**An Investigation on the Impact of COVID-19 on the Housing Market of London**

**Jianqiang Li**

**Student ID:16018626**

**Abstract**

1. **Introduction**

In 2020, a global pandemic named COVID-19 stormily spread across the world and caused one of the biggest economic downturns in the 21st century. By the June 2022, this global pandemic has recorded more than 500 million cases and 6 million death globally, and in the first year of this global pandemic, it led the global economy with a global economic contraction of -3.3%, which is the worst downturn since the Great Depression(IMF, 2021). The impact of COVID-19 led to a global financial crisis which changes our daily life with terms that were unknown to us, such as lockdowns, social distancing, the terms that led to new solutions for life, throughout this pandemic, we find ways to work and study remotely, tourism becomes luxury in many regions.

Meanwhile, with government intervention and global recession, shrinking supply and uncertain demand, the impact of COVID-19 on the housing market is likely to be significant. The need for dwelling is one of the most basic need of humanity, it is essential to have a sound, safe, and large enough space as the shelter for all mankind, no matter rich or poor. The existence of housing market satisfies this basic need, it provide opportunities for people to have a house of their own. Nowadays, house is not simply just a place to stay, it is a legal property, a estate, a commodity, an investment, access to employment and school, service around the area(Bergenstråhle, 2016). As one of the most essential sector to our daily life, the size of housing market occupies a great percentage of the GDP for many countries. At the same time, real estate is also one of the most popular investment choice, many people prefer to purchase houses to hold on to their asset rather than in financial assets(Zhu, 2014). It is safe to say that the housing market is an indispensable part of our life, as well as for the global economy.

However, depending on the level of impact COVID-19 had on the local economy and different environment and infrastructures of the area, the housing market of each area is likely to receive completely different influence from this global pandemic, and it is hard to examine the global housing market as a whole and conclude on how COVID-19 changed the global housing market. Thus, this paper selects London as the researching city, as one of the biggest city in the world, the infection level of COVID-19 in London is also one of the highest in the world, which means if COVID-19 do had influence on the trend of the housing market, it is likely to be significant. This paper aims to provide novel evidence of why it is important to review the possible consequence of global recession caused by this global pandemic on the housing market, as well as the possible consequence of COVID-19 on London housing market and the reason behind it. Moreover, we uses the related dataset of each London boroughs to prove the predictions on the housing market base on literatures and examine the impact of infection level on the housing market. To better identify the research purpose, we define them into three major research question:

*1/Is possible to predict the trend of the housing market under this global recession caused by COVID-19?*

As the saying goes, history is a mirror, and if we look back at the history of modern economy, there are many remarkable global recessions base on various reasons. Although COVID-19 as a new pandemic, it has its unique characteristics, but it still need to follow the basic principle of an ecosystem. So, is it possible that we can theoretical predict the trend of a typical housing market suffering from COVID-19?

*2/Does different levels of infection rate in each boroughs lead to different level of impact on the price level and sales level of London housing market?*

This cases study examine the relationship between COVID-19 and the housing market at boroughs level in London, however, due to difference in position, environment, and other infrastructure such as education, pharmacy, it is likely that the housing market for each boroughs are totally different to other boroughs, and at the same time, the infection level of each boroughs are also likely to be distinct depending on factors such as population, regulation policy. In this case, how does the housing market in different boroughs react to the change in infection level of that area? Is there any linear relationship between key variables?

*3/Does the result we received based on data analysis match our theoretical prediction on London housing market?*

Through literature review and data analysis, we are able to receive two different outcome, however, does the prediction we made base one literatures match the result we received from analysing London boroughs dataset? If not, what is the reason behind it? What could be the better solution of predicting the trend of London housing market?

1. **Literature Review**

Since COVID-19 spread across the globe, the impact of COVID-19 on the global economy is significant, and it does no difference for the housing market. However, why should we pay attention to the changes in the housing market? How does COVID-19 impact on the housing market? Is it able to reflect the changes in the housing market from past financial crisis? We will explore the influence of the housing market by answering those questions in the following sections.

**2.1 The importance of the housing market**

Back to when human first appeared on earth, the first human kind started to live in caves, the need for dwelling started to engrave in our genes. And in the modern society, the housing market plays an enabler of economic activity, as the housing market satisfies the need of space for both individual and companies, by providing individuals with houses and apartments and companies with workspace with facilities, these basic need created the housing marketing and led it to the size it is today.

And as an essential part of the real estate industry, the transactions in the housing market attributes a great value to the financial statistic of an economy, and by combining the markets derived from the housing market as well as the supply chain of real estate industry, it created the real estate ecosystem. In 2019, the real estate as a whole contributed more than £100 billions to the UK economy, which represents about 7 % of the total UK GVA, at the same time, the whole sector also employed more than 1.2 million people in the year of 2019(BPF). From statistical perspective, the real estate sector has always been one of the most important sector contributing to the economic growth of an economy, thus, it also mean that it is essential to remain the stability of the housing market in order to maintain sustainable economic growth.

The fundamental of housing market indeed is to supply legal property to people in need, however, as the major commodities in the housing market: houses, apartments, office buildings, they are not only just a home or workplace, the property itself is a commodity as well as an investment. From long time ago, owning properties is a sign of wealth and status, and in the current society, investing in real estate is one of the most common accepted investment method(Glickman, 2014).

**2.2 The Influence of Housing Market on Stabilising Economy**

Through out history, changes in housing market have often been detrimental to the financial stability and the real economy(Zhu, 2014), the size of housing market and the unique characteristics of real estate property determine the distinctiveness of housing market for an economy. The assets of real estate are buildings, and the large size of the built environment makes the housing marketing asset intensive, at the same time, investing property often requires a large amount of capital, and the process of aggregating the required capital become a critical part of our financial system(Glickman, 2014).

In the process of aggregating capital for property investment, the use of loans is one of the most frequently used method, according to the Office of National Statistics of UK(ONS, 2019), from April 2016 to March 2018, the total household debt in the UK was £1.28 trillion, and this number is only going to increase. However, the reason behind the statistics of household debt can be complex, there could be two major reasons for using debt to purchase a property, the first one is obversely when a family need a house, in this case, it is a compulsory need, which means these needs are likely to be price inelastic or rate inelastic, the rate stand for the rate of debt; another reason of purchasing property using debt is when the investor believes that the value added to the property they purchase will outruns the rate of debt, in this case, it is not a compulsory need, and the housing market fluctuates, this type of debt is likely to change significantly. Thus when the housing market faces crisis, the decrease in the size of household debt could cause serious problem for the banking sectors, and for the investors, they may not be able to pay for their debt due to the recession in the housing market and the decrease of rental income.

For all the transactions in the housing market, tax is an essential part of the process. In some areas, the tax for property transaction can be very high, and the requirement of taxation could be a major driver for decisions in the housing market(Glickman, 2014). The types of taxation included in property transactions could be complex,but the principle behind taxation is simple, the aims of property taxation is to regulate the housing market and control the amount of transactions in the housing market, the logic behind it is to avoid the manipulation of housing price with frequent transactions. At the same time, the taxation for property transactions also contributes a great proportion to the tax income of government, which means that low level of transactions in the housing market could lead to financial issues for the local government.

According to the economist Arthur Pigou, there is a phase named “the Wealth Effect” between consumption and the value of housing and financial assets, he stated that when the asset values are high, the consumption level goes up, when the assets value are low, consumers reduces spending, this is simply because the changes in asset values could strongly influence consumer’s confidence on spending(Glickman, 2014). This is an interesting theory introducing the influence of asset value on the level of consumption, based on that, considering that the value of housing accounts more than half of the total asset values for many families, it is likely that price fluctuations in the housing market is likely to impact significantly on the consumption level of that economy. Thus, it is essential for an economy to maintain the stability of its housing market in order to boost the consumption level and achieve sustainable economic growth.

**2.3 COVID-19 and the Housing Market**

In 2020, the first COVID-19 case appeared in Wuhan China, due to the high level of infectiousness of COVID-19, it soon spreads across the world and became a global pandemic. COVID-19 as one of the most influential virus in human history, it has changed our daily life in many aspects. Throughout this global pandemic, term such as lockdowns, social distancing, quarantines, these are used to be unknown to our life now became a big part of life. The regulation took by the governments indeed helped reducing the infection level of COVID-19, however, it also leads to a global recession as it reduces consumption significantly(Chetty et al., 2020; Horvath et al., 2020), increases in unemployment (Beland et al., 2020; Borjas & Cassidy, 2020; Dingel & Neiman, 2020; Koren & Pető, 2020). From consumer’s perspective, the regulation policy and government interventions due to COVID-19 altered consumer behaviour(Alexander & Karger, 2020; Coibion et al., 2020), including the housing market.

