**Data cleaning:**

1. Check the datatype of the columns (.info)
2. Finding null values in the columns (.info)
3. Finding number of different categories in the columns – AgeCategory, Race, Diabetic, GenHealth
4. Converting categorical data to numerical

**EDA:**

1. Can a non-smoker get heart disease?
2. Can an alcohol drinker get heart disease?
3. Is a person with stroke, likely to get heart disease?
4. Which age group has heart disease the most?
5. Is there any relationship between sleep time and heart disease?
6. Does race affect heart disease?
7. Does sex affect heart disease?
8. What is the ideal BMI to not get heart disease?
9. How much does asthma affect heart disease?
10. How much does diabetes affect heart disease?

**Model:**

1. Which model will be the best? (Accuracy)
2. Logistic regression
3. KNN

**Model analysis:**

1. Accuracy
2. Confusion matrix

**Deployment:**

Streamlit interface