

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА**

Факультет комп'ютерних наук та кібернетики
Кафедра теорії та технології програмування

Кваліфікаційна робота
на здобуття ступеня бакалавра
за спеціальністю 122 «Комп'ютерні науки»

на тему:

**ВИЯВЛЕННЯ ШАХРАЙСТВА У СФЕРІ ТАКСІ
МЕТОДАМИ АНАЛІЗУ ДАНИХ**

Виконала студентка 4-го курсу
Зверева Тетяна Данилівна _____

Науковий керівник:
доцент, кандидат фіз.-мат. наук
Панченко Тарас Володимирович _____

Засвідчую, що в цій роботі немає запозичень
з праць інших авторів без відповідних
посилань

Студентка _____

Роботу розглянуто й допущено до захисту
на засіданні кафедри теорії та
технології програмування
« ____ » _____ 2021 р.,

протокол № ____
Завідувач кафедри
М. С. Нікітченко _____

РЕФЕРАТ

Обсяг роботи 39 сторінок, 6 ілюстрацій, 21 таблиця та 6 джерел.

АНАЛІЗ ДАНИХ, ТАКСОМОТОРНА ДІЯЛЬНІСТЬ, ШАХРАЙСТВО.

Об'єктом дослідження є дані про транзакції та виконані замовлення водіїв таксопарку для різних міст України.

Предметом дослідження є критерії завчасного виявлення шахрайських транзакцій.

Метою роботи є проведення порівняльного аналізу транзакцій водіїв в межах одного таксопарку для різних міст, ідентифікація існуючих шахрайських платежів, розробка критеріїв, що найкраще виявляють шахрайство.

Методи розроблення: методи попередньої обробки даних, аналізу даних та побудови графіків.

Інструменти розроблення: для аналізу даних була використана мова Python. Середовище розробки – Jupyter Notebook.

Результати роботи: проаналізовано транзакції та замовлення водіїв в межах одного таксопарку для різних міст. Ідентифіковано існуючі шахрайські транзакції водіїв у системі. Розроблено критерії для завчасного виявлення шахрайства. Виявлено відсоток заблокованих шахрайських транзакцій при об'єднанні критеріїв блокування у розглянутих містах.

Сфера застосування: дані критерії можуть використовуватися різними таксопарками для попередження та блокування шахрайських транзакцій.

ЗМІСТ

РЕФЕРАТ	2
ЗМІСТ	3
СКРОЧЕННЯ ТА УМОВНІ ПОЗНАЧЕННЯ	4
ВСТУП	5
РОЗДІЛ 1. АНАЛІЗ ТАКСОМОТОРНОЇ ДІЯЛЬНОСТІ	7
1.1 Історія розвитку таксі	7
1.2 Принцип роботи агрегаторів таксі	8
1.3 Основні види платежів	10
1.4 Передумови виникнення шахрайства	11
1.5 Бібліотеки для обробки даних	12
РОЗДІЛ 2. ВИЯВЛЕННЯ КРИТЕРІЇВ БЛОКУВАННЯ	13
2.1 Опис даних	13
2.2 Попередня обробка та об'єднання даних	15
2.3 Аналіз аномалій	17
2.4. Об'єднання критеріїв блокування.....	30
ВИСНОВКИ	37
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ	38

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАЧЕННЯ

Фрод - шахрайство, мета якого зазвичай полягає в тому, щоб привласнити грошові кошти жертви. Для цього часто використовуються несанкціоновані дії і неправомірне користування ресурсами. У сфері таксі передбачає придбання права на чуже майно шляхом введення, видалення, модифікації інформації або іншого втручання в роботу засобів обробки або передачі даних.

ВСТУП

Оцінка сучасного стану об'єкта дослідження. Життя міста сьогодні важко уявити без стабільного функціонування транспортного комплексу. Комфортне сучасне життя вимагає високого рівня мобільності населення. І одним з головних завдань є доступ до основних пунктів призначення пасажирів

Таксі - послуга, яка надійно увійшла в життя сучасної людини. Сфера таксомоторної діяльності розвивається з кожним днем. Особливо актуальним це є в період пандемії, коли транспорт не працює, не залишається іншої альтернативи окрім таксі або власного транспорту, щоб дістатися до бажаного місця призначення.

Але не всі знають як саме влаштована робота всередині агрегаторів таксі, та з якими проблемами можуть зіткнутися не лише пасажирів, а й таксопарки при роботі з водіями.

Актуальність роботи та підстави для її виконання. Шахрайство у сфері таксі доволі розповсюджене явище. Деякі водії обманюють пасажирів з метою більше заробити, а деякі намагаються обдурити таксопарк, щоб отримати ще більше коштів. Тому необхідно мати певні інструменти для запобігання розповсюдження цієї проблеми у світі.

Мета й завдання роботи. Метою курсової роботи є аналіз виконаних замовлень та здійснених транзакцій в межах одного таксопарку для різних міст України, а також розробка критеріїв, за допомогою яких можливо завчасно виявити підозрілі на шахрайство транзакції. Для досягнення цієї мети поставлено такі завдання:

- дослідити принцип роботи агрегатору таксі;
- дослідити існуючі зони шахрайства у сфері таксі;
- провести аналіз транзакцій за виконаними замовленнями у таксопарку;
- розробити критерії для виявлення шахрайських транзакцій.

- Перевірити результат застосування критеріїв блокування до існуючих транзакцій і виявити відсоток заблокованого фроду

Об'єкт, методи й засоби дослідження. Об'єктом дослідження є дані про транзакції та виконані замовлення водіїв таксопарку для різних міст України. Предметом дослідження є критерії завчасного виявлення шахрайських транзакцій. Аналіз є необхідним для утворення системи антифроду для завчасного блокування шахрайських транзакцій.

Для виявлення критеріїв, було розглянуто три набори даних, що містять інформацію про нарахування, списання та виконані замовлення водіїв, виконано попередню обробку та об'єднання даних.

Для аналізу використано інтерактивне середовище Jupyter Notebook. Він об'єднує код і його вивід в єдиний документ, що є дуже зручним при розробці.

Можливі сфери застосування. Розроблені критерії можуть застосовуватися у різних таксопарках, що мають схожий принцип роботи, як основа для виявлення шахрайства та прийняття рішень щодо оплати водію.

РОЗДІЛ 1. АНАЛІЗ ТАКСОМОТОРНОЇ ДІЯЛЬНОСТІ

1.1 Історія розвитку таксі

Виникнення таксі датується 1605 роком, коли в Парижі та Лондоні працювали прокатні екіпажі, що їздили на конях.

Через два століття Джозеф Хансом спроектував та запатентував швидку та легку кабіну Hansom. Для цих кабін потрібні були лише два колеса, тобто їх міг тягнути один кінь і пересуватися в пробках по всьому Лондону. Поїздки стали швидшими та дешевшими, і в результаті їх популярність зросла. Потім кабіни Hansom поширилися по Європі, з'явившись у Парижі, Берліні та Санкт-Петербурзі.

