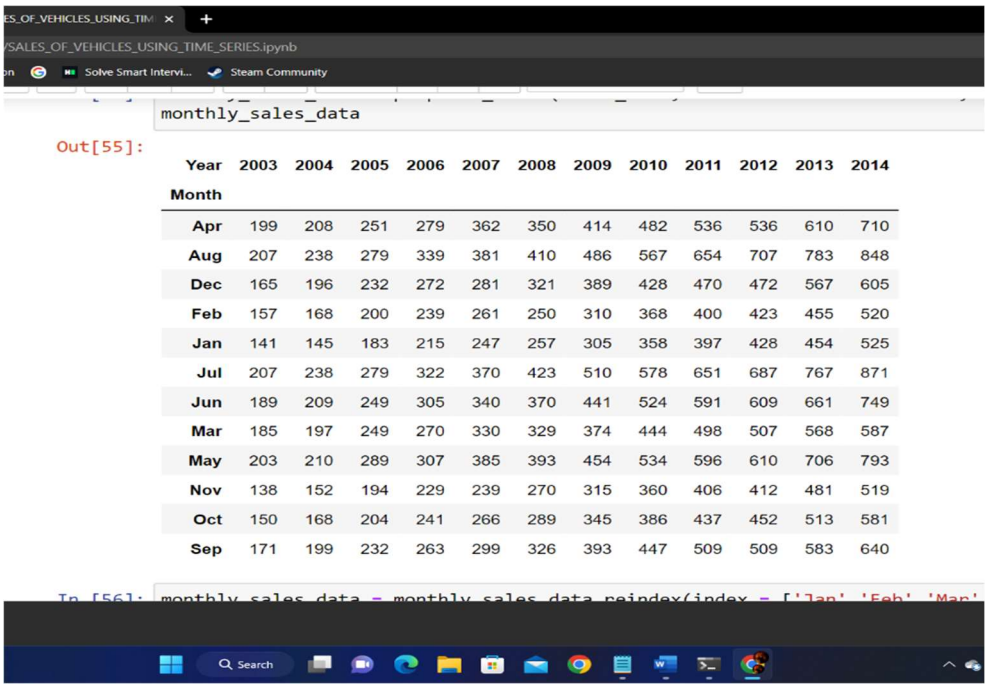
SYSTEM RESULT

The performance of ARIMA is checked by plotting the graph for the confusion matrix, which is generated for the temperature against years. A confusion matrix can help visualize the results of a ARIMA classification algorithm. In this project the implementation is done in the programming language R. Here, we can show that results by compare to other models ARIMA can perform better.

ORIGINAL DATA

Original data that is available in the excel sheet.

