

# Credit Risk

*Filip Mordarski & Mateusz Wasielewski*

*5 12 2020*

## Spis treści

Wstęp 1

Tworzenie zmiennych objaśniających 1

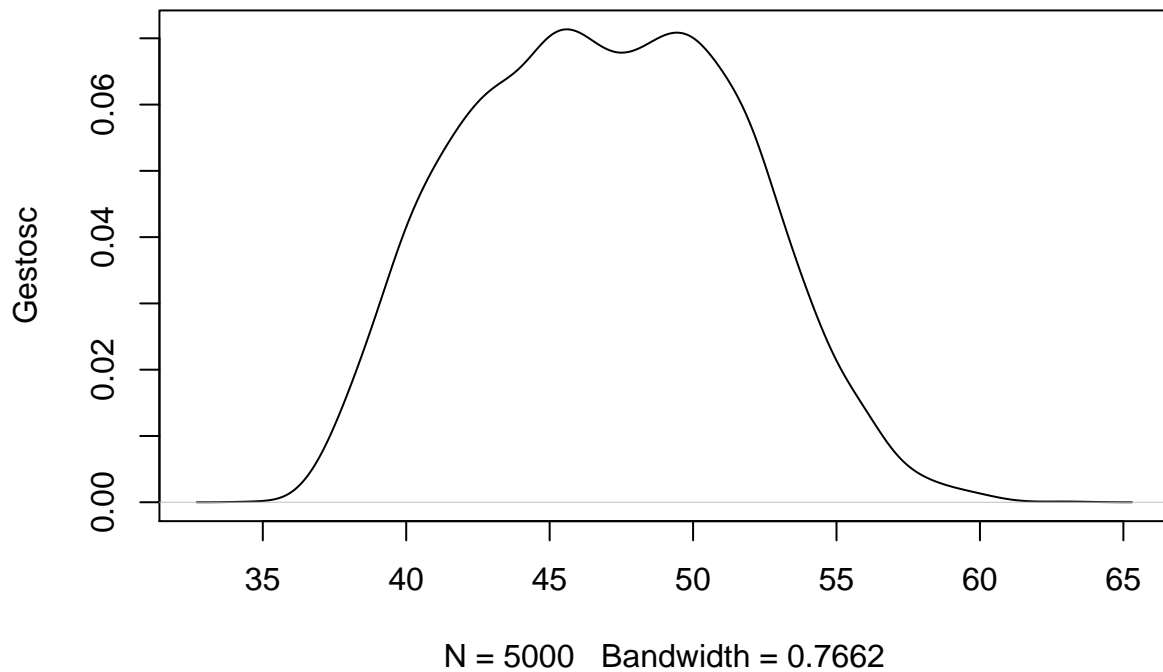
## Wstęp

Poniższy raport będzie zawierał analizę modelu ryzyka kredytowego.

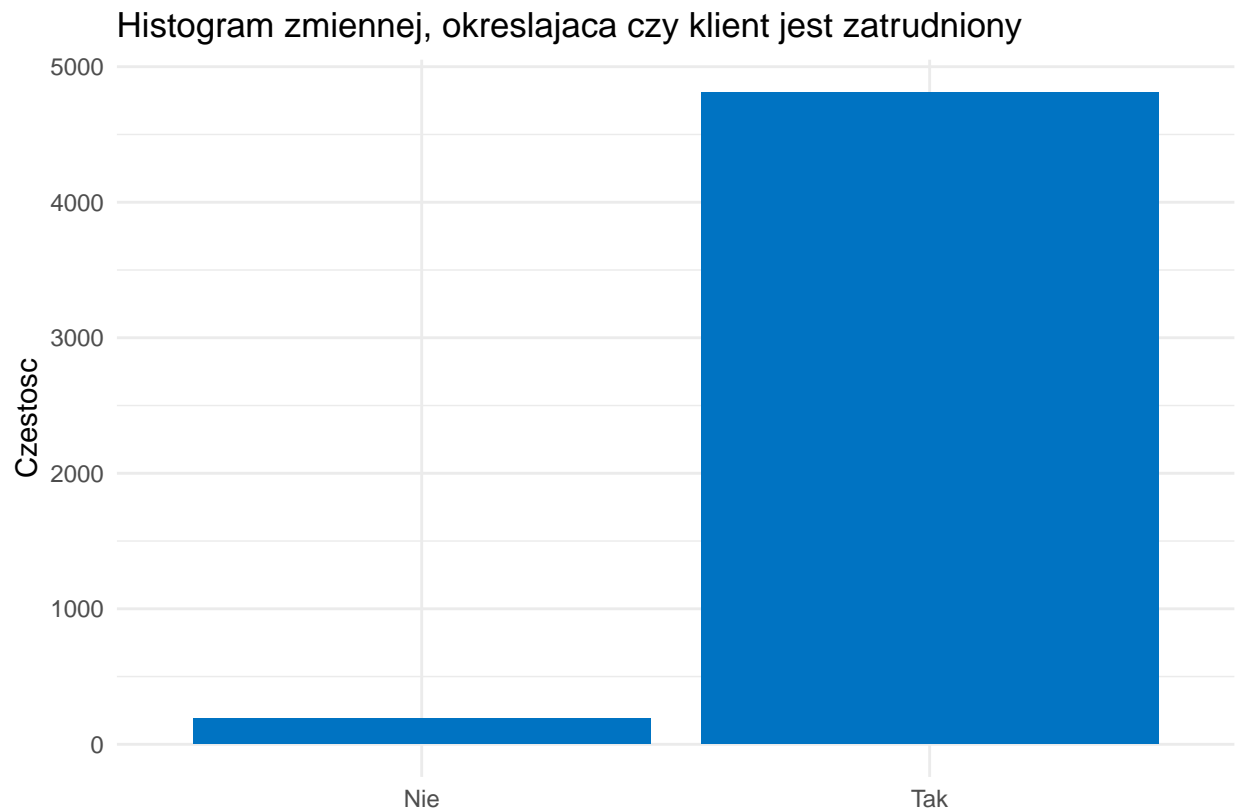
## Tworzenie zmiennych objaśniających

W pierwszej kolejności zostały utworzone zmienne objaśniające potrzebne do modelu PD. Pierwszą zmienną, która została wygenerowana na podstawie utworzonych wcześniej zmiennych jest wiek. Został wyliczony okres trwania umowy w latach na podstawie różnicy między obecną datą a wartością w zmiennej *agreement\_start*. Następnie został wygenerowany wektor wartości z rozkładu gamma z parametrem kształtu równym 3 oraz parametrem skali równym 2. Wiek został wyznaczony dodając do siebie: czas trwania umowy, liczbę 18 (wiek kiedy człowiek może podpisać wiążącą umowę kredytową) oraz wylosowaną wartość z rozkładu gamma, oznaczającą różnicę w latach pomiędzy datą podpisania umowy a osiągnięciem pełnoletności. Poniższy wykres przedstawia gęstość tej utworzonej zmiennej w naszym zbiorze.

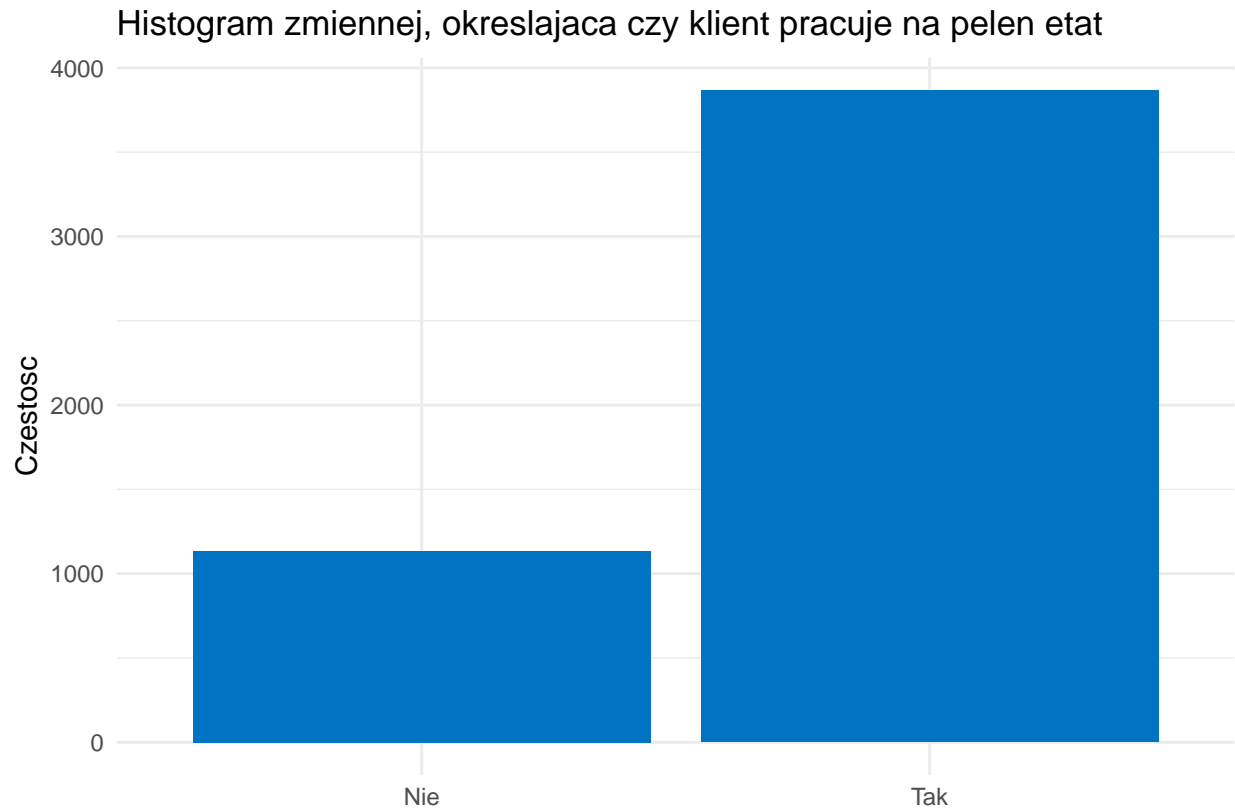
### Wykres gestosci wieku



Następnie została wygenerowana zmienna, która określa czy dany pracownik jest zatrudniony, czy też nie. Prawdopodobieństwo bezrobocia zostało ustalone na poziomie 3.6 %. Wartość ta odzwierciedla średnią stopę bezrobocia w 2019 roku w Stanach Zjednoczonych. Poniższy wykres przedstawia histogram tej zmiennej w zbiorze.

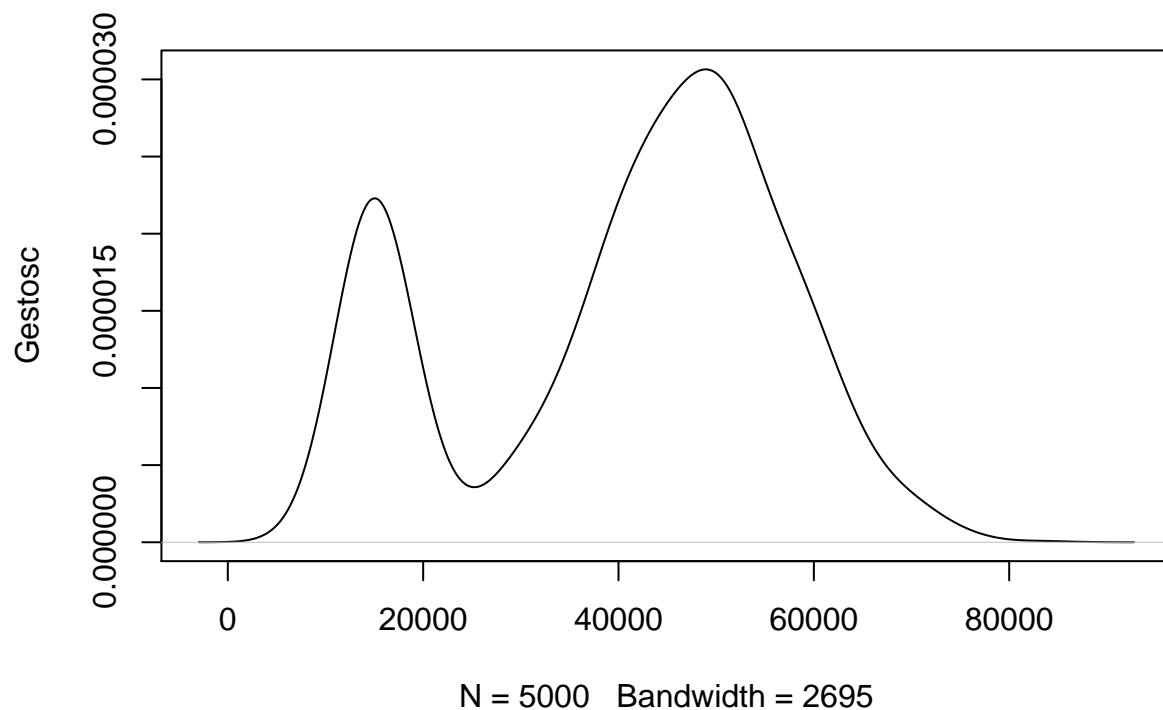


Na podstawie zmiennej, określającej czy dana osoba jest zatrudniona, wygenerowano zmienną czy dana osoba jest zatrudniona na pełny etat. Prawdopodobieństwo tego wynosi 80 %. Poniższy wykres przedstawia histogram tej zmiennej.



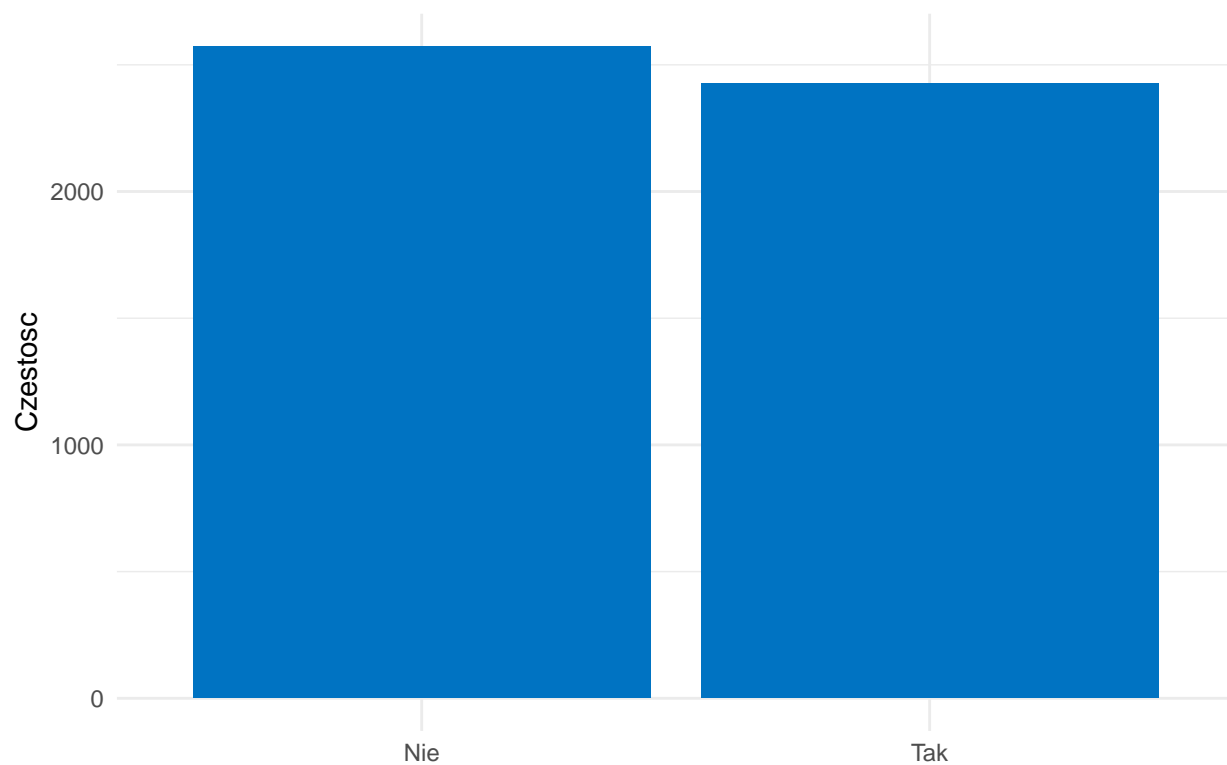
Zmienną, która z pewnością może okazać się istotna w tworzeniu modelu PD jest dochód roczny danej osoby. Wartości te zostały wylosowane z rozkładu normalnego. Średnia dla osób zatrudnionych na pełen etat została ustalona na poziomie 48000 USD z odchyleniem standardowym na poziomie 10000 USD. Dla osób niezatrudnionych na pełen etat wartość średnia została ustalona na poziomie 15000 USD, natomiast odchylenie 3000 USD. Poniżej zaprezentowano wykres gęstości tej zmiennej.

### Wykres gestosci zarobków

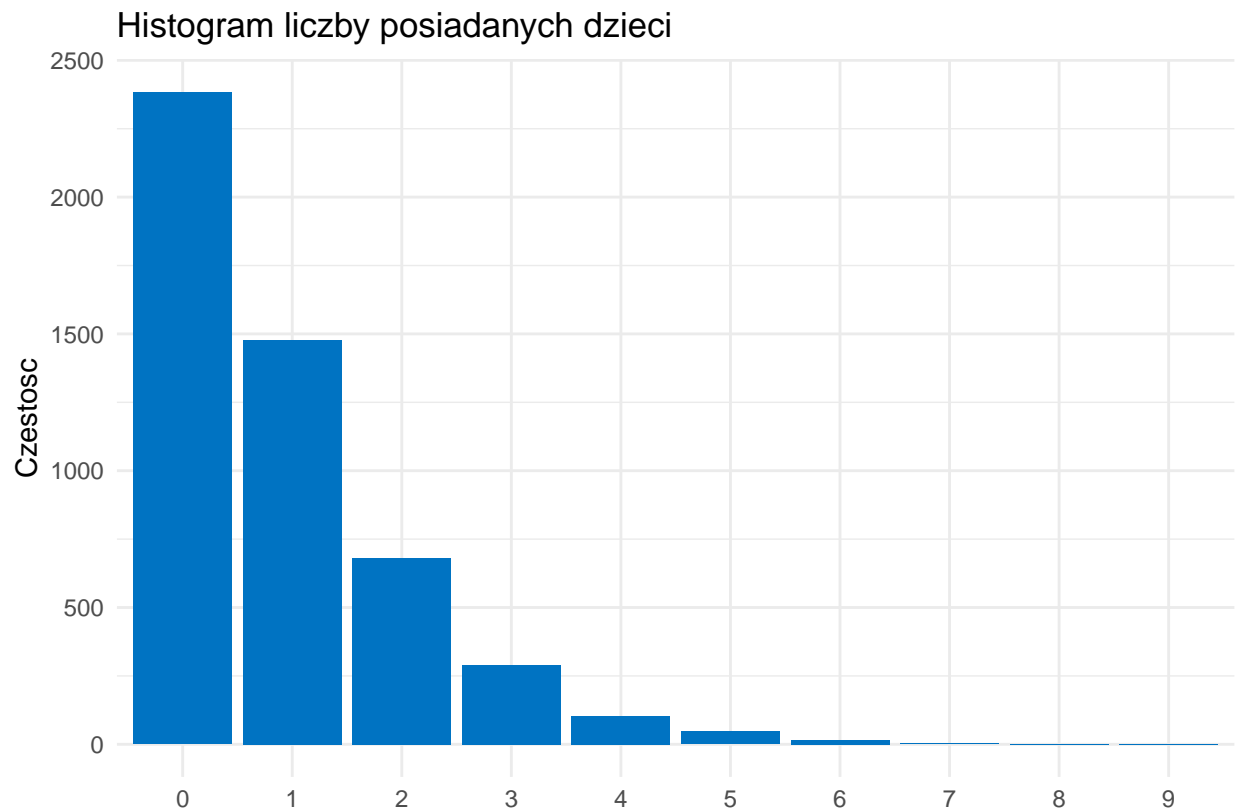


Kolejno, została wygenerowana zmienna, mówiąca o tym czy dana osoba jest singlem, czy żyje w związku z inną osobą. Prawdopodobieństwo, że ktoś jest singlem w Stanach Zjednoczonych wynosi 50.2 %. Przedstawiono histogram tej zmiennej w zbiorze.

Histogram zmiennej, określająca czy klient jest singlem

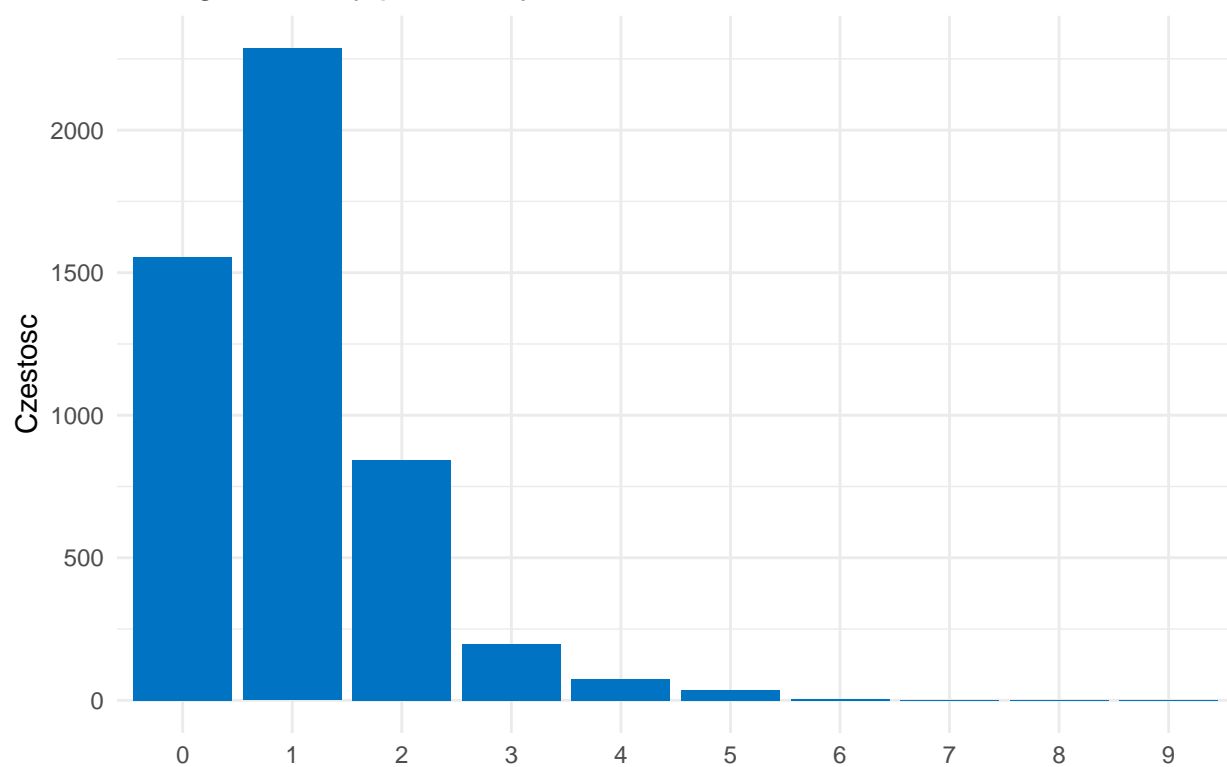


Następnie, została wygenerowana liczba posiadanych dzieci przez daną osobą. Zmienna ta została wygenerowana na podstawie zmiennej, określającej czy dana osoba jest singlem czy nie jest. Poniżej przedstawiono histogram tej wygenerowanej zmiennej.



Kolejno, została wygenerowana zmienna, mówiąca o liczbie posiadanych samochodów przez klienta. Dla osób nieposiadających dzieci lub mających jedno dziecko, liczba ta została wylosowana z następującego zakresu:  $[0, 1, 2]$  z prawdopodobieństwami równymi kolejno:  $[40\%, 50\%, 10\%]$ . Dla klientów mających więcej niż jedno dziecko, liczba samochodów zależy od liczby posiadanych dzieci i jest obliczona za pomocą następującej formuły:  $\text{liczba\_dzieci} - [\text{wartość z losowania liczb } [1,2] \text{ z } 50\% \text{ prawdopodobieństwami}] + 1$ . Poniżej przedstawiono histogram liczby posiadanych samochodów przez klientów.

Histogram liczby posiadanych samochodów



Następnie, została wygenerowana zmienna, określająca sektor gospodarki, w którym pracują klienci. Prawdopodobieństwo wystąpienia następujących sektorów [rolnictwo, przemysł, usługi] wśród zatrudnionych wynosi kolejno: [20%, 30%, 50%]. Poniżej przedstawiono histogram sektorów gospodarki.



Histogram sektorów gospodarki

