Case Study 2

This is a group project so it’s the responsibility of the group members to collaborate accordingly. Team work is important.

The due date for submission is 1 hour before live session 15. During live session 15, the group will have up to 20 minutes to present their case study report. The goal is to communicate the findings of the project in a clear, concise and scientific manner. Tell me the story.

Create a GitHub repository with an RMarkdown file containing an executive summary, introduction to the project, all supporting code and analysis, and the group presentation.

Submit a link to the GitHub repository via the space provided for the Case Study 02 page in 2DS.

Since you are getting your own data. I would like the following

1. Research Question – Clearly stated
2. Hypothesis
3. A Code book
4. READme file
5. Exploratory Data
   1. Please provide (in table format or similar), descriptive statistics on at least 7 variables (age, Income, etc.). Create a simple histogram for *two* of them. Comment on the shape of the distribution in your markdown.
   2. Give the frequencies (in table format or similar) for Gender, Education, and Occupation. They can be separate tables, if that’s your choice
6. One Linear regressions
7. One model of your choice (can be linear regression, thus 2 linear regressions total)
8. Results
9. Discussion
10. Conclusion

Your Presentation should be in this outline

* Title
  + Authors all listed
* Presentation Outline
  + Business Objectives
  + Data Sourced
  + Methodology
  + Evaluation/Results
  + Summary
* Business Objectives
* Data Source
  + Where you got it
  + Basic Statistics (EDA)
* Methodology
  + Steps
  + Workflow
* Evaluation/Results
  + Tell me the percentages and why
  + Show me graphs with explanations
  + Was your hypothesis correct
  + Provide any other interesting trends and observations from your analysis
  + Other things to consider?
* Summary
  + Insights
  + Recommendations
  + Improvements
* Questions

Have your presentation ready to present on during live session 15.

NO CODE should be present in your presentation. (ie variable names is Mon\_Age this should in your presentation as age). You code should be in your GitHub and RMarkdown File.

Remember, no late assignments will be accepted, as this is the final. It has been a pleasure teaching you this semester! Congratulations on making it through your first semester in the Masters of Data Science program. Good luck!