Project 4 Proposal

Jennifer Tran

Khristopher Prince

Geraldo Bedoya

Sumain Hemani

Dataset:

<https://www.kaggle.com/datasets/prosperchuks/health-dataset?resource=download>

We are interested in evaluating a patient’s risk for the disease based on predisposition to a disease, lifestyle factors such as physical activity, food consumption, and drinking alcohol, and risk depending on age. The dataset chosen evaluates 70,000+ survey responses on Diabetes, Hypertension and Stroke Prediction with factors such as age, sex, high cholesterol, physical activity, BMI, fruit consumption, smoker status, and more.

To start, we plan on using MongoDB to store the data, create tables, run SQL queries, use ML to predict the disease risk, and visualize our chosen interests in Python and/or Tableau.