horizontal line

Caesar’s - Staffing Problem

# OVERVIEW

The age old staffing problem has been at the heart of operations research since inception. This is a continuation of our project from *Programming for Analytics* from the previous semester. *Caesar’s staffing problem* is a real life case study from caesars’ resorts, where the objective is to forecast guest arrivals on any given day to determine staffing requirements.

***What’s the problem?***

Front desk staff is unionized in the Las Vegas region, as a result there are some strict guidelines to follow for staffing procedures. The staffing plan needs to be finalized two weeks in accordance with union guidelines. As such, Caesar’s corporate office needs to forecast arrivals at least two weeks in advance.

This becomes an efficiency problem. If front desk is understaffed, it creates unpleasant check in experience for customers, resulting in potential lost revenues. On the other hand, if the front desk is over staffed, Caesar’s loses money in extra wages. Hence, it becomes very important to forecast guest arrivals with reasonable accuracy.

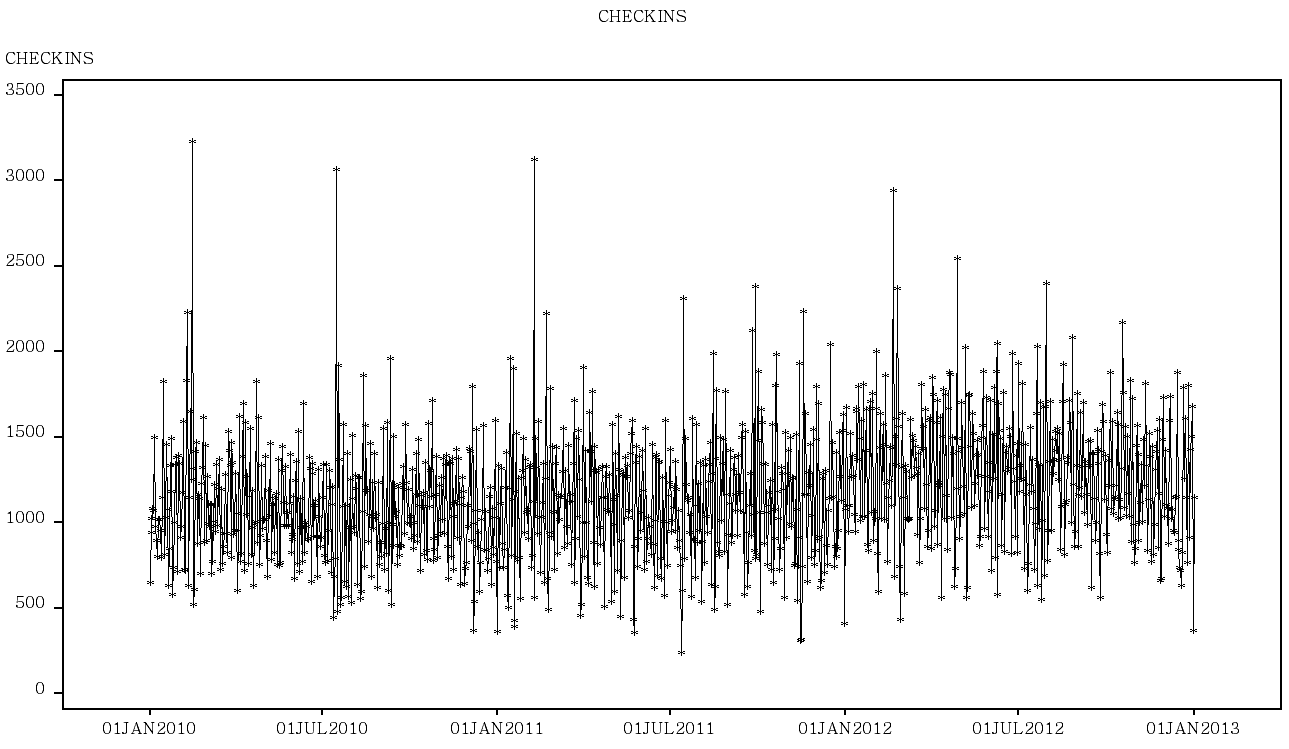
***What was the project last semester?***

The project last semester focused on **demonstrating programming skills** for a data exploration and causal forecast model to get more insights into what variables / holidays were affecting checkins. In fact, one of our recommendations was to carry out a time series forecast model. This is a perfect opportunity to develop such model to build upon our findings from previous semester’s work.