Тоді таксі рухалось набагато швидшими темпами. Електричні кабіни, що працюють від акумуляторів, вперше стали доступними наприкінці 1800-х років, і Готліб Даймлер винайшов першу у світі таксі-кабіну на бензині в 1897 році [\[1\]](#).

До 1907 року ці бензинові та оснащені таксометром кабіни працювали в Нью-Йорку. Того ж року Гаррі Н. Аллен заснував свою компанію New York Taxicab Company, представивши кабіни на бензині з таксометрами, які стягуватимуть плату з клієнтів на основі пробігу.

В Києві перший автомобіль-таксомотор з'явився в 1913 році. Вартість поїздки становила 1 рубль 20 копійок за 10 км або за одну годину їзди. Оплата здійснювалася за показаннями таксометра.

На початку 2000-х років українські міста почали заповнюватися великою кількістю служб таксі. Кожна така служба самостійно набирала водіїв, виставляла свої ціни, а також займалася усіма бізнес-процесами. Але для замовника це не приносило зручності, оскільки виникала необхідність звертатися до великої кількості служб, щоб знайти найближче вільне авто [\[2\]](#).

На сьогодні, служби таксі мають повну свободу в у виборі цін, дизайну, сервісу, проте незмінною залишається основна місія служб таксі - виконувати свою роботу швидко і з комфортом для пасажера.

Стрімке зростання впливу інтернету на життя людей і нові програмні продукти, розробка нових мобільних застосунків, дали відмінну основу для розвитку агрегаторів таксі.

1.2 Принцип роботи агрегаторів таксі

Агрегатор таксі – це компанія або сервіс, що виступає мостом між пасажиром, що розміщує своє замовлення, та водієм який може його виконати. Найчастіше цей процес відбувається за допомогою мобільного додатку [3]. Офіційно агрегатори таксі перевезеннями не займаються, а виконують роль посередника між водієм і пасажиром таксі. Бурхливий розвиток агрегаторів таксі почався з ростом популярності мобільних застосунків і програмного забезпечення.

На сьогодні найпопулярнішими агрегаторами таксі є Uber, Uklon, Bolt.

Агрегатори таксі працюють за наступними принципами:

- Замовлення утворюються через мобільний додаток. Таким чином пасажир створює нове замовлення, а водій його приймає.
- Агрегатори працюють з офіційно зареєстрованими таксопарками або партнерами, а ті своєю чергою підключають водіїв для роботи.
- Виступаючи посередником, агрегатори й таксопарки беруть свій відсоток від замовлення. Наприклад, від кожної виконаної водієм поїздки агрегатор отримує близько 20%, а таксопарк до 5-10% від вартості замовлення.

Чому ця система лише набирає популярність?

- Через те, що агрегатори таксі мають певні вимоги до автомобіля і водія, це забезпечує пасажиру більшу безпеку і комфортні умови поїздки.
- Пасажиру заздалегідь відома вартість поїздки
- Можна обрати комфортний клас автомобіля: від економа до бізнес-класу, з дитячим кріслом або мінівен
- Швидка подача авто. Оскільки до агрегатора під'єднано багато машин, велика ймовірність що одна з них знаходиться зовсім поруч, і тому не доводиться довго чекати на таксі. Зазвичай машина прибуває в інтервалі 10 хвилин.

Як нараховуються виплати водіям?

Агрегатор таксі щодня нараховує заробіток в партнерські таксопарки, які потім розплачуються з водіями. Моментально, раз в тиждень, кілька разів на місяць або на вимогу – у кожному таксопарку по-різному.

На картинці нижче проілюстрована схема роботи та оплати замовлень. Тобто, пасажир створює замовлення у програмі, і таким чином зв'язується з водієм. Після оплати замовлення, якщо це безготівковий розрахунок, гроші переводяться на рахунок агрегатора таксі. Далі з суми замовлення знімається певна комісія і гроші переводяться на рахунок таксопарку. Після чого, таксопарк-партнер додатково знімає свою комісію і виплачує залишок водію.

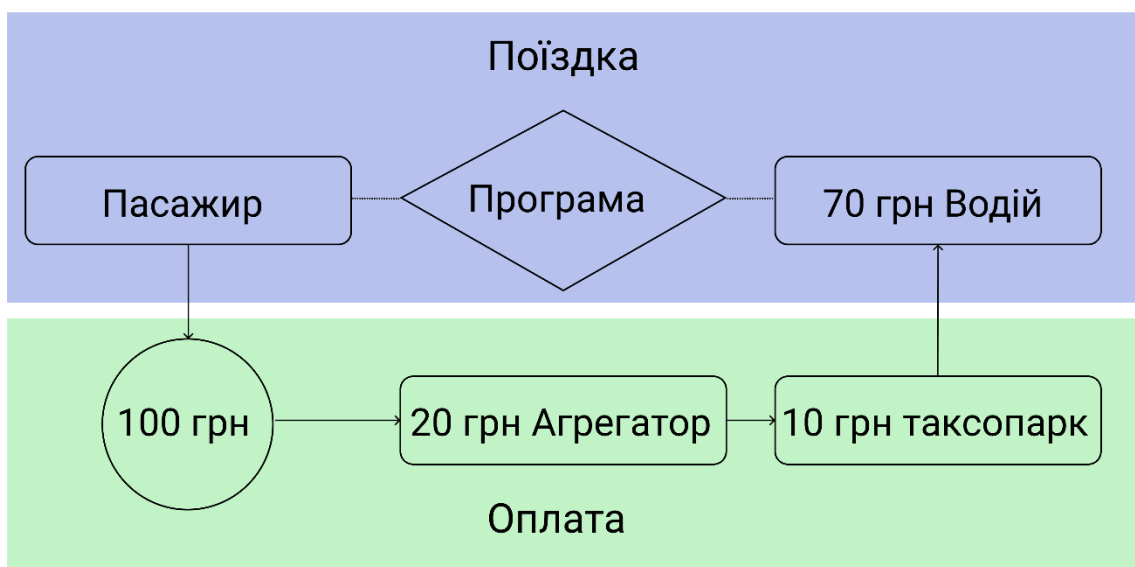


Рисунок 1 – Принцип роботи агрегаторів

Для таксопарків у яких існує система моментальних виплат на рахунок водія, це створює певні «дірки» у безпеці. Адже, партнери можуть виплачувати заробіток водію раніше, ніж самі отримують ці кошти від агрегатора.

1.3 Основні види платежів

Існують три різновиди основних платіжок, які може отримати водій, виконуючи замовлення:

- Фіксована плата за замовлення
- Бонусні виплати
- Чайові

Бонуси використовуються для заохочення водіїв виконувати короткі, незручні та недорогі поїздки. Бонуси складають додатковий прибуток отриманий водієм таксі при наданні онлайн послуги з перевезення пасажирів. Тож водій може отримати більший ніж передбачений заробіток. У цьому випадку водій таксі більше схильний надавати послугу перевезення.

Вони є однією з можливих зон для шахрайства.

Бонус виплачується, якщо:

- водій виконує замовлення по тарифу «Економ»
- у водія гарний показник активності
- поїздка триває понад хвилину
- замовлень потрапило під час, в який діють бонусні виплати

Види бонусів:

- Гарантія мінімальної вартості поїздки

Це бонус, який гарантує мінімальну вартість замовлення. Навіть якщо пасажир їде на сусідню вулицю, вартість замовлення для водія компенсується.

