BAN 673

TIME SERIES PROJECT REPORT

Victor Florencius
Abhishek Yadav
Subramanian Tirunelveli Padmanabhan

Contents

Contents 1
Summary 4
Introduction 5
Eight Steps of Forecasting 6
Models Accuracies 9
  Training & Validation Partition Forecast 9
  Entire Data Set Forecast 10
ARIMA Model 12
  Automated ARIMA Model 12
    Training & Validation Data Summary 12
    Plot for Training & Validation Forecast 13
    Forecast Validation Data 14
    Accuracy of Training & Validation Data 14
    Entire Data Set Summary 14
    Plot for Entire Data Set Forecast 16
    Forecast Future Period 17
    Autocorrelation for Residuals 17
    Accuracy Measure of Entire Data Set 18
  ARIMA (2,1,2) (1,1,0) Model 18
    Training & Validation Data Summary 18
    Plot for Training & Validation Forecast 20
    Forecast for Validation Period 20
    Accuracy Measure for Training & Validation 21
    Entire Data Set Summary 21
    Plot for Entire Data Set Forecast 22
    Forecast for Validation Period 23
    Autocorrelation for Residuals 23
    Accuracy Measure of Entire Data Set 23
Exponential Smoothing (SES) 24

Simple Exponential Smoothing (SES) 24
Training & Validation Data Summary 24
Forecast 25
Accuracy Measure for Training & Validation 25
Entire Data Set Summary 26
Plot for Entire Data Set Forecast 26
Forecast 27
Accuracy Measure for Entire Data set 27
Conclusion 28
Analysis On Results 28
Recommendation 28
Remarks 29
Limitations 29
Benefits 29
Bibliography 29

Summary

The goal of this study is to predict Tesla revenue in 2021 and 2022. The data used for this project is exported from ycharts.com. Tesla dataset has two columns, Quarter and Revenue (converted to millions), and has 50 records from Q4 2008 to Q1 2021. We used plot(), ACF() and stl() functions to visualize and identify components of data.

The revenue vs. quarter chart shows an upward trend from 2008 with few dips in Q1 2019, Q1 2020. In addition, the autocorrelation graph shows a significant autocorrelation coefficient at lag 1, indicating the presence of a trend. Also, considerable lag values at lag 4, lag 5, lag 6, and lag 8 reveal autocorrelation of data and seasonality.

For analysis, we divided data into a training set and a validation set. The training set has 72% (36 records) of records; similarly, validation set has 28% (14 records) of dataset records. To make credible forecasts, we applied multiple time series models to fit the data. Unfortunately, models like moving average (trailing moving average), exponential smoothing (simple, Holt's and Holt-Winter's), regression model(simple linear, quadratic) performed poorly on training or overall data set. Two-level models Quadratic trend with seasonality plus AR(1) and Quadratic trend with seasonality plus Trailing MA(2) recorded better accuracy measures (RMSE and MAPE).Autoregressive models like Auto-Arima performed better on the whole dataset. Model ARIMA(2,1,2)(1,1,0) has the best accuracy model, and hence it fits our data best.

Introduction

Tesla, Inc. is an American electric vehicle and clean energy company based in Palo Alto, California. Tesla is accelerating the world's transition to sustainable energy with electric cars, solar and integrated renewable energy solutions for homes and businesses. Tesla's sales and revenues have been phenomenal in the last four years. Tesla revenue is significantly upwards trends from 2017. In Q4 2020, Tesla reported the all-time highest revenue of 10.74 billion. In addition, Tesla produced 509,737 cars in 2020, which is 40% higher than the year 2019. Tesla cheaper model 3 accounted for 89% of vehicles built-in 2020. The availability of more affordable models and Government support through policies, subsidies, and programs have boosted the sales of electric cars.

Countries' commitment to clean energy and recent charging and battery technology developments is expected to keep this trend upwards.

Dataset used in this project is exported from ycharts.com. We exported Quarter and corresponding Revenue from Q4 2008 to Q1 2021 for analysis. Data is downloaded in CSV format and imported in R script for use.

This project's scope is to forecast the quarterly revenue of Tesla for the years 2021 and 2022. Although project analysis utilizes historical data, external factors directly impacting the revenue (Ex: Government Policies, new Tesla car models, and technology advancements) are out of scope. We will use various time series models and compare the accuracy to pick the best predicting model.

Eight Steps of Forecasting

Step 1: Define Goal

This project aims to create a credible model to forecast Tesla's quarterly revenues for 2021 and 2022. The dataset has 50 revenue records from Q4 2008 to Q1 2021. We will compare the accuracy measures (RMSE and MAPE) to find the best fit model. Some external factors like production data help to increase accuracy. In this project, we are only relying on historical records for predictions.

Step 2: Get data

This study utilizes Tesla revenue data from ycharts.com for analysis. Dataset used in this project has 50 records from Q4 2008 to Q1 2021. The dataset has two fields Quarter (Quarter and Year of revenue reported) and Revenue (in millions)

Step 3: Explore and Visualize Series

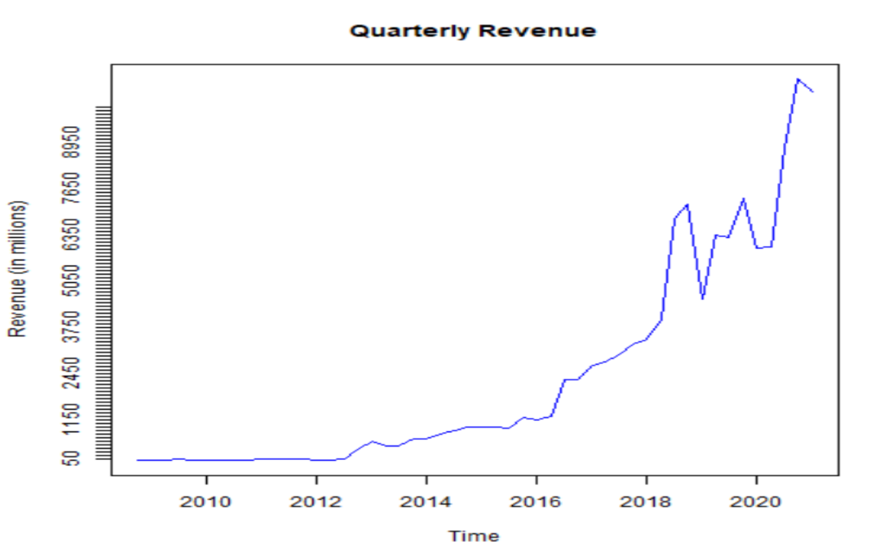

Figure:1

Quarterly revenue data shows an upward trend in revenue from the year 2008 to 2021. Revenue is significantly higher in the years 2019, 2020, and 2021. Q4 revenue is higher for each year except years 2009, 2011and, 2015; high Q4 revenue indicates the presense of seasonality in data. Revenue data is lowest in Q1 for most of the years (exception years 2015,2013, and 2012) and shows an upward trend till Q4.

