DSA 8030 – R Project Details

In an effort to provide you with the concepts of Importing, Tidying, Transforming, Visualizing, Model, and Communicating in a Data Science project we are going to do a Data Science Project of your choosing.

Why is this useful for my Data Analyst career?

To give you a little usefulness to this assignment, it may be helpful to notice for this assignment we are covering **2 of the** *"4 R project to form a core data analyst portfolio"* found in this article here: https://www.r-bloggers.com/2020/11/4-r-projects-to-form-a-core-data-analyst-portfolio/Links to an external site..

How does this project apply to this course?

The purpose of this R project is for you to put together all the tools that you have learned in this course about R (or most of them). The project will allow you to explore your skills with real data and synthesize information used from this class and other courses. The goal is for you to produce a professional looking report. One that you can be proud of and use to future employers as an example of the work that you do.

This project should encompass most all of the learning outcomes for the course as they pertain to R:

- Import data into R
- Clean, transform, and subset data in R
- Perform descriptive and inferential statistics with R
- Produce informative summary tables and graphs using R
- Create reports using R Markdown

Guidelines:

Report

- Create an R Markdown file for your report (you will submit an html file)
- Introduction –
    - Provide a description of your topic of interest
    - Provide a short explanation of how you plan to address the question of interest.
    - Include a "cover image" that is relevant to your topic (be sure to cite the source)
- Dataset –
    - Provide a clear reference and description of the data
    - Thoroughly explain what the data is (what years, what are the observations, what subset of people, etc.)
    - Provide explanation for relevant variables (what is the variable, what does it represent (units?), any content knowledge about variables (remember audience)
- Analysis –
    - Import Dataset
    - Tidy Dataset – Provide explanation of how data was made "Tidy"
    - Transform – Provide a full explanation of how you transformed the data

- o Visualize – Provide appropriate exploratory data analysis for your variables (graphs, frequency tables, descriptive statistics, etc.)
    - ▪ Properly label any graphs
  - o Model – Explore your questions of interest – uncover something about your data that is apparent from the basic information, this can be done with multivariate graphs, statistical tests, regression models
- Communicate –
  - o Provide a summary of your findings
  - o Discuss any limitations of your analysis. How could it be improved in the future? What are the next steps in analysis (ie., additional research should be done in *xyz*)?
- Flexdashboard –
  - o Provide a link to your flexdashboard (more information below)
- R code
  - o Provide your code either as an appendix or throughout your report (using hidden coding)
  - o Somewhere in your report (can be appendix or within your report) discuss the packages you used in the analysis and a general description of what the package was used for.
- Biography
  - o Provide an appropriately sized picture of you
  - o Provide a short biography (include why the subject matter for your project is of interest to you)

Things to keep in mind for Report:
- Report should be professional looking – something you can hand to a potential employer
- Your audience is general public (future employers)
- You should include color, titles, styling, be grammatically correct, do not use conversational language, etc.
- Make the report your own – go above and beyond what is minimally required (be creative) (extraordinary effort, tools not addressed in class, application of tools, creativeness, etc.)
- Use reference where appropriate (the style of reference is your choice)

Presentation

- Provide a 5-7 minute video explaining the highlights of your project.
- You may use whatever means you would like to record – I will provide some options.

Shiny Dashboard

- Create a flexdashboard for your user to explore your dataframe.
- There should be multiple ways to explore the data

Things to Consider

- Choose a topic of interest to you.
- Decide what story you want to tell.
- Your final report should explore/investigate your data and logically explain this investigation such that investigators could reproduce your work.
- Make it meaningful – your project should provide insight. It doesn't have to be globally impactful but it should be interesting.
- Be creative!