INSIGHTS FROM THE NEW YORK

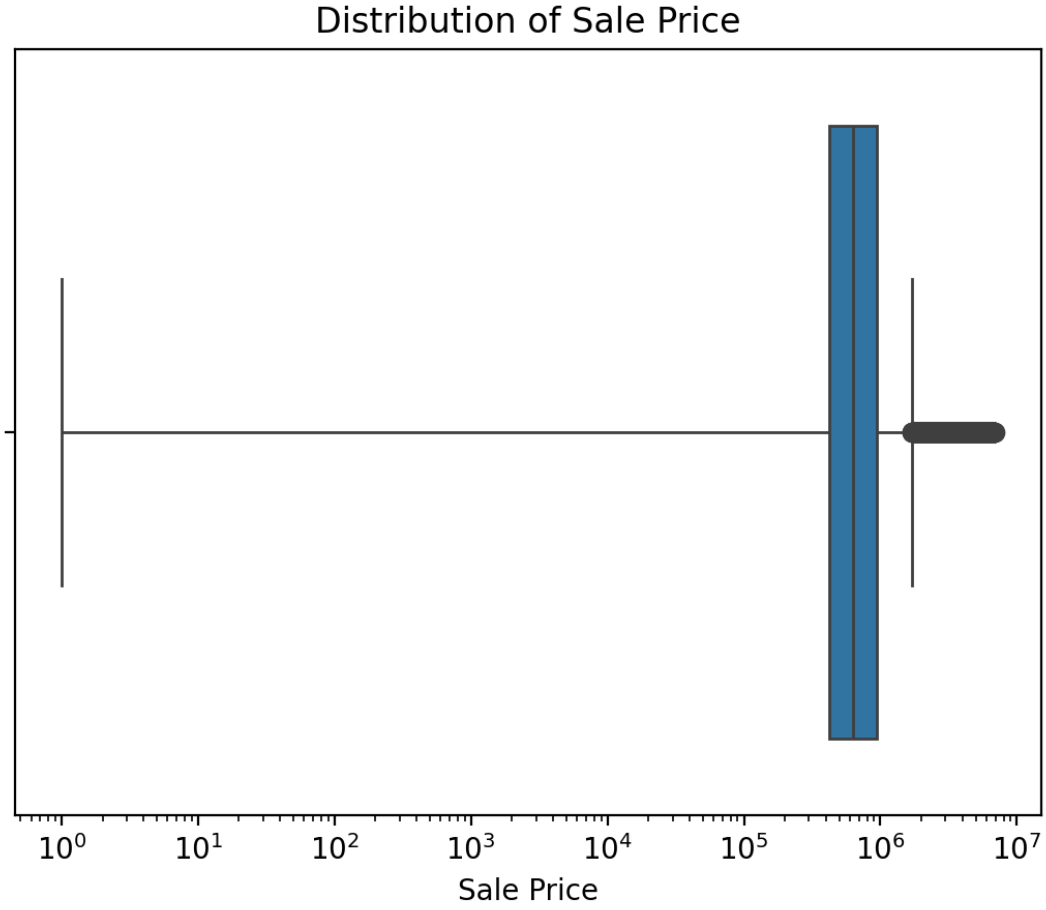
PROPERTY SALES DATASET

This dataset contains details of property sales in New York over a year. The columns in the dataset are:

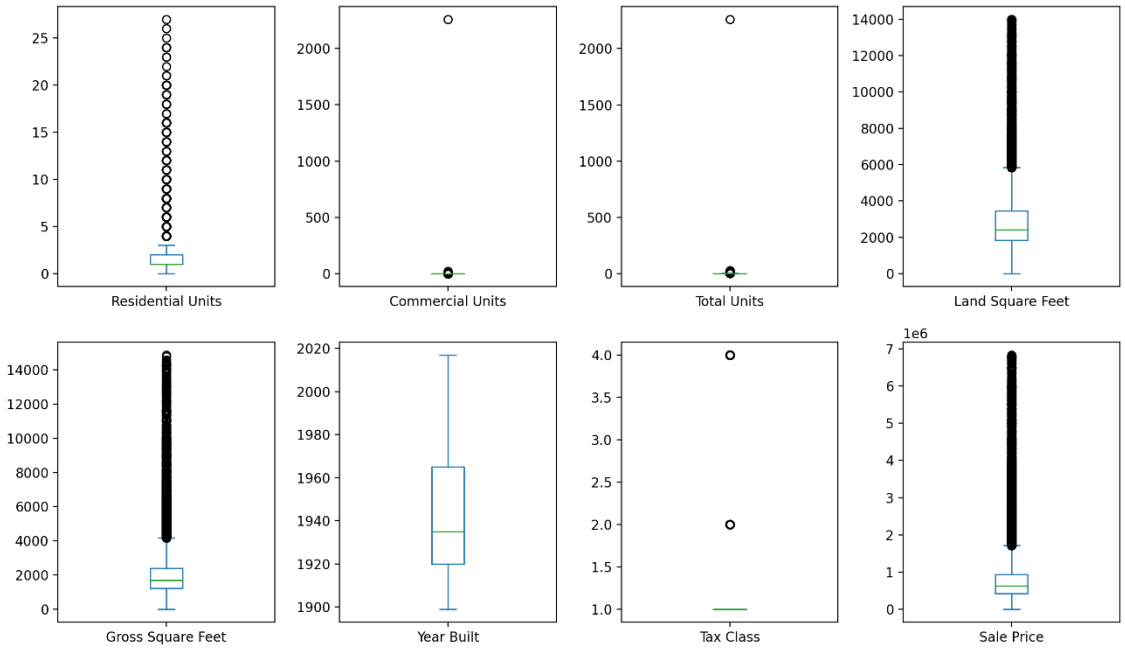
'Borough', 'Building Class Category', 'Residential Units', 'Commercial Units', 'Total Units', 'Land Square Feet', 'Gross Square Feet', 'Year Built', 'Tax Class', 'Sale Price', 'Sale Date', 'Sale Month', 'Sale Season'.