Demand and supply are the basic principle that creates the balance of an economy, and there is no difference for the housing market. On the demand side, we can distinguish consumers by their intentions of buying, dwelling or investing, the intention of buying in some ways determines the impact of other factors that influences consumer’s decision. During this pandemic, families that intend to buy new house due to dwelling need are less likely to be influenced by COVID-19 and the price fluctuations in the housing market, however, the chance of investors to change their mind are likely to be higher. Social distancing is one of the most common regulation throughout this global pandemic, it may introduce frictions and transaction costs to the buyer-seller matching process that may in turn affect prices and liquidity(D’Lima and Lopez, 2021). Meanwhile, lockdowns as one of the most effective way to reduce the infection rate of COVID-19 in one area, it is implemented by governments across the globe in national level and local level. For the housing market, the implementation of lockdowns simply avoided face to face interactions among buyer, agency and seller, and the communication cost is likely to increase dramatically, and due to lack of information buyers are able to receive without visiting the property, it indeed impacts on the incentive of purchasing for buyers. Overall, the impact of COVID-19 on the consumer behaviour of the housing market is likely to be significant, the lack of information exchange opportunity and trust issue could temporarily reduce the demand of housing. The impact of regulations on demand is only temporarily, demand still exists, but on the supply side, the impact of COVID-19 on the construction sector and related supply chain is significant, this reduces the supply of new properties dramatically in this pandemic(Duca and Murphy, 2021). In an ecosystem, when demand is constant, supply falls, price goes up, this is no difference for the housing marketing. In the housing market, the adjustment of the stock of dwelling is generally held to be quite low, it is often argued that the supply of housing can be considered as fixed(Kenny, 1999). However, the high level of infectiousness forces many government to implement lockdowns with the region, as lockdown forces people to stay at home, it increases the construction duration and causes supply chain disruptions, which in turn reduces the supply of new houses during this pandemic. For another major suppler of housing market, previously-owned homes accounts for a great percentage of housing transactions, and due to COVID-19, shrinking demand in the rental housing market could encourage more property owners to sell their property, however, the chances are likely to be low, at the same time, reduced house trading agency and agent could also lead to a illusion that supply is reduced, and this illusion will recover immediately when the economy recovers from COVID-19. Overall, the impact of COVID-19 is likely to causes obvious decrease on the demand in the housing market, and potential reduction in the supply of dwelling, which means the price level and sales volume of housing market is likely to fall during this pandemic, but only temporarily.

The prediction base on demand and supply relationship excludes all other factors that would influence the trend of the housing market, and in reality, no one is expecting a continuous economic downturn, especially for local governments. During this pandemic, in order to minimise the impact of COVID-19 on the housing market and stimulating economic activity by encouraging consumption, the UK government announced the stamp duty holiday in July 2020(Scanlon, 2021). The stamp duty holiday raised the threshold for tax payment to £500,000 until end-March 2021, and it was extended several times afterwards. The decision of government announcing this holiday strongly reduces the transaction cost in the housing market, and on the consumer side, the announcement of the stamp duty holiday is likely to boost the incentive of purchasing, and on the other hand, increased number of transactions could also create and sustain jobs, which strongly reduces the negative impact of COVID-19 in the housing market.

In the housing market, there is one other factor that influences price level regardless of time, which is inflation. High levels of inflation could lead to increases in the overall price level in the housing market, however, since the impact of inflation is on the whole economy, the price change in the housing market due to inflation does not change the true value of a property, at the same time,it may cause illusion for consumers that housing marketing is increase above expectation. Thus, if inflation level is predicted to increase, this would encourage people to shift consumption from future to present(Armantier et al., 2021). Throughout this global pandemic, due to restrictions and economic crisis, the consumption level for consumer are at a very low level, many government are finding ways to encourage consumption to maintain economic growth, such as reduce interest rate, print more cash. And as result, the increase in inflation level can easily be predicted, which meets the objective of encouraging consumption for government. For the housing market, upcoming inflation and low interest rate encourages investors to move their money out of bank, and investing in real estate becomes a great choice and solution for inflation, which boosts the demand in the housing market.

**2.4 Conclusion**

In conclusion, the existence of housing market not only satisfies the most basic human needs of dwelling, it is also an essential component of an economy. From economical perspective, the housing market is one of the major contributors to the GDP for most countries, and it also contributes to the employment of a nation as it provided more than 1.2 million job for the UK in 2019. Meanwhile, with the unique characteristics of housing property, it is also one of the most popular investment in the global economy, upon that, as the transactions in the housing market often requires large amount of capitals, the great amount of housing debt again contributes to its importance on stabilizing economic.

Thus, it is the essential maintain the stability of the housing market in order to achieve sustainable economic growth.

As the spread of COVID-19 has led the global economy under second biggest recession since 2007 banking crisis, the performance of the housing market could potentially determine the economic performance during this pandemic. London as the biggest city in the UK as well as the biggest housing market in the UK, it is likely to be a major influencing factor on the economic performance for the whole nation. And during this global pandemic, it is critical to investigate the impact of COVID-19 on the housing market to support the regulation policy against COVID-19 for the UK government. Base on literature research and the basic demand and supply principle, we assume that the spread of COVID-19 is likely to cause decrease in both demand and supply in London housing market, regulation policies like lockdowns and social distancing not only increase the construction duration, it also lead to lack of opportunity for information exchange as face to face interaction is not encouraged. Since the influence of COVID-19 is likely to be more significant on the demand side of the housing market, it is likely to cause fall in the price level and sales volume for London market. However, during this pandemic, regulation policy against pandemic is not the not factor that influences the price level of London housing market, the announcement of Stamp Duty Holiday is likely to boost the demand level significantly as the cost of SDLT is strongly reduced during the holiday. Meanwhile, The increasing level of inflation could also lead to the increase in housing prices as the currency itself becomes less valuable.

1. **Data**

Base on the research questions we stated above, we decide to obtained data from three perspectives, dataset for London housing market, dataset for COVID-19 in London and dataset of London that consist factor that may influence consumer decision in the housing market. For the uniformity and reliability of the dataset, all the dataset we obtained in this paper are originally from London DataStore.

For the housing dataset, we choose to use a dataset that consist of four major type of data for London housing market from January 1995 to March 2022, the four data types are: average price at boroughs level, index price at boroughs level, sales volume at boroughs level, and average price and index price of different property types at London level. According to the information in the original dataset, the index housing price is calculated by taking the average (geometric mean) price in January 2015 and then recalculating it in accordance with the index change back in time and forward to the present day, which means it is an adjusted price index that represents the housing price level of an region.

For the COVID-19 dataset we obtained in this paper, we selected a dataset that consist of daily new cases and daily counter of total cases in each London boroughs from 2022/2/3 to 2022/6/13. The dataset of daily COVID-19 cases can help visualise a more detailed trend of how each London boroughs controls the spread of COVID-19, but since the population of each boroughs is completely different, we also uses the population data from 2021 census in order to convert case numbers to infection rate, which can help us to find a more accurate infection level of each London boroughs. Meanwhile, since other dataset included are all monthly data, we also need to covert this COVID-19 dataset into monthly format.

Apart from datasets for London housing market and COVID-19, there are also other factors that may influence consumer’s decision in the housing market and distinguish the housing market of one area. In reality, price is certainly not the only factor that consumers considers when they are picking a property, it could be education resources, environment, security and many attributes that are related to our daily life. And in this paper, we selected education resource and environment as the representative factors, since we expects those attributes are not likely to change significantly in the past years, we uses datasets of average GCSE score in 2020/2021 and the percentage of green spaces in 2014 of London boroughs as the attributes of education resource level and environment level of London boroughs.

1. **Methodology**

**4.1 Data Processing and Cleaning**

At the first stage of analysis, we need to process and clean the original dataset we downloaded from the London DataStore and convert them into the format we needed. Since COVID-19 dataset is downloaded as CSV format, we directly read in the dataset using pd.read\_csv. The original dataset of COVID-19 has five columns: *area\_name, area\_code, date, new\_cases, total\_cases*, and for descriptive analysis purpose and converting it to monthly data, we choose to covert the original dataset to borough vs time format separately for each London boroughs, and use pd.merge to join them together afterwards. Then, we set date index for the processed dataset, and we use resample('M').sum() to calculate the monthly new cases and use resample('M').max() to calculate the total cases on the last day of each month. At this step, the column of time for COVID-19 data is still in yyyy/mm/dd format, in order to prepare for further analysis, we convert the column of time to yyyy/mm and added two new columns of year and month. Before finalise the data format for COVID-19, we need to calculate the infection rate for each boroughs, and to do that we manually copied the London population data at boroughs level from 2021 census, and calculate the monthly new cases of each boroughs in every 1000 people. The specific formula is as follow:

*I = n/0.001p*

*I = Infection rate n = monthly new cases in a borough p = population*

Then,for further analysis, we convert the data format from boroughs VS time to the original format. Moreover, we also manually copied the data of average GCSE score and percentage of green space from the original dataset and joined them into the final COVID-19 data frame.

