**ABSTRACT**

The study of Exploring the Impact of Location and Property Attributes on Housing Prices in Perth: A Comprehensive Analysis investigates the interplay between location, property characteristics, and housing prices in Perth, Australia. Employing advanced statistical methodologies, the research aims to uncover meaningful insights for stakeholders and policymakers. Chapter 3 introduces the CRISP-DM methodology, detailing stages from defining research goals to deploying insights. In Chapter 4, data analysis and visualization techniques reveal intricate patterns and correlations. This aids decision-making for buyers, sellers, and policymakers, contributing to urban planning. Chapter 5 delves into the complex dynamics influencing housing prices, highlighting the roles of location, property attributes, and sustainability features. The research provides valuable insights, guiding investment strategies, policy formulation, and urban development in a dynamic housing market context.

Keywords: Comprehensive Analysis, Urban Development, Dynamic Context, Statistical Methodologies, Stakeholders, Policymakers, CRISP-DM Methodology, Data Analysis, Visualization Techniques, Correlations, Investment Strategies, Housing Market

Table of Contents

[Chapter 1: Introduction 8](#_Toc144216126)

[1.1: Introduction 8](#_Toc144216127)

[1.2: Research Background 8](#_Toc144216128)

[1.3 Aim and Objectives 9](#_Toc144216129)

[1.4 Rationale 9](#_Toc144216130)

[1.5 Research Significance 10](#_Toc144216131)

[1.6 Summary 10](#_Toc144216132)

[Chapter 2: Literature Review 11](#_Toc144216133)

[2.1 Introduction 11](#_Toc144216134)

[2.2 Geographical Effects on Housing Prices 11](#_Toc144216135)

[2.3 Impact of the Social Economy on Housing Prices 13](#_Toc144216136)

[2.4 Housing Price Trends over Time 15](#_Toc144216137)

[2.5 Infrastructure’s Effect on Housing Prices 16](#_Toc144216138)

[2.6 Impact of Property Attributes on Housing Prices 17](#_Toc144216139)

[2.7 Housing Prices in Perth Supply and Demand Affect Prices 18](#_Toc144216140)

[2.8 Machine Learning Algorithms in House Price Prediction 20](#_Toc144216141)

[2.9 Literature Gap 21](#_Toc144216142)

[2.9 Summary 21](#_Toc144216143)

[Chapter 3: Methodology 22](#_Toc144216144)

[3.1 Introduction 22](#_Toc144216145)

[3.2 CRISP-DM 22](#_Toc144216146)

[3.2.1 Business Understanding: 23](#_Toc144216147)

[3.2.2 Data Understanding: 23](#_Toc144216148)

[3.2.3 Data Preparation 24](#_Toc144216149)

[3.2.4 Modelling 25](#_Toc144216150)

[3.2.5 Evaluation 25](#_Toc144216151)

[3.2.6 Deployment 25](#_Toc144216152)

[3.3 Tools and technologies used 26](#_Toc144216153)

[Chapter 4: Data Description and Exploration 26](#_Toc144216154)

[4.1 Introduction 26](#_Toc144216155)

[4.2.1 Data Source 27](#_Toc144216156)

[4.2.2 Data Overview 27](#_Toc144216157)

[4.2.3 Data Cleaning 28](#_Toc144216158)

[4.3 Exploratory Data Analysis (EDA) 29](#_Toc144216159)

[4.3.1 Descriptive Statistics 29](#_Toc144216160)

[4.3.2 Data Distribution 30](#_Toc144216161)

[4.3.3 Correlation Analysis 31](#_Toc144216162)

[4.3.4 Geographic Analysis 33](#_Toc144216163)

[4.4 Location Analysis 34](#_Toc144216164)

[4.5 Property Attributes Analysis 35](#_Toc144216165)

[4.6 Model Comparison and Selection 37](#_Toc144216166)

[4.6.1 Hyperparameter Tuning 37](#_Toc144216167)

[4.6.2 Evaluation 38](#_Toc144216168)

[4.7 Interpretation and Visualization 40](#_Toc144216169)

[4.7.1 Feature Importance 40](#_Toc144216170)

[4.7.2 Predictions vs. Actual Prices 41](#_Toc144216171)

[4.8 Deployment 42](#_Toc144216172)

[Chapter 5: Conclusion 42](#_Toc144216173)

[5.1 Conclusion 42](#_Toc144216174)

[5.2 Recommendation 43](#_Toc144216175)

[5.3 Future scope 43](#_Toc144216176)

[References 45](#_Toc144216177)

[Appendices 47](#_Toc144216178)

[Appendix A: Glossary of Terms 47](#_Toc144216179)

[Appendix B: Import necessary library for the task 48](#_Toc144216180)

Table of Figure

[Figure 1 : Indirect variables that affect real estate values. 11](#_Toc144115418)

[Figure 2 : Population Migration’s Effect on Urban Housing Prices 14](#_Toc144115419)

[Figure 3 : The possible effects of overvaluation in the US property markets and unpriced climate risk 17](#_Toc144115420)

[Figure 4 : CRISP-DM Process 23](#_Toc144115421)

[Figure 5 : Data Information 27](#_Toc144115422)

[Figure 6 : The steps for null functions 28](#_Toc144115423)

[Figure 7 : The result of the descriptive statistics 29](#_Toc144115424)

[Figure 8 : The histogram plot 30](#_Toc144115425)

[Figure 9 : Correlation heatmap showing relationships between key variables. 31](#_Toc144115426)

[Figure 10 : The geographical representation 33](#_Toc144115427)

[Figure 11 : Scatter plot of Price Vs Nearest School, Station, and CBD distances 34](#_Toc144115428)

[Figure 12 : Scatter plot of Price Vs Land and Floor Area 35](#_Toc144115429)

[Figure 13 : Scatter plot of Price Vs number of Bathroom, Bedroom, and Garage 36](#_Toc144115430)

[Figure 14 : Average house price by build year 36](#_Toc144115431)

[Figure 15 : Random Forest Feature Importance 40](#_Toc144115432)

[Figure 16 : Random Forest Predictions vs Actual Prices 41](#_Toc144115433)

# Chapter 1: Introduction

## 1.1: Introduction

The introduction to the research provides a framework for a thorough assessment of the elements affecting Perth real estate prices, which serve as Australia’s hub for international commerce. The research makes use of a large dataset that was gathered from several sources to investigate the link between location, property characteristics, and housing costs. The research takes use of contemporary statistical analysis and machine learning tools in order to uncover patterns and relationships and deliver helpful information to real estate purchasers, investors, decision-makers, and urban planners. The research will advance our knowledge of housing and real estate economics, as well as urban planning and decision-making.

## 1.2: Research Background

By making educational resources accessible to the general public, decision-makers, and real estate investors, the project aims to address this knowledge gap. Examining the considerable contributions that location and property characteristics make to home prices in Perth is the main goal of this thorough investigation. It is crucial to comprehend how geography affects price discrepancies in housing since various geographic regions within a city may have various features. The closeness of amenities, job centres, transportation hubs, and natural features may have a big influence on property prices. The investigation will include other property characteristics including size, number of bedrooms, bathrooms, and modifications like garages to ascertain how these factors impact house values. A large and varied dataset will be acquired from a variety of sources, including real estate firms, property listings, governmental records, and demographic information, in order to accomplish the project’s objectives. The data acquired will comprise information on houses that were sold in certain Perth neighbourhoods throughout a specified time frame (Li, Hu, and Liu, 2020). The research searches for connections between house prices and location/property parameters using complex statistical approaches, data visualisation, and machine learning algorithms.

Numerous consequences stem from the investigation’s conclusions. Based on the relevant criteria that have been found, homebuyers and investors may make better educated judgements regarding real estate purchases and investments. In order to encourage fair and sustainable urban expansion, policymakers and urban planners may utilise the information to improve zoning regulations, infrastructure construction, and housing legislation (Lieber, 2022). By being more cognizant of these elements, real estate agents may enhance their client suggestions and pricing tactics. By serving as a springboard for further investigation into the details of housing markets in other locations, the findings may deepen our knowledge of the issue. All things considered, this research has the potential to significantly clarify the intricate relationships that exist in Perth between location, property characteristics, and housing prices. The study aims to enhance evidence-based decision-making, explain the dynamics of the housing market, and develop a more stable and effective urban real estate market by carefully evaluating these components.

## 1.3 Aim and Objectives

With an emphasis on the impacts of location and property quality, the dissertation’s objective is to perform a detailed investigation of the variables that affect home costs in Perth, Australia.

**Objectives**

* To evaluate the effect of various property features on housing costs
* To give city planners and decision-makers data-based guidance so they can create effective housing policies and urban development plans.
* To provide practical insights for homebuyers, real estate investors, and agents

## 1.4 Rationale

There are issues with the rationale project. In order to comprehend the dynamics of the city’s real estate market, the study has closely investigated the elements affecting Perth property prices. It has looked at the intricate connection between geographical features and architectural traits. Data from multiple sources have been combined to build a large and diverse dataset that offers specific information on homes that were sold in various Perth areas over a long period of time. The research team looked for relationships and trends between home prices and location/property parameters using cutting-edge machine learning methods, in-depth statistical analysis, and data visualisation. In order to determine how they could affect property price, a variety of geographic variables, including proximity to services, transportation hubs, employment centres, and natural features, as well as a number of property characteristics, including size, number of bedrooms, bathrooms, and extras, have been examined.

The findings of this thorough analysis may very well provide direction and suggestions to various parties. The process of selecting the residences and establishing their pricing may be delegated to brokers, investors, and purchasers with greater knowledge. By changing the rules governing housing and infrastructure, policymakers and urban planners may promote fair and sustainable urban expansion (Lowies et al., 2022). The initiative aims to promote relevant data-driven current research and significantly broaden our knowledge of urban studies, housing economics, and real estate. The rationale’s main goals are to simplify the Perth real estate market, enable sustainable market development, and make it easier to make choices based on accurate information.

