**DATA 624 Summer 2020, Project #1**

**Format**: Group Effort, Group Representative will turn in your assignment. No conversations across groups regarding this project.

**DUE**: 6/27/20 by Midnight ET

**Submission**: Via Email – scott.burk@sps.cuny.edu

**Submission**: Word Readable Document for Report (all in one), Excel Readable (all in one, separate sheets).

**File Convention:** Group#\_Project1\_Summer624, example Group1\_Project1\_Summer624

**GRADE:** 70% Report, 30% Forecast Accuracy

**Overview**

Your data is a de-identified Excel spreadsheet. Your assignment is to perform the appropriate analysis to forecast several series for 140 periods. You will have 1622 periods for your analysis. See Requirement #2 for more details.

**Requirement #1**

You will turn in a written report. You need to write this report as if it the report will be routed in an office to personnel of vary different backgrounds. You need to be able to reach readers that have no data science background to full fledge data scientists. So, you need to explain what you have done, why and how you did it with both layman and technical terminology. Do not simply write this with me in mind. Visuals and output are expected, but it is not necessary to include every bit of analysis. In fact, a terse report with simple terminology is preferred to throwing in everything into a long, ad nausem report. Story telling is really taking on for data science, so please flex your muscles here. The report is part 1 of 2 requirements.

*NOTE:* We have covered a lot of material. I don’t want you to try every method to create your forecast. But you should perform fundamentals, visualize your data, and perform exploratory data analysis as appropriate.

**Requirement #2**

Your second requirement is to produce forecasts. This workbook will contain at least 6 sheets where I will calculate your error rates. There will be one sheet (tab) for each Group – S01, S02, S03, SO4, S05, S06. You should order each sheet by the variable SeriesIND (low to high). Your source data is sorted this way, except there are all 6 groups present in one sheet which you must break out into 6 tabs. You will submit the data I sent AND the forward forecast for 140 periods. I want you to forecast the following

S01 – Forecast Var01, Var02

S02 – Forecast Var02, Var03

S03 – Forecast Var05, Var07

S04 – Forecast Var01, Var02

S05 – Forecast Var02, Var03

S06 – Forecast Var05, Var07

**Grading**

You team project will be graded in two parts. Seventy percent will be the report and 30 percent will be the forecast accuracy. Your report will be graded on value of information presented, brevity, readability (easy to read), accuracy and demonstration of content knowledge. Forecasts, sending your Excel file in the **right** format is very important as I will calculate measures of accuracy (ask me if needed). You need to include training and forecast data.

Your **final**, **individual** **score** will be weighted by your teammates. Each member will rate their fellow team members. Make sure your contribution is equal or better than others in your group.