

Міністерство освіти і науки України  
Національний технічний університет України «Київський політехнічний  
інститут імені Ігоря Сікорського»

Факультет інформатики та обчислювальної техніки

Кафедра інформатики та програмної інженерії

Звіт

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

«Введення в data science»

Варіант 7

Виконав студент:        ПІ-12 Васильєв Єгор Костянтинович

Перевірив:                Нестерук Андрій Олександрович

Київ 2022

- |    | A                                             | B       | C            | D      | E      | F        | G       | H       | I       | J       | K       | L        | M       | N        | O       |
|----|-----------------------------------------------|---------|--------------|--------|--------|----------|---------|---------|---------|---------|---------|----------|---------|----------|---------|
| 1  | Товари (послуги) -представники                | Одиниця | Грудень 2021 | Счень  | Лютий  | Березень | Квітень | Травень | Червень | Липень  | Серпень | Вересень | Жовтень | Листопад | Грудень |
| 2  | Рис                                           | кг      | 28,74        | 29,30  | 30,19  | 33,12    | 40,11   | 46,63   | 50,41   | 52,95   | 56,07   | 58,46    | 59,56   | 59,78    | 59,65   |
| 3  | Хліб пшеничний з борошна вищого гатунку       | кг      | 33,08        | 33,61  | 34,96  | 36,20    | 37,32   | 37,86   | 38,06   | 38,52   | 38,95   | 39,10    | 39,48   | 40,00    | 40,94   |
| 4  | Хліб пшеничний з борошна першого гатунку      | кг      | 26,29        | 27,11  | 28,29  | 29,50    | 29,77   | 30,64   | 31,65   | 32,10   | 32,00   | 32,20    | 32,76   | 32,63    | 33,23   |
| 5  | Хліб житній, житньо-пшеничний                 |         | 27,46        | 29,73  | 30,93  | 32,28    | 31,50   | 31,84   | 33,01   | 33,44   | 33,58   | 33,65    | 34,46   | 34,53    | 35,11   |
| 6  | Макарони виробні з м'яких сортів пшениці      | 500 г   | 17,88        | 18,10  | 18,63  | 18,93    | 19,67   | 19,74   | 20,27   | 20,64   | 20,74   | 20,86    | 21,06   | 21,25    | 21,48   |
| 7  | Борошно пшеничне                              | кг      | 24,36        | 24,95  | 25,16  | 26,85    | 30,13   | 31,05   | 32,34   | 32,55   | 32,35   | 32,50    | 32,36   | 31,79    | 31,86   |
| 8  | Борошно пшеничне                              | кг      | 15,01        | 15,48  | 15,65  | 16,81    | 18,34   | 18,08   | 18,29   | 17,41   | 16,82   | 16,73    | 16,83   | 17,28    | 17,79   |
| 9  | Крупа манна                                   | кг      | 18,18        | 18,85  | 19,10  | 20,07    | 22,69   | 23,67   | 24,34   | 24,05   | 24,24   | 23,65    | 23,47   | 22,80    | 22,43   |
| 10 | Крупа гречана                                 | кг      | 47,77        | 47,61  | 48,66  | 54,22    | 61,72   | 69,25   | 81,23   | 93,23   | 96,01   | 89,81    | 82,69   | 79,16    | 74,69   |
| 11 | Крупа ячмінна                                 | кг      | 15,80        | 16,41  | 16,96  | 18,30    | 21,05   | 21,61   | 21,70   | 21,09   | 19,81   | 19,47    | 18,62   | 17,95    | 17,70   |
| 12 | Пшоно                                         | кг      | 18,42        | 18,74  | 19,45  | 21,48    | 25,10   | 28,15   | 28,10   | 27,95   | 28,85   | 29,27    | 29,78   | 29,27    | 29,78   |
| 13 | Крупа пшенична                                | —       | —            | 17,42  | 17,65  | 19,02    | 21,06   | 21,24   | 21,16   | 20,91   | 19,98   | 20,05    | 20,02   | 19,16    | 18,77   |
| 14 | Яловичина                                     |         | 182,84       | 195,61 | 208,02 | 221,45   | 216,39  | 211,23  | 205,59  | 200,63  | 200,20  | 200,92   | 203,45  | 200,65   | 203,18  |
| 15 | Свинина                                       |         | 123,75       | 123,95 | 123,50 | 130,60   | 130,85  | 132,26  | 140,24  | 144,35  | 152,84  | 164,29   | 174,45  | 166,87   | 168,56  |
| 16 | Птиця (тушки курчат)                          | кг      | 70,82        | 73,52  | 74,12  | 74,81    | 76,54   | 80,83   | 82,79   | 82,43   | 82,93   | 83,29    | 83,49   | 82,93    | 83,29   |
| 17 | Філе курчаче                                  | кг      | 117,10       | 122,65 | 123,63 | 127,65   | 127,28  | 128,34  | 131,97  | 133,60  | 136,38  | 137,66   | 138,81  | 135,72   | 136,58  |
| 18 | Ковбаси варені першого гатунку                | кг      | 120,15       | 124,41 | 125,50 | 130,85   | 133,28  | 131,14  | 136,90  | 137,94  | 142,64  | 146,51   | 151,61  | 154,08   | 156,69  |
| 19 | Ріба морожена                                 | кг      | 87,86        | 103,70 | 105,83 | 114,47   | 123,17  | 128,74  | 130,60  | 133,09  | 139,09  | 144,14   | 150,00  | 150,53   | 155,11  |
| 20 | Молоко пастеризоване жирністю до 2,6% включно | 1000 г  | 28,55        | 28,93  | 29,37  | 30,53    | 31,46   | 31,92   | 32,29   | 32,61   | 32,40   | 32,90    | 33,32   | 33,86    | 34,28   |
| 21 | Сирів м'яких жирні                            | кг      | 125,98       | 133,60 | 139,45 | 144,58   | 147,35  | 149,68  | 146,76  | 147,15  | 148,40  | 148,12   | 150,24  | 150,24   | 150,24  |
| 22 | Сметана жирністю до 15% включно               | кг      | 74,11        | 75,12  | 76,56  | 80,75    | 82,38   | 83,97   | 85,00   | 85,07   | 86,60   | 87,99    | 89,38   | 92,77    | 94,11   |
| 23 | Яйця                                          | десяток | 33,79        | 35,46  | 33,44  | 34,26    | 24,80   | 19,39   | 22,72   | 24,70   | 28,53   | 35,08    | 57,01   | 63,33    | 62,03   |
| 24 | Масло вершкове                                | 200 г   | 50,59        | 52,17  | 54,56  | 56,41    | 58,76   | 59,24   | 59,97   | 60,13   | 60,59   | 61,53    | 63,06   | 63,67    | 65,98   |
| 25 | Олія соняшникова                              | л       | 61,43        | 61,05  | 61,61  | 65,49    | 67,37   | 67,87   | 69,87   | 69,87</ |         |          |         |          |         |