## 1.5 Research Significance

Because it could provide useful information for buyers, investors, and real estate brokers, the research is noteworthy. The study’s conclusions may aid in their better evaluation of real estate investments, acquisitions, and pricing tactics, resulting in more informed and successful transactions. The discoveries have important ramifications that planners and politicians must consider. People might make better judgements about urban growth, infrastructure development, and housing policy if they were aware of the effects that location and property features have on house prices. This may encourage ecologically responsible and fair growth, improving Perth’s overall liveability and quality of life. Researchers and academicians working in the fields of real estate, housing economics, and urban studies may find the study’s conclusions to be of great use. It offers a thorough examination of the variables affecting real estate expenditure and serves as a basis for further research and comparison studies in related areas.

## 1.6 Summary

The research looks at the complex interplay between property quality, community characteristics, and housing pricing in Perth, Australia. The research intends to provide beneficial insights for homebuyers, investors, legislators, and urban planners via in-depth data analysis and cutting-edge statistical methodologies. Due to its contributions to the domains of real estate and housing economics, as well as urban planning and decision-making, the research is significant. The results might be useful to decision-makers, and the predictive models that were created could help forecast probable changes in the housing market in the future.

# Chapter 2: Literature Review

## 2.1 Introduction

This chapter has been examined the research on the influences of various factors on Perth housing prices and analyse six research themes: geographic impact, socio economic impact, historical housing price trends, infrastructure impact, and property attribute impact. This chapter’s literature review has been looked at the most current research on each topic and offers details on how these factors could impact Perth property prices. This chapter has been gone through the findings from a range of theoretical and empirical research to provide readers with a better understanding of the present body of literature on this issue.

## 2.2 Geographical Effects on Housing Prices

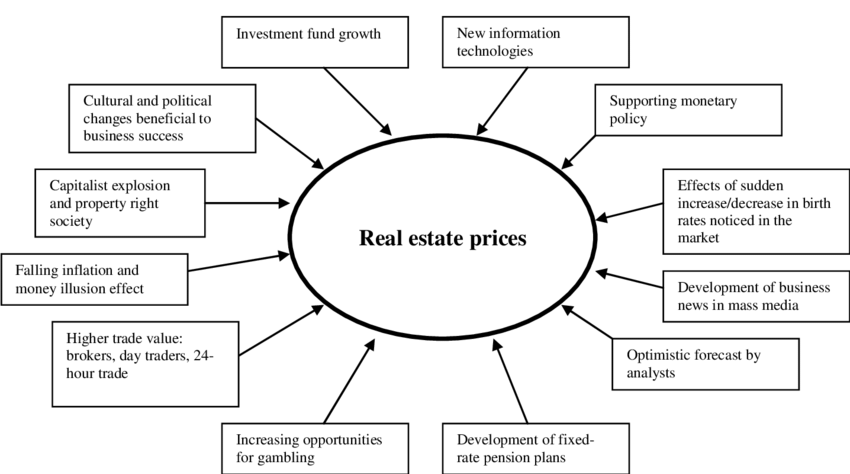


Figure 1 : Indirect variables that affect real estate values.

(Source: https://www.researchgate.net/)

There has been research on the relationship between location and housing prices, with an emphasis on Perth. A multiscale spatially and chronologically weighted regression model was employed by Wu *et al.* (2019) to investigate the temporal and spatial variables that influence Perth house prices. According to the study, property prices were much higher in areas with lush landscaping and excellent school catchment areas that were near one kilometre of a public transit stop. Additionally, it was shown that neighbourhoods with more restaurants and bars had higher house prices. Yang *et al.* (2020) examined the effects of BRT on Perth property prices with an emphasis on ridership and proximity. The findings showed that property prices increased close to BRT stations and decreased considerably further away, with the biggest decline taking place between one and two km from the station. The study discovered that the positive accessibility impact BRT would have on property prices would make it advantageous to include it in urban transportation improvements.

Karahan and Rhee (2019) used Geographic Reallocation (GR) to analyse the effect of the housing slump on unemployment in Perth and found that although the effect was significant initially, it diminished over time. Results showed that in order to lessen the effects of economic shocks, appropriate planning and investment were required. In 2019, Cloyne et al. investigated the impact of Perth housing prices on household debt. Their data suggest that increasing housing prices increase household leverage, particularly in the suburbs. It was discovered that the structure of the housing market, namely the availability of less expensive properties with lower potential capital returns, was a significant factor associated with consumer borrowing behaviour.

According to the study, location has a significant impact on Perth property prices. The value of a house or flat may be significantly raised by its convenient location and closeness to restaurants, public transportation, schools, and other facilities. Properties close to BRT stations often sell for a premium, as do those in more desirable locations or regions with higher median earnings. Historical trends also seem to influence property prices in the aftermath of a recession, when GR appears to play an important role in both cutting unemployment and protecting against negative economic shocks. Other factors, such as the availability of less priced residences with lower potential for capital gains, have an impact on consumer borrowing as well. According to the study, location is by far the most crucial factor in determining Perth property prices. Urban development initiatives, zoning regulations, and tax policies are only a few more possible variables that need further research. Further research is also required to ascertain the combined effects of location and infrastructure on housing costs. The results may be used to influence policy objectives and provide information on areas where investment may have the most effect on housing costs.

## 2.3 Impact of the Social Economy on Housing Prices

Neighbourhood income and racial demographics have been emphasised as socioeconomic factors that are essential for determining how property prices affect people’s lives. In previous literature studies on this topic, immigration, segregation, and labour market integration are only a few of the topics that have drawn the greatest attention.

According to Massey and Tannen (2018), between 1970 and 2010, suburbanization had a negative impact on racial segregation in the United States. They note that during this time, both Hispanic and African American families were moving more often to suburban areas, which contributed to the reinforcement of residential segregation. This is a crucial example of how the socioeconomic characteristics of a neighbourhood may influence housing prices and, in this case, have a long-term impact on racial segregation.

Wittowsky *et al.* (2020) conducted an empirical study to better understand how residential property costs in Dortmund, Germany is influenced by house characteristics, accessibility, and surrounding flats. The poll found that location and closeness to amenities and public transport were the main determinants of property prices. This study serves as an example of how a community’s socioeconomic mix may have a big influence on housing prices.

Research has also looked at the relationship between a migrant’s sense of community and housing expenses. Liu *et al.* (2022) looked at this subject in Beijing, which has a long and complex history of immigration. According to their study, immigrants’ levels of acceptance and integration differed, which in turn affected the price of housing they were willing to pay. This provides as an example of how the presence of immigrants may impact housing costs in a town.

Wimark *et al.* (2019) also investigated how an immigrant’s original location, subsequent employment, and income affected their ability to integrate into the labour market. They found that if immigrants initially chose to reside in places with a higher population of immigrants, they were more likely to benefit from greater economic and employment market opportunities. This raises the idea that a neighbourhood’s socioeconomic characteristics may have a long-lasting impact on property prices since recent immigrants often struggle to integrate into their new community.

These results suggest that the cost of housing may be significantly influenced by regional socioeconomic characteristics. Decision-makers and urban planners must thus consider the possible impacts of socioeconomic factors, such as neighbourhood income, racial demography, immigration, and labour market integration, on housing prices. For example, steps should be developed to ensure that recent immigrants have the resources and opportunities to adapt into their new communities in order to prevent some of the housing-related challenges. Additionally, laws that address gentrification and economic inequality should be implemented since they may have a significant long-term impact on property values.

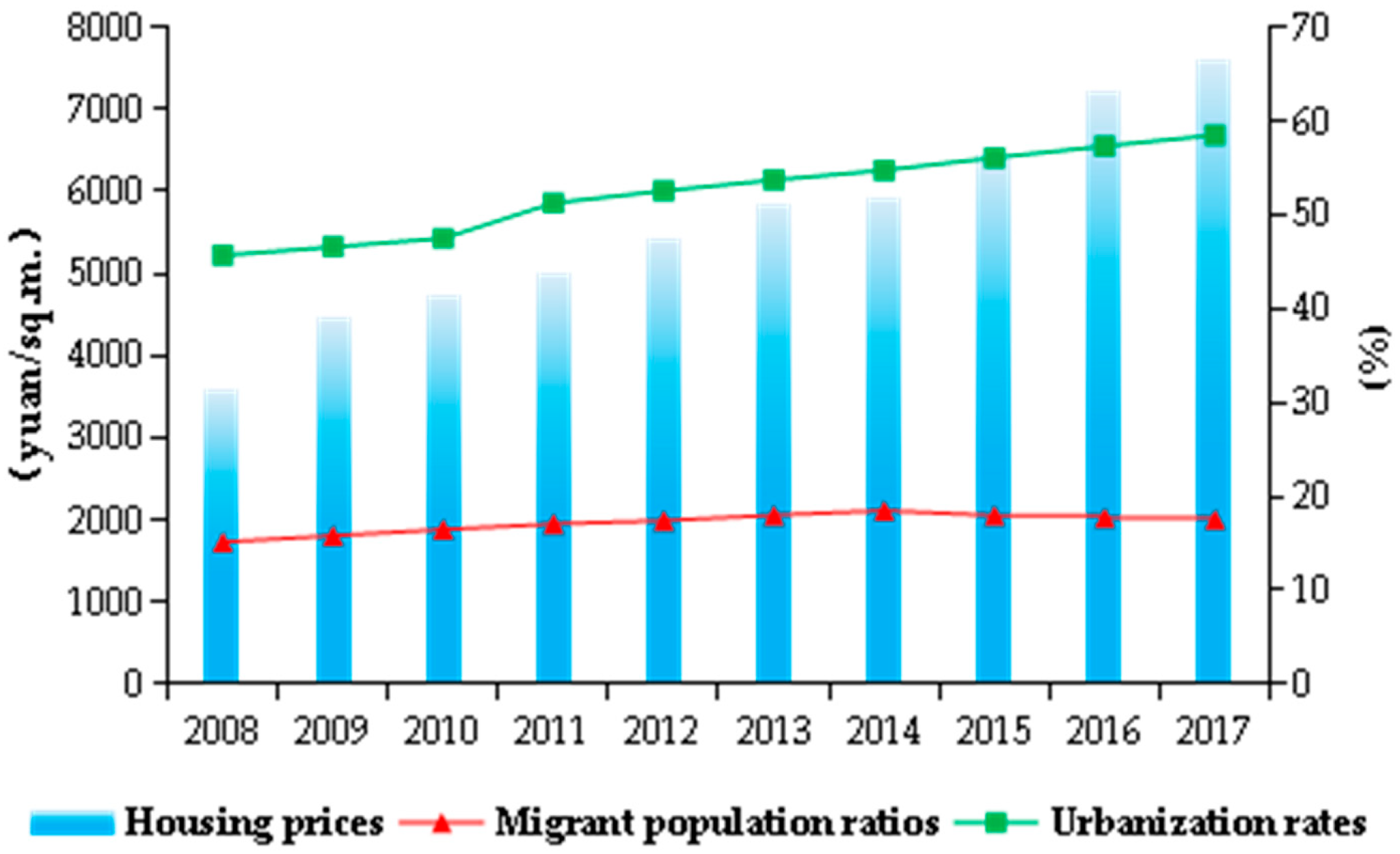


Figure 2 : Population Migration’s Effect on Urban Housing Prices

(Source: <https://www.mdpi.com/>)

