

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)

Кафедра Штучного інтелекту
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти другий (магістерський)

Дослідження та прогнозування міграційних тенденцій громадян України
методами аналізу даних
(тема)

Виконав:

здобувач другого року навчання,

групи ДСМ-23-1

Моргун О.Ю.

(прізвище, ініціали)

Спеціальність 122 Комп'ютерні науки

(код і повна назва спеціальності)

Тип програми освітньо-професійна

(освітньо-професійна або освітньо-наукова)

Освітня програма Науки про дані (Data Science)

(повна назва спеціалізації)

Керівник доц. Волощук О.Б.

(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри

(підпис)

О.В. Золотухін

(прізвище, ініціали)

2025 р.

Харківський національний університет радіоелектроніки

Факультет _____ Комп'ютерних наук
(повна назва)
Кафедра _____ Штучного інтелекту
(повна назва)
Рівень вищої освіти _____ другий (магістерський)
Спеціальність _____ 122 Комп'ютерні науки
(код і повна назва)
Тип програми _____ освітньо-професійна
(освітньо-професійна або освітньо-наукова)
Освітня програма _____ Науки про дані (Data Science)
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____

(підпис)

«_____» _____ 20 25_ р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві _____ Моргун Олесі Юріївні
(прізвище, ім'я, по батькові)

1. Тема роботи _____ Дослідження та прогнозування міграційних тенденцій громадян України методами аналізу даних

затверджена наказом університету від 22 листопада 20 24 р. № 1238Ст

2. Термін подання здобувачем роботи до екзаменаційної комісії 21 січня 20 25 р.

3. Вихідні дані до роботи _____ Дані з офіційного ресурсу Європейської статистичної агенції Eurostat про заявки на отримання тимчасового захисту (датасет migr_asytpfm), Таблиця географічних, соціальних та економічних показників приймаючих країн (destination countries), створена для аналітичних цілей, дані з порталу Statista для заповнення прогалин у статистиці, Інструменти для візуалізації даних: Python (бібліотеки matplotlib, seaborn, plotly), Power BI, Tableau.

4. Перелік питань, що потрібно опрацювати в роботі _____

1) Аналіз предметної галузі

2) Теоретичні дослідження

3) Методологія аналізу даних

4) Аналіз міграційних даних

5) Прогнозування міграційних потоків

РЕФЕРАТ

Пояснювальна записка: 96 с., 54 рис., 14 табл., 1 дод., 20 джерел.

АНАЛІЗ ДАНИХ, ВІЗУАЛІЗАЦІЯ ДАНИХ, КЛАСТЕРІЗАЦІЯ, МІГРАЦІЙНІ ПОТОКИ, ПРОГНОЗУВАННЯ, ПРОГНОСТИЧНІ МОДЕЛІ, УКРАЇНА, ЧАСОВІ РЯДИ.

Об'єкт дослідження – міграційні процеси громадян України в умовах війни та повоєнного відновлення.

Предмет дослідження – закономірності, тенденції та фактори, що впливають на міграційні потоки українців, а також методи їх аналізу та прогнозування.

Мета роботи – дослідження та прогнозування тенденцій міграційних потоків громадян України з використанням сучасних методів аналізу даних для формування рекомендацій щодо управління цими процесами.

Методи дослідження:

- кореляційний аналіз для виявлення взаємозв'язків між параметрами приймаючих країн та інтенсивністю міграційних потоків;
- кластеризація для групування країн за подібними характеристиками впливу на міграцію;
- аналіз часових рядів для дослідження динаміки міграційних потоків і створення прогностичних моделей;
- інструменти візуалізації даних (Python, Power BI) для ідентифікації трендів і представлення результатів.

ABSTRACT

Master's thesis contains: 96 pp., 54 fig., 14 tabl., 1 ann., 20 references.

CLUSTERIZATION, DATA ANALYSIS, DATA VISUALIZATION, FORECASTING, FORECASTING MODELS, MIGRATION FLOWS, TIME SERIES, UKRAINE.

The object of the study is the migration processes of Ukrainian citizens in the conditions of war and post-war reconstruction.

The subject of the study is the patterns, trends and factors affecting the migration flows of Ukrainians, as well as methods of their analysis and forecasting.

The purpose of the work is to study and forecast the trends of migration flows of Ukrainian citizens using modern data analysis methods to formulate recommendations for managing these processes.

Research methods:

- correlation analysis to identify relationships between parameters of host countries and the intensity of migration flows;
- clustering to group countries according to similar characteristics of the impact on migration;
- time series analysis to study the dynamics of migration flows and create predictive models;
- data visualization tools (Python, Power BI) to identify trends and present results.

ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів	8
Вступ.....	9
1 Аналіз предметної галузі.....	10
1.1 Міграційні процеси	10
1.2 Аналіз даних міграційних процесів.....	11
1.2.1 Методи аналізу	11
1.3 Методи прогнозування міграційних потоків.....	12
1.4 Актуальність використання аналізу даних у вивченні міграційних процесів	13
2 Теоретичні дослідження.....	14
2.1 Вплив війни на міграційні процеси в Україні	14
2.1.1 Хронологія подій, що вплинули на міграцію.....	14
2.1.2 Демографічні й економічні наслідки міграції.....	15
2.2 Візуалізація даних у міграційних дослідженнях.....	15
2.2.1 Інструменти для візуалізації (Python, Power BI, Tableau).....	16
2.3 Використання кластеризації для соціальних явищ.....	17
2.4 Метод аналізу часових рядів	18
2.5 Огляд існуючих систем для аналізу міграційних потоків.....	19
2.6 Проблематика та виклики аналізу міграційних процесів під час війни	20
3 Методологія аналізу даних	22
3.1 Вибір джерел даних і параметрів.....	22
3.1.1 Джерела даних	22
3.1.2 Параметри для аналізу	24
3.2 Очищення, обробка та підготовка даних	25
3.3 Вибір методів аналізу (кореляція, кластеризація, часові ряди).....	27
3.3.1 Кореляційний аналіз для оцінки взаємозв'язків	27
3.3.2 Кластеризація для аналізу просторових закономірностей ...	28

3.3.3	Аналіз часових рядів для виявлення динаміки	28
4	Аналіз міграційних даних	29
4.1	Візуалізація даних	29
4.2	Кореляційний аналіз.....	34
4.3	Кластеризація даних.....	37
4.4	Аналіз часових рядів	42
4.4.1	Кластеризація часових рядів.....	49
4.4.1	Мультифакторна кластеризація часових рядів	56
4.5	Вибір прогностичних моделей.....	59
5	Прогнозування міграційних потоків	61
5.1	Модель Sarima.....	61
5.2	Модель ETS.....	61
5.3	Модель Prophet	62
5.4	Тестування та оцінка точності моделей.....	63
5.4.1	Тестування моделі ETS.....	63
5.4.3	Тестування моделі Sarima	70
5.4.1	Тестування моделі Prophet	75
5.5	Візуалізація результатів прогнозування	83
5.5.1	Модель ETS.....	83
5.5.2	Модель Prophet	84
5.5.3	Модель Sarima	87
5.6	Інтерпретація прогнозів	88
	Висновки	91
	Перелік джерел посилання	93
	Додаток А Відомість кваліфікаційної роботи	96

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

ВВП – внутрішній валовий продукт;

ІІР – індекс людського розвитку.

ADF – Augmented Dickey–Fuller test – тест Дікі–Фуллера;

ARIMA – Autoregressive Integrated Moving Average – авторегресійне інтегроване ковзне середнє;

SARIMA – Seasonal Autoregressive Integrated Moving Average – сезонне авторегресійне інтегроване ковзне середнє.

ВСТУП

У сучасному світі аналіз даних є важливим інструментом для вирішення складних завдань у багатьох галузях, зокрема соціальних наук, економіки та державного управління. Використання сучасних методів аналізу дозволяє працювати з великими обсягами інформації, виявляти закономірності й розробляти прогностичні моделі, що є незамінними для вивчення складних соціально-економічних процесів, таких як міграційні потоки.

Особливого значення аналіз міграційних процесів набуває в умовах сучасної України, яка через російсько-українську війну стикається зі значними внутрішніми та зовнішніми переміщеннями населення. Такі процеси суттєво впливають на демографічну, економічну та соціальну структуру як України, так і приймаючих країн. Це зумовлює необхідність створення інструментів для моніторингу та прогнозування міграційних потоків, які можуть слугувати основою для ухвалення ефективних управлінських рішень.

Метою даної кваліфікаційної роботи є дослідження та прогнозування міграційних тенденцій громадян України за допомогою методів аналізу даних, таких як кластеризація, кореляційний аналіз та аналіз часових рядів. У межах роботи проведено аналіз статистичних даних, досліджено прогностичні моделі та оцінено їхню ефективність. Дослідження також спрямоване на вдосконалення практичних навичок у використанні методів Data Science у соціально-економічних дослідженнях.

Результати дослідження спрямовані на вдосконалення практичних навичок у використанні методів Data Science у соціально-економічних дослідженнях і мають потенційне практичне застосування для вдосконалення управління міграційними потоками на державному рівні.

1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ

1.1 Міграційні процеси

Міграційні процеси є результатом складних соціально-економічних, політичних і демографічних чинників, що взаємодіють у глобальному та локальному контекстах [1]. У сучасних умовах війни в Україні ключовими напрямками міграцій є:

- масова внутрішня міграція, що стала одним із головних проявів кризи. Населення активно переміщувалося з регіонів, охоплених бойовими діями, до відносно безпечних областей Західної України. Згідно з офіційними даними, більшість внутрішньо переміщених осіб обирали місця, які мали доступ до базових соціальних послуг і можливостей для працевлаштування [2]. Це створило значний тиск на місцеву інфраструктуру, включаючи школи, лікарні та житлові приміщення;

- зовнішня міграція. Мільйони українців виїхали до країн ЄС (Європейський Союз), зокрема Польщі, Німеччини, Чехії, Угорщини та Румунії. Програми тимчасового захисту, запроваджені ЄС, надали можливість біженцям отримати доступ до ринку праці, освіти та медичних послуг. Польща стала головним напрямком, прийнявши понад 30% усіх біженців з України. Інші країни, такі як Німеччина, також значно збільшили свої ресурси для підтримки українських мігрантів.

Дослідження міграційних процесів є важливим для розуміння їхнього впливу на демографічну структуру, економіку та соціальну інтеграцію як у країнах походження, так і в приймаючих регіонах. Наприклад, дослідження дозволяють оцінити, які чинники впливають на тривалість перебування біженців у певних країнах чи їхню можливу репатріацію. Ці процеси визначають подальші підходи до управління міграцією та адаптації населення.

1.2 Аналіз даних міграційних процесів

Аналіз міграційних процесів є складним багатофакторним завданням, що вимагає системного підходу до роботи з такого типу даними. Вивчення потоків міграції є особливо актуальним у контексті сучасних подій в Україні, які спричинили масове переселення громадян через війну. Масштаби міграції, зокрема її вплив на соціально-економічну структуру, потребують ретельного вивчення за допомогою сучасних методів аналізу.

1.2.1 Методи аналізу

Методи аналізу міграційних даних – це основа для виявлення закономірностей та прогнозування трендів. Вони дозволяють досліджувати взаємозв'язки між ключовими факторами, такими як демографічні, економічні чи географічні показники [3]. Сучасні підходи охоплюють статистичний аналіз (кластеризація, регресія) та інструменти машинного навчання, забезпечуючи багатовимірний погляд на динаміку міграції. Основні методи, які використовуються для вивчення міграційних потоків, включають:

- кластеризація – інструмент, що дає змогу групувати об'єкти за схожими характеристиками;
- регресійний аналіз – метод для оцінки впливу різних факторів, таких як економічний розвиток, географічна близькість і наявність мовного бар'єру, на обсяги міграції;
- аналіз часових рядів – підхід для вивчення змін у часі, що дозволяє прогнозувати динаміку міграційних потоків;
- кореляційний аналіз – метод, що допомагає визначати силу та напрямок зв'язків між змінними.

Поєднання цих методів дозволяє здійснити комплексний підхід до аналізу, забезпечуючи точність і надійність отриманих результатів.

1.3 Методи прогнозування міграційних потоків

Прогнозування міграційних потоків є багатовимірним процесом, що включає аналіз соціально-економічних, демографічних і політичних чинників [4]. Міграція залежить від багатьох факторів, таких як економічна ситуація, політична стабільність, соціальна інфраструктура в країні призначення та фактори «виштовхування» в країні походження. Для її аналізу та прогнозування застосовуються різні підходи, які можна класифікувати за основними методами:

- екстраполяційні моделі базуються на історичних даних, аналізуючи часові ряди для прогнозування майбутніх трендів. Цей підхід підходить для короткострокового прогнозування, коли тренди в міграційних потоках є стабільними. Проте ці моделі чутливі до зовнішніх шоків, таких як військові конфлікти чи економічні кризи, які не враховуються в історичних трендах;

- гравітаційні моделі базуються на ідеї взаємозалежності між розміром економічної «маси» країн (населення, ВВП) та «відстанню» між ними (географічною, культурною, мовною). Гравітаційні моделі забезпечують якісну оцінку потоків між країнами та є ефективними для середньострокового прогнозування. Водночас вони мають обмеження у врахуванні політичних чи соціальних факторів, що динамічно змінюються;

- економетричні моделі зосереджуються на аналізі зв'язків між міграційними потоками та соціально-економічними факторами, такими як рівень безробіття, заробітна плата, рівень ВВП у країнах призначення та походження;

- методи машинного навчання, такі як Random Forest, Gradient Boosting та нейронні мережі, дозволяють обробляти великі обсяги даних та виявляти складні нелінійні залежності між змінними. Основною перевагою є здатність адаптуватися до змін у реальному часі, проте ці моделі вимагають високої обчислювальної потужності та складні у впровадженні;

– сценарний підхід дозволяє моделювати декілька можливих варіантів розвитку подій, враховуючи як позитивні, так і негативні фактори. Такий аналіз корисний для стратегічного планування та довгострокового прогнозування.

1.4 Актуальність використання аналізу даних у вивченні міграційних процесів

Використання сучасних методів аналізу даних дозволяє не лише оцінити масштаби міграції, але й побудувати прогностичні моделі, які допомагають державним органам та міжнародним організаціям розробляти ефективну політику. Такі моделі можуть відіграти ключову роль у досягненні цілей, надаючи інструменти для:

- оцінки масштабів міграції: аналіз даних дозволяє точно оцінити кількість мігрантів, їх розподіл за країнами, регіонами та демографічними характеристиками;
- виявлення закономірностей та трендів: за допомогою аналізу даних можна виявити приховані закономірності та тренди в міграційних потоках;
- побудови прогностичних моделей: аналіз даних дозволяє створювати прогностичні моделі для передбачення майбутніх міграційних тенденцій;
- розробки ефективних стратегій управління міграцією: результати аналізу даних можуть бути використані для розробки ефективних стратегій управління міграційними процесами.

2 ТЕОРЕТИЧНІ ДОСЛІДЖЕННЯ

2.1 Вплив війни на міграційні процеси в Україні

Війна в Україні, яка розпочалася у лютому 2022 року, спричинила значні зміни у міграційних потоках, як внутрішніх, так і зовнішніх. Масштабна релокація населення, як вимушений захід, значною мірою вплинула на демографічну, економічну та соціальну структуру країни [6].

2.1.1 Хронологія подій, що вплинули на міграцію (2022–2024 роки)

Лютий 2022 року: початок повномасштабного вторгнення. Перше значне переміщення населення було зафіксоване впродовж кількох днів після початку військових дій. Мільйони людей залишили свої домівки, шукаючи безпечні місця як в Україні, так і за її межами [5].

Березень 2022 року: ескалація бойових дій. Інтенсивність військових операцій на сході та півдні України призвела до збільшення кількості вимушених мігрантів. Основними пунктами переселення стали західні регіони України, а також сусідні країни Європейського Союзу, зокрема Польща та Румунія.

Середина – кінець 2022 року: стабілізація потоків. Після початкового сплеску кількість переміщених осіб зменшилася, проте залишалася стабільною. Більшість мігрантів шукали довгострокові рішення, включаючи інтеграцію у приймаючих країнах або повернення до відносно безпечних регіонів України.

2023–2024 рік: часткове повернення населення. З початком реконструкції та стабілізації у деяких регіонах України частина мігрантів повертається додому. Водночас залишаються значні потоки людей, які обирають залишитися за кордоном через невизначеність майбутнього.

2.1.2 Демографічні й економічні наслідки міграції

Демографічні наслідки міграції:

- втрата населення у східних та південних регіонах України через інтенсивну міграцію призвела до зменшення щільності населення та дисбалансу в його структурі;
- зовнішня міграція призвела до втрати переважно молодого працездатного населення, що може негативно вплинути на темпи економічного зростання в довгостроковій перспективі;
- водночас у західних регіонах України спостерігається приріст населення, що створює нові демографічні виклики.

Економічні наслідки:

- зменшення робочої сили у східних регіонах призвело до скорочення економічної активності, зокрема у промисловості та сільському господарстві;
- зовнішня міграція сприяла формуванню значних грошових переказів від українців за кордоном, що стало важливим джерелом підтримки економіки країни;
- витрати на соціальну допомогу та інтеграцію переселенців створили додаткове навантаження на бюджети як українських регіонів, так і приймаючих країн.

Загалом, міграційні процеси, викликані війною, мають тривалий вплив на демографічний, економічний та соціальний стан як України, так і країн, які приймають українських мігрантів.

2.2 Візуалізація даних у міграційних дослідженнях

Візуалізація даних є одним із ключових інструментів у дослідженні міграційних процесів. Вона дозволяє перетворювати складні аналітичні результати на зрозумілі для широкого кола користувачів графічні

представлення. Це сприяє ідентифікації трендів, аномалій і ключових факторів, що впливають на міграційні потоки.

Одним із основних завдань візуалізації є представлення міграційних потоків у динаміці. Це досягається за допомогою таких підходів:

- лінійні графіки. Лінійні графіки використовуються для відображення змін кількості мігрантів у часі. Вони допомагають виявити ключові періоди зростання чи спаду міграційної активності, наприклад, різке збільшення кількості біженців у березні 2022 року;

- географічні карти. Карти є незамінним інструментом для аналізу напрямків міграції. Вони дозволяють відобразити просторовий розподіл мігрантів за країнами чи регіонами, а також оцінити вплив географічних факторів. Наприклад, карти можуть показати, як географічна близькість до України впливає на кількість прийнятих біженців у різних країнах;

- теплові діаграми. Теплові діаграми дозволяють оцінити інтенсивність міграційних потоків за різними категоріями, такими як вік, стать або країна походження. Вони є ефективними для аналізу великих обсягів даних;

- кругові та стовпчасті діаграми. Ці типи діаграм використовуються для порівняння часток мігрантів за окремими категоріями. Наприклад, вони можуть показати демографічний розподіл мігрантів за статтю чи віком;

- анімовані візуалізації. Для відображення змін у динаміці за кілька років створюються анімовані графіки чи карти, які ілюструють, як змінюються міграційні потоки з часом. Вони особливо ефективні для демонстрації сезонних змін або наслідків значущих подій;

2.2.1 Інструменти для візуалізації (Python, Power BI, Tableau)

У дослідженні міграційних процесів використовуються різноманітні інструменти для створення графіків і діаграм. Найпопулярнішими є:

- Python (бібліотеки `matplotlib`, `seaborn`, `plotly`). Python надає широкий спектр інструментів для створення кастомізованих графіків. Бібліотека `matplotlib` використовується для базових візуалізацій, `seaborn` дозволяє створювати статистичні графіки, а `plotly` – інтерактивні візуалізації;
- Tableau та Power BI. Ці інструменти дозволяють швидко створювати динамічні та інтерактивні дашборди для аналізу даних. Вони є особливо корисними для представлення результатів у реальному часі;
- GIS-системи. Інструменти, такі як QGIS чи ArcGIS, використовуються для геопросторового аналізу та створення карт. Вони дозволяють інтегрувати дані про міграцію з географічними координатами для оцінки впливу просторових факторів;
- Excel. Попри обмежену функціональність у порівнянні з іншими інструментами, Excel залишається популярним завдяки простоті використання. Він дозволяє створювати базові графіки для початкового аналізу даних.

2.3 Використання кластеризації для соціальних явищ

Кластеризація – це один із найпоширеніших методів аналізу даних, що дозволяє об'єднувати об'єкти у кластери на основі подібності їхніх характеристик [7]. Основна мета кластеризації – знайти приховані закономірності в даних, виділивши групи, які мають схожі властивості, але відрізняються від інших груп [8]. У соціальних дослідженнях цей метод використовується для:

- групування країн. На основі показників, таких як економічний розвиток, рівень інтеграції мігрантів чи демографічні особливості, країни можуть бути поділені на кластери для аналізу їхньої ролі у міграційних процесах;

- аналізу мігрантів. Кластеризація допомагає ідентифікувати групи мігрантів із подібними потребами чи характеристиками, наприклад за рівнем освіти, віком або соціальним статусом;

- оптимізації ресурсів. Використання кластеризації дозволяє приймаючим країнам ефективніше розподіляти ресурси, спрямовуючи їх на регіони чи групи з подібними викликами.

Методи кластеризації, такі як K-Means, DBSCAN або ієрархічна кластеризація, забезпечують широкий спектр можливостей для дослідження складних соціальних явищ [9].

2.4 Метод аналізу часових рядів

Аналіз часових рядів є одним із ключових підходів до вивчення динаміки міграційних потоків. Він дозволяє розглядати міграційні процеси в динаміці, виявляти закономірності та тенденції змін у часі, а також прогнозувати майбутні значення [10]. Основними завданнями цього методу є:

- виявлення трендів. Часові ряди дозволяють визначити довгострокові тенденції в міграційних потоках [11], які можуть бути пов'язані зі змінами економічних, політичних чи соціальних умов. Наприклад, можна виявити зростання або спад міграційної активності протягом певного періоду. Для цього використовуються методи згладжування та апроксимації часових рядів, такі як ковзне середнє, експоненційне згладжування та метод Хольта-Вінтерса [12];

- сезонні коливання. Аналіз часових рядів допомагає виявити періодичні зміни в міграційних потоках, які можуть бути обумовлені сезонними факторами, такими як кліматичні умови, свята або періоди відпусток. Для аналізу сезонності використовуються методи спектрального аналізу та автокореляції;

- прогнозування. На основі часових рядів створюються моделі для короткострокових і середньострокових прогнозів міграційних потоків [12]. Це дозволяє приймаючим країнам планувати ресурси та розробляти ефективні стратегії управління міграцією. Для прогнозування використовуються різні моделі, такі як ARIMA, SARIMA, Prophet та моделі машинного навчання, такі як рекурентні нейронні мережі (RNN);
- виявлення структурних змін. Аналіз часових рядів дозволяє виявити точки структурних змін, тобто моменти, коли відбуваються різкі зміни в динаміці міграційних потоків. Це може бути пов'язано з певними подіями, такими як військові конфлікти, економічні кризи або зміни в міграційному законодавстві. Для виявлення структурних змін використовуються такі методи, як CUSUM та Chow test;
- аналіз аномалій. За допомогою часових рядів можна виявити аномальні значення, які значно відрізняються від загальної тенденції. Це можуть бути викиди, спричинені помилками в даних або незвичайними подіями. Для аналізу аномалій використовуються різні методи, такі як Isolation Forest, One-Class SVM та DBSCAN.

2.5 Огляд існуючих систем для аналізу міграційних потоків

Існують кілька систем і платформ, які використовуються для аналізу та прогнозування міграційних потоків:

- UNHCR Population Statistics. Ця система забезпечує доступ до даних про міграційні потоки, біженців та переміщених осіб. Вона використовується для оцінки глобальних трендів і розробки політик;
- Eurostat. Платформа надає детальні дані про міграційні процеси в Європі, включаючи інформацію про заяви на тимчасовий захист, демографічні показники та економічну інтеграцію мігрантів;

- IOM Migration Data Portal. Ця система містить дані про міжнародну міграцію, включаючи аналітичні звіти та інтерактивні інструменти для візуалізації;

- Google Mobility Data. Ця платформа дозволяє аналізувати мобільність населення, що може бути корисним для оцінки міграційних потоків у кризових ситуаціях.

Огляд існуючих систем показує, що інтеграція даних із різних джерел є ключовою для створення точних моделей прогнозування міграції. Системи, які поєднують статистичні дані, аналітичні інструменти та інтерактивні візуалізації, забезпечують ефективніший аналіз міграційних процесів.

2.6 Проблематика та виклики аналізу міграційних процесів під час війни

Аналіз міграційних процесів у періоди військових конфліктів стикається з низкою значних проблем і викликів, які обмежують точність досліджень і ускладнюють формулювання обґрунтованих висновків. Основні проблеми включають:

- недостатність даних. Військові конфлікти створюють хаотичні умови, за яких збір даних є ускладненим або неможливим. Часові затримки в оновленні статистики, розрізненість джерел інформації та відсутність стандартизації ускладнюють створення повної картини міграційних потоків. Крім того, деякі держави не надають повних даних або надають їх із запізненням, що впливає на репрезентативність результатів;

- неоднорідність міграційних потоків. У кризових умовах мігранти належать до різних соціально-економічних груп, відрізняються за віком, статтю та рівнем освіти. Така неоднорідність ускладнює аналіз, оскільки різні групи можуть мати специфічні потреби та моделі поведінки, які потребують індивідуального вивчення;

- відсутність достовірної інформації про нелегальну міграцію. Під час війни значна частина міграційних потоків може бути нелегальною або неофіційно зафіксованою, що виключає її з офіційних статистичних даних. Це значно зменшує точність оцінок обсягів і напрямків переміщення населення;
- географічна складність аналізу. Зміни в міграційних потоках під час війни відбуваються в умовах змінних геополітичних обставин, таких як окупація територій, порушення транспортної інфраструктури або закриття кордонів. Це ускладнює створення просторових моделей міграції;
- динамічність ситуації. Міграційні потоки можуть змінюватися щодня залежно від інтенсивності бойових дій, введення нових політичних рішень чи змін у безпековій ситуації. Це створює труднощі у прогнозуванні та побудові моделей для тривалого аналізу;
- психологічні та соціальні фактори. Під час війни багато рішень про міграцію приймаються під впливом стресу, страху чи невизначеності. Традиційні моделі, які базуються на раціональних передумовах вибору, можуть бути недостатньо ефективними для аналізу таких умов;
- перевантаження соціальних систем у приймаючих регіонах. Значний приплив біженців до певних країн чи регіонів створює додатковий тиск на місцеву інфраструктуру, що може спотворювати дані через нерівномірний розподіл ресурсів або їхню нестачу;
- етичні аспекти збору даних. Під час конфліктів важливо дотримуватися конфіденційності та етичних норм у зборі інформації про мігрантів. Недотримання цих принципів може погіршити ситуацію або створити ризики для безпеки переміщених осіб.

3 МЕТОДОЛОГІЯ АНАЛІЗУ ДАНИХ

3.1 Вибір джерел даних і параметрів

3.1.1 Джерела даних

Для виконання дослідження використовуються дані з надійних і офіційних джерел, які відповідають темі дослідження та забезпечують необхідну повноту й релевантність.

Основне джерело даних: датасет `migr_asytpfm` (рисунок 3.1). Набір даних з офіційного ресурсу Європейської статистичної агенції [13], який містить інформацію про кількість заявок на отримання тимчасового захисту в країнах Європи.

