Group 9

1) Lorna Wanjiku
2) Dennis Wainaina
3) Mark Nyumba
4) Valary Thairu
5) Wendy Wairimu
6) Peris Wamoni

USING MACHINE LEARNING ALGORITHMS TO IMPROVE THE EFFICIENCY OF CARDIOVASCULAR DISEASE DIAGNOSIS

INDEX

Introduction

- Cardiovascular disease (CVD) presents a significant public health challenge, impacting millions of people resulting in over 17 million deaths annually. Early detection of CVD is critical however, conventional diagnostic methods for CVD are prone to errors and time consuming.

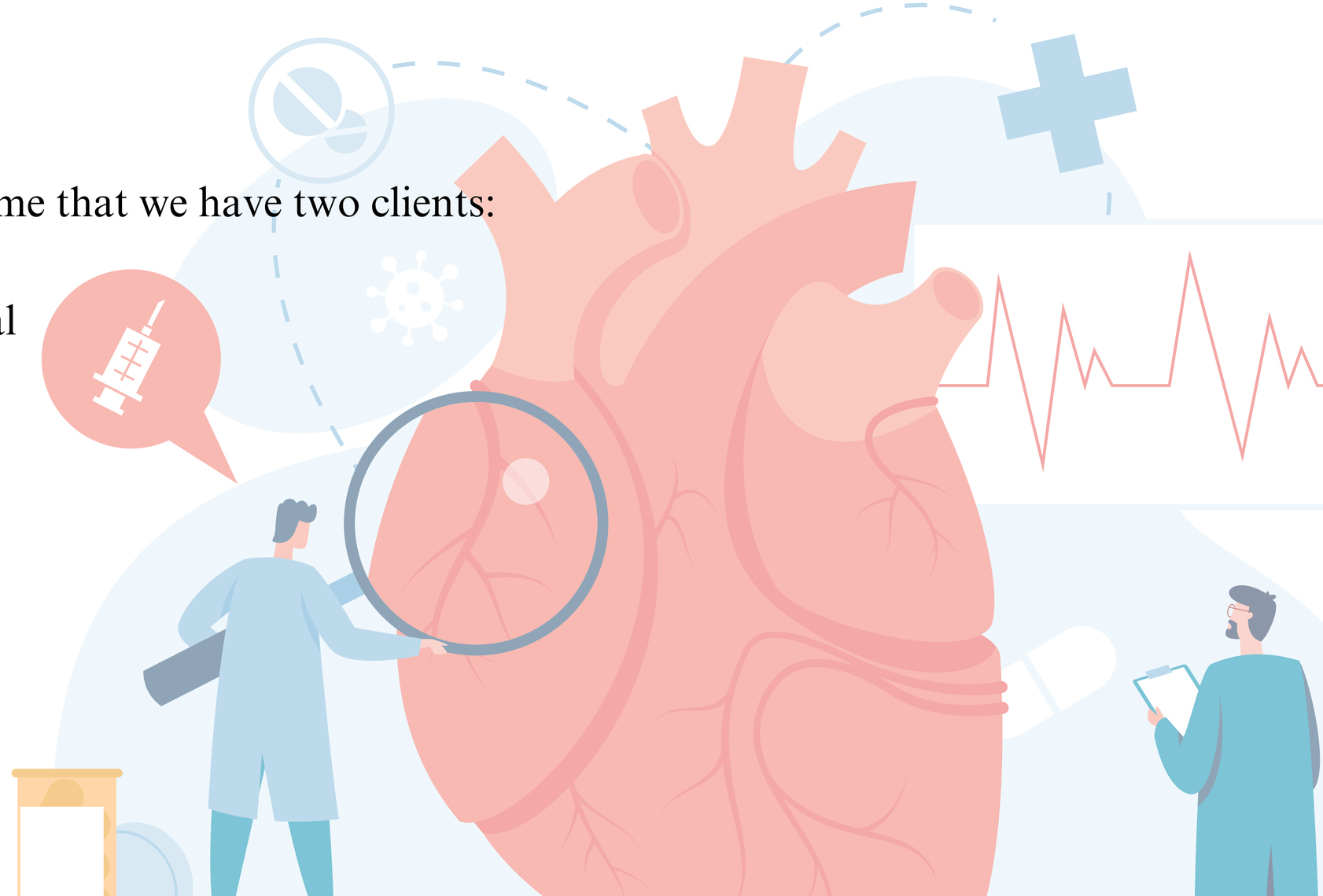
- This project explores the potential of machine learning algorithms for CVD diagnosis as machine learning has gained significant attention in medical research with promising results.

