CardioData Insights:

A Heart Disease Risk Visualization

Team Name: BeatBytes
Team Members: Isaac Ferry, Josh Melchor-Nieto, and Jamie Almeida

Team Introduction

Isaac: Lineplot, Correlation HeatMap, and a Radar Chart

Josh: Bar Charts, Box plot, and Interactive 3D Surface Plot

Jamie: Violin Chart, Interactive 3D Scatterplot, and a Pie Chart

Why We Chose This Topic

In the USA its the leading cause of death for men, women, and people of most racial and ethnic groups.
Every 33 seconds someone dies of cardiovascular disease

In 2022 1 and 5 people died from Coronary artery disease (most common type of heart disease)

Source: CDC

(also because it would look good on a resume)

Number of deaths(%) in 2021 (USA)

Race or Ethnic Group	# of Deaths
American Indian or Alaska Native	15.5
Asian	18.6
Black (Non-Hispanic)	22.6
Native Hawaiian / Pacific Islander	18.3
White (Non-Hispanic)	18.0
Hispanic	11.9
All	17.4

Project Overview

- **Objectives:**
  - Identify key risk factors for heart disease.
  - Validate clinical knowledge (e.g., BMI vs. blood pressure).
- **Significance:**
  - Assists in preventive healthcare by highlighting modifiable risks (exercise, smoking, and blood pressure).
- **Dataset:**
  - Source: Kaggle (Oktay Ördekçi), updated 5 months ago.
  - Metrics: Age, gender, cholesterol, exercise, family history, etc.

Implementation Method

Tools and Libraries:

- Jupyter Notebook
- Pandas
- Matplotlib
- Seaborn
- Numpy

Workflow:

- Research different visualizations
- Import dataset using pandas
- Clean data using dropna
- Used categorical variables or calculated metrics
- Selected visualization chart

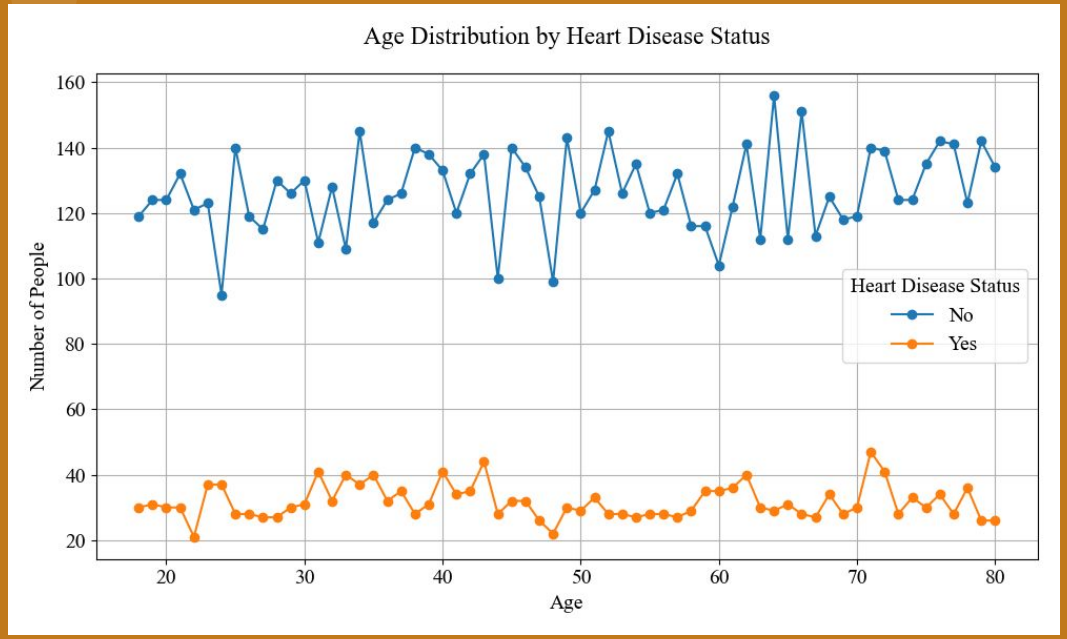
Visualization Techniques

- Line Plot
- Correlation Heatmap
- Radar Chart

Justification:

- Line Plot: Chosen to show the amount of people with and without Heart Disease in the dataset
- Correlation Heatmap: Chosen to show the correlation between different variables and their relation to Heart Disease
- Radar Chart: Chosen to show a easy to read view of the average levels of different variables of people with Heart Disease

Visualization 1: Age Distribution by Heart Disease



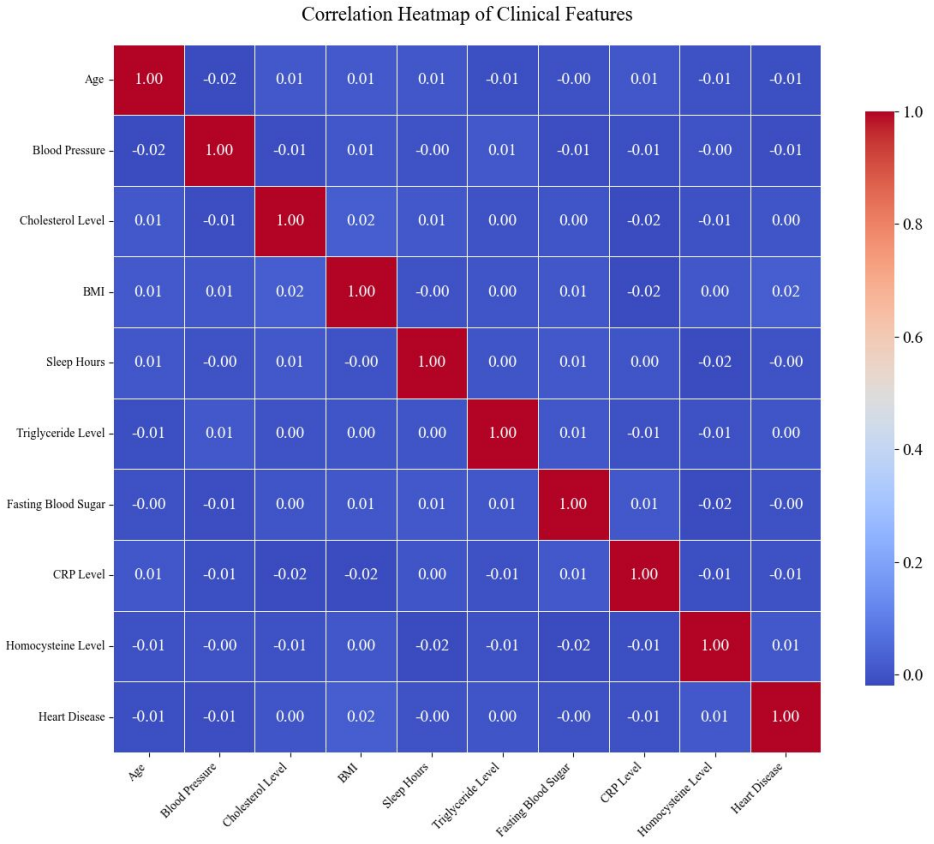
Observations:

- This dataset has many more people without heart disease
- Heart disease line (red) surges sharply in the early 70s Notable "bumps" in heart disease frequency at **30s and 40s**—likely tied to lifestyle or genetic risks
- From mid 70s to 80s heart disease average is less than ages 30-40

Why This Matters:

- Challenges the idea that only the elderly are affected by heart disease

Visualization 2: Correlation of Clinical Features



Observations:

- Little to no correlation with heart disease or each other
- This implies that not a single value is a strong contributor to heart disease

Why This Matters:

- This data suggests that heart disease is caused by multiple variable.
- It can indicate that the data is artificially created.

Visualization 3: Health Metrics by Heart Disease Status

Observations:

- Individuals with heart disease exhibit significantly elevated triglyceride and cholesterol levels.
- Fasting blood sugar levels are moderately high, suggesting a link to prediabetes or diabetes.

