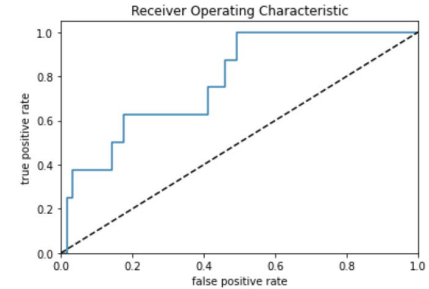**H0:** There is no relationship between the independent variables (X = numerical data) and the dependent variable (y = categorical data = loan status).

H1: There is a relationship between the independent variables (X = numerical data) and the dependent variable (y = categorical data = loan status).

There is a 78% chance that the model will be able to distinguish between status A (good loan) and status B (bad loan).

However, the model fails to predict status B completely.

Receiver Operating Characteristic

| Hypotheses | Business Impact | Accuracy of Model | Further Steps to Improve the Model |

- The model provides an assessment of loans based on demographic data and increases understanding by showing which factors most affect specific outcomes
- In the long run, the model can reduce the amount of bad loans and secures financial stability of the bank

- More data needed to increase size of dataset
- Data needs to be balanced
  (e.g. using SMOTE)
- Drop unnecessary columns
  (e.g. 95_umep, crime_95)
- Run statistical tests, e.g. Likelihood ratio test