Detecting Structural Breaks and Forecasting Trends in London Borough House Prices

Ashan Shepherd, 23044640, ass0770@my.londonmet.ac.uk

***Abstract*— The housing market in the UK is said to reflect the country's current economic and social situations. Property owners, developers, investors, and buyers ask questions such as when and where to buy or invest. The motivation behind this study is to research the evolution of house prices across boroughs as it offers insights into economic and social shifts over the last decades, which can help policymakers, investors, and economists understand broader house market trends. A Time Series machine learning model will be executed to forecast future house prices. The Linear Regression model will also be executed to test against the explanatory variables with house prices. The Time Series Model forecasts that by 2029, the average house price in London will be £659,597. Salary, GDP, and population variables positively correlate with the target variable. The research gives an overall insight into the housing market in London and suggests certain variables which correlate with house prices, allowing stakeholders to make decisions with more confidence through machine learning.**

Keywords*— *Time Series, Linear Regression, House Prices, Structural Breaks

I. Introduction

The housing market in the UK is said to reflect the country's current economic and social situations [1]. Property owners, developers, investors, and buyers ask questions such as when and where to buy or invest [2]. It is a big investment when buying in the housing market, and stakeholders need to be confident that their purchases are worth the investment. Questions have been asked about the unaffordability of housing, in which there are many variables as to why price changes occur. These reasons are said to be diversified and complex [3].

One reason for a change in house prices can be due to the structural features of the property itself, with the number of bedrooms and bathrooms affecting their value and other amenities like accessibility to work and public transport [4]. Fully detached properties will likely be more expensive when compared with semi-detached houses or flats around the same area. The location also influences property value. When relating specifically to London, over the past 40 years, the 32 boroughs which make up Greater London have seen a great fluctuation in house prices [5]. London is a popular City with a high economic status, causing a high supply and demand for houses. The high prices within Greater London reflect the strong demand and limited supply, which has led to increased interest in previously undervalued neighbourhoods, inciting gentrification as investors and new residents seek more affordable options within Greater London [6]. Fluctuations in prices can also be partly due to structural breaks, which are abrupt changes in the data pattern by external factors like economic crises, policy changes, political events or climate risks [7]. These occur at a specific time, affecting the housing market for a short or long period.

All these factors make house pricing a complex and dynamic economic entity, which can be analysed in many ways. There is considerable public debate and concern over the affordability of house prices in London, and the risk of instability in the housing market makes it hard to judge when and where to purchase [8]. Recent developments in machine learning offer new ways of modelling complex socio-spatial processes, permitting accurate predictions about how and where they might appear in the future [9].

The motivation behind this study is to research the evolution of house prices across boroughs as it offers insights into economic and social shifts over the last decades, which can help policymakers, investors, and economists understand broader house market trends. The UK housing market has been influenced by various macroeconomic events such as COVID-19, recessions and policy changes [10]. Detecting these events will allow researchers to pinpoint how these impacted different regions within London. Also, the study aims to have significance in revealing key points where house market conditions notably changed to uncover differences in price evolution across London boroughs, inferencing housing affordability, gentrification, and socioeconomic dynamics. Finally, accurate forecasting of house prices will be important for stakeholders, helping them make informed decisions before investing. Based on previous data, predictions can be made about house prices across the different London boroughs, allowing greater confidence and understanding when investing.

II. Literature Review

The goal of the literature review is to examine research previously done on London housing price trends, structural breaks and the use of machine learning techniques on this topic.

The UK Housing market has seen a near-continuous price increase since the mid-1990s, which has led to decreasing homeownership rates and an increasing number of individuals turning into long-term renters [11]. This is especially present in the Southeast of England and the London urban area, which have recorded the highest house price increase [12]. The volatility of the housing market is one of the highest in the world, indicating its tendency to change quickly in an unpredictable way [13]. However, government data shows that house prices were almost unchanged between 1953 and 1995, contrary to recent times and that by 2002, house prices had risen by almost 30% due to inflation [14]. Based on these reasons, it is said that housing affordability in London is a key concern.

