

Розрахунково-графічна робота
з дисципліни «Методи та засоби комп'ютерного навчання»
для студентів групи КНСП-11

Викладач:
Андрущак Н.А.

2021

Вступ до розрахункової графічної роботи

Із початку 2020 року світ потерпає від коронавірусної інфекції (COVID-2019), наслідки якої забрали життя сотень тисяч людей, а статистика по захворюваності і летальних випадках змінюється динамічно кожного дня.

Жодна країна не знає загальної кількості людей, інфікованих COVID-19. Ми знаємо лише стан зараженості тих, хто пройшов тестування. Усі, хто має лабораторно підтверджену інфекцію, зараховуються до підтверджених випадків. Це означає, що кількість підтверджених випадків залежить від того, скільки насправді тестує країна. Без тестування даних немає.

Тестування – це наше вікно до пандемії та її поширення. Без даних про зараження вірусом ми не можемо зрозуміти пандемію. Без цих даних ми не можемо знати, у яких країнах все добре, а які просто занижують випадки та смертність.

Для інтерпретації будь-яких даних про підтверджені випадки ми повинні знати, скільки насправді проводить тестування на COVID-19.

Тестування і розуміння отриманих даних – це один з найважливіших інструментів у боротьбі за уповільнення та зменшення поширення та впливу вірусу. Тести дозволяють ідентифікувати інфікованих осіб, керуючись медичним лікуванням, яке вони отримують. Це забезпечує ізоляцію інфікованих, а також відстеження та карантинування їхніх контактів. А також може допомогти ефективніше розподілити медичні ресурси та персонал.

Без даних про COVID-19 ми не можемо зрозуміти, як прогресує пандемія. Без даних ми не можемо належним чином реагувати на загрозу; ні як окремі особи, ні як суспільство. Також ми не можемо дізнатися, де діють контрзаходи проти пандемії. Кількість підтверджених випадків – це те, що інформує нас про розвиток пандемії. Тому надійні дані про тестування необхідні для оцінки надійності даних, які повідомляють нас про поширення пандемії: дані про випадки та смертність.

Вихідні дані

Для виконання розрахункової роботи вам доступні два файли, в яких зберігають дані про перебіг тестування, летальних випадків та статистики для кожної країни:

- **owid-covid-data-set1.xlsx** – файл, в якому зберігається інформація про випадки COVID-19 від початку пандемії і до 2 жовтня 2020 року.
- **owid-covid-data-set2.xlsx** – файл, в якому зберігається інформація про випадки COVID-19 від початку пандемії і до 25 грудня 2020 року.

Усі дані отримані із сайту - <https://ourworldindata.org/coronavirus-source-data>.

Пояснення до даних в таблиці:

- **iso_code** – міжнародний код країни;
- **continent** – континент, на якому знаходиться країна;
- **location** – назва країни;
- **date** – поточна дата (коли дані були внесені в таблицю);
- **total_cases** – загальна кількість випадків захворювання на COVID-19;
- **new_cases** – кількість нових випадків захворювання на COVID-19;
- **new_cases_smoothed** – нові випадки згладжені;
- **total_deaths** – загальна кількість смертей;
- **new_deaths** – кількість нових смертей (за день спостереження);
- **new_deaths_smoothed** – нових смертей згладжені;
- **total_cases_per_million** – загальна кількість випадків на 1 млн населення;
- **new_cases_per_million** – нові випадки на 1 млн населення;
- **new_cases_smoothed_per_million** – нові випадки згладжені на 1 млн населення;

