

Università degli Studi di Milano – Bicocca  
Data Science Master's Degree



# NEW YORK COUNTY HOUSING MARKET ANALYSIS

Students:  
Induni Sandapiumi Nawarathna Pitiyage – 906451  
Sara Campolattano – 906453



# INTRODUCTION

The real estate market in New York County is dynamic and influenced by various economic, social, and infrastructural factors.

The objective of this study is to analyze housing trends, price distributions, and property availability.

To achieve this goal, the following data-driven techniques were leveraged:

- Data Acquisition
- Optimization
- Storage
- Analysis

# DATA ACQISITION



Two of the most commonly used techniques were employed for the data acquisition phase.

## ❖ Kaggle API

The New York Housing Market dataset, available on the Kaggle platform, was obtained through Kaggle's API. This dataset includes details on 4,801 properties across 17 attributes, such as:

BROKERTITLE - Title of the broker

> TYPE - Type of the house

> PRICE - Price of the house

> ADDRESS - Complete address of the house

> STATE - State of the house.

## ❖ OpenStreetMap (OSM) API

It was used to acquire geographic coordinates (longitude, latitude) of the essential amenities:

> Schools

> Public stop positions (bus-stops, train-stops)

> Hospitals

## ❖ Web Scraping

It was used to get insights further about the neighborhood of the listed house locations. Specifically, the extracted data contains information about:

- > Walk Score – it measures the walkability of a specific location which evaluates how convenient to access everyday amenities.
- > Transit score – it evaluates the availability and usefulness of public transportation.
- > Bike score – it measures how bike-friendly the area is.

# OPTIMIZING DATA MANAGEMENT & DATA STORAGE

Effective data management is crucial for ensuring accuracy and reliability in analysis.

## - Data Quality & Profiling

To maintain data integrity, duplicate records were identified and removed, reducing the dataset from 4,801 to 1,231 entries. Geospatial coordinates were validated to ensure properties were correctly mapped within New York County. Key attributes, such as the number of bedrooms and bathrooms, were standardized to maintain uniformity. Additionally, missing or incomplete address details were reviewed to provide a more comprehensive dataset.

## - Data Integration & Enrichment

To enhance the dataset, information from multiple sources—including Kaggle, OpenStreetMap, and web scraping—was integrated. Neighborhood insights such as Walk Score, Transit Score, and Bike Score were added to evaluate accessibility. Proximity metrics were also calculated, determining the distance from each property to schools, hospitals, and public transportation stops. These enhancements provided a more detailed and insightful foundation for analyzing housing trends in New York County.