AIM: To generate insights that can assist Real Estate Companies in making informed decisions.

DISTRIBUTION ANALYSIS



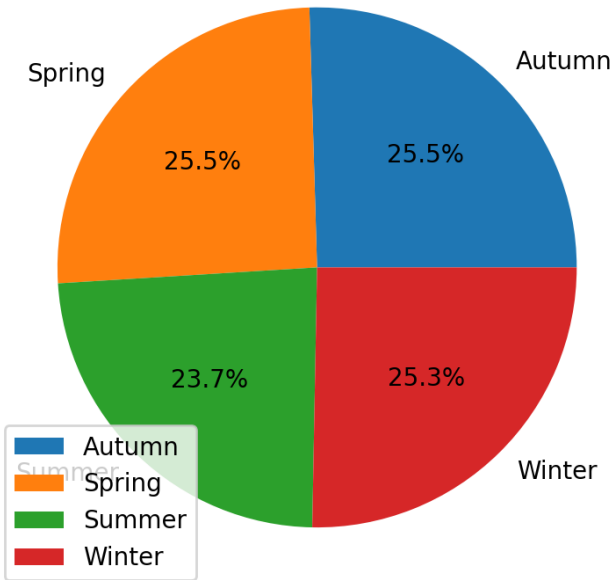
Most of the data points of *Sale Price* Fall between $600,000 and $1,000,000. The outliers are indications of expensive buildings/properties such as Office Buildings.



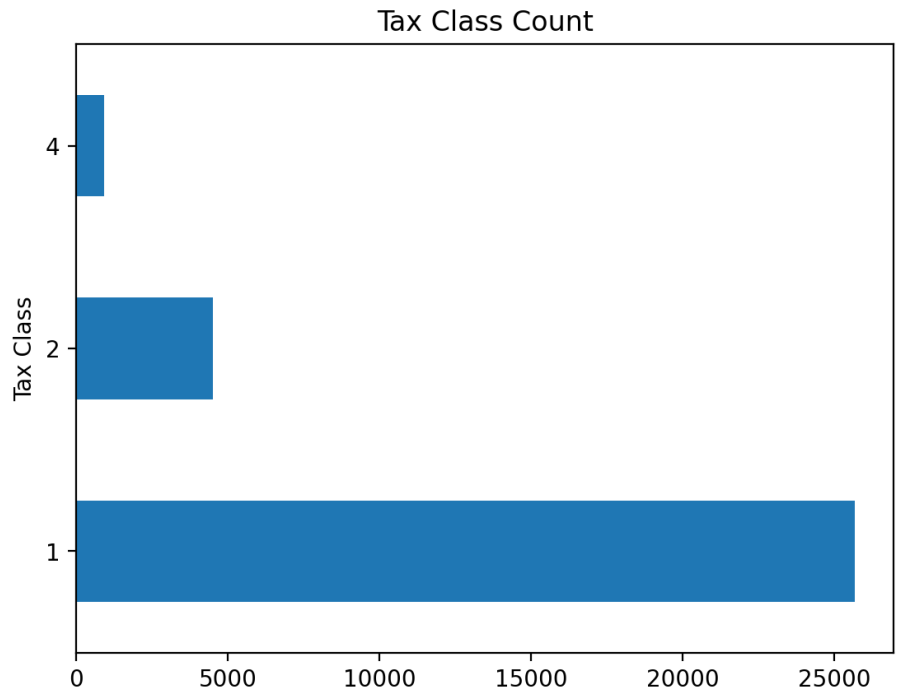
Most data points in *Land Square Feet fall between* **1800 sq ft – 3000 sq ft.** Outliers here could possibly indicate large land mass areas such as a Warehouse.

Most data points in *Gross Square Feet* fall between **1800 sq ft – 2100 sq ft.** Outliers here are an indication of elevator buildings and office buildings, probably.

