

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «ЗАПОРІЗЬКА ПОЛІТЕХНІКА»

Інститут інформатики та радіоелектроніки
Факультет комп'ютерних наук і технологій
Кафедра системного аналізу та обчислювальної математики

Пояснювальна записка
до дипломного проекту (роботи)

бакалавр

(ступінь вищої освіти)

на тему Вплив карантинних заходів на динаміку захворюваності на
коронавірусну інфекцію

Виконав: студент 4_ курсу, групи КНТ-818сп

Спеціальності 124 – Системний аналіз

Освітня програма (спеціалізація)

«Інтелектуальні технології та прийняття рішень в
складних системах»

Слободяник О.М.

(прізвище та ініціали)

Керівник Бахрушин В.Є.

(прізвище та ініціали)

Рецензент Ігнахіна М.О.

(прізвище та ініціали)

2021

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Національний університет «Запорізька політехніка»
 (повне найменування закладу вищої освіти)

Факультет комп'ютерних наук та технологій
 Кафедра системного аналізу та обчислювальної математики
 Ступінь вищої освіти бакалавр
 Спеціальність 124 – Системний аналіз
 (код і найменування)
 Освітня програма (спеціалізація) Інтелектуальні технології та прийняття рішень в складних системах
 (назва освітньої програми (спеціалізації))

ЗАТВЕРДЖУЮ

Завідувач кафедри _____

« 8 » червня 2021 року

З А В Д А Н Н Я
НА ДИПЛОМНИЙ ПРОЄКТ (РОБОТУ) СТУДЕНТА(КИ)

Слободянику Олександр Максимовичу
 (прізвище, ім'я, по батькові)

1. Тема проєкту (роботи) Вплив карантинних заходів на динаміку захворюваності на коронавірусну інфекцію

Керівник проєкту (роботи) Бахрушин В.Є., д.ф.-м.н., професор
 (прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Затверджені наказом закладу вищої освіти від «26» травня 2020 року №215

2. Строк подання студентом проєкту (роботи) 05 червня 2021

3. Вихідні дані до проєкту (роботи) Дані відкритих джерел щодо захворюваності на коронавірусну інфекцію у різних країнах, заходи, яких вживали ці країни для запобігання поширенню інфекції, та їх вплив на мобільність населення

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити): Опис методів та інструментів дослідження, загальний опис даних, що характеризують динаміку пандемії та заходи щодо запобігання її поширенню, аналіз отриманих даних і гіпотез щодо впливу запобіжних заходів.

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень) Ілюстративний матеріал щодо динаміки захворюваності та інших досліджуваних показників.

6. Консультанти розділів проєкту (роботи)

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	прийняв виконане завдання
1	Біхрушин В.Е. д.ф. - м.н., проф.		
2	Біхрушин В.Е. д.ф. - м.н., проф.		
3	Біхрушин В.Е. д.ф. - м.н., проф.		
4	Біхрушин В.Е. д.ф. - м.н., проф.		
	Нормоконтроль Широкогод Д.В. к.ф. - м.н. ст. доц.		

7. Дата видачі завдання « 08 » лютого 2020 року.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломного проєкту (роботи)	Строк виконання етапів проєкту (роботи)	Примітка
1	Визначення проблеми, завдань та методів дослідження	01.03 - 05.04	
2.	Пошук джерел інформації та збирання даних	05.04 - 23.04	
3.	Аналіз зібраних даних	23.04 - 20.05	
4.	Написання тексту пояснювальної записки	20.05 - 05.06	

Студент(ка)

(підпис)

Григоренко О.М.
(прізвище та ініціали)

Керівник проєкту (роботи)

(підпис)

Біхрушин В.Е.
(прізвище та ініціали)

РЕФЕРАТ

Обсяг: 48 сторінок, 31 рисунок, 14 джерел

Об'єкт дослідження — вплив карантинних заходів на динаміку захворюваності на коронавірусну інфекцію.

Мета дослідження — Визначити, чи існує взаємозв'язок між карантинними заходами та динамікою захворюваності.

Метод дослідження — аналіз відкритих даних.

Ціль дослідження створити комп'ютерну програму для збирання та аналізу даних щодо захворюваності на коронавірусну інфекцію, та впливу карантинних заходів на динаміку захворюваності. Через те що людство до цього часу не зустрічало цей вірус, на даний час не існує однозначної стратегії, тому потрібно аналізувати існуючі, щоб можливо було порівняти, та зроби припущення, про оптимальну стратегію.

КОРОНОВІРУС, АНАЛІЗ ДАНИХ, ВІДКРИТІ ДАНІ, PYTHON.

ЗМІСТ

Вступ	6
1 Методи дослідження.....	7
1.1 Об'єкт спостереження.....	8
1.2 Засоби спостереження.....	8
2 Інструменти дослідження.....	9
2.1 Python	9
2.2 Jupyter notebook.....	10
2.3 Matplotlib	10
2.4 Pandas	12
2.5 Seaborn	13
2.6 Scikit-learn.....	14
3 Первинна розвідка даних	15
3.1 Базовий набір даних	15
3.2 Карантинні заходи	16
3.3 Тестування.....	20
3.4 Нові випадки захворювання по дням	21
3.5 Смерті по днях	22
3.6 Оцінка коефіцієнту смертності (CFR)	23
3.7 Оцінка жорсткості обмежень	24
3.8 Переміщення громадян	25
4 Аналіз та побудова припущень	29
4.1 Підготування функцій Python для аналізу	29
4.2 Дослідження графіків мобільності/захворюваності.....	35
4.3 Дослідження видів мобільності.....	43
Висновок	46
Перелік посилань	47

ВСТУП

У даній роботі ми розглянемо існуючі карантинні заходи в різних країнах світу, проведемо невеличкий аналіз і первинну розвідку матеріалу. Список країн обмежимо 5ю, більша кількість країн потребує значно більшої кількості зусиль. Виберемо країни з найрізноманітнішими карантинними стратегіями, щоб розглянути більше варіантів реакцій держави, і скласти більш повний список. Далі за цим списком сформуємо дані, побудуємо графіки і будемо аналізувати ефективність карантинних заходів.

Під час аналізу ми будемо розглядати проблему лише з однієї сторони, розповсюдження хвороби серед людей, і не будемо аналізувати фінансову сторону та вплив на економіку країни.

Перша країна зі списку Україна. Її ми виберемо тому що це країна в якій пишеться курсова робота, і цікаво порівняти її з іншими.

Іспанія, вже має досвід боротьби з великими спалахами зараження, і можемо взяти її як один з можливих варіантів.

Швеція, цікава тим, що держава не обмежувало своїх громадян, і все носить рекомендаційний характер, без жорстких обмежень. Думаю це пов'язано з високим рівнем довіри громадян до уряду.

Ісландія, країна з низькою щільністю населення, плюс холодний клімат, роблять її цікавою для аналізу, і порівняння з іншими.

Німеччина, звичайна країна, з високим рівнем життя, і відсутністю великих спалахів.

Головною метою цієї роботи є комплексно проаналізувати країни та їх карантинні стратегії.

1 МЕТОДИ ДОСЛІДЖЕННЯ

Головним методом дослідження було обрано спостереження та побудова гіпотез.

Спостереження — метод наукового дослідження, що полягає в активному (систематичному, цілеспрямованому, планомірному) та навмисному сприйнятті об'єкта, в ході якого здобувається знання про зовнішні сторони, властивості й відносини досліджуваного об'єкта. Спостереження містить у собі як елементи: спостерігача (суб'єкта), об'єкт спостереження та засоби спостереження. Як останні в розвинених формах спостереження використовуються різноманітні спеціально створені прилади, що виступають як продовження й посилення органів чуття людини, а також використовувані як знаряддя впливу на об'єкт (що перетворює спостереження на складову частину експериментальної діяльності).

Гіпотеза — недоведене твердження, припущення або здогад.

Як правило, гіпотеза висловлюється на основі ряду спостережень, які її підтверджують і тому виглядає правдоподібно. Гіпотезу згодом або доводять, перетворюючи її у встановлений факт (див. теорема, теорія), або ж заперечують (наприклад, вказуючи контрприклад), переводячи в розряд помилкових тверджень.

Недоведена і неспростована гіпотеза називається відкритою проблемою. [7]

У нашій роботі ми найбільш вірогідно не зможемо довести, або спростувати гіпотезу, тому вона залишиться відкритою проблемою. Причина це не достовірність даних, та велика варіативність збору цих даних, від країни до країни.

1.1 Об'єкт спостереження

В якості об'єкта спостереження було обрано відкриті данні по захворюваності та мобільності, що були детально описані розділом вище.

1.2 Засоби спостереження

В якості основних засобів спостереження, та подальшого побудування гіпотез було обрано:

1. Побудову графіків.
2. Порівняння графіків.
3. Кросскореляція та побудування дерева рішень.

2 ІНСТРУМЕНТИ ДОСЛІДЖЕННЯ

Головним інструментом для досліджень було обрано мову Python, через наявність широких можливостей у області аналізу даних, та досвіду у використанні цієї мови. Також були використані супутні інструменти, такі як Jupyter notebook, matplotlib, pandas, seaborn, scikit-learn. Роздивимося кожний інструмент детальніше.

2.1 Python

Python – інтерпретована об'єктно-орієнтована високорівнева мова програмування з динамічною семантикою. Вбудовані високорівневі структури даних у поєднанні з динамічними типізацією і зв'язуванням роблять мову привабливим для швидкої розробки додатків (RAD, Rapid Application Development). Крім того, його можна використовувати як сценарну мову для зв'язку програмних компонентів. Синтаксис Python простий у вивченні, в ньому надається особливе значення читабельності коду, а це скорочує витрати на супровід програмних продуктів. Python підтримує модулі та пакети, заохочуючи модульність і повторне використання коду. Інтерпретатор Python і велика стандартна бібліотека доступні безкоштовно у вигляді вихідних і виконуваних кодів для всіх основних платформ і можуть вільно поширюватися. Python – це універсальна мова програмування. Він має свої переваги і недоліки, а також сфери застосування. У комплект постачання Python входить велика стандартна бібліотека для вирішення широкого кола завдань. В Інтернеті доступні якісні бібліотеки для Python в різних предметних областях: засоби обробки текстів і технології Інтернет, обробка зображень, інструменти для створення додатків, механізми доступу до баз даних, пакети для наукових

обчислень, бібліотеки побудови графічного інтерфейсу і т.п. Крім того, Python має досить прості засоби для інтеграції з мовами C, C++ (і Java) як шляхом вбудовування (embedding) інтерпретатора в програми на цих мовах, так і навпаки, за допомогою використання бібліотек, написаних на цих мовах, в Python-програмах. Мова Python підтримує кілька парадигм програмування: імперативне (процедурний, структурний, модульний підходи), об'єктно-орієнтоване і функціональне програмування. Можна вважати, що Python – це ціла технологія для створення програмних продуктів (і їх прототипів). Вона доступна майже на всіх сучасних платформах (як 32-бітових, так і на 64-бітових) з компілятором C і на платформі Java.[14]

2.2 Jupyter notebook

Блокнот Jupyter - це веб-додаток, який дозволяє користувачу писати коди та текстові елементи різноманітної форми. Всередині нього можна писати абзаци, рівняння, заголовки, додавати посилання, фігури тощо. Блокнот корисний для того, щоб ділитися інтерактивними алгоритмами зі своєю аудиторією, зосереджуючись на навчанні чи демонстрації техніки. Dodatok Jupyter Notebook — інтерфейс в якому можемо писати сценарії та коди у веб-браузері. Dodatok використовується локально, тобто без доступу до Інтернету, або до віддаленого сервера.[9]

2.3 Matplotlib

Matplotlib — бібліотека на мові програмування Python для візуалізації даних двовимірною 2D графікою (3D графіка також підтримується). Отримувані зображення можуть бути використані як

ілюстрації в публікаціях. Matplotlib написана і підтримується в основному Джоном Хантером і поширюється на умовах BSD-подібної ліцензії. Зображення, які генеруються в різних форматах, можуть бути використані в інтерактивній графіці, наукових публікаціях, графічному інтерфейсі користувача, веб-додатках, де потрібно будувати діаграми (англ. plotting). В документації автор зізнається, що Matplotlib починався з імітування графічних команд MATLAB, але є незалежним від нього проектом. Бібліотека Matplotlib побудована на принципах ООП, але має процедурний інтерфейс pylab, який надає аналоги команд MATLAB.

Matplotlib є гнучким, легко конфігурованим пакетом, який разом з NumPy, SciPy і IPython надає можливості, подібні до MATLAB. В даний час пакет працює з декількома графічними бібліотеками, включаючи wxWindows і PyGTK.