- By training machine learning algorithms on vast datasets of patient information, such as demographics, lifestyle factors, and medical history, patterns can be identified, and predictions about the likelihood of a patient having CVD.

Stakeholders

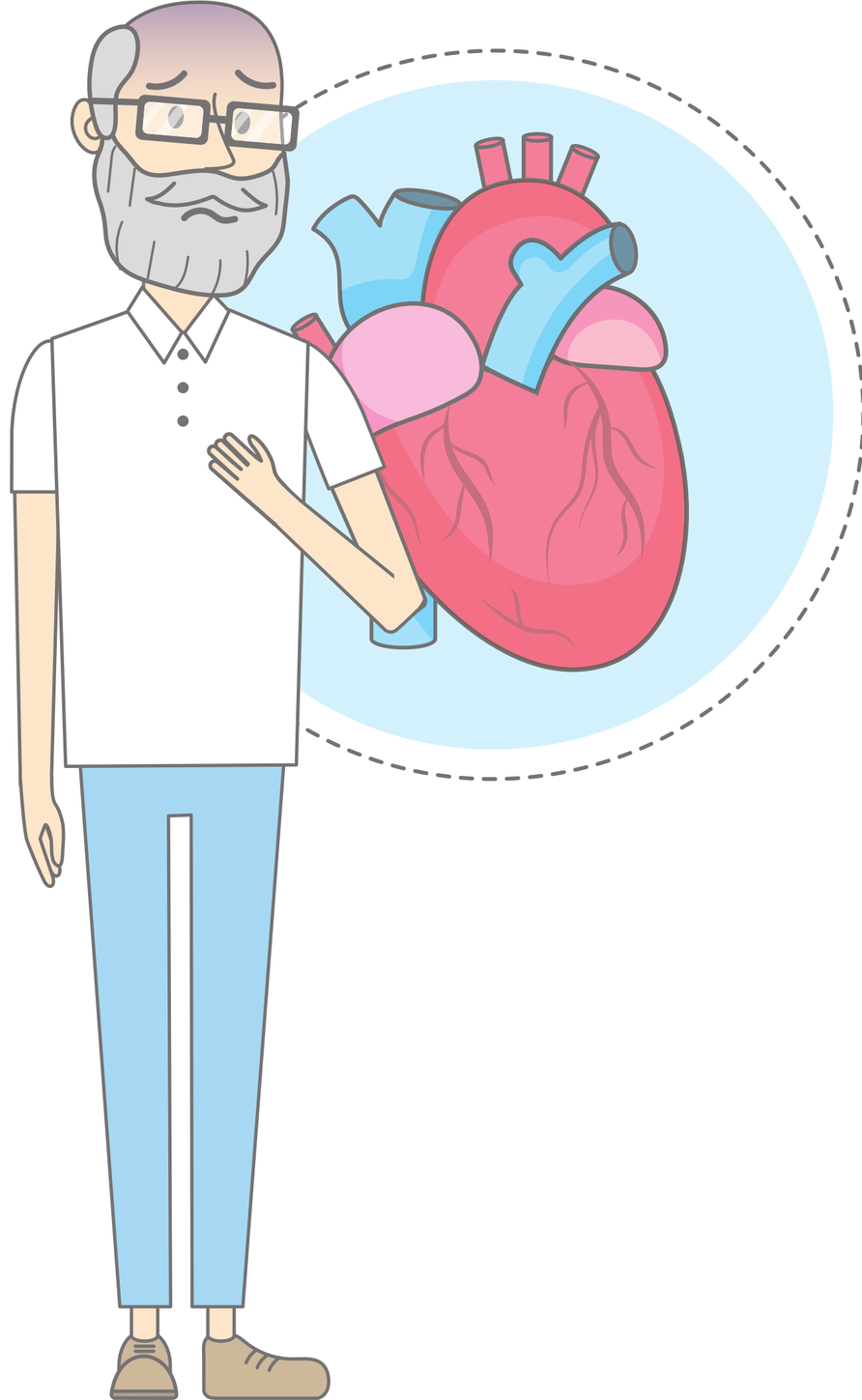
For our project, we will assume that we have two clients:

1) The Nairobi West Hospital
2) The patients



Problem Statement

The Nairobi West Hospital is in search of a more streamlined and dependable approach to diagnose CVD, aiming to provide patients with expedited diagnostics and mitigate the potential for adverse outcomes.



Objectives

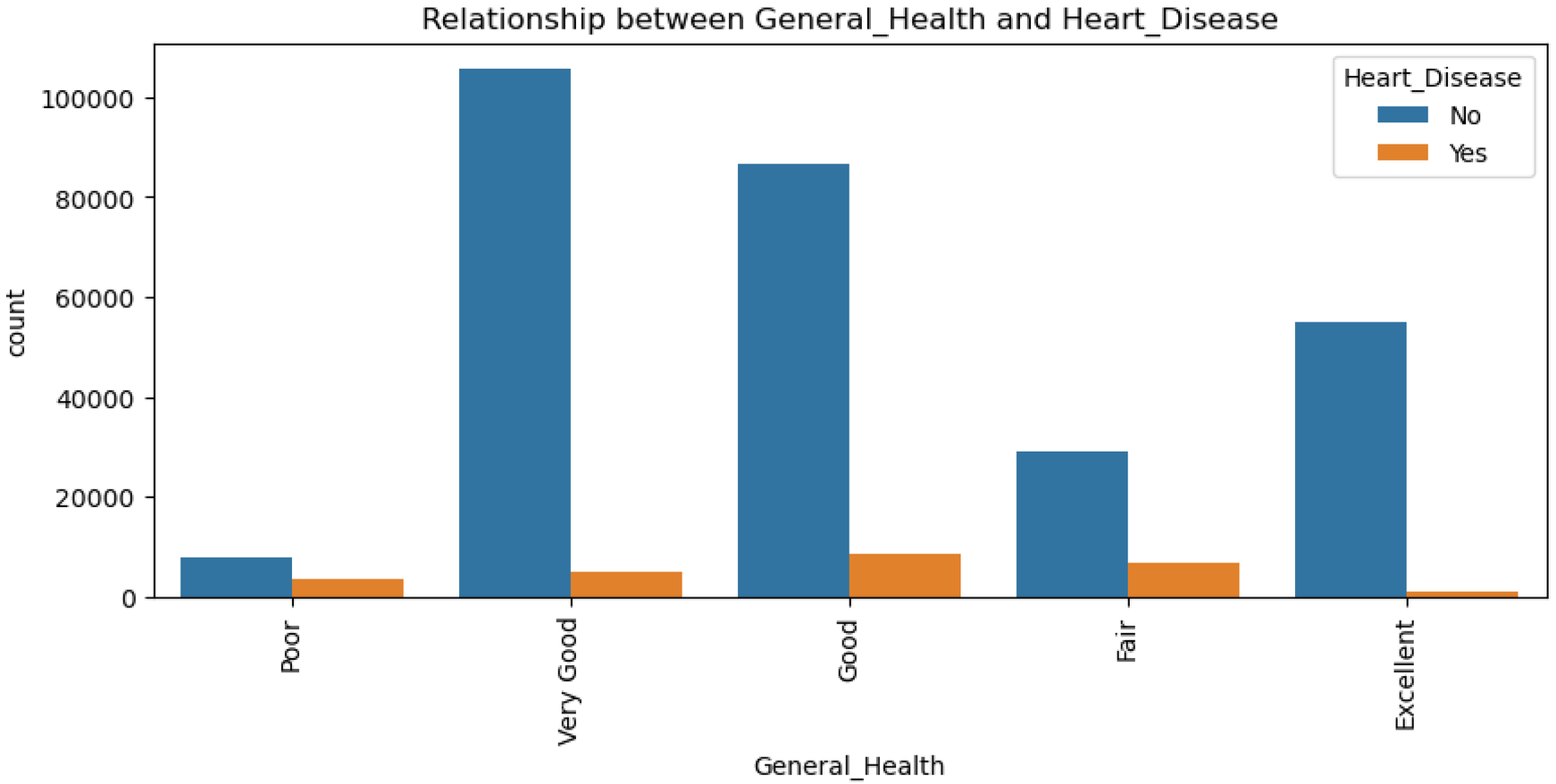
1. **Effective Communication of Insight**
    a. Summarize research findings in a clear and concise manner that aligns with the data-driven initiatives of the Nairobi West Hospital.
    b. Create detailed reports and documentation to communicate insights gained from the predictive model to the Lead Data Scientist.
2. **Development and Application of a Predictive Model**
    a. Apply machine learning algorithms to build a predictive model that accurately assesses the risk level of patients for developing cardiovascular diseases.
    b. Optimize model parameters to ensure high diagnostic precision and minimize false positives and false negatives.

