FORECASTING

Individual Project Report
By Inderpreet Singh Rana

30 April 2022

IESEG School of Management, Lille

Contents

Project Overview:

The objective of this project is to apply the forecasting techniques learned in class to a real-world dataset.

In this project we experiment with the following models, a short description of the methods used is outlined below:

Naïve Method: Estimating technique in which the last period's actuals are used as this period's forecast, without adjusting them or attempting to establish causal factors

Seasonal Naïve: Based on naïve method which takes the seasonality aspect into account as well.

Holt-Winters: Holt-Winters forecasting is a way to model and predict the behavior of a sequence of values over time (a time series). Holt-Winters is one of the most popular forecasting techniques for time series forecasting.

STL/STLF: STL is a versatile and robust method for decomposing time series. STL is an acronym for "Seasonal and Trend decomposition using Loess", while Loess is a method for estimating nonlinear relationships. The STL method was developed by R. B. Cleveland, Cleveland, McRae, & Terpenning (1990).

ETS: Its an exponential smoothing method of time series forecasting.
Exponential smoothing is a rule of thumb technique for smoothing time series data using the exponential window function. Whereas in the simple moving average the past observations are weighted equally, exponential functions are used to assign exponentially decreasing weights over time

ARIMA: An autoregressive integrated moving average, or ARIMA, is a statistical analysis model that uses time series data to either better understand the data set or to predict future trends. A statistical model is autoregressive if it predicts future values based on past values.

Evaluation Metrics:

MASE: In statistics, the mean absolute scaled error (MASE) is a measure of the accuracy of forecasts. It is the mean absolute error of the forecast values, divided by the mean absolute error of the in-sample one-step naive forecast.

MAPE: The mean absolute percentage error (MAPE), also known as mean absolute percentage deviation (MAPD), is a measure of prediction accuracy of a forecasting method in statistics. It usually expresses the accuracy as a ratio defined by the formula:

Ljung–Box test: The Ljung–Box test (named for Greta M. Ljung and George E. P. Box) is a type of statistical test of whether any of a group of autocorrelations of a time series are different from zero. Instead of testing randomness at each distinct lag, it tests the "overall" randomness based on several lags and is therefore a portmanteau test.

ACF: In the analysis of data, a correlogram is a chart of correlation statistics. For example, in time series analysis, a plot of the sample autocorrelations, versus, (the time lags) is an auto correlogram. If cross-correlation is plotted, the result is called a cross-correlogram.

1) Forecasting for Belgium Airlines Passengers

1) Data Overview:

Forecasting of International intra-EU air passenger transport by reporting country (BE) and EU partner country.

Data Period: JAN 2003 – JUL 2021
Data Frequency: Monthly

Data Split:
Training set: January 2003 up to December 2017
Test set: January 2018 up to February 2020

1.1) Data Exploration:

We can see from the initial time series plot of the raw data that there is seasonality and an upwards trend. Important to note that there is a huge dip in 2020 due to COVID-19.

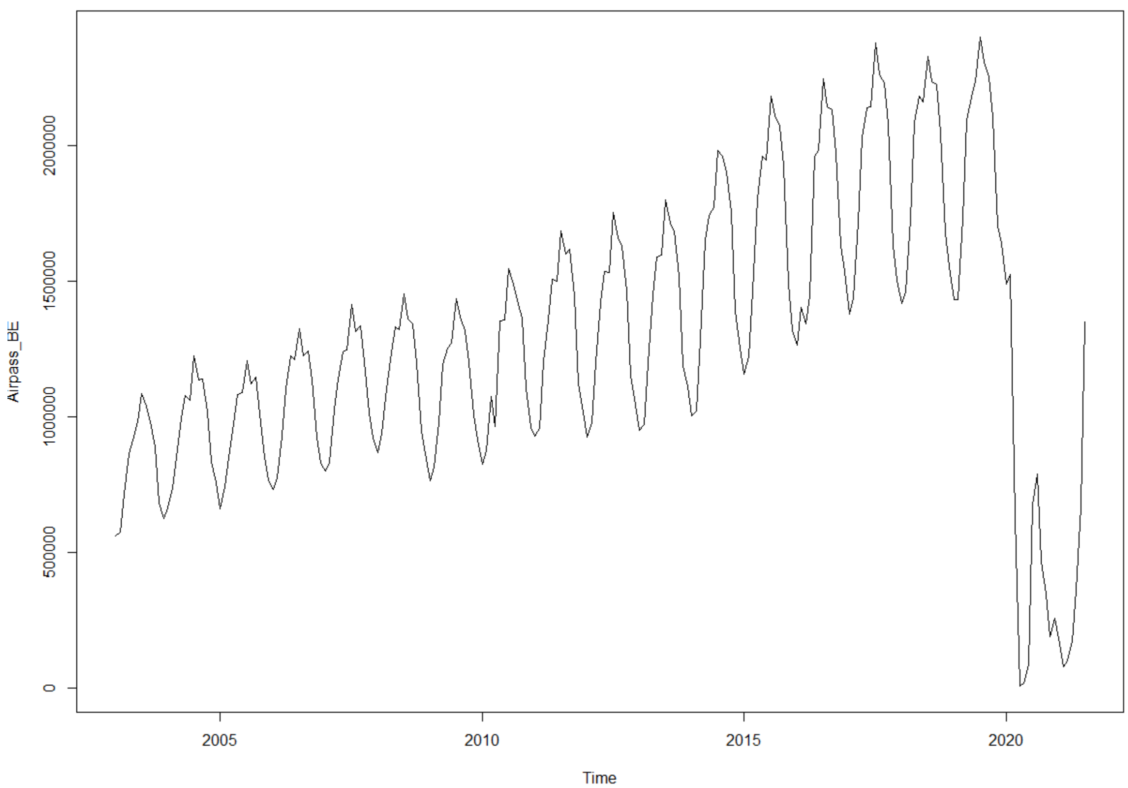

Fig 1.1.1

For further exploratory analysis, we will remove the observations after February 2020 (the Covid period). The seasonality and trend look constant in our data at the first glance.

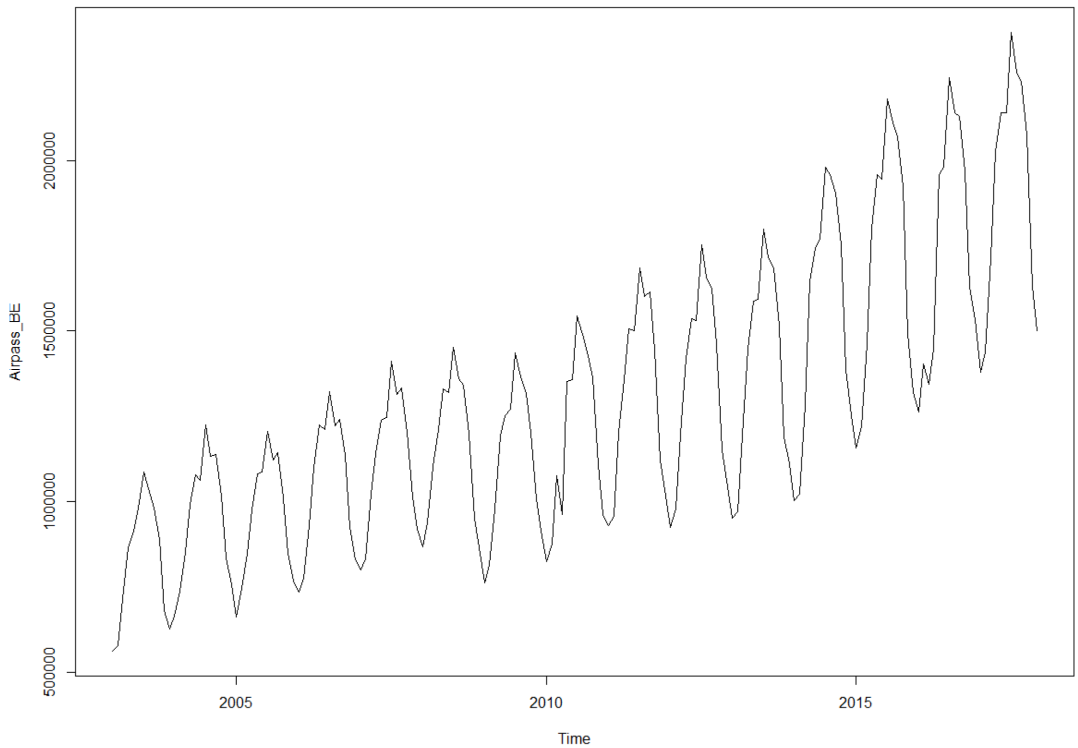

Fig 1.1.2

The Autocorrelation and Partial Autocorrelation confirm that our data indeed has seasonality and some upwards trend

