

Міністерство освіти і науки України
Національний університет «Львівська політехніка»
Кафедра ІСМ



Звіт
про виконання лабораторної роботи № 8
«Візуалізація та обробка даних за допомогою
спеціалізованих бібліотек Python»
з дисципліни
«Спеціалізовані мови програмування»

Виконав:
Студент групи ІТ-32,
Вольвенко І. Р.

Прийняв:
Щербак С.С

Львів 2023

Мета роботи: Розробка додатка для візуалізації CSV-наборів даних за допомогою Matplotlib та базових принципів ООП (наслідування, інкапсуляція, поліморфізм).

Завдання:

Завдання 1: Вибір CSV-набору даних

Оберіть CSV-набір даних, який ви хочете візуалізувати. Переконайтеся, що він містить відповідні дані для створення змістовних візуалізацій.

Завдання 2: Завантаження даних з CSV

Напишіть код для завантаження даних з CSV-файлу в ваш додаток Python. Використовуйте бібліотеки, такі як Pandas, для спрощення обробки даних.

Завдання 3: Дослідження даних

Визначте екстремальні значення по стовцям

Завдання 4: Вибір типів візуалізацій

Визначте, які типи візуалізацій підходять для представлення вибраних наборів даних. Зазвичай це може бути лінійні графіки, стовпчикові діаграми, діаграми розсіювання, гістограми та секторні діаграми.

Завдання 5: Підготовка даних

Попередньо обробіть набір даних за необхідністю для візуалізації. Це може включати виправлення даних, фільтрацію, агрегацію або трансформацію.

Завдання 6: Базова візуалізація

Створіть базову візуалізацію набору даних, щоб переконатися, що ви можете відображати дані правильно за допомогою Matplotlib. Розпочніть з простої діаграми для візуалізації однієї змінної.

Завдання 7: Розширені візуалізації

Реалізуйте більш складні візуалізації, виходячи з характеристик набору. Поекспериментуйте з різними функціями Matplotlib та налаштуваннями.

Завдання 8: Декілька піддіаграм

Навчіться створювати кілька піддіаграм в межах одного малюнка для відображення декількох візуалізацій поруч для кращого порівняння.

Завдання 9: Експорт і обмін

Реалізуйте функціональність для експорту візуалізацій як зображень (наприклад, PNG, SVG) або інтерактивних веб-додатків (наприклад, HTML)

Код:

diagrams_service.py:

```
from collections import Counter
import abc
from matplotlib import pyplot as plt
from config import (DIFFERENCE_IN_THOUSANDS_HISTOGRAM,
SEX_PIE_CHART_PHOTO,
STATE_BAR_CHART_PHOTO, COMBINED_DIAGRAM_PHOTO, logger)
from entity.user import User
from shared.file_processors import CsvProcessor as csv_processor

class DiagramService(abc.ABC):

    def __init__(self):
        self._users = []

class DiagramServiceImpl(DiagramService):

    def __init__(self, file_path: str):
```

```

super().__init__()
logger.info("Loading user data from CSV")
users_dataframe = csv_processor.read(file_path)
for data in users_dataframe.values:
    self._users.append(User(data))

def get_difference_in_thousands(self):
    return [user.followers_count / 1000 for user in self._users]

def get_sex(self):
    return [user.sex for user in self._users]

def get_state(self):
    return [user.location for user in self._users]

def create_difference_in_thousands_histogram(self, has_to_be_downloaded=False):
    logger.info("Creating histogram of difference in thousands")
    difference_in_thousands = self.get_difference_in_thousands()
    plt.hist(difference_in_thousands, bins=20, edgecolor='black')
    plt.title('Histogram of Followers Count in Thousands')
    plt.xlabel('Thousands of Followers')
    plt.ylabel('Frequency')
    if has_to_be_downloaded:
        plt.savefig(DIFFERENCE_IN_THOUSANDS_HISTOGRAM)
    plt.show()

def create_sex_pie_chart(self, has_to_be_downloaded=False):
    logger.info("Creating sex pie chart")
    sex = self.get_sex()
    sex_counter = Counter(sex)
    plt.pie(list(sex_counter.values()), labels=list(sex_counter), startangle=90, colors=['blue',
'pink'])
    plt.title('Pie Chart of User Sex')
    if has_to_be_downloaded:
        plt.savefig(SEX_PIE_CHART_PHOTO)

