Sheth L.U.J. College Of Arts & Sir M.V. College of Science & Commerce Department Of Science Jayesh mali T084 Pract\_ 8 Data Science

#Jayesh Mali T084 import pandas as pd import matplotlib.pyplot as plt import numpy as np from sklearn.cluster import KMeans import matplotlib.cm as cm from kneed import KneeLocator

pip install kneed

Collecting kneed

Downloading kneed-0.8.5-py3-none-any.whl (10 kB)

Requirement already satisfied: numpy>=1.14.2 in /usr/local/lib/python3.10/dist-packages (from kneed) (1.25.2) Requirement already satisfied: scipy>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from kneed) (1.11.4)

Installing collected packages: kneed

Successfully installed kneed-0.8.5

#Jayesh Mali T084 file\_location = "/content/CGPA.csv" # Download the CGPA.csv file and replace this String variable with the file location of the download df = pd.read\_csv(file\_location) arr = np.array(df['CGPA']) arr

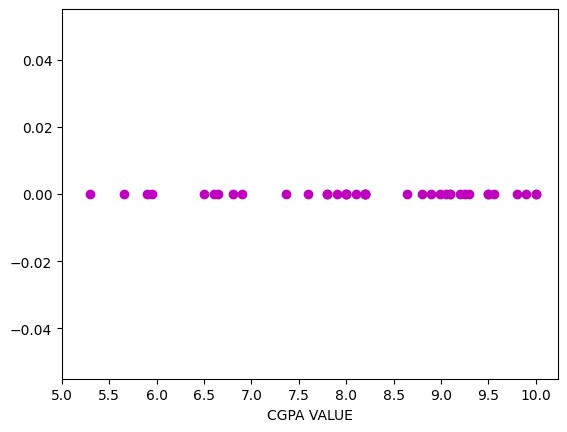
array([ 9.1 , 9.2 , 8.2 , 8. , 8.2 , 5.9 , 9.05, 8.9 , 8. , 6.9 , 8.2 , 8.1 , 10. , 9.5 , 7.9 , 5.3 , 6.5 , 9.3 , 8.2 , 9.8 , 8.8 , 6.6 , 9.5 , 7.8 , 8. , 9.9 , 10. ,

6.8 , 7.6 , 7.8 , 8.64, 9.56, 8.99, 9.25, 9.1 , 5.65,

5.95, 6.65, 7.36])

#Jayesh Mali T084 y = np.zeros(len(arr)) plt.scatter(arr,y,color = 'm') plt.xticks(np.arange(5, 10.5, .5)) plt.xlabel('CGPA VALUE')

Text(0.5, 0, 'CGPA VALUE')



|  |
| --- |
| #Jayesh Mali T084 k\_range = range(1,11) inertias = [] for k in k\_range:  km = KMeans(n\_clusters = k) km.fit(df[['CGPA']]) inertias.append(km.inertia\_) y = np.zeros(len(inertias)) inertias |

/usr/local/lib/python3.10/dist-packages/sklearn/cluster/\_kmeans.py:870: FutureWarning: The default value of `n\_init` will change fro warnings.warn(

/usr/local/lib/python3.10/dist-packages/sklearn/cluster/\_kmeans.py:870: FutureWarning: The default value of `n\_init` will change fro warnings.warn(

/usr/local/lib/python3.10/dist-packages/sklearn/cluster/\_kmeans.py:870: FutureWarning: The default value of `n\_init` will change fro warnings.warn(

/usr/local/lib/python3.10/dist-packages/sklearn/cluster/\_kmeans.py:870: FutureWarning: The default value of `n\_init` will change fro warnings.warn(

/usr/local/lib/python3.10/dist-packages/sklearn/cluster/\_kmeans.py:870: FutureWarning: The default value of `n\_init` will change fro warnings.warn(

/usr/local/lib/python3.10/dist-packages/sklearn/cluster/\_kmeans.py:870: FutureWarning: The default value of `n\_init` will change fro warnings.warn(

