Time Series Analysis

CLIMATE CHANGE: EARTH SURFACE TEMPERATURE DATA

Some say climate change is the biggest threat of our age, while others say it's a myth based on dodgy science."

AN ATTEMPT TO STUDY LAND SURFACE DATA THROUGH TIME SERIES ANALYSIS.

PROBLEM DEFINITION:

An attempt to study trends after studying Global Land Surface Temperature Data with the use of Time Series Analysis.

DATA SOURCE

This data was repackaged data from a newer compilation put together by Berkeley Earth, which is affiliated with Lawrence Berkeley National Laboratory.

It is nicely packaged and allows for slicing into interesting subsets (for example by country).

The link to this database can be found here.

EXPLORATORY DATA ANALYSIS

- After giving a cursory look to the data it was found that it was divided into 577,462 rows and 6 columns, namely:
  - Date.
  - Average Temperature.
  - City.
  - Country.
  - Latitude.
  - Longitude.

These data was further downsized and we concentrated our findings on Germany.

- The columns were then downsized and we concentrated on the Date and Average Temperature columns in our data frame.

- We further downsized the data from the year 1970 to the year 2013 as the original data, dated from the year 1743.

- After doing so, our data then was comprised of **525 rows and 1 column** (final size of the data).

- Then the Null values were accounted for. All the null values were replaced using the Pandas ‘pad’ method from the previous data.

MODEL BUILDING

- Model building was then begun by splitting the data between Training Data and Test Data.

- A plot was generated to study trends, etc.

ARIMA MODEL

- We then focused on building an Arima Model on our data using the statsmodel module in Python.
- We attempted to see if the data had any stationarity.
- Here we found the P-value was less than 0.05, this gave us an idea that the data was stationary.
- We then plotted the ACF (Autocorrelation Function) and PACF (Partial Autocorrelation function) plots.
- After this we trained our efforts on the Auto Arima method.

AUTO ARIMA METHOD

- The Auto Arima module using the **pmdarima** library was then attempted.
- We then fit our model and ran predictions for our forecast model and also plotted the same.
- It gave us the Best Model and also plotted the prediction.
- Then we printed our Model Summary which gave us the AIC value (Akaike Information Criteria).
- The AIC basically helped us quantify the goodness of fit and the simplicity of the model into a single statistic.

ARMA MODEL

Then using statsmodel module in Python we attempted to build the ARMA model.

We ran this model on our Germany Data Frame.

We then proceeded to calculate the AIC value as well as we got the plotting done.

MODEL COMPARISON METRICS

MODEL (Comparison Metrics = RMSE) Model RMSE Holt's Winter exponential smoothing. 1.8253 and S-ARIMA 1.9866.

BEST MODEL: Holt's Winter Exponential Smoothing.