NYC DATA SCIENCE ACADEMY

Web Scraping Project Rubric

Use this rubric as a guide to completing a successful web scraping project. These are the criteria the instructors will use to evaluate your project.

	Excellent (A) **5 points** Exceeds expectations	**Good (B)** **4 points** Meets expectations	**Fair (C)** **3 points** Meets lowest acceptable standards	**Poor (F)** **1 point** Doesn't meet acceptable standards
Project Submission Criterion related to timely submission and presentation of the project. **Weight: 25%**	Project presented on time, presentation shared correctly, code is hosted on GitHub repository, and there is strong evidence of version control.	Project presented on time, presentation shared correctly, code is hosted on GitHub repository, but there is no evidence of version control.	Project presented on time, but presentation not shared, and code not hosted on GitHub repository.	Project not presented on time, no presentation shared, and code not hosted on GitHub repository.
Purpose Criterion related to the purpose of the project. **Weight: 15%**	Objective of the problem statement is clearly defined and demonstrates a value proposition. The data collected is relevant to answering the question asked.	Objective of the problem statement is defined and loosely demonstrates a value proposition. The data collected is relevant to answering the question asked.	Objective of the problem statement is vaguely defined and does not demonstrate a value proposition. The data collected is somewhat relevant to answering the question asked.	No objective of the problem statement is defined. The data collected is not relevant to the project goals.
Tools and Methods Criterion related to tools and methods used in the project. **Weight: 15%**	Demonstrates correct usage of tools and methods learned in the coursework and selects those appropriate to solve tasks at hand. Data is collected using a reasonable approach with one or more of the technologies learned in the coursework.	Demonstrates correct usage of tools and methods learned in the coursework and selects those which are somewhat appropriate to solve tasks at hand. Data is collected using a reasonable approach with one or more of the technologies learned in the coursework.	Usage of tools and methods learned in the coursework is often incorrect or misunderstood. Data is downloaded through an API or from a server rather than through the use of web scraping tools learned in the coursework.	Very limited to no usage of tools and methods learned in the coursework and usage is mostly incorrect or misunderstood. Data is downloaded through an API or from a server rather than through the use of web scraping tools learned in the coursework.

	Excellent (A) **5 points** Exceeds expectations	**Good (B)** **4 points** Meets expectations	**Fair (C)** **3 points** Meets lowest acceptable standards	**Poor (F)** **1 point** Doesn't meet acceptable standards
Outcome Criterion related to the outcome of the project. **Weight: 15%**	Project achieves stated objectives and evidence-based actionable steps for future recommendations are given.	Project somewhat achieves stated objectives and limited actionable steps for future recommendations are given.	Project fails to achieve stated objectives, but some actionable steps for future recommendations are given.	Project fails to achieve stated objectives and no actionable steps for future recommendations are given.
Presentation Criterion related to the presentation of the project. **Weight: 30%**	The student effectively presents their project goals, analyses, and outcomes in a logical and easy-to-follow way. Visualizations used effectively convey the intended messages. Presentation clearly indicates that time was dedicated for development and practice of the presentation.	The student presents their project goals, analyses, and outcomes in a logical way. Visualizations used mostly convey the intended messages. Presentation indicates that some time was dedicated for development and practice of the presentation.	The student presents their project goals, analyses, and outcomes in a way that is difficult to follow. Visualizations used have difficulty conveying the intended messages. Presentation indicates that minimal time was dedicated for development and practice of the presentation.	The student presents their project goals, analyses, and outcomes in a way that is hard to follow. Few to no visualizations used. Presentation indicates that little to no time was dedicated for development and practice of the presentation.

Note: This criterion may be used at the instructor's discretion to alter the overall grade of the project. These categories are designed to help guide students to follow common best practices and produce high quality code.

	Excellent (A) **5 points** Exceeds expectations	**Good (B)** **4 points** Meets expectations	**Fair (C)** **3 points** Meets lowest acceptable standards	**Poor (F)** **1 point** Doesn't meet acceptable standards
Code (For Discretionary Consideration) Criterion related to the usage of proper coding techniques and practices.	Code is organized into scripts, each with a clear purpose, and are given meaningful file names. Sufficient commenting/docstring is used to describe the functionality of the code. In Jupyter Notebooks, markdown is used to partition the code into logical sections. Code is modular, robust, efficient and demonstrates an understanding of best practices (such as using helper functions). Code has no syntax errors and follows the standard formatting style.	Code is organized into scripts, each with a vague purpose. Some commenting is used to describe the functionality of the code. In Jupyter Notebooks, markdown is used occasionally to partition the code into sections. Code is somewhat robust and efficient but has room for improvement. Code may have minor syntax errors and mostly follows standard formatting style.	Code is not organized into scripts. Minimal commenting is used to describe the functionality of the code. In Jupyter Notebooks, markdown is not used to partition the code into sections. Code is repetitive and inefficient. Code has several syntax errors and often ignores standard formatting style.	Code does not run successfully, and no commenting is used to describe the functionality of the code. Code is riddled with syntax errors and completely ignores standard formatting style.
Data Cleaning/ Use of Regex (For Discretionary Consideration) Criterion related to the usage of proper data cleaning and regex techniques and practices.	Data is cleaned and stored in a structured way that is logical and easily interpretable. Appropriate use of functions for data cleaning. Appropriate use of regex.	Data is cleaned and stored in a somewhat structured way, however the logic may not be clear. Functions are sometimes defined for data cleaning. Mostly appropriate use of regex.	Data is not fully cleaned and is stored in a somewhat structured way, however the logic is unclear. Functions are not defined for data cleaning. Regex used incorrectly.	Data is not cleaned or there are significant issues with the data that can be resolved which prevent the use of analysis techniques. No use of regex when there is a clear need for it.