ANALYSIS & FORECAST OF CALL DROPS IN CELLULAR NETWORK

Background

1. Cellular communication operators are expected to provide a noiseless, uninterrupted voice and data service quality to its customers. This class of noiseless, uninterrupted service is technically ensured through provisioning Quality of Service (QoS) parameters across a cellular telecommunication network. The QoS parameters provide for the overall satisfactory performance of the network

2. Dropped Call Rate (DCR) is one of the parameters of QoS of a network and refers to the telephone calls which, due to technical reasons, were cut off before the speaking parties had finished their conversational tone and before one of them had physically disconnected the call or hung up. Technically speaking, it represents the service provider's inability to maintain a call, either incoming or outgoing, once it has been correctly established. It is one of the major Key Performance Indices (KPIs) of a network and provides an insight into technical quality of the network.

3, Call drops being experienced by customers' results in customer dissatisfaction, customer churn (discontinuation of service) and erosion of brand value. Ultimately it affects the revenue of a company and its growth since a telecom company needs to add more fresh customers to its subscriber base than it suffers through churn.

Business Problem

4. The cellular network needs to guarantee a satisfactory quality of voice and data service to its customers, which is the primary product of the company. In the current competitive cellular telecommunication landscape wherein the customer has lucrative alternatives in terms of cellular service providers, inability to ensure quality and service leads to customer dissatisfaction and therefore churn. This has an adverse impact on the growth of the company as well as its brand value and reputation. The current business problem is as follows:-

(a) Currently call drops are being experienced across certain Base Transceiver Stations (BTS) in the network due to adverse weather situations.
(b) High DCR due to adverse weather conditions is one of the major compelling reasons for churn and needs to be identified and addressed.
(c) Deterioration in the product quality caused by a high DCR is unacceptable, hence this aspect needs to be mitigated/ minimized effectively.

5. Data on the call drops in these BTS (or cellular towers) primarily due to weather conditions is available. The data needs to be studied for identifying patterns so that technical measures are undertaken to address call drops pre-emptively or when call drops are experienced to prevent further incidences.

Business Objective

6. The objectives have been defined as follows:-

(a) Study data related to the call drops as experienced across identified cellular towers of the network and ascertain patterns, probable causes and mitigation measures.

(b) Having identified the probable reasons, forecast prospective call drop situations for towers in the network.

(c) Suggest technical pre-emptive measures to be undertaken to mitigate the build-up of causes which lead to call drop.

7. A high performing network with superior QoS and devoid of call drops even in adverse weather conditions will help the company provide a qualitative edge to the customers over the competition. This will attract fresh customers to the services offered by the company and hence allow expansion in customer base as well as generate revenue to expand the network itself.

Dataset

8. Source. https://github.com/IBM/icp4d-telco-manage-ml-project#2-obtain-your-data-from-data-virtualisation

Exploratory Data Analysis

9. The original dataset consists of 6157 row. It provides call drop data for six sites or Bas Transceiver Stations (BTS). The data has the following coloumns:-

(a) outgoing_site_id = col_double(),

(b) Start_Time_MM_DD_YYYY = col_double(),

(c) Start_Time_HH_MM_SS_s = col_character(),

(d) Weather = col_character(),

` (e) Total Calls` = col_double(),

(f) Traffic = col_character(),

(g) lat = col_double(),

(h) long = col_double(),

(i) Call_Dropped = col_double()

10. A visual analysis of data reveals the following aspects:-

(a) Call drop values have been provided six time periods of four hours each for each day of Jan, 2017.

(b) The values are repeated for each cellular site/ tower.

(c) Each day has six values for Call Drops corresponding to a four hourly period in a 24 hour cycle, namely one value each for 00:00 to 04:00, 04:00 to 08:00, 08:00 to 12:00, 12:00 to 16:00, 16:00 to 20:00, 20:00 to 23:59.

(d) Some values are missing for particular days.

11. Data Cleaning.

(a) The analysis is being carried out for the site ID 1717 only.

(b) All duplicate values for site ID 1717 have been removed.

(c) The series of six values per day has been completed through manual imputation thus creating 186 rows for a period of 31 days. (Total of 15 rows were missing).

12. <u>Graphical Analysis of the Data</u>.

(a) Weather as Causative factor for Call Drops and related to Time.