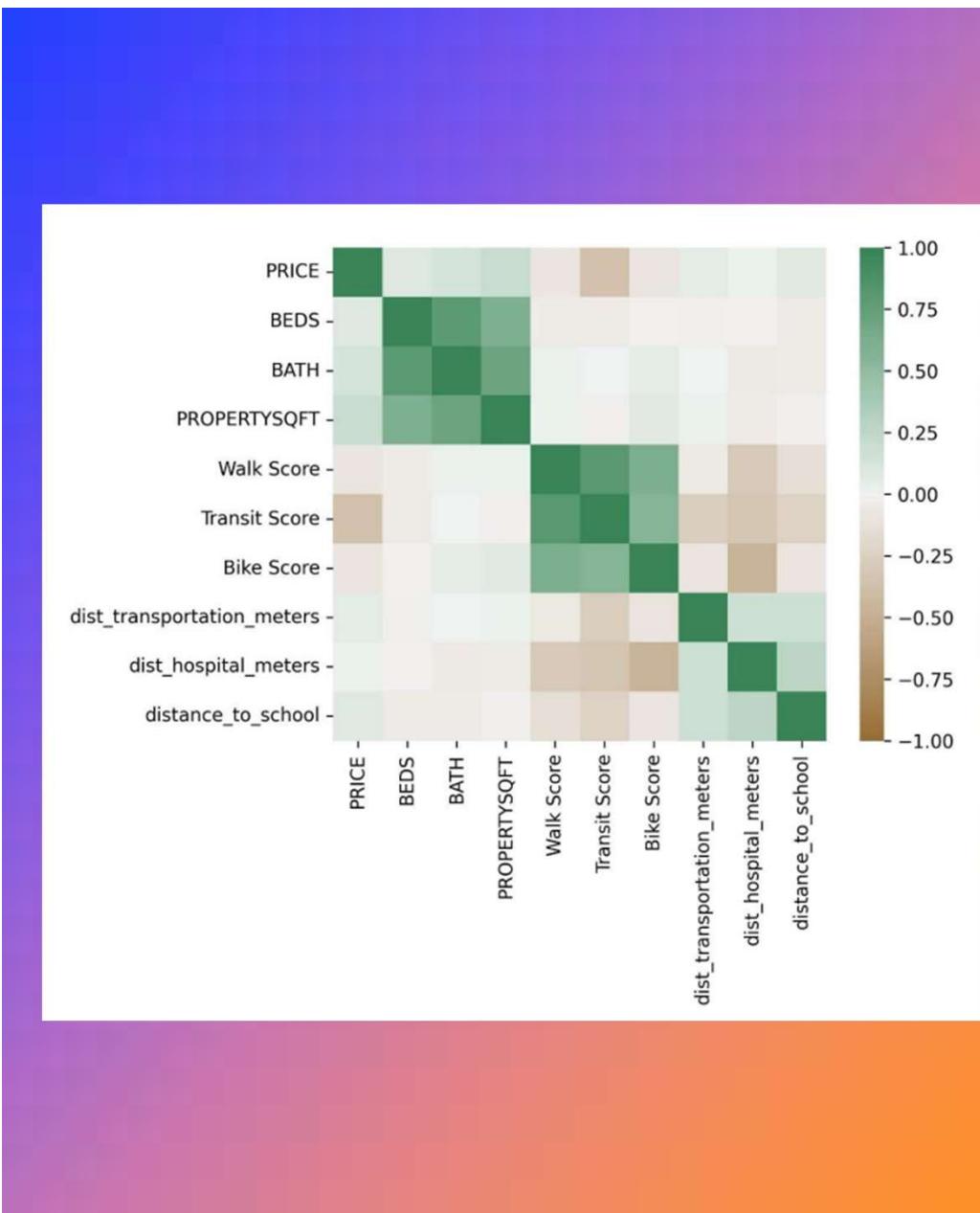
## - Data Storage

To efficiently handle the scale and diversity of the dataset, MongoDB, a NoSQL database, was chosen for storage. Its flexibility in managing both unstructured and semi-structured data allows for seamless integration and efficient querying.



# DATA STORAGE & DATA ANALYSIS

- ❖ In real estate, data analysis plays a crucial role in determining property valuation, identifying key pricing factors, and understanding market trends.
- ❖ The **descriptive statistics** reveal important insights:
  - Average property price: ~\$5.6 million
  - High standard deviation: Indicates extreme outliers, with a maximum price of \$2.1 billion
  - Typical property features:
    - Size: ~2,357 square feet
    - Bedrooms: 3
    - Bathrooms: 2
- ❖ Further **correlation analysis** highlights that:
  - Larger properties tend to have more bedrooms and bathrooms
  - Property size generally correlates with price
  - Wealthier neighborhoods prioritize private transport, resulting in fewer transit options.
  - High transit accessibility is common in high-density, lower-income areas.



# NEIGHBORHOOD ANALYSIS

To gain deeper insights into the distribution of property prices and density across New York County, an analysis was conducted using MongoDB queries. This involved calculating the number of properties, average price, and average property size for each neighborhood.

## ❖ Highest-Priced Areas:

- > SoHo and Tribeca have the most expensive properties, with average prices of \$8.09M and \$8.07M, respectively.
- > These neighborhoods also offer some of the largest properties, averaging over 3,100 sqft.

## ❖ Mid-Range Neighborhoods:

- > Hell's Kitchen, Upper East Side, and Chelsea have property prices ranging from \$5M to \$6.67M.
- > Property sizes in these areas vary between 2,300 and 2,663 sqft. East Village has the largest average property size (4,393 sqft), despite a moderate price of \$5.26M.

## ❖ More Affordable Neighborhoods:

- > Central Harlem, Roosevelt Island, Washington Heights, and Inwood offer properties at lower price points (\$1.05M - \$2.1M), with sizes ranging from 1,800 to 2,130 sqft.
- > Kingsbridge and Gravesend have the lowest prices, at \$0.46M and \$0.19M, respectively, with smaller property sizes and limited availability.

# NEIGHBORHOOD ANALYSIS

Manhattan, with its high concentration of properties, presents a wide range of prices and sizes.

The **average property price is \$3.25M**, with an **average size of 2,192 sqft**, reflecting the diverse nature of the city's real estate market.

This variation highlights **the contrast** between high-end luxury areas and more affordable housing options within the borough.



Figure 1.2: Average Neighbourhood locations with average price per square feet.

# DATA ANALYSIS – PROXIMITY TO KEY AMENITIES

Proximity to transportation, schools, and healthcare is crucial for buyers and investors.

