**Name: Liam Liew**

**Grab AI for SEA Challenge 1 (Traffic Management)**

In this challenge, I found out it is a Time series problem and I used Seasonal Auto Regressive Integrated SARIMA model.

In this project, since the traffic demand for the following time will be affected by the previous traffic demand and traffic is seasonal for every day (24 hours, 96 data in 15 min time interval), instead of judging by the position of the location, we forecast based on the historical data by using SARIMA model with 96 data lag.

The demand data is normalized to [0, 1] and it can be easily noticed the seasonal pattern in the graph.

In this challenge, I will first convert the ‘day’ and ‘time stamp’ to ‘time\_order’.

For Example:

*Day 1 00:00 = 1, Day 1 00:15 = 2, Day 1 01:00 = 4, Day 2 00:00 = 97, Day 2 13:15 = 150 and so on.*

It is basically split the hour\*60 + minute, then divided by 15 minutes to get the number. Also for the number of day, I will add (24 hours \* 60 minutes) / 15 minutes to get the ‘time\_order’ and add to the previous one. Eventually it will tidy up and left the ‘time\_order’ and the ‘demand’ columns, store it in a dictionary with key value = geohash6 name.

**SARIMA Models**

The traffic data is seasonal data with 96 data per day. Thus I used SARIMA(p,d,q)(P,D,Q)96 model to forecast the demand for the next T+1 to T+5 interval.

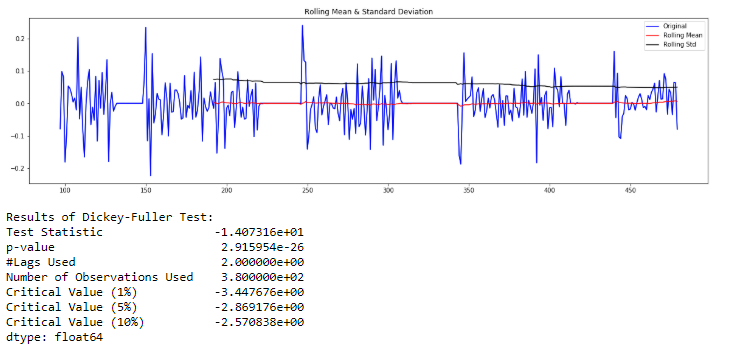
For the time series problem, geohash6 in this challenge is not that critical, it is just that to tell us the demand and the time is in one particular place, that is why it is removed after I group the location separately in ‘separated\_data’ in my program.

For some location, it is obvious that there is an uptrend from Saturday to next week Friday. Hence, I do two differencing here in my code.

First = data - data.shift(1) -- non seasonal differencing ( to remove the trend)

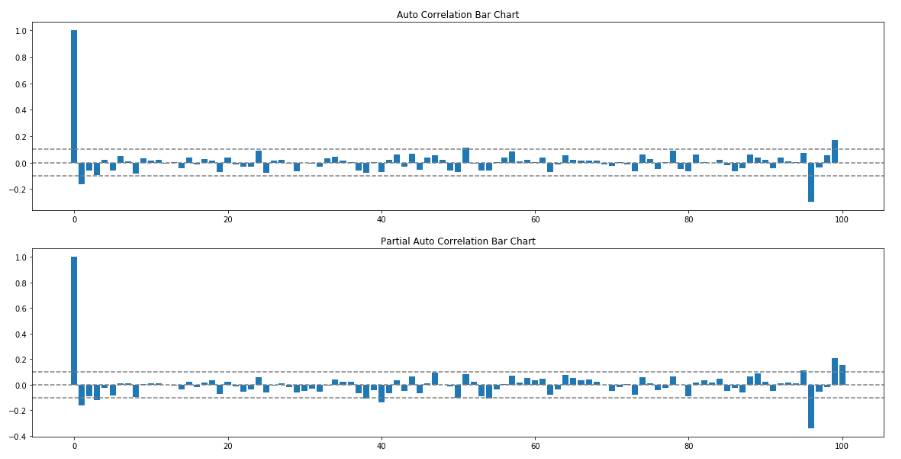
Second = First - First.shift(96) -- seasonal differencing (to remove the seasonal)

After that, this data is stationarity and it can be proceeded to SARIMA model to forecast the next 5 time interval demand prediction.