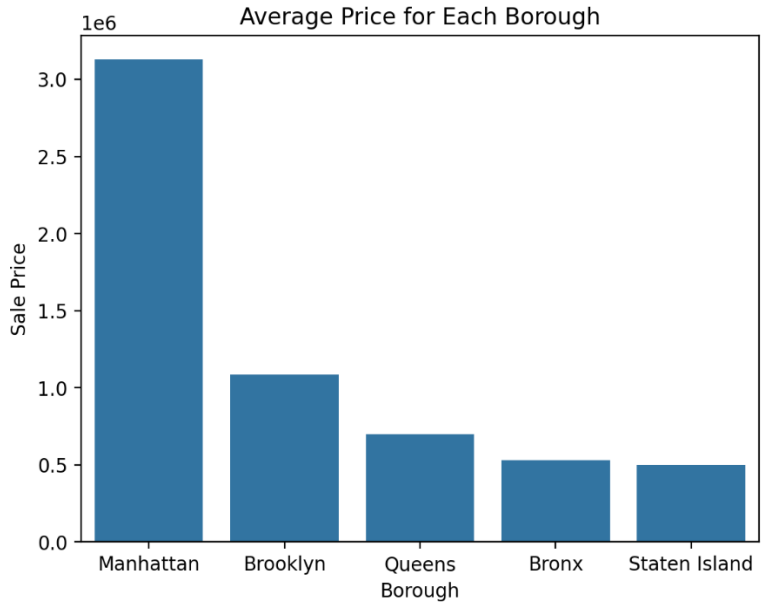
Outliers in *Commercial Units* are indications of tall buildings, like Office Buildings.



The dataset contains a roughly even proportion of Spring, Autumn, Summer and Winter *Sale Season.*



