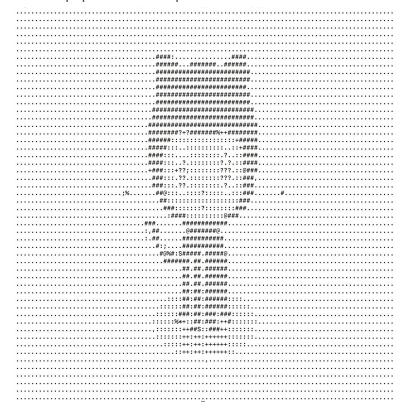**Jack Sandberg**

INFO 3401

Prof. Szafir

9/16/2018

Monday:

1. Use the command line to run asciify.py (included in the assignment repo). What does this script do? Include a screenshot of your results.

   a. This script prints an ascii picture of a cat.

   b.
```
.............................................................................
:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
:::::::::::::::::::::::::.####:...........####:.............................
:::::::::::::::::::::::.######...######..######..........................
:::::::::::::::::::::::.#####################.......................
:::::::::::::::::::::::.#####################.......................
:::::::::::::::::::::::.#####################.......................
:::::::::::::::::::::::.#####################.......................
:::::::::::::::::::::::.#####################.......................
:::::::::::::::::::::::.#####################..........................
:::::::::::::::::::::::.#####################.......................
:::::::::::::::::::::::.######################......................
:::::::::::::::::::::.########?+?#######%++########.................
:::::::::::::::::::::.######:::::::::::::::::::+#####.................
:::::::::::::::::::::.#####:::..:::::::::::..::+####.................
:::::::::::::::::::::.####:::...:::::::::::.?..::####...............
:::::::::::::::::::::.####::::..?.:::::::::?.?.::####...............
:::::::::::::::::::::.+###:::+??;:::::::::???.::@###...............
:::::::::::::::::::::.###:::.??.:::::::::???.::###,...............
:::::::::::::::::::::.###:::.??.:::::::::.?..::###...............
::::::::::::::::::;%......##@:::..:::::?:::::..:::###.......#............
:::::::::::::::::::::::.....##::::::::::::::::::::::###..................
:::::::::::::::::::::::::.....###:::::::::?:::::::::###.................
:::::::::::::::::::::::::....:####:::::::::::@###..................
:::::::::::::::::::::.###......###########.....................
:::::::::::::::::::::::,##.......@#######@.....................
:::::::::::::::::::::.:.##......###########...................
:::::::::::::::::::::::.#:;...###########...................
:::::::::::::::::::::::.#@%#:S#####.#####@...................
:::::::::::::::::::::::::.#######.##.######.................
:::::::::::::::::::::::::::.##.##.######.................
:::::::::::::::::::::::::::.##.##.######.................
:::::::::::::::::::::::::::.##.##.######.................
:::::::::::::::::::::::::::.##:##:######.................
:::::::::::::::::::::::::::::##:##:#####:::::..............
:::::::::::::::::::::::::.::::##:##:#####:::::::...........
:::::::::::::::::::::::.::::::###:##:###:###:::::::.........
::::::::::::::::::::::::.:::::%*+::##:###:++#:::::::.......
::::::::::::::::::::::::,:::::::++##S::###++:::::::.......
:::::::::::::::::::::::::::::::++:++:++++++:::::::.........
:::::::::::::::::::::::::::::::++:++:+++++:::::............
:::::::::::::::::::::::::::::::++:++:++++++::..............
:::::::::::::::::::::::::::::::::::::::::,:::::::::::::::::::::::
:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
::::::::::::::::::::::::::::::::::::::::::_::::::::::::::::::::::::::::::::::::
```

2. Install the Delorean Python package. What command did you use? Then, from the command line, launch Python. Use it to execute the following commands:

   a. I used the command pip install Delorean to install it.

   b.
```
from delorean import Delorean
EST = "US/Eastern"
d = Delorean(timezone=EST)
print(d)
```

      c.   What prints to the command line?
           i.   Delorean(datetime=datetime.datetime(2018, 9, 10, 16, 22, 27, 677358), timezone='US/Eastern')

3. Use the command line to download http://www.colorado.edu/. What command did you use? Make sure to include this file as part of your submission.
   a. wget http://www.colorado.edu/
   b. Wget wasn't working for me, but wget is the command used to download this file. I instead used the 'index.html' file posted on Canvas.

4. Search the downloaded file from #3 for the word "Colorado." What command did you use? How many lines did this find?
   a. I used 'grep Colorado index.html' to find Colorado
   b. Using 'grep Colorado index.html -c' returned that there were 17 instances/lines of it

5. Now, print the lines containing the word "Colorado" to a file to "search_results.html". What command did you use? Include this file as part of your submission.
   a. cat index.html | grep Colorado > search_results.html

Wednesday:

CU has recently come across a variety of findings stating that regular gym attendance can lead to a higher GPA. In an effort to understand this effect and how they might update their services and facilities to encourage students to be active, they want to turn to the data to both understand the current state of gym use, its potential effects on students, and how they might target specific facilities or programs to improve attendance.

First, identify three datasets that CU can provide you to support this effort.

1. A dataset of the GPA's of students that regularly go to the gym (2x+ per week)
2. A dataset of the frequency of machines used at the Rec
3. A dataset of responses of students answering questions about the gym, such as which machines they find the most useful for de-stressing

Then, use the CoNVO method to scope a project that will solve this problem. Provide sufficient detail in your response such that someone could reason about and execute your plan.