- Бонус за кожну поїздку

З будь-якого успішно виконаного замовлення водій отримує додаткову оплату.

- Щоденний бонус

Бонус нараховується за добу, протягом яких водій виконав певне число поїздок. Як правило, такі акції діють недовго, але дозволяють отримати відчутну прибавку до заробітку.

- Персональний бонус

У цього бонусу свої умови для кожного водія, а механізм дії однаковий для всіх: водій виконує певне число поїздок за зазначений час - і отримує надбавку до заробітку [\[4\]](#).

1.4 Передумови виникнення шахрайства

Для таксопарків, що мають моментальні виплати, підвищується вірогідність шахрайства зі сторони водіїв. Річ у тому, що вони виплачують гроші водіям відразу після виконання замовлення - при тому, що гроші від агрегатора приходять з затримкою від трьох днів до місяця. Будь-яка спроба обману з боку водія – це їх ризик не отримати гроші від провайдера [\[5\]](#).

Багато водіїв працюють як незалежні підрядники, вони можуть реєструвати фальшиві профілі з фальшивими документами. Через те, що це погано

відстежується, зловити шахрая стає майже нереально і водії користуються цим не боячись покарання.

Саме тому так необхідно мати якийсь інструмент, щоб швидше провайдерів дізнаватися про те, що водій намагається обдурити та заробити нечесним шляхом.

1.5 Бібліотеки для обробки даних

Pandas - це пакет, який зазвичай використовується для аналізу даних. Це спрощує завантаження даних із зовнішніх джерел, таких як текстові файли та бази даних, а також забезпечує способи аналізу та керування даними після їх завантаження на комп'ютер. Функції, що надаються у pandas, автоматизують та спрощують багато загальних завдань, для реалізації яких необхідно було б створювати багаторядковий код базовою мовою Python [6].

Популярність pandas зростає з кожним днем, бібліотека використовується в різних дисциплінах - від екологічних та кліматичних наук до соціальних наук, лінгвістики, біології, а також для ряду застосувань у галузі, таких як аналіз даних, фінансова торгівля та багато інших.

Найкраще pandas підходить для структурованих, маркованих даних, іншими словами, табличних даних, які мають заголовки, пов'язані з кожним стовпцем даних.

Matplotlib - це пакет Python, який використовується для побудови графічних даних та візуалізації. Це корисне доповнення до Pandas, і дуже багата на функції бібліотека, яка може створювати велику кількість графіків, діаграм, карт та інших візуалізацій.

РОЗДІЛ 2. ВИЯВЛЕННЯ КРИТЕРІЇВ БЛОКУВАННЯ

2.1 Опис даних

Для аналізу використано дані замовлень водіїв в межах одного таксопарку.

Ці дані охоплюють наступні таблиці:

- Orders – список усіх замовлень;
- Incomes – список нарахованих транзакцій;
- Outcomes – список списаних з рахунку транзакцій;

Розберемо кожну таблицю більш детально. В табл. 1 наведено детальний опис кожної таблиці і її атрибутів.

Таблиця 1. Опис таблиць даних

Таблиця	Атрибут	Тип	Опис
Orders	Id	String	Id замовлення
	Taximeter_id	Int	Id таксометру
	Driver_id	String	Id водія
	Number	Int	Номер замовлення
	Payment_method_id	Int	Метод оплати
	Date	Datetime	Дата замовлення
	Amount	Float	Сума замовлення
	Park_comission	Float	Комісія парку-партнеру

	Agregator_Comission	Float	Комісія агрегатору
	Rule_id	Int	Умови за якими працює водій
	City_id	Int	ID міста у якому працює водій
	Is_Canceled	Bool	Чи відмінено замовлення
Incomes	Income_id	String	Id нарахування
	Income_taximeter_id	Int	Id таксометру
	Income_number	Int	Номер нарахування
	Income_date	Datetime	Дата замовлення
	Income_payment_type_id	Int	Тип оплати (фіксована оплата, бонус і т.д)
	Income_amount	Float	Сума нарахування
Outcomes	Outcome_id	String	Id нарахування
	Outcome_taximeter_id	Int	Id таксометру
	Outcome_number	Int	Номер нарахування
	Outcome_date	Datetime	Дата замовлення
	Outcome_payment_type_id	Int	Тип оплати (фіксована оплата, бонус і т.д)
	Outcome_amount	Float	Сума нарахування

2.2 Попередня обробка та об'єднання даних

Для обробки даних необхідно правильно зчитати дані з файлу. Спочатку створимо масив з назвами колонок та визначимо їх типи. Для стовпчиків, що містять інформацію про вартість та комісію замовлення, перевіримо трансформацію в чисельний тип, у іншому випадку, замінимо нулями.

Оскільки дані зберігаються у форматі .csv, зручно використовувати бібліотеку pandas і її метод `read_csv` в аргументи якого ми записуємо усі необхідні параметри й створені масиви з назвами та типами колонок.

Наступним етапом є об'єднання таблиць для зручного аналізу. Оскільки кожна таблиця має спільні елементи, ми можемо зв'язати їх за номером таксометру (`taximeter_id`) та номером замовлення (`number`).

Зробимо наступні перетворення:

- 1) З таблиці `Incomes` утворюємо і додаємо до таблиці `Orders` дві колонки:
 - `Payment_amount` – Фіксовані нарахування за замовленням
 - `Bonus_amount` – Бонусні нарахування за замовленням
- 2) З таблиці `Outcomes` додаємо до таблиці `Orders` дві колонки:
 - `Payment_return` – Списання фіксованої оплати за замовленням
 - `Bonus_return` – Списання бонусів за замовленням
- 3) Створюємо нові колонки:
 - `Is_payment_fraud` – чи є повні списання фіксованої оплати за замовленням
 - `Is_bonus_fraud` – чи є повні списання бонусів за замовленням
 - `Is_fraud` – чи істина хоча б одна з попередніх двох колонок (`Is_payment_fraud` або `Is_bonus_fraud`)
 - `Card_payment` – сума безготівкового платежу

Після обробки з'являється нова таблиця, що містить усі описані дані:

Таблиця 2. Опис об'єднаної таблиці даних Processed_orders

Атрибут	Тип	Опис
Id	String	Id замовлення
Taximeter_id	Int	Id таксометру
Driver_id	String	Id водія
Number	Int	Номер замовлення
Payment_method_id	Int	Метод оплати (готівковий, безготівковий, корпоративний)
Date	Datetime	Дата замовлення
Amount	Float	Сума замовлення
Park_comission	Float	Комісія парку-партнеру
Agregator_Comission	Float	Комісія агрегатору
Rule_id	Int	Умови за якими працює водій
City_id	Int	ID міста у якому працює водій
Is_Canceled	Bool	Замовлення відмінено
Payment_amount	Float	Фіксовані нарахування за замовленням
Bonus_amount	Float	Бонусні нарахування за замовленням

Payment_return	Float	Списання фіксованої оплати за замовленням
Bonus_return	Float	Списання бонусів замовленню
Is_payment_fraud	Bool	Існують повні списання фіксованої оплати за замовленням
Is_bonus_fraud	Bool	Існують повні списання бонусів за замовленням
Is_fraud	Bool	Хоча б одна з колонок Is_bonus_fraud або Is_payment_fraud істина
Card_payment	Float	Сума безготівкового платежу

2.3 Аналіз аномалій

Для того, щоб завчасно виявляти та блокувати шахраїв, необхідно вміти розпізнавати фродові транзакції. З цією метою я обрала декілька критеріїв, щоб перевірити, наскільки вдало вони можуть розпізнавати фрод.

