

Прогнозування відтоку клієнтів

у телекомунікаційній компанії Vodafone



Безрукавий Ярослав

BigDataLab

Проблематика

Зменшення доходів

Клієнти, які покидають компанію, більше **не приносять їй доход**. Це може значно вплинути на фінансові показники компанії.

Збільшення витрат

За даними різних досліджень, залучення нових клієнтів може коштувати в **5 разів** дорожче, ніж утримання існуючих

Негативний імідж

Високий рівень відтоку може сигналізувати про **проблеми в послугах**, наданих компанією, це негативно впливає на репутацію

Припущення

Вартість утримання

5 грн

Витрати компанії на
utrимання існуючих клієнтів

Вартість залучення

25 грн

Витрати компанії на
залучення нових клієнтів

За умови що залучення нових клієнтів в **5 разів більше**, ніж утримання існуючих

Цілі

Модель прогнозування

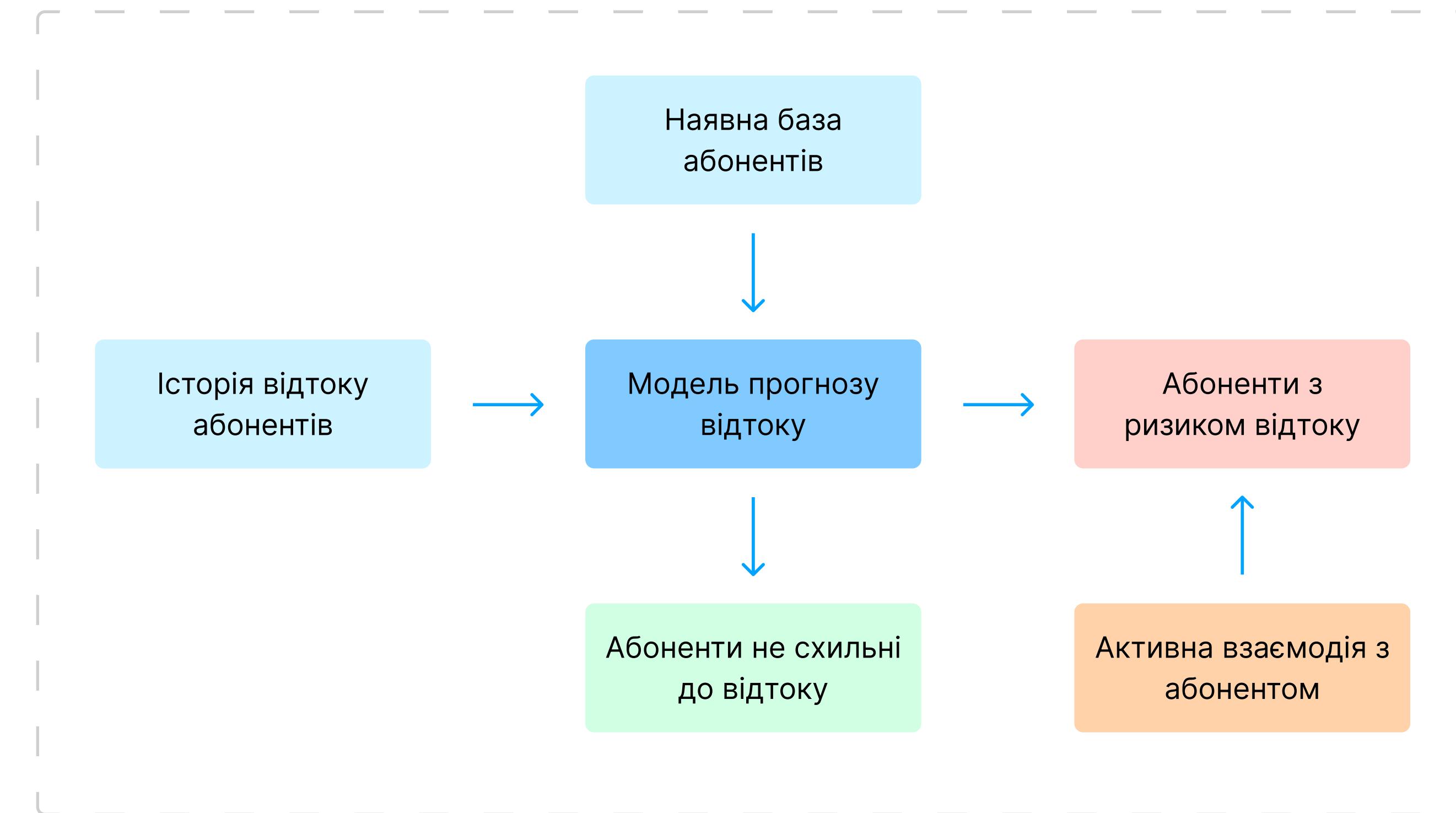
Розробка та впровадження моделі машинного навчання, для прогнозування ймовірності відтоку клієнтів

Аналіз відтоку

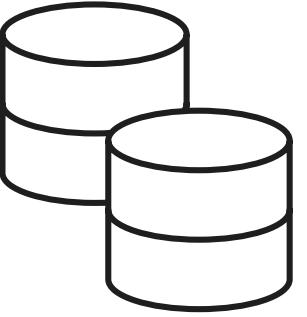
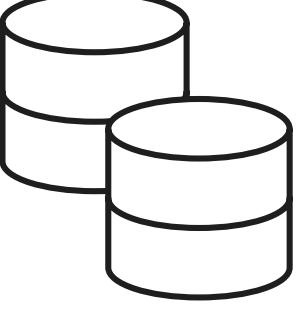
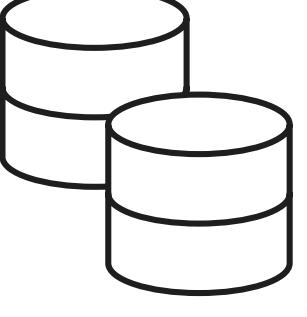
Проаналізувати отримані дані та визначити фактори, що впливають на відтік абонентів

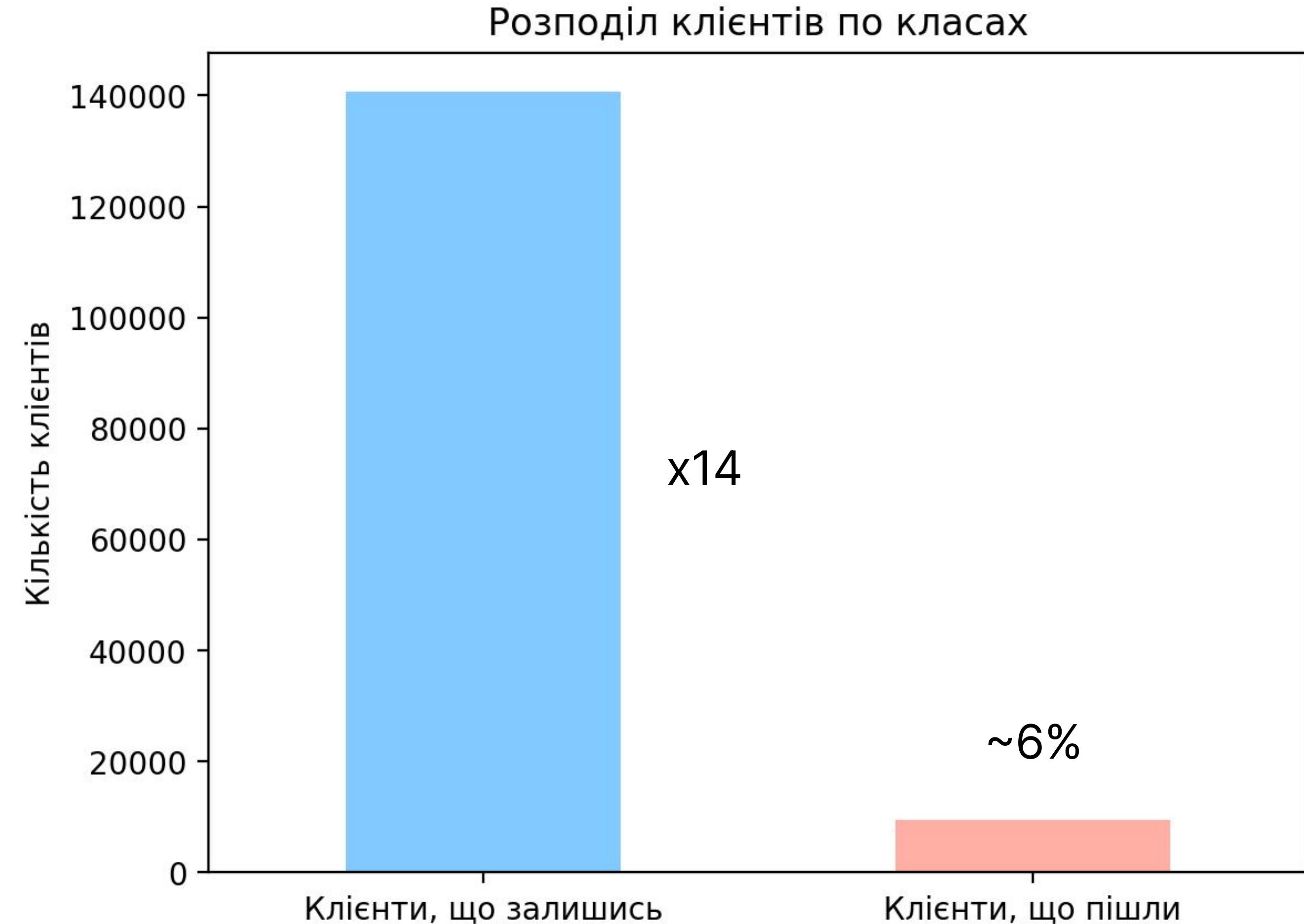
Мінімізація витрат компанії

Використати отриману інформацію для проведення більш успішної кампанії взаємодії з абонентами склонними до відтоку та оптимізації процесів компанії у наданні послуг абонентам



Початкові дані

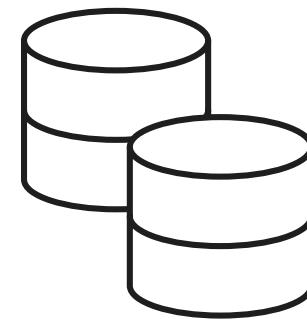
- 🔑 abon_id
-  **FE** - профіль клієнта
2x **150k** записів
817 параметрів
-  **B_NUM** - активність з іншими номерами
2x **~669k** записів
8 параметрів
-  **DPI** - трафік клієнта по додатками
2x **~6.7m** записів
6 параметрів



