

Міністерство освіти і науки України
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Факультет інформатики та обчислювальної техніки

Звіт

З лабораторної роботи №1 з дисципліни
«Прикладні задачі машинного навчання»

На тему:

«Введення в data science»

Виконали студенти	<u>ІП-24 Новиков Гліб Костянтинович</u>	<u>13.02.2024</u>
	(шифр, прізвище, ім'я, по батькові)	
	<u>ІК-21 Насібов Рафаель Сафадович</u>	
	(шифр, прізвище, ім'я, по батькові)	
Перевірів	<u>.Нестерук А. О.</u>	<u> </u>
	(прізвище, ім'я, по батькові)	

Київ 2024

Лабораторна робота №1

«Введення в data science»

Мета:

Засвоїти основи роботи з даними, включаючи їх збір, аналіз та візуалізацію, а також використання бібліотеки Python для ефективної обробки та представлення даних.

Постановка задачі лабораторної роботи №1:

1. На сайті <http://www.ukrstat.gov.ua/> обрати дані які для Вас є цікавими, можна використати будь-який ресурс з відкритими даними, та завантажте дані.
2. Знайти математичне сподівання, медіану, моду, дисперсію, середньоквадратичне відхилення (поясніть їх зміст).
3. Візуалізувати завантажені дані за допомогою гістограм.
4. Для цих даних проробити всі дії з пункту колекції Series і DataFrame бібліотеки pandas.
5. Виконати первинну обробку даних
6. Прочитати набір даних катастрофи «Титаніка»
Набір даних катастрофи «Титаніка» належить до числа найпопулярніших наборів даних машинного навчання і доступний в багатьох форматах, включаючи CSV:
<https://vincentarelbundock.github.io/Rdatasets/datasets.html>.
7. Завантажити набір даних катастрофи «Титаніка» за URL адресою. Якщо у вас є URL-адресу, що представляє набір даних в форматі CSV, то ви можете завантажити його в DataFrame функцією `read_csv` - припустимо, з GitHub.
8. Переглянути рядки набору даних катастрофи «Титаніка».
Набір даних містить понад 1300 рядків, кожен рядок це інформація про одного пасажера. За даними «Вікіпедії», на борту було приблизно 1317 пасажирів, а 815 з них загинули. Для великих наборів даних при виведенні DataFrame показуються тільки перші 30 рядків, потім йде

три крапки «...» і останні 30 рядків.

Для економії місця перегляньте перші і останні п'ять рядків за допомогою методів `head` і `tail` колекції `DataFrame`. Обидва методи за замовчуванням повертають п'ять рядків, але число виведених рядків можна передати в аргументі.

9. Налаштувати назви стовпців

Назва першого стовпчика в наборі даних виглядає досить дивно ('Unnamed: 0'). Цю проблему можна вирішити налаштуванням імен стовпців.

Замініть 'Unnamed: 0' на 'name' і скоротіть 'passengerClass' до 'class'.

10. Провести простий аналіз даних: визначте наймолодшого пасажирів, найстаршого, який був середній вік пасажирів та статистику по пасажирів які вижили. Відсортуйте всіх жінок з кают 1-го класу, знайдіть наймолодшу та найстаршу серед них, кількість виживших. Зверніть увагу на розбіжності в значенні `count` (1046) і кількості рядків даних в наборі даних (1309 - при виклику `tail` індекс останнього рядка дорівнював 1308). Тільки 1046 рядків даних (значення `count`) містили значення `age`. Решта результатів були відсутні і були помічені `NaN`, як в рядку 1305. При виконанні обчислень бібліотека `pandas` за замовчуванням ігнорує відсутні дані (`NaN`).

11. Побудувати гістограму віку пасажирів.

Візуалізація - хороший спосіб ближче познайомитися з даними. `Pandas` містить багато вбудованих засобів візуалізації, реалізованих на базі `Matplotlib`. Щоб використовувати їх, спочатку включіть підтримку `Matplotlib` в `Ipython`.

Гістограма наочно показує розподіл числових даних за діапазоном значень. Метод `hist` колекції `DataFrame` автоматично аналізує дані кожного числового стовпця і будує відповідну гістограму. Щоб переглянути гістограми по кожному числовому стовпці даних, викличте `hist` для своєї колекції `DataFrame`.