DATAFLOW	LAST UPDATE	freq	unit	citizen	sex	age	geo	TIME_PERIOD	OBS_VALUE	OBS_FLAG
ESTAT:MIGR_ASYTPFM(1.0)	10.12.2024 23:00:00	Monthly	Person	Andorra	Females	Total	Austria	01.03.2022	0	
ESTAT:MIGR_ASYTPFM(1.0)	10.12.2024 23:00:00	Monthly	Person	Andorra	Females	Total	Austria	01.04.2022	0	
ESTAT:MIGR_ASYTPFM(1.0)	10.12.2024 23:00:00	Monthly	Person	Andorra	Females	Total	Austria	01.05.2022	0	
ESTAT:MIGR_ASYTPFM(1.0)	10.12.2024 23:00:00	Monthly	Person	Andorra	Females	Total	Austria	01.06.2022	0	
ESTAT:MIGR_ASYTPFM(1.0)	10.12.2024 23:00:00	Monthly	Person	Andorra	Females	Total	Austria	01.07.2022	0	
ESTAT:MIGR_ASYTPFM(1.0)	10.12.2024 23:00:00	Monthly	Person	Andorra	Females	Total	Austria	01.08.2022	0	
ESTAT:MIGR_ASYTPFM(1.0)	10.12.2024 23:00:00	Monthly	Person	Andorra	Females	Total	Austria	01.09.2022	0	
ESTAT:MIGR_ASYTPFM(1.0)	10.12.2024 23:00:00	Monthly	Person	Andorra	Females	Total	Austria	01.10.2022	0	
ESTAT:MIGR_ASYTPFM(1.0)	10.12.2024 23:00:00	Monthly	Person	Andorra	Females	Total	Austria	01.11.2022	0	
ESTAT:MIGR_ASYTPFM(1.0)	10.12.2024 23:00:00	Monthly	Person	Andorra	Females	Total	Austria	01.12.2022	0	
ESTAT:MIGR_ASYTPFM(1.0)	10.12.2024 23:00:00	Monthly	Person	Andorra	Females	Total	Austria	01.01.2023	0	
ESTAT:MIGR_ASYTPFM(1.0)	10.12.2024 23:00:00	Monthly	Person	Andorra	Females	Total	Austria	01.02.2023	0	
ESTAT:MIGR_ASYTPFM(1.0)	10.12.2024 23:00:00	Monthly	Person	Andorra	Females	Total	Austria	01.03.2023	0	
ESTAT:MIGR_ASYTPFM(1.0)	10.12.2024 23:00:00	Monthly	Person	Andorra	Females	Total	Austria	01.04.2023	0	
ESTAT:MIGR_ASYTPFM(1.0)	10.12.2024 23:00:00	Monthly	Person	Andorra	Females	Total	Austria	01.05.2023	0	
ESTAT:MIGR_ASYTPFM(1.0)	10.12.2024 23:00:00	Monthly	Person	Andorra	Females	Total	Austria	01.06.2023	0	
ESTAT:MIGR_ASYTPFM(1.0)	10.12.2024 23:00:00	Monthly	Person	Andorra	Females	Total	Austria	01.07.2023	0	
ESTAT:MIGR_ASYTPFM(1.0)	10.12.2024 23:00:00	Monthly	Person	Andorra	Females	Total	Austria	01.08.2023	0	
ESTAT:MIGR_ASYTPFM(1.0)	10.12.2024 23:00:00	Monthly	Person	Andorra	Females	Total	Austria	01.09.2023	0	

Рисунок 3.1 – Прев'ю датасету `migr_asytpfm`

Дані охоплюють період із березня 2022 року по листопад 2024 року включно та містять показники: країну подання заявки, часові інтервали, чисельність заявників, гендер, вікову категорію тощо. Початковий набір

даних охоплює понад 6.5 мільйонів записів, що містять інформацію про мігрантів з різних країн від початку введення терміну «тимчасовий захист» (temporary protection).

Також для інтеграції географічних, соціальних, економічних та культурних показників приймаючих країн, що мають безпосередній вплив на міграційні потоки була створена таблиця (рисунок 3.2) destination_countries. Її мета – забезпечити аналітичну основу для дослідження, яка враховує ключові фактори, що впливають на вибір країни мігрантами.

Geo	Distance_to_Ukr	Maritime_Comp	Borders_Ukr	Slavic	GDP	HDI_2022	Monthly_Paym	Cost_of_Living
Austria	1336	0	0	0	673525	0.926	260	1070
Belgium	2043	0	0	0	832972	0.942	1100	940
Bulgaria	1260	0	0	1	248788	0.799	255	600
Switzerland	2080	0	0	0	822862	0.967	450	1550
Cyprus	1710	1	0	0	53033	0.907	0	900
Czechia	1120	0	0	1	585187	0.895	205	760
Germany	1360	0	0	0	5857856	0.950	449	1000
Denmark	1930	0	0	0	456059	0.952	323	1080
Estonia	1680	0	0	1	66932	0.899	200	880
Greece	1620	0	0	0	426748	0.893	0	800
Spain	3010	0	0	0	2553107	0.911	400	715
Finland	1830	1	0	0	363315	0.942	555	950
France	2000	0	0	0	4169071	0.91	400	950
Croatia	1390	0	0	1	176899	0.878	250	750
Hungary	955	0	1	0	440579	0.851	56	630
Ireland	3030	1	0	0	671603	0.95	412	1000
Iceland	4500	1	0	0	30530	0.959	400	1250
Italy	1930	0	0	0	3452505	0.906	0	870

Рисунок 3.2 – Прев'ю датасету destination_countries

Атрибути таблиці:

- Geo: назва країни;
- Distance_to_Ukraine: відстань найближчого великого міста до столиці України (у км);
- Maritime_Component: наявність морського шляху (булеве значення, 1/0);
- Borders_Ukraine: наявність спільного кордону з Україною (булеве значення, 1/0);

- Slavic: більша частина населення знає одну з слов'янських мов (булеве значення, 1/0);
- GDP: валовий внутрішній продукт (ВВП) країни (в мільйонах);
- HDI_2022: індекс людського розвитку (ІЛР);
- Monthly_Payment: середній місячний платіж біженцям (у євро);
- Cost_of_Living: середня вартість проживання (у євро).

3.1.2 Параметри для аналізу

Таблиця `destination_countries` сформована таким чином, щоб забезпечити всебічний аналіз ключових факторів, які впливають на вибір приймаючих країн мігрантами. До неї включено атрибути, що охоплюють географічні, соціальні, економічні та культурні аспекти, оскільки вони є критично важливими для розуміння динаміки міграційних потоків. Наприклад, відстань до України та наявність спільного кордону (`Distance_to_Ukraine` і `Borders_Ukraine`) були обрані через їхню вагомую роль у доступності країн для біженців, особливо в кризових умовах. Наявність морського компонента (`Maritime_Component`) також враховано, адже морські маршрути можуть стати визначальними у випадку закриття сухопутних шляхів. Економічні показники, такі як ВВП (GDP), ІЛР (HDI_2022), середні виплати біженцям (Monthly_Payment) і вартість проживання (Cost_of_Living), дозволяють оцінити здатність країн підтримувати мігрантів та рівень комфорту їхнього життя. Включення таких соціально-культурних атрибутів, як приналежність до слов'янської мовної групи (Slavic), було зумовлено впливом мовної та культурної близькості на інтеграцію біженців. Таким чином, кожен з цих атрибутів є важливим для багатовимірного аналізу міграційних процесів, оскільки вони спільно враховують доступність, економічну привабливість та можливості соціальної адаптації у приймаючих країнах.

3.2 Очищення, обробка та підготовка даних

На етапі очищення даних було виконано низку кроків, спрямованих на забезпечення точності аналізу та уникнення дублювання даних. У межах цього дослідження основна увага була зосереджена на громадянах України, тому вибірка була обмежена лише записами, де значення атрибута `citizenship==Ukraine`. Це дозволило зосередитися виключно на релевантній частині даних, яка складає близько 30 тисяч записів. Далі були видалені нерелевантні атрибути, які не несли аналітичної цінності, наприклад, технічні показники, що дублювали вже наявну інформацію або не використовувалися в аналізі. Зокрема:

- `DATAFLOW` – технічна інформація про джерело даних (внутрішній код Eurostat);

- `LAST_UPDATE` – дата останнього оновлення даних на сервері. Оскільки у дослідженні використовуються вже завантажені актуальні дані, цей стовпець не несе аналітичної цінності;

- `freq` – періодичність даних (місячна чи квартална). Дослідження вже оперує параметром `TIME_PERIOD`, тому цей стовпець є зайвим;

- `unit` – одиниця виміру, що завжди має значення `Person` у цьому датасеті і не містить варіативності;

- `OBS_FLAG` – допоміжний стовпець для позначення спеціальних випадків або винятків у даних. У наявному датасеті він є порожнім і не використовується у розрахунках;

Особливу увагу приділили атрибутам `sex` та `age`, де були виключені агреговані значення `total`, оскільки вони перекривалися з деталізованими даними й створювали зайві дублювання. З атрибута `age` також були видалені значення «less than 14» і «from 14 to 17», які дублювали групу «less than 18». Було вирішено залишити саме «less than 18», адже це значення більш чітко представляє дітей до повноліття, що відповідає потребам аналізу.

Крім того, для уникнення дублювання заявок на рівні регіонів було виключено агреговане значення «European Union – 27 countries (from 2020)» з атрибута geo. Це забезпечило коректність даних у просторовому вимірі, оскільки цей показник є сумарним для окремих країн і міг би спотворити результати. Усі ці кроки були спрямовані на те, щоб зробити дані максимально чистими, структурованими та готовими для проведення подальших аналітичних процедур, таких як кластеризація, аналіз часових рядів і прогнозування.

Після очищення у датасеті залишено наступні колонки:

- geo – країна, що приймає тимчасових мігрантів;
- sex – категорія статі мігрантів (Males, Females, Unknown);
- age – вікові групи мігрантів (0–18, 18–34, 35–64, 65+);
- TIME_PERIOD – часовий вимір подання заявки;
- OBS_VALUE – кількість осіб, яким надано тимчасовий захист.

Під час аналізу основного набору даних було виявлено прогалини у статистиці аплікантів на тимчасовий захист у Німеччині за період з 1 березня по 1 липня 2022 року включно. Оскільки Німеччина є однією з ключових країн призначення для українських біженців, відсутність цих даних могла суттєво вплинути на точність аналізу та прогнозів.

Для відновлення даних було використано статистику, надану порталом Statista [14], що містить інформацію про кількість біженців з України у Німеччині за місяцями у 2022–2024 роках.

Ці значення інтегрувалися в основний датасет без додаткової модифікації, щоб забезпечити максимально наближене відтворення реальної динаміки міграційних потоків.

Додані дані використовувалися у всіх подальших етапах аналізу, включаючи візуалізацію і побудову прогностичних моделей.

Використані дані з порталу Statista могли містити певну похибку через відмінності у методології збору статистики, однак вони є найбільш репрезентативними для зазначеного періоду.

Заповнення прогалин було необхідним для забезпечення повноти та коректності дослідження. Зазначений період охоплює пікові значення міграційних потоків, спричинені ескалацією війни в Україні, що робить ці дані ключовими для вивчення тенденцій.

Результати з доданими даними дозволяють отримати більш точну картину змін у міграційних потоках і підвищують надійність побудованих моделей прогнозування.

3.3 Вибір методів аналізу (кореляція, кластеризація, часові ряди)

У дослідженні для аналізу міграційних процесів було обрано методи, які враховують просторові, часові та демографічні аспекти. Основна увага приділяється кореляційному аналізу, кластеризації та аналізу часових рядів. Ці підходи дозволяють глибше зрозуміти динаміку міграційних потоків, оцінити вплив характеристик приймаючих країн та розробити прогностичні моделі.

3.3.1 Кореляційний аналіз для оцінки взаємозв'язків

Кореляційний аналіз використовується для виявлення взаємозв'язків між різними характеристиками приймаючих країн та інтенсивністю міграційних потоків. За допомогою цього методу визначаються, наприклад, сила і напрямок зв'язку між кількістю мігрантів та економічними показниками країни, такими як ВВП чи ІЛР. Результати кореляційного аналізу також демонструють роль культурної спорідненості (наприклад, приналежність до слов'янської мовної групи) та географічної близькості в процесі вибору мігрантами країни для переселення. Цей підхід дає змогу ідентифікувати ключові чинники, які впливають на міграційні потоки, і закладає основу для подальшого прогнозування та розробки рекомендацій.

3.3.2 Кластеризація для аналізу просторових закономірностей

Метод кластеризації дозволяє групувати країни за схожими характеристиками, такими як географічна близькість до України, наявність спільного кордону, культурна спорідненість чи рівень економічного розвитку. Це дає змогу визначити типові групи країн із подібними умовами, що впливають на вибір мігрантів. Наприклад, країни зі спільним кордоном або високим індексом людського розвитку часто утворюють окремі кластери, що допомагає виявити закономірності у міграційних потоках. Такий підхід дозволяє формулювати рекомендації для ефективнішого розподілу ресурсів та розробки міграційної політики.

3.3.3 Аналіз часових рядів для виявлення динаміки

Аналіз часових рядів спрямований на дослідження змін міграційних потоків у часі. Цей метод дає змогу виявити ключові тренди, сезонні коливання та реакції на значущі соціально-економічні чи політичні події. Наприклад, різкий сплеск міграції після початку війни в Україні може бути відображений як аномалія в часовому ряді. Прогнозування на основі часових рядів дозволяє оцінити динаміку міграційних потоків у майбутньому, що є критично важливим для планування інтеграційних заходів у приймаючих країнах. У цьому дослідженні часові ряди побудовані на основі квартальних даних, що забезпечує достатню деталізацію для короткострокового прогнозування.

4 АНАЛІЗ МІГРАЦІЙНИХ ДАНИХ

4.1 Візуалізація даних

Під час аналізу міграційних потоків було виявлено кілька ключових закономірностей, які стали основою для створення серії візуалізацій.

Для створення візуалізацій було використано Power BI, що дозволило побудувати інтерактивні графіки, карти та діаграми. Завдяки цьому інструменту вдалося не лише представити результати аналізу у зрозумілому вигляді, але й забезпечити можливість динамічного дослідження даних, таких як порівняння країн, часових періодів чи демографічних груп.

Графік (рисунок 4.1) відображає вартість проживання у різних країнах, відсортовану за спаданням. Найвищі значення демонструють країни з найдорожчим рівнем життя, тоді як нижчі свідчать про більш доступні умови. Така візуалізація дозволяє порівняти країни за цим показником і виділити ті, де умови проживання є найдорожчими чи найдешевшими.

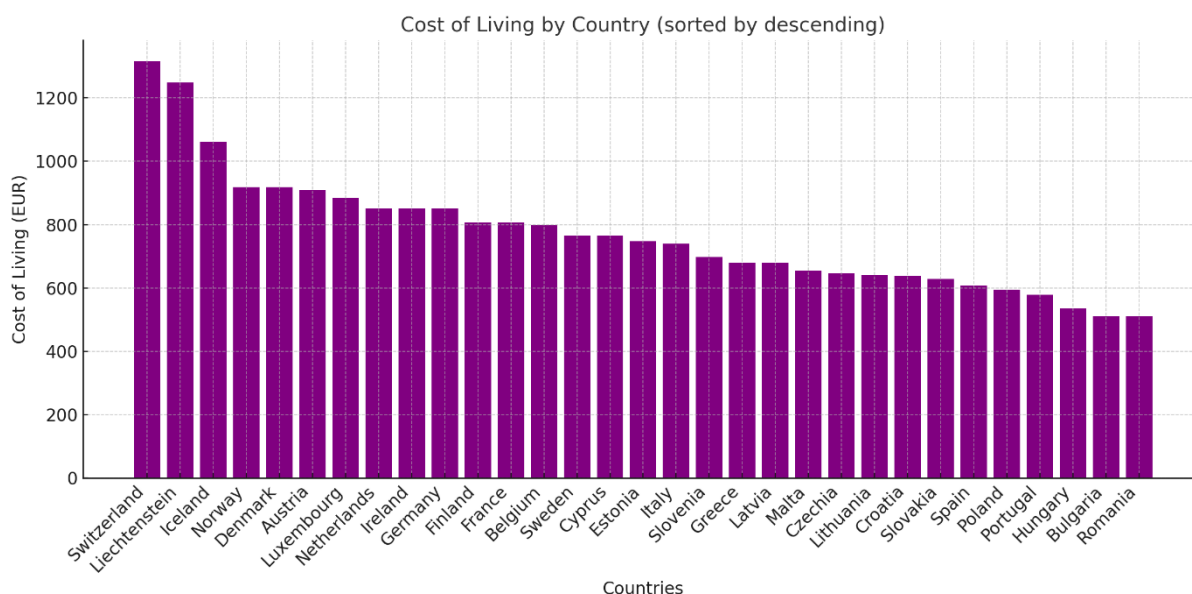


Рисунок 4.1 – Графік вартості проживання за спаданням

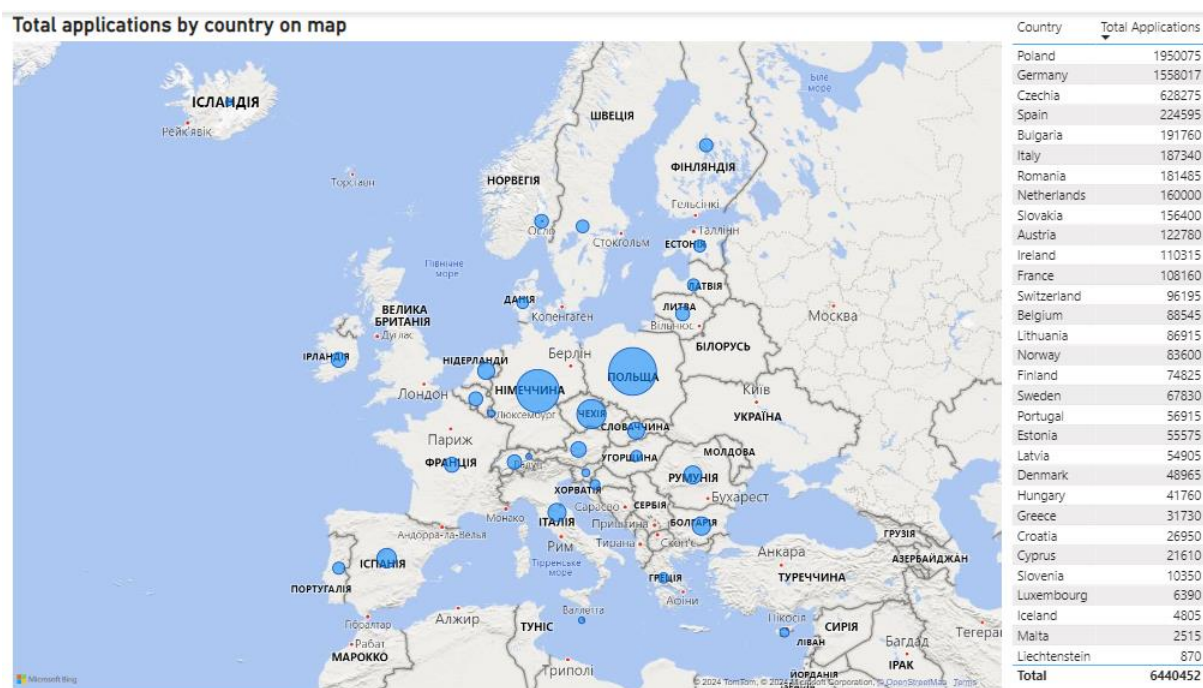


Рисунок 4.3 – Географічна карта та список суми заявок у країнах

Демографічний аналіз показав, що структура заявників на тимчасовий захист складається з кількох важливих груп. Першим кроком було дослідження загального розподілу заявок за статтю, де більшість заявників складають жінки (53,64%), чоловіки становлять 31,38%, а невизначені заявки (unknown) – 14,98%. Ця тенденція відображає соціальні наслідки війни, зокрема мобілізацію чоловіків в Україні, яка суттєво обмежила їхній виїзд за кордон.

Наступним етапом став аналіз розподілу заявок за віковими групами. Найбільшу частину складають особи віком від 35 до 64 років (29,91%), за ними йдуть діти до 18 років (28,78%). Ці показники підкреслюють характер війни, коли основними заявниками є жінки з дітьми, а також працездатні особи, які шукають можливості забезпечити сім'ї у нових умовах.

Для поглибленого аналізу було проведено розподіл кожної вікової групи за статтю, що дозволило краще зрозуміти структуру заявників. Найбільше жінок зафіксовано у віковій категорії 35–64 років, що відповідає припущенню, що переважно мігрують матері, які супроводжують дітей. У

молодших вікових категоріях (до 18 років) домінують жінки, що також свідчить про виїзд переважно материнських родин. У категорії 65 років і старше заявники розподілені більш рівномірно між статями, що пояснюється меншою залежністю від мобілізації або інших обмежень.

Для відображення цих результатів було створено кілька візуалізацій (рисунок 4.4). Перша стовпчаста діаграма демонструє загальний розподіл заявок за статтю. Друга візуалізація ілюструє поділ заявок за віковими групами. Третя діаграма комбінує обидва аспекти, показуючи розподіл статей у кожній віковій категорії. Ці візуалізації забезпечують глибше розуміння демографічної структури заявників і дозволяють детально дослідити особливості кожної групи.

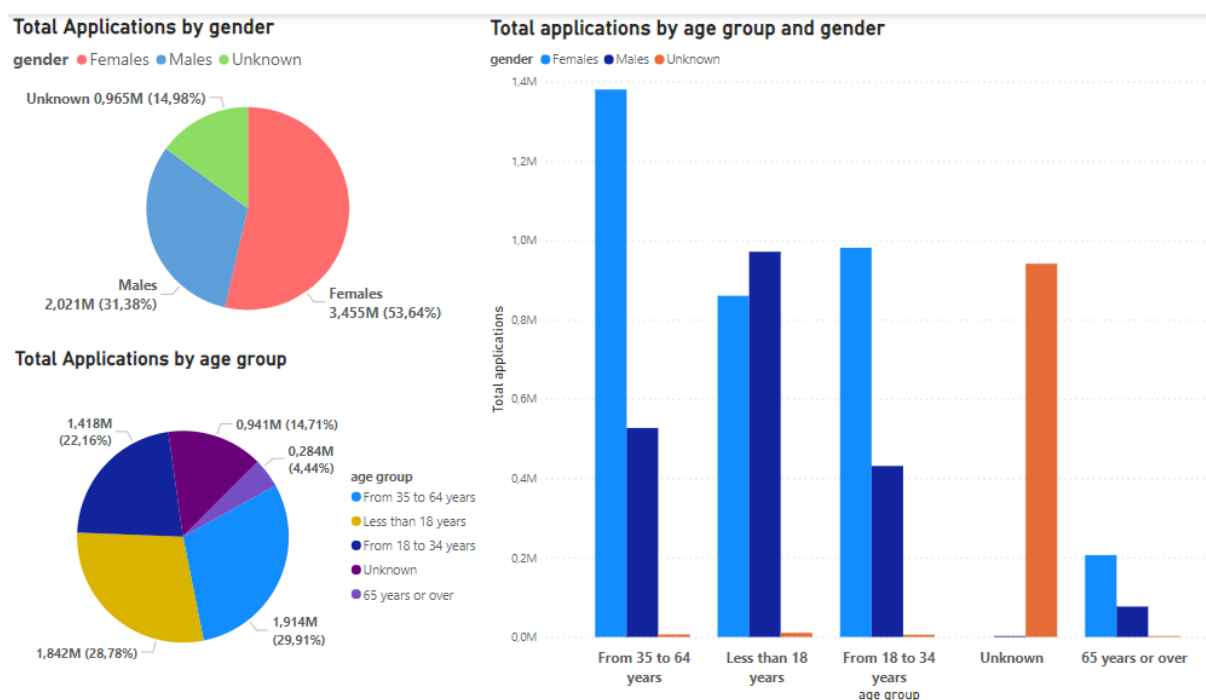


Рисунок 4.4 – Дашборд розподіл за гендером, віковою групою та перетин категорій

Аналіз розподілу заявок за мовами приймаючих країн показав, що польська та німецька мови домінують серед країн, які прийняли найбільшу кількість біженців. Це корелює з даними опитування, проведеного у 2023

році [15], яке свідчить, що 22% українців володіють польською мовою, а 14% – німецькою.

Однак англійська, яка є найбільш популярною іноземною мовою серед українців (51% мають певні знання), не відіграє ключової ролі у виборі країн-приймачів. Це зумовлено географічним розташуванням: поруч із Україною немає англійськомовних країн, тому вибір мігрантів обмежується країнами, де розмовляють мовами, близькими за культурою чи знаннями.

Дані опитування підкреслюють, що знання польської та німецької мов сприяло масовому переміщенню до Польщі (30,29% заявок) і Німеччини (27,08%). Для ілюстрації цієї тенденції було створено кругову діаграму (рисунок 4.5), яка демонструє домінування заявок у країнах із польською та німецькою мовами. Це підтверджує, що мовна компетентність і географічна близькість є ключовими факторами при виборі країни для міграції.

Distribution of Temporary Protection Applications by Official Languages in Host Countries

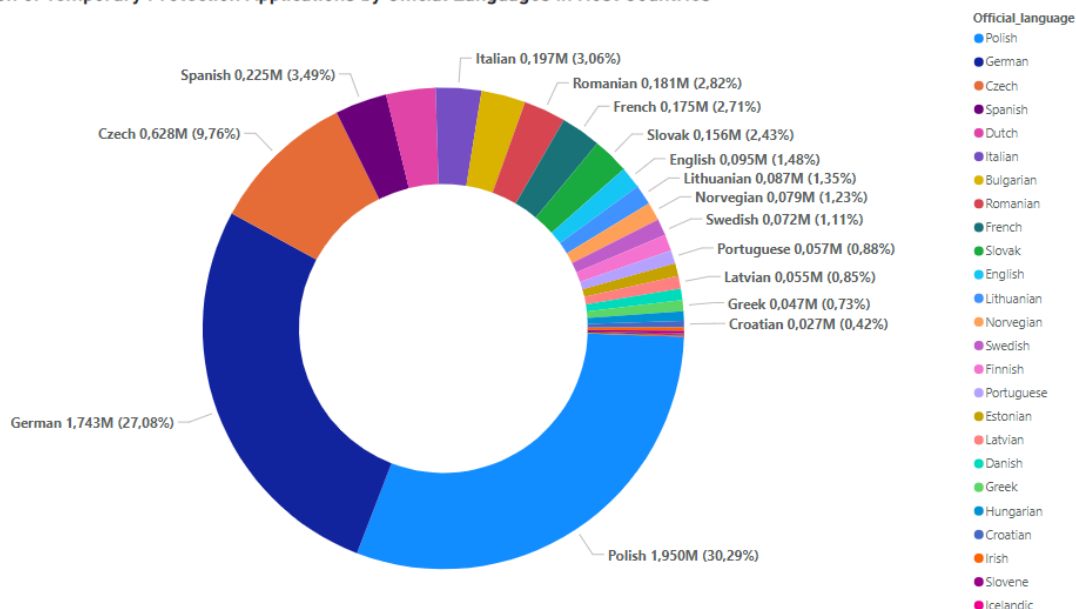


Рисунок 4.5 – Кругова діаграма розподілу заявок за мовами

4.2 Кореляційний аналіз

Для проведення аналізу використовувалися дані про кількість заявок на тимчасовий статус українських біженців, ВВП, ІРЛ, відстань до України, вартість проживання, культурну схожість через належність до слов'янської мовної групи та кількість заявок за країнами (OBS_VALUE). Було застосовано коефіцієнт кореляції Пірсона, який дозволяє оцінити силу та напрямок лінійного зв'язку між змінними. Попередньо всі дані були стандартизовані для забезпечення коректності розрахунків.

На візуалізації кореляційної матриці (рисунок 4.6) чітко видно взаємозв'язки між змінними.

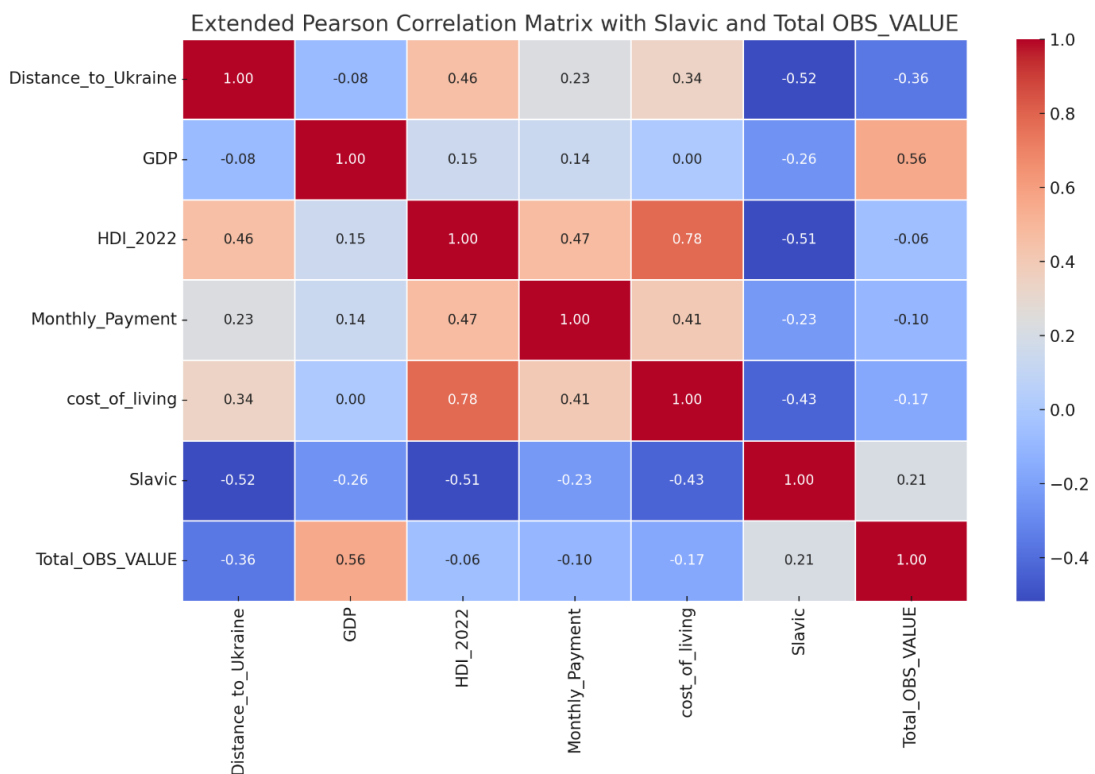


Рисунок 4.6 – Кореляційна матриця Пірсона

Результати кореляційного аналізу продемонстрували значний вплив економічних показників на кількість заявок. Зокрема:

- кількість заявок (OBS_VALUE) має позитивну кореляцію з ВВП, що свідчить про те, що економічно розвинені країни мають більше ресурсів для прийому біженців;
- ІЛР також демонструє помірний позитивний зв'язок із кількістю заявок, що підкреслює важливість якості життя в країні;
- вартість проживання (cost of living) має слабкий зворотний зв'язок із кількістю заявок;
- приналежність до слов'янської мовної групи (Slavic) показує слабку позитивну кореляцію;
- відстань до України має слабкий негативний зв'язок із кількістю заявок, підтверджуючи, що мігранти надають перевагу країнам, розташованим ближче до їхньої батьківщини.

