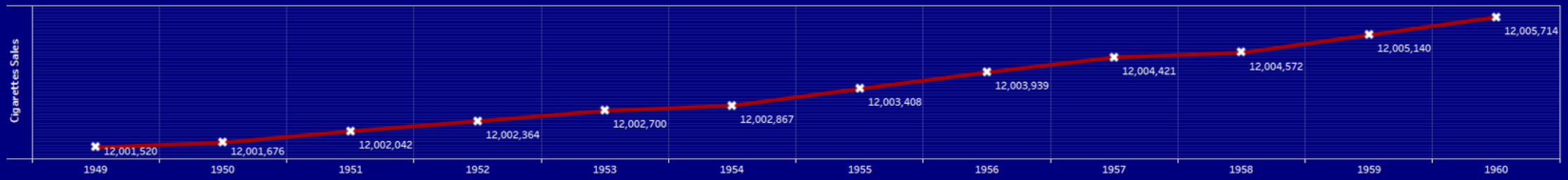
Advanced Time series - Cowboy Cigarettes Sales: Seasonality per Quarter per Month powered by Tableau.

Cowboy Cigarettes Sales per Year

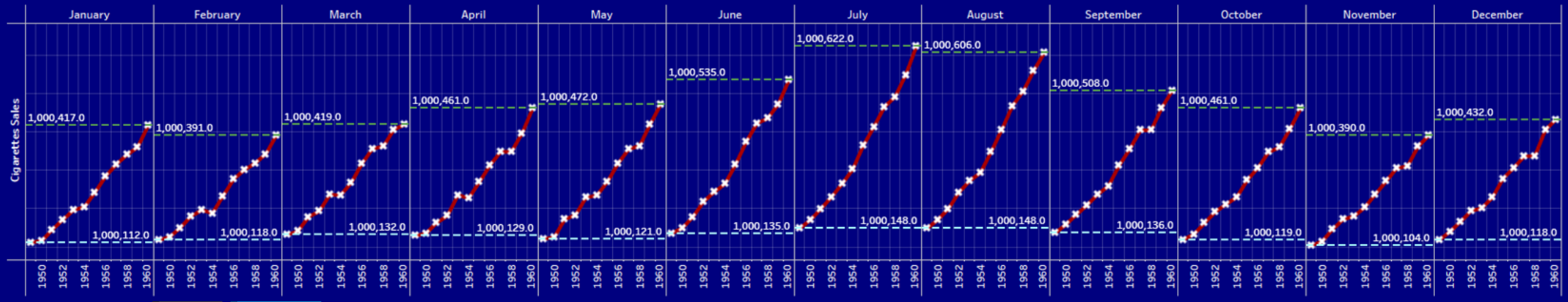


We can see that, generally, there is a trend upwards in cigarette sales from Cowboy Cigarettes.

Cowboy Cigarettes Sales per Year per Quarter (Seasonality)

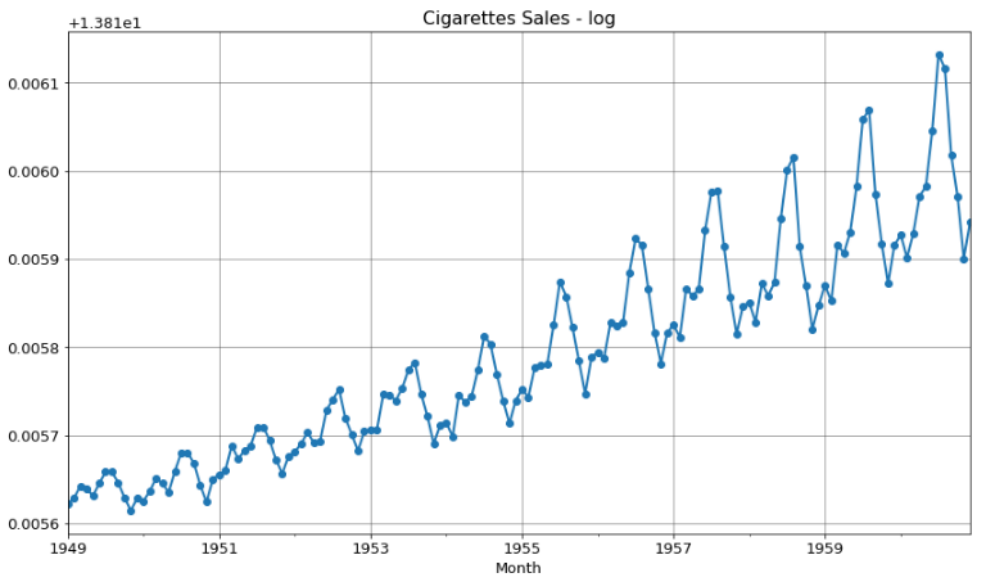
Cigarette sales are always higher in Q3 of every year compared to other quarters between 1949-1960.