Finally, it’s crucial to consider socioeconomic aspects while establishing land use and zoning regulations. Giving the general public access to amenities like parks, public transportation, and amenities, as well as establishing incentives for developments that provide affordable housing, are examples of this. These laws might help address some of the socioeconomic factors that affect residential housing prices and make sure that everyone in society has access to safe, affordable housing.

## 2.4 Housing Price Trends over Time

The value of a house in Perth is heavily influenced by the market conditions there during the last many decades. According to Gharahighehi *et al.* (2021), a better understanding of property price trends over time might aid in making wise investments. The performance of the market in the past may be used as a proxy for its success in the future, according to research (Nethercote, 2020). Historical home price patterns, in addition to variables like population growth, infrastructural development, and economic activity, may provide light on the future of a property market (Bangura and Lee, 2020).

There is a dearth of published data on the condition of the Perth property market. There is a need for more regional data on housing price inflation and deflation, as suggested by a literature review (Gharahighehi *et al.* 2021). The present predicament remains despite the availability of data revealing historical trends in home price fluctuations. Economic recessions are only one time-specific aspect that has to be accounted for when analysing home prices in Perth (Wei *et al.* 2022). We still don’t know enough about how historical home price patterns in an economy could affect present choices in the housing market, especially in the years after the global financial crisis (Nethercote, 2020).

Historical property price trends in Perth may be better understood by comparing data from many sources. Deed and mortgage records, taxation records, and transaction records are all examples of public and private data sets that might be used in this way (Gharahighehi *et al.* 2021). Personalised indices that provide light on the historical performance of the Perth real estate market are possible to derive from such data sets (Wei *et al.* 2022). Housing markets at the regional level may be examined, as can the effect that time-specific events, such as economic recessions, have on property values.

Previous modifications in the Perth real estate market’s pricing structure need careful analysis. Market performance may be anticipated using these kinds of patterns (Bangura and Lee, 2020). More study of the housing market’s past tendencies in Perth is needed due to its complexity and the wide range of elements that affect it. It may be helpful to examine data from many sources concurrently in order to better understand how property values in the community have evolved over time. Deed and mortgage records, taxation records, and transaction records are all examples of public and private data sets that might be used in this way (Gharahighehi *et al.* 2021). Personalised indices derived from such data sets might provide insight into the workings of the Perth real estate market (Wei *et al.* 2022). More research is needed to understand how past events and trends have affected property values in Perth throughout the years. This kind of study might provide light on how the present economic downturns have affected the real estate market. The results of this study might potentially be utilised to improve housing market forecasts and guide residential real estate investment choices.

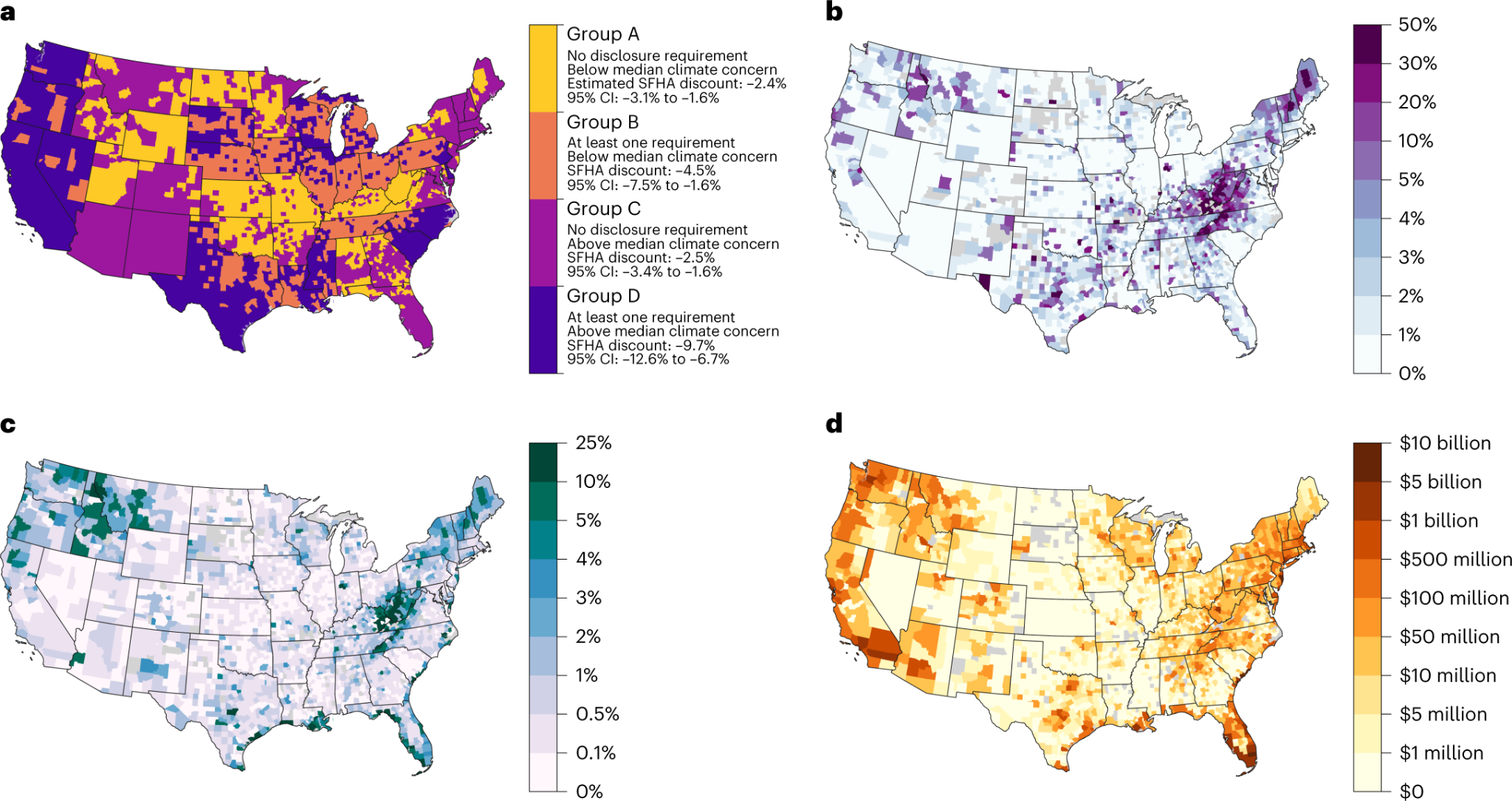
## 2.5 Infrastructure’s Effect on Housing Prices

Numerous academic fields, including economics, sociology, geography, and urban planning, have studied how infrastructure affects housing prices. Since infrastructure refers to factors connected to public services, such as roads, bridges, sewage systems, electrical networks, and other public infrastructure, research has shown that certain infrastructure components may have an impact on real estate values. Nozhati *et al.* (2019) found that the results of food security after a disaster depend on the infrastructure. They paid particular attention to the recovery of families in Joplin, Missouri, after the storm in 2011, and found that transportation and health facility accessibility significantly influenced community food security outcomes. Overall, their findings suggest that, in addition to housing prices, infrastructure has an impact on other socio-economic parameters.

The role of private equity in the management of this system was stressed by Bayliss *et al.* in their research on the financialization of infrastructure in Britain in 2023. They said that the public sector was being underinvested in infrastructure projects, which led to an uneven distribution of resources and higher housing prices. This study shows how spending on private infrastructure may lead to increased home prices and worsening living conditions. Allegrante and Sleet (2021) looked at the need for spending on infrastructure and public health in order to solve the problems associated with homelessness. They found that community-based systems of care, such as shelters, street outreach, and staff-monitored houses, had better outcomes for persons who were suffering homelessness than traditional housing schemes. Their study did not particularly address housing prices, but it did demonstrate how improving infrastructure, such as access to healthcare, may help improve the lives of underprivileged groups. Ficek (2018) assessed the effect of infrastructure on housing prices in Puerto Rico during Hurricane Maria. This study found that the cost of living and housing prices increased as a result of the destruction of old infrastructure. The unequal allocation of resources that followed the disaster, which expanded disparities between different socioeconomic strata, was also examined by Ficek.

These studies often demonstrate how the prices of housing may be impacted by infrastructure elements such as roads, bridges, and sewage systems. Research suggests that expanding access to certain services and amenities, such public transportation, healthcare options, and career opportunities, may raise housing prices. A lack of infrastructure or the destruction of already-existing infrastructure, on the other hand, can lead to higher living costs and property prices. This demonstrates the need for governments to invest in infrastructure systems to ensure equal housing prices and living standards.

## 2.6 Impact of Property Attributes on Housing Prices



**Figure 3**

Figure 3 : The possible effects of overvaluation in the US property markets and unpriced climate risk

**:**

(Source: <https://www.nature.com/>)

The literature has a wealth of information on the potential effects of property attributes including size, proximity to amenities, and condition on home values. According to research, homes with more amenities or larger rooms have been demonstrated to have a greater level of price premiums. Many property markets have reported on this phenomenon, including those in developed nations like the United States and the United Kingdom as well as emerging economies like China and India.

