Time series prediction

You are given the electricity power units (MegaWatts) consumed on a daily basis by Himachal Pradesh as a csv file (dataset1_HP.csv). To study the power consumption in the light of COVID19 (lockdown), data is recorded in the form of a time series for a period of 17 months beginning from 2nd Jan 2019 till 23rd May 2020. Rows are indexed with dates, first column represents the date and second column represent power consumed in Himachal Pradesh. Rows and columns put together, each data point in second column reflects the power consumed in Mega Units (MU) by the Himachal Pradesh at the given date.

1. Create a line plot with x-axis as index of the day and y-axis as power consumed in mega units (MU).

2. Plot the autocorrelation of the given sequence at values of time delay (p)=0,1,2,....,10. (You can use plot_acf() function for this.)

3. Plot the partial autocorrelation of the sequence for these values of lags (use plot_pacf() function).

4. A general autoregression model estimates the unknown data values as a linear combination of given lagged data values. For example, data value at (t+1) instant, denoted by x(t+1) can be estimated from its previous instance values, such as x(t+1) = w0 + w1*x(t) + w2*x(t-1) + ... +wp*x(t-p+1). The coefficients w0, w1, ... wp can be estimated while training the autoregression model on training dataset.

   a. Split the data into two parts for training and testing. Choose the first 375 days as training data and last remaining days as test data.

   Build an autoregression (AR) model using AutoReg(). This function generates an AR model with the specified training data and lagged values (given as its input). Use 5 lagged values as its input (p=5). Train/Fit the model onto the training dataset. Use the trained AR model to predict the values for the test dataset and compute RMSE between original and the predicted values.

   Repeat the procedure for lag (p)= 1,5,10,15,20,50,100,150. Find the RMSE between original test data and the predicted values in each case. Display the value of the lag for which the RMSE is minimum.

4. Build an ARIMA model on the train data and test it on the test data. Give the RMSE between the original and predicted values for different values of p (lag), d (difference), and q (size of moving average window).