



PES University, Bangalore
(Established under Karnataka Act No. 16 of 2013)
UE16CS322 – DATA ANALYTICS
Assignment 4 - Time Series Analysis

Date of Submission: 17 / 10 / 2019

Max. Marks: 20

Note: *This is a Group Assignment and one submission per group is accepted. Submissions to this Assignment will only be accepted as Kernels in Kaggle. Please mention complete Team Member Details in your Kaggle Kernel!*

TOPIC: Time Series Analysis

Weather warnings are important forecasts because they are used to protect life and property. Forecasts based on temperature and precipitation are important to agriculture, and therefore to traders within commodity markets. Temperature forecasts are used by utility companies to estimate demand over coming days. On an everyday basis, people use weather forecasts to determine what to wear on a given day. Since outdoor activities are severely curtailed by heavy rain, snow and wind chill, forecasts can be used to plan activities around these events and to plan ahead and survive them. Source:

<https://en.wikipedia.org/>

[Kaggle Link to DataSet - Assignment 4](#)

Question 1 (5 points)

- a) Plot the *meantemp* variable across the dataset. **(1 point)**

Note: Make sure you interpolate the data (linear interpolation) before plotting. (Use this interpolated data for further use as well)

- b) Calculate a Moving Average filter for the mean temperature with

- i) A 10-tap $\frac{1}{10}[1, 1, 1, 1, 1, 1, 1, 1, 1, 1]$ filter
- ii) A weighted average $\frac{1}{10}[1.5, 1.5, 1, 0.5, 0.5, 0.5, 0.5, 1, 1.5, 1.5]$ filter.

Plot the graph after smoothing it with the above two filters. What are your observations?

(2 points)

- c) How does resampling help a time-series data? Resample the *meantemp* data

- i) Hourly
- ii) Weekly
- iii) Monthly
- iv) Quarterly

Which of these capture the essence of the dataset? Why? **(2 points)**

Questions 2 (7 points)

- a)
 - i) Decompose the Monthly resampled data into trend, seasonality and residual. Is it an additive or multiplicative series? **(2 points)**
 - ii) Plot the ACF and PACF plots for the *meantemp* series. What can you conclude from the plots? **(2 points)**
- b) Is the *meantemp* a stationary time series? **(3 points)**
 - i) Provide one **statistical** and one **non-statistical** test to support your claim.
 - ii) If it isn't stationary convert it to stationary using 1st order differentials. Perform *Augmented Dickey Fuller test* to verify that it is indeed stationary.

Question 3 (5 points)

- a) What is the major problem with ARIMA? How does SARIMA help overcome the problem? Get the best parameter for the SARIMA model using AIC metric.
(Hint: Use `statsmodels.tsa.statespace.SARIMAX()`) **(3 points)**
Note: Use the insights from previous ACF and PACF plots to fill up model parameters for SARIMA. Other unknown parameters can be found out through *grid search*.
- b) Use the best parameter obtained to forecast the values for the testing dataset given. Plot the forecasted and actual values together in the same graph. Did the model perform as per expectations, why / why not? Report the testing RMSE of the model. **(2 points)**

Question 4 (3 points)

- a) How would you make use of the other features in the dataset to predict the weather on a particular day in Delhi? What new features would you engineer? **(2 points)**
 - b) LSTM (Long Short term memory network) are popularly used to forecast time series data. Why do you think they are used? **(1 point)**
-