

Міністерство освіти і науки України
Національний технічний університет
«Дніпровська політехніка»

Факультет інформаційних технологій
(факультет)

Кафедра системного аналізу та управління
(повна назва)

ПОЯСНЮВАЛЬНА ЗАПИСКА

кваліфікаційної роботи ступеню магістра

Студента Осипенко Олексія

академічної групи 124м-23-1

спеціальності 124 Системний аналіз

на тему: Аналіз шкоди цивільному населенню в Україні на основі інцидентів, зафіксованих у період російської агресії: дослідження на основі відкрити даних

Керівники кваліфікаційної роботи розділів	Прізвище, ініціали	Оцінка за шкалою		Підпис
		рейтинговою	інституційною	
Інформаційно-аналітичний розділ	к.ф.-м.н., доц. Хом'як Т.В.			
Спеціальний розділ	к.ф.-м.н., доц. Хом'як Т.В.			
Рецензент	к.ф.-м.н., доц. Хом'як Т.В.			
Нормоконтролер	к.ф.-м.н., доц. Хом'як Т.В.			

Дніпро

2024

ЗАТВЕРДЖЕНО:

завідувач кафедри

Системного аналізу та управління

(повна назва)

к.т.н., доц. Желдак Т.А.

(підпис)

(прізвище, ініціали)

«___» _____ 20__ року

ЗАВДАННЯ

на кваліфікаційну роботу
ступеню магістра

студента Осипенко О. академічної групи 124м-23-1

спеціальності: 124 Системний аналіз

на тему: Аналіз шкоди цивільному населенню в Україні на основі інцидентів, зафіксованих у період російської агресії: дослідження на основі відкрити даних

затверджену наказом ректора НТУ «Дніпровська політехніка»

від 16.10.2024 № 1388-с

Розділ	Зміст виконання	Терміни виконання
1. Інформаційно-аналітичний розділ	<i>Проаналізувати структуру об'єкта дослідження. Визначити предметну область дослідження та проблему, що розв'язується. Обґрунтувати методи виконання поставлених завдань</i>	10.09.2024 – 01.11.2024
2. Спеціальний розділ	<i>Розв'язати поставлені задачі: зібрати необхідні дані та виконати часово-просторовий аналіз інцидентів, що дозволить краще оцінювати ризики та розробляти практичні рекомендації для гуманітарної допомоги населенню.</i>	01.11.2024 – 30.12.2024

Завдання видано

_____ (підпис керівника)

доц. Хом'як Т.В.

_____ (прізвище, ініціали)

Дата видачі: 06.09.2024

4

Дата подання до екзаменаційної комісії: _____

Прийнято до виконання

_____ Осипенко О.М.

РЕФЕРАТ

Пояснювальна записка: 61 с., 23 рис., 2 додатки, 24 джерела.

Об'єкт розробки: методи аналізу інцидентів, що завдали шкоди цивільному населенню України.

Предмет дослідження: методи та засоби статистичного і геопросторового аналізу інцидентів із використанням відкритих даних.

Мета кваліфікаційної роботи: аналіз динаміки, частотності та географічної концентрації інцидентів, що вплинули на цивільне населення, для підвищення ефективності моніторингу та прийняття рішень щодо заходів безпеки.

Методи дослідження: використання статистичних моделей (Пуассонівський процес, автокореляція), кластеризація (K-means) та аналіз часових рядів.

Наукова новизна: робота виявляє закономірності в розподілі інцидентів у часі та просторі, вперше поєднуючи аналіз часових рядів із геопросторовою кластеризацією для моніторингу конфліктів.

Практичне значення роботи: результати дослідження можуть бути застосовані для гуманітарного реагування, оцінки ризиків, покращення моніторингу зон конфліктів та планування заходів із безпеки цивільного населення.

Ключові слова: ІНЦЕНДЕНТИ, СТАТИСТИЧНИЙ АНАЛІЗ, ГЕОПРОСТОРОВІ ДАНІ, КЛАСТЕРИЗАЦІЯ, ЧАСОВІ РЯДИ, ВІДКРИТІ ДАНІ, МОНІТОРИНГ КОНФЛІКТУ, ПУАССОНІВСЬКИЙ РОЗПОДІЛ, K-MEANS, АВТОКОРЕЛЯЦІЯ.

ABSTRACT

Explanatory note: 61 pages, 23 pictures, 2 appendices, 24 sources.

Object of research: analysis methods for incidents affecting civilians during armed conflicts.

Subject of research: statistical and geospatial analysis methods for incidents using open-source data.

Purpose of Master's thesis: to analyze the dynamics, frequency, and geographic concentration of incidents affecting civilians to enhance monitoring and decision-making for safety measures.

Research methods: statistical modeling (Poisson process, autocorrelation), clustering (K-means), and time series analysis.

Originality of research: the study identifies patterns in the temporal and spatial distribution of incidents, uniquely combining time series analysis with geospatial clustering for conflict monitoring.

Practical value of the results: the findings can be applied to humanitarian response, risk assessment, conflict zone monitoring, and planning safety measures for civilians.

Keywords: INCIDENTS, STATISTICAL ANALYSIS, GEOSPATIAL DATA, CLUSTERING, TIME SERIES, OPEN-SOURCE DATA, CONFLICT MONITORING, POISSON PROCESS, K-MEANS, AUTOCORRELATION.

ЗМІСТ

СПИСОК УМОВНИХ ПОЗНАЧЕНЬ	7
ВСТУП	8
РОЗДІЛ 1. ІНФОРМАЦІЙНО-АНАЛІТИЧНИЙ	11
1.1 Актуальність дослідження	11
1.2 Мета та задачі роботи	13
1.3 Основні задачі роботи	14
1.4 Опис об'єкту дослідження	15
1.4.1 Аналіз цивільних об'єктів у контексті війни.	15
1.4.2 Основні категорії цивільних об'єктів	16
1.4.3 Географічне та часово-історичне охоплення об'єкту	16
1.4.4 Типологія уражень цивільних об'єктів.....	17
1.4.5 Інформаційне підґрунтя для опису об'єкту.....	17
1.4.6 Особливості та специфіка об'єктів аналізу.....	18
1.5 Проблеми та виклики, пов'язані з об'єктом.....	19
1.5.1 Технічні та методологічні виклики	20
1.5.2 Етичні аспекти документування.....	20
1.5.3 Організаційні та ресурсні обмеження.....	20
1.6 Порівняльний аналіз та обґрунтування унікальності підходу	21
1.7 Методи, засоби і моделі аналізу	24
1.8 Висновки до інформаційно-аналітичного розділу	26
РОЗДІЛ 2. ПРИКЛАДНИЙ	28
2.1. Дані та їх джерела	28
2.1.1. Опис використаних даних.....	28

2.1.2 Важливість параметрів для дослідження	29
2.1.3. Джерела даних та період дослідження	29
2.1.4 Часові рамки аналізу.....	30
2.1.5 Надійність та репрезентативність джерел.....	31
2.1.6 Роль джерел у дослідженні	31
2.3 Гістограма частоти ураження	35
2.4 Розрахунок середнього значення та дисперсії.....	36
2.4.1 Середнє значення кількості інцидентів на день	37
2.4.2 Дисперсія кількості інцидентів на день.....	38
2.5 Розрахунок середнього квадратичного відхилення.....	39
2.6 Розподіл кількості інцидентів.....	40
2.6 Пуассонівський розподіл.....	42
2.7 Тест Колмогорова-Смирнова (K-S тест).....	44
2.8 Лінійна регресія.....	45
2.9 Ковзна середня	46
2.10 Автокореляційна функція	48
2.11. Часткова автокореляційна функція (PACF)	50
2.11 Деталізація розрахунків для K-means кластеризації	53
2.12 Пуассонівський процес для аналізу частоти інцидентів.....	56
2.12.1 Основна модель Пуассонівського процесу:	56
2.12.2 Ймовірність кількості інцидентів на одиницю площі:.....	57
ВИСНОВКИ.....	60
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	62
Додаток А. Відомість матеріалів кваліфікаційної роботи.....	606
Додаток Б. ВІДГУК КЕРІВНИКА.....	607

СПИСОК УМОВНИХ ПОЗНАЧЕНЬ

OSINT – Open-Source Intelligence (розвідка на основі відкритих джерел)

ГІС – Геоінформаційні системи

ПТСР – Посттравматичний стресовий розлад

ACF – Autocorrelation Function (автокореляційна функція)

PACF – Partial Autocorrelation Function (часткова автокореляційна функція)

K-Means – Метод кластеризації (середні значення)

Survival analysis – Аналіз виживання

Hazard Model – Модель ризику

ГІС-аналіз – Геопросторовий аналіз даних

ММС – Мультимедійні свідчення

ПМП – Пункти медичної допомоги

СНД – Супутникові дані

ВСТУП

Сучасна історія України позначена серйозними викликами, серед яких особливо трагічним є збройний конфлікт, що триває внаслідок російської агресії. Від початку бойових дій значні території країни зазнали систематичних руйнувань, а населення – масштабних втрат, зокрема ураження цивільних об'єктів, масового переміщення людей та суттєвого погіршення якості життя у постраждалих регіонах. Наслідки цих подій не обмежуються економічними втратами та інфраструктурними збитками – вони безпосередньо позначаються на фізичному й психоемоційному стані мирних мешканців. Водночас, встановлення реальних масштабів, структури та характеру шкоди, завданої цивільному населенню, ускладнено через неповноту, фрагментарність та недоступність значної частини даних, а також дезінформаційні кампанії та пропагандистські впливи.

На тлі цих обставин окреслюється нагальна потреба у системному, науково обґрунтованому дослідженні шкоди цивільному населенню, що дозволить отримати більш повну та якісну картину реальної ситуації. Для цього доцільно спиратися не лише на офіційні звіти, статистичні дані та матеріали міжнародних організацій, а й на потужний потенціал відкритих джерел інформації (Open Source Intelligence – OSINT). Залучення розгалуженої мережі верифікованих фактів, супутникових знімків, геолокаційних сервісів, а також машиннонавчальних алгоритмів і статистичних методів надає можливість ідентифікувати закономірності, виявляти територіальну та часову динаміку уражень, оцінювати структурні особливості ушкоджень та прогнозувати потенційні ризики.

В умовах, коли класичні підходи до документування втрат цивільного населення та об'єктів виявляються недостатніми, інноваційні методики аналізу великих масивів даних, машинного навчання, геопросторового моделювання й статистичного прогнозування стають ключовим інструментом розширення аналітичних можливостей дослідників. Такий

підхід дозволяє не лише формувати кількісні оцінки масштабів шкоди, а й глибше інтерпретувати отримані результати, зіставляючи їх із контекстом конкретних географічних районів, певних фаз конфлікту та використаних видів озброєнь. Це, у свою чергу, створює основу для формування цілісної картини та вироблення рекомендацій, спрямованих на відновлення пошкоджених інфраструктурних об'єктів, реабілітацію постраждалих спільнот, мінімізацію ризиків для цивільного населення в майбутньому.

Таким чином, актуальність даного дослідження зумовлена поєднанням кількох факторів: по-перше, суттєвий вплив збройного конфлікту на цивільне населення та інфраструктуру України; по-друге, недостатня повнота і узгодженість інформаційної бази, доступної для дослідників; по-третє, потенціал застосування інноваційних аналітичних методів та технологій з метою підвищення якості та достовірності результатів аналізу.

Мета цього дослідження – здійснити комплексний аналіз шкоди цивільному населенню України в період російської агресії, спираючись на відкриті дані та застосовуючи сучасні інструменти обробки, верифікації та статистико-математичного моделювання. Для досягнення цієї мети передбачено розв'язання таких завдань, як:

- систематизація даних про інциденти, пов'язані з ураженням цивільних об'єктів;

- застосування геопросторового аналізу для виявлення територіальних концентрацій руйнувань;

- використання статистичних та машиннонавчальних методів для виявлення закономірностей, перевірки гіпотез про розподіл інцидентів, оцінки динаміки та прогностичних характеристик;

- аналіз типів об'єктів, які найчастіше зазнають пошкоджень, характеру та інтенсивності застосованої зброї, а також визначення факторів, що впливають на ризик та тривалість негативних наслідків.

Результати дослідження будуть корисні як для наукової спільноти, що досліджує конфліктні ситуації, гуманітарне право, міжнародні відносини та соціологію війни, так і для практиків – фахівців державних установ, міжнародних організацій, гуманітарних місій, що мають на меті зниження негативних наслідків бойових дій для мирного населення та планують діяльність з післяконфліктного відновлення.

Таким чином, у межах цієї роботи будуть закладені основи для створення комплексної системи аналізу шкоди цивільним об'єктам і населенню, що спиратиметься на відкриті дані та сучасні технології. Це дозволить підвищити ефективність отримання та обробки релевантної інформації, сприятиме формуванню більш прозорої та верифікованої картини поточної ситуації, а також стане надійним підґрунтям для розробки дієвих заходів із захисту цивільних та відновлення постраждалих територій.

РОЗДІЛ 1 ІНФОРМАЦІЙНО-АНАЛІТИЧНИЙ

1.1 Актуальність дослідження

Після початку повномасштабного вторгнення Росії на територію України у 2022 році світ зіткнувся з катастрофічними наслідками для цивільного населення. Це вторгнення призвело до значних руйнувань інфраструктури, численних жертв серед мирного населення та вимушеного переселення мільйонів людей, які були змушені залишити свої домівки в пошуках безпеки. Збір та аналіз даних про інциденти, що спричинили шкоду цивільним особам, є не лише важливим елементом документування воєнних злочинів, але й основою для розробки гуманітарних ініціатив і юридичних процедур, які можуть допомогти у відновленні справедливості та забезпеченні підтримки постраждалим [14].

У сучасних дослідженнях, спрямованих на аналіз шкоди цивільному населенню в умовах збройних конфліктів, особливу увагу приділено опрацюванню відкритих даних, що дозволяє отримати більш повну картину подій. Застосування методологій відкритих джерел (OSINT) та використання сучасних аналітичних інструментів і технологій великої даних (Big Data) дозволяють дослідникам ефективніше виявляти, аналізувати та верифікувати факти, пов'язані з впливом бойових дій на цивільне населення. Це, в свою чергу, сприяє більш точному розумінню масштабів і характеру шкоди, завданої мирним жителям.

В контексті російської агресії проти України фіксація, систематизація та оцінювання інцидентів, пов'язаних із впливом бойових дій на цивільне населення, набули виняткової актуальності. Дослідження з даної тематики базуються на міждисциплінарному підході, поєднуючи методи соціології, міжнародних відносин, безпекознавства, правознавства, а також інформаційних технологій. Такий підхід дозволяє не лише виявляти та документувати факти шкоди, заподіяної цивільним жителям, але й розробляти ефективні стратегії для їх захисту та підтримки в умовах конфлікту.

