**Unveiling Hidden Heart Attack Risks**

**Varshini Vaisnavi Srinivasan,** [**vsrinivasan3@student.gsu.edu**](mailto:vsrinivasan3@student.gsu.edu)

**Introduction**

Heart attacks have become an alarmingly common occurrence, afflicting even young people when least expected. This project aims to reveal insights into the prevention of heart attacks through predictive data analysis. Using a dataset of numerous patient health factors like blood pressure, heart rate, lifestyle behaviours, and medical history, we will develop a multivariate risk model to estimate an individual's likelihood of experiencing a heart attack. By uncovering the complex interplay between the many demographic, genetic, and behavioural variables that influence cardiovascular outcomes, our goal is an actionable risk calculator to empower at-risk individuals to lead proactive, healthy lives. This undertaking promises life-saving impact - identifying those in danger early enough to make pivotal lifestyle changes, seek medical care, and avoid becoming another tragic statistic. Through compassionate data science, we endeavour to illuminate the path towards a future of vibrant, thriving heart.

**Goals:**

The goal of this project is to use data analysis to develop a predictive model that estimates an individual's risk of experiencing a heart attack. By analysing various health factors, the aim is to empower individuals with a risk calculator, allowing early intervention and lifestyle changes to prevent heart attacks and ultimately save lives

**Dataset**

The dataset contains 8,764 rows and 26 columns. Each column represents the lifestyle, health data of each patient.

* **Patient ID** - Unique identifier for each patient
* **Age** - Age of the patient
* **Sex** - Gender of the patient (Male/Female)
* **Cholesterol** - Cholesterol levels of the patient
* **Blood Pressure** - Blood pressure of the patient (systolic/diastolic)
* **Heart Rate** - Heart rate of the patient
* **Diabetes** - Whether the patient has diabetes (Yes/No)
* **Family History** - Family history of heart-related problems (1: Yes, 0: No)
* **Smoking** - Smoking status of the patient (1: Smoker, 0: Non-smoker)
* **Obesity** - Obesity status of the patient (1: Obese, 0: Not obese)
* **Alcohol Consumption** - Level of alcohol consumption by the patient (None/Light/Moderate/Heavy)
* **Exercise Hours Per Week** - Number of exercise hours per week
* **Diet** - Dietary habits of the patient (Healthy/Average/Unhealthy)
* **Previous Heart Problems** - Previous heart problems of the patient (1: Yes, 0: No)
* **Medication Use** - Medication usage by the patient (1: Yes, 0: No)
* **Stress Level** - Stress level reported by the patient (1-10)
* **Sedentary Hours Per Day** - Hours of sedentary activity per day
* **Income** - Income level of the patient
* **BMI** - Body Mass Index (BMI) of the patient
* **Triglycerides** - Triglyceride levels of the patient
* **Physical Activity Days Per Week** - Days of physical activity per week
* **Sleep Hours Per Day** - Hours of sleep per day
* **Country** - Country of the patient
* **Continent** - Continent where the patient resides
* **Hemisphere** - Hemisphere where the patient resides
* **Heart Attack Risk** - Presence of heart attack risk (1: Yes, 0: No)

