

Вінницький національний технічний університет  
Факультет комп'ютерних систем і автоматики  
Кафедра системного аналізу та інформаційних технологій

## МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА

на тему:

### **«Інформаційна технологія аналізу та прогнозування якості атмосферного повітря у місті Вінниці за даними громадського моніторингу (комплексна МКР) Частина I. Прогнозування даних»**

Виконав: студент 2 курсу, групи 2ІСТ-20м спеціальності 126 – «Інформаційні системи та технології»

\_\_\_\_\_ Бондарчук А. Р.

Керівник: д.т.н., проф. каф. САІТ

\_\_\_\_\_ Мокін О. Б.

«\_\_» \_\_\_\_\_ 2021 р.

Опонент: к.т.н., доц. каф. АІТ

\_\_\_\_\_ Іванов Ю.Ю.

«\_\_» \_\_\_\_\_ 2021 р.

**Допущено до захисту**

Завідувач кафедри САІТ

\_\_\_\_\_ д.т.н., проф. Мокін В. Б.

«\_\_» \_\_\_\_\_ 2021 р.

Вінниця ВНТУ – 2021 рік

Вінницький національний технічний університет  
Факультет комп'ютерних систем і автоматики  
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

\_\_\_\_\_ д.т.н., проф. Мокін В. Б.

«\_\_\_» \_\_\_\_\_ 2021 р.

**ЗАВДАННЯ  
НА МАГІСТЕРСЬКУ КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ  
Бондарчуку Артуру Руслановичу**

1. Тема роботи: «Інформаційна технологія аналізу та прогнозування якості атмосферного повітря у місті Вінниці за даними громадського моніторингу (комплексна МКР) Частина I. Прогнозування даних», керівник роботи: Мокін О. Б., д.т.н., проф. каф. САІТ, затверджені наказом закладу вищої освіти від «\_\_\_» \_\_\_\_\_ 2021 року №\_\_
2. Строк подання студентом роботи «\_\_\_» \_\_\_\_\_ 2021 року
3. Вихідні дані до роботи:  
Перелік атрибутивних даних про якість повітря, взятих з ресурсів Eco City та SaveEcoBot.
4. Зміст текстової частини:
  - аналіз предметної області;
  - вибір оптимальних налаштувань та огляд набору даних;
  - проектування задачі;
  - економічна частина.
5. Перелік ілюстративного матеріалу (з точним зазначенням обов'язкових креслень):
  - візуалізація значень з поста EcoCity;
  - візуалізація значень з поста EcoCity\_848;
  - графік з аномальними даними з поста EcoCity\_337;
  - графік з аномальними даними з поста EcoCity\_848;
  - графік з прогнозом кількості пилу для поста EcoCity\_337;
  - графік з прогнозом кількості пилу для поста EcoCity\_848.

## 6. Консультанти розділів МКР

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
3	Мокін В.Б., д.т.н., проф. каф. САІТ		
4	Ратушняк О.Г., к.е.н., доц. каф. ЕПВМ		

7. Дата видачі завдання « \_\_\_ » \_\_\_\_\_ 2021 року

## КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів МКР	Строк виконання етапів роботи	Примітка
1	Аналіз предметної області	09.2021	
2	Огляд проблем створення інформаційної технології прогнозу якості повітря в місті Вінниця	10.2021	
3	Розробка інформаційної технології	10.2021	
4	Економічна частина	11.2021	
5	Оформлення матеріалів до захисту МКР	11.2021	

Студент

\_\_\_\_\_

Бондарчук А. Р.

Керівник роботи

\_\_\_\_\_

Мокін О. Б.

## АНОТАЦІЯ

УДК 004.09

Бондарчук А.Р. Інформаційна технологія аналізу та прогнозування якості атмосферного повітря у місті Вінниці за даними громадського моніторингу (комплексна МКР) Частина I. Прогнозування даних. Магістерська кваліфікаційна робота зі спеціальності 126 – інформаційні системи та технології, освітньо-професійна програма – інформаційні технології аналізу даних та зображень. Вінниця: ВНТУ, 2021. 112 с.

На укр. мові. Бібліогр.: 22 назв; рис.: 77; табл.: 6.

В магістерській кваліфікаційній роботі звернено увагу на проблему прогнозування якості повітря. Запропоновані технології, які допоможуть оптимально реалізувати дану технологію з точки зору можливості її використання на персональних комп'ютерах із вільним для розповсюдження програмним забезпеченням. Здійснено реалізацію прогнозу, а також розроблено рекомендації по її впровадженню та експлуатації. Об'єкт досліджень – процес моніторингу якості атмосферного повітря у місті Вінниця.

Ілюстративна частина складається з 6 плакатів із результатами моделювання.

У розділі економічної частини розглянуто питання про доцільність розробки та впровадження інформаційної технології прогнозування якості повітря.

Ключові слова: інформаційна технологія, моніторинг, повітря, якість повітря, прогноз, база даних.

## **ABSTRACT**

Bondarchuk A.R. Information technology of analysis and forecasting of air quality in the city of Vinnytsia according to the data of public monitoring (complex MCR) Part I. Forecasting of data. Master's thesis in specialty 126 - information systems and technologies, educational and professional program - information technology data and image analysis. Vinnytsia: VNTU, 2021. 112 p.

In Ukrainian language. Bibliogr .: 22 titles; fig .: 77; table: 6.

In the master's qualification work attention is paid to the problem of air quality forecasting. Technologies are proposed that will help to optimally implement this technology in terms of the possibility of its use on personal computers with free software for distribution. The forecast has been implemented, and recommendations for its implementation and operation have been developed. The object of research is the process of monitoring air quality in the city of Vinnytsia.

The illustrative part consists of 6 posters with simulation results.

In the section of the economic part the question of expediency of development and introduction of information technology of air quality forecasting is considered.

Key words: information technology, monitoring, air, air quality, forecast, databas

## ЗМІСТ

ВСТУП.....	4
1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ.....	6
1.1 Суть забруднення повітря .....	6
1.1.1 Дрібні частки .....	7
1.1.3 Чадний газ .....	8
1.1.3 Діоксид сірки .....	8
1.1.4 Діоксид азоту .....	9
1.1.5 Озон .....	9
1.1.6 Свинець .....	10
1.1.7 Парникові гази.....	10
1.2 Глобальне охоплення забруднення повітря.....	11
1.3 Джерела даних про забруднення навколишнього повітря.....	12
1.3.1 Стаціонарні станції екологічного моніторингу. ....	12
1.3.2 Бездротові сенсорні мережі.....	13
1.3.3 Супутникові дані .....	14
1.4 Сучасний стан моніторингу у місті Вінниця.....	15
1.5 Висновки .....	18
2 ВИБІР ОПТИМАЛЬНИХ НАЛАШТУВАНЬ ТА ОГЛЯД НАБОРУ ДАНИХ.....	19
2.1 Опис FbProphet .....	19
2.2 Вибір мови програмування .....	30
2.3 Огляд вхідного набору даних.....	35
2.4 Висновки .....	37
3 ПРОЕКТУВАННЯ ЗАДАЧІ.....	38
3.1 Щоденне прогнозування даних .....	38
3.1.1 Імпорт бібліотек та завантаження даних .....	38
3.1.2 Прогнозування даних.....	41
3.2 Щогодинне прогнозування даних .....	55
3.2.1 Імпорт бібліотек та завантаження даних .....	55

3.2.2 Прогнозування даних.....	58
3.3 Висновки .....	73
4 ЕКОНОМІЧНА ЧАСТИНА.....	74
4.1 Оцінювання комерційного потенціалу розробки.....	74
4.2 Прогнозування витрат на виконання науково-дослідної роботи.....	77
4.3 Розрахунок економічної ефективності науково-технічної розробки.....	82
4.4 Розрахунок ефективності вкладених інвестицій та періоду їх окупності...	83
4.5 Висновки .....	86
ВИСНОВКИ.....	87
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	89
Додаток А (обов'язковий). Технічне завдання .....	93
Додаток Б (обов'язковий). Протокол перевірки кваліфікаційної роботи .....	95
Додаток В (довідниковий). Лістинг програмного коду .....	96
Додаток Г (обов'язковий). Ілюстративна частина .....	106

## ВСТУП

**Актуальність теми.** В даний час для здоров'я людини першорядне значення має моніторинг та прогнозування якості повітря. Це особливо стосується великих міст, а також територій з несприятливими географічними та метеорологічними умовами з великою кількістю джерел викидів, включаючи неконтрольовані джерела. У всьому світі велику увагу держави приділяють покращенню екологічних обставин у густонаселених районах.

**Мета і завдання роботи.** Метою даної роботи є підвищення точності прогнозування якості атмосферного повітря у м. Вінниці за даними громадського моніторингу. За допомогою спеціального середовища, можливо здійснити аналіз даних та прогноз рівня забрудненості атмосферного повітря.

Розробка інформаційної технології аналізу та прогнозування якості атмосферного повітря передбачає виконання таких задач:

- розвідувальний аналіз даних, отриманих з постів розміщених у місті Вінниця за певний період;
- вибір середовища та оптимальних технологій;
- підготовка набору даних;
- пошук аномальних даних;
- здійснення прогнозу даних.

**Об'єктом дослідження** магістерської кваліфікаційної роботи є процес моніторингу якості атмосферного повітря у місті Вінниця.

**Предметом дослідження** магістерської кваліфікаційної роботи є моделі та методи прогнозування забрудненості атмосферного повітря у місті Вінниця.

**Новизна одержаних результатів.** Дістала подальший розвиток інформаційна технологія прогнозування якості атмосферного повітря у регіоні за даними громадського моніторингу за рахунок урахування аномалій, що дозволяє підвищити точність такого прогнозування.

**Практичне значення** роботи сприятиме покращенню прогнозу якості атмосферного повітря у місті. Результати роботи мають цінність для



екологічного контролю забруднення атмосферного повітря та для людей з алергічною реакцією на пил.

**Апробація результатів магістерської кваліфікаційної роботи.**

Результати кваліфікаційної роботи доповідались на всеукраїнській науково-практичній інтернет-конференції: «Молодь в науці: дослідження, проблеми, перспективи» (Вінниця, 2021–2022 рр.).

**Публікації результатів магістерської кваліфікаційної роботи.**

В ході виконання магістерської кваліфікаційної роботи опубліковано тези у збірнику матеріалів конференції «Молодь в науці: дослідження, проблеми, перспективи» (Вінниця, 2021–2022 рр.) [1].

# 1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

## 1.1 Суть забруднення повітря

Забруднення повітря є наявність речовин в атмосфері, які шкідливі для здоров'я в організмі людини та інших живих істот, або завдали шкоди клімату або матеріалам. Існує багато різних типів забруднювачів повітря, таких як гази (включаючи аміак, оксид вуглецю, діоксид сірки, оксиди азоту, метан, діоксид вуглецю), тверді частинки (як органічні, так і неорганічні) і біологічні молекули [2].

Чисте сухе повітря складається в основному з азоту та кисню – 78% та 21% відповідно за обсягом. 1 відсоток є сумішшю інших газів, в основному аргону (0,9 відсотка), разом із незначною кількістю вуглекислого газу, метану, водню, гелію і т. д. Водяна пара також є нормальним, хоча і дуже мінливим компонентом атмосфери, зазвичай становить від 0,01 до 4 відсотків за обсягом. У дуже вологих умовах вологість повітря може сягати 5 відсотків.

Існує шість основних забруднювачів повітря, які були визначені як критерії забруднювачів – критерії, які означають, що концентрація цих забруднювачів в атмосфері корисні як індикатори загальної якості повітря.

До газоподібних забруднювачів повітря, що викликають найбільше занепокоєння в міських умовах, відносяться діоксид сірки, діоксид азоту та монооксид вуглецю. Вони викидаються безпосередньо у повітря з копалин видів палива, таких як мазут, бензин і природний газ, які спалюються на електростанціях, автомобілях та інших джерелах горіння. Озон (ключовий компонент смогу) також є газоподібним забруднювачем, він утворюється в атмосфері в результаті складних хімічних реакцій, що протікають між діоксидом азоту та різними леткими органічними сполуками (наприклад, парами бензину) [3].

Завислі надзвичайно дрібних твердих або рідких частинок, званих «макрочастинками» (наприклад, сажа, пил, дим, дим, туман), особливо розміром менше 10 мікрометрів (мільйонних часток метра), є значними забруднювачами повітря через їх дуже шкідливий вплив на здоров'я людини. Вони викидаються різними промисловими процесами, електростанціями, що працюють на вугіллі чи мазуті, системами опалення житлових будинків та автомобілями. Пари свинцю (частинки в повітрі розміром менше 0,5 мкм) є особливо токсичними і є важливим забруднювачем багатьох дизельних палив.

За винятком свинцю, зазначені в умовах забруднювачі викидаються у промислово розвинених країнах з дуже високими темпами, які зазвичай вимірюються мільйонами тонн на рік. Всі, крім озону, викидаються прямо в атмосферу з різних джерел. Вони регулюються, насамперед, шляхом встановлення стандартів якості навколишнього повітря, які є максимально допустимими концентраціями кожного з критеріїв забруднювача в атмосфері, незалежно від його походження. Шість критеріїв забруднювачів по черзі описані нижче.

### 1.1.1 Дрібні частки

Дуже маленькі фрагменти твердих матеріалів або рідких крапель, зважених у повітрі, називаються тверді частки. За винятком свинцю, що переноситься повітрям, який розглядається як окрема категорія, вони характеризуються за розміром і фазою (тобто тверді або рідкі), а не за хімічним складом. Наприклад, тверді частинки діаметром від 1 до 100 мкм називаються частинками пилу, а тверді частинки діаметром менше 1 мкм – пари [4].

Найбільш небезпечними з точки зору їх впливу на здоров'я людини є тверді частинки діаметром менше 10 мкм, оскільки вони можуть вдихатися глибоко у легені та потрапляти у пастку нижніх дихальних шляхів. Деякі частинки, такі як волокна азбесту, відомі канцерогени (агенти, що викликають рак) і багато вуглецевих частинок, наприклад, сажа, вважаються

канцерогенними. Основними джерелами викидів твердих частинок є електростанції, що працюють на викопному паливі, виробничі процеси, системи опалення житлових приміщень, що працюють на викопному паливі, та автомобілі, що працюють на бензині.

### 1.1.3 Чадний газ

Чадний газ (окис вуглецю) – це невидимий газ без аромату, що виникає внаслідок неповного згорання. Це найвідоміший із критеріїв забруднювачів. Дорожні транспортні методи, що працюють на бензині, є основним джерелом цього газу, але системи опалення житлових будинків і деякі виробничі процеси ще викидають важливі числа цього газу. Електростанції виділяють порівняно мало чадного газу, так як вони ретельно спроектовані і експлуатуються для граничного збільшення продуктивності згорання. Вплив окису вуглецю може бути досить небезпечним, тому що він просто тіснить повітря в кровотоку, що призводить до задухи при досить високих концентраціях і тривалості впливу [5].

### 1.1.3 Діоксид сірки

Безбарвний газ з різким запахом, діоксид сірки утворюється при згоранні вугілля або нафти, що містять сірку в якості домішки. Більшість викидів діоксиду сірки походить від електростанцій, дуже мало надходить із мобільних джерел. Цей їдкий газ може спричинити подразнення очей та горла, а також пошкодити легеневу тканину при вдиханні.

Діоксид сірки також вступає в реакцію з киснем та водяною парою в повітрі, утворюючи туман сірчаної кислоти, що досягає землі як компонент кислотного дощу. Вважається, що кислотні дощі пошкодили або знищили рибу та рослини у багатьох тисячах озер та струмків у деяких частинах Європи, на північному сході США, на південному сході Канади та у деяких частинах

Китаю. Це також викликає корозію металів та пошкодження відкритих поверхонь будівель та громадських пам'яток.

#### 1.1.4 Діоксид азоту

З кількох форм оксидів азоту найбільше занепокоєння викликає діоксид азоту – їдкий, драгівливий газ. Відомо, що він викликає набряк легенів – скупчення надмірної кількості рідини у легенях. Діоксид азоту також вступає у реакцію в атмосфері з утворенням азотної кислоти, що посилює проблему кислотних дощів. Крім того, діоксид азоту відіграє роль в утворенні фотохімічного смогу – червонувато-коричневого серпанку, який часто спостерігається в багатьох міських районах і створюється реакціями, викликаними сонячним світлом у нижніх шарах атмосфери [6].

Оксиди азоту утворюються, коли температура згоряння досить висока, щоб викликати реакцію молекулярного азоту повітря з киснем. Стаціонарні джерела, такі як вугільні електростанції є основними джерелами цього забруднювача, хоча бензинові двигуни та інші мобільні джерела також мають велике значення.

#### 1.1.5 Озон

Озон, ключовий компонент фотохімічного смогу, утворюється внаслідок складної реакції між діоксидом азоту та вуглеводнями у присутності сонячного світла. Він вважається основним забруднювачем у тропосфері – нижньому шарі атмосфери, але не у верхніх шарах атмосфери, де він зустрічається природним чином і служить для блокування шкідливих ультрафіолетових променів від Сонця. Оскільки діоксид азоту та вуглеводні у значних кількостях викидаються автомобілями, фотохімічний зміг поширений у таких містах, як Лос-Анджелес, де багато сонячного світла та інтенсивний рух шосе. Певні географічні об'єкти, такі як гори, що перешкоджають руху повітря, та погодні умови, наприклад,

температурні інверсії в тропосфері, сприяють уловлюванню забруднювачів повітря та утворенню фотохімічного смогу [7].

#### 1.1.6 Свинець

Частинки свинцю, що вдихаються, у вигляді диму і пилу особливо шкідливі для дітей, у яких навіть незначне підвищення рівня свинцю в крові може викликати порушення навченості, судоми або навіть смерть. Джерела частинок свинцю, що переносяться по повітря, включають нафтопереробку, плавку та іншу промислову діяльність. У минулому спалювання бензину, що містить антидетонаційну присадку на основі свинцю, звану тетраетилсвинцем, було основним джерелом твердих частинок свинцю. У багатьох країнах зараз діє повна заборона використання свинцю в бензині.

#### 1.1.7 Парникові гази

Глобальне потепління визнається майже всіма дослідниками атмосфери як серйозна екологічна проблема, викликана збільшенням рівнів деяких газових домішок у атмосфері початку промислової революції у середині 18 століття. Ці гази, разом звані парниковими газами, включають діоксид вуглецю, органічні хімічні речовини, звані хлорфторвуглецями (CFC), метан, закис азоту, озон і багато інших. Великих обсягів, що викидаються повітря при спалюванні викопного палива (наприклад, бензину, нафти, вугілля) [8].

Вуглекислий газ вважається нормальним компонентом атмосфери і до промислової революції середній рівень цього газу становив близько 280 частин на мільйон (ppm). До 2020 року рівні вуглекислого газу досягли 417 частин на мільйон і продовжують зростати зі швидкістю майже 3 частин на мільйон на рік. Міжнародна співпраця та угоди, такі як Паризька угода 2015 року, будуть необхідні для скорочення викидів вуглекислого газу в усьому світі.

## 1.2 Глобальне охоплення забруднення повітря

Оскільки деякі забруднювачі повітря зберігаються в атмосфері та переносяться вітрами на великі відстані, забруднення повітря виходить за межі місцевих, регіональних та континентальних кордонів, а також може впливати на глобальний клімат та погоду. Наприклад, кислотні дощі привернули увагу всього світу з 1970-х років як регіональна та навіть континентальна проблема. Кислотні дощі виникають, коли діоксид сірки та оксиди азоту, що утворюються при спалюванні вичерпного палива, з'єднуються з водяною парою в атмосфері, утворюючи тумани сірчаної та азотної кислоти. Кислотні осадки, що утворилася, завдають шкоди водним, лісовим і ґрунтовим ресурсам. Це призвело до зникнення риби з багатьох озер у горах Адірондак у Північній Америці, повсюдної загибелі лісів у горах Європи та пошкодження зростання дерев у Сполучених Штатах та Канаді. Кислотний дощ може викликати корозію будівельних матеріалів і бути небезпечним для здоров'я людини. Ці проблеми не стримуються політичними кордонами. Викиди від спалювання вичерпного палива в середніх частинах США та Канади випадають у вигляді кислотних дощів у східних регіонах цих країн та кислотних дощів у Норвегії. Викиди від спалювання вичерпного палива в середніх районах Сполучених Штатів і Канади випадають у вигляді кислотних дощів у східних регіонах цих країн, а кислотні дощі в Норвегії надходять переважно з промислових районів Великобританії та континентальної Європи. Міжнародний масштаб проблеми призвів до підписання міжнародних угод про обмеження викидів сірки та оксидів азоту [9].

Ще одна глобальна проблема, викликана забрудненням повітря, – це виснаження озонового шару в стратосфері. На рівні землі (тобто в тропосфері) озон є забруднювачем, але на висоті понад 12 км (7 миль) він грає вирішальну роль у поглинанні та блокуванні ультрафіолетового випромінювання (УФ) від Сонця до того, як воно досягне землі. Вплив УФ-випромінювання пов'язують із раком шкіри та іншими проблемами зі здоров'ям.

### 1.3 Джерела даних про забруднення навколишнього повітря

#### 1.3.1 Станіонарні станції екологічного моніторингу.

На даний момент у світі більш високоякісні дані про забруднення повітря надходять від мереж автоматичних станіонарних (рис. 1.1) [10] та мобільних станцій контролю забруднення атмосферного повітря. Вони втілюють безперервні виміри концентрацій забруднювачів і є ґрунтом надання даних у завданнях зведення моделей прогнозування і карт високого просторово-часового дозволу.



Рисунок 1.1 – Станіонарна станція моніторингу повітря



В даний час у світі є більше 30 000 популярних станцій прогнозу якості повітря. Складання станції моніторингу залежить як від щільності населення, наприклад, і від обсягу забудованої землі, і має гарантувати репрезентативність, безперервність, захищеність і функціональність [11].

Класичні станції прогнозу можуть вимірювати якість повітря тільки в межах 1–3 км, а тимчасова перерва вимірювання може займати від декількох хвилин до дня і ночі. Проте якість повітря в місті змінюється нелінійно в залежності від області, наприклад що він не може бути правильно помічений обмеженою кількістю станцій прогнозу [12]. Ще одним обмеженням цього методу є те, що детектори мають короткий термін служби і вимагають постійного технічного обслуговування. Тому застосування цих датчиків – дуже дороге рішення для непроникного розгортання і охоплення будь-якого куточка міста. Тобто, люди, що живуть і працюють далеко від станцій прогнозу, не можуть отримувати чіткі доповіді про якість повітря в реальному часі.

Впровадження класичних станцій прогнозу повітря для отримання своєчасних даних вважаються важливою, але не необхідною умовою для чіткого картографування забруднення навколишнього повітря.

### 1.3.2 Бездротові сенсорні мережі

Використання недорогих мініатюрних датчиків (рис. 1.2) гарантує значне розширення просторового і тимчасового дозволу отриманої інформації. Тому в останні роки активно розвивається система прогнозу забруднення повітря, що об'єднує дешеві мініатюрні вимірювальні детектори в бездротову сенсорну мережу (WSN). Якість одержуваних даних незрівнянно нижче проти стаціонарних станцій, хоча мініатюрні сенсори забезпечують мобільність і можливість великомасштабного розгортання вузлів бездротової мережі WSN, дозволяючи отримувати дані у час.



Рисунок 1.2 – Мініатюрна станція моніторингу повітря

Такі системи допомагають дослідникам більш ефективно та точно будувати розподіл рівня забруднення повітря, і, як наслідок, моделі оцінки та прогнозування якості повітря.

### 1.3.3 Супутникові дані

Останнім часом проводиться функціональна робота зі створення інтеграції існуючих наземних сіток датчиків з іншими методами збору даних, у тому числі з допомогою супутників [13].

Існує різноманітність моделей для визначення певного забруднюючого препарату за допомогою даних, одержуваних за допомогою супутників, але однією з провідних моделей для інтерпретації того, що вимірюється супутниковим пристроєм є модель радіаційного перенесення (тобто, поширення електричної енергії через атмосферу Землі). Модель передбачає поглинання, випромінювання та розсіювання світла хмарами, земною поверхнею, аерозолями та всіма газами, при проходженні крізь атмосферу Землі до супутника.

Метод прямого розрахунку line-by-line є стандартним методом обчислення функції пропускання і дозволяє з високою точністю проводити розрахунки атмосферних радіаційних даних. Даний спосіб передбачає обчислення коефіцієнта поглинання на частоті  $\nu$  підсумовуванням вкладів від усіх рядів, що розглядаються, в спектральному спектрі.

Втім, супутникові прилади можуть давати інформацію тільки в світлу погоду і з низькою тимчасовою роздільною здатністю. Ще супутникові прилади можуть дати інформацію не про всі забруднюючі частинки. Наприклад, вони не визначають PM2.5, але цей розмір корелюється з аерозольною оптичною глибиною (AOD) [14].

Дані, що отримуються за допомогою дистанційного зондування Землі, можуть працювати додатковим джерелом, доповнюючи дані, зібрані наземними вимірювальними станціями.

Однією з можливих методик застосування цих даних є тест епізодів транскордонного перенесення, наприклад, перенесення жорстких частинок при пильних бурях в Африці на південь Європи.

#### 1.4 Сучасний стан моніторингу у місті Вінниця

Основні проблеми з моніторингом якості атмосферного повітря України пов'язані з декількома факторами:

- відсутністю систематичного фінансування;
- відсутністю стратегії розвитку моніторингових спроможностей;
- відсутністю комунікаційної стратегії.

Натомість різні громадські організації створюють проекти до яких можливо долучитись або ж зробити свій внесок у проект.

Наприклад у місті Вінниця є такі проекти, як SaveDnipro, Eco City, Український гідрометеорологічний центр та SaveEcoBot. На рисунку 1.3 можемо бачити сайт проекту SaveEcoBot.

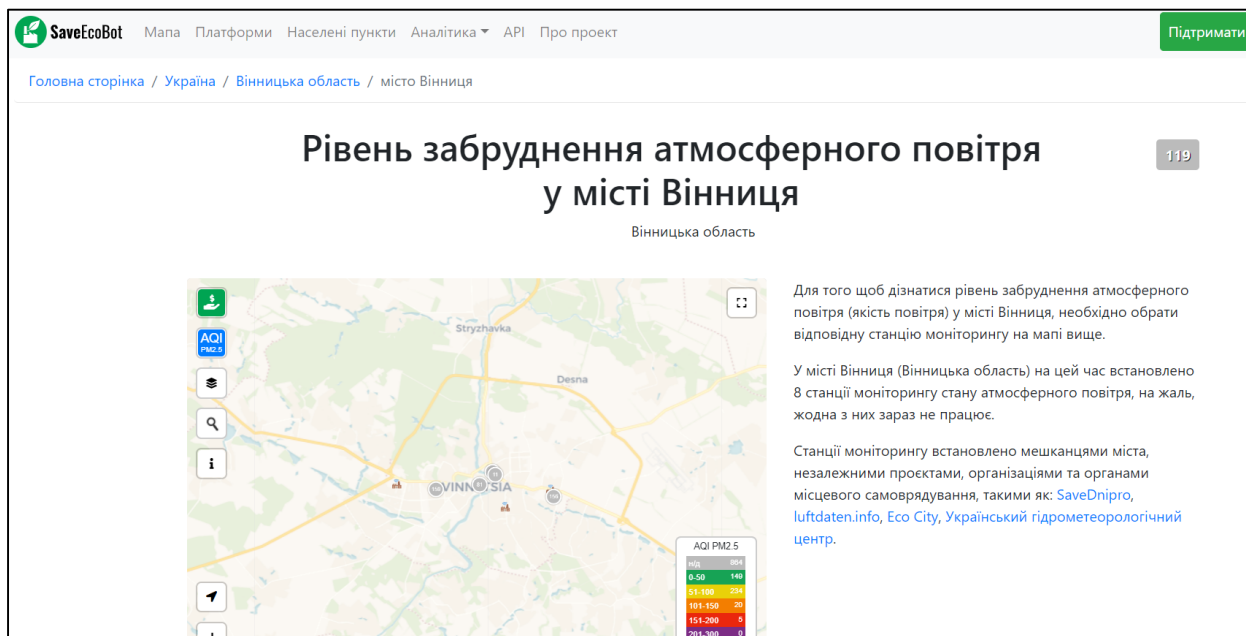


Рисунок 1.3 – Сайт проекту моніторингу якості повітря SaveEcoBot

Станції даного проекту було становлено в таких містах, як Вінниця, Хмельник та Ладижин. Станції встановлюють мешканці міста та органи місцевого самоврядування [15].

Також можна відзначити проєкт Eco City, на рисунку 1.4 зображена головна сторінка сайту, де власне відображається одразу ж карта моніторингу якості повітря з постами по місту та області та показники вимірювання за певними ознаками.

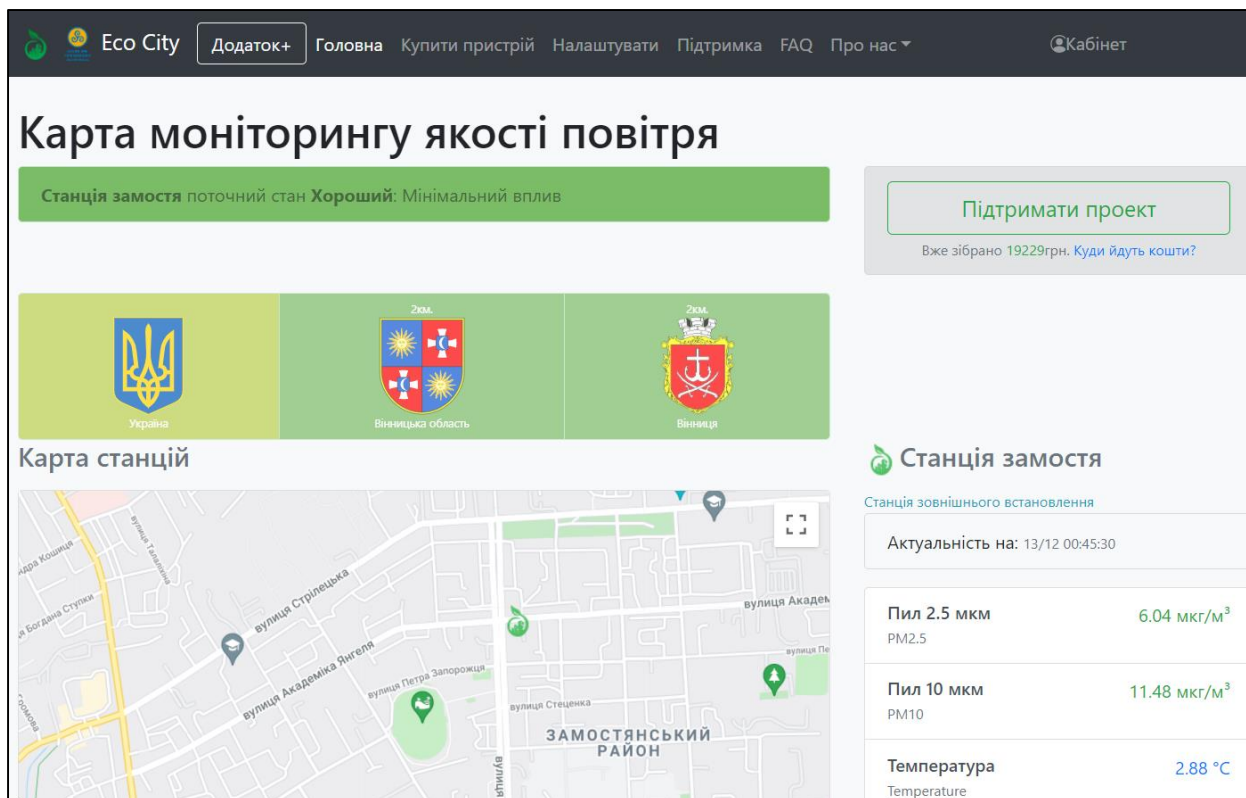


Рисунок 1.4 – Головна сторінка сайту проекту Eco City

Громадський моніторинг якості повітря України EcoCity є соціальним, екологічним, незалежним проектом, який реалізовується неприбутковою громадською організацією "Фрі Ардуіно".

Дані постійно оновлюються та розміщені на головній сторінці того чи іншого проекту, також дані є у вільному доступі, тобто їх можна завантажити або ж оновлювати на сайті.

На рисунку 1.5 зображена таблиця AQI – індексу якості повітря.

AQI	Категорія	Наслідки для здоров'я
0-50	<b>Хороший</b>	Мінімальний вплив
51-100	<b>Задовільний</b>	Може викликати незначний дискомфорт при диханні у чутливих людей. Не рекомендується займатись спортом на вулиці.
101-150	<b>Помірно забруднений</b>	Може спричинити дискомфорт при диханні у людей із захворюваннями дихальних шляхів, таких як астма, а також у людей з серцевими захворюваннями. Діти та літні люди зазнають найбільшого негативного впливу. Заняття спортом на вулиці заборонені.
151-200	<b>Високий</b>	Спричиняє дискомфорт при диханні у здорових людей. Людям із серцево-судинними і легеневиими захворюваннями категорично заборонено перебувати на вулиці. Заняття спортом на вулиці заборонені.
201-300	<b>Дуже високий</b>	Може викликати респіраторні захворювання у здорових людей при тривалому впливі. Людям із серцево-судинними і легеневиими захворюваннями категорично заборонено перебувати на вулиці. При такому забрудненні існує ризик інсульту та інфаркту у людей похилого віку. Заняття спортом на вулиці заборонені.
300+	<b>Небезпечний</b>	Впливає навіть на здорових людей, і може спричинити серйозні наслідки для їх здоров'я. Людям із серцево-судинними захворюваннями категорично заборонено перебувати на вулиці. Негативні наслідки можуть виникнути навіть під час легкої фізичної активності. Існує ризик інсульту та інфаркту.

Рисунок 1.5 – Таблиця AQI

Наразі у Вінниці розміщено 9 постів спостереження моніторингу якості атмосферного повітря.

## 1.5 Висновки

У даному розділі було визначено суть і види забруднень атмосферного повітря, вказано, що таке пости спостереження та які вони бувають. Також, розглянуто сучасний стан моніторингу атмосферного повітря у місті Вінниця, визначено основні проблеми та проекти, що займаються встановленням постів по місту.

## 2 ВИБІР ОПТИМАЛЬНИХ НАЛАШТУВАНЬ ТА ОГЛЯД НАБОРУ ДАНИХ

### 2.1 Опис FbProphet

Prophet – це процедура прогнозування даних часових рядів, заснована на адитивній моделі, в якій нелінійні тренди відповідають річній, щотижневої і щоденної сезонності, а також впливу свят. Це найкраще працює з тимчасовими рядами, які мають сильні сезонні ефекти і кілька сезонів історичних даних. Prophet стійкий до відсутніх даних і змін тренду, і зазвичай добре справляється з викидами.

Facebook розробив Prophet з відкритим вихідним кодом, інструмент прогнозування, доступний як на Python, так і R. Він надає інтуїтивно зрозумілі параметри, які легко налаштувати. Навіть той, кому не вистачає глибоких знань у моделях прогнозування часових рядів, може використати це для створення значущих прогнозів для різних проблем у бізнес-сценаріях.

Основні особливості Facebook Prophet:

- Дуже швидко, оскільки він побудований на Stan, мовою програмування для статистичного виводу, написаною на C++.
- Модель адитивної регресії, у якій нелінійні тренди відповідають річній, щотижневій та денній сезонності, а також святковим ефектам:
  - 1) Шматково-лінійний або логістичний тренд кривої росту. Prophet автоматично виявляє зміни у тенденціях, вибираючи точки зміни даних.
  - 2) Річна сезонна складова змодельована з використанням ряду Фур'є.
  - 3) Щотижнева сезонна складова з використанням фіктивних змінних.
  - 4) Список важливих свят, що надається користувачем [16].
- Стійкий до відсутності даних та зсувів у тренді та зазвичай обробляє викиди.

