

Аналіз прохідності із запропонованими гіпотезами щодо чинників, які впливають на неї.

У цьому дослідженні здійснюється аналіз прохідності транзакцій (acceptance rate) на основі даних за шість місяців.

Мета роботи: виявити та дослідити чинники, що можуть впливати на успішність платежів.

Для досягнення цієї мети сформульовано гіпотези щодо впливу окремих характеристик транзакцій на рівень успішних оплат.

Гіпотези

Часові характеристики(created_day)

- Гіпотеза 1: Прохідність змінюється залежно від місяця.
- Гіпотеза 2: Прохідність змінюється залежно від дня тижня.

Характеристики продукту(channel_id, order_type, merchant_token_type):

- Гіпотеза 3: Різні продукти бізнесу (channel_id) можуть мати різну прохідність
- Гіпотеза 4: Тип платежу (order_type) впливає на ймовірність успішної оплати.
- Гіпотеза 5: Тип токена (merchant_token_type) впливає на прохідність.

Джерело оплати (payment_source):

- Гіпотеза 6: Прохідність при оплаті звичайною карткою може бути нижчою ніж при оплаті Apple Pay/Google Pay

Захист платежу (**is_secured**):

- Гіпотеза 7: Наявність 3D Secure підвищує прохідність.

Сума транзакції (**bin_amount_usd**):

- Гіпотеза 8: Дуже великі платежі мають нижчу ймовірність успішної оплати.

Характеристики картки (**card_brand**, **card_type**):

- Гіпотеза 9: Прохідність може бути вищою на деяких видах карток
- Гіпотеза 10: Прохідність може бути вищою на деяких типах карток

Країна картки (**card_country**, **ip_country**):

- Гіпотеза 11: Картки з певних країн (**card_country**) мають вищу або нижчу прохідність.
- Гіпотеза 12: Країна IP користувача та країни з якої випускалася карта можуть відрізнятися, а це може бути ризик шахрайства, який впливає на прохідність.

Історія платежів (**invoice_number**, **retry_number**):

- Гіпотеза 13: Чим вищий номер інвойсу (**invoice_number**) тим користувач може бути досвідченішим, а відповідно його прохідність може бути вищою
- Гіпотеза 14: Більша кількість спроб (**retry_number**) зменшує ймовірність успішної оплати.

Загальні характеристики (**count_orders**, **count_emails**):

- Гіпотеза 15: Чим більше замовлень (**count_orders**) на одного користувача — тим вища загальна прохідність.

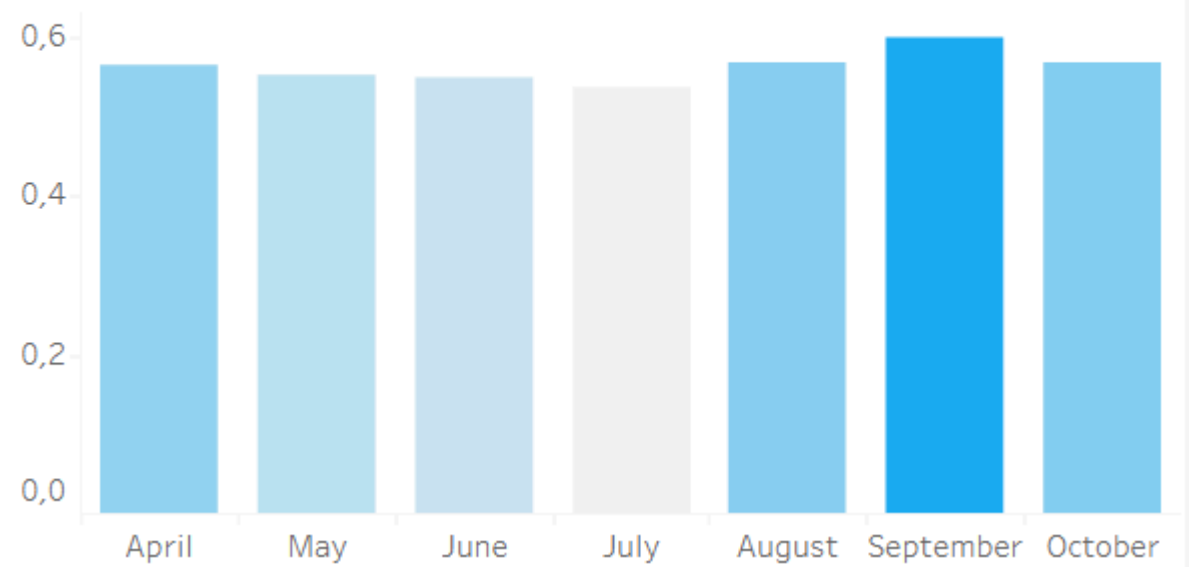
- **Гіпотеза 16:** Велика кількість emails(`count_emails`) може свідчити про створення багатьох пошт, а відповідно і шахрайства, яке призводить до низької прохідності

