**README**

Our coding environment was based on Colab. Please upload the .ipynb file to Colab and run the code. And remember change the data files’ path to yours.

There two parts of code in the **Main** notebook:

1. Code to get the statistics used to complement the descriptive analysis of the dataset.
2. Code to train models on the training set and evaluate them on the test set (regression models only).

For part 1, please run all cells in ‘Data preparation’ and T1 from top to bottom.

For part 2, please restart the runtime after you executed part 1 code. And follow the instruction below to run the supervised learning and neural network models:

**Regression models:**

The regression models are very simple to run, although they may require more RAM than the standard google Collaboration has.

1. Run all of the cells in ‘Data preparation’
2. Skip T1
3. Run all cells in T2 before ‘Preprocessing for Neural Network’ and then the cells under ‘Preprocessing for Supervised Learning Models’
4. Run the first 2 cells under the T3 - Implementation + Results block
5. Run the cells between 10% regression models and Feature Selection - Univariate Selection
6. Run all cells between KNN Regression Model and Decision Tree Classifier
7. Lastly run Decision Regression Model until the end.

**Neural Networks:**

The steps/Requirements to run the RNN models on the dataset are as follows

 REQUIREMENTS

Colab Pro or similar environment with RAM upto 25 GBs

PACKAGES

1. Tensorflow
2. Keras
3. Sklearn
4. Numpy
5. Pandas
6. Matplotlib

INSTRUCTIONS

For running only the neural network implementation please follow these steps

* Run all “Data preparation” cells followed by “Import the dataset”
* Skip T1
* Run the “Outlier treatment” and “Check Building outlier” cells
* Run all the sub cells from “Preprocessing for neural networks”
  + The “groupby” statement needs almost 22 GBs of RAM to run\*
* In T3 do the following
  + Load all the libraries
  + Goto GRU and LSTM cells
  + Run all the cells upto Early\_stop command line
  + After that run model\_gru or model\_lstm to run GRU or LSTM models

\**The RAM usage after the groupby commands remains at 20 - 22 GBs*

TROUBLESHOOTING

 If at any point the notebook crashes please run all the steps again after restarting the runtime.

In addition to make predictions by neural networks on the test set and do Kaggle submission, we created another notebook for it due to the memory shortage.

Below is the instruction for the ‘Neural Networks Test code for Kaggle’ notebook.

**Neural Networks Test code for Kaggle:**

The steps/Requirements to run the code for generating csv for kaggle submission for model trained only half of the data is as follows

REQUIREMENTS

Colab Pro or similar environment with RAM upto 25 GBs

PACKAGES

1. Keras
2. Tensorflow
3. Sklearn
4. Numpy
5. Pandas
6. Matplotlib

INSTRUCTIONS

For running only the neural network implementation please follow these steps

* Run all “Data preparation” cells followed by “Import the dataset”
* Run the “Outlier treatment” and “Check Building outlier” cells
* Run all the sub cells from “Preprocessing for neural networks”
  + The “groupby” statement needs almost 22 GBs of RAM to run\*
* In T3 do the following
  + Load all the libraries.
  + GRU : run all the sub code cells in it.
  + Run the test set prediction (kaggle).

\**The RAM usage after the groupby commands remains at 20 - 22 GBs*

TROUBLESHOOTING

 If at any point the notebook crashes please run all the steps again after restarting the    runtime.