Predicting Default of Credit Cards

Yi Wang

Brown University
10/22/2019

Github:
https://github.com/yahowang/Data1030Project

Introduction

- Problem to solve: target potential customers who will default

- It is challenging yet vital for banks to minimize the risk of capital loss by

  determining whether their customers would pay off their credits.

- Type of problem: classification

- Data Source: UCI Machine Learning Repository

Data Preprocessing

- 30,000 instances with 24 features
- Imbalanced
- No missing values
- 4 sets of features:
  - Demographical data (One hot & Standard)
    - Gender, Education Level, Age, etc.
  - Monthly billing amount (Standard)
  - Monthly payment amount (Standard)
  - Monthly delay status of payment (One hot & Minmax)
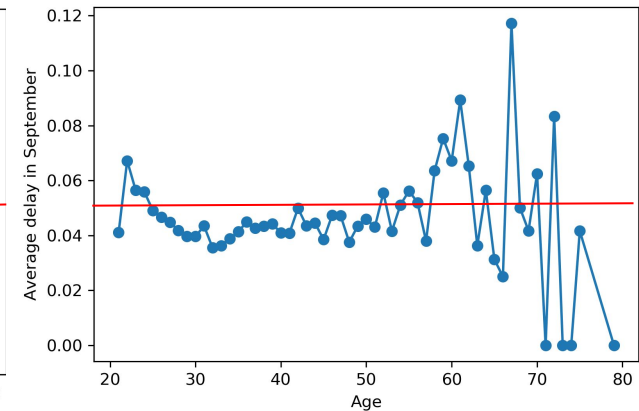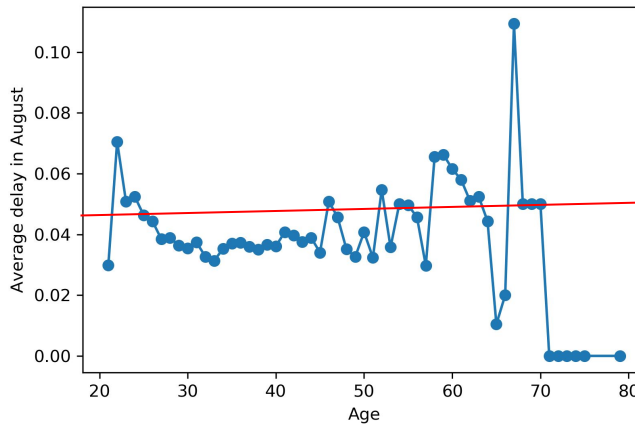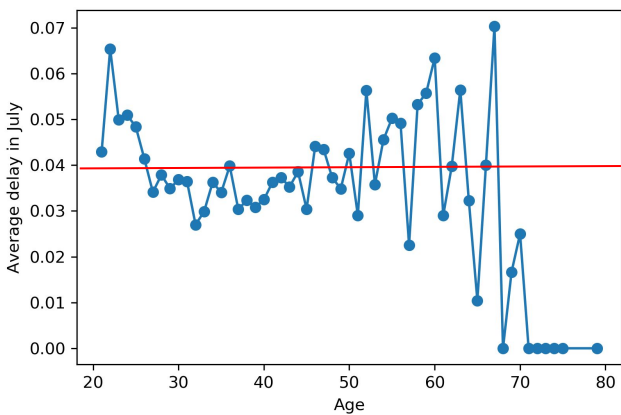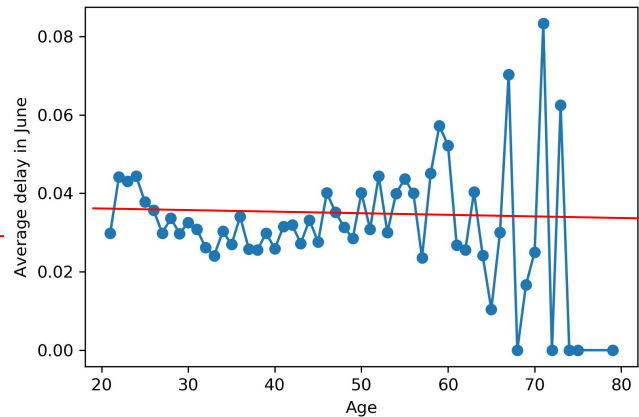- 52 features after preprocessing
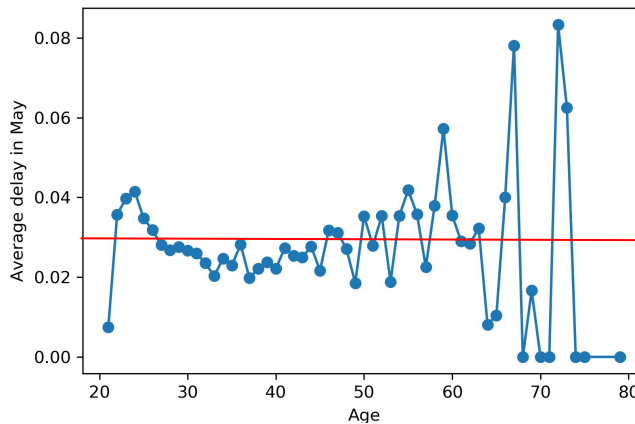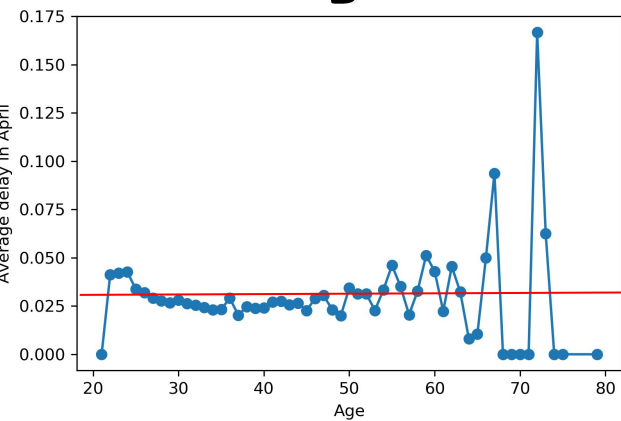