Також було досліджено кореляційний аналіз за методом Кендалла (рисунок 4.7), який дозволяє оцінити монотонні зв'язки між змінними.

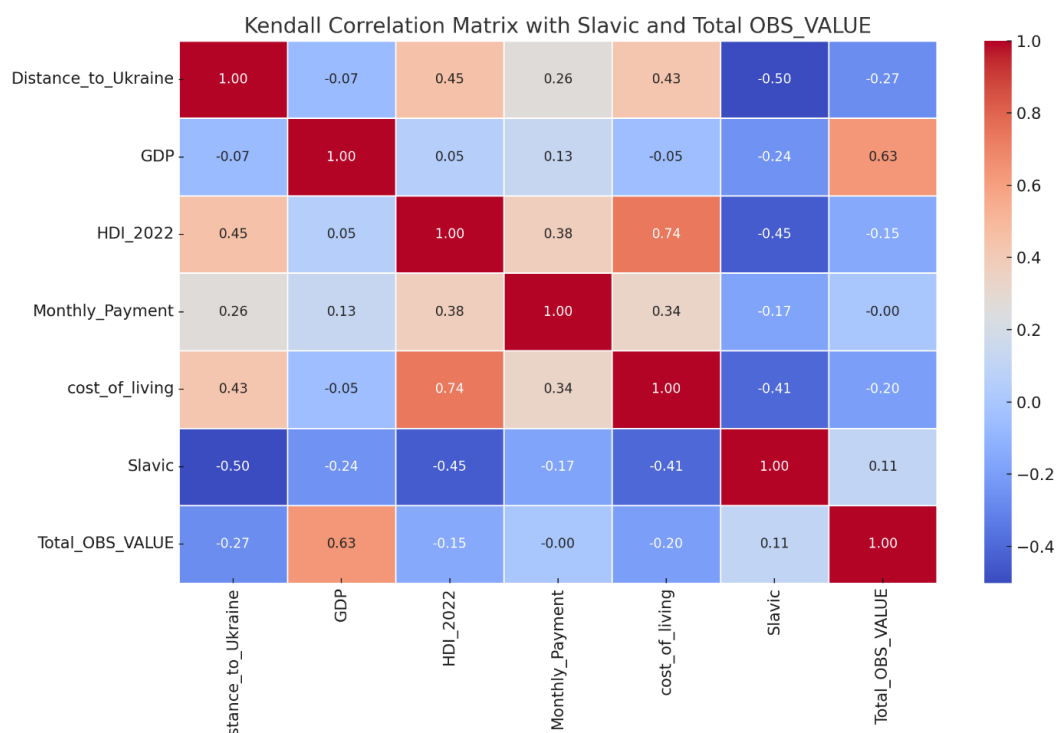


Рисунок 4.7 – Кореляційна матриця Кендалла

Результати показують, що між кількістю заявок на тимчасовий захист та ВВП країн існує помірний позитивний зв'язок, що підтверджує гіпотезу про перевагу економічно розвинених країн серед мігрантів. Залежність між якістю життя, вираженою через ІРЛ, та кількістю заявок також є позитивною, але менш вираженою. Це свідчить про те, що хоча рівень розвитку країни має вплив, він не є домінуючим фактором для прийняття рішення мігрантами.

Культурна близькість, визначена через належність до слов'янської мовної групи, показує позитивний, хоча і слабкий, зв'язок із кількістю заявок. Це свідчить про певну роль культурних і мовних чинників у виборі країни. Водночас вартість проживання демонструє слабку негативну кореляцію, що може вказувати на те, що високі витрати можуть стримувати мігрантів від переїзду до певних країн.

Географічні фактори, такі як відстань до України, демонструють негативний зв'язок із кількістю заявок, що є очікуваним результатом, оскільки ближчі країни частіше обираються мігрантами. Проте цей зв'язок не є сильним, що вказує на те, що географічна віддаленість є лише одним із багатьох факторів, які впливають на вибір країни.

У цілому, кореляційний аналіз за методом Кендалла підкреслює складність взаємозв'язків між економічними, соціальними, культурними та географічними факторами. Він показує, що вибір країни для міграції є результатом впливу багатьох чинників, серед яких ключовими залишаються економічні умови та географічна близькість, тоді як культурна схожість і вартість проживання відіграють допоміжну роль.

Результати кореляційного аналізу за методами Пірсона і Кендалла демонструють схожі тенденції, але з певними відмінностями. Метод Пірсона виявляє сильні лінійні зв'язки, зокрема між кількістю заявок, ВВП і ІЛР, підтверджуючи, що економічно розвинені країни приваблюють більше мігрантів. Водночас метод Кендалла краще виявляє монотонні

залежності, такі як негативний вплив відстані до України та слабкий позитивний вплив культурної близькості.

Обидва методи узгоджуються в тому, що економічні та географічні чинники є визначальними, проте Кендалл краще враховує нелінійності та стійкіший до викидів.

4.3 Кластеризація даних

У ході дослідження міграційних потоків було застосовано метод кластеризації для аналізу груп країн, які приймали українських біженців. Цей підхід дозволив виявити подібності між країнами за такими показниками, як відстань до України, ВВП (GDP), ІЛР (HDI), кількість заявок на тимчасовий статус (OBS_VALUE), а також наявність слов'янської мови в офіційних або поширених мовах.

Для кластеризації було використано алгоритм K-means із поділом на п'ять кластерів. Для визначення оптимальної кількості кластерів було використано метод ліктя (рисунок 4.8).

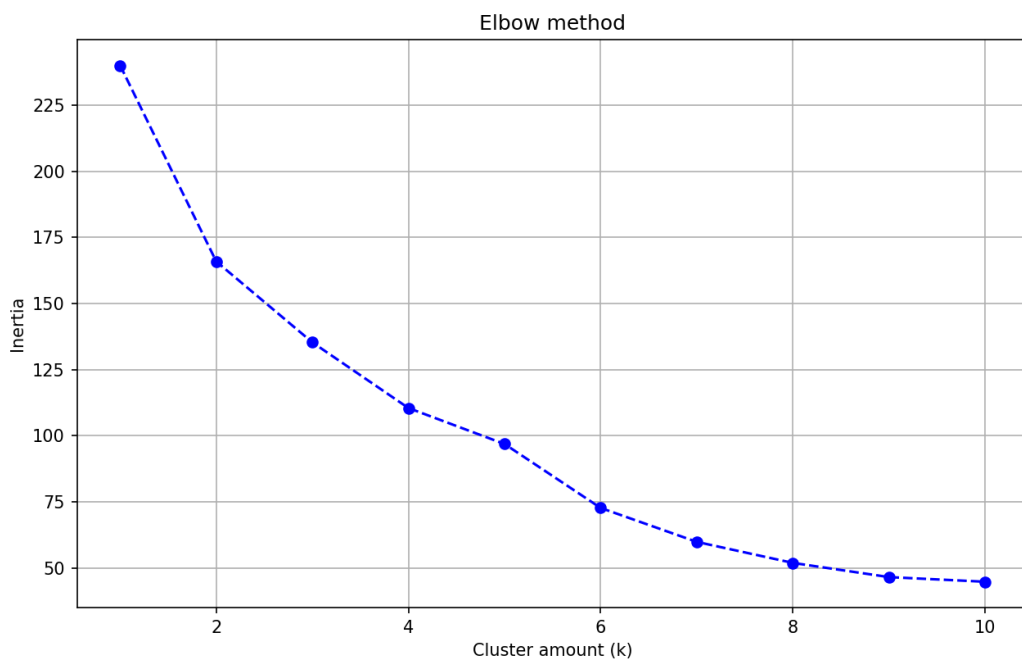


Рисунок 4.8 – Графік методу ліктя

Перед кластеризацією дані були стандартизовані, щоб уникнути домінування окремих змінних у результатах. В результаті було створено кілька візуалізацій, які наочно демонструють характеристики кожного кластера. Кожен кластер було детально проаналізовано, щоб виділити ключові особливості країн у ньому (таблиця 4.1).

Таблиця 4.1 – Таблиця результатів кластеризування

Показник	Кластер 0	Кластер 1	Кластер 2	Кластер 3
Середня відстань до України (км)	2274	1207	986	2532
Середній ВВП (млн)	942 336	286 496	3 836 242	938 397
Середній HDI	0.9436	0.8708	0.917	0.9168
Середній щомісячний платіж (євро)	481.5	197.1	224.5	130.7
Середня вартість проживання, євро	1126	744.9	850	861.7
Середня кількість мігрантів	81 219	154 737	1 748 094	75 416
Кількість країн	10	9	2	9

Кластеризація дозволила виділити групи країн із подібними характеристиками, що спрощує їх аналіз. Зокрема:

- кластер 0: кластер включає 10 економічно розвинених країн із високим рівнем ВВП (942 336 млн) і ІЛР (0.9436). Вони розташовані на значній відстані від України (2274 км), мають високу вартість проживання (1126 євро) і середні соціальні виплати (481.5 євро). Ці країни приймають помірну кількість мігрантів (81 219 в середньому) і характеризуються стабільними економічними умовами, хоча географічна віддаленість може стримувати міграцію;

- кластер 1: до цього кластеру входять 9 країн із середнім рівнем ВВП (286 496 млн) і ІЛР (0.8708). Вони розташовані ближче до України (1207 км), що робить їх доступними для мігрантів. Країни цього

кластеру мають нижчу вартість проживання (744.9 євро) та середній рівень соціальних виплат (197.1 євро). Вони приймають більшу кількість мігрантів (154 737 в середньому), що свідчить про їхню популярність через географічну близькість і доступні умови;

– кластер 2: кластер представлений лише 2 країнами, які мають найвищий ВВП (3 836 242 млн) і середній ІЛР (0.917). Відстань до України є найменшою серед усіх кластерів (986 км). Вартість проживання в цих країнах становить 850 євро, а соціальні виплати – 224.5 євро. Ці країни приймають найбільшу кількість мігрантів (1 748 094 в середньому), що пояснюється їхньою економічною привабливістю та близькістю до України;

– кластер 3: у кластер входять 9 країн із середнім рівнем ВВП (938 397 млн) і ІЛР (0.9168). Вони розташовані найдалі від України (2532 км), мають помірну вартість проживання (861.7 євро) та низькі соціальні виплати (130.7 євро). Ці країни приймають меншу кількість мігрантів (75 416 в середньому), що може бути пов'язано з їхньою географічною віддаленістю та обмеженими соціальними програмами.

Також було візуалізовано графік, який представляє порівняння країн за їх кластерною приналежністю на основі кількості заяв (рисунок 4.9).

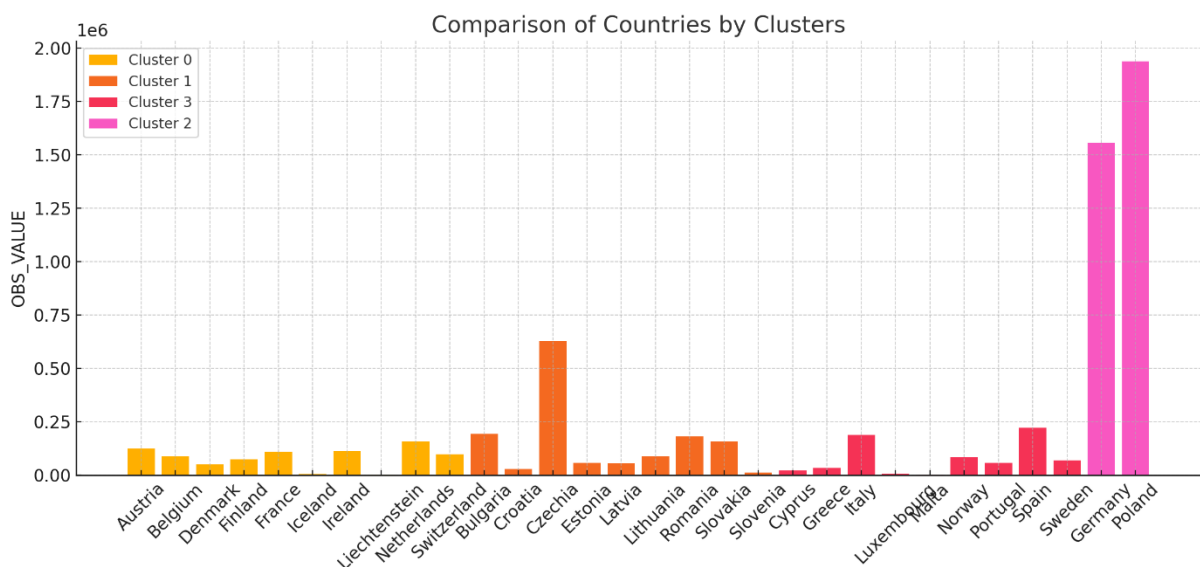


Рисунок 4.9 – Візуалізація за кластерами

Для наочності кластеризації були створені кілька візуалізацій, які ілюструють ключові особливості розподілу країн між кластерами:

– взаємозв'язок між ВВП (GDP) та ІЛР (HDI) (рисунок 4.10). Графік демонструє, як кластеризація розділяє країни за рівнем ВВП (GDP) та ІЛР (HDI). Це дозволяє чітко побачити економічні та соціальні відмінності між групами;

– відстань до України та вартість проживання (рисунок 4.11). Візуалізація показує вплив географічного розташування та економічної доступності на кластеризацію. Кластери добре розподіляються залежно від близькості до України та рівня витрат на життя;

– ІЛР (HDI) та середньої кількості мігрантів (рисунок 4.12). Графік ілюструє, як ІЛР впливає на міграційні потоки в кожному кластері.



Рисунок 4.10 – Кластери на основі Взаємозв'язок між GDP та HDI



Рисунок 4.11 – Кластери на основі відстані до України та вартість проживання

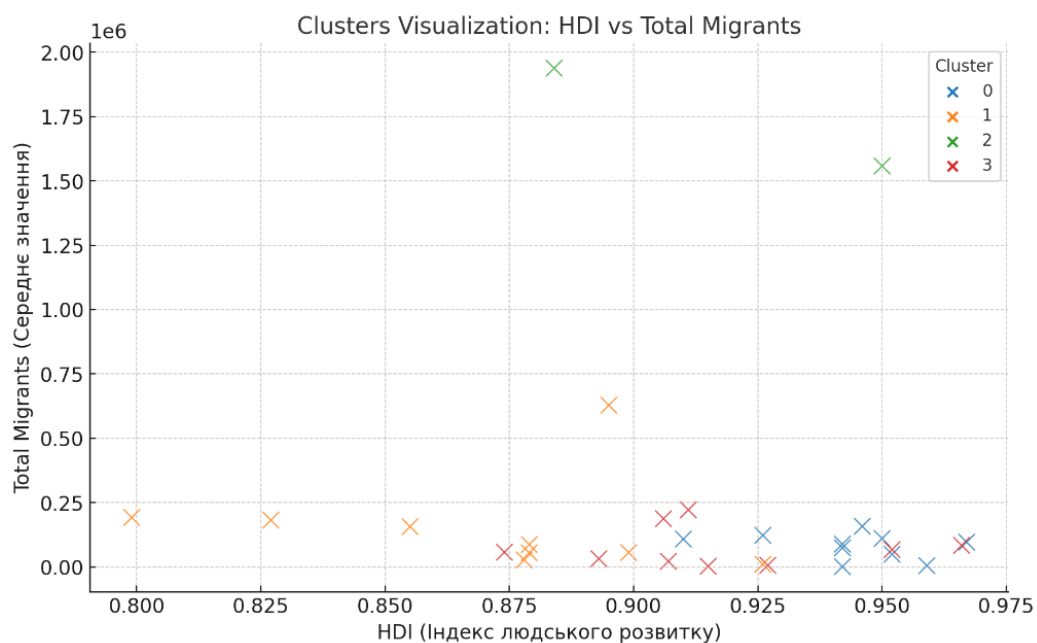


Рисунок 4.12 – Кластери на основі HDI та середньої кількості мігрантів

Візуалізації допомагають не лише зрозуміти структуру даних, але й визначити ключові характеристики кожного кластеру, що важливо для подальшого аналізу та прийняття рішень.

4.4 Аналіз часових рядів

Першим етапом аналізу було створення трендового графіка, який демонструє загальну картину змін міграційних потоків (рисунок 4.13). Такий графік допомагає оцінити масштаби та ключові періоди активності.

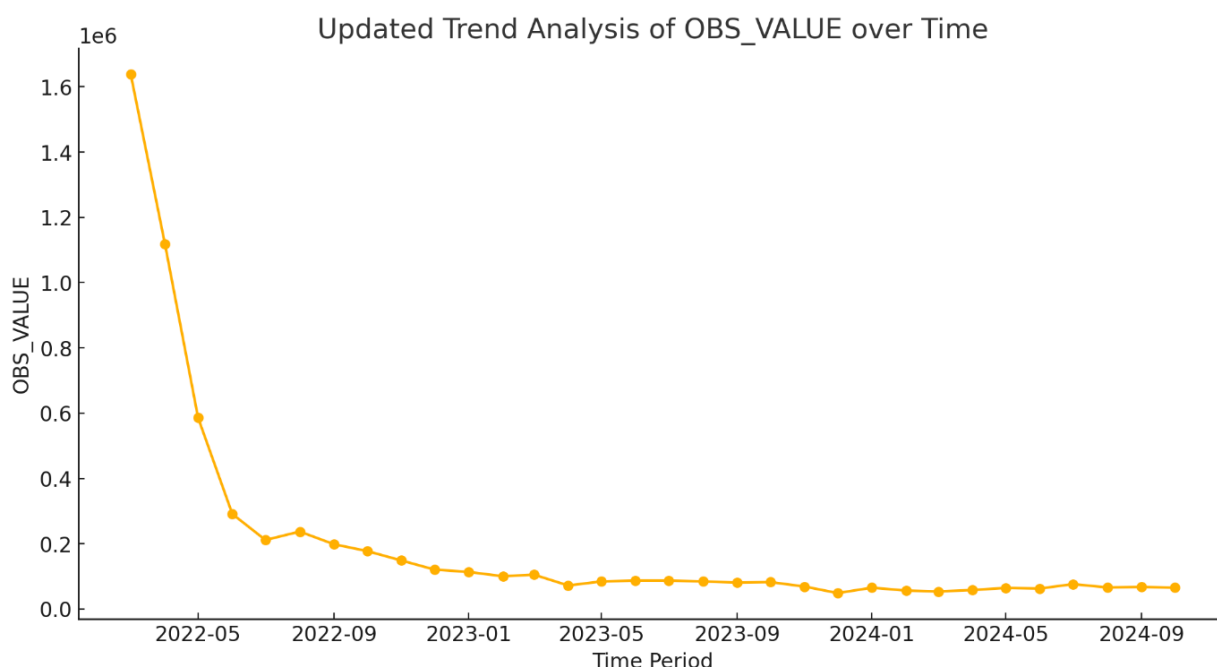


Рисунок 4.13 – Часовий ряд агрегованої вибірки

Для побудови графіка було використано дані про кількість заявок на тимчасовий захист (OBS_VALUE), згруповані за часовими періодами (TIME_PERIOD). Це дозволило створити безперервну лінію змін, яка демонструє, як кількість заявок змінювалася протягом досліджуваного періоду. Основна мета такого підходу – зрозуміти загальні закономірності, включаючи періоди різких змін чи стабілізації.

Агреговані дані для всіх країн були декомпозовані на тренд, сезонність і залишкові значення (рисунок 4.14). Такий підхід дозволив виділити довгострокову тенденцію, що відображає поступове зниження міграційних потоків після пікових значень. Сезонність, хоч і не є яскраво вираженою, демонструє слабкі коливання, пов'язані з періодами активізації. Залишкові значення дозволяють оцінити аномальні відхилення, які не пояснюються трендом чи сезонністю.

Для виявлення таких аномалій було застосовано алгоритм Isolation Forest. Він дозволив ідентифікувати кризові періоди, зокрема початок війни, коли кількість заявок різко перевищувала середній рівень. На графіку (рисунок 4.15) ці періоди позначені червоними точками.

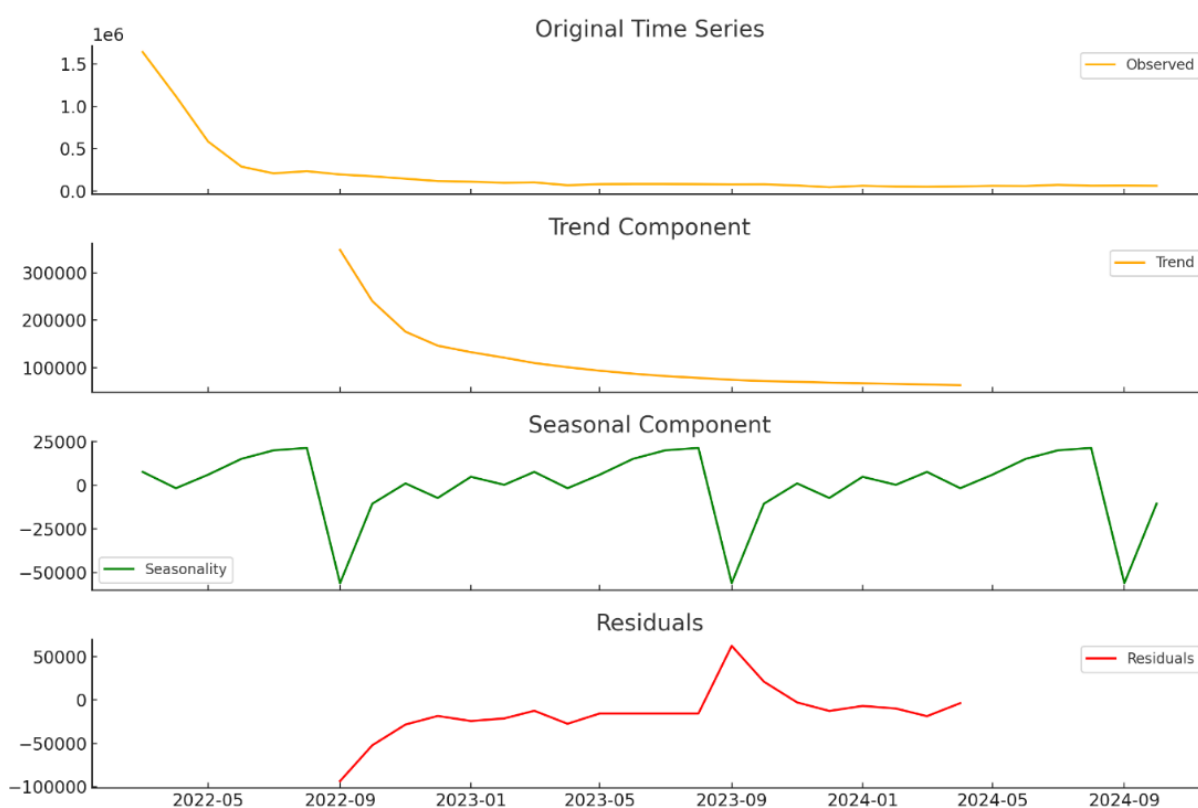


Рисунок 4.14 – Декомпозиція агрегованого часового ряду

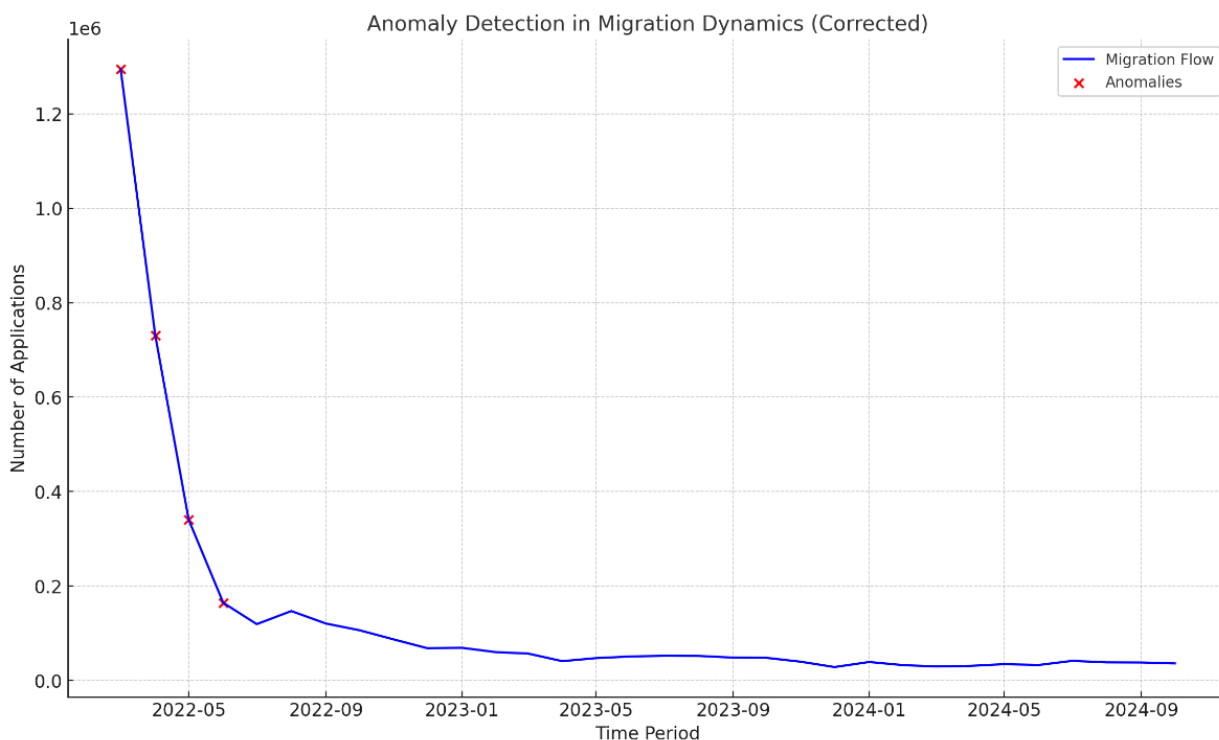


Рисунок 4.15 – Графік ідентифікації аномалій часового ряду

Для того, щоб проаналізувати динаміку міграційних потоків за країнами, ми побудували часові ряди, які відображають кількість міграційних подій за країнами за певний період. На рисунку 4.16 можна побачити здебільшого найпопулярніші країни міграції, де кожна лінія представляє зміну міграційної активності для окремої країни в часі. Таким чином, графік надає змогу оцінити як загальні тенденції, так і індивідуальні особливості деяких країн.

Як і на агрегованому графіку (рисунок 4.16) спостерігаються піки у березні 2022 року пов'язані з повномасштабним вторгненням Росії в Україну.

Ми бачимо різке зростання міграційної активності у країнах, що межують з Україною або традиційно приймають велику кількість біженців, таких як Польща, Німеччина та Чехія, для яких було створено окремий порівняльний графік (рисунок 4.17), який накладає часові ряди для країн на одну площину.

