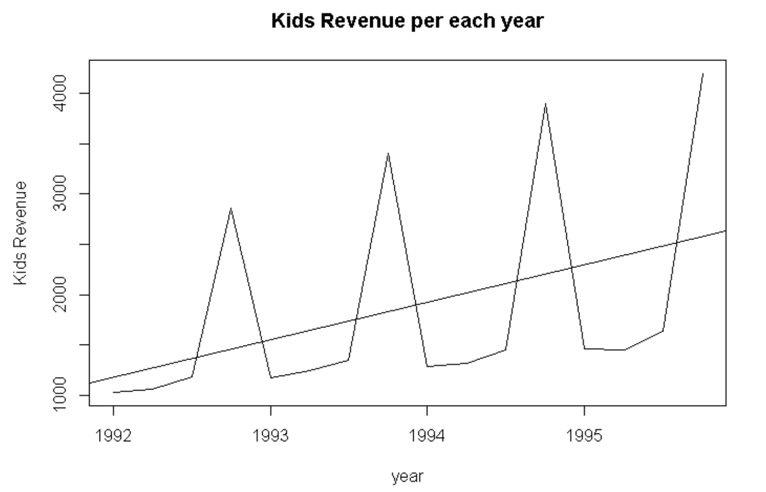
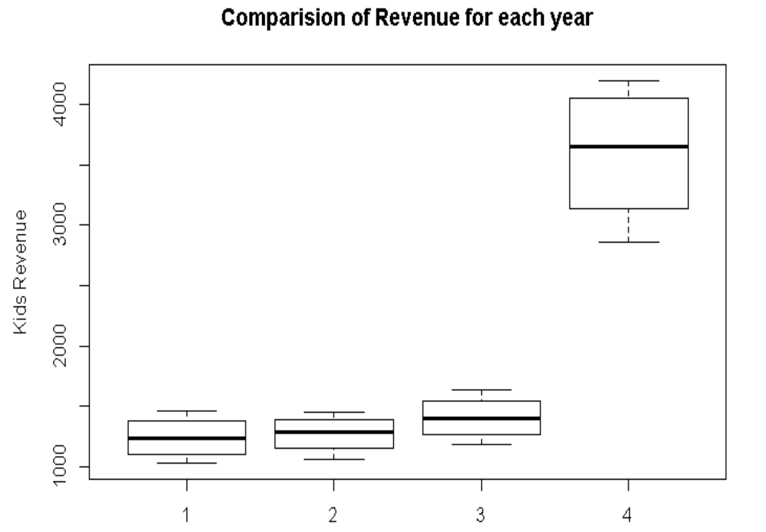
Forecasting of Kids Revenue

By

Niranjan	(D18023)
Sasidhar	(D18014)
Sravanya	(D18039)
Narayana	(D18030)
Anil	(D18035)
Satish	(D18036)





Exploratory Data Analysis

- The year on year trend clearly shows that the revenue has been increasing without fail.
- A Spike in revenue in fourth quarter of every year is observed.

Exploratory Data Analysis

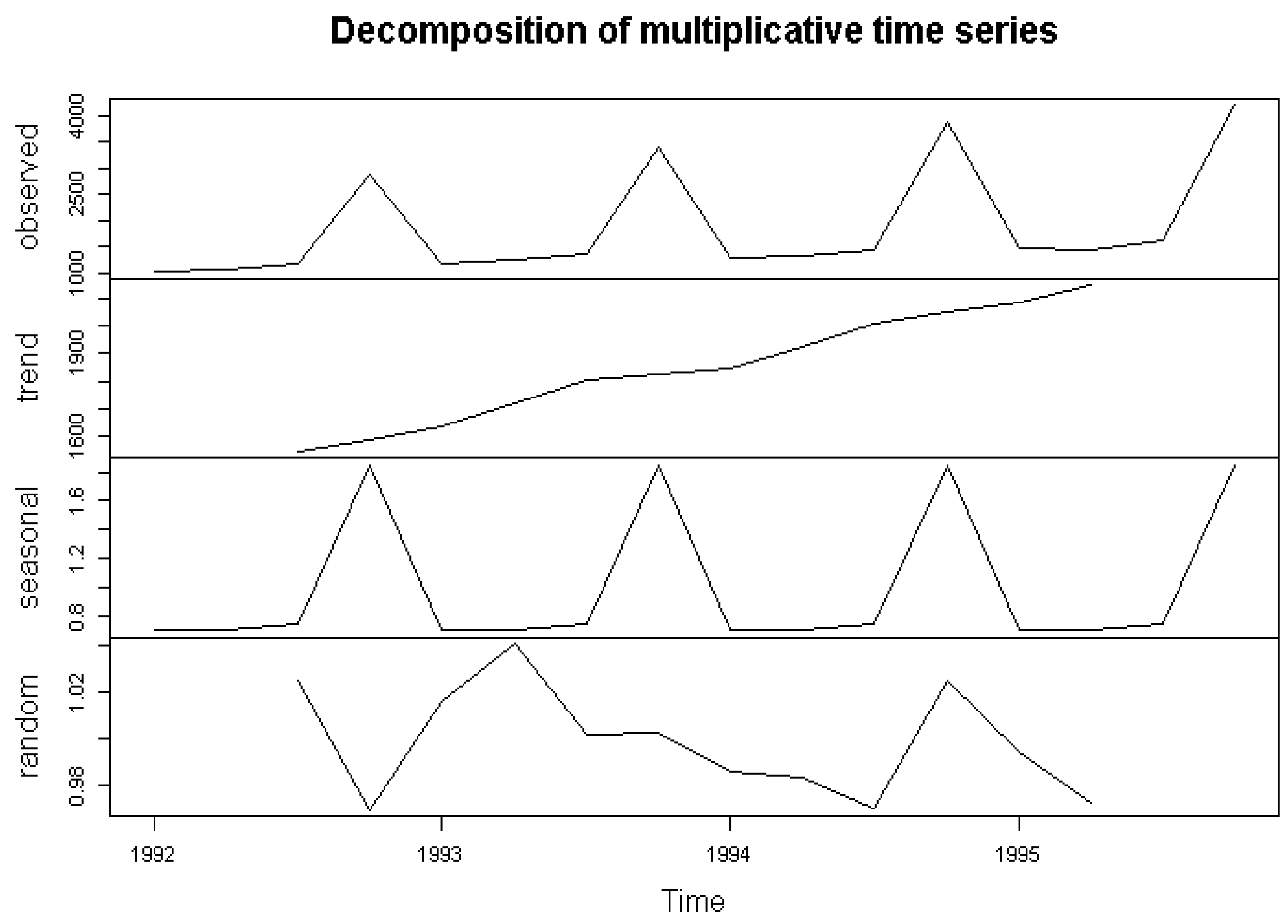
- Decomposition is a combination of level, trend, seasonality, and noise components.