Prophet використовує модель розкладених часових рядів із трьома основними компонентами моделі: тренд, сезонність та свята. Вони об'єднані в наступне рівняння:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t, \quad (2.1)$$

де  $g(t)$  – шматково-лінійна або логістична крива росту для моделювання неперіодичних змін часових рядів;

$s(t)$  – періодичні зміни (наприклад, щотижнева/річна сезонність);

$h(t)$  – вплив свят (надається користувачем) з нерегулярним графіком;

$\varepsilon_t$  – термін помилки враховує будь-які незвичайні зміни, не враховані моделлю.

Використовуючи час в якості регресора, Prophet намагається вмістити кілька лінійних і нелінійних функцій часу як компоненти. Prophet формулює проблему прогнозування як вправу з припасування кривої, а не розглядає явно тимчасову залежність кожного спостереження в часовому ряду [17].

Проста процедура настроювання та коригування прогнозів при додаванні знань у предметній галузі або бізнес-ідей. Як приклад розглянемо моделі Prophet взяті з офіційного сайту.

На рисунку 2.1 зображено прогноз щоденних переглядів сторінок журналу для сторінок Wikipedia.

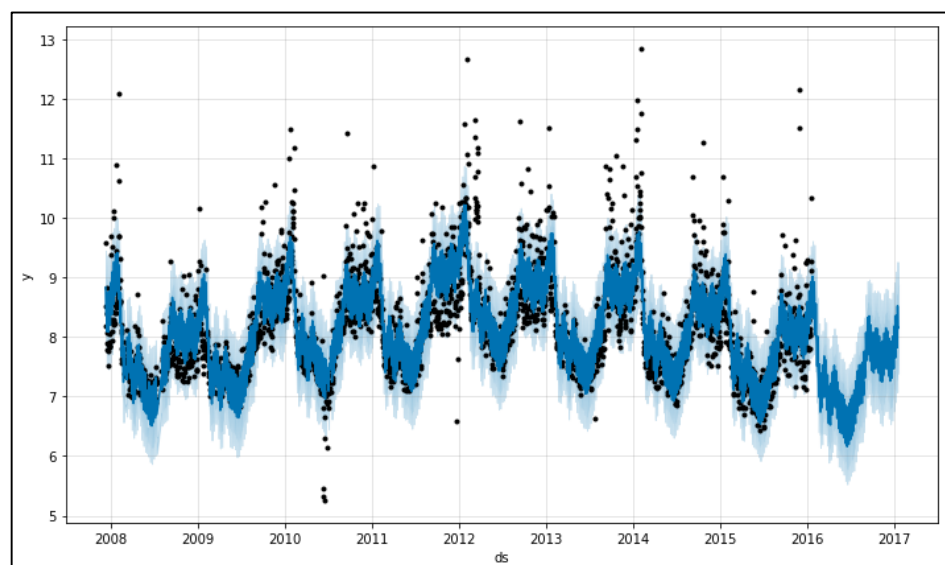


Рисунок 2.1 – Прогноз щоденних переглядів



Також цей графік можна розбити на компоненти, якщо використовувати метод `Prophet.plot_components`. Тоді ми отримаємо наступні графіки, що зображені на рисунку 2.2.

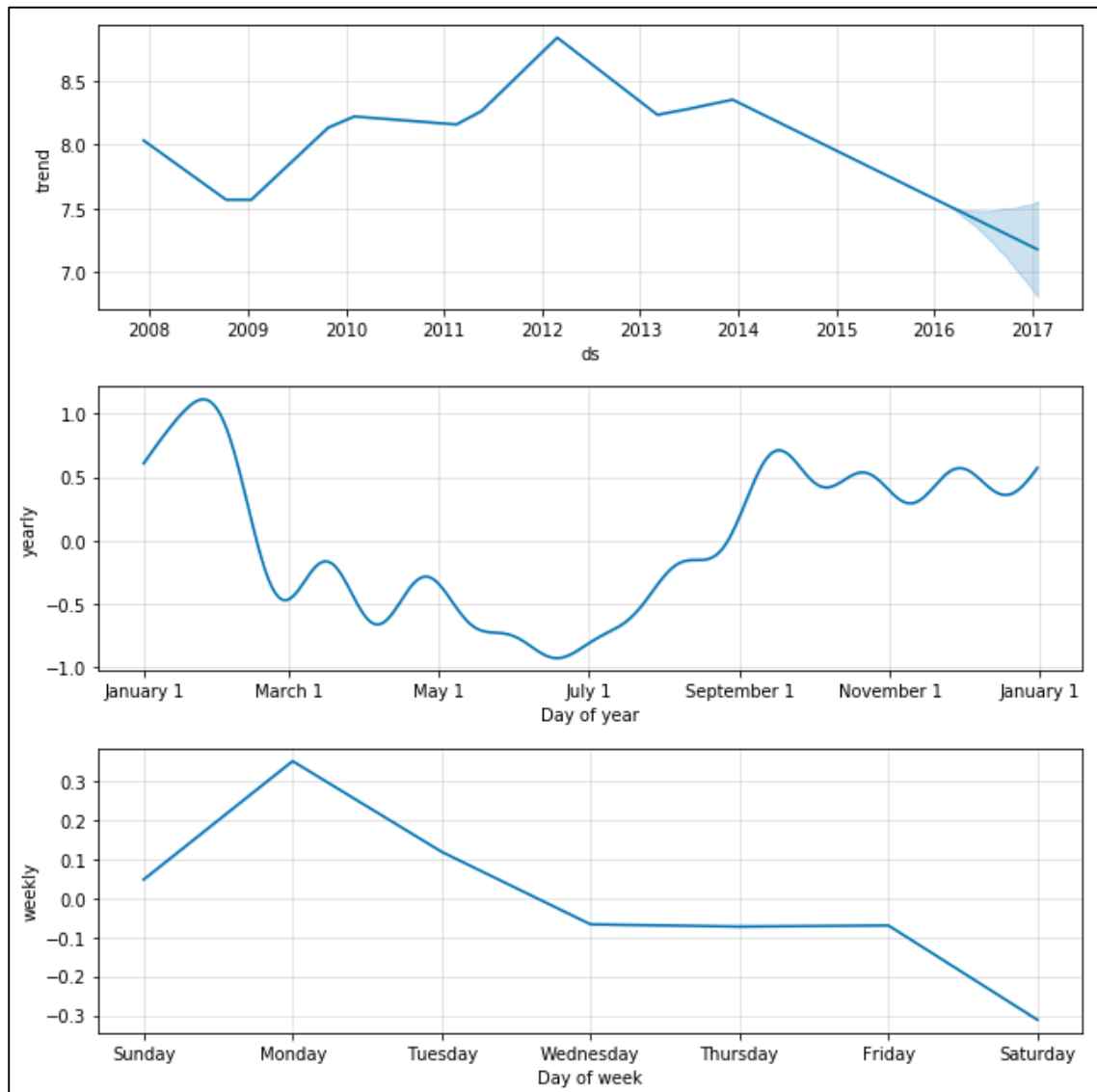


Рисунок 2.2 – Компоненти прогнозу

Також в `Prophet` можна виконувати прогноз на тривалий період часу, використовуючи щомісячні дані (рис 2.3). Однак можливі невпевненості в сезонності, що відображено на рисунку 2.4.

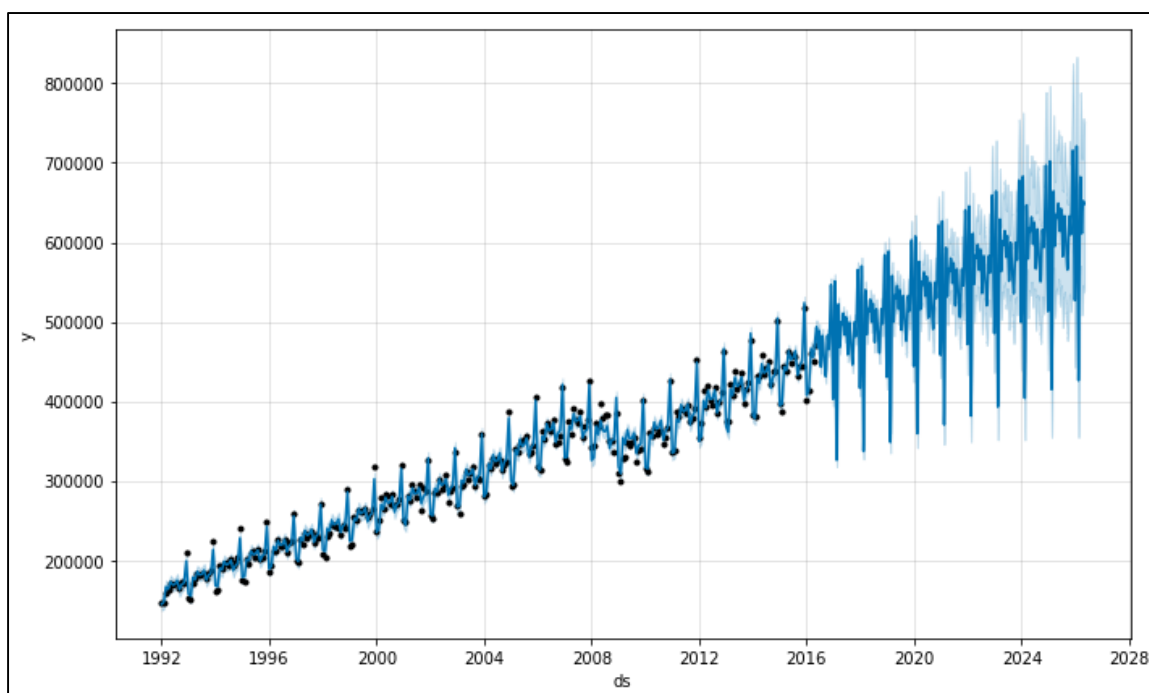


Рисунок 2.3 – Прогноз з використанням щомісячних даних

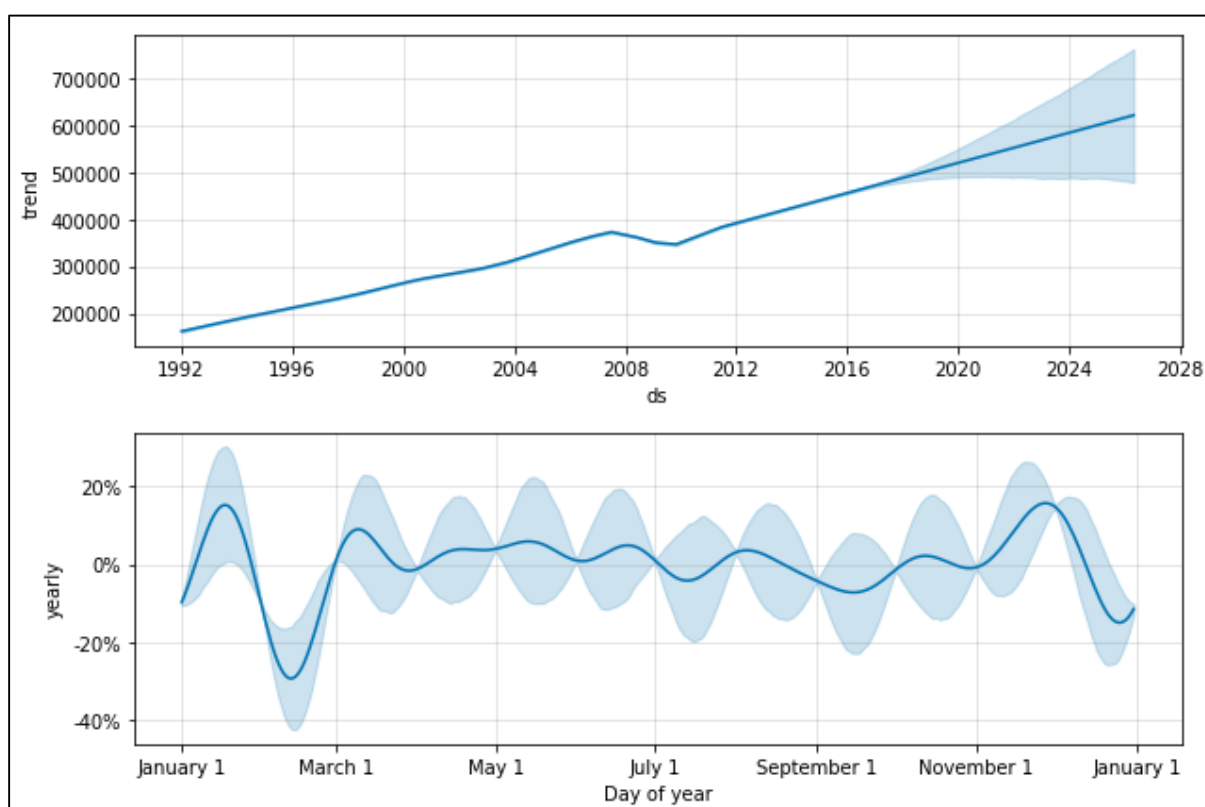


Рисунок 2.4 – Сезонна невпевненість

При використанні даних за день щоденна сезонність буде автоматично

відповідати їм, що ми можемо побачити на рисунку 2.5. Компоненти щоденної сезонності будуть відображатися на рисунку 2.6.

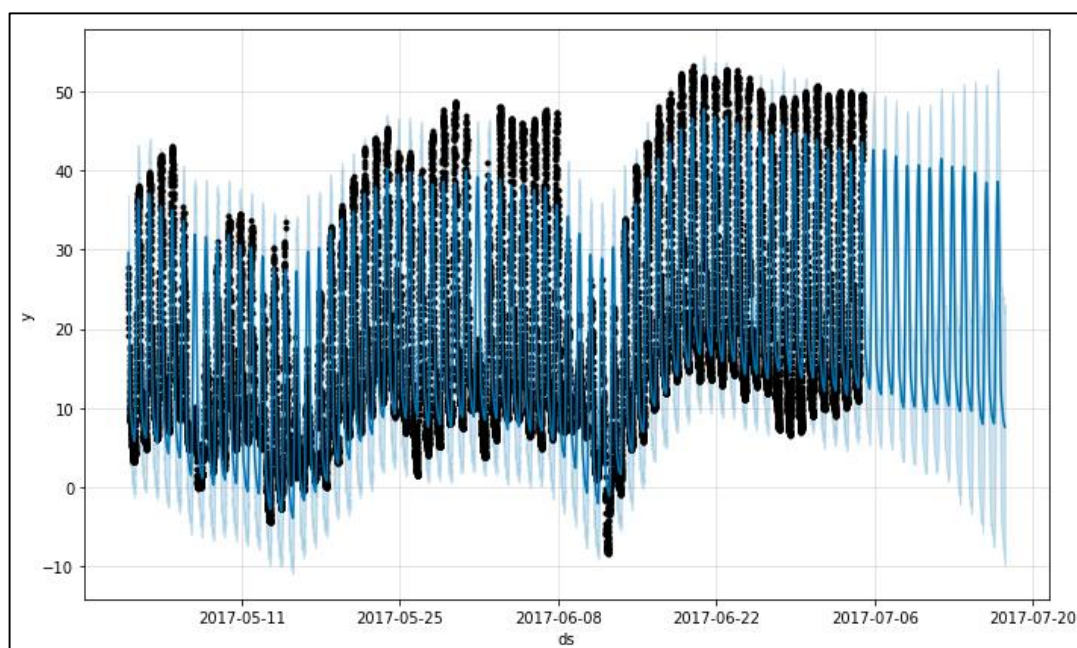


Рисунок 2.5 – Прогноз з використанням щоденних даних

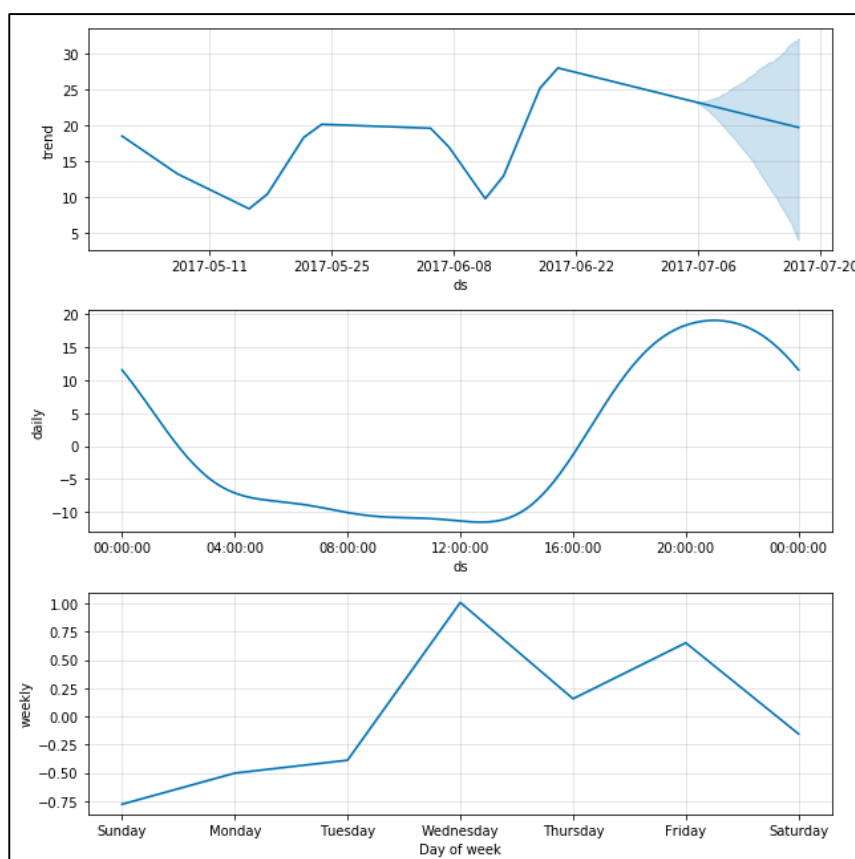


Рисунок 2.6 – Компоненти щоденної сезонності

Окрім цього Prophet зручно використовувати ще й тому, що він може аналізувати дані і на їх основі заповнювати відсутні дані. Це можна побачити на рисунку 2.7

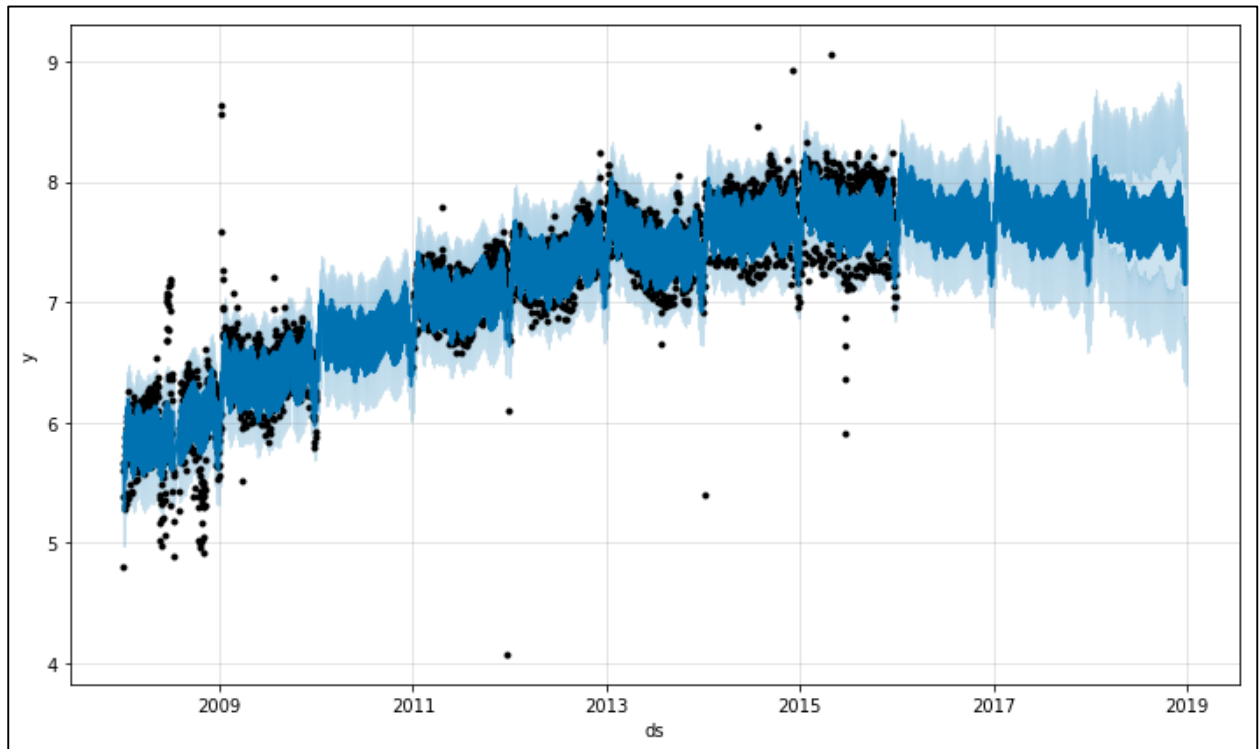


Рисунок 2.7 – Прогнозування з частково відсутніми даними

Prophet також може робити аналіз даних на основі вихідних і свят (рис 2.8). Для цього можна використовувати як уже створену в самому Prophet бібліотеку, так і самому створити їх [18].

Додаткові регресори можна додати до лінійної частини моделі з використанням `add_regressor` метода або функції. На графіку компонентів цей ефект буде відображатися на графіку `extra_regressors` (рис. 2.9).

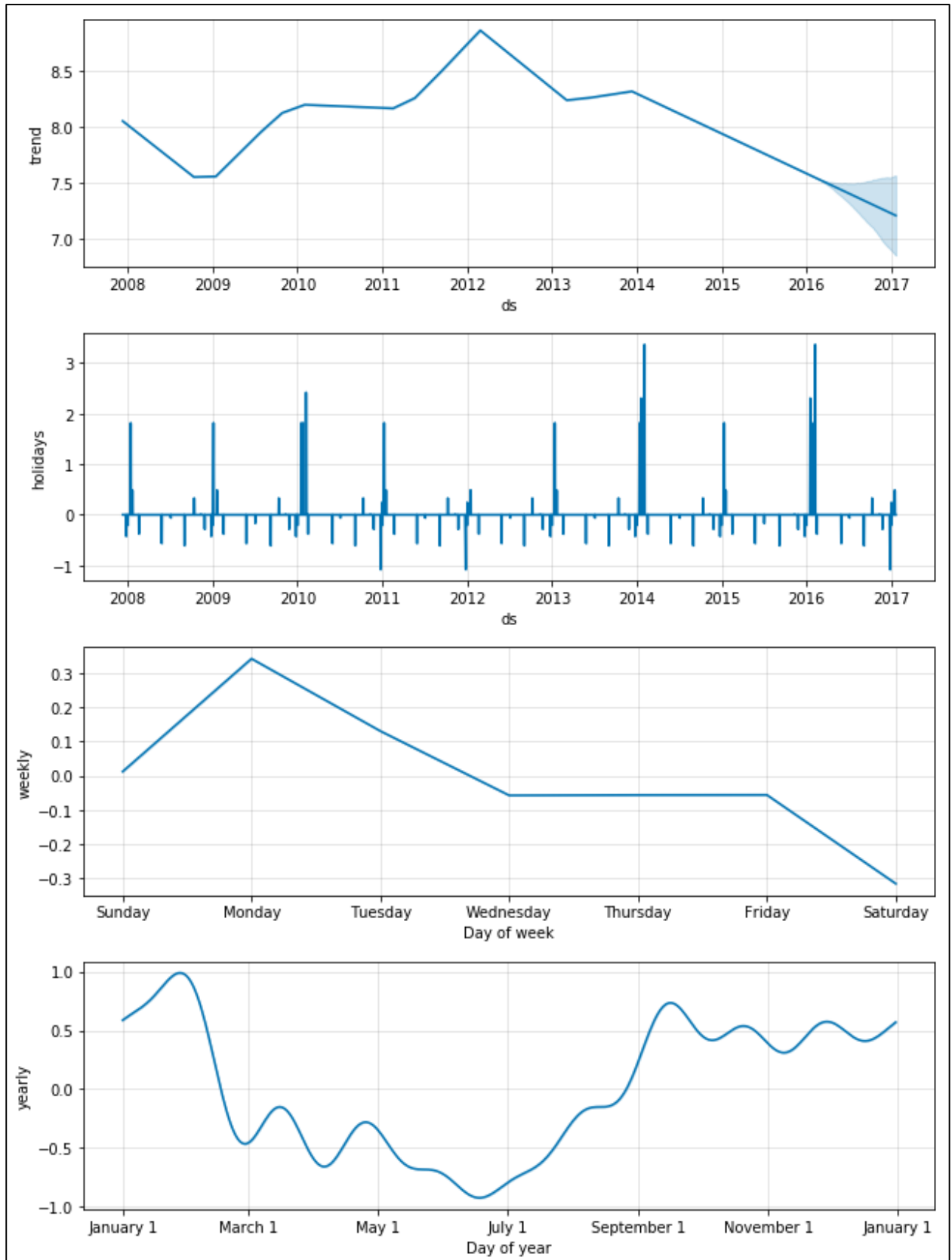


Рисунок 2.8 – Використання вихідних та свят

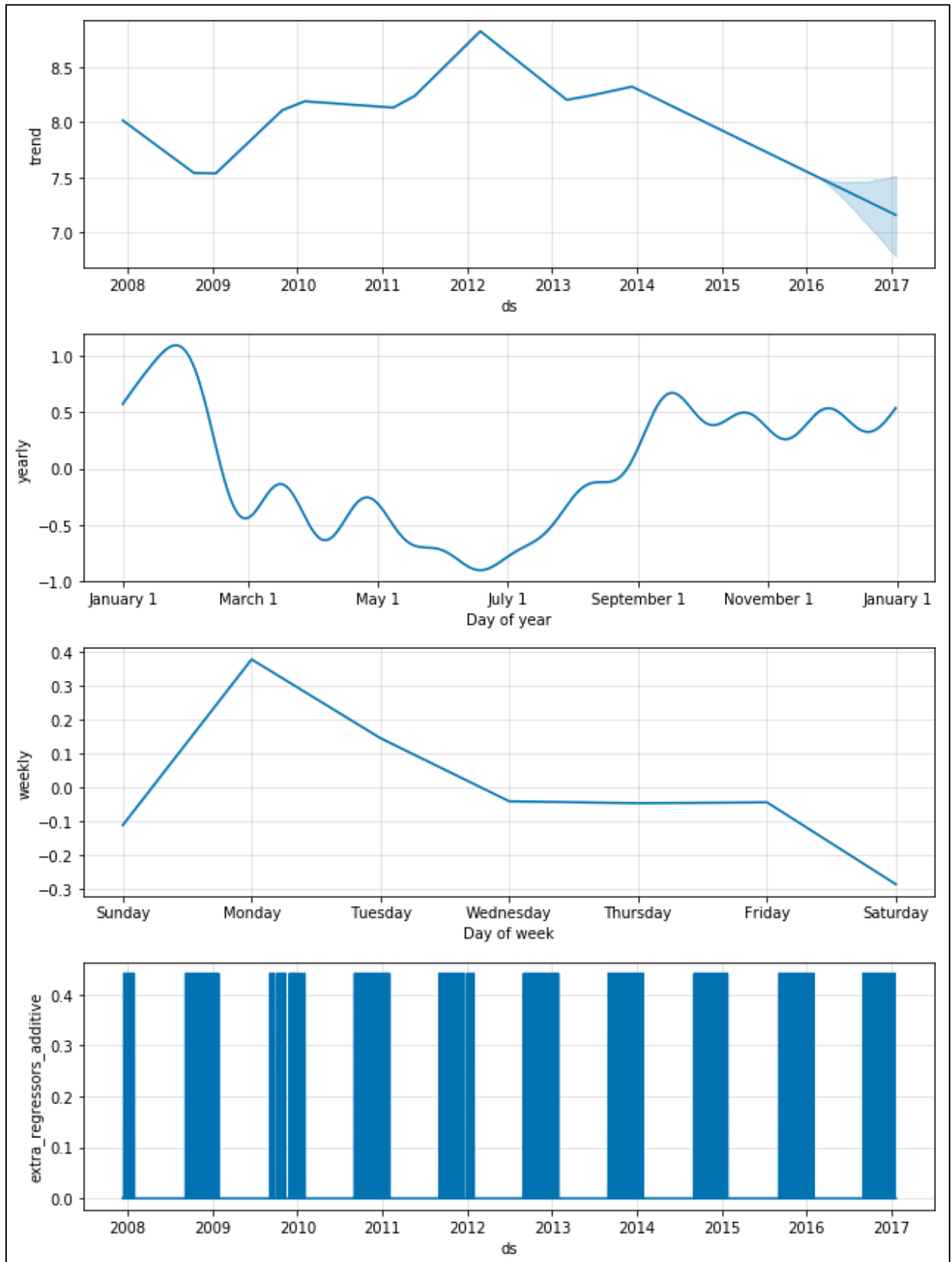


Рисунок 2.9 – Додаткові регресори

На рисунках 2.10 – 2.16 зображено ще декілька прикладів моделей Prophet.

