

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ ІМЕНІ
ВАДИМА ГЕТЬМАНА**

Інститут інформаційних технологій в економіці
(назва навчально-наукового інституту / факультету)

Кафедра математичного моделювання та статистики
(назва кафедри)

ЕКОНОМІЧНА КІБЕРНЕТИКА І ДАТА САЙНС

Галузь знань

05 «Соціальні та поведінкові науки»

Спеціальність

051 «Економіка»

Форма навчання: _____
дистанційна
очна (денна), заочна, дистанційна

КВАЛІФІКАЦІЙНА МАГІСТЕРСЬКА РОБОТА

на тему **«Аналіз та прогнозування ціни на нерухомість в США»**
(назва теми)

здобувача _____
Поліщук Ольга Вадимівна
(ПІБ, підпис)

Науковий керівник: кандидат економічних наук, доцент Ішук Я. В.
(науковий ступінь, учене звання, ПІБ)

(підпис)

**Робота допущена до захисту перед екзаменаційною комісією
з атестації здобувачів вищої освіти (ЕК)**

В. о. завідувача кафедри: _____
кандидат фізико-математичних наук,
професор Великоіваненко Г.І.
(науковий ступінь, учене звання, ПІБ)

(підпис)

Київ 2024

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ВАДИМА ГЕТЬМАНА**

**Навчально-науковий інститут
«Інститут інформаційних технологій в економіці»
Кафедра математичного моделювання та статистики**

ОСВІТНЬО-ПРОФЕСІЙНА ПРОГРАМА	ЕКОНОМІЧНА КІБЕРНЕТИКА І ДАТА САЙНС
ГАЛУЗЬ ЗНАНЬ	05 СОЦІАЛЬНІ ТА ПОВЕДІНКОВІ НАУКИ
СПЕЦІАЛЬНІСТЬ	051 ЕКОНОМІКА

ПОГОДЖЕНО

Керівник проектної групи (гарант)
освітньо-професійної програми

О.В. Піскунова

(підпис)

20__ р.

ЗАТВЕРДЖУЮ

В.о. завідувача кафедри
математичного моделювання та
статистики

Г. І. Великоіваненко

(підпис)

20__ р.

ІНДИВІДУАЛЬНЕ ЗАВДАННЯ

**здобувачу вищої освіти Поліщук Ользі Вадимівні
дистанційної форми навчання
на підготовку кваліфікаційної магістерської роботи**

на тему «Аналіз та прогнозування ціни на нерухомість в США»

Тему затверджено наказом ректора Університету від «__» 2023 р. №__

**Кваліфікаційна магістерська робота виконується на матеріалах
взятих з відкритих джерел на сайті Kaggle, National association of realtors та інші
План кваліфікаційної магістерської роботи**

Розділ 1	Теоретико-методичні основи дослідження ціни на нерухомість в США
(назва розділу)	
Розділ 2	Аналіз факторів, що впливають на ціни на нерухомість в США
(назва розділу)	
Розділ 3	Прогнозування цін на нерухомість в США

(назва розділу)	
Об'єкт дослідження:	ринок нерухомості в Сполучених Штатах Америки
Предмет дослідження:	моделі та методи аналізу формування та прогнозування цін на нерухомість
Мета кваліфікаційної магістерської роботи:	проведення комплексного аналізу факторів, які впливають на ціни нерухомості в США та прогнозування середньої ціни на нерухомість в США на наступний період

Конкретні завдання, які здобувач повинен виконати для досягнення поставленої мети:

У розділі 1	Дослідити основні фактори, які діють на формування ціни на ринку нерухомості, провести огляд сучасних тенденцій ринку нерухомості США, окреслити методи економіко-математичного дослідження цін на ринку нерухомості.
У розділі 2	Проаналізувати визначальні фактори формування ціни на ринку нерухомості США на основі регресійного аналізу, розбити сукупність об'єктів нерухомості на однорідні групи зі схожими характеристиками за допомогою кластерного аналізу.
У розділі 3	Визначити моделі, які можуть бути використані для прогнозування цін нерухомості США, провести їх порівняльну характеристику і на основі отриманих даних побудувати прогноз на наступний період.

**Завдання підготував
науковий керівник**

(підпис)

Я. В. Іщук

(ініціали, прізвище)

« ____ » _____ 20__ р.

**Завдання одержав
здобувач**

(підпис)

О. В. Поліщук

(ініціали, прізвище)

« ____ » _____ 20__ р.

Реферат

Кваліфікаційна магістерська робота містити 77 сторінок, 31 рисунок, 2 таблиці, список використаних джерел з 51 найменувань.

«Аналіз та прогнозування ціни на нерухомість в США»

Об'єктом дослідження є ринок нерухомості в Сполучених Штатах Америки.

Предметом дослідження є моделі та методи аналізу, формування та прогнозування цін на нерухомість.

Метою роботи є проведення комплексного аналізу факторів, які впливають на ціни нерухомості в США та прогнозування середньої ціни на нерухомість в США на наступний період.

Досягнення визначеної мети передбачає необхідність вирішення таких завдань:

- проаналізувати стан сучасного ринку нерухомості в США;
- дослідити економічні та інші чинники, що впливають на ціну нерухомості;
- розробити моделі для прогнозування змін у цінах на ринку нерухомості за допомогою різних методів;
- реалізувати побудовані моделі у програмному середовищі та проаналізувати їх точність на основі реальних даних;
- сформулювати висновки на основі створених моделей та отриманих прогнозів.

В роботі присутня *теоретико-методична значущість*, отримані результати прогнозування ціни нерухомості, врахування факторів, що впливають на неї, можуть допомогти більш точно визначати основні тенденції на ринку нерухомості. Також в роботі присутня і *практична значущість*, яка пов'язана з впровадженням методів прогнозування, що допоможе, як і інвесторам, так і продавцям та покупцям нерухомості.

Рік виконання кваліфікаційної магістерської роботи – 2023.

Рік захисту роботи – 2024.

Ключові слова: прогнозування ціни на нерухомість в США, регресійний метод LASSO, модель експоненціального згладжування, кластерний аналіз, регресійний аналіз, авторегресійне інтегроване ковзне середнє з сезонним компонентом (Sarima).

В і д г у к
про кваліфікаційну магістерську роботу
здобувача навчально-наукового інституту
«Інститут інформаційних технологій в економіці»
освітньо-професійної програми «Економічна кібернетика і Дата Сайнс»

Поліщук О. В.

(прізвище, ініціали)

на тему «Аналіз та прогнозування ціни на нерухомість в США»

(назва теми)

1. *Актуальність теми.* Аналіз та прогнозування цін на нерухомість є актуальною темою дослідження, оскільки зміни в економіці, попиті та пропозиції, а також в правовому середовищі та технологічному прогресі продовжують впливати на ринок нерухомості. Врахування цих факторів та використання сучасних методів аналізу дозволяють фахівцям у сфері нерухомості робити точніші прогнози, визначати тенденції та адаптуватися до змін в галузі.
2. *Позитивні риси кваліфікаційної роботи.* Кваліфікаційна магістерська робота демонструє, що здобувач здатний дібрати та опрацювати необхідну наукову літературу, зібрати необхідну для дослідження інформацію, аналізувати її на основі обраних методів дослідження, а також виявляє його навички самостійної роботи з використанням наукових методів та підходів. Робота також вказує на його вміння чітко та логічно викладати результати своїх досліджень.
3. *Наявність самостійних розробок автора.* Автор роботи продемонстрував навички самостійних розробок, дослідивши сучасний стан та перспективи розвитку ринку нерухомості США. Автор врахував широкий спектр факторів, що впливають на зміну цін нерухомості США. Побудована економіко-математична модель є достатньо точною, щоб застосовувати її для прогнозування цін нерухомості.
4. *Цінність теоретичних висновків та практичних рекомендацій.* На основі отриманих результатів, можна зробити висновок, що використання розроблених моделей, де враховані основні фактори впливу зміну цін нерухомості, може буде доцільно, як на національному рівні, так і на рівні окремих штатів та забезпечить стейкхолдерів необхідною інформацією для прийняття стратегічних рішень, що сприятимуть стабільності та ефективності ринку нерухомості.
5. *Наявність недоліків.* Регресійна модель формування ціни під впливом неекономічних факторів має невелику точність. Доцільно було обрати іншу комбінацію факторів для підвищення ефективності моделі.
6. *Загальна оцінка кваліфікаційної магістерської роботи та її допущення до захисту перед ЕК.* Проведений аналіз відповідає в цілому вирішенню поставлених завдань, є окремі недоліки та некоректності в опрацюванні отриманих результатів, текст оформлено з незначними порушеннями вимог оформлення таблиць, рисунків. Робота допущена до захисту.

Науковий керівник

доцент кафедри математичного моделювання та статистики, к.е.н.

Іщук Я.В.

“23” січня 2024 р.

ЗМІСТ

ВСТУП.....	8
РОЗДІЛ 1 ТЕОРЕТИКО-МЕТОДИЧНІ ОСНОВИ ДОСЛІДЖЕННЯ ЦІНИ НА НЕРУХОМІСТЬ В США	11
1.1. Сутність та значення ринку нерухомості.....	11
1.2. Теорії та концепцій, які визначають фактори впливу на ринок нерухомості.....	15
1.3. Економічна ситуація на ринку нерухомості в США.....	18
1.4. Методи дослідження цін на ринку нерухомості.....	24
РОЗДІЛ 2 АНАЛІЗ ФАКТОРІВ, ЩО ВПЛИВАЮТЬ НА ЦІНИ НА НЕРУХОМІСТЬ В США	35
2.1. Опис вибраних даних для аналізу факторів, що впливають на ціну нерухомості.....	35
2.2. Розгляд факторів, які мають найбільший та найменший вплив на ціну квартири в США.....	40
2.3. Кластерний аналіз об'єктів нерухомості США.....	48
2.4. Регресійний аналіз цін нерухомості в США.....	50
РОЗДІЛ 3 ПРОГНОЗУВАННЯ ЦІН НА НЕРУХОМІСТЬ В США.....	57
3.1. Підготовка даних до прогнозування.....	57
3.2. Вибір найкращого методу прогнозування цін на нерухомість в США.....	61
3.3. Прогноз ціни на нерухомість на майбутній період.....	68
ВИСНОВОК.....	71
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	74

ВСТУП

США - це високорозвинена країна, яка пропонує високий рівень життя, тому велика кількість молодого покоління бажає іммігрувати в цю країну, а в зв'язку з війною в нашій країні для жінок це стало більш можливим. Для багатьох людей власне житло є великою справою, не лише у фінансовому плані, а й у всьому життєвому шляху. Однак отримати будинок, який відповідає вашим особистим і фінансовим потребам, нелегко. Одним з головних факторів, який слід враховувати, буде ціна будинку. Сучасний ринок нерухомості в США визначається багатьма факторами, які впливають на його динаміку та стабільність, тому в роботі проаналізовано фактори, які впливають на ціну нерухомості в США, та прогнозовано середню ціну на нерухомість на наступний період.

Дослідження впливу факторів на ціни нерухомості проводяться вже багато років. У цих дослідженнях виділено ряд факторів, які мають найбільший вплив на ціну нерухомості, включаючи економічні фактори (ВВП, інфляція, безробіття, процентні ставки), соціально-демографічні фактори (доходи домогосподарств, демографічні зміни, міграція), розташування нерухомості та інші. Наприклад дослідження, проведене Федеральною резервною системою США у 2023 році, показало, що низькі процентні ставки є одним з основних факторів, що сприяють зростанню цін на нерухомість в США. Згідно з дослідженням, при зниженні процентних ставок на 1% ціни на нерухомість зростають у середньому на 2-3%. Це пояснюється тим, що низькі процентні ставки роблять житло більш доступним для покупки, що збільшує попит на нерухомість [1].

Дослідження наукової літератури показало, що питання ціноутворення на ринку нерухомості досліджували такі вчені: А. Асаула, С.В. Грибовський, А. М. Іванченко, Е. Божко, М. В. Кірносорова, М. М. Білаш, Джозеф К. Еккерт, М. А. Федотова, Л. А. Різва, К. В. Загребельна та інші. Однак в ході аналізу було виявлено, що питання економіко-математичного моделювання ціноутворення в житловому секторі в США є не достатньо дослідженим.

Метою магістерської роботи є проведення комплексного аналізу факторів, які впливають на ціни нерухомості в США та прогнозування середньої ціни на нерухомість в США на наступний період. Для досягнення цієї мети поставлені наступні *завдання*:

- проаналізувати стан сучасного ринку нерухомості в США;
- дослідити економічні та інші чинники, що впливають на ціну нерухомості;
- розробити моделі для прогнозування змін у цінах на ринку нерухомості за допомогою різних методів;
- реалізувати побудовані моделі у програмному середовищі та проаналізувати їх точність на основі реальних даних;
- сформулювати висновки на основі створених моделей та отриманих прогнозів.

Об'єктом дослідження є ринок нерухомості в Сполучених Штатах Америки.

Предметом дослідження є моделі та методи аналізу формування та прогнозування цін на нерухомість.

У роботі використовуються різноманітні *методи дослідження*: метод статистичних показників, статистичних спостережень, табличний та графічний методи, а також методи економіко-математичного моделювання.

Робота має *теоретико-методична значущість*, отримані результати прогнозування ціни нерухомості, врахування факторів, що впливають на неї, можуть допомогти більш точно визначати основні тенденції на ринку нерухомості. Також в роботі має і практична значущість, яка пов'язана з впровадженням методів прогнозування. Результати роботи мають значення для звичайних людей, які замислюються щодо імміграції в США та бажають оцінити ризики переїзду, ціни на покупку житла, а також прикладне значення для професіоналів у галузі нерухомості та економіки, дозволяючи їм зробити більш обґрунтовані рішення на ринку нерухомості в США.

Дослідження базується на аналізі наукових праць вітчизняних та зарубіжних науковців і статистичних даних офіційних джерел.

Робота складається зі вступу, трьох розділів, висновків та списку використаних джерел. Кожен розділ відповідає конкретному завданню та включає аналітичний та дослідницький матеріал, спрямований на досягнення поставленої мети роботи.

Робота виконана в середовищі Jupiter notebook на мові програмування Python.

РОЗДІЛ 1

ТЕОРЕТИКО-МЕТОДИЧНІ ОСНОВИ ДОСЛІДЖЕННЯ ЦІНИ НА НЕРУХОМІСТЬ В США

1.1. Сутність та значення ринку нерухомості

Ринок нерухомості — це економічне просторове середовище, де відбувається оренда, купівля, продаж та інші операції з нерухомістю. Нерухомість охоплює житлові будівлі, комерційні об'єкти, земельні ділянки та інші види власності. Ринок нерухомості включає в себе різні сегменти, такі як ринок житла, офісних приміщень, земельних ділянок, торгівельних площ та інші.

Ринок нерухомості є важливим для будь-якої національної економіки, оскільки нерухомість є значною частиною національного багатства, на яку припадає понад 50% світового багатства. Без ринку нерухомості взагалі не було б ринку, тому що ринки праці, капіталу, товарів, послуг тощо мають володіти або орендувати місця, необхідні для свого існування [2]:

- ринок нерухомості забезпечує житлові та комерційні простори для населення та підприємств. Це важливий елемент для задоволення житлових та бізнес-потреб;
- ринок нерухомості привертає інвесторів, які розглядають нерухомість, як інвестиційний об'єкт. Інвестиції в нерухомість можуть приносити прибуток через зростання її вартості та орендні доходи;
- розвиток ринку нерухомості сприяє економічному розвитку через будівництво та створення робочих місць, це може включати будівництво нових житлових комплексів, комерційних центрів та інших об'єктів;

- нерухомість може слугувати як забезпечення для кредитів та іпотеки. Банки та фінансові установи використовують нерухомість як засіб забезпечення кредитів та інших фінансових операцій;
- ринок нерухомості може впливати на стабільність економіки. Зміни в цінах на нерухомість можуть вказувати на стан економіки та впливати на споживчу активність;
- ринок нерухомості є важливим фактором у формуванні та реалізації містобудівної політики. Уряд та місцеві органи використовують регулювання ринку для забезпечення раціонального використання території та сталого розвитку;
- ринок нерухомості може визначати доступність житла для різних соціальних груп. Питання доступності та соціальної справедливості у власності нерухомості важливі для суспільства.

Ринок нерухомості є динамічним та складним сегментом економіки, який взаємодіє з багатьма іншими секторами. Його стан і функціонування мають значущий вплив на економічну активність та якість життя громадян.

Ринок нерухомості складається з таких елементів, як об'єкти нерухомості (земля, будівлі, споруди, інші об'єкти, пов'язані з ґрунтом, а також права на них), суб'єкти ринку (особи, які беруть участь у ринкових відносинах з нерухомістю), спосіб обміну (купівля-продаж, оренда, будівництво та інші форми обігу нерухомості), ціна (грошовий вираз вартості об'єкта нерухомості) [3].

Ринок нерухомості має ряд особливостей, що відрізняють його від інших ринків:

- нерухомість є обмеженим ресурсом: кількість землі на земній кулі є незмінною, а кількість будівель і споруд обмежена їхньою вартістю і терміном служби;
- нерухомість є локалізованим ресурсом: нерухомість має певне місце розташування, яке впливає на її вартість;

- нерухомість є складним продуктом: нерухомість є комплексним продуктом, який складається з землі, будівлі, споруд, інженерних мереж та іншого обладнання;
- нерухомість є дорогим продуктом: вартість нерухомості може бути дуже високою, що обмежує її доступність для більшості населення;
- нерухомість має сезонні коливання цін: весною та літом погода сприяє активнішому огляду, переїзду та покупці нерухомості. Люди відчують себе більш зручно під час огляду будинків та квартир, коли переїжджають за умов теплої погоди. Також сім'ї з дітьми, наприклад, шукають нерухомість переважно в літній період часу, щоб була можливість перевести дитину з початку нового навчального року в інший садочок чи школу.

На формування ціни нерухомості впливають різні фактори і їх сутність та значення можуть варіюватися залежно від контексту ринку нерухомості та економічних умов [4].

Значення цін на нерухомість можна розглядати з різних точок зору.

Економічний аспект: ціни на нерухомість визначаються взаємодією попиту та пропозиції. Якщо попит перевищує пропозицію, ціни можуть зростати і навпаки. Також грає роль економічна ситуація у країні або регіоні, оскільки вона може впливати на рівень інвестицій та здатність людей купувати нерухомість.

Соціокультурний аспект: наприклад, зміни в населенні, розподілі населення за різними категоріями, а також зміни в суспільних уподобаннях можуть впливати на попит конкретного виду нерухомості та їх ціни.

Правовий аспект: законодавство, яке регулює купівлю, продаж та власність нерухомості, може впливати на ціни і стабільність ринку.

Географічний аспект: нерухомість в центрі міста, зазвичай, коштує дорожче через зручний доступ до інфраструктури, бізнес-центрів та розваг. Нерухомість в менш розвинених районах може бути дешевшою, але це може змінитися з розвитком інфраструктури та зростанням попиту. Також ціна залежить від міста,

наприклад, якщо це велике місто, як Київ, то ціна на нерухомість буде значно більшою ніж наприклад в якомусь селі чи маленькому містечку.

Будівельний та технічний аспект: стан будинку, його розмір, матеріали, використані при будівництві, інженерні системи - всі ці фактори можуть впливати на ціну нерухомості. Звичайно, що ціна на квартиру в новобудові, де використані якісні матеріали буде дорожчою ніж ціна квартири в старому будинку, яка буде потребувати можливого ремонту.

