

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).

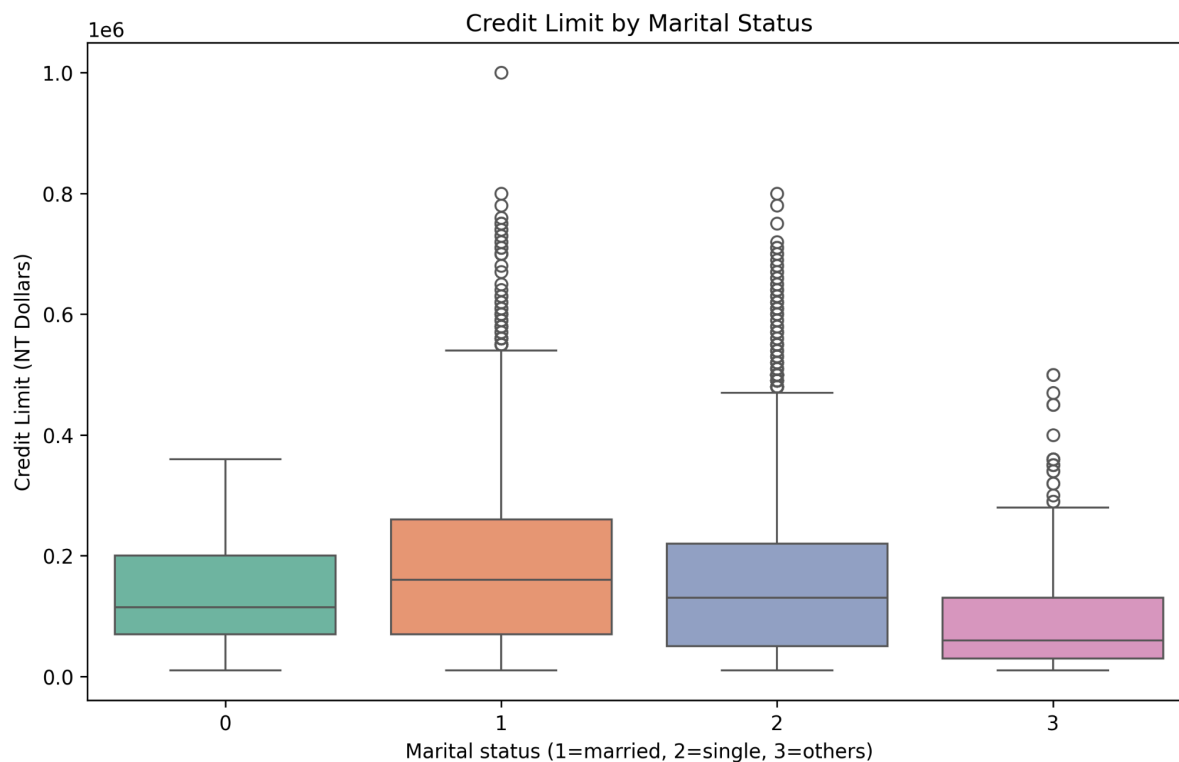
Statystyki opisowe kluczowych zmiennych:

- **LIMIT_BAL (Limit kredytowy):** Średnia: 167 484 NT\$, Mediana: 140 000 NT\$, Min: 10 000 NT\$, Max: 1 000 000 NT\$, Odchylenie standardowe: 129 747 NT\$, 5. percentyl: 20 000 NT\$, 95. percentyl: 430 000 NT\$.
- **AGE (Wiek):** Średnia: 35,5 lat, Mediana: 34 lata, Min: 21 lat, Max: 79 lat, Odchylenie standardowe: 9,2 lat, 5. percentyl: 23 lata, 95. percentyl: 53 lata.
- **BILL_AMT1 (Rachunek we wrześniu):** Średnia: 51 223 NT\$, Mediana: 22 381 NT\$, Min: -165 580 NT\$, Max: 964 511 NT\$, Odchylenie standardowe: 73 635 NT\$, 5. percentyl: 0 NT\$, 95. percentyl: 201 203 NT\$.
- **PAY_AMT1 (Spłata we wrześniu):** Średnia: 5 663 NT\$, Mediana: 2 100 NT\$, Min: 0 NT\$, Max: 873 552 NT\$, Odchylenie standardowe: 16 563 NT\$, 5. percentyl: 0 NT\$, 95. percentyl: 18 428 NT\$.
- **SEX (Płeć):** Proporcje: Kobiety (2): 60,4%, Mężczyźni (1): 39,6%.
- **EDUCATION (Wykształcenie):** Proporcje: Uniwersytet (2): 46,8%, Szkoła wyższa (1): 35,3%, Szkoła średnia (3): 16,4%, Inne kategorie (4, 5, 6, 0): <1%.
- **MARRIAGE (Stan cywilny):** Proporcje: Samotni (2): 53,2%, Żonaci/Zamężne (1): 45,5%, Inne (3, 0): <2%.
- **default.payment.next.month (Niewypłacalność):** Proporcje: Wypłacalni (0): 77,9%, Niewypłacalni (1): 22,1%