In econometrics and statistics, the definition of a structural break is an unexpected change over time, often caused by an external event [15]. Examples of what can cause structural breaks are pandemics, protests or economic changes such as tax rates [16]. The impact of policy and economic events on house prices can be analysed from past experiences. It has been argued that UK house prices increased between 1996 and 2007 due to increased interest rates [17]. Not all research agrees with this view; some argue there is little relationship between house prices and interest rates [18]. COVID-19 has caused much economic uncertainty, and most studies have focused on GDP and industrial production [19]. Few have looked at the housing market and the pandemic's effect on it. One study concluded that the average UK house price slightly declined during the lockdown until the beginning of May 2020. After that, it increased back at the national level [20].

Many studies have looked at whether identifying structural breaks creates a more accurate forecast when predicting the housing market. The goal of carrying out predictive machine learning models may be to understand the effects of volatility and structural breaks on the prediction rather than just having data predicted with little understanding [21]. Different learning models, such as neural network models, may be more suitable for this.

Other methods have been used to analyse house pricing, such as Linear Regression, Decision Tree Regression, K-means regression and Random Forest Regression models [22]. One study used a support vector model (SVM) to analyse housing. The study concluded that the SVM performs well at making accurate predictions of property prices. However, the random forest and the gradient-boosting machine models performed better [23]. In another study by Satish et al., linear regression, LASSO regression algorithm and gradient boosting algorithms models were used to predict house prices. Their results show that the LASSO regression algorithm outperforms other algorithms in terms of accuracy [24].

Machine learning algorithms generate many benefits in research. They provide more flexible and accurate estimation procedures which can analyse a substantial amount of data. These large datasets sometimes contain complex relationships between variables that traditional estimation techniques may not be able to identify [25]. Machine learning methodologies are said to provide many advantages in estimating property prices. What would have taken a long time in the past can now be completed quickly, with the use of this technique today being less costly or time-consuming. Time-series machine learning models aim to make good predictions and forecasts [26]. Accurate forecasting can provide useful analysis and information based on historical data on house prices in London Boroughs.

III. METHODOLOGY

A. *Discover*

During the Discover phase, activities include framing the business problem as an analytical problem that can be addressed [27]. The issue must be investigated and understood with an initial hypothesis to be tested against the data. Therefore, for this journal, domain knowledge was developed through past research and information on the housing market and machine learning. The main topic objectives were made, and the correct data was gathered to answer the research question. The availability of the required data must be considered and assessed if it is sufficient to support the project's goals [28]. Framing is the process of stating the analytics problem to be solved [29]. For this project, the statement is to analyse the evolution of average house prices in London boroughs and forecast future house prices using time series machine learning techniques. Correlations will also be analysed between multiple explanatory variables and the target variable.

B. *Data Preparation*

The second phase of the Data Analytics Lifecycle involves data preparation, including the steps to explore, preprocess, and condition data before modelling and analysis. The housing dataset was extracted, transformed and loaded. The data for this project is from the public data site London Datastore [30]. This public data website uploads real statistics on different sectors in the UK, such as finance, marketing, and economics, suggesting the information is credible and reliable, which will ensure the validity of this research. During the preparation phase, it was ensured that the data had no null values or duplicated rows. Data cleaning will prevent false conclusions from being made as the data's integrity is assured. A lot of time can be spent during this phase as the data must be reliable and appropriate for analysing and answering the business questions. It must be sure that the data being used is adequate in allowing the models to be executed properly. Data was transformed by combining tables into one. The explanatory variables were found in different tables, which were combined with the housing market table based on date so that machine learning could be carried out and compared on all the required attributes.

C. *Plan Model*

The third stage of the methodology involves identifying the model or models to apply to the data. This depends on the goal of the research and the type of data available in the dataset. Based on the variables in the housing dataset after cleaning and transformation, time series machine learning can be executed to predict future house prices. Time series machine learning can be executed because the dataset has a date variable, the months from 1995 to 2024. The Linear Regression model can also be executed as there are multiple explanatory variables to test against house prices.

D. *Build Model*

During the 4th stage, the datasets for training and testing must be developed. With this, an analytical model can be used because of its robustness. Once it is evaluated that the model is powerful enough to solve the research question, the next step can be started after the model is executed. The models selected will give the most accurate answers based on the objectives and allow for clear communications to those who benefit from the research.