This is the way that I used to test the stationarity for my data after differencing. We can noticed that p-value is very small and most importantly the mean and variation is constant for the time period.

After we are confident that it is stationarity, we will proceed to the next step to focus on p, q, P, Q value based on Auto Correlation Function plot (ACF) and Partial Auto Correlation Function Plot (PACF).



Based on the chart, it can be seen that p is possible to be [0,1,2] and P, q, Q will be either 0,1. In this SARIMA model, I used AIC value to determine which model is the best to predict the demand in the next 5 time intervals. The most common **lowest AIC** in below testing are **SARIMAX (1,1,1)(1,1,0)96 OR SARIMAX (1,1,0)(1,1,0)96.** Thus, we will calculate the RMSE for this problem by using different number of days to forecast the traffic demand.

3 Continuous day

RESULTS:

SARIMAX model: non-seasonal (0,1,0) seasonal (0,1,1) AIC value: NAN

SARIMAX model: non-seasonal (0,1,1) seasonal (0,1,0) AIC value: -486.1440159785346

SARIMAX model: non-seasonal (0,1,1) seasonal (0,1,1) AIC value: NAN

SARIMAX model: non-seasonal (0,1,0) seasonal (1,1,0) AIC value: -498.9262513294788

SARIMAX model: non-seasonal (0,1,0) seasonal (1,1,1) AIC value: NAN

SARIMAX model: non-seasonal (0,1,1) seasonal (1,1,0) AIC value: -501.4890479227734

SARIMAX model: non-seasonal (0,1,1) seasonal (1,1,1) AIC value: NAN

SARIMAX model: non-seasonal (1,1,0) seasonal (0,1,0) AIC value: -484.30637245567243

SARIMAX model: non-seasonal (1,1,0) seasonal (0,1,1) AIC value: NAN

SARIMAX model: non-seasonal (1,1,1) seasonal (0,1,0) AIC value: NAN

SARIMAX model: non-seasonal (1,1,1) seasonal (0,1,1) AIC value: NAN

SARIMAX model: non-seasonal (1,1,0) seasonal (1,1,0) AIC value: -500.3756903957691

SARIMAX model: non-seasonal (1,1,0) seasonal (1,1,1) AIC value: NAN

SARIMAX model: non-seasonal (1,1,1) seasonal (1,1,0) AIC value: NAN

SARIMAX model: non-seasonal (1,1,1) seasonal (1,1,1) AIC value: NAN

SARIMAX model: non-seasonal (2,1,0) seasonal (0,1,0) AIC value: -485.3685323814775

SARIMAX model: non-seasonal (2,1,0) seasonal (0,1,1) AIC value: NAN

SARIMAX model: non-seasonal (2,1,1) seasonal (0,1,0) AIC value: NAN

SARIMAX model: non-seasonal (2,1,1) seasonal (0,1,1) AIC value: NAN

SARIMAX model: non-seasonal (2,1,0) seasonal (1,1,0) AIC value: -500.7435319263062

4 Continuous day

RESULTS:

SARIMAX model: non-seasonal (0,1,0) seasonal (0,1,1) AIC value: NAN

SARIMAX model: non-seasonal (0,1,1) seasonal (0,1,0) AIC value: -765.7187410173

SARIMAX model: non-seasonal (0,1,1) seasonal (0,1,1) AIC value: NAN

SARIMAX model: non-seasonal (0,1,0) seasonal (1,1,0) AIC value: -795.9567784943

SARIMAX model: non-seasonal (0,1,0) seasonal (1,1,1) AIC value: NAN

SARIMAX model: non-seasonal (0,1,1) seasonal (1,1,0) AIC value: -802.7405272198

SARIMAX model: non-seasonal (0,1,1) seasonal (1,1,1) AIC value: NAN

SARIMAX model: non-seasonal (1,1,0) seasonal (0,1,0) AIC value: -763.3985275688

SARIMAX model: non-seasonal (1,1,0) seasonal (0,1,1) AIC value: NAN

SARIMAX model: non-seasonal (1,1,1) seasonal (0,1,0) AIC value: NAN

SARIMAX model: non-seasonal (1,1,1) seasonal (0,1,1) AIC value: NAN

SARIMAX model: non-seasonal (1,1,0) seasonal (1,1,0) AIC value: -801.1174633330

SARIMAX model: non-seasonal (1,1,0) seasonal (1,1,1) AIC value: NAN

SARIMAX model: non-seasonal (1,1,1) seasonal (1,1,0) AIC value: NAN

SARIMAX model: non-seasonal (1,1,1) seasonal (1,1,1) AIC value: NAN

SARIMAX model: non-seasonal (2,1,0) seasonal (0,1,0) AIC value: -765.8073659865

SARIMAX model: non-seasonal (2,1,0) seasonal (0,1,1) AIC value: NAN

SARIMAX model: non-seasonal (2,1,1) seasonal (0,1,0) AIC value: NAN

SARIMAX model: non-seasonal (2,1,1) seasonal (0,1,1) AIC value: NAN

SARIMAX model: non-seasonal (2,1,0) seasonal (1,1,0) AIC value: -801.3547442389

5 Continuous day

RESULTS:

SARIMAX model: non-seasonal (0,1,0) seasonal (0,1,1) AIC value: NAN

SARIMAX model: non-seasonal (0,1,1) seasonal (0,1,0) AIC value: -1064.146295402

SARIMAX model: non-seasonal (0,1,1) seasonal (0,1,1) AIC value: NAN

SARIMAX model: non-seasonal (0,1,0) seasonal (1,1,0) AIC value: -1103.106697499

SARIMAX model: non-seasonal (0,1,0) seasonal (1,1,1) AIC value: NAN

SARIMAX model: non-seasonal (0,1,1) seasonal (1,1,0) AIC value: -1116.522174986

SARIMAX model: non-seasonal (0,1,1) seasonal (1,1,1) AIC value: NAN

SARIMAX model: non-seasonal (1,1,0) seasonal (0,1,0) AIC value: -1061.768212845

SARIMAX model: non-seasonal (1,1,0) seasonal (0,1,1) AIC value: NAN

SARIMAX model: non-seasonal (1,1,1) seasonal (0,1,0) AIC value: -1068.995298844

SARIMAX model: non-seasonal (1,1,1) seasonal (0,1,1) AIC value: NAN

SARIMAX model: non-seasonal (1,1,0) seasonal (1,1,0) AIC value: -1114.141709672

SARIMAX model: non-seasonal (1,1,0) seasonal (1,1,1) AIC value: NAN

SARIMAX model: non-seasonal (1,1,1) seasonal (1,1,0) AIC value: -1115.430444606

SARIMAX model: non-seasonal (1,1,1) seasonal (1,1,1) AIC value: NAN

SARIMAX model: non-seasonal (2,1,0) seasonal (0,1,0) AIC value: -1062.668841866

SARIMAX model: non-seasonal (2,1,0) seasonal (0,1,1) AIC value: NAN

SARIMAX model: non-seasonal (2,1,1) seasonal (0,1,0) AIC value: NAN

SARIMAX model: non-seasonal (2,1,1) seasonal (0,1,1) AIC value: NAN

SARIMAX model: non-seasonal (2,1,0) seasonal (1,1,0) AIC value: -1115.077815594

6 Continuous day

RESULTS:

SARIMAX model: non-seasonal (0,1,0) seasonal (0,1,1) AIC value: NAN

SARIMAX model: non-seasonal (0,1,1) seasonal (0,1,0) AIC value: -1064.146295402

SARIMAX model: non-seasonal (0,1,1) seasonal (0,1,1) AIC value: NAN

SARIMAX model: non-seasonal (0,1,0) seasonal (1,1,0) AIC value: -1103.106697499

SARIMAX model: non-seasonal (0,1,0) seasonal (1,1,1) AIC value: NAN

SARIMAX model: non-seasonal (0,1,1) seasonal (1,1,0) AIC value: -1116.522174986

SARIMAX model: non-seasonal (0,1,1) seasonal (1,1,1) AIC value: NAN

SARIMAX model: non-seasonal (1,1,0) seasonal (0,1,0) AIC value: -1061.768212845

SARIMAX model: non-seasonal (1,1,0) seasonal (0,1,1) AIC value: NAN

SARIMAX model: non-seasonal (1,1,1) seasonal (0,1,0) AIC value: -1068.995298844

SARIMAX model: non-seasonal (1,1,1) seasonal (0,1,1) AIC value: NAN

SARIMAX model: non-seasonal (1,1,0) seasonal (1,1,0) AIC value: -1114.141709672

SARIMAX model: non-seasonal (1,1,0) seasonal (1,1,1) AIC value: NAN

SARIMAX model: non-seasonal (1,1,1) seasonal (1,1,0) AIC value: -1115.430444606

SARIMAX model: non-seasonal (1,1,1) seasonal (1,1,1) AIC value: NAN

SARIMAX model: non-seasonal (2,1,0) seasonal (0,1,0) AIC value: -1062.668841866

SARIMAX model: non-seasonal (2,1,0) seasonal (0,1,1) AIC value: NAN

SARIMAX model: non-seasonal (2,1,1) seasonal (0,1,0) AIC value: NAN

SARIMAX model: non-seasonal (2,1,1) seasonal (0,1,1) AIC value: NAN

SARIMAX model: non-seasonal (2,1,0) seasonal (1,1,0) AIC value: -1115.077815594

SARIMAX model: non-seasonal (0,1,0) seasonal (0,1,1) AIC value: -1364.2501126375676

SARIMAX model: non-seasonal (0,1,1) seasonal (0,1,0) AIC value: -1267.7772377999704

SARIMAX model: non-seasonal (0,1,1) seasonal (0,1,1) AIC value: -1373.3876768815144

SARIMAX model: non-seasonal (0,1,0) seasonal (1,1,0) AIC value: -1323.676068939083

SARIMAX model: non-seasonal (0,1,0) seasonal (1,1,1) AIC value: -1362.7460984329487

SARIMAX model: non-seasonal (0,1,1) seasonal (1,1,0) AIC value: -1330.7986508851636

SARIMAX model: non-seasonal (0,1,1) seasonal (1,1,1) AIC value: -1372.8524170943265

SARIMAX model: non-seasonal (1,1,0) seasonal (0,1,0) AIC value: -1266.2772673550155

SARIMAX model: non-seasonal (1,1,0) seasonal (0,1,1) AIC value: -1372.163321949272

SARIMAX model: non-seasonal (1,1,1) seasonal (0,1,0) AIC value: -1271.881030920224

SARIMAX model: non-seasonal (1,1,1) seasonal (0,1,1) AIC value: -1376.0649956840964

SARIMAX model: non-seasonal (1,1,0) seasonal (1,1,0) AIC value: -1329.7662728593791

SARIMAX model: non-seasonal (1,1,0) seasonal (1,1,1) AIC value: -1371.4976021513444

SARIMAX model: non-seasonal (1,1,1) seasonal (1,1,0) AIC value: -1332.5549278576568

SARIMAX model: non-seasonal (1,1,1) seasonal (1,1,1) AIC value: -1375.7224087709646