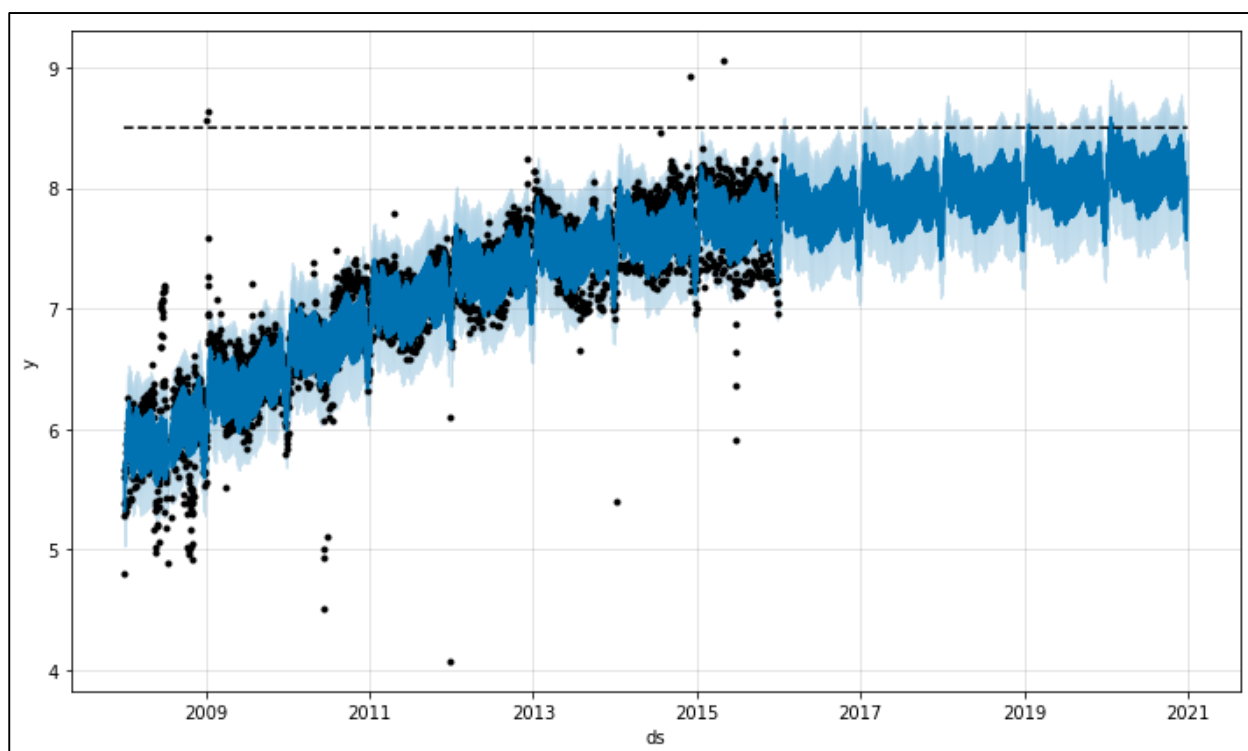


Рисунок 2.10 – Прогноз росту

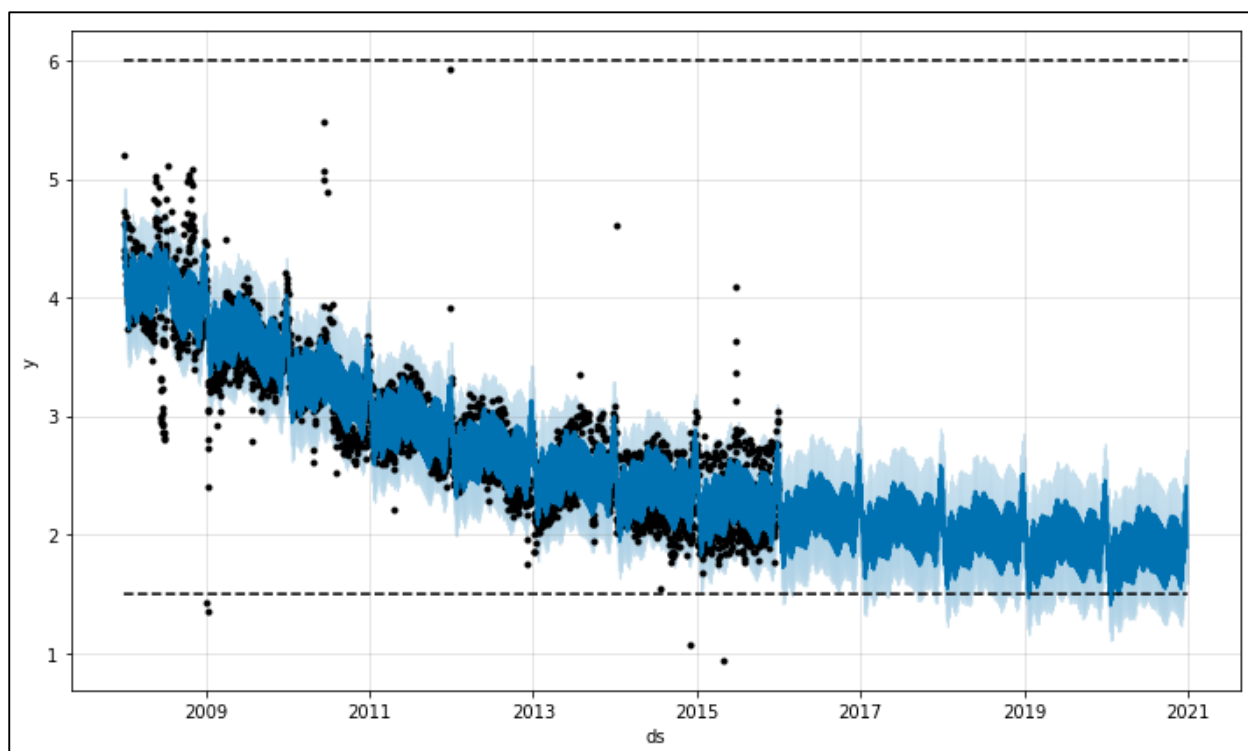


Рисунок 2.11 – Насичений мінімум

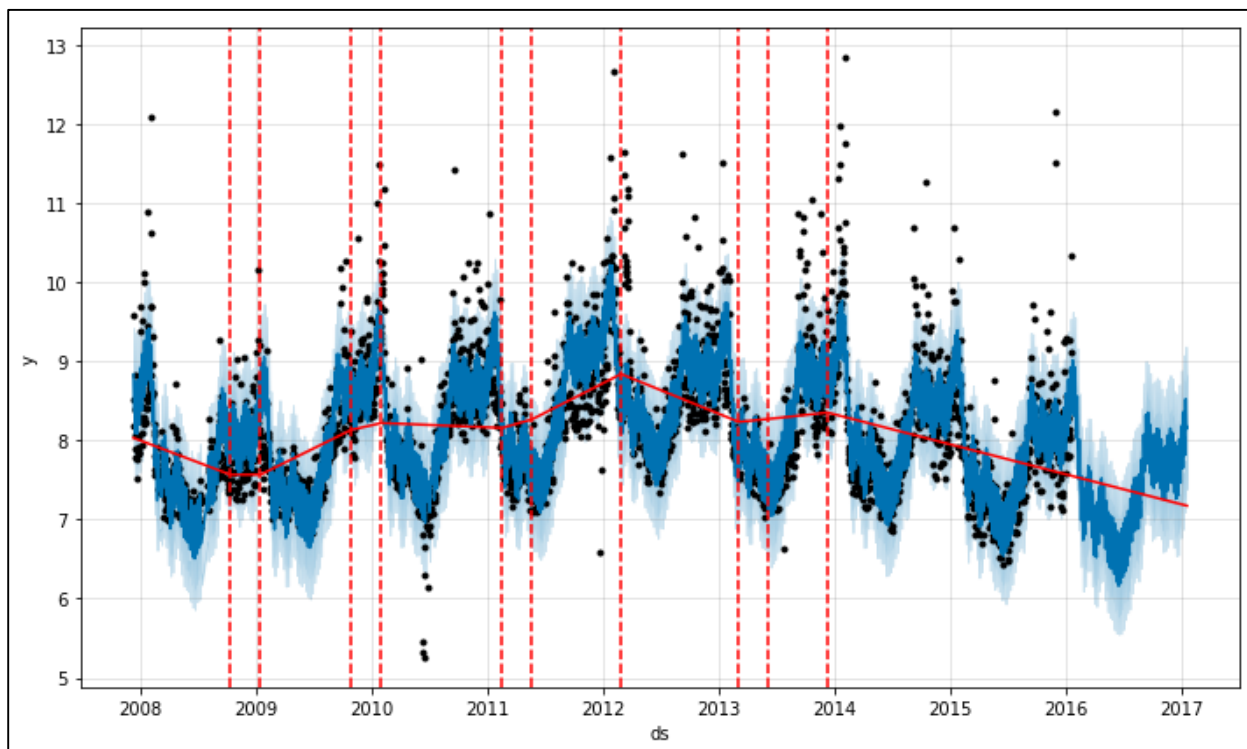


Рисунок 2.12 – Візуалізація точок зміни тренду

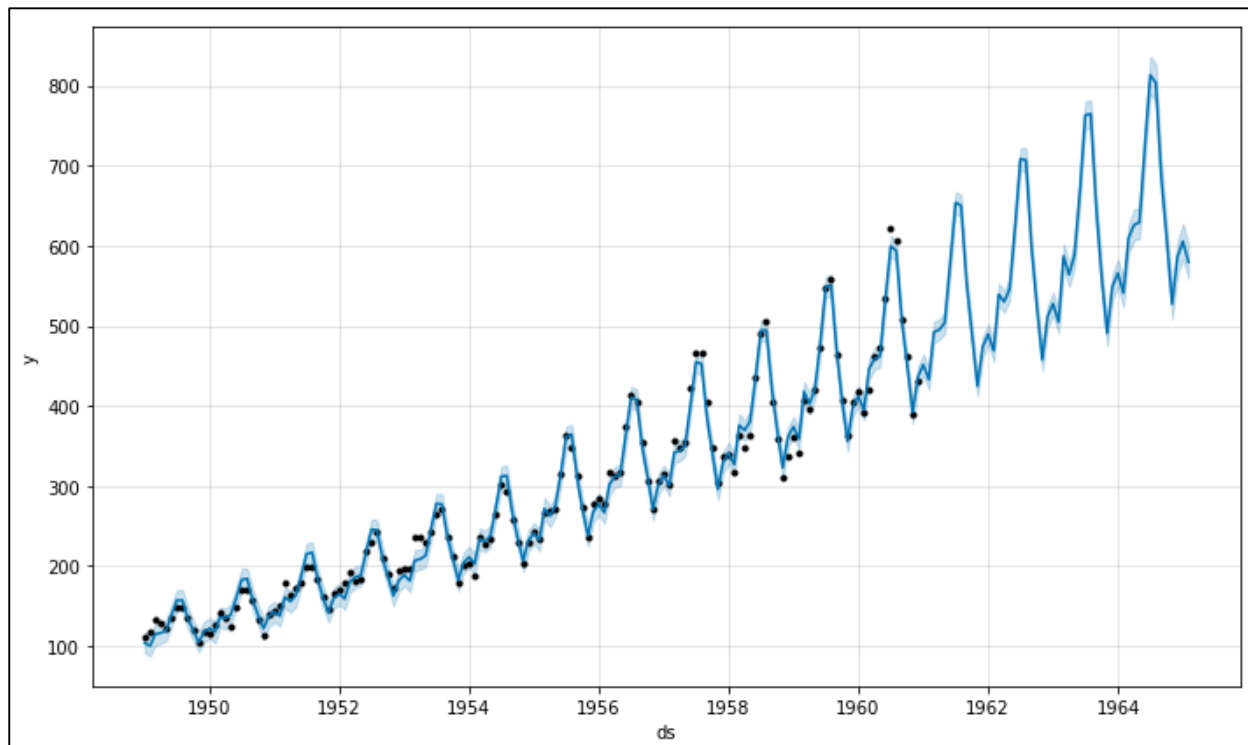


Рисунок 2.13 – Мультиплікативна сезонність



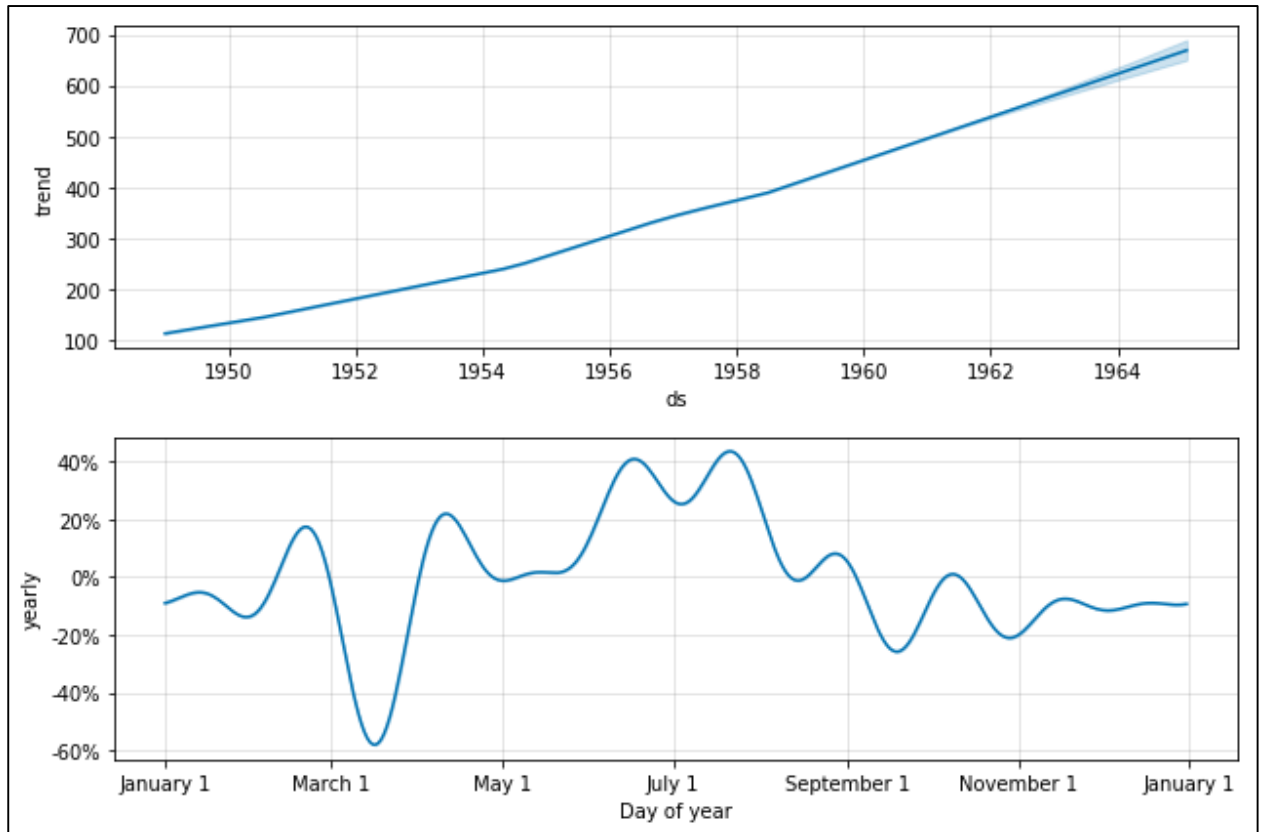


Рисунок 2.14 –Сезонність в відсотках від тенденції

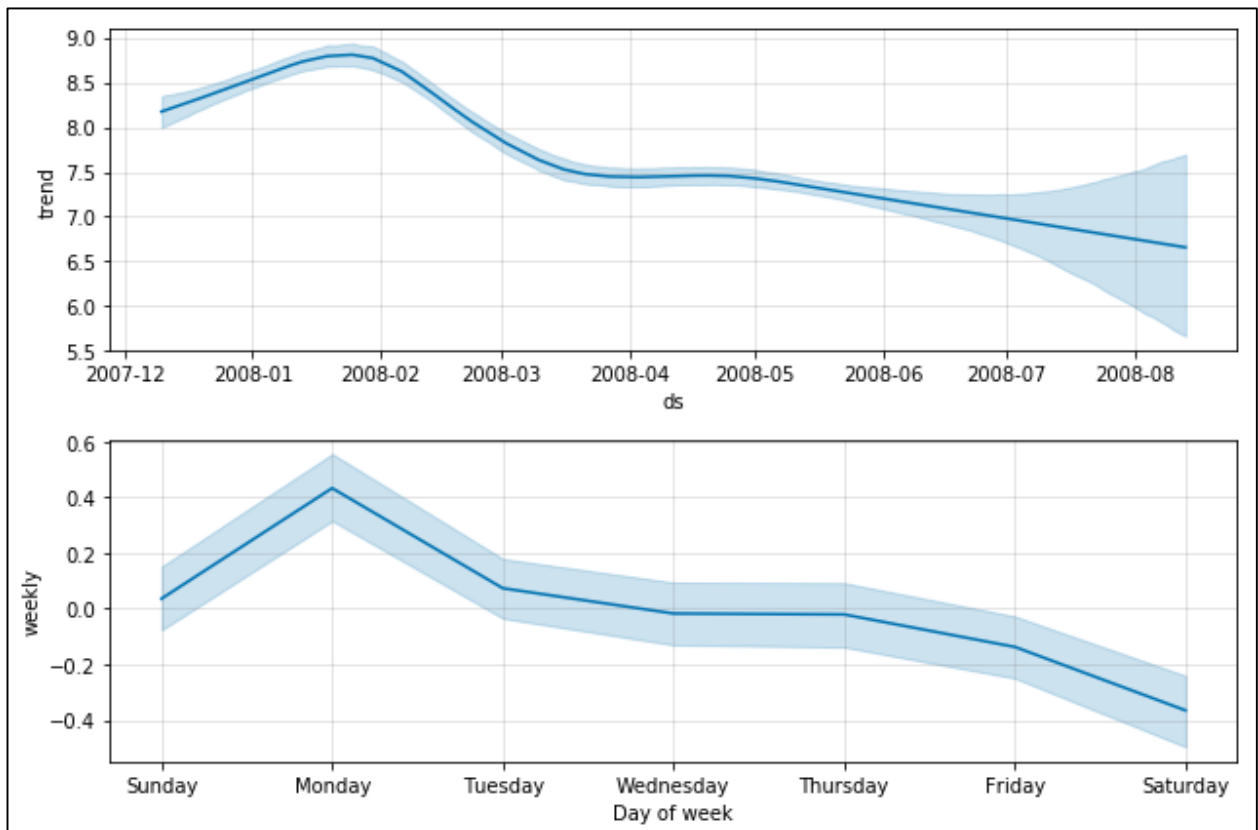


Рисунок 2.15 – Невизначеність в тренді і сезонності

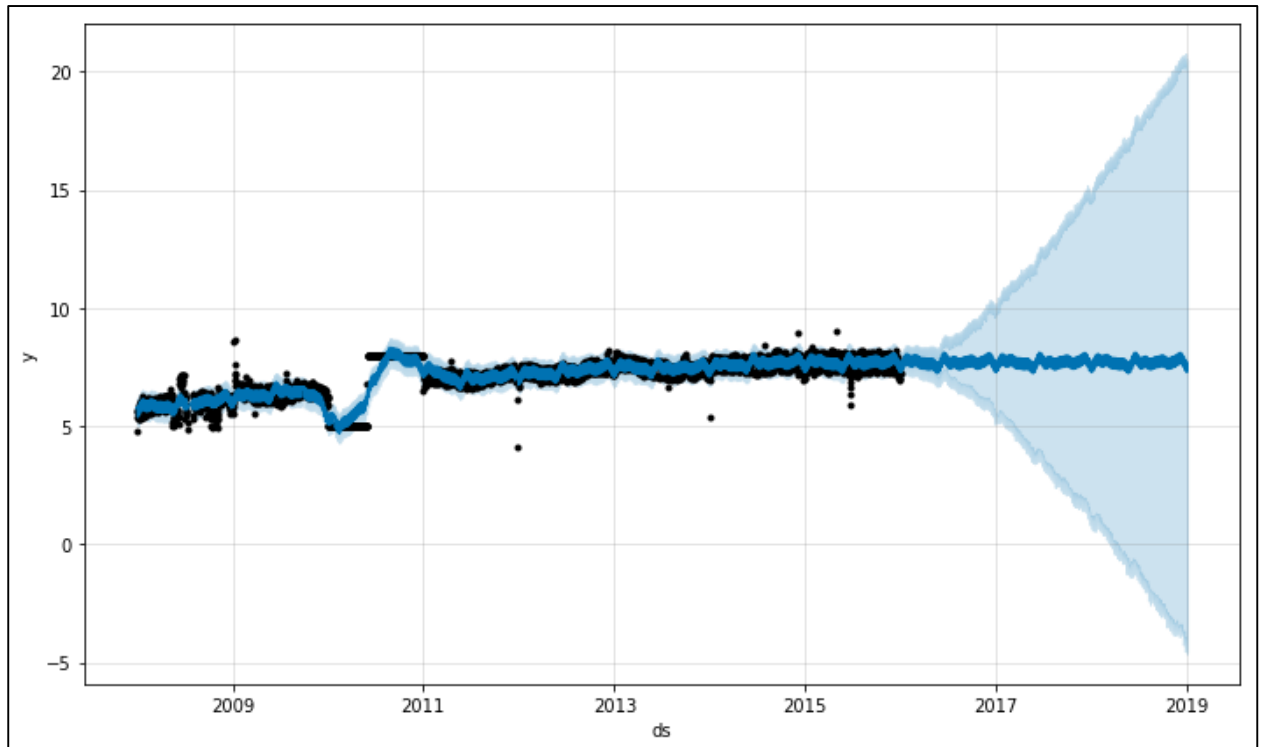


Рисунок 2.16 – Прогноз з блоком невірних даних

## 2.2 Вибір мови програмування

Для роботи з Prophet використовуються дві мови програмування: R і Python. Тому для подальшої роботи необхідно обрати одну з них.

R та Python є мовами програмування з відкритим вихідним кодом і мають велику спільноту. Нові бібліотеки або інструменти постійно додаються до відповідних каталогів. R в основному використовується для статистичного аналізу, в той час як Python забезпечує більш загальний підхід до науки даних.

R та Python – це сучасні мови програмування, орієнтовані на науку про дані. Зрозуміло, вивчення їх обох – ідеальне рішення. R та Python вимагають витрат часу, а така розкіш доступна не кожному. Python – це мова загального призначення з синтаксисом, що читається. Проте R створюється статистиками і охоплює їхню специфічну мову [19].

Python – це об'єктно-орієнтована мова програмування загального призначення, яка підкреслює зручність читання коду за рахунок щедрого використання прогалин. Випущений у 1989 році, Python простий у освоєнні та

любимо програмістами та розробниками. Фактично, Python – одна з найпопулярніших мов програмування у світі, відразу після Java та C.

Декілька бібліотек Python підтримують завдання науки про дані, у тому числі такі:

- NumPy для обробки великих розмірних масивів;
- Панди для обробки та аналізу даних;
- Matplotlib для побудови візуалізацій даних.

Крім того, Python особливо добре підходить для великомасштабного розгортання машинного навчання. Його набір спеціалізованих бібліотек для глибокого навчання та машинного навчання включає такі інструменти, як scikit-learn, Keras та TensorFlow, які дозволяють фахівцям за даними розробляти складні моделі даних, що підключаються безпосередньо до виробничої системи. Jupyter Notebooks – це веб-додаток з відкритим вихідним кодом для простого обміну документами, які містять ваш живий код Python, рівняння, візуалізації та пояснення науки про дані.

R – це мова програмування з відкритим вихідним кодом, оптимізована для статистичного аналізу та візуалізації даних. Розроблений у 1992 році, R має багату екосистему зі складними моделями даних та елегантними інструментами для створення звітів. За останніми підрахунками, більш ніж 13000 пакетів R були доступні через Comprehensive R Archive Network (CRAN) для глибокої аналітики [20].

Популярний серед науковців та дослідників у галузі науки про дані, R надає широкий спектр бібліотек та інструментів для наступного:

- очищення та підготовка даних;
- створення візуалізацій;
- навчання та оцінка алгоритмів машинного навчання та глибокого навчання.

R зазвичай використовується в RStudio, інтегрованому середовищі розробки (IDE) для спрощеного статистичного аналізу, візуалізації та звітності.

Програми R можна використовувати безпосередньо та в інтерактивному режимі в Інтернеті через Shiny.

Основна відмінність між двома мовами полягає в їхньому підході до науки про дані. Обидві мови програмування з відкритим вихідним кодом підтримуються великими спільнотами, які постійно розширюють свої бібліотеки та інструменти. Але в той час, як R в основному використовується для статистичного аналізу, Python забезпечує більш загальний підхід до обробки даних.

Python – це багатоцільова мова, дуже схожа на C++ і Java, з синтаксисом, який легко вивчити. Програмісти використовують Python для аналізу даних або використовують машинне навчання у масштабованих виробничих середовищах. Наприклад, ви можете використовувати Python для вбудовування розпізнавання облич у свій мобільний API або для розробки програми машинного навчання.

R, з іншого боку, створений статистиками та багато в чому спирається на статистичні моделі та спеціалізовану аналітику. Фахівці з обробки даних використовують R для глибокого статистичного аналізу, що підтримується лише кількома рядками коду та гарною візуалізацією даних. Наприклад, можна використовувати R для аналізу поведінки клієнтів або дослідження геноміки.

Інші ключові відмінності:

- Збір даних: Python підтримує всі види форматів даних, від файлів з роздільниками-комами (CSV) до JSON з Інтернету. Ви також можете імпортувати таблиці SQL прямо в код Python. Для веб-розробки бібліотека запитів Python дозволяє легко виймати дані з Інтернету для створення наборів даних. На відміну від цього, R призначено для аналітиків даних, щоб імпортувати дані з Excel, CSV і текстових файлів. Файли, створені в Minitab або у форматі SPSS, також можуть бути перетворені на кадри даних R. Хоча Python більш універсальний для вилучення даних з Інтернету, сучасні пакети R, такі як Rvest, призначені для базового пошуку в Інтернеті.

– Дослідження даних: у Python ви можете досліджувати дані за допомогою Pandas, бібліотеки аналізу даних для Python. Ви можете фільтрувати, сортувати та відображати дані за лічені секунди. R, з іншого боку, оптимізований для статистичного аналізу великих наборів даних та пропонує ряд різних варіантів дослідження даних. За допомогою R ви можете будувати розподіл ймовірностей, застосовувати різні статистичні тести та використовувати стандартні методи машинного навчання та інтелектуального аналізу даних.

– Моделювання даних: Python має стандартні бібліотеки для моделювання даних, включаючи NumPy для аналізу чисельного моделювання, SciPy для наукових обчислень та обчислень та scikit-learn для алгоритмів машинного навчання. Для конкретного аналізу моделювання R вам іноді доведеться покладатися на пакети, що виходять за рамки основних функцій R. Але спеціальний набір пакетів, відомий як Tidyverse, спрощує імпорт, обробку, візуалізацію і створення звітів за даними.

– Візуалізація даних: Хоча візуалізація не є сильною стороною Python, ви можете використати бібліотеку Matplotlib для створення основних графіків та діаграм. Крім того, бібліотека Seaborn дозволяє малювати на Python привабливіші та інформативніші статистичні графіки. Однак R був створений для демонстрації результатів статистичного аналізу, а базовий графічний модуль дозволяє легко створювати базові діаграми та графіки. Ви також можете використовувати ggplot2 для складніших графіків, таких як складні графіки розсіювання з лініями регресії [21].

Також на рисунку 2.17 можна побачити порівняння популярності Python і R серед програмістів.

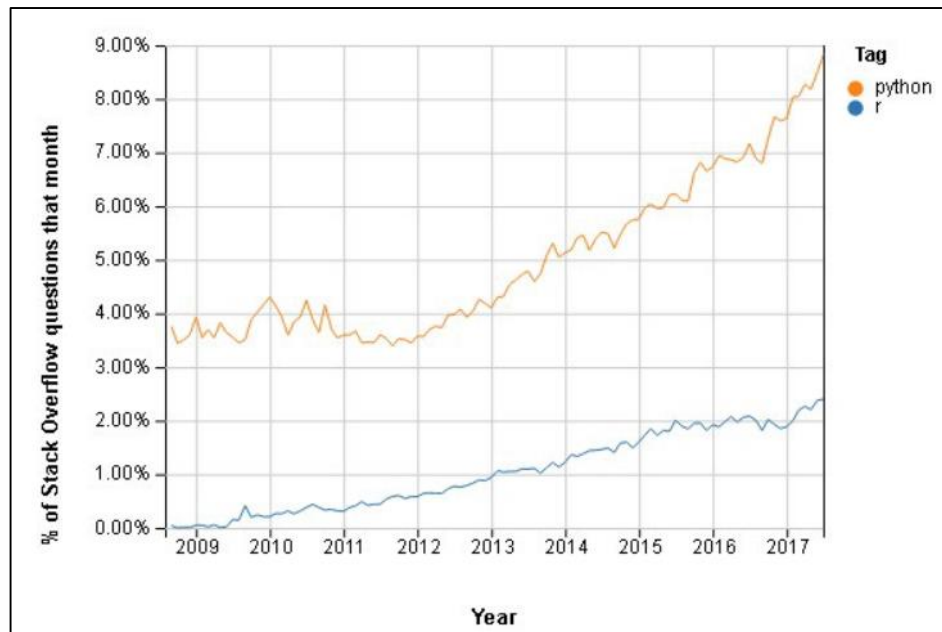


Рисунок 2.17 – Графік популярності Python і R

Ось деякі речі, які слід враховувати при виборі між цими мовами програмування:

- Завдяки легкому для читання синтаксису Python має лінійну та плавну криву навчання. Вважається гарною мовою для програмістів-початківців. З R новачки можуть виконувати завдання аналізу даних за лічені хвилини. Але складність розширених функцій у R ускладнює розвиток експертних знань.

- R – це статистичний інструмент, використовуваний академіками, інженерами та вченими без будь-яких навичок програмування. Python – це готова до роботи мова, яка використовується в широкому спектрі промислових, дослідних та інженерних робочих процесів.

- Програмування на R найкраще підходить для статистичного навчання з неперевершеними бібліотеками для дослідження даних та експериментів. Python – найкращий вибір для машинного навчання та великомасштабних програм, особливо для аналізу даних у веб-додатках.

– Програми R ідеально підходять для візуалізації ваших даних у гарній графіці. Навпаки, програми Python простіше інтегрувати у середовище розробки.

Проаналізувавши все вищеперераховане, можна зробити висновок, що мова програмування Python більш підходить для вирішення даної задачі.

### 2.3 Огляд вхідного набору даних

Як було вказано у розділі 1.4, у Вінниці наразі встановлено 9 постів моніторингу якості атмосфери повітря. Тож для розв'язання даної задачі ресурсами Eco City та SaveEcoBot було надано 9 наборів даних по цих постах. У таблиці 2.1 наведені стовпці з такими даними:

- `id_station` – унікальний номер станції,
- `locality` – назвою тої чи іншої станції,
- `address` – адресою станції,
- `source` – сервіс який надав дані по тій чи іншій станції.

Таблиця 2.1 – Вхідний набір даних

<b>id_station</b>	<b>locality</b>	<b>address</b>	<b>source</b>
337	Олієжир	Чехова, 21	Eco City
767	Соборна 36	Соборна, 36	Eco City
774	Станція Замостя	Замостянська, 7/35	Eco City
790	Станція Вишенька	Космонавтів, 30	Eco City
848	Верхня Слов'янка	Константиновича, 4	Eco City
1315	ВНТУ	Хмельницьке шосе, 95 (буд. 7)	Eco City

## Продовження таблиці 2.1

id_station	locality	address	source
1183	Авалон	Трамвайна, 3	Eco City
761	Тріумф	600-річчя, 66в	Eco City
1008	Сеценка 75	Сеценка, 75	SaveEcoBot

Як бачимо з таблиці (2.1) було взято дані по 8 станціях з Eco City, та 1 станція з SaveEcoBot.

Після того, як дані були оброблені в MS Access (рис 2.18) вони були завантажені в програмне середовище Kaggle.

id_station	datetime	CO2	Humidity	PM1_0	PM10	PM2_5	Pressure	Temperature
1183	11.01.2021 09:29	550.0	32.48	69.11	128.11	113.0	99265.84	24.11
1183	11.01.2021 09:36	550.0	37.06	70.0	125.9	112.8	99273.22	25.96
1183	11.01.2021 09:41	550.0	32.75	67.9	124.9	111.2	99263.88	26.76
1183	11.01.2021 09:42	550.0	44.64	63.4	116.8	106.7	99281.64	25.14
1183	11.01.2021 10:22	550.0	46.47	62.5	109.8	99.3	99294.29	21.09
1183	11.01.2021 10:23	550.0	43.37	59.33	109.78	100.89	99290.63	21.1
1183	11.01.2021 10:24	1451.0	43.03	61.5	112.5	100.7	99296.45	21.29
1183	11.01.2021 10:26	1428.0	43.1	60.4	106.6	96.0	99292.97	21.43
1183	11.01.2021 10:27	1407.0	42.81	60.73	108.0	99.36	99295.47	21.59
1183	11.01.2021 10:28	1396.0	42.59	60.2	109.4	96.5	99293.63	21.75
1183	11.01.2021 10:29	1375.0	42.24	63.27	109.27	98.18	99293.27	21.93
1183	11.01.2021 10:30	1447.0	42.67	58.56	108.89	95.78	99287.98	22.14
1183	11.01.2021 10:42	550.0	44.37	57.8	100.8	89.5	99298.53	22.59
1183	11.01.2021 10:43	550.0	44.25	57.7	102.1	89.8	99300.92	22.32
1183	11.01.2021 10:44	2456.0	44.82	54.33	105.11	93.44	99296.34	22.26
1183	11.01.2021 10:45	2286.0	46.73	56.11	99.56	89.11	99299.64	22.28
1183	11.01.2021 10:47	1950.0	43.19	55.7	93.0	86.4	99298.34	22.24
1183	11.01.2021 10:48	1819.0	43.6	56.2	99.8	88.8	99295.59	22.17

Рисунок 2.18 – Обробка даних в MS Access

На рисунку 2.18, бачимо завершений варіант таблиці, де дані було транспоновано та відсортовано по ознаках. Загалом усі станції фіксували значення за такими показниками:

- pm1 – PM1.0 (частинки пилу розміром 1 мкм), мкг/куб.м;
- pm25 – PM2,5 (частинки пилу розміром 2,5 мкм), мкг/куб.м;
- pm10 – PM10 (частинки пилу розміром 10 мкм), мкг/куб.м;
- CO2 – вуглекислий газ, мкг/куб.м;



- temperature – температура, в градусах Цельсія;
- вологість – Вологість, %;
- тиск – Тиск, міліметри ртутного стовпа.

## 2.4 Висновки

У даному розділі було здійснено вибір оптимальних інформаційних технологій та мови програмування, оглянуто вхідні дані, та проаналізовано показники, що фіксуються постами моніторингу якості атмосферного повітря.

## 3 ПРОЕКТУВАННЯ ЗАДАЧІ

### 3.1 Щоденне прогнозування даних

#### 3.1.1 Імпорт бібліотек та завантаження даних

Спочатку завантажимо основні бібліотеки, необхідні для коректної роботи з даними (рис. 3.1). Також додамо бібліотеки необхідні для наступного: завантаження та візуалізація даних (рис. 3.2), моделювання та прогнозування даних (рис. 3.3).

```
import numpy as np
import pandas as pd
from scipy.interpolate import Rbf, interp2d
import datetime
```

Рисунок 3.1 – Бібліотеки для роботи з даними

```
# For import data
import os

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

Рисунок 3.2 – Бібліотеки для завантаження та візуалізація даних

```
# Modeling and Prediction
from fbprophet import Prophet
from sklearn.metrics import r2_score, mean_absolute_error
```

Рисунок 3.3 – Бібліотеки для моделювання та прогнозування даних

Далі встановлюються параметри для роботи з даними в fbprophet (рис. 3.4).

```
N = 30
forecasting_period = 2
indicator_names = ['PM2.5', 'PM10']
indicator_name = indicator_names[0]
type_agg='mean'
time_interval = 'D' # day
Q = 14 # 2 weeks = 14 days
```

Рисунок 3.4 – Встановлення параметрів

Наступним кроком імпортуємо дані з датасету Kaggle, так як це зображено на рисунку 3.5.

