Final Weekly Project Report

Project- Long Call Timer

Mentor- Mr. Sridharan Natarajan

Report Submitted by-

Abhishek Raj Permani
MAGPIE Intern'20

Problem Statement

Most of service providers have a threshold value of each call, once the call duration exceeds this threshold value, the call automatically gets disconnected, and this value cannot be very small (because this will end call very frequently, which will leave users unsatisfied) also cannot be very large(because this will increase the load on network lines), it should be optimal.
We have to find that value of time after which the user has to redial again in order to continue. Also, a fix percentage of users must be satisfied with that time value.
Satisfied Percentage of users- is the percentage of total users, who ends the call before that end call time value, which we have to calculate.

Weekly Work Done-

Week 1-

Problem statement was shared with me. I started learning basics of python and time series analysis.

Week 2-

Started working on the data set using python libraries. Also, was studying different statical models of time series analysis, like-

- Autoregression (AR)
- Moving Average (MA)
- Autoregressive Moving Average (ARMA)
- Autoregressive Integrated Moving Average (ARIMA)

Week 3-

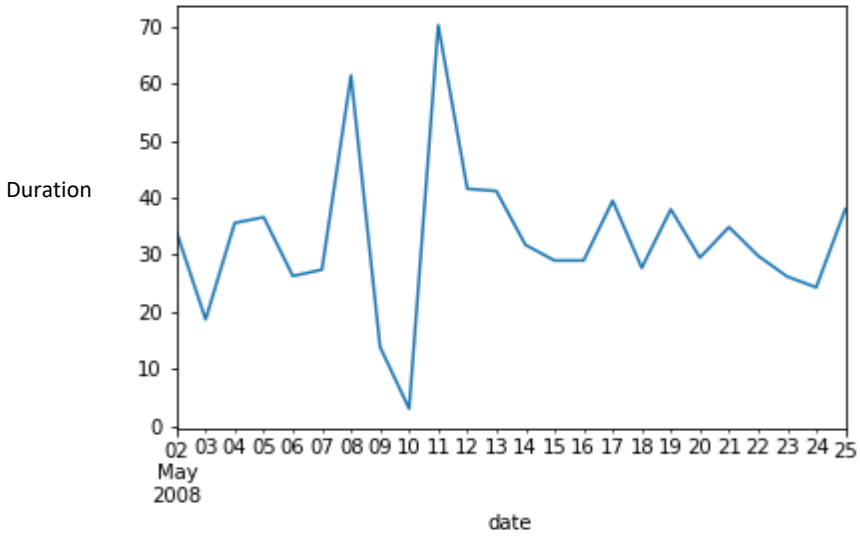
Completed the part of Dataset , and started working on its analysis part.

	A	B	C	D
1	User ID	Date	Time	Duration
2	1	01-05-2008	02:50	32
3	1	02-05-2008	08:45	23
4	1	03-05-2008	10:30	17
5	1	04-05-2008	18:48	29.4
6	1	05-05-2008	20:43	28.57143
7	1	06-05-2008	21:29	23.5
8	1	07-05-2008	01:38	22.16667
9	1	08-05-2008	20:47	30.66667
10	1	09-05-2008	22:41	38.75

Part of Dataset for User with User ID-1

Week 4-

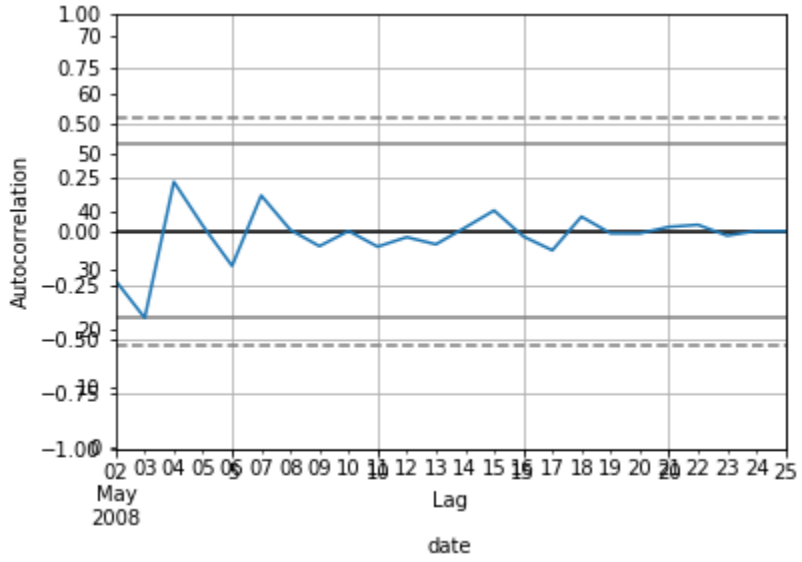
Started with considering the average duration of each user for each single day, and got the below plot.



