Analyzing Predictors of Heart Disease

By:

Alaa Aldubaie, Saksham Motwani, Jin Zhang

ISYS 812 - Professor Guillaume Faddoul | Fall 2022 | San Francisco State University

Table of Contents

Introduction

Today, data is used to derive valuable insights in all spheres. We wanted to use the power of data to analyze various aspects related to a lethal problem that the world, especially the U.S., is facing: Heart Disease. The name of the dataset is *Personal Key Indicators of Heart Disease* [1]. The data is from the 2020 annual Center for Disease Control survey of 319,796 adults about their health status. The survey is conducted annually by telephone to gather data on the health status of U.S. residents in all 50 states, the District of Columbia, and three U.S. territories.

The motivation for choosing this dataset was its relevance and impact in the real world. Cardiovascular diseases are the leading cause of death globally and are the number one cause of death in the U.S. One in five deaths in the U.S. in 2020 was due to heart disease. In terms of economic impact, from 2017 to 2018, it cost the country $229 billion per year in healthcare services, medicines and lost productivity [2]. Hence, detecting and preventing factors that have the closest connection to heart disease is a critical public health and economic issue. On a personal note, one of the team members of this project has a heart disease condition. The analysis may help them gain insight into their condition and consider possible lifestyle changes to minimize future health risks.

The dataset is structured in 18 columns/variables and 319,796 rows and contains a wide range of data types - numerical values/floating-point numbers, strings, and booleans. There are no missing values, as the authors of the dataset have already selected the most relevant variables from the original dataset of more than 300 variables [3]. A large number of them were of little use as they contained a lot of missing values.

The primary response variable is HeartDisease, defined as either Coronary Artery Disease or Myocardial Infarction. The other 17 variables are explanatory and can be broadly classified as

1. Physical Attributes: BMI, Sex, AgeCategory, and Race
2. Habits: Smoking, AlcoholDrinking, PhysicalActivity, and SleepTime
3. Health conditions: PhysicalHealth, MentalHealth, DiffWalking, Diabetic, GenHealth, Asthma, KidneyDisease, SkinCancer

The table below provides a detailed explanation of all variables in the dataset.

Variable Name	Description	Values
HeartDisease	Respondents that have ever reported having Coronary Heart Disease (CHD) or Myocardial Infarction (MI)	Yes/No
BMI	Body Mass Index (BMI)	Numeric
Smoking	Have you smoked at least 100 cigarettes in your entire life?	Yes/No
AlcoholDrinking	Heavy drinkers. Adult men having more than 14 drinks per week and adult women having more than 7 drinks per week	Yes/No
Stroke	Have you ever had a stroke?	Yes/No
PhysicalHealth	For how many days during the past 30 days was your physical health not good? (This includes physical illness and injury).	0 - 30 (days)
MentalHealth	For how many days during the past 30 days was your mental health not good?	0 - 30 (days)
DiffWalking	Do you have serious difficulty walking or climbing stairs?	Yes/No
Sex	Are you male or female?	Female/Male

AgeCategory	Fourteen-level age categories.	'18-24' '25-29' '30-34' '35-39' '40-44' '45-49' '50-54' '55-59' '60-64' '65-69' '70-74' '75-79' '80 or older'
Race	Ethnicity	'White' 'Black' 'Asian' 'American Indian/Alaskan Native' 'Hispanic' 'Other'
Diabetic	Are you diabetic?	'Yes' 'Yes (during pregnancy)’ 'No' 'No, borderline diabetes'
PhysicalActivity	Adults who reported doing physical activity or exercise during the past 30 days other than their regular job.	Yes/No
GenHealth	Would you say that in general your health is	'Excellent' 'Very good' 'Good' 'Fair' 'Poor'
SleepTime	On average, how many hours of sleep do you get in a 24-hour period?	0 - 24 (hours)
Asthma	Do you have Asthma?	Yes/No
KidneyDisease	Do you have Kidney Disease? This does not include kidney stones, bladder infection, or incontinence	Yes/No
SkinCancer	Do you have Skin Cancer?	Yes/No

Driving Question

The driving question that guided us throughout our analysis was to examine any relationships and correlations between the occurrence of heart disease and the 17 predictor variables that may or may not be contributing factors.