Since the original housing data are in XLSX format, we need firstly move the part of data we need into separate CSV file. And after splitting them into four different CSV file: Average Price, Index Price, Sales Volume, Housing Data by Type, we deleted the row of area codes and added the title for the column of Time, meanwhile, we also changed the “Time” column to yyyy/mm format and all the numerical cells to general format. However, since the data of Housing Data by Type is only at London level, we only use it for descriptive analysis.Until this process, we found that the list of boroughs among all the dataset used in this paper are not the same, the dataset for COVID-19 combined the borough of City of London and Hackney into one borough, and for the rest of datasets, they are two separate boroughs, considering the City of London is a very small area and to avoid outliers in the dataset, we decide to combine City of London and Hackney into one boroughs for all dataset. Therefore, for further calculation on average price dataset and index price dataset, we firstly combined the sales volume of City of London and Hackney and added a new column of Hackney and City of London to the sales volume data frame. And after read in all the housing datasets, we firstly calculated the average price and index price for the new added borough “Hackney and City of London”, the calculation is as follow:

*HC = h\*(hs/hcs)+c\*(cs/hcs)*

*HC = average or index price for Hackney and City of London,*

*h = average or index price for Hackney,*

*c = average or index price for City of London,*

*hs = Hacney sales volume, cs = City of London sales volume*

*Hcs = Hackney and City of London sales volume*

We use the proportion of sales volume to value the contribution the combined data. Then, we need to change the data format to match the precessed data for COVID19, we added new column of year and month and calculate the monthly change in percentage for Average Price, Index Price and Sales Volume. At this stages, all the data processing and cleaning are almost finished, we join the data of COVID-19 and London housing market using pd.merge. However, considering that the impact of COVID-19 may take time to emerge on the housing market, we use .shift(3) to lag the COVID-19 data backward for 3 months as the time for COVID-19 to impact on the housing market.

**4.2 Data Analysis**

Since all the data are precessed and cleaned, we uses line charts to visualise the trend of COVID-19 cases and London housing market, and we uses scatter plots to examine the relationship of between variables and visualise the distribution between variables. However, in order to determine whether COVID-19 infection level do have significant impact on the housing market, more analysis need to be made apart from descriptive analysis. In this paper, we decide to run OLS regression, VIF, Lasso, and Regression tree as the major data analysis solutions for this investigation.

For the dependent variable, we plan to select Average Price, Index Price, Sales Volume as the dependent variables for regressions, and we will run three different set of regression models to explain the research question. On the left hand side, since infection level is the index that is most appropriate to represent the infection level in one area, it is selected as the only COVID-19 related variable in the list of independent variables, and apart from the infection rate, GCSE score and green space percentage are also selected as independent variables. Moreover, considering that the different stage of COVID-19 and the other influential factors for the housing market, as the price level of dwelling decreases with distance from employment due to the trade-off between transport cost and rents(Alonso, 1964; Mills, 1967; Muth, 1969), we also adds a year dummy and a area dummy to the linear regression model.

**4.2.3 Linear Regression**

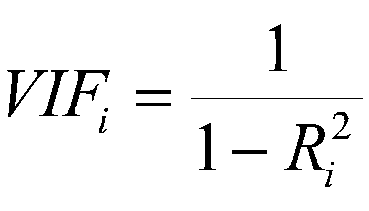
Since the research purpose of this paper is to investigate the relationship between COVID-19 and London housing market, linear regression is likely to be the most appropriate statistical method. The use of linear regression is to study the relationship between a dependent variable and one or more independent variable, in a linear regression model, all the calculation is based on the assumption of linearity between dependent variable and independent variables(Sluijmers, 2020).

The first regression model is OLS regression model, as one of the most commonly used regression model, it is a method that estimates the parameter of the linear model by finding the parameters that minimise the sum of squared residuals(Sargent & Stachurski, 2021). In this paper, we intend to use the OLS model to predict the price and sale volume in London housing market from a dataset of COVID-19 infection level and area informations, and the model writes as：

*Y = β0 + Σj=1..p βjXj + ε*

Where Y is the dependent variable, β0, is the intercept of the model, X j corresponds to the jth explanatory variable of the model (j= 1 to p), and *ε* is the random error with expectation 0 and variance σ²(XLSAT).

The second model is the Variance Inflation Factor(VIF), VIF is a tool to help measure the degree of multicollinearity. Multicollinearity is a measure of the degree of correlation between variables, and in a regression models, the multicollinearity between independent variables strongly affects the performance of the model, this is also the reason we only have infection rate as the independent variable(Gupta, 2021). The formula of VIF is simple:



The R²value is the statistical measure of how well the regression line approximates the real data points, and after calculation, we receive an index of the magnitude of collinearity, where 1-5 is considered small, and 5-10 or 10+ is considered high or extreme(Gakovic,2020).

The third model involved is the Lasso model, it is a example of regularized regression, it’s a regularization technique that reduces the complexity of a model by shrinking some of the coefficients to zero(Pykes, 2022; Tayo, 2020). The advantage of Lasso regression is that it considers all predictors simultaneously, and the process of Lasso regression is to perform feature selection using a penalty function:

min {SSE + λ ∑||β||1

β,β0

Where λ is a hyper-parameter controlling the strength of penalty effect, and ||β||1 is called L1-norm. And in this paper, we will use Lasso path as the visualisation plot to see how the coefficients change with the lambda value.

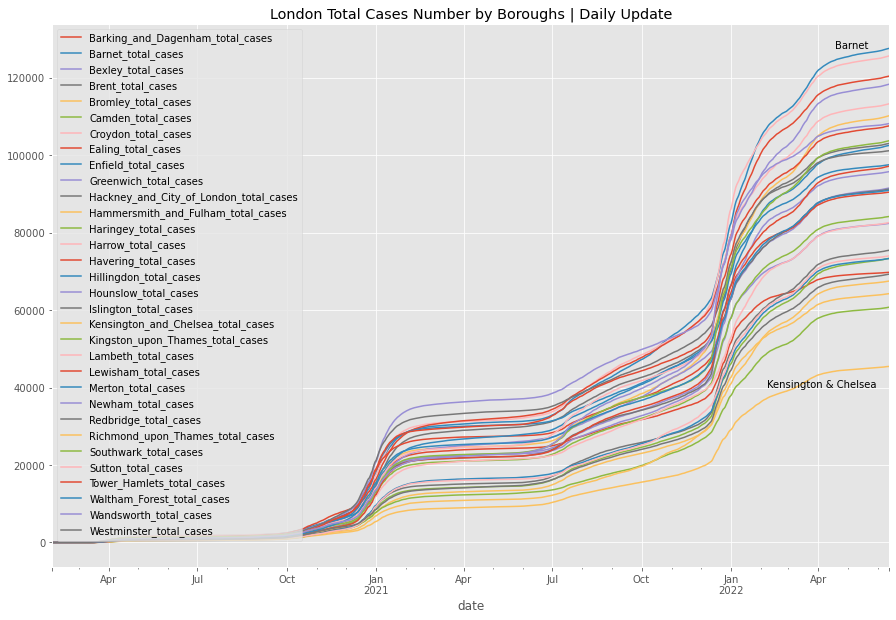
Regression tree is the final regression model in this paper, also known as the decision tree regression, it is simply a flow diagram or a ‘tree’ of decisions about the x variables of a dataset, and there are three type of nodes on the decision tree: root node, interior nodes, leaf nodes. The root node is the initial node, it connects all the interior nodes, the interior nodes represents the feature of a data set and the branches represent the decision rule, and the leaf nodes represents the outcome(Salunkhe, 2021). The advantage of regression tree is its Flexibility and Interpretability, it is relatively easy to visualise and faster on processing data as there is no assumption and transformation needed in this model, but the tendency of overfitting could cause issues in the following analysis.

The selection of four different regression models aims to explore the relationship between COVID-19 and London housing market in a more diverse way, and the use of four different model also increases the reliability of the conclusion we find through regression. Meanwhile, in order to avoid randomness of the data set, we will also created the training and testing dataset for all above regression models.

