CMTH642 Data Analytics: Advanced Methods Lab 1

1. Read the **train.csv** file from the following website:
   https://raw.githubusercontent.com/agconti/kaggle-titanic/master/data/train.csv

   Have a look at the dataset by using the **head**, **tail**, **str** and **summary** commands.

2. Change the **Pclass** and **Survived** attributes to factors.

3. Check the missing values of the **Age** and **Name** attributes. The **Name** attribute consists of titles such as **Miss.**, **Mrs.**, **Mr.** and **Dr.** For a title containing a missing value, assign the mean age value for each title not containing a missing value. After these imputations, check the missing values of **Age**.

4. List the distribution of Port of Embarkation. Replace empty entries with NA for **Embarked** attribute. Assign the two missing values to the most counted port, which is **Southampton** in this case.

This is the end of lab 1
Ceni Babaoglu, PhD