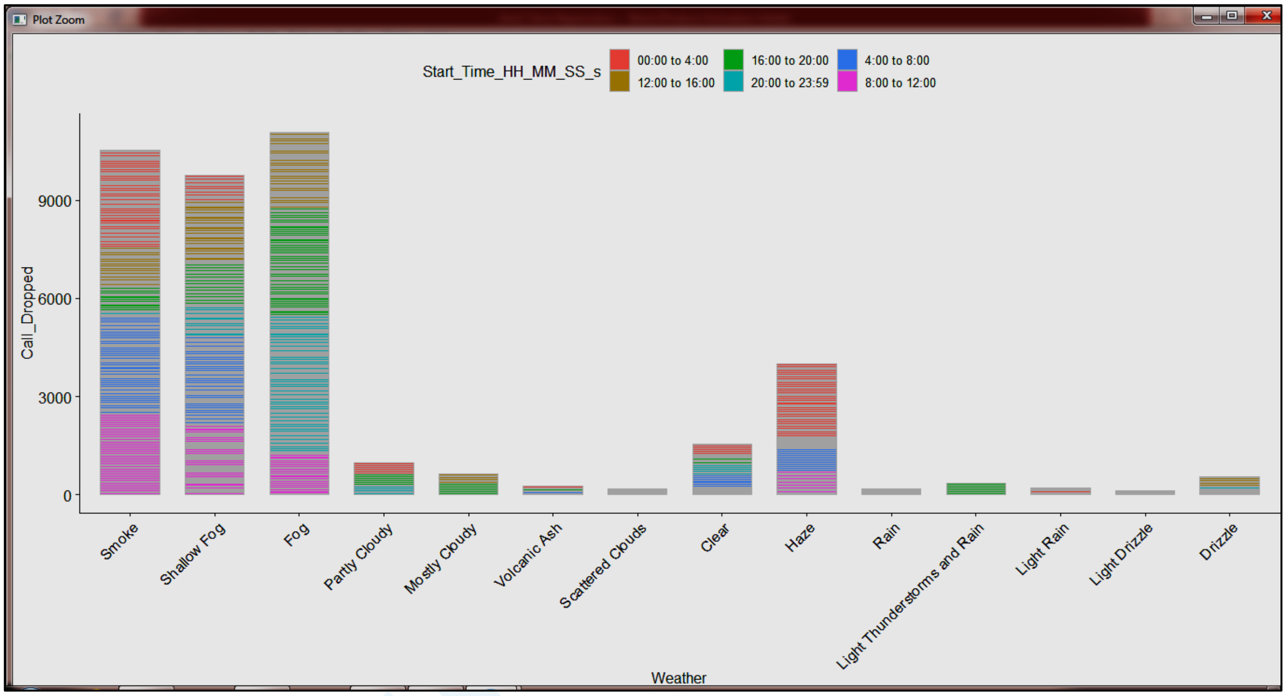


Fig 1. Plot of Weather Vs Call_Dropped with Time Intervals

(i) Maximum contributors to call drop is Fog, Smoke and Shallow Fog followed by Haze. Maximum effect is early morning and and after 20:00. Fog effects from 16:00 to 23:59 hours.

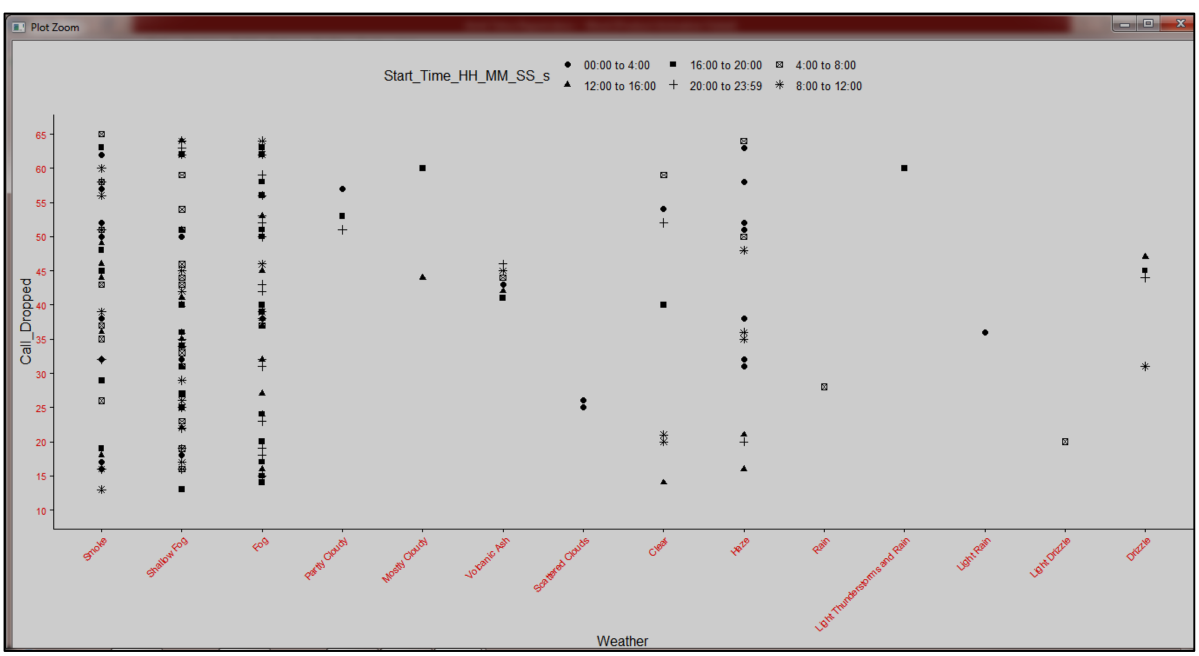
(ii) Effect of rain is felt primarily between 16:00 to 20:00.

(iii) Haze effects primarily in early morning hours.

(iv) Call drops take place in clear weather as well. Maximum during early morning and late night hours.