/usr/local/lib/python3.10/dist-packages/sklearn/cluster/\_kmeans.py:870: FutureWarning: The default value of `n\_init` will change fro warnings.warn(

/usr/local/lib/python3.10/dist-packages/sklearn/cluster/\_kmeans.py:870: FutureWarning: The default value of `n\_init` will change fro warnings.warn(

/usr/local/lib/python3.10/dist-packages/sklearn/cluster/\_kmeans.py:870: FutureWarning: The default value of `n\_init` will change fro warnings.warn(

/usr/local/lib/python3.10/dist-packages/sklearn/cluster/\_kmeans.py:870: FutureWarning: The default value of `n\_init` will change fro warnings.warn(

[62.789758974358975,

19.271322402597406,

6.100243749999999, 3.855219391025642,

1.855979047619047,

1.3019846031746019, 0.9283388888888878, 0.7150055555555541,

0.5600499999999995,

0.378185714285714]

([<

matplotlib.axis.XTick at 0x7b5aeeda3d60>,

<

matplotlib.axis.XTick at 0x7b5aeeda3d30>,

<

matplotlib.axis.XTick at 0x7b5aeeda3c40>,

<

matplotlib.axis.XTick at 0x7b5aec4704f0>,

<

matplotlib.axis.XTick at 0x7b5aec470fa0>,

<

matplotlib.axis.XTick at 0x7b5aec4714b0>,

<

matplotlib.axis.XTick at 0x7b5aec471f60>,

matplotlib.axis.XTick at 0x7b5aec472a10>,

<

matplotlib.axis.XTick at 0x7b5aec4734c0>],

<

Text(0, 0, '0'),

[

Text(1, 0, '1'),

Text(2, 0, '2'),

Text(3, 0, '3'),

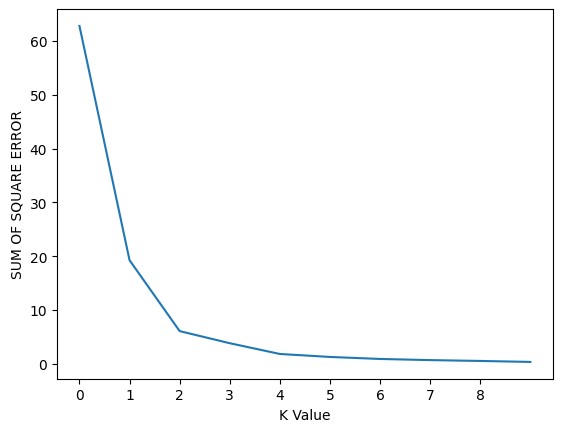
Text(4, 0, '4'),

Text(5, 0, '5'),

Text(6, 0, '6'),

Text(7, 0, '7'),

Text(8, 0, '8')])



plt.plot(inertias)

plt.xlabel(

'K Value'

)

plt.ylabel(

'SUM OF SQUARE ERROR'

)

plt.xticks(np.arange(

0

,

9

,

1

))

kn = KneeLocator(k\_range, inertias, curve='convex', direction='decreasing') elbow\_point = kn.knee - 1 elbow\_point

2

# elbow\_point = 3 (for experiment purpose you can change the value of elbow point manually and see the output) km = KMeans(n\_clusters = elbow\_point) y\_predicted = km.fit\_predict(df[['CGPA']]) y\_predicted

/usr/local/lib/python3.10/dist-packages/sklearn/cluster/\_kmeans.py:870: FutureWarning: The default value of `n\_init` will change fro warnings.warn( array([0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1], dtype=int32)