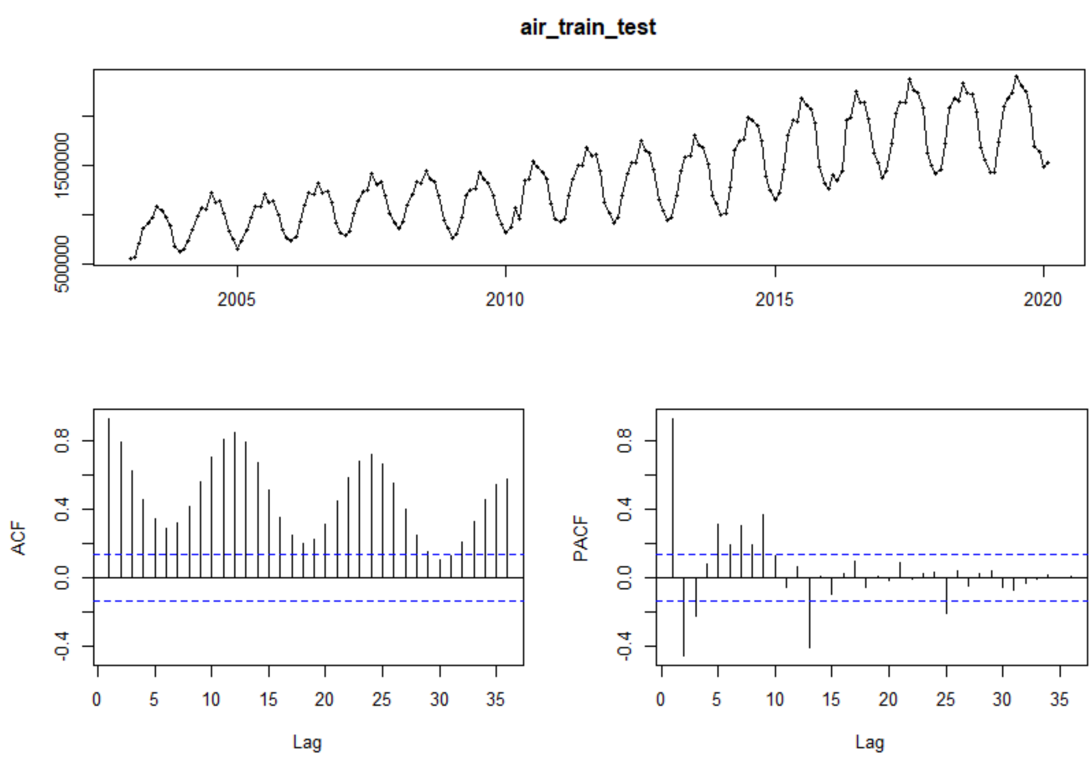

Fig 1.1.3

Looking at the following figure, we notice that the seasonality is at its peak during the summer, which could be explained due to the fact a lot of people travel for vacations during these months.

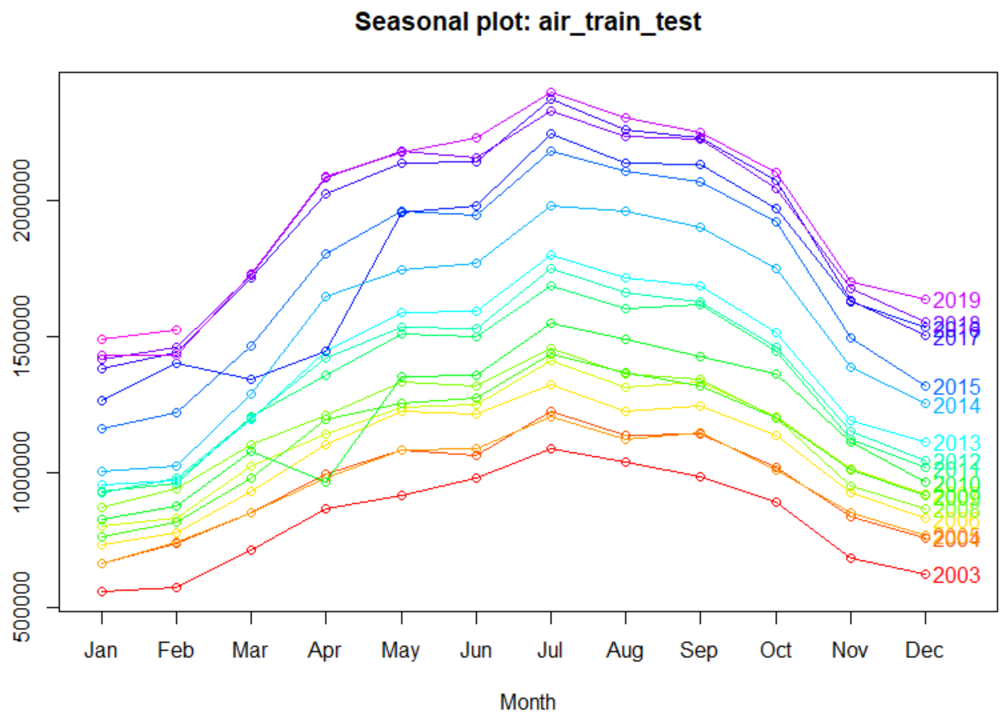

Fig 1.1.4

1.2) Data Transformation:

We use the Box-Cox function to find the optimal lambda, which in the case of our data is 0.014 which is low and close to 0, applying the box cox transformation doesn't have much impact on our time series, it reduces the trend to some extent, the same can be said for log transformation.

The following two figures compare the time series before and after the log transformation.
Based on this, it could be said that the model performance will be similar with or without log or box-cox transformation.

We will see further the impact of transformation while comparing models.

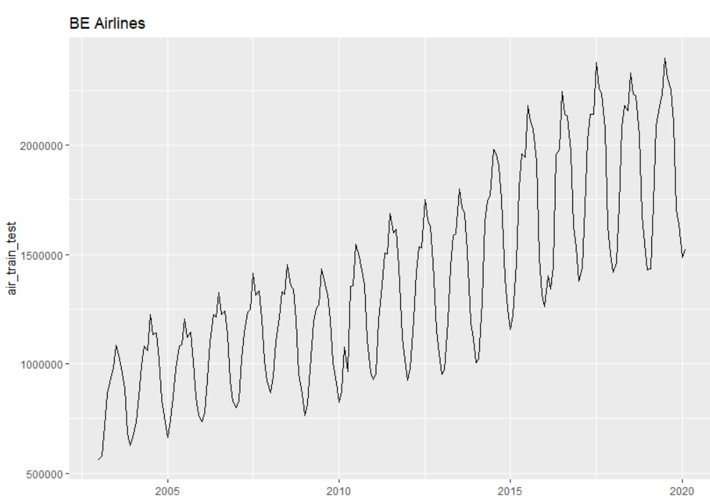

Fig 1.2.1 (Time Series without data transformation

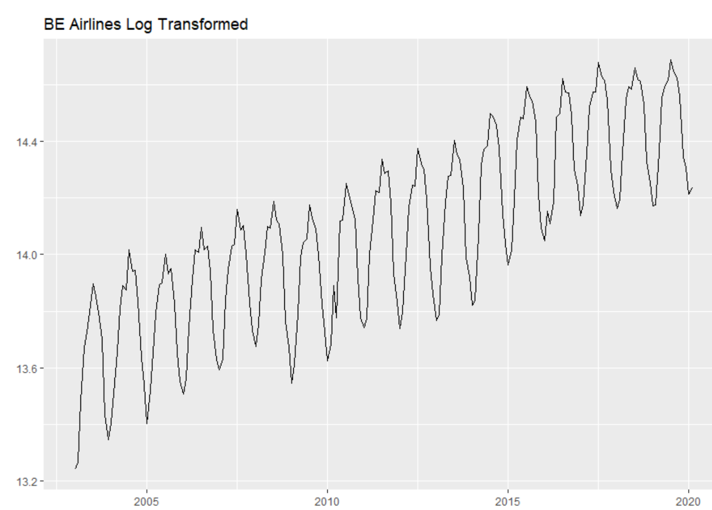

Fig 1.2.2 (Time series after log transformation)

