iPhone Sales Forecasting for 2024

Introduction:

About the data:

We are provided with the history of iPhone data from 2007 to 2023, which includes iPhone sales, number of units sold, and sales per unit. The provided data is quarterly.

Sales: iPhone revenue in million dollars

Units Sold: Number of units sold every quarter in millions.

Sales Per Unit: Average Cost of each unit (Sales/Unit Sold).

There is missing data for Units Sold and Sales Per Unit from 2019 to 2023.

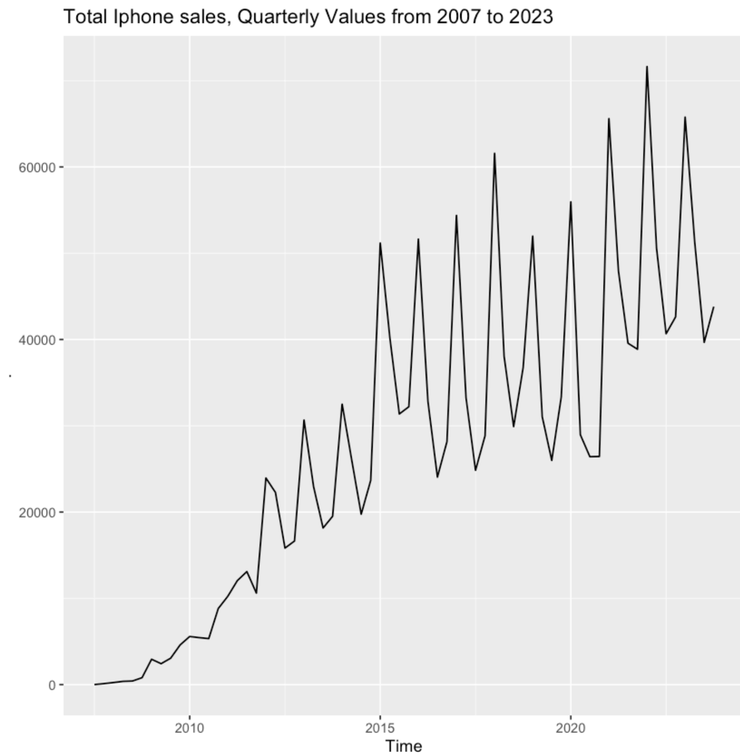
Aim:

→ We need to fill in the missing data for Units Sold and Sales Per Unit.

→ We need to Forecast iPhone sales for 2024.

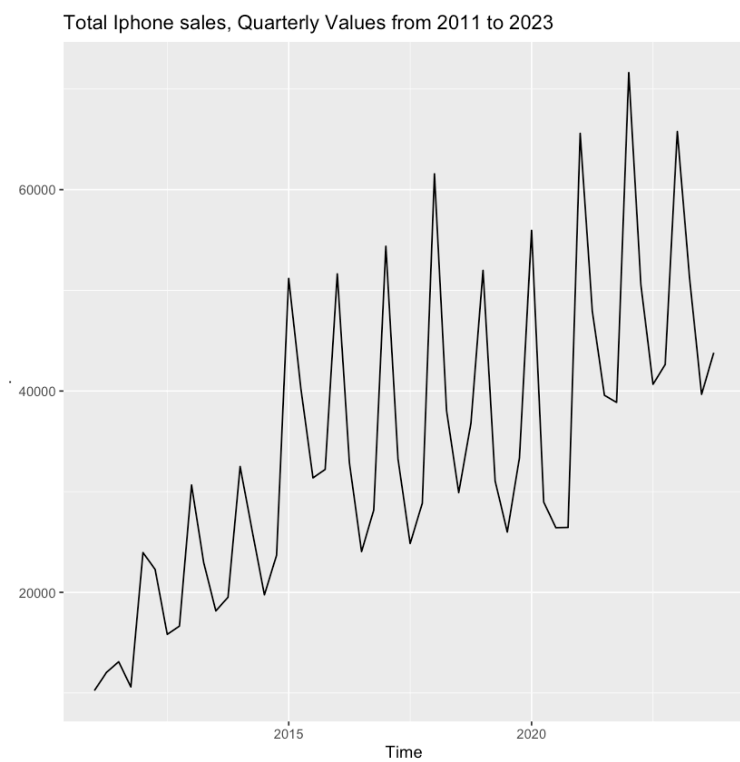
→ We need to do a Sensitivity Analysis with a 10% increase and decrease in the Average Selling price.

Basic Plots of iPhone Sales:



→ This graph is plotted to show iPhone sales for all available data from 2007 to 2023.

→ We can observe that in the initial years, from 2007 to 2010, there were very few sales, and the data was not stable.

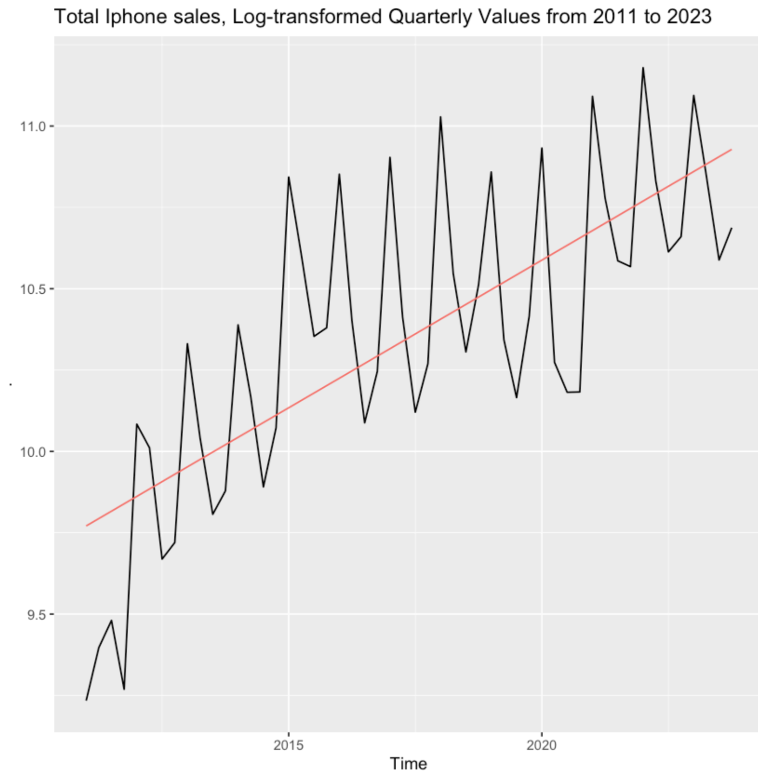


→ We considered data from 2011 for better understanding and future predictions, as the data has been mostly stable since 2011.