(b) Plot of Weather Vs Call Drop with respect to time intervals is shown in more specific details in Fig 2.

Fig 2. Plot of Weather Vs Call_Dropped with Specific Time Intervals marked



(c) Plot of Weather Vs Call Drops with respect to traffic indicates the following:-

(i) For Smoke, Shallow Fog and Fog conditions, the Low, Medium, and High Traffic leads to call drops in a band of 0-25, 25-50 and 50 to 65. Maximum call drops seem to be in the 25-50 band during medium traffic conditions.

(ii) Light thunderstorm causes call drops in heavy traffic situations. Drizzle in medium traffic situations.

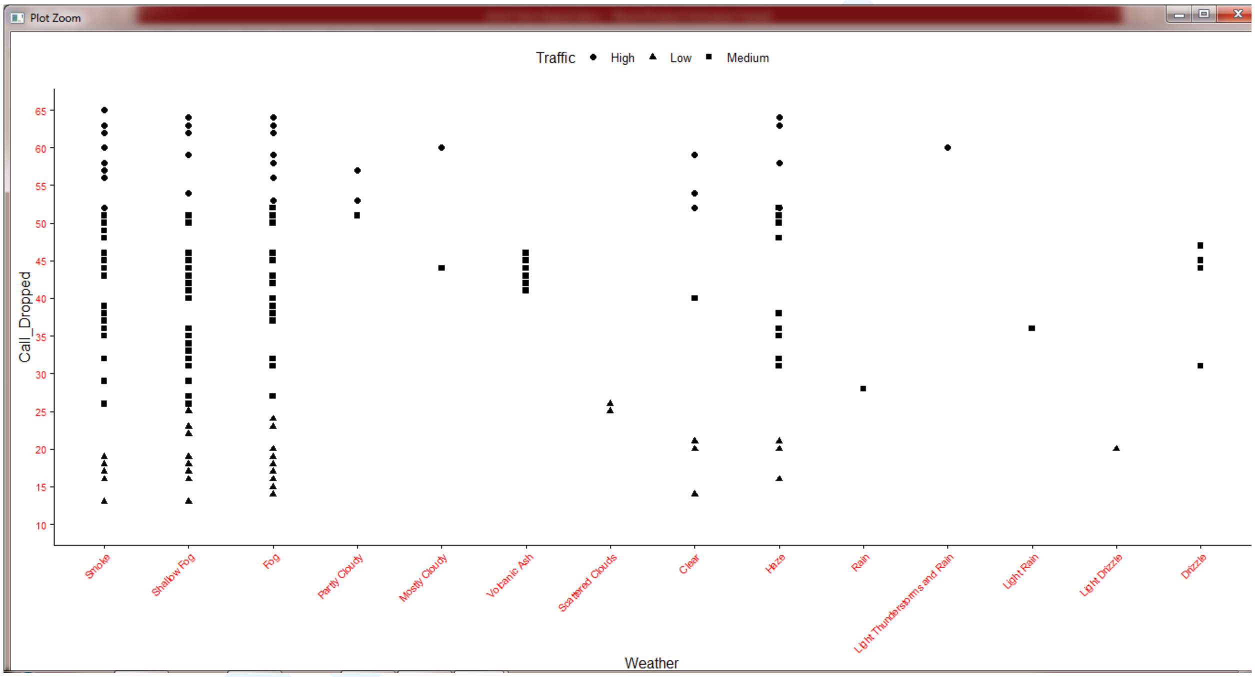


Fig 3. Weather Vs Call Drops with respect to traffic

(ii) Clear sky causes equal call drops during High and Low traffic conditions.

(d) Plot of Total Calls Vs Call Dropped does not indicate any specific pattern.