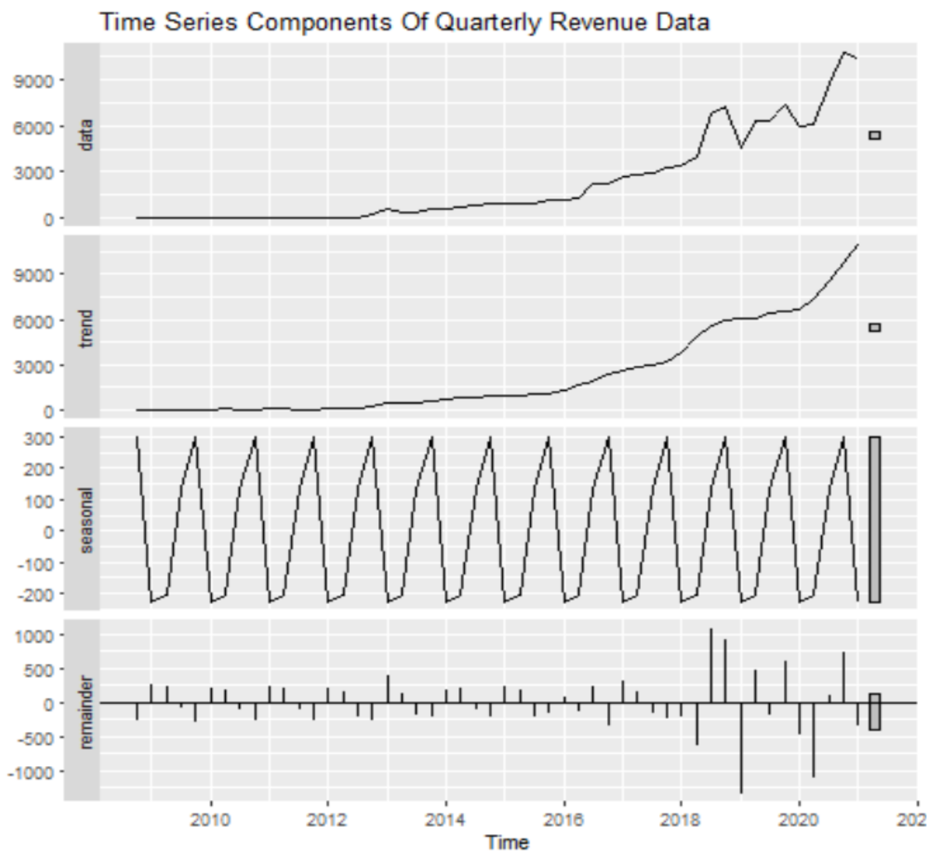

Figure:2

The time series component chart shows an upward non-linear trend in revenue from 2008 to 2021. Seasonal components peeks are not changing over time which indicates the presence of additive seasonality. The seasonal component is consistent throughout the analysis period. Significant levels are recorded in the years 2018, 2019, and 2020. Q1 2019 reported the highest negative level.

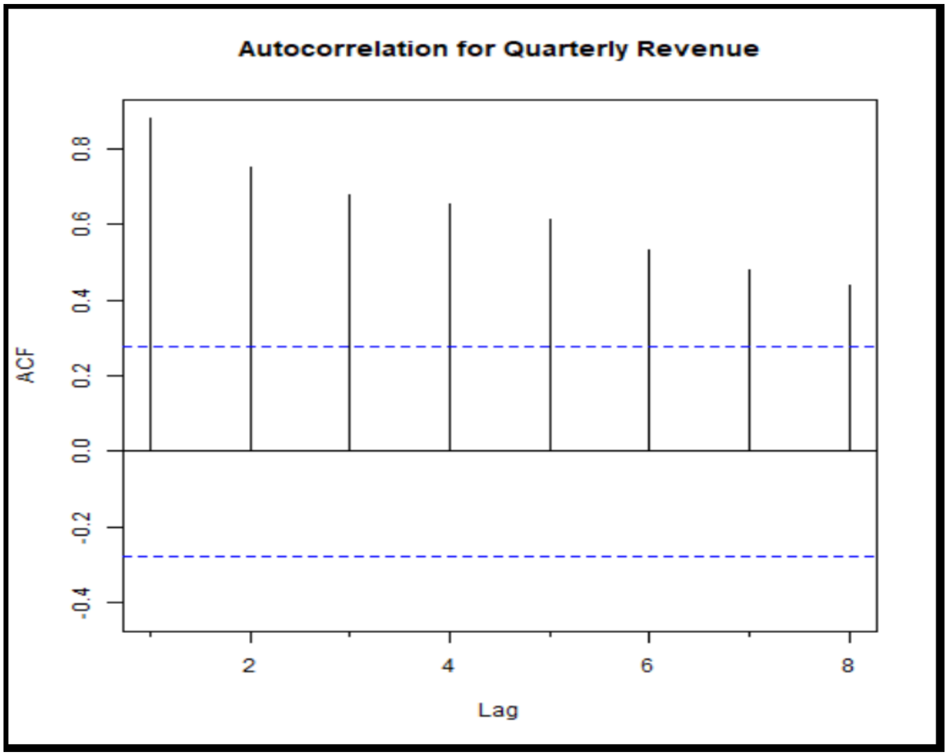

Figure:3

A high autocorrelation coefficient at lag1 indicates the presence of a trend in the data. Significant autocorrelation at lag4 points toward quarterly seasonality. Autocorrelation is also substantial at other lags (lag2, lag3 , lag5, lag6 ,lag7 and lag8).

Step 4: Data Preprocessing

Data offered by ycharts.com is from Q4 2008 to Q1 2021; we utilized all the records present on the website. Quarterly revenue is available in Billions($) converted to Millions($) for analysis. Quarter column value is used as is in the project.

Step 5: Partition Series

Dataset is partitioned into two sets training and validation. The training set has 72% (36 records), and the validation set has 28% (14 records). Tesla revenues have increased significantly from Q2 of 2017. Therefore, we included a slow-moving revenue trend from Q4 2017 and Q1 2018 in validation data to better assess the accuracy.

<u>Step 6 & 7: Apply Forecasting & Comparing Performance</u>

Models Accuracies

Several methods were used to find the best fit for the Revenue projection of Tesla Data set. The forecasting on validation and on entire data set was required to decide on the preferable methods.

