WELOCME

DATA ANALYTICS SQL

PRESENTED BY SEHBA KHAN

TABLE OF CONTENT

INTRODUCTION

Objective: To provide insights into the New York real market through analysis of the dataset.

Significance: Understanding trends in the housing market can inform investment decisions, help buyers find suitable properties, and aid brokers in serving their clients effectively.

Overview of Presentation Structure: Broker Information PropertyType Analysis Pricing Analysis Bedroom and Bathroom Analysis Property Size Analysis Geographical Distribution Address Information Analysis Let's dive into the data to uncover valuable insights.

Key Columns:

BROKERTITLE:	Title of the broker
TYPE:	Type of the house
PRICE:	Price of the house
BEDS:	Number of bedrooms
BATH:	Number of bathrooms
PROPERTYSQFT:	Square footage of property
ADDRESS:	Full address of the house
STATE:	State of the house
MAIN_ADDRESS:	Main address information
ADMINISTRATIVE AREA LEVEL 2:	Administrative area level 2 information
LOCALITY:	Locality information
SUBLOCALITY:	Sublocality information
STREET_NAME:	Street name
LONG_NAME:	Long name
FORMATTED_ADDRESS:	Formatted address
LATITUDE:	Latitude coordinate of the house
LONGITUDE:	Longitude coordinate of the house

SQL QUERY

Utilized SQL queries to extract and manipulate data from the "NY_House" table. Queries included selection, filtering, aggregation to obtain relevant insights.

DATA CLEANING

Conducted data cleaning to handle missing values, outliers, and inconsistency. Ensured data integrity and reliability for analysis.

STATISTICAL ANALYSIS

Applied statistical techniques to analyze numerical data (e.g., prices, square footage) and category data (e.g., property types, states).

GEOSPATIAL ANALYSIS

Latitude and longitude coordinates for geospatial analysis, including mapping property location.

METHODOLOGY

INSIGHT – 1

Property Type Distribution

SELECT type , count(*) AS Property_Count
FROM NY_House
GROUP BY type
ORDER BY Property_Count DESC;