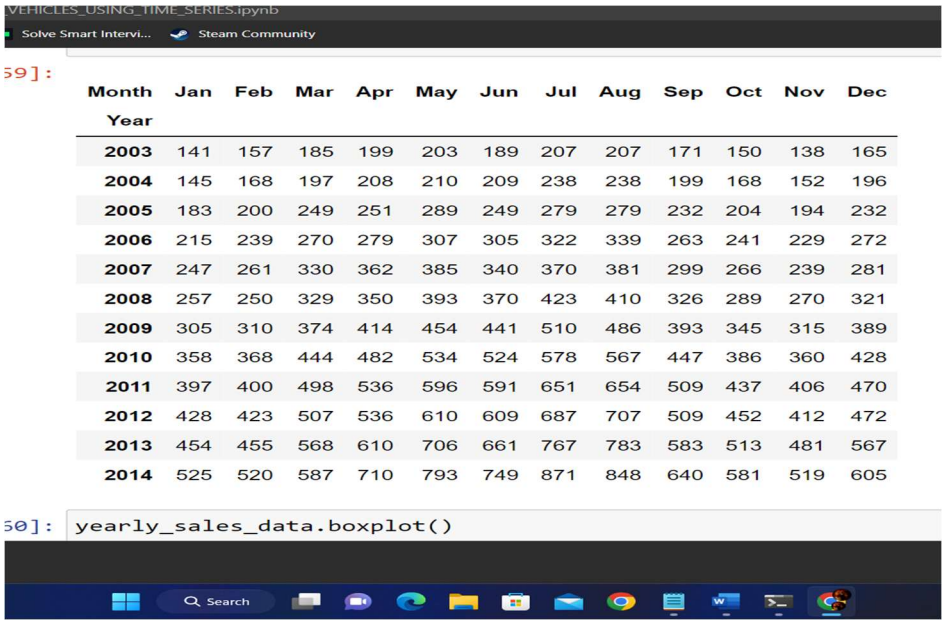


Month	Year 2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Apr	199	208	251	279	362	350	414	482	536	536	610	710
Aug	207	238	279	339	381	410	486	567	654	707	783	848
Dec	165	196	232	272	281	321	389	428	470	472	567	605
Feb	157	168	200	239	261	250	310	368	400	423	455	520
Jan	141	145	183	215	247	257	305	358	397	428	454	525
Jul	207	238	279	322	370	423	510	578	651	687	767	871
Jun	189	209	249	305	340	370	441	524	591	609	661	749
Mar	185	197	249	270	330	329	374	444	498	507	568	587
May	203	210	289	307	385	393	454	534	596	610	706	793
Nov	138	152	194	229	239	270	315	360	406	412	481	519
Oct	150	168	204	241	266	289	345	386	437	452	513	581
Sep	171	199	232	263	299	326	393	447	509	509	583	640

SCREENSHOT 8.1 ORIGINAL DATA

DATA PRE-PROCESSING

Original data is pre-processed before applying algorithms.

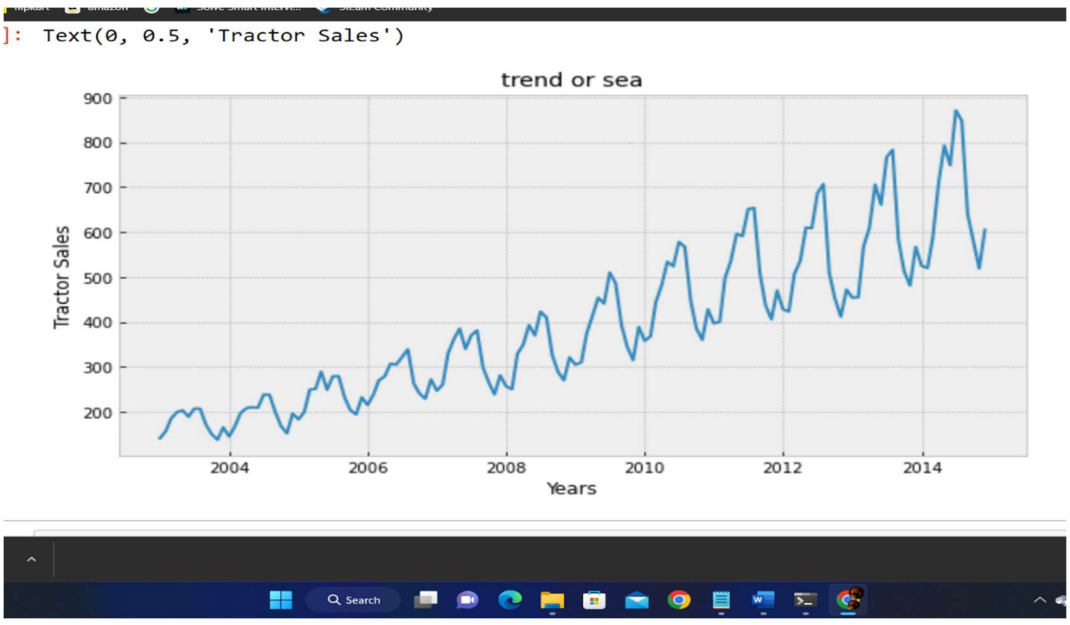


Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Year												
2003	141	157	185	199	203	189	207	207	171	150	138	165
2004	145	168	197	208	210	209	238	238	199	168	152	196
2005	183	200	249	251	289	249	279	279	232	204	194	232
2006	215	239	270	279	307	305	322	339	263	241	229	272
2007	247	261	330	362	385	340	370	381	299	266	239	281
2008	257	250	329	350	393	370	423	410	326	289	270	321
2009	305	310	374	414	454	441	510	486	393	345	315	389
2010	358	368	444	482	534	524	578	567	447	386	360	428
2011	397	400	498	536	596	591	651	654	509	437	406	470
2012	428	423	507	536	610	609	687	707	509	452	412	472
2013	454	455	568	610	706	661	767	783	583	513	481	567
2014	525	520	587	710	793	749	871	848	640	581	519	605

