```
<h3> Submitted By : <br> Siddarth G <br> Reg No: 19122013 </h3>
```

The following dataset is a monthly count of riders for the Portland public transportation system. The website states that it is from January 1960 through June 1968

In [104]:
```
data<-read.table(file="C:/Users/Downloads/Temp/portland-oregon-aver
```

In [105]:
```
class(data)
```

'data.frame'

In [106]:
```
head(data)
```

	average_monthly_ridership
1960-01	648
1960-02	646
1960-03	639
1960-04	654
1960-05	630
1960-06	622

In [107]:
```
tail(data)
```

	average_monthly_ridership
1968-07	1258
1968-08	1214
1968-09	1326
1968-10	1417
1968-11	1417
1968-12	1329

The dataset contains 2 columns one represents the year_month and other represents the total number of riders for the Portland public transportation system

In [108]:
```
str(df)
```

```
'data.frame':	116 obs. of  2 variables:
 $ Month                     : Factor w/ 116 levels "","1960-01","1960-02",..: 2 3 4 5 6 7 8 9 10 11 ...
 $ average_monthly_ridership: Factor w/ 113 levels ""," n=114","1020",..: 85 84 83 86 82 81 80 79 87 90 ...
```

Convertig the dataframe into time series table

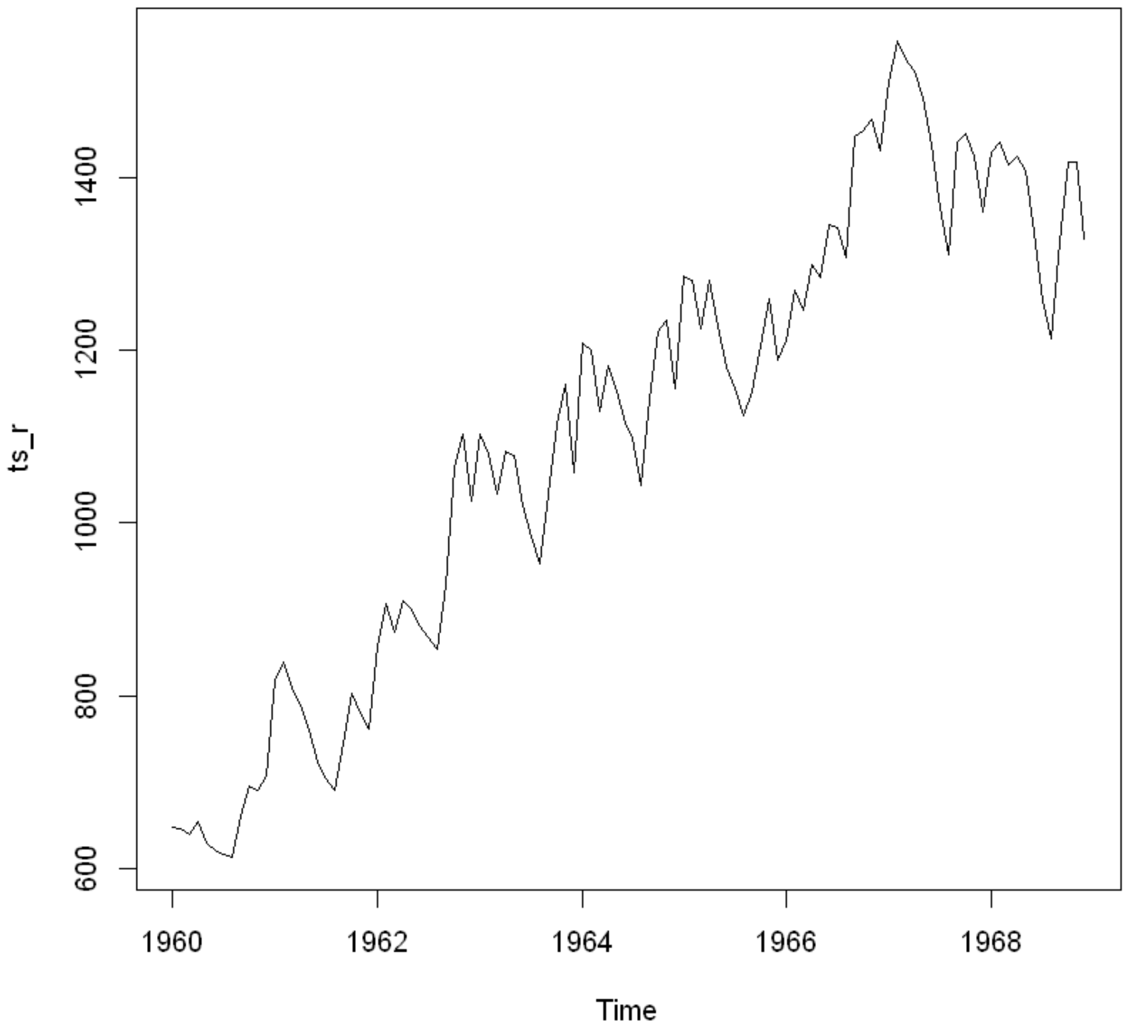
In [109]:
```
ts_r<-ts(data$average_monthly_ridership, start=1960, frequency = 12
```

