Heart Disease Investigation In the United States

Team Members:

Abiodun, Dev, Kikelomo, Lyvia, & Taylor

Project Description

Heart disease is the number one killer of people in the United States. Our goal is to analyze a multivariable demographic through describing and comparing data to answer the following questions:

- What increases the chances of heart disease?
- What are the leading causes of heart disease?
- Where did we find our information?

Healthcare is always changing and advancing in today's world. Being able to solve and understand what increases the likelihood of having America’s number one killer can help save lives in the future.

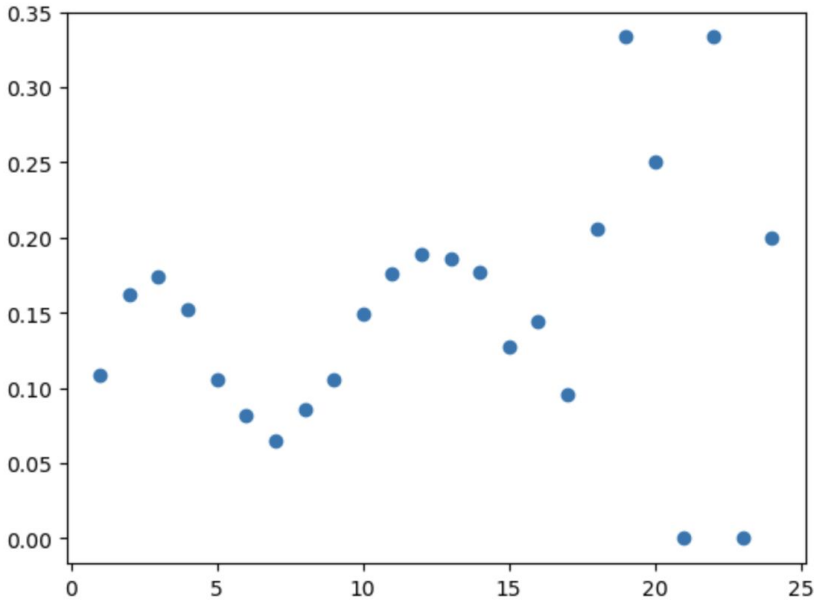
Data Set

	HeartDisease	BMI	Smoking	AlcoholDrinking	Stroke	PhysicalHealth	MentalHealth	DiffWalking	Sex	AgeCategory	Race	Diabetic	PhysicalActivity	GenHealth	SleepTime	Asthma	KidneyDisease	SkinCancer
0	No	16.60	Yes	No	No	3.0	30.0	No	Female	55-59	White	Yes	Yes	Very good	5.0	Yes	No	Yes
1	No	20.34	No	No	Yes	0.0	0.0	No	Female	80 or older	White	No	Yes	Very good	7.0	No	No	No
2	No	26.58	Yes	No	No	20.0	30.0	No	Male	65-69	White	Yes	Yes	Fair	8.0	Yes	No	No
3	No	24.21	No	No	No	0.0	0.0	No	Female	75-79	White	No	No	Good	6.0	No	No	Yes
4	No	23.71	No	No	No	28.0	0.0	Yes	Female	40-44	White	No	Yes	Very good	8.0	No	No	No

Our dataset is taken from Kaggle and consists of 319,795 adults. Per the CDC, about 47% of Americans have at least one of the top medical conditions and life choices that puts people at a higher risk for heart disease.