SARIMAX model: non-seasonal (2,1,0) seasonal (0,1,0) AIC value: -1266.2022407273312

SARIMAX model: non-seasonal (2,1,0) seasonal (0,1,1) AIC value: -1371.6126643722923

SARIMAX model: non-seasonal (2,1,1) seasonal (0,1,0) AIC value: NAN

SARIMAX model: non-seasonal (2,1,1) seasonal (0,1,1) AIC value: NAN

SARIMAX model: non-seasonal (2,1,0) seasonal (1,1,0) AIC value: -1329.4735712175675

SARIMAX model: non-seasonal (2,1,0) seasonal (1,1,1) AIC value: -1371.126636871673

SARIMAX model: non-seasonal (2,1,1) seasonal (1,1,0) AIC value: NAN

After running my data training, the RMSE for the random input in my results are:

For qp03wf geohash6:

RMSE for using 3 continuous date for the next 5 time interval

0.01767602572413616

RMSE for using 4 continuous date for the next 5 time interval

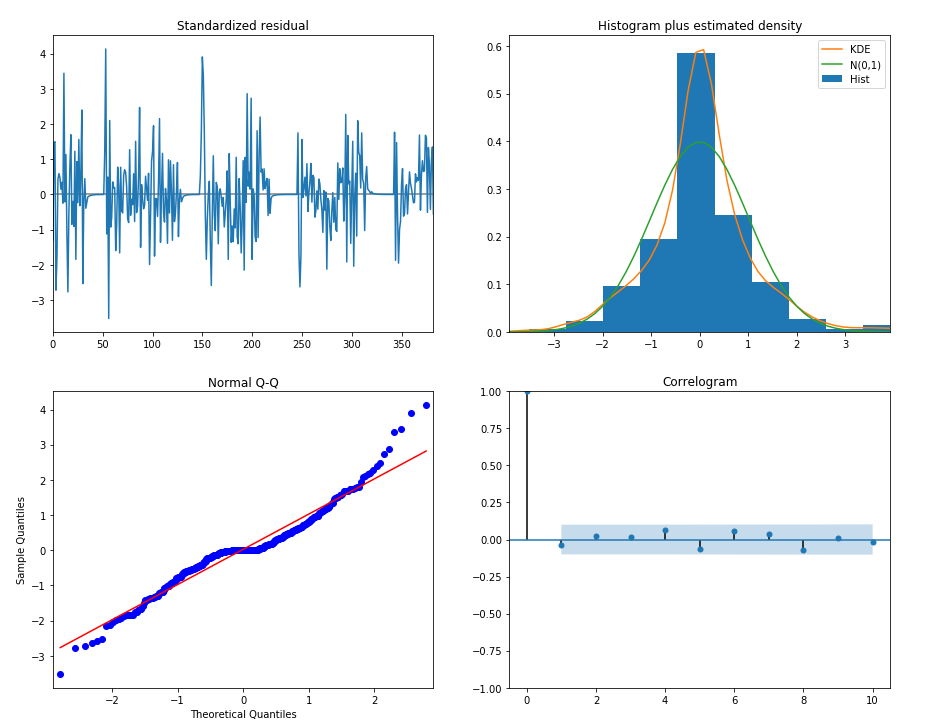
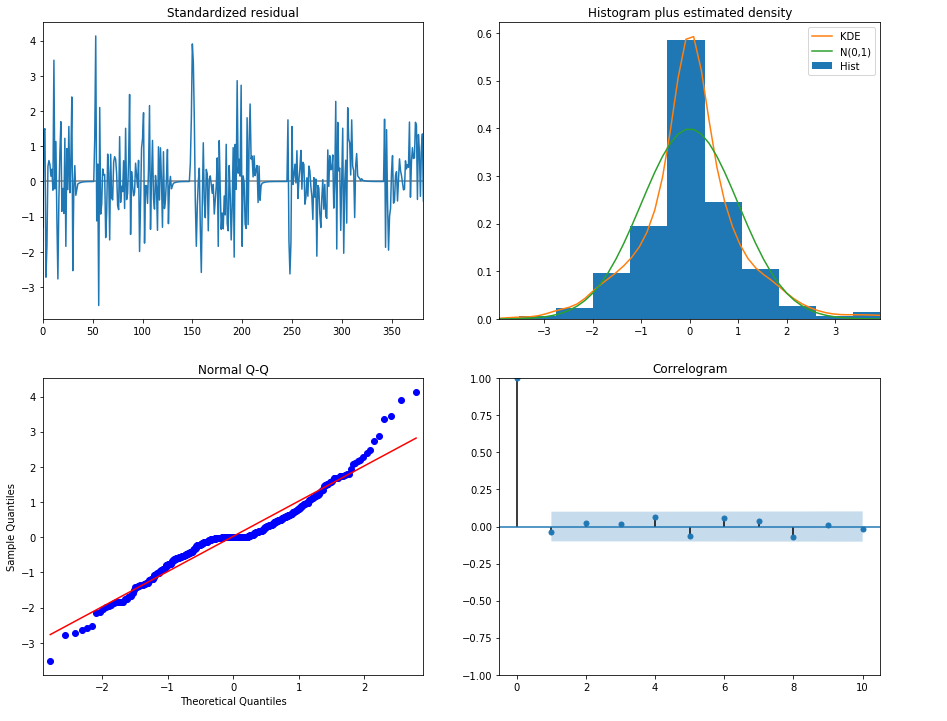
0.004601231856513220

