Feedback Exercise

Group we met with: *Squirrel census - NYC - Squirrels in My Park (SIMP)*

Feedback we gave/questions we had:

- For the census recorder - is there bias when recording data?
- Potentially look at different years (2018 vs post-pandemic) to compare behavior and add project complexity.
- Look into Python packages for GIS - mapping squirrels to a map
- Are there species classifications? Can you classify them into species with available data?
- Do you have access to the weather data? Fewer squirrels on a rainy day? When there's bad AQI?
- For exploratory analysis - Using geographical data, are there squirrel hotspots?

Feedback we received:

Additional data we could include:

- Do we have access to diet prior to admission?

The dataset does not contain data on the patient diet prior to admission, however it does contain a value for HbA1c (glycosylated hemoglobin). While this lab information gives an indication of treatment compliance and can be tracked over time following dietary changes, it is not an exact proxy for specific diet and does not directly correlate to dietary choices.

- Were each of the patients already diagnosed with diabetes prior to admission?

No, each of the patients were not already diagnosed with diabetes prior to admission. Some of the patients were (which is captured in the DIABETESHIST.csv file). This information is not relevant to our study, however, as each patient in the dataset presented in DKA - a prior diagnosis does not change the physical exam findings. Eventually we may be able to do a subgroup analysis for patients with a previous diagnosis, to determine whether they were more or less likely to experience cerebral edema or require admission, but that is not the question we are aiming to answer at this time.

- How many variables will be included? Is it robust enough?

We are including several variables including the following organ systems: cardiac, respiratory, HEENT (head, eyes, ears, nose, throat), GI, dermatology, GU, renal, neurology, psychology, and musculoskeletal. Within those categories, we have broken them down even further to individual physical exam findings. The collection of these findings has led us to assess a total of 14 variables against the development of cerebral edema, and admission to the hospital. These 14 variables are the most common

physical exam findings within the respective organ systems/categories. We currently feel that we have captured an appropriate amount of variables to make our assessment, however, we will not fully be able to answer this question until we have completely our preliminary analysis in its entirety.

Ethical considerations:

- Something to add to our current ethical considerations:
  - If parents could see this would it cause undue stress if they thought their child was going to go into a more severe state. There are a lot of ethical considerations when delivering really intense news to families.

While it is possible that stress would be caused by suggesting their child may progress to a more severe/concerning disease state, this may also allow for intervention at an earlier time point to prevent decompensation or prevent the development of cerebral edema.

There is very little ethical consideration when delivering intense news to families. Withholding information is far more inappropriate, and may have ethical considerations as families have the right to know that their child may progress. There is the potential for added stress, especially if their child does not progress to a more severe state. However, it is often better to be incorrect about disease progression, than having progression without disclosing that there was the potential for it to occur.

Data Cleaning:

- Create a dictionary of all possible iterations of words and then create a function that will match these descriptions to classify to simplify cleaning steps.
  - Like the example of "tachy", "tachycardia", "tach"

This portion of the data cleaning was conducted using regex to ensure that we captured all variants and misspellings of physical exam findings. This is outlined and commented thoroughly in our attached notebook file.

Analysis:

- Should we do a linear regression for the Glasgow score? And logistic regression for binary outcomes? (Rather than only looking at logistic regressions)

Please refer to the analysis portion of our attached notebook file for full discussion of our intended approach to analysis. We plan to do a multi-linear regression for the Glasgow coma scale, need for CT scan/imaging, development of cerebral edema/treatment for cerebral edema, and hospital admission to determine if there are any physical exam findings that can predict these outcomes.