- Decomposition procedures are used in time series to describe the trend and seasonal factors.

- Multiplicative Model

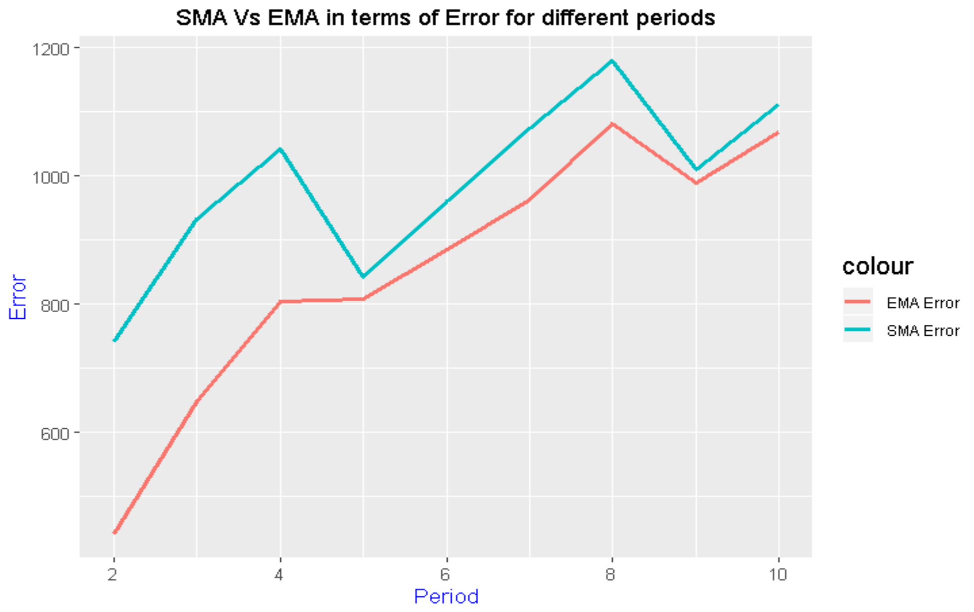
y(t) = Level * Trend * Seasonality * Noise