Обрані дані

Обраними даними для лабораторної роботи були «Регіональні обсяги зовнішньої торгівлі товарами за січень–вересень 2023 року»:

Regional volumes of foreign trade in Goods, January–September 2023 ¹								
	Експорт/Exports			Імпорт/Imports			Сальдо/ Balance	
	тис.дол. США/ thsd. USD	у % до 9 місяців 2022/in % to January- September 2022	у % до загального обсягу/ % of the total volume	тис.дол. США/ thsd. USD	у % до 9 місяців 2022/in % to January- September 2022	у % до загального обсягу/ % of the total volume		
Україна	27144776,9	82,2	100,0	46547149,3²	119,2	100,0	#####	Ukraine
у тому числі								of which
Вінницька	1133651,2	104,5	4,2	672785,7	151,3	1,4	460865,5	Vinnitsya
Волинська	645737,7	81,9	2,4	1900030,1	132,4	4,1	-1254292,4	Volyn
Дніпропетровська	3610917,8	69,3	13,3	3386500,0	118,3	7,3	224417,8	Dnipropetrovsk
Донецька	175022,9	73,7	0,6	39227,8	32,9	0,1	135795,1	Donetsk
Житомирська	418267,6	80,2	1,5	904711,9	121,5	1,9	-486444,2	Zhytomyr
Закарпатська	1124405,6	73,0	4,1	1053379,2	83,5	2,3	71026,4	Zakarpattya
Запорізька	1057208,0	42,3	3,9	426065,2	41,9	0,9	631142,8	Zaporizhzhya
Івано– Франківська	474722,1	85,6	1,7	395948,9	82,8	0,9	78773,2	Ivano– Frankivsk
Київська	1325004,2	84,8	4,9	3397623,6	131,9	7,3	-2072619,4	Kyiv
Кіровоградська	612892,8	100,5	2,3	203150,4	113,2	0,4	409742,5	Kirovohrad
Луганська	141,2	1,7	0,0	1028,3	9,5	0,0	-887,1	Luhansk
Львівська	2003860,4	88,8	7,4	4302555,9	112,5	9,2	-2298695,5	Lviv
Миколаївська	766576,3	62,1	2,8	398592,0	64,1	0,9	367984,3	Mykolayiv
Одеська	1369283,8	91,9	5,0	1652727,6	107,9	3,6	-283443,8	Odesa
Полтавська	997829,9	66,1	3,7	631241,8	67,1	1,4	366588,1	Poltava
Рівненська	457773,4	82,8	1,7	573021,8	155,9	1,2	-115248,4	Rivne
Сумська	525873,0	104,5	1,9	404945,9	97,3	0,9	120927,1	Sumy
Тернопільська	555512,2	105,4	2,0	501334,0	138,7	1,1	54178,2	Ternopil
Харківська	573763,0	87,4	2,1	1101910,5	125,9	2,4	-528147,5	Kharkiv
Херсонська	114129,1	214,4	0,4	13988,8	46,2	0,0	100140,3	Kherson
Хмельницька	568266,1	106,3	2,1	490491,9	94,9	1,1	77774,2	Khmelnytskyi
Черкаська	913789,1	106,5	3,4	536814,4	128,0	1,2	376974,7	Cherkasy
Чернівецька	150856,9	72,8	0,6	367854,7	113,0	0,8	-216997,8	Chernivtsi
Чернігівська	592111,4	118,4	2,2	227151,6	106,7	0,5	364959,8	Chernihiv
м.Київ	6892657,6	92,2	25,4	20101776,8	132,3	43,2	-13209119,2	Kyiv city

¹ Дані наведено без урахування тимчасово окупованих російською федерацією територій та частини територій, на яких ведуться (велися) бойові дії/ ¹ Data exclude the territories which are temporarily occupied by the russian federation and part of territories where the military actions are/were conducted.

² З урахуванням не розподілених за регіонами поставок газу природного./ ² Including natural gas supply undistributed among regions.

Примітка. В окремих випадках сума складових може не дорівнювати підсумку у зв'язку з округленням даних./**Note.** In some cases, the sum of the components may not be equal to the result due to rounding data.

Таблиця 1 – зміст xls-файлу «Регіональні обсяги зовнішньої торгівлі товарами за січень–вересень 2023 року»