Перевірка гіпотез

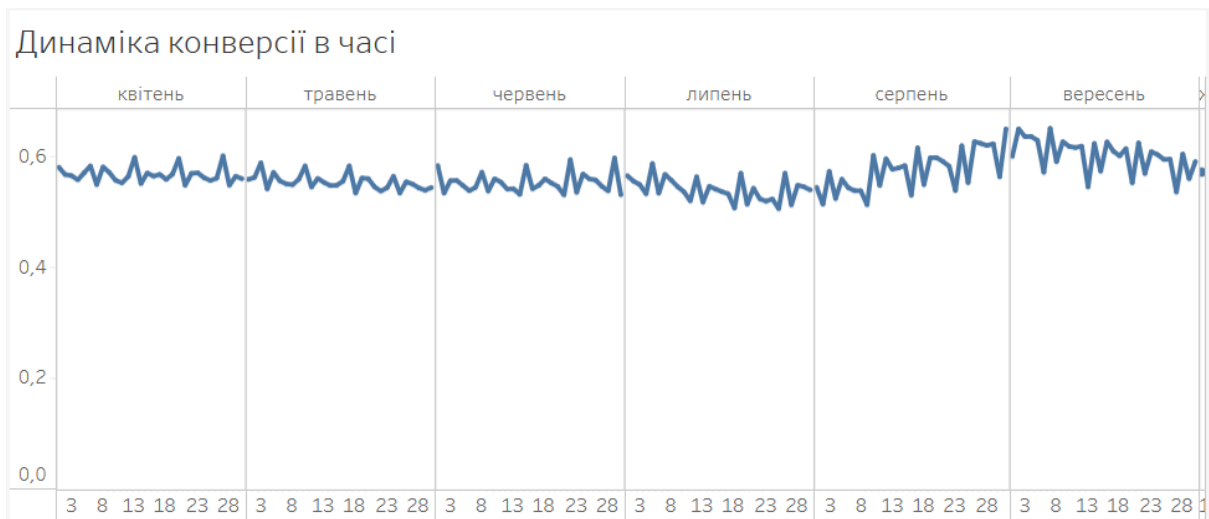
Перевіряти гіпотези будемо на основі візуалізацій в Tableau, та результатів кореляцій знайдених в Python коді.

Гіпотеза 1: Прохідність змінюється залежно від місяця.

Розподіл показників транзакцій і середня конверсія



- Прохідність по місяцях коливається слабо.
- Найнижча середня прохідність у липні.
- Найвища середня прохідність у вересні.
- Всі інші місяці мають дуже близькі значення прохідності($\sim 0.55-0.6$).



- З квітня по липень видно легкий спад прохідності.
- Починаючи із серпня та особливо у вересні видно підвищення і стабілізацію прохідності на вищому рівні.
- Прохідність коливається доволі стабільно без різких стрибків.

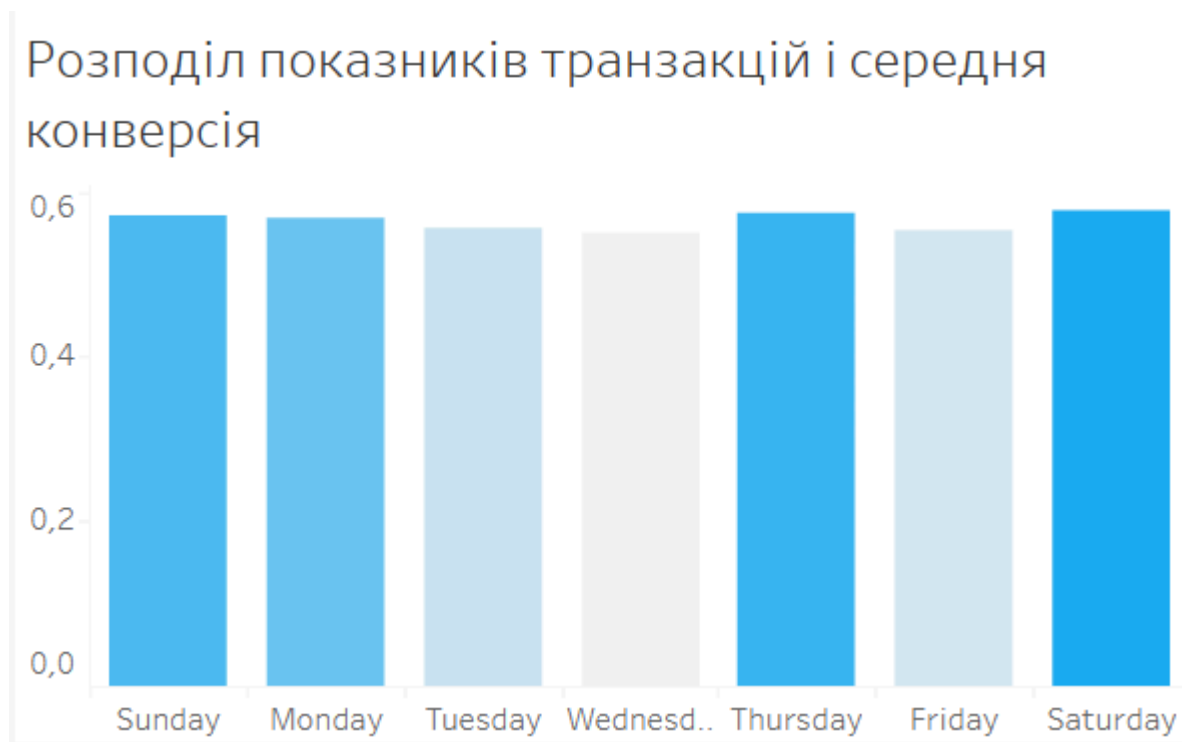
Кореляція між місяцем і конверсією = 0.04:

- 0.04 — дуже слабка кореляція.
- Це означає, що місяць майже не впливає на конверсію напряду.

