**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ**

**імені ТАРАСА ШЕВЧЕНКА**

Факультет інформаційних технологій

**Кафедра прикладних інформаційних систем**

****

**Звіт**

до виконання лабораторної роботи №1

з дисципліни «Data Science та Big Data»

на тему:

## «Агрегація, обробка пропусків та візуалізація даних пакетами Python»

(*тема роботи*)

**Виконано:**

студ. групи ПП- 41

Рижука Андрія

**Перевірено:**

Білий Р.О.

**Київ – 2023**

**Мета роботи:**

Метою лабораторної роботи є отримання практичних навичок у роботі з raw data, використовуючи пакети jupyter, pandas, seaborn.

**Завдання:**

### Завантажити файли з даними у папку проекту з посилання:

### <https://github.com/oreillymedia/doing_data_science>

### Створіть нову змінну age\_group, яка агрегує користувачів як <18, 18–24, 25–34, 35–44, 45–54, 55–64 та 65+.

### Зафіксуйте на діаграмі кількість показів та показник переходів (CTR = #clicks/#inmpressions) для цих шести вікових категорій.

### Вивчіть дані та проведіть візуальні та кількісні порівняння між сегментами користувачів/демографічними групами (наприклад, чоловіки старше 18 років у порівнянні з жінками старше 18 років або авторизовані та неавторизовані користувачі).

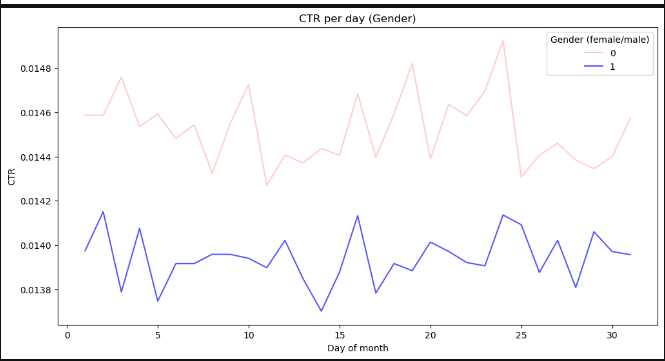
### Створіть метрики/вимірювання/статистику, які підсумовують дані. Приклади можливих метрик включають CTR, квантил, середнє значення, медіану, дисперсію та максимальне значення. Ці показники потрібно розрахувати за різними сегментами користувачів. Подумайте про елементи, які важливо відстежувати з часом - що стискає дані, але, як і раніше, захоплює поведінку користувача.

### Результати статистичного дослідження подати у вигляді результуючого ДатаФрейма (одного), дивлячись на який можна зрозуміти і порівнювати дані за віковими підкатегоріями.

### Опишіть та інтерпретуйте будь-які закономірності, які знайдете.

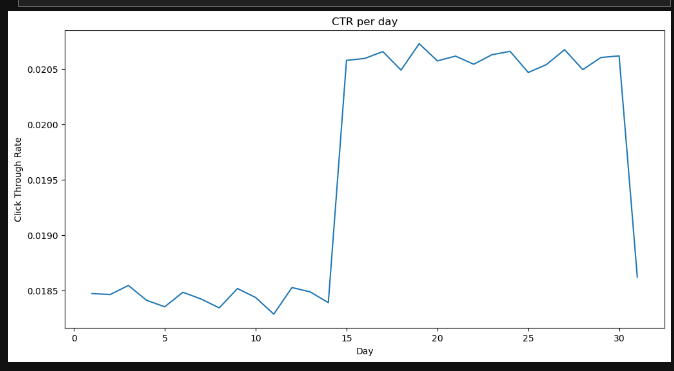
### Завантажити файл ipynb з виконаними завданнями на git в окрему папку з відповідною назвою лабораторної роботи

Висновки:



На основі даних, наданих отриманих у лабораторній роботі , можна зробити такі висновки:

* Кількість показів рекламних оголошень збільшується протягом місяця. Це може бути пов'язано з рядом факторів, таких як:
* Збільшення кількості користувачів, які переглядають веб-сайти або використовують програми, де розміщуються рекламні оголошення.
* Збільшення частоти показу рекламних оголошень.
* Зміна стратегії рекламування, яка спрямована на показ рекламних оголошень більшій кількості користувачів.
* Жінки та чоловіки в рівних кількостях переглядають рекламні оголошення. Це означає, що рекламодавці можуть ефективно звертатися до обох груп населення.
* Клікабельність рекламних оголошень залишається відносно стабільною протягом місяця. Це означає, що рекламні оголошення ефективні в тому, щоб залучити користувачів до взаємодії з ними.
* Компанія успішно залучає користувачів до свого веб-сайту.
* Компанія може підвищити ефективність своєї рекламної кампанії, запровадивши додаткові заходи, такі як таргетинг на конкретні аудиторії або використання більш привабливих рекламних оголошень.
* Звичайно, для того, щоб зробити більш точні висновки, необхідно провести додатковий аналіз даних. Наприклад, можна розглянути такі фактори, як:
* Тип рекламних оголошень, які розміщуються.
* Контент веб-сайту або програми, де розміщуються рекламні оголошення.
* Цільова аудиторія рекламних оголошень.
* Аналіз цих додаткових факторів може допомогти краще зрозуміти, чому кількість показів рекламних оголошень збільшується, і як можна підвищити ефективність рекламної кампанії.



* Кількість показів рекламних оголошень зображень збільшується протягом місяця. Це може бути пов'язано з рядом факторів, таких як:
* Збільшення кількості користувачів, які переглядають веб-сайти або використовують програми, де розміщуються рекламні оголошення зображень.
* Збільшення частоти показу рекламних оголошень зображень.
* Зміна стратегії рекламування, яка спрямована на показ рекламних оголошень зображень більшій кількості користувачів.
* Клікабельність рекламних оголошень зображень становить в середньому близько 2,05%. Це означає, що приблизно 2 користувачі з кожних 100, які бачать рекламне оголошення зображення, натискають на нього.
* Клікабельність рекламних оголошень зображень залишається відносно стабільною протягом місяця. Це означає, що рекламні оголошення зображень ефективні в тому, щоб залучити користувачів до взаємодії з ними.
* Конкретні висновки, які можна зробити, залежать від контексту цих даних. Наприклад, якщо ці дані стосуються рекламних оголошень зображень, які розміщуються на веб-сайті компанії, можна зробити такі додаткові висновки:
* Компанія успішно залучає користувачів до свого веб-сайту за допомогою рекламних оголошень зображень.
* Компанія може підвищити ефективність своєї рекламної кампанії, запровадивши додаткові заходи, такі як таргетинг на конкретні аудиторії або використання більш привабливих рекламних оголошень зображень.