```
dataset_files = []
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        dataset_files.append(os.path.join(dirname, filename))
dataset_files
```

Рисунок 3.5 – Імпорт даних

Після чого ми можемо завантажити набори даних по усіх станціях у Вінницькій області. Дані завантажуються з середовищ SaveEcoBot (рис. 3.6) та EcoCity (рис. 3.7).

	id_saveecobot	id_ecocity	network	locality	address	start_date
0	14627	848.0	Eco-City	Vinnytsia city	2nd Konstantynovycha Ln, 4	14.12.2020
1	4152	767.0	Eco-City	Vinnytsia city	Soborna St, 36	26.06.2020
2	1008	NaN	SaveDnipro	Vinnytsia city	Stetsenka St, 75	25.05.2019
3	17796	NaN	luftdaten.info	Vinnytsia city	Kyivs'ka St, 5	16.05.2021
4	15952	337.0	Eco-City	Vinnytsia city	3rd Chekhova Passage, 21	23.01.2021

Рисунок 3.6 – Завантаження даних середовища SaveEcoBot

	id_saveecobot	id_ecocity	network	locality	address	start_date	lat	lng
0	15952.0	337	Eco-City	Vinnytsia city	3rd Chekhova Passage, 21	23.01.2021	49.2267	28.5259
1	4152.0	767	Eco-City	Vinnytsia city	Soborna St, 36	26.06.2020	49.2329	28.4704
2	NaN	774	Eco-City	Vinnytsia city	Vinnytsia, Zamostyanska Street, 7/35	NaN	49.2455	28.4936
3	NaN	790	Eco-City	Vinnytsia city	Vinnytsia, Cosmonauts Avenue, 30	NaN	49.2268	28.4194
4	14627.0	848	Eco-City	Vinnytsia city	2nd Konstantynovycha Ln, 4	14.12.2020	49.2307	28.4371
5	NaN	1315	Eco-City	Vinnytsia city	Vinnytsia, str. Khmelnytsky shose, 95, VNTU, b...	16.11.2021	49.2330	28.4100
6	NaN	1183	Eco-City	Vinnytsia city	Vinnytsia, str. Tram, 3	NaN	49.2311	28.4371
7	NaN	761	Eco-City	Vinnytsia city	Vinnytsia, 600th anniversary, 66V	NaN	49.2222	28.4242

Рисунок 3.7 – Завантаження даних середовища Eco City

Проаналізувавши та об'єднавши ці дві таблиці, а також вилучивши значення, що дублюються, ми отримаємо таблицю з загальними даними, яка зображена на рисунку 3.8.

	ds	value	network	id_station	lat	lng
0	2019-05-25	10.131005	SaveDnipro	SaveEcoBot_1008	49.2393	28.4799
1	2019-06-04	4.244945	SaveDnipro	SaveEcoBot_1008	49.2393	28.4799
2	2019-06-05	7.482687	SaveDnipro	SaveEcoBot_1008	49.2393	28.4799
3	2019-06-06	3.982741	SaveDnipro	SaveEcoBot_1008	49.2393	28.4799
4	2019-06-07	5.130192	SaveDnipro	SaveEcoBot_1008	49.2393	28.4799
...	...	...	...	...	...	...
1663	2021-11-20	1.552563	Eco-City	EcoCity_761	49.2222	28.4242
1664	2021-11-24	5.259661	Eco-City	EcoCity_761	49.2222	28.4242
1665	2021-11-25	6.709394	Eco-City	EcoCity_761	49.2222	28.4242
1666	2021-11-26	4.930761	Eco-City	EcoCity_761	49.2222	28.4242
1667	2021-11-27	3.560560	Eco-City	EcoCity_761	49.2222	28.4242

Рисунок 3.8 – Таблиця з загальними даними

### 3.1.2 Прогнозування даних

Перш ніж почати прогнозування, потрібно обрати декілька постів з значеннями, які підходять для більш кращої візуалізації. Значення отримуємо з попередньої таблиці (рис. 3.8).

На рисунку 3.9 зображена візуалізація значень з поста SaveEcoBot\_1008. По осі ординат взято параметр value – значення, а по осі абсцис – порядковий номер в таблиці. Також там зображені дві таблиці з значеннями, частина з яких використовується при побудові графіку.

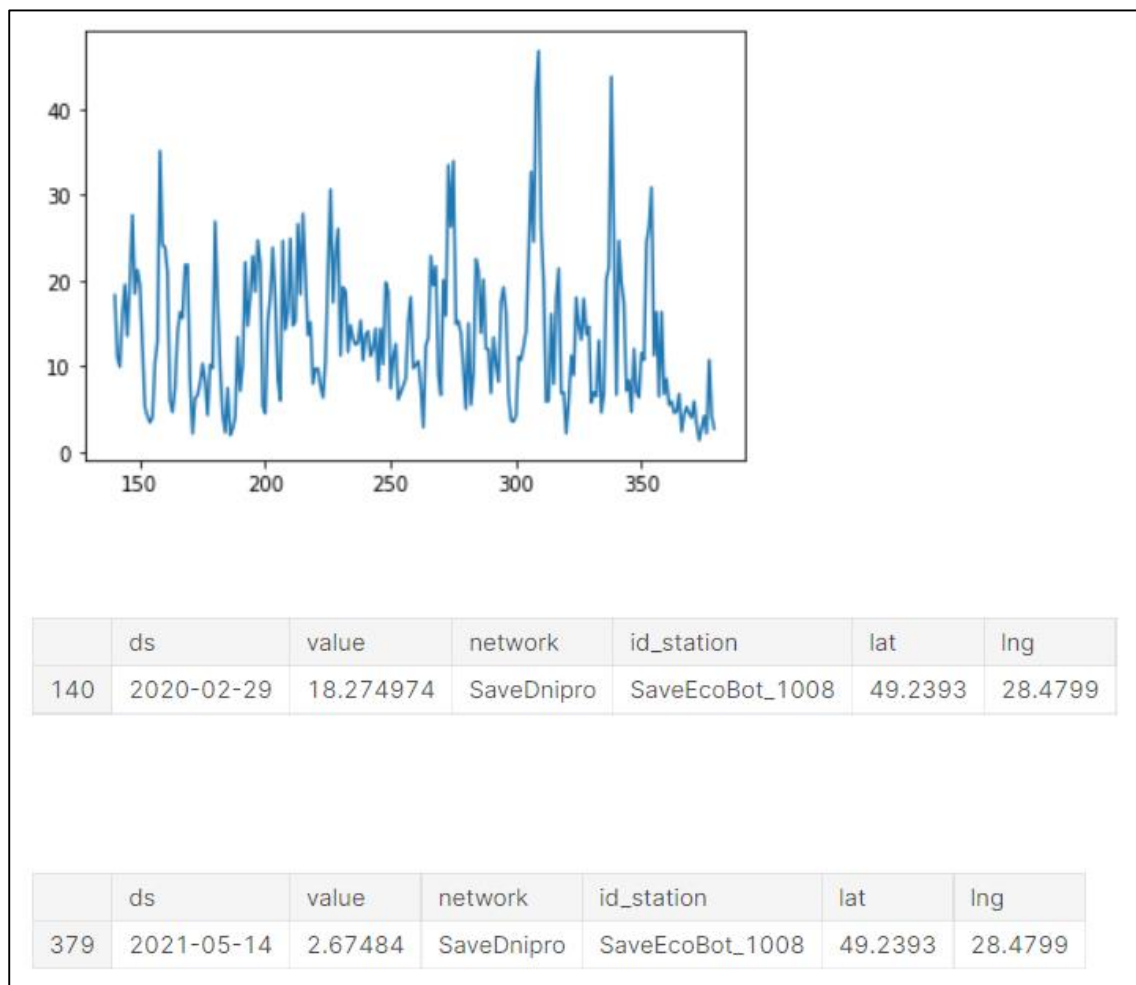


Рисунок 3.9 – Візуалізація значень з поста SaveEcoBot\_1008

На рисунку 3.10 зображена візуалізація значень з поста SaveEcoBot\_17796. По осі ординат взято параметр value – значення, а по осі

абсцис – порядковий номер в таблиці. Також там зображені дві таблиці з значеннями, частина з яких використовується при побудові графіку.

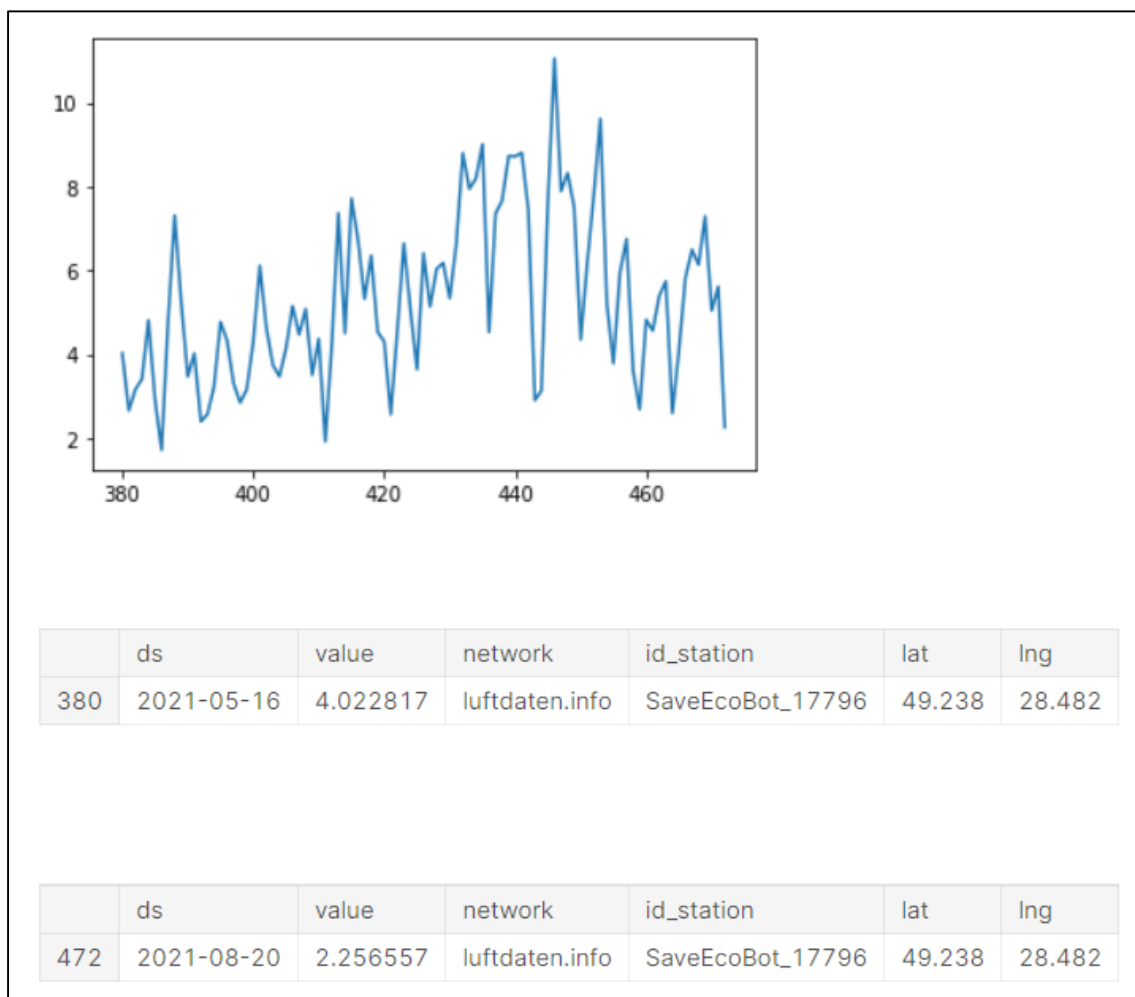


Рисунок 3.10 – Візуалізація значень з поста SaveEcoBot\_17796

На рисунку 3.11 зображена візуалізація значень з поста EcoCity\_337. По осі ординат взято параметр value – значення, а по осі абсцис – порядковий номер в таблиці. Також там зображені дві таблиці з значеннями, частина з яких використовується при побудові графіку.

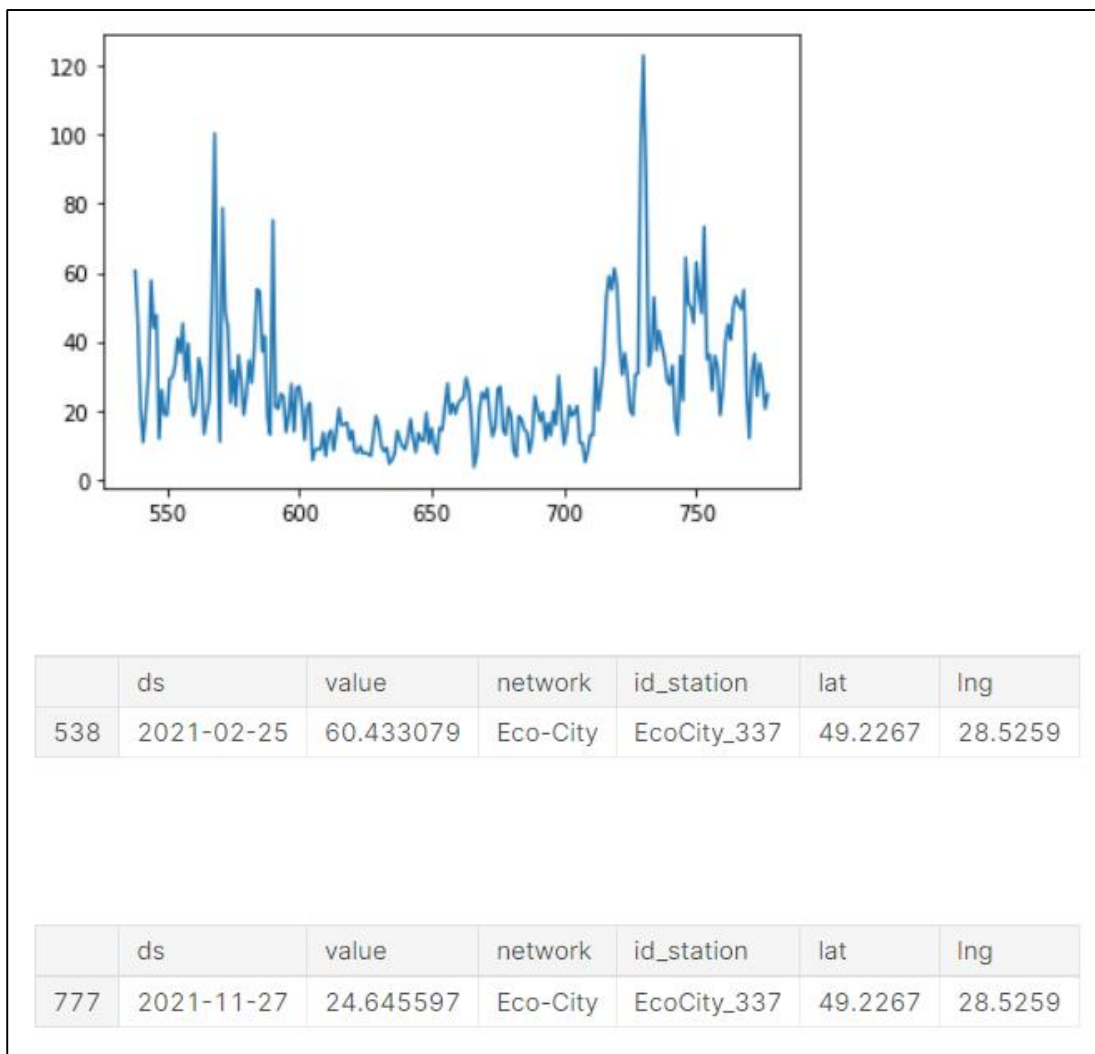


Рисунок 3.11 – Візуалізація значень з поста EcoCity\_337

На рисунку 3.12 зображена візуалізація значень з поста EcoCity\_767. По осі ординат взято параметр value – значення, а по осі абсцис – порядковий номер в таблиці. Також там зображені дві таблиці з значеннями, частина з яких використовується при побудові графіку.

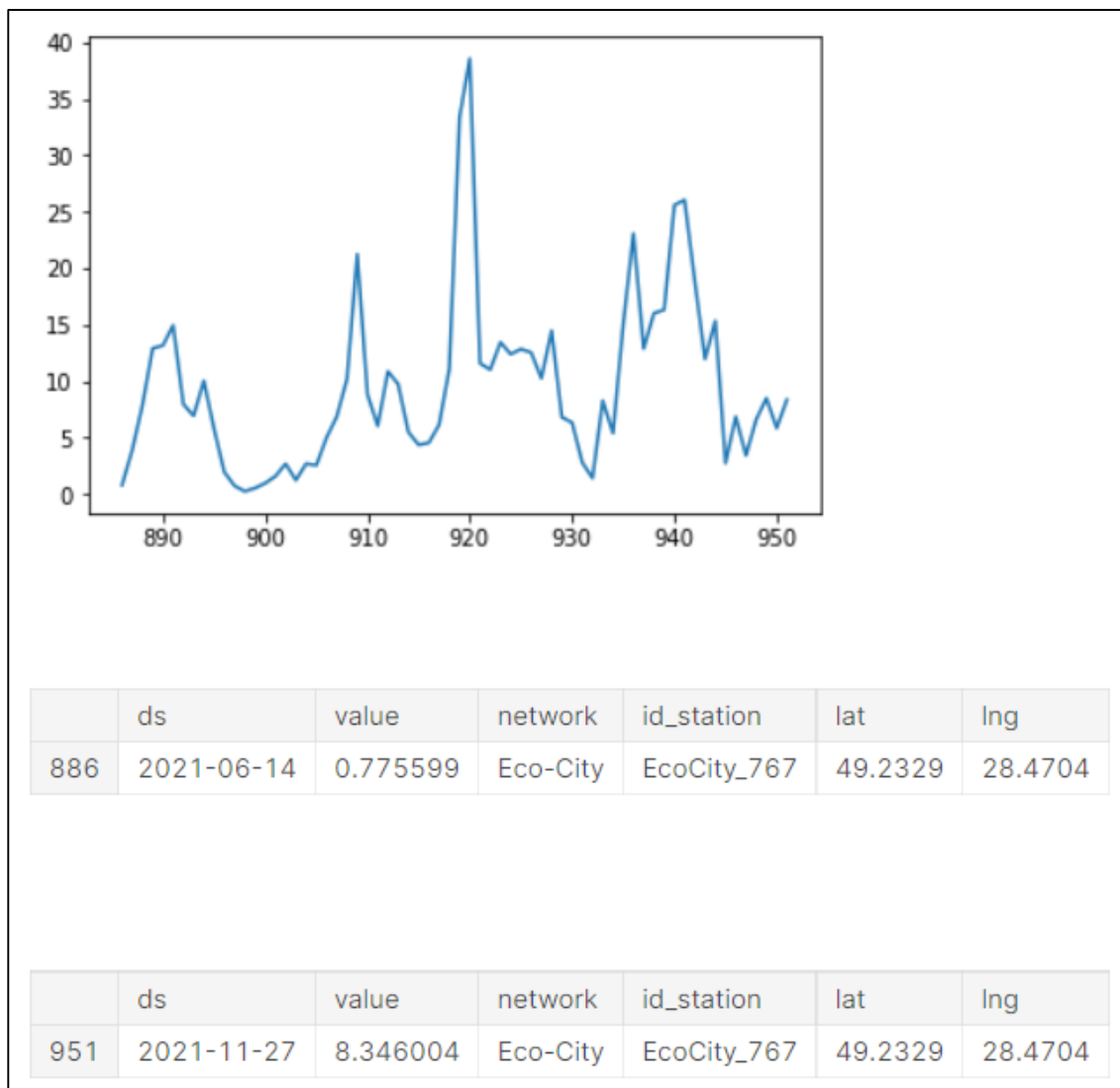


Рисунок 3.12 – Візуалізація значень з поста EcoCity\_767

На рисунку 3.13 зображена візуалізація значень з поста EcoCity\_774. По осі ординат взято параметр value – значення, а по осі абсцис – порядковий номер в таблиці. Також там зображені дві таблиці з значеннями, частина з яких використовується при побудові графіку.



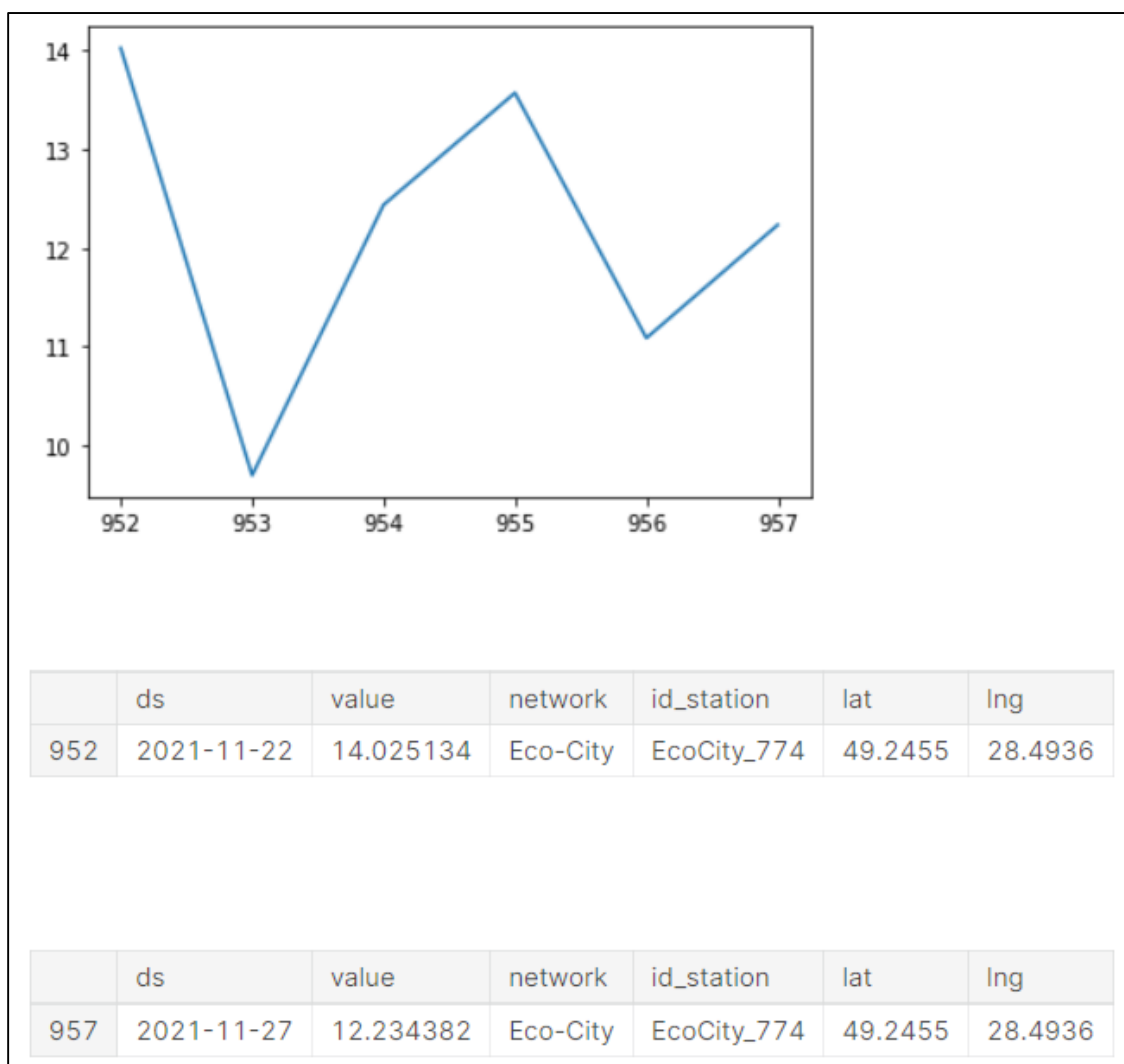


Рисунок 3.13 – Візуалізація значень з поста EcoCity\_774

На рисунку 3.14 зображена візуалізація значень з поста EcoCity\_790. По осі ординат взято параметр value – значення, а по осі абсцис – порядковий номер в таблиці. Також там зображені дві таблиці з значеннями, частина з яких використовується при побудові графіку.

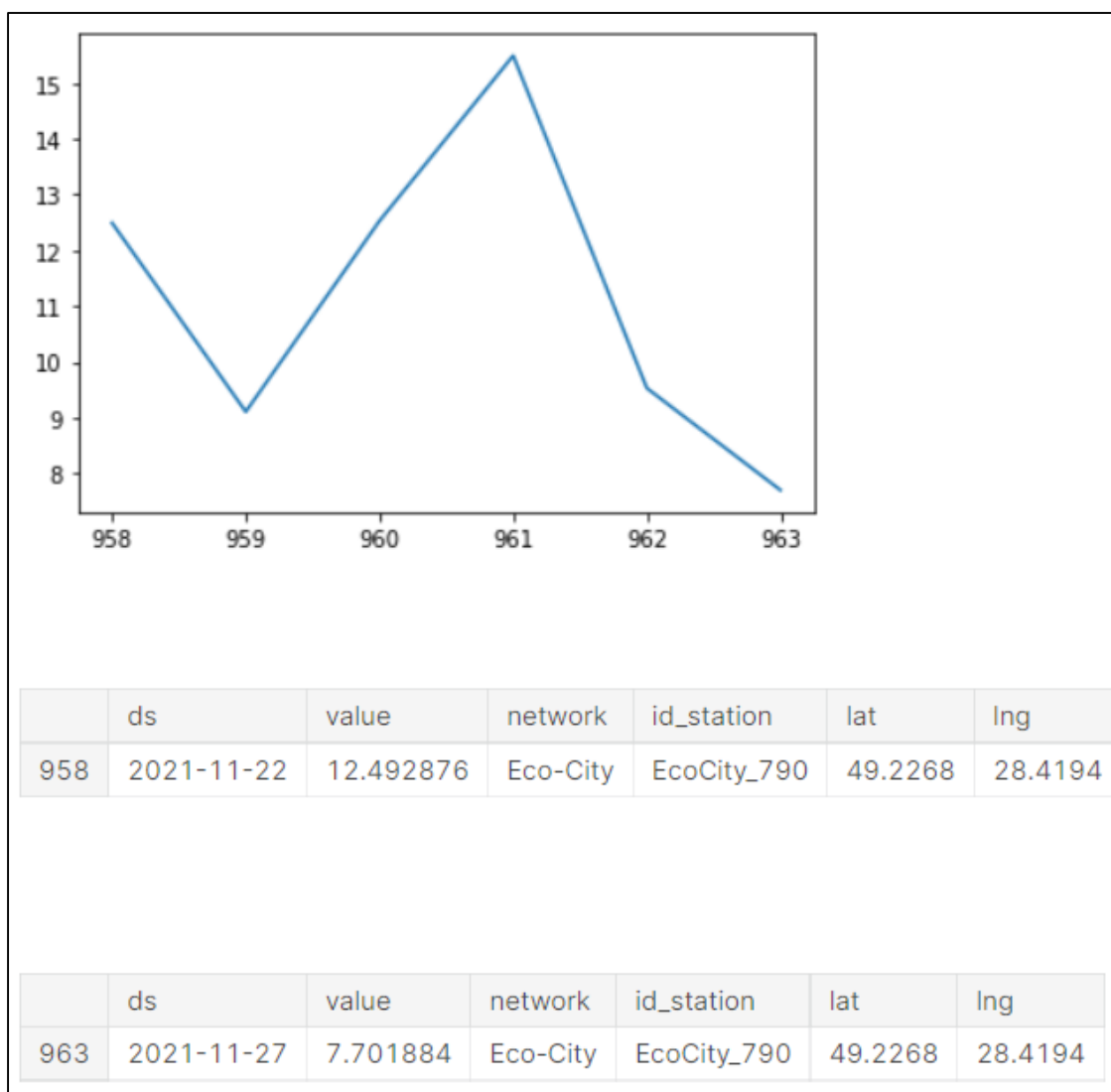


Рисунок 3.14 – Візуалізація значень з поста EcoCity\_790

На рисунку 3.15 зображена візуалізація значень з поста EcoCity\_848. По осі ординат взято параметр value – значення, а по осі абсцис – порядковий номер в таблиці. Також там зображені дві таблиці з значеннями, частина з яких використовується при побудові графіку.

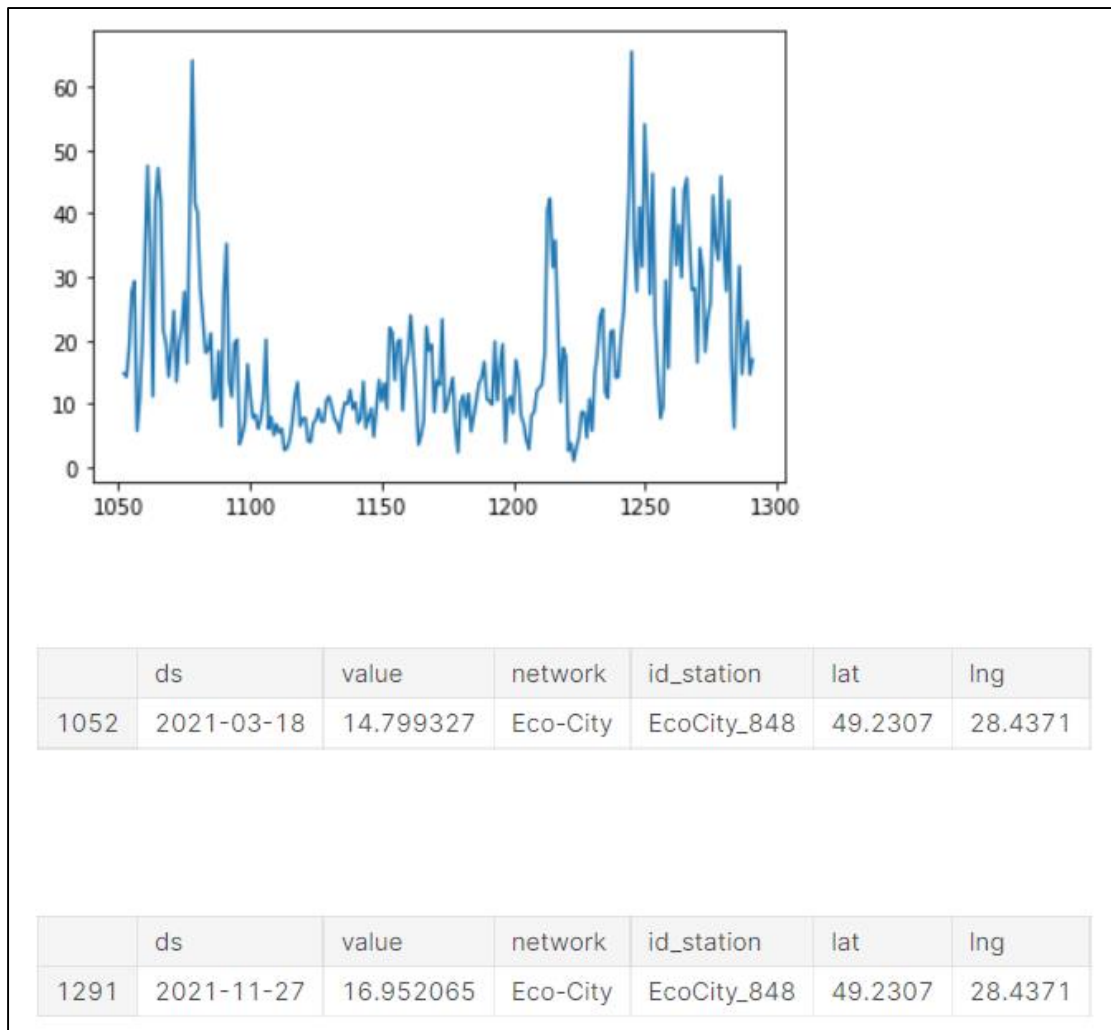


Рисунок 3.15 – Візуалізація значень з поста EcoCity\_848

На рисунку 3.16 зображена візуалізація значень з поста EcoCity\_1315. По осі ординат взято параметр value – значення, а по осі абсцис – порядковий номер в таблиці. Також там зображені дві таблиці з значеннями, частина з яких використовується при побудові графіку.

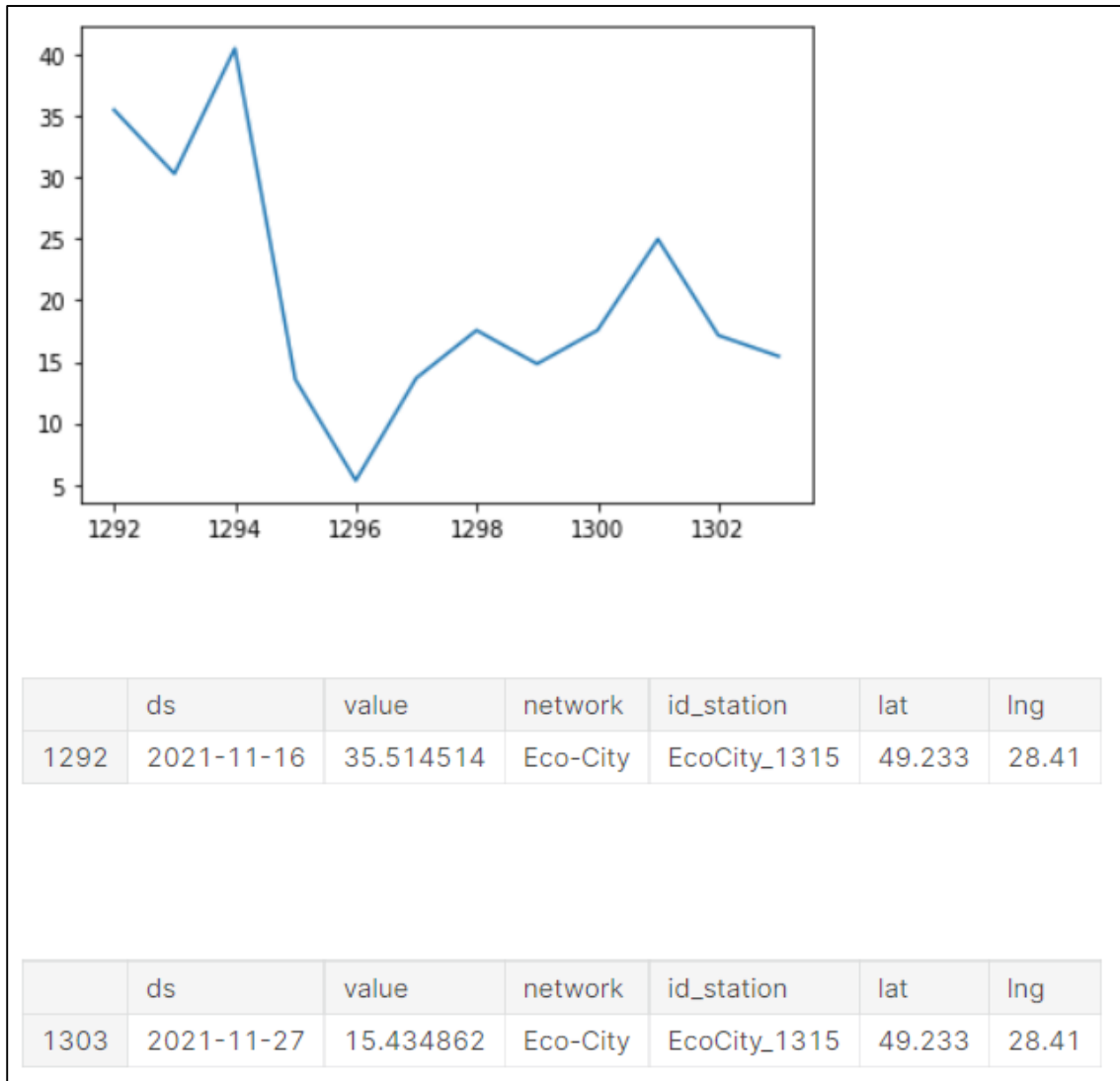


Рисунок 3.16 – Візуалізація значень з поста EcoCity\_1315

На рисунку 3.17 зображена візуалізація значень з поста EcoCity\_1183. По осі ординат взято параметр value – значення, а по осі абсцис – порядковий номер в таблиці. Також там зображені дві таблиці з значеннями, частина з яких використовується при побудові графіку.

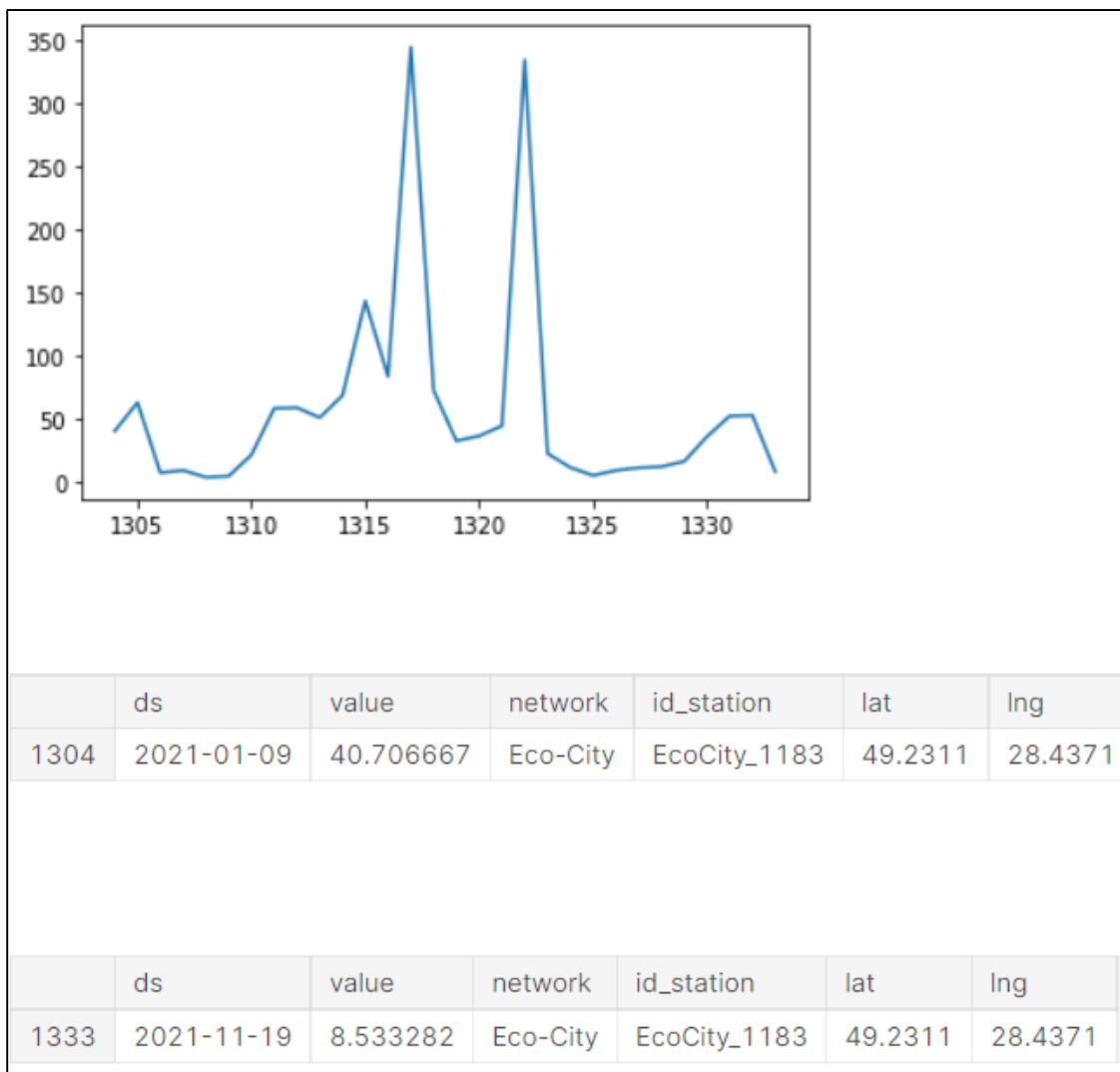


Рисунок 3.17 – Візуалізація значень з поста EcoCity\_1183

На рисунку 3.18 зображена візуалізація значень з поста EcoCity\_761. По осі ординат взято параметр value – значення, а по осі абсцис – порядковий номер в таблиці. Також там зображені дві таблиці з значеннями, частина з яких використовується при побудові графіку.

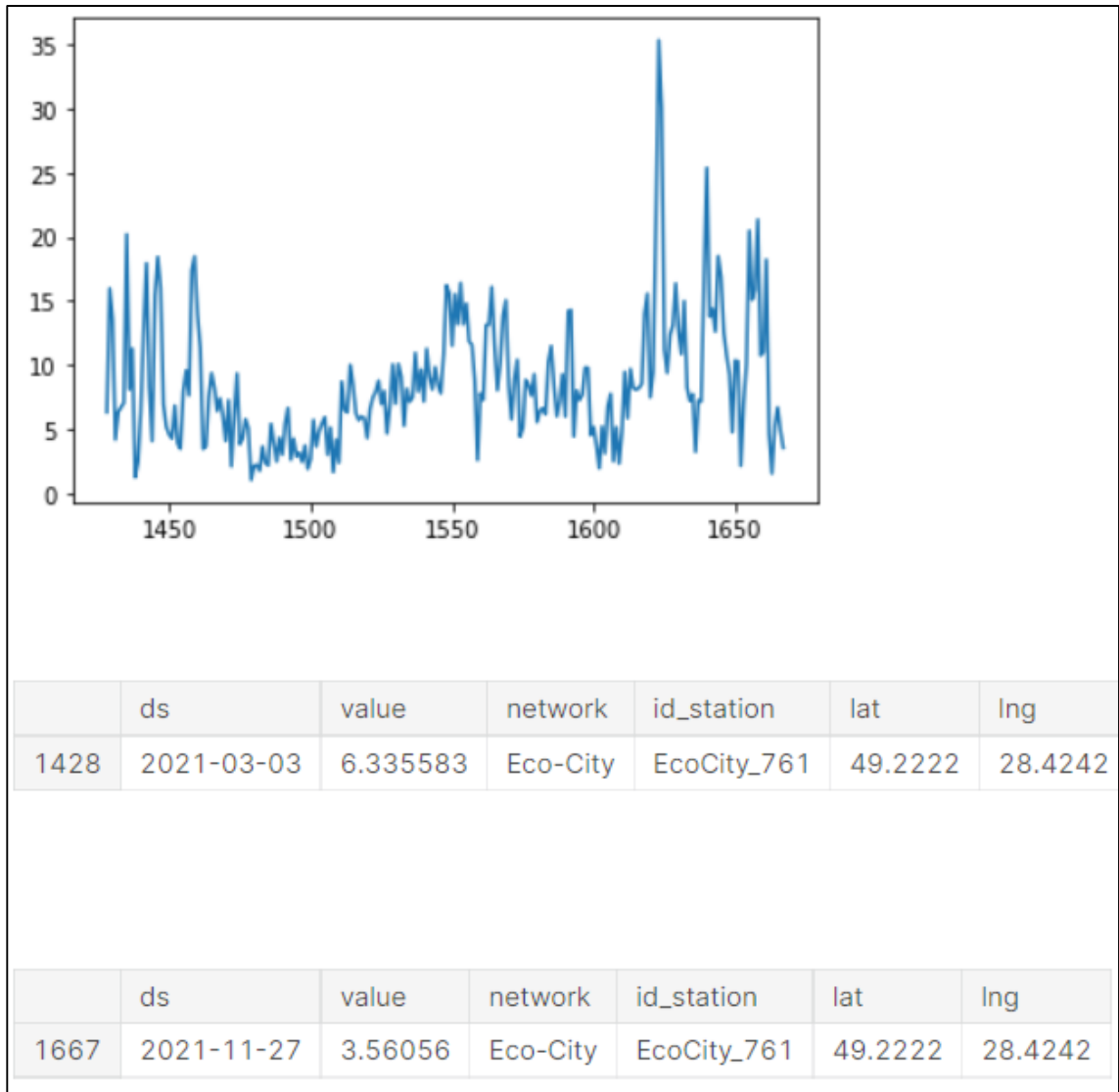


Рисунок 3.18 – Візуалізація значень з поста EcoCity\_761

Проаналізувавши графіки, наведені вище, можна зробити висновок, що найбільше нам підходять дані з таких постів: EcoCity\_337, EcoCity\_848, EcoCity\_1315.

Тепер, перевіримо дані з цих постів, на наявність аномальних даних, це необхідно для створення коректного прогнозу.

На рисунку 3.19 зображено графік з позначенням аномальних значень з поста EcoCity\_337. По осі ординат взято параметр value – значення, а по осі абсцис – дата.

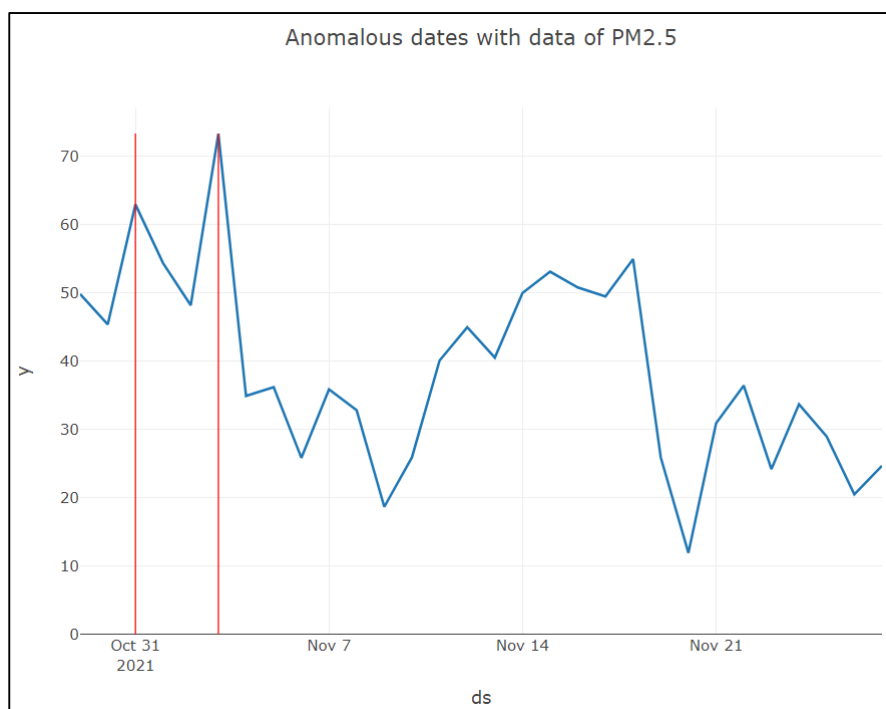


Рисунок 3.19 – Графік з аномальними даними з поста EcoCity\_337

На рисунку 3.20 зображено графік з позначенням аномальних значень з поста EcoCity\_848. По осі ординат взято параметр value – значення, а по осі абсцис – дата.

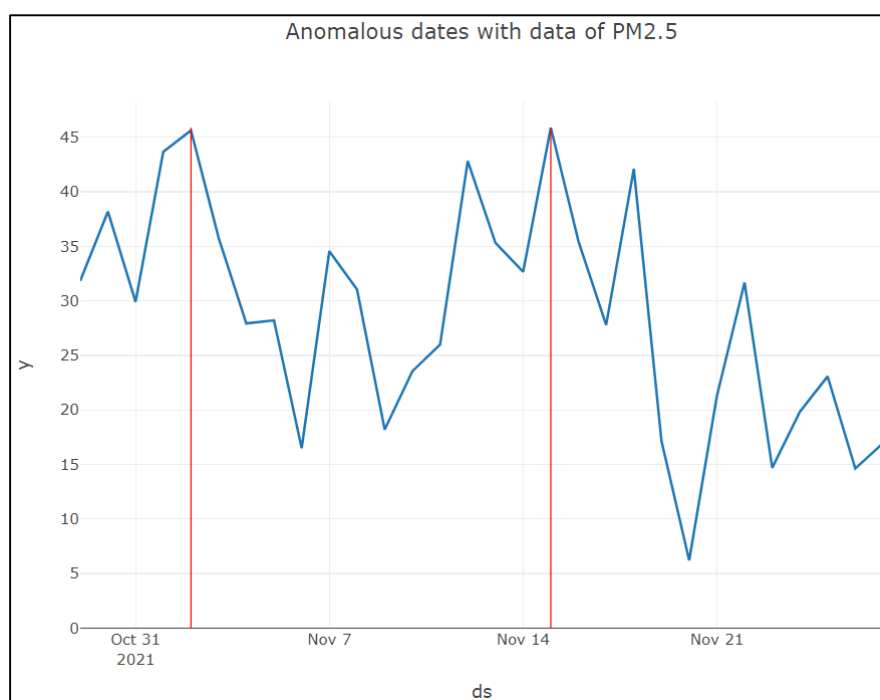


Рисунок 3.20 – Графік з аномальними даними з поста EcoCity\_848

Аналогічно на рисунку 3.21 зображено графік з позначенням аномальних значень з поста EcoCity\_1315. По осі ординат взято параметр value – значення, а по осі абсцис – дата.

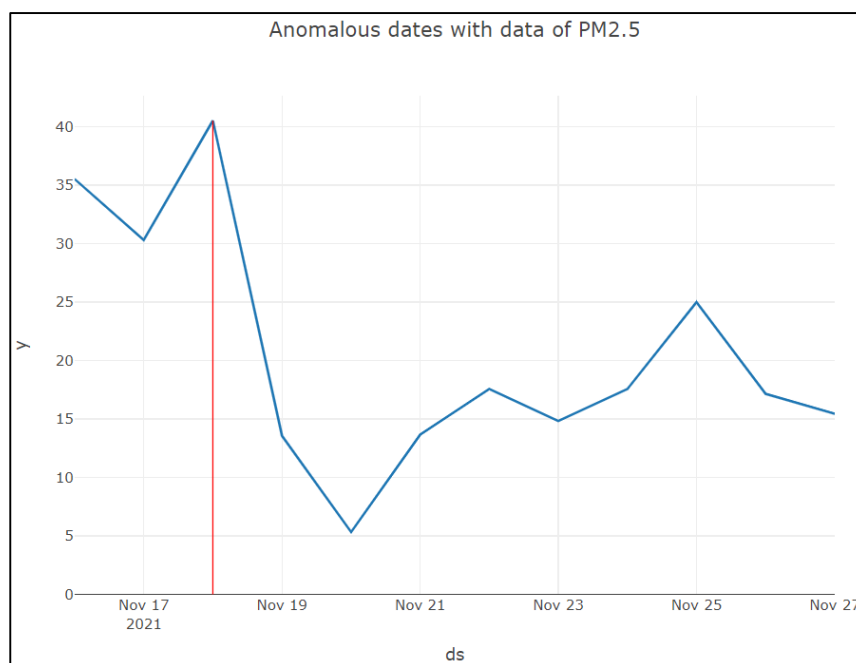


Рисунок 3.21 – Графік з аномальними даними з поста EcoCity\_1315

Далі було побудовано графіки з прогнозом кількості пилу в атмосферному повітрі на наступні дві доби, для кожного з цих постів (рис. 3.22–3.24). І також був побудований тренд до них, що зображено на рисунках 3.25–3.27. По осі ординат показана кількість частинок пилу розміром 2,5 мікрметра, а по осі абсцис – дата.



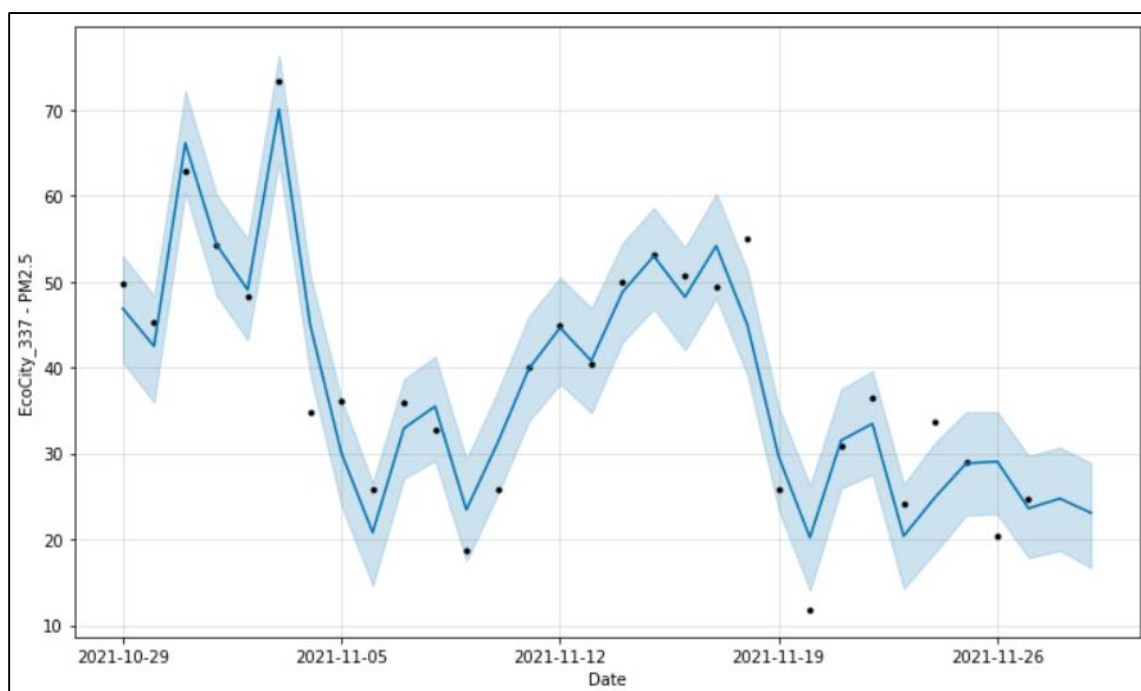


Рисунок 3.22 – Графік з прогнозом кількості пилу для поста EcoCity\_337

Відносна похибка для поста EcoCity\_337 становить 3,58 – 3,64%.

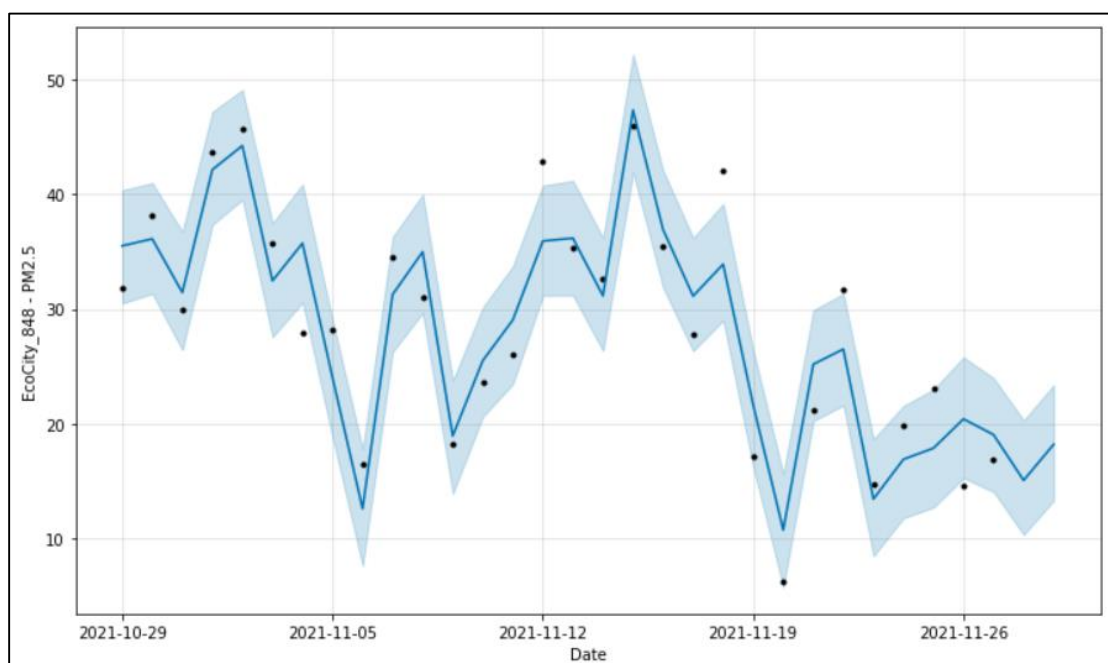


Рисунок 3.23 – Графік з прогнозом кількості пилу для поста EcoCity\_848

Відносна похибка для поста EcoCity\_848 становить 3,38 – 4,08%.

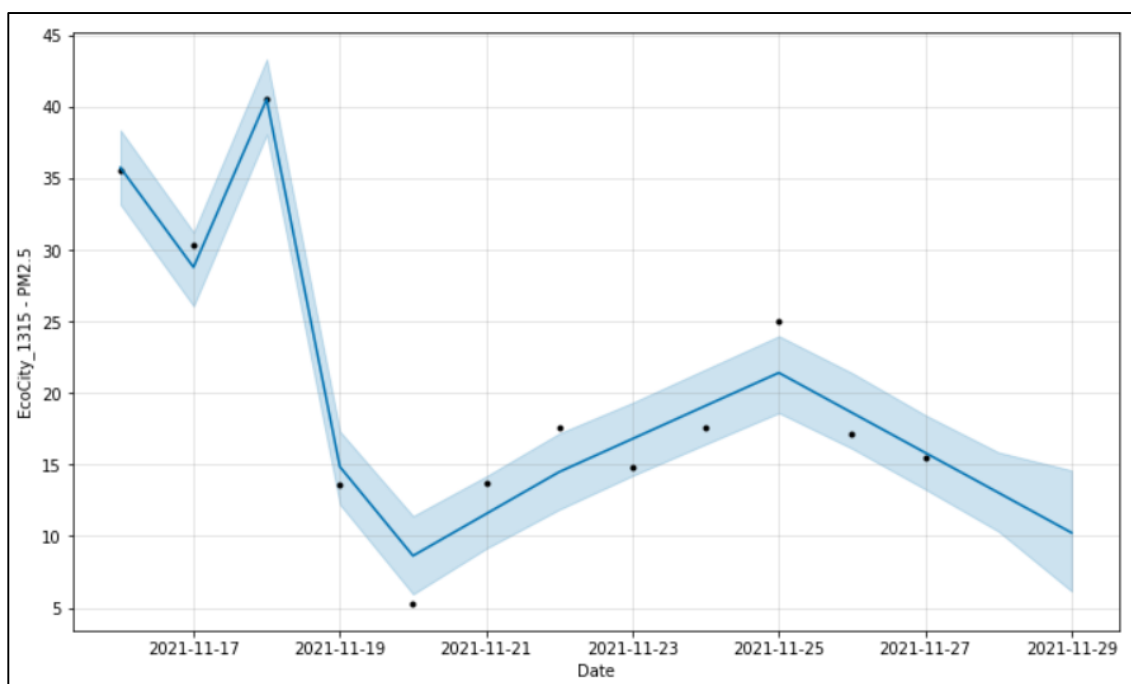


Рисунок 3.24 – Графік з прогнозом кількості пилу для поста EcoCity\_1315

Відносна похибка для поста EcoCity\_1315 становить 1,71%.

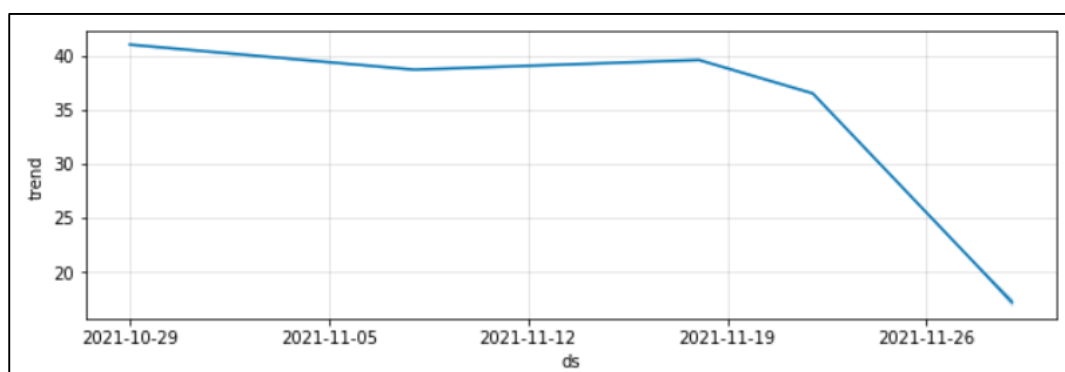


Рисунок 3.25 – Графік з трендом для поста EcoCity\_337

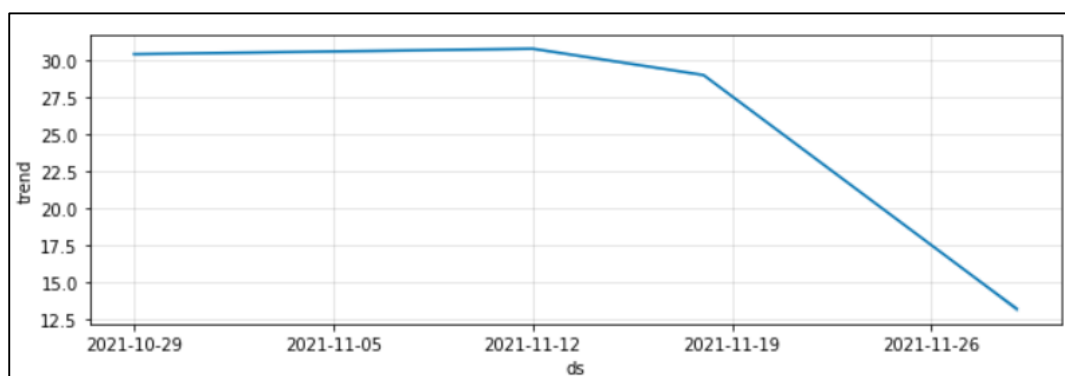


Рисунок 3.26 – Графік з трендом для поста EcoCity\_848

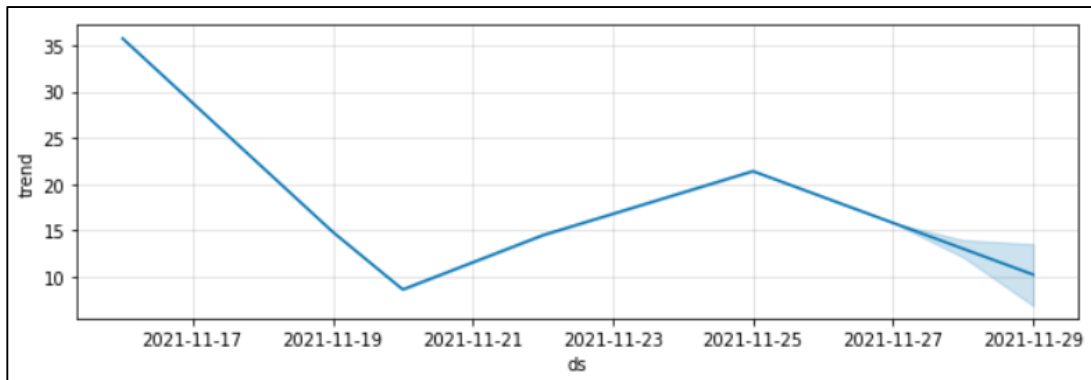


Рисунок 3.27 – Графік з трендом для поста EcoCity\_1315

## 3.2 Щогодинне прогнозування даних

### 3.2.1 Імпорт бібліотек та завантаження даних

Аналогічно з розділом 3.1.1, спочатку завантажимо основні бібліотеки, необхідні для коректної роботи з даними (рис. 3.28), потім додамо бібліотеки необхідні для наступного: завантаження та візуалізація даних (рис. 3.29), моделювання та прогнозування даних (рис. 3.30).

