# Programming for Data Analytics

# Application Of Python to Data Analytics

Summited by: Abusomwan Amadevborho

Date: 02/12/2021

Application of Python to Data analysis has been an interesting journey for myself in the past few months, I have become acclimatized with its applications. Furthermore, I would be discussing some of my experience learning with the various packages in python (NumPy and Pandas) and the usefulness of the software itself. I would be using the dataset “crypto-markets.csv” which is also attached to this assignment in the jupyter notebook as a case study.

Firstly, I would discuss my experience with the use of NumPy and Pandas also, how easy we can get our summary statistics without going through any stress of analysing our dataset manually with the use of formula. Given my background in Economics and Statistics it was a big task getting summary statistics, plotting trends and its interpretation, in particular has always been a stressful task because this was normally done manually using its various formulas for data processing to build a model for visualization, However with the use of python libraries like pandas getting summary statistics such as mean, standard deviation, minimum, maximum value and percentile can be so easy and can be done in less than a minute with the use of pandas library “describe()” function in contrast to manual calculations as seen in the attached crypto Jupiter project analysed in this project. Moreover, getting information from our dataset which is a large time series csv file with 13 column and 942,297 rows ranging from 28-04-2013 to 21-02-2018 was done in few minutes using the pure python codes and pandas libraries.

Secondly, with the help of python for data analysis, forecasting or prediction in the context of time series analysis or categorical data is easy to analyse with the use of trends and plotting of graphs. Also as illustrated in the attached jupyter notebook we can clearly see the time trends using the pure python codes and the pandas code despite the continues increase in the dataset, the pandas runs faster than the pure python code. This was done using following the “import time” syntax to know the time durations the various codes are executed compared while using libraries. Nevertheless, after getting our time comparison of the various codes, a line chart was plotted for a proper visualization and interpretation.

Thirdly, another interesting fact I learned while using python is how information from a dataset can be passed into a function. This can be user defined function when we can add various argument in the function as seen in the crypto dataset manipulation where we made use of the “elif statement” to predict if our market is a bull market or a beer market and the “ifelse” statement to predict if the crypto market is bullish or bearish. Also, we can also get loops using python from a dataset as seen when using “while loop” in the Jupyter notebook project.

Lastly, it was interesting to notice that codes and syntax generated using a function in a dataset in some cases cannot work for another dataset except the function becomes defined in the new dataset and this was clearly seen in the attached crypto jupyter file when I constantly got an error trying to use task 1 codes for task 3 dataset, however some codes were also able to run, with this I came into conclusion that a code can work more effectively and efficiently when it is defined in the dataset.