Research Questions

1. What is the relationship between patients' general health and heart diseases?
2. To what extent does exercise determine whether an individual is diagnosed with heart disease?
3. What is the prevalence of heart disease in male and female sexes?
4. What relationship exists between different age group categories and heart disease diagnosis?
5. Does smoking increase the likelihood of a heart disease diagnosis?

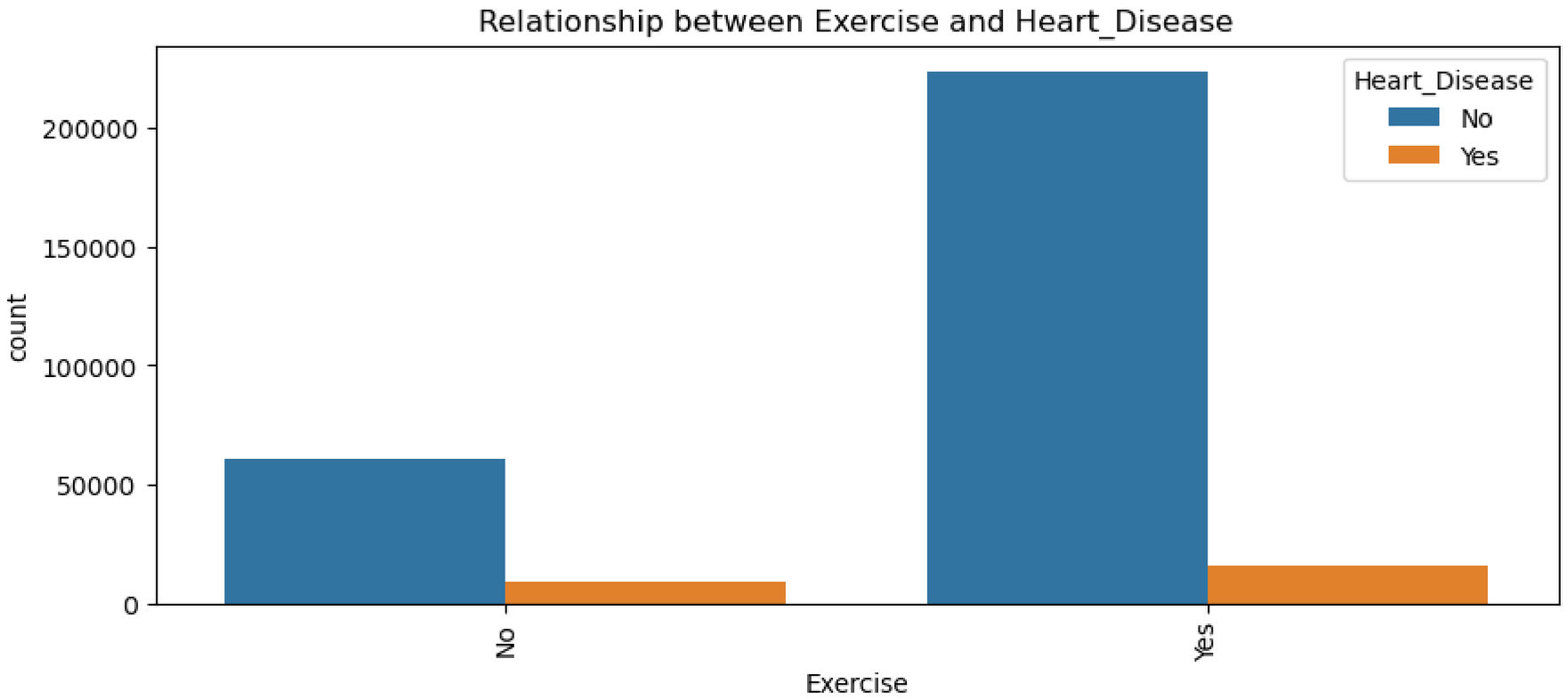
What is the relationship between patients general health and heart disease diagnosis?