1.3) Benchmark Model:

We first run the seasonal naive model to get a benchmark for our modelling.

We can see from the following metrics that log transformation does have some impact on forecasting performance.

Model	RMSE	MAE	MAPE	MASE
Snaive	55167.45	45156.54	2.518184	0.4752746
Snaive (Log transformed)	46096.78	35825.84	2.017294	0.3770686

Fig 1.3.1

Next, we look at the residuals, ACF & PACF to further evaluate the model.

From the following figure, we can see that the residuals are not good; there could be a better model for our forecast.

We also look at the Ljung-Box test:

- Snaive without transformation: Q* = 207.7, df = 24 and p-value = 2.2e-16
- Snaive with log transformation: Q* = 225.73, df = 24, p-value = 2.2e-16

Based p-value and residuals, we can say that this model is significant, and we can accept the null hypothesis that there is no or very low white noise, however, we do see that there is a room for improvement as based on the ACF plot below, the model does not capture the entire data points, so we continue to experiment to find a better model.

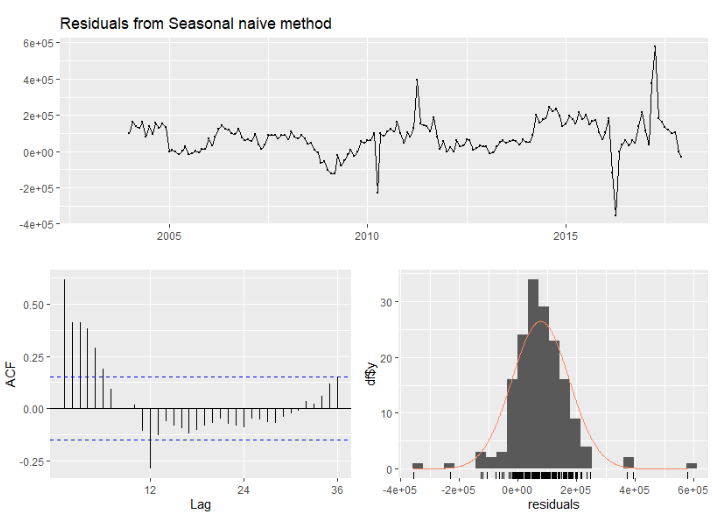

Fig 1.3.2 (Snaive without transformation)

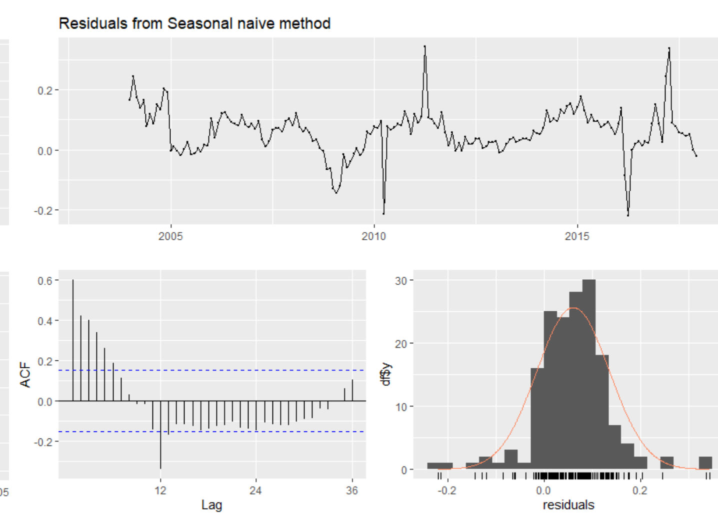

Fig 1.3.3 (Snaive with log transformation)

1.4) Decomposition:

Next, we try seasonal decomposition methods to see if the model performance is better after applying decomposition.

1.4.1) Comparing variations of deposition models

At the first glance of the metrics in the table below, the seasonal naive with lambda still has the best MASE. We will look at the residuals and p-values next.

Model	RMSE	MAE	MAPE	MASE
Seasonal Naive Lambda	46096.78	35825.84	2.017294	0.377069
STL rwdrift	60510.12	41900.51	2.319262	0.441005
Seasonal Naive	55167.45	45156.54	2.518184	0.475275
STL naïve lambda	72108.91	61211.46	3.168164	0.644253
STL arima	75612.79	65012.92	3.697211	0.684264
STL ets	78937.4	67413.63	3.824808	0.709532
STL naive	95954.42	80868.85	4.111746	0.851148
STL Periodic	118389.1	100047.9	4.979355	1.053009
STL arima lambda	155561.4	128532.1	6.350925	1.352806
STL ets lambda	158865.1	131761.4	6.528619	1.386795
STL rwdrift lambda	180895.4	151634.8	7.557488	1.595963

Fig 1.4.1 (Comparing benchmark models with STL models, sorted by MASE)
(Benchmark models in green)

On comparing the Ljung-Box test, we see that the STL model with decomposition and rwdrift is better than our benchmark models, however, the MASE and MAPE do not show us the same picture.

Model	Q*	df	p-value	p-value Eval
STL Periodic	73.4661	23	0	Significant
Seasonal Naive	207.7022	24	0	Significant
Seasonal Naive with Lambda	225.7336	24	0	Significant
STL rwdrift	43.6689	23	0.0058	Significant
STL naive	43.6689	24	0.0083	Significant
STL rwdrift lambda	41.4131	23	0.0106	Significant
STL naive lambda	41.4131	24	0.015	Significant
STL ets	29.743	20	0.0741	Not Significant
STL ets lambda	23.1571	20	0.2811	Not Significant
STL arima	18.9574	21	0.5879	Not Significant
STL arima lambda	18.1386	21	0.6402	Not Significant

Fig 1.4.2 (Comparing the residuals, p-value and degrees of freedom)

On looking at the ACF and PACF plots in the figure below, we see that there is still a lot of room for improvement, so we continue to find a better model.

