


Вінницький національний технічний університет  
Факультет інтелектуальних інформаційних технологій та автоматизації  
Кафедра системного аналізу та інформаційних технологій

**Комплексна магістерська кваліфікаційна робота на тему:**

**«ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ СТАНУ  
АТМОСФЕРНОГО ПОВІТРЯ МІСТА ВІННИЦІ. ЧАСТИНА 1.  
ФОРМУВАННЯ ДАТАСЕТУ ТА СИСТЕМНИЙ АНАЛІЗ ДАНИХ»**

Виконав: студент 2 курсу, групи 2ІСТ-22м  
спеціальності 126 – «Інформаційні  
системи та технології»

 Андрій ПІРОГОВСЬКИЙ

Керівник: к.т.н., доц. каф. САІТ

 Євгеній КРИЖАНОВСЬКИЙ

« 01 » 12 2023 р.

Рецензент: к.т.н., доц. каф. КН

 Володимир ОЗЕРАНСЬКИЙ

« 05 » 12 2023 р.

**Допущено до захисту**

Завідувач кафедри САІТ

 д.т.н., проф. Віталій МОКІН

« 01 » 12 2023 р.

Вінниця ВНТУ – 2023 рік

Вінницький національний технічний університет

Факультет інтелектуальних інформаційних технологій та автоматизації

Кафедра системного аналізу та інформаційних технологій

Рівень вищої освіти – II-й (магістерський)

Галузь знань – 12 Інформаційні технології

Спеціальність – 126 Інформаційні системи та технології

Освітньо-професійна програма – Інформаційні технології аналізу даних та зображень

ЗАТВЕРДЖУЮ

Завідувач кафедри САІТ

*Мокін* д.т.н., проф. Віталій МОКІН

«01» 09 2023 року

**ЗАВДАННЯ  
НА КОМПЛЕКСНУ МАГІСТЕРСЬКУ КВАЛІФІКАЦІЙНУ РОБОТУ  
СТУДЕНТУ**

Піроговському Андрію Віталійовичу

1. Тема роботи: «Інформаційна технологія аналізу та прогнозування стану атмосферного повітря міста Вінниці. Частина 1. Формування датасету та системний аналіз даних»,

керівник роботи: Євгеній КРИЖАНОВСЬКИЙ к.т.н., доц. каф. САІТ,

затверджені наказом закладу вищої освіти від «16» 09 2023 року № 247

2. Строк подання студентом роботи «30» 11 2023 року

3. Вихідні дані до роботи:



– Kaggle dataset “Air quality in Vinnytsia 2023 (September)”  
<https://www.kaggle.com/datasets/pirogovskiy/air-quality-in-vinnytsia-2023-september>;

– набір статистичних даних державного моніторингу атмосферного повітря в Вінницькій області за 2016-2021 роки.

4. Зміст текстової частини:







- загальна характеристика об’єкту досліджень;
- вибір оптимальних інформаційних технологій для здійснення системного аналізу та прогнозування даних;
- збирання та підготовка до системного аналізу даних забруднення атмосферного повітря міста Вінниці;
- системний аналіз та прогнозування даних;

- економічна частина.
5. Перелік ілюстративного матеріалу:
- тематична карта забруднення повітря пилом;
  - візуалізація значень забруднюючих показників;
  - графіки з аномальними даними забруднюючих показників;
  - графіки з прогнозом значень забруднюючих показників.
6. Консультанти розділів МКР

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
5	Наталія БУРЕННІКОВА д.е.н., проф. каф. ЕПВМ	 01.11.23	 10.11.23

7. Дата видачі завдання « 01 » 09 2023 року

### КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва та зміст етапу	Термін виконання		Примітка
		початок	закінчення	
1	Загальна характеристика об'єкту досліджень	04.09	20.09	
2	Вибір оптимальних інформаційних технологій для здійснення системного аналізу та прогнозування даних	20.09	10.10	
3	Збирання та підготовка до системного аналізу даних забруднення атмосферного повітря міста Вінниці	10.10	20.10	
4	Системний аналіз та прогнозування даних	20.10	01.11	
5	Економічна частина	01.11	10.11	
6	Оформлення матеріалів до захисту МКР	10.11	30.11	

Студент



Андрій ПІРОГОВСЬКИЙ

Керівник роботи



Євгеній КРИЖАНОВСЬКИЙ

## АНОТАЦІЯ

УДК 004.09

Піроговський А. В. Інформаційна технологія аналізу та прогнозування стану атмосферного повітря міста Вінниці. Частина 1. Формування датасету та системний аналіз даних. Комплексна магістерська кваліфікаційна робота зі спеціальності 126 – інформаційні системи та технології, освітньо-професійна програма – інформаційні технології аналізу даних та зображень. Вінниця: ВНТУ, 2023. 112 с.

На укр. мові. Бібліогр.: 30 назв; рис.: 86; табл.: 11.

В магістерській кваліфікаційній роботі звернено увагу на проблематику забруднення атмосферного повітря, яка є актуальною та важливою у сучасному світі. Досліджено методи та підходи до збору та обробки даних, що дозволяють виявити тенденції та прогнозувати зміни в якості повітря в майбутньому.

Запропоновані технології системного аналізу та прогнозування спрямовані на розв'язання проблем забруднення атмосферного повітря шляхом оперативного виявлення факторів, що впливають на його якість, та адаптації стратегій. Результати досліджень можуть бути використані для розробки більш ефективних програм зменшення забруднення та покращення екологічної стійкості міст. Об'єкт досліджень – процес створення інформаційної технології аналізу та прогнозування стану атмосферного повітря. Галузь застосування – станції моніторингу повітря та веб-сервіси, які спеціалізуються на моніторингу якості повітря.

Ілюстративна частина складається з 9 плакатів із результатами моделювання.

Ключові слова: системний аналіз, інформаційні технології, прогнозування, моніторинг, атмосферне повітря, машинне навчання, якість повітря.

## ABSTRACT

Pirohovskiy A. V. Information technology of analysis and forecasting of the state of atmospheric air in the city of Vinnytsia. Part I. Dataset formation and systematic data analysis. Comprehensive master's qualification thesis in specialty 126 – information systems and technologies, educational and professional program – information technologies of data and image analysis. Vinnytsia: VNTU, 2023. 112 p.

In Ukrainian speech. Bibliography.: 30 titles; fig.: 86; tab.: 11.

In the master's qualification work, attention is paid to the problem of atmospheric air pollution, which is relevant and important in the modern world. Methods and approaches to data collection and processing, allowing to identify trends and predict changes in air quality in the future, were studied.

The proposed technologies of system analysis and forecasting are aimed at solving the problems of atmospheric air pollution by promptly identifying factors affecting its quality and adapting strategies. Research results can be used to develop more effective programs to reduce pollution and improve the environmental sustainability of cities. The object of research is the process of creating information technology for analyzing and forecasting the state of atmospheric air. Field of application – air monitoring stations and web services specializing in air quality monitoring.

The illustrative part consists of 9 posters with simulation results.

Keywords: system analysis, information technologies, forecasting, monitoring, atmospheric air, machine learning, air quality

## ЗМІСТ

ВСТУП.....	4
1 ЗАГАЛЬНА ХАРАКТЕРИСТИКА ОБ’ЄКТУ ДОСЛІДЖЕНЬ .....	7
1.1 Характеристика проблематики забруднення атмосферного повітря.....	7
1.2 Стан моніторингу повітря у місті Вінниця.....	10
1.3 Аналіз поточного стану атмосферного повітря за даними державного моніторингу в місті Вінниця .....	14
1.4 Висновки .....	23
2 ВИБІР ОПТИМАЛЬНИХ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ДЛЯ ЗДІЙСНЕННЯ СИСТЕМНОГО АНАЛІЗУ .....	24
2.1 Вибір мови програмування та середовища розробки.....	24
2.2 Вибір бібліотек Python та їх опис .....	27
2.3 Висновки .....	28
3. ЗБИРАННЯ ТА ПІДГОТОВКА ДО СИСТЕМНОГО АНАЛІЗУ ДАНИХ ЗАБРУДНЕННЯ АТМОСФЕРНОГО ПОВІТРЯ МІСТА ВІННИЦЯ .....	30
3.1 Збирання та опис ознак вхідного набору даних.....	30
3.2 Завантаження датасету до системи Kaggle та підготовка до роботи.....	33
3.3 Висновки .....	36
4 СИСТЕМНИЙ АНАЛІЗ ТА ПРОГНОЗУВАННЯ ДАНИХ.....	38
4.1 Системний аналіз стану атмосферного повітря за даними громадського моніторингу в місті Вінниця.....	38
4.2 Прогнозування стану атмосферного повітря за даними громадського моніторингу в місті Вінниця .....	66
4.3 Висновки .....	76
5 ЕКОНОМІЧНА ЧАСТИНА.....	77
5.1 Проведення комерційного та технологічного аудиту науково-технічної розробки .....	77
5.2 Визначення рівня конкурентоспроможності розробки .....	82
5.3 Розрахунок витрат на проведення науково-дослідної роботи.....	85
5.3.1 Витрати на оплату праці.....	86
5.3.2 Відрахування на соціальні заходи .....	90
5.3.3 Сировина та матеріали.....	90
5.3.4 Розрахунок витрат на комплектуючі.....	92

5.3.5 Спецустаткування для наукових (експериментальних) робіт .....	92
5.3.6 Програмне забезпечення для наукових (експериментальних) робіт .....	93
5.3.7 Амортизація обладнання, програмних засобів та приміщень .....	95
5.3.8 Паливо та енергія для науково-виробничих цілей .....	96
5.3.9 Службові відрядження.....	98
5.3.10 Витрати на роботи, які виконують сторонні підприємства, установи і організації.....	99
5.3.11 Інші витрати.....	99
5.3.12 Накладні (загальновиробничі) витрати.....	100
5.4 Розрахунок економічної ефективності науково-технічної розробки при її можливій комерціалізації потенційним інвестором.....	101
5.5 Висновки .....	106
ВИСНОВКИ.....	107
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	110
Додаток А (обов'язковий). Технічне завдання.....	113
Додаток Б (обов'язковий). Протокол перевірки кваліфікаційної роботи на наявність текстових запозичень.....	116
Додаток В (Довідковий). Лістинг програми.....	117
Додаток Г (обов'язковий). Ілюстративна частина.....	125

## ВСТУП

**Актуальність теми.** Забруднення повітря є однією з найактуальніших екологічних проблем, яка виходить за межі простого впливу на природу. Це питання також має прямий і суттєвий вплив на здоров'я людей, визначаючи стан громадського здоров'я та якість життя. Останнім часом відзначається зростаюча свідомість щодо ролі технологій у редукції рівня забруднення повітря та забезпеченні сталих та здорових умов для життя нашого суспільства [1].

Забруднення повітря становить серйозну загрозу екосистемам та природним ресурсам, порушуючи рівновагу природного середовища. Однак його вплив на людське здоров'я надзвичайно важливий. Забруднене повітря містить токсичні речовини, які можуть викликати або погіршувати різноманітні захворювання, зокрема захворювання дихальних шляхів, серцево-судинні захворювання та інші проблеми зі здоров'ям. Велика кількість адаптованих до міського життя людей стикається з негативним впливом забрудненого повітря на їхню якість та тривалість життя.

Зростаюча увага до екологічних питань підкреслює необхідність розробки та впровадження новітніх технологій, спрямованих на зменшення викидів та покращення якості повітря в містах. Технологічні інновації, такі як використання відновлюваних джерел енергії, розробка ефективних систем фільтрації та програми стимулювання для переходу до екологічно чистих викидів транспорту, стають ключовими компонентами стратегії зменшення забруднення повітря та збереження здоров'я громадян.

Таким чином, ефективне вирішення проблеми забруднення повітря вимагає комплексного підходу, що поєднує наукові дослідження, розвиток технологій та активну участь громадян у створенні здорового та екологічно безпечного майбутнього.

**Мета і задачі дослідження.** Метою роботи є підвищення вивченості стану атмосферного повітря міста Вінниці шляхом проведення розвідувального та системного аналізу з використанням інформаційних технологій.



Для досягнення поставленої мети необхідно розв'язати такі задачі:

- провести аналіз проблеми стосовно забруднення атмосферного повітря та стану моніторингу повітря у місті;
- збирання та формування даних за інформацією громадського й державного моніторингу;
- здійснити системний аналіз та прогнозування даних, користуючись мовою програмування Python та іншими допоміжними інформаційними технологіями.

**Об'єктом дослідження** є процес аналізу та прогнозування стану атмосферного повітря міста Вінниці.

**Предметом дослідження** є методи та інформаційна технологія аналізу та прогнозування даних щодо якості атмосферного повітря міста Вінниці.

**Новизна одержаних результатів.** Дістав подальший розвиток підхід до аналізу та прогнозування стану атмосферного повітря міста Вінниці, за рахунок застосування сучасних методів та інформаційних технологій системного аналізу та моделей прогнозування даних.

**Практичне значення** роботи сприятиме покращенню аналізу якості моніторингу та прогнозу атмосферного повітря у місті Вінниця. Результати роботи мають велике значення для подальшого аналізу даних у сфері екології, а саме контролю за рівнем забруднення атмосферного повітря в місті.

**Апробація результатів магістерської кваліфікаційної роботи.** Результати роботи доповідались на LIII Науково-технічній конференції факультету інтелектуальних інформаційних технологій та автоматизації (м. Вінниця, 2023-2024 рр.).

**Публікації результатів магістерської кваліфікаційної роботи.** Згідно з результатами даної роботи було зроблено доповідь на тему «Просторовий аналіз даних забруднення атмосферного повітря Вінницької області за даними державного моніторингу у 2020 рік» на LI Науково-технічній конференції факультету інтелектуальних інформаційних технологій та автоматизації Вінницького національного технічного університету (2022) з публікацією тез [2].

Також, опубліковано тези на LIII Всеукраїнській науково-технічній конференції підрозділів Вінницького національного технічного університету факультету інтелектуальних інформаційних технологій та автоматизації (2023-2024) [3].

## **1 ЗАГАЛЬНА ХАРАКТЕРИСТИКА ОБ'ЄКТУ ДОСЛІДЖЕНЬ**

### **1.1 Характеристика проблематики забруднення атмосферного повітря**

Проблематика забруднення атмосферного повітря, на сьогоднішній день, є особливо важливою та складною проблемою, яка помітно впливає на наше навколишнє середовище, та задля вирішення якої потрібно неабияких зусиль, часу і коштів.

Як правило, найбільше забруднення повітря відбувається в великих містах, де досить великий рівень населення, відповідно значно більше автомобілів, промислових підприємств та інше, порівнюючи з менш населеними містами.

Для прикладу, наведемо декілька причин, чому більші міста мають вищий рівень забруднення повітря:

- утилізація відходів;
- густість населення;
- промислова діяльність;
- викиди транспортних засобів;
- житлове опалення та загальне використання енергії;
- обмежені зелені зони (сукупність різних деревних, трав'янистих рослин тощо);
- будівництво, демонтаж та інші процеси, які є чинниками здійснення високої концентрації пилу й інших шкідливих речовин.

Забруднене повітря впливає на здоров'я будь-якого живого організму, в особливості на людське здоров'я. Коли ми вдихаємо шкідливі речовини, вони можуть потрапити в нашу кров і спричинити кашель, свербіж в очах та ряд інших симптомів та захворювань, наприклад, захворювань органів дихання та легенів, що може привести до стимулювання клітин раку або навіть смерті [4].

Варто пам'ятати, що незалежно від того, де ви живете, будь-хто може бути підданий впливу забруднення повітря вихлопними газами транспортних засобів,

димом, викидами промислових підприємств, пилом, хімікатами та іншими речовинами, які ми використовуємо в нашому сьогодні.

Опираючись на дані звіту State of Global Air 2020, забруднення повітря посідає четверте місце за величиною фактором ризику передчасної смерті у світі – даний показник становить понад 6,67 мільйони смертей через забруднене повітря. Детальніше про глобальний рейтинг факторів ризику за загальною кількістю смертей в 2019 році можна розглянути на рисунку 1.1.

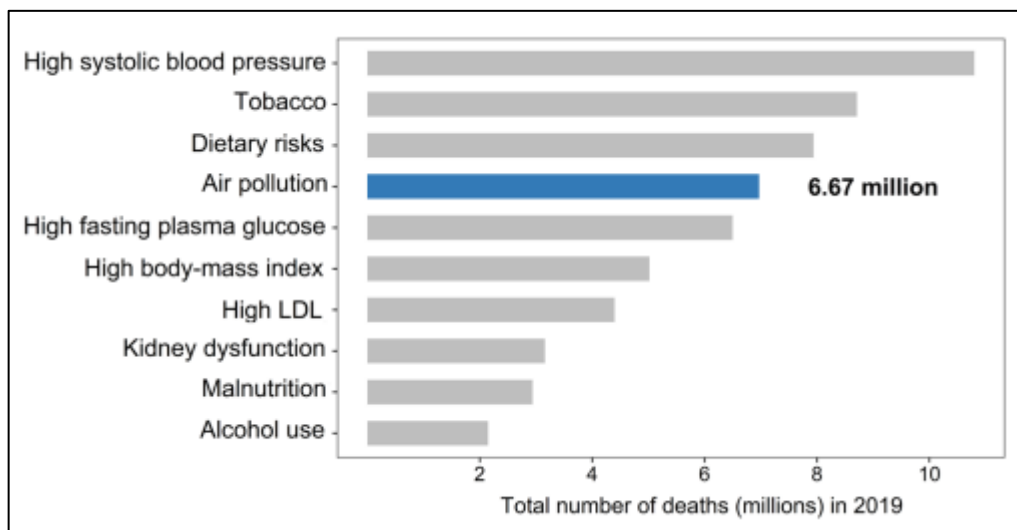


Рисунок 1.1 – Глобальний рейтинг факторів ризику за загальною кількістю смертей в 2019 р. згідно звіту State of Global Air 2020

Для підтвердження даних показників, на рисунку 1.2 продемонструємо карту з середнім рівнем показнику частинок пилу розміром 2,5 мікрметрів (PM<sub>2,5</sub>) згідно звіту з State of Global Air 2020 в різних частинах світу [5].

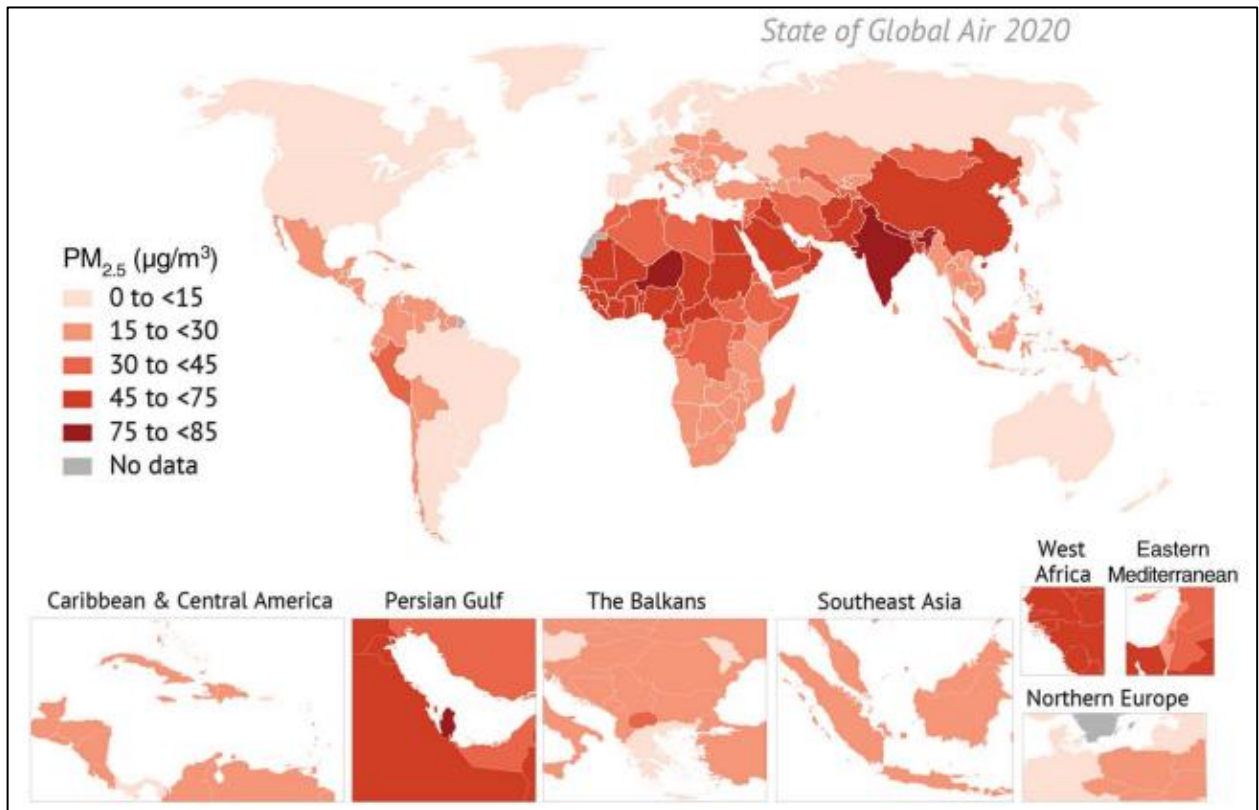


Рисунок 1.2 – Карта середнього рівня показнику частинок пилу розміром 2,5 мікрометрів (PM<sub>2,5</sub>) у різних країнах світу

На рисунку 1.2 зображено карту згідно звіту з State of Global Air 2020 стосовно середнього рівня вмісту в повітрі показнику PM<sub>2,5</sub>, згідно якої видно, що країни Азії, Африки та Близького Сходу продовжують переживати найвищі рівні концентрації частинок пилу розміром 2,5 мкм в атмосферному повітрі. Тобто, на основі даної статистики, можна навести топ-10 країн світу, які мають найбільший середній рівень частинок пилу розміром 2,5 мікрометрів в повітрі. На рисунку 1.3 можна побачити таблицю, яка відображає 10 найбільш забруднених країн показником PM<sub>2,5</sub>.

Country	PM <sub>2.5</sub> Concentration (µg/m <sup>3</sup> )	95% Uncertainty Intervals*
India	83.2	76.1 to 90.7
Nepal	83.1	62.9 to 107
Niger	80.1	42.2 to 145
Qatar	76.0	59.2 to 96.6
Nigeria	70.4	45.4 to 105
Egypt	67.9	47.8 to 92.8
Mauritania	66.8	37.6 to 108
Cameroon	64.5	43.8 to 92.6
Bangladesh	63.4	55.1 to 73.8
Pakistan	62.6	49.9 to 77.5

Рисунок 1.3 – Топ-10 країн світу з найвищим середнім рівнем показнику PM<sub>2.5</sub> в повітрі

На рисунку 1.3 зображена таблиця з 10-ма найбільш забрудненими країнами частинками пилу розміром 2,5 мікрометрів, де стовбець під назвою “PM<sub>2.5</sub> Concentration (µg/m<sup>3</sup>)” відображає середню концентрацію показнику частинок пилу розміром 2,5 мікрометрів на метр кубічний, а стовбець “95% Uncertainty Intervals” – діапазон значень від 2,5 до 97,5 перцентилів можливого розподілу значень, у межах якого, ймовірно, потраплятиме справжня концентрація.

Через це, згідно рисунків 1.1-1.3 видно, що проблема забруднення атмосферного повітря є надзвичайно вагомим фактором від якого залежить не лише здоров’я людини, а й стан усього нашого навколишнього середовища.

## 1.2 Стан моніторингу повітря у місті Вінниця

На сьогоднішній день, наше місто стикається з досить великою кількістю різних викликів різних масштабів, але не зважаючи на це, місто Вінниця, залишається одним із найбільш розвинених міст в Україні, особливо в сфері моніторингу навколишнього середовища, у нашому випадку, моніторингу

атмосферного повітря. Місто постійно оновлює існуючі станції для моніторингу повітря, будує нові та розширює їх загальні можливості за допомогою використання різних інформаційних технологій, зокрема технологій для якісного аналізу даних, які надаються безпосередньо від самих моніторингових станцій.

В самому місті Вінниці, станції моніторингу встановлено рядом груп мешканців міста, незалежних організацій тощо, наприклад, такими як [6]:

- Вінницька міська рада;
- ЛУН Місто;
- Eco City;
- SaveDnipro.

Також, існують сервіси, за допомогою яких можна здійснювати моніторинг щодо рівня забруднення повітря міста онлайн, наприклад, як в SaveEcoBot, чи Eco City (рис. 1.4-1.5).

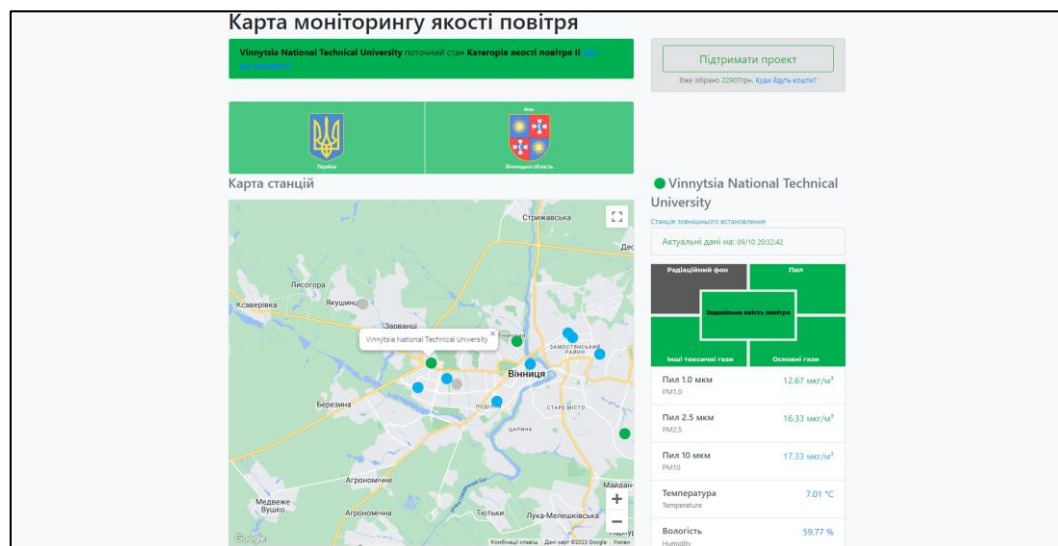


Рисунок 1.4 – Сайт для здійснення моніторингу повітря Eco City

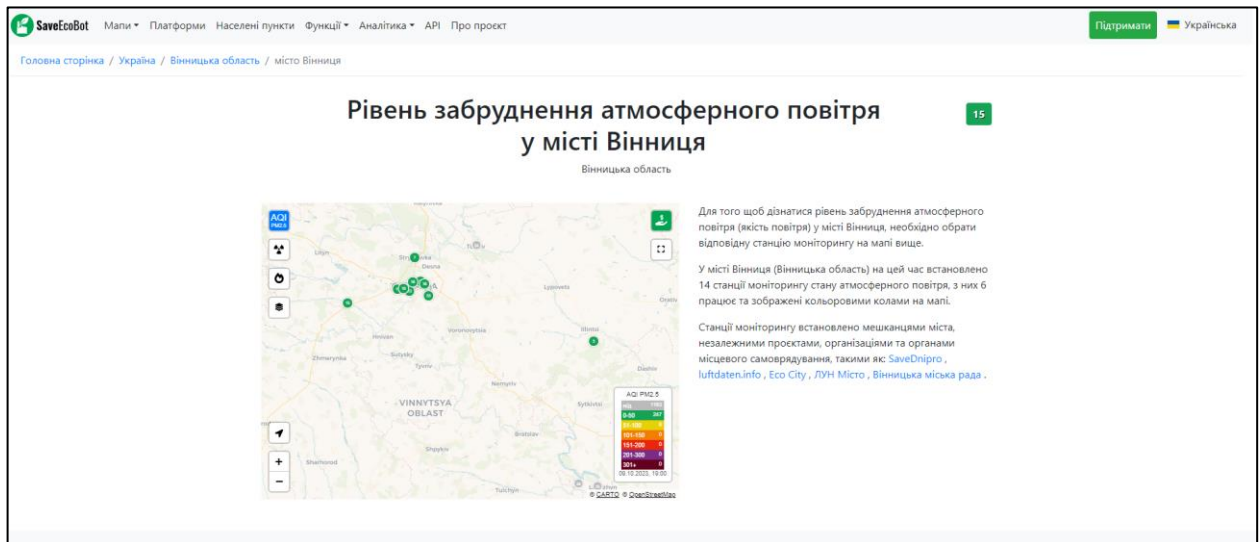


Рисунок 1.5 – Онлайн-сервіс з моніторингу якості атмосферного повітря SaveEcoBot

Через динамічну зміну даних у вище згаданих онлайн-сервісів для моніторингу повітря, можна досить чітко відслідковувати будь-які зміни у повітрі, особливо концентрації таких речовин, як: частинки пилу розміром 1 мкм –  $PM_1$ , частинки пилу розміром 2,5 мкм –  $PM_{2,5}$  та частинки пилу розміром 10 мкм –  $PM_{10}$ . Також деякі сервіси можуть надавати інформацію стосовно різних газів, наприклад: діоксид сірки ( $SO_2$ ), діоксид азоту ( $NO_2$ ), оксид вуглецю (CO) та озон ( $O_3$ ).

Повертаючись до моніторингу даних за допомогою онлайн-сервісів, важливо згадати про наявність такої функції, як відображення індексу якості повітря (AQI).

Air Quality Index (AQI) — це шкала, яка демонструє наскільки забруднене повітря в області, що вимірюється. Він був встановлений Агентством з охорони навколишнього середовища США (EPA), як стандартизований спосіб інформування населення про якість повітря. AQI вперше був представлений у 1968 році, щоб забезпечити простий і послідовний метод звітування населення про рівні якості повітря щодня. Згодом, кількість країн, які почали використовувати індекс якості повітря, запровадженого науковцями з США, сильно збільшувалося [7]. Для прикладу, продемонструємо графік з



відображенням індексу якості повітря в реальному часі, який було взято з вище згаданого сервісу SaveEcoBot (рис. 1.6)

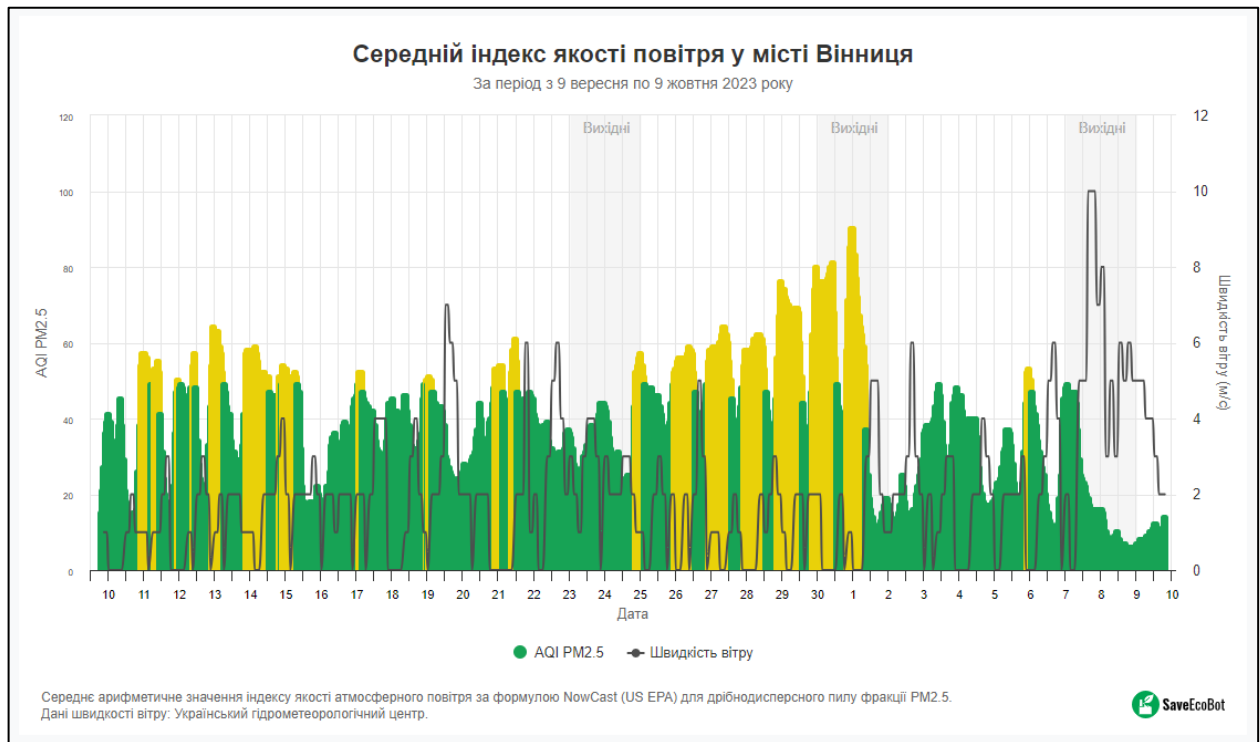


Рисунок 1.6 – Графік індексу якості повітря (AQI) в м. Вінниця

На рисунку 1.6 можна розглядити AQI-графік, який відображує актуальну інформацію стосовно стану якості повітря в місті Вінниця, у нашому випадку за період з 9 вересня по 9 жовтня 2023 року.

Значення кожної речовини, яка забруднює повітря та впливає на організм людини, AQI перетворює значення цих концентрацій в одне число, яке згодом підлягає до певної категорії, де в свою чергу категорії, безпосередньо показують рівень впливу забрудненого повітря на здоров'я людини.

Зазвичай значення індексу якості повітря коливається від 0 до 500 (чим менше тим краще), а самі категорії відділяють різними нормативними порогами (діапазонами) й кольорами. На рисунку 1.7 наведено приклад категорій AQI [8].

Категорія AQI (Діапазон)	PM <sub>10</sub> (24год)	PM <sub>2.5</sub> (24год)	NO <sub>2</sub> (24год)	O <sub>3</sub> (8год)	CO (8год)	SO <sub>2</sub> (24год)	NH <sub>3</sub> (24год)	Pb (24год)
Хороший (0-50)	0-50	0-30	0-40	0-50	0-1.0	0-40	0-200	0-0.5
Задовільний (51-100)	51-100	31-60	41-80	51-100	1.1-2.0	41-80	201-400	0.5-1.0
Помірно забруднений (101-200)	101-250	61-90	81-180	101-168	2.1-10	81-380	401-800	1.1-2.0
Високий (201-300)	251-350	91-120	181-280	169-208	10-17	381-800	801-1200	2.1-3.0
Дуже високий (301-400)	351-430	121-250	281-400	209-748	17-34	801-1600	1200-1800	3.1-3.5
Небезпечний (401-500)	430+	250+	400+	748+	34+	1600+	1800+	3.5+

Рисунок 1.7 – Категорії індексу якості повітря

Дивлячись на рисунок 1.7 видно, що в залежності від значення індексу, виділяється 6 категорій, де кожна з них має свій діапазон допустимих значень. Тобто, якщо діапазон якості повітря у місті коливається від 0 до 50, це означає, що поточна якість повітря ніяк негативно не впливає на людський організм, й не заважає в побутовому житті, але, наприклад, при діапазоні значень від 201 до 300, якість повітря суттєво погіршується та може призвести до дискомфорту в диханні, а також значне погіршення самопочуття у людей з захворюваннями серця.

### 1.3 Аналіз поточного стану атмосферного повітря за даними державного моніторингу в місті Вінниця

Станом на сьогодні, місто Вінниця визначається як одне з найбільш розвинених міст в Україні. Завдяки цьому статусу, Вінниця володіє добре розвиненими цифровими та моніторинговими системами, зокрема системами моніторингу якості повітря. Це надає можливість відстежувати та досліджувати поточний стан атмосферного повітря у місті й не тільки.

Поточний аналіз стане важливим етапом, оскільки він надасть можливість отримати комплексне уявлення про сучасний стан атмосферного повітря в місті. Крім того, цей аналіз служитиме основою для подальшого порівняння отриманих результатів нашого дослідження. Такий порівняльний аналіз дозволить більш якісно проаналізувати тенденції концентрацій різних показників у повітрі та

відобразити їх відмінності в якості повітря між декількома різними періодами: 2015-2023 роками.

Порівнюючи дані з року в рік, з'являється можливість не лише спостерігати зміни у рівні забруднення повітря, але й аналізувати їхні можливі причини та тенденції. Цей підхід дозволить нам краще зрозуміти динаміку забруднення повітря та його вплив на здоров'я мешканців, враховуючи різні фактори, такі як промисловість, транспорт та інші джерела забруднення.

Згідно інформації програми державного моніторингу у галузі охорони атмосферного повітря агломерації «Вінниця» на 2021-2025 роки, відомо, що за останні 5 років, кількість хворих на хвороби органів дихання коливалась в районі від 24-41% від усієї кількості населення агломерації, однак, за останні два роки спостерігається тенденція зниження даного відсотку до 24% [9].

Щоб розпочати досліджувати забруднюючі повітря показники, спершу розглянемо наявність промислових підприємств в м. Вінниця, які можуть слугувати стаціонарними джерелами забруднення повітря. Перелік даних підприємств можна розглядати на рисунку 1.8.

№ п/п	Назва підприємства	Фактичне місцезнаходження, адреса	Код ЄРДПОУ	Загальний обсяг викидів забруднювальних речовин, т				
				2016	2017	2018	2019	2020
1	ПрАТ «Вінницький олійножировий комбінат»	м. Вінниця, вул. Немпівське шосе, 26	00373758	998,3	904,6	928,8	976,4	787,9
2	КП «Вінницяоблтеплоенерго»	м. Вінниця, вул. Вишнева, 24	33649363	233,3	282,9	82,5	93,7	88,6
3	ПрАТ «Будматеріали»	м. Вінниця, вул. Данила Нечая, 56	06959287	125,7	137,3	120,4	128,2	170,8
4	КП ВМР «Вінницяміськтеплоенерго»	м. Вінниця, вул. 600-річчя, 13	33126849	657,2	630,9	519,4	679,7	477,3
5	ТОВ «Барлінек інвест»	м. Вінниця, вул. Чехова, 7 В	34004579	118,1	116,8	91,8	138,1	222,5
6	ТОВ «АВІС»	м. Вінниця, вул. Пирогова, 150	13304871	177,5	200,0	37,4	16,2	13,6
7	ПрАТ «Вінницька кондитерська фабрика»	м. Вінниця, вул. Гліба Успенського, 9	00382013	55,4	40,7	55,5	69,0	81,9
8	КП «Вінницяоблводоканал»	м. Вінниця, вул. Київська, 173	03339012	23,1	33,1	63,6	65,7	69,4

Рисунок 1.8 – Список основних підприємств що забруднюють повітря

Державний моніторинг повітря агломерації “Вінниця” здійснюється лабораторією спостережень за забрудненням атмосфери (ЛСЗА) Вінницького обласного ЦГМ на двох стаціонарних постах (ПЗС), де перший пост розташований по вулиці Київська, 25, а другий – на вулиці Немирівське шосе, 29. Дані пости спостережень здатні вимірювати наступні показники:

- аміак;
- формальдегід;
- пил;
- діоксид сірки;
- діоксид азоту;
- оксид азоту;
- оксид вуглецю;
- фтористий водень.

