Linear Models Final Project – Spring 2019

Due Date: May 8, 2019

For your Linear Models course final project, you will apply what you've learned in the course to use linear modeling in the field of data science. You will build a regression model and report on the model results and diagnostics.

You will work on the project throughout the second half of the semester, beginning in Week 6. Your final deliverables for the project will be a report and the RStudio scripts for all your analyses. Information on Project Evaluation & the Grading Rubric is provided below.

Final Project Instructions:

A. Select Dataset
B. Complete 7 Project Steps
C. Compile & Submit Final Report

A. Select a Dataset

Select a dataset from two options provided in the table below:

Data Description	Response Variable	Notes	Download link
Stock Portfolio	Annual Return (or another as listed)	There is data for four periods; choose only one to analyze	Download Stock Portfolio dataset (.xlsx file)
Communities and Crime	ViolentCrimesPerPop: total number of violent crimes per 100K	Need to remove rows with missing data	Download Communities and Crime dataset (.txt file)

B. Complete 7 Project Steps

Beginning in Week 6 of the course, each week (except the week of the Midterm Exam) you will complete parts of the final project, culminating with your final project report which is due at the end of the course.

You will complete your drafts for the final project components in Steps 1–7 according to the schedule below with opportunity to revise before submitting the final report .

Completing drafts of Steps 1–7 on time and satisfactorily throughout the semester will count for 7% of your total grade for the course. Each step is described in greater detail within the course during the week due. **See the Course Schedule for dates.**

	Project Step	During Week
1	Exploratory data analysis (EDA)	Week 6
2	Fit a linear model	Week 8
3	Perform model selection via automated selection	Week 9
4	Apply diagnostics to the model	Week 10
5	Investigate fit for individual observations	Week 11
6	Apply transformations to model as needed	Week 12
7	Report inferences and make predictions using your final model	Week 13
Project Report (instructions below)		**Week 14**

C. Compile & Submit Final Project Report

For your project report you will revise and finalize Steps 1–7, and compile them into a final report. Along with these project steps, the final report should include an Introduction, Conclusion, and Appendix with the R scripts the code of your analyses.

Report Guidelines

The main narrative of your report should explain what you set out to accomplish, the steps you took, and the results of those steps. It should also document any intricacies of the data that impacted the analysis. For example, were outliers removed? If so what criteria were used to determine which would be removed.

Where appropriate, you should describe your methods for fitting, tuning and refining the model. You should describe important decisions made during the analysis process.

- Evaluate your audience and target the report to them. However, you should also consider secondary audiences.
    - For example, you may be writing your report for a product manager who has only a limited background in data science. That manager may pass on your report to more technical data scientists within their organization who will want to see more details.
- Your report should be well-organized and concise.
- The report should include graphics and output components in line with the text when appropriate and relevant to the story you are telling.
- It goes without saying that the report is a business document and should be free of spelling and grammatical errors. It is your responsibility to review and proofread the work.

Submit your final project as a ZIP file in Nexus containing:

- final version of the project report as a PDF document (10-15 pages in length for text and graphs only, not including code)

R Code Scripts

- The code should be human readable to allow a person to quickly find sections that are relevant. Anyone that has access to the original data and your code should be able to completely repeat your analysis.

Project Grading & Evaluation Rubric

The final project counts for **25% of your course grade** (Steps 1–7 will be evaluated throughout the semester and together, their grade will count for 7% of your course grade with the final project report grade making up 18%). See the Grading Rubric that follows for specifics on how your project will be evaluated.

- The drafts for Steps 1-7 will be scored out of 10 points each for a total of 70 points. The drafts submitted throughout the semester count for 7% of your course grade.

- The final project report will be scored out of 100 points. The final report counts for 18% of your course grade.

Scoring for the components of the final project is broken down as follows:

Project Grade Breakdown (25% of grade)

Project Component	Raw Points	Grade Weight	Weighted Points
Project Draft Steps	**70**	**7%**	**7**
Final Report • Introduction • Steps 1–7 • Conclusion • R Code (in Appendix) • Writing quality & grammar	**100** 15 7@8pts each 15 4 10	**18%**	**18**
Total possible		*25%*	*25*

Example: You earn 56 raw points for the draft steps. This contributes 56/70*7 = 5.6 points to your course grade. You receive 95 raw points for the final report. This contributes 95/100*18 = 17.1 points to your course grade. Your total weighted score for the project is thus 5.6 + 17.1 = 22.7. That is, your project has earned you 22.7% (out of 25% maximum) for your overall course grade.

Grading Rubric for Project and Steps

	Excellent	Good	Satisfactory	Below Satisfactory
Draft of Project Steps 1–7	There is a clear and comprehensive understanding of the work that was carried out. Excellent quality code. Submitted on time.	There is a clear understanding of the work that was carried out. Good quality code. Submitted on time.	There is some understanding of the work that was carried out. Poor quality code. Submitted on time.	No clear rationale as to why or how work was carried out. Not submitted on time.
Final Report: *Introduction*	Clear and concise explanation of the study's objective. Demonstrates interesting and nuanced insights into	Clear and concise explanation of the study's objective. Demonstrates some insights into the dataset.	Unclear explanation of the study's objective. Few insights into the dataset.	No explanation of the study's objective. No insights into the datasets.

	the dataset.			
Final Report: *Steps 1-7*	Provides a complete description of the analysis process, good insights into the analysis including what worked and what didn't, and an overall accurate and comprehensive understanding of the analysis.	There is a semi-complete description of the analysis process, decent insights into the analysis including what worked and didn't work, and an accurate understanding of the analysis.	Parts of the description of the analysis process are missing. Lacking in number and quality of insights into the analysis.	Little to no description of the analysis process. An inaccurate understanding of the analysis.
Final Report: *Conclusion*	Excellent summary of the results and findings with relevant concluding thoughts.	Good summary of the results and findings with relevant concluding thoughts.	Poor summary of the results and findings. Irrelevant concluding thoughts.	No summary or concluding thoughts provided.
Final Report: *Code*	Excellent quality code. Good usage of spaces and comments. Code can be rerun and produce the same results reported.	Good quality code. Adequate usages of spaces and comments. Code can be rerun and produce the same results reported.	Poor quality code. Little to no usage of spaces and comments.	No code or the code does not compile due to errors.
Overall Quality of Writing	Written in a tone appropriate for the audience. No errors in spelling or grammar.	Written in a tone appropriate for the audience. Few errors in spelling or grammar.	Written in a tone appropriate for the audience. Some grammatical and spelling errors throughout.	Inappropriate tone for intended audience. Extensive grammatical and spelling errors throughout.