Аналіз даних

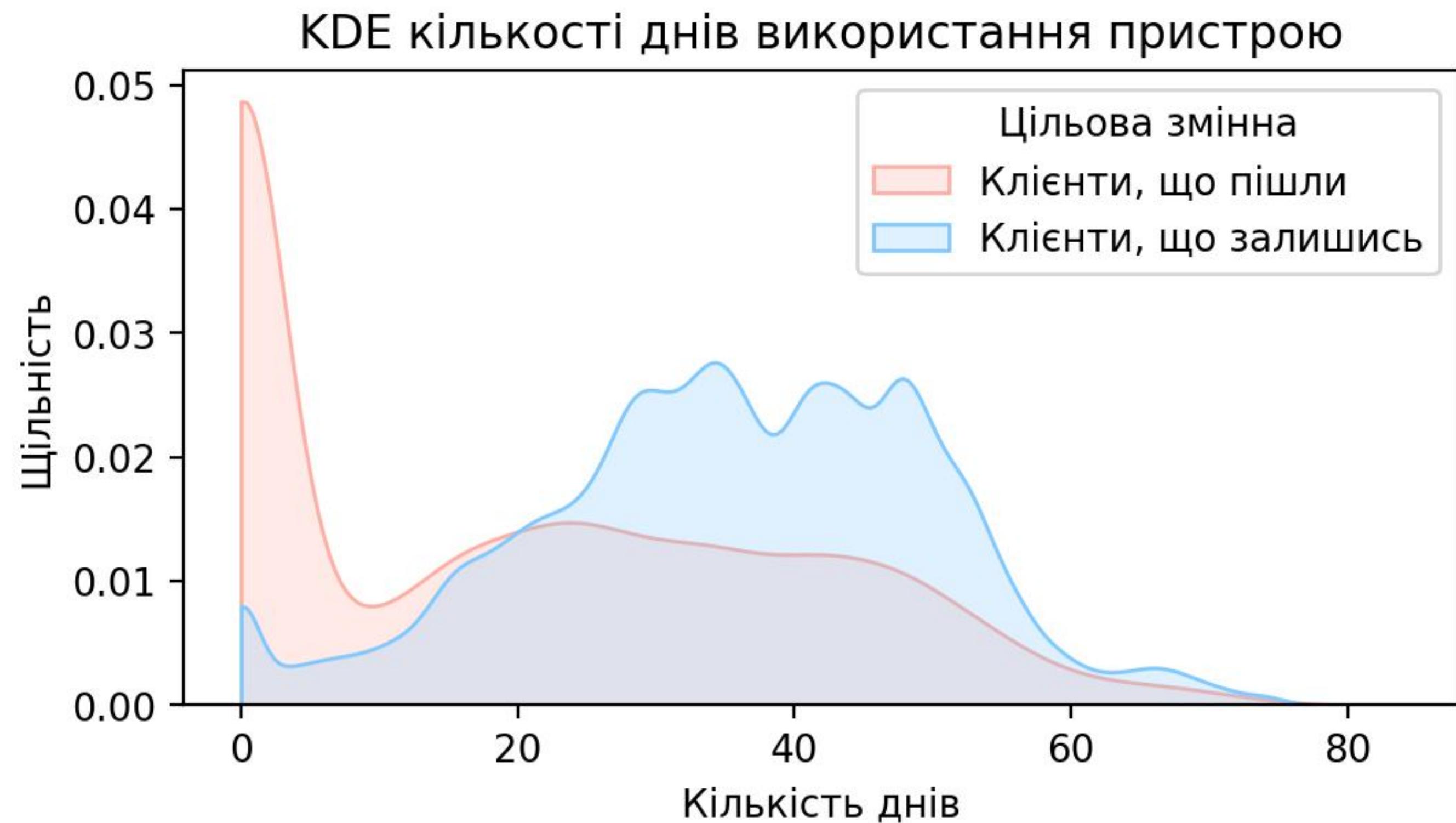
FE

Профіль абонента



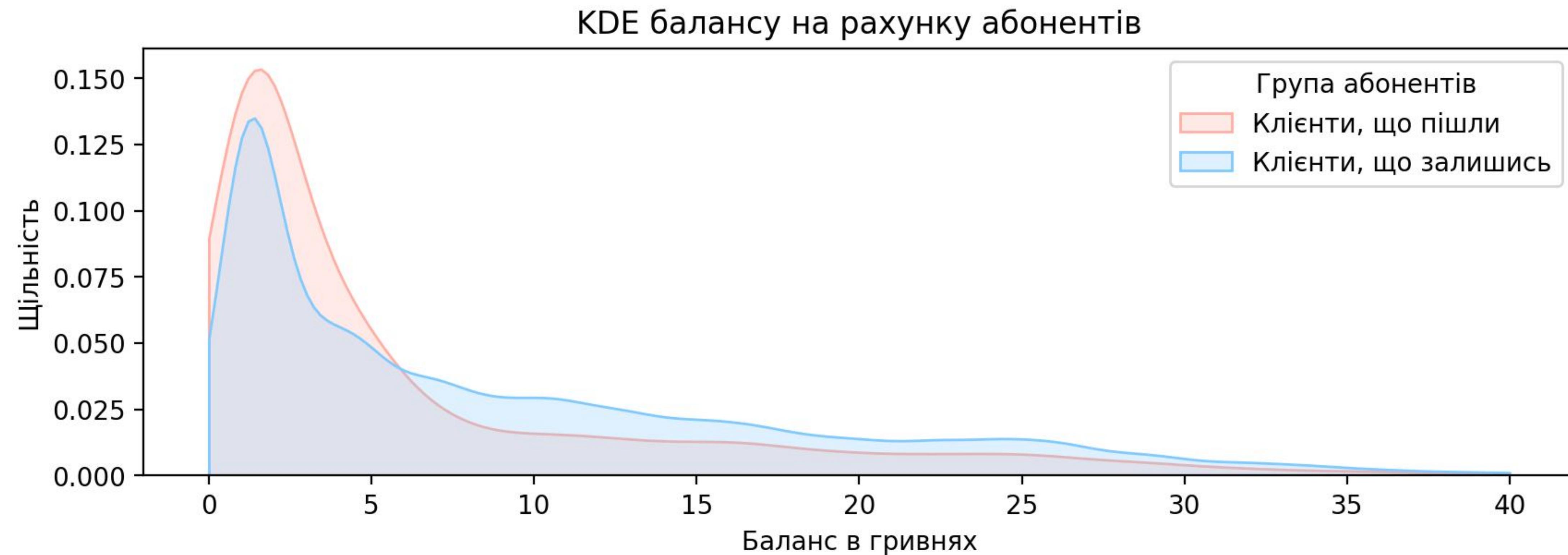
FE - профіль абонента

2x **150k** записів
817 параметрів



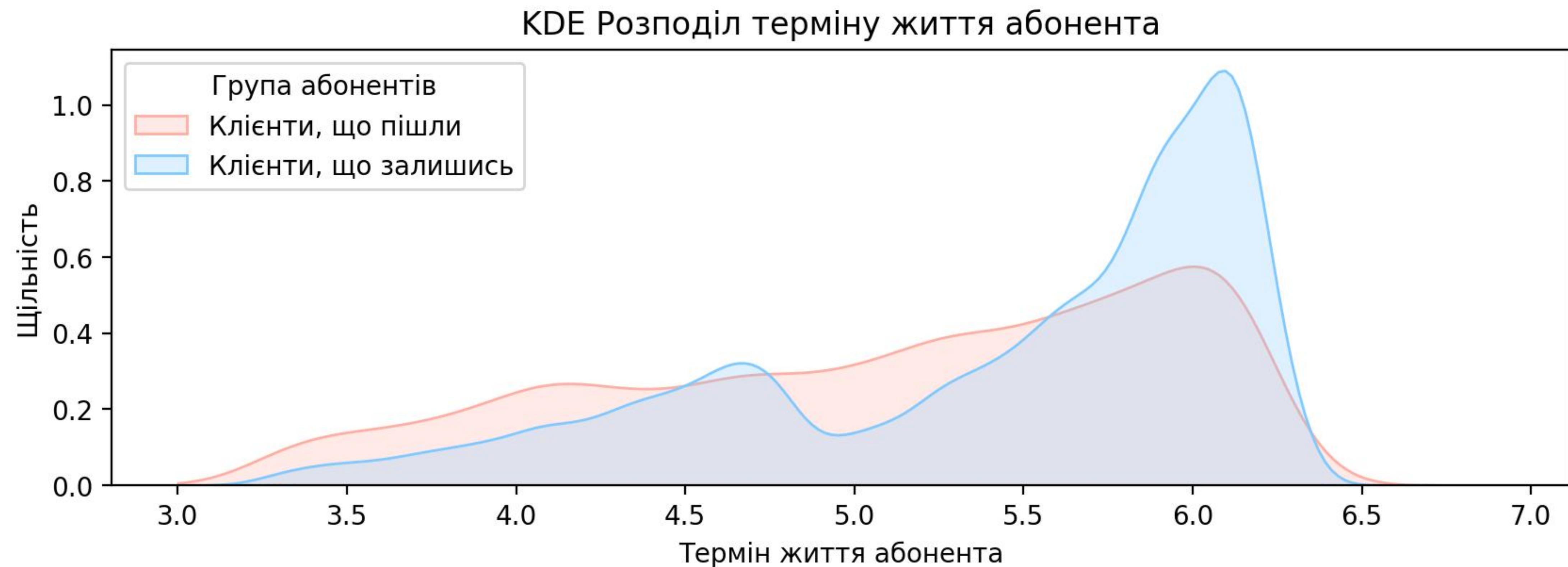
FE

Розподіл балансу на рахунку абонентів



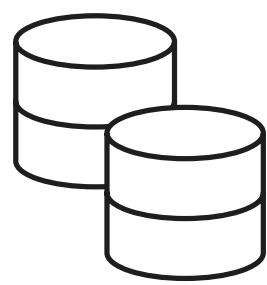
FE

Розподіл терміну життя абонента

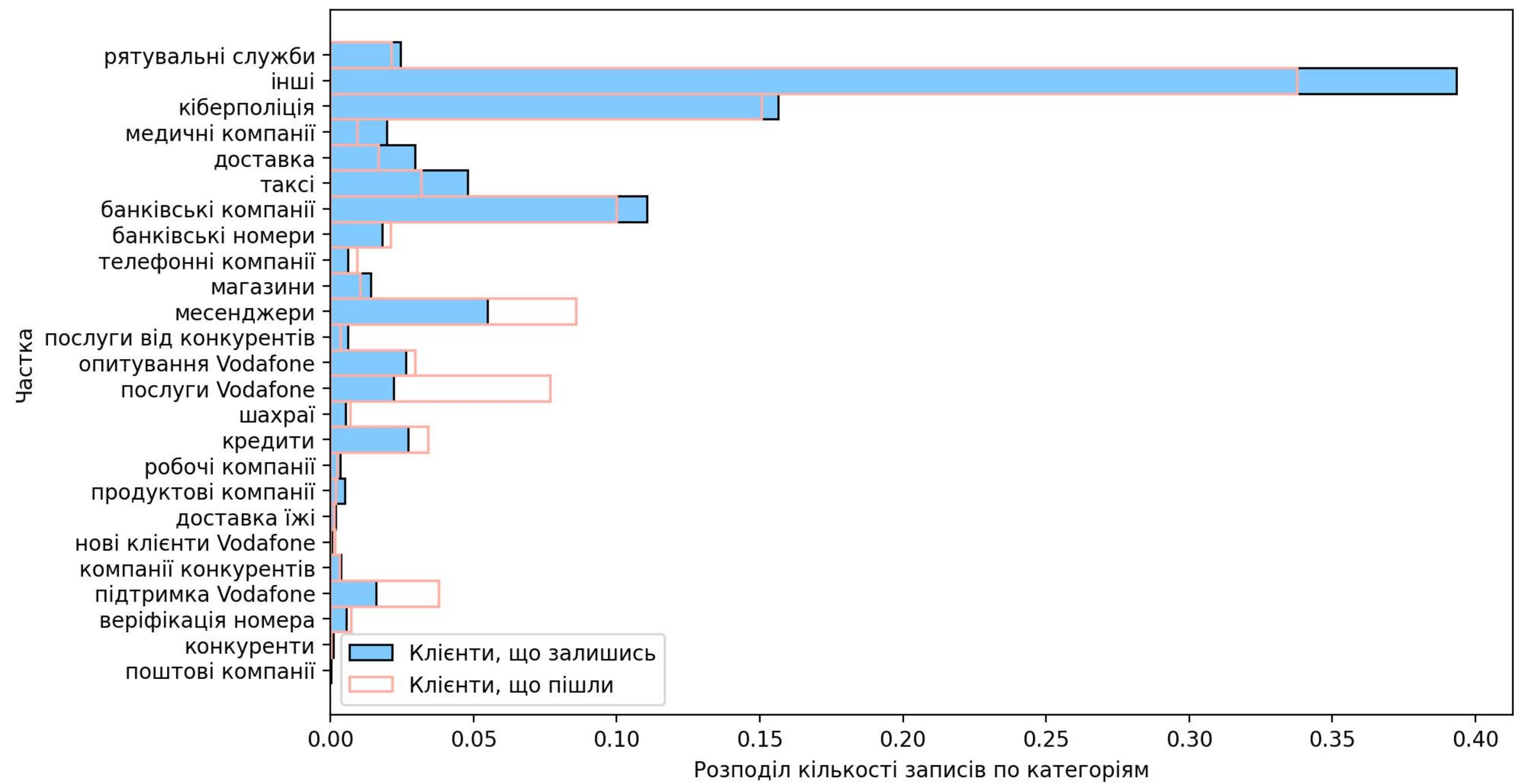


B_NUM

Взаємодія з іншими номерами



B_NUM - активність
з іншими номерами
2x ~669k записів
8 параметрів

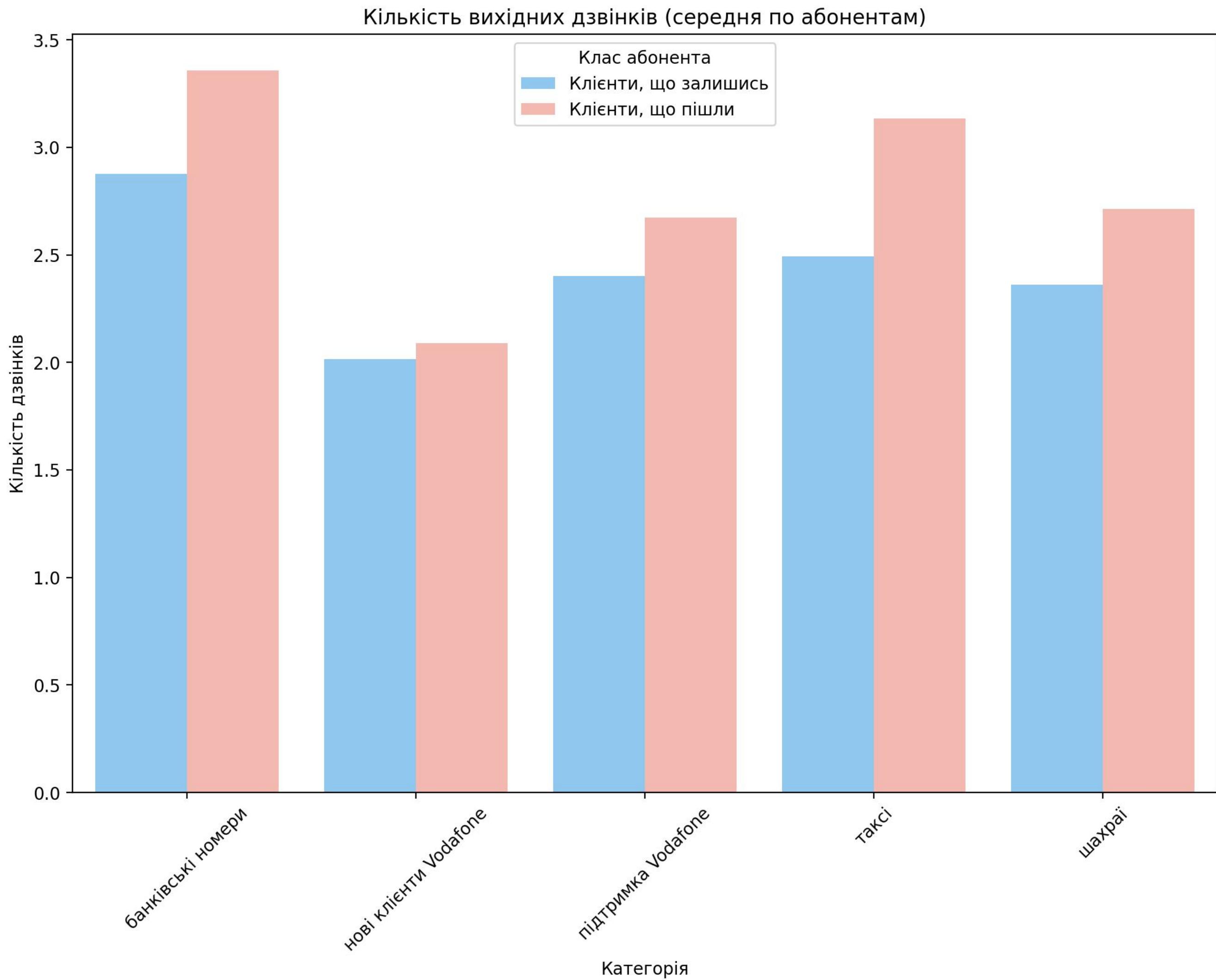


B_NUM

Кількість вихідних дзвінків
(середня по абонентам)

В середньому клієнти з групу відтоку
частіше взаємодіють з:

- Підтримкою Vodafone
