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,

Садовенко А. Ю.

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,rujlh\_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

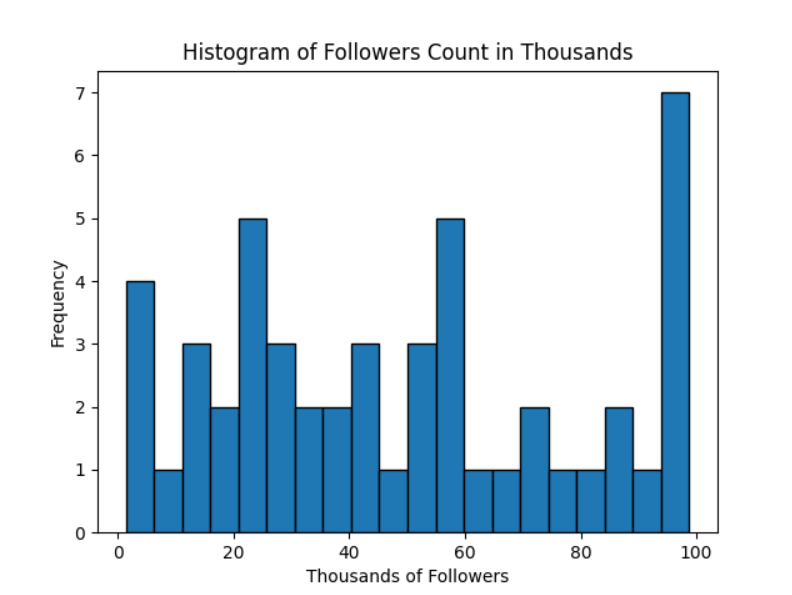
48,66943,qikrt\_618,64825,Michigan,Male,25/06/2014

