IS-733 HOMEWORK 1 – UB01976

GPT – Usage Description

1. The GPT helped me understand the correlation heatmap generated by my code, the prompt used is as follows "**Can you explain me how to read the values in the correlation heatmap and what do they mean for based on values**". This helped me understand and interpret the results, but then I used the insights provided from the heatmap to make my conclusions about water quality.
2. To improve my data visualization skills, I have asked GPT the following prompt "**What are the best visualization tools available? And how to use Data profiling library in python?**". It provided me with the available tools presently in market like Tableau for visualization and how libraries like Data Profiling (y-data profiling) can be used for explorative analysis of data, and how it can be used in my code.
3. To understand which data mining technique would be best suited for the dataset, I have given GPT the following prompt "**Based on the Water Quality Dataset, Can I use Random Forest or SVM?**". It gave me clear analysis on which one to prefer based on the AUC, Accurracy, and clear understanding. As a result, I have chosen Random Forest to analyze the water quality from the given metrics in dataset.
4. In the second question I had problem with the date, as I couldn't generate it in proper format, So I have GPT the following prompt "**Dates are not taken properly, and I am unable to plot them on the temporal plots? Can you explain to me what's wrong with my code**". It clearly explained to me how to extract the date into proper distributions of year, month, and day of week. By this I could make changes in my code and could successfully generate temporal plots.
5. While generating the dashboard, the plot was becoming too clumsy and shabby, as a result I could not analyze or visualize trends clearly. So, I have asked GPT the prompt that "**The plot in the dashboard looks clumsy and shabby, can you suggest any ways to improve it?**". Based on my question, it gave me several options, out of which Rolling average window size was good enough, to improve the way on how to visualize the plot.