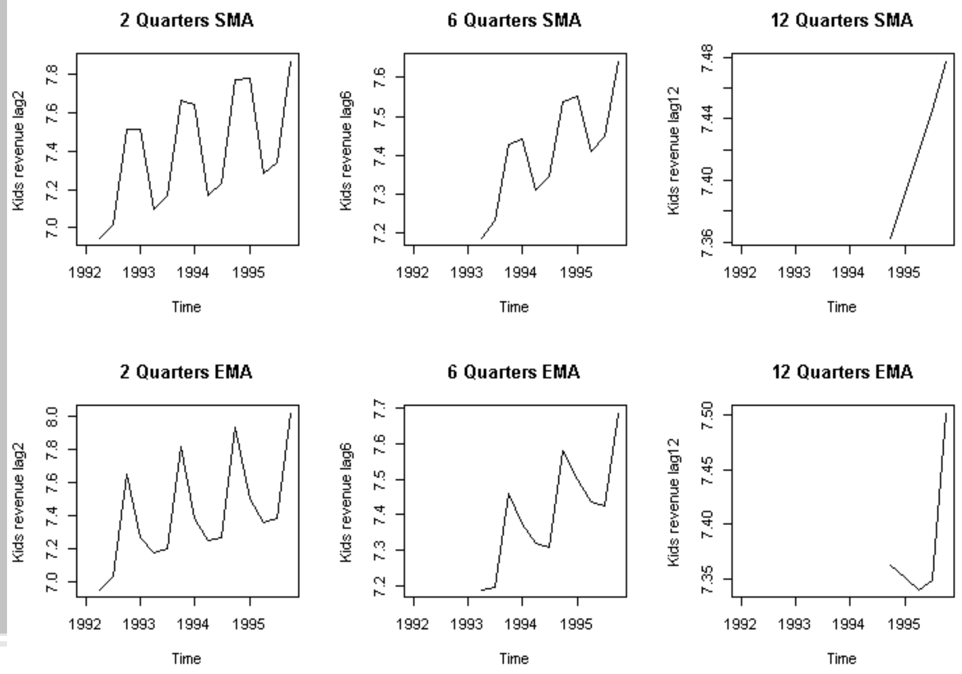
- We can see that the trend and seasonality information extracted from our data shows that revenue increases overtime.

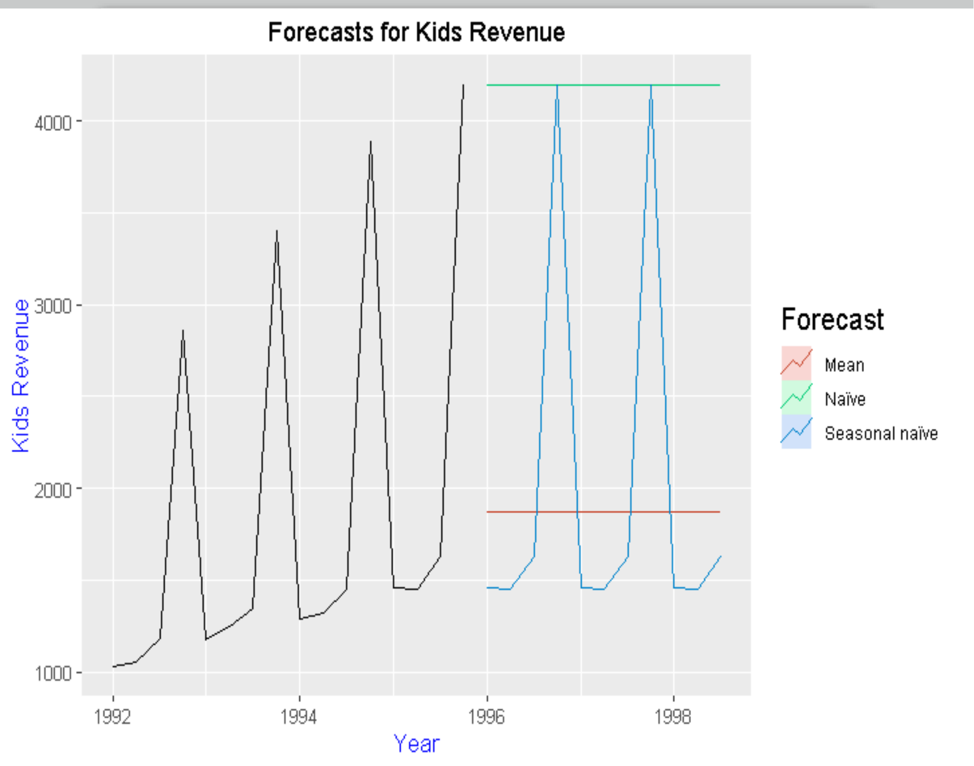


Using Moving Average

- SMA gives overall trend based on average of historical values in a given time frame.
- Moving averages smoothens the kids revenue data to form a trend, they do not predict price direction, but rather define the current direction with a lag.
- EMA gives higher weightage to recent revenue, while the SMA assigns equal weightage to all values.








```
> forecast_seasonalnaive
        Point Forecast     Lo 80    Hi 80     Lo 95    Hi 95
1996 Q1           1462 1126.1704 1797.830  948.3929 1975.607
1996 Q2           1452 1116.1704 1787.830  938.3929 1965.607
1996 Q3           1631 1295.1704 1966.830 1117.3929 2144.607
1996 Q4           4200 3864.1704 4535.830 3686.3929 4713.607
1997 Q1           1462  987.0652 1936.935  735.6499 2188.350
1997 Q2           1452  977.0652 1926.935  725.6499 2178.350
1997 Q3           1631 1156.0652 2105.935  904.6499 2357.350
1997 Q4           4200 3725.0652 4674.935 3473.6499 4926.350
```

