Call Center Forecasting Case Study

By Simon Liu (May, 2018)

Background: use historical call center data (Jan, 2015 to May, 2018) to forecast number of calls offered for next two years and next 8 weeks for a telecom company in Canada.

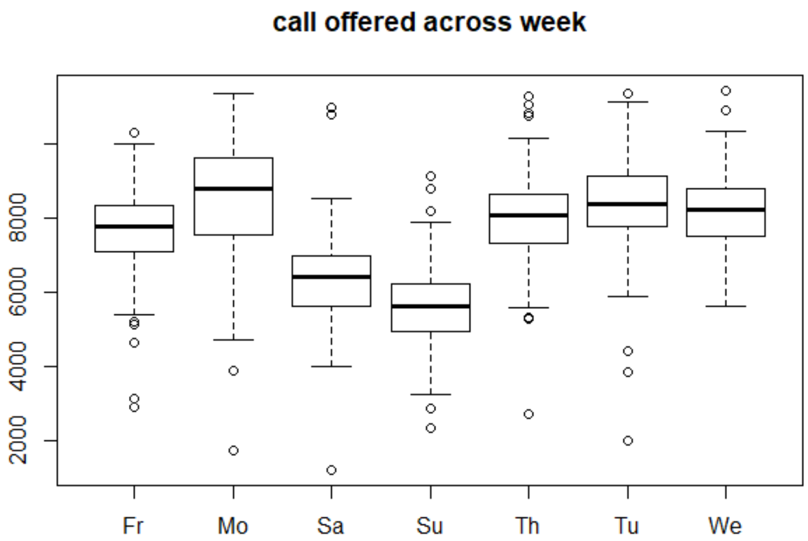
- Using R library: TSA, forecast, dplyr, lubridate, xts, stats
- Steps:
  - Data preparation, outliner imputation.
  - Testing stationarity, selecting model.
  - Model diagnosis.
  - Forecasting and discussion.

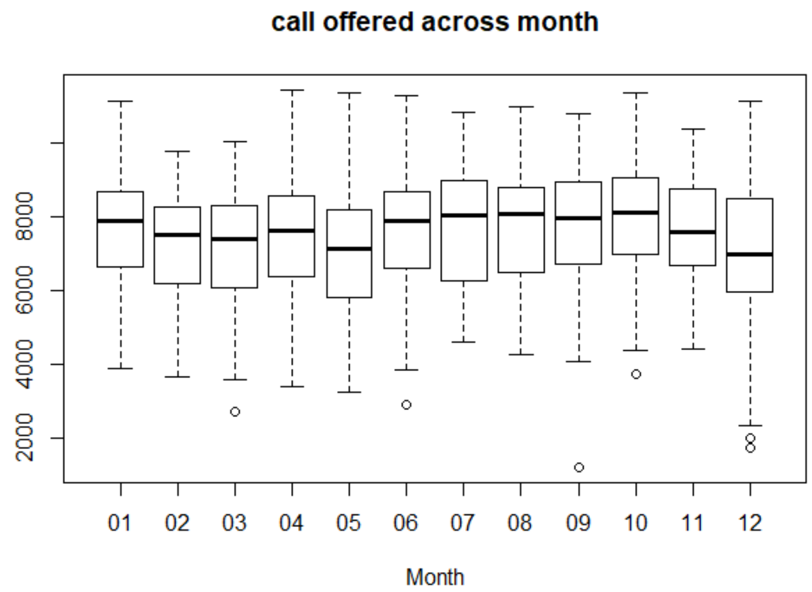
Monthly Forecasting: assumption

- Assumptions of ARIMA model.
  - The mean of error term is zero.
  - The error terms are identically independently distributed (i.i.d.).
  - Error terms follow standard normal distribution (white noise).
  - Model parameters are constant over time.
  - The relationship between variable is linear.

Monthly Forecasting: data preparation

- Exploring the data
  - Monday has most call offered, Sunday has fewest call offered.
  - May and December have fewer call offered compared with other months.