Link: <https://nbviewer.jupyter.org/github/bosea3000/Projects/blob/master/Caesars-Staffing-Problem/Checkin-Forecast.ipynb>

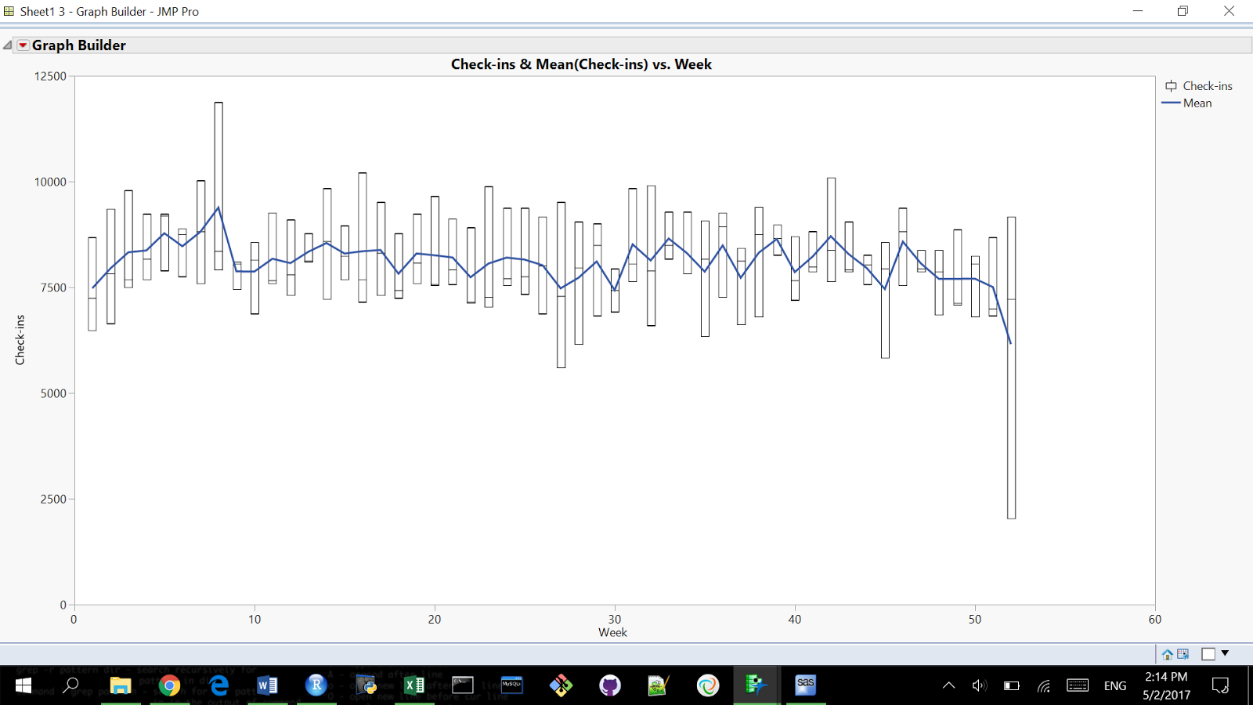
The series

Number of daily check-ins from January 2010 to January 2013

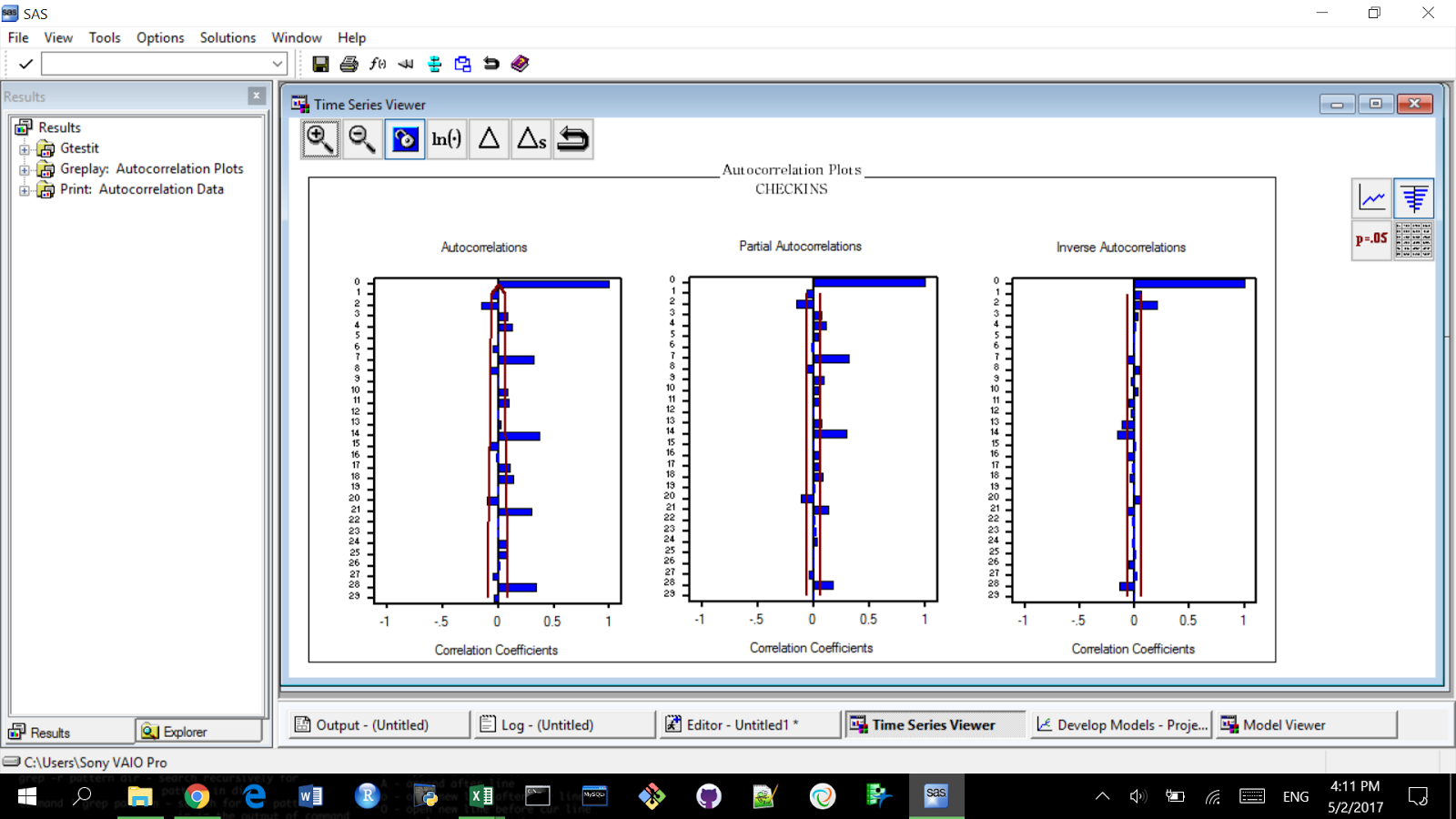


There is no visible trend, as the number of rooms remains constant over this relatively short period of time.