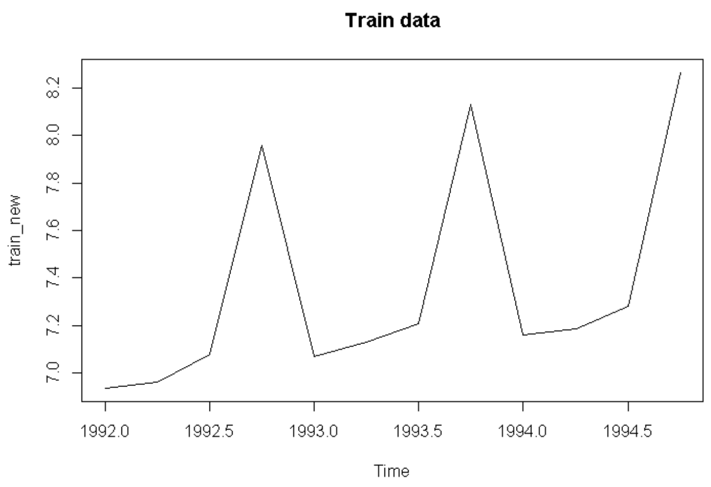
Naïve Methods

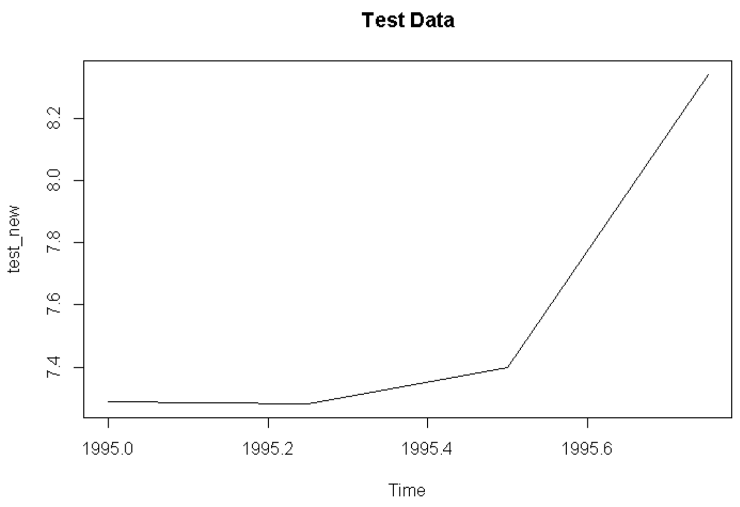
- **Naive Forecast :** Method uses the most recent observation for forecasting. In our case forecast value for the revenue is $4200 (in millions) . Since it is the last observation.

- **Seasonal Naive :** Method uses the last value from the same season for forecasting

- **Mean :** Method is the Mean of historical data, mean revenue forecasted for current data is $1874 (in million)

