**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ**

**імені ТАРАСА ШЕВЧЕНКА**

****

**ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ**

**Кафедра прикладних інформаційних систем**

**Звіт до лабораторної роботи №2**

**з курсу**

**«Data Science та Big Data»**

*студента 4 курсу*

*групи ПП-32*

*спеціальності 122 «Комп'ютерні науки»*

*ОП «Прикладне програмування»*

*Симоніка Д.О*

*Викладач:*

*Шолохов О.В*

*Білий Р.О*

**Київ – 2022**

**Тема:** "**Розвідувальний аналіз даних (EDA). Складання аналітичного звіту**"

**Мета:** метою лабораторної роботи є отримання практичних навичок виконання розвідувального аналізу даних, використовуючи пакети jupyter, pandas, seaborn. Ознайомлення з методологією складання аналітичного звіту для зовнішнього користувача інформаційного продукту.

**Контекст:**

Ви – щойно нанятий data analyst у великій американській компанії, яка працює на ринку нерухомості США. На черговому засіданні ваш бос дав вам завдання зробити аналітичний звіт по цікавому йому сегменту ринку - Нью-Йорку

**Завдання:**

1. Виконайте дослідження domain experience стосовно американського ринку нерухомості. Ознайомтесь з декількома прикладами аналітичних продуктів від топових гравців на американському ринку, направлених на інвесторів. Питання, які потрібно опрацювати:
   1. Як топові компанії на ринку складають звіти по нерухомості?
   2. Які графіки використовуються для донесення інформації?
   3. Які співвідношення між якими даними по ринку є показовими для інвесторів / керівників агенцій нерухомості?
   4. Яка термінологія використовується для опису закономірностей на ринку нерухомості?
2. Завантажити файли з даними у папку проекту з посилання:

<https://www1.nyc.gov/site/finance/taxes/property-rolling-sales-data.page>

1. Очистити дані.
2. Виконайте розвідувальний аналіз, щоб дізнатися, де є викиди або відсутні значення, вирішіть, як ви їх обробляти, переконайтеся, що дати відформатовані правильно, значення, які ви вважаєте числовими, розглядаються як такі і т.д.
3. Виконайте аналіз розвідувальних даних (отриманих результатів) для візуалізації та зіставлення за житловими масивами та за часом. Почніть шукати осмислені закономірності у цьому наборі.
4. Зберіть висновки до невеликий звіт для генерального директора (графіки, висновки з текстом у окремому файлі), який потребує належного оформлення висновків, структури тощо.

**Хід роботи:**

**Завдання 1**

*a)* *Як топові компанії на ринку складають звіти по нерухомості?*

У світлі зростаючої важливості даних для прийняття обґрунтованих рішень у сфері нерухомості, вивчення звітів провідних компаній стає ключовим етапом для інвесторів та фахівців галузі.

Великі компанії на ринку нерухомості вирізняються високоорганізованим та методичним підходом до складання звітів, спрямованих на надання ключової інформації інвесторам та учасникам ринку. Перший етап цього процесу - це обширний збір даних, що охоплюють різні аспекти ринку нерухомості, такі як цінові тенденції, ринкова активність, характеристики об'єктів, географічні параметри та демографічні дані.

Після збору даних наступний етап - їхній аналіз. Великі компанії використовують розуміння ринкових тенденцій та використовують аналітичні інструменти для визначення ключових показників та факторів, що впливають на нерухомість. Під час створення звітів акцент робиться на прогнозуванні змін та наданні стратегічних рекомендацій для інвесторів.

Важливим елементом є також забезпечення доступності та зрозумілості отриманої інформації. Звіти представляються у вигляді чіткої та структурованої документації, яка дозволяє зацікавленим сторонам легко здобути необхідний уявлення про стан ринку та можливості для інвестицій. Такий комплексний та професійний підхід допомагає великим компаніям забезпечувати найвищий ступінь достовірності та корисності своїх звітів на ринку нерухомості.

*b) Які графіки використовуються для донесення інформації?*

У процесі створення звітів по нерухомості використовуються різні графічні елементи, які роблять дані більш доступними та легко зрозумілими. Лінійні графіки дозволяють інвесторам визначити тенденції цін з часом, стовпчасті графіки використовуються для порівняння різних районів за ціновою динамікою, а географічні карти вказують на розташування об'єктів у контексті інфраструктурних особливостей.

Подальший аналіз таких графіків може допомогти інвесторам швидше виявити перспективні напрямки для інвестицій або ризики в конкретних регіонах. Це візуальне підкріплення дозволяє учасникам ринку краще розуміти та швидше реагувати на зміни у сфері нерухомості.

*с)**Які співвідношення між якими даними по ринку є показовими для інвесторів / керівників агенцій нерухомості?*

У світі нерухомості визначення правильних стратегій і прийняття обґрунтованих рішень для інвесторів та керівників агенцій вимагає уважного розгляду певних ключових показників.

Відношення ціни до доходу - цей показник вказує на те, наскільки доступне житло для місцевого населення в порівнянні з їхнім доходом. Високе відношення може свідчити про можливу переоціненість ринку.

Відношення ціни до орендної плати - оцінює, якість інвестиції в орендну нерухомість, допомагаючи визначити, чи вигідно придбати чи орендувати житло.

Відношення попиту до пропозиції показує, як попит взаємодіє з пропозицією на ринку, що визначає його життєздатність та ймовірність вигідних угод.

Вартість життя та рівень доходу в регіоні показує співвідношення між вартістю життя та рівнем доходу, що визначає платоспроможність місцевого населення.

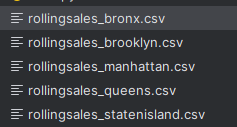
*d) Яка термінологія використовується для опису закономірностей на ринку нерухомості?*

У сфері нерухомості використовується специфічна термінологія для опису закономірностей та явищ. ***Цінова еластичність попиту*** означає наскільки змінюється кількість людей, які хочуть купити будинок, коли змінюється його ціна. ***Попит та пропозиція*** вказують на те, скільки людей хоче купити нерухомість та скільки її є на ринку для продажу. ***Цикл нерухомості*** описує різні етапи активності на ринку, включаючи часи зростання, спаду та стабільності. ***Вартість замовлення*** - це оцінка вартості будинку, а ***ринкова вартість*** - фактична ціна, за яку його можна продати. Аналіз порівняння допомагає визначити ціну, порівнюючи її з аналогічними об'єктами.

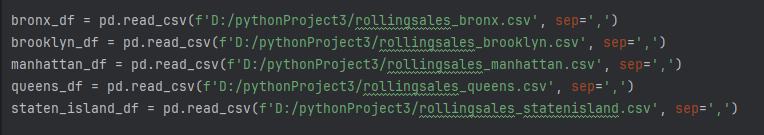
Щодо ***іпотеки та процентної ставки***, це стосується умов фінансування для покупки нерухомості, а ***дозвіл на забудову*** вказує на те, чи можна будувати на конкретній ділянці землі.

**Завдання 2**

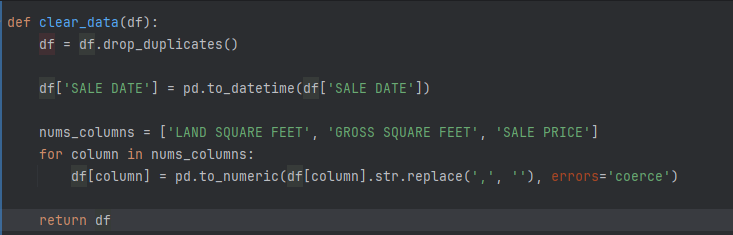
Завантажимо дані з посилання <https://www1.nyc.gov/site/finance/taxes/property-rolling-sales-data.page> у папку проекту. Попердньо переформатуємо xlsx в csv.