Training & Validation Partition Forecast

Accuracy of other Models in Validation Partition were checked to determine the best model

```
> #All training & validation data sets accuracy for other Models
>
> #Trailing MA WIDTH 2
> round(accuracy(ma.trailing_2.pred$mean, valid.ts), 3)
               ME    RMSE      MAE    MPE   MAPE  ACF1 Theil's U
Test set 3467.272 4087.96 3467.272 47.304 47.304 0.568     2.121
> #Trailing MA WIDTH 3
> round(accuracy(ma.trailing_3.pred$mean, valid.ts), 3)
               ME     RMSE      MAE    MPE   MAPE  ACF1 Theil's U
Test set 1971.392 2503.349 1971.392 25.764 25.764 0.405     1.329
> #Trailing MA WIDTH 4
> round(accuracy(ma.trailing_4.pred$mean, valid.ts), 3)
              ME    RMSE     MAE    MPE   MAPE  ACF1 Theil's U
Test set 2334.02 2846.88 2334.02 31.355 31.355 0.437     1.511
>
> # SIMPLE EXPONENTIAL SMOOTHING
> round(accuracy(ses.tesla.pred$mean, valid.ts), 3)
               ME    RMSE      MAE    MPE   MAPE ACF1 Theil's U
Test set 3532.805 4182.42 3532.805 47.972 47.972 0.58     2.159
>
> #HOLTS MODEL
> round(accuracy(tesla.holts.opt.pred$mean, valid.ts), 3)
               ME     RMSE      MAE    MPE   MAPE  ACF1 Theil's U
Test set 3413.616 4053.076 3413.616 46.241 46.241 0.571     2.094
>
> #HOLTS WINTER MODEL
> round(accuracy(tesla.holts.w.opt.pred$mean, valid.ts), 3)
               ME     RMSE      MAE    MPE   MAPE  ACF1 Theil's U
Test set 2309.679 3321.995 2808.361 26.361 41.378 0.477     1.823
>
> # Linear trend model
> round(accuracy(tesla.trend.pred$mean, valid.ts), 3)
              ME     RMSE     MAE    MPE   MAPE  ACF1 Theil's U
Test set 3983.95 4456.745 3983.95 57.425 57.425 0.537     2.405
>
> #Linear trend with seasonality
> round(accuracy(tesla.trend.seas.pred$mean, valid.ts), 3)
               ME     RMSE      MAE   MPE  MAPE  ACF1 Theil's U
Test set 3984.174 4453.409 3984.174 57.47 57.47 0.537     2.405
```

Figure:4

```
> #Quadratic trend          #
> round(accuracy(tesla.quad.trend.pred$mean, valid.ts), 3)
               ME     RMSE      MAE    MPE   MAPE  ACF1 Theil's U
Test set 1912.655 2397.244 1912.655 25.474 25.474 0.361     1.307
>
> #Quadratic trend with seasonality
> round(accuracy(tesla.quad.seas.trend.pred$mean, valid.ts), 3)
               ME     RMSE      MAE    MPE   MAPE ACF1 Theil's U
Test set 1913.863 2393.495 1913.863 25.527 25.527 0.36     1.305
>
> ## Two level Model with Regression Quadractic trend and seasonality +AR(1) model
> round(accuracy(trainq.trend.season.pred$mean + res.ar1.pred$mean, valid.ts), 3)
               ME     RMSE      MAE    MPE   MAPE  ACF1 Theil's U
Test set 1892.213 2386.382 1892.213 24.962 24.962 0.367     1.298
>
> ## Two level Model with Regression Quadractic  trend  +AR(1) model
> round(accuracy(tesla.quad.trend.pred$mean + res1.ar1.pred$mean, valid.ts), 3)
               ME     RMSE      MAE    MPE   MAPE  ACF1 Theil's U
Test set 1887.767 2388.777 1887.767 24.832 24.832 0.368     1.299
>
> #Two level Model with Regression Quadractic  trend and seasonality +Trailing MA(1)
> round(accuracy(fst.2level, valid.ts), 3)
               ME     RMSE      MAE    MPE   MAPE ACF1 Theil's U
Test set 1652.373 2189.015 1658.452 20.986 21.165 0.36     1.179
>
> #Auto ARIMA model.
> round(accuracy(train.auto.arima.pred$mean, valid.ts), 3)
               ME     RMSE      MAE    MPE   MAPE  ACF1 Theil's U
Test set 2113.212 2680.446 2113.212 27.355 27.355 0.444     1.401
>
> #ARIMA(2,1,2)(1,1,2) model;
> round(accuracy(train.arima.seas.pred$mean, valid.ts), 3)
               ME     RMSE      MAE    MPE   MAPE  ACF1 Theil's U
Test set 1847.625 2354.972 1847.625 24.119 24.119 0.392     1.242
```

Figure:5

Entire Data Set Forecast

After validating the validation forecast same methods were used to forecast for the entire data set

```
> ## Performance measure for ACCURACY OF THE TWO SES WITH  OPTIMAL ALPHA.
> round(accuracy(ses.opt.pred$fitted, tesla.ts),3)
             ME    RMSE     MAE   MPE   MAPE   ACF1 Theil's U
Test set 207.24 839.977 400.253 4.907 23.987 -0.058         1
>
> # Performance measures for HW forecast.
> round(accuracy(HW.ZZZ.pred$fitted, tesla.ts), 3)
             ME     RMSE     MAE    MPE   MAPE   ACF1 Theil's U
Test set 19.333 1310.291 713.201 -3.647 37.995 -0.287     0.983
>
> #Performance measures for Quadratic trend + SEASONALITY for entire data set
> round(accuracy(trend.season.pred$fitted, tesla.ts),3)
         ME    RMSE     MAE    MPE    MAPE  ACF1 Theil's U
Test set  0 674.026 499.444 -80.23 400.484 0.375     7.375
>
> #Performance measures for Quadratic trend  for entire data set
> round(accuracy(trend.quad.pred$fitted, tesla.ts),3)
         ME    RMSE     MAE     MPE    MAPE  ACF1 Theil's U
Test set  0 708.695 512.371 -78.534 330.217 0.353     8.152
>
> #Performance measures for Quadratic trend + Seasonality with AR(1) FOR ENTIRE DATA SET
> round(accuracy(tesla.total.quad.seas.trend.pred$fitted + res.total.ar1.pred$fitted, tesla.ts), 3)
            ME    RMSE     MAE     MPE    MAPE  ACF1 Theil's U
Test set 5.848 624.989 438.158 -59.152 335.944 0.114     5.493
>
```

Figure:6

```
> #Performance measures for Quadratic trend + with AR(1) FOR ENTIRE DATA SET
> round(accuracy(tesla.total.quad.trend.pred$fitted + res.total.ar1.pred$fitted, tesla.ts), 3)
            ME    RMSE     MAE     MPE    MAPE  ACF1 Theil's U
Test set 5.848 660.119 437.346 -57.457 228.691 0.109     4.523
>
> #Performance measures for Quadratic trend + TMA
> round(accuracy(tot.trend.seas.pred$fitted+tot.ma.trail.res, tesla.ts), 3)
             ME    RMSE     MAE    MPE    MAPE  ACF1 Theil's U
Test set 47.096 578.112 393.298 64.548 182.652 0.285     5.568
>
> #Performance measures for Linear trend model
> round(accuracy(tesla.lin.trend$fitted, tesla.ts), 3)
         ME     RMSE      MAE     MPE     MAPE  ACF1 Theil's U
Test set  0 1465.961 1218.945 974.846 1240.746 0.763    33.992
>
> #Performance measures for Linear trend with seasonality
> round(accuracy(tesla.lin.trend.seas$fitted, tesla.ts), 3)
         ME     RMSE      MAE     MPE     MAPE  ACF1 Theil's U
Test set  0 1442.067 1202.118 950.343 1193.824 0.777    36.925
```

