----What is this README and REPO?---

This repository provides python codes to reproduce the plots and tables from

the paper Exploratory Analysis of Machine Learning Methods in Geothermal Energy Research.

The file formats are in Jupyter Notebook IDE format (.ipnyb).

If you wish to obtain a .py version, please email to seho0808@vt.edu.

README is in two formats: .pdf and .txt. They contain equivalent information.

----Prerequisite Datasets----

AASG Dataset - This file is used throughout the codes and contains the well data throughout North-Eastern USA.

=> Go to https://gdr.openei.org/submissions/638

=> Click Download next to

ThermalQualityAnalysisThermalModelDataFilesStateWellTemperatureDatabases.zip

=> If you unzip, there should be AASG_Thermed_AllThicksAndConds.xlsx

Well Dataset - This file is cleaned-up version of new well data (2016~) and used in some of the files.

=> Go to https://drive.google.com/open?id=12EqQ3etRl3uu7yyPWwBJ8E7SIOVv3_Yu

=> You should change any import of this csv file directory within the .ipnyb files.

For example, in "0306_Comparing Well Prediction.ipnyb",

```
df2 = pd.read_csv('Past+New_Wells/clean_new_well_data.csv')
```

above line of python code should be,

```
df2 =pd.read_csv('<Insert Your Relative Directory>/clean_new_well_data.csv')
```

But this "0306_Comparing Well Prediction.ipnyb" file is currently the only known

file that imports this .csv file. Therefore, you only have to change a single line in the

whole document. If you place your clean_new_well_data.csv in the same directory level

as "0306_Comparing Well Prediction.ipnyb," you don't have to insert your relative directory,

but just say

```
df2 = pd.read_csv('clean_new_well_data.csv')
```

----Documentation----

Many of the figures and tables are omitted as they are not produced by ourselves, or of less importance in terms of code documentation.

Figure 1 - 0305_Compare_New_and_Old_LONG_LAT.ipynb

Figure 5 - _Organized Tests.ipynb*

*This is a legacy version of 0310_HPTuning+ExcelMetric+NewWellMetric.ipynb. Therefore, you should ignore everyting except the importance plot.

Figure 6 - 0316_Q_map.ipynb

Figure 7 - 0314_depth_maps.ipynb & 0315_map_excel_data.ipynb

Figure 10 - 0306_Comparing Well Prediction.ipynb

Figure 11 - 0316_Q_map.ipnyb

----Appendix----

Cross Validation, Model Comparison, and Hyperparameter tuning for XGBoost, DNN, Ridge Regression, and Random Forest is in 0310_HPTuning+ExcelMetric+NewWellMetric.ipynb