Analiza eksploracyjna danych

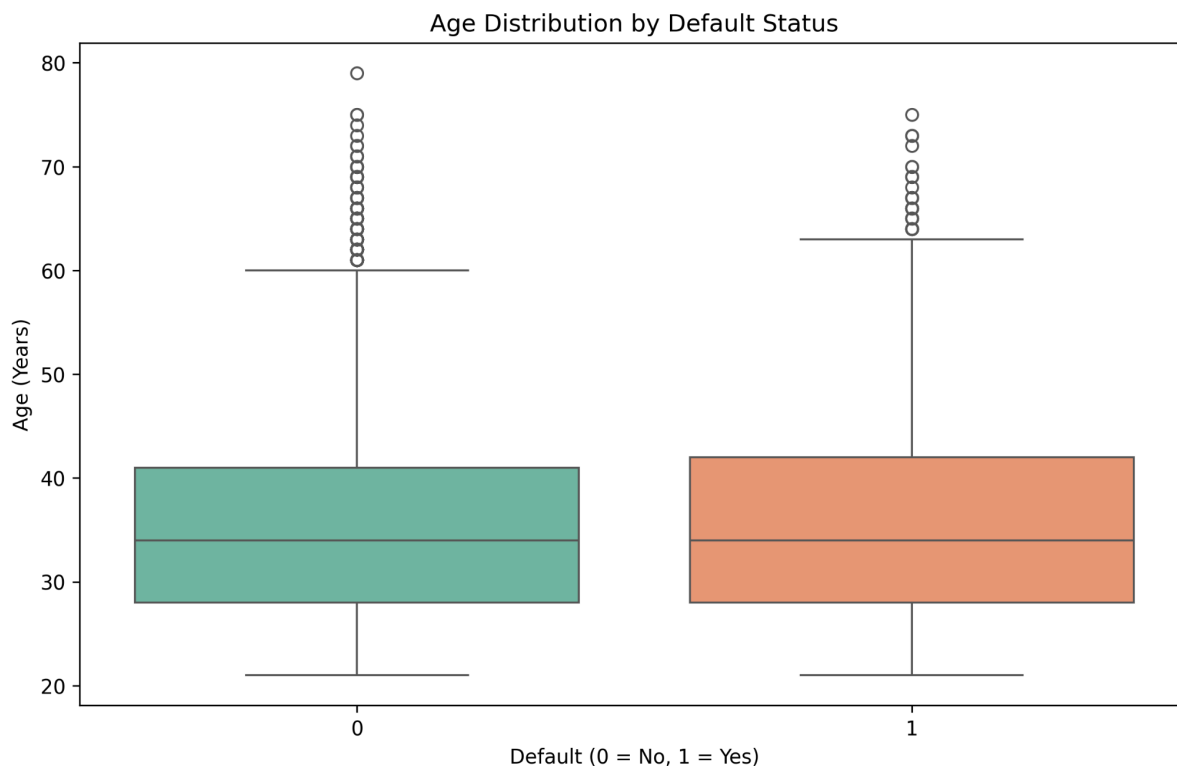
Wykresy pudełkowe (Box Plots)

- **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.



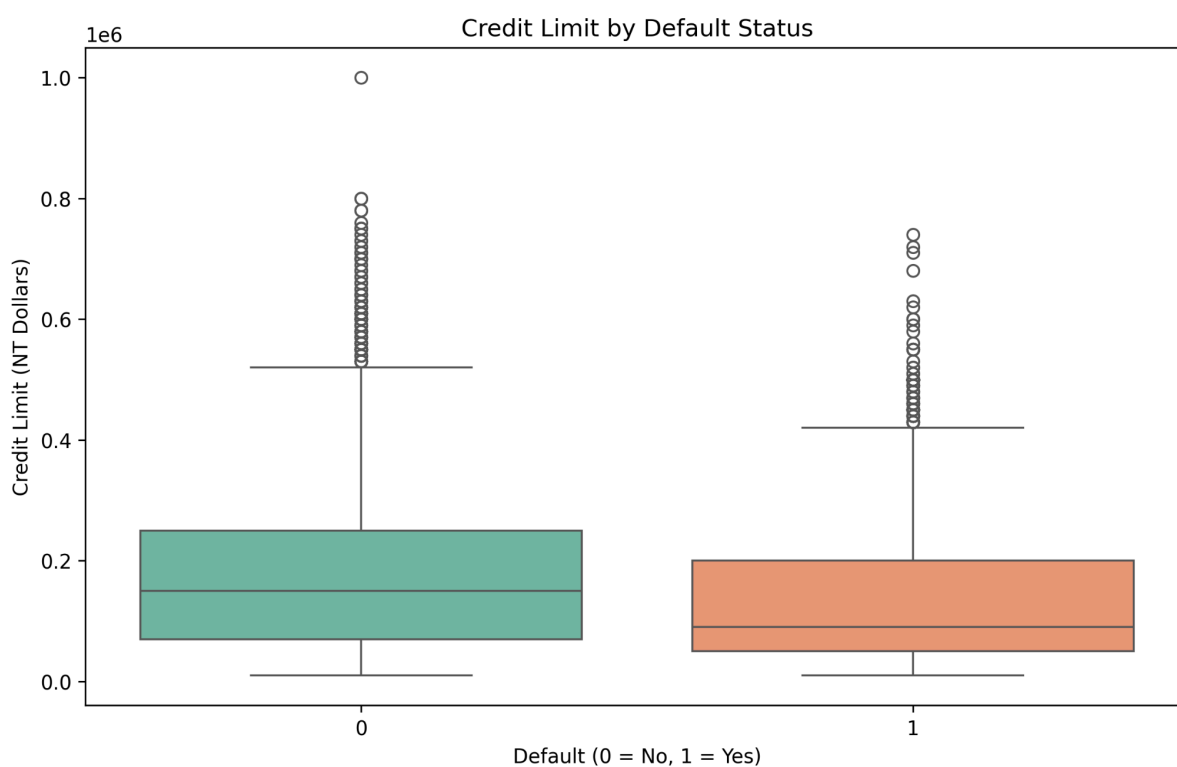
- **Rozkład wieku względem statusu niewypłacalności**

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.