**Data Types of Features:**

Patient ID object

Age int64

Sex object

Cholesterol int64

Blood Pressure object

Heart Rate int64

Diabetes int64

Family History int64

Smoking int64

Obesity int64

Alcohol Consumption int64

Exercise Hours Per Week float64

Diet object

Previous Heart Problems int64

Medication Use int64

Stress Level int64

Sedentary Hours Per Day float64

Income int64

BMI float64

Triglycerides int64

Physical Activity Days Per Week int64

Sleep Hours Per Day int64

Country object

Continent object

Hemisphere object

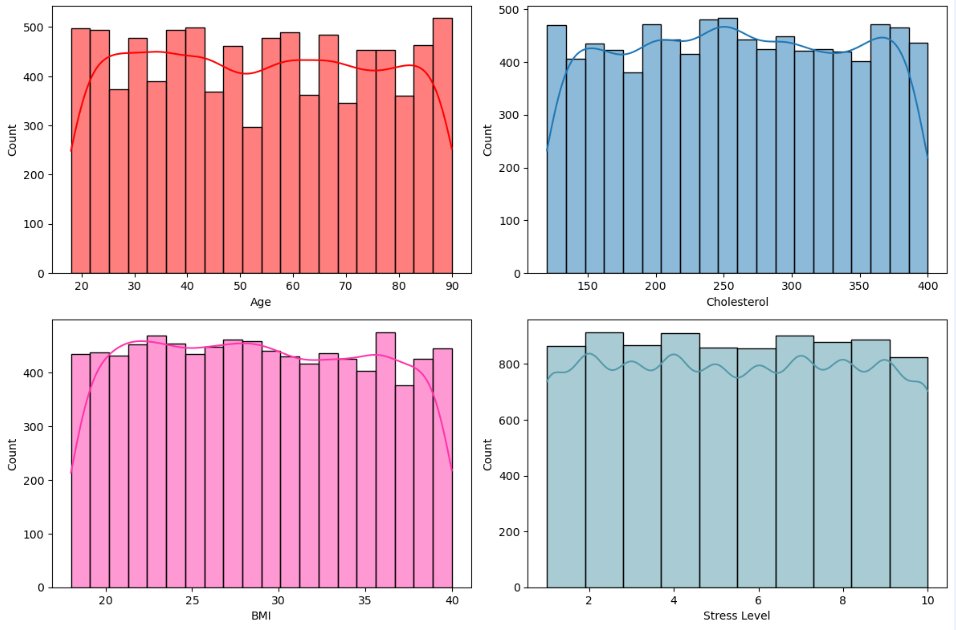
Heart Attack Risk int64

**Data Collection**

Downloaded the dataset from the Kaggle online source

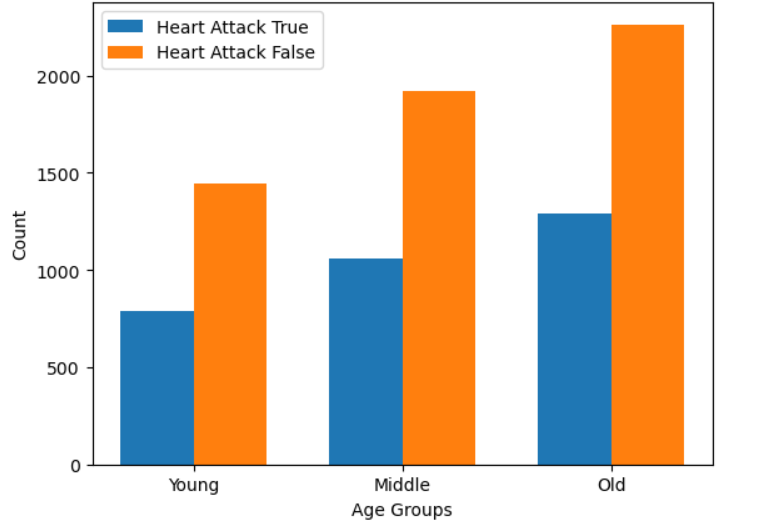
Source: Sourav Banerjee curated and prepared the dataset, leveraging synthetic data created with ChatGPT.