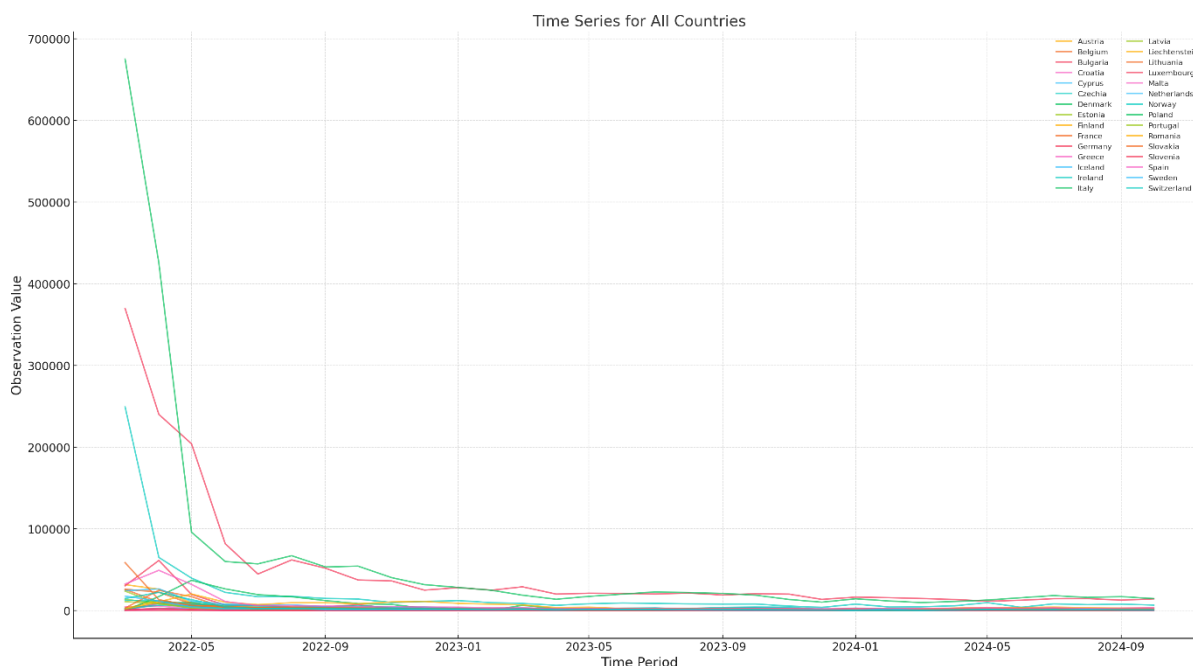


Рисунок 4.16 – Графік динаміки міграції за країнами

Дані для кожної країни були відфільтровані, згруповані за часом і агреговані.

Польща продемонструвала різке зростання кількості заявок на початку війни в березні 2022 року, після чого спостерігалось поступове зниження.

Німеччина, натомість, показала стабільно високий рівень активності протягом усього аналізованого періоду. Чехія характеризується помірним, але рівномірним прийомом мігрантів.

Це дозволило наочно продемонструвати схожі періоди пікової активності (березень 2022 року) та відмінності у динаміці між країнами. Наприклад, Німеччина має довший період стабільно високого рівня прийому біженців, тоді як у Чехії й Польщі спостерігається більш виражене зниження після початкового сплеску.

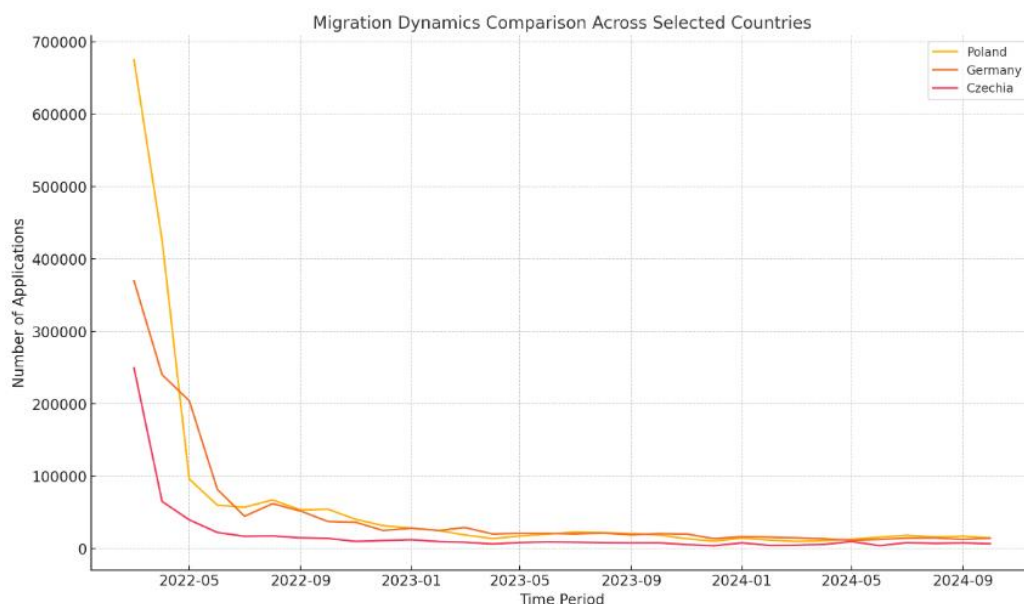


Рисунок 4.17 – Порівняльний графік Польщі, Німеччини та Чехії

Для додаткового дослідження потоків було обрано ще три європейські країни: Францію, Італію та Швейцарію. Усі вони знаходяться на приблизно однаковій відстані від України, що дозволяє оцінити вплив інших чинників на масштаби прийому мігрантів. Аналіз базується на часових рядах, представлених на графіку «Migration Dynamics Comparison» (рисунок 4.18).

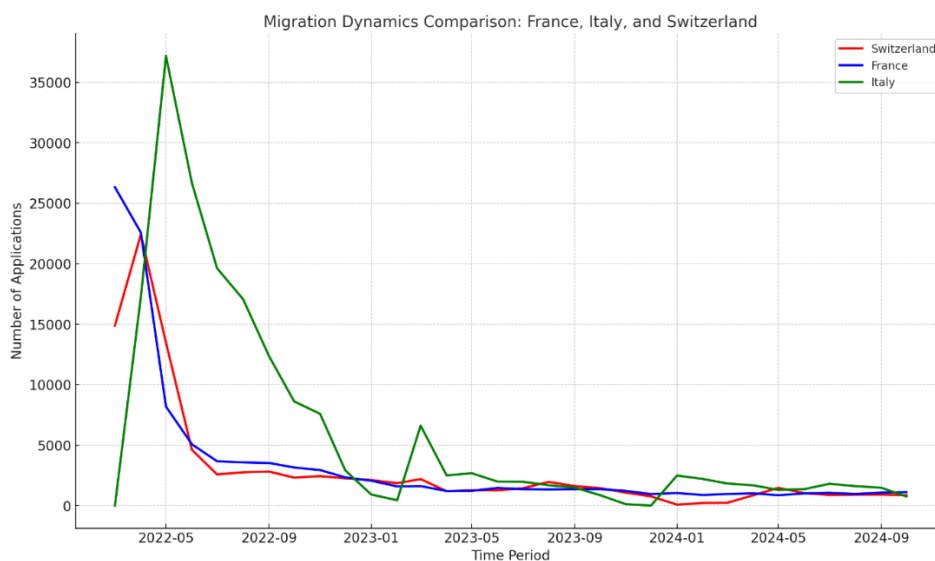


Рисунок 4.18 – Часовий ряд Швейцарії, Італії та Франції

Декомпозиція часових рядів для окремо Франції (рисунок 4.19), Італії (рисунок 4.20) та Швейцарії (рисунок 4.21) дозволяє глибше зрозуміти природу міграційних потоків.

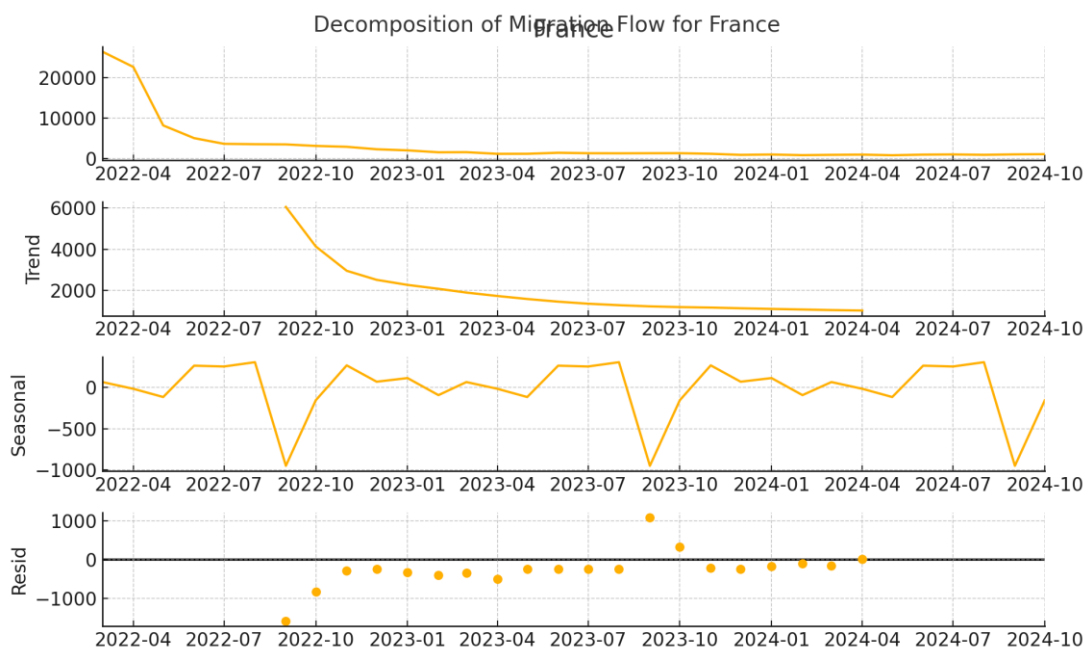


Рисунок 4.19 – Декомпозиція часового ряду Франції

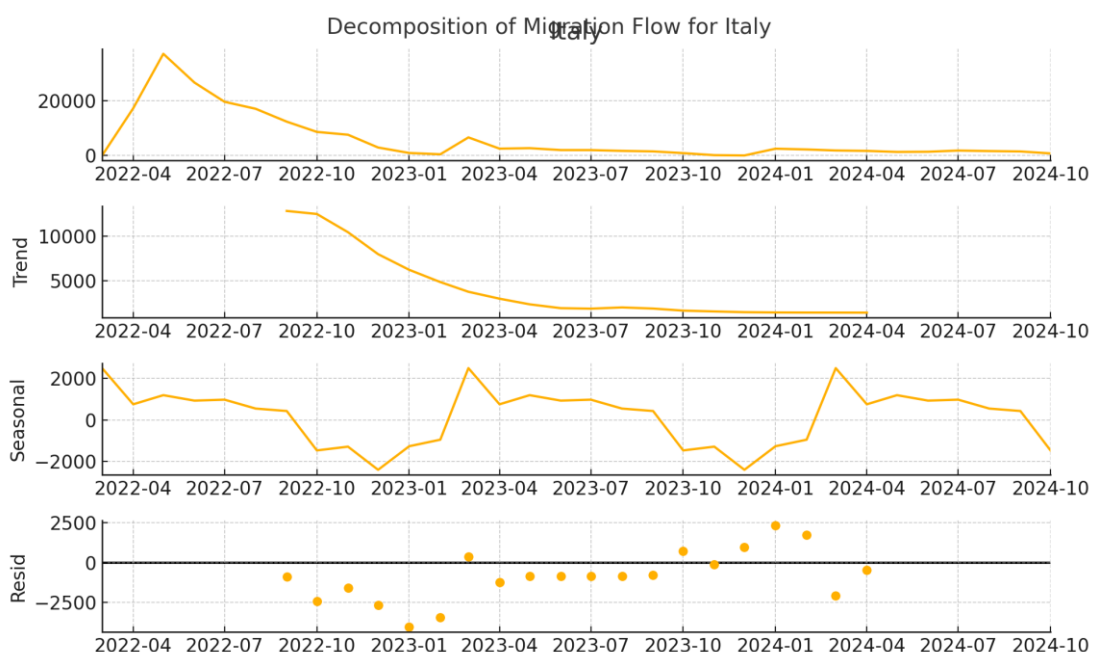


Рисунок 4.20 – Декомпозиція часового ряду Італії

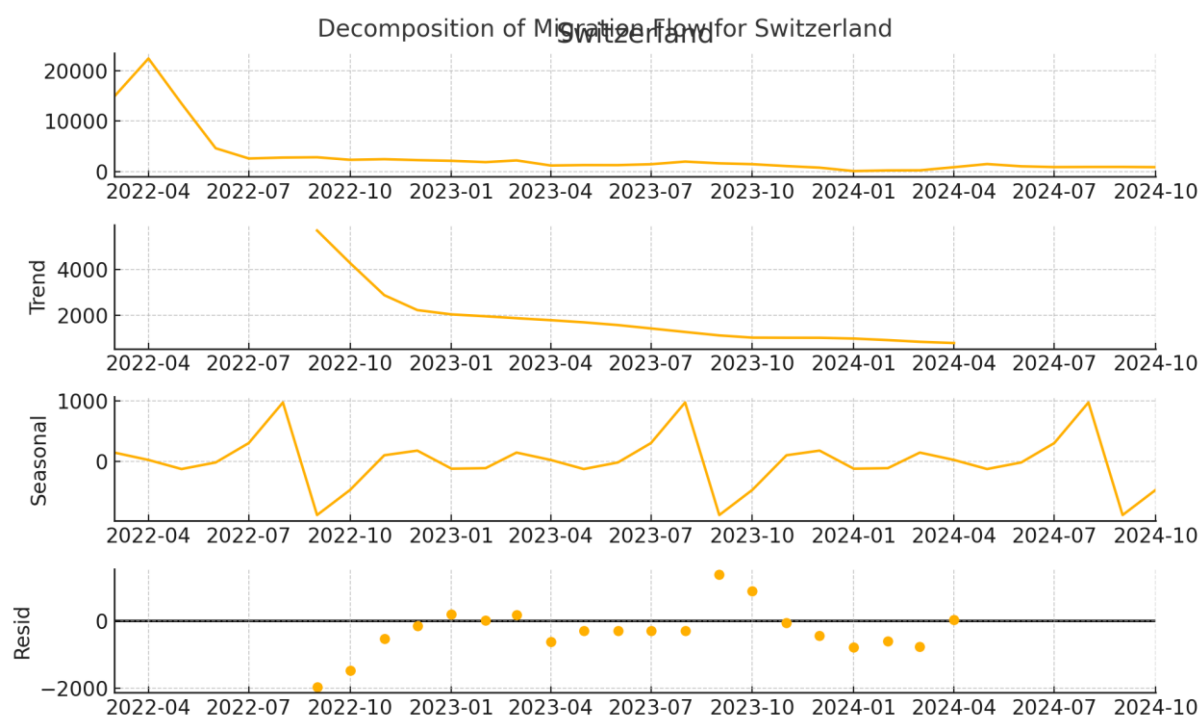


Рисунок 4.21 – Декомпозиція часового ряду Швейцарії

Тренд міграційних потоків до Італії демонструє виражені пікові значення, особливо в перші місяці війни в Україні. Це підтверджує її роль як основного пункту для новоприбулих мігрантів. Зростання міграції в цей період обумовлено не лише гуманітарною кризою, але й наявністю великої української діаспори, яка функціонує як соціальна та економічна підтримка. Після пікових періодів тренд поступово стабілізується, що свідчить про перехід до інтеграційної фази, коли більшість новоприбулих або залишаються в країні, або переселяються далі.

Тренд міграційних потоків до Франції залишається стабільним і не демонструє різких коливань. Це свідчить про суворішу політику прийому та меншу роль Франції як країни первинного в'їзду. Невелика кількість пікових значень також вказує на те, що французькі імміграційні служби не створюють умов для швидкої легалізації нових мігрантів. Водночас стабільність тренду може бути наслідком економічних можливостей для інтеграції мігрантів, але лише в межах уже встановлених квот.

Швейцарія демонструє найстабільніший тренд серед трьох країн. Кількість заявок на тимчасовий захист залишається відносно постійною, без значних піків навіть у кризові періоди. Це може бути обумовлено її географічним положенням, високими стандартами життя та ретельним контролем міграційних потоків. Така стабільність тренду підтверджує, що Швейцарія функціонує як країна кінцевого призначення для мігрантів, які планують довготривале перебування.

На об'єднаному графіку часових рядів (рисунок 4.18) чітко видно, що Італія демонструє найбільші коливання через кризові періоди, в той час як Франція і Швейцарія залишаються стабільними. Ці тренди відображають різну роль кожної країни в системі міграції: Італія як країна первинного в'їзду, Франція як країна зі стабільною міграційною політикою, та Швейцарія як пункт остаточного осідання.

Така різниця в трендах підтверджує, що навіть при схожих географічних умовах, міграційні потоки визначаються політикою, економічною привабливістю, рівнем підтримки діаспори та доступністю соціальних програм.

Графік часових рядів та декомпозицій підтверджує, що географічна відстань є лише одним із факторів, а вирішальними залишаються комплексні соціально-економічні та політичні аспекти.

4.4.1 Кластеризація часових рядів

Кластеризація часових рядів міграції була проведена з метою виявлення груп країн із подібною динамікою міграційних потоків у часі. На першому етапі ми підготували дані, нормалізувавши часові ряди кожної країни. Це було необхідно, щоб усунути вплив різниці в масштабах міграційної активності між країнами, дозволивши зосередитися на формах їхньої динаміки.

Далі ми застосували алгоритм K-Means, який розділив країни на чотири кластери на основі схожості їхніх часових рядів. Це дало змогу виявити групи країн із найбільш схожими трендами міграції.

Візуалізація кластеризації (рисунок 4.22) дозволяє побачити ключові особливості динаміки кожного кластера та їхніх країн.

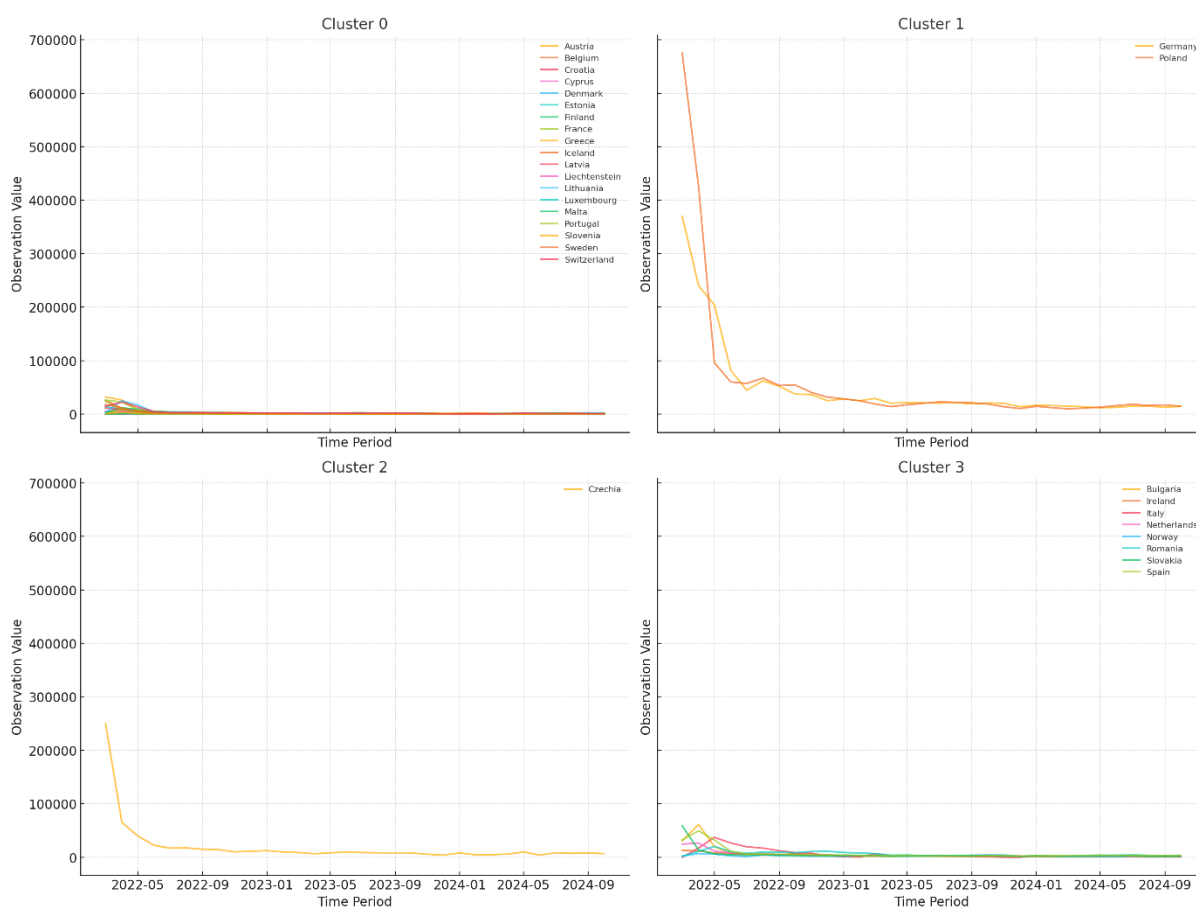


Рисунок 4.22 – Результат кластеризації часових рядів за країнами

Дослідження кластерів показало, що кожна група має специфічні характеристики. Кластер 0 об'єднує країни, що мають стабільно низькі значення протягом усього періоду. Їхні часові ряди не показують значних коливань або піків, і спостереження залишаються на мінімальному рівні.

Кластер 1 характеризується країнами, де значення є найвищими, і в часі спостерігається тенденція до стабілізації або поступового зменшення.

Їхні часові ряди демонструють відносно рівномірний характер без різких стрибків, що може вказувати на стабільну, високу міграційну активність.

У кластері 2 зібрано країни, де часові ряди відзначаються значними коливаннями, часто з різкими піками.

Кластер 3 включає країни з високими стабільними значеннями протягом усього періоду. Їхні часові ряди демонструють високу активність без значних коливань, що свідчить про їхню стійку роль у прийомі мігрантів. Це можуть бути країни з добре розвиненими міграційними політиками або ті, що є привабливими для мігрантів через економічні чи соціальні фактори.

Також над часовими рядами кожного кластеру було проведено декомпозицію що можна побачити на рисунках: cluster 0 (рисунок 4.23), cluster 1 (рисунок 4.24), cluster 2 (рисунок 4.25), cluster 3 (рисунок 4.26).

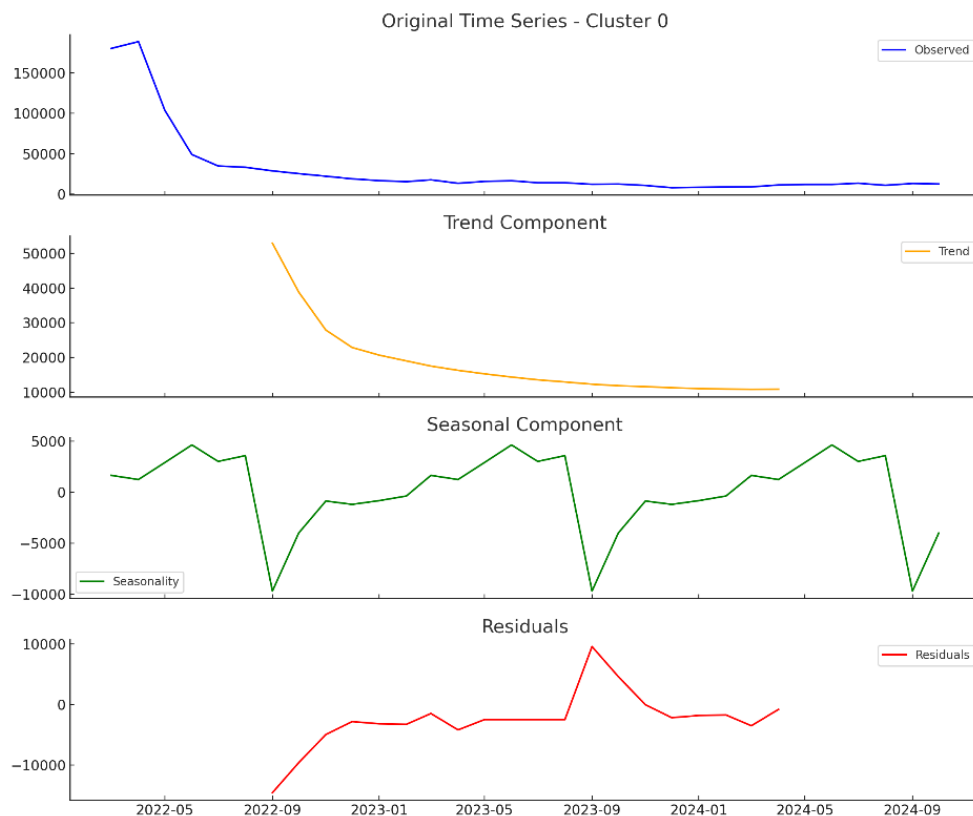


Рисунок 4.23 – Декомпозиція кластеру 0

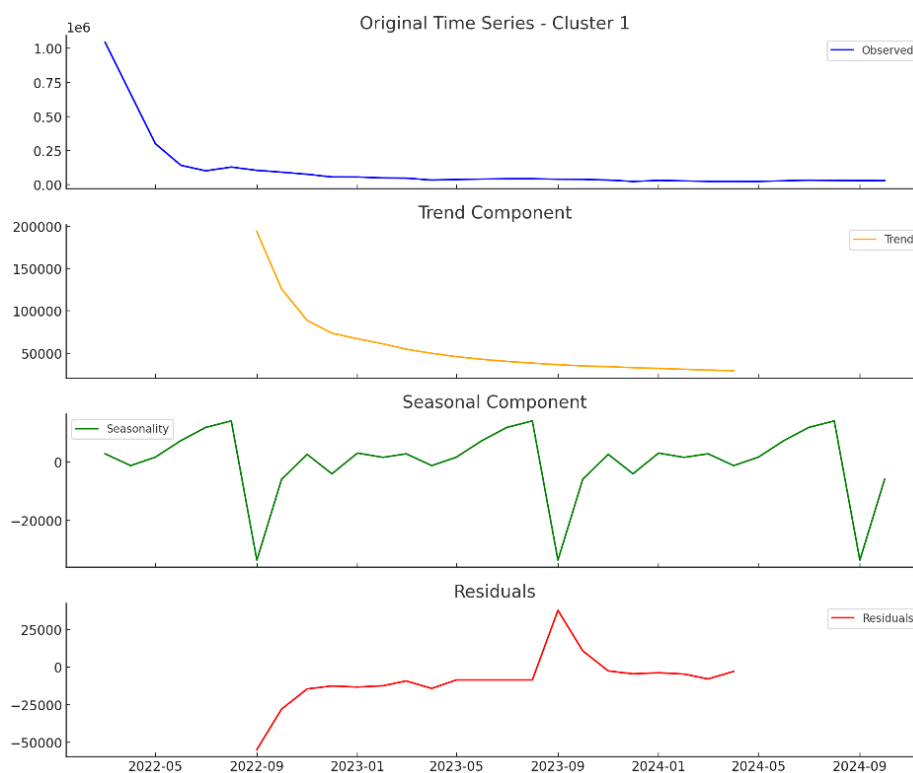


Рисунок 4.24 – Декомпозиція кластеру 1



Рисунок 4.25 – Декомпозиція кластеру 2



Рисунок 4.26 – Декомпозиція кластеру 3

Кластер 0 демонструє низькі рівні міграційної активності, які залишаються стабільними протягом усього періоду. Тренд вказує на незначне зростання у певні періоди, але загалом динаміка залишається стабільною. Сезонність майже відсутня, що свідчить про відсутність регулярних коливань у потоках.

Кластер 1 характеризується помірною міграційною активністю з періодами стабілізації. У спостережуваних даних інколи виникають незначні сплески. Тренд залишається відносно стабільним із короткими періодами помірного зростання. Сезонність слабка, її вплив на потоки незначний.

У кластері 2 спостерігаються значні коливання міграційної активності з помітними піками, що вказує на реакцію на окремі події чи фактори. Тренд демонструє стабільне зростання міграційних потоків, що свідчить про важливість цих країн у прийомі мігрантів. Сезонність виражена сильніше, що може бути ознакою певної циклічності у потоках. Залишкові значення

вказують на наявність аномалій під час пікових періодів, які, ймовірно, пов'язані з кризовими ситуаціями.

Кластер 3 має найвищий рівень міграційної активності серед усіх кластерів. Спостережувані дані відображають високу стабільність потоків. Тренд також показує стійкість і значні постійні значення, що підтверджує значну роль цих країн у прийомі мігрантів. Сезонність майже відсутня, потоки не піддаються регулярним коливанням.

У залишкових значеннях усіх 4 кластерів спостерігаються незначні аномалії, зокрема найбільша у вересні 2022 року, що може бути наслідком локальних подій.

Під час аналізу часових рядів кластерів для кожної країни була проведена перевірка на наявність сезонної компоненти.

Для аналізу сезонності та стаціонарності часових рядів, що відображають динаміку міграційних процесів, було проведено перевірку стаціонарності рядів за допомогою тесту Діккі-Фуллера (ADF).

Аналіз включав декомпозицію часових рядів на трендову, сезонну компоненти та залишки. Це дозволило оцінити внесок сезонної компоненти в загальну варіацію даних. Для перевірки наявності сезонності використовувалась автокореляційна функція (ACF) з фокусом на лагу 12.

Результати перевірки наведені в таблиці 4.2.

