Predicting Income

Problem Statement

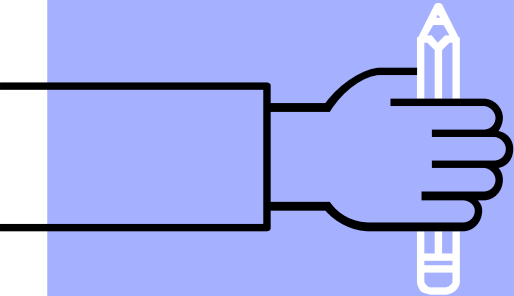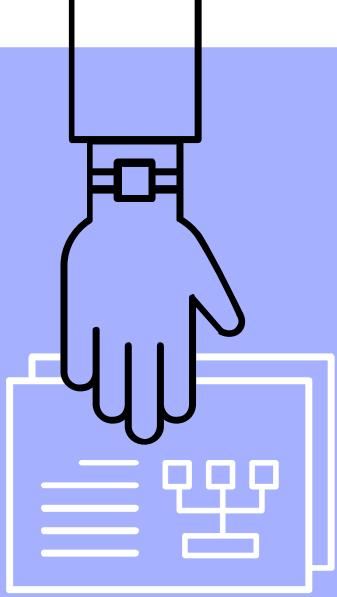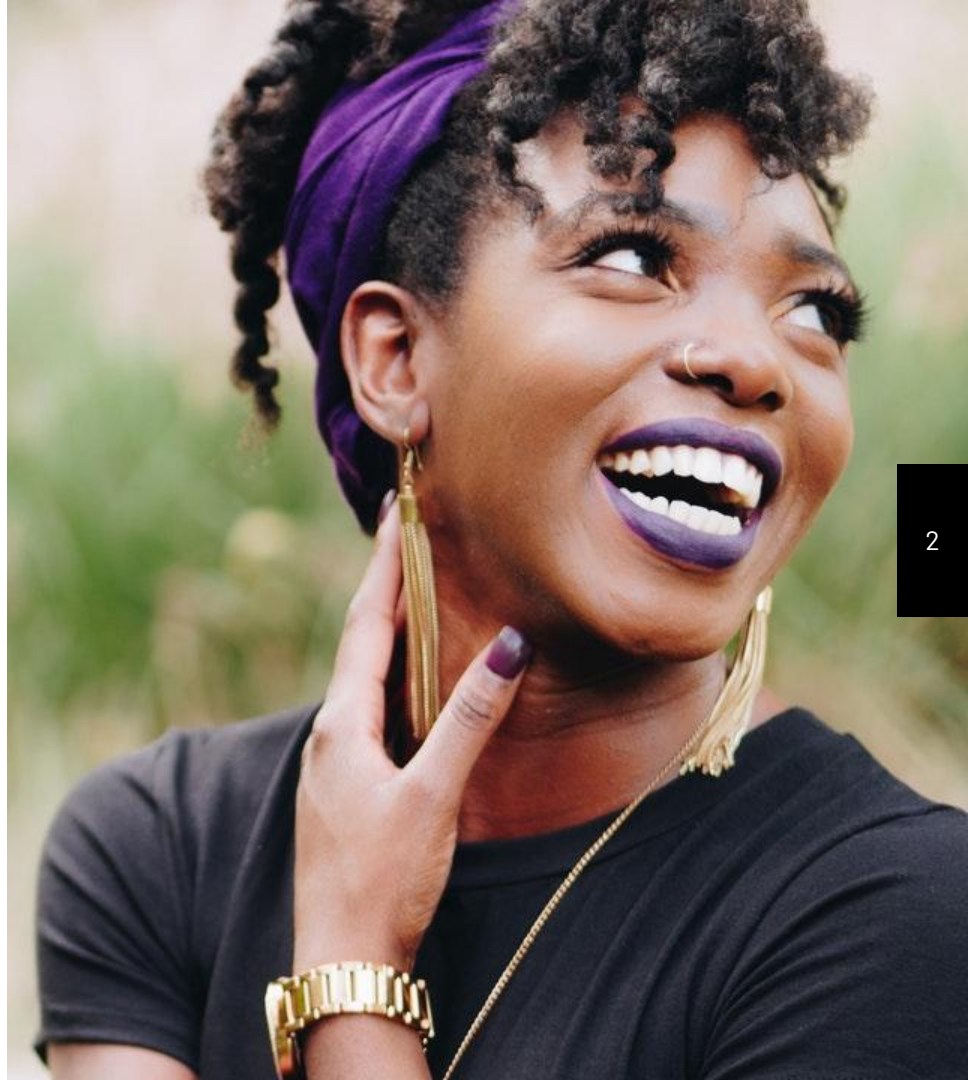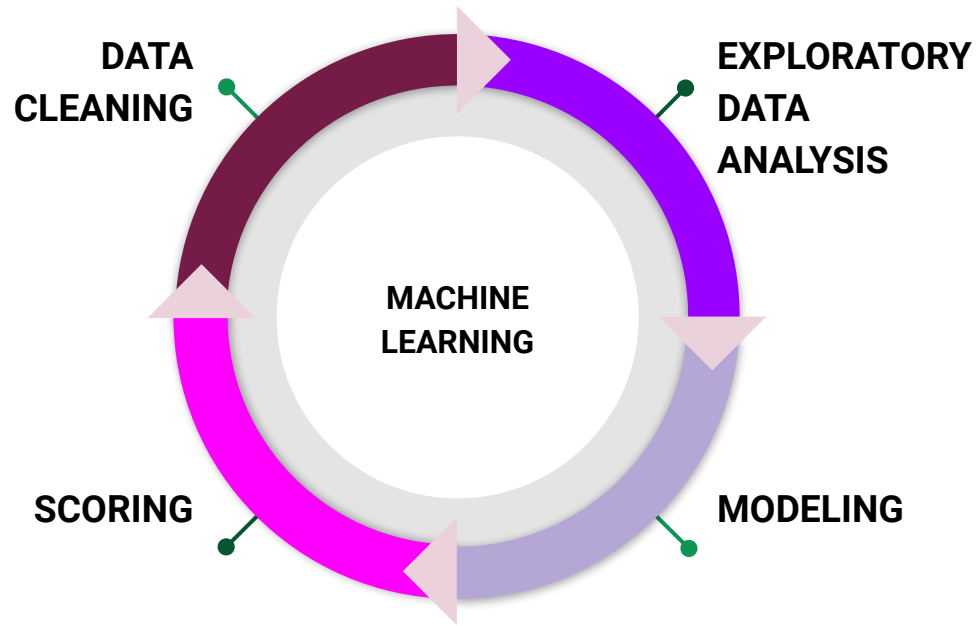Build a model to predict if someone will make greater than or less than $50,000 given their:

▷ Age
▷ Education level
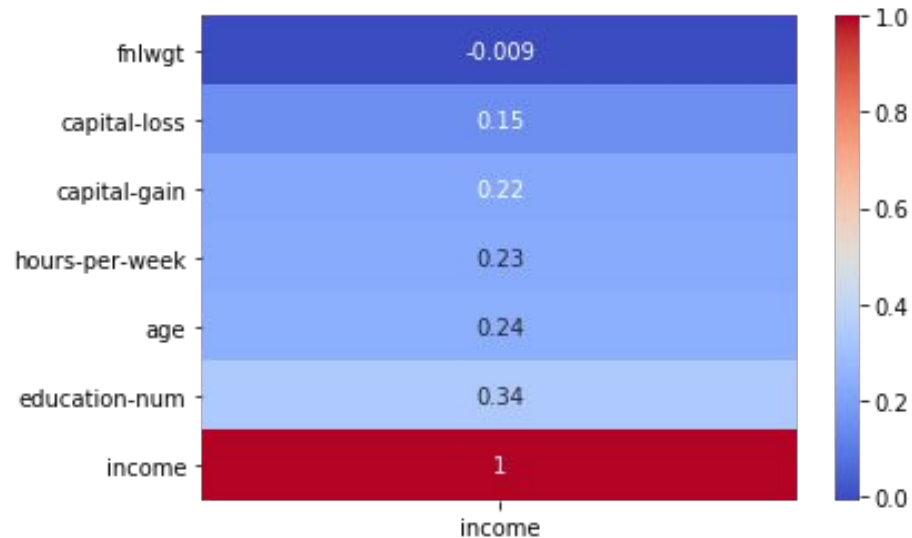▷ Occupation, working class, hours per week
▷ Marital status, race, sex

