Dr. Tony Diana DATA 602 Introduction to Machine Learning Homework | Week 3

Time Series Analysis

- Use the multiTimeline.csv data.
- Import the main libraries (numpy, pandas, seaborn, matplotlib)
- Check the information about the data.
- Next, we need to convert the 'month' column into a 'DateTime' data type and make it the index of the DataFrame. We need to do this because when you generated the information about the data, the 'Month' column was actually an 'of data type' object. That generic data type encapsulates everything from strings to integers.
- We need to use '.to_datetime()' to convert the 'month' column in the DataFrame to a DateTime.
- Make sure to include the inplace argument when you are setting the index of the DataFrame df so that you actually alter the original index and set it to the 'month' column.
- Check with 'df.head()'.
- Visualize the data with matplotlib.
- Plot diet only.
- Use the rolling average of 'diet' using pandas' 'rolling' method.
- Do the same with 'gym'.
- Plot the trend of both 'diet' and 'gym'.
- One way to think about the seasonal components to the time series of your data is to remove the trend from a time series, so that you can more easily investigate seasonality. To remove the trend, you can subtract the trend you computed above (rolling mean) from the original signal. Another way to remove the trend is to use 'differencing': You consider the difference between successive data points (called "first-order differencing", because you're only looking at the difference between one data point and the one before it).
- Check the correlation among the variables.
- Start off by plotting the first-order differences with the help of .diff() and .plot().
- The variables 'diet' and 'gym' are highly correlated once you remove the trend. Now, compute the correlation coefficients of the first-order differences of these time series.
- Use the plotting interface of pandas, which has the autocorrelation_plot() function. You can use this function to plot the time series 'diet'.
- What do the dotted lines represent?
- What is the seasonality?