House Prices in Beijing

GGR377 Final Project, Sizhang Lyu

Introduction

Big data refers to data containing a large variety, volume, and velocity. A high variety indicates there are many types of available data coming into the database; a high volume means a gigantic amount of data coming into the database; and a high velocity refers to the high rate at which new data comes into the database, for example, millions of transactions may be processed within a single data by a credit card company (Favaretto et al., 2020). One of the most prominent methods to work with big data is machine learning, a set of techniques where the computer learns information from the data. In other words, the performance of such models improves as the size of training data increases, and such a unique property is ideal in the context of big data, where a high variety, volume, and velocity data is present (Nichols et al., 2019). With the enormous growth in computational power in the past two decades, methods that were only used to appear on research papers and discussed among computer scientists are now feasible for mass production, and these methodologies are so robust that they can be utilized in almost every field or industry. For example, hedge funds use machine learning algorithms to predict stock prices, banks use machine learning algorithms to detect fraudulent transactions, and video sites such as YouTube use machine learning algorithms to recommend new videos that align with the users' appeals more closely and plenty of other instances. As UofT's world-renowned Jeoffrey Hinton said, "All you need is lots and lots of data and lots of information about what the right answer is, and you will be able to train a big neural net to do what you want (BrainyQuote, n.d.)." Urban big data is an extensive and unprecedented branch of big data that has revolutionized our understanding of city dynamics, particularly in the realm of real estate markets.

The primary purpose of this study is to examine housing prices in Beijing, China's capital, for 800 years, using machine learning. Predicting and inferring housing prices has always been a hot topic among machine learning practitioners and other statisticians. When you look at Kaggle, one of the largest online communities for data science competitions and machine learning enthusiasts in the world, the first-ever competition beginners take is to predict housing prices in the San Francisco Bay Area using regression models (https://www.kaggle.com/competitions/house-prices-advanced-regression-techniques). However, some researchers study housing prices in Northern America but not in many other places worldwide, especially in East Asia. Beijing is currently the world's eighth-largest city in population and the world's twelfth-largest city in GDP (OECD). The city has an exceptionally high population density, and rapid urban development has resulted in a prosperous real estate market during the past ten years. Housing prices are critical in determining economic growth and social stability. The focus will be on Beijing's dynamic urban environment, where economic, social, and spatial factors influence the housing market. Understanding housing price fluctuations is crucial for various stakeholders, including policymakers, urban planners, investors, and residents. Policymakers can leverage this knowledge to create equitable and sustainable urban development plans. Investors and real estate developers can make informed decisions based on the insights gained. Furthermore, this research can shed light on the factors that impact housing affordability, a critical issue affecting many urban dwellers.

Research Design

There are two main objectives in this essay. The first objective is to understand the factors influencing housing prices in various neighborhoods in Beijing, such as location, adjacent resources, or property characteristics. All relevant features will be included in the data that will be used and supported by some macro-scale data, namely from the open data platforms by the government. The initial hypothesis is that Housing prices in Beijing are significantly influenced by factors such as neighborhood amenities, local economic conditions, purchasing power, and accessibility to public transportation. These factors also appear significant in other studies on housing prices in other areas (Jiang & Qiu, 2022), so more emphasis will be placed on these factors accordingly. The second objective is to model the relationship between housing prices and the influential factors discovered in the first objective. Different machine learning and especially probabilistic models will be applied to make inferences and predictions about the housing price. Factors will be examined based on their importance in the final selected models, and visualizations will also be utilized to present some models, such as a spatial visualization of the clustered properties. The initial hypothesis for this objective would be invoking a machine learning model incorporating variables like average income levels, crime rates, proximity to city centers, and public infrastructure that can accurately predict housing prices in different neighborhoods of Beijing and invoking a simpler statistical model at the same time to provide more interpretability.

The main framework for the methodologies used in this essay arises from the theoretical basis and the data source found by web scraping and will further be elaborated on via data analytics, such as descriptive statistics, geospatial visualizations, and regression models. Multiple available algorithms will be tested and will be selected based on the performance of each. The study will explore the relationship between real estate markets and urban development. It will incorporate spatial economics and geography to analyze location-specific factors that affect housing costs. Data will be sourced from both open data, such as official government publications and open data portals, and big data, such as web-scraped data from the listings of real estate trading companies or platforms. These sources will provide detailed information on housing prices, demographics, local amenities, and other relevant variables. Descriptive statistics will be provided and visualized as an initial data exploration, which will aim to understand the fundamental trends and patterns in housing prices across different neighborhoods. Geospatial analysis will form the second part of the data analysis to examine the spatial distribution of housing prices, where geospatial analytical techniques, namely clustering, will be employed and will allow for the visualization and analysis of how prices vary across different geographical locations in Beijing. Regression models, the third and final stage of data analysis, will invoke complex and flexible algorithms like random forest, support vector machine, or gradient-boosting to yield the maximum predictive power, while simpler models like regularized linear regression or statistical tests will be utilized to test the hypotheses and quantify the relationship between housing prices and various influencing factors, but this will depend on the nature and structure of the data. Model selection and evaluation will also be vital to the analysis to see the performance. Criterion such as the accuracy score and goodness-of-fit will be used to assess the models' ability to handle the complexities of urban data, and techniques such as cross-validation and model assumption checks will be used to validate the model's robustness. Finally, and arguably the most importantly, ethical considerations, especially regarding data privacy and the responsible use of data, will be maintained throughout the research process. The study will comply with data protection regulations and ensure the utmost confidentiality of any sensitive information.

Data Acquisition, Cleaning, and Wrangling

Both open data and big data will be used throughout the study. The primary data used for modeling were retrieved from Kaggle, a data science platform known for its diverse data sets ready for machine learning projects. The dataset was scraped by the author from LianJia, one of China's largest real estate brokerage sites, and includes detailed specifications of every house sold through the company. The data ranges from 2013 to 2017 and contains more than 300000 observations and 26 variables. The supplementary data were mainly open data, with sources primarily from governmental websites, including the Beijing Municipal Bureau of Statistics and the Beijing Public Data Open Platform. The platforms have updated census data on property prices each month since July 2013, the same as the earliest records in the primary data source. These sources were carefully chosen for their reliability, comprehensiveness, and relevance to the housing market in Beijing. The combination of government statistics, machine learning-ready datasets, and accurate transaction data provides a well-rounded foundation for analyzing variations in housing prices.