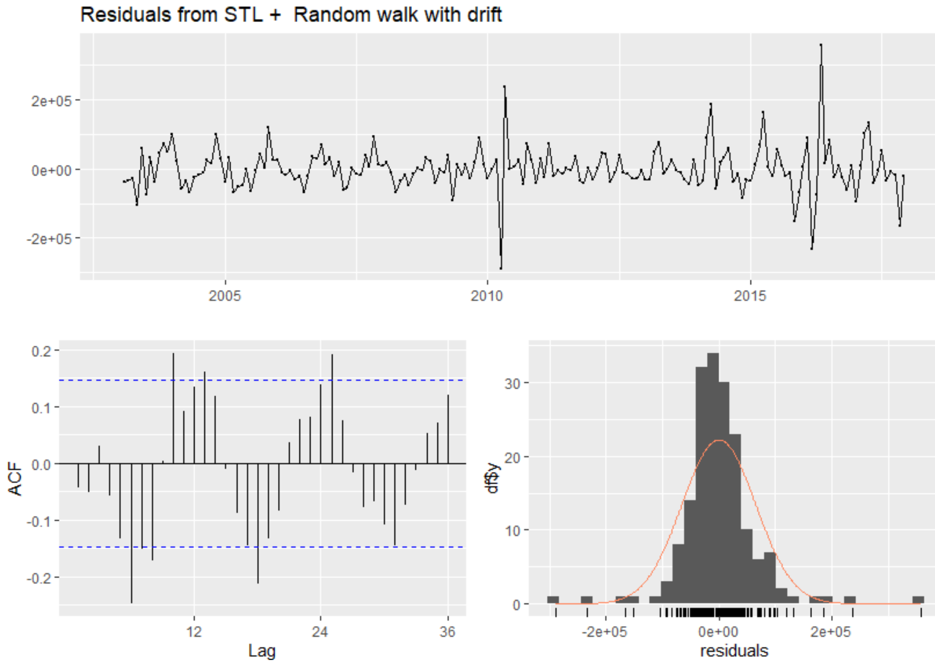

Fig 1.4.3 (Residuals, ACF and PACF for STL with rwdrift)

1.5) ETS & Auto ETS Models:

Next, we look at some of the ETS models, we expect them to perform better due to the way they handle seasonality and trending data.

The following table shows the comparison between our benchmark models and the variations of the ETS models. MASE is still the best for the seasonal naïve model, however upon comparing MAPE, we notice that the difference is quite low, so we will look at the residuals next to check if this model is better than the seasonal naïve and STL models

Model	RMSE	MAE	MAPE	MASE
Seasonal Naive with Lambda	46096.78	35825.84	2.017294	0.377069
Seasonal Naive	55167.45	45156.54	2.518184	0.475275
ETS - AAdA lambda	64270.91	51535.06	2.663379	0.542409
ETS - MAdM	69085.62	56001.95	2.81248	0.589423
ETS - AAA	112266.5	98133.01	5.059773	1.032854
ETS - MAA	112700.7	98303.23	5.059191	1.034646
ETS - AAA lambda	140475.5	115174.6	5.736623	1.212218
ETS - MAM	161295	133042.5	6.526182	1.400278
ETS - MAdA	168084.8	142833.8	6.910386	1.503332
ETS - AAdA	168077	142881.9	6.914275	1.503838

1.5.1 (Comparison of ETS models)

Looking at the residuals and the significance level, we see that all the ETS models are significant, based on the combination of the significance level and the evaluation metrics in table 1.4.4, the ETS model with AAdA & lambda transformation seems to be a good candidate for our data.

Model	Q*	df	p-value	Significance
Seasonal Naive	207.7022	24	0	Significant
Seasonal Naive with Lambda	225.7336	24	0	Significant
ETS - AAA	87.4316	8	0	Significant
ETS - MAA	96.5668	8	0	Significant
ETS - MAM	19.9687	8	0.0105	Significant
ETS - AAdA	86.3902	7	0	Significant
ETS - MAdA	98.9526	7	0	Significant
ETS - MAdM	16.1686	7	0.0236	Significant
ETS - AAA lambda	27.3932	8	0.0006	Significant
ETS - AAdA lambda	14.276	7	0.0465	Significant

1.5.2 (Comparing the residuals and p values for ETS models)

We now look at the residual plot to visualize the ACF, residuals & degrees of freedom, based on the below figure, we can see that the ETS model is the best one we have so far for our dataset.

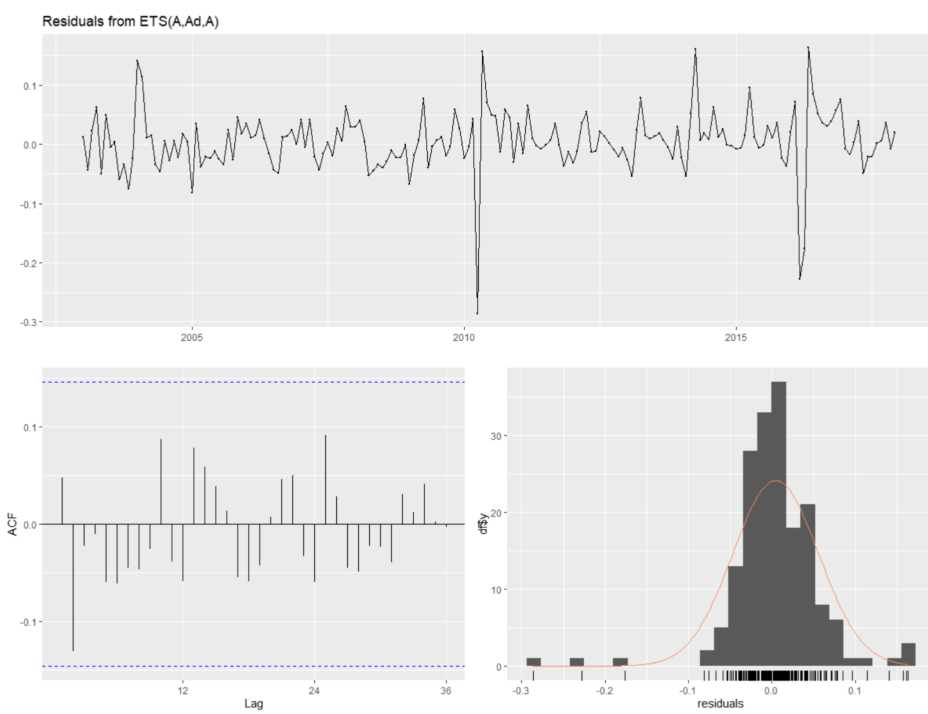

Fig 1.5.3

The parameters of this ETS model are as follows:

model = "AAA", damped = TRUE, lambda = 0.0146, biasadj = TRUE

Smoothing parameters: alpha = 0.5609 , beta = 8e-04 , gamma = 1e-04 , phi = 0.9783

Fig 1.5.4

1.6) ARIMA & Auto-ARIMA Models:

Although we have a model that's performing a lot better than our benchmark model, we will also run ARIMA models to see if they can beat the performance of the ETS model for our dataset.

At the 1st glance, we see that ARIMA does not perform better than our benchmark or ETS models, we will see the residuals and significance level next.

Model	RMSE	MAE	MAPE	MASE
Seasonal Naive with Lambda	46096.78	35825.84	2.017294	0.3770686
Seasonal Naïve	55167.45	45156.54	2.518184	0.4752746
ARIMA	145139.11	127891.13	6.730419	1.3460599
ARIMA dD - Backward Shift	145139.11	127891.13	6.730419	1.3460599
ARIMA dD lambda	146048.82	128704.08	6.567899	1.3546162
ARIMA lambda	162334.63	143075.63	7.311609	1.5058775

Fig 1.6.1 (Performance of ARIMA models compared to the benchmark models)

From the following metrics, we see that none of the ARIMA models we ran is significant, so we proceed to reject the ARIMA models and choose the ETS model shown in fig 1.5.4 on page 9.

Model	Q*	df	p-value	Significance
Seasonal Naive	207.7022	24	0	Significant
Seasonal Naive with Lambda	225.7336	24	0	Significant
ARIMA	18.044	19	0.5195	Not Significant
ARIMA lambda	6.1312	18	0.9956	Not Significant
ARIMA dD	18.044	19	0.5195	Not Significant
ARIMA dD lambda	8.689	20	0.9862	Not Significant

Fig 1.6.2 Comparing the residuals, degrees of freedom and significance level

The results for the auto-ARIMA were very similar to the 3rd model in the table above (ARIMA).

1.7) Forecasting

Now that we have our best model, we proceed to generate a forecast up to December 2022.

To do this, we re-fit the model on the complete dataset (train + test) and then forecast the period up to December 2022.

It is also important to note that our re-fit data does not include the Covid period since it was a rare event, and we will instead look at the forecast to see how things could have been if there was no covid and forecast how the volume of the airlines' passengers is predicted till December 2022.

