Diabetes, Hypertension and Stroke Health Prediction

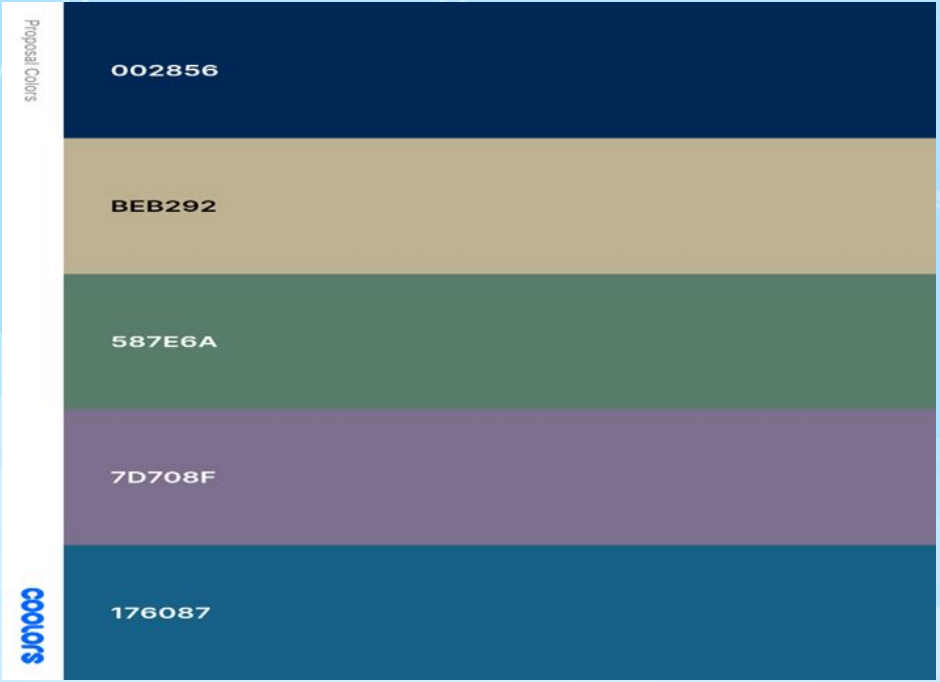
By Eugenio Elizondo, Jason Wloszek, William Parker and Kenny Berry

Project Purpose & Inspiration

- Purpose: Our primary goal was to create an interactive visualization dashboard and machine learning models that would give us a better understanding of these illnesses.

- Inspiration: As heart diseases is the leading cause of death in the United States of America we should be looking into the causes/correlations to our deteriorating health. This data sheet provided a lot of such info

Design Concepts



- Our dashboard design is based off of healthcare color palettes commonly used for research
- Each bold and contrast from each other well

Research Questions

The dashboard allows us to explore several research questions:

1. How at risk you may be for the specified conditions based on input criteria.
2. What are some of the most common risk factors for diabetes/hypertension/stroke?
3. How accurately are we able to predict someone's risk of developing one of these conditions?
4. Are there specific age groups that see a higher rate of development of a specific condition?
5. Does fruit/vegetable consumption have an impact on diagnosis rates?
6. Does gender have an impact on diagnosis rates?

Live Demo

https://wrp94.pythonanywhere.com/

Conclusions

- Diabetes:
  - The higher the BMI and general health (poorer) the higher probability there is for you to have diabetes
- Stroke:
  - The higher the glucose level and BMI, the higher the probability there is for a patient to have diabetes
  - Noticable is that there is a higher count of stroke patients for married compared to a fewer count for not ever married
  - The jobs with the highest stroke counts are Private jobs, Govt Jobs, and Self-employed. Oddly enough, a job that had no stroke count is children
- Hypertension:
  - It appears if the slope is downsloping there is a higher probability of having hypertension, vise versus if the slope is flat then lower probability of having hypertension.
  - If the resting ECG is abnormal there is a higher chance of having hypertension

Bias & Limitations

- Some of the data is generated, referred to in the description as ‘augmented’
- Certain columns within the data are based off of the patient/users discretion, which would leave room for people to not be giving the most accurate reflection of their actual health, highlighted in the ‘GenHlth’, ‘HvyAlcoholConsumption’, ‘MentHlth’, ‘PhysHlth’ columns
- Most of the columns measured are binary, 1 for yes, 0 for no for the corresponding column. Age is bracketed, meaning 1 = ages 18-24 and so on for the diabetes dataset. Columns will be dropped based on heatmap of correlation
- Not all of the columns between the 3 datasets are consistent with each other

Future Works

- There are more health based datasets that would be very possible to add on and possible work into the website

- Since some of our data is augmented, it would be more ideal to find a dataset that is both complete and real

Works Cited

News-Medical. (n.d.). What is type 2 diabetes? News-Medical. Retrieved October 1, 2024, from https://www.news-medical.net/health/What-is-Type-2-Diabetes.aspx

National Heart, Lung, and Blood Institute. (n.d.). Stroke. National Institutes of Health. Retrieved October 1, 2024, from https://www.nhlbi.nih.gov/health/stroke

Tableau Public. (n.d.). Data science capstone - healthcare. Retrieved October 1, 2024, from https://public.tableau.com/app/profile/jay4006/viz/DataScienceCapstone-Healthcare/DATASCIENCECAPSTONE-HEALTHCARE

World Health Organization. (n.d.). Diabetes. World Health Organization. Retrieved October 1, 2024, from https://www.who.int/health-topics/diabetes#tab=tab_1

World Health Organization. (n.d.). Hypertension. World Health Organization. Retrieved October 1, 2024, from https://www.who.int/news-room/fact-sheets/detail/hypertension

Kaggle. (n.d.). Health dataset [Data set]. Kaggle. Retrieved October 1, 2024, from https://www.kaggle.com/datasets/prosperchuks/health-dataset

Q & A