Heart Diseases Analysis

BY ISBELIS CASTRO, GAVIN PLEMON, SAM HOEMANN, STEPHEN FERRIER

Content Layout

- Purpose
- Inspiration
- Design concepts
- Research Questions
- PIVOT TO LIVE DEMO
- Conclusions
- Limitations/Bias
- Future work

Purpose

Heart Disease

Leading cause of death in the United States

Inspiration

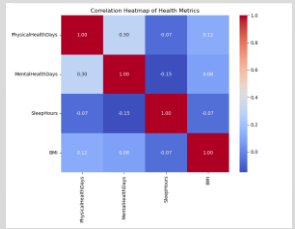
- Kaggle
  - Code recommend and learning in class.
- Tableau free

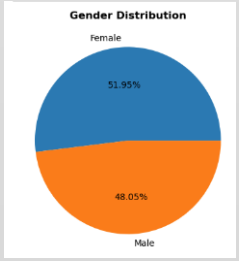
Designs and analysis type

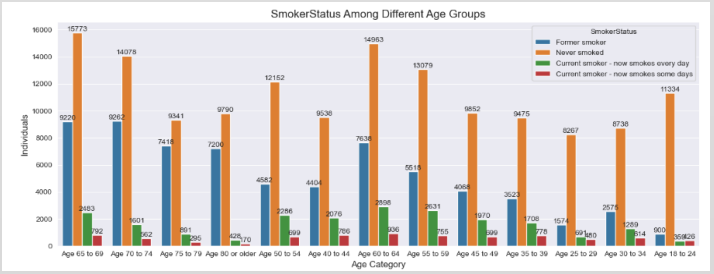
- Cleaning data
- Visualization
- Color Design
- Machine Learning
- Tableau Dashboard

Outputs

- Notebook: Cleaning data
- Notebook: Visualization
- Story Dashboard Tableau (n=2)
- Notebook: ML model
- Webapp



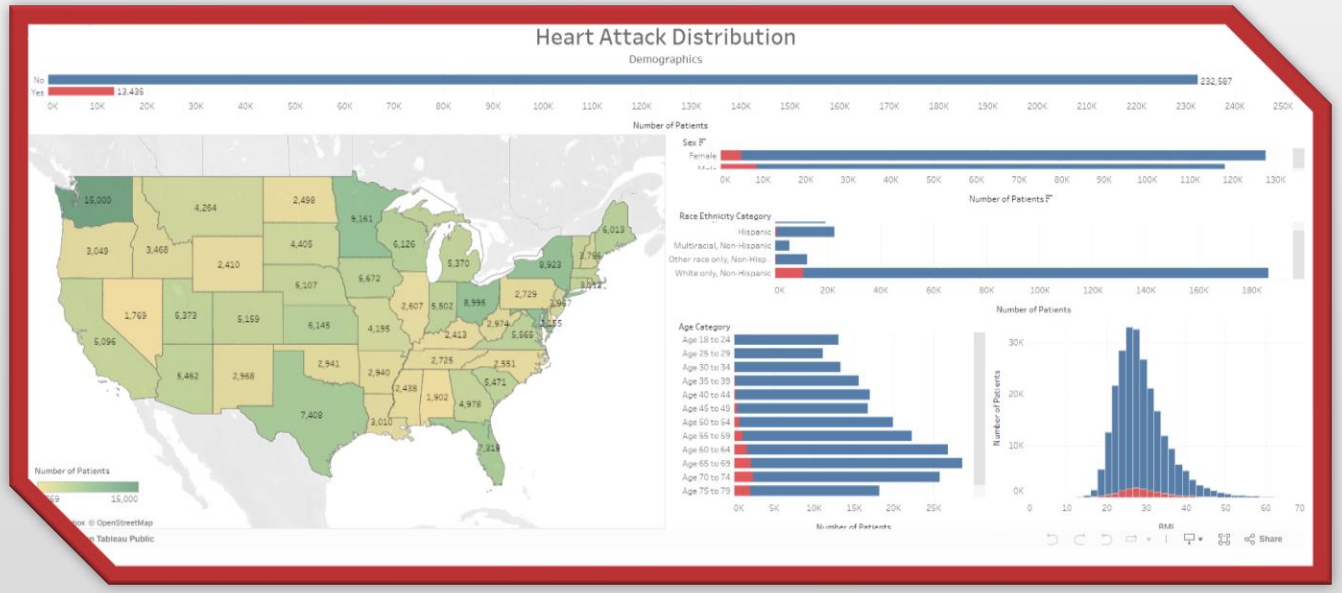




Inspiration

- Kaggle

- Tableau



Design concepts Heart Disease analysis

Color design

- Autumn
- Orange
- Red
- Brown
- Bootswatch 4.5.2 –united
- Style “darkgrid” (python)