Кількість кімнат та площа: однокімнатні квартири з маленькою квадратурою, в більшості випадків, коштують менше і можуть бути популярними серед одиноких людей чи молодих пар. Багатокімнатні квартири, зазвичай, дорожчі і привабливі для сімей або тих, хто шукає більше простору.

Інфраструктура та середовище: розвиток інфраструктури (транспорту, комунікацій, освіти) може підвищити ціну нерухомості в певних районах. Стан навколишнього середовища та екологічні аспекти також можуть впливати на ціни.

Інвестиційний аспект: інвестори можуть оцінювати перспективи зростання вартості нерухомості в майбутньому, що впливає на їх рішення щодо інвестування [5].

Ціни на нерухомість мають велике значення для економіки та життя людей, оскільки вони визначають доступність житла, впливають на рівень життя та є важливим елементом фінансового плану багатьох осіб. Окрім того, вони служать індикатором стану ринку та економіки загалом.

Вплив цін на нерухомість для економіки країни є значним. Ціни мають вплив на такі показники, як інфляція, ВВП, зайнятість та зовнішній баланс.

Ціни на нерухомість є фактором інфляції, якщо вони ростуть швидше, ніж зростає реальний ВВП. Це відбувається тому, що ціни на нерухомість є одним із компонентів споживчої кошика, яка використовується для розрахунку інфляції.

Ціни на нерухомість також можуть впливати на ВВП, якщо вони пов'язані з інвестиціями. Наприклад, якщо ціни на нерухомість ростуть, це може стимулювати інвестиції в будівництво нових будинків та квартир.

Ціни на нерухомість також можуть впливати на зайнятість, якщо вони пов'язані з будівельним сектором. Наприклад, якщо ціни на нерухомість ростуть, це може призвести до збільшення попиту на робочі місця в будівельному секторі.

Нарешті, ціни на нерухомість можуть впливати на зовнішній баланс країни, якщо вони пов'язані з іноземними інвестиціями. Наприклад, якщо ціни на нерухомість ростуть в одній країні, це може призвести до збільшення припливу іноземних інвестицій в цю країну[6].

1.2. Теорії та концепції, які визначають фактори впливу на ринок нерухомості

У вивченні ринку нерухомості ключовою метою є розуміння та аналіз факторів, які визначають його динаміку та поведінку. Цей підрозділ роботи розглядає ряд теоретичних підходів та концепцій, які допомагають розкрити основні чинники, що впливають на ринок нерухомості.

Існує кілька теорій, які пояснюють вплив факторів на ринок нерухомості.

1. Класична теорія пропозиції та попиту

Альфред Маршалл (1842-1924) - британський економіст, який вважається одним із засновників неокласичної школи в економічній теорії. Його робота "Принципи економіки" (1890) є одним із найважливіших творів економічної теорії.

У своїй праці Маршалл розробив класичну теорію пропозиції та попиту, яка є основою сучасної економічної теорії. Ця теорія стверджує, що ціна товару визначається рівновагою між попитом і пропозицією.

Класична економічна теорія пропозиції та попиту є основою для розуміння динаміки ціни на ринку нерухомості. Згідно з цією теорією, ціна визначається взаємодією пропозиції та попиту. Фактори, такі як зміна доходів населення, рівень банківських ставок та регіональні економічні показники, впливають на обсяг попиту та пропозиції на ринку нерухомості [7].

2. Теорія циклів нерухомості

Теорія циклів нерухомості або теорія циклу нерухомості є складною, і вона не пов'язана з конкретним автором або однією конкретною школою економічної думки. Замість того, це більше концептуальний підхід до розуміння того, як ринок нерухомості може проходити через періоди зростання та спаду.

Теорія циклів нерухомості базується на ідеї, що ринок нерухомості може демонструвати циклічні коливання через різні етапи економічного циклу. Наприклад, під час економічного зростання може збільшуватися попит на нерухомість, що веде до підвищення цін і розвитку нових проектів будівництва. З іншого боку, під час економічного спаду може зменшуватися попит і ринок нерухомості може стикатися з проблемами, такими як зниження цін і недофінансовані будівельні проекти.

Ці цикли можуть бути пов'язані з економічними циклами, змінами відсоткових ставок, а також змінами в поведінці споживачів. Розуміння цих циклів дозволяє прогнозувати та адаптуватися до змін на ринку нерухомості [8].

3. Теорія місцевого розвитку та планування

Теорія місцевого розвитку та планування представляє собою розгляд питань організації та розвитку простору на місцевому рівні. Ця теорія може бути пов'язана з роботами багатьох авторів, і багато експертів внесли свій внесок у цей напрям.

Один із важливих авторів, пов'язаних із теорією місцевого розвитку та планування, - Джейн Джейкобс. У своїй книзі "Смерть і життя великих американських міст" ("The Death and Life of Great American Cities") [9], вона висловила концепції щодо розвитку місцевості, зокрема, важливість збереження та розвитку різноманітних та живих міських спільнот. Її підхід акцентує на потребі активного участі місцевого населення в процесах прийняття рішень та плануванні.

Додаткові внески у теорію місцевого розвитку та планування внесли також інші дослідники, такі як Кевін Лінч, Пітер Голдман, та інші, які розвивали концепції сталого розвитку та інші аспекти місцевого планування.

Теорія місцевого розвитку вказує на те, що фактори, такі як зміни в інфраструктурі, розвиток нових об'єктів соціального призначення та зон

планування, впливають на цінову динаміку та попит на нерухомість в конкретному регіоні. Планування міст та регулювання використання землі є ключовими аспектами цієї концепції.

4. Теорія фінансової інвестиції в нерухомість

Теорія фінансової інвестиції в нерухомість має численних авторів, і розвиток цієї теорії включає в себе внесок економістів, фінансистів, та фахівців з нерухомості. Один із важливих внесків в цю область був зроблений у 1960-70-х роках.

Одним із ключових авторів теорії фінансової інвестиції в нерухомість є Джек Сірез. У своїй книзі "Інвестиції в нерухомість та управління портфелем" ("Real Estate Investment and Portfolio Management") [10], Сірез розглядав питання відділення нерухомості як активу та вивчав стратегії інвестування в цей вид активів.

Зараз теорія фінансової інвестиції в нерухомість постійно розвивається, і нові дослідження та публікації розширюють розуміння того, як керувати та інвестувати в нерухомість з фінансової точки зору.

Ринок нерухомості великою мірою визначається інвестиційним попитом. Інвестори розглядають нерухомість як засіб диверсифікації портфеля та забезпечення прибутковості. Зміни в інвестиційних стратегіях та фінансових ринкових умовах впливають на забезпеченість та вартість нерухомості.

5. Теорія соціокультурних систем

Теорія соціокультурних факторів у контексті ринку нерухомості включає різноманітні внески вчених з різних галузей, і немає конкретного автора, який би її однозначно формулював. Замість того, ця теорія виникає з розуміння впливу соціокультурних та демографічних факторів на попит та пропозицію на ринку нерухомості.

Деякі з основоположників концепції соціокультурних факторів у нерухомості включають економістів, соціологів та географів. Дослідження з демографії, зокрема, може допомагати визначити тенденції населення, такі як зміни

в розмірі сім'ї, структурі населення, міграційні потоки та інші аспекти, які впливають на попит на житло та комерційну нерухомість.

Соціокультурні фактори також включають в себе аналіз соціальних та культурних тенденцій, які можуть впливати на вибір місця проживання, тип житла та споживчі уподобання.

Узагальнити автора теорії соціокультурних факторів у нерухомості може бути складно, оскільки ця концепція є результатом комплексного аналізу різних досліджень та дисциплін.

Ця концепція розглядає вплив соціальних та культурних факторів на ринок нерухомості. Зміни в демографічних характеристиках, смаках та уподобаннях споживачів, а також соціокультурні тенденції, можуть впливати на попит на ринку та формування цін.

Розглянуті теорії та концепції надають комплексне уявлення про фактори, які визначають ринок нерухомості. Теорії, які пояснюють вплив факторів на ринок нерухомості, допомагають зрозуміти, як ці фактори впливають на ціноутворення на ринку нерухомості. Розуміння цих аспектів дозволяє ефективно прогнозувати зміни, а також розробляти стратегії для управління нерухомістю в умовах постійної динаміки ринку [11].

1.3. Економічна ситуація на ринку нерухомості в США

США - це одна з найбільш розвинених країн у світі. Мабуть, саме тому ця країна є такою привабливою для іммігрантів.

У III кварталі 2023 року економіка США продемонструвала швидше зростання, ніж прогнозувалось, а от ВВП за цей період збільшився до 5,2% у річному обчисленні. Це пов'язано з різким збільшенням кількості складів та збільшення кількості машинного обладнання, а також включаючи сильний попит на товари та послуги, що спричинило зростання інвестицій. Щодо доходів країни,

експерти говорять, що вони зростають у повільнішому темпі ніж ВВП. Таке зростання ВВП приваблює до США більшу кількість інвесторів, які вкладаються у розбудову медичних центрів та бізнес-інфраструктури. Інвестиції в житлове будівництво також зросли завдяки проектам односімейних будинків, що допомогло покласти край скороченню в цій сфері протягом дев'яти кварталів [12].

У 2023 році США прогнозували невелику рецесію, що була спричинена збільшенням показника інфляції. У травні 2023 року річна інфляція була найвищою за останні 40 років і становила 8,5%. Це зростання інфляції було зумовлене декількома факторами, включаючи високі ціни на енергоносії та продовольство, а також зриви в глобальних ланцюжках поставок.

Щоб позбутись такого великого рівня інфляції та стримати підвищення цього рівня Федеральна резервна система США збільшила процентні ставки, на 2023 рік вони становили 1,75%, що є найвищим показником з 2008 року.

У червні 2023 року рівень безробіття становив 3,6% це є найнижчим показником за останні 70 років. Падіння рівня безробіття було зумовлене тим, що почали з'являтися нові робочі місця, та ринок праці став сильнішим.

Середній дохід домогосподарств через посилення інфляції та збільшення заробітної плати також зріс і був найвищим показником за останні 60 років.

Нетто міграції в США також зросло за останній рік, це зумовлено політичними проблемами в інших країнах. На фоні війни в Україні сталась криза у багатьох країнах Європи, тому велика кількість українців виїхало до США, оскільки США надає гарну програму для підтримки українців, а також велика кількість європейців, також вирішили іммігрувати до США, оскільки ця країна їм здається більш стабільнішою в плані економіки.

Зростання будівництва в США, зростання цін на нерухомість та збільшення процентних ставок зумовило зростання пропозиції та зниження попиту на житло, оскільки не всі іммігранти можуть собі дозволити купівлю нерухомості в США [13].

На основі цієї інформації ми можемо зробити висновок, що економіка США знаходиться в стані переходу, оскільки з одного боку ми бачимо високі темпи

зростання ВВП, зниження рівня безробіття, тобто економіка зростає і США має велику кількість робочих місць. А з іншого боку ми бачимо зростання інфляції та збільшення процентної ставки ФРС для того щоб стримати цю інфляцію.

За даними МВФ, на відміну від інших країн світу, американська економіка, відновила до рівня який існував до пандемії, яка почалась в 2019 році, та продовжує рости [14].

У 2008 році ринок нерухомості в США зазнав одну з найсерйозніших фінансових криз у сучасній історії. Вона почалася з краху ринку субстандартної іпотеки, який призвів до банкрутства декількох великих фінансових установ і дефолту на ринку облігацій східної частини США.

Криза на ринку нерухомості в США була викликана низкою факторів, зокрема:

- зростання попиту на житло: у період з 2000 по 2008 рік ціни на житло в США зросли на 120%. Це було спричинено низькими відсотковими ставками, зростанням багатства домогосподарств і попитом на житло з боку інвесторів;
- легкість отримання іпотеки: кредитні установи почали активно кредитувати навіть позичальників з низьким кредитним рейтингом. Це призвело до збільшення кількості субстандартних іпотек, які були більш ризикованими для кредиторів;
- поява нових фінансових інструментів: кредитні установи почали створювати нові фінансові інструменти, засновані на субстандартних іпотеках. Ці інструменти були продані інвесторам по всьому світу.

У 2007 році почалося зростання кількості прострочених платежів за субстандартними іпотеками. Це призвело до падіння цін на житло і дефолту на ринку облігацій східної частини США. У вересні 2008 року банкрутство американського інвестиційного банку Lehman Brothers стало каталізатором глобальної фінансової кризи.

Криза 2008 року мала серйозні наслідки для економіки США і всього світу. Вона призвела до падіння ВВП, зростання безробіття і зниження життєвого рівня [15].

Після кризи 2008 року ринок нерухомості в США відновився. Ціни на житло почали зростати і до 2018 року досягли рівня до кризи. Однак, у період пандемії COVID-19 ринок нерухомості в США знову пережив період стрімкого зростання цін. Серед факторів, які сприяли зростанню цін є перехід на дистанційну роботу та низькі відсоткові ставки.

Багато людей, які працювали в офісі, перейшли на дистанційну роботу під час пандемії, що призвело до підвищення попиту на житло в районах із сприятливими умовами для роботи вдома, наприклад, у сільській місцевості та в містах із низькими цінами на житло.

Федеральна резервна система США знизилася відсоткові ставки до рекордно низького рівня в період пандемії. Це зробило іпотеки більш доступними для покупців, що також сприяло зростанню попиту на житло.

Ці фактори призвели до того, що ціни на житло в США зросли на 20% у період з 2020 по 2022 рік. Це було найбільшим зростанням цін на житло за 30 років. Однак, у 2023 році ринок нерухомості США почав втрачати динаміку. Це було викликано низкою факторів, зокрема стан ринку нерухомості в США наразі можна охарактеризувати як такий, що перебуває в процесі нормалізації. Після різкого зростання цін на житло в період пандемії, у 2022 році спостерігалось їхнє поступове зниження. За даними Національної асоціації ріелторів (NAR), у листопаді 2022 року середня ціна на однокімнатний будинок у США становила 425 000 доларів, що на 5% нижче, ніж у листопаді 2021 року.

На даний момент, стан ринку нерухомості в США визначався кількома ключовими тенденціями:

- ціни на нерухомість (рис 1.1): багато регіонів США спостерігають стабільні або зростаючі ціни на нерухомість, за останні роки було відзначено зростання цін, особливо в міських областях та популярних локаціях;

- низький рівень вільних вакансій: наразі в США низький рівень вільних вакансій. Великий попит на житло, обмежена пропозиція та низький рівень вільних вакансій у багатьох регіонах призводять до конкуренції між покупцями та підвищують ціни;
- зростання процентних ставок: у 2021 році було відзначено зростання процентних ставок за іпотечними кредитами. Це вплине на спроможність покупців отримувати іпотечні позики та впливати на попит;
- дефіцит житла в деяких регіонах: в деяких місцевостях, зокрема великих містах та їхніх передмістях, спостерігається дефіцит житла, що може підтримувати зростання цін та збільшувати конкуренцію серед покупців;
- збільшений інтерес до передмість: у зв'язку зі змінами в робочих практиках (наприклад, віддалена робота), багато людей виявляють інтерес до життя в передмістях, що призводить до зростання попиту на нерухомість в цих областях;
- стабільні показники будівництва: будівельна галузь вказує на стабільний або певний ріст у певних регіонах, проте, інфраструктурні проблеми, такі як дефіцит будівельних матеріалів, можуть впливати на темпи будівництва.
- великий попит на ринку оренди: в деяких місцевостях попит на оренду житла залишається високим, що впливає на ціни оренди та вигідність інвестицій у нерухомість з метою оренди [16].



Рисунок 1.1 - Динаміка цін на нерухомість в США за 2007-2023рр.

Джерело: [17]

Міністерство житлово-комунального господарства та розвитку міського середовища у своїй праці «Документи з питань демографічних тенденцій,

важливих для житлового сектора» зробили висновки про те, як демографічні зміни в США будуть впливати на ринок житла протягом наступних десятиліть. Визначається чотири ключових демографічних тенденції, які будуть мати найбільший вплив на ринок житла: імміграція, старіння населення, зростання меншин, зменшення розміру домогосподарств. Такі тенденції призведуть до зростання попиту на житло, зміна складу попиту на житло, зміна розташування попиту на житло [18].

Звичайно, демографічні зміни можуть призвести до зростання попиту на житло, але це не обов'язково так. Ринок житла може бути чутливим до інших факторів, таких як економічні умови, політика уряду та технологічні зміни.

Майкл Вілкерсон у своїй статті «Ринки житла в США переживають хвилювання» стверджує, що зростання ставок Федерального резерву є одним із головних факторів, які впливають на ринок житла. Зростання ставок робить житло більш дорогим для покупки, оскільки збільшує суму, яку люди повинні платити за іпотеку. Це, в свою чергу, зменшує попит на житло, що призводить до зниження цін [19].

Іншими факторами, які впливають на ринок житла є інфляція, геополітична напруженість і загальний стан економіки. Інфляція підвищує вартість життя, що зменшує реальний дохід домогосподарств. Це зменшує їхню здатність дозволити собі житло, що також призводить до зниження попиту на житло.

Геополітична напруженість, наприклад, війна в Україні, може призвести до економічної нестабільності, що також може негативно вплинути на ринок житла. Загальний стан економіки, такий як рівень безробіття та темпи зростання економіки, також може впливати на ринок житла.

Падіння ринку житла може мати значний вплив на економіку в цілому. Воно може призвести до зниження споживчих витрат, інвестицій та зайнятості.

Зниження споживчих витрат може призвести до того, що компанії продаватимуть менше товарів і послуг, що може призвести до звільнень працівників. Зниження інвестицій може призвести до того, що компанії

вкладатимуть менше коштів у нові проекти, що також може призвести до звільнень працівників.

Падіння ринку житла також може призвести до того, що люди втрачатимуть свої будинки. Це може призвести до соціальної нестабільності та збільшити витрати уряду на допомогу бездомним [20].

1.4. Методи дослідження цін на ринку нерухомості

На основі аналізу літературних джерел були відібрані наступні методи дослідження цін на ринку нерухомості.

Для більш наочного представлення даних в роботі був використаний графічний метод. Візуалізація даних – це процес, за допомогою якого дані можна представити у вигляді графіків, гістограм, діаграм або інших методів, для виявлення тенденцій та обґрунтування висновків.

У роботі були використані графіки для візуалізації часового ряду та показу результатів прогнозування. Це допомогло більш детально побачити сезонність та тенденції часового ряду, його закономірності та зрівняти прогнозовані дані з реальними, оцінити наскільки якісним вийшов прогноз.

У роботі були використані гістограми для перегляду залежності ціни нерухомості відносно розташування, кількості спалень, ванних кімнат та розміру будинку. Це допомогло зрозуміти, в якому регіоні або місті найдорожча ціна на нерухомість та яка квартира буде мати найвищу ціну відносно кількості спалень, ванних кімнат та розміру будинку.

У роботі було використано також кругові діаграми. Даний метод візуалізації був застосований для більш детального аналізу даних, а саме перегляду залежності ціни від кількості спалень. Візуалізація за допомогою кругових діаграм змогла підтвердити теорію, яка була запропонована, виходячи з результатів гістограм.

Графік розсіювання - це один із методів візуалізації, що був застосований у роботі саме для перегляду результатів кластеризації, що допомогло побачити кластери на які були розбиті дані, а також для відображення результатів регресійного аналізу, а саме розуміння чи вдасться як найточніше спрогнозувати ціну на нерухомість за допомогою цих даних [21].