Об’єднаємо дані в один датафрейм



Очистимо дані. Для цього очистимо дані від дублікатів, приведемо дані до правильних типів даних, видалимо дані у яких відсутні необхідні значення (район,мікрорайон, ціна продажу, дата продажу). Для цього створимо функцію clear\_data:

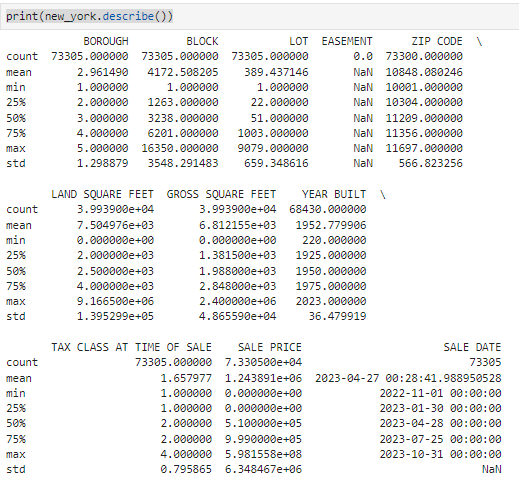


Перевіримо типи даних:

|  |  |
| --- | --- |
| До: | Після: |
|  | у |

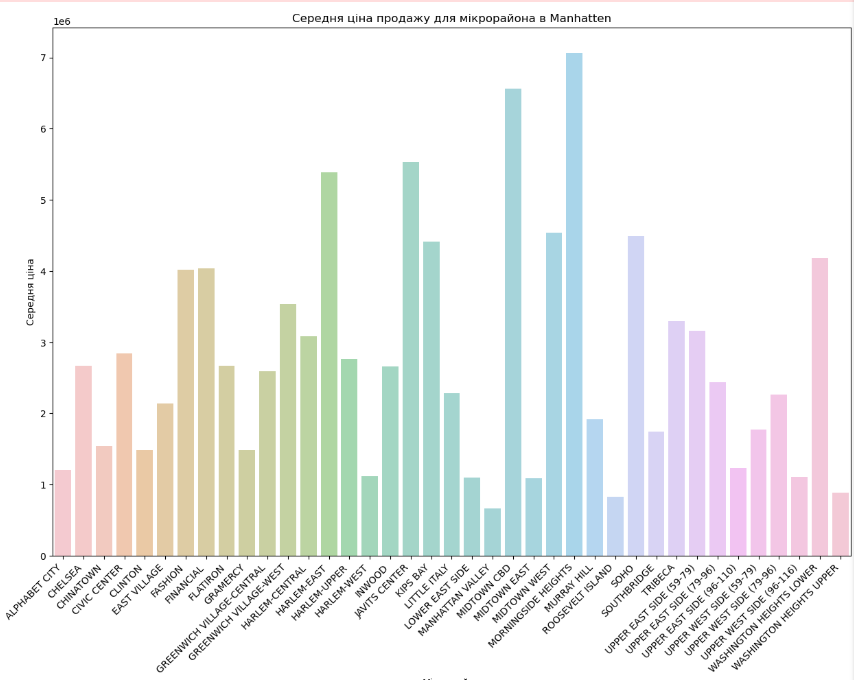
Проведемо розвідувальний аналіз даних

Отримаємо попередню оцінку даних (к-сть ненульових значень, максимум, мінімум, середнє значення, стандартне відхилення, процентиль 25%, 50%, 75% )

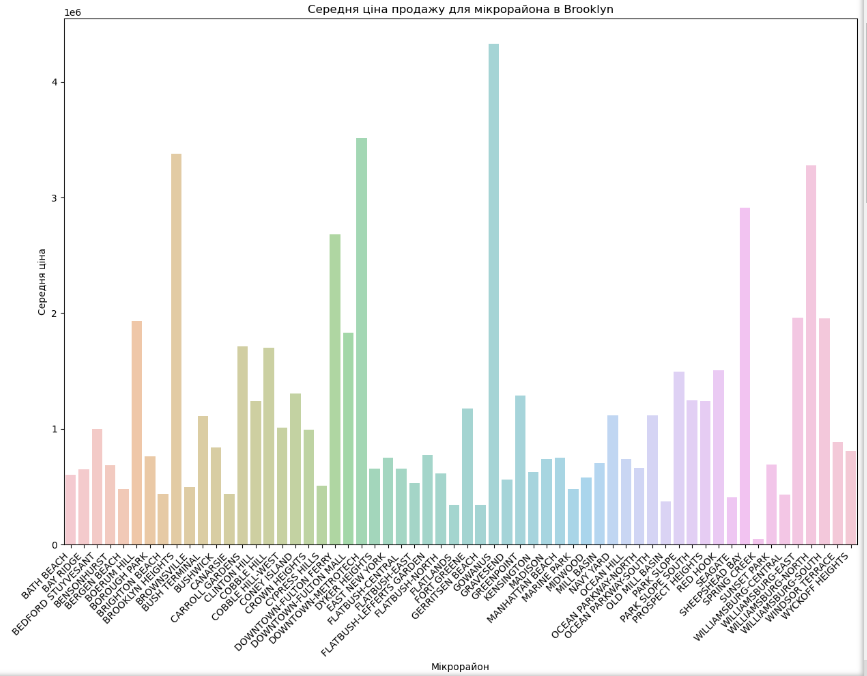


Далі дослідимо середню ціну для кожного мікрорайону района Нью-Йорка:

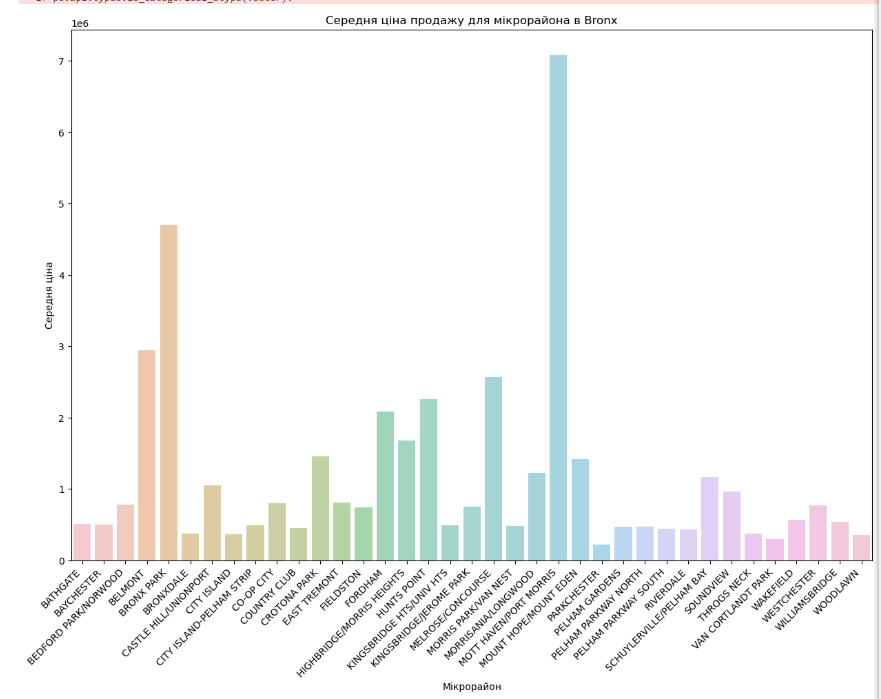




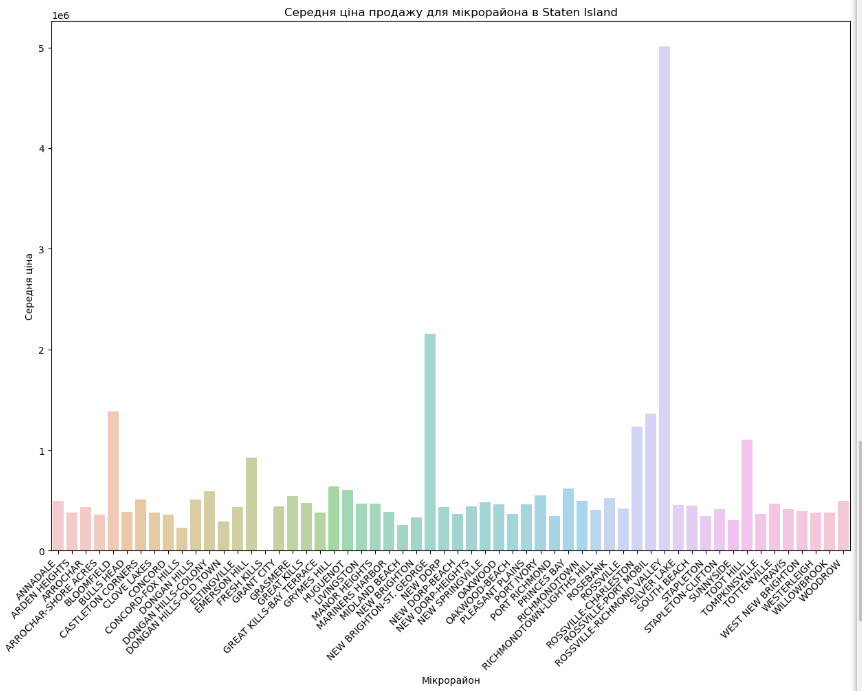
*Манхеттен*



*Brooklyn*



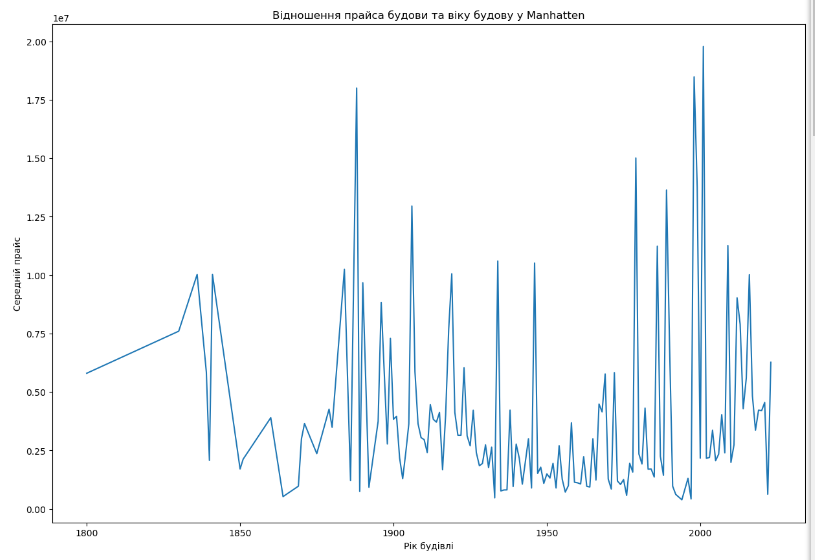
*Bronx*



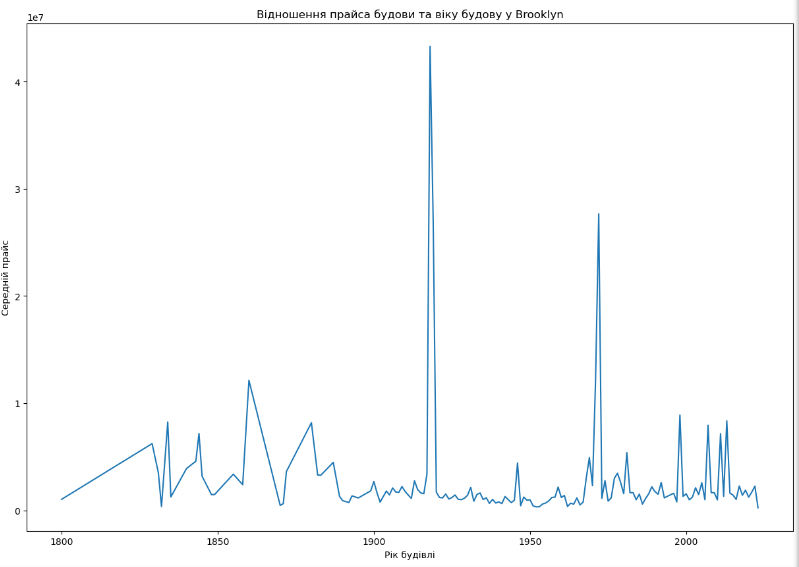
*Staten Island*

З візуалізації бачимо що найдорожчими мікрорайонами є Morningside Height у Манхетені, Govanus в Брукліні,Mott Haven/Port Morris в Бронксі та Rossville-Richmong Valley в Staten Island.

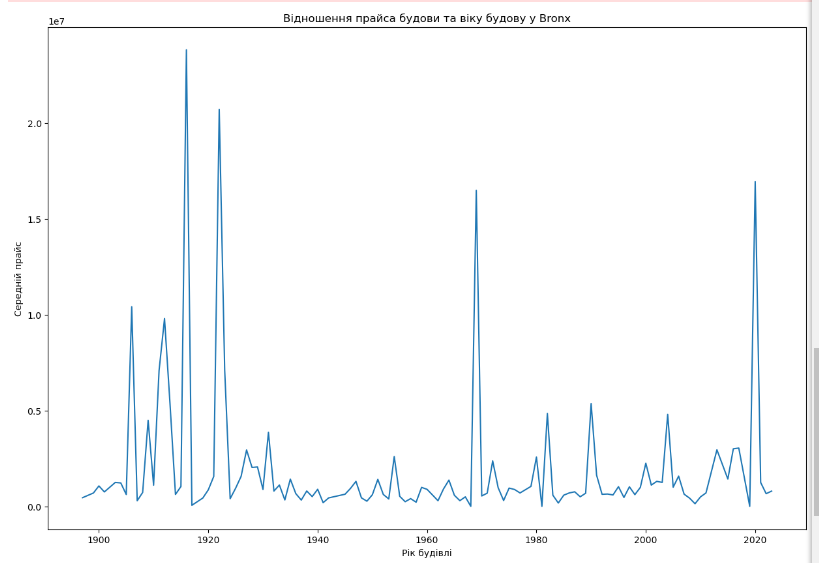
Дослідимо відношення ціни на будівлю до її віку