В аналізі брало учать п'ять міст України: Дніпро, Одеса, Київ, Харків та Львів. Для кожного міста розраховувався певний критерій за верхніми та нижніми квантилями.

Результати досліджень вийшли найкращими для кожного критерію при 0,95 верхньому та 0,05 нижньому квантілі.

Верхньому квантилю виставлялися параметри «more», тобто, розглядалися дані, що попали у проміжок, більший заданого значення (наприклад, більше 0,95 квантіля). Результати для 0,05 та 0,95 квантіля наведені нижче у таблицях.

При аналізі розглядалися наступні критерії:

- Відсоток бонусу від замовлення
- Вартість замовлення

- Вартість замовлення з заблокованими бонусами
- Вартість бонусу
- Середня вартість замовлення
- Середня вартість бонусу
- Кількість замовлень за годину
- Вартість замовлень годину
- Щоденний ліміт без врахування бонусів
- Щоденний бонусний ліміт

Для кожного критерію розглянемо декілька з наступних значень:

- відсоток суми шахрайських платежів, що попадають під цей квантиль, від суми усіх шахрайських платежів
- відсоток кількості шахрайських платежів, що попадають під цей квантиль, від кількості усіх шахрайських платежів
- відсоток суми реальних платежів, що попадають під цей квантиль, від суми усіх реальних платежів
- відсоток кількості реальних платежів, що попадають під цей квантиль, від кількості усіх реальних платежів
- відсоток суми шахрайських бонусів, що попадають під цей квантиль, від суми усіх шахрайських бонусів
- відсоток кількості шахрайських бонусів, що попадають під цей квантиль, від кількості усіх шахрайських бонусів
- відсоток суми реальних бонусів, що попадають під цей квантиль, від суми усіх реальних платежів
- відсоток кількості реальних бонусів, що попадають під цей квантиль, від кількості усіх реальних платежів

1) Відсоток бонусу від замовлення:

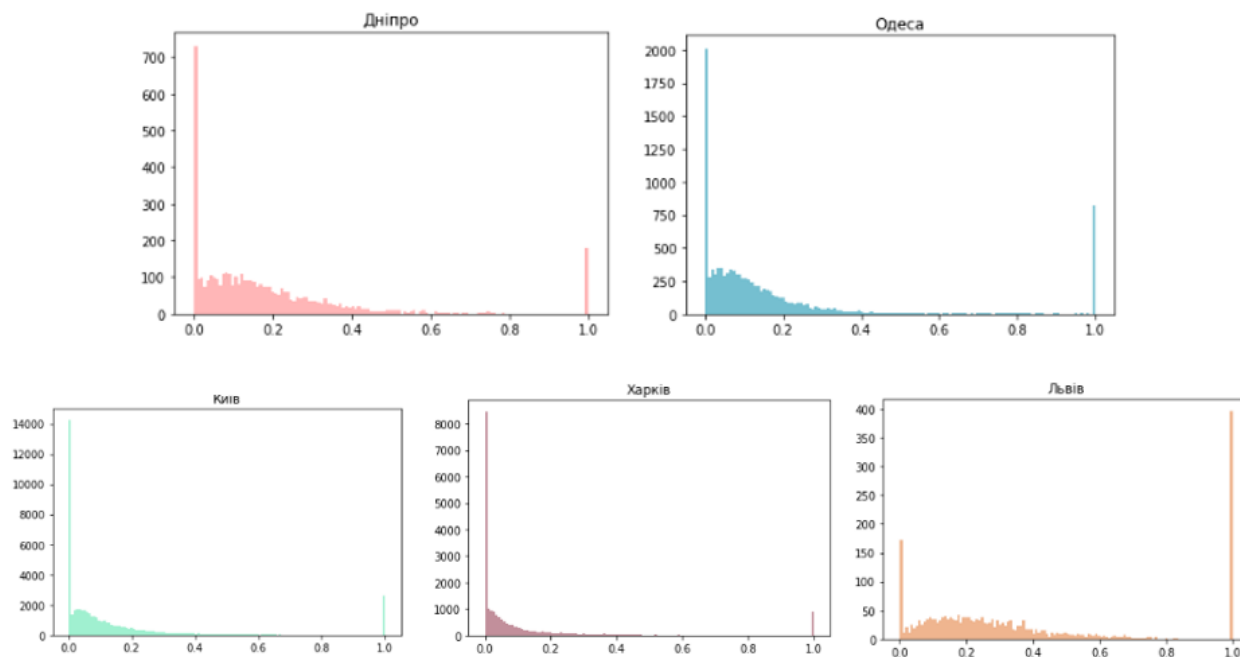


Рисунок 2 – Відсоток бонусу від замовлення

Таблиця 3. Відсоток бонусу від замовлення за 0,05 квантилем

Місто	Шахрайський платіж		Реальний платіж		Шахрайський бонус		Реальний бонус	
	Сума, %	К-ть, %	Сума, %	К-ть, %	Сума, %	К-ть, %	Сума, %	К-ть, %
Дніпро	15,21	10,00	16,53	13,62	0,00	0,00	0,81	2,25
Одеса	28,35	22,86	16,81	14,02	0,00	0,00	0,38	2,11
Київ	37,61	21,82	28,72	21,90	0,52	1,66	0,43	1,74
Харків	50,71	32,20	38,22	31,74	0,56	3,14	0,46	2,28
Львів	11,12	16,67	14,50	11,198	2,37	2,90	2,67	3,77

Таблиця 4. Відсоток бонусу від замовлення за 0,95 квантилем

Місто	Шахрайський платіж		Реальний платіж		Шахрайський бонус		Реальний бонус	
	Сума, %	К-ть, %	Сума, %	К-ть, %	Сума, %	К-ть, %	Сума, %	К-ть, %
Дніпро	4,16	6,67	0,60	1,09	24,81	17,65	4,02	2,88
Одеса	0,00	0,00	0,03	0,05	0,00	0,00	0,29	0,45
Київ	0,00	0,00	0,01	0,02	0,68	0,61	0,50	0,34
Харків	0,12	0,56	0,04	0,09	4,49	4,41	1,68	1,42
Львів	0,00	0,00	0,17	0,22	0,00	0,00	0,91	0,77

2) Вартість замовлення:

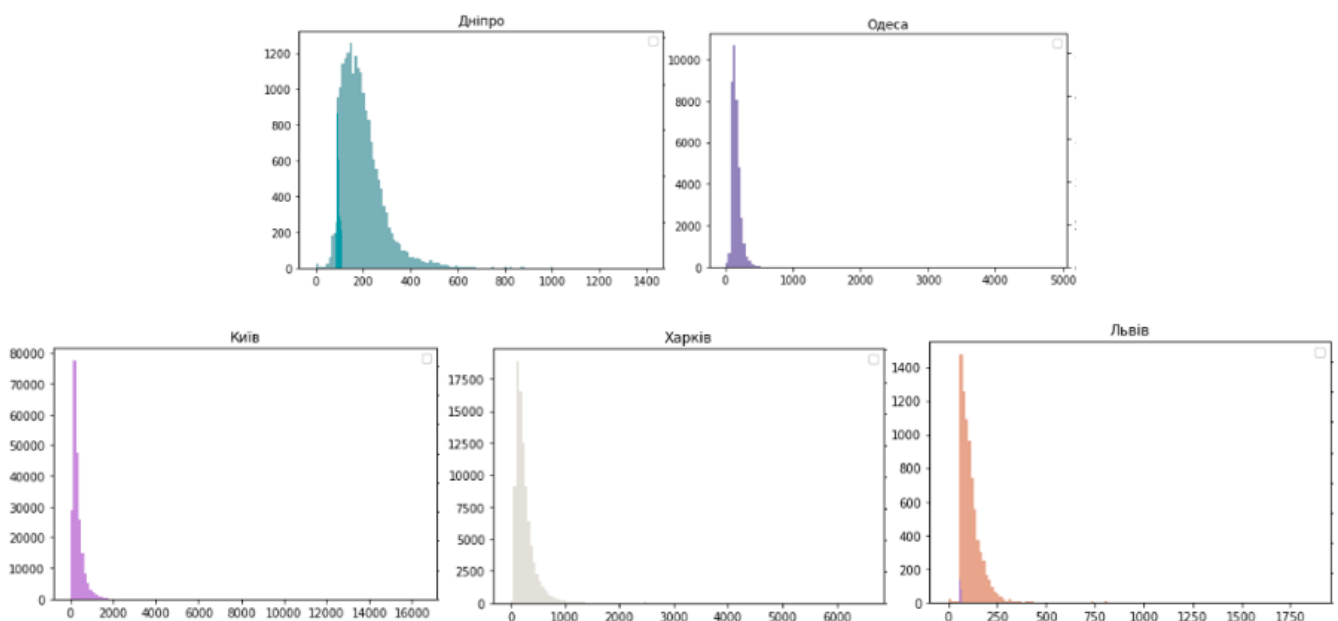


Рисунок 3 – Вартість замовлення

Таблиця 5. Вартість замовлення за 0,05 квантилем

Місто	Шахрайський платіж		Реальний платіж	
	Сума, %	К-ть, %	Сума, %	К-ть, %
Дніпро	8,69	16,67	2,29	5,39
Одеса	5,55	11,43	2,41	5,34
Київ	3,96	17,64	1,24	5,15
Харків	3,45	19,21	1,54	5,21
Львів	3,86	16,67	3,44	6,99

Таблиця 6. Вартість замовлення за 0,95 квантилем

Місто	Шахрайський платіж		Реальний платіж	
	Сума, %	К-ть, %	Сума, %	К-ть, %
Дніпро	13,37	6,67	11,47	5,17
Одеса	14,25	7,14	10,87	5,05
Київ	39,88	8,55	16,78	5,01
Харків	51,63	9,60	16,56	5,03
Львів	20,44	8,33	13,69	5,44

3) Вартість замовлення з заблокованим бонусом:

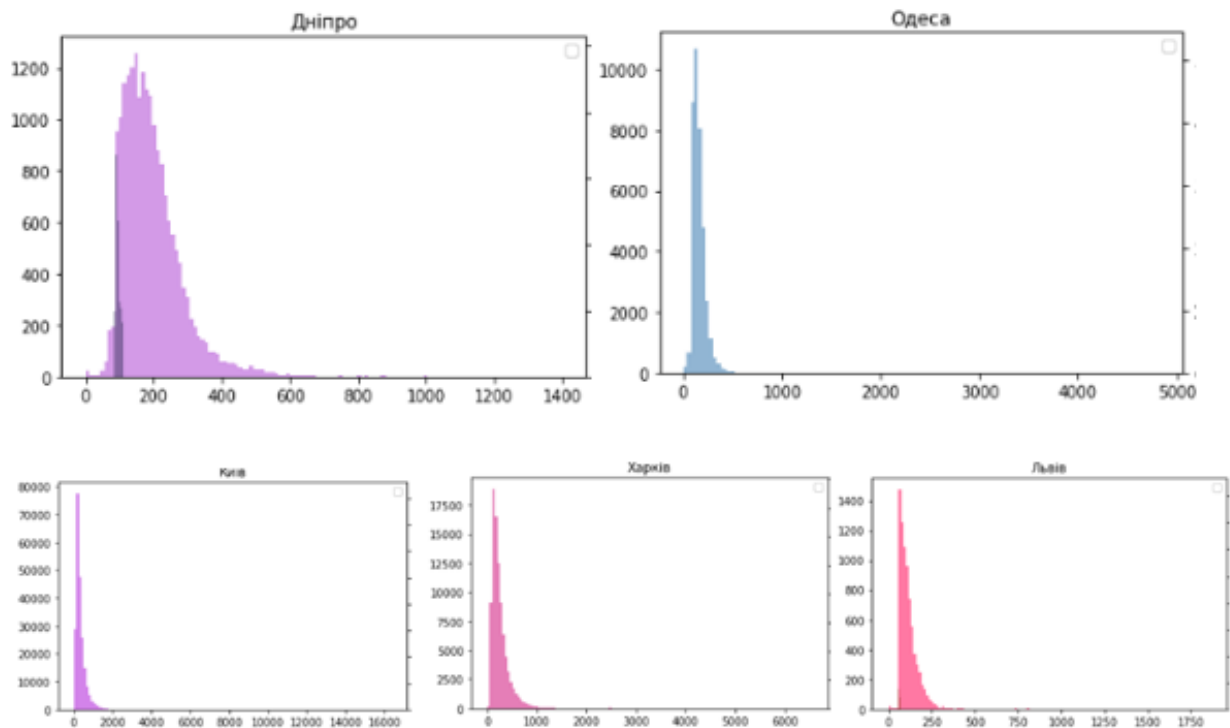


Рисунок 4 – Вартість замовлення з заблокованим бонусом

Таблиця 7. Вартість замовлення з заблокованим бонусом за 0,05 квантилем

Місто	Шахрайський платіж		Реальний платіж	
	Сума, %	К-ть, %	Сума, %	К-ть, %
Дніпро	22,96	13,16	17,16	10,21
Одеса	51,01	22,45	19,49	15,95
Київ	11,14	12,75	12,39	14,30
Харків	13,42	22,63	28,13	21,78
Львів	9,32	7,14	21,86	13,58