Пакет підтримує багато видів графіків і діаграм:

1. Графіки (line plot)
2. Діаграми розсіювання (scatter plot)
3. Стовпчасті діаграми (bar chart) і гістограми (histogram)
4. Секторні діаграми (pie chart)
5. Діаграми «Стовбур-листя» (stem plot)
6. Контурні графіки (contour plot)
7. Поля градієнтів (quiver)
8. Спектральні діаграми (spectrogram)

Користувач може вказати осі координат, сітку, додати підписи і пояснення, використовувати логарифмічну шкалу або полярні координати.

Нескладні тривимірні графіки можна будувати з допомогою набору інструментів (toolkit) mplot3d. Існують і інші набори інструментів: для картографії, для роботи з Excel, утиліти для GTK та інші.

З допомогою Matplotlib можна створювати і анімовані зображення.

Набір підтримуваних форматів зображень, векторних і растрових, можна отримати з словника `FigureCanvasBase.filetypes`. Типові підтримувані формати:

1. Encapsulated PostScript (EPS)
2. Enhanced Metafile (EMF)
3. JPEG
4. PDF
5. PNG
6. PostScript
7. RGBA («сирий» формат)
8. SVG
9. SVGZ
10. TIFF

Крім того, на основі класів пакету можна створювати й інші модулі. Наприклад, для генерування спарклайнів. [13]

2.4 Pandas

Pandas — програмна бібліотека, написана для мови програмування Python для маніпулювання даними та їхнього аналізу. Вона, зокрема, пропонує структури даних та операції для маніпулювання чисельними таблицями та часовими рядами. `pandas` є вільним програмним забезпеченням, що випускається за трипунктовою ліцензією BSD. Ця назва походить від терміну «панельні дані» (англ. *panel data*), який в економетрії позначає багатовимірні структуровані набори даних.

Можливості бібліотеки:

1. Об'єкт `DataFrame` із вбудованим індексуванням для маніпулювання даними.

2. Інструменти для зчитування та записування даних між структурами даних у пам'яті та різними форматами файлів.
3. Вирівнювання даних та вбудована підтримка пропущених даних.
4. Переформатовування для отримання зведених наборів даних.
5. Отримання зрізів за мітками, індексування з розширеними можливостями та отримання піднаборів з великих наборів даних.
6. Вставляння та вилучення стовпчиків у структурах даних.
7. Рушій групування, що дозволяє робити з наборами даних операції розділення-зміни-об'єднання (англ. split-apply-combine).
8. Злиття та з'єднання наборів даних.
9. Ієрархічне індексування осей для роботи з даними високої вимірності в структурі даних нижчої вимірності.
10. Функціональність для часових рядів: породження діапазонів дат та перетворення частоти, статистики рухливого вікна, лінійні регресії рухливого вікна, зсування дат та запізнювання.

Цю бібліотеку сильно оптимізовано за продуктивністю, критичні ланцюжки коду написано мовами Cython та C.[11]

2.5 Seaborn

Seaborn - це бібліотека візуалізації даних Python, заснована на matplotlib. Вона забезпечує інтерфейс високого рівня для малювання привабливих та інформативних статистичних графічних зображень.[10]

2.6 Scikit-learn

Scikit-learn (також відома як `sklearn` або `scikits.learn`) — це безкоштовна програмна бібліотека машинного навчання для мови програмування Python, яка надає функціональність для створення та тренування різноманітних алгоритмів класифікації, регресії та кластеризації, таких як лінійна регресія, random forest, градієнтний бустинг, і працює у зв'язці з бібліотеками NumPy та SciPy. Scikit-learn є однією з найбільш популярних бібліотек машинного навчання. [8]

З цієї бібліотеки будемо використовувати функціональність для побудування дерева рішень, щоб оцінити який тип мобільності найбільше впливає на захворюваність.

3 ПЕРВИННА РОЗВІДКА ДАНИХ

3.1 Базовий набір даних

Основним джерелом інформації є <https://ourworldindata.org>. Вибрав його так як він містить всю необхідну мені інформацію, таку як кількість тестів на тисячу чоловік, кількість смертей, кількість денних випадків зараження за всіма необхідними країнам. На тому ж сайті, всі ці дані представлені у вигляді інформативних графіків, з можливістю аналізу по днях.

Ще одне джерело даних вибрав репозиторій GitHub https://github.com/ActiveConclusion/COVID19_mobility. Він містить необхідні дані за весну по переміщенню громадян. Так само дані відсортовані, і сформовані в CSV файл, за яким зручно проводити аналіз засобами Python.

Додаткова інформація щодо обмежень в країнах була взята з офіційних сайтів.

<https://www.gov.uk/foreign-travel-advice/spain/coronavirus> - сайт уряду Великобританії, на якому розташована пам'ятка подорожуючим в Іспанію

<https://www.integrationsbeauftragte.de/ib-de/service/fragen-und-antworten/1731242-1731242?id=01-Intro-en> - німецький сайт, на якому розташовані обмеження Німеччини.

<https://www.covid.is/english> - урядовий сайт Ісландії, з якого були взяті дані по Ісландії.

<https://www.kmu.gov.ua/npas/pro-vstanovlennya-karantinu-z-metoyu-zapobigannya-poshirennyu-na-teritoriyi-ukrayini-gostroyi-respiratornoyi-hvorobi-covid-19-sprichinenoyi-koronavirusom-sars-cov-i200520-392> - сайт з

постановою Кабінету Міністрів України про введення карантин на початок весни.

<https://www.folkhalsomyndigheten.se/the-public-health-agency-of-sweden/communicable-disease-control/covid-19/regulations-and-general-guidelines/> - сайт містить правила, по карантину в Швеції.

Всі перераховані вище сайти містять необхідну інформацію для моїх досліджень, і відмінно підходять для аналізу.

3.2 Карантинні заходи

Будемо розглядати весняні піки в країнах, дані за весну досить повні, і можна буде провести нормальний аналіз. Почнемо зі складання списку обмежень і карантинних заходів введених в державах.

Іспанія [2]

1. Громадські збори обмежені максимум 6 людьми.
2. Соціальна дистанція 1,5 метра.
3. Обов'язкове використання масок в громадських місцях.
4. Всі магазини, підприємства і транспортні компанії зобов'язані зберігати контактну інформацію клієнтів (якщо така є) протягом 4 тижнів для відстеження.
5. Обмеження місткості на пляжах або в інших громадських місцях, і використання систем бронювання
6. При виявленні у відвідувачів позитивного результату або у них появи симптомів під час їх перебування в Іспанії, вони можуть бути переведені в спеціально відведене місце для запобігання подальшого поширення.
7. Часте і ретельно миття рук і використання дезінфікуючий засіб для рук, якщо мило і вода недоступні.

Німеччина [3]

1. Салони краси, тату-салони, масажні салони закриті. Перукарні залишаються відкритими.
2. Ночівля в готелях і т. д. Тільки при необхідності, а не в туристичних цілях.
3. Театри, опери, концертні зали закриті.
4. Торгові ярмарки, кінотеатри, тематичні парки, гральні заклади закриті.
5. Любительський спорт на приватних і громадських об'єктах, таких як доріжки, басейни, фітнес-студії, закриті.
6. Ресторани, бари, клуби, дискотеки, паби і т. д. закрито; дозволена доставка їжі і їжа на винос.
7. Роботодавцям рекомендується дозволяти співробітникам працювати з дому, де це можливо.
8. Школи і дитячі сади залишаються відкритими.

Ісландія [4]

1. Раннє виявлення зон ризику.
2. Всі жителі Ісландії, які прибувають із зони ризику, зобов'язані пройти карантин та/або пройти тестування.
3. Всіх, хто контактував з інфікованими, відправили на карантин.
4. Обмеження на зібрання.
5. Середні школи та університети, які використовують дистанційне навчання і обмежена кількість дитячих садків і початкових шкіл.

Україна [5]

На період дії карантину забороняється:

1. перебування в громадських будинках і спорудах, громадському транспорті без вдягнутих засобів індивідуального захисту, зокрема

респіраторів або захисних масок, що закривають ніс та рот, у тому числі виготовлених самостійно;

2. перебування на вулицях без документів, що посвідчують особу, підтверджують громадянство чи її спеціальний статус;

3. самовільно залишати місця самоізоляції, обсервації;

4. відвідування закладів освіти її здобувачами, крім участі у пробному зовнішньому незалежному оцінюванні, державній підсумковій атестації у формі зовнішнього незалежного оцінювання, вступних випробуваннях, єдиному вступному іспиті з іноземної мови, єдиному фаховому вступному випробуванні, атестації у формі тестових екзаменів ліцензійних інтегрованих іспитів “КРОК”, а також здійснення підтвердження кваліфікації моряків, проведення підготовки моряків, членів екіпажів торговельних суден, у тому числі тих, що здійснюють плавання внутрішніми водними шляхами України, та проведення підготовки судноводіїв малих/маломірних суден;

5. проведення масових (культурних, розважальних, спортивних, соціальних, релігійних, рекламних та інших) заходів з кількістю учасників більше ніж 10 осіб, крім заходів, необхідних для забезпечення роботи органів державної влади та органів місцевого самоврядування, навчально-тренувальних зборів спортсменів національних збірних команд України з олімпійських, неолімпійських, національних видів спорту та видів спорту осіб з інвалідністю, навчально-тренувального процесу спортсменів командних ігрових видів спорту професійних спортивних клубів;

6. робота закладів громадського харчування (ресторанів, кафе тощо), торговельно-розважальних центрів (крім розміщених у них магазинів), діяльність закладів, що надають послуги з розміщення, закладів розважальної діяльності, фітнес-центрів, закладів культури

Швеція [6]

В Швеції обмеження носять рекомендаційний характер:

1. Обмежте будь-які нові близькі контакти.
2. Тримайтеся на відстані від інших і уникайте багатолюдних місць.
3. Працюйте з дому якомога частіше.
4. Переконайтеся, що ви подорожуєте таким чином, щоб звести до мінімуму ризик зараження.
5. Займайтеся спортом і проводите дозвілля таким чином, щоб звести до мінімуму ризик зараження.
6. Обмежте кількість пасажирів в автомобілі.
7. Обмежте кількість клієнтів, яким дозволено перебувати в їх приміщеннях одночасно.
8. Уникайте всіх інших зустрічей або проводите їх в цифровому форматі.

Дані за кількістю заражених дуже нестабільні / недостовірні, і не можна покладатися на них. Це пов'язано з різними видами тестування, аналізу, і піднесення інформації. Деякі країни наприклад можуть приховувати результати тестування, а деякі можуть використовувати свої (відмінні від усіх інших) алгоритми фіксації випадків зараження. Це знижує рівень довіри до таких даних як кількість підтверджених випадків в день. Візьмемо ще два додаткових джерела інформації, а саме моделювання випадків, і кількість смертей. Також додаймо інформацію з тестування, невелика кількість тестів може значно спотворити картину зараження.