The emphasis of Feng and Humphreys’ (2018) study was on how sports facilities affect residential property prices in the US. The study’s findings indicate that when sports stadiums and fields are situated within a 3-kilometer radius of residential properties, the perceived convenience they provide raises property values. The relationship between walking accessibility and real estate prices in Beijing was examined by researchers Yang *et al.* (2018). In order to assess their potential influence on the identified link, the study included geographic features and air pollution levels as controlled variables. The study’s findings demonstrated that families living in walkable neighbourhoods made more money than those living in less walkable neighbourhoods. It was shown that this link became even more significant when the distances to key facilities, such as marketplaces and hospitals, were taken into account.

Housing expenses are significantly influenced by a property’s condition in addition to the presence of amenities and conveniences. Finding out how low-income property development affects housing prices was the main goal of Diamond and McQuade’s, (2019) study. The researchers also investigated how a property’s desirability to potential buyers was influenced by its condition. The study’s findings indicate that consumers seeking to purchase a home choose properties with contemporary amenities and outstanding condition over those with out-of-date features. The size of a property may also have an impact on the expense of living there. A study was conducted by Kang *et al.* in (2021) to look at how property size affected the rise in house prices. They used machine learning techniques and multi-source big data. According to the study’s results, residential property dimensions were important in determining housing costs in well-known international cities including London, New York City, and Tokyo.

This study found that a variety of property attributes, including amenities, condition, and size, had a significant influence on housing expenses. It goes without saying that properties with larger interior spaces, convenient access to services, and better upkeep are more in demand and hence sell for more money. As a result, it’s crucial to carefully consider these factors in order to completely understand how location and property attributes impact home prices in Perth.

## 2.7 Housing Prices in Perth Supply and Demand Affect Prices

The relationship between housing market supply and demand and the resulting real estate price has long been studied in economic literature. Housing markets throughout the globe have lately transformed as a result of population growth and economic development. The housing market has been strained, particularly in Perth, as a result of rapid development, increasing housing demand, and rising prices. In this article, we look at recent studies on the factors of supply and demand that have impacted Perth, Australia’s housing prices over the last several years.

Kendall and Tulip (2018) examined the impact of zoning on housing prices in the Perth region. They found that zoning was associated with higher home prices, along with other factors including population density, economic growth, and employment levels. Costello *et al.* (2019) found that factors like population growth that had boosted demand and lowered supply had a substantial influence on housing prices in their examination of geographical changes in housing submarkets in relation to demand. Both owner-occupied and rental housing have shown an increase in demand over time, according to Ma *et al.* (2019), who examined this trend in numerous Australian states.

Debelle, (2019) discussed the impact of Perth’s real estate market on the whole Australian economy. The author described how a difficult housing market was caused by an increase in housing demand, which led to higher prices since there wasn’t enough inventory. However, more wealthy buyers were able to take advantage of the situation and buy comparatively cheap properties.

Shi *et al.* (2020) assert that the fundamentals and speculation both had a role in the expansion of the Australian housing market. They said that the lack of wage growth and the cheap interest rate environment had increased demand, which had resulted in higher prices. The authors also hypothesised that the increase may be brought on by a growth in investor and international participation.

The difficulty of renting or buying a home in or around Perth has typically increased due to the mismatch between supply and demand. Numerous socioeconomic factors are contributing to this imbalance, which is increasing housing prices. Slow wage growth and low borrowing costs have boosted the market, enabling more wealthy buyers to benefit from the situation via capital gains. Rising foreign investment has also increased speculation, which has raised prices.

The research shows that in order to better understand how supply and demand affect housing prices, it is essential to take into account the many factors impacting the phenomena. Zoning could limit supply growth, raising prices as a result. Demand volume may be affected in a similar way to how economic growth, population size, and service accessibility are affected. High amounts of foreign involvement and speculation might make things more difficult. Policymakers must take a wide range of problems, both public and private, into account in order to make informed decisions and put in place appropriate mechanisms to support the Perth housing market.

Since they influence Perth real estate prices via the supply of available houses and the consequent demand, policy makers need to have a full understanding of the many underlying trends that influence supply and demand in the housing market. Examining the many public as well as private initiatives, such as zoning regulations, foreign investment, and speculation, is necessary to ensure that the right processes are in place to stabilise housing prices. Regulations like tax breaks, improved infrastructure, and improved access to social services and amenities should be implemented to encourage foreign investment, increase housing supply, stop speculation, and provide potential buyers fair alternatives.

## 2.8 Machine Learning Algorithms in House Price Prediction

Machine learning algorithms are a subset of artificial intelligence that can analyse large datasets, learn patterns, and make predictions without explicit programming. In real estate, these algorithms are used to predict house prices based on various features and attributes associated with properties. Common machine learning algorithms for house price prediction include Linear Regression, Decision Trees, Random Forest, Gradient Boosting, Support Vector Machines (SVM), and Neural Networks (Allegrante and Sleet, 2021). Despite the growing popularity of machine learning algorithms in predicting house prices, there are still challenges and gaps in existing research. These include data quality and availability, feature engineering, interpretable models, domain knowledge, spatial analysis, handling heterogeneous markets, and model validation.

Data quality and availability are crucial for predicting house prices, while feature engineering ensures the selection and transformation of relevant features. Interpretable models can provide insights into factors affecting housing prices, while domain knowledge can enhance model development. Spatial analysis can improve predictive power by capturing the unique characteristics of different housing markets. Handling heterogeneous markets requires developing models that capture the unique characteristics of different regions and cities (Bangura and Lee, 2020). Addressing these gaps has been enhanced the effectiveness and practicality of machine learning algorithms in predicting housing prices, providing valuable insights to buyers, sellers, and policymakers in the real estate market. Support Vector Machines (SVM) is a supervised learning algorithm for regression tasks, working well with high-dimensional data and handling complex relationships between features and target variables. Neural networks, particularly deep learning models, have shown promising results in house price prediction but may require more data and computational resources.

## 2.9 Literature Gap

Although the literature is still scant, machine learning algorithms are increasingly being used to forecast property values. Some of the areas where there are still gaps include feature engineering, interpretable models, domain knowledge, geographic analysis, targeting various markets, and model validation. Data availability and quality are essential for accurate projections, and feature engineering and selection are essential for identifying relevant traits. Understanding the unique characteristics of diverse property markets is crucial for models to accurately anticipate house values. Domain expertise, geographic analysis, and market management are essential for enhancing the prediction power of machine learning algorithms. To be robust and comprehensive, predictive models need to go through rigorous model validation and testing. Addressing these gaps has been improve the effectiveness and applicability of machine learning algorithms in predicting housing prices, assisting buyers, sellers, and policymakers in making informed decisions in the real estate market.

## 2.9 Summary

According to the literature review, following supply and demand for housing, location and the state of the economy are the two main variables influencing property prices in Perth. Aspects like neighbouring facilities, public transit, and schools have a direct impact on how important a property’s location is. For instance, two social economic elements that could affect housing costs are neighbourhood income and racial demography. Infrastructure may increase or decrease living expenses, and historical patterns may provide light on how Perth real estate prices have been impacted by earlier occurrences. Housing supply and demand, as well as variables like size, location, and condition, all have an impact on home prices. Each of these elements must be taken into consideration when trying to understand how and why Perth’s property prices are what they are.

# Chapter 3: Methodology

## 3.1 Introduction

In this section, the methodology serves as a guide for exploring the impact of location and property attributes on housing prices in Perth through a comprehensive analysis. The main objective of this research is to understand the intricate relationships between these variables and their contribution to housing price dynamics. To achieve this, the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology is employed to provide a structured and systematic framework for the data analysis process.

## 3.2 CRISP-DM

CRISP-DM (Cross-Industry Standard Process for Data Mining) is a widely used methodology for guiding the process of data analysis and mining. It provides a structured framework that encompasses various stages of data exploration, preparation, modelling, evaluation, and deployment. The rationale for choosing the CRISP-DM model in this research is its proven effectiveness in guiding systematic and organized data analysis processes. Its well-defined stages ensure a comprehensive approach to understanding the relationships between housing prices, location factors, and property attributes in Perth. This methodology promotes a clear and logical flow of tasks, making it easier to manage the complexities of the research and draw meaningful insights from the data.

The CRISP-DM methodology comprises six distinct phases,

* Business Understanding
* Data Understanding
* Data Preparation
* Modelling
* Evaluation
* Deployment

Figure 4 : CRISP-DM Process**A diagram of a data processing process

Description automatically generated**

### 3.2.1 Business Understanding:

In this phase involves clarifying the research's purpose: to comprehensively explore how housing prices in Perth are influenced by both location-related factors (e.g., distance from CBD, nearest amenities) and property attributes (e.g., bedrooms, bathrooms). By defining clear research goals and specific objectives, the study aims to uncover patterns, correlations, and trends that can provide valuable insights to real estate professionals, policymakers, and investors. This phase sets the stage for the subsequent analysis, focusing on understanding the intricate interplay between these variables and their impact on housing prices.

### 3.2.2 Data Understanding:

The process initiates by importing the data into the tools. A housing price dataset employed in this research is sourced from <https://www.kaggle.com/datasets/syuzai/perth-house-prices>, consisting of 9733 records and 19 attributes. These attributes encompass housing particulars and location factors. Employing an analogous approach to a telecom churn dataset, the analysis embarks on Exploratory Data Analysis (EDA) and subsequent pre-processing stages. EDA facilitates comprehension of data nuances, identification of missing values, and evaluation of attribute distribution. This foundational step is instrumental in unravelling the interplay between property attributes, geographical location, and housing prices within the Perth Metropolitan Area.