```
import numpy as np
import pandas as pd
from scipy.interpolate import Rbf, interp2d
import datetime
```

Рисунок 3.28 – Бібліотеки для роботи з даними

```

# For import data
import os

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

```

Рисунок 3.29 – Бібліотеки для завантаження та візуалізація даних

```

# Modeling and Prediction
from fbprophet import Prophet
from sklearn.metrics import r2_score, mean_absolute_error

```

Рисунок 3.30 – Бібліотеки для моделювання та прогнозування даних

Далі встановлюються параметри для роботи з даними в fbprophet (рис. 3.31).

```

N = 15*24
forecasting_period = 2 # hours
time_interval='H'
Q = 12 # Seasonality is less than a day, hours
indicator_names = ['PM2.5', 'PM10']
indicator_name = indicator_names[0]
type_agg='mean'

```

Рисунок 3.31 – Встановлення параметрів

Наступним кроком імпортуємо дані з датасету Kaggle, так як це зображено на рисунку 3.32.

```

dataset_files = []
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        dataset_files.append(os.path.join(dirname, filename))
dataset_files

```

Рисунок 3.32 – Імпорт даних

Після чого ми можемо завантажити набори даних по усіх станціях у Вінницькій області. Дані завантажуються з середовищ SaveEcoBot (рис. 3.33) та EcoCity (рис. 3.34).

	id_saveecobot	id_ecocity	network	locality	address	start_date
0	14627	848.0	Eco-City	Vinnytsia city	2nd Konstanynovycha Ln, 4	14.12.2020
1	4152	767.0	Eco-City	Vinnytsia city	Soborna St, 36	26.06.2020
2	1008	NaN	SaveDnipro	Vinnytsia city	Stetsenka St, 75	25.05.2019
3	17796	NaN	luftdaten.info	Vinnytsia city	Kyivs'ka St, 5	16.05.2021
4	15952	337.0	Eco-City	Vinnytsia city	3rd Chekhova Passage, 21	23.01.2021

Рисунок 3.33 – Завантаження даних середовища SaveEcoBot

	id_saveecobot	id_ecocity	network	locality	address	start_date	lat	lng
0	15952.0	337	Eco-City	Vinnytsia city	3rd Chekhova Passage, 21	23.01.2021	49.2267	28.5259
1	4152.0	767	Eco-City	Vinnytsia city	Soborna St, 36	26.06.2020	49.2329	28.4704
2	NaN	774	Eco-City	Vinnytsia city	Vinnytsia, Zamosyanska Street, 7/35	NaN	49.2455	28.4936
3	NaN	790	Eco-City	Vinnytsia city	Vinnytsia, Cosmonauts Avenue, 30	NaN	49.2268	28.4194
4	14627.0	848	Eco-City	Vinnytsia city	2nd Konstanynovycha Ln, 4	14.12.2020	49.2307	28.4371
5	NaN	1315	Eco-City	Vinnytsia city	Vinnytsia, str. Khmelnytsky shose, 95, VNTU, b...	16.11.2021	49.2330	28.4100
6	NaN	1183	Eco-City	Vinnytsia city	Vinnytsia, str. Tram, 3	NaN	49.2311	28.4371
7	NaN	761	Eco-City	Vinnytsia city	Vinnytsia, 600th anniversary, 66V	NaN	49.2222	28.4242

Рисунок 3.34 – Завантаження даних середовища Eco City

Проаналізувавши та об'єднавши ці дві таблиці, а також вилучивши значення, що дублюються, ми отримаємо таблицю з загальними даними, яка зображена на рисунку 3.35.

	ds	value	network	id_station	lat	lng
0	2019-05-25 06:00:00	21.555000	SaveDnipro	SaveEcoBot_1008	49.2393	28.4799
1	2019-05-25 09:00:00	9.730000	SaveDnipro	SaveEcoBot_1008	49.2393	28.4799
2	2019-05-25 10:00:00	12.286087	SaveDnipro	SaveEcoBot_1008	49.2393	28.4799
3	2019-05-25 11:00:00	12.141739	SaveDnipro	SaveEcoBot_1008	49.2393	28.4799
4	2019-05-25 12:00:00	12.543636	SaveDnipro	SaveEcoBot_1008	49.2393	28.4799
...	...	...	...	...	...	...
33314	2021-11-27 05:00:00	4.478070	Eco-City	EcoCity_761	49.2222	28.4242
33315	2021-11-27 06:00:00	5.415690	Eco-City	EcoCity_761	49.2222	28.4242
33316	2021-11-27 07:00:00	5.672105	Eco-City	EcoCity_761	49.2222	28.4242
33317	2021-11-27 08:00:00	5.674035	Eco-City	EcoCity_761	49.2222	28.4242
33318	2021-11-27 09:00:00	5.964091	Eco-City	EcoCity_761	49.2222	28.4242

Рисунок 3.35 – Таблиця з загальними даними

### 3.2.2 Прогнозування даних

Перш ніж почати прогнозування, потрібно обрати декілька постів з значеннями, які підходять для більш кращої візуалізації. Значення отримуємо з попередньої таблиці (рис. 3.35).

На рисунку 3.36 зображена візуалізація значень з поста SaveEcoBot\_1008. По осі ординат взято параметр value – кількість частинок пилу розміром 2,5 мікрметра, а по осі абсцис – порядковий номер в таблиці. Також там зображені дві таблиці з значеннями, частина з яких використовується при побудові графіку.

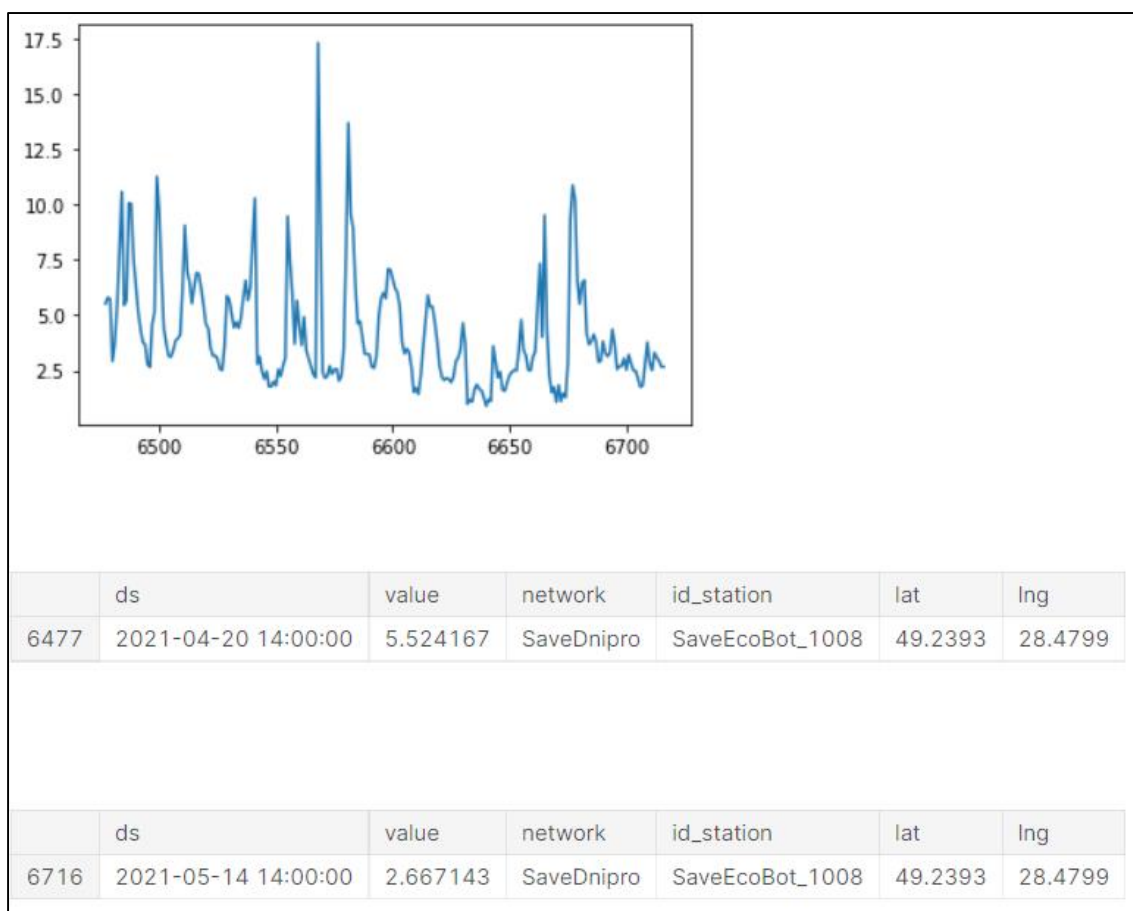


Рисунок 3.36 – Візуалізація значень з поста SaveEcoBot\_1008

На рисунку 3.37 зображена візуалізація значень з поста SaveEcoBot\_17796. По осі ординат взято параметр value – кількість частинок пилу розміром 2,5 мікрметра, а по осі абсцис – порядковий номер в таблиці. Також там зображені дві таблиці з значеннями, частина з яких використовується при побудові графіку.

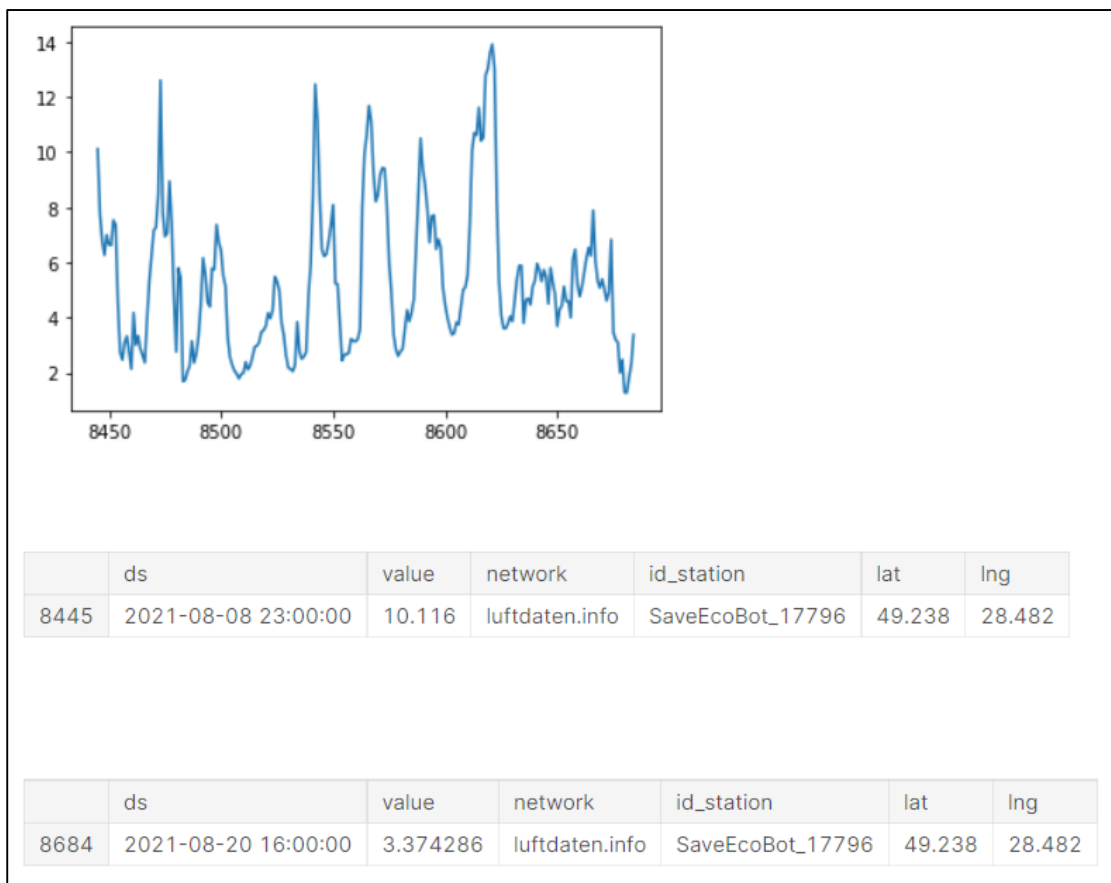


Рисунок 3.37 – Візуалізація значень з поста SaveEcoBot\_17796

На рисунку 3.38 зображена візуалізація значень з поста EcoCity\_337. По осі ординат взято параметр value – кількість частинок пилу розміром 2,5 мікрметра, а по осі абсцис – порядковий номер в таблиці. Також там зображені дві таблиці з значеннями, частина з яких використовується при побудові графіку.



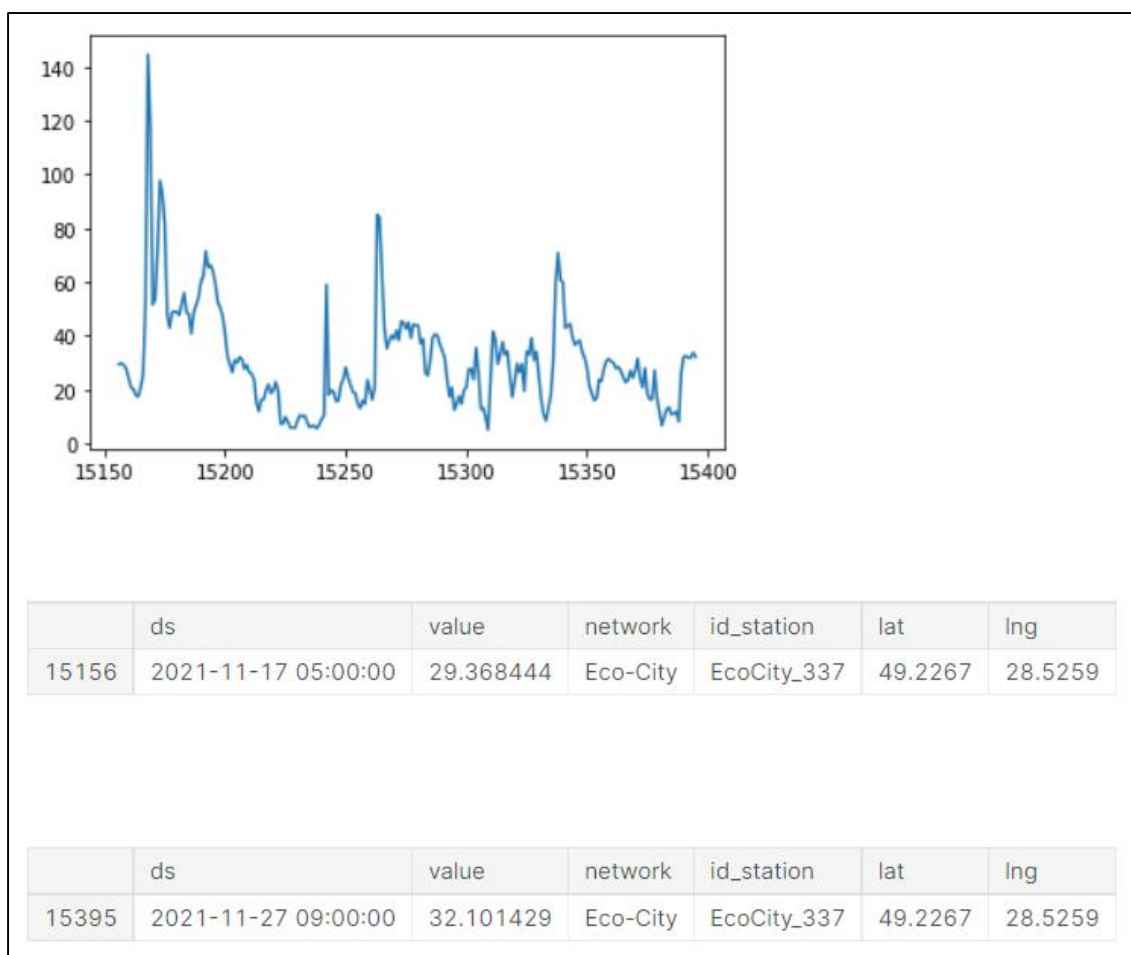


Рисунок 3.38 – Візуалізація значень з поста EcoCity\_337

На рисунку 3.39 зображена візуалізація значень з поста EcoCity\_767. По осі ординат взято параметр value – кількість частинок пилу розміром 2,5 мікрметра, а по осі абсцис – порядковий номер в таблиці. Також там зображені дві таблиці з значеннями, частина з яких використовується при побудові графіку.

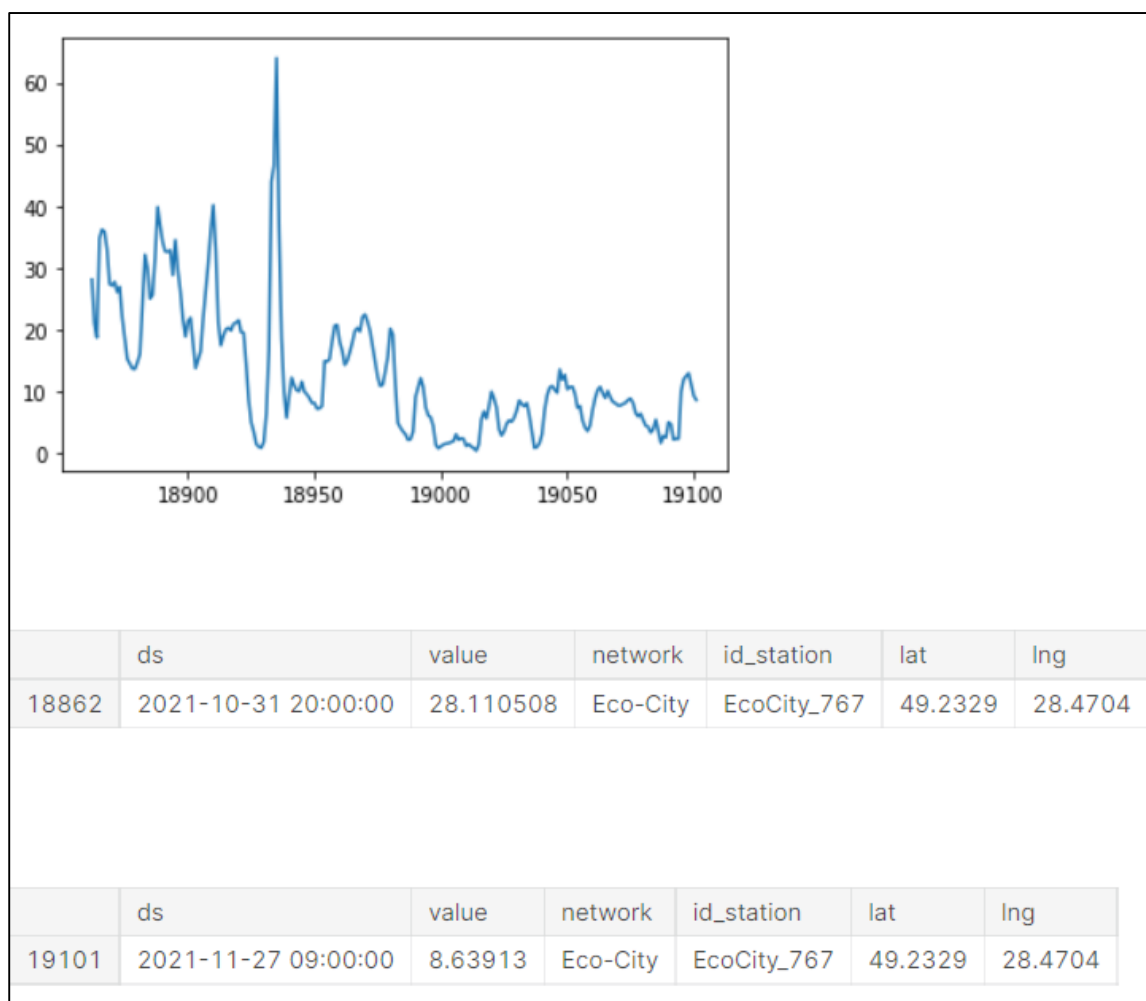


Рисунок 3.39 – Візуалізація значень з поста EcoCity\_767

На рисунку 3.40 зображена візуалізація значень з поста EcoCity\_774. По осі ординат взято параметр value – кількість частинок пилу розміром 2,5 мікрметра, а по осі абсцис – порядковий номер в таблиці. Також там зображені дві таблиці з значеннями, частина з яких використовується при побудові графіку.



Рисунок 3.40 – Візуалізація значень з поста EcoCity\_774

На рисунку 3.41 зображена візуалізація значень з поста EcoCity\_790. По осі ординат взято параметр value – кількість частинок пилу розміром 2,5 мікрметра, а по осі абсцис – порядковий номер в таблиці. Також там зображені дві таблиці з значеннями, частина з яких використовується при побудові графіку.

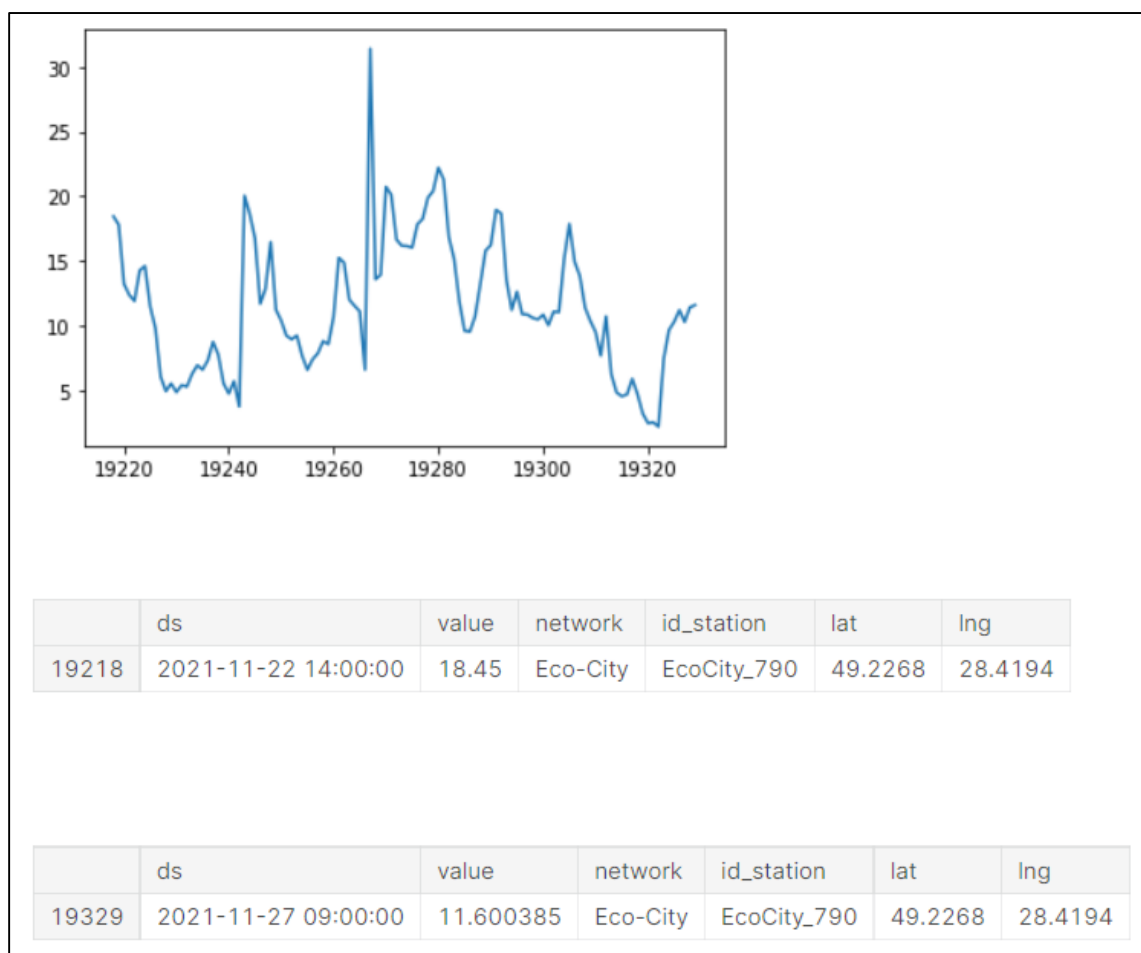


Рисунок 3.41 – Візуалізація значень з поста EcoCity\_790

На рисунку 3.42 зображена візуалізація значень з поста EcoCity\_848. По осі ординат взято параметр value – кількість частинок пилу розміром 2,5 мікрметра, а по осі абсцис – порядковий номер в таблиці. Також там зображені дві таблиці з значеннями, частина з яких використовується при побудові графіку.

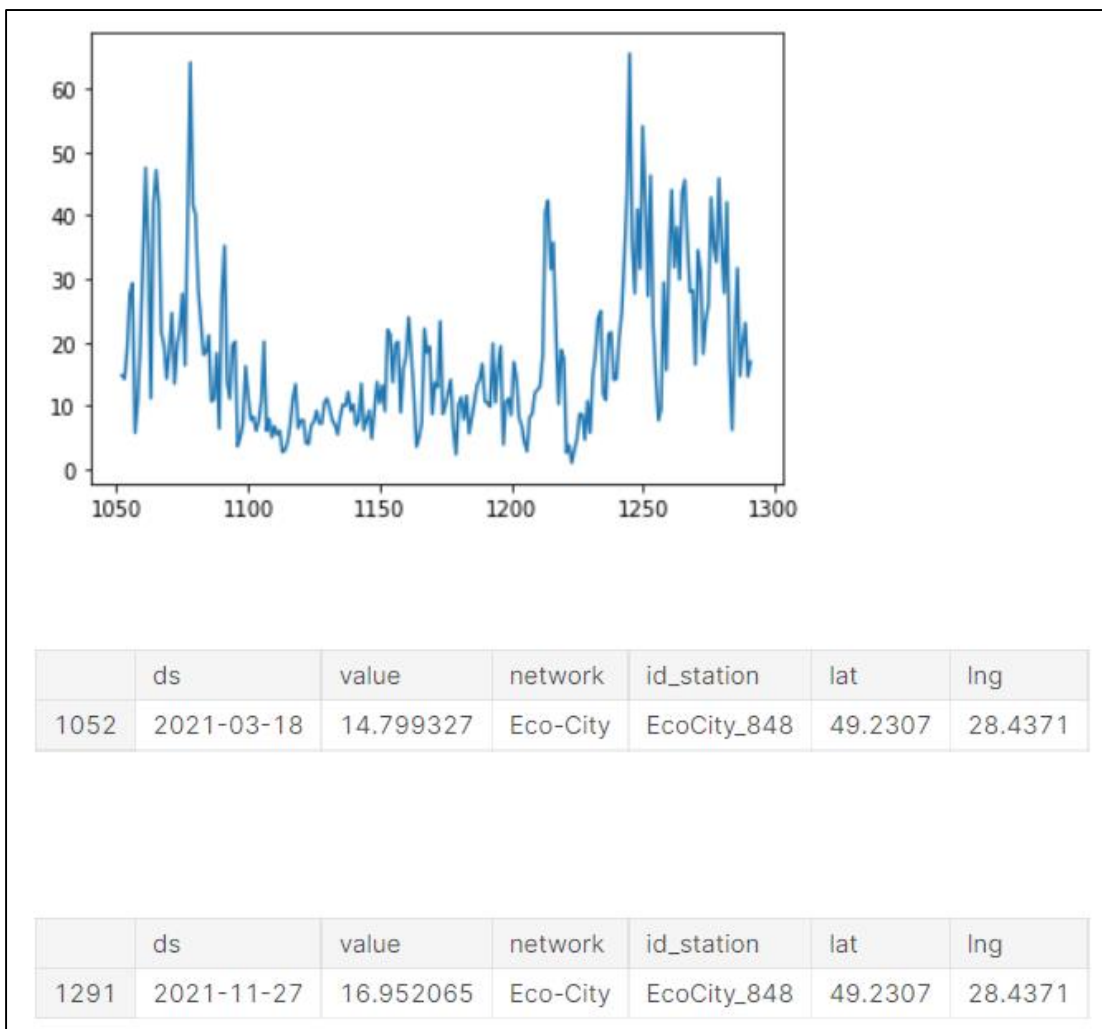


Рисунок 3.42 – Візуалізація значень з поста EcoCity\_848

На рисунку 3.43 зображена візуалізація значень з поста EcoCity\_1315. По осі ординат взято параметр value – кількість частинок пилу розміром 2,5 мікрметра, а по осі абсцис – порядковий номер в таблиці. Також там зображені дві таблиці з значеннями, частина з яких використовується при побудові графіку.



Рисунок 3.43 – Візуалізація значень з поста EcoCity\_1315

На рисунку 3.44 зображена візуалізація значень з поста EcoCity\_1183. По осі ординат взято параметр value – кількість частинок пилу розміром 2,5 мікрметра, а по осі абсцис – порядковий номер в таблиці. Також там зображені дві таблиці з значеннями, частина з яких використовується при побудові графіку.

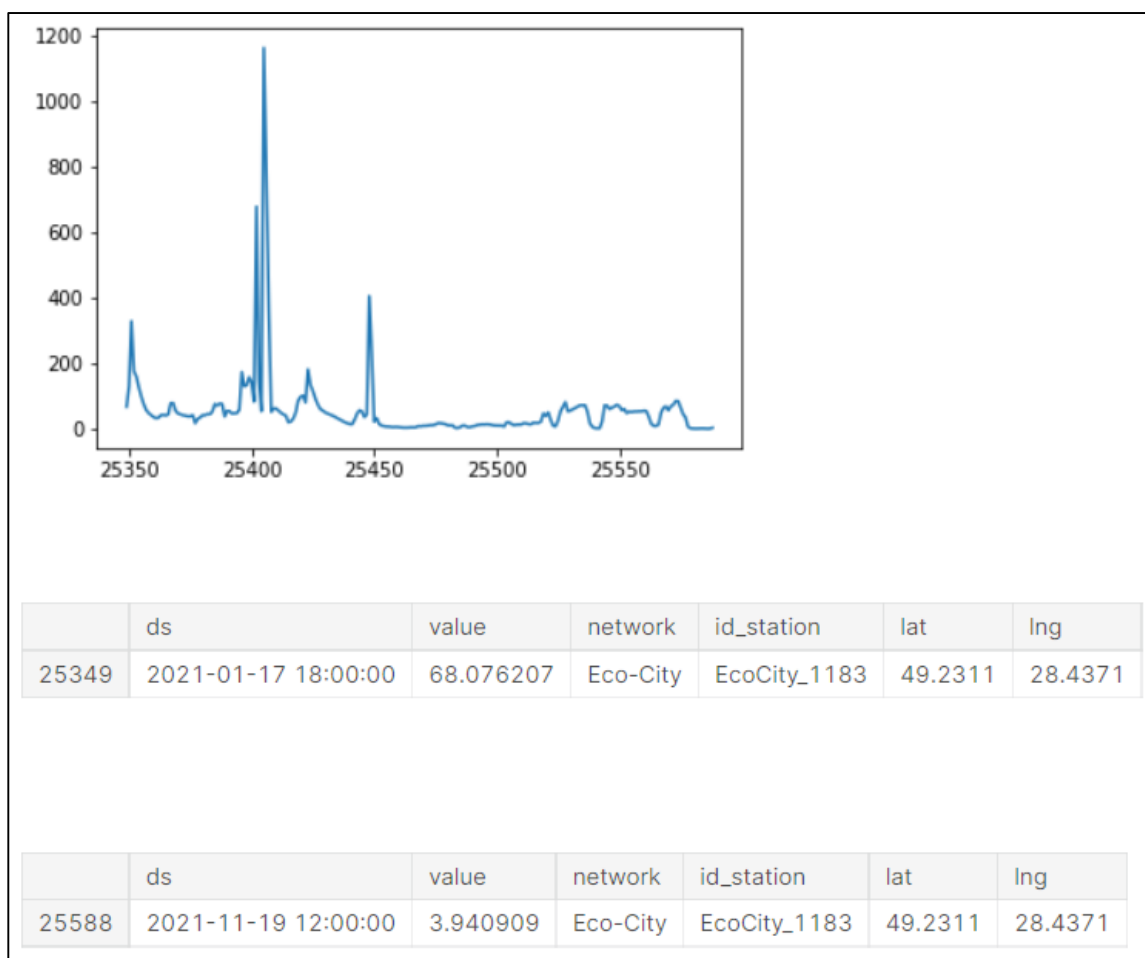


Рисунок 3.44 – Візуалізація значень з поста EcoCity\_1183

На рисунку 3.45 зображена візуалізація значень з поста EcoCity\_761. По осі ординат взято параметр value – кількість частинок пилу розміром 2,5 мікрметра, а по осі абсцис – порядковий номер в таблиці. Також там зображені дві таблиці з значеннями, частина з яких використовується при побудові графіку.

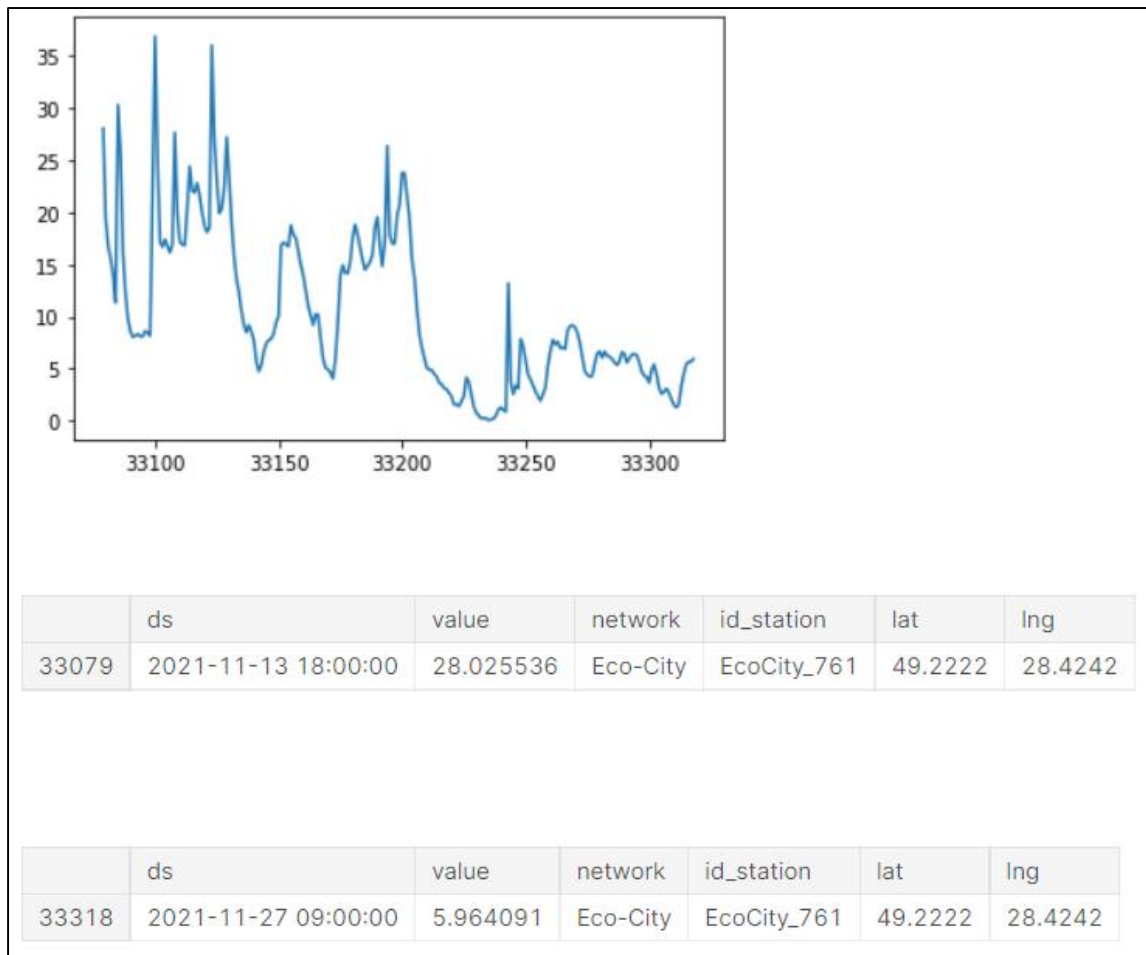


Рисунок 3.45 – Візуалізація значень з поста EcoCity\_761

Проаналізувавши графіки, наведені вище, можна зробити висновок, що найбільше нам підходять дані з таких постів: EcoCity\_337, EcoCity\_848, EcoCity\_1315.

Тепер, перевіримо дані з цих постів, на наявність аномальних даних, це необхідно для створення коректного прогнозу.

На рисунку 3.46 зображено графік з позначенням аномальних значень з поста EcoCity\_337. По осі ординат взято параметр value – значення, а по осі абсцис – дата.



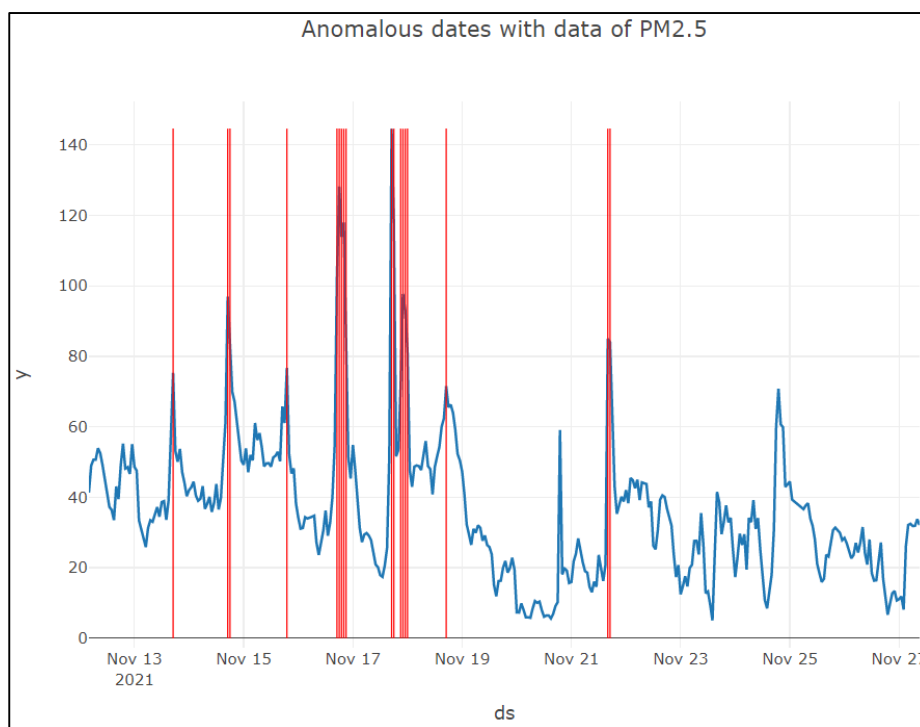


Рисунок 3.46 – Графік з аномальними даними з поста EcoCity\_337

На рисунку 3.47 зображено графік з позначенням аномальних значень з поста EcoCity\_848. По осі ординат взято параметр value – значення, а по осі абсцис – дата.

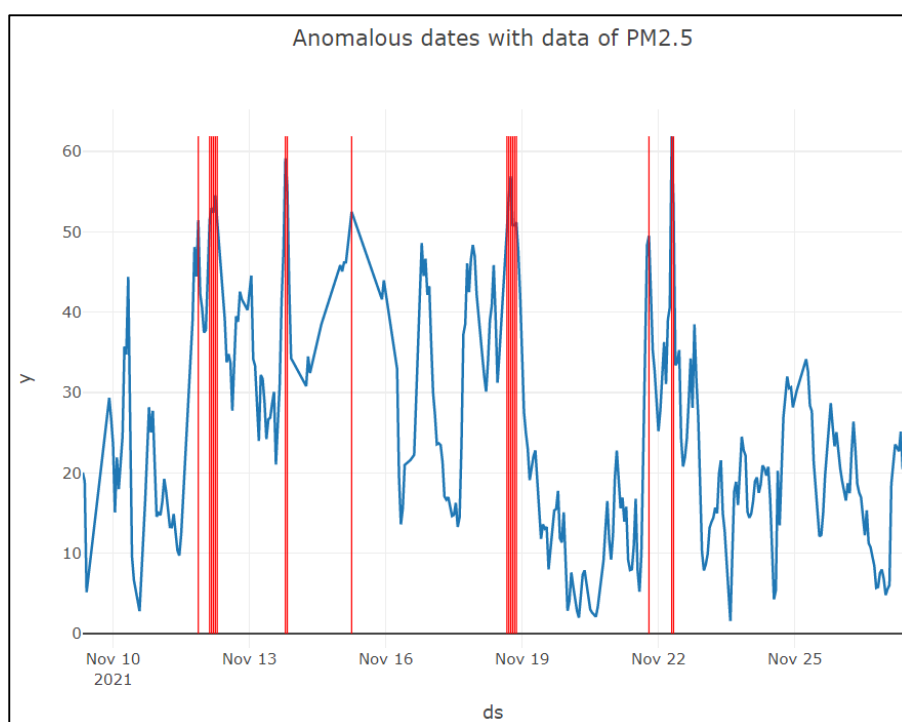


Рисунок 3.47 – Графік з аномальними даними з поста EcoCity\_848

Аналогічно на рисунку 3.48 зображено графік з позначенням аномальних значень з поста EcoCity\_1315. По осі ординат взято параметр value – значення, а по осі абсцис – дата.

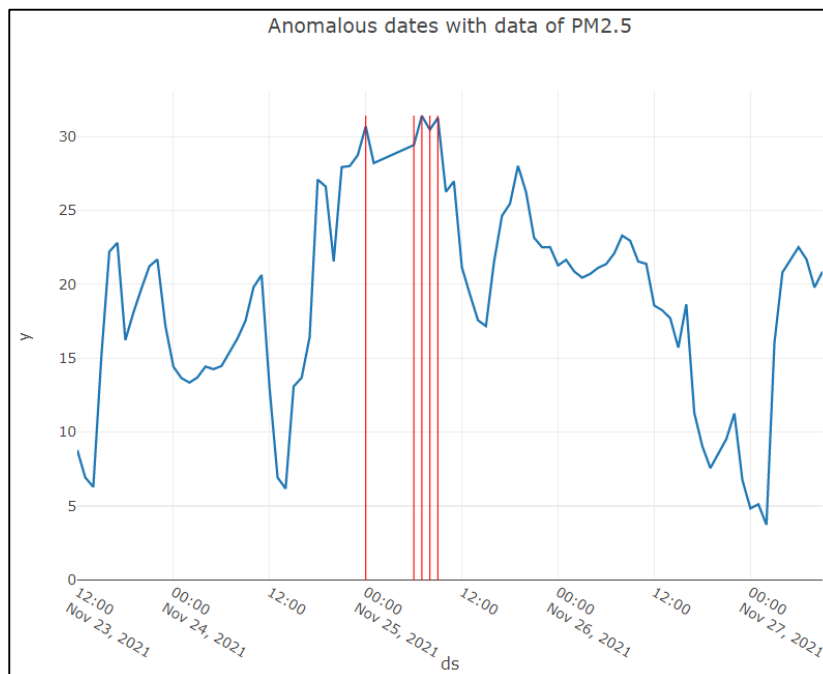


Рисунок 3.48 – Графік з аномальними даними з поста EcoCity\_1315

Далі було побудовано графіки з прогнозом кількості пилу в атмосферному повітрі на наступні дві години, для кожного з цих постів (рис. 3.49– 3.51). І також був побудований тренд до них, що зображено на рисунках 3.52– 3.54. По осі ординат показана кількість частинок пилу розміром 2,5 мікрметра, а по осі абсцис – дата.

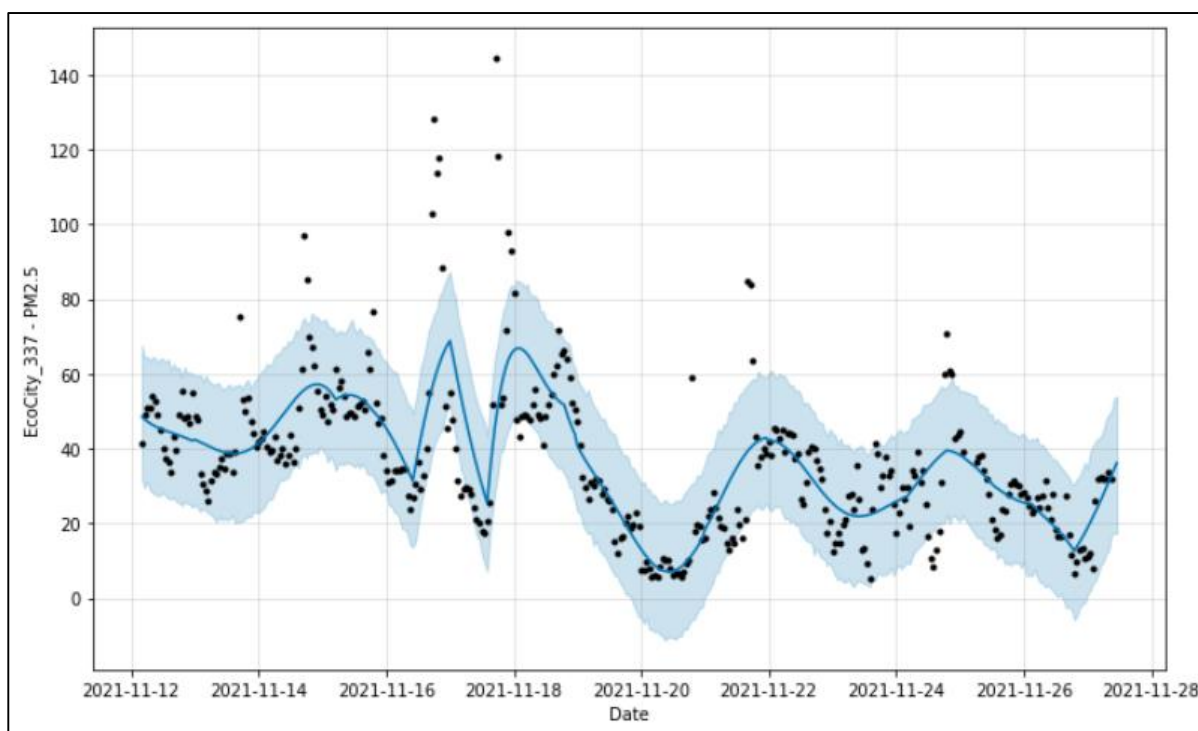


Рисунок 3.49 – Графік з прогнозом кількості пилу для поста EcoCity\_337

Відносна похибка для посту EcoCity\_337 становить 8,95 – 9,93%.

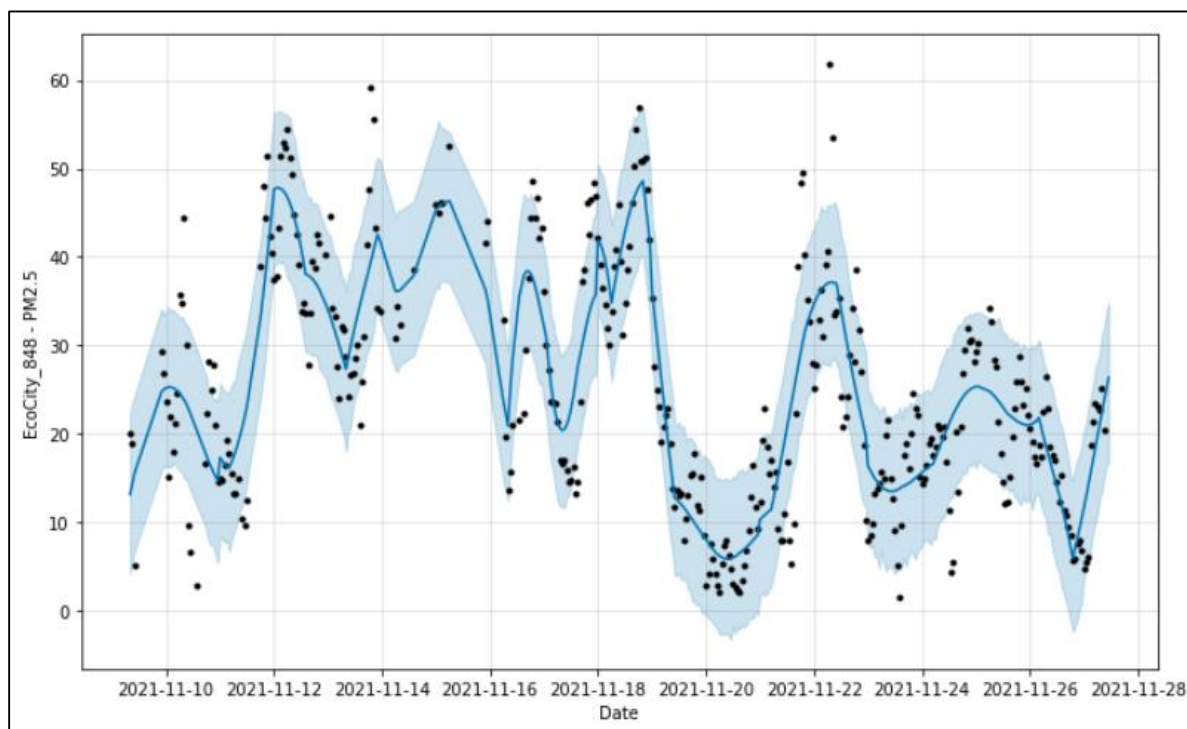


Рисунок 3.50 – Графік з прогнозом кількості пилу для поста EcoCity\_848

Відносна похибка для посту EcoCity\_848 становить 5,27 – 6,2%.

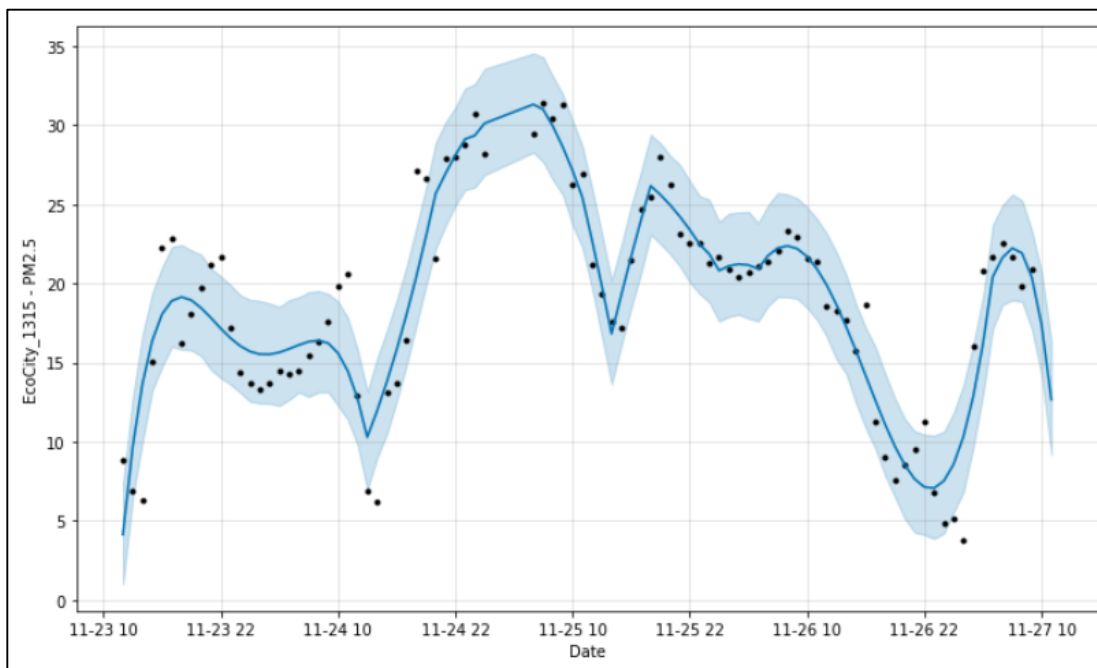


Рисунок 3.51 – Графік з прогнозом кількості пилу для поста EcoCity\_1315

Відносна похибка для посту EcoCity\_1315 становить 1,84 – 2,9%.

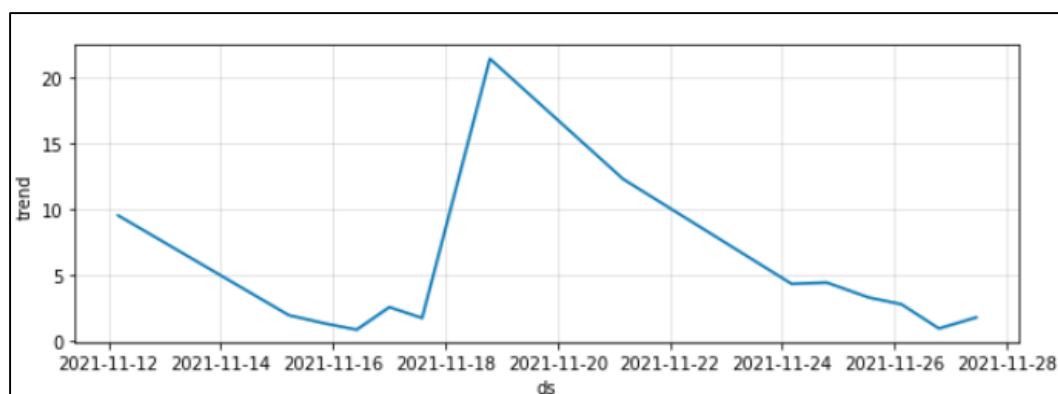


Рисунок 3.52 – Графік з трендом для поста EcoCity\_337

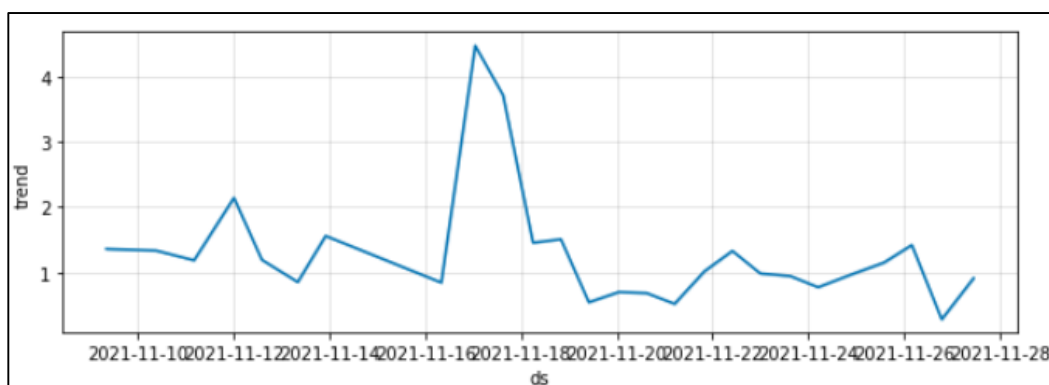


Рисунок 3.53 – Графік з трендом для поста EcoCity\_848

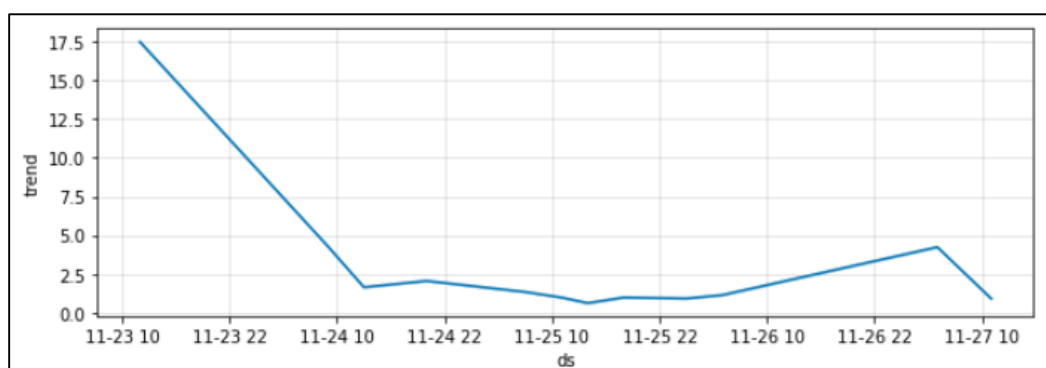


Рисунок 3.54 – Графік з трендом для поста EcoCity\_1315

### 3.3 Висновки

У даному розділі було проведено аналіз і підготовка даних для коректної роботи прогнозу. А також був створений сам прогноз, який показав поступове зменшення вмісту пилу в повітрі.

Проаналізувавши графіки, можна зробити висновок, що Prophet працює правильно та його прогноз досить точний. Також проаналізувавши відносні похибки між щогодинним і щоденним прогнозом, можна зробити висновок, що для якіснішого результату, варто використовувати щоденний прогноз.

## 4 ЕКОНОМІЧНА ЧАСТИНА

### 4.1 Оцінювання комерційного потенціалу розробки

Проведення комерційного та технологічного аудиту полягає в оцінюванні комерційного потенціалу інформаційної технології аналізу та прогнозування якості атмосферного повітря у м. Вінниці за даними громадського моніторингу.

Для отримання об'єктивної оцінки щодо комерційного потенціалу розробки було залучено 3-х незалежних експертів Вінницького національного технічного університету з кафедри системного аналізу та інформаційних технологій: к.т.н., доц. Козачко О.М., к.т.н., доц. Крижановський Є. М., к.т.н., доц. Варчук І. В. Експерти проводили оцінку комерційного потенціалу згідно таблиці 4.1 [22] за 12 критеріями використовуючи п'ятибальну шкалу.

Таблиця 4.1 – Рекомендовані критерії оцінювання комерційного потенціалу розробки та їх можлива бальна оцінка

Критерії оцінювання та бали (за 5-ти бальною шкалою)					
Кри-терій	0	1	2	3	4
Технічна здійсненність концепції:					
1	Достовірність концепції не підтверджена	Концепція підтверджена експертними висновками	Концепція підтверджена розрахунками	Концепція перевірена на практиці	Перевірено роботоздатність продукту в реальних умовах
Ринкові переваги (недоліки):					
2	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 ринку
3	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

## Продовження таблиці 4.1

Критерії оцінювання та бали (за 5-ти бальною шкалою)					
Кри- тері й	0	1	2	3	4
4	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
5	Експлуатаційні 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					
6	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
7	Активна конкуренція великих компаній на ринку	Активна конкуренція	Помірна конкуренція	Незначна конкуренція	Конкуренція немає
Практична здійсненність					
8	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
9	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
10	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

## Продовження таблиці 4.1

Критерії оцінювання та бали (за 5-ти бальною шкалою)					
Кри- тері й	0	1	2	3	4
11	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
12	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

В таблиці 4.2 наведено результати оцінювання експертами комерційного потенціалу розробки.

Таблиця 4.2 – Результати оцінювання комерційного потенціалу розробки

Критерії	Прізвище, ініціали, посада експерта		
	Козачко О.М.	Крижановський Є.М.	Варчук І.В.
	Бали, виставлені експертами:		
1	3	2	3
2	3	2	2
3	2	4	3
4	4	4	3
5	2	3	3
6	1	2	2
7	2	1	1
8	4	3	3
9	1	3	1
10	3	4	3
11	4	4	4
12	2	3	3
Сума балів	СБ <sub>1</sub> =31	СБ <sub>2</sub> =34	СБ <sub>3</sub> =34
Середньоарифметична сума балів $\overline{СБ}$	$\overline{СБ} = \frac{\sum_{i=1}^3 СБ_i}{3} = \frac{31 + 34 + 34}{3} = 33$		



На основі експертних оцінок середньоарифметична сума балів дорівнює 33. Згідно таблиці 4.3 вважається, що рівень комерційного потенціалу проведених досліджень щодо впровадження програмного засобу для прогнозування якості повітря є вище середнього.

Таблиця 4.3 – Рівні комерційного потенціалу розробки

Середньоарифметична сума балів СБ, розрахована на основі висновків експертів	Рівень комерційного потенціалу розробки
0-10	Низький
11-20	Нижче середнього
21-30	Середній
31-40	Вище середнього
41-48	Високий

Інформаційні технології аналізу та прогнозування якості атмосферного повітря у м. Вінниці за даними громадського моніторингу будуть цікаві міській громаді, а також іншим містам у яких є потреба оцінити якість повітря та прийняти рішення щодо покращення екологічної ситуації в своєму регіоні.

#### 4.2 Прогнозування витрат на виконання науково-дослідної роботи

Витрати, пов'язані з проведенням науково-дослідної роботи групуються за такими статтями: витрати на оплату праці, витрати на соціальні заходи, матеріали, паливо та енергія для науково-виробничих цілей, витрати на службові відрядження, програмне забезпечення для наукових робіт, інші витрати, накладні витрати.

1. Основна заробітна плата кожного із дослідників  $Z_0$ , якщо вони працюють в наукових установах бюджетної сфери визначається за формулою:

$$Z_0 = \frac{M}{T_p} * t \text{ (грн)}, \quad (4.1)$$

де  $M$  – місячний посадовий оклад конкретного розробника (інженера, дослідника, науковця тощо), грн.;

$T_p$  – число робочих днів в місяці; приблизно  $T_p \approx 21...23$  дні;

$t$  – число робочих днів роботи дослідника.

Для розробки програми, яка оцінює якість повітря необхідно залучити програміста з посадовим окладом 12000 грн. Кількість робочих днів у місяці складає 30, а кількість робочих днів програміста складає 22. Зведемо сумарні розрахунки до таблиця 4.4.

Таблиця 4.4 – Заробітна плата дослідника в науковій установі бюджетної сфери

Найменування посади	Місячний посадовий оклад, грн.	Оплата за робочий день, грн.	Число днів роботи	Витрати на заробітну плату грн.
Керівник	15000	681,8	5	3409
Програміст	12000	545,5	30	16364
Всього				19773

## 2. Розрахунок додаткової заробітної плати робітників

Додаткова заробітна плата  $Z_d$  всіх розробників та робітників, які приймали участь в розробці нового технічного рішення розраховується як 10 - 12 % від основної заробітної плати робітників.

На даному підприємстві додаткова заробітна плата начисляється в розмірі 10% від основної заробітної плати.

$$Z_d = (Z_o + Z_p) * \frac{N_{\text{дод}}}{100\%}, \quad (4.2)$$

$$Z_d = 0,11 * 19773 = 2175 \text{ (грн).}$$

3. Нарахування на заробітну плату  $N_{3П}$  дослідників та робітників, які брали участь у виконанні даного етапу роботи, розраховуються за формулою:

$$H_{3П} = (Z_o + Z_d) * \frac{\beta}{100} \text{ (грн)}, \quad (4.3)$$

де  $Z_o$  – основна заробітна плата розробників, грн.;

$Z_d$  – додаткова заробітна плата всіх розробників та робітників, грн.;

$\beta$  – ставка єдиного внеску на загальнообов'язкове державне соціальне страхування, % .

Дана діяльність відноситься до бюджетної сфери, тому ставка єдиного внеску на загальнообов'язкове державне соціальне страхування буде складати 22%, тоді:

$$H_{3П} = (19773 + 2175) * \frac{22}{100} = 4828,5 \text{ (грн)}.$$

4. Витрати на комплектуючі вироби (таблиця 4.5), які використовують при виготовленні одиниці продукції, розраховуються, згідно їх номенклатури, за формулою:

$$K = \sum_{i=1}^n H_i * C_i * K_i, \quad (4.5)$$

де  $H_i$  – кількість комплектуючих  $i$ -го виду, шт.;

$C_i$  – покупна ціна комплектуючих  $i$ -го найменування, грн.;

$K_i$  – коефіцієнт транспортних витрат (1,1...1,15).

Таблиця 4.5 – Комплектуючі, що використані на розробку

Найменування матеріалу	Ціна за одиницю, грн.	Витрачено	Вартість витраченого матеріалу, грн.
Папір	135	1	135
Ручка	15	1	15
CD-диск	15	1	15
Флешка	150	1	150
Блокнот	30	1	30
Всього			345

З врахуванням коефіцієнта транспортування	379,5
---	-------

5. Програмне забезпечення для наукової роботи включає витрати на розробку та придбання спеціальних програмних засобів і програмного забезпечення необхідного для проведення дослідження.

Для написання магістерської роботи використовувалися інтернет середовище Kaggle, які є безкоштовними.

6. Амортизація обладнання, комп'ютерів та приміщень, які використовувались під час виконання даного етапу роботи

Дані відрахування розраховують по кожному виду обладнання, приміщенням тощо.

$$A = \frac{Ц \cdot T}{T_{\text{кор}} \cdot 12}, \quad (4.6)$$

де Ц – балансова вартість даного виду обладнання (приміщень), грн.;

$T_{\text{кор}}$  – час користування;

T – термін використання обладнання (приміщень), цілі місяці.

Згідно пункту 137.3.3 Податкового кодекса амортизація нараховується на основні засоби вартістю понад 2500 грн. В нашому випадку для написання магістерської роботи використовувався персональний комп'ютер вартістю 25000 грн.

$$A = \frac{25000 \cdot 1}{2 \cdot 12} = 1041,67.$$

7. До статті «Паливо та енергія для науково-виробничих цілей» відносяться витрати на всі види палива й енергії, що безпосередньо використовуються з технологічною метою на проведення досліджень.

$$B_e = \sum_{i=1}^n \frac{W_{yt} \cdot t_i \cdot Ц_e \cdot K_{\text{впі}}}{\eta_i}, \quad (4.7)$$

де  $W_{yt}$  – встановлена потужність обладнання на певному етапі розробки, кВт;

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

$C_e$  – вартість 1 кВт-години електроенергії, грн;

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

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

Для написання магістерської роботи використовується персональний комп'ютер для якого розраховуємо витрати на електроенергію.

$$V_e = \frac{0,3 \cdot 170 \cdot 4,1 \cdot 0,5}{0,8} = 130,69.$$

Витрати на службові відрядження, витрати на роботи, які виконують сторонні підприємства, установи, організації та інші витрати в нашому дослідженні не враховуються оскільки їх не було.