- Сервісами Таксі
- Шахраями

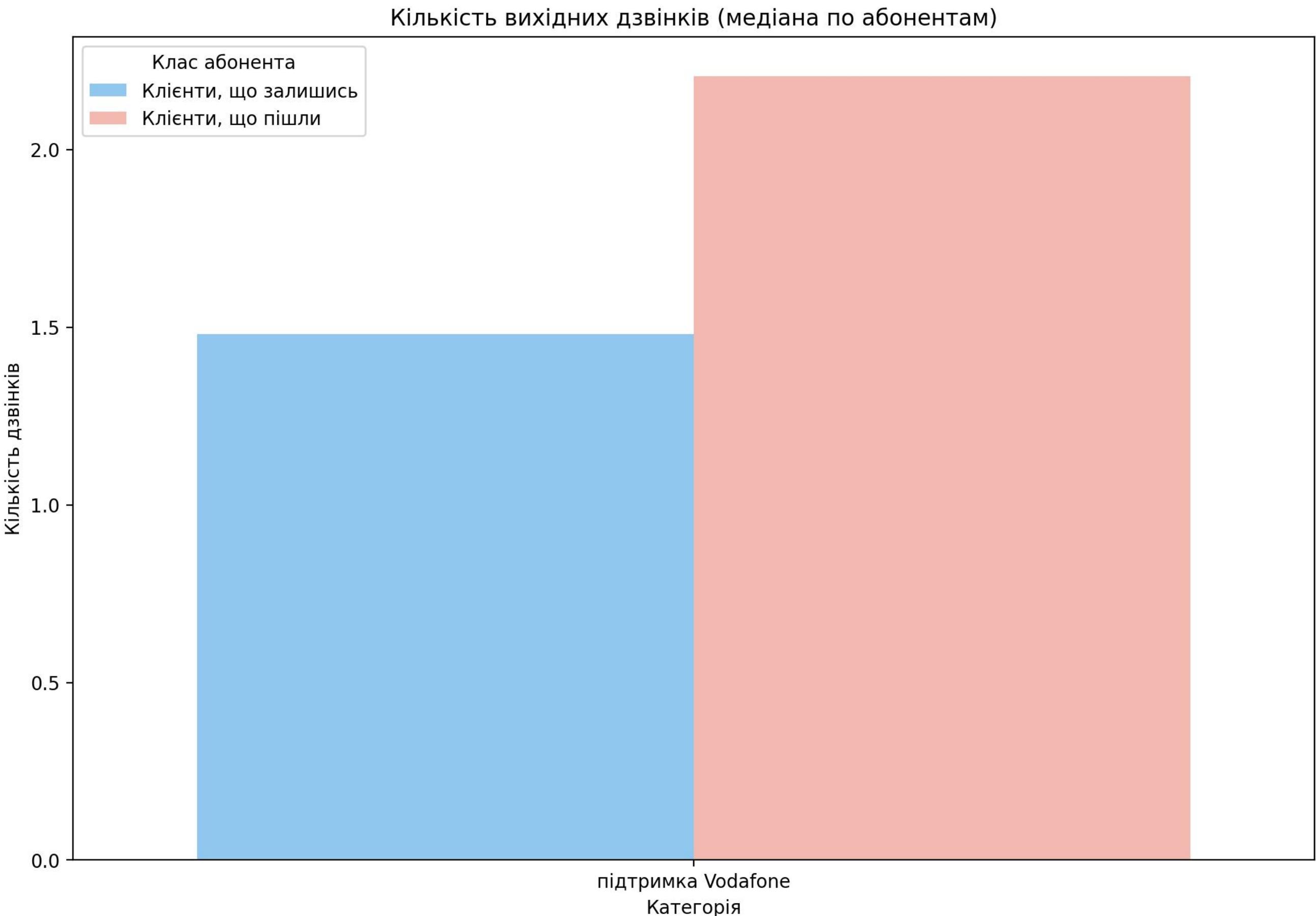


B_NUM

Кількість вихідних дзвінків
(медіана по абонентам)

Медіана також свідчить про те що
абоненти, з групи відтоку, частіше
звертались до підтримки компанії
Vodafone

Рекомендація: Дослідити, чи стикалися
клієнти з групи відтоку з певними
проблемами та чи вдалося підтримці
Vodafone їх вирішити

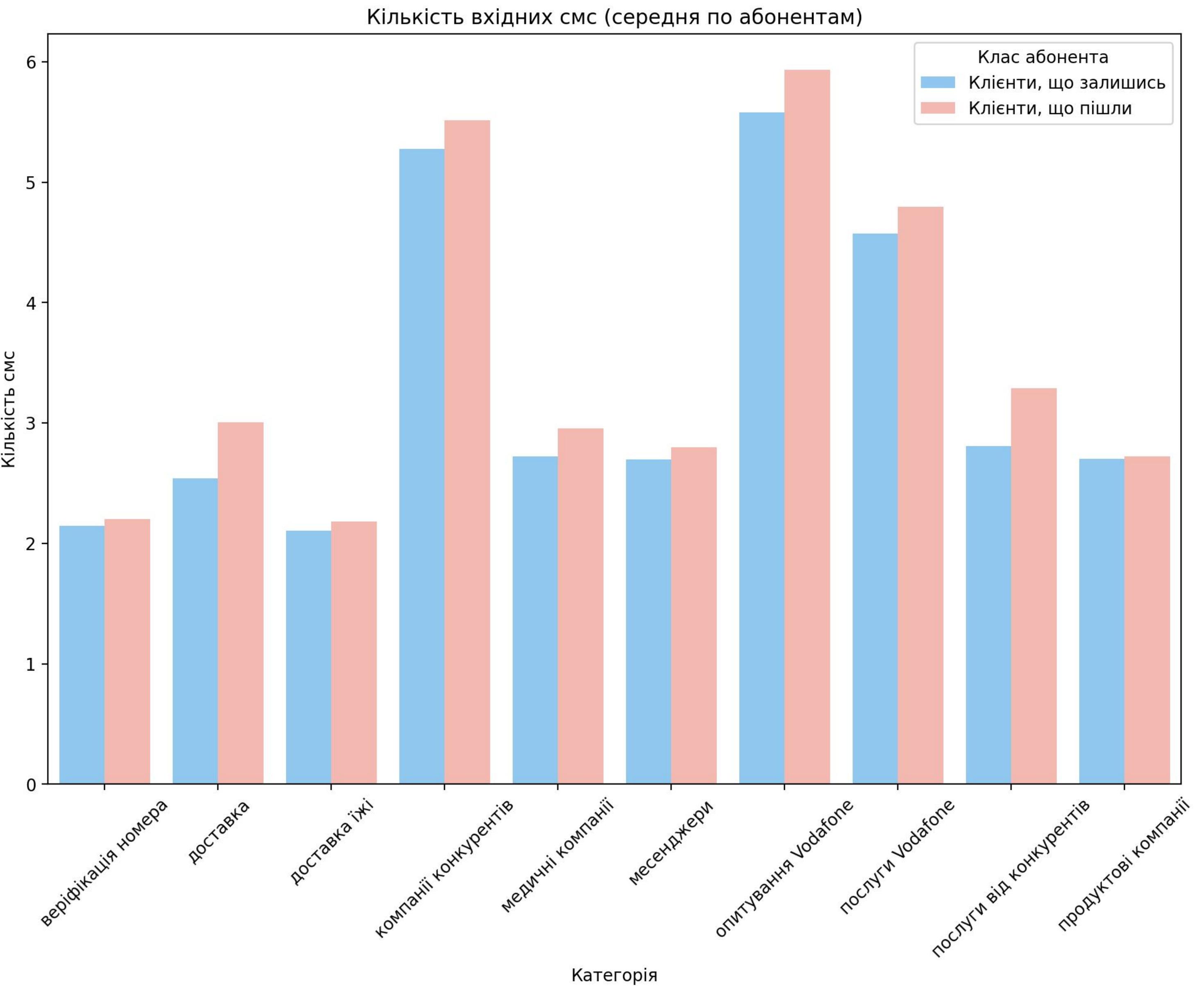


B_NUM

Кількість вхідних СМС
(середня по абонентам)

В середньому клієнти з групи “відтік”
отримували більше СМС від:

- Vodafone (опитування та послуги)
- Послуги від конкурентів

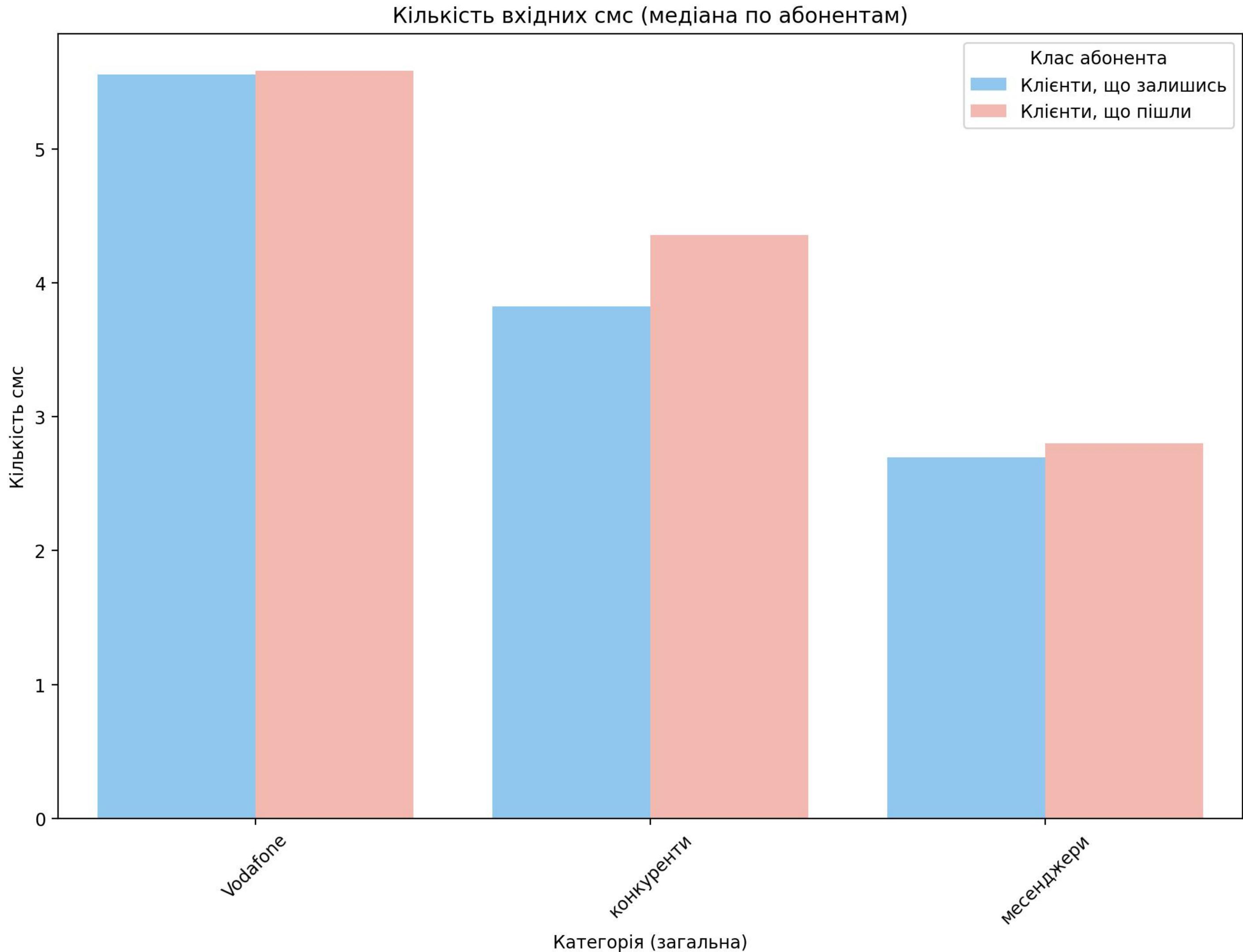


B_NUM

Кількість вхідних СМС
(медіана по абонентам)