The dataset provides information on various attributes related to housing properties. These attributes include**:**

* ADDRESS: The address of the property.
* SUBURB: The suburb in which the property is located. [here](https://www.homely.com.au/find-suburb-by-region/perth-greater-western-australia)
* PRICE: The price of the property.
* BEDROOMS: The number of bedrooms on the property.
* BATHROOMS: The number o
* f bathrooms on the property.
* GARAGE: The number of garage spaces available.
* LAND\_AREA: The land area of the property (m2).
* FLOOR\_AREA: The floor area of the property (m2).
* BUILD\_YEAR: The year in which the property was built.
* CBD\_DIST: The distance of the property from the central business district (CBD) (m).
* NEAREST\_STN: The nearest train station to the property.
* NEAREST\_STN\_DIST: The distance of the property from the nearest station (m).
* DATE\_SOLD: The date on which the property was sold.
* POSTCODE: The postal code of the property location.
* LATITUDE: The latitude coordinates of the property.
* LONGITUDE: The longitude coordinates of the property.
* NEAREST\_SCH: The nearest school to the property.
* NEAREST\_SCH\_DIST: The distance of the property from the nearest school (m).
* NEAREST\_SCH\_RANK: The ranking of the nearest school.

### 3.2.3 Data Preparation

In the data preparation phase, raw housing price data will be transformed into a structured format ready for analysis. This involves handling missing values, detecting, and addressing outliers, creating new informative attributes, and converting categorical data. Tools such as Python programming and libraries like Pandas will be used for data manipulation and cleaning, along with Scikit-Learn for outlier detection. Attributes will be normalized or scaled using Scikit-Learn to ensure fair representation across features. The process can be documented using platforms like Google Colab for transparency. Thorough data quality checks and dataset splitting for training and testing, employing tools like Scikit-Learn, will ensure the analysis's reliability.

By employing these techniques and tools, the data will be made accurate, consistent, and well-prepared for subsequent analysis, enhancing the thesis's exploration of the impact of location and property attributes on housing prices in Perth.

### 3.2.4 Modelling

In the modelling phase, the primary objective is to develop predictive models that reveal the interplay between location attributes, property characteristics, and housing prices within the Perth housing market. This phase involves employing advanced statistical techniques and machine learning algorithms to establish a robust understanding of the factors that drive housing prices. By leveraging attributes such as number of bedrooms, bathrooms, garage availability, land, and floor areas, build year, and proximity to key points of interest like CBD distance, nearest station, and nearest school, the models will aim to capture the complex relationships inherent in the dataset.

The modelling process will encompass various steps, including feature selection, model selection, model training, hyperparameter tuning, and model evaluation. Linear regression models, as well as more sophisticated machine learning methods like decision trees, random forests, and gradient boosting, will be explored to capture both linear and nonlinear relationships. This modelling effort aims to reveal insights into the complex interplay between location, property attributes, and housing prices, contributing to a deeper understanding of the Perth housing market dynamics.

### 3.2.5 Evaluation

In the evaluation phase, the performance of these models will be thoroughly evaluated using appropriate metrics such as Root Mean Squared Error (RMSE) and R-squared, allowing for a comprehensive assessment of their predictive capabilities. Furthermore, the evaluation process extended beyond quantitative metrics to encompass a qualitative analysis of the model coefficients and their interpretability. This analysis shed light on the relative influence of different property attributes on housing prices, enabling a deeper understanding of the specific factors that drive price fluctuations in Perth's housing market.

### 3.2.6 Deployment

In this phase, the insights gained from the comprehensive analysis of housing prices in Perth are translated into practical solutions. The predictive models developed to understand the impact of location and property attributes on housing prices are made available for real estate professionals, buyers, and policymakers. These models can assist in estimating property values, aiding in informed decision-making during transactions. Additionally, the findings of the analysis offer valuable insights into the dynamics of the housing market and can contribute to policy discussions aimed at ensuring sustainable urban development.

## 3.3 Tools and technologies used

The successful execution of the comprehensive analysis of housing prices in Perth required the utilization of various tools and technologies. These resources were instrumental in data preprocessing, modelling, evaluation, and result visualization. The following list provides an overview of the key tools and technologies employed:

IDE (Integrated Development Environment):

* Google Colab: Google Colab served as the primary environment for coding, analysis, and documentation, allowing for an interactive and well-documented workflow.

Programming Languages:

* Python: Python served as the primary programming language for data manipulation, analysis, and modeling. Libraries such as NumPy, pandas, Urllib, and scikit-learn facilitated efficient data handling and machine learning implementations.

Data Visualization:

* Matplotlib: Matplotlib was utilized to create informative data visualizations, including scatter plots, histograms, and trend graphs, aiding in the exploration of relationships between housing attributes and prices.
* Seaborn: Seaborn was employed for creating visually appealing and informative statistical visualizations.
* Plotly: Plotly library enables interactive and dynamic data visualization, allowing users to create interactive charts, graphs, and dashboards for better understanding and exploration of data.

Machine Learning and Modelling:

* scikit-learn: The scikit-learn library provided a wide array of machine learning algorithms, including linear regression, decision trees, and random forests, gradient boosting enabling the development of predictive models.

# Chapter 4: Data Description and Exploration

## 4.1 Introduction

In this chapter, data description and exploration will be stated by using the Perth housing dataset, aiming to uncover valuable insights regarding the impact of various location-based and property attribute factors on housing prices. The foundation of any meaningful analysis lies in understanding the data itself, and this chapter serves as a crucial step in our journey. By describing the dataset's characteristics, preprocessing steps, and conducting exploratory data analysis (EDA), we lay the groundwork for the subsequent analytical chapters.

**4.2 Data Collection and Preprocessing**

### 4.2.1 Data Source

The dataset used for this analysis was sourced from Kaggle, a prominent platform for sharing and discovering datasets. The dataset titled "Perth House Prices" is available at the following URL: <https://www.kaggle.com/datasets/syuzai/perth-house-price> . It contains a comprehensive collection of housing-related information from the Perth area, a prominent city in Western Australia known for its diverse real estate landscape.

### 4.2.2 Data Overview

The data snippet comprises information related to housing properties, including attributes such as address, suburb, price, bedrooms, bathrooms, garage spaces, land, and floor areas, build year, distance to CBD, nearest train station, date of sale, postcode, geographical coordinates, nearest school, and its attributes. This dataset consists of 9,733 rows and 19 columns. It offers insights into property characteristics, location, and proximity to key amenities.

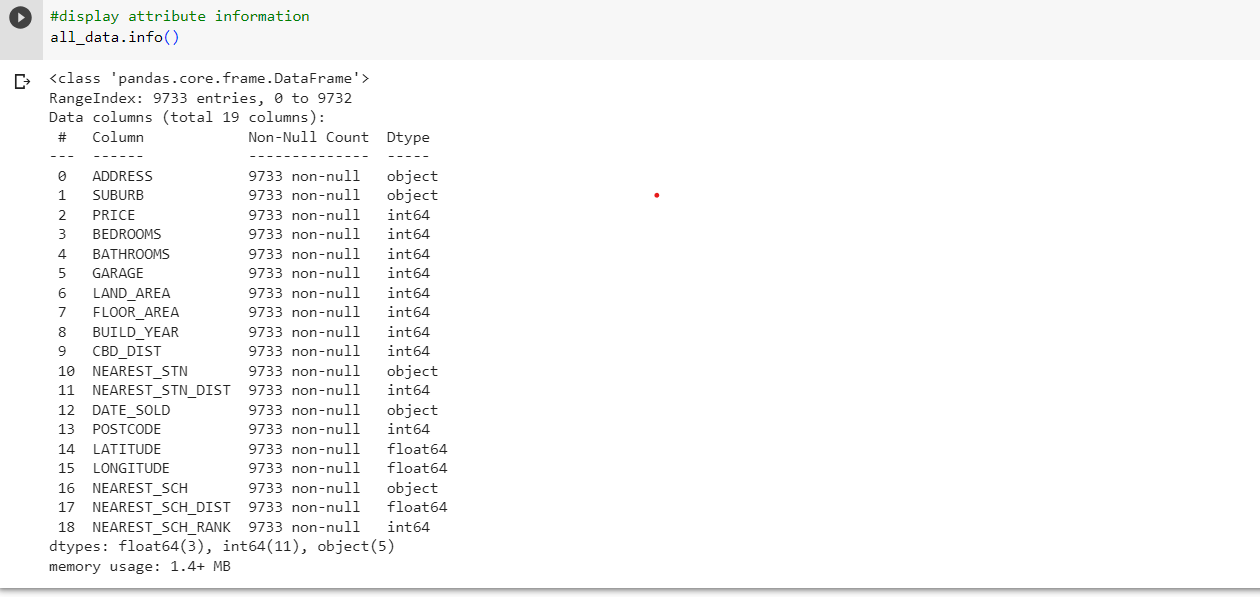


Figure 5 : Data Information

### 4.2.3 Data Cleaning

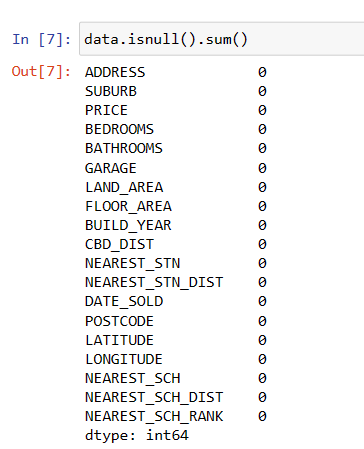
****

Figure 6 : The steps for null functions

The code snippet utilizes the isnull().sum() function to calculate and display the sum of missing values for each column in the dataset. This aids in identifying the extent of missing data, which is crucial for data quality assessment and potential data imputation or cleaning processes. The dataset contains no missing values, as indicated by the zero counts in each column. All 19 columns, including attributes like address, price, bedrooms, bathrooms, and geographical coordinates, have complete data. This data quality is essential for accurate and reliable analysis.