Proportion in 2 classes
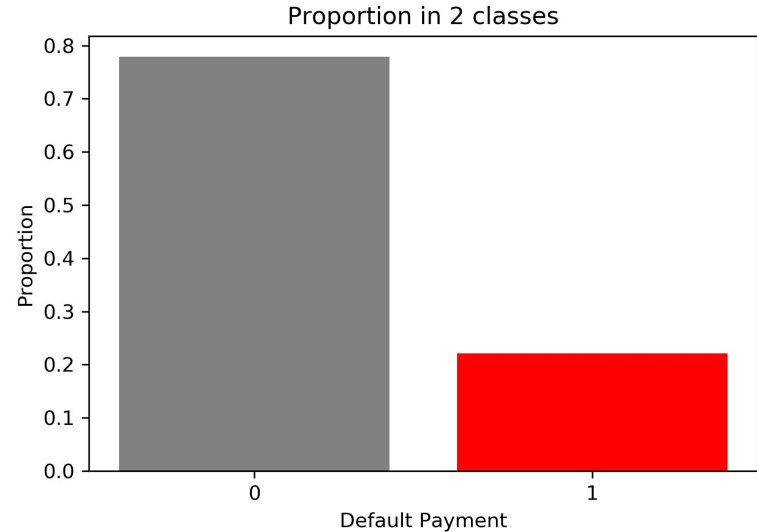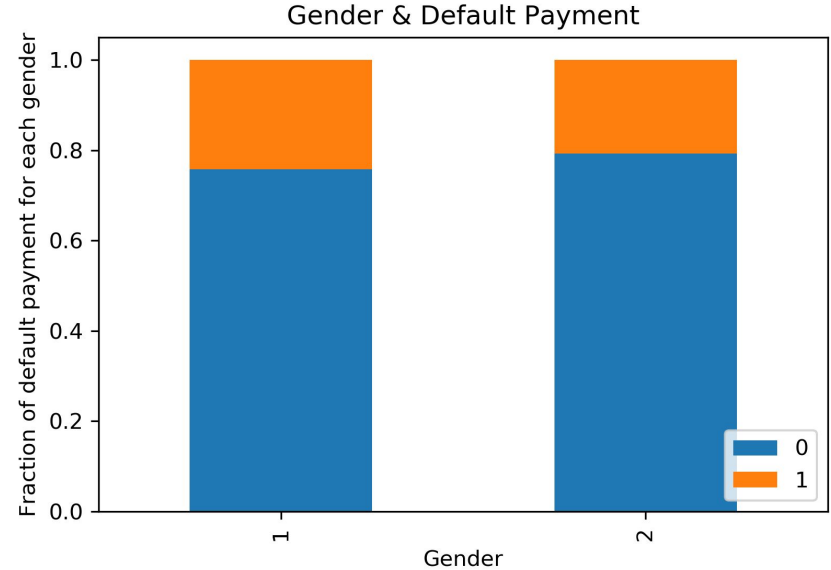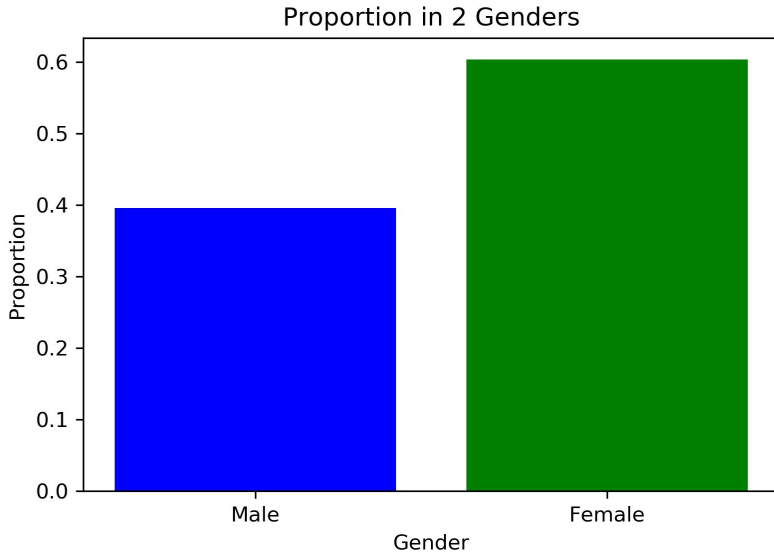
EDA: Age and Default Payment Status
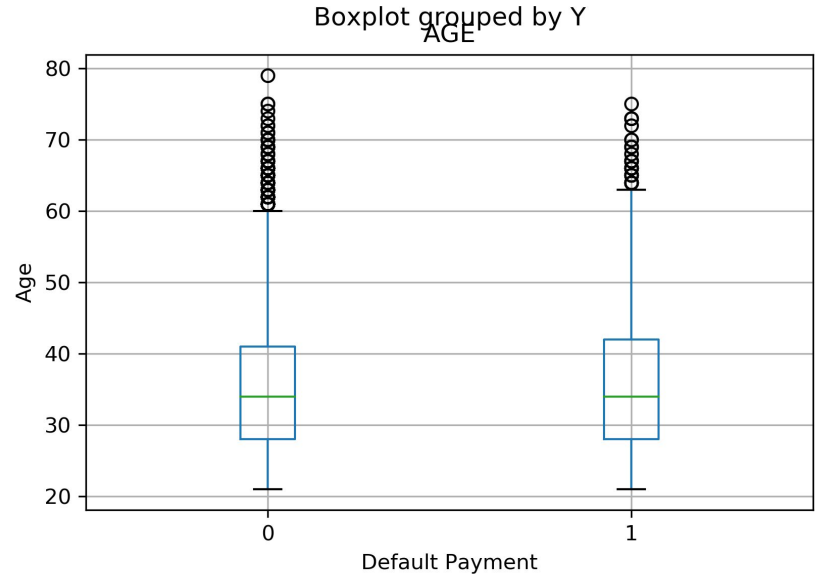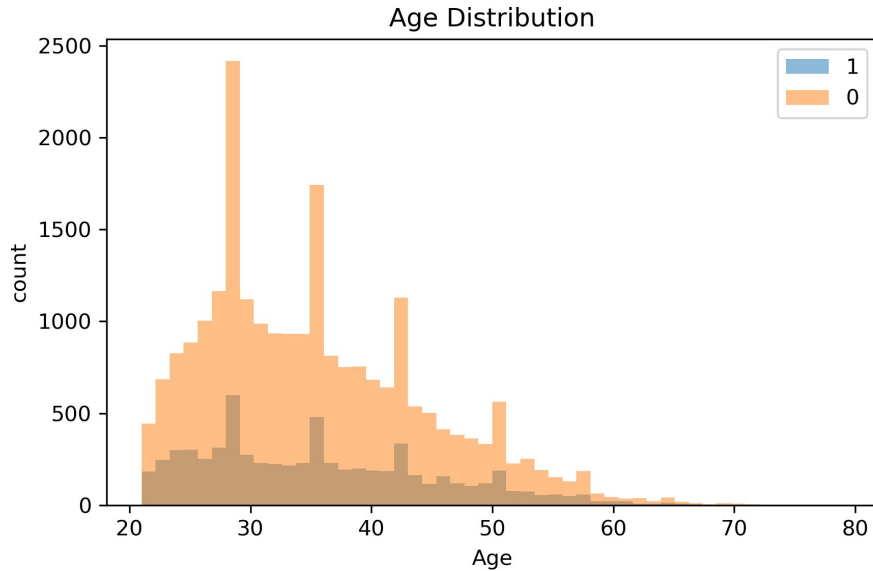
- Age is right-skewed; most of the people are young or mid-aged.
- No significant difference between default payment status and age.

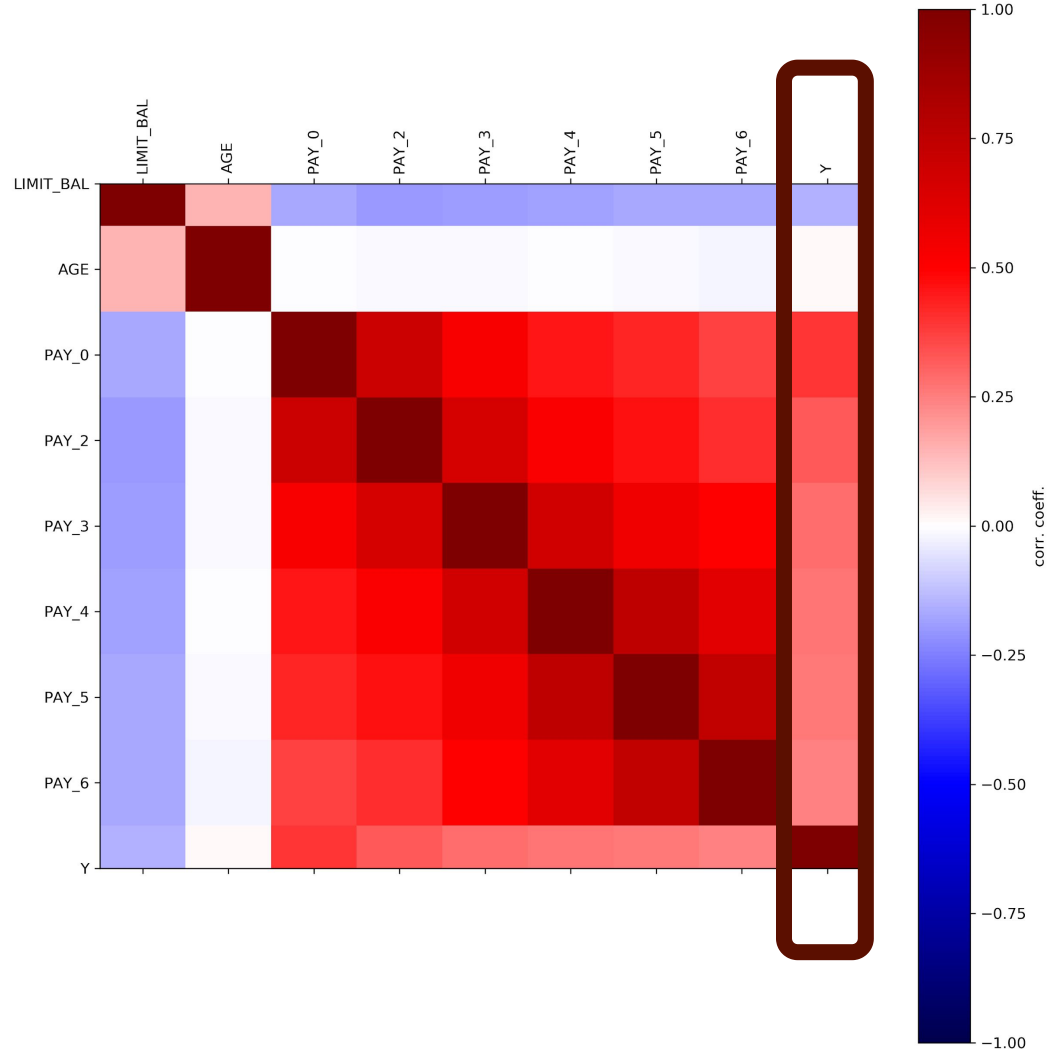# EDA: Age vs. Average Duration of Delayed Payment

EDA: Gender and Default Payment Status

Proportion in 2 Genders

Gender & Default Payment

- More females than males in the dataset
- Males tend to have more default payment

EDA: Correlation Matrix

Key point:
More mount of given credit (limited balance) leads to less chance of default.

Thank You