Завдяки коробчастої діаграми ("ящик з вусами") у роботі були перевірені дані на викиди та нормальність розподілу. За допомогою цих знань були видалені викиди з даних, що допомогло для більш точного майбутнього прогнозу, оскільки велика кількість викидів провокує похибки в прогнозі [22].

За допомогою теплової матриці кореляції, яка відображає інтенсивність явища або значення змінної з використанням кольорів на карті для виявлення взаємозв'язку між змінним. Сама кореляція у роботі була визначена за допомогою коефіцієнта кореляції Пірсона [23]. Він визначається за формулою:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

де n — розмір вибірки; x_i, y_i — окремі точки вибірки з індексом i ; $\bar{x}$ — середнє значення вибірки й аналогічно для $\bar{y}$.

Кореляція є важливим інструментом для аналізу взаємозв'язків між змінними та для розуміння ступеня залежності між ними. Важливо розуміти, що кореляція не завжди означає причинно-наслідкові зв'язки. Наприклад, в роботі визначено, що найбільша кореляція між ціною на нерухомість в США та ВВП на душу населення, це означає, що при високому ВВП на душу населення, як правило є висока ціна на нерухомість, але це зовсім не означає, що високий показник ВВП на душу населення є причиною високої ціни. Можливо, є інші фактори, які впливають на обидві змінні [24].

Графік автокореляції - ще один із способів візуалізувати дані, який використаний в роботі. За допомогою графіку автокореляції у роботі було визначено ступінь кореляції між значеннями часового ряду та його зсувами в часі.

Для вивчення взаємозв'язку між ціною на нерухомість в США, тобто між залежною змінною та незалежними змінними, як, наприклад, ВВП на душу населення, рівень безробіття, кількість спальних кімнат та ванн, площі нерухомості та інше, був використаний регресійний аналіз. Залежна змінна, тобто ціна на нерухомість США є тією, яку ми намагаємося передбачити, а незалежні змінні - це ті, які ми можемо використовувати для прогнозування залежної змінної.

Регресійний аналіз - це статистичний метод для кількісної оцінки зв'язку та виявлення залежності між залежною змінною та однією або кількома незалежними змінними. Він допомагає зрозуміти, як зміни в одній або декількох незалежних змінних пов'язані зі змінами в залежній змінній [25].

Регресійний аналіз був використаний в роботі для оцінки того, чи є причинно-наслідковий зв'язок між ціною на нерухомість в США та залежними змінними. Також зазначений аналіз був застосований для прогнозування значення ціни на нерухомість в США для нових значень незалежних змінних.

Регресійний аналіз можна розділити на декілька типів, включаючи одиночну лінійну регресію, множинну регресію, логістичну регресію та поліноміальну регресію. Відповідна регресійна модель визначається характером даних і предметом дослідження, що розглядається [26].

У роботі використана множинна лінійна регресія, яка є розширеною версією лінійної регресії та дозволяє використовувати більше, ніж одну незалежну змінну. У роботі обраний саме такий вид регресії, оскільки при аналізі виявлено, що залежну змінну можна описати лінійною функцією незалежних змінних, а найкращий зв'язок з залежною змінною мають декілька незалежних змінних. На графіку лінійної регресії є значення, які зосереджуються в основному впродовж прямої лінії [27].

Коефіцієнт асиметрії - це міра асиметрії (несиметричності) розподілу випадкової величини. Він вказує на те, наскільки величина відхиляється від нормального розподілу (симетричного відносно середнього значення).

Коефіцієнт асиметрії визначається як третій центральний момент величини, поділений на стандартне відхилення до куба. Математично це виражається так:

$$\gamma_1 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2)^{\frac{3}{2}}} \quad (2)$$

де x_i — значення випадкової величини; $\bar{x}$ — середнє значення величини; n — кількість спостережень.

Знак коефіцієнта асиметрії вказує на те, в якому напрямку (вправо чи вліво) величина відхиляється від нормального розподілу:

- якщо коефіцієнт додатний, то розподіл скошений вправо, тобто хвіст розподілу розташований праворуч від середнього;
- якщо коефіцієнт від'ємний, то розподіл скошений вліво, тобто хвіст розподілу розташований ліворуч від середнього;
- якщо коефіцієнт дорівнює 0, то розподіл симетричний (без асиметрії).

Отриманий коефіцієнт асиметрії допомагає визначити форму розподілу величини і виявляти відхилення від нормального розподілу [28].

Коефіцієнт ексцесу (ексцес) - це міра "товстості" або "тонкості" хвостів розподілу випадкової величини. Він вказує, наскільки високі чи низькі, в порівнянні з нормальним розподілом, великі значення випадкової величини.

Для нормального розподілу ексцес дорівнює 0. Позитивне значення ексцесу вказує на те, що розподіл має більше важких (товстих) хвостів, тобто більше великих значень, ніж передбачено нормальним розподілом. Негативне значення ексцесу вказує на те, що розподіл має менше важких хвостів, тобто менше великих значень, ніж передбачено нормальним розподілом [29].

Коефіцієнт ексцесу можна розрахувати за допомогою формули:

$$\gamma_2 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2)^2} \quad (3)$$

де x_i — значення випадкової величини; $\bar{x}$ — середнє значення величини; n — кількість спостережень.

Для поділу набору даних на групи, які називаються кластерами в роботі використано кластерний аналіз.

Кластерний аналіз є методом багатовимірного статистичного дослідження, який включає збір даних, що містять інформацію про вибіркові об'єкти, та їх упорядкування в порівняно однорідні, схожі між собою групи. Його суть полягає в класифікації об'єктів дослідження за допомогою обчислювальних процедур, що призводять до формування "кластерів" або груп об'єктів, які є дуже схожими між собою. Відмінності від інших методів полягають в можливості класифікації об'єктів за декількома ознаками одночасно.

Мета кластерного аналізу полягає в виявленні структур в даних через утворення груп схожих об'єктів – кластерів. Цей метод також сприяє введенню структури в досліджувані об'єкти. Кластери формуються з об'єктів, які подібні між собою, при цьому об'єкти з різних кластерів відрізняються. Використання кластерного аналізу важливо для групування подібних об'єктів і розкриття структури даних, що інколи складно визначити візуально чи за допомогою експертів.

Застосування кластерного аналізу важливе для групування схожих об'єктів та виявлення структури даних, зокрема в розділенні квартир на групи. Це дозволить ріелторам з легкістю визначити групу квартир, що відповідає конкретним потребам клієнтів. Враховуючи це, ріелтор може пропонувати квартири з певної групи, що збільшить шанси на успішну угоду купівлі-продажу нерухомості як для покупця, так і для власника [30].

Перед розбиттям на кластери було використано метод ліктя. Метод ліктя (англ. "elbow method") — це алгоритм вибору оптимальної кількості кластерів для кластеризації даних. Цей метод базується на аналізі деякої міри внутрішньокластерної дисперсії (наприклад, суми квадратів відстаней між об'єктами та центрами їхніх кластерів) для різної кількості кластерів. Визначення найоптимальнішої кількості кластерів визначається тоді, коли додаткове збільшення кластерів не призводить до значущого зменшення внутрішньокластерної дисперсії.

У роботі використовується один з методів кластеризації який, має назву К-середні (k-means). К-середні (k-means) є одним із найпоширеніших методів

кластерного аналізу. Цей метод спрямований на розділення набору даних на k кластерів, де k - задане число кластерів, яке визначається перед початком алгоритму.

Перевагами даного метода є те, що він простий у реалізації та розумінні та ефективний саме для великої кількості даних.

Недоліками методу K -середніх є те що він залежний від вибору початкових центроїдів, що може вплинути на кінцевий результат, чутливий до шуму та викидів у даних, а також працює краще, коли кластери є сферичними та рівними за розміром.

Перед тим, як робити прогнозування потрібно перевірити дані на мультиколінеарність, автокореляцію та гетероскедастичність та позбутись цих явищ, якщо вони будуть виявлені.

Мультиколінеарність може виникнути, коли дві або більше змінних мають високий ступінь лінійної залежності одна від одної, що може призвести до серйозних проблем при оцінці параметрів моделі, інтерпретації результатів та узагальненні моделі на нові дані. Мультиколінеарність вважається присутньою, якщо коефіцієнт кореляції між змінними більше 0,9.

Негативними наслідками мультиколінеарності вважаються те, що є можливість зробити некоректною інтерпретацію важливості окремих змінних, при мультиколінеарності є великі стандартні похибки, що може вплинути на подальше прогнозування даних, також мультиколінеарність робить оцінку параметрів моделі менш стійкою та менш точною.

Загальні ознаки мультиколінеарності включають високі значення коефіцієнтів кореляції між змінними, саме для того щоб виявити мультиколінеарність була застосована кореляційна матриця [31].

Існують різні способи за допомогою яких можна позбутись мультиколінеарності, наприклад, видалення факторів, які сильно корелюються з іншими даними, використання спеціальної LASSO, використання методу головних компонент та інші. Модель LASSO - це один із способів регуляризації лінійної регресії, який може допомогти вирішити проблему мультиколінеарності. LASSO

додає до функції втрат (задачі оптимізації) штраф за абсолютні значення коефіцієнтів, що призводить до автоматичного відбору ознак. В роботі було використано модель LASSO для позбавлення моделі від деяких коефіцієнтів ознак. Коли LASSO встановлює коефіцієнт для якої-небудь ознаки на нуль, це означає, що ця ознака не враховується у моделі [32].

Гетероскедастичність вказує на те, що дисперсія помилок в регресійній моделі не є стала по всіх рівнях незалежних змінних. Наслідками гетероскедастичності є нестійкі та неточні оцінки параметрів моделі, несправедливі стандартні помилки та неправильні висновки з тестів гіпотез.

Для визначення гетероскедастичності є багато тестів, саме у роботі було використано тест Бройша-Пагана, Гольфреда-Квандта.

Для обчислення теста Бройша-Пагана є формула [33]:

$$BP = n * R^2 \quad (4)$$

де n – кількість спостережень у вибірці; R^2 – коефіцієнт детермінації з регресії з квадратами залишкових змінних на незалежні змінні.

Інтерпретація результатів:

- якщо BP-статистика більше критичного значення χ^2 -квадрат розподілу з $k-1$ ступенями свободи та заданим рівнем значущості (наприклад, 0,05), то гетероскедастичність вважається суттєвою.
- якщо BP-статистика менше критичного значення, то гетероскедастичність не вважається суттєвою.

Тест Гольфреда-Квандта обчислюється за формулою [34]:

$$F = ESS2 / ESS1 \quad (5)$$

де ESS1 - сума квадратів залишків регресії, розрахована за першою підвибіркою (зазвичай це спостереження з нижчими значеннями незалежної змінної); ESS2 - сума квадратів залишків регресії, розрахована за другою підвибіркою (зазвичай це спостереження з вищими значеннями незалежної змінної).

Інтерпретація результатів:

- якщо F більше критичного значення F -розподілу з (k, k) ступенями свободи та заданим рівнем значущості (наприклад, 0,05), то гетероскедастичність вважається суттєвою;
- якщо F менше критичного значення, то гетероскедастичність не вважається суттєвою.

Тести, що були описані використовуються для перевірки гіпотези про гомоскедастичність (постійну дисперсію) помилок в регресійній моделі. В основі тестів лежить порівняння дисперсій помилок для різних значень незалежних змінних.

Є дві основні гіпотези:

- нульова гіпотеза (H_0): гомоскедастичність - дисперсія помилок однакова для всіх значень незалежних змінних.
- альтернативна гіпотеза (H_1): гетероскедастичність - дисперсія помилок різна для різних значень незалежних змінних.

Для того щоб позбутись гетероскедастичності використовують різні способи, такі як видалення викидів, збільшення обсягу вибірки, видалення викидів, використання ваг та інше [35].

Автокореляція вказує на кореляцію між значеннями однієї та тієї ж самої змінної в різних часових або просторових точках. Це означає, що кореляція спостерігається між значеннями змінної та її лагованими (відстроченими) копіями. Автокореляція може впливати на точність оцінок параметрів у моделях.

Існує позитивна та негативна автокореляція. Позитивна автокореляція виникає тоді, коли змінні в часі або просторі корелюють позитивно зі своїми попередніми значеннями, іншими словами, велике значення супроводжується великими значеннями у попередні моменти часу чи простору, і навпаки. Негативна автокореляція вказує на те, що між значеннями змінної в різні моменти часу (або простору) існує негативна залежність, тобто, коли значення змінної в один момент часу збільшуються, значення в інший момент часу зменшуються, і навпаки.

Причини виникнення автокореляції можуть бути циклічні коливання, тренди, спільні фактори та моделі з підгонкою.

Для виявлення автокореляції було використано графіки автокореляції (ACF). Ці графіки відображають кореляцію між значеннями змінної та її лагованими копіями на різних затримках (лагах) [36].

Також для більш точного визначення автокореляції було використано статистичний тест, а саме тест Дарбіна-Уотсона[37]. Основна ідея тесту полягає в перевірці гіпотези про відсутність автокореляції у залишках:

- нульова гіпотеза (H_0): відсутність автокореляції (некорельованість) в залишковому члені;
- альтернативна гіпотеза (H_1): при наявності автокореляції в залишковому члені.

Для перевірки на автокореляцію залишків за тестом Дарбіна-Уотсона існує два кроки. Перший крок це визначення d-статистики, яка обчислюється за формулою:

$$DW = d = \frac{\sum_{t=2}^n (u_t - u_{t-1})^2}{\sum_{t=1}^n u_t^2} \quad (6)$$

де u_t – залишковий член у момент часу t ; n – кількість спостережень.

Наступний кроком задаємо рівень значущості α . Далі за таблицею Дарбіна-Уотсона при заданому рівні значущості α , кількості факторів m і кількості спостережень n знаходимо два значення DW_1 і DW_2 .

Статистика Дарбіна-Уотсона приймає значення від 0 до 4. Результати можна інтерпретувати так (рис 1.2):



Рисунок 1.2 - Інтерпритація автокореляції за тестом Дарбіна-Уотсона

Джерело: [38]

- якщо $0 < DW < DW_1$, то наявна додатна автокореляція;

- якщо $DW1 < DW < DW2$ або $4-DW2 < DW < 4-DW1$, то ми не можемо зробити висновки ані про наявність, ані про відсутність автокореляції (DW потрапляє в зону невизначеності);
- якщо $4-DW1 < DW < 4$, маємо від'ємну автокореляцію;
- якщо $DW2 < DW < 4-DW2$, то автокореляція відсутня.

Для того щоб позбутись мультиколінеарності використовують різні способи, наприклад, введення у модель фактору часу, включення в модель нових змінних та інше. Саме в роботі було використано модель експоненціального згладжування [38].

Для прогнозування також потрібно виявити, чи є в наших даних сезонність, оскільки ці знання зможуть допомогти в подальшому у виборі методів прогнозування для більшої точності прогнозу. Для цього був використаний метод визначення сезонності за допомогою аналізу сезонної декомпозиції. Цей метод є потужним інструментом для розкриття тренду, сезонності та залишкових компонентів в часових рядах. Він включає в себе декомпозицію часового ряду на три компоненти: тренд, сезонність та залишкові значення.

Одним із методів прогнозування, які використані в роботі є експоненціальне згладжування. Це метод аналізу та прогнозу часових рядів, який використовує зважений підхід для прогнозних значень на основі попередніх спостережень у часі. Цей метод особливо корисний для рядів, де присутня трендова або сезонна структура. Оскільки дані мають сезонність та тренд, то цей метод буде доцільно використати для прогнозу ціни на нерухомість в США.

Основна ідея експоненціального згладжування полягає в тому, щоб приділити більшу вагу більш новим спостереженням та меншу вагу старшим. Це досягається за допомогою згладжувального коефіцієнта (зазвичай позначається як α), який визначає, наскільки сильно ви згладжуєте дані.

Згладжування може бути експоненціальним зваженим середнім (EWMA), де кожне наступне прогнозне значення обчислюється за допомогою попереднього прогнозного значення та актуального спостереження.

Переваги експоненціального згладжування включають його простоту та ефективність у визначенні змін у трендах чи сезонних коливаннях. Однак важливо правильно вибрати параметр α , оскільки він впливає на рівень згладжування [39, 40].

Ще один із методів прогнозування який використаний прогнозування застосовуючи SARIMA, яка є розширення моделі ARIMA. Модель ARIMA (AutoRegressive Integrated Moving Average) – це модель, яка об'єднує три основні компоненти: авторегресія (AR), інтегрованість (I) та ковзне середнє (MA). В свою чергу модель SARIMA враховує ще й сезонні коливання в часовому ряді. Зазвичай позначається як SARIMA (p,d,q)(P,D,Q)_s, де:

- p,d,q - порядки для компонент AR, I та MA в основній моделі ARIMA.
- P,D,Q - порядки для сезонних компонент AR, I та MA.
- s - період сезонності.

Цей підхід особливо ефективний для прогнозування часових рядів з вираженими сезонними змінами. Оскільки дані, які використовуються у роботі мають сезонність та тренд, то використання даної моделі є доцільним. Важливо правильно визначити параметри моделі, оцінити її ефективність та забезпечити, щоб вхідний часовий ряд був достатньо стаціонарним [41].

РОЗДІЛ 2

АНАЛІЗ ФАКТОРІВ, ЩО ВПЛИВАЮТЬ НА ЦІНИ НА НЕРУХОМІСТЬ В США

2.1. Опис вибраних даних для аналізу факторів, що впливають на ціну нерухомості

Для роботи взято два датасети, один з яких був знайдений на платформі Kaggle [42], створений для змагань у сфері науки про дані, машинного навчання та аналізу даних. Цей датасет складається з 10 колонок та 904966 значень. У датасеті (рис. 2.1) є такі змінні:

```

RangeIndex: 904966 entries, 0 to 904965
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   status                 904966 non-null object  
1   bed                    775126 non-null float64  
2   bath                   791082 non-null float64  
3   acre_lot               638324 non-null float64  
4   city                   904894 non-null object  
5   state                  904966 non-null object  
6   zip_code               904762 non-null float64  
7   house_size             612080 non-null float64  
8   prev_sold_date         445865 non-null object  
9   price                  904895 non-null float64  
dtypes: float64(6), object(4)
memory usage: 69.0+ MB

```

Рисунок 2.1 – Дані по нерухомості з неекономічними факторами

Джерело: складене автором на основі [42]

- status – статус квартири на момент зібрання даних (для продажу або готовий для будівництва);
- bed – кількість спалень;
- bath - кількість ванних кімнат;
- acre_lot - розмір земельної ділянки;
- city- місто в якому знаходиться нерухомість;

- state – штат в якому знаходиться нерухомість;
- zip_code- поштовий індекс;
- house_size – розмір нерухомості, м²;
- prev_sold_date – дата попереднього продажу;
- price – ціна на нерухомість, долари США.

Перед початком аналізу даних була проведена передобробка даних, а саме позбавлення від пропущених значень та видалення викидів в цільовому показнику, тобто в змінній ціна. На викиди цільова змінна була перевірена за допомогою коробчастої діаграми, яка показала викиди у змінній price (рис 2.2).