SCREENSHOT 8.2 PRE-PROCESSED DATA

EXTRACTING TIME SERIES

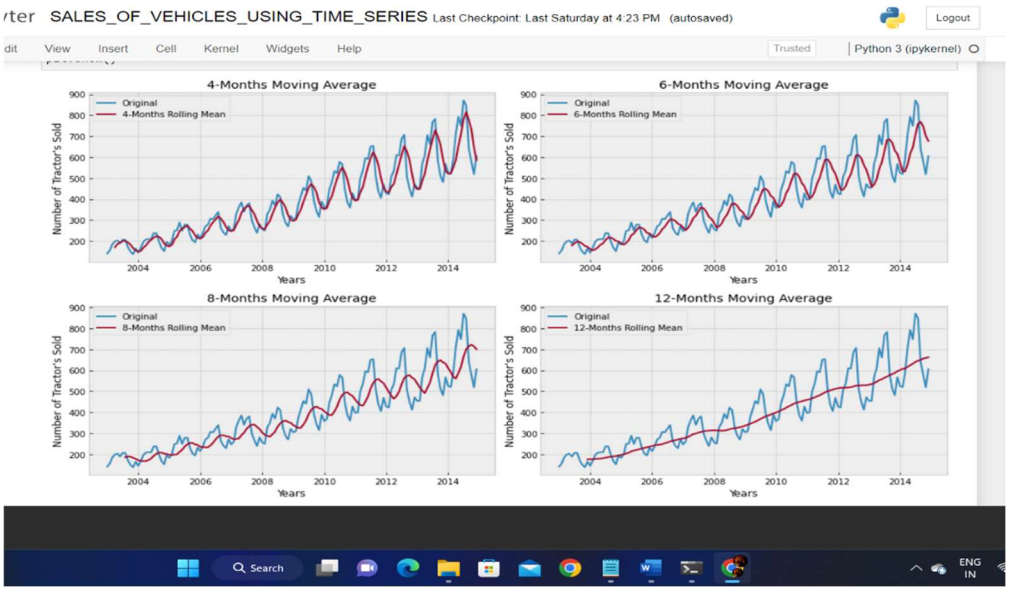
Extracting the time series from the pre-processed data to apply arima.



Screenshot 8.3 Extracting Time Series

REMOVING WRINKLES

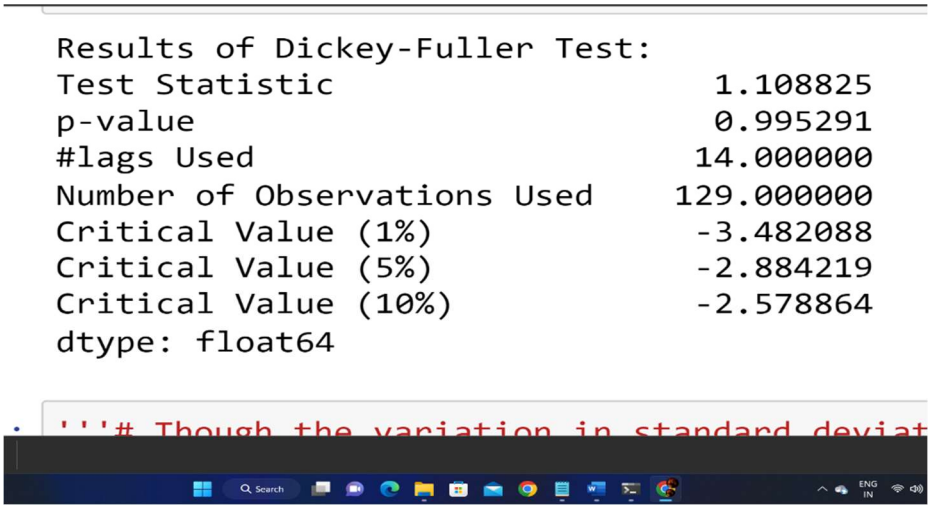
Removing Wrinkles from our time series using moving average.