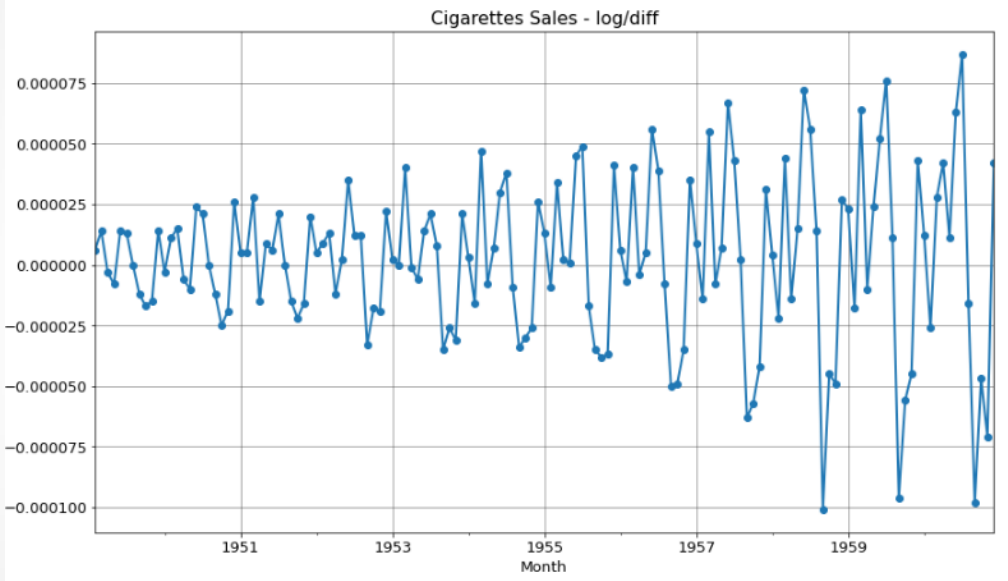
Cowboy Cigarettes Sales per Year per Month (Seasonality)



Cigarette sales are always higher in the Months of July and August of every year compared to other Months between 1949-1960.

