Healthcare Worker Shortage Project Proposal

By Crystal Han

Potential Client:

- Cities or towns looking to mitigate the shortage of healthcare workers/professionals will seek to find places with an influx of these workers/professionals. The places with shortages should be able to determine how many or the density of healthcare workers in an area and be able to allocate resources from another area to an area in need. By going through these areas, the client should be able to:
    - Determine if an area has a shortage of healthcare workers/professionals
    - Determine whether an area can allocate healthcare workers/professionals to another area in need.
    - What is considered not a shortage?
    - How many healthcare workers/professionals are needed based on the population size of city/town?

Assignment Objective

- What is the density/number of healthcare workers in an area?
- Classify which areas are in need of healthcare workers and which are not
- What is the ratio between population and number of healthcare workers/professionals? Scaling required to have an average?
- What is considered not a shortage and what is a shortage?

Data Description

- In this data exploration, I plan to use data sets provided by the google BigQuery datasets for healthcare workers shortage. I want to use these dataset to determine if an area has a shortage through the degree of shortage, status code, HPSA score, and other features.
- In this analysis, I expect to be able to classify areas of shortages and use these features to label each area with a 0 for no shortage and a 1 for shortage.
- Through analysis and classification, I should be able to classify an area with no degree of shortage (null value) with a number from low (1) to high(5) based on their HPSA score or other features.

Tools

- The raw data will be put into an SQL database where I will query from that database into Python. Libraries such as pandas will be used to read in the data and create a dataframe, then visualized through Matplotlib (for basic plotting of

traffic) and Seaborn (for comparison plotting, distribution of traffic, identify traffic patterns). The data will be provided by google BigQuery.
- If I require additional tools beyond what is required, I will use them.

MVP Goal

- Datasets will be turned into a dataframe where we will trim down columns that are not needed and keep the columns of interest and then use a decision tree to determine shortage. I will visualize this through matplotlib and provide insight into areas of shortage with scatter plots and more.