STA302H1F/STA1001HF: Final Project 1 Due on 22nd August, 2021 11:59 PM Sharp on Crowd mark

Final Project

The final project will be due on Aug. 22, 2021 by 11:59PM EST. It will not be possible to extend the due date since we have to grade all the projects and submit the course grades. Students will be required to demonstrate their understanding of the methods taught in lecture by developing a reasonable regression model using the techniques taught in class. The students will be responsible for choosing the correct methods to apply and providing appropriate justifications where necessary. **This is a formal report and therefore it must contain the following sections:**

- Introduction section: provides details regarding the question you wish to address, why the model is being developed, how you intend to go about developing the model, and finally how the model meets the purpose mentioned earlier.
- Exploratory data analysis section: a detailed description of the variables in the data set with appropriate tables or figures that highlight certain characteristics of your variables that you deem important to mention.
- Model development section: a detailed discussion of the process used to come to the final model. Justifications may be both statistical and empirical in nature.You should also have as well as in-depth diagnostics to illustrate the 'goodness' of the model.
- Conclusion section: restate why the model is useful in the context of the data, provide an interpretation of the final model in non-technical language (i.e explain how the variables work, discuss predictions), and discuss any limitations/problems remaining with the model and how they might impact its use in the real world.

The final project will be done alone or in groups of 2, and must be typed and submitted by the stated deadline. **If you work in a group then only one of you should hand in the project with your group members names on the front page**. In marking your project, I will take into consideration the difficulty of the analysis that you attempt. A simple analysis with few errors may be graded higher than a more ambitious analysis with more errors.

The breakdown of the report will be as follows:

Item	**Break down**
Introduction and data description	10
Exploratory analysis and Model Development	25
Conclusions and discussion	10
Quality of writing	5

For the data set, we tracked how many hours students studied for each assessment, how many hours they thought about COVID, and their performance.We also determined where each student is currently stationed as they participate in STA302, due to the impact it may have on the hours a student considered COVID-19.

We interested in addressing the following question:

- *What are the factors that predict student performance on the final STA302 assessment (i.e. quiz 4)?*

There are several things to consider. You may want to look at the descriptive statistics (i.e. charts and graphs) to determine what to do pertaining to the following things below.

- You may wish to take an average of a person's three earlier quiz scores to create a new variable, or you may chose to do something else.
- You may wish to take an average of the number of hours where a person considered COVID-19, or you may choose to do something else.
- You may want to think about whether you have different models for different regions, or if the countries should be added as a predictor variable.
- If you do add the country as a predictor variable rather than creating separate models, you may want to aggregate similar regions together to reduce the number of categories.

Your report should be as long as is needed to be a convincing exploration of the question at hand, but not so long that it becomes excessively so. However, when we (the graders and I) read your report it needs to have a logical flow. **Any code should be displayed in the Appendix section of the report; please do not display this in the body of the formal report.**

While you will be using R for your analysis, should consider using R Markdown. If you use LaTex this is fine as well. At the end make sure the report is either in a .doc file or a pdf file. In order to pass the course, *everyone* must submit the final project. At the end of the submission include a detailed paragraph of each group members contribution to the project (if you work in groups). **This will be used to determine individual final grades**.