



МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»
ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ
Кафедра Інформаційної Безпеки

Засоби підготовки та аналізу даних

Лабораторна робота №2 **Наука про дані: обмін результатами та початковий аналіз**

Мета: ознайомитися з системою контролю версій GitHub, навчитися створювати прості веб-додатки для обміну результатами досліджень із використанням модуля `spurge`.

Основні поняття: система контролю версій, репозиторій, інтерактивний веб-додаток.

Перевірив:

Виконав:

студент II курсу

групи ФБ-01

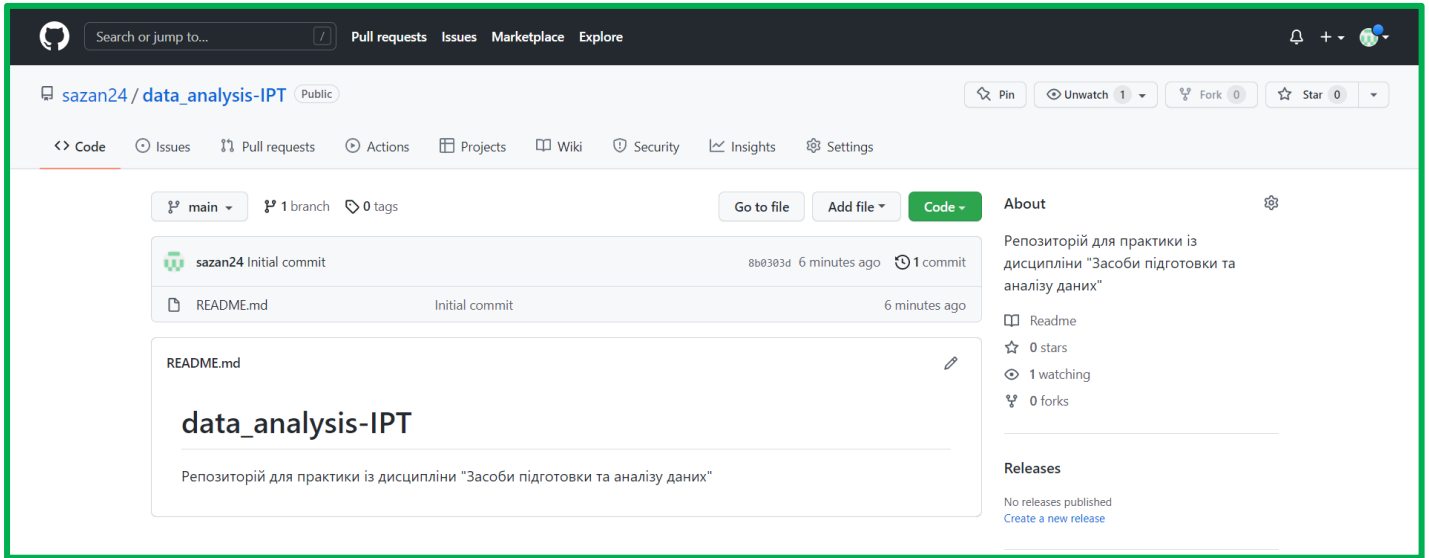
Сахній Н.Р.