Data set:

- heart_2020_cleaned
- heart_2022_no_nans
- US_GeoCode.csv
- world_country_and_usa_states_latitude_and_longitude_values.csv (add Guam and Virgin Islands)

Visual:

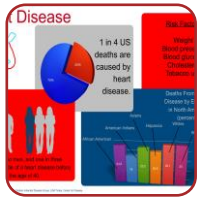
1. Map
2. Line Chart
3. Bar Chart
4. Heatmap
5. Pie chart
6. Horizontal bar
7. Packed bubbles
8. Dual combination.

Visualization (python)

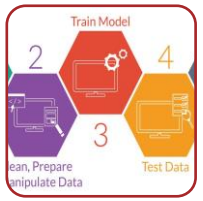
Tableau

Webpage

Research question:



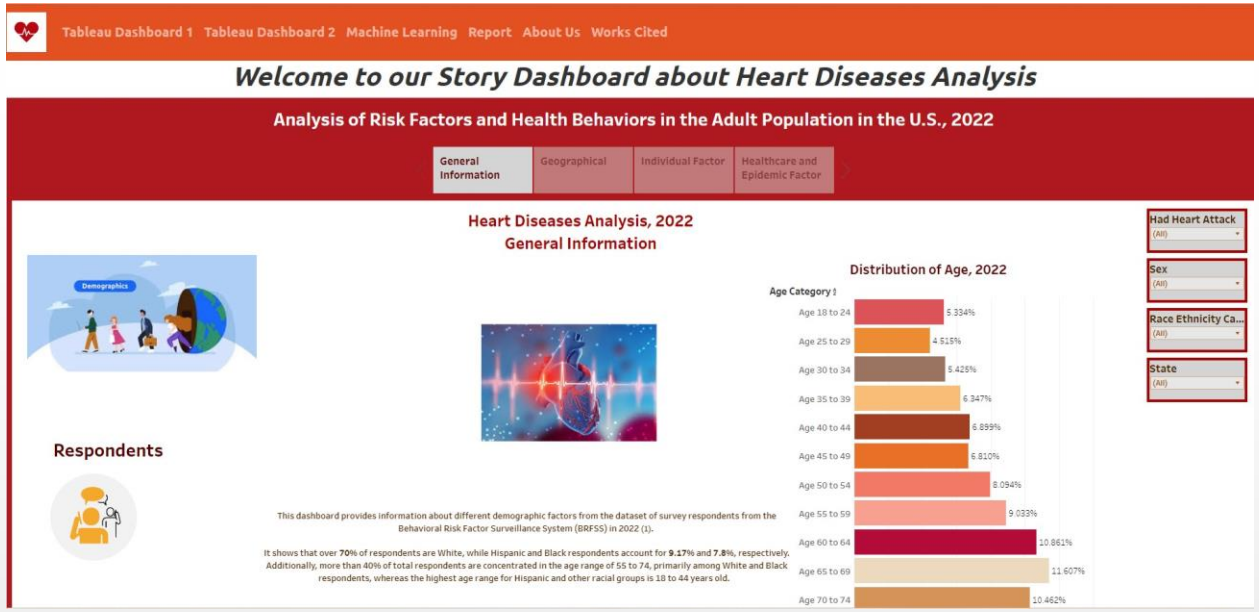
What factors indicate the highest likelihood for heart disease?