- Patients in excellent health showed very little count of heart disease,
- Patients in very good health showed a very high count of no heart disease
- Patients in poor health showed that about half of those had heart disease.



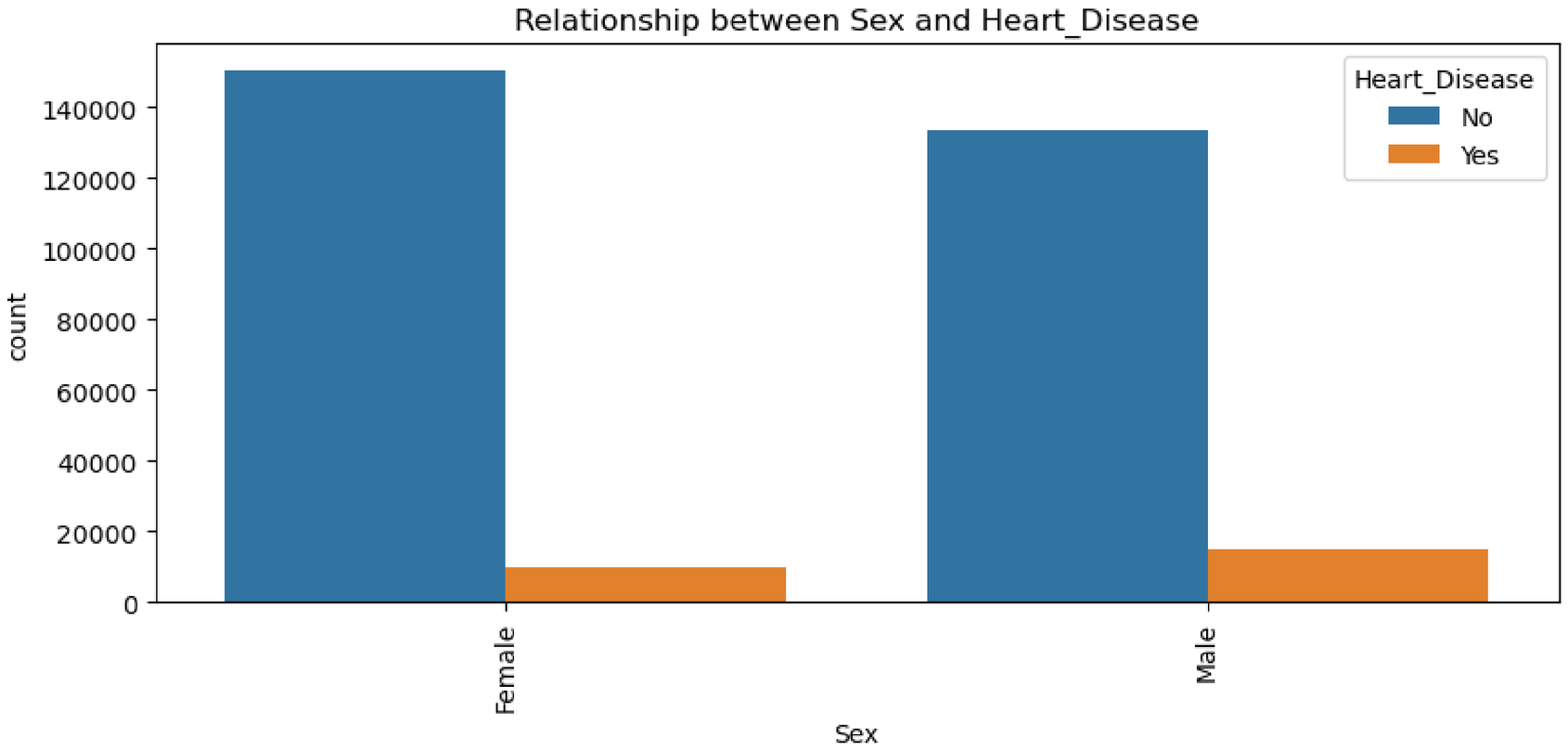
To what extent does exercise determine whether an individual is diagnosed with CVD or not?