Explaination::
This query calculates the average price of properties grouped by their types, allowing stakeholders to understand the average price range for each property type and identify potential areas of investment or market trends.
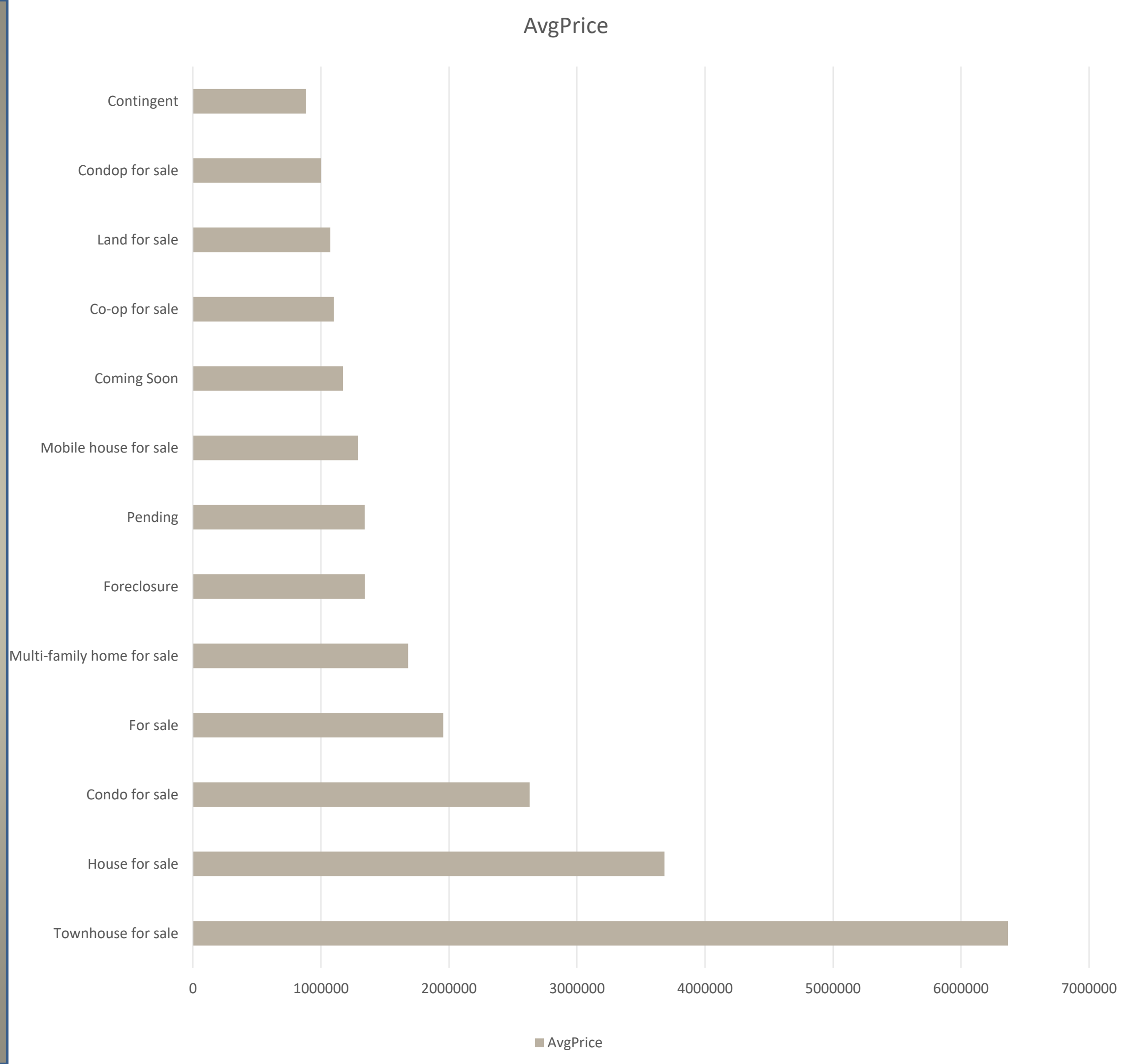