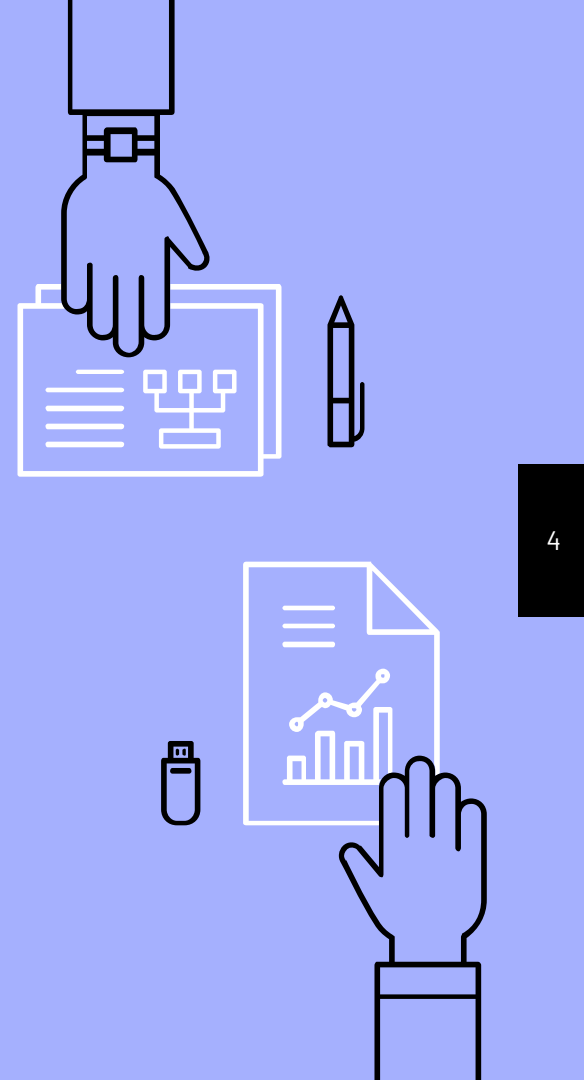# OUR PROCESS IS EASY

DATA CLEANING

EXPLORATORY DATA ANALYSIS
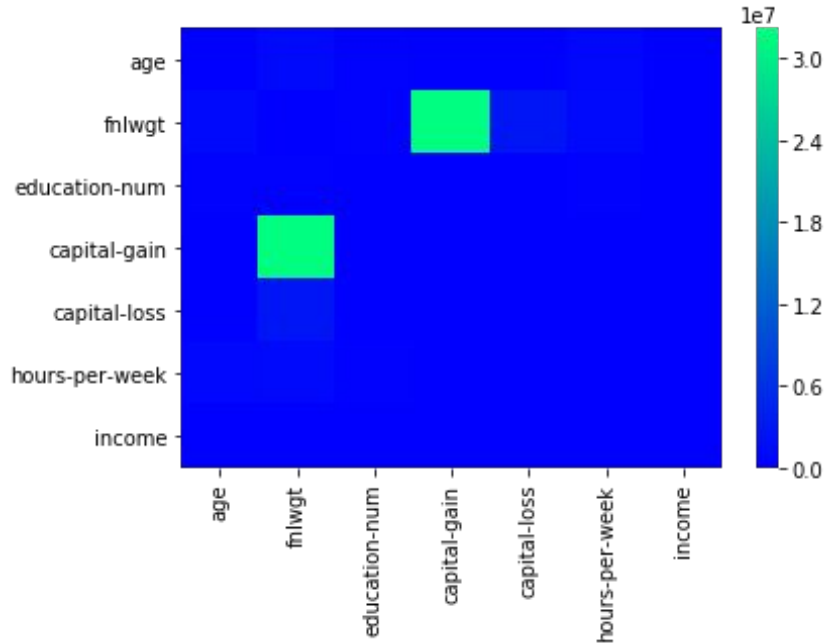
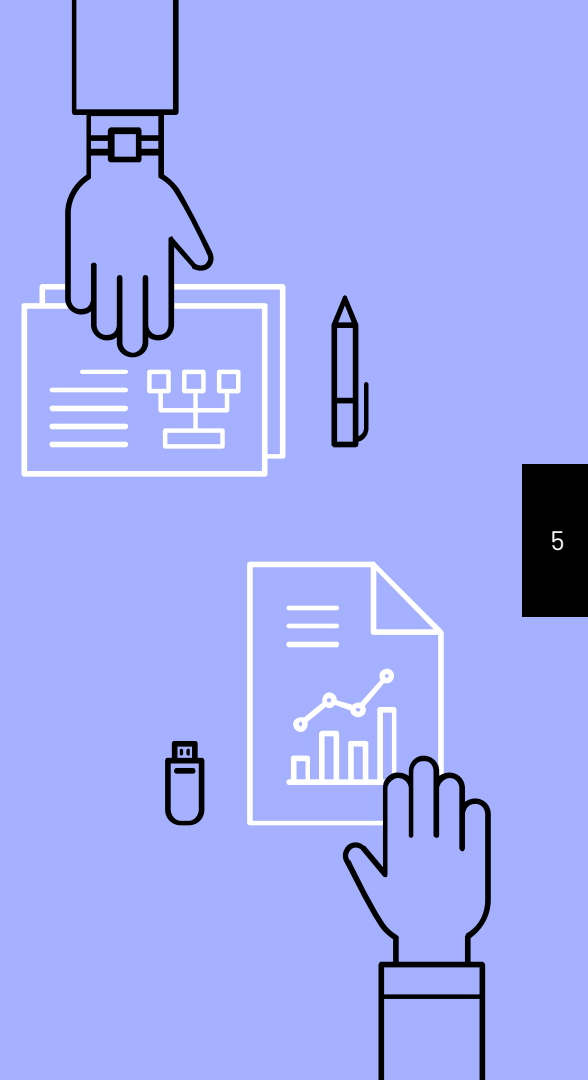MACHINE LEARNING

