**Problem Statement II-Model Building and Evaluation**

Now that you have found out the most profitable market segment, the next challenge is to forecast the sales for the next 6 months (test data) for that market segment. For forecasting this, you need to check which time series model will work the best. So you decide to apply all the techniques in the Smoothing methods and the ARIMA set of methods and decide to find that out. But even before you start applying each of the techniques, remember the flow chart you learnt at the end of Time Series Forecasting - II.

It helped understand which models to choose based on different problems or datasets. You can simply plot the sales for the concerned market segment, observe the plot and decompose the data into the trend, seasonal and residual components. Based on these insights, you can go back to the flow chart taught in the conceptual lectures and conclude which method as per the flow chart should suit best from the smoothing technique and similarly which method will work best for predicting the sales using the ARIMA set of techniques. Obviously this depends on whether you observe any seasonality or not and other factors after you decompose the time series data for that particular market segment.

After checking this, you should apply all the models in the smoothing techniques and ARIMA set of techniques (except ARIMAX and SARIMAX) and forecast the sales for the next 6 months. You will compare their forecast plots and also find their MAPE values and keep comparing the MAPE values by adding them in the same table (this is similar to what the SME used in the notebook used in the lecture).

**IMP NOTE:** Before you proceed to the next step, we would like to highlight an important point that is must to note: In the dataset, the column **Order-Date** which is now the **48 months** data is given in the **DateTime format**. This might give some errors while you apply the modelling techniques. Hence, before you start applying the methods, convert the **Order-Date** column from "**DateTime**" to "**timestamp**" and **then start building the models**. Thus, **you will get a timestamp for each order date now as well**.

**Model building and evaluation**

Although we have explained the steps above, here’s summing up the steps you need to do after you have prepared the dataset for the 21 market segments, calculated CoV and found the most profitable market segment.

1. You can drop the data for the rest 20 market segments and keep only the one for the market segment that you found to be the most consistently profitable.
2. **Perform the train-test split** for that market segment that takes the **42 months data** as the train data and the **6 months data** as the test data.
3. **Decompose the time-series** to get an idea of the trend, seasonality and residual parts of the data (refer to the lecture notes for this).
4. Go to the **flow chart** and see what method in the smoothing technique and the ARIMA set of methods will work here.
5. Now you will simply apply all the below methods and see the hypothesis you formed from the flowchart was correct or not based on the forecast plots and their MAPE values
   1. Simple exponential smoothing
   2. Holt’s exponential smoothing
   3. Holt-Winters’ exponential smoothing - Additive
   4. Holt-Winters’ exponential smoothing - Multiplicative
6. **Check the forecast plot** calculated on the **test**data and also the **MAPE values**for **each** of the above **methods**. (keep adding the MAPE values in a single table to compare them)
7. Conclude the method in the smoothing techniques whose forecast is able to predict the sales closer to the actual values and whose MAPE values is the least among all the methods done above. Check if this matches with what you found out from the flow chart.

Next, you are to apply the ARIMA set of techniques and check the sales forecast there. But here you will have to **use the box-cox transformation** and **differencing** to make the **time-series stationery**. For the box-cox transformation, you should use the value of **lambda = 0**. Note, you **perform the box-cox on the entire dataset of that market segment** and **then train-test split it again** before proceeding to the ARIMA set of methods. (Refer to the notebook used in the lectures).

Also, you need **NOT** plot the PACF and the ACF values and **can directly use p=1, q=1 and d=1**as the forecasts are relatively better for these values. After the above, you can start forecasting the sales for the most profitable market segment using the following:

1. AR model
2. MA model
3. ARMA model
4. ARIMA model
5. SARIMA model
6. **Check the forecast plot** calculated on the **test**data and also the **MAPE**values for each of the above methods. (keep adding the MAPE values in a single table to compare them)
7. **Conclude the method** whose forecast is able to predict the sales closer to the actual values and whose MAPE values is the least among all the methods done in the ARIMA above. Check if this matches what you found from the flow chart.
8. Overall, we want the **best forecasting method** in the **smoothing**technique **as well as**the **ARIMA**set of techniques.