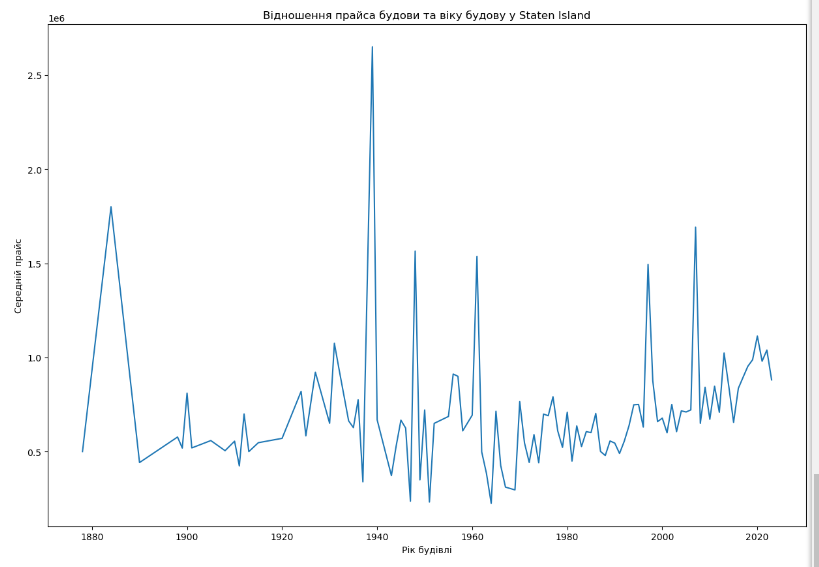
*Manhatten*



*Brooklyn*



*Bronx*

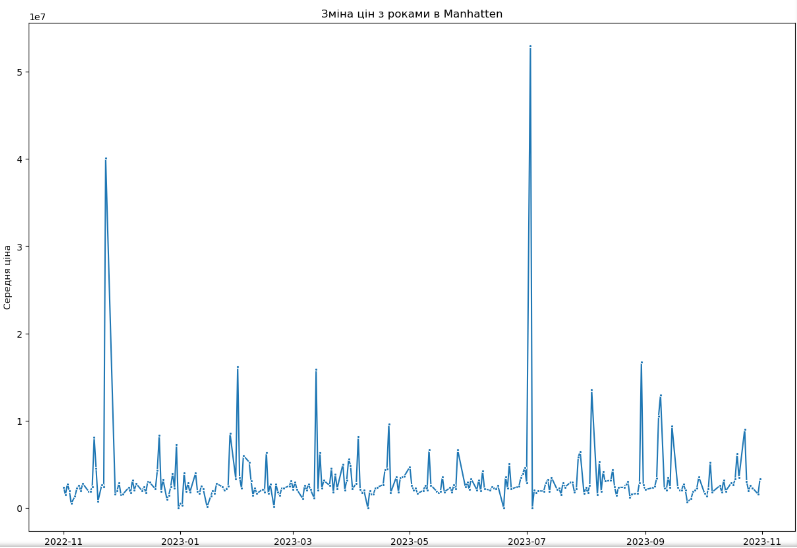


*Staten Island*

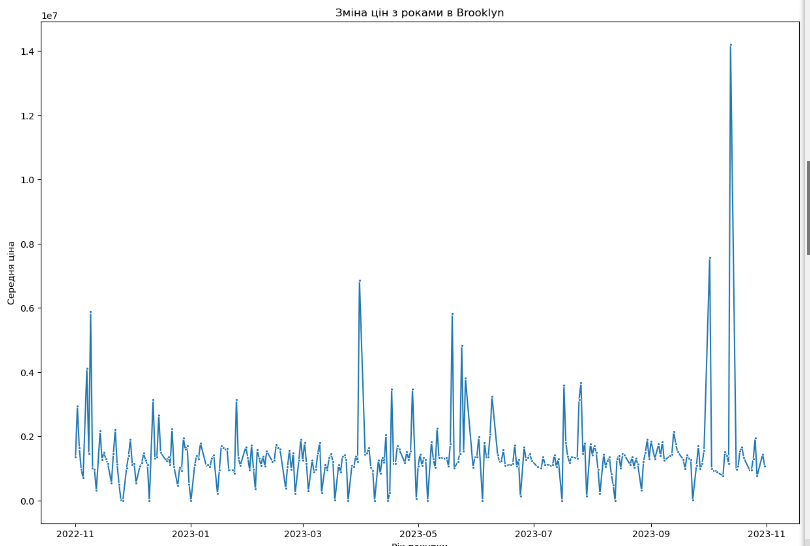
З графіків можна побачити що у Манхетені найбільш вартують будівлі кінця 19 століття та 1960х-70х років, у Брукліні – 1920х та 1970х років, у Бронксі в загальному найвищу ціну мають будівлі першох половини 19 століття, а у Стейтен Айленді – 1890 років, та будівлі періоду 1935-1965 років.