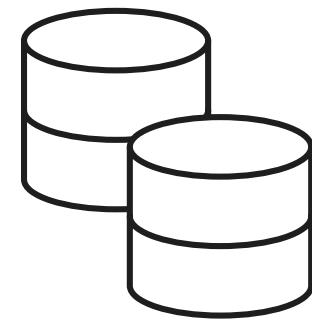
Медіана показує що абоненти, з групу відтоку, частіше отримували СМС від компаній конкурентів

З наявних даних варто дослідити, які послуги від конкурентів були запропоновані абонентам з групи “відтоку”



DPI

Трафік клієнта по додатками

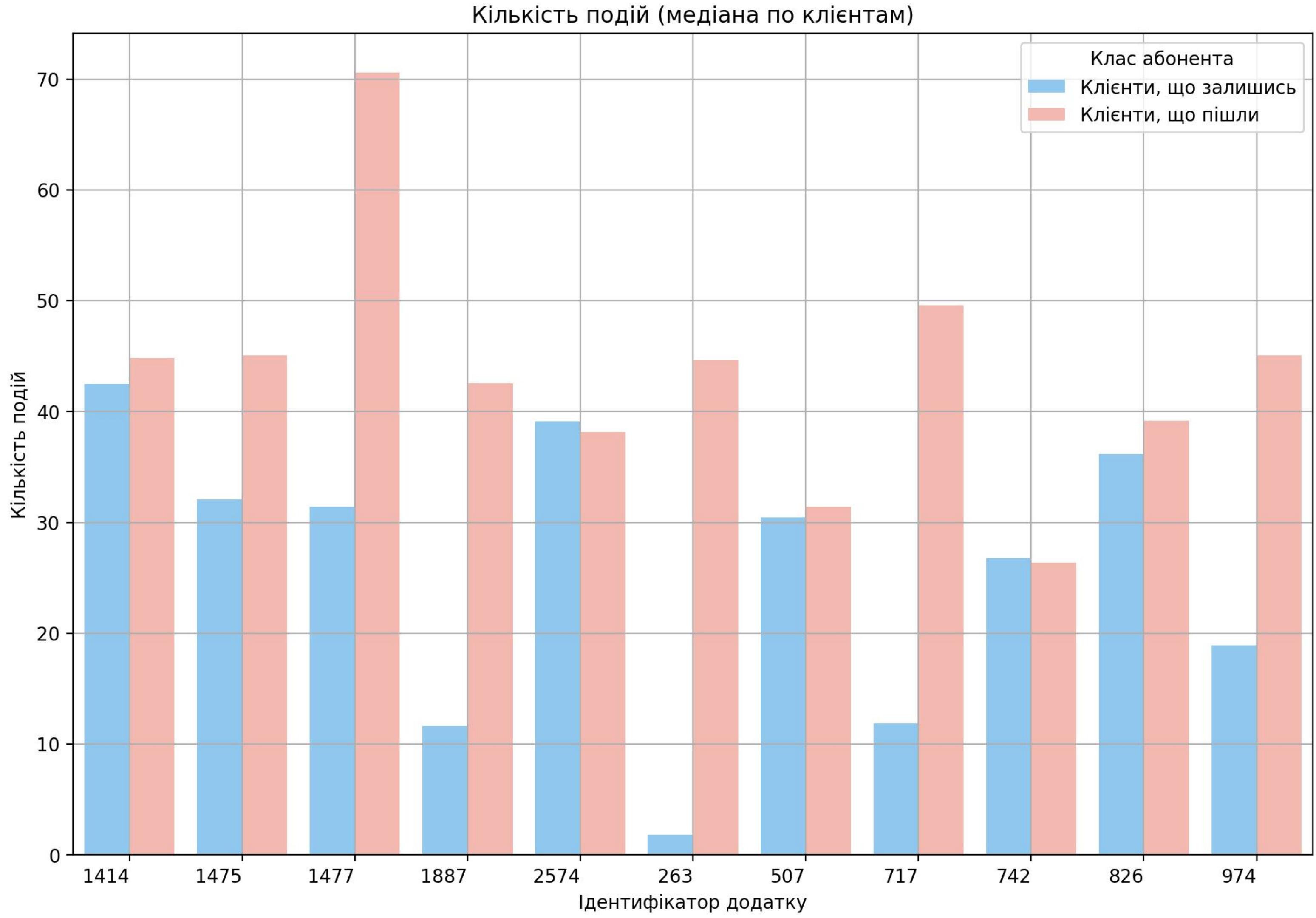


DPI - трафік клієнта по додатками

2x ~6.7m записів

6 параметрів

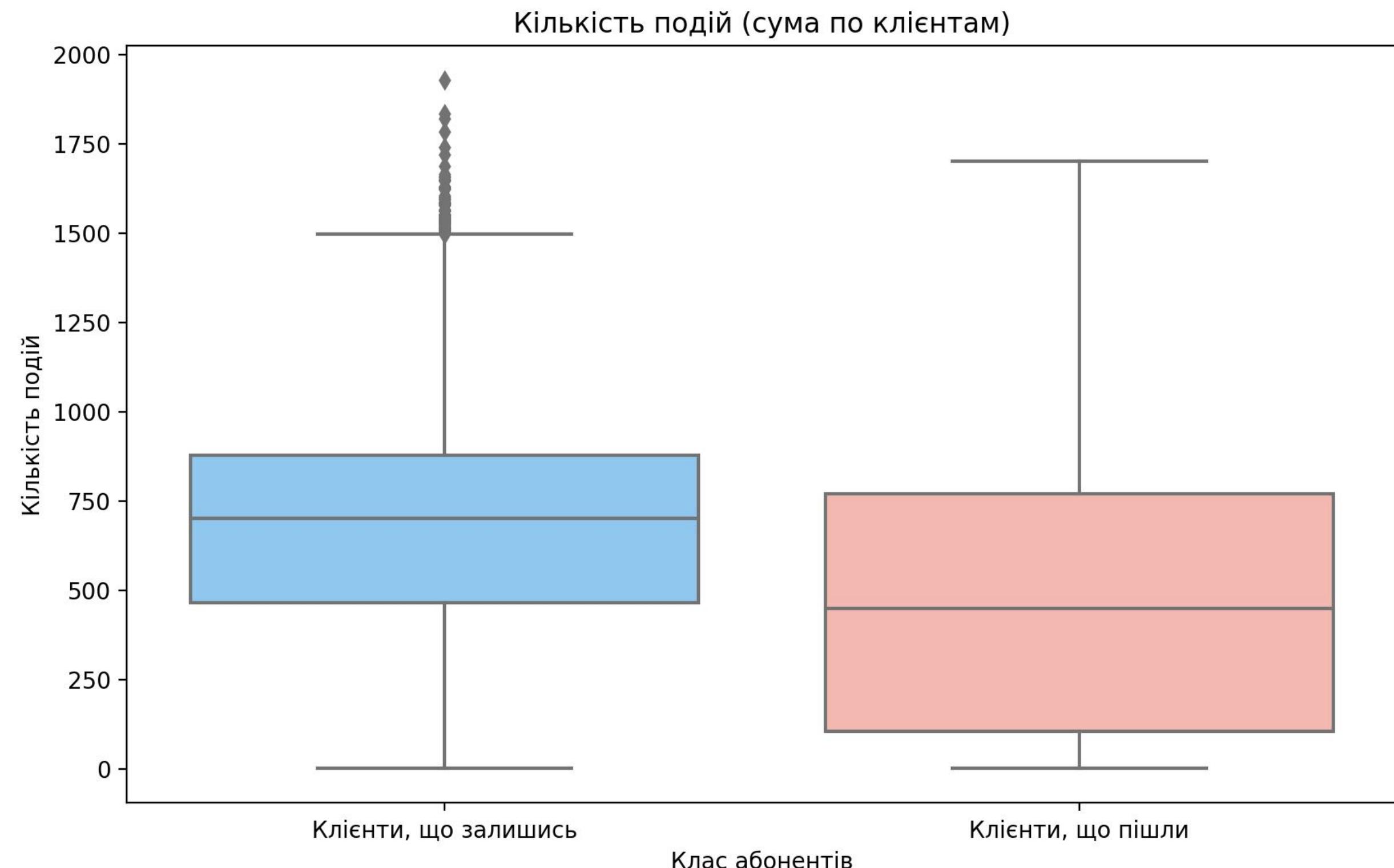
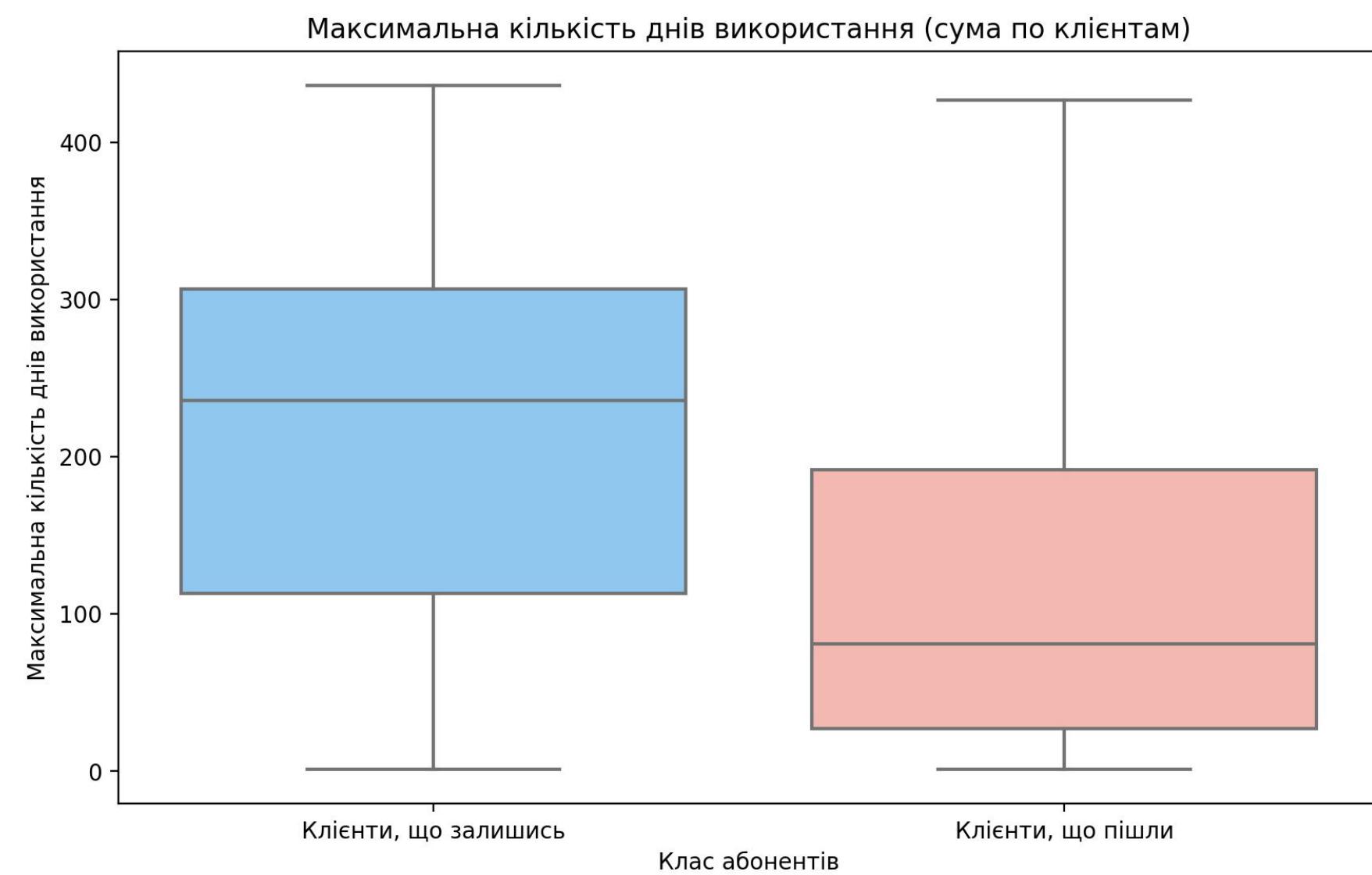
Було помічено певні додатки, які переважають серед клієнтів з групи відтоку