**Exploratory Data Analysis**

****

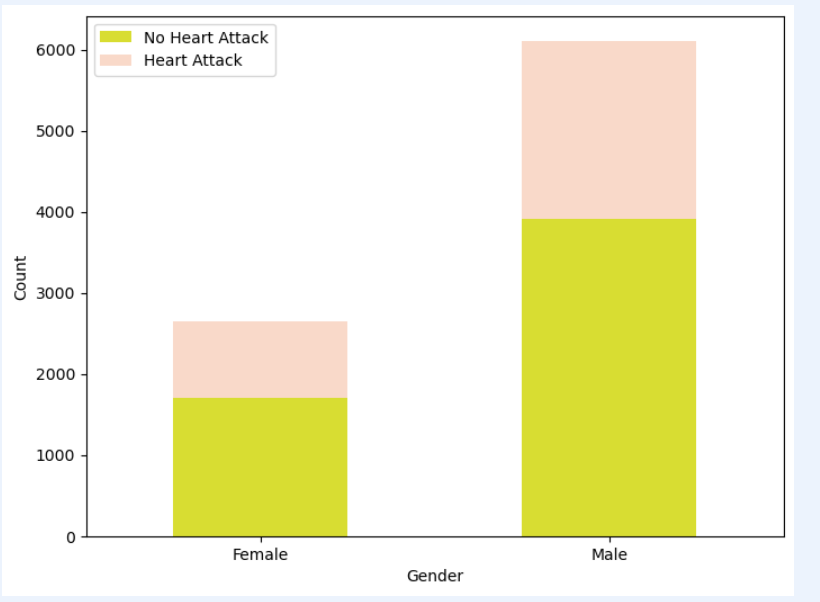
*Distribution of Age, Cholesterol, BMI, Stress Level in the dataset*

The graph shows that the dataset contains a mix of people of different ages, cholesterol levels, BMIs, and stress levels. There are more older people than younger people in the dataset, and more people with overweight or obesity than people with a healthy weight. However, most people in the dataset have cholesterol levels and stress levels within the normal range.

****

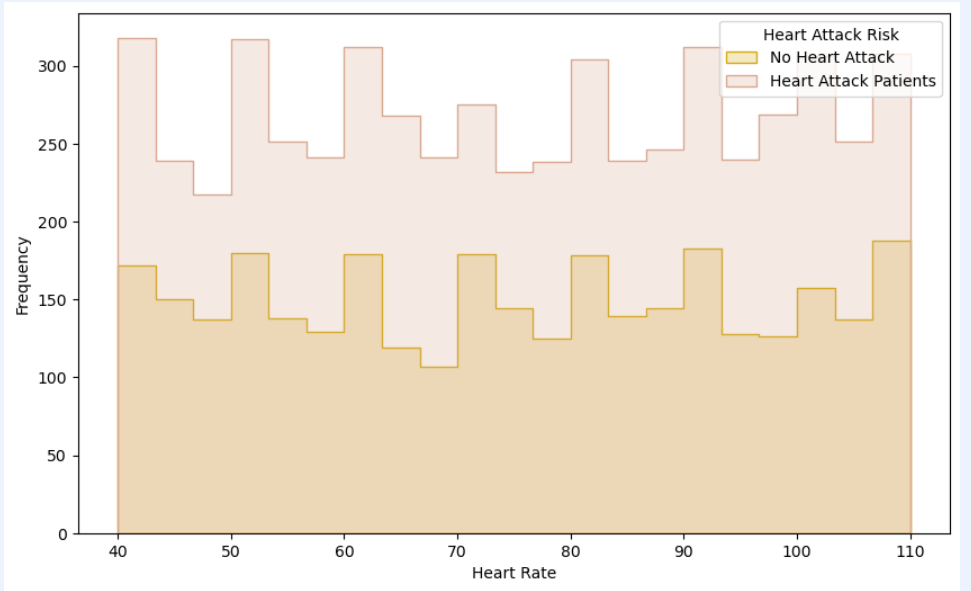
*Grouped Bar plot for Age with respect to heart attack risk*

The graph shows that age is a significant risk factor for heart attack, and that the risk increases with increasing age. People of all ages should take steps to reduce their risk of heart attack, such as exercising regularly, eating a healthy diet, and managing stress. However, people in the oldest age groups should be especially vigilant about taking steps to reduce their risk of heart attack.

