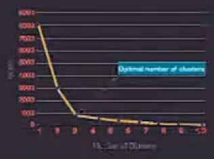
There are no dissimilar data points on either side of the line, which means our model is formed

8



This method uses the concept of WCSS value. Within Cluster Sum of Squares (WCSS)

6 Finding best K Value

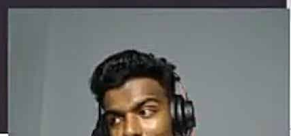
Elbow Method

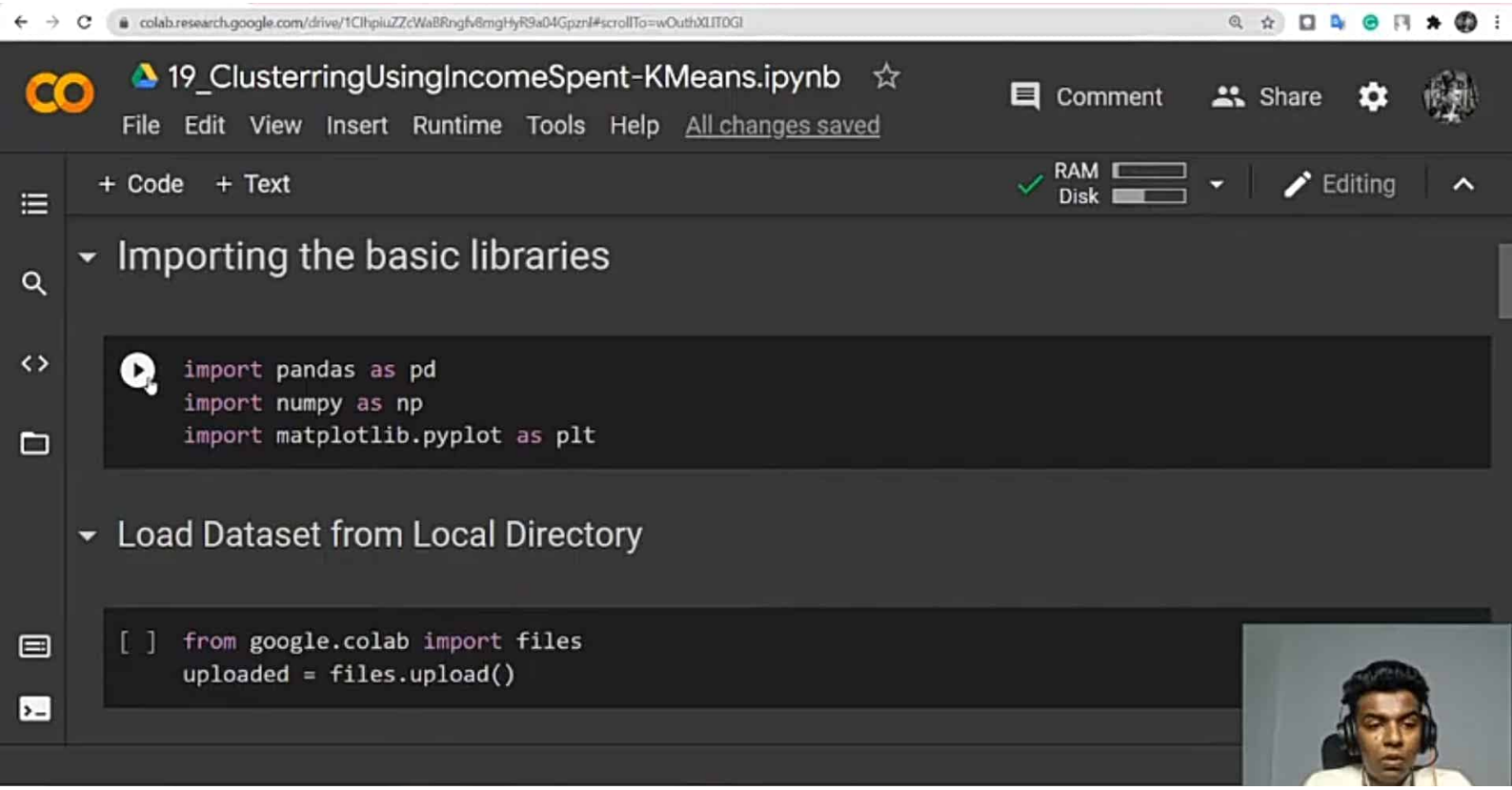
It is calculated by measuring the distance between each data point and its centroid, squaring this distance, and summing these squares across one cluster.