Таблиця 8. Вартість замовлення з заблокованим бонусом за 0,95 квантилем

Місто	Шахрайський платіж		Реальний платіж	
	Сума, %	К-ть, %	Сума, %	К-ть, %
Дніпро	0,00	0,00	0,00	0,01
Одеса	0,00	0,00	2,87	0,13
Київ	0,25	0,45	1,11	0,37
Харків	0,20	0,21	0,68	0,29
Львів	0,00	0,00	0,00	0,00

4) Вартість бонусу:

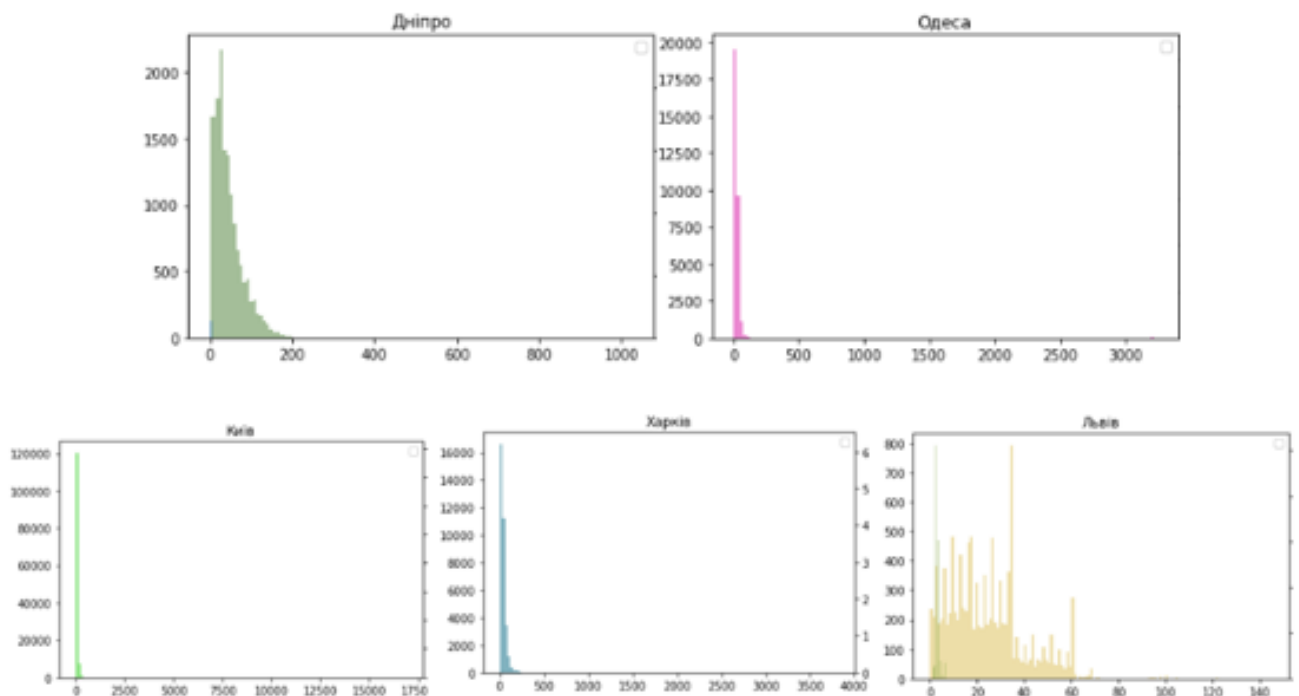


Рисунок 5 – Вартість бонусу

Таблиця 9. Вартість бонусу за 0,05 квантилем

Місто	Шахрайський бонус		Реальний бонус	
	Сума, %	К-ть, %	Сума, %	К-ть, %
Дніпро	0,29	5,88	0,35	5,67
Одеса	0,31	5,66	0,28	6,18
Київ	0,20	3,95	0,25	5,14
Харків	0,21	4,24	0,31	5,39
Львів	0,17	2,90	0,51	6,25

Таблиця 10. Вартість бонусу за 0,95 квантилем

Місто	Шахрайський платіж		Реальний платіж	
	Сума, %	К-ть, %	Сума, %	К-ть, %
Дніпро	26,30	11,76	14,84	5,11
Одеса	19,15	8,96	44,89	4,98
Київ	24,95	6,32	31,43	5,01
Харків	19,55	6,20	20,08	5,09
Львів	17,81	8,70	11,26	4,53

5) Середня вартість замовлення:

Таблиця 11. Середня вартість замовлення за 0,05 квантилем

Місто	Шахрайський платіж		Реальний платіж	
	Сума, %	К-ть, %	Сума, %	К-ть, %
Дніпро	8,03	13,33	2,1	3,81
Одеса	3,25	5,71	1,58	2,69
Київ	3,45	12,55	1,13	3,49
Харків	1,95	7,91	1,11	2,83
Львів	5,59	8,33	2,00	3,29

Таблиця 12. Середня вартість замовлення за 0,95 квантилем

Місто	Шахрайський платіж		Реальний платіж	
	Сума, %	К-ть, %	Сума, %	К-ть, %
Дніпро	11,09	6,67	6,11	3,48
Одеса	3,19	2,86	4,99	2,74
Київ	26,56	5,27	6,78	2,29
Харків	46,23	7,34	7,42	2,59
Львів	0,00	0,00	8,02	4,24

б) Середня вартість бонусу

Таблиця 13. Середня вартість бонусу за 0,05 квантилем

Місто	Шахрайський бонус		Реальний бонус	
	Сума, %	К-ть, %	Сума, %	К-ть, %
Дніпро	17,5	9,41	8,59	3,7
Одеса	12,45	9,91	22,3	3,47
Київ	15,21	6,14	23,66	4,38
Харків	14,16	5,01	14,54	4,17
Львів	21,63	14,49	8,41	4,88

Таблиця 14. Середня вартість бонусу за 0,95 квантилем

Місто	Шахрайський платіж		Реальний платіж	
	Сума, %	К-ть, %	Сума, %	К-ть, %
Дніпро	0,00	0,00	0,63	2,59
Одеса	0,17	0,47	0,26	2,27
Київ	0,17	1,34	0,29	2,31
Харків	0,25	2,29	0,28	2,53
Львів	0,00	0,00	1,62	3,58