SCORING

MODELING

Exploratory Data Analysis

Education, age, hours worked, and capital gain have the highest correlation with Income
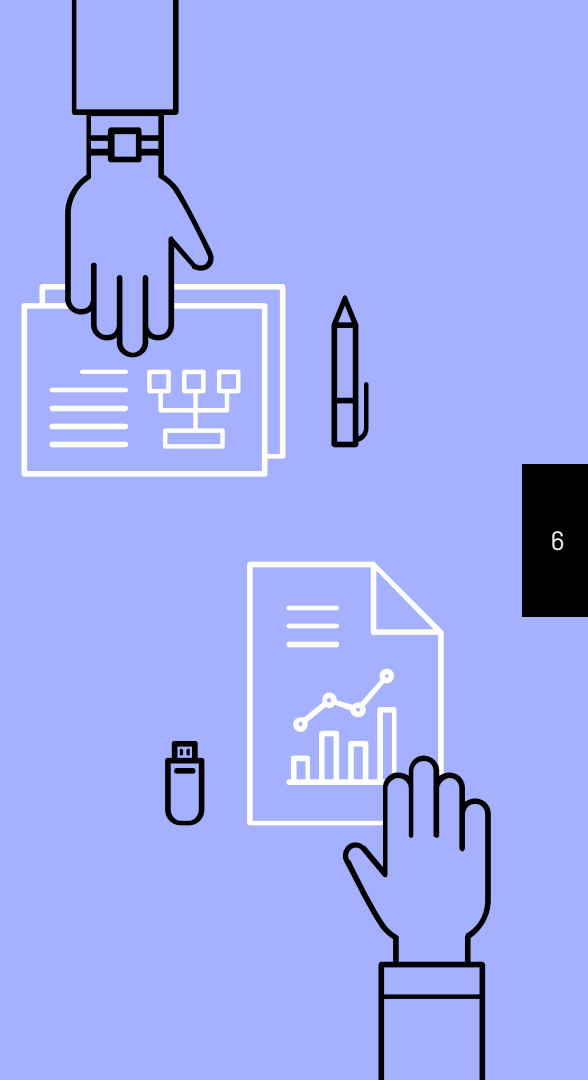
Exploratory Data Analysis