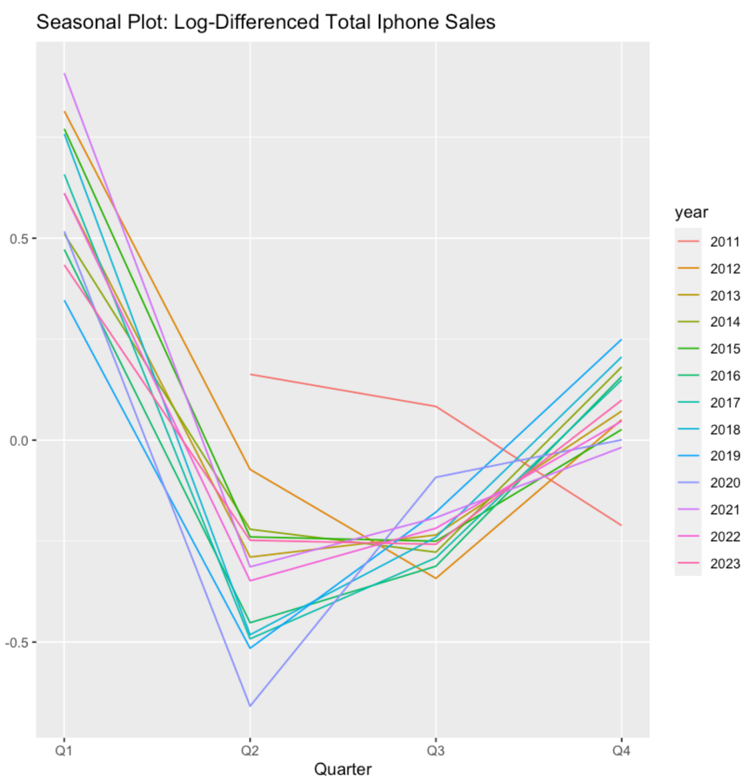
→ We see from the graph that there is an increasing trend and seasonality.

→ Around 2019 and 2020 there was a slight dip in sales when compared to previous years this may be due to the COVID

Trend and Seasonality plots of iPhone Sales:



→ From the graph we can clearly see that there is an upward trend in the iPhone sales.



→ We can clearly observe there is a seasonality from the graph.

→Every year, in the first quarter, there is an increase in iPhone sales. Many people buy more iPhones as they want to start their year with a brand-new phone, or because iPhone releases generally happen in the last quarter of the year, so people wait for reviews before buying and buy in first quarter.

→ Sales decrease every year in the second quarter, probably because most people have already bought the new iPhone or are waiting for the next release.

→ Sales gradually increase in the third and fourth quarters.

Estimating the missing Units Sold column values:

→ In the given data we have information about iPhone sales quarterly, the number of units sold quarterly, and the average selling price per unit (avg_selling_price_per_unit = sales/units_sold).

→We have full data about iPhone sales from 2007 to 2023 but we only have data from 2007 to 2018 regarding units_sold and data is missing from 2019.

→ We will be using different models to estimate the missing values data.

Initial estimation for 2019 and 2020 years:

Training data report:

A tibble: 6 × 5

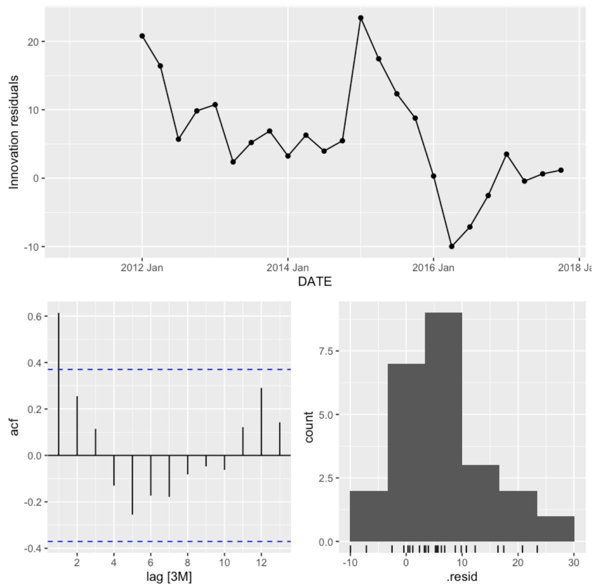
.model	ME	RMSE	MAPE	ACF1
<chr>	<dbl>	<dbl>	<dbl>	<dbl>
NAIVE Model	1.127407e+00	15.523964	24.896269	-0.2188700
Trend	5.075305e-16	11.615038	22.978142	0.1101399
Seasonal Trend	2.410770e-15	7.199230	16.018901	0.4304254
SNAIVE Model	6.019167e+00	9.951756	17.913321	0.6140132
Random Walk	1.480297e-15	15.482972	25.037063	-0.2188700
ES Model	-1.624057e+00	5.371352	9.495352	0.2636509

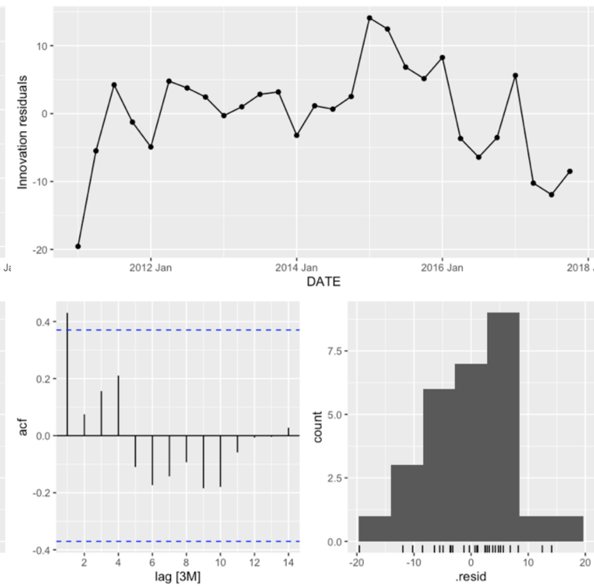
→ Above is the report for the training data which is till 2017.

→ Exponential smoothing has the lowest MAPE (mean absolute percent error) value followed by Seasonal Naive and Seasonal Trend models.

