**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ**

**імені ТАРАСА ШЕВЧЕНКА**

****

**ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ**

**Кафедра прикладних інформаційних систем**

**Звіт до лабораторної роботи №2**

**з курсу «Data Science»**

*студента 4 курсу*

*групи ПП-43*

*спеціальності 122 «Комп'ютерні науки»*

*ОП «Прикладне програмування»*

*Трофимчука Мирослава Ігоровича*

*Викладачі:*

*ас. проф. Білий Р. О.*

**Київ – 2023**

**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.

***Виконання***

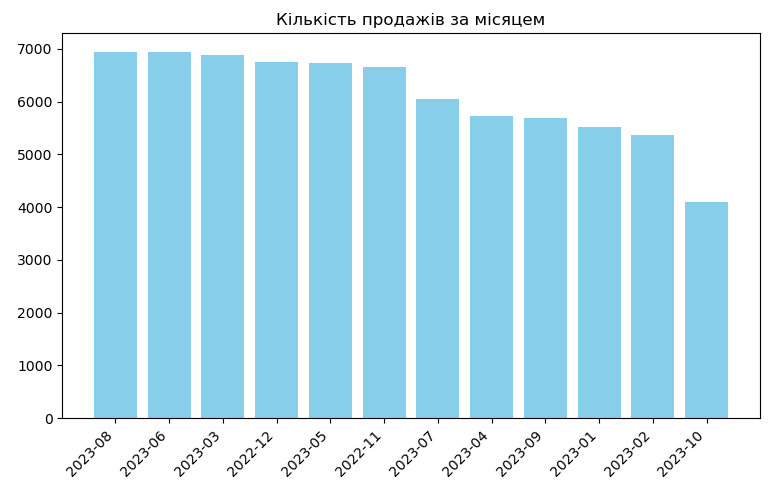
1. Проаналізуємо основний (початковий) датафрейм.

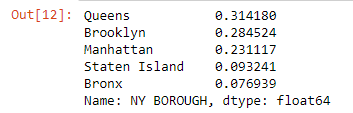


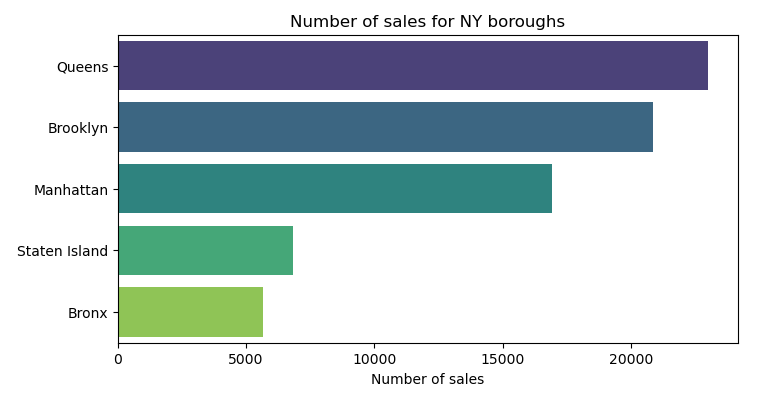
Можна побачити що всі наші дані розбиті на 19 стовпців і мають 73305 рядків. Саме стільки було здійснено операцій купівлі нерухомості в Нью-Йорку за період з листопада 2022 року по кінець жовтня 2023.

1. Проведемо розподіл продажів за часом, а саме за 12 місяців.

Як можна побачити найбільшими за к-стю продажів були місяці літа і весни 2023. Найменшими по к-сті продажів буди місяці зими та осені 2023 року, а саме січень, лютий та жовтень.