Property_Count
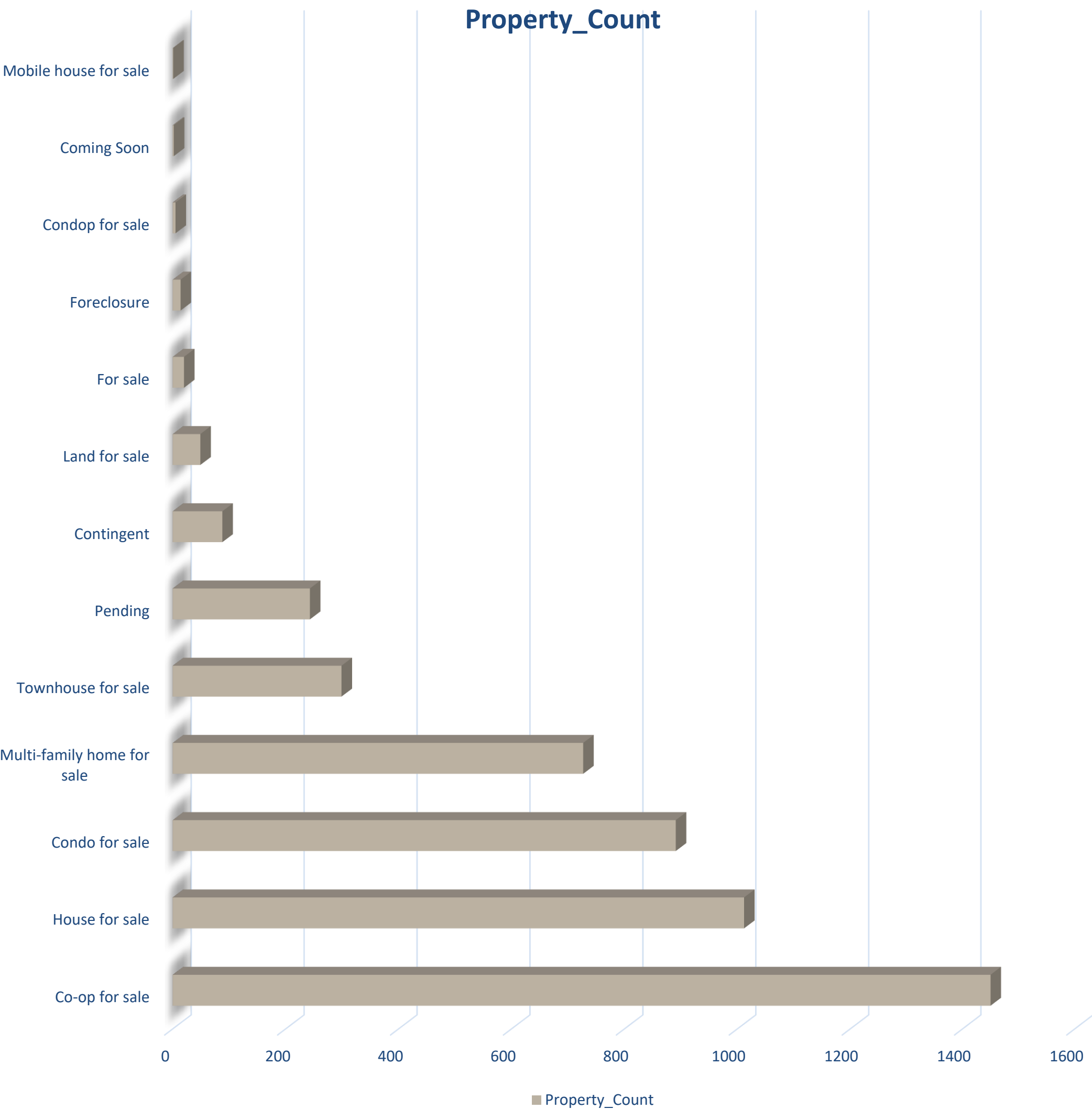
INSIGHT – 2
Average Price by Property

SELECT type , AVG(PRICE) AS Avg_Price
FROM NY_House
GROUP BY type
ORDER BY Avg_Price DESC;