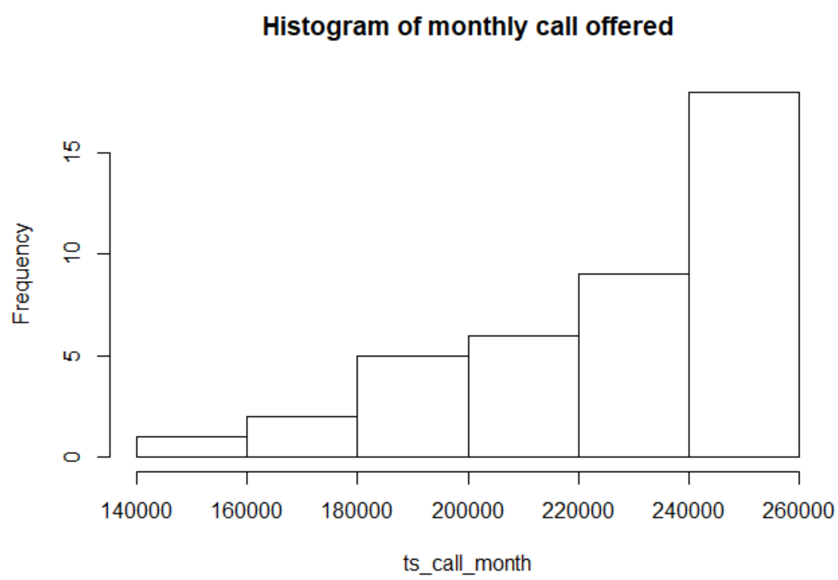




Monthly Forecasting: data preparation

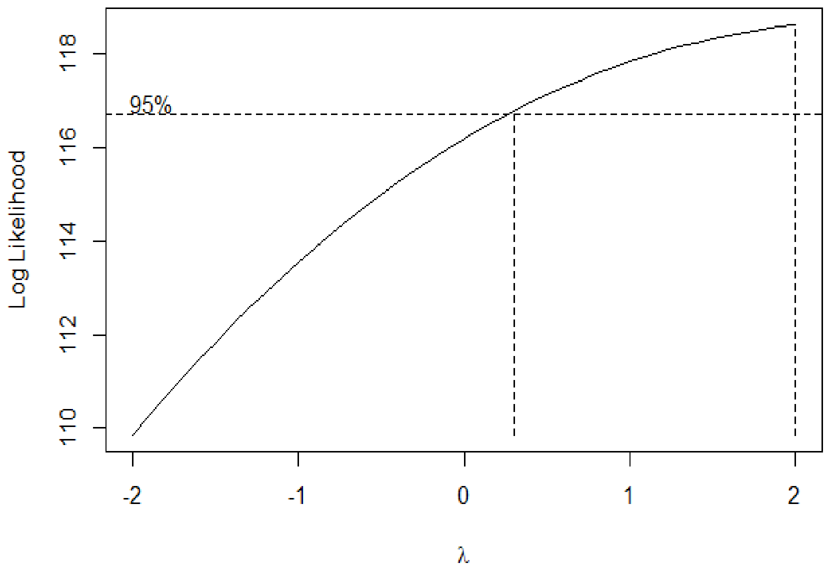
- Summarize daily data into monthly data.
  - Use dplyr() group by, summarize, arrange function.

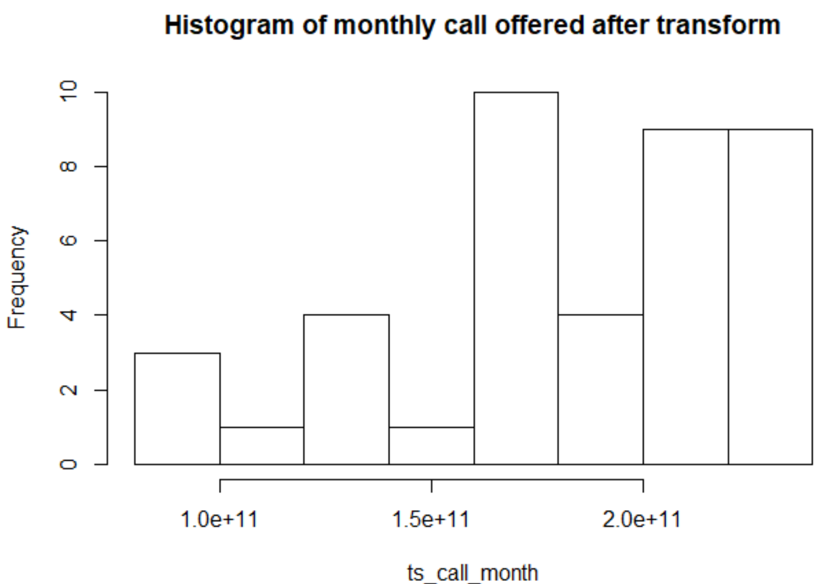
- Checking distribution of monthly data
  - Number of call is left skewed, need to do transformation (BoxCox).



Monthly Forecasting: data preparation

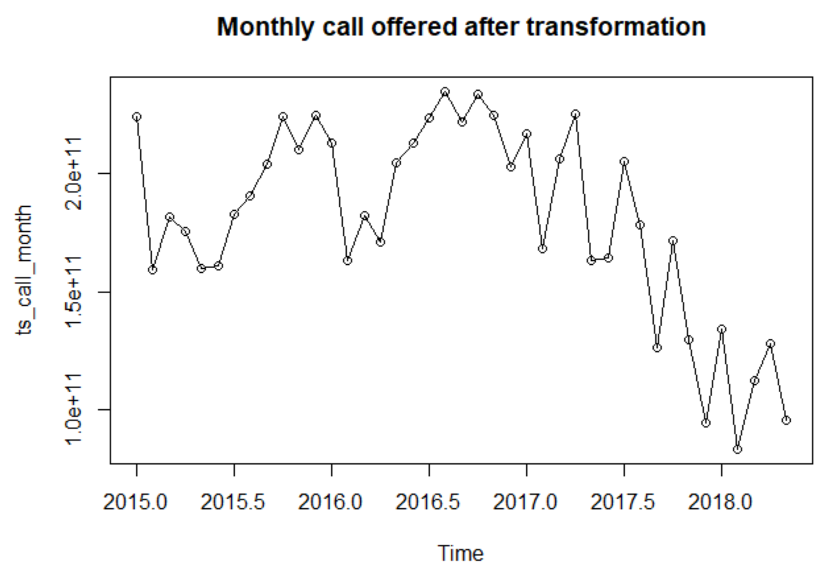
- Data transformation
  - BoxCox: transform data to normal by raise power to 2.1