DPI

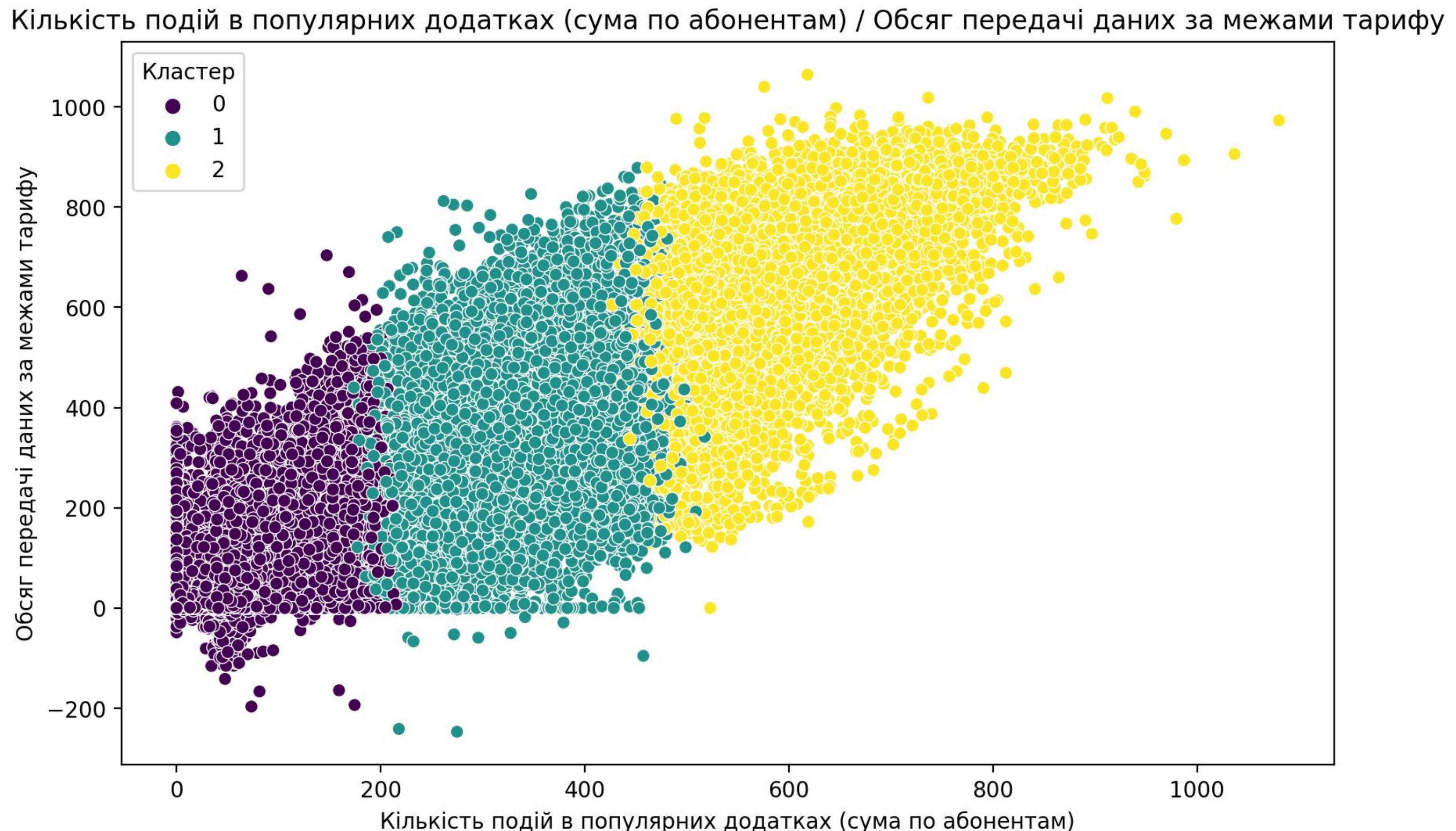
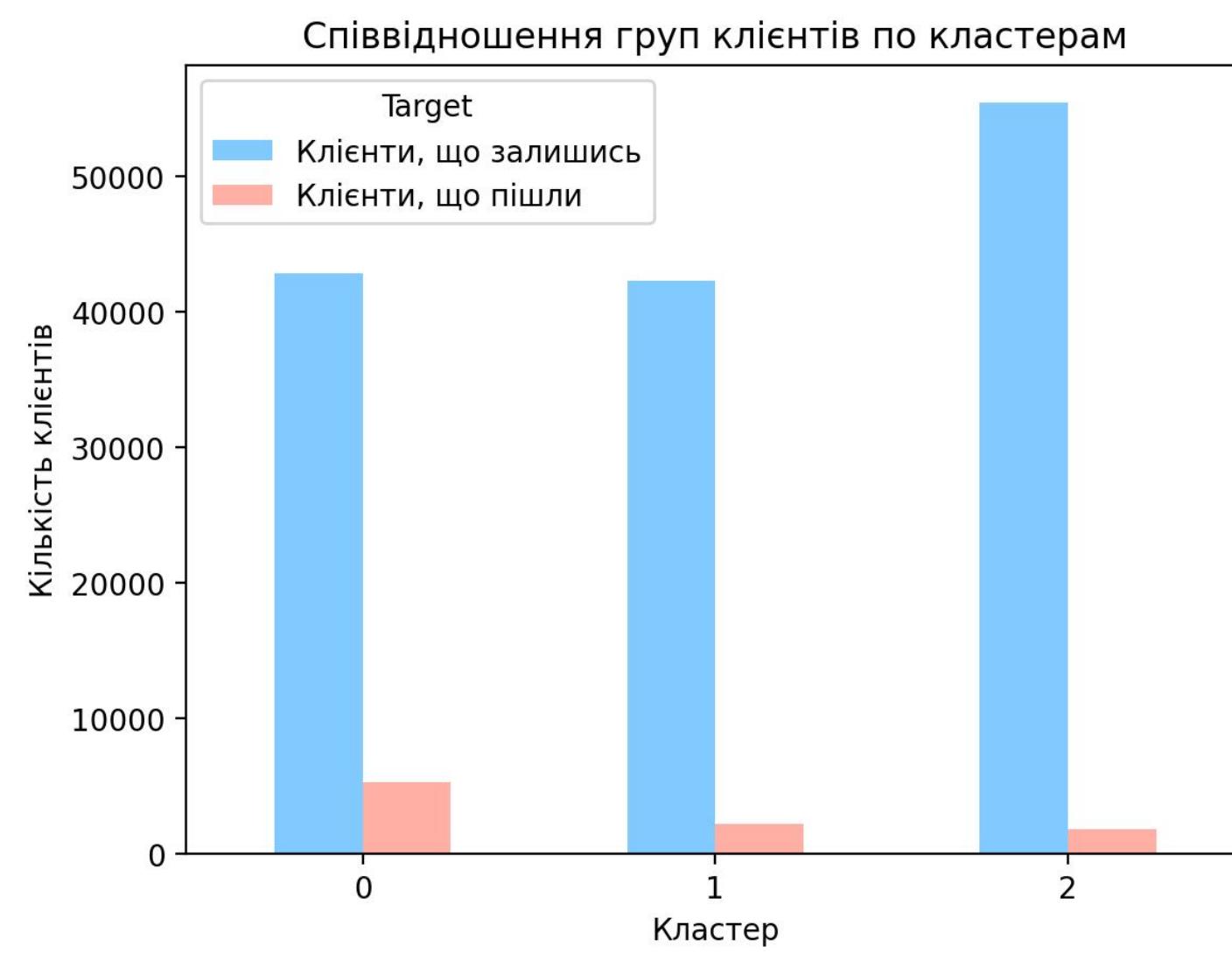
Кількість подій серед популярних додатків

Популярні додатки рідше використовуються клієнтами з групи відтоку



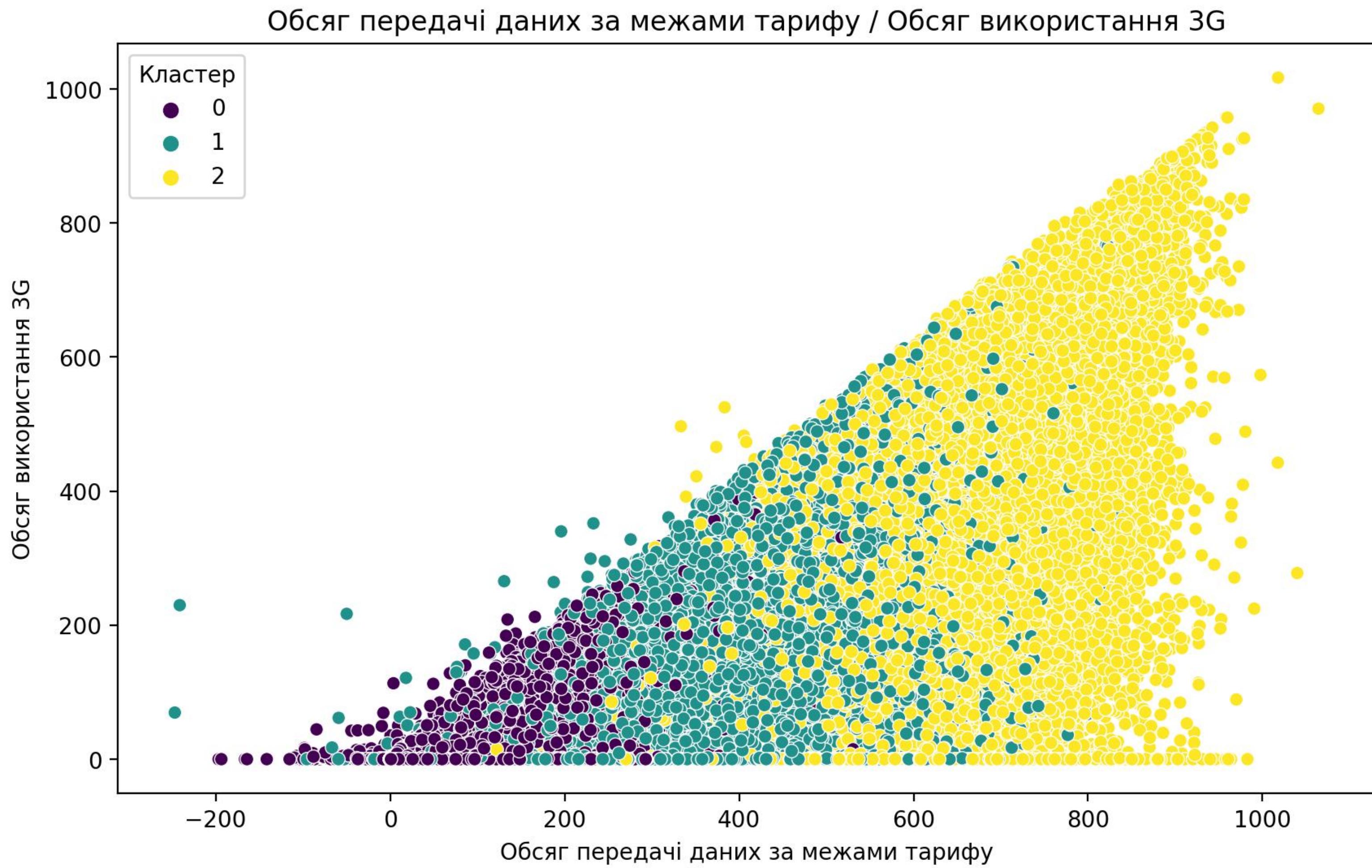
Кластерізація

56% абонентів з групи відтоку
знаходиться в "Кластері 0"



Кластерізація

У майбутньому це можна використати в рекламній кампанії, щоб запропонувати абонентам найбільш підходящі для них тарифи та послуги в залежності від їх потреб

