# DS 2500 - Spring 2026
# Milestone 1: Team formation & topic report requirements

**Deadline:** February ~~17~~ 20, 2026, 11:59pm ET

**Total points**: 260 pts

## Finding team members

You are free to find team members on your own. Please note that all team members must be from the same lecture section.

If you would like to be *randomly assigned to a team*, please fill this [Google Form](#) by **Wednesday, February 11, 2026 11:59pm**. If you fill the Google Form, we expect that you will not join a team outside of the one that we assign you. We will send your team assignments via email by February 12 to give you plenty of time to work on the first milestone.

## Requirement

1.  **# of team members:** Teams must be of 2-4 people (only 2, 3, or 4 are allowed). All team members must be in the same section.
2.  **Report format**: Use the template provided below for the report. You can copy and paste the template to a Google Doc or Word Document and fill in all the answers. Submit as PDF on Canvas.
3.  **Report submission**: Only one team member submits on Canvas, but make sure to add the full names of all your team members on the PDF.

## Evaluation criteria

- Problem statement is clear and specific (80 points)
- Data sources are realistic and reasonable for the problem/topic. (80 points)
- All the required information is provided including all team member information (40 points)
- Report uses the provided template (30 points)
- Writing is clear, well-organized, and professional (30 points)

Reports will be graded on completeness and above evaluation criteria. TAs will provide feedback.

## Template

**Question 1: What is your team's section?**
> Your section: 1

**Question 2: Who are your team members?**
> Name 1: Jonathan Chamberlin
> Name 2: Anuhya Mandava
> Name 3: Min Yu Huang
> Name 4: Tsion Teklaeb

**Question 3: What are you trying to do? (problem or topic) Describe in a couple sentences**
Brainstormed answers
- We aim to predict the risk of chronic disease using variables from the U.S. Chronic Disease Indicators dataset. We will analyze the demographic and clinical health indicators such as:
    - age
    - cholesterol level
    - blood pressure
    - lifestyle factors
    - other related factors
- We aim to use this data to build a model that identifies patients at high risk of developing chronic disease. Our goal will be analyzing patterns and trends in chronic disease prevalence and understanding factors associated with long term health outcomes.

**Question 4: Why is this important? Short summary (2-3 sentences) of who cares about this problem, what impact it has, what implications better solutions might have.**
- Chronic diseases like diabetes and heart disease negatively impact millions of US citizens over the course of their lives, leading to worse quality of life and causing early death. By determining the patients most at risk, resources can be directed towards improving preventative care in those areas and public health funding could be concentrated in high risk populations. Otherwise, resources may be squandered unnecessarily, depleting money, time, and leading to an unhealthier nation.

**Question 5: What dataset(s) are you interested in working with? Add links if possible.**

- We're interested in working with data sets related to chronic disease indicators to determine what states have the highest levels of chronic disease and what factors are driving it.
- https://catalog.data.gov/dataset/u-s-chronic-disease-indicators - chronic disease indicators
- https://arg.org/center/national-alcohol-surveys/ - data relating to alcohol usage
- https://www.icpsr.umich.edu/web/HMCA/studies/29801/summary - data relating to management of chronic conditions

- [https://addhealth.cpc.unc.edu/data/#public-use](https://addhealth.cpc.unc.edu/data/#public-use) - contains public health data sets that consider a variety of environmental factors.