Практична робота № 6

**Тема:** ***Візуалізація аналізу даних в Google Colab*.**

Мета: формування практичних навичок виконання задач аналізу данихв Google Colab засобами бібліотек **Matplotlib** та **Seaborn**.

**Теоретичні відомості**

Щоб правильно візуалізувати дані, потрібно:

* переконатися, що ви розумієте типи графіків і їхні функції;
* отримати дані;
* вибрати інструмент;
* дослідити дані;
* показати найцікавіші висновки на графіках.

Граматика графіки — розкладання графіка на компоненти, з яких і складається результат. Значення компонентів граматики таке:

* Масив даних (data);
* Схема того, як змінні з масиву відповідають візуальним засобам (aesthetics) — яка змінна буде відповідати якій осі;
* Геометричний об'єкт (geometries) — які геометричні об'єкти відповідатимуть яким даним;
* Панелі (facets) — зі скількох частин буде складатися графік;
* Статистичне перетворення (statistics) — наприклад, кількість спостережень в кожній групі на стовпчастій діаграмі;
* Координатна система (coordinates);
* Художнє оформлення (theme).

Якщо певні компоненти граматики не визначені точно, можна взяти стандартні значення, що відповідають типу графіка.

Є кілька корисних методів, про які варто пам'ятати під час створення графіків. Це допоможе легше сприймати інформацію.

1. Використовувати графік для порівняння величин.
2. Порівнювати можна за допомогою візуальної ієрархії (розміру елементів), наприклад, висоти стовпців або величини сегментів.
3. Мінімізувати «шум» і нагромадження даних. Скажімо, якщо аномальні значення не суттєві для аналізу або представлення даних, їх не обов'язково показувати на графіку. Якщо даних для одного графіка занадто багато, їх можна представляти у вигляді кластерів (наприклад, як на бульбашковому графіку).
4. Вводити колір останнім. Колір — це фактично ще одна змінна, останній вимір, який можна додати в графік. Після додавання кольору, як правило, вже не залишається графічних засобів уявлення, тому варто приберегти його наостанок (ну, або зовсім не використовувати).

**Зміст роботи.**

1. Використовуючи набори даних за посиланням, <https://data.gov.ua/dataset>, (flights) створити датафрейм.
2. Провести аналіз отриманих даних використовуючи матеріали попередніх занять. Це може бути розрахунок статистичних даних (середнього значення, мінімального та максимального….), побудова графіків і діаграм використовуючи бібліотеки **Matplotlib** та **Seaborn.**
3. ***У звіті відобразити основні етапи виконання роботи. На перевірку прислати звіт і посилання на блокнот.***

**Контрольні запитання.**

1. Як розпочати роботу з Google Colab?
2. Як створити DataFrame?
3. Як додати нову колонку, рядок?
4. Як отримати серію з DataFrame?
5. Як отримати доступ до колонки, рядка?
6. Як записати DataFrame у файл?
7. Що представляє собою набір даних і де його можна взяти?
8. Яке ключове слово використовується перед функцією бібліотеки Seaborn, Matplotlib?
9. Яка команда дозволить відображати графіки і діаграми на екрані при використанні бібліотек Seaborn, Matplotlib?
10. Які існують елементи графіка створеного засобами Seaborn та Matplotlib?
11. Які параметри можна задавати для графіків?