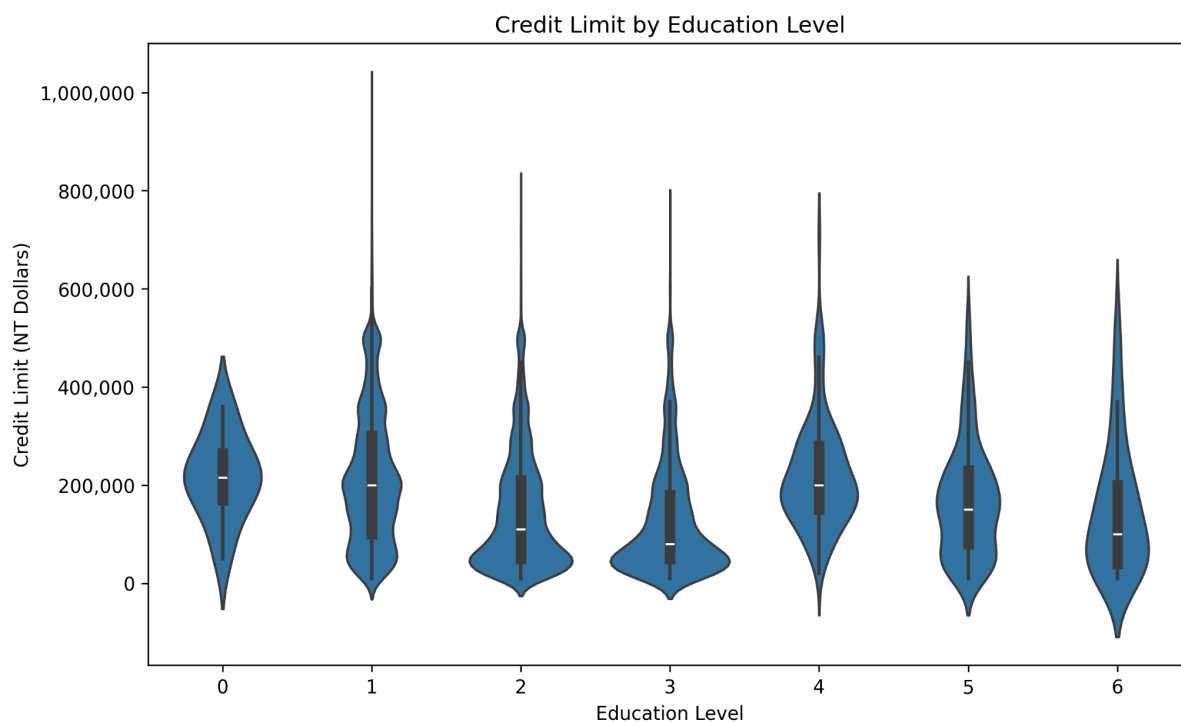
- **Limit kredytowy względem statusu niewypłacalności**

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.



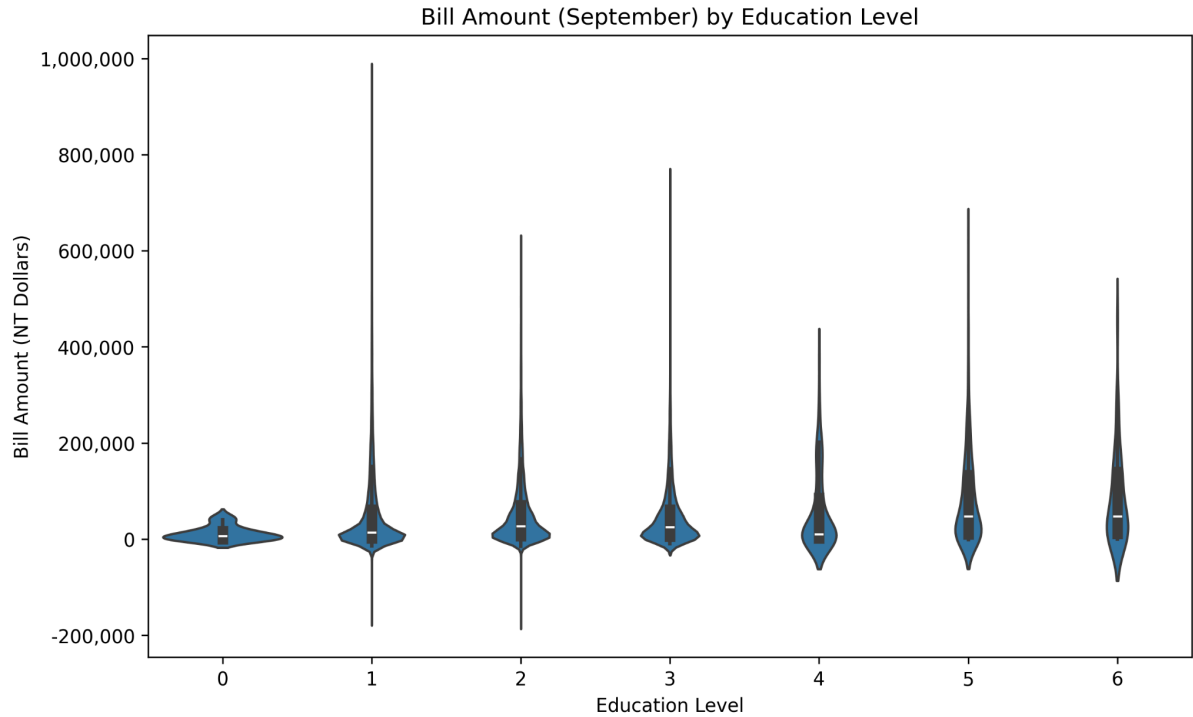
- **Limit kredytowy względem poziomu wykształcenia**

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.



