

**Applied Regression and Time Series Analysis  
(DATASCI W271)  
2016 Fall**

**Professors Devesh Tiwari and Jeffrey Yau**

**Lab 3**

**Instructions:**

- Unlike the homework assignments, lab 1, and lab 2, this is a semester-long lab that must be done in groups; form a group between 3 to 5 students.
- Lab 3 is worth **30% of your final grade**
- Your group will
  1. propose a data science question (or a set of related questions) that can be addressed using classical linear regression models or multivariate time series models
  2. find a data set that is suitable for answering the question your group propose
    - state the source of your data
    - the data must be approved by the instructor
    - **Note: It cannot be a proprietary data set, such as a dataset that owns by your company, owns by the clients or partners of your company, provided by data vendors, etc.**
  3. Conduct a comprehensive quantitative analysis using classical linear regression models or multivariate time series models and the techniques, such as EDA, model diagnostic, formal assumptions testing, taught in the course to answer the question(s) your group propose
  4. Write your analysis in a report submitted in a pdf file alongside with a well-written, easily understood R script (in the form of jupyter notebook, Rmd file, or R file) that is used to generated all the analyses included in your report
    - More details about the report will be provided later
  5. Make a 15-minute presentation presented in the live session in the final week of the course (i.e. the week of December 12<sup>th</sup>)
- Deliverables and Due Dates:
  1. **Due September 30<sup>th</sup>: Proposal**
    - Points worth: 0 (**However, late submission will receive 10% reduction in the final grade of the project.**)
    - The proposal will not be graded; however, the instructor will provide comments to your proposal to ensure that your question is not too broad, too narrow/simple, or otherwise not appropriate to be addressed by techniques used in this course.
    - The proposal should include the questions, a proposed dataset(s), and the identification of the dependent and in primary independent variables in your dataset. Note that analysis of the dataset(s) is not needed at this point.
  2. **Due October 28<sup>th</sup> : Revised Proposal**
    - Points worth: 0 (**However, late submission will receive 10% reduction in the final grade of the project.**)
    - The revised proposal will not be graded; however, the instructor will provide comments to your revised proposal
    - The revised proposal should include a finalized dataset(s), but analysis of the finalized dataset(s) is not needed at this point.

**3. Due December 9<sup>th</sup>: Final Report and the associated R script/Jupyter notebook/Rmd file**

- **Points worth: 80 (out of 100)**
- Provide the rationale of the choice of your methods (e.g. why is classical linear regression a suitable methodology to answer the question(s) using your dataset?)
- As in lab 1 and lab 2, all the steps used to arrive at your final model and final answers (to your proposed question(s)) need to be shown clearly; these steps are as important as the final answer. Do not just print a set of graphs, tables, and statistical model results without any explanation.
- Conduct EDA and use insights generated from your EDA to build your statistical model
- The steps used to build the models must be shown and explained clearly; multiple candidate models should be considered and explanations of the choice of the final model be given
- Conduct model diagnostics
- Conduct model performance evaluation
- Conduct formal testing of the statistical assumptions
- Provide remedies when underlying one or more of the assumptions are violated
- Most importantly, use your model results to answer the data science question proposed

**4. Due at the beginning of the live session in the week of December 12<sup>th</sup>: Presentation**

- **Points worth: 20 (out of 100)**
- When submitting your files, use the following naming convention of your file:
  - SectionName\_TypeOfFile\_LastNameFirstLetterofFirstNameOfStudent1\_LastNameFirstLetterofFirstNameOfStudent2\_LastNameFirstLetterofFirstNameOfStudent3.fileExtention
  - For example, the proposal submitted by Jeffrey Yau and Devesh Tiwari from Section 1. Their pdf report will have the file name “Section1\_Proposal\_YauJ\_TiwariD.pdf”