Monthly Forecasting: testing stationarity

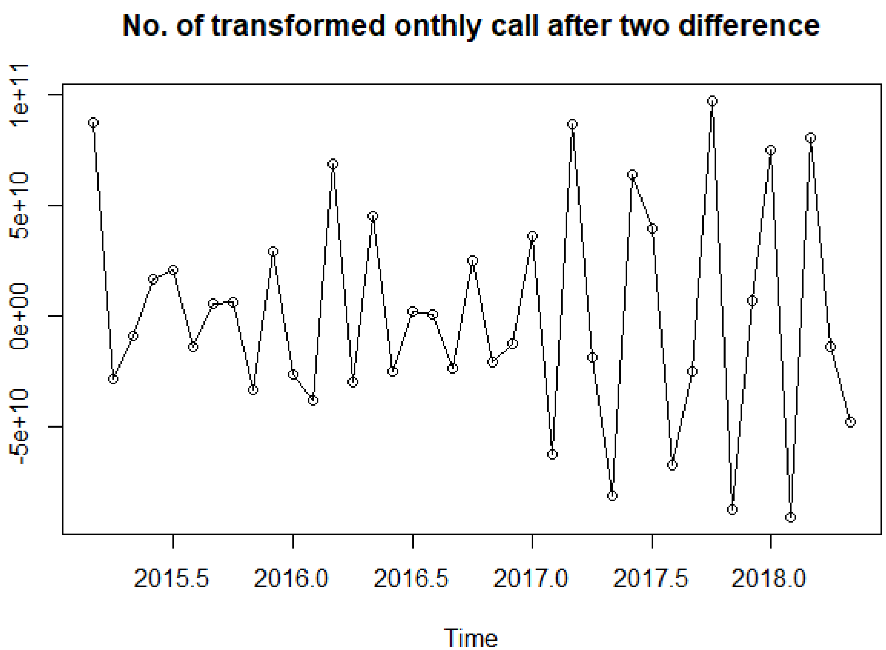
- Test stationarity : in order to use ARIMA must make sure data is stationary.
  - Look at the data, seems not stationary (downward trend).
  - Perform unit root test, which shows data is not stationary.




```
        Augmented Dickey-Fuller Test

data:  ts_call_month
Dickey-Fuller = -1.8795, Lag order = 3, p-value =
0.6209
alternative hypothesis: stationary
```

Monthly Forecasting: testing stationarity

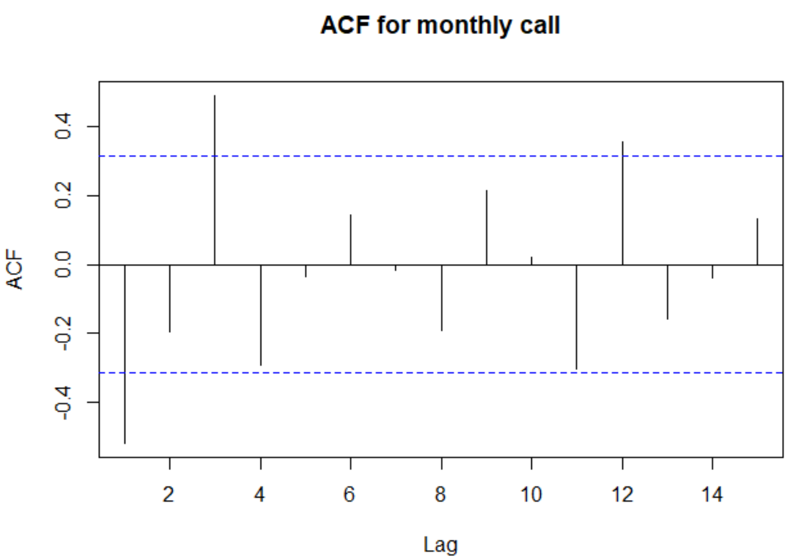
- Since data is not stationary, need to take difference and test again.
  - After 1st diff, data set is still not stationary, so take 2nd diff.
  - Unit root test is significant. Null hypothesis of not stationary is rejected.
  - Data now seems have no trend.

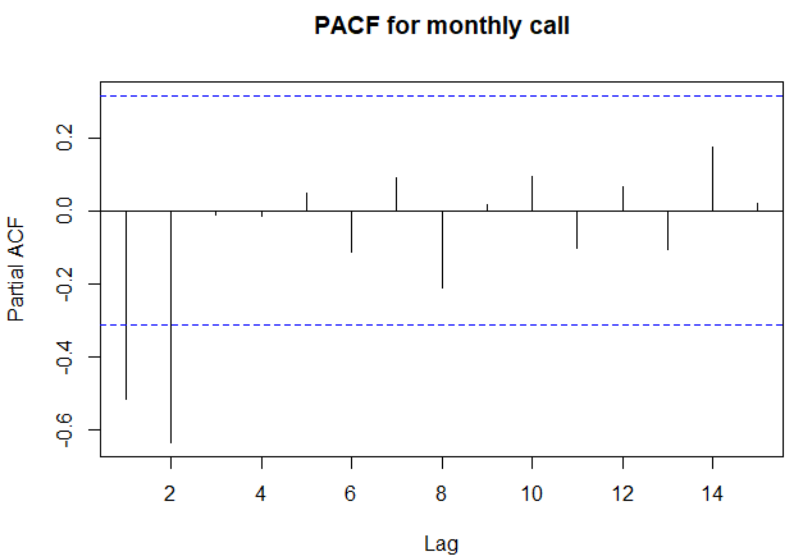



```
        Augmented Dickey-Fuller Test

data:  diff(diff(ts_call_month))
Dickey-Fuller = -4.5305, Lag order = 3, p-value =
0.01
alternative hypothesis: stationary
```

Monthly Forecasting: seasonality and trends.

- Check ACF and PACF to see if there is seasonality trend or pattern.
  - ACF shows a sudden increase in t(3), t(6), t(9), and t(12), suggesting a quarterly seasonality.
  - Maybe there are certain customers whose contract is quarterly based so they will call in to re-negotiate their contracts.