1. Проведемо розподіл за основними районами міста Нью-Йорк:

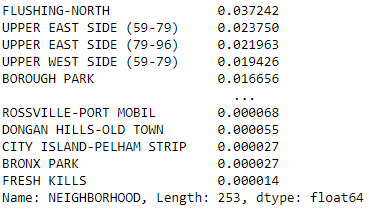


Можна побачити що найбільше нерухомості було продано в районах Queens та Brooklyn, 31% та 28% відповідно. Найменше в Бронксі – 0,77%. Для того щоб побачити реальні цифри можна скористатися діаграмою barplot:

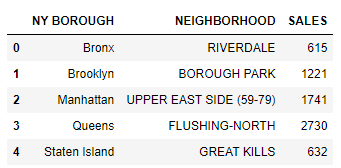
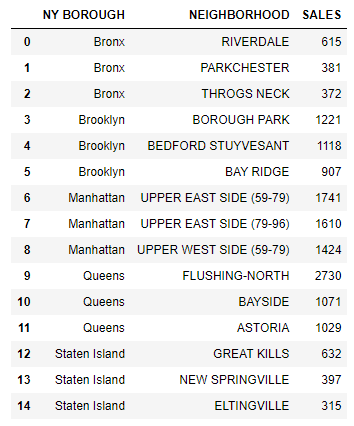
На долю Queens та Brooklyn випадає більше половини всіх продажів

(60% 43250), що робить ці 2 райони найпопулярнішими на ринку нерухомості Нью-Йорка.

1. Проведемо розподіл за мікрорайонами (neighbohood) Нью-Йорка:

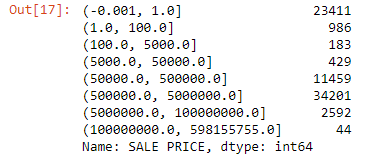


Можна побачити що у Нью-Йорку є 253 мікррайони. Найбільше операцій з продажу нерухомості припадає на Fushing-North, Upper East&West Side.

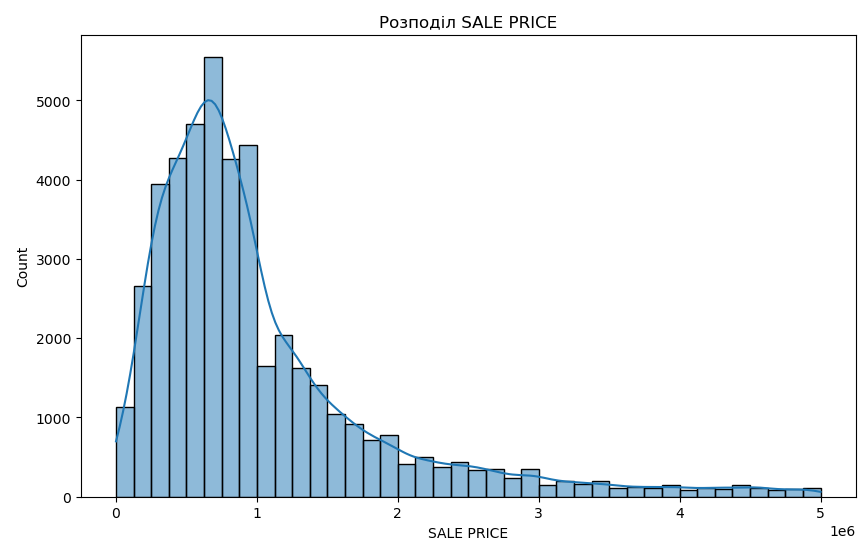
Для більше детального ознайомлення було створено спеціальні датафрейми, який відображають найпопулярніші мікрорайони кожного района за к-стю продажів:

Flushing-North - 2730 операції з продажу майна та Upper East & West Side – 4800 продажів для всіх.

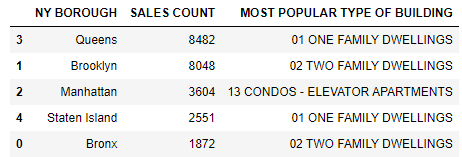
1. Тепер проведемо розподіл цін



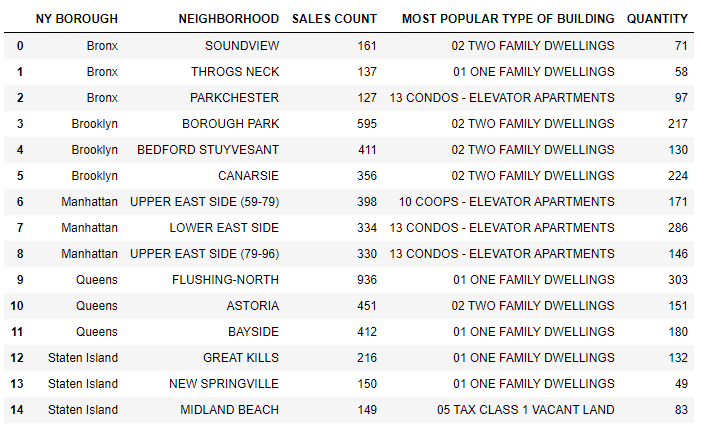
Можна побачити що 24600 (або 33,5%) операцій було здійснено за нульовими або близькими до нуля цінами. Одна з можливих причин - це відсутність даних або неповна інформація про реальну ціну продажу. Також, у деяких випадках, можливо використовували 0 або маленькі значення для позначення випадків, де об'єкт був переданий або проданий за символічну ціну.

