DIABETES HEALTH INDICATORS DATASET

Project 4 -- Group 7

Sierra Sarkis, Neyda Morales, Hannah Lashway, Thripura Pakala

OVERVIEW

1. Project background and description

This project aims to utilize machine learning, create Tableau visuals, and enhance our presentation skills.

Dataset: The Diabetes Health Indicators Dataset includes columns such as HighBP, CholCheck, BMI, Smoker, Stroke, Heart Disease, Physical Activity, Fruits and Vegetables Consumption, Age, Sex, Healthcare Access, General Health, Mental Health, Physical Health, Education, Income, and Difficulty Walking.

Inspiration/Reasoning: Considering our shared interest, we have chosen to work with the Diabetes Health Indicators Dataset to explore how various health indicators and lifestyle factors contribute to diabetes prevalence and management. Understanding these relationships can help inform public health strategies and guide individual health decisions.

You can find Diabetes Health Indicators Dataset at the following link:
https://www.kaggle.com/datasets/alexteboul/diabetes-health-indicators-dataset

2. Project scope

This project focuses on analyzing a CSV dataset with a maximum size of 50MB. While the dataset can be sourced from various platforms, we recognize that Kaggle.com provides an extensive collection of valuable datasets and prior analyses to aid our research.

Our analysis will feature five core visualizations, with the possibility of including additional visualizations if time permits. The planned visualizations will encompass a variety of formats, including bar charts, line charts, bubble charts, horizontal bar charts, and maps, all designed to effectively convey the insights derived from the data.

To ensure visual consistency and enhance the overall presentation, we will incorporate a cohesive color palette across all graphs. Additionally, roles and responsibilities within the team will be clearly defined and outlined in our project proposal, collaboration and accountability throughout the project.

Possible visualizations for the dashboard:

- **Bubble chart:** Correlation between BMI and diabetes prevalence, with bubble size representing the number of individuals in each category.
- **Bar chart:** Comparison of diabetes prevalence among different age groups and sex.
- **Stacked bar chart:** Distribution of health conditions (HighBP, Stroke, Heart Disease) among individuals with diabetes.
- **Heatmap:** Visualization of the relationship between physical activity levels and diabetes prevalence across different income levels.

- **Map:** Geographic distribution of diabetes prevalence across states, with filters for age and sex.
- **Table:** Summary of health indicators by demographic characteristics (age, sex, education)

3. High-level requirements

- Project Proposal
- Dataset and Inspiration Links
- Visualizations
- Color Pallet Consistency across the chart
- Roles and Responsibilities
- Analysis and Presentation of the data
- Machine Learning, Flask, Tableau Dashboards, and HTML Pages

4. Deliverables

- Project Proposal
- How do lifestyle factors (such as physical activity, smoking, and diet) correlate with diabetes prevalence across different demographics?
- What is the impact of socioeconomic factors (education and income) on diabetes health indicators?
- How do health conditions (such as high blood pressure and heart disease) relate to diabetes management and outcomes?
- Can survey questions from the BRFSS provide accurate predictions of whether an individual has diabetes?
- What risk factors are most predictive of diabetes risk?
- Can we use a subset of the risk factors to accurately predict whether an individual has diabetes?
- Can we create a short form of questions from the BRFSS using feature selection to accurately predict if someone might have diabetes or is at high risk of diabetes?

5. Color Palette

We chose this color palette which suits better to show the mild and intensity for the Diabetes Health Indicators Dataset.

https://coolors.co/palette/8ecae6-219ebc-023047-ffb703-fb8500

6. Work Responsibilities

Sierra Sarkis: Review project proposal, working together with dataset cleaning and analysis-python files, Machine learning.

Neyda Morales: Review project proposal, working together with dataset cleaning and analysis-python files, Machine Learning.

Hannah Lashway: Review project proposal, working together with dataset cleaning and analysis-python files and work with Tableau Dashboard visuals.

Thripura Pakala: Created project proposal, working together with dataset and work with Tableau Dashboard visuals.

The team will share responsibilities of organizing the daily standups.

Git Hub link: https://github.com/hannahlashwayyy/project04-group-07