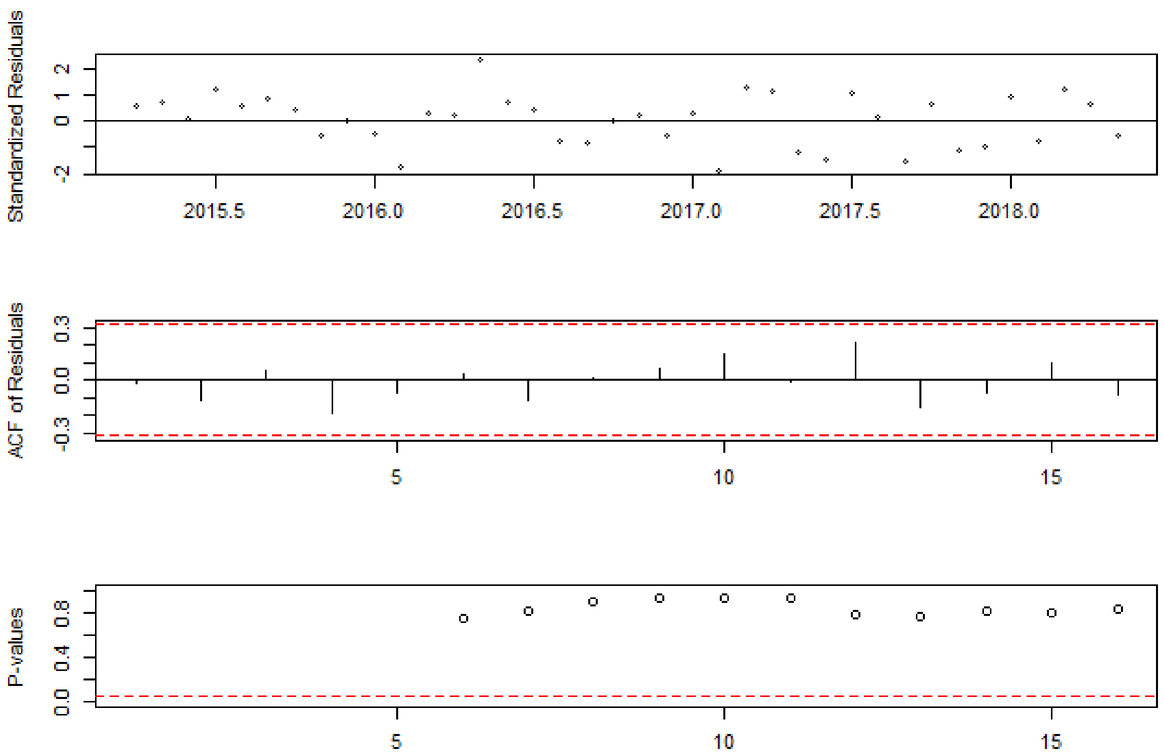
Monthly Forecasting: model selection

- Use auto.arima to test a few models and see which one performs better.
  - ARIMA(2,2,2) * $(1,0,0)_{12}$ model is selected. But still need to test a few others.
  - Seems model ARIMA(1,2,1) * $(0,0,1)_{3}$ works best (has lowest AIC, coefficients are significant).
  - Still need to do model diagnostics.

```
auto.arima(ts_call_month,d=2,D=NA,max.p=14,max.q=14,max.P=14,max.Q=14)

Series: ts_call_month
ARIMA(2,2,2)(1,0,0)[12]

Coefficients:
          ar1      ar2      ma1      ma2    sar1
      -0.9270  -0.5387  -0.3108  -0.6697  0.4874
s.e.   0.1994   0.1387   0.2754   0.2813  0.1599

model01 <- arima(ts_call_month, order=c(1,2,1),seasonal=list(
                 order=c(0,0,1),period=3))

Coefficients:
          ar1      ma1    sma1
      -0.3346  -1.0000  0.5758
s.e.   0.1630   0.0703  0.1361
```

Monthly Forecasting: model diagnosis

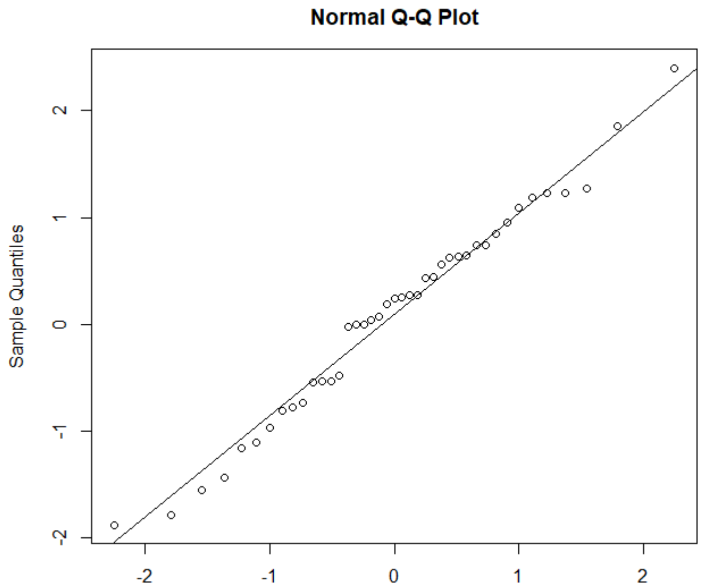
- Check if the model is appropriate.
  - Residuals are normally distributed and independent (i.i.d).
  - Model seems appropriate.

