Case Study Rubric

DS 4002 - Fall 2024 - Maansi Taori
Due: TBD
Submission Format: Upload link to GitHub Repository on UVA Canvas

Individual Assignment

Why am I doing this?
The case study allows you to leverage your accumulated technical and conceptual data science skills in a comprehensive project. By working through this project, you will be exposed to different methodologies of modeling analyses on a hands-on scenario that can be applied to real-world contexts with potential implications in the fields of cancer research and health policy.

What am I going to do?
The GitHub repository can be found at https://github.com/tpq4ew/DS4002---CS3. You will obtain selected data from the NCI that includes pertinent information including the project abstracts. After obtaining this CSV file, you will use Python to ultimately determine whether funding inequality exists across cancer research grants, as proven through the use of statistical hypothesis testing. You will ultimately provide a deliverable that covers all requirements, including significant results and conclusions. The final deliverable will include:

- Written portion - includes Data Dictionary and reference page
- Well documented, commented source code
- GitHub Repository - containing all materials used

How will I know I have succeeded?
You will meet expectations on the Case Study when you follow the criteria in the rubric below.

Spec Category	Spec Details
Formatting	Submit each component listed in the rest of this rubric in one GitHub Repository that is submitted on Canvas: ● Create a new GitHub repository for this assignment called ‘Case Study 3 - [Last Name, First Name]’ that contains ○ README.md

	○ LICENSE.md ○ Source Code File ○ Dataset ○ REFERENCES.md
Written Portion of README.md	● Brief goal statement and summary of what was produced from the resulting case study. Include analysis plan and discuss significance of results. 1-2 paragraphs ● Allow for people who haven't seen this to understand the project at a glance
Source Code	● Well documented Jupyter notebook with comments so viewer can understand the thought process and flow of work ● Script must contain the code to execute the keyword and statistical analyses. It must include: ○ TF-IDF model ○ LDA model ○ ANOVA statistical testing ○ Any initial exploratory plots
References	● Please include all references in IEEE format at the bottom of the README.md. Additionally, include all articles and supporting materials in REFERENCES.md