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**.

**Зміст роботи.**

***У звіті відобразити основні етапи виконання роботи. На перевірку прислати звіт і файли результатів.***

Робота з Seaborn Seaborn — це пакет Python для візуалізації даних, який також забезпечує високорівневий інтерфейс для Matplotlib. З Seaborn легше працювати, ніж Matplotlib, і насправді він розширює Matplotlib, але майте на увазі, що Seaborn не такий потужний, як Matplotlib.

Seaborn вирішує дві проблеми Matplotlib. Перша включає параметри Matplotlib за замовчуванням. Seaborn працює з різними параметрами, що забезпечує більшу гнучкість, ніж відтворення графіків Matplotlib за замовчуванням. Seaborn усуває обмеження значень за замовчуванням Matplotlib для таких функцій, як кольори, галочки на верхній і правій осях і стиль (серед інших). Крім того, Seaborn полегшує побудову цілих фреймів даних (подібно до Pandas), робити це в Matplotlib важче.

 Тим не менш, оскільки Seaborn розширює Matplotlib, знання Matplotlib є вигідним і спростить вашу криву навчання.

*Можливості Seaborn*

Деякі з можливостей Seaborn включають:

 • масштабування графіка Seaborn

• задання стилю графіка

• встановлення розміру фігури

• обертання тексту напису

• встановлення xlim або ylim

• встановлення логарифмічного масштабу

• додавання заголовку

Деякі корисні методи:

• plt.xlabel()

• plt.ylabel()

• plt.annotate()

• plt.legend()

• plt.ylim()

• plt.savefig()

Seaborn підтримує різні вбудовані набори даних, як і NumPy і Pandas, включаючи набір даних Iris і Titanic.

Як відправну точку, трирядковий зразок коду в наступному розділі показує, як відобразити рядки у вбудованому наборі даних «tips».

*Вбудовані набори даних Seaborn*

У наступному листингі показано вміст файлу seaborn\_tips.py, який ілюструє, як читати набір даних tips у фрейм даних і відображати перші п’ять рядків набору даних

import pandas as pd

**import seaborn as sns**

**%matplotlib inline** # для відображння графіків в блокноті

**df = sns.load\_dataset("tips")**

**print(df.head()) # або df.head()**

Цей лiстинг дуже простий: після імпорту seaborn змінна df ініціалізується даними з вбудованого набору даних tips, а оператор print() відображає перші п’ять рядків df. Зверніть увагу, що API load\_dataset() шукає онлайн або вбудовані набори даних. На виході отримаємо:



1. Для вибору набору даних можна прейти за посиланням https://www.kaggle.com/datasets.

Для завантаження Google Colab перейти за посиланням. [**https://colab.research.google.com/**](https://colab.research.google.com/)Google Colab

1. Завантажити набір даних **"tips"** та створюємо датасет. Бажано роздрукувати щоб переконатись в правильності формування датасету. Інформацію про набір даних **tips** можна прочитати за посиланням: https://www.angela1c.com/projects/tips-project-files/part1/
2. Отримати інформацію про створений датасет.
3. Побудувати гістограми розподілу кількості клієнтів по дням в загальному, у співвідношенні чоловіки/жінки, палять/не палять.
4. За допомогою функції jointplot() роілюструйте спільний розподіл за двома змінними.
5. Побудувати теплову карту і проаналізувати отриману інформацію.
6. Побудувати звичайний графік розсіювання і розподілом даних за статтю.

**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?

# <https://www.youtube.com/watch?v=1CFX5MMp2mQ&t=120s> Python Seaborn графіки. Аналізуємо DataSet. Візуалізація на прикладі вбудованих даних.

1. <https://www.youtube.com/watch?v=RZm8TzRfNCE>
2. <https://devzone.org.ua/post/vizualizatsiia-danykh-i-metryky-v-testuvanni>
3. <https://seaborn.pydata.org/tutorial/distributions.html> Seaborn графіки з кодом