Таблиця 4.2 – Результати перевірки часових рядів на сезонну компоненту

Geo	Stationary	Seasonal Contribution	Lag 12 ACF	Seasonal 12 months
Austria	FALSE	0.32431658224953025	0.5369600561420027	Yes
Belgium	TRUE	0.11207155818711785	0.5929833477420069	Yes
Bulgaria	FALSE	0.19721363598624675	0.5133686427643628	Yes
Croatia	TRUE	0.19432031755297477	0.5274460363711349	Yes
Cyprus	FALSE	1.0094374102691286	0.5989156477879976	Yes
Czechia	FALSE	0.5450742229537899	0.500560816686168	Yes

Продовження таблиці 4.2

Geo	Stationary	Seasonal Contribution	Lag 12 ACF	Seasonal 12 months
Denmark	FALSE	0.5045176966173823	0.6175528785731302	Yes
Estonia	TRUE	0.04085147537642552	0.580705070131275	Yes
Finland	TRUE	0.3051823667051963	0.6038125466237139	Yes
France	TRUE	0.04051187899657569	0.5643778050588707	Yes
Germany	TRUE	0.024002310609216756	0.547807572796116	Yes
Greece	TRUE	0.5738310497144287	0.5211278643434124	Yes
Hungary	TRUE	0.03159085418617404	0.5368635534171994	Yes
Iceland	FALSE	0.9394216101386379	0.5742403158513029	Yes
Ireland	FALSE	0.6603637751657389	0.5823535878815044	Yes
Italy	TRUE	0.03772913430683614	0.5099106165916781	Yes
Latvia	FALSE	1.4293328722505612	0.5021260195801882	Yes
Liechtenstein	TRUE	0.9372848509975372	0.5834966756673499	Yes
Lithuania	TRUE	0.11483885768124565	0.5137278954020786	Yes
Luxembourg	TRUE	0.2797305424321095	0.5570964490566929	Yes
Malta	FALSE	1.0209073715835335	0.5364871875096637	Yes
Netherlands	TRUE	0.19577683147352407	0.5770637935436802	Yes
Norway	TRUE	0.958976070158676	0.5842909831758957	Yes
Poland	FALSE	0.38910975482175325	0.5082968482697899	Yes
Portugal	TRUE	0.0644461699847918	0.5869558144025824	Yes
Romania	FALSE	0.25894735470277785	0.534097803510488	Yes
Slovakia	FALSE	0.6047362702759604	0.5066112404506392	Yes
Slovenia	TRUE	0.10003154124471406	0.5317423741679334	Yes
Spain	TRUE	0.041714587201328444	0.562517267322351	Yes
Sweden	TRUE	0.2592393475254237	0.5352676742677108	Yes
Switzerland	FALSE	0.9047135737243595	0.5766059747593307	Yes
Aggregated Data	FALSE	0.26691532334479934	0.5016377017472143	Yes

Результати аналізу показали, що сезонна компонента мала різну значущість у різних країнах. Наприклад, та Мальті, Кіпрі та у Латвії

сезонність пояснює понад 90% загальної варіації, тоді як у Франції та Німеччині її вплив був мінімальним.

Значення автокореляційної функції підтвердили наявність стабільного сезонного патерну в більшості країн. Найвища сезонність спостерігалася в країнах із високими значеннями АСФ. Для агрегованого ряду сезонність була менш вираженою через згладжування різних патернів.

Отримані результати дозволяють зробити висновки про наявність і характер сезонності міграційних процесів у кожній країні. Це відкриває можливості для використання цих даних у подальших прогностичних моделях.

4.4.1 Мультифакторна кластеризація часових рядів

Для комплексного аналізу міграційних потоків українських біженців було реалізовано мультифакторну кластеризацію, яка поєднала динамічні та статичні характеристики країн (кластеризацію та кластеризацію часових рядів). Цей підхід дозволив врахувати як часові зміни потоків (середнє значення, стандартне відхилення, тренд), так і географічні, економічні та соціальні фактори приймаючих країн (ВВП, ІЛР, вартість життя тощо).

Перший етап передбачав обчислення часових характеристик потоків для кожної країни. Для цього були використані середні значення місячних потоків, їхні стандартні відхилення та тренди, розраховані як різниця між початковими та кінцевими значеннями потоку за аналізований період. Ці динамічні показники були об'єднані зі статичними характеристиками, такими як відстань до України, наявність морського кордону, ВВП на душу населення, ІЛР, середні виплати біженцям і вартість життя.

Другим етапом стало нормалізування всіх характеристик для забезпечення їх порівнянності, після чого було проведено кластеризацію за допомогою алгоритму K-means із вибором чотирьох кластерів. Кожен кластер об'єднав країни зі схожими динамічними та статичними

характеристиками, що дозволило виявити чіткі патерни, що можна побачити на графіках (рисунок 4.27).

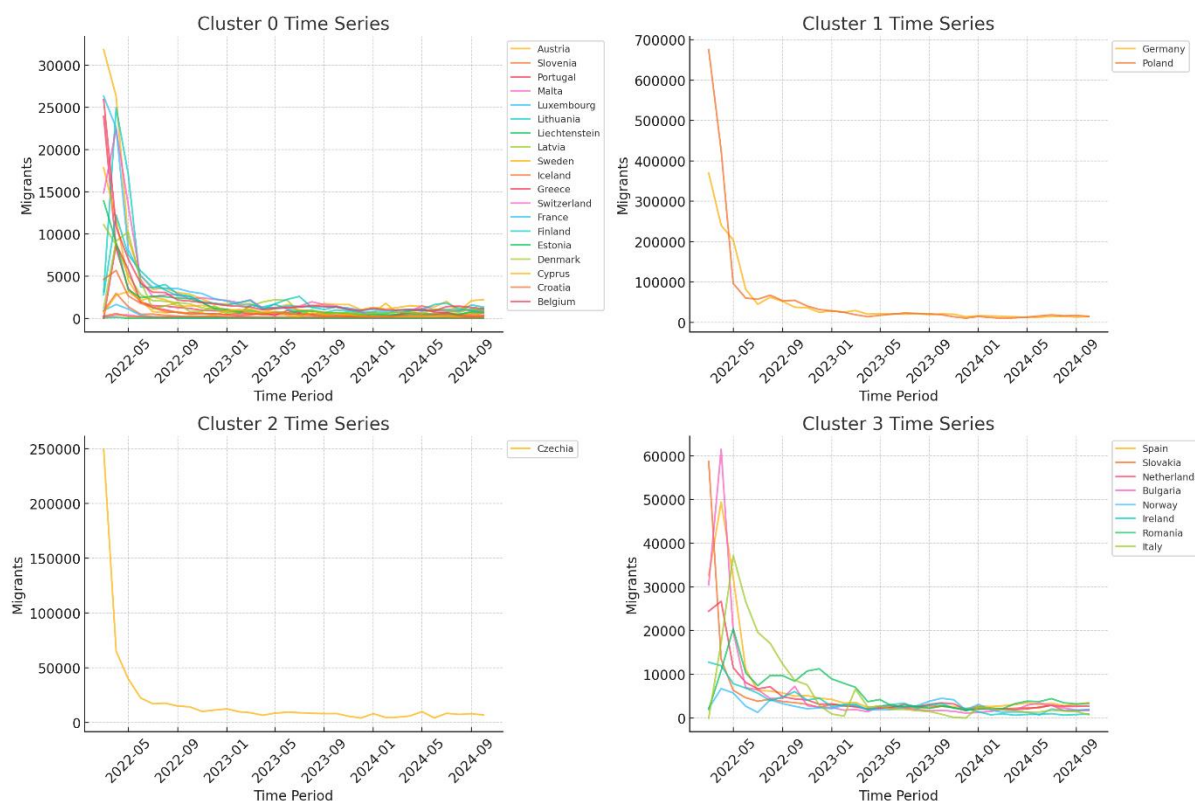


Рисунок 4.27 – Візуалізація мультифакторної кластеризації

Кластер 0 об'єднує країни, розташовані на значній відстані від України, із середньою відстанню понад 2000 км. Ці країни, такі як Ірландія, Іспанія та Швейцарія, не мають спільних кордонів із Україною, а лише 21% із них мають морські кордони. Культурна близькість також обмежена: лише 26% країн кластеру є слов'янськими. Проте вони вирізняються високим рівнем економічного розвитку, із середнім ВВП 516 тисяч і ІЛР 0.92. Вартість життя в цих країнах є однією з найвищих (середній показник – 966), що робить їх менш привабливими для першочергового переміщення біженців. Потоки до цих країн стабільні та відносно низькі, що свідчить про роль цих держав як кінцевого пункту міграції для певних категорій населення.

Кластер 1 охоплює країни, які є найближчими до України, із середньою відстанню близько 950 км. До них належать Польща, Німеччина та інші держави, які відіграли ключову роль у прийомі найбільшої кількості біженців у перші місяці війни. Понад половина країн цього кластеру мають спільний кордон із Україною, і стільки ж є слов'янськими. Економічно цей кластер найсильніший: середній ВВП становить 3.7 мільйона, а ІЛР – 0.91. Попри відносно помірні середні витрати на життя (838), потоки мігрантів до цих країн характеризуються яскраво вираженими піками на початку війни (березень–квітень 2022 року), що свідчить про їхню важливу роль як пунктів першого притулку.

Кластер 2 представлений єдиною країною – Чехією. Вона демонструє унікальну динаміку потоків, яка відрізняється від інших кластерів. Чехія розташована відносно близько до України (1120 км), але не має спільного кордону. Вона є слов'янською країною з середнім рівнем економічного розвитку: ВВП становить 585 тисяч, а ІЛР – 0.895. Динаміка міграційних потоків до Чехії характеризується помірною інтенсивністю, що дозволяє виділити її як окремий випадок.

Кластер 3 включає країни, розташовані на середній відстані від України (близько 1900 км). До нього входять, наприклад, Італія, Словаччина та Румунія. Лише чверть цих країн мають спільний кордон із Україною, а морські кордони є у 12.5%. Економічно цей кластер помірний, із середнім ВВП 1.2 мільйона та HDI 0.895. Вартість життя становить 825, а середні виплати – 225. Потоки мігрантів до цих країн з часом стабілізувалися, що свідчить про роль цих країн як резервних напрямків для мігрантів.

Кластеризація дозволила чітко визначити групи країн зі схожими патернами міграційних потоків, що відображає їхні економічні, географічні та соціальні особливості. Зокрема, близькі до України країни (кластер 1) зіграли ключову роль у прийомі біженців у перші місяці, тоді як віддалені держави (кластер 0) стали кінцевим пунктом для стабільніших потоків. Унікальні особливості Чехії (кластер 2) та помірна динаміка кластеру 3

свідчать про необхідність диференційованого підходу до політики міграції, враховуючи особливості кожної групи. Ці результати можуть бути корисними для прогнозування майбутніх потоків та розробки ефективних заходів підтримки мігрантів.

4.5 Вибір прогностичних моделей

Для вибору оптимальних прогностичних моделей враховувалися характеристики наявних даних, такі як стаціонарність, сезонність, трендові зміни та вплив аномальних періодів. Особливу увагу було приділено структурі часових рядів, адже різкі сплески, характерні для перших місяців війни в Україні, могли суттєво вплинути на результати аналізу. Для забезпечення точності прогнозування було розроблено підхід, що включає обробку аномалій, аналіз ключових властивостей даних та тестування моделей.

Першим етапом стало визначення впливу аномальних періодів, зокрема різкого зростання міграційних потоків у березні–травні 2022 року. Було протестовано кілька підходів до виключення цих періодів: виключення п'яти, чотирьох і трьох місяців. Оптимальним виявилось виключення перших чотирьох місяців 2022 року, що дозволило забезпечити стабільність тренду та зберегти достатню кількість даних для аналізу.

Далі проведено аналіз властивостей часових рядів за допомогою тесту ADF для перевірки стаціонарності. У 17 країн дані виявилися стаціонарними, що дало змогу застосовувати моделі SARIMA. Для решти країн, де спостерігалися тренди або сезонність, рекомендовано використання моделей, які враховують ці компоненти, наприклад Prophet. У разі значної нестаціонарності передбачено використання попереднього диференціювання.

Для оцінки сезонності було виявлено середню амплітуду змін на рівні 1 626 осіб, що є важливою характеристикою для окремих країн. Це

дозволило виділити категорії даних, для яких сезонні моделі, такі як ETS, можуть бути ефективнішими. Аналіз трендових змін показав середній приріст у 84 827 осіб, що свідчить про чітку динаміку, яку важливо враховувати під час прогнозування.

Наявність лише 28 часових точок у даних є обмежувальним фактором для використання складних моделей, таких як LSTM. Для таких моделей зазвичай потрібно кілька сотень точок, щоб коректно відобразити патерни даних. У зв'язку з цим основну увагу зосереджено на класичних підходах: ARIMA/SARIMA, ETS і Prophet, які ефективно працюють із невеликими обсягами даних.

Результати аналізу свідчать, що ARIMA/SARIMA підходять для стаціонарних рядів із мінімальним трендом, ETS забезпечує якісний прогноз для сезонних даних, а Prophet виявляється оптимальним для рядів із вираженими трендами та змінними сезонними компонентами. Для оцінки точності прогнозів заплановано використання метрик RMSE, MAE та MAPE [16], що дозволить порівняти ефективність кожної моделі в різних контекстах.

Загалом, обраний підхід передбачає дослідження кількох моделей, адаптованих до специфіки даних кожної країни чи кластеру. Це забезпечить комплексне охоплення усіх характеристик міграційних потоків і підвищить точність прогнозів.

5 ПРОГНОЗУВАННЯ МІГРАЦІЙНИХ ПОТОКІВ

5.1 Модель Sarima

SARIMA (Seasonal Autoregressive Integrated Moving Average) є однією з ключових моделей, яка використовувалася в рамках цього дослідження для прогнозування міграційних потоків. Ця модель інтегрує три ключові аспекти: автокореляцію, сезонність і диференціювання для досягнення стаціонарності. SARIMA дозволяє моделювати динаміку даних із річними або місячними циклами, що робить її надзвичайно ефективною для аналізу міграційних потоків, які зазвичай мають сезонний характер.

Модель SARIMA використовує параметри:

- p, d, q : для моделювання автокореляції (p), диференціювання (d) та ковзного середнього (q) у несезонній частині часового ряду;
- P, D, Q, s : для моделювання автокореляції (P), сезонного диференціювання (D), сезонного ковзного середнього (Q), де s визначає період сезонності.

Ця модель поєднує здатність працювати зі стаціонарними часовими рядами, описувати як короткострокові, так і сезонні патерни, та адаптуватися до змін у динаміці даних.

Також важливу роль у застосуванні SARIMA в нашому дослідженні відіграє диференціювання адже це дозволило обробити ряди, де спостерігався поступовий тренд у даних через зміну демографічної чи соціальної ситуації.

5.2 Модель ETS

ETS (Exponential Smoothing State Space Models) – це модель, яка дозволяє аналізувати часові ряди шляхом розкладання їх на три основні компоненти: тренд, сезонність і залишки. У нашому дослідженні ETS була

застосована для вивчення даних із чітко вираженими сезонними циклами й трендами. Завдяки використанню експоненціального згладжування ETS моделює залежності в даних із високою точністю, особливо якщо дані демонструють стабільну структуру сезонності та плавні зміни тренду.

ETS підтримує різноманітні конфігурації компонентів. Адитивні моделі застосовуються для рядів із постійною амплітудою сезонних коливань, тоді як мультиплікативні моделі враховують зміну амплітуди пропорційно до рівня ряду [17]. Гнучкість ETS у виборі структури робить її універсальним інструментом для аналізу різних типів даних.

Важливою перевагою ETS стала її здатність адаптуватися до різних типів даних без необхідності попереднього перетворення. Це дозволило ефективно використовувати модель для часових рядів, які відображають не лише сезонність, а й довгострокові соціально-економічні тенденції.

5.3 Модель Prophet

Prophet – це потужна модель для аналізу й прогнозування часових рядів, розроблена компанією Meta [18]. Її головною особливістю є здатність автоматично виявляти тренди й сезонні компоненти у даних, навіть якщо вони мають нерегулярний характер або змінюються з часом. Модель базується на адитивному підході до комбінування трендів, сезонності й залишків, а також може враховувати зовнішні фактори, такі як свята чи інші події.

Prophet ефективно працює з часовими рядами, які мають нестационарні тренди, точки змін (changepoints) і множинні сезонності [19]. Завдяки цьому вона ідеально підходить для аналізу міграційних потоків, які можуть змінюватися через вплив різних політичних, економічних або соціальних чинників. Prophet містить функціонал для перехресної перевірки часових рядів для вимірювання похибки прогнозу на основі історичних даних [20].

Модель Prophet дозволяє гнучко налаштовувати параметри для адаптації до особливостей часового ряду. У цьому дослідженні було використано кілька ключових гіперпараметрів, які забезпечили високу точність прогнозування міграційних потоків:

- `seasonality_mode`. Цей параметр визначає тип сезонності в даних: `additive/multiplicative`;
- `changepoint_prior_scale`. Цей параметр контролює чутливість моделі до змін тренду (`changepoints`). Високе значення цього параметра дозволяє моделі адаптуватися до різких змін у тренді, тоді як низьке згладжує тренд;
- `seasonality_prior_scale`. Параметр відповідає за ступінь регулярності сезонної компоненти. Вище значення робить модель більш чутливою до сезонних змін;
- `fourier_order`. Цей параметр визначає кількість гармонік, що використовуються для моделювання сезонності [21].

5.4 Тестування та оцінка точності моделей

5.4.1 Тестування моделі ETS

Під час використання моделі Exponential Smoothing буде виконано прогнозування даних з для кожної країни окремо, кластерів та агрегованого ряду. Це дозволить оцінити, наскільки ефективно модель може передбачати кількість заявок на тимчасовий захист з урахуванням індивідуальних особливостей кожної країни, таких як масштаби міграції, адміністративні фактори або специфічні сезонні патерни.

Було проведено прогнозування для кожного окремого кластера країн. Кластеризація була виконана заздалегідь в розділі роботи 4.4.1. Такий підхід дозволяє визначити, чи покращує згруповане прогнозування результати

порівняно з аналізом кожної країни окремо, а також виявити можливі спільні закономірності в межах кожного кластера.

На завершальному етапі буде виконано агреговане прогнозування всіх заявок. Це передбачає створення загального прогнозу кількості українців, які отримують тимчасовий захист у будь-якій європейській країні, незалежно від їх розподілу по окремих державах. Такий підхід дозволить оцінити загальну картину міграційних процесів і, за потреби, порівняти результати з прогнозами для окремих країн або кластерів.

Підхід, що поєднує аналіз на рівні країн, кластерів і агрегованих даних, забезпечує комплексне бачення проблематики та дозволяє оцінити сильні й слабкі сторони моделі ETS у різних умовах застосування.

В процесі тестування моделі було досліджено використання аддитивного та мультиплікативного трендів. Для оцінки результатів прогнозування тестування проводилося на останніх відомих двох місяцях даних. Результати тестування з використанням аддитивного тренду можна спостерігати в таблиці 5.1.

Таблиця 5.1 – Таблиця аналізу прогнозування моделлю ETS, аддитивний тренд

Country	MAE	Mean	MAPE	MSE
Austria	1179.088769	2160.0	49.939277	1.407902e+06
Belgium	255.232097	1217.5	22.155686	6.621873e+04
Bulgaria	612.416987	1747.5	7.892773	5.962809e+05
Croatia	86.867895	242.5	22.467947	9.033044e+03
Cyprus	358.253545	397.5	75.705983	1.309018e+05
Czechia	1309.364577	7400.0	21.698413	1.885409e+06
Denmark	369.174138	1067.5	24.373756	1.444066e+05
Estonia	540.032816	735.0	70.646534	2.916914e+05
Finland	249.004815	950.0	34.716006	6.771722e+04
France	239.444654	1095.0	10.765418	7.277297e+04
Germany	1651.206379	13665.0	9.745756	2.882876e+06

Продовження таблиці 5.1

Country	MAE	Mean	MAPE	MSE
Greece	61.488882	312.5	10.378276	4.671018e+03
Hungary	340.056448	567.5	62.126614	1.156411e+05
Iceland	50.418326	77.5	92.362963	2.567002e+03
Ireland	498.830390	850.0	28.889034	3.228438e+05
Italy	1132.420294	1115.0	71.592996	1.288776e+06
Latvia	188.065511	642.5	21.456742	3.863812e+04
Liechtenstein	9.583365	22.5	60.417468	9.809153e+01
Lithuania	1157.709168	1455.0	79.905325	1.352126e+06
Luxembourg	15.013161	67.5	43.653381	3.063266e+02
Malta	4.999372	32.5	16.900573	2.499873e+01
Netherlands	680.889557	2677.5	21.819169	4.739312e+05
Norway	1185.373391	1822.5	65.004251	1.407395e+06
Poland	3756.616037	15997.5	18.457840	1.445073e+07
Portugal	188.578123	555.0	27.773163	3.776460e+04
Romania	381.322073	3222.5	18.572909	1.911767e+05
Slovakia	374.418278	1837.5	30.795648	1.638211e+05
Slovenia	13.653542	152.5	0.102353	3.683933e+02
Spain	639.192881	3330.0	11.217219	4.918891e+05
Sweden	185.936767	807.5	24.588215	3.522385e+04
Switzerland	134.204414	885.0	0.575588	3.465805e+04

Далі проведемо прогноз по агрегованим даним (таблиця 5.2).

Таблиця 5.2 – Тестування на агрегованому ряді, аддитивний тренд

MAE	Mean	MAPE	MSE
8357.685228514092	67107.5	12.42915527895202	7.11e+7

Наступним кроком була перевірка тих самих часових рядів з використанням мультиплікативного методу. Результати тестування було записано в таблицю (таблиця 5.3).

Таблиця 5.3 – Таблиця аналізу прогнозування моделлю ETS, мультиплікативний тренд

Country	MAE	Mean	MAPE	MSE
Austria	1095.658432	2160.0	45.899308	1.218442e+06
Belgium	149.873571	1217.5	17.204312	2.790594e+04
Bulgaria	1036.324556	1747.5	12.098988	1.743981e+06
Croatia	440.154123	242.5	127.679371	2.211734e+05
Cyprus	276.415159	397.5	62.425578	8.008776e+04
Czechia	1007.124613	7400.0	18.705890	1.242902e+06
Denmark	213.552360	1067.5	12.256356	5.096535e+04
Estonia	463.226815	735.0	67.999491	2.186456e+05
Finland	162.216030	950.0	28.591966	3.736401e+04
France	184.475151	1095.0	10.104646	3.986123e+04
Germany	322.086816	13665.0	1.217385	1.310398e+05
Greece	44.312015	312.5	8.888498	2.259467e+03
Hungary	221.744536	567.5	39.843322	4.917743e+04
Iceland	59.238178	77.5	105.125420	3.523885e+03
Ireland	841.302302	850.0	57.167427	8.618756e+05
Italy	411.365417	1115.0	35.348980	1.809427e+05
Latvia	808.878505	642.5	117.937837	6.622871e+05
Liechtenstein	8.574570	22.5	56.008569	8.042513e+01
Lithuania	787.758893	1455.0	68.000507	7.046926e+05
Luxembourg	200.591161	67.5	230.093560	4.571869e+04
Malta	3.908987	32.5	7.425978	1.810659e+01
Netherlands	602.910023	2677.5	18.416983	3.764826e+05
Norway	1924.169880	1822.5	92.684383	3.793750e+06
Poland	1486.503122	15997.5	9.332888	2.223794e+06
Portugal	181.016448	555.0	30.582109	3.339434e+04
Romania	287.657068	3222.5	4.229164	1.058848e+05
Slovakia	374.061492	1837.5	30.503987	1.621474e+05
Slovenia	268.752058	152.5	94.784760	8.593925e+04
Spain	606.352349	3330.0	9.513977	4.631692e+05
Sweden	250.865995	807.5	31.886306	6.347926e+04
Switzerland	62.559554	885.0	0.045897	7.775883e+03

Проведемо агрегацію по кластерам і виконаємо прогнозування даних по кожному кластеру (таблиця 5.4).

Таблиця 5.4 – Результати прогнозування, по кластерам

Cluster	MAE	MEAN	MAPE	MSE
0	3537.2767	12842.5	27.487	1.28e+7
1	239.6330	29695	0.796	1.05e+5
2	1007.1246	7400	13.208	1.24e+6
3	2229.0667	16602.5	13.399	5.54e+6

Далі проведемо прогноз по агрегованим даним (таблиця 5.5).

Таблиця 5.5 – Результат прогнозування агрегованого ряду

MAE	MEAN	MAPE	MSE
3645.151824460234	67107.5	5.380905038989927	1.76e+7

Можна спостерігати, що мультиплікативний тренд, в середньому, працює точніше і краще, ніж аддитивний, тому далі будемо використовувати саме таку конфігурацію.

Для сезонності було обрано для всіх варіацій саме аддитивний тренд, бо він стабільно дає кращий результат, у порівнянні з відсутністю сезонності та мультиплікативною сезонністю.

Агреговане прогнозування для всіх країн Європи показало, що середня абсолютна помилка (MAE) становила 3645, що становить близько 5.38% від середнього значення реальних даних (MAPE).

Аналіз окремих країн показав, що ETS добре справляється з прогнозуванням у випадках, коли дані мають чітку структуру та стабільні сезонні компоненти. Наприклад, для Німеччини середня абсолютна помилка становила лише 322, при середньому значенні даних 13665, що

демонструє високу точність прогнозування. Проте для країн із більш нестабільними даними, таких як Ісландія або Словенія, помилки були значно більшими, що вказує на наявність нерегулярних коливань у даних, які модель не змогла врахувати.

У результаті можна зробити висновок, що ETS-модель з мультиплікативним трендом та адитивною сезонністю є ефективним інструментом для аналізу та прогнозування часових рядів у рамках даної задачі.

Для Польщі (рисунок 5.2), Німеччини (рисунок 5.1), Швейцарії (рисунок 5.3) та Ліхтенштейну (рисунок 5.4) було створено графіки тестування мультиплікативного методу.