3.3 Тестування

Розглянемо кількість тестів на тисячу чоловік. Гарну порівняльну статистику по кількості тестів в країнах можна отримати на <https://ourworldindata.org/coronavirus/country/spain?country=~ESP#how-many-tests-are-performed-each-day>

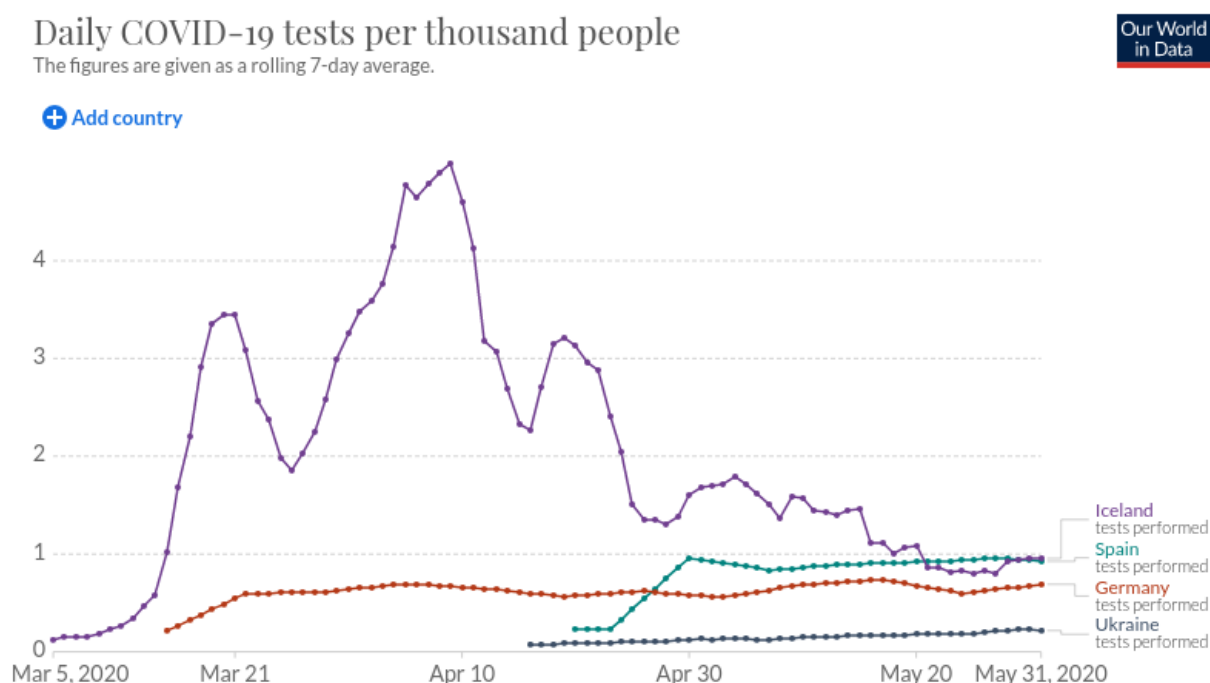


Рисунок 3.1 – Графік тестів по країнам на тисячу чоловік

На отриманому графіку можливо порівняти країни, і зробити припущення про те, в яких даних можуть бути спотворені через брак тестів. Також можна зробити припущення, що чим більше кількість тестів, тим точніші дані ми будемо мати, тому що через банальну нестачу тестування може занижуватиметься статистика. На даному графіку Україна має найменшу кількість тестів, врахуємо це в подальших дослідженнях. Інформація по Швеції відсутня на даному графіку.

3.4 Нові випадки захворювання по дням

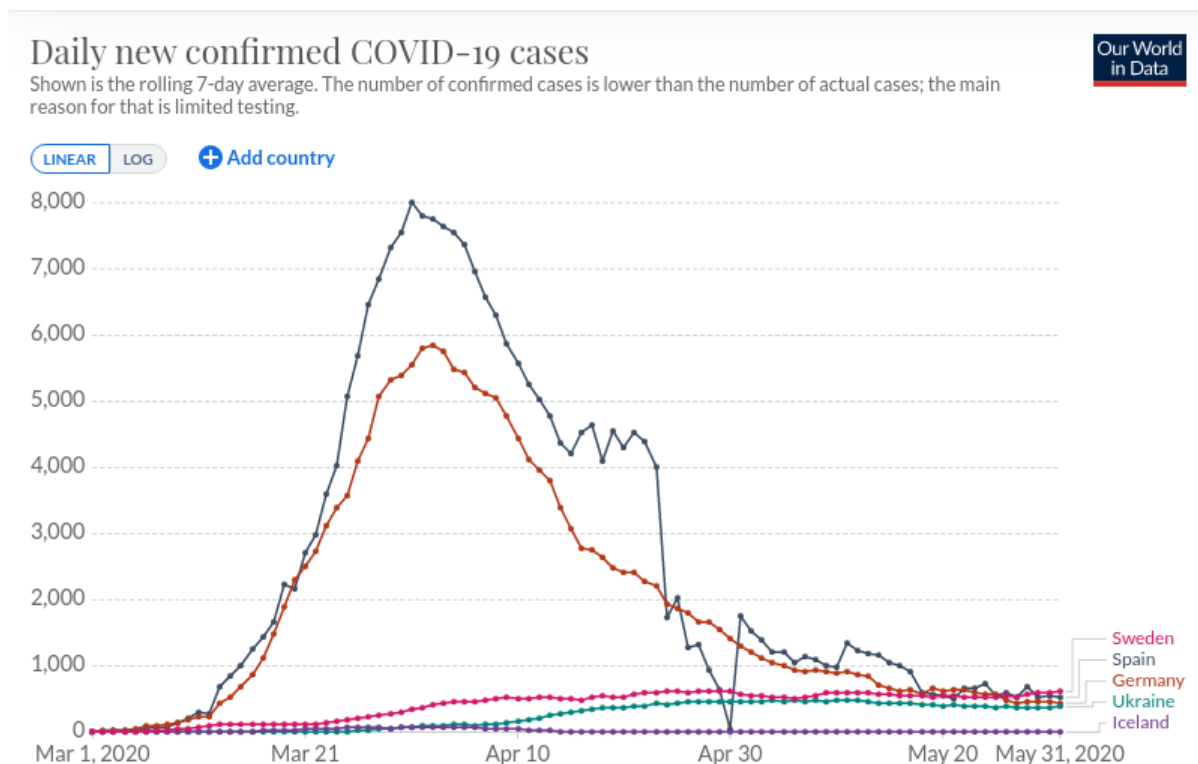


Рисунок 3.2 – Графік нових випадків захворювання по днях

На графіку можна спостерігати спалах заражень в Німеччині та Іспанії. В інших країнах графіки значно більш пологі, і можливо це результат дій уряду по введенню карантинних заходів. Або ж відмінності в алгоритмах фіксування хворих.

Також потрібно враховувати відмінність в кількості населення, країни з більшою кількістю населення мають більше хворих. Наприклад Німеччина має майже в два рази більше населення ніж Іспанія, але графік щоденних захворювань нижчий. Це свідчить про більш ефективну карантинну політику ніж в Іспанії.

3.5 Смерті по днях

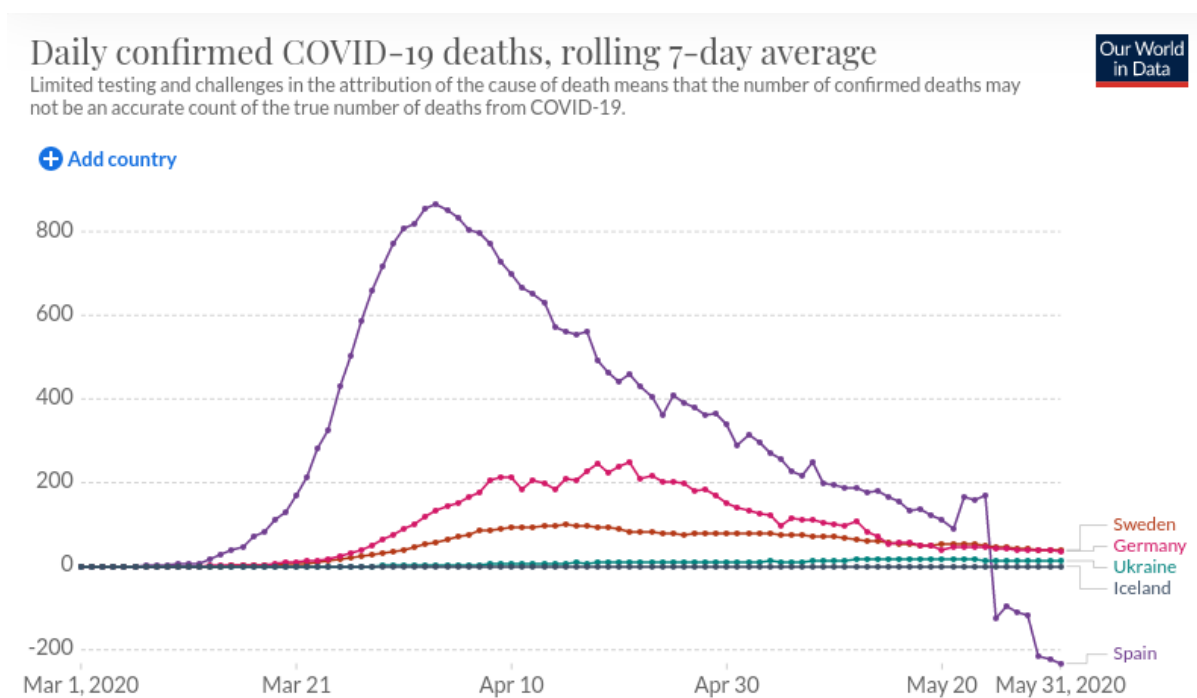


Рисунок 3.2 – Графік нових смертей по днях

Найбільшу кількість смертей ми можемо спостерігати в Іспанії. На другому місці розташувалася Німеччина. І вона має більш пологий графік смертності, при досить гострому графіку зараження. Подивимося на різницю між піками зараження, вона дорівнює десь 25%, кількість заражень в Німеччині на 25% менше ніж в Іспанії. А ось кількість смертей менше вже на 74%. Що наводить на думки, або про дуже хороше лікування в Німеччині, або відмінний від інших країн алгоритм фіксування смертей. Також помітне зростання смертей в Швеції, на відміну від України, і Ісландії, але при цьому графік зараження у них + - однаковий.

При додаванні такої величини як кількість населення при аналізі, можливо зробити висновки, що в Швеції кількість хворих на 1млн чоловік, була б значно вища ніж наприклад в Україні. Майже в 4 рази.

3.6 Оцінка коефіцієнту смертності (CFR)

Case fatality rate of the ongoing COVID-19 pandemic

The Case Fatality Rate (CFR) is the ratio between confirmed deaths and confirmed cases. During an outbreak of a pandemic the CFR is a poor measure of the mortality risk of the disease. We explain this in detail at OurWorldInData.org/Coronavirus

Our World
in Data

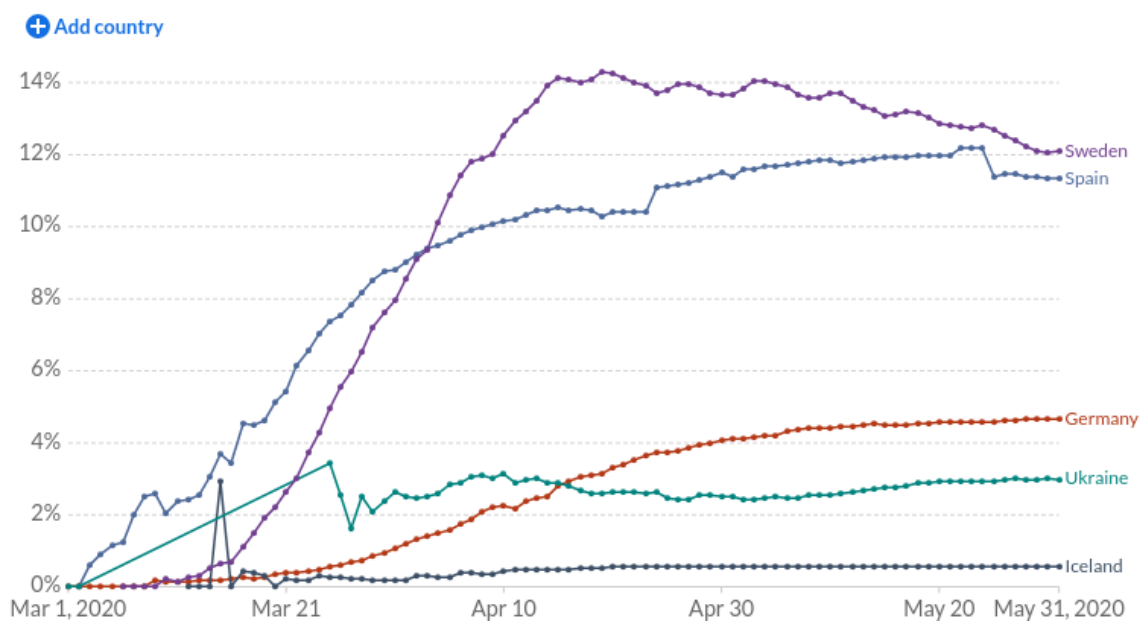


Рисунок 3.3 – Графік CFR по днях

Ці дані мають пряму кореляцію з кількістю заражених, і кількістю смертей, тому немає сенсу розглядати їх окремо. Графік просто наведено для порівняння, і так само як доказ того що, дані сильно нестабільні, і країни мають дуже великий розкид CFR.