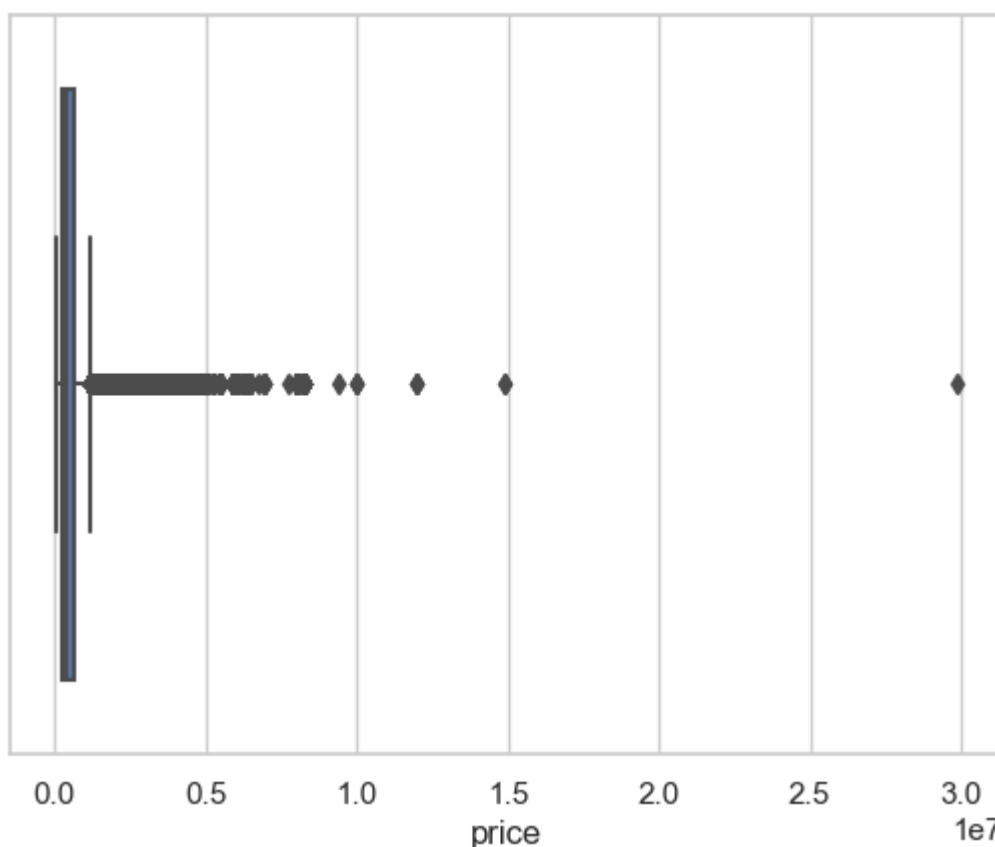


Рисунок 2.2 - Коробчаста діаграма для цільового показника price до видалення викидів

Джерело: складено автором на основі [42]

Для кращої візуалізації змінної price було видалено верхні 10% найвищих цін (рис. 2.3). На коробчастій діаграмі помічено правосторонню асиметрію, що вказує на те, що в даних є значення, які відступають від типового розподілу та скупчені вздовж більших значень.

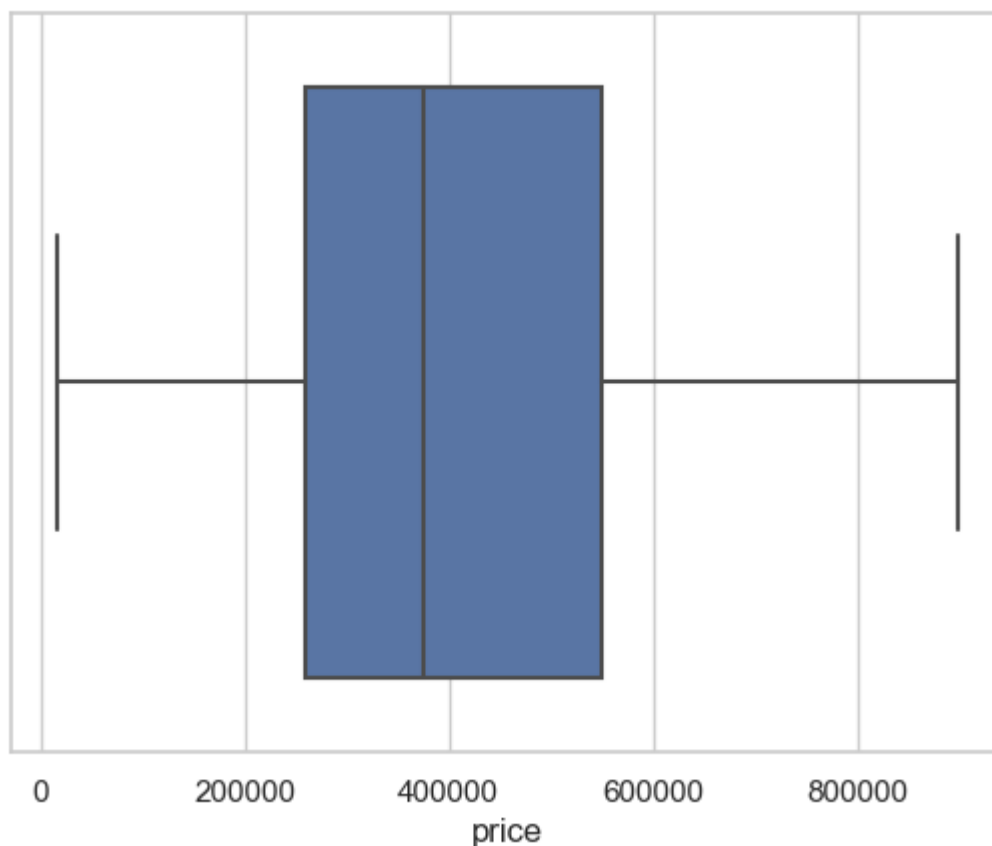


Рисунок 2.3 - Коробчаста діаграма для цільового показника `price` після видалення викидів.

Джерело: складено автором на основі [42]

Інші змінні, такі як `bed`, `bath` та `house_size`, також були перевірені на викиди та видалені.

Також в датасеті є деякі змінні в числовому форматі, але насправді мають бути категоріальними змінними, як-от поштовий індекс. Кількість спалень і ванних кімнат також слід розглядати як категоріальні змінні, оскільки ці значення є дискретними, а не безперервними.

Також із цільової змінної були видалені пропущені значення, для того аби в подальшому побудувати як найточнішу модель (рис 2.4).

```

Index: 149467 entries, 17 to 192631
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype
---  -
0   status                 149467 non-null object
1   bed                    149467 non-null float64
2   bath                   149467 non-null float64
3   acre_lot               149467 non-null float64
4   city                   149467 non-null object
5   state                  149467 non-null object
6   zip_code               149467 non-null float64
7   house_size             149467 non-null float64
8   prev_sold_date         149467 non-null object
9   price                  149467 non-null float64
dtypes: float64(6), object(4)
memory usage: 12.5+ MB

```

Рисунок 2.4 – Дані по нерухомості з неекономічними факторами після обробки

Джерело: складено автором на основі [42]

Датасет обраний для того, аби проаналізувати, як розташування, розмір будинку, кількість спалень та ванних кімнат впливає на ціну нерухомості, також для проведення кластерного аналізу, тобто розбиття даних на групи. Це дослідження допомогло б ріелторам, покупцям та продавцям, тому що ріелтори дізнавшись потреби покупців, могли б запропонувати найбільш відповідний до потреб клієнтів варіант, як по ціні, так і по іншим характеристикам.

Інший датасет складений самостійно за допомогою даних із різних ресурсів [43-46]. В другий датасет входить також ціна на нерухомість але уже зовнішні фактори, які впливають на ціну нерухомості. Дані взяті за останні 5 років помісячно. Датасет складається з 10 змінних в кількості 65. В датасет входять такі змінні (рис 2.5):

```

RangeIndex: 65 entries, 0 to 64
Data columns (total 10 columns):
#   Column                                                                 Non-Null Count  Dtype
---  -
0   Дата                                                                    65 non-null    datetime64[ns]
1   Середня ціна на нерухомість (млн доларів США)                       65 non-null    float64
2   ВВП на душу населення                                                 65 non-null    float64
3   Річна інфляція                                                         65 non-null    float64
4   Рівень безробіття                                                      65 non-null    float64
5   Процентна ставка                                                      65 non-null    float64
6   Середній дохід домогосподарства                                       65 non-null    int64
7   Нетто міграції(осіб)                                                  65 non-null    int64
8   Пропозиція нерухомості (млн будинків)                                65 non-null    float64
9   Попит на нерухомість (млн будинків)                                   65 non-null    float64
dtypes: datetime64[ns](1), float64(7), int64(2)
memory usage: 5.2 KB

```

Рисунок 2.5 – Дані по нерухомості з економічними факторами

Джерело: складено автором на основі [43-46]

- дата – дата на момент якої були зафіксовані інші змінні, представлені саме помісячні дані;
- середня ціна на нерухомість, млн доларів США;
- ВВП на душу населення, доларів США ;
- річна інфляція, %;
- рівень безробіття, %;
- процентна ставка, %;
- середній дохід домогосподарства, доларів США;
- нетто міграції, осіб;
- пропозиція нерухомості, млн будинків;
- попит на нерухомість, млн будинків.

Оскільки, дані були зібрані вручну з декількох джерел, то пропусків не мають, але цільова зміна була перевірена на викиди за допомогою коробчастої діаграми. Яка показала що у змінній викидів немає, а розподіл даних має незначну правосторонню асиметрію, що може вплинути в подальшому на вибір моделі прогнозування (рис. 2.6).

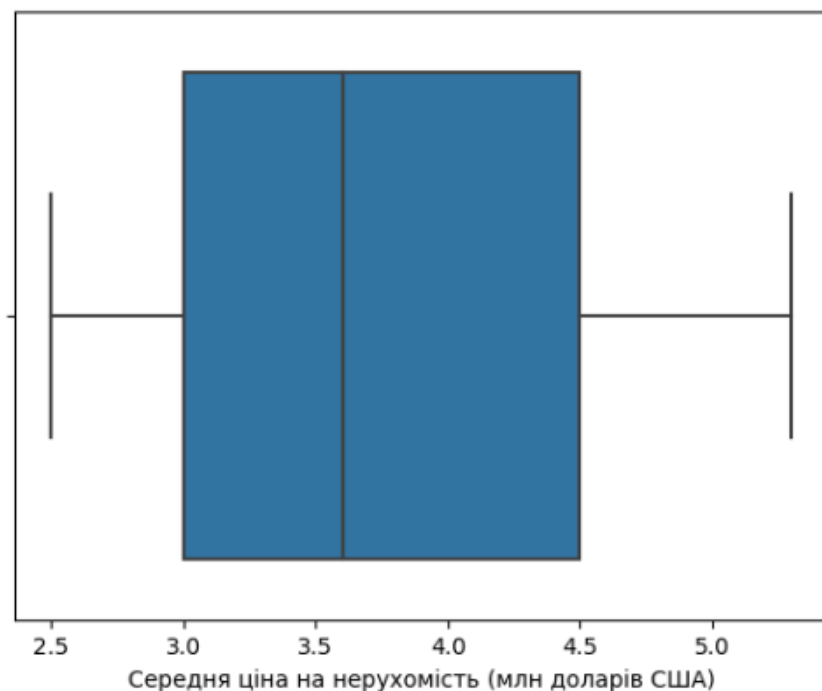


Рисунок 2.6 – Коробчаста діаграма для цільового показника Середня ціна на нерухомість

Джерело: складено автором на основі [43-46]

У роботі датасет використовується задля розгляду економічних факторів, які впливають на ціну нерухомості в США, аналізу цих факторів та в подальшому прогнозування ціни на наступний рік. В датасет було взято саме ті змінні, які вважались найбільш впливовішими на думку експертів в проаналізованих статтях в галузі досліджень ринку нерухомості в США.

2.2. Розгляд факторів, які мають найбільший та найменший вплив на ціну квартири в США

Для початку розглянемо залежність ціни від розташування кількості спалень та ванних кімнати.

Найдорожчою нерухомість є в штаті Нью-Йорк (рис. 2.7). Це можна пояснити тим, що Нью-Йорк є одним із найбільших і найпопулярніших міст у світі.

Місто вважається фінансовим та економічним центром світу, Нью-Йорк є домом для багатьох міжнародних компаній і організацій, а також приваблює мільйони туристів щороку. Це створює великий попит на житло в місті, що підвищує ціни.

Ще одними із лідерів по ціні є штат Нью-Гемпшир і Массачусетс, вони є популярними місцями для проживання, як для місцевих жителів, так і для людей, які переїжджають з інших штатів. Це створює великий попит на житло, що підвищує ціни. Кількість доступного житла в Нью-Гемпширі та Массачусетсі не відповідає попиту. Це також сприяє підвищенню цін. У Нью-Гемпширі та Массачусетсі середній дохід домогосподарства значно вищий, ніж у середньому по США. Це означає, що люди в цих штатах можуть собі дозволити платити більше за житло.

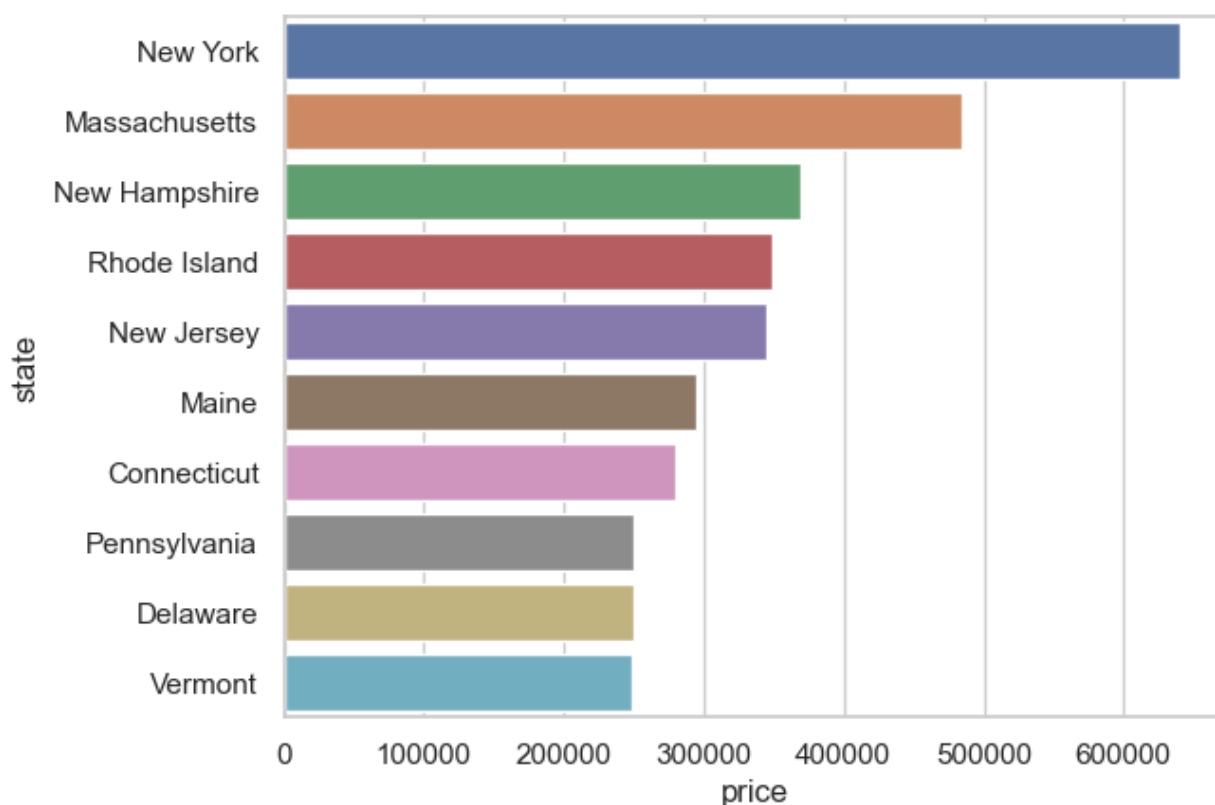


Рисунок 2.7 – Середня ціна на нерухомість в США по штатам

Джерело: складено автором на основі [42]

Щодо міст (рис. 2.8) у топ-10 найдорожчих міст два міста з Массачусетсу, шість з Нью-Йорка, одне з Нью-Джерсі, одне з Вермонта та одне Род-Айленда. Але не помічено ні одного міста зі штату Нью-Гемпшир. Це спричинено тим, що в даних подано невелика кількість міст зі штату Нью-Гемпшир з досить високою

ціною на квартири. Відповідно до цього, по середній ціні даний штат стоїть на третьому місці, а інші, оскільки мають більшу кількість міст мають і нижчу середню ціну, але входять в топ-10 найдорожчих міст.

Враховуючи цю інформацію можемо вважати, що найдорожчими штатами є Нью Йорк та Массачусетс, що є зрозумілим, оскільки ці штати є густонаселеними та мають високий попит на житло.

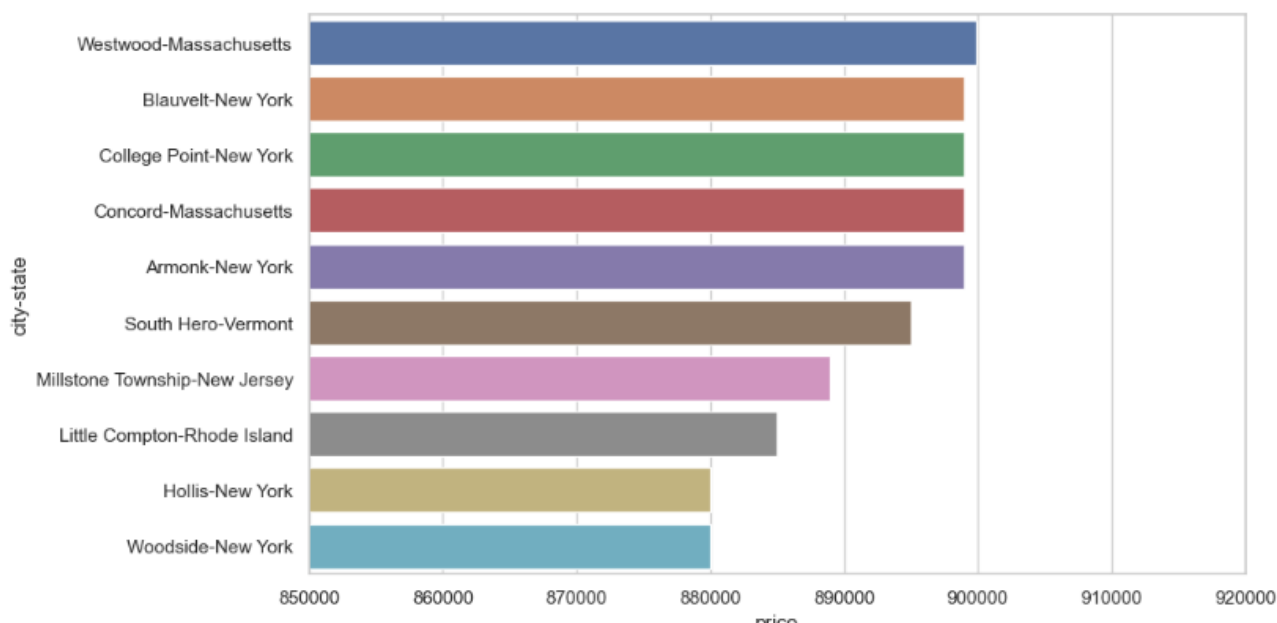


Рисунок 2.8 – Середня ціна на нерухомість в США по містам

Джерело: складено автором на основі [42]

Розглянемо, як залежить ціна квартир від кількості спальних кімнат (рис. 2.9).