Explaination:
This query helps analyze the distribution of property types and identifies the most common types of properties listed in the dataset, providing insights into market preferences.
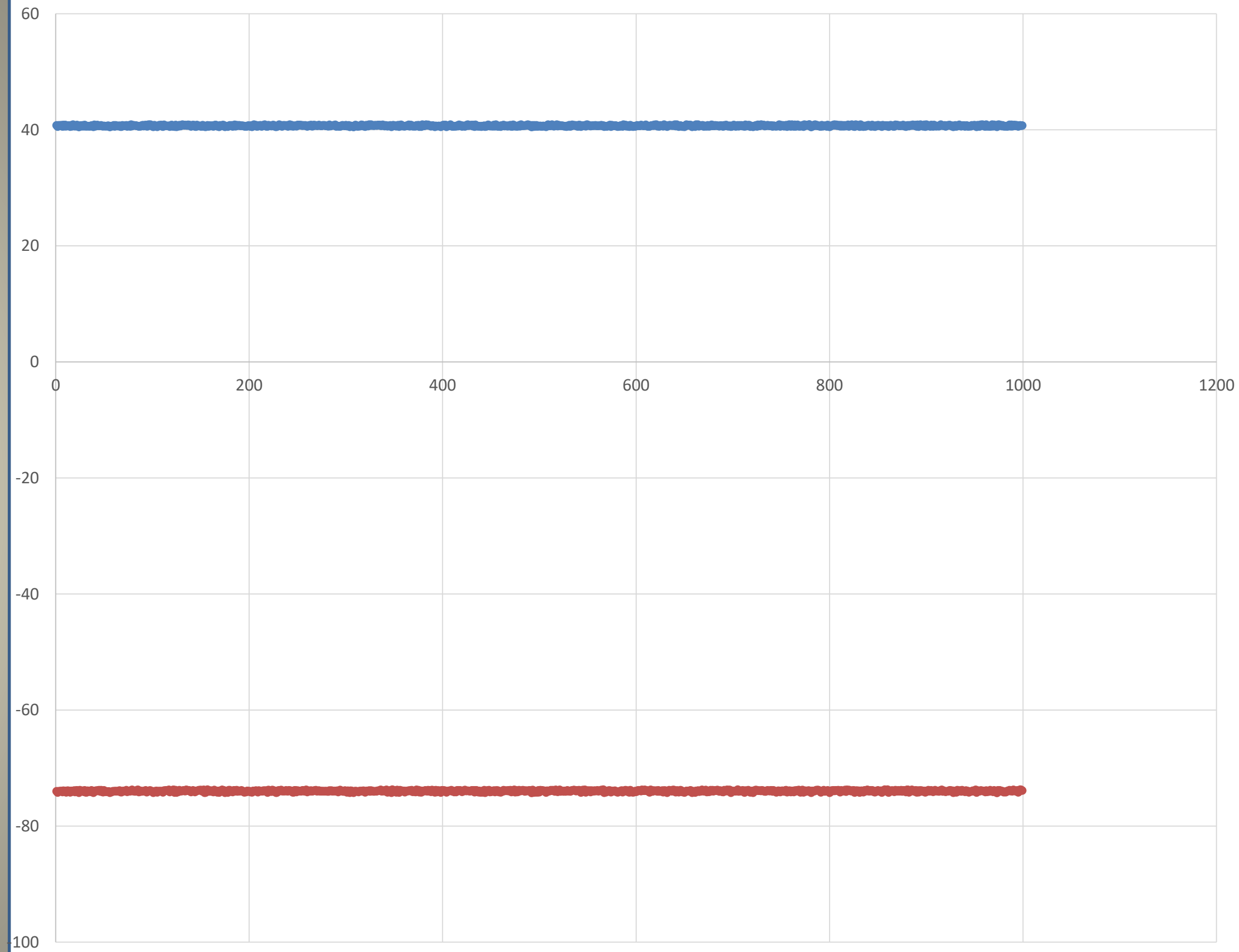

AvgPrice

INSIGHT – 3

Geospatial Analysis

SELECT address , latitude , longitude
FROM NY_House;