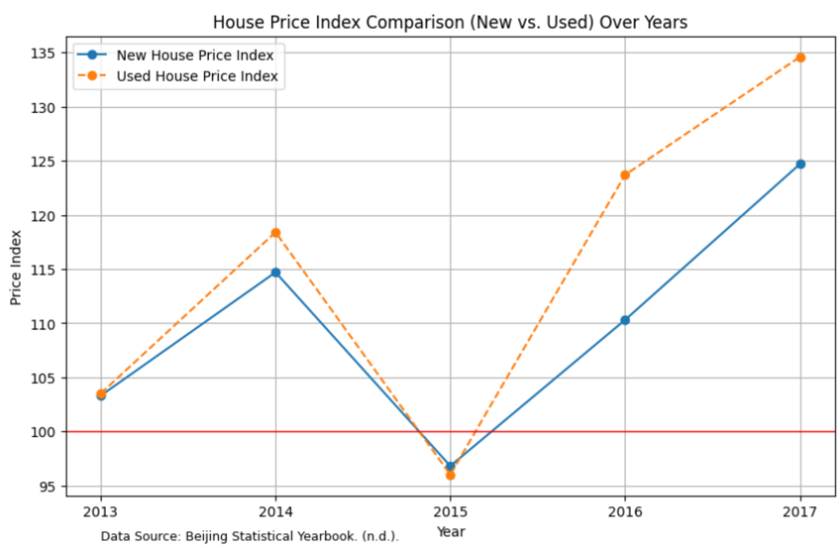
However, even though the primary data source is "machine learning ready," some flaws exist for us to clean. First, we take an initial assessment of the data. The 26 columns in the data set provide comprehensive information about each transaction, such as the coordinates, which are handy for spatial visualizations, total price, price per square meter, number of living rooms and bathrooms, whether the building includes an elevator, etc. A preliminary analysis was conducted to identify data patterns and anomalies. The inspection revealed missing values and potential outliers, typical in large datasets. In particular, a substantial fraction (around 50%) of observations in the "DOM" column, which refers to the active days on the market, is missing. I decided to remove this column as any imputation method would inevitably introduce a significant bias due to the high missing rate, even though this column could be crucial in predicting house prices. The other features only contain a small number of missing observations (roughly 30), so I decided to impute them with the column mean. We must also convert the data types to ensure these variables are interpreted correctly. For example, the number of living rooms is a count instead of a continuous number so that I will convert it into an ordinal category. Also, some columns are recorded as text labels that must be converted into ordinal categories. Similar columns include the number of kitchens and bathrooms, building type, whether there is an adjacent subway station and renovation condition type. Outliers were also inspected in this data set. The main feature suffering from outliers is the price per square meter and total price of the property, which is reasonable, as there exist scarce and extremely luxurious houses that deviate far from the majority, so we will not remove or impute them to make the model interpretation more realistic.

After these cleaning steps, the data can be split into training and testing sets. Standardizing continuous numerical features is another essential preprocessing step in any machine learning project. Several machine learning models, especially for probabilistic models such as linear regression, contain a distributional assumption for the input and the output variables. For example, in the case of linear regression, both the independent and dependent variables are assumed to have a normal distribution with some variance, so ensuring the assumptions are met is the first and foremost step in ensuring accurate results from the model. The typical practice of statistics is subtracting each observation from the sample mean and dividing it by the sample standard deviation. Note that this procedure should only be done using the summary statistics from the training set. Otherwise, the results will likely suffer from overfitting, a common phenomenon of data leakage.

After carefully conducting the cleaning and wrangling steps, the primary data was structured into a format suitable for statistical and geospatial analysis, enabling a comprehensive examination of the spatial heterogeneity of housing prices. Now, it is ready to facilitate descriptive, inferential, and predictive analyses. The government already processed the supplementary data, and multiple visualizations are available. Some will be selected, modified, and presented in the next section to support the argument.

Data Analysis

Let us first look at the overall trend of housing prices in Beijing. Each month, the government publishes the house pricing index compared to the last month among all the houses sold in the city. The house pricing index is defined as the percentage change in the average price of all houses in each category compared to the last month. I have calculated the aggregate trend for each year from 2013 to 2017 and plotted them. As we can see from the figure, except for 2015,

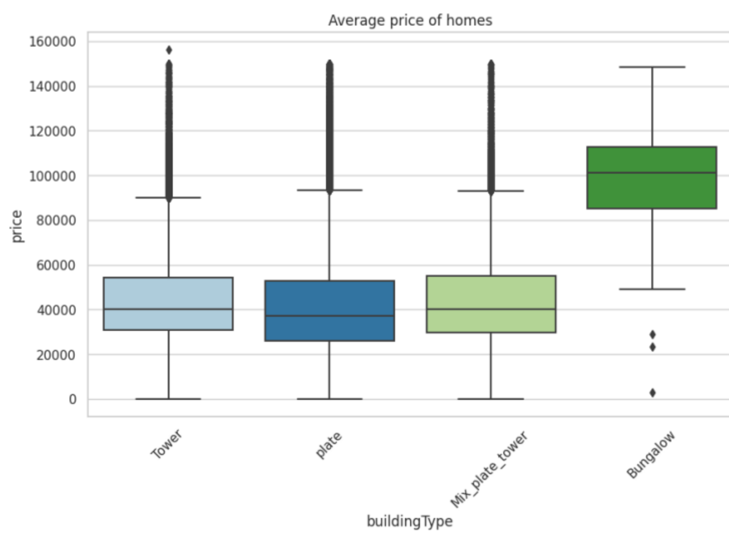


the overall housing price increased dramatically. This can be due to multiple factors. Since 1978, and especially since 2002, the rapid urbanization happening in mainland China has brought millions of people each year from the countryside to the cities. The skilled laborers quickly filled the jobs in the factories, and a cycle of economic boost has since begun. The development in infrastructure has awakened the real estate demand and supply, and vice versa. The rapid development in real estate brought up even more of an urge to build up infrastructures, which raised property prices even higher. This can be reflected in the table below; a substantial improvement in investment and income accompanies the skyrocketing increase in real estate prices.

