**Time Series Project: Spring 2020**

**Due Dates:**

**EDA:**

**Saturday March 14 at 11:59pm**

**Final Project Documentation and Presentation:**

**Saturday, April 11 at 11:59pm**

**The project:**

1. Pick a data set that has two or more variables recorded over time (similar to the Schumway LA air quality data from Unit 12. (dataset: *lap* from package: *astsa*) (can’t use this one ☺ )

2. Select a response from the data set.

3. Be creative and come up with a scenario as to why a client would want to analyze this data and why this response is important! Or better yet, use a real problem that you are interested in!

4. Fit at least one model from ***each*** of the following four categories (provide all plots and tables needed to ID these models: acfs, spectral density, factor tables, etc.):

a. ARMA / ARIMA / ARUMA / Signal Plus Noise (univariate analysis)

b. VAR with at least one explanatory model.

c. Neural Network

d. Ensemble model using at least two of the above. (this model does not have to “beat” your

other models.

5. Compare all models with the ASE… this does not mean you have to choose the model with the lowest ASE.

6. Pick a forecast horizon based on your “problem” from part 3 above and provide the forecasts and prediction limits.

7. Create a ppt and a 7-minute video describing your analysis and steps 1 – 6 above.

8. Post that video to you-Tube and the (private) link to the Google-Doc and submit your ppt and Rmd File (or Jupyter notebook) to 2DS. Please leave the link on the Google Doc for a week so others can learn from your presentation. Please check out at least 3 of your peer’s presentations and please watch your own presentation as well. It is often very useful (although always a bit awkward for me at least ;) to watch yourself present!

**Groups:**

This is either a group project of 2 or less or you may complete it alone (group of 1). To be clear, groups must be 2 members or less, no groups of three or more. Communication between groups is encouraged although it is assumed that the analysis, write up (RMD or equivalent) and presentation materials are performed by the members of the group. Also, you may partner with someone from the other class as well if you like.

**Rubric:**

Knit RMD: 20%

Well organized and documented RMD knit to pdf or html.

Results / Analysis: 40%

Correct Interpretation

Creating Useful Models

Performing a Complete Analysis: Model ID, Model Building, Forecasting, Cross Validation

Presentation: 40%

* Communication and presentation of your findings are critical to being a successful data scientist. You will be graded on:
  + Voice inflection
  + Slide Organization / Content
  + Visualization
  + Composure: This will include **not reading** off of the slides and not smoothness of delivery.
  + Pace: Not going a second over 7 minutes. Your client is very strict on this point.

**Deliverables:**

Please submit

a. your pptx (or slides in whatever form (pdf, Prezi, etc.) (Put in “Final Project Presentation.”)

b. an R markdown or Jupyter notebook or equivalent (Put in “Final Project Documentation.”)

this file should contain all of your EDA, modeling and forecasting code and be very organized and well commented.

c. Please make sure your link is on your pptx and the Google Doc.