Now we can see we caonverted the data into the time series

In [110]:
```
class(ts_r)
```

'ts'

In [114]:
```
plot(ts_r)
```



The above figure represents the trend in the data

The following R package forecast provides methods and tools for displaying and analysing univariate time series forecasts including exponential smoothing via state space models and automatic ARIMA modelling.

Here we are using the auto.arima function from the forecast R package to fit the best model and coefficients, given the default parameters including seasonality as TRUE

In [121]:
```
library('forecast')
mymodel = auto.arima(ts_r)
mymodel
```

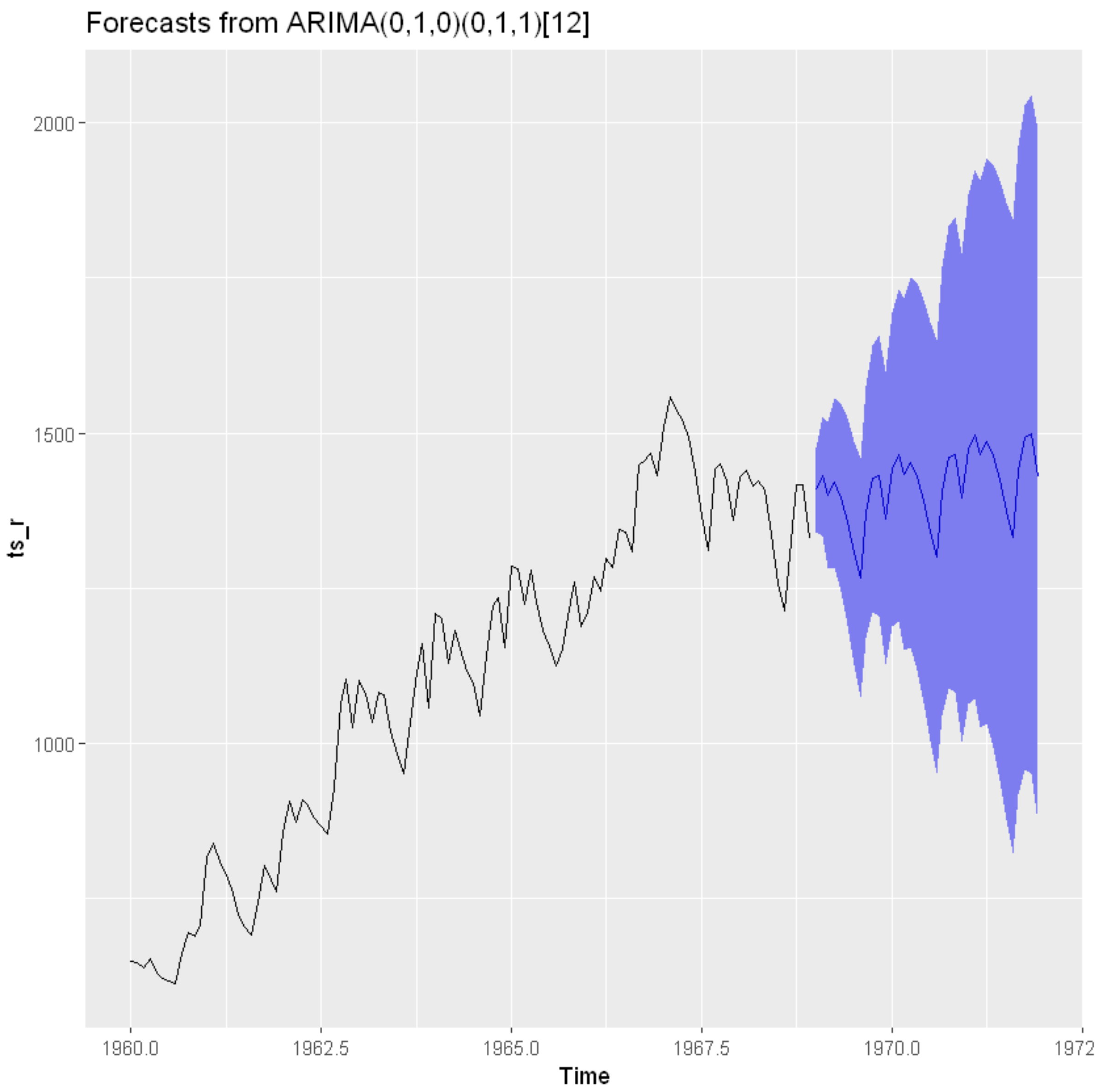
```
Series: ts_r
ARIMA(0,1,0)(0,1,1)[12]

Coefficients:
         sma1
      -0.6586
s.e.   0.1583

sigma^2 estimated as 1202:  log likelihood=-474.57
AIC=953.14   AICc=953.27   BIC=958.25
```