3.7 Оцінка жорсткості обмежень

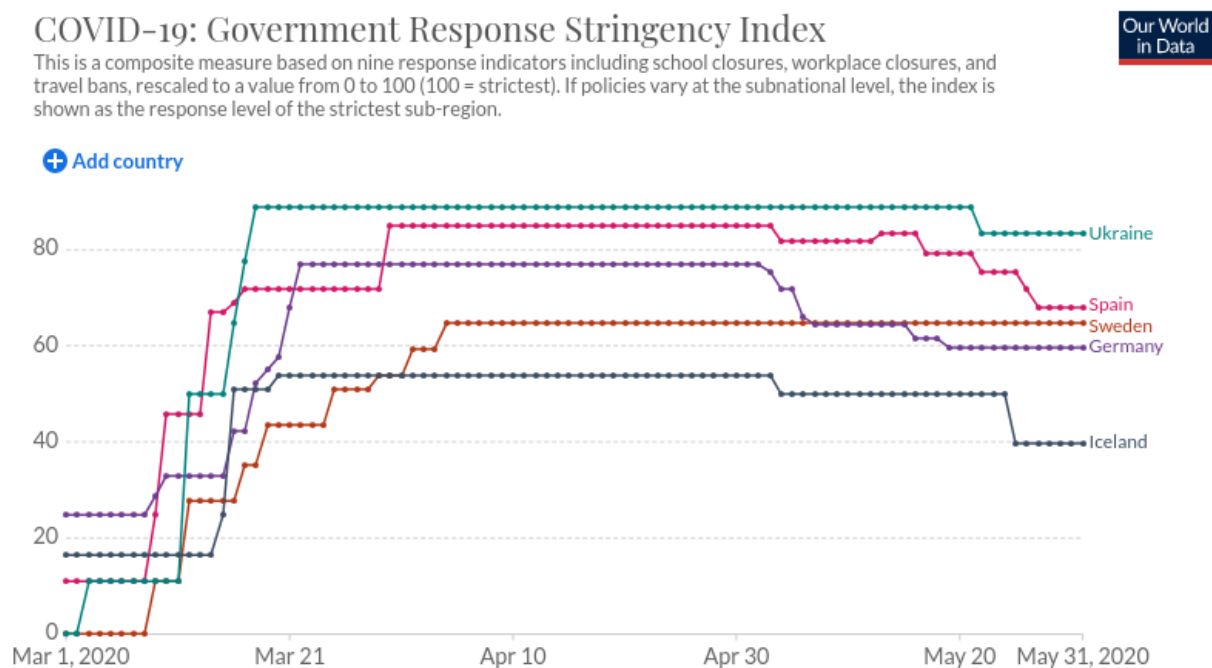


Рисунок 3.4 – Графік оцінки жорсткості обмежень по днях

На даному графіку Україна має найбільш жорсткі обмеження, це дозволило уникнути спалаху під час весни. Так само введення карантину в Іспанії і Німеччини дозволило звести нанівець спалах і стабілізувати ситуацію.

3.8 Переміщення громадян

Дані були отримані з

https://github.com/ActiveConclusion/COVID19_mobility. Потім за допомогою Python і jupyter notebook були обрані дані за весну, і побудовані графіки по ним. Нульова лінія, це середні дані за лютий, до введення карантину.

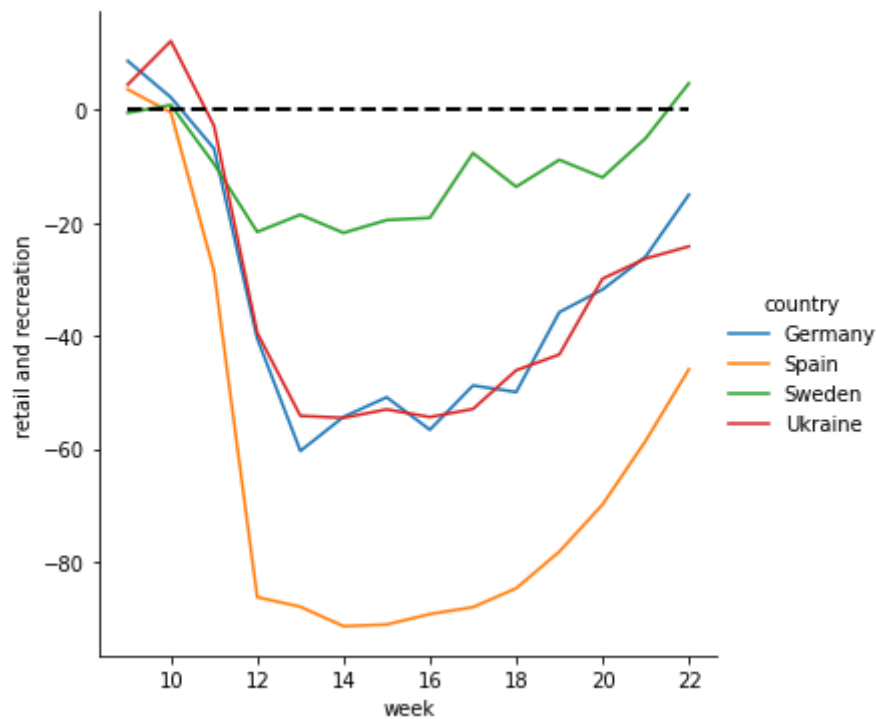


Рисунок 3.5 – Графік змін перебувань громадян у зонах торгівлі та відпочинку

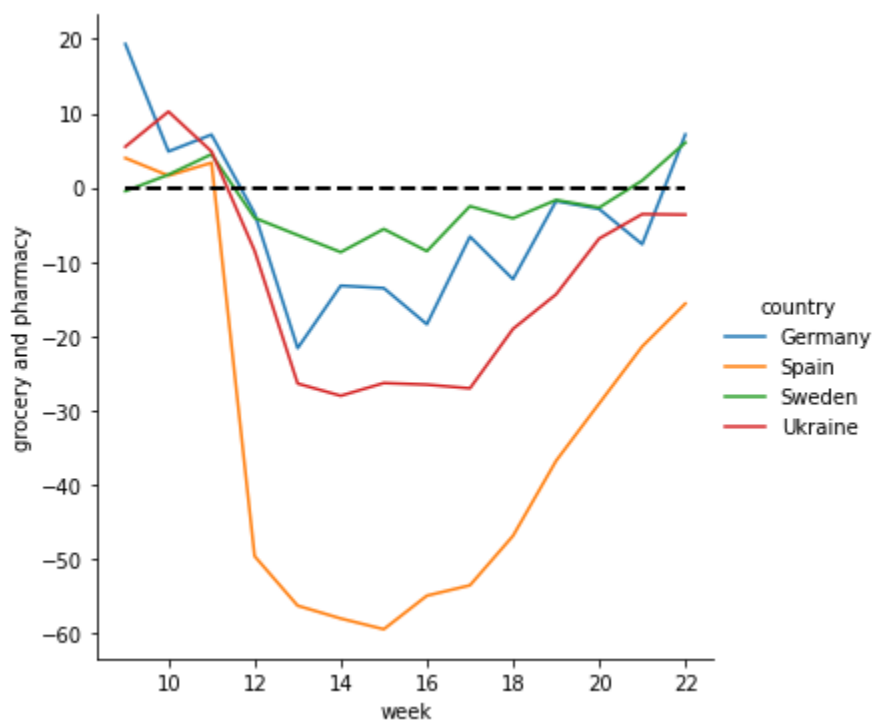


Рисунок 3.6 – Графік змін перебувань громадян у продовольчих магазинах та аптеках