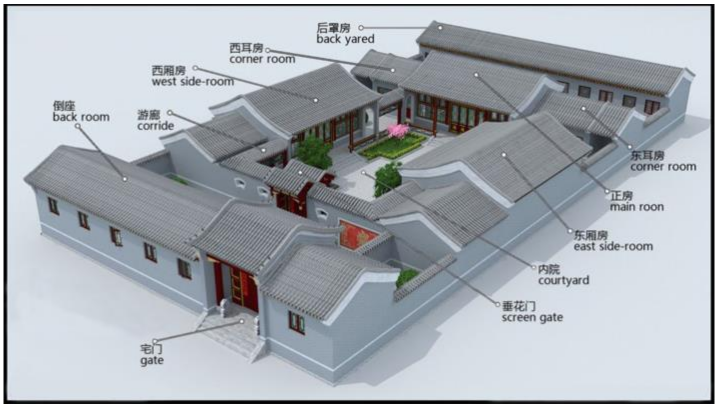
Increase in Real Estate and Income Each Year (2013-2017)

Year	2013	2014	2015	2016	2017
Total Real Estate Investment (% increase)	204.6	12.3	8.1	4.3	7.4
Total Built Area (% Increase)	7.1	24.5	16.0	13.8	25.7
Disposable Income (% Increase)	9.9	9.8	8.9	8.5	9.0

Data Source: Beijing Statistical Yearbook. (n.d.).



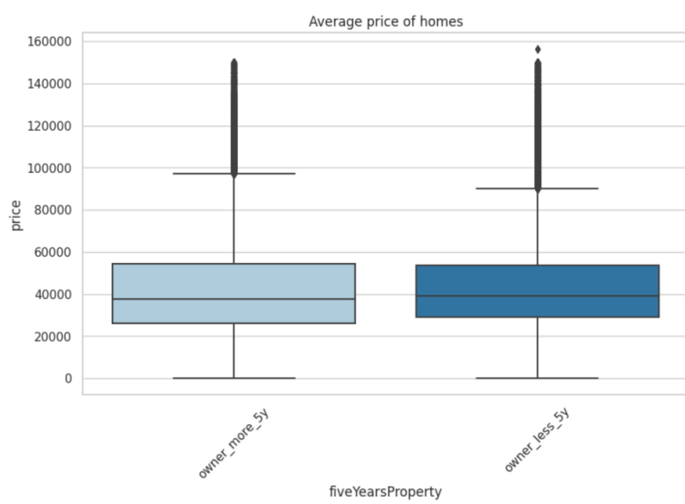
Let us then delve into the primary data to explore its trend. First, we will inspect some of the categorical features to see if there are any differences within each category. The most notable one among all is building type. As we can see from the plot, the rest of the three building types appear to have little difference in numerical and spatial distribution. However, bungalows are almost twice the price and are located entirely in the city's center. I must point out that Beijing bungalows are unique to the city and differ far from those in Europe and the Americas. They are usually located in the old town and consist of multiple connected houses surrounding a central garden, and the image on the bottom left perfectly illustrates their structures. Another

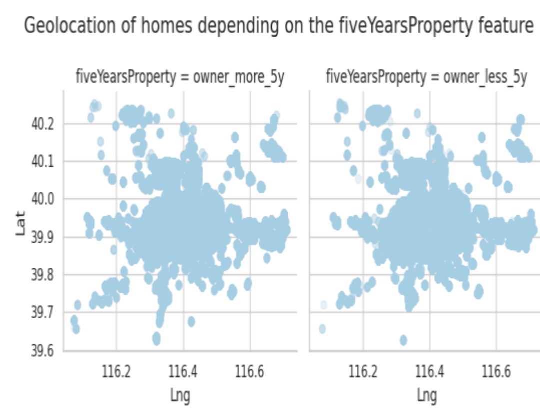


interesting feature is whether the property has been owned for more than 5 years. Usually, one would suspect that when a property turns older, its value naturally depreciates, and normally these old buildings

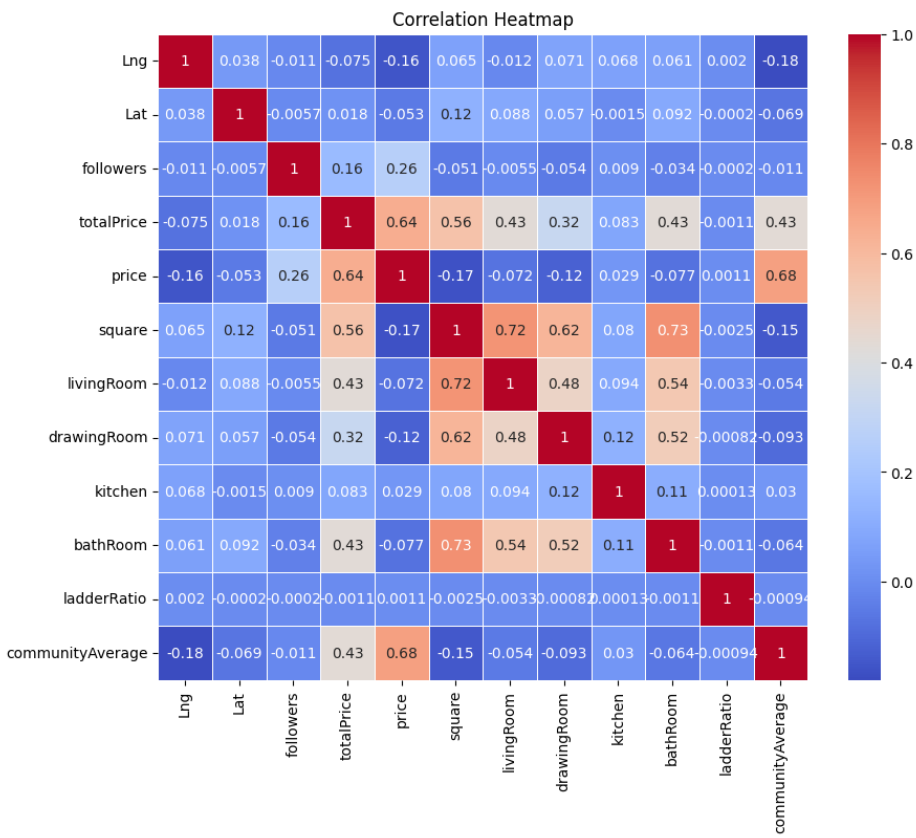
Image Source: https://www.fangruan.com/news/detail/8795