Explaination:
This query retrieves the address along with latitude and longitude coordinates of properties, enabling geospatial analysis to identify hotspots, clusters, or patterns in property distribution across New York City.
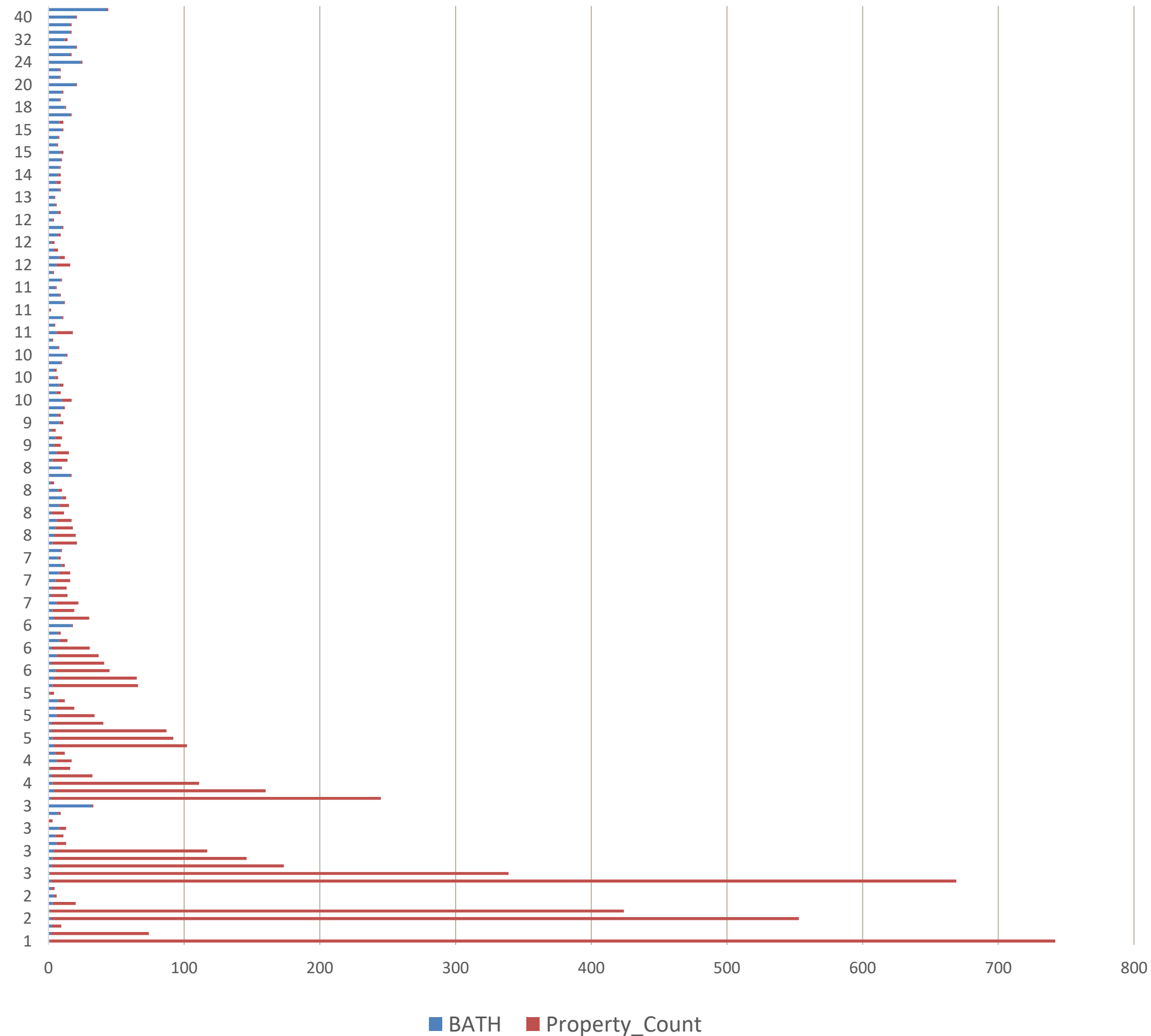
INSIGHT – 4

Bedroom-Bathroom Ratio

SELECT beds, bath , count(*) AS Property_Count
FROM NY_House
GROUP BY beds , bath
ORDER BY Property_Count DESC;