- Первинна обробка даних (заміщення пропущених даних «—», даними типу NaN)

	Грудень 2021	Січень	Лютий
Рис/кг	28.74	29.30000	30.19000
Хліб пшеничний з борошна вищого ґатунку...	33.08	33.61000	34.96000
Хліб пшеничний з борошна першого ґату...	26.29	27.11000	28.29000
Хліб житній, житньо-пшеничний/кг	27.46	29.73000	30.93000
Батон /500 г	17.88	18.10000	18.63000
Макаронні вироби з м'яких сортів пшениці...	24.36	24.95000	25.16000
Борошно пшеничне/кг	15.01	15.48000	15.65000
Крупи манні/кг	18.18	18.85000	19.10000
Крупи гречані/кг	47.77	47.61000	48.66000
Крупи ячні /кг	15.8	16.41000	16.96000
Пшоно/кг	18.42	18.74000	19.45000
Крупи пшеничні/кг	—	17.42000	17.65000
Яловичина/кг	182.84	195.61000	208.02000
Свинина/кг	123.75	123.95000	123.50000
Птиця (тушки курячі)/кг	70.82	73.52000	74.12000
Філе куряче/кг	117.1	122.65000	123.63000
Ковбаси варені першого ґатунку/кг	120.15	124.41000	125.50000
Риба морожена/кг	87.86	103.70000	105.83000
Молоко пастеризоване жирністю до 2,6% в...	28.55	28.93000	29.37000
Сири м'які жирні/кг	125.98	133.60000	139.55000