Figure:7

```
> #Performance measure for Trailing MA with width of 2
> round(accuracy(ma.tesla.trailing_2.pred$fitted, tesla.ts), 3)
              ME    RMSE     MAE     MPE   MAPE  ACF1 Theil's U
Test set 299.274 944.616 483.241 -10.797 33.994 0.233     1.069
>
> #Performance measure for Trailing MA with width of 3
> round(accuracy(ma.tesla.trailing_3.pred$fitted, tesla.ts), 3)
              ME     RMSE     MAE    MPE   MAPE  ACF1 Theil's U
Test set 409.035 1012.653 533.369 -0.948 32.245 0.387     1.074
>
> #Performance measure for Trailing MA with width of 4
> round(accuracy(ma.tesla.trailing_4.pred$fitted, tesla.ts), 3)
              ME     RMSE     MAE   MPE   MAPE  ACF1 Theil's U
Test set 504.217 1063.414 574.315 6.532 32.303 0.472     1.078
>
> # Performance measure for Seasonal ARIMA (2,1,2)(1,1,2) Model,
> round(accuracy(arima.seas.pred$fitted, tesla.ts), 3)
              ME    RMSE     MAE   MPE   MAPE   ACF1 Theil's U
Test set 117.537 657.978 353.527 3.839 17.802 -0.025     0.943
>
> # Performance measure for Auto ARIMA Model,
> round(accuracy(auto.arima.pred$fitted, tesla.ts), 3)
              ME    RMSE     MAE   MPE   MAPE  ACF1 Theil's U
Test set 144.906 731.519 373.314 4.033 20.759 -0.13     0.989
>
> # Performance measure for Seasonal naive Model,
> round(accuracy((snaive(tesla.ts))$fitted, tesla.ts), 3)
              ME     RMSE    MAE   MPE   MAPE  ACF1 Theil's U
Test set 778.902 1390.706 818.35 28.05 39.159 0.541     1.155
>
> # Performance measure for NAive model
> round(accuracy((naive(tesla.ts))$fitted, tesla.ts), 3)
              ME    RMSE     MAE  MPE   MAPE  ACF1 Theil's U
Test set 211.752 848.501 408.123 7.16 22.324 -0.06         1
```

Figure:8

Looking at the above forecasts for the entire data set, following three models had the best accuracy performance and same will be used to gain more insights:

- 1. ARIMA(2,1,2)(1,1,0)
- 2. Auto Arima
- 3. Simple Exponential Smooting (SES model (A,N,N)

ARIMA Model

Autoregressive Integrated Moving Average (ARIMA) is a popular model in time series forecasting that can present any time series component – level (stationary), trend, and seasonality – or a combination of the same. It has a Complex structure, may include up to 6 parameters.

Automated ARIMA Model

Automated Arima model can be developed by the auto.Arima() function which gets the automated order of level trend and seasonality.

Training & Validation Data Summary

The output from the Auto ARIMA model for the training partition using summary function:

```
Series: train.ts
ARIMA(0,2,1)

Coefficients:
          ma1
      -0.8603
s.e.   0.0838

sigma^2 estimated as 40548:  log likelihood=-228.78
AIC=461.57   AICc=461.96   BIC=464.62

Training set error measures:
                   ME     RMSE      MAE        MPE     MAPE      MASE       ACF1
Training set 36.52531 192.7929 93.91746 -0.9223501 22.05049 0.2784514 -0.2486889
```

Figure:9

The above model incorporates a second order step of differencing and is used to forecast data with only level and trend components. It has moving average component (ma1) with the value of -0.8603. **s.e** stands for standard error of estimate, and it is 0.0838.

Equation: yt - yt-1 = -0.8603 et-1

Here, there no mean or intercept as the differencing eliminates the same.

- yt - yt-1 - difference is the outcome variable.
- et-1 is the error variable for residuals of the order 1 of moving average for error lag

Auto Arima model with model structure for respective time series data from 2008 Q4 to 2021 Q1

p	0	No autoregressive model (AR)
d	2	order 2 differencing to remove linear trend
q	1	order 1 moving average (MA2) model for error lags

Table:1

Plot for Training & Validation Forecast

From the plot below we can see that the model is fitting well into the data in training partition. They are taking trend into consideration. The plot is under forecasting the validation predictions.

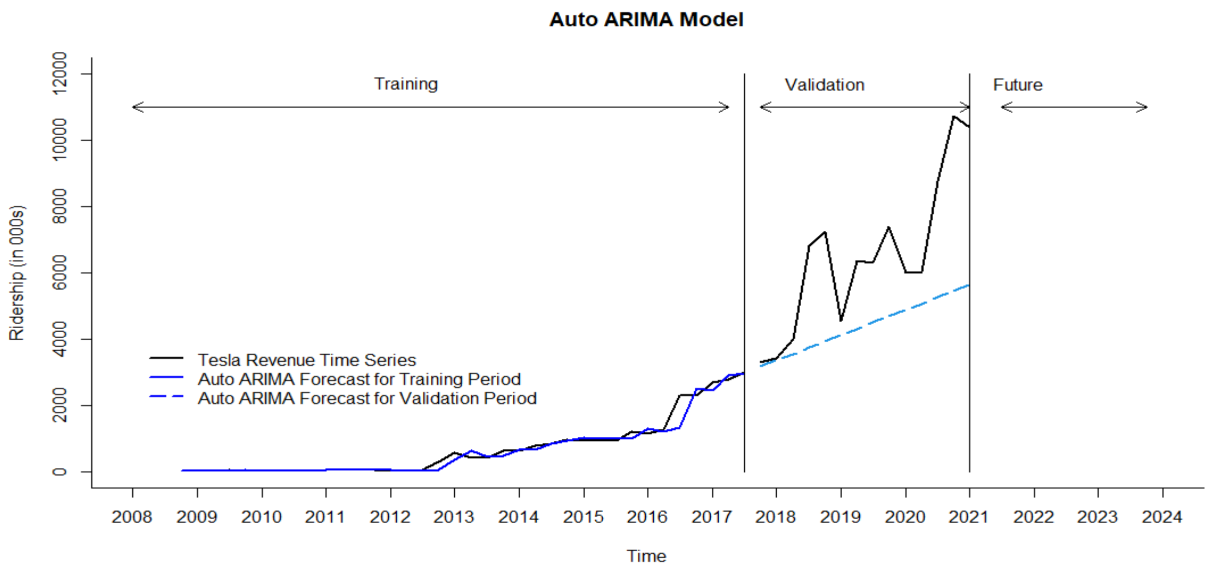

Figure:10

Forecast Validation Data

→ Forecast revenue for the validation period using the Auto Arima Model:

```
        Point Forecast     Lo 0     Hi 0
2017 Q4       3174.277 3174.277 3174.277
2018 Q1       3363.553 3363.553 3363.553
2018 Q2       3552.830 3552.830 3552.830
2018 Q3       3742.106 3742.106 3742.106
2018 Q4       3931.383 3931.383 3931.383
2019 Q1       4120.659 4120.659 4120.659
2019 Q2       4309.936 4309.936 4309.936
2019 Q3       4499.212 4499.212 4499.212
2019 Q4       4688.489 4688.489 4688.489
2020 Q1       4877.766 4877.766 4877.766
2020 Q2       5067.042 5067.042 5067.042
2020 Q3       5256.319 5256.319 5256.319
2020 Q4       5445.595 5445.595 5445.595
2021 Q1       5634.872 5634.872 5634.872
```

Figure:11

Accuracy of Training & Validation Data

Training RMSE	Validation RMSE	Training MAPE	Validation MAPE
192.793	2680.446	22.05	27.355

Table:2

From the above metrics, there is increase in RMSE and MAPE value from training to validation period and the model presents an overfit.