Код програми на мові Python

import-export.py

```
import pandas as pd
import statistics as s
import numpy as np
import matplotlib.pyplot as plt
import num4 as n4
from matplotlib.gridspec import GridSpec
import os

#1
df=pd.read_excel("stat.xls",skiprows=6,skipfooter=3)
saldo=df["Unnamed: 7"]
names_saldo=df.iloc[:,[0,7]]

print('-----Сальдо по областям України за 2023 рік-----')
print(names_saldo)

#2
print('-----Статистика по сальдо-----')
mean = saldo.mean()
print("mean", mean)
mode = saldo.mode()
print("mode", mode if mode.size < saldo.size else "no mode")
median = s.median(sorted(saldo))
print("median",median)
dysp = s.pvariance(saldo)
print("pvariance",dysp)
std = s.pstdev(saldo)
print("std",std)

#3
gs = GridSpec(1, 5)
fig = plt.figure(figsize=(10, 6))
def hui(plot_column_num:int, column_dict_name:str, column_dict_num:int,
x_label:str, label:str, multiplier=1.0):
    ax = fig.add_subplot(gs[0, plot_column_num])
    cities_data = df.iloc[:, [0, column_dict_num]]
    cities_data_hist = cities_data.to_dict('list')
    ys = cities_data_hist[column_dict_name]
    if multiplier != 1.0:
        temp = []
        for i in cities_data_hist[column_dict_name]:
            temp.append(i * multiplier)
        ys = temp
    ax.barh(cities_data_hist['у тому числі'], ys)
    ax.set_yticklabels(cities_data_hist['у тому числі'], rotation=0)
    xticks = np.linspace(round(min(ys), 2), round(max(ys), 3), num=5)
    ax.set_xticks(xticks)
    ax.set_xticklabels(xticks,rotation=90)
    ax.set_ylabel(x_label)
    ax.set_title(label)
    ax.grid(axis='x')
    return ax

# Сальдо - гістограма
hui(0, 'Unnamed: 7', 7, 'Міста', 'Сальдо, млрд. долл', 10e-7)

# експорт - гістограма
hui(2, 'Unnamed: 1', 1, 'Міста', 'Експорт, млрд. долл', 10e-7)
```

```

# Імпорт - гістограма
hui(4, 'Unnamed: 4', 4, 'Міста', 'Імпорт, млрд. долл', 10e-7)

plt.show()

# 4
print('\n-----')
print('-----Колекція Series-----')
print('-----\n')
n4.do_series(df)

df.index = [df["у тому числі"]]
df = df.iloc[:, 1:7]

print('\n-----')
print('-----Колекція DataFrame-----')
print('-----\n')
n4.do_dataframe(df=df, num_of_column=1, num_of_row_start=1, num_of_row_end=15)

```

num4.py

```

#4
import pandas as pd

def do_series(df:pd.DataFrame):
    print('-----Створення Series з індексами за замовчуванням-----')
    print(df.head())

    print('-----Звернення до елементів Series-----')
    \n'
        (Звернення до колонки)')
    print(df['у тому числі'].head())

    print('-----Обчислення описових статистик для Series-----')
    saldo = df["Unnamed: 7"]
    mean = saldo.mean()
    print("mean", mean)
    mode = saldo.mode()
    print("mode", mode if mode.size < saldo.size else "no mode")
    min = saldo.min()
    print("min", min)
    max = saldo.max()
    print("max", max)
    std = saldo.std()
    print("std", std)

    print('-----Створення колекції Series з нестандартними індексами-----')
    df.index = [df["у тому числі"]]
    df = pd.DataFrame(df.iloc[:, 1:])
    df.columns = ['експ. тис. дол.', 'експ. у % до 9 міс.', 'експ. у % до заг.',
                  'імп. тис. дол.', 'імп. у % до 9 міс.', 'імп. у % до заг.',
                  'сальдо', 'region']
    print(df.head())

    print('-----Інші дії-----')
    print('Тип даних: ', df.dtypes)
    print('\nПовернення базової колекції:\n', df.values, '\n\n\n')

def

```

```

do_dataframe(df:pd.DataFrame,num_of_cloumn:int,num_of_row_start:int,num_of_row_end:int):
    print('-----Звернення до колонки-----')
    print(df[f'Unnamed: {num_of_cloumn}'])
    df=df.rename(columns={'Unnamed: 1':'|у тому числі|',
                           'Unnamed: 2':'|експорт, тис USD|',
                           'Unnamed: 3':'|експорт у % до 2022|',
                           'Unnamed: 4':'|експорт у % до загального|',
                           'Unnamed: 5':'|імпорт, тис USD|',
                           'Unnamed: 6':'|імпорт у % до 2022|',
                           'Unnamed: 7':'|імпорт % до загального|'})
    print('-----Звернення до стовпців-----')
    print(df.iloc[num_of_row_start])
    print('-----Вібір підмножини-----')
    print(df.iloc[num_of_row_start:num_of_row_end])
    print('-----Логічне індексування-----')
    print(df[df>=45])
    print('-----Звернення до конкретного осередку[4,5]-----')
    print(df.iat[4,5])
    print('-----Описова статистика-----')
    print(df.describe())
    print('-----Транспонування-----')
    print(df.T)
    print('-----Сортування рядків за індексами (спадання)-----')
    print(df.sort_index(ascending=False))
    print('-----Сортування стовпців за індексами (спадання)-----')
    print(df.sort_index(ascending=False,axis=1))