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

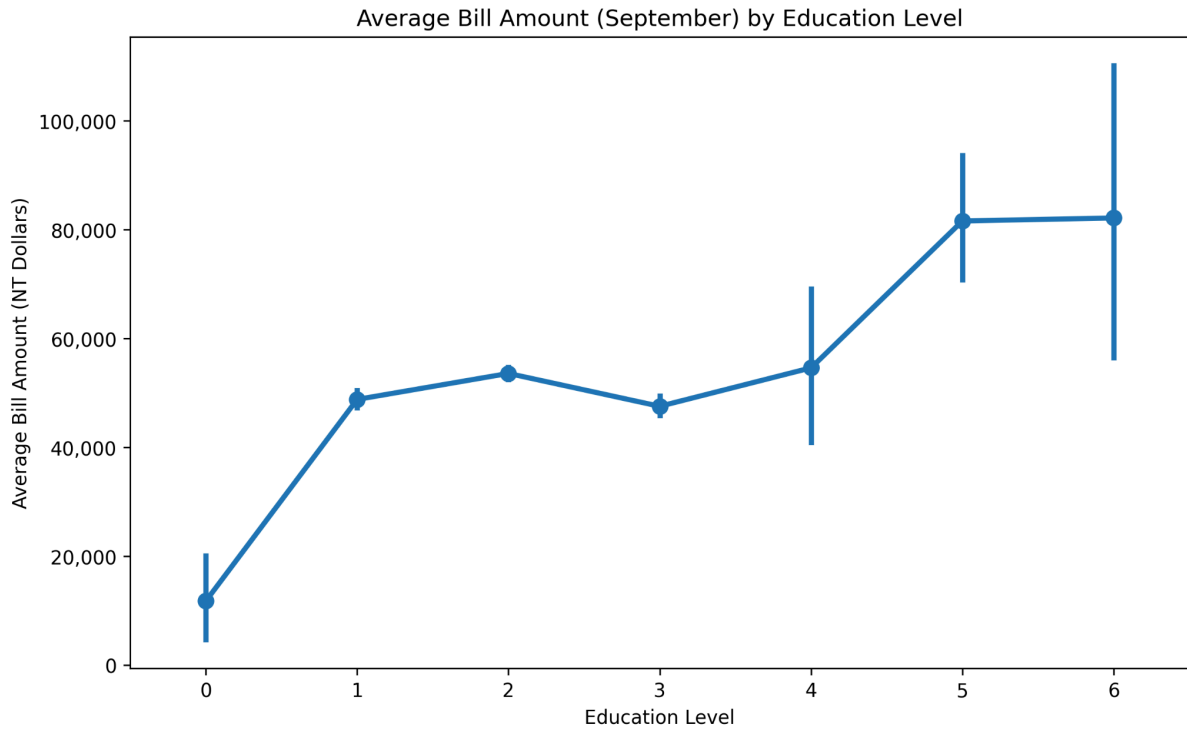
Rozkład kwot rachunków we wrześniu pokazuje, że osoby z wyższym wykształceniem mają wyższe rachunki, co może być związane z większymi wydatkami lub zdolnością kredytową.



Wykresy słupkowe z przedziałami ufności (Error Bars)

- Średnia kwota rachunku (wrzesień) względem poziomu wykształcenia

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

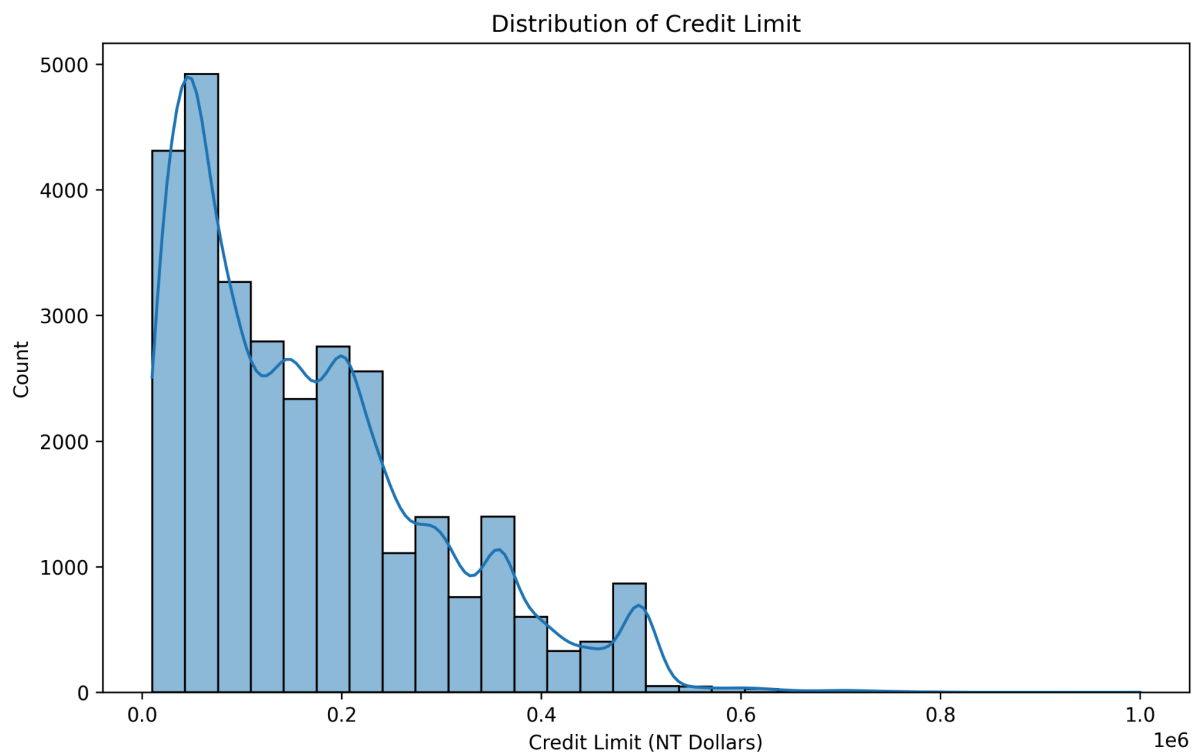


Histogramy

- **Rozkład limitu kredytowego**

Histogram pokazuje, że limity kredytowe są prawostronnie skośne:

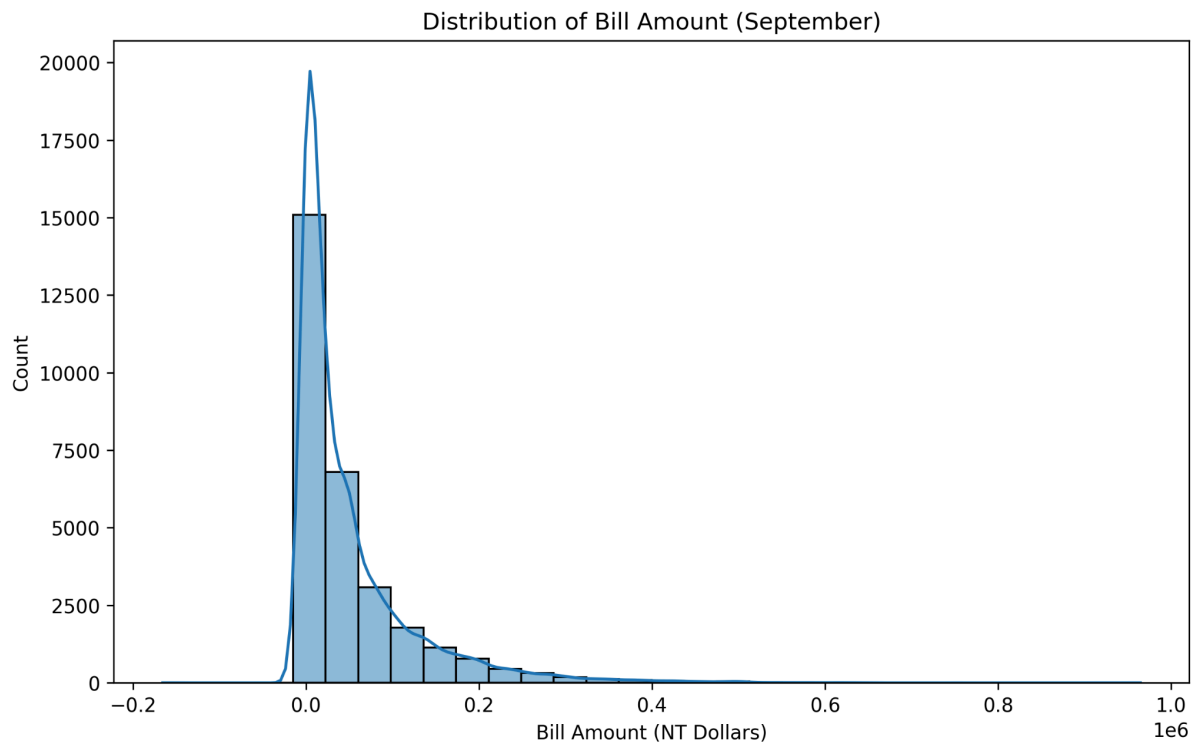
- Większość klientów ma niskie limity.
- Tylko nieliczni mają bardzo wysokie limity.



- **Rozkład kwoty rachunku (wrzesień)**

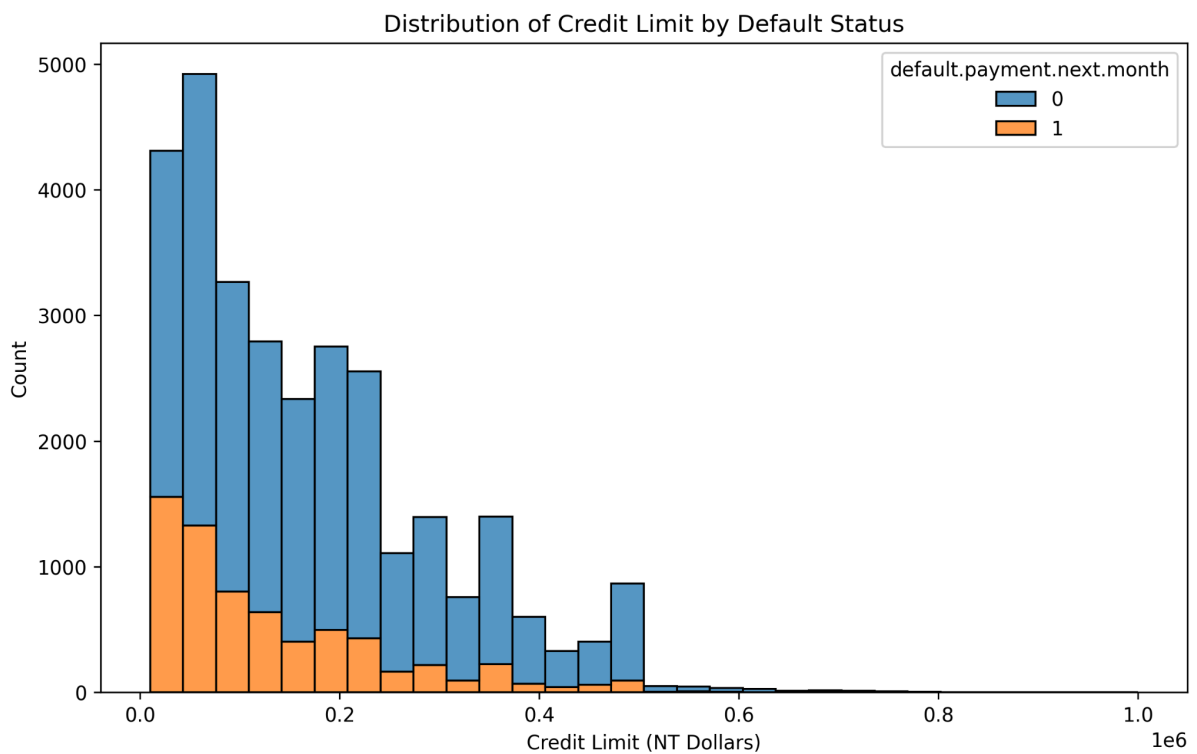
Rozkład kwot rachunków we wrześniu jest prawostronnie skośny:

- Większość klientów ma niskie rachunki.
- Nieliczni klienci mają bardzo wysokie kwoty.