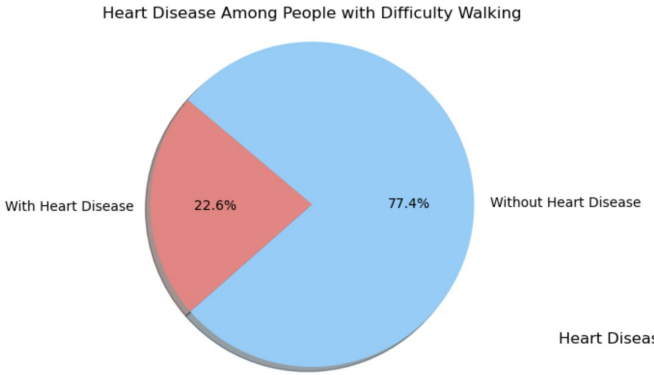
Sleep Time vs. Heart Disease

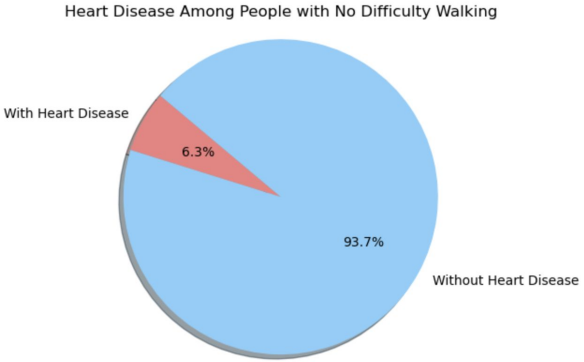
- Graph to the right shows number of hours slept compared to percentage of heart disease
- 7 hours of sleep correlates to 8.6% of people having heart disease
- 12 hours of sleep correlates to 18.6% of people having heart disease
- Shows relation that not enough hours of sleep and too much sleep compared to the average amount can lead to higher heart disease rates



Difficulty Walking vs. Heart Disease

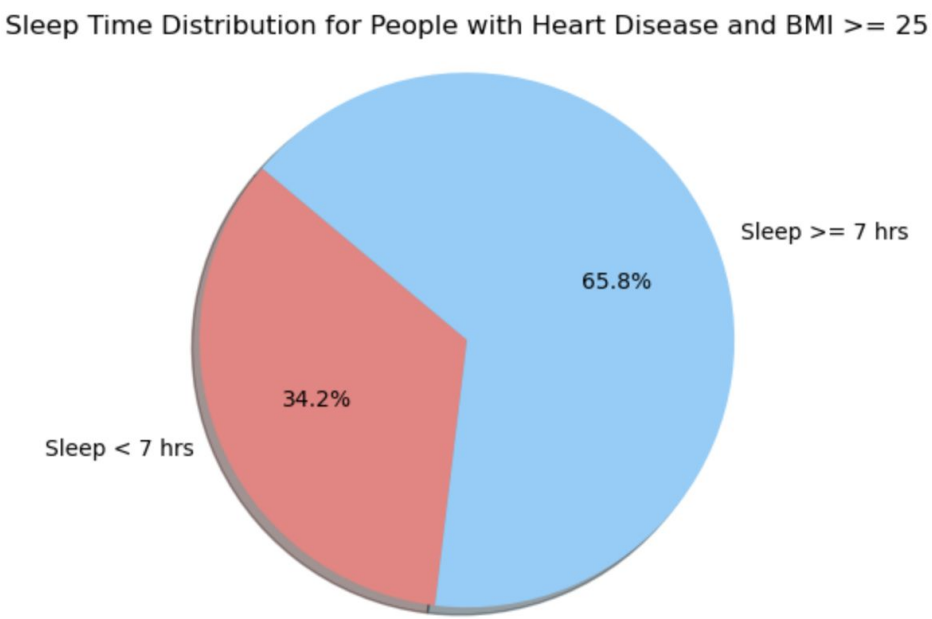
- Shows that there is a strong correlation of difficulty walking and heart disease as 22.6% is more than triple 6.3%





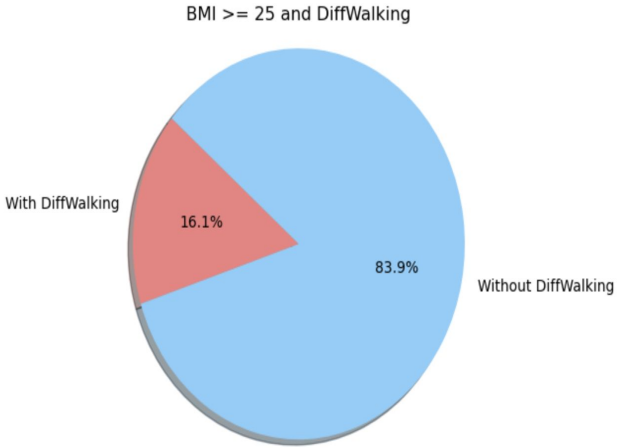
Sleep Time Distribution for People With Heart Disease and BMI >=25

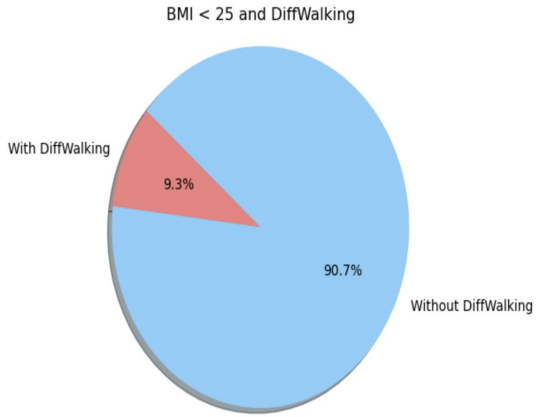
- Compared sleep time and unhealthy BMI to heart disease
- Correlation that the heart disease count of people sleeping over 7 hours with a 25+ BMI is almost double than sleeping under 7 hours



