**COMP5310 Presentation Script**

Slide 1:

Hello everyone, welcome to my presentation for COMP5310. In this presentation we will be answering the research question: what are the most important factors in predicting the presence of heart disease, and how can these factors be incorporated into a predictive model?

Slide 2:

Here is the following agenda we will be covering, from the introduction and the problem statement to highlighting the data analysis tools and techniques that will assist in bringing useful insights.

Slide 3:

* Cardiovascular diseases are responsible for 43% fatalities worldwide (Bhatt, 2023).
* What is more beneficial than merely detecting the presence or absence of heart diseases is to classify these diseases and understand the relationship between patients from this dataset on a molecular level, including their cholesterol and blood sugar levels.
* By performing data analysis, misdiagnosis of these diseases can be reduced through these vastly accurate methods using several classification and clustering methods in machine learning.

Slide 4 :

* The dataset contains 12 attributes such as age, sex, blood pressure, cholesterol levels and maximum heart rate
* The dataset contains a total of 918 instances as well, which provides a sufficient amount of data to train and evaluate for heart failure detection
* The outcome variable is binary, with 1 indicating the presence of heart disease and 0 indicating the absence of heart disease
* We will be conducting thorough data preprocessing and exploratory data analysis prior to developing our models

From here, don’t have to read from script