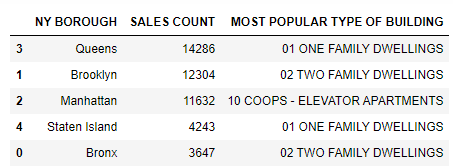
Але найбільш цікавим є те о найбільшу к-сть продажів було здійснено в діапазоні цін від 50 тис. до 2 млн. доларів. Діаграма histplot тому підтвердження:

1. Тепер подивимося в яких районах було здійснено найбільше операцій купівлі за символічними цінами (<5000$):



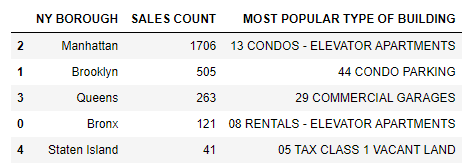
Абсолютними лідерами є Queens та Brooklyn. На них припадає 67% усіх продажів. Найпоширенішим типом нерухомості, що продавалися по таким цінам, було одно- та дво-сімейне житло (приватний будинок). На Манхетені це квартири у багатоповерхових будівлях з ліфтом.

1. Тепер проаналізуємо нерухомість за середніми цінами (від 5000$ до 5 млн $).



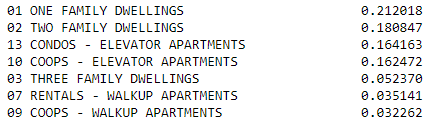
Найпоширенішою нерухомістю є знову приватні будинки та квартири.

1. Тепер нас цікавить нерухомість що була продана за високими цінами:



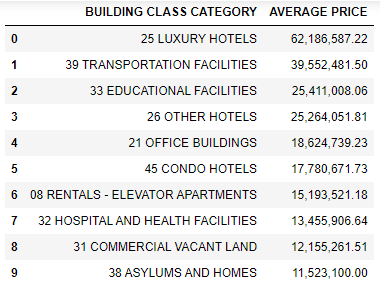
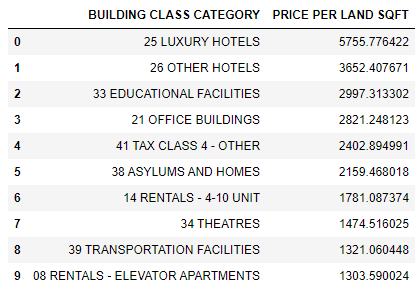
Її надзвичайно мало порівняно з іншою нерухомістю. І це не дивно адже ціни тут стартують від 5 і до 400 млн доларів. Найбільше продажів має Манхеттен ( 1700). Найпоширенішим типом нерухомості є квартири (в історичних будівлях), парковки (підземні), гаражі та інше майно.

1. Якщо проаналізувати продажі за типом будівель то ми приходимо до висновку що найбільше продаються в Нью-Йорку одно-, дво-сімейні приватні будинки та квартири (як з ліфтом так і без).

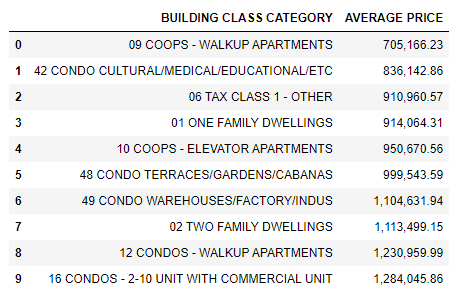
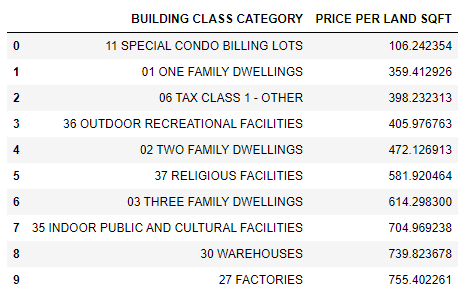


1. Далі зробимо порівняльний аналіз для топ-10 найдорожчих та топ-10 найдешевших типів нерухомості (будівель).

***Топ-10 найдрожчих***

 ***Без врахування площі Враховуючи площу***

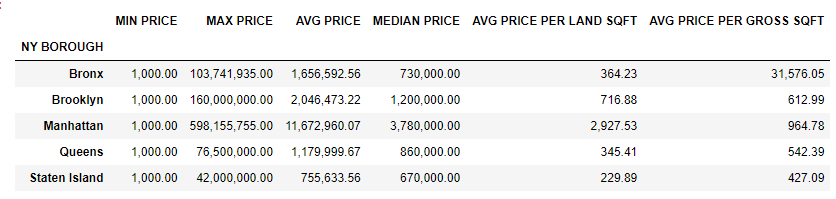
***Топ-10 найдешевших***

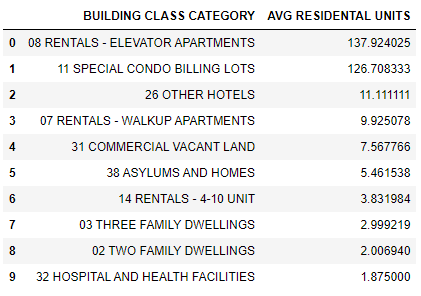
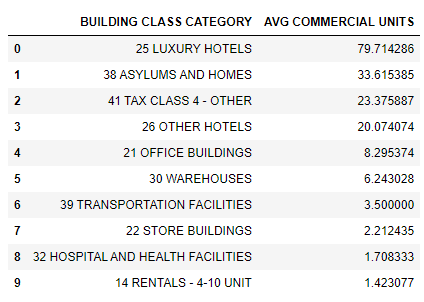
 ***Без врахування площі Враховуючи площу***

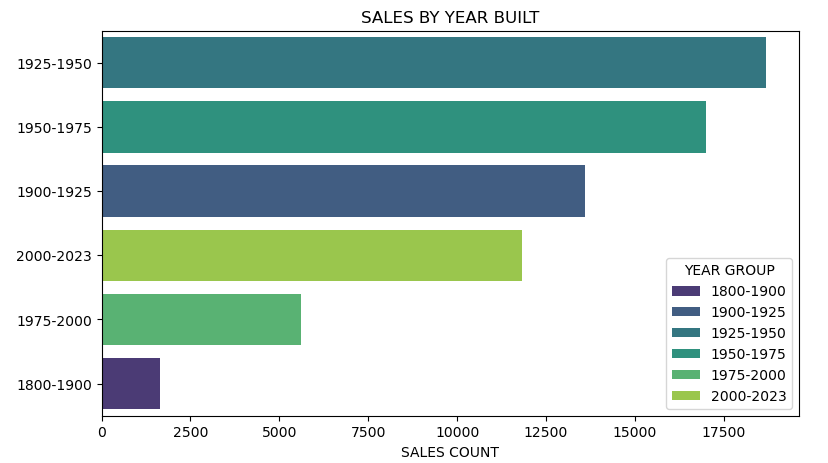
Можна побачити що найдорожчим майном є готелі, навчальні заклади, офісні будівлі, жд вокзали та автомобільні зупинки/станції.

Найдешевшими є квартири, приватні будинки та інше майно.