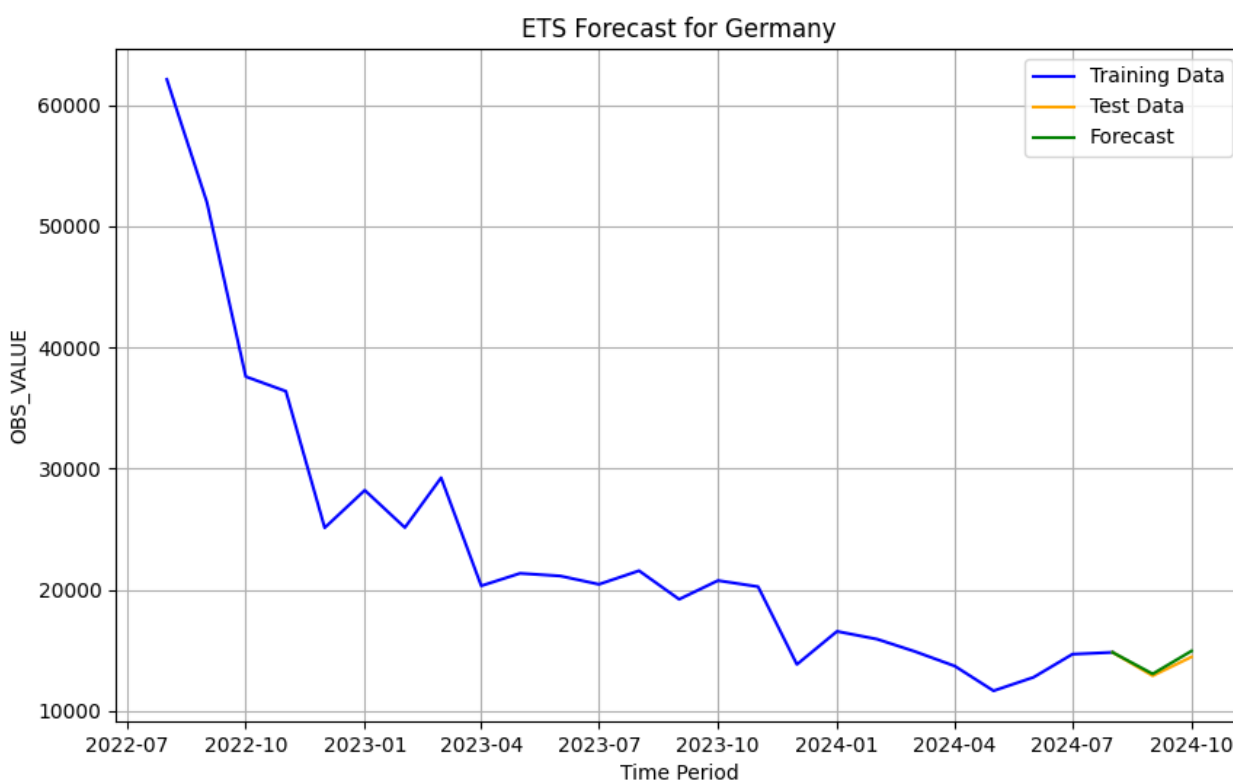


Рисунок 5.1 – Графік тестування для часового ряду Німеччини

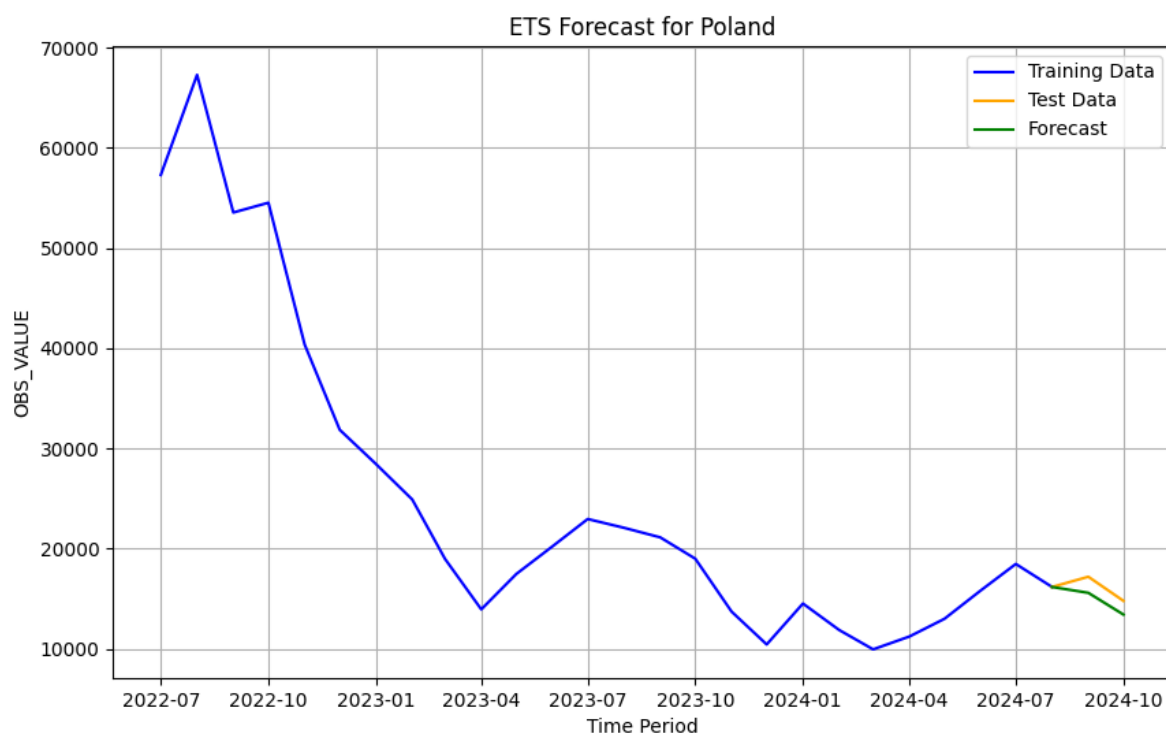


Рисунок 5.2 – Графік тестування для часового ряду Польщі

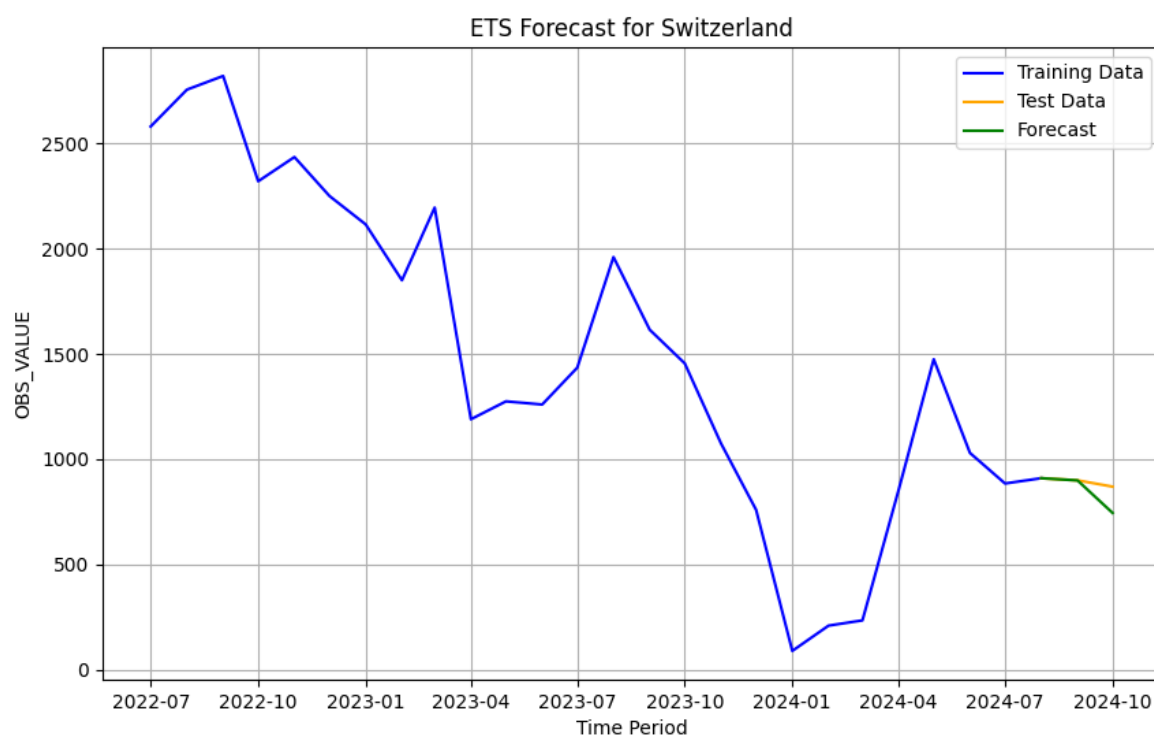


Рисунок 5.3 – Графік тестування для часового ряду Швейцарії

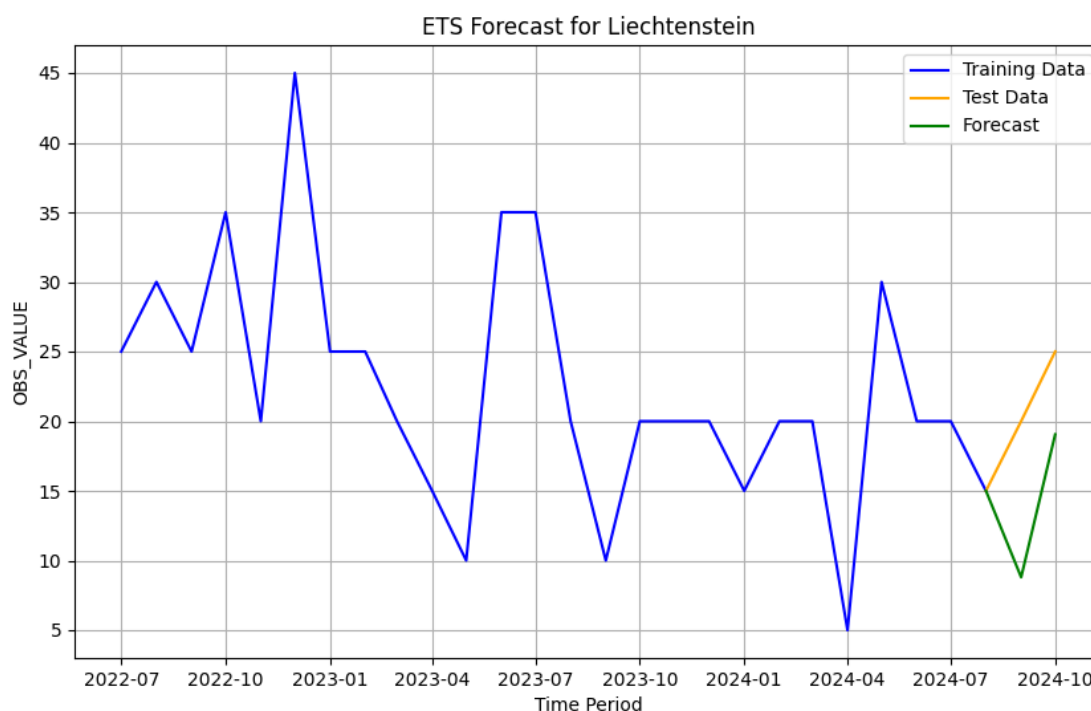


Рисунок 5.4 – Графік тестування для часового ряду Ліхтенштейну

5.4.3 Тестування моделі Sarima

Для моделювання моделі Sarima було використано параметри $p=1$ (вказує на кількість лагів у компоненті) авторегресії, $d=0/1$ (вказує на кількість різниць для досягнення стаціонарності), $q=1$ (визначає кількість лагів у компоненті ковзного середнього). Сезонні компоненти моделі були визначені параметрами $p=1$, $d=1$, $q=1$, і сезонний період дорівнює 12 місяцям.

Параметр d , що відповідає за інтеграцію даних, визначається індивідуально для кожної країни на основі перевірки стаціонарності тренувального набору даних. Для цього використовується тест ADF. Якщо дані є стаціонарними ($p\text{-value} < 0.05$), то $d=0$, інакше $d=1$. Цей підхід дозволяє адаптувати модель до часових рядів із різною структурою.

Після отримання прогнозу обчислюються метрики точності, зокрема MAE (mean absolute error), MAPE (mean absolute percentage error), MAPE для першого прогнозного періоду та MSE (mean squared error). Результати

прогнозу супроводжуються розрахунком довірчих інтервалів, які враховують можливу невизначеність моделі.

В процесі тестування даної моделі було створено таблиці з метриками точності для кожної країни (таблиця 5.6), а також для кластерів (таблиця 5.7) і агрегованого часового ряду (табл.5.8).

Таблиця 5.6 – Таблиця аналізу прогнозування моделлю SARIMA

Country	d	MAE	Mean	MAPE	MSE
Austria	1	521.142319	2160.0	23.562632	4.483093e+05
Belgium	0	47.659515	1217.5	4.101248	3.181237e+03
Bulgaria	1	402.848067	1747.5	23.213525	1.688944e+05
Croatia	0	37.920130	242.5	15.998118	1.448953e+03
Cyprus	1	397.500000	397.5	100.000000	1.783125e+05
Czechia	1	814.136235	7400.0	10.462516	1.026925e+06
Denmark	0	657.434141	1067.5	61.215380	4.426210e+05
Estonia	0	353.452580	735.0	47.681272	1.304210e+05
Finland	0	163.844960	950.0	17.397301	3.457488e+04
France	0	11.143800	1095.0	1.009488	1.417415e+02
Germany	0	944.780803	13665.0	6.816288	9.748580e+05
Greece	0	23.305513	312.5	7.553924	6.860507e+02
Hungary	0	192.235679	567.5	33.911524	3.695545e+04
Iceland	1	42.823354	77.5	62.951216	2.070840e+03
Ireland	1	291.252887	850.0	33.210223	1.041032e+05
Italy	0	54.509669	1115.0	5.702212	3.039277e+03
Latvia	1	494.201107	642.5	75.649602	2.787389e+05
Liechtenstein	0	22.500000	22.5	100.000000	5.125000e+02
Lithuania	0	679.115580	1455.0	45.194587	5.510612e+05
Luxembourg	0	12.688850	67.5	22.224574	2.553048e+02

Продовження таблиці 5.6

Country	d	MAE	Mean	MAPE	MSE
Malta	1	10.036981	32.5	29.876936	1.257172e+02
Netherlands	0	707.843476	2677.5	26.398675	5.172207e+05
Norway	0	1981.878490	1822.5	108.075501	4.076144e+06
Poland	1	2257.318958	15997.5	14.163716	5.098732e+06
Portugal	1	170.882422	555.0	14.163716	2.991697e+04
Romania	1	2624.801580	3222.5	81.351579	7.264296e+06
Slovakia	1	720.024883	1837.5	40.516920	6.192228e+05
Slovenia	0	9.751780	152.5	6.600402	1.297449e+02
Spain	1	65.988513	3330.0	2.034480	4.950493e+03
Sweden	0	156.750829	807.5	19.485918	2.457189e+04
Switzerland	1	1093.700086	885.0	123.188898	1.246222e+06

Таблиця 5.7 – Таблиця аналізу прогнозування на агрегованому ряді

MAE	Mean	MAPE	MSE
3318.929126886345	67107.5	4.993228130876276	1.43e+7

Таблиця 5.8 – Таблиця аналізу прогнозування моделлю SARIMA, по кластерам

Cluster	MAE	Mean	MAPE	MSE
0	4090.993117	12842.5	27.487	1.28e+7
1	4023.472258	29695	0.796	1.05e+5
2	814.136235	7400	13.208	1.24e+6
3	3108.504447	16602.5	13.399	5.54e+6

Було також реалізовано кілька візуалізацій у виді графіків результатів моделювання: Німеччина (рисунок 5.5), Ліхтенштейн (рисунок 5.6), Польща (рисунок 5.7) та Швейцарія (рисунок 5.8).

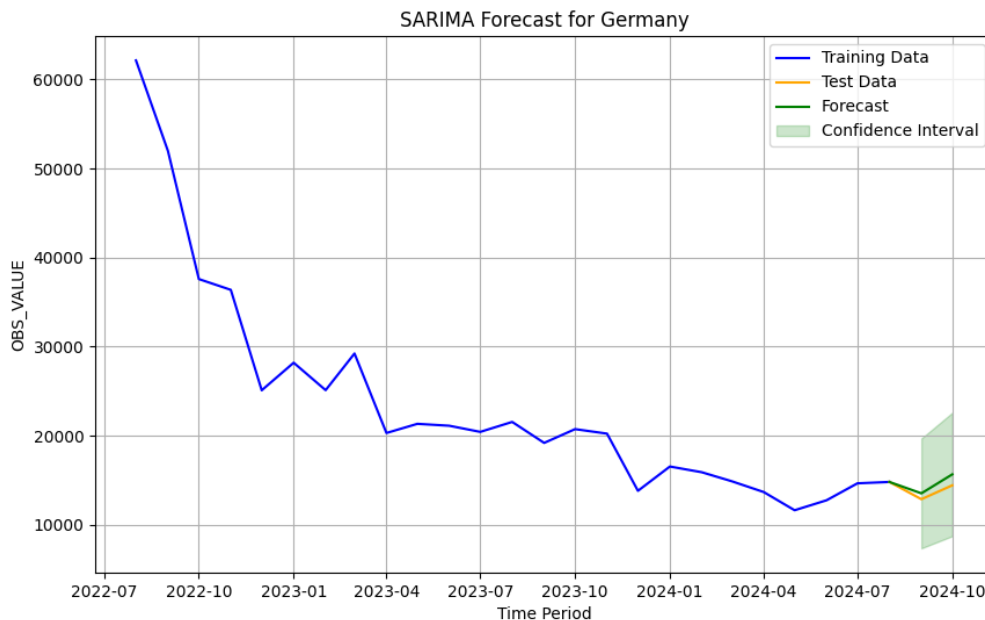


Рисунок 5.5 – Тестування моделі на часовому ряді Німеччини

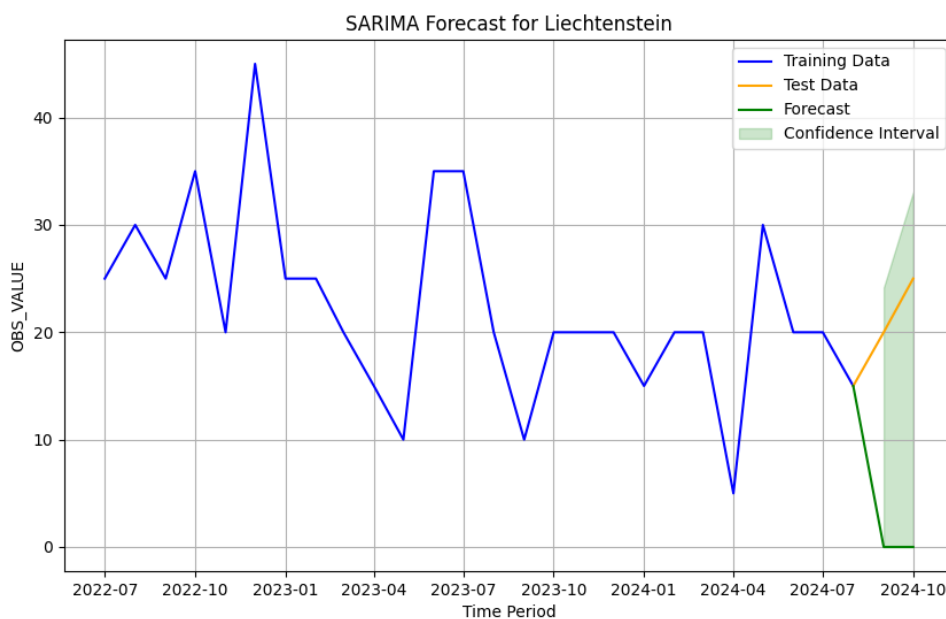


Рисунок 5.6 – Тестування моделі на часовому ряді Ліхтенштейну

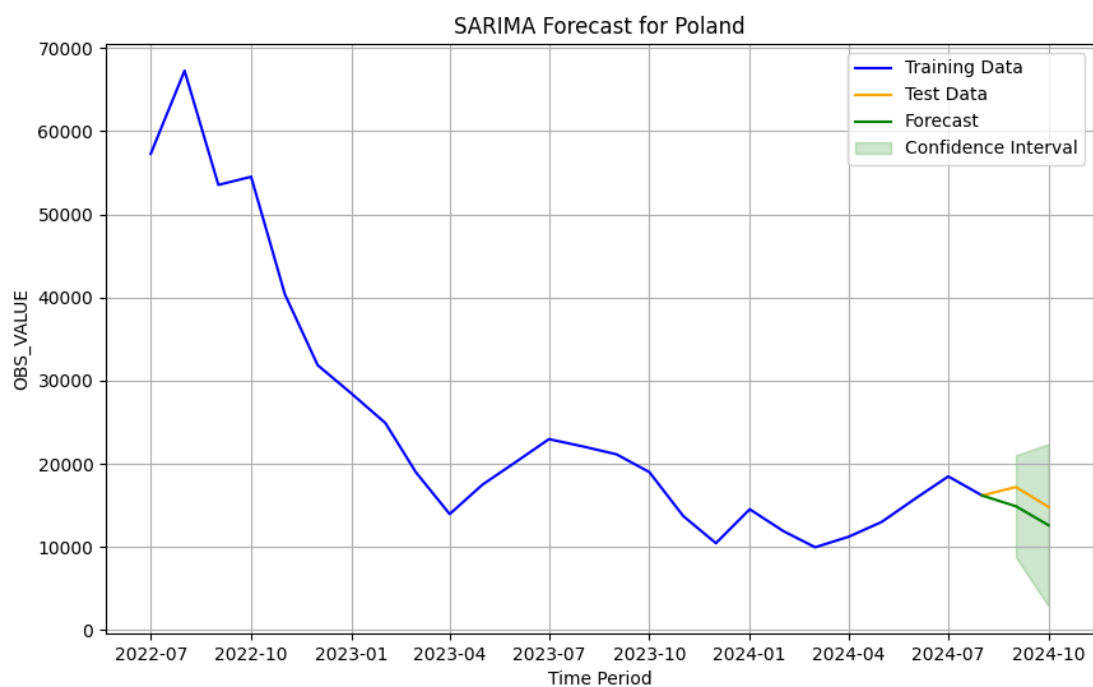


Рисунок 5.7 – Тестування моделі на часовому ряді Польщі

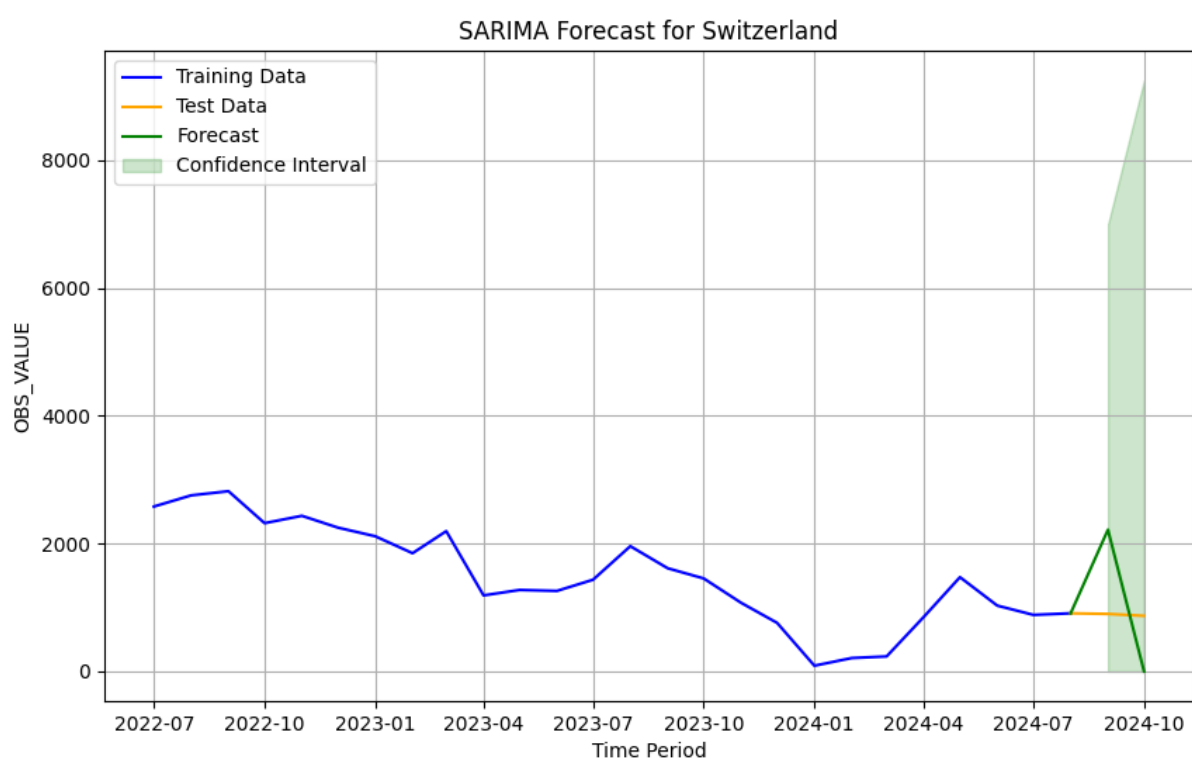


Рисунок 5.8 – Тестування моделі на часовому ряді Швейцарії

5.4.1 Тестування моделі Prophet

Метою тестування було перевірити якість прогнозу моделі Prophet шляхом мінімізації помилок на тестовій вибірці. Модель навчалася на історичних даних, а результати оцінювалися на незалежній тестовій вибірці.

Для першого тестування моделі було обрано однакові гіперпараметри для усіх країн . seasonality_prior_scale: 1, changepoint_prior_scale: 0.05, seasonality_mode: 'multiplicative', seasonality_period: 6, fourier_order: 5.

Результати тестування можна побачити в таблиці (таблиця 5.9). Можна спостерігати що деякі з країн показують хороші результати, деякі – добрі але здебільшого результати погані. Після цього тестування була зроблена спроба змінити комбінацію гіперпараметрів для отримання кращих результатів тестування, проте жоден з можливих варіантів не був вдалим. При зміні гіперпараметру деякі з результатів ставали кращі , в той час як інші погіршувались.

Таблиця 5.9 – Метрики оцінювання тестових прогнозів

Country	MAE	MEAN	MSE	MAPE
Austria	837.6500174017397	2157.5	706011.789980355	38.83%
Belgium	220.17410247631358	1212.5	63705.23096166988	18.16%
Bulgaria	1292.4725829878173	1750.0	1688280.7403586279	73.86%
Croatia	173.349599048358	242.5	30147.47566407374	71.48%
Cyprus	218.93221931388103	400.0	66590.86047197139	54.73%
Czechia	697.1092988836867	7397.5	935631.8040568745	9.42%
Denmark	270.4765768945761	1062.5	81767.76981089424	25.46%
Estonia	228.71577722956005	735.0	54321.8471856489	31.12%
Finland	304.58711973584104	945.0	93026.3487881104	32.23%
France	149.85589548569823	1097.5	22877.450778961902	13.65%
Germany	836.2897719683897	13667.5	1316995.947253894	6.12%
Greece	521.8088025466509	312.5	272353.3000181682	166.98%
Iceland	21.72470394644707	85.0	608.7249717871778	25.56%

Продовження таблиці 5.9

Country	MAE	MEAN	MSE	MAPE
Ireland	770.2120971375398	852.5	607652.2119947791	90.35%
Italy	372.13667603376155	1117.5	269022.0752327309	33.30%
Latvia	177.65141779205373	645.0	32866.56125794908	27.54%
Liechtenstein	7.685096835466062	17.5	92.28659377041014	43.91%
Lithuania	413.63124860088	1460.0	191735.9590059523	28.33%
Luxembourg	13.939439517865985	70.0	246.51071051912396	19.91%
Malta	8.794299371923781	32.5	80.58526961347721	27.06%
Netherlands	147.79910429430788	2680.0	23325.216630226387	5.51%
Norway	1126.9134159342484	1825.0	1297711.8984000906	61.75%
Poland	2923.135670331456	15995.0	12003391.746311741	18.28%
Portugal	64.47791522826947	555.0	5516.042758909438	11.62%
Romania	1377.3359032757712	3225.0	1914423.930501402	42.71%
Slovakia	451.2042894565991	1835.0	219504.6234682187	24.59%
Slovenia	56.41711463498088	152.5	3260.9568006089958	36.99%
Spain	157.93463814929373	3330.0	49326.48625071273	4.74%
Sweden	349.780552490589	807.5	127831.14879046526	43.32%
Switzerland	292.1950968121994	887.5	88865.12172856393	32.92%
Aggregated Data	4653.494156200948	68898.75	54575024.48753035	6.754%

Наступним кроком було проведено підбір гіперпараметрів з використанням визначеного діапазону значень:

- changepoint_prior_scale: [1, 5, 10];
- seasonality_prior_scale: [0.05, 0.07, 0.1, 0.2, 0.5, 0.7];
- fourier_order: ['additive', 'multiplicative'];
- seasonality_mode: [3, 6, 12];
- seasonality_period: [3, 5, 7].

У процесі роботи було протестовано декілька комбінацій параметрів, включаючи варіанти сезонності з різними періодами та ступенем гнучкості. Параметри оптимізувалися шляхом порівняння метрик точності, зокрема MAE (Mean Absolute Error) та MSE (Mean Squared Error).

Для агрегованого ряду найкращий результат був отриманий з такими параметрами: `seasonality_prior_scale: 5`, `changepoint_prior_scale: 0.05`, `seasonality_mode: 'multiplicative'`, `seasonality_period: 3`, `fourier_order: 3`. Ця комбінація забезпечила значення $MAE = 5251.72$, $MSE = 61396701.50$ та має відсоткову середню абсолютну похибку (MAPE), що становить 7.62% від середнього фактичного значення даних. Це свідчить про відносно високий рівень точності прогнозу, оскільки похибка складає менше 10%. Таким чином, модель може вважатися достатньо надійною для використання в прогнозуванні, особливо в контексті міграційних потоків, які характеризуються складністю та нестабільністю, і відносно високу точність прогнозу на тестовій вибірці. Візуалізація результатів (рисунок 5.9) також підтвердила, що прогноз моделі відповідає фактичним значенням.

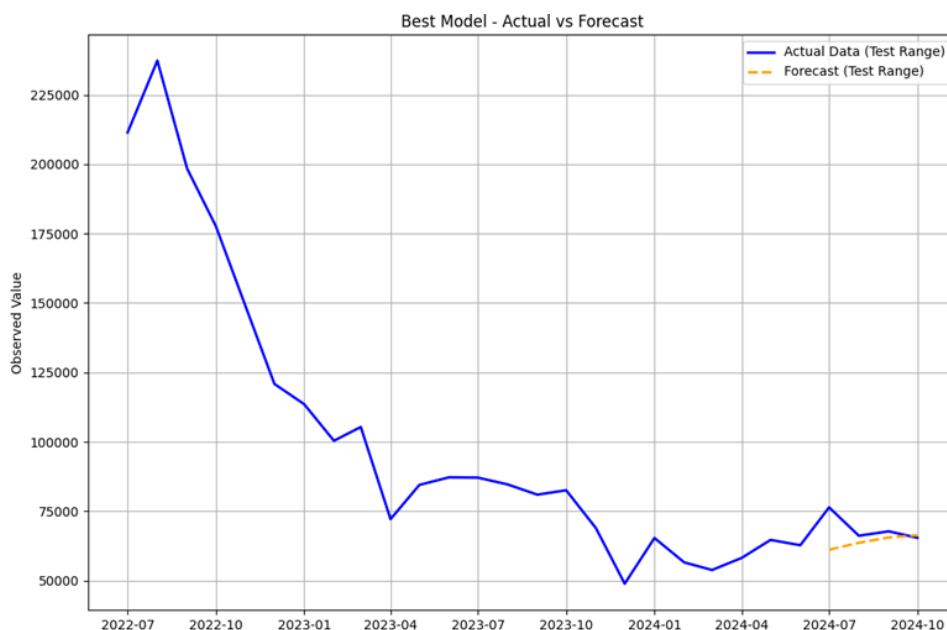


Рисунок 5.9 – Тестування на агрегованому ряді

Тестування моделі з підбором гіперпараметрів на основі мультифакторної кластеризації можна спостерігати на рисунках: cluster 0 (рисунок 5.10), cluster 1 (рисунок 5.11), cluster 2 (рисунок 5.12), cluster 0 (рисунок 5.13).