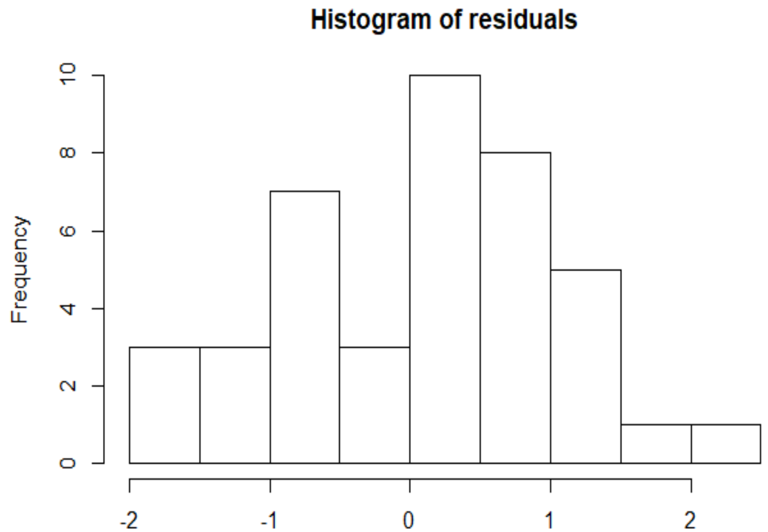


Monthly Forecasting: model diagnosis

- Check if the model is appropriate.
  - Again residuals are normally distributed.
  - Shapiro test doesn't reject null hypothesis, so residuals are normally distributed.
  - Again model seems reasonable.

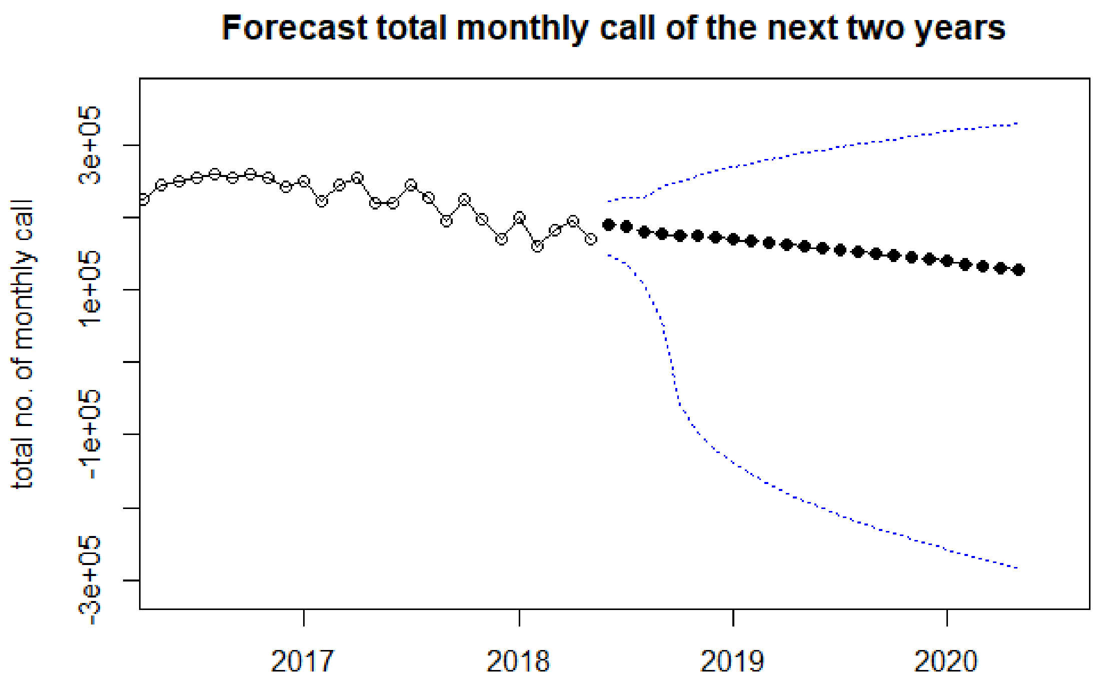
```
        Shapiro-Wilk normality test

data:  rstandard(model01)
W = 0.98197, p-value = 0.7497
```





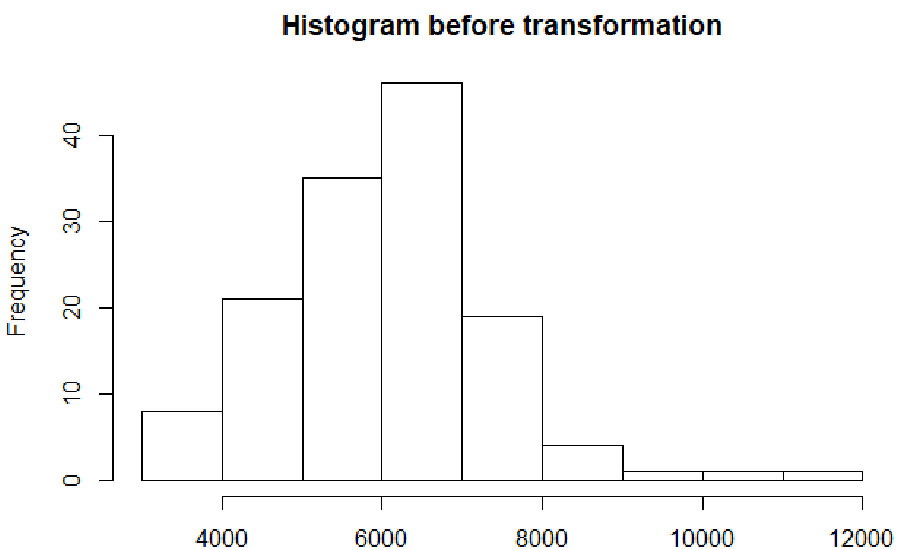
Monthly Forecasting: predict two years ahead.