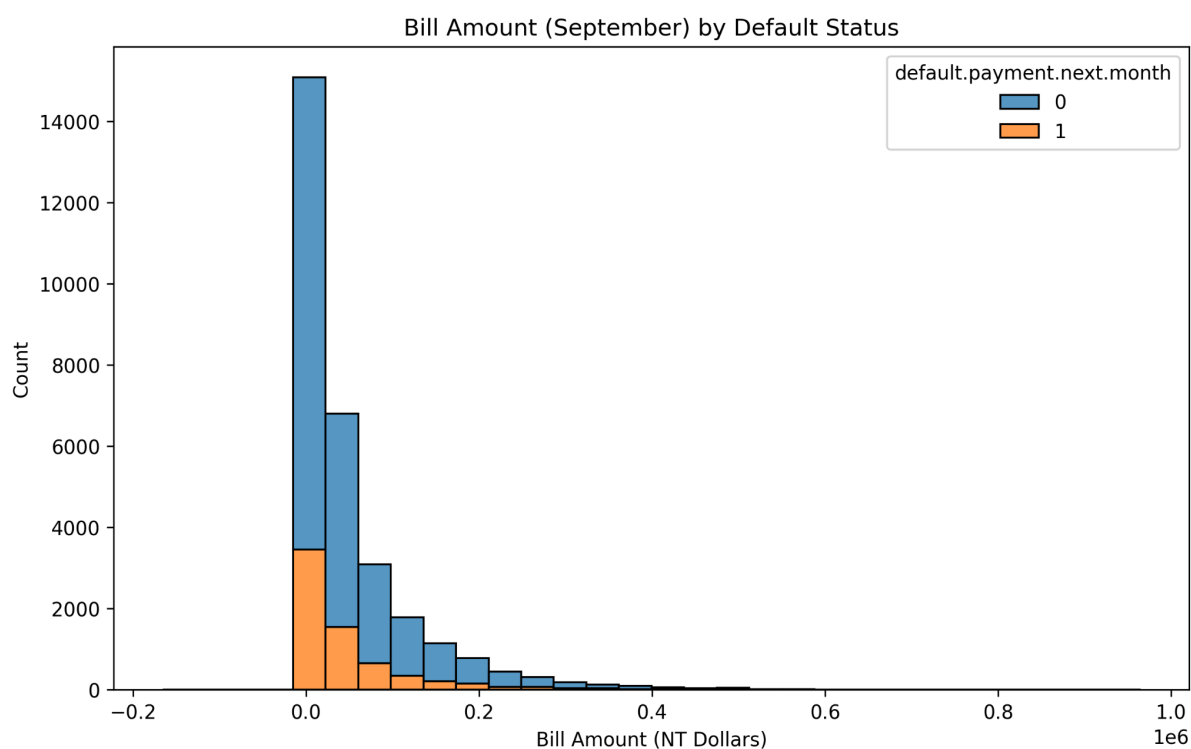
Histogramy z odcieniem (Hue)

- **Rozkład limitu kredytowego w zależności od statusu spłaty:**
- Klienci niewypłacalni mają niższe limity kredytowe niż ci, którzy spłacają dług.
- Różnicę widać w przesunięciu rozkładów – grupa z problemami finansowymi skupia się przy niższych kwotach.



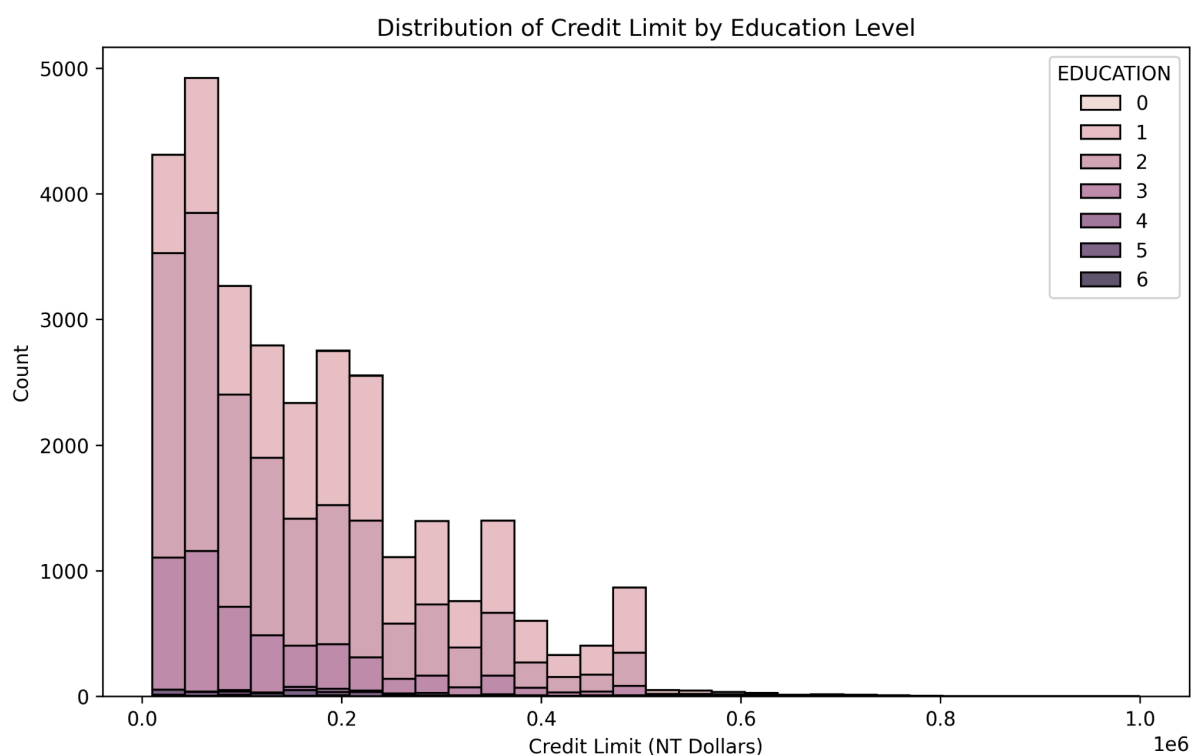
- **Kwota rachunku (wrzesień) względem statusu niewypłacalności**