7) Кількість замовлень за годину

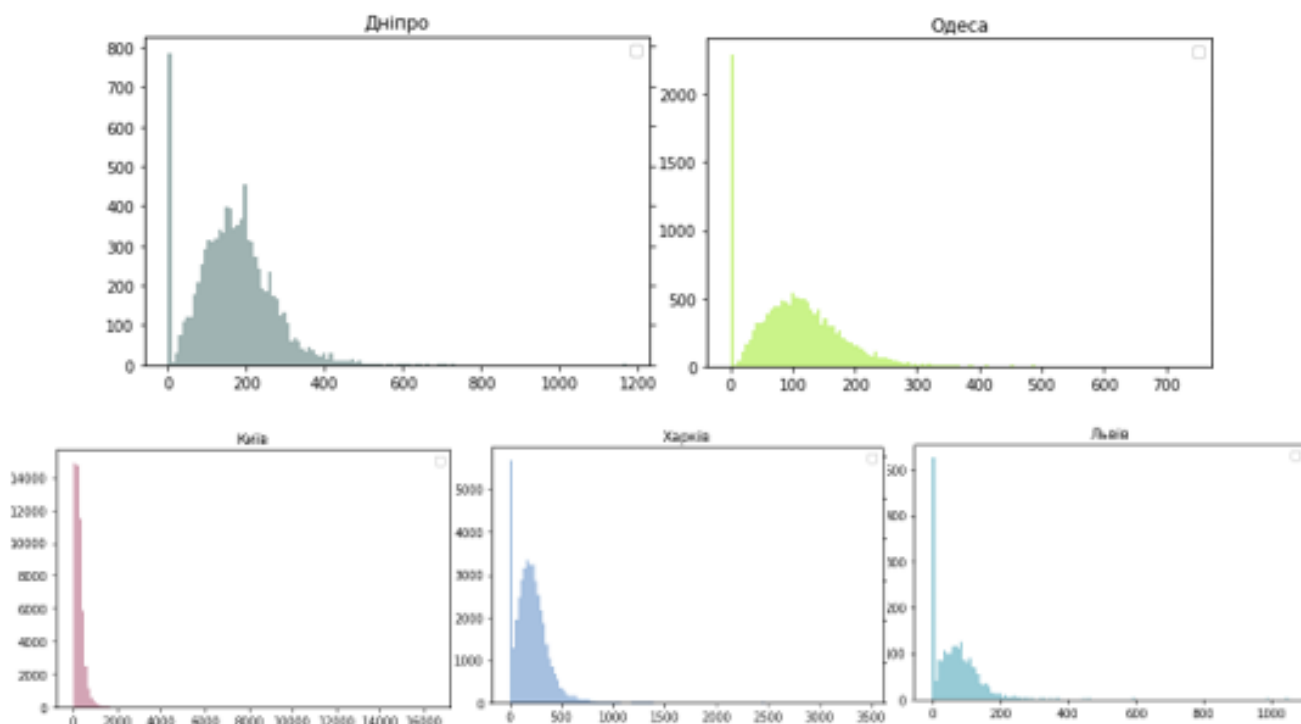


Рисунок 6 – Кількість замовлень за годину

Таблиця 15. Кількість замовлень за годину за 0,05 квантилем

Місто	Шахрайський платіж		Реальний платіж	
	Сума, %	К-ть, %	Сума, %	К-ть, %
Дніпро	5,78	3,33	4,60	3,96
Одеса	6,14	7,14	4,53	4,11
Київ	22,18	11,64	11,16	7,37
Харків	31,39	11,86	9,06	6,52
Львів	0,00	0,00	3,89	3,44

Таблиця 16. Кількість замовлень за годину за 0,95 квантилем

Місто	Шахрайський платіж		Реальний платіж	
	Сума, %	К-ть, %	Сума, %	К-ть, %
Дніпро	0,00	0,00	5,13	6,12
Одеса	7,78	8,57	5,66	6,17
Київ	5,61	14,00	4,19	7,15
Харків	1,67	1,69	4,88	6,78
Львів	0,00	0,00	8,27	8,89

8) Вартість замовлень за годину

Таблиця 17. Вартість замовлень за годину за 0,05 квантилем

Місто	Шахрайський бонус		Реальний бонус	
	Сума, %	К-ть, %	Сума, %	К-ть, %
Дніпро	0,00	0,00	0,13	0,17
Одеса	0,00	0,00	0,00	0,00
Київ	0,00	0,00	0,00	0,00
Харків	0,00	0,00	0,00	0,00
Львів	0,00	0,00	0,00	0,00

Таблиця 18. Вартість замовлень за годину за 0,95 квантилем

Місто	Шахрайський платіж		Реальний платіж	
	Сума, %	К-ть, %	Сума, %	К-ть, %
Дніпро	5,31	3,33	7,92	6,38
Одеса	6,06	2,86	8,49	6,86
Київ	29,71	7,45	9,26	4,39
Харків	47,51	11,86	9,54	4,85
Львів	25,06	16,67	13,81	10,78

9) Щоденний ліміт без врахування бонусів

Таблиця 19. Щоденний ліміт без врахування бонусів за 0,08 квантилем

Місто	Шахрайський платіж		Реальний платіж	
	Сума, %	К-ть, %	Сума, %	К-ть, %
Дніпро	7,13	3,33	16,02	14,62
Одеса	15,79	14,29	15,57	14,61
Київ	28,87	14,00	18,49	13,73
Харків	47,21	15,25	19,39	14,78
Львів	50,39	41,67	21,26	17,98

10) Щоденний бонусний ліміт

Таблиця 20. Щоденний бонусний ліміт за 0,90 квантилем

Місто	Шахрайський платіж		Реальний платіж	
	Сума, %	К-ть, %	Сума, %	К-ть, %
Дніпро	39,40	29,41	12,31	8,64
Одеса	18,83	13,21	42,83	9,18
Київ	21,77	11,81	28,98	9,65
Харків	21,74	14,43	21,70	11,65
Львів	0,00	0,00	9,63	8,09

2.4. Об'єднання критеріїв блокування

Розглянемо об'єднання певних критеріїв і результат який отримаємо.

Для цього візьмемо нашу початкову таблицю, у якій зберігаються дані по усім транзакціям і приєднаємо до неї нові колонки з критеріями– результат, чи є дана транзакція шахрайською за певним критерієм.

Після обробки з'являється нова таблиця, що містить усі описані дані:

Таблиця 21. Опис об'єднаної таблиці даних

Атрибут	Тип	Опис
Id	String	Id замовлення
Taximeter_id	Int	Id таксометру

Driver_id	String	Id водія
Number	Int	Номер замовлення
Payment_method_ID	Int	Метод оплати (готівковий, безготівковий, корпоративний)
Date	Datetime	Дата замовлення
Amount	Float	Сума замовлення
Park_comission	Float	Комісія парку-партнеру
Agregator_comission	Float	Комісія агрегатору
Rule_id	Int	Умови за якими працює водій
City_id	Int	ID міста у якому працює водій
Is_Canceled	Bool	Замовлення відмінено
Payment_amount	Float	Фіксовані нарахування за замовленням
Bonus_amount	Float	Бонусні нарахування за замовленням
Payment_return	Float	Списання фіксованої оплати за замовленням
Bonus_return	Float	Списання бонусів замовленню