Data Cleaning

The steps we took to check how ready our data is for analysis are as follows:

1. Check for null values in each column: There were no null or missing values in the dataset.

```
# Check if columns have any missing values
print(master_data.isnull().any())
```

```
HeartDisease        False
BMI                 False
Smoking             False
AlcoholDrinking     False
Stroke              False
PhysicalHealth      False
MentalHealth        False
DiffWalking         False
Sex                 False
AgeCategory         False
Race                False
Diabetic            False
PhysicalActivity    False
GenHealth           False
SleepTime           False
Asthma              False
KidneyDisease       False
SkinCancer          False
dtype: bool
```

2. Check for misspelled values in categorical columns: String values in datasets can be misspelled due to human error. We searched for unique values in each categorical column and found no errors in this dataset.

```
# There are no missing values, checking for unusual values in the categorical columns (any misspelled) values
print ('Unique Values of all categorical columns:')
for col in master_data.select_dtypes(include = 'object'):
    print(col,':', master_data[col].unique())
```

```
Unique Values of all categorical columns:
HeartDisease : ['No' 'Yes']
Smoking : ['Yes' 'No']
AlcoholDrinking : ['No' 'Yes']
Stroke : ['No' 'Yes']
DiffWalking : ['No' 'Yes']
Sex : ['Female' 'Male']
AgeCategory : ['55-59' '80 or older' '65-69' '75-79' '40-44' '70-74' '60-64' '50-54'
 '45-49' '18-24' '35-39' '30-34' '25-29']
Race : ['White' 'Black' 'Asian' 'American Indian/Alaskan Native' 'Other'
 'Hispanic']
Diabetic : ['Yes' 'No' 'No, borderline diabetes' 'Yes (during pregnancy)']
PhysicalActivity : ['Yes' 'No']
GenHealth : ['Very good' 'Fair' 'Good' 'Poor' 'Excellent']
Asthma : ['Yes' 'No']
KidneyDisease : ['No' 'Yes']
SkinCancer : ['Yes' 'No']
```

3. Check for unusual values in the numerical columns: We checked for any unusual values in the numerical columns using the describe() method. For example, we found that 30 observations had a SleepTime value of 24 hours, which did not make sense; however, we chose not to remove these observations as they were minimal in number.

```
# Summarize numerical data, checking for unusual values
master_data.describe()
```

	BMI	PhysicalHealth	MentalHealth	SleepTime
count	319795.000000	319795.00000	319795.000000	319795.000000
mean	28.325399	3.37171	3.898366	7.097075
std	6.356100	7.95085	7.955235	1.436007
min	12.020000	0.00000	0.000000	1.000000
25%	24.030000	0.00000	0.000000	6.000000
50%	27.340000	0.00000	0.000000	7.000000
75%	31.420000	2.00000	3.000000	8.000000
max	94.850000	30.00000	30.000000	24.000000

4. One Hot Encoding categorical variables: Since our dataset was already very clean, this was perhaps the only "real" cleaning/transformation we did. We transformed non-numerical values into numerical ones. First, we created a mapping dictionary for columns having (string) binary values. Then, we used the replace() function to replace those (string) values with numeric ones. Further, we used the get_dummies() function from the pandas library to transform the columns that have more than two categories (Age, Race, and Diabetes) into dummy variables, so every category becomes a variable having a value of 0 or 1. Lastly, we renamed some columns for better comprehension to help us answer our analysis questions.

```python
# For features with Yes/No binary values, Sex - Male/Female values, and General Health - having 5 values
# create a mapping dictionary that contains each column to process as well as a dictionary of the values to transform
encoding_dict = {"HeartDisease":      {"No": 0, "Yes": 1},
                 "Smoking":           {"No": 0, "Yes": 1},
                 "AlcoholDrinking":   {"No": 0, "Yes": 1},
                 "Stroke":            {"No": 0, "Yes": 1},
                 "DiffWalking":       {"No": 0, "Yes": 1},
                 "Sex":               {"Female": 0, "Male": 1},
                 "PhysicalActivity":  {"No": 0, "Yes": 1},
                 "Asthma":            {"No": 0, "Yes": 1},
                 "KidneyDisease":     {"No": 0, "Yes": 1},
                 "SkinCancer":        {"No": 0, "Yes": 1},
                 "GenHealth":         {"Poor": 5, "Fair": 4, "Good": 3, "Very good": 2, "Excellent": 1},
}
data = data.replace(encoding_dict)
```

```python
# Using Pandas getdummies function to convert the categorical variables 'AgeCategory' , 'Race', 'Diabetic' into indicator variables
# After that just dropping the original columns, renaming columns for better understanding
data = pd.concat([data, pd.get_dummies(data['AgeCategory'])], axis=1)
data = pd.concat([data, pd.get_dummies(data['Race'])], axis=1)
data = pd.concat([data, pd.get_dummies(data['Diabetic'])], axis=1)
```

```python
data.drop(['AgeCategory', 'Race', 'Diabetic', 'No'], axis = 1, inplace = True)
data.rename (columns= {            '18-24': 'Age: 18-24',
                                   '25-29': 'Age: 25-29',
                                   '30-34': 'Age: 30-34',
                                   '35-39': 'Age: 35-39',
                                   '40-44': 'Age: 40-44',
                                   '45-49': 'Age: 45-49',
                                   '50-54': 'Age: 50-54',
                                   '55-59': 'Age: 55-59',
                                   '60-64': 'Age: 60-64',
                                   '65-69': 'Age: 65-69',
                                   '70-74': 'Age: 70-74',
                                   '75-79': 'Age: 75-79',
                                  '80 or older': 'Age: 80+',
                                  'White': 'Race: White',
                                  'Black': 'Race: Black',
                                  'Asian': 'Race: Asian',
```

