Goal: Explain the purpose, task, and criteria. This is where the details live to answer questions for the student.

● Use this document's format. There is no need to reinvent the wheel.

● The goal is about 2 pages.

● A rubric in this style and describing the deliverable you want the 2nd year to produce (about 2 pages in pdf format)

COI Case Study Rubric

DS 4002 – Spring 2025 – Elle Park
4/29

Individual Assignment

Why am I doing this?
This case study allows you to leverage your data science knowledge by using time series data to accurately predict child opportunity scores of Richmond, VA counties from the child opportunity index (COI) collected from diversitydatakids foundation's 2012-2021 census. As you work through this assignment, you will be exposed to the ways that data analysis can be used in a real-world context with potential implications for public policy and adolescent development research. Additionally, this case allows you to reflect and reproduce the work of a past student.

- Course Learning Objectives: recreate results from previous students and appy learned skills.

What am I going to do?
The GitHub repository for this case study can be found at https://github.com/ElliotRoosevelt/CS3-DS4002/tree/main. With your accumulated technical and conceptual skills in data science, you will be able to successfully go through the life cycle of data collection, exploratory data analysis, predictive modeling, and multilinear regression analysis. Download the necessary datasets from diversitydatakids.org as listed in the GitHub repository. Merge the two datasets. Then, perform EDA like boxplots and pivot tables. Code for correlation heat maps and cross-tabulation. Then, train your model with the census information from 2012 to 2020 to predict 2021 COI scores with multilinear regression. Lastly, conclude how well your model did and interpret what you found.

Your final deliverables should include:
- Multilinear regression model trained to predict 2021 COI scores.
- Case Study Review and Reflection
- Case Study Results Reproduction

Tips for success:
- You will be working in both Python and R. Familiarize yourself with both languages to the best of your ability before starting this case study in order to complete it more efficiently
- Ask your TA and professor about any questions. They are paid to help you.

- Talk to your classmates and compare case studies.

How will I know I have succeeded?
You will meet expectations on this case study when you successfully follow and complete the criteria in the rubric below:

Spec Category	Spec Details
Formatting	● Github repository ○ Create new GitHub repo that contains ■ README.md ■ CASE STUDY folder ■ REFERENCES folder ○ Pdf format if possible, code/data in appropriate format
README.md	● Orient the user to the repo and link to every folder and its content
CASE STUDY folder	● Case study review ● Case study reproducibility ○ Write a project overview summarizing this project's motivation/ context ○ List the steps needed to replicated and follow the output of results ○ Highlight challenges encountered throughout the process. How did you fix or solve them? ● Case study reflection ○ Provide a positive and why you would include it in future projects. ○ Provide constructive feedback on one element of the case study and why you would not include it in future projects. Highlight what you would do instead. And deliver it in a way

	that someone would be open and feel supported to improve next time.
REFERENCES folder	● Write out any references used

Acknowledgements: Thank you to Professor Alonzi for providing the rubric structure!