Is_payment_fraud	Bool	Існують повні списання фіксованої оплати за замовленням
Is_bonus_fraud	Bool	Існують повні списання бонусів за замовленням
Is_fraud	Bool	Хоча б одна з колонок Is_bonus_fraud або Is_payment_fraud істина
Card_payment	Float	Сума безготівкового платежу
Subsidies_percentage_05_fraud	Bool	Відсоток бонусу від замовлення за 0,05 квантилем
Subsidies_percentage_95_fraud	Bool	Відсоток бонусу від замовлення за 0,95 квантилем
Orders_amount_05_fraud	Bool	Вартість замовлення за 0,05 квантилем
Orders_amount_95_fraud	Bool	Вартість замовлення за 0,95 квантилем
bonus_amount_05_fraud	Bool	Вартість бонусу за 0,05 квантилем
bonus_amount_95_fraud	Bool	Вартість бонусу за 0,95 квантилем
avg_orders_amount_05_fraud	Bool	Середня вартість замовлення за 0,05 квантилем

avg_orders_amount_95_fraud	Bool	Середня вартість замовлення за 0,95 квантилем
avg_bonus_amount_05_fraud	Bool	Середня вартість бонусу за 0,05 квантилем
avg_bonus_amount_95_fraud	Bool	Середня вартість бонусу за 0,95 квантилем
hour_trips_05_fraud	Bool	Кількість замовлень за годину за 0,05 квантилем
hour_trips_95_fraud	Bool	Кількість замовлень за годину за 0,95 квантилем
hour_amount_05_fraud	Bool	Вартість замовлень за годину за 0,05 квантилем
hour_amount_95_fraud	Bool	Вартість замовлень за годину за 0,95 квантилем
day_limit_80_fraud	Bool	Щоденний ліміт без врахування бонусів за 0,08 квантилем
day_limit_without_bonuses_80_fraud	Bool	Щоденний бонусний ліміт за 0,08 квантилем

Для виявлення шахрайства застосуємо критерії, що показали кращий результат:

- вартість замовлень за 0,95 квантилем
- середня вартість замовлень за 0,95 квантилем
- кількість замовлень за годину за 0,05 квантилем
- сума замовлень за годину за 0,95 квантилем

- щоденний ліміт без врахування бонусів за 0,8 квантилем

Якщо хоча б за одним з критеріїв транзакція вважається шахрайською – вона блокується і позначається як шахрайський платіж.

Для отримання і порівняння результатів розглянемо наступні показники:

- Кількість транзакцій у таблиці для обраного міста

Для отримання цих даних використано функцію `len()` для підрахунку рядків у таблиці: `len(city_data)`

- Загальна сума транзакцій

Використано функцію `sum()` по колонці `'amount'` у таблиці: `sum(city_data['amount'])`

- Сума реальних транзакцій:

Фільтруємо дані в яких статус `'is_payment_fraud' = False`:
`city_data[~city_data['is_payment_fraud']]`

- Сума шахрайських транзакцій:

Фільтруємо дані в яких статус `'is_payment_fraud' = True`:
`city_data[city_data['is_payment_fraud']]`

- Заблокована сума шахрайських платіжок:

Фільтруємо дані по відібраним критеріям і статусу `'is_payment_fraud'=True`, рахуємо суму платежів

- Заблокована сума реальних платіжок:

Фільтруємо дані по відібраним критеріям і статусу `'is_payment_fraud'=False`, рахуємо суму платежів

За цими показниками маємо наступний результат для кожного міста:

Для міста Дніпро:

- Всього транзакцій у таблиці: 19 625
- Загальна сума транзакцій: 3 816 211
- Сума шахрайських транзакцій: 5 121
- Сума реальних транзакцій: 3 811 090
- Заблокована сума шахрайських платіжок: 3 654 із 5 121 (71%)
- Заблокована сума реальних платіжок: 1 956 038 із 3 811 090 (51%)

Для міста Одеса:

- Всього транзакцій у таблиці: 37 892
- Загальна сума транзакцій: 5 954 967
- Сума шахрайських транзакцій: 10 986
- Сума реальних транзакцій: 5 943 981
- Заблокована сума шахрайських платіжок: 3 654 із 10 986 (33%)
- Заблокована сума реальних платіжок: 1 956 038 із 5 943 981 (32,9%)

Для міста Києва:

- Всього транзакцій у таблиці: 216 894
- Загальна сума транзакцій: 72 623 092
- Сума шахрайських транзакцій: 189 301
- Сума реальних транзакцій: 72 433 791
- Заблокована сума шахрайських платіжок: 109 833 із 189 301 (57%)
- Заблокована сума реальних платіжок: 28 417 737 із 72 433 791 (38%)

Для міста Харків:

- Всього транзакцій у таблиці: 89 852

- Загальна сума транзакцій: 23 273 246
- Сума шахрайських транзакцій: 60 744
- Сума реальних транзакцій: 23 212 502
- Заблокована сума шахрайських платіжок: 38 793 із 60 744 (63,8 %)
- Заблокована сума реальних платіжок: 8 899 737 із 23 212 502 (38,3%)

Для міста Львів:

- Всього транзакцій у таблиці: 7 692
- Загальна сума транзакцій: 895 047
- Сума шахрайських транзакцій: 1 556
- Сума реальних транзакцій: 893 491
- Заблокована сума шахрайських платіжок: 856 із 1 556 (55%)
- Заблокована сума реальних платіжок 358 015 із 893 491 (40%)

ВИСНОВКИ

У даній роботі було проведено аналіз замовлень та транзакцій водіїв в межах одного таксопарку для п'яти міст України: Дніпра, Києва, Львова, Одеси та Харкова. Досліджено критерії для завчасного виявлення фродових транзакцій. Більше детально про результат виконаної роботи:

- Проведено підготовку даних до аналізу, що включає у себе дослідження даних, попередню обробку, зміну назв та типів значень, створення нових стовпців для більш детального аналізу та об'єднання даних з різних таблиць.
- Проведено дослідження критеріїв для завчасного виявлення можливого шахрайства.
- Виявлено, що найкращими квантилями для розглянутих критеріїв є 0,95 та 0,05 квантиль
- Проведено аналіз результату об'єднання декількох критеріїв
- Виявлено відсотки заблокованих транзакцій при об'єднаних критеріях в різних містах
- Розглянуто принцип роботи агрегаторів таксі та існуючі зони шахрайства

Результат виконаного дослідження дозволив виявити первинні критерії, що можуть допомогти таксопаркам, особливо тим, що мають моментальні виплати, більш якісно боротися з шахрайськими транзакціями. Це дозволить партнерам швидше провайдерів дізнаватися про те, що водій намагається обдурити та заробити нечесним шляхом.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Diletant.media [Електронний ресурс] – Режим доступу до ресурсу: <https://diletant.media/articles/25683415/>
2. taxiservice.com.ua [Електронний ресурс] – Режим доступу до ресурсу: <https://taxiservice.com.ua/ukraine/2158-agregator-taksi-hto-eto-takoe-i-osnovnye-printsipy-raboty-agregatorov>
3. Агрегатор таксі [Електронний ресурс] – Режим доступу до ресурсу: https://ru.wikipedia.org/wiki/%D0%90%D0%B3%D1%80%D0%B5%D0%B3%D0%B0%D1%82%D0%BE%D1%80_%D1%82%D0%B0%D0%BA%D1%81%D0%B8
4. Uber-driver [Електронний ресурс] – Режим доступу до ресурсу: https://uber-driver.ru/bonusy_uber
4. Fraud methods that taxi drivers use [Електронний ресурс] – Режим доступу до ресурсу: <https://medium.com/@ahmhasan947/10-methods-of-fraud-that-taxi-drivers-use-5687eb1b7deb>
5. rabotavtaxi.com [Електронний ресурс] – Режим доступу до ресурсу: <http://rabotavtaxi.com/ua/news/728>
6. Fandango A. Python Data Analysis - Second Edition / Armando Fandango., 2008.