***Dealing with missing values:***

***Step-1: Data imputation:***

* Are there missing values? If there are, impute the missing blocks by the mean of all of the other values in that column.

***Feature scaling:***

***Step-2: Normalization:***

* Rescale values to fall in the range [0, 1]
* ***Why is this necessary?*** 
  + Normalization helps ensure that all of the features (independent variables) fall in the same range.
  + If the value of one particular feature is way larger than all of the rest, such a scenario is analogous to having higher importance being designated to that feature. In other words, that one feature would be instrumental in driving the prediction decision above all the rest.
  + We will have a greater intuition of this once we design a simple logistic regression model later on in this project.
* ***Min-max normalization:***

The new features are obtained as follows:

* ***Code shortcut in Matlab:***

***Step-3: Standardization:***

* This transforms the data distribution to have a mean value of 0, and a standard deviation of 1.
* ***Intuition:*** We have a feature with zero mean; and all of the samples are centred around the mean with a uniform standard deviation.
* ***Formula for standard deviation:***
* ***Code shortcut in Matlab:***