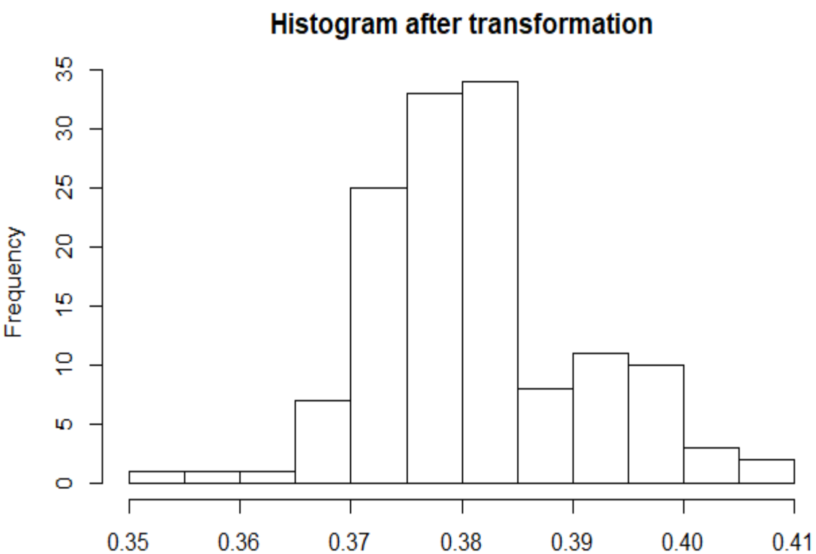
- Forecast monthly call offered in 24 month with 95% confidence intervals.
  - Seems there is a decrease trend for monthly calls after May 2018.
  - The longer the model forecasts, the bigger the variance and wider the confidence intervals.



Daily Forecasting : data preparation

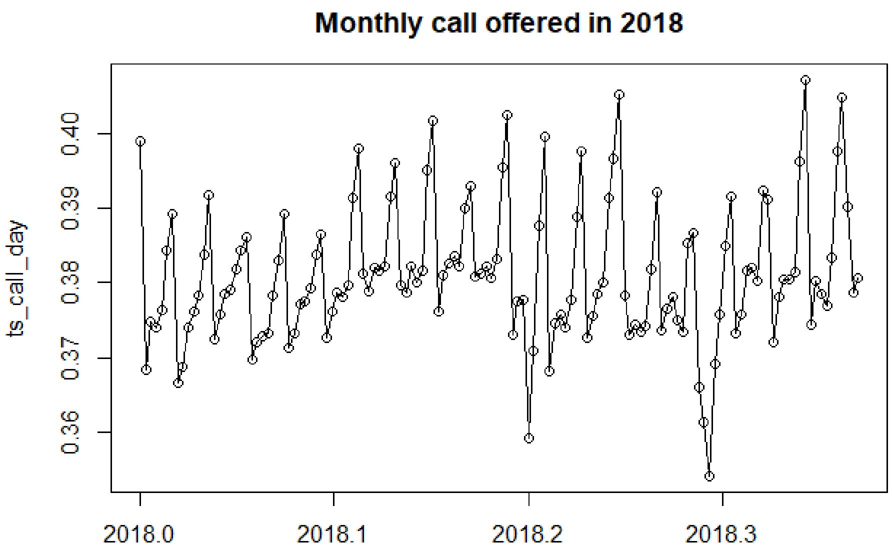
- Checking distribution of daily data
  - Since data is till May, 2018, use Jan to May data to forecast June and July.
  - Number of call is right skewed, need to do transformation (BoxCox).
  - All data is taken to the power of -1/9 to be normalized.





Daily Forecasting: testing stationarity

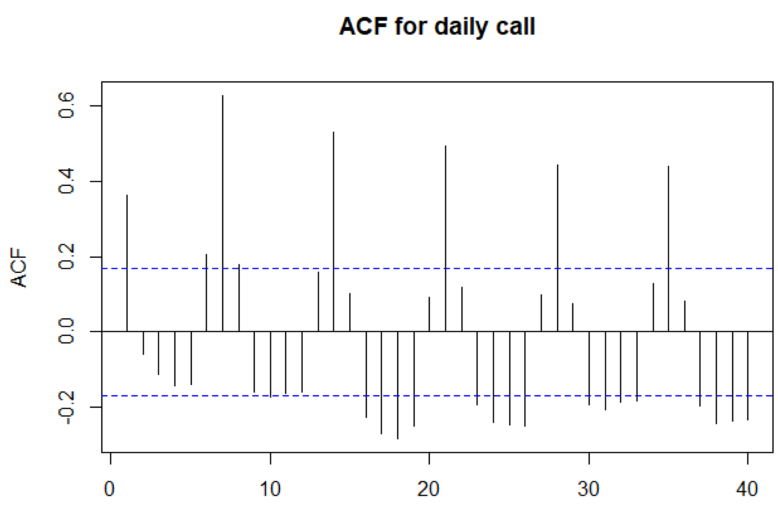
- Test stationarity : in order to use ARIMA must make sure data is stationary.
  - The data is stationary already ! No need to take difference.
  - Unit Root test shows data is stationary.

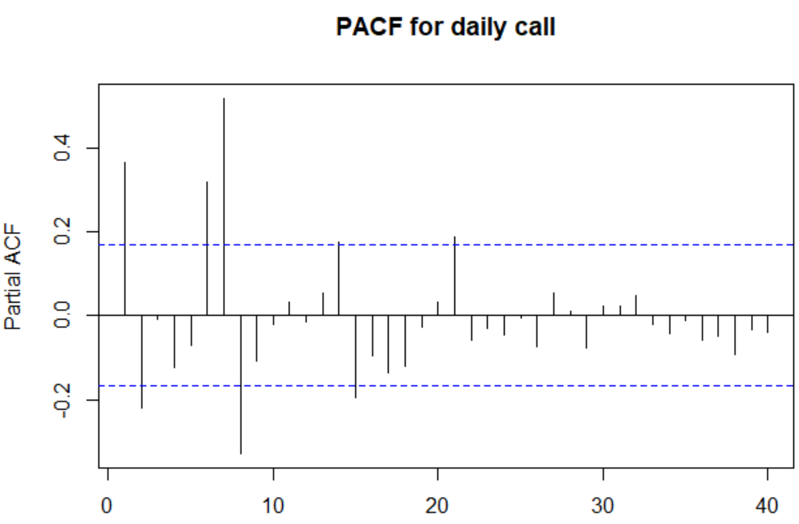



```
        Augmented Dickey-Fuller Test

data:  ts_call_day
Dickey-Fuller = -3.5799, Lag order = 5, p-value =
0.03778
alternative hypothesis: stationary
```

Daily Forecasting: seasonality and trends.