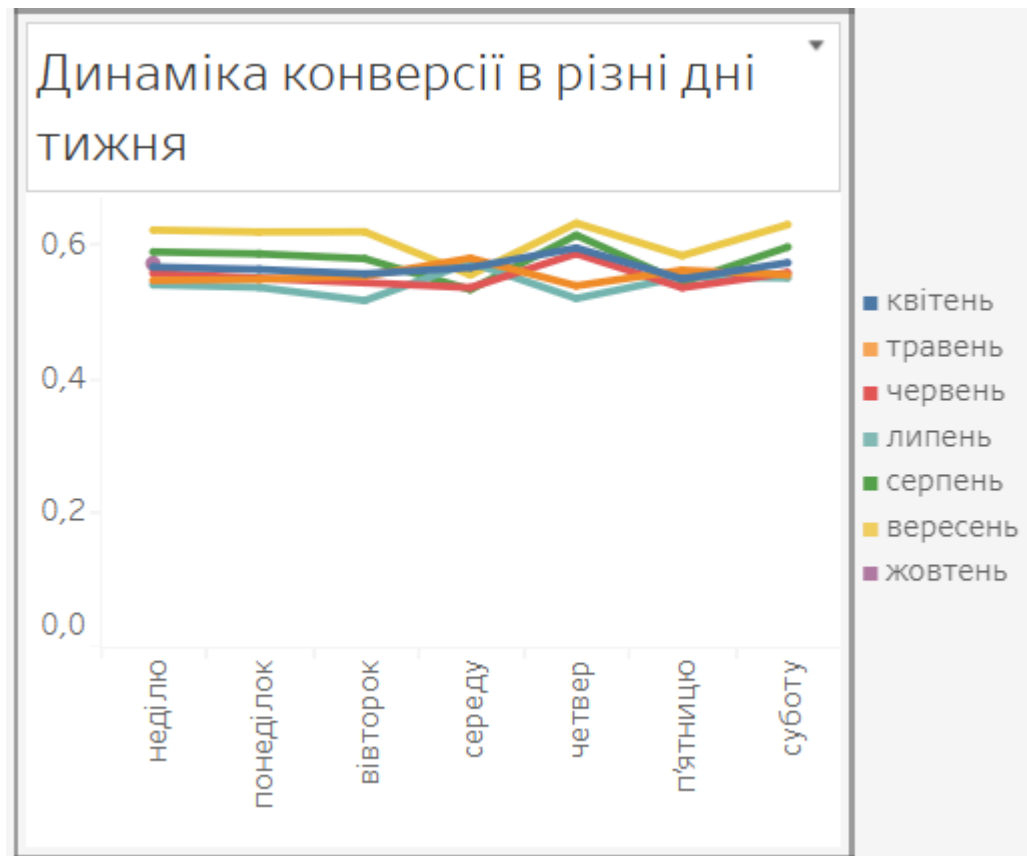
Висновок: Гіпотеза не підтверджується.

Хоча на графіках спостерігаються невеликі коливання прохідності між місяцями, статистична кореляція 0.04 вказує на відсутність реального зв'язку між місяцем і успішністю оплат.

Гіпотеза 2: Прохідність змінюється залежно від дня тижня.



- Конверсія найнижча у середу (Wednesday).
- Найвища конверсія у четвер (Thursday), неділю (Sunday) та понеділок (Monday).
- Інші дні коливаються дуже слабо навколо середнього рівня.



- У більшості місяців конверсія трохи просідає в середині тижня (вівторок–середа).
- Піки конверсії здебільшого припадають на четвер або вихідні (субота/неділя).
- Загальні коливання невеликі — графіки майже паралельні.

Кореляція між днем тижня і конверсією = 0.0159:

- 0.0159 — практично нульова кореляція.
- Це означає, що день тижня майже не впливає на конверсію в статистичному сенсі.

Висновок: Гіпотеза не підтверджується.

Незважаючи на незначні коливання прохідності між різними днями тижня, статистична кореляція 0.0159 вказує на **відсутність реального впливу** дня тижня на успішність оплат.

Гіпотеза 3: Різні продукти бізнесу (channel_id**) можуть мати різну прохідність**



- Є чітка різниця в середній прохідності між різними **channel_id**.
- **channel_id = 1** має найвищу прохідність (~80%).
- **channel_id = 2** і **channel_id = 3** теж мають високу прохідність (~75-78%).
- **channel_id = 5** і **channel_id = 6** мають значно нижчу прохідність (~50%).