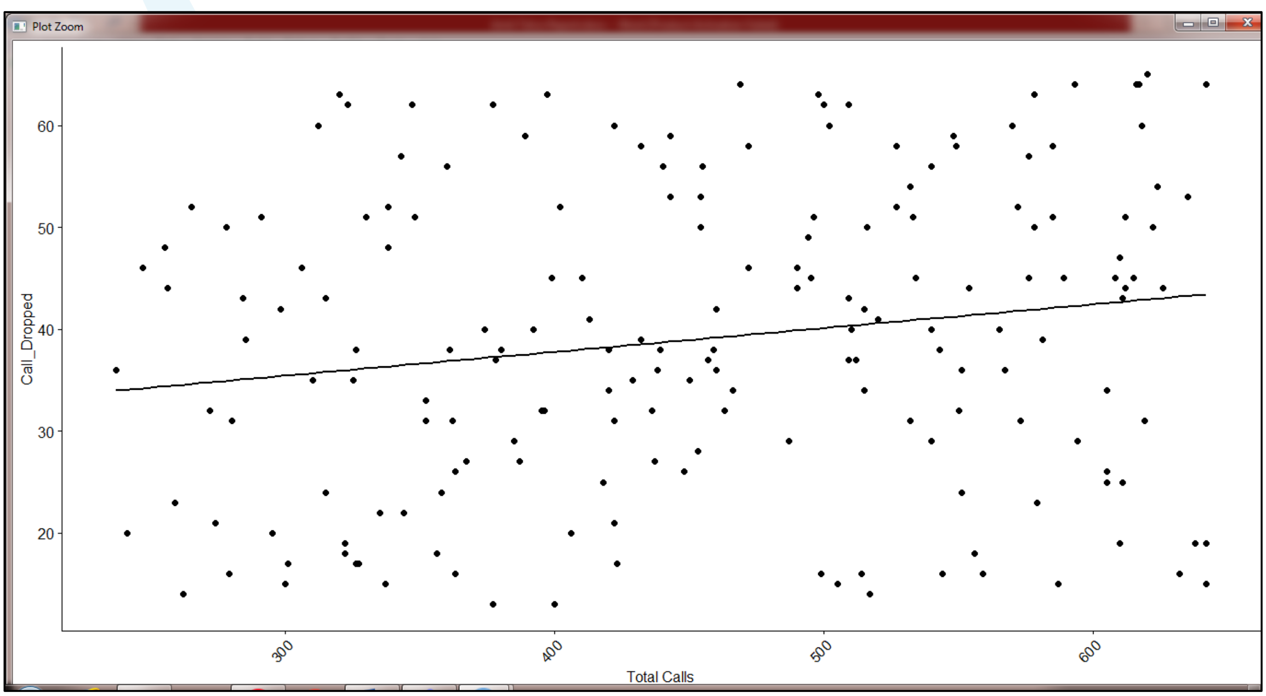


Fig 4. Total Calls Vs Call Dropped

(e) Plot of Traffic Vs Call Dropped indicates the findings as given in para 12(C)(i).

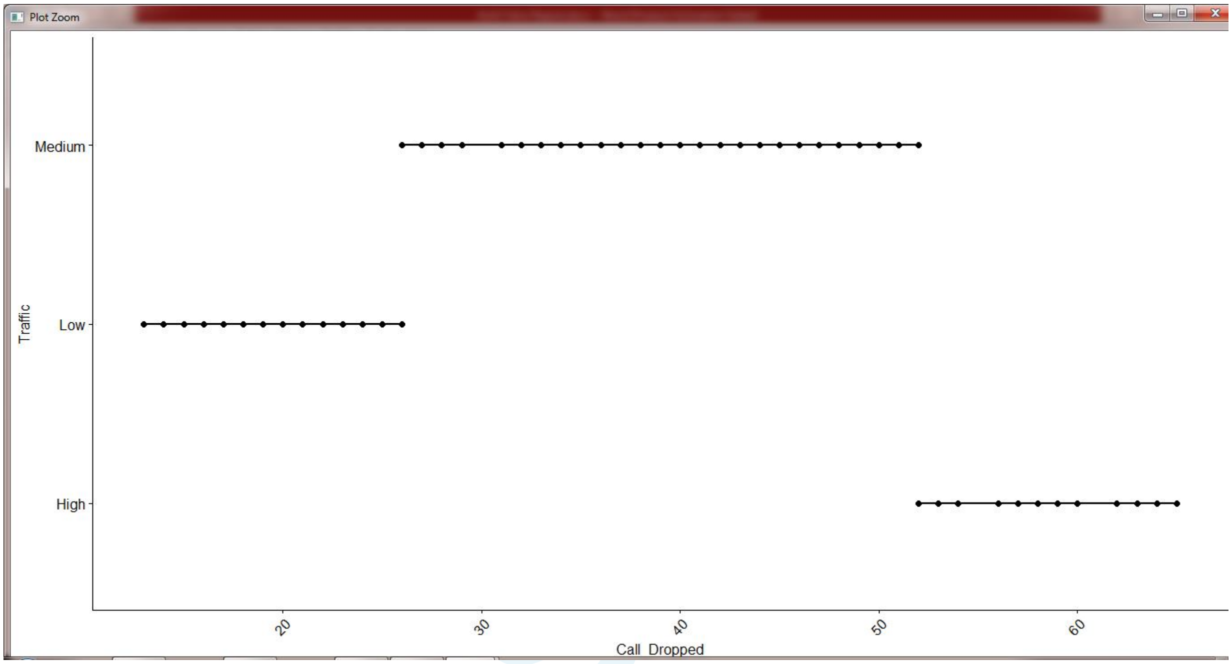


Fig 5. Traffic Vs Call Dropped

(f) Plot of Start Time Vs Call Dropped does not indicate any pattern.

