# Analityk kredytowy w Dream Housing Finance Company

Dream Housing Finance Company oferuje wszelkiego rodzaju pożyczki i kredyty. Obecne są na wszystkich obszarach miejskich, pół miejskich i wiejskich. Klient najpierw aplikuje o pożyczkę bądź kredyt, a następnie analityk kredytowy podejmuje ostateczną decyzją o przyznaniu klientowi pożyczki.

Firma chce zautomatyzować proces przyznawania pożyczek (w czasie rzeczywistym) na podstawie szczegółów klienta dostarczonych podczas wypełniania formularza aplikacyjnego online. Te szczegóły to płeć, stan cywilny, edukacja, liczba osób pozostających na utrzymaniu, dochód, kwota pożyczki, historia kredytowa i inne. Aby zautomatyzować ten proces, dostarczyli zestaw danych o dotychczasowych decyzjach przyznania i odrzucenia wniosku kredytowego dla ponad 600 klientów. Potrzebują, aby przeanalizować dane i przygotować raport jak poszczególne zmienne wpływają na decyzję kredytową.

# Dane

**Train:**PlikCSVzawierający dane o klientach z nadanym statusem przyznania bądź nie pożyczki 'Loan\_Status'

|  |  |
| --- | --- |
| **Zmienna** | **Opis** |
| Loan\_ID | ID wniosku |
| Gender | Płeć (Male/ Female) |
| Married | Czy wnioskodawca jest w związku małżeńskim (Y/N) |
| Dependents | Liczba osób na utrzymaniu (0,1,2,+3) |
| Education | Poziom wykształcenia (Graduate/ Under Graduate) |
| Self\_Employed | Praca na własny rachunek (Y/N) |
| ApplicantIncome | Dochód wnioskodawcy |
| CoapplicantIncome | Dochód współwnioskodawcy |
| LoanAmount | Wartość pożyczki w tys. |
| Loan\_Amount\_Term | Okres kredytowania w miesiącach |
| Credit\_History | Pozytywna historia kredytowa (Y/N) |
| Property\_Area | Lokalizacja nieruchomości (Urban/ Semi Urban/ Rural) |
| Loan\_Status | Czy przyznano kredyt (Y/N) |

# Zadania

Scharakteryzuj zmienne *ApplicantIncome*, *CoapplicantIncome*, *LoanAmount* w zakresie: minimum, pierwszy kwartyl, medina, średnia, trzeci kwartyl, maksimum, odchylenie standardowe, wybrana miarę asymetrii i koncentracji w każdej z grup: według płci (*Gender*) i według stanu cywilnego (*Married*).

Oblicz jak często pożyczkę dostawały osoby *Married=Y* a jak często osoby *Married=N.* (tabela liczności).

Oblicz jak często pożyczkę dostawały osoby *Self\_employed=Y* a jak często osoby *Self\_employed=N.* (tabela liczności).

Dla wybranych zmiennych zrób wykres ramka-wąsy (boxplot). Niech zmienna *Gender* będzie tworzyła serie. Dodał tytuł, opis osi, legendę.

Dla wybranych zmiennych zrób wykres ramka-wąsy (boxplot). Niech zmienna *Married* będzie tworzyła serie. Dodał tytuł, opis osi, legendę.

Dla pary zmiennych *ApplicantIncome*, *CoapplicantIncome* zrób wykres rozrzutu. Niech zmienna *Self\_employed* będzie tworzyła serie. Dodał tytuł, opis osi, legendę.

Dla par zmiennych *ApplicantIncome*, *LoanAmount* zrób wykres rozrzutu. Niech zmienna *Married* będzie tworzył serie. Dodał tytuł, opis osi, legendę.

Zbuduj 95% przedział ufności dla średniej i wariancji dla zmiennych *ApplicantIncome*, *LoanAmount* osobno dla kobiet i mężczyzn.

Sprawdź czy średni poziom zmiennej *ApplicantIncome* różni się pomiędzy grupą osób, które otrzymały pożyczkę i nie (zmienna *Loan\_status*)

Sprawdź czy wariancja zmiennej *ApplicantIncome* różni się pomiędzy grupą osób, które otrzymały pożyczkę i nie (zmienna *Loan\_status*)

# Model

Zbuduj model wstępnego decydowania czy firma powinna przyznać pożyczkę czy też nie. Oceń jakość modelu stosując walidację krzyżową, macierz pomyłek i statystyki na jej bazie tj. accuracy, precision, recall i f1-score. Spróbuj różnych modeli.