MT Assignment

Course: IT Applications to Banking and Finance – ML53

Using Python to download data from an online data base, i.e., The World Bank Database, The IMF Database, The Federal Reserves System, and many other databases. Those databases support API access. If you try with a recent method to scrape data from Facebook or other social networks, you project will be awarded a premium.

* Provide data cleaning techniques by which any possible data errors such as NAs, missing values, outliers, … can be identified and treated by a relevant method.
* Use the cleaned dataset and provide descriptive statistics of this dataset. Use tables, graphs and summary statistics to provide your comments on the main features of this dataset by comparing and contrasting the descriptive statistics, highlight your finding on the interesting features of the dataset.
* Using this dataset to perform statistical calculations on means, variances, covariances and correlations of variables in that dataset. Write your comments on those statistics to show your understandings on these statistical instruments in discovering features and tendencies within the dataset.
* Based on your data set, decide to use Random Forest Classification or Regression to perform data analysis. Report your results in comparison to the standard econometric method. Give comments to explain the RF algorithm and to explain the criterion used in model testing.
* Students need to submit all require plots, comments and Jupyter Notebook in a PDF file and a Jupyter Notebook. All codes implemented in Jupyter notebook will earn full mark. Codes that are mainly based on available python libraries will earn 50% of the mark allocated. Students submit code file only without any comment will earn zero mark.
* Deadline: 9am, 26.02.2025 on MsTeams.
* Late submission will be deducted by 10% of the mark. If submission is late more than 3 days, it will be marked zero.
* Assignments partially generated by ChatGPT or other AI agents will be marked zero.