Screenshot 8.4 Removing Wrinkles

DICKEY-FULLER TEST

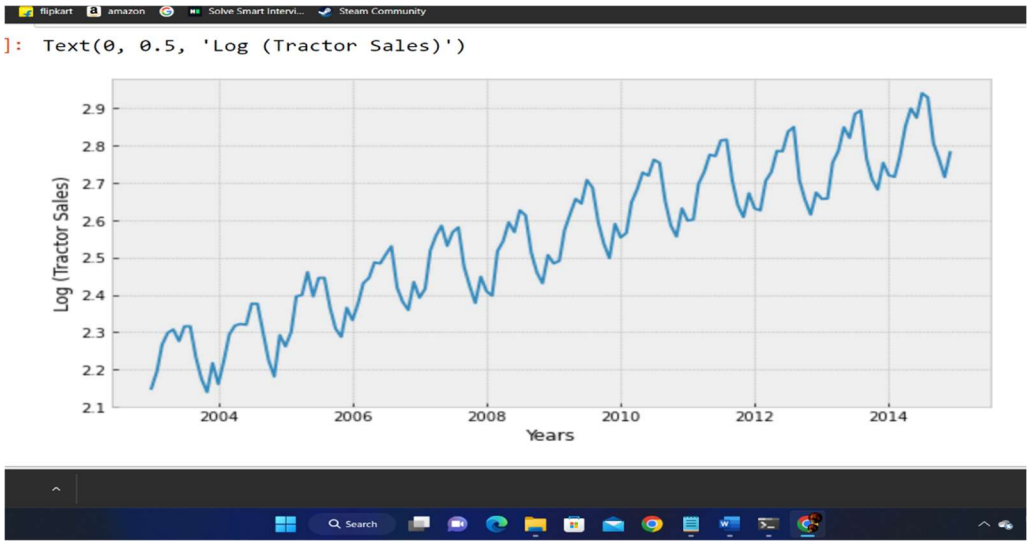
Applying Dickey-Fuller test on our time series to verify the null hypothesis that our timeseries is Non-Stationary.



Screenshot 8.5 Dickey-Fuller Test

APPLYING LOG

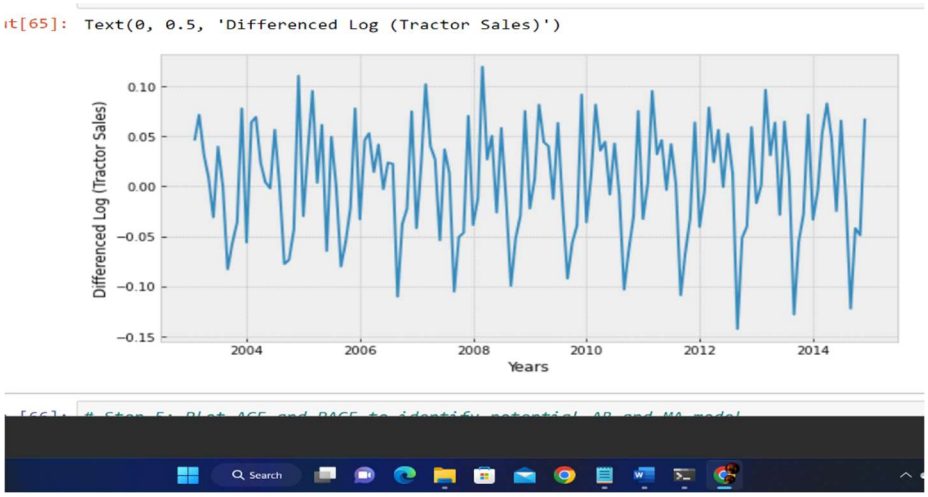
Applying log to make our time series stationary on variance.