Значну роль у формуванні аналітичної бази відіграє використання офіційних звітів міжнародних організацій, що здійснюють моніторинг порушень прав людини та фіксацію втрат серед цивільного населення. Наприклад, Управління Верховного комісара ООН з прав людини (OHCHR) публікує розгорнуті звіти щодо цивільних втрат та порушень міжнародного гуманітарного права у східних регіонах України [3]. Подібні дослідження доповнюються даними від Спеціальної моніторингової місії ОБСЄ в Україні, яка також здійснює фіксацію та верифікацію інцидентів, використовуючи механізми польових спостережень та документування доказової бази [4]. Поряд із цим, неурядові організації та незалежні дослідницькі інститути, зокрема Bellingcat, застосовують відкриті джерела інформації, геолокаційні сервіси та інструменти візуального аналізу для перевірки автентичності фото- й відеоматеріалів, що дозволяє побудувати більш цілісну картину подій [20]. Схожі завдання розв'язують і науково-дослідні центри, які працюють з базами даних на кшталт ACLED (Armed Conflict Location & Event Data Project), забезпечуючи фахівців систематичними наборами геоприв'язаних даних щодо конфліктних інцидентів [4].

Сучасні методи, які використовуються для аналізу та обробки таких даних, включають складні алгоритми машинного навчання, технології обробки природної мови (NLP) та геоінформаційні системи (ГІС). Застосування методів кластеризації, семантичного аналізу текстів, автоматизованого визначення геопросторових координат та виявлення закономірностей у великих масивах даних сприяє оперативнішому виявленню тенденцій та потенційних порушень норм міжнародного гуманітарного права. Крім того, порівняльний аналіз результатів різних досліджень та підходів, що застосовуються при вивченні цієї предметної галузі, дозволяє виокремити оптимальні методики для підвищення точності й надійності отриманих результатів.

Інноваційні технології, такі як автоматизоване вилучення текстової інформації з аудіо- та відеоматеріалів, розпізнавання зображень та нейронні мережі для аналізу супутникових знімків, сприяють формуванню нових

методичних підходів та розширенню аналітичного інструментарію. Залучення різномірних джерел, включаючи медіа-платформи, соціальні мережі, офіційні прес-релізи, а також архіви міжнародних організацій, створює можливість формувати повніший і більш достовірний інформаційний масив.

Таким чином, сучасний стан досліджуваної предметної галузі характеризується широким використанням відкритих даних, міждисциплінарним підходом, активною взаємодією з міжнародними організаціями, застосуванням інноваційних інструментів та технологій аналізу. Це дозволяє підвищити рівень достовірності та повноти інформації, необхідної для всебічного розуміння масштабів і характеру шкоди, завданої цивільному населенню в Україні в умовах російської агресії, а також сформулювати науково обґрунтовані рекомендації для подальших досліджень та вироблення політик з мінімізації таких втрат.

1.2 Мета та задачі роботи

Мета даної роботи полягає в аналізі часово-просторового розподілу інцидентів, пов'язаних із шкодою цивільному населенню, з метою ідентифікації закономірностей, оцінки ризиків та розробки рекомендацій для покращення методів моніторингу зон конфлікту. Ця мета обумовлена потребою в отриманні надійних даних для зменшення гуманітарних наслідків конфліктів та підвищення ефективності гуманітарних і державних реагувань. Аналіз базується на інтеграції геоінформаційних систем, статистичних методів і сучасних OSINT-технологій, що дозволяють врахувати складність та динаміку досліджуваних процесів.

Для досягнення мети роботи сформульовано кілька ключових задач. По-перше, необхідно систематизувати дані про інциденти, пов'язані зі шкодою цивільному населенню, включаючи їх часово-просторову прив'язку, типи пошкоджень та вплив на різні групи населення. По-друге, провести детальний аналіз часової та просторової динаміки інцидентів із використанням геоінформаційних систем, супутникових знімків та мультимедійних OSINT-

джерел. По-третє, необхідно розробити статистичні моделі для оцінки ризиків ураження цивільної інфраструктури та побудови прогнозів щодо можливих зон ризику. По-четверте, слід ідентифікувати фактори, які визначають розподіл інцидентів, зокрема типи озброєнь, густоту населення, доступність інфраструктури та географічні характеристики територій.

Важливо інтегрувати дані з різних джерел, включаючи офіційні звіти, верифіковані супутникові знімки та результати OSINT-аналізу. Це дозволить створити комплексну базу для аналізу та виявлення закономірностей. Крім того, розробка прогнозних моделей допоможе оцінити потенційні ризики для цивільних об'єктів у майбутньому, що стане основою для розробки практичних рекомендацій щодо підвищення точності моніторингу зон конфлікту. Очікувані результати включають виявлення закономірностей у розподілі інцидентів, створення ефективних прогнозних моделей ризиків та надання конкретних рекомендацій для державних і гуманітарних організацій щодо зменшення впливу конфліктів на цивільне населення.

1.3 Основні задачі роботи

Задачі цієї роботи спрямовані на реалізацію комплексного аналізу часово-просторового розподілу інцидентів, пов'язаних із шкодою цивільному населенню, для ідентифікації закономірностей, оцінки ризиків і покращення моніторингу зон конфлікту. Основним кроком є збір та підготовка даних, що включає очищення, нормалізацію та інтеграцію з різних джерел, таких як офіційні звіти, OSINT, супутникові знімки та мультимедійні матеріали. Цей етап формує основу для подальшого аналізу.

Частотний аналіз та оцінка динаміки змін кількості інцидентів у часі дозволяють ідентифікувати ключові періоди активності, виявити піки та спад динаміки подій. Для цього використовуються часові ряди та гістограми, що дозволяють оцінити зміни в часі та створити основу для прогнозування.

Геопросторовий аналіз з кластеризацією допомагає виявити зони підвищеної інтенсивності інцидентів, що є критичним для розуміння просторових закономірностей.

Часовий аналіз на основі автокореляційної та часткової автокореляційної функцій (ACF, PACF) визначає залежності між інцидентами, дозволяючи оцінити періодичність та повторюваність подій. Ці результати слугують основою для побудови математичних моделей, таких як Пуассонівський процес, які забезпечують прогнозування інтенсивності подій у майбутньому.

Інтерпретація отриманих результатів включає виявлення ключових закономірностей, оцінку ризиків для цивільної інфраструктури та формування рекомендацій для покращення моніторингу зон конфлікту. Розроблені рекомендації можуть використовуватися для підвищення ефективності дій гуманітарних і державних організацій. Таким чином, виконання поставлених задач забезпечує комплексний аналіз наслідків конфліктів, спрямований на ідентифікацію закономірностей та мінімізацію ризиків для цивільного населення.

1.4 Опис об'єкту дослідження

1.4.1 Аналіз цивільних об'єктів у контексті війни.

Цивільними об'єктами в умовах збройних конфліктів визнаються всі ті матеріальні об'єкти, які не використовуються для військових цілей. Відповідно до норм міжнародного гуманітарного права, насамперед статей Додаткового протоколу I до Женевських конвенцій, цивільними об'єктами є житлові будинки, навчальні, медичні та культурні установи, об'єкти енергетики, водопостачання, транспорту та інші будівлі і споруди, не пов'язані з безпосереднім веденням бойових дій [21]. Відмежування цивільних об'єктів від військових цілей базується на принципі розмежування, що вимагає від сторін конфлікту постійно відрізняти цивільні об'єкти від військових.

1.4.2 Основні категорії цивільних об'єктів

До ключових категорій цивільних об'єктів належать:

- **Житлові будинки та житлові комплекси:** безпосередньо пов'язані з виживанням та базовими потребами населення, їх руйнування призводить до втрати житла, збільшення внутрішньо переміщених осіб та підвищення вразливості населення.
- **Медичні установи, освітні заклади, культурні й релігійні об'єкти:** пошкодження лікарень і поліклінік ускладнює надання медичної допомоги, тоді як знищення шкіл, університетів та інших закладів освіти перериває освітній процес. Руйнування бібліотек, музеїв, храмів негативно впливає на культурну спадщину та соціальну згуртованість.
- **Об'єкти критичної інфраструктури:** енергетичні, водопостачальні, каналізаційні та транспортні мережі є основою функціонування суспільства. Їх пошкодження призводить до перебоїв у постачанні електроенергії, води, тепла, палива, ускладнює логістику, постачання гуманітарної допомоги та евакуацію населення.
- **Промислові, торговельні та адміністративні будівлі, що не мають військового призначення:** їх руйнування дестабілізує місцеву економіку, спричиняє втрату робочих місць, зниження доходів населення та підвищення соціальної напруги.

1.4.3 Географічне та часово-історичне охоплення об'єкту

У дослідженні розглядається шкода, завдана цивільним об'єктам на території України, що зазнали обстрілів та ракетних ударів. Часові рамки охоплюють період від початку повномасштабного вторгнення 24 лютого 2022 року до 20 лютого 2024 року. Такий проміжок часу дозволяє простежити динаміку та характер змін у використанні зброї, інтенсивність атак та масштаб пошкоджень цивільної інфраструктури.

1.4.4 Типологія уражень цивільних об'єктів

Характер уражень варіюється від повного знищення до часткових пошкоджень, які ускладнюють або унеможливають функціонування об'єкта. Використання артилерійських систем, ракетних комплексів, авіаційних бомб та інших видів зброї має різний вплив на структуру будівель. Повне руйнування означає втрату об'єкта як такого, тоді як часткові пошкодження можуть вимагати капітального ремонту або реконструкції. Не менш важливим є непрямий вплив, коли ураження однієї ділянки інфраструктури спричиняє системні збої та викликає ланцюгову реакцію проблем у суміжних сферах життєзабезпечення [13].

1.4.5 Інформаційне підґрунтя для опису об'єкту

Опис об'єкту ґрунтується на даних з різних джерел, зокрема звітах міжнародних організацій (ОНЧР, ОБСЄ, МКЧХ, Amnesty International, UNOCHA), аналітичних матеріалах неурядових організацій, журналістських розслідуваннях та проектах з відкритих джерел (OSINT) [19–20]. Ключове значення має база даних, підготовлена Bellingcat та Global Authentication Project, де задокументовано понад 1 000 підтверджених інцидентів пошкоджень цивільних об'єктів, зокрема житлових будинків, медичних і освітніх установ та інфраструктурних елементів [7]. Це дозволяє провести детальний геопросторовий та часовий аналіз для встановлення тенденцій та закономірностей.

Верифікація, яка включає геолокаційний аналіз, перехресну перевірку з офіційними звітами, супутникові знімки, фото- та відеоматеріали, забезпечує високу ступінь надійності зібраних даних [14]. Таким чином, сформована інформаційна база виступає фундаментом для системного аналізу й забезпечує якісний опис досліджуваного об'єкту.

1.4.6 Особливості та специфіка об'єктів аналізу

Цивільні об'єкти, які стають мішенню під час збройного конфлікту, відзначаються різноманітністю будівельних матеріалів, архітектурних рішень і конструктивних типів. Житлові будинки, інколи побудовані без врахування стандартів захисту від вибухів чи осколків, вразливі до артилерійських обстрілів, ракетних та авіаударів. Медичні й освітні заклади зазвичай не передбачають посилені конструкції для протидії ударам, тому будь-які прямі ураження призводять до значних пошкоджень конструкцій та обладнання. Об'єкти критичної інфраструктури (наприклад, електричні підстанції, системи водопостачання) часто містять ключові інженерні вузли, руйнування яких викликає каскадні збої інших систем [24]. Таким чином, характер ушкоджень значною мірою визначається специфікою конструкції кожного окремого об'єкта та його призначенням.

Цивільні об'єкти виконують функції, які є фундаментальними для стабільності та добробуту суспільства. Житловий фонд забезпечує базову потребу в притулку, освітні заклади гарантують розвиток людського капіталу, а медичні установи забезпечують охорону здоров'я. Об'єкти інженерної інфраструктури, включаючи енергетику, водопостачання, каналізацію та транспортні вузли, формують основу нормального функціонування соціуму. Їх пошкодження призводить до прямих матеріальних втрат і непрямих наслідків, таких як гуманітарна криза, масова міграція та економічна дестабілізація [12].

Просторове розміщення цивільних об'єктів переважно орієнтоване на густонаселені райони – міста, містечка, селища. Урбанізовані зони, щільно забудовані житловими будинками та промисловими спорудами, створюють складні умови для документування шкоди та надання невідкладної допомоги. Натомість у сільській місцевості інфраструктура часто є більш розосередженою, що впливає на ступінь вразливості до певних типів озброєнь та ускладнює логістику евакуації та відновлення пошкоджених об'єктів. Таким чином, геопросторові особливості впливають на інтенсивність уражень, тип пошкоджень та оперативність реагування на гуманітарні виклики.

Характер руйнувань цивільних об'єктів не є статичним і змінюється відповідно до фаз конфлікту та застосовуваних систем озброєнь. На початкових етапах агресії пошкодження могли бути фрагментарними, тоді як із загостренням конфлікту, застосуванням важчого та дальнобійного озброєння руйнування стають масштабнішими та системнішими. Використання некерованих артилерійських систем або касетних боєприпасів призводить до хаотичних пошкоджень значної кількості об'єктів на великих територіях [6]. Ці динамічні зміни ускладнюють прогнозування масштабів шкоди та потребують постійного оновлення інформаційних баз даних.

Специфіка об'єкту дослідження ускладнює процес документування. Активні бойові дії, обмежений доступ до деяких районів, пропагандистський вплив та недостовірність окремих свідчень ускладнюють отримання повної картини. Використання супутникових знімків, геолокаційних сервісів, OSINT-методологій, а також співставлення візуальних доказів з офіційними звітами та свідченнями очевидців дає змогу підвищити надійність і точність даних. Зібрані таким чином відомості дозволяють краще зрозуміти реальні масштаби руйнувань, типи застосованого озброєння та характер наслідків для цивільних об'єктів [6].

1.5 Проблеми та виклики, пов'язані з об'єктом

Однією з головних проблем при аналізі шкоди, завданої цивільним об'єктам, є неповнота даних та труднощі з доступом до постраждалих територій. Збройний конфлікт, активні бойові дії, ризики для дослідників і журналістів, а також відсутність належної інфраструктури для фіксації ушкоджень ускладнюють отримання репрезентативної та точної інформації. Інформаційне поле переповнене пропагандою та дезінформацією, що підважує довіру до багатьох джерел та ускладнює встановлення істинних масштабів руйнувань [20].

1.5.1 Технічні та методологічні виклики

У відсутності уніфікованих стандартів збору та структурування даних складно порівнювати результати різних досліджень. Також динамічність бойових дій, зміна лінії фронту, використання нових видів озброєнь ускладнюють геолокацію та атрибуцію конкретних атак. Чимало методів OSINT потребують високого рівня технічної експертизи, застосування алгоритмів машинного навчання й спеціалізованого програмного забезпечення. Відсутність стандартизованих методик та складність аналізу великих масивів даних ускладнюють оперативне прийняття рішень [14].

1.5.2 Етичні аспекти документування

Опублікування даних про руйнування цивільних об'єктів може бути використане у пропагандистських цілях. Водночас, недостатня або надто деталізована інформація може поставити під загрозу безпеку очевидців, дослідників та місцевого населення. Важливим є дотримання принципів етичної журналістики та розумний баланс між публічністю інформації та безпекою її джерел.

1.5.3 Організаційні та ресурсні обмеження

Систематичний моніторинг ушкоджень цивільних об'єктів вимагає значних людських та фінансових ресурсів. Обмеженість коштів та відсутність необхідної матеріально-технічної бази стримують можливості проведення масштабних досліджень. Брак спеціалістів (аналітиків, ІТ-експертів, спеціалістів з міжнародного права) знижує ефективність процесу обробки, верифікації та аналізу даних.