Forecast Results Interpretation:

- The following plot shows the forecasted number of passengers till December 2022 (in blue)
- The black line is the actual number of passengers from our dataset.
- We can see a huge drop starting in February 2020 due to the impact of COVID-19
- We can also see that if there was no covid, we would have expected the trend and seasonality to continue where there would have been an upward seasonality during summer the overall trend is also upwards.

Legends:

Black → Training data
Green → Test Data
Orange → Covid data
Blue → Forecasted data

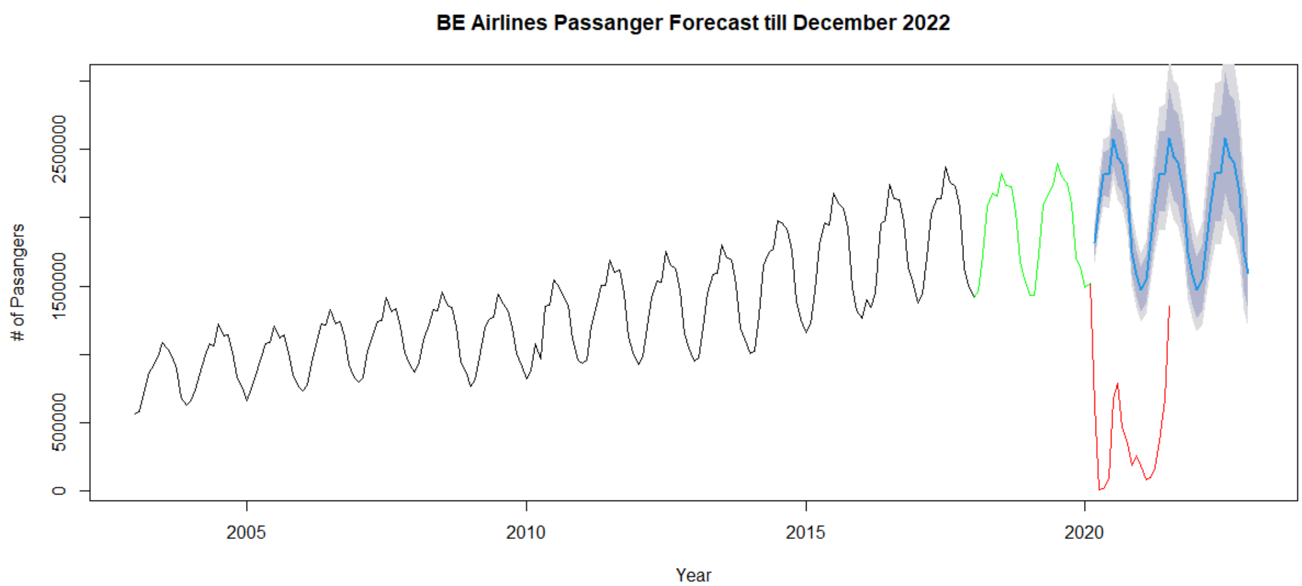

Fig 1.7.1 (Forecast till December 2022)

We can also say based on the residuals, lags, and the degrees of freedom that the model seems to be working well. All values are within an acceptable range in the ACF plot as well.

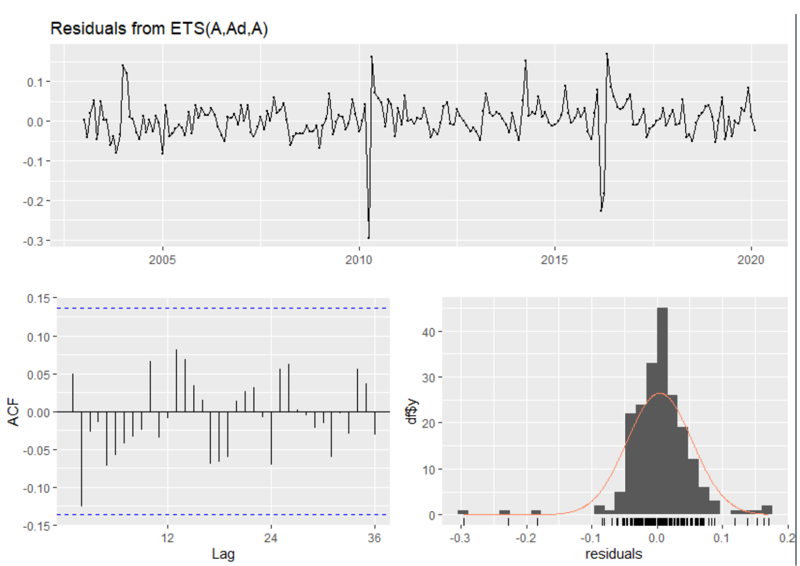

Fig 1.7.2

2) Forecasting for US Electronic Sales Revenue (Net Sales)

2.1) Data Overview

The dataset is sourced from the US census database and contains the Quarterly Financial Reports.
For this project, I have chosen to analyze and forecast the 'Net Sales' for 'Computer and Electronic Products in the US'.

Data Source: https://www.census.gov/econ/currentdata/datasets/index

- **Data Period:** 2001 – 2021
- **Data Frequency:** Quarterly
- **Data Values:** Net sales in USD (Million)

2.2) Data Exploration

We start with looking at the raw time series data to understand the trend, pattern, and seasonality. From the initial view, we can notice some upwards trend and seasonality. We can further confirm the same using the ACF, PACF and seasonal plots.

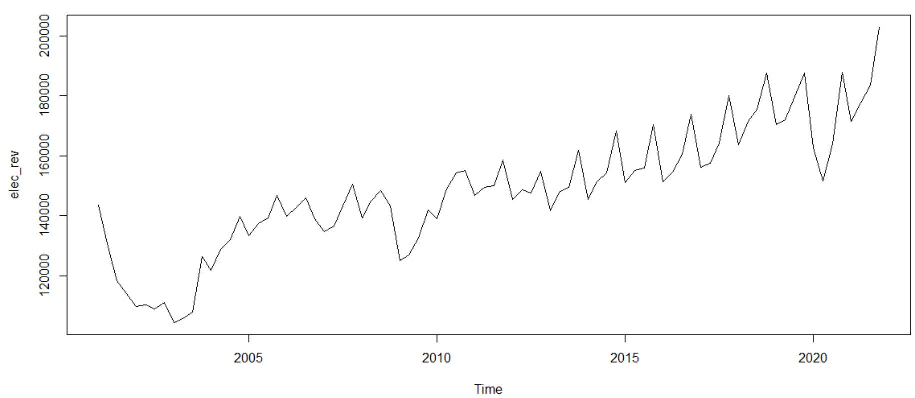

Fig 2.2.1

Looking at the ACF, PACF and the seasonal plots in the following figures (2.2.2, 2.2.3 and 2.2.4), we can confirm that the data indeed has a trend and seasonality.

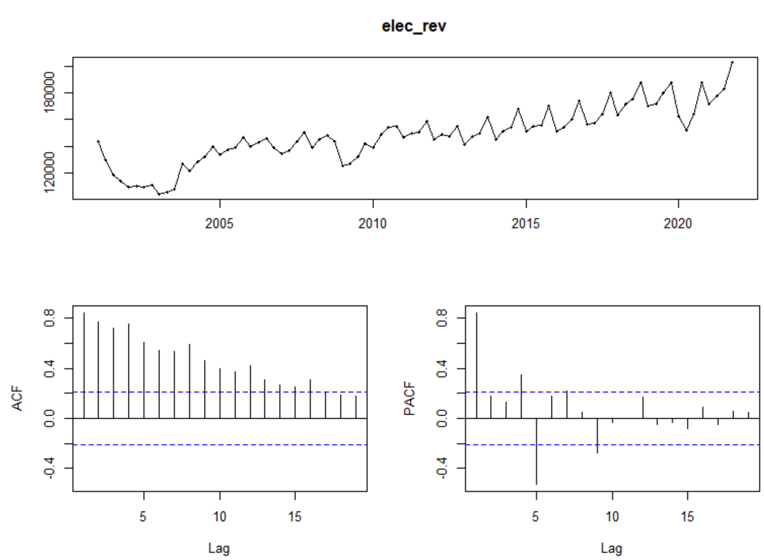

Fig 2.2.2

From the following figure, we see that the seasonality is at its peak in Q4, which could be justified by the fact that a lot of people buy electronics during the holiday season and there are a lot of promotions on these product categories that run during the 4th quarter. (Cyber Monday Sales etc.)

