Guide to setup folder and run Python code to perform Data Preprocessing and Interpretation

Course ID: WQD7005

Course Name: Data Mining

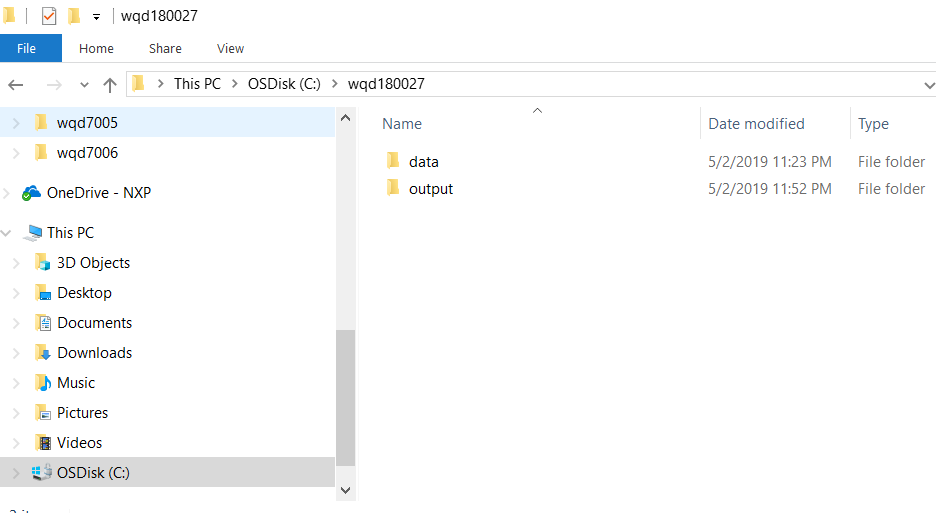
Lecturer: Assoc. Prof. Dr. Teh Ying Wah

Student Name: Lim Shien Long

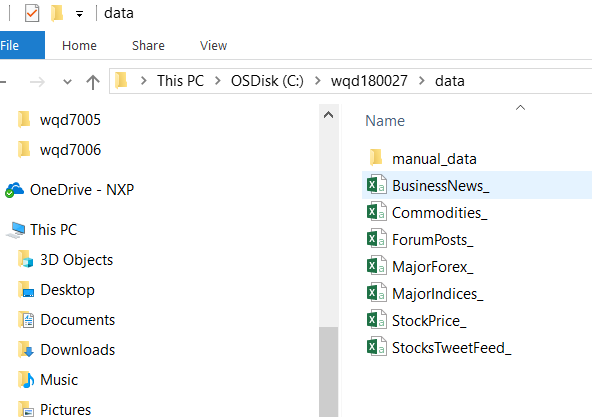
Student Matrix: WQD180027

Steps to Run Code

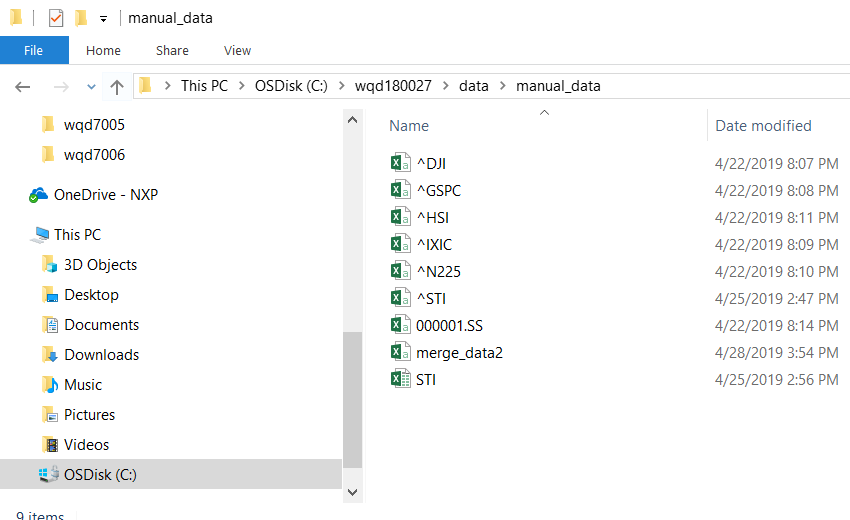
1. Unzip ‘wqd180027.zip’.
2. Place folder in ‘C:/’
3. Folder path should be ‘C:/wqd180027’



1. ‘C:/wqd180027/data’ folder should contain below files and folder



1. ‘C:/wqd180027/data/manual\_data’ folder should contain below files



1. Python codes are in different zip folder (‘wqd180027\_python.zip’)
2. Codes are run in this order:
   1. Cleaning.py – clean missing data, pivot data, dropping unused columns, merging data
   2. Cleaning\_merged.py – imputing data and some manual data added to merge2.csv
   3. Ins\_tgt.py – create new columns for target and assigning nominal value based on ChgPct of stock
   4. Corr\_Analysis.py – simple correlation analysis, removed CrudeOil and Nasdaq due to correlation to BrentCrude and SP500 respectively
   5. PCA.py – PCA analysis on MYEG stock, work in progress to understand how to use PCA for dimension reduction and then using reduced dimension data for prediction.
   6. DecisionTree.py – decision tree on MYEG data, tree diagram saved as PNG file in ‘C:/wqd180027’
   7. Arima.py – ARIMA prediction on MYEG data