Переходячи до аналізу показників, продемонструємо на основі отриманих даних графіки, які відобразять середньомісячну концентрацію показників, які були зазначені вище, та їх гранично-допустимі значення (ГДК) протягом 2018-2021 років (рис. 1.9-1.15).

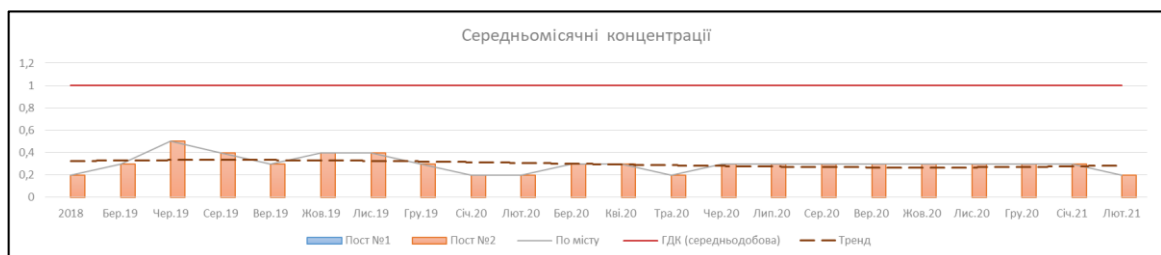


Рисунок 1.9 – Графік середньомісячних концентрацій аміаку

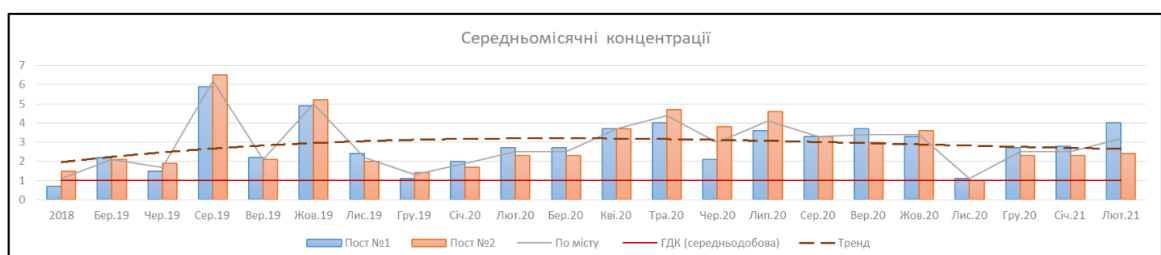


Рисунок 1.10 – Графік середньомісячних концентрацій діоксиду азоту



Рисунок 1.11 – Графік середньомісячних концентрацій діоксиду сірки

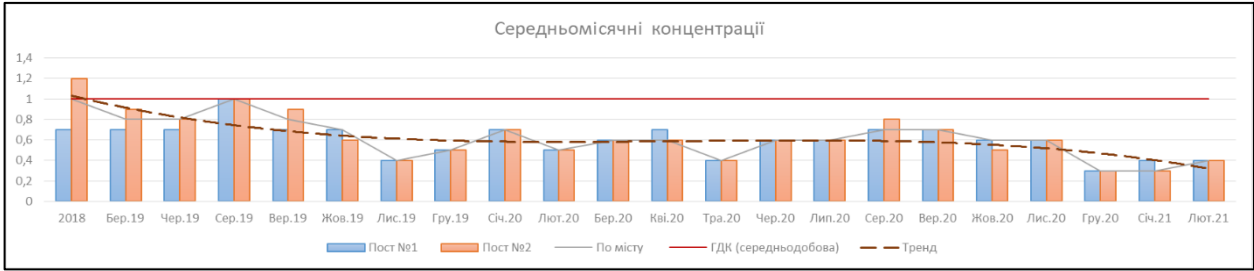


Рисунок 1.12 – Графік середньомісячних концентрацій пилу

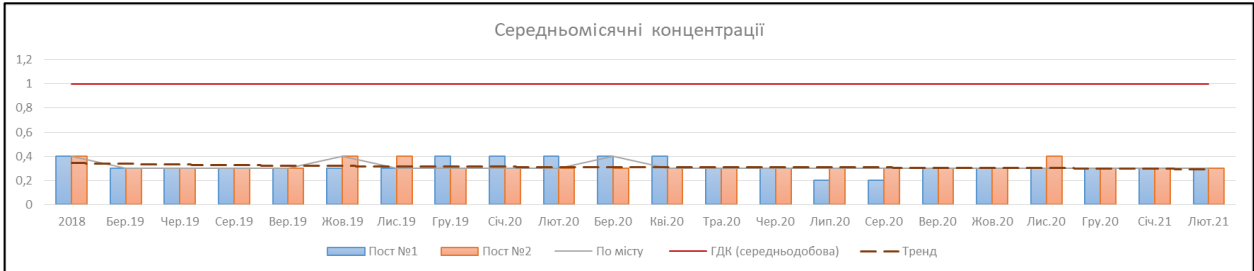


Рисунок 1.13 – Графік середньомісячних концентрацій оксиду вуглецю

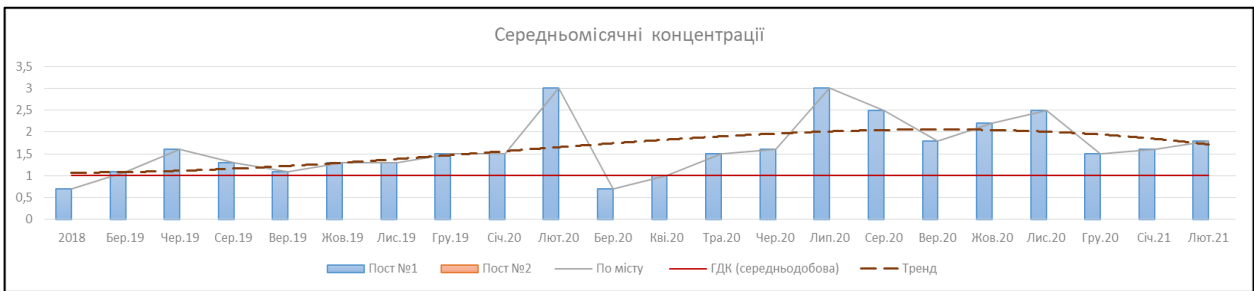


Рисунок 1.14 – Графік середньомісячних концентрацій формальдегіду

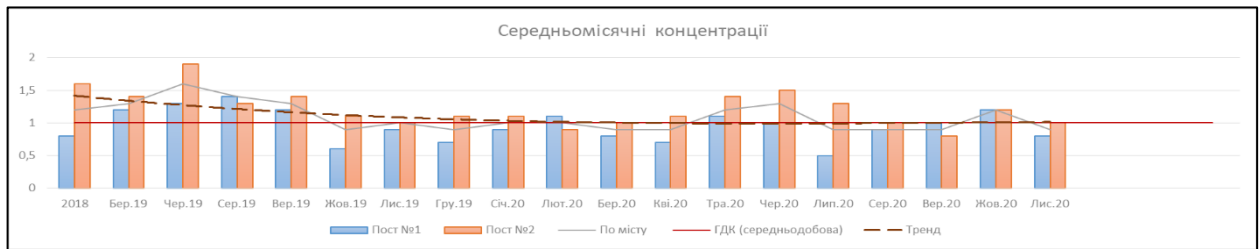


Рисунок 1.15 – Графік середньомісячних концентрацій фтористого водню

Отже, розглядаючи графіки, які зображені на рисунках 1.9-1.15, можна побачити загальну якість повітря за період 2018-2021 років. Більшість показників, а саме: оксид вуглецю, пил, діоксид сірки, аміак, мають стабільні значення та, що найголовніше, не перевищують гранично-допустимі концентрації (ГДК). Щодо решти показників, ситуація інакша, адже, наприклад, показник діоксид азоту має постійне перевищення гранично-допустимої норми, що потребує більш ретельної уваги в подальших дослідженнях. Беручи до уваги значення показників формальдегіду – вони теж мають стабільне перевищення норми, але повертаючись до аналізу цього показнику, який був здійснений за період вересня місяця 2023 року, можна побачити, що загальна кількість викидів даного чинника зменшилася і, на даний момент, не перевищує свої нормовані значення зрівняно з іншими показниками, наприклад, діоксиду азоту чи пилу.

Також продемонструємо вищевказані концентрації у вигляді побудованих тематичних карт, що дозволить візуально краще відобразити різницю в концентраціях за роки (рис. 1.16-1.19)

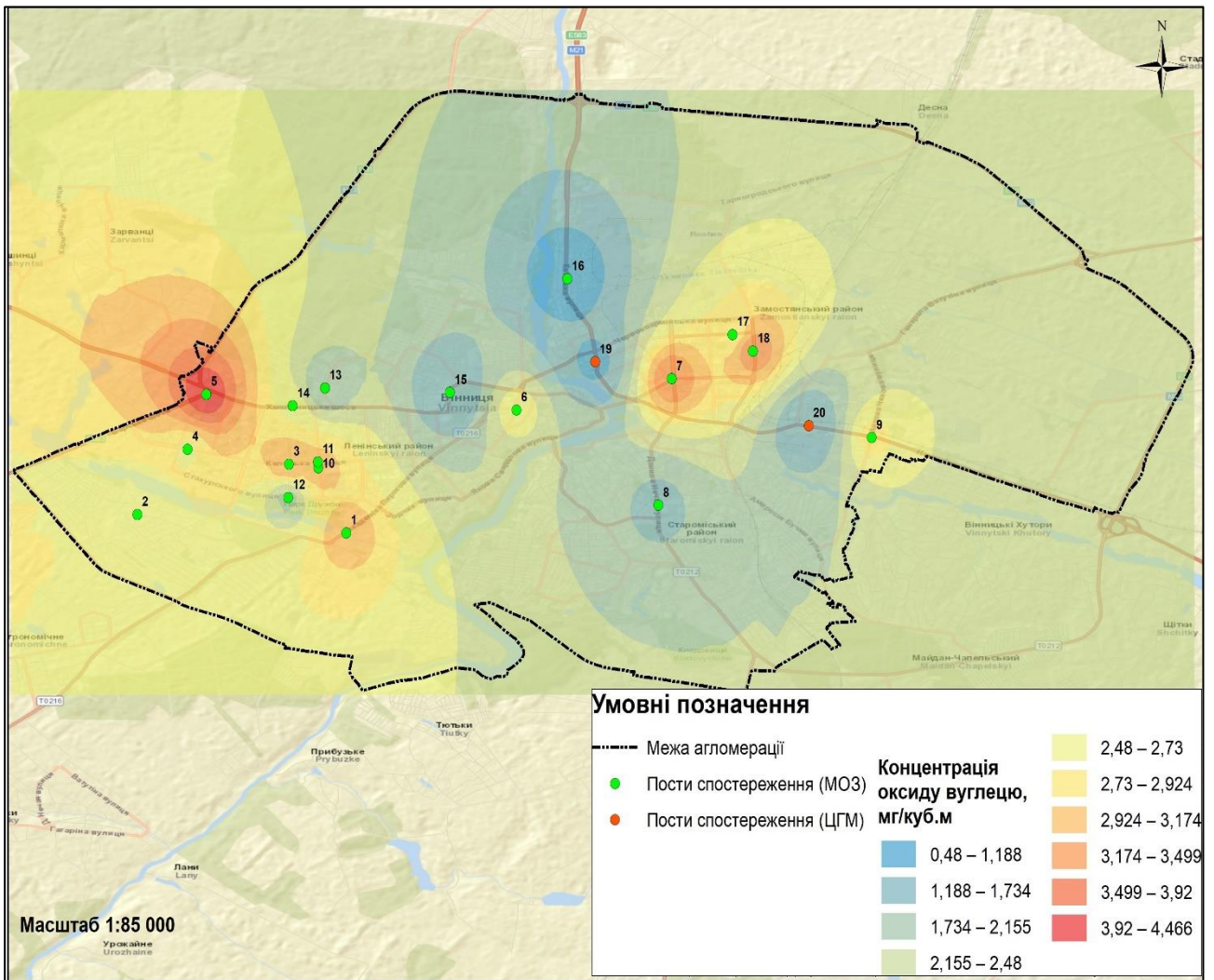


Рисунок 1.16 – Карта просторового розподілу концентрації оксиду вуглецю

На рисунку 1.16, зображено карту просторового розподілу вмісту оксиду вуглецю в агломерації Вінниця. На ньому відзначається високий рівень забруднення повітря оксидом вуглецю, зокрема в західній частині міста, де спостерігається певне погіршення якості повітря.

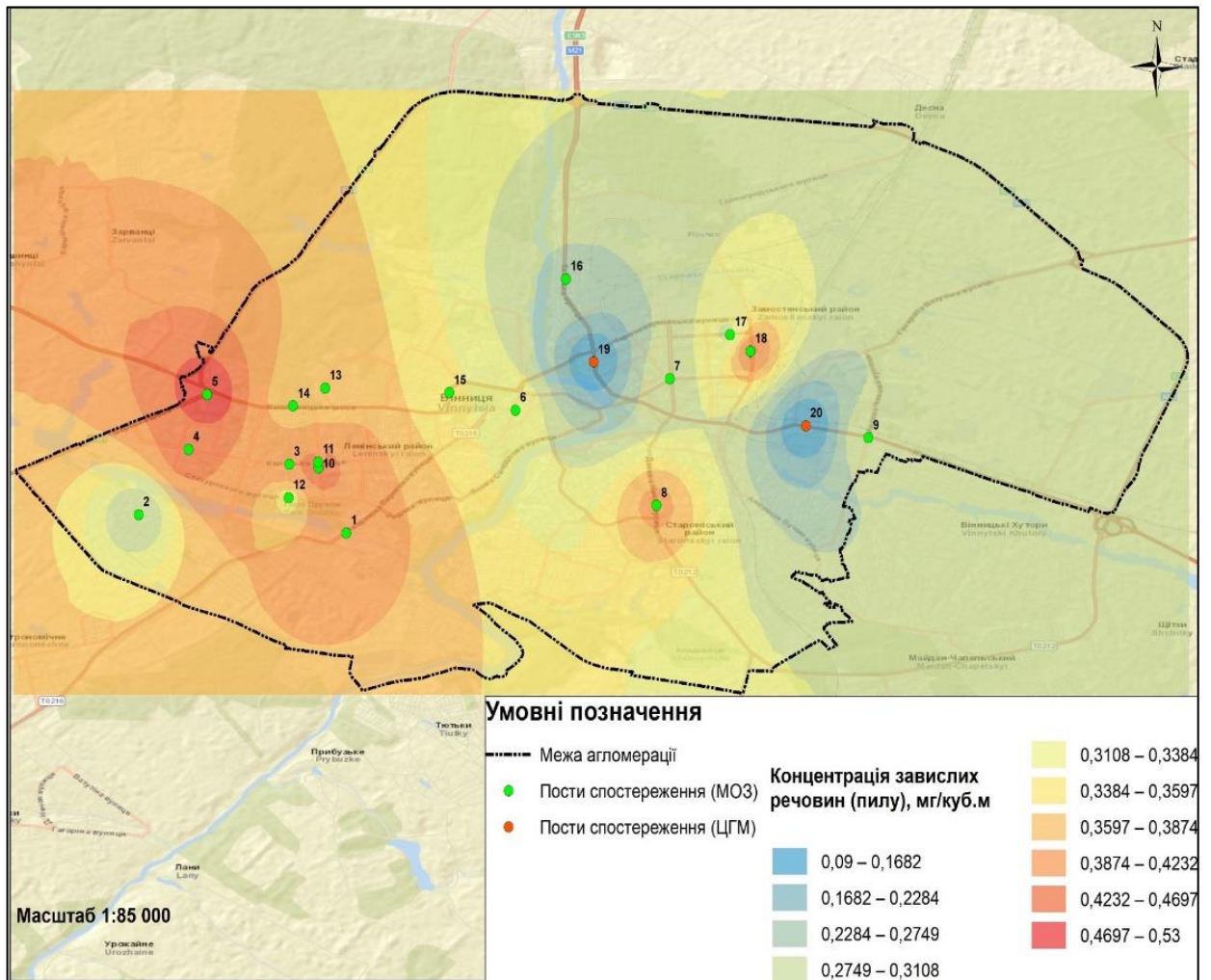


Рисунок 1.17 – Карта просторового розподілу концентрації пилу

На рисунку 1.17, зображено аналогічну карту просторового розподілу, але за вмістом пилу в агломерації Вінниця. Згідно даного рисунку, можна спостерігати схожу ситуацію, а саме погіршення якості повітря в західній частині міста.



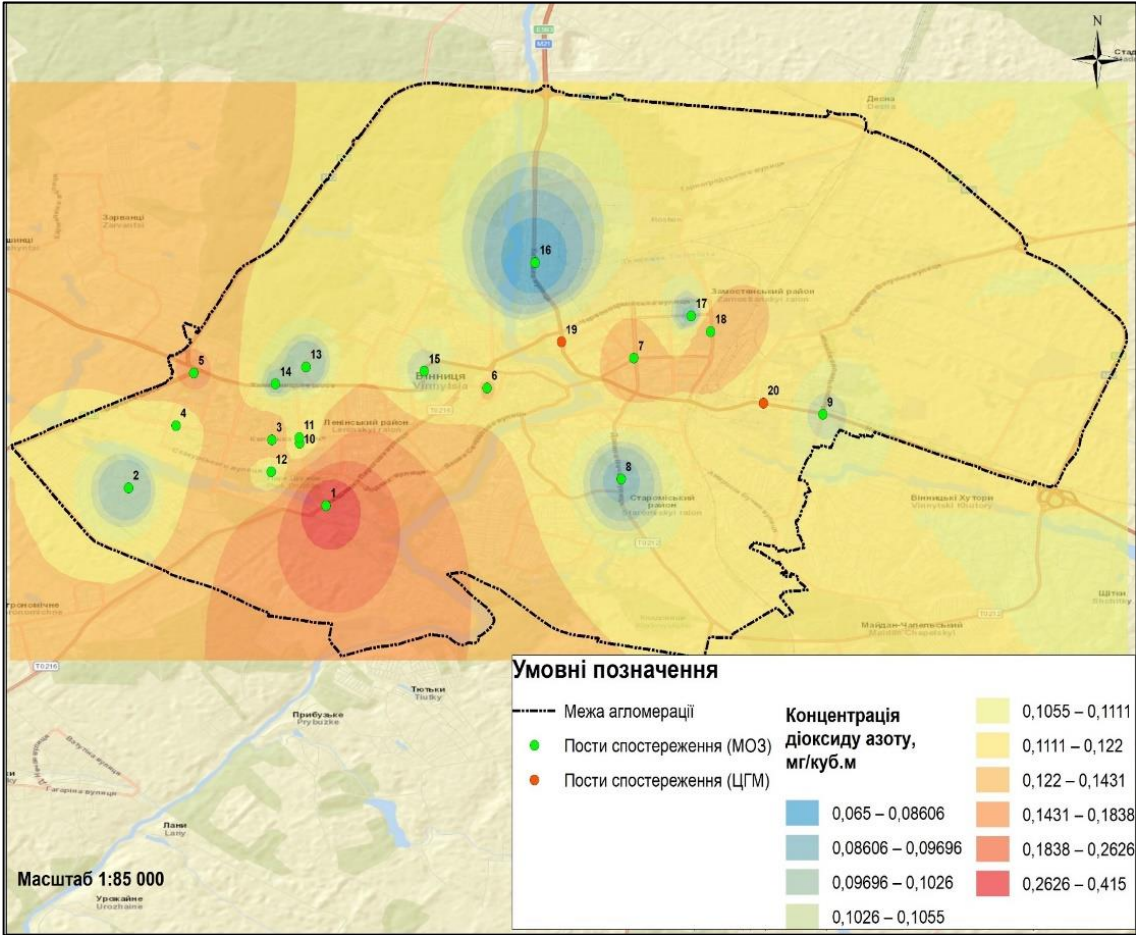


Рисунок 1.18 – Карта просторового розподілу діоксиду азоту

На рисунку 1.18, зображено карту просторового розподілу, проте вона відображає концентрацію діоксиду азоту в алгомерації Вінниця. На відмінно від попередній рисунків, ситуація з діоксидом азоту відрізняється певними змінами, зокрема, що найвищий рівень забруднення спостерігається у південній частині міста.

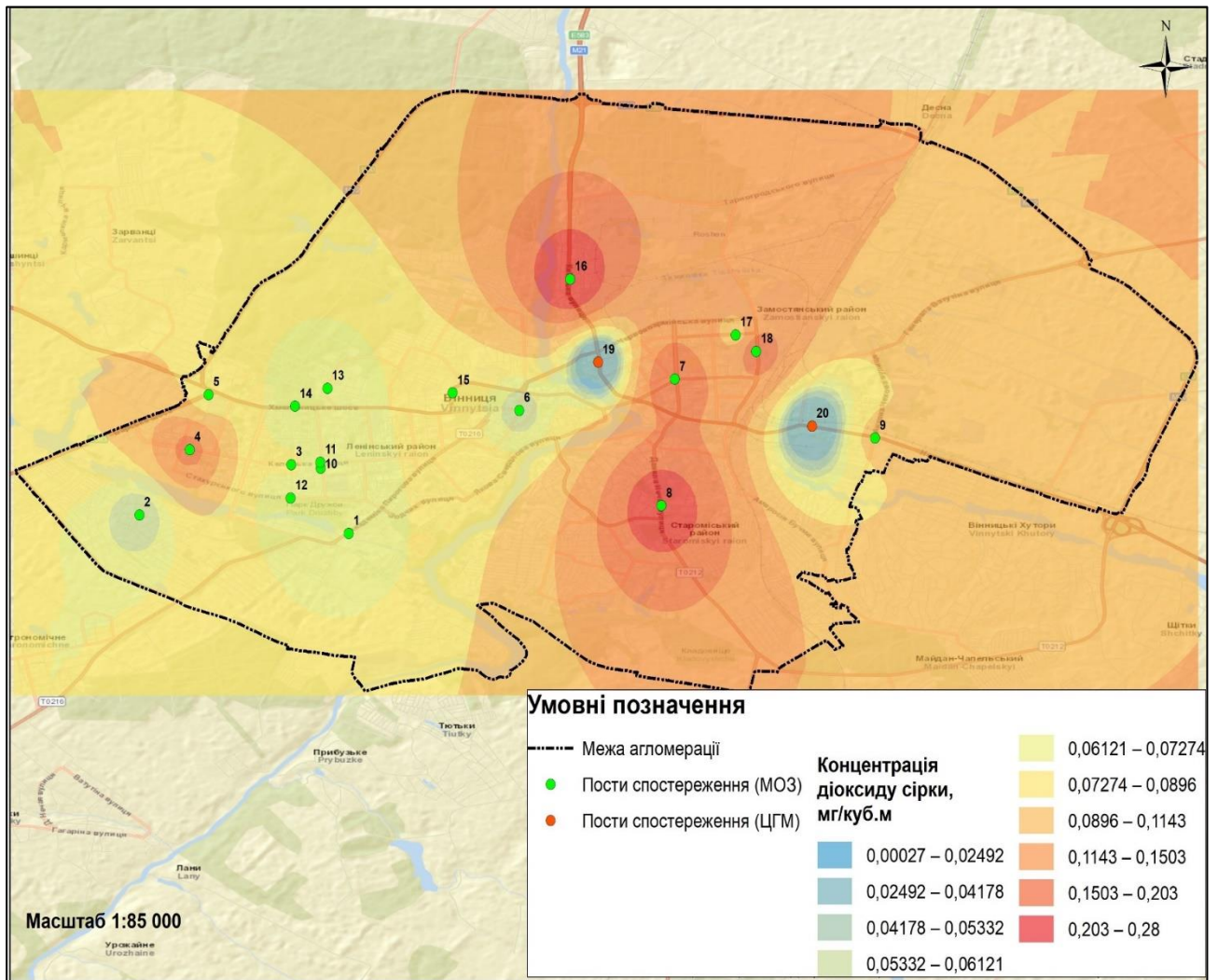


Рисунок 1.19 – Карта просторового розподілу концентрації діоксиду сірки

На рисунку 1.19 представлена карта просторового розподілу за показником діоксиду сірки в місті Вінниця. Розглядаючи цей рисунок, можна визначити, що найвищий рівень забруднення цим показником спостерігається у східній частині міста.

Тобто, порівнюючи дані за вказані періоди, можна спостерігати позитивні зміни в атмосфері повітря, адже станом на 2023 рік, якість повітря хоча і не значно, але краща ніж за минулі роки. На це вплинуло досить багато різних факторів, починаючи від збільшення попиту електрокарів та закінчуючи на побудові нових підприємств, які здійснюють мінімальну кількість викидів в атмосферу базуючись на сучасних вимогах щодо “Зеленого курсу”.

## **1.4 Висновки**

В даному розділі надано загальний огляд об'єкту досліджень, описано проблему забруднення повітря та ретельно проаналізовано моніторинг його стану у м. Вінниця.

Також проведено короткий аналіз поточного стану атмосферного повітря за даними державного моніторингу в місті Вінниця. Це допомогло визначити важливість вирішення цієї проблеми та посприяло більш глибокому вивченню обраної теми для її повного розуміння.

Під час написання розділу, отримано знання та дані, які будуть корисні для подальшої розробки нових та ефективних заходів, спрямованих на поліпшення якості повітря в нашому оточенні у майбутньому.

## **2 ВИБІР ОПТИМАЛЬНИХ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ДЛЯ ЗДІЙСНЕННЯ СИСТЕМНОГО АНАЛІЗУ**

### **2.1 Вибір мови програмування та середовища розробки**

Інформаційні технології керують більшою частиною того, що ми робимо в особистому та професійному житті. Це основа нашого спілкування, технологічного прогресу, інновацій, сталого розвитку та відпочинку. Ми використовуємо інформаційні технології на особистому рівні, щоб спілкуватися з іншими, ділитися медіафайлами та навіть просто робити покупки, чи розважуватися [10].

З точки зору кар'єри, інформаційні технології значною мірою відповідають за більшість наших операцій і охоплюють майже всі галузі. Від охорони здоров'я до громадського харчування, виробництва та продажів тощо, ми покладаємося на ІТ, які допомагають зв'язувати нас з іншими незалежно від відстані, зберігати та керувати майже будь-якою інформацією та створювати автоматизовані системи, тим самим роблячи бізнес-процеси більш ефективнішими.

На даний момент існує безліч різних інформаційних технологій, адже їх на стільки багато, що відома фраза “На будь-який смак” найбільше підходить до процесу вибору інформаційних технологій, які необхідні для вирішування поставлених задач.

В нашому випадку, а саме для вирішення задачі стосовно аналізу забруднення атмосферного повітря міста Вінниці, було обрано мову програмування Python.

Python — це інтерпретована об'єктно-орієнтована мова програмування високого рівня з динамічною семантикою, яка була розроблена Гвідо ван Россумом. Дана мова програмування має репутацію мови, зручної для початківців, замінюючи Java як найпоширенішу початкову мову, оскільки вона справляється з більшою частиною завдань і водночас має простіший синтаксис, що дозволяє початківцям більше зосередитися на вивченні самої мови та

розробці початкових проєктів без вдавання в зайві деталі на ранніх стадіях програмування [11].

Особливості та переваги мови Python:

- сумісність з різними платформами, наприклад: Microsoft Windows, Mac, Linux та інші;
- використовує легкий для розуміння синтаксис, який постійно можна оптимізувати та скорочувати;
- може обробляти код, процедурним, функціональним або об'єктно-орієнтованим способами;
- гнучкість мови;
- наявність великої кількості бібліотек, за допомогою яких можна вирішувати досить широкий спектр задач, наприклад, прогнозування даних або створення штучного інтелекту та його навчання.

Велика функціональність та призначення цієї мови, робить її, мабуть, чи не однією із найбільш універсальних мов в світі. Python використовують для веб-розробки, а саме back-end частини, розробки програмного забезпечення, в особливості розробки швидких додатків. Завдяки легкому освоєнню та простому синтаксису, втрати на технічне обслуговування програми написаної на мові Python значно зменшується.

Не зважаючи на усі позитивні аспекти даної мови програмування, вона, як і будь-які інші, має свої недоліки, а саме:

- повільно працює;
- Python вважається менш придатним до створення мобільних додатків;
- велике споживання пам'яті та інших технічних складових комп'ютера в цілому.

Тобто, зваживши усі “за” та “проти”, можна з впевненістю сказати, що мова програмування Python чудово підходить для вирішення нашого завдання – здійснення системного аналізу даних забруднення повітря в м. Вінниця.

Визначившись з мовою програмування, варто не забувати ще один невід’ємний аспект, який необхідний під час реалізації будь-якої інформаційної технологій, а саме вибір середовища розробки.

Середовище розробки — це так звані програмні додатки, призначені для полегшення програмування та розробки програмного забезпечення. Ці середовища надають великий набір інструментів і функцій, які допомагають розробникам при створенні програмних продуктів [12].

На даний момент, існує багато середовищ розробки, в особливості для обраної мови програмування Python. Ось деякі з них:

- Visual Studio Code;
- PyCharm;
- Jupyter Notebook;
- Atom;
- Spyder;
- Google Colab;
- Kite та інші.

Опираючись на сьогоднішні виклики з якими стикається наша країна та загальні сучасні тенденції, буде доречно використовувати онлайн-середовище розробки.

В нашому випадку, маючи досвід використання онлайн-середовищ розробки, було обрано середовище від системи Kaggle.

Kaggle Notebooks — це онлайн-середовище розробки, яке інтегроване в популярній системі Kaggle. Ноутбуки у Kaggle — це інтерактивні веб-обчислювальні середовища, які дозволяють користувачам писати, редагувати та виконувати код для аналізу даних, їх прогнозування тощо [13].

Можна визначити деякі ключові ознаки використання середовища Kaggle Notebooks:

- інтеграція з Jupyter Notebook;
- наявність безкоштовних обчислювальних ресурсів;
- інтеграція різних наборів даних;

- наявність історії версій, що дуже зручно в випадку, коли потрібно швидко повернутися до попередніх версій своєї роботи;
- велика спільнота, яка активно бере участь в обговоренні різних тем, проблем тощо.

Через те, що це онлайн-середовище розробки, воно має деякі свої обмеження стосовно технічних характеристик та ресурсів, що надаються.

Ресурси, які надає система Kaggle:

- максимальне виконання сеансів – 12 годин;
- 20 ГБ автоматичного збереження на диску (в директорії “/Kaggle/working”);
- додатковий простір на диску, але поза межею директорії “/Kaggle/working”.

Технічні характеристики центрального процесора (ЦП):

- 4 ядра ЦП;
- 30 ГБ оперативної пам’яті.

Після завершення оцінки різних середовищ розробки, в кінцевому результаті, було вибрано онлайн-середовище Kaggle Notebooks. Воно ідеально підходить для вже обраної мови програмування Python.

## 2.2 Вибір бібліотек Python та їх опис

Мова програмування Python має величезну кількість допоміжних бібліотек, які містять в собі різні функції, класи, значення тощо.

Для виконання поставленого завдання, а саме системного аналізу даних забруднення повітря у м. Вінниця, в роботі будуть використовуватися наступні бібліотеки:

NumPy — це базовий пакет для наукових обчислень за допомогою мови програмування Python [14].

Pandas — потужна та проста у використанні бібліотека, яка призначена для аналізу, обробки даних [15].

Matplotlib — це бібліотека, яка призначена для побудови 2D-графіків. Її можна використовувати для створення графіків, гістограм, кругових діаграм, діаграм розсіювання тощо [16].

Seaborn — ще одна бібліотека в Python, призначена для візуалізації статистичних даних на основі Matplotlib. Вона забезпечує інтерфейс високого рівня для створення привабливої та інформативної графіки [17].

SciPy — бібліотека, яка базується на NumPy і надає велику кількість функцій, які працюють з масивами NumPy [18].

Scikit-Learn — ефективний інструмент для інтелектуального аналізу даних. Базуючись на основі NumPy, SciPy та Matplotlib, даний інструмент надає прості та ефективні функції для інтелектуального аналізу даних. Бібліотека Scikit-Learn являється одною із найбільш популярних бібліотек для реалізації машинного навчання [19].

StatsModel — бібліотека в Python, яка дозволяє досліджувати дані, оцінювати статистичні моделі та виконувати статистичні тести. Вона доповнює модуль SciPy [20].

Folium — ще одна бібліотека, яка доступна на мові програмування Python. Вона дозволяє легко створювати інтерактивну візуалізацію в вигляді карти. Базуючись на іншій бібліотеці Leaflet.js, створює веб-карти у вигляді html файлу [21].

Усі вище зазначені бібліотеки активно використовують в системному аналізі даних задля вирішення багатьох завдань, включаючи перевірку та очищення даних, їх дослідження та візуалізація.

## 2.3 Висновки

У даному розділі проведено аналіз та вибір мови програмування й середовища розробки для виконання подальших поставлених завдань. Також здійснено огляд вже існуючих бібліотек, зокрема тих, які мають важливе значення для системного аналізу даних забруднення повітря в місті Вінниця.



Для нашого дослідження була обрана мова програмування Python через її потужність, гнучкість та велику кількість явних бібліотек для аналізу даних. Також, у зв'язку з цим, вирішено використовувати онлайн-середовище розробки Kaggle Notebooks. Вона надає можливість спільно працювати над проєктами, надаючи зручність та легкість у розробці та виборі найбільш оптимальних бібліотек для власних потреб.

Щодо бібліотек, які будуть використовуватися, привернуто уваго на такі, як Pandas для обробки та аналізу даних, NumPy для роботи з числовими масивами, Matplotlib і Seaborn для візуалізації результатів, а також інші спеціалізовані бібліотеки для статистичного аналізу та машинного навчання. Ці інструменти стануть основою нашого дослідження та допоможуть в розгортанні зручної та ефективної системи аналізу даних щодо забруднення повітря в м. Вінниця.

### **3. ЗБИРАННЯ ТА ПІДГОТОВКА ДО СИСТЕМНОГО АНАЛІЗУ ДАНИХ ЗАБРУДНЕННЯ АТМОСФЕРНОГО ПОВІТРЯ МІСТА ВІННИЦЯ**

#### **3.1 Збирання та опис ознак вхідного набору даних**

Для виконання системного аналізу даних забруднення повітря, було взято дані з громадського моніторингу якості повітря міста Вінниці, де на даний момент встановлено 16 станцій по всьому місту та поза його межами. Дані зі станцій, щотижня автоматично розміщуються на порталі відкритих даних Вінницької міської ради, які відображають дані стосовно якості повітря у м. Вінниці за останні 7 днів.

Портал відкритих даних ВМР є спеціально розробленою платформою для публікації інформації у відкритому форматі. Це дозволяє легко обробляти дані за допомогою комп'ютерних програм і забезпечує безкоштовний та вільний доступ до них. Всі дані на порталі мають чітку структуру і формат, що спрощує їх завантаження та використання в подальшому [22].

У 2017 році Вінниця приєдналась до Міжнародної Хартії відкритих даних, підписавши Меморандум про співпрацю в рамках пілотного проекту з відкритими даними. Рішення виконавчого комітету міської ради від 20.07.2017 р. №1627 “Про приєднання Вінницької міської ради та її виконавчих органів до Міжнародної Хартії відкритих даних” було прийняте. План дій для реалізації принципів Міжнародної Хартії відкритих даних був запроваджений відповідно до розпорядження міського голови від 18.10.2017 р. №181-Р “Про заходи щодо впровадження принципів Міжнародної Хартії відкритих даних та розвитку відкритих даних” [23].

Для швидкого, зручного та простого збору даних, було використано програму MS Excel.

Microsoft Excel — це програма, розроблена компанією Microsoft та є частиною пакету програм Microsoft Office. Це потужний інструмент, призначений для створення та редагування електронних таблиць, а також їх

форматування. Програма має широкі можливості у галузі математичних та статистичних обчислень, що дозволяє користувачам швидко та ефективно обробляти дані, які містяться у таблицях. Однією з основних переваг цього продукту є те, що майже всі сучасні програми підтримують формати, які використовує MS Excel. Це значно спрощує імпорт даних та їх подальшу обробку, оскільки користувачі можуть легко обмінюватися інформацією між різними програмами, забезпечуючи зручність та сумісність у використанні даних. Microsoft Excel залишається невід’ємною частиною роботи в багатьох сферах, де потрібні точні обчислення та організація даних [24].

Отже, зібравши усі найактуальніші дані з громадського моніторингу якості повітря в м. Вінниця, а саме дані за вересень місяць 2023 року, перейдемо до опису ознак вхідного набору даних. На рисунку 3.1 можна побачити початковий вигляд даних забруднення атмосферного повітря м. Вінниця за вересень 2023 року у \*.csv форматі.

_id	stations...	stations...	Lat	Long	stations...	stations...	stations...	stations...	stations...	stations...	stations...	stations...	stations...
2	650	дім	49.343484	28.72983	2023-09-...	0	A3	PM2.5	PM2.5	Пил 2.5 ...	ug/m3	мкг/м³	4.8740
3	650	дім	49.343484	28.72983	2023-09-...	0	C1	VOC (Н...	VOC (Н...	ЛОС (Ф...	ppb	мкг/м³	10.0000
4	650	дім	49.343484	28.72983	2023-09-...	0	E1	Tempera...	Tempera...	Темпера...	°C	°C	13.5580
5	650	дім	49.343484	28.72983	2023-09-...	0	E2	Humidity	Humidity	Вологість	%	%	57.2250
6	650	дім	49.343484	28.72983	2023-09-...	0	E3	Pressure	Pressure	Атмосф...	Pa	мм.рт.ст.	99272.2
7	767	Соборн...	49.232859	28.470453	2023-09-...	0	A2	PM10	PM10	Пил 10 ...	ug/m3	мкг/м³	13.4511
8	767	Соборн...	49.232859	28.470453	2023-09-...	0	A3	PM2.5	PM2.5	Пил 2.5 ...	ug/m3	мкг/м³	5.7542
9	767	Соборн...	49.232859	28.470453	2023-09-...	0	E1	Tempera...	Tempera...	Темпера...	°C	°C	15.9537
10	767	Соборн...	49.232859	28.470453	2023-09-...	0	E2	Humidity	Humidity	Вологість	%	%	56.4974
11	767	Соборн...	49.232859	28.470453	2023-09-...	0	E3	Pressure	Pressure	Атмосф...	Pa	мм.рт.ст.	99021.2
12	790	станція ...	49.2270...	28.4189...	2023-09-...	0	A2	PM10	PM10	Пил 10 ...	ug/m3	мкг/м³	21.7440
13	790	станція ...	49.2270...	28.4189...	2023-09-...	0	A3	PM2.5	PM2.5	Пил 2.5 ...	ug/m3	мкг/м³	9.0650
14	790	станція ...	49.2270...	28.4189...	2023-09-...	0	E1	Tempera...	Tempera...	Темпера...	°C	°C	14.7475
15	790	станція ...	49.2270...	28.4189...	2023-09-...	0	E2	Humidity	Humidity	Вологість	%	%	53.4355
16	790	станція ...	49.2270...	28.4189...	2023-09-...	0	E3	Pressure	Pressure	Атмосф...	Pa	мм.рт.ст.	99054.5
17	1811	EcoCity ...	49.558012	27.956897	2023-09-...	0	A2	PM10	PM10	Пил 10 ...	ug/m3	мкг/м³	22.5850
18	1811	EcoCity ...	49.558012	27.956897	2023-09-...	0	A3	PM2.5	PM2.5	Пил 2.5 ...	ug/m3	мкг/м³	9.3400
19	1811	EcoCity ...	49.558012	27.956897	2023-09-...	0	D1	RAD	RAD	Радіацій...	uR/h	мкР/год	13.2820
20	1811	EcoCity ...	49.558012	27.956897	2023-09-...	0	E1	Tempera...	Tempera...	Темпера...	°C	°C	13.4920
21	1811	EcoCity ...	49.558012	27.956897	2023-09-...	0	E1	Tempera...	Tempera...	Темпера...	°C	°C	14.3150
22	1811	EcoCity ...	49.558012	27.956897	2023-09-...	0	E2	Humidity	Humidity	Вологість	%	%	65.6835
23	1811	EcoCity ...	49.558012	27.956897	2023-09-...	0	E3	Pressure	Pressure	Атмосф...	Pa	мм.рт.ст.	99278.7

Рисунок 3.1 – Вхідний набір даних забруднення атмосферного повітря в м. Вінниці за вересень 2023 року

У вхідному наборі даних, який відображено на рисунку 3.1 щодо забруднення атмосфери повітря в м. Вінниця, присутні наступні атрибути:

- “PM1.0” — частинки пилу розміром 1 мікромметр ( $PM_1$ );
- “PM2.5” — частинки пилу розміром 2,5 мікромметрів ( $PM_{2,5}$ );
- “PM100” — частинки пилу розміром 10 мікромметрів ( $PM_{10}$ );
- “Pressure” — атмосферний тиск;
- “Temperature” — температура повітря;
- “Humidity” — відсоток вологості;
- “VOC (H<sub>2</sub>CO)” — Формальдегід;
- “RAD” — радіаційний фон.

