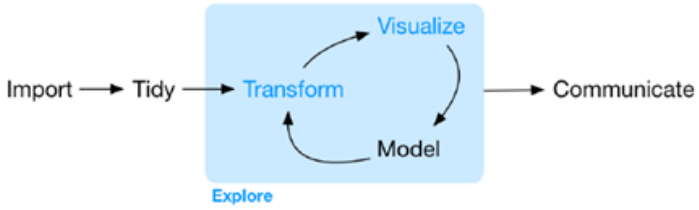
Description and Grading Rubric for Project

Data Science Project Report

The primary purpose of this assignment is for students to gain experience carrying out and reporting an original data science project that utilizes some of the various methods and techniques learned in class. Although we have had in-class and homework assignments throughout the semester, learning is best solidified by an independent project.

Students need to identify a data set to use for this project. This data can be data already collected by the student or their lab, or this data can be a public data set retrieved from online. If the latter, the data just need to be "real". That is, it cannot be simulated data that was created specifically for data science training (as an example). If you are using prior data that you have collected and reported on, your class project just cannot simply be a recreation of a prior report (e.g., your thesis). Finally, if a student is unable to locate an appropriate data set, they should let me know as soon as possible and I will help facilitate this.

For their project, students will write a detailed but brief report (see below of guidelines). There are no set or required subsections of the report, because every data set and data science project is different. However, keep in mind the usual steps of a data science project as visualized in the textbook:



A data science report often consists of first describing the goal of the data science project. For this project, this does not need to be a confirmatory hypothesis or anything like that. It can be exploratory. Then, a brief description of the data set that will be used for the project. After that, it is up to you to walk the reader (i.e., me) through the various steps you needed to complete to accomplish this goal. For instance, you can describe any issues you had to overcome with importing and tidying your data, necessary data cleaning or transformations that needed to be done, assumptions testing that were performed, any statistical modeling that was carried out, important data visualizations that were created, and any conclusions that were drawn from the analyses.

My hope is that students will use this assignment as an opportunity to craft code and a project report that they can store in their research portfolio. When applying to data science positions, having example projects like these can help demonstrate your proficiency.

Below, I have listed guidelines and policies to clarify my expectations about this assignment. However, if anything about the project assignment is not clear, please do not hesitate to reach out to me.

Deadline

The project report is due by **May 12th at 11:59 PM**. Please submit your project on Canvas. Your final submission will include:

(1) The written report (as a Word document)

(2) Your final code (e.g., an R script file)
(3) Your data

For the data, it can simply be the necessary data or variables that are needed to produce the output you describe in your written report. Make sure all data files are de-identified if relevant. Also, if the data is coming from your lab, please confirm with your advisor that it is okay to share the de-identified data with me. Please consult with me as soon as possible if this is an issue.

Project Length

The maximum length of your project report is 5 pages (single spaced). This length includes a maximum of 2 figures or tables. You are free to include additional figures, but please include them in an appendix at the end of the document (these appended figures would not contribute to your page total).

For your submitted code, there are no set requirements for the number of lines of code. However, it is hard to imagine that the entire data science project could be completed under 100 lines of code. Though that may even seem a lot to you, think back to the homework assignments. Many of those have ended up with a few dozen lines of code when there were just a few targeted problems.

Grading Rubric

You will be graded according to the following rubric. Please see below for details about each component.

Component	Excellent	Very Good	Good	Satisfactory	Unsatisfactory
Scope of the Project (20%)	19-20	17-18	15-16	13-14	≤ 12
Effectiveness of Data Science Methods (40%)	37-40	33-36	29-32	25-28	≤ 24
Accuracy and Clarity of Reporting (20%)	19-20	17-18	15-16	13-14	≤ 12
General Writing & Formatting (20%)	19-20	17-18	15-16	13-14	≤ 12

Note. Half-point grades (e.g., 19.5) will be allotted as appropriate.

Overview of Each Component

Scope of the Project

By scope of the project, I simply mean the amount of effort and/or complexity that would be expected to go into completing the project. For instance, if your project simply consists of you importing a data set and renaming some of the columns, even if you do everything correct it will not score you a lot of points for this component. However, I am also aware that this is a class project, so I am not expecting a data science project that would be expected for a comprehensive thesis or anything like that. If you have any questions or concerns about the scope of your planned project, please feel free to consult with me beforehand.

Effectiveness of Data Science Methods

This component covers issues such as: Were appropriate data science methods or techniques utilized to accomplish the various goals/tasks of the project? Were these methods and techniques implemented correctly in the code? Were errors made in the code? If so, what were the impact of these errors?

Accuracy and Clarity of Reporting

This component covers issues such as: Were the conclusions of the data science methods and results appropriately and accurately reported in the document? Were there errors in how any output or outcomes were interpreted? If so, what was the scope of these errors in reporting?

Writing and Formatting

Your paper should be clearly written and well organized, with evidence of careful proofreading (i.e., an absence of typographical and grammatical errors). In addition, it should closely adhere to APA formatting whenever relevant.

Any questions? Please email bixterm@montclair.edu.