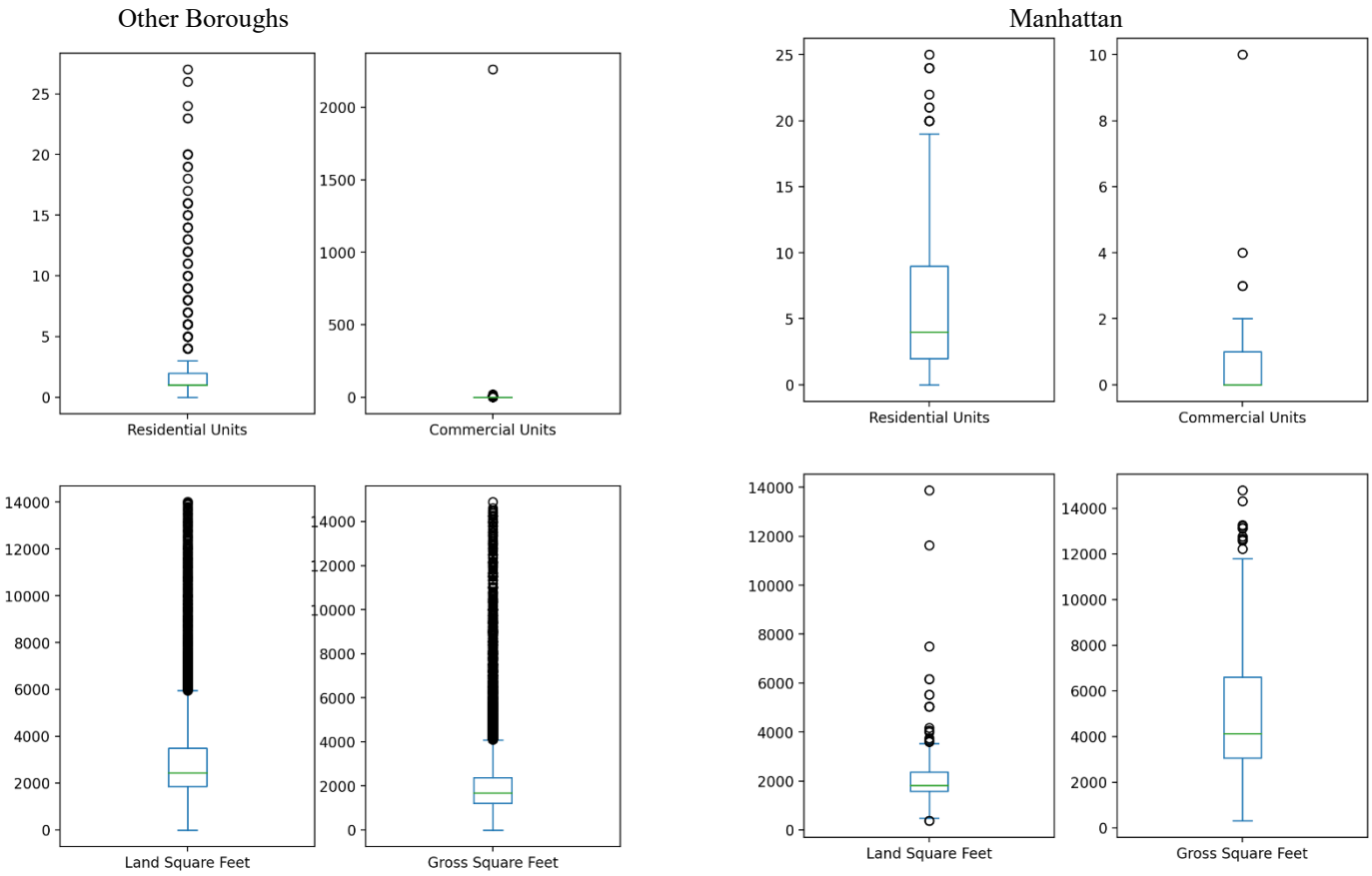
EXPLORATORY DATA ANALYSIS



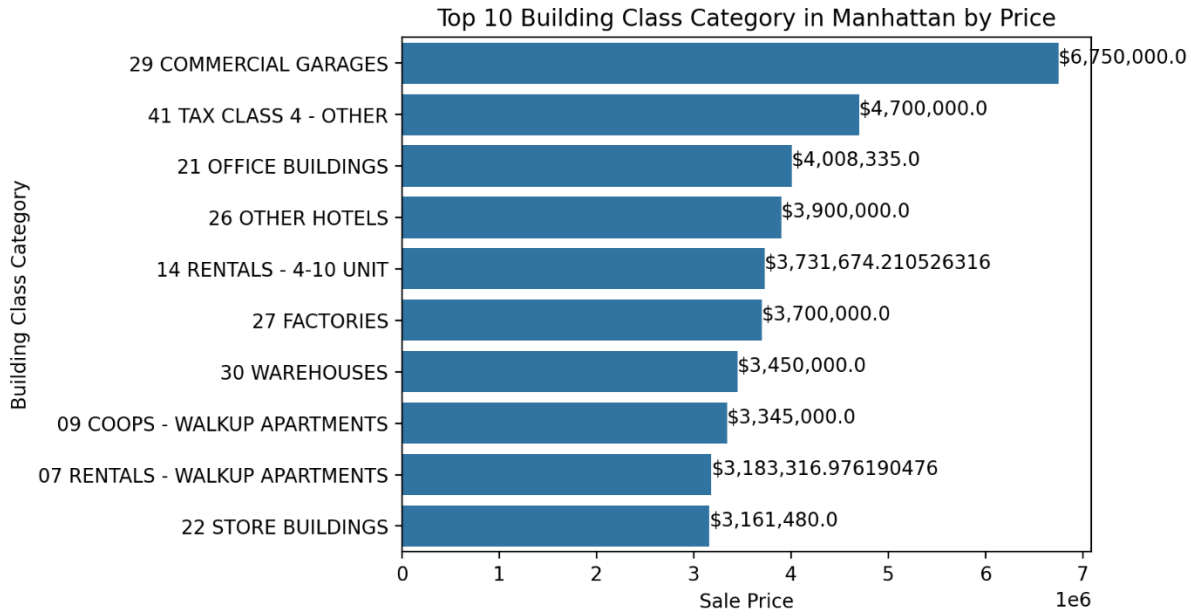
The chart above shows that the average price of property sold in Manhattan is significantly