CASE STUDY 031

[Python]

Global Temperature Change over Time – PART II

Here are some clues in case you are stuck with the case study:

1. To extract the series from the seasonal decompose, use:

```
dec = seasonal_decompose(ts)

ts_resid = dec.resid

ts_trend = dec.trend

ts_seasonal = dec.seasonal
```

2. To exclude the nulls from the series, you can use this kind of syntax:

```
ts_resid = ts_resid[~dec.resid.isnull()]
```

And you can use the same index to filter all the series, like:

```
ts_seasonal = ts_seasonal[~dec.resid.isnull()]
```

3. It is easy to plot the boxplot with seaborn. To extract the month from the data, the code is:

```
ts_resid.index.month
```

4. To execute the ACF:

Statistical correlation summarizes the strength of the relationship between two variables.

We can calculate the correlation for time series observations with observations with previous time steps, called lags. Because the correlation of the time series observations is calculated with values of the same series at previous times, this is called a serial correlation, or an autocorrelation.

```
from statsmodels.graphics.tsaplots import plot_acf

plot_acf(ts, lags = 40)

pyplot.show()
```

5. To execute the PACF:

A partial autocorrelation is a summary of the relationship between an observation in a time series with observations at prior time steps with the relationships of intervening observations removed.

```
from statsmodels.graphics.tsaplots import plot_pacf

plot_pacf(ts_wo_seasonal, lags=40)

pyplot.show()
```

6. To create the ARIMA model and fit the parameters:

```
from statsmodels.tsa.arima_model import ARIMA

model = ARIMA(ts_wo_seasonal, order=(5, 0, 5))

results_AR = model.fit(trend ='c')
```

What does ARIMA(5, 0, 5) mean?

Specifically for this model, ARIMA(5, 0, 5) means that it you are describing some response variable (Y) by combining a 5th order Auto-Regressive model and a 5th order Moving Average model. A good way to think about it is (AR, I, MA). This makes your model look the following, in simple terms:

Y = (Auto-Regressive Parameters) + (Moving Average Parameters)

The 0 in the between the 5's represents the 'I' part of the model (the Integrative part) and it signifies a model where you're taking the difference between response variable data.

In basic terms, the ACF and PACF give us information about the parameters.

7. To predict, the function is:

```
forecast = results_AR.predict(start = start_date, end = end_date)
```