Київ 2022

ФБ-01 Сахній Назар. Варіант 14

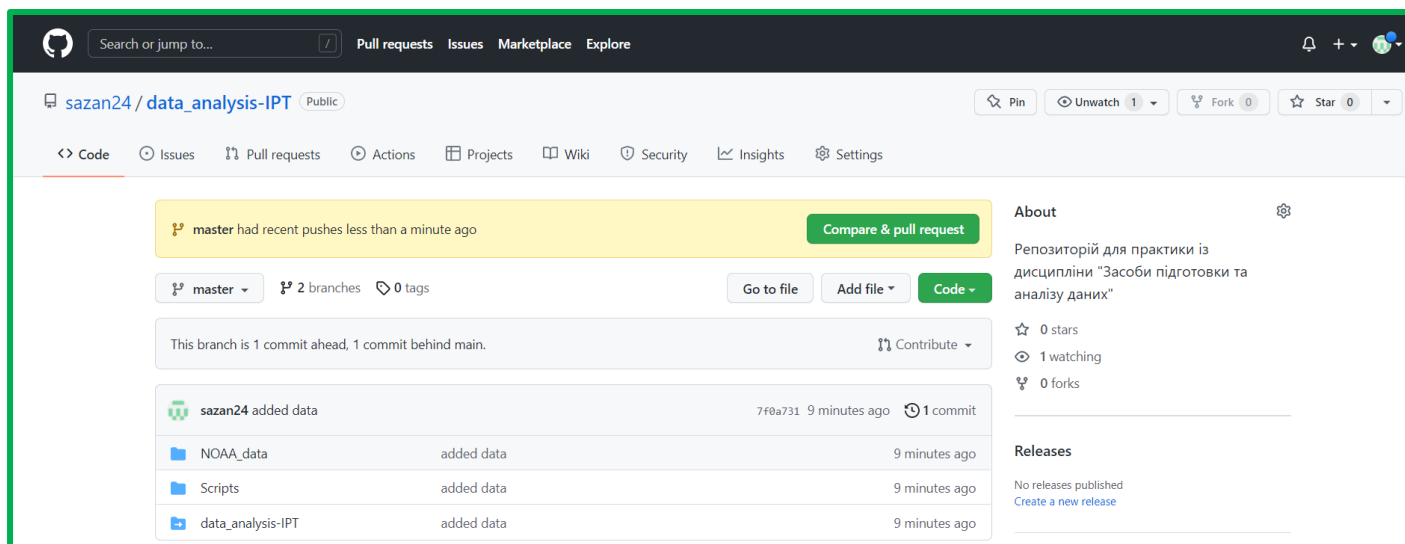
Хід виконання роботи

1. Зареєструватися у GitHub і створити репозиторій для першої лабораторної роботи;



2. Додати в репозиторій код та дані із лабораторної роботи №1, продемонструвати навички роботи з системою контролю версій git на роботі з проектом GitHub;

```
Windows PowerShell
PS D:\KPI\Data Analysis\data_analysis-git> git clone https://github.com/sazan24/data_analysis-IPT
Cloning into 'data_analysis-IPT'...
remote: Enumerating objects: 3, done.
remote: Counting objects: 100% (3/3), done.
remote: Compressing objects: 100% (2/2), done.
remote: Total 3 (delta 0), reused 0 (delta 0), pack-reused 0
Receiving objects: 100% (3/3), done.
PS D:\KPI\Data Analysis\data_analysis-git> git init
Initialized empty Git repository in D:\KPI\Data Analysis\data_analysis-git/.git/
PS D:\KPI\Data Analysis\data_analysis-git> git add .
PS D:\KPI\Data Analysis\data_analysis-git> git commit -m "added data"
[master (root-commit) 7f0a731] added data
 31 files changed, 54916 insertions(+)
 create mode 100644 "NOAA_data/NOAA_ID1 (\320\222\321\226\320\275\320\275\320\270\321\206\321\214\320\272\320\260).csv"
 create mode 100644 "NOAA_data/NOAA_ID10 (\320\232\321\200\320\276\320\277\320\270\320\262\320\275\320\270\321\206\321\214\320\272\320\260).csv"
 create mode 100644 "NOAA_data/NOAA_ID11 (\320\233\321\203\320\263\320\260\320\275\321\201\321\214\320\272\320\260).csv"
 create mode 100644 "NOAA_data/NOAA_ID12 (\320\233\321\214\320\262\321\226\320\262\321\201\321\214\320\272\320\260).csv"
 create mode 100644 "NOAA_data/NOAA_ID13 (\320\234\320\270\320\272\320\276\320\273\320\260\321\227\320\262\321\201\321\214\320\272\320\260).csv"
 create mode 100644 "NOAA_data/NOAA_ID14 (\320\236\320\264\320\265\321\201\321\214\320\272\320\260).csv"
 create mode 100644 "NOAA_data/NOAA_ID15 (\320\237\320\276\320\273\321\202\320\260\320\262\321\201\321\214\320\272\320\260).csv"
PS D:\KPI\Data Analysis\data_analysis-git> git remote add data_analysis https://github.com/sazan24/data_analysis-IPT
PS D:\KPI\Data Analysis\data_analysis-git> git push data_analysis
Enumerating objects: 34, done.
Counting objects: 100% (34/34), done.
Delta compression using up to 8 threads
Compressing objects: 100% (33/33), done.
Writing objects: 100% (34/34), 784.25 KiB | 3.83 MiB/s, done.
Total 34 (delta 0), reused 0 (delta 0), pack-reused 0
remote:
remote: Create a pull request for 'master' on GitHub by visiting:
remote:   https://github.com/sazan24/data_analysis-IPT/pull/new/master
remote:
To https://github.com/sazan24/data_analysis-IPT
 * [new branch]      master -> master
```