greater than in other Boroughs with an average price of roughly **$3,200,000.**

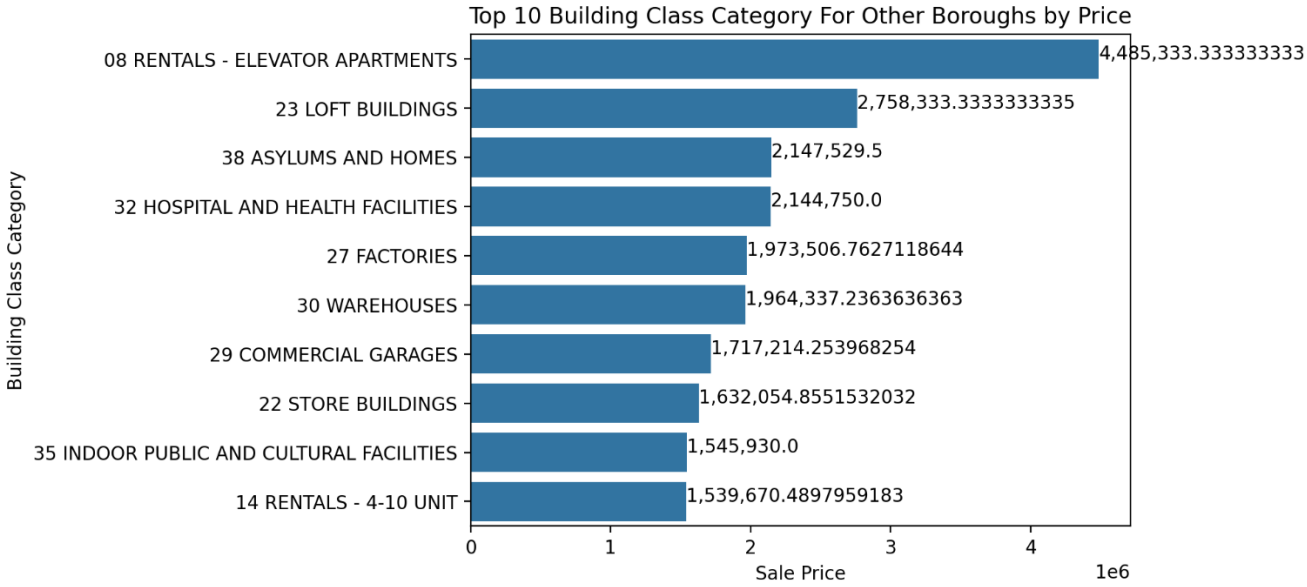
DOES MANHATTAN HAVE MORE EXPENSIVE PROPERTIES THAN OTHER BOROUGHS?