E. *Communicate*

Once the model has been carried out, the outcomes must be compared with the objective of the project. It should be determined if it succeeded or failed the objective based on the results. Key findings are documented during this phase, along with the major insights found from the analysis. The machine-learning techniques will produce graphs and charts which will communicate the results of the analysis onto a dashboard. This allows for clear communication of the findings for all people produced on PowerBI.

F. *Measure effectiveness/Apply Live*

The last phase is the deployment of the model, ensuring its operational success. Ethical considerations must also be discussed during this phase. The data contained no private or personal data and was available to the public, therefore being acceptable to use.

Overall, based on the data analytics cycle used in this project, a good plan was implemented. This allowed for the creation of an objective and the finding of the necessary data to clean and transform. With this, the appropriate machine-learning techniques can be chosen and executed to find key results and represent them in graphs and figures through PowerBI.

IV. Results and Discussion

Time series supervised machine learning was used with PowerBI. More specifically, the time series model carried out on this software is the Exponential Triple Smoothing model (ETS). Other types, such as the Autoregressive or Neural Network models, are unavailable in the software unless Python or R code is incorporated externally.

A time series model is an ordered sequence of equally spaced values over time. It uses the equation Y= a+bX to provide plots for each row of data. The three main components of the model are trend, seasonality and error. Seasonality describes the fixed fluctuations in the observation over time, which can be adjusted on PowerBI depending on the data. The trend component describes how the model deals with upward and downward movement over time. Error addresses how the model deals with residual [31]. The ETS model uses multiplicative seasonality and an additive trend, represented mathematically as:

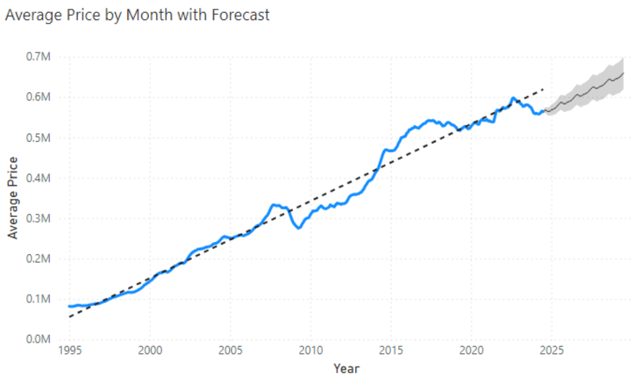
$$y_t=(l_{t-1}+b_{t-1})\cdot s_{t-p}+\epsilon_t$$

which exhibits both non-linear growth and seasonal variability.

The key variable for forecasting is the date variable, which has monthly data between 1995 and 2024. The model can accurately forecast housing prices in the next five years with consistent historical data available. The shaded area shows these forecasts with a 95% confidence interval [32]. The data model integrates with the objectives set as forecasts will be produced for the London housing market, and trends can be identified.

The immediate trends and patterns apparent from the analysis were that house prices increased consistently across all London Boroughs from 1995 to 2024. Figure 1 shows a positive correlation between London's average salary and house prices. This agrees with previous research which carried out similar analysis. The figure also shows a forecast for house prices for the next five years. This forecasts that by 2029, the average house price in London will be £659,597. Compared to the most present average price in the data, the time series models forecast a £94,175 rise in housing prices over the next five years.