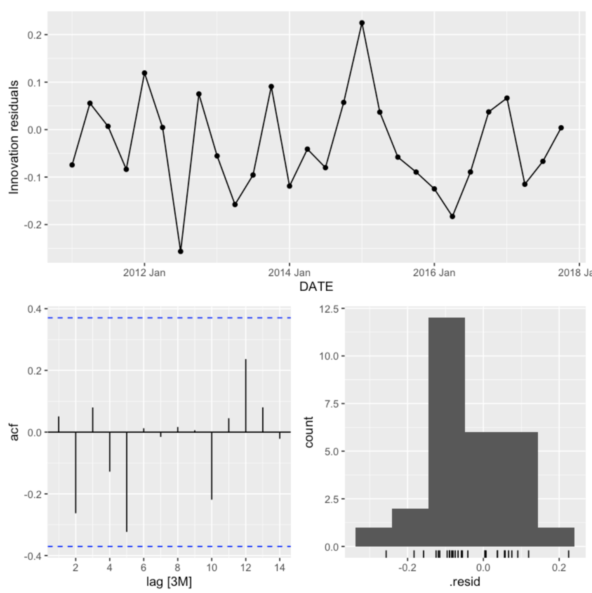
SNAIVE Model residuals:

Seasonal Trend model Residuals:





Exponential Smoothing Model Residuals:



→ From the above residual graphs, we can see that the Seasonal Naïve model captures the residual data without any clear pattern, which is good, but the Exponential Smoothing model has a pattern, which isn't as good.

→ When it comes to serial correlation, all three models are performing well. The Exponential Smoothing model shows no serial correlation, and both the Seasonal Naïve and Seasonal Trend models have one lag crossing the confidence interval. This indicates that all the models effectively account for seasonality with little to no correlation.

→ Among all the other model's distribution of residual is good for Seasonal Naïve which is near to the normal distribution which indicates it captures the underlying pattern well when compared to other models.

Test data reports:

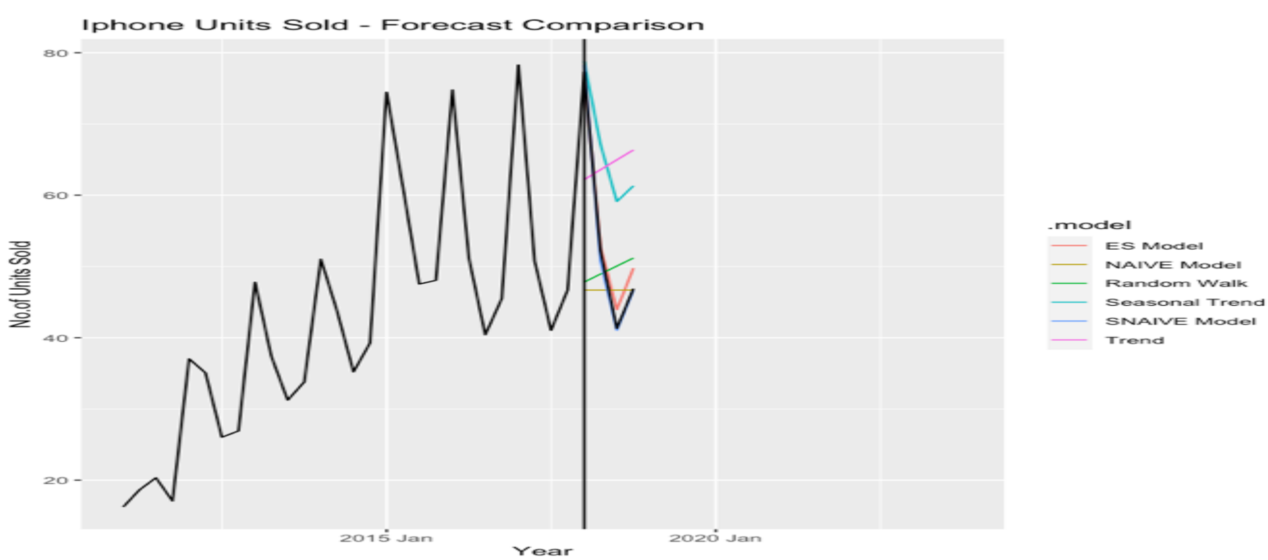
A tibble: 6 × 5

.model	ME	RMSE	MAPE	ACF1
<chr>	<dbl>	<dbl>	<dbl>	<dbl>
ES Model	-1.435960	1.9791636	3.515282	0.300215067
NAIVE Model	7.745000	15.7880303	15.920583	0.101797796
Random Walk	4.926481	15.6177136	18.692732	0.123024791
SNAIVE Model	0.235000	0.8909265	1.290545	-0.485587827
Seasonal Trend	-12.171429	13.7010846	26.114776	-0.007536082
Trend	-9.847343	18.0043470	35.035458	0.127407403

→ Above is the report for the test data which is for 2018.

→ Seasonal Naïve has the lowest MAPE (mean absolute percent error) value followed by Exponential smoothing.

Comparing Forecasting models on Test data:



→We can see that the Seasonal Naïve Model is doing well on the test data forecasting when compared to others as it almost merges with the actual test data.

Forecasting data for 2019 and 2020:

→ Based on the above results from reports, residuals, and test forecasting, I choose Seasonal Naïve to forecast the data for 2019 and 2020.

→ Below is the forecasted data for 2019 and 2020.

A tbl_ts: 6 × 2

DATE	UNITS_SOLD
<mth>	<dbl>
2019 Jul	41.03
2019 Oct	46.68
2020 Jan	78.29
2020 Apr	50.76
2020 Jul	41.03
2020 Oct	46.68

Estimation of missing values for 2021,2022, and 2023:

Training data report:

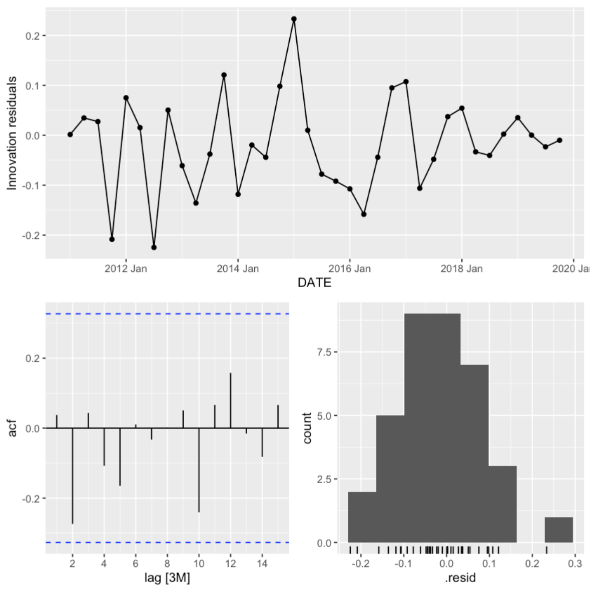
A tibble: 6 × 5

.model	ME	RMSE	MAPE	ACF1
<chr>	<dbl>	<dbl>	<dbl>	<dbl>
NAIVE Model	8.697143e-01	17.133260	26.658790	-0.2310585
Trend	1.924387e-15	13.082400	25.157890	0.1392055
Seasonal Trend	3.454027e-15	8.320187	17.527423	0.5066850
SNAIVE Model	4.514375e+00	8.618474	13.434991	0.6496214
Random Walk	-1.141944e-15	17.111171	26.737002	-0.2310585
ES Model	-7.349882e-01	4.629227	7.594775	0.2134866

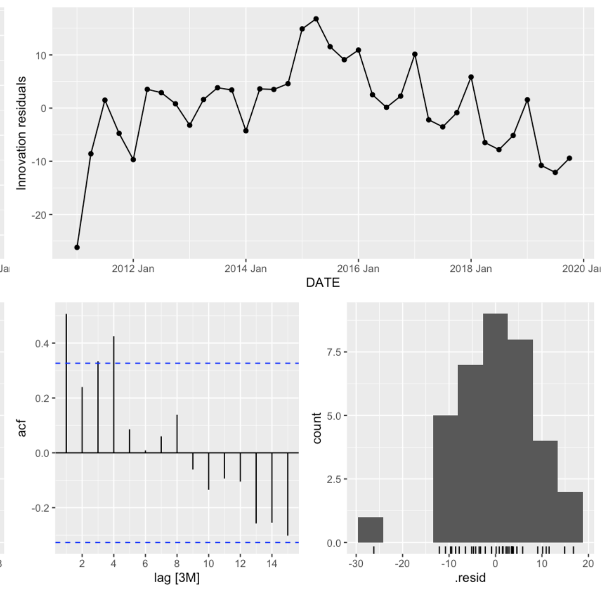
→ Above is the report for the training data which is till 2019.

→ Exponential smoothing has the lowest MAPE (mean absolute percent error) value followed by Seasonal Naive and Seasonal Trend models.

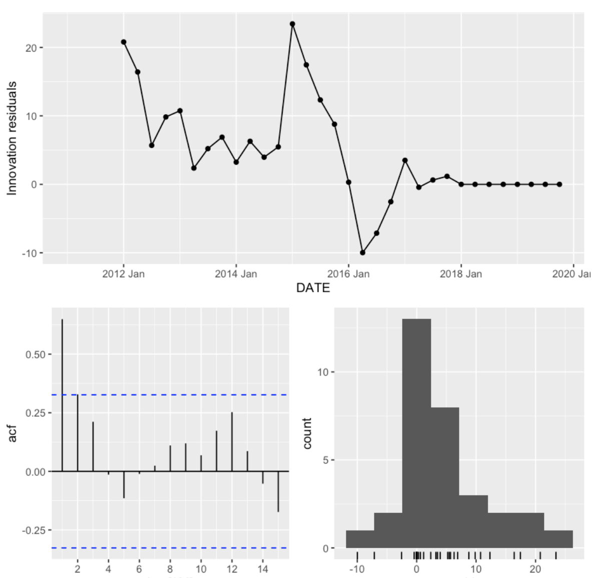
Exponential Smoothing Model Residuals:



Seasonal Trend Model Residuals:



Seasonal Naïve Model residuals:



→ From the above residual graphs, we can see that the Exponential Smoothing model captures the residual data better when compared to other models.

→ When it comes to serial correlation, all three models are performing well. The Exponential Smoothing model shows no serial correlation where all the lags lie within the confidence interval which is a good sign.

→ For the Exponential Smoothing model, the distribution of residuals forms an almost normal distribution curve which indicates that the model captures the underlying pattern well.

Test data reports:

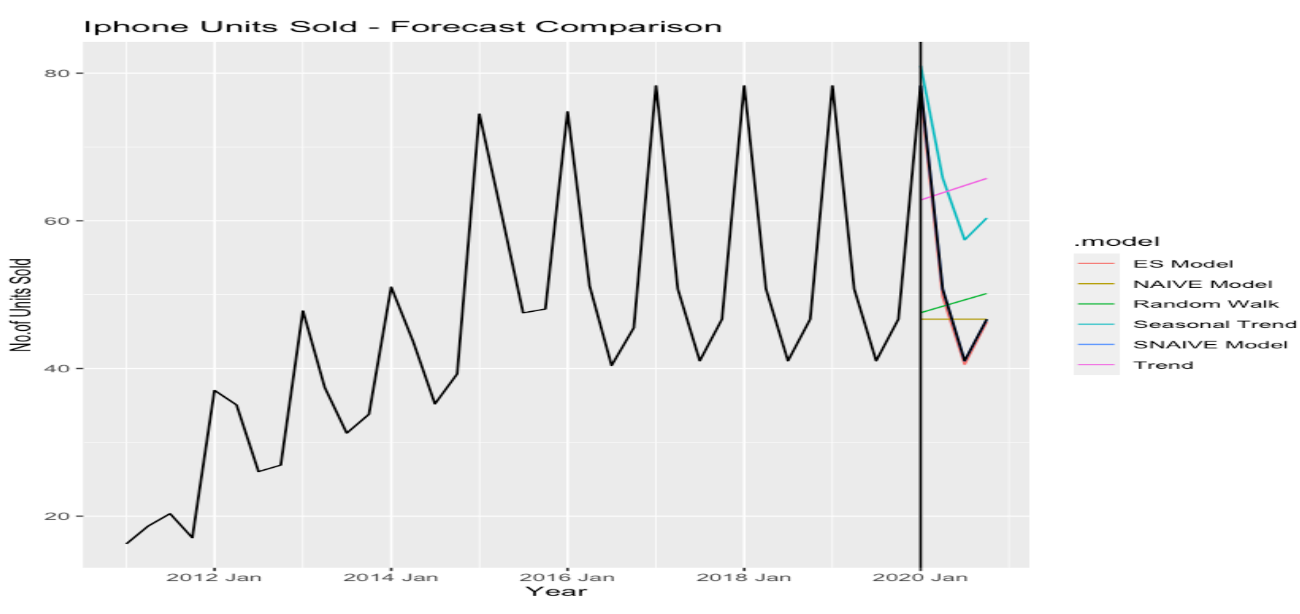
A tibble: 6 × 5

.model	ME	RMSE	MAPE	ACF1
<chr>	<dbl>	<dbl>	<dbl>	<dbl>
ES Model	0.7792973	0.8620768	1.394619	0.30798939
NAIVE Model	7.5100000	16.1845683	15.545941	0.07456864
Random Walk	5.3357143	16.0527204	17.864445	0.09211790
SNAIVE Model	0.0000000	0.0000000	0.000000	NaN
Seasonal Trend	-11.9681944	13.1391000	25.610410	-0.06165238
Trend	-10.0922179	18.2796497	36.038663	0.09413603

