**Name: Tingyang Wei**

**Matriculation Number: G2202458H**

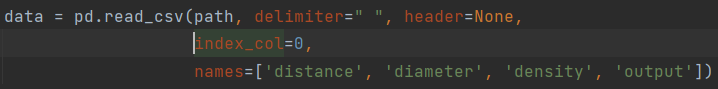
Serial Number: 32

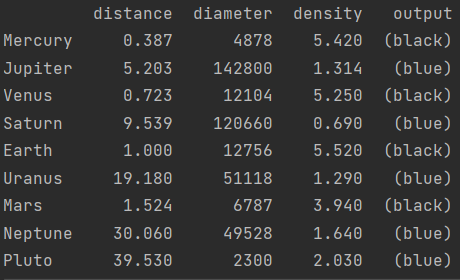
**Problem Description:**

Determine the Linear Transform to ensure a diagonalised covariance matrix in the transformed space. Determine the Vk values for k=1 and 2. Draw the scatter plot for PCA-1 and PCA-2.

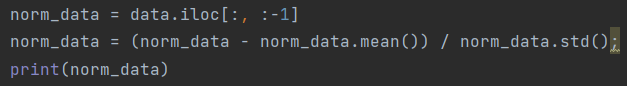
**Solution:**

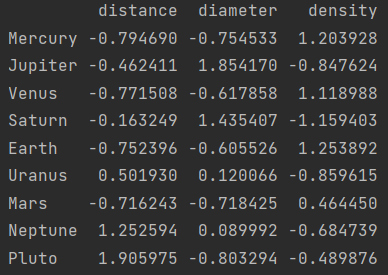
1. First we put all the data in planet datasets in a txt file and fetch the data using Pandas package in Python.



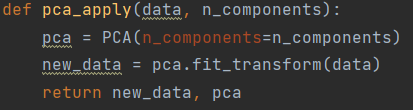


1. Prior to the diagonalization process, a whitening process is needed to enforce the mean of the data distribution to be zero, so that data distribution is invariant to scaling. Also rescale the data in the dataset so that the standard deviation of each dimension is one to remove measurement units.

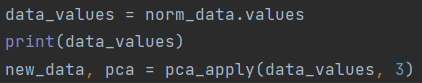




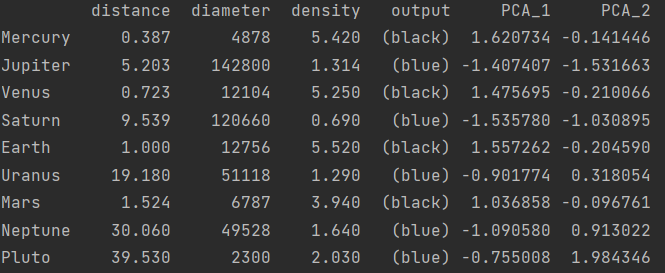
1. Then we conduct matrix decomposition using PCA decomposition package in sklearn of Python. Using the interface of this package, the transformed new data distribution as well as the PCA object can be obtained as follows:



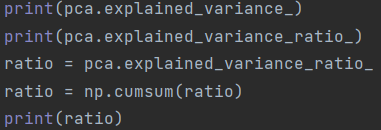
1. Since the requirement of the problem do not include dimension reduction, herein all the components are maintained and only the transformation is applied. To this end, the parameter “***n\_components***” should be set to 3 since the original distribution has 3 dimensions.

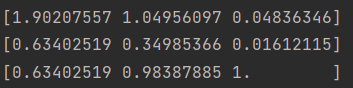


1. To implement the visualization we store the transformed dimension using PCA\_1 and PCA\_2 into the original dataset like below:



1. Also, according to the obtained PCA object in step 3, the variance, the relative ratio, the cumulative ratio V\_k of the resultant diagonal matrix, can be stated as follows:





**So V\_1 is 0.63402519 and V\_2 is 0.98387885.** This phenomenon happens and explains how PCA is constructing the direction with maximum variance.

1. Finally we obtained the transformed dimension values in the direction of PCA\_1 and PCA\_2, which stores more information than PCA\_3. We can plot the diagram with categorical information as below:

