

Project Report / Camilo Toro Gonzalez

Data Analytics for Finance Course UCD

GitHub URL

https://github.com/kamtoro/UCDPA_Camilo/

Abstract

This final project investigates the trends and insights in the Greater London property market using data publicly available from the HM Land Registry, which covers all property sold in Greater London from January 2018 to February 2023.

The primary objective is to identify investment opportunities and analyse the market's evolution during significant events, such as the COVID-19 pandemic and rising inflation. The dataset, comprising almost 500.000 sales during this period and has over 15 different types of meta data for each record of property sold within Greater London County.

The analysis reveals a general upward trend in the property market index over the last five years, with significant variations in year-over-year trends and property prices across different regions of Greater London. The most sold property types are flats/maisonettes, accounting for nearly half of all property sales. The results suggest that data analytics can provide valuable insights for potential investors and homebuyers looking to make informed decisions in an increasingly unpredictable global property market.

Introduction

The global property market is very unstable nowadays due to multiple external factors, as businesses and people increasingly rely on data to make informed decisions, using data analytics can be a crucial tool to decide where and when to buy your next property. This written assignment will explore some of the data provided by HM Land Registry institution and contains data for all property sold and registered since Jan 2018 up to Feb 2023.

Through the analysis of real-world property transactions data, we will examine the insights that can be gained from these data collected for the last 5 years and what might be the best areas of investment according with different type of budgets.

Dataset

So many people are looking to buy a house in the UK or even from abroad, and what I wanted to do is using my final assignment to assess how is the property marking evolving since important events such as covid, inflation and where and when would be great opportunities for investment. I am very interested on the Property Market trends for the past few years in the UK and wanted to use this as an opportunity to learn and gain real value out of this research on this complex type of market.

The dataset was extracted from (HM Land Registry, 2023) on an article named **“Guidance on How to access HM Land Registry Price Paid Data”**. The article pointed me out to another website portal from the same entity “(landregistry.data.gov.uk, n.d.)”, what I wanted

to do, was downloading data from all the UK counties to produce a large data set to be analysed, however, this site was having performance issues when tried to retrieve the full data from 2018 up to Feb 2023 for all sales in the UK. What I had to do was applying some filtering to reduce the data scope, by only focusing on “**Greater London**” county and extracting one data set per year (1st from Jan 2018 until 31st Dec 2022), then using pandas to read the CSV data stored and merge all of them in just one single **Pandas Dataframe**.

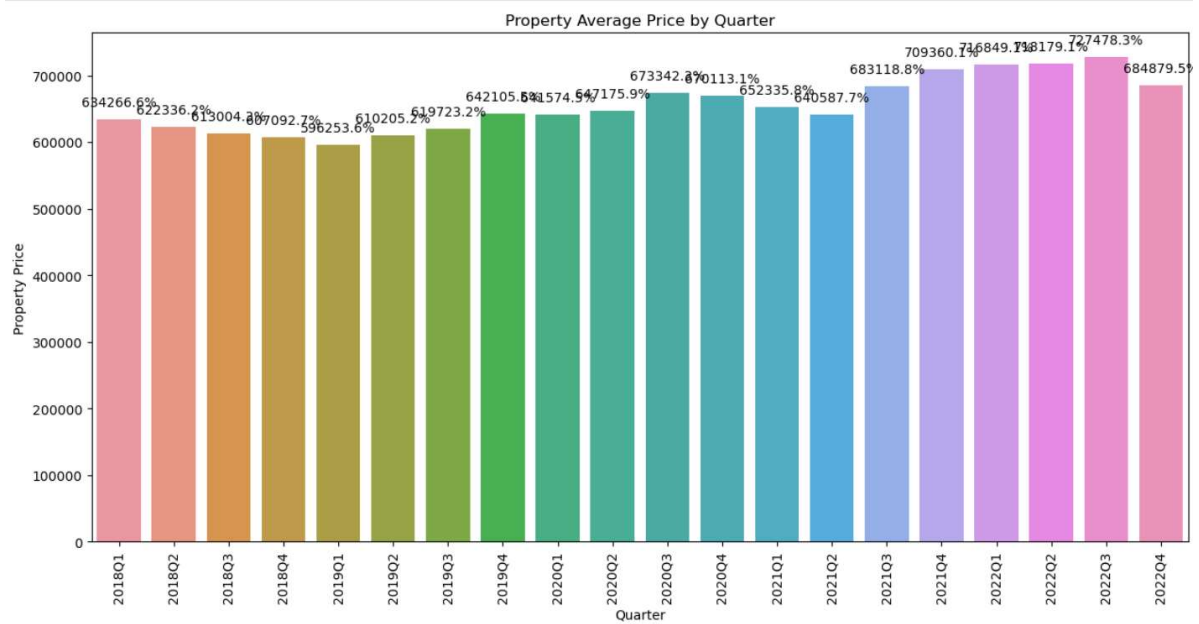
Based on (HM Land Registry, 2023), the time between a property is sold and registered can typically take between 2 weeks to two months, hence, we won't analyse data for Jan and Feb 2023 as it is incomplete.

