Read me: SARIMAX model

For predicting the overall **sales** next year and specify future sales for each category of products we applied the SARIMA model which takes into account the past experience (autoregressive model) and seasonality patterns. The model includes the following parameters:

p - order of the autoregressive part;

d - degree of first differencing involved;

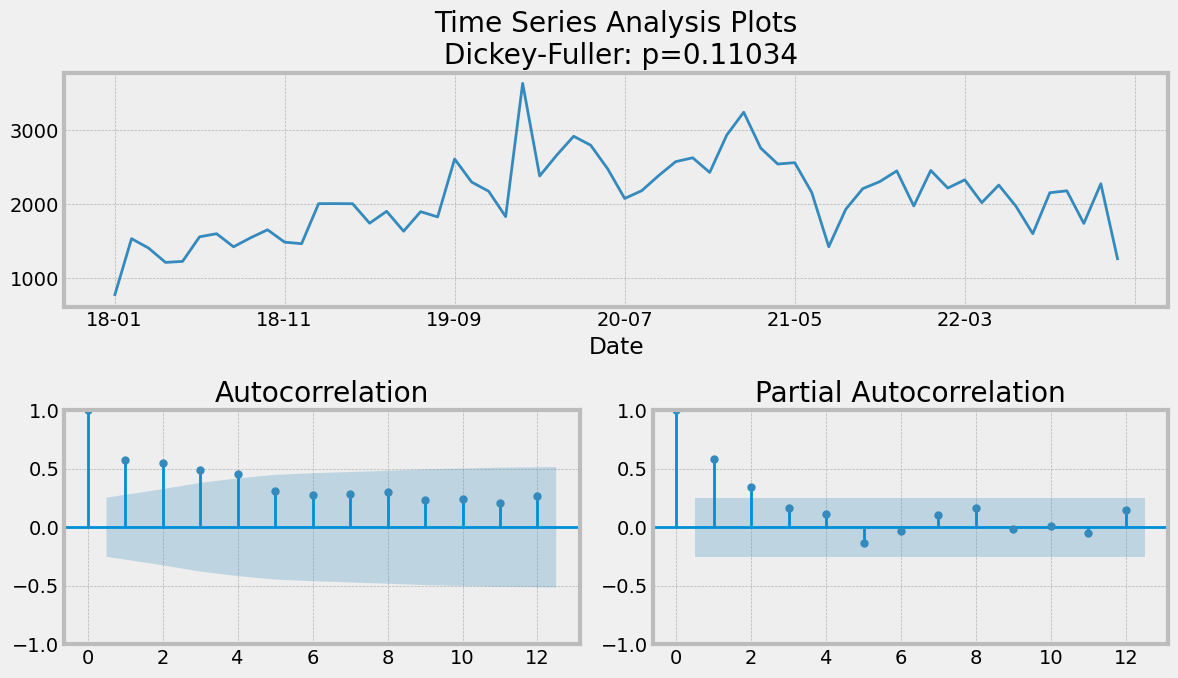
q - order of the moving average part.

P, D, Q - all previous characteristics with seasonal factors

s - seasonal length in the data

We followed the steps below considering the example of overall sales:

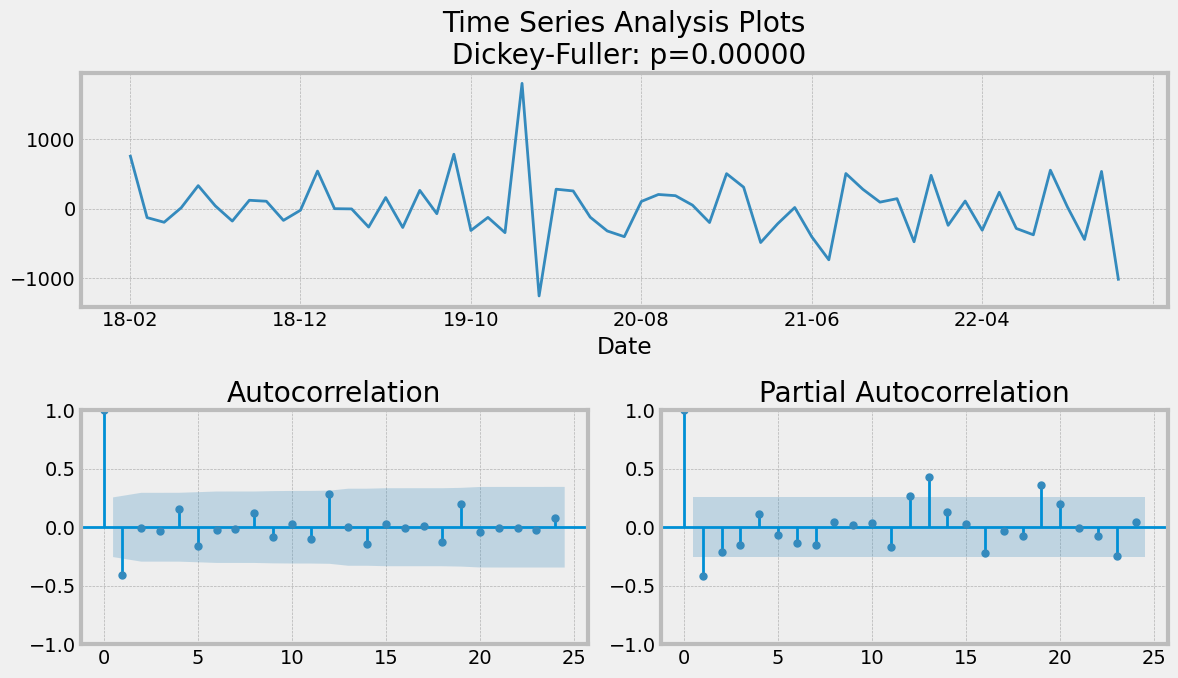
1. Identifying the stationarity and difference of the time series:

Based on the results of Dickey-Fuller test p-value = 0.11 > 0.05, we can suggest that our time series is not stationary. 

Analyzing the ADF chart we can suggest that parameter p = 0 since coefficients of autocorrelation slowly decrease over time lags, so autocorrelation is not significant.

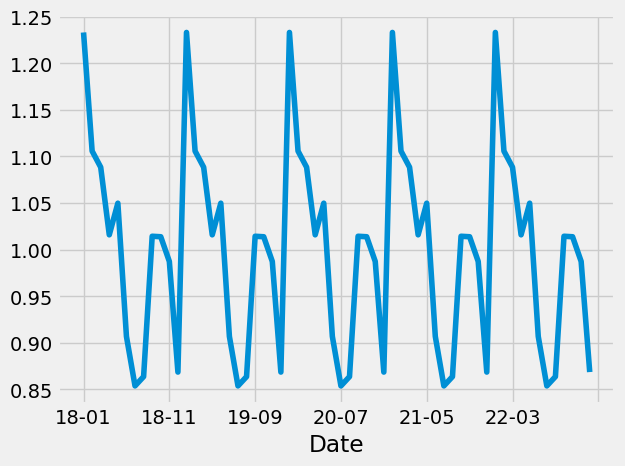
1. Suggesting the initial parameters:

Since our series are not stationary, for further analysis we need to find the difference series with lag=1: *y\_dif(t+1)=y(t+1)- y(t)* . The p-value = 0.00 of Dickey-Fuller test, so it is stationary. We can suggest the other parameters: q=1 (the first lag has a spike).



We can see little spikes at lag 12 and 24 on ACF, so P can be equal to 1.

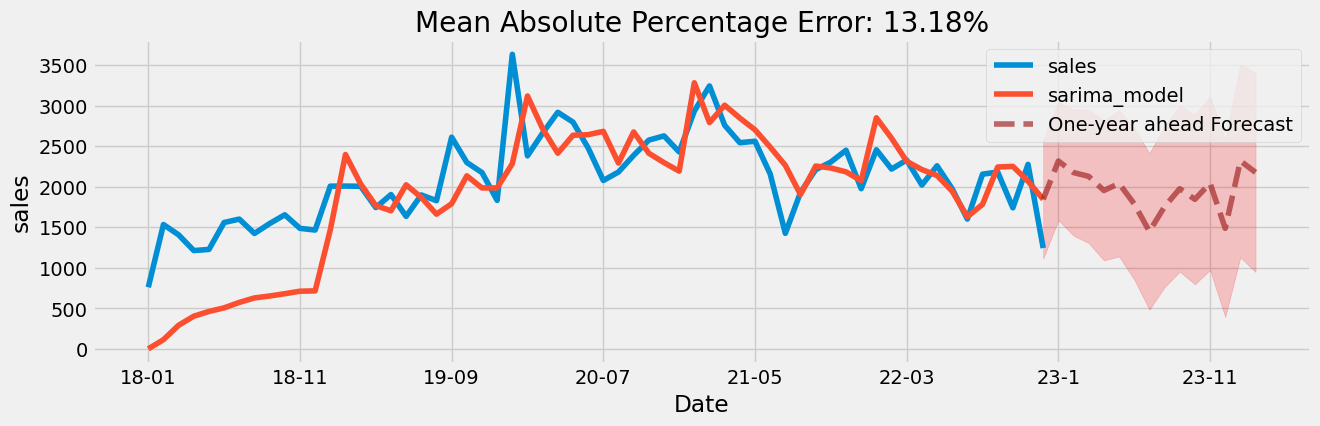
By decomposition of the time series we can see the seasonal trend, so s =12:



1. Generating the final parameters for the model:

In order to find the best parameters for our model, we generated different series of parameters (p, d, q, P, D, Q) with the lowest AIC(Akaike’s Information Criterion).

1. Finally, we run the SARIMAX model:



Analysis of residuals shows, that the model fits well:

