Project 1 - Proposal

Group 1

- Natalie Ericson
- Dawson Michael Marcade
- Alex Stewart
- Cortney Cantwell
- Darnyesha Willis

Description

For our project, we are analyzing multiple datasets in search of potential predictors of diabetes. There are many factors commonly known to be associated with having diabetes, such as a genetic predisposition to developing type 1 diabetes as a child, or developing type 2 diabetes as an adult due to lifestyle choices. Our group is going to take a deeper dive into features such as BMI, age, gender, physical activity, smoker status, and more to evaluate if there is one or multiple factors that are statistically correlated to the presence of diabetes.

Research Questions

We are still in the process of finalizing our research questions, however we have narrowed down our specific areas of focus and will fine tune these questions during class on Wednesday, July 26. Currently, our research questions are:

1. What are potential predictors of developing diabetes?
2. If there are multiple predictors, is there one that is the most reliable? If so, which one?
3. Can we use any of these factors to accurately predict if an individual has diabetes?

Datasets

1. Diabetes Health Indicators Dataset - binary 50/50 split
    a. Sourced from Kaggle, updated 2 years ago
    b. Linked HERE
2. Diabetes Prediction Dataset
    a. Sourced from Kaggle, updated 4 months ago
    b. Linked HERE

Task Breakdown

We have not yet determined a breakdown of project tasks and assigned them to members of the group. We will be tackling this and solidifying it during class on Wednesday, July 26. We will update this section after doing so.

Each person will create their own analysis for the relationship between Diabetes and one of the indicators. The breakdown is:

- Alex: BMI
- Dawson: Smoking Habits
- Natalie: Physical Activity
- Cortney: High Blood Pressure
- Darnyesha: High Cholesterol

Hypothesis

H_O : Any of High Blood Pressure, high cholesterol, smoking habits and activity rating are best indicators of whether or not someone has diabetes

H_A : Out of high blood pressure, high cholesterol, high BMI, smoking habits and physical activity rating, BMI is the most likely indicator for risk of Diabetes

Comparisons:

1) Run Pie chart comparison on column value vs diabetes value (5 charts)
2) Bar chart of all values + having diabetes, bar chart of all values + not having diabetes (2 charts)