```

titanic.py

```

import pandas as pd
import matplotlib.pyplot as plt

# 8
titanic =
pd.read_csv('https://vincentarelbundock.github.io/Rdatasets/csv/carData/TitanicSurvival.csv')

# 9
titanic.columns = ['name', 'survived', 'sex', 'age', 'class']

print('-----Зчитані дані-----')
print(titanic)

# 10
youngest = titanic.loc[titanic['age'].idxmin()]
eldest = titanic.loc[titanic['age'].idxmax()]
average = titanic['age'].mean()

first_class_women = titanic[(titanic['sex'] == 'female') & (titanic['class'] == '1st')].sort_values(by='name')
youngest_fcw = first_class_women.loc[first_class_women['age'].idxmin()]
eldest_fcw = first_class_women.loc[first_class_women['age'].idxmax()]
count_survived = first_class_women[first_class_women['survived'] == 'yes'].shape[0]

print('-----youngest person-----')
print(youngest)
print('-----eldest person-----')
print(eldest)

```

```

print('-----average age-----')
print(average)
print('-----women of 1st class-----')
print(first_class_women)
print('-----youngest women of 1st class-----')
print(youngest_fcw)
print('-----eldest women of 1st class-----')
print(eldest_fcw)
print('-----quantity of survived women of 1st class-----')
print(count_survived)

#11
titanic.hist()
plt.xlabel("Вік пасажирів")
plt.ylabel("Частота")
plt.show()

```

Результат виконання програми

```

-----Сальдо по областях України за 2023 рік-----
      у тому числі      Unnamed: 7
0      Вінницька      4.608655e+05
1      Волинська      -1.254292e+06
2      Дніпропетровська      2.244178e+05
3      Донецька      1.357951e+05
4      Житомирська      -4.864442e+05
5      Закарпатська      7.102640e+04
6      Запорізька      6.311428e+05
7      Івано-Франківська      7.877320e+04
8      Київська      -2.072619e+06
9      Кіровоградська      4.097425e+05
10     Луганська      -8.870604e+02
11     Львівська      -2.298696e+06
12     Миколаївська      3.679843e+05
13     Одеська      -2.834438e+05
14     Полтавська      3.665881e+05
15     Рівненська      -1.152484e+05
16     Сумська      1.209271e+05

17     Тернопільська      5.417822e+04
18     Харківська      -5.281475e+05
19     Херсонська      1.001403e+05
20     Хмельницька      7.777420e+04
21     Черкаська      3.769747e+05
22     Чернівецька      -2.169978e+05
23     Чернігівська      3.649598e+05
24     м.Київ      -1.320912e+07

-----Статистика по сальдо-----
mean -664984.2119012
mode no mode
median 77774.2
pvariance 7066897262119.849
std 2658363.6436950923

```

Рис. 1 – Вивід «очищених» даних із вхідного файлу та вивід статистичних даних по сальдо.