The ARIMA(0,1,0)(0,1,1)[12] model parameters are lag 1 differencing (d), an autoregressive term of second lag (p) and a moving average model of order 1 (q). Then the seasonal model has an autoregressive term of first lag (D) at model period 12 units, in this case months.

In [182]:
```
forecast <- forecast(mymodel, level = c(95), h = 36)
autoplot(forecast)
```



In the above we plotted a forecast of the time series using the forecast function, again from the forecast R package, with a 95% confidence interval where h is the forecast horizon periods in months.

Building the ARIMA model

In [162]:
```
fit<-arima(log(ts_r), c(0,1,1), seasonal = list(order=c(1,0,0), per
```

In [164]:
```
pred<-predict(fit, n.ahead = 2*12)
print(pred)
```

```
$pred
          Jan      Feb      Mar      Apr      May      Jun      Ju
l      Aug
1969 7.225891 7.231925 7.217587 7.223133 7.214241 7.173523 7.12559
5 7.097579
1970 7.195327 7.200075 7.188792 7.193156 7.186159 7.154117 7.11640
1 7.094354
          Sep      Oct      Nov      Dec
1969 7.167022 7.219255 7.219255 7.168801
1970 7.149001 7.190105 7.190105 7.150401

$se
            Jan        Feb        Mar        Apr        May
Jun
1969 0.03118884 0.04832178 0.06080497 0.07113012 0.08013579 0.0882
2695
1970 0.14047724 0.15518917 0.16862236 0.18106165 0.19269963 0.2036
7368
            Jul        Aug        Sep        Oct        Nov
Dec
1969 0.09563600 0.10251095 0.10895295 0.11503476 0.12081079 0.1263
2299
1970 0.21408594 0.22401476 0.23352181 0.24265667 0.25145990 0.2599
6520
```