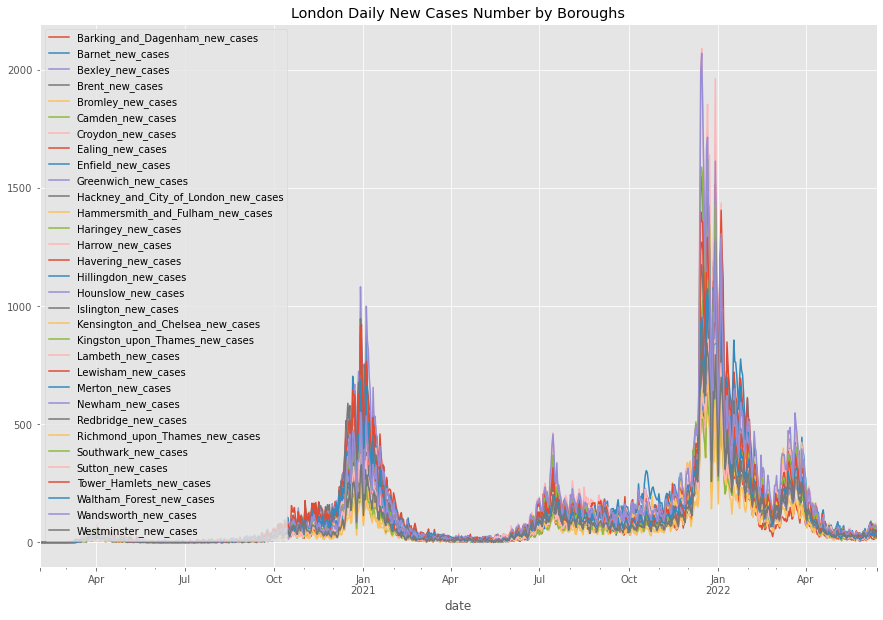
1. **Result**

**5.1 Descriptive Analysis**

Since February 3rd 2020, the first COVID-19 case in London was confirmed in the borough of Merton, the total COVID-19 cases number in London has exceed 3 millions. As Figure 1 and Figure 2 shows, the trend of COVID-19 cases for London borough are nearly identical, and there are two huge waves of infection occurred in late 2020 and late 2021. For the first wave in 2020, it occurred when the first national lockdown was eased in the UK, as people just came off a long term lockdowns, the level of social activity was rebounded, and it led to a surge in COVID-19 cases started at the end of 2020. For the second wave, it occurred at the time where a new variant named Omicron was found, and the extremely high level of infectiousness of Omicron soon led to another surge in case numbers in London boroughs at the end of 2021 and the start of 2020. And to the day we acquire this COVID-19 dataset, the borough of Barnet recorded the highest total cases number among all London boroughs with the total cases of 127,611 on June 13th 2022, and the borough of Kensington and Chelsea recorded the lowest total cases number, it actually is the only borough that had a total COVID-19 cases number below 50,000, which is about 15,000 lower than the second lowest London borough.

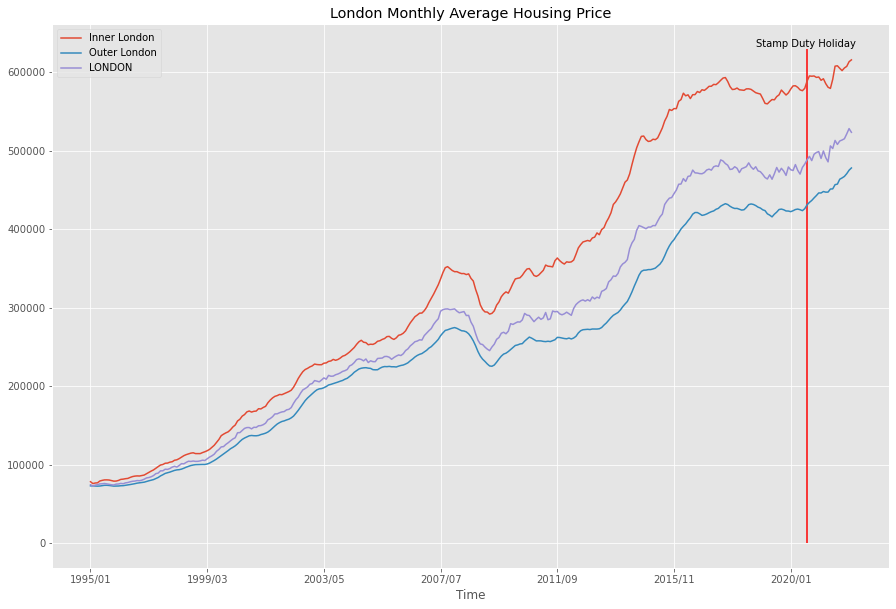


*Figure 1: London Total Cases Number by Boroughs* | *Daily Update*



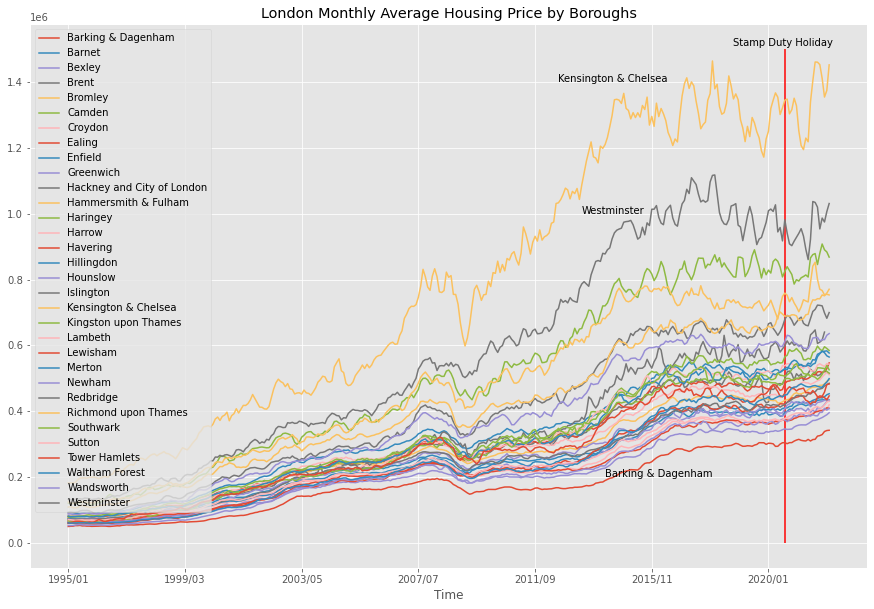
*Figure 2: London Daily New Cases Number by Boroughs*

Move on to the plots for London housing market, as figure 3 illustrates, the trend of average housing price for London are generally on a upward trend since 1995, the biggest downturn occurred in the second half of 2007 as the world was under the biggest economic recession in 21st century by far. However, as COVID-19 caused another global recession in 2020, the average housing price for London was still rising since this global pandemic started, and since the Stamp Duty Holiday was announced, we can significant increase in average housing price for both Inner and Outer London. Moreover, as figure 3 shows, the different in average housing price between Inner London and Outer London is extreme, and the huge gap between Inner London and Outer London indicate that it is likely Inner London boroughs and Outer London boroughs are two separate housing market.



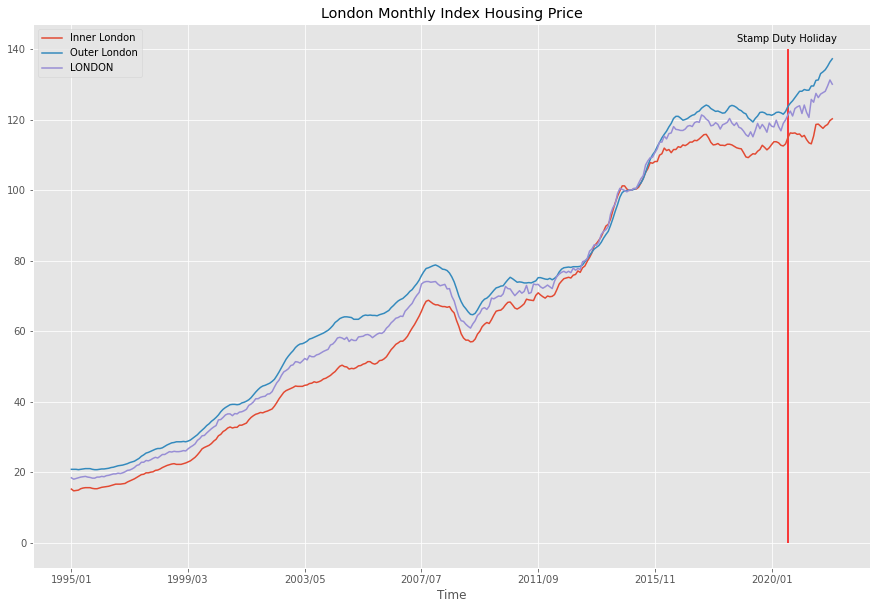
*Figure 3: London Monthly Average Housing Price*

Figure 4 represents the average housing price trend by boroughs, and we can see that the trend for boroughs in central London are more significant than other boroughs, and since 1995, the borough of Kensington and Chelsea has always been the most expensive borough in the housing market, followed by Westminster being the second, as they are also the only two borough among London boroughs recorded average housing price more than £1,000,000. As the borough of Kensington and Chelsea recorded a average housing price more than £1,400,000, the borough of Barking and Dagenham has not exceed more than £400,000 in the history of London housing market, which is the cheapest borough in London housing market. Meanwhile, looking at the trend after Stamp Duty Holiday, it is clear that some Inner London boroughs demonstrates a more significant upward trend than many Outer London boroughs, which again confirmed the statement that Inner London boroughs and Outer London boroughs are two separate housing market.



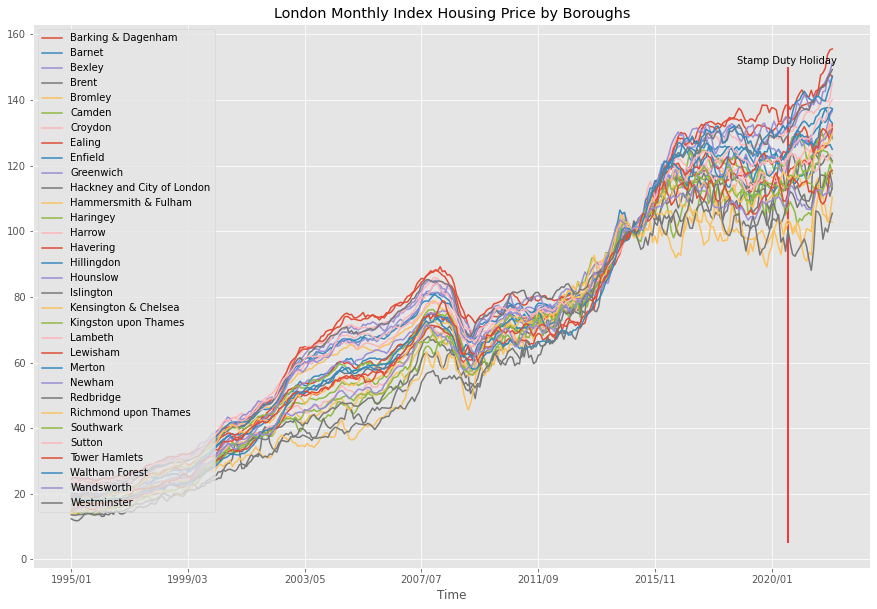
*Figure 4: London Monthly Average Housing Price by Boroughs*

Figure 5 demonstrates the trend of Index Housing Price for Inner and Outer London, compare to average housing price, the index price focus more on the changes of price level rather than price itself. And as the figure shows, the trend of index housing price is similar with the trend for average housing price, however, we can see that the line of Outer London is actually always on the top of other two lines, which means that the index price for Outer London boroughs has been growing faster than Inner London boroughs, and the trend after Stamp Duty Holiday also proves it as the increasing trend for Outer London is clearly more significant than Inner London. This also indicates the difference between index price and average price, since the index price is calculated base on the average price in January 2015, we can see there is a intersection for all three line at 100 index value and 2015/01, prior to January 2015, the higher index price is, the faster the average price grows, and after January 2015, the higher index value is, the faster average price grows. Therefore, as the figure shows, the average price for Outer London boroughs has been rising faster than boroughs in Inner London since 2015, and prior to 2015, housing price boroughs in Inner London was rising faster.



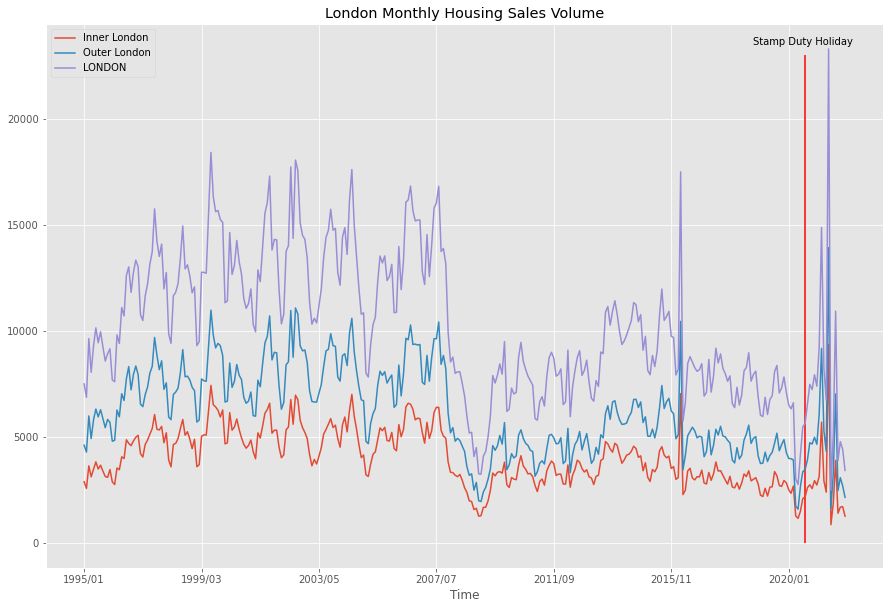
*Figure 5: London Monthly Index Housing Price*

In figure 6, this figure shows the monthly index housing price for London at boroughs level, same with the trend for Inner London and Outer London, all lines cross at index value 100 and January 2015, and compare to average price trend in figure 4, we can see that there is no outliers in figure 5, and in March 2022, the borough of Barking and Dagenham has the highest index housing price among all London boroughs, which means Barking and Dagenham is the borough that increased the most in percentage of the average price in 2015, and the lowest two are actually two Inner London boroughs, Hammersmith and Fulham, and Westminster. The trend of index price for London boroughs again proves our finding in figure 5, as the housing price for Inner London boroughs have been rising slower than Outer London boroughs since 2015 as well as the changes during the pandemic.



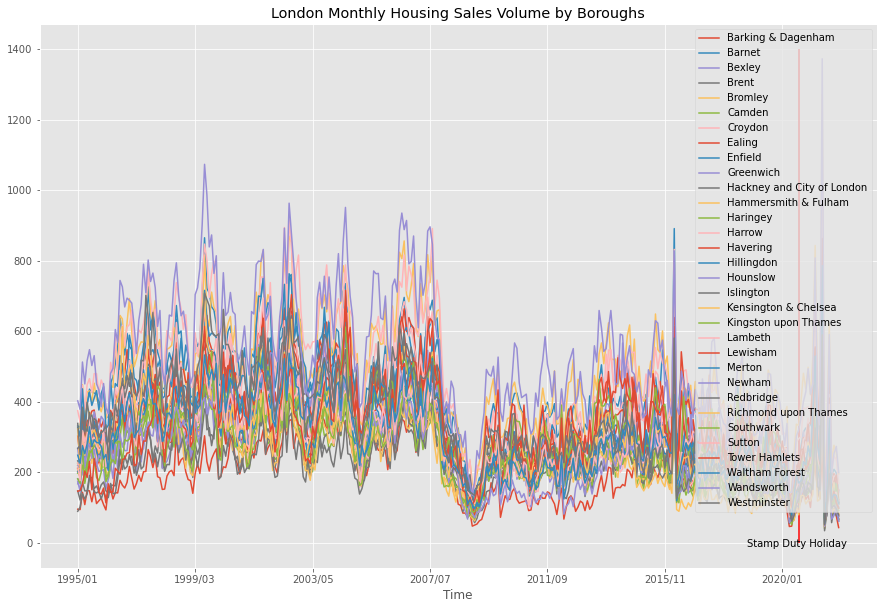
*Figure 6: London Monthly Index Housing Price by Boroughs*

Move on to the sales volume for London housing market, as figure 7 shows, the sales volume for Outer London has always been higher than Inner London, this may because the number of boroughs in Outer London are about double of the number of boroughs in Inner boroughs. As for the trend of sales volume, there is a significant downturn for the whole London during 2007 banking crisis, and the sales volume has not recovered since then. However, during this pandemic, with the announcement of Stamp Duty Holiday, the monthly sale volume for both Inner and Outer London reached a peak since 1995 in June 2021, which is also the last month of Stamp duty holiday with no SDLT charge for transactions below £500,000.



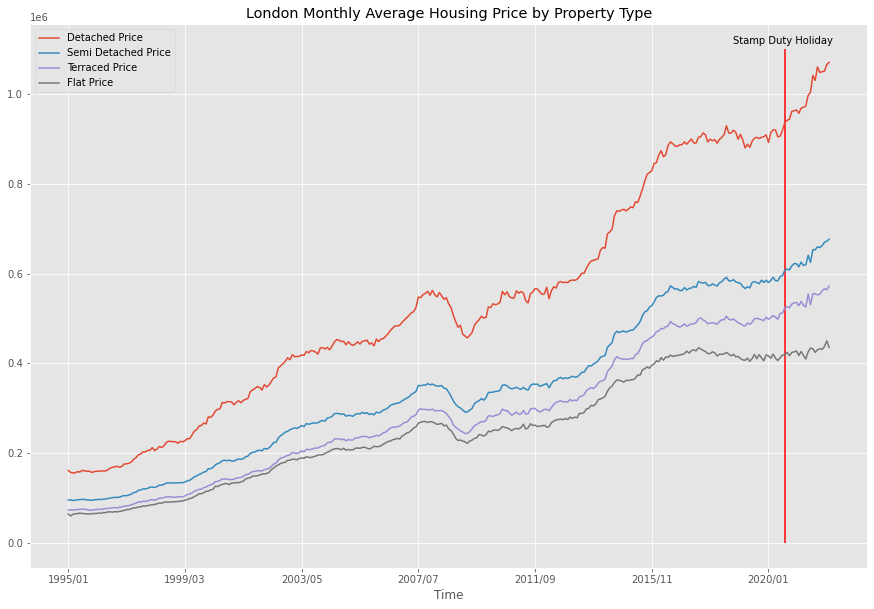
*Figure 7: London Monthly Housing Sales Volume*

Figure 8 shows the monthly sales volume for London housing market by boroughs, similar to the trend for Inner London and Outer London, but the gap between top boroughs and bottom boroughs are huge. In June 2021, the most active month for London housing market since 1995, the borough with the highest sales volume is Wandsworth, which is the purple line on the top of all other boroughs for the most of time, and for the least active boroughs, it has to be Barking and Dagenham, as it only 279 transactions in June 2021, which is the only London borough who recorded less than 400 transactions in that month.

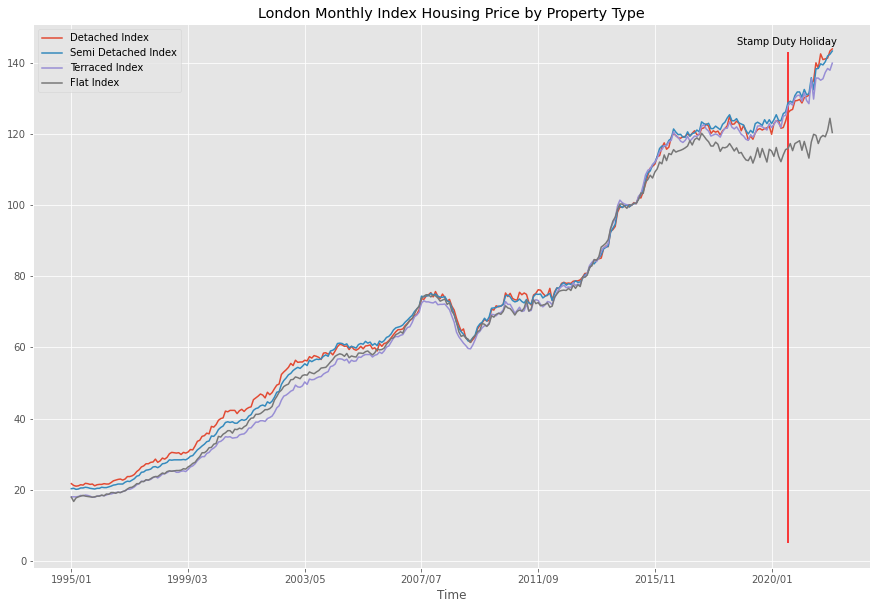


*Figure 8: London Monthly Housing Sales Volume by Boroughs*

In a housing market, there are different types of property, and in London housing market, there are four major types of property: Detached, Semi Detached, Terraced, and Flat. As figure 9 shows, the detached property is definitely the most expensive property type and potentially the most popular property type since 1995, and flats is clearly the cheapest among them. And as figure 10 shows, there are no significant different between the index price value for all four property types, but since 2016, the index price trend for flat has became steady compare to other three property types, this indicates that the price for flats in London has not changed much in recent years as it is also can be seen on figure 9.



*Figure 9: London Monthly Average Housing Price by Property Type*

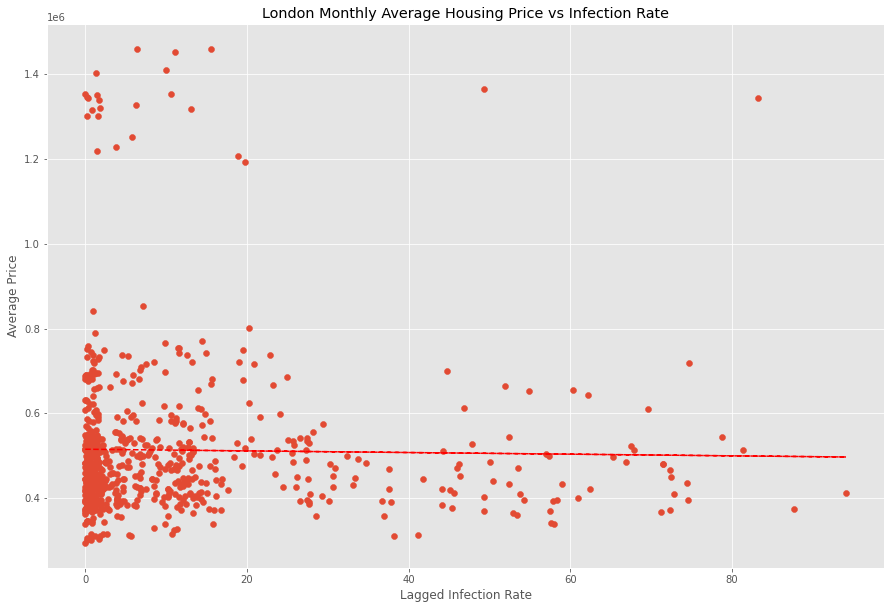


*Figure 10: London Monthly Index Housing Price by Property Type*

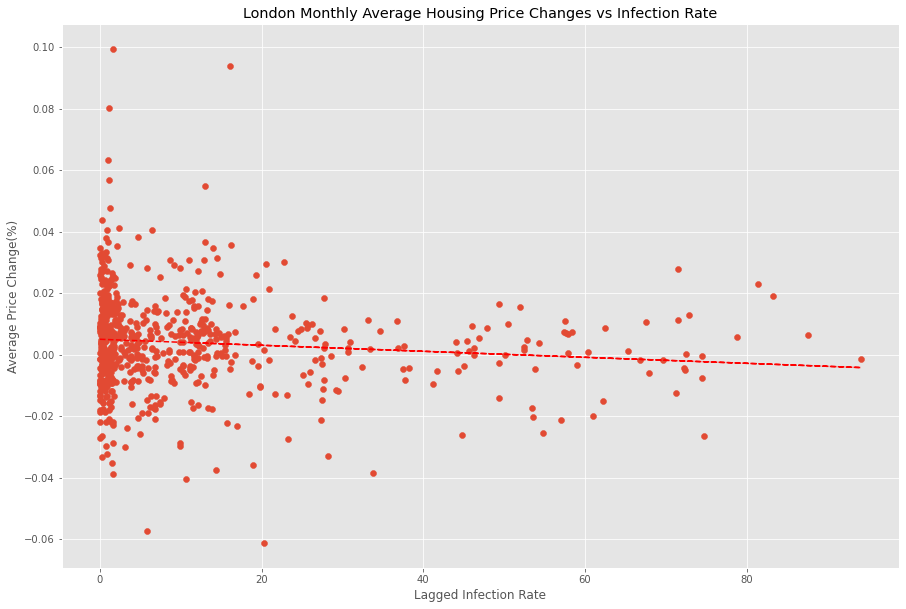
If we compare the trend for prices and sales volume, we can find that they actually have similar trends throughout time, if sales volume falls dramatically, price falls as well, if sale volume increases, price increases as well, this can be clearly found on the trend after 2007 banking crisis and Stamp Duty Holiday. Meanwhile, the economic recession in 2007 is the turning point for London housing market, the average sales volume decreased significantly after the banking crisis and remained at a relatively low level compare to prior to banking crisis, and this situation did not change until the announcement of Stamp Duty Holiday. However, compare the trend for London housing market between 2007-2009 and 2020-2022, the two period for banking crisis and COVID-19, the trend for London housing market under two different economic downturn period are totally different, during 2007 banking crisis, average price for all London areas falls dramatically, and so does the sales volume, but during the pandemic, price level and sales level both surged in the past few years. This may because of the announcement of Stamp Duty Holiday, but it is more because of the difference in fundamental causes for two recessions, and it also proves that it is not possible to predict the trend of housing market base on the trend in past economic recessions.

After the analysis of the trend of COVID-19 and London Housing Market, we uses scatter plots to examine the potential relationship between variables before the start of regression. And since infection rate is selected as the representative variable for COVID-19, we will only examine of visualisation for infection rate and three dependent variables, considering that the impact of COVID-19 may take time to show on the housing market, the infection level data we used for scatter plots is lagged 3 month backward.

As figure 11 visualises the distribution of variables between infection rate and average housing price, we can find that the distribution of variables are not normally distributed, there are many outliers on the top of the scatter plot. And as the trend line for figure 11 shows a relatively flat curve, it means that as the changes in infection rate is not likely to influence the changes in average price. However, since the average price level for each boroughs are different and the outliers on the top, this trend line may not represent the true relationship between two variables. Figure 12 shows the visualisation between monthly percentage changes for average housing price in each boroughs and the lagged infection rate, and compare to the distribution in figure 11, it is much more normally distributed, and only a few outliers spreads on the side. As the distribution between percentage changes in average price and infection level is different from just average price and infection level, the trend line is also changed, as it shows a negative gradient, it indicates that as the infection level increases, it is more likely that the change in average price is negative.