are more likely to be located at the center of the city. However, the figure below shows that neither the spatial nor numerical distribution between properties owned for less or more than five years have any discernable differences. This is supported by the overall trend plot for each year, as newly built and previously owned houses share almost the same trend in price. The reason could be Beijing's public school policy, which states that a pupil should attend the school nearest to their address, so old homes near a prestigious school may still hold their value.

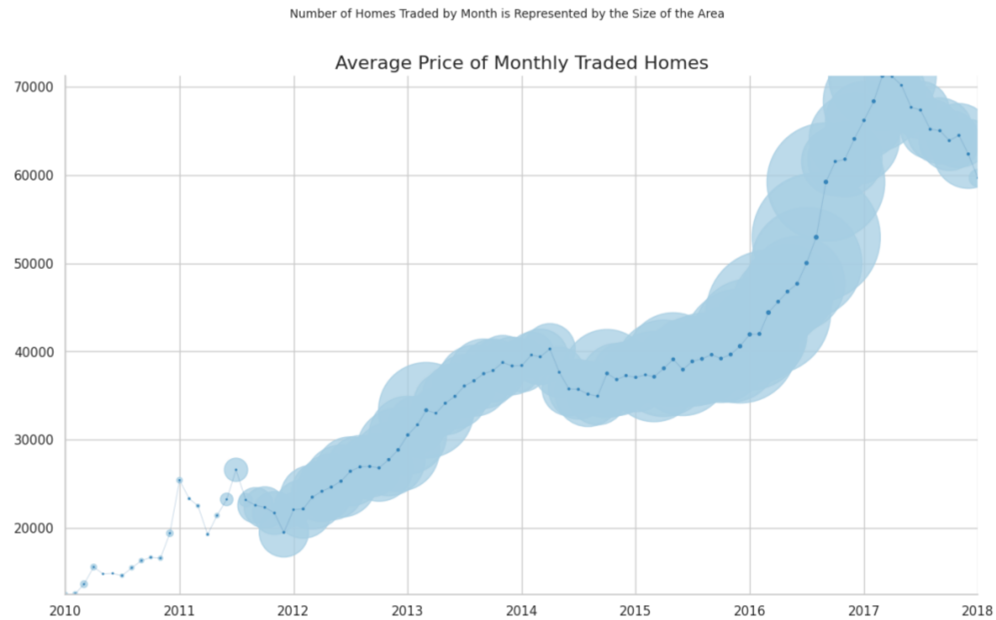




Next, we will look at the correlation between the numerical features, especially with our variable of interest – price. Below is a visualization of the correlation matrix. As we can see, the price per square meter is highly positively correlated with the total price but negatively correlated to the area in square meters of the property. This means that larger houses tend to have even lower prices per square meter, but houses with high total prices often have high prices per square meter. This could imply that houses with the highest total prices tend to be smaller. This coincides with my discussion in the last paragraph that some old and small apartments may still be the most expensive in the city simply because they are close to a prestigious school. Another interesting observation is that both total price and price per square meter are relatively highly correlated with the community average, which could suggest that property prices in Beijing are clustered, which I will refer to in the next section. The number of followers also seems positively correlated with price, reflecting the most fundamental economic principle – prices will rise as demand rises. I also included latitude and longitude as a very high-level summary of potential hot spots in the city, and it appears that the southern part of the city tends to have lower property prices.