А також, є наступні колонки:

- “id” — унікальний ідентифікатор;
- “station\_id” — ідентифікатор станції моніторингу повітря;
- “station\_name” — назва станції моніторингу повітря;
- “lat” — координати (широта);
- “long” — координати (довгота);
- “station\_params\_id” — унікальний ідентифікатор вимірювального показнику;
- “station\_params\_key” — ключова назва вимірювального параметру;
- “station\_params\_name” — назва вимірювального параметру;
- “station\_params\_localName” — назва параметру українською мовою;
- “station\_params\_unit” — одиниця вимірювання параметру;
- “station\_params\_value” — значення параметру;
- “stations\_time” — час на станції;
- “stations\_params\_time” — час подачі показнику.

### 3.2 Завантаження датасету до системи Kaggle та підготовка до роботи

Перш ніж здійснювати системний аналіз даних щодо забруднення атмосферного повітря в м. Вінниця, необхідно завантажити наявний датасет до онлайн-середовища розробки, у нашому випадку Kaggle Notebooks, та підготувати простір до роботи – завантажити необхідні для роботи бібліотеки та перевірити коректність завантажених даних. Виконання усіх вище описаних операції можна розглядіти на рисунках 3.2-3.3.

× Upload Data

DATASET TITLE  
Air quality in Vinnytsia 2023 (September)  
www.kaggle.com/datasets/pirogovskiy/air-quality-in-vinnytsia-2023-september

VISIBILITY  
Public

LICENSE  
Apache 2.0

FILES  
Air quality in September 2023.csv (171.92 MB)

Drag & drop additional files to upload  
Consider zipping large directories for faster uploads  
or  
+ Upload another file

Public Quota: 171.92 MB / 107.37 GB

Reset Create

Рисунок 3.2 – Завантаження датасету в систему Kaggle

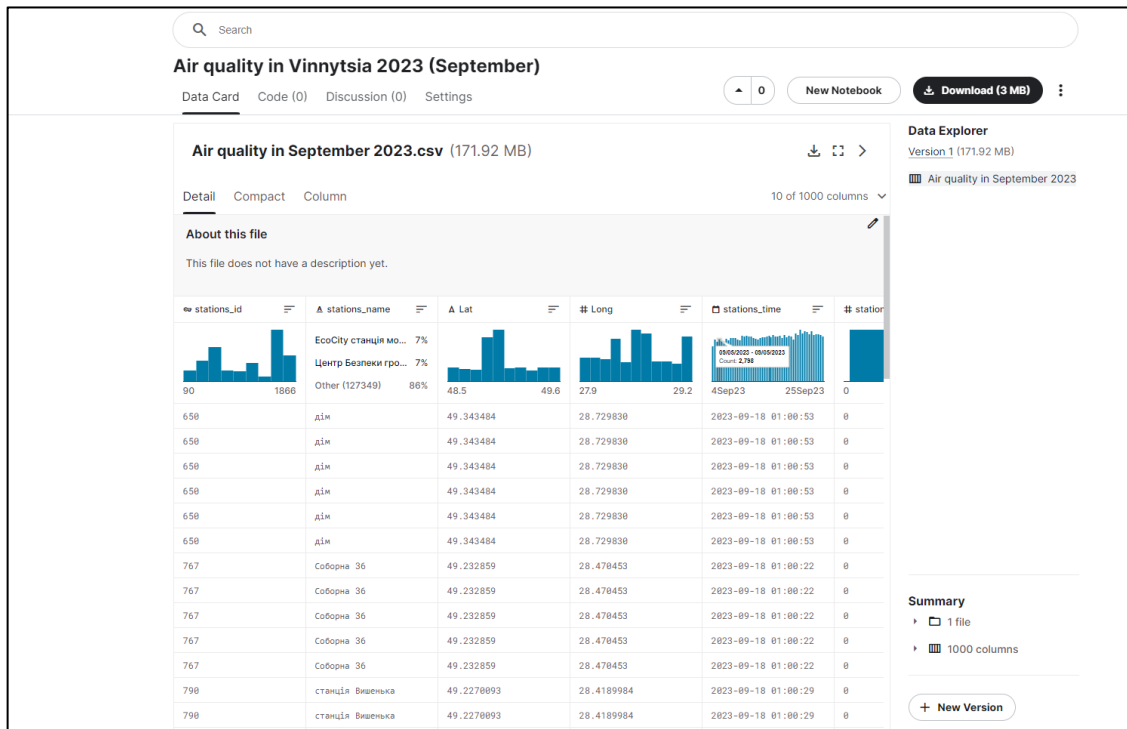


Рисунок 3.3 – Вигляд датасету, який було завантажено у систему Kaggle

На рисунках 3.2-3.3 зображено процес завантаження датасету у систему Kaggle та перевірка його наявності [25].

Далі, успішно завантаживши дані у онлайн-середовище Kaggle Notebooks, перейдемо до створення нового ноутбуку та підключення вже створеного датасету до нього. На рисунках 3.4-3.5 продемонстровано виконання вище зазначених процесів.

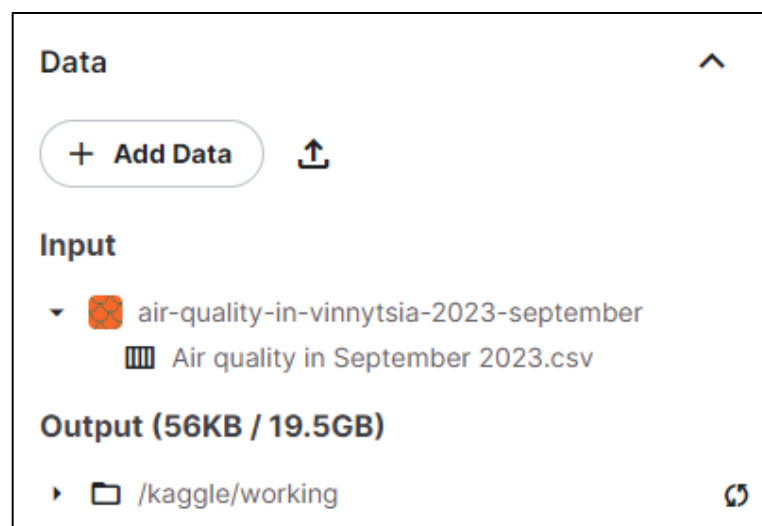


Рисунок 3.4 – Підключення даних до створеного ноутбуку

```
# Thanks to the users for the help to make this notebook:
# https://www.kaggle.com/code/vbmokin/air-quality-city-mapping

# Work with data - the main Python libraries
import numpy as np
import pandas as pd

# For import data
import os
import gc
import requests
from pprint import pprint
import json

from datetime import datetime

# Visualization
import matplotlib.pyplot as plt
import plotly.express as px
import plotly.graph_objects as go

# Others
import seaborn as sns
import scipy
from sklearn import *
import statsmodels.api as sm
import torch
import itertools
from pathlib import Path
from scipy.stats import zscore
from sklearn.metrics import r2_score
from statsmodels.tsa.arima.model import ARIMA
from prophet import Prophet

import folium #for building maps
from IPython.display import IFrame

import warnings
warnings.simplefilter('ignore')
```

Рисунок 3.5 – Завантаження необхідних для подальшої роботи бібліотек

Дивлячись на рисунки 3.4-3.5, можна побачити процес створення ноутбуку, підключення до нього раніше створеного датасету та завантаження усіх бібліотек, які знадобляться під час системного аналізу даних забруднення повітря в м. Вінниця.

Для того, аби перевірити успішність під'єднання датасету, за допомогою мови програмування Python, здійснимо його завантаження та виведемо перші 10 рядків датасету (рис. 3.6)

```
[2]: # Downloading data
path = Path('../input/air-quality-in-vinnitsia-2023-september')
df_air = pd.read_csv(path/'Air quality in September 2023.csv')

[3]: # Data view
df_air.head(10)
```

	stations_id	stations_name	Lat	Long	stations_time	stations_offset	stations_params_id	stations_params_key	stations_params_name	stations_params_localName	...	Unnamed: 991	Unnamed: 992	Unnamed: 993
0	650	дім	49.343484	28.72983	2023-09-18 01:00:53	0	A2	PM10	PM10	Пил 10 мкм ...	...	NaN	NaN	NaN
1	650	дім	49.343484	28.72983	2023-09-18 01:00:53	0	A3	PM2.5	PM2.5	Пил 2.5 мкм ...	...	NaN	NaN	NaN
2	650	дім	49.343484	28.72983	2023-09-18 01:00:53	0	C1	VOC (H <sub>2</sub> CO)	VOC (H <sub>2</sub> CO)	ЛОС (Формальдегід) ...	...	NaN	NaN	NaN
3	650	дім	49.343484	28.72983	2023-09-18 01:00:53	0	E1	Temperature	Temperature	Температура ...	...	NaN	NaN	NaN
4	650	дім	49.343484	28.72983	2023-09-18 01:00:53	0	E2	Humidity	Humidity	Вологість ...	...	NaN	NaN	NaN
5	650	дім	49.343484	28.72983	2023-09-18 01:00:53	0	E3	Pressure	Pressure	Атмосферний тиск ...	...	NaN	NaN	NaN
6	767	Соборна 36	49.232859	28.470453	2023-09-18 01:00:22	0	A2	PM10	PM10	Пил 10 мкм ...	...	NaN	NaN	NaN
7	767	Соборна 36	49.232859	28.470453	2023-09-18 01:00:22	0	A3	PM2.5	PM2.5	Пил 2.5 мкм ...	...	NaN	NaN	NaN
8	767	Соборна 36	49.232859	28.470453	2023-09-18 01:00:22	0	E1	Temperature	Temperature	Температура ...	...	NaN	NaN	NaN
9	767	Соборна 36	49.232859	28.470453	2023-09-18 01:00:22	0	E2	Humidity	Humidity	Вологість ...	...	NaN	NaN	NaN

10 rows × 1001 columns

Рисунок 3.6 – Завантаження та виведення даних

На рисунку 3.6 можна побачити, що процес завантаження датасету та вивід перших 10-ти його рядків було успішно виконано. Тому, після збирання даних, їх налаштування, підключення тощо, можна перейти до виконання поставленого завдання, а саме здійснити системний аналіз даних забруднення атмосферного повітря в м. Вінниця, у нашому випадку за вересень 2023 року.

### 3.3 Висновки

Під час формування цього розділу, акцентовано на зборі та обробці даних, пов'язаних із забрудненням атмосферного повітря в місті Вінниця.

Даний розділ почато з проведення збирання даних, визначення джерела, з якого отримано актуальну інформацію, а також більш глибоко вивчено середовище та методи, за допомогою яких ці дані були зібрані та структуровані.



Для підготовки до системного аналізу забруднення повітря в місті Вінниця, здійснено кілька ключових кроків, а саме:

- Створення нового ноутбуку та завантаження даних в систему Kaggle;
- Завантаження необхідних бібліотек;
- Підключення та налаштування даних, їх демонстрація (виведення перших 10-ти рядків основного завантаженого датасету).

Усі ці кроки були важливою передумовою для подальшого аналізу та вивчення забруднення повітря в місті Вінниця. Даний підхід до підготовки даних формує основу для глибокого розуміння проблеми та розвинених аналітичних висновків, які будуть формуватися в подальших наших досліджень.

## 4 СИСТЕМНИЙ АНАЛІЗ ТА ПРОГНОЗУВАННЯ ДАНИХ

### 4.1 Системний аналіз стану атмосферного повітря за даними громадського моніторингу в місті Вінниця

Після успішного виконання всіх умов, необхідних для початку системного аналізу даних щодо забруднення повітря в м. Вінниця, перейдемо до самого аналізу та всіх його аспектів. Наш наступний крок – докладне вивчення проблематики, пов'язаної із забрудненням атмосферного повітря.

Під час цього аналізу, розглянемо різні аспекти забруднення повітря, включаючи рівні різних забруднюючих речовин, вплив погодних умов на якість повітря, можливі взаємозв'язки між різними забруднюючими речовинами та їх вплив на здоров'я населення. Ми також розглянемо часові та просторові зміни у рівнях забруднення повітря та спробуємо знайти закономірності у цих даних.

Цей аналіз дозволить глибше зрозуміти проблему забруднення повітря в м. Вінниця та визначити можливі шляхи вирішення цієї проблеми для покращення якості повітря та забезпечення здоров'я та безпеки мешканців міста.

Отже, на першому етапі, проведемо аудит даних, щоб підготувати їх для подальшого докладного аналізу та візуалізації, використовуючи мову програмування Python. Цей крок включає в себе перевірку наявності та обробку можливих пропущених значень, видалення непотрібних стовпців або рядків, а також перевірку правильності форматів даних (рис. 4.1-4.4).

Після завершення цієї підготовки, дані будуть повністю готові до подальшого дослідження. Використовуючи бібліотеки Python, такі як Pandas, чи Seaborn, можна буде ретельно аналізувати дані щодо забруднення повітря в місті Вінниця.

```
[165]: # Drop columns named 'Unnamed: 1' through 'Unnamed: 1000'
df_air_cleaned = df_air.drop(df_air.columns[df_air.columns.str.startswith('Unnamed: ')], axis = 1)

# Cheking the cleaned dataset
df_air_cleaned.head(10)
```

	stations_id	stations_name	Lat	Long	stations_time	stations_offset	stations_params_id	stations_params_key	stations_params_name	stations_params_localName	stations_params_unit	stations_params_loca
0	650	д/м	49.343484	28.72983	2023-09-18 01:00:53	0	A2	PM10	PM10	Пил 10 мкм	ug/m3	м
1	650	д/м	49.343484	28.72983	2023-09-18 01:00:53	0	A3	PM2.5	PM2.5	Пил 2.5 мкм	ug/m3	м
2	650	д/м	49.343484	28.72983	2023-09-18 01:00:53	0	C1	VOC (H2CO)	VOC (H2CO)	ЛОС (Формальдегід)	ppb	м
3	650	д/м	49.343484	28.72983	2023-09-18 01:00:53	0	E1	Temperature	Temperature	Температура	°C	
4	650	д/м	49.343484	28.72983	2023-09-18 01:00:53	0	E2	Humidity	Humidity	Вологість	%	
5	650	д/м	49.343484	28.72983	2023-09-18 01:00:53	0	E3	Pressure	Pressure	Атмосферний тиск	Pa	мм.
6	767	Соборна 36	49.232859	28.470453	2023-09-18 01:00:22	0	A2	PM10	PM10	Пил 10 мкм	ug/m3	м
7	767	Соборна 36	49.232859	28.470453	2023-09-18 01:00:22	0	A3	PM2.5	PM2.5	Пил 2.5 мкм	ug/m3	м
8	767	Соборна 36	49.232859	28.470453	2023-09-18 01:00:22	0	E1	Temperature	Temperature	Температура	°C	
9	767	Соборна 36	49.232859	28.470453	2023-09-18 01:00:22	0	E2	Humidity	Humidity	Вологість	%	

Рисунок 4.1 – Видалення зайвих стовпців з нульовими значеннями

Дивлячись на рисунок, який показаний у третьому розділі, а саме рисунок 3.6, можна побачити, що дані містять в собі багато стовпців, більшість з яких мають нульові значення. Для того, щоб видалити їх, використаємо метод `drop()`, що зображено на рисунку 4.1.

```
[166]: # Dataset description
df_air_cleaned.describe()
```

	stations_id	stations_name	Lat	Long	stations_time	stations_offset	stations_params_id	stations_params_key	stations_params_name	stations_params_localName	stations_params_unit	station
<b>count</b>	148489	148489	148489.000000	148489.000000	148489	148489	148489	148489	148489	148489	148489	148489
<b>unique</b>	37	19	37.000000	37.000000	14044	3	11	12	12	11	9	
<b>top</b>	1811	ЕсоСіті Станція моніторингу Хмельник	49.558012	27.956897	2023-09- 07T03:54:00Z	0	E1	Temperature	Temperature	Температура	ug/m3	
<b>freq</b>	10423	10570	10423.000000	10423.000000	42	146441	32836	32836	32836	32836	57402	

Рисунок 4.2 – Опис датасету

Видаливши непотрібні дані, за допомогою метода `describe()`, здійснимо опис загальних статистичних показників утвореного датасету. На рисунку 4.2, видно такі статистичні показники:

- “count — показник, що демонструє загальну кількість даних стовпців;
- “unique” — кількість унікальних значень в стовпці;
- “top” — найбільше число в стовпці;

– “freq” — вказує на кількість значень, які найчастіше зустрічаються в поточному стовпці.

```
[167]: # Basic information of dataset
df_air_cleaned.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 148489 entries, 0 to 148488
Data columns (total 17 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   stations_id                            148489 non-null object
1   stations_name                           148489 non-null object
2   Lat                                      148489 non-null object
3   Long                                     148489 non-null object
4   stations_time                           148489 non-null object
5   stations_offset                          148489 non-null object
6   stations_params_id                       148489 non-null object
7   stations_params_key                      148489 non-null object
8   stations_params_name                     148489 non-null object
9   stations_params_localName                148489 non-null object
10  stations_params_unit                     148489 non-null object
11  stations_params_localUnit                148489 non-null object
12  stations_params_value                    145548 non-null object
13  stations_params_cr                       87559 non-null object
14  stations_params_time                     148489 non-null object
15  stations_params_offset                   148489 non-null object
16  stations_params_level                    148489 non-null object
dtypes: object(17)
memory usage: 19.3+ MB
```

Рисунок 4.3 – Загальна інформація про датасет

```
[168]: # Find null values
df_air_cleaned.isnull().sum()

[168... stations_id                0
stations_name              0
Lat                        0
Long                       0
stations_time              0
stations_offset            0
stations_params_id        0
stations_params_key       0
stations_params_name      0
stations_params_localName 0
stations_params_unit      0
stations_params_localUnit 0
stations_params_value     2941
stations_params_cr        60930
stations_params_time      0
stations_params_offset    0
stations_params_level     0
dtype: int64

[169]: # Drop rows with null values from dataset
df_air_cleaned_2 = df_air_cleaned.dropna()
```

Рисунок 4.4 – Знаходження та видалення нульових значень в наборі даних

На рисунках 4.3 та 4.4, зображено процес виведення загальної інформації стосовно датасету та, за допомогою методу `isnull()`, процес знаходження та видалення усіх нульових значень, які присутні в наборі даних.

По завершенні перевірки наявності даних та їх обробки (включаючи видалення стовпців та нульових значень), наступним кроком буде побудова мапи розташування станцій моніторингу повітря, за допомогою мови програмування Python [26].

Даний процес, а також саму мапу, можна буде переглянути на рисунках 4.5-4.6.

```
[192]: # Creating new dataset for building map (List of stations)
df_air_for_map = pd.DataFrame({
    'station_unique_id': ['650', '767', '790', '1811', '1813', '1825', '1830', '1834', '1866', '246', '256', '271', '274', '281', '315', '90'],
    'station_name': ['дім', 'Соборна 36', 'станція Вишенька', 'EcoCity станція моніторингу Хмільник', 'Якушинці', 'Центральна площа', 'Центр Безпеки грк
    'Eco Stryzh', 'Северинівка', 'vinnytsia-246', 'vinnytsia-256', 'vinnytsia-271', 'vinnytsia-274', 'vinnytsia-281', 'vinnytsia-315',
    'vinnytsia-90'],
    'Lat': ['49.343484', '49.232859', '49.2270093', '49.558012', '49.2571196', '49.105637', '49.6288314', '49.3064814', '48.516638', '49.2234644', '49.24
    '49.23688714', '49.22477769', '49.24192687', '49.24368759', '49.21773365'],
    'Long': ['28.72983', '28.470453', '28.4189984', '27.956897', '28.3671704', '29.206617', '27.89884', '28.4828855', '29.184421', '28.40095116', '28.52
    '28.51330549', '28.42491984', '28.46202092', '28.49643223', '28.44979464']
})

# Converting 'Lat' and 'Long' columns to float
df_air_for_map['Lat'] = df_air_for_map['Lat'].astype(float)
df_air_for_map['Long'] = df_air_for_map['Long'].astype(float)

df_air_for_map
```

	station_unique_id	station_name	Lat	Long
0	650	дім	49.343484	28.729830
1	767	Соборна 36	49.232859	28.470453
2	790	станція Вишенька	49.227009	28.418998
3	1811	EcoCity станція моніторингу Хмільник	49.558012	27.956897
4	1813	Якушинці	49.257120	28.367170
5	1825	Центральна площа	49.105637	29.206617
6	1830	Центр Безпеки громадян Війтівецької сільської ...	49.628831	27.898840
7	1834	Eco Stryzh	49.306481	28.482885
8	1866	Северинівка	48.516638	29.184421
9	246	vinnytsia-246	49.223464	28.400951
10	256	vinnytsia-256	49.204856	28.528835
11	271	vinnytsia-271	49.236887	28.513305
12	274	vinnytsia-274	49.224778	28.424920

Рисунок 4.5 – Створення нового датасету для побудови карти розташування станцій моніторингу повітря

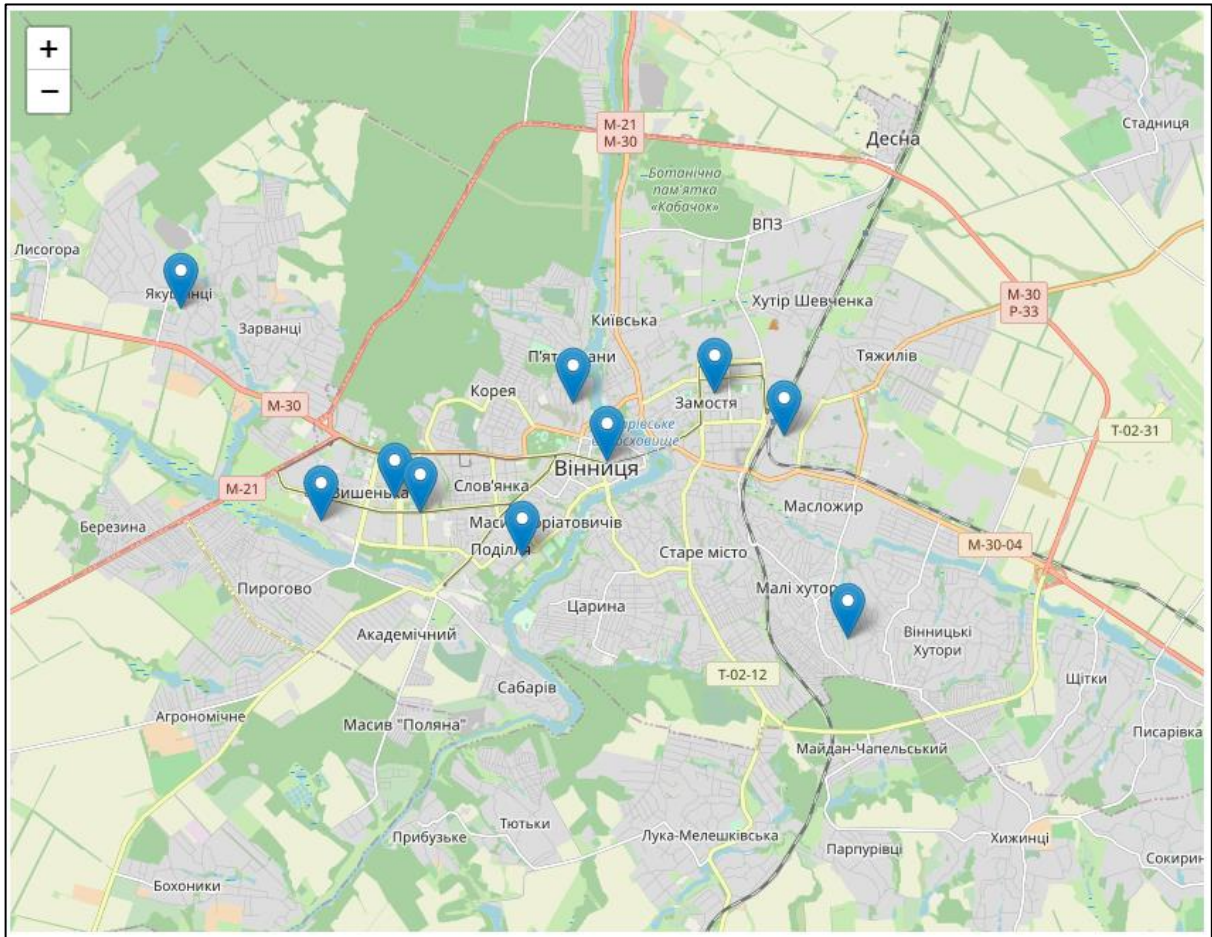


Рисунок 4.6 – Побудована мапа розташування станцій моніторингу повітря

На рисунку 4.6 представлено карту розташувань станцій моніторингу повітря. Як зазначалося в розділах, описаних вище, в місті Вінниці та поза його межами, на даний момент встановлено 16 станцій моніторингу повітря, які щоденно працюють та надають інформацію.

Наступним етапом є проведення розвідувального аналізу даних та створення візуалізації щодо забруднення повітря в м. Вінниця. Перед тим як розпочинати побудову графіків, спершу, створимо вибірки. Це дозволить більш зручно та ефективно організувати та відображувати дані на графіках, забезпечуючи простоту сприйняття й аналізу. На рисунку 4.7 можна побачити, основну, новостворену, вибірку.

	stations_id	stations_params_time	stations_params_name	stations_params_value
0	650	2023-09-18 01:00:53	PM10	4.663
1	650	2023-09-18 01:00:53	PM2.5	4.164
2	650	2023-09-18 01:00:53	VOC (H <sub>2</sub> CO)	10.0
3	650	2023-09-18 01:00:53	Temperature	15.413
4	650	2023-09-18 01:00:53	Humidity	62.951
...	...	...	...	...
148442	1866	2023-09-18 00:40:04	RAD	13.0089
148443	1866	2023-09-18 00:40:04	Temperature	11.3168
148444	1866	2023-09-18 00:40:04	Temperature	10.1279
148445	1866	2023-09-18 00:40:04	Humidity	75.3489
148446	1866	2023-09-18 00:40:04	Pressure	100136.4511

87559 rows × 4 columns

Рисунок 4.7 – Основна вибірка даних

На рисунку 4.7 можна побачити вибірку “df\_air\_selected”, в яку входять такі параметри, як:

- “station\_id” — унікальний номер станції з якої було взято показники;
- “station\_params\_time” — час, коли дані показники були виміряні;
- “station\_params\_name” — назва показнику;
- “station\_params\_value” — значення вимірювального показнику.

Далі, для полегшення аналізу, розділимо новостворену вибірку за різними ознаками, групуючи та класифікуючи її за конкретними характеристиками. Результати утворених вибірок можна побачити на рисунках 4.8-4.15.

	stations_id	stations_params_time	stations_params_name	stations_params_value
17894	1315	2023-09-20 12:39:09	PM1.0	17.25
18011	1315	2023-09-20 12:44:43	PM1.0	16.4
18128	1315	2023-09-20 13:19:38	PM1.0	15.8686
18245	1315	2023-09-20 13:39:52	PM1.0	16.0194
18362	1315	2023-09-20 13:59:06	PM1.0	16.7819
...	...	...	...	...
51870	1315	2023-09-24 23:19:03	PM1.0	22.6665
51980	1315	2023-09-24 23:38:00	PM1.0	23.2194
52090	1315	2023-09-24 23:59:19	PM1.0	22.6559
52200	1315	2023-09-25 00:19:37	PM1.0	21.8335
52304	1315	2023-09-25 00:38:43	PM1.0	19.6044

325 rows × 4 columns

Рисунок 4.8 – Вибірка за ознакою частинок пилу розміром 1 мкм

	stations_id	stations_params_time	stations_params_name	stations_params_value
1	650	2023-09-18 01:00:53	PM2.5	4.164
7	767	2023-09-18 01:00:22	PM2.5	3.39
12	790	2023-09-18 01:00:29	PM2.5	7.16
17	1811	2023-09-18 01:00:10	PM2.5	6.395
24	1813	2023-09-18 01:00:26	PM2.5	6.9
...	...	...	...	...
148413	1813	2023-09-18 00:40:19	PM2.5	6.89
148420	1825	2023-09-18 00:40:01	PM2.5	15.9294
148427	1830	2023-09-18 00:40:47	PM2.5	9.6125
148434	1834	2023-09-18 00:40:50	PM2.5	8.325
148441	1866	2023-09-18 00:40:04	PM2.5	4.8579

13306 rows × 4 columns

Рисунок 4.9 – Вибірка за ознакою частинок пилу розміром 2,5 мкм

На рисунку 4.8 та 4.9 представлені вибірки, які були сформовані за різними ознаками, а саме, на рисунку 4.8 зображена вибірка з частинок пилу розміром 1 мкм, а на рисунку 4.9 – вибірка з частинок пилу розміром 2,5 мкм.



	stations_id	stations_params_time	stations_params_name	stations_params_value
0	650	2023-09-18 01:00:53	PM10	4.663
6	767	2023-09-18 01:00:22	PM10	6.4879
11	790	2023-09-18 01:00:29	PM10	14.024
16	1811	2023-09-18 01:00:10	PM10	14.35
23	1813	2023-09-18 01:00:26	PM10	14.78
...	...	...	...	...
148412	1813	2023-09-18 00:40:19	PM10	14.095
148419	1825	2023-09-18 00:40:01	PM10	33.5706
148426	1830	2023-09-18 00:40:47	PM10	21.3688
148433	1834	2023-09-18 00:40:50	PM10	17.195
148440	1866	2023-09-18 00:40:04	PM10	10.0737

13306 rows × 4 columns

Рисунок 4.10 – Вибірка за ознакою частинок пилу розміром 10 мкм

	stations_id	stations_params_time	stations_params_name	stations_params_value
2	650	2023-09-18 01:00:53	VOC (H <sub>2</sub> CO)	10.0
102	650	2023-09-18 01:01:56	VOC (H <sub>2</sub> CO)	10.0
296	650	2023-09-18 02:00:44	VOC (H <sub>2</sub> CO)	10.0
396	650	2023-09-18 02:02:49	VOC (H <sub>2</sub> CO)	10.0
590	650	2023-09-18 03:00:19	VOC (H <sub>2</sub> CO)	10.0
...	...	...	...	...
147897	650	2023-09-17 23:00:03	VOC (H <sub>2</sub> CO)	10.0
147997	650	2023-09-17 23:20:46	VOC (H <sub>2</sub> CO)	10.0
148097	650	2023-09-17 23:40:26	VOC (H <sub>2</sub> CO)	10.0
148197	650	2023-09-18 00:00:08	VOC (H <sub>2</sub> CO)	10.0
148297	650	2023-09-18 00:17:45	VOC (H <sub>2</sub> CO)	10.0

1619 rows × 4 columns

Рисунок 4.11 – Вибірка за ознакою формальдегіду

Аналогічним методом, було сформовано вибірки, які вказані на рисунках 4.10-4.11, де перша вибірка стосується частинок пилу розміром 10 мкм, а друга – формальдегіду.

	stations_id	stations_params_time	stations_params_name	stations_params_value
3	650	2023-09-18 01:00:53	Temperature	15.413
8	767	2023-09-18 01:00:22	Temperature	17.6468
13	790	2023-09-18 01:00:29	Temperature	16.8595
19	1811	2023-09-18 01:00:10	Temperature	15.411
20	1811	2023-09-18 01:00:10	Temperature	16.226
...	...	...	...	...
148430	1830	2023-09-18 00:40:47	Temperature	12.3475
148436	1834	2023-09-18 00:40:50	Temperature	16.3915
148437	1834	2023-09-18 00:40:50	Temperature	15.3355
148443	1866	2023-09-18 00:40:04	Temperature	11.3168
148444	1866	2023-09-18 00:40:04	Temperature	10.1279

22681 rows × 4 columns

Рисунок 4.12 – Вибірка за ознакою температури

	stations_id	stations_params_time	stations_params_name	stations_params_value
4	650	2023-09-18 01:00:53	Humidity	62.951
9	767	2023-09-18 01:00:22	Humidity	50.8926
14	790	2023-09-18 01:00:29	Humidity	52.397
21	1811	2023-09-18 01:00:10	Humidity	59.884
28	1813	2023-09-18 01:00:26	Humidity	65.0785
...	...	...	...	...
148417	1813	2023-09-18 00:40:19	Humidity	63.6175
148424	1825	2023-09-18 00:40:01	Humidity	73.5235
148431	1830	2023-09-18 00:40:47	Humidity	77.6219
148438	1834	2023-09-18 00:40:50	Humidity	67.5495
148445	1866	2023-09-18 00:40:04	Humidity	75.3489

13631 rows × 4 columns

Рисунок 4.13 – Вибірка за ознакою вологості

Дивлячись на рисунки 4.12-4.13, можна розглянути схожі до попередніх вибірки, але з ознаками вологості та температури.

	stations_id	stations_params_time	stations_params_name	stations_params_value
5	650	2023-09-18 01:00:53	Pressure	99461.903
10	767	2023-09-18 01:00:22	Pressure	99208.84
15	790	2023-09-18 01:00:29	Pressure	99242.8715
22	1811	2023-09-18 01:00:10	Pressure	99464.194
29	1813	2023-09-18 01:00:26	Pressure	99038.8745
...	...	...	...	...
148418	1813	2023-09-18 00:40:19	Pressure	99040.1785
148425	1825	2023-09-18 00:40:01	Pressure	99923.1788
148432	1830	2023-09-18 00:40:47	Pressure	99238.2312
148439	1834	2023-09-18 00:40:50	Pressure	99586.3685
148446	1866	2023-09-18 00:40:04	Pressure	100136.4511

13306 rows × 4 columns

Рисунок 4.14 – Вибірка за ознакою атмосферного тиску

	stations_id	stations_params_time	stations_params_name	stations_params_value
18	1811	2023-09-18 01:00:10	RAD	13.6995
25	1813	2023-09-18 01:00:26	RAD	11.9545
32	1825	2023-09-18 01:00:10	RAD	14.3163
39	1830	2023-09-18 01:00:28	RAD	11.976
46	1834	2023-09-18 01:00:57	RAD	13.578
...	...	...	...	...
148414	1813	2023-09-18 00:40:19	RAD	12.0585
148421	1825	2023-09-18 00:40:01	RAD	13.9471
148428	1830	2023-09-18 00:40:47	RAD	10.8756
148435	1834	2023-09-18 00:40:50	RAD	14.2685
148442	1866	2023-09-18 00:40:04	RAD	13.0089

8733 rows × 4 columns

Рисунок 4.15 – Вибірка за ознакою радіаційного фону

На рисунках 4.14-4.15 відображено останні з сформованих вибірок, які містять інформацію стосовно ознак атмосферного тиску та радіаційного фону.

Отже, після формування вибірок, можна перейти до візуалізації даних. Спершу, порівняємо кожний показник з його встановленою нормою, щоб оцінити загальну якість повітря, зокрема в місті Вінниця станом на вересень 2023 року. Отримані результати в вигляді графіків можна буде розгледіти на рисунках 4.16-4.20.

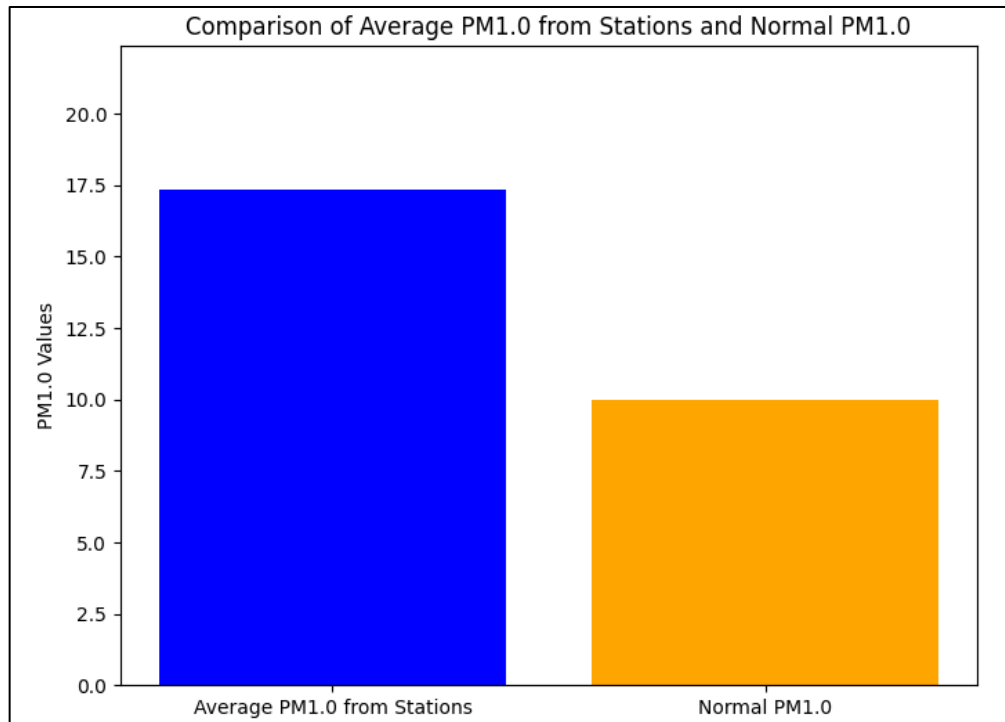


Рисунок 4.16 – Зрівняння показників частинок пилу розміром 1 мкм з встановленою нормою

На зображенні 4.16 представлено порівняння середнього рівня частинок пилу розміром 1 мікромметр з встановленою нормою. Згідно з цим порівнянням, можна виявити значне перевищення встановленої норми.

