

## Порівняльний розвідувальний аналіз даних про стан атмосферного повітря за даними мережі EcoCity з використанням бібліотеки Sweetviz

<sup>1</sup>Вінницький національний технічний університет

### Анотація

Досліджено можливості порівняльного розвідувального аналізу даних про стан атмосферного повітря за даними мережі EcoCity з використанням бібліотеки Sweetviz. Показано, що більш ефективним є саме багатофакторний аналіз, який дозволяє виявити важливі закономірності як у просторі, так і у часі.

**Ключові слова:** розвідувальний аналіз даних, якість атмосферного повітря, часові ряди, EcoCity, Sweetviz.

### Abstract

The paper aims to overview the capabilities of the exploratory data analysis library Sweetviz for air quality data provided by the air quality monitoring system EcoCity. The result has proven that multivariate analysis is more effective, as it reveals important regularities both in space and in time.

**Keywords:** exploratory data analysis, air quality, time series, EcoCity, Sweetviz.

### Вступ

Мінлива якість атмосферного повітря постійно стимулює шукати причини таких змін з метою їх кращого передбачення та завчасного реагування та прийняття певних управлінських рішень.

В Україні реалізується україно-чеський проєкт «Чисте повітря для України». Автори даного дослідження мають авторизований доступ від ВНТУ до даних моніторингу якості атмосферного повітря від громадського проєкту EcoCity у межах цього проєкту [1] та їх сервісу «Кабінет дослідника» [2] (рис. 1).

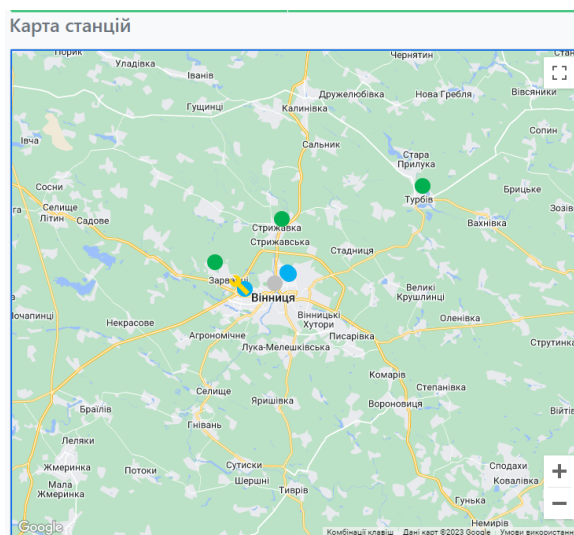


Рис. 1. Карта станцій моніторингу у Вінницькій області

Система формує доволі великі часові ряди, які потребують аналізу.

Метою даного дослідження є дослідити можливості порівняльного розвідувального аналізу даних

про стан атмосферного повітря за даними мережі EcoCity з використанням бібліотеки Sweetviz.

### Огляд бібліотек Sweetviz та AutoViz

SweetViz [11] – це бібліотека для візуалізації та аналізу даних в середовищі Python. Вона автоматично будує гистограми для всіх числових та категоріальних змінних. Є можливість аналізу зв'язку між числовою та категоріальною ознаками, будуючи графіки відносності. Це дозволяє визначити, як числові характеристики змінюються в залежності від категоріальних ознак. SweetViz надає зручні інтерфейси для порівняння двох різних датасетів. Це може бути корисно для аналізу різниці між наборами даних, наприклад, навчальним та тестовим, або даними, зібраними у різний час.

AutoViz [11] – це бібліотека для автоматичної візуалізації даних у Python. AutoViz автоматично визначає тип графіка для кожної змінної в залежності від її характеристик. Наприклад, числові змінні можуть бути відображені у вигляді гистограм, діаграм розсіювання або лінійних графіків, а категоріальні – у вигляді кругових діаграм чи стовпчикових графіків – намагається надати корисні візуалізації для кожного типу даних. Може робити згруповані аналізи, наприклад, враховуючи залежності між змінними або розподіл значень змінних за певною категорією. Може генерувати інтерактивні графіки.

Узагальнені можливості цих бібліотек див. у табл. 1

Таблиці 1 – Порівняльна характеристика аналітичних можливостей EDA-бібліотек AutoViz, SweetViz

Бібліотека	Види графіків	Закономірності
AutoViz	Гистограми, діаграми розсіювання, лінійні графіки, ящикові графіки, кругові діаграми, залежності між змінними.	Швидка автоматична візуалізація даних без великої кількості коду, автоматичний вибір типу графіка в залежності від характеристик даних.
SweetViz	Розподіл змінних, графіки розсіювання, порівняльні графіки, графіки відносності, таблиці з описом даних.	Взаємозв'язок між змінними, аналіз внутрішнього вигляду змінних, порівняння розподілів даних між різними наборами.

Бібліотека SweetViz є більш зручною саме для порівняльного аналізу даних.