## 4.3 Exploratory Data Analysis (EDA)

### 4.3.1 Descriptive Statistics

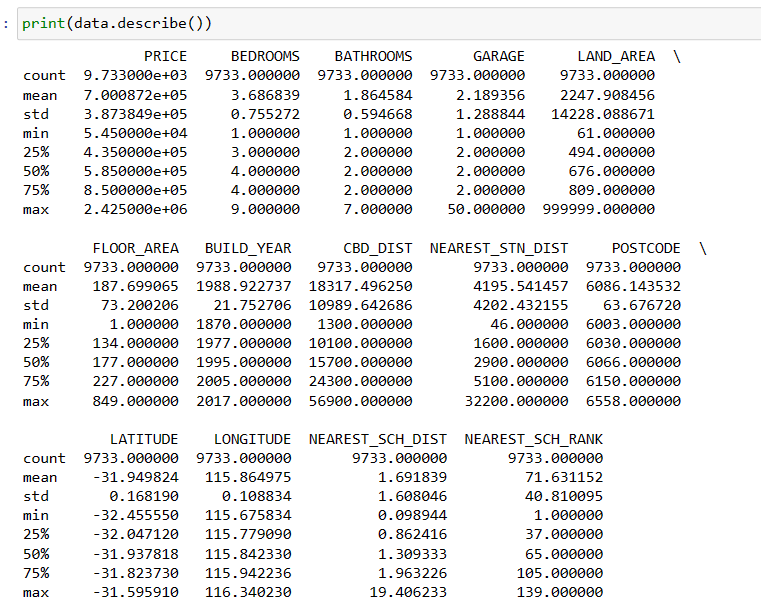
****

Figure 7 : The result of the descriptive statistics

The descriptive statistics summary of the dataset includes essential housing attributes such as price, bedrooms, bathrooms, garage, land and floor areas, build year, and distances. This summary provides an overview of the data distribution and key statistical measures. The summary statistics for various attributes in the dataset are presented. These include housing prices, bedroom and bathroom counts, garage spaces, land and floor areas, build year, distances to CBD and nearest train station, postcodes, geographical coordinates, nearest school distance, and ranking. The statistics provide insights into central tendencies, distributions, and variations of these attributes.

### 4.3.2 Data Distribution

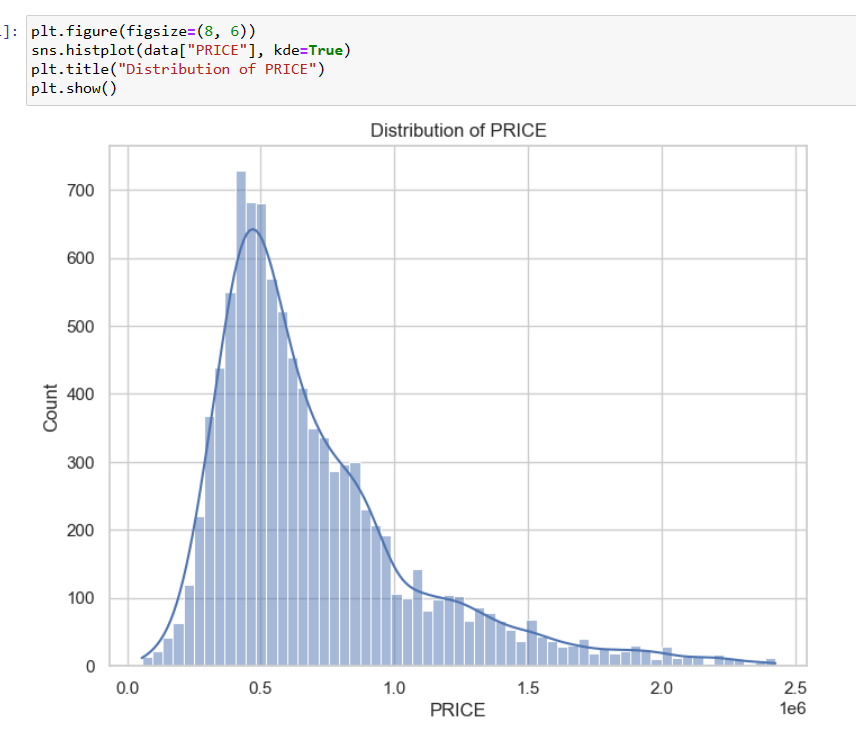
**`**

Figure 8 : The histogram plot

The histogram plot visualizes the distribution of housing prices (PRICE) in the dataset. The x-axis represents the price ranges, while the y-axis indicates the frequency of properties falling within each range. The plot showcases the spread of housing prices and highlights the concentration of properties within specific price intervals. Relating this visualization to the comprehensive analysis on Housing Prices in Perth, it provides a fundamental understanding of the price distribution across different property attributes and locations.

This graphical representation aids in identifying the typical price ranges for housing properties in Perth. By analysing the histogram, insights can be gained into whether most properties fall within a certain price bracket, potential outliers at higher price ranges, and the overall market segmentation based on housing prices. Furthermore, this histogram offers an essential facet of the larger analysis, which could encompass correlations between property attributes, geographical factors (like suburb or distance to key amenities), and their influence on housing prices.

### A colorful squares with white text Description automatically generated4.3.3 Correlation Analysis

Figure 9 : Correlation heatmap showing relationships between key variables.

A pivotal aspect of our exploratory data analysis involves understanding the correlations between various attributes and the 'PRICE' of housing properties in Perth. To achieve this, we have constructed a correlation heatmap that vividly illustrates the relationships between attributes and the target variable.

In the heatmap, attributes are represented along both the x and y axes, forming a grid of color-coded squares. The color intensity of each square reflects the strength and direction of correlation. Specifically, attributes are categorized based on their correlation coefficients into three groups: strong positive correlation, weak positive correlation, and no significant correlation.

Correlation Categories:

**Attributes with Correlation > 0.2 (Strong Positive Correlation):**

* 'BATHROOMS', 'BEDROOMS', 'FLOOR\_AREA'

These attributes exhibit a strong positive correlation with the 'PRICE' of housing properties. An increase in the values of these attributes is associated with higher housing prices. For instance, properties with more bathrooms, bedrooms, and larger floor areas tend to command higher prices.

**Attributes with Correlation ≥ 0.1 (Weak Positive Correlation):**

* 'GARAGE'

While the correlation for 'GARAGE' is weaker than the group, it still holds a positive relationship with housing prices. This suggests that properties with a garage might fetch slightly higher prices compared to those without one.

**Attributes with Correlation ≤ 0 (No Significant Correlation):**

* 'NEAREST\_SCH\_DIST', 'NEAREST\_STN\_DIST', 'BUILD\_YEAR', 'POSTCODE', 'LONGITUDE', 'CBD\_DIST', 'NEAREST\_SCH\_RANK', 'LATITUDE', 'LAND\_AREA'

These attributes exhibit either no significant correlation or correlations close to zero with the 'PRICE' of housing properties. The color shades in the heatmap for these attributes indicate a

lack of substantial influence on housing prices.

### 4.3.4 Geographic Analysis

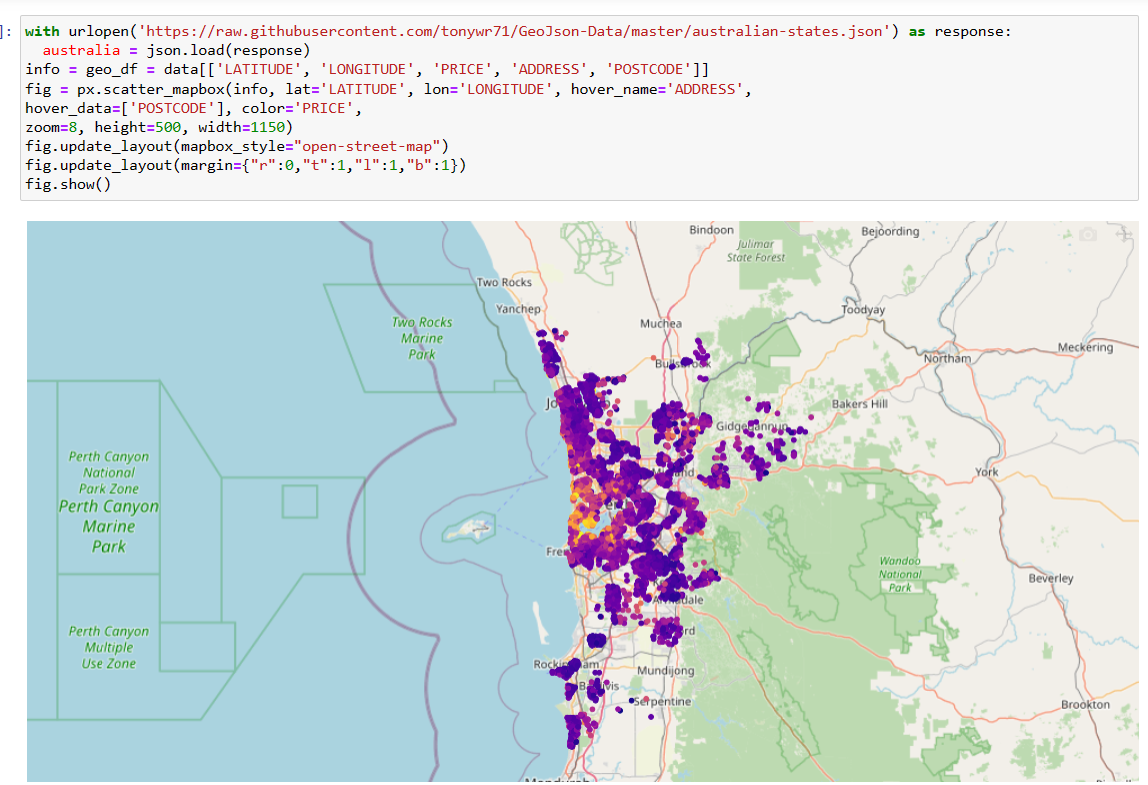
****

Figure 10 : The geographical representation

Geographical representation involves visually displaying data on a map to reveal the spatial distribution and patterns of specific variables, like housing prices, across a designated geographic area. The following points provide a summary of observations derived from map visualizations:

**Homes with prices above $1.5 million:** Concentrated close to the ocean and the central area of Perth. Suggests that higher property prices are linked with locations near the ocean and city centre.

**Houses under $0.5 million:** More commonly found on the right side of town or in areas farther from the central core of Perth. Indicates that more affordable housing tends to be situated in suburban or outlying regions.

