

# MDA 9159A: Data analysis project

## Project description

Your group project is to perform a regression analysis and write a report of your analysis. The final product of the project will be a written report. The report should include the following:

- Introduction: Motivate the analysis with clear goals.
- A general description of the data set and variables in the data
- Analysis: Describe how you used regression to analyze the data set. Specifically, discuss how you carried out the steps (model definition, estimation, interactions, checking model assumptions, variable selection etc.) in analysis discussed in class.
- Summary of the output and results
- Discussion of the findings and inferences
- Limitations and further questions raised by your study

## Group Selection

You can form your own group (Each group must have two members). An email that requests a group formation must be sent to [hgweon@uwo.ca](mailto:hgweon@uwo.ca) by **October 24, 2021, 11:59 PM**. You will be assigned a group by the instructor if you skip this task.

The request email must contain the following:

- Originate from your UWO email account.
- Have your group member CCed on the email.
- The subject should be “[MDA 9159] Project Group – Name”. Replace “Name” with some group name.
- Contain a list of names and IDs of all group members.

Please send one email per group.

## Data selection

You may select any dataset, so long as the response variable is a continuous variable (valid for regression modelling) and the data contains at least 8 predictors and 1,000 observations. The data set may be relevant to any field of your interest. If you are not sure whether your data is appropriate, please do not hesitate to ask (you can send the data set or the link to [hgweon@uwo.ca](mailto:hgweon@uwo.ca)). The two most common open sources of data are:

- UCI Machine Learning Repository (<https://archive.ics.uci.edu/ml/datasets.php>)
- Kaggle (<https://www.kaggle.com/datasets>)

The data set of your choice must be approved by the instructor no later than November 7, 2021, 11:59 PM. Send me an email ([hgweon@uwo.ca](mailto:hgweon@uwo.ca)) that includes a description of the data set (e.g., domain, # obs, response variable, # predictors etc.) you will analyze and the URL of the data (or actual file).

If your group skips this data selection task, a data set will be assigned by the instructor.

## Final Report

- The report for the project should be a 5 – 20 pages paper (this does not include the R codes).
- **The due date for your final report is December 1.** It should be submitted to [hgweon@uwo.ca](mailto:hgweon@uwo.ca) by a single group member. Aside from the report, you also need to submit your R codes (.R file).
- More details about project evaluation will be posted later.