Working Process of Time Series

When you get the data, you need to do four things: -

(I) Keep only 2 variables (Date / Target): one is the datetime variable, and the other is the target variable that you want to predict.

(II) If the datetime variable is in object data type, then convert it to datetime data type: This is because you will need to apply datetime functions to it, such as fetching the date, month, or year from the variable. You cannot apply these functions to object data type variables. There is a direct function available in Python for this task. The ‘to\_datetime()’ function is used to convert object data type variables to datetime data type variables.

(III) Make datetime variable as index.

(IV) Make non-stationary data into stationary

\* After doing all these steps we perform other things like visualization and EDA.

Methods by which we can identify that data is stationarity or non-stationarity

(I) The first way is **to calculate the mean and variance or standard deviation of the data points.** Then, check the highest and lowest values. If the difference between the highest and lowest values is more than 10%, then the data is stationary.

(II) Ad Fuller Test: The second way is to use the **Augmented Dickey-Fuller test** (**ADF test**). This is a statistical test that can be used to determine if data is stationary. If the p-value of the ADF test is less than 0.05, then the data is stationary.

Methods by which we can convert non-stationarity data into stationarity

(I) Differencing

(II) There are several **transformations** that can be used to make data stationary. These transformations include logarithm transformation, exponential transformation, square root transformation, and cube root transformation. After doing the transformation, we try to make the data stationary.

Steps of EDA

**This is known as Exploratory Data Analysis (EDA) for time series. The process involves these steps:**

1) Begin by visualizing the data using the pandas plot function.

2) Check if the date variable is in the right datetime format. If not, change its data type.

3) Set the date variable as the index.

4) Plot the distribution of original data, such as high, low, close, open prices.

5) Consider plotting all variable distributions in a single chart for a clearer view.

6) If you notice sudden changes in product prices or sales, your role as a data scientist is to find the reasons behind these changes.

7) Also, examine whether there's a pattern of seasonality, where sales are low in the same month of different years. This is usually not a major concern.

8) If there's no clear seasonality, investigate to find the reasons behind sudden variations.

9) Begin your analysis with a broader time span, like five years.

10) Then, delve deeper by focusing on one year's data.

11) Further analyse month by month to gather insights.

12) Resampling rules can aid your analysis:

(I) 'BQS' (Business Quarter Starting)

(II) 'A' (last day of each year)

(III) 'B' (last working day of each year)

(IV) 'M' (data for every month)