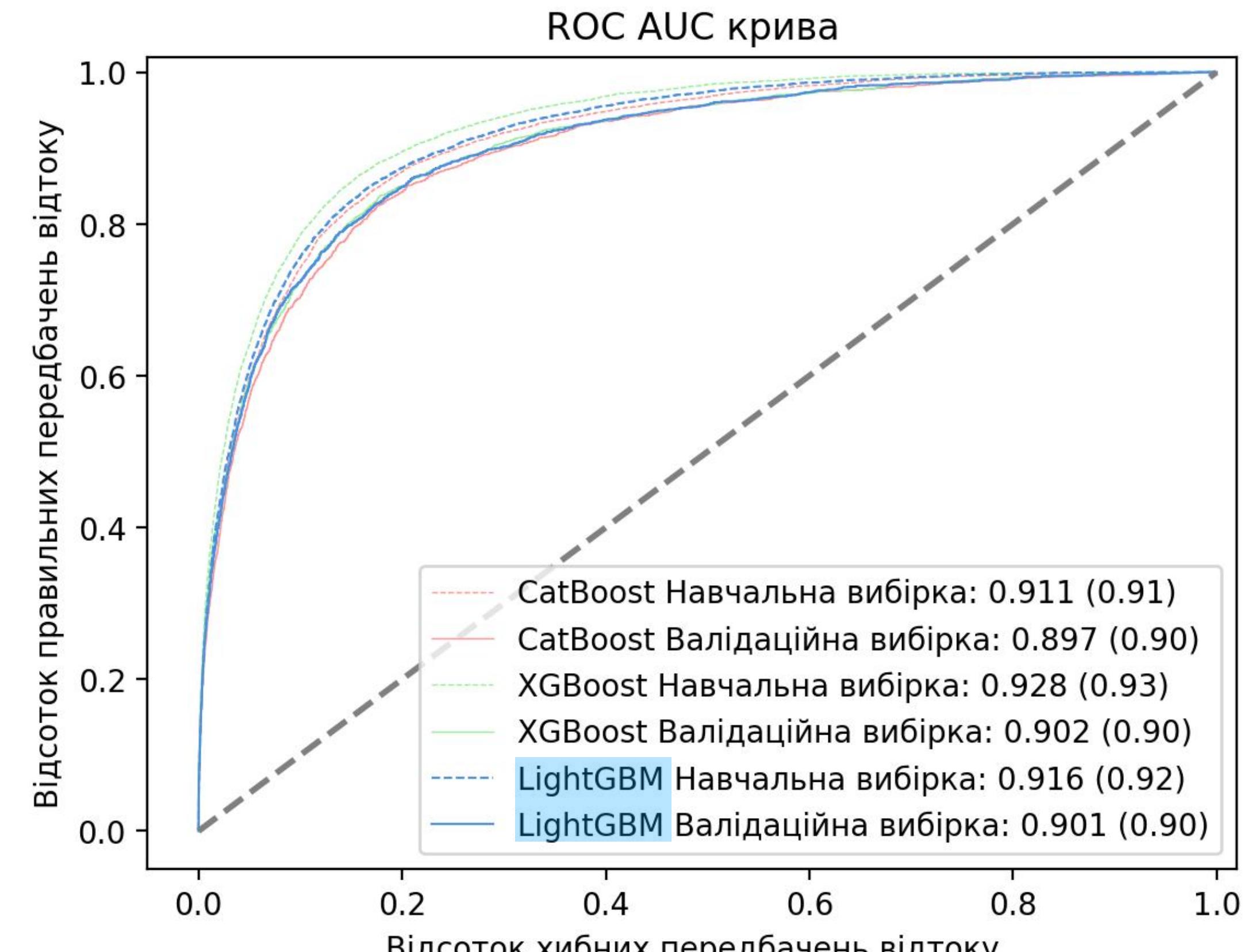


Вибір моделі та навчання

Навчання моделі



Розділені випадковим чином,
зі збереженням пропорцій класів



Метрики

- ↑ **ROC AUC** - Загальна ефективність моделі
- ↑ **F score (beta = 0.5)** - Точність має пріорітет

Калібрування гіперпараметрів

```
model_params = {
    "lambda_l1": 2.91,
    "lambda_l2": 5.57,
    "learning_rate": 0.075,
    "num_leaves": 92,
    "feature_fraction": 0.3888218905277871,
    "bagging_fraction": 0.26715088946500626,
    "max_depth": 17,
    "n_estimators": 100,
    "lambda_l1": 8,
    "lambda_l2": 5,
    "learning_rate": 0.018,
    "num_leaves": 14,
    "feature_fraction": 0.6803603979260223,
    "bagging_fraction": 0.6735621254996546,
    "max_depth": 11,
    "n_estimators": 350,
    "random_state": 42,
    "seed": 42,
    "objective": "binary",
```

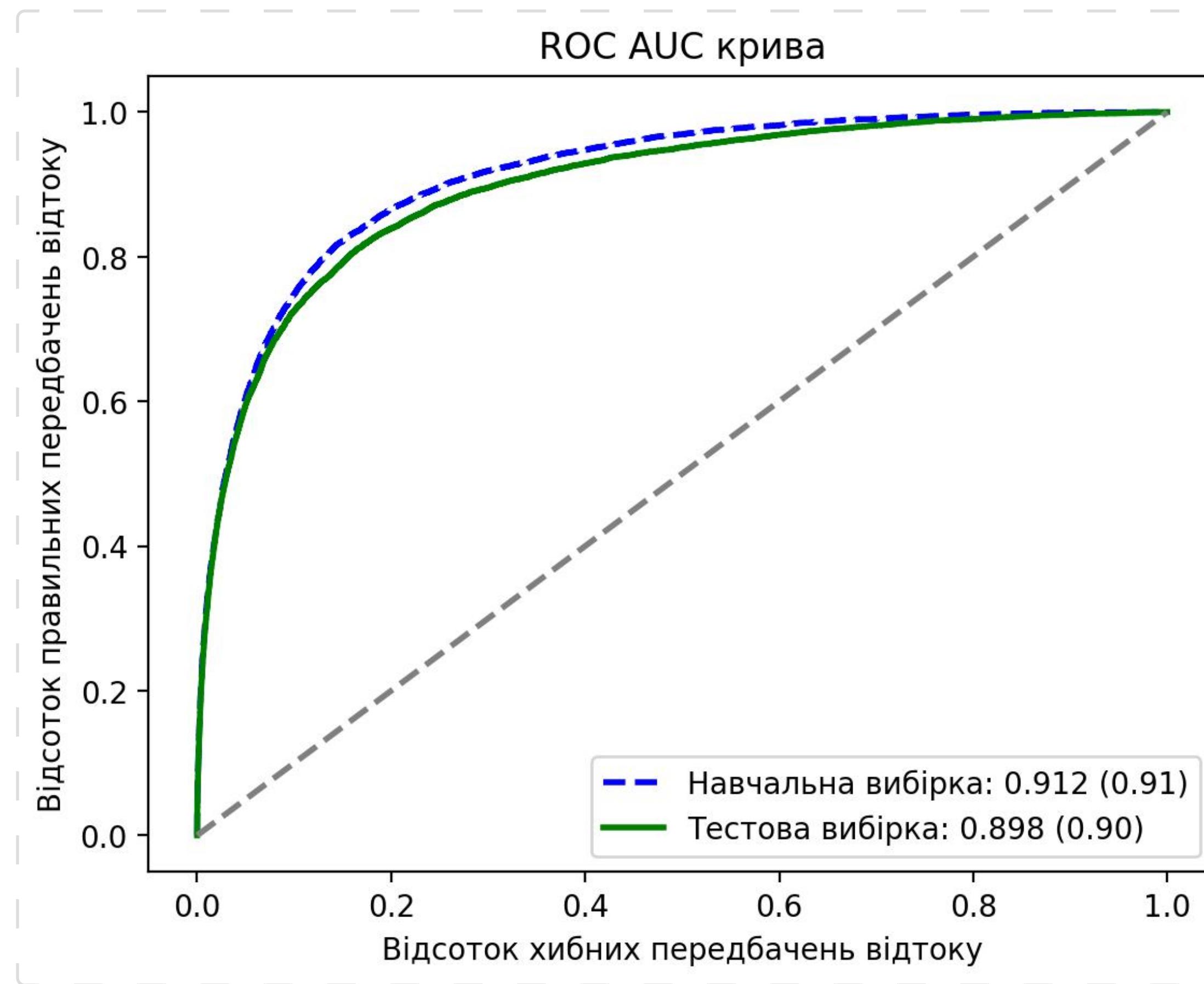


0.6 * np.mean(auc_scores)
+ 0.4 * np.mean(fbeta_scores)
- (10 * np.std(auc_scores))
- (10 * np.std(fbeta_scores))



Фінальні метрики

Результати



Тренувальна

0.912

ROC AUC

0.560

F beta=0.5 score

0.909

Поріг класифікації

0.95

Accuracy (акуратність)

Тестова

0.898

ROC AUC

0.548

F beta=0.5 score

0.895

Поріг класифікації

0.95

Accuracy (акуратність)

>1%

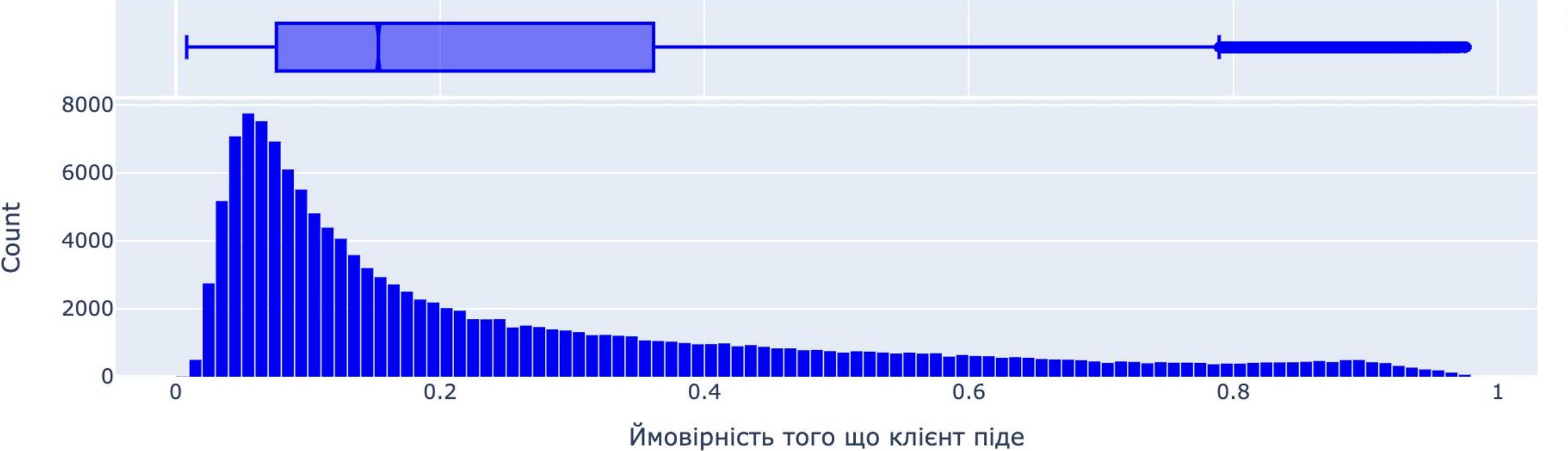
>1%

>1%

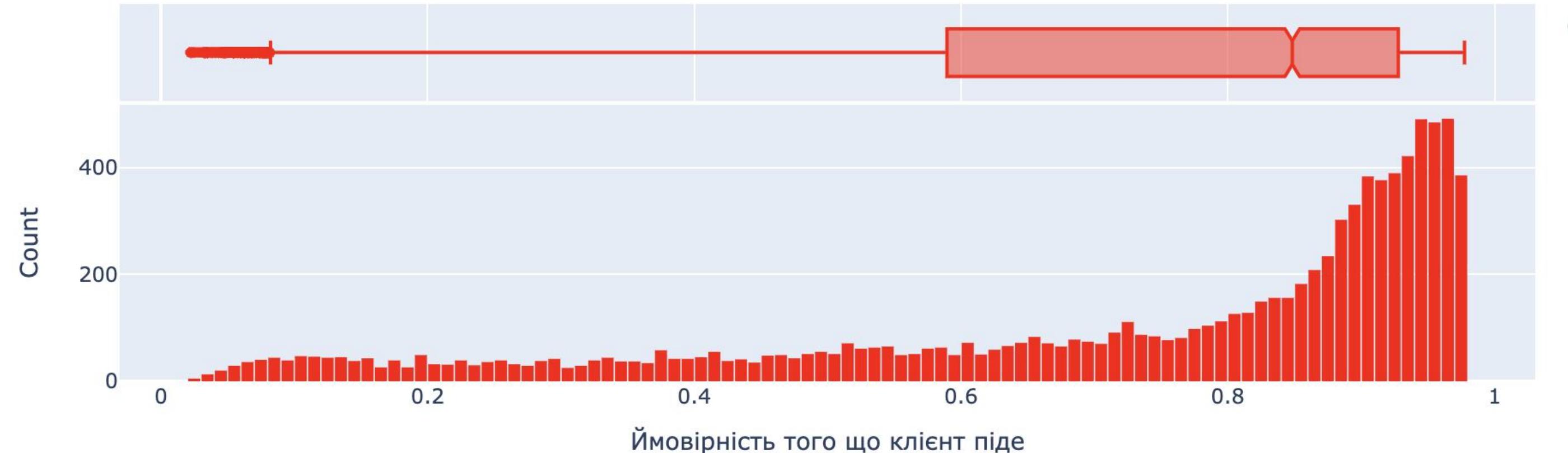
Розподіл ймовірностей

Модель досить добре розрізняє клієнтів, які залишаються, від тих, які підуть, за ймовірністю відтоку