При аналізі взаємозв'язку середньої ціни житла і кількості спалень спостерігається загальна тенденція до підвищення ціни зі збільшенням кількості спалень у будинку. Однак цікавим явищем є те, що середня ціна для будинків із однією спальнею виявляється вищою, ніж для будинків із двома і трьома спальнями.

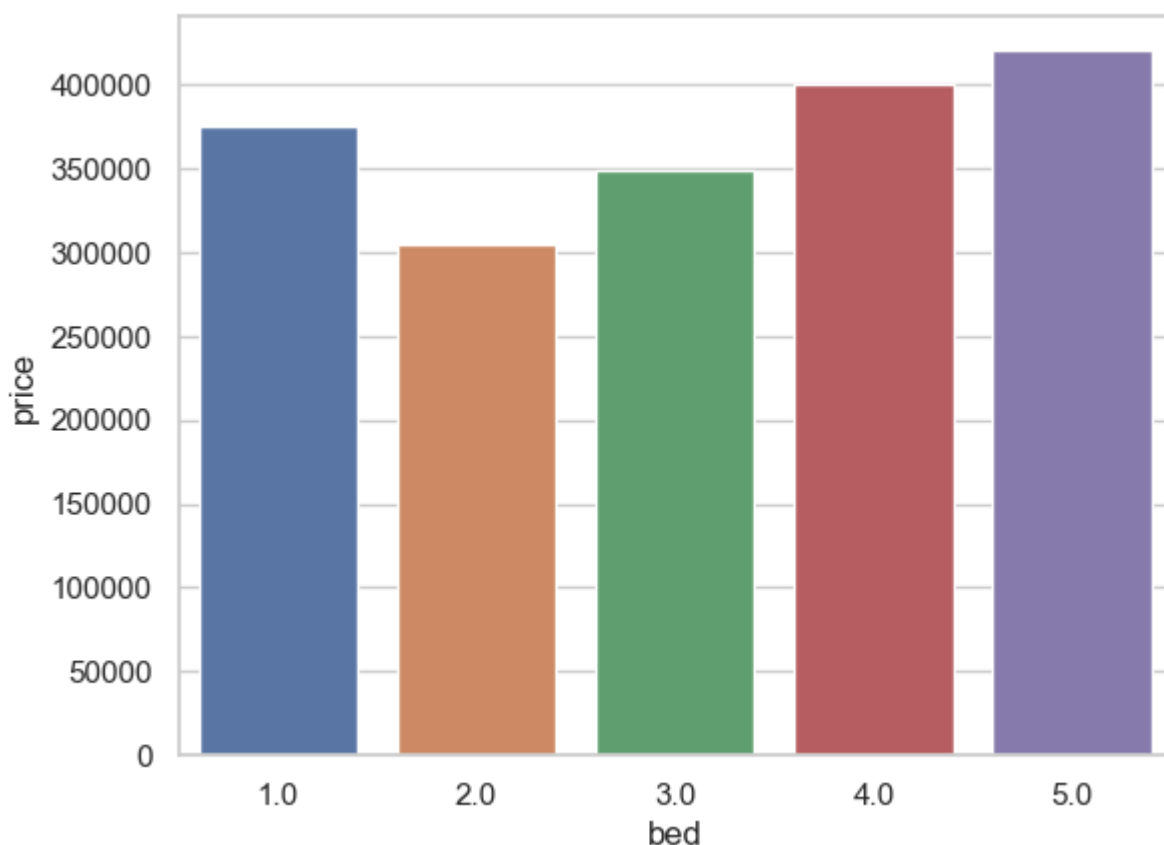


Рисунок 2.9 – Залежність ціни від кількості спалень

Джерело: складено автором на основі [42]

Припустимо, що у вхідних даних визначено кілька міст із високою вартістю життя, де, як правило, переважають однокімнатні квартири чи будинки. Це може впливати на те, чому середня ціна за одну спальню може виявитися вищою. З метою ретельнішого аналізу цього явища була створена кругова діаграма, яка вказує на розподіл кількості однокімнатних, двокімнатних та трикімнатних квартир у цих містах (рис. 2.10).

З двох кругових діаграм ми можемо побачити, що Нью-Йорк займає набагато більший потенціал на ринку однокімнатних будинків порівняно з ринком двокімнатних або трикімнатних будинків. І з попереднього аналізу ми знаємо, що Нью-Йорк є найдорожчим штатом для житла в наших даних, і це добре пояснює, чому ми бачимо, що середня ціна житла з однією спальнею вища, ніж будинки з двома або трьома спальнями.

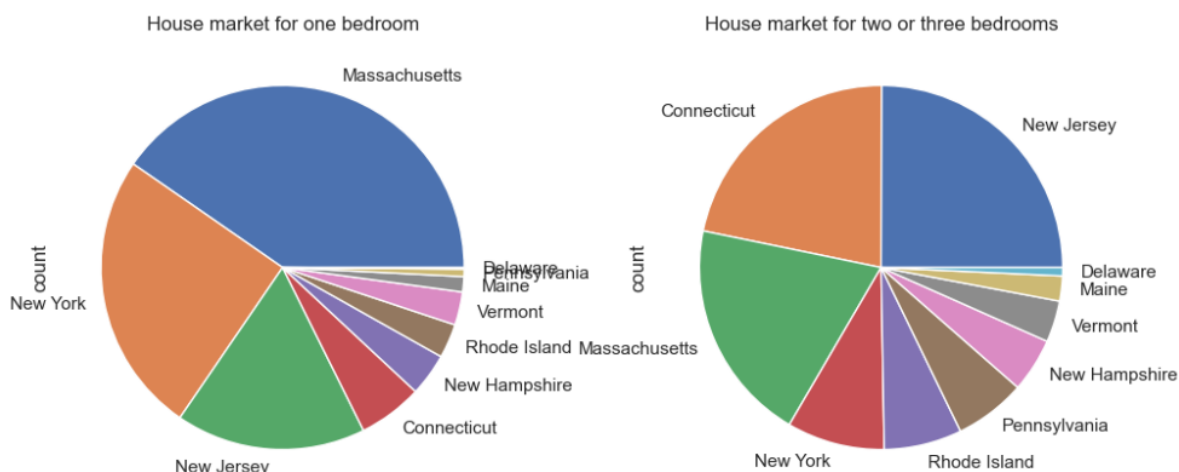


Рисунок 2.10 – Кількість однокімнатних, двокімнатний та трикімнатних квартир в кожному штаті

Джерело: складено автором на основі [42]

Для аналізу залежності ціни від числових факторів було виконано побудову теплової кореляційної матриці (рис. 2.11). За результатами матриці кореляції відзначено, що найвищий рівень кореляції із ціною спостерігається у відношенні до кількості ванних кімнат та розміру будинку. Відомо, що вартість нерухомості тісно пов'язана із її площею, і чим більший розмір будинку, тим вища його ціна. Водночас, мінімальний рівень кореляції із ціною спостерігається відносно розміру земельної ділянки. Це може бути пов'язано з тим, що більшість об'єктів нерухомості у датасеті є квартирами, де розмір земельної ділянки відносно малий. У такому контексті вартість землі виявляється менш визначальною для ціноутворення, оскільки більшість квартир розташована в великих містах, де цінова динаміка має свої особливості.

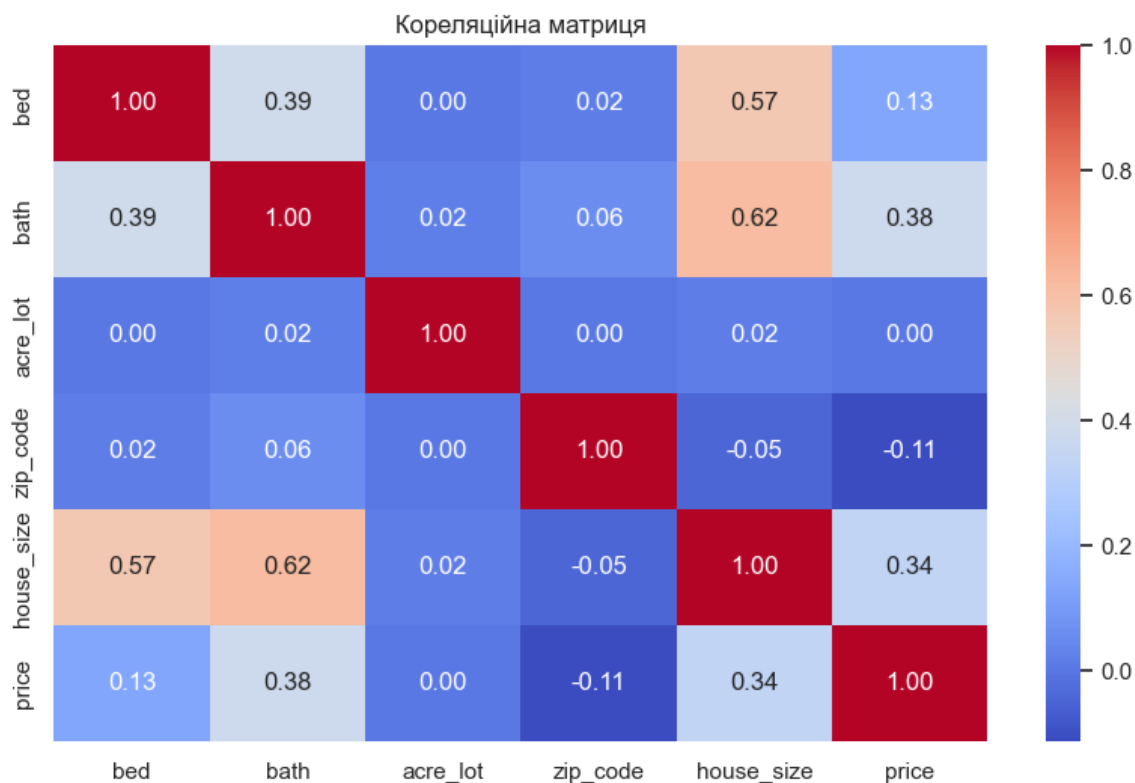


Рисунок 2.11 – Кореляційна матриця неекономічних факторів

Джерело: створено автором на основі [42]

Щодо економічних факторів (рис. 2.12), то найбільшу додатню кореляцію з середньою ціною на нерухомість США, мають ВВП на душу населення, річна інфляція, попит на нерухомість, а також відносно високу кореляцію має середній дохід домогосподарств. Найбільшу від’ємну кореляцію має змінна пропозиція на нерухомість.

Попит на ціну нерухомості має велику додатню кореляцію, оскільки за законом попиту і пропозиції, коли попит перевищує пропозицію то ціна має тенденцію зростати. Як ми бачимо пропозиція на ціну нерухомості має від’ємний характер, тобто при зменшенні пропозиції та зростання попиту ціна росте [47].

Кореляція між ВВП на душу населення та ціною нерухомості може бути такою високою, оскільки регіони з високим ВВП на душу населення володіють великою економічною активністю, що і приваблює велику кількість покупців, а відповідно і призводить до зростання ціни на нерухомість.

Ще однією із причин такої високої кореляції між ВВП на душу населення та середньою вартістю житла може бути те, що вищий ВВП впливає на можливість

людей придбати житло, оскільки при високому ВВП спостерігається стабільність на ринку праці та зростання зайнятості. Також регіон з високим ВВП на душу населення часто привертають більше інвестицій, що може відобразитися на розвитку нерухомості, оскільки інвесторів привертає можливість отримати високий дохід від нерухомості [48].

Інфляція може мати додатню кореляцію з середньою ціною на нерухомість в США, оскільки у періоди високої інфляції люди та інвестори можуть шукати способи збереження вартості своїх активів та вкладатись, наприклад, у нерухомість, оскільки це є одним із способів збереження своїх активів. Також дії центральних банків та урядів можуть впливати на інфляцію та ціни на нерухомість, оскільки банки запускають заходи, що можуть підтримати економіку країни, наприклад низькі ставки відсоткового кредиту та інші стимулюючі програми, що може впливати на ціни.

Ще однією з причин такої кореляції може бути і те що при інфляції ціни на всі товари зростають, не виключенням є і будматеріали, що у подальшому і призводить до збільшення ціни на нерухомість [49].

Щодо доходу домогосподарств, то велика кореляція може бути спричинена тим, що при збільшенні доходу домогосподарств зростає купівельна спроможність людей, відповідно і зростає попит на житло, а з ним і ціна на нерухомість. З вищим рівнем доходу люди можуть шукати більш комфортне житло або розглядати вищий клас житла [50].

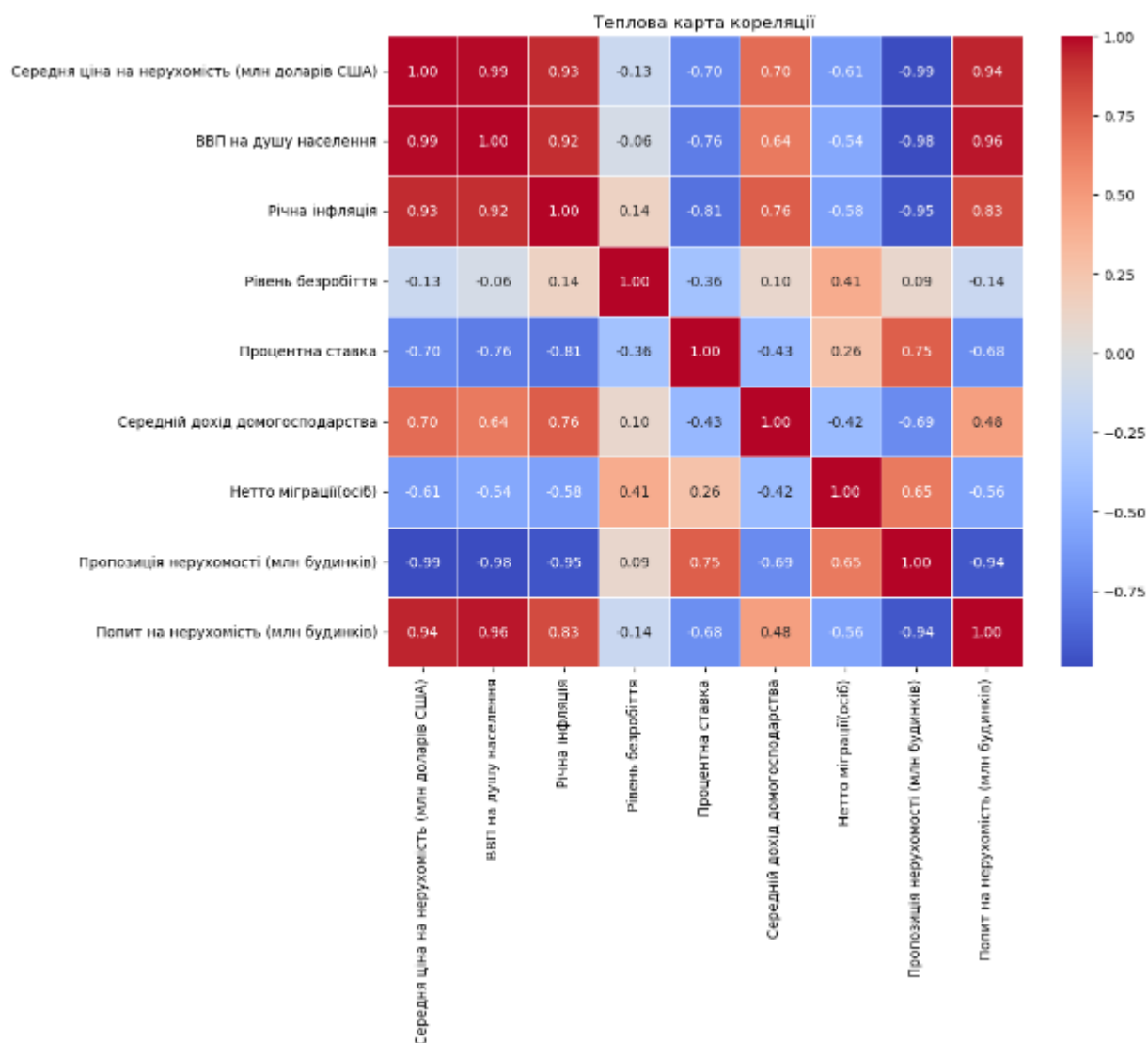


Рисунок 2.12 – Кореляційна матриця економічних факторів

Джерело: складено автором на основі [43-46]

Рівень безробіття має від’ємний і найнижчий показник кореляції з нерухомістю. У розвинених країнах, де економічний розвиток високий, рівень безробіття може бути низьким, оскільки є велика кількість робочих місць і попит на працю. Водночас, високий рівень економічного розвитку може також супроводжуватися збільшеним попитом на нерухомість, що може призвести до підвищення цін.

2.3. Кластерний аналіз об'єктів нерухомості США

Для розбиття нерухомості на категорії був використаний кластерний аналіз за допомогою K-means методу.

Для початку за допомогою методі ліктя був побудований графіку зміни внутрішньо-кластерної суми квадратів (рис. 2.13). Отже, найбільш оптимальна сума кластерів дорівнює 5, оскільки після 5 зміна внутрішньо-кластерної суми квадратів почала ставати меншою. Отже, найдоцільніше буде розділити наші дані на 5 груп.

