

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 №3 з дисципліни
«Аналіз даних в інформаційних системах»

«Описова статистика»

Варіант 7

Виконав студент ПІ-12 Васильєв Єгор Костянтинович
(шифр, прізвище, ім'я, по батькові)

Перевірив Ліхоузова Тетяна Анатоліївна
(прізвище, ім'я, по батькові)

Київ 2023

Лабораторна робота № 3

Тема: Описова статистика

Для виконання лабораторної роботи було обрано мову програмування Python, та бібліотеки Numpy, Pandas і Matplotlib. Спочатку дані було завантажено та імпортовано до датафрейму.

	Country Name	Region	GDP per capita	Population	CO2 emission	Area
0	Afghanistan	South Asia	561,7787463	34656032.00000	9809,225	652860
1	Albania	Europe & Central Asia	4124,98239	2876101.00000	5716,853	28750
2	Algeria	Middle East & North Africa	3916,881571	40606052.00000	145400,217	2381740
3	American Samoa	East Asia & Pacific	11834,74523	55599.00000	nan	200
4	Andorra	Europe & Central Asia	36988,62203	77281.00000	462,042	470
5	Angola	Sub-Saharan Africa	3308,700233	28813463.00000	34763,16	1246700
6	Antigua and Barbuda	Latin America & Caribbean	14462,17628	100963.00000	531,715	440
7	Argentina	Latin America & Caribbean	12440,32098	43847430.00000	204024,546	2780400
8	Armenia	Europe & Central Asia	3614,688357	2924816.00000	5529,836	29740
9	Aruba	Latin America & Caribbean	nan	104822.00000	872,746	180
10	Australia	East Asia & Pacific	49755,31548	24127159.00000	361261,839	7741220
11	Austria	Europe & Central Asia	44757,6349	8747358.00000	58712,337	83879
12	Azerbaijan	Europe & Central Asia	3878,709257	9762274.00000	37487,741	86600
13	Bahamas, The	Latin America & Caribbean	28785,47767	391232.00000	2416,553	13880
14	Bahrain	Middle East & North Africa	22579,09342	1425171.00000	31338,182	771
15	Bangladesh	South Asia	1358,779029	162951560.00000	73189,653	147630
16	Barbados	Latin America & Caribbean	15891,62655	284996.00000	1272,449	430
17	Belarus	Europe & Central Asia	4989,427763	9507120.00000	63497,772	207600
18	Belgium	Europe & Central Asia	41271,48215	11348159.00000	93350,819	30530
19	Belize	Latin America & Caribbean	4744,736397	366954.00000	495,045	22970
20	Benin	Sub-Saharan Africa	789,4404107	10872298.00000	6318,241	114760

Рисунок 3.1 – імпортовані дані у data frame

Було досліджено структуру даних та знайдено наступні помилки: граматична помилка у назві стовпця Population, числові дані, які частково зберігаються у вигляді рядкових, наявність від'ємних значень та пропущені значення, що не дозволяють коректно побудувати діаграму розмаху. Після виправлення цих помилок датафрейм отримав наступний вигляд:

	Country Name	Region	GDP per capita	Population	CO2 emission	Area
0	Afghanistan	South Asia	561.77875	34656032.00000	9809.22500	652860.00000
1	Albania	Europe & Central Asia	4124.98239	2876101.00000	5716.85300	28750.00000
2	Algeria	Middle East & North Africa	3916.88157	40606052.00000	145400.21700	2381740.00000
3	American Samoa	East Asia & Pacific	11834.74523	55599.00000	165114.11634	200.00000
4	Andorra	Europe & Central Asia	36988.62203	77281.00000	462.04200	470.00000
5	Angola	Sub-Saharan Africa	3308.70023	28813463.00000	34763.16000	1246700.00000
6	Antigua and Barbuda	Latin America & Caribbean	14462.17628	100963.00000	531.71500	440.00000
7	Argentina	Latin America & Caribbean	12440.32098	43847430.00000	204024.54600	2780400.00000
8	Armenia	Europe & Central Asia	3614.68836	2924816.00000	5529.83600	29740.00000
9	Aruba	Latin America & Caribbean	13445.59342	104822.00000	872.74600	180.00000
10	Australia	East Asia & Pacific	49755.31548	24127159.00000	361261.83900	7741220.00000
11	Austria	Europe & Central Asia	44757.63490	8747358.00000	58712.33700	83879.00000
12	Azerbaijan	Europe & Central Asia	3878.70926	9762274.00000	37487.74100	86600.00000
13	Bahamas, The	Latin America & Caribbean	28785.47767	391232.00000	2416.55300	13880.00000
14	Bahrain	Middle East & North Africa	22579.09342	1425171.00000	31338.18200	771.00000
15	Bangladesh	South Asia	1358.77903	162951560.00000	73189.65300	147630.00000
16	Barbados	Latin America & Caribbean	15891.62655	284996.00000	1272.44900	430.00000
17	Belarus	Europe & Central Asia	4989.42776	9507120.00000	63497.77200	207600.00000
18	Belgium	Europe & Central Asia	41271.48215	11348159.00000	93350.81900	30530.00000
19	Belize	Latin America & Caribbean	4744.73640	366954.00000	495.04500	22970.00000
20	Benin	Sub-Saharan Africa	789.44041	10872298.00000	6318.24100	114760.00000

