Principal Component Analysis

1. Download the "applicant" dataset from the Homework 3 page on the course website. This dataset contains various features on applicants applying for a job at NCSA. The description of each feature is as follows:

FL	Form of their letter application
APP	Appearance
AA	Academic ability
LA	Likeability
SC	Self-confidence
LC	Lucidity
HON	Honesty
SMS	Salesmanship
EXP	Experience
DRV	Drive
AMB	Ambition
GSP	Ability to grasp concepts
POT	Potential
KJ	Keenness to join
SUIT	Suitability for position

2. Use Python to solve the problems below. Submit your code and output of the code tabularized in a readable format. You do not need to submit a separate report.
    a. Find correlation (**R**) and covariance **($\hat{\sum}$)** matrix of the applicant dataset.
    b. Should principal component analysis be performed on the covariance or the correlation matrix? Explain your choice.
    c. Find the eigenvalues and eigenvectors of both **R** and $\hat{\sum}$. Show that all pairs of eigen vectors are orthogonal to one another.
    d. Show that sum of the eigen values of $\hat{\sum}$ is equal to the trace of $\hat{\sum}$**.**
    e. Run PCA analysis on the dataset based on your choice in part 2.b.
        i. How many principal components are required to describe the space in which the data actually falls?
        ii. Show the scree plot for your PCA analysis
        iii. Plot the first principal component against the second principal component and include an interpretation
        iv. Explain the benefits of running PCA analysis on this dataset