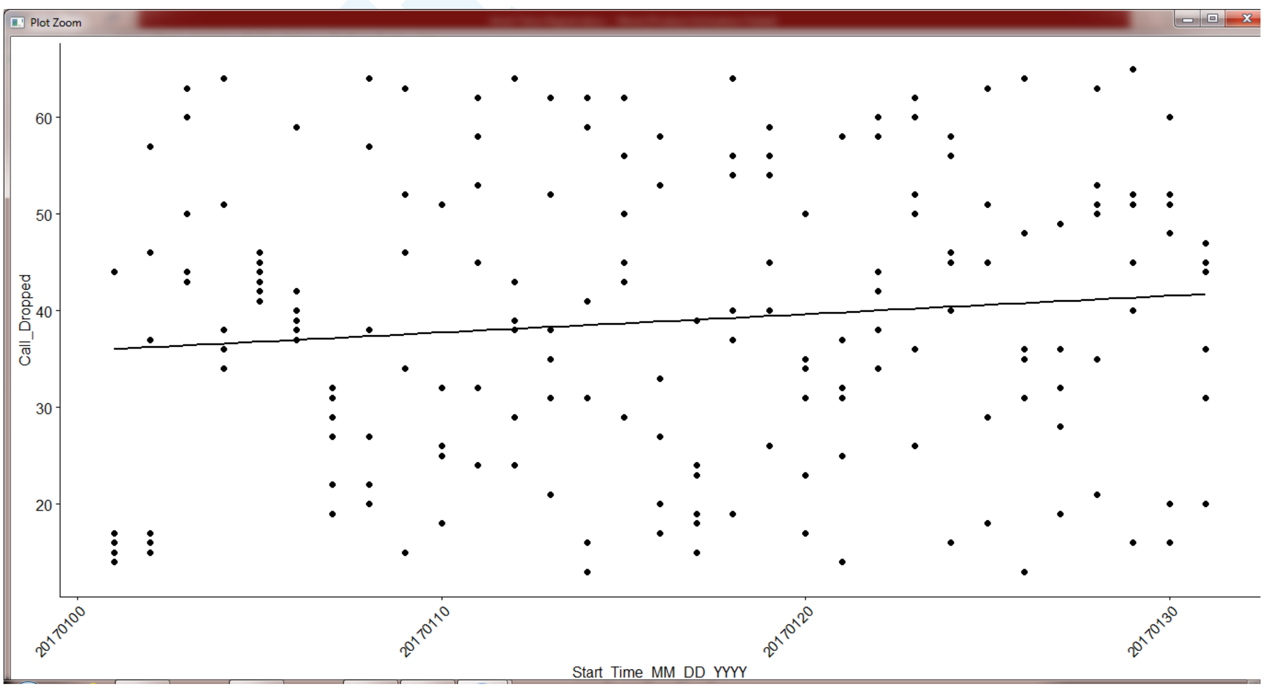


Fig 6. Start Time Vs Call Dropped

13. **Data Preparation**. The analysis will be a univariate analysis using the Time and Call Dropped parameters. The time series analysis requires a time – series data coloumn with corresponding values in another coloumn. The data is in terms of strings and hence needs to be parsed and prepared as time series data. The data was prepared for time series forecasting through the following process:-

(a) Removing duplicate rows for site ID 1717.

(b) Removing columns which are not required.

(c) Converting or parsing “Start_Time_MM_DD_YYYY” col to datatime format.

(d) Splitting the “Start_Time_HH_MM_SS_s” coloumn into three parts to extract the start of the four hour time period.

(e) Uniting the “Start_Time_MM_DD_YYYY” coloumn with the start time of the four hourly period to create a single “time_pd” coloumn which will be used as the time series. This coloumn will represent the time series for the call drops.

(f) Convert the col “time_pd” into POSIXct class and creating a separate col called ‘time_pd_posix”.

14. **<u>Time Series Plot</u>**.

(a) The plot does not show any trends. The Level is about 39 call dropped and there does not appear any additive or multiplicative aspects as well. The series data appears purely random. The plot and data is for a period of a month only, hence it is difficult to ascertain seasonal trends.