Drilling down further into other features and how they interact with the price of the properties, we observe that there are more properties with Commercial Units in Manhattan compared to other boroughs.

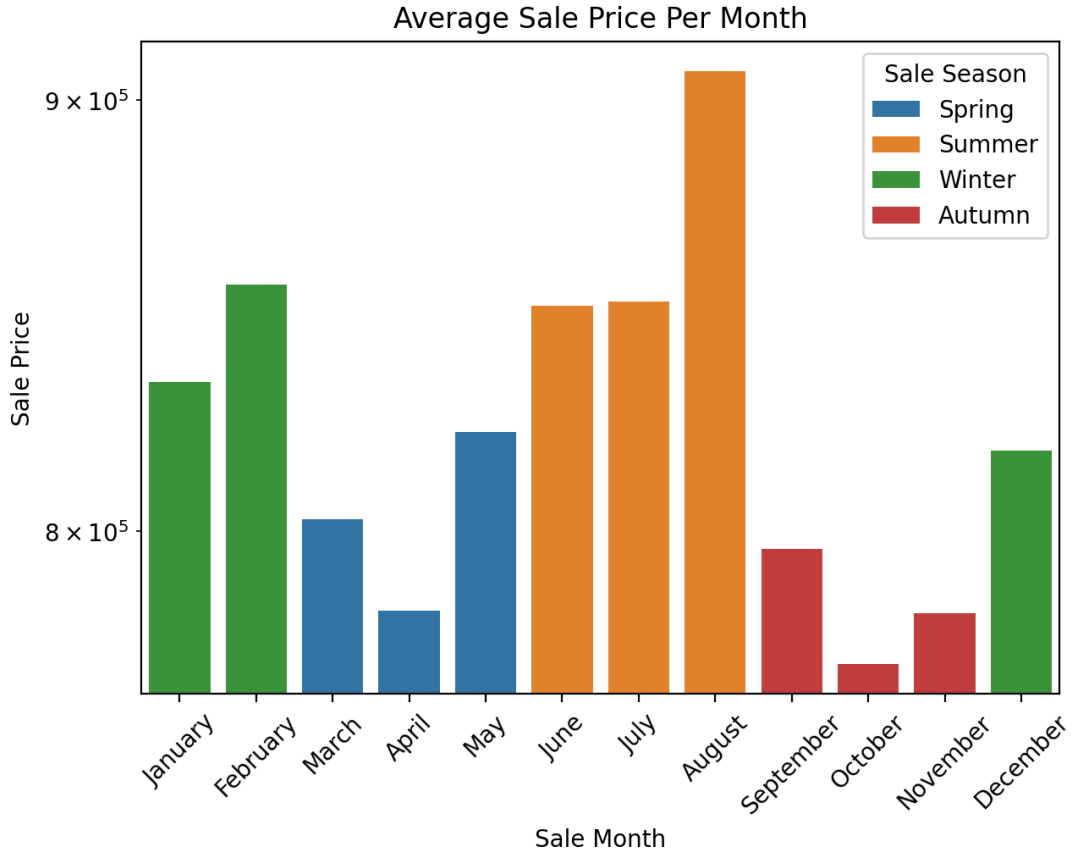


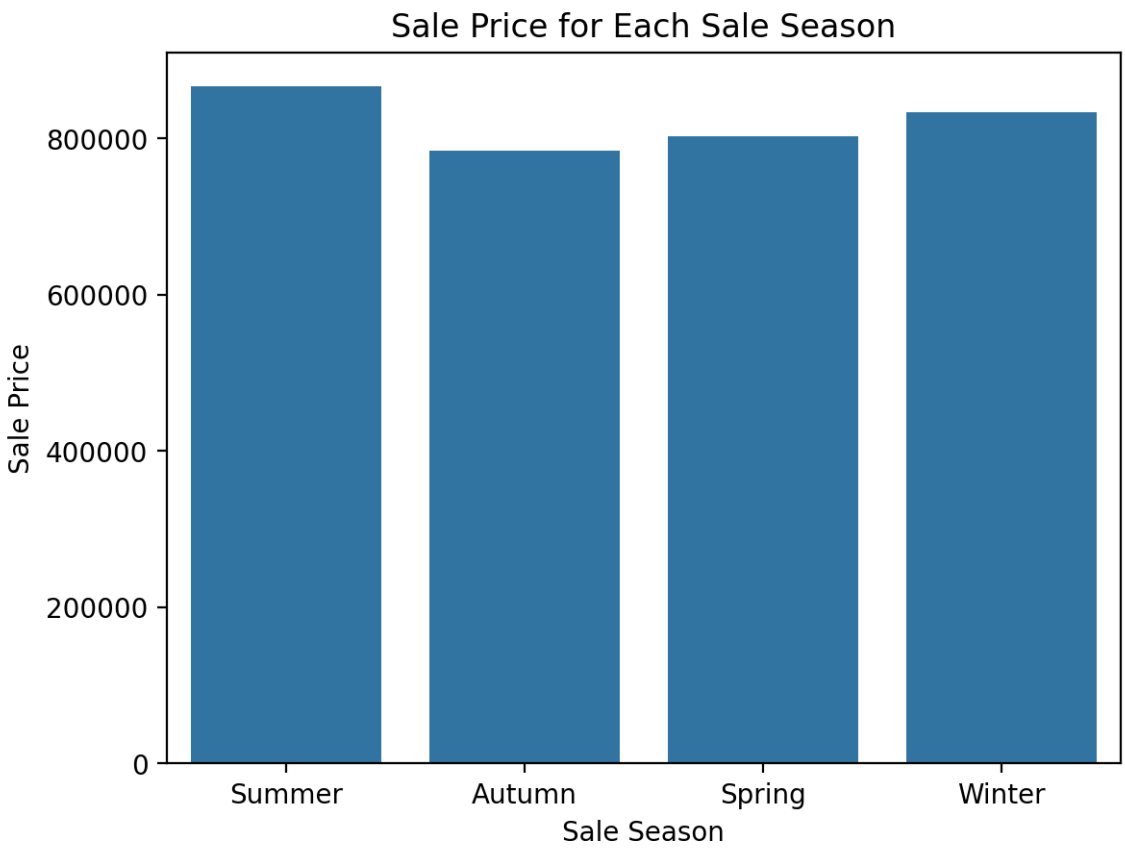
From the charts above we can also confirm that the average *Gross Square Feet* in Manhattan is greater than other Boroughs. **In conclusion,** Manhattan has more expensive properties than other boroughs due to the presence of more commercial properties (such as Hotels, Warehouses etc.) than residential properties and we can see this in the charts below.