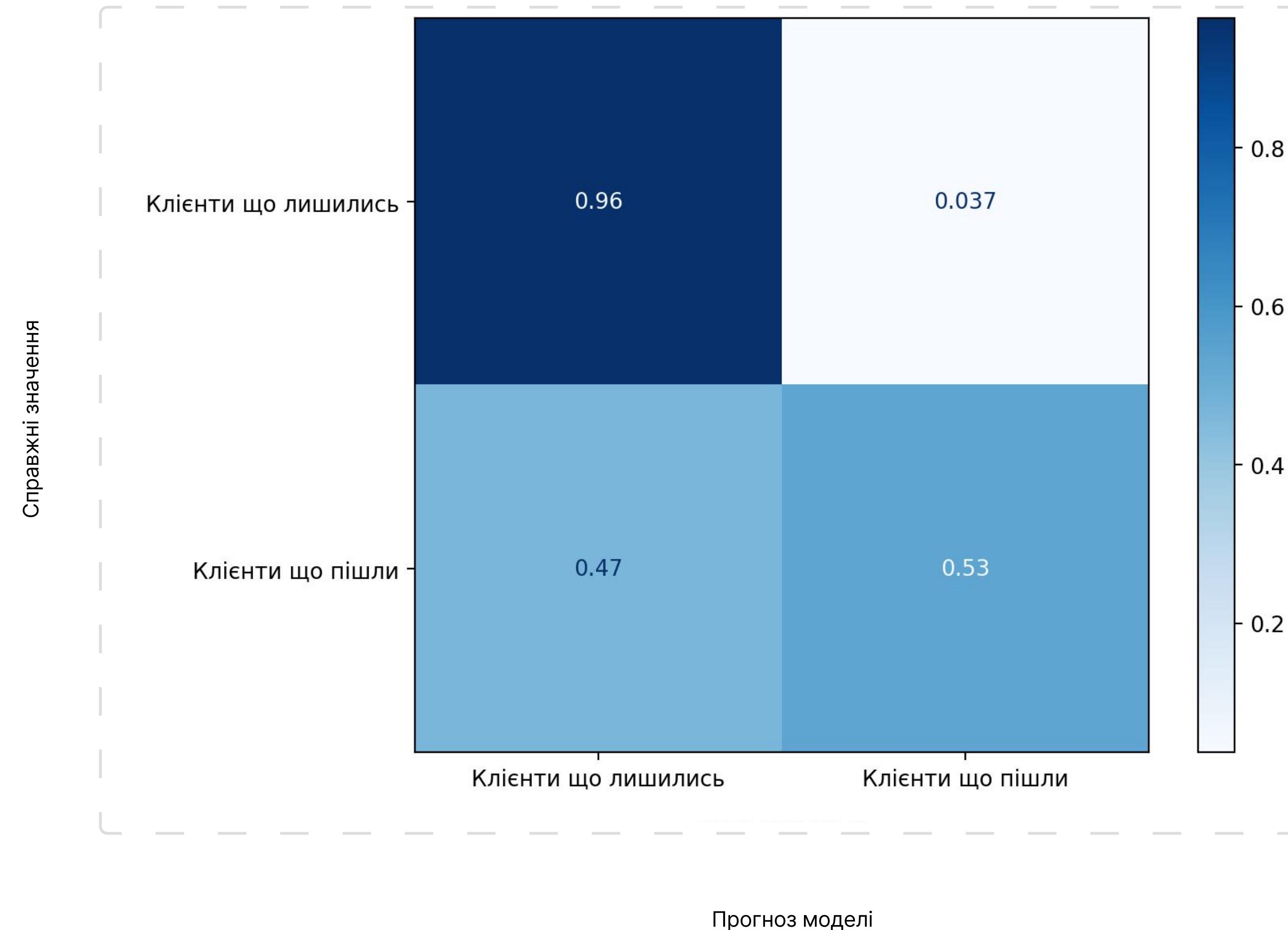
Розподіл ймовірностей для класу 'Клієнти що залишились'



Розподіл ймовірностей для класу 'Клієнти що пішли'



Матриця помилок



Тестова вибірка

0.94

Accuracy (загальна точність)

0.49 / 0.53

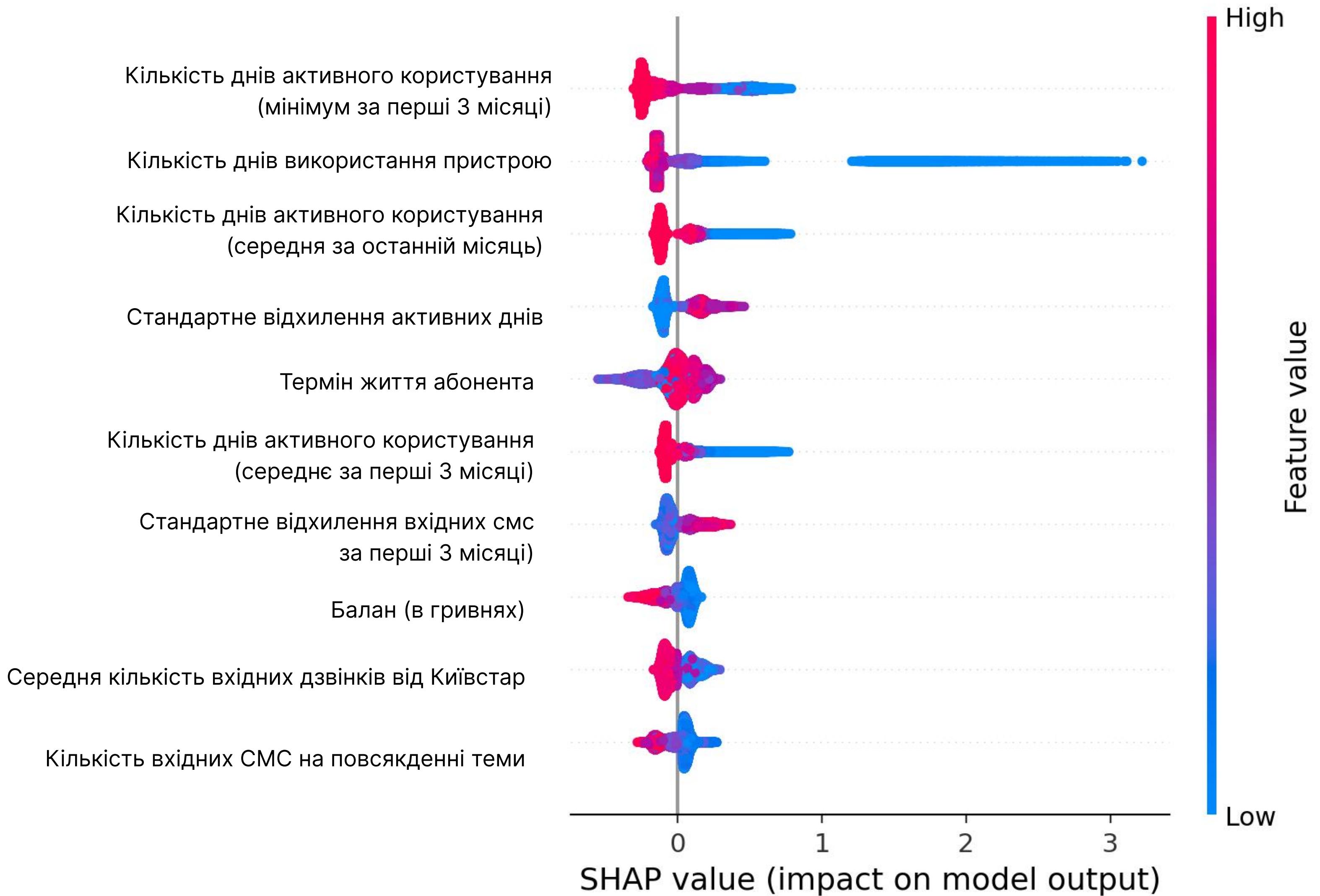
Precision / Recall для класу 1 (відтік)

0.97 / 0.96

Precision / Recall для класу 0 (не відтік)

Поріг 0.83 (максимум в F1 score)

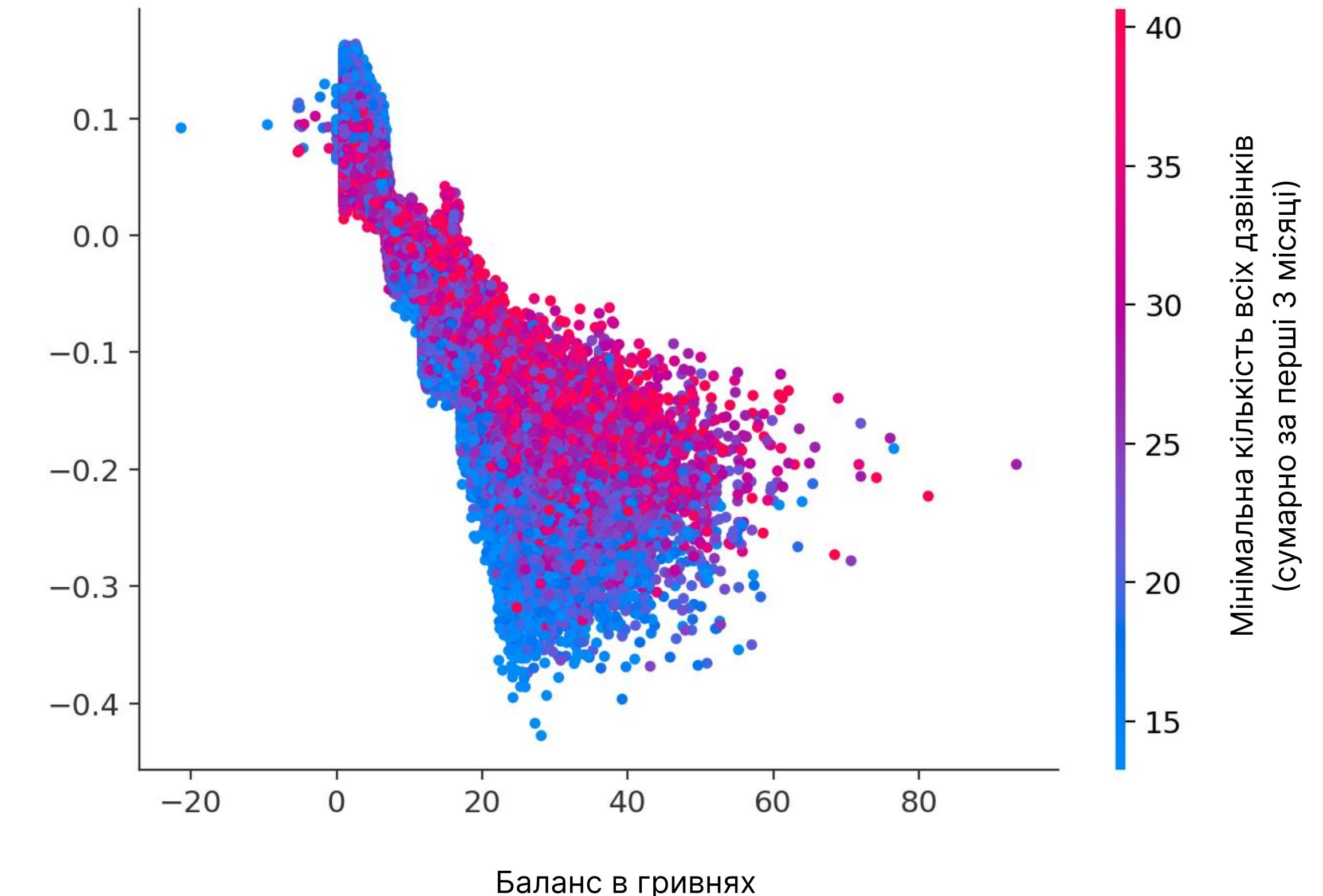
Вплив характеристик



Вплив характеристик

Зі зменшенням балансу ймовірність відтоку абонента збільшується

Можна реагувати на певний поріг балансу на рахунку і пропонувати абоненту схильному до відтоку певні пропозиції