→ Above is the report for the test data which is for 2020.

→ Exponential smoothing has the lowest MAPE (mean absolute percent error) among all other models (ignoring seasonal naïve models as it has 0 values).

Comparing Forecasting models on Test data:



→ From the above comparison graphs we can see that the Exponential smoothing model is doing well on the test data forecasting when compared to others as it almost merges with the actual test data.

Forecasting data for 2021, 2022, and 2023:

→ Based on the above results from reports, residuals, and test forecasting, I choose the Exponential smoothing model to forecast the data for 2021, 2022, and 2023.

→ Below is the forecasted data for 2021, 2022, and 2023.

A tbl_ts: 12 × 2

DATE	UNITS_SOLD
<mth>	<dbl>
2021 Jan	76.86629
2021 Apr	49.82943
2021 Jul	40.67580
2021 Oct	46.45524
2022 Jan	77.23511
2022 Apr	50.05453
2022 Jul	40.84945
2022 Oct	46.64335
2023 Jan	77.53264
2023 Apr	50.23876
2023 Jul	40.99359
2023 Oct	46.80166

→ By using the sales values and units_sold values I have filled the sales_per_unit column using the formula sales/units_sold.

Data after filling in the missing values:

A grouped_ts: 8 × 4

DATE	SALES	UNITS_SOLD	SALES_PER_UNIT
<mth>	<dbl>	<dbl>	<dbl>
2022 Jan	71628	77.24	927.4021
2022 Apr	50570	50.05	1010.2981
2022 Jul	40665	40.85	995.4847
2022 Oct	42626	46.64	913.8709
2023 Jan	65775	77.53	848.3523
2023 Apr	51334	50.24	1021.8006
2023 Jul	39669	40.99	967.6880
2023 Oct	43805	46.80	935.9711

Forecasting iPhone Sales for 2024:

Training data reports:

A tibble: 7 × 3

.model	r_squared	AICc
<chr>	<dbl>	<dbl>
Trend	0.4301813	899.1332
Seasonal Trend	0.8034365	855.5486
Seasonal Trend Units_Sold	0.9140033	818.6193
SNAIVE Model	NA	NA
ES Model	NA	1014.5785
Naive	NA	NA
MEAN	NA	NA

A tibble: 7 × 5

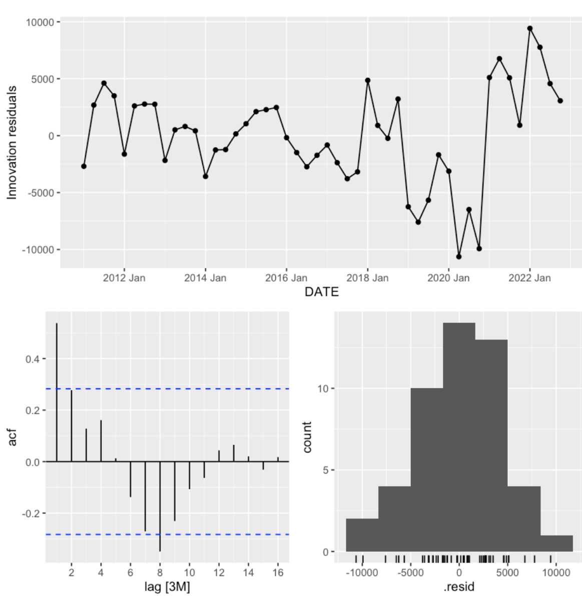
.model	ME	RMSE	MAPE	ACF1
<chr>	<dbl>	<dbl>	<dbl>	<dbl>
Trend	-9.094947e-13	10913.865	27.69364	0.02765682
Seasonal Trend	1.534772e-12	6410.054	18.27891	0.38152914
Seasonal Trend Units_Sold	9.000208e-13	4239.856	11.64804	0.53793265
SNAIVE Model	3.624795e+03	7626.580	17.80287	0.58159330
ES Model	-8.842947e+02	5067.103	11.93997	0.22443028
Naive	6.890851e+02	15332.138	31.42664	-0.26261829
MEAN	-3.221127e-12	14458.068	44.75577	0.41869495

→ Above are the reports for the train data which is till 2022.

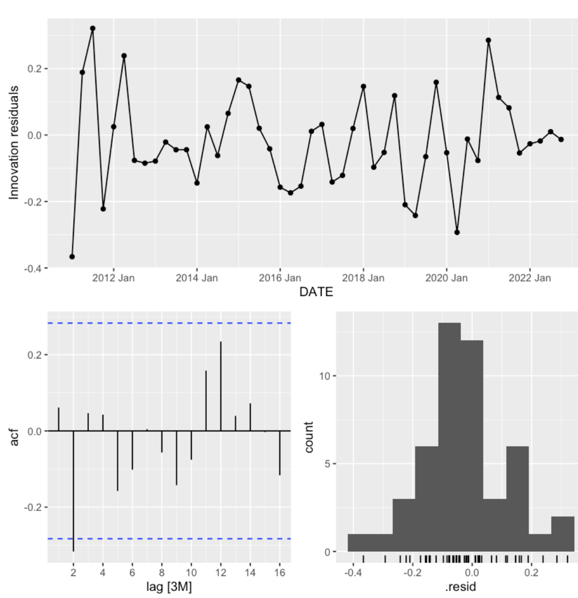
→ We can see the Seasonal Trend Units_Sold has the highest r_squared value and low AICc value when compared to other models.

→ Seasonal Trend Units_Sold has the lowest MAPE (mean absolute percent error) when compared to other models.

