import pandas as pd

import matplotlib.pyplot as plt

from sklearn.cluster import KMeans

data = {'x': [25,34,22,27,33,33,31,22,35,34,67,54,57,43,50,57,59,52,65,47,49,48,35,33,44,45,38,43,51,46],

'y': [79,51,53,78,59,74,73,57,69,75,51,32,40,47,53,36,35,58,59,50,25,20,14,12,20,5,29,27,8,7]

}

#create a dataframe with the data above

df = pd.DataFrame(data, columns=['x', 'y'])

#create a K-means model

model = KMeans(n\_clusters=3).fit(df)

#Get the coordinates of the centroids

centroids = model.cluster\_centers\_

print("The centroids for the groups are:")

print(centroids)

#plot the graph with the three groups

plt.scatter(df['x'], df['y'], c= model.labels\_.astype(float), s=50, alpha=0.5)

plt.scatter(centroids[:, 0], centroids[:, 1], c='red', s=50)

plt.show()

#Show the labels for each of the 3 centroids for each observation above in the dataframe

#in other words, this shows which group each x y pair belongs to

model.labels\_

#Predict which group a new observation belongs to

new\_observation=[[78,97]]

print("For an x value of 78 and y value of 97, the predicted group is:")

model.predict(new\_observation)