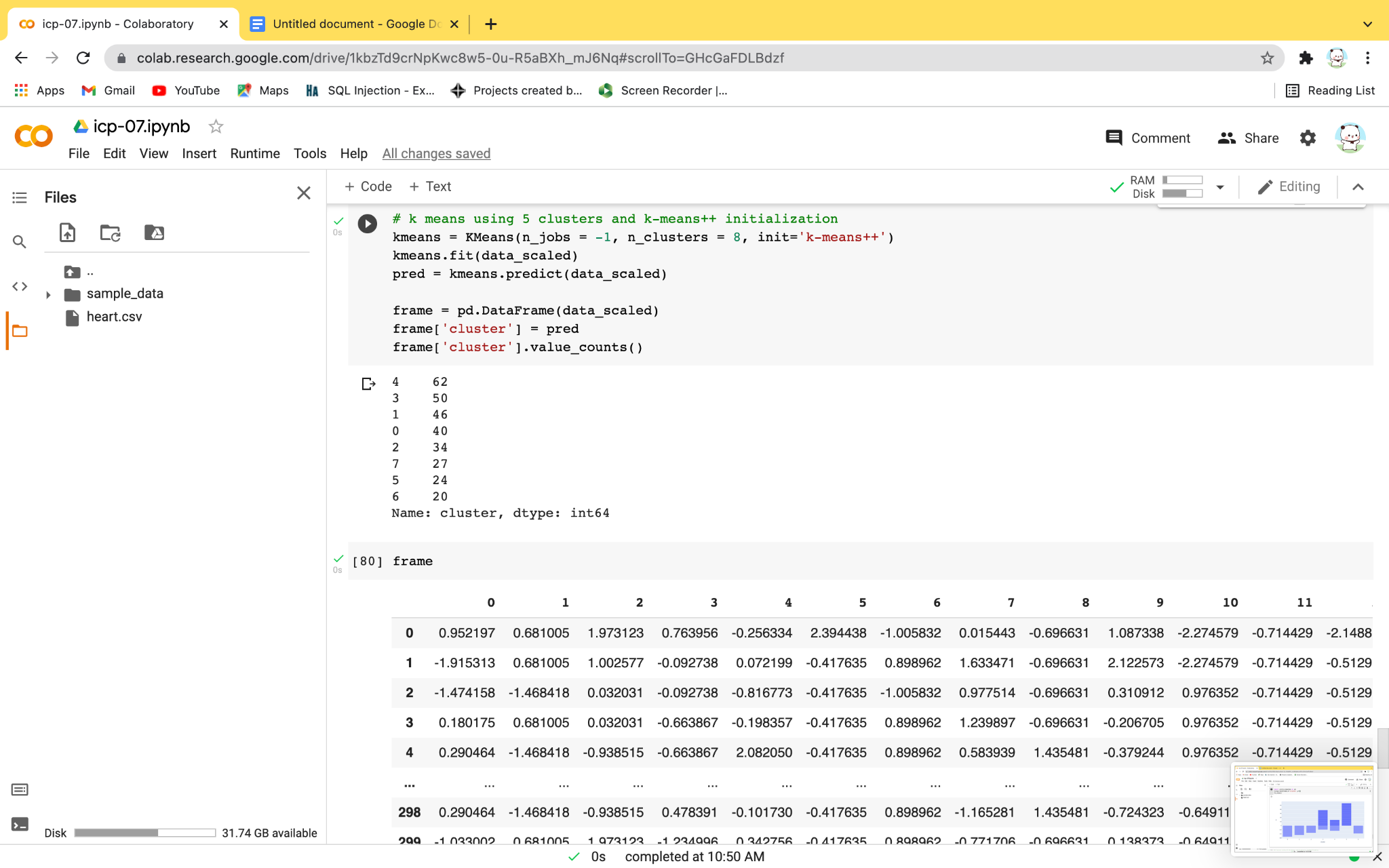
****

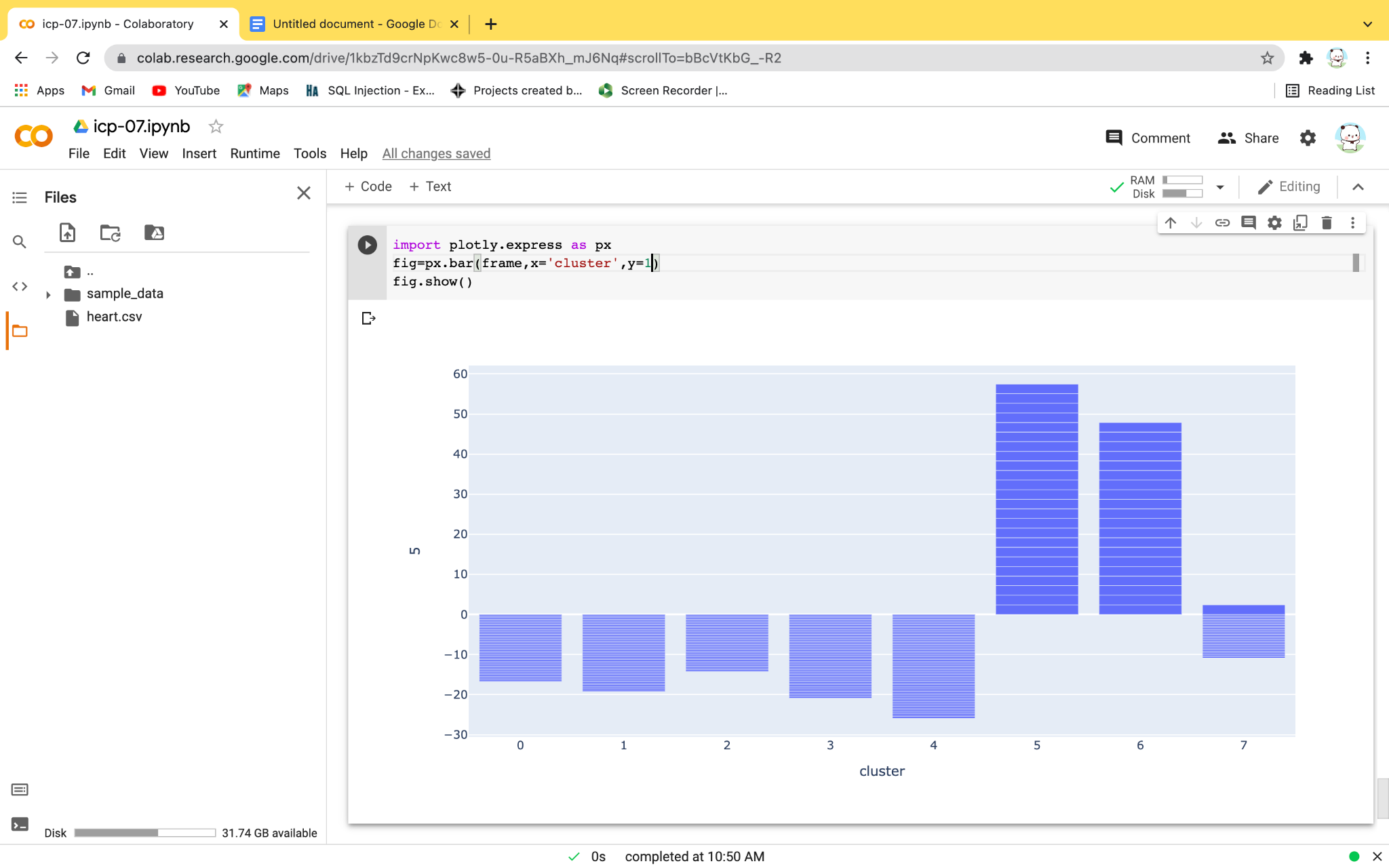
****

****

****

****

****

****

**F)VIDEO LINK:https://github.com/saidurga-kanuganti/icp07/blob/main/Mon%20Oct%2018%202021%2011\_49\_39.webm**

**G)**In general icp is about performing k-means algorithm using different k values.