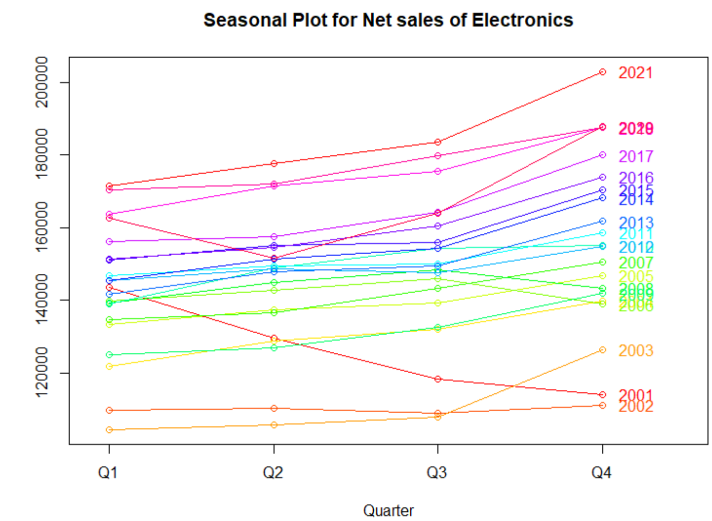

Fig 2.2.3

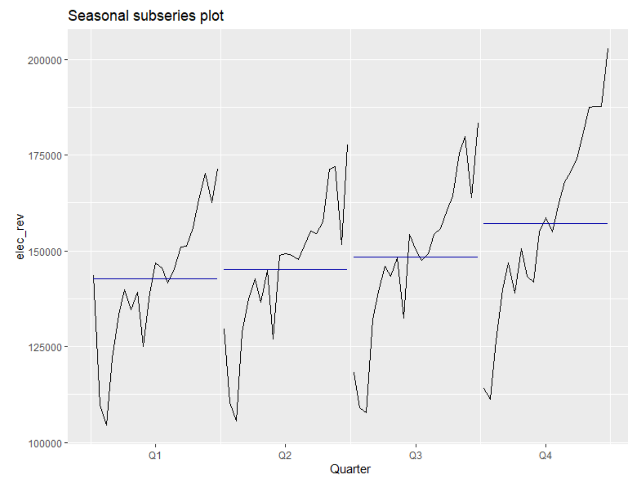

Fig 2.2.4

2.3) Modelling & Performance

To choose the best model for our dataset, we 1st set a benchmark model, for the benchmark I choose the Holt-Winters model.

In the following table, we can see the model performance on the test dataset

Model	RMSE	MAE	MAPE	MASE
Holt Winters - Additive	8852.01	6156.903	3.63246	0.77717
Holt Winters - Multiplicative	9101.03	5951.668	3.591634	0.751264

Fig 2.3.1

Looking at the residuals and ACF plot from the Holt winter methods, we can say that the model seems to capture most of the data, however there is a minor white noise based on the ACF plot below.

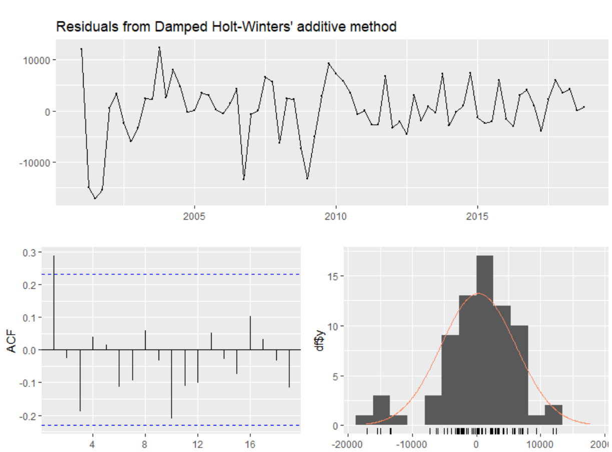

Fig 2.3.2

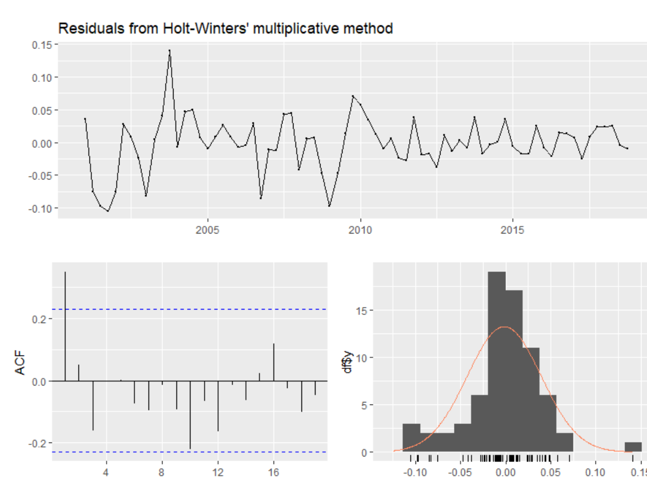

Fig 2.3.3

2.4) Model Selection

Next, we run and compare different models to choose the best one that gives us the optimal performance, to choose the best model, we use the combination of MASE & Ljung-Box test.

2.4.1) STL Models

First, we compare variations of the STL models, since our data has seasonality STL may be a good model to evaluate.

Looking at the table below (2.3.4), the holt winters still seem to be the best model based on MASE alone, however, the difference in the MASE value is very minute, so we can look at the Ljung-Box test to evaluate the model further.

Models	RMSE	MAE	MAPE	MASE
Holt Winters - Multiplicative	9101.03	5951.67	3.59	0.751
STL ets lambda	8904.52	5957.68	3.57	0.752
STL rwdrift lambda	9012.43	6015.10	3.61	0.759
STL arima lambda	8616.86	6077.71	3.58	0.767
Holt Winters - Additive	8852.01	6156.90	3.63	0.777
STL rwdrift	9406.60	6216.74	3.73	0.785
STL arima	9194.33	6265.21	3.69	0.791
STL naive lambda	8804.96	6279.56	3.69	0.793
STL naive	9170.23	6337.26	3.72	0.800
STL ets	9170.23	6337.27	3.72	0.800
STL Periodic	12280.32	9220.38	5.54	1.164

Fig 2.3.4

Looking at the significance level, degrees of freedom and the residuals, the STL model with ETS and lambda transformation seems to be a good candidate.

Model	Q*	df	p-value	Significance Level
STL naive lambda	15.9895	8	0.0425	Significant
STL ets	14.1151	6	0.0284	Significant
STL rwdrift lambda	15.9895	7	0.0252	Significant
STL Periodic	16.6593	7	0.0197	Significant
STL ets lambda	10.395	3	0.0155	Significant
Holt Winters - Additive	17.1526	3	0.0007	Significant
Holt Winters - Multiplicative	18.037	3	0.0004	Significant
STL arima	4.3758	7	0.7356	Not Significant
STL arima lambda	5.9727	7	0.5429	Not Significant
STL naive	13.565	8	0.0938	Not Significant
STL rwdrift	13.565	7	0.0595	Not Significant

Fig 2.3.5

In the following residual plots, we can see compare the impact that the lambda transformation has on the model, transformation can cover more data points compared to non-transformed data.