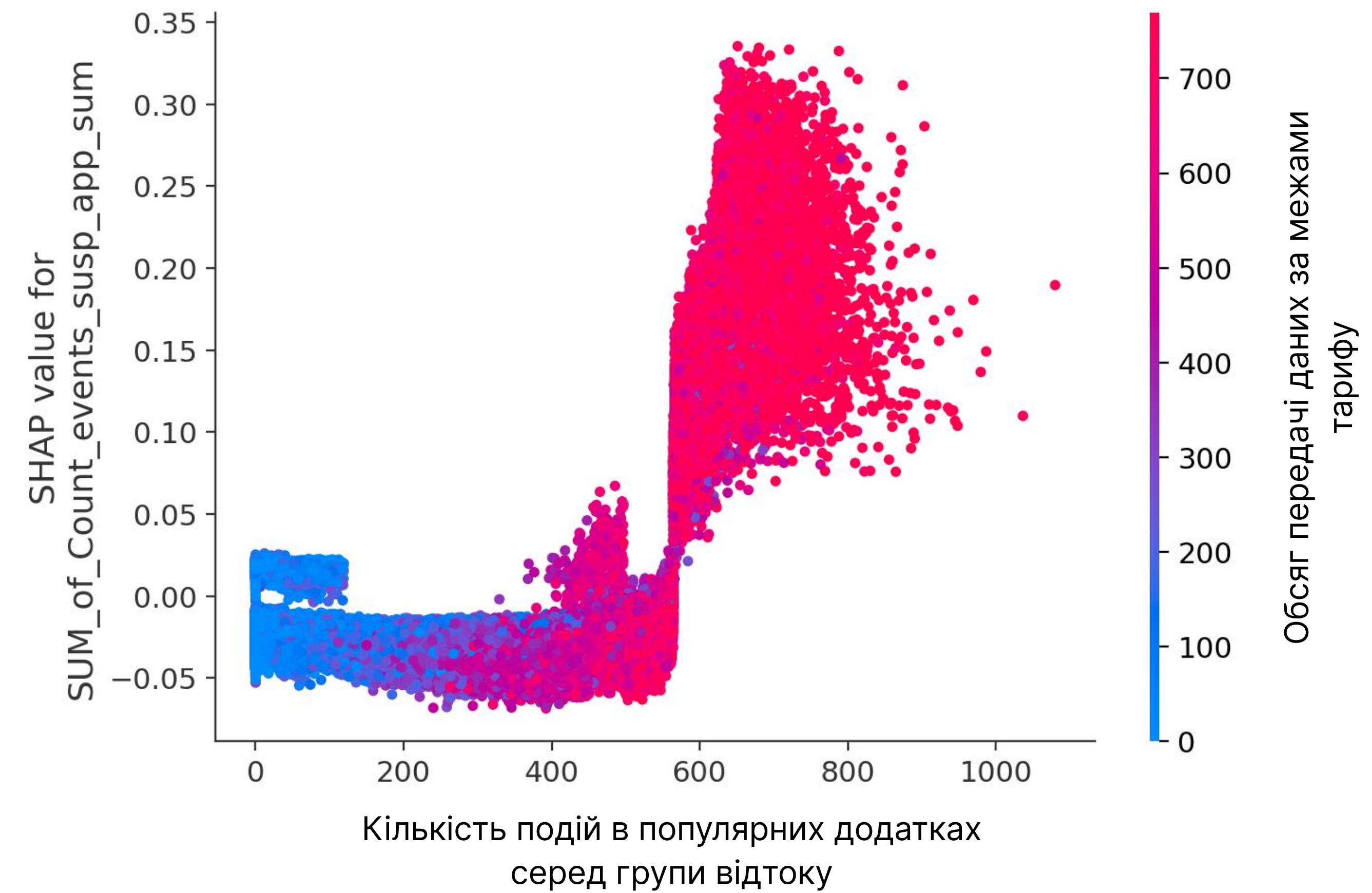


Вплив характеристик

Кількість подій в популярних додатках серед групи відтоку
до обсягу передачі даних за межами тарифу

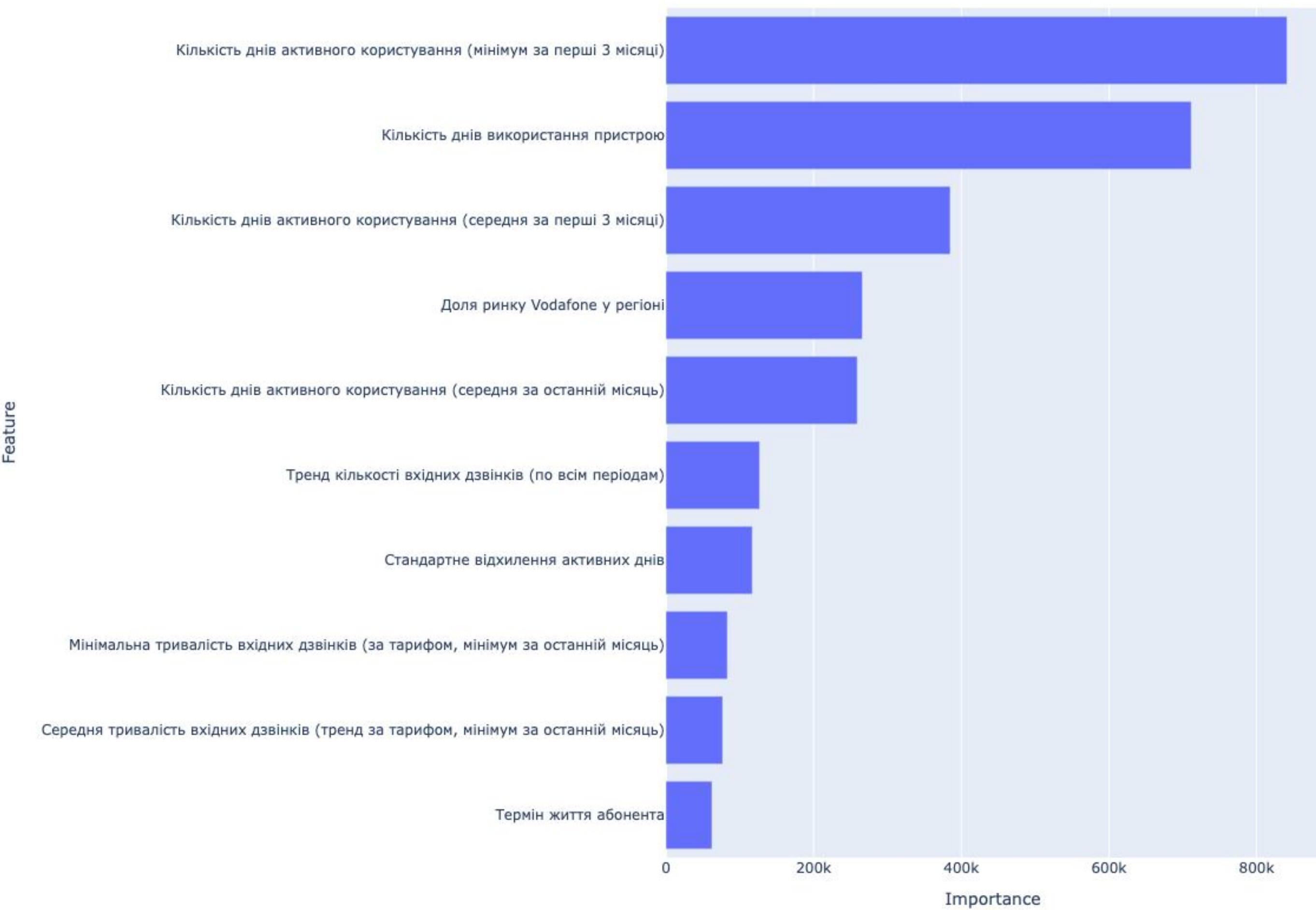
Високе використання певних додатків
підвищує ймовірність того що клієнт може
піти у відтік

Можна висунути припущення, що певних
активних користувачів може не
задовольняти їх поточний тариф, що, в
свою чергу, може підвищувати
ймовірність відтоку



Найважливіші характеристики

Важливість характеристик



Економіка

Результати за ймовірністю значень моделі

Rate	Клієнти з групи, що залишились	Клієнти з групи відтоку	Покриття абонентів з групи відтоку	Покриття абонентів, що залишились	Точність (%)	Lift	Витрати
5	3270	4230	45%	2%	56%	9	166 750,00 грн.
10	9000	6010	64%	6%	40%	6	159 800,00 грн.
15	15530	6970	74%	11%	31%	5	173 250,00 грн.
20	22430	7570	80%	16%	25%	4	195 750,00 грн.
25	29550	7950	85%	21%	21%	3	223 750,00 грн.
30	36740	8260	88%	26%	18%	3	253 500,00 грн.
35	44030	8470	90%	31%	16%	3	285 750,00 грн.
40	51360	8640	92%	37%	14%	2	319 000,00 грн.
45	58730	8770	93%	42%	13%	2	353 250,00 грн.
50	66110	8890	95%	47%	12%	2	387 750,00 грн.
55	73520	8980	96%	52%	11%	2	423 000,00 грн.
60	80930	9070	96%	58%	10%	2	458 250,00 грн.
65	88350	9150	97%	63%	9%	1	493 750,00 грн.
70	95790	9210	98%	68%	9%	1	529 750,00 грн.
75	103240	9260	98%	73%	8%	1	566 000,00 грн.
80	110700	9300	99%	79%	8%	1	602 500,00 грн.
85	118160	9340	99%	84%	7%	1	639 000,00 грн.
90	125630	9370	100%	89%	7%	1	675 750,00 грн.
95	133110	9390	100%	95%	7%	1	712 750,00 грн.
100	140600	9400	100%	100%	6%	1	750 000,00 грн.

64%

зайдених клієнтів з відтоку
при **6%** хибних рішень з групи, що
лишилась

x6

разів краще ніж
випадкове передбачення

Економіка

Вхідні дані

140 000

абоненти, що лишилися

10 000

абоненти, що пішли

5 грн

вартість утримання

25 грн

вартість залучення

Випадковість

75 000

абонентів потрібно утримати

$75\ 000 * 5 \text{ грн} = 375\ 000 \text{ грн}$

5 000

абонентів потрібно залучити

$5\ 000 * 25 \text{ грн} = 125\ 000 \text{ грн}$

525 000 грн

Нічого

10 000

абонентів потрібно залучити

$10\ 000 * 25 \text{ грн} = 250\ 000 \text{ грн}$

250 000 грн

-34%

Модель

6 400 + 8 400

абонентів потрібно утримати

$8\ 100 * 5 \text{ грн} = 74\ 000 \text{ грн}$

3 600

потрібно буде залучити

$3\ 600 * 25 \text{ грн} = 90\ 000 \text{ грн}$

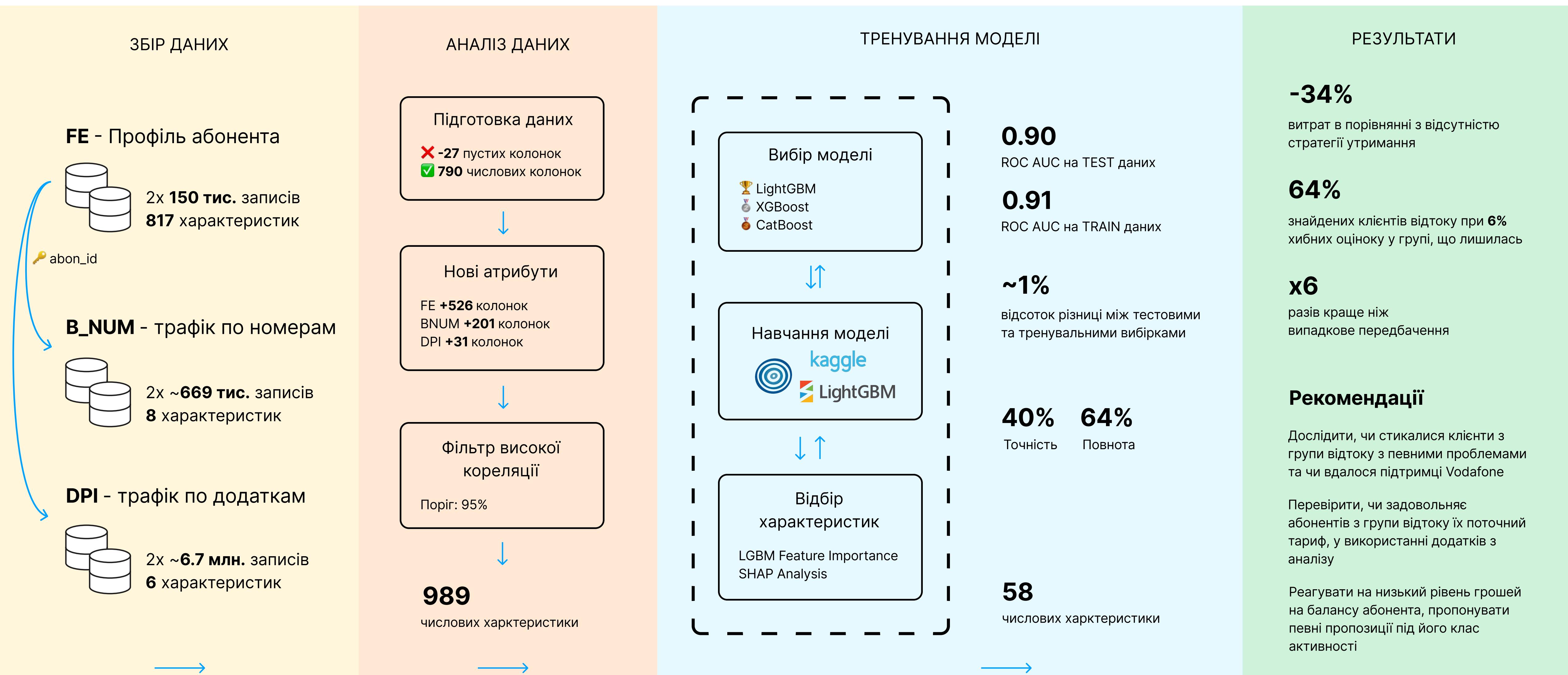
6 400 - модель знайшла

8 400 - модель помилилась

3 600 - модель не знайшла

164 000 грн

Підсумки



Дякую за увагу



Безрукавий Ярослав

<https://www.linkedin.com/in/bezrukavyi>

<https://github.com/bezrukavyi>

yaroslav.bezrukavyi@gmail.com