Кореляція між **channel_id і конверсією = -0.2476:**

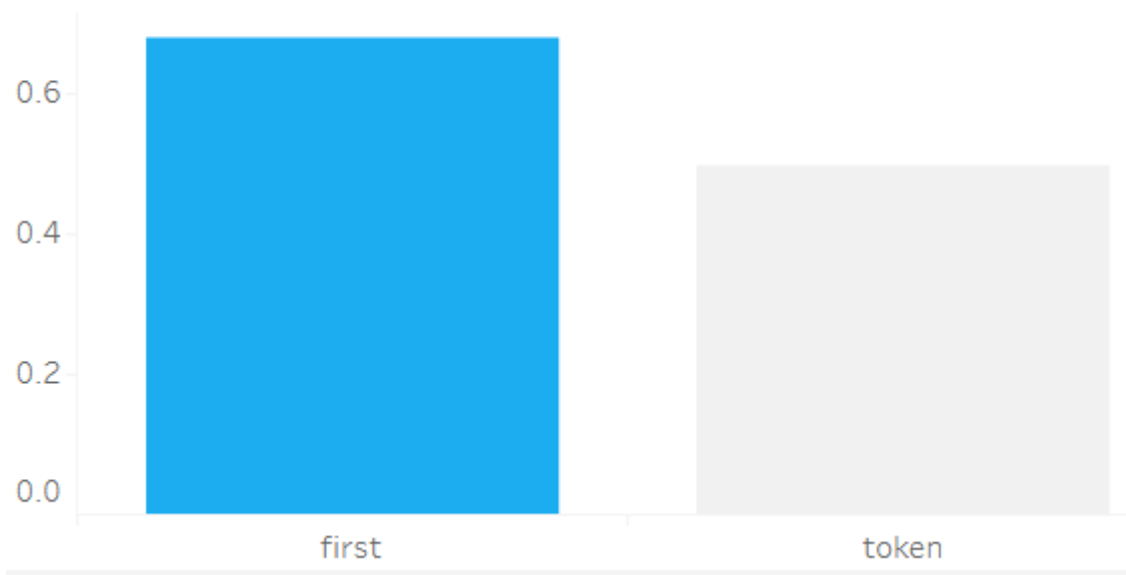
- Кореляція помітна і від'ємна (приблизно -0.25), тобто є зв'язок, який показує, що чим більший номер **channel_id**, тим нижча середня конверсія.

Висновок: Гіпотеза частково підтверджується.

Дані показують, що продукт бізнесу (`channel_id`) впливає на прохідність: продукти з меншими номерами мають вищу прохідність. Кореляція -0.2476 вказує на **слабкий від'ємний зв'язок** між `channel_id` і успішністю оплат.

Гіпотеза 4: Тип платежу (`order_type`) впливає на ймовірність успішної оплати.

Розподіл показників транзакцій і середня конверсія



- При першій оплаті (`first`) прохідність помітно вища (~ 0.68).
- При оплаті через токен (`token`) прохідність нижча (~ 0.50).

Кореляція між `order_type` і прохідністю = -0.2863 :

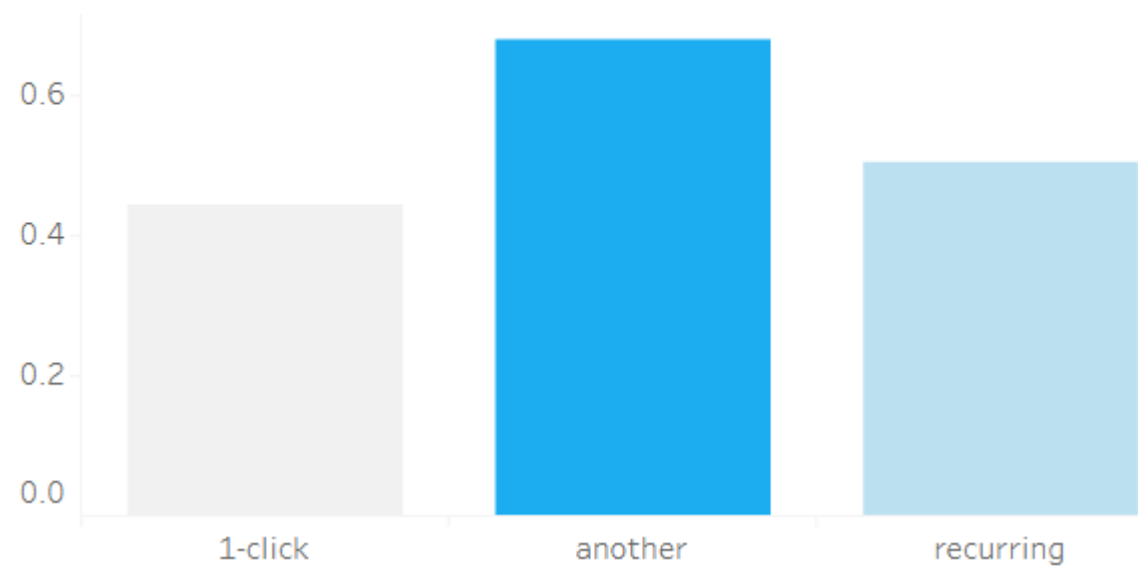
- Кореляція від'ємна і помітна (майже -0.29), тобто є зв'язок, який показує, що оплата через токен пов'язана з нижчою прохідністю.
- Тип платежу справді впливає на успішність транзакції.

Висновок: Гіпотеза підтверджується.

Прохідність оплат залежить від типу платежу (**order_type**): перші оплати мають вищу прохідність, ніж оплати через токен. Кореляція -0.2863 вказує на помітний слабкий від'ємний зв'язок.

Гіпотеза 5: Тип токена (merchant_token_type**) впливає на прохідність.**

Розподіл показників транзакцій і середня конверсія



- another має найвищу прохідність (~ 0.68).
- recurring має прохідність трохи нижчу (~ 0.52).
- 1-click має найнижчу прохідність (~ 0.45).

