

# Introduction

After providing descriptive statistics of the working dataset, applicable attributes, unusual patterns, etc., and understanding the scope of the datasets, you will think about simple replications/simulations based on the literature review report and the tentative methodology. You can also come up with new approaches, apply them, and document your findings. This stage is an iterative and incremental process. To enable the versioning of your analyses and solutions, you need to use the [GitHub](#) account that you created earlier to check in the conducted investigations as you advance in this project. The designated supervisor may ask you to walk them through the analyses to ensure the validity of the applied techniques and the feasibility of your project by the end of the semester. Given this, it would be best to check in a commented and structured set of source code files.

## Topics and Learning Objectives

### Topics

- Data analyses, including summarizing and visualizing data
- Data preparation, including selecting, preprocessing, and transforming data
- Models evaluation, including testing options, exploring algorithms, and reporting results

### Learning Objectives

By the end of the module, you will be able to:

- Provide a proof of concept ensuring the feasibility of the proposed solution
- Summarize and visualize data features
- Format, clean, sample, scale, decompose, or aggregate data
- Build and evaluate at least one model answering question
- Document and check in your codes using the GitHub repository

## Assessments

### Initial Results and the Code (10% of your final grade)

A link to a repository on GitHub is expected to be shared with your designated supervisor and uploaded as a text-based file to the D2L course shell assignment link. This repository should include:

- A read.me file listing the tentative stages of the project and outlining the content of the repository
- A set of technical reports either in R/rmarkdown (R/RMD) or Python/JupyterNotebook (PY/IPYNB) format
- A compilation of the uploaded technical reports, either in HTML or PDF format
- The working dataset, or a link to it

Please ensure the validity of the links, the availability of the working datasets, and the readability of the source code files.

**This assignment is due by 11:59 p.m. EST on Monday of Week 9.** Submit your assignment to the D2L “Initial Results and the Code” assignment link. Late submissions will be penalized with 10% loss a day, up to a maximum of five business days. After five business days, it is up to the designated supervisor to accept or reject your submission. In the case of acceptance, the submission should be marked out of 50%.

## Resources

Your supervisor might suggest a set of references for you to review based on your project's theme.