1.6 Порівняльний аналіз та обґрунтування унікальності підходу

Порівняння базується на роботах, що досліджують медичні, соціологічні, геопросторові та інфраструктурні аспекти збройних конфліктів, зокрема дослідження Мейтленд, Болтуць та Азі що опубліковані в журналах *Conflict and Health* і *EClinicalMedicine*.

У дослідженні Мейтленд та співавторів [9], опублікованому в *EClinicalMedicine*, проаналізовано поранення цивільних осіб, спричинені вибуховими та балістичними механізмами, з метою створення шаблону навантаження на хірургічну допомогу. Робота базується на аналізі 983 випадків первинних травматичних хірургічних втручань, здійснених у військовому медичному закладі Великобританії в Афганістані. Основна увага приділяється порівнянню впливу вибухових і балістичних поранень на характер травм і обсяг необхідних хірургічних втручань серед дорослих і дітей.

Методологія дослідження базується на аналізі даних, зібраних у період з 2009 по 2014 рік. У вибірку увійшли лише цивільні особи, виключаючи комбатантів та перекладачів. Найбільш поширеними втручаннями при вибухових пораненнях були дебридмент, ампутації та лапаротомії, тоді як при балістичних пораненнях – дебридмент, лапаротомії та судинні процедури. Аналіз показав, що вибухові травми частіше призводять до ампутацій, тоді як балістичні поранення потребують частішого встановлення дренажів грудної клітки.

Результати підкреслюють, що дорослі частіше зазнавали травм верхніх і нижніх кінцівок, тоді як у дітей спостерігався вищий рівень травм голови та живота. Автори відзначають, що отримані дані можуть бути використані для планування медичних ресурсів у масових випадках травм, спричинених терористичними атаками або конфліктами, наприклад, в Україні. Хоча дослідження базується на військових даних, його результати мають значення і для цивільного медичного планування.

Робота робить вагомий внесок у розуміння специфіки медичних потреб у зоні конфлікту та може стати основою для покращення підготовки медичних

працівників, розподілу ресурсів і формування системи оперативного реагування у випадках масових поранень.

У магістерській роботі Болтуць [10], виконаній в Утрехтському університеті, досліджено переміщення українського населення внаслідок збройного конфлікту через використання інструментів Open-Source Intelligence (OSINT) і географічних інформаційних систем (ГІС). Основна увага приділяється аналізу просторових патернів внутрішнього переміщення, ідентифікації найбільш уразливих регіонів та оцінці масштабів мобільності населення. Авторка використовує мультимедійні докази, супутникові знімки, верифіковані джерела OSINT, а також офіційні звіти та статистичні дані для побудови детальної картини динаміки переміщення.

Робота базується на геопросторовому аналізі, який дозволяє ідентифікувати регіони з найбільшими потоками переселенців, визначити головні напрямки руху населення та оцінити причини такої мобільності. Використання ГІС-технологій дає змогу створювати інтерактивні карти, які візуалізують переміщення людей у часі та просторі. Окрема увага приділяється виявленню ключових факторів, що впливають на вибір напрямку переселення, таких як доступ до ресурсів, рівень безпеки, наявність інфраструктури та транспортної доступності.

Авторка аналізує, як конфлікт вплинув на соціальну структуру регіонів, зокрема через зміну демографічного складу населення. Визначається, які групи населення найбільше зазнали переміщення, та досліджуються їхні стратегії адаптації до нових умов. Використання OSINT забезпечує високу точність верифікації даних, що підвищує довіру до отриманих результатів.

Попри те, що робота пропонує детальний аналіз переміщення населення, вона обмежена у фокусі та не розглядає інші аспекти впливу конфлікту, такі як руйнування цивільної інфраструктури або динаміка бойових дій. Відсутність прогнозних моделей також ускладнює оцінку майбутніх тенденцій.

Унікальність цієї роботи полягає у міждисциплінарному підході, який поєднує ГІС і OSINT для аналізу переміщення людей, що робить її важливим внеском у дослідження демографічних наслідків збройних конфліктів

У дослідженні Азі та співавторів [11], опублікованому в журналі *Conflict and Health*, проведено детальний просторовий аналіз руйнувань критичної цивільної інфраструктури в секторі Газа під час військової кампанії з 7 жовтня по 22 листопада 2023 року. Основна увага приділяється документуванню та оцінці пошкоджень об'єктів, таких як медичні заклади, школи, системи водопостачання та інші критичні об'єкти, які забезпечують базові потреби цивільного населення. Автори використовують супутникові знімки, дані OSINT, а також матеріали, надані гуманітарними організаціями, для створення детальної картини наслідків бойових дій.

Методологія дослідження включає ГІС-технології, які дозволяють візуалізувати масштаби пошкоджень та ідентифікувати географічні патерни руйнувань. Особливу увагу приділено територіям, які зазнали найбільших втрат, з метою визначення ключових тенденцій у розподілі пошкоджень. Автори також аналізують наслідки руйнувань для доступу населення до базових послуг, таких як охорона здоров'я, освіта та водопостачання. Результати демонструють, що інфраструктурні втрати мають не лише фізичний, а й соціально-економічний вплив, посилюючи гуманітарну кризу та ускладнюючи відновлення регіону.

Одним із ключових досягнень дослідження є інтеграція супутникових знімків із іншими джерелами даних, що дозволяє підтверджувати інформацію та оцінювати її достовірність. Проте автори визнають певні обмеження, такі як відсутність аналізу часової динаміки пошкоджень або прогнозних моделей для оцінки майбутніх ризиків. Робота також має географічне обмеження, зосереджуючись виключно на секторі Газа.

Це дослідження є важливим внеском у вивчення наслідків військових дій для цивільної інфраструктури, пропонуючи ґрунтовний аналіз із використанням сучасних аналітичних інструментів. Однак його вузький географічний і

тематичний фокус залишає простір для розширення підходу, який міг би враховувати багатогранні аспекти впливу конфліктів на цивільні об'єкти та довгострокові наслідки. Незважаючи на це, робота демонструє важливість інтеграції різнорідних даних і сучасних методів аналізу для оцінки гуманітарної ситуації у зонах конфлікту.

На відміну від розглянутих робіт, моє дослідження має комплексний характер, охоплюючи різноманітні типи інфраструктури, включаючи житлові будинки, школи, транспортні, комерційні та медичні заклади. Використання супутникових знімків, OSINT-даних, мультимедійних доказів та офіційних звітів забезпечує інтеграцію різнорідних джерел. Застосування сучасних аналітичних методів, таких як статистичні моделі, кластеризація, Survival analysis та прогнозні моделі, дозволяє глибше зрозуміти динаміку та тенденції пошкоджень. Геопросторовий аналіз додає широти охоплення, дозволяючи ідентифікувати територіальні патерни руйнувань, а часові моделі дають змогу прогнозувати майбутні ризики. Таким чином, наша робота об'єднує сильні сторони попередніх досліджень, долаючи їхні обмеження, і пропонує унікальний міждисциплінарний підхід до аналізу наслідків збройних конфліктів.

1.7 Методи, засоби і моделі аналізу

У сучасних дослідженнях, спрямованих на аналіз шкоди цивільним об'єктам, застосовуються як традиційні підходи, так і сучасні інструменти аналізу даних, моделювання та машинного навчання. Це забезпечує можливість оперативного збору, верифікації, структуризації та статистичної обробки великого масиву інформації з різноманітних джерел.

Використання офіційної статистики, звітів міжнародних організацій (ООН, ОБСЄ, МКЧХ, Amnesty International), польових досліджень, інтерв'ю та свідчень очевидців слугують початковим базисом для розуміння ситуації. Незважаючи на обмежений доступ до зон конфлікту, ці методи дозволяють отримати базові відомості про характер і масштаби уражень.

Аналітичні підходи на базі відкритих джерел (OSINT)

Збір даних із соціальних мереж, медіа, відео- та фотоматеріалів ідентифікованих свідків, застосування геолокації, верифікаційних платформ та методологій (наприклад, Bellingcat) допомагають підтвердити факти атак, визначити точне місце, час і характер інцидентів.

Геопросторовий аналіз та картографування. Застосування ГІС-технологій та створення тематичних карт дозволяє встановити просторові патерни, визначити найбільш уразливі регіони, проаналізувати вплив топографії та інфраструктури на масштаб і тип уражень.

Технології машинного навчання та обробки великих даних. Автоматизована класифікація, NLP, алгоритми для розпізнавання об'єктів на супутникових і фотознімках дають змогу оперативно опрацьовувати значні обсяги даних. Машинне навчання допомагає виявляти кореляції, закономірності та взаємозв'язки між характеристиками інцидентів і параметрами об'єктів.

Моделі статистичного аналізу та прогнозування. Частотний аналіз визначає повторюваність інцидентів. Побудова гістограм розподілу частоти допомагає оцінити варіативність. Розрахунок середнього значення, дисперсії та середньоквадратичного відхилення дозволяє кількісно оцінити коливання. Тести на відповідність даних теоретичним розподілам (Пуассонівський, Колмогорова-Смирнова) перевіряють якість припасування. Лінійна регресія показує наявність трендів. Ковзна середня згладжує часові ряди. ACF і PACF розкривають часові залежності. Кластеризація K-Means групує інциденти за спільними ознаками. Survival analysis та Hazard Model оцінюють вплив параметрів (наприклад, "Type of area affected", "Weapon System") на тривалість і ймовірність негативних ефектів.

Інструменти візуалізації та дашборди. Інтерактивні дашборди й інфографіка спрощують сприйняття результатів, полегшуючи порівняння різних показників та виявлення тенденцій у часі та просторі.

1.8 Висновки до інформаційно-аналітичного розділу

У першій частині дослідження було сформульовано актуальність проблеми, яка полягає у необхідності детального аналізу шкоди цивільним об'єктам у контексті збройного конфлікту. Цей аналіз є критично важливим для гуманітарної сфери, забезпечення безпеки цивільного населення, а також для міжнародних організацій, які займаються документуванням наслідків військових дій. Було визначено, що дослідження спрямоване на заповнення прогалин у часово-просторовому аналізі інцидентів, а також на створення методології, що дозволить ідентифікувати закономірності та оцінювати ризики.

Розглянуто мету та задачі дослідження, які включали збір даних, їх підготовку, аналіз часової динаміки та географічного розподілу інцидентів, а також розробку математичних моделей для прогнозування ризиків. Усі задачі були чітко узгоджені із загальною метою роботи, що забезпечило структурований підхід до вивчення об'єкта.

Проведений опис об'єкта дослідження дозволив деталізувати його структуру. Було розглянуто визначення цивільних об'єктів, основні категорії, такі як житлові будинки, школи, медична інфраструктура та транспортні об'єкти, а також географічне охоплення та часово-історичний контекст. Виявлення типології уражень, зокрема пошкоджень, руйнувань і втрат функціональності, стало важливим етапом для подальшого аналізу. Інформаційне підґрунтя для вивчення об'єкта, що включало офіційні звіти, дані супутникових знімків і OSINT-матеріали, забезпечило достовірність отриманих результатів.

Окрему увагу було приділено викликам, пов'язаним із дослідженням. Технічні та методологічні проблеми, такі як неповнота даних і складність верифікації, були враховані для мінімізації похибок. Етичні аспекти документування, включаючи конфіденційність даних, отримали відповідне методологічне вирішення. Організаційні та ресурсні обмеження було враховано при визначенні масштабів і пріоритетів дослідження.

Порівняльний аналіз із іншими дослідженнями, зокрема роботами, що вивчають медичну інфраструктуру, переміщення населення чи локалізовані впливи конфлікту, дозволив визначити унікальність підходу. У цьому дослідженні забезпечено інтеграцію різнорідних даних із застосуванням сучасних аналітичних інструментів, що робить його вагомим внеском у вивчення наслідків збройного конфлікту.

Методи, засоби і моделі, використані у цьому розділі, продемонстрували свою ефективність для ідентифікації часово-просторових закономірностей. Використання ГІС-технологій, статистичних методів і машинного навчання забезпечило глибокий аналіз даних, дозволивши створити основу для подальших етапів дослідження.

РОЗДІЛ 2 СПЕЦІАЛЬНИЙ

2.1. Дані та їх джерела

2.1.1. Опис використаних даних

Для аналізу шкоди, завданої цивільному населенню України внаслідок військової агресії, використано відкритий набір даних, який містить інформацію про інциденти, що призвели до потенційного впливу на цивільних осіб. Цей набір даних був зібраний і верифікований дослідниками Bellingcat та учасниками Global Authentication Project, і включає наступні ключові параметри:

Унікальний ідентифікатор запису (ID): Кожен запис має унікальний ID, що дозволяє точно визначити та аналізувати інциденти окремо, запобігаючи дублюванню та систематизуючи дані.

Дата інциденту (Date): Цей параметр визначає час події, важливий для аналізу динаміки конфлікту, виявлення трендів і залежностей.

Географічні координати (Longitude, Latitude): Довгота і широта забезпечують точну локалізацію інцидентів, що є основою для геоаналізу, визначення зон активності та ідентифікації гарячих точок.

Місце (Location): Назва населеного пункту або регіону, де стався інцидент, доповнює координати, полегшуючи інтерпретацію даних і зв'язок з адміністративними територіями.

Тип ураженого об'єкта (Type of area affected): Вказує, чи був інцидент спрямований на цивільну інфраструктуру, житлові квартали, громадські установи чи інші об'єкти, допомагаючи оцінити вплив на цивільне населення.

Тип вогневої системи (Weapon system): Інформація про вид зброї, що застосовувалась (наприклад, артилерія, авіаудари, ракетні системи), дозволяє аналізувати специфіку використання озброєння в різних регіонах та періодах.

2.1.2 Важливість параметрів для дослідження

Кожен із зазначених параметрів виконує критичну роль у забезпеченні всебічного аналізу наслідків військових дій.

Виявлення просторово-часових патернів. Аналіз даних про час і місце інцидентів допомагає ідентифікувати періоди ескалації, зональну концентрацію подій та можливі тенденції у веденні військових дій.

Класифікація впливу на цивільних. Інформація про тип ураженого об'єкта дозволяє оцінити, які категорії інфраструктури чи житлових будинків найчастіше страждали, що є важливим для гуманітарних ініціатив.

Оцінка інтенсивності конфлікту. Тип зброї та частота її використання в конкретних регіонах можуть вказувати на інтенсивність військових дій, а також наявність змін у стратегії воюючих сторін.

Ідентифікація гарячих точок. Географічні координати та аналіз концентрації подій дозволяють визначити регіони з найбільшою кількістю інцидентів, що є критичним для організації гуманітарної допомоги.

Такі структуровані та верифіковані дані є основою для кількісного й якісного аналізу впливу війни на цивільне населення. Вони дозволяють не лише зафіксувати масштаб наслідків, але й зробити висновки про характер та інтенсивність конфлікту, а також оцінити ризики для населення в різних регіонах. Дані слугують базисом для ухвалення обґрунтованих рішень щодо захисту цивільного населення, планування гуманітарних операцій і документування порушень міжнародного гуманітарного права.

2.1.3. Джерела даних та період дослідження

Основними джерелами даних для дослідження стали проєкти Bellingcat та Global Authentication Project, які спеціалізуються на зборі та аналізі інформації про інциденти, пов'язані з військовими конфліктами. Ці організації використовують передові методи розслідувань з відкритих джерел (OSINT), поєднуючи дані з соціальних мереж, супутникових знімків, новинних ресурсів та свідчень очевидців.

