INSTRUCTIONS

You can find the incomplete datasets in the folder entitled "Incomplete Datasets".
You can find the complete datasets in the folder entitled "Original Datasets" to compute NRMS and AE.
You should report the attained NRMS/AE results in excel file entitled "Table-NRMS-AE".
You need to save each imputed dataset and deliver them at the date of final demo.
You must present the final results, completed Table-NRMS-AE, explanation and conclusions in your final report.

Evaluation of imputation for the numerical datasets can be done in terms of the normalized root mean square difference (NRMS) that is defined as follows:

$$NRMS = \frac{\left\| X^{estimate} - X^{original} \right\|_F}{\left\| X^{original} \right\|_F}$$

where $X^{original}$ and $X^{estimate}$ are the originally complete dataset and the imputed dataset, respectively; and ||.|| stands for the Frobinus norm

$$\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}$$

To evaluate the imputation performance on the datasets with categorical values, you should use the AE function instead of NRMS as follows:

$$AE = \frac{1}{n}\sum_{i=1}^{n} I(\hat{x}_i = x_i)$$

where $I(.)$ stands for function that returns 1 if the estimated value $\hat{x}_i$ and real value x_i are the same (i.e., if the condition is true), but 0 otherwise.

Thus,
1) If the dataset is Numerical then use NRMS.
2) If the dataset is Categorical then use AE.
3) If the dataset contains both Numerical and Categorical attributes then for its numerical features use NRMS and for the categorical features use AE.

Note: Each group must study the "computational complexity" of the method/algorithm and discuss it in CP, PD, FD and FR. A proper/improper computational complexity analysis will significantly change your grade. This will be discussed in the class.