Below is the snapshot of the first 10 rows of our dataset after cleaning

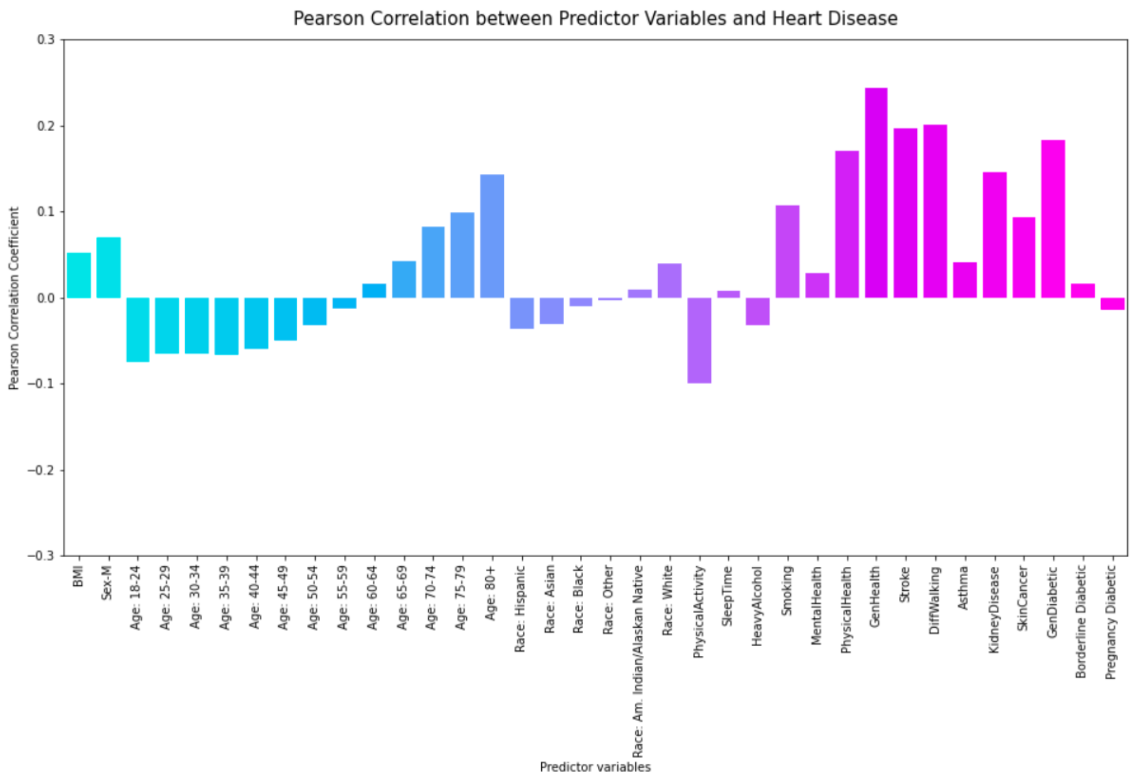
	HeartDisease	BMI	Smoking	HeavyAlcohol	Stroke	PhysicalHealth	MentalHealth	DiffWalking	Sex	PhysicalActivity	...	Age: 80+	Race: Am. Indian/Alaskan Native	Race: Asian	Race: Black	Race: Hispanic	Race: Other	Race: White	GenDiabetic	Borderline Diabetic	Pregnancy Diabetic
0	0	16.60	1	0	0	3.0	30.0	0	0	1	...	0	0	0	0	0	0	1	1	0	0
1	0	20.34	0	0	1	0.0	0.0	0	0	1	...	1	0	0	0	0	0	1	0	0	0
2	0	26.58	1	0	0	20.0	30.0	0	1	1	...	0	0	0	0	0	0	1	1	0	0
3	0	24.21	0	0	0	0.0	0.0	0	0	0	...	0	0	0	0	0	0	1	0	0	0
4	0	23.71	0	0	0	28.0	0.0	1	0	1	...	0	0	0	0	0	0	1	0	0	0
5	1	28.87	1	0	0	6.0	0.0	1	0	0	...	0	0	0	1	0	0	0	0	0	0
6	0	21.63	0	0	0	15.0	0.0	0	0	1	...	0	0	0	0	0	0	1	0	0	0
7	0	31.64	1	0	0	5.0	0.0	1	0	0	...	1	0	0	0	0	0	1	1	0	0
8	0	26.45	0	0	0	0.0	0.0	0	0	0	...	1	0	0	0	0	0	1	0	1	0
9	0	40.69	0	0	0	0.0	0.0	1	1	1	...	0	0	0	0	0	0	1	0	0	0