## ❖ Public Transportation:

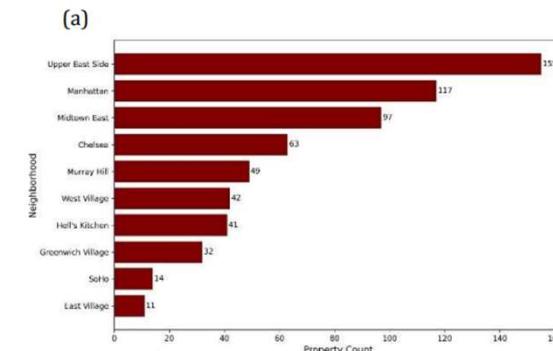
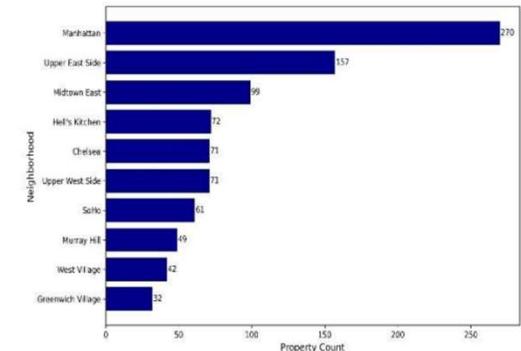
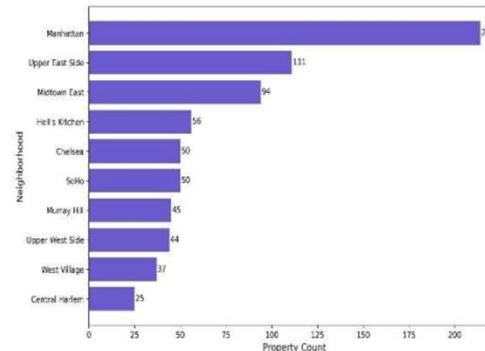
- 866 properties are within 200 meters of transit stops, with walk and transit scores above 70.
- Manhattan leads with 214 such properties, reflecting its dense connectivity.

## ❖ Schools:

- 1,071 properties are within 200 meters of a school, maintaining high accessibility scores.
- Manhattan (270 properties), Upper East Side (157), and Midtown East (99) are the top areas for school proximity.

## ❖ Hospitals:

- 638 properties are within 1,000 meters of a hospital, ensuring easy medical access.
- Upper East Side (155), Manhattan (117), and Midtown East (97) have the highest concentration.



(c)

# LUXURY REAL ESTATE ANALYSIS

New York City's luxury real estate market features some of the most prestigious and expensive properties. The dataset highlights the highest-priced listings:

## ❖ Top Listings:

- Hell's Kitchen – Most expensive listing at \$195M (Brokered by Serhant).
- Upper East Side – \$60M listing (Brokered by Compass).
- Other notable properties:
  - Hell's Kitchen (\$56M) & Manhattan (\$55M) – Both brokered by Douglas Elliman.
  - SoHo (\$50M) – Brokered by Nest Seekers International.

## ❖ Neighborhood Trends:

- Upper East Side leads with 7 luxury listings, followed by Chelsea and SoHo (4 each).
- Many high-end properties are located along Central Park, offering premium skyline and park views.

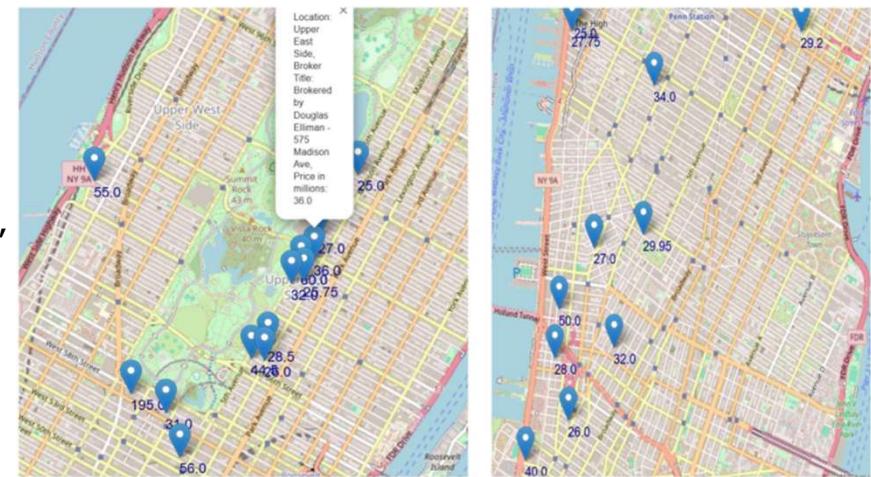


Figure 1.6: Highest- priced property locations.

# CONCLUSIONS

- ❖ This study aims to demonstrate the comprehensive process of acquiring, merging, and analyzing real estate data. Through the use of APIs and web scraping, extensive data was gathered, and effective data management allowed for meaningful analysis.
- ❖ The findings highlight key variations in pricing, property sizes, and availability across New York City's neighborhoods.
- ❖ SoHo and Tribeca emerge as the priciest areas with luxury properties, while neighborhoods like Central Harlem, Washington Heights, and Inwood offer more affordable options for buyers.
- ❖ Manhattan remains a central hub for diverse property prices and sizes. Beyond pricing, factors like proximity to public transportation, schools, and healthcare significantly influence property desirability. Manhattan leads in well-connected properties, making it ideal for those prioritizing convenience.
- ❖ This analysis also reveals a strong demand for luxury properties, particularly Central Park-facing apartments, with the Upper East Side standing out for its high-end listings.
- ❖ Ultimately, the study emphasizes the importance of location, connectivity, and amenities in shaping real estate trends and provides valuable insights for prospective homeowners and investors alike.

**THANK YOU FOR YOUR  
ATTENTION!**