Implementation Process

The implementation process for data analytics in finance involves a series of steps to ensure the data insights produces are as accurate as possible, so the analysis done are close enough to reality allowing the final user to take informed decisions in case they would like to invest on a property within the Greater London County area. The following are some of the key steps that had to be taken for an effective implementation:

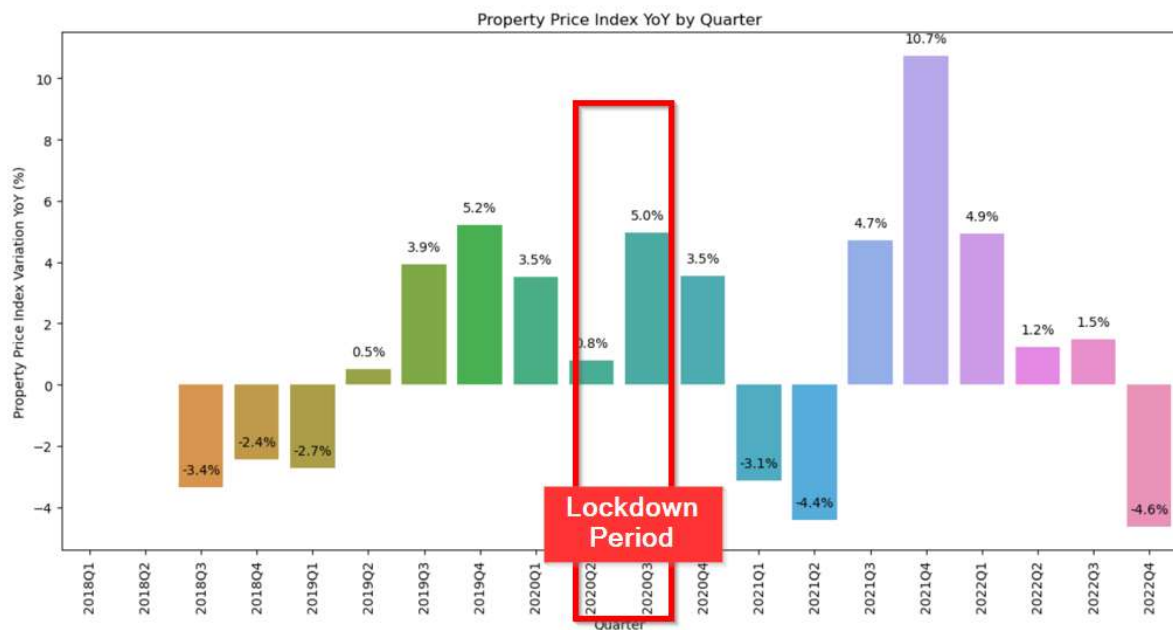
1. **Defining objectives:** The first step is to define the objectives of the data analytics project. This involves understanding the main goal and activities aligned to the goal.
2. **Understanding the Data needed/gathered:** I had to be able to understand the data collected, and what were other potential sources to complement.
3. **Data collection and preparation:** The next step was collecting and preparing the data required for the analysis. The best place found was the (HM Land Registry, 2023) website, they have all information available for all property transactions done in the UK. However, collecting the data was not an easy process as I could not query all the 5 years in the past, due to some technical restrictions on their site.
4. **After collecting the data, I had to wrangle and transforming to make it usable and preparing it for analysis.** The dataset, comprising 489,352 rows and 16 columns, was filtered and processed to focus on the Greater London area. I personally even had to manually identify and find the Locations of each borough from the data set (Inner London, South, North, West, East) I had to (Wikipedia-Contributors, 2023) and map the different boroughs (Districts on the data set) to be able to get more insights. I have tried Web Scrapping techniques to gather the data in python, but it was not possible due to some technical restrictions.
5. **Data analysis:** The data was analysed using statistical and machine learning techniques to identify patterns, trends, and insights.
6. **Visualization and reporting:** The insights obtained from the analysis are then visualized and reported in a way that is understandable and actionable for the potential audience/users of this information. For this, what I have used different python libraries available that were presented during the Data Analytics for Finance course.

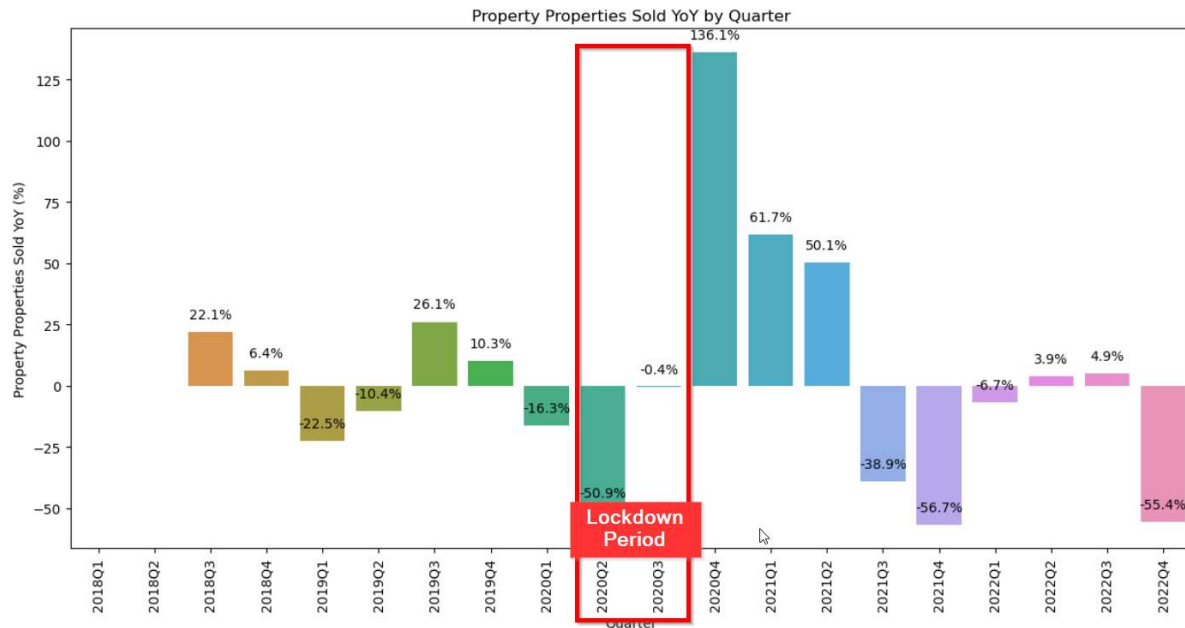
Results



Overall trend: The property market index has increased from 634,266 in 2018-Q1 to 684,879 in 2022-Q4, indicating a general upward trend during the last 5 years.

Year-over-year trends and what has happened during and after the Covid Lockdown?





2018: The market started with a decline from Q1 to Q4, with the most significant drop of -3.3% YoY in 2018-Q3.

2019: The market experienced a recovery with a YoY increase of 0.1% in 2019Q2, 3.9% in 2019-Q3, and a notable 5.3% in 2019-Q4.

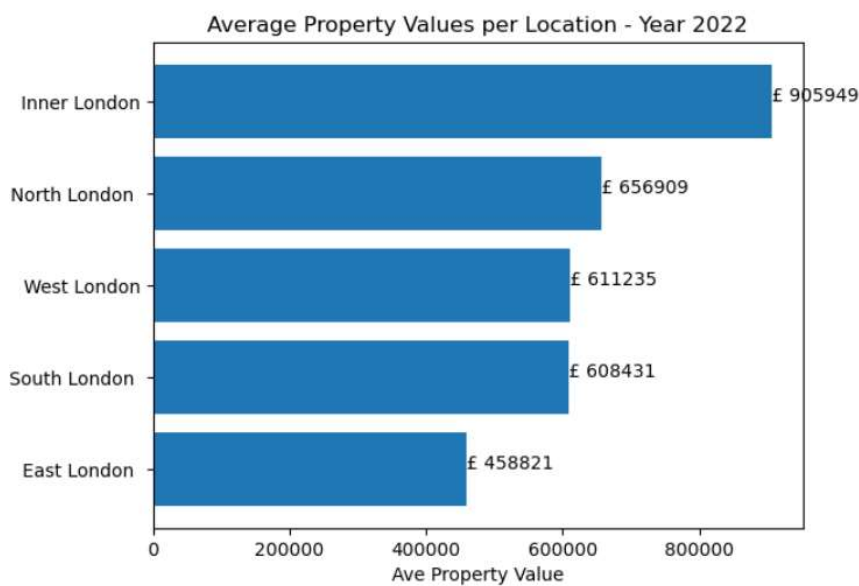
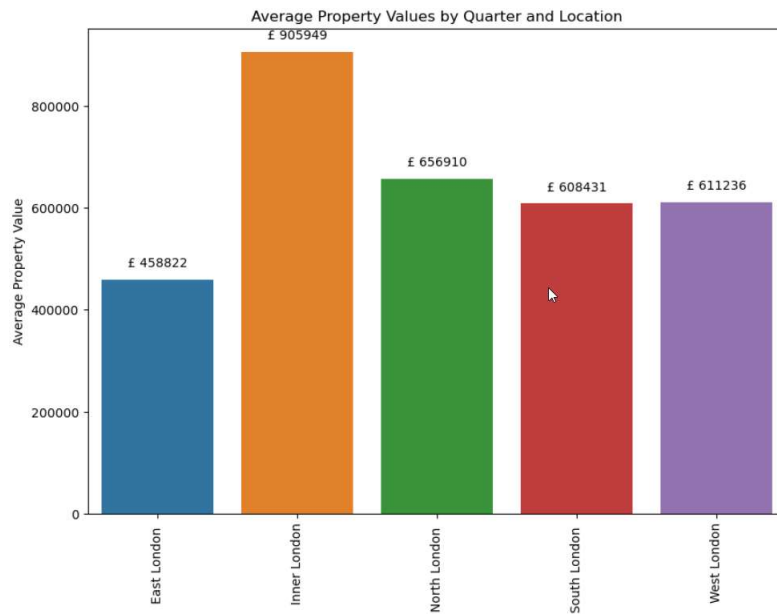
2020: The market demonstrated continued growth with a YoY increase of 3.5% in 2020-Q1 and 4.9% in 2020-Q3, but with a slight slowdown in 2020-Q2 (0.7%) and 2020-Q4 (3.5%).

2021: The market had a mixed performance with a YoY decrease of -3.1% in 2021Q1 and -4.4% in 2021-Q2. However, it rebounded strongly with a YoY increase of 4.7% in 2021-Q3 and a significant 10.7% in 2021-Q4.

2022: The market growth slowed down, with YoY increases of 4.9% in 2022-Q1, 1.2% in 2022-Q2.

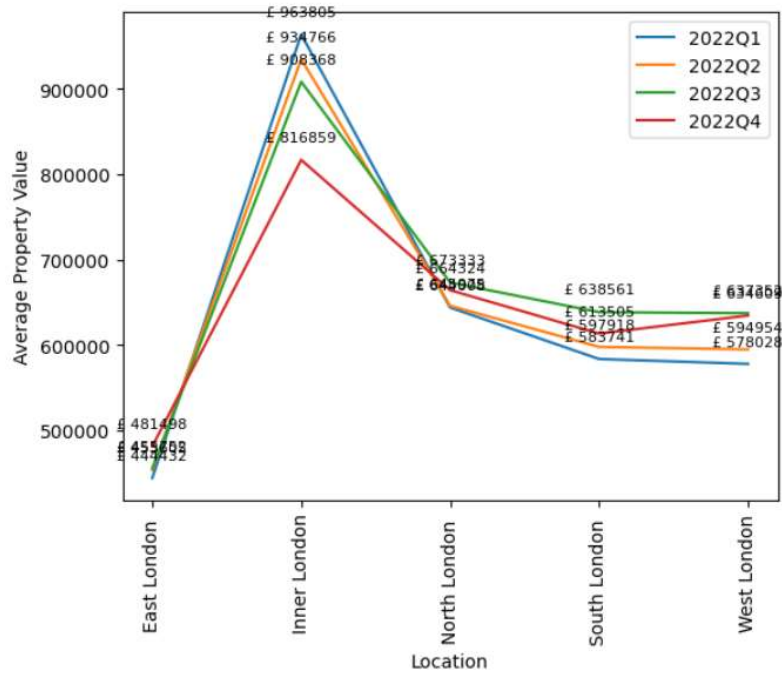
By comparing both graphs above, the number of properties sold are almost inversely proportional to the increase of the prices, and it's basically due to the increased/decreased demand of properties and availability in the market. And if checking during the Lockdown period due to covid, trends are extreme, we can see that properties sales have dropped by over 50%, and slowly started to recover when it has finished and people could be able to go out to buy their properties, this caused a spike on the average price of properties due to high demand for these.

Property Price per Location - 2022

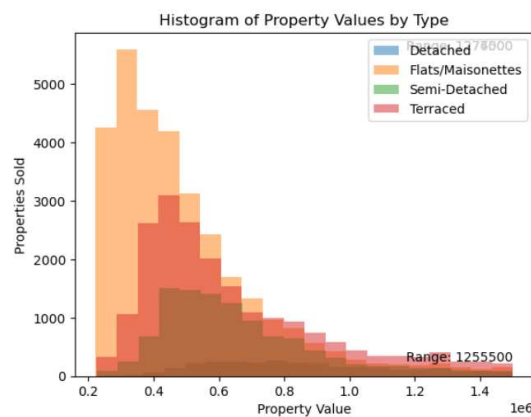


Based on the provided data on the average property prices in Greater London for year 2022, it can be observed that Inner London has the highest property price at £905,949, while East London has the lowest average property price at £458,821. The other regions fall in between, with South London at £608,431, West London at £611,235, and North London at £656,909.

This data suggests that there are significant variations in property prices across different regions of Greater London, with Inner London being the most expensive area to purchase property. It could also indicate that there may be differences in the types of properties available in each area or differences in demand for properties in those locations.



Having a look at the average per quarter and location, it can be observed that there are some variations in property prices across different regions of Greater London even within the same quarter. This data allows for comparisons of property prices between different regions and quarters, which could be useful for various purposes such as investment decisions for current year, for instance, we can conclude that East London Prices have not suffered much increase, compare to average prices in Inner London, where the average price have been dropped significantly on the last quarter, by almost £100.000.

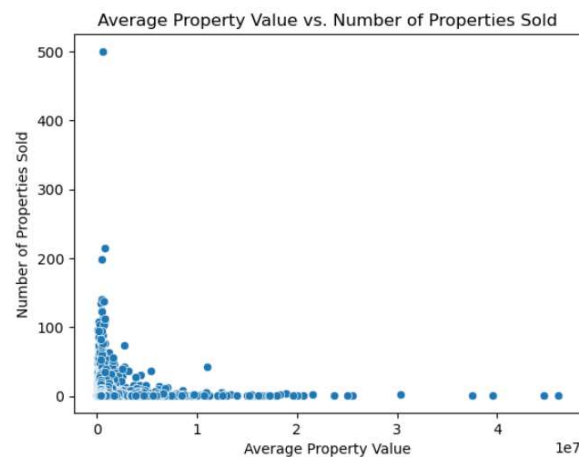


Having the above histogram, we can conclude the most sold property type are "Flats/Maisonettes" concentrating almost half of the type of properties sold for the last 5 years, on the second most common property type are "Terraced houses", having around half of the first property types. The least common property type is Detached, and it is precisely because these are the most expensive types.

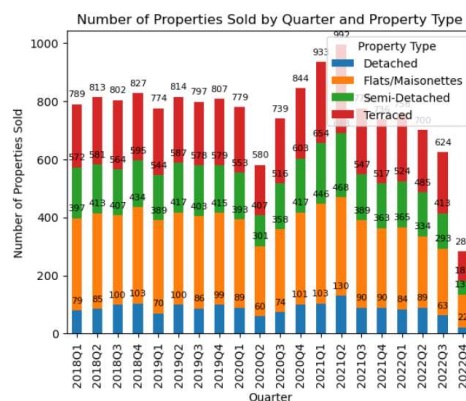
Predictability Trends Using Modelling and Machine Learning Techniques

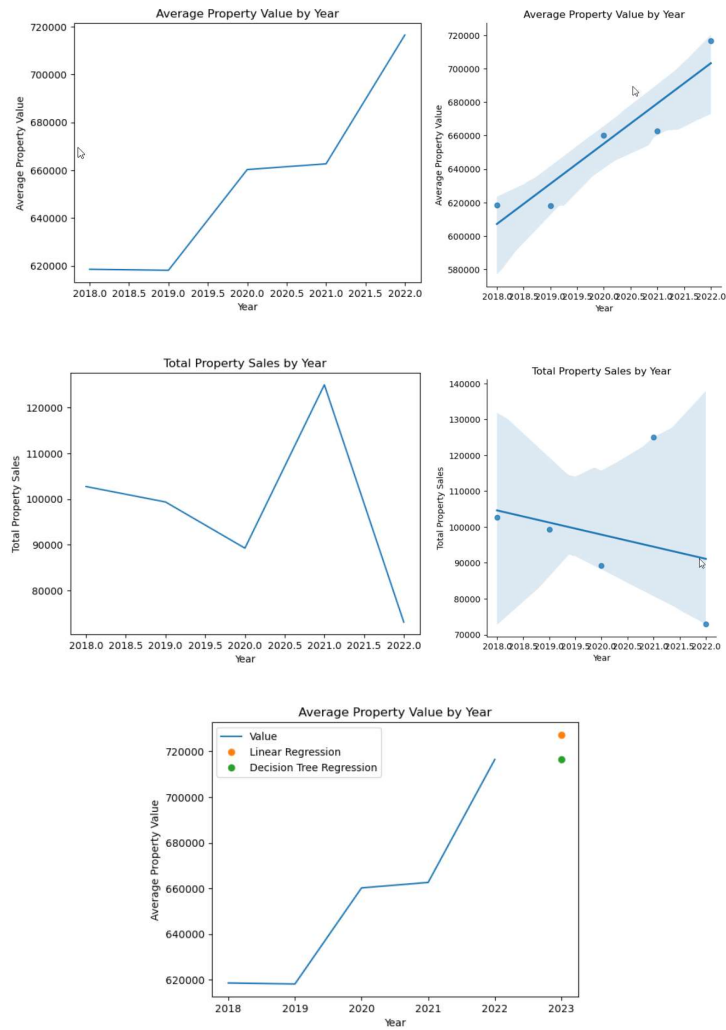


Having the above Heatmap type of graph, we can identify where and what are the type of properties more expensive or cheaper of the Greater London. As it is remarkable shown, Houses in Inner London, detached type is the most expensive, with over 3 million GBP average. The cheapest price to buy is East London if these are Flat/Maisonettes type.

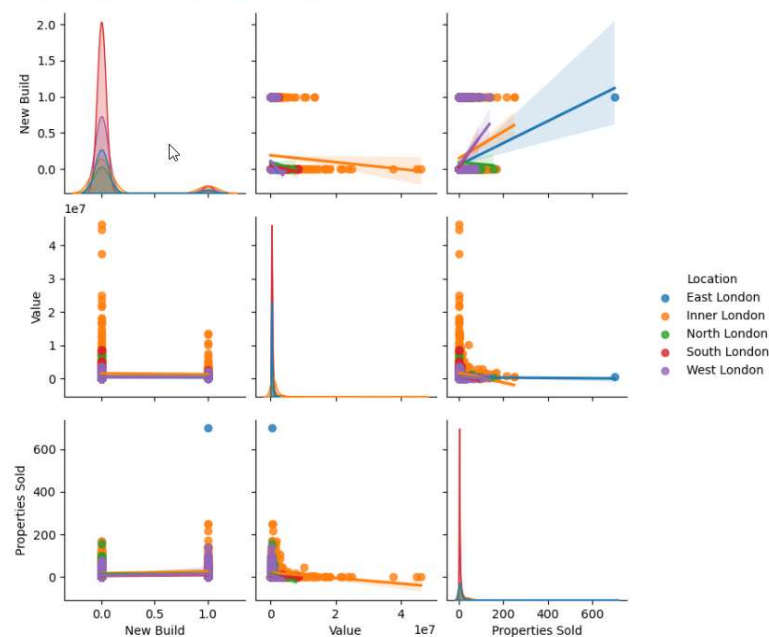


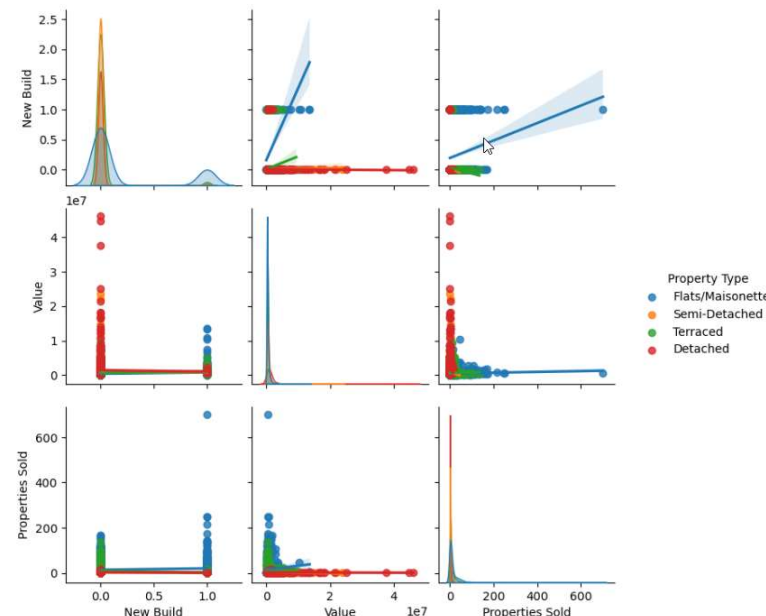
Having the above scatterplot graph, we can see the distribution of number of properties sold vs average price, it's clear that most numbers are below 1 million GBP, having very extreme cases of properties sole above 3 million.



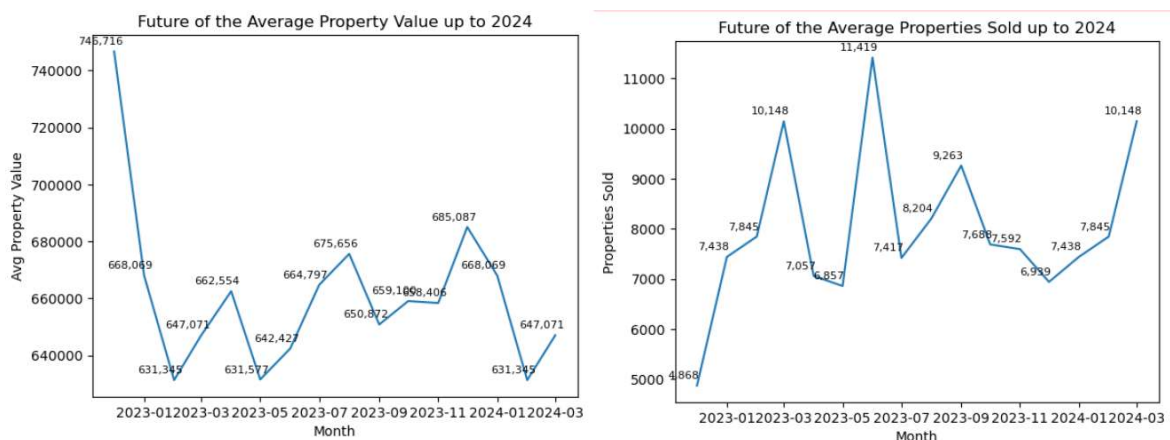


```
4 sns.pairplot(temp, hue='Location', kind='reg');
```





Regrettably, the **pairplot** graphs presented above look saturated due to the substantial number of records and the complexity of the data loaded. As a result, data points appear quite scattered, making visualization and interpretation challenging, even when aggregated by quarter. The graphs employ two variables: Property Type, which considers the number of properties sold, and Location. From the available information, it is evident that properties in Inner London are considerably more expensive than those in other areas, particularly in comparison to the East Side of London.



Last but not least, the above graph displays a trend on future property average values using data from past years. It's likely that the average price of property will decrease around 2 percent in 2023, which seems to be positive trend for those families who would like to buy a house. For the second quarter of 2023, it is predicted that the number of properties that will be sold are increasing, so their average prices, however, towards the end of the year, this trend will be reversed.

Insights

Having all the different type of graphs, below some of the key insights that we can identify.

1. Overall trend: The data shows that the number of flats/maisonettes sold in Greater

London was consistently higher than any other property type during the last five years, while detached properties had the lowest sales numbers. However, the sales of all property types fluctuated over the years.

2. Quarterly trend: The sales of all property types fluctuated from quarter to quarter over the last five years. In general, the sales of all property types were lower during the first quarter of the year and higher during the second and third quarters. The fourth quarter of the year was somewhat mixed in terms of sales numbers.
3. Year-on-year trend: Although there were fluctuations, there was a general trend of increasing sales for all property types from 2018 to 2021, with a decrease in sales numbers observed in 2022.
4. Seasonal trend: There is a seasonal pattern in the sales of different property types, with higher sales numbers in the spring and summer quarters (Q2 and Q3) and lower sales in the winter quarters (Q1 and Q4).
5. Property type trend: Flats/maisonettes were consistently the most popular property type during the last five years, with a clear increase in sales numbers from 2018 to 2021. Semi-detached properties had the second highest sales numbers and showed a slight increase in sales from 2018 to 2021. Detached and terraced properties had lower sales numbers and showed more fluctuations.
6. The impact of the COVID-19 pandemic: The COVID-19 pandemic had a clear impact on property sales in Greater London, with a sharp decrease in sales observed during the second quarter of 2020 (Q2) and a gradual recovery in subsequent quarters. The decrease in sales was more pronounced for flats/maisonettes, while the sales of other property types were also affected to a lesser extent.

References

- HM Land Registry, 2023. *How to access HM Land Registry Data*. [Online]
Available at: <https://www.gov.uk/guidance/about-the-price-paid-data#:~:text=The%20amount%20of%20time%20between,2%20weeks%20and%202%20months.>
[Accessed 24 02 2023].
- HM Land Registry, 2023. *How to access HM Land Registry Price Paid Data*. [Online]
Available at: <https://www.gov.uk/guidance/about-the-price-paid-data#explanations-of-column-headers-in-the-ppd>
[Accessed 23 February 2023].
- landregistry.data.gov.uk, n.d. *HM Land Registry Open Data*. [Online]
Available at: <https://landregistry.data.gov.uk/>