10 rows × 37 columns

Data Analysis

We formulated five questions to focus our analysis on in line with our driving question.

Question 1: What is the overall correlation across all factors?



The most appropriate way was to calculate the correlation coefficient of all predictor variables (in our cleaned dataset) with the response variable HeartDisease. Below are the findings:

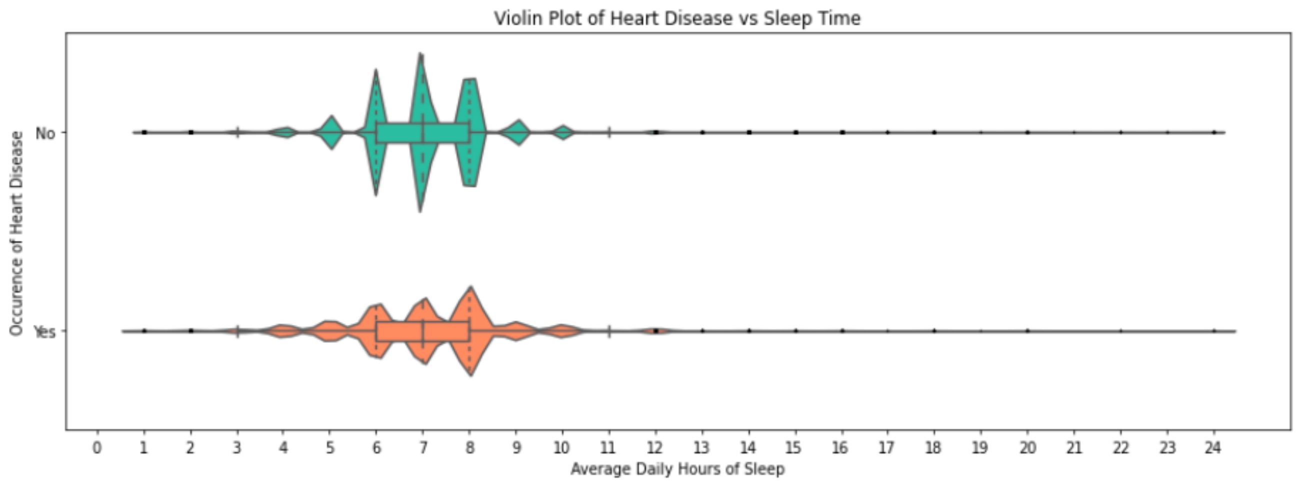
- Age: The correlation between Age and HeartDisease starts negative but then reverts to positive as Age increases (-.08 for Ages 18-24 and 0.14 for Ages 80+).
- Race: The White population has the highest positive correlation with HeartDisease, while the Hispanic population has the highest negative correlation of equal magnitude (0.04).

- Sex: Males have a positive correlation (0.07) with the occurrence of HeartDisease compared to Females.
- BMI: BMI is a value that allows us to assess the degree of correspondence between a person's weight and height and thereby shows whether a person is underweight, normal, overweight, or obese. Hence, a higher BMI value is usually considered unhealthy. Surprisingly, there was just a slight positive correlation (0.05) between BMI and HeartDisease, as we thought it would be higher.
- Stroke, Asthma, KidneyDisease, SkinCancer, and Diabetes: All positively correlate with HeartDisease. However, Stroke has the highest correlation out of all of them, while Asthma has the lowest. Therefore, Stroke and Diabetes are important risk factors for HeartDisease. This is also a medical fact.
- Smoking: The question in the survey was "Have you smoked at least 100 cigarettes in your life?" and it had a positive correlation of 0.11 with HeartDisease.
- Alcohol: Specifically, those who have heavy drinking habits have a lower correlation with HeartDisease (-0.03) than those who do not have HeartDisease, which was very surprising to see.
- PhysicalActivity: The people surveyed were asked a yes or no question asking if they have done any physical activity during the past 30 days other than doing their regular job. Our graph showed a negative correlation with HeartDisease (-0.10) for those who answered Yes. This implies that for people who do physical activity, it is less likely that they have or will get heart disease. This is one of the variables that we have decided to drop from our latter analysis questions because it is recommended by medical

professionals that people with heart disease, depending on the severity of their case, should not do intensive physical activities such as cardio or lifting weights. While for some, it is recommended that they be active to strengthen the heart muscle. That said, we decided to drop PhysicalActivity (in later analysis questions) because the answers were too vague and did not tell us enough about the people in the survey.