- Check ACF and PACF to see if there is seasonality trend or pattern.
  - Both ACF and PACF show there is a sudden increase between t(1), t(7), t(14), and t(21), suggesting a weekly seasonality.
  - Seems AR model may fit, but need to test a few to see.





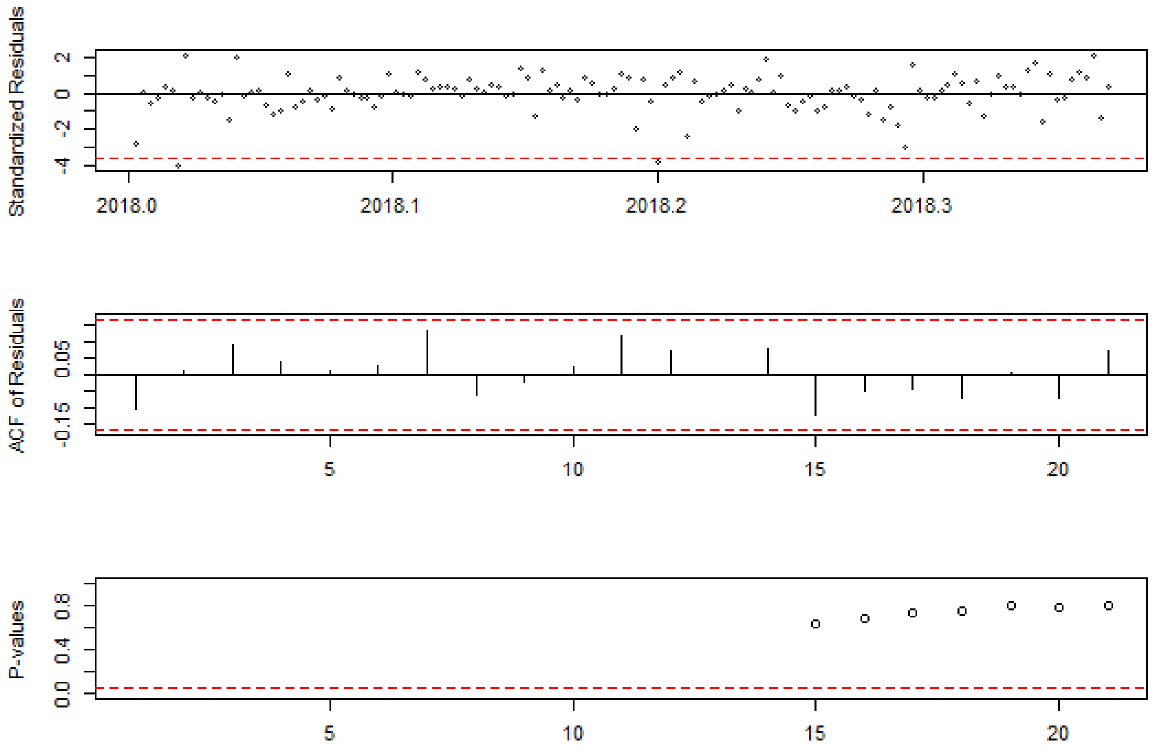
Daily Forecasting: model selection

- Use auto.arima to test a few models and see which one performs better.
  - ARIMA(2,0,1) model is selected. But still need to test a few others.
  - Seems model ARIMA(1,0,0) * $(1,0,1)_7$ works best (has lowest AIC, coefficients are significant).
  - Still need to do model diagnostics.

```
auto.arima(ts_call_day,d=NA,D=NA,max.p=15,max.q=15,max.P=15,max.Q=15)

Series: ts_call_day
ARIMA(2,0,1) with non-zero mean

Coefficients:
         ar1      ar2     ma1    mean
      0.3106  -0.1856  0.1671  0.3812
s.e.  0.4958   0.2155  0.5124  0.0010

model02 <- arima(ts_call_day, order=c(1,0,0),seasonal=list(
           order=c(1,0,1),period=7))

Coefficients:
         ar1    sar1     sma1  intercept
      0.6179  0.9976  -0.8951     0.3817
s.e.  0.0742  0.0055   0.1158     0.0075
```

Daily Forecasting: model diagnosis

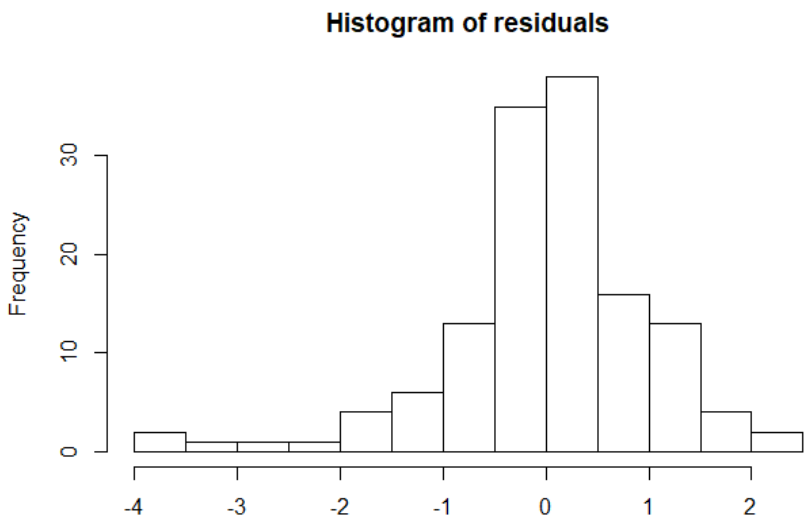
- Check if the model is appropriate.
  - Residuals seem normally distributed and independent (i.i.d).
  - Model seems appropriate.

