# **Key Steps to follow while modeling Time Series Problems**

## **1.Visualize**

1. Visualize the dependent variable by plotting to see the various features in the data such as Trend, Seasonality and Cyclical components.
2. Before deciding which type of model we want to use, we must decide if we want to work with Univariate Time Series model or a Multivariate Time Series model
   1. Univariate models - ARIMA/SARIMAX, Prophet, Average and Smoothing Models, DeepAR and Supervised Learning models
   2. Multivariate models - VARMAX, LSTM and Supervised Learning models such as XGBoost, Random Forest,etc.

## **2. Data imputing for missing values:**

We can check how correlated the missing value is with the previous day or previous seven days or previous month or previous year. If there is good correlation with previous day sales we can use back or forward fill. We could also use previous years data for filling if there is correlation between previous years data. We could also use mean of previous days to fill the missing value.

Refer youtube link for more info : <https://www.youtube.com/watch?v=GEytNZVjZNU>

Linear interpolation and Spline interpolation rely on the assumption that adjacent observations are similar to one another. These methods do not work well when this assumption is not valid, especially when the presence of strong seasonality. We deseasonalize the data first, and then do interpolation on the data. Once the missing values are imputed, we need to reseasonalize the data.

Refer link below for more info: https://www.kaggle.com/code/juejuewang/handle-missing-values-in-time-series-for-beginners/report

## 

## **3. Data Preprocessing and Testing**

1. Always try to add new columns in the dataset of lags of the target column and check the correlation between them to see if they may help in the modeling.
   1. Add columns like t-1, t-2 and so on to check for its influence on the predictions. This is also a step to be taken to convert a time series problem into a supervised learning model.
2. If you are working with the ARIMA/SARIMAX model, then check for stationarity in the data using the combinations of ADF and KPSS testing.
   1. If the data has stationarity then, continue directly with the data to fit a model.
   2. If the data does not have stationarity, we can make it have stationarity by various methods like detrending and differencing.
3. For Multivariate models like VARMAX, the data has to have stationarity too, which can be checked similarly to above models.

**Visualize the data once more to confirm the data is now similar to white noise**

1. For other Univariate models, we don’t need the data to have stationarity and the data can be used as it is without even preprocessing it as we only have one column of data to work on.
   1. Average and Smoothing models use the preprocessing of the data as part of fitting the model.
   2. The Prophet can also use the data without preprocessing.
2. For Multivariate models, we first check if all the Features (Independent variables) in the data have high correlation with the Label (Dependent variable).
   1. We can plot a correlation plot of all the features with label and select the best features that have high correlation with the label.
   2. We can also use the SelectKBest library to directly find the K-best features to work with.
3. Once we select the features to make a model from, we normalize or standardize the data before working on the model as we will be building multiple parallel time series models and the features will have different orders of magnitude.
   1. For categorical data, we Label Encoding, Ordinal Encoding or One-Hot Encoding to convert the string values into numerical values
   2. For numerical data, we can use StandardScaler, MinMaxScaler, RobustScaler, etc.

## **4.Splitting Data**

1. Once the preprocessing of the data, we split the data into either Training and Testing data or Training, Validation and Testing data depending on the model we are working on.
   1. Since we have chronological data here, we will be splitting the data in the same order and not randomly.
   2. For ARIMA/SARIMAX, VARMAX, Prophet and Smoothing models, we only use Training and Testing data.
   3. For Supervised Learning Models and LSTM we split the data into X(features) and y(label) and then into Training, Validation and Testing data for each X and y.

## **5.Model Selection**

1. Once the splitting is done, we try to find the best hyperparameter values for each model to find a good fitting model.
   1. For ARIMA/SARIMAX and VARMAX, we use the acf and pacf plots to find the maximum lag values and then using that information, we use the **auto\_arima** library to find the best (p,q,P,Q) values to fit a good model.
   2. For a Deep Learning model like LSTM, we try to find the optimum value for the number of nodes in the hidden layers and the best corresponding learning\_rate using the **Keras Tuner.**
   3. For Supervised Learning models, where we use XGBoost and Random Forest, we can use hyperparameters tuners like **GridSearchCV**, **RandomSearchCV** and **HyperOPT** to find the best values of parameters such as learning\_rate, max\_depth, n\_estimators, etc.

## **6.Model Fitting**

1. Once all of the above steps are done, we can finally run and fit the model with the hyperparameter values we found.
   1. Check for the assumptions of the regression model i.e. normality of errors and no autocorrelation between the errors, after the fitting of the model.
2. After checking that the model is a good fit, we can use the forecast/predict function in the model library to forecast the future value.
   1. Plot the predicted vs actual value to see how the model is predicting the data.
   2. Print the various Evaluation metrics to also get a specific value of the accuracy of the model.

## 

## **Special Notes**

* For the Prophet model, we only need to change the Label column to **‘y’** and the date columns to **‘ds’** and then change it to a DateTime object. We can directly fit the model after splitting the data afterwards.
* When working with SARIMAX or VARMAX model for the time series problem, try using exogenous variables in the data, it may increase the accuracy of the final model

# 

# 

# 

# 

# **Key steps in ARIMA/SARIMA modeling**

1. Plot the time series data and observe the trend and seasonality
2. Use statistical tests like ADF and KPSS to check for stationarity
3. If the series is not stationary, use transformations like differencing or detrending. For seasonal data, one may require seasonal differencing in addition to non seasonal differencing.
4. Once the order of differencing is determined, we need to decide on the optimal AR and MA terms.
5. Use the ACF and PACF plots to get a range within which the parameters p,P,q,Q should lie. Use auto.arima and specify the range of these parameters.
6. Check for the assumptions of the regression model i.e. normality of errors and no autocorrelation between the errors.
7. **Scenario I — Both tests conclude stationarity. Therefore, the series is stationary.**
8. **Scenario II — Both tests conclude non stationarity. Therefore, the series is non stationary.**
9. **Scenario III — ADF test concludes stationarity but KPSS test concludes non stationarity. This indicates that one needs to use differencing to make the series stationary. Here, a new series is created where the value of the observation at time t is given by the difference between the actual value at time t and the value at time t-1. Thus,**
10. **Scenario IV — ADF tests concludes non stationarity but KPSS test confirms stationarity. This indicates that one needs to use detrending to make the series stationary. Common methods of detrending include using log transformation, square root transformation of the original series.**

# **References:**

**SMOOTHING Models, ARIMA/SARIMAX and VARMAX**

* <https://towardsdatascience.com/introduction-to-time-series-forecasting-part-1-average-and-smoothing-models-a739d832315>
* <https://towardsdatascience.com/introduction-to-time-series-forecasting-part-2-arima-models-9f47bf0f476b>
* <https://towardsdatascience.com/lets-forecast-your-time-series-using-classical-approaches-f84eb982212c>