

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ імені ІГОРЯ СІКОРСЬКОГО”
ФАКУЛЬТЕТ ІНФОРМАТИКИ ТА ОБЧИСЛЮВАЛЬНОЇ ТЕХНІКИ

Кафедра інформатики та програмної інженерії

Звіт до лабораторної роботи №2

з курсу

«Машинне навчання»

студента 2 курсу
групи ІТ-02
Макарова Іллі Сергійовича

Викладач:
Оніщенко В.

Київ – 2022

Тема: Введення в data science 2

Завдання:

1. Для даних з лабораторної роботи 1 проробити всі дії з пункту 1. колекції Series і DataFrame бібліотеки pandas
2. Завантажити набір даних катастрофи «Титаніка» за URL-адресою
3. Переглянути рядки набору даних катастрофи «Титаніка»
4. Налаштувати назви стовпців
5. Проведіть простий аналіз даних:
6. Побудуйте гістограму віку пасажирів

Виконання:

Імпортуємо бібліотеки

```
In [26]: import pandas as pd
import matplotlib.pyplot as plt

pd.set_option('precision', 2)

In [27]: plt.rcParams['figure.figsize'] = [30, 10]
plt.rcParams.update({'font.size': 25})
plt.rcParams.update({'xtick.color': 'white'})
```

Першою частиною завдання є виконання завдань з 1ї лаби, за допомогою pandas

Laba 1 solution with pandas

```
In [28]: df = pd.read_csv('data/GDP by Country in 2017.csv').rename(columns={'GDP(in US$)': 'gdp'})

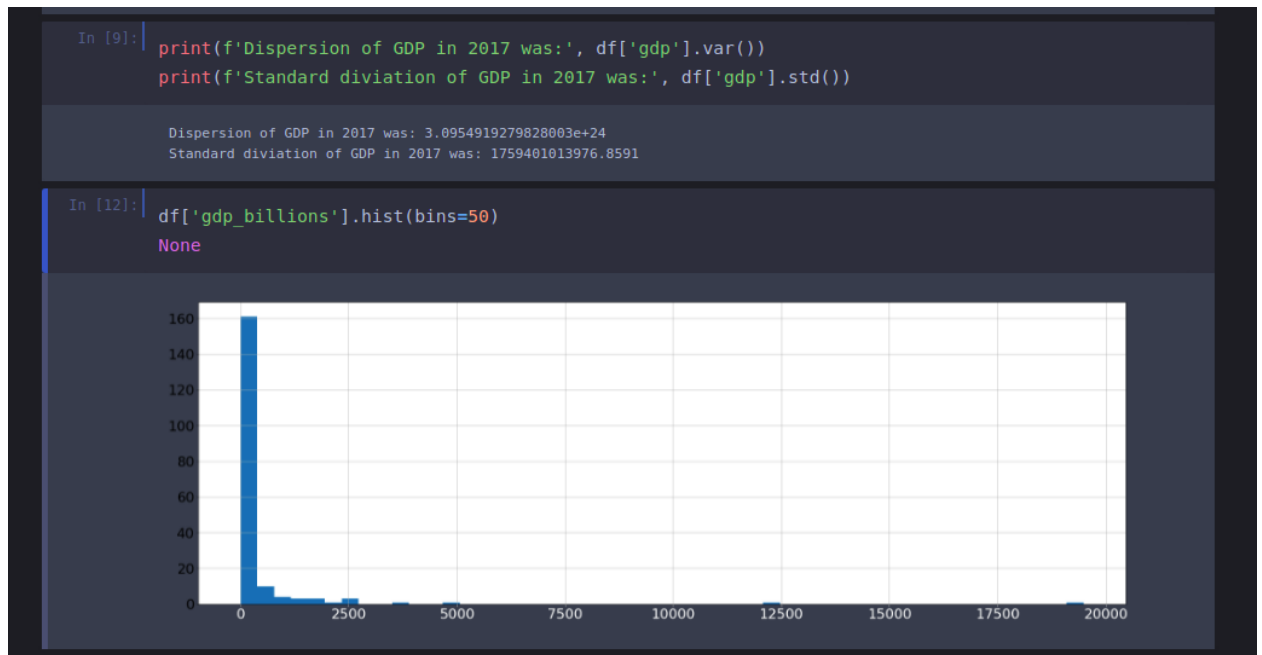
In [29]: df['gdp_billions'] = (df['gdp'] // 1_000_000_000).round()

In [30]: print(f'Median of GDP in 2017 was:', df['gdp'].median())
print(f'Mean of GDP in 2017 was:', df['gdp'].mean())

Median of GDP in 2017 was: 37353276059.0
Mean of GDP in 2017 was: 421023956456.77246

In [31]: print(f'Mode of GDP in 2017 was:', df['gdp_billions'].mode()[0], 'billion')

Mode of GDP in 2017 was: 1 billion
```



Так, з завданням першої лаби покінчили, тепер далі, до аналізу Titanic Stats

Titanic analysis

```
In [2]: df = pd.read_csv('https://vincentarelbundock.github.io/Rdatasets/csv/carData/TitanicSurvival.csv')
```

```
In [3]: df.head(5)
```

	Unnamed: 0	survived	sex	age	passengerClass
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

```
In [4]: df.shape
```

(1309, 5)

4.Налаштувати назви стовпців

```
In [5]: df = df.rename(columns={'Unnamed: 0': 'name', 'passengerClass': 'class'})
```

```
In [6]: df.head()
```

	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

5. Проведіть 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

```
In [7]: print(f'The youngest passenger was: {df["age"].min()} years old')
        print(f'The oldest passenger was: {df["age"].max()} years old')
        print(f'The mean age of passengers was: {df["age"].mean()}')
```

```
The youngest passenger was: 0.166700006 years old
The oldest passenger was: 80.0 years old
The mean age of passengers was: 29.881134512434034
```

Lets look at the statistics of the passengers who survived

```
In [8]: df[df['survived'] == 'yes'].describe()
```

	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

```
In [9]: first_class_females = df[(df['class'] == '1st') & (df['sex'] == 'female')]
```

```
In [10]: print(f'The youngest passenger was: {first_class_females["age"].min()} years old')
        print(f'The oldest passenger was: {first_class_females["age"].max()} years old')
        print(f'Number of survived: {first_class_females[first_class_females["survived"] == "yes"]["name"]
```

```
The youngest passenger was: 2.0 years old
The oldest passenger was: 76.0 years old
Number of survived: 139
```

Sorting females from the first class by their age

```
In [12]: first_class_females.sort_values(by='age')
```

	name	survived	sex	age	class
2	Allison, Miss. Helen Loraine	no	female	2.0	1st
55	Carter, Miss. Lucile Polk	yes	female	14.0	1st
193	Madill, Miss. Georgette Alexand	yes	female	15.0	1st
195	Maioni, Miss. Roberta	yes	female	16.0	1st
159	Hippach, Miss. Jean Gertrude	yes	female	16.0	1st
...
176	Kenyon, Mrs. Frederick R (Mario	yes	female	NaN	1st
204	Meyer, Mrs. Edgar Joseph (Leila	yes	female	NaN	1st
277	Spencer, Mrs. William Augustus	yes	female	NaN	1st
293	Taylor, Mrs. Elmer Zebley (Juli	yes	female	NaN	1st
297	Thorne, Mrs. Gertrude Maybelle	yes	female	NaN	1st

144 rows x 5 columns

6. Побудуйте гістограму віку

```
In [41]: df.hist(column='age', bins=50, figsize=(25, 15), xlabelsize=25, ylabelsize=25)  
None
```