In [165]:
```
pred1<-2.718^pred$pred
print(pred1)
```

```
          Jan      Feb      Mar      Apr      May      Jun      Ju
l      Aug
1969 1373.533 1381.846 1362.176 1369.750 1357.626 1303.463 1242.47
0 1208.147
1970 1332.192 1338.532 1323.516 1329.304 1320.036 1278.414 1231.10
0 1204.258
          Sep      Oct      Nov      Dec
1969 1295.017 1364.449 1364.449 1297.322
1970 1271.891 1325.254 1325.254 1273.672
```

Testing model

For this we created a time series data from 1960 to 1967. Which is the subset of the acutual time series data and we leaved the 1968 data which is we are going to predict and we will compare the results with the original values.

In [170]:
```
datawide = ts(ts_r, frequency = 12, start = c(1960,1), end= c(1967,
print(datawide)
```

```
      Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
1960  648  646  639  654  630  622  617  613  661  695  690  707
1961  817  839  810  789  760  724  704  691  745  803  780  761
1962  857  907  873  910  900  880  867  854  928 1064 1103 1026
1963 1102 1080 1034 1083 1078 1020  984  952 1033 1114 1160 1058
1964 1209 1200 1130 1182 1152 1116 1098 1044 1142 1222 1234 1155
1965 1286 1281 1224 1280 1228 1181 1156 1124 1152 1205 1260 1188
1966 1212 1269 1246 1299 1284 1345 1341 1308 1448 1454 1467 1431
1967 1510
```

In [171]:
```
fit<-arima(log(datawide), c(0,1,1), seasonal = list(order=c(1,0,0),
pred<-predict(fit, n.ahead = 2*12)
print(pred)
```