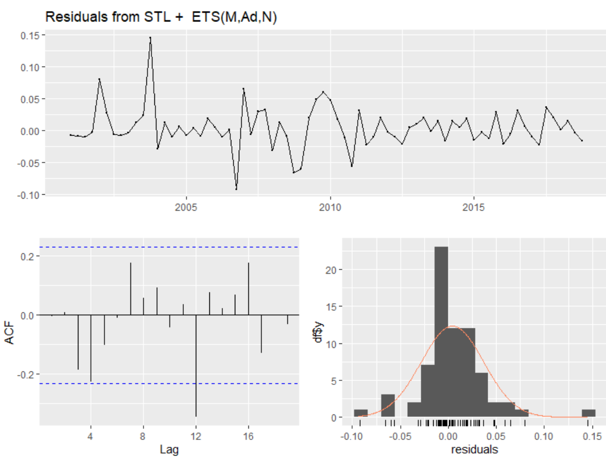

Fig 2.3.6 (STL With lambda)

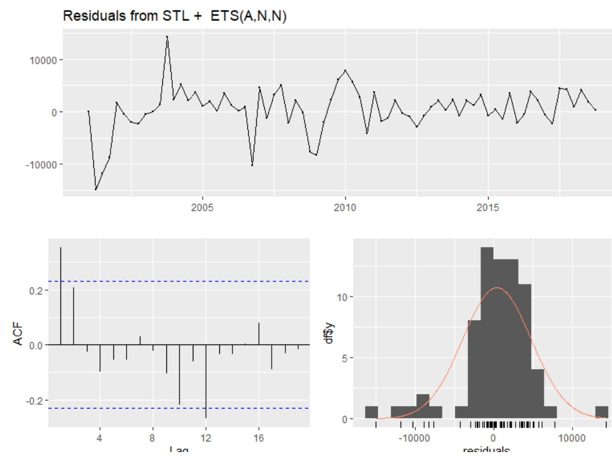

Fig 2.3.7 (STL without lambda)

2.4.2) ETS Models

Next, we look at the ETS models to check if they can give us better results because they consider the seasonality and trend of the data. Just by looking at the MASE alone, Holt winters is still the best model, let's compare the residuals next to see if they give us a different picture.

Model	RMSE	MAE	MAPE	MASE
Holt Winters - Multiplicative	9101.03	5951.67	3.59	0.75
ETS - AAdA	8852.24	6156.86	3.63	0.78
ETS - Holt Winters - Additive	8852.01	6156.90	3.63	0.78
ETS - AAA	9840.11	6642.89	4.01	0.84
ETS - MAdM	10704.00	7422.61	4.46	0.94
ETS - AAdA lambda	10581.59	7442.38	4.49	0.94
ETS - MAdA	11634.33	8244.27	4.95	1.04
ETS - AAA lambda	11589.08	9393.99	5.51	1.19
ETS - MAM	12818.70	9871.22	5.93	1.25
ETS - MAA	14361.23	11522.89	6.90	1.45

Fig 2.4.1

Based on the significance level and the residuals, most of the ETS models are significant, so we can further evaluate the model with an acceptable residual value & MASE. Which in our case could be ETS with AAA.

Model	Q*	df	p-value	Significance
Holt Winters - Additive	17.1	3	0.0007	Significant
Holt Winters - Multiplicative	18.0	3	0.0004	Significant
ETS - AAA	16.2	3	0.001	Significant
ETS - MAA	21.7	3	0.0001	Significant
ETS - MAM	13.6	3	0.0035	Significant
ETS - AAdA	17.1	3	0.0007	Significant
ETS - MAdA	21.6	3	0.0001	Significant
ETS - MAdM	14.1	3	0.0027	Significant
ETS - AAA lambda	14.0	3	0.0028	Significant
ETS - AAdA lambda	17.8	3	0.0005	Significant

Fig 2.4.2

Looking at the residuals and ACF plot for the ETS model with parameters AAA, it seems to capture most of the data, seasonality, and the trend, however, we can see there is a minor part of the data at lag ~1 that it seems to miss out on.

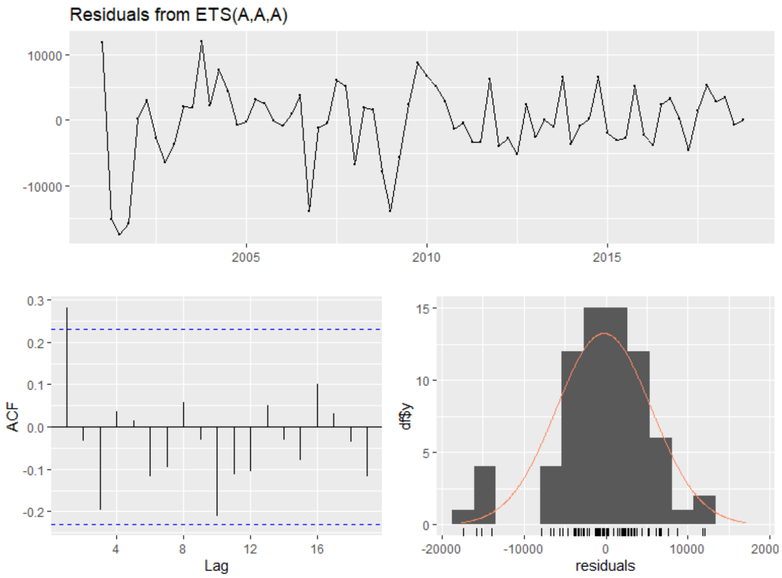

Fig 2.4.3

2.4.3) ARIMA

Lastly, we run some of the ARIMA models before we compare all the models and choose the best one.

The MASE & MAPE for these models are not too good when compared with our benchmark model, so we again look at the Ljung-Box test for residual and significance levels along with the degrees of freedom.

Model	RMSE	MAE	MAPE	MASE
Holt Winters - Multiplicative	9101.03	5951.668	3.591634	0.7512639
Holt Winters - Additive	8852.01	6156.903	3.63246	0.7771703
ARIMA dD	11630.958	8173.675	4.934558	1.0317423
ARIMA lambda	10074.079	8406.817	4.770595	1.0611712
ARIMA	10630.415	8716.117	4.913548	1.1002133
ARIMA dD lambda	14850.038	12185.685	7.119804	1.5381682

Fig 2.4.4

None of the ARIMA models has an acceptable significance level, so the null hypothesis is rejected and there is no white noise, the residuals seem to be quite low, so let us look at the ACF plot to understand if this captured all the data or not.

Model	Q*	df	p-value	Significance
Holt Winters - Additive	17.1526	3	0.0007	Significant
Holt Winters - Multiplicative	18.037	3	0.0004	Significant
ARIMA	3.8709	3	0.2757	Not Significant
ARIMA lambda	5.666	3	0.129	Not Significant
ARIMA dD	2.4228	4	0.6585	Not Significant
ARIMA dD lambda	2.5432	3	0.4675	Not Significant

Fig 2.4.5

Based on the following plot (Fig 2.4.6), the model seems to be capturing all the data points, we can see the side-by-side comparison of this model with the Holt winter's multiplicative model below (Fig 2.4.7).
Where holt winter seems to miss out on a small part of the data, ARIMA can capture everything, so we can trade off a slightly higher MASE with more convergence.