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



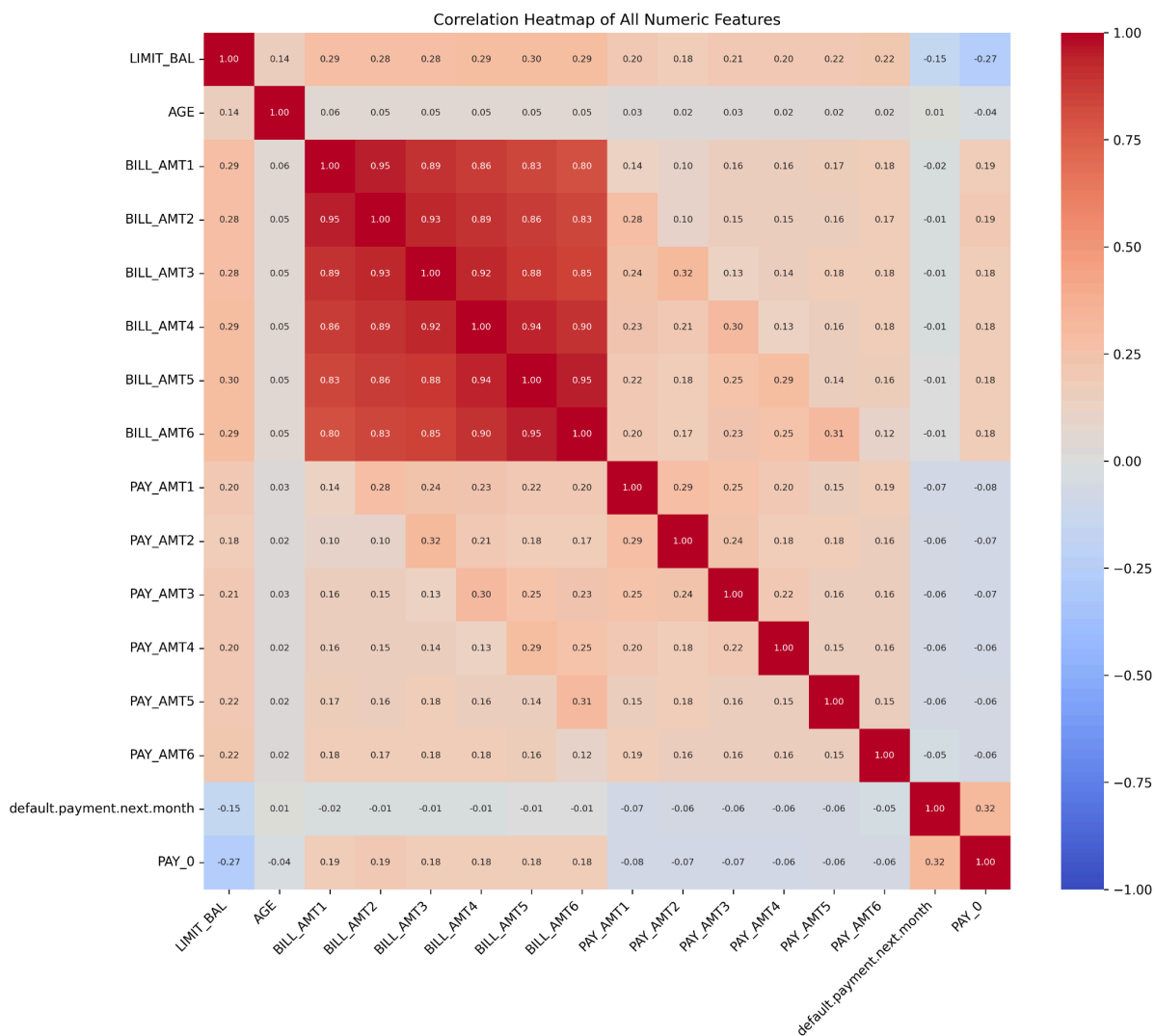
- **Rozkład limitu kredytowego względem poziomu wykształcenia**

Wyższe poziomy wykształcenia korelują z wyższymi limitami kredytowymi.



Mapa termiczna (Heatmap)