Screenshot 8.1.1 Applying Log

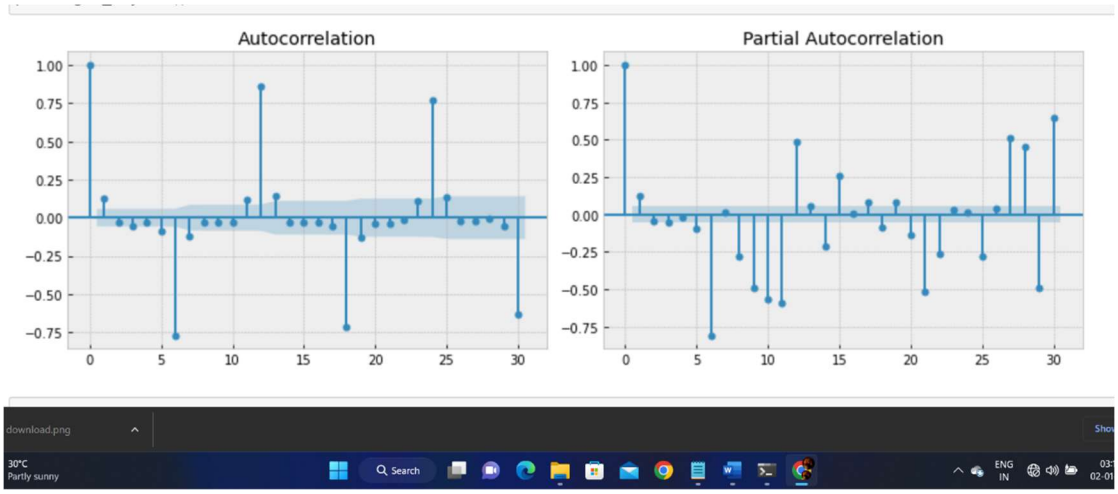
APPLYING DIFFERENCE LOG

Applying difference log to make our time series stationary on both variance and mean.



Screenshot 8.7 Applying Difference log

ACF AND PACF OF ORIGINAL DATA



Screenshot 8.8 ACF and PACF of Original Data

From the two ACF and PACF graphs we have to find the potential ARIMA fit from which we are going to predict the sales forecasting. When choosing the model which is best fit it should be based on AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion). Normally, the model which has lowest AIC value is taken is best fit.

Best Fit ARIMA Model

Predict sales on in-sample date using the best fit ARIMA model

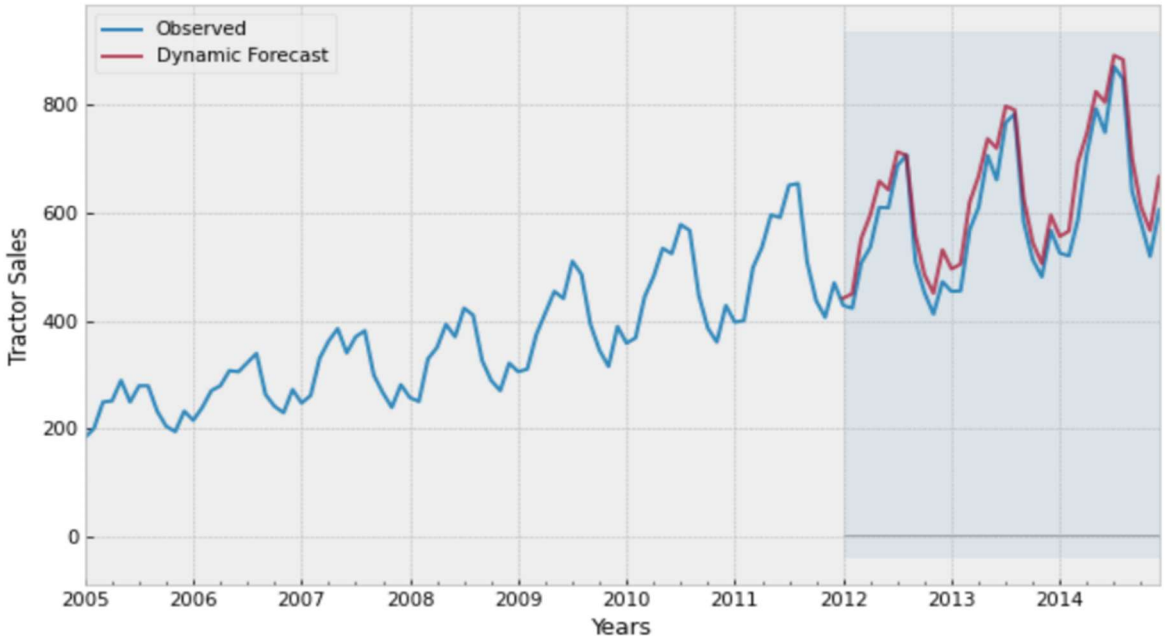
```
                                SARIMAX Results
==========================================================================================
Dep. Variable:                      Tractor-Sales   No. Observations:                  144
Model:             SARIMAX(0, 1, 1)x(1, 0, 1, 12)   Log Likelihood                 370.887
Date:                            Sat, 31 Dec 2022   AIC                           -733.774
Time:                                    16:26:19   BIC                           -721.923
Sample:                                01-01-2003   HQIC                          -728.958
                                     - 12-01-2014
Covariance Type:                              opg
==========================================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.3571      0.069     -5.191      0.000      -0.492      -0.222
ar.S.L12       0.9933      0.006    175.722      0.000       0.982       1.004
ma.S.L12      -0.5524      0.097     -5.723      0.000      -0.742      -0.363
sigma2         0.0003   2.73e-05      9.222      0.000       0.000       0.000
==========================================================================================
```

