

Assignment Date: 24/01 Assignment Due Date: 7th Feb,2018 (2 weeks)

Note: Why are we doing this? We want to become comfortable in Pandas (if we are not)

Marks: 5

Instruction:

1. Mail a Jupyter notebook (use Python 3.0)

2. Skills required is: Pandas in Python

1. Download the following dataset and understand the dataset

<https://archive.ics.uci.edu/ml/datasets/Parkinsons+Telemonitoring>

2. Load the dataset into data frame

3. Provide statistical information about the data (various variables)

4. Total_UPDRS column is chosen to be your dependent variable. Drop the other candidate dependent variable motor_UPDRS.

5. Look at the data distributions with box plots and histograms. What does the boxplot say about skewed nature of the distribution of each variable?

7. Address skewed nature of the distribution by transforming to log values using numpy.

8. Create a scatter matrix and comment.

9. Do scatter plot analysis of possible independent-dependent variable pair. What will be the possible independent variables candidates for multiple regression?

10. Attempt simple regression analysis of the above candidate independent variables using Seaborn library **(optional)**