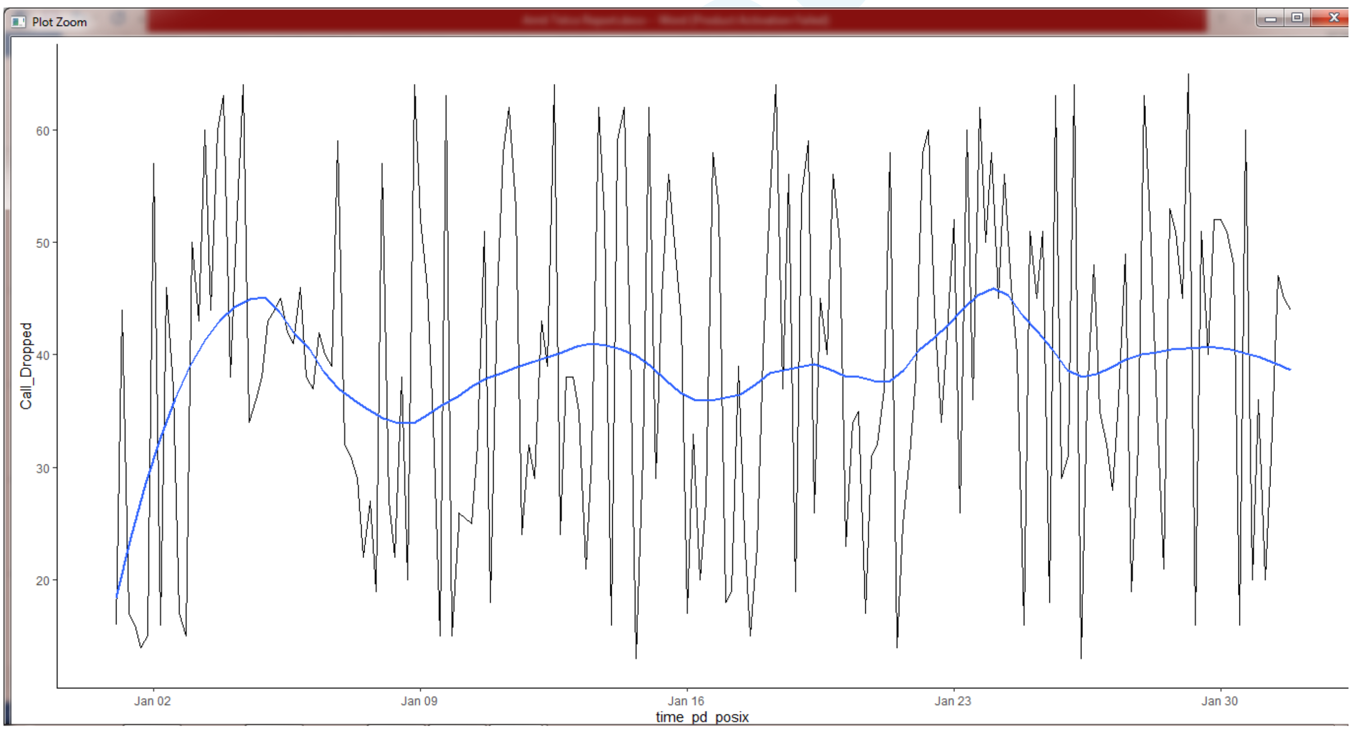


Fig 7. Time Series Plot

(b) **<u>Geometrical Smoothing Curve</u>**. The curve does not show a specific cyclic trend as well within a monthly period. Although an approximate 5 day cycle can be imagined.

15. **<u>Scatter Plot of Time Series Vs Call Dropped</u>**. Does not show any trend.

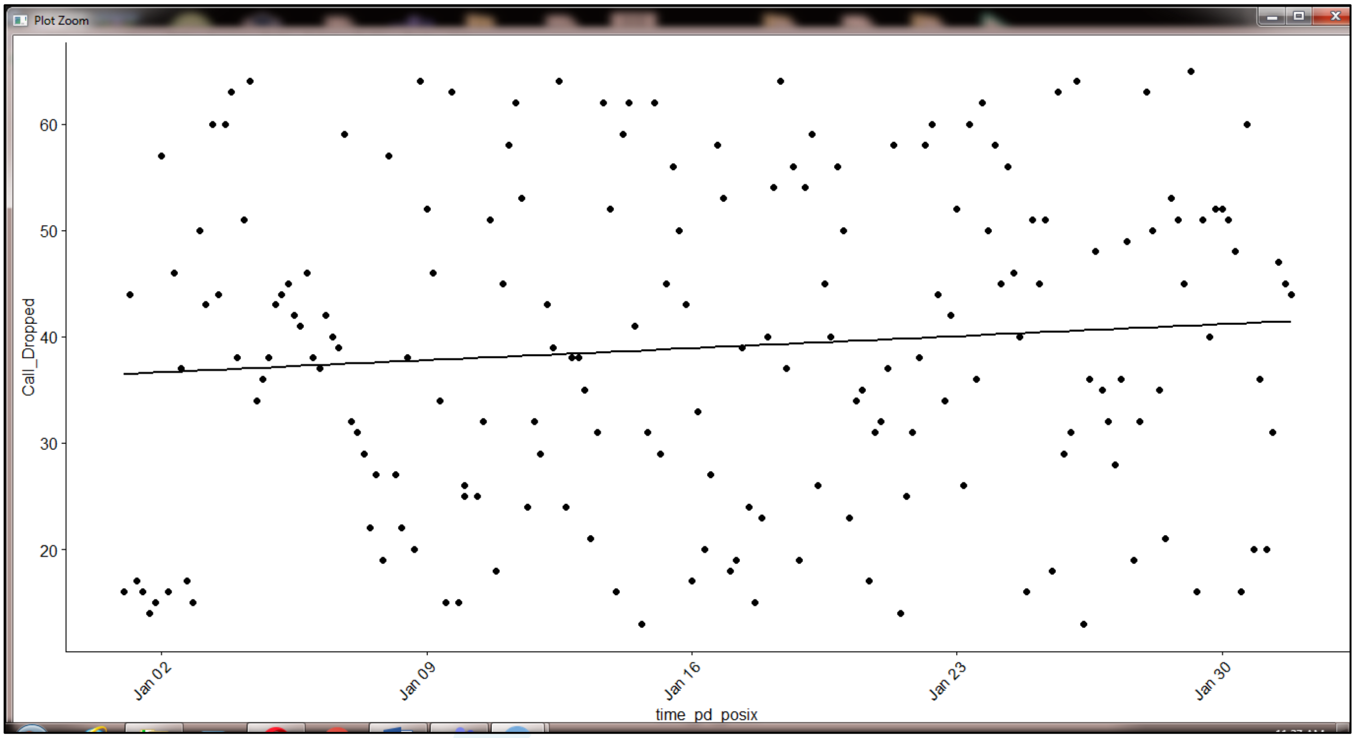


Fig 8. Time Series Scatter Plot

16. **<u>The STL plot</u>**. Indicates that there are no cyclic or seasonal trends.