Screenshot 8.9 Predict sales on sample data

The Mean Squared Error of our forecasts is **0.0011**. By this value we choose best fit model

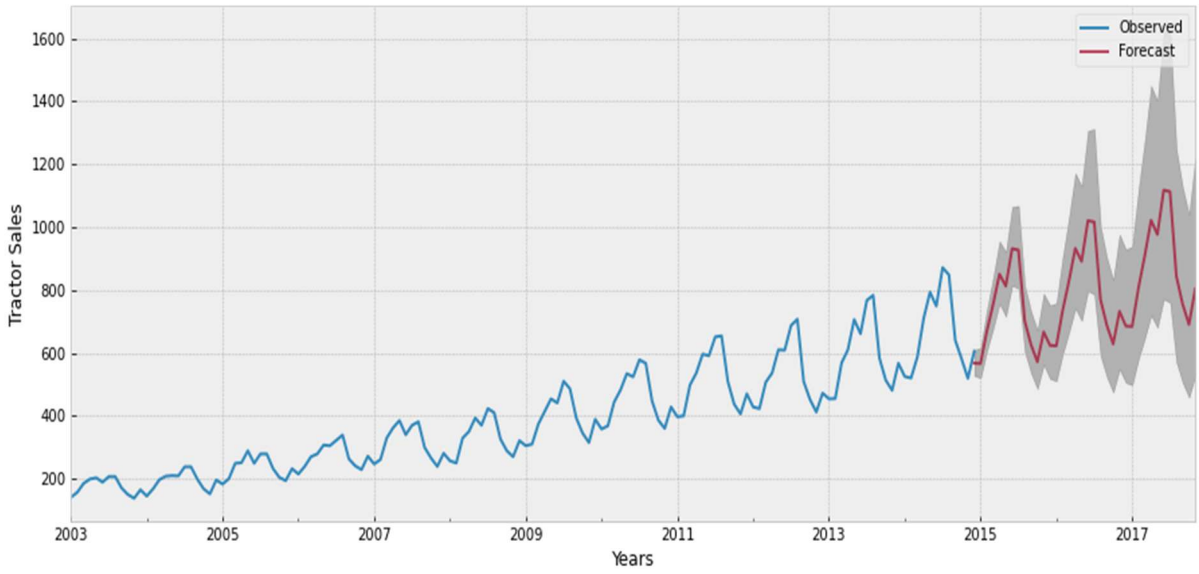
BEST FIT ARIMA MODEL

As expected, our model has I (or integrated) component equal to 1. There is additional differencing of lag 12 in the above best fit model.