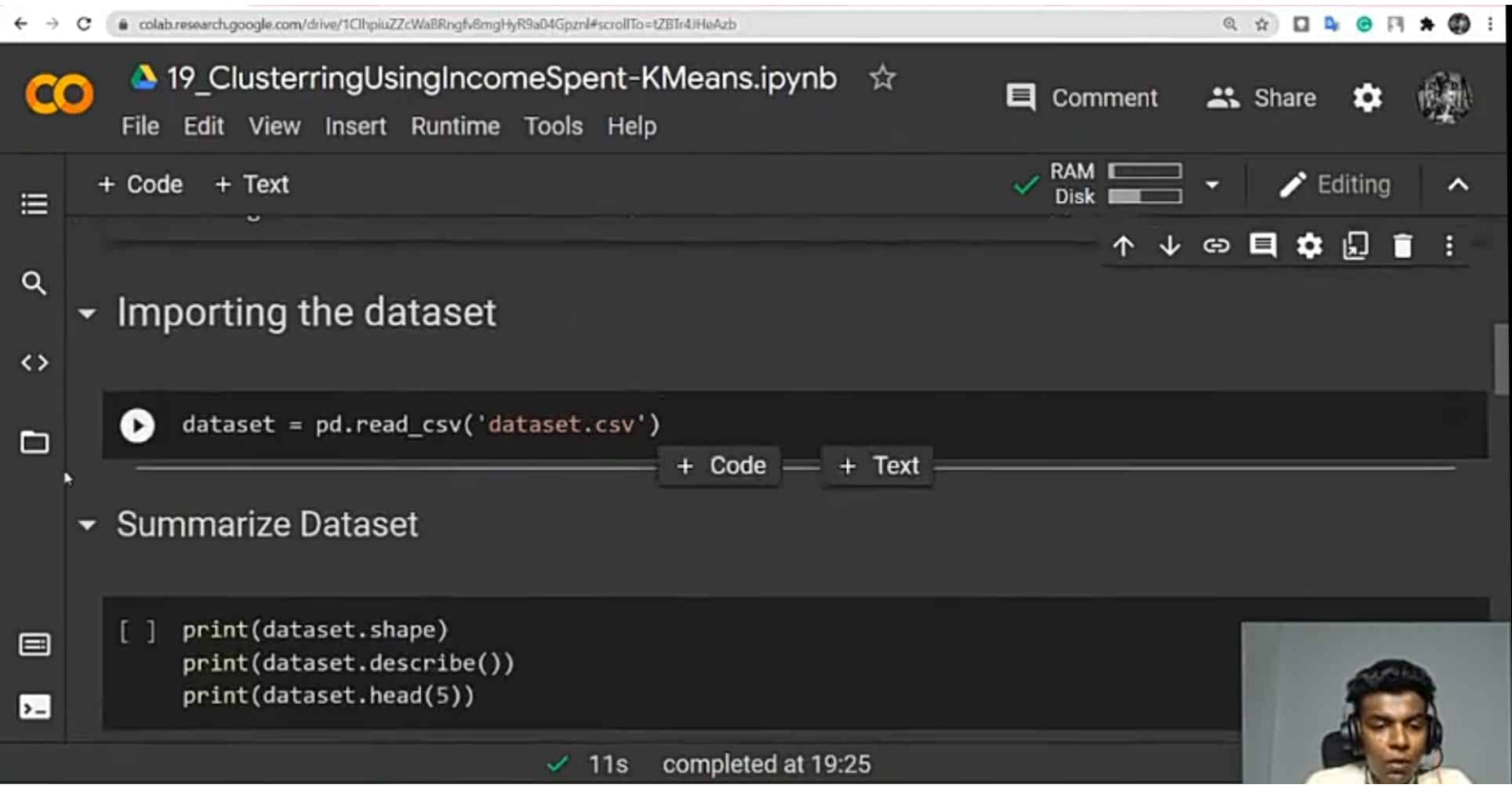
Inertia

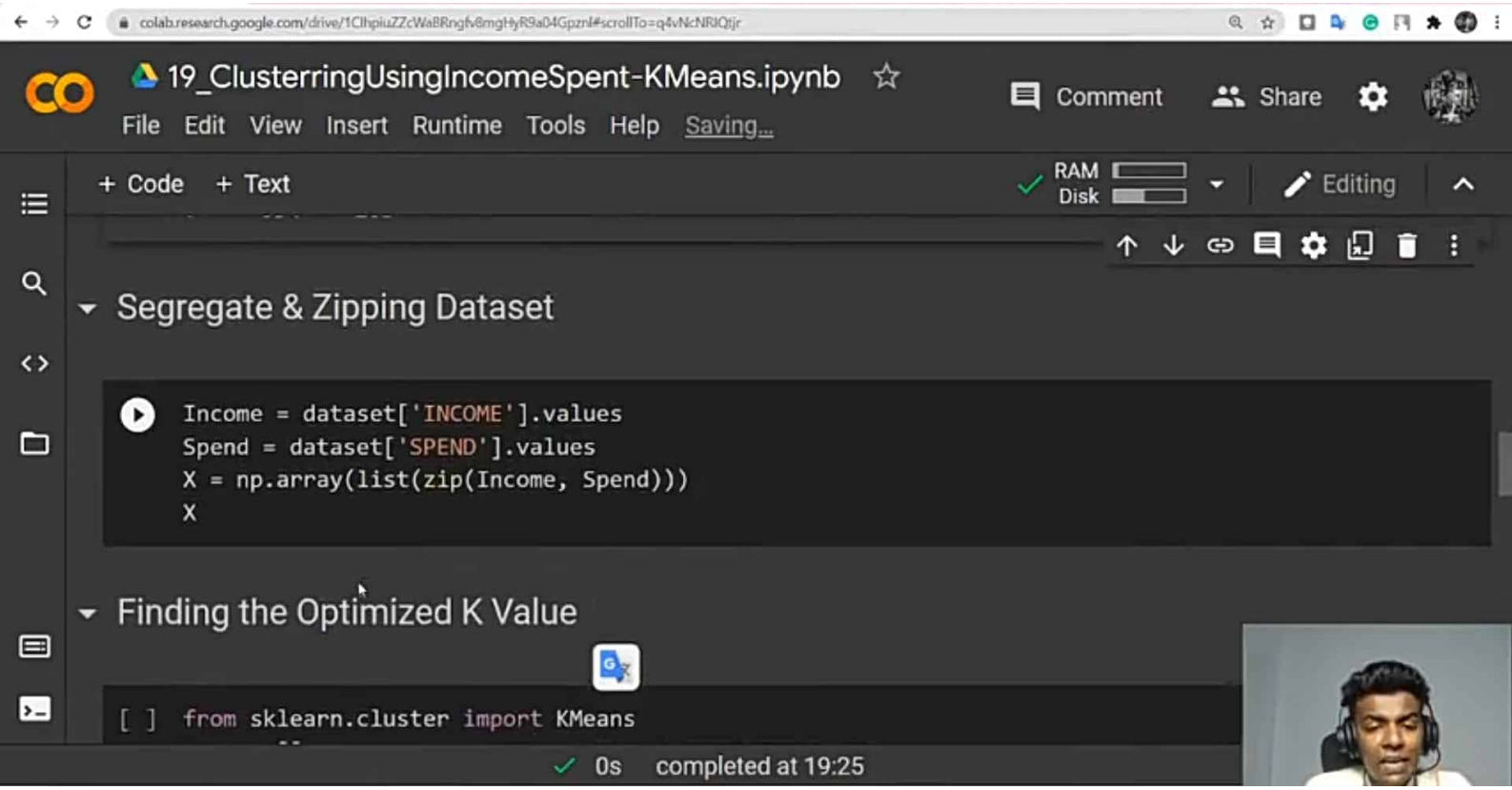
A good model is one with low inertia AND a low number of clusters (K)

7 Fitting Model to Optimized K-Value




colab.research.google.com/drive/1ClhpiuZZcWaBRngfv8mgHyR9a04GpznI#scrollTo=wOuthXLIT0GI
19_ClusterringUsingIncomeSpent-KMeans.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Editing
Importing the basic libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
Load Dataset from Local Directory
[ ] from google.colab import files
uploaded = files.upload()


colab.research.google.com/drive/1ClhpiuZZcWaBRngfv8mgHyR9a04Gpznl#scrollTo=tZBTr4JHeAzb
19_ClusterringUsingIncomeSpent-KMeans.ipynb
File Edit View Insert Runtime Tools Help
Comment
Share
+ Code + Text
RAM
Disk
Editing
Importing the dataset
dataset = pd.read_csv('dataset.csv')
+ Code + Text
Summarize Dataset
[ ] print(dataset.shape)
print(dataset.describe())
print(dataset.head(5))
11s completed at 19:25


