DATS 6101 – Fall 2018 – Project 2 Instructions

Goal

In this project 2, we focus on modeling building as well as presenting and communicating the model results effectively. We still will include developing the SMART question and brief EDA to complete the tasks. Your team can either continue to use the same dataset and topic from Project 1 and now move to model building. If your original dataset and project 1 does not work well with building a model using techniques that we learn, you are absolutely free to use a new dataset for project 2. 0 – 9

Instructions

Come up with a question independent of data gathering, gather dataset(s), conduct EDA and develop a model/models that answers your question(s). Address at least the following questions:

1. How did you develop your question and what relevant research has already been completed on this topic?
2. How did you gather and prepare the data for analysis?
3. How did you select and determine the correct model to answer your question?
4. How reliable are your results?
5. What predictions can you make with your model? Examples
6. What additional information or analysis might improve your model results or work to control limitations?

Develop a presentation that provides an overview of your results, inclusive of the limitations and be prepared to demonstrate your knowledge in class. **25 minute presentation max**

Deliverables

Please turn in your team's final copy of each of these items to me directly in email:

- Powerpoint or other presentation files that you used (emailed to me by end-of-day of presentation).
- The dataset that you used. If the file is too large, you may simply direct me to the source of the dataset.
- Rmarkdown file that shows the codes you used for your analysis.
- Written report, in Word, pdf, RMarkdown, LaTeX or any other common formats, that is in "publication form" intended to be submitted to a journal.
- Other than the first item, all others are due on the Friday midnight after the presentation, so that I can have the final grades for everyone on time.

Grading (See Rubric)

1. 33% Summary Report
2. 33% Model Development, Selection and Usage
3. 33% Presentation