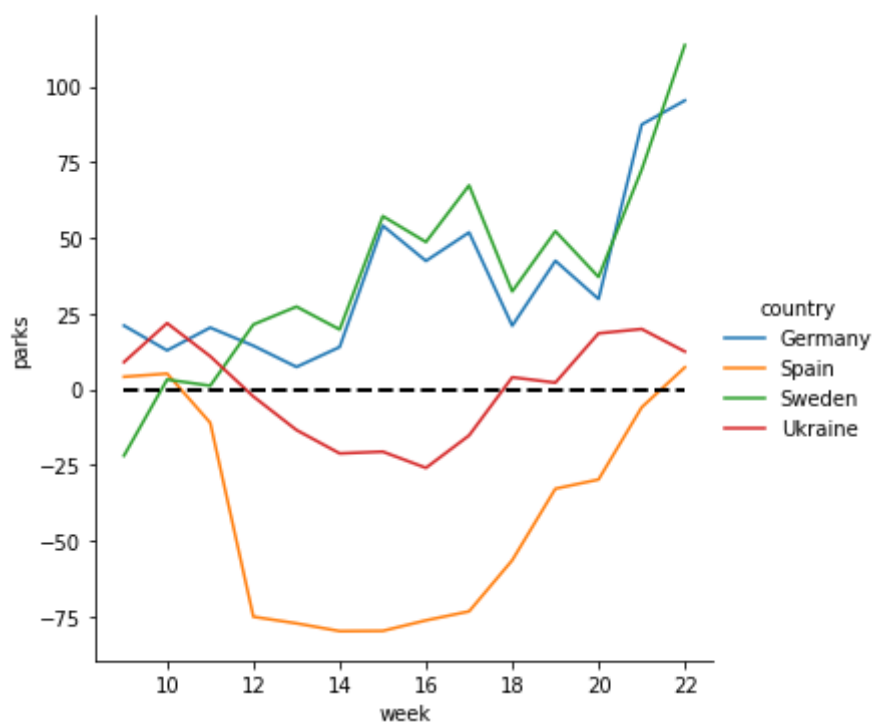


Рисунок 3.7 – Графік змін перебувань громадян у парках

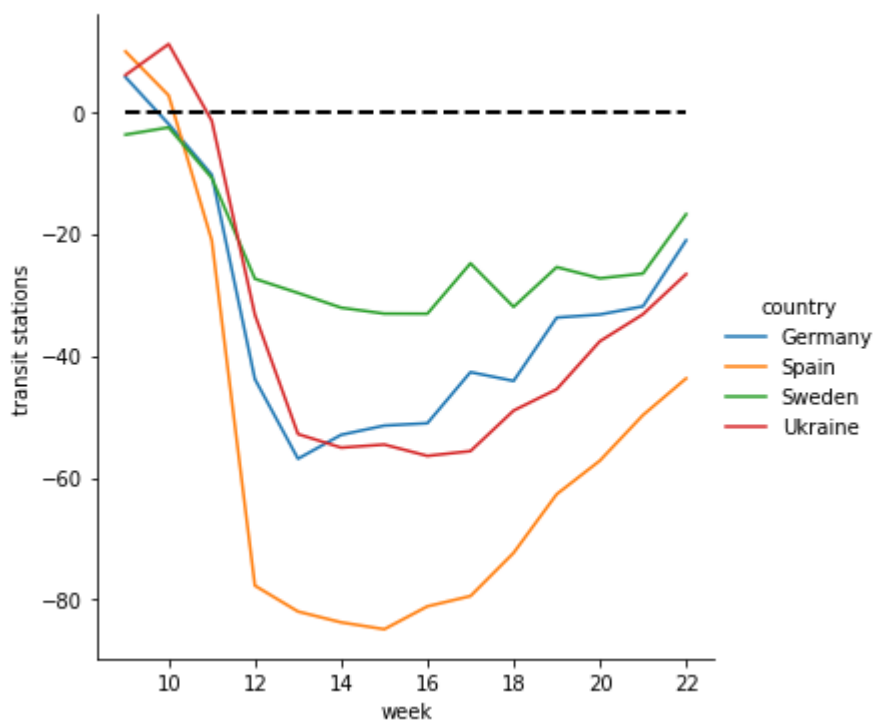


Рисунок 3.8 – Графік змін перебувань громадян на зупинках транспорту

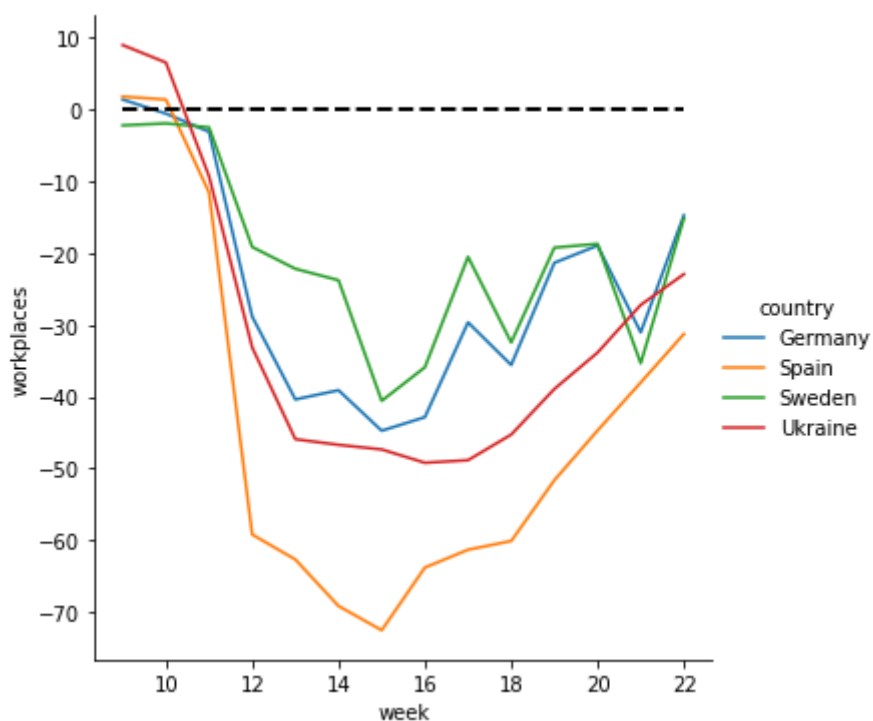


Рисунок 3.9 – Графік змін перебувань громадян у місцях роботи

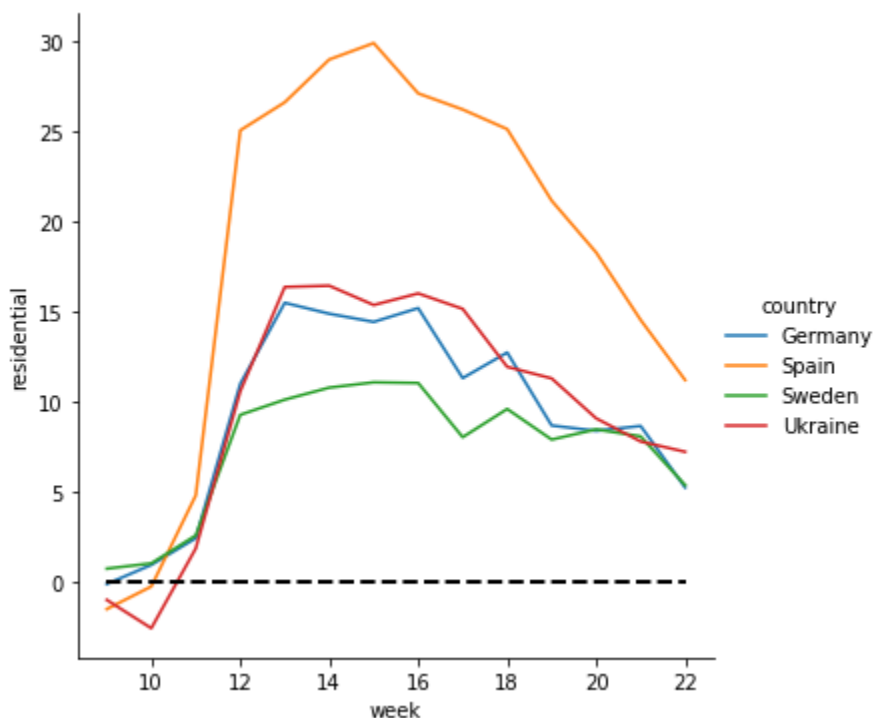


Рисунок 3.10 – Графік змін перебувань громадян у власному житлі

За графіками можна спостерігати, що жителі Іспанії добре дотримуються карантин, і завдяки цьому було можливо придушити спалах заражень. Хоча в Україні і вище рівень обмежень, але цифри значно відрізняються від Іспанії, думаю тут сильно впливає менталітет і недовіру до уряду.

4 АНАЛІЗ ТА ПОБУДОВА ПРИПУЩЕНЬ

Дослідження умовно можна розділити на декілька частин, перша це підготування інструментів для аналізу — написання власних функцій. Друга частина аналіз та порівняння базових графіків мобільності/захворюваності. Третя частина побудування кроскореляційних мап та дерев рішень, щоб зробити припущення, який вид мобільності найбільше впливає на захворюваність.

4.1 Підготування функцій Python для аналізу

Створимо новий Jupyter notebook та імпортуємо усі необхідні бібліотеки.

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

from typing import Tuple

import warnings
warnings.simplefilter(action='ignore', category=FutureWarning)
```

Рисунок 4.1 – Імпорт бібліотек

Підготуємо дані для подальшого аналізу. Завантажимо набір даних по мобільності.

```
def calculate_week_num(row):
    if row['year'] == 2021 and row['week'] < 53:
        return row['week']+52
    return row['week']

data_mobility = pd.read_csv('mobility_report_europe.csv')
data_mobility['year'] = pd.DatetimeIndex(data_mobility['date']).year
data_mobility['week'] = pd.DatetimeIndex(data_mobility['date']).week
data_mobility['week'] = data_mobility.apply(calculate_week_num, axis=1)

data_mobility = data_mobility.drop('year', axis=1)
```

Рисунок 4.2 – Завантаження набору даних по мобільності

Завантажимо набір даних по новим випадкам захворювання.

```
cases_new = pd.read_csv('owid-covid-data.csv')
cases_new = cases_new[['date', 'location', 'new_cases']]
cases_new['week'] = pd.DatetimeIndex(cases_new['date']).week
cases_new['year'] = pd.DatetimeIndex(cases_new['date']).year

cases_new['week'] = cases_new.apply(calculate_week_num, axis=1)

cases_new = cases_new.drop('year', axis=1)
```

Рисунок 4.3 – Завантаження набору даних по новим випадкам захворювання

Будуємо перетин тижнів з одного датасета та іншого, таким чином ми отримаємо доступний діапазон тижнів для аналізу. Будуємо функції для вибірки даних по країні.

```
supported_weeks = set(data_mobility.week.unique()).intersection(set(cases_new.
    week.unique()))
def get_data_for_country_v2(country: str) -> Tuple[pd.DataFrame, pd.DataFrame]:
    mobility = data_mobility[data_mobility['week'].isin(supported_weeks)]
    cases = cases_new[cases_new['week'].isin(supported_weeks)]

    mobility = mobility[mobility['country'] == country].groupby('week').mean()
    cases = cases[cases['location'] == country].groupby('week').sum()

    return mobility, cases
```

Рисунок 4.4 – Функція для вибору даних по країні

Будуємо функцію на базі функції описаної вище, для побудови графіків по країні.

```
def build_two_graphic_for_country(country, week_start=None, week_end=None):
    mobility, cases = get_data_for_country_v2(country)

    f, ax = plt.subplots(nrows=2, figsize=(14,14))

    cases.iloc[week_start:week_end].plot(ax=ax[0],
                                         title='Cases')

    mobility.iloc[week_start:week_end].plot(ax=ax[1],
                                             title='Mobility',
                                             ylim=[-100,50])
```

Рисунок 4.5 – Функція для побудови графіків по країні

Також підготуємо функцію для побудови кроскореляційної мапи, яка буде використовуватися в подальшій роботі.

```
def corr(df_1, df_2):
    a = pd.merge(df_1, df_2, how='left', left_on=['week'], right_on = ['week'])
    corr_result = a.corr()
    corr_result = corr_result.drop(['new_cases'])
    corr_result = corr_result.drop(['retail and recreation',
                                    'grocery and pharmacy',
                                    'parks',
                                    'transit stations',
                                    'workplaces',
                                    'residential'], axis=1)
    corr_result = corr_result.T
    return corr_result
```

Рисунок 4.6 – Функція для побудови кореляції між двома наборами даних (мобільність та нові випадки захворювання)

```
def build_cross_corr_for_country_v2(country, shift_mobility=False,
    week_start=None, week_end=None):
    """
    week_start - num week when quarantine start
    week_end - num week when quarantine end
    """
    mobility, cases = get_data_for_country_v2(country)

    stable = mobility
    shifted = cases[['new_cases']]
    if shift_mobility:
        stable, shifted = shifted, stable

    cross_corr = [corr(stable.iloc[week_start:week_end], shifted.shift(-i)
        week_start:week_end)]
        for i in range(len(stable.index)-(week_end or 2))]

    week_result_limit = 10
    cross_corr = pd.concat(cross_corr[:week_result_limit])
    cross_corr = cross_corr.reset_index(drop=True)

    f, ax = plt.subplots(figsize=(15, 7))
    a = sns.heatmap(cross_corr, linewidths=.1, cmap="YlGnBu", annot=True)
```

Рисунок 4.7 – Функція для побудови кросскореляції для обраної країни

Створюємо дерево рішень для аналізу даних, на аналізі якого в подальшому побудуємо припущення про те, який вид мобільності найбільше впливає на захворюваність.

```
from sklearn.model_selection import cross_val_score
from sklearn.tree import DecisionTreeRegressor
from sklearn.tree import export_graphviz
from IPython.display import Image
```

Рисунок 4.8 – Імпортуємо необхідні модулі для використання дерева рішень

Функція для побудови дерева рішень по обраній країні, також опціонально можливо задати зсув даних, щоб проаналізувати дію карантину впродовж часу.

```
def build_tree_for_country(country: str, shift: int =0):
    mobility, cases = get_data_for_country_v2(country)

    if shift:
        mobility = mobility[:-shift]
        cases = cases.shift(shift)[shift:]

    x = mobility.values
    y = cases.values
    y[y < 1] = 1
    y = np.log(cases.values)
    y[y == -inf] = 0

    tree = DecisionTreeRegressor(random_state=10, max_depth=3)
    tree.fit(x, y)
    score = tree.score(x, y)

    export_graphviz(tree, feature_names=mobility.columns,
out_file='./tree.dot', filled=True)
    !dot -Tpng './tree.dot' -o './tree.png'

    feature_importance = dict(zip(mobility.columns, tree.feature_importances_))
    return score, feature_importance
```

Рисунок 4.9 – Функція для побудови дерева рішень, та отримання коефіцієнтів важливості ознак

Створимо функцію, яка побудує дерева рішень, по країнам (32 країни), доступним з двох датасетів, та знайде середнє значення важливості ознак

```
def calculate_mean_feature_importance(shift: int = 0):
    feature_importance_for_country = []
    for country in supported_countries:
        score, feature_impotance = build_tree_for_country(country, shift)
        feature_importance_for_country.append(feature_impotance)

    feature_importance_mean = {}
    for feature_name in feature_importance_for_country[0].keys():
        a = [country_feature[feature_name] for country_feature in
↪feature_importance_for_country]
        feature_importance_mean[feature_name] = sum(a)/len(a)

    feature_importance_mean = {k: v
↪for k, v in sorted(feature_importance_mean.
items(),
                    key=lambda item: item[1],
                    reverse=True)}

    return feature_importance_mean
```

Рисунок 4.10 – Функція для аналізу важливості ознак

Маленька функція для виводу важливості різних типів мобільності у вигляді стовпчастої діаграми.

```
def build_feature_plot(feature_importance: dict):
    f, ax = plt.subplots(figsize=(15, 7))

    plt.bar(range(len(feature_importance)), list(feature_importance.values()),
↪align='center')
    plt.xticks(range(len(feature_importance)), list(feature_importance.keys()))