**Price and population density**: Appears to exhibit a correlation between housing prices and population density. Areas with higher population density tend to demonstrate higher property prices. This connection might stem from heightened housing demand in densely populated regions, resulting in increased property prices.

## 4.4 Location Analysis

In this location analysis, we explore the impact of three key geographic factors - CBD distance, nearest station distance, and nearest school distance - on property prices. By examining how these distances relate to housing prices, we gain insights into the influence of location on real estate values.

A group of graphs showing different colored dots

Description automatically generated with medium confidence

Figure 11 : Scatter plot of Price Vs Nearest School, Station, and CBD distances

The scatter plots reveal a significant relationship between certain geographic factors and property prices. Specifically, properties located within CBD distances of less than 30,000 meters, station distances less than 15,000 meters, and school distances less than 5 meters are associated with higher prices.

This suggests that properties situated closer to the central business district (CBD), conveniently located near train stations, and in immediate proximity to schools tend to command higher market values. The close correlation between these key locations and higher prices underscores the influence of accessibility, convenience, and educational amenities on property valuation.

## 4.5 Property Attributes Analysis

The Property Attributes Analysis delves into the impact of various property-specific characteristics on housing prices in Perth. This section investigates how factors such as the number of bedrooms, bathrooms, garage spaces, land area, floor area, and the year of construction influence property prices.

**Land and Floor Area**

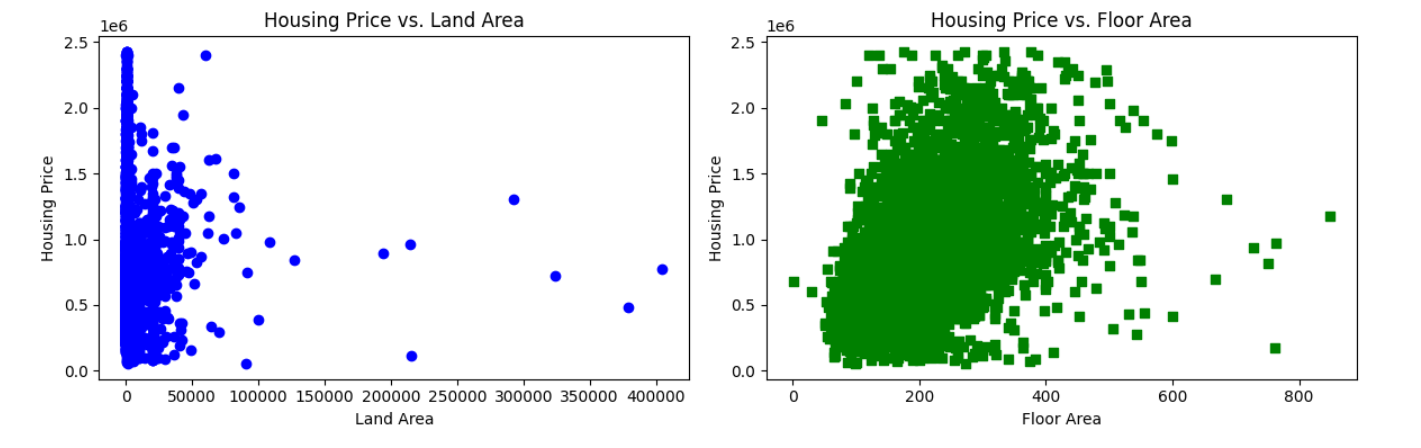


Figure 12 : Scatter plot of Price Vs Land and Floor Area

We generate two scatter plots using the matplotlib.py plot library. Each plot visualizes the relationship between housing prices and different features: land area and floor area. The blue circles on the first plot represent the correlation between housing prices and land area, while the green squares on the second plot represent the relationship between housing prices and floor area. This visualization aids in understanding whether there's a connection between these features and housing prices.

Moreover, based on additional insights, properties with less than 50,000 m2 of land are popular. This observation suggests that people might prefer properties with smaller land areas, possibly because they are easier to manage, especially in urban areas. Additionally, there is a preference for houses ranging from 100 to 400 square meters in size. This finding indicates that buyers are inclined towards homes that strike a balance between not being too small or too large.

**Bedrooms, Bathrooms, and Garage**

A screenshot of a graph

Description automatically generated

Figure 13 : Scatter plot of Price Vs number of Bathroom, Bedroom, and Garage

Properties featuring up to 5 garages are in high demand across different price ranges, showing that having enough parking is important to many buyers. Similarly, Houses with 1, 2, and 3 bathrooms experience increased demand across all price ranges, implying that buyers appreciate having bathroom options. Similarly, properties with 3, 4, and 5 bedrooms are highly desired in all price categories, suggesting that having sufficient bedrooms is important to potential buyers.

**Property Age**

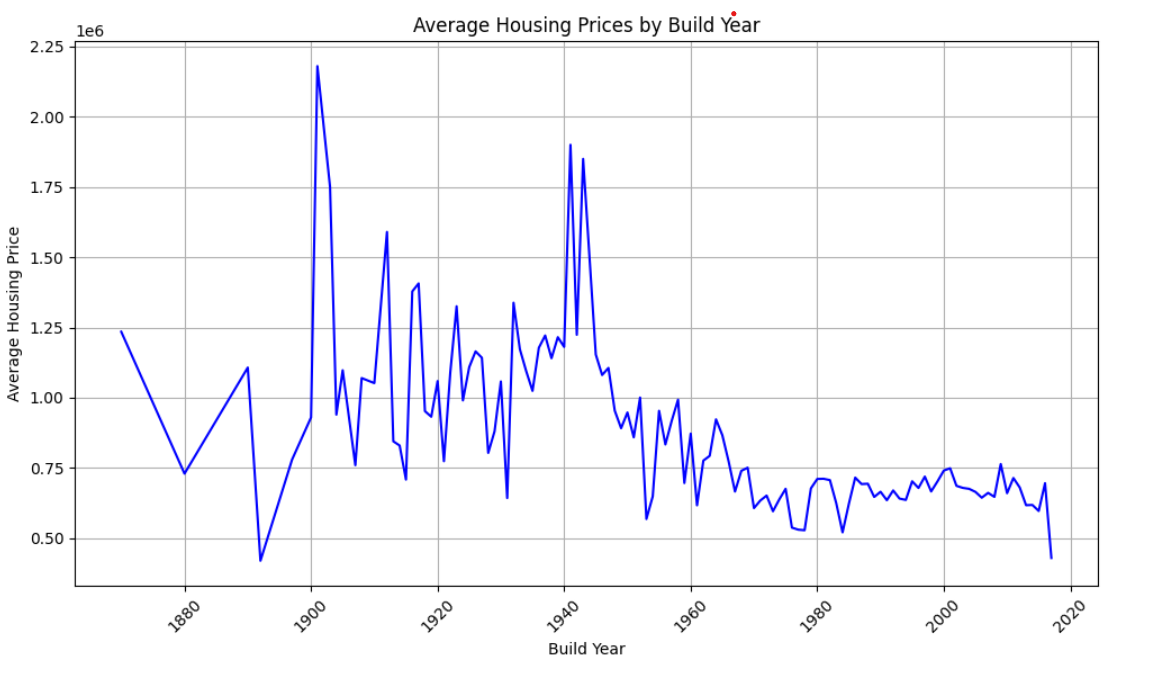
****

Figure 14 : Average house price by build year

The line graph provides a historical overview of average house prices spanning from 1880 to 2020. Noteworthy trends emerge as we analyze the data. The peak average house price occurred in approximately 1900, with values ranging from 2 to 2.25 million AUD. This period likely reflects a confluence of economic prosperity and housing demand. The subsequent years up to 1950 witnessed relatively high prices, maintaining a range between 2 million and 0.75 million. However, a noticeable shift began around 1950, leading to a gradual decline in average prices, eventually dipping to less than 0.5 million by 2020.

## 4.6 Model Comparison and Selection

In this section, we compare the performance of the Linear Regression, Random Forest Regression, and Gradient Boosting Regression models. The objective is to identify the model that best captures the relationships between property attributes and housing prices.

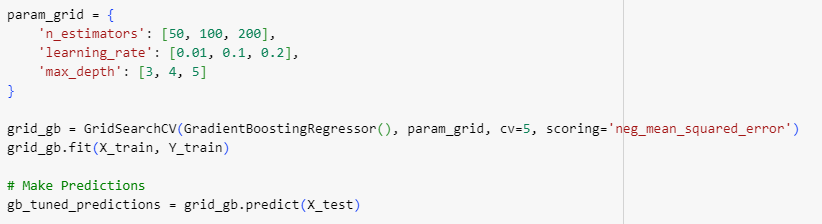
### 4.6.1 Hyperparameter Tuning

To enhance the performance of the machine learning models used for predicting housing prices, hyperparameter tuning is employed. Hyperparameters are configuration settings that control the behaviour of the models during training. The process of hyperparameter tuning involves systematically evaluating various combinations of hyperparameters to identify the ones that result in the most optimal model performance.

* **Linear Regression**

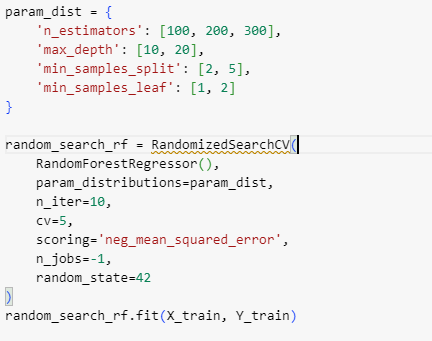
In the case of Linear Regression, a preprocessing step involving feature scaling is applied to ensure that the features have similar scales. The StandardScaler is fitted to the training data and then applied to both the training and test datasets. This standardized data is then used to train the Linear Regression model. Hyperparameter tuning is not typically performed for Linear Regression since it has minimal hyperparameters.