Identification – Making Time Series Stationary





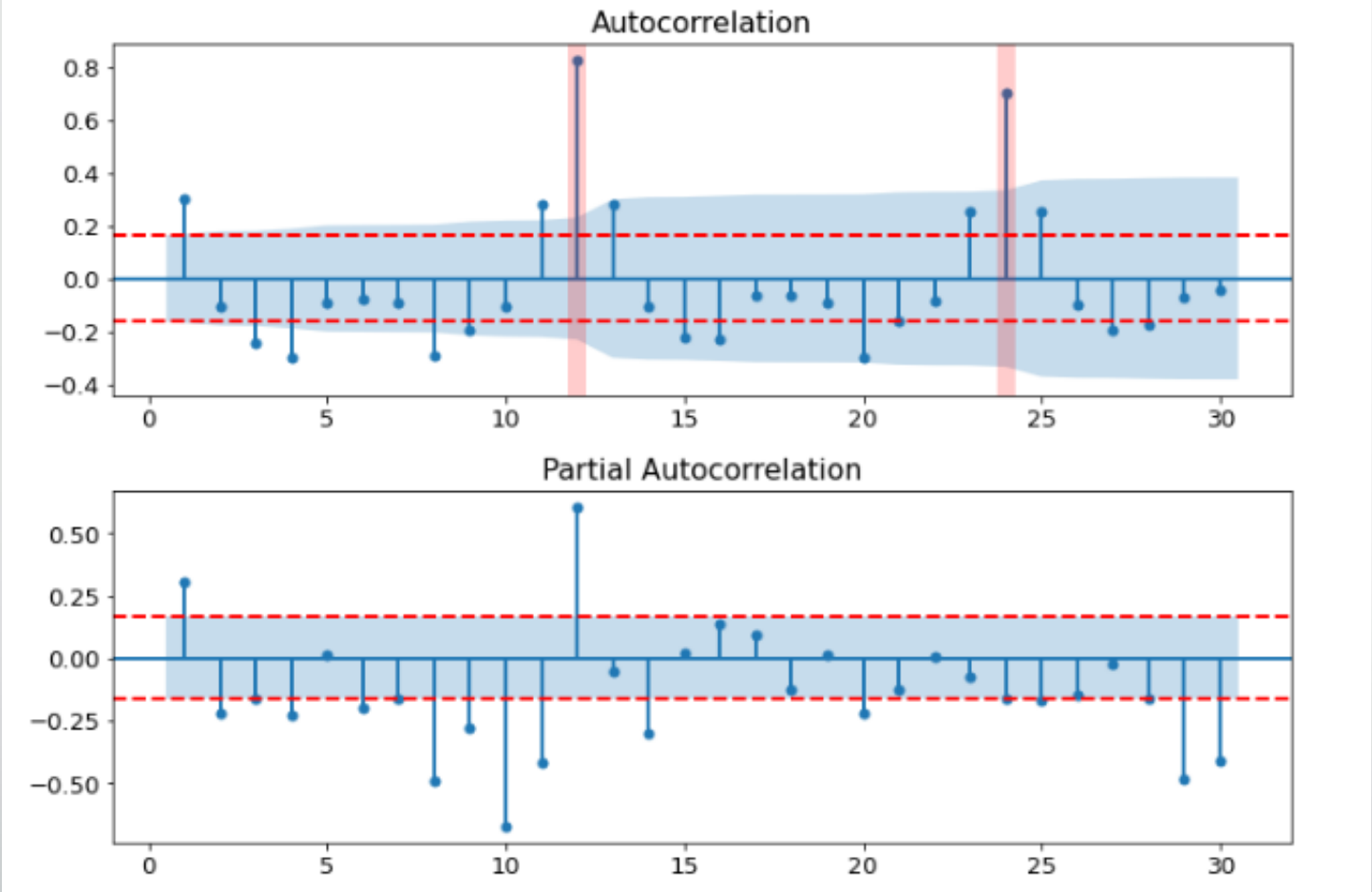

```
*******Results of kpss Test*******
Test Statistic            1.052175
p-value                   0.010000
Lags Used                14.000000
Critical Value (10%)      0.347000
Critical Value (5%)       0.463000
Critical Value (2.5%)     0.574000
Critical Value (1%)       0.739000
dtype: float64
**********************************
```

Since our p-value is less than 0.05, we *should reject the Null hypothesis* and deduce the non-stationarity of our data.

```
*******Results of kpss Test*******
Test Statistic            0.053011
p-value                   0.100000
Lags Used                14.000000
Critical Value (10%)      0.347000
Critical Value (5%)       0.463000
Critical Value (2.5%)     0.574000
Critical Value (1%)       0.739000
dtype: float64
**********************************
```

Our p-value is now greater than 0.05, so we can *accept the null hypothesis* that our data is stationary.