### Результати дослідження

Використовуючи датасет, створений авторами у Kaggle за даними EcoCity, було проаналізовано:

- один показник з однієї станції за 2022 і 2023 роки (рис. 2);
- показник pm2.5 та pm10 з однієї станції за 2022-2023 роки (рис. 3);
- показник pm10 з двох станцій за 2023 рік (рис. 4).

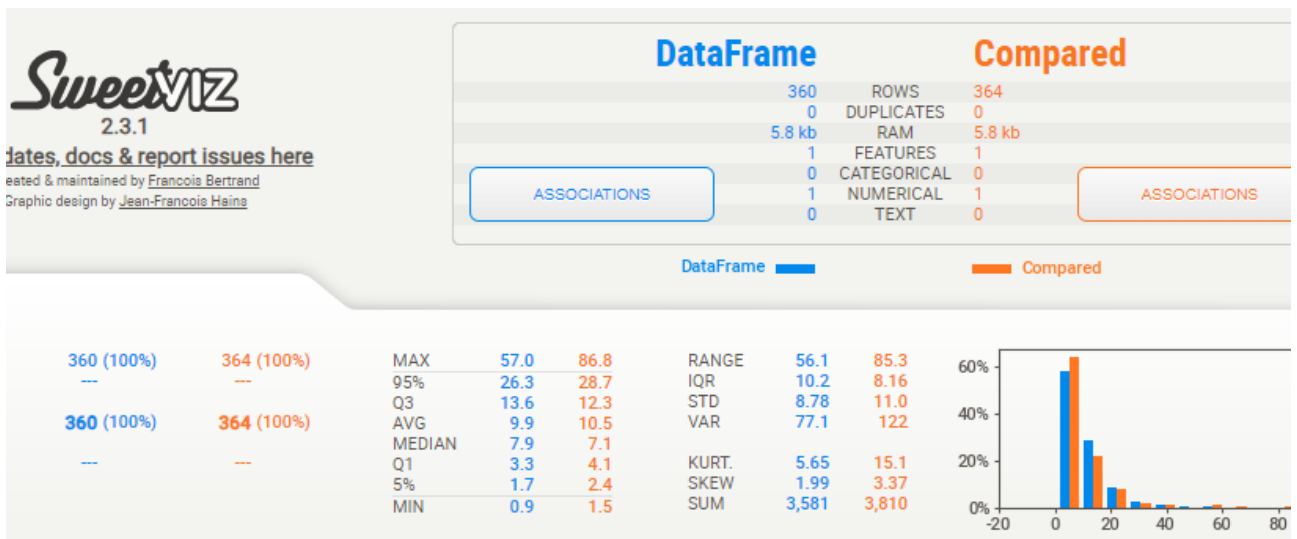


Рис. 2. Один показник з однієї станції за 2022 і 2023 роки

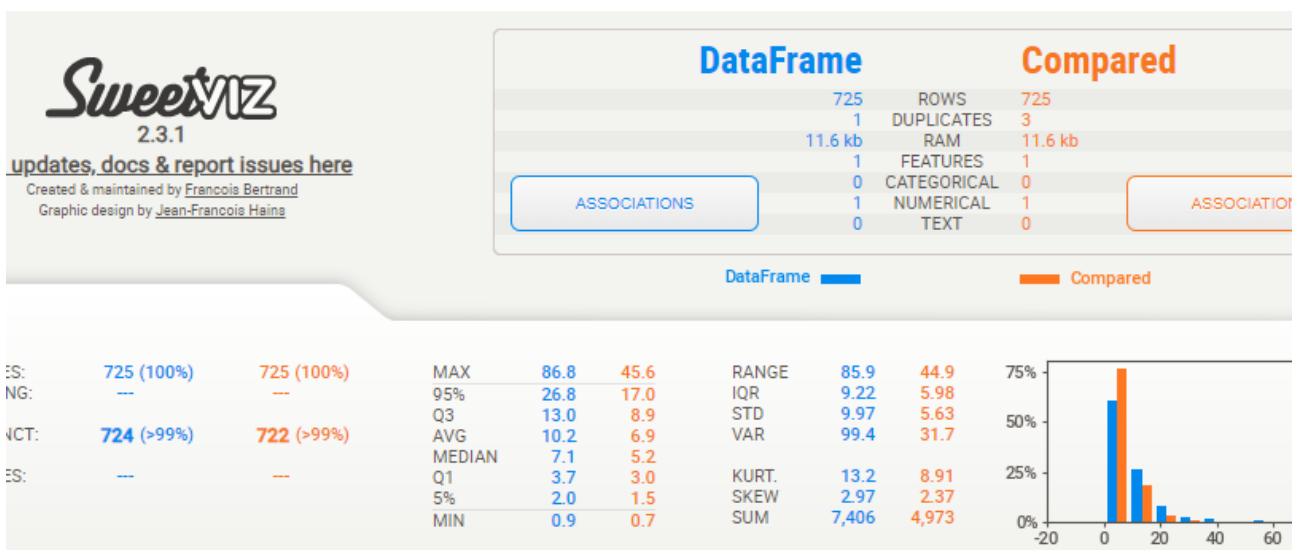


Рис. 3. Показник pm2.5 та pm10 з однієї станції за 2022-2023 роки

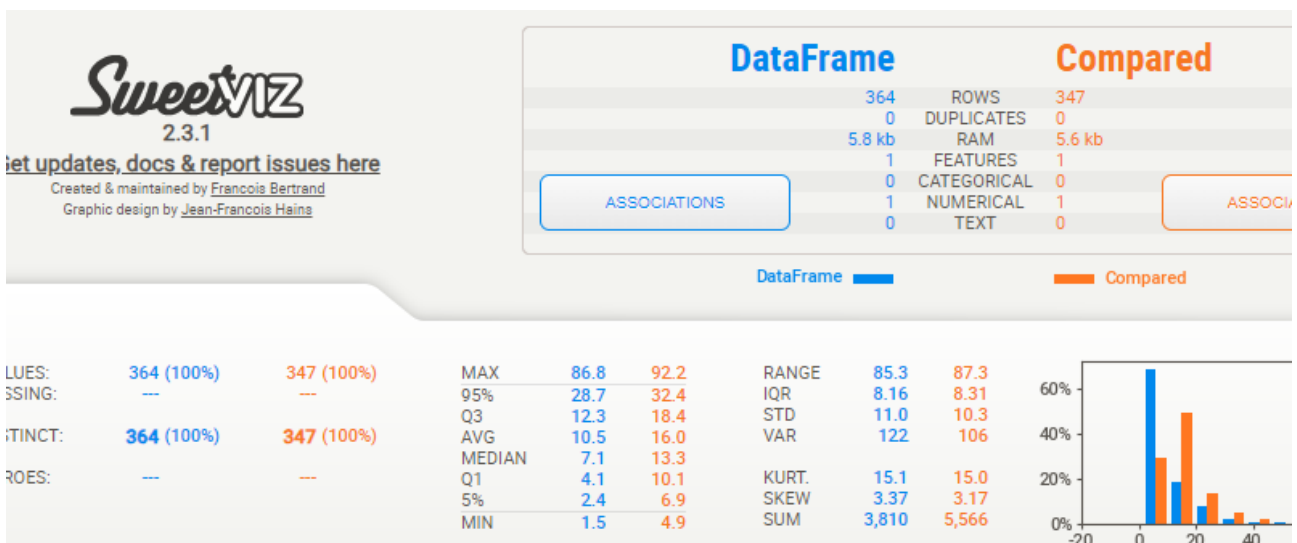


Рис. 4. Показник pm10 з двох станцій за 2023 рік

Аналіз рисунків дозволяє зробити такі висновки:

- 1) один показник на одній мало змінюється по роках;
- 2) різні показники однієї станції мають схожу динаміку;
- 3) один показник з різних станцій є різним, оскільки на них впливають різні чинники.

Загалом, результати інтуїтивно очікувані, але аналіз дозволяє отримати кількісні характеристики у розрізі конкретних періодів, станцій, показників, що є дуже цінним.

### Висновки

Досліджено можливості порівняльного розвідувального аналізу даних про стан атмосферного повітря за даними мережі EcoCity з використанням бібліотеки Sweetviz. Показано, що більш ефективним є саме багатofакторний аналіз, який дозволяє виявити важливі закономірності як у просторі, так і у часі. Це дозволить краще виявляти причини появи таких закономірностей з метою їх кращого передбачення та завчасного реагування та прийняття певних управлінських рішень.

### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Eco-City Громадський моніторинг стану якості повітря [Електронний ресурс] – Режим доступу до ресурсу: <https://eco-city.org.ua/>.
2. Eco-City Кабінет дослідника [Електронний ресурс] – Режим доступу до ресурсу: <https://archive.eco-city.org.ua/>.
3. Kaggle Notebook «50 Advanced Tips: Data Science for tabular data» : веб-сайт. URL: <https://www.kaggle.com/code/vbmokin/50-advanced-tips-data-science-for-tabular-data/notebook>

**Мокін Віталій Борисович** – д-р. техн. наук, проф., завідувач кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, Вінниця, e-mail: [ybmokin@vntu.edu.ua](mailto:ybmokin@vntu.edu.ua)

**Шмундяк Дмитро Олександрович** — аспірант кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, Вінниця, [dimashmund@gmail.com](mailto:dimashmund@gmail.com)

**Копняк Володимир Євгенович** — аспірант кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, Вінниця, [vkopnyak@gmail.com](mailto:vkopnyak@gmail.com)

**Mokin Vitalii B.** – Dr. tech. Sciences, Prof., Head of the Department of System Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia, e-mail: [ybmokin@vntu.edu.ua](mailto:ybmokin@vntu.edu.ua)

**Shmundiak Dmytro O.** – Postgraduate student of the Department of System Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia, [dimashmund@gmail.com](mailto:dimashmund@gmail.com).

**Kopniak Volodymyr Y.** – Postgraduate student of the Department of System Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia, [vkopnyak@gmail.com](mailto:vkopnyak@gmail.com)