EDA for Default of Credit Card Clients

Helnaz Soltani

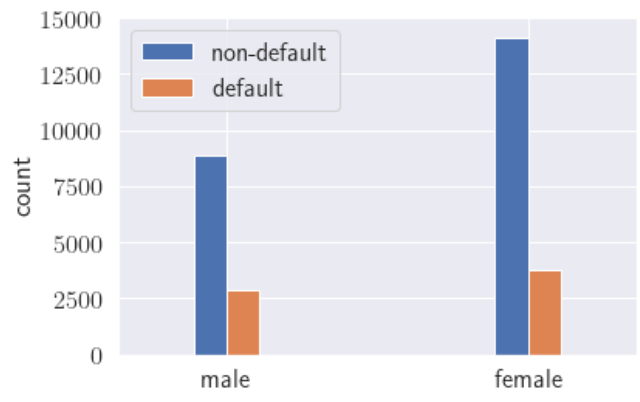
April 10, 2020

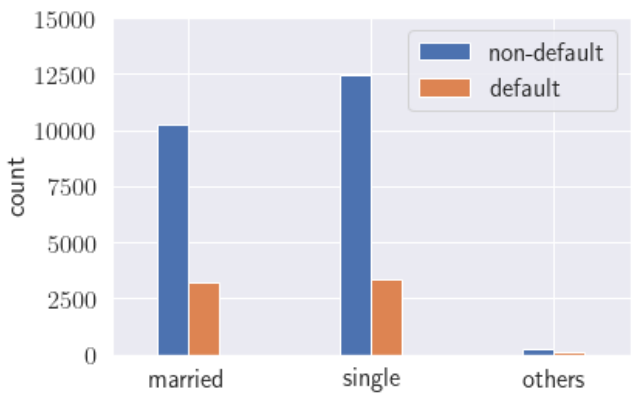
Galvanize DSI

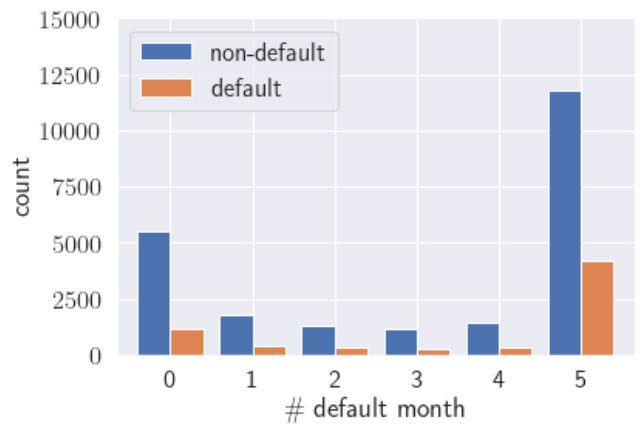
Project Overview

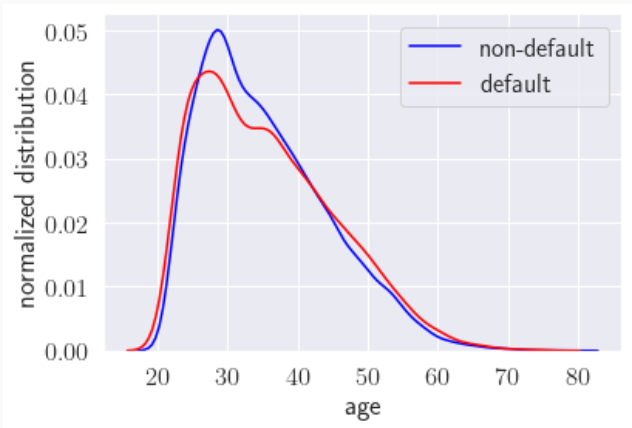
- Objectives: To perform an EDA on the dataset in order to predict the default of the credit card clients.
- Dataset information:
  - Number of instances: 30000
  - Number of attributes: 24
  - Data analysis tasks: Classification
- Data cleaning issues:
  - Duplicated entries
  - Inconvenient columns name
  - Incorrect datatype
  - Undocumented labels for some of the features
  - Inconsistent values for some of the columns
- Feature engineering:
  - Credit utilization
  - Number of defaults in the previous 5 months

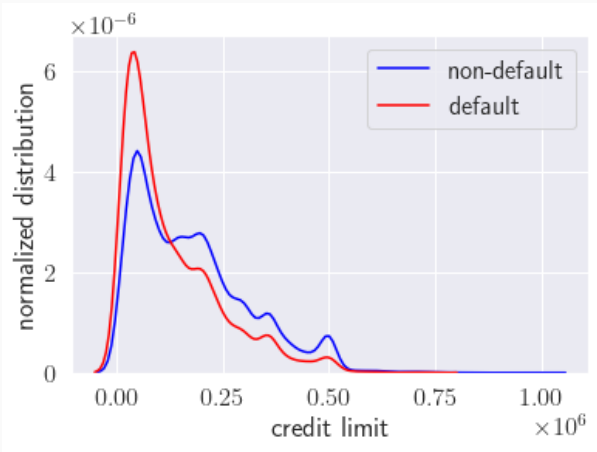
Descriptive Statistics

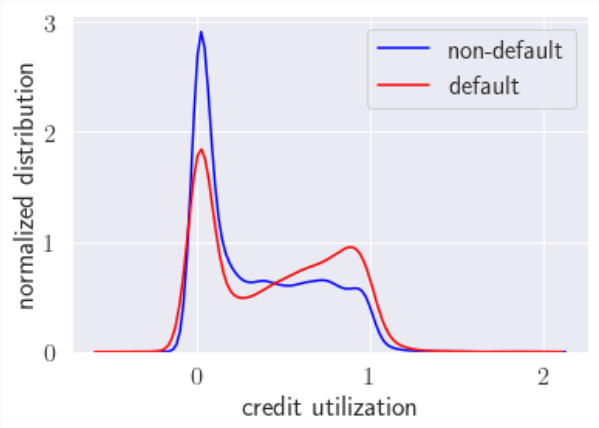








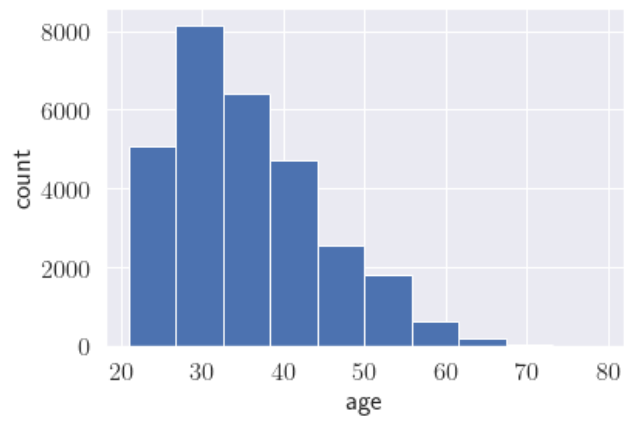


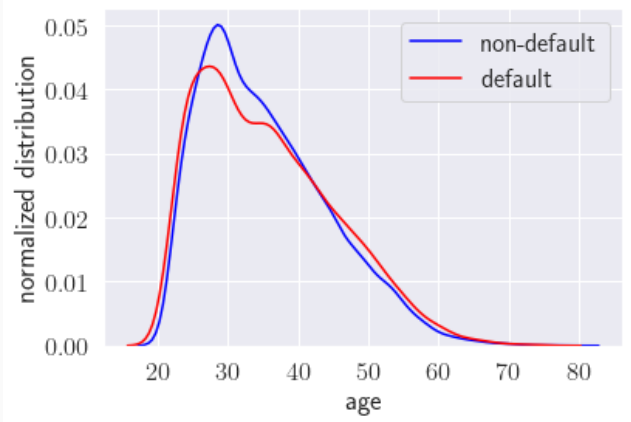


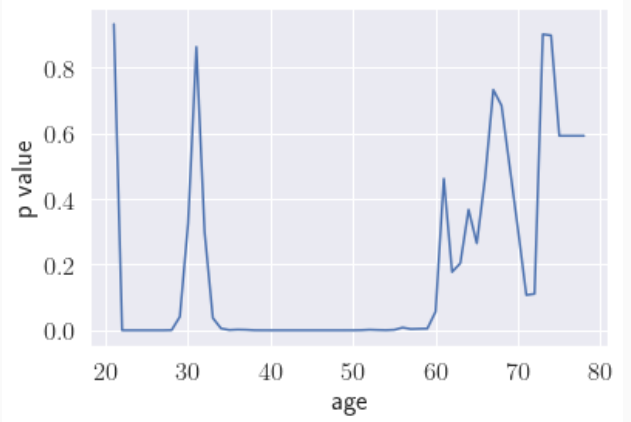
Inferential Statistics - Hypothesis Testing

Hypothesis testings have been performed to determine if there is statistically significant difference between the default means for the following features:

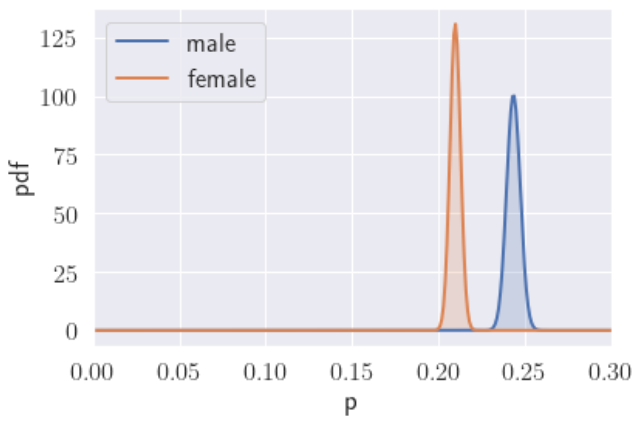
- Sex (male vs. female)
- Marital Status (married vs. single)
- Education (educated vs. non-educated)
- Age (below and above a certain age)

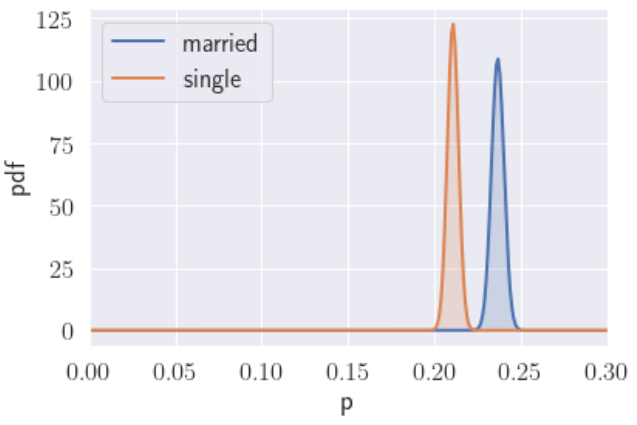


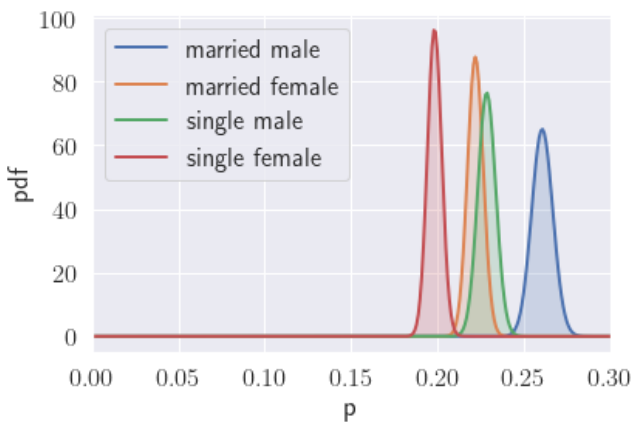


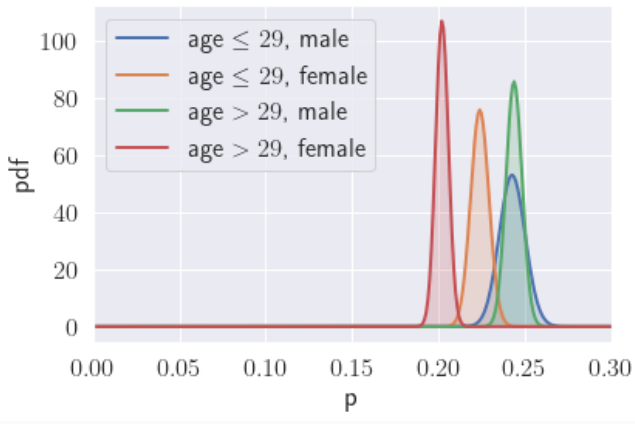


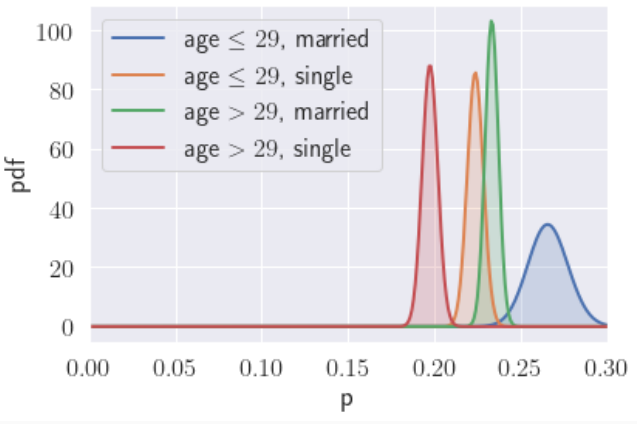
Inferential Statistics - Bayesian Testing

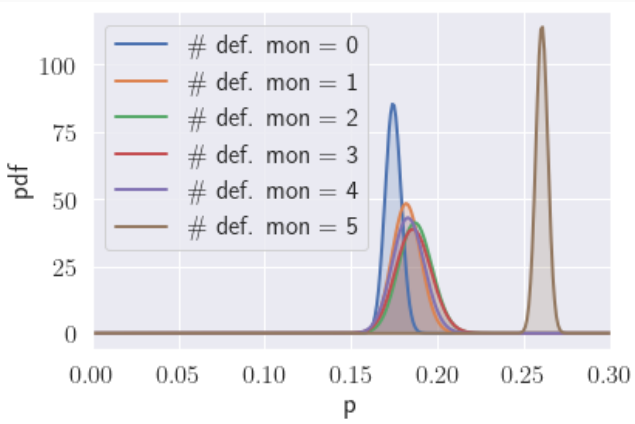












Thank you!