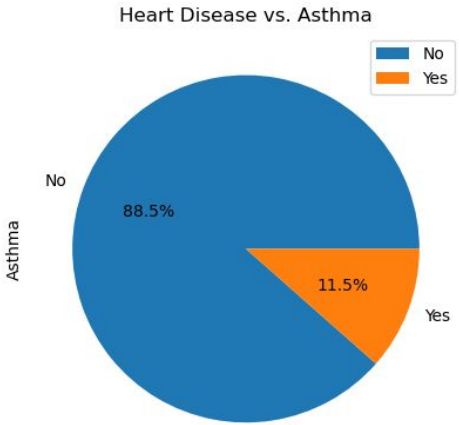
Difficulty Walking and BMI vs. Heart Disease

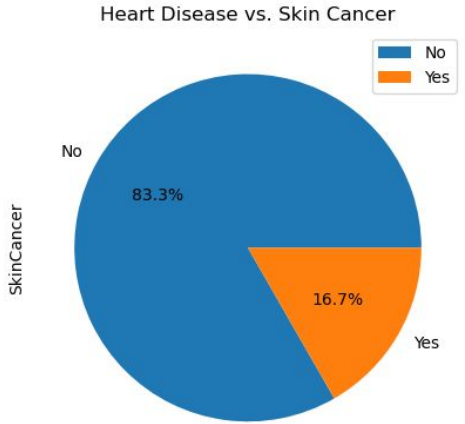
- Charts show that there is enough correlation between BMI being over and under 25 to difficulty walking
- Almost double the amount of people with heart disease with a BMI of 25 or more and difficulty walking





Heart Disease vs. Asthma & Skin Cancer





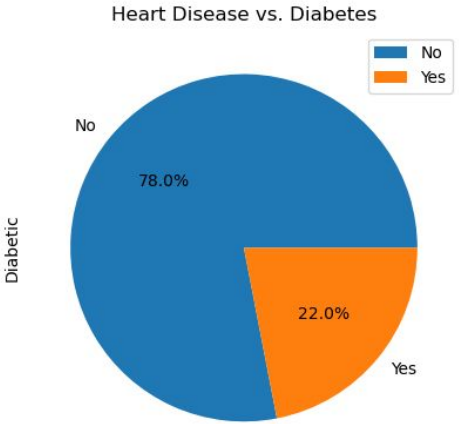
11.5% of people with heart disease also have asthma.

Population (Yes) = 5,000

16.7% of people with heart disease also have skin cancer.

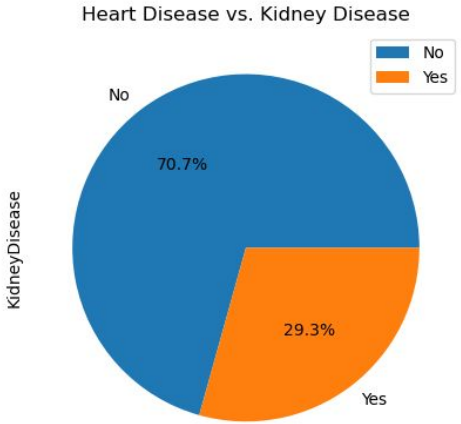
Population (Yes) = 5,000

Heart Disease vs. Diabetes & Kidney Disease



22.0% of people with heart disease also have diabetes.

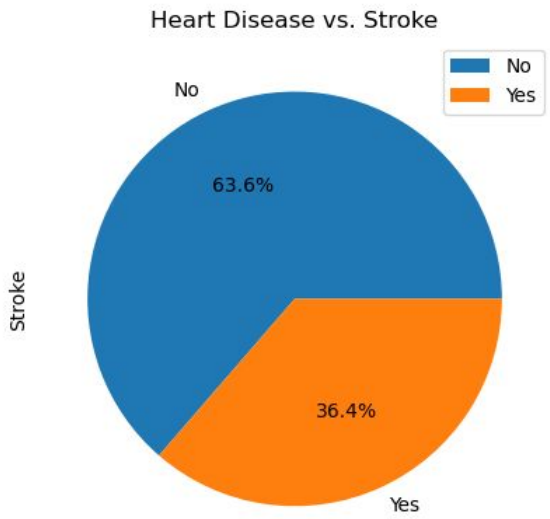
Population (Yes) = 10,000



29.3% of people with heart disease also have kidney disease.