- Patients who exercised more frequently recorded a high count of patients with no heart disease.



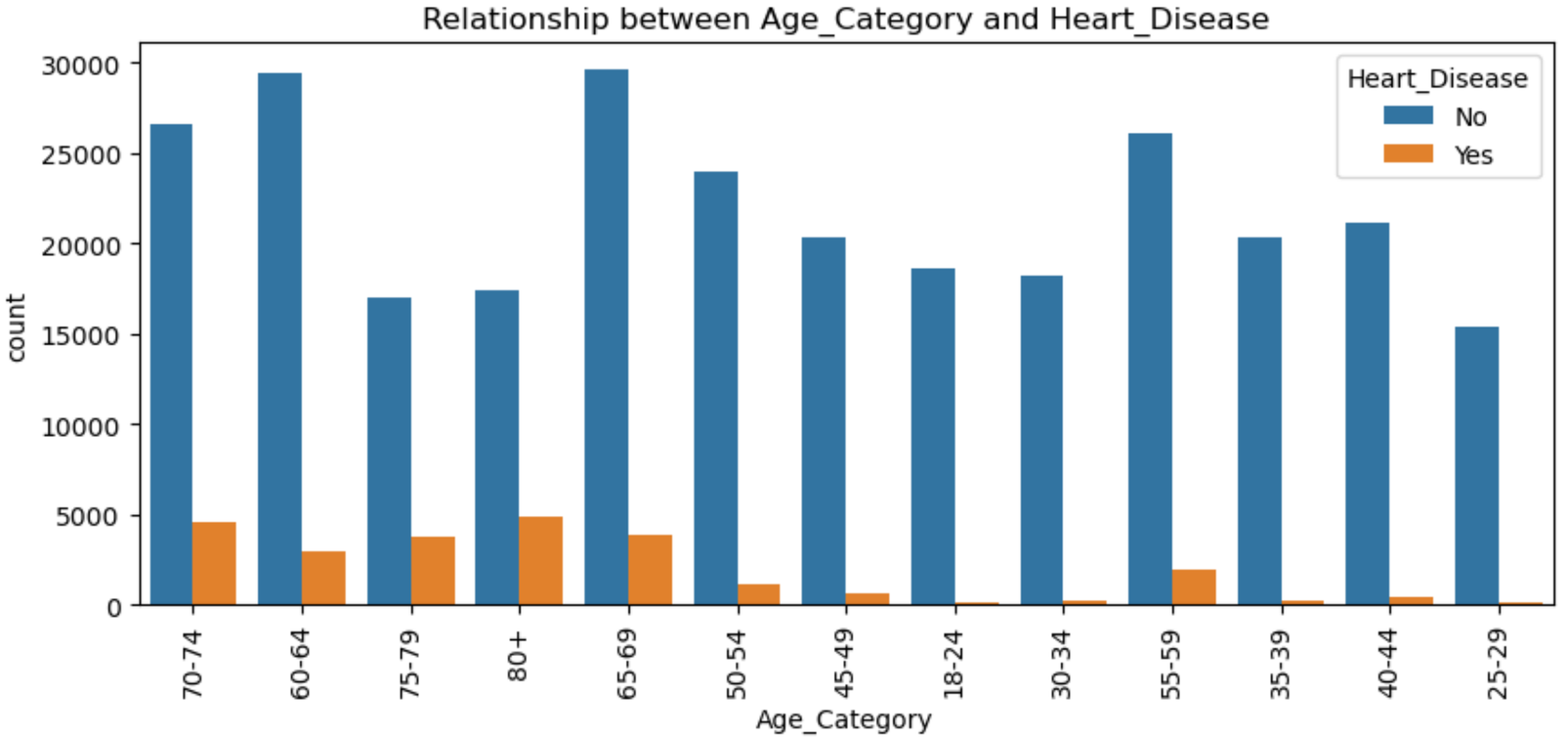
What is the prevalence of heart disease in female and male sexes?