Weekly representation

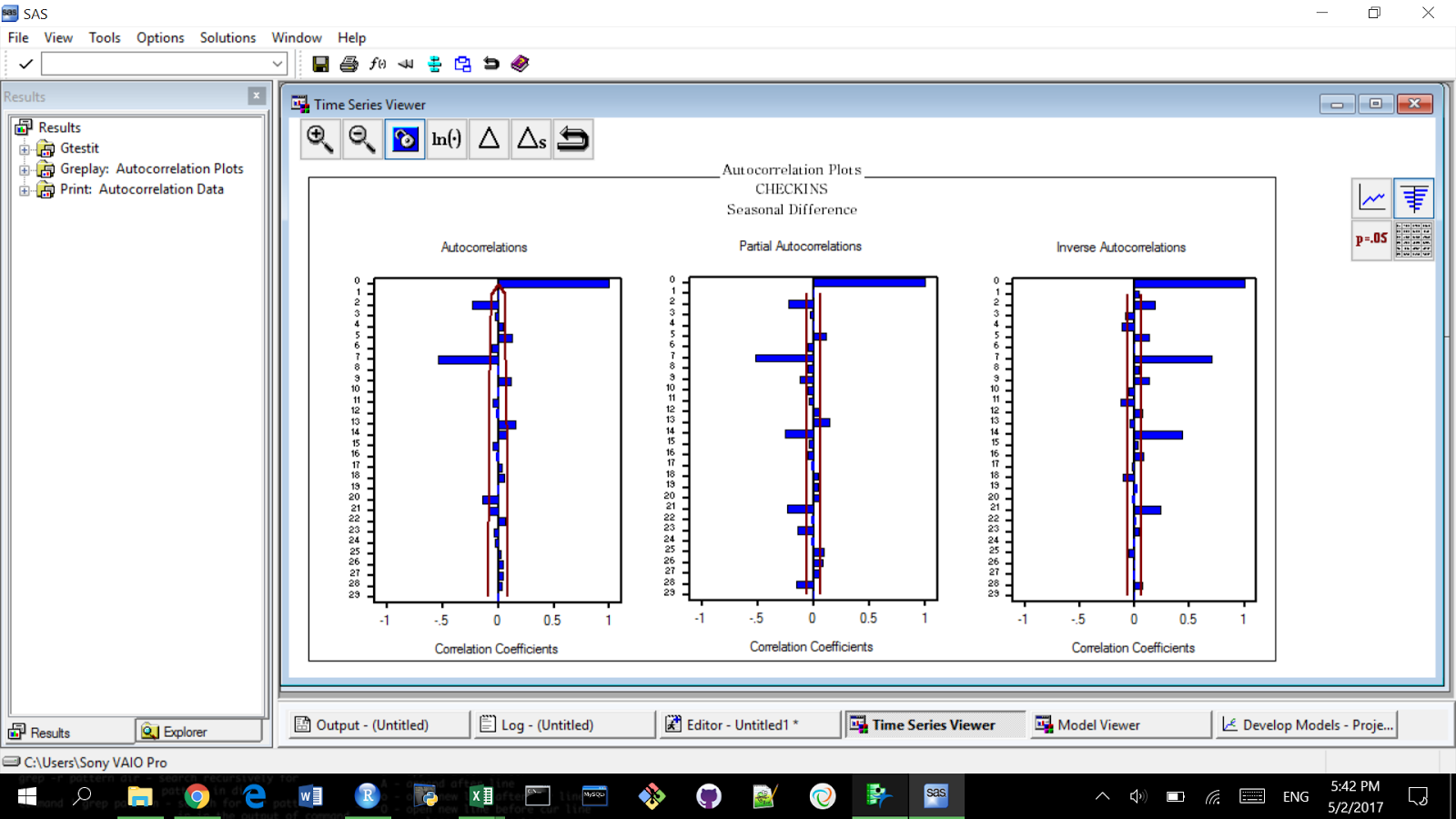


ACF | PACF - Original Data



ACF is clearly beyond the s.e. on lags 7, 14, 21 etc. We