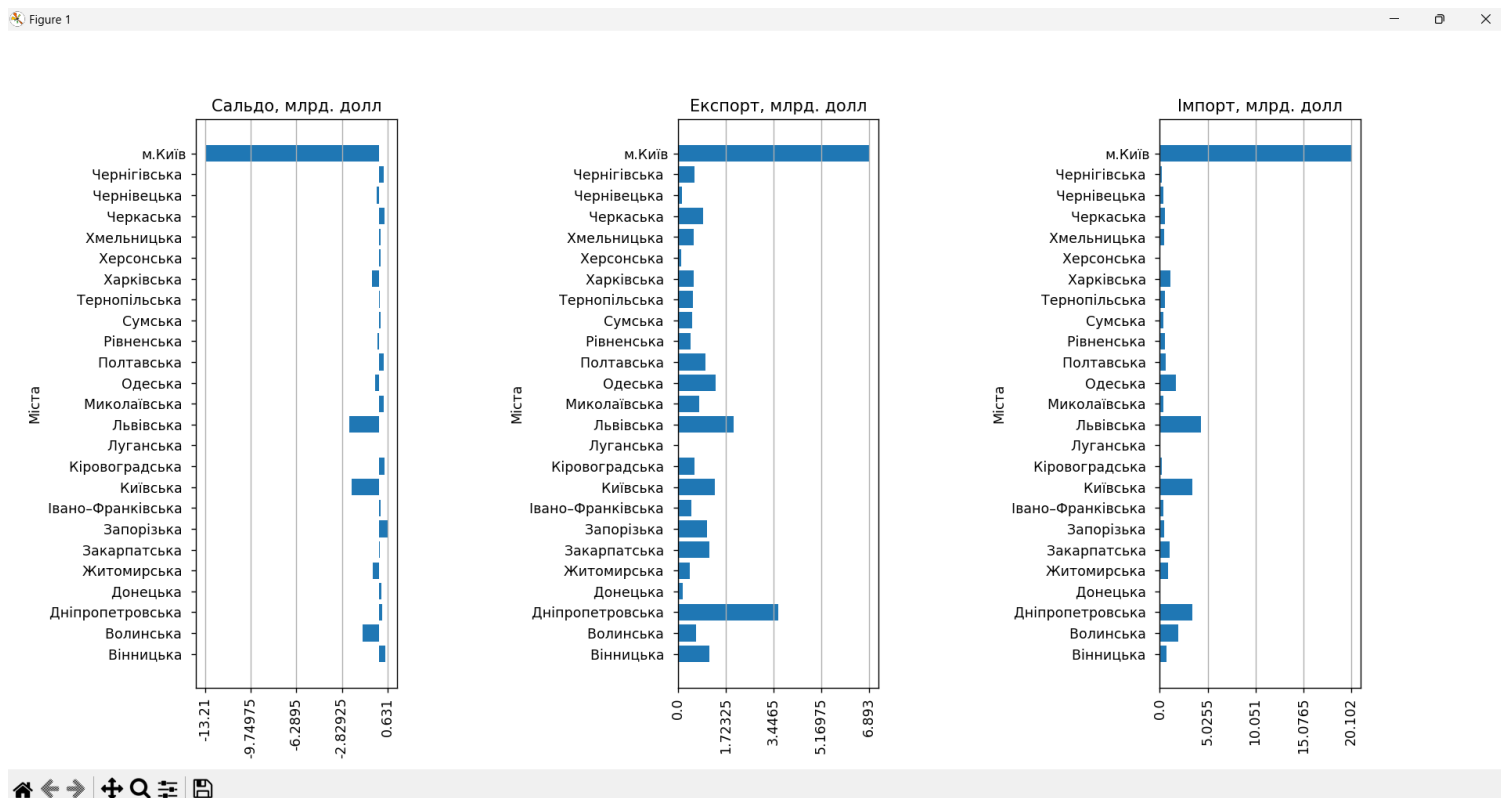


Рис. 2 – Гістограми по сальдо, експорту та імпорту

Далі слідє вивід для виконання 4 пункту завдання лабораторної роботи «Для цих даних проробити всі дії з пункту колекції Series і DataFrame бібліотеки pandas.»

```

-----
-----Колекція Series-----
-----

-----Створення Series з індексами за замовчуванням-----
      у тому числі      Unnamed: 1  ...      Unnamed: 7      of which
0      Вінницька      1.133651e+06  ...      4.608655e+05      Vinnytsya
1      Волинська      6.457377e+05  ...      -1.254292e+06      Volyn
2      Дніпропетровська      3.610918e+06  ...      2.244178e+05      Dnipropetrovsk
3      Донецька      1.750229e+05  ...      1.357951e+05      Donetsk
4      Житомирська      4.182676e+05  ...      -4.864442e+05      Zhytomyr

[5 rows x 9 columns]
-----Звернення до елементів Series-----
      (Звернення до колонки)
0      Вінницька
1      Волинська
2      Дніпропетровська
3      Донецька
4      Житомирська
Name: у тому числі, dtype: object

```

Рис. 3 – Series з індексами за замовчуванням

```

-----Обчислення описових статистик для Series-----
mean -664984.2119012
mode no mode
min -13209119.18835
max 631142.76522
std 2713181.0324245184
-----Створення колекції Series з нестандартними індексами-----
      експ. тис. дол.  ...      region
у тому числі      ...
Вінницька      1.133651e+06  ...      Vinnytsya
Волинська      6.457377e+05  ...      Volyn
Дніпропетровська      3.610918e+06  ...      Dnipropetrovsk
Донецька      1.750229e+05  ...      Donetsk
Житомирська      4.182676e+05  ...      Zhytomyr

[5 rows x 8 columns]

```

Рис. 4 – Статистика описових даних та нестандартні індекси

```

-----Інші дії-----
Тип даних:  експ. тис. дол.          float64
експ. у % до 9 міс.      float64
експ. у % до заг.        float64
імп. тис. дол.           float64
імп. у % до 9 міс.       float64
імп. у % до заг.         float64
сальдо                    float64
region                    object
dtype: object