ARIMA Model

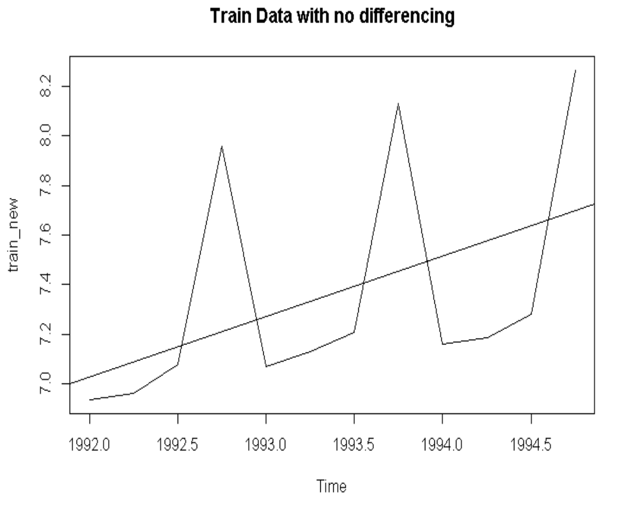


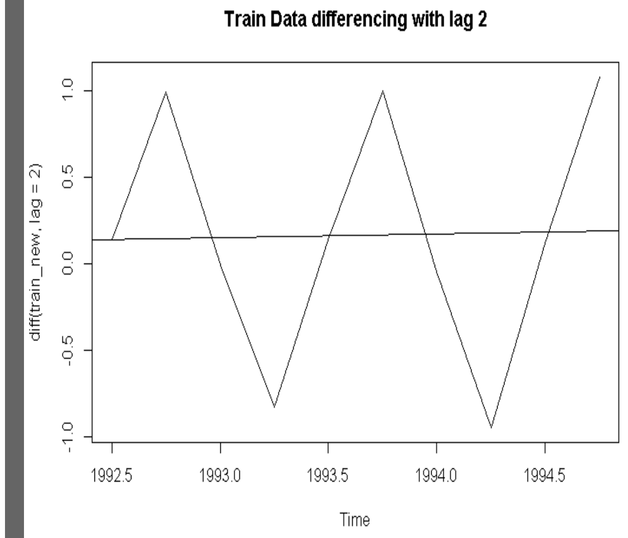


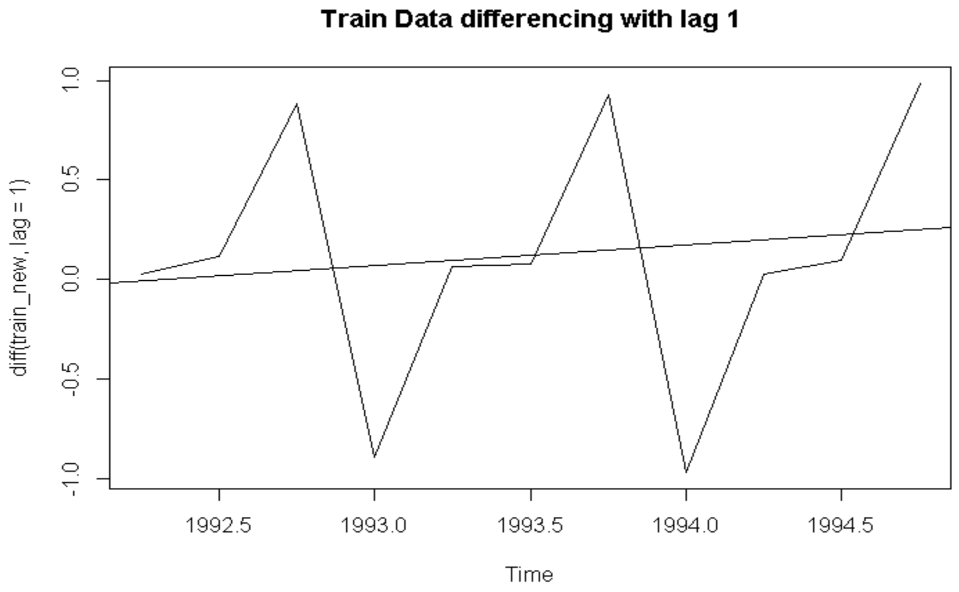
- **ARIMA :**Autoregressive Integrated Moving Averages
- Present data contains revenue information of 4 years across the 4 quarters.
- First 3 years data has been taken for training and last 1 year data has been taken for testing.
- Calculating AR and MA part required stationary time series data which is achieved in I through differencing.

ARIMA Model: Integrated

- We have applied logarithmic transformation to stabilize the variance in the data
- Based on the plots we could see that the data shows the variation in the plot.
- Differencing can help to stabilize the time series data by eliminating the trend and seasonality.
- In Kids revenue we converted data to stationary. By differencing the data with lag 1
- ADF Test with p – value (0.01) less than 0.05, also confirms that differenced data is stationary.

