**Files**

|  |  |
| --- | --- |
| **Folder Name** | **Purpose** |
| Dashboard UI | VS solution of the dashboard |
| Data | Data used in the Machine Learning Life Cycle |
| Data Preprocessing | EDA, Heatmap, PCA |
| Training Model | Different algorithms used to train model |

**Database Setup**

1. Run the “Database Script” in SQL server
2. Run the “Table Script” in SQL server

**Data Preprocessing**

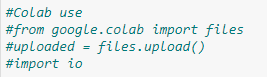
The python scripts can be directly used in Jupyter Notebook with all the required libraries.

|  |  |
| --- | --- |
| **Library** | **Version** |
| numpy | 1.21.2 |
| pandas | 1.2.4 |
| matplotlib | 3.3.4 |
| seaborn | 0.11.1 |

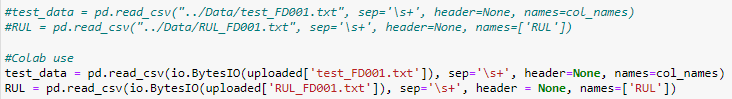
Another option to run the python scripts is to upload the files in Google Colab (<https://colab.research.google.com/>)

In order to run **Exploratory Data Analysis (EDA)** in Colab, please follow the steps below:

1. Uncomment the code shown below in the first cell



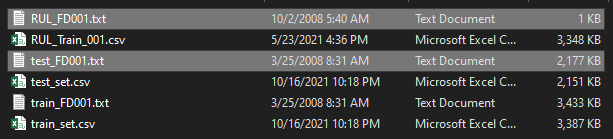
1. Comment and uncomment the code to looks like the figure below



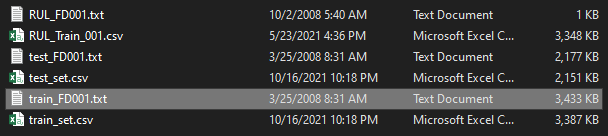
1. When the script is executed, Colab will prompt to select files



1. Click “Choose Files” and select the 2 files below in the Data folder for EDA on test set

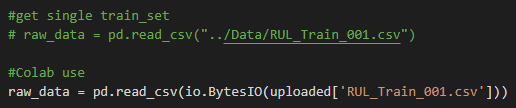


1. Select the file below in the Data folder for EDA on train set

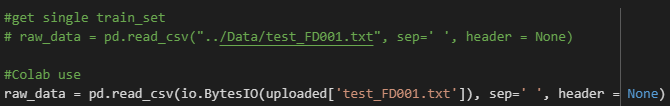


In order to run **Heatmap** and **PCA** in Colab, please follow the steps below:

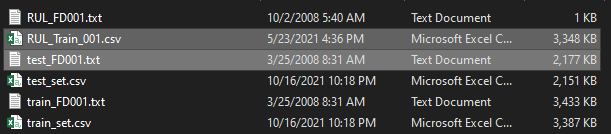
1. Follow the step 1 in **EDA**
2. Comment and uncomment the code to looks like the figure below



Test Data cell



1. Follow the step 3 in **EDA**
2. Click “Choose Files” and select the 2 files below in the Data folder



**Training Predictive Models**

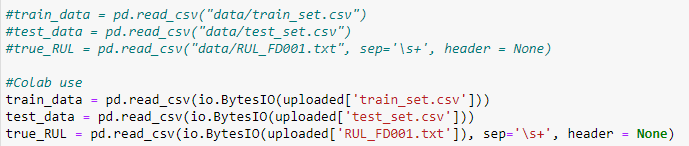
The python scripts can be directly used in Jupyter Notebook with all the required libraries.

|  |  |
| --- | --- |
| **Library** | **Version** |
| numpy | 1.21.2 |
| pandas | 1.2.4 |
| matplotlib | 3.3.4 |
| seaborn | 0.11.1 |
| pickleshare | 0.7.5 |
| scikit-learn | 0.24.1 |

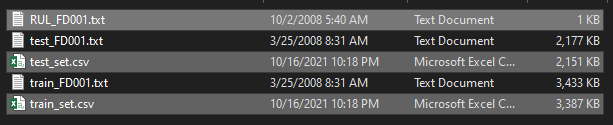
Another option to run the python scripts is to upload the files in Google Colab (<https://colab.research.google.com/>)

In order to run in Colab, please follow the steps below:

1. Follow the step 1 in **EDA**
2. Comment and uncomment the code in the “Load Data” cell to looks like the figure below



1. Follow the step 3 in **EDA**
2. Click “Choose Files” and select the 3 files below in the Data folder



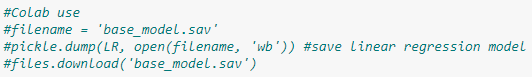
**Saving Predictive Models**

***#The standardization and PCA model can be saved in the Gradient Boosting Regression Script.***

In order to save the model in Jupyter Notebook, please uncomment the code below



In order to save the model in Colab, please uncomment the code below.



***#Please save the model in “Dashboard UI/Dashboard UI/model”***

***Dashboard setup***



Please change the file path of the Python executable in **App.config** file. Make sure that the python executable has the libraries below

|  |  |
| --- | --- |
| **Library** | **Version** |
| numpy | 1.21.2 |
| pandas | 1.2.4 |
| pickleshare | 0.7.5 |



Please change the connection string in the **App.config** file.