- Створити веб-додаток із використанням модуля `Spyre`, який дозволить: обрати часовий ряд VCI, TCI, VHI для набору даних із лабораторної роботи 1 (випадаючий список);

```

69 class WebAPI(server.App):
70     title = "NOAA data vizualization" # Заголовок веб-додатку
71
72     inputs = [{"type": "dropdown", # Випадаючий список для показників: VCI, TCI та VHI
73               "label": "NOAA data dropdown",
74               "options": [{"label": "VCI", "value": "VCI"},
75                           {"label": "TCI", "value": "TCI"},
76                           {"label": "VHI", "value": "VHI"}],
77               "key": "type",
78               "action_id": "update_data"},
79

```

- Створити ще один випадаючий список для вибору області;

```

56
57 regions = ["АР Крим", "Вінницька", "Волинська", "Дніпропетровська", "Донецька", "Житомирська", "Закарпатська",
58            "Запорізька", "Івано-Франківська", "Київ", "Київська", "Кропивницька", "Луганська", "Львівська",
59            "Миколаївська", "Одеська", "Полтавська", "Рівненська", "Севастополь", "Сумська", "Тернопільська",
60            "Харківська", "Херсонська", "Хмельницька", "Черкаська", "Чернівецька", "Чернігівська"]
61
80
81 {"type": "dropdown", # Випадаючий список для вибору області
82   "label": "Region dropdown",
83   "options": [{"label": f"{name}", "value": f"{name}"} for name in regions],
84   "key": "region",
85   "action_id": "update_data"},

```

5. Створити текстове поле для введення інтервалу років та тижнів;

```
85
86     {"type": "text", # Текстове поле для введення інтервалу років
87       "label": "Year interval",
88       "key": "year",
89       "value": '1982-2022',
90       "action_id": "simple_html_output"},
91
92     {"type": "text", # Текстове поле для введення інтервалу тижнів
93       "label": "Week interval",
94       "key": "week",
95       "value": '1-52',
96       "action_id": "simple_html_output"}]
```

6. Створити кілька вкладок для відображення таблиці із даними та графіку руху відповідних індексів;

```
98 controls = [{"type": "button", # Кнопка, яка оновлює сторінку з новими значеннями
99               "label": "Завантажити дані",
100               "id": "update_data"}]
101
102 tabs = ["Plot", "Table"] # Дві вкладки для відображення графіку та таблиці
103
104 outputs = [{"type": "plot", # Дані демонструються на графіку
105              "id": "plot_id",
106              "control_id": "update_data",
107              "tab": "Plot"},
108
109              {"type": "table", # Дані демонструються у таблиці
110                "id": "table_id",
111                "control_id": "update_data",
112                "tab": "Table",
113                "on_page_load": True}]
```

7. Запустити програму (веб-додаток)

```
140
141 app = WebAPI()
142 app.launch()
143
```

