The aim of this project is to build a predictive model to assess the likelihood of individuals experiencing heart-related conditions, such as heart attacks, based on their lifestyle and general health information. Heart disease remains one of the most pressing global health concerns, and I believe that early identification of individuals at high risk can lead to timely prevention and intervention strategies. By using this model, healthcare providers and individuals alike will have access to insights that can aid in reducing heart disease occurrences and improving overall public health.

The prevalence of heart disease has put a significant strain on healthcare systems worldwide. Through the use of health and lifestyle data, such as physical activity, sleep patterns, and general health status, I aim to create a model that can predict heart-related conditions before they occur. By identifying key indicators that contribute to heart health, this analysis can lead to better decision-making for both healthcare professionals and individuals looking to improve their health outcomes.

Success in this project will be determined by how well the predictive model identifies individuals at high risk of experiencing heart-related conditions. I plan to evaluate the model based on its accuracy, precision, recall, and F1 score. These metrics will help ensure that the model's predictions are meaningful and reliable. It's also important to me that the model remains interpretable so that I can clearly understand which health and lifestyle factors are most strongly associated with heart disease risk.

The scope of my solution involves several steps, including data preprocessing, feature selection, and predictive modeling. Initially, I will focus on cleaning the dataset, handling any missing values, and selecting the most relevant features for prediction. From there, I will apply various machine learning algorithms, such as logistic regression and decision trees, to develop the model. To ensure optimal performance, I will experiment with hyperparameter tuning and evaluate the models through cross-validation. The end goal is not just to create an accurate model but to identify key health indicators that contribute to heart disease, which can be valuable for future healthcare strategies.

There are some constraints I need to be mindful of. The quality and completeness of the data may impact the model's performance, so I will need to handle missing data carefully. Additionally, I must ensure that the data is balanced so that the model doesn't become biased toward predicting the majority class. Ethical considerations will also be at the forefront of my work, especially since this data relates to sensitive health information.

The main stakeholders for this project include healthcare providers who will benefit from understanding which factors to monitor more closely in their patients. Public health organizations also have a vested interest, as this model can help them target interventions more effectively to reduce the prevalence of heart disease. Ultimately, I believe this project will also benefit individuals by providing insights that can help them make more informed lifestyle choices to reduce their risk of heart-related conditions.

I'm using a dataset that contains detailed lifestyle and health information from individuals across various states. Some key variables in this dataset include general health status, physical and mental health days, physical activity, sleep hours, and whether the individual has had a heart attack. This data provides a comprehensive foundation for building my predictive model. I acquired this data from Kaggle, and I will be using it to perform all analyses and model building for this project.

To address this problem, my approach will involve several key steps. First, I will clean and preprocess the data, making sure to handle any missing values and properly encode categorical data like "GeneralHealth" and "HadHeartAttack." Next, I will conduct exploratory data analysis (EDA)

to identify the most important features and visualize correlations. Feature engineering will also play a role, as I plan to create new features like BMI categories and physical activity levels. I will then proceed to the modeling phase, where I'll implement several machine learning models, including logistic regression, decision trees, and random forests. By using k-fold cross-validation, I will ensure that the model is evaluated thoroughly for performance. Finally, I will interpret the results, identifying key predictors of heart disease and providing actionable insights for healthcare professionals.
This project will help me gain a deeper understanding of the factors that contribute to heart disease and give me the opportunity to apply machine learning techniques to a real-world health issue.