- **total_deaths_per_million** – загальна кількість смертей на 1 млн населення;
- **new_deaths_per_million** – нових смертей на 1 млн населення;
- **new_deaths_smoothed_per_million** – нових смертей згладжені на 1 млн населення;
- **reproduction_rate** – коефіцієнт відтворення;
- **icu_patients** – кількість пацієнтів в інтенсивній терапії;
- **icu_patients_per_million** – кількість пацієнтів в інтенсивній терапії на 1 млн населення;
- **hosp_patients** – кількість пацієнтів в лікарні;
- **hosp_patients_per_million** – кількість пацієнтів в лікарні на 1 млн населення;
- **weekly_icu_admissions** – щотижнева кількість нових пацієнтів в інтенсивній терапії;
- **weekly_icu_admissions_per_million** – щотижнева кількість нових пацієнтів в інтенсивній терапії на 1 млн населення;
- **weekly_hosp_admissions** – щотижнева кількість пацієнтів в лікарнях;
- **weekly_hosp_admissions_per_million** – щотижнева кількість пацієнтів в лікарнях на 1 млн населення;
- **new_tests** – кількість нових тестів на COVID-19;
- **total_tests** – загальна кількість проведених тестів на COVID-19;
- **total_tests_per_thousand** – загальна кількість проведених тестів на COVID-19 на 1 тис населення;
- **new_tests_per_thousand** – кількість нових проведених тестів на COVID-19 на й тис населення;
- **new_tests_smoothed** – кількість нових проведених тестів на COVID-19 згладжені;
- **new_tests_smoothed_per_thousand** – кількість нових тестів на COVID-19 на 1 тис населення;
- **positive_rate** – позитивний показник;
- **tests_per_case** – тестів на випадок;
- **tests_units** – кількість місць для тестування;
- **total_vaccinations** – загальна кількість вакцинованих;
- **total_vaccinations_per_hundred** – загальна кількість вакцинованих на 1 тис населення;
- **stringency_index** – індекс жорсткості;
- **population** – населення;
- **population_density** – густина населення;
- **median_age** – середній вік населення;
- **aged_65_older** – відсоток населення старших за 65 років;
- **aged_70_older** – відсоток населення старших за 70 років;
- **gdp_per_capita** – ВВП на душу населення;
- **extreme_poverty** – рівень бідності;
- **cardiovasc_death_rate** – смертність від серцево-судинної системи;
- **diabetes_prevalence** – поширеність діабету;
- **female_smokers** – кількість жінок, які палять;
- **male_smokers** – кількість чоловіків, які палять;
- **handwashing_facilities** – кількість місця для миття рук;
- **hospital_beds_per_thousand** – кількість ліжко-місць в лікарні на 1 тис населення;
- **life_expectancy** – ймовірна тривалість життя;
- **human_development_index** – індекс людського розвитку.

Завдання на виконання

Для вибраного варіанту (варіант вибирається у відповідності до номеру студента в списку – такий список доступний в ВНС в шапці курсу), потрібно виконати такі завдання:

1. Вивести вміст файлу **owid-covid-data-set1.xlsx** на екран (для заданої країни), де в 1 вікні на 3 різних графіках показати:
 - 1.1. Залежність кількості випадків захворювання на COVID-19 від поточної дати.
 - 1.2. Залежність загальної кількості смертей від поточної дати.
 - 1.3. Залежність загальної кількості проведених тестів на COVID-19 від поточної дати.
2. З використанням регресійного аналізу для даних з файлу **owid-covid-data-set1.xlsx** передбачити та вивести на екран у вигляді графіків:
 - 2.1. Кількість випадків захворювання на COVID-19 від поточної дати до 25 грудня 2020 року.
 - 2.2. Загальну кількість смертей від поточної дати до 25 грудня 2020 року.
3. Порівняти результати отриманих даних передбачення із пункту 2 із даними із файлу **owid-covid-data-set2.xlsx** (до 25 грудня 2020 року), шляхом виводу на екран у вигляді графіків результат такого порівняння.
4. Провести кореляційний аналіз для перевірки, чи (1) кількість чоловіків та жінок, які палять, (2) кількість місця для миття рук, (3) середній вік населення впливають на (1) загальну кількість смертей та (2) загальна кількість випадків захворювання на COVID-19. Результат аналізу обґрунтувати.
5. Для даних із файлу **owid-covid-data-set2.xlsx** для своєї країни для аналізу та трьох країн для порівняння (які задані для індивідуального варіанту), проаналізувати та показати в 1 вікні на 4 різних графіках:
 - 5.1. Загальна кількість проведених тестів на COVID-19 від поточної дати.
 - 5.2. Загальна кількість випадків захворювання на COVID-19 від поточної дати.
 - 5.3. Загальна кількість смертей від поточної дати.
 - 5.4. Кількість нових смертей від поточної дати.
6. Серед 4 заданих країн для індивідуального варіанту, знайти до якої країни належить громадянин, коли відомо, що 20 грудня 2020 року в цій країні було зафіксовано дані, які показані в таблиці для індивідуального завдання, з використанням таких методів машинного навчання:
 - 6.1. Метод k-найближчих сусідів.
 - 6.2. Метод опорних векторів.
 - 6.3. Метод машинного навчання на вибір (студент сам вибирає та реалізує метод машинного навчання, який вважає, що найбільше підходить для рішення поставленого завдання).

