**Requirements**

**Software Version Control (10 points)**

* Repository is created on GitHub (2 points).
* Files are frequently committed to the repository (3 points).
* Commit messages include an appropriate level of detail (2 points).
* Repository is organized and includes relevant information and project files (3 points).

**Documentation (10 points)**

* Code is well commented with concise, relevant notes (3 points).
* GitHub README file includes a concise project overview (2 points).
* GitHub README file includes detailed usage and installation instructions (2 points).
* GitHub README includes either examples of the application, or the results and a summary of the analysis (3 points).

**Analysis and Conclusion (30 points)**

* Findings are strongly supported with numbers and visualizations (10 points).
* Write-up summarizes major findings and implications at a professional level (10 points).
* Each question in the project proposal is answered with precise descriptions and findings (5 points).
* Each question response is supported with a well-discerned statistical analysis from lessons, such as aggregation, correlation, comparison, summary statistics, sentiment analysis, and time series analysis (5 points).

**Visualizations (20 points)**

* 6–8 visualizations of data (at least two per question) (10 points).
* Clear and accurate labeling of images (5 points).
* Visualizations supported with ample and precise explanation (5 points).

**Presentation Requirements (30 points)**

Your presentation should cover the following:

* An executive summary or overview of the project and project goals (5 points).
* An overview of the data collection, cleanup, and exploration processes (5 points).
* The approach that your group took in achieving the project goals (5 points).
* Any additional questions that surfaced, what your group might research next if more time was available, or share a plan for future development (5 points).
* The results and conclusions of the application or analysis (5 points).
* Slides effectively demonstrate the project (3 points).
* Slides are visually clean and professional (2 points).

This project will be evaluated against the requirements and assigned a grade according to the following table:

| **Grade** | **Points** |
| --- | --- |
| A (+/-) | 90+ |
| B (+/-) | 80–89 |
| C (+/-) | 70–79 |
| D (+/-) | 60–69 |
| F (+/-) | < 60 |

**Project Guidelines**

The following project guidelines focus on teamwork, your project proposal, data sources, and data cleanup and analysis.

**Collaborating with Your Team**

Remember that these projects are a group effort. The experience of close collaboration will create better project outcomes and help you in your future careers. Specifically, you’ll learn collaborative workflows that will enable you to approach and solve complex problems. Working in groups allows you to work smart and dream big. Take advantage!

**Project Proposal**

Before you start writing any code, your group should outline the scope and purpose of your project. This will help provide direction and safeguard against **scope creep** (the tendency for projects to become more complex after work begins).

The proposal is essentially a brief summary of your interests and intent. Be sure to include the following details:

* The kind of data you’d like to work with and the field you’re interested in (finance, healthcare surveys, etc.);
* The questions you’ll ask of the data; and
* Possible source for the data.

Use the following example for guidance:

The aim of our project is to uncover patterns in credit card fraud. We’ll examine relationships between transaction types and location, purchase prices and times of day, purchase trends over the course of a year, and other related relationships derived from the data.

**Finding Data**

Once your group has written a proposal, it’s time to start searching for data. We recommend the following curated sources of high-quality data:

* [data.worldLinks to an external site.](https://www.data.world/)
* [KaggleLinks to an external site.](https://www.kaggle.com/)
* [Data.govLinks to an external site.](https://www.data.gov/)
* [Awesome Public DatasetsLinks to an external site.](https://github.com/awesomedata/awesome-public-datasets)

**IMPORTANT**

Whenever you use a dataset or create a new dataset based on other sources (such as existing datasets or information scraped from websites), make sure to use the following guidelines:

1. Check for copyright protections, and make sure that the way you plan to use this dataset is within the bounds of fair use.
2. Document how you intend to use this dataset now and in the future. Find any licenses or terms of use associated with the dataset, and review them to confirm that your intended use is in compliance.
3. Investigate how the dataset was collected. Identify any indicators that the data was obtained from a source that the compilers were not authorized to access.

You’ll likely have to adjust your project plan as you explore the available data. That’s okay! This is all part of the process. Just make sure that everyone in the group is aligned on the project’s goals as you make changes.

Make sure that your datasets are not too large for your personal computer. Big datasets are difficult to manage locally, so consider using data subsets or different datasets altogether.

**Data Cleanup and Analysis**

Now that you’ve picked your data, it’s time to tackle development and analysis. This is where the fun starts!

The analysis process can be broken into two broad phases: (1) exploration and cleanup, and (2) analysis.

As you’ve learned, you’ll need to explore, clean, and reformat your data before you can begin answering your research questions. We recommend keeping track of these exploration and cleanup steps in a dedicated Jupyter notebook to stay organized and make it easier to present your work later.

After you’ve cleaned your data and are ready to start crunching numbers, you should track your work in a Jupyter notebook dedicated specifically to analysis. We recommend focusing your analysis on multiple techniques, such as aggregation, correlation, comparison, summary statistics, sentiment analysis, and time-series analysis. Don’t forget to include plots during both the exploration and analysis phases. Creating plots along the way can reveal insights and interesting trends in the data that you might not notice if you wait until you’re preparing for your presentation. Presentation requirements will be further explained in the next module.

**Presentation Guidelines**

This section lists the Project 1 presentation guidelines. Each group will prepare a formal, 10-minute presentation that covers the following points.

* An executive summary or overview of the project and project goals:
  + Explain how the project relates to the industry you selected.
* An overview of the data collection, cleanup, and exploration processes:
  + Describe the source of your data and why you chose it for your project.
  + Describe the collection, exploration, and cleanup process.
* The approach that your group took to achieve the project goals:
  + Include any relevant code or demonstrations of the application or analysis.
  + Discuss any unanticipated insights or problems that arose and how you resolved them.
* The results/conclusions of the application or analysis:
  + Include relevant images or examples to support your work.
  + If the project goal was not achieved, discuss the issues and how you attempted to resolve them.
* Next steps:
  + Briefly discuss potential next steps for the project.

It’s crucial that you find time to rehearse before presentation day.

On the day of your presentation, each member of your group is required to submit the URL of your GitHub repository for grading.

**Presentation Day**

Your group will have a total of 10 minutes—7 minutes for the presentation followed by a 3-minute question-and-answer session. It’s crucial that you find time to rehearse before presentation day.

On the day of your presentation, each member of your group is required to submit the URL of your GitHub repository for grading.

**NOTE**

Projects are requirements for graduation. While you are allowed to miss up to two Challenge assignments and still earn your certificate, projects cannot be skipped.

[Previous](https://bootcampspot.instructure.com/courses/4941/modules/items/1117446)[Next](https://bootcampspot.instructure.com/courses/4941/modules/items/1117450)