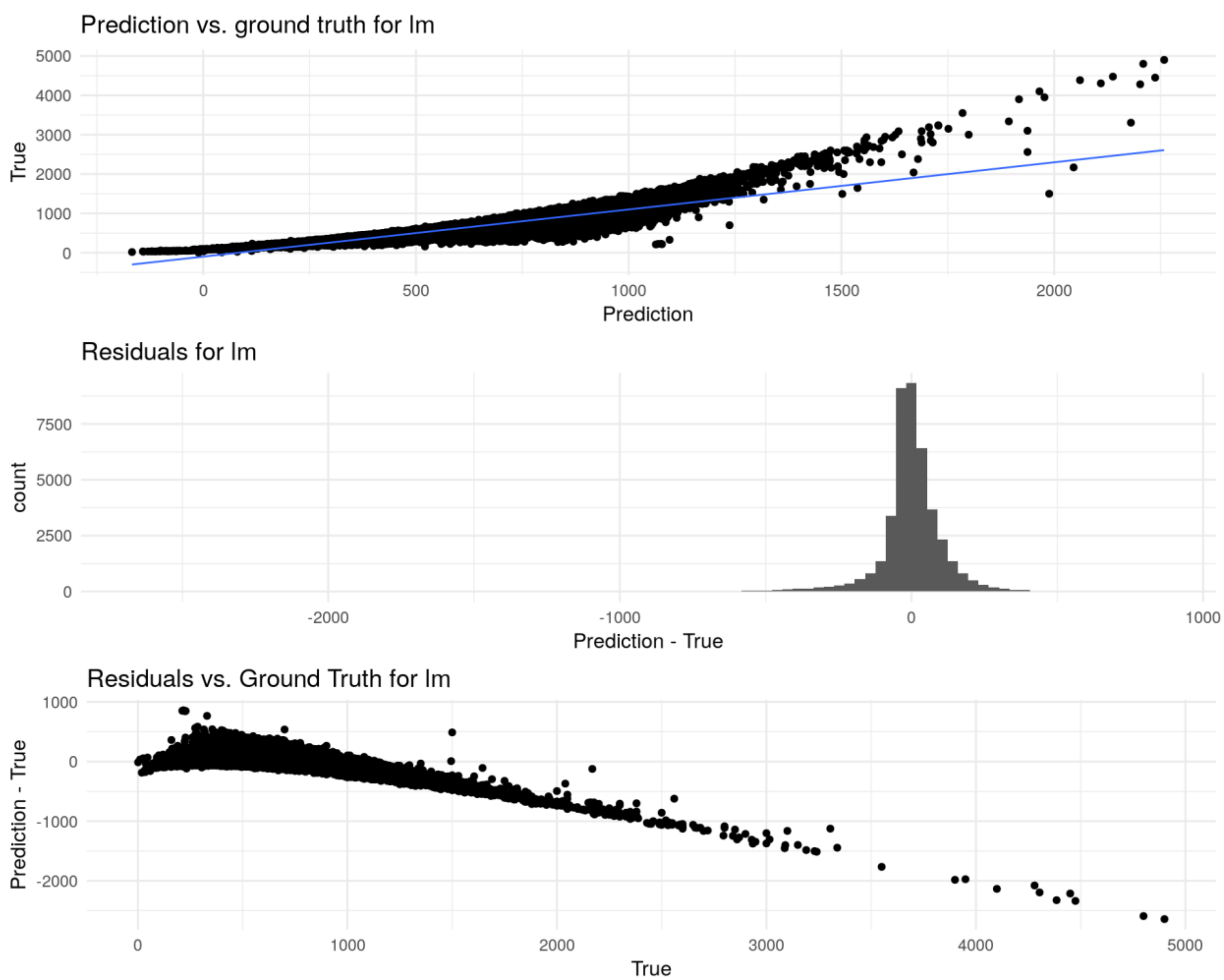


After a thorough understanding of the structure of the data, we can start to build a model, but first, we have to inspect the trend of the outcome variable with respect to time. If we do observe some seasonality over time, then classical machine learning methods would not be appropriate for this time series data (Jiang & Qiu, 2022). From the trend plot below, we can see that, generically, the number of trades and the average monthly price both increase as time progresses, with minor fluctuations. We can also observe that the average price increased until about mid-2017 and then quickly decreased, and at the same time, the number of homes traded also decreased with their price. The data points appear to be the most concentrated between late 2014 and mid-2017, so we will split them on Jan 1st, 2017, and use the observation after the date as the validation data.

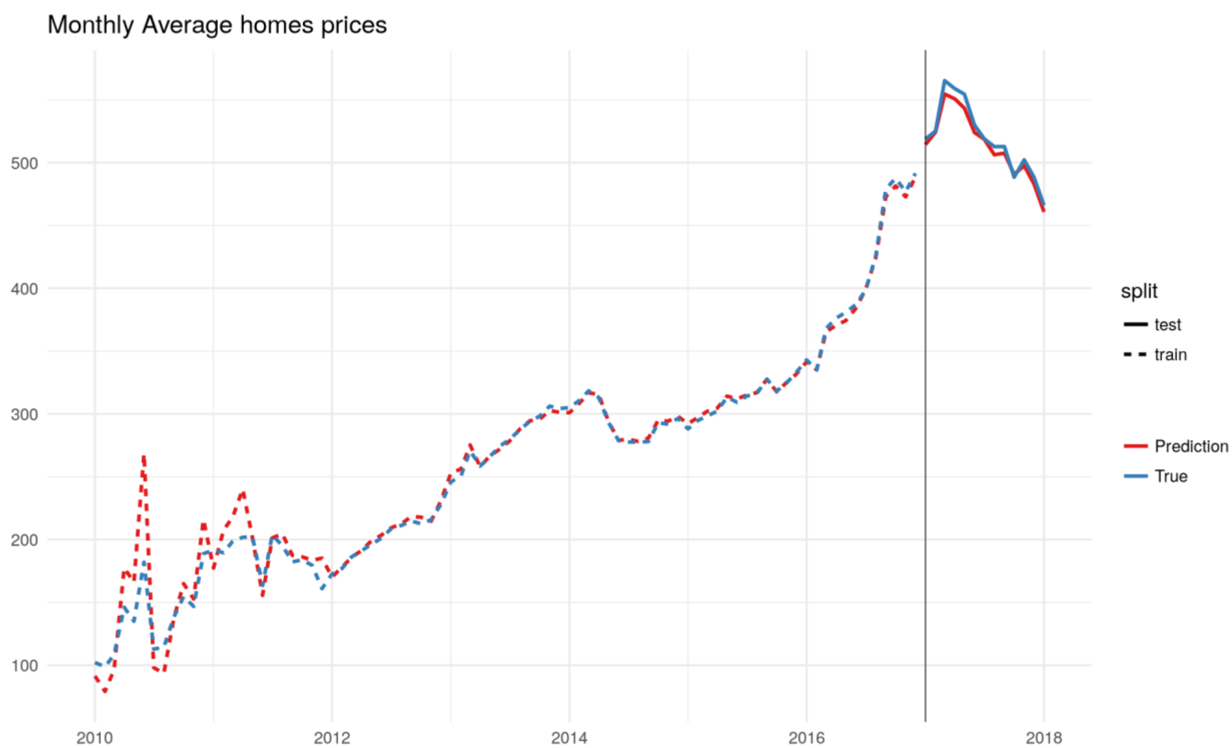


As discussed in the research design section, a linear regression model and a more flexible model will be used for both interpretability and prediction accuracy.