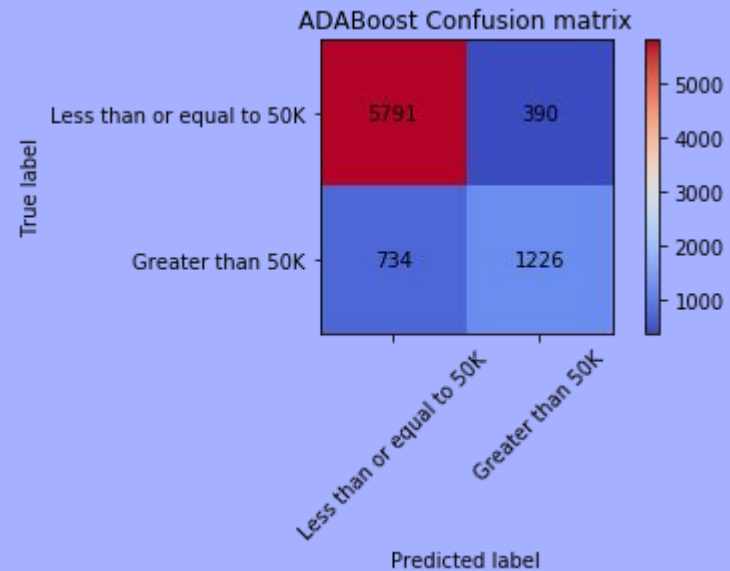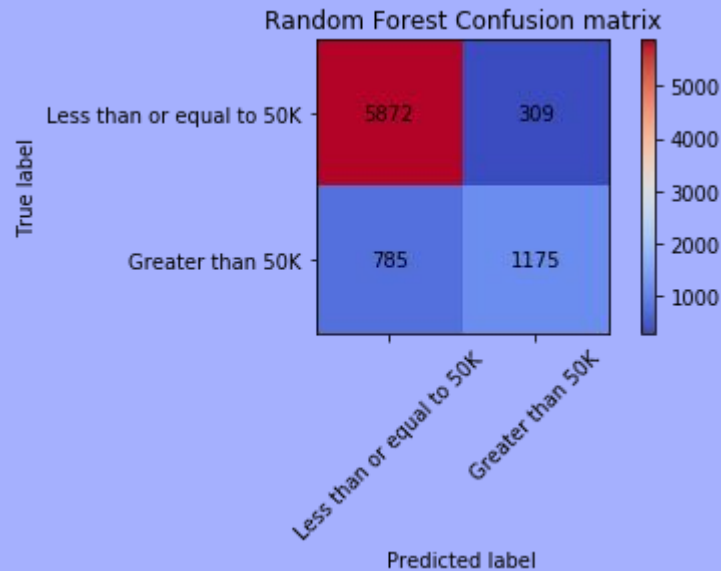'capital_gain' and 'fnlwgt' were highly correlated so we decided to use only one of those variables in the model

Statistical Models

	Logistic Regression	**Random Forest**
Taining	0.863	0.876
Test	0.854	.865

Evaluating our Models

Random Forest Confusion matrix

ADABoost Confusion matrix
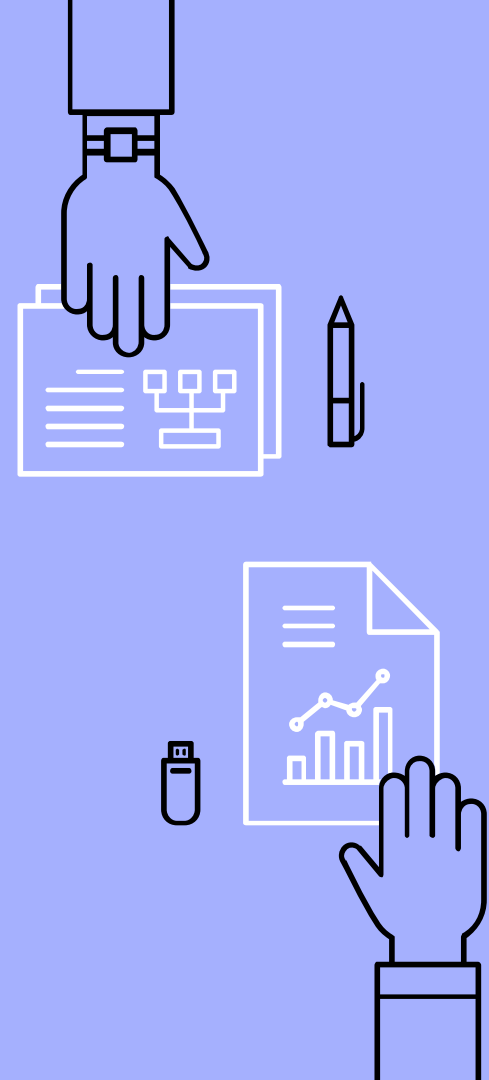
By using AdaBoost Classifier, we were
able to reduce False Positives

Executive Summary

▷ Data Wrangling
  ▪ Deal with incomplete dataframes
▷ Modeling
  ▪ Imbalanced data
  ▪ Multi-collinearity
▷ Internal discussion on features

THANKS!

Any questions?