```

```
plt.show()
```

```
def create_state_bar_chart(self, has_to_be_downloaded=False):
    logger.info("Creating state bar chart")
    state = self.get_state()
    state_counter = Counter(state)
    plt.figure(figsize=(10, 6))
    plt.bar(state_counter.keys(), state_counter.values(), color='green')
    plt.title('Bar Chart of Users per State')
    plt.xlabel('State')
    plt.ylabel('Number of Users')
    plt.xticks(rotation=45)
    if has_to_be_downloaded:
        plt.savefig(STATE_BAR_CHART_PHOTO)
    plt.show()
```

```
def create_combined_diagram(self, has_to_be_downloaded=False):
    logger.info("Creating combined diagram")
    fig, (ax1, ax2, ax3) = plt.subplots(3, 1, figsize=(10, 15))

    difference_in_thousands = self.get_difference_in_thousands()
    ax1.hist(difference_in_thousands, bins=20, color='blue', edgecolor='black')
    ax1.set_title('Histogram of Followers Count in Thousands')

    sex = self.get_sex()
    sex_counter = Counter(sex)
    ax2.pie(list(sex_counter.values()), labels=list(sex_counter), startangle=90, colors=['blue',
'pink'])
    ax2.set_title('Pie Chart of User Sex')

    state = self.get_state()
    state_counter = Counter(state)
    ax3.bar(state_counter.keys(), state_counter.values(), color='green')
    ax3.set_title('Bar Chart of Users per State')
    ax3.tick_params(axis='x', labelrotation=45)
```

```
plt.tight_layout()
```

```
if has_to_be_downloaded:
```

```
    plt.savefig(COMBINED_DIAGRAM_PHOTO)
```

```
plt.show()
```

```
logger.info("Combined diagram created successfully")
```

diagrams_menu.py:

```
import sys
```

```
from config import USERS_DATA
```

```
from service.lab8.diagrams_service import DiagramServiceImpl
```

```
class DiagramMenu:
```

```
    def run(self):
```

```
        service = DiagramServiceImpl(USERS_DATA)
```

```
        while True:
```

```
            print(
```

```
                "1. Display difference in thousands histogram\n"
```

```
                "2. Display sex pie chart\n"
```

```
                "3. Display state bar chart\n"
```

```
                "4. Display combined diagram\n"
```

```
                "0. Exit\n"
```

```
            )
```

```
        choice = input("Enter your choice: ")
```

```
        match choice:
```

```
            case "1":
```

```
                self.display_diagram(service.create_difference_in_thousands_histogram)
```

```
            case "2":
```

```
                self.display_diagram(service.create_sex_pie_chart)
```

```

case "3":
    self.display_diagram(service.create_state_bar_chart)
case "4":
    self.display_diagram(service.create_combined_diagram)
case "0":
    sys.exit(0)
case _:
    print("Invalid choice. Enter again!")

```

@staticmethod

```

def display_diagram(diagram_function):
    has_to_be_downloaded = input(
        "Do you want to download the diagram? Enter 'y' or "
        "anything else not to download: ") == "y"

    diagram_function(has_to_be_downloaded)

```

users.csv:

Index,User Id,Username,Followers Count,Location,Sex,Date of Creation

```

1,85007,elbqu_841,38267,Alaska,Male,25/08/2011
2,83797,pgeda_510,55943,Arizona,Male,11/04/2019
3,77090,yfrpp_833,47917,South Dakota,Female,06/06/2019
4,48177,jltuj_342,54693,Massachusetts,Female,06/08/2010
5,94473,izqgz_572,56892,Rhode Island,Male,15/12/2021
6,32587,iukcy_202,74695,Ohio,Male,22/11/2015
7,48477,iuxcc_837,41371,Wyoming,Male,03/04/2023
8,73777,esvrs_962,3630,California,Female,27/07/2012
9,19774,gydew_118,4022,Illinois,Male,16/06/2012
10,30198,cpkzh_642,27082,Michigan,Female,06/06/2014
11,70528,jxgzm_707,31623,Nevada,Male,23/01/2020
12,53202,ybcxi_420,18402,Nevada,Male,09/11/2021
13,89906,hnwwu_723,56843,New Mexico,Female,05/08/2011
14,97025,bewhl_533,96479,Louisiana,Female,16/04/2020
15,81523,vudhm_650,41013,Mississippi,Male,07/08/2013
16,25286,jxogb_779,17195,Wyoming,Male,20/06/2015

```

17,57045,fesql_565,1346,New Mexico,Female,29/08/2021
18,17177,ygkif_266,98776,Illinois,Male,12/11/2015
19,40141,wovyx_517,13765,Connecticut,Male,03/06/2011
20,98640,kaumj_917,98554,Delaware,Female,03/09/2016
21,75745,frmjb_947,54587,Vermont,Male,10/09/2023
22,54989,snyag_568,24089,South Dakota,Male,22/05/2018
23,29995,fihel_183,96525,West Virginia,Male,05/07/2018
24,87028,gjnos_389,2381,Montana,Male,28/11/2023
25,72758,denmr_403,96071,South Carolina,Male,17/05/2020
26,86653,yknud_347,26561,California,Female,31/10/2021
27,69906,mrtai_141,71204,Hawaii,Male,06/08/2022
28,91032,rjzog_831,31130,Oregon,Female,01/06/2012
29,55635,okqhv_957,59158,Kansas,Female,09/11/2014
30,23744,xivmg_436,37789,Maryland,Male,23/11/2016
31,20340,nbejf_403,25504,Alabama,Female,29/05/2022
32,62090,uwyqk_727,24316,Idaho,Female,14/10/2010
33,84615,qzqkx_282,98338,Texas,Female,06/03/2022
34,31901,yghxq_731,63019,Georgia,Female,07/09/2012
35,36508,klhuz_614,30167,New York,Male,01/10/2010
36,50460,zzgxl_791,25644,Tennessee,Female,28/03/2012
37,28370,dzkiv_685,12005,North Dakota,Male,24/05/2016
38,39356,jvqdc_207,86370,Utah,Female,02/02/2012
39,55773,dqzxc_345,81391,New Hampshire,Female,19/07/2014
40,17981,lwekm_918,15210,Maine,Male,07/03/2019
41,98858,jymxw_253,91128,Missouri,Female,12/12/2021
42,50970,rujllh_144,96120,Arizona,Male,06/06/2016
43,40892,gqeyh_853,41649,Indiana,Female,24/12/2017
44,74550,rffun_318,21050,Iowa,Female,04/08/2018
45,96846,zwoqe_235,56494,Florida,Male,23/11/2022
46,70064,jtyql_792,71789,Montana,Male,04/07/2018
47,73325,pzhek_885,7427,Hawaii,Female,13/05/2019
48,66943,qikrt_618,64825,Michigan,Male,25/06/2014
49,11933,vslon_632,52911,Pennsylvania,Female,12/06/2023
50,93516,gwmjb_519,85125,New Jersey,Female,09/11/2014

Результати виконання програми:

1. Гістограма, що показує різницю кількості підписників користувачів, у тисячах, наведена на рисунку 1:

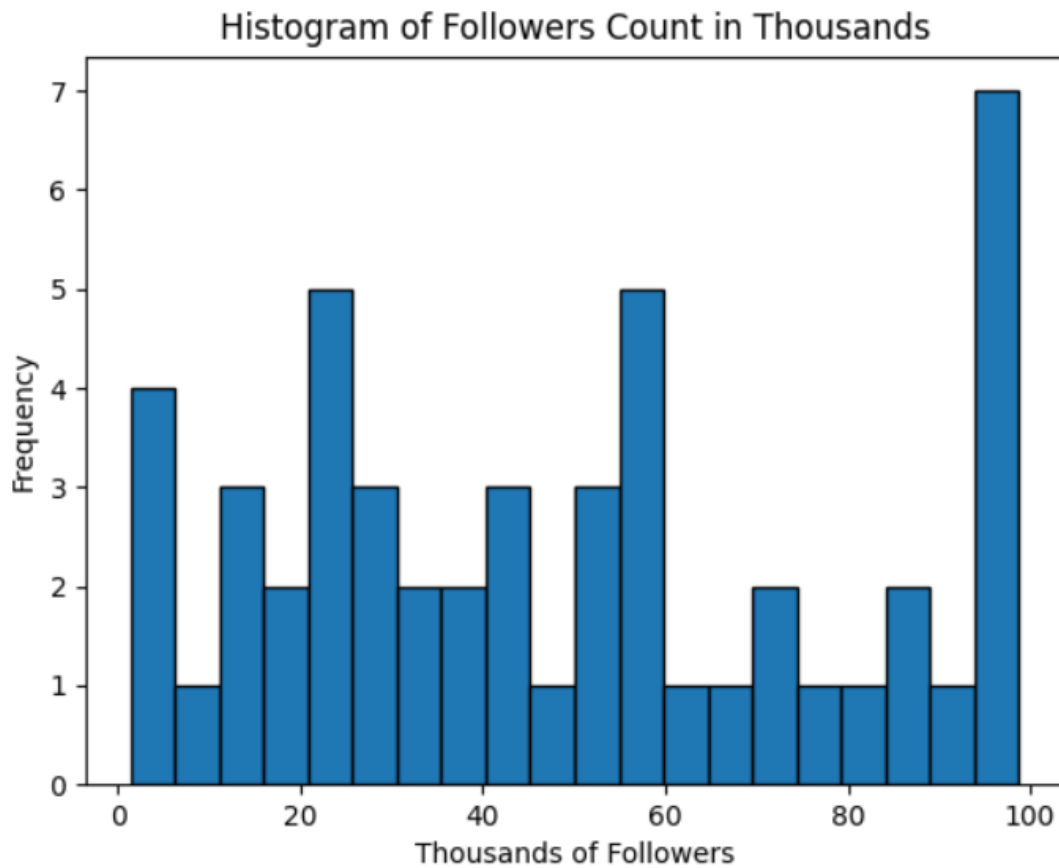


Рис. 1 Гістограма

2. Кругова діаграма, що показує співвідношення жінок до чоловіків, наведена на рисунку 2:

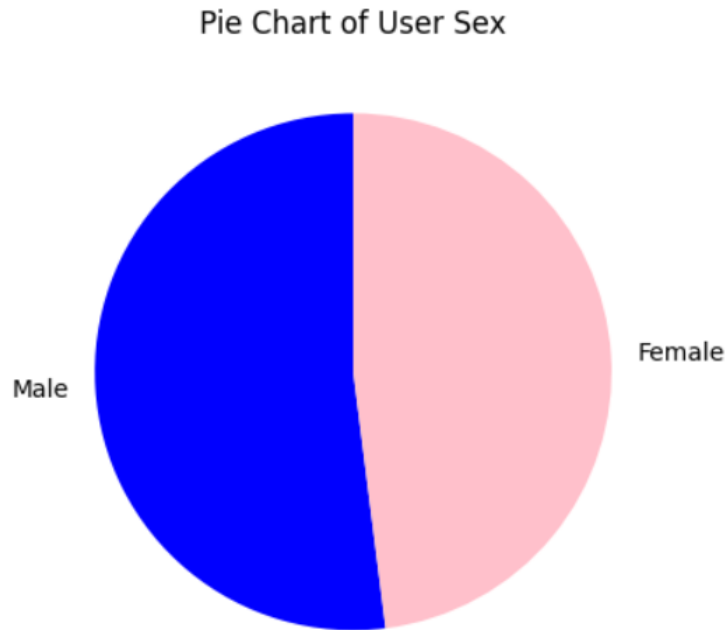


Рис. 2 Кругова діаграма

3. Стовпчикова діаграма, що порівнює кількість користувачів за локаціями, наведена на рисунку 3:

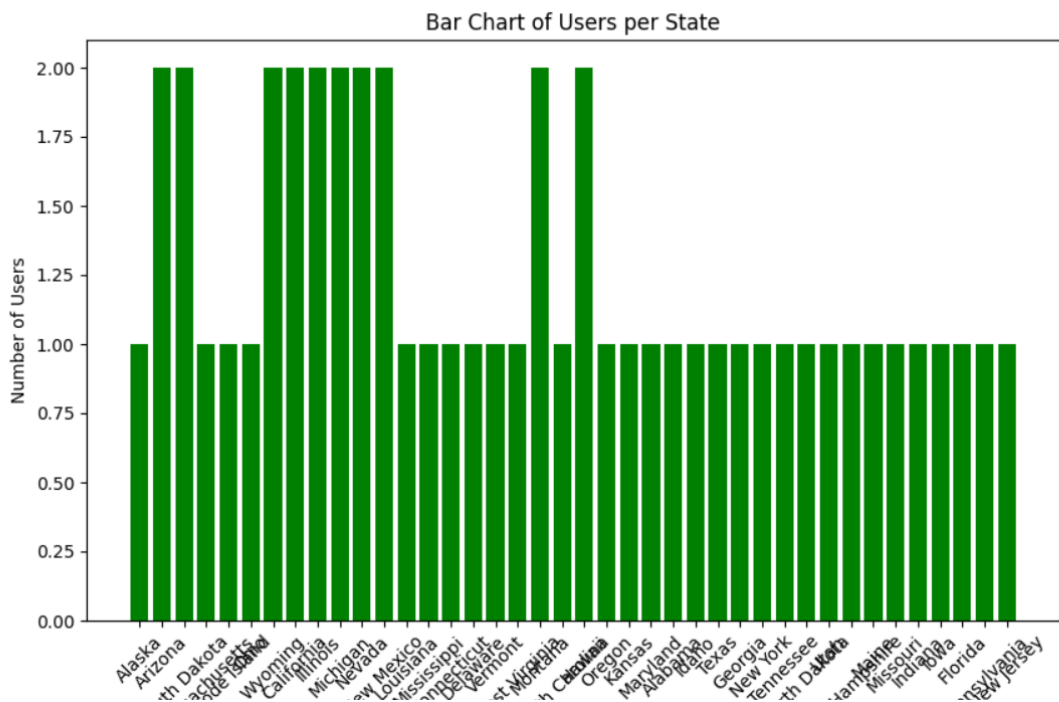


Рис. 3 Стовпчикова діаграма

4. Комбінована діаграма наведена на рисунку 4:

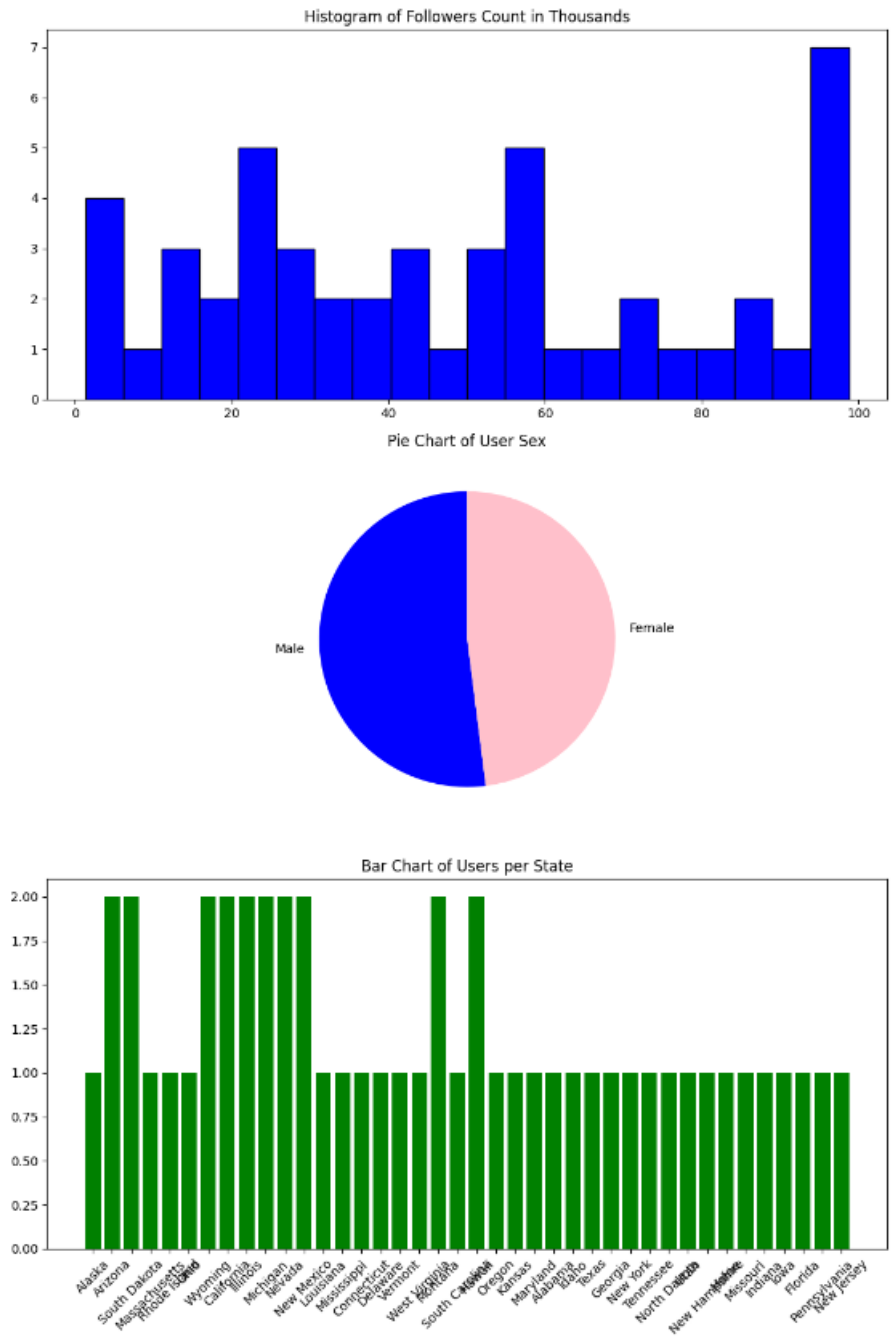


Рис. 4 Комбінована діаграма

Демонстрація виконання програми наведена на рисунку 5:

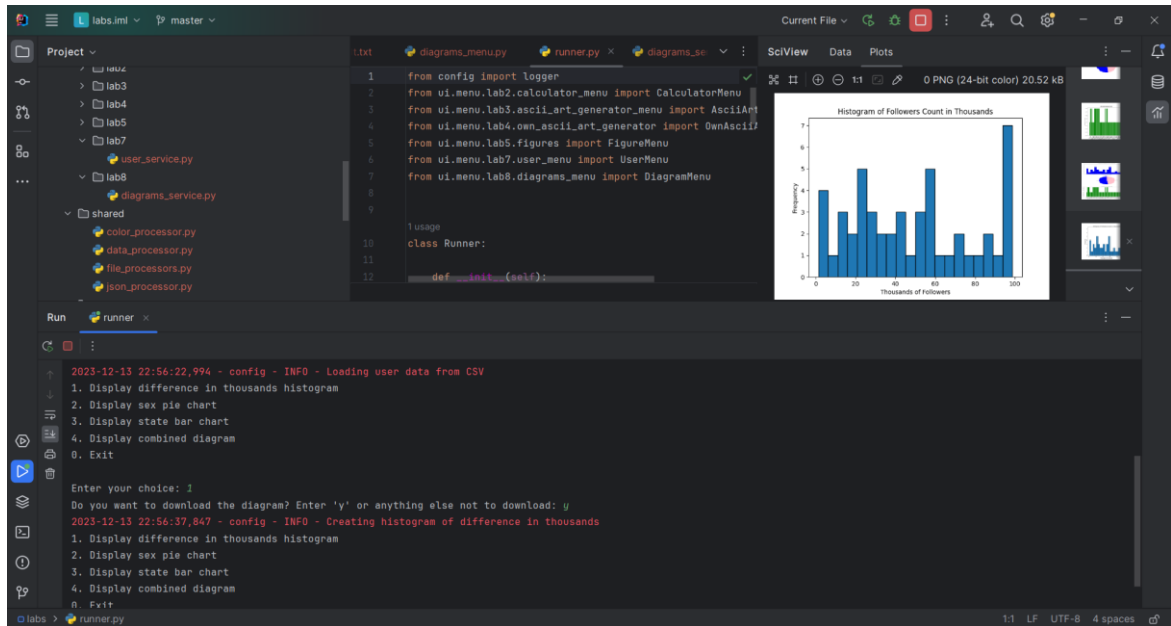


Рис. 5 Виконання програми

Посилання на GitHub репозиторій: <https://github.com/Deadmarvald/smp>

Висновки: Виконавши ці завдання, я створив багатофункціональний додаток для візуалізації CSV-наборів даних за допомогою Matplotlib. Цей проект покращив мої навички візуалізації даних, дозволив досліджувати результати з різноманітними наборами даних.