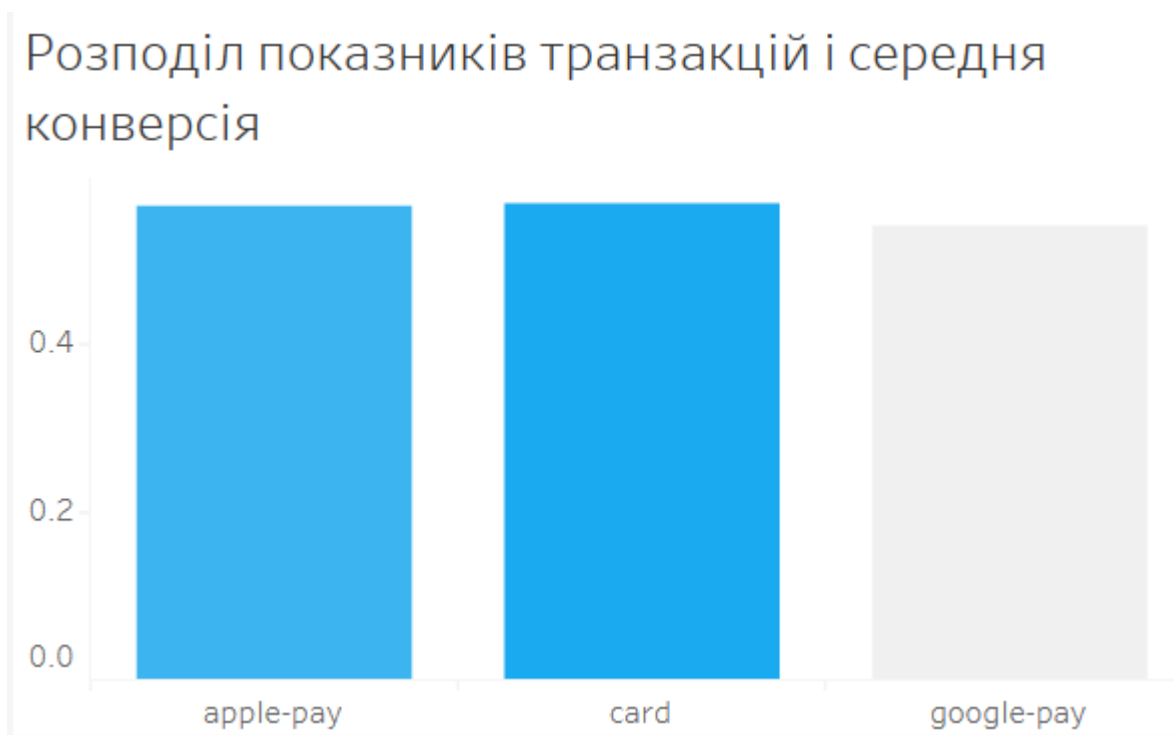
Кореляція = -0.1334 :

- Дуже слабка від'ємна кореляція. Формально зв'язок існує, але він дуже слабкий.

Висновок: Гіпотеза підтверджується частково.

Тип токена (`merchant_token_type`) має вплив на прохідність, але цей вплив дуже слабкий (кореляція -0.1334). Видно, що прохідність різниться між типами токенів: найкращі результати у категорії "another", найгірші — у "1-click", однак сила зв'язку статистично мінімальна.

Гіпотеза 6: Прохідність при оплаті звичайною картою може бути нижчою ніж при оплаті Apple Pay/Google Pay



- Прохідність при оплаті картою і Apple Pay практично однакова ($\sim 0.54-0.55$).

- Google Pay має трохи нижчу прохідність (~ 0.50).

Кореляція = -0.0161 :

- Це практично нульовий зв'язок.
- Спосіб оплати не має статистично значущого впливу на прохідність.

Висновок: Гіпотеза не підтверджується.

Прохідність при оплаті Apple Pay, Google Pay та звичайною картою практично однакова. Кореляція -0.0161 вказує на відсутність реального статистичного зв'язку між способом оплати і успішністю транзакції.

Гіпотеза 7: Наявність 3D Secure підвищує прохідність



- Коли `is_secured = false`, прохідність вища (~ 0.60).
- Коли `is_secured = true`, прохідність нижча (~ 0.48).

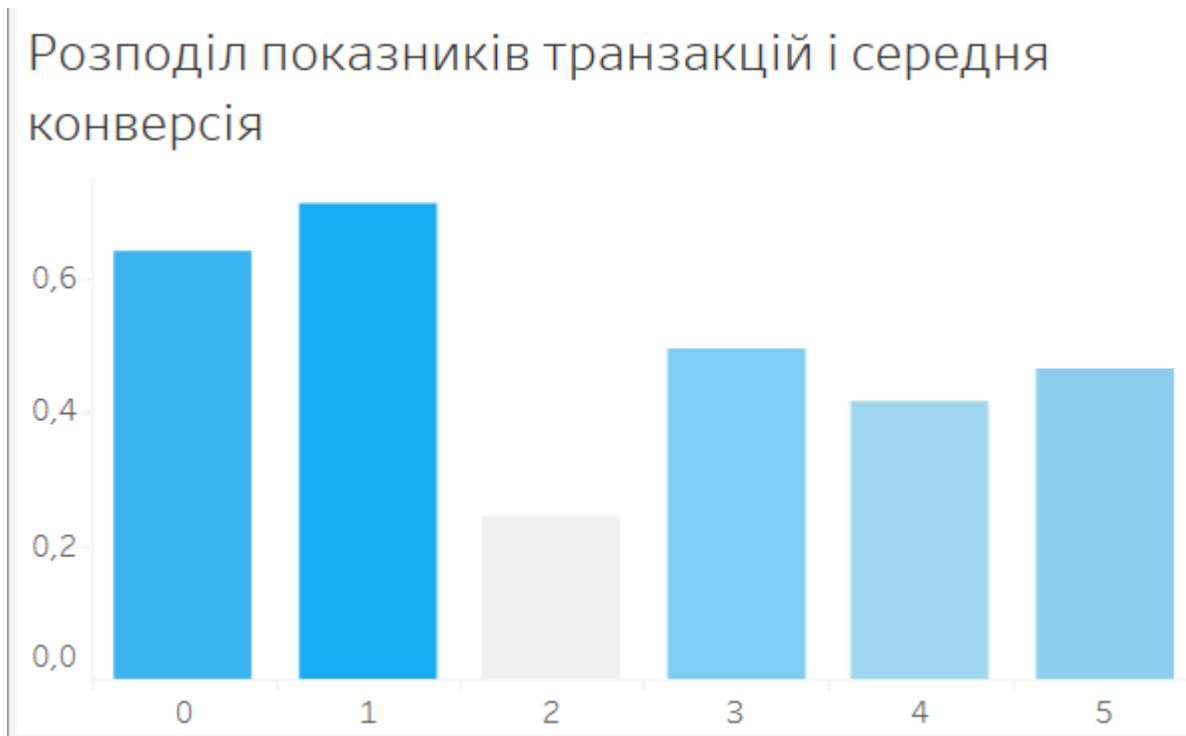
Кореляція = -0.0996 :