```

Повернення базової колекції:

```

[[1133651.20004 104.5 4.17631429602384 672785.7 151.25 1.44538564355433
 460865.5 'Vinnytsya']
[645737.71718 81.9 2.37886543907462 1900030.1 132.4 4.08194597670641
-1254292.4 'Volyn']
[3610917.81017 69.26 13.302440531839 3386500.0 118.26 7.27541891183632
224417.8 'Dnipropetrovsk']
[175022.92193 73.7 0.644775686703575 39227.80408 32.94
0.0842754168084239 135795.11785 'Donetsk']
[418267.64174 80.16 1.54087706315779 904711.9 121.53 1.9436456515193
-486444.2 'Zhytomyr']
[1124405.59622 73.04 4.14225395417665 1053379.19719 83.51
2.26303697039649 71026.3990299999 'Zakarpattya']
[1057207.9587 42.31 3.89470121994596 426065.19348 41.94

```

Рис. 5 – інші дії з колекції Series

```

-----
-----Колекція DataFrame-----
-----

-----Зернення до колонки-----
у тому числі
Вінницька      1.133651e+06
Волинська      6.457377e+05
Дніпропетровська 3.610918e+06
Донецька      1.750229e+05
Житомирська   4.182676e+05
Закарпатська  1.124406e+06
Запорізька    1.057208e+06
Івано-Франківська 4.747221e+05
Київська      1.325004e+06
Кіровоградська 6.128928e+05
Луганська     1.412289e+02
Львівська     2.003860e+06
Миколаївська  7.665763e+05
Одеська       1.369284e+06
Полтавська    9.978299e+05
Рівненська    4.577734e+05
Сумська       5.258730e+05
Тернопільська 5.555122e+05
Харківська    5.737630e+05
Хмельницька   1.141201e+06

```

Рис. 6 – Робота з колекцією DataFrame

```

-----Звернення до стовпців-----
|у тому числі|          6.457377e+05
|експорт, тис USD|      8.190000e+01
|експорт у % до 2022|   2.378865e+00
|експорт у % до загального| 1.900030e+06
|імпорт, тис USD|       1.324000e+02
|імпорт у % до 2022|    4.081946e+00
Name: (Волинська ,), dtype: float64
-----Відбір підмножини-----
          |у тому числі|  ...  |імпорт у % до 2022|
у тому числі          ...
Волинська             6.457377e+05  ...             4.081946
Дніпропетровська     3.610918e+06  ...             7.275419
Донецька             1.750229e+05  ...             0.084275
Житомирська         4.182676e+05  ...             1.943646
Закарпатська         1.124406e+06  ...             2.263037
Запорізька           1.057208e+06  ...             0.915341
Івано-Франківська    4.747221e+05  ...             0.850640
Київська             1.325004e+06  ...             7.299314
Кіровоградська       6.128928e+05  ...             0.436439
Луганська            1.412289e+02  ...             0.002209
Львівська            2.003860e+06  ...             9.243437
Миколаївська         7.665763e+05  ...             0.856319

```

Рис. 7 – Звернення до стовпців та відбір підмножини

-----Логічне індексування-----			
	у тому числі	...	імпорт у % до 2022
у тому числі		...	
Вінницька	1.133651e+06	...	NaN
Волинська	6.457377e+05	...	NaN
Дніпропетровська	3.610918e+06	...	NaN
Донецька	1.750229e+05	...	NaN
Житомирська	4.182676e+05	...	NaN
Закарпатська	1.124406e+06	...	NaN
Запорізька	1.057208e+06	...	NaN
Івано-Франківська	4.747221e+05	...	NaN
Київська	1.325004e+06	...	NaN
Кіровоградська	6.128928e+05	...	NaN
Луганська	1.412289e+02	...	NaN
Львівська	2.003860e+06	...	NaN
Миколаївська	7.665763e+05	...	NaN
Одеська	1.369284e+06	...	NaN
Полтавська	9.978299e+05	...	NaN
Рівненська	4.577734e+05	...	NaN
Сумська	5.258730e+05	...	NaN
Тернопільська	5.555122e+05	...	NaN
Харківська	5.737630e+05	...	NaN
Херсонська	1.141291e+05	...	NaN