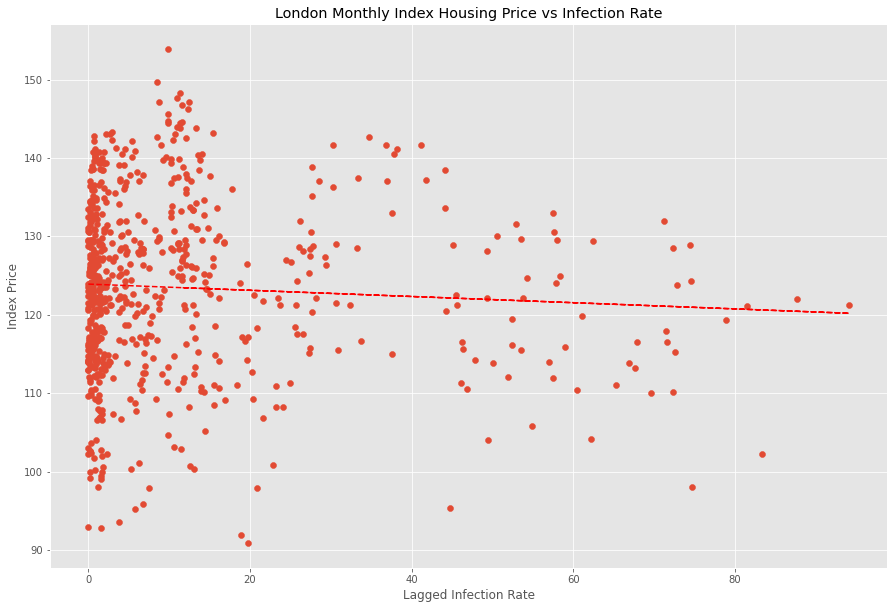


*Figure 11: London Monthly Average Housing Price vs Infection Rate*

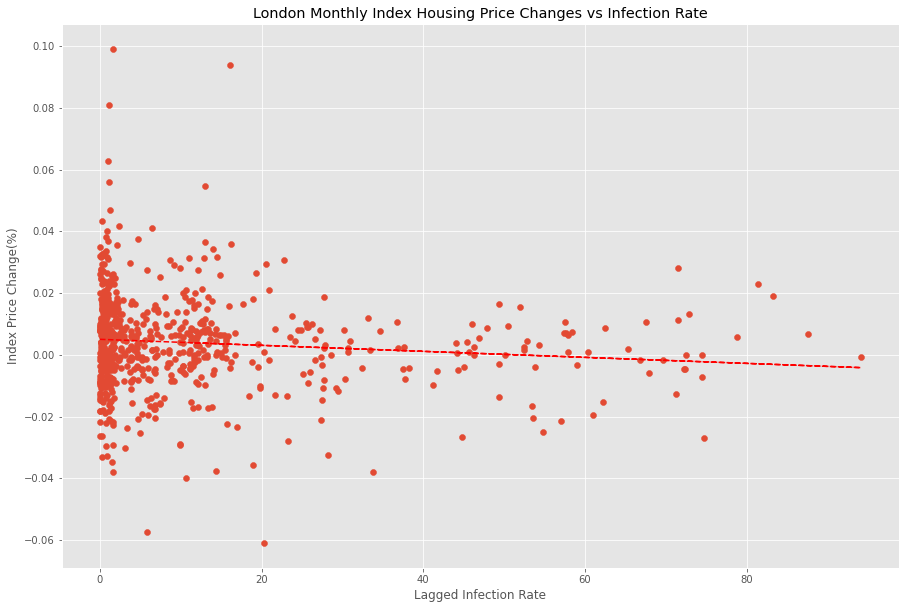


*Figure 12: London Monthly Average Housing Price Changes vs Infection Rate*

Figure 13 shows the scatter plots for index housing price and infection rate, and compare to the scatter plot for average housing price and infection, it is much more normally distributed, but they also have similarity, as the coordinates are all clustered on the left side of scatter plots. And as the trend line shows, the negative gradient of the trend line indicates that high infection rate is likely to lower index price. Move on to the scatter plots between index housing price changes and infection rate, the distribution in figure 14 is similar to the distribution in figure 13, but more narrowed to the line of no index price change, and same with the trend line for index price and infection rate, the relationship between index housing price change also indicates a negative relationship.

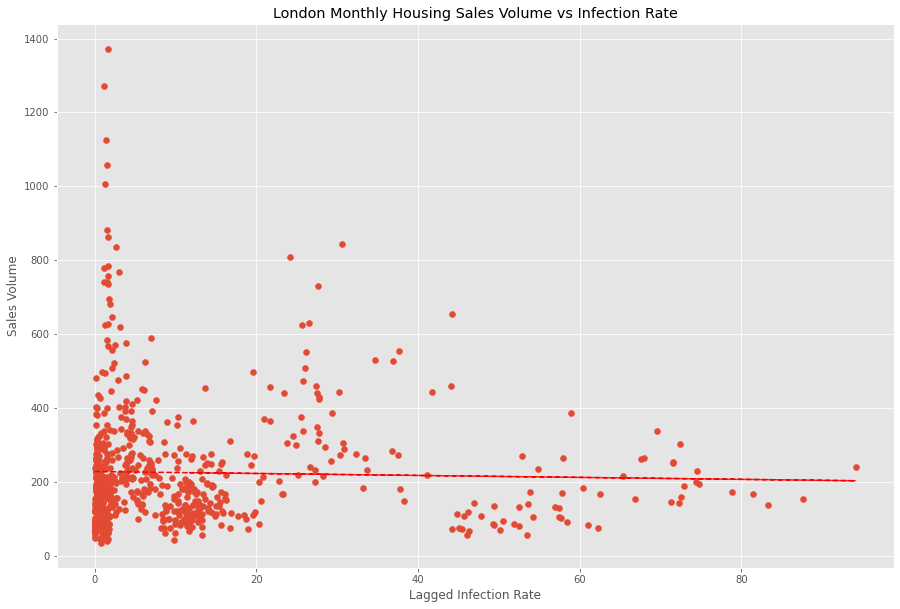


*Figure 13: London Monthly Index Housing Price vs Infection Rate*

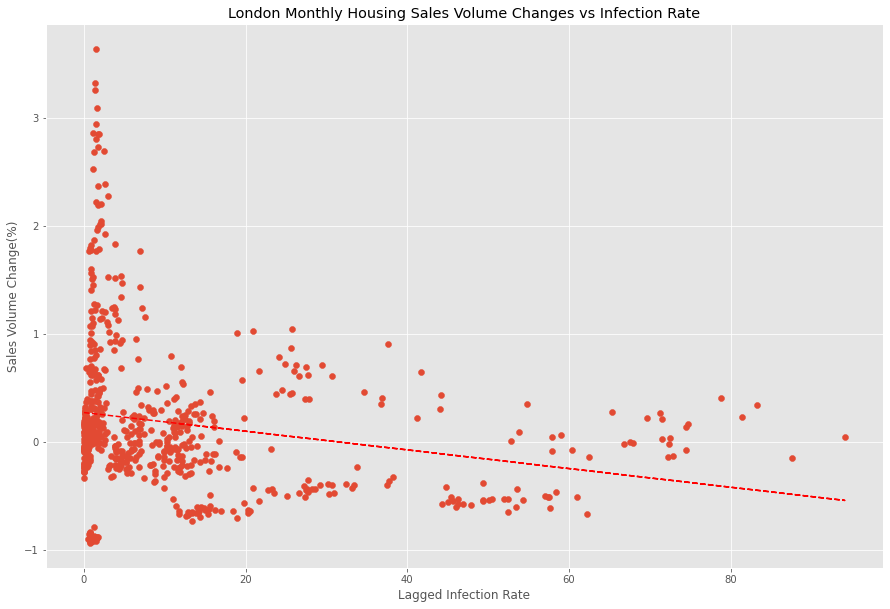


*Figure 14: London Monthly Index Housing Price Changes vs Infection Rate*

Move on to the scatter plots for sales volume, as figure 15 shows, the scatter plot of London monthly housing sales volume vs infection rate is somewhat normally distributed, as most coordinates are clustered at bottom left, the trend line is relatively flat, but it still shows slight negative gradient indicating the negative relationship between sales volume and infection rate. Since the relationship shown in figure 15 is not significant, we move on to the scatter plot for sales volume changes vs infection rate, as figure 16 shows a more normally distributed plot, the trend line between sales volume changes shows a more significant negative curve than the one in figure 15, which again proves the negative relationship between sales volume and infection rate.



*Figure 15: London Monthly Housing Sales Volume vs Infection Rate*

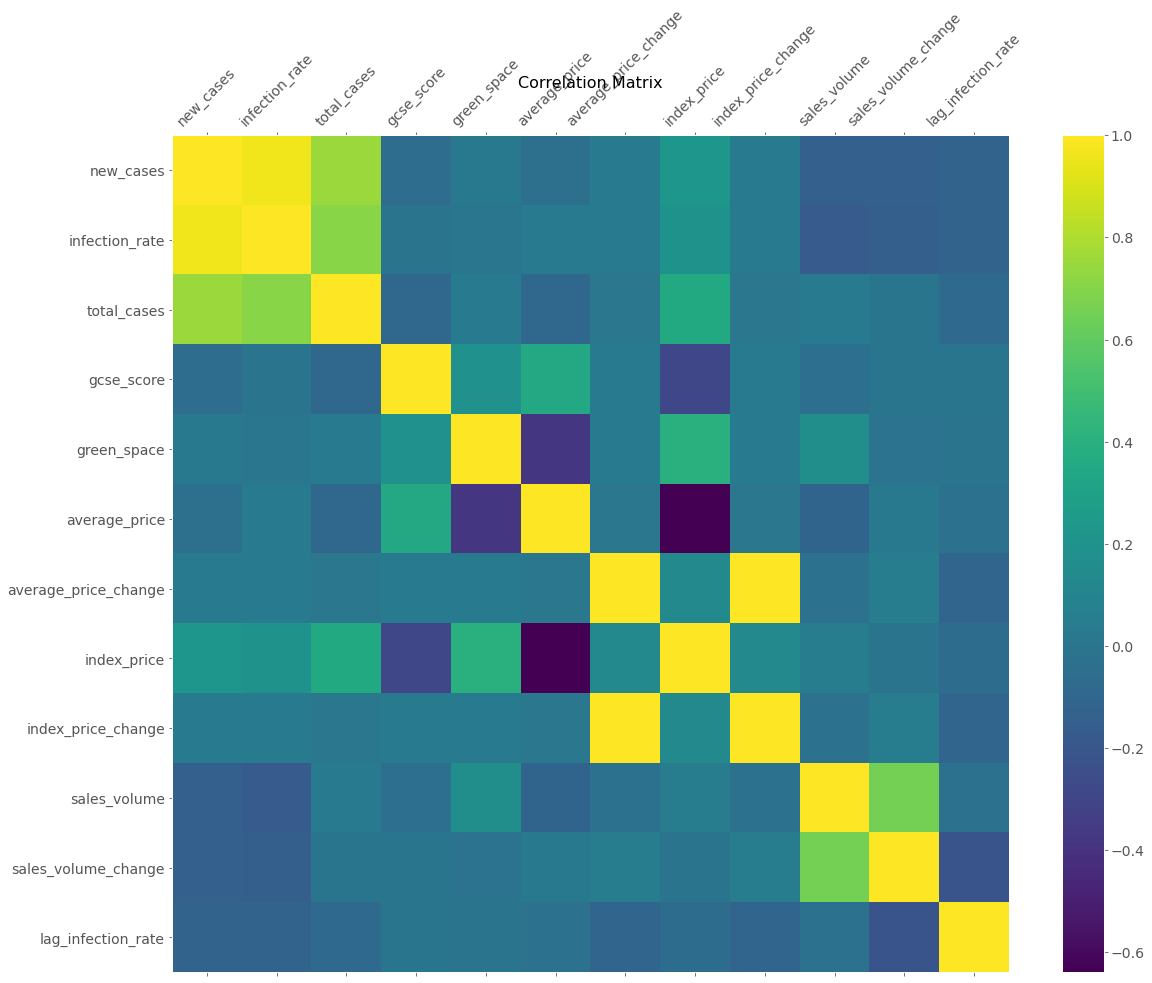


*Figure 16: London Monthly Housing Sales Volume Changes vs Infection Rate*

After visualising the scatter plots for infection rate and dependent variables, it is interesting that all the scatter plots indicated a negative gradient trend line, the negative relationship between infection rate and dependent variable is an indication of negative impact of high infection level on London housing market, which the borough with high level of infection rate likely to be less active in the housing market after 3 month time and potentially lead to a price reduction. Since the supply in the housing market are relatively stable, the major influence of high infection level is likely to be on the demand side of the housing market, as the high level of infection is likely to lead to high level of regulation in the area as well as causing panic among residences in that area, it may reduces the incentive to purchase within the area and results a temporarily reduction in demand.

**5.2 Regression Results**

Before the result of the regression, we firstly use correlation matrix to visualise the linear correlation between the variables in the dataset. As the correlation matrix shown below, it is obvious that there are some high level of correlations between COVID-19 related variables, but the correlation index between dependent variables and independent variables are relatively low, as all the correlation bars between dependent variables and independent variables are with in the rage of 0 to 0.4. However, the level of correlation between variables can not fully represent the relationship between variables, further analysis need to be made to provide a more precise conclusion.



*Figure 17: Correlation Matrix*

**5.2.1 Regression for Average Housing Price**

Since we selected three important data of London housing market as the dependent variables for this investigation, we will run three sets of regression with different dependent variables respectively, and the first set of regression is for average housing price.

The first regression model is the OLS linear regression model, and for the accuracy of the model, we also added a year dummy and area dummy on the left hand side of the formula. As the result shows, the R² value represents how well the model fit the data, and the R² value of 0.577 means that the model only explains 57.7% of the variability of independent variables, which is a relatively high value as the model explains more than half of independent variables. Move on to the p-values for independent variables, the p-value for both gcse score and green space percentage are 0, which means they are statistically significant to the dependent variable, and for the only COVID-19 in the model, the p=value for lagged infection rate is 0.787, the large p-value for lagged infection rate means that it is not statistically significant, and the changes in the predictor are not associated with changes in the response.

|  |  |  |  |
| --- | --- | --- | --- |
| **Dep. Variable:** | average\_price | **R-squared:** | 0.577 |
| **Model:** | OLS | **Adj. R-squared:** | 0.564 |
| **Method:** | Least Squares | **F-statistic:** | 43.37 |
| **Date:** | Thu, 11 Aug 2022 | **Prob (F-statistic):** | 4.79e-28 |
| **Time:** | 14:49:38 | **Log-Likelihood:** | -2024.7 |
| **No. Observations:** | 165 | **AIC:** | 4061. |
| **Df Residuals:** | 159 | **BIC:** | 4080. |
| **Df Model:** | 5 |  |  |
| **Covariance Type:** | nonrobust |  |  |

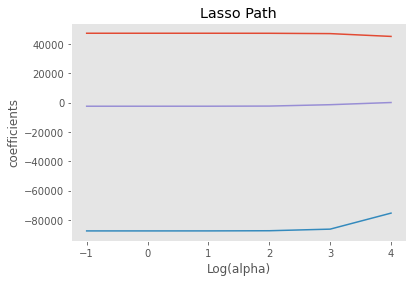
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **coef** | **std err** | **t** | **P>|t|** | **[0.025** | **0.975]** |
| **Intercept** | -5.221e+05 | 6.6e+04 | -7.908 | 0.000 | -6.52e+05 | -3.92e+05 |
| **C(year\_index)[T.2]** | 2.061e+04 | 8393.263 | 2.456 | 0.015 | 4033.619 | 3.72e+04 |
| **C(year\_index)[T.3]** | 3.767e+04 | 2.08e+04 | 1.811 | 0.072 | -3415.926 | 7.88e+04 |
| **lag\_infection\_rate** | -69.0614 | 254.811 | -0.271 | 0.787 | -572.312 | 434.190 |
| **gcse\_score** | 1.88e+04 | 1303.752 | 14.419 | 0.000 | 1.62e+04 | 2.14e+04 |
| **green\_space** | -1558.3397 | 416.180 | -3.744 | 0.000 | -2380.293 | -736.386 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Omnibus:** | 45.440 | **Durbin-Watson:** | 0.074 |
| **Prob(Omnibus):** | 0.000 | **Jarque-Bera (JB):** | 17.907 |
| **Skew:** | -0.607 | **Prob(JB):** | 0.000129 |
| **Kurtosis:** | 1.936 | **Cond. No.** | 1.06e+03 |

*Table 1: OLS Regression for Average Housing Price*

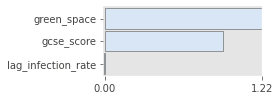
Move on to the VIF result for average housing price, we splits the dataset into training and testing data in the proportion of 75:25, this aims to avoid randomness of the dataset, however, it also removes the year dummy and area dummy for the following models. And as we received a maximum VIF of 1.035, the VIF value below 10 indicates no sign of significance among variables included in this model.

In the Lasso regression model for average housing price, we received a score value of 0.342 for the training data and 0.267 for the testing data, which means that the Lasso Regression attained an accuracy of 34.2% and 26.7% with the given Dataset. And as the Lasso Path figure shows below, there is clearly no signs of high level of correlation among variables, which could be an indication that our dataset does not fit Lasso model very well.



*Figure 18: Lasso Path for Average Housing Price*

For the last model with average housing price being the dependent variable, the regression tree model, for the R² value, we received 0.999 for the training data, and 0.968 for the testing data, this could be a sign of the dataset is fitting this model perfectly, but it also could be a sign of overfitting. And for the RMSE value, we received 2253.3 for the training data and 30661.8 for the testing data, the high value of RMSE for both training and testing data indicates that this model is not able to fit the dataset very well, and it also proves the sign of overfitting in R² values. And for importance plot of this regression tree model, as figure 19 shows, the importance for infection rate is at the bottom of the list, which also indicates the lack of statistical significance between infection rate and average price for London housing market.



*Figure 19: Importance Plot for Average Housing Price*

**5.2.2 Regression for Index Housing Price**

For the second set of regression, we selected the index price for London housing market as the dependent variable.