RMSE for using 5 continuous date for the next 5 time interval

0.007118088544773944

RMSE for using 6 continuous date for the next 5 time interval

0.009125540871183672

**The Plot analysis for the SARIMA model**

**To run program**

1. Change the “training.csv” to whatever the “test.csv”. Run every row in Jupyter Notebook and it will automatically tidy up the data format to match with my program.

**Note: For testing in "training.csv", after pointing to the correct file, simply run all the code and this program will show all the related results.**

**By default it is grabData(separated\_data[geohash6[3]],100,4), which location = geohash6[3], start from time\_order 100 and continuous fir 4 days, meaning 100+(4\*96) = 484, it predicts the demand of 485 - 489 time\_order.**

1. There are some parts that to uncomment and comment out to focus, so that the testing can be done accordingly. (more explanation in the Jupyter notebook for the code)

*A)*

*# comment out this when doing testing*

*final\_test = grabData(separated\_data[geohash6[3]],,100,4)*

*stationarity = stationarity(final\_test)*

*# uncomment it when doing testing*

*# final\_test = grabDataTesting(separated\_data["geohash6--Name"])*

*# stationarity = stationarity(final\_test)*

*B)*

*# comment out this for testing*

*final\_test = grabData(separated\_data[geohash6[3]],100,4)*

*# ---------------------------uncomment this for testing-------------------*

*# second parameter is suggested to be 4 continuos day to predict the next 5 interval*

*# final\_test = grabDataTesting(separated\_data[geohash6[3]],4)*

*C)*

*# comment out this for testing*

*# plt.plot(grabData(separated\_data[geohash6[3]],100,5)[:-91])*

*# uncomment this for testing*

*# plt.plot(grabDataTesting(separated\_data[geohash6[3]],4))*

*D)*

*# comment out this for testing*

*RMSE = np.sum((final\_demand.as\_matrix() - grabData(separated\_data[geohash6[3]],100,5)[-96:-91].values.reshape(-1))\*\*2)/len(final\_demand)*

*# uncomment this for testing*

*# RMSE = np.sum((final\_demand.as\_matrix() - grabDataTesting(separated\_data[geohash6[3]],4)[-96:-91].values.reshape(-1))\*\*2)/len(final\_demand)*