# Team Project

## Overview

Throughout this project you will be practicing the life cycle of analyzing data. The goal is to use data to answer questions of interest and importance.

## Requirements

The project should start with a business question, a question of societal or personal importance, a scientific question, etc. The goal is to enforce the fact that the question comes first and then we look at ways to answer it.

Assuming the question can be answered by data, you have to collect **data from multiple sources**. So asking a question you can answer by collecting and analyzing data in 7 weeks is important (i.e., choosing the right scope). There will be opportunities to update your scope but you will only be able to do it once.

You need to be able to defend that the data you have collected can answer your question(s) and your model and its assumptions are suitable for the type of analysis you want to do.

## Potential Topics

You can pick any topic you want based on your interests. Here are some examples:
- any type of firm performance analysis using cross sectional data.
- social networks (measuring the impact of influencers on sales)
- any type of A/B experiment

### Topic 1

The impact of local and national policies on controlling the effect of the pandemic (caused by Covid). Do vaccines work? Do other interventions such as mask mandates, crowd control, etc. work?

There is quite a bit of data about the pandemic but understanding what you can and cannot answer by the data available is important.

### Topic 2

Conduct your own campus survey to answer a question of interest at UF.

- You can study factors that impact learning outcomes. For example, online teaching vs. in-class teaching, working, amount of homework, time spent studying, etc. Note that some of these factors are controlled by students, some are imposed by instructors or university rules.

- You can look at predictive models of learning outcomes. For example, how well the admission criteria predict future student success.

### Topic 3

What factors affect chief executive officer salaries? The files CEOSAL1.DTA and CEOSAL2.DTA are data sets that have various firm performance measures, as well as information such as tenure and education. Though a good starting point, you need update these data files and look for other interesting factors. Rose and Shepard (1997) considered firm diversification as one important determinant of CEO compensation. (You can download the data sets from Canvas).

- Rose, N. L., and A. Shepard (1997), "Firm Diversification and CEO Compensation: Managerial Ability or Executive Entrenchment?" Rand Journal of Economics, 28, 489–514.

## Topic 4

General public policy questions (including healthcare).

Do changes in traffic laws affect traffic fatalities? McCarthy (1994) contains an analysis of monthly time series data for the state of California. A set of dummy variables can be used to indicate the months in which certain laws were in effect. The file TRAFFIC2.DTA contains the data used by McCarthy. An alternative is to obtain a panel data set on states in the United States, where you can exploit variation in laws across states, as well as across time. (See the file TRAFFIC1.DTA.)

- Mullahy and Sindelar (1994) used individual-level data matched with state laws and taxes on alcohol to estimate the effects of laws and taxes on the probability of driving drunk (You can download the data sets from Canvas).
- McCarthy, P. S. (1994), "Relaxed Speed Limits and Highway Safety: New Evidence From California," Economics Letters, 46, 173–179.
- Mullahy, J., and J. L. Sindelar (1994), "Do Drinkers Know When to Say When? An Empirical Analysis of Drunk Driving," Economic Inquiry, 32, 383–394.

Did Affordable Care Act (ACT) lead to better outcomes? You need to refine this question a bit before digging in. There is quite a bit of public data about healthcare before and after ACT.

## Deliverables and Deadlines

The summary of the deliverables and their deadlines are provided in the table below (Table 1). Detailed descriptions follow.

| Deadline | Deliverable |
|----------|-------------|
| Mar-4 | #0: Member Names |
| March-25 | #1: One-page proposal |
| Apr-25 | #2: Final Paper (due after your presentation) |
| Apr-19 | **#3: Presentation on zoom** |

*Table 1- Deadlines*

**Please ensure that the first page of each submission is a cover sheet** (team name, members, and project title) and that all submissions are presented in a professional style. All written text must be **11pt** font and **1.5 spaced** (this does not apply to figures). **Handwritten or hand-drawn material is not acceptable.**

## Deliverable #0: Teams                                           0 points

For the first deliverable, you only need to form your teams on Canvas. Maximum of 6 students per team.

## Deliverable #1: One Page Project Proposal          60 points

For this deliverable, you will submit your research idea on <u>one page</u>. The importance of posing a very specific research question cannot be overstated. Without being explicit about the goal of your analysis, you cannot know where to even begin. You should clearly state what your research question is and why it is important to study.

The widespread availability of rich data sets makes it tempting to launch into data collection based on half-baked ideas, but this is often counterproductive. It is likely that, without carefully formulating your hypotheses and the kind of model you will need to estimate, you will forget to collect information on important variables, obtain a sample from the wrong population, or collect data for the wrong time period.

**You will submit this document using the Assignment Tool on Canvas by the deadline listed in Table 1.**

## Deliverable #2: Final Paper                                        150 points

A successful paper combines a careful, convincing data analysis with good explanations and exposition. Your research paper should have the following sections.

1. The <u>introduction</u> should answer two questions **what** and **why**. <u>What you will study</u> and <u>why it is important</u>. State the basic objectives of the study and explain why it is important and how the results will be used. Your goal is to convince the user that this is an important problem worth studying.
2. <u>Describe the data</u>. Whether you used an existing database, collected your own data by scraping the web, by experiments, or by surveys, describe the contents of the data. Walk the reader through why you think this data can answer your question. Details of surveys and experiments can be presented in an appendix.
3. The <u>analysis</u> section is where you answer **how**. Document your efforts to reach a final model that you believe answers your questions with overwhelming evidence. This is where you use everything we discussed in class and of course any additional tools, methodologies you know.  The discussion should include the interpretation and strength of your empirical results. Do the coefficients have the expected signs? Are they statistically significant? If a coefficient is statistically significant but has a counterintuitive sign, why might this be true?
   If you use a technique we did not discuss, please provide a short description and a reference.
4. The <u>results</u> (or conclusions) section is where you summarize what your models tell you and what that means in the context of your objectives. Did you solve the problem you set out to solve? If not, what did you find?

**You will submit this document using the Assignment Tool on Canvas by the deadline listed in Table 1.**

## Deliverable #3:  Presentation                                        90 points

Prepare a 15-20 minute presentation to walk us through the 4 sections I have highlighted above, with particular emphasis on your analysis and conclusions. This should have enough detail to explain not only what and why but also how.
**Submit the presentation files (pdf, pptx, ..) using the Assignment Tool on Canvas by the deadline listed in Table 1.**

## Additional Guidelines

**No late submissions will be accepted.**

You are free to meet with the instructor to discuss the project as required. I recommend that the entire group attends such meetings. Please do not discuss your group's project with individuals other than your group members.

One of the learning outcomes of this project is to expose students to the experience of working in groups. Please meet with your entire group at least once every other week. In addition, divide up the work such that all group members work on each deliverable.

**Peer evaluation forms may be given after each deliverable is submitted.** Group members will be able to evaluate each other using this form. Based on the feedback provided on these forms, students who make extra-ordinary contributions to their group's project may receive additional points. Similarly, students who do not contribute sufficiently to their group's project, their group project score may be diminished.

# Academic Integrity

Plagiarism and Cheating of any kind will not be tolerated. For any academic class activity, students must follow the University of Florida Student Honor Code. Any violation of the honor code **will** automatically result in a grade of E (Fail) for this course and further sanctions that may include a suspension or expulsion from the University through the Dean of Students Office. All incidents **will** be reported to Student Conduct and Conflict Resolution at the University of Florida.

A list covering some examples of actions that break the honor code is given below:

• Copying, submitting, or viewing the work of a student(s) other than your team members.

• Submitting work that is copied from information posted on the Internet or any other source.

• Working in any capacity with or helping a student(s) other than your team members.

**Team members are jointly responsible for the academic honesty and integrity of teamwork. All team members will be equally penalized for any violation of the honor code in this project.**