Raport z analizy danych — UCI Credit Card

# 1. Cel analizy

Celem projektu jest analiza danych kredytowych w celu zrozumienia zależności między cechami klienta a ryzykiem niespłacenia kredytu (default).

# 2. Opis zbioru danych

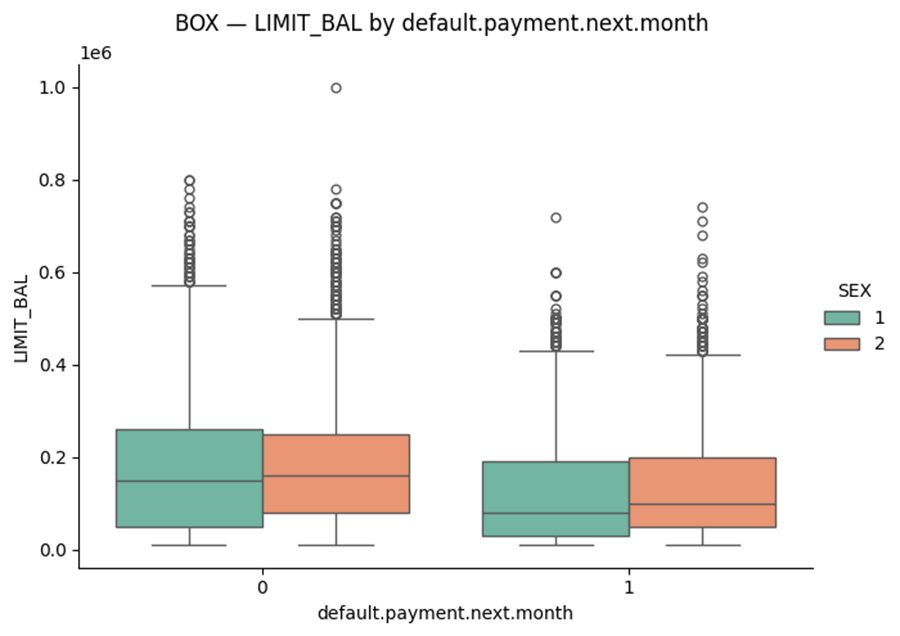
Zbiór danych zawiera 30 000 rekordów opisujących klientów kart kredytowych w Tajwanie. Zawiera informacje o limicie kredytowym, historii spłat, danych demograficznych, a także informację, czy klient zalegał ze spłatą w następnym miesiącu.

# 3. Wybrane cechy do analizy

Do szczegółowej analizy wybrano następujące cechy:

* • LIMIT\_BAL — wysokość limitu kredytowego
* • AGE — wiek klienta
* • EDUCATION — poziom wykształcenia
* • default.payment.next.month — zmienna docelowa (0 = brak zaległości, 1 = zaległość)

# 4. Wizualizacje i wnioski



**Boxplot: LIMIT\_BAL by default**

* Klienci bez zaległości mają średnio wyższy limit kredytowy.
* Cecha LIMIT\_BAL może być użyteczna w predykcji defaultu.

Obraz zawierający diagram, Wykres, origami

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

**Violinplot: AGE by default**

* Rozkład wieku dla obu klas defaultu jest podobny.
* Młodsi klienci dominują w zbiorze.
* Nie widać wyraźnej różnicy między grupami default = 0 i default = 1.
* Obraz zawierający tekst, diagram, zrzut ekranu, Wykres

  Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

**Countplot: EDUCATION**

* Najwięcej klientów ma wykształcenie na poziomie 1 i 2(graduate school, university).
* Kobiety dominują w grupie z wykształceniem poziomu 2 (university).

Obraz zawierający tekst, diagram, zrzut ekranu, Wykres

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

**Histogram: LIMIT\_BAL z podziałem na default**

* Klienci z niższym LIMIT\_BAL częściej znajdują się w grupie default = 1.
* Rozkład dla default = 0 przesunięty w stronę wyższych limitów.
* Sugeruje to zależność między wysokością limitu a wiarygodnością klienta.

Obraz zawierający tekst, zrzut ekranu, linia, wzór

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

**Heatmap**

# 5. Korelacja między cechami

Na podstawie mapy korelacji (heatmapy) zidentyfikowano najciekawsze zależności pomiędzy cechami numerycznymi. Poniżej przedstawiono 5 najważniejszych obserwacji wraz z krótką interpretacją:

• BILL\_AMT1 vs BILL\_AMT2 (i kolejne) (korelacja: ~0.95)

Bardzo silna dodatnia korelacja. Sugeruje, że klienci mają stabilne salda zadłużenia na przestrzeni miesięcy.

• PAY\_0 vs PAY\_2, PAY\_3 itd. (korelacja: ~0.7 – 0.8)

Wysoka korelacja świadczy o konsekwencji w opóźnieniach – jeśli klient spóźni się raz, prawdopodobnie zrobi to ponownie.

• default.payment.next.month vs LIMIT\_BAL (korelacja: -0.15)

Ujemna korelacja oznacza, że klienci z niższym limitem kredytowym częściej mają zaległości.

# • default.payment.next.month vs PAY\_0 (korelacja: ~0.32)

# Dodatnia korelacja wskazuje, że opóźnienia w ostatnim miesiącu zwiększają szansę na niespłacenie kolejnej raty.

# • default.payment.next.month vs PAY\_AMT1 (korelacja: -0.06)

# Słaba ujemna korelacja – osoby, które zapłaciły mniej, nieco częściej trafiają do grupy zalegających.

# 6. Regresja liniowa

Obraz zawierający tekst, zrzut ekranu, linia, diagram

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

📈 Regresja LIMIT\_BAL względem AGE

• Występuje lekka dodatnia zależność między wiekiem a limitem kredytowym.

• Klienci młodsi częściej mają niższe limity.

# 7. Wnioski końcowe

• Niektóre cechy są silnie skorelowane między sobą.

• Default nie jest zmienną zbalansowaną (większość klientów spłaca zobowiązania).

• Występują wartości odstające, które mogą mieć wpływ na analizę.

• Cecha LIMIT\_BAL oraz PAY\_0 mogą być dobrymi predyktorami defaultu.