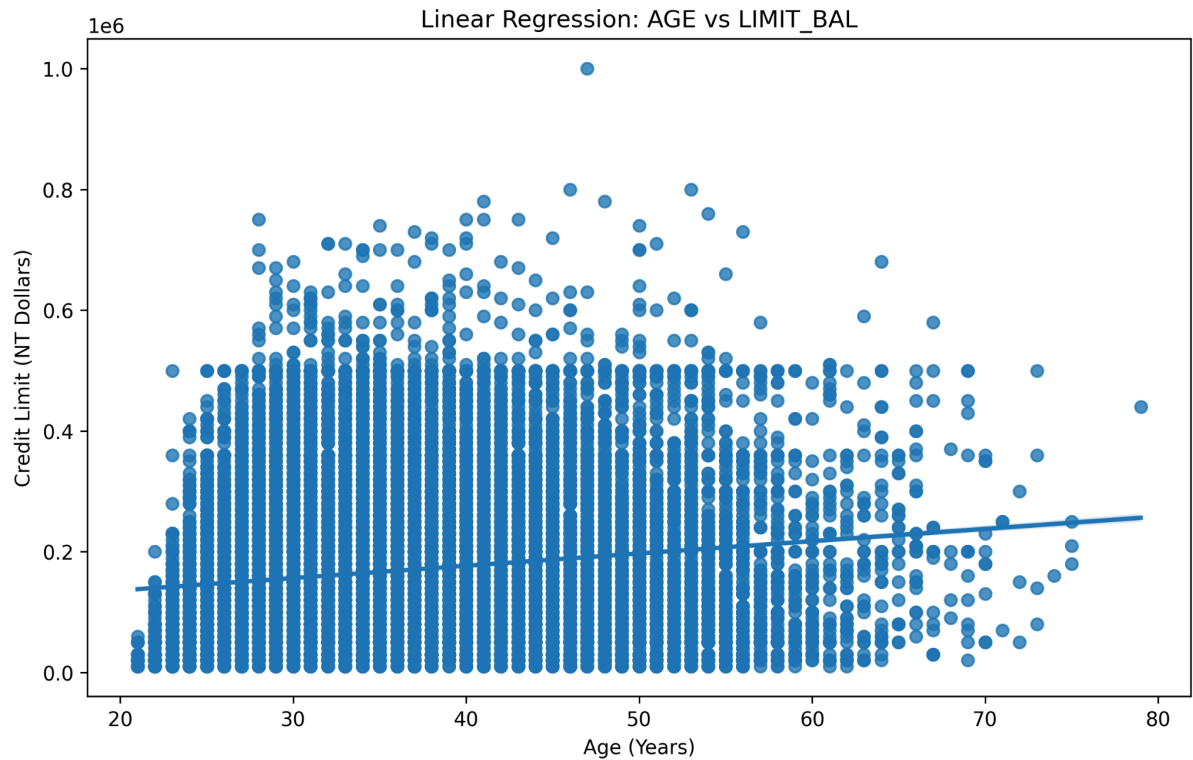
Mapa termiczna korelacji między wszystkimi zmiennymi liczbowymi wskazuje na silne dodatnie korelacje między kwotami rachunków w kolejnych miesiącach, co sugeruje spójność w wydatkach klientów. Istnieje również umiarkowana korelacja między limitem kredytowym a kwotami rachunków, co oznacza, że klienci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



Wykresy regresji liniowej (Regression Plots)

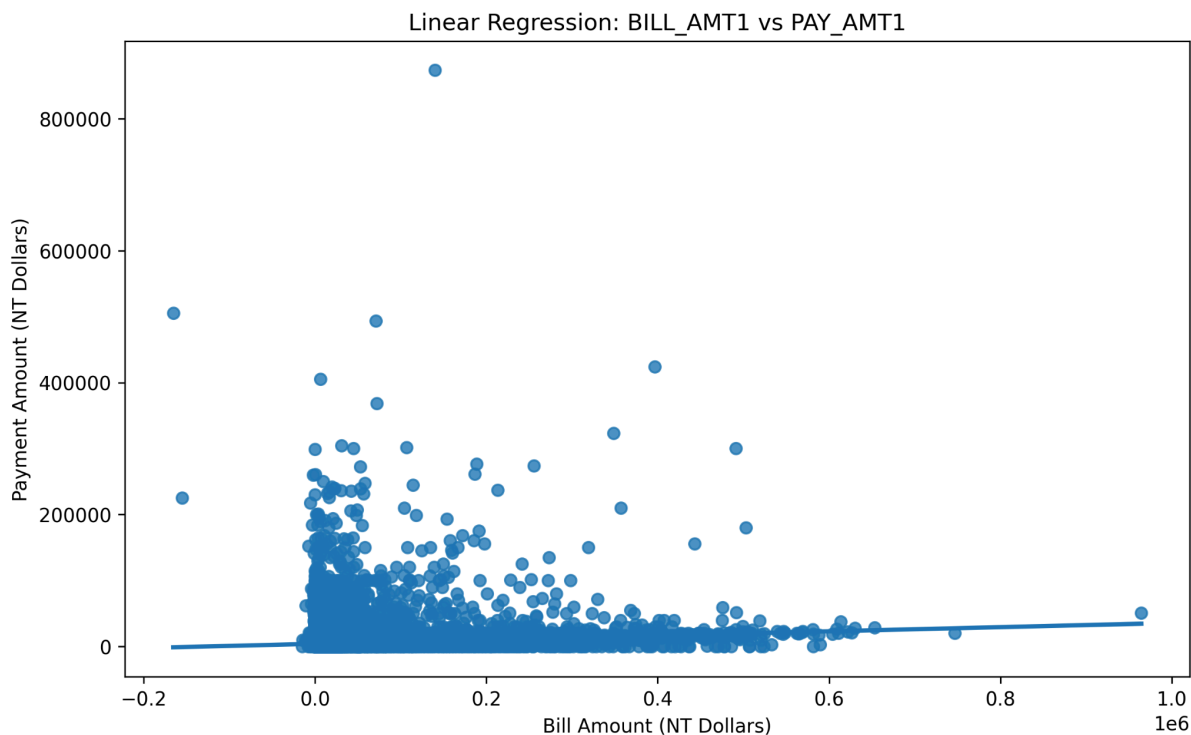
- AGE vs LIMIT_BAL**

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



- **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

Kluczowe wnioski z analizy danych:

1. **Wpływ demografii** – Osoby zamężne/żonate i z wyższym wykształceniem mają większe limity kredytowe i wydają więcej.
2. **Ryzyko niewypłacalności** – Młodszy klienci i osoby z niskim limitem kredytowym częściej mają problemy ze spłatą. Wysokie rachunki i niskie spłaty we wrześniu zwiększają ryzyko defaultu.
3. **Stale wzorce wydatków** – Kwoty rachunków w kolejnych miesiącach są do siebie podobne, a wyższy limit kredytowy pozwala na większe wydatki.

Wyniki wskazują, że dane demograficzne i historia płatności pomagają przewidzieć ryzyko niewypłacalności. Warto stworzyć model, który lepiej zidentyfikuje klientów zagrożonych defaultem. Wszystkie wykresy są zapisane w katalogu „outputs”.