Figure 1: Average London House Price from 1995-2024 with Forecast for Five Years

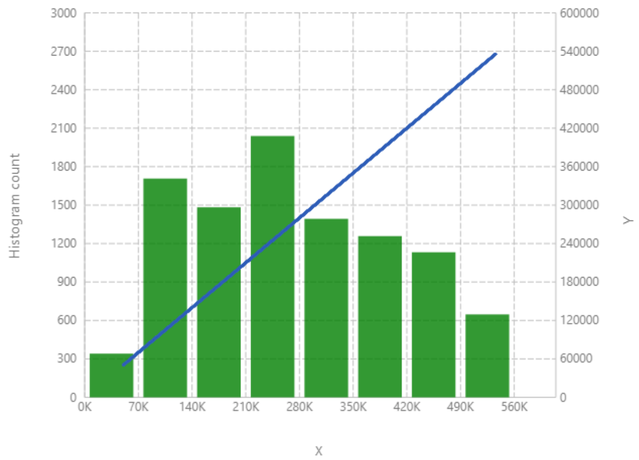


When comparing all explanatory variables with the target variable, some strongly correlate with London house prices. Looking at each linear regression in Figure 2, the inflation (CPIH) and property sales variables have little to no correlation with house prices. Salary, GDP, and population variables positively correlate with the target variable. This does not mean there is a causation, but that the explanatory variables have increased in value along with the target variable between the time present in the dataset.

Figure 2: Linear Regressions Model on all Explanatory Variables Against the Target Variable House Price

Outliers were assessed as part of the steps taken to validate the accuracy and reliability of the results obtained through the models. Boxplots and histograms were produced to visualise the accuracy of the data. The data is distributed but slightly skewed for the average house price variable (Figure 3). The Boxplots identified some outliers in the data. It was decided to keep these outliers in the data for machine learning as they may identify structural breaks in the data and help with forecasting.

Figure 3: Histogram of Average House Price Variable in Dataset



The most expensive London Borough found in the analysis was Camden, which is Forecasted to reach an average price of £1,007,159 by 2029, a 17.34% increase from the latest price in the dataset (July 2024). In contrast, Barking & Dagenham had the lowest average house prices on the market and is forecasted to reach an average price of £392,605 in 2029, a 15.25% increase from July 2024.

Figure 4: Average House Price from 1995-2024 with Forecast for Five Years in Camden

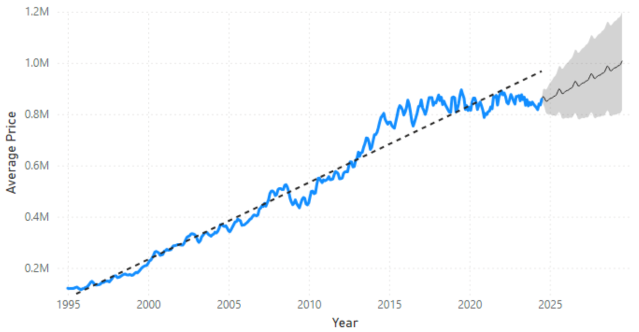
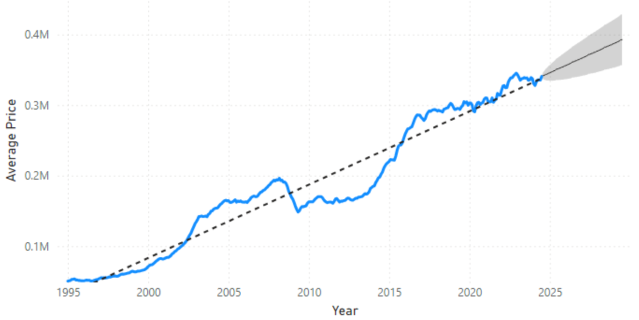
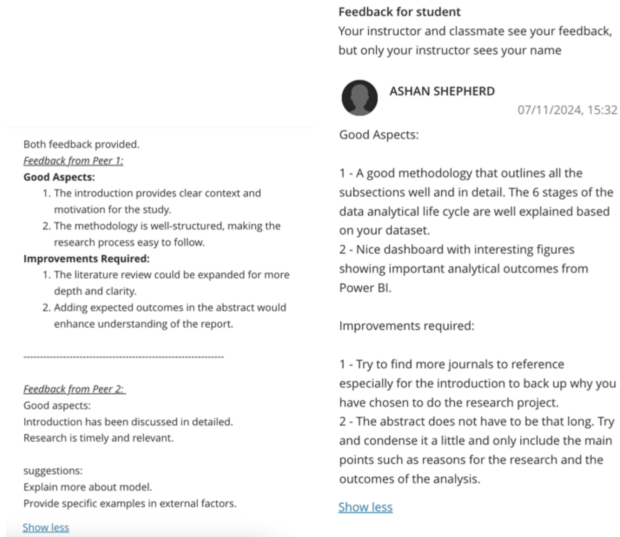