Entire Data Set Summary

Below is the summary output when auto.Arima() function is run on the entire time series dataset

```
Series: tesla.ts
ARIMA(0,1,0)(2,0,0)[4]

Coefficients:
        sar1    sar2
      0.1802  0.4532
s.e.  0.1370  0.1507

sigma^2 estimated as 569277:  log likelihood=-394.33
AIC=794.67   AICc=795.2   BIC=800.34

Training set error measures:
                   ME     RMSE      MAE      MPE     MAPE     MASE       ACF1
Training set 144.9063 731.5192 373.3143 4.033117 20.75939 0.456179 -0.1304743
```

Figure:12

The above model incorporates an additional seasonal term in the ARIMA model. The first set of parameters has first order step of differencing. The second set of parameters has order 2 of auto regressive seasonal model. The auto regressive coefficient of sar1 and sar2 is 0.1802 and 0.4532.

s.e stands for standard error of estimate, and it is 0.1370.

Equation: $\mathbf{y_t - y_{t-1} = 0.1802\ (y_{t-1} - y_{t-5}) + 0.4532\ (y_{t-1} - y_{t10})}$

Here, there no mean or intercept as the differencing eliminates the same.

- **yt - yt-1** - difference is the outcome variable.
- **yt-1 -yt-5** - difference variable for first predictor in seasonal AR model of order 1 with yt-5 indicating quarterly seasonality
- **yt-1 -yt-10** - difference variable for first predictor in seasonal AR model of order 2 with yt-10 indicating quarterly seasonality

Auto Arima model with model structure for entire time series data from 2008 Q4 to 2021 Q1

p	0	No autoregressive model (AR)
d	1	order 1 differencing to remove linear trend
q	0	order 0 moving average (MA2) model for error lags
P	2	order 2 autoregressive model (AR1) for seasonality
D	0	order 0 differencing to remove linear trend
Q	0	order 0 moving average (MA2) model for error lags
m	4	For quarterly seasonality

Table:3

Plot for Entire Data Set Forecast

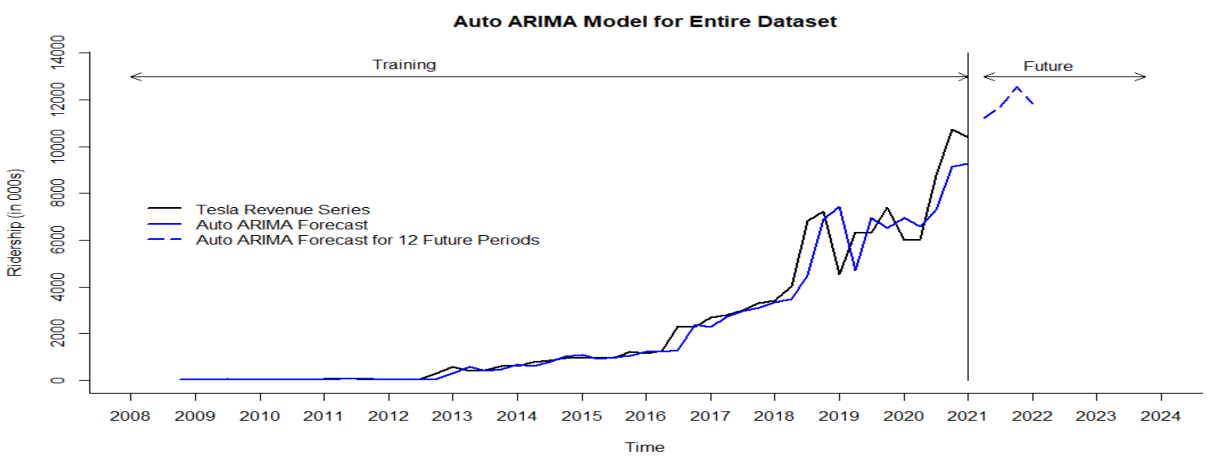

Figure:13

From the plot below we can see that the model is fitting well into the historical data. They are taking trend into consideration with higher revenue being predicted for future quarters.

Forecast Future Period

→ Forecast revenue for the future 4 period using the Auto Arima Model:

```
        Point Forecast     Lo 0     Hi 0
2021 Q2       11219.07 11219.07 11219.07
2021 Q3       11690.52 11690.52 11690.52
2021 Q4       12535.21 12535.21 12535.21
2022 Q1       11838.09 11838.09 11838.09
```

Figure:14

Autocorrelation for Residuals

Acf() function can be used to create autocorrelation chart for entire data set.

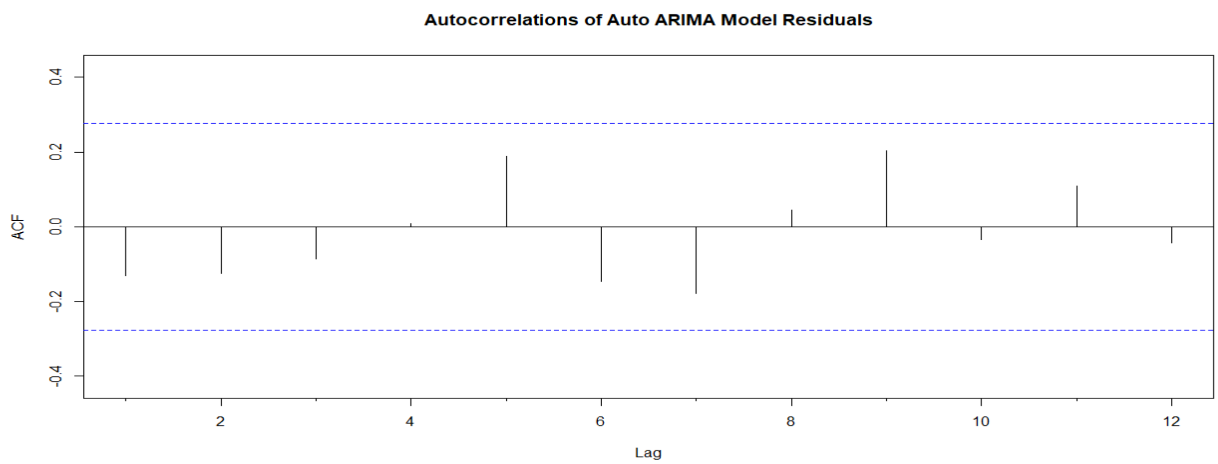

Figure:15

The auto correlation coefficient for lags 1 to 12 are below the upper and lower threshold limits. They are not statistically significant, nor they are indicating any successive relationships and are just noise or randomness.