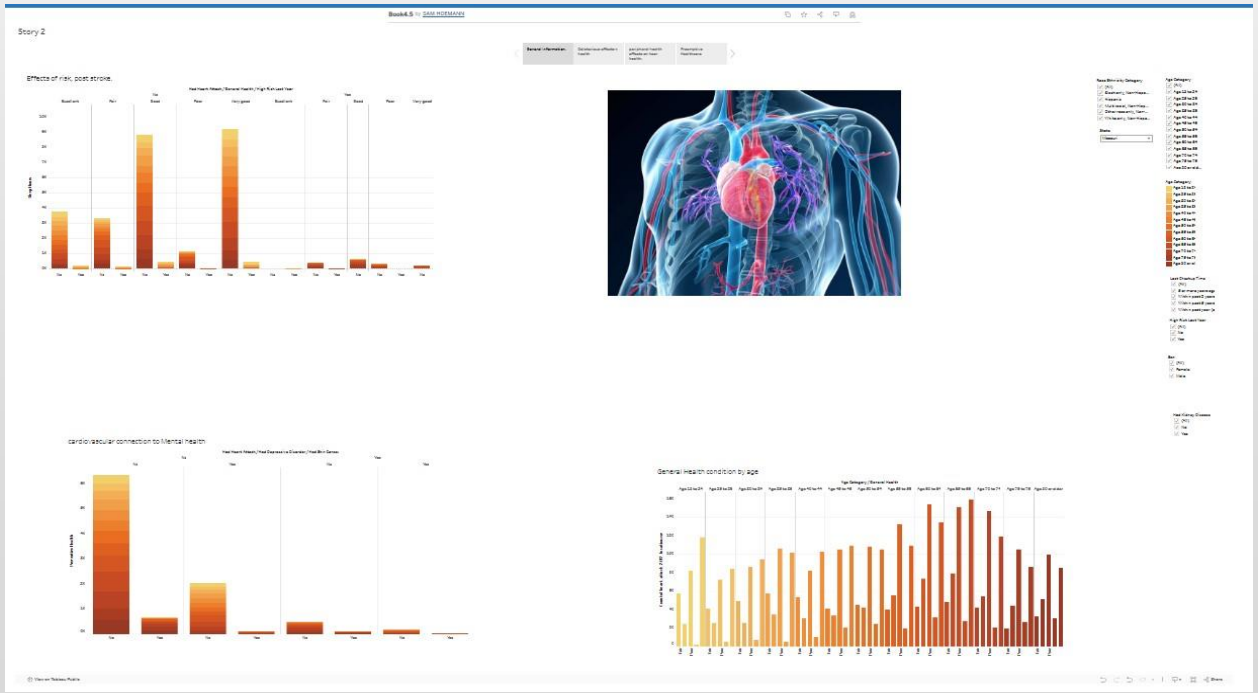


Can we use a machine learning model to accurately predict based on health factors whether someone has or is likely to get heart disease?

What are the key demographic factors (age, gender, ethnicity) that significantly influence heart attack risk?

How do lifestyle factors (smoking, physical activity, diet) and medical history (Musculoskeletal, Infectious, Respiratory, Mental Health, Cardiovascular, Metabolic disorders, Cancers, Kidney diseases) correlate with heart attack occurrences in different demographic groups?

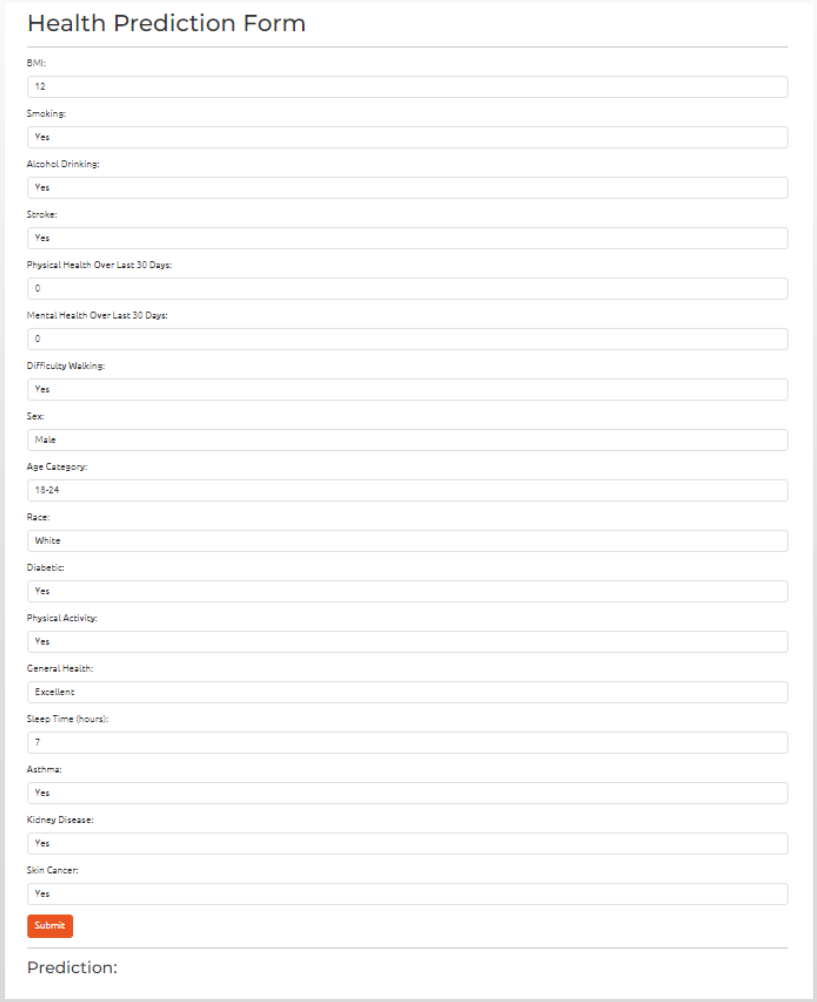




Story dashboards

Tableau – Heart Diseases Analysis.