- SleepTime: This was a controversial variable we decided to further understand in the second question of the analysis report. The question was, “How many hours on average do you sleep in a 24-hour period?” It showed a slight positive correlation with HeartDisease (0.01).
- PhysicalHealth and MentalHealth: Both variables asked respondents the question: In the past 30 days, how many days have you been feeling physically/mentally unwell? Both showed a positive correlation with Heart Disease (0.17 and 0.03, respectively).
- GenHealth: Respondents rated their health on a Likert scale from poor to excellent, and we mapped this variable accordingly, i.e., 5 - poor, 1 - excellent. This positively correlates with Heart Disease (0.24), which makes sense because higher GenHealth means poorer health.
- DiffWalking: This asked the respondent if they have difficulty climbing stairs. This had a high correlation to Heart Disease (0.20).

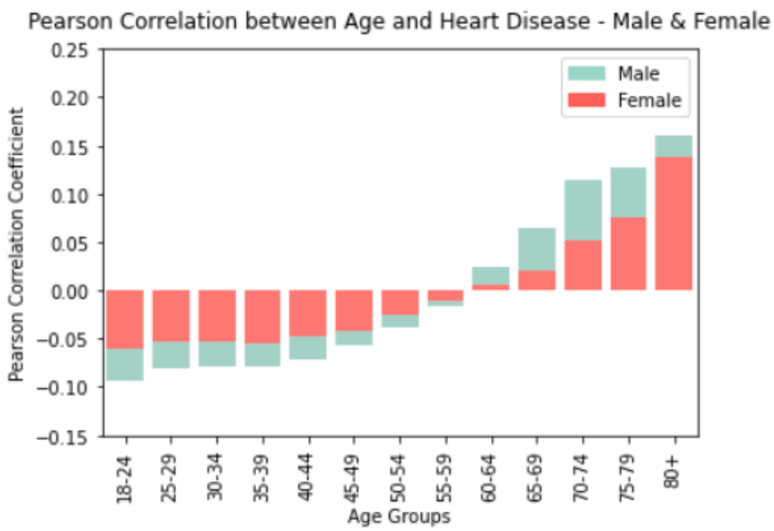
Question 2: Does the distribution of SleepTime for people with HeartDisease and those who don't have HeartDisease vary?



This question was visually answered by creating a Violin plot (A combination of a Kernel Density Estimation plot and a Boxplot). Violin plots are used when we want to observe the distribution of numeric data and are especially useful when comparing distributions between multiple groups. The analysis of the values of people with HeartDisease and without HeartDisease shows a slightly positive trend of people having HeartDisease sleeping for more hours. However, the difference might not be enough to produce an apparent correlation. The amount of SleepTime in a day does not seem to contribute to HeartDisease (like we saw in the previous question); however, as we can see from the plot, the people without HeartDisease sleep for 7 hours. Whereas of the people with HeartDisease, most sleep for 8 hours.