are definitely dealing with seasonal data.

Seasonal difference



Univariate model

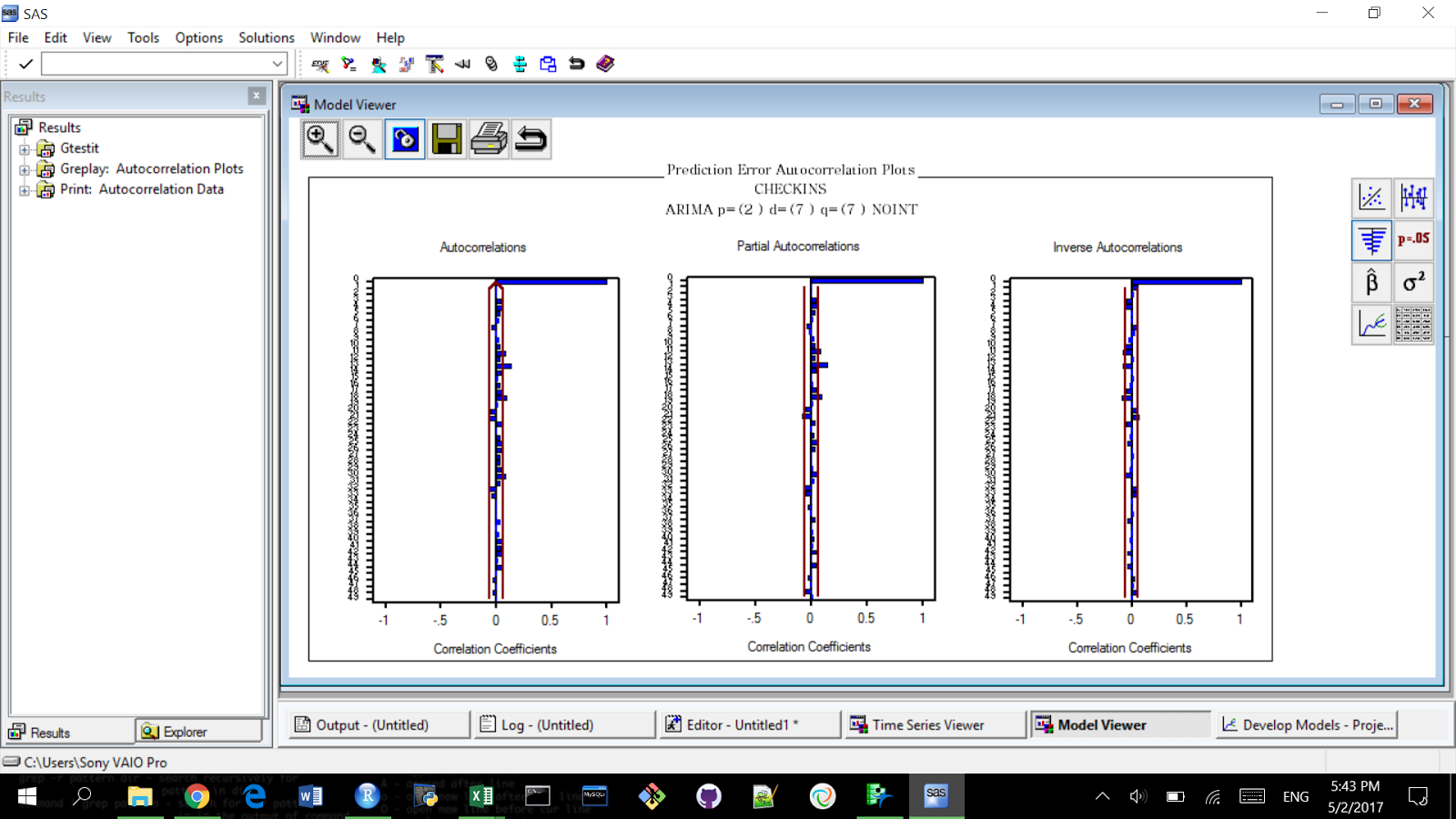
Reasoning for the next step:

ACF cuts off after lag 7 - MA(7) process

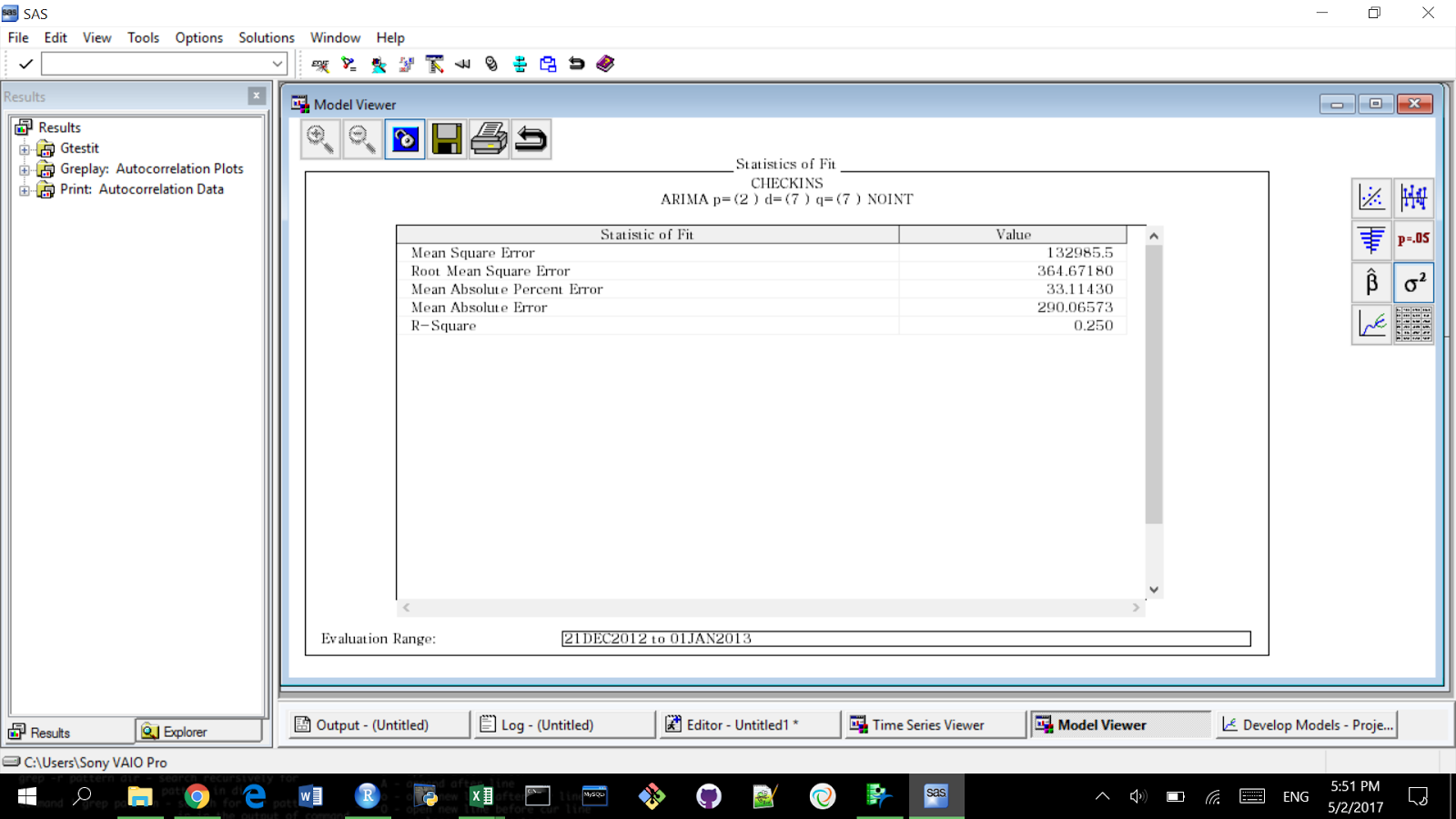
Differencing at the lag 7 - weekly seasonality