df['cluster']=y\_predicted df

**Name**

**CGPA**

**cluster**

**0**

KAUSHAL AGARWAL

9.10

0

**1**

ADARSH PANIKAR

9.20

0

**2**

LAKSHAY HIRANI

8.20

0



|  |  |  |
| --- | --- | --- |
| **3** SIDDHANT HADA | 8.00 | 0 |
| **4** SUDHANSHU KUMAR | 8.20 | 0 |
| **5** KALYAN PILLI | 5.90 | 1 |
| **6** KUNAL GUPTA | 9.05 | 0 |
| **7** SARTHAK SINGHAL | 8.90 | 0 |
| **8** YASH SALUJA | 8.00 | 0 |
| **9** NITIN KAUSHIK | 6.90 | 1 |
| **10** SANAT THAKUR | 8.20 | 0 |
| **11** DIWIK GOYAL | 8.10 | 0 |
| **12** RAM SHARMA | 10.00 | 0 |
| **13** SARTHAK GOYAL | 9.50 | 0 |
| **14** ABHINAV SINGH | 7.90 | 0 |
| **15** JOHN | 5.30 | 1 |
| **16** SIMRAN | 6.50 | 1 |
| **17** NEHA | 9.30 | 0 |
| **18** SHRUTI | 8.20 | 0 |
| **19** SRISHTI | 9.80 | 0 |
| **20** ATHARVA | 8.80 | 0 |
| **21** PRATHMESH | 6.60 | 1 |
| **22** CHAITANYA | 9.50 | 0 |
| **23** RUCHIRA | 7.80 | 0 |
| **24** ASHWIN | 8.00 | 0 |
| **25** VIBHA | 9.90 | 0 |
| **26** VISHVESH | 10.00 | 0 |
| **27** PAWAN | 6.80 | 1 |
| **28** OMKAR | 7.60 | 1 |
| **29** PIYUSH | 7.80 | 0 |
| **30** ASHU | 8.64 | 0 |
| **31** GAYATRI | 9.56 | 0 |
| **32** MARK | 8.99 | 0 |
| **33** DHONI | 9.25 | 0 |
| **34** VIRAT | 9.10 | 0 |
| **35** JADEJA | 5.65 | 1 |
| **36** BHUVI | 5.95 | 1 |
| **37** BOOMBOOM | 6.65 | 1 |
| **38** YUVI | 7.36 | 1 |

Next steps:

Generate code with

df



View recommended plots

km.cluster\_centers\_

array([[8.82107143],

[6.47363636]])

color = ['tomato','limegreen','midnightblue','blueviolet', 'g', 'r', 'c', 'm', 'wheat', 'k'] for i in range(elbow\_point): df0 = df[df.cluster == i] y = np.zeros(len(df0['CGPA'])) plt.scatter(df0.CGPA,y,color = color[i],label = 'CLUSTER ' + str(i)) y = np.zeros(len(km.cluster\_centers\_)) plt.scatter(km.cluster\_centers\_[:],y,color = 'yellow',marker = '\*',label = 'CENTRIOD',s = 500) plt.legend() plt.xlabel("CGPA VALUES") plt.xticks(np.arange(5, 10.5, .5))

([<matplotlib.axis.XTick at 0x7b5aeb2acbe0>, <matplotlib.axis.XTick at 0x7b5aeb2acc10>,

<matplotlib.axis.XTick at 0x7b5aeb2acd00>,

<matplotlib.axis.XTick at 0x7b5aeb2923b0>,

<matplotlib.axis.XTick at 0x7b5aeb290370>,

<matplotlib.axis.XTick at 0x7b5aeb291240>,

<matplotlib.axis.XTick at 0x7b5aeb290460>,

<matplotlib.axis.XTick at 0x7b5aeb256bc0>,

<matplotlib.axis.XTick at 0x7b5aeb255a50>,

<matplotlib.axis.XTick at 0x7b5aeb254700>,

<matplotlib.axis.XTick at 0x7b5aeb255480>],

[Text(5.0, 0, '5.0'), Text(5.5, 0, '5.5'),

Text(6.0, 0, '6.0'),

Text(6.5, 0, '6.5'),

Text(7.0, 0, '7.0'),

Text(7.5, 0, '7.5'),

Text(8.0, 0, '8.0'),

Text(8.5, 0, '8.5'),

Text(9.0, 0, '9.0'),

Text(9.5, 0, '9.5'),

Text(10.0, 0, '10.0')])