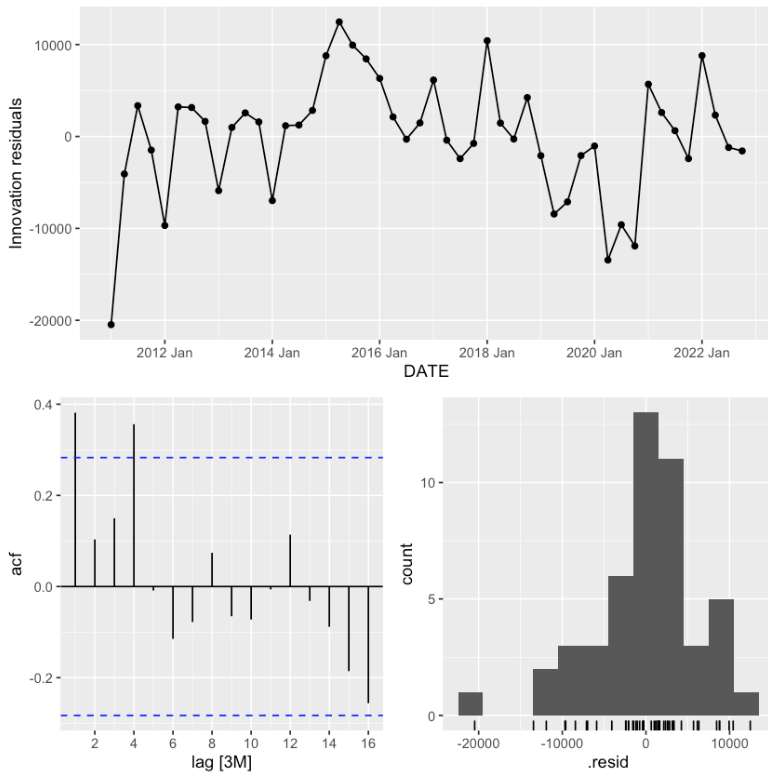
Seasonal Trend Units Sold Model Residuals:



Exponential Smoothing Model Residuals:



Seasonal Trend Model Residuals:



→ From the above residual graphs, we can see that the Seasonal Trend Units_Sold model captures the residual data better when compared to other models.

→ When it comes to serial correlation, all three models are performing well. All the models have 1-2 lags over the confidence interval but not in any seasonal pattern way which is a good sign.

→ For the Seasonal Trend Units_Sold model, the distribution of residuals forms an almost normal distribution curve which indicates that the model captures the underlying pattern well.

Test data reports:

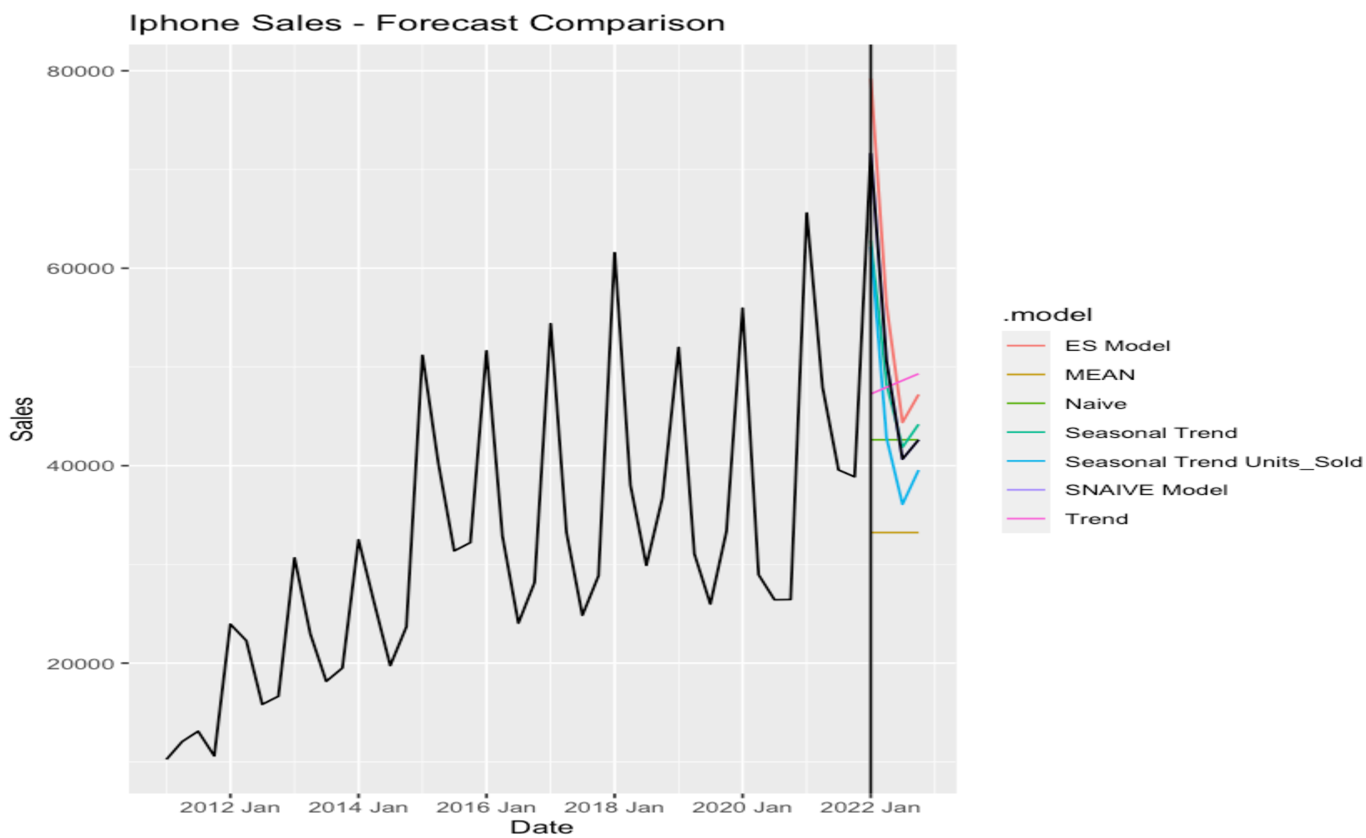
A tibble: 7 × 5

.model	ME	RMSE	MAPE	ACF1
<chr>	<dbl>	<dbl>	<dbl>	<dbl>
ES Model	-5400.150	5592.006	10.443788	0.1795503
MEAN	18144.896	21903.368	32.061173	0.1428180
Naive	8746.250	15067.089	15.255250	0.1428180
SNAIVE Model	0.000	0.000	0.000000	NaN
Seasonal Trend	2088.660	4655.764	5.875631	0.1851422
Seasonal Trend Units_Sold	6208.715	6699.699	11.740254	0.3002199
Trend	3085.680	13312.082	18.621260	0.1539743

→ Above is the report for the test data which is for 2023.