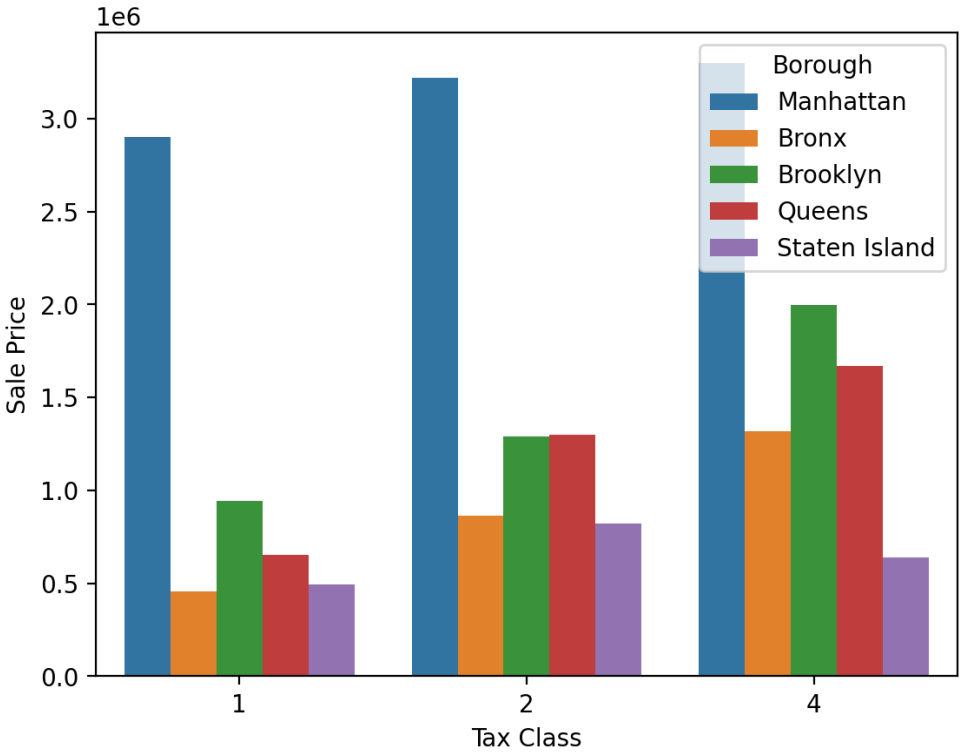


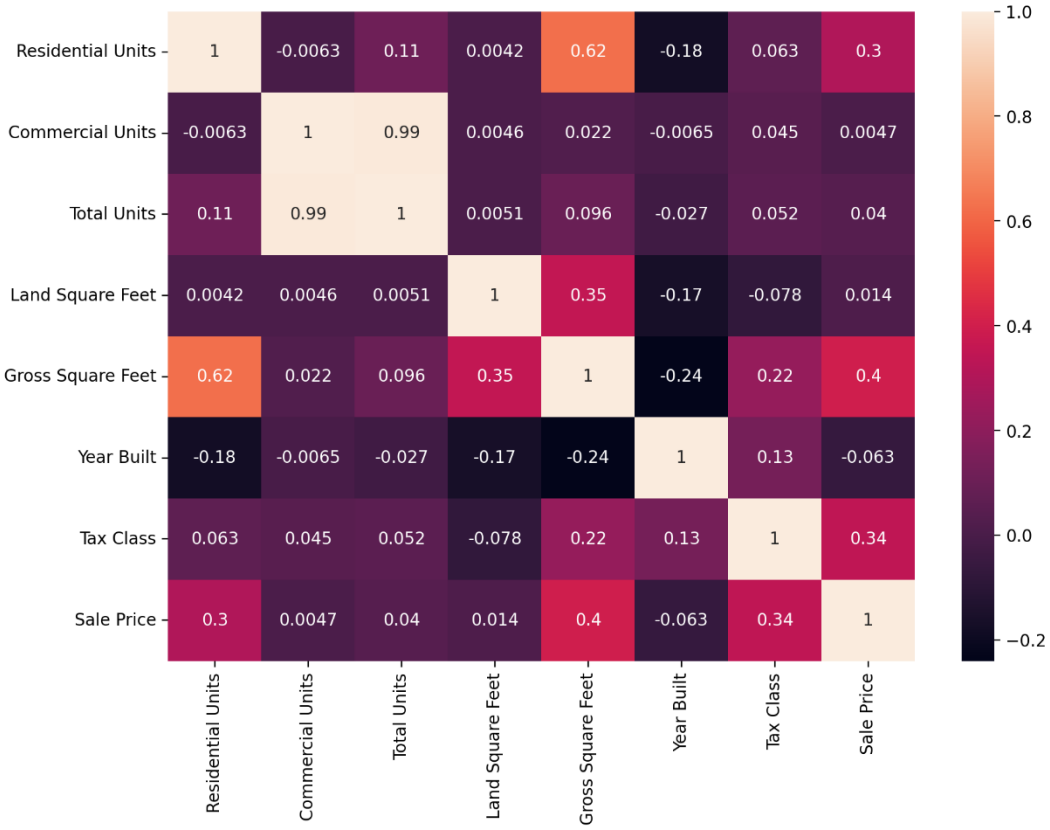


MORE INSIGHTS



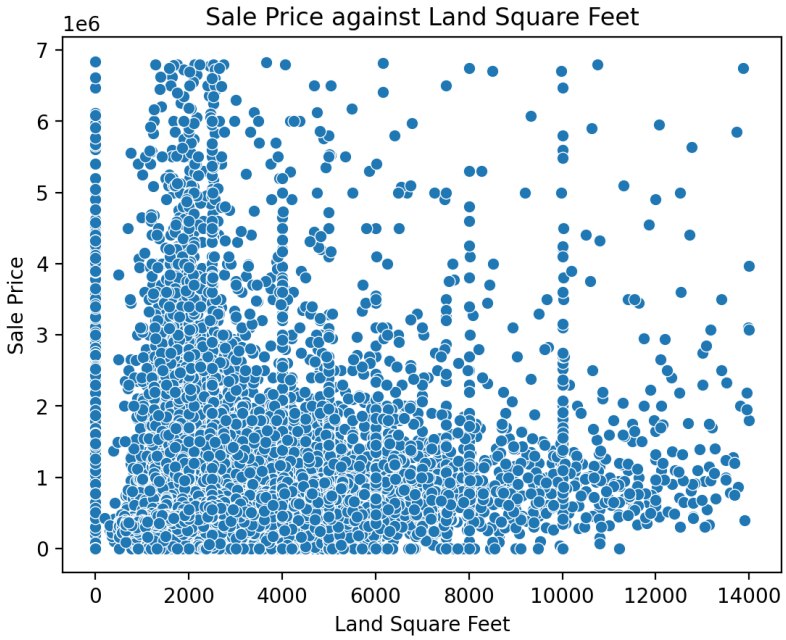

Sale Price for Each Sale Season
800000
600000
400000
200000
0
Sale Price
Summer
Autumn
Spring
Winter
Sale Season


1e6
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Sale Price
Borough
Manhattan
Bronx
Brooklyn
Queens
Staten Island
1
2
4
Tax Class