Plot of Data of User ID-1 (Duration vs Date)

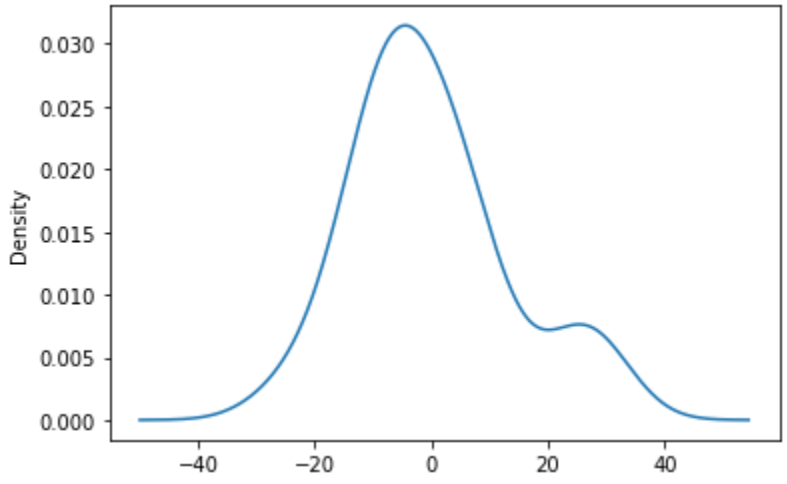
Week 5-

Applied different methods like AR model, LSTM on the generated data set ,but faced problems, so finally decided and discussed with Mentor , to go with ARIMA model, and obtained the below autocorrelation plot from generated datset-



Autocorrelation Plot of User ID-1 Dataset

Also, obtained the Density plot-



ARIMA Fit Residual Error Density Plot

Week 6-

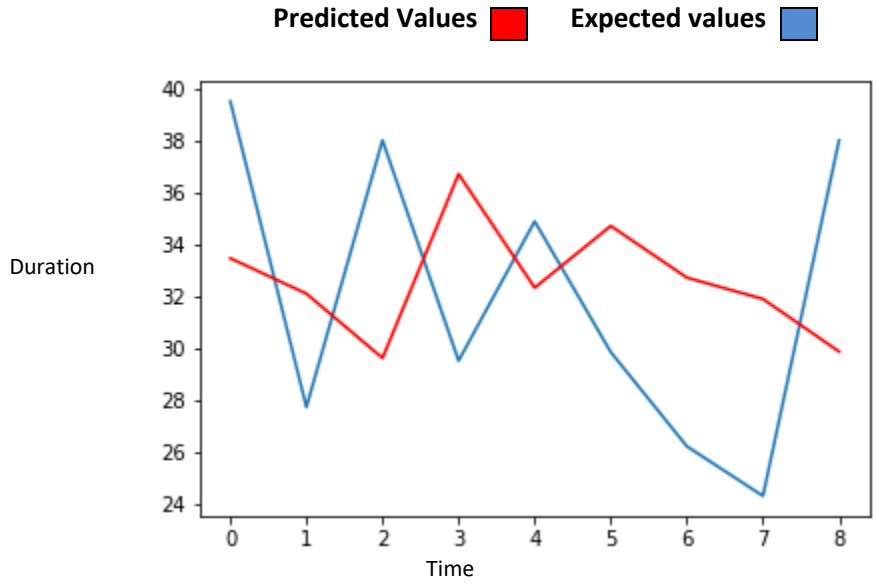
Applied ARIMA model and fit the mode on generated dataset, and obtained the below summary-

```
                             ARIMA Model Results
==============================================================================
Dep. Variable:             D.duration   No. Observations:                   23
Model:                 ARIMA(3, 1, 0)   Log Likelihood                -94.005
Method:                       css-mle   S.D. of innovations             13.979
Date:                Sun, 28 Jun 2020   AIC                            198.010
Time:                        13:00:09   BIC                            203.687
Sample:                    05-03-2008   HQIC                           199.437
                         - 05-25-2008
==============================================================================
                     coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const              -0.0116      1.027     -0.011      0.991      -2.025       2.002
ar.L1.D.duration   -0.8802      0.199     -4.432      0.000      -1.269      -0.491
ar.L2.D.duration   -0.8155      0.198     -4.128      0.000      -1.203      -0.428
ar.L3.D.duration   -0.2953      0.190     -1.554      0.120      -0.668       0.077
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1           -0.3260           -1.2244j            1.2670           -0.2914
AR.2           -0.3260           +1.2244j            1.2670            0.2914
AR.3           -2.1096           -0.0000j            2.1096           -0.5000
-----------------------------------------------------------------------------
                0
count  23.000000
mean   -0.193324
std    14.562287
min   -24.618742
25%    -8.905928
50%    -2.114163
75%     7.635549
max    29.031727
predicted=32.704698, expected=26.200000
predicted=31.880349, expected=24.285714
predicted=29.856820, expected=38.000000
Test MSE: 7.445
```

summary of the fit model

Week 7-

Forecasted the generated dataset using ARIMA model and obtained the plot below-



ARIMA Rolling Forecast Line Plot

(A line plot is created showing the expected values (blue) compared to the rolling forecast predictions (red). We can see the values show some trend and are in the correct scale.)

Week 8-

Asked for the percentage of satisfaction from Mentor and calculated the RMSE value and also the required Threshold Time Duration value for that percentage of Satisfaction- (Result for 95% Satisfaction Percentage)

RMSE Value= 15.879966352 minutes

Threshold Time Duration Value= 55.09 minutes