colab.research.google.com/drive/1ClhpiuZZcWaBRngfv8mgHyR9a04GpznI#scrollTo=q4vNcNRiQtjr
19_ClusterringUsingIncomeSpent-KMeans.ipynb
File Edit View Insert Runtime Tools Help Saving...
Comment
Share
+ Code + Text
RAM
Disk
Editing
Segregate & Zipping Dataset
Income = dataset['INCOME'].values
Spend = dataset['SPEND'].values
X = np.array(list(zip(Income, Spend)))
X
Finding the Optimized K Value
[ ] from sklearn.cluster import KMeans
0s completed at 19:25

colab.research.google.com/drive/1ClhpiuZZcWaBRngfv8mgHyR9a04GpznI#scrollTo=2PuOMjABSCXw

19_ClusterringUsingIncomeSpent-KMeans.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

RAM Disk

Editing

```
from sklearn.cluster import KMeans
wcss = []
for i in range(1,11):
    km=KMeans(n_clusters=i, random_state=0)
    km.fit(X)
    wcss.append(km.inertia_)
plt.plot(range(1,11),wcss,color="red", marker ="8")
plt.title('Optimal K Value')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.show()
```

1s completed at 19:27

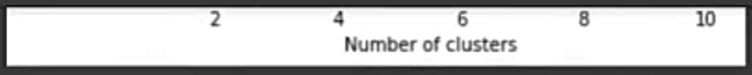
19_ClusterringUsingIncomeSpent-KMeans.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

RAM Disk Editing



Fitting the k-means to the dataset with k=4

```
model=KMeans(n_clusters=4, random_state=0)
y_means = model.fit_predict(X)
```

+ Code + Text

Visualizing the clusters for k=4

Cluster 1: Customers with medium income and low spend

Cluster 2: Customers with high income and medium to high spend

1s completed at 19:27

colab.research.google.com/drive/1ClhpiuZZcWaBRngfv8mgHyR9a04Gpznl#scrollTo=0K7mijvHUW6Z

19_ClusterringUsingIncomeSpent-KMeans.ipynb

File Edit View Insert Runtime Tools Help

Comment Share

+ Code + Text

RAM Disk

Editing

Cluster 2: Customers with high income and medium to high spend

Cluster 3: Customers with low income

Cluster 4: Customers with medium income but high spend

```
[ ]  plt.scatter(X[y_means==0,0],X[y_means==0,1],s=50, c='brown',label='1')
     plt.scatter(X[y_means==1,0],X[y_means==1,1],s=50, c='blue',label='2')
     plt.scatter(X[y_means==2,0],X[y_means==2,1],s=50, c='green',label='3')
     plt.scatter(X[y_means==3,0],X[y_means==3,1],s=50, c='cyan',label='4')
     plt.scatter(model.cluster_centers_[:,0], model.cluster_centers_[:,1],s=100,marker='s', c='red', l
     plt.title('Income Spent Analysis')
     plt.xlabel('Income')
     plt.ylabel('Spent')
     plt.legend()
     plt.show()
```

0s completed at 19:31

19_ClusterringUsingIncomeSpent-KMeans.ipynb

File Edit View Insert Runtime Tools Help All changes saved

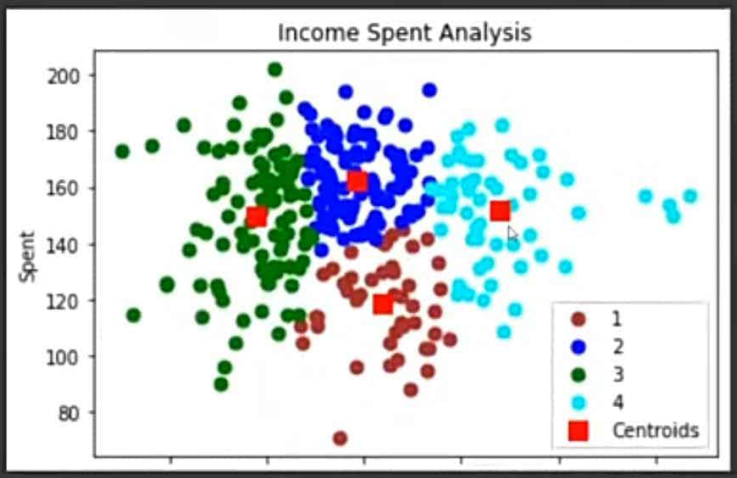
Comment Share

+ Code + Text

RAM Disk

Editing

```
plt.legend()
plt.show()
```



0s completed at 19:32

Type here to search

29°C Haze