- Male patients had a higher count of patients with heart disease than female patients



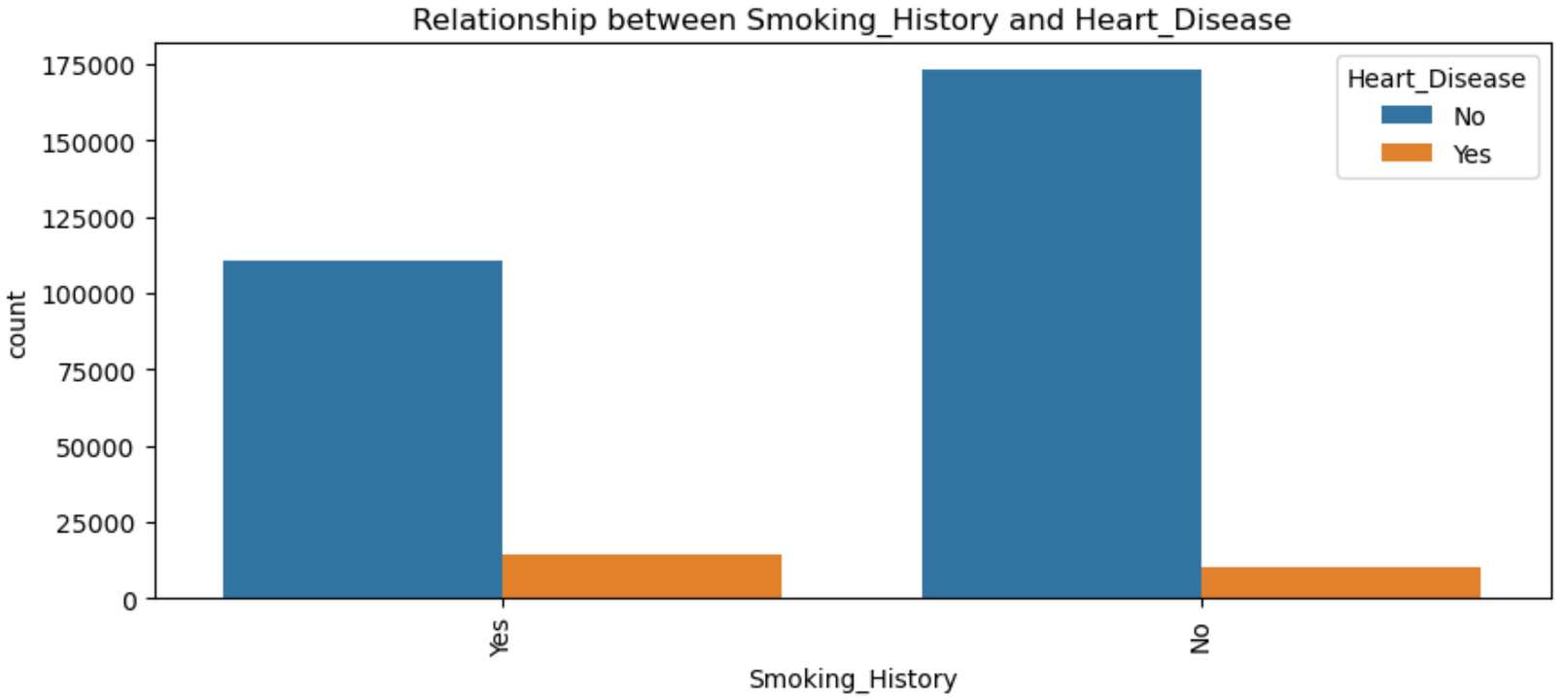
What relationship exists between different age group categories and heart disease diagnosis?