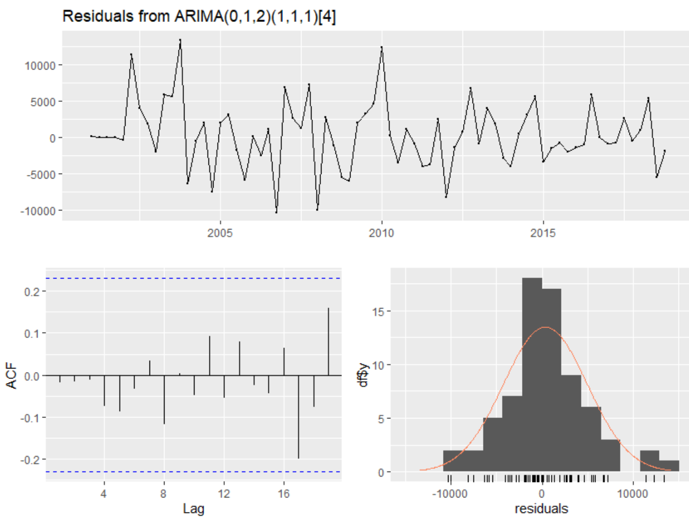

Fig 2.4.6

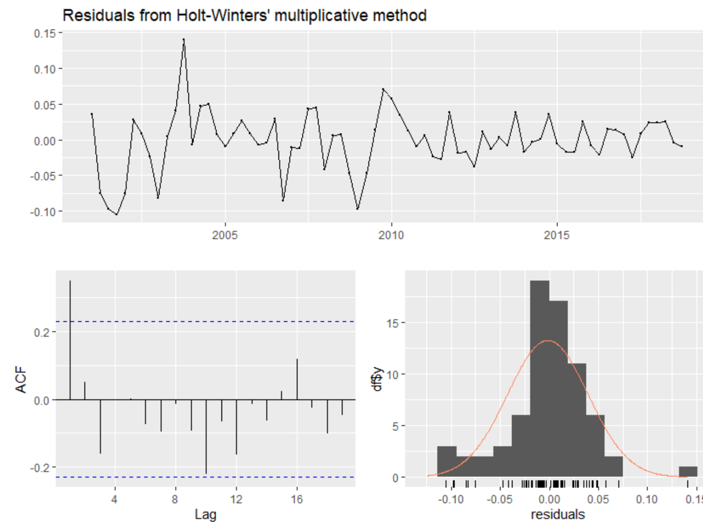

Fig 2.4.7

So, we choose the ARIMA(2,1,2)(1,1,0) model as our final model and forecast the net sales of electronics in the US market till 2024 Q4.

2.5) Forecasting

We can notice that the forecast seems to capture the seasonality very well, and looking at the trend, it also seems to be somewhat aligned with the rest of the data.

Legend:
Black → Train data
Green → Test data
Blue → Forecasted data

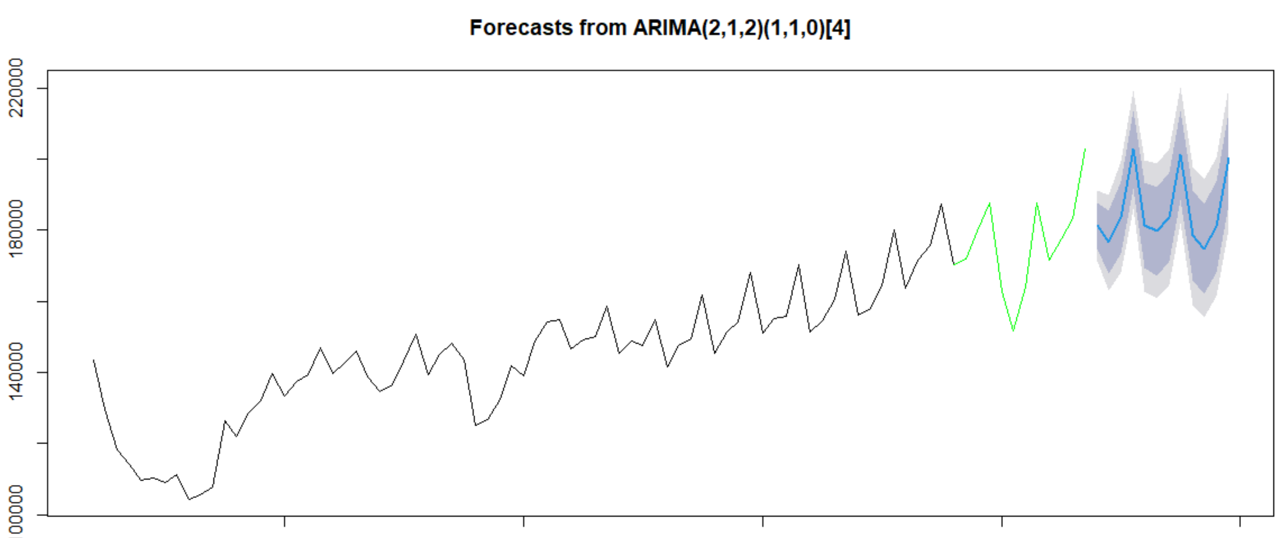

Fig 2.5.1

2.6) Conclusion

We did have to make a trade-off between choosing the best MASE or the model that captured the complete data, seasonality, and trend, however, the results look promising.

From the following plot, we can see a similar trend as the historical data, every 4th quarter has an upwards trend, and it drops in Q1 of every year. The lag also seems to be in line with the historical data.

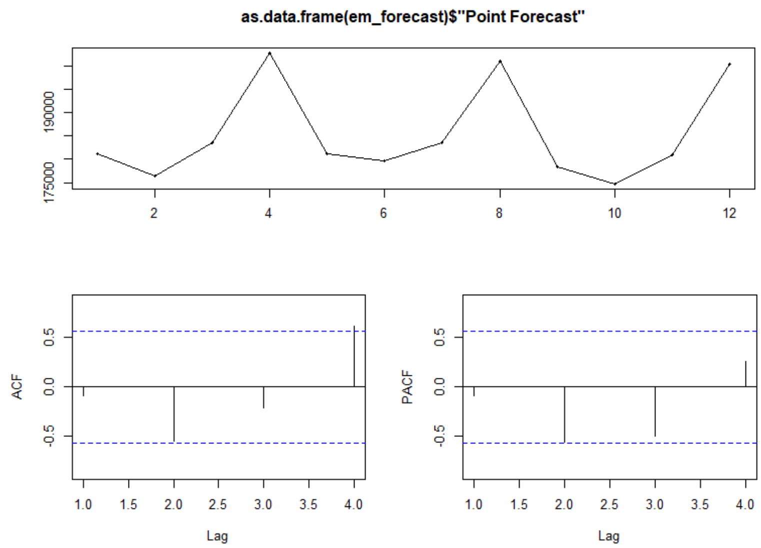

Fig 2.5.2

--------------------------------------End of Report-------------------------------------

References

- Forecasting Class notes, slides by Prof. Filip Van den Bossche (KU Leuven/IESEG School of Management)
- R code from class exercises and solutions by Prof. Filip Van den Bossche (KU Leuven/IESEG School of Management)
- https://en.wikipedia.org/wiki/Ljung%E2%80%93Box_test
- https://www.statisticshowto.com/ljung-box-test/
- https://www.statology.org/ljung-box-test/
- https://www.kaggle.com/code/iamleonie/time-series-interpreting-acf-and-pacf/notebook
- https://towardsdatascience.com/identifying-ar-and-ma-terms-using-acf-and-pacf-plots-in-time-series-forecasting-ccb9fd073db8
- https://www.analyticsvidhya.com/blog/2021/08/holt-winters-method-for-time-series-analysis/
- Data Camp Forecasting Course
- https://www.datacamp.com/courses/forecasting-product-demand-in-r
- https://www.youtube.com/watch?v=dBNy_A6Zpcc
- https://orangematter.solarwinds.com/2019/12/15/holt-winters-forecasting-simplified/#:~:text=Holt%2DWinters%20forecasting%20is%20a,forecasting%20techniques%20for%20time%20series.
- https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp
- https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average
- https://en.wikipedia.org/wiki/Exponential_smoothing
- https://docs.rapidminer.com/9.4/studio/operators/modeling/time_series/forecasting/holt-winters_trainer.html
- https://otexts.com/fpp2/stl.html
- https://en.wikipedia.org/wiki/Correlogram
- https://en.wikipedia.org/wiki/Ljung%E2%80%93Box_test
- https://en.wikipedia.org/wiki/Mean_absolute_scaled_error
- https://en.wikipedia.org/wiki/Mean_absolute_percentage_error