Рис. 8 – Логічне індексування

```

-----Звернення до конкретного осередку[4,5]-----
1.9436456515193
-----Описова статистика-----
      |у тому числі| ... |імпорт у % до 2022|
count  2.500000e+01 ...          25.000000
mean   1.082410e+06 ...          3.754031
std     1.414882e+06 ...          8.566603
min     1.412289e+02 ...          0.002209
25%     4.747221e+05 ...          0.850640
50%     6.128928e+05 ...          1.153270
75%     1.124406e+06 ...          2.367299
max     6.892658e+06 ...          43.185839

[8 rows x 6 columns]
-----Транспонування-----
у тому числі          Вінницька ...          м.Київ
|у тому числі|          1.133651e+06 ...  6.892658e+06
|експорт, тис USD|      1.045000e+02 ...  9.224000e+01
|експорт у % до 2022|    4.176314e+00 ...  2.539221e+01
|експорт у % до загального| 6.727857e+05 ...  2.010178e+07
|імпорт, тис USD|        1.512500e+02 ...  1.322500e+02
|імпорт у % до 2022|     1.445386e+00 ...  4.318584e+01

```

Рис. 9 – Звернення до осередку, описова статистика та транспонування

-----Сортування рядків за індексами (спадання)-----			
	у тому числі	...	імпорт у % до 2022
у тому числі		...	
м.Київ	6.892658e+06	...	43.185839
Чернігівська	5.921114e+05	...	0.488004
Чернівецька	1.508569e+05	...	0.790283
Черкаська	9.137891e+05	...	1.153270
Хмельницька	5.682661e+05	...	1.053753
Херсонська	1.141291e+05	...	0.030053
Харківська	5.737630e+05	...	2.367299
Тернопільська	5.555122e+05	...	1.077046
Сумська	5.258730e+05	...	0.869970
Рівненська	4.577734e+05	...	1.231057
Полтавська	9.978299e+05	...	1.356133
Одеська	1.369284e+06	...	3.550651
Миколаївська	7.665763e+05	...	0.856319
Львівська	2.003860e+06	...	9.243437
Луганська	1.412289e+02	...	0.002209
Київська	1.325004e+06	...	7.299314
Кіровоградська	6.128928e+05	...	0.436439
Запорізька	1.057208e+06	...	0.915341
Закарпатська	1.124406e+06	...	2.263037
Житомирська	4.182676e+05	...	1.943646
-----Сортування стовпців за індексами (спадання)-----			
	імпорт, тис USD	...	експорт у % до 2022
у тому числі		...	
Вінницька	151.25	...	4.176314
Волинська	132.40	...	2.378865
Дніпропетровська	118.26	...	13.302441
Донецька	32.94	...	0.644776
Житомирська	121.53	...	1.540877
Закарпатська	83.51	...	4.142254
Запорізька	41.94	...	3.894701
Івано-Франківська	82.77	...	1.748852
Київська	131.87	...	4.881249
Кіровоградська	113.16	...	2.257866
Луганська	9.47	...	0.000520
Львівська	112.46	...	7.382121
Миколаївська	64.08	...	2.824029
Одеська	107.90	...	5.044373
Полтавська	67.14	...	3.675956
Рівненська	155.85	...	1.686414
Сумська	97.33	...	1.937290
Тернопільська	138.69	...	2.046479
Харківська	125.86	...	2.113714

Рис. 10 – Сортування рядків

Далі слідує результат виконання програми для даних катастрофи Титаніка.

```
-----Зчитані дані-----
      name survived      sex      age class
0      Allen, Miss. Elisabeth Walton      yes  female  29.0000  1st
1      Allison, Master. Hudson Trevor      yes   male   0.9167  1st
2      Allison, Miss. Helen Loraine      no   female   2.0000  1st
3      Allison, Mr. Hudson Joshua Crei      no    male  30.0000  1st
4      Allison, Mrs. Hudson J C (Bessi      no   female  25.0000  1st
...
1304      Zabour, Miss. Hileni      no   female  14.5000  3rd
1305      Zabour, Miss. Thamine      no   female   NaN   3rd
1306      Zakarian, Mr. Mapriededer      no    male  26.5000  3rd
1307      Zakarian, Mr. Ortin      no    male  27.0000  3rd
1308      Zimmerman, Mr. Leo      no    male  29.0000  3rd

[1309 rows x 5 columns]
```

Рис. 11 – Зчитані дані

```

-----youngest person-----
name      Dean, Miss. Elizabeth Gladys M
survived                                     yes
sex                                              female
age                                             0.1667
class                                          3rd
Name: 763, dtype: object
-----eldest person-----
name      Barkworth, Mr. Algernon Henry W
survived                                     yes
sex                                              male
age                                             80.0
class                                          1st
Name: 14, dtype: object
-----average age-----
29.881134512434034

-----women of 1st class-----

