1. Feature Engineering

Time series Component:

* Trend 🡪 increasing or decreasing in overall value
* Seasonality 🡪 repeating pattern in a certain cycle
* Level 🡪 average value in the series
* Noise 🡪 random variation in Series

Data Stationarity:

* Constant mean
* Constant variance
* Autocovariance does not depend on time

How to test Stationary?

* Augmented Dicky Fuller Test (ADF)

How does the test work?

* the null hypothesis is the time series possesses a unit root and is non-stationary. So, if the P-Value in ADF test is less than the significance level (0.05), you reject the null hypothesis. (We have to reject the null hypothesis to make it stationary)

How to make our data stationary: (there are 5 ways)

1. Transformation 🡪 Log, Square, Root
2. Smoothing 🡪 weekly avg, monthly avg, rolling avg
3. Differencing 🡪 First Order Differencing
4. Polynomial Fitting 🡪 fitting regression model (not use in our project)
5. Decomposition 🡪 Detrending and Deseasonalizing, okay this have longer explanation

So, every time series data contains trend, seasonality, level, noise. We use seasonal\_decompose() function it’ll decompose our data and return us the seasonal, trend, and residual (level + noise). Since residual is our data that doesn’t have the trend and seasonal anymore, so residual most likely a stationary data and we can use that one.

Extra: decompose model have additive and multiplicative model, if using additive method etherPrice = resid+trend+seasonality, is using multiplicative then etherPrice = resid\*trend\*seasonality. The one that we apply on our project is the additive one.

=====================================

class statsmodels.tsa.statespace.sarimax.SARIMAX(endog, exog=None, order=(1, 0, 0), seasonal\_order=(0, 0, 0, 0), enforce\_stationarity=True, enforce\_invertibility=True)

• Endog: The observed time-series process

• Order: The (p,d,q) order of the model for the number of AR parameters, differences, and MA parameters. d must be an integer indicating the integration order of the process

• Seasonal\_Order: adds periodicity which is an integer giving the periodicity (number of periods in season). Default is no seasonal effect.

• enforce\_stationarity: Whether or not to transform the AR parameters to enforce stationarity in the autoregressive component of the model. Default is True.

• enforce\_invertibility: Whether or not to transform the MA parameters to enforce invertibility in the moving average component of the model. Default is True.

class statsmodels.tsa.arima\_model.ARIMA(endog, order, exog=None)

AIC: Akaike information criterion. Given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models. Thus, AIC provides a means for model selection.

When a statistical model is used to represent the process that generated the data, the representation will almost never be exact; so some information will be lost by using the model to represent the process. AIC estimates the relative amount of information lost by a given model: the less information a model loses, the higher the quality of that model. In estimating the amount of information lost by a model, AIC deals with the trade-off between the goodness of fit of the model and the simplicity of the model. So, AIC deals with both the risk of overfitting and the risk of underfitting.

=

Prophet: parameters:

Growth - no real saturation insight

Holidays - Holidays are periods of time where the days have the same sort of effect each year. E.g people migrate over the festive periods

Changepoints - are the points in your data where there are sudden and abrupt changes in the trend. Automatic changepoint detection in Prophet - By default, Prophet specifies 25 potential changepoints which are uniformly placed in the first 80% of the time series.The number of potential changepoints can be set using the argument n\_changepoints, but this is better tuned by adjusting the regularization. The locations of the signification changepoints can be visualized with:

daily\_seasonality –

changepoint\_prior\_scale - If the trend changes are being overfit (too much flexibility) or underfit (not enough flexibility), you can adjust the strength of the sparse prior using the input argument changepoint\_prior\_scale. By default, this parameter is set to 0.05. Increasing it will make the trend more flexible