**Project Title**

Diabetes Predictor

**Team Members**

Kenway

Tim

Ipsita

**Project Description/Outline**

Run through research to predict groups who are likely to have diabetes based on various health factors.

**Research Questions**

|  |
| --- |
|  |
| * Is there a correlation between age of the patient and hypertension and heart disease? (Done) |
| * Is there a correlation between gender/sex of the patient and hypertension and heart disease? (Ipsita) * Find the correlation between bmi and blood\_sugar level and whether a person is diabetic? (Ipsta) |
| * What age group is diabetic? (Create a bar chart for each age group) (Tim) |
| * What age group suffers from heart disease and hypertension more? (Tim) |
|  |
| * What's the correlation between HbA1c level and diabetes, heart disease and hypertension? (Tim) |
| * What's the correlation between smoking history and the three diseases? (Clean up: drop the “No Info”) (Kenway) * Compare census population data to the healthcare dataset. (Using the census API) (Data) (gender/age distribution) (Kenway) (compare for how accurate our study is compared to population) |
|  |
|  |

**Datasets to be used**

Censu.gov package in python:

c = Census(census\_key, year=2019)

c.acs5.get()

Healthcare dataset: <https://www.kaggle.com/datasets/nanditapore/healthcare-diabetes>

**Rough Breakdown of Tasks**

* Clean up the data for each research question
* Create a repository for our group project with a main branch and separate branches for each.