```

	name	survived	sex	age	class
0	Allen, Miss. Elisabeth Walton	yes	female	29.0	1st
2	Allison, Miss. Helen Loraine	no	female	2.0	1st
4	Allison, Mrs. Hudson J C (Bessi	no	female	25.0	1st
6	Andrews, Miss. Kornelia Theodos	yes	female	63.0	1st
8	Appleton, Mrs. Edward Dale (Cha	yes	female	53.0	1st
..
311	Wick, Mrs. George Dennick (Mary	yes	female	45.0	1st
314	Widener, Mrs. George Dunton (El	yes	female	50.0	1st
315	Willard, Miss. Constance	yes	female	21.0	1st
319	Wilson, Miss. Helen Alice	yes	female	31.0	1st
322	Young, Miss. Marie Grice	yes	female	36.0	1st

```

[144 rows x 5 columns]
-----youngest women of 1st class-----
name      Allison, Miss. Helen Loraine
survived                                     no
sex                                              female
age                                             2.0
class                                          1st
Name: 2, dtype: object
-----eldest women of 1st class-----
name      Cavendish, Mrs. Tyrell William
survived                                     yes
sex                                              female
age                                             76.0
class                                          1st
Name: 61, dtype: object
-----quantity of survived women of 1st class-----
139

```

Рис. 12 – Проведення простого аналізу даних із умови завдання

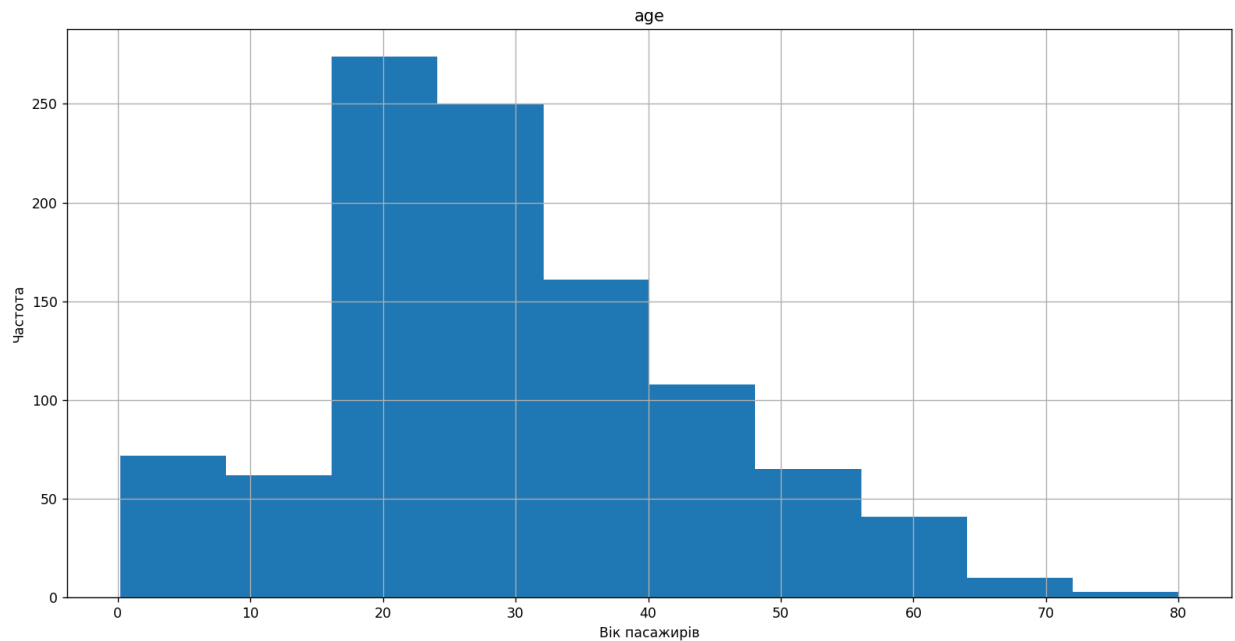


Рис. 13 – Побудована гістограма віку пасажирів

Висновок: У результаті виконання лабораторної роботи №1 «Введення в data science» було засвоєно основи роботи з даними, включаючи їх збір, аналіз та візуалізацію, а також використання бібліотеки Python для ефективної обробки та представлення даних.

Були обрані та завантажені дані про зовнішню торгівлю України за січень–вересень 2023 року та набір даних катастрофи «Титаніка».

Проведено аналіз обраних даних, зокрема розраховано основні статистичні показники, такі як математичне сподівання, медіана, мода, дисперсія, середньоквадратичне відхилення.

Було візуалізовано дані за допомогою гістограм.

Отже, лабораторна робота дозволила вивчити та практично застосувати основні концепції та інструменти data science на конкретних даних, що відкриває шлях для подальших досліджень та аналізу в цій області, і, найголовніше, відкриває шлях до машинного навчання.