- **Bellingcat:** незалежна група розслідувачів, яка відома своєю діяльністю у сфері документування порушень міжнародного гуманітарного права. Їхній підхід ґрунтується на перевірці достовірності даних за допомогою геолокації, хронологічного аналізу та кореляції з іншими джерелами.
- **Global Authentication Project:** міжнародна ініціатива, яка фокусується на перевірці автентичності інформації, використовуючи інструменти штучного інтелекту та аналітичні методики для виявлення фейків і забезпечення високої якості даних.

Обидва джерела забезпечують надійність та репрезентативність інформації завдяки суворим стандартам перевірки й багатоканальному збору даних. Дані, зібрані цими організаціями, включають деталі інцидентів, які стали основою для систематичного аналізу шкоди цивільному населенню.

2.1.4 Часові рамки аналізу

Період дослідження охоплює час із 24 лютого 2022 року (дата початку повномасштабного вторгнення Росії в Україну) до 20 лютого 2024 року. Обрання цих часових рамок обумовлено такими факторами:

- **Повний період ескалації конфлікту:** цей часовий інтервал охоплює як початкову фазу агресії, так і подальші етапи бойових дій.
- **Наявність структурованих даних:** період із 2022 до 2024 року характеризується систематичним збором даних про інциденти, що забезпечує їхню цілісність і точність.
- **Динаміка бойових дій:** цей період дозволяє аналізувати зміни в тактиці та інтенсивності бойових дій, їхній вплив на цивільне населення та інфраструктуру.

2.1.5 Надійність та репрезентативність джерел

- Надійність: обидві організації дотримуються високих стандартів перевірки даних, включаючи використання сучасних технологій верифікації. Завдяки багаторівневій перевірці, імовірність потрапляння недостовірної інформації до дослідження є мінімальною.
- Репрезентативність: набір даних охоплює широкий спектр інцидентів за географічним (різні регіони України) і часовим (постійний моніторинг) критеріями. Це забезпечує можливість узагальнення результатів дослідження на весь конфлікт.

2.1.6 Роль джерел у дослідженні

Дані, зібрані Bellingcat та Global Authentication Project, слугують основою для об'єктивного аналізу наслідків російської агресії. Вони дозволяють:

- Виявляти тренди та закономірності у динаміці конфлікту.
- Географічно локалізувати найбільш уражені райони.
- Оцінювати масштаби шкоди для цивільного населення та інфраструктури.

Надійність та репрезентативність цих даних забезпечують високу якість дослідження, що має важливе значення для гуманітарних, юридичних і стратегічних цілей.

2.2 Підготовка та обробка даних

Підготовка та обробка даних перед їх аналізом є критично важливими етапами для забезпечення коректності результатів і надійності висновків.

Першим кроком є завантаження та підготовка даних. Оскільки ми маємо справу з датами інцидентів, необхідно переконатися, що колонки з датами мають правильний формат, і що дані готові для подальшого аналізу. Крім того, дані було представлено у вигляді 4х csv файлів, для проведення повного аналізу, нам потрібно згрупувати дані в один датасет.

Для виконання цієї роботи використаємо Tableau.

Tableau - це система інтерактивної аналітики, що дає змогу в найкоротші строки проводити глибокий і різнобічний аналіз великих масивів інформації. Першим кроком завантажуюмо файли для аналітики в окрему папку.

Завантажуємо дані в Tableau.

Після цього поєднуємо дані, надані в окремих файлах в один датасет, використовуючи опцію UNION в Tableau (рис.2.1).

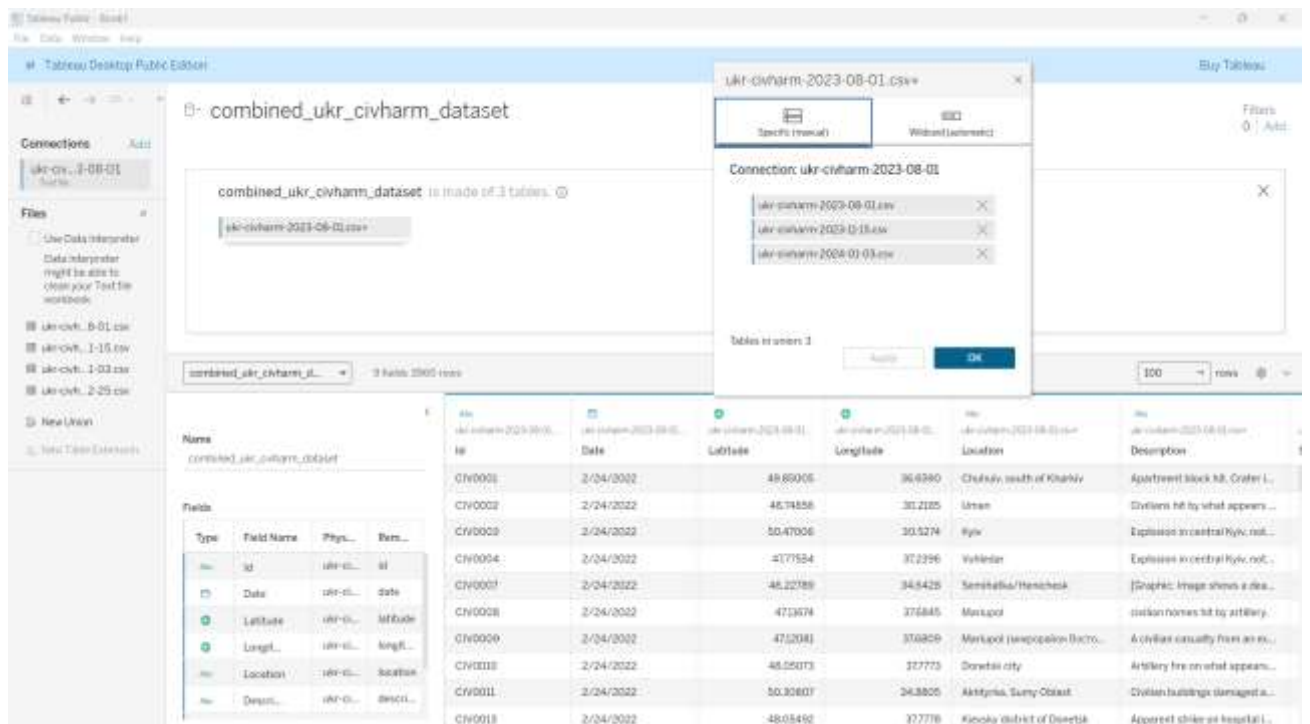


Рис. 2.1. Створення датасету в Tableau

Впевнюємося в правильності відображення завантажених даних.

Розбиваємо колонку 'Associations' щоб мати змогу відобразити розподіл по типам уражених об'єктів, та за типами систем ураження.

Для цього використовуємо опцію 'Custom Split' з сепаратором ',', (рис.2.2).

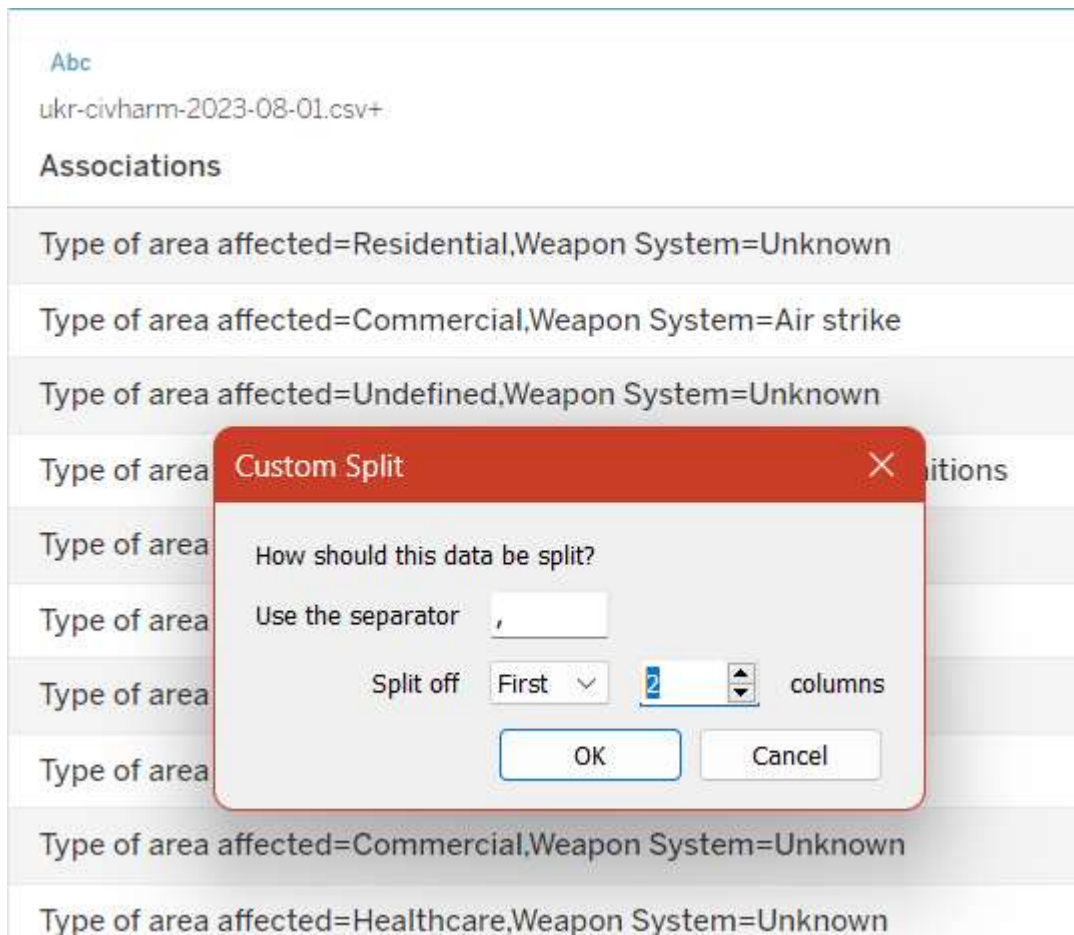


Рис. 2.2. Налаштування розділення колонок за типами

Після цього розділяємо новостворені колонки ще раз, аби позбутися заголовків.

Заголовки даних, створені автоматично переіменовуємо відповідно (рис.2.3).

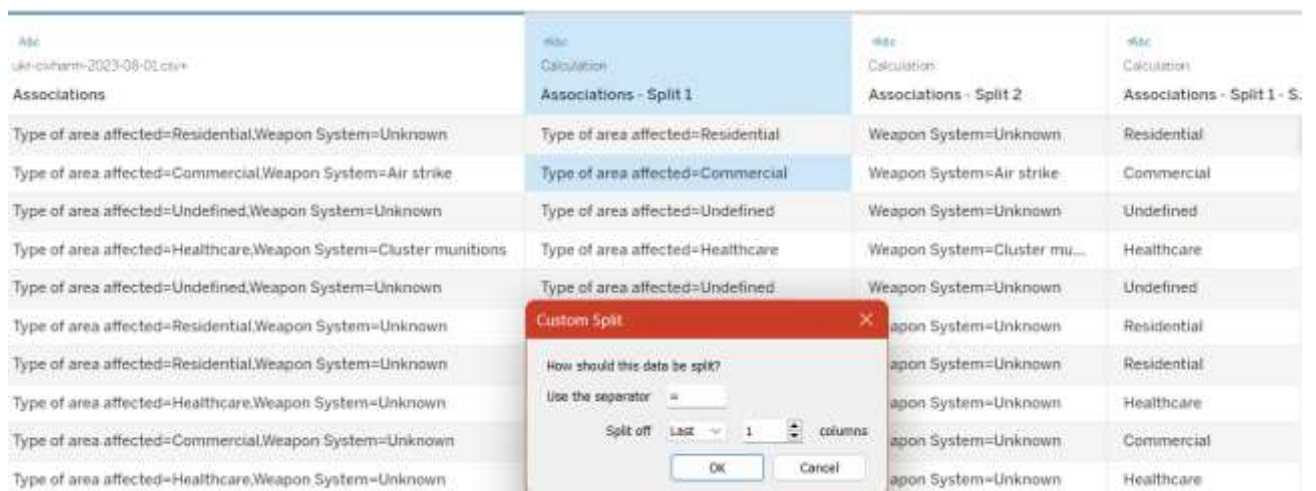


Рис. 2.3. Відділення заголовків колонок від даних

Наступним кроком проводимо перевірку пропущених значень, що є поширеною проблемою у великих наборах даних, зокрема при роботі з відкритими джерелами. У цьому дослідженні пропуски стосувалися ключових параметрів, таких як географічні координати, дата події або тип об'єкта. Виявлення пропусків проводилося через перевірку наявності значень у кожному критичному стовпці.

Повторювані записи можуть спотворювати результати аналізу, особливо при підрахунку кількості подій або оцінці їх частоти. Для ідентифікації дублікатів проводився пошук збігів у ключових стовпцях, таких як ідентифікатор запису, координати та дата події. Усі дублікати були видалені для забезпечення унікальності записів. Це особливо важливо для статистичних моделей, які враховують частотність подій.

Після виконання основних етапів обробки даних проведено перевірку на узгодженість ключових параметрів. Дані були перевірені на відсутність аномалій у часових рядах, коректність географічних меж та відповідність форматів. Цей етап забезпечив додаткову гарантію якості даних перед їхнім використанням у моделях.

Підготовка даних дозволяє усунути помилки, пов'язані з відсутніми значеннями, некоректними координатами або надмірністю записів, що критично важливо для точного аналізу часово-просторового розподілу інцидентів. Всі заходи обробки створюють основу для проведення коректного аналізу, підвищуючи достовірність результатів і обґрунтованість висновків.

Кінцевий датасет для аналізу має наступний вигляд (рис.2.4)

ID	Date	Latitude	Longitude	Location	Type of area affected	Weapon System
OIV0001	2/24/2022	49.85000	36.0090	Chukotka, south of Khatyrka	Residential	Unknown
OIV0002	2/24/2022	48.74856	38.2185	Uman	Commercial	Air strike
OIV0003	2/24/2022	50.47006	30.5274	Kyiv	Unaffected	Unknown
OIV0004	2/24/2022	47.77664	37.3396	Vuhledar	Healthcare	Casualty report
OIV0007	2/24/2022	46.22769	24.6428	Serdyukivka/Venetskivka	Unaffected	Unknown
OIV0008	2/24/2022	47.12674	37.6845	Manushev	Residential	Unknown
OIV0009	2/24/2022	47.12081	37.6809	Manushev (unpopulated St...)	Residential	Unknown
OIV0010	2/24/2022	48.09073	37.7773	Odesa city	Healthcare	Unknown
OIV0011	2/24/2022	50.30807	34.8005	Alitrykha, Sunny Dniest...	Commercial	Unknown
OIV0013	2/24/2022	48.66492	37.7778	Kievsky district of Odesa	Healthcare	Unknown

Рис. 2.4. Відображення датасету готового для подальшого аналізу

2.3 Гістограма частоти ураження

Гістограма частоти ураження дозволяє візуалізувати та аналізувати динаміку подій у різні періоди часу, забезпечуючи розуміння того, як частота інцидентів змінюється в залежності від дня тижня, календарного дня чи місяця. Такі дані можуть мати критичне значення для планування гуманітарних операцій, оцінки ризиків та розробки заходів з мінімізації шкоди.