Identification – *check what values of (p,d,q)(P,D,Q)s are most promising?*



In this plot, the two red dotted lines on either sides of 0 are representing the confidence intervals. These can be used to determine the 'p' and 'q' values as:

(1) q – The lag value where the ACF chart crosses the upper confidence interval for the first time. If you notice closely, in this case q=1.
(2) p – The lag value where the PACF chart crosses the upper confidence interval for the first time. If you notice closely, in this case p=1 or 2.
(3) S - in SARIMA is 12 as shown from ACF plot

Identification – *Parameters Selection for the SARIMA Model (Grid Search using AIC & BIC)*

Top 10 performance based in AIC

	param	param_seasonal	aic	bic
57	(2, 1, 1)	(1, 1, 1, 12)	1022.26	1039.51
56	(2, 1, 1)	(1, 1, 0, 12)	1023.24	1037.62
60	(2, 1, 1)	(2, 1, 0, 12)	1024.84	1042.09
16	(1, 1, 0)	(1, 1, 0, 12)	1025.29	1033.92
61	(2, 1, 1)	(2, 1, 1, 12)	1025.65	1045.78
20	(1, 1, 0)	(2, 1, 0, 12)	1026.8	1038.3
17	(1, 1, 0)	(1, 1, 1, 12)	1026.98	1038.48
21	(1, 1, 0)	(2, 1, 1, 12)	1027.23	1041.61
24	(1, 1, 1)	(1, 1, 0, 12)	1027.28	1038.78
48	(2, 1, 0)	(1, 1, 0, 12)	1027.29	1038.79

The output of our code suggests that SARIMAX(2, 1, 1)x(1, 1, 1, 12) yields the lowest AIC value of 1022.26 (This confirmed what we predicted from the ACF and PACF. We should therefore consider this to be optimal option out of all the models we have considered.

Model Diagnostics - *Summary Statistics*

(1) Summary Statistics:
Some of these plots also have accompanying test statistics in results.summary() tables. Prob(Q) is the p-value associated with the null hypothesis that the residuals have no correlation structure. Prob(JB) is the p-value associated with the null hypothesis that the residuals are Gaussian normally distributed. If either p-value is less than 0.05 we reject that hypothesis.

•**Prob(Q):** p-value for null hypothesis that residuals are uncorrelated. Reject the null hypothesis and the residuals are correlated.
•**Prob(JB):** p-value for null hypothesis that residuals are normally distributed. for our Model Prob(JB) is Zero , which means you should reject the null hypothesis that the residuals are normally distributed (This time the JB value was thrown off by the one outlying point in the time series).