    plt.show()
```

Рисунок 4.11 – Функція для аналізу важливості ознак

4.2 Дослідження графіків мобільності/захворюваності

Усього для візуального дослідження було обрано 5 країн, які розглядалися вище, та за допомогою функцій були побудовані графіки. Перша країна за якою були побудовані графіки це Україна.

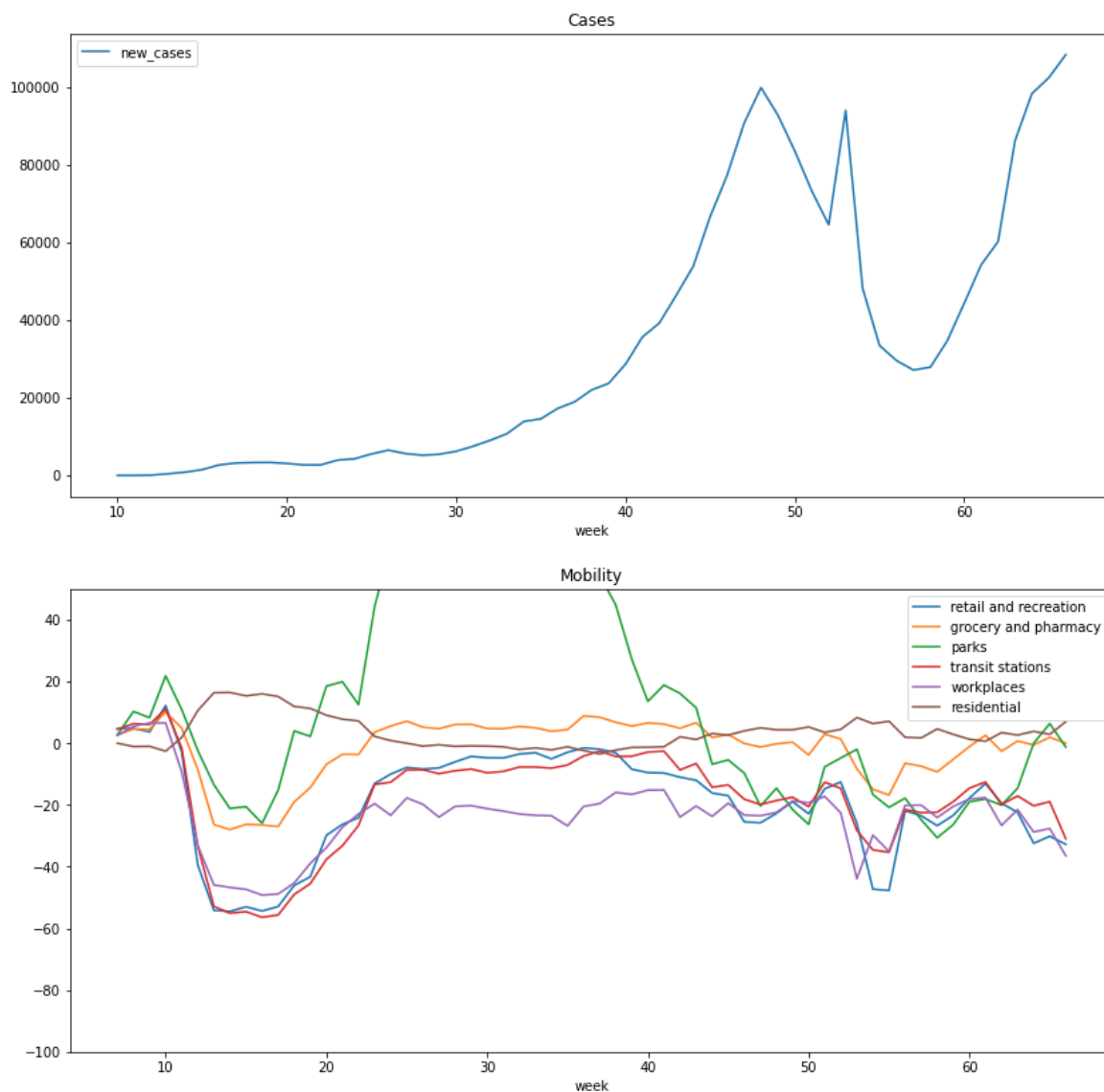


Рисунок 4.12 – Графік мобільності та захворюваності протягом року в Україні

Нумерування тижнів на графіку починається з 2020 року, а в 2021 додається 52, щоб продовжити графік.

Перша частина графіку припадає на час, коли був впроваджений весняний карантин. Графік захворюваності дуже пологий, відносно осіннього піку. Одна з можливих причин, введення карантину. Далі, в проміжку з 25-50 тиждів, були скасовані карантинні обмеження, і після цього відбулося зростання переміщень громадян, що цілком ймовірно призвело до зростання захворюваності. На базі цих спостережень сформуємо першу гіпотезу — зростання мобільності(відносно значень під час карантину) спонукає підвищенню кількості нових випадків захворювання.

Наступним буде графік Іспанії

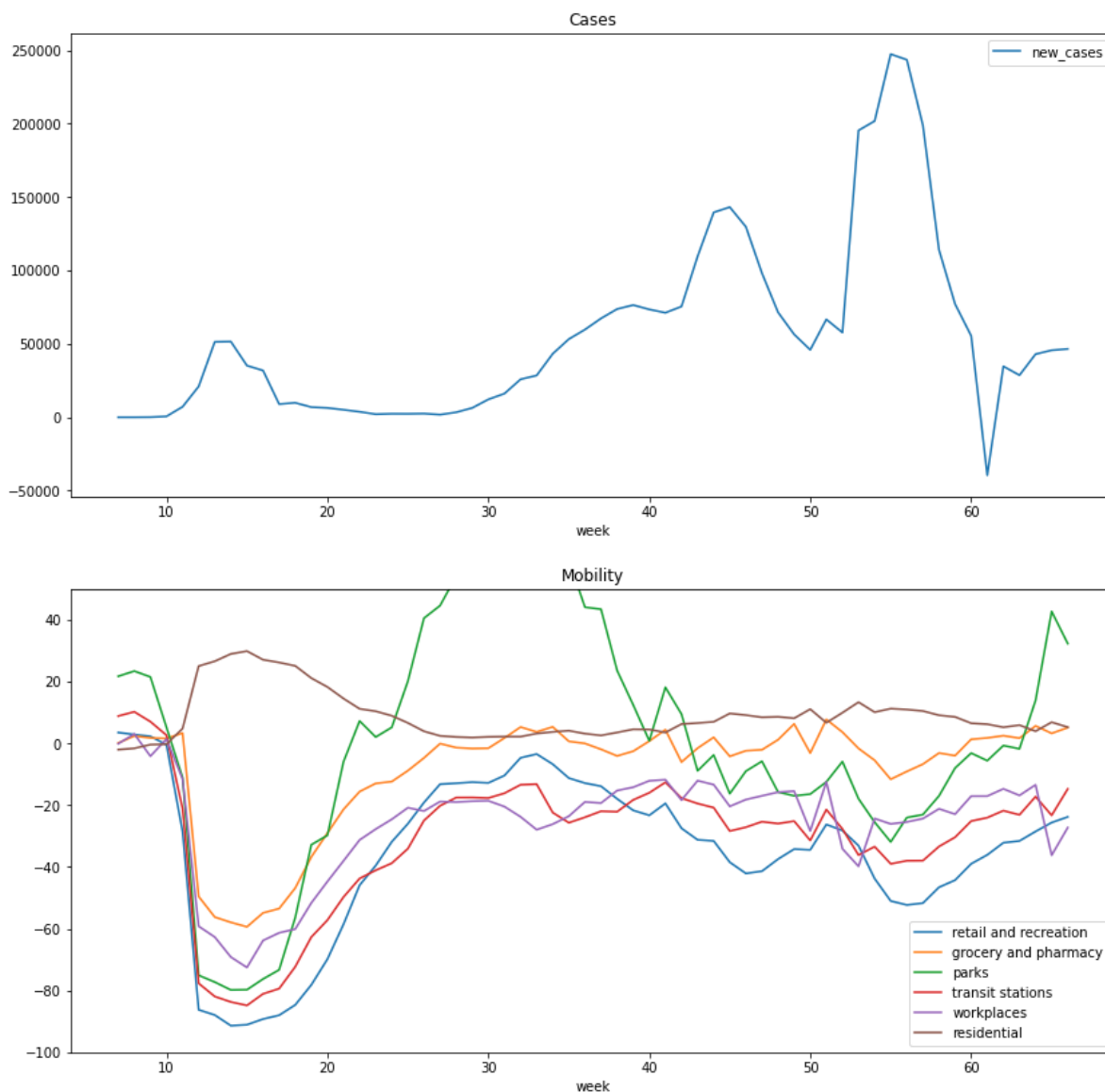


Рисунок 4.13 – Графік мобільності та захворюваності протягом року в Іспанії.

В Іспанії також простежується динаміка, зниження числа нових випадків захворювання, при зниженні мобільності. На загальному графіку можливо побачити, що зниження мобільності в проміжок 10 - 20 тижднів, дозволило зупинити зростання захворюваності. Подальше повернення мобільності на нульовий рівень призвело до осіннього піку. Але паттерн,

зниження мобільності призводить до зменшення захворюваності, можливо простежити як навесні, так і восени.

Таким чином на основі вище приведених даних, отриманих, з результатів попереднього аналізу, засобами описаними у підрозділі 4.1, сформулюємо гіпотезу — зниження мобільності, сприяє зниженню захворюваності.

А у нас на черзі Німеччина. Загальний графік за рік.

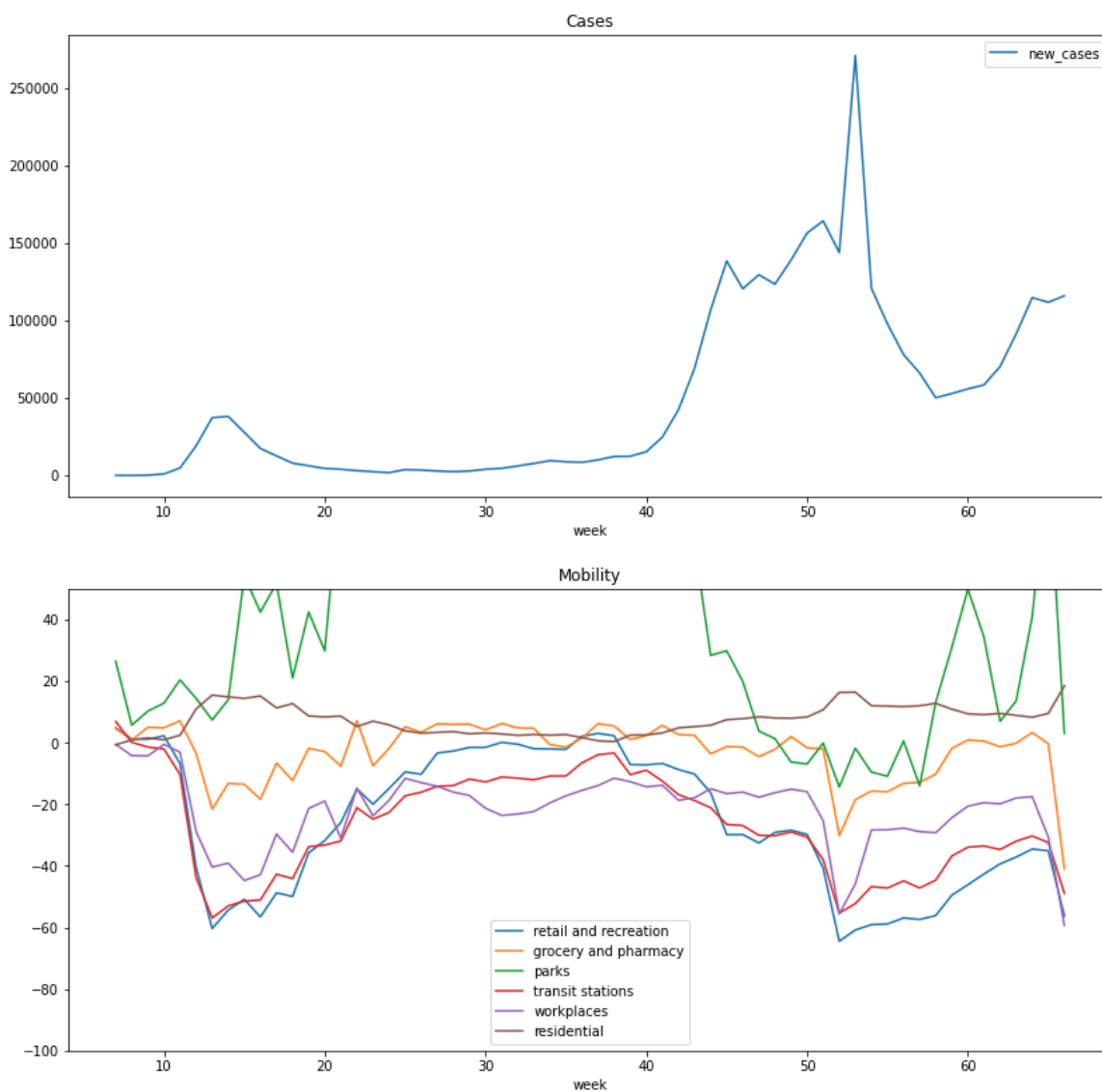


Рисунок 4.14 – Графік мобільності та захворюваності протягом року в Німеччині.

На даному графіку добре простежуються гіпотези, про вплив мобільності, на захворюваність. Видно що зниження мобільності на проміжку з 10-го по 20-й тиждень, дало в результаті значне зниження рівня захворюваності, пік до 20й тижня досяг майже мінімального значення.

Дані по Ісландії на жаль відсутні. Тому остання в цьому списку буде Швеція.

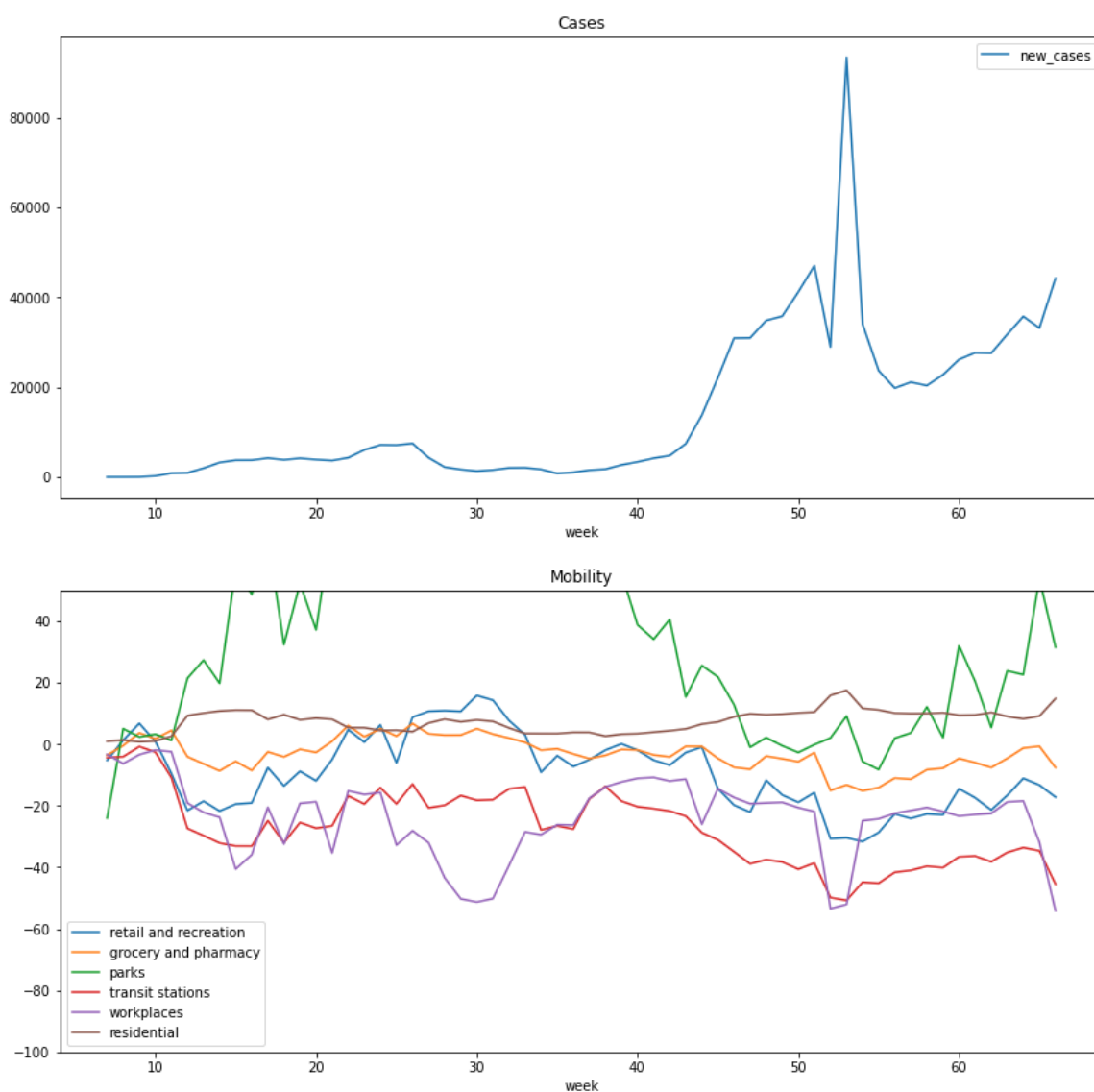


Рисунок 4.15 – Графік мобільності та захворюваності протягом року в Швеції.

На цьому графіку також простежується та ж динаміка, що й на графіках вище. Хоча за графіком мобільності Швеції, можна спостерігати, що він хоч і не досяг відміток Німеччини в орієнтовно 50% зниження мобільності, за основними напрямками, але має місце зниження в 20%, а у напрямку retail and recreation і transit station до 40%. Що в свою чергу позитивно вплинуло, на зниження кількості захворілих.

Також додатково розглянув наступні країни

1. Austria
2. Finland
3. France
4. Ireland
5. Norway
6. Switzerland
7. United Kingdom

Цікаві дані вдалося отримати, за допомогою цього розширення, наших досліджень. На всіх графіках простежується один і той же патерн, з подвійним спалахом за рік. Осінь, весна. та другий спалах більш розтягнутий.

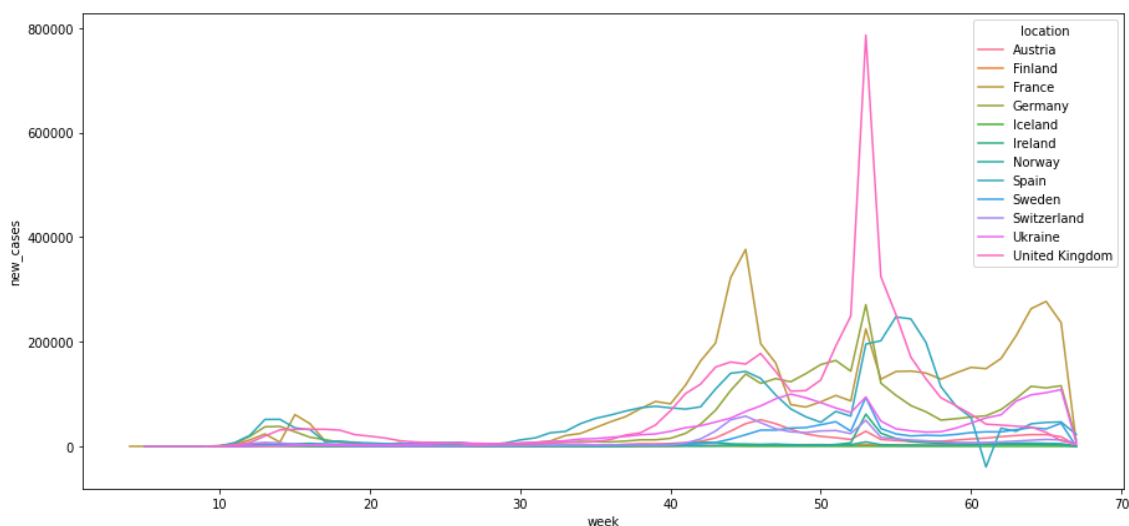


Рисунок 4.16 – Графік захворюваності різних країн

Також на базі побудованих графіків, можливо зробити припущення, що зниження мобільності, впливає на зниження кількості нових випадків захворювання з затримкою, а саме 5 тижнів. Це можливо простежити на усіх побудованих графіках. Також, якщо додати зсув захворюваності, на 5 тижнів, відносно мобільності, зменшення значень будуть майже одне над одним, з похибкою ± 2 тижні. Наведемо приклад графіків із зсувом. Ця інформація буде використана в подальшому аналізі.