Накладні (загальновиробничі) витрати  $V_{нзв}$  охоплюють: витрати на управління організацією, оплата службових відряджень, витрати на утримання, ремонт та експлуатацію основних засобів, витрати на опалення, освітлення, водопостачання, охорону праці тощо. Накладні (загальновиробничі) витрати  $V_{нзв}$  можна прийняти як (100...150)% від суми основної заробітної плати розробників та робітників, які виконували дану МКНР, тобто:

$$V_{нзв} = (Z_o + Z_p) \cdot \frac{N_{нзв}}{100\%}, \quad (4.8)$$

де  $N_{нзв}$  – норма нарахування за статтею «Інші витрати».

$$V_{нзв} = 19773 \cdot \frac{100}{100\%} = 19773 \text{ грн.}$$

Сума всіх попередніх статей витрат дає витрати, які безпосередньо стосуються даного розділу МКНР:

$$V = 19773 + 2175 + 4828,5 + 379,5 + 1041,67 + 130,69 + 19773 = 48100,8.$$

Прогнозування загальних втрат  $ZB$  на виконання та впровадження результатів виконаної МКНР здійснюється за формулою:

$$ZB = \frac{V}{\eta}, \quad (4.9)$$

де  $\eta$  – коефіцієнт, який характеризує стадію виконання даної НДР.

Оскільки, робота знаходиться на стадії науково-дослідних робіт, то коефіцієнт  $\beta = 0,9$ .

Звідси:

$$ЗВ = \frac{48100,8}{0,9} = 53445,44 \text{ грн.}$$

#### 4.3 Розрахунок економічної ефективності науково-технічної розробки

У даному підрозділі кількісно спрогнозуємо, яку вигоду, зиск можна отримати у майбутньому від впровадження результатів виконаної наукової роботи. Розрахуємо збільшення чистого прибутку підприємства  $\Delta\Pi_i$ , для кожного із років, протягом яких очікується отримання позитивних результатів від впровадження розробки, за формулою:

$$\Delta\Pi_i = \sum_1^n (\Delta\Pi_0 * N * \Pi_0 * \Delta N)_i * \lambda * \rho * \left(1 - \frac{v}{100}\right), \quad (4.10)$$

де  $\Delta\Pi_0$  – покращення основного оціночного показника від впровадження результатів розробки у даному році.

$N$  – основний кількісний показник, який визначає діяльність підприємства у даному році до впровадження результатів наукової розробки;

$\Delta N$  – покращення основного кількісного показника діяльності підприємства від впровадження результатів розробки:

$\Pi_0$  – основний оціночний показник, який визначає діяльність підприємства у даному році після впровадження результатів наукової розробки;

$n$  – кількість років, протягом яких очікується отримання позитивних результатів від впровадження розробки:

$\lambda$  – коефіцієнт, який враховує сплату податку на додану вартість. Ставка податку на додану вартість дорівнює 20%, а коефіцієнт  $\lambda = 0,8333$ .

$\rho$  – коефіцієнт, який враховує рентабельність продукту.  $\rho = 0,25$ ;

$x$  – ставка податку на прибуток. У 2021 році – 18%.

Припустимо, що ціна за програмний продукт по оцінці якості повітря зросте на 500 грн. Кількість одиниць реалізованої продукції також збільшиться: протягом першого року на 45 шт., протягом другого року – на 35 шт., протягом третього року на 25 шт. Реалізація продукції до впровадження розробки складала 1 шт., а її ціна до складає 9000 грн. Розрахуємо прибуток, яке отримає підприємство протягом трьох років.

$$\Delta\Pi_1 = [500 \cdot 1 + (9000 + 500) \cdot 45] \cdot 0,833 \cdot 0,25 \cdot \left(1 + \frac{18}{100}\right) = 73113,74 \text{ грн.}$$

$$\begin{aligned} \Delta\Pi_2 &= [500 \cdot 1 + (9000 + 500) \cdot (45 + 35)] \cdot 0,833 \cdot 0,25 \cdot \left(1 + \frac{18}{100}\right) \\ &= 130328,14 \text{ грн.} \end{aligned}$$

$$\begin{aligned} \Delta\Pi_3 &= [500 \cdot 1 + (9000 + 500) \cdot (45 + 35 + 25)] \cdot 0,833 \cdot 0,25 \cdot \left(1 + \frac{18}{100}\right) \\ &= 170899,43 \text{ грн.} \end{aligned}$$

#### 4.4 Розрахунок ефективності вкладених інвестицій та періоду їх окупності

Розрахуємо основні показники, які визначають доцільність фінансування наукової розробки певним інвестором, є абсолютна і відносна ефективність вкладених інвестицій та термін їх окупності.

Розрахуємо величину початкових інвестицій  $PV$ , які потенційний інвестор має вкласти для впровадження і комерціалізації науково-технічної розробки.

$$PV = k_{\text{інв}} \cdot ЗВ, \quad (4.11)$$

$k_{\text{інв}}$  – коефіцієнт, що враховує витрати інвестора на впровадження науково-технічної розробки та її комерціалізацію. Це можуть бути витрати на підготовку приміщень, розробку технологій, навчання персоналу, маркетингові заходи тощо ( $k_{\text{інв}} = 2 \dots 5$ ).

$$PV = 2 \cdot 53445,44 = 106890,69.$$

Розрахуємо абсолютну ефективність вкладених інвестицій  $E_{abc}$  згідно наступної формули:

$$E_{abc} = (ПП - PV), \quad (4.12)$$

де ПП – приведена вартість всіх чистих прибутків, що їх отримає підприємство від реалізації результатів наукової розробки, грн.;

$$ПП = \sum_1^T \frac{\Delta\Pi_i}{(1+\tau)^t}, \quad (4.13)$$

де  $\Delta\Pi_i$  – збільшення чистого прибутку у кожному із років, протягом яких виявляються результати виконаної та впровадженої НДДКР, грн.;

$T$  – період часу, протягом якого виявляються результати впровадженої НДДКР, роки;

$\tau$  – ставка дисконтування, за яку можна взяти щорічний прогнозований рівень інфляції в країні; для України цей показник знаходиться на рівні 0,2;

$t$  – період часу (в роках).

$$ПП = \frac{73113,74}{(1 + 0,2)^1} + \frac{130328,14}{(1 + 0,2)^2} + \frac{170899,43}{(1 + 0,2)^3} = 250793,91 \text{ грн.}$$

$$E_{abc} = (250793,91 - 109301,33) = 141492,58 \text{ грн.}$$

Оскільки  $E_{abc} > 0$ , то вкладання коштів на виконання та впровадження результатів НДДКР може бути доцільним.

Розрахуємо відносну (щорічну) ефективність вкладених в наукову розробку інвестицій  $E_v$ . Для цього користуються формулою:

$$E_v = \sqrt[T_{ж}] {1 + \frac{E_{abc}}{PV}} - 1, \quad (4.14)$$



де  $T_{ж}$  – життєвий цикл наукової розробки, роки.

$$E_B = \sqrt[3]{1 + \frac{141492,58}{106890,69}} - 1 = 0,55 = 55\%.$$

Визначимо мінімальну ставку дисконтування, яка у загальному вигляді визначається за формулою:

$$\tau = d + f, \quad (4.15)$$

де  $d$  – середньозважена ставка за депозитними операціями в комерційних банках; в 2021 році в Україні  $d = (0,14 \dots 0,2)$ ;

$f$  – показник, що характеризує ризикованість вкладень; зазвичай, величина  $f = (0,05 \dots 0,1)$ .

$$\tau_{min} = 0,18 + 0,5 = 0,23.$$

Так як  $E_B > \tau_{min}$  то інвестор може бути зацікавлений у фінансуванні даної наукової розробки.

Розрахуємо термін окупності вкладених у реалізацію наукового проекту інвестицій за формулою:

$$T_{ок} = \frac{1}{E_B}, \quad (4.16)$$

$$T_{ок} = \frac{1}{0,55} = 1,8 \text{ роки.}$$

Так як  $T_{ок} \leq 3 \dots 5$ -ти років, то фінансування даної наукової розробки в принципі є доцільним.

#### 4.5 Висновки

Було проведено оцінку комерційного потенціалу інформаційної технології аналізу та прогнозування якості атмосферного повітря у м. Вінниці за даними громадського моніторингу, який є на вище середньому рівні.

Прогнозування витрат на виконання науково-дослідної роботи по кожній з статей витрат складе 48100,8 грн. Загальна ж величина витрат на виконання та впровадження результатів даної НДР буде складати 53445,44 грн.

Вкладені інвестиції в даний проект окупляться через 1,8 роки, приведена вартість всіх чистих прибутків, що їх отримає підприємство від реалізації результатів наукової розробки склала 250793,91 грн., абсолютна ефективність вкладених інвестицій склала 141492,58 грн.

## ВИСНОВКИ

В роботі була розроблена інформаційна технологія аналізу та прогнозування якості атмосферного повітря у м. Вінниці за даними громадського моніторингу. Таким чином, мету роботи було досягнуто.

В ході виконання роботи було вирішено всі поставлені завдання, а саме:

- розвідувальний аналіз даних, отриманих з постів розміщених у місті Вінниця за певний період;
- вибір середовища та оптимальних технологій;
- підготовка набору даних;
- пошук аномальних даних;
- здійснення прогнозу даних.

У першому розділі, було визначено суть і види забруднень атмосферного повітря, вказано, що таке пости спостереження та які вони бувають. Також, розглянуто сучасний стан моніторингу атмосферного повітря у місті Вінниця, визначено основні проблеми та проекти, що займаються встановленням постів по місту.

У другому розділі, було здійснено вибір оптимальних інформаційних технологій та мови програмування, оглянуто вхідні дані, та проаналізовано показники, що фіксуються постами моніторингу якості атмосферного повітря.

В третьому розділі, було проведено аналіз і підготовка даних для коректної роботи прогнозу. А також був створений сам прогноз, який показав поступове зменшення вмісту пилу в повітрі.

Проаналізувавши результати, можна зробити висновок, що Prophet працює правильно та його прогноз досить точний. Також проаналізувавши відносні похибки між щогодинним і щоденним прогнозом, можна зробити висновок, що для якіснішого результату, варто використовувати щоденний прогноз.

В наступному розділі було проведено оцінку комерційного потенціалу інформаційної технології аналізу та прогнозування якості атмосферного

повітря у місті Вінниці за даними громадського моніторингу, який є на вище середньому рівні.

Прогнозування витрат на виконання науково-дослідної роботи по кожній з статей витрат складе 48100,8 грн. Загальна ж величина витрат на виконання та впровадження результатів даної НДР буде складати 53445,44 грн.

Вкладені інвестиції в даний проект окупляться через 1,8 роки, приведена вартість всіх чистих прибутків, що їх отримає підприємство від реалізації результатів наукової розробки склала 250793,91 грн., абсолютна ефективність вкладених інвестицій склала 141492,58 грн.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Бондарчук А. Р., Мокін В. Б., Мокін О. Б., Інформаційна технологія прогнозування даних про якість атмосферного повітря у м. Вінниці за даними громадського моніторингу. *Всеукраїнська науково-практична інтернет-конференція: «Молодь в науці: дослідження, проблеми, перспективи»*, м. Вінниця, 2021-2022 рр., [Електронний ресурс]. URL: <https://conferences.vntu.edu.ua/index.php/mn/mn2022/paper/view/14243>
2. Мокін В. Б. Розробка імітаційної моделі та комп'ютерної програми для автоматизації процесу еколого-економічного планування управління природокористуванням у регіоні із дотриманням вимог екологічної безпеки. *XIV Міжнародна науково-практична конференція: Сучасні інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях: 5-9 жовтня 2015 р. Київ, 2015. С. 135-139.*
3. Технології обробки та моделювання екологічної та економічної інформації [Електронний навчальний посібник] / В. Б. Мокін, А. В. Поплавський, М. П. Боцула, А. Р. Ящолт. Вінниця : ВНТУ, 2015. 120 с.
4. Боцян В. В., Мокін В. Б. Моніторинг стану забруднення атмосферного повітря міста оксидом вуглецю з використанням мобільної аналітичної інформаційно-вимірювальної системи. *Молодь в технічних науках: дослідження, проблеми, перспективи* : матеріали Міжнародної науково-практичної Інтернет-конференції, 23-26 квітня 2015 року. ВНТУ. Вінниця, 2015. С. 31-33.
5. Frederik Lundh. Python Standard Library./ Frederik Lundh - O'Reilly & Associates, 2001. 304 pp.
6. Мокін В. Б., Ящолт А. Р., Новий підхід до розробки інформаційних моделей автоматизованих систем екоінспекційного контролю. *Зб. матеріалів X Міжнародної науково-технічної конференції «Системний аналіз та інформаційні технології»*. – 20-24 травня 2008 р. К.: Інститут прикладного системного аналізу НТУУ «КПІ», 2008. 385 с.

7. Практична реалізація мобільної аналітичної комп'ютерної системи моніторингу стану атмосферного повітря з підсистемою веб-аналізу та виведення даних на геопортал / В. Б. Мокін, Д. Ю. Дзюняк, К. О. Бондалетов, Г. В. Горячев, В. В. Боцян // *V-й Всеукраїнський з'їзд екологів з міжнародною участю (Екологія/Ecology-2015)*, 23-26 вересня, 2015. Збірник наукових праць. Вінниця: ТОВ «Нілан-ЛТД», 2015. С. 106.

8. Мокін В. Б., Дзюняк Д. Ю., Бондалетов К. О., Технологія оперативного екологічного моніторингу стану повітряного середовища та пересувних джерел його забруднення. *XIV Міжнародна науково-практична конференція: Сучасні інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях: 5-9 жовтня 2015 р. К., 2015. С. 126-130.*

9. Автоматизована система екоінспекційного контролю стану забруднення довкілля України та викидів, скидів і відходів "Екоінспектор". Частина I. Підсистема "Викиди.": методичний посібник. Вид. 2-ге, змінене та доповнене / [В. Б. Мокін, Б. І. Мокін, М. М. Костров та ін.]; під ред. Г. В. Горячева. ВНТУ, Держ. екологічна інспекція Мінприроди України. Вінниця : УНІВЕРСУМ-Вінниця, 2009. 192 с.

10. Метод оцінювання параметрів стаціонарного джерела викиду на основі моделі Гауса за даними оперативного моніторингу зони розсіювання / В. Б. Мокін, Д. Ю. Дзюняк // *Математичне моделювання в економіці*. 2016. № 3-4 (7). С. 27-35.

11. Information measuring systems with mobile devices for identification of air pollution parameters caused by transport / Vitalii B. Mokin ; Georgii V. Goriachev ; Dmytro Y. Dziuniak ; Konstantin O. Bondaletov ; Serhii O. Zhukov ; Mariusz Duk ; Saltanat Sailarbek // *Proc. SPIE 10031, Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2016*, 1003128 (September 28, 2016), 8 pages; doi:10.1117/12.2249202 (Scopus, USA). [Електронний ресурс]. URL:

<http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=2564871>

12. Метод та технологія моніторингу стану атмосферного повітря за допомогою універсальної інформаційно-вимірювальної системи з використанням мобільних пристроїв / В. Б. Мокін, Д. Ю. Дзюняк, К. О. Бондалетов, В. В. Олійник // Наукові праці Вінницького національного технічного університету [Електронний ресурс]. 2015. № 4. URL:

<http://praci.vntu.edu.ua/index.php/praci/article/view/456/454>

13. Технологія оцінювання параметрів викидів зі стаціонарних джерел за даними оперативного моніторингу / В. Мокін, Д. Дзюняк // *Матеріали III Міжнародної науково-практичної конференції "Summer InfoCom Advanced Solutions 2016"*, 1-2 грудня 2016 року, м. Київ. Київ: ТОВ «Інженеринг», 2016. С. 66-68.

14. A. Sweigart. Core Python Applications Programming./ A. Sweigart. - Albert\Sweigart, 2012. 888 с.

15. Роберт Кабаков. R в действии = R in Action./ Роберт Кабаков. ДМК-Пресс, 2014. 588 с.

16. Хэдли Уикем. Язык R в задачах науки о данных: импорт, подготовка, обработка, визуализация и моделирование данных = R for Data Science: Visualize, Model, Transform, Tidy, and Import Data./ Хэдли Уикем, Гарретт Гроулмунд. Вильямс, 2017. 592 с.

17. Підвищення точності моделювання забруднення атмосферного повітря міста з використанням мобільних інформаційно-вимірювальних систем / В. Б. Мокін, Г. В. Горячев, Д. Ю. Дзюняк, К. О. Бондалетов // *XLV Регіональна науково-технічна конференція професорсько-викладацького складу, співробітників та студентів університету з участю працівників науково-дослідних організацій та інженерно-технічних працівників підприємств м. Вінниці та області, електронне наукове видання матеріалів конференції*, м. Вінниця, 23-24 березня 2016. – [Електронний ресурс]. URL: <http://conferences.vntu.edu.ua/index.php/all-ebmd/all-ebmd-2016/paper/view/465/866>

18. Мокін В. Б., Яцолт А. Р., Новий метод розробки інформаційних моделей систем екологічного контролю. *Міжнародний науково-технічний журнал «Інформаційні технології та комп'ютерна інженерія»*. 2008. № 1(11). С. 37–43.
19. Alan Gauld. *Learn to Program Using Python: A Tutorial for Hobbyists, Self-Starters, and Those Who Want to Learn the Art of Programming*. Addison-Wesley Professional./ Alan Gauld - O'Reilly & Associates, 2001. 288 pp.
20. Марк Саммерфилд. *Python на практике*. Перевод с английского./ Марк Саммерфилд - М.: ДМК Пресс, 2014. 338 с.
21. Марк Лутц. *Программирование на Python, 4-е издание*. Пер. с англ./ Марк Лутц - СПб.: Символ-Плюс, 2011. 992 с.
22. Методичні вказівки до виконання економічної частини магістерських кваліфікаційних робіт / В. О. Козловський, О. Й. Лесько, В. В. Кавецький. Вінниця : ВНТУ, 2021. 42 с.



Додаток А  
(обов'язковий)

Міністерство освіти і науки України  
Вінницький національний технічний університет  
Факультет комп'ютерних систем і автоматики

ЗАТВЕРДЖУЮ

Завідувач кафедри САІТ

\_\_\_\_\_ д.т.н., проф. Мокін В. Б.

«\_\_» \_\_\_\_\_ 2021 р.

ТЕХНІЧНЕ ЗАВДАННЯ

на магістерську кваліфікаційну роботу

«ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ  
ЯКОСТІ АТМОСФЕРНОГО ПОВІТРЯ У МІСТІ ВІННИЦІ ЗА ДАНИМИ  
ГРОМАДСЬКОГО МОНІТОРИНГУ (КОМПЛЕКСНА МКР) ЧАСТИНА І.  
ПРОГНОЗУВАННЯ ДАНИХ»

08-53.МКР.001.02.000.ТЗ

Керівник: д.т.н., проф.

\_\_\_\_\_ Мокін О. Б.

«\_\_» \_\_\_\_\_ 2021 р.

Розробив: студент гр. 2ІСТ-20м

\_\_\_\_\_ Бондарчук А. Р.

«\_\_» \_\_\_\_\_ 2021 р.

Вінниця 2021

### 1. Підстава для проведення робіт

Підставою для виконання роботи є наказ № \_\_ по ВНТУ від «\_\_» \_\_\_\_\_ 2021 р., та індивідуальне завдання на МКР, затверджене протоколом № \_\_ засідання кафедри САІТ від «\_\_» \_\_\_\_\_ 2021 р.

### 2. Джерела розробки:

- Технології обробки та моделювання екологічної та економічної інформації [Електронний навчальний посібник] / В. Б. Мокін, А. В. Поплавський, М. П. Боцула, А. Р. Ящолт. — Вінниця : ВНТУ, 2015. — 120 с.
- Технологія оцінювання параметрів викидів зі стаціонарних джерел за даними оперативного моніторингу / В. Мокін, Д. Дзюняк // *Матеріали III Міжнародної науково-практичної конференції "Summer InfoCom Advanced Solutions 2016"*, м. Київ, 1-2 грудня 2016 року. — Київ: ТОВ «Інженеринг», 2016. — С. 66-68.

### 3. Мета і призначення роботи:

Метою даної роботи є підвищення точності прогнозування якості атмосферного повітря у м. Вінниці за даними громадського моніторингу.

### 4. Вихідні дані для проведення робіт:

Перелік атрибутивних даних про якість повітря, взятих з ресурсів Eco City та SaveEcoBot.

### 5. Методи дослідження:

- розвідувальний аналіз;
- прогнозування даних.

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

1. Аналіз предметної області ..... \_\_\_\_\_ — \_\_\_\_\_
2. Розробка інформаційної технології..... \_\_\_\_\_ — \_\_\_\_\_
3. Реалізація інформаційної технології..... \_\_\_\_\_ — \_\_\_\_\_
4. Оформлення пояснювальної записки..... \_\_\_\_\_ — \_\_\_\_\_

### 7. Очікувані результати та порядок реалізації:

Отримання технології прогнозу якості атмосферного повітря у м. Вінниці.

### 8. Вимоги до розробленої документації

Пояснювальна записка оформлена у відповідності до вимог «Методичних вказівок до виконання та оформлення магістерських кваліфікаційних робіт для студентів спеціальності 126 – «Інформаційні системи та технології» денної форми навчання».

### 9. Порядок приймання роботи

Публічний захист ..... «\_\_» \_\_\_\_\_ 2021 р.

Початок розробки..... «\_\_» \_\_\_\_\_ 2021 р.

Граничні терміни виконання МКР ..... «\_\_» \_\_\_\_\_ 2021 р.

Розробив студент групи 2ІСТ-20м \_\_\_\_\_ Бондарчук А. Р.

## Додаток Б

(обов'язковий)

## Протокол перевірки кваліфікаційної роботи

Назва роботи: «Інформаційна технологія аналізу та прогнозування якості атмосферного повітря у місті Вінниці за даними громадського моніторингу (комплексна МКР) Частина І. Прогнозування даних»

Тип роботи: магістерська кваліфікаційна робота

Підрозділ: кафедра САІТ

Науковий керівник: Мокін О.Б. д.т.н., проф.

## Показники звіту подібності

Unicheck	
Оригінальність	89,2 %
Схожість	10,8 %

## 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 доопрацювання.
- 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.

Заявляю, що ознайомлений з повним звітом подібності, який був згенерований системою щодо роботи

Автор \_\_\_\_\_ Бондарчук А.Р.  
(підпис)

## Опис прийнятого рішення

Робота допускається до захисту \_\_\_\_\_

Особа, відповідальна за перевірку \_\_\_\_\_ Жуков С. О.  
(підпис)

Додаток В  
(довідниковий)

Лістинг програмного коду