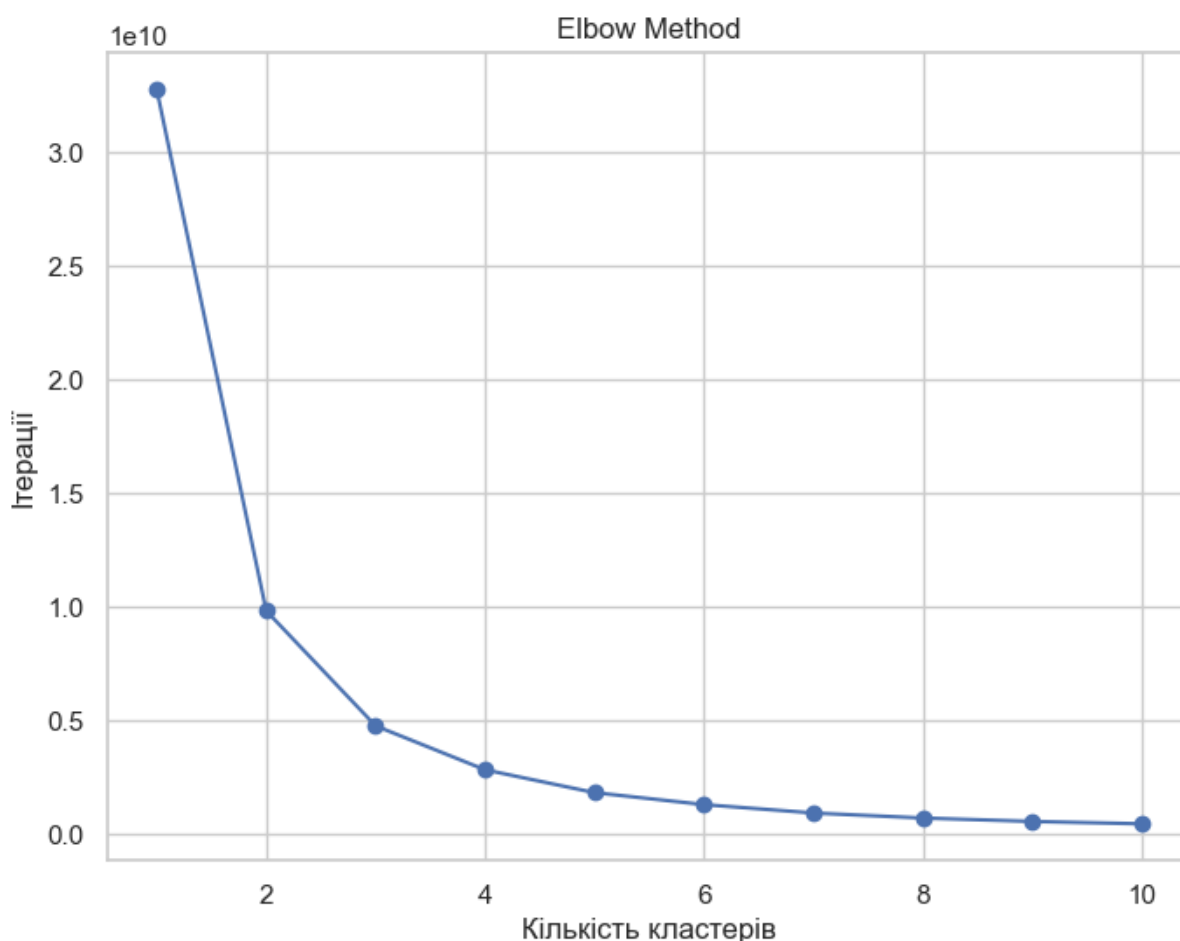


Рисунок 2.13 – Метод ліктя для визначення найоптимальнішої кількості кластерів

Джерело: складено автором на основі [42]

Результати кластерного аналізу представлені у вигляді статистики (рис. 2.14). Маємо такі кластери:

- кластер 0 характеризується невеликою кількістю спалень (в середньому 2,91), невеликою кількістю ванних кімнат (в середньому 1,55), невеликою площею (в середньому 1293,1 кв.м.) та відносно низькою ціною (в середньому 174606,6 доларів). Цей кластер можна віднести до категорії "квартири для молодих сімей або людей, які живуть самотійно". Ці квартири можуть бути привабливими для молодих сімей або людей, які живуть самотійно, оскільки вони пропонують достатній простір для комфортного життя за відносно низькою ціною.
- кластер 1 характеризується великою кількістю спалень (в середньому 3,15), великою кількістю ванних кімнат (в середньому 2,34), великою площею (в середньому 1712,7 кв.м.) та дуже високою ціною (в середньому 591705,7 доларів). Цей кластер можна віднести до категорії "розкішні квартири". Ці квартири є найдорожчими в наборі даних. Вони можуть бути привабливими для людей з високим рівнем доходу, які шукають розкішне житло.
- кластер 2 схожий на кластер 1, але має більшу кількість спалень (в середньому 3,16), меншу кількість ванних кімнат (в середньому 2,25), є меншим за розміром та не таким дорогим, як квартири в кластері 1. Цей кластер можна віднести до категорії "квартири преміум-класу". Вони можуть бути привабливими для людей, які шукають комфортне житло за трохи нижчою ціною.
- кластер 3 характеризується великою кількістю спалень (в середньому 3,26), великою кількістю ванних кімнат (в середньому 2,43), великою площею (в середньому 1751,6 кв.м.) та дуже високою ціною (в середньому 782622,6 доларів). Цей кластер можна віднести до категорії "великі дома". Вони можуть бути привабливими для великих сімей з високим рівнем доходу, або людей які люблять приймати велику кількість гостей.
- кластер 4 характеризується середньою кількістю спалень (в середньому 3,03), середньою кількістю ванних кімнат (в середньому 1,93), середньою площею (в середньому 1457,3 кв.м.) та відносно низькою ціною (в середньому

297891,5 доларів). Цей кластер можна віднести до категорії "квартири для середнього класу". Ці квартири є найбільш доступними в наборі даних. Вони можуть бути привабливими для людей зі середнім рівнем доходу, які шукають просторе житло за помірною ціною.

Між кластерами 0 і 1 спостерігається значна різниця за ціною. Це може бути пов'язано з тим, що ці кластери представляють різні сегменти ринку нерухомості. Між кластерами 1, 2 і 3 спостерігається менша різниця за ціною. Це може бути пов'язано з тим, що ці кластери представляють більш схожі сегменти ринку нерухомості. Між кластерами 0, 4 і 1, 2, 3 спостерігається значна різниця за розміром. Це може бути пов'язано з тим, що ці кластери представляють різні типи будинків, які призначені для різних цілей.

cluster	bed mean	bath mean	house_size mean	price mean
0	2.905496	1.553759	1293.130009	174606.608252
1	3.151670	2.335472	1712.690062	591705.696389
2	3.159227	2.249552	1674.303083	430043.584998
3	3.260479	2.433391	1751.648076	782622.584841
4	3.031873	1.930978	1457.332072	297891.542490

Рисунок 2.14 – Статистичні дані по кластерам

Джерело: складено автором на основі [42]

Отже, виконавши кластерний аналіз ми розділили наші дані по нерухомості на 5 груп. Це зроблено для того аби виявивши потреби клієнтів ріелтори могли пропонувати клієнтам певну групу будинків. Це допоможе більш швидше клієнтам визначитись, яку саме нерухомість вони бажають придбати та вибирати саме ту нерухомість, яка підходить під їх стиль життя та забаганки.

2.4. Регресійний аналіз цін нерухомості в США

Було проведено регресійні аналізи, неекономічних показників та економічних, на основі множинної регресії.

Регресійний аналіз неекономічних показників використовує модель лінійної множинної регресії для пояснення залежної змінної "price" (ціна нерухомості) за допомогою трьох незалежних змінних: "bed" (кількість спалень), "bath" (кількість ванних кімнат) і "house_size" (розмір будинку). Були обрані саме ці показники оскільки вони найтісніше корелюються з цільовою змінною price.

Маємо рівняння регресії:

$$Y = 183218.8892 - 23026.1379 * X1 + 70733.4802 * X2 + 87.8422 * X3$$

де X1- це кількість спалень, X2- це кількість ванних кімнат, X3- це розміри будинку.

Коефіцієнт константи в рівнянні регресії (183218.8892) представляє ціну квартири з нулем спалень, нулем ванних кімнат та нульовою площею. Це означає, що квартира з такими характеристиками буде мати ціну в середньому 183218.8892 доларів. Цей коефіцієнт є досить високим, що може свідчити про те, що на ціну квартири в основному впливають фактори, крім кількості спалень, кількості ванних кімнат та площі будинку. Наприклад, розташування, стан будинку, наявність інфраструктури тощо..

Стосовно кількість спалень маємо від'ємний коефіцієнт (-23026.1379), що означає, що зі збільшенням кількості спалень на одну ціна нерухомості зменшується на 23026.1379 долари США. У підрозділі 2.2 було розглянуто цю ситуацію, тобто це зумовлено тим, що в наших даних існує більше однокімнатних квартир в дорогих регіонах США, через це ціна відповідна. Отож, таке значення коефіцієнта спалень зумовлено тим, що більші будинки із багатьма спальнями розташовані в менш цінних районах.

Щодо кількість ванних кімнат маємо позитивний коефіцієнт (70733.4802), що вказує на те, що зі збільшенням кількості ванних кімнат на одну ціна нерухомості також збільшується на 70733.4802 долари США. Тому можемо зробити висновок, що нерухомості із більшою кількістю ванних кімнат можуть бути більш привабливими для покупців.

Якщо розглядати змінну розмір будинку, то маємо також позитивний коефіцієнт (87.8422), що означає, що зі збільшенням розміру будинку на 1м² ціна

нерухомості також зростає на 87.8422 долари США. Це спонукає нас до висновку, що більші будинки часто мають вищу ціну через більшу житлову площу.

Отож, найбільше наша ціна зросте саме від збільшення кількості ванних кімнат, але при збільшенні ванних кімнат на нашу думку може зростати і квадратура.

Розглянемо критерії точності нашої моделі (рис. 2.15).

OLS Regression Results						
=====						
Dep. Variable:	price		R-squared:	0.168		
Model:	OLS		Adj. R-squared:	0.168		
Method:	Least Squares		F-statistic:	1.004e+04		
Date:	Sat, 20 Jan 2024		Prob (F-statistic):	0.00		
Time:	15:29:10		Log-Likelihood:	-2.0135e+06		
No. Observations:	149467		AIC:	4.027e+06		
Df Residuals:	149463		BIC:	4.027e+06		
Df Model:	3					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	1.832e+05	1846.543	99.223	0.000	1.8e+05	1.87e+05
bed	-2.303e+04	652.592	-35.284	0.000	-2.43e+04	-2.17e+04
bath	7.073e+04	747.076	94.680	0.000	6.93e+04	7.22e+04
house_size	87.8422	1.349	65.094	0.000	85.197	90.487
=====						
Omnibus:	13904.836		Durbin-Watson:	0.775		
Prob(Omnibus):	0.000		Jarque-Bera (JB):	18092.144		
Skew:	0.836		Prob(JB):	0.00		
Kurtosis:	3.330		Cond. No.	6.84e+03		
=====						

Рисунок 2.15 – Результати регресійного аналізу з неекономчними змінними

Джерело: складено автором на основі [42]

Розглянемо R-squared, тобто коефіцієнт детермінації, його значення становить 0.168, що вказує на те, що лише 16.8% варіації ціни може бути пояснена моделлю, а інші 83,2% пояснюються іншими факторами, що не входять в нашу модель. Це може вказувати на те, що інші фактори або змінні також впливають на ціну, які не враховані у моделі. Тобто за коефіцієнтом детермінації уже можемо зробити висновок, що наша модель є не досить точною, а отже в подальшому використовувати її для прогнозування буде не доцільним.

F-статистика (F-statistic) має значення $1.004e+04$, а ймовірність (Prob (F-statistic)) близька до нуля. Це свідчить про те, що модель в цілому є статистично значущою.

Маємо коефіцієнти t-статистики для кількості спалень (-35.284), кількості ванних кімнат (94.680) і розміру будинку (65.094). Значення кожної t-статистики в абсолютному значенні велике, всі вони мають низькі p-значення, які менше 0,05. Це вказує на те, що кожен із коефіцієнтів є статистично значущим. Зокрема, константа та всі три ознаки значущі для залежної змінної.

Значення асиметрії вказує на те, наскільки симетричний (або асиметричний) розподіл даних. Значення асиметрії в моделі 0.836 позитивне, що вказує на правосторонню асиметрію. Такий розподіл має довгий правий хвіст, і більше даних сконцентровані ліворуч від середнього значення. Зазвичай, позитивне значення асиметрії свідчить про те, що є більше вибірових значень, які вищі за середнє.

Значення ексцесу вказує на "товстість" хвостів розподілу та на наявність викидів. Значення ексцесу 3.330 є вищим за нормальний розподіл (який має ексцес рівний 0). Позитивне значення ексцесу вказує на те, що розподіл має більш товсті та вищі хвости, ніж нормальний розподіл.

Отже, ця модель надає базове розуміння впливу кількості спалень, ванних кімнат і розміру будинку на ціну нерухомості. Проте модель має недостатню точність, асиметрію та не нормальний розподіл даних. Загалом модель є статистично значущою та має статистично значущі коефіцієнти. Через недоліки моделі було вирішено не намагатись виконати прогнозування, оскільки він буде не якісним через асиметрію, та низьку пояснювальність залежної змінної незалежними факторами.

Регресійний аналіз економічних показників використовує модель лінійної множинної регресії для пояснення залежної змінної "Середня ціна на нерухомість (млн доларів США)" за допомогою ВВП на душу населення, річної інфляції, попиту та пропозиції на нерухомість та середнього доходу домогосподарств (рис 2.16).

OLS Regression Results

Dep. Variable:	Середня ціна на нерухомість (млн доларів США)	R-squared:	0.990
Model:	OLS	Adj. R-squared:	0.988
Method:	Least Squares	F-statistic:	869.2
Date:	Wed, 17 Jan 2024	Prob (F-statistic):	2.57e-44
Time:	21:26:37	Log-Likelihood:	64.673
No. Observations:	52	AIC:	-117.3
Df Residuals:	46	BIC:	-105.6
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	2.7216	4.163	0.654	0.516	-5.657	11.100
ВВП на душу населення	0.1601	0.049	3.249	0.002	0.061	0.259
Річна інфляція	-0.0671	0.017	-4.028	0.000	-0.101	-0.034
Середній дохід домогосподарства	1.466e-05	9.51e-06	1.542	0.130	-4.48e-06	3.38e-05
Пропозиція нерухомості (млн будинків)	-7.2117	1.052	-6.853	0.000	-9.330	-5.094
Попит на нерухомість (млн будинків)	-0.0656	0.103	-0.636	0.528	-0.273	0.142

Omnibus:	1.730	Durbin-Watson:	0.483
Prob(Omnibus):	0.421	Jarque-Bera (JB):	1.568
Skew:	-0.303	Prob(JB):	0.457
Kurtosis:	2.403	Cond. No.	2.66e+07

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.66e+07. This might indicate that there are strong multicollinearity or other numerical problems.

Рисунок 2.16 – Результати регресійного аналізу з економічними змінними

Джерело: складено автором на основі [43-46]

Рівняння регресії має вигляд:

$$Y = 2.7216 + 0.1601 * X1 - 0.0671 * X2 + 0.0000 * X3 - 7.2117 * X4 - 0.0656 * X5$$

де $X1$ - ВВП на душу населення, $X2$ - річна інфляція, $X3$ - Середній дохід домогосподарства, $X4$ - Пропозиція нерухомості (млн будинків), $X5$ - Попит на нерухомість (млн будинків).

Коефіцієнт константи дорівнює 2.7216, це може говорити про те що, якщо всі інші фактори будуть рівні нулю, ціна нерухомості буде мати значення 2.7216 млн. доларів США. Р-значення даної змінної досить велике (0.552), що вказує на те, що він може бути нестатистично значущим.

Розглянемо незалежні змінні, які входять до нашої моделі.

Коефіцієнт ВВП на душу населення дорівнює 0.1601, Коефіцієнт для ВВП на душу населення складає 0.1601, і його стандартна помилка є 0.049. Цей коефіцієнт статистично значущий ($P\text{-value} < 0.05$), що свідчить про те, що зі

зростанням ВВП на душу населення на 1 долар США середня ціна на нерухомість також зростає на 0.1601 млн. доларів США.

Щодо коефіцієнту інфляції, то він складає -0.0671, і його стандартна помилка є 0.017. Цей коефіцієнт також є статистично значущим ($P\text{-value} < 0.05$), що вказує на зворотний зв'язок між річною інфляцією і середньою ціною на нерухомість: зі зростанням річної інфляції на 1% ціна на нерухомість зменшується на 0.0671 млн. доларів США.

Коефіцієнт середнього доходу домогосподарства є додатним та близьким до 0 ($1.466e-05$) і не є статистично значущим ($P\text{-value} > 0.05$). Це може свідчити про те, що середній дохід домогосподарства має незначний вплив на середню ціну на нерухомість, тобто при зростанні середнього доходу домогосподарств на 1 долар США, середня ціна нерухомості буде зростати на $1.466e-05$ млн. доларів США.

Маємо ще одну змінну - пропозиція нерухомості її коефіцієнт від'ємний (-7.2117), що може означати, що при збільшенні пропозиції на нерухомість на 1 млн. будинків, ціна на нерухомість буде зменшуватись на 7.2117 млн. доларів США. Щодо стандартної помилки, вона дорівнює 1.052. Цей коефіцієнт $P\text{-value} < 0.05$, що свідчить про статистично значущий вплив на залежну змінну.

Коефіцієнт змінної попиту на нерухомість є також від'ємним (0.0656), з чого випливає, що при збільшенні попиту на нерухомість на 1 млн будинків, ціна буде зменшуватись на 0.0656 млн. доларів США. Що стосується р-значення, воно мале дорівнює 0.103. Однак цей коефіцієнт не є статистично значущим ($P\text{-value} > 0.05$), що може вказувати на те, що попит на нерухомість не має значущого впливу на середню ціну нерухомості.

Розглянемо параметри для оцінки нашої моделі.

R-squared, тобто коефіцієнт детермінації, має високе значення (0.990) вказує на те, що використані фактори на 99% пояснюють варіацію залежної змінної, інший 1% пояснюється факторами, які не враховані моделлю.

R-squared скоригований (Adjusted R-squared) - це виправлена версія коефіцієнта детермінації (R-squared), який враховує кількість предикторів у моделі

та компенсує можливі перекоси, що виникають при додаванні нових змінних. R-squared скоригований дорівнює 0,998, що свідчить про те, що велика частина варіації залежної змінної (середня ціна на нерухомість) пояснюється використовуваною моделлю лінійної регресії.

Значення F-статистики досить високим (869,2), а ймовірність (Prob (F-statistic)) надзвичайно мала ($2.57e-44$). Це свідчить про те, що модель регресії в цілому статистично значима. Тобто, є значуща різниця між позначеною моделлю та моделлю, де всі коефіцієнти регресії рівні нулю. Загалом, модель регресії є статистично значуща.

Значення асиметрії дорівнює -0.303, що вказує на те, що розподіл може мати невеликий лівий зсув, але загалом він не дуже асиметричний.

Значення ексцесу (Kurtosis) дорівнює 2.403. Зазвичай, значення ексцесу більше 3 свідчило б про "остріший" хвіст розподілу, але значення 2.403 може вважатися поміркованим. Це може вказувати на те, що розподіл має нормальну форму, без виражених відхилень у великих чи малих хвостах.

Отже, модель має високий рівень пояснювальної здатності (R-squared). F-статистика вказує на статистичну значущість моделі в цілому. Однак існують певні аспекти, які потребують уваги. Деякі змінні мають дуже малі коефіцієнти та не є статистично значущими, що може підкреслювати меншу важливість цих факторів, а також в моделі присутня лівостороння асиметрія.

РОЗДІЛ 3

ПРОГНОЗУВАННЯ ЦІН НА НЕРУХОМІСТЬ В США

3.1. Підготовка даних до прогнозування

Прогнозування ціни на нерухомість в США може бути важливим для різних стейкхолдерів та галузей економіки з різних причин. Наприклад, інвестори використовують прогнози цін для прийняття рішень щодо продажу, придбання чи утримання. Гарячі точки ринку та прогнози можуть допомогти їм визначити оптимальні часи для інвестування. Також прогнозування буде корисним для осіб, які розглядають можливість купівлі нерухомості, оскільки прогноз допоможе їм прийняти рішення коли і де купувати, для оптимального збереження вартості, або здешевлення нерухомості.

Перед тим, як розпочинати прогнозування потрібно перевірити наші дані на мультиколінеарність, гетероскедастичність та автокореляцію. Така перевірка потрібна для виявлення цих явищ в даних, та для того, аби позбутись цих аномалій. Це зробить наш прогноз більш якісним та надійнішим.

Для того аби перевірити дані на мультиколінеарність була використана теплова матриця кореляції (рис. 2.12).

За допомогою матриці кореляції можемо зробити висновок про те, що існує мультиколінеарність між змінними ВВП на душу населення та річною інфляцією(0,92), попитом на нерухомість(0,96) і пропозицією нерухомості(-0,98).