→ Seasonal Trend has the lowest MAPE (mean absolute percent error) among all other models followed by Exponential smoothing and Seasonal Trend Units_Sold model.

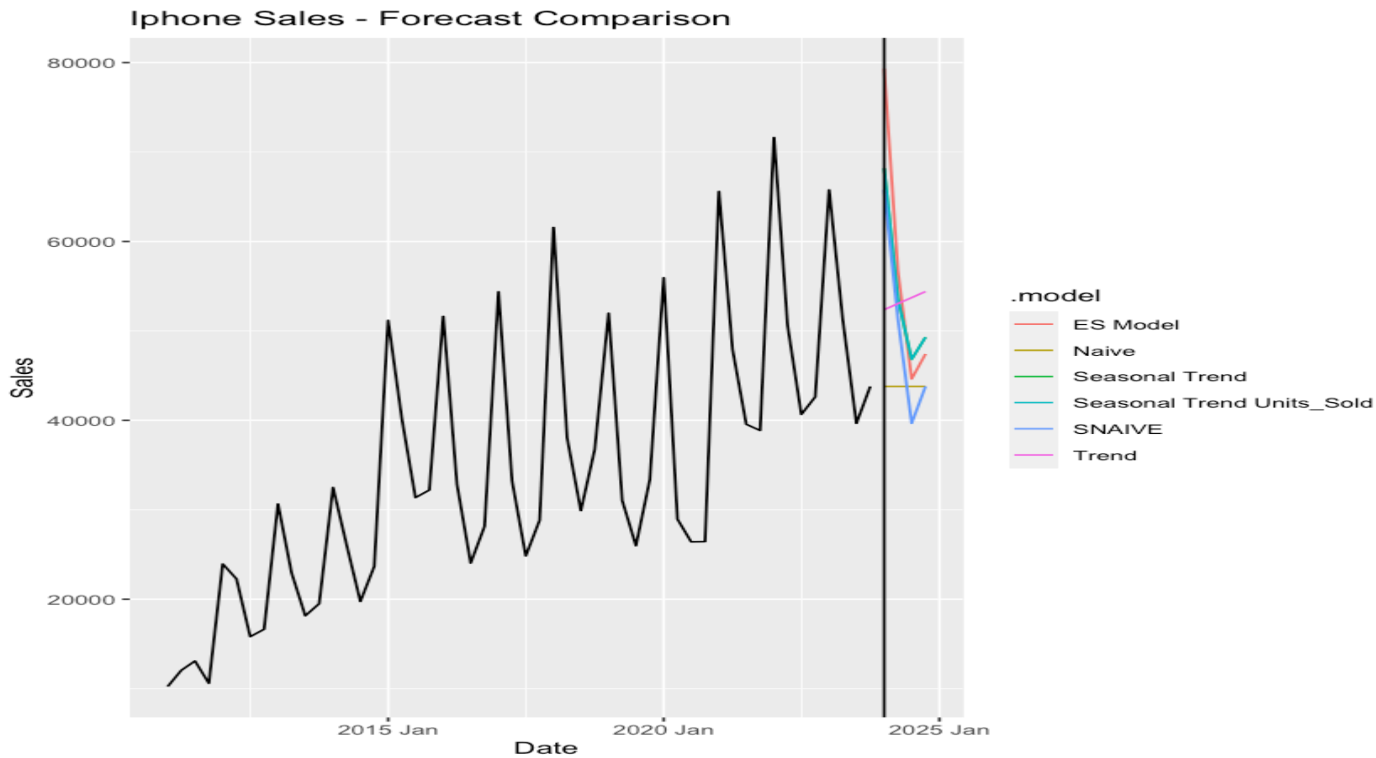
Comparing Forecasting models on Test data:



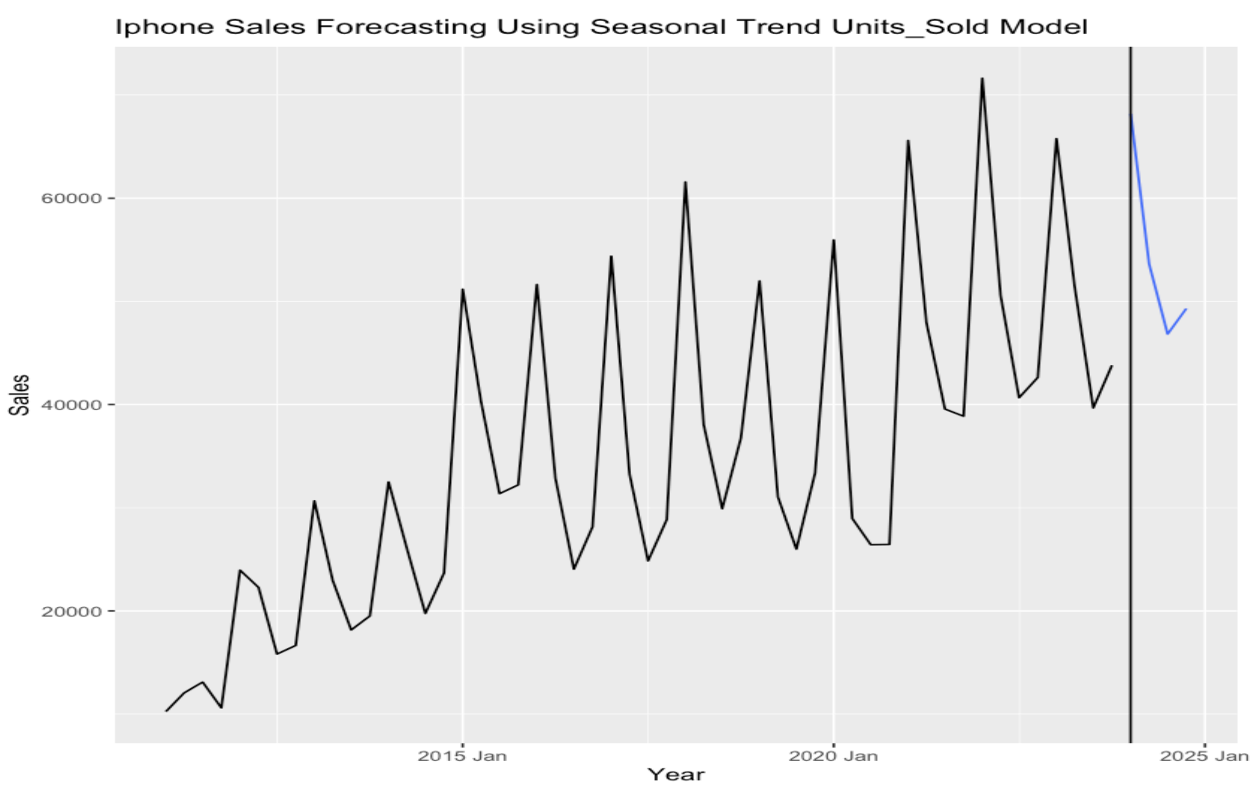
→ From the above comparison graphs we can see that the Seasonal Trend model and Seasonal Trend Units_Sold model are doing well on the test data forecasting when compared to other models.