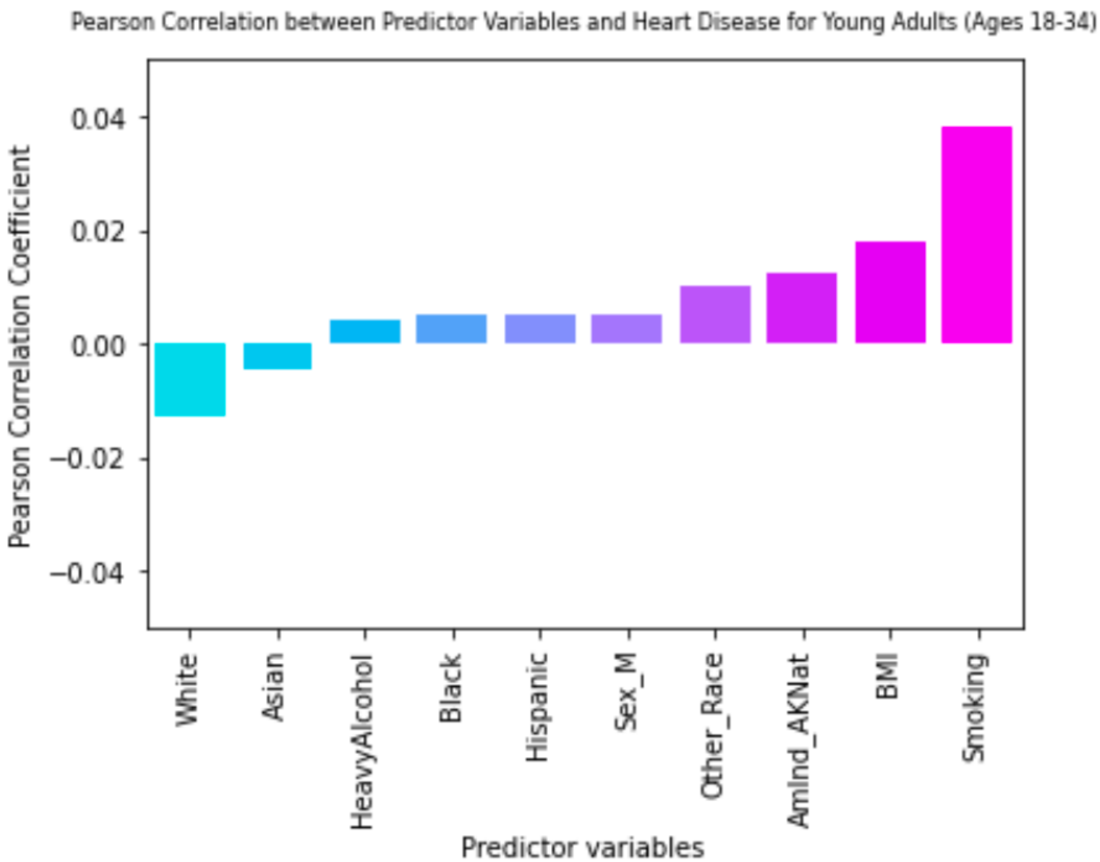
According to the American Heart Association, studies have found that most people need six to eight hours of sleep per day and that too little or too much sleep can increase the risk of heart disease [5]. The plot above shows that people with HeartDisease are more likely to sleep 12 hours or more in comparison to people who don't have HeartDisease. It also depicts that more people with HeartDisease tend to sleep 4 hours or less compared to people without HeartDisease. Thus, abnormal sleep duration is more prevalent in patients with heart disease.

<u>Question 3: What is the extent to which Age explains the occurrence of HeartDisease for men versus women?</u>



To properly know the effect of Age on HeartDisease in the male subset versus the female subset, we have eliminated the other variables and focused on Sex and Age. We found that across all age groups, there is a larger correlation in magnitude between men and HeartDisease than between women and HeartDisease. Additionally, men experience the highest correlation jump between ages 65-69 and 70-74, while women experience the highest correlation jump between ages 75-79 and 80+. Lastly, the inflection point at which Age becomes positively correlated with the occurrence of HeartDisease is the same for both men and women, between ages 55-59 and 60-64. However, according to research, cardiovascular disease develops 7 to 10 years later in women than in men [6].

<u>Question 4: What are the highest risk factors for HeartDisease amongst the 18 - 34 age group?</u>



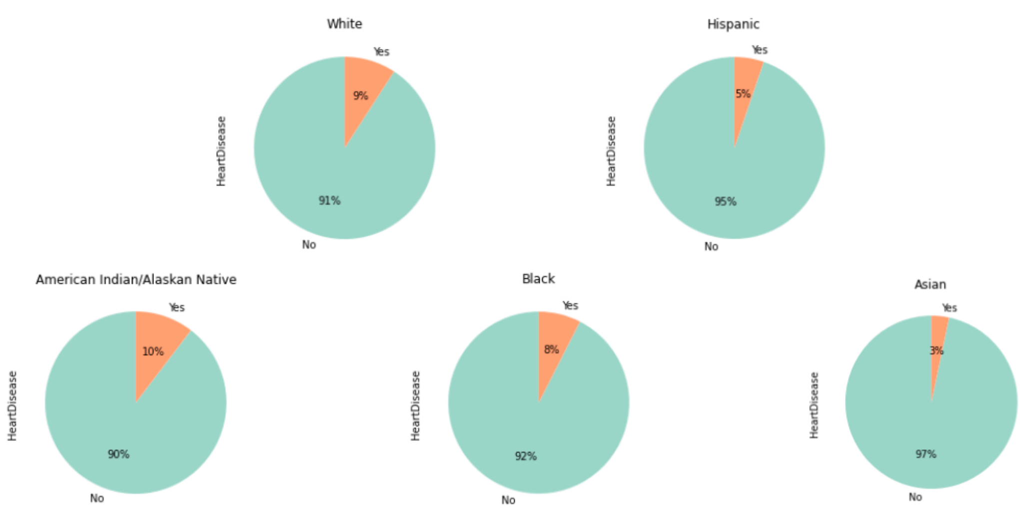
Analyzing the age group under the "Young Adults" category seemed interesting. We were curious to know what factors to HeartDisease come into play and how to minimize or eliminate bad habits for better health. We chose to study only the following explanatory variables for this analysis: Race, AlcoholDrinking (name changed to HeavyAlcohol), Smoking, Sex, and BMI.