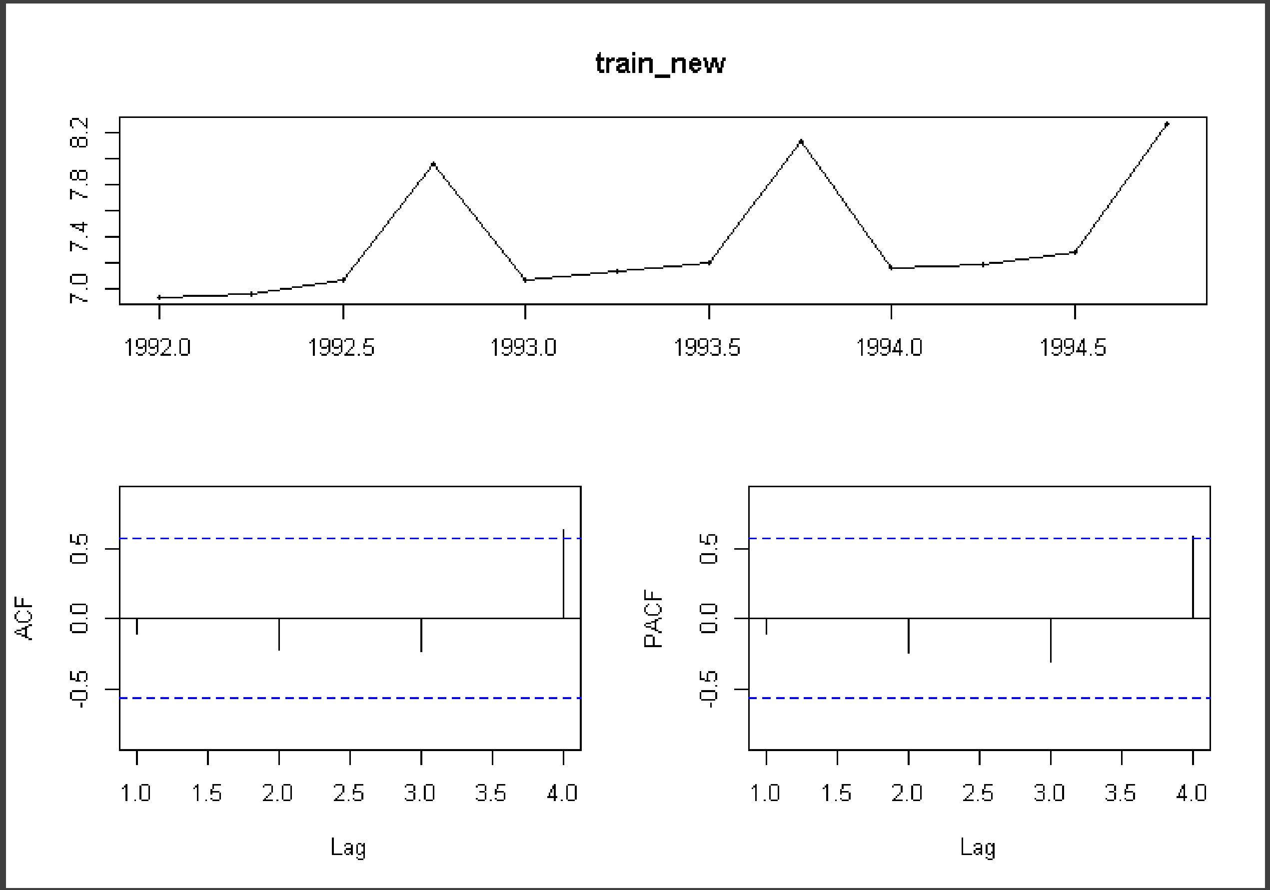






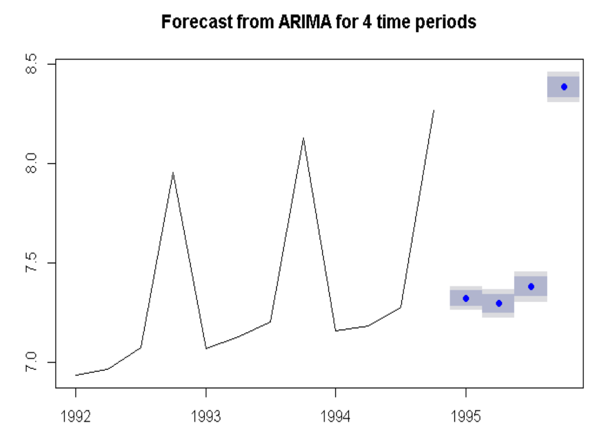
ACF and PACF Values:

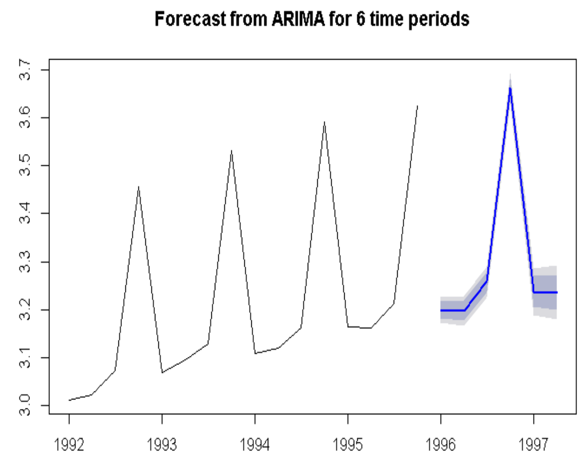
- **p:** The number of lag observations included in the model, also called the lag order.
- **q:** The size of the moving average window, also called the order of moving average.
- From ACF we can see that lag of 4 has strong correlation between values of the same variable across observations.
- From PACF, we can see that lag of 4 shows that there is strong association between two variables while adjusting for effect.
- ARIMA Model – (4,1,4)



ARIMA Model Forecasting

- Using p, d, q Values, we calculated the Test RMSE on year 1995 for 4 quarters.
- We have forecasted revenue for the next 6 quarters Q1 1996 – Q2 1997.