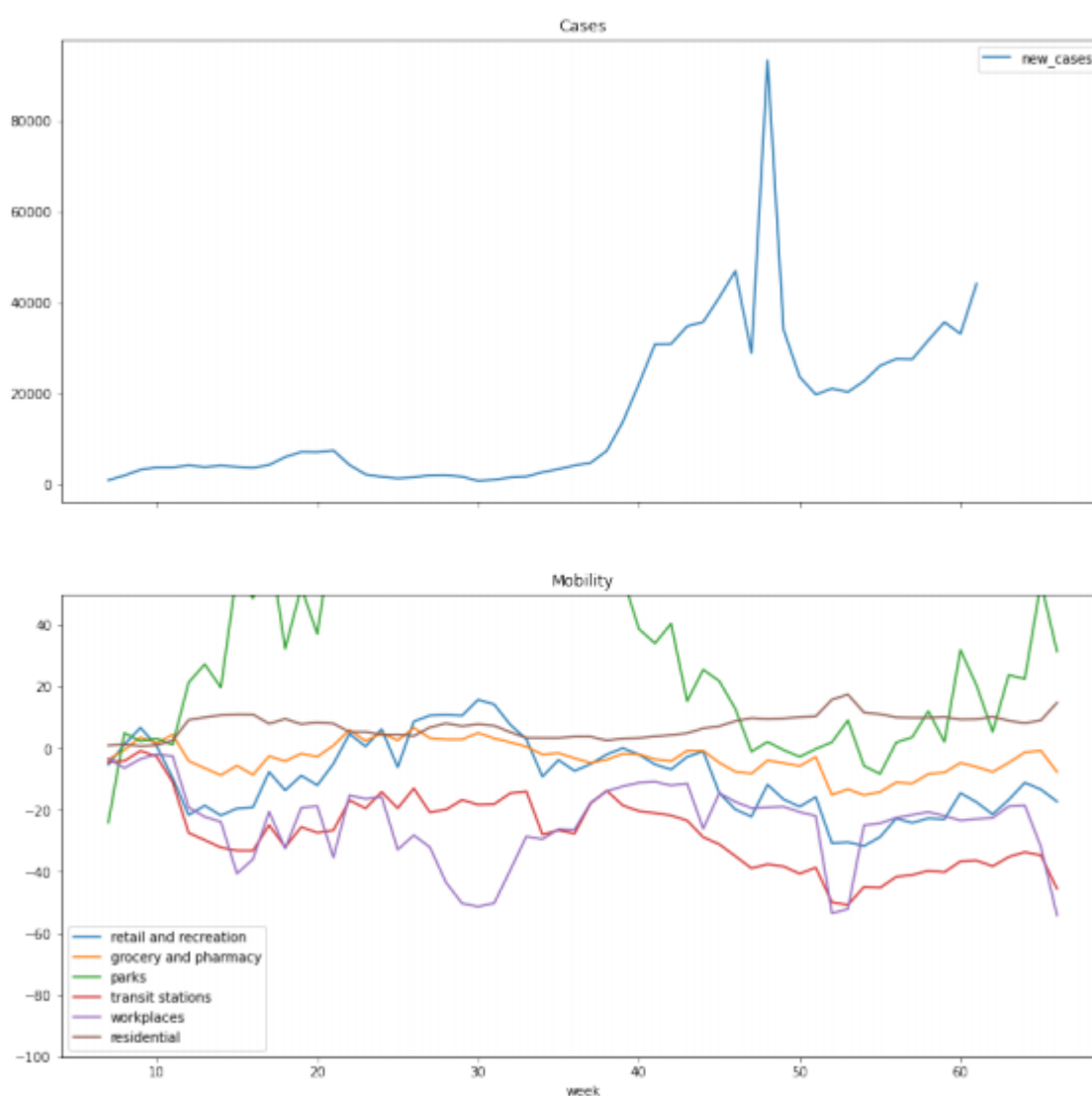


Рисунок 4.17 – Графік Швеції із зсувом 5 тижнів

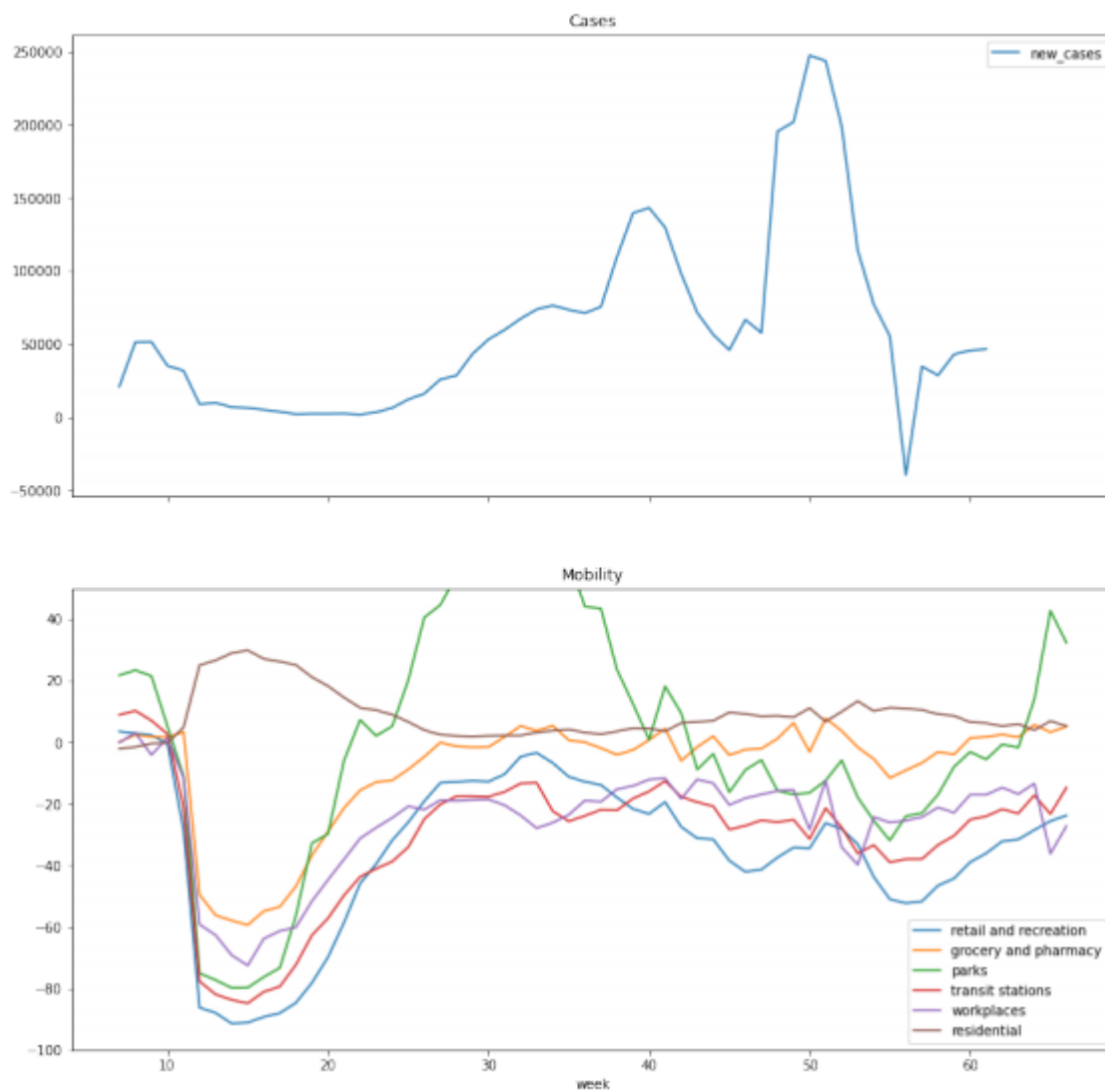


Рисунок 4.18 – Графік Іспанії із зсувом 5 тижнів

4.3 Дослідження видів мобільності

Додатковим інструментом, для аналізу гіпотез, була використана крос-кореляція. Був обраний осінній пік для розгляду. Випадки захворюваності зсувалися щодо мобільності, за цей період, щоб підтвердити, або спростувати раніше запропоновані гіпотези. Вийшов неоднозначний результат, в половині країн вона показувала позитивні результати, а в іншій половині негативні. Тому не став наводити приклади.

Другий інструмент використаний для аналізу — дерева рішень. Деревя рішень дозволяють нам використовувати їх з неоднорідними даними, їх легко візуалізувати та пояснити, та за допомогою них, можливо отримати інформацію, яка ознака найбільше впливає на результат.

В нашому випадку дерево рішень використовується не для прогнозування, а лише для аналізу різних видів мобільності, щоб зробити припущення, який з них, найбільше впливає на розповсюдження інфекції.

Для роботи з деревом дані були попередньо підготовлені, кількість нових захворювань були переведені в логарифмічні значення.

Приклад побудованого дерева для України.

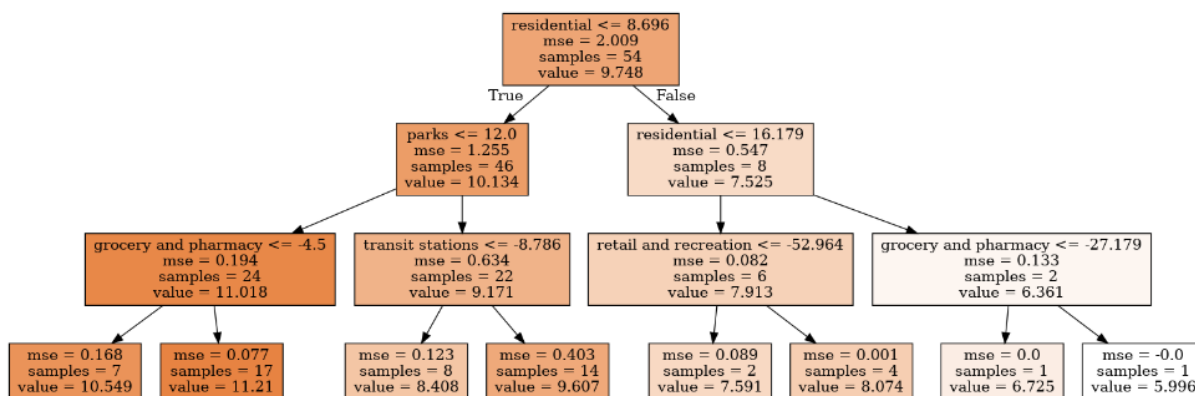


Рисунок 4.20 – Дерево рішень для України

У даному дереві ми бачимо, що перше розділення йде по residential, далі по всім іншим. Таким чином, ми за допомогою дерева отримуємо порядок у якому різні види мобільності впливають на кількість нових захворюваних.

Як це працює?

У двох словах, на кожній ітерації обирається випадкова ознака, та випадкове значення, з яким ця ознака порівнюється, і потім після розділення набору даних за цією ознакою, рахується ентропія(наскільки гарно розділені дані, одного типу зліва, іншого справа). На базі цього розрахунку обирається найкращий варіант розділення, і процес повторюється.

У нашому випадку не важлива результуюча точність, важливий лише порядок ознак.

Також потрібно виключити парки з дерева, бо дані по паркам дуже сильно корелюють із статистикою захворювань, також накладається сезонність цієї мобільності. Наприклад влітку, дуже зросла мобільність по паркам, але кількість захворювань була невелика, тому ми можемо побачити умову в дереві, що коли мобільність в парках нижче 12, велика кількість нових випадків хвороби.

Але у наведеному дереві, можливо побачити і дивні ознаки, які воно побудувало. Пояснення цьому, невеликий набір даних, та дуже великий розбіг.

Створимо таке дерево для кожної з 32х доступних країн, та побудуємо діаграму важливості ознак. Цифри на діаграмі не несуть корисної інформації, через те, що дані дуже різноманітні, та побудування дерева відбувалося без контрольної виборки, але порядок зрозумілий.

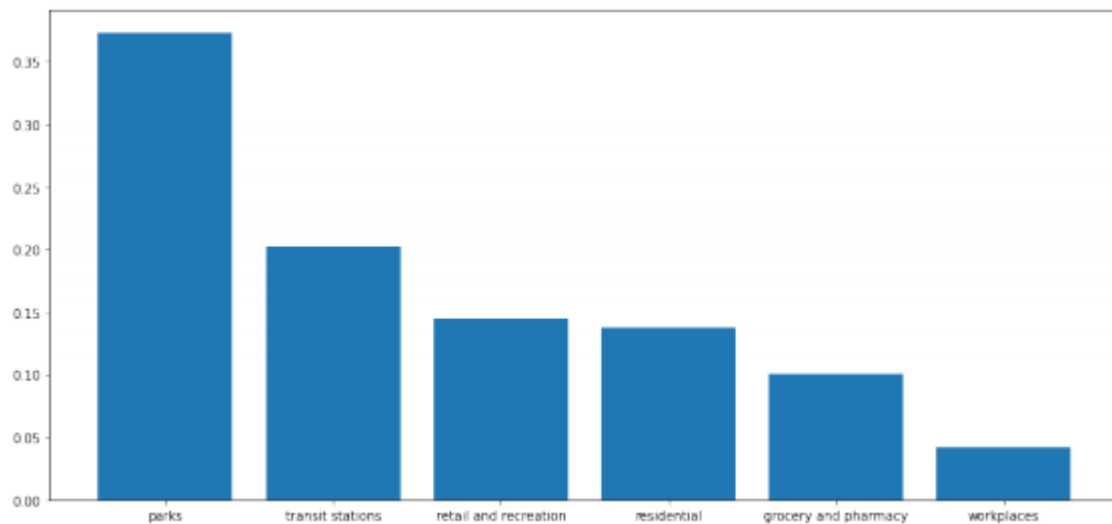


Рисунок 4.21 – Діаграма важливості ознак, за 32 країнами із зсувом на 5 тижнів

На графіку бачимо що парки на першому місці, це саме та кореляція про яку йшла мова вище. Але наступний порядок ознак, не суперечить здоровому глузду, тому створимо на базі нього припущення. Переміщення на автобусних зупинках в першу чергу мають в плив на захворюваність, далі місця відпочинку, та кількість часу проведеного вдома.

ВИСНОВОК

Стратегія країни складається з двох частин, карантинних заходів та часу їх впровадження. Тому подивимось до цих двох пунктів. Бо приділення недостатньої уваги часу впровадження карантинних заходів, потребує в подальшому жорсткіших карантинних заходів, через сплеск захворюваності.