Для зменшення мультиколінеарності рекомендується виключити деякі змінні, які сильно корелюються або застосування методів для зменшення мультиколінеарності у моделі. Оскільки змінні які сильно корелюються між собою мають також вагомий вплив на залежну змінну, тобто середню ціну на нерухомість США, то метод виключення цих змінних з моделі вважається недоречним, тому що

це повипливає в подальшому на прогноз. Було прийнято рішення використовувати модель LASSO, яка ігнорує коефіцієнти, які сильно між собою корелюються.

Щодо гетероскидастичності, було проведено два тести, а саме тест Бройша-Пагана та тест Гольфреда-Квандта (рис. 3.1).

Тест Бройша-Пагана : 8.854167343348927
P-value: 0.11502423101102839
Тест Гольфреда-Квандта : 1.1875484232310392
P-value: 0.3267760002347241

Рисунок 3.1 – Результати тесту Бройша-Пагана та Гольфреда-Квандта

Джерело: складено автором на основі [43-46]

Статистика тестів є високою, але P-value є досить великим (більшим за 0.05). Це означає, що хоча є деякі ознаки гетероскедастичності, їхній ступінь вираженості не є досить сильним, щоб вважати результат статистично значущим.

Загальною тенденцією є те, що обидва тести не надають достатніх доказів про гетероскедастичність на рівні статистичної значущості 0.05 або навіть на більш високому рівні значущості. Це означає, що нульову гіпотезу про відсутність гетероскедастичності можна прийняти для цих даних. Можемо зробити висновок, що в даних є рівномірність або постійність дисперсії залишкових помилок моделі регресії. Це означає, що розкид значень залишкових помилок не може змінюватися від одного спостереження до іншого, що є дуже добре для подальшого прогнозування.

Що стосується перевірки даних на присутність в них автокореляції був побудований графік лагових затримок (рис.3.2) було виявлено закономірність і зроблено припущення щодо присутності автокореляції у даних.

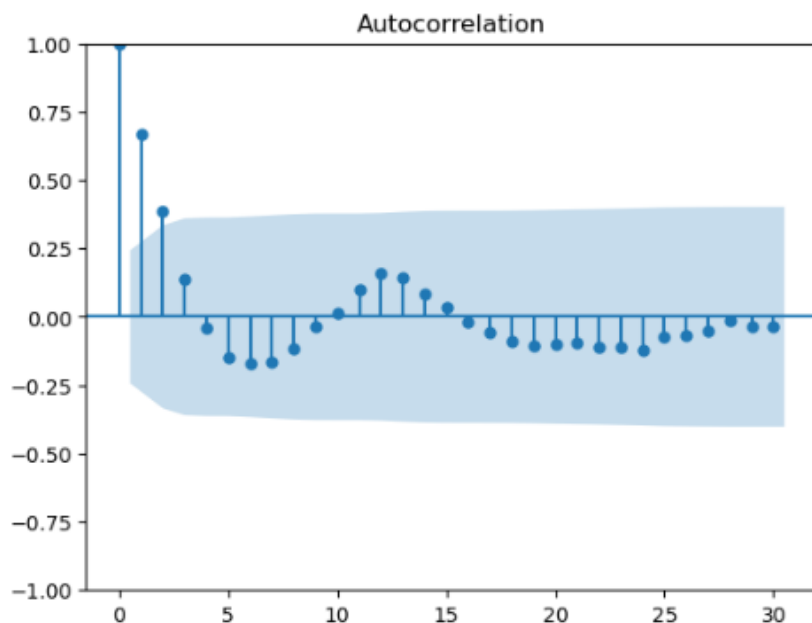


Рисунок 3.2 – Графік лагових затримок

Джерело: складено автором на основі [43-46]

Для більш точного визначення автокореляції був проведений тест Дарбіна-Уотсона (рис.3.3), який показав наявність позитивної автокореляції.

d1: 1.334, d2: 1.883 4-d1: 2.666, 4-d2: 2.117
Статистика Дарбіна-Уотсона: 0.47758967211465514
Позитивна автокореляція

Рисунок 3.3 – Результати тесту Дарбіна-Уотсона

Джерело: складено автором на основі [43-46]

Позитивна автокореляція вказує на те, що значення залежної змінної на певному часовому проміжку пов'язані позитивно з попередніми значеннями цієї ж змінної. Іншими словами, якщо виникає позитивна автокореляція, то високі (або низькі) значення змінної в певний момент часу супроводжуються високими (або низькими) значеннями в наступний момент часу.

Це може мати важливе значення в аналізі часових рядів та прогнозуванні, оскільки дозволяє передбачати, які зміни можуть відбутися в майбутньому на основі попередніх спостережень. Але при моделюванні важливо враховувати автокореляцію та можливі її впливи на надійність прогнозів.

Через присутність автокореляції в даних в подальшому для прогнозування будуть застосовуватись спеціальні методи, які враховують автокореляцію в даних, а саме модель ARIMA та експоненціальне згладжування.

Перед початком прогнозування, також хотілося б впевнитись щодо наявності сезонності та тренду в часовому ряду, тому було побудовано графік з якого було виділено тренд, сезонність та залишки (рис 3.4).

За графіком можемо побачити, що часовий ряд має тренд до зростання, що означає, що ціни на нерухомість з часом зростають. Щодо сезонності, то чітко бачимо що часовий ряд її має, а період сезонності дорівнює 12 місяцям. На початку року ціна на нерухомість знижується, що може бути спричинено податковою сезонністю. Багато покупців можуть очікувати на податковий повернення, яке часто надходить в перший квартал року. Покупці можуть відкладати свої рішення про покупку до отримання податкового повернення, що може призвести до тимчасового падіння попиту на початку року, а відповідно і ціни [51].



Рисунок 3.4 – Часовий ряд середньої ціни на нерухомість США з додаванням тренду, сезонності та залишків

Джерело: складено автором на основі [43-46]

Щодо залишків в часовому ряді є нестабільним, що означає - залишки не мають постійного середнього значення або дисперсії. Крім того, залишки мають

значну дисперсію. Це може бути ознакою того, що модель не враховує деякі випадкові фактори, які впливають на залежну змінну, тому відповідно до цього буде використана відповідна модель для прогнозування.

3.2. Вибір найкращого методу прогнозування цін на нерухомість в США

Вибір методу прогнозування є дуже важливим для подальшого прогнозу, оскільки від конкретного методу будуть залежити прогнозовані значення та їх точність. Тому враховуючи всі особливості нашої моделі були обрані такі методи:

- модель SARIMA, що врахує нашу сезонність та присутню автокореляцію в даних;
- модель експоненціального згладжування, що врахує автокореляцію в наших даних;
- модель LASSO, що враховує мультиколінеарність у наших даних.

За моделлю SARIMA прогноз вийшов досить точний. Прогноз добре вписується в дані, і він не має значних відхилень.

Показник MSE, що вимірює середню кількість квадратів відхилень між прогнозами та фактичними значеннями, є досить низьким для нашої моделі (0.009498). Це вказує на те, що середня кількість квадратів відхилень між прогнозованими значеннями та фактичними значеннями є невеликою. Це свідчить про те, що модель добре відтворює дані і має високу точність.

Значення MAE також є низьким (0.0851), що вказує на те, що середня абсолютна помилка між прогнозованими та фактичними значеннями є невеликою. Це ще раз підтверджує високу точність моделі.

Значення коефіцієнта детермінації дорівнює 0.8267, що означає, що близько 82.67% варіації ціни на нерухомість може бути пояснена за допомогою

використовуваної моделі. Високе значення свідчить про те, що модель добре підходить для даних та має високу прогностичну здатність (рис. 3.5).

Mean Squared Error (MSE): 0.009497718785970378
 Mean Absolute Error (MAE): 0.08509529702543248
 R²: 0.8266615037981648

Рисунок 3.5 – Показники якості моделі SARIMA

Джерело: складено автором на основі [43-46]

Загалом, прогноз SARIMA є досить точним і може бути використаний для планування майбутніх дій. Однак, важливо враховувати статистичну природу прогнозу і ймовірність того, що реальний результат буде відрізнятися від прогнозу (рис. 3.6).

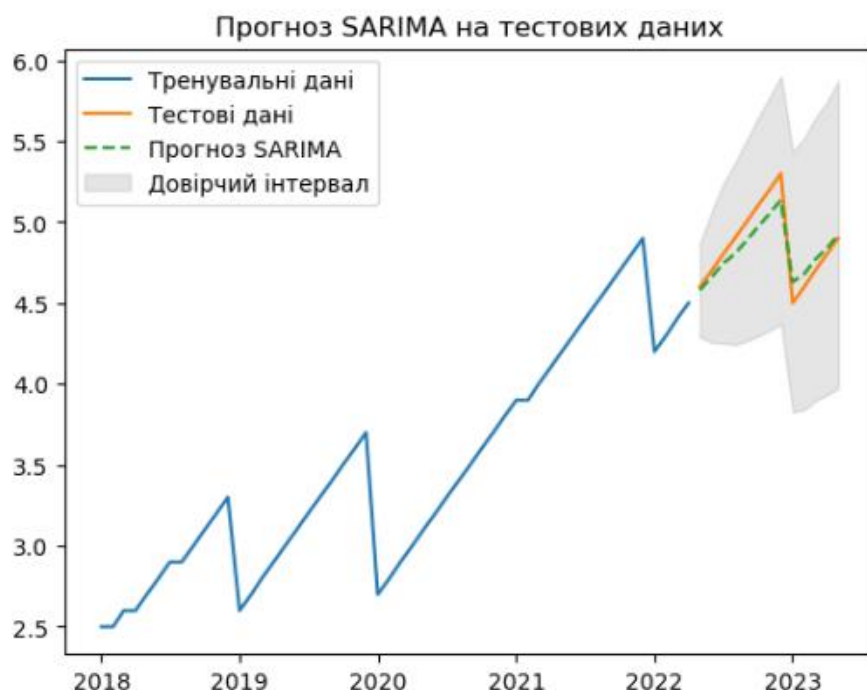


Рисунок 3.6 – Прогнозування за допомогою SARIMA

Джерело: складено автором на основі [43-46]

Інша модель, яка була використана для прогнозування середньої ціни на нерухомість - це модель експоненціального згладжування. Воно може допомогти зменшити вплив автокореляції, шляхом застосування згладжування наявних значень в часовому ряді. Оскільки в даних наявна автокореляція, яка виникає, коли значення в часовому ряді корелюють зі значеннями, що знаходяться на певній відстані в минулому, і може впливати на точність прогнозування, тому використати модель експоненціального згладжування буде доцільним для наших даних.

Результати прогнозування за допомогою цієї моделі є також досить точними (рис. 3.7), оскільки модель враховує також сезонність та тренд, а коефіцієнт експоненціального згладжування визначається автоматично за допомогою навчання моделі. Прогноз не має суттєвих відхилень.

```
Mean Squared Error (MSE): 0.010196132618053917
Mean Absolute Error (MAE): 0.07226280634905077
R^2: 0.8139150742493401
```

Рисунок 3.7 – Показники якості моделі експоненціального згладжування

Джерело: складено автором на основі [43-46]

Показник MSE є досить низьким для нашої моделі (0.0102), що вказує на те, що середня кількість квадратів відхилень між фактичними та прогнозованими значеннями є невеликою. Отже, модель добре відтворює дані і має високу точність.

Значення MAE також є низьким (0.07226), що вказує на те, що середня абсолютна помилка між прогнозованими та фактичними значеннями є невеликою. Що дає підстави нам зробити висновок про високу точність моделі.

Значення коефіцієнта детермінації дорівнює 0.8139, що означає, що близько 81.39% зміна ціни на нерухомість може бути пояснена за допомогою використовуваних у моделі незалежних змінних, а решта 18,61% пояснюється впливом інших факторів. Високе значення свідчить про те, що модель добре підходить для даних та має високу прогностичну здатність.

Отже, прогноз за допомогою експоненціального згладжування є досить точним, якщо порівнювати з моделлю SARIMA, то по коефіцієнту детермінації та MSE кращий показники має модель SARIMA, а інші показники кращі у моделі експоненціального згладжування (рис. 3.8).



Рисунок 3.8 – Прогнозування за допомогою експоненціального згладжування

Джерело: складено автором на основі [43-46]

Наступний метод прогнозування, який був використаний – це модель LASSO. Цей метод лінійної регресії враховує мультиколінеарність шляхом включення штрафу на суму абсолютних значень коефіцієнтів в функціонал втрат, використовуючи гіперпараметр α . Цей штраф допомагає зменшити або виключити вплив малозначущих ознак, що дозволяє моделі краще враховувати мультиколінеарність. LASSO надає перевагу обранню одного з корелюючих коефіцієнтів, зробивши інший нульовим.

Для початку дана модель визначила ваги кожної змінної (рис. 3.9).

```
const: 0.0
ВВП на душу населення(X1): 0.34045169729948743
Річна інфляція(X2): 0.0
Рівень безробіття(X3): -0.052762519311961575
Процентна ставка(X4): 0.0
Середній дохід домогосподарства(X5): 0.07718527241214077
Нетто міграції(осі6)(X6): 0.0
Пропозиція нерухомості (млн будинків)(X7): -0.35563903470734837
Попит на нерухомість (млн будинків)(X8): 0.05806493434334207
```

Рисунок 3.9 – Вагові коефіцієнти кожної змінної

Джерело: складено автором на основі [43-46]

Для більшої наочності ваги кожної змінної відобразили на графіку (рис. 3.10). Отож, стосовно коефіцієнтів:

- const: коефіцієнт дорівнює 0.0, що означає, що константа або початкове значення моделі відсутнє в результатах LASSO. Можливо, LASSO виключив це значення як несуттєве;
- ВВП на душу населення (X1): коефіцієнт становить 0.340, що свідчить про позитивний вплив ВВП на ціну нерухомості. Зі зростанням ВВП на душу населення на одиницю, середня ціна на нерухомість зростає на 0.340 млн. доларів США;
- річна інфляція (X2): коефіцієнт дорівнює 0.0, що може вказувати на те, що річна інфляція не має суттєвого впливу на ціни на нерухомість за умови використання LASSO;
- рівень безробіття (X3): коефіцієнт дорівнює -0.053, що свідчить про негативний вплив рівня безробіття на ціни на нерухомість. Зі зростанням рівня безробіття на одиницю, середня ціна на нерухомість зменшується на 0.053 млн. доларів США;
- процентна ставка (X4): коефіцієнт дорівнює 0.0, що може вказувати на те, що процентна ставка не має суттєвого впливу на ціни на нерухомість за умови використання LASSO;
- середній дохід домогосподарства (X5): коефіцієнт становить 0.077, що свідчить про позитивний вплив середнього доходу домогосподарства на ціни на нерухомість. Зі зростанням середнього доходу домогосподарства на одиницю, середня ціна на нерухомість зростає на 0.077 млн. доларів США;
- нетто міграція (осіб) (X6): коефіцієнт дорівнює 0.0, що може вказувати на те, що нетто міграція не має суттєвого впливу на ціни на нерухомість за умови використання LASSO;
- пропозиція нерухомості (млн будинків) (X7): коефіцієнт становить -0.356, що свідчить про негативний вплив пропозиції нерухомості на ціни на нерухомість. Зі зростанням пропозиції на нерухомість на одиницю, середня ціна на нерухомість зменшується на 0.356 млн. доларів США;

- попит на нерухомість (млн будинків) (X8): коефіцієнт становить 0.058, що свідчить про позитивний вплив попиту на нерухомість на ціни на нерухомість. Зі зростанням попиту на нерухомість на одиницю, середня ціна на нерухомість зростає на 0.058 млн. доларів США.

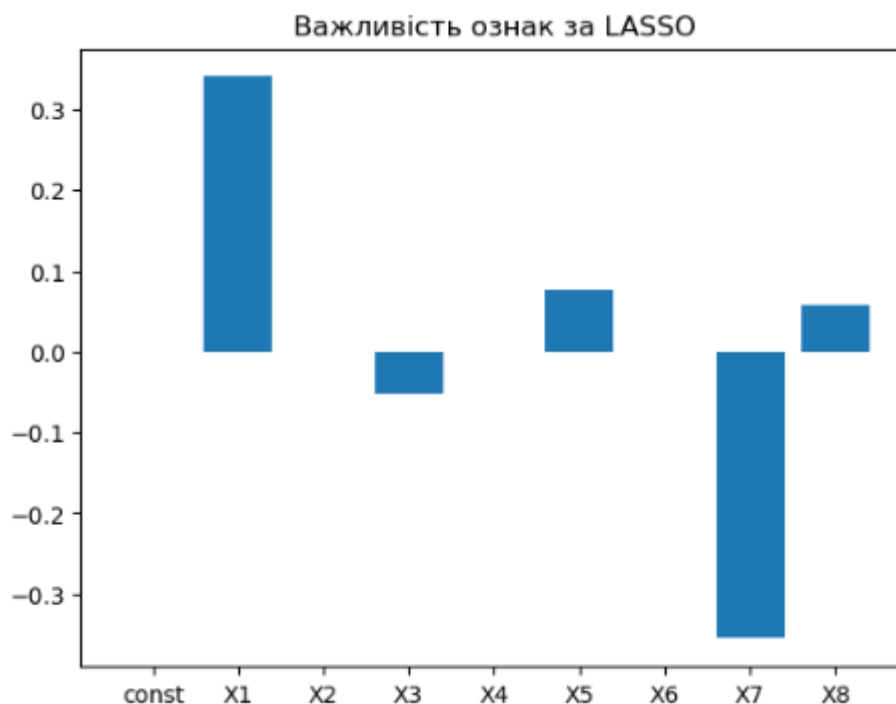


Рисунок 3.10 – Важливість ознак за LASSO

Джерело: складено автором на основі [43-46]

Результати прогнозування за допомогою цієї моделі є також досить точними (рис. 3.11), оскільки модель враховує мультиколінеарність. Прогноз не має суттєвих відхилень.

```
Mean Squared Error (MSE): 0.003823110490596913
Mean Absolute Error (MAE): 0.0513808048103265
R^2: 0.9948819259116692
```

Рисунок 3.11 – Показники якості моделі LASSO

Джерело: складено автором на основі [43-46]

Показник MSE є найнижчим відносно інших вже побудованих моделей (0.00382), що вказує на те, що середня кількість квадратів відхилень між фактичними та прогнозованими значеннями є невеликою. Отже, модель добре відтворює дані і має високу точність, практично без відхилень.

Значення MAE також є найнижчим з поміж інших моделей (0.05138), що вказує на те, що середня абсолютна помилка між прогнозованими та фактичними значеннями є найнижчою. Отож, це підтверджує, що модель має високу точність.

Значення коефіцієнта детермінації дорівнює 0.9949, що означає, що близько 99,49% зміна ціни на нерухомість може бути пояснена за допомогою використовуваних у моделі незалежних змінних, а решта 0,51% пояснюється впливом інших факторів. Значення коефіцієнта детермінації також є найвищим.

Отже, прогноз за допомогою моделі LASSO є найбільш точним в порівнянні з попередніми (рис. 3.12).