Explaination:
This query provides insights into the distribution of bedroom and bathroom combinations in listed properties, helping stakeholders understand housing preferences and market demand for different configurations.
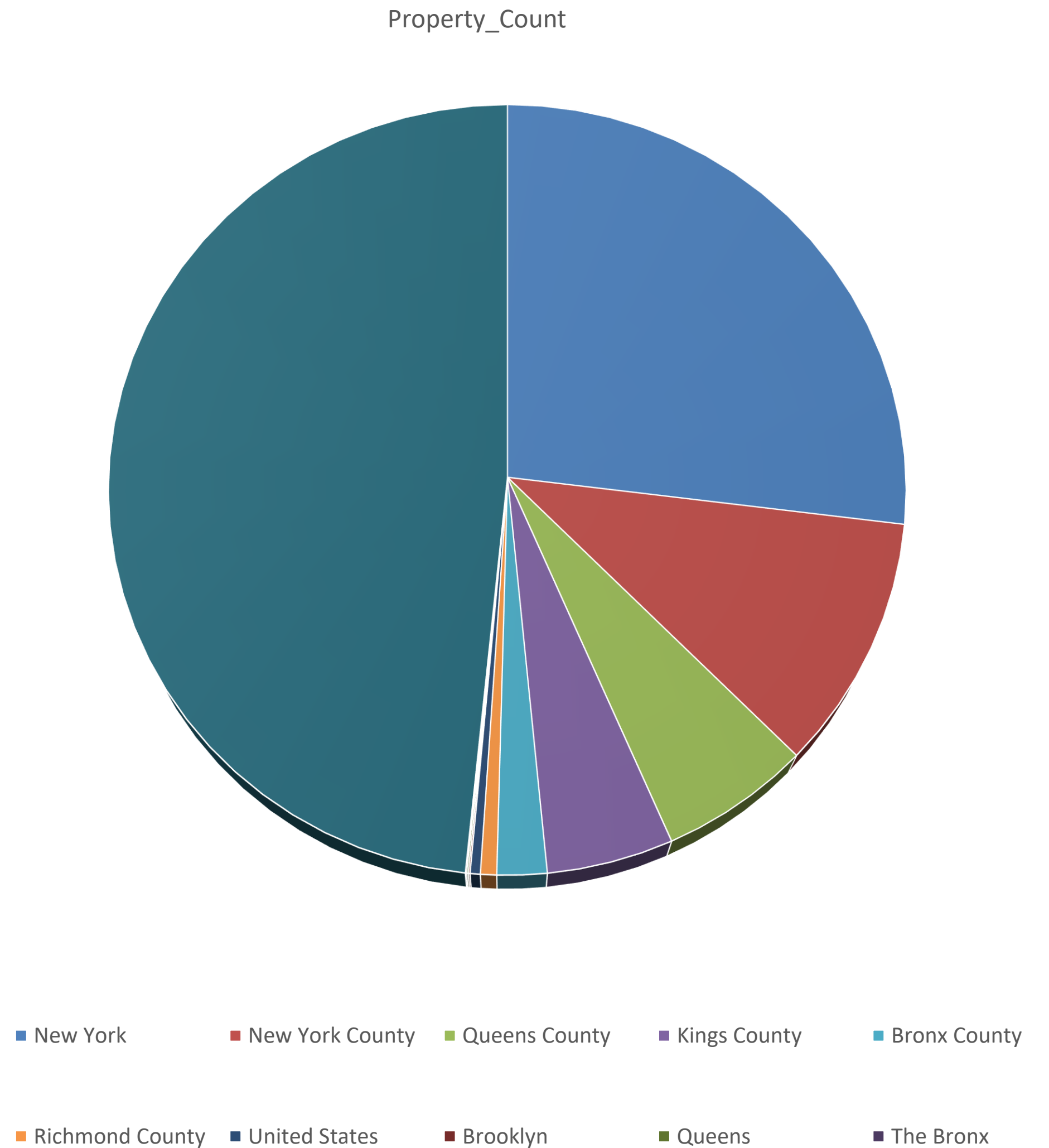
INSIGHT – 5
Price Distribution

SELECT locality, count(*) AS Property_Count
FROM NY_House
GROUP BY LOCALITY
ORDER BY Property_Count DESC
LIMIT 10;

Explaination:
This query retrieves the top 10 localities with the highest number of properties listed in the dataset. It provides insights into the distribution of properties across different neighborhoods or localities within New York City.

Property_Count

- New York
- New York County
- Queens County
- Kings County
- Bronx County
- Richmond County
- United States
- Brooklyn
- Queens
- The Bronx

Main Findings:

o Overview of key insights gained from the analysis, such as: Distribution of broker titles and property types.
o Trends in pricing, bedroom and bathroom configurations, and property sizes.
o Geographical distribution of properties across New York.

Challenges and Lessons Learned:

o Discussion of challenges encountered during the analysis process, such as: Data cleaning complexities.
o Interpretation of geospatial data.
o Lessons learned and potential improvements for future analysis.

Future Directions:

- Further exploration of the dataset to uncover additional insights.
- Integration of external data sources for enhanced analysis(e.g., demographic data, market trends).
- Implementation of predictive modeling to forecast housing market trends or property prices.
- Incorporation of advanced visualization techniques for more interactive and dynamic presentations.

THANKS

Sehba

9214981586

sehbakhan43@gmail.com