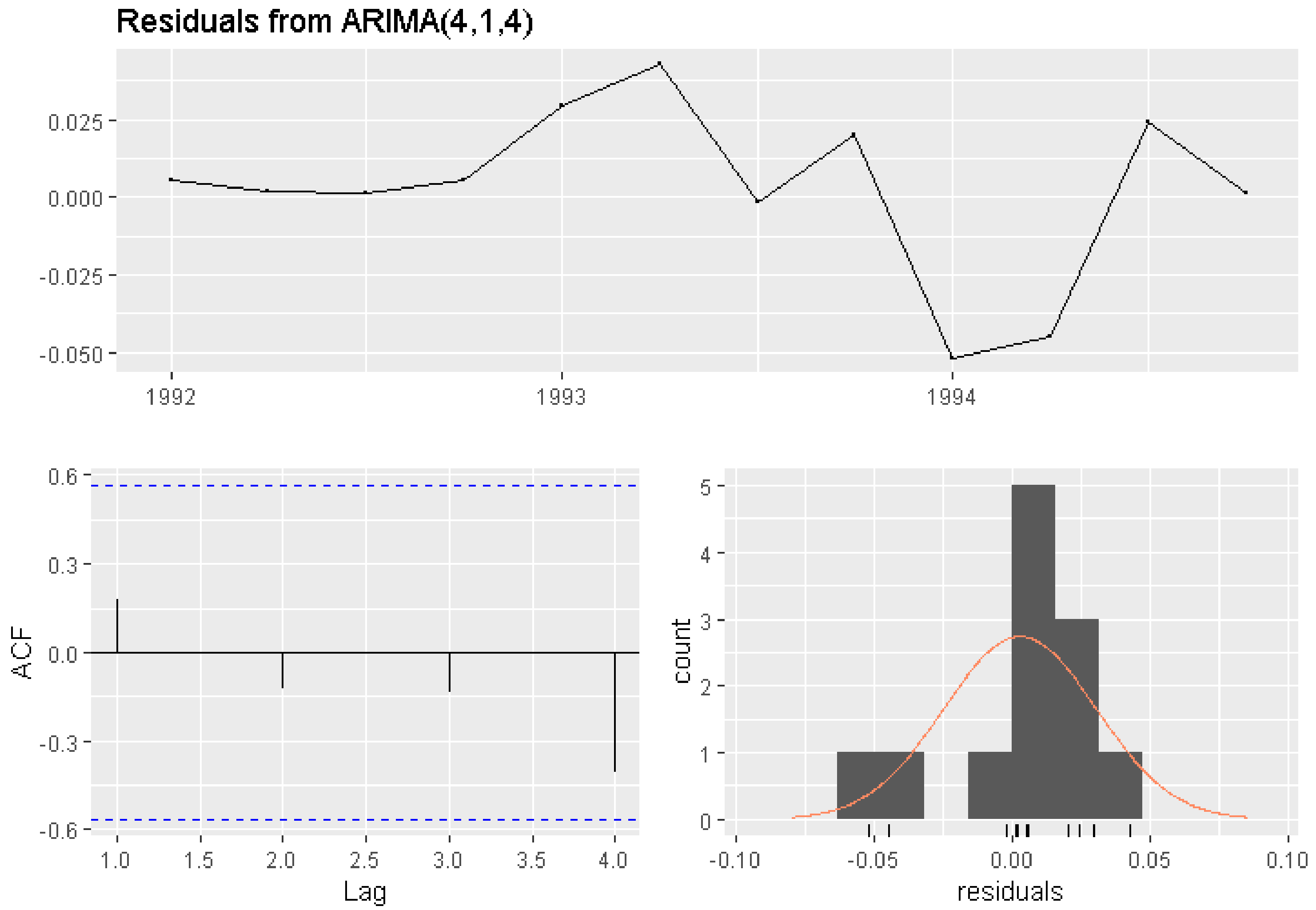


QuarterYear	Revenue(in million $)	Predicted
Q1-95	1462	1481.312
Q2-95	1452	1440.49
Q3-95	1631	1568.2
Q4-95	4200	4278.52
	RMSE	51.51375

```
> 10^pred$pred
         Qtr1     Qtr2     Qtr3     Qtr4
1996 1582.233 1577.702 1812.492 4583.581
1997 1727.559 1719.674
>
```

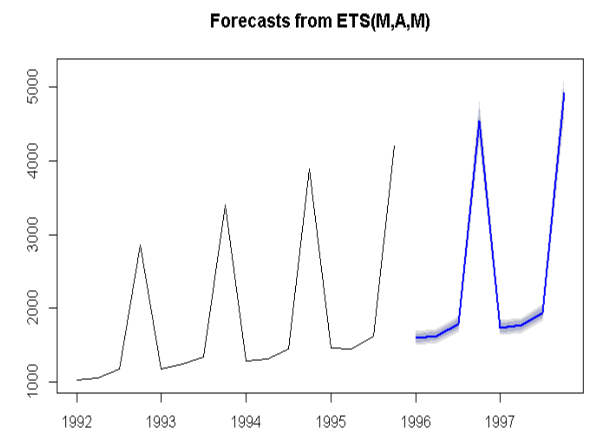
Using ARIMA Model

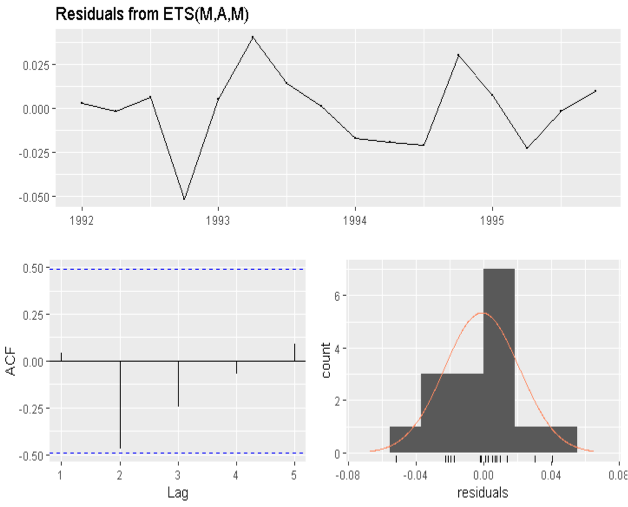
- **Residual errors:**
  - Amount of variability in a dependent variable (DV) that is "left over" after accounting for the variability explained by the predictors
  - We can see that the distribution does have a Gaussian look, that shows that errors are randomly distributed.
  - From the residual ACF plot, we could see that there is no correlation among the residuals.
  - As there are no spikes outside the insignificant zone for ACF plots, we can conclude that residuals are random with no information or juice in them.