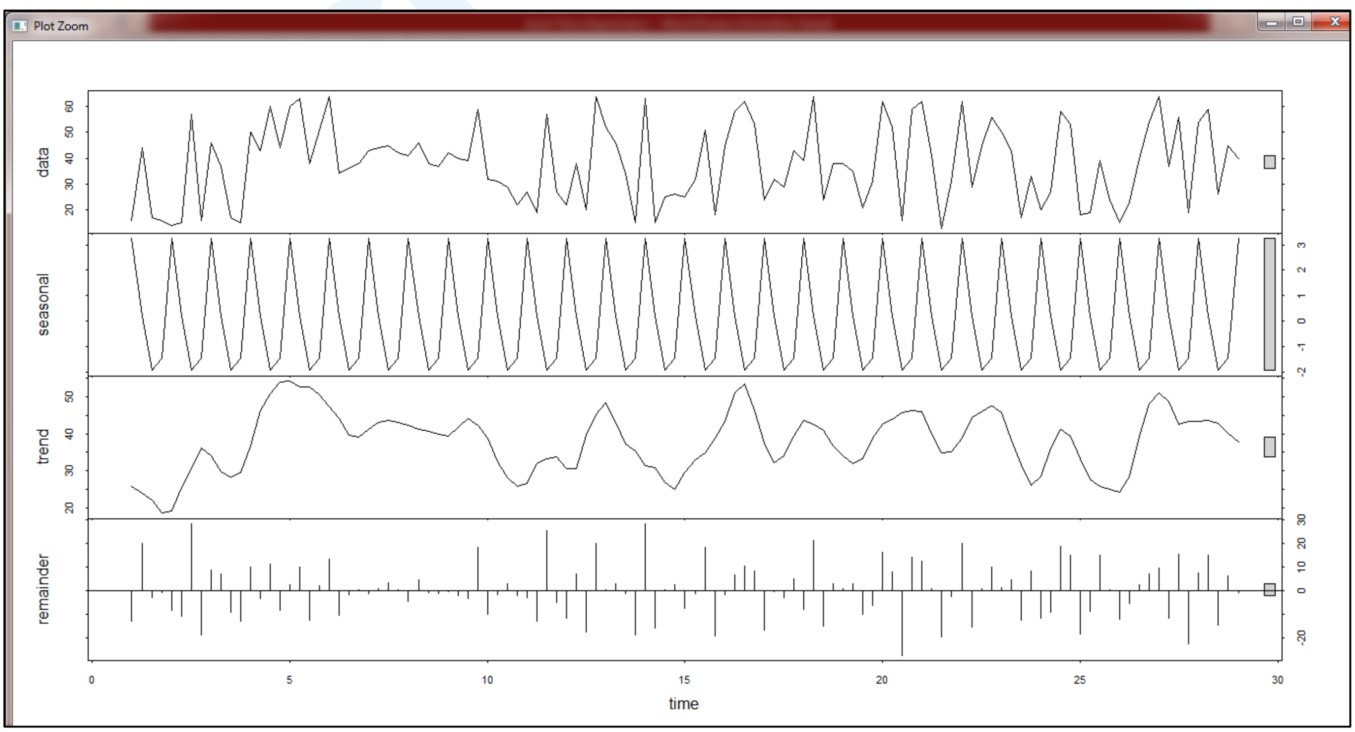


Fig 9. STL Plot

17. **Training & Test Dataset**.

(a) Train and Test Data. The model was trained for data of 29 days and then tested on the last two days of the month.

Forecast Models

18. **Naïve Model**.

```
Forecast method: Naive method

Model Information:
Call: naive(y = telco_1717_train_ts, h = 12)

Residual sd: 20.3838

Error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE       ACF1
Training set 0.2142857 20.38382 15.80357 -18.74106 51.63102 0.9206784 -0.4402511

Forecasts:
      Point Forecast      Lo 80     Hi 80       Lo 95     Hi 95
29 Q2             40  13.877087  66.12291   0.0484526  79.95155
29 Q3             40   3.056623  76.94338 -16.5000202  96.50002
29 Q4             40  -5.246212  85.24621 -29.1981099 109.19811
30 Q1             40 -12.245825  92.24583 -39.9030948 119.90309
30 Q2             40 -18.412609  98.41261 -49.3343758 129.33438
30 Q3             40 -23.987807 103.98781 -57.8609056 137.86091
30 Q4             40 -29.114731 109.11473 -65.7018589 145.70186
31 Q1             40 -33.886755 113.88675 -73.0000404 153.00004
31 Q2             40 -38.368738 118.36874 -79.8546422 159.85464
31 Q3             40 -42.607903 122.60790 -86.3378858 166.33789
31 Q4             40 -46.639900 126.63990 -92.5042925 172.50429
32 Q1             40 -50.492424 130.49242 -98.3962199 178.39622

RMSE – 14.33 %
```