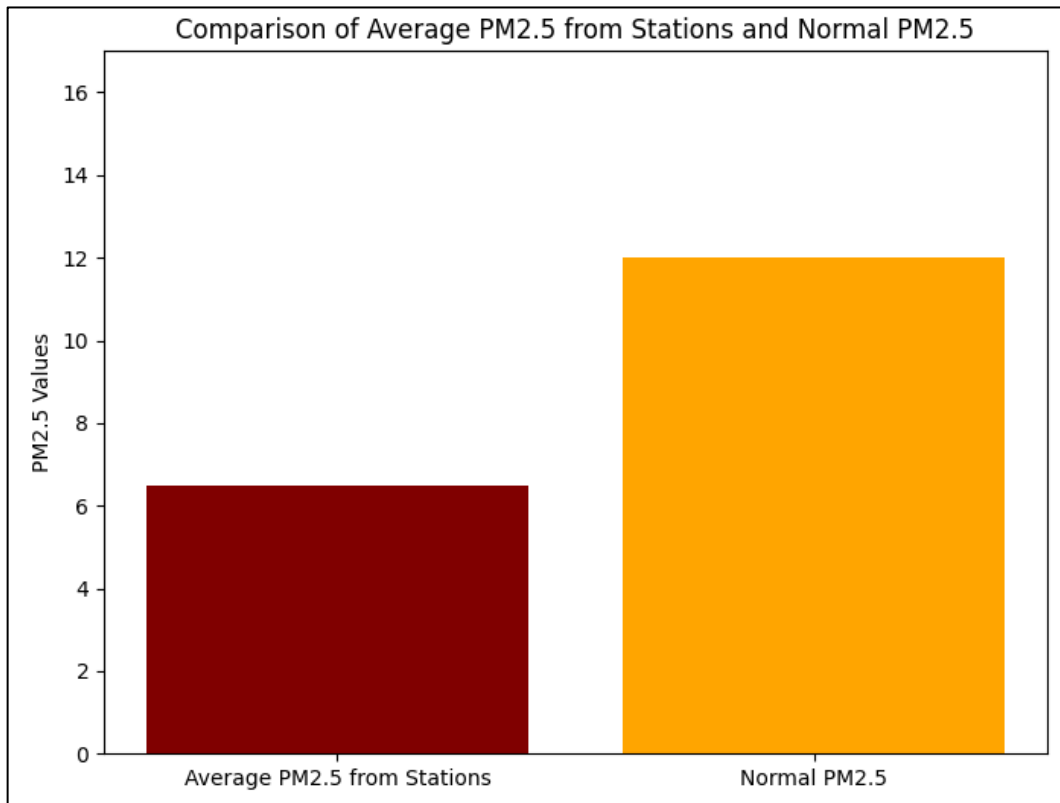


Рисунок 4.17 – Зрівняння показників частинок пилу розміром 2,5 мкм з встановленою нормою

На зображенні 4.17 показано порівняння середнього рівня частинок пилу розміром 2,5 мікрметра з встановленою нормою, де можна визначити, що середній рівень пилу розміром 2,5 мікрметра на момент вересня 2023 року не перевищує встановленої норми.

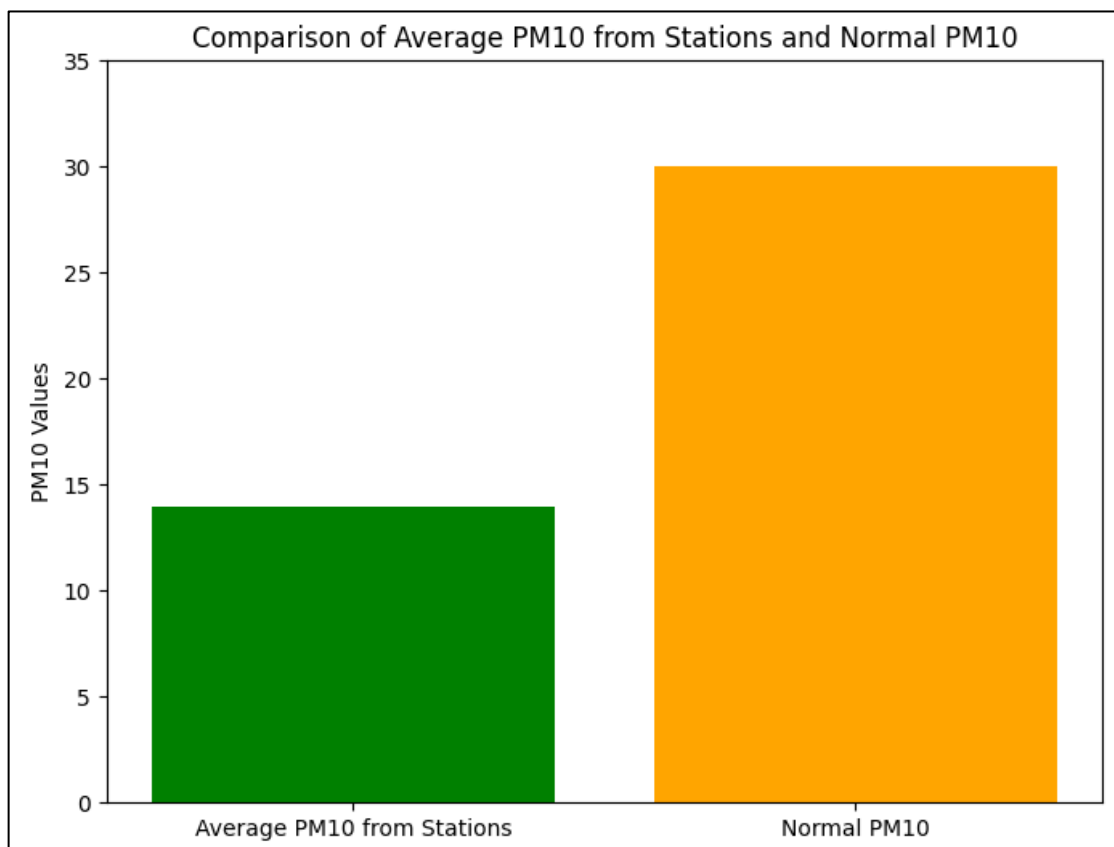


Рисунок 4.18 – Зрівняння показників частинок пилу розміром 10 мкм з встановленою нормою

На малюнку 4.18 представлено зрівняння середнього рівня частинок пилу діаметром 10 мікрметрів з встановленою нормою. З цим порівнянням можна визначити, що середній рівень пилу розміром 10 мікрметрів на даний момент не перевищує встановленої норми.

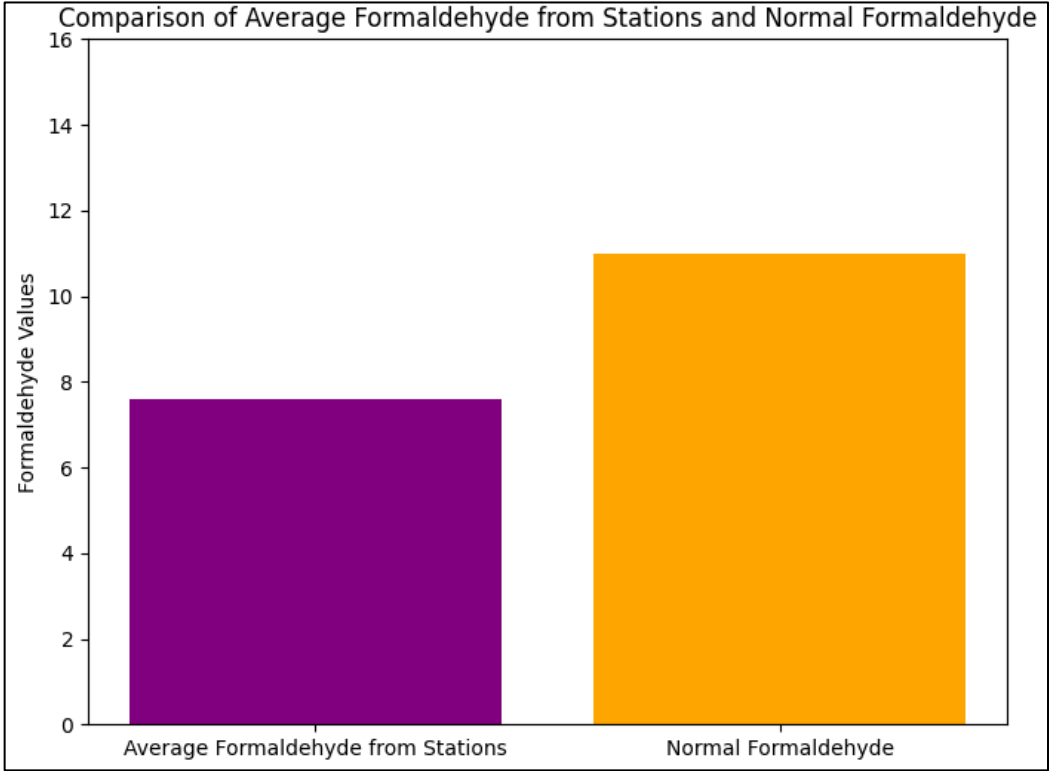


Рисунок 4.19 – Зрівняння показників формальдегіду з встановленою нормою

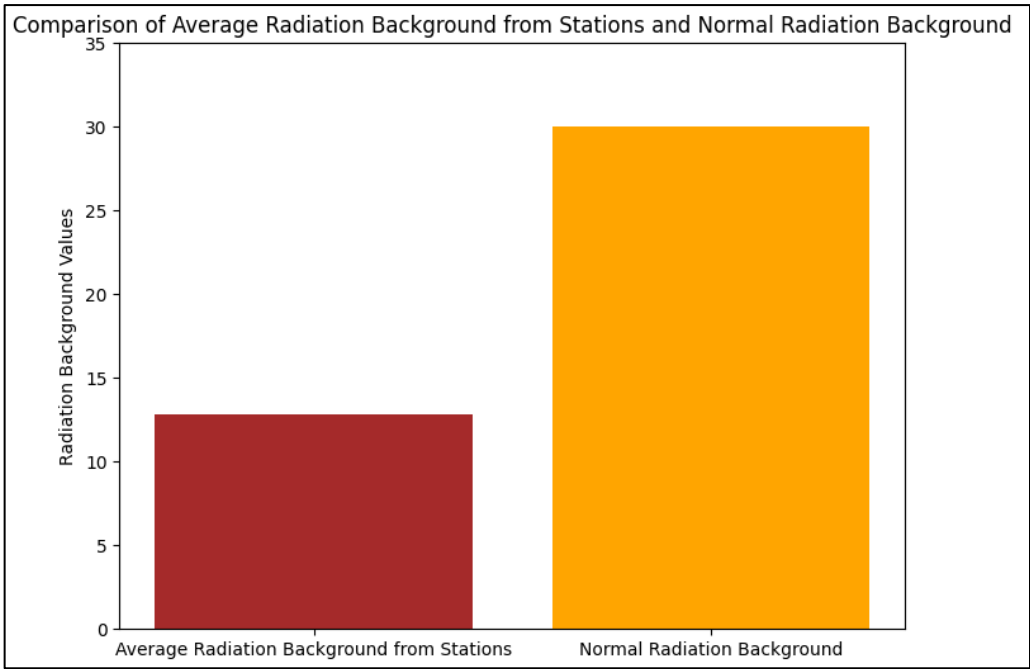


Рисунок 4.20 – Зрівняння показників радіаційного фону з встановленою нормою

На рисунках 4.16-4.20 можна знайти результати порівняння середніх показників різних речовин та їх норми. За цими даними видно, що станом на

вересень 2023 року якість повітря в місті в цілому є задовільною, за винятком показника  $PM_{10}$  – частинок пилу розміром 10 мікромметр. На рисунку 4.16 видно значне перевищення норми цього показника, практично удвічі.

Згідно цієї інформації, можна зробити висновок, що загальний стан повітря в м. Вінниця досить добрий, окрім показника  $PM_{10}$ , на який потрібно буде звернути увагу в подальших дослідженнях.

Далі, створимо нові вибірки та проведемо їх візуалізацію для більш детальнішого дослідження. Результати створенної візуалізації можна побачити на рисунках 4.21-4.28.

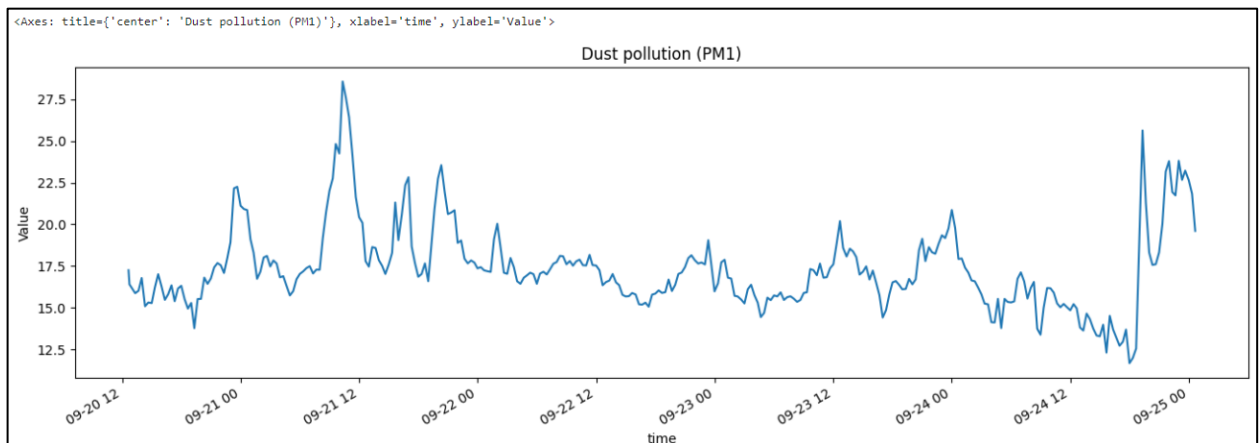


Рисунок 4.21 – Графік з відображенням динаміки частинок пилу розміром 1 мкм у часі

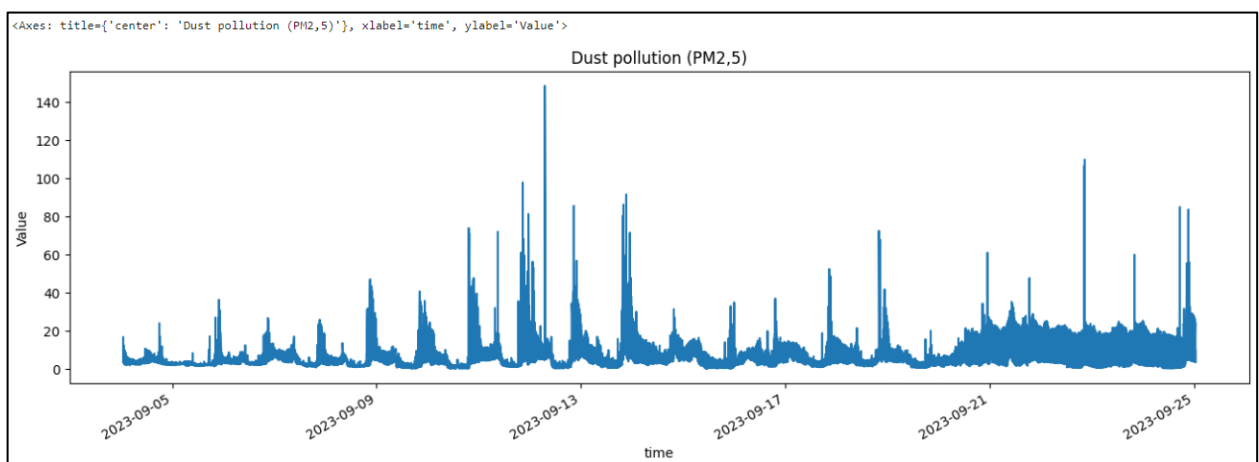


Рисунок 4.22 – Графік з відображенням динаміки частинок пилу розміром 2,5 мкм у часі



На зображеннях 4.21-4.22 представлені графіки, які відображають динаміку частинок пилу розміром 1 та 2,5 мікрометрів відповідно. Аналізуючи ці графіки, можна спостерігати за їхньою динамікою значень у вересні 2023 року.

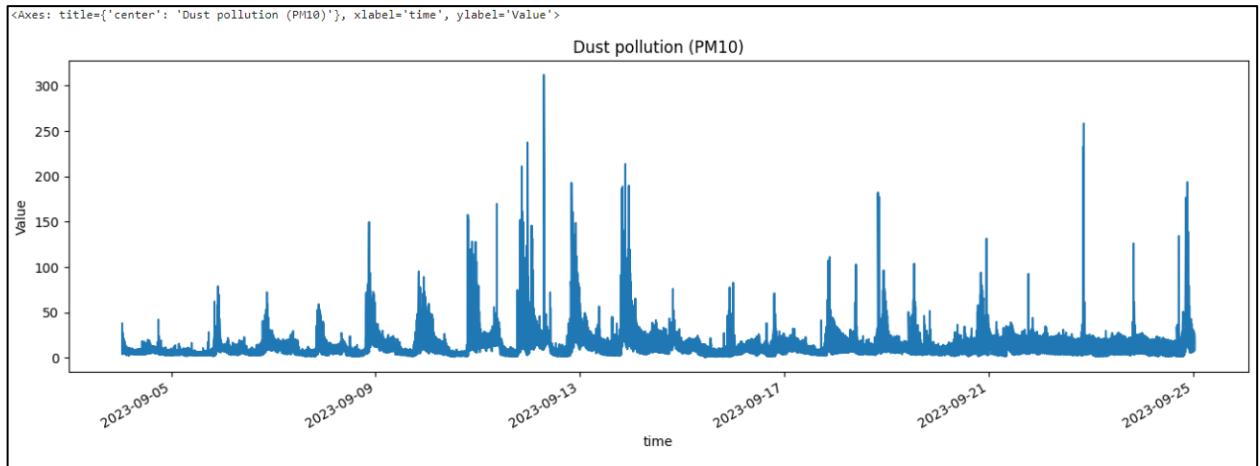


Рисунок 4.23 – Графік з відображенням динаміки частинок пилу розміром 10 мкм у часі

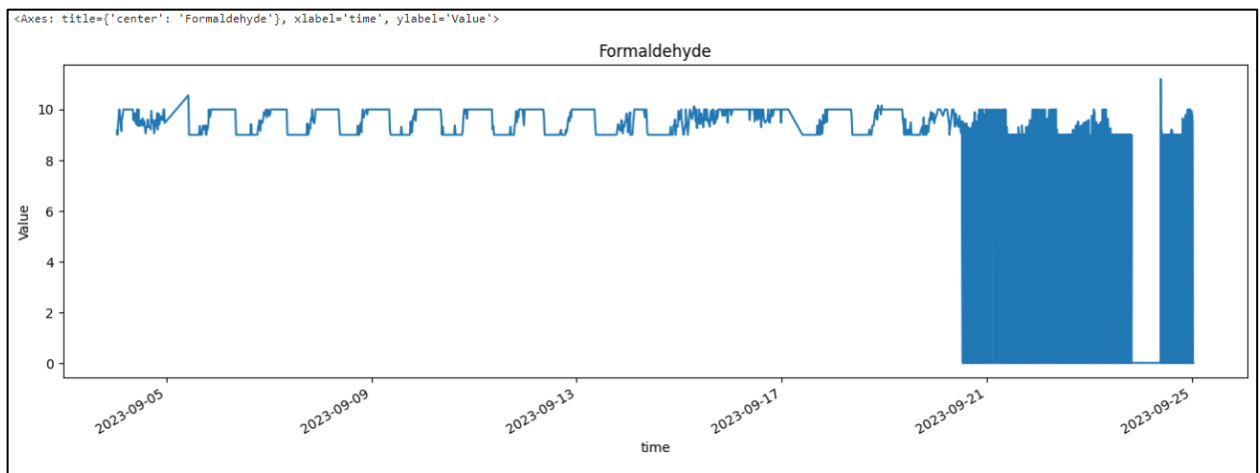


Рисунок 4.24 – Графік з відображенням динаміки формальдегіду у часі

На малюнках 4.23-4.24 подані графіки, що ілюструють зміни у концентрації частинок пилу розміром 10 мікрометрів та формальдегіду. Проводячи аналіз цих графіків, можна відслідкувати їхню динаміку за цей ж самий проміжок часу, а саме, вересень 2023 року.

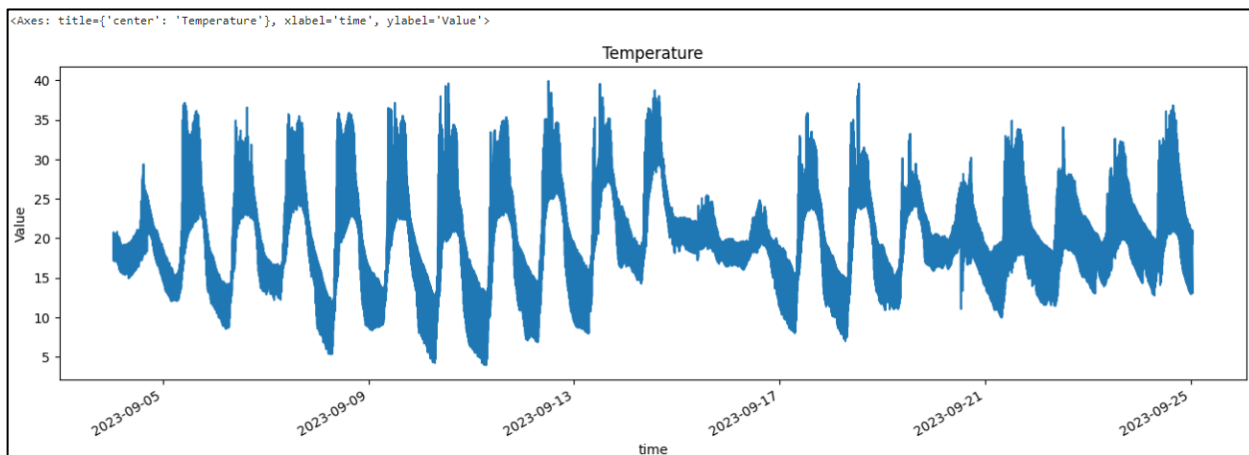


Рисунок 4.25 – Графік з відображенням динаміки температури у часі

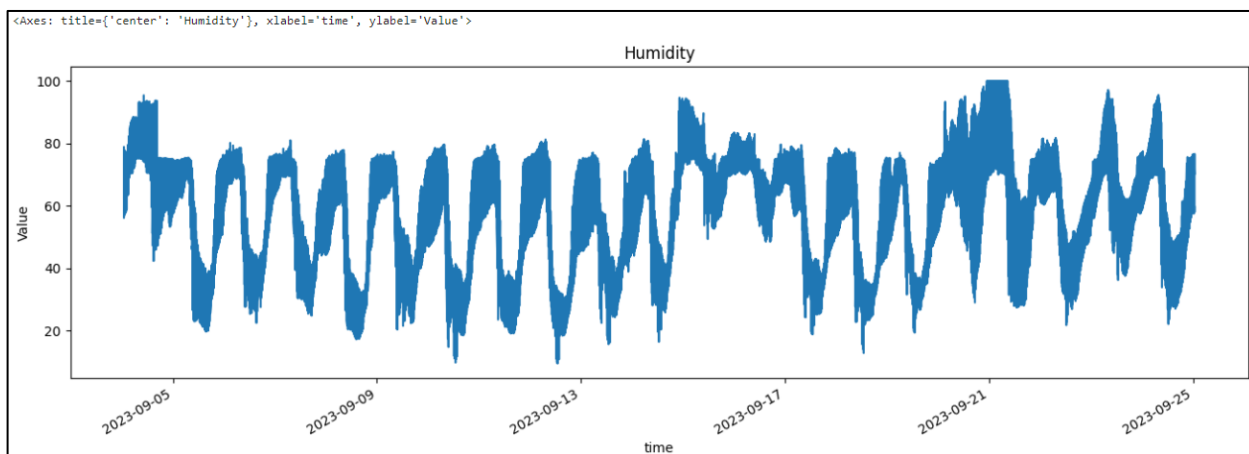


Рисунок 4.26 – Графік з відображенням динаміки вологості у часі

На малюнках 4.25-4.26 представлені графіки, що відображають зміни у рівнях вологості та значення показнику температури повітря. За даними вищевказаних рисунків можна відслідкувати загальну динаміку значень показників вологості та температури повітря протягом вересня 2023 року.

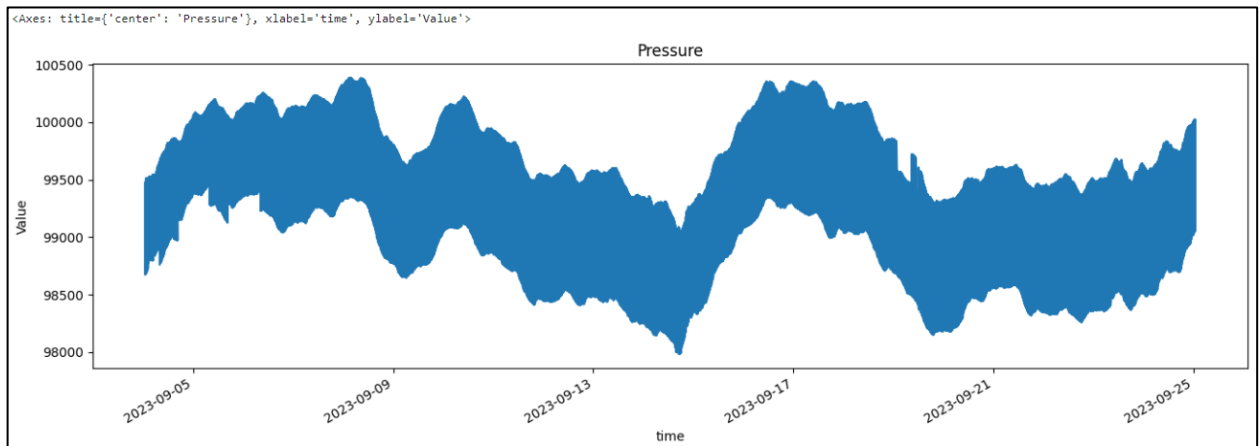


Рисунок 4.27 – Графік з відображенням динаміки атмосферного тиску у часі

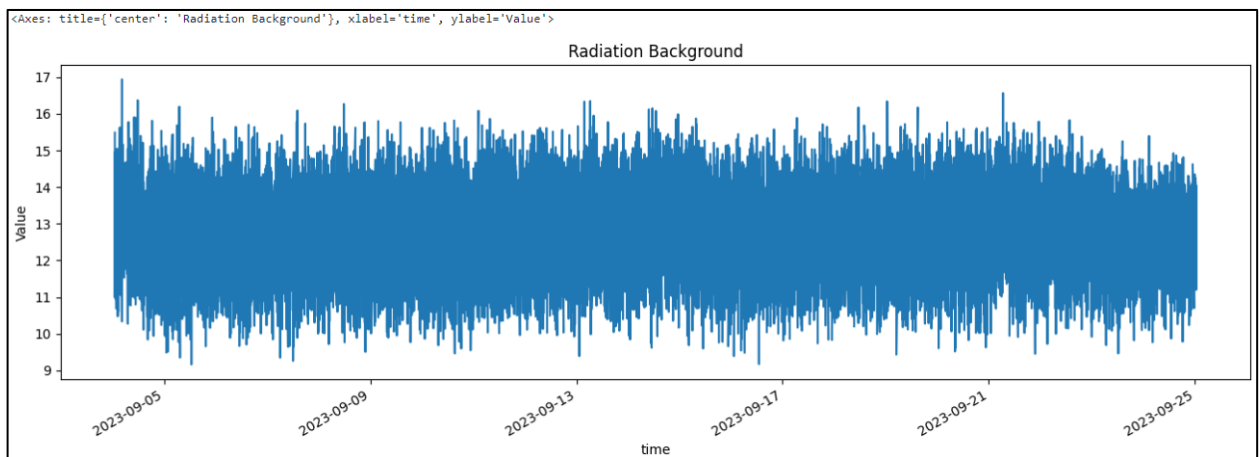


Рисунок 4.28 – Графік з відображенням динаміки радіаційного фону у часі

Аналізуючи графіки на рисунках 4.21-4.28, можна відзначити динаміку вимірювальних показників. Наприклад, значення показників  $PM_{2.5}$  та  $PM_{10}$  поступово збільшується у період з 09.09.2023 до 14.09.2023 та досягає найвищих значень за весь місяць. Щодо радіаційного фону, то, хоча значення цього показника часто коливаються, вони завжди залишаються в межах норми.

Однак найцікавішим для дослідження був графік, який відображає динаміку температури повітря у часі (рис. 4.25). У цьому році світ спостерігав аномально високі температури, що підтверджується даною візуалізацією. Графік показує, що температури протягом дня майже завжди перевищує 30 градусів Цельсія, навіть у перший місяць осені. У зв'язку з цим, побудуємо новий графік,

в якому буде відображено порівняння середньої температури за вересень місяць 2022 та 2023 року відповідно (рис. 4.29).

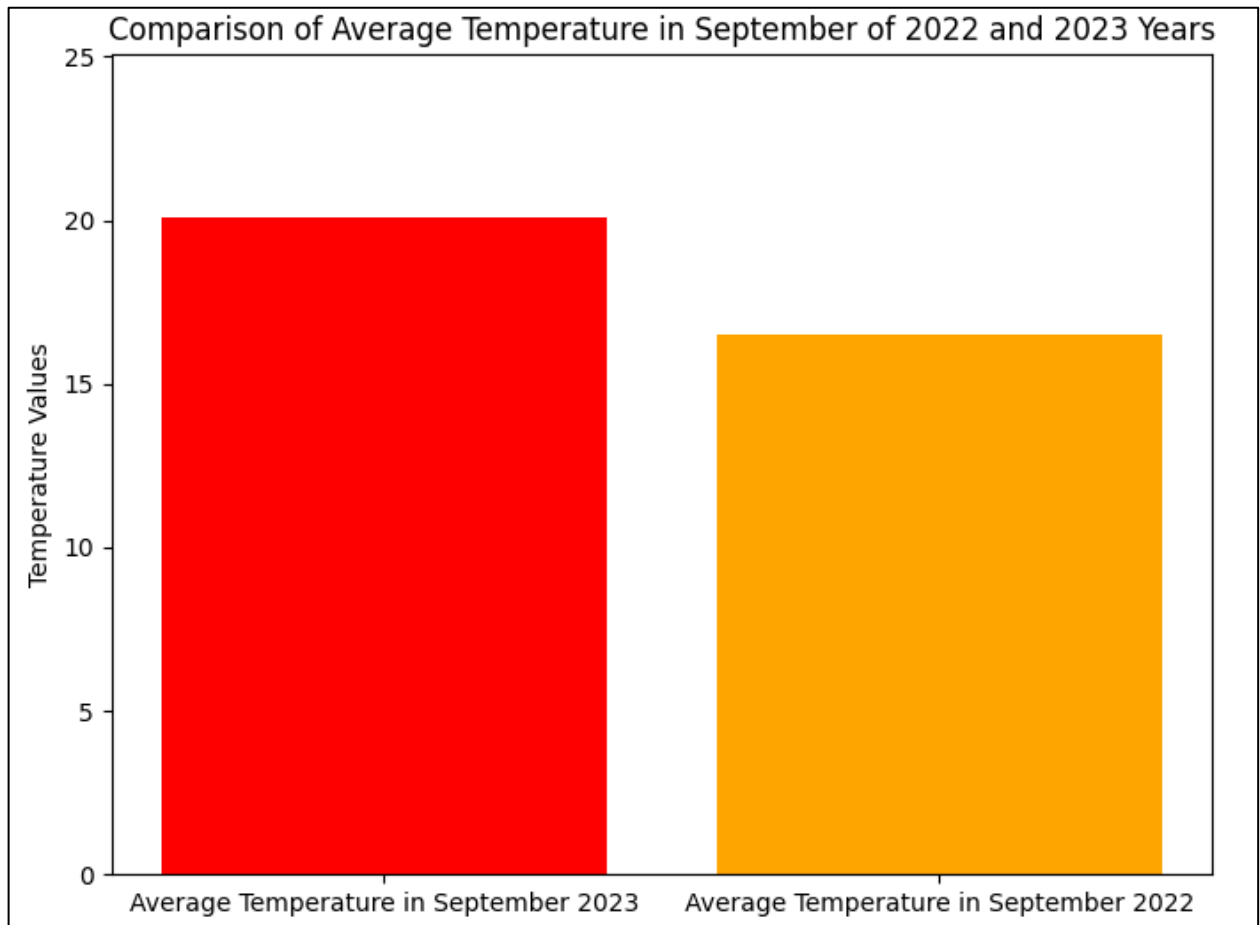


Рисунок 4.29 – Порівняння середньої температури повітря в вересні місяці за 2022 та 2023 роки в м. Вінниця

Тобто, як вже було відзначено раніше, середній рівень температури повітря відчутно підвищився. У нашому випадку, різниця між вереснем 2022 року та вереснем 2023 року становить 3,5 градусів Цельсія. Це досить значне збільшення температури може мати серйозний вплив на клімат та середовище міста. Такі зміни в температурному режимі можуть вплинути на ріст різних видів рослин, водні ресурси, а також здоров'я мешканців міста. Ці високі температури можуть також призвести до збільшення використання кондиціонерів та інших систем охолодження, що в свою чергу збільшить енергоспоживання та кількість викидів в атмосферу. Таким чином, ці зміни в кліматі потребують уважності та можливих

заходів для збалансування впливу на навколишнє середовище та комфорт мешканців міста.

Наступним етапом є побудова графіків, аналогічних тим, які представлені на рисунках 4.21-4.28, з включенням ліній, що відобразять встановлені норми для кожного з показників. Під час порівняння середніх значень із встановленими нормами було пропущено деякі важливі деталі, такі як аномальні значення та мінімуми. Тому тепер наші графіки будуть більш повними і точними, дозволяючи краще оцінити відхилення від стандартів та виявити будь-які незвичайні тенденції в даних (рис. 4.30-4.34).

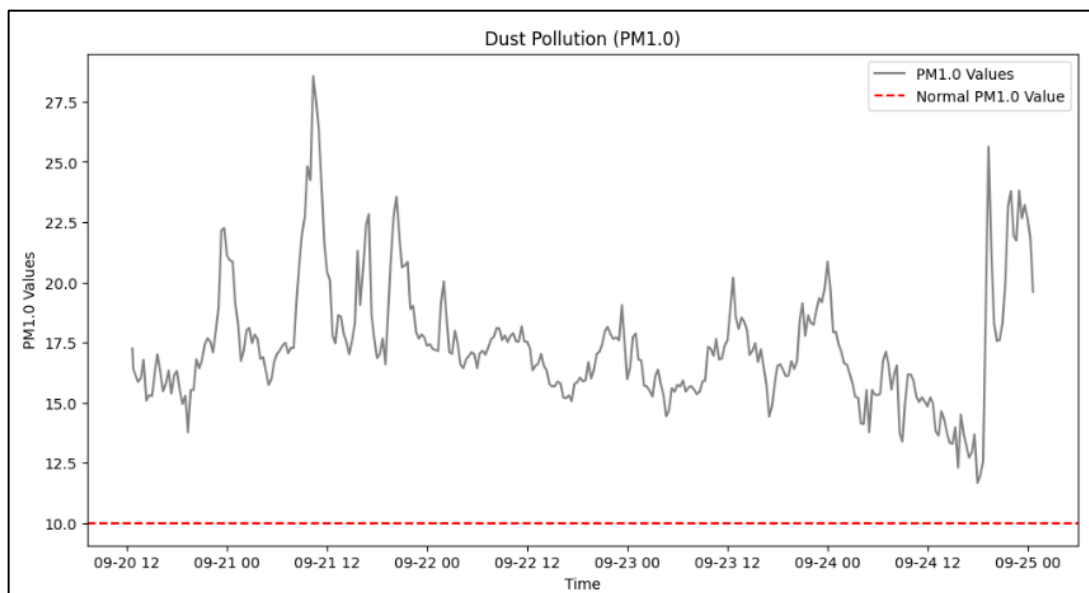


Рисунок 4.30 – Значення частинок пилу розміром 1 мкм відносно встановленої норми

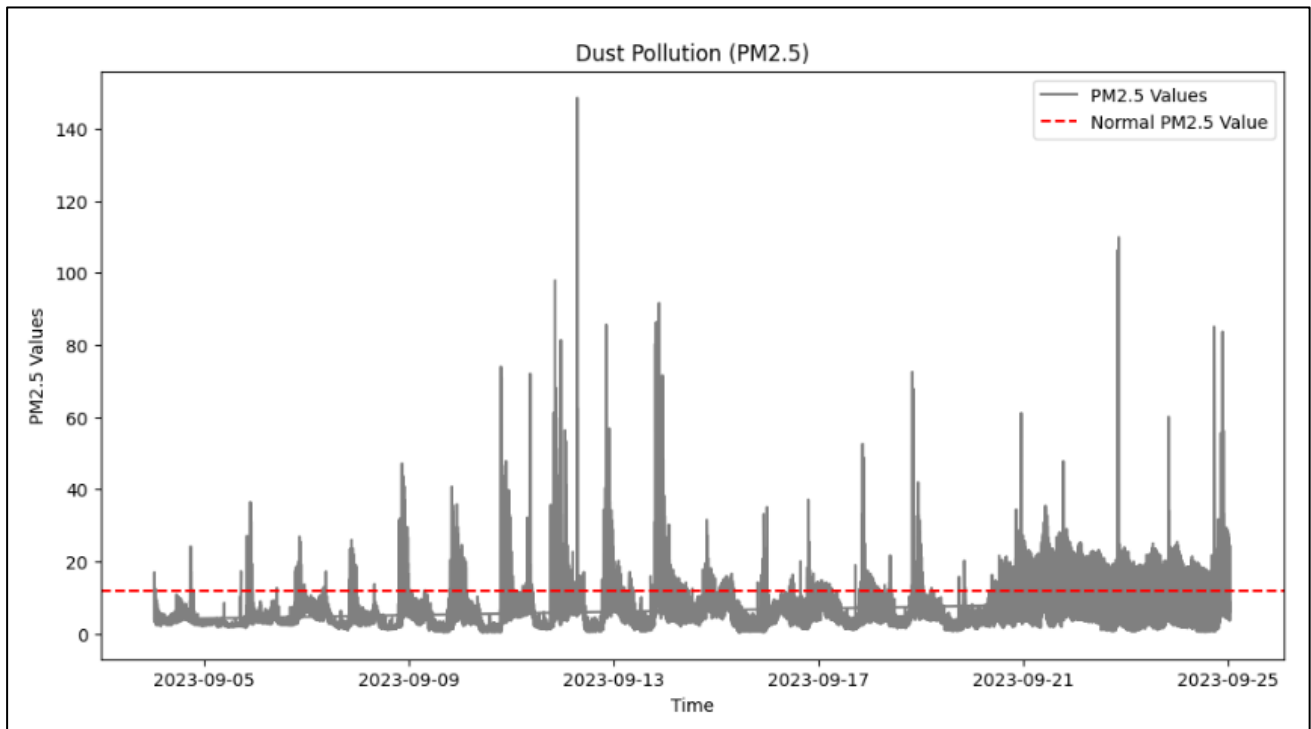


Рисунок 4.31 – Значення частинок пилу розміром 2,5 мкм відносно встановленої норми

Дивлячись на рисунки 4.30-4.31, можна зауважити, що протягом місяця значення показника  $PM_{10}$  мають постійне значне перевищення своєї встановленої норми. Проте, щодо показника  $PM_{2.5}$ , то ситуація трохи краща, адже через постійні коливання, значення цього показнику постійно варіюються, тобто мають як нормальні значення, так і перевищення норми.