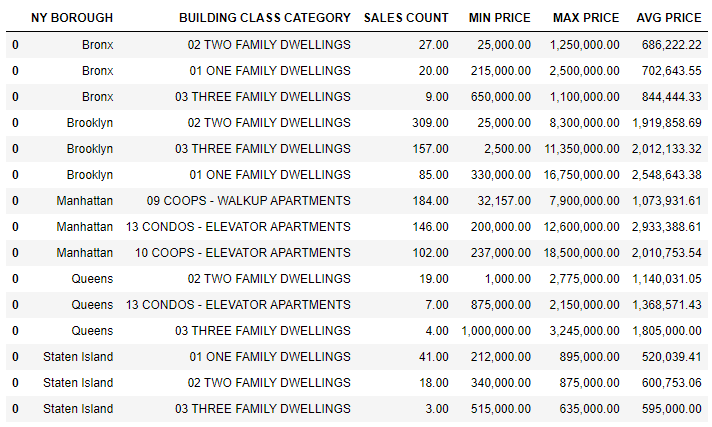
1. Проаналізуємо найпопулярніші мікрорайони. Обрахуємо для кожного з них середню, мінімальну та максимальну вартість майна. А також знайдемо вартість одиниці площі (фут квадратний) нерухомості.

Найдорожче майно є на Манхетені. Ціни там вимірюються в мільйонах. А середня вартість 1 фута квадратного в районі 2 тис доларів. Найдешевше майно в Бронксі. Там без проблем можна купити квартиру чи дім в якому фут квадратний = 400$.

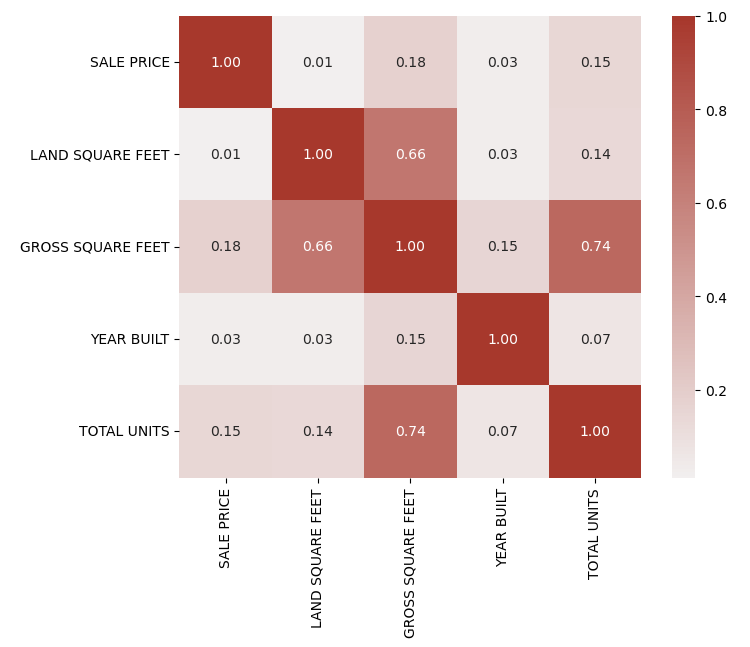
1. Проаналізуємо нерухомість яка має найбільше і найменше житлових та комерційних кімнат (units):
2. Проведемо розподіл цін на нерухомість за роком будівництва.

Найбільше було продано будівель, які були збудовані у 1925-1975 роках.

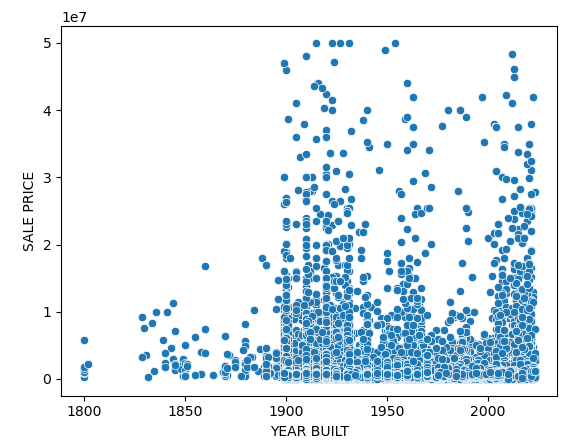
1. Можна знайти топ-3 типа будівель, що були збудовані у 1800-х роках.

Також подивитися ціни на них і їню к-сть за районами.

1. Знадейдемо к-сть власників що змінили тип оподаткування після купівлі нерухомості. Таких всього 6741.
2. Проведемо кореляційний аналіз за найважливішими стовпцями:



Можна побачити між ціною продажу і роком зведення будівлі немає зв’зку. Діаграма розподілу це підтверджує.



Також можна побачити, що є зв’язок між площею нерухомості та к-стю приміщень (кімнат).