Рисунок 3.2 – виправлені дані у data frame

За допомогою вбудованих методів бібліотеки Matplotlib було створено функцію для побудови діаграм розмаху та гістограм по заданому стовпцю датафрейму.

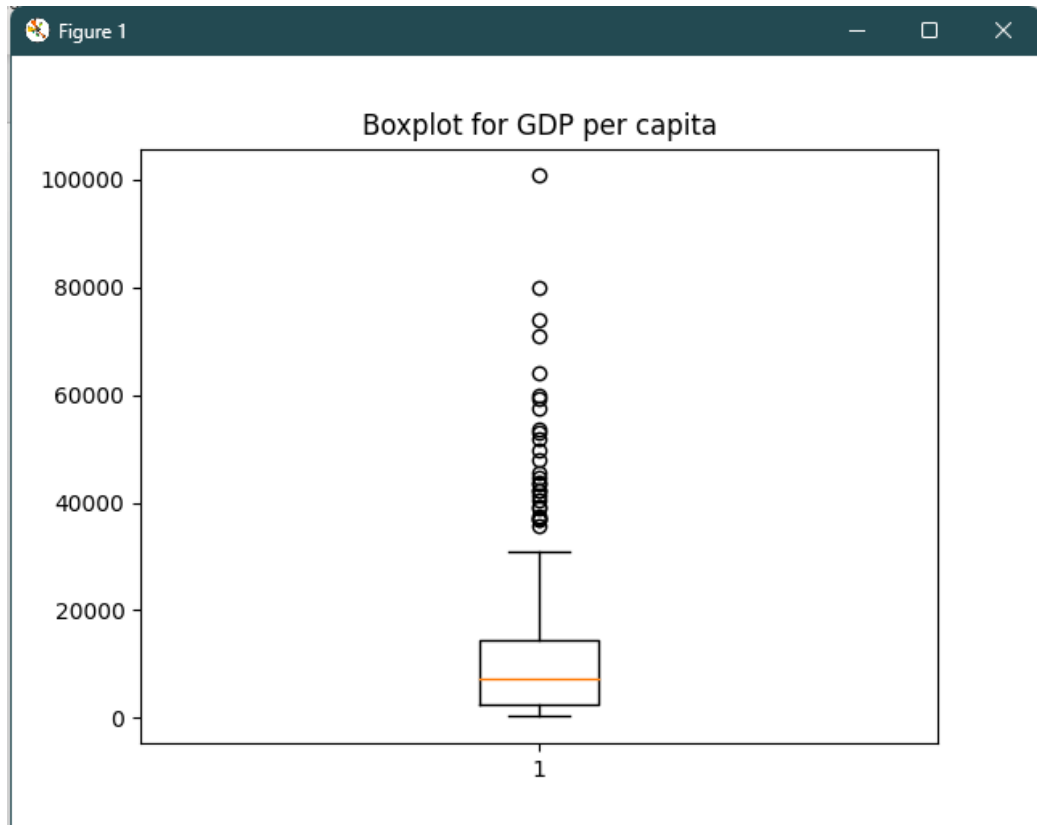


Рисунок 3.3 – діаграма розмаху колонки GDP per capita

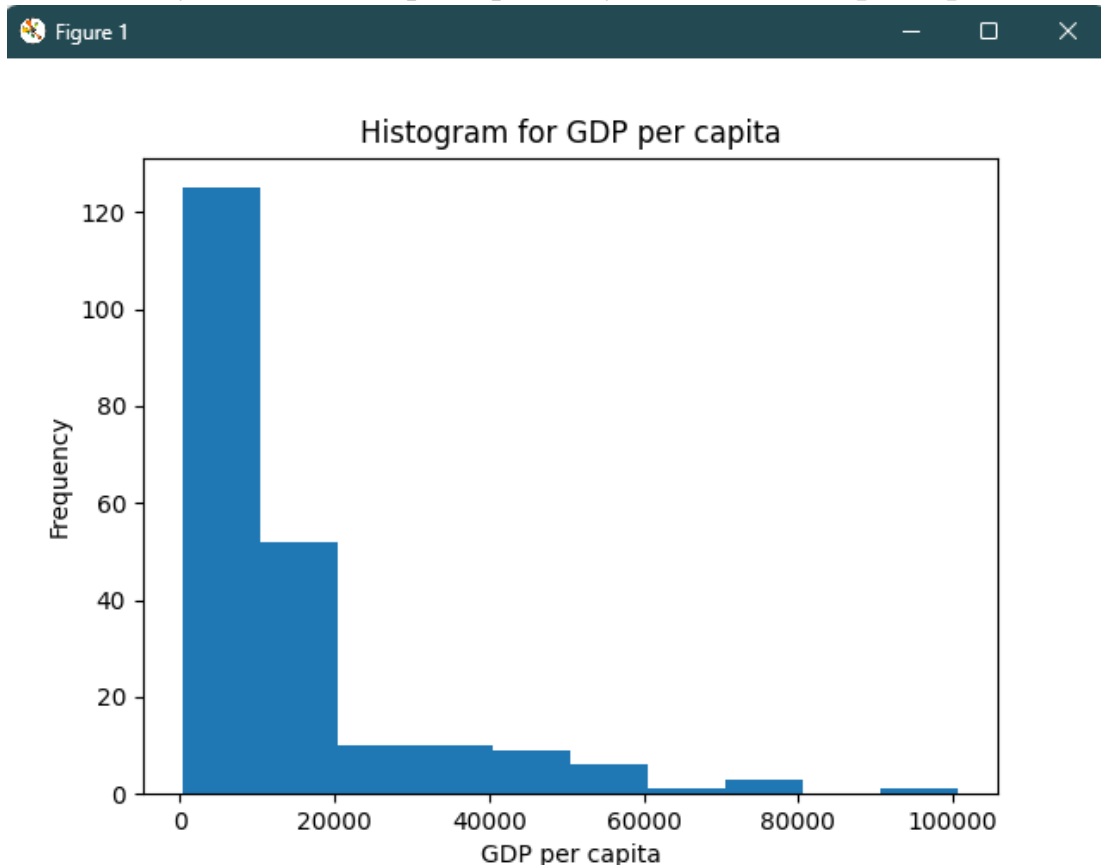


Рисунок 3.4 – гістограма колонки GDP per capita

До датафрейму було додано стовпчик з щільністю населення, який обраховується як частка населення та площі території.

	Country Name	Region	GDP per capita	Population	CO2 emission	Area	Population density
0	Afghanistan	South Asia	561.77875	34656032.00000	9809.22500	652860.00000	53.08341
1	Albania	Europe & Central Asia	4124.98239	2876101.00000	5716.85300	28750.00000	100.03830
2	Algeria	Middle East & North Africa	3916.88157	40606052.00000	145400.21700	2381740.00000	17.04890
3	American Samoa	East Asia & Pacific	11834.74523	55599.00000	165114.11634	200.00000	277.99500
4	Andorra	Europe & Central Asia	36988.62203	77281.00000	462.04200	470.00000	164.42766
5	Angola	Sub-Saharan Africa	3308.70023	28813463.00000	34763.16000	1246700.00000	23.11179
6	Antigua and Barbuda	Latin America & Caribbean	14462.17628	100963.00000	531.71500	440.00000	229.46136
7	Argentina	Latin America & Caribbean	12440.32098	43847430.00000	204024.54600	2780400.00000	15.77019