	Грудень 2021	Січень	Лютий
Рис/кг	28.74000	29.30000	30.19000
Хліб пшеничний з борошна вищого ґатунку...	33.08000	33.61000	34.96000
Хліб пшеничний з борошна першого ґату...	26.29000	27.11000	28.29000
Хліб житній, житньо-пшеничний/кг	27.46000	29.73000	30.93000
Батон /500 г	17.88000	18.10000	18.63000
Макаронні вироби з м'яких сортів пшениці...	24.36000	24.95000	25.16000
Борошно пшеничне/кг	15.01000	15.48000	15.65000
Крупи манні/кг	18.18000	18.85000	19.10000
Крупи гречані/кг	47.77000	47.61000	48.66000
Крупи ячні /кг	15.80000	16.41000	16.96000
Пшоно/кг	18.42000	18.74000	19.45000
Крупи пшеничні/кг	nan	17.42000	17.65000
Яловичина/кг	182.84000	195.61000	208.02000
Свинина/кг	123.75000	123.95000	123.50000
Птиця (тушки курячі)/кг	70.82000	73.52000	74.12000
Філе куряче/кг	117.10000	122.65000	123.63000
Ковбаси варені першого ґатунку/кг	120.15000	124.41000	125.50000
Риба морожена/кг	87.86000	103.70000	105.83000
Молоко пастеризоване жирністю до 2,6% в...	28.55000	28.93000	29.37000
Сири м'які жирні/кг	125.98000	133.60000	139.55000

- Обчислення математичного сподівання, медіани моди, дисперсії та середньоквадратичного відхилення

Грудень 2021	162.63660
Січень	165.39500
Лютий	166.56944
Березень	177.20056
Квітень	185.75926
Травень	186.41278
Червень	190.65370
Липень	192.81741
Серпень	196.31130
Вересень	201.18093
Жовтень	204.56704
Листопад	208.10259
Грудень	209.70296

математичне сподівання

Грудень 2021	30.47000
Січень	32.53500
Лютий	33.68500
Березень	36.26500
Квітень	37.58000
Травень	39.50500
Червень	45.81000
Липень	44.76000
Серпень	44.87000
Вересень	44.96500
Жовтень	48.40000
Листопад	48.85500
Грудень	50.65000