→ Based on the above reports I choose the Seasonal Trend Units_Sold model because it has the highest r-squared and low AICc value and it is performing well on the residuals when compared to other models.

Different forecasting model graph for 2024 iPhone Sales:

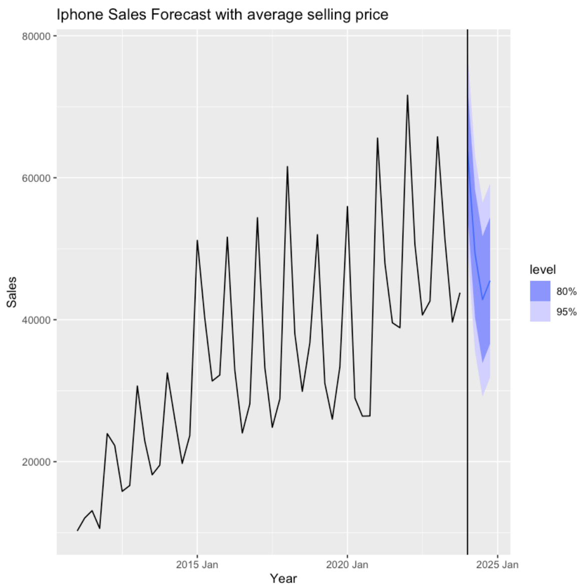


Seasonal Trend Units_Sold Model Forecasting for 2024 iPhone sales:

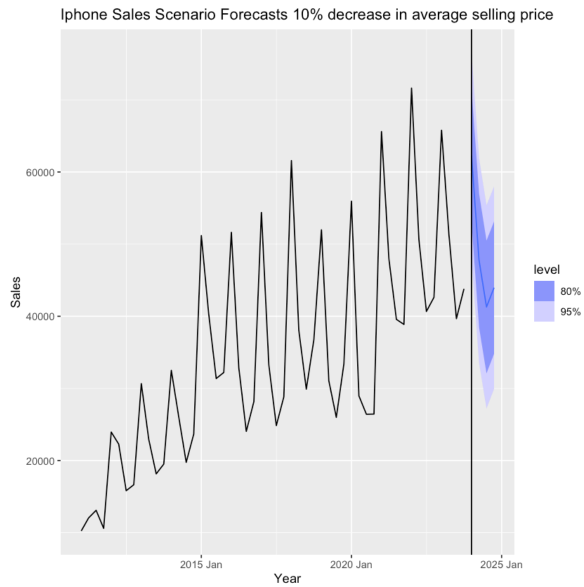


Sensitivity Analysis with 10% increase and 10% decrease in Average Selling price:

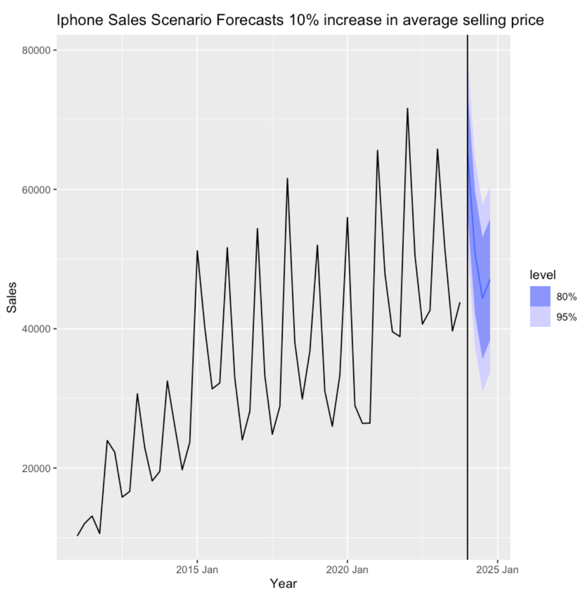
Graph with Average Selling price:



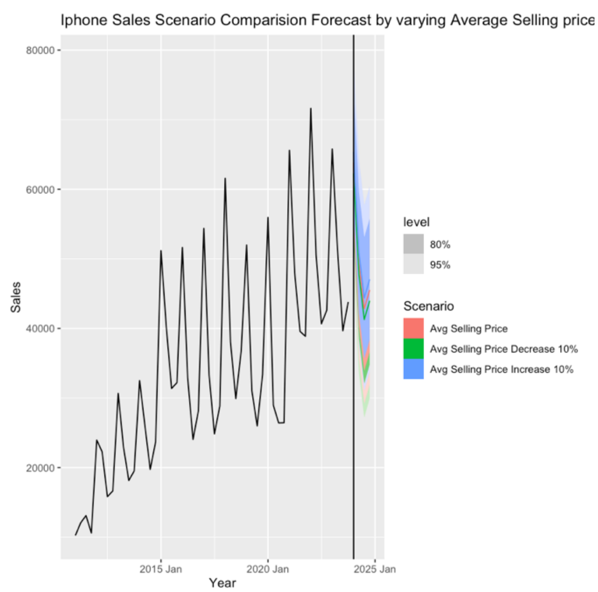
Graph with 10% decrease in Average Selling price:



Graph with 10% increase in Average Selling price:



Scenarios Comparison Graph:



→ We can see from the above graphs that sales decrease when the average selling price decreases by 10% and the sales increase when the average price increases by 10%.

→ There is only a slight change in sales when we increase or decrease the average selling price by 10%.