Why This Matters:

- Identifying these clustered risk factors can help target early interventions to help reduce heart disease

Implementation Method

Tools and Libraries:

- Jupyter Notebook
- Pandas
- Matplotlib
- Plotly
- Seaborn

Visualization Techniques

- Violin Plot
- Interactive 3D Scatterplot
- Pie Chart

Workflow:

- Import dataset using pandas
- Handle missing values with “dropna()’
- Used categorical variables or calculated metrics
- Selected visualization chart

Justification:

- Violin Plot: Chosen to show distribution density and outliers
- Interactive 3D Scatterplot: Reveals multidimensional correlations
- Pie Chart: Simplified comparison of prevalence rates

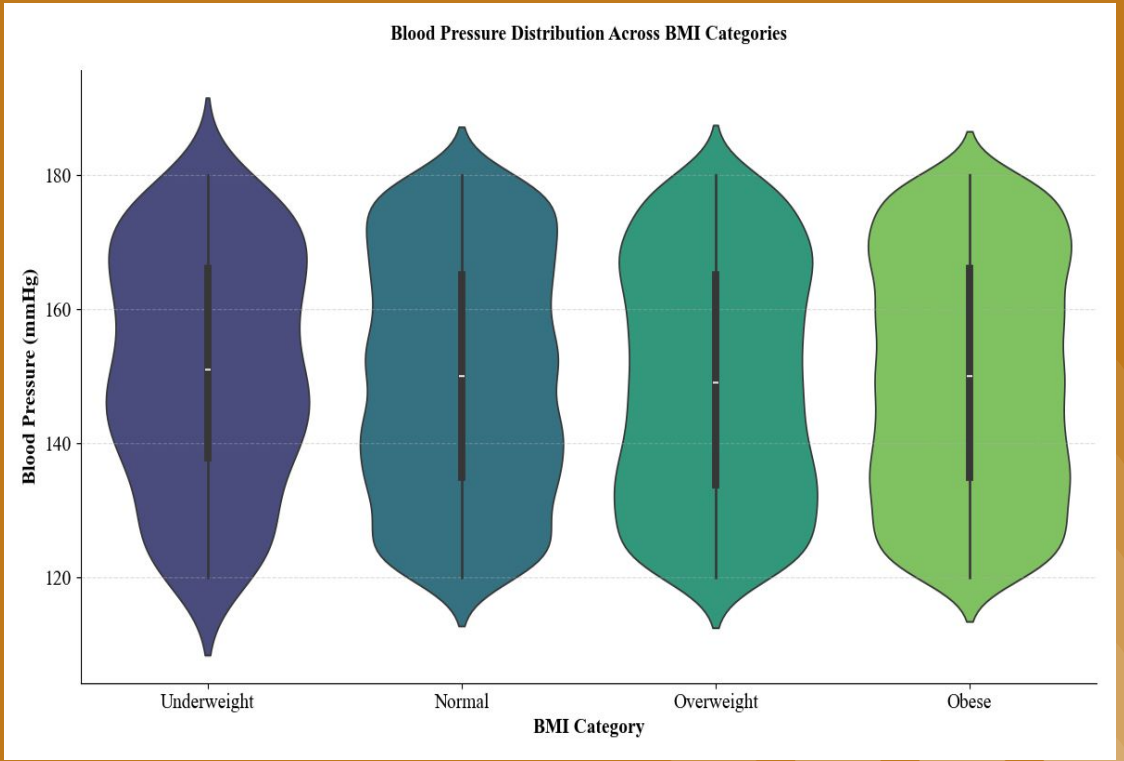
Visualization 4: Blood Pressure Distribution

Observations:

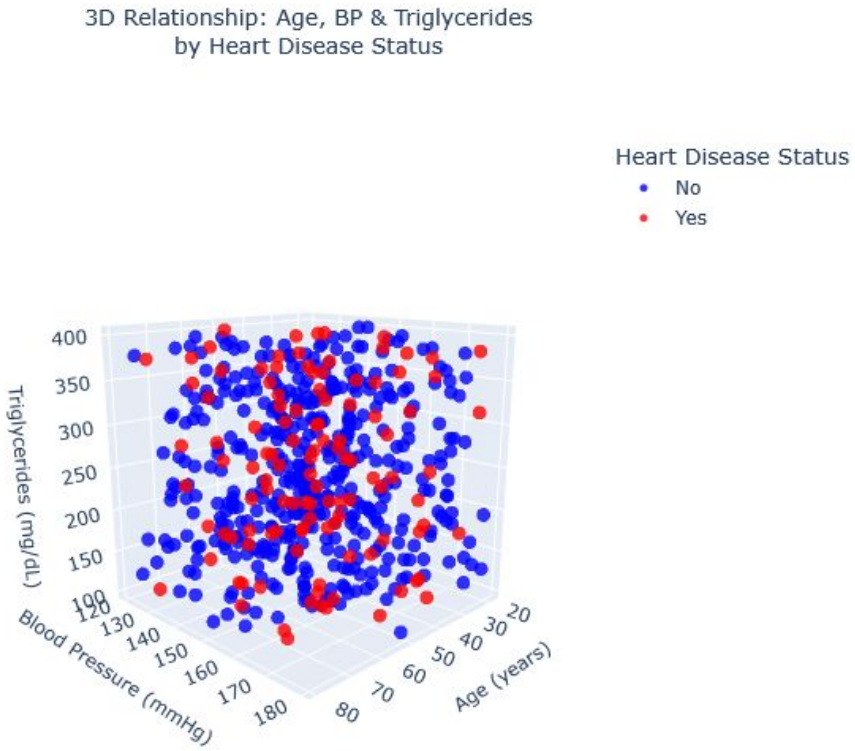
- Unexpected median trend (underweight and obese share a close blood pressure value)
- Overweight group has the largest IQR while the Underweight group has the smallest

Why This Matters:

- Public Health: Reinforces obesity prevention
- Clinical Insight: Confirms the link between obesity and hypertension



Visualization 5: Cardiovascular Risk Factors



Observations:

- Red dots were not clustered in expected areas
- Blue dots dominate lower ranges but appear at higher values.

Why This Matters:

- **Research Validation:** Challenges assumptions that risk factors act in isolation
- **Clinical Utility:** Helps health professionals identify high-risk profiles (elderly patients with rising triglyceride levels).

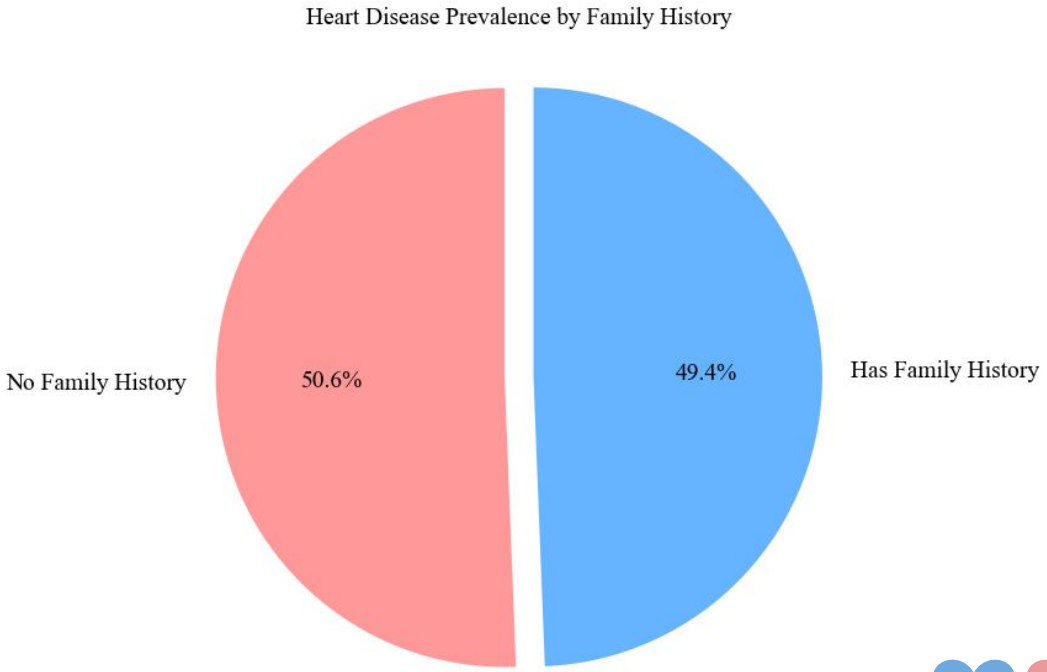
Visualization 6: Family Medical History of Heart Disease

Observations:

- Near equal split meaning that genetics are a risk factor
- Lifestyle choices in “No Family History” group are greatly offsetting genetic advantage.