Population (Yes) = 3,500

Heart Disease vs. Stroke



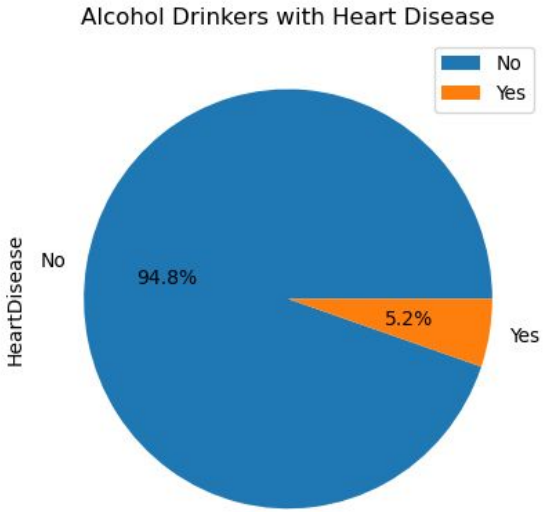
Population (Yes) = 4,500

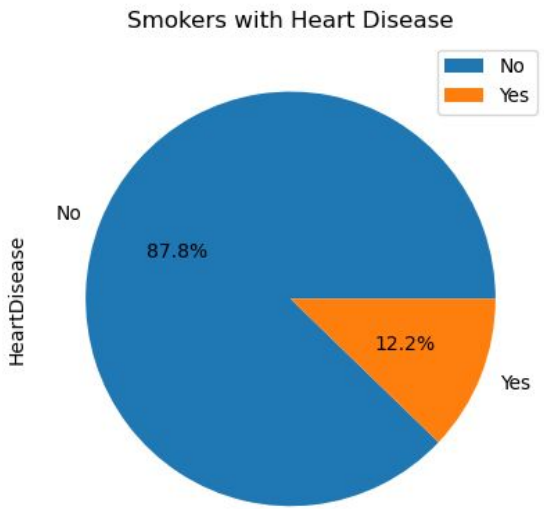
- 36.4% of our population that have suffered a stroke also deals with heart disease.
- Per the AHA, it is common for patients to suffer from a heart condition within a month of having a stroke.

Pre-existing Conditions vs. Heart Disease

- 61% of our population answered no to having heart disease, stroke, diabetes, asthma, kidney disease, and skin cancer.
- 0.00009% of our population answered yes to having all of the conditions listed above.
- Smoking, diet, physical inactivity, excessive alcohol use, overweight, and obesity are all the top contenders that leads to heart disease. We will dive deeper with what that looks like from our dataset.