To enhance the accuracy of our regression model in reflecting the time series nature of real estate data, I integrated temporal dynamics. I acknowledged the significance of temporal patterns in property price fluctuations and, thus, preprocessed our dataset, temp_train. This involved extracting crucial temporal features such as year, month, and day from the date column (tradeTime). This preprocessing step transformed the date information into a format conducive to regression analysis, allowing a nuanced understanding of how property prices vary over time. To maintain the accuracy of our time series analysis, I utilized the TimeSeriesSplit function from the scikit-learn library to conduct cross-validation. This technique deviates from conventional cross-validation methods by preserving the sequence of observations, thus preventing any unintentional influence from future information that could skew the model's training phase. Every fold within this split function comprised a continuous section of the data, which respects the dataset's time-based structure. My linear regression model was trained on a set of temporally ordered folds, which allows us to make accurate predictions of property prices while accounting for temporal trends and seasonality. I measured the model's performance using negative mean squared error, a standard metric for regression tasks, which helps us quantify how much the model's predictions differ from the actual values. We also monitored the computational efficiency of the training process by recording the time taken for the training. This allowed us to gain valuable insights into the computational demands of our time series approach, ensuring that our methodology was both statistically sound and practically feasible. The recorded computational time was 12.7 seconds, which is relatively fast due to the model's simplicity. This approach uses the temporal structure within the data, enabling a more sophisticated and temporally aware regression analysis. This methodology is particularly well-suited for datasets like ours, where time plays a critical role in shaping the underlying patterns and trends. The model returned with an adjusted R squared (the proportion of variability in the data explained by the fitted model) of around 0.88 and a mean square error (the average square squared distance between actual and predicted value) of around 15000, which is pretty good. The residuals are plotted as follows.



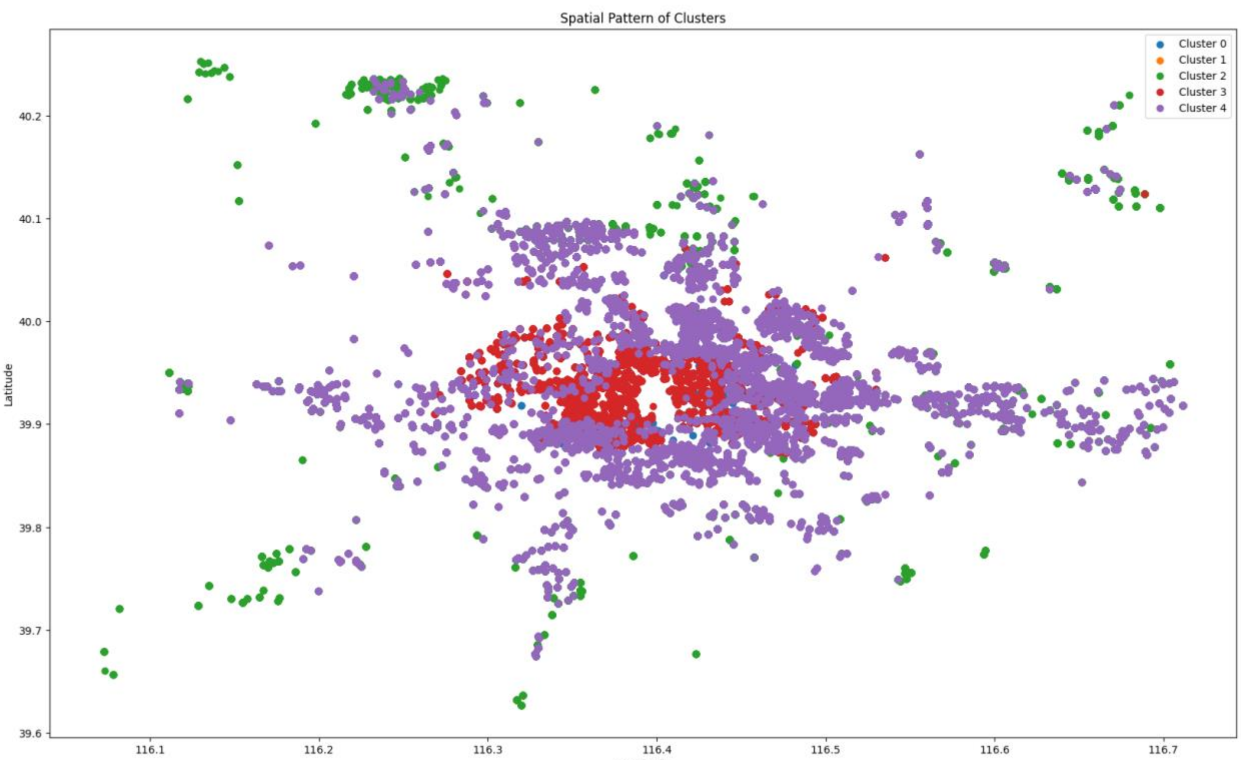
As we can see, the distribution of the residuals is centered around 0 and appears to have a non-skewed or normal-like distribution, but the model does not predict well on the lower end of the price. Therefore, based on this linear regression model, we can repeat and group the prediction for all months to see the accuracy of predicting the monthly mean. The target would be the mean price of the homes, and we will plot the predicted values against the actual values as displayed in the plot below. The vertical line represents Jan 1st, 2017, which is the point where the training and testing sets are separated. As we can see, the predictions are still not doing so



well for years before 2012, but the overall trend still aligns. This could be because of the lack of training data during this period, as shown in a previous plot. The predicted value follows closely with the actual values from 2012 to 2017, corresponding to the months we have enough data points. However, surprisingly, for months after 2017, the model still seems to predict well, following closely the trend. The model still predicted the monthly average even though the trend shifted dramatically. The model is robust against drastic variations over time. The linear regression has achieved satisfactory predictive power, and there is no need to fit a more complicated model.

