Article Writing

Visualization Techniques for Data Analysis utilizing *.ipynb* modules

Techniques covered:

Pandas dataframe

Plotly

Seaborn

Matplotlib

PowerBI

For a beginner in the data science field, the breakdown of activities can be considered as (in ascending order):

Problem Identification

Strategy Approach

Data Collection

Data Analysis & Manipulation

Prediction Modelling

Model Deployment

Continual Feedback & Improvement

In an interesting fashion, it may not be necessary to go through each of these steps to develop the necessary solution for the problem at hand. However, the chronological order would have to be maintained, for it is highly unlikely that a lower-level step is able to be carried out successfully without consideration of its predecessor.

The first three levels are quite subjective in whichever problem case. The following 4 steps constitute the general activity breakdown of solving the problem, & may have some similarities between completely variant scenarios.

In this article, I wish to cover a small section on the data analysis step, which focuses on data visualization in order to develop insights based on the data collected.

For a beginner, I found that there were multiple ways to carry out visualizations when utilizing Python notebooks, powered by the various packages that are currently in trend. However, there was little comparative articles on which way to go, only a series of responses online on how to develop specific plots, dependent on what was googled.

Since I was on a learning sprint, my policy was:

1. I wanted to figure out which plotting strategy would be ‘simplest’,
2. Plotting strategy that would be most intuitive for use, so that I don’t have to cram the code lines, but really understand what each declaration referred to.
3. Adequate online support
4. Easy integration with final reporting techniques at different levels of professionalism.

For the comparison, I considered 4 techniques of plotting data on Python Notebooks:

Pandas.plotting

Seaborn

Matplotlib

Plotly

Additionally, I considered utilizing Dashboard visualization softwares:

Microsoft Power BI

Tableau

Datasets

Two comparisons are offered:

Large single dataset,

Multiple datasets using join functions

The following features are compares:

Complexity of design

Speed of plot