****

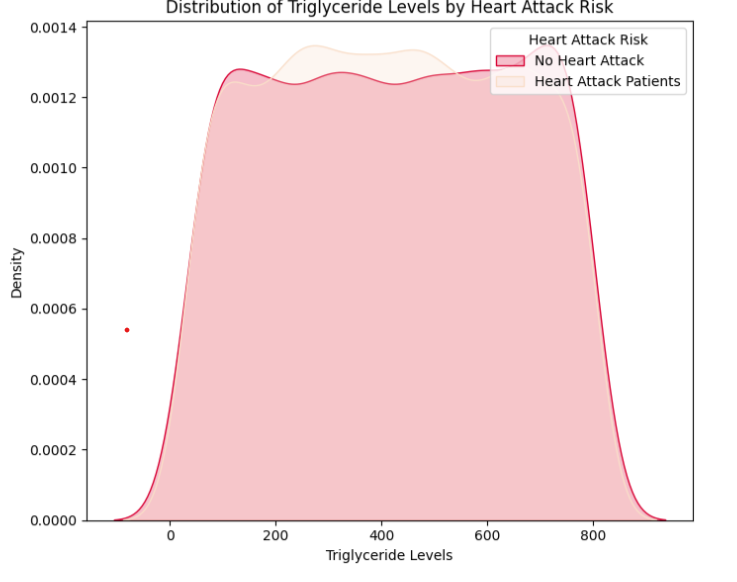
*The data bar plot shows that men high risk of heart attack*

The EDA shows that heart rate is a significant risk factor for heart attack, and that the risk increases with increasing heart rate.

****

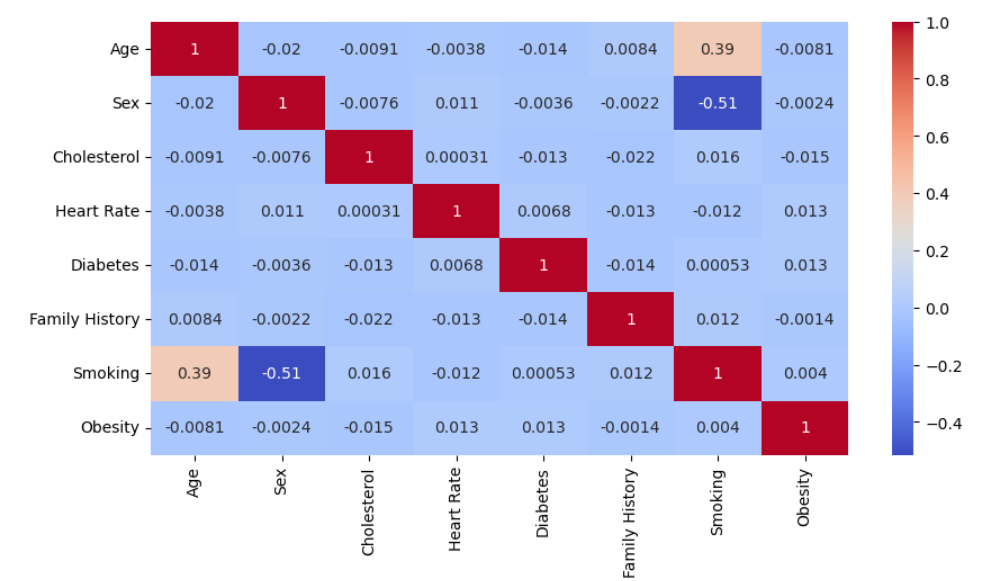
*Distribution of the heart rate with respect to heart attack risk*

The EDA shows that heart rate is a significant risk factor for heart attack, and that the risk increases with increasing heart rate.

****

This graph means that some people with high triglyceride levels do not have heart attack risk, and some people with low triglyceride levels do have heart attack risk.

Correlation Heat Map:

****

**Methodology:**

1. **Pre-processing:** Importing CSV file from Kaggle. Removing duplicates and checking for null values.
2. **Feature Engineering:** Handling categorical variables like diet with one hot coding.
3. **Exploratory Data Analysis:** Various visualization is created using Seaborn and Matplotlib to understand the distribution of data
4. **Feature Selection:** Cholesterol, Heart rate, BP, Previous Heart Problems are the important features
5. **Model Building:** Various classification model including KNN, Logistic Regression, SVM are used.
6. **Model Evaluation:** F-score, precision, recall and confusion matrix is used to measure the accuracy
7. **Final Model Training and Testing:** SVM model is finalized, trained on a combined train and validation set and tested on the test set

**Conclusion:**

The project is using the health data of various patients and analysis the factors affecting the heart using Machine learning model which predicts heart attack using the health and behaviours of the patient