Heart Disease vs. Alcohol Drinking & Smoking



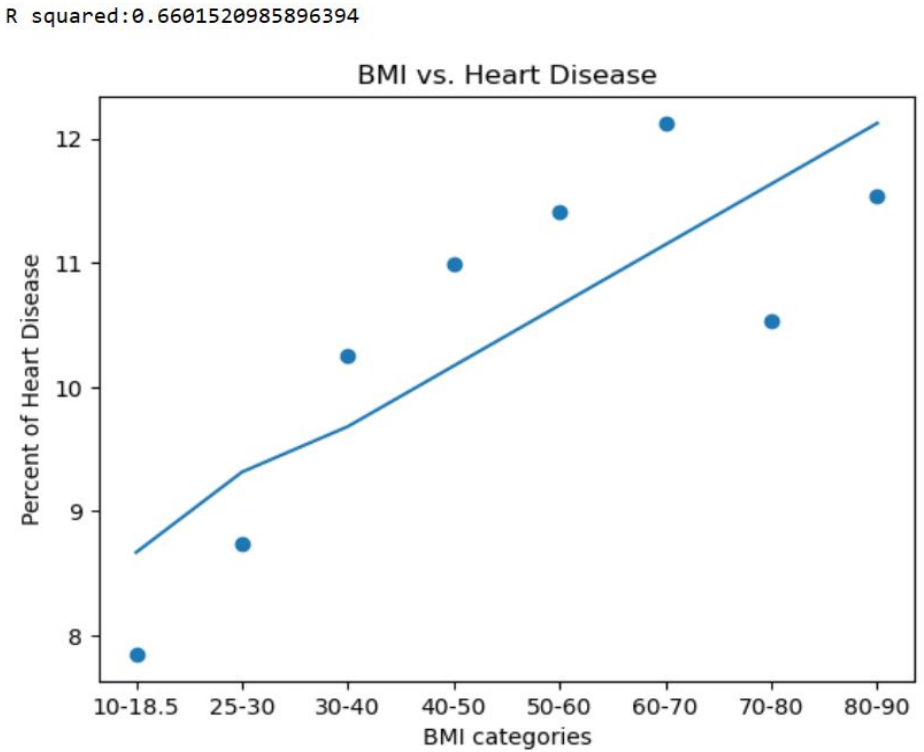


5.2% of alcohol drinkers have heart disease

12.2 % of smokers have heart disease

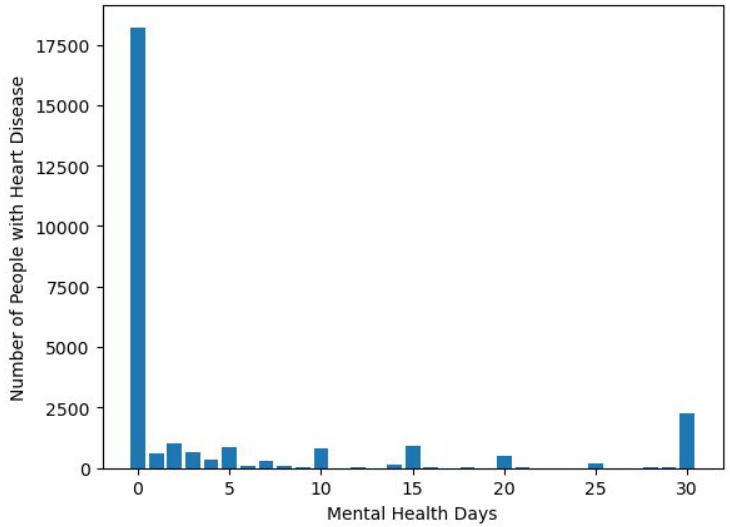
Heart Disease vs. BMI

- This graph shows the prediction of heart disease based on BMI
- R squared value is 0.66
- BMI has a weak correlation to heart disease



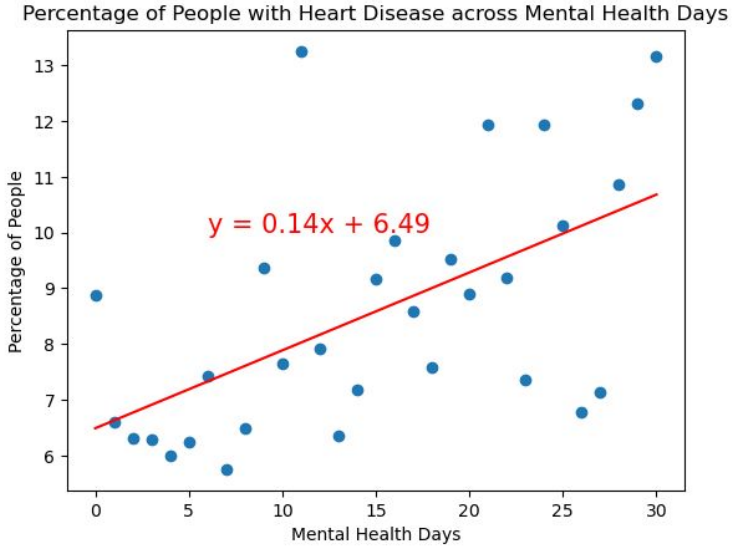
Mental Health vs. Heart Disease

```
Out[6]: Text(0, 0.5, 'Number of People with Heart Disease')
```



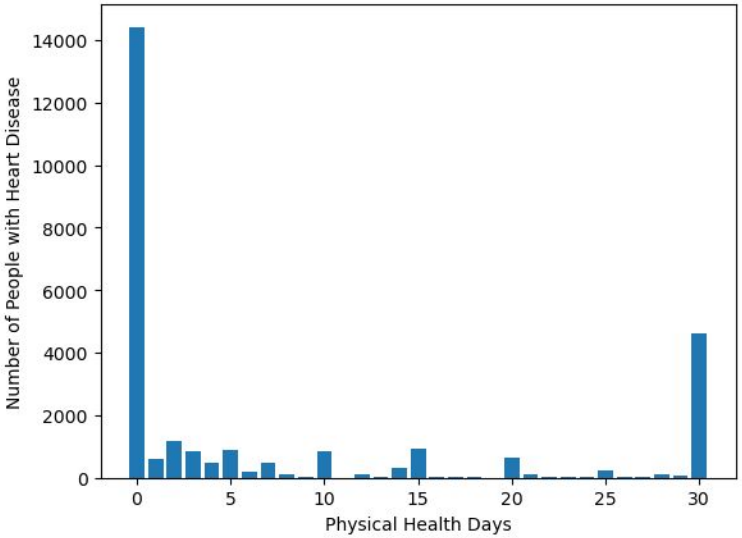
- Mental health days depicts how many days during the **past 30 days was the mental health bad.**
- 0 = No mental health issue in 30 days (Great)
- 30= Bad mental health in 30 days (Worst) .