Figure 5: Average House Price from 1995-2024 with Forecast for Five Years in Barking & Dagenham



All 33 London Boroughs have a similar percentage increase over time, suggesting that London house prices in its entirety are moving at a similar rate. When looking at the time series graph in Figure 1, a steep dip can be seen between the years of 2008 and 2009. This backs past research, which states that there was a housing market crash between those years [33]. This suggests that the model does a good job of showing these structural breaks within the data and, therefore, can make reliable forecasts. Research also says that there was a sharp incline in house prices during the COVID-19 pandemic, which accords with the time series model as an increased incline can be seen from the end of 2019 till October 2022 [34].

V. Peer Review



Both feedback provided.

Feedback from Peer 1:

Good Aspects:

1. The introduction provides clear context and motivation for the study.
2. The methodology is well-structured, making the research process easy to follow.

Improvements Required:

1. The literature review could be expanded for more depth and clarity.
2. Adding expected outcomes in the abstract would enhance understanding of the report.

Feedback from Peer 2:

Good aspects:

Introduction has been discussed in detailed.

Research is timely and relevant.

suggestions:

Explain more about model.

Provide specific examples in external factors.

Show less

Feedback for student

Your instructor and classmate see your feedback, but only your instructor sees your name

ASHAN SHEPHERD

07/11/2024, 15:32

Good Aspects:

1 - A good methodology that outlines all the subsections well and in detail. The 6 stages of the data analytical life cycle are well explained based on your dataset.

2 - Nice dashboard with interesting figures showing important analytical outcomes from Power BI.

Improvements required:

1 - Try to find more journals to reference especially for the introduction to back up why you have chosen to do the research project.

2 - The abstract does not have to be that long. Try and condense it a little and only include the main points such as reasons for the research and the outcomes of the analysis.

Show less

Upon reflection on the peer review, I gained important insights about the progress of my work, which I could then improve to ensure the quality of my project was maximised. More focus on the explanation of the machine models used, and more previous journals included were part of the feedback, which I managed to incorporate to improve the quality of the project. The peer review also contributed to the continuing of my professional development as providing feedback to fellow students may be like what is required in other work settings. Being able to read other projects also provided me an insight into the other types of research being carried out, broadening my knowledge of the different machine learning models used in data visualisation.

VI. Conclusion and Recommendation

Overall, the key insights derived from the time series model were that the housing market in London is forecasted to rise to an average price of £659,597 by 2029. Based on the Linear Regression Model, the three main explanatory variables with the highest correlation were salary, GDP, and population. Further research is recommended to find the level of causation these attributes have on the housing market through deeper machine learning. The initial hypothesis was that London house prices would see a continuous increase in prices, which aligns with the results of the project. Specific projections and forecasts were made with Time Series. A more complex Time Series Model could be implemented with this data, such as a Neural Network, to provide a forecast with greater detail. Also, different explanatory variables could be measured against the average London house price, such as geographical factors like public transport accessibility and employment hubs or social and lifestyle factors like crime rate and school quality. Analysing if the average salary in London is sufficient to get into the property market would be beneficial research as well.

REFERENCES

[1] A. Ng, "Machine Learning for a London Housing Price Prediction Mobile Application," Imperial College London, London, 2015.

[2] L. Stephen, P. Brooks and R. Chris, "Take a Look Around: Using Street View and Satellite Images to Estimate House Prices," ACM Journals, vol. 10, no. 5, pp. 1-19, 2019.

[3] G. Yiyang, "What are the most important factors that influence the changes in London Real Estate Prices? How to quantify them?," The Bartlett Centre for Advanced Spatial Analysis, 2018.

[4] A. A. AWONAIKE, "A Machine Learning Framework for House Price Estimation.," 1st International Conference on Intelligent Systems Design and Applications, 2022.

[5] A. Matthew, "Which London Borough Will See the Greatest Increase in House Prices Over the Next Five Years?," UNIVERSITY OF EXETER, Exeter, 2021.

