

DSA 610, Project #1, Spring 2025

Instructions: Use the **BW_MBA_data.xlsx** dataset for this project. The guidelines for this project are quite broad. Explore the data. Create some summary tables. Discuss the statistics of any variables of interest. Make some graphs. You may use Excel or Python for the data exploration, but all final analysis should be completed in Python (though, you may do some data cleaning in Excel before importing it into Python). Create a minimum of five graphs and related tables.

Write up your analysis. The written analysis should include (not necessarily in this order):

1. All graphs and tables, and explanations of each.
2. An explanation of the process. Include some sample Python code snippets, but these should be illustrative, not the focus of the report. (If you want to refer to more than examples, use an appendix for more extensive code and related explanations.)
3. Any supplemental research done to support your analysis (and citations/references) in a proper format (though the style is of your choice). For example, to validate the data is reasonable, etc.
4. Equations should be formatted in an equation editor or in LaTeX (if needed).
5. A research question and a conclusion: what is your goal in analyzing this data set? What are you interested in learning?
6. Discuss how the data and analysis connect to data lifecycles discussed in class. Are there any ethical considerations with collecting or using this data? How was this data likely collected?

The written portion should be a minimum of five pages, including graphs. You should have appropriate headers. You can create this in Jupyter notebook, if you like, or work in a word processing tool. The report should be able to be read on its own. Everything you want me to read should be in the body of the report. Do not send me to file attachments.

Submit all supporting exploratory analysis, including your Excel file (if used) and Jupyter notebook (saved as a pdf).

The expectation is that you can complete this analysis project using primarily what we have done in class so far. You may, however, go beyond what we have done in class according to your interests and experience. You may complete this project entirely in Python. The use of Excel or other spreadsheets is optional.