PACF cuts off at lag 2 - AR(2) process

**We choose to work with Factored ARIMA(2,7,7)**

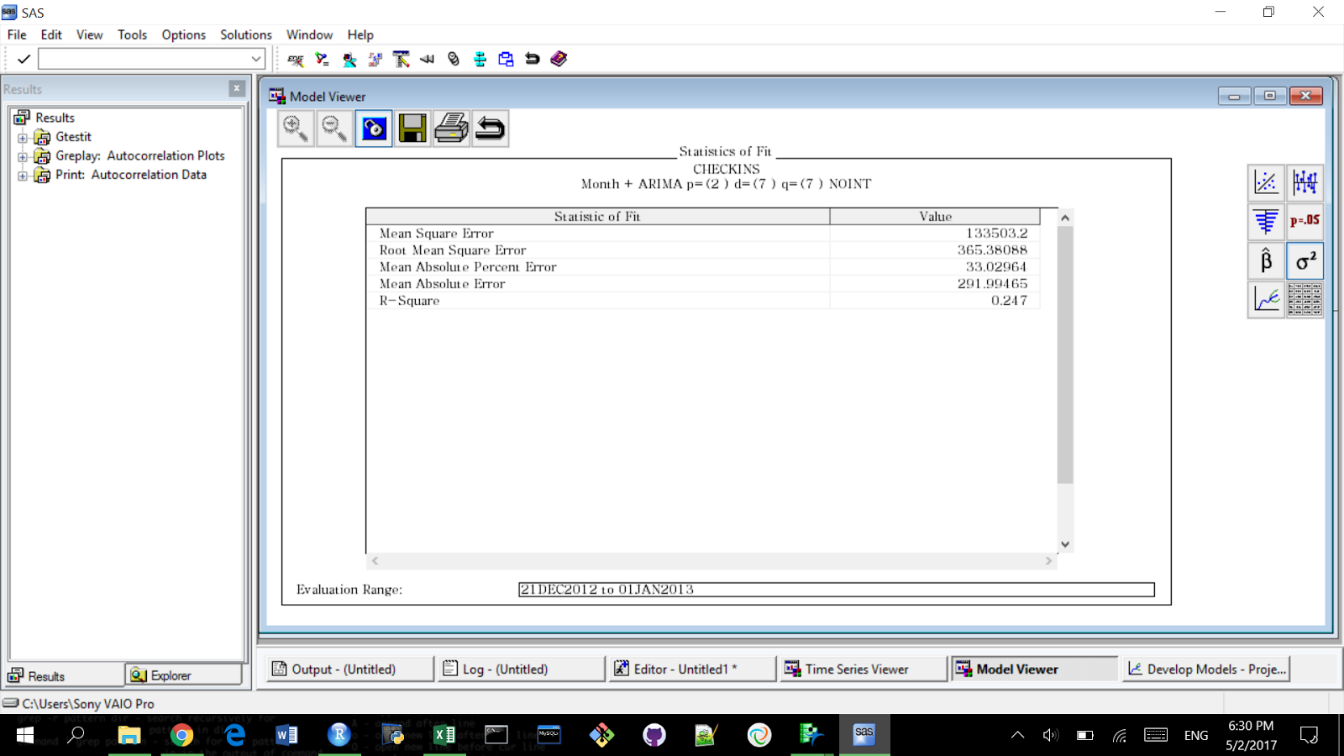


Factored ARIMA model removed insignificant terms and gave us a lower MAPE than the models we initially used in the draft.



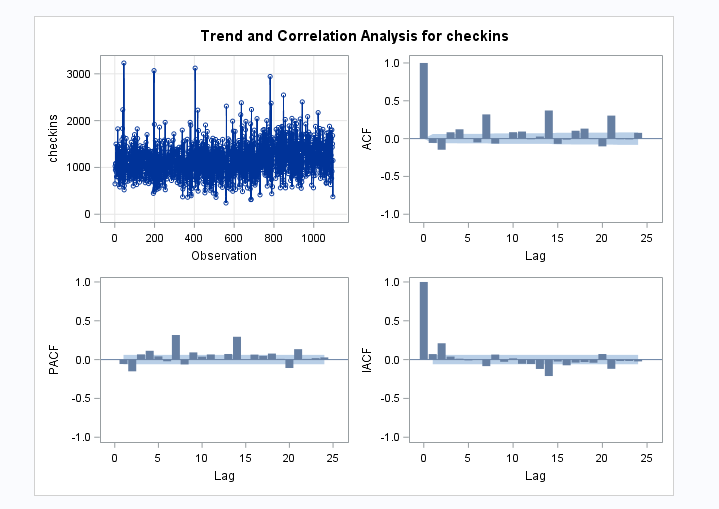
MAPE is at 33%

Adding a regressor MONTH into the model (with months with the highest number of checkins being 1 and others being 0) gives us the following results:

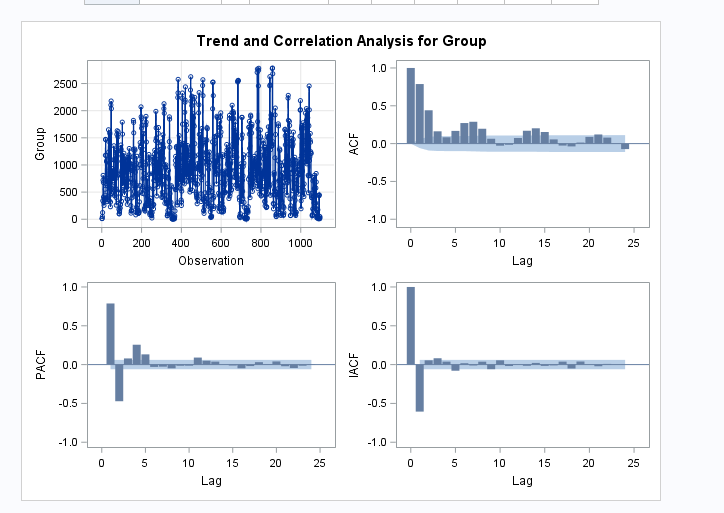


MAPE is slightly lower, and so is the R-Square.

Multivariate model

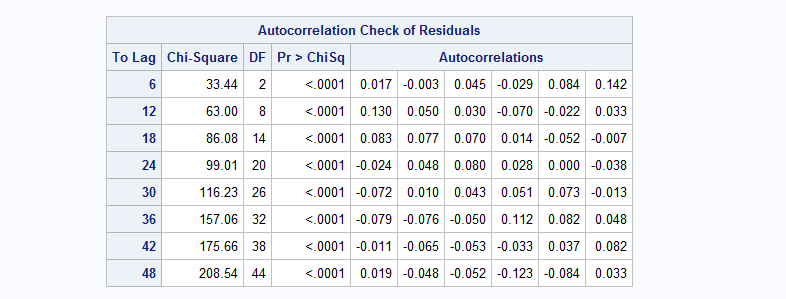


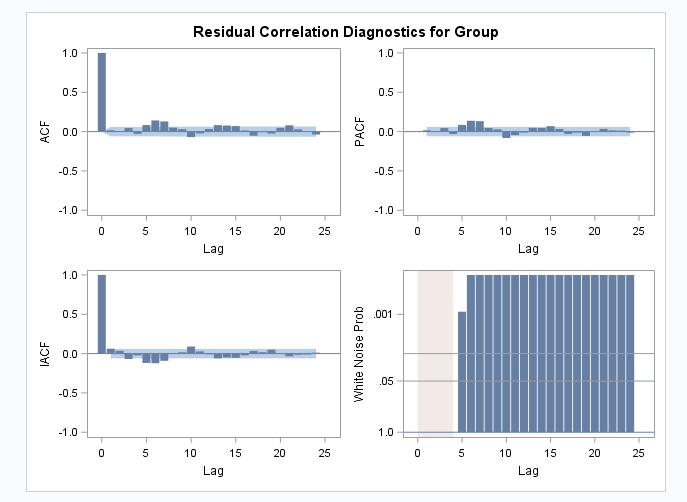
Check-ins above can clearly be seen as a non-stationary series with seasonal components



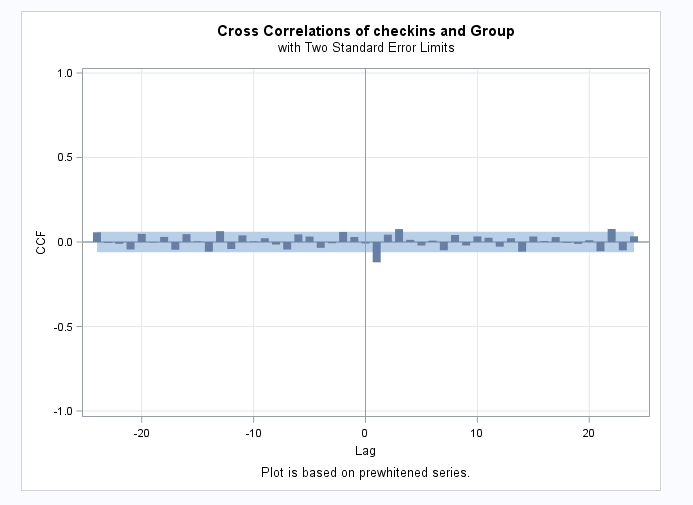
Similarly, for Group (a variable showing the number of people who checked in as a part of a group). They can be seen as non-stationary TS since the ACF decays slowly and is not chopped off.

As a result it is important to prewhiten the process using ARIMA modelling.