[6][8] J. Marsden, "House prices in London – an economic analysis of London's housing market," GLA Economics, London, 2015

[7][9] R. Jonathan, D. S. Jordan and H. Phil, "Understanding urban gentrification through machine learning," Sage Journals, vol. 56, no. 5, pp. 922-942, 2019.

[10] P. Vasilios, G. Rangan, K. Constantinos and W. Mark E, "Time-varying role of macroeconomic shocks on house prices in the US and UK: evidence from over 150 years of data," Empirical Economics, vol. 58, p. 2249–2285, 2020.

[11] V. D. Rossall, "Solving the UK Housing Crisis," The Bow Group, London, London, 2015.

[12][13] C. Hilber, "UK Housing and Planning Policies: the evidence from economic research," Centre for Economic Performance, London, 2015.

[14] D. Baker, "The housing bubble and the financial crisis," real-world economics review, no. 46, pp. 73-81, 2008.

[15] J. Antoch, J. Hanousek, L. Horváth, M. Hušková and S. Wang, "Structural breaks in panel data: large number of panels and short length time series," Taylor & Francis , London, 2019.

[16] J. A. Ndako, T. T. Kumeka, F. F. Adedoyin and S. Asongu, "Structural breaks in global stock markets: Are they caused by pandemics, protests or other factors?," AGDI Working Paper, Yaoundé, 2022.

[17] J. E. Levin and G. B. J. Pryce, "What Determines the Price Elasticity of House Supply? Real Interest Rate Effects and Cyclical Asymmetries," Housing Studies, vol. 24, no. 6, p. 713–736, 2009

[18] C.-B. Tse, T. Rodgers and J. Niklewski, "The 2007 financial crisis and the UK residential housing market: Did the relationship between interest rates and house prices change?," Economic Modelling, vol. 37, pp. 518-530, 2014.

[19][20] J.-C. Bricongne, B. Meunier and S. Pouget, "Web-scraping housing prices in real-time: The Covid-19 crisis in the UK," *Journal of Housing Economics,* vol. 59, 2023.

[21] K. Choi, W. C. Yu and E. Zivot, "Long memory versus structural breaks in modelling and forecasting realized volatility.," J. Int. Money Finance, vol. 29, no. 5, p. 857–875, 2010.

[22] A. G. Rawool, D. V. Rogye, S. G. Rane and V. A. Bharadi, "House price prediction using machine learning," IRE Journals, vol. 4, no. 11, pp. 29-33, 2021.

[23] W. K. O. Ho, B.-S. Tang and S. W. Wong, "Predicting property prices with machine learning algorithms," Journal of Property Research, vol. 38, no. 1, pp. 48-70, 2021.

[24] G. N. Satish, C. V. Raghavendran, M. D. S. Rao and C. Srinivasulu, "House price prediction using machine learning," International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9, p. 717–722, 2019.
[25][26] W. K. O. Ho, B.-S. Tang and S. W. Wong, "Predicting property prices with machine learning algorithms," Journal of Property Research, vol. 38, no. 1, pp. 48-70, 2021.

[27][28][29] W. John, Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, Indianapolis: John Wiley & Sons, 2015.

[30] "UK house price index," London Datastore News, https://data.london.gov.uk/dataset/uk-house-price-index (accessed Oct. 10, 2024).

[31][32] J. Flores, R. Silva, D. Machado and A. Silva, "PREDICTIONS OF THE NUMBER OF SUSPENDED SURGERIES USING SOFTWARE POWER BI® IN A UNIVERSITY HOSPITAL," R Pesq Cuid Fundam, 2023.

[33][34] J. Robert, "propertyinvestmentsuk," 2024. [Online]. Available:https://www.propertyinvestmentsuk.co.uk/housing-market-crash-2008-09/#:~:text=Summarising%20the%202008%2F09%20housing,by%20the%20middle%20of%202023..[Accessed 24 November 2024].

VI. Appendix

https://londonmet-my.sharepoint.com/:u:/g/personal/ass0770_my_londonmet_ac_uk/EZpKQKLP1XpIoMe4PnosFogBA08MpzqXNY3jjsfyvY9rUg?e=kDvVLB