медіана

0	27.11000
---	----------

мода за місяць січень

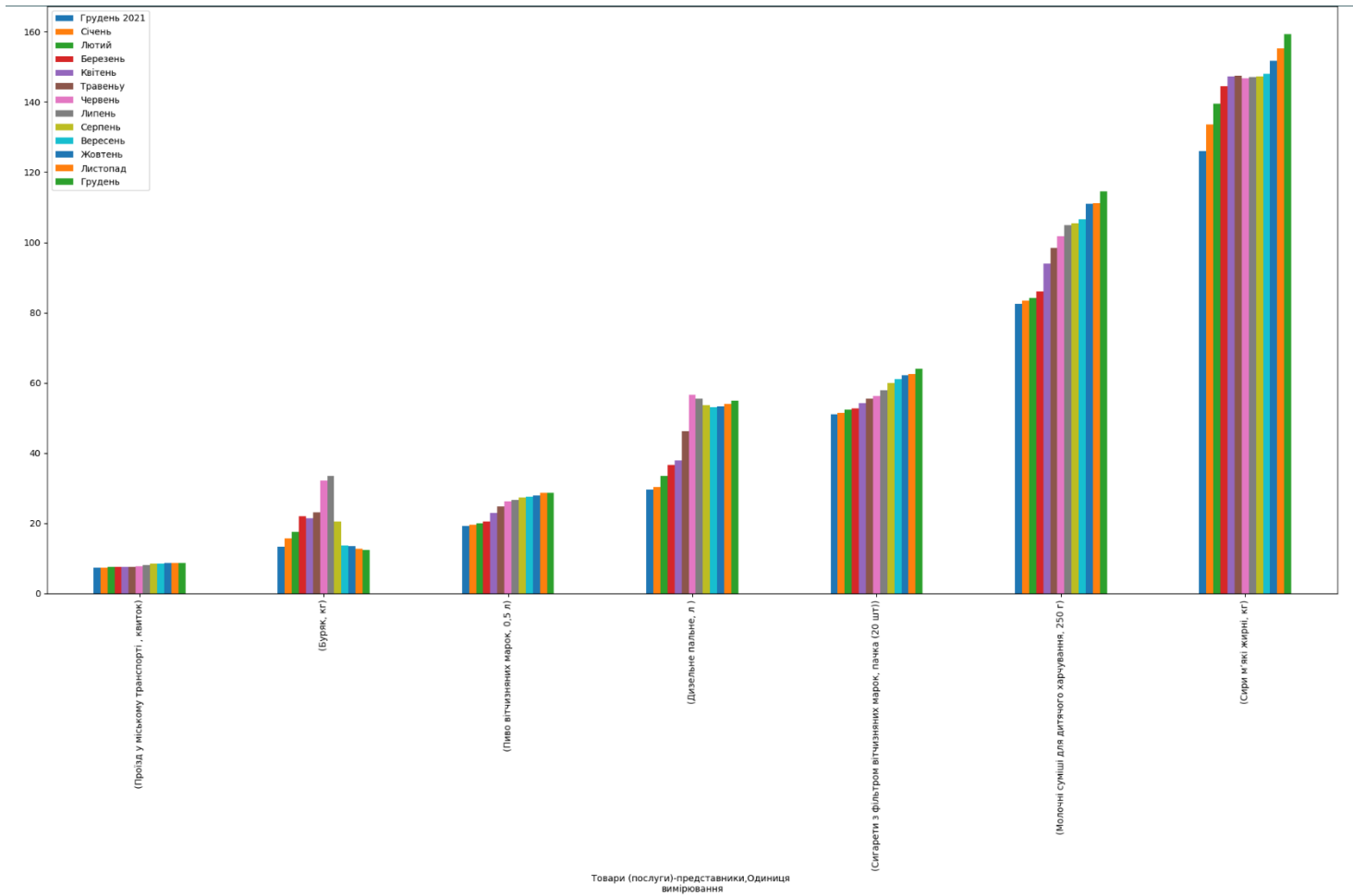
Грудень 2021	373229.11138
Січень	380780.24538
Лютий	378700.43963
Березень	441346.82522
Квітень	499148.76415
Травень	489452.21410
Червень	503773.65597
Липень	504415.75649
Серпень	530680.35263
Вересень	573055.91215
Жовтень	589794.10604
Листопад	622523.49736
Грудень	632873.35901

дисперсія

Грудень 2021	610.92480
Січень	617.07394
Лютий	615.38641
Березень	664.33939
Квітень	706.50461
Травень	699.60861
Червень	709.77014
Липень	710.22233
Серпень	728.47811
Вересень	757.00457
Жовтень	767.98054
Листопад	789.00158
Грудень	795.53338

середньоквадратичне відхилення

- Візуалізація даних за допомогою гістограм



## Вихідний код:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

pd.set_option('display.precision', 2)

initial_df = pd.read_excel(r'F:\Egor\Уроки\Машинне навчання\вартість товарів – копія.xlsx', index_col=[0, 1])
df = initial_df.copy()
df.replace('-', np.nan, inplace=True)
# print(df)