19. Simple Exponential Smoothing Model.

```
Forecast method: Simple exponential smoothing

Model Information:
Simple exponential smoothing

Call:
 ses(y = telco_1717_train_ts, h = 12)

  Smoothing parameters:
    alpha = 1e-04

  Initial states:
    l = 37.9997

  sigma:  15.3554

     AIC     AICc      BIC
1155.489 1155.709 1163.671

Error measures:
                      ME     RMSE     MAE       MPE     MAPE      MASE      ACF1
Training set 0.003321073 15.21894 12.8151 -22.22674 45.30798 0.7465773 0.1015642

Forecasts:
      Point Forecast  Lo 80    Hi 80    Lo 95    Hi 95
29 Q2       37.99978 18.321 57.67856 7.903681 68.09587
29 Q3       37.99978 18.321 57.67856 7.903681 68.09587
29 Q4       37.99978 18.321 57.67856 7.903680 68.09587
30 Q1       37.99978 18.321 57.67856 7.903680 68.09587
30 Q2       37.99978 18.321 57.67856 7.903680 68.09587
```

```
30 Q3        37.99978 18.321 57.67856 7.903680 68.09587
30 Q4        37.99978 18.321 57.67856 7.903680 68.09587
31 Q1        37.99978 18.321 57.67856 7.903680 68.09587
31 Q2        37.99978 18.321 57.67856 7.903680 68.09587
31 Q3        37.99978 18.321 57.67856 7.903679 68.09587
31 Q4        37.99978 18.321 57.67856 7.903679 68.09587
32 Q1        37.99978 18.321 57.67856 7.903679 68.09587

RMSE – 12.337 %
```

20. <u>Holt's Trend Method.</u>

```
Forecast method: Holt's method

Model Information:
Holt's method

Call:
 holt(y = telco_1717_train_ts, h = 12)

  Smoothing parameters:
    alpha = 0.1315
    beta  = 1e-04

  Initial states:
    l = 19.6718
    b = 0.1534

  sigma:  15.9492

     AIC     AICc      BIC
1166.009 1166.570 1179.646

Error measures:
                    ME     RMSE      MAE      MPE     MAPE      MASE       ACF1
Training set 0.2692822 15.66441 13.16694 -18.7748 43.16255 0.7670748 0.04301466

Forecasts:
      Point Forecast    Lo 80    Hi 80     Lo 95    Hi 95
29 Q2       41.92251 21.48274 62.36229 10.662574 73.18245
29 Q3       42.07893 21.46296 62.69490 10.549528 73.60834
29 Q4       42.23535 21.44442 63.02629 10.438360 74.03234
30 Q1       42.39177 21.42707 63.35647 10.329024 74.45451
30 Q2       42.54819 21.41089 63.68549 10.221475 74.87490
30 Q3       42.70461 21.39584 64.01337 10.115667 75.29355
30 Q4       42.86103 21.38191 64.34014 10.011561 75.71049
31 Q1       43.01744 21.36907 64.66582  9.909114 76.12577
31 Q2       43.17386 21.35729 64.99044  9.808288 76.53944
31 Q3       43.33028 21.34654 65.31403  9.709046 76.95152
31 Q4       43.48670 21.33680 65.63660  9.611351 77.36205
32 Q1       43.64312 21.32805 65.95819  9.515169 77.77107
```

RMSE – 11.575%

21. <u>ARIMA Model</u>.

```
Series: telco_1717_train_ts
 ARIMA(0,0,0) with non-zero mean

 Coefficients:
          mean
       38.0000
 s.e.   1.4316

 sigma^2 estimated as 233.7:  log likelihood=-467.98
 AIC=939.96   AICc=940.07   BIC=945.42

 Training set error measures:
              ME     RMSE      MAE       MPE     MAPE      MASE      ACF1
 Training set  0 15.21818 12.81416 -22.23685 45.30968 0.7465224 0.1015667
```

RMSE – 12.33 %

Conclusion

22. The following can be inferred:-

(a) The dataset is inadequate to give a string forecast.

(b) The call drops has a strong dependence on weather conditions. Smoke, Fog and Haze play a major role in call drops.

(c) Heavy, medium or Low traffic does not have major role in call drops.

(d) The immediate pre-emptive measure suggested are as follows:-

(i) Weather forecasts of fog, haze and smoke should be obtained for cell towers experiencing call drops.

(ii) Transmission power of the microwave systems should be enhanced during these weather conditions to reduce call drops.

(iii) Ideally, towers where fog and smoke is a regular occurrence should be connected on optical fibre.

(f) Call drops will continue to occur in specified weather conditions as forecasted and hence a study of the effect of transmission power in different weather conditions and corresponding call drops needs to be carried out to create an actionable template of pre-emptive measure to be taken.

(e) Detailed study needs to be done and more data needs to be collected on effect of weather on cellular tower connectivity and call drops so as to make the forecast more accurate.

(f) ARIMA has given a forecast of 38 drops while Holt;s has given a forecast of 41 to 43 drops wherein the weather conditions are light thunderstorm, haze, drizzle. These are plausible forecasts, however it differs from actual values by about 15 percent. More data will improve the forecast further.