Why This Matters:

- **Research Validation:** Reinforces the need for universal screening
- **Clinical Utility:** Highlights the importance of personalized prevention



Implementation Method

Tools and Libraries:

- Jupyter Notebook
- Pandas
- Matplotlib
- Seaborn
- Plotly
- Numpy

Visualization Techniques

- Clustered Bar Chart
- Boxplot
- Interactive 3D Surface plot

Workflow:

- Import dataset using pandas.
- Gather group only necessary information and filter out unwanted variables.
- Find averages based on different groups.
- Selected visualization chart

Justification:

- **Clustered Bar Plot:** allows to plot multiple grouped variables as one.
- **Box Plot**: shows averages for different variables to see where majorities lie.
- **Interactive 3D Surface Plot:** Reveals multi variable patterns while reducing noise

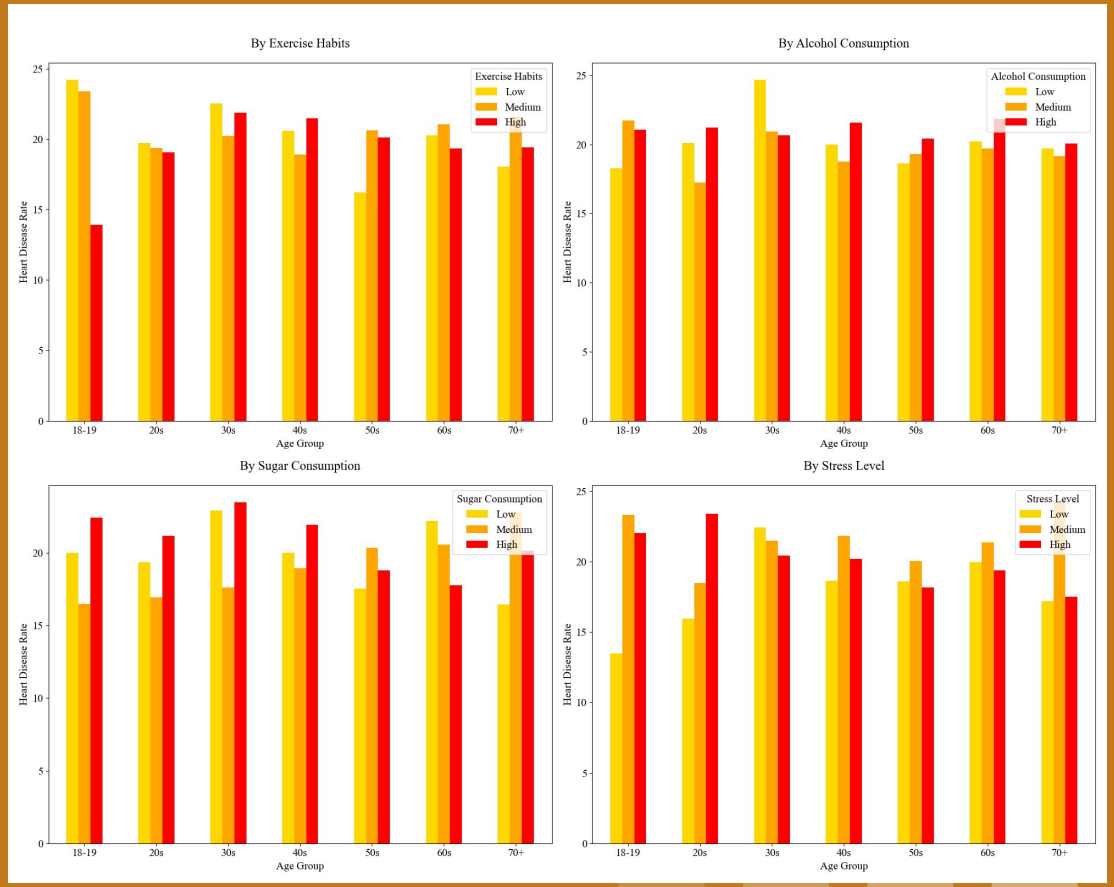
Visualization 7: Heart Disease Likelihood Across Age and Lifestyle Habits

Observations:

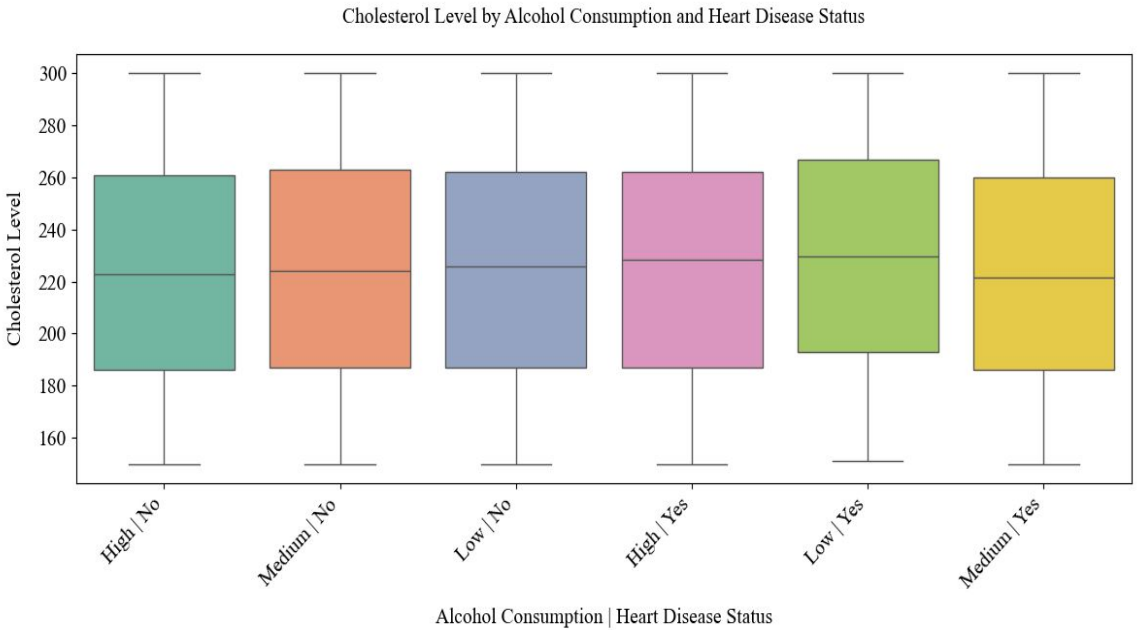
- Good habits overall reduces the risk of heart disease.
- Sugar consumption has less impact than the other habits.
- Stress impacts younger people the most.

Why This Matters:

- Visual proof that lifestyle habits play a role in overall health
- Enforces claims about healthy dieting and exercise.



Visualization 8: Cholesterol Variability by Alcohol Intake



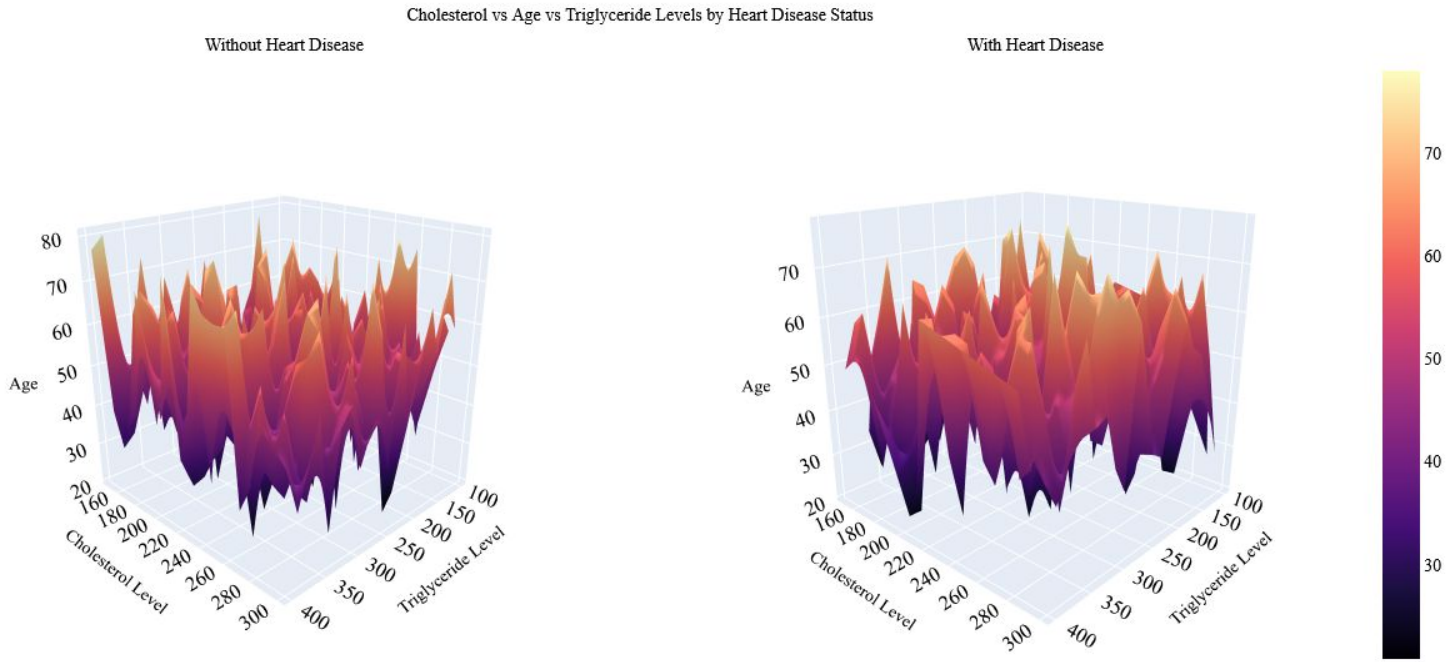
Observations:

- The average cholesterol level for drinking across the board is about 220.
- There may be other contributing factors to cholesterol, or other variables that contribute to cholesterol more than alcohol.

Why This Matters:

- **Public Health:** Alcohol is a major contributing factor to heart disease.
- **Clinical Action:** Identifies medium alcohol intake as highest-risk zone for cholesterol spikes in at-risk patients.

Visualization 9: Health Factors by Heart Disease Status



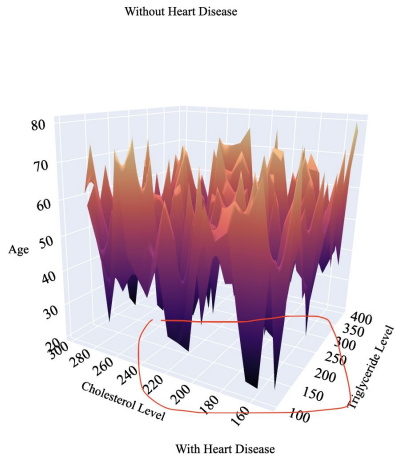
Cont. of Visualization 9

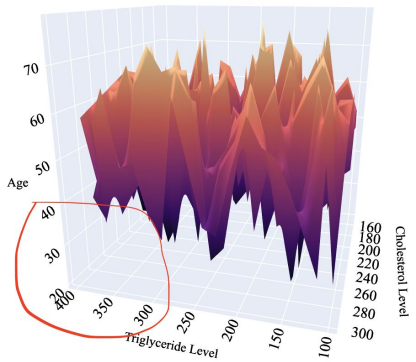
Observations:

- **No Heart Disease:** Most people below the age of 30 have lower triglyceride and cholesterol levels
- **With Heart Disease:** Only those between the ages 40-70 year old have the highest levels of cholesterol and triglyceride

Why This Matters:

- **Public Health:** Visualizes the "metabolic storm" period (40s-50s) needing targeted education.
- **Clinical Strategy:** Helps with both prevention and treatment as individuals should watch their triglyceride and cholesterol levels and those between 40-50s need to have a carefully planned diet.





Key Findings & Conclusion

1. **Validated Clinical Knowledge**
   a. Obesity correlates with hypertension
   b. Increased risk as an individual ages
2. **Valuable Insights**
   a. Habits such as poor exercise and high sugar diets increase an individual's risk
   b. Medium to High alcohol consumption worsens an individual's cholesterol
3. **Impact**
   a. Health experts need to advocate for regular screenings
   b. Ages of 20-30 can avoid worsening health through lifestyle choices
   c. Doctors should investigate each and every anomaly they come across

Questions?

Thank you!

Stay Healthy!