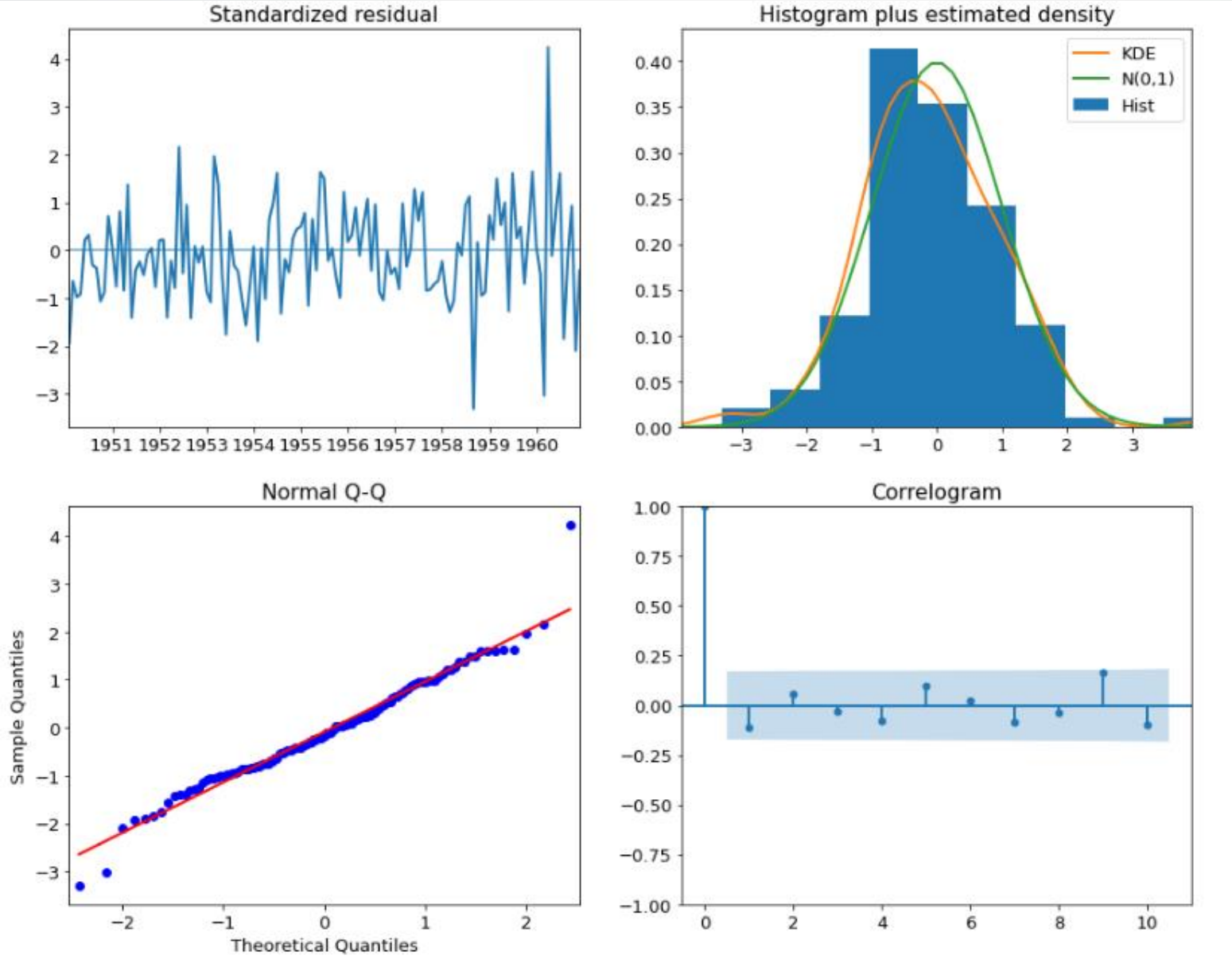
SARIMAX Results

Dep. Variable:	CigSales	No. Observations:	144
Model:	SARIMAX(2, 1, 1)x(1, 1, 1, 12)	Log Likelihood	-505.099
Date:	Sun, 07 Mar 2021	AIC	1024.198
Time:	04:24:51	BIC	1044.325
Sample:	01-01-1949 - 12-01-1960	HQIC	1032.377
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
drift	0.0003	0.001	0.352	0.725	-0.001	0.002
ar.L1	0.7000	0.065	10.753	0.000	0.572	0.828
ar.L2	0.1185	0.062	1.910	0.056	-0.003	0.240
ma.L1	-0.9953	0.167	-5.958	0.000	-1.323	-0.668
ar.S.L12	-0.9973	0.287	-3.471	0.001	-1.561	-0.434
ma.S.L12	0.9835	0.913	1.077	0.281	-0.806	2.773
sigma2	111.5275	65.648	1.699	0.089	-17.139	240.194

Ljung-Box (Q):	45.31	Jarque-Bera (JB):	17.81
Prob(Q):	0.26	Prob(JB):	0.00
Heteroskedasticity (H):	2.02	Skew:	0.29
Prob(H) (two-sided):	0.02	Kurtosis:	4.71

Model Diagnostics – *Plot diagnostics*



(2) Plot diagnostics

Application - Forecasting