49,11933,vslon\_632,52911,Pennsylvania,Female,12/06/2023

50,93516,gwmjb\_519,85125,New Jersey,Female,09/11/2014

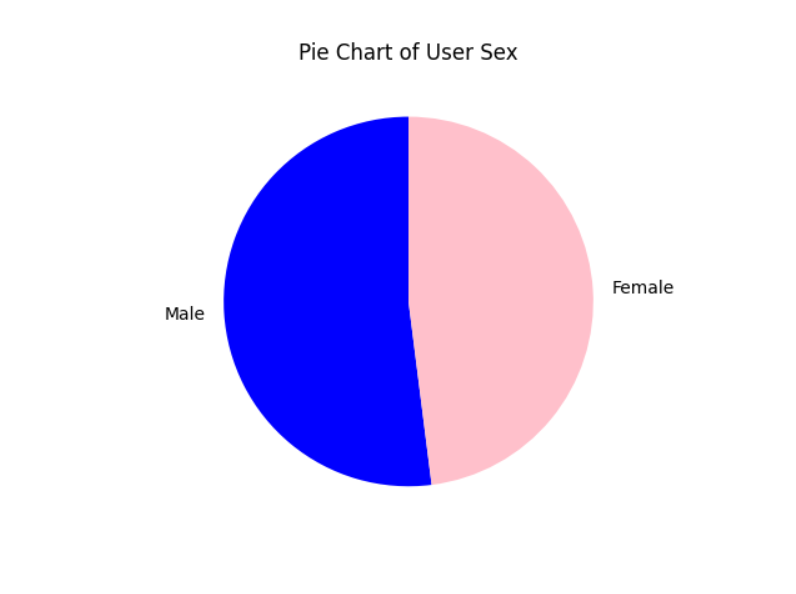
**Результати виконання програми:**

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



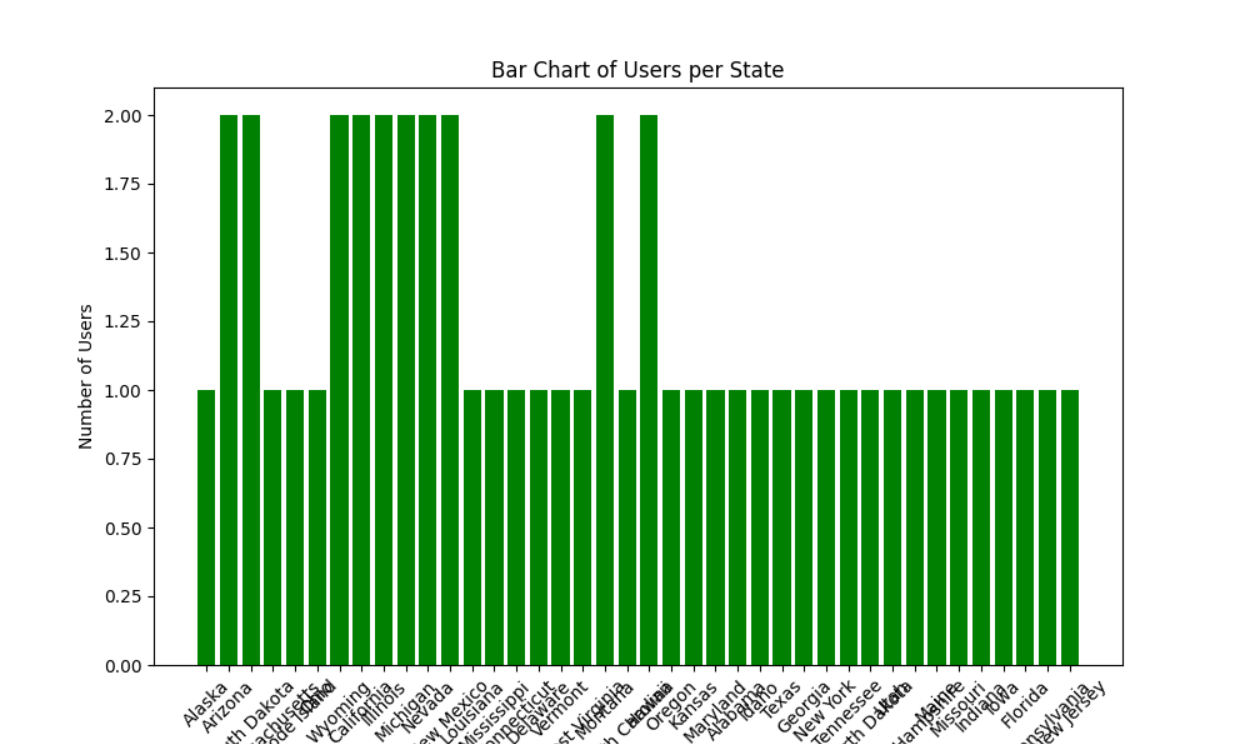
*Рис. 1 Гістограма*

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



*Рис. 2 Кругова діаграма*

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



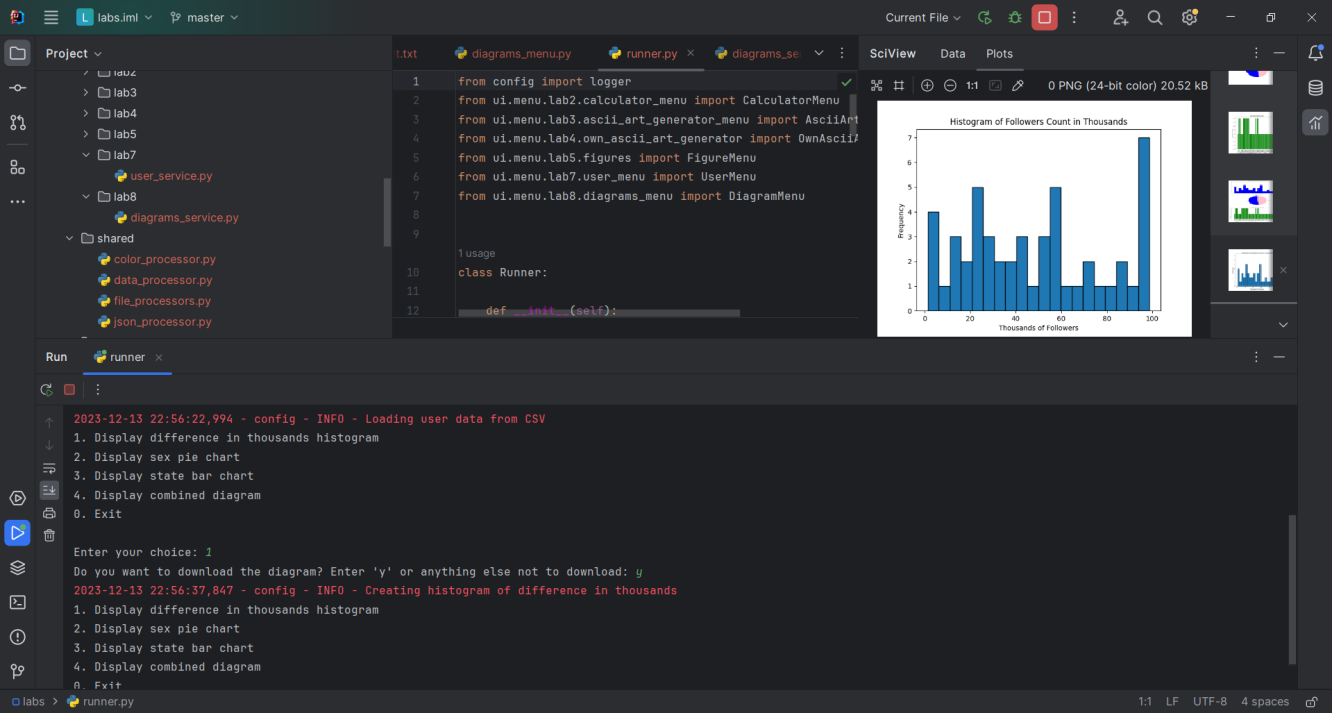
*Рис. 3 Стовпчикова діаграма*

1. 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/ArtemSadovenko/smp/tree/master>

**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.