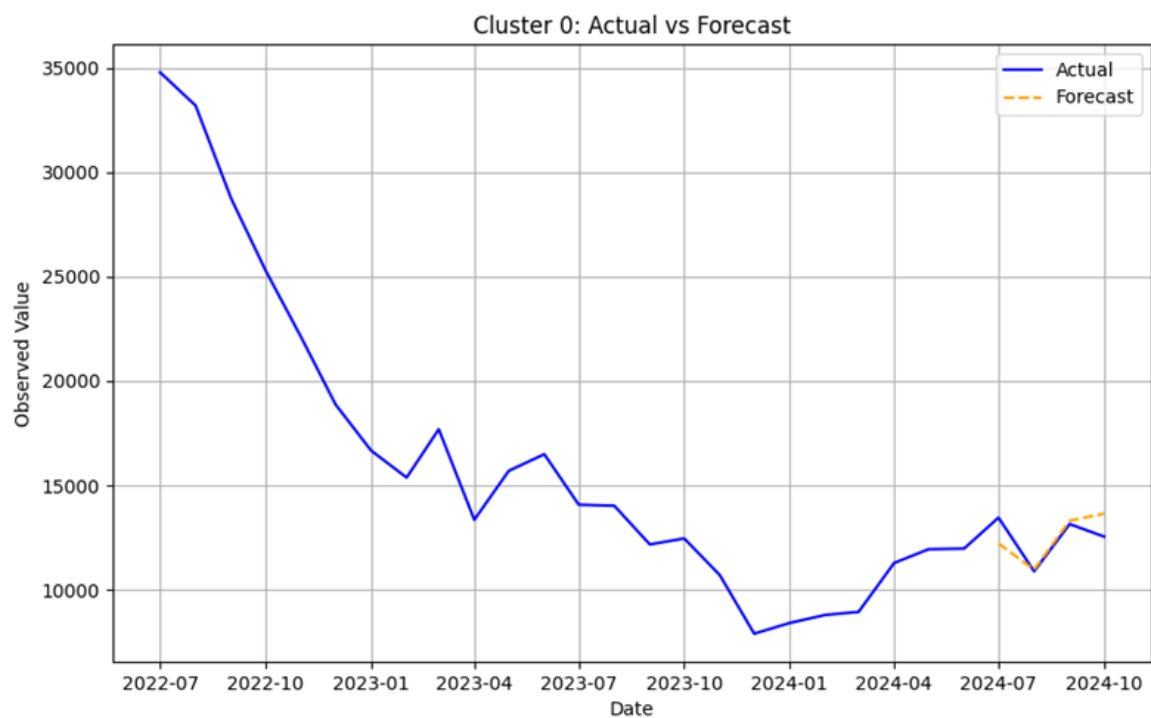


Рисунок 5.10 – Тестування даних кластеру 0

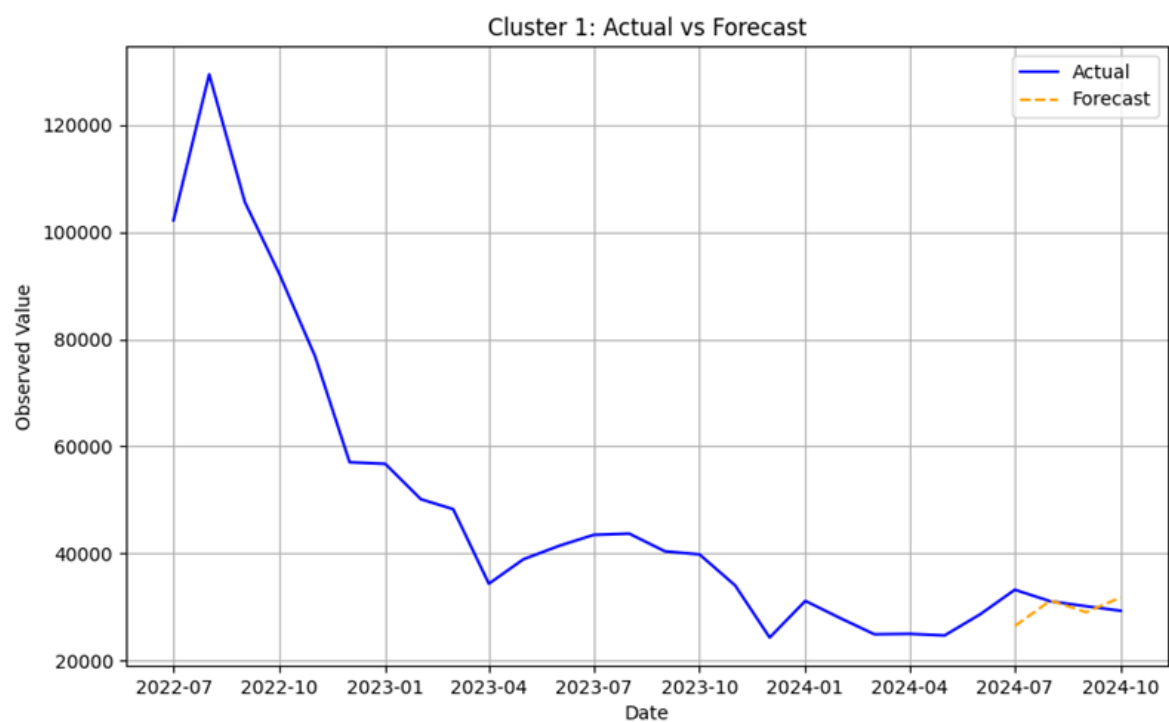


Рисунок 5.11 – Тестування даних кластеру 1

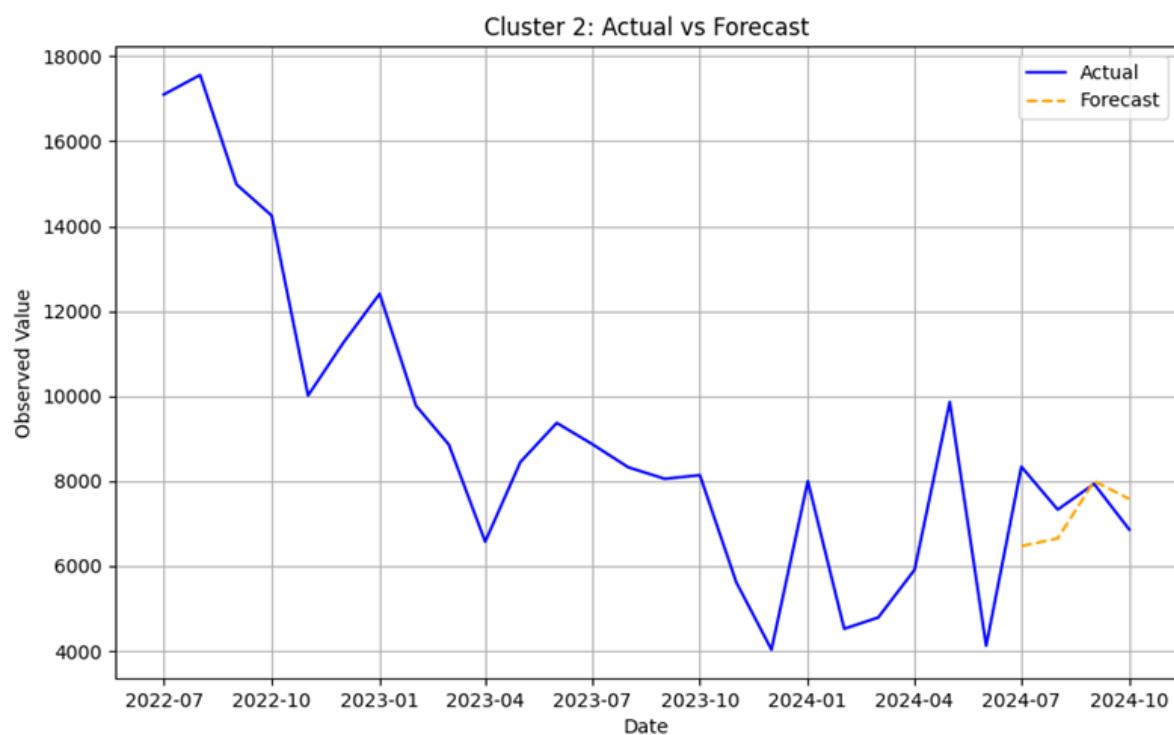


Рисунок 5.12 – Тестування даних кластеру 2

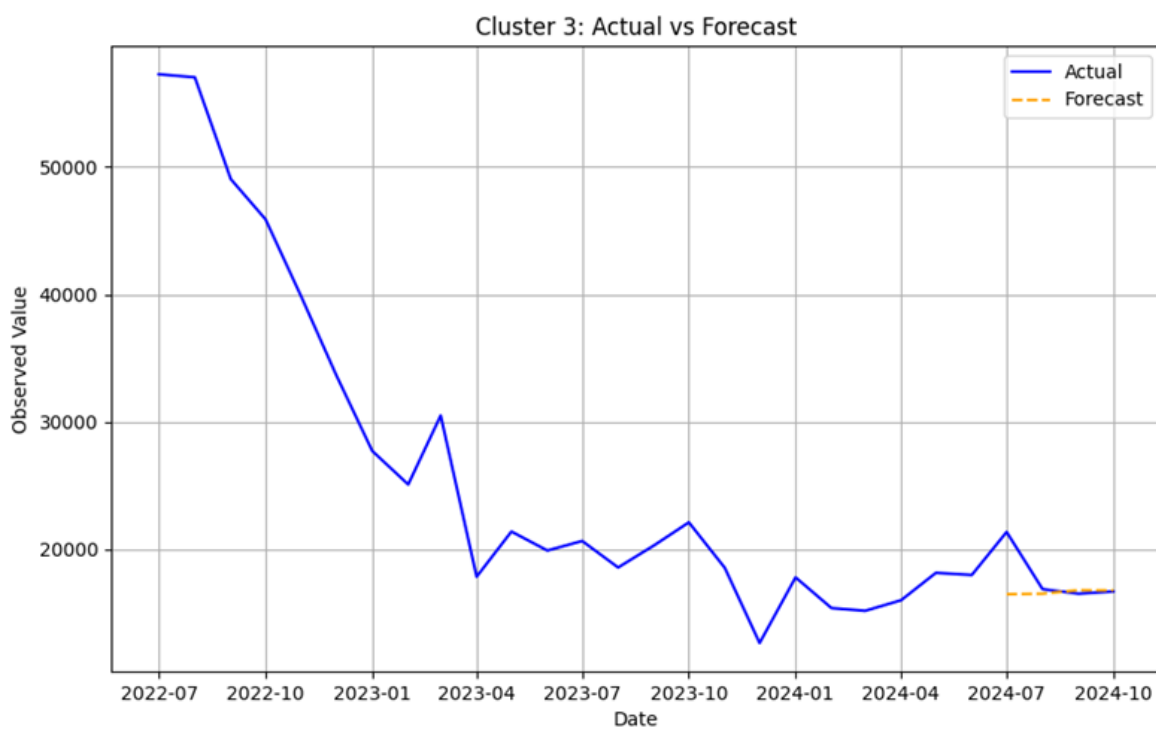


Рисунок 5.13– Тестування даних кластеру 3

Було прораховано оцінки точності для кожного кластера (таблиця 5.10). Кластери 0, 1 і 2 мають хороший результат (MAPE від 5-10%), а кластер 1 має задовільний результат (MAPE від 10-20%).

Таблиця 5.10 – Оцінка точності моделі за кластерами

Cluster	MAE	MSE	MAPE
Cluster 0	645.1002497554223	965518.2069480058	5.16%
Cluster 1	2634.6821835725186	13377728.873122936	8.53%
Cluster 2	833.203704163066	1119123.9882832228	10.94%
Cluster 3	1398.057869898491	5996875.835789702	7.82%

Також моделлю було протестовано кілька окремих країн, та зроблено візуалізації: Німеччина (рисунок 5.14), Польща (рисунок 5.15), Швейцарія (рисунок 5.16) та Ліхтенштейн (рисунок 5.17).

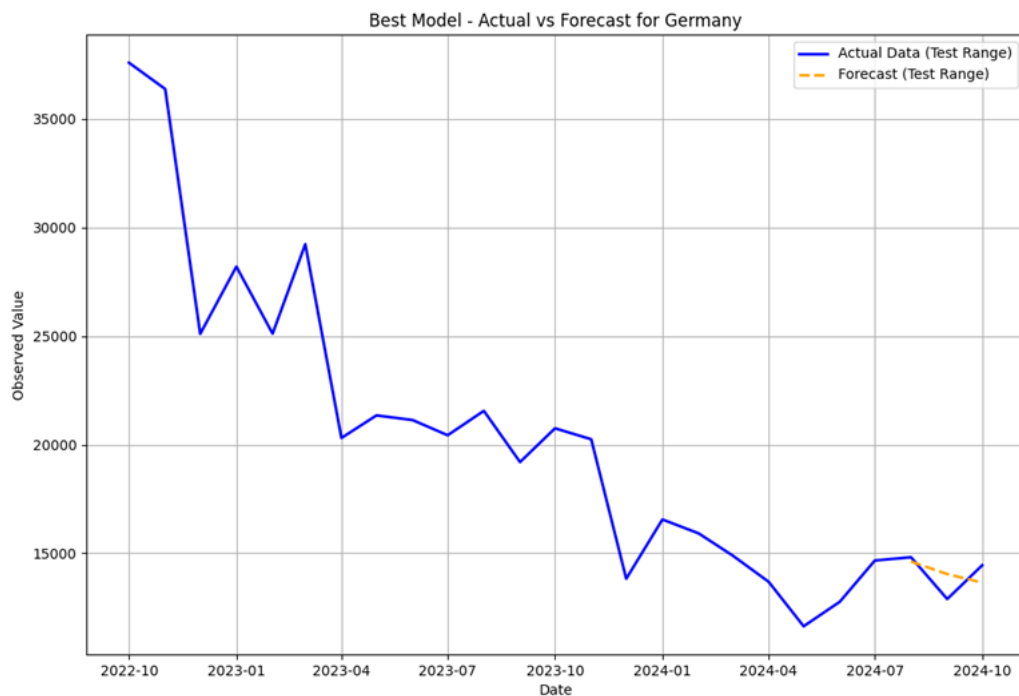


Рисунок 5.14 – Тестування часового ряду Німеччини

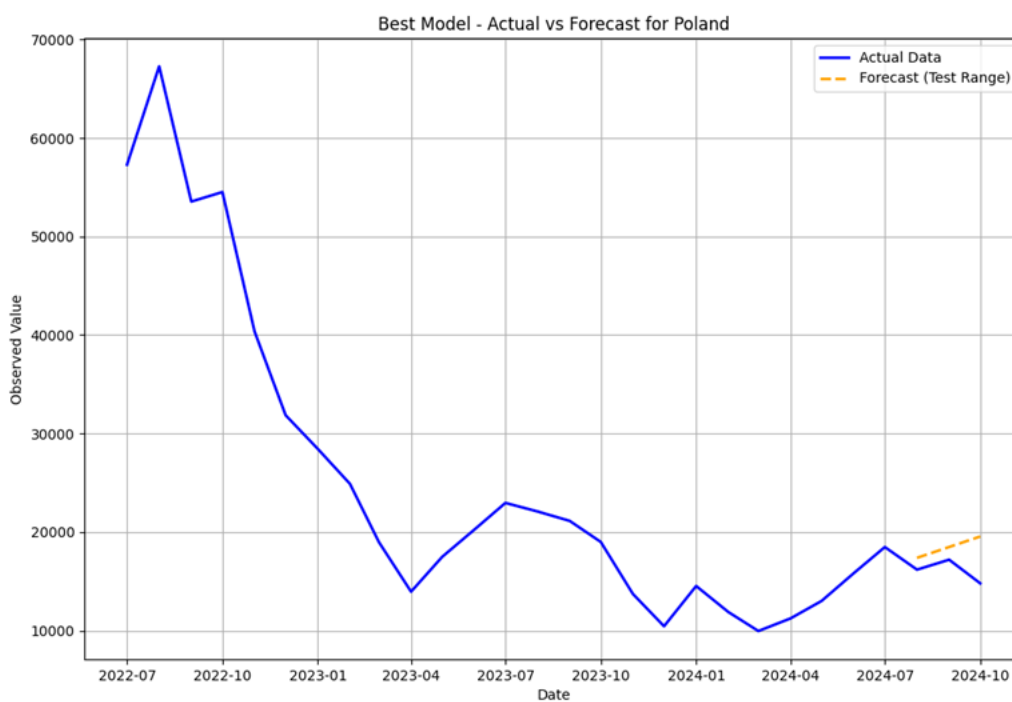


Рисунок 5.15 – Тестування часового ряду Польщі

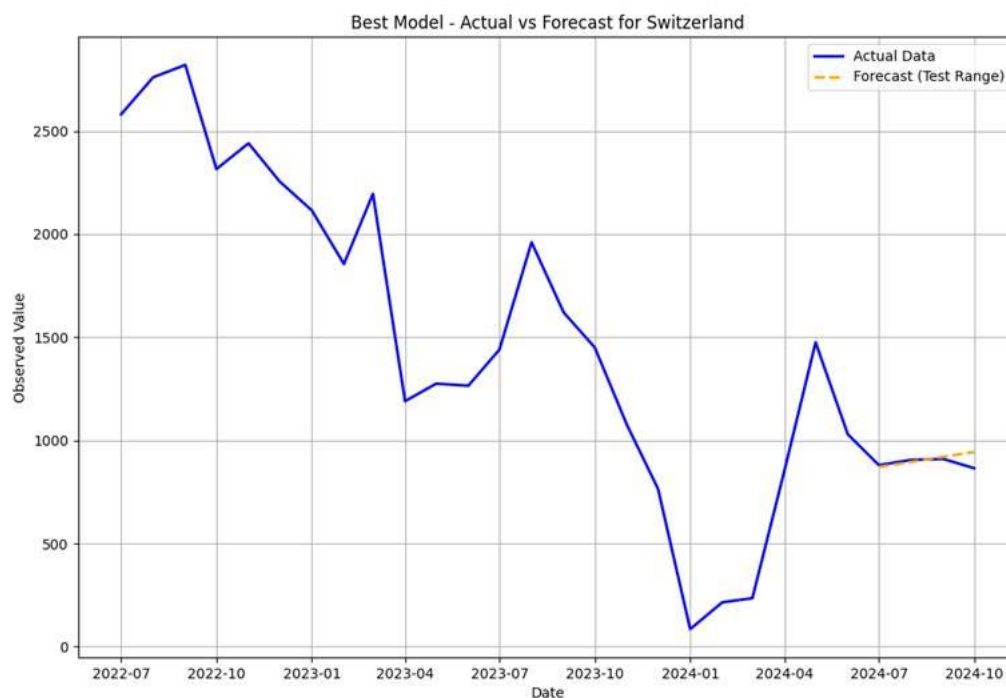


Рисунок 5.16 – Тестування часового ряду Швейцарії

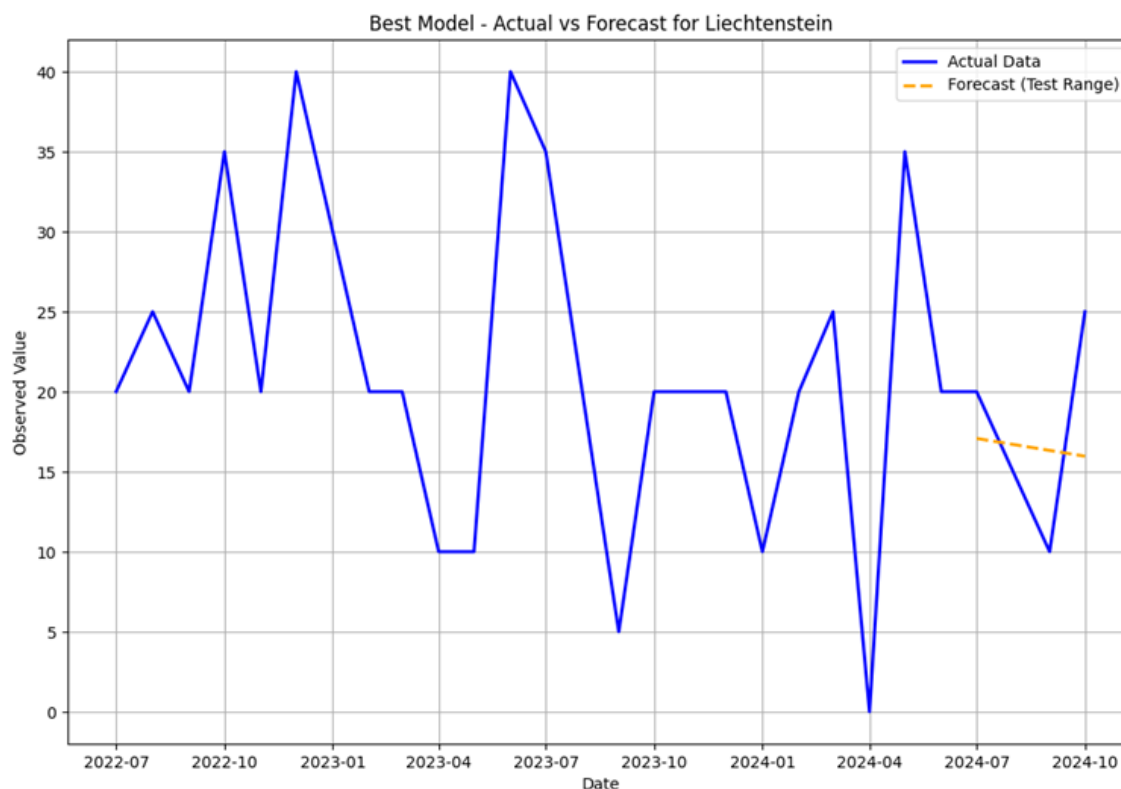


Рисунок 5.17 – Тестування часового ряду Ліхтенштейну

Результати тестування було занесено у таблицю (таблиця 5.11).

Таблиця 5.11 – Оцінка тестування з підбором гіперпараметрів

Country	MAE	MEAN	MSE	MAPE
Germany	716.0954597254507	14050.0	666490.6251963256	5.10%
Poland	1077.8117403117383	16058.33	2963543.1944330838	6.71%
Switzerland	29.715552643210714	890.0	1917.8665431269974	3.34%
Liechtenstein	4.164733894850187	17.5	29.927770648951547	23.80%

Дивлячись на результати можна сказати що для країн з дуже малою кількістю заяв підбір гіперпараметрів не змінює оцінку прогнозу і вона є недостатньою для стабільного прогнозу, в той час як в країнах з високою та середньою кількістю заявок ми можемо побачити суттєве покращення результатів метрик.

5.5 Візуалізація результатів прогнозування

5.5.1 Модель ETS

На основі проведеного тестування ETS-моделі встановлено, що комбінація мультиплікативного тренду та адитивної сезонності забезпечує найкращі результати для прогнозування кількості заявок на тимчасовий захист від громадян України в європейських країнах. Ця комбінація дозволяє більш точно враховувати як загальну тенденцію зростання або зменшення кількості заявок, так і вплив регулярних сезонних коливань, які притаманні даним.

Візуалізація прогнозу агрегованого ряду на 6 місяців вперед (рисунок 5.18).

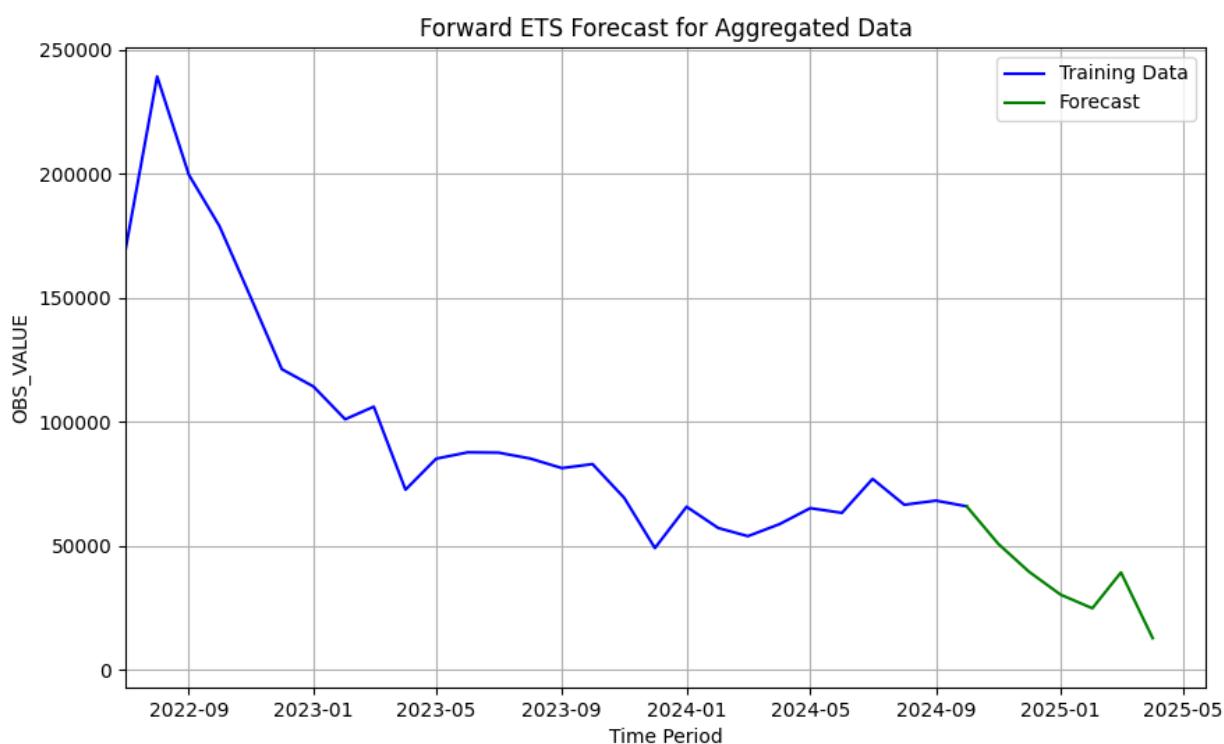


Рисунок 5.18 – Прогнозування агрегованого ряду на 6 місяців

5.5.2 Модель Prophet

Після тестування моделі з підбором гіперпараметрів та без підбору, що показало більш точні результати були саме під час індивідуального підбору параметрів. Цей метод було обрано для прогнозування майбутніх дат, а саме прогнозування на 6 місяців вперед. Результати прогнозування можна спостерігати нижче на графіках: Польща (рисунок 5.19), Німеччина (рисунок 5.21), Швейцарія (рисунок 5.20) та Ліхтенштейн (рисунок 5.22). А також було візуалізовано прогнозування агрегованого ряду (рисунок 5.23).

Середні широти довірчих інтервалів було записано до таблиці (таблиця 5.12) відповідно до кожного прогнозу, а також результати прогнозування кластерів.

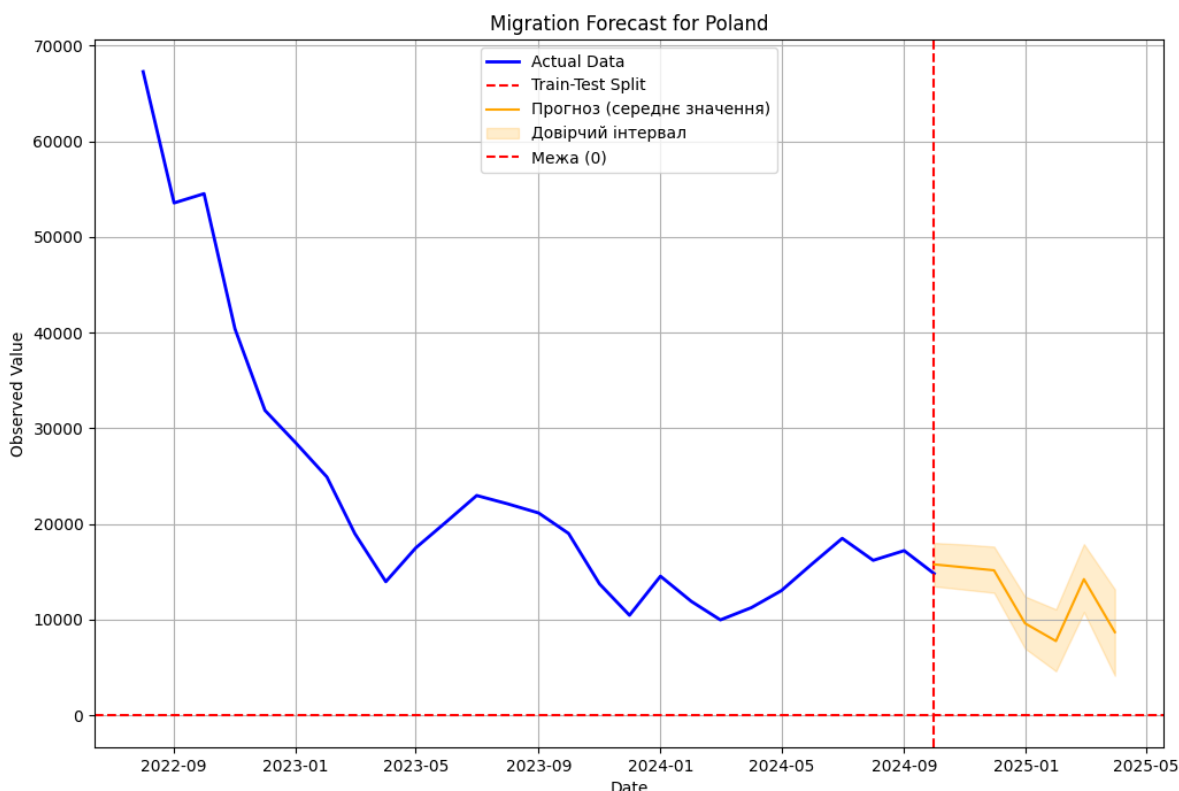


Рисунок 5.19 – Прогнозування на 6 місяців для часового ряду Польщі

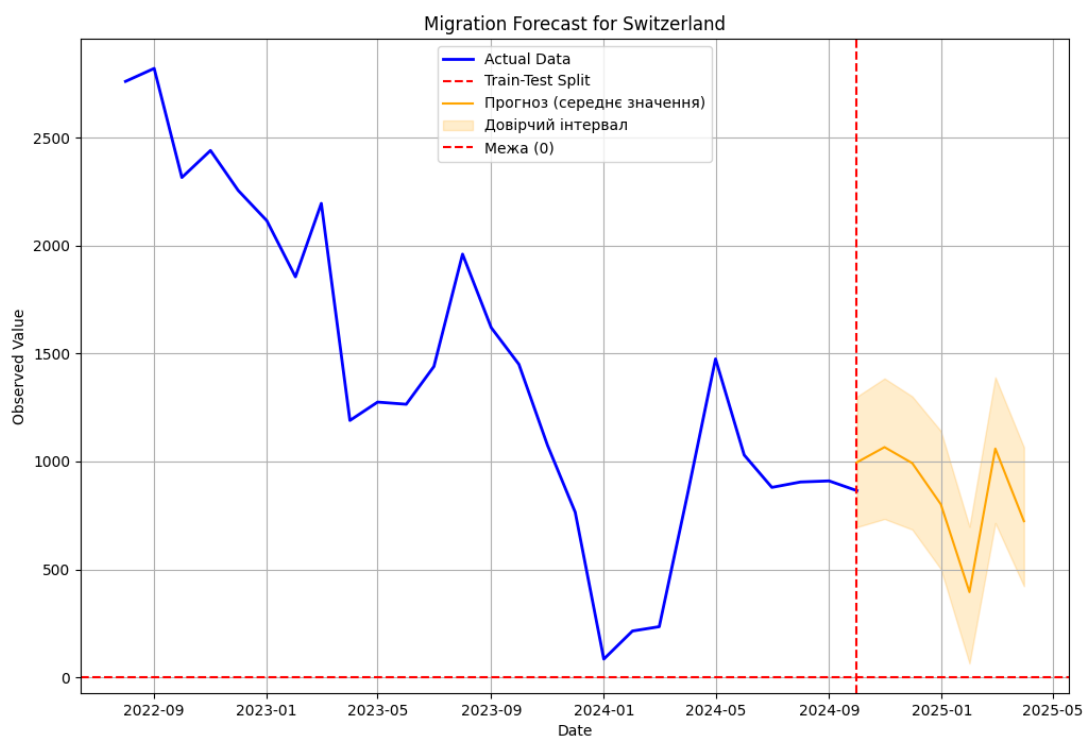


Рисунок 5.20 – Прогнозування на 6 місяців для часового ряду Швейцарії

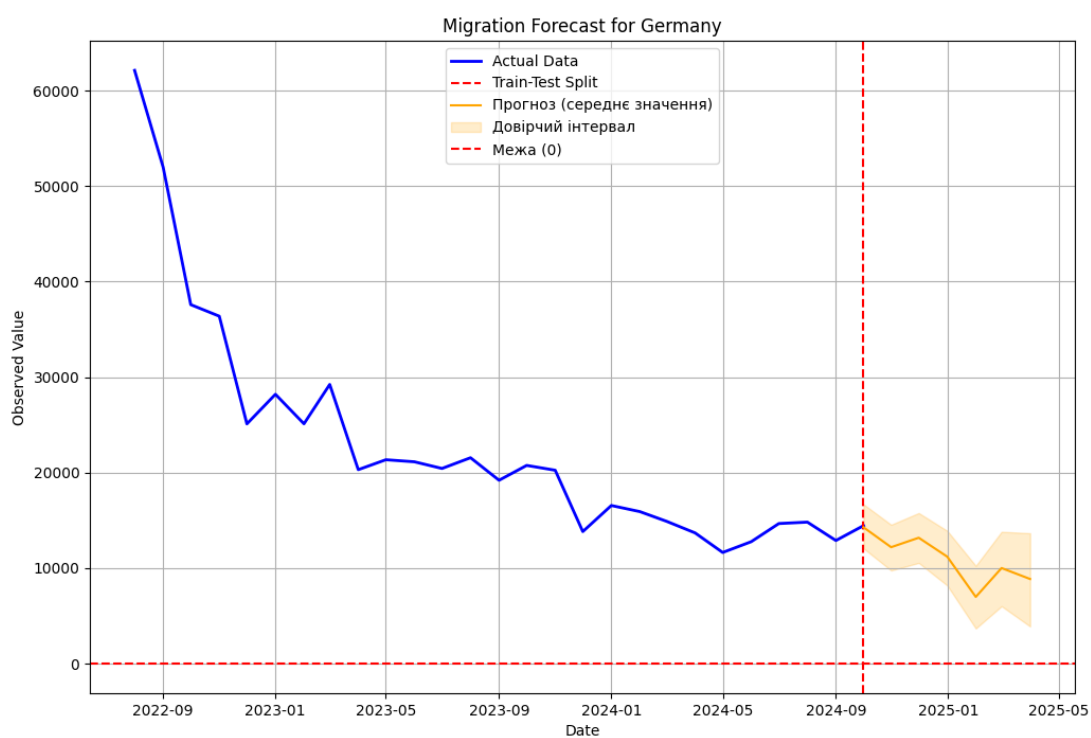


Рисунок 5.21 – Прогнозування на 6 місяців для часового ряду Німеччини

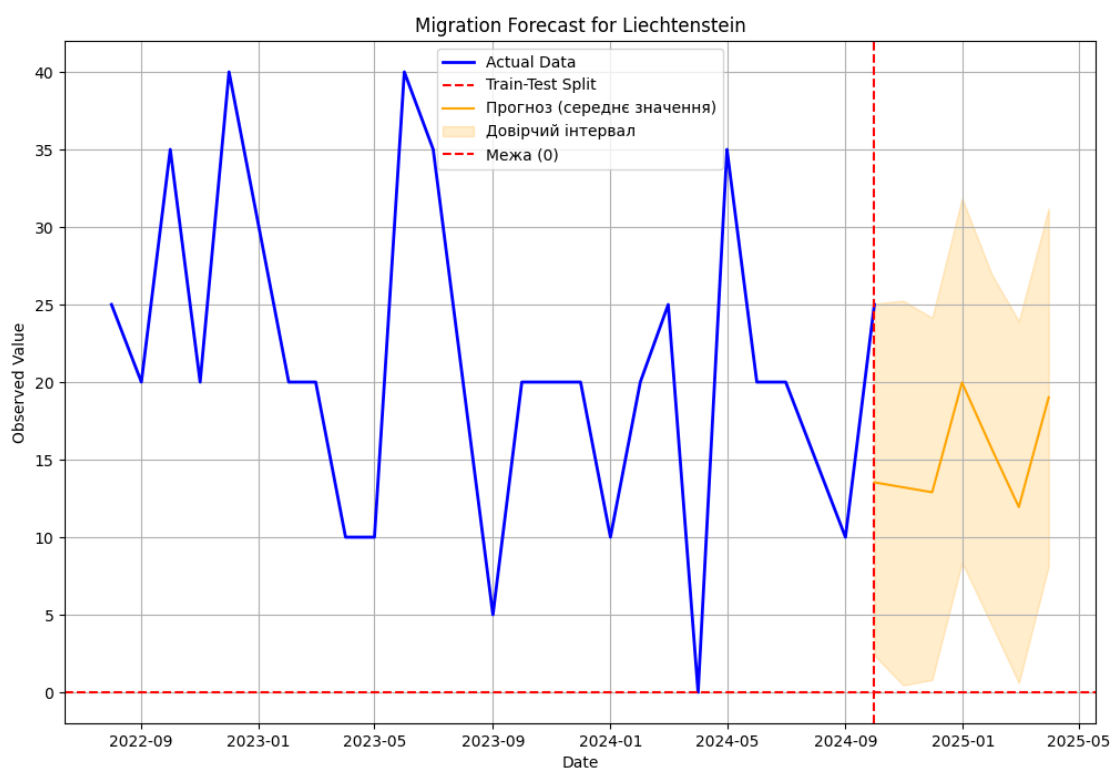


Рисунок 5.22 – Прогнозування на 6 місяців для часового ряду Ліхтенштейн

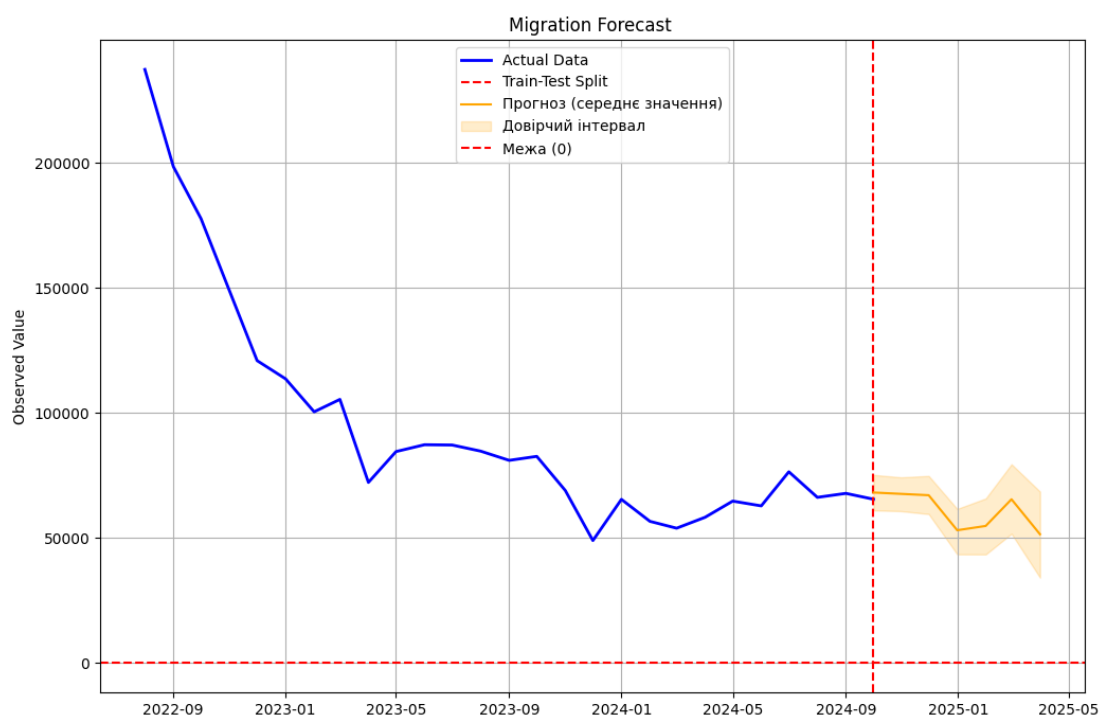


Рисунок 5.23 – Прогнозування на 6 місяців для агрегованого часового ряду

Таблиця 5.12 – Результати середньої ширини довірчого інтервалу прогнозів

Country	Average width of the confidence interval
Aggregated Data	20814.15
Cluster 0	3269.81
Cluster 1	3193.08
Cluster 2	6596.86
Cluster 3	6015.30
Switzerland	637.29
Poland	9551.66
Germany	6356.78
Liechtenstein	23.44

5.5.3 Модель Sarima

Після завершення тестування моделі SARIMA та визначення оптимальних параметрів для кожного часового ряду було виконано прогнозування на періоди 6 (рисунок 5.24) та 12 місяців (рисунок 5.25) уперед.

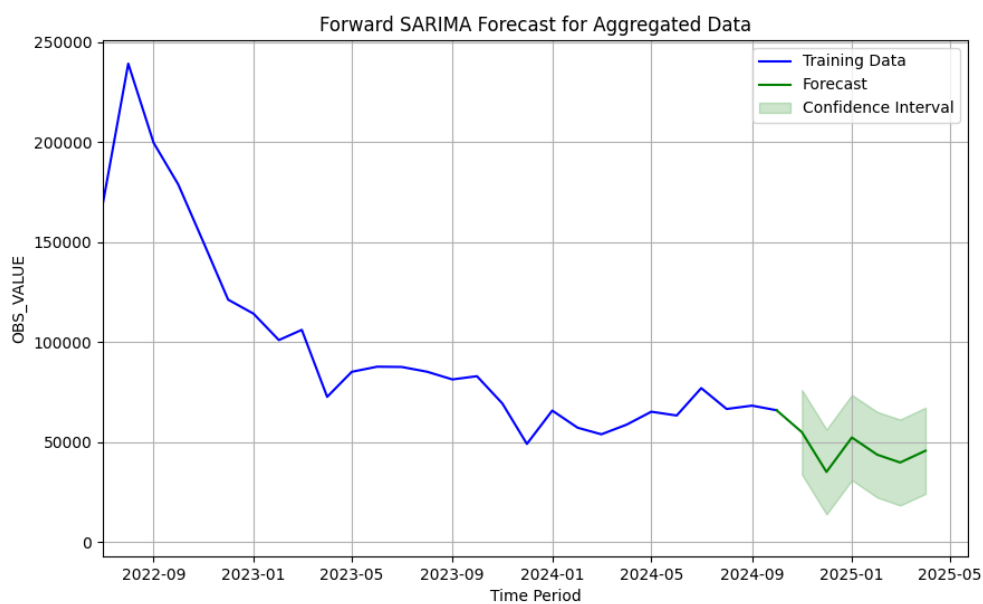


Рисунок 5.24 – прогнозування агрегованого ряду на 6 місяців

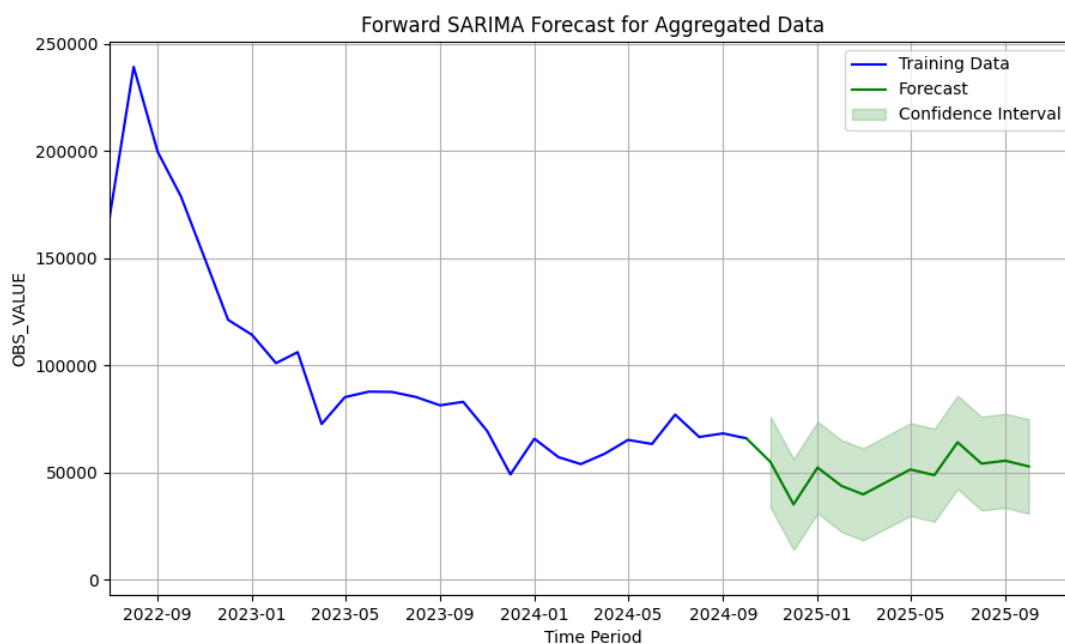


Рисунок 5.25 – прогнозування агрегованого ряду на 12 місяців

5.6 Інтерпретація прогнозів

Прогнозування міграційних потоків українських біженців до європейських країн відобразило кілька ключових трендів, що визначають характер змін у найближчі періоди.

Агреговане прогнозування для всіх країн Європи демонструє поступове зниження загальної кількості заявок на тимчасовий захист. Пікові значення, що спостерігалися на початку війни, поступово згладжуються, що свідчить про стабілізацію потоків. Це відповідає очікуванню, що початкова хвиля міграції через ескалацію конфлікту в Україні вже пройшла, а потоки наразі формуються за більш стабільними темпами.

Сезонні компоненти, виявлені у прогнозах, свідчать про незначні періодичні коливання, які можуть бути пов'язані із соціально-економічними чинниками, такими як міграційна активність у літній період або підготовка до холодної пори року.

Прогнози для кластерів країн відображають різні динаміки залежно від їхніх характеристик:

– кластер 1 (країни з високими обсягами міграції, наприклад, Польща, Німеччина): очікується поступове зниження кількості заявок, але збереження стабільно високого рівня порівняно з іншими країнами. Польща, зокрема, демонструє тренд на спад після початкового піку, що може свідчити про зменшення ролі цієї країни як першочергового пункту притулку;

– кластер 3 (країни зі стабільною низькою активністю): потоки залишатимуться майже незмінними, без значних коливань. Це свідчить про те, що ці країни відіграють роль кінцевих пунктів для довготривалого перебування;

– кластер 0 (віддалені країни з високою вартістю проживання): динаміка прогнозу показує поступове вирівнювання потоків на низькому рівні. Це відображає роль цих країн як менш привабливих для первинної хвилі міграції.

Прогнози для окремих країн, таких як Німеччина, Польща, Швейцарія та Ліхтенштейн, відображають унікальні тенденції:

– Німеччина: прогноз показує стабільно високий рівень заявок із поступовим спадом. Це вказує на тривалу привабливість країни для біженців через економічну стабільність і доступність соціальних програм;

– Польща: тренд демонструє чітке зниження після різкого піку в перші місяці війни, що свідчить про завершення етапу масової міграції до цієї країни;

– Швейцарія: потоки залишаються стабільними на низькому рівні, без суттєвих сезонних коливань. Це вказує на роль країни як пункту довготривалого осідання для певних категорій мігрантів;

– Ліхтенштейн: прогнози демонструють найнижчі значення серед усіх країн, з незначними коливаннями, що відповідає її обмеженим можливостям для прийому біженців.

У дослідженні та прогнозуванні міграційних потоків було виявлено кілька ключових трендів, які стосуються як історичних даних, так і вірогідних майбутніх змін:

- загальне зниження потоків після початкового піку. Найвищі значення міграційної активності спостерігалися у березні-квітні 2022 року. Подальший аналіз виявив, що ці показники поступово зменшуються, а прогноз підтверджує стабілізацію потоків на значно нижчому рівні в найближчі місяці;

- географічна близькість і економічна привабливість як ключові фактори. Потоки демонструють залежність від поєднання чинників, як відстань до України, наявність спільного кордону, рівень ВВП і підтримка для біженців. Країни, які відповідають цим критеріям, продовжують приймати більші обсяги мігрантів;

- зменшення ролі країн-сусідів як первинних напрямків. Польща, яка прийняла найбільшу частину біженців на початку війни, поступово втрачає свою роль як основний пункт прибуття. Це пояснюється як перенасиченням ресурсів, так і стабілізацією потоків;

- різна динаміка у кластерах. Прогноз виявив різні траєкторії для груп країн. Наприклад, для кластеру 1 (країни з високою активністю) характерний тренд до помірною зниження, тоді як кластер 3 (стабільні країни) демонструє стабільність потоків на низькому рівні;

- прогноз на найближчі 6 місяців вказує на продовження загальної тенденції до зменшення кількості заявок у більшості країн. Сезонні коливання залишаються незначними, що свідчить про стабільність у середньостроковій перспективі.

ВИСНОВКИ

Прогнозування міграційних потоків українських біженців виявило важливі соціально-економічні тенденції. Зокрема, очікується поступове зниження кількості заявок на тимчасовий захист у країнах Європи. Це відображає стабілізацію ситуації після пікових міграційних потоків, викликаних початком війни. Економічно розвинені країни, такі як Німеччина та Польща, залишаються привабливими для довготривалої міграції завдяки своїм соціальним програмам і доступу до ринку праці.

Результати роботи підкреслюють необхідність розширення часових рядів та інтеграції додаткових показників, таких як доступність житла, рівень інтеграції та мовний бар'єр та інших демографічних показників.

Обмеження цього дослідження включають недостатню кількість часових точок, що впливає на точність прогнозів і обмежує застосування складніших моделей. Аномальні періоди на початку війни також створюють виклик для адекватного відображення динаміки і відсікання аномальних часових точок робить і ряд ще меншим. Крім того, через обмеженість соціально-економічних показників у даних дослідження деякі важливі аспекти могли залишитися не врахованими. Відсутність змінних, які враховують політичні рішення або регіональні програми підтримки біженців, також може вплинути на точність моделювання.

Отримані результати підтвердили поставлені цілі, включаючи аналіз трендів міграції, оцінку ключових впливових факторів та розробку прогнозів. У перспективі робота сприяє подальшому вдосконаленню інструментів для прогнозування соціальних явищ і підтримки ефективних управлінських рішень у контексті міграції.

Підсумовуючи результати дослідження міграційних трендів, можна зробити висновок, що війна в Україні суттєво вплинула на динаміку міграційних потоків, створивши значний тиск на приймаючі країни Європи. Основні тренди вказують на стабілізацію після початкового піку міграції,

що супроводжувалась переважно короткостроковими потоками до країн-сусідів України. Польща, Німеччина та Чехія стали головними напрямками для біженців завдяки географічній близькості, економічній привабливості та доступним соціальним програмам.

Економічні фактори, такі як ВВП та ІЛР приймаючих країн, значно вплинули на вибір біженцями місць тимчасового захисту. Моделі прогнозування підтвердили, що найближчим часом потоки до більшості країн знижуватимуться, досягаючи стабільного рівня, тоді як країни з меншою доступністю й надалі прийматимуть обмежену кількість мігрантів. Виявлено, що сезонні й географічні фактори продовжують формувати динаміку, але їхній вплив поступово зменшується в міру стабілізації ситуації.

Результати дослідження підтверджують, що поєднання сучасних методів аналізу даних із врахуванням соціально-економічних, демографічних і культурних чинників дозволяє глибше зрозуміти природу міграційних потоків і створити ефективні інструменти для їхнього прогнозування. Отримані прогнози є цінним ресурсом для ідей вдосконалення державної політики, розподілу ресурсів і планування інтеграційних заходів у приймаючих країнах.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Ukraine - migration overview. *Migration data portal*. URL: <https://www.migrationdataportal.org/ukraine/migration-overview> (date of access: 25.11.2024).
2. Сімахова А. О., Церковний І. О. Міграційні процеси в Україні в умовах війни: соціальний аспект. Економіка, управління та адміністрування. 2022. С. 61–64. URL: [https://doi.org/10.26642/ema-2022-4\(102\)-61-64](https://doi.org/10.26642/ema-2022-4(102)-61-64) (дата звернення: 25.11.2024).
3. Willekens F. Migration Flows: Measurement, Analysis and Modeling. *International Handbook of Migration and Population Distribution*. Dordrecht, 2016. P. 225–241. URL: https://doi.org/10.1007/978-94-017-7282-2_11 (date of access: 27.11.2024).
4. Raymer J., Smith P. W. F. Modelling migration flows. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 2010. Vol. 173, no. 4. P. 703–705. URL: <https://doi.org/10.1111/j.1467-985x.2010.00660.x> (date of access: 01.12.2024).
5. Вимушена міграція і війна в Україні (24 березня – 10 червня 2022) | Cedos. *Cedos*. URL: https://cedos.org.ua/researches/vymushena-migracziya-i-vijna-v-ukrayini-24-bereznya-10-chervnya-2022/?utm_source=chatgpt.com (дата звернення: 27.11.2024).
6. Situation Ukraine Refugee Situation. *Operational Data Portal*. URL: <https://data.unhcr.org/en/situations/ukraine> (date of access: 26.11.2024).
7. Hadi A. S., Kaufman L., Rousseeuw P. J. Finding Groups in Data: An Introduction to Cluster Analysis. *Technometrics*. 1992. Vol. 34, no. 1. P. 111. URL: <https://doi.org/10.2307/1269576> (date of access: 15.12.2024).
8. Cluster Analysis / B. S. Everitt et al. Wiley & Sons, Incorporated, John, 2012. (date of access: 17.12.2024).

9. Jain A. K., Murty M. N., Flynn P. J. Data clustering. *ACM Computing Surveys*. 1999. Vol. 31, no. 3. P. 264–323. URL: <https://doi.org/10.1145/331499.331504> (date of access: 18.12.2024).
10. GeeksforGeeks. Time Series Analysis and Forecasting - GeeksforGeeks. *GeeksforGeeks*. URL: <https://www.geeksforgeeks.org/time-series-analysis-and-forecasting/> (date of access: 18.12.2024).
11. Time Series Analysis: Forecasting and Control / G. M. Jenkins et al. Wiley & Sons, Incorporated, John, 2015. 712 p. (date of access: 20.12.2024)
12. Seymour L., Brockwell P. J., Davis R. A. Introduction to Time Series and Forecasting. *Journal of the American Statistical Association*. 1997. Vol. 92, no. 440. P. 1647. URL: <https://doi.org/10.2307/2965440> (date of access: 21.12.2024).
13. Decisions granting temporary protection by citizenship, age and sex - monthly data. *Eurostat*. URL: https://ec.europa.eu/eurostat/databrowser/product/page/MIGR_ASYTPFM (date of access: 26.11.2024).
14. Refugees from Ukraine in Germany 2024 | Statista. *Statista*. URL: <https://www.statista.com/statistics/1333488/refugees-ukraine-germany/> (date of access: 10.12.2024).
15. Прес-релізи та звіти - Рівень володіння англійською та іншими іноземними мовами в Україні: результати кількісного соціологічного дослідження проведеного у грудні 2022 – січні 2023. *Домашня сторінка КМІС*. URL: <https://kiis.com.ua/?lang=ukr&cat=reports&id=1210&page=1> (дата звернення: 15.12.2024).
16. Hodson T. O. Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geoscientific Model Development*. 2022. Vol. 15, no. 14. P. 5481–5487. URL: <https://doi.org/10.5194/gmd-15-5481-2022> (date of access: 27.12.2024).
17. Richardson J. G., Newbury F. D. Business Forecasting: Principles and Practice. *Southern Economic Journal*. 1953. Vol. 19, no. 4. P. 531. URL: <https://doi.org/10.2307/1054108> (date of access: 28.12.2024).

18. Prophet. *Prophet*. URL: <https://facebook.github.io/prophet/> (date of access: 18.12.2024).
19. Taylor S. J., Letham B. Forecasting at Scale. *The American Statistician*. 2018. Vol. 72, no. 1. P. 37–45. URL: <https://doi.org/10.1080/00031305.2017.1380080> (date of access: 18.12.2024).
20. Diagnostics. *Prophet*. URL: <https://facebook.github.io/prophet/docs/diagnostics.html#cross-validation> (date of access: 19.12.2024).
21. Seasonality, Holiday Effects, And Regressors. *Prophet*. URL: https://facebook.github.io/prophet/docs/seasonality_holiday_effects_and_regressors.html#fourier-order-for-seasonalities (date of access: 28.12.2024).