Додаткові необхідні сегменти коду (підключення бібліотек, створення функцій):

```
1 import os
2 import re
3 from spyre import server
4
5 import matplotlib.pyplot as plt
6 import seaborn as sns
7
8 import pandas as pd
9
10 from dateutil.relativedelta import relativedelta
11 import datetime
12
13
14 # Функція для встановлення першого дня кожного року (н/д 1 січня 2004)
15 def df_datetime_date(year):
16     return datetime.date(year, 1, 1)
17
18
19 # Функція, яка дасть змогу додавати до поточного дня року певну кількість тижнів, які будуть переведені у дні та місяці
20 def df_relativedelta(week):
21     return relativedelta(weeks=+week)
22
23
24 # Функція для зміни формату дат, щоб можна було поплудувати графік, на якому значення показників залежать від часу
25 def change_data_format(df):
26     df["Week"] = df["Week"].astype("int")
27     df["Year"] = df["Year"].astype("int")
28     df["date"] = df["Year"].apply(df_datetime_date) + df["Week"].apply(df_relativedelta)
29     df["date"] = pd.to_datetime(df["date"], format="%Y-%m-%d")
30     pass
31
32
33 # Функція, яка видаляє HTML-теги з початкового файлу (н/д такі, як <br>)
34 def remove_tags(string):
35     return re.sub("<.*?>", "", string)
36
37
38 # Функція для створення та підготовки єдиного файлу даних (конкатенація усіх файлів .csv)
39 def prepare_dataframe(path):
40     concatenate_files = []
41     for filename in os.listdir(path):
42         filepath = os.path.join(path, filename)
43         if os.path.isfile(filepath):
44             headers = ['Year', 'Week', 'SMN', 'SMT', 'VCI', 'TCI', 'VHI', 'empty', 'Area']
45             df = pd.read_csv(filepath, header=1, names=headers)
46
47             df["Year"] = df["Year"].apply(lambda cw: remove_tags(cw)) # Застосовуємо ф-ію видалення HTML-тегів
48             df = df.loc[df["VHI"] != -1] # Обрати лише ті дані, де показник "VHI" не дорівнює -1
49             df = df[df["Week"].notna()] # Використовувати лише існуючі дані
50             df = df.dropna(axis=1) # Видалити стовпчик порожніх значень (empty)
51
52             df["Area"] = regions[int(filename.split("_")[2]) - 1]
53             concatenate_files.append(df)
54     return pd.concat(concatenate_files, axis=0, ignore_index=True)
55
56
```

Результат виконання програмного коду:

1) Таблиця

NOAA DATA VIZUALIZATION

NOAA data dropdown:
VHI

Region dropdown:
АР Крим

Year interval:
1996-2004

Week interval:
1-52

Завантажити дані

PlotTable

	Year	Week	SMN	SMT	VCI	TCI	VHI	Area	date
20467	1996	1	0.032	265.31	5.19	52.90	29.05	АР Крим	1996-01-08
20468	1996	2	0.031	265.06	5.62	52.73	29.17	АР Крим	1996-01-15
20469	1996	3	0.031	265.41	6.26	51.40	28.83	АР Крим	1996-01-22
20470	1996	4	0.032	266.23	7.20	50.39	28.80	АР Крим	1996-01-29
20471	1996	5	0.034	266.90	8.17	50.21	29.19	АР Крим	1996-02-05
20472	1996	6	0.036	267.63	8.79	52.73	30.76	АР Крим	1996-02-12
20473	1996	7	0.040	268.83	10.33	55.79	33.06	АР Крим	1996-02-19
20474	1996	8	0.043	270.32	11.71	57.22	34.46	АР Крим	1996-02-26
20475	1996	9	0.049	272.68	12.89	56.65	34.77	АР Крим	1996-03-04
20476	1996	10	0.060	275.37	15.37	54.78	35.08	АР Крим	1996-03-11
20477	1996	11	0.073	278.27	17.34	51.01	34.17	АР Крим	1996-03-18
20478	1996	12	0.089	280.93	18.88	52.44	35.66	АР Крим	1996-03-25
20479	1996	13	0.108	283.53	19.73	58.06	38.90	АР Крим	1996-04-01
20480	1996	14	0.137	287.07	23.12	51.83	37.48	АР Крим	1996-04-08
20481	1996	15	0.172	290.00	27.91	44.38	36.15	АР Крим	1996-04-15
20482	1996	16	0.210	292.55	31.94	36.73	34.34	АР Крим	1996-04-22
20483	1996	17	0.253	295.28	36.06	24.93	30.49	АР Крим	1996-04-29
20484	1996	18	0.292	297.36	39.95	17.11	28.53	АР Крим	1996-05-06
20485	1996	19	0.329	298.88	43.49	13.79	28.64	АР Крим	1996-05-13
20486	1996	20	0.358	300.19	48.15	11.83	29.99	АР Крим	1996-05-20
20487	1996	21	0.372	300.86	52.23	14.62	33.42	АР Крим	1996-05-27
20488	1996	22	0.376	301.64	55.41	16.39	35.90	АР Крим	1996-06-03
20489	1996	23	0.374	302.89	57.83	13.57	35.70	АР Крим	1996-06-10
20490	1996	24	0.361	304.08	56.15	11.22	33.68	АР Крим	1996-06-17

2) Графік