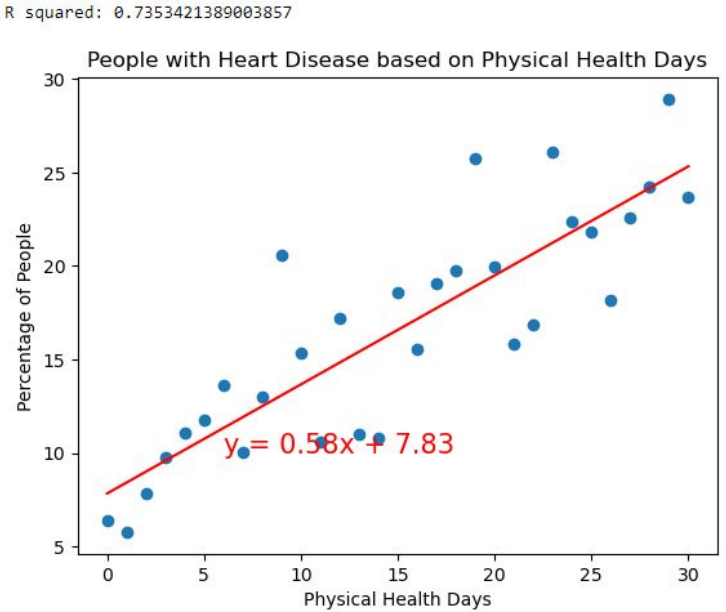
```
R squared: 0.3292861895340285
```



- **Used scatter plot, regression and line of fit** to establish the relationship between percentage of people with heart disease for each physical health day record.
- The **R squared value is 0.329.**
- **Mental health has a weak correlation with heart disease** as R squared value is closer to zero than one.

Physical Health vs. Heart Disease



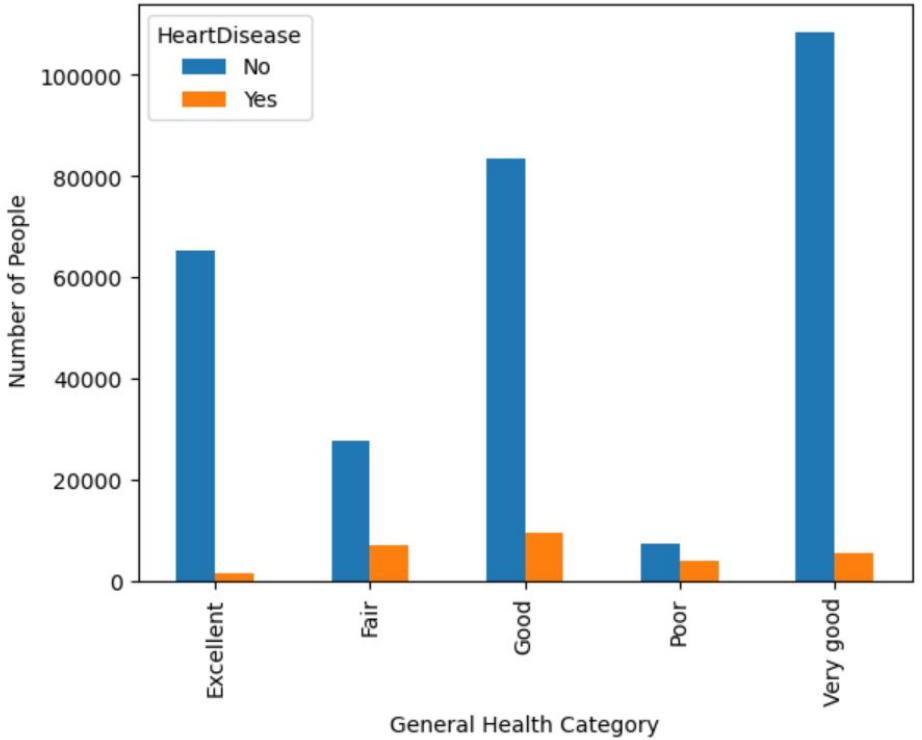


- Physical health days depicts how many days during the **past 30 days was any physical health issue recorded.**
- Physical health issues includes physical illness and injury.
- Used scatter plot, regression and line of fit to establish the relationship between percentage of people with heart disease for each physical health day record.
- The **R squared value is 0.735**

- **Physical health is a better indicator of heart disease compared to mental health** but not very strong as R squared value is not quite close to 1.

