PROGRAMMING FOR DATA

INSIGHTS GAINED FROM APPLICATION OF PYTHON TO DATA ANALYTICS.

By: Osa Gilbert Enofe

27/11/2022

Over the past few months, I have become accustomed to Python's applications for data analysis. Additionally, I will discuss some of my experience learning various Python packages (NumPy and Pandas) and why the software is useful. I will be using the “pizza\_sales.csv” dataset which is attached to this assignment in the Jupyter notebook as a case study.

To begin with, I would like to speak about my experience with NumPy and Pandas, and how we can easily get summary statistics without going through the stress of manually analyzing our dataset with formulas. With a background in Economics and Statistics, I found it challenging to get summary statistics, plot trends, and interpret them, which has been a stressful task since data processing and visualization have always been done manually using various formulas. However, with the help of Python libraries like pandas, summary statistics such as mean, standard deviation, minimum value, maximum value and percentile can now be easily obtained and this is done in less than a minute with the use of pandas library “describe()” function in contrast to manual calculations. In addition, getting summary statistics from our dataset which is a large csv file with 12 columns and 48,620 rows was done in few minutes using the pure python codes and pandas libraries.

Secondly, with Python for data analysis, forecasting or prediction of data is simple to analyze using trends and graph plotting. Also, as shown in the attached Jupyter notebook, pandas run faster than pure Python code.

Thirdly, another fascinating fact I discovered while using Python is how to pass information from a dataset into a function. we can get loops from a dataset using Python.

Finally, it is interesting to notice that the codes and syntax of a dataset in some cases cannot work for another dataset unless the function becomes defined in the new dataset. This was the case when I constantly got an error trying to use task 1 codes for the task 3 dataset. We can therefore conclude that a code can work more effectively and efficiently when it is defined in the dataset.