ETS Model

- ETS is an approach for detecting additive errors and seasonal structures with prediction of interval coverages.
- In the current data, we built ETS (M,A,M) Multiplicative Holt-Winter's method with Multiplicative Errors.
- We have compared Error for ETS and ARIMA Model using cross validation, ARIMA Model gives less error compared to ETS.

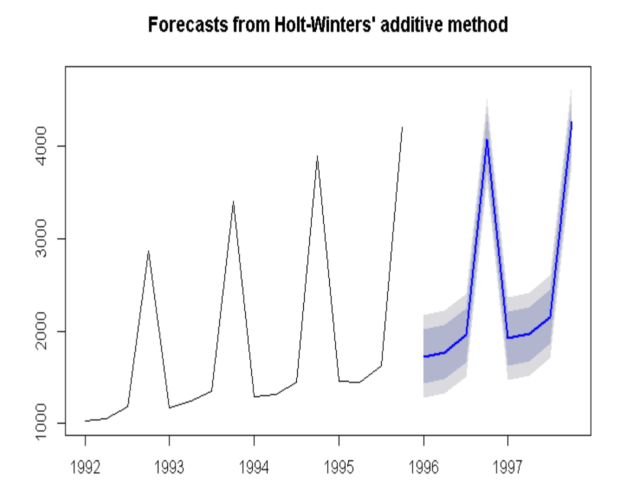


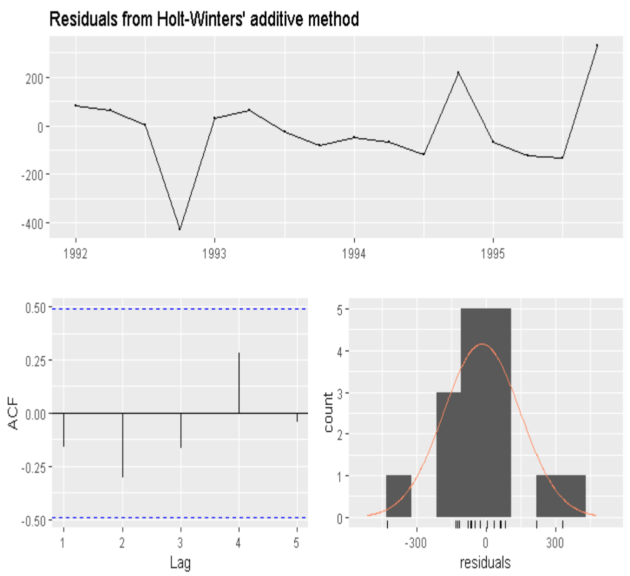



```
> mean(etserror^2, na.rm=TRUE)
[1] 8405885
> mean(arimaerror^2, na.rm=TRUE)
[1] 93955.99
>
```

Holt Winter Model

- Holt Winter also known as Triple Exponential Smoothing Method.
- Holt Winter additive is an extension of Holt exponential smoothing that captures seasonality.






```
> forecast(hw(kids_ts, h = 8))
        Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
1996 Q1       1723.594 1430.954 2016.235 1276.040 2171.149
1996 Q2       1772.144 1479.503 2064.784 1324.589 2219.698
1996 Q3       1956.510 1663.870 2249.151 1508.956 2404.065
1996 Q4       4066.387 3773.746 4359.028 3618.832 4513.942
1997 Q1       1918.367 1625.726 2211.008 1470.811 2365.922
1997 Q2       1966.916 1674.275 2259.557 1519.360 2414.472
1997 Q3       2151.283 1858.642 2443.924 1703.727 2598.839
1997 Q4       4261.160 3968.518 4553.801 3813.603 4708.717
>
```

Recommendations

- For given set of data , we have built and compared multiple time series forecasting techniques.
- From the error comparison table we can arrive to the conclusion that **ARIMA** model gives less RMSE, so it's the best fit for the given data.

Model	RMSE	MAPE
Moving Average - 2	740	55.2
Exponential Average - 2	441	25.4
Naïve	1481	51.1
SNaive	262	10.9
Mean	1032	46.7
ETS	53.2	2.34
HW - Additive	161	5.7
ARIMA	51.5	1.96

Code

The following files has the code for entire project –

- Code to Build ARIMA Model - Forecasting_Kids_ARIMA.R
- Code for EDA and other modelling techniques (Naïve, SMA, EMA, ETS and HW) at Forecasting_KidsRevenue.R