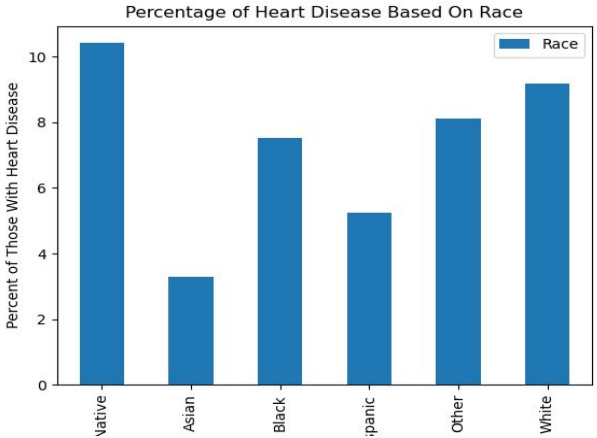
General Health vs. Heart Disease

HeartDisease	No	Yes
GenHealth		
Excellent	65342	1500
Fair	27593	7084
Good	83571	9558
Poor	7439	3850
Very good	108477	5381

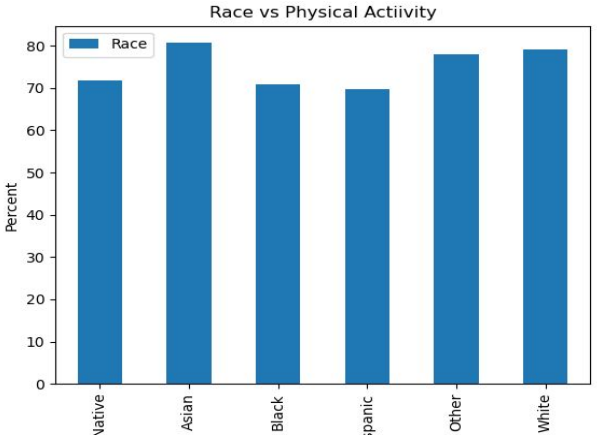


- Excellent health category have very low number of people with heart disease.
- Excellent , very good, good reduces the chance of having heart disease by at least 8X.
- There is about 51.7% chance people in the poor category will have a heart disease.

Race

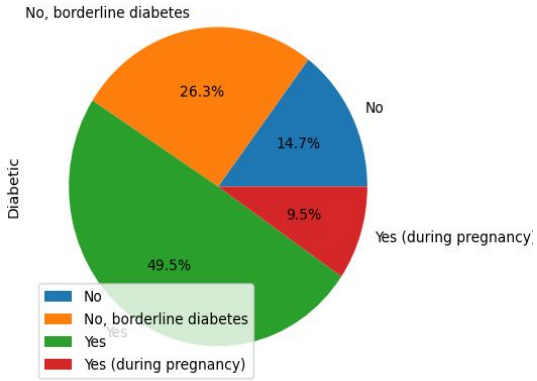


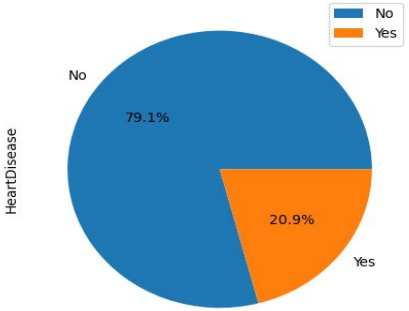
- American Indians/Alaskan Natives have the highest rate of heart disease with 10.42%.
- Asians have the lowest with 3.30%



- Conversely, Asians are the most physically active in the sample with 80.68%.
- Hispanics were the least physically active with 69.83%.

