#IMPORT THE LIBRARIES

%matplotlib inline

import pandas as pd

import matplotlib.pyplot as plt

import numpy as np

import seaborn as sns

plt.style.use("ggplot")

plt.rcParams["figure.figsize"]=(12,8)

from google.colab import files

uploaded = files.upload()

import io

df = pd.read\_csv(io.BytesIO(uploaded['netflix\_titles.csv']))

df.dropna(how='all',inplace=True)-------drop all rows with NA values

df.head()

#Visualize the Data

sns.scatterplot(x=df.SepalLengthCm,y=df.SepalWidthCm,hue=df.Species,style=df.Species)

#Standardize the Data

Separating the Target and Predictor variable and then standardizing them

x=df.iloc[:,0:4]

y=df.Species.values

from sklearn.preprocessing import StandardScaler

x=StandardScaler().fit\_transform(x)

# SVD Method to calculate Eigen values and Eigen vectors

#Eigen value decomposition Method to calculate eigen values and eigen vectors

covariance\_matrix=np.cov(x.T)

print("covariance\_matrix:n",covariance\_matrix)

eigen\_values,eigen\_vectors=np.linalg.eig(covariance\_matrix)

#Choosing/selecting the principal components using the explained variance

Each eigen value corresponds to one PC line.

for val in eigen\_values:

  print(val)

variance\_explained=[(i/sum(eigen\_values))\*100 for i in eigen\_values]

so we can choose eigen value 1 and eigen value 2

#Get the transformed dataset by projecting the points onto the new PC’s selected

projection\_matrix=(eigen\_vectors.T[:][:])[:2].T

projection\_matrix

x\_pca=x.dot(projection\_matrix)

x\_pca is the transformed dataset.