Screenshot 8.10 Best Fit ARIMA Model

Sales Predictions (2015-2017)

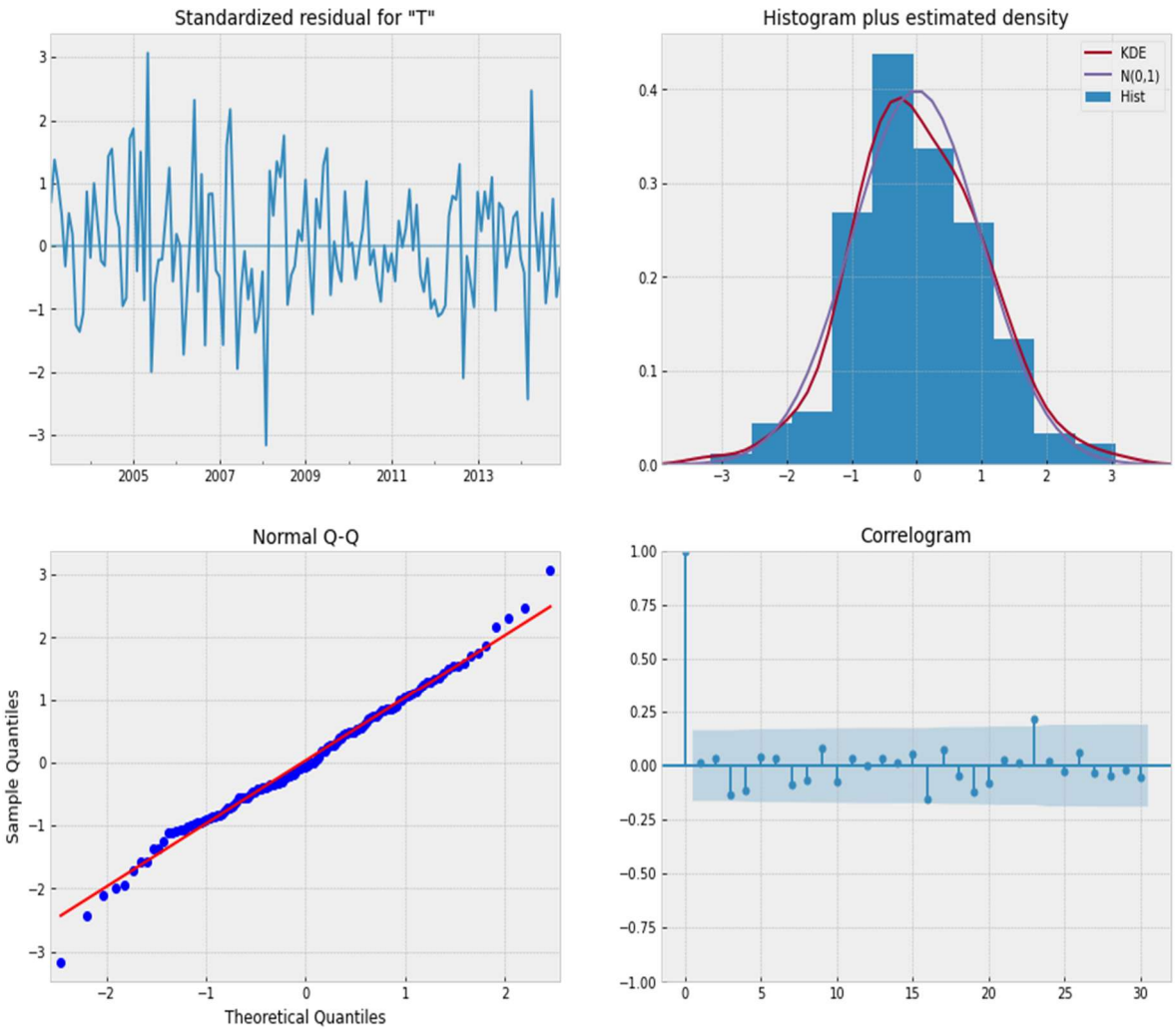


Screenshot 8.11 ARIMA Model Sales Prediction up to 2017

The final graph is plotted for the best fit ARIMA model of number of sales next following years. The following is the output with forecasted values of tractor sales in blue. Also, the range of expected error is displayed with orange lines on either side of predicted blue line.

ACF AND PACF OF PREDICTED SALES

After predicting the number of sales, ensure that there is no more information is left for prediction that is there are no residuals in the ARIMA model.



Screenshot 8.11 ACF and PACF of Predicted Sales