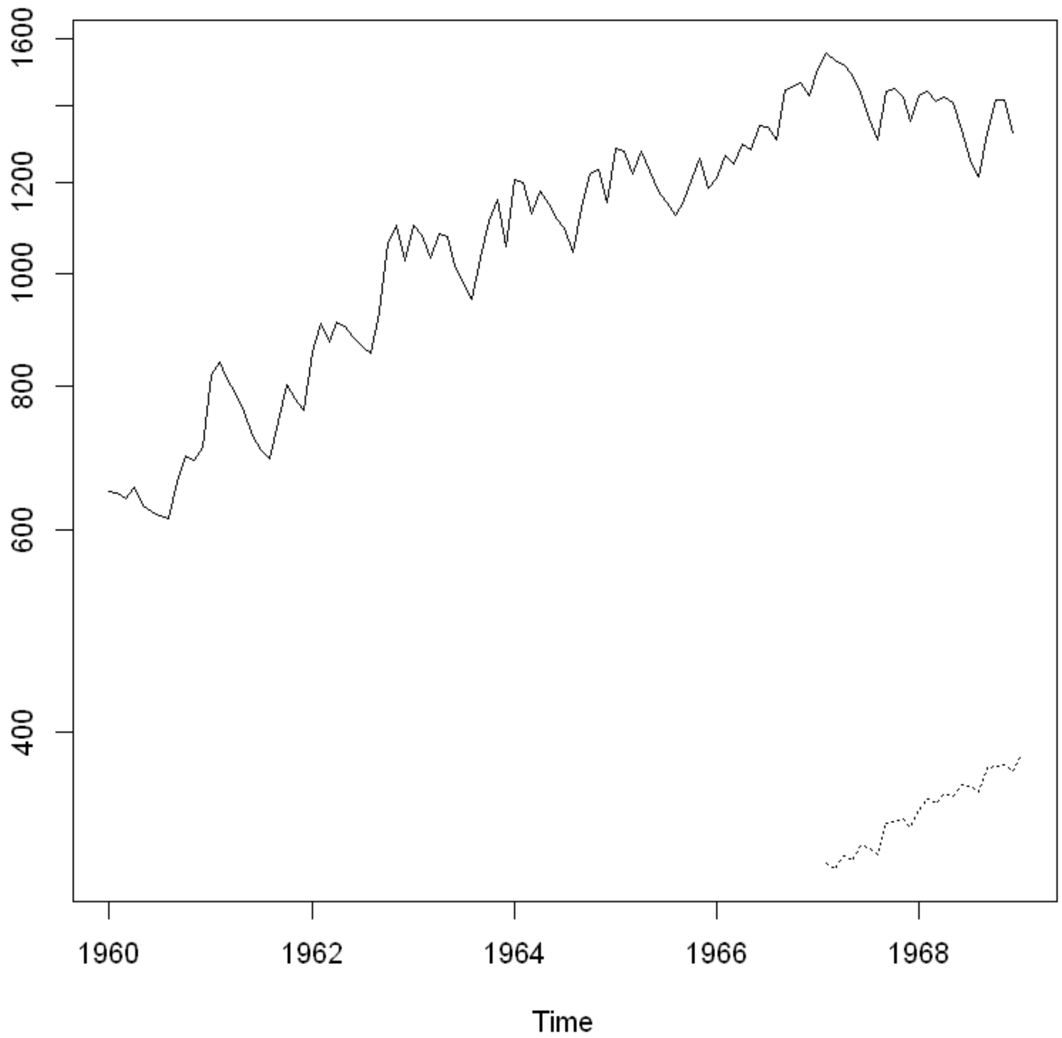
```
$pred
          Jan      Feb      Mar      Apr      May      Jun      Ju
l      Aug
1967          7.359943 7.345714 7.378120 7.369084 7.405191 7.40287
4 7.383491
1968 7.495209 7.526387 7.515318 7.540527 7.533498 7.561586 7.55978
4 7.544705
1969 7.631614
          Sep      Oct      Nov      Dec
1967 7.462594 7.465810 7.472735 7.453406
1968 7.606241 7.608744 7.614130 7.599094
1969

$se
            Jan        Feb        Mar        Apr        May
Jun
1967            0.03239287 0.04869898 0.06077732 0.07082496 0.0796
1450
1968 0.12487800 0.13922811 0.15350654 0.16656545 0.17867244 0.1900
0956
1969 0.25566027
            Jul        Aug        Sep        Oct        Nov
Dec
1967 0.08752577 0.09477897 0.10151525 0.10783152 0.11379776 0.1194
6641
1968 0.20070731 0.21086303 0.22055160 0.22983211 0.23875216 0.2473
5073
1969
```

In [172]:
```
pred1<-2.718^pred$pred
print(pred1)
```

```
          Jan      Feb      Mar      Apr      May      Jun      Ju
l      Aug
1967          1570.548 1548.361 1599.353 1584.969 1643.237 1639.43
4 1607.966
1968 1798.003 1854.938 1834.521 1881.350 1868.175 1921.387 1917.92
7 1889.227
1969 2060.744
          Sep      Oct      Nov      Dec
1967 1740.312 1745.919 1758.049 1724.398
1968 2009.121 2014.155 2025.032 1994.815
1969
```

In [173]:
```
ts.plot(ts_r, 2.178^pred$pred, log = "y", lty= c(1,3))
```



In [174]:
```
data1<-head(pred1,12)
print(data1)
```

```
          Jan      Feb      Mar      Apr      May      Jun      Ju
l      Aug
1967          1570.548 1548.361 1599.353 1584.969 1643.237 1639.43
4 1607.966
1968 1798.003
          Sep      Oct      Nov      Dec
1967 1740.312 1745.919 1758.049 1724.398
1968
```

In [175]:
```
predict_1968<-round(data1, digits = 0)
print(predict_1968)
```

```
      Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
1967      1571 1548 1599 1585 1643 1639 1608 1740 1746 1758 1724
1968 1798
```

In [176]:
```
original_1968<-tail(ts_r,12)
print(original_1968)
```

```
      Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
1968 1429 1440 1414 1424 1408 1337 1258 1214 1326 1417 1417 1329
```

As we can see our prediction is little bit higher than the original values.