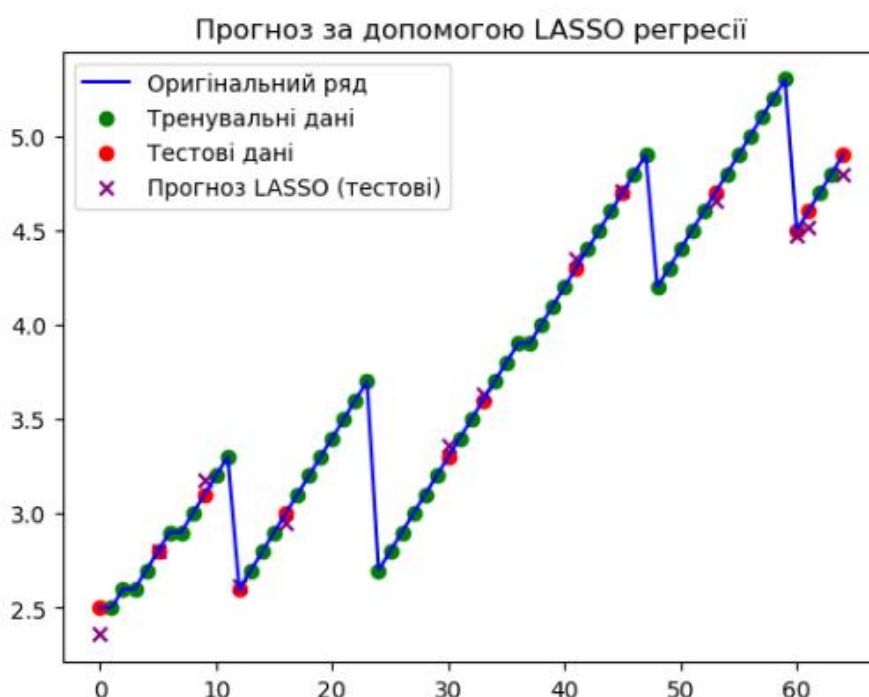


Рисунок 3.12 – Прогнозування за допомогою LASSO

Джерело: складено автором на основі [43-46]

Для кращого порівняння моделей побудуємо таблицю.

Таблиця 1

Вибір найкращої моделі для подальшого прогнозування

	SARIMA	Експоненціальне згладжування	LASSO	Яка модель найкраща?
R^2	0.8267	0.81391	0.9949	LASSO
MSE	0.009498	0.0101	0.00382	LASSO
MAE	0,0851	0.07226	0.05138	LASSO

Отож, по оцінкам якості моделі найкращою є LASSO регресія. На нашу думку, це саме через те, що даний тип моделі, підбирає найбільш значимі коефіцієнти для нашої залежної змінної та враховує незалежні змінні. Але для того, щоб спрогнозувати ціну на нерухомість на наступний рік, доцільніше буде обрати SARIMA, оскільки для прогнозування ціни на нерухомість по моделі LASSO нам потрібно знати всі показники на наступний період. Тобто модель LASSO доцільно застосовувати в тому випадку коли ми хочемо спрогнозувати ціну на нерухомість знаючи показники факторів, які впливають на ціну на нерухомість, а саме ВВП на душу населення, рівень інфляції, безробіття та інше, в модель навіть буде доцільно додати інші змінні, що допоможе клієнтам спрогнозувати ціну на квартиру за певних умов, та визначитись стосовно покупки цієї квартири.

3.3. Прогноз ціни на нерухомість на майбутній період

Обравши модель прогнозу SARIMA було спрогнозовано ціну на нерухомість США на наступні 12 місяців. Оскільки крайні дані були за травень 2023 роком, то прогноз був створений з червня 2023 року по травень 2024 року.

За графіком(рис. 3. 13) можемо побачити, що до початку 2024 року ціна на нерухомість піднімалась а на початку 2024 знову знизилась але до травня 2024 року знову зростає.



Рисунок 3.13 – Прогноз ціни нерухомості на наступні 12 місяців

Джерело: складене автором на основі [43-46]

Щоб більш детально розглянути дані виведемо таблицю з прогнозованими значеннями на наступні 12 місяців.

Таблиця 2

Прогноз ціни на нерухомість в США з 06.2023 по 05.2024

Дата	Середня ціна на нерухомість (млн доларів США)
2023-06-01	5.004418
2023-07-01	5.088715
2023-08-01	5.151791
2023-09-01	5.236731
2023-10-01	5.321970
2023-11-01	5.407494
2023-12-01	5.493290
2024-01-01	4.989929
2024-02-01	5.042232
2024-03-01	5.130372
2024-04-01	5.201154
2024-05-01	5.285968

Отож, бачимо що все таки на початку 2024 року спостерігалось падіння ціни, оскільки ціна на нерухомість є сезонним фактором. Зниження ціни на нерухомість в США на початку року може бути спричинена податковою сезонністю, а також взимку попит на нерухомість може бути меншим через несприятливі погодні умови, та через те, що багато сімей утримуються від переїзду через шкільний рік. Зменшення попиту може призвести до зменшення цін.

Модель прогнозування SARIMA показала дуже гарні результати та її можна використовувати в подальшому для прогнозування ціни нерухомості на наступні періоди.

ВИСНОВОК

У результаті проведеного дослідження було ретельно проаналізовано ринок нерухомості в Сполучених Штатах Америки, визначено та досліджено ключові фактори, які впливають на формування цін на нерухомість. Виявлено, що економічна ситуація, категорії нерухомості та інші параметри відіграють суттєву роль у цьому процесі. Це дало можливість сформулювати такі висновки:

1. Ринок нерухомості є одним із найбільших складових економічного багатства країни, на нього припадає близько 50% всього багатства країни. Ринок нерухомості має сезонну складову, та велику кількість факторів, що можуть на нього вплинути. Окрім цього ринок нерухомості привертає інвесторів, що є дуже добрим для економіки країни. Саме тому прогнозування ціни на ринку нерухомості є дуже важливим.
2. З неекономічних факторів, що впливають на ціну нерухомості найбільший вплив має кількість ванних кімнат. Щодо розташування, то найдорожчими районами є Массачусетс, Нью-Йорк та Нью-Гемпшир, оскільки мають велику кількість робочих місць у різних галузях, є густонаселеними і великими центрами економіки, культури та розваг, що створює попит на нерухомість.
3. З економічних факторів, що впливає на ціну на нерухомість найбільший вплив має ВВП на душу населення, річна інфляція, попит та пропозиція нерухомості, менший але також високий середній дохід домогосподарств. ВВП на душу населення, річна інфляція та дохід домогосподарств мають такий сильний вплив, оскільки ці фактори впливають на купівельну спроможність покупців, що в свою чергу формує попит на нерухомість, що уже й впливає безпосередньо на ціну нерухомості.
4. Проведено кластерний аналіз за допомогою якого вдалось розділити дані на 5 категорій, які відповідають потребам споживачів. За допомогою

кластерного аналізу та виявлення потреб конкретного споживача, можна пропонувати конкретну категорію. Це допоможе в подальшому ріелторам, покупцям та продавцям нерухомості, для того аби якнайшвидше знайти потрібну нерухомість для покупців та як найшвидше продати нерухомість для продавців.

5. За допомогою регресійного аналізу було виявлено, як саме кожна зі змінних впливає на варіацію ціни та був визначений датасет для прогнозування, а саме датасет який враховує саме економічні показники.
6. Розроблено різні моделі прогнозування, враховуючи специфіку датасету. Найкращою була визначена модель LASSO, яка враховує всі показники моделі та мультиколінеарність. Але дана модель підходить для прогнозування саме тоді, коли нам відомі вже інші фактори, які впливають на ціну, тому для прогнозування ціни саме на наступний період було обрано модель SARIMA. Вона враховує сезонність, тренд та автокореляцію в наших даних та є також досить точною.
7. Проведений прогноз за допомогою моделі SARIMA на період червень 2023-травень 2024. Він показав, що на початку 2024 року ціна на нерухомість буде дещо знижена, а потім зростає, на початку року ціна може бути знижена через процентну сезонність, що зменшує попит, а відповідно до цього і ціну на нерухомість.

Отже, у результаті роботи виконано всі поставлені задачі, а саме:

- проаналізовано стан сучасного ринку нерухомості в США;
- досліджено економічні чинники, розташування та інші параметри, що впливають на ціни на нерухомість;
- розроблено моделі для прогнозування змін у цінах на ринку нерухомості за допомогою різних методів;
- реалізовано побудовані моделі у програмному середовищі та проаналізувати їх точність на основі реальних даних;

- сформульовано висновки на основі створених моделей та отриманих прогнозів.

В подальшому є доцільним покращити модель LASSO, для того аби за заданими параметрами покупці та продавці могли побачити та скоригувати ціну на нерухомість.

Щодо моделі SARIMA, її також можна покращити, наприклад, розширити обсяг даних та врахувати інші фактори, які будуть доцільними. Це буде корисно для інвесторів, оскільки допоможе їм визначити періоди, коли краще збувати, а коли купляти нерухомість. Наразі можна сказати, що купувати нерухомість потрібно на початку року, оскільки в цей період найнижча ціна, а продавати можна вкінці року, оскільки тоді ціна є найвищою.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. The impact of today's higher interest rates on the housing market. Wealth Management U.S. Bank. URL: <https://www.usbank.com/investing/financial-perspectives/investing-insights/interest-rates-impact-on-housing-market.html>.
2. Сутність і значення ринку нерухомості. Studfiles. URL: <https://studfile.net/preview/8899125/page:38/>.
3. McMurray S. Real Estate. Wallstreetmojo. URL: <https://www.wallstreetmojo.com/real-estate/#h-real-estate-meaning>.
4. Global Peace Index 2022 - World. ReliefWeb. URL: https://reliefweb.int/report/world/global-peace-index-2022?gad_source=1&gclid=Cj0KCQiA2eKtBhDcARIsAEGTG42e98KAdHp4EtMVb7A_D3Ar9ko4a-UyTjrwSo1CY5FlxenEONgwPuMaAsgBEALw_wcB.
5. Стерненко М. Стейкхолдери: що таке і чому важливі. Smartik. URL: <https://smartik.kiev.ua/stejkkholdery-shcho-take-i-chomu-vazhlyvi/>.
6. U.S. Housing Market Conditions, First Quarter 2003: Summary, The importance of demographic trends to housing. Redirect. URL: https://www.huduser.gov/periodicals/usmc/spring03/summary_2.html#:~:text=Demographic%20changes%20in%20the%20population,housing%20units%20demanded%20and%20supplied.
7. Principles of Economics (Marshall book). Wikipedia, the free encyclopedia. URL: [https://en.wikipedia.org/wiki/Principles_of_Economics_\(Marshall_book\)](https://en.wikipedia.org/wiki/Principles_of_Economics_(Marshall_book)).
8. Шкурупій О. В., Білоброва Т. О. Циклічність розвитку ринку нерухомості : сучасний контекст : монографія. Полтава : ВНЗ Укоопспілки «Полтавський університет економіки і торгівлі» (ПУЕТ), 2012. 190 с. URL: <https://core.ac.uk/download/pdf/71925482.pdf>.
9. The Death and Life of Great American Cities. Wikipedia, the free encyclopedia. URL: https://en.wikipedia.org/wiki/The_Death_and_Life_of_Great_American_Cities.

10. Benjamin J. D., Sirmans G. S., Zietz E. N. Returns and risk on real estate and other investments: more evidence. Washington, 2016. 214p. URL: https://www.researchgate.net/publication/228690242_returns_and_risk_on_real_estate_and_other_investments_more_evidence.
11. Теорія соціокультурних систем, як основа вивчення особливостей і стандартів рекреації. Київ, 2021. С. 1–16. URL: https://moodle.znu.edu.ua/pluginfile.php/448686/mod_resource/content/1/скоіср%20к%20тема%201%20ml.pdf.
12. Економіка США зростає швидше, ніж очікувалося. Укрінформ - актуальні новини України та світу. URL: <http://surl.li/pclnj>.
13. Bachman D. United States economic forecast. Deloitte Insights. URL: <https://www2.deloitte.com/xe/en/insights/economy/us-economic-forecast/united-states-outlook-analysis.html>.
14. Hyndman R. J., Athanasopoulos G. Forecasting: principles and practice. 2nd ed. Australia : Otexts, 2018. 382 p. URL: <https://otexts.com/fpp2>.
15. Santarelli M. Housing market crash 2008 explained: Causes & Effects. Norada real estate investments. URL: <https://www.noradarealestate.com/blog/housing-market-crash-2008/#:~:text=the%20housing%20market%20collapse%20of,and%20many%20businesses%20went%20bankrupt>.
16. Ринок нерухомості в США продовжив нормалізацію наприкінці 2022 року. Академія біржевої торгівлі в Україні - компанія Double Case. URL: <https://double-case.com.ua/analitika/novosti/rinok-nerukhomosti-v-ssha-prodovzhiv-normalizatsiyu-naprikintsi-2022-roku>.
17. 30-Year Fixed Rate Mortgage Average in the United States. Federal Reserve Economic Data. URL: <https://fred.stlouisfed.org/series/MORTGAGE30US>.
18. Issue papers on demographic trends important to housing. U.S. department of housing and urban development office of policy development and research, 2020. 159p. URL: https://www.novoco.com/public-media/documents/demographic_trends.pdf.

19. Wilkerson M. US housing markets are in turmoil. The epoch times. URL: <https://www.theepochtimes.com/bright/us-housing-markets-are-in-turmoil-4882804?welcomeuser=1>.
20. CFI Team. Real Estate. Corporate Finance Institute. URL: <https://corporatefinanceinstitute.com/resources/commercial-real-estate/real-estate/>.
21. Візуалізація даних: що це таке і для чого вона потрібна. Ресурсний центр ГУРТ. URL: <https://gurt.org.ua/articles/37609/>.
22. Коробковий графік Вікіпедія. URL: https://uk.wikipedia.org/wiki/Коробковий_графік.
23. Коефіцієнт кореляції Пірсона Вікіпедія. URL: https://uk.wikipedia.org/wiki/коефіцієнт_кореляції_пірсона.
24. Як читати кореляційну матрицю. Statorials. URL: <https://statorials.org/uk/як-читати-кореляційну-матрицю/>.
25. Аббадія Дж. Використання регресійного аналізу для розуміння складних взаємозв'язків. Mind the graph. URL: <https://mindthegraph.com/blog/uk/перспективний-аналіз/>.
26. Regression model accuracy (mae, mse, rmse, r-squared). Datatechnotes. URL: <https://www.datatechnotes.com/2019/02/regressionmodel-accuracy-mae-mse-rmse.html>.
27. Фінансове моделювання. Лінійні моделі множинної регресії. Elib. URL: https://elib.lntu.edu.ua/sites/default/files/elib_upload/електронний%20посібник%20фм/page8.html.
28. Коефіцієнти асиметрії та варіації асиметрії. Pidru4niki. URL: https://pidru4niki.com/86411/ekonomika/koeffitsiyenti_asimetriyi_variatsiyi_asimetriyi.
29. Коефіцієнт ексцесу та варіації ексцесу. Pidru4niki. URL: https://pidru4niki.com/86411/ekonomika/koeffitsiyenti_asimetriyi_variatsiyi_asimetriyi.
30. Кластерний аналіз. Pidru4niki. URL: https://pidru4niki.com/11800912/ekonomika/klasterniy_analiz.
31. Мультиколінеарність Вікіпедія. URL: <https://uk.wikipedia.org/wiki/мультиколінеарність>.

32. Great Learning Team. A Complete understanding of LASSO Regression. Great Learning. URL: <https://www.mygreatlearning.com/blog/understanding-of-lasso-regression/>.
33. Тест Бройша-Пагана. Вікіпедія. URL: https://uk.wikipedia.org/wiki/Тест_Бройша-Пагана.
34. Параметричний тест Гольдфельда – Квандта. Studfiles. URL: <https://studfile.net/preview/8244516/page:2/>.
35. Гетероскедастичність залишків. Elib LNTU. URL: [https://elib.lntu.edu.ua/sites/default/files/elib_upload/Електронний%20підручник%20Кондіус%20\(MMICM\)/page11.html](https://elib.lntu.edu.ua/sites/default/files/elib_upload/Електронний%20підручник%20Кондіус%20(MMICM)/page11.html).
36. Автокореляція. Вікіпедія. URL: <https://uk.wikipedia.org/wiki/автокореляція>.
37. Тест Дарбіна-Уотсона. Stud.com. URL: https://stud.com.ua/151121/ekonomika/test_darbina_uotsona.
38. Автокореляція залишків. Elib LNTU. Url: [https://elib.lntu.edu.ua/sites/default/files/elib_upload/електронний%20підручник%20кондіус%20\(ммісм\)/page12.html](https://elib.lntu.edu.ua/sites/default/files/elib_upload/електронний%20підручник%20кондіус%20(ммісм)/page12.html).
39. Сутність методу прогнозування методом експоненціального згладжування. Бібліотека buklib.net. URL: <https://buklib.net/books/31392/#:~:text=одним%20з%20методів%20адаптивного%20прогнозування,мірі%20віддалення%20від%20кінця%20ряду>.
40. Модель авторегресії – ковзного середнього. Вікіпедія. URL: https://uk.wikipedia.org/wiki/модель_авторегресії_–_ковзного_середнього.
41. Brownlee J. A Gentle Introduction to SARIMA for Time Series Forecasting in Python. Machine Learning Mastery. URL: <https://machinelearningmastery.com/sarima-for-time-series-forecasting-in-python/>.
42. USA Real estate dataset. Kaggle: your machine learning and data science community. URL: <https://www.kaggle.com/datasets/ahmedshahriarsakib/usa-real-estate-dataset>.
43. National association of realtors. URL: <https://www.nar.realtor/>.
44. Federal reserve board. URL: <https://www.federalreserve.gov/default.htm>.
45. U.S. Bureau of labor statistics. URL: <https://www.bls.gov/home.htm>

46. Homeland security. URL: <https://www.dhs.gov/>.
47. Нестеренко І. Пропозиція і попит – основні чинники впливу на вартість нерухомості. Шляхи підвищення ефективності будівництва в умовах формування ринкових відносин : зб. наук. пр., м. Київ. 2015. С. 118–122. URL: <https://core.ac.uk/reader/322454540>.
48. Habinshuti P., Patrick M. Influence of gross domestic product (gdp) on real estate prices in rwanda. International journal of innovative research&development. 2020. Vol. 9, no. 4. P. 77–82.
49. Nguyen J. Is there a correlation between inflation and home prices?. Investopedia. URL: <https://www.investopedia.com/ask/answers/correlation-inflation-houses.asp>.
50. Податки в США у 2024 році для юридичних та фізичних осіб. Migrant.biz.ua - робота, міграція, туризм і навчання в країнах світу. URL: <https://migrant.biz.ua/ssha/finansy-ssha/podatky-v-ssha.html>.
51. Kiburz S. Real estate prices vs. Income. Crews bank & trust. URL: <https://www.crews.bank/blog/real-estate-prices-vs.-income#:~:text=in%20the%20past,%20an%20average,the%20ratio%20has%20a%20limit>.



Ім'я користувача:
Математичного моделювання та статистики Даценк...

ID перевірки:
1016085698

Дата перевірки:
27.01.2024 15:34:15 EET

Тип перевірки:
Doc vs Internet + Library

Дата звіту:
27.01.2024 15:38:37 EET

ID користувача:
100005744

Назва документа: Поліщук_Іщук_плагіат

Кількість сторінок: 70 Кількість слів: 13033 Кількість символів: 94373 Розмір файлу: 2.36 MB ID файлу: 1015797925

4.08%
Схожість

Найбільша схожість: 0.59% з Інтернет-джерелом (<https://foundry.kpi.ua/wp-content/uploads/2020/05/samaraj-vp-ekon...>)

2.98% Джерела з Інтернету 187 Сторінка 72

2.88% Джерела з Бібліотеки 434 Сторінка 73

0% Цитат

Вилучення цитат вимкнене

Вилучення списку бібліографічних посилань вимкнене

0%
Вилучень

Немає вилучених джерел