Аналіз за днями тижня дозволяє виявити циклічність інцидентів, пов'язану з оперативною активністю у конфліктних зонах. Наприклад, деякі типи атак можуть частіше відбуватися у робочі дні через наявність більшої кількості людей у громадських місцях, таких як ринки, адміністративні будівлі чи освітні заклади. Інший тип атак може спостерігатися у вихідні, що може бути пов'язано зі зміною патернів пересування цивільного населення або специфікою бойових дій.

Гістограми за календарними днями дозволяють виявити короткострокові тренди, які можуть бути пов'язані з певними подіями, наприклад, інтенсивністю атак під час святкових днів, значущих дат або впродовж періодів активізації бойових дій. Аналіз цих даних дозволяє ідентифікувати дні, коли ризик для цивільного населення був найбільшим, що сприяє більш точному визначенню часових інтервалів для підвищеної уваги гуманітарних організацій.

Аналіз частоти інцидентів за календарними місяцями дає змогу вивчити довгострокові закономірності, такі як сезонні зміни в інтенсивності конфлікту. Наприклад, збільшення кількості атак може спостерігатися в літні місяці через покращені погодні умови, які сприяють мобільності військових підрозділів, або у зимовий період, коли атаки можуть бути спрямовані на руйнування об'єктів критичної інфраструктури, таких як електростанції чи системи опалення. Такий аналіз дозволяє визначати стратегічні періоди підвищеного ризику для цивільної інфраструктури та населення.

Важливість гістограм розподілу частоти ураження полягає у створенні цілісної картини динаміки інцидентів. Вони дозволяють гуманітарним організаціям і державним органам не лише розуміти поведінку конфлікту, але й

ефективніше планувати заходи з реагування та мінімізації шкоди. Аналіз часових патернів також є основою для побудови прогнозних моделей, що сприяє більш точному передбаченню майбутніх інцидентів та оцінці ризиків (рис.2.5).

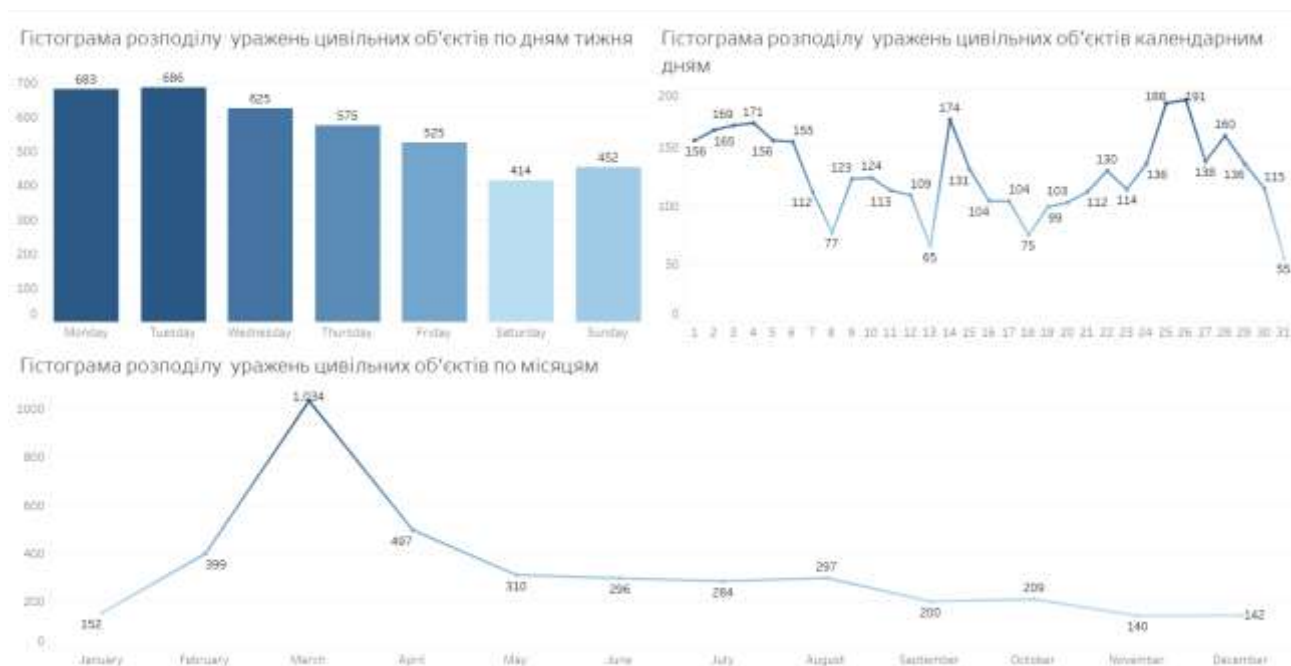


Рис. 2.5. Гістограми розподілу частоти ураження

2.4 Розрахунок середнього значення та дисперсії

Середнє значення та дисперсія є одними з базових статистичних характеристик, які дозволяють описати розподіл даних та оцінити їхню варіативність. У контексті дослідження наслідків збройного конфлікту, зокрема аналізу шкоди цивільним об'єктам, ці показники мають фундаментальне значення для формування уявлення про динаміку подій та ідентифікацію закономірностей.

Середнє значення покаже, скільки інцидентів відбувається в середньому щодня, тоді як дисперсія вказує на варіативність цієї кількості.

2.4.1 Середнє значення кількості інцидентів на день

Середнє значення (або математичне сподівання) — це основний статистичний показник, який показує середню кількість інцидентів за день.

Це дозволяє сформулювати базові припущення про інтенсивність інцидентів у досліджуваних умовах. Наприклад, середнє значення кількості атак за день може бути використане для оцінки типового рівня ризику для цивільного населення. Воно може слугувати основою для порівняння між різними регіонами, часовими інтервалами або категоріями об'єктів.

Формула середнього значення:

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i$$

Де:

- μ — середня кількість інцидентів на день.
- n — загальна кількість днів у вибірці.
- x_i — кількість інцидентів у день i .

Алгоритм розрахунку:

1. Підсумовуємо кількість інцидентів за кожен день.
2. Ділимо загальну кількість інцидентів на кількість днів.

У Python розрахунок середнього виглядає так (рис.2.6):

```
1 # Розрахунок середнього значення кількості інцидентів на день
2 mean_incidents_per_day = daily_incidents['incident_count'].mean()
3
```

Рис. 2.6. Програмний код розрахунку середнього числа на мові Python

Середня кількість інцидентів на день становить **11.15**. Це показник того, скільки інцидентів відбувається в середньому кожного дня.

2.4.2 Дисперсія кількості інцидентів на день

Дисперсія є важливим статистичним показником, який демонструє ступінь розкиду значень навколо середнього. Вона дозволяє оцінити, наскільки значення кількості інцидентів відрізняються одне від одного, що є критично важливим для розуміння стабільності або варіабельності процесу. Висока дисперсія вказує на значні коливання, тоді як низька свідчить про відносну стабільність. Це може бути корисним для аналізу даних у різних сферах, таких як економіка, соціологія, або управління ризиками, де важливо розуміти, як часто і наскільки значення можуть відхилятися від середнього.

Формула дисперсії:

$$D = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

Де:

- D — дисперсія кількості інцидентів на день.
- μ — середня кількість інцидентів.
- n — загальна кількість днів у вибірці.
- x_i — кількість інцидентів у день i .

Алгоритм розрахунку:

1. Для кожного дня обчислюємо різницю між кількістю інцидентів x_i і середнім значенням μ .
2. Квадратуємо ці відхилення для кожного дня.
3. Підсумовуємо отримані квадрати відхилень і ділимо на загальну кількість днів n .

У Python розрахунок дисперсії виглядає так (рис.2.7):

```
1 # Розрахунок дисперсії
2 variance_incidents_per_day = daily_incidents['incident_count'].var()
3
```

Рис. 2.7. Програмний код для розрахунку дисперсії на мові Python

- Дисперсія становить **201.76**. Вона вказує на те, що кількість інцидентів значно коливається від дня до дня.
- Велика дисперсія свідчить про нестабільність у частоті подій: є дні з великою кількістю інцидентів і дні з незначною активністю. Це може бути пов'язано з різними фазами конфлікту, локальними ескалаціями або з різними військовими стратегіями.

2.5 Розрахунок середнього квадратичного відхилення

Середнє квадратичне відхилення (стандартне відхилення) є коренем із дисперсії і дозволяє оцінити, наскільки типово кількість інцидентів на день відхиляється від середнього значення.

Формула середнього квадратичного відхилення:

$$\sigma = \sqrt{D}$$

Де:

- σ — середнє квадратичне відхилення.
- D — дисперсія кількості інцидентів.

Алгоритм розрахунку:

1. Беремо квадратний корінь з дисперсії.

У Python стандартне відхилення можна обчислити так (рис.2.8):

```

1  #Розрахунок середнього квадратичного відхилення
2  std_dev_incidents_per_day = daily_incidents.std()
3

```

Рис. 2.8. Програмний код для розрахунку стандартного відхилення

2.6 Розподіл кількості інцидентів

На гістограмі розподілу кількості інцидентів ми можемо побачити, як часто певна кількість інцидентів трапляється на день. Це допоможе виявити закономірності і сплески.

У Python це реалізується за допомогою функції `hist` із бібліотеки `matplotlib` (рис.2.9):

```
1 import matplotlib.pyplot as plt
2
3 # Налаштуємо розмір графіка
4 plt.figure(figsize=(10,6))
5
6 # Створимо гістограму з 30 бінарними інтервалами
7 plt.hist(daily_incidents['incident_count'], bins=30, edgecolor='black')
8
9 # Додаємо заголовок та підписи осей
10 plt.title('Частота інцидентів на день')
11 plt.xlabel('Кількість інцидентів на день')
12 plt.ylabel('Частота')
13
14 # Відображаємо сітку для кращої читабельності
15 plt.grid(True)
16
17 # Відображаємо графік
18 plt.show()
19
```

Рис. 2.9. Програмний код для відображення гістограми за допомогою Matplotlib

Інтерпретація результатів

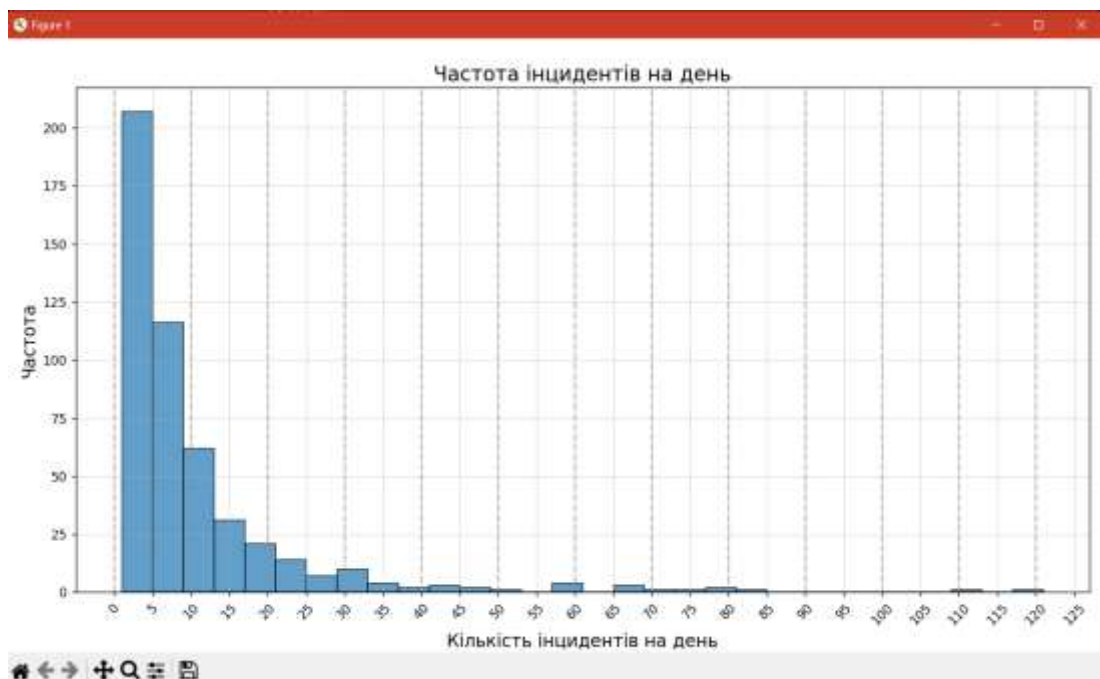


Рис. 2.10. Графік частоти інцидентів на день


```
C:\Users\Alex\PycharmProjects\ImageExtractionAutomatization\.venv\Scripts\python.exe C:\Users\Alex\PycharmProjects\ImageExtractionAutomatization\test.py
Середня кількість інцидентів на день: 11.153846153846153
Дисперсія: 201.75721641441773
Середнє квадратичне відхилення: 14.204126738888851
```

Рис. 2.11. Результати розрахунків

Після розрахунку середнього значення та дисперсії, а також побудови гістограми, можна зробити наступні висновки:

Гістограма частоти інцидентів за день

Гістограма демонструє, як часто зафіксовано певну кількість інцидентів за один день. На графіку видно, що більшість днів має кількість інцидентів близьку до середнього значення, але є також значна кількість днів із меншим або більшим числом інцидентів, що створює доволі широкий розподіл (рис.2.10). Це може свідчити про певні сплески у військовій активності або про періоди інтенсивнішого конфлікту.

Середнє значення

- Середня кількість інцидентів на день становить **11.15**. Це показник того, скільки інцидентів відбувається в середньому кожного дня (рис.2.11).

Дисперсія

- Дисперсія становить **201.76**. Вона вказує на те, що кількість інцидентів значно коливається від дня до дня.
- Велика дисперсія свідчить про нестабільність у частоті подій: є дні з великою кількістю інцидентів і дні з незначною активністю. Це може бути пов'язано з різними фазами конфлікту, локальними ескалаціями або з різними військовими стратегіями.

Ці статистичні показники є основою для подальшого аналізу динаміки інцидентів та їх геопросторового розподілу.

2.6 Пуассонівський розподіл

Пуассонівський розподіл описує ймовірність рідкісних подій, які відбуваються у певний проміжок часу або на певному просторі. Його часто використовують для моделювання таких явищ, як кількість інцидентів на день.

Функція ймовірності Пуассонівського розподілу:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Де:

- $P(X = k)$ — ймовірність того, що трапиться k подій;
- λ — середнє число подій (у нашому випадку — середня кількість інцидентів на день);
- e — експоненціальна функція.

Тест Колмогорова-Смирнова в нашому випадку показує, що дані про кількість інцидентів на день **не відповідають нормальному розподілу**, це може означати кілька важливих моментів:

- **Нерівномірність розподілу:** Інциденти не відбуваються стабільно кожного дня. Можливо, є дні з великою кількістю подій та дні з відносно малим числом інцидентів, що створює сильні відхилення від нормального розподілу.
- **Періоди ескалації:** Можливо, є дні з піками активності, що створюють відхилення від нормального розподілу. Це свідчить про те, що певні події (наприклад, військові операції) можуть спричиняти значне збільшення кількості інцидентів у короткий період.
- **Наявність аномалій:** Деякі дні можуть бути "аномальними" за кількістю подій. Наприклад, дні з масованими атаками чи великими інцидентами можуть суттєво впливати на розподіл.