HeartDise	BMI	Smoking	AlcoholDr	Stroke	PhysicalH	MentalHe	DiffWalkir	Sex	AgeCategc	Race	Diabetic	PhysicalA	GenHealth	SleepTime	Asthma	KidneyDis	SkinCance
No	16.6	Yes	No	No	3	30	No	Female	55-59	White	Yes	Yes	Very good	5	Yes	No	Yes
No	20.34	No	No	Yes	0	0	No	Female	80 or olde	White	No	Yes	Very good	7	No	No	No
No	26.58	Yes	No	No	20	30	No	Male	65-69	White	Yes	Yes	Fair	8	Yes	No	No
No	24.21	No	No	No	0	0	No	Female	75-79	White	No	No	Good	6	No	No	Yes
No	23.71	No	No	No	28	0	Yes	Female	40-44	White	No	Yes	Very good	8	No	No	No
Yes	28.87	Yes	No	No	6	0	Yes	Female	75-79	Black	No	No	Fair	12	No	No	No
No	21.63	No	No	No	15	0	No	Female	70-74	White	No	Yes	Fair	4	Yes	No	Yes
No	31.64	Yes	No	No	5	0	Yes	Female	80 or olde	White	Yes	No	Good	9	Yes	No	No
No	26.45	No	No	No	0	0	No	Female	80 or olde	White	No, borde	No	Fair	5	No	Yes	No
No	40.69	No	No	No	0	0	Yes	Male	65-69	White	No	Yes	Good	10	No	No	No
Yes	34.3	Yes	No	No	30	0	Yes	Male	60-64	White	Yes	No	Poor	15	Yes	No	No



Model

Creation

- Encoded Categorical variables
- Scaled numerical variables
- Decision Trees Classifier model was used

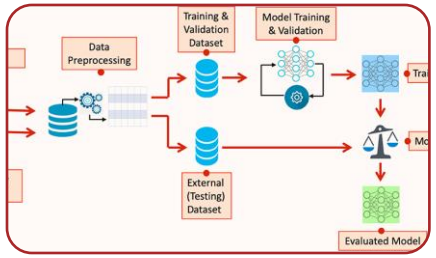
Prediction Form

- Form Takes in user inputs
- Logic and app.py takes in and casts the variables
- ModelHelper function runs model and returns a result to the user

PIVOT TO LIVE DEMO

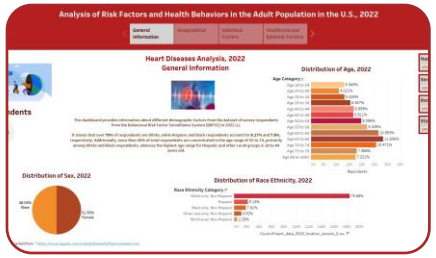
- https://isbelis.pythonanywhere.com/tableau

Conclusions



ML

- The most important predictor to the model was BMI at 0.364 compared to the next highest of 0.103
- Sleep time was number 2 on the list with a score 0.103
- Overall hard to make accurate predictions with a model from this data



Story Dashboard

- Highlight the critical demographic factors linked to heart attack occurrences (e.g., age, race).
- Emphasize that a significant portion of heart attack patients are aged 60 and older.
- Note the notable differences in heart attack prevalence across various racial categories, indicating the need for tailored public health strategies.

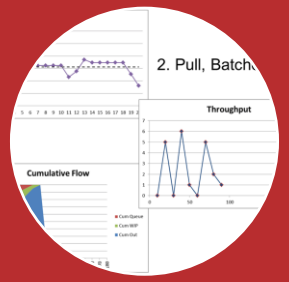
Call to Action:

- Encourage stakeholders to use the insights from the dashboard to develop targeted interventions.
- Suggest collaboration among public health officials, policymakers, and community organizations to address identified disparities.

Limitations/Bias

Limitations:

Data is only for US



Limited time to run tests

Bias:

Anchoring Bias:

- With limited time we located and used the first piece of evidence we could find, which may influence results.

Future work

Continuing work on this project would include:

- Expanding the dataset to include more variables or larger sample sizes for more comprehensive analysis.
- Utilizing different or additional machine learning models to enhance predictive accuracy and uncover deeper insights.
- Highlighting the potential for further research to investigate the root causes of disparities and assess the effectiveness of various interventions.

This Photo by Unknown Author is licensed under CC BY-NC-ND