* **Gradient Boosting Regression**

For Gradient Boosting Regression, a grid of hyperparameters is defined. This grid includes values for n\_estimators, learning\_rate, and max\_depth. The GridSearchCV technique exhaustively searches through this grid to find the combination of hyperparameters that minimizes the negative mean squared error (neg\_mean\_squared\_error) on a cross-validated basis. This approach ensures that the selected hyperparameters result in the best possible predictive performance.

* **Random Forest Regression**

Similarly, for Random Forest Regression, a randomized search over a defined range of hyperparameters is conducted using the RandomizedSearchCV technique. The parameters considered include n\_estimators, max\_depth, min\_samples\_split, and min\_samples\_leaf. RandomizedSearchCV performs a random search through the defined parameter space, evaluating the model's performance using cross-validation. This approach helps identify hyperparameters that lead to the optimal performance in terms of mean squared error.

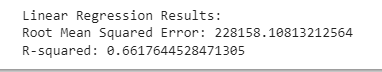


### 4.6.2 Evaluation

In the research we used two vital performance metrics: Root Mean Squared Error (RMSE) and R-squared (R2). RMSE quantifies the average prediction error's magnitude, indicating how closely model predictions align with actual prices. On the other hand, R2 measures the proportion of variance in housing prices explained by the models, providing insights into their explanatory power. These metrics collectively allow us to assess the accuracy and effectiveness of predictive models.

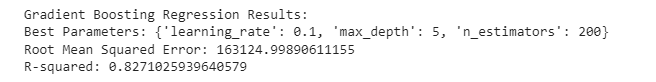
**Linear Regression Model**

In the Linear Regression model. The calculated RMSE for this model is 228,158.11, indicating the average difference between the predicted housing prices and the actual prices. The R2 value of 0.662 suggests that the linear regression model explains approximately 66.2% of the variance in housing prices. While this model provides a baseline understanding of the data, its relatively higher RMSE and moderate R2 indicate room for improvement.



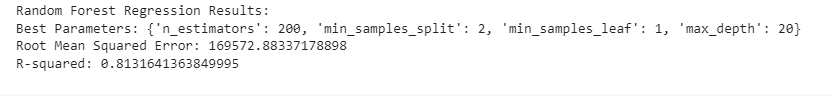
**Gradient Boosting Regression Model**

In the Gradient Boosting Regression model, which has been fine-tuned using GridSearchCV to optimize its hyperparameters. The obtained parameters—learning rate of 0.1, max depth of 5, and 200 estimators—contributed to an improved RMSE of 163,125.00 and an impressive R2 of 0.827. These outcomes signify a notable enhancement over the linear regression model, with the gradient boosting approach capturing approximately 82.7% of the variance in housing prices.



**Random Forest Regression Model**

Similarly, we evaluated the Random Forest Regression model using RandomizedSearchCV to identify optimal hyperparameters. The selected parameters—200 estimators, minimum samples split of 2, minimum samples leaf of 1, and a maximum depth of 20—yielded an RMSE of 169,572.88 and an R2 of 0.813. These results position the random forest model between linear regression and gradient boosting in terms of performance.

****

The assessment highlights the strong predictive capabilities of the Gradient Boosting Regression model for housing prices in Perth. With the lowest RMSE and the highest R2 compared to the other models, gradient boosting delivers the most precise and reliable predictions. These results align with the overarching aim of thoroughly examining the influence of location and property characteristics on housing prices. The employment of cross-validation techniques played a pivotal role in ensuring unbiased evaluation and model comparison, thus instilling confidence in the conclusions drawn.

## 4.7 Interpretation and Visualization

In this section, we delve into the interpretation of the chosen Gradient Boosting Regression model's results. Additionally, we present visualizations that aid in understanding the relationships between property attributes and housing prices.

### 4.7.1 Feature Importance

To gain insights into the importance of each property attribute in predicting housing prices, we visualize the feature importance determined by the Gradient Boosting Regression model. The feature importance highlights the relative contribution of each attribute to the model's predictive power.

The following bar chart displays the feature importance:

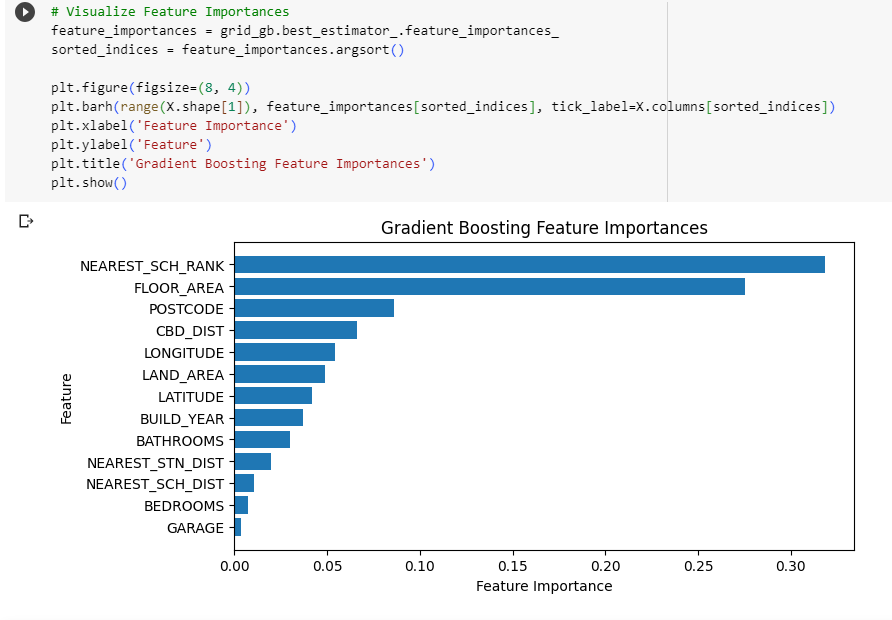


Figure 15: Gradient Boosting Feature Importance

The graph reveals those attributes such as 'Nearest School Rank,' 'Floor area,' ‘Postcode’, and

‘CBD distance’ hold significant importance in predicting housing prices, while other attributes contribute to a lesser extent. This information provides a clear understanding of which attributes play a pivotal role in influencing housing prices in the Perth area.

### 4.7.2 Predictions vs. Actual Prices

Visualizing the predictions made by the Gradient Boosting model against the actual housing prices offers a tangible comparison of the model's performance. The scatter plot below showcases this comparison.

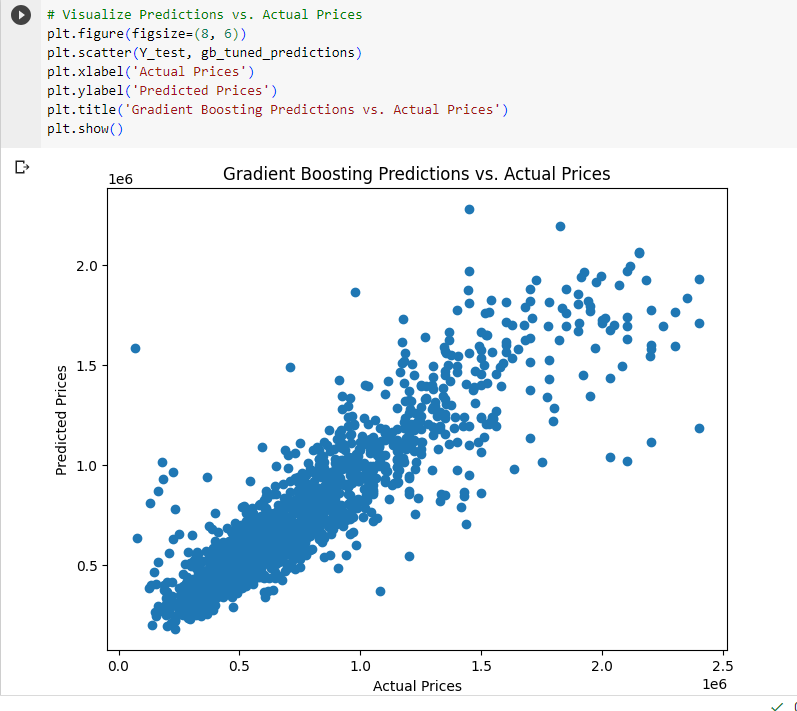


Figure 16: Gradient Boosting Predictions vs Actual Prices

The scatter plot demonstrates that the model's predictions align reasonably well with the actual prices. While there are deviations, particularly for higher-priced properties, the overall pattern indicates that the model captures the underlying trends in the data effectively.

The combination of feature importance and predictions versus actual prices allows us to draw insightful conclusions. We can affirm that attributes such as land area, build year, and distance to the nearest school are key drivers of housing prices in Perth. Furthermore, the scatter plot illustrates that the model's predictive performance is especially strong for properties within certain price ranges, highlighting its suitability for guiding pricing decisions and market analysis.

## 4.8 Deployment

Deploying the developed Gradient Boosting Regression model for predicting housing prices in Perth involves crucial steps. First, the model is serialized and saved for future use, preserving its parameters and structure. Incoming data undergoes preprocessing to match the model's format, including handling missing values and scaling features. Integration into an application or service allows users to input property attributes and receive predicted prices. Continuous monitoring ensures performance consistency, with periodic retraining to stay accurate. User-friendly outputs and interpretability, along with security measures and scalability considerations, enhance usability. Successful deployment hinges on technical, operational, and ethical factors, ensuring the model operates effectively, maintains privacy, and meets user needs.

# Chapter 5: Conclusion

## 5.1 Conclusion

In conclusion, this study delves into the dynamic interplay between property attributes and housing prices within the distinctive context of the Perth real estate market. An in-depth examination of diverse property features has revealed intricate insights that shape our understanding of property valuation dynamics. Attributes like bathrooms, bedrooms, floor area, and garage emerge as pivotal factors that exert significant influence on housing prices. Notably, larger and more feature-rich properties tend to command higher prices, underlining the significance of these attributes in the pricing equation.

The study's distinctiveness lies in its innovative approach of employing advanced modelling techniques, particularly the Gradient Boosting Regression. In contrast to initial linear correlation analysis, this technique uncovers nuanced non-linear relationships and interaction effects, offering a richer comprehension of the intricate relationships intrinsic to real-world data. The revealed feