

# Raport o analizie danych klientów kart kredytowych

Niniejszy raport przedstawia analizę zestawu danych dotyczących klientów kart kredytowych z Tajwanu, który zawiera informacje o ich zachowaniach płatniczych i demograficznych. Zestaw danych obejmuje około 30 000 rekordów i 25 atrybutów, takich jak limit kredytowy, płeć, wykształcenie, stan cywilny, wiek, historia płatności, kwoty rachunków oraz kwoty spłat, a także zmienną docelową wskazującą, czy klient zalegał z płatnością w następnym miesiącu. Celem analizy jest zidentyfikowanie wzorców i zależności, które mogą pomóc w przewidywaniu ryzyka niewypłacalności.

## Przegląd danych

Zestaw danych zawiera następujące kluczowe zmienne:

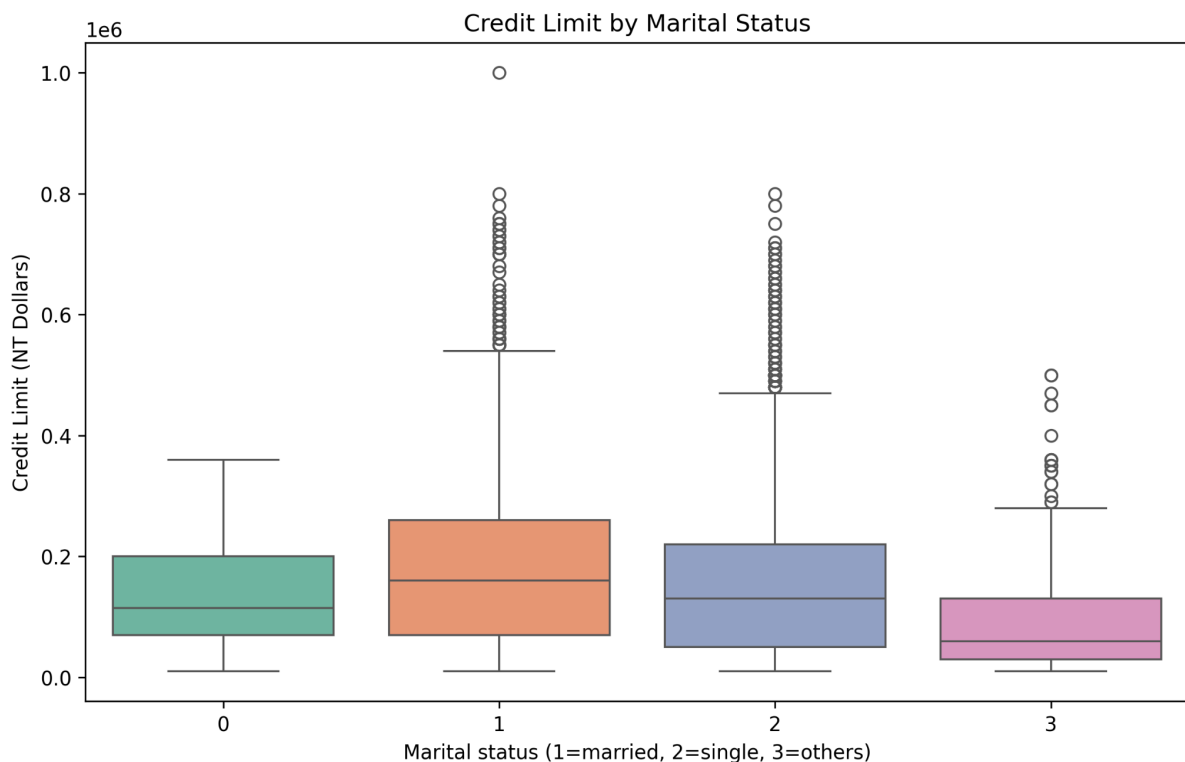
- **LIMIT\_BAL**: Limit kredytowy w dolarach NT.
- **SEX**: Płeć (1 = mężczyzna, 2 = kobieta).
- **EDUCATION**: Poziom wykształcenia (1 = szkoła wyższa, 2 = uniwersytet, 3 = szkoła średnia, 4 = inne).
- **MARRIAGE**: Stan cywilny (1 = żonaty/zamężna, 2 = samotny, 3 = inne).
- **AGE**: Wiek w latach.
- **PAY\_0 do PAY\_6**: Status spłaty od kwietnia do września (-1 = terminowa spłata, 1 = opóźnienie o jeden miesiąc itp.).
- **BILL\_AMT1 do BILL\_AMT6**: Kwoty rachunków od kwietnia do września.
- **PAY\_AMT1 do PAY\_AMT6**: Kwoty spłat od kwietnia do września.
- **default.payment.next.month**: Czy klient zalegał z płatnością w następnym miesiącu (1 = tak, 0 = nie).

# Analiza eksploracyjna danych

## Wykresy pudełkowe (Box Plots)

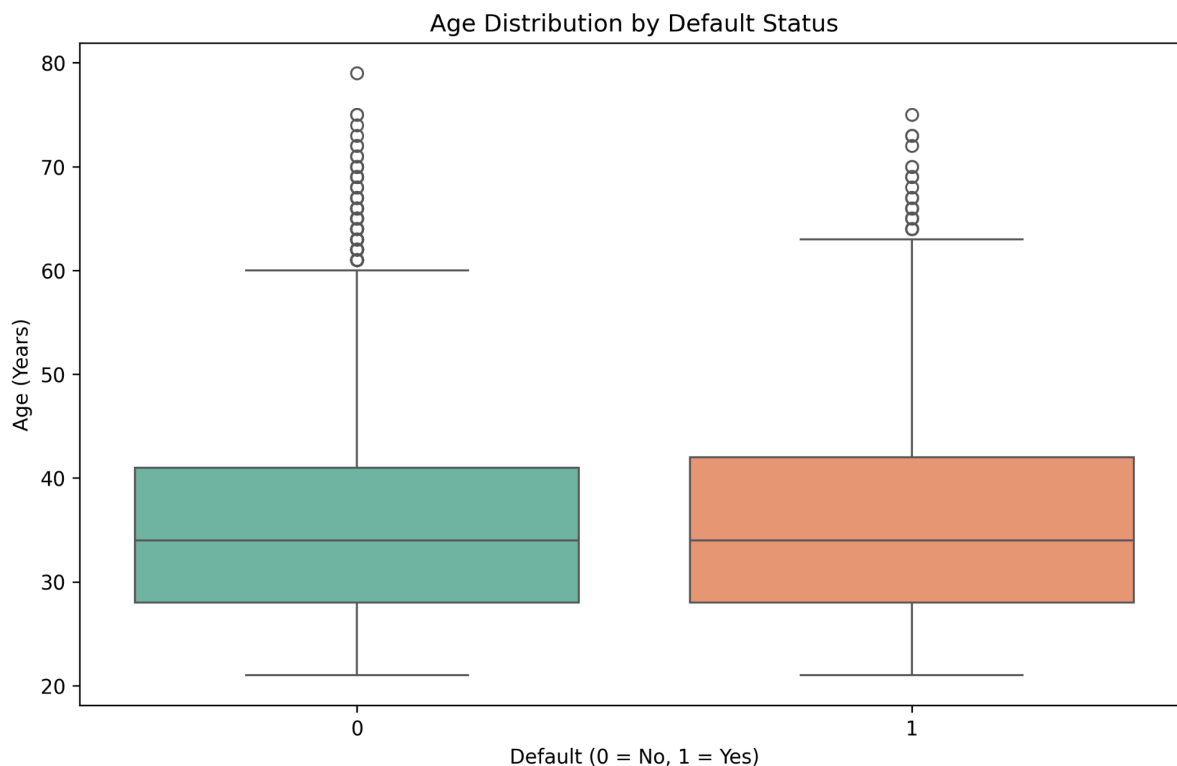
- Limit kredytowy względem stanu cywilnego**

Wykres pudełkowy przedstawia rozkład limitów kredytowych w zależności od stanu cywilnego. Osoby zamężne/żonate (MARRIAGE = 1) mają tendencję do posiadania wyższych median limitów kredytowych w porównaniu do osób samotnych (MARRIAGE = 2) i innych (MARRIAGE = 3). Może to sugerować, że stan cywilny wpływa na zdolność kredytową, być może ze względu na stabilność finansową w małżeństwie.



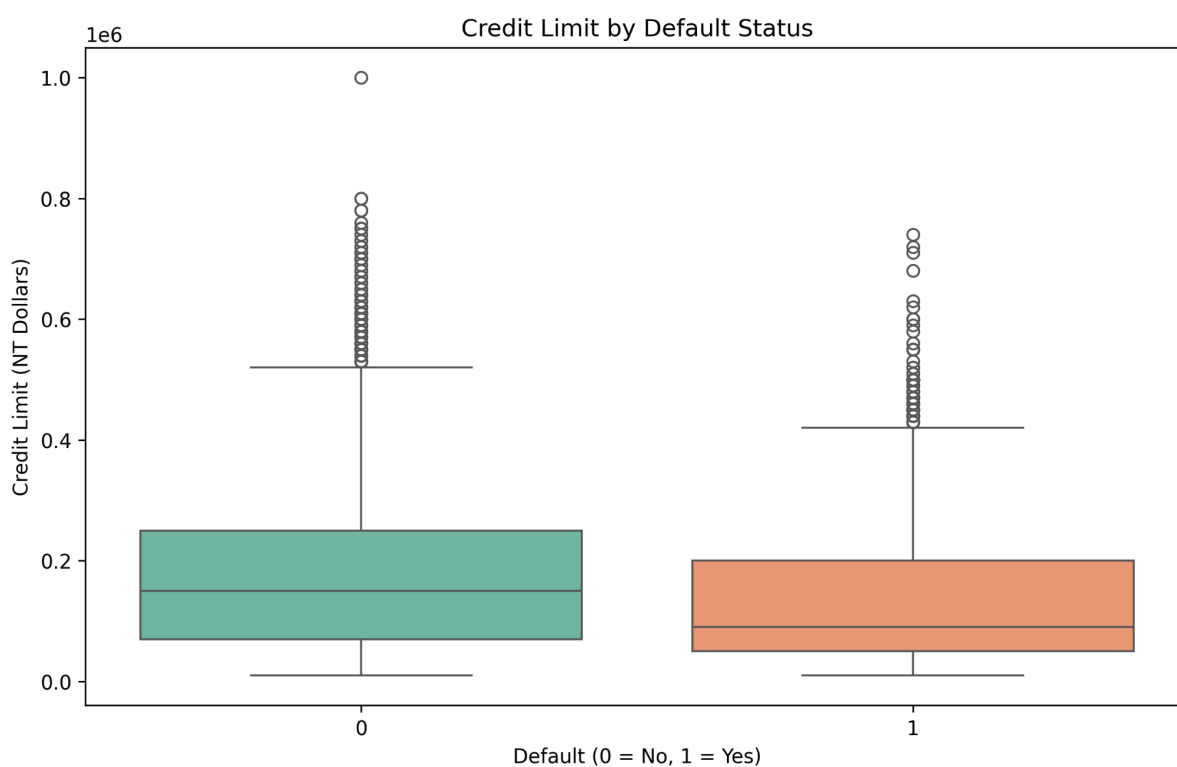
- Rozkład wieku względem statusu niewypłacalności**

Wykres porównuje rozkład wieku między klientami, którzy zalegali z płatnościami (default = 1), a tymi, którzy tego nie robili (default = 0). Mediana wieku wśród niewypłacalnych jest niższa, co może wskazywać, że młodszy klienci są bardziej narażeni na ryzyko niewypłacalności.



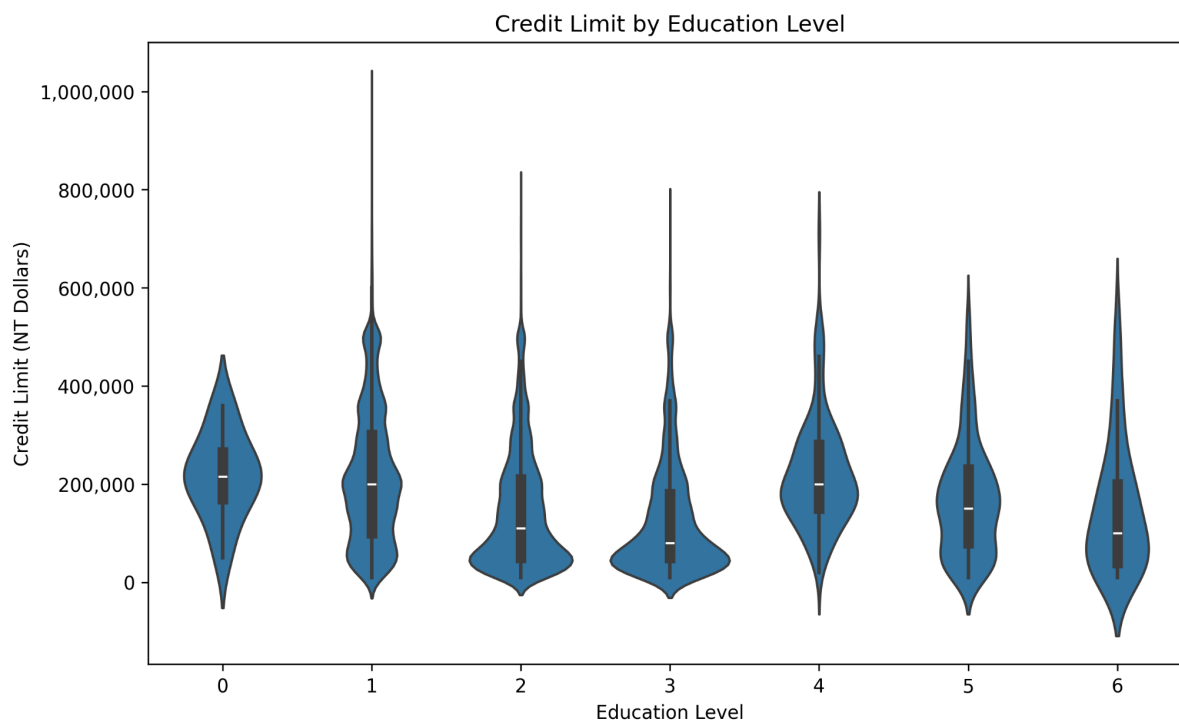
- **Limit kredytowy względem statusu niewypłacalności**

Wykres pokazuje, że klienci, którzy nie zalegali z płatnościami, mają wyższe mediany limitów kredytowych w porównaniu do niewypłacalnych. Może to oznaczać, że wyższe limity kredytowe są przyznawane klientom o niższym ryzyku defaultu.



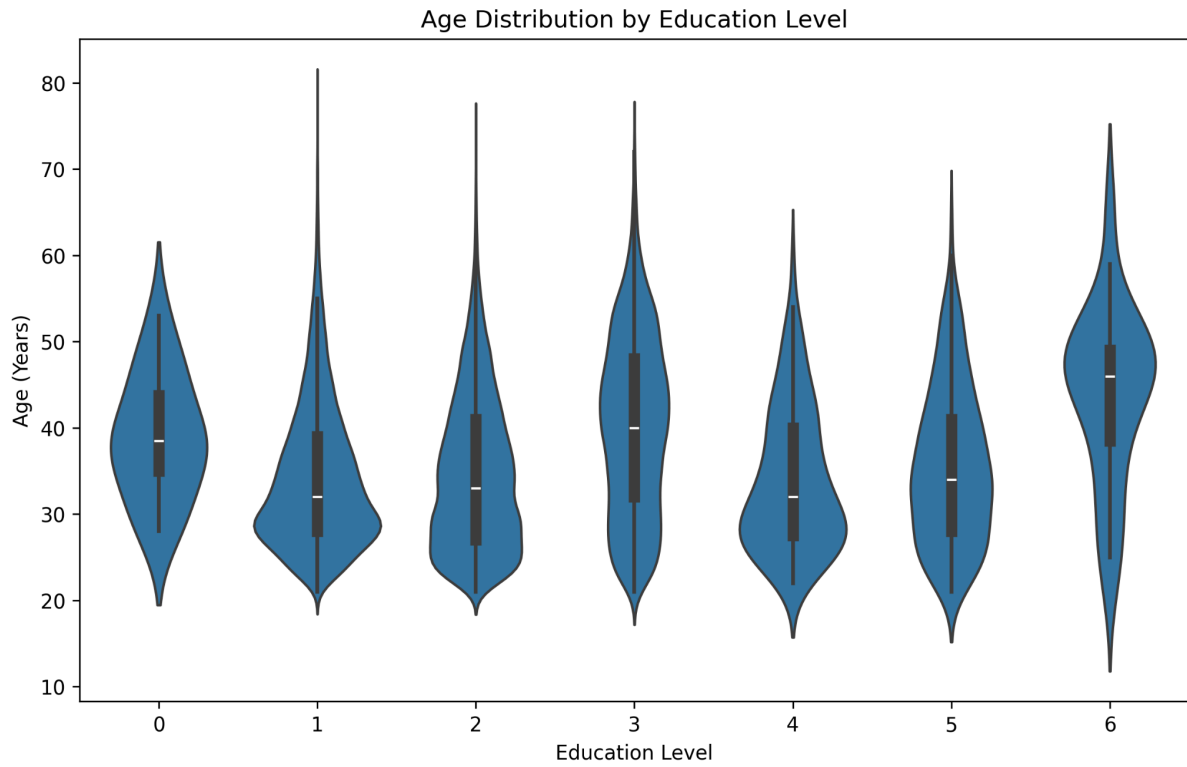
- **Limit kredytowy względem poziomu wykształcenia**

Wykres skrzypcowy ukazuje rozkład limitów kredytowych w zależności od poziomu wykształcenia. Klienci z wyższym wykształceniem (np. szkoła wyższa, EDUCATION = 1) mają szerszy rozkład i wyższe limity kredytowe, co może odzwierciedlać ich lepszą sytuację finansową.



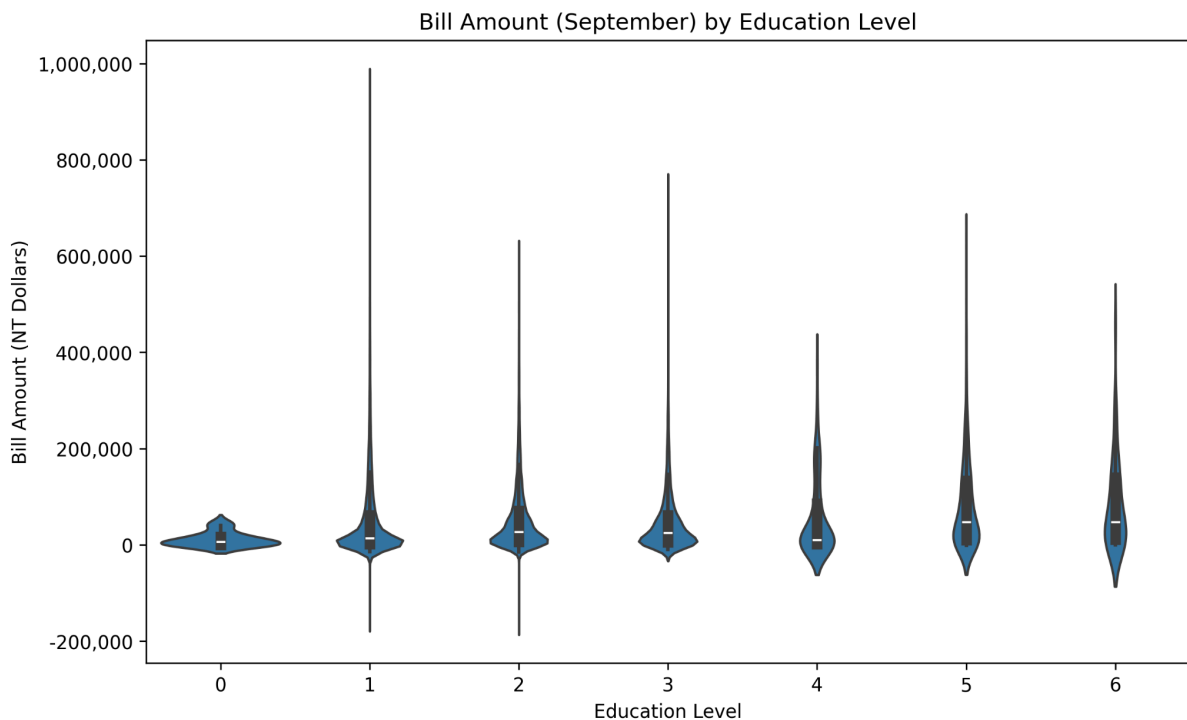
- **Rozkład wieku względem poziomu wykształcenia**

Wykres wskazuje, że klienci z wyższym wykształceniem są zazwyczaj starsi, co może wynikać z dłuższego czasu spędzonego na edukacji.



- **Kwota rachunku (wrzesień) względem poziomu wykształcenia**

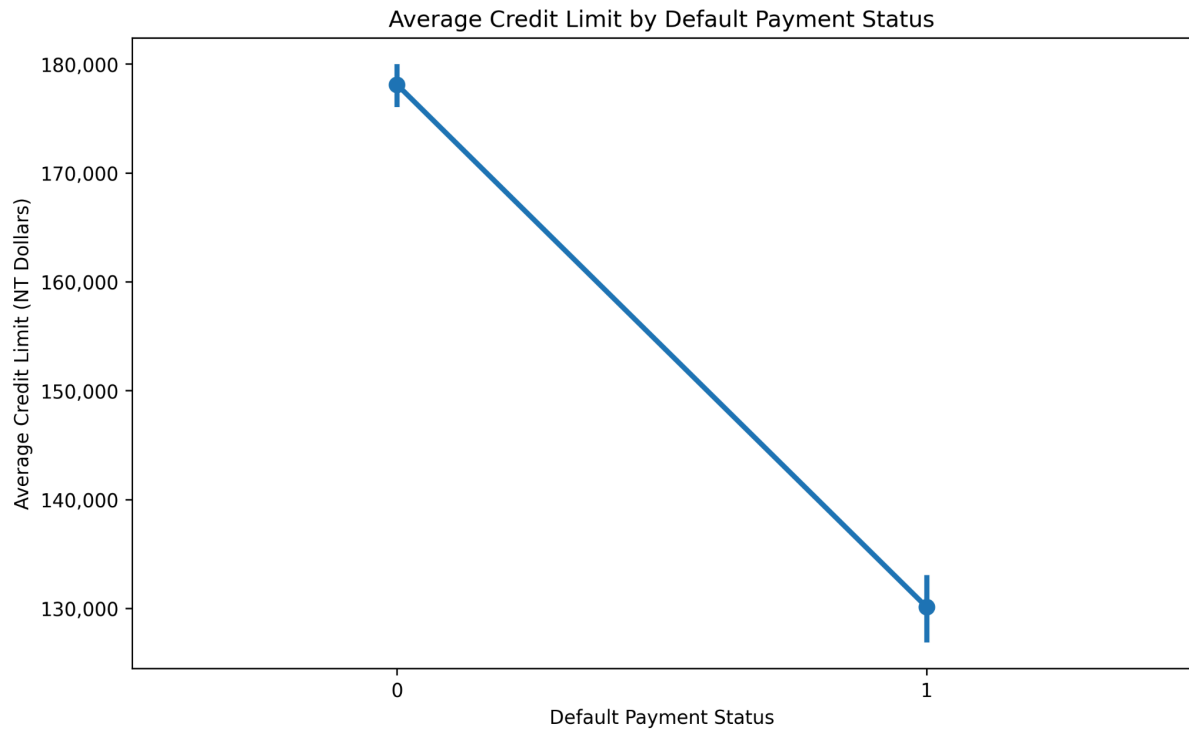
Rozkład kwot rachunków we wrześniu pokazuje, że osoby z wyższym wykształceniem mają wyższe rachunki, co może być związane z większymi wydatkami lub zdolnością kredytową.



## Wykresy słupkowe z przedziałami ufności (Error Bars)

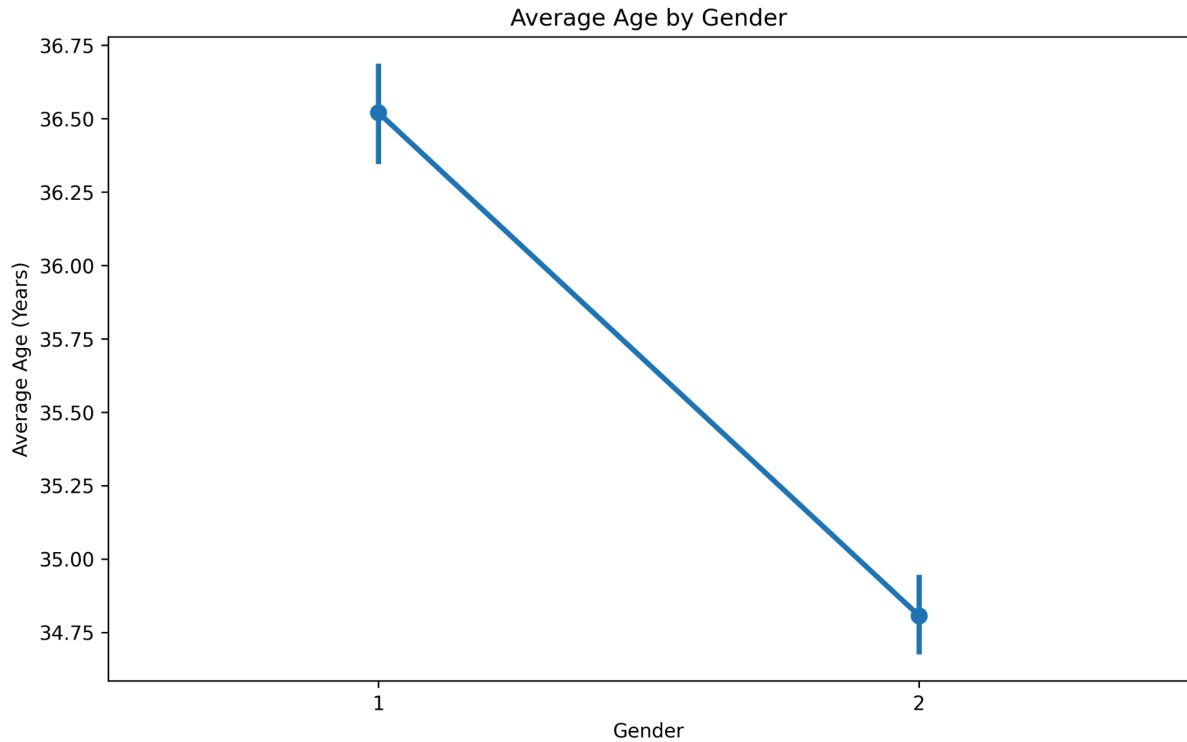
- Średni limit kredytowy względem statusu niewypłacalności

Wykres pokazuje średni limit kredytowy dla niewypłacalnych i niewypłacalnych z przedziałami ufności. Klienci niewypłacalni mają niższą średnią, co potwierdza wcześniejsze obserwacje.



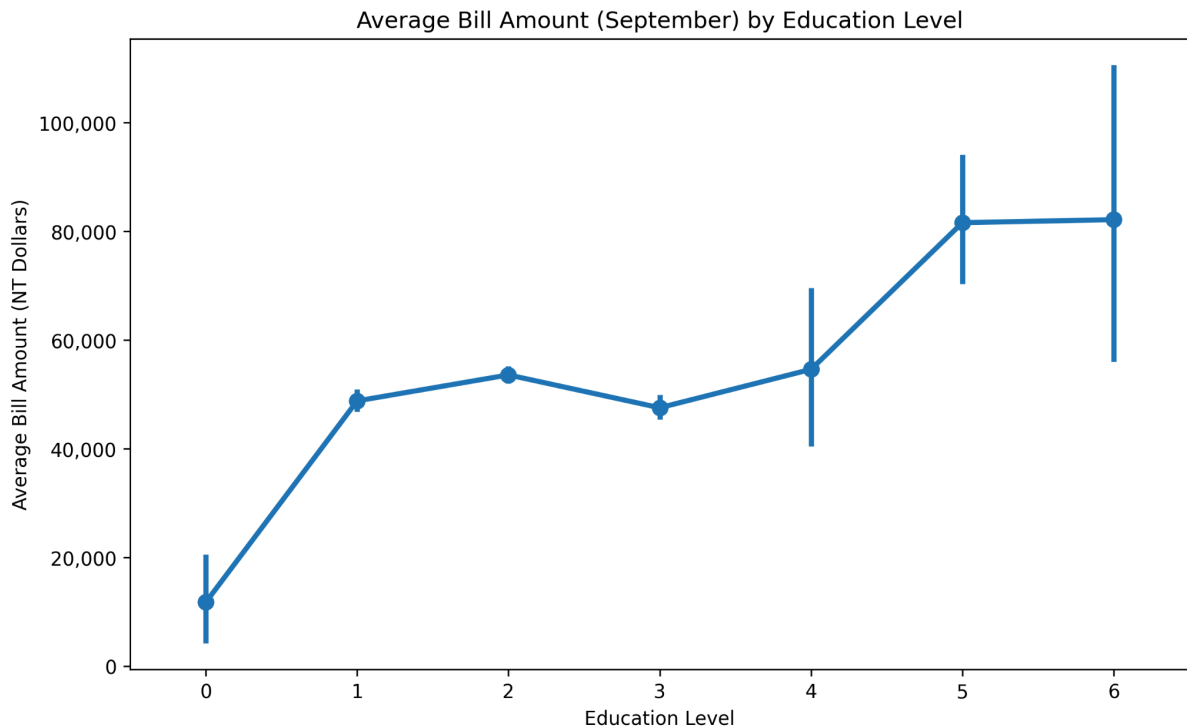
- **Średni wiek względem płci**

Średni wiek mężczyzn i kobiet jest podobny, z niewielkimi różnicami widocznymi w przedziałach ufności.



- **Średnia kwota rachunku (wrzesień) względem poziomu wykształcenia**

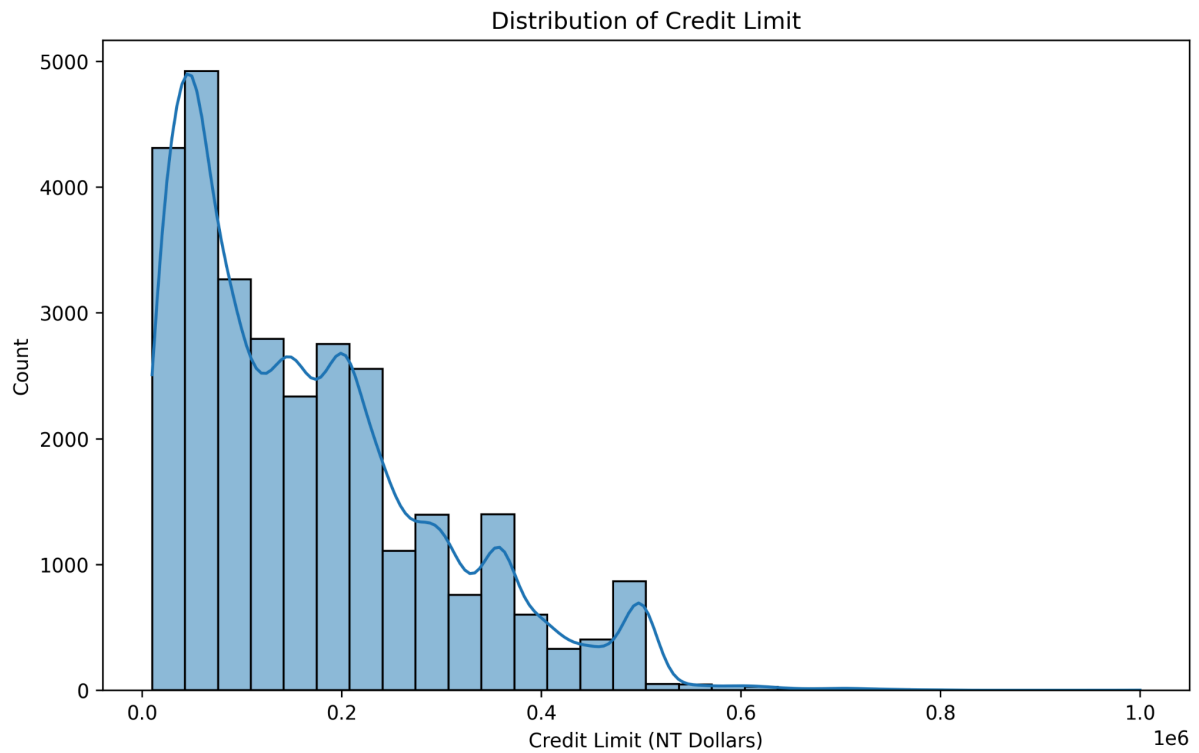
Wykres wskazuje, że średnie kwoty rachunków rosną wraz z poziomem wykształcenia, co jest spójne z wcześniejszymi wnioskami.



## Histogramy

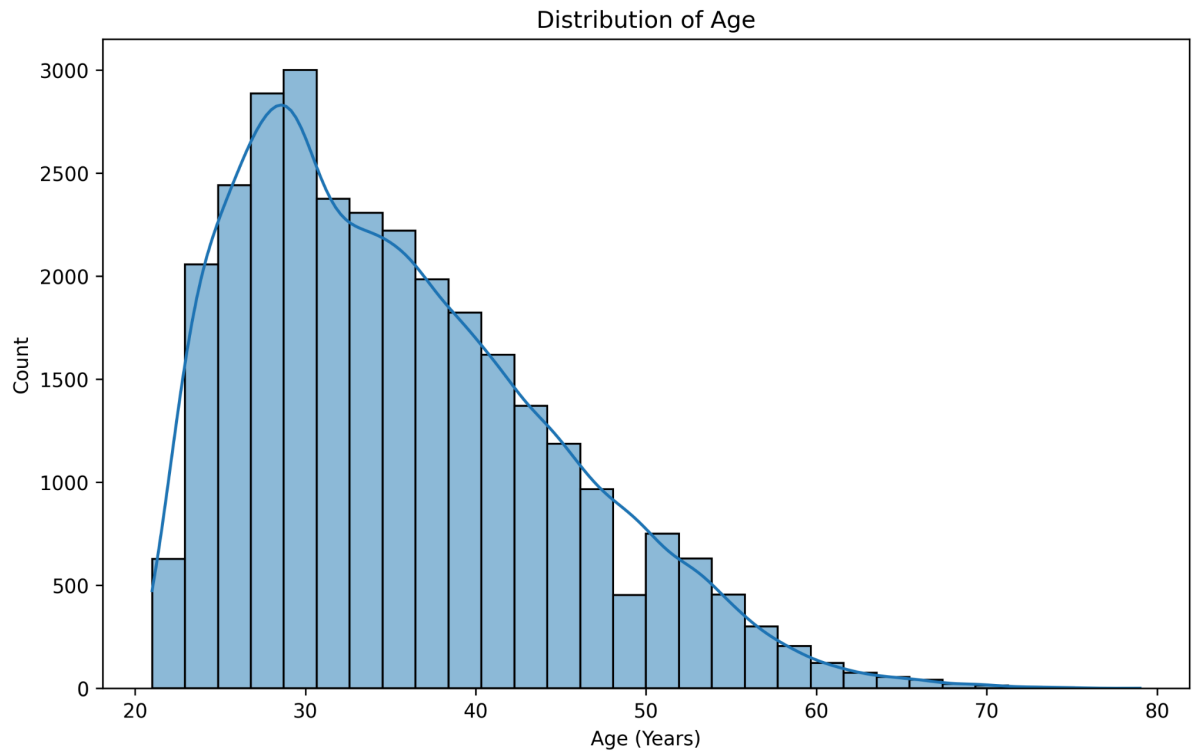
- **Rozkład limitu kredytowego**

Histogram pokazuje, że limity kredytowe są skewed w prawo, z większością klientów posiadających niższe limity i nielicznymi z bardzo wysokimi limitami.



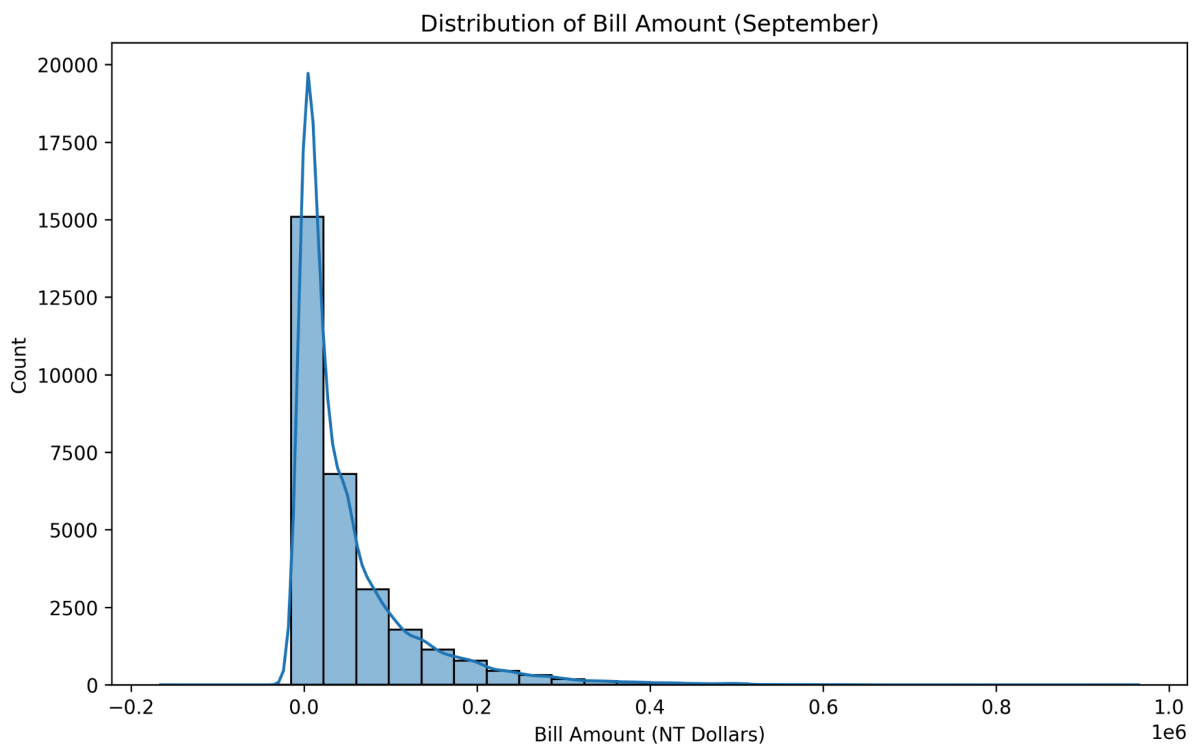


- Rozkład wieku jest zbliżony do normalnego, z największą koncentracją między 20 a 60 lat.



- **Rozkład kwoty rachunku (wrzesień)**

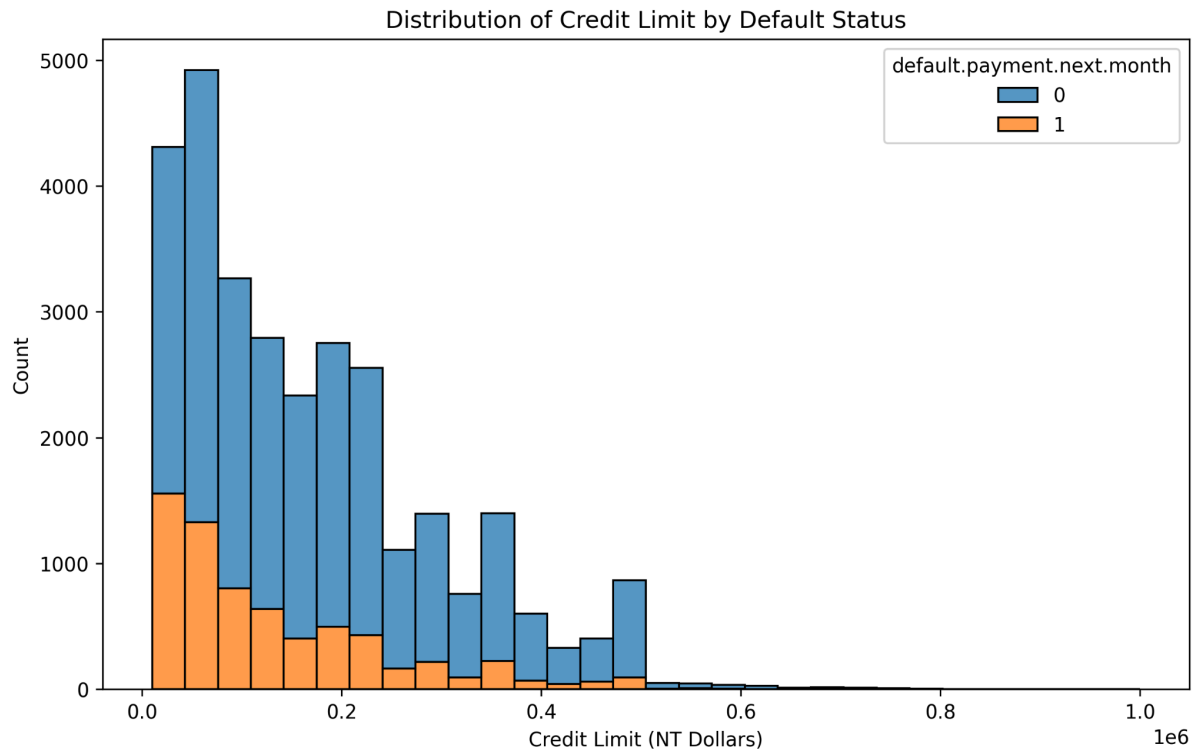
Kwoty rachunków we wrześniu są również skewed w prawo, z wieloma klientami mającymi niskie kwoty i nielicznymi z wysokimi rachunkami.



## Histogramy z odcieniem (Hue)

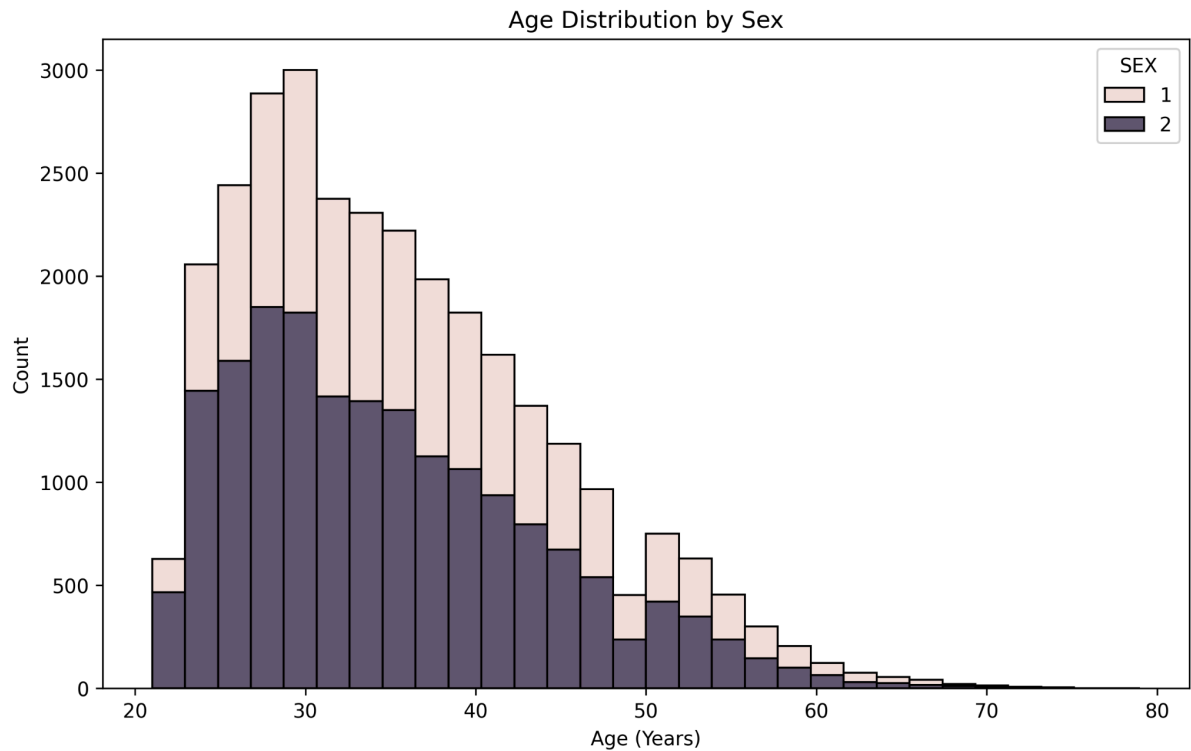
- Rozkład limitu kredytowego względem statusu niewypłacalności**

Niewypłacalni klienci mają niższe limity kredytowe w porównaniu do niewypłacalnych, co widoczne jest w przesunięciu rozkładów.



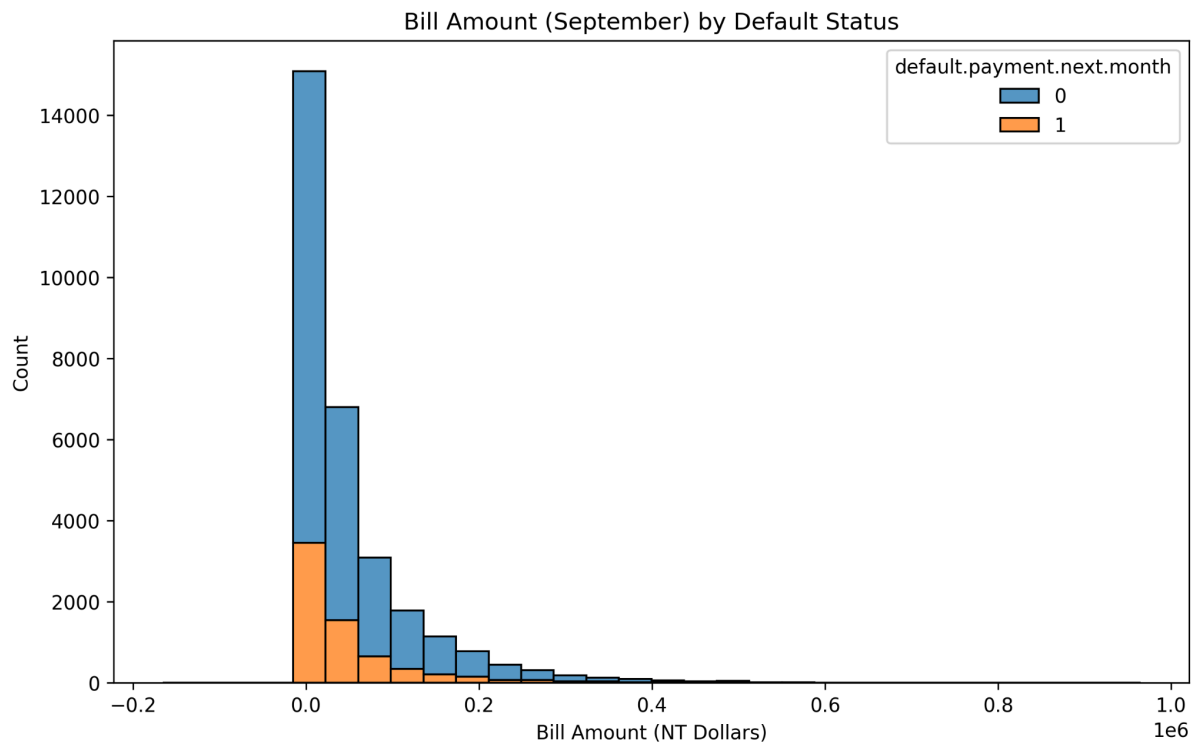
- **Rozkład wieku względem płci**

Rozkłady wieku dla mężczyzn i kobiet są podobne, z niewielkimi różnicami w szczytach.



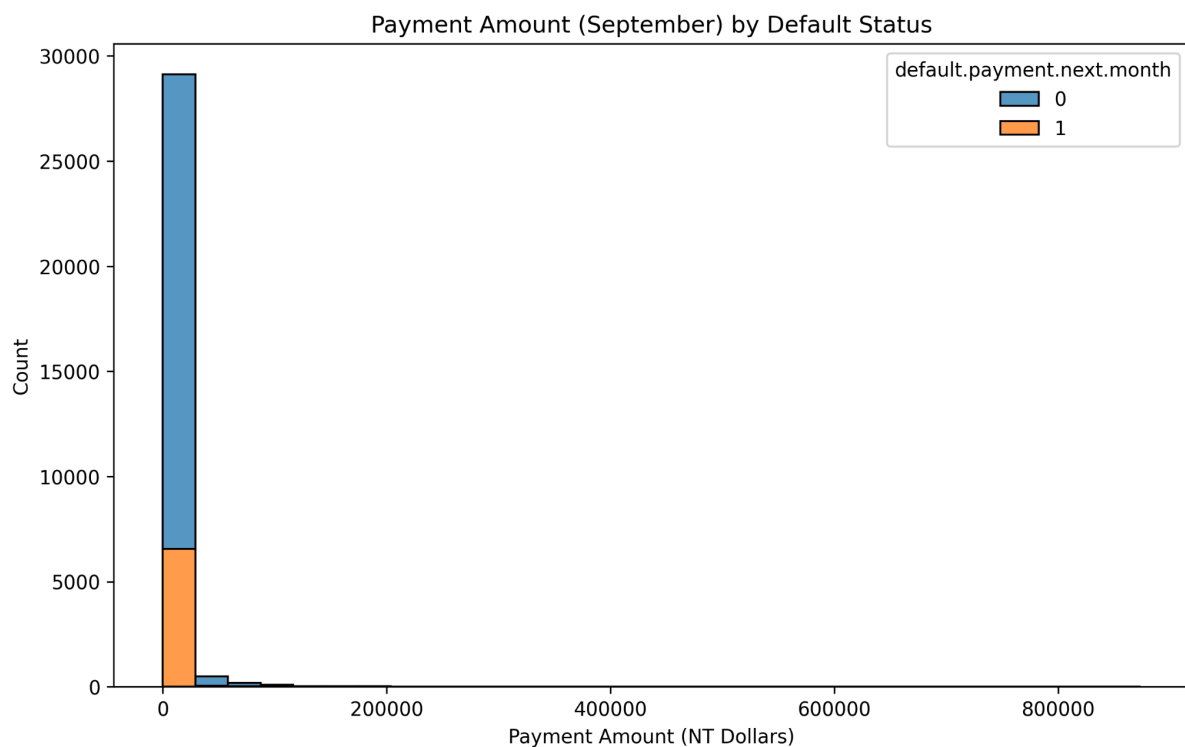
- **Kwota rachunku (wrzesień) względem statusu niewypłacalności**

Niewypłacalni mają wyższe kwoty rachunków, co może wskazywać na trudności w zarządzaniu długiem.



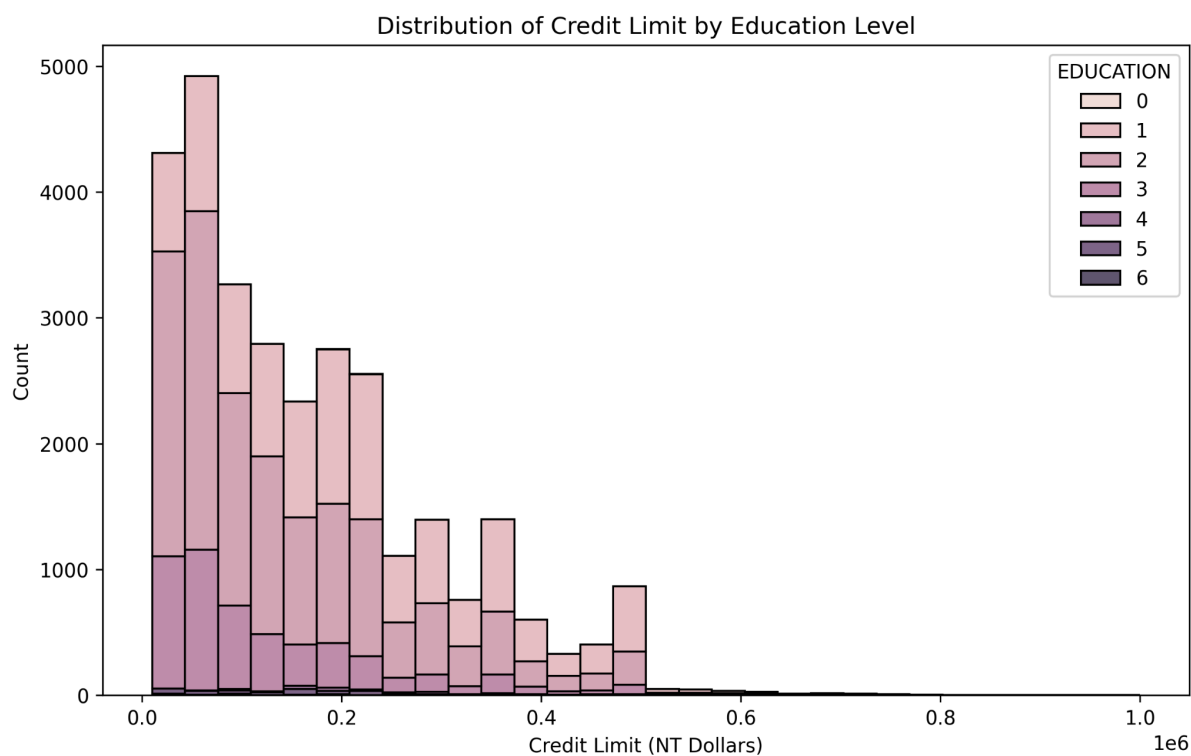
- **Kwota spłaty (wrzesień) względem statusu niewypłacalności**

Klienci niewypłacalni dokonują mniejszych spłat we wrześniu w porównaniu do niewypłacalnych.



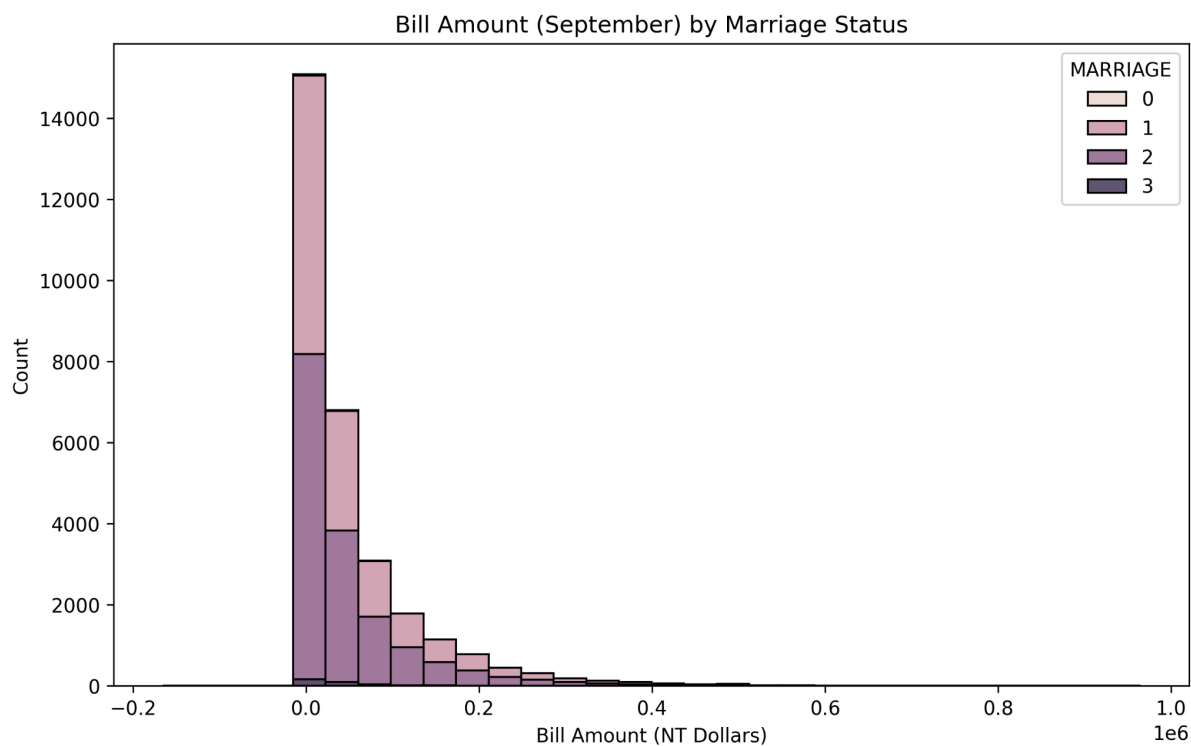
- **Rozkład limitu kredytowego względem poziomu wykształcenia**

Wyższe poziomy wykształcenia korelują z wyższymi limitami kredytowymi.



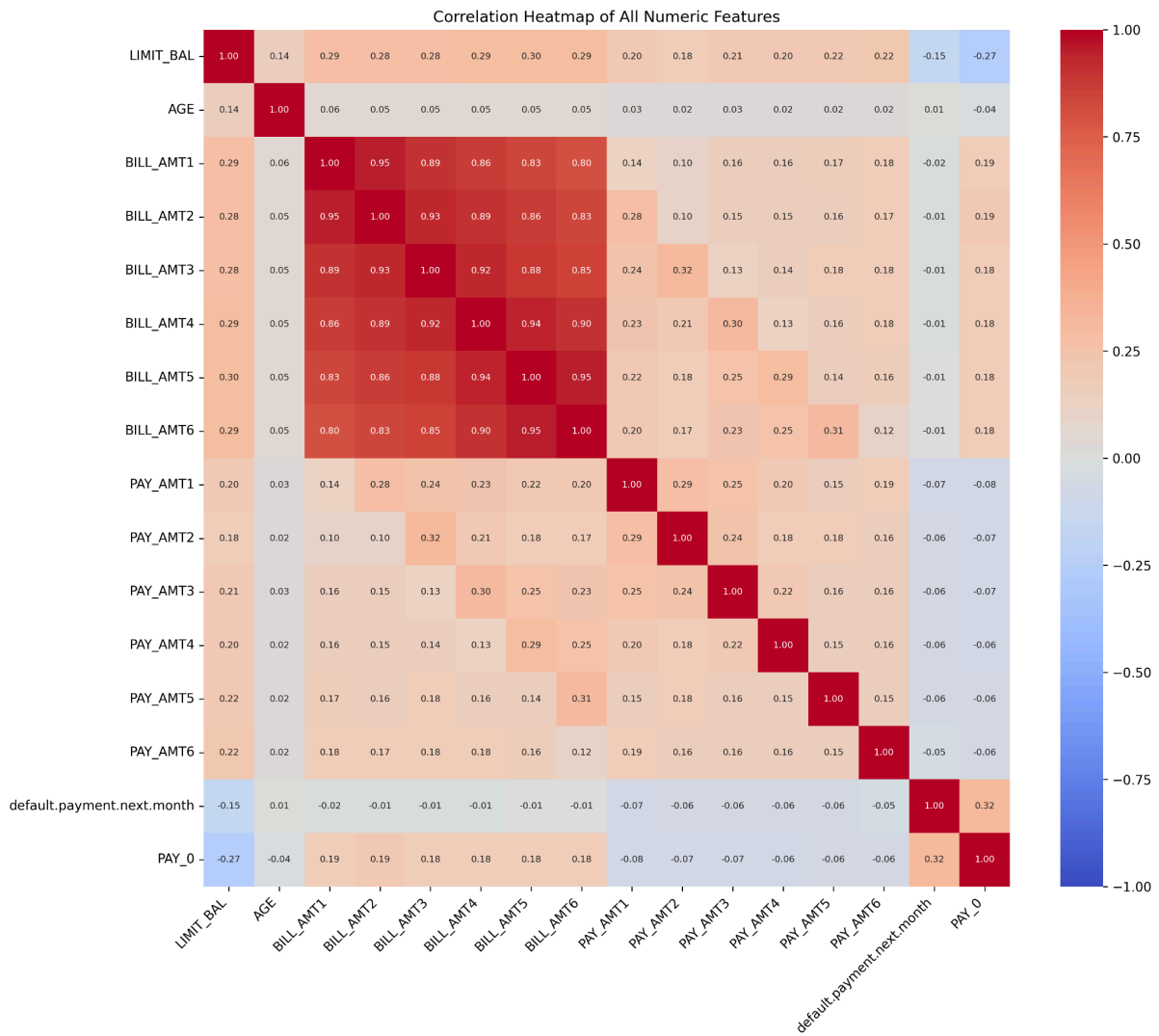
- **Kwota rachunku (wrzesień) względem stanu cywilnego**

Osoby zamężne/żonate mają wyższe kwoty rachunków, co może być związane z większymi wydatkami domowymi.



## Mapa termiczna (Heatmap)

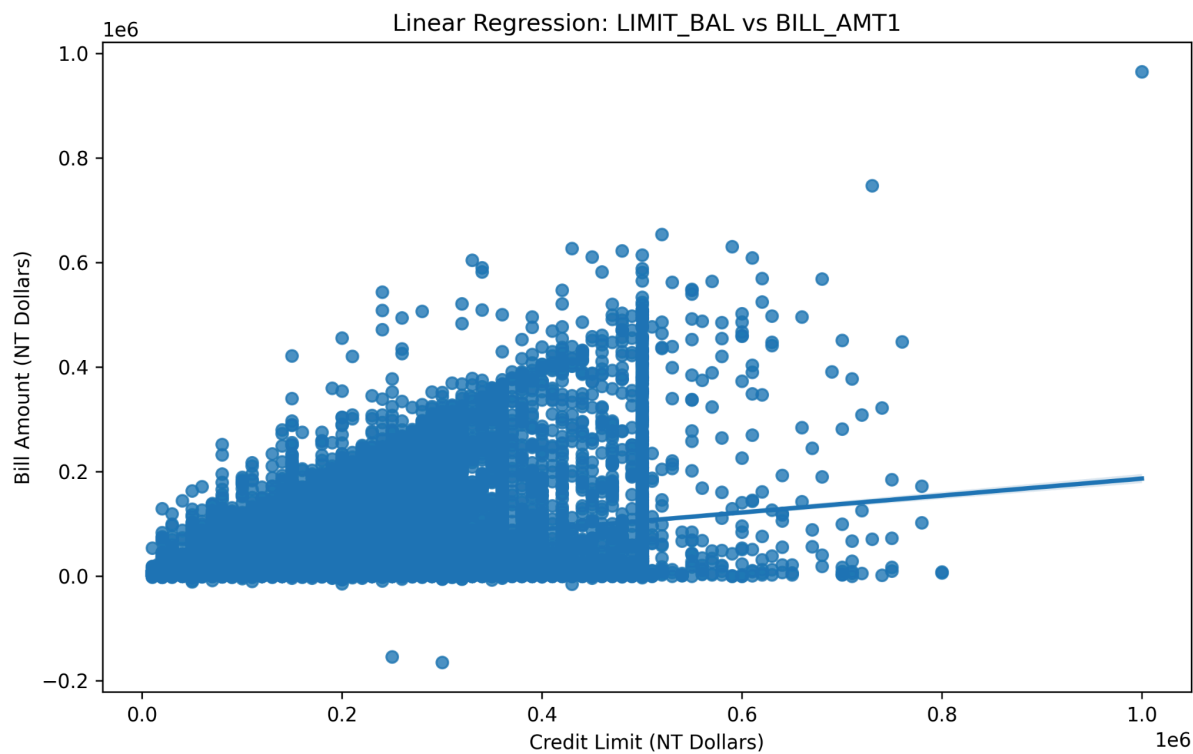
Mapa termiczna korelacji między wszystkimi zmiennymi liczbowymi wskazuje na silne dodatnie korelacje między kwotami rachunków w kolejnych miesiącach, co sugeruje spójność w wydatkach klientów. Istnieje również umiarkowana korelacja między limitem kredytowym a kwotami rachunków, co oznacza, że klienci z wyższymi limitami więcej wydają.



## Wykresy regresji liniowej (Regression Plots)

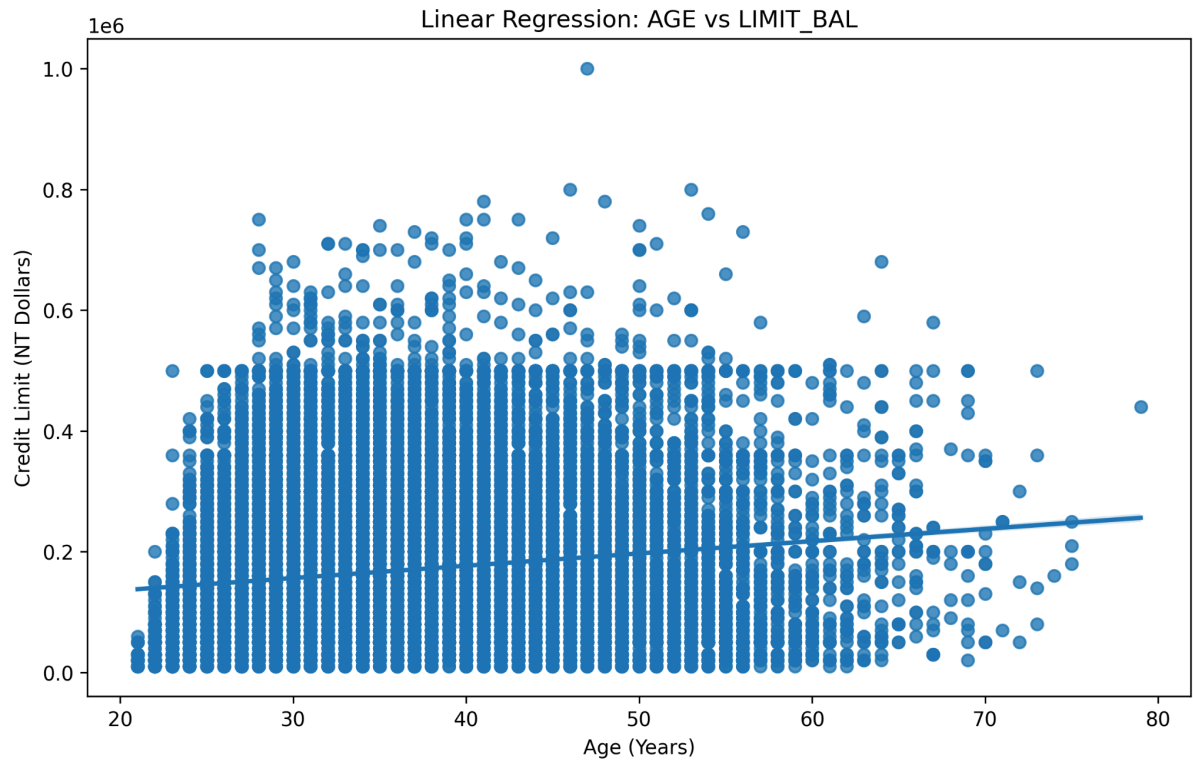
- **LIMIT\_BAL vs BILL\_AMT1**

Wykres pokazuje słabą dodatnią zależność między limitem kredytowym a kwotą rachunku we wrześniu, co sugeruje, że wyższe limity mogą prowadzić do większych wydatków.



- **AGE vs LIMIT\_BAL**

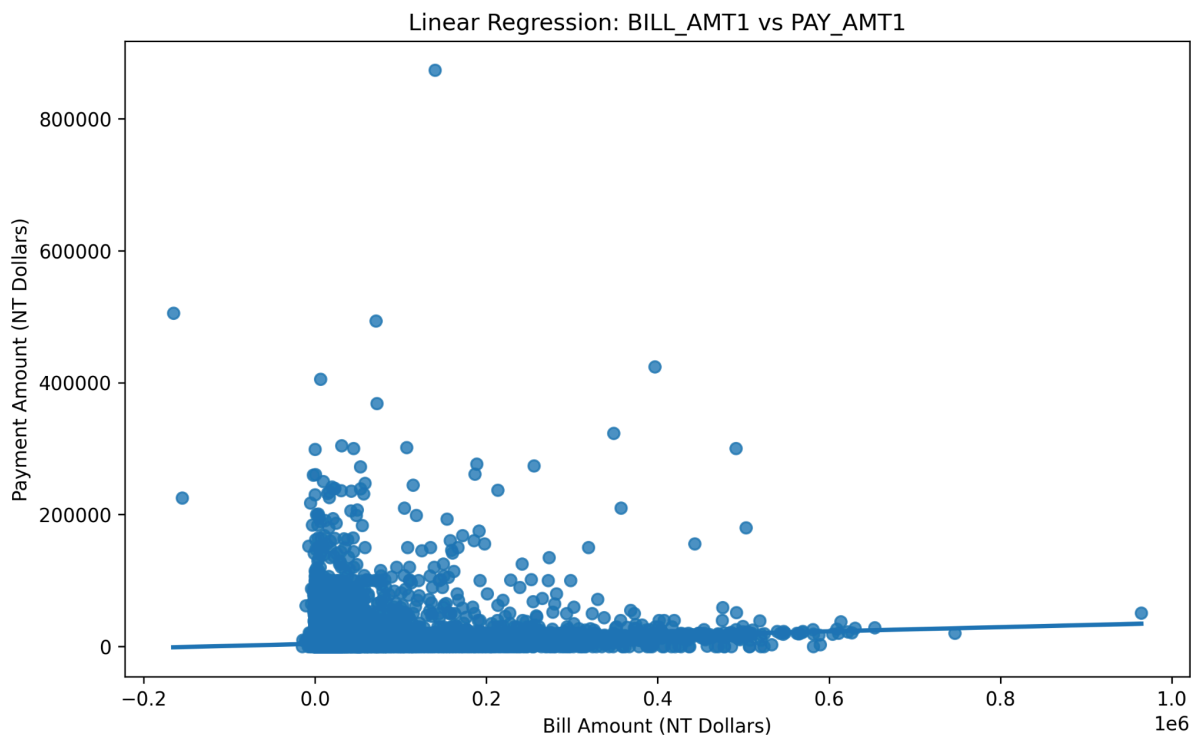
Słaba dodatnia zależność wskazuje, że starsi klienci mogą mieć nieco wyższe limity kredytowe, choć związek nie jest siln



- **BILL\_AMT1 vs PAY\_AMT1**

Wykres ujawnia dodatnią zależność między kwotą rachunku a kwotą spłaty we wrześniu, co jest oczekiwane, ponieważ klienci z wyższymi rachunkami zwykle dokonują większych spłat.





## Wnioski i podsumowanie

Analiza eksploracyjna danych ujawniła kilka kluczowych wniosków:

- **Czynniki demograficzne:** Stan cywilny i poziom wykształcenia znacząco wpływają na limity kredytowe i kwoty rachunków. Osoby zamężne/żonate oraz te z wyższym wykształceniem mają wyższe limity i większe rachunki.
- **Ryzyko niewypłacalności:** Młodszy klienci oraz ci z niższymi limitami kredytowymi są bardziej narażeni na niewypłacalność. Wyższe kwoty rachunków i niższe spłaty we wrześniu również zwiększają ryzyko defaultu.
- **Korelacje:** Silne korelacje między kwotami rachunków w czasie wskazują na stałe wzorce wydatków, podczas gdy umiarkowane korelacje między limitem kredytowym a rachunkami sugerują, że wyższe limity umożliwiają większe wydatki.

Te obserwacje wskazują, że zmienne demograficzne i historia płatności są istotnymi predyktorami ryzyka niewypłacalności. Dalsze prace powinny skupić się na budowie modeli predykcyjnych wykorzystujących te wnioski do identyfikacji klientów zagrożonych defaultem. Wszystkie wykresy zostały zapisane w katalogu „outputs” i mogą być wykorzystane do szczegółowej analizy.