Для тестування відповідності Пуассонівському розподілу, ми використовуємо K-S тест (рис.2.12):

```
1 # Тест Колмогорова-Смирнова для перевірки на розподіл Пуассона
2 poisson_test_stat, poisson_p_value = stats.kstest(
3     daily_incidents['incident_count'],
4     'poisson',
5     args=(daily_incidents['incident_count'].mean())
6 )
7 print(f"Пуассон p-value: {poisson_p_value}")
8
```

Рис. 2.12. Програмний код з перевіркою відповідності розподілу

Результат: Пуассон p-value: 1.28

Інтерпретація результатів:

- р-значення менше 0.05 означає, що розподіл не є нормальним.
- р-значення більше 0.05 означає, що дані можуть відповідати нормальному розподілу.

В нашому випадку **відповідають пуассонівському розподілу**, це свідчить про кілька речей:

- **Події є випадковими:** Пуассонівський розподіл описує події, які виникають випадковим чином у часі та просторі. Якщо дані відповідають цьому закону, то можна сказати, що інциденти є випадковими подіями, які виникають із певною частотою без чіткої закономірності.

Можливість використання Пуассонівських моделей: Це дозволяє використовувати моделі на основі Пуассона для прогнозування кількості подій у майбутньому. Наприклад, можна оцінити ймовірність того, що протягом наступних кількох днів трапиться певна кількість інцидентів

2.7 Тест Колмогорова-Смирнова (K-S тест)

Тест Колмогорова-Смирнова порівнює розподіл ваших даних із теоретичним нормальним розподілом. Він вимірює відстань між кумулятивною функцією розподілу (CDF) даних і нормальної теоретичної моделі.

Статистика тесту Колмогорова-Смирнова:

$$D_n = \sup_x |F_n(x) - F(x)|$$

Де:

- D_n — максимальна відстань між емпіричною функцією розподілу $F_n(x)$ та теоретичною функцією $F(x)$
- \sup_x — супремум (максимальне значення) абсолютної різниці.

В коді, написаного мовою Python це виглядає так (рис.2.13):

```
1 # Тест Колмогорова-Смирнова для перевірки нормального розподілу
2 ks_test_stat, ks_p_value = stats.kstest(
3     daily_incidents['incident_count'],
4     'norm',
5     args=(daily_incidents['incident_count'].mean(),
6           daily_incidents['incident_count'].std())
7 )
8 print(f"Колмогоров-Смирнов p-value: {ks_p_value}")
9
```

Рис. 2.13. Програмний код для перевірки нормального розподілу

Результат: Колмогоров-Смирнов p-value: 6.1

Тест Колмогорова-Смирнова в нашому випадку показує, що дані про кількість інцидентів на день **не відповідають нормальному розподілу**, це може означати кілька важливих моментів:

- **Нерівномірність розподілу:** Інциденти не відбуваються стабільно кожного дня. Можливо, є дні з великою кількістю подій та дні з відносно малим числом інцидентів, що створює сильні відхилення від нормального розподілу.

- **Періоди ескалації:** Можливо, є дні з піками активності, що створюють відхилення від нормального розподілу. Це свідчить про те, що певні події (наприклад, військові операції) можуть спричинити значне збільшення кількості інцидентів у короткий період.
- **Наявність аномалій:** Деякі дні можуть бути "аномальними" за кількістю подій. Наприклад, дні з масованими атаками чи великими інцидентами можуть суттєво впливати на розподіл.

2.8 Лінійна регресія

Лінійна регресія використовується для визначення загальної тенденції (тренду) зміни кількості інцидентів у часі. Основна ідея полягає в тому, щоб знайти найкращу пряму, яка представляє зміну залежної змінної (кількість інцидентів) у часі (незалежна змінна — день).

Формула лінійної регресії:

$$Y = \beta_0 + \beta_1 X$$

Де:

- Y — прогнозована кількість інцидентів;
- X — номер дня (час);
- β_0 — константа (перетин із віссю Y);
- β_1 — коефіцієнт нахилу лінії (показує, як зміна часу впливає на кількість інцидентів).

Алгоритм розрахунку:

1. **Збір даних:** Ми беремо кількість інцидентів на кожен день і додаємо колонку з номером дня X .
2. **Підбір параметрів:** Лінійна регресія мінімізує квадрат відхилень між фактичними значеннями Y і прогнозованими значеннями \hat{Y} за формулою:

$$\min \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Параметри β_0 і β_1 підбираються таким чином, щоб сума квадратів відхилень була мінімальною.

Приклад розрахунків на Python (рис.2.14):

```

1  # Підготовка даних для регресії
2  X = np.arange(len(daily_incidents)).reshape(-1, 1)
3  y = daily_incidents['incident_count'].values
4
5  # Створення моделі лінійної регресії
6  model = LinearRegression()
7  model.fit(X, y)
8
9  # Прогноз тренду
10 trend = model.predict(X)
11

```

Рис. 2.14. Програмний код, що демонструє використання Лінійної Регресії

2.9 Ковзна середня

Ковзна середня — це метод згладжування даних, який допомагає усунути короткострокові коливання та виявити більш тривалі тренди. Для цього ми використовуємо середнє значення певної кількості попередніх днів.

Формула ковзної середньої:

$$MA_t = \frac{1}{n} \sum_{i=t-n+1}^t Y_i$$

Де:

- MA_t — ковзна середня на день t ;
- n — кількість днів у ковзному вікні;
- Y_i — фактична кількість інцидентів у день i .

Алгоритм розрахунку:

1. Визначаємо вікно ковзної середньої (наприклад, 7 днів). Це означає, що для кожного дня ми будемо обчислювати середнє значення кількості інцидентів за попередні 7 днів.
2. Обчислюємо ковзну середню для кожного дня на основі середнього значення за вибране вікно.

Приклад розрахунків на Python (рис.2.15):

```
1 # Обчислення ковзної середньої (7 днів)
2 daily_incidents['moving_average'] = daily_incidents['incident_count'].rolling(window=7).mean()
3
```

Рис. 2.15. Програмний код розрахунку ковзної середньої за 7 днів

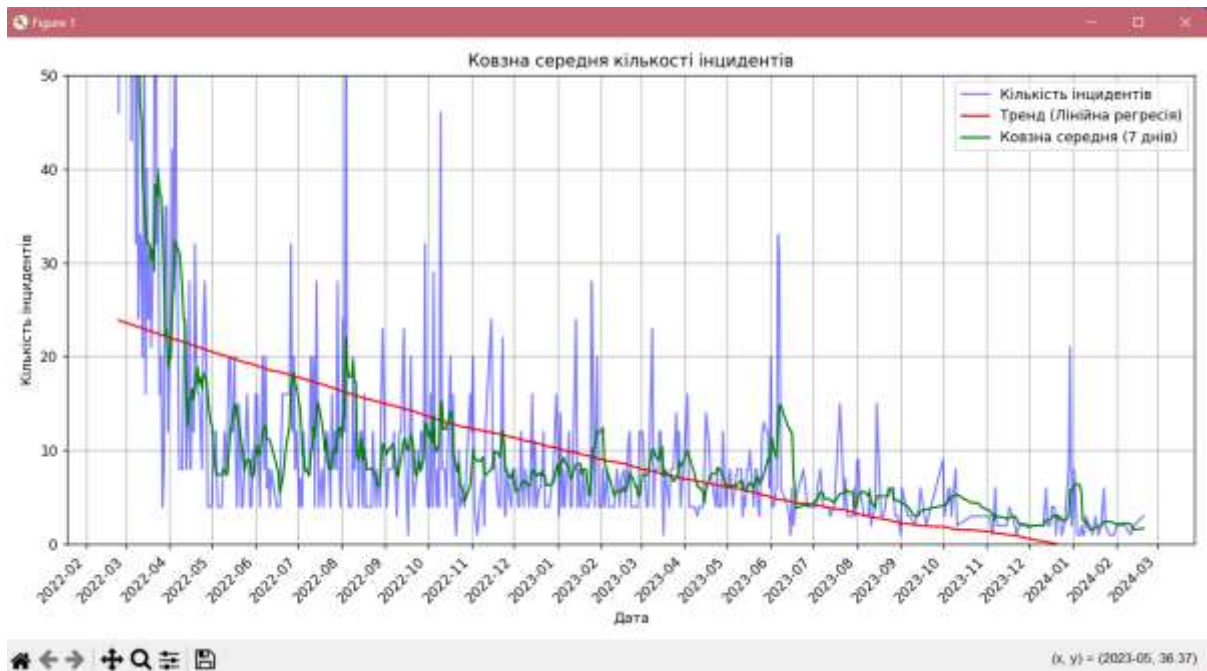


Рис. 2.16. Сукупний графік частоти подій, регресії і тренду

Інтерпретація результатів зображень на графіку (рис.2.16):

- Лінійна регресія (червона лінія) демонструє загальний спад кількості інцидентів протягом аналізованого періоду. Тренд, який показує ця червона лінія, вказує на те, що кількість інцидентів поступово зменшується з часом. Негативний нахил трендової лінії свідчить про можливі стабілізаційні процеси або ж про зниження інтенсивності бойових дій після найбільш активних фаз

конфлікту. Можна також припустити, що активність конфлікту в певні періоди суттєво спадала.

– Ковзна середня (зелена лінія), обчислена за 7 днів, виконує функцію згладжування короткострокових коливань. Вона усуває вплив випадкових піків, які спостерігаються на фактичних даних (синій лінії), і відображає більш стабільний тренд. На основі ковзної середньої можна зробити висновок, що кількість інцидентів значно зменшувалася з початку періоду, хоча окремі сплески активності все ж мали місце. Особливо помітними є сплески в середині 2022 року та на початку 2023 року, після яких ковзна середня підтверджує загальний тренд до зменшення.

– Фактичні дані, представлені синьою лінією, ілюструють щоденну кількість інцидентів. У перші місяці 2022 року інциденти траплялися значно частіше, із досягненням пікових значень — до 50 інцидентів на день. У подальшому, починаючи з середини 2022 року, можна спостерігати поступове зниження кількості інцидентів, хоча окремі сплески активності все ще мали місце. Ці сплески відображають локальні інтенсивні періоди бойових дій, але вони не змінюють загальної тенденції до зменшення частоти інцидентів.

2.10 Автокореляційна функція

Автокореляція дозволяє виявити, чи впливає кількість інцидентів у минулі дні на їх кількість у наступні дні. Для цього використовуються автокореляційна функція (ACF) і часткова автокореляційна функція (PACF), які допомагають зрозуміти структуру залежностей у часовому ряді.

Формула для автокореляції:

$$\rho_k = \frac{\sum_{t=k+1}^T (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^T (Y_t - \bar{Y})^2}$$

Де:

- ρ_k — автокореляція для лагу k ;
- Y_t — значення кількості інцидентів на день t ;
- \bar{Y} — середнє значення кількості інцидентів;
- T — загальна кількість днів у часовому ряді.

Алгоритм розрахунку:

1. Для кожного лагу k обчислюється, наскільки кількість інцидентів на день t пов'язана з кількістю інцидентів на день $t-k$.
2. ACF показує, чи є зв'язок між поточною кількістю інцидентів і подіями в минулі дні.

Інтерпретація:

- Якщо значення ACF значно відрізняється від нуля для певного лагу k , це вказує на наявність залежності між кількістю інцидентів на день t і кількістю інцидентів $t-k$.
- Наприклад, якщо автокореляція для лагу 7 є значущою, це може свідчити про тижневий цикл інцидентів.

2.11. Часткова автокореляційна функція (PACF)

PACF розраховується за допомогою регресійної моделі, яка усуває вплив проміжних значень між лагами (рис.2.17).

Алгоритм розрахунку:

1. Для кожного лагу k розраховується часткова кореляція між поточним значенням кількості інцидентів і значенням $t-k$, усуваючи вплив усіх проміжних днів.

2. PACF показує прямий вплив інцидентів на день t на інциденти на день $t-k$, виключаючи ефект усіх інших днів.

Інтерпретація:

– Якщо значення PACF значно відрізняється від нуля на певному лагу, це означає, що між інцидентами на день t і $t-k$ є пряма залежність, незалежно від того, що відбувалося між цими днями.

– Якщо значення PACF різко падає після кількох лагів, це вказує на те, що залежність слабшає після певної кількості днів.

Пояснення результатів аналізу

– ACF показує загальну кореляцію для всіх лагів. Якщо є тривала кореляція (зростання чи зменшення кореляцій), це свідчить про довготривалі залежності. Наприклад, якщо інциденти відбуваються регулярно через певний проміжок часу (тиждень, місяць), ACF покаже періодичні піки.

– PACF допомагає краще зрозуміти точну кількість днів, протягом яких існує залежність. Якщо PACF різко спадає, це означає, що після кількох днів (наприклад, після 7 днів) подальша кореляція є незначною, і інциденти є незалежними (рис..

```

1 # --- ACF ---
2 plt.figure(figsize=(12, 6))
3 plot_acf(daily_incidents['incident_count'], lags=50)
4 plt.title('Автокореляційна функція (ACF) для інцидентів')
5 plt.show()
6
7 # --- PACF ---
8 plt.figure(figsize=(12, 6))
9 plot_pacf(daily_incidents['incident_count'], lags=50)
10 plt.title('Часткова автокореляційна функція (PACF) для інцидентів')
11 plt.show()
12

```

Рис. 2.17. Приклад коду для демонстрації ACF та PACF функцій

Висновки на основі автокореляційної функції (ACF)

Значущі кореляції спостерігаються на перших лагах, особливо до 10 днів. На графіку ACF чітко видно високі значення кореляції на цих інтервалах, що свідчить про сильний вплив кількості інцидентів у певний день на наступні кілька днів. Найвищий рівень кореляції, майже 1.0, спостерігається на лагу 1, що вказує на тісний зв'язок між кількістю інцидентів у сусідні дні. Це може пояснюватися подіями, які провокують підвищену активність протягом кількох днів поспіль (рис.2.18).

Зменшення кореляції спостерігається після 10-го лагу, хоча значущі значення зберігаються до 20-го дня. Це свідчить про те, що вплив подій триває до трьох тижнів, поступово слабшаючи. Після 25-го дня кореляція наближається до нуля, що вказує на зникнення залежності від подій, які відбулися більше місяця тому.

Періодичні шаблони залишаються невираженими. Хоча кореляція є значущою у перші кілька днів, періодичних піків, що вказували б на регулярність подій (наприклад, щотижня чи щомісяця), не виявлено. Це свідчить про відсутність чіткої періодичності, хоча події можуть впливати на кількість інцидентів протягом кількох днів після початкового інциденту.

