Identificar el minimo numero de variables

1. COVARIANZA
2. PCA

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.preprocessing import StandardScaler

iris = sns.load\_dataset('iris')

scaler = StandardScaler()

scaled = scaler.fit\_transform(

          iris[['sepal\_length', 'sepal\_width','petal\_length','petal\_width']].values

            )

covariance\_matrix = np.cov(scaled.T)

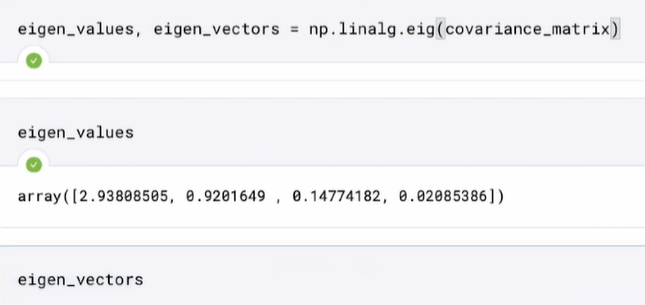
sns.pairplot(iris)

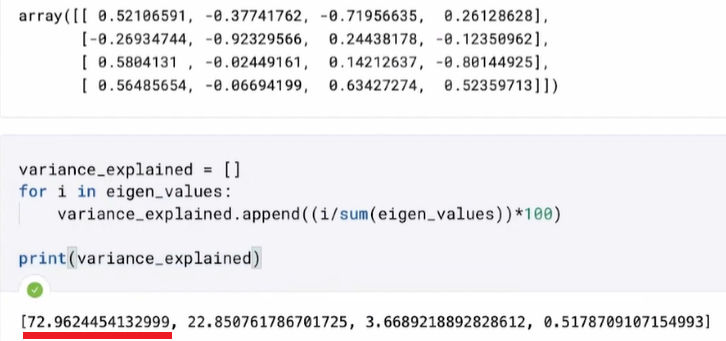
sns.pairplot(iris , hue='species')

|  |  |
| --- | --- |
| sns.pairplot(iris) | sns.pairplot(iris , hue=’species’) |
| Chart, histogram  Description automatically generated |  |



|  |  |
| --- | --- |
|  |  |





LO QUE HEMOS VISMO ANTES CON VARIANZA Y VECTORES SE RESUME USANDO LO SIGUIENTE,

LLEGAMOS A LA MISMA CONCLUSION QUE 2 COMPONENTES SONLOS PRINCIPALES, QUE APORTAN 72.9% Y 22.85%

from sklearn.decomposition import PCA

pca =PCA(n\_components=2)

pca.fit(scaled)

pca.explained\_variance\_ratio\_

#0.7296 0.2284