Accuracy Measure of Entire Data Set

The MAPE measure shows 20.759% overfitting with ACF being well incorporated for forecasting and with a high RMSE value of 731.519.

```
                 ME    RMSE     MAE   MPE   MAPE  ACF1 Theil's U
Test set 144.906 731.519 373.314 4.033 20.759 -0.13     0.989
```

Figure:16

ARIMA (2,1,2) (1,1,0) Model

Arima (2,1,2)(1,1,0) model can be developed by feeding in the details of the order of the autoregressive model, differencing order to remove trend, moving average order for error lags. The same set of seasonal parameters can also be developed.

Training & Validation Data Summary

```
Series: train.ts
ARIMA(2,1,2)(1,1,0)[4]

Coefficients:
         ar1      ar2      ma1     ma2     sar1
      0.9335  -0.6634  -1.1186  1.0000  -0.7253
s.e.  0.2132   0.1643   0.1515  0.1787   0.1462

sigma^2 estimated as 47192:  log likelihood=-210.74
AIC=433.48   AICc=436.98   BIC=442.09

Training set error measures:
                   ME     RMSE      MAE      MPE     MAPE      MASE        ACF1
Training set 36.66545 184.6157 98.96827 2.812535 18.97441 0.2934262 -0.08445483
```

Figure:17

ARIMA (2,1,2)(1,1,0) [4] model is seasonal ARIMA model where:

1.the first set of parameters (2,1,2) indicates:

- 2-order of the auto regressive model with number of auto correlation lags

- 1-order of lag-1 differencing to remove trend
- 2-order of 2 moving average for number of residuals' auto correlation lags

2. the second set of parameters (1,1,0) indicates:

- 1-order of the auto regressive seasonal model with number of auto correlation lags
- 1-order of lag-1 differencing in seasonal auto regressive model to offset/remove any newly developed trend or correlation coefficient
- 0-order of 0 seasonal moving average for number of residuals' auto correlation lags
- [4] -quarterly seasonality.
- s.e stands for standard error of estimate, and it is 0.0838.

Equation: yt - yt-1 = 0.9335 $\mathbf{(y_{t-1} - y_{t-2})}$ **-0.6634** $\mathbf{(y_{t-2} - y_{t-3})}$ **-1.1186** $\mathbf{e_{t-1}}$**+1** $\mathbf{e_{t-2}}$ **-0.7253** $\mathbf{(y_{t-1} - y_{t-5})}$

Here, there no mean or intercept as the differencing eliminates the same.

- 0.9335 is the auto regressive coefficient of the AR model of order 2.
- **-0.6634** is the auto regressive coefficient of the AR model of order 2.
- **-1.1186**is the coefficient of the order 2 of moving average for error lags.
- 1 is the coefficient of the order 2 of moving average for error lags.
- **0.7253** is the auto regressive coefficient of the seasonal AR model of order 1.

Plot for Training & Validation Forecast

From the plot below we can see that the model is fitting well into the data in training partition. They trend seems upward. The plot is under forecasting the validation predictions.

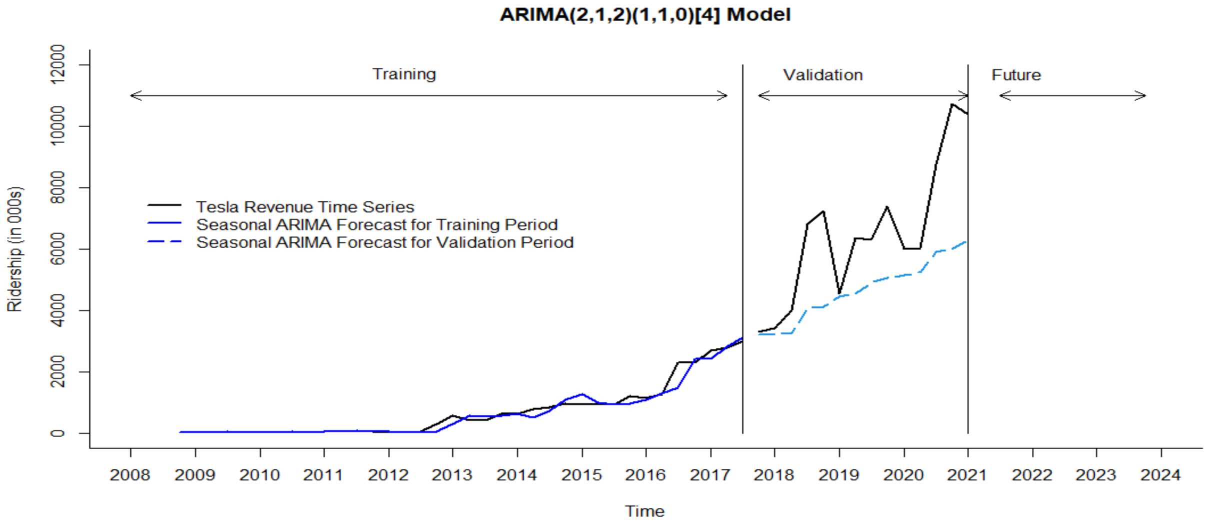

Figure:18

Forecast for Validation Period

→ Forecast revenue for the validation period using the Arima (2,1,2)(1,1,0) Model:

	Point Forecast	Lo 0	Hi 0
2017 Q4	3202.613	3202.613	3202.613
2018 Q1	3218.586	3218.586	3218.586
2018 Q2	3275.121	3275.121	3275.121
2018 Q3	4051.730	4051.730	4051.730
2018 Q4	4119.868	4119.868	4119.868
2019 Q1	4453.875	4453.875	4453.875
2019 Q2	4555.229	4555.229	4555.229
2019 Q3	4905.575	4905.575	4905.575
2019 Q4	5066.260	5066.260	5066.260
2020 Q1	5157.758	5157.758	5157.758
2020 Q2	5226.063	5226.063	5226.063
2020 Q3	5892.914	5892.914	5892.914
2020 Q4	5993.706	5993.706	5993.706
2021 Q1	6262.951	6262.951	6262.951

Figure:19

Accuracy Measure for Training & Validation

Training RMSE	Validation RMSE	Training MAPE	Validation MAPE
184.616	2354.972	18.974	24.119

Table:4

From the above metrics, there is very high increase in RMSE and small increase in MAPE value from training to validation period and the model presents some overfit.

Entire Data Set Summary

The output from summary of the model however results with model parameters as ARIMA (2,1,2)(1,1,0).

```
Series: tesla.ts
ARIMA(2,1,2)(1,1,0)[4]

Coefficients:
         ar1      ar2      ma1    ma2     sar1
      1.5719  -0.6099  -1.9663  1.000  -0.6146
s.e.  0.1421   0.1571   0.1104  0.108   0.1334

sigma^2 estimated as 541169:  log likelihood=-361.26
AIC=734.51   AICc=736.73   BIC=745.35

Training set error measures:
                   ME     RMSE      MAE      MPE     MAPE      MASE        ACF1
Training set 117.5369 657.978 353.5268 3.839026 17.80211 0.4319993 -0.02530699
```

Figure:20

s.e stands for standard error of estimate, and it is 0.1108.