After observing the variables above in our data for individuals aged 18 - 34, we have concluded that Smoking has the highest positive correlation (0.038). However, compared to all age groups (0.11), the correlation is much lower, with a difference of 97%.

As for Race and our selected age group, the correlation with the occurrence of Heart Disease was so small that you could barely think there was a correlation. Especially among the Black, Hispanic and Asian populations, with around 0.004. It was interesting to see that the White population has the highest negative correlation with HeartDisease, opposite of what we observed from the overall data covering all ages. In contrast, the American Indian/Alaskan Native population had the highest positive correlation, of equal magnitude (0.013).

Lastly, Sex has a much smaller correlation with HeartDisease in this age group than the population data. Going through all these variables assures us that HeartDisease amongst this young adult age group mostly comes from Smoking more so than attributes we analyzed.

<u>Question 5: What is the percentage of people with HeartDisease versus those without within each Race from the dataset?</u>



Race does not contribute to HeartDisease. It's not a cause, and people of any race could get HeartDisease depending on other factors. However, we have wanted to observe the pattern across all races with HeartDisease. First, we saw how many people are diagnosed with HeartDisease, grouped by Race.

```
# First, let us see how many people are diagnosed with heart disease, grouped by race
master_data.groupby(['Race', 'HeartDisease']).HeartDisease.count()
```

```
Race                           HeartDisease
American Indian/Alaskan Native  No              4660
                                Yes              542
Asian                           No              7802
                                Yes              266
Black                           No             21210
                                Yes             1729
Hispanic                        No             26003
                                Yes             1443
Other                           No             10042
                                Yes              886
White                           No            222705
                                Yes            22507
Name: HeartDisease, dtype: int64
```

We then depicted these numbers as percentages. The pie charts above show the percentages of people with HeartDisease and without for all races (except the Other's race category). In the White population, 9% of the population reported having HeartDisease, while 91% of them do not have HeartDisease. In the Hispanic population, 5% of them reported having HeartDisease, while 95% of them do not have HeartDisease. In the American Indian/Alaskan Native population, 10% of them reported having HeartDisease, while 90% of them do not have HeartDisease. In the Black population, 8% of them reported having HeartDisease, while 92% of them do not have HeartDisease. Lastly, In the Asian population, 3% of their population reported having HeartDisease, while 97% of them do not have HeartDisease. According to the findings, American Indian/Alaskan Native have the highest percentage of people with HeartDisease. Asians have the lowest.

Data Modeling

As a continuation of our analysis in question 4, we wanted to establish a model that could predict the probability of a young adult (aged 18 - 34) having HeartDisease or not based on the predictor variables we chose for question 4.

Our response variable was discrete and contained binary values (Yes/No or 1/0) for HeartDisease or not. Linear regression was not appropriate for this type of response variable, so we used logistic regression. Below is a snapshot of the code:

```
[ ] import statsmodels.formula.api as smf
    model = smf.logit (' HeartDisease ~ HeavyAlcohol + Sex_M + BMI + Smoking + White + Asian + Hispanic + Black + AmInd_AKNat', data = ya)
```

```
results = model.fit()
print(results.summary())
```

```
Optimization terminated successfully.
         Current function value: 0.048598
         Iterations 9
                           Logit Regression Results
==============================================================================
Dep. Variable:           HeartDisease   No. Observations:                56772
Model:                          Logit   Df Residuals:                    56762
Method:                           MLE   Df Model:                            9
Date:                Wed, 30 Nov 2022   Pseudo R-squ.:                 0.01878
Time:                        22:21:00   Log-Likelihood:                -2759.0
converged:                       True   LL-Null:                       -2811.8
Covariance Type:            nonrobust   LLR p-value:                 1.141e-18
================================================================================
                   coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------
Intercept       -5.3016      0.251    -21.086      0.000      -5.794      -4.809
HeavyAlcohol    -0.0447      0.158     -0.283      0.777      -0.354       0.265
Sex_M            0.0716      0.092      0.774      0.439      -0.110       0.253
BMI              0.0205      0.006      3.314      0.001       0.008       0.033
Smoking          0.7998      0.094      8.499      0.000       0.615       0.984
White           -0.4693      0.174     -2.695      0.007      -0.811      -0.128
Asian           -0.3981      0.284     -1.401      0.161      -0.955       0.159
Hispanic        -0.1787      0.197     -0.909      0.363      -0.564       0.206
Black           -0.1127      0.223     -0.506      0.613      -0.549       0.324
AmInd_AKNat      0.2097      0.289      0.725      0.468      -0.357       0.776
================================================================================
```