- Дуже слабкий від'ємний зв'язок. Тобто наявність 3D Secure не лише не підвищує прохідність, а пов'язана з її зниженням, хоча зв'язок слабкий.

Висновок: Гіпотеза не підтверджується.

Наявність 3D Secure (`is_secured = true`) пов'язана з нижчою прохідністю (~ 0.48 проти ~ 0.60). Кореляція -0.0996 вказує на дуже слабкий від'ємний зв'язок між 3D Secure і успішністю транзакції.

Гіпотеза 8: Дуже великі платежі мають нижчу ймовірність успішної оплати.



- Прокідність найбільша у категоріях 0 і 1 (тобто невеликі суми платежів).
- Прокідність помітно падає у категоріях 2, 3, 4, 5 (тобто у більших сумах платежів).
- Тобто [0,10) та [10,20) мають найбільшу кількість транзакцій і найвищу прохідність. А [40,60) та [60,100) — суми більші, транзакцій менше, і прохідність гірша.

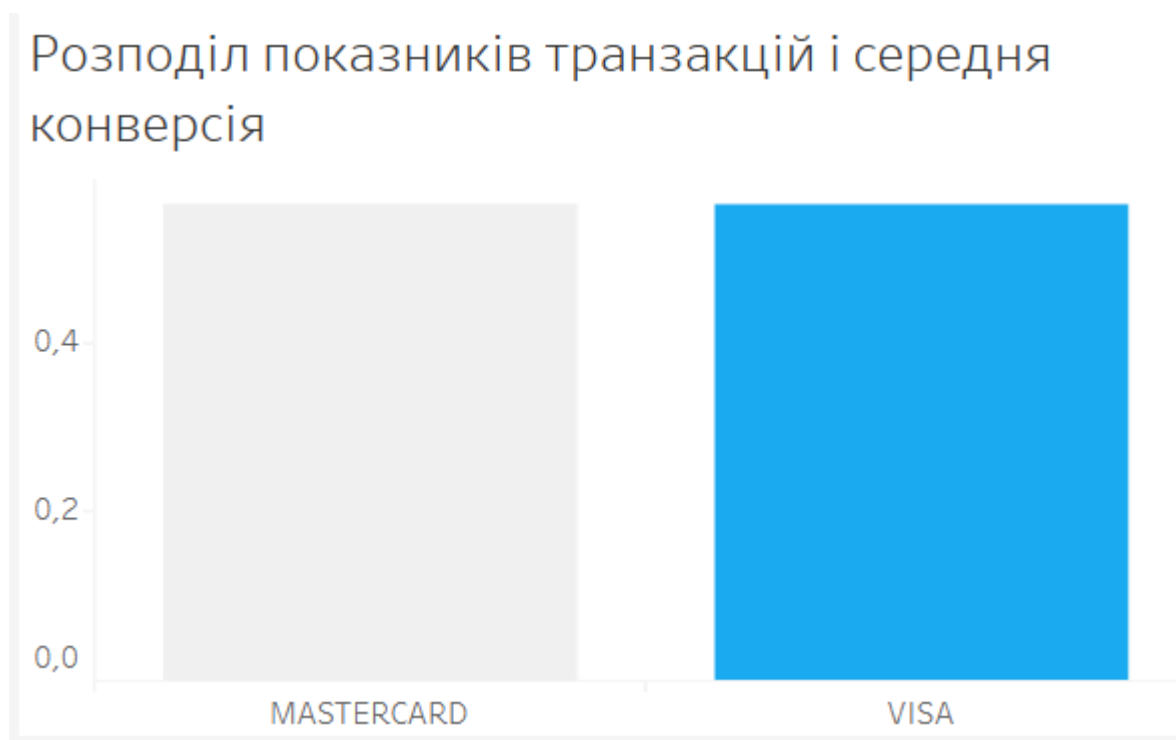
Кореляція = -0.3005 :

- Середньої сили від'ємний зв'язок. Тобто більша сума платежу дійсно пов'язана зі зниженням прохідності.

Висновок: Гіпотеза підтверджується.

Дуже великі платежі мають нижчу прохідність у порівнянні з малими сумами. Кореляція -0.3005 вказує на середній від'ємний зв'язок між розміром платежу і ймовірністю успішної оплати.