Equation: yt - yt-1 = 1.5719 $\mathbf{(y_{t-1} - y_{t-2})}$ **-0.6099** $\mathbf{(y_{t-2} - y_{t-3})}$ **-1.9663**$\mathbf{e_{t-1}}$

$+1\ e_{t-2}$ $-0.6146\ (y_{t-1}\ -y_{t-5})$

Here, there no mean or intercept as the differencing eliminates the same.

- 1.5719 is the auto regressive coefficient of the AR model of order 2.
- **-0.6099** is the auto regressive coefficient of the AR model of order 2.
- **-1.9663** is the coefficient of the order 2 of moving average for error lags.
- 1 is the coefficient of the order 2 of moving average for error lags.
- **-0.6146** is the coefficient of the order 1 of seasonal moving average for error lags.

Plot for Entire Data Set Forecast

From the plot below we can see that the model is fitting well into the historical data. They are taking trend into consideration with higher revenue being predicted for future quarters.

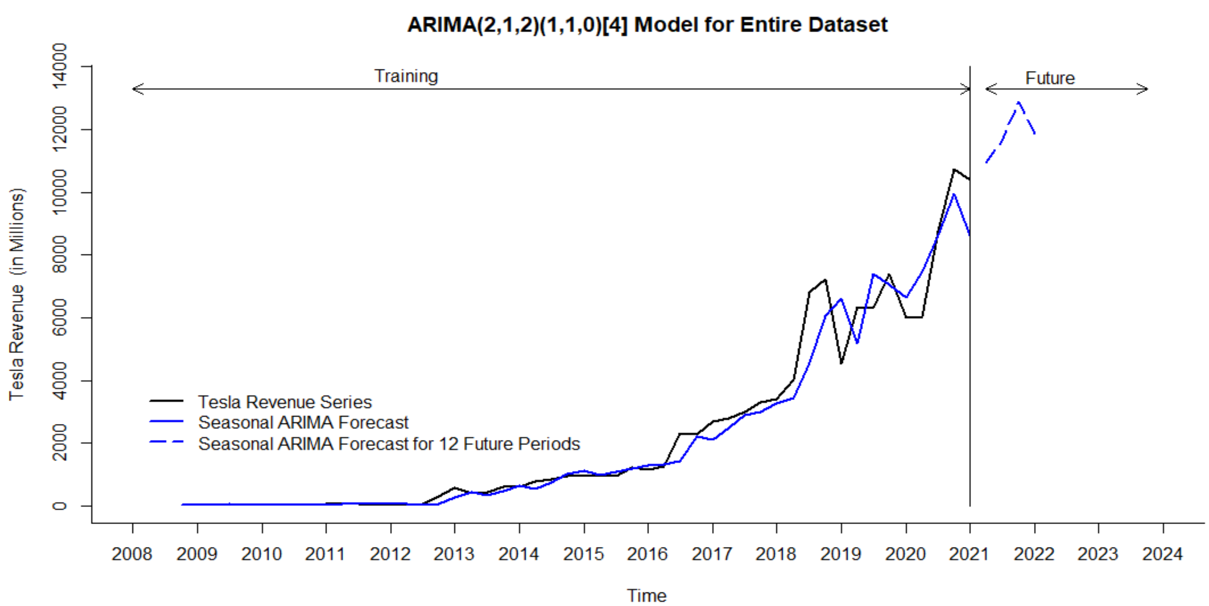

Figure:21

Forecast for Validation Period

→ Forecast revenue for the validation period using the Auto Arima Model:

```
        Point Forecast    Lo 0     Hi 0
2021 Q2       10928.85 10928.85 10928.85
2021 Q3       11622.57 11622.57 11622.57
2021 Q4       12886.36 12886.36 12886.36
2022 Q1       11843.12 11843.12 11843.12
```

Figure:22

Autocorrelation for Residuals

Acf() function can be used to create autocorrelation chart for entire data set.

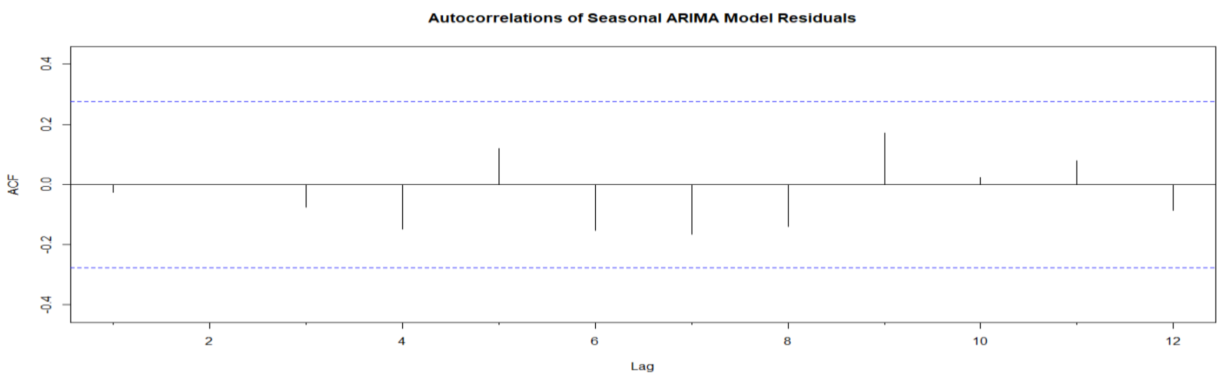

Figure:23

The auto correlation coefficient for lags 1 to 12 are below the upper and lower threshold limits. They are not statistically significant, nor they are indicating any successive relationships and are just noise or randomness. Hence, ACF is well incorporated for forecasting.

Accuracy Measure of Entire Data Set

The MAPE measure shows 17.802% overfitting with ACF being well incorporated for forecasting and with a high RMSE value of 657.978.

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Test set	117.537	657.978	353.527	3.839	17.802	-0.025	0.943

Figure:24

Exponential Smoothing (SES)

Exponential smoothing (SES) is method that estimate time series components from the data and smoothens out the noise in a series and average series values over multiple time periods.

Simple Exponential Smoothing (SES)

Simple exponential smoothing (SES) is flexible method that averages historical data periods to smooth out noise. It takes a weighted average of past historical data periods, so that the weights decrease exponentially into the past and gives more weight to recent historical data periods.

Training & Validation Data Summary

A summary of the SES model with automated selection of the smoothing parameters:

```
ETS(A,N,N)

Call:
 ets(y = train.ts, model = "ANN")

  Smoothing parameters:
    alpha = 0.9999

  Initial states:
    l = 29.0981

  sigma:  217.1128

     AIC     AICc      BIC
520.3390 521.0890 525.0896
```

Figure:25

This SES model has the (A, N, N) options, i.e., additive error, no trend, and no additive seasonality. The optimal value for exponential smoothing constant (alpha) is 0. 9999.The alpha value of this model indicates that the model's level component tends to be more local.

