Name: Ruchir M. Adnaik

Srn: 202200282

Roll: 08

Batch: B1

DWDM ASSIGNMENT 07

Code:

import pandas as pd

import numpy as np

from sklearn.cluster import KMeans

from sklearn.preprocessing import StandardScaler

import matplotlib.pyplot as plt

import seaborn as sns

# Load the data

df = pd.read\_csv("student\_performance\_data.csv")

print(df)

# Convert categorical variables to dummy/indicator variables

df1 = pd.get\_dummies(df)

print(df1)

# Display info about the DataFrame

df.info()

# Scatter plot before scaling

sns.scatterplot(x=df['Age'], y=df['StudyHoursPerWeek'])

plt.title('Scatter Plot Before Scaling')

plt.show()

# Apply StandardScaler to 'Age' and 'StudyHoursPerWeek' columns

std\_scalar = StandardScaler()

scaled\_values = std\_scalar.fit\_transform(df[['Age', 'StudyHoursPerWeek']])

# Convert the scaled array back to a DataFrame for further use

scaled\_df = pd.DataFrame(scaled\_values, columns=['Age', 'StudyHoursPerWeek'])

# Scatter plot after scaling

sns.scatterplot(x=scaled\_df['Age'], y=scaled\_df['StudyHoursPerWeek'])

plt.title('Scatter Plot After Scaling')

plt.show()

# Apply KMeans clustering

kmean\_model = KMeans(n\_clusters=3)

y\_pred = kmean\_model.fit\_predict(scaled\_df)  # Use scaled\_df for clustering

# Scatter plot with KMeans clustering results

plt.scatter(x=scaled\_df['Age'], y=scaled\_df['StudyHoursPerWeek'], c=y\_pred, cmap='rainbow')

plt.title('KMeans Clustering Results')

plt.xlabel('Age')

plt.ylabel('Study Hours Per Week')

plt.show()

OUTPUTS:

A screen shot of a computer screen

Description automatically generated

A screen shot of a computer screen

Description automatically generated

A screen shot of a computer screen

Description automatically generated