Гіпотеза 9: Прохідність може бути вищою на деяких видах карток



- Прохідність для Mastercard і VISA практично однакова (~ 0.54).

Кореляція = 0.0030:

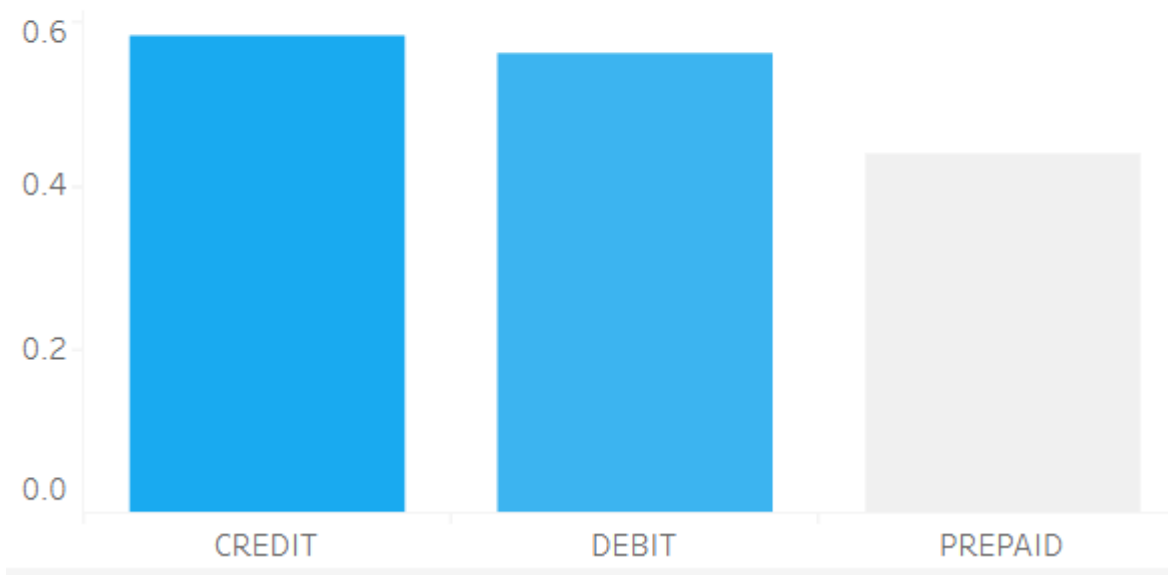
- Практично нульова кореляція, тому статистичного зв'язку між типом картки і прохідністю немає.

Висновок: Гіпотеза не підтверджується.

Вид картки (`card_brand`) практично не впливає на прохідність. Кореляція 0.0030 вказує на відсутність зв'язку між брендом картки і успішністю оплати.

Гіпотеза 10: Прохідність може бути вищою на деяких типах карток

Розподіл показників транзакцій і середня конверсія



- Найвища прохідність у CREDIT карток (~0.60).
- Трохи нижча прохідність у DEBIT карток (~0.57).
- Значно нижча прохідність у PREPAID карток (~0.45).

Кореляція = -0.0731 :

- Дуже слабкий від'ємний зв'язок. Формально різниця є, але статистично зв'язок практично відсутній.

Висновок: Гіпотеза частково підтверджується.

Різниця в прохідності між типами карток справді існує: кредитні картки мають дещо вищу прохідність, ніж дебетові і передплачені. Проте кореляція -0.0731 вказує на дуже слабкий зв'язок, який може бути випадковим і не є сильним фактором впливу.

Гіпотеза 11: Картки з певних країн (**card_country**) мають вищу або нижчу прохідність.

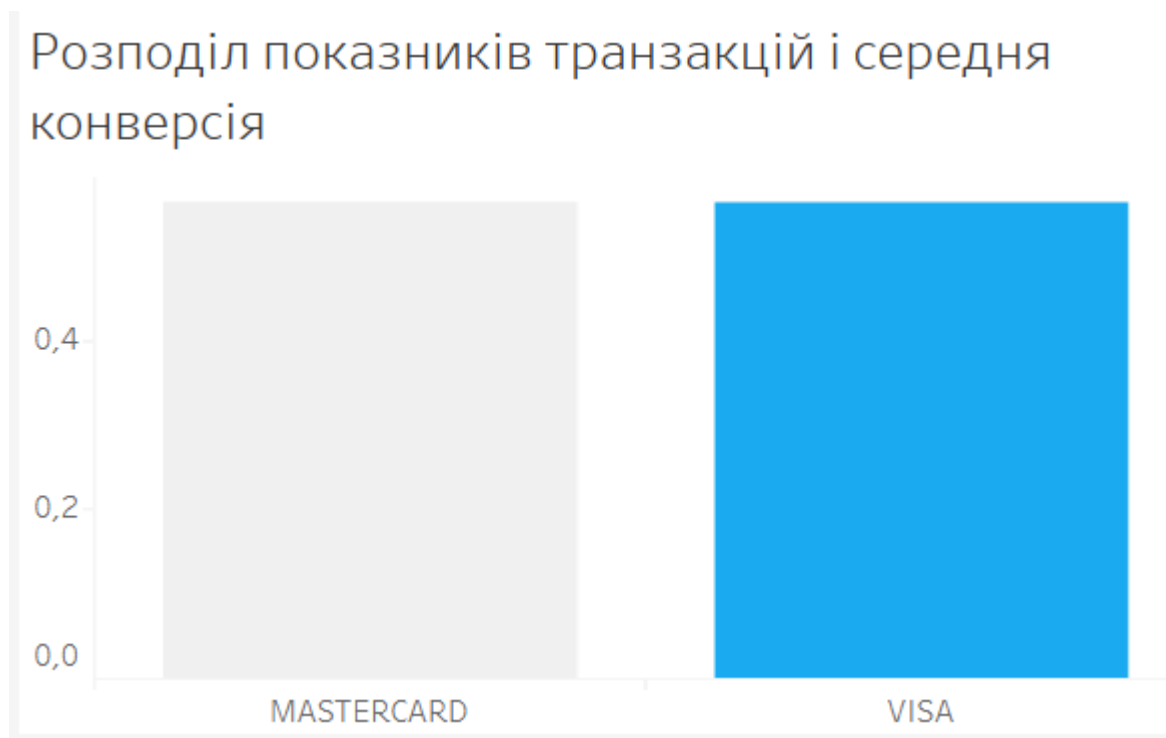
Кореляція = -0.0208 :