Also, the coefficients in the model confirmed the exploratory data analysis. For instance, the coefficients for the area, the number of living rooms, the kitchen, and the drawing room are all negative, confirming that a larger house does not necessarily have a higher price. In addition, the building structure also played a vital role in house prices. Namely, a building made out of steel and concrete or just concrete has significantly higher mean prices than that made out of brick or mixed ingredients, and the coefficient for brick-wood buildings is negative, which means they are the least popular. Moreover, the coefficients for whether the property contains an elevator and proximity to a subway station are close to zero, meaning that they do not contribute much to the property prices.

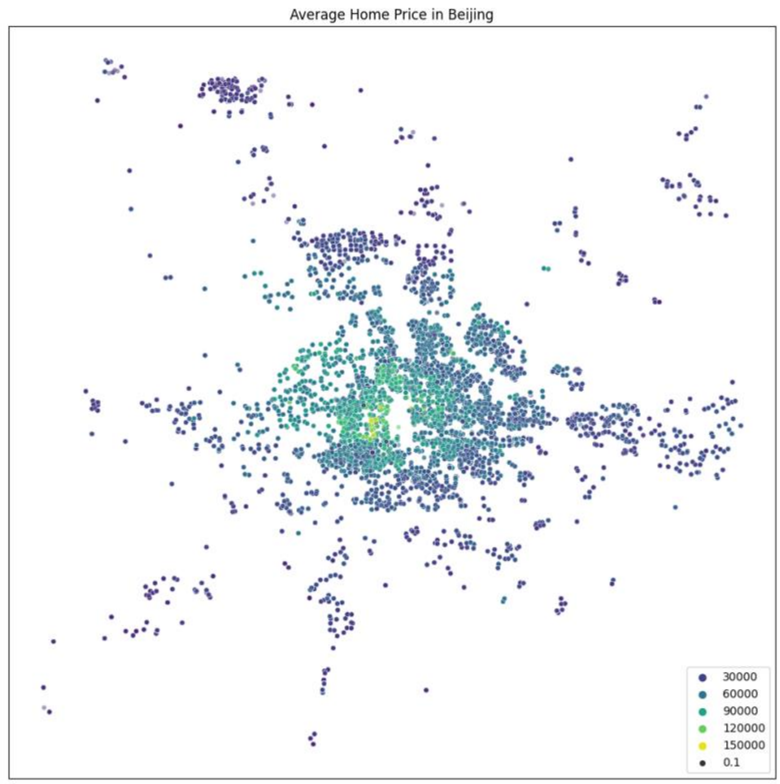
The spatial relationship is also examined. K-means clustering with 5 clusters is used to group the data points based on all their features except for the latitude and longitude, and the clustered points are then plotted to visualize the spatial relationship between them. K-Means clustering is a widely used unsupervised machine learning algorithm for partitioning a dataset into distinct groups or clusters. It operates on the principle of minimizing the within-cluster variance, thereby grouping similar data points into the same cluster while ensuring distinctness between



different clusters. The algorithm initializes with a predefined number of clusters, denoted as 'K.' It iteratively assigns each data point to the nearest cluster centroid based on a chosen distance metric, typically Euclidean distance. Initially chosen at random, these centroids are recalculated in each iteration as the mean of all points in the cluster, leading to a convergence where the cluster assignments no longer change significantly. The simplicity and efficiency of K-Means make it a popular choice for various applications, ranging from market segmentation to image compression, despite its susceptibility to initial centroid placement and its assumption of spherical clusters of similar size. The plot above is a spatial visualization of the clustered points using K equal to five, and we can see a clear layering pattern in which the clusters form a “growth ring” from the inside to the city's suburban areas. This shows that the characteristics of housing are distinct at different stages. For example, at the very center of the city is the old town, where the traditional bungalows are typical, corresponding to clusters zero and one. As the city expanded during the 1960s, it was heavily influenced by Soviet designs. Thus, Soviet architecture was prominent, which corresponds to cluster three. With the economic boost since 1978, more and more modern high-rises were built, which occupied a majority of the outer space of the city, corresponding to cluster four. Finally, in the modern era, as urbanization peaked, more and more suburban areas developed in real estate and infrastructures, corresponding to cluster two. The clustered points told an exciting story on the history and development of the city.

The clustered points also reflect the actual price distribution. As we can see from the plot, which plots the price per square meter of each property, the price distribution also forms a layering pattern, the same as the clusters. The average price per square meter is the highest at the center and gradually decreases when we move to the outer parts of the city. Also, the western and northern parts of the city have higher average prices than the eastern and southern parts, even at the northwestern boundaries. This was also reflected precisely by the clusters as cluster three extends to the northwestern boundaries while not much into the eastern and southern boundaries. The spatial distribution of the prices consolidates our previous conclusion that the house prices are distinctive with the characteristics of the properties themselves, which explains part of why an ordinary linear regression suffices in predicting the house prices.

(The above analyses are all conducted via Python and R)



Discussion

Through the in-depth investigation of housing prices in different neighborhoods across Beijing, I have uncovered a multifaceted web of interdependent variables. Ranging from broad macroeconomic trends to nuanced micro-level neighborhood characteristics, these variables paint a complex picture of the city's housing market. This research absorbs established theories in urban geographics while revealing previously unknown factors unique to Beijing's urban landscape. The results of this research align with many previous studies that highlight the role of property characteristics and economic circumstances in shaping housing costs. Nevertheless, specific demographic components appear to have a more pronounced effect in Beijing than in other major cities worldwide. This distinction emphasizes the need to acknowledge regional, cultural, and societal factors in creating urban economic frameworks. The spatial analysis sheds light on the geographical dimension of variations in housing prices. It confirms the idea that urban development patterns and access to critical urban infrastructure are crucial factors, as suggested by Cellmer et al. in their 2020 study. This highlights the importance of geospatial analysis in urban housing market research.