df.sort_values(by="Грудень 2021", inplace=True)
mean = df.mean()
print("\nMean:\n", mean)
median = df.median()
print("\nMedian:\n", median)
mode = df["Січень"].mode() # coz another columns have 54 modes
print("\nMode:\n", mode)
variance = df.var()
print("\nVariance:\n", variance)
RMSD = np.sqrt(variance)
print("\nRoot mean square deviation:\n", RMSD)
df.iloc[0:-1:8].plot(kind='bar')
plt.show()
```

- Завантаження набору даних катастрофи «Титаніка»

	A	B	C	D	E	F
1	,"survived","sex","age","passengerClass"					
2	Allen, Miss. Elisabeth Walton,"yes","female",29,"1st"					
3	Allison, Master. Hudson Trevor,"yes","male",0.916700006,"1st"					
4	Allison, Miss. Helen Loraine,"no","female",2,"1st"					
5	Allison, Mr. Hudson Joshua Crei,"no","male",30,"1st"					
6	Allison, Mrs. Hudson J C (Bessi,"no","female",25,"1st"					
7	Anderson, Mr. Harry,"yes","male",48,"1st"					
8	Andrews, Miss. Kornelia Theodos,"yes","female",63,"1st"					
9	Andrews, Mr. Thomas Jr,"no","male",39,"1st"					
10	Appleton, Mrs. Edward Dale (Cha,"yes","female",53,"1st"					
11	Artagaveytia, Mr. Ramon,"no","male",71,"1st"					
12	Astor, Col. John Jacob,"no","male",47,"1st"					
13	Astor, Mrs. John Jacob (Madelei,"yes","female",18,"1st"					
14	Aubart, Mme. Leontine Pauline,"yes","female",24,"1st"					
15	Barber, Miss. Ellen Nellie,"yes","female",26,"1st"					
16	Barkworth, Mr. Algernon Henry W,"yes","male",80,"1st"					
17	Baumann, Mr. John D,"no","male",NA,"1st"					
18	Baxter, Mr. Quigg Edmond,"no","male",24,"1st"					
19	Baxter, Mrs. James (Helene DeLa,"yes","female",50,"1st"					
20	Bazzani, Miss. Albina,"yes","female",32,"1st"					
21	Beattie, Mr. Thomson,"no","male",36,"1st"					
22	Beckwith, Mr. Richard Leonard,"yes","male",37,"1st"					
23	Beckwith, Mrs. Richard Leonard,"yes","female",47,"1st"					

- Перейменування стовпців та перегляд частини набору даних

	name	survived	sex	age	class
0	Allen, Miss. Elisabeth Walton	yes	female	29.00	1st
1	Allison, Master. Hudson Trevor	yes	male	0.92	1st
2	Allison, Miss. Helen Loraine	no	female	2.00	1st
3	Allison, Mr. Hudson Joshua Crei	no	male	30.00	1st
4	Allison, Mrs. Hudson J C (Bessi	no	female	25.00	1st
	name	survived	sex	age	class
1304	Zabour, Miss. Hileni	no	female	14.5	3rd
1305	Zabour, Miss. Thamine	no	female	NaN	3rd
1306	Zakarian, Mr. Mapriededer	no	male	26.5	3rd
1307	Zakarian, Mr. Ortin	no	male	27.0	3rd
1308	Zimmerman, Mr. Leo	no	male	29.0	3rd

- Пошук наймолодшого та найстаршого пасажиру, загального середнього віку й обчислення статистики по пасажирам, які вижили

```

Youngest passenger:
  name      Dean, Miss. Elizabeth Gladys M
survived      yes
sex          female
age          0.17
class        3rd
Name: 763, dtype: object

Oldest passenger:
  name      Barkworth, Mr. Algernon Henry W
survived      yes
sex          male
age          80.0
class        1st
Name: 14, dtype: object

Average age: 29.88

Statistics on passengers who survived:
      age
count  427.00
mean   28.92
std    15.06
min     0.17
25%    20.00
50%    28.00
75%    38.00
max    80.00

```

- Пошук наймолодшої та найстаршої жінки з першого класу й загальної кількості виживших серед них