Далі дослідимо як мінялась середня вартість будівлі з часом

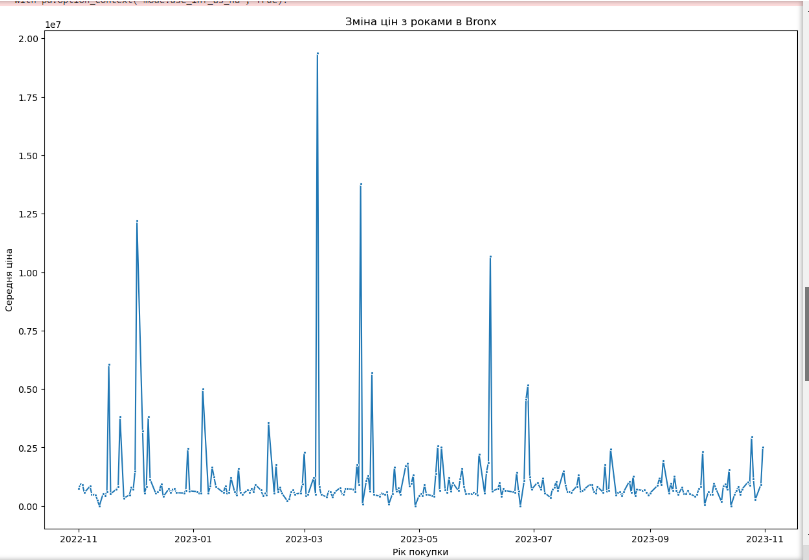




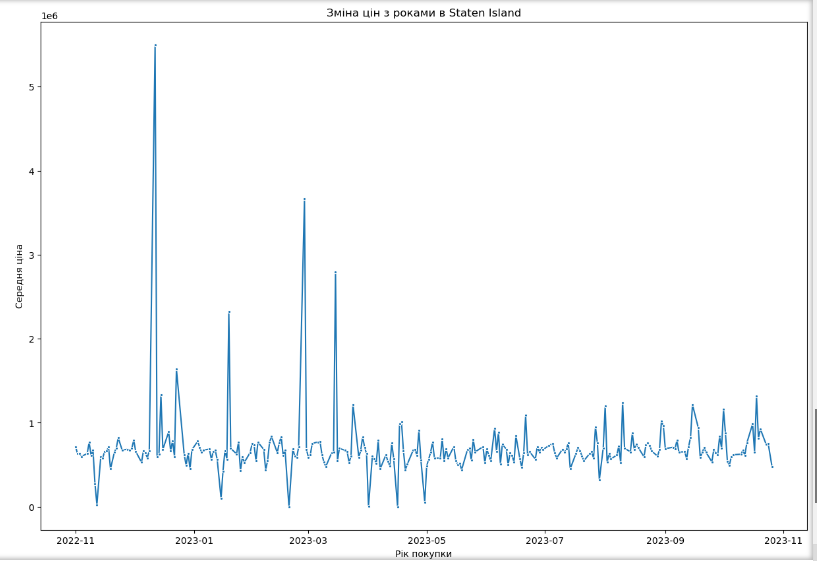
*Manhatten*



*Brooklyn*



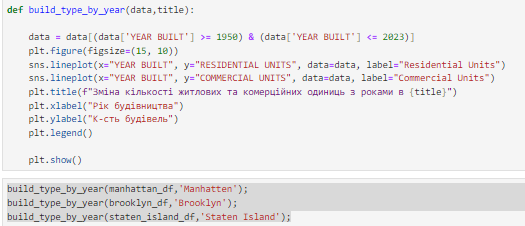
*Bronx*

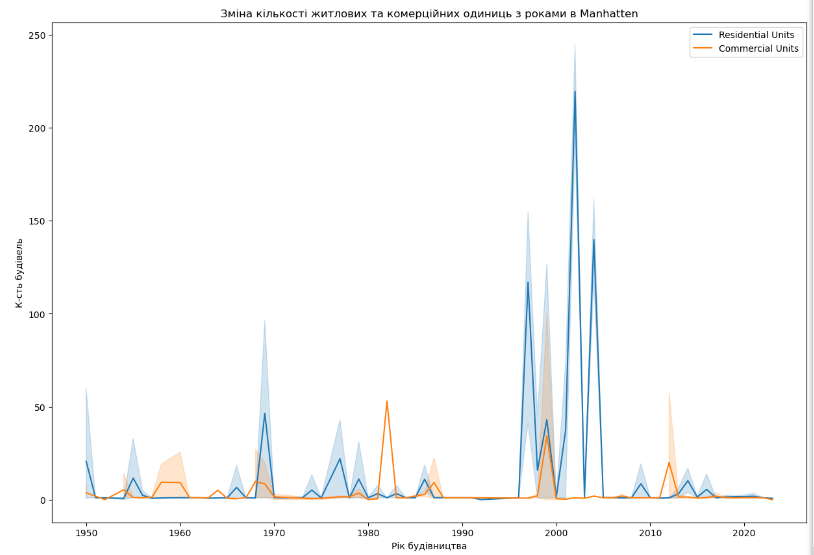


*Staten Island*

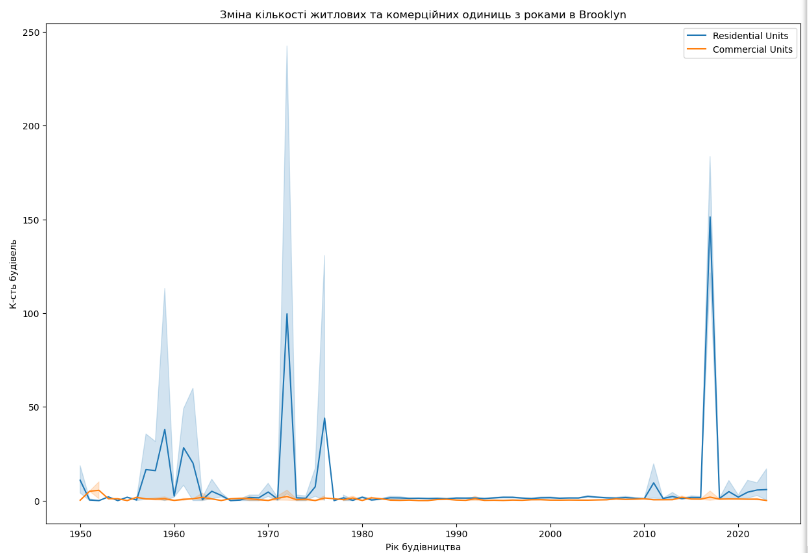
Згідно графіків, може зробити висновок, що середня сума, яку люди віддавали за будівлі у Манхетені досягла максимума у листопаді 2022 та липні 2023, у Брукліні досягла свого максимума за осінь 2023 року, виросши у 7 разів, у Броксі – максимальна середня сума була отримана у період 2023-03 по 2023-05, а у Стейтен Айленді з 2022-11 по 2023-01

Далі дослідимо зміну кількості комерційних та житлових будівель з часом:

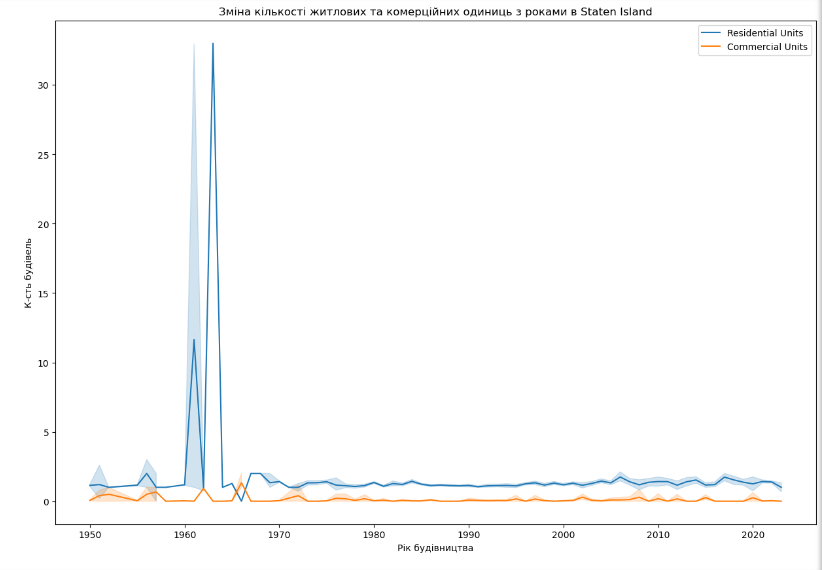




*Manhattern*



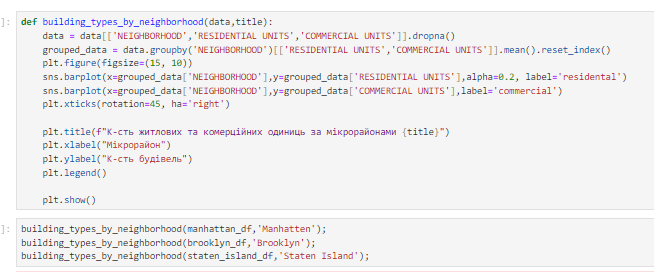
*Brooklyn*

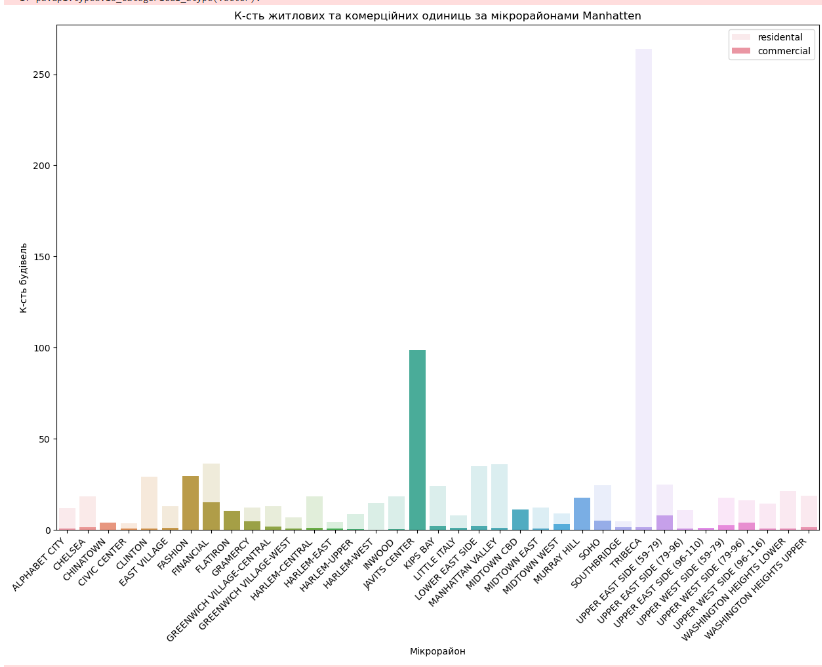


*Staten Island*

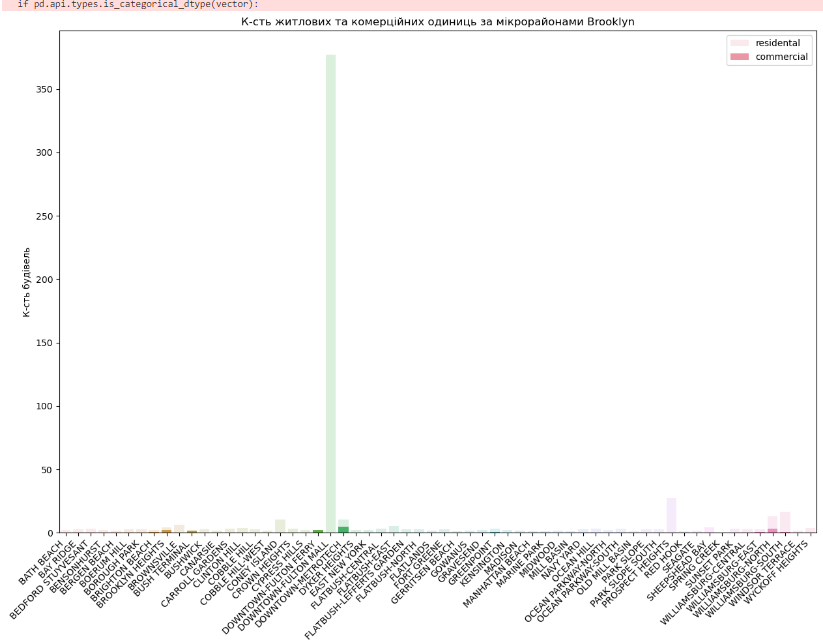
Згідно графіків, можемо сказати що у всіх районах, окрім Манхетена, переважає будування житла, а у Манхетені упродовж даного проміжку часу показник будування житла чергується з будування комерції, щоправда у період з 1995 по 2005 будування комерційних будівель значно випередило будівництво житла.

Дослідимо к-сть комерційних та житлових споруд по мікрорайонах

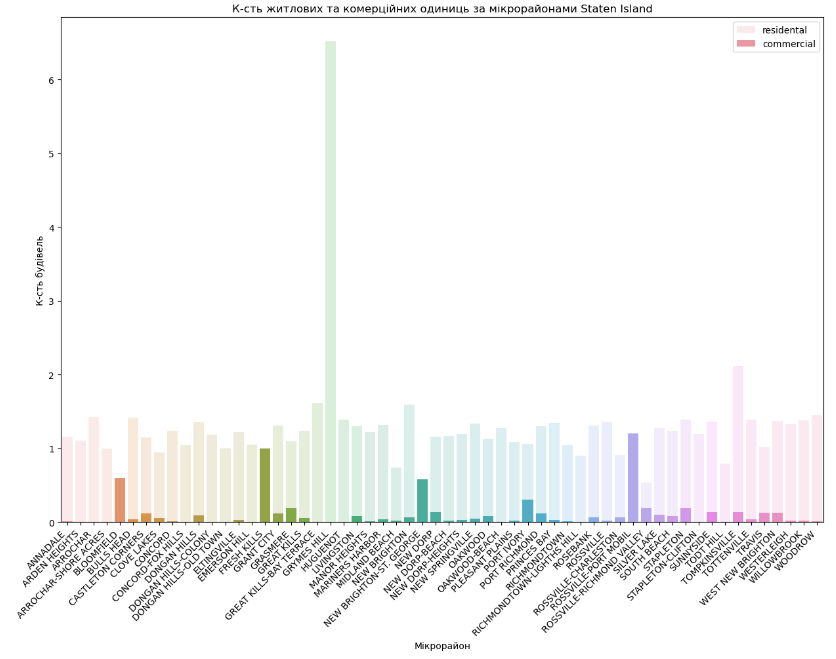




*Manhatten*



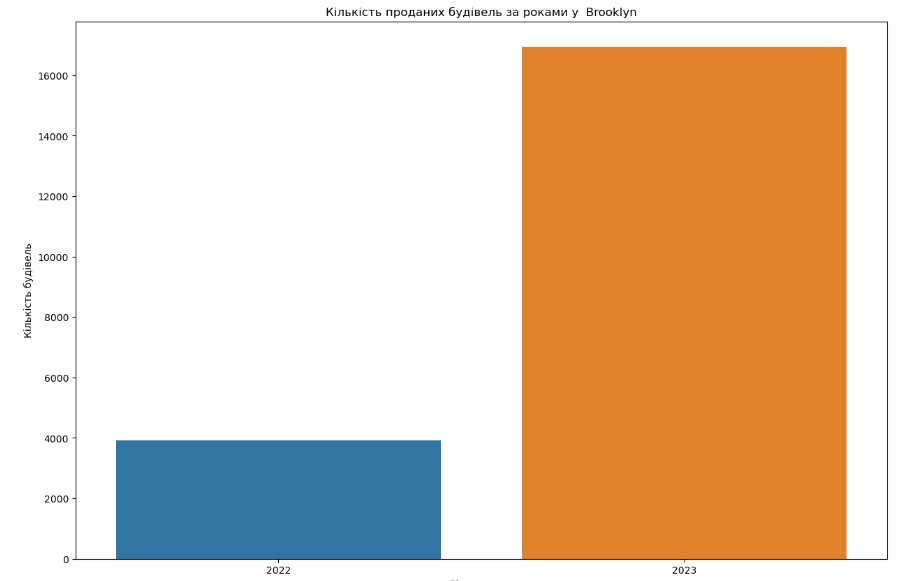
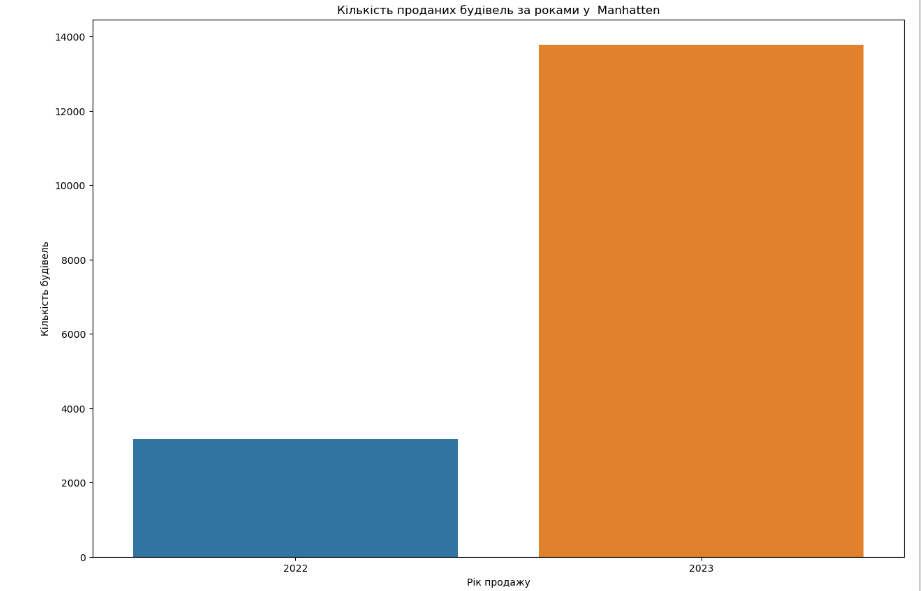
*Brooklyn*

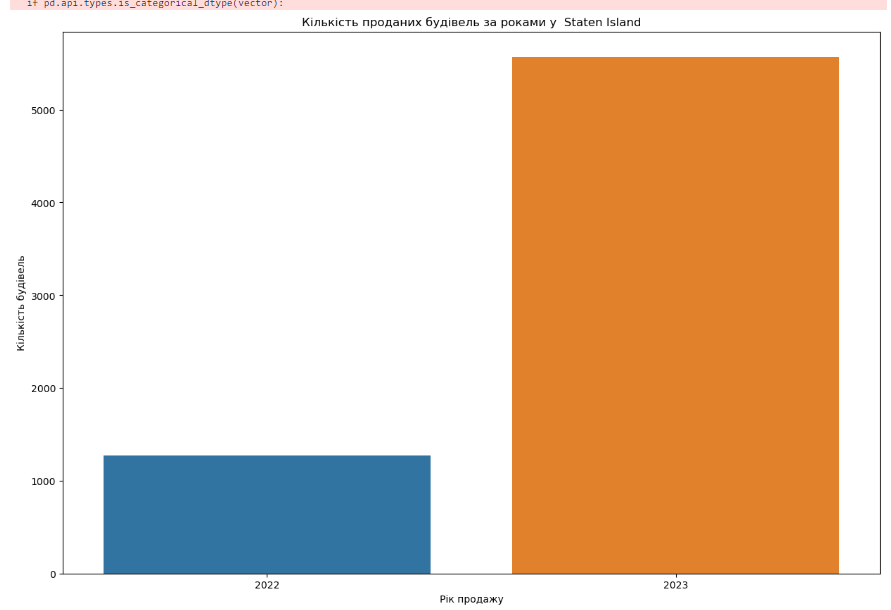
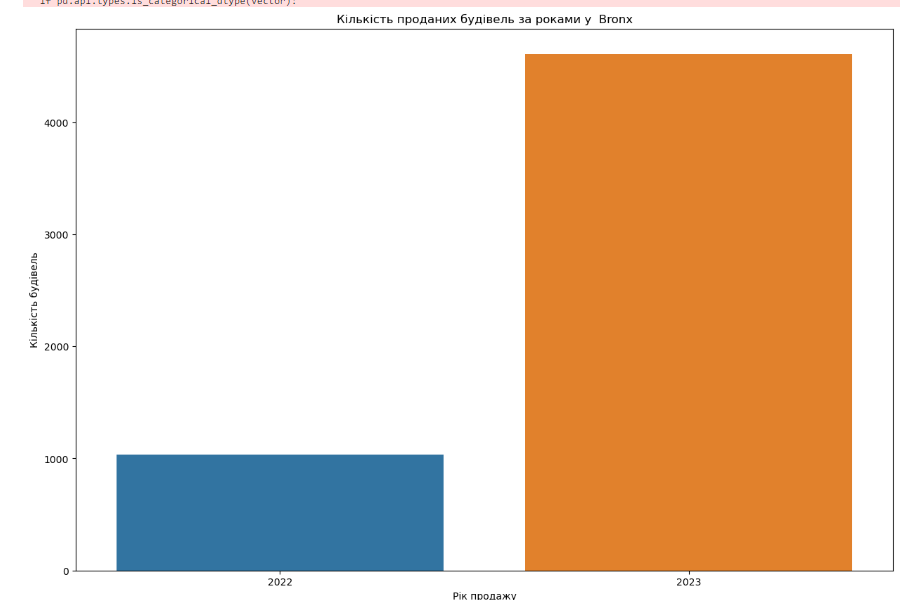


*Staten Island*

Бачимо, що у всіх районах, переважають житлові будівлі, проте у Манхетені виділяється мікро Javits Center, де переважають комерційні будівлі.

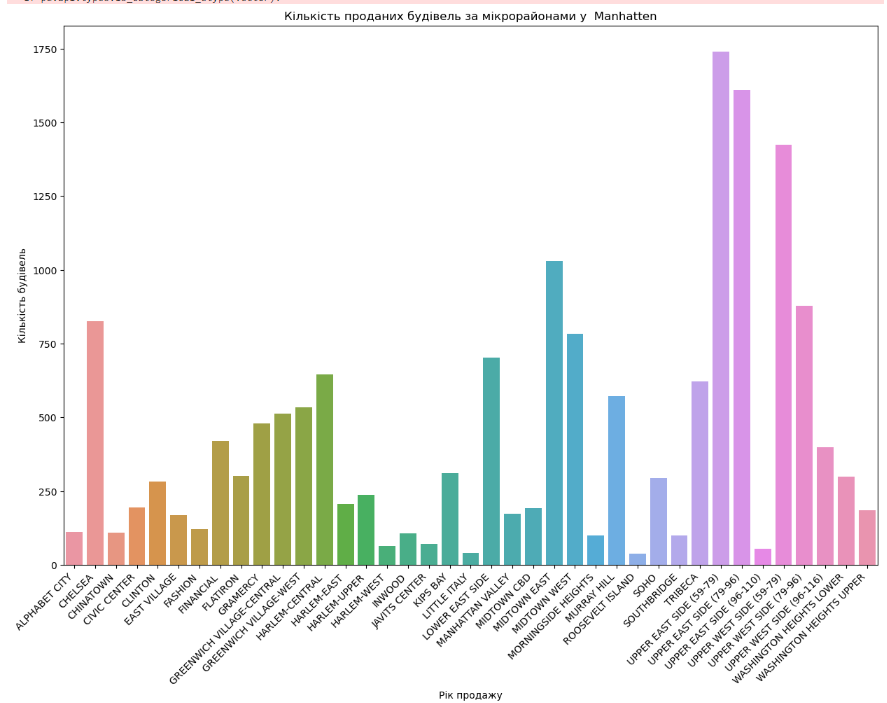
Відобразимо к-сть продажів будівель по мікрорайонам



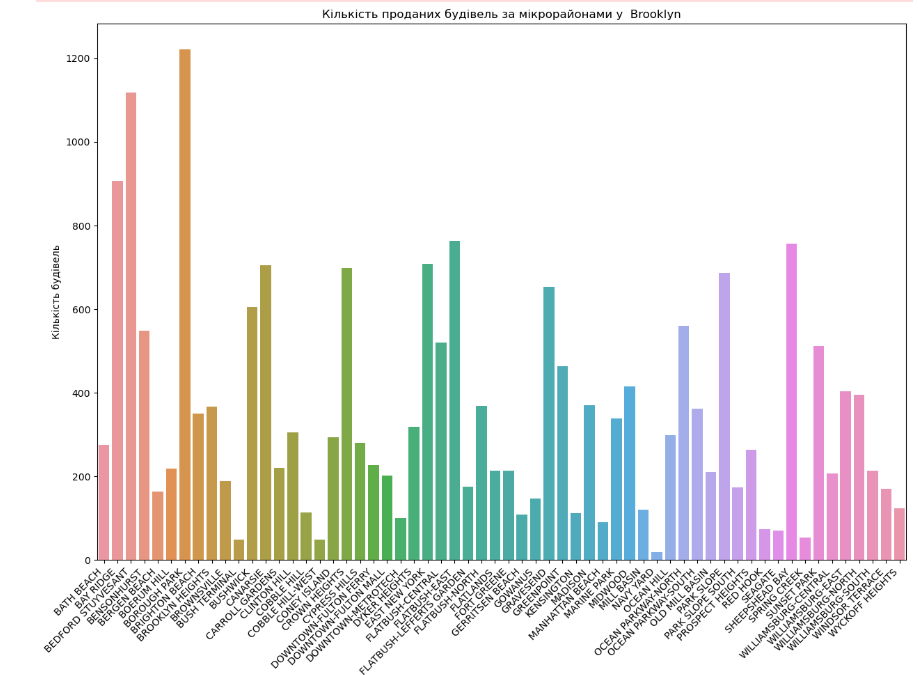


Бачимо, що продажі у 2023 році виросли у всіх районах, це пов’язано з реабіліацією світової економіки від кризи та з підвищення платіжеспроможності як компаній, так і людей.