While this study provides a comprehensive analysis of Beijing's housing market, some limitations must be considered. Firstly, the data used in the study primarily focuses on the house properties, such as neighborhood characteristics, economic conditions, and amenities nearby. This means the analysis may only capture certain variables affecting house prices. Factors such as future urban planning policies, unexpected economic shifts, or societal trends that could significantly impact the housing market are beyond the scope of this study. They may have yet to be accounted for. Secondly, the reliance on historical data might only partially account for rapid changes or future trends in the real estate market, particularly in a fast-evolving city like Beijing, even though the model has already predicted well on unseen data. Another area for improvement is the model's potential bias due to omitted variable bias, where unobserved or unmeasured variables could skew the results. Lastly, while the study employs advanced statistical and machine learning methods, the accuracy of these models is contingent upon the quality and completeness of the data. Any data collection or processing inaccuracies could lead to less reliable model predictions. This is primarily a concern with the primary data set used in this study, as around twenty variables are used to fit the model. These limitations underscore the importance of carefully interpreting the study's findings and suggest areas for further research and data enhancement.

Another crucial aspect to consider is the ethical implications of this research. Acknowledging how the findings could affect policy-making and urban planning is essential. While our study provides valuable insights, it raises questions about housing affordability and social equity in urban development. It is essential to ensure that the benefits of urban development are distributed fairly to everyone. The potential influence of data-driven analyses on urban policy requires a thoughtful and responsible approach. The big data research emphasizes the importance of strict measures to protect data privacy. Although the data used in the research was anonymized and aggregated, the increasing availability of detailed urban data necessitates continued attentiveness to safeguard individual privacy. This study has important implications for urban policy, especially in housing regulation and urban planning. Policymakers can use this information to develop strategies that tackle the inequalities in housing affordability and enhance the overall quality of urban living. In order to better understand the changes in housing prices in Beijing over time, I suggest taking a longitudinal approach in future research. More recent data can be scraped and analyzed using the same approach as this paper. Comparative studies with other significant cities globally could provide a broader urban housing market dynamics perspective. Exploring the effects of emerging trends, such as remote work and environmental sustainability, on housing markets could also lead to valuable insights worth investigating.

Conclusion

This study has thoroughly analyzed the factors that affect housing prices in Beijing, using a combination of statistical and machine learning methods. The findings demonstrate that the characteristics of a neighborhood, the economic conditions of the local area, and the proximity to amenities are the key factors influencing housing prices. The advanced data analysis techniques used in the study have provided a detailed understanding of these relationships, revealing the complexity of the real estate market in Beijing. The study findings have significant implications for stakeholders such as policymakers, urban planners, investors, and residents. Policymakers and urban planners can use their knowledge about the impact of various factors on housing prices to create more effective urban development and housing policies. Investors and real estate professionals can also benefit from these insights to make more informed decisions in the property market. Furthermore, this study adds to comprehending urban housing markets in significant cities. It provides a structure that can be used in similar investigations in other situations. The study emphasizes the significance of utilizing big data and machine learning in real estate analysis, presenting a prototype for future research. It is crucial to recognize that this study has some limitations, such as its dependence on past data and probable biases in the modeling procedure. These constraints emphasize the importance of conducting continuous research and gathering data to enhance our comprehension of the factors influencing the housing market. This study provides a comprehensive analysis of the Beijing housing market, offering valuable insights and paving the way for further urban real estate economics research.

References

1. Favaretto, M., Clercq, E. D., Schneble, C. O., & Elger, B. S. (2020). What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. PLoS ONE, 15(2). https://doi.org/10.1371/journal.pone.0228987
2. Nichols, J. A., Herbert Chan, H. W., & B. Baker, M. A. (2019). Machine learning: Applications of artificial intelligence to imaging and diagnosis. *Biophysical Reviews, 11(1)*, 111–118. https://doi.org/10.1007/s12551-018-0449-9
3. BrainyQuote. (n.d.). Top 10 Geoffrey Hinton Quotes. Retrieved from https://www.brainyquote.com/lists/authors/top-10-geoffrey-hinton-quotes
4. Oecd. (n.d.). *Metropolitan areas*. https://stats.oecd.org/Index.aspx?DataSetCode=CITIES
5. Jiang, Y., & Qiu, L. (2022). Empirical study on the influencing factors of housing price —— Based on cross-section data of 31 provinces and cities in China. *Procedia Computer Science*, *199*, 1498-1504. https://doi.org/10.1016/j.procs.2022.01.191
6. Ruiqurm. (2018, July 7). *Housing price in Beijing*. Kaggle. https://www.kaggle.com/datasets/ruiqurm/lianjia/data
7. *Beijing*. Beijing Public Data Open Platform. (n.d.). https://data.beijing.gov.cn/
8. Beijing Statistical Yearbook. (n.d.). https://nj.tjj.beijing.gov.cn/nj/main/2018-tjnj/zk/indexeh.htm
9. Chen, D., & Li, R. Y. (2022). Predicting housing price in Beijing via Google and Microsoft automl. *Current State of Art in Artificial Intelligence and Ubiquitous Cities*, 105–115. https://doi.org/10.1007/978-981-19-0737-1_7
10. University of Michigan. China Data Center.. Beijing Hua tong ren shi chang xin xi you xian ze ren gong si. Beijing China Province Boundary, 2000. [Shapefile]. All China Marketing Research Co.. Retrieved from https://earthworks.stanford.edu/catalog/sde-columbia-cdc_census2000_beijingprov
11. Jiang, Y., & Qiu, L. (2022). Empirical study on the influencing factors of housing price —— Based on cross-section data of 31 provinces and cities in China. *Procedia Computer Science*, *199*, 1498-1504. https://doi.org/10.1016/j.procs.2022.01.191
12. Cellmer, R., Cichulska, A., & Bełej, M. (2020). Spatial Analysis of Housing Prices and Market Activity with the Geographically Weighted Regression. *ISPRS International Journal of Geo-Information*, *9*(6), 380. https://doi.org/10.3390/ijgi9060380
13. Zhang, Y., Zhang, D., & Miller, E. J. (2021b). Spatial autoregressive analysis and modeling of housing prices in city of Toronto. *Journal of Urban Planning and Development*, *147*(1). https://doi.org/10.1061/(asce)up.1943-5444.0000651