Name: Abhishek N. Singh

#Student Number: 322079
#Assignment Project 1
#Course: Matrix Decomposition for Data Science
#Date: 2nd February 2021
#Description: We do analysis of SVD and PCA

Note: **The PDF file with the code and the explanations of assignment questions as comments should be considered as the main document PDF.**

First the cartopy package was installed that helps visualizing the world map with various python functionalities. SVD decomposition was done for the raw data that was read. Plotting was done using cartopy. Plot the first column of U so that the color indicates the value.
This is the first left singular vector. The plot is shown in the second PDF file which has codes as well as the plots. The data was then normalized using minimum and maximum values, as well as using the formula for z-score.

We know that the first column of U is more important than 2nd column of U and so on and so forth, in d escribing the data matrix. In other words, the left singular vector 1 is more important than the left singul ar vector 2, and the importance is tapped
in the S matrix 2.67803546e+05 being importance factor for 1st left singular vector and 1.24366268e+05 being the importance factor for 2nd left singular vector. For instance the first singular vector could be ta pping in the precipitation distribution
as we can see by the similarity in the color in the geographies with similar longitudes. The second left sin gular vector can perhaps be tapping the temperature as per latitude.

I also tried plotting with higher order singular vectors, and it confirmed that the maximum variation was tapped by initial vectors. Most part of the script is well documented and annotated, such that the examiner can well relate the chain of thoughts and analysis the coder is doing.

Different methods have been used to get the rank of the truncated SVD. Rank of 6 as per the Cattell's Scree plot is chosen to be the best, given that it takes into account the shape of the data.

Later on clustering by k-means with 10 random initialization is done, and plots are formed. The clustering well segments into climatic region as a combination of temperature and precipitation, where it seems that the temperature factor dominates much more than precipitation. The scatter plot is obtained using the 1st and 2nd left singular values of SVD decomposition, and interestingly the plot comes out to be very close to that obtained later by PCA method. The codes are well annotated to address the various concerns raised in the assignment. The PDF file with the code and the explanations of assignment questions as comments should be considered as the main document PDF.