- Практично нульовий зв'язок, тому статистично країна випуску картки практично не впливає на прохідність.

Висновок: Гіпотеза не підтверджується.

Аналіз показав, що країна випуску картки (**card_country**) майже не впливає на успішність оплат. Кореляція -0.0208 вказує на відсутність реального статистичного зв'язку.

Гіпотеза 12: Країна IP користувача та країни з якої випускалася карта можуть відрізнятися, а це може бути ризик шахрайства, який впливає на прохідність.



- **true** (країна IP збігається з країною картки) має трохи вищу прохідність (~0.56).

- **false** (країна IP не збігається з країною картки) має трохи нижчу прохідність (~0.53).

Кореляція = -0.0208:

- Практично нульова кореляція, тому статистично відмінність між IP-країною та країною картки майже не впливає на прохідність.

Висновок: Гіпотеза не підтверджується.

Невідповідність між країною IP користувача та країною випуску картки не має суттєвого впливу на прохідність транзакцій. Кореляція -0.0208 свідчить про відсутність реального статистичного зв'язку.

Гіпотеза 13: Чим вищий номер інвойсу (invoice_number**) тим користувач може бути досвідченішим, а відповідно його прохідність може бути вищою**



- При збільшенні номера інвойсу (**invoice_number**) середня прохідність поступово зростає.

- На низьких інвойсах багато нестабільності та нижчі значення.
- Починаючи з інвойсів близько 50–60 прохідність стабілізується і тримається на високому рівні (~0.8–0.9).

Кореляція = 0.2392:

- Слабкий позитивний зв'язок. Тобто номер інвойсу має вплив на прохідність, але вплив помірний і слабкий.

Висновок: Гіпотеза частково підтверджується.

Дані показують, що із збільшенням номера інвойсу (*invoice_number*) середня прохідність зростає, що може вказувати на більший досвід або лояльність користувачів. Кореляція 0.2392 вказує на слабкий позитивний зв'язок, тому вплив існує, але не є дуже сильним.

Гіпотеза 14: Більша кількість спроб (**retry_number**) зменшує ймовірність успішної оплати.



- Найвища прохідність при **retry_number = 0** (~0.70).
- При збільшенні **retry_number** (1, 2, 3, 4) прохідність помітно нижча (~0.43–0.45).
- Значення **-1** має найнижчу прохідність (~0.20), але це окреме некоректне значення.

Кореляція = 0.0142:

- Практично нульова кореляція. Хоч графік показує падіння прохідності, статистично зв'язок дуже слабкий і майже не зафіксований через особливості даних.

Висновок: Гіпотеза візуально підтверджується, але статистично не підтверджується.

Зі збільшенням кількості спроб прохідність дійсно зменшується, що видно на графіку. Проте кореляція 0.0142 вказує на практично нульовий статистичний зв'язок, тому формально гіпотеза не підтверджена за кореляційним аналізом.

Гіпотеза 15: Чим більше замовлень (count_orders**) на одного користувача — тим вища загальна прохідність.**

Кореляція = -0.0476:

- Дуже слабкий від'ємний зв'язок. Тобто фактично кількість замовлень на користувача не має реального впливу на загальну прохідність.

Висновок: Гіпотеза не підтверджується.

Кількість замовлень (**count_orders**) на одного користувача практично не впливає на прохідність оплат. Кореляція -0.0476 вказує на відсутність статистично значущого зв'язку.

Гіпотеза 16: Велика кількість emails(count_emails**) може свідчити про створення багатьох пошт, а відповідно і шахрайства, яке призводить до низької прохідності**

Кореляція = 0.0659:

- Дуже слабкий позитивний зв'язок. Тобто кількість email'ів практично не впливає на прохідність.

Висновок: Гіпотеза не підтверджується.

Велика кількість email'ів (`count_emails`) не має суттєвого впливу на прохідність оплат. Кореляція 0.0659 вказує на відсутність значущого статистичного зв'язку між кількістю emails та рівнем успішності транзакцій.

Висновок

На основі проведеного аналізу прохідності було виявлено, що найбільший вплив на успішність транзакцій мають тип продукту бізнесу (`channel_id`), тип платежу (`order_type`), сума транзакції (`bin_amount_usd`) та певною мірою номер інвойсу (`invoice_number`). Інші фактори, такі як місяць, день тижня, тип картки чи наявність 3D Secure, показали слабкий або відсутній вплив на прохідність. Зокрема, розбіжність між IP країною та країною картки, а також кількість emails або замовлень практично не впливають на успішність оплат.