Forecast

SES model's forecast in the validation period is presented below:

```
        Point Forecast   Lo 0    Hi 0
2017 Q4        2984.98 2984.98 2984.98
2018 Q1        2984.98 2984.98 2984.98
2018 Q2        2984.98 2984.98 2984.98
2018 Q3        2984.98 2984.98 2984.98
2018 Q4        2984.98 2984.98 2984.98
2019 Q1        2984.98 2984.98 2984.98
2019 Q2        2984.98 2984.98 2984.98
2019 Q3        2984.98 2984.98 2984.98
2019 Q4        2984.98 2984.98 2984.98
2020 Q1        2984.98 2984.98 2984.98
2020 Q2        2984.98 2984.98 2984.98
2020 Q3        2984.98 2984.98 2984.98
2020 Q4        2984.98 2984.98 2984.98
2021 Q1        2984.98 2984.98 2984.98
```

Figure:26

Accuracy Measure for Training & Validation

Training RMSE	Validation RMSE	Training MAPE	Validation MAPE
210.996	4182.42	26.243	47.972

Table:5

From the above metrics, there is very high increase in RMSE and in MAPE value from training to validation period and the model presents some overfit.

Entire Data Set Summary

A summary of the SES model with automated selection of the smoothing parameters:

```
ETS(A,N,N)

Call:
 ets(y = tesla.ts, model = "ANN")

  Smoothing parameters:
    alpha = 0.9999

  Initial states:
    l = 29.1025

  sigma:  857.298

     AIC     AICc      BIC
874.9386 875.4603 880.6747
```

Figure:27

This SES model has the (A, N, N) options, i.e., additive error, no trend, and no additive seasonality. The optimal value for exponential smoothing constant (alpha) is 0.9999. The alpha value of this model indicates that the model's level component tends to be more local.

Plot for Entire Data Set Forecast

The plot below suggests that the model is fitting well into the historical data. They are taking trend into consideration with revenue for future quarters equal to the latest historical quarters.

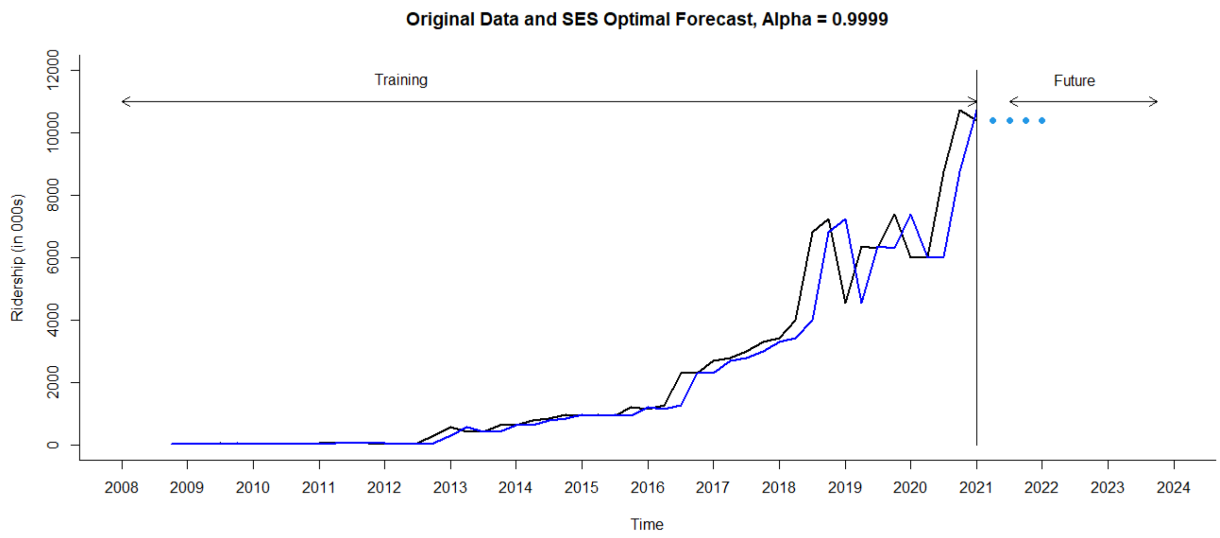

Figure:28

Forecast

SES model's forecast in the validation period is presented below:

	Point Forecast	Lo 0	Hi 0
2021 Q2	10390.04	10390.04	10390.04
2021 Q3	10390.04	10390.04	10390.04
2021 Q4	10390.04	10390.04	10390.04
2022 Q1	10390.04	10390.04	10390.04

Figure:29

Accuracy Measure for Entire Data set

The MAPE measure shows 23.987% overfitting with ACF being well incorporated for forecasting and with a high RMSE value of 839.977.

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Test set	207.24	839.977	400.253	4.907	23.987	-0.058	1

Figure:30

Conclusion

After developing the models, it is necessary to compare the accuracy performance measures.

1. Auto Arima Model

```
> round(accuracy(auto.arima.pred$fitted, tesla.ts), 3)
              ME    RMSE     MAE   MPE   MAPE  ACF1 Theil's U
Test set 144.906 731.519 373.314 4.033 20.759 -0.13     0.989
```

Figure:36

2. ARIMA(2,1,2)(1,1,0)

```
> round(accuracy(arima.seas.pred$fitted, tesla.ts), 3)
              ME    RMSE     MAE   MPE   MAPE   ACF1 Theil's U
Test set 117.537 657.978 353.527 3.839 17.802 -0.025     0.943
```

Figure:37

3. Simple Exponential Smoothing

```
> round(accuracy(ses.opt.pred$fitted, tesla.ts),3)
             ME    RMSE     MAE   MPE   MAPE   ACF1 Theil's U
Test set 207.24 839.977 400.253 4.907 23.987 -0.058         1
```

Figure:38

4. Naïve Forecast

```
> round(accuracy((naive(tesla.ts))$fitted, tesla.ts), 3)
              ME    RMSE     MAE  MPE   MAPE  ACF1 Theil's U
Test set 211.752 848.501 408.123 7.16 22.324 -0.06         1
```

Figure:39

Analysis On Results

On comparing the above 4 models, ARIMA(2,1,2)(1,1,0) is the best model with the lowest MAPE and RMSE values (APE:17.802% & RMSE :657.978)

Recommendation

The fact that ARIMA(2,1,2)(1,1,0) has a seasonal component included is beneficial in deciding a model for forecast. We can add external factors like production time series data to make final moel more accurate.

Remarks

There were few observations during the development of the Models.

- Most of the future forecast results indicated higher revenue than the latest historical data
- Auto correlation was well incorporated into the modelling without any significant relationships left behind

Limitations

- Most of the Regression Models displayed very high RMSE values and MAPE values indicating overfitting
- Models like 2 level model using Trailing MA and AR(1) with optimal performance measure in Training and validation partitions had high MAPE and RMSE values while forecasting for the entire data set

Benefits

- Models like Auto Arima and Arima (2,1,2)(1,1,0) had performance accuracies very close to one another indicating that more than 1 model could confirm the future predictions
- The above 3 mentioned models showed consistent measures in training and validation measurement and with entire data set.
- Both the ARIMA models handled the seasonality component along with trend which helped in better forecasting

Bibliography

- To get better understanding of measuring Forecasting error below URL was referred:

  https://towardsdatascience.com/forecast-kpi-rmse-mae-mape-bias-cdc5703d242d

- The data set was obtained from the below URL:

  https://ycharts.com/companies/TSLA/revenues