Diabetic Status

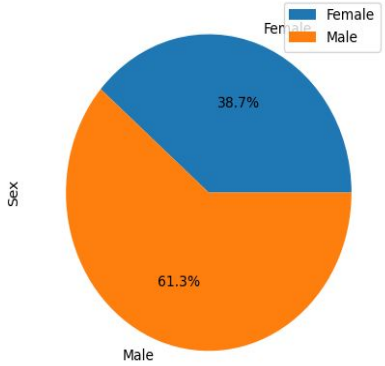


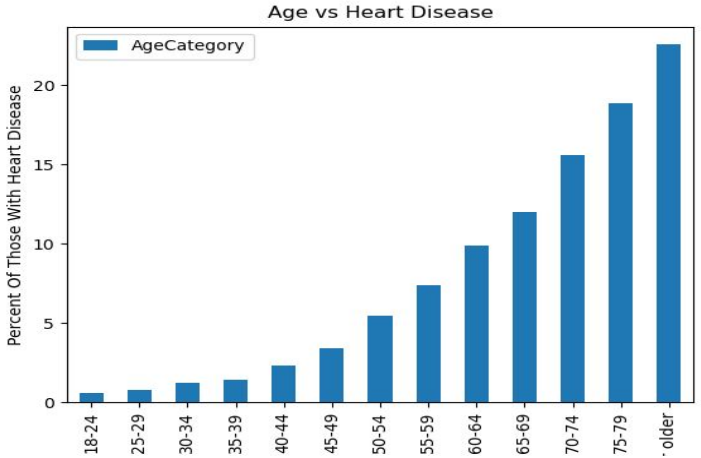


- Diabetic Status is split into 2 main groups (Yes and No)
- However, those groups are also split into subgroups.

- A minority of folks with diabetes (only 20.9%) have heart diseases in this sample.
- However, as we will see shortly, diabetes does increase the chances of heart disease.

Sex and Age





- Even though this sample had more females (167,805) than males (151,990), the males were more affected by heart disease.
- There is a very strong positive correlation between age and percent of odds of contacting a heart disease.

Cross Sectional Data

	Race	Diabetic	PhysicalActivity	OddsHeartDisease
0	White	Yes	Yes	21.089754
1	White	Yes	No	28.147399
2	White	No	Yes	6.015302
3	White	No	No	11.745422
4	White	No, borderline diabetes	Yes	12.041250
5	White	No, borderline diabetes	No	18.001652
6	White	Yes (during pregnancy)	Yes	3.176796
7	White	Yes (during pregnancy)	No	8.641975
8	Black	Yes	Yes	14.477612
9	Black	Yes	No	21.343874
10	Black	No	Yes	4.306110
11	Black	No	No	7.452749
12	Black	No, borderline diabetes	Yes	6.502242
13	Black	No, borderline diabetes	No	7.182320
14	Black	Yes (during pregnancy)	Yes	7.086614
15	Black	Yes (during pregnancy)	No	8.000000
16	Asian	Yes	Yes	11.270125
17	Asian	Yes	No	11.659193
18	Asian	No	Yes	2.317704
19	Asian	No	No	2.981467
20	Asian	No, borderline diabetes	Yes	1.136364
21	Asian	No, borderline diabetes	No	8.000000
22	Asian	Yes (during pregnancy)	Yes	0.000000

- Presence of diabetes and absence of physical activity increased the likelihood of heart disease.
- 18-24 year old American Indian/Alaskan Native males were the least likely to have heart disease with 0% frequency in the data.
- 80 year old or older American Indian/Alaskan Native males were the most likely, with 33.33% or 1 in 3 to have heart disease.

	Race	Sex	AgeCategory	OddsHeartDisease
0	White	Female	55-59	0.054412
1	White	Female	80 or older	0.178739
2	White	Female	65-69	0.076527
3	White	Female	75-79	0.130307
4	White	Female	40-44	0.020217
...	...	...	...	...
151	Hispanic	Male	45-49	0.049223
152	Hispanic	Male	18-24	0.010955
153	Hispanic	Male	35-39	0.023007
154	Hispanic	Male	30-34	0.005556
155	Hispanic	Male	25-29	0.012085

Results and Conclusion

- Strong correlation in variables like difficulty walking and sleep time
- There is an inverse relationship between physical activity and likelihood of heart disease.
- There are some uncontrollable factors such as race, age, and sex that do contribute to greater odds of heart disease
    - Men have higher odds than women
    - The older, the more likely to have heart disease
    - American Indians/Alaskan Natives have the highest likelihood, especially the elderly.

Resources

https://www.cdc.gov/heartdisease/about.htm

https://newsroom.heart.org/news/heart-complications-after-a-stroke-increase-the-risk-of-future-cardiovascular-events#:~:text=A%20large%2C%20observational%20study%20found,another%20stroke%20within%20five%20years