Проаналізувати отримані результати різних методів. Оцінити точність прогнозування. Зробити висновки.

Зміст звіту розрахункової роботи

1. Звіт, де показано результати викання завдання. Звіт оформляється у вигляді статті у відповідності до заданих вимог, де мають бути обов'язково такі розділи:
 - 1.1. Опис проблематики проведеного дослідження.
 - 1.2. Опис методів та способів, які застосовувалися для аналізу даних.
 - 1.3. Результати виконання аналізу (графіки, діаграми, т.д.).
 - 1.4. Порівняльний аналіз отриманих даних у відповідності до завдань із даними рекомендаціями.
 - 1.5. Блок-схеми методів, які використовувалися для аналізу даних. Для кожного із методів має бути нарисована своя блок-схема, яка показує логіку виконання поставленого завдання із застосування вибраного методу.
 - 1.6. Висновки
2. Код для реалізації необхідних завдань на мові Python.

Важливо, звіт і код завантажуються окремо на ВНС для оцінювання.

Крайній термін надсилання розрахункової роботи – **12 грудня 2021 року (18:00)**. Якщо розрахункова робота надіслана після вказаного крайнього терміну, то така робота вважається незданою.

Варіанти індивідуальних завдань

№	Країна для аналізу	Країни для порівняння			Загальна кількість випадків захворювання на COVID-19	Загальна кількість смертей від COVID-19	Кількість нових смертей від COVID-19	Загальна кількість проведених тестів на COVID-19
1	Austria	Argentina	Ukraine	Hungary	338854	5351	142	3593276
2	Belgium	Estonia	Japan	Ukraine	21794	174	4	585874
3	Bulgaria	Ukraine	Slovakia	Switzerland	403989	6622	20	3106426
4	Canada	Italy	Ukraine	Ireland	1953185	68799	352	25129125
5	Croatia	Ukraine	Germany	Estonia	1514962	26400	229	33708381
6	Ukraine	France	Spain	Canada	511748	14245	74	13115616
7	Denmark	Canada	Ukraine	Japan	199270	2784	35	3999409
8	Estonia	Italy	Japan	Ukraine	1953185	68799	352	25129125
9	Finland	Ukraine	Germany	Norway	43905	404	0	2637381
10	France	Spain	Estonia	Ukraine	21794	174	4	585874
11	Germany	Ukraine	Norway	Portugal	374121	6134	71	5285351
12	Greece	Japan	Ukraine	France	199270	2784	35	3999409
13	Hungary	Italy	Norway	Ukraine	1953185	68799	352	25129125
14	Ireland	Ukraine	Japan	China	199270	2784	35	3999409
15	Italy	China	Germany	Ukraine	1953185	68799	352	25129125
16	Japan	Ukraine	Canada	Spain	511748	14245	74	13115616
17	Lithuania	Italy	Ukraine	Estonia	1953185	68799	352	25129125
18	Norway	Ukraine	France	Portugal	374121	6134	71	5285351
19	Poland	France	Spain	Ukraine	1202700	25397	143	6654502
20	Portugal	Ukraine	Estonia	Greece	21794	174	4	585874
21	Spain	China	Norway	Ukraine	43905	404	0	2637381
22	Switzerland	Spain	Ukraine	Portugal	374121	6134	71	5285351
23	Sweden	Ukraine	Canada	Germany	511748	14245	74	13115616
24	Turkey	China	Estonia	Ukraine	21794	174	4	585874
25	United States	Germany	Ukraine	France	1514962	26400	229	33708381