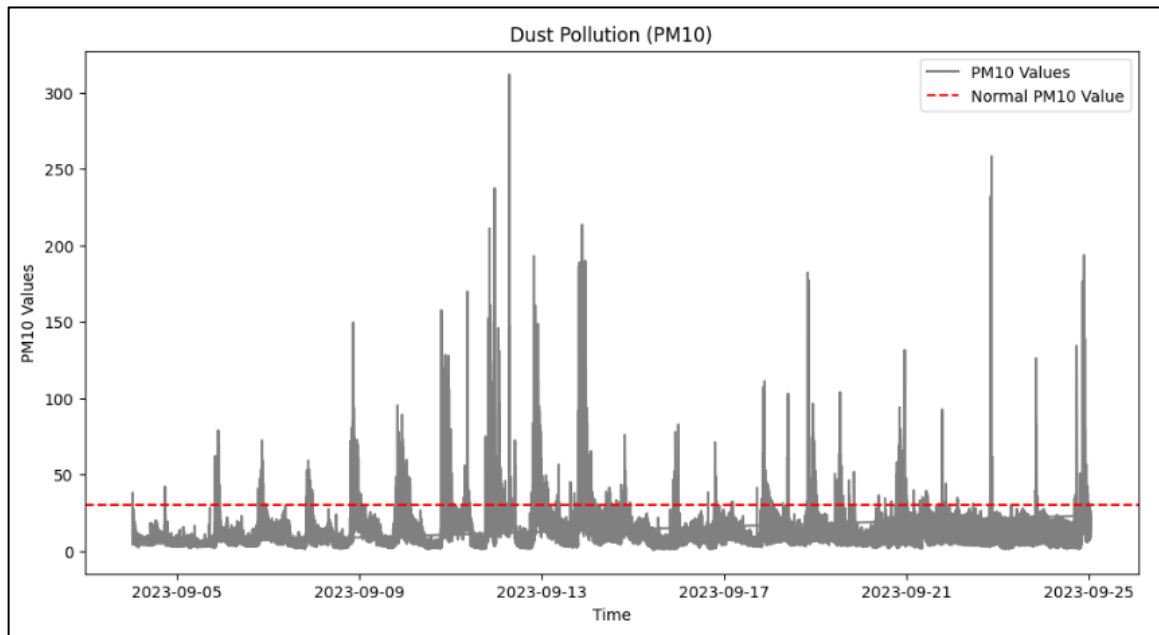


Рисунок 4.32 – Значення частинок пилу розміром 10 мкм відносно встановленої норми

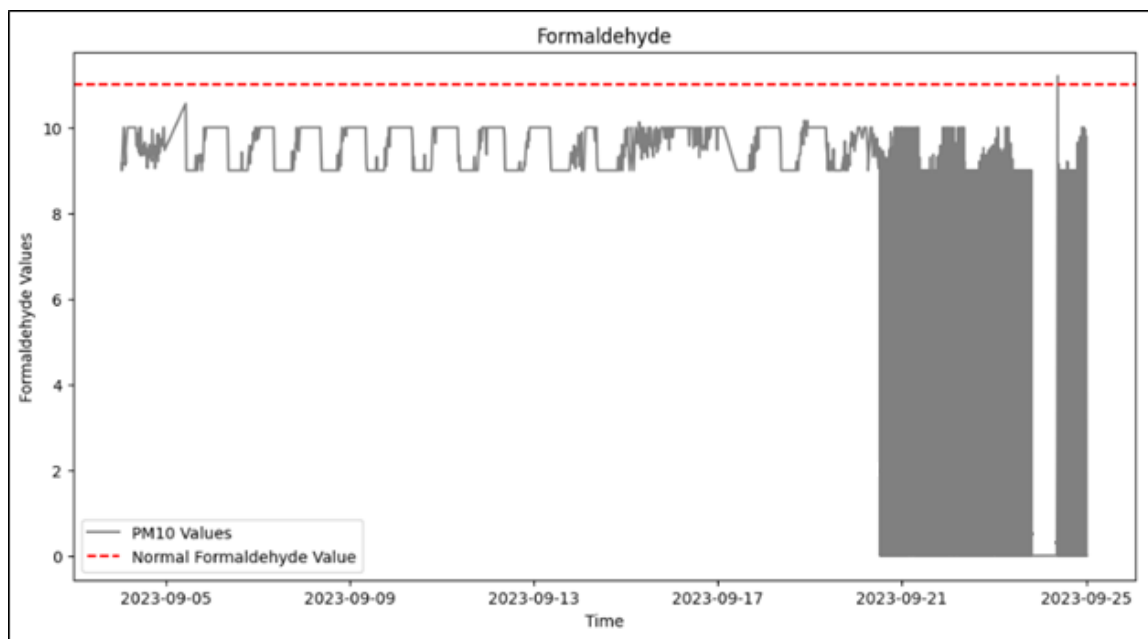


Рисунок 4.33 – Значення формальдегіду відносно встановленої норми

Після побудови аналогічних графіків для показників  $PM_{10}$  та формальдегіду (рис. 4.32-4.33), можна відзначити, що ситуація з показником пилу, де розмір частинок 10 мікрометрів, схожа на ситуацію з показником  $PM_{2.5}$ . Щодо формальдегіду, то спостерігаються досить позитивні значення, які не перевищують своєї норми.

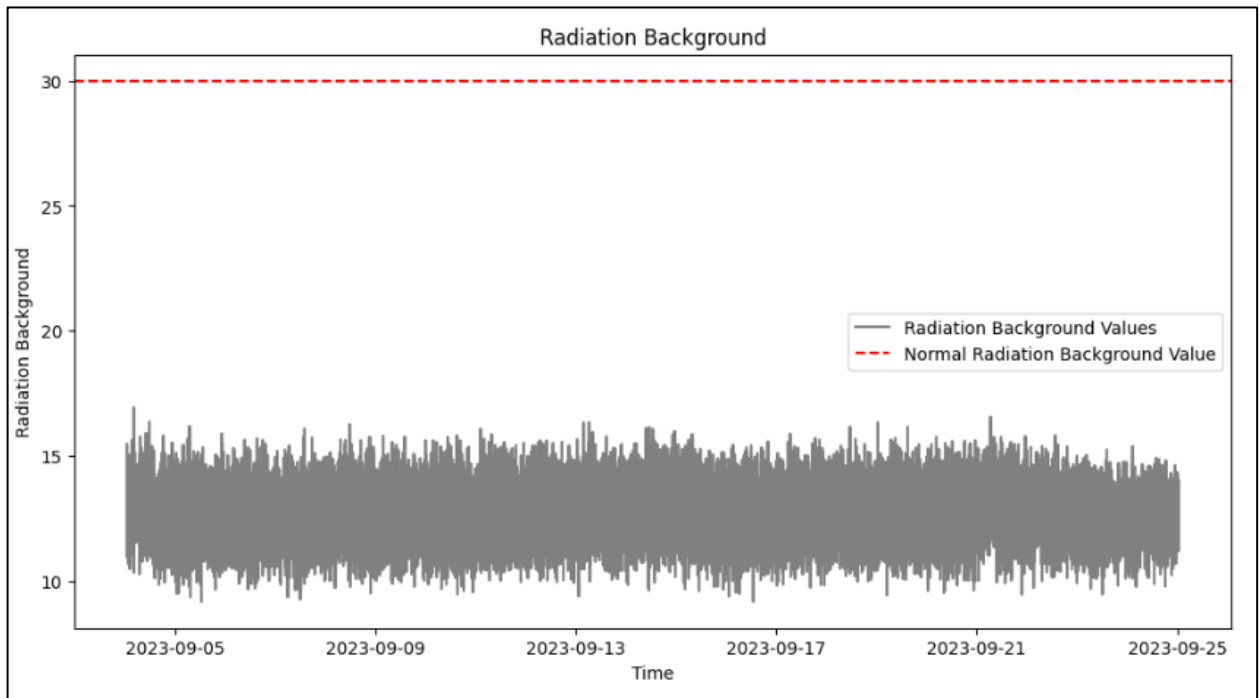


Рисунок 4.34 – Значення радіаційного фону відносно встановленої норми

Що ж, як зазначалося раніше, на рисунку 4.30 видно, що значення частинок пилу розміром 1 мкм значно перевищують встановлені норми для цього показника. Що стосується  $PM_{2,5}$  та  $PM_{10}$  (див. рис. 4.31-4.32), чітких змін не спостерігається – більшість даних знаходяться в межах норми, але спостерігається їх постійне різке коливання. Щодо інших досліджених показників, вони залишаються на задовільному рівні, не перевищуючи свої нормативи.

Додатковий аналіз показників є ключовим кроком у нашому дослідженні. Знаходження аномалій у значеннях може розкрити незвичайні тенденції або події, які можуть мати значущий вплив на якість повітря в місті Вінниця. Використовуючи мову програмування Python, було заплановано дослідити дані більш глибоко, виявляючи виняткові або відхилені значення, які впливають на розуміння загального контексту забруднення повітря. Ця детальна перевірка дозволить нам здійснити більш точні та обґрунтовані висновки щодо стану атмосферного повітря в місті та спрямує нас на правильний шлях для подальших досліджень (рис. 4.35-4.43).



```

# Calculate Z-scores
pm1_data['Z_Score'] = zscore(pm1_data['stations_params_value'])

# Define a threshold for Z-score (e.g., 2 or 3)
z_score_threshold = 2

# Identify anomalous data points
anomalous_pm1_data = pm1_data[abs(pm1_data['Z_Score']) > z_score_threshold]

# Plotting the entire dataset
plt.figure(figsize=(12, 7))
plt.scatter(pm1_data['stations_params_time'], pm1_data['stations_params_value'], label='PM1.0 Values', color='green', alpha=0.5)

# Plotting the anomalous data points
plt.scatter(anomalous_pm1_data['stations_params_time'], anomalous_pm1_data['stations_params_value'], label='Anomalous PM1.0 Values', color='red')

plt.xlabel('Time')
plt.ylabel('PM1.0 Values')
plt.title('PM1.0 Values with Anomalous Data Points Highlighted')
plt.legend()
plt.show()

```

Рисунок 4.35 – Приклад побудови графіку з відображенням аномальних значень

На рисунку 4.35 можна побачити приклад побудови графіку з відображенням аномалій, використовуючи статистичний метод Z-score. Цей метод визначає відстань між сумнівними точками і центром розподілу даних. Застосування Z-score дозволило точно виявити ті значення, які відхиляються від нормального розподілу та вважаються аномальними в контексті нашого дослідження.

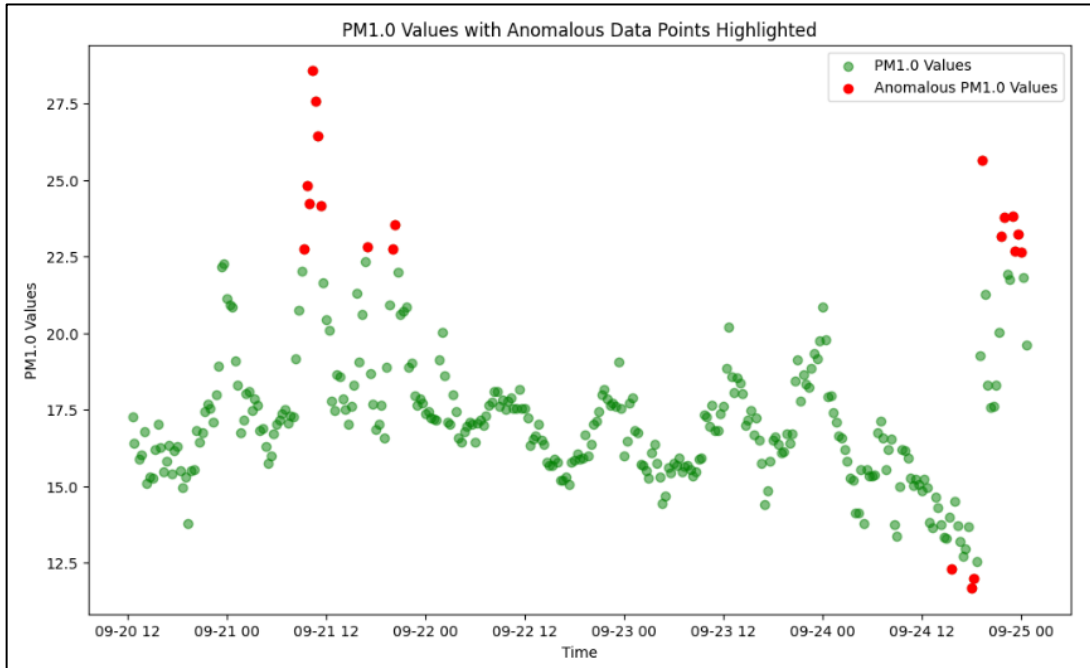


Рисунок 4.36 – Графік з відображенням аномальних значень показнику  $PM_1$

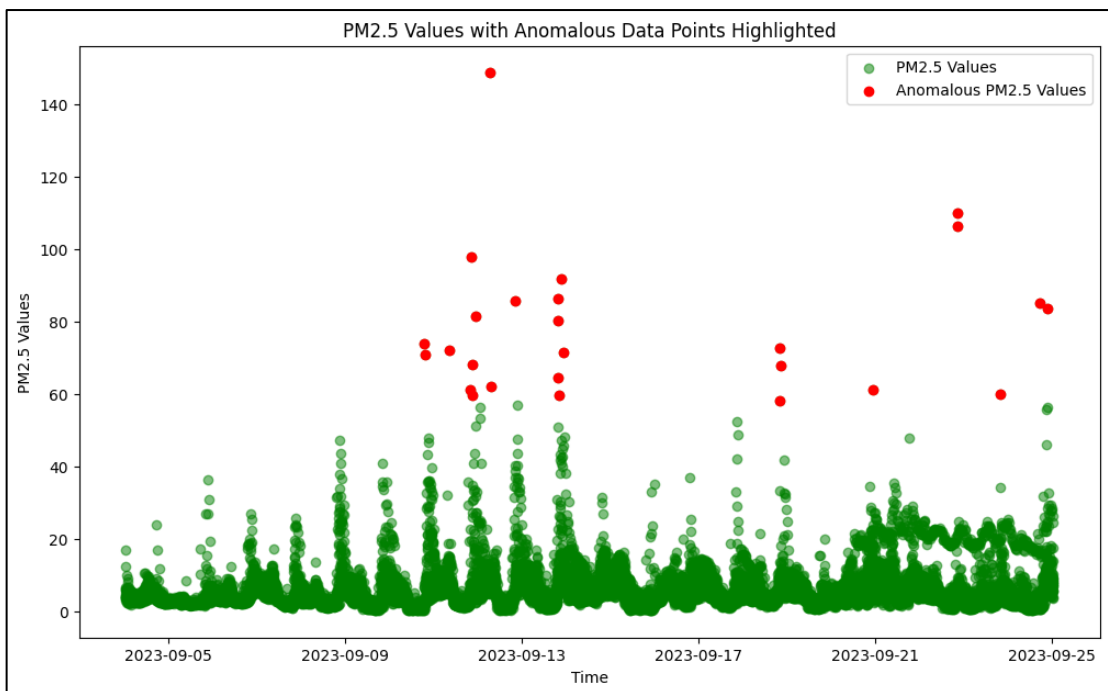


Рисунок 4.37 – Графік з відображенням аномальних значень показнику  $PM_{2,5}$

На рисунках 4.36-4.37 представлені графіки, побудовані з використанням мови програмування Python і методу Z-score. Ця візуалізація вказує на аномальні значення для кожного з показників (позначені червоним кольором), у нашому випадку,  $PM_1$  та  $PM_{2,5}$ .

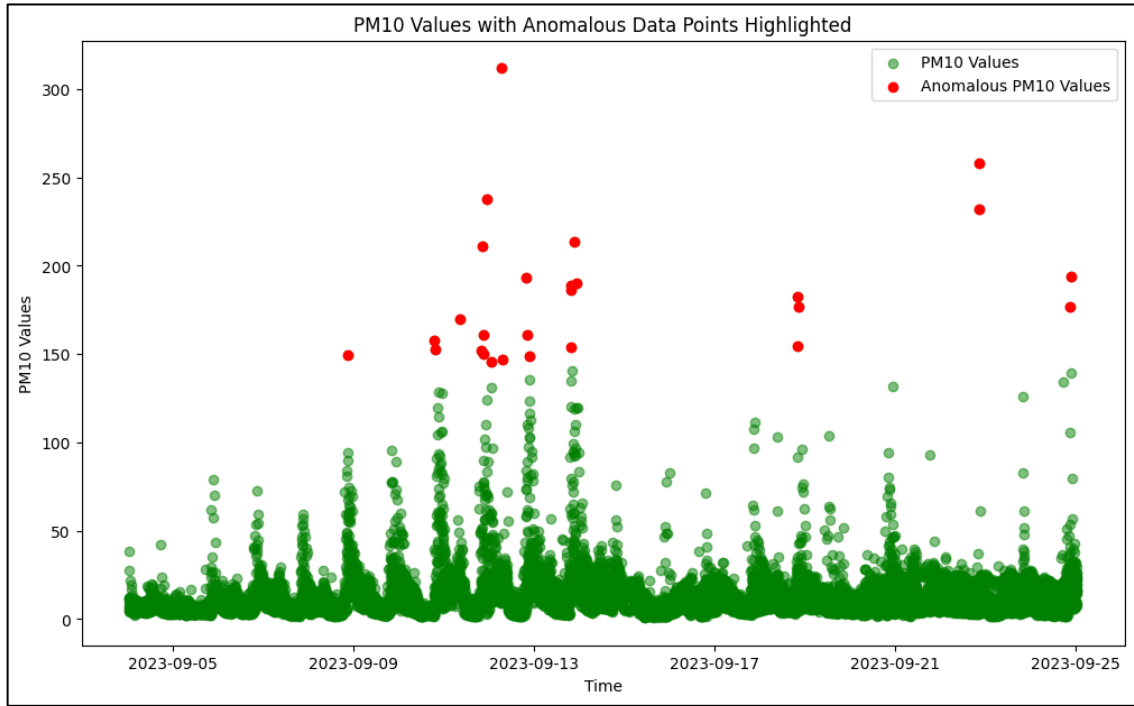


Рисунок 4.38 – Графік з відображенням аномальних значень показнику  $PM_{10}$

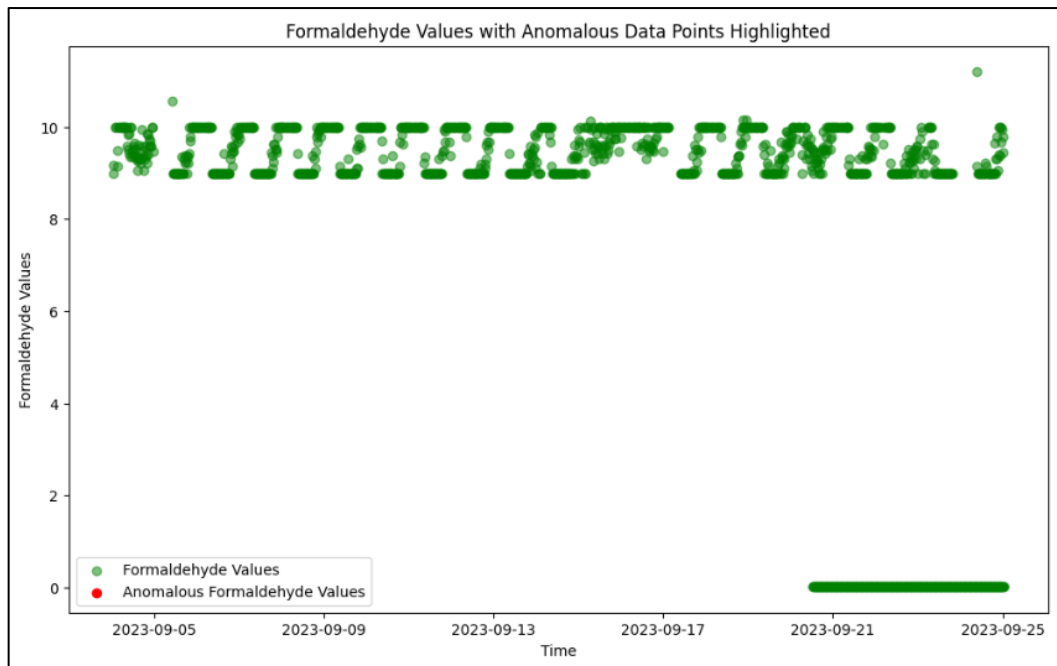


Рисунок 4.39 – Графік з відображенням аномальних значень показнику формальдегіду

На рисунках 4.38-4.39 зображені аналогічні графіки з відображенням аномальних значень, але для показників  $PM_{10}$  та формальдегіду.

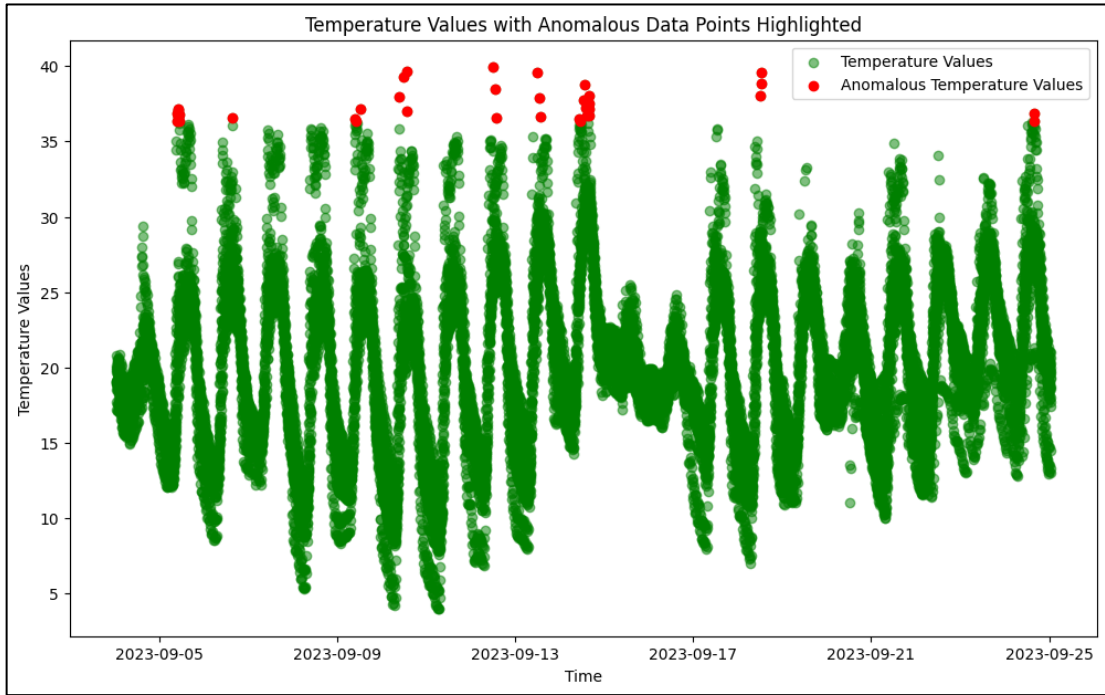


Рисунок 4.40 – Графік з відображенням аномальних значень показнику температури

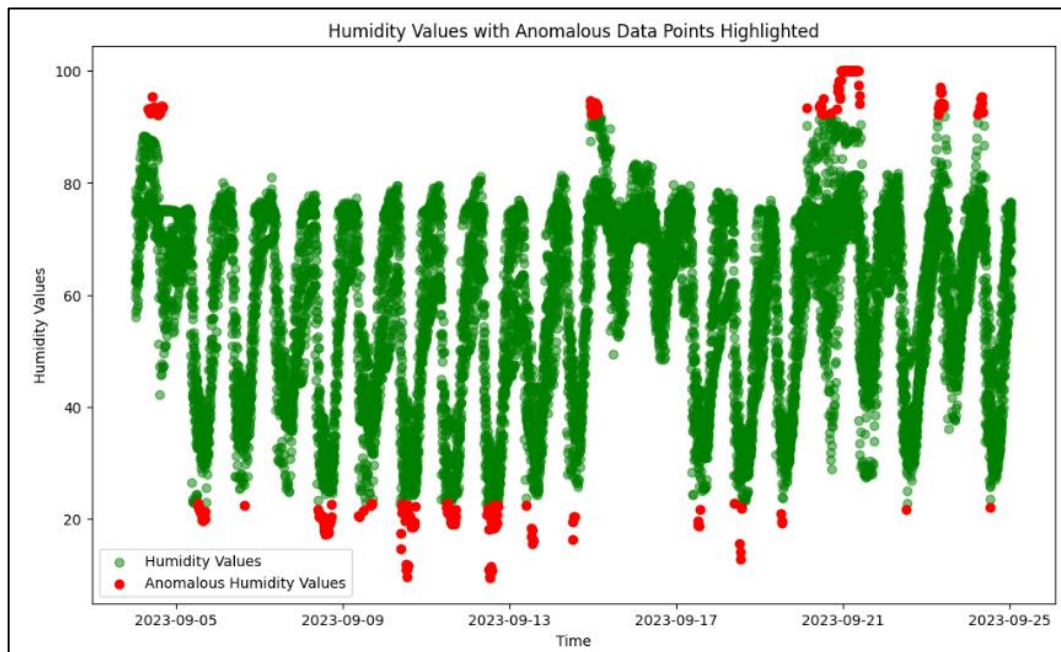


Рисунок 4.41 – Графік з відображенням аномальних значень показнику вологості

Спостерігаючи на рисунки 4.40-4.41, можна побачити графіки з відображенням аномальних значень показників вологості та температури повітря в особливості, ще раз доводячи про наявність аномальної жари цього року.

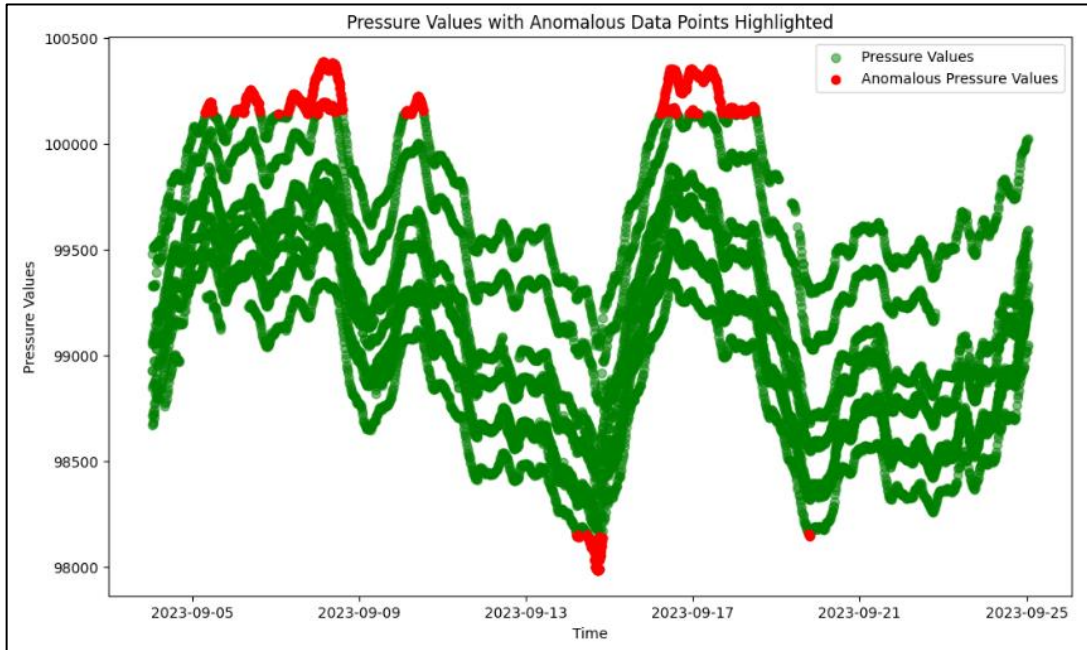


Рисунок 4.42 – Графік з відображенням аномальних значень показнику атмосферного тиску

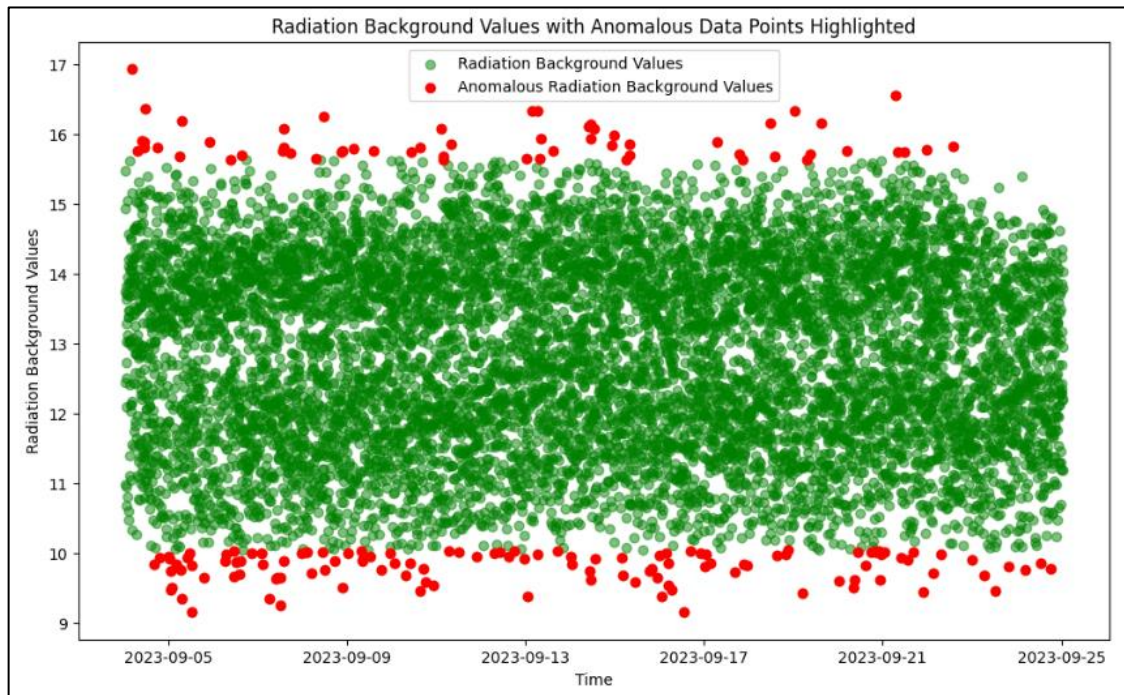


Рисунок 4.43 – Графік з відображенням аномальних значень показнику радіаційного фону

Спостерігаючи за рисунками 4.36-4.43, явно видно графіки, які ілюструють аномалії у різних досліджуваних показниках. Цікаво, що серед усіх показників

найстабільнішим є рівень формальдегіду, який не має жодних аномальних значень, але як було зазначено під час попередніх аналізів, цьогорічна аномальна спека призвела до зростання температури, і це відобразилося на даних, зображення яких можна побачити на рисунку 4.40. Знаходячи аномалії (див. рис. 4.40), можна підтвердити, що їх було надзвичайно багато цього року. Щодо інших показників, спостерігається більш стабільна ситуація, на відмінну від показників пилу, які, як і температура, мають значну кількість аномальних значень (див. рис. 4.36-3.38). Це може свідчити про те, що ці значення не є окремими випадками, але, навпаки, вказують на їх стабільні та системні перевищення. Отже, забруднення повітря пилом залишається однією з найгостріших проблем серед всіх інших показників, які були розглянуті до цього моменту.

## **4.2 Прогнозування стану атмосферного повітря за даними громадського моніторингу в місті Вінниця**

Після глибокого аналізу даних та врахування результатів наших досліджень, перейдемо до етапу прогнозування, який дозволить нам створити модель, яка здатна імітувати дані на певний період у майбутньому. Це надасть можливість приблизно оцінити якість повітря в місті Вінниця на наступний місяць, тим самим, показуючи нам, які можуть бути рівні концентрацій певних забруднюючих речовин в повітрі.

Для виконання прогнозування даних було використано модель ARIMA. Дана модель є досить популярною в сфері прогнозу часових рядів та яка часто використовується в статистиці та економетриці [27]. Використовуючи мову Python, побудуємо модель ARIMA та за її допомогою спрогнозуємо ряд показників (рис. 4.44-4.54).

```

# Create the ARIMA model
model = sm.tsa.ARIMA(df_air_pm1['values'], order = (1, 0, 1))

# Fit the model
model_fit = model.fit()
print(model_fit.summary)

<bound method SARIMAXResults.summary of <statsmodels.tsa.arima.model.ARIMAResults object at 0x7f
aa26fab700>>

# Get the forecasted values
predictions = model_fit.predict()

# Plot the actuals and forecasted values
plt.figure(figsize = (12, 5), dpi = 100)
plt.plot(df_air_pm1.index, df_air_pm1['values'], label = 'Actual')
plt.plot(df_air_pm1.index, predictions, label = 'Forecast')
plt.title('Forecast vs Actuals by ARIMA')
plt.legend(loc = 'upper left', fontsize = 8)
plt.show()

```

Рисунок 4.44 – Створення ARIMA моделі та побудова графіка прогнозу значень на прикладі показнику  $PM_1$

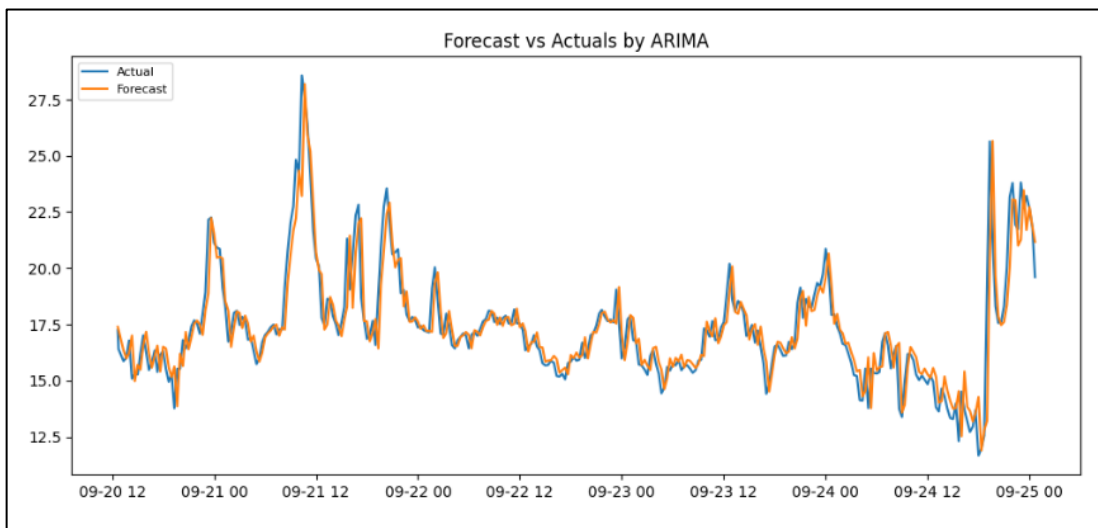


Рисунок 4.45 – Графік прогнозу частинок пилу розміром 1 мкм

Згідно рисунку 4.45, можна побачити результат побудови графіку прогнозу, у данному випадку за показником частинок пилу розміром 1 мкм.

```
# Calculate R-squared score
r2 = r2_score(df_air_pm1['values'], predictions)
rounded_r2_pm1 = round(r2, 2)

print("R-squared score:", rounded_r2_pm1)

R-squared score: 0.82
```

Рисунок 4.46 – Показник точності прогнозу частинок пилу розміром 1 мкм

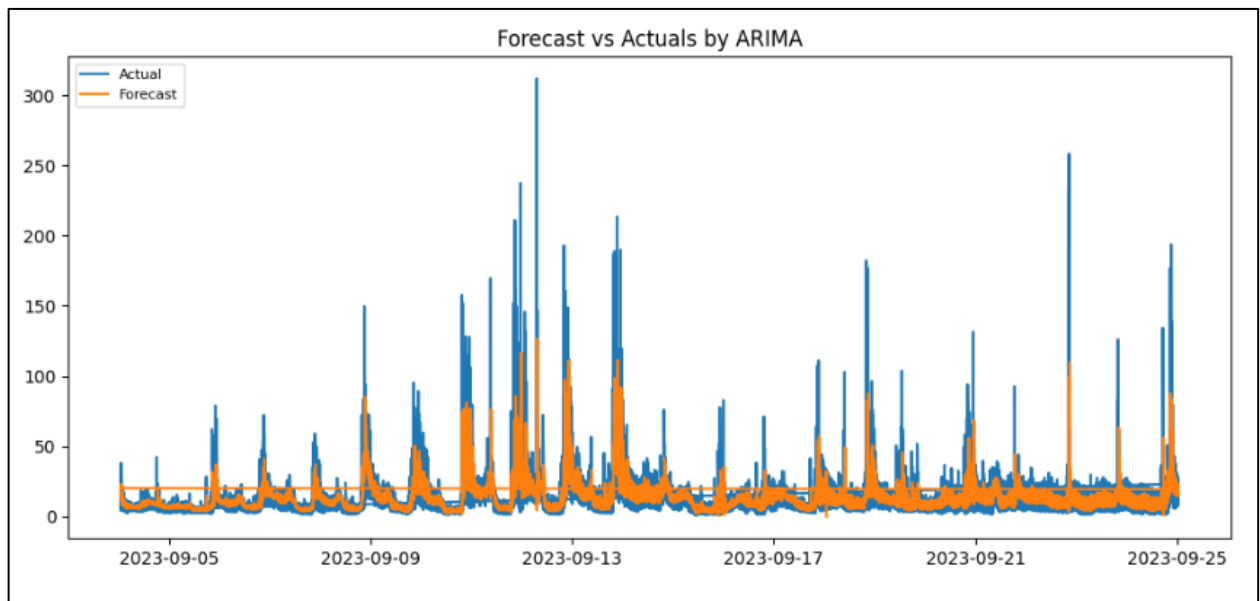


Рисунок 4.47 – Графік прогнозу частинок пилу розміром 10 мкм

```
# Calculate R-squared score
r2 = r2_score(df_air_pm10['values'], predictions)
rounded_r2_pm10 = round(r2, 2)

print("R-squared score:", rounded_r2_pm10)

R-squared score: 0.41
```

Рисунок 4.48 – Показник точності прогнозу частинок пилу розміром 10 мкм

На вищевказаних рисунках 4.46-4.48, продемонстровано графік прогнозу частинок пилу розміром 10 мкм та результати точності, у вигляді метрики R2 score, яка демонструє показник ефективності роботи моделі ARIMA.



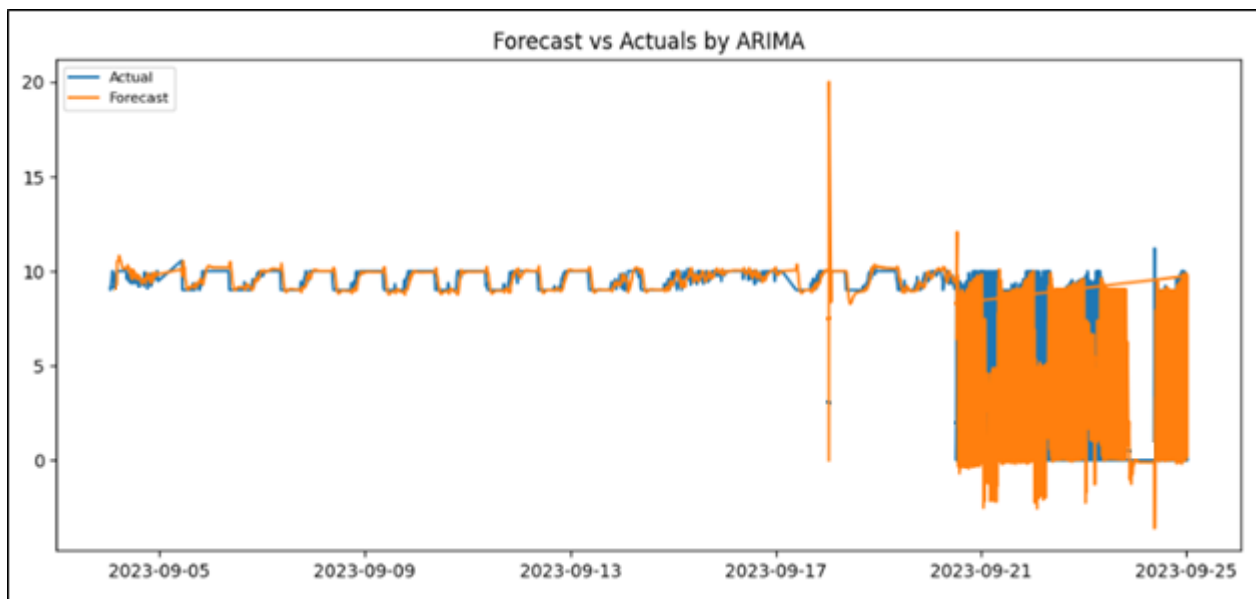


Рисунок 4.49 – Графік прогнозу формальдегіду

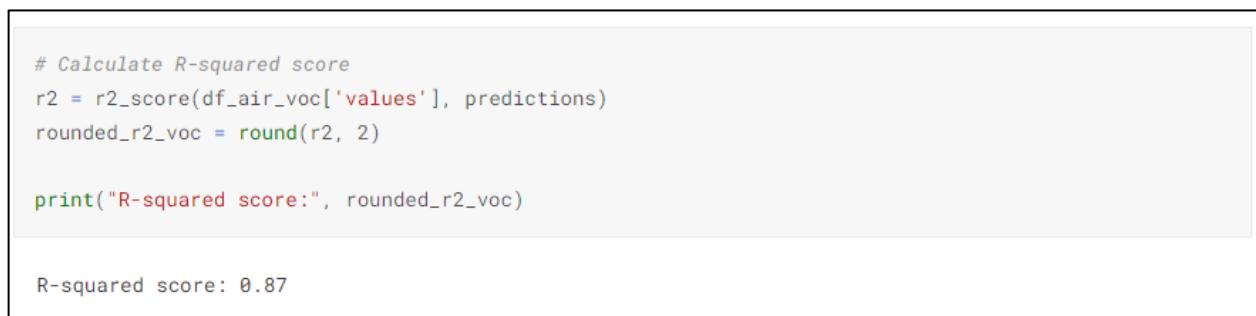


Рисунок 4.50 – Показник точності прогнозу формальдегіду

На графіку вище (рис. 4.49) зображено результат прогнозування за показником формальдегіду, де точність роботи моделі ARIMA стосовно даного показнику можна побачити на рисунку 4.50.

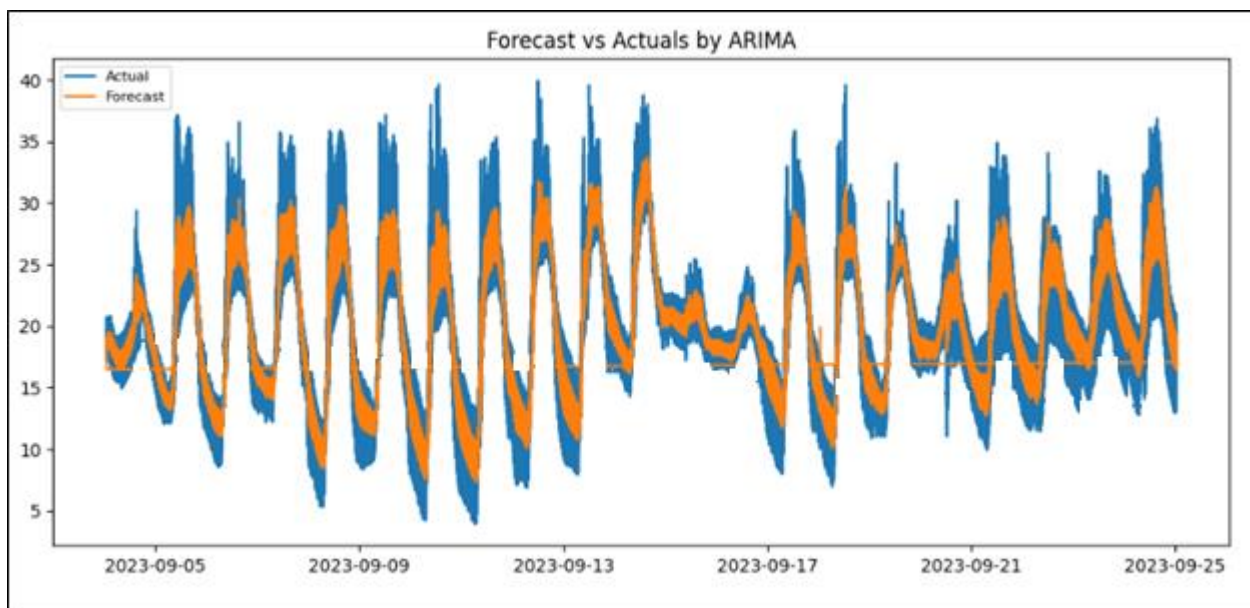


Рисунок 4.51 – Графік прогнозу температури

```
# Calculate R-squared score
r2 = r2_score(df_air_temperature['values'], predictions)
rounded_r2_temperature = round(r2, 2)

print("R-squared score:", rounded_r2_temperature)

R-squared score: 0.82
```

Рисунок 4.52 – Показник точності прогнозу температури

Таким же чином, на рисунку 4.51 можна побачити побудований графік прогнозу, але стосовно показника температури, де точність поточного прогнозу становить 0,82 (див. рис. 4.52).

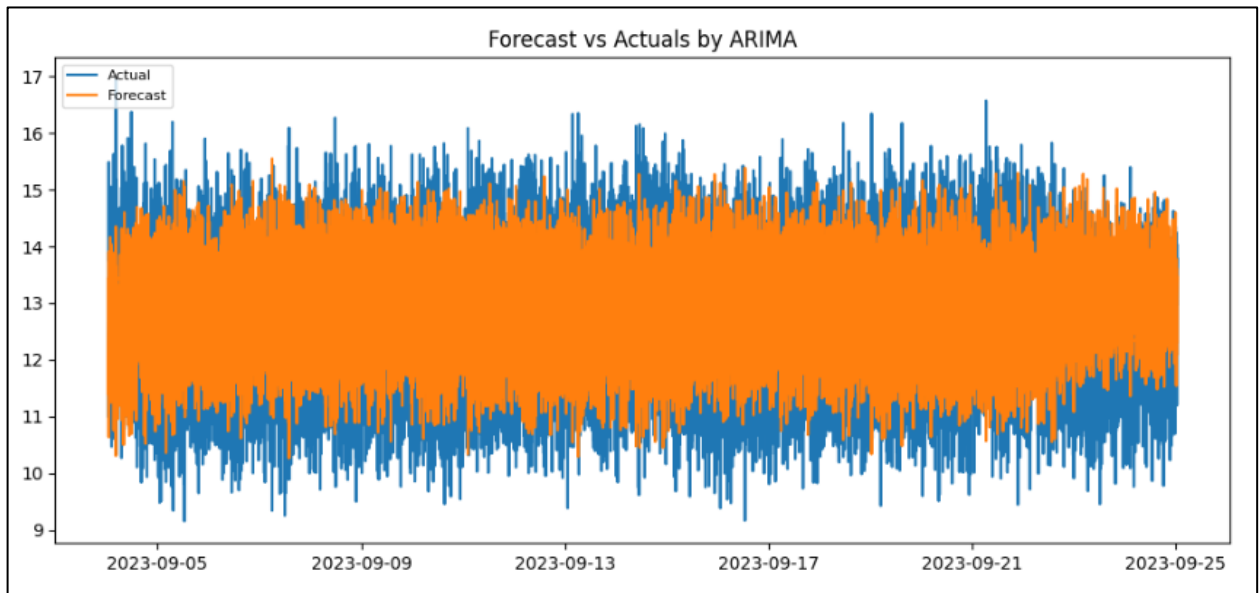


Рисунок 4.53 – Графік прогнозу радіаційного фону

```
# Calculate R-squared score
r2 = r2_score(df_air_radiation['values'], predictions)
rounded_r2_radiation = round(r2, 2)

print("R-squared score:", rounded_r2_radiation)
```

R-squared score: 0.52

Рисунок 4.54 – Показник точності прогнозу радіаційного фону

На рисунках 4.44-4.54 представлені результати прогнозування показників, використовуючи модель прогнозування часових рядів ARIMA. У висновках можна виділити наступне: найнижче прогнозоване значення складає 0,41 для показника  $PM_{10}$ , тоді як найвище значення – 0,87 для показника формальдегіду. Важливо відзначити, що середня точність ARIMA з оптимальними налаштуваннями складає 0,68, що є дуже задовільним результатом для прогнозування даних.

Хоча загальна точність нашої моделі сягає приблизно 70%, аналіз результатів дозволяє вбачити незначне, але зменшення концентрації вивчених речовин в наступному місяці. Важливо врахувати, що з приходом холодів спостерігається зниження кількості пилу в повітрі, а також концентрації інших

речовин. Це дає підстави вважати, що із значною вірогідністю якість повітря в місті покращиться в наступному місяці.

Далі, для порівняння ефективності моделі ARIMA, виконаємо прогноз за допомогою іншої моделі, а саме Prophet. Дана модель прогнозування реалізована не лише в Python, а й в мові програмування R [28]. Prophet є досить швидкою моделлю і забезпечує повністю автоматизовані прогнози, які налаштовуються вручну під час обробки даних. Результати прогнозування за допомогою моделі Prophet можна буде побачити на рисунках 4.55-4.61.

```
df_prophet = pd.DataFrame()
df_prophet['ds'] = df_air_pm1_prophet['time'] # Replace your column with the actual column name
df_prophet['y'] = df_air_pm1_prophet['values'] # Replace your column with the actual column name

# Build Prophet model with parameters and structure
model = Prophet(daily_seasonality = True, weekly_seasonality = False, yearly_seasonality = False,
                changepoint_range = 0.7, changepoint_prior_scale = 0.4, seasonality_mode = 'multiplicative')

# Training model for df_air
model.fit(df_prophet)

07:50:37 - cmdstanpy - INFO - Chain [1] start processing
07:50:38 - cmdstanpy - INFO - Chain [1] done processing

<prophet.forecaster.Prophet at 0x7f7ea3f06d40>

# Create a DataFrame with future dates for forecasting
future = model.make_future_dataframe(periods = 20) # Adjust the number of periods as needed

# Generate forecast
forecast = model.predict(future)

# Plot the forecast
fig = model.plot(forecast)
plt.show()
```

Рисунок 4.55 – Побудова моделі прогнозування Prophet, на прикладі показнику

PM1

На рисунку 4.55 зображено процес побудови та тренування моделі Prophet, використовуючи показник пилу, а саме частинки пилу розміром 1 мікромметр. Варто відзначити, що в налаштуваннях обрано прогнозування на наступні 20 днів. З метою максимально точного прогнозу, враховуючи наявні дані за вересень, модель Prophet дозволяє нам передбачити значення майже на місяць вперед. Хоча дана модель не обмежує час прогнозування, в нашому випадку 20 днів є найбільш оптимальним терміном, враховуючи початкові дані. Це дозволяє спрогнозувати значення частинок пилу розміром 1 мкм. На 20 днів вперед, забезпечуючи при цьому більш точний прогноз.

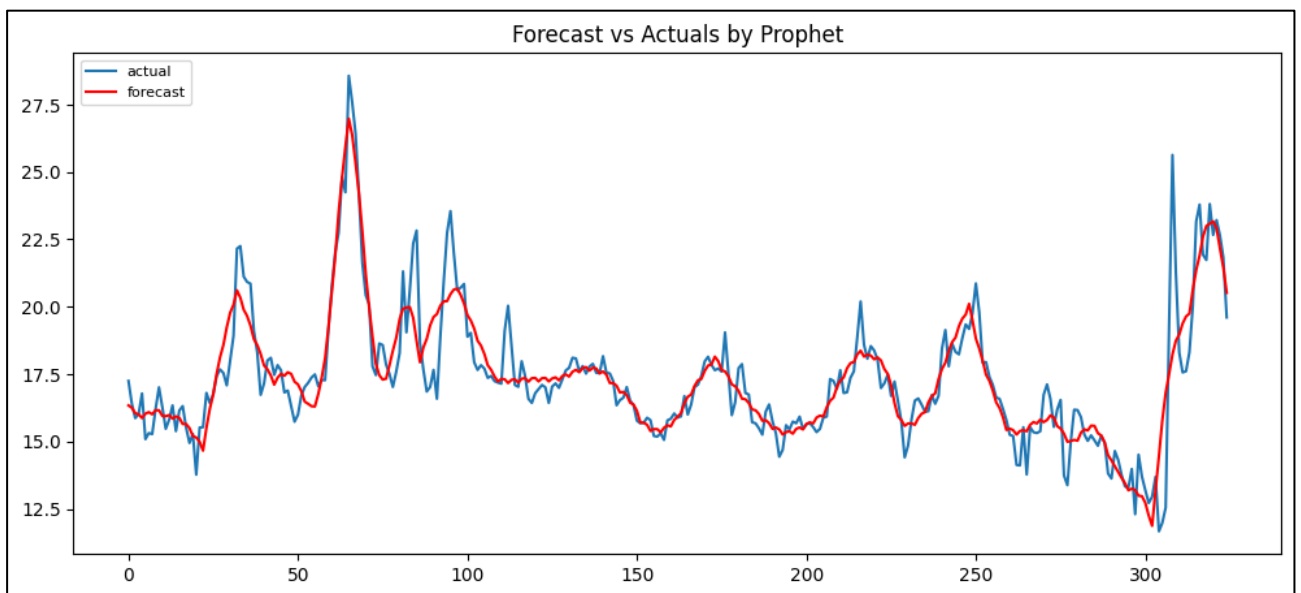


Рисунок 4.56 – Графік прогнозування значення частинок пилу розміром 1 мкм за допомогою моделі Prophet

```

# Assuming df_air_radiation_prophet contains your actual training data
# Merge forecasted values with actual values based on the date column (ds)
df_comparison = pd.merge(df_forecast, df_air_pm1_prophet, left_on = 'ds', right_on = 'time', how =
'left')

# Drop rows with missing actual values (NaN)
df_comparison = df_comparison.dropna(subset = ['values'])

# Calculate R-squared score
r2 = r2_score(df_comparison['values'], df_comparison['yhat'])
print("R-squared score:", round(r2, 2))

```

R-squared score: 0.83

Рисунок 4.57 – Показник точності прогнозу частинок пилу розміром 1 мкм

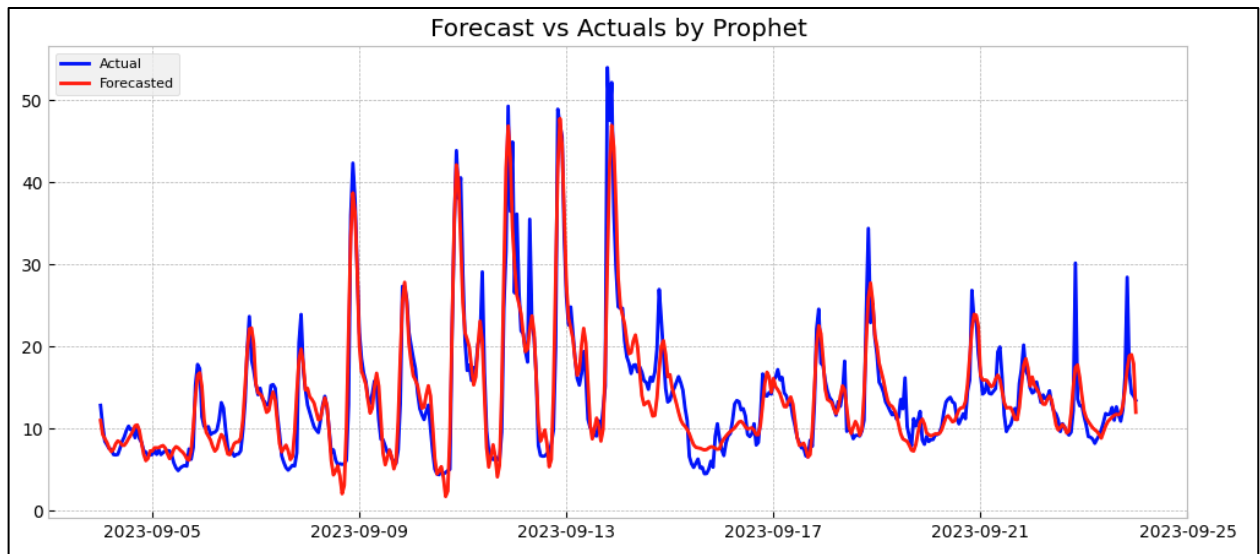


Рисунок 4.58 – Графік прогнозування значення частинок пилу розміром 10 мкм за допомогою моделі Prophet

Глянувши на рисунки 4.56 та 4.58, можна спостерігати графіки прогнозування за показниками пилу розміром частинок яких сягає 1 та 10 мікрометрів відповідно, але при цьому користуючись іншою моделю, а саме Prophet.

```
r2 = r2_score(pm10_train_data['y'], b['yhat'])
print(f'R-squared score: {r2}')
```

R-squared score: 0.873570184599277

Рисунок 4.59 – Показник точності прогнозу частинок пилу розміром 10 мкм

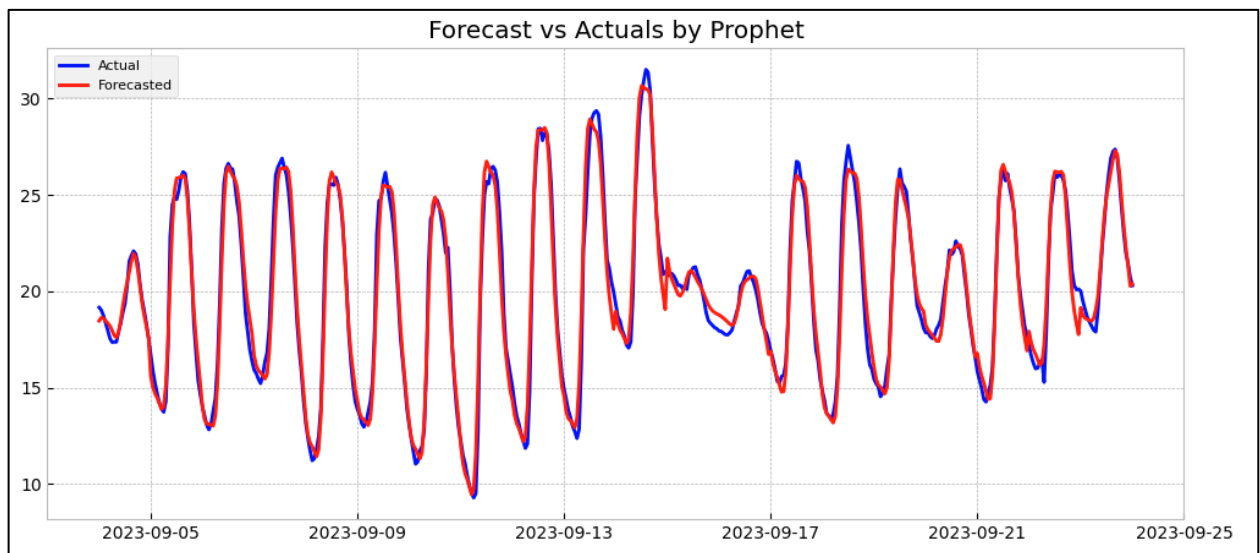


Рисунок 4.60 – Графік прогнозування значення температури за допомогою моделі Prophet

На графіку 4.60 відображено прогнозні результати для показника температури, які були здійснені за допомогою моделі Prophet.

```
r2 = r2_score(df_air_pm10_prophet['y'], b['yhat'])
print(f'R-squared score: {r2}')
```

R-squared score: 0.971896779716758

Рисунок 4.61 – Показник точності прогнозу температури

Аналізуючи рисунки 4.55-4.61, можна оцінити результати роботи моделі Prophet. За отриманими даними видно, що найточніший прогноз було отримано для показника температури (0,97), тоді як найменший прогноз спостерігався для радіаційного фону (0,23). Середній показник точності прогнозування складає приблизно 80%. З цього можна зробити висновок, що на основі наших даних

більш виправдано використовувати модель Prophet, адже різниця в роботі цих моделей становить приблизно 5%, з середнім показником точності прогнозування 70% для ARIMA і 75% для Prophet відповідно.

### 4.3 Висновки

У цьому розділі проведено системний аналіз забруднення атмосферного повітря в м. Вінниця. В ході дослідження були розглянуті різні забруднюючі показники, і наразі загальний рівень якості повітря у місті є задовільним, за винятком одного показника –  $PM_{10}$ . Серед різних аспектів забруднення, саме частинки пилу розміром 1 мкм перевищують норму найбільше, що вказує на необхідність удосконалення заходів щодо зменшення пилу та чинників, які провокують збільшення пилу в повітрі загалом.

Крім того, виявлено надзвичайне потепління, оскільки середня температури у вересні 2023 року виявилася на 3,5 градусів Цельсія вищою, ніж у минулому році того ж місяця. Ці результати є важливими для подальшого вдосконалення стратегій збереження якості повітря та взаємодії з навколишнім середовищем.

Також, проведено короткий порівняльний аналіз даних за даними державного моніторингу в місті Вінниця та Вінницькій області, й помічено, що зрівнюючи усі отримані значення показників за період з 2015-2023 роки, попри усі вище зазначені проблеми, можна спостерігати покращення якості повітря як в місті Вінниця, так і поза його межами.

При створенні інформаційної технології аналізу та прогнозування стану атмосферного повітря міста Вінниці, використовуючи мову програмування Python, побудовано модель ARIMA та Prophet, де точність роботи яких під час нашого дослідження, а саме прогнозуванні ряду досліджуваних показників, становила близько 75%. Згідно отриманих з прогнозу результатів, спостерігається можливе покращення якості повітря в місті на наступний місяць.



## **5 ЕКОНОМІЧНА ЧАСТИНА**

Науково-технічна розробка має право на існування та впровадження, якщо вона відповідає вимогам часу, як в напрямку науково-технічного прогресу та і в плані економіки. Тому для науково-дослідної роботи необхідно оцінювати економічну ефективність результатів виконаної роботи.

Магістерська кваліфікаційна робота з розробки та дослідження «Інформаційна технологія аналізу та прогнозування стану атмосферного повітря міста Вінниці. Частина 1. Формування датасету та системний аналіз даних» відноситься до науково-технічних робіт, які орієнтовані на виведення на ринок (або рішення про виведення науково-технічної розробки на ринок може бути прийнято у процесі проведення самої роботи), тобто коли відбувається так звана комерціалізація науково-технічної розробки. Цей напрямок є пріоритетним, оскільки результатами розробки можуть користуватися інші споживачі, отримуючи при цьому певний економічний ефект. Але для цього потрібно знайти потенційного інвестора, який би взявся за реалізацію цього проекту і переконати його в економічній доцільності такого кроку.

Для наведеного випадку нами мають бути виконані такі етапи робіт:

- 1) проведено комерційний аудит науково-технічної розробки, тобто встановлення її науково-технічного рівня та комерційного потенціалу;
- 2) розраховано витрати на здійснення науково-технічної розробки;
- 3) розрахована економічна ефективність науково-технічної розробки у випадку її впровадження і комерціалізації потенційним інвестором і проведено обґрунтування економічної доцільності комерціалізації потенційним інвестором.

### **5.1 Проведення комерційного та технологічного аудиту науково-технічної розробки**

Метою проведення комерційного і технологічного аудиту дослідження за темою «Інформаційна технологія аналізу та прогнозування стану атмосферного

повітря міста Вінниці. Частина 1. Формування датасету та системний аналіз даних» є оцінювання науково-технічного рівня та рівня комерційного потенціалу розробки, створеної в результаті науково-технічної діяльності.

Оцінювання науково-технічного рівня розробки та її комерційного потенціалу рекомендується здійснювати із застосуванням 5-ти бальної системи оцінювання за 12-ма критеріями, наведеними в табл. 5.1 [29].

Таблиця 5.1 – Рекомендовані критерії оцінювання науково-технічного рівня і комерційного потенціалу розробки та бальна оцінка

Бали (за 5-ти бальною шкалою)					
	0	1	2	3	4
Технічна здійсненність концепції					
	Достовірність концепції не підтверджена	Концепція підтверджена експертними висновками	Концепція підтверджена розрахунками	Концепція перевірена на практиці	Перевірено працездатність продукту в реальних
Ринкові переваги (недоліки)					
	Багато аналогів на малому ринку	Мало аналогів на малому ринку	Кілька аналогів на великому ринку	Один аналог на великому ринку	Продукт не має аналогів на великому
	Ціна продукту значно вища за ціни аналогів	Ціна продукту дещо вища за ціни аналогів	Ціна продукту приблизно дорівнює цінам аналогів	Ціна продукту дещо нижче за ціни аналогів	Ціна продукту значно нижче за ціни аналогів
	Технічні та споживчі властивості продукту значно гірші, ніж в аналогів	Технічні та споживчі властивості продукту трохи гірші, ніж в аналогів	Технічні та споживчі властивості продукту на рівні аналогів	Технічні та споживчі властивості продукту трохи кращі, ніж в аналогів	Технічні та споживчі властивості продукту значно кращі, ніж в аналогів

Продовження таблиці 5.1

Бали (за 5-ти бальною шкалою)					
	0	1	2	3	4
	Експлуатаційні витрати і витрати значно вищі, ніж в аналогів	Експлуатаційні витрати дещо вищі, ніж в аналогів	Експлуатаційні витрати на рівні експлуатаційних витрат аналогів	Експлуатаційні витрати трохи нижчі, ніж в аналогів	Експлуатаційні витрати значно нижчі, ніж в аналогів
Ринкові перспективи					
	Ринок малий і не має позитивної динаміки	Ринок малий, але має позитивну динаміку	Середній ринок з позитивною динамікою	Великий стабільний ринок	Великий ринок з позитивною динамікою
	Активна конкуренція великих компаній на ринку	Активна конкуренція	Помірна конкуренція	Незначна конкуренція	Конкуренція немає
Практична здійсненність					
	Відсутні фахівці як з технічної, так і з комерційної реалізації ідеї	Необхідно наймати фахівців або витратити значні кошти та час на навчання наявних фахівців	Необхідне незначне навчання фахівців та збільшення їх штату	Необхідне незначне навчання фахівців	Є фахівці з питань як з технічної, так і з комерційної реалізації ідеї
	Потрібні значні фінансові ресурси, які відсутні. Джерела фінансування ідеї відсутні	Потрібні незначні фінансові ресурси. Джерела фінансування відсутні	Потрібні значні фінансові ресурси. Джерела фінансування є	Потрібні незначні фінансові ресурси. Джерела фінансування є	Не потребує додаткового фінансування

Продовження таблиці 5.1

Бали (за 5-ти бальною шкалою)					
	0	1	2	3	4
0	Необхідна розробка нових матеріалів	Потрібні матеріали, що використовуються у військово-промисловому комплексі	Потрібні дорогі матеріали	Потрібні досяжні та дешеві матеріали	Всі матеріали для реалізації ідеї відомі та давно використовуються у виробництві
1	Термін реалізації ідеї більший за 10 років	Термін реалізації ідеї більший за 5 років. Термін окупності інвестицій більше 10-ти років	Термін реалізації ідеї від 3-х до 5-ти років. Термін окупності інвестицій більше 5-ти років	Термін реалізації ідеї менше 3-х років. Термін окупності інвестицій від 3-х до 5-ти років	Термін реалізації ідеї менше 3-х років. Термін окупності інвестицій менше 3-х років
2	Необхідна розробка регламентних документів та отримання великої кількості дозвільних документів на виробництво та реалізацію продукту	Необхідно отримання великої кількості дозвільних документів на виробництво та реалізацію продукту, що вимагає значних коштів та часу	Процедура отримання дозвільних документів для виробництва та реалізації продукту вимагає незначних коштів та часу	Необхідно тільки повідомлення відповідним органам про виробництво та реалізацію продукту	Відсутні будь-які регламентні обмеження на виробництво та реалізацію продукту

Результати оцінювання науково-технічного рівня та комерційного потенціалу науково-технічної розробки потрібно звести до таблиці (табл. 5.2).

Таблиця 5.2 – Результати оцінювання науково-технічного рівня і комерційного потенціалу розробки експертами

Критерії	Експерт (ПІБ, посада)		
	1	2	3
	Бали:		
1. Технічна здійсненність концепції	4	5	5
2. Ринкові переваги (наявність аналогів)	2	2	3
3. Ринкові переваги (ціна продукту)	3	3	3
4. Ринкові переваги (технічні властивості)	2	3	3
5. Ринкові переваги (експлуатаційні витрати)	2	2	2
6. Ринкові перспективи (розмір ринку)	3	3	3
7. Ринкові перспективи (конкуренція)	2	2	2
8. Практична здійсненність (наявність фахівців)	5	5	5
9. Практична здійсненність (наявність фінансів)	2	3	2
10. Практична здійсненність (необхідність нових матеріалів)	4	5	5
11. Практична здійсненність (термін реалізації)	3	4	5
12. Практична здійсненність (розробка документів)	4	5	4
Сума балів	36	42	42
Середньоарифметична сума балів $СБ_c$	40,0		

За результатами розрахунків, наведених в таблиці 5.2, зробимо висновок щодо науково-технічного рівня і рівня комерційного потенціалу розробки. При цьому використаємо рекомендації, наведені в табл. 5.3.

Таблиця 5.3 – Науково-технічні рівні та комерційні потенціали розробки

Середньоарифметична сума балів $СБ_c$ , розрахована на основі висновків експертів	Науково-технічний рівень та комерційний потенціал розробки
41...48	Високий
31...40	Вище середнього
21...30	Середній

Продовження таблиці 5.3

Середньоарифметична сума балів СБ , розрахована на основі висновків експертів	Науково-технічний рівень та комерційний потенціал розробки
11...20	Нижче середнього
0...10	Низький

Згідно проведених досліджень рівень комерційного потенціалу розробки за темою «Інформаційна технологія аналізу та прогнозування стану атмосферного повітря міста Вінниці. Частина 1. Формування датасету та системний аналіз даних» становить 40,0 бала, що, відповідно до таблиці 5.3, свідчить про комерційну важливість проведення даних досліджень (рівень комерційного потенціалу розробки вище середнього).

## 5.2 Визначення рівня конкурентоспроможності розробки

В процесі визначення економічної ефективності науково-технічної розробки також доцільно провести прогноз рівня її конкурентоспроможності за сукупністю параметрів, що підлягають оцінюванню.

Одиничний параметричний індекс розраховуємо за формулою:

$$q_i = \frac{P_i}{P_{базі}}, \quad (5.1)$$

де  $q_i$  – одиничний параметричний індекс, розрахований за  $i$ -м параметром;

$P_i$  – значення  $i$ -го параметра виробу;

$P_{базі}$  – аналогічний параметр базового виробу-аналога, з яким проводиться порівняння.

Загальні технічні та економічні характеристики розробки представлено в таблиці 5.4.

Таблиця 5.4 – Основні техніко-економічні показники аналога та розробки, що проектується

Показники (параметри)	Одиниця вимірювання	Аналог	Проектований пристрій	Відношення параметрів нової розробки до аналога	Питома вага показника
Базове підвищення ефективності роботи інформаційної системи аналізу та прогнозування	%	48	73	1,52	0,3
Щільність заповнення сформованої бази даних	%	85	95	1,12	0,12
Доступність інтерфейсу	бал	7	9	1,28	0,15
Кешовані інформаційні блоки	шт.	12	18	1,5	0,25
Захищеність БД від зовнішнього впливу (до 10 балів)	бал	7	7	1	0,18
Експлуатаційні витрати	грн	2100	1500	0,71	0,45
Вартість системи	грн	280000	170000	0,61	0,55

Нормативні параметри оцінюємо показником, який отримує одне з двох значень: 1 – пристрій відповідає нормам і стандартам; 0 – не відповідає.

Груповий показник конкурентоспроможності за нормативними параметрами розраховуємо як добуток частинних показників за кожним параметром за формулою:

$$I_{HP} = \prod_{i=1}^n q_i, \quad (5.2)$$

де  $I_{HP}$  – загальний показник конкурентоспроможності за нормативними параметрами;

$q_i$  – одиничний (частинний) показник за  $i$ -м нормативним параметром;

$n$  – кількість нормативних параметрів, які підлягають оцінюванню.

За нормативними параметрами розроблюваний пристрій відповідає вимогам ДСТУ, тому  $I_{HP} = 1$ .

Значення групового параметричного індексу за технічними параметрами визначаємо з урахуванням вагомості (частки) кожного параметра:

$$I_{TP} = \sum_{i=1}^n q_i \cdot \alpha_i, \quad (5.3)$$

де  $I_{TP}$  – груповий параметричний індекс за технічними показниками (порівняно з виробом-аналогом);

$q_i$  – одиничний параметричний показник  $i$ -го параметра;

$\alpha_i$  – вагомість  $i$ -го параметричного показника,  $\sum_{i=1}^n \alpha_i = 1$ ;

$n$  – кількість технічних параметрів, за якими оцінюється конкурентоспроможність.

Проведемо аналіз параметрів згідно даних таблиці 5.4.

$$I_{HP} = 1,52 \cdot 0,3 + 1,12 \cdot 0,12 + 1,28 \cdot 0,15 + 1,5 \cdot 0,25 + 1 \cdot 0,18 = 1,34.$$

Груповий параметричний індекс за економічними параметрами розраховуємо за формулою [29]:



$$I_{EP} = \sum_{i=1}^m q_i \cdot \beta_i, \quad (5.4)$$

де  $I_{EP}$  – груповий параметричний індекс за економічними показниками;

$q_i$  – економічний параметр  $i$ -го виду;

$\beta_i$  – частка  $i$ -го економічного параметра,  $\sum_{i=1}^m \beta_i = 1$ ;

$m$  – кількість економічних параметрів, за якими здійснюється оцінювання.

Проведемо аналіз параметрів згідно даних таблиці.

$$I_{EP} = 0,71 \cdot 0,45 + 0,61 \cdot 0,55 = 0,66.$$

На основі групових параметричних індексів за нормативними, технічними та економічними показниками розрахуємо інтегральний показник конкурентоспроможності за формулою [30]:

$$K_{INT} = I_{HP} \cdot \frac{I_{TP}}{I_{EP}}, \quad (5.5)$$

$$K_{INT} = 1 \cdot 1,34 / 0,66 = 2,04.$$

Інтегральний показник конкурентоспроможності  $K_{INT} > 1$ , отже розробка переважає відомі аналоги за своїми техніко-економічними показниками.

### 5.3 Розрахунок витрат на проведення науково-дослідної роботи

Витрати, пов'язані з проведенням науково-дослідної роботи на тему «Інформаційна технологія аналізу та прогнозування стану атмосферного повітря міста Вінниці. Частина 1. Формування датасету та системний аналіз даних», під

час планування, обліку і калькулювання собівартості науково-дослідної роботи групуємо за відповідними статтями.

### 5.3.1 Витрати на оплату праці

До статті «Витрати на оплату праці» належать витрати на виплату основної та додаткової заробітної плати керівникам відділів, лабораторій, секторів і груп, науковим, інженерно-технічним працівникам, конструкторам, технологам, креслярам, копіювальникам, лаборантам, робітникам, студентам, аспірантам та іншим працівникам, безпосередньо зайнятим виконанням конкретної теми, обчисленої за посадовими окладами, відрядними розцінками, тарифними ставками згідно з чинними в організаціях системами оплати праці.

Основна заробітна плата дослідників

Витрати на основну заробітну плату дослідників ( $Z_o$ ) розраховуємо у відповідності до посадових окладів працівників, за формулою [30]:

$$Z_o = \sum_{i=1}^k \frac{M_{ni} \cdot t_i}{T_p}, \quad (5.6)$$

де  $k$  – кількість посад дослідників залучених до процесу досліджень;

$M_{ni}$  – місячний посадовий оклад конкретного дослідника, грн;

$t_i$  – число днів роботи конкретного дослідника;

$T_p$  – середнє число робочих днів в місяці,  $T_p=22$  дні.

$$Z_o = 17800,00 \cdot 34 / 22 = 27509,09 \text{ (грн)}.$$

Проведені розрахунки зведемо до таблиці (табл. 5.5).

Таблиця 5.5 – Витрати на заробітну плату дослідників

Найменування посади	Місячний посадовий оклад, грн	Оплата за робочий день, грн	Число днів роботи	Витрати на заробітну плату, грн
1. Керівник проекту з формування датасету та системного аналізу даних	17800,00	809,09	34	27509,09
2. Інженер-програміст	17500,00	795,45	34	27045,45
3. Консультант (провідний фахівець служби екологічного нагляду)	12860,00	584,55	9	5260,91
4. Технік	8650,00	393,18	32	12581,82
Всього				72397,27

#### Основна заробітна плата робітників

Витрати на основну заробітну плату робітників ( $Z_p$ ) за відповідними найменуваннями робіт НДР на тему «Інформаційна технологія аналізу та прогнозування стану атмосферного повітря міста Вінниці. Частина 1. Формування датасету та системний аналіз даних» розраховуємо за формулою:

$$Z_p = \sum_{i=1}^n C_i \cdot t_i, \quad (5.7)$$

де  $C_i$  – погодинна тарифна ставка робітника відповідного розряду, за виконану відповідну роботу, грн/год;

$t_i$  – час роботи робітника при виконанні визначеної роботи, год.

Погодинну тарифну ставку робітника відповідного розряду  $C_i$  можна визначити за формулою:

$$C_i = \frac{M_M \cdot K_i \cdot K_c}{T_p \cdot t_{зм}}, \quad (5.8)$$

де  $M_M$  – розмір прожиткового мінімуму працездатної особи, або мінімальної місячної заробітної плати (в залежності від діючого законодавства), прийmemo  $M_M=6700,00$  грн;

$K_i$  – коефіцієнт міжкваліфікаційного співвідношення для встановлення тарифної ставки робітнику відповідного розряду [30];

$K_c$  – мінімальний коефіцієнт співвідношень місячних тарифних ставок робітників першого розряду з нормальними умовами праці виробничих об'єднань і підприємств до законодавчо встановленого розміру мінімальної заробітної плати.

$T_p$  – середнє число робочих днів в місяці, приблизно  $T_p = 22$  дня;

$t_{зм}$  – тривалість зміни, год.

$$C_i = 6700,00 \cdot 1,10 \cdot 1,35 / (22 \cdot 8) = 56,53 \text{ (грн)}.$$

$$Z_{pl} = 56,53 \cdot 7,50 = 423,98 \text{ (грн)}.$$

Таблиця 5.6 – Величина витрат на основну заробітну плату робітників

Найменування робіт	Тривалість роботи, год	Розряд роботи	Тарифний коефіцієнт	Погодинна тарифна ставка, грн	Величина оплати на робітника грн
1. Встановлення офісних меблів, техніки, обладнання	7,50	2	1,10	56,53	423,98
2. Інсталяція програмного забезпечення	8,50	4	1,35	69,38	589,72
3. Компіляція програмних модулів	15,00	5	1,70	87,37	1310,50
4. Відлагодження програмних модулів	8,00	5	1,70	87,37	698,93
5. Формування бази даних	32,00	2	1,10	56,53	1809,00
6. Налаштування серверного обладнання	6,00	5	1,70	87,37	524,20
Всього					5356,34

Додаткова заробітна плата дослідників та робітників

Додаткову заробітну плату розраховуємо як 10 ... 12% від суми основної заробітної плати дослідників та робітників за формулою:

$$Z_{\text{дод}} = (Z_o + Z_p) \cdot \frac{H_{\text{дод}}}{100\%}, \quad (5.9)$$

де  $H_{\text{дод}}$  – норма нарахування додаткової заробітної плати. Прийmemo 11%.

$$Z_{\text{дод}} = (72397,27 + 5356,34) \cdot 11 / 100\% = 8552,90 \text{ (грн)}.$$

### 5.3.2 Відрахування на соціальні заходи

Нарахування на заробітну плату дослідників та робітників розраховуємо як 22% від суми основної та додаткової заробітної плати дослідників і робітників за формулою:

$$Z_n = (Z_o + Z_p + Z_{\text{доод}}) \cdot \frac{H_{zn}}{100\%}, \quad (5.10)$$

де  $H_{zn}$  – норма нарахування на заробітну плату. Приймаємо 22%.

$$Z_n = (72397,27 + 5356,34 + 8552,90) \cdot 22 / 100\% = 18987,43 \text{ (грн)}.$$

### 5.3.3 Сировина та матеріали

До статті «Сировина та матеріали» належать витрати на сировину, основні та допоміжні матеріали, інструменти, пристрої та інші засоби і предмети праці, які придбані у сторонніх підприємств, установ і організацій та витрачені на проведення досліджень за темою «Інформаційна технологія аналізу та прогнозування стану атмосферного повітря міста Вінниці. Частина 1. Формування датасету та системний аналіз даних».

Витрати на матеріали ( $M$ ), у вартісному вираженні розраховуються окремо по кожному виду матеріалів за формулою:

$$M = \sum_{j=1}^n H_j \cdot C_j \cdot K_j - \sum_{j=1}^n B_j \cdot C_{ej}, \quad (5.11)$$

де  $H_j$  – норма витрат матеріалу  $j$ -го найменування, кг;

$n$  – кількість видів матеріалів;

$C_j$  – вартість матеріалу  $j$ -го найменування, грн/кг;

$K_j$  – коефіцієнт транспортних витрат, ( $K_j = 1,1 \dots 1,15$ );

$V_j$  – маса відходів  $j$ -го найменування, кг;

$C_{vj}$  – вартість відходів  $j$ -го найменування, грн/кг.

$$M_1 = 3,0 \cdot 184,00 \cdot 1,1 - 0,0 \cdot 0,0 = 607,20 \text{ (грн)}.$$

Проведені розрахунки зведемо до таблиці (табл. 5.7).

Таблиця 5.7 – Витрати на матеріали

Найменування матеріалу, марка, тип, сорт	Ціна за 1 кг, грн	Норма витрат, кг	Величина відходів, кг	Ціна відходів, грн/кг	Вартість витраченого матеріалу, грн
Папір канцелярський офісний (A4)	184,00	3,0	0,0	0,0	607,20
Папір для заміток (A5)	111,00	3,0	0,0	0,0	366,30
Начиння канцелярське	196,00	3,0	0,0	0,0	646,80
Органайзер офісний	200,00	3,0	0,0	0,0	660,00
Картридж для принтера (Canon A725-fdx01)	1206,00	1,0	0,0	0,0	1326,60
Диск оптичний (CD-R)	27,00	3,0	0,0	0,0	89,10

## Продовження таблиці 5.7

Найменування матеріалу, марка, тип, сорт	Ціна за 1 кг, грн	Норма витрат, кг	Величина відходів, кг	Ціна відходів, грн/кг	Вартість витраченого матеріалу, грн
Диск оптичний (CD-RW)	32,00	3,0	0,0	0,0	105,60
USB-FLASH-пам'ять (32 ГБ)	139,00	1,0	0,0	0,0	152,90
Всього					3954,50

### 5.3.4 Розрахунок витрат на комплектуючі

Витрати на комплектуючі ( $K_e$ ), які використовують при проведенні НДР на тему «Інформаційна технологія аналізу та прогнозування стану атмосферного повітря міста Вінниці. Частина 1. Формування датасету та системний аналіз даних» відсутні.

### 5.3.5 Спецустаткування для наукових (експериментальних) робіт

До статті «Спецустаткування для наукових (експериментальних) робіт» належать витрати на виготовлення та придбання спецустаткування необхідного для проведення досліджень, також витрати на їх проектування, виготовлення, транспортування, монтаж та встановлення.

Балансову вартість спецустаткування розраховуємо за формулою:

$$B_{\text{спец}} = \sum_{i=1}^k C_i \cdot C_{\text{пр.і}} \cdot K_i, \quad (5.12)$$

де  $C_i$  – ціна придбання одиниці спецустаткування даного виду, марки, грн;



$C_{np.i}$  –кількість одиниць устаткування відповідного найменування, які придбані для проведення досліджень, шт.;

$K_i$  – коефіцієнт, що враховує доставку, монтаж, налагодження устаткування тощо, ( $K_i = 1,10...1,12$ );

$k$  – кількість найменувань устаткування.

$V_{спец} = 35680,00 \cdot 1 \cdot 1,1 = 39248,00$  (грн).

Отримані результати зведемо до таблиці (табл. 5.8):

Таблиця 5.8 – Витрати на придбання спецустаткування по кожному виду

Найменування устаткування	Кількість, шт	Ціна за одиницю, грн	Вартість, грн
Сервер - Компютер Expert PC Balance (I91F8H1S115E429)	1	35680,00	39248,00
Імітатори термінальних пристроїв	1	4259,00	4684,90
Всього			43932,90

### 5.3.6 Програмне забезпечення для наукових (експериментальних) робіт

До статті «Програмне забезпечення для наукових (експериментальних) робіт» належать витрати на розробку та придбання спеціальних програмних засобів і програмного забезпечення, (програм, алгоритмів, баз даних) необхідних для проведення досліджень, також витрати на їх проектування, формування та встановлення.

Балансову вартість програмного забезпечення розраховуємо за формулою:

$$V_{npg} = \sum_{i=1}^k C_{inpg} \cdot C_{np.g.i} \cdot K_i , \quad (5.13)$$

де  $C_{inpr}$  – ціна придбання одиниці програмного засобу даного виду, грн;

$C_{prg.i}$  – кількість одиниць програмного забезпечення відповідного найменування, які придбані для проведення досліджень, шт.;

$K_i$  – коефіцієнт, що враховує інсталяцію, налагодження програмного засобу тощо, ( $K_i = 1, 10 \dots 1, 12$ );

$k$  – кількість найменувань програмних засобів.

$$B_{prg} = 239,00 \cdot 2 \cdot 1,1 = 525,80 \text{ (грн)}.$$

Отримані результати зведемо до таблиці (табл. 5.9):

Таблиця 5.9 – Витрати на придбання програмних засобів по кожному виду

Найменування програмного засобу	Кількість, шт	Ціна за одиницю, грн	Вартість, грн
Доступ до мережі Internet (високошвидкісний) грн/місяць	2	239,00	525,80
Емулятор серверної інтернет-платформи для моделювання поведінки інформаційного ресурсу	1	5800,00	6380,00
Середовище програмування Python та її бібліотеки	1	8520,00	9372,00
Система керування реляційними базами даних MySQL	1	8459,00	9304,90
Система Kaggle	1	7599,00	8358,90
Всього			33941,60

### 5.3.7 Амортизація обладнання, програмних засобів та приміщень

В спрощеному вигляді амортизаційні відрахування по кожному виду обладнання, приміщень та програмному забезпеченню тощо, розраховуємо з використанням прямолінійного методу амортизації за формулою:

$$A_{обл} = \frac{Ц_{\bar{o}}}{T_{\bar{e}}} \cdot \frac{t_{вик}}{12}, \quad (5.14)$$

де  $Ц_{\bar{o}}$  – балансова вартість обладнання, програмних засобів, приміщень тощо, які використовувались для проведення досліджень, грн;

$t_{вик}$  – термін використання обладнання, програмних засобів, приміщень під час досліджень, місяців;

$T_{\bar{e}}$  – строк корисного використання обладнання, програмних засобів, приміщень тощо, років.

$$A_{обл} = (44999,00 \cdot 2) / (3 \cdot 12) = 2499,94 \text{ (грн)}.$$

Проведені розрахунки зведемо до таблиці (табл. 5.10).

Таблиця 5.10 – Амортизаційні відрахування по кожному виду обладнання

Найменування обладнання	Балансова вартість, грн	Строк корисного використання, років	Термін використання обладнання, місяців	Амортизаційні відрахування, грн
Програмно-аналітичний модуль на основі Qbox I21831	44999,00	3	2	2499,94

Продовження таблиці 5.10

Найменування обладнання	Балансова вартість, грн	Строк корисного використання, років	Термін використання обладнання, місяців	Амортизаційні відрахування, грн
Серверне обладнання з підтримкою DATA BASE 2.0	25340,00	3	2	1407,78
Програмний комплекс розробника програмного забезпечення	10840,00	3	2	602,22
Місце оператора спеціалізоване	9650,00	5	2	321,67
Офісна оргтехніка	11780,00	4	2	490,83
Дослідницька лабораторія	286000,00	25	2	1906,67
Мультисервісна платформа зберігання інформації	12599,00	4	2	524,96
ОС Windows 11	7400,00	3	2	411,11
Прикладний пакет Microsoft Office 2021	6700,00	3	2	372,22
Всього				8537,40

### 5.3.8 Паливо та енергія для науково-виробничих цілей

Витрати на силову електроенергію ( $B_e$ ) розраховуємо за формулою:

$$B_e = \sum_{i=1}^n \frac{W_{yi} \cdot t_i \cdot C_e \cdot K_{eni}}{\eta_i}, \quad (5.15)$$

де  $W_{yi}$  – встановлена потужність обладнання на визначеному етапі розробки, кВт;

$t_i$  – тривалість роботи обладнання на етапі дослідження, год;

$C_e$  – вартість 1 кВт-години електроенергії, грн; (вартість електроенергії визначається за даними енергопостачальної компанії), прийmemo  $C_e = 7,50$  грн;

$K_{eni}$  – коефіцієнт, що враховує використання потужності,  $K_{eni} < 1$ ;

$\eta_i$  – коефіцієнт корисної дії обладнання,  $\eta_i < 1$ .

$$B_e = 0,36 \cdot 250,0 \cdot 7,50 \cdot 0,95 / 0,97 = 675,00 \text{ (грн)}.$$

Проведені розрахунки зведемо до таблиці (табл. 5.11).

Таблиця 5.11 – Витрати на електроенергію

Найменування обладнання	Встановлена потужність, кВт	Тривалість роботи, год	Сума, грн
Програмно-аналітичний модуль на основі Qbox I21831	0,36	250,0	675,00
Серверне обладнання з підтримкою DATA BASE 2.0	0,25	250,0	468,75
Програмний комплекс розробника програмного забезпечення	0,25	250,0	468,75
Місце оператора спеціалізоване	0,08	250,0	150,00

Продовження таблиці 5.11

Найменування обладнання	Встановлена потужність, кВт	Тривалість роботи, год	Сума, грн
Офісна оргтехніка	0,45	1,5	5,06
Мультисервісна платформа зберігання інформації	0,07	250,0	131,25
Сервер - Компютер Expert PC Balance (I91F8H1S115E429)	0,32	180,0	432,00
Імітатори термінальних пристроїв	0,03	180,0	40,50
Всього			2371,31

### 5.3.9 Службові відрядження

До статті «Службові відрядження» дослідної роботи на тему «Інформаційна технологія аналізу та прогнозування стану атмосферного повітря міста Вінниці. Частина 1. Формування датасету та системний аналіз даних» належать витрати на відрядження штатних працівників, працівників організацій, які працюють за договорами цивільно-правового характеру, аспірантів, зайнятих розробленням досліджень, відрядження, пов'язані з проведенням випробувань машин та приладів, а також витрати на відрядження на наукові з'їзди, конференції, наради, пов'язані з виконанням конкретних досліджень.

Витрати за статтею «Службові відрядження» розраховуємо як 20...25% від суми основної заробітної плати дослідників та робітників за формулою:

$$B_{cv} = (Z_o + Z_p) \cdot \frac{H_{cv}}{100\%}, \quad (5.16)$$

де  $H_{cb}$  – норма нарахування за статтею «Службові відрядження», прийmemo  $H_{cb} = 20\%$ .

$$B_{cb} = (72397,27 + 5356,34) \cdot 20 / 100\% = 15550,72 \text{ (грн)}.$$

### **5.3.10 Витрати на роботи, які виконують сторонні підприємства, установи і організації**

Витрати за статтею «Витрати на роботи, які виконують сторонні підприємства, установи і організації» розраховуємо як 30...45% від суми основної заробітної плати дослідників та робітників за формулою:

$$B_{cn} = (Z_o + Z_p) \cdot \frac{H_{cn}}{100\%}, \quad (5.17)$$

де  $H_{cn}$  – норма нарахування за статтею «Витрати на роботи, які виконують сторонні підприємства, установи і організації», прийmemo  $H_{cn} = 35\%$ .

$$B_{cn} = (72397,27 + 5356,34) \cdot 35 / 100\% = 27213,76 \text{ (грн)}.$$

### **5.3.11 Інші витрати**

До статті «Інші витрати» належать витрати, які не знайшли відображення у зазначених статтях витрат і можуть бути віднесені безпосередньо на собівартість досліджень за прямими ознаками.

Витрати за статтею «Інші витрати» розраховуємо як 50...100% від суми основної заробітної плати дослідників та робітників за формулою:

$$I_g = (Z_o + Z_p) \cdot \frac{H_{ig}}{100\%}, \quad (5.18)$$

де  $H_{ib}$  – норма нарахування за статтею «Інші витрати», прийmemo  $H_{ib} = 50\%$ .

$$I_e = (72397,27 + 5356,34) \cdot 50 / 100\% = 38876,80 \text{ (грн)}.$$

### 5.3.12 Накладні (загальновиробничі) витрати

До статті «Накладні (загальновиробничі) витрати» належать: витрати, пов'язані з управлінням організацією; витрати на винахідництво та раціоналізацію; витрати на підготовку (перепідготовку) та навчання кадрів; витрати, пов'язані з набором робочої сили; витрати на оплату послуг банків; витрати, пов'язані з освоєнням виробництва продукції; витрати на науково-технічну інформацію та рекламу та ін.

Витрати за статтею «Накладні (загальновиробничі) витрати» розраховуємо як 100...150% від суми основної заробітної плати дослідників та робітників за формулою:

$$B_{нзв} = (Z_o + Z_p) \cdot \frac{H_{нзв}}{100\%}, \quad (5.19)$$

де  $H_{нзв}$  – норма нарахування за статтею «Накладні (загальновиробничі) витрати», прийmemo  $H_{нзв} = 110\%$ .

$$B_{нзв} = (72397,27 + 5356,34) \cdot 110 / 100\% = 85528,97 \text{ (грн)}.$$

Витрати на проведення науково-дослідної роботи на тему «Інформаційна технологія аналізу та прогнозування стану атмосферного повітря міста Вінниці. Частина 1. Формування датасету та системний аналіз даних» розраховуємо як суму всіх попередніх статей витрат за формулою:

$$B_{заг} = Z_o + Z_p + Z_{доо} + Z_n + M + K_в + B_{снец} + B_{прз} + A_{обл} + B_e + B_{св} + B_{сн} + I_e + B_{нзв}. \quad (5.20)$$



$$B_{\text{заг}} = 72397,27 + 5356,34 + 8552,90 + 18987,43 + 3954,50 + 0,00 + 43932,90 + 33941,60 + 8537,40 + 2371,31 + 15550,72 + 27213,76 + 38876,80 + 85528,97 = 365201,91 \text{ (грн)}.$$

Загальні витрати  $ZB$  на завершення науково-дослідної (науково-технічної) роботи та оформлення її результатів розраховується за формулою:

$$ZB = \frac{B_{\text{заг}}}{\eta}, \quad (5.21)$$

де  $\eta$  - коефіцієнт, який характеризує етап (стадію) виконання науково-дослідної роботи, прийmemo  $\eta=0,95$ .

$$ZB = 365201,91 / 0,95 = 384423,06 \text{ (грн)}.$$

#### **5.4 Розрахунок економічної ефективності науково-технічної розробки при її можливій комерціалізації потенційним інвестором**

Науково-дослідна робота складається з двох взаємопов'язаних частин «Інформаційна технологія аналізу та прогнозування стану атмосферного повітря міста Вінниці. Частина 1. Формування датасету та системний аналіз даних» та «Інформаційна технологія аналізу та прогнозування стану атмосферного повітря міста Вінниці. Частина 2. Прогнозування даних» які досліджувались окремими групами науковців. Сумарні витрати на проведення відповідних досліджень складають:

$$ZB = 384423,00 + 479656,00 = 864079,00 \text{ (грн)}.$$

Результати дослідження проведені за темою «Інформаційна технологія аналізу та прогнозування стану атмосферного повітря міста Вінниці» передбачають комерціалізацію протягом 4-х років реалізації на ринку.

В цьому випадку основу майбутнього економічного ефекту будуть формувати:

$\Delta N$  – збільшення кількості споживачів яким надається відповідна інформаційна послуга у періоди часу, що аналізуються;

Показник	1-й рік	2-й рік	3-й рік	4-й рік
Збільшення кількості споживачів, осіб	6000	8500	9200	6800

$N$  – кількість споживачів яким надавалась відповідна інформаційна послуга у році до впровадження результатів нової науково-технічної розробки, прийmemo 33000 осіб;

$C_o$  – вартість послуги у році до впровадження інформаційної системи, прийmemo 250,00 (грн);

$\pm \Delta C_o$  – зміна вартості послуги від впровадження результатів, прийmemo 57,50 (грн).

Можливе збільшення чистого прибутку у потенційного інвестора  $\Delta \Pi_i$  для кожного із 4-х років, протягом яких очікується отримання позитивних результатів від можливого впровадження та комерціалізації науково-технічної розробки, розраховуємо за формулою [30]:

$$\Delta \Pi_i = (\pm \Delta C_o \cdot N + C_o \cdot \Delta N)_i \cdot \lambda \cdot \rho \cdot \left(1 - \frac{\rho}{100}\right), \quad (5.22)$$

де  $\lambda$  – коефіцієнт, який враховує сплату потенційним інвестором податку на додану вартість. У 2023 році ставка податку на додану вартість складає 20%, а коефіцієнт  $\lambda = 0,8333$ ;

$\rho$  – коефіцієнт, який враховує рентабельність інноваційного продукту).  
Прийmemo  $\rho = 45\%$ ;

$\vartheta$  – ставка податку на прибуток, який має сплачувати потенційний інвестор, у 2023 році  $\vartheta = 18\%$ ;

Збільшення чистого прибутку 1-го року:

$$\Delta\Pi_1 = (57,50 \cdot 33000,00 + 307,50 \cdot 6000) \cdot 0,83 \cdot 0,45 \cdot (1 - 0,18/100\%) = 1146215,48$$

(грн).

Збільшення чистого прибутку 2-го року:

$$\Delta\Pi_2 = (57,50 \cdot 33000,00 + 307,50 \cdot 14500) \cdot 0,83 \cdot 0,45 \cdot (1 - 0,18/100\%) = 1946728,69$$

(грн).

Збільшення чистого прибутку 3-го року:

$$\Delta\Pi_3 = (57,50 \cdot 33000,00 + 307,50 \cdot 23700) \cdot 0,83 \cdot 0,45 \cdot (1 - 0,18/100\%) = 2813166,52$$

(грн).

Збільшення чистого прибутку 4-го року:

$$\Delta\Pi_4 = (57,50 \cdot 33000,00 + 307,50 \cdot 30500) \cdot 0,83 \cdot 0,45 \cdot (1 - 0,18/100\%) = 3453577,09$$

(грн).

Приведена вартість збільшення всіх чистих прибутків  $ПП$ , що їх може отримати потенційний інвестор від можливого впровадження та комерціалізації науково-технічної розробки:

$$ПП = \sum_{i=1}^T \frac{\Delta\Pi_i}{(1 + \tau)^t}, \quad (5.23)$$

де  $\Delta\Pi_i$  – збільшення чистого прибутку у кожному з років, протягом яких виявляються результати впровадження науково-технічної розробки, грн;

$T$  – період часу, протягом якого очікується отримання позитивних результатів від впровадження та комерціалізації науково-технічної розробки, роки;

$\tau$  – ставка дисконтування, за яку можна взяти щорічний прогнозований рівень інфляції в країні,  $\tau = 0,12$ ;

$t$  – період часу (в роках) від моменту початку впровадження науково-технічної розробки до моменту отримання потенційним інвестором додаткових чистих прибутків у цьому році.

$$\begin{aligned} ПП &= 1146215,48/(1+0,12)^1 + 1946728,69/(1+0,12)^2 + 2813166,52/(1+0,12)^3 + \\ &+ 3453577,09/(1+0,12)^4 = 1023406,67 + 1551920,19 + 2002356,36 + 2194810,67 = \\ &= 6772493,90 \text{ (грн)}. \end{aligned}$$

Величина початкових інвестицій  $PV$ , які потенційний інвестор має вкласти для впровадження і комерціалізації науково-технічної розробки:

$$PV = k_{инв} \cdot ЗВ, \quad (5.24)$$

де  $k_{инв}$  – коефіцієнт, що враховує витрати інвестора на впровадження науково-технічної розробки та її комерціалізацію, приймаємо  $k_{инв} = 1$ ;

$ЗВ$  – загальні витрати на проведення науково-технічної розробки та оформлення її результатів, приймаємо 864079,00 (грн).

$$PV = k_{инв} \cdot ЗВ = 1 \cdot 864079,00 = 864079,00 \text{ (грн)}.$$

Абсолютний економічний ефект  $E_{абс}$  для потенційного інвестора від можливого впровадження та комерціалізації науково-технічної розробки становитиме:

$$E_{абс} = ПП - PV, \quad (5.25)$$

де  $ПП$  – приведена вартість зростання всіх чистих прибутків від можливого впровадження та комерціалізації науково-технічної розробки, 6772493,90 (грн);

$PV$  – теперішня вартість початкових інвестицій, 864079,00 (грн).

$$E_{абс} = ПП - PV = 6772493,90 - 864079,00 = 5908414,90 \text{ (грн)}.$$

Внутрішня економічна дохідність інвестицій  $E_e$ , які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки:

$$E_e = \sqrt[Tж]{1 + \frac{E_{абс}}{PV}} - 1, \quad (5.26)$$

де  $E_{abc}$  – абсолютний економічний ефект вкладених інвестицій, 5908414,90 (грн);

$PV$  – теперішня вартість початкових інвестицій, 864079,00 (грн);

$T_{ж}$  – життєвий цикл науково-технічної розробки, тобто час від початку її розробки до закінчення отримання позитивних результатів від її впровадження, 4 роки.

$$E_g = \sqrt[T_{ж}]{1 + \frac{E_{abc}}{PV}} - 1 = (1 + 5908414,90/864079,00)^{1/4} = 0,67.$$

Мінімальна внутрішня економічна дохідність вкладених інвестицій  $\tau_{min}$ :

$$\tau_{min} = d + f, \quad (5.27)$$

де  $d$  – середньозважена ставка за депозитними операціями в комерційних банках; в 2023 році в Україні  $d = 0,12$ ;

$f$  – показник, що характеризує ризикованість вкладення інвестицій, прийmemo 0,3.

$\tau_{min} = 0,12 + 0,3 = 0,42 < 0,67$  свідчить про те, що внутрішня економічна дохідність інвестицій  $E_g$ , які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки вища мінімальної внутрішньої дохідності. Тобто інвестувати в науково-дослідну роботу за темою «Інформаційна технологія аналізу та прогнозування стану атмосферного повітря міста Вінниці» доцільно.

Період окупності інвестицій  $T_{ок}$  які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки:

$$T_{ок} = \frac{1}{E_g}, \quad (5.28)$$

де  $E_g$  – внутрішня економічна дохідність вкладених інвестицій.

$$T_{ок} = 1 / 0,67 = 1,49 \text{ р.}$$

$T_{ок} < 3$ -х років, що свідчить про комерційну привабливість науково-технічної розробки і може спонукати потенційного інвестора профінансувати впровадження даної розробки та виведення її на ринок.

## **5.5 Висновки**

Згідно проведених досліджень рівень комерційного потенціалу розробки за темою «Інформаційна технологія аналізу та прогнозування стану атмосферного повітря міста Вінниці» є високим, що, свідчить про комерційну важливість проведення даних досліджень (рівень комерційного потенціалу розробки вище середнього).

Також термін окупності становить 1,49 року, що менше 3-х років, що свідчить про комерційну привабливість науково-технічної розробки і може спонукати потенційного інвестора профінансувати впровадження даної розробки та виведення її на ринок.

Отже, можна зробити висновок про доцільність проведення даної роботи за темою «Інформаційна технологія аналізу та прогнозування стану атмосферного повітря міста Вінниці».

## ВИСНОВКИ

Під час виконання магістерської кваліфікованої роботи проведено глибоке дослідження щодо якості повітря в місті Вінниця на момент 2023 року.

В ході дослідження виконано аналіз проблематики забруднення атмосферного повітря, використовуючи звіт State of Global Air 2020, де визначено найбільш забруднюючі країни та висвітлено причини забруднення. Також розглянуто поточний стан моніторингу якості повітря в місті.

Ще однією необхідною складовою для виконання даної роботи став вибір оптимальних інформаційних технологій для здійснення системного аналізу даних, де визначено, що мова програмування Python є відмінним інструментом в галузі штучного інтелекту та системного аналізу. Обрана мова програмування відрізняється своєю потужністю та гнучкістю, надаючи можливість ефективно вирішувати завдання різного рівня складності.

Крім того, середовище розробки Kaggle Notebooks обрано за свою зручність та наявність інтегрованих бібліотек, необхідних для виконання поставлених завдань.

Використання Python у поєднанні з Kaggle Notebooks створює оптимальне робоче середовище для проведення дослідження та розробки нових підходів до аналізу якості повітря у місті.

Також, в цій роботі акцент був зроблений на систематизації та аналізі даних, пов'язаних із забрудненням повітря в м. Вінниця. Для цього використано мову програмування Python та відповідні бібліотеки.

Обрана мова програмування Python забезпечила не лише швидку та ефективну обробку великих обсягів даних, але й надала можливість використовувати різноманітні бібліотеки для статистичного аналізу та візуалізації результатів.

Створення нового ноутбуку у Kaggle та завантаження даних були ключовими кроками для підготовки до подальшого аналізу. Це дозволило

створити зручне робоче середовище, де кожен крок обробки та аналізу даних міг бути ефективно реалізований.

В даній частині роботи зроблено перший крок у вивченні та розумінні даних про якість повітря в місті Вінниця, що є ключовим етапом для подальшого розроблення нових стратегій та рішень для покращення ситуації з якістю повітря.

Магістерську кваліфікаційну роботу присвячено детальному аналізу рівня забруднення повітря в місті Вінниця. Загальний висновок полягає в тому, що рівень якості повітря вважається задовільним, за винятком показника  $PM_{10}$ . Цей показник свідчить про перевищення нормативів частинок пилу розміром 1 мікромметр, що вказує на необхідність прийняття заходів для зменшення концентрації пилу та інших факторів, що впливають на якість повітря в цілому.

Окрім того, виявлено надзвичайне потепління, що може суттєво вплинути на стратегії збереження якості повітря та взаємодії з навколишнім середовищем. Зафіксоване збільшення середньомісячної температури вказує на необхідність адаптації існуючих стратегій до змін кліматичних умов та розроблення нових заходів для збереження стабільної якості повітря в умовах постійного змінення клімату.

На завершення, під час прогнозування стану атмосферного повітря у місті Вінниця за використанням моделей ARIMA та Prophet в мові програмування Python, була досягнута точність прогнозування близько 75%. Цей результат вказує на можливе покращення якості повітря в місті протягом наступних місяців.

Усі ці висновки підкреслюють важливість не тільки постійного моніторингу якості повітря, але і регулярного перегляду та удосконалення заходів для збереження здоров'я громадян та екологічної стабільності міста.

У роботі враховано результати науково-дослідної роботи ВНТУ "Розробка розділів програми державного моніторингу у галузі охорони атмосферного повітря Вінницької області" на замовлення ТОВ "ІННОВІНПРОМ" та Управління розвитку територій та інфраструктури Вінницької обласної державної адміністрації (номер у ВНТУ - 2859, номер держреєстрації



0121U114085) (листопад-грудень 2021 року) та опубліковано тези на ЛШ Всеукраїнській науково-технічній конференції підрозділів Вінницького національного технічного університету факультету інтелектуальних інформаційних технологій та автоматизації (2023-2024).

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Роль технологій в зменшенні забруднення атмосферного повітря. URL: <https://airly.org/en/the-role-of-technology-in-reducing-air-pollution/> (дата звернення: 10.09.2023).
2. Крижановський Є. М., Мокін В. Б, Піроговський А. В., Просторовий аналіз даних забруднення атмосферного повітря Вінницької області за даними державного моніторингу у 2020 р. *LI Науково-технічна конференція факультету інтелектуальних інформаційних технологій та автоматизації*: зб. матеріалів конференції. Вінниця, 2022. URL: <https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2022/paper/view/14910>
3. Крижановський Є. М., Піроговський А.В., Формування датасету та прогнозування стану атмосферного повітря міста Вінниці. *LIII Науково-технічна конференція факультету інтелектуальних інформаційних технологій та автоматизації (2023-2024)*. Вінниця, 2023-2024. URL: <https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2024/paper/view/19696>
4. Якість повітря та здоров'я. URL: <https://www.pca.state.mn.us/air-water-land-climate/air-quality-and-health#:~:text=Air%20pollution%20also%20increases%20the,underlying%20health%20and%20other%20factors> (дата звернення: 10.09.2023).
5. State of Global Air. URL: <https://www.stateofglobalair.org/> (дата звернення: 13.09.2023).
6. Рівень забруднення атмосферного повітря у м. Вінниця. URL: <https://www.saveecobot.com/maps/vinnitsia> (дата звернення: 16.09.2023).
7. Індекс якості повітря. URL: [https://en.wikipedia.org/wiki/Air\\_quality\\_index](https://en.wikipedia.org/wiki/Air_quality_index) (дата звернення: 16.09.2023).
8. Таблиця класифікації індексу якості повітря. URL: [https://uk.wikipedia.org/wiki/%D0%86%D0%BD%D0%B4%D0%B5%D0%BA%D1%81\\_%D1%8F%D0%BA%D0%BE%D1%81%D1%82%D1%96\\_%](https://uk.wikipedia.org/wiki/%D0%86%D0%BD%D0%B4%D0%B5%D0%BA%D1%81_%D1%8F%D0%BA%D0%BE%D1%81%D1%82%D1%96_%)

[D0%BF%D0%BE%D0%B2%D1%96%D1%82%D1%80%D1%8F](#) (дата звернення: 16.09.2023).

9. Програма державного моніторингу у галузі охорони атмосферного повітря агломерації «Вінниця» на 2021-2025 роки. URL: <https://vmr.gov.ua/media/%D0%A1%D0%B5%D1%80%D0%B2%D1%96%D1%81%D0%B8/Ecology/%D0%9F%D1%80%D0%BE%D0%B3%D1%80%D0%B0%D0%BC%D0%B0%20%D0%BE%D1%81%D1%82%D0%B0%D1%82%D0%BE%D1%87%D0%BD%D0%B8%D0%B9%20%D0%B2%D0%B0%D1%80%D1%96%D0%B0%D0%BD%D1%82%2026.11.21.pdf> (дата звернення: 17.09.2023).

10. Інформаційні технологій простими словами. URL: <https://www.ukraine-lifehacker.com/informatsiini-tekhnohii-prostymy-slovamy> (дата звернення: 17.09.2023).

11. Мова програмування Python. URL: <https://www.teradata.com/Glossary/What-is-Python> (дата звернення: 17.09.2023).

12. Середовище розробки. URL: <https://umbraco.com/knowledge-base/development-environment/> (дата звернення: 18.09.2023).

13. Kaggle Notebooks. URL: <https://www.kaggle.com/docs/notebooks> (дата звернення: 20.09.2023).

14. Бібліотека NumPy. URL: <https://uk.wikipedia.org/wiki/NumPy> (дата звернення: 20.09.2023).

15. Бібліотека Pandas. URL: <https://uk.wikipedia.org/wiki/Pandas> (дата звернення: 20.09.2023).

16. Бібліотека Matplotlib. URL: <https://uk.wikipedia.org/wiki/Matplotlib> (дата звернення: 20.09.2023).

17. Бібліотека Seaborn. URL: <https://seaborn.pydata.org/> (дата звернення: 20.09.2023).

18. Бібліотека SciPy. URL: <https://uk.wikipedia.org/wiki/SciPy> (дата звернення: 20.09.2023).

19. Бібліотека Scikit-Learn. URL: <https://uk.wikipedia.org/wiki/Scikit-learn> (дата звернення: 20.09.2023).

20. Бібліотека Statsmodels. URL: <https://en.wikipedia.org/wiki/Statsmodels> (дата звернення: 20.09.2023).
21. Folium. URL: <https://python-visualization.github.io/folium/latest/> (дата звернення: 20.09.2023).
22. Портал відкритих даних Вінницької міської ради. URL: <https://opendata.gov.ua/> (дата звернення: 11.11.2023).
23. Портал відкритих даних ВМР та його нормативні документи. URL: <https://www.vmr.gov.ua/open-data> (дата звернення: 02.11.2023).
24. Редактор таблиць Microsoft Office. URL: [https://uk.wikipedia.org/wiki/Microsoft\\_Excel](https://uk.wikipedia.org/wiki/Microsoft_Excel) (дата звернення: 02.10.2023).
25. Data of the air quality in Vinnytsia 2023. URL: <https://www.kaggle.com/datasets/pirogovskiy/air-quality-in-vinnytsia-2023-september> (дата звернення: 08.12.2023).
26. Kaggle Notebook: Air Quality City - Mapping. URL: <https://www.kaggle.com/code/vbmokin/air-quality-city-mapping> (дата звернення: 20.10.2023).
27. ARIMA model Python. URL: <https://towardsdatascience.com/machine-learning-part-19-time-series-and-autoregressive-integrated-moving-average-model-arima-c1005347b0d7> (дата звернення: 01.10.2023).
28. Робота з моделю прогнозування Prophet. URL: [https://facebook.github.io/prophet/docs/quick\\_start.html](https://facebook.github.io/prophet/docs/quick_start.html) (дата звернення: 01.10.2023).
29. Козловський В. О., Лесько О. Й., Кавецький В. В. Методичні вказівки до виконання економічної частини магістерських кваліфікаційних робіт : Вінниця: ВНТУ, 2021. 42 с.
30. Кавецький В. В., Козловський В. О., Причепка І.В., Економічне обґрунтування інноваційних рішень : приктикум / за ред. В. В. Кавецький. Вінниця: ВНТУ, 2016. 113 с.

## Додаток А

Міністерство освіти і науки України  
Вінницький національний технічний університет  
Факультет інтелектуальних інформаційних технологій та автоматизації

ЗАТВЕРДЖУЮ

Завідувач кафедри САІТ

\_\_\_\_\_ д.т.н., проф. Віталій МОКІН

«\_\_» \_\_\_\_\_ 2023 р.

## ТЕХНІЧНЕ ЗАВДАННЯ

на комплексну магістерську кваліфікаційну роботу  
«ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ  
СТАНУ АТМОСФЕРНОГО ПОВІТРЯ МІСТА ВІННИЦІ. ЧАСТИНА 1.  
ФОРМУВАННЯ ДАТАСЕТУ ТА СИСТЕМНИЙ АНАЛІЗ ДАНИХ»  
08-34.МКР.002.02.000.ТЗ

Керівник: к.т.н., доц. каф. САІТ

\_\_\_\_\_ Євгеній КРИЖАНОВСЬКИЙ

«\_\_» \_\_\_\_\_ 2023 р.

Розробив: студент гр. 2ІСТ-22м

\_\_\_\_\_ Андрій ПІРОГОВСЬКИЙ

«\_\_» \_\_\_\_\_ 2023 р.

Вінниця 2023

### 1. Підстава для проведення робіт

Підставою для виконання роботи є наказ № \_\_ по ВНТУ від «\_\_» \_\_\_\_\_ 2023 р., та індивідуальне завдання на МКР, затверджене протоколом № \_\_ засідання кафедри САІТ від «\_\_» \_\_\_\_\_ 2023 р.

### 2. Джерела розробки:

- Крижановський С. М., Мокін В. Б, Піроговський А. В., Просторовий аналіз даних забруднення атмосферного повітря Вінницької області за даними державного моніторингу у 2020 р. *LI Науково-технічна конференція факультету інтелектуальних інформаційних технологій та автоматизації*: зб. Матеріалів конференції. Віння, 2022. URL: <https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2022/paper/view/14910>
- Технологія проектування мережі спостережень якості атмосферного повітря регіону на основі методу аналізу ієрархій / В. Б. Мокін, С. М. Крижановський, А. Р. Ящолт, Д. О. Шмундяк // Наукові праці ВНТУ. – 2021. – № 4. URL: <https://praci.vntu.edu.ua/index.php/praci/article/view/642>

### 3. Мета і призначення роботи:

Метою роботи є підвищення вивченості стану атмосферного повітря міста Вінниці шляхом проведення розвідувального та системного аналізу з використанням інформаційних технологій.

### 4. Вихідні дані для проведення робіт:

- Kaggle dataset “Air quality in Vinnytsia 2023 (September)” <https://www.kaggle.com/datasets/pirogovskiy/air-quality-in-vinnytsia-2023-september>;
- набір статистичних даних державного моніторингу атмосферного повітря в Вінницькій області за 2016-2021 роки.

### 5. Методи дослідження:

- розвідувальний аналіз;
- прогнозування даних;
- методи штучного інтелекту, а саме використання моделей: ARIMA та Prophet.

### 6. Етапи роботи і терміни їх виконання:

- a) Загальна характеристика об’єкту досліджень..... \_\_\_\_\_ – \_\_\_\_\_
- b) Вибір оптимальних інформаційних технологій для здійснення системного аналізу та прогнозування даних..... \_\_\_\_\_ – \_\_\_\_\_
- c) Збирання та підготовка до системного аналізу даних забруднення атмосферного повітря міста Вінниці..... \_\_\_\_\_ – \_\_\_\_\_
- d) Системний аналіз та прогнозування даних ..... \_\_\_\_\_ – \_\_\_\_\_
- e) Економічна частина ..... \_\_\_\_\_ – \_\_\_\_\_
- f) Оформлення пояснювальної записки..... \_\_\_\_\_ – \_\_\_\_\_

### 7. Очікувані результати та порядок реалізації:

Визначити найкраще місце розташування для встановлення нових постів моніторингу атмосферного повітря у місті Вінниця.

#### 8. Вимоги до розробленої документації

Текстова та ілюстративна частини роботи оформлені у відповідності до вимог «Методичних вказівок до виконання магістерських кваліфікаційних робіт для студентів спеціальності 126 «Інформаційні системи та технології» (освітня програма «Інформаційні технології аналізу даних та зображень»).

#### 9. Порядок приймання роботи

Публічний захист ..... «\_\_» \_\_\_\_\_ 2023 р.

Початок розробки ..... «\_\_» \_\_\_\_\_ 2023 р.

Граничні терміни виконання МКР ..... «\_\_» \_\_\_\_\_ 2023 р.

Розробив студент групи 2ІСТ-22м \_\_\_\_\_ Андрій ПРОГОВСЬКИЙ

## Додаток Б

Протокол перевірки кваліфікаційної роботи на наявність текстових запозичень

Назва роботи: «Інформаційна технологія аналізу та прогнозування стану атмосферного повітря міста Вінниці. Частина 1. Формування датасету та системний аналіз даних»

Тип роботи: магістерська кваліфікаційна робота


Підрозділ: кафедра САІТ

## Показники звіту подібності Unicheck

Оригінальність 95,5 % Схожість 4,5 %

Аналіз звіту подібності (відмітити потрібне)

- Запозичення, виявлені у роботі, оформлені коректно і не містять ознак плагіату.
- Виявлені у роботі запозичення не мають ознак плагіату, але їх надмірна кількість викликає сумніви щодо цінності роботи і самостійності її автора. Роботу направити на розгляд експертної комісії кафедри.
- Виявлені у роботі запозичення є недобросовісними і мають ознаки плагіату та/або в ній містяться навмисні спотворення тексту, що вказують на спроби приховування недобросовісних запозичень.

Особа, відповідальна за перевірку  Сергій ЖУКОВ  
(п.п.мс)

Ознайомлені з повним звітом подібності, який був згенерований системою Unicheck щодо роботи.

Автор роботи

  
(п.п.мс)

Андрій ПІРОГОВСЬКИЙ

Керівник роботи

  
(п.п.мс)

Святослав КРИЖАНОВСЬКИЙ



## Додаток В

## Фрагмент лістингу програмни

```
# Thanks to the users for the help to make this notebook:  
# https://www.kaggle.com/code/vbmokin/air-quality-city-mapping  
  
# Work with data - the main Python libraries  
import numpy as np  
import pandas as pd  
  
# For import data  
import os  
import gc  
import requests  
from pprint import pprint  
import json  
  
from datetime import datetime  
  
# Visualization  
import matplotlib.pyplot as plt  
import plotly.express as px  
import plotly.graph_objects as go  
  
# Others  
import seaborn as sns  
import scipy  
from sklearn import *  
import statsmodels.api as sm  
import torch  
import itertools  
from pathlib import Path  
from scipy.stats import zscore  
from sklearn.metrics import r2_score  
from statsmodels.tsa.arima.model import ARIMA  
from prophet import Prophet  
  
import folium #for building maps  
from IPython.display import IFrame  
  
import warnings  
warnings.simplefilter('ignore')
```

```
# Downloading data
```

```
path = Path('../input/air-quality-in-vinnytsia-2023-september')
df_air = pd.read_csv(path/'Air quality in September 2023.csv')
```

```
# Data view
```

```
df_air.head(10)
# Drop columns named 'Unnamed: 1' through 'Unnamed: 1000'
df_air_cleaned = df_air.drop(df_air.columns[df_air.columns.str.startswith('Unnamed: ')], axis = 1)
```

```
# Cheking the cleaned dataset
```

```
df_air_cleaned.head(10)
```

```
# Dataset description
```

```
df_air_cleaned.describe()
```

```
# Basic information of dataset
```

```
df_air_cleaned.info()
```

```
# Find null values
```

```
df_air_cleaned.isnull().sum()
```

```
# Drop rows with null values from dataset
```

```
df_air_cleaned_2 = df_air_cleaned.dropna()
```

```
# Creating new dataset for building map (List of stations)
```

```
df_air_for_map = pd.DataFrame({
    'station_unique_id': ['650', '767', '790', '1811', '1813', '1825', '1830', '1834', '1866', '246', '256', '271', '274', '281', '315', '90'],
    'station_name': ['дiм', 'Соборна 36', 'станція Вишенька', 'ЕсоСity станція моніторингу Хмільник', 'Якушинці', 'Центральна площа', 'Центр Безпеки громадян Війтівецької сільської ради', 'Есо Stryzh', 'Северинівка', 'vinnytsia-246', 'vinnytsia-256', 'vinnytsia-271', 'vinnytsia-274', 'vinnytsia-281', 'vinnytsia-315', 'vinnytsia-90'],
    'Lat': ['49.343484', '49.232859', '49.2270093', '49.558012', '49.2571196', '49.105637', '49.6288314', '49.3064814', '48.516638', '49.2234644', '49.20485607', '49.23688714', '49.22477769', '49.24192687', '49.24368759', '49.21773365'],
    'Long': ['28.72983', '28.470453', '28.4189984', '27.956897', '28.3671704', '29.206617', '27.89884', '28.4828855', '29.184421', '28.40095116', '28.5288355', '28.51330549', '28.42491984', '28.46202092', '28.49643223', '28.44979464']
})
```

```

# Converting 'Lat' and 'Long' columns to float
df_air_for_map['Lat'] = df_air_for_map['Lat'].astype(float)
df_air_for_map['Long'] = df_air_for_map['Long'].astype(float)

df_air_for_map

import folium
from IPython.display import IFrame

# Create a base map using mean of 'Lat' and 'Long' columns
map_stations = folium.Map(location=[df_air_for_map['Lat'].mean(), df_air_for_map[
'Long'].mean()], zoom_start = 9)

# Add markers for each station
for index, station in df_air_for_map.iterrows():
    folium.Marker([float(station['Lat']), float(station['Long'])], popup=str(station['stati
on_name'])).add_to(map_stations)

# Save the map to an HTML file
map_path = 'stations_map.html'
map_stations.save(map_path)

# Display the map in the Jupyter Notebook using IFrame
IFrame(map_path, width=1400, height=650)

# Data selection
df_air_selected = df_air_cleaned_2[['stations_id', 'stations_params_time', 'stations_pa
rams_name', 'stations_params_value']]
df_air_selected

# Create separate DataFrames for each parameter
pm1_data = df_air_selected[df_air_selected['stations_params_name'] == 'PM1.0']

pm25_data = df_air_selected[df_air_selected['stations_params_name'] == 'PM2.5']

pm10_data = df_air_selected[df_air_selected['stations_params_name'] == 'PM10']

voc_data = df_air_selected[df_air_selected['stations_params_name'] == 'VOC (H2CO
)']

temperature_data = df_air_selected[df_air_selected['stations_params_name'] == 'Tem
perature']

```

```

humidity_data = df_air_selected[df_air_selected['stations_params_name'] == 'Humidity']

pressure_data = df_air_selected[df_air_selected['stations_params_name'] == 'Pressure']

radiation_data = df_air_selected[df_air_selected['stations_params_name'] == 'RAD']

# Data view
pm1_data
# Calculate average PM1.0 value from stations
pm1_data['stations_params_value'] = pd.to_numeric(pm1_data['stations_params_value'], errors='coerce')

average_pm1_value = pm1_data['stations_params_value'].mean()

# Define normal PM1.0 value
normal_pm1_value = 10.0

# Plotting
plt.figure(figsize=(8, 6))
plt.bar(['Average PM1.0 from Stations', 'Normal PM1.0'], [average_pm1_value, normal_pm1_value], color=['blue', 'orange'])
plt.ylabel('PM1.0 Values')
plt.title('Comparison of Average PM1.0 from Stations and Normal PM1.0')
plt.ylim(0, max(normal_pm1_value, average_pm1_value) + 5) # Set y-axis limit for better visualization
plt.show()

# Set date as date & time types
pm1_data['stations_params_time'] = pd.to_datetime(pm1_data['stations_params_time'], format = '%Y-%m-%d %H:%M:%S', errors = 'ignore')

# Renaming
df_air_pm1 = pm1_data[['stations_params_time', 'stations_params_value']].reset_index(drop = True)
df_air_pm1.columns = ['time', 'values']

# Dropping the columns
df_air_pm1.index = df_air_pm1['time']
df_air_pm1 = df_air_pm1.drop(columns=['time'])
df_air_pm1

```

```

# Build the plot (Values and Time)
df_air_pm1['values'] = df_air_pm1['values'].astype('float')
plt.figure().set_figwidth(16)
plt.xlabel('Time')
plt.ylabel('Value')
plt.title('Dust Pollution (PM1)')

df_air_pm1['values'].plot()

pm1_data['stations_params_time'] = pd.to_datetime(pm1_data['stations_params_time']
)

# Set your normal PM1.0 value
normal_pm1_value = 10.0 # Replace this with your actual normal PM1.0 value

# Plotting
plt.figure(figsize=(12, 6))
plt.plot(pm1_data['stations_params_time'], pm1_data['stations_params_value'], label
='PM1.0 Values', color='gray')
plt.axhline(y=normal_pm1_value, color='red', linestyle='--', label='Normal PM1.0 Va
lue')
plt.xlabel('Time')
plt.ylabel('PM1.0 Values')
plt.title('Dust Pollution (PM1.0)')
plt.legend()
plt.show()

# Calculate Z-scores
pm1_data['Z_Score'] = zscore(pm1_data['stations_params_value'])

# Define a threshold for Z-score (e.g., 2 or 3)
z_score_threshold = 2

# Identify anomalous data points
anomalous_pm1_data = pm1_data[abs(pm1_data['Z_Score']) > z_score_threshold]

# Plotting the entire dataset
plt.figure(figsize=(12, 7))
plt.scatter(pm1_data['stations_params_time'], pm1_data['stations_params_value'], lab
el='PM1.0 Values', color='green', alpha=0.5)

# Plotting the anomalous data points
plt.scatter(anomalous_pm1_data['stations_params_time'], anomalous_pm1_data['stati
ons_params_value'], label='Anomalous PM1.0 Values', color='red')

```

```

plt.xlabel('Time')
plt.ylabel('PM1.0 Values')
plt.title('PM1.0 Values with Anomalous Data Points Highlighted')
plt.legend()
plt.show()

# Create the ARIMA model
model = sm.tsa.ARIMA(df_air_pm1['values'], order = (1, 0, 1))

# Fit the model
model_fit = model.fit()
print(model_fit.summary)

# Get the forecasted values
predictions = model_fit.predict()

# Plot the actuals and forecasted values
plt.figure(figsize = (12, 5), dpi = 100)
plt.plot(df_air_pm1.index, df_air_pm1['values'], label = 'Actual')
plt.plot(df_air_pm1.index, predictions, label = 'Forecast')
plt.title('Forecast vs Actuals by ARIMA')
plt.legend(loc = 'upper left', fontsize = 8)
plt.show()

# Calculate R-squared score
r2 = r2_score(df_air_pm1['values'], predictions)
rounded_r2_pm1 = round(r2, 2)

print("R-squared score:", rounded_r2_pm1)

# Set date as date & time types
pm1_data['stations_params_time'] = pd.to_datetime(pm1_data['stations_params_time']
], format = '%Y-%m-%d %H:%M:%S', errors = 'ignore')

# Renaming
df_air_pm1_prophet = pm1_data[['stations_params_time', 'stations_params_value']].r
eset_index(drop = True)
df_air_pm1_prophet.columns = ['time', 'values']
df_air_pm1_prophet

df_prophet = pd.DataFrame()
df_prophet['ds'] = df_air_pm1_prophet['time'] # Replace your column with the actual
column name
df_prophet['y'] = df_air_pm1_prophet['values'] # Replace your column with the actu
al column name

```

```

# Build Prophet model with parameters and structure
model = Prophet(daily_seasonality = True, weekly_seasonality = False, yearly_seasonality = False,
                changepoint_range = 0.7, changepoint_prior_scale = 0.4, seasonality_mode = 'multiplicative')

# Training model for df_air
model.fit(df_prophet)

# Create a DataFrame with future dates for forecasting
future = model.make_future_dataframe(periods = 20) # Adjust the number of periods as needed

# Generate forecast
forecast = model.predict(future)

# Plot the forecast
fig = model.plot(forecast)
plt.show()

# Draw plot with the components (trend and seasonalities) of the forecasts
figure_component = model.plot_components(forecast)

# Output the forecast for the next time
forecast[['yhat_lower', 'yhat', 'yhat_upper']] = forecast[['yhat_lower', 'yhat', 'yhat_upper']].round(2)
df_forecast = forecast[['ds', 'yhat_lower', 'yhat', 'yhat_upper']]
df_forecast.tail(5)

forecasting_period = 4 # The number of steps in the forecast, in hours, integer more 1

# Calculation training data by the model
y_training = forecast['yhat'][:-forecasting_period]
y_training[-5:]

# Assuming df_air_radiation_prophet contains your actual training data
# Merge forecasted values with actual values based on the date column (ds)
df_comparison = pd.merge(df_forecast, df_air_pm1_prophet, left_on = 'ds', right_on = 'time', how = 'left')

# Drop rows with missing actual values (NaN)
df_comparison = df_comparison.dropna(subset = ['values'])

```

```
# Calculate R-squared score  
r2 = r2_score(df_comparison['values'], df_comparison['yhat'])  
print("R-squared score:", round(r2, 2))
```



Додаток Г

## ІЛЮСТРАТИВНА ЧАСТИНА

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ СТАНУ  
АТМОСФЕРНОГО ПОВІТРЯ МІСТА ВІННИЦІ. ЧАСТИНА 1.  
ФОРМУВАННЯ ДАТАСЕТУ ТА СИСТЕМНИЙ АНАЛІЗ ДАНИХ

Нормоконтроль: к.т.н., доцент

\_\_\_\_\_ Сергій ЖУКОВ

«\_\_\_\_» \_\_\_\_\_ 2023 р.

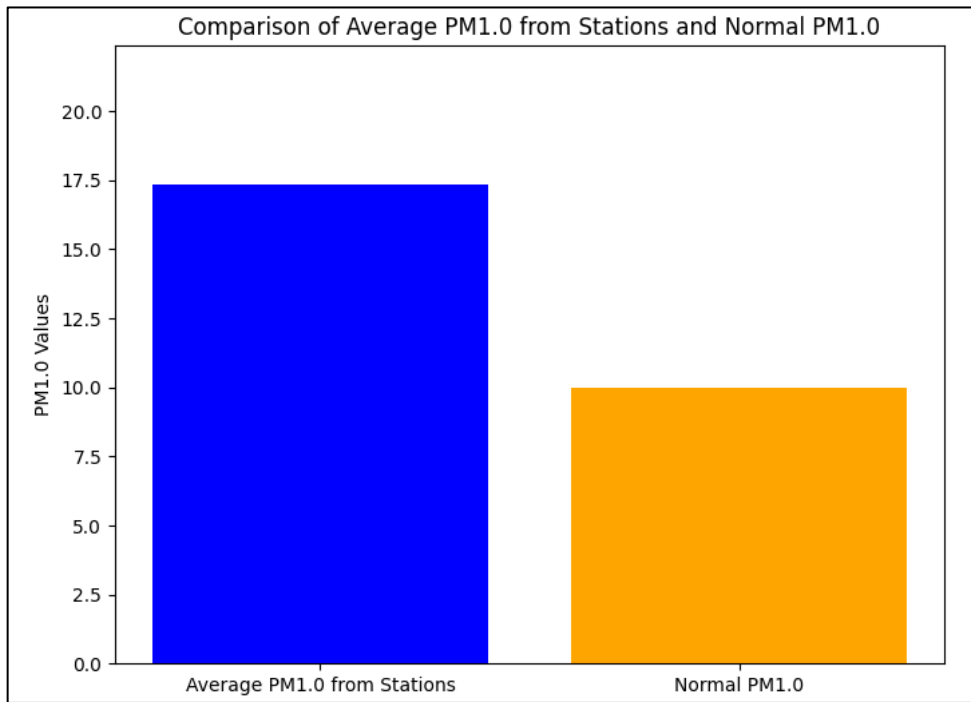


Рисунок Г.1 – Зрівняння показників частинок пилу розміром 1 мкм з встановленою нормою

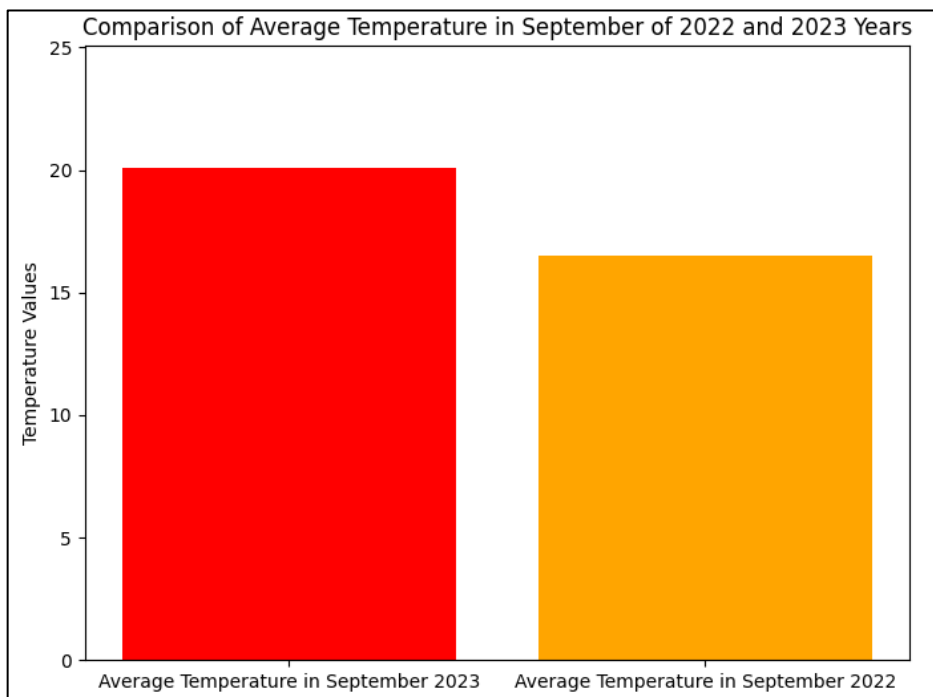


Рисунок Г.2 – Порівняння середньої температури повітря в вересні місяці за 2022 та 2023 роки в м. Вінниця

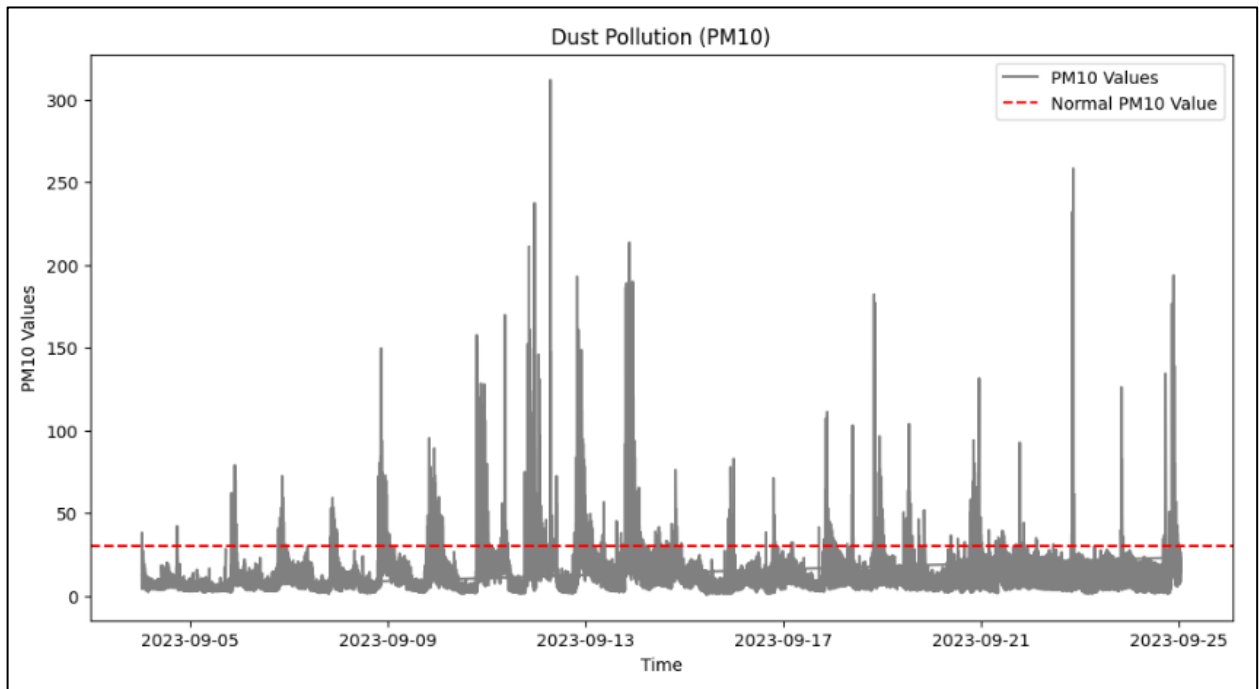


Рисунок Г.3 – Значення частинок пилу розміром 10 мкм відносно встановленої норми

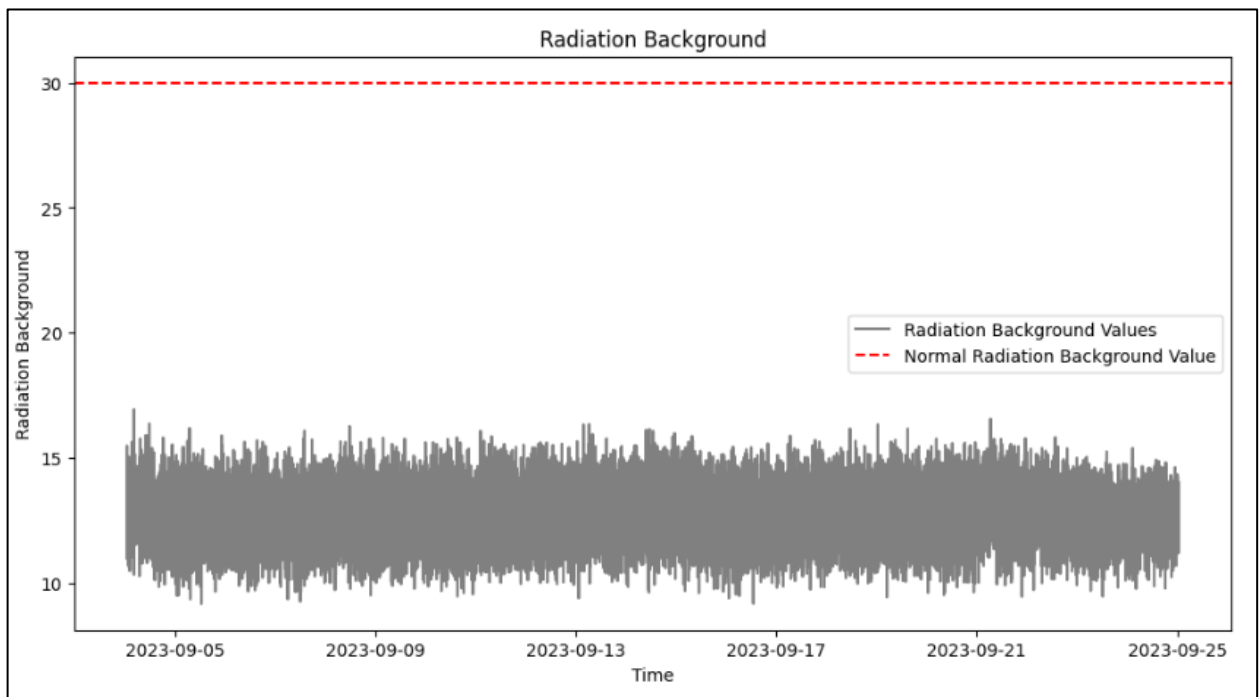


Рисунок Г.4 – Значення радіаційного фону відносно встановленої норми

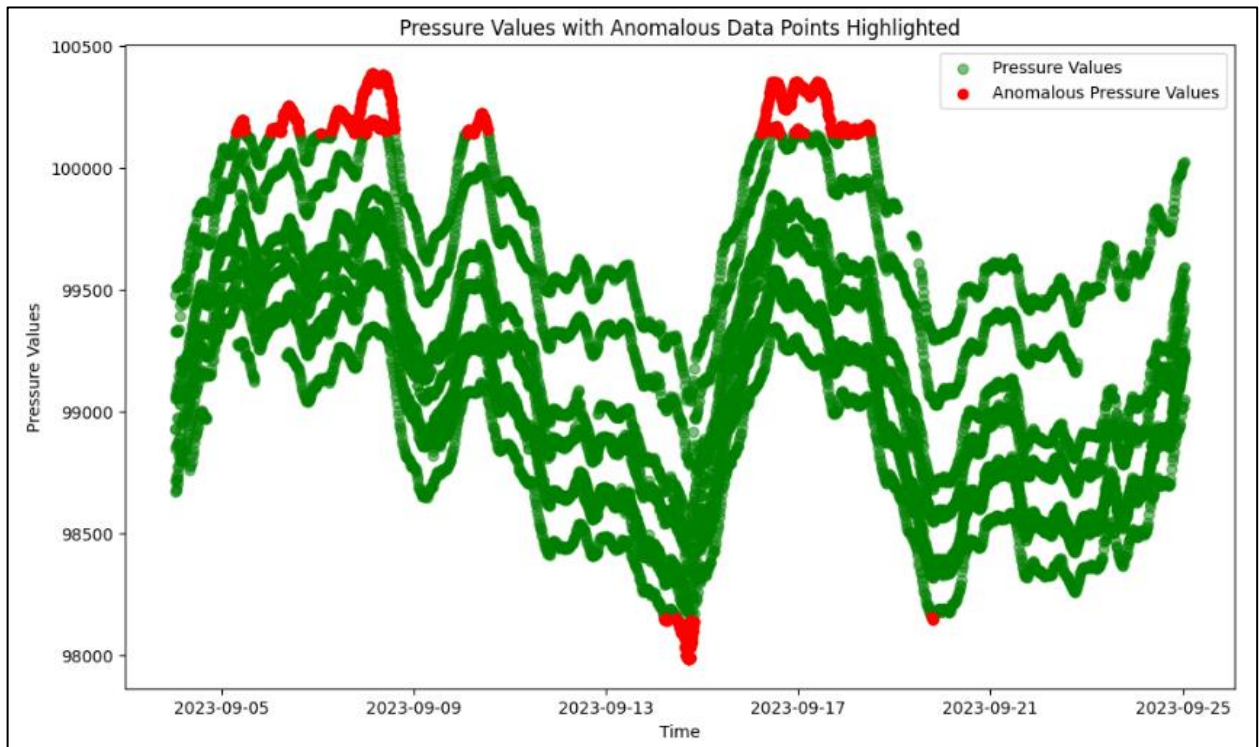


Рисунок Г.5 – Графік з відображенням аномальних значень показнику атмосферного тиску

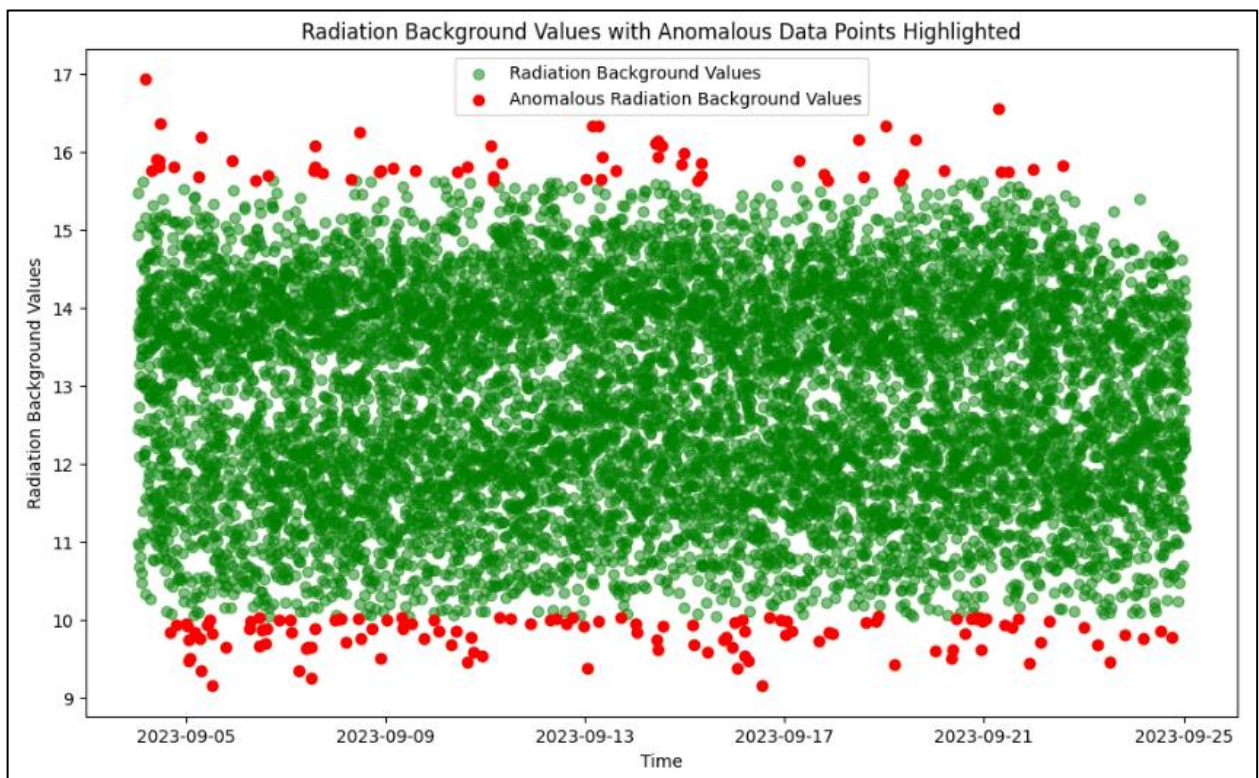


Рисунок Г.6 – Графік з відображенням аномальних значень показнику радіаційного фону

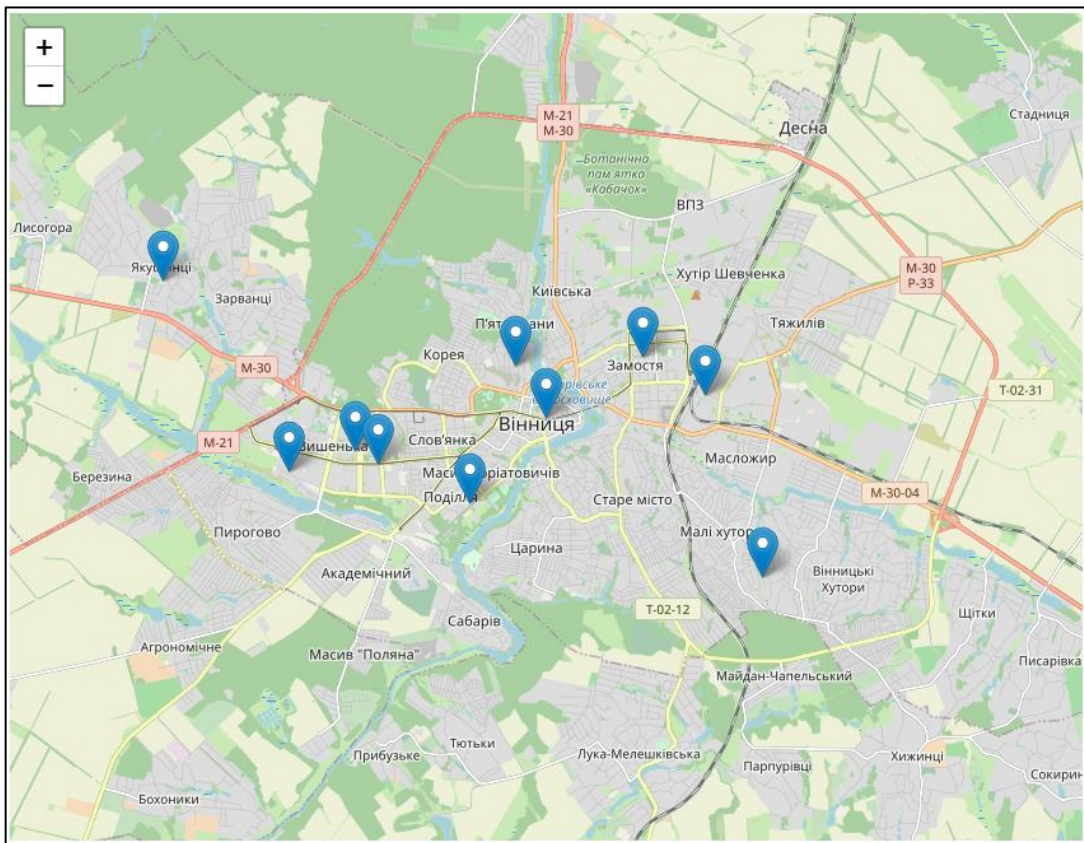


Рисунок Г.7 – Побудована мапа розташування станцій моніторингу повітря

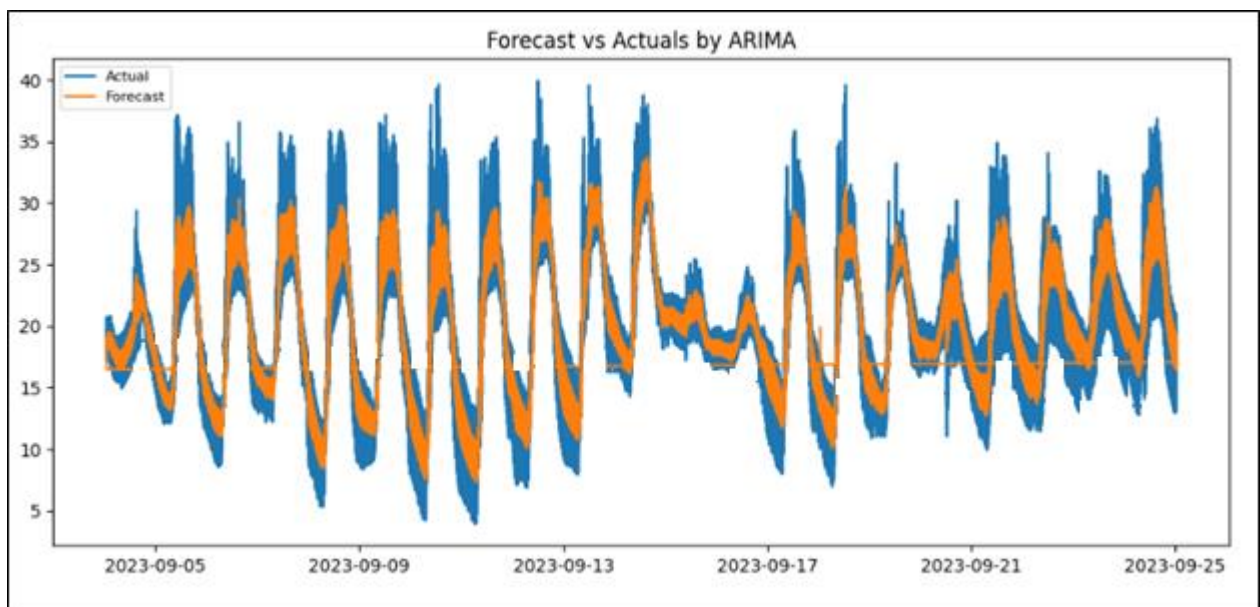


Рисунок Г.8 – Графік прогнозу температури

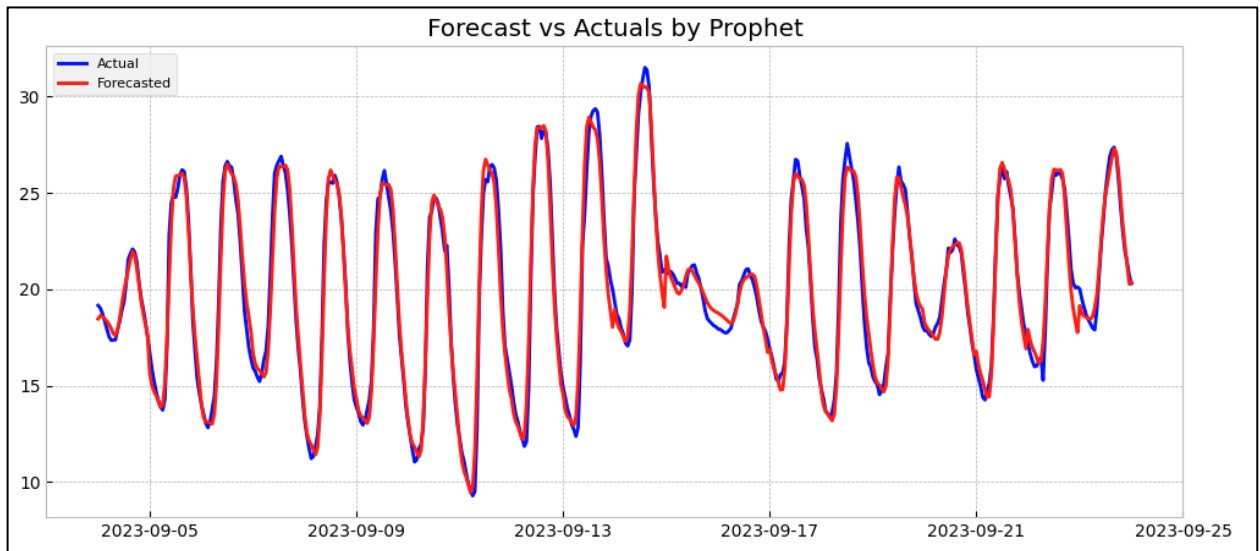


Рисунок Г.9 – Графік прогнозування значення температури за допомогою моделі Prophet