- Patients 80 years and older showed the highest prevalence of heart disease.
- Individuals 55 and above are observed to have a higher likelihood of being diagnosed with CVD.



Does smoking increase the likelihood of a heart disease diagnosis?

- Patients who did not smoke recorded a lower count of heart disease compared to the patients who did smoke.



Modelling

Model	Logic
Logistic Regression	Logistic regression is well-suited for binary classification problems like predicting the presence or absence of cardiovascular disease. It calculates the probability of an event occurring (in this case, disease presence) based on input features.
Decision Tree	Decision trees are versatile and can handle both classification and regression tasks
Gaussian	Gaussian Naive Bayes is suitable for probabilistic classification tasks.
KNN	K-Nearest Neighbors (KNN) is a non-parametric algorithm that can capture complex patterns in data.

MODEL	BASELINE ACCURACY	TUNED ACCURACY
Logistic Regression	81.59%	81.59%
Decision Tree	89.00%	90.78%
Gaussian	76.28%	77.86%
KNN	91.00%	94.00%

Conclusion

- The number of heart disease cases increases with age category. 'Older' people tend to be affected more by heart disease compared to young people.

- There is little difference between the male and female category. Both genders are susceptible to heart disease with the male gender more affected.

- Based on 'BMI Category', Underweight people are less likely to be affected by heart disease with overweight and obese people are more likely to have heart disease. Normal-weight people are also likely to be affected by heart disease.

- Our best model was KNN, with an accuracy of 94% after hyperparameter tuning. However, it was computationally expensive as it took more than 6 hrs to return results. Our selected model is therefore the `decision tree` which gave an accuracy of 90.78%. This accuracy conveys that the model will give the right predictions over 90% of the time

Recommendations

- With an accuracy of above 90%, we recommend that health practitioners use our algorithm as the first screening to diagnose patients before they can perform a second consultation increasing the hospital's efficiency in diagnosing patients.
- We recommend that patients maintain good overall health through healthy habits and regular check-in in order to lower the risk of heart disease.
- Patients who are 55 and older are advised to attend regular screenings to aid in the early detection and management of heart conditions.
- It is recommended that the Nairobi West Hospital promote smoking cessation programs and continue to raise awareness about the cardiovascular risks associated with smoking

Overall, these findings emphasize the importance of adopting a healthy lifestyle, regular exercise, and risk factor management to reduce the risk of heart disease. Tailoring interventions based on gender, age, and health status can contribute to better heart health outcomes in the population. Furthermore, raising awareness and providing education on the impact of lifestyle choices on heart health is crucial for preventing heart disease