**CS 108L, Spring 2024**

**Final Homework / Project**

The goal with your final homework is to give you time to extend and further explore a previous homework with agent-based modeling or a previous homework with K-Means clustering.

This homework/project will take the place of your final exam. That is, there will be no final exam, and this homework/project will compose 15% of your final grade.

**Due Date:** Wednesday, May 8th, 11:59pm

**Submit** your final homework/project on the course website as a single Python Notebook (IPYNB) that includes all of your code, plots, and writing.

**Timeline and Rubric:**

1. Friday, April 26, Due In-Class  
   Bring your topic to class  
   15%

**What to hand-in:** Bring a short paragraph to class on paper describing your topic. Clearly identify (i) the question or problem that you are solving, (ii) how you will address or solve this question or problem, and (iii) how you will evaluate your results. For (iii), you could for instance describe the plots or data that you plan to generate and what they will tell you.

1. Wednesday, May 1, Due In-Class  
   Bring Preliminary Results to Class, ready to discuss

15%

**What to hand-in:** Bring your current Python Notebook (IPYNB) to class, that is, have your Python Notebook on Google Colab so that you can show it to the instructor and discuss your current progress. You should already be running some experiments and be able to discuss them.

1. Wednesday, May 8th, Due 11:59pm on course webpage  
   Submit your final homework/project on the course webpage   
   70%

**What to hand-in:** Submit a single Python Notebook (IPYNB) that includes all of your code, output, plots, and writing.

At the end of your Python Notebook (IPYNB), provide a Text Cell (or multiple Text Cells) that contains **750-1000 words** following this outline. Remember, that you are generating and evaluating quantitative information.

Definition Reminder: Quantitative information is information represented numerically, i.e., represented with numbers.

* This can include anything that can be measured or counted to that it has a numerical value.
* This includes our agent-based models, which provide measurable outcomes like percent of forest burned or number of apples that each trader has.
* And, this also includes the K-Means algorithm, which provides measurable numerical output, such as the group or cluster for each data point or the inertia of the grouping.

1. Problem Description:
   1. What is the question or problem that you are solving?
      1. This could for instance be, what is the effect of various levels of dishonesty in apple trading, or what is the effect of a fire break for various sizes on forest fires, or how can I develop better metrics for evaluating the quality of my K-Means clustering?
   2. What model will you use to address the problem/question, e.g., agent-based or K-Means?
      1. Describe that model, so that a reader which knows nothing will understand at a high-level what you are doing. Including a picture or two can help.
      2. How will you address your question/problem in your implementation? For instance, how will you implement a fire break or dishonest apple trading?
2. Results: Communication/Representation of Quantitative Information
   1. Give plots or other output showing quantitative information relevant to the question/problem that you are asking.
   2. This could for instance be percent of forest burned over many experiments, or the inertia of a clustering, or the distribution of apples among traders over many experiments.
3. Conclusion: Quantative Analysis
   1. Discuss your output (plots and/or data) as it pertains to your question/problem
   2. This could for instance be how the distribution of two types of trees affects forest burning, or how a fire break affects a forest fire, and so on.
   3. What is still unexplained?
4. Ethical Reasoning
   1. Consider topics such as sustainability, human impact, and algorithmic bias, similar to what you did for the last three homeworks on agent-based modeling and K-Means.

Here are some possible ideas for this last homework, how you could extend our agent-based models or K-Means homeworks. Feel free to come up with your own ideas, too.

1. Forest-fire model ideas:
   1. Consider the impact of including a fire break through part of the forest. This could for instance be a fire break that goes through 50% or 33% of the forest. Consider fire breaks of various size.
   2. Consider the impact of two types of trees. Perhaps this new type of tree is invasive and is more likely to burn than the original tree type.
   3. Consider forest density in either of these two contexts.
2. Apple trading ideas: Explore the effects of the following.
   1. Effect of adding consumption of apples. Here each trader could receive a new apple every K turns and eat an apple every M turns, where you need to choose K and M to be reasonable.
   2. Effect of dishonest trading. Here, some traders would skip giving apples to other traders sometimes (but perhaps not all the time).
   3. Effect of implementing a tax that takes away apples from some traders and gives those apples to other traders. You will need to design the tax.
   4. Effect of having two items or commodities to trade (say, apples and bananas).
3. K-Means ideas with the loan data:
   1. Implement and use a way to consider more information during the clustering. This could for instance be a way to use the Urban/Rural or Marital Status information.  
        
      Make sure to discuss the ethics of considering such information.
   2. Implement and use a way to quantitatively measure how good your clusters are for predicting loan status. For instance, you could compute the percent of people in a cluster that are all “yes” for their loan and the percent that is all “no” for their loan.   
        
      Give a percent correct metric for all the clusters, that is the percent of people in a cluster that are all “yes” and all “no”.

Consider how this metric improves (or does not improve) as you increase the number of clusters.  
  
Make sure to discuss the ethics of using clustering to classify people.