Project 4 Group 7

1. What is your dataset and why

   a. Pima Indians Diabetes Database, the objective of the dataset is to diagnostically predict whether a patient has diabetes, based on certain diagnostic measurements included in the dataset. Several constraints were placed on the selection of these instances from a larger database. All patients here are females at least 21 years old of Pima Indian heritage.

   b. With how our lifestyles are, in terms of what we eat diabetes is very prevalent for all.

   c. https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database

2. 1-3 research guiding questions

   a. High level

      1. Which features are more helpful in training/predicting diabetes

      2. Can a limited data set be used to predict diabetes (limitations: all females and specific demographic)

   b. Each should be able to be answered in 1-2 visualizations and/or leaderboards (Tableau)

3. Inspiration - other code/visualizations

   a. https://www.kaggle.com/code/shrutimechlearn/step-by-step-diabetes-classification

   b. Tableau Public

4. Visuals (we would be using a variety of charts some include)

   a. Scatterplot BMI vs Blood Pressure

   b. Heat map correlation between age, insulin, bmi, glucose, pedigree

   c. Bar chart Age vs BMI

   d. Line chart avg insulin for pregnancies in specific age groups

5. what are you PREDICTING/RECOMMENDING?

   a. We are predicting whether a person has diabetes based on pregnancy, age, glucose, blood pressure, BMI, pedigree function, insulin, and skin thickness.

    b. Machine learning logistic regression

6. color palette

    a. https://coolors.co/122c34-224870-2a4494-4ea5d9-44cfcb

7. roles & responsibilities

    a. Roles

        1. Data cleaning - all

        2. Research questions – all

        3. ML – Ed

        4. Tableau – Khalil & Ugur

        5. Slides – All

        6. Flask - Victor