We are observing the likelihood of the outcome variable to be 1 (HeartDisease occurrence = Yes). Below are the interpretations of the coefficients:

- A young adult who identifies as a heavy drinker (coefficient = -0.04) is less likely than a young adult who identifies as not a heavy drinker to have heart disease.
- A young adult who identifies as male (coefficient = 0.07) is more likely than a young adult who identifies as female to have heart disease.

- An increase in a young adult's BMI (coefficient = 0.02) increases the chances of having heart disease.
- A young adult who identifies as a smoker (coefficient = 0.80) is more likely than a young adult who identifies as a non-smoker to have heart disease.
- A young adult who identifies as White (coefficient = -0.5) is less likely than a young adult who does not identify as White to have heart disease.
- A young adult who identifies as an Asian (coefficient = -0.40) is less likely than a young adult who does not identify as Asian to have heart disease.
- A young adult who identifies as a Hispanic (coefficient = -0.18) is less likely than a young adult who does not identify as Hispanic to have heart disease.
- A young adult who identifies as a Black (coefficient = -0.11) is less likely than a young adult who does not identify as Black to have heart disease.
- A young adult who identifies as an American Indian/Alaskan Native (coefficient = 0.21) is more likely than a young adult who does not identify as American Indian/Alaskan Native to have heart disease.

The probability (of having or have had heart disease), P, can be calculated based on the coefficients:

P = 1/(1+e^(-x)) where:

x = -0.04 *HeavyAlcohol + 0.07 *Sex_M + 0.02 *BMI + 0.80 *Smoking - 0.47 *White - 0.40 *Asian - 0.18 *Hispanic - 0.11 *Black + 0.21 *AmInd_AKNat - 5.30

Conclusion

Benefits we gained from the project:

- Technical Learnings:
    - In general - working with Python and pandas dataframes for data analysis.
    - Data Cleaning: Checking missing data, converting categorical variables into numeric.
    - Data Manipulation: Rename, drop columns in dataframes and correlation matrices in pandas.
    - Plotting: Using matplotlib and seaborn for barplots, boxplots, and violin plots. Using pandas plots to plot pie charts.
    - Data Modeling: Using statsmodels to perform logistic regression
- Analytical Learnings:
    - What to analyze: The thought process of coming up with a driving question and sub-questions that will ultimately help in answering the driving question.

Challenges we faced:

- The data was collected via a telephone survey. Some questions asked were vague and did not truly reveal what they intended. For instance, it would have been better if the questions related to Smoking and PhysicalActivity were framed as “Are you a regular smoker?” and “Do you often do some physical exercise?”. Hence, post question 1, we had to remove some predictors that didn’t make sense to us.
- The code to plot the correlation plots could have been inside a function since it was used repeatedly, but doing this was somewhat complex, and hence was left.

References

[1] "Personal Key Indicators of Heart Disease." *Kaggle*, 16 Feb. 2022, https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease.

[2] "Heart Disease Facts." *Centers for Disease Control and Prevention*, 14 Oct. 2022, https://www.cdc.gov/heartdisease/facts.htm.

[3] "CDC - 2020 BRFSS Survey Data and Documentation." *Centers for Disease Control and Prevention*, 27 Oct. 2022, https://www.cdc.gov/brfss/annual_data/annual_2020.html.

[4] "Heart Disease and Stroke." *Centers for Disease Control and Prevention*, 8 Sept. 2022, https://www.cdc.gov/chronicdisease/resources/publications/factsheets/heart-disease-stroke.htm.

[5] "Sleep, Women and Heart Disease." *Www.heart.org*, 2 June 2022, https://www.heart.org/en/healthy-living/go-red-get-fit/sleep-women-and-heart-disease.

[6] Maas, A.H.E.M., and Y.E.A. Appelman. "Gender Differences in Coronary Heart Disease." *Netherlands Heart Journal: Monthly Journal of the Netherlands Society of Cardiology and the Netherlands Heart Foundation*, U.S. National Library of Medicine, https://pubmed.ncbi.nlm.nih.gov/21301622/.