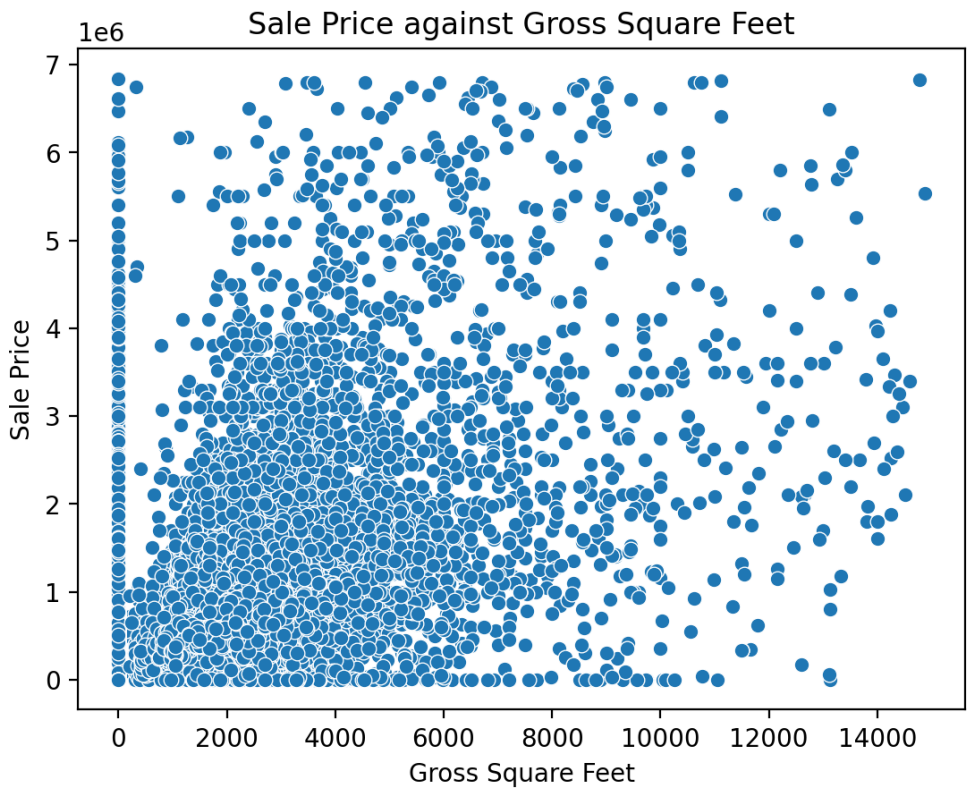


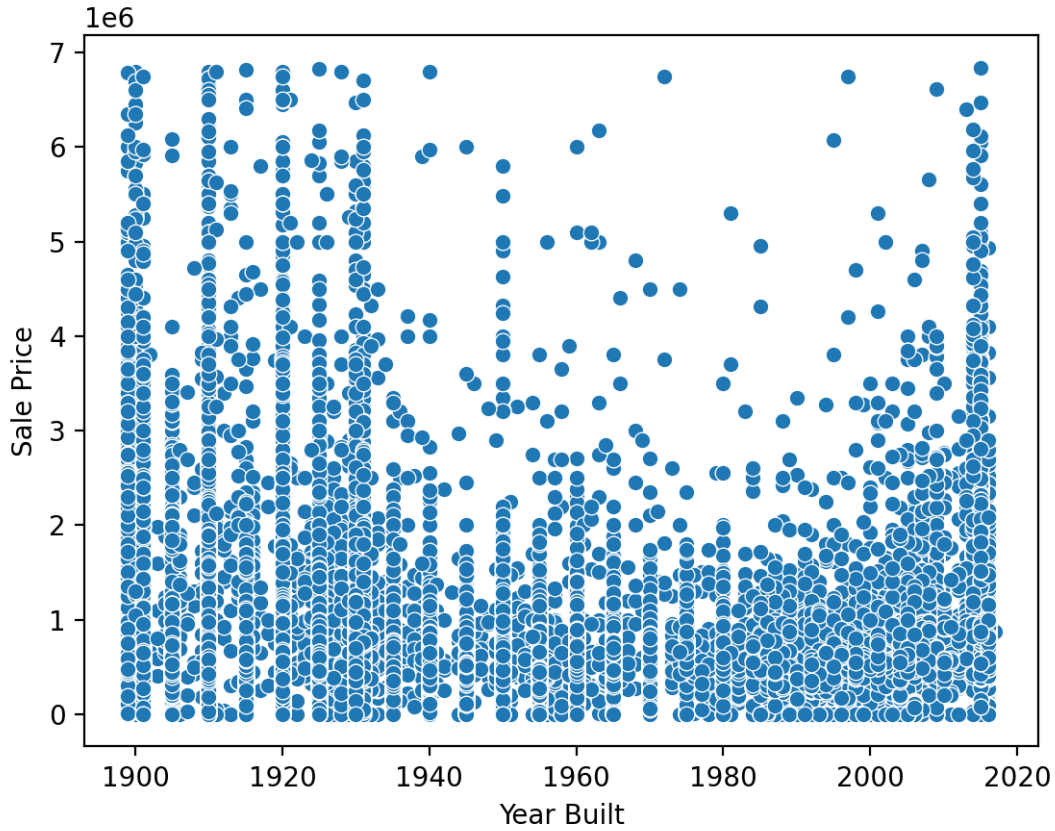
Analysing the correlation matrix above, it is observed that there is a low correlation between most features with a few exceptions. *Gross Square Feet* has a moderate correlation with *Sale Price*, this can be used as a predictive feature for machine learning models to predict Sale Price of property. *Residential Units* have a correlation of 0.3 which is moderate but still has a bit of an effect on the *Sale Price*

VISUALISING SOME THESE RELATIONSHIPS WITH A SCATTERPLOT



We observe very low correlation between *Sale Price* and *Land Square Feet*


Sale Price against Gross Square Feet
1e6
7
6
5
4
3
2
1
0
Sale Price
0
2000
4000
6000
8000
10000
12000
14000
Gross Square Feet


1e6
7
6
5
4
3
2
1
0
Sale Price
1900
1920
1940
1960
1980
2000
2020
Year Built

IN CONCLUSION

- The price of property in Manhattan is more expensive compared to other Boroughs.
- The average Sale Price for each Sale Season (Winter, Spring, Summer, Autumn) are pretty much the same
- The year the property was built has little to no effect on Sale Price
- The greater the Gross Square Feet of a property they greater the Sale Price
- Gross Square Feet is the best predictor for Sale Price

LIMITATIONS TO ANALYSIS

The quality of the dataset is not the best, a lot of missing values where present in the dataset and thus we cannot fully rely on the results of the analysis as there will be a great amount of bias due to the nature of the dataset.