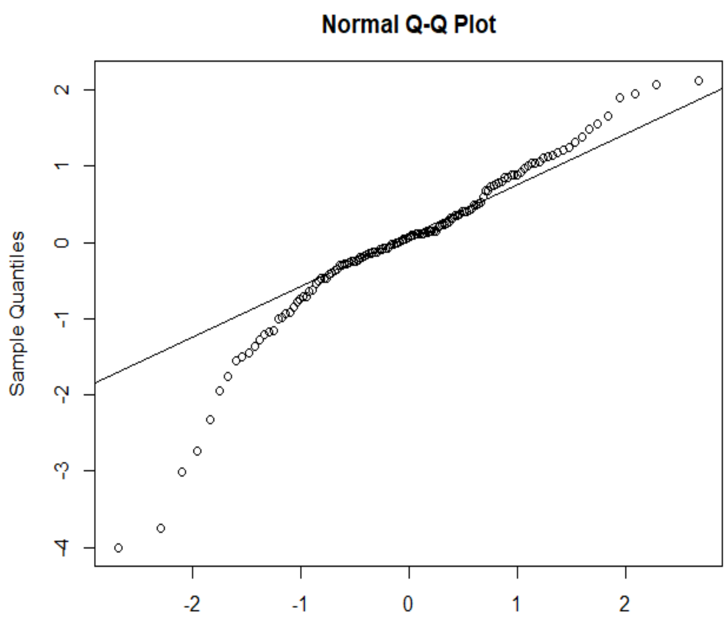


Daily Forecasting: model diagnosis

- Check if the model is appropriate.
  - Residuals seem have some left skew.
  - Shapiro test reject null hypothesis of normality.
  - Except for normality, all other looks good.

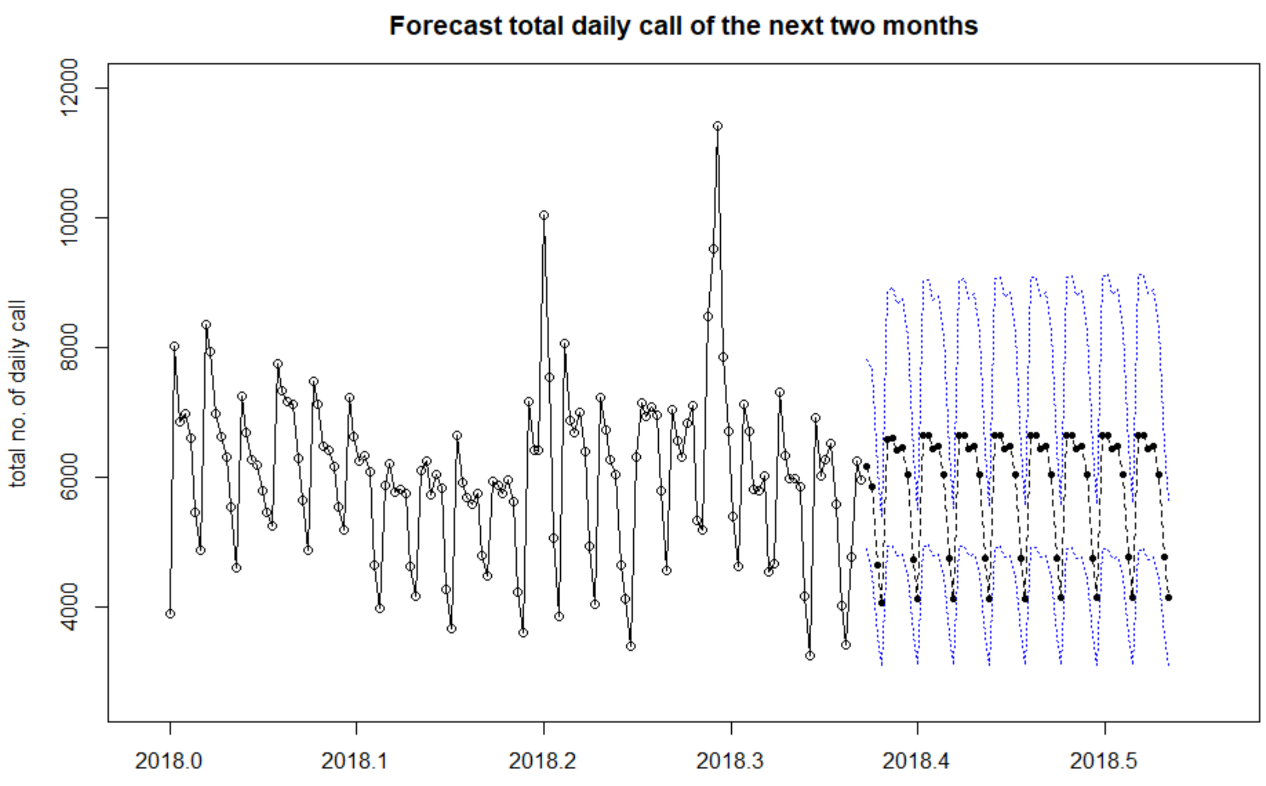
```
        Shapiro-Wilk normality test

data:  rstandard(model02)
W = 0.92716, p-value = 1.817e-06
```





Daily Forecasting: predict 8 weeks ahead.

- Forecast daily call offered in 8 weeks with 95% confidence intervals.
  - Daily call has a weekly seasonality.
  - Same trend continues to the future weeks.



THE END, THANK YOU!