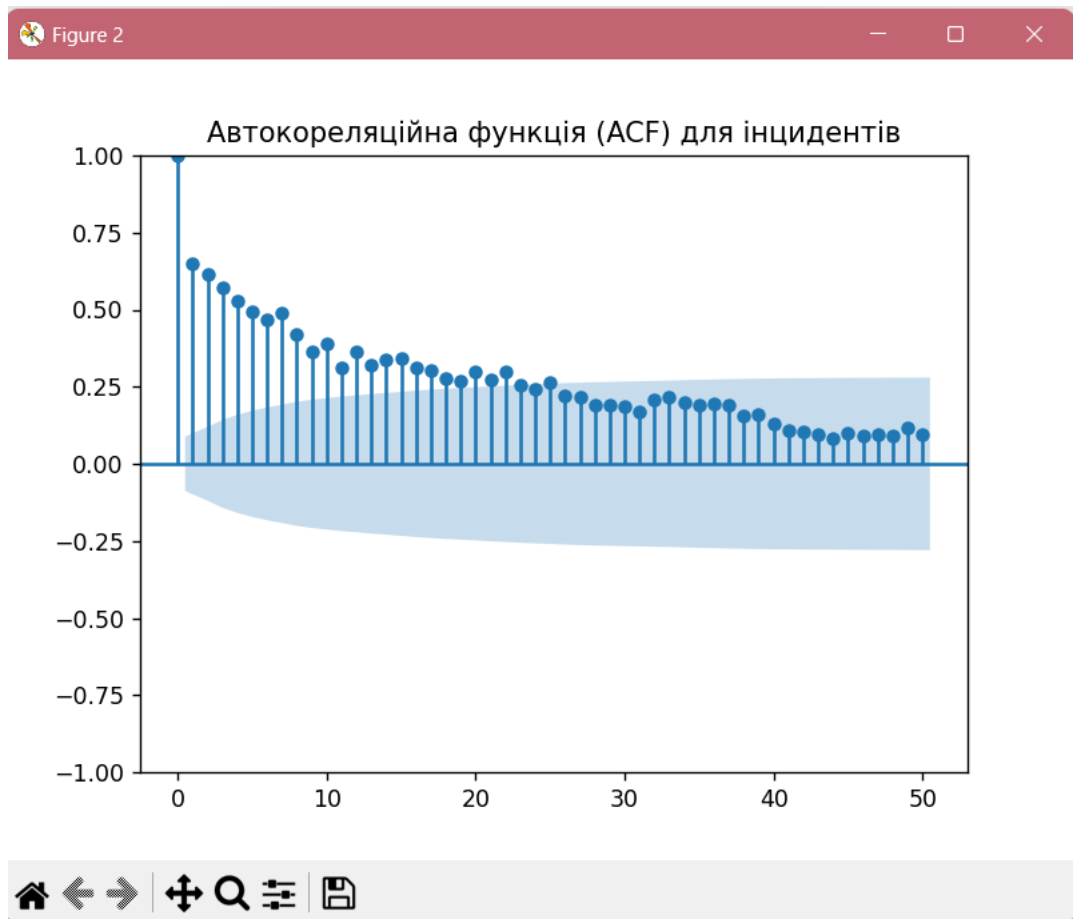


Рис. 2.18. Автокореляційна функція для інцидентів

Висновки на основі часткової автокореляційної функції (PACF)

Сильна часткова автокореляція спостерігається на першому і другому лагах. Перший лаг має майже максимальне значення (близько 1.0), що свідчить про сильний вплив кількості інцидентів на певний день на кількість наступного дня. Другий лаг також демонструє значну кореляцію, що вказує на вплив подій на два дні вперед.

Після другого лагу часткова автокореляція різко зменшується і стає незначною. Це свідчить про короткострокову залежність між інцидентами: інтенсивність подій триває кілька днів, після чого їхній вплив суттєво слабшає. Події після третього дня вже не мають суттєвого впливу на поточну кількість інцидентів.

На лагах більше 5 днів значуща кореляція відсутня, що свідчить про незалежність подій, які трапляються через більш ніж 5 днів. Це підтверджує відсутність довготривалого впливу подій або циклічності у даних (рис.2.19).

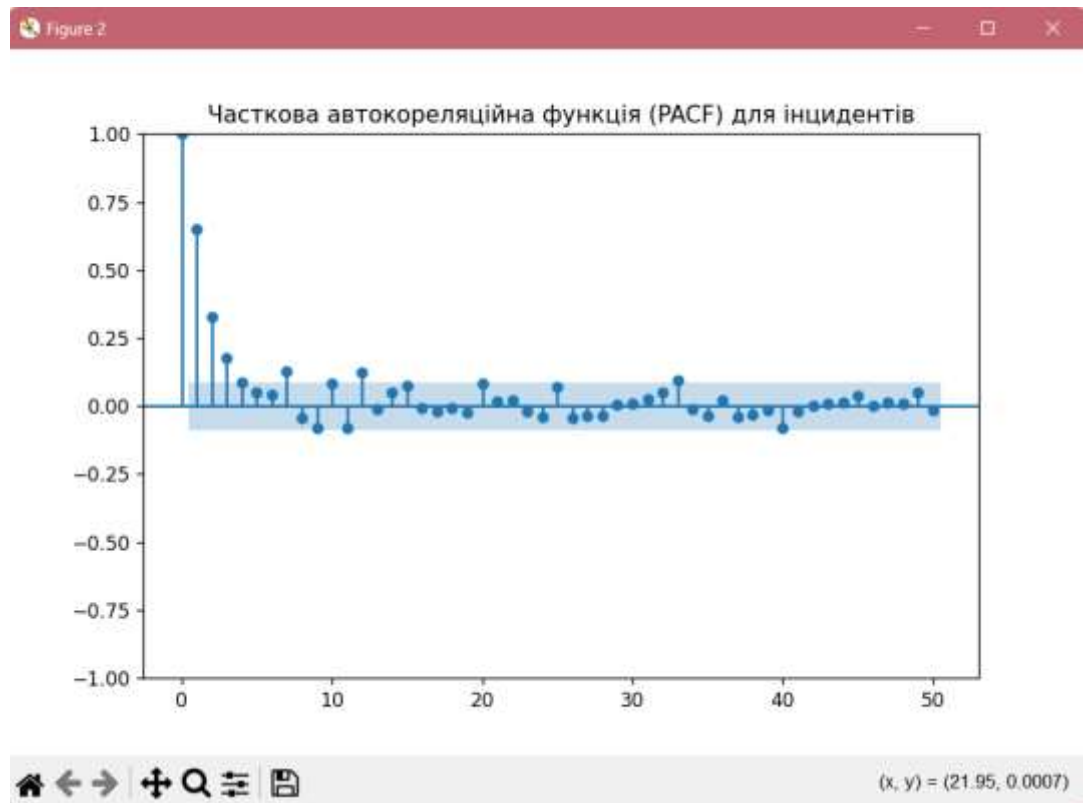


Рис. 2.19 Часткова автокореляційна функція для інцидентів

2.11 Деталізація розрахунків для K-means кластеризації

K-means — це один із найпоширеніших алгоритмів кластеризації, який групує інциденти на основі їх географічних координат (широта та довгота). Основна ідея полягає в тому, щоб мінімізувати відстані між точками в кожному кластері та їхнім центроїдом.

Алгоритм працює в кілька кроків (рис.2.21):

Крок 1: Вибираємо кількість кластерів k , яку потрібно створити (в нашому випадку це може бути 3, 5 або більше).

Крок 2: Алгоритм випадковим чином ініціалізує k центрів кластерів (центроїдів).

Крок 3: Для кожної точки (координати інциденту) обчислюється відстань до кожного центроїду, і точка відноситься до найближчого кластера.

Крок 4: Після присвоєння всіх точок до кластерів центроїди переміщуються до середніх координат точок у їхньому кластері.

Крок 5: Процес повторюється, поки центроїди більше не змінюють свої позиції або зміни дуже незначні (досягнуто умов зупинки).

Математично K-means мінімізує суму квадратичних відстаней між кожною точкою x_i і центроїдом кластера μ_j :

$$\min \sum_{i=1}^n \sum_{j=1}^k \|x_i - \mu_j\|^2$$

Де:

- x_i — точка (широта, довгота) інциденту;
- μ_j — координати центроїда кластера;
- k — кількість кластерів;
- n — кількість інцидентів.

Вибір кількості кластерів

Метод ліктя - це один із методів визначення оптимальної кількості кластерів k . Для кожного можливого k обчислюється сума квадратних відстаней між точками та їхніми центроїдами. Оптимальне значення k - це точка, де сума відстаней різко зменшується (утворюється "лікоть").

Розрахунок відстаней і центроїдів

Для кожної точки обчислюється **евклідова відстань** до кожного центроїда:

$$d(x_i, \mu_j) = \sqrt{(x_i^{(1)} - \mu_j^{(1)})^2 + (x_i^{(2)} - \mu_j^{(2)})^2}$$

Де:

$x_i^{(1)}, x_i^{(2)}$ — координати (широта і довгота) точки x_i ;

$\mu_j^{(1)}, \mu_j^{(2)}$ — координати центроїда μ_j

Центроїди обчислюються як середні значення всіх точок у кожному кластері:

$$\mu_j^{(1)} = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i^{(1)}, \quad \mu_j^{(2)} = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i^{(2)}$$

Де:

- C_j — сукупність точок у кластері j ;
- $|C_j|$ — кількість точок у кластері j .

Візуалізація результатів

Після проведених розрахунків було отримано мапу з відображенням інцидентів (рис.2.20):

- Точки інцидентів відображаються на карті, кожна з яких має свій колір залежно від кластера.
- Центроїди кластерів (середні точки) показують центральну позицію для кожного кластера і відображаються окремо на графіку червоними хрестиками.

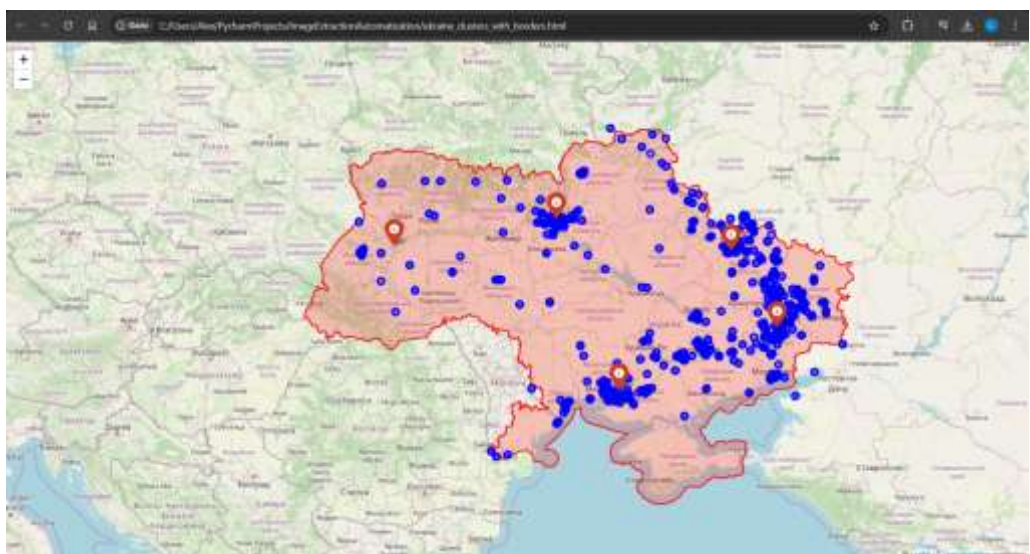


Рис. 2.20. Демонстрація результатів K-means кластеризації

характеристикою Пуассонівського процесу є **інтенсивність** λ , що відповідає середній кількості подій на одиницю площі або часу.

Формула Пуассонівського розподілу:

$$P(k) = \frac{(\lambda A)^k e^{-\lambda A}}{k!}$$

Де:

- $P(k)$ — ймовірність того, що на певній площі A станеться k подій.
- λ — середня кількість інцидентів на одиницю площі (інтенсивність).
- A — площа (в нашому випадку, площа квадрата на карті).
- k — кількість подій (інцидентів).

Кроки розрахунку

Оцінка інтенсивності λ

Інтенсивність λ — це середня кількість інцидентів на одиницю площі.

Вона обчислюється як кількість інцидентів у регіоні, поділена на площу цього регіону.

Формула для інтенсивності:

$$\lambda = \frac{\text{incident_count}}{\text{region_area}}$$

Де:

- `incident_count` — кількість інцидентів у даному квадраті.
- `region_area` — площа регіону (квадрата), яку визначаємо на карті.
Для квадрата 100x100 км це 10,000 км².

2.12.2 Ймовірність кількості інцидентів на одиницю площі:

Після оцінки інтенсивності λ для кожного квадрата, можемо обчислити ймовірність того, що у даному квадраті станеться певна кількість подій k .

Формула для ймовірності:

$$P(k) = \frac{(\lambda A)^k e^{-\lambda A}}{k!}$$

Де:

- $P(k)$ — ймовірність того, що на площі A станеться k подій.
- A — площа квадрата.

Опис і пояснення коду (рис.2.22):

```

1 import numpy as np
2 import pandas as pd
3 from scipy.stats import poisson
4 import matplotlib.pyplot as plt
5
6 # Завантаження даних
7 df = pd.read_csv('path_to_combined_ukr_civharm_dataset.csv')
8
9 # Припускаємо, що в колонках є широта (latitude) і довгота (longitude)
10 df = df.dropna(subset=['latitude', 'longitude'])
11
12 # Розмір квадрата для розбиття території (наприклад, 100 км)
13 grid_size = 100 # в км
14
15 # Розрахунок координат для квадратів, до яких належать інциденти
16 df['x_grid'] = (df['longitude'] // grid_size) * grid_size
17 df['y_grid'] = (df['latitude'] // grid_size) * grid_size
18
19 # Групування інцидентів за координатами квадратів
20 region_incidents = df.groupby(['x_grid', 'y_grid']).size().reset_index(name='incident_count')
21
22 # Площа кожного квадрата (100x100 км = 10,000 км²)
23 region_area = grid_size ** 2
24
25 # Оцінка інтенсивності подій (λ) для кожного регіону
26 region_incidents['intensity'] = region_incidents['incident_count'] / region_area
27
28 # --- Моделювання Пуассонівського процесу ---
29 # Для кожного регіону можемо змоделювати ймовірність кількості подій на основі інтенсивності
30 for i, row in region_incidents.iterrows():
31     lambda_region = row['intensity'] * region_area # Іntenсивність на одиниці площі
32     k_values = np.arange(0, 20) # Кількість подій
33     probabilities = poisson.pmf(k_values, lambda_region) # Ймовірності для кожного k
34     print(f"Region: ({row['x_grid']}, {row['y_grid']})")
35     print(f"Ймовірність різної кількості подій: {probabilities}")
36

```

Рис. 2.22. Програмний код для розрахунку ймовірності інцидентів у регіонах
Алгоритм роботи програмного коду:

1. Завантаження та підготовка даних:

Завантажуються дані про інциденти, фільтруються рядки з відсутніми координатами (широтою і довготою).

2. Розбиття на квадрати:

Територія розбивається на квадрати розміром 100x100 км для оцінки кількості інцидентів у кожному квадраті.

3. Оцінка інтенсивності:

Для кожного квадрата обчислюється інтенсивність подій λ , що є середньою кількістю інцидентів на 1 км².

4. Моделювання Пуассонівського процесу:

Для кожного квадрата обчислюється ймовірність того, що в ньому станеться певна кількість подій k за допомогою розподілу Пуассона. Результати відображаються для кожного квадрата на основі його інтенсивності (рис.2.23).

Інтерпретація результатів

- Інтенсивність подій: Якщо в різних квадратах інтенсивність λ відрізняється, це свідчить про те, що події не є випадковими і зосереджені в певних регіонах.
- Ймовірність кількості подій: Для кожного квадрата можна обчислити ймовірність того, що в ньому станеться певна кількість подій (наприклад, від 0 до 20 подій).

Цей підхід дозволяє глибше зрозуміти, як події розподіляються в просторі і які регіони мають вищу ймовірність виникнення інцидентів.