8	Armenia	Europe & Central Asia	3614.68836	2924816.00000	5529.83600	29740.00000	98.34620
9	Aruba	Latin America & Caribbean	13445.59342	104822.00000	872.74600	180.00000	582.34444
10	Australia	East Asia & Pacific	49755.31548	24127159.00000	361261.83900	7741220.00000	3.11671
11	Austria	Europe & Central Asia	44757.63490	8747358.00000	58712.33700	83879.00000	104.28543
12	Azerbaijan	Europe & Central Asia	3878.70926	9762274.00000	37487.74100	86600.00000	112.72834
13	Bahamas, The	Latin America & Caribbean	28785.47767	391232.00000	2416.55300	13880.00000	28.18674
14	Bahrain	Middle East & North Africa	22579.09342	1425171.00000	31338.18200	771.00000	1848.47082
15	Bangladesh	South Asia	1358.77903	162951560.00000	73189.65300	147630.00000	1103.78351
16	Barbados	Latin America & Caribbean	15891.62655	284996.00000	1272.44900	430.00000	662.78140
17	Belarus	Europe & Central Asia	4989.42776	9507120.00000	63497.77200	207600.00000	45.79538
18	Belgium	Europe & Central Asia	41271.48215	11348159.00000	93350.81900	30530.00000	371.70518
19	Belize	Latin America & Caribbean	4744.73640	366954.00000	495.04500	22970.00000	15.97536
20	Benin	Sub-Saharan Africa	789.44041	10872298.00000	6318.24100	114760.00000	94.73944

Рисунок 3.5 – доданий стовпчик у data frame

Пропущені значення у всьому датафреймі були замінені середнім значенням по колонці.

#	Column	Non-Null Count	Dtype	#	Column	Non-Null Count	Dtype
0	Country Name	217 non-null	object	0	Country Name	217 non-null	object
1	Region	217 non-null	object	1	Region	217 non-null	object
2	GDP per capita	190 non-null	float64	2	GDP per capita	217 non-null	float64
3	Population	216 non-null	float64	3	Population	217 non-null	float64
4	CO2 emission	205 non-null	float64	4	CO2 emission	217 non-null	float64
5	Area	217 non-null	float64	5	Area	217 non-null	float64

Рисунок 3.6 – кількість пропущених значень до та після

Далі було досліджено країни з найбільшими та найменшими показниками за різними критеріями. Для дослідження показників по регіонам, було згруповано датафрейм за відповідним стовпчиком, та проведено вже його аналіз.

```
Country with biggest GDP per capita:
Country Name      Luxembourg
Region            Europe & Central Asia
GDP per capita     100738.684
Population        582972.0
CO2 emission      9658.878
Area              2590.0
Population density 225.086
```

Рисунок 3.7 – країна з найбільшим ВВП на людину

```
Country with smallest area:
Country Name          Monaco
Region                Europe & Central Asia
GDP per capita        13445.593
Population            38499.0
CO2 emission         165114.116
Area                  2.0
Population density    19249.5
```

Рисунок 3.7 – країна з найменшою площею

```
Region with highest average country area:
GDP per capita        3.776e+04
Population            1.198e+08
CO2 emission         1.931e+06
Area                  6.605e+06
Population density    4.477e+02
Name: North America, dtype: float64
```

Рисунок 3.8 – регіон з найбільшою середньою площею країни

```
Country with biggest population density:
Country Name          Macao SAR, China
Region                East Asia & Pacific
GDP per capita        74017.185
Population            612167.0
CO2 emission         1283.45
Area                  30.3
Population density    20203.531
```

Рисунок 3.9 – країна з найбільшою щільністю населення у світі

```
Country with biggest population density in Europe and Central Asia:
Country Name          Monaco
Region                Europe & Central Asia
GDP per capita        13445.593
Population            38499.0
CO2 emission         165114.116
Area                  2.0
Population density    19249.5
```

Рисунок 3.10 – країна з найбільшою щільністю населення Європи та
центральній Азії

The average and median of GDP are different in all regions.

Рисунок 3.11 – співпадіння середнього та медіанного значення ВВП на душу населення по регіонам

Top 5 countries by GDP:	Last 5 countries by GDP:
Luxembourg: 100738.6842	Madagascar: 401.7422696
Switzerland: 79887.51824	Central African Republic: 382.2131743
Macao SAR, China: 74017.18471	Mozambique: 382.0693304
Norway: 70868.1225	Malawi: 300.3076651
Ireland: 64175.43824	Burundi: 285.7274421

Рисунок 3.12 – Топ 5 та останні 5 країн по ВВП на душу населення

Top 5 countries by CO2 emission:	Last 5 countries by CO2 emission:
China: 10291926.88	Marshall Islands: 102.676
United States: 5254279.285	Kiribati: 62.339
India: 2238377.137	Nauru: 47.671
Russian Federation: 1705345.684	Liechtenstein: 44.004
Japan: 1214048.358	Tuvalu: 11.001

Рисунок 3.13 – Топ 5 та останні 5 країн по кількості CO2

Висновок: я дослідив економічні та географічні показники різних країн світ та впевнився в наступних твердженнях:

- Монако є найменшою серед усіх досліджуваних країн та при цьому має найбільшу щільність населення у Європі й другу у світі.
- Регіон Східної Азії та Тихого океану має найбільшу за країнами середню щільність населення.
- Люксембург має найбільше ВВП на людину у світі, однак регіон Північної Америки перевершує Європу за цим показником (незважаючи на те, що топ 5 країн за ВВП також європейські).
- Південна Азія очікувано має найбільшу середню кількість населення по країнам.
- Регіон Африки, південніше від Сахари, має як найменшу щільність населення, так і практично найменше ВВП на душу населення (Південна Азія має показник лише на 83 одиниці менший)