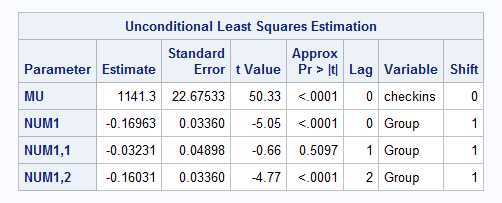


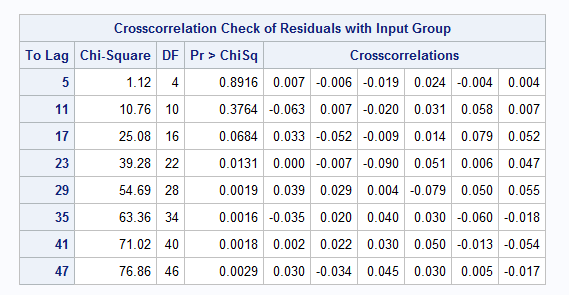
The P value for the chi-sq test is significant, however the autocorrelations are negligible. The reason that at 2 lags the ACF is out of 2-se’s is that with almost 1500 observation, the SE becomes really small, and even negligible values become significant. So for practical purposes we consider this to a pre-whitened series.



The above cross-correlation functions shows that X(1) precedes Y(1) and hence a TF model would be suitable.

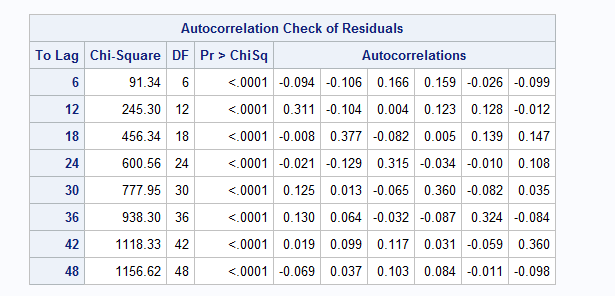
Running an estimation of b=1,r=1,s=2 we get the below model:



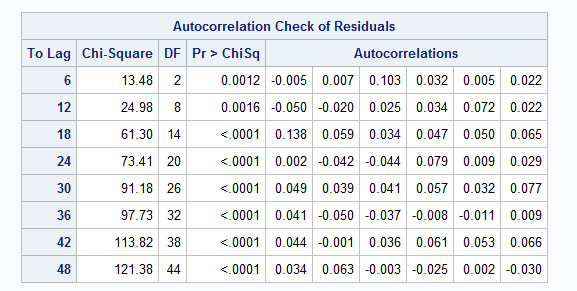


The cross-correlation p value is insignificant stating that the transfer model is adequate. Some of the values at higher lags are significant, but that is because of the high number of observations, the SE is very small.

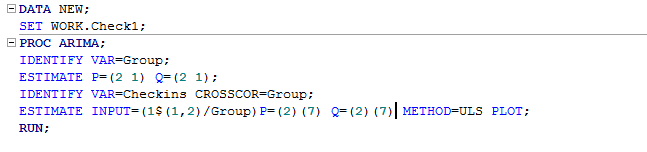
However as below the terms are significant we will use a TF noise model to get rid of it.



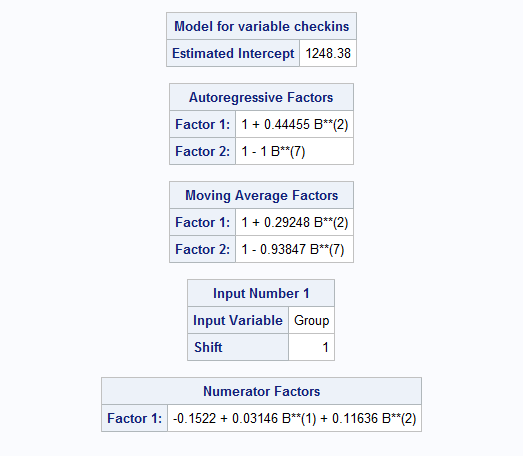
We use a P=(2 7) Q=(2 7) noise model to get rid of the auto-correlation. At higher lags, it is still significant but it is because of the low SE because of a very high number of observations.



The final code is as follows



The final TF model comes out to be



Conclusion:

After analyzing different models in the first and the second parts (at a certain point by trial and error), we have concluded that a factored ARIMA model is our best choice for univariate analysis. By adding an additional regressor into this model we saw that it does not change the overall picture. In other words, adding variable month into our model did not contribute into its predictive power.

In the third part a major issue we faced was prewhitening of the series. Working directly with SAS editor has proven to be somewhat of a challenge as we did not initially know all the terms we should have been using for creating a TF model.

In general, this project was highly valuable for us in terms of learning new analysis techniques and we think that we will refer to it more in the future when we face a challenge of analyzing a time series.

Thank you!