```
import numpy as np
import pandas as pd
from scipy.interpolate import Rbf, interp2d
import datetime
import os
import matplotlib.pyplot as plt
import plotly.express as px
import plotly.graph_objects as go
from fbprophet import Prophet
from sklearn.metrics import r2_score, mean_absolute_error
import warnings
warnings.simplefilter('ignore')

N = 30
forecasting_period = 2 # in time_intervals
indicator_names = ['PM2.5', 'PM10']
indicator_name = indicator_names[0]
type_agg='mean' # 'mean' or 'max'
time_interval = 'D' # day
Q = 14 # 2 weeks = 14 days

dataset_files = []
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        dataset_files.append(os.path.join(dirname, filename))
```

```
dataset_files
```

```
stations_about = pd.read_csv('./input/air-quality-
monitoring/saveecobot_city_about_stations.csv', header=0, sep=';')
```

```
stations_about = stations_about[stations_about['locality']=='Vinnytsia
city'].reset_index(drop=True)
```

```
stations_about
```

```
def get_data_for_indicator_of_station_from_saveecobot(stations_about,
indicator_name, num):
```

```
    # Get data for given indicator_name for station in num-th row in the dataframe
saveecobot files
```

```
    # with parameters about stations from the dataframe stations_about
```

```
    # Transform indicator to SaveEcoBot variants
```

```
    if indicator_name=='PM2.5':
```

```
        indicator_name = 'pm25'
```

```
    elif indicator_name=='PM10':
```

```
        indicator_name = 'pm10'
```

```
    # Get codes station
```

```
    id_station_saveecobot = int(stations_about.loc[num,'id_saveecobot'])
```

```
    id_station_ecocity = stations_about.loc[num,'id_ecocity']
```

```
    if not np.isnan(id_station_ecocity):
```

```
        id_station_ecocity = int(id_station_ecocity)
```

```
        id_station = "EcoCity_" + str(id_station_ecocity)
```

```
    else: id_station = "SaveEcoBot_" + str(id_station_saveecobot)
```

```
    #print(num, id_station_saveecobot, id_station_ecocity, id_station)
```

```
    df = pd.read_csv(f"./input/air-quality-
monitoring/data_saveecobot_{id_station_saveecobot}.csv")
```

```

#display(df.head())
df = df[df['indicator_code']==indicator_name]
#display(df.head())
df['ds'] = pd.to_datetime(df['datetime'], format='%Y-%m-%d %H:%M:%S',
errors='ignore')
df = df[['ds', 'value']]
#df = df.dropna().reset_index(drop=True)
df.index = df['ds']
df = df.drop(columns=['ds'])
#display(df.head())

# Data processing - converting data to average or maximum values per given
time_interval
if type_agg == 'mean':
    df = df.resample(time_interval).mean()
else:
    # type_agg == 'max'
    df = df.resample(time_interval).max()
df = df.reset_index(drop=False)
df = df.dropna().reset_index(drop=True)

if df.shape[1] > 1:
    # Resample is successfull
    #display(df.head())
    df['network'] = str(stations_about.loc[num,'network'])
    df['id_station'] = id_station
    df['lat'] = float(stations_about.loc[num,'lat'])
    df['lng'] = float(stations_about.loc[num,'lng'])
    print(f"Number of data for {num}th station #{id_station} is {len(df)}")
    #display(df)

```

```

    #print(df.info())
else:
    print(f>Data for {num}th station #{id_station} is bad")
    df = pd.DataFrame()
return df

%%time

df = pd.DataFrame()

ln = 0
for i in range(len(stations_about)):
    df_i = get_data_for_indicator_of_station_from_saveecobot(stations_about,
indicator_name, i)
    #df_i.info()
    if len(df) > 0:
        #ln += len(df_i)
        #print('\n',ln)
        df = pd.concat([df, df_i], ignore_index=True)
    else: df = df_i

df = df.dropna().reset_index(drop=True)

df

ecocity_stations_about = pd.read_csv('./input/air-quality-monitoring-from-
ecocity/ecocity_about_stations_2021.csv', header=0, sep=';')

ecocity_stations_about

ecocity_stations_about_region =
ecocity_stations_about[ecocity_stations_about['locality']=='Vinnytsia
city'].reset_index(drop=True)

ecocity_stations_about_region['id_ecocity'] =
ecocity_stations_about_region['id_ecocity'].astype('int')

ecocity_stations_about_region

def get_data_for_indicator_of_station_from_ecocity(stations_about,
indicator_name, num):

```

```
# Get data for given indicator_name for station in num-th row in the dataframe
saveecobot files
```

```
# with parameters about stations from the dataframe stations_about
```

```
#id_station_saveecobot = int(stations_about.loc[num,'id_saveecobot'])
```

```
id_station_ecocity = int(stations_about.loc[num,'id_ecocity'])
```

```
# Find file name
```

```
for i in range(len(dataset_files)):
```

```
    if dataset_files[i].find(str(id_station_ecocity))>0:
```

```
        file_name = dataset_files[i]
```

```
df = pd.read_csv(file_name)
```

```
#display(df)
```

```
df = df[df['indicator_name']==indicator_name]
```

```
df['ds'] = pd.to_datetime(df['datetime'], format='%Y-%m-%d %H:%M:%S',
errors='ignore')
```

```
df = df[['ds', 'value']]
```

```
df.index = df['ds']
```

```
df = df.drop(columns=['ds'])
```

```
df = df.resample(time_interval).mean()
```

```
df = df.reset_index(drop=False)
```

```
df = df.dropna().reset_index(drop=True)
```

```
df['network'] = str(stations_about.loc[num,'network'])
```

```
#df['id_station_ecocity'] = id_station_ecocity
```

```
df['id_station'] = "EcoCity_" + str(id_station_ecocity)
```

```
df['lat'] = float(stations_about.loc[num,'lat'])
```

```
df['lng'] = float(stations_about.loc[num,'lng'])
```



```

    #print(f"Number of data for {num}th station #{id_station_saveecobot} in
SaveEcoBot and #{id_station_ecocity} in EcoCity is {len(df)}")

    #display(df)

    return df

linkcode

%%time

df2 = pd.DataFrame()

for i in range(len(ecocity_stations_about_region)):

    df_i =
get_data_for_indicator_of_station_from_ecocity(ecocity_stations_about_region,
indicator_name, i)

    if len(df2) > 0:

        df2 = pd.concat([df2, df_i], ignore_index=True)

    else: df2 = df_i

df2

# Drop data on stations of the EcoCity network from SaveEcoBot
# with datetime which equal datetime of data from EcoCity
if len(df)>0:

    len_before = len(df)

    for id_station in df2['id_station'].unique().tolist():

        print(id_station)

        ds_list = df2[df2['id_station']==id_station]['ds'].tolist()

        df = df.drop(df[(df.id_station == id_station) & (df.ds.isin(ds_list))].index)

    len_after = len(df)

    print(f"Number of data before the dropping duplicates - {len_before}, after -
{len_after}")

def get_data_for_prediction(df, id_station_name, num_last_data=30):

    # Get data for given id_station_name with the last num_last_data data

    #id_station_name = 'EcoCity_848'

```

```

df_i = df[df['id_station']==id_station_name]
df_i = df_i[(len(df_i)-num_last_data):].reset_index(drop=True)
display(df_i['value'].describe())
print(len(df_i))
return df_i

def plot_with_anomalies(df, cols_y_list, cols_y_list_name, dates_x,
anomalous_dates, log_y=False):
    fig = px.line(df, x=dates_x, y=cols_y_list[0],
title=cols_y_list_name[cols_y_list[0]], log_y=log_y, template='gridon',width=800,
height=600)

    y_max = df[cols_y_list[0]].max()
    for i in range(len(cols_y_list)-1):
        fig.add_trace(go.Scatter(x=df[dates_x], y=df[cols_y_list[i+1]], mode='lines',
name=cols_y_list_name[cols_y_list[i+1]]))
        max_i = df[cols_y_list[i+1]].max()
        y_max = max_i if max_i > y_max else y_max

    y_min = min(df[cols_y_list[0]].min(),0)
    for i in range(len(anomalous_dates)):
        anomal_date = anomalous_dates[i]
        #print(anomal_date, y_min, y_max)

        fig.add_shape(dict(type="line", x0=anomal_date, y0=y_min, x1=anomal_date,
y1=y_max, line=dict(color="red", width=1)))
    fig.show()

def get_model_err(df, res, model, id_station_name, forecasting_period,
Q_order, weekly_order, monthly_order, quarterly_order):
    # Data prediction and score calculation

    # Make a forecast

```

```

future = model.make_future_dataframe(periods = forecasting_period,
freq=time_interval)

forecast = model.predict(future)

# Output the prediction for the next time

forecast[['yhat_lower', 'yhat', 'yhat_upper']] = forecast[['yhat_lower', 'yhat',
'yhat_upper']].round(2)

y_pred = forecast['yhat'][:-forecasting_period]

# Calculation r2_score (accuracy of prediction for training data)

r2 = round(r2_score(df['y'], y_pred),2)

mae_err = round(mean_absolute_error(df['y'], y_pred),2)

# Save results

if res.empty:
    num = 0
else: num = len(res)

res.loc[num, 'id_station'] = id_station_name
res.loc[num, 'Q_order'] = Q_order
res.loc[num, 'weekly_order'] = weekly_order
res.loc[num, 'monthly_order'] = monthly_order
res.loc[num, 'quarterly_order'] = quarterly_order
res.loc[num, 'r2_score'] = r2
res.loc[num, 'mae'] = mae_err

print(id_station_name, Q_order, weekly_order, monthly_order, quarterly_order,
r2, mae_err)

#display(res)

# Draw plot of the values with forecasting data

```

```

#label_str = " - ".join([id_station_name, indicator_name, str(Q_order),
str(daily_order), str(weekly_order), str(quarterly_order), str(r2), str(mae_err)])

#figure = model.plot(forecast, xlabel = 'Date', ylabel = label_str)

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

return res

```

In [24]:

```

def Prophet_tuning(df, anomalous, number_steps_in_day, Q_order, weekly_order,
monthly_order, quarterly_order):

```

```

    # Prophet tuning

```

```

    model = Prophet(daily_seasonality=False, weekly_seasonality=False,
yearly_seasonality=False,

```

```

                    changepoint_range=1, changepoint_prior_scale = 0.5,

```

```

                    holidays=anomalous, seasonality_mode = 'multiplicative')

```

```

    if (weekly_order > 0):

```

```

        model.add_seasonality(name='weekly', period=7*number_steps_in_day,
fourier_order=weekly_order, mode = 'multiplicative')\

```

```

    if (Q_order > 0):

```

```

        model.add_seasonality(name='Q_days', period=Q,
fourier_order=Q_order, mode = 'multiplicative')

```

```

    if (monthly_order > 0):

```

```

        model.add_seasonality(name='monthly', period=1*number_steps_in_day,
fourier_order=monthly_order, mode = 'multiplicative')

```

```

if (quarterly_order > 0):
    model.add_seasonality(name='quarterly',
period=365.25/12*4*number_steps_in_day,
                            fourier_order=quarterly_order, mode = 'multiplicative')
    model.fit(df)

return model

```

In [25]:

linkcode

```

def model_tuning(res, df, anomalous, id_station_name, number_steps_in_day,
forecasting_period):

```

```

    # Prophet model with parameters and structure tuning

```

```

    Q_order_from = -1

```

```

    monthly_order_from = -1

```

```

    weekly_order_from = -1

```

```

    quarterly_order_from = -1

```

```

    Q_order_to = 0

```

```

    monthly_order_to = 0

```

```

    weekly_order_to = 0

```

```

    quarterly_order_to = 0

```

```

    order_min = 3

```

```

    order_max = 6

```

```

if len(df)/number_steps_in_day > 31*4*2:

```

```

    # All data more 2 quarters

```

Додаток Г  
(обов'язковий)

## ІЛЮСТРАТИВНА ЧАСТИНА

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ ЯКОСТІ  
АТМОСФЕРНОГО ПОВІТРЯ У МІСТІ ВІННИЦІ ЗА ДАНИМИ  
ГРОМАДСЬКОГО МОНІТОРИНГУ (КОМПЛЕКСНА МКР) ЧАСТИНА І.  
ПРОГНОЗУВАННЯ ДАНИХ

Виконав: студент гр. 2ІСТ-20м

\_\_\_\_\_ Бондарчук А. Р.

« \_\_\_ » \_\_\_\_\_ 2021 р.

Керівник: д.т.н., проф.

\_\_\_\_\_ Мокін О. Б.

« \_\_\_ » \_\_\_\_\_ 2021 р.

Нормоконтроль: к.т.н., доцент

\_\_\_\_\_ Жуков С. О.

« \_\_\_ » \_\_\_\_\_ 2021 р.

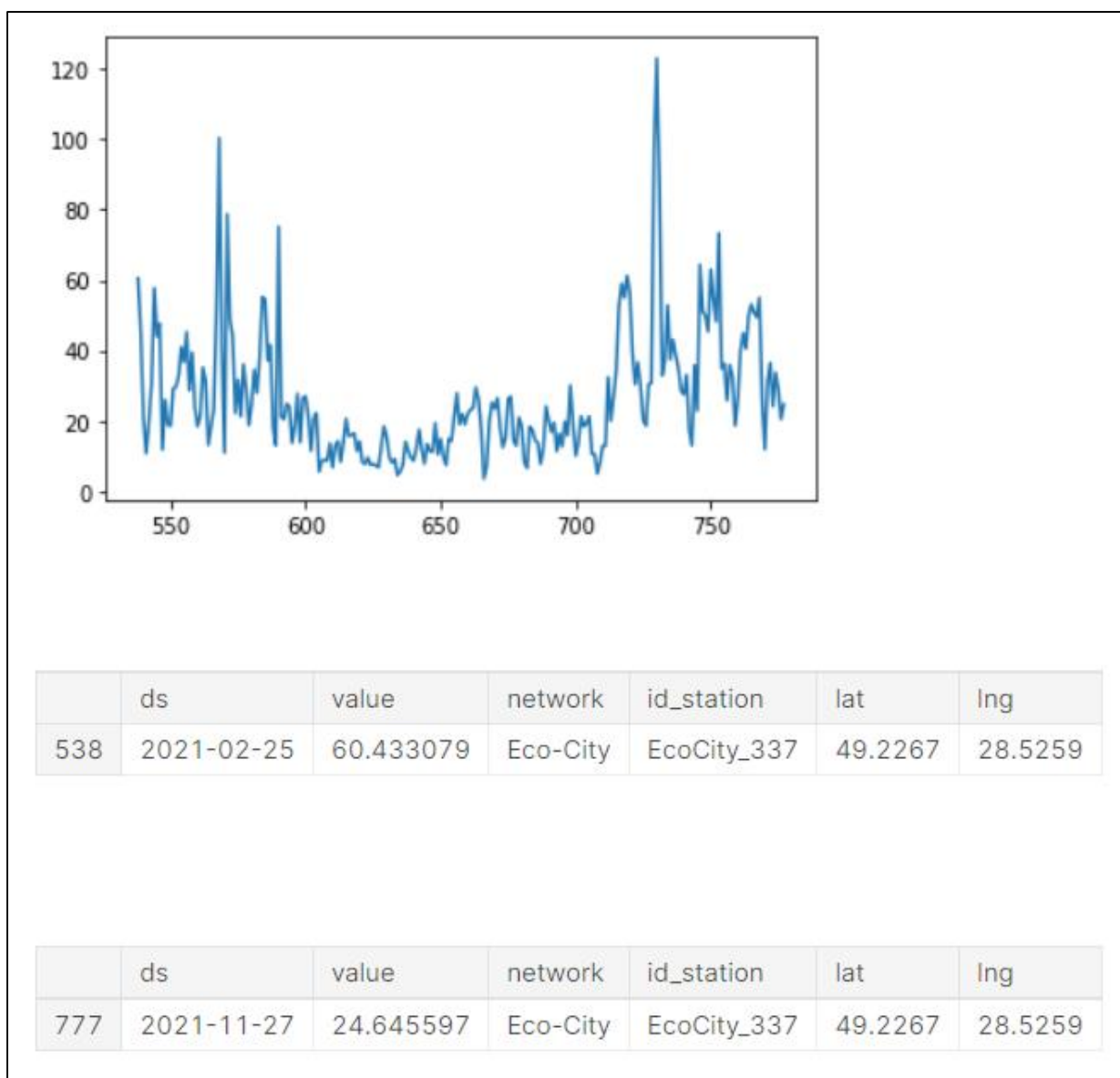


Рисунок Г.1 – Візуалізація значень з поста EcoCity\_337

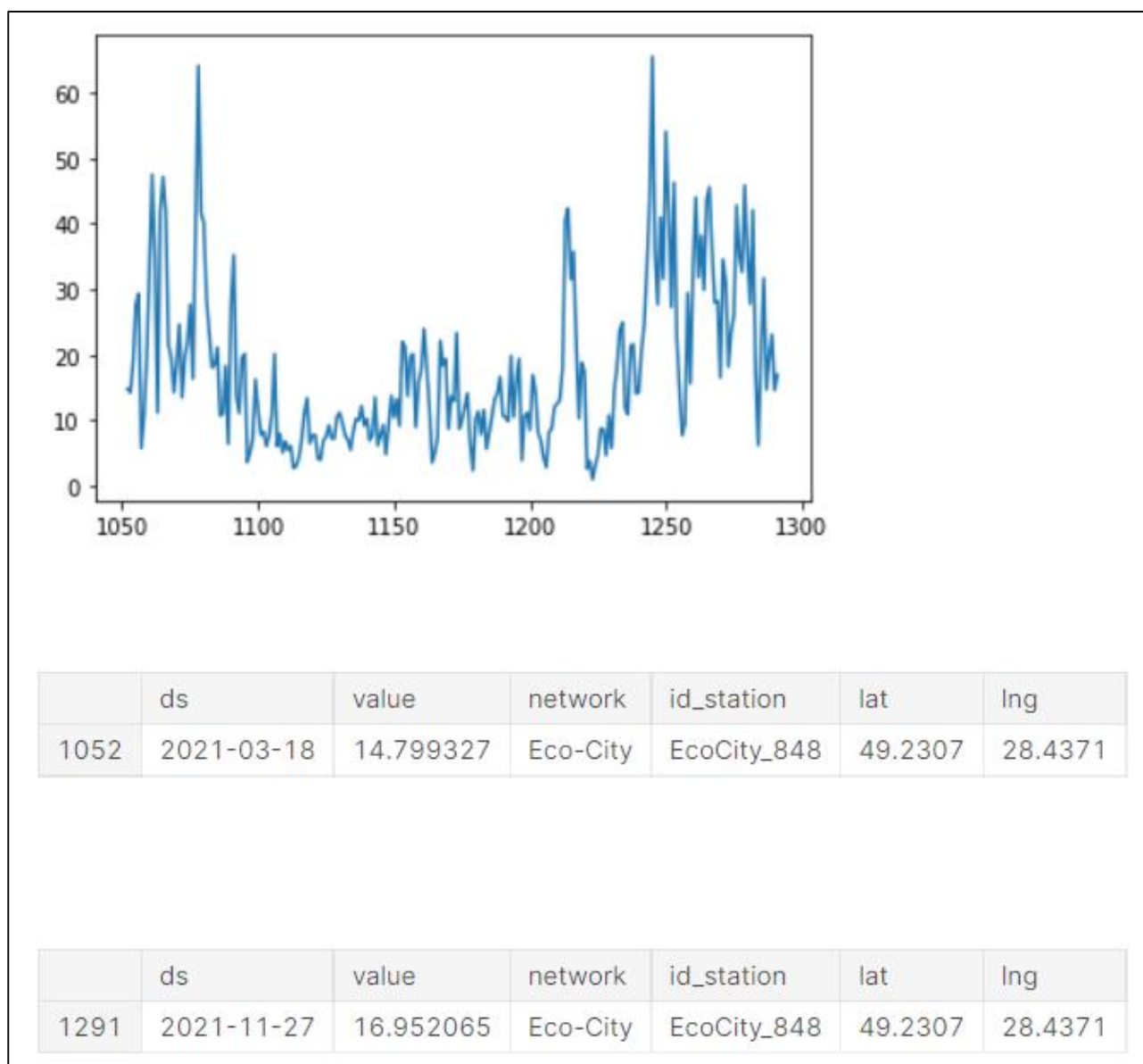


Рисунок Г.2 – Візуалізація значень з поста EcoCity\_848



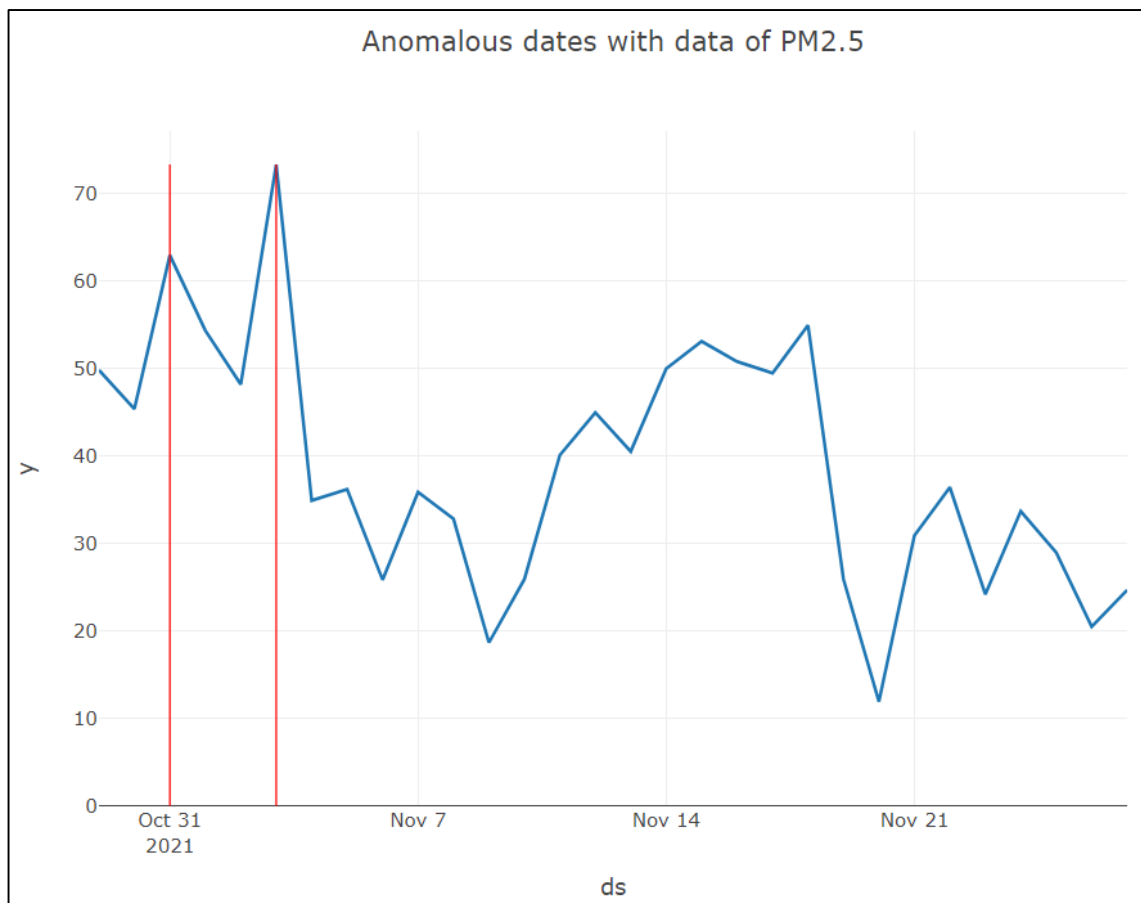


Рисунок Г.3 – Графік з аномальними даними з поста EcoCity\_337

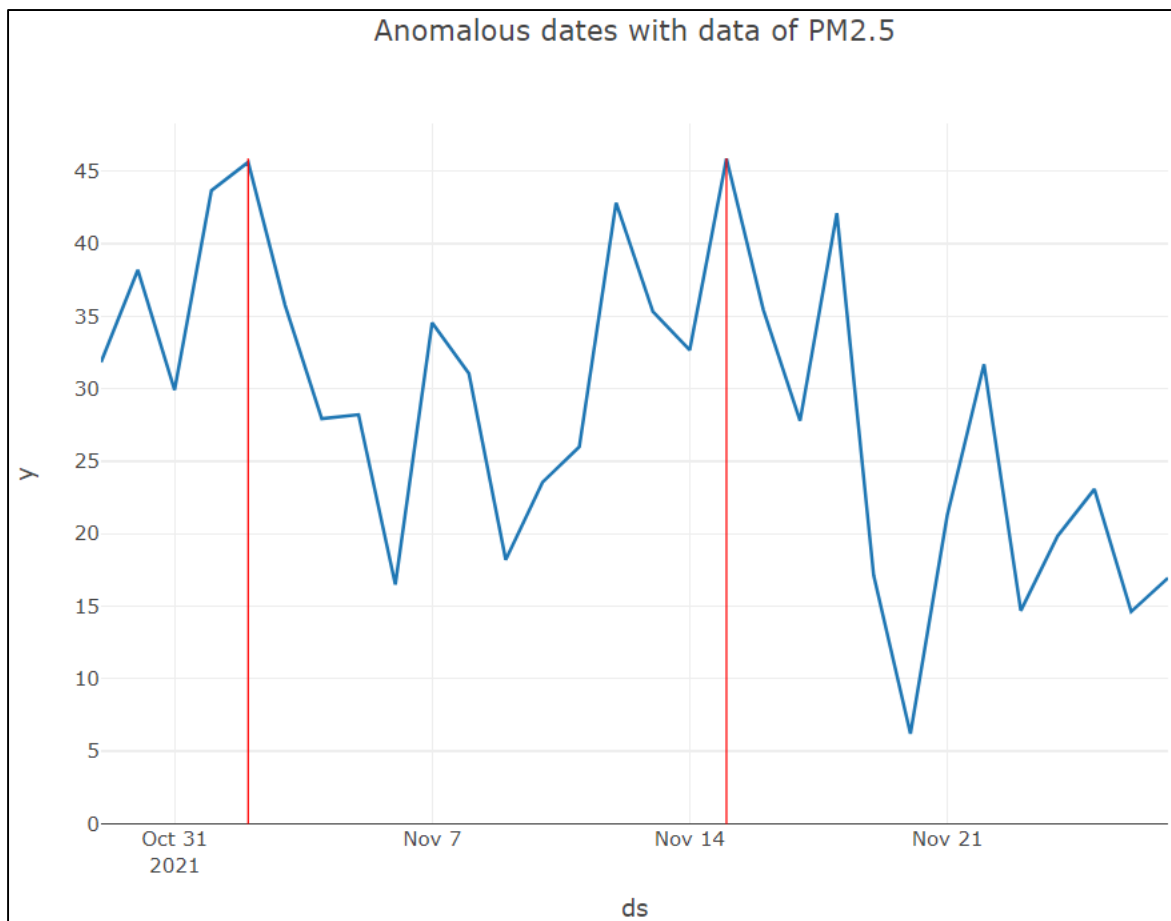


Рисунок Г.4 – Графік з аномальними даними з поста EcoCity\_848

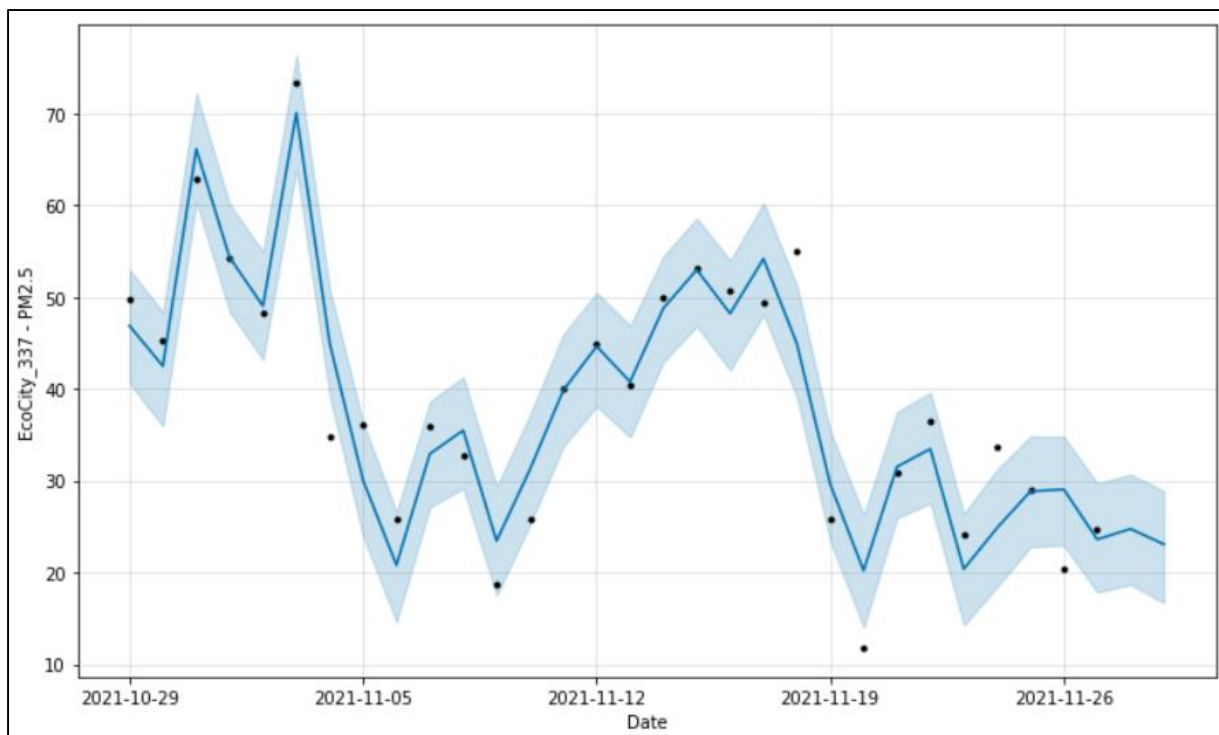


Рисунок Г.5 – Графік з прогнозом кількості пилу для поста EcoCity\_337

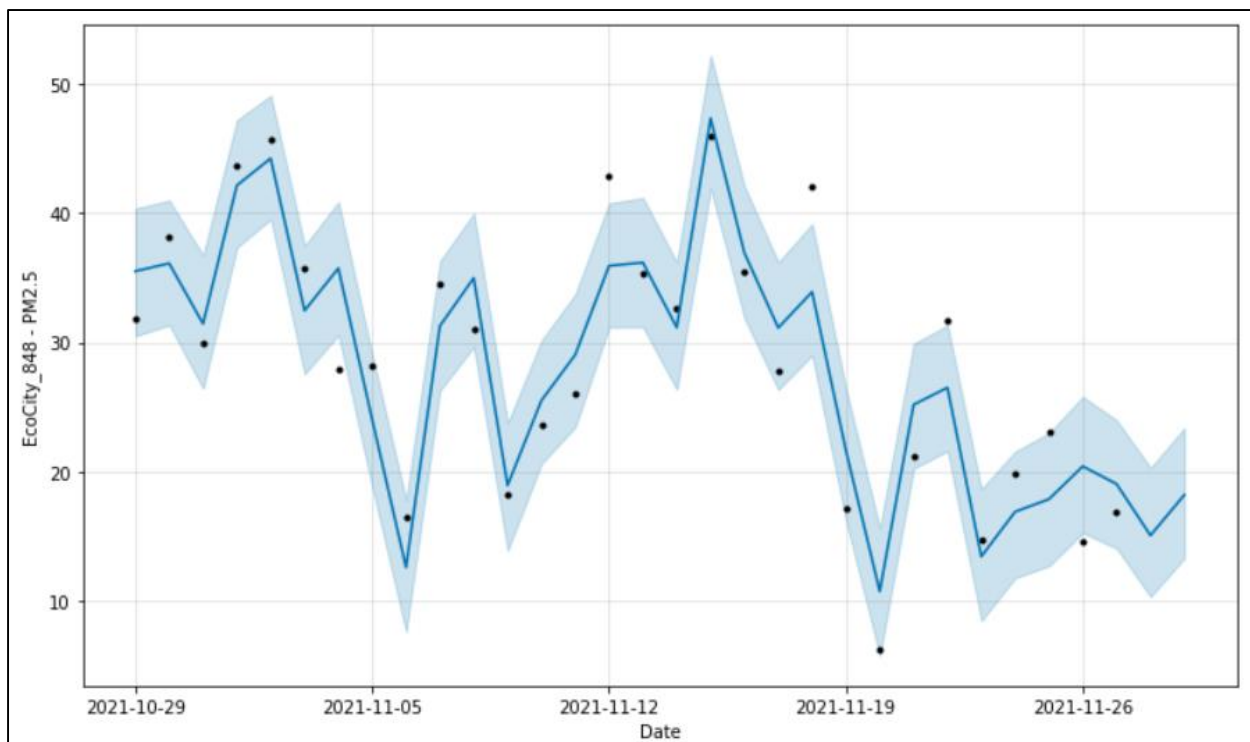


Рисунок Г.6 – Графік з прогнозом кількості пилу для поста EcoCity\_848