

# DSCI 510: Comprehensive Rubric and Guidelines for DSCI 510 Final Project Evaluation

This document specifies the rubric for the final project. It is divided into 2 main sections: **Final Report** and **Code Base**.

## Final Report (40 points)

The final report is worth 40 points, it should include the following details:

### 1. Name, Team Members (2 points)

- Include the name of the project.
- List team members, providing their USC emails and IDs.

### 2. Short Description (4 points)

- Provide a concise overview of your project, capturing the main goals and purpose.

### 3. Data (8 points)

- **List all the sources (4):** Clearly specify the sources of your data, whether scraped, retrieved via APIs or collected using any other way.
- **Specify the number of Data Samples (4):** State how many data samples were collected.

### 4. Data Cleaning, Analysis & Visualization (18 points)

This section defines the rubric for the data cleaning, analysis and visualization process. Your report should clearly describe the steps you took for each process. If your project skips a step, please mention that as well.

- **Well-Defined Process (12):** Detail each step of the data cleaning, analysis, and visualization process. The explanation should provide clarity and a logical flow.
- **Hypothesis/Premise & Conclusions (6):** State the initial hypothesis or premise of your project. Discuss the conclusions drawn from the analysis and visualizations. Ensure that your report communicates the findings effectively through a compelling narrative.

## 5. Changes from Original Proposal (4 points)

- This section should explain what changes you had to make from your original proposal and why. You should mention the challenges faced and how they were addressed.
- If your project has no changes from the original proposal, clearly state this in that case.

## 6. Mention of Future Work (2 points)

- Discuss potential directions for extending or improving the project if more time or resources were available.

## 7. File Name, Format, and Page Limit (2 points)

- The final report should be called **final\_report.pdf**.
- The final report should be a pdf file.
- Ensure the final report is between 2 to 5 pages.

# Code Base (60 points)

## 1. README (6 points)

- The README should include your project name and team member details.
- The README should include detailed instructions for:
  - Creating a virtual environment.
  - Installing the required libraries.
  - Running the scripts for data collection, cleaning, analysis, and visualization.

## 2. Requirements File (4 points)

- Ensure the `requirements.txt` file lists all external libraries used.
- The file should enable a user to install all dependencies with `pip install -r requirements.txt`.

## 3. Code Compatibility (10 points)

- Follow coding standards for readability and maintainability.
- Ensure consistent formatting, proper indentation, and adherence to naming conventions.

## 4. Function Documentation (10 points)

- Include **docstrings** for all functions.
- Use **type hints** to clarify input and output types.

## 5. Code Execution (16 points)

- All and any code you use, including the jupyter notebooks should be included with your submission.
- Verify that all code runs without errors.
- Ensure all dependencies are correctly specified and compatible with your scripts.

## 6. Jupyter Notebooks Usage (4 points)

- Use Jupyter notebooks **only** for creating and presenting visualizations.
- Core processing code should reside in `.py` files and be imported into notebooks as necessary.
- Your jupyter notebooks may include more visualizations than your final project.
- If you do not use jupyter notebooks at all, make sure all visualizations you use in the final project are replicable by running the `.py` files.

## 7. Data Upload (10 points)

- Include both raw and processed data in appropriate subfolders under the `data/` directory.
  - `data/raw/` for the original data collected.
  - `data/processed/` for the cleaned and structured data ready for analysis and visualization.