

DSCI 511 – Project Proposal

Roy Phelps

Shad Scarboro

Leland Weeks

Evan Wessel

October 5, 2025

1 Team Formation and Roles

Our team consists of four members with complementary strengths. Below are our self-identified skills and anticipated contributions:

Roy Phelps

Skills: Python, Jupyter Notebooks, basic data cleaning, visualization

Contribution: Code development, documentation, Git/GitHub organization

Shad Scarboro

Skills: Data sourcing, research, presentation formatting

Contribution: Lead on data acquisition planning and writeup support

Leland Weeks

Skills: Python scripting, exploratory analysis, math/stats foundations

Contribution: Support on initial dataset exploration and preprocessing

Evan Wessel

Skills: Writing, data summaries, editing

Contribution: Drafting sections of the proposal and summarizing findings

2 Data Source of Interest (To Be Finalized)

We are in the process of selecting a dataset topic. The dataset will be chosen based on the following criteria:

- Real-world relevance and potential impact
- Feasibility of acquisition within the course timeline
- Opportunities for exploratory analysis and coding
- Clear connection to potential users or applications

Potential areas we are considering:

- Health & wellness data
- Environmental or climate data
- Sports statistics
- Social media or public sentiment
- Education and academic performance
- Transportation or mobility data

Once we finalize a topic, we will specify:

- The exact source (API, website, files, etc.)
- The expected format (CSV, JSON, text, etc.)
- The licensing or access constraints, if any

3 Potential Users and Applications

We intend to collect a dataset that can be useful beyond this class. Once our topic is chosen, we will identify:

- Who might use this data (e.g., researchers, policy analysts, educators, businesses, developers)
- Example use cases (e.g., modeling, dashboards, comparison studies, predictive insights)
- Academic or commercial relevance

4 Plan for Acquiring the Data

Our plan will depend on the chosen topic, but we anticipate one of the following acquisition methods:

- Downloading CSV/JSON files from open data portals or repositories
- Using a public API and writing Python scripts to fetch the data
- Web scraping (if allowed and ethical)
- Combining multiple smaller datasets into a single structured source

After finalizing the dataset, we will document:

- Tools we plan to use (e.g., Python, BeautifulSoup, requests, pandas)
- Steps to handle cleaning and preprocessing
- Any potential obstacles (e.g., limits, formatting issues, permissions)

5 Sample of the Dataset (Placeholder)

A preliminary data sample will be included once our source is selected. This may include:

- A few sample rows
- A screenshot, snippet, or small extracted file
- A short block of test code if we access via API or scraping

Example placeholder until the dataset is confirmed:

[Sample data will be inserted here once the source is chosen.]
Possible formats: CSV snippet, JSON object, or scraped output

6 Submission Plan

We will submit the proposal as a Markdown or PDF document (or via Jupyter Notebook if preferred) including:

- All required sections
- Team member names and roles
- Dataset description and acquisition plan

- Sample or preview of the dataset

Once our topic is locked in, we will revise this template and finalize the language.