     In order to tackle this project, its important to think about it holistically. First, it's important to recognize the context of this problem. We are trying to figure out a way to

encourage the students of CU Boulder to use the gym and/or workout more often. This is to figure out if there is a correlation between gym attendance and GPA, and if there is, to encourage higher attendance. This will require a lot of learning about the students' needs and preferences in regards to the gym. This is being done to help the students, as the ideal outcome would be to have higher gym attendance, and as a result, higher GPAs across the board.

Secondly, it's important to recognize the needs for the problem. We are trying to address questions in regards to GPA-attendance correlations, student preferences for gym utilities and programs, the status of the gym and its usage, and what CU can do to improve its facilities, utilities and programs to achieve higher attendance. By researching data about all of these potential driving questions, choices could be made to improve specific aspects of the gym, as well as the image of the gym, in order to attract more students.

In order to achieve results and understand this problem, the most important step is to receive feedback and use data to make decisions. This is a multi-step project, but there are some important steps. First, looking at the GPA of those who already attend the gym regularly would help us understand how impactful the gym can be. Secondly, using data about the machines, utilities, and programs about the gym itself could allow the gym to alter its current arrangements and priorities to more accurately reflect and help the students. Understand the usage of the gym is critical. Thirdly, by reflecting and analyzing information given directly by the students, important input could be used to have more people come. For instance, if CU knew which things students used to de-stress, the gym could put more of an emphasis on those, to have students feel more comfortable using the gym when they need it. All of this data, if collected and analyzed properly, would directly allow the gym to improve itself, by making new, informed, intentional decisions about its function. To gather even more helpful input, the gym could make pamphlets saying that it is going to change x, y, and z, and asking students of their thoughts on said changes. These mock-ups would be a helpful communication tool for understanding the scope and context of the problem.

Lastly, as the entire point of the problem is to make educated decisions to have more students use the facilities, it's important to consider the outputs. Any data collected should be used to make informed decisions, while keeping the scope of the problem non-abstract: the intent is to help students. The data collected should be kept, labelled, and stored, as this data would probably need to be in-depth and extensive. Storing the data would allow future people to use it for their own needs, and although the needs of students may change, the gym is one aspect of the university that could remain consistent in its usage and image over a long time. All of the data would translate very well back into the context of the problem, and could be analyzed in many ways to help students use the gym more often and fully.

Once the data is analyzed, the final step would be to change the rec center, in its functions, utilities, and programs, so that it would be more beneficial to the student. Adding/removing various equipment, changing the lessons offered, and making new facilities/courts would all be possible changes that could better help students.

Friday:

In class today, you learned about a historical instance of data science in practice. Find three additional problems in history that were addressed using data. For each of these problems, describe the following:

1. During World War 2, the British Air Force (RAF) was trying to decide how to improve their fighters. They had access to fighters that returned from battle, but data that let them improve was that many fighters returned shot, but still functioning. The British at first made adjustments to the airplane model by reinforcing, with additional metal, the areas that had been shot. However, upon further reflection and finding that a lower proportion of planes were returning after being shot, they realized that the areas that on planes that had been shot but survived were areas of little importance, as they realized that they were not critical areas regarding function of the aircraft. They lowered the reinforcement on those areas, and increased it elsewhere. Immediately, a higher proportion of fighters returned with bullet holes in them, meaning that their strategy had worked. There is not much that could be done differently about how this problem was solved. The only thing that could have been done better is doing analysis, in the beginning, of locations that needed focused protection from bullets, rather than focusing on putting further protection on all places in which planes that were returning had not been shot. E.g. focusing on the engine and cockpit.

2. During the end of World War 2, the Allies were trying to predict how many tanks the Germans had left, so that they could gauge how much resistance to expect as they pushed into German-Austrian territory. The Allies realized that there were serial numbers left on the parts of captured or nonfunctional German tanks. After battles with the Germans, the Allies would record this data, and perform statistical analysis based on the recorded serial numbers. They found that they could, with high accuracy, predict the number of tanks that were left in the German arsenal. There were enough serial numbers recorded with not enough variation that this method worked. If performed today, the analysis could have gone further by trying to find patterns or tendencies in the serial numbers, to predict if the tanks were being spread thinly or not. This could be done by comparing serial numbers from different regions, as there would have had to have been multiple sites that built tanks, using different serial code imprinting methods.

3. During World War 2, the Germans were using a machine to heavily encrypt war messages. The messages themselves were available to the allies: it is thought that the Germans knew we could intercept them, but they simply weren't worried about them being decrypted, as the encryption method changed every day. Therefore, they didn't care. The Allies put tons and tons of resources into decrypting these messages: they

tried doing it by hand, but eventually started inventing new devices that would help: the Allies literally invented computers to decrypt war messages from the Germans. Cycle groups and patterns were written down and tracked, and since the Germans weren't worried about the messages themselves being tracked, there was a plethora of information and options to decrypt. Eventually, it worked. With enough time and effort, the messages were decrypted using various techniques and the Allies were able to decrpy all further messages, primarily due to the Polish and British. If this problem presented itself today, there would be a much heavier reliance on computers: they have already been invented and have had a lot of work put into them, therefore they could be harnessed to do a lot of the work for us.