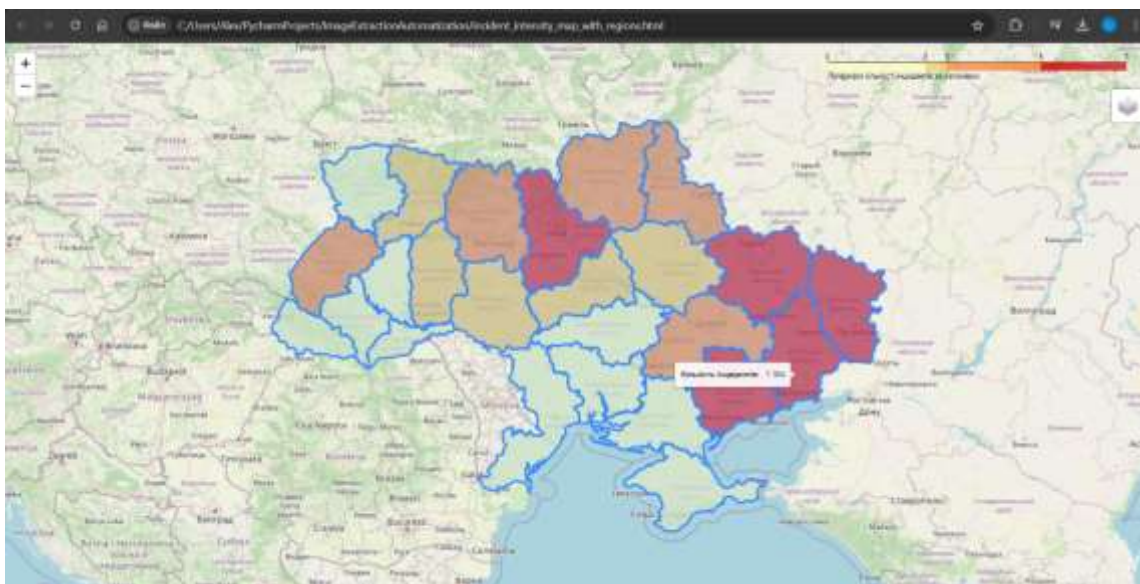


Рис. 2.23 Мапа розподілу інцидентів за регіонами

ВИСНОВКИ

У ході виконання дослідження було реалізовано комплексний підхід до аналізу інцидентів, пов'язаних із шкодою цивільному населенню України під час російської агресії в період з 24 лютого 2022 року по 20 лютого 2024 року. Актуальність роботи обумовлена значними гуманітарними наслідками конфлікту, що вимагають детального вивчення закономірностей і факторів, які впливають на безпеку цивільного населення. Основною метою дослідження було проведення аналізу часово-просторового розподілу інцидентів для ідентифікації ключових закономірностей, оцінки ризиків і розробки практичних рекомендацій для покращення моніторингу зон конфлікту. Для досягнення цієї мети були поставлені задачі, які охоплювали збір і підготовку даних, статистичний аналіз, побудову моделей і розробку висновків.

Ключовим результатом дослідження стало виявлення суттєвих часових і просторових закономірностей у розподілі інцидентів. Середня кількість подій на день становила 11.15, проте значна дисперсія свідчить про нерівномірність розподілу. Аналіз часових рядів показав, що інциденти мають короткострокову залежність, що зникає через кілька днів, без прояву чіткої довгострокової періодичності. Це свідчить про швидке згасання ефекту ескалації, що необхідно враховувати при розробці стратегій оперативного реагування. Геопросторовий аналіз дозволив визначити регіони з підвищеною концентрацією подій, особливо у східних регіонах України, які залишаються найбільш уразливими до інтенсивних атак. Метод кластеризації K-means допоміг виділити гарячі точки, що є критично важливими для гуманітарних та безпекових заходів.

Застосування статистичних моделей, зокрема Пуассонівського процесу, надало можливість оцінити ймовірність виникнення певної кількості інцидентів на одиницю площі. Лінійна регресія виявила загальну тенденцію до зменшення кількості інцидентів у часі, що може свідчити про стабілізацію конфлікту або зміну його характеру. Водночас коварна середня демонструє періодичні сплески активності, що корелюють із локальними ескалаціями бойових дій. Аналіз

частотного розподілу також підтвердив наявність значної варіативності, що дозволило виявити періоди найбільшої інтенсивності конфлікту.

Науковий внесок роботи полягає у застосуванні інтегрованого підходу, який поєднує часові, просторові та статистичні методи аналізу. Використання геоінформаційних систем, OSINT, супутникових даних та сучасних моделей дозволило досягти глибокого рівня розуміння динаміки конфлікту. Унікальність підходу підкреслюється здатністю поєднувати різноманітні джерела даних для побудови комплексних моделей аналізу.

Практичне значення отриманих результатів полягає у можливості застосування висновків для покращення систем моніторингу зон конфлікту, планування гуманітарних операцій та розробки стратегій мінімізації ризиків для цивільного населення. Ідентифіковані закономірності можуть бути використані для прогнозування ризиків і розробки заходів із захисту населення у найбільш уразливих регіонах. Крім того, результати роботи можуть бути корисними для міжнародних організацій та дослідницьких груп, які займаються оцінкою наслідків збройних конфліктів.

Дослідження також виявило низку обмежень, зокрема залежність від доступності даних, географічні та часові рамки, а також обмеження методів аналізу. Ці виклики створюють перспективи для подальших досліджень, які можуть включати інтеграцію нових джерел даних, розширення аналітичних підходів і поглиблення аналізу впливу конфлікту на соціально-економічну сферу.

Загалом, проведене дослідження підтвердило високу цінність комплексного підходу до аналізу інцидентів у зоні конфлікту. Отримані результати створюють основу для подальшого розвитку методів моніторингу зон конфлікту, забезпечуючи надійну базу для прийняття рішень у сфері безпеки, гуманітарної допомоги та захисту цивільного населення.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Кваліфікаційна робота магістра [Електронний ресурс] : методичні рекомендації для здобувачів ступеня магістра освітньо-професійної програми «Системний аналіз» зі спеціальності 124 Системний аналіз / уклад.: Т.А. Желдак, Т.В. Хом'як, А.В. Малієнко ; М-во освіти і науки України, Нац. техн. ун-т «Дніпровська політехніка». – Дніпро : НТУ «ДП», 2024. – 33 с.
<https://ir.nmu.org.ua/handle/123456789/167921>
2. Хом'як Т. В. Бази даних у професійних задачах аналітики [Електронний ресурс] : навч. наочн. посіб. / Т. В. Хом'як, К. С. Хабарлак, Д.М. Гаранжа; М-во освіти і науки України, Нац. техн. ун-т «Дніпровська політехніка». – Дніпро : НТУ «ДП», 2024. – 192с.
3. OSCE SMM. Civilian casualties in the conflict-affected regions of eastern Ukraine.
4. ACLED. Armed Conflict Location & Event Data Project.
<https://acleddata.com/>
5. Research article. Typology and implications of verified attacks on health care in Ukraine in the first 18 months of war. May 2024
6. HRW. Ukraine: Russia's Use of Cluster Munitions Harms Civilians. June, 2023
7. Bellingcat. Hospitals Bombed and Apartments Destroyed: Mapping Incidents of Civilian Harm in Ukraine
8. OSCE. Thematic Report: Protection of Civilians in Ukraine. March 2021.
9. Maitland L, Middleton L, Veen H, et al. Analysis of 983 civilian blast and ballistic casualties and the generation of a template of injury burden: An observational study. *EClinicalMedicine*. December 2022.
10. Bořtuć K. Examining the Displacement of the Ukrainian Population through the Lens of Open-Source Intelligence and Geographic Information Systems. Master's thesis, Utrecht University. 2023.

11. Asi Y, Mills D, Greenough PG, et al. 'Nowhere and no one is safe': spatial analysis of damage to critical civilian infrastructure in the Gaza Strip during the first phase of the Israeli military campaign, 7 October to 22 November 2023. *Conflict and Health*. 2024
12. Молоканова, В. М., & Шевченко, Ю. О. (2024). Управління проектною командою. <https://ir.nmu.org.ua/handle/123456789/167646>
13. UNESCO. In the face of war, UNESCO's action in Ukraine. 2023
<https://unesdoc.unesco.org/ark:/48223/pf0000384454>
14. Rogatinska N., Halahan O., Protsiuk O., Galagan S., Fierieva N. War crimes and crimes against humanity in Ukraine: Legal qualification and features of documentation. *Cuestiones Politicas*, September 2023
15. Шрамко С.С. Правові механізми відшкодування шкоди жертвам воєнного конфлікту в Україні. Том 4 № 84 (2024): Науковий вісник Ужгородського національного університету.
16. Коряшкіна, Л. С., Станіна, О. Д., & Шевченко, Ю. О. (2024). Практикум з диференційних рівнянь. <https://ir.nmu.org.ua/handle/123456789/167658>
17. Хасті, Т., Тібширані, Р., Фрідман, Дж. *Елементи статистичного навчання: дані, прогнози та алгоритми*. Нью-Йорк: Springer 2009
18. Шевченко, Ю. О. (2022). Обробка і аналіз даних з використанням електронних таблиць. Частина I «Обробка даних». <https://ir.nmu.org.ua/handle/123456789/162623>
19. UNOCHA. Ukraine: Humanitarian Needs Overview. <https://www.unocha.org/ukraine>
20. Bellingcat Ukraine Project. Civilian Harm in Ukraine.
21. ICRC. Geneva Conventions and their Additional Protocols.
22. Statista Research Department. Civilian deaths related to the Russia-Ukraine conflict 2014-2024.
23. Bellingcat. Documenting and Debunking Dubious Footage from Ukraine's Frontlines

24. Customary International Humanitarian Law. Cambridge University Press, 2005. P.25—36 <http://dx.doi.org/10.1017/cbo9780511804700.006>

25. Аналіз та обробка великих даних [Електронний ресурс] : методичні рекомендації до виконання практичних робіт для здобувачів ступеня магістра освітньо-професійної програми «Системний аналіз» зі спеціальності 124 Системний аналіз / М-во освіти і науки України, Нац. техн. ун-т «Дніпровська політехніка». – Дніпро : НТУ «ДП», 2024. – 82 с. <https://ir.nmu.org.ua/handle/123456789/167968>

26. Машинне навчання [Електронний ресурс] : методичні рекомендації до виконання практичних робіт для здобувачів ступеня магістра освітньо-професійної програми «Системний аналіз» зі спеціальності 124 Системний аналіз / уклад.: Т.А. Желдак, О.Б. Владико, А.В. Малієнко, Д.М. Гаранжа ; М-во освіти і науки України, Нац. техн. ун-т «Дніпровська політехніка». – Дніпро : НТУ «ДП», 2024. – 48 с. <https://ir.nmu.org.ua/handle/123456789/167920>

27. Практикум з диференційних рівнянь [Електронний ресурс] : навчальний посібник / Л.С. Коряшкіна, О.Д. Станіна, Ю.О. Шевченко; М-во освіти і науки України, Нац. техн. ун-т «Дніпровська політехніка» - Дніпро : НТУ «ДП», 2024 – 178 с. <https://ir.nmu.org.ua/handle/123456789/167658>

28. Аналіз та обробка великих даних [Електронний ресурс] : методичні рекомендації до виконання практичних робіт для здобувачів ступеня магістра освітньо-професійної програми «Системний аналіз» зі спеціальності 124 Системний аналіз / М-во освіти і науки України, Нац. техн. ун-т «Дніпровська політехніка». – Дніпро : НТУ «ДП», 2024. – 82 с. <https://ir.nmu.org.ua/handle/123456789/167968>

29. Шевченко, Ю. О. (2022). Обробка і аналіз даних з використанням електронних таблиць. Частина I «Обробка даних». <https://ir.nmu.org.ua/handle/123456789/162623>

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
«ДНІПРОВСЬКА ПОЛІТЕХНІКА»

Факультет інформаційних технологій
Кафедра системного аналізу та управління

ВІДГУК

Наукового керівника к.ф.- м.н., доцента Хом'як Тетяни Валеріївни

на кваліфікаційну роботу магістра

студента групи *124М-23-1* Осипенко О.М.

спеціальності *124 Системний аналіз*

Тема кваліфікаційної роботи: Аналіз шкоди цивільному населенню в Україні на основі інцидентів, зафіксованих у період російської агресії: дослідження на основі відкрити даних

Обсяг кваліфікаційної роботи: 61 стор.

Мета кваліфікаційної роботи: Здійснити комплексний аналіз шкоди цивільному населенню України в період російської агресії, використовуючи відкриті дані та сучасні аналітичні інструменти, для ідентифікації ключових закономірностей, оцінки ризиків та розробки практичних рекомендацій щодо моніторингу зон конфлікту.

Актуальність теми дослідження обумовлена значними наслідками збройного конфлікту, які включають руйнування інфраструктури, втрати серед мирного населення та гуманітарну кризу. Робота спрямована на вирішення задачі системного аналізу інцидентів, що дозволяє отримати достовірну картину ситуації за допомогою інноваційних технологій та методів обробки даних.

Тема кваліфікаційної роботи безпосередньо пов'язана з об'єктом діяльності магістра спеціальності 124 Системний аналіз, оскільки передбачає застосування статистичних методів, геопросторового аналізу, машинного навчання та OSINT для оцінки ризиків і розробки рекомендацій у сфері безпеки та гуманітарної допомоги.

Виконані в кваліфікаційній роботі завдання відповідають вимогам другого (магістерського) рівня вищої освіти. Оригінальність наукових рішень полягає в застосуванні комплексного підходу, який дозволяє інтегрувати відкриті дані, супутникові знімки та сучасні моделі аналізу для визначення ключових закономірностей уражень цивільного населення.

Практичне значення результатів кваліфікаційної роботи полягає в можливості їх застосування для покращення моніторингу зон конфлікту, планування гуманітарних операцій, оцінки ризиків та розробки стратегій мінімізації шкоди для цивільного населення. Результати роботи також можуть бути використані міжнародними організаціями та гуманітарними місіями для планування заходів із відновлення постраждалих територій.

Висновки підтверджують можливість використання результатів роботи в моніторингових системах, які забезпечують безпеку цивільного населення у зонах конфлікту, а також у плануванні гуманітарних операцій та заходів із мінімізації ризиків, подальших дослідженнях у сфері аналізу ризиків, розробки стратегій реагування на кризові ситуації та захисту цивільного населення. Матеріали роботи можуть бути використаними у розробці інтегрованих систем прогнозування, моделювання конфліктних ситуацій та створенні рекомендацій для органів державної влади та міжнародних партнерів.

Оформлення пояснювальної записки та демонстраційного матеріалу до неї виконано згідно з вимогами. Роботу виконано самостійно, відповідно до завдання та у повному обсязі.

У роботі відзначено такі недоліки: Робота базується на відкритих даних, які можуть містити прогалини або бути неповними, що може впливати на точність прогнозів. Хоча робота добре охоплює аналіз безпеки, соціально-економічні наслідки не отримали достатньої уваги.

Ці недоліки мають не знижують загальної наукової цінності роботи. Вони створюють перспективи для подальших досліджень і вдосконалення підходів.

Кваліфікаційна робота в цілому заслуговує оцінки: Відмінно

З урахуванням висловлених зауважень автор (не) заслуговує присвоєння кваліфікації «магістр з системного аналізу».

Науковий керівник к.ф.- м.н., доцент Хом'як Тетяна Валеріївна

«___» _____ 20__ р.

(підпис)