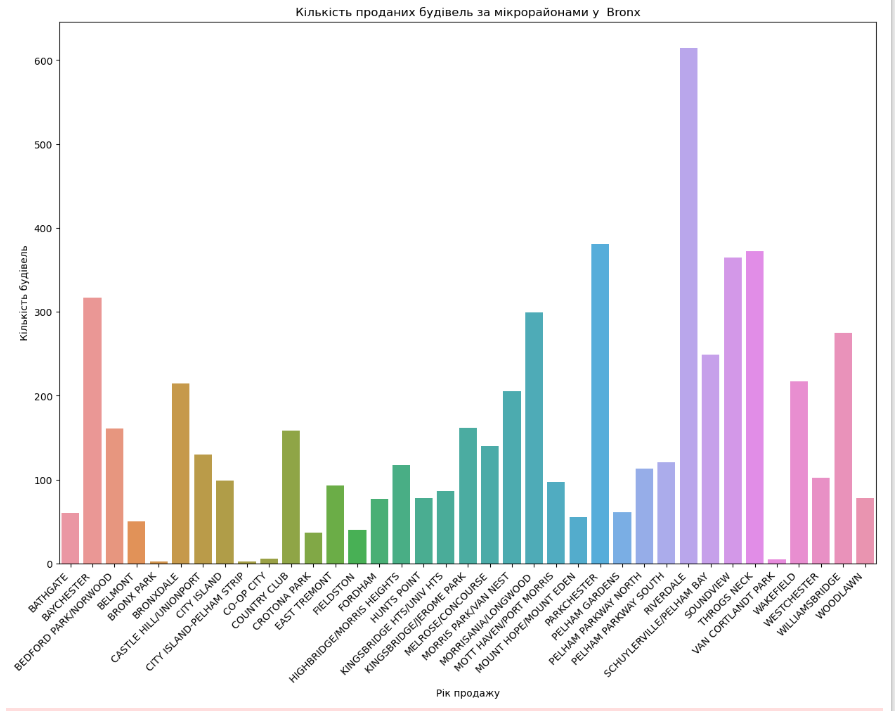
Дослідимо продажі будівель за мікрорайонами



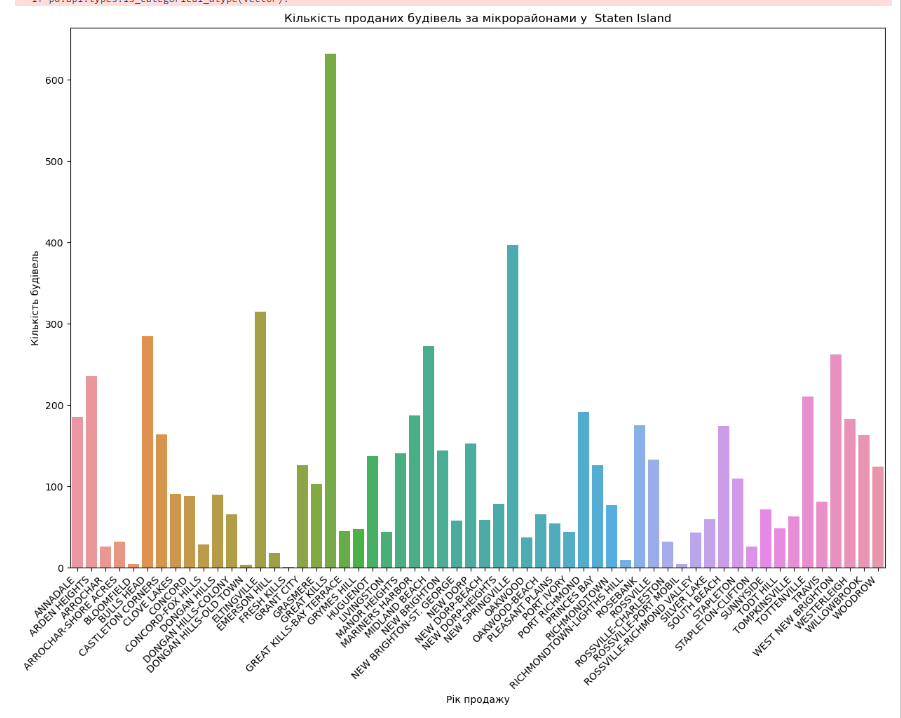
*Manhatten*



*Brooklyn*



*Bronx*



*Staten Island*

По мікрорайонах за найбільшою к-стю проданих будівель виділяються Upper East Side(79-96), Upper East Side(96-110), Upper West Side(59-96), Upper West Side(79-96) в Манхетені; Borough Park, Bedford Stuyvesant, Bay Ridge – в Брукліні; Riverdale - в Бронксі; Great Kills, New Springvill – в Стейтен Айленді.

**Висновок**

На основі проведеного ретельного дослідження ринку нерухомості в Нью-Йорку було виведено висновки, які мають ключове значення для прийняття важливих стратегічних рішень нашої компанії.

*Географічний аспект та його вплив на ціни*

Результати аналізу свідчать, що вартість нерухомості суттєво залежить від географічного розташування об'єктів. Ми можемо виділити окремі мікрорайони, такі як Morningside Height у Манхеттені, Govanus в Брукліні, Mott Haven/Port Morris в Бронксі та Rossville-Richmond Valley в Staten Island, де ціни перевищують середні значення.

*Важливість архітектурної спадщини*

Вивчення відношення ціни до року будівництва підкреслює, що історичні періоди великою мірою впливають на вартість нерухомості. Об'єкти з архітектурною спадщиною дуже цінується, і це слід враховувати при розробці наших стратегій.

*Динаміка будівництва та ринку*

Зміна кількості комерційних та житлових будівель з часом свідчить про гнучкість та реагування ринку на змінюючіться умови. Важливо враховувати цю динаміку при формуванні наших інвестиційних та розвиткових стратегій.

*Середня вартість будівлі та попит*

Дослідження зміни середньої вартості будівлі з роками вказує на періоди активності на ринку. Враховуючи ці піки, ми можемо адаптувати наші стратегії для оптимальних результатів.

*Відновлення попиту в 2023 році*

Позитивний тренд у продажах у всіх районах у 2023 році свідчить про відновлення економіки та підвищення платіжеспроможності. Це може відкривати нові можливості для наших інвестицій та проектів.

*Загальний висновок*

Ринок нерухомості у Нью-Йорку виявляється різноманітним та динамічним. На основі отриманих даних рекомендую взяти до уваги врахування географічних, архітектурних та економічних факторів при розробці наших подальших стратегій. Це дозволить нам ефективно реагувати на зміни на ринку та максимізувати наш потенціал у цій конкурентній сфері.