Individual Assignment 4

MSCI 718

Instructions: Complete this assignment individually and submit online to the LEARN Dropbox as a ***PDF (max 2 pages)****. Also upload any source files, for example, and any .R or .Rmd files that you used. If you discussed problems with other students, please describe the nature of that discussion (e.g., on Team, in a study group). Remember, you can discuss approaches to problems with other people, but* ***the work you submit must be your own.***

Learning Objectives: In this assignment, you will perform a multiple linear regression, including ***at least one comparison between two models.*** *You will estimate the coefficients and assess the accuracy of coefficient estimates. Finally, you will need to assess and describe the accuracy of your model, and argue for choosing the set of parameters you chose either because of its predictive power, or because it led to a valuable insight.*

For your report, include the following sections:

1. Problem statement and data used: You are interested in getting into real estate, and hear that Australia's markets are hot right now. You've picked a city - Melbourne - but before you think about what home to buy, you want to understand the market a bit better. **Analyze this Melbourne housing dataset** to answer a question that will give better insight into your purchase, using at least one model comparison or method of variable selection: https://www.kaggle.com/dansbecker/melbourne-housing-snapshot. Some ideas:
    a. The first rule of real estate is "location, location, location". Is this true? Compare the influence of location vs. features of the home. You may need to refine this into a more specific question.
    b. Create and evaluate the most accurate model predicting a home's price (pick just one of a house, a unit, or townhome, and compare within that set).
    c. Find the areas in town that will give you the biggest house for the price.
    d. Imagine you are selling a 2-bedroom, 2-bathroom home in a given location. What's the best upgrade you can make to it before selling?
    e. Determine the three most important factors determining house price.
2. Planning: Plan your analysis based on the problem statement. Include data wrangling, assumption tests, and any other analyses you may need to conduct.
3. Analysis: Conduct a multiple linear regression, along with any appropriate assumption and accuracy checks. You will need to choose from two or more predictor variables, and compare your final model against other possible options (either through variable selection or to answer a research question).
4. Conclusion: Write up your analyses in a report to your manager (who has not taken MSCI 718, and does not understand statistics) and explain your conclusions.

Grading Criteria You will be graded against the same rubric that was used for pair assignments 1 and 2. We will be looking for a number of things:

1. Data selection: How you chose data and your problem statement. We will consider how your problem statement fits both the data, and the method (regression) that you need to apply.
2. Process and clarity of thought: Did you clearly apply a step by step process? Why did you chose the steps you did? Did you miss out on anything? Some examples would be ensuring you wrangled your data appropriately, checked the data against any assumptions required for the analysis, and did not do any work that was not required for this specific test.

3. Presentation: Use bullets, subtitles, and the APA style for your report. Use simple language, do not use complicated words unless they help you describe a complex idea. Use graphs judiciously. Show your work - if you do not include something in the main report, the teaching staff will not grade it.