

# DSCI 510: Principles of Programming for Data Science

Fall 2022

Homework 2

Plan for Final Project

**Due: 11/18/2022 23:59 PM PT**

## Task:

- Please submit a PDF file answering the following 3 questions.
  - Please be concise and informative.
  - Use of bullet points is encouraged.
  - No more than 1 page.
1. What's the name of your final project? Please describe it as a research question and provide a short description [5 points]
  2. What data sources are available? Could you find multiple data sources? How are you going to collect them? How many data samples are you going to collect? [5 points]
  3. What kind of analyses or visualizations do you want to do? [5 points]

## Rubric:

1. The project name should:
  - a. Reads like a research project. 1 points.
  - b. Have a short description. 2 points.
  - c. Indicates the data source. 1 point.
  - d. Indicates the analysis method. 1 point.

what kind of analysis method do we have?

2. The data sources should:

- a. Specify the general nature of data, e.g. weather, hotel price, airfare, etc. 1 point.
- b. Specify exact data sources, e.g. a link to the dataset, a link to an API, or a link to a website to parse. 1 point. [If we use Kaggle, how many points will be deducted?](#)
- c. Have 2 or more data sources. 1 point. [How to build connection between two data sources?](#)
- d. Includes a short description and your plan. 1 point. [Difference between analysis and plan](#)
- e. The data source should be large and recent. It should contain at least 100 samples (recommends 1000 samples or more), and is no older than 10 years from now. 1 point.

3. The analyses and visualizations should:

- a. Feasible. It should be an analysis that could be done with your data source. 2 points.
- b. Meaningful. It should be something that is not obvious, non-trivial, not in common sense. 2 points.
- c. Impactful. It should be something having an impact and is interesting to the audience. 1 point.