```

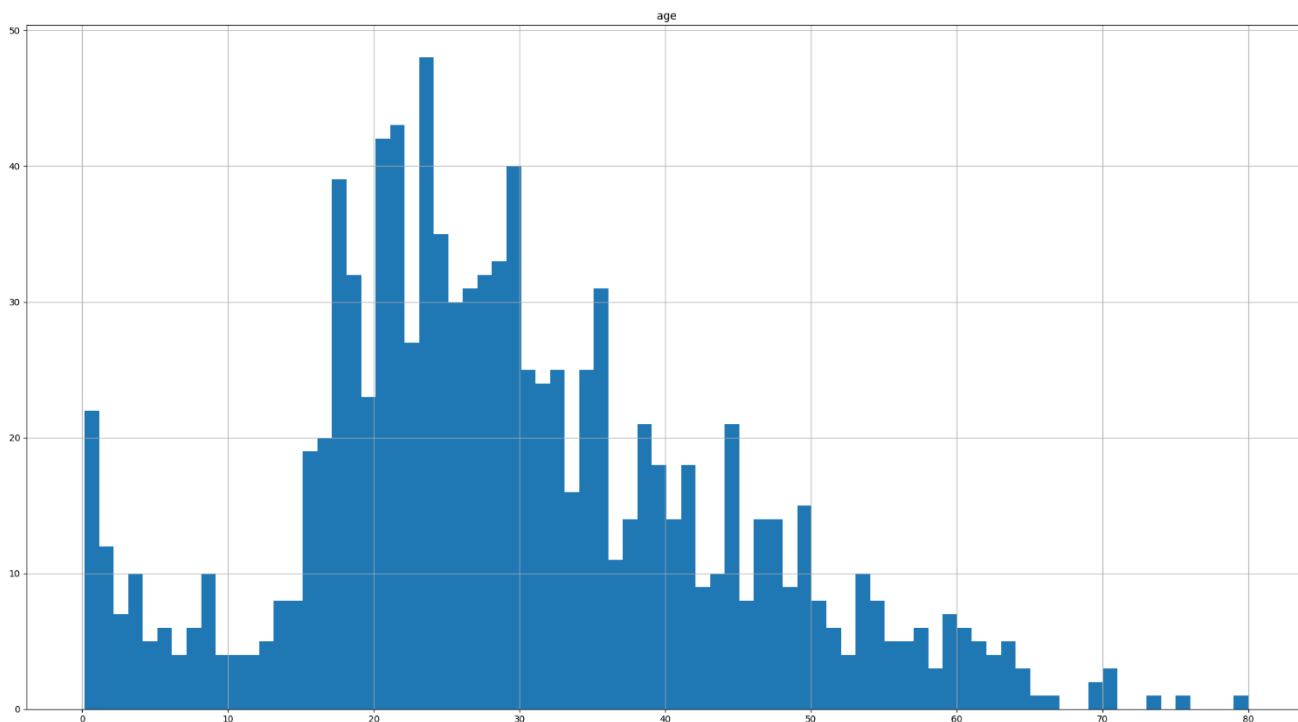
Youngest woman from the 1st class:
  name      Carter, Miss. Lucile Polk
survived      yes
sex          female
age          14.0
class        1st
Name: 25, dtype: object

Oldest woman from the 1st class:
  name      Snyder, Mrs. John Pillsbury (Ne
survived      yes
sex          female
age          23.0
class        1st
Name: 122, dtype: object

Total female survivors amount: 139

```

- Побудова гістограми віку пасажирів



**Вихідний код:**

```
initial_df = pd.read_csv(r'F:\Egor\Уроки\Машинне навчання\TitanicSurvival.csv')
df = initial_df.copy()
df.columns = ['name', 'survived', 'sex', 'age', 'class']
print(df.head(), end="\n\n")
print(df.tail(), end="\n\n")
print("Youngest passenger:\n", df.iloc[df['age'].idxmin()], end="\n\n")
print("Oldest passenger:\n", df.iloc[df['age'].idxmax()], end="\n\n")
print("Average age:", df['age'].mean().round(2), end="\n\n")
print("Statistics on passengers who survived:\n", df[df['survived'] == 'yes'].describe(), end="\n\n")
women_first_class = df[(df['class'] == '1st') & (df['sex'] == 'female')].copy()
women_first_class.reset_index(drop=True, inplace=True)
women_first_class.sort_values(by='age', inplace=True)
print("Youngest woman from the 1st class:\n", women_first_class.iloc[women_first_class['age'].idxmin()], end="\n\n")
print("Oldest woman from the 1st class:\n", women_first_class.iloc[women_first_class['age'].idxmax()], end="\n\n")
print("Total female survivors amount:", women_first_class.loc[women_first_class['survived'] == 'yes'].shape[0])
df.hist(bins=80)
plt.show()
```

**Висновок:**

Було досліджено можливості бібліотек pandas, numpy й matplotlib та використано деякі з них на практиці для різних наборів даних; було вивчено методи імпорту даних як у форматі xlsx, так і у форматі csv; було вивчено основні математичні характеристики статистики та використано вище згадані модулі для їх пошуку; було досліджено методи візуалізації наборів даних, та побудовано гістограми обраних наборів даних для закріплення знань; таким чином було вивчено основні інструменти, що знадобляться при подальшому вивченні дисципліни «Прикладні задачі машинного навчання» та виконано вступ в data science.