# CMP 142 - Programming Assignment #1 *Collecting and Preparing Data*

You are to work in your teams to solve the following problem.

**PROBLEM DESCRIPTION**

Over the course of the semester, your team will carry out a large-scale data science project. You will perform the following steps:   
  
1) **Analyze a real-world data science problem**,  
2) **Collect and prepare data**,  
3) Explore the data using current tools and techniques,  
4) Build machine learning models to analyze the data,  
5) Analyze the models to select the best models, and  
6) Draw conclusions about the problem based on the data model.

The first step is to identify a data science problem. Here are some examples for your team to consider: movie reviews, disease diagnosis, stock analysis, sports analytics, and many more. Please consult with your teammate to decide which problem you want to analyze.

After deciding on a problem, your team needs to decide what the **target** is for the problem. In other words, what is it that you want to predict? The target value could be either numerical (e.g., salaries) or categorical (e.g., iris species).

Your team will find and download a dataset that will allow you to explore the problem from a data-centric point of view. Your instructor will point you to possible datasets to download.

**Task 1.** Decide on a data science problem. Describe the problem. What is the target variable for this problem? Name and explain the features that will be used to predict the target variable. Done

**Task 2.** Download a dataset in a CSV format for the stated problem. Done

**Task 3.** Clean and prepare the data in the dataset for future analysis. Done

### WHAT TO TURN IN

Upload a Word document with descriptions to the three tasks mentioned above. For Task 2, provide a link to the dataset that your team will be using for this project. Upload your document and you CSV file containing your dataset to the Programming Assignment #1 - Drop Box on Moodle. Make sure you include the names of your team members at the top of each of your files.    
  
Turn everything in by **5pm on Friday, February 25**.