Під час дипломної роботи була створена програма для аналізу даних отриманих з відкритих джерел. Вона будує графіки за цими даними, та дозволяє проводити аналіз карантинних стратегій різних країн.

Найкращі стратегії за результатами аналізу, мають Німеччина з Іспанією.

Гарні результати Німеччини, були досягнуті, завдяки приділенню гарної уваги як часу впровадження карантинних заходів, так і самим карантинним заходам.

У Ісландії також були своєчасно впровадженні заходи, та в додаток до цього, це острівна країна, з невеликою кількістю населення, що дозволило виявляти спалахи та зводити нанівець.

Швеція напроти, показала не такі гарні результати як очікувалось, відмова від жорстких заходів, та застосування лише обмежень дало приріст захворюваності.

Україна вводила жорсткий карантин навесні, що дозволило уникнути піку, але можливо теж самий результат міг бути досягнутий, і менш жорсткішими заходами.

Також окремо були розглянутий вплив мобільності на захворюваність, та отримано припущення, що в першу чергу, впливають переміщення громадян по зупинках загального транспорту та місцях для відпочинку.

ПЕРЕЛІК ПОСИЛАНЬ

1. Агреговані дані по мобільності [Електронний ресурс] — Режим доступу: https://github.com/ActiveConclusion/COVID19_mobility. — Назва з екрану.
2. Карантинні обмеження Іспанії [Електронний ресурс] — Режим доступу: <https://www.gov.uk/foreign-travel-advice/spain/coronavirus>. — Назва з екрану.
3. Карантинні обмеження Німеччини [Електронний ресурс] — Режим доступу: <https://www.integrationsbeauftragte.de/ib-de/service/fragen-und -antworten/1731242-1731242?id=01-Intro-en>. — Назва з екрану.
4. Карантинні обмеження Ісландії [Електронний ресурс] — Режим доступу: <https://www.covid.is/english>. — Назва з екрану.
5. Постанова про встановлення карантину з метою запобігання поширенню на території України гострої респіраторної хвороби COVID-19 [Електронний ресурс] — Режим доступу: <https://www.kmu.gov.ua/npas/pro-vstanovlennya-karantinu-z-metoyu-zapobigannya-poshirennyu-na-teritoriyi-ukrayini-gostroyi-respiratornoyi-hvorobi-covid-19-sprichinenoyi-koronavirusom-sars-cov-i200520-392https://www.folkhalsomyndigheten.se/the-public-health-agency-of-sweden/communicable-disease-control/covid-19/regulations-and-general-guidelines/>. — Назва з екрану.
6. Кількість протестованих за день, огляд по країнах [Електронний ресурс] — Режим доступу: <https://ourworldindata.org/coronavirus/country/spain?country=~ESP#how-many-tests-are-performed-each-day>. — Назва з екрану.
7. Навчальний посібник з методів наукових досліджень [Електронний ресурс] — Режим доступу: <https://core.ac.uk/download/pdf/159817923.pdf> — Назва з екрану.

8. Документація з використання бібліотеки Scikit-learn [Електронний ресурс] — Режим доступа: https://scikit-learn.org/stable/user_guide.html — Назва з екрану.

9. Документація з використання jupyter notebook [Електронний ресурс] — Режим доступа: <https://jupyter.org/documentation>. — Назва з екрану.

10. Документація з використання Seaborn [Електронний ресурс] — Режим доступа: <https://seaborn.pydata.org/tutorial.html><https://uk.wikipedia.org/wiki/Pandas>. — Назва з екрану.

11. Документація pandas [Електронний ресурс] — Режим доступа: https://pandas.pydata.org/docs/user_guide/index.html#user-guide. — Назва з екрану.

12. Документація з оформлення коду на Python [Електронний ресурс] — Режим доступа: <https://www.python.org/dev/peps/pep-0008/>. — Назва з екрану.

13. Документація з бібліотеки matplotlib [Електронний ресурс] — Режим доступа: <https://matplotlib.org/stable/tutorials/index.html>. — Назва з екрану.

14. Путівник мовою Python [Електронний ресурс] — Режим доступа: <https://pythonguide.rozh2sch.org.ua/>. — Назва з екрану.