MA544 - Programming Assignment - 3

Name:

```
In [ ]:
# Import required packages
```

Question 1

(An open ended asignment) Visit the University of California open source [data repository](). Pick up an appropriate dataset of your choice preferably one with default-task 'classification'. Follow the examples provided during the class to get some insight from the data that you have selected. In the process of doing so utilize any tools and techniques at your disposal including (and not limited to)...

- Observation of covariance and correlation between different features (columns, explanatory variables).
- Distribution of different features and observations.
- Utilize PCA to describe low-dimension representations.
- Use plots such as Scree-plot, Bi-plot etc.
- Discuss and clustering behaviours around the classes exposed by the first few PCs.
- Anything else that you find interesting.

```
In [ ]:
# Your code starts here
```

Question 2

Use the Python code for image compression to compress the national flags of different countries of your choice. The smallest rank r gives the numerical rank of the flag. Pick up any 5 national flags and arrange them in the order of decreasing numerical rank*.

*: the numerical rank could be given by the number of singular values larger than $\square \times \|A\|_F$ where $\square = 10^{-16}$ is the machine epsilon.

```
In [ ]:
# Your code comes here
```

Question 3

Use your understanding of singular value decomposition, matrix completion and recommendation systems (from your last reading, R2) to build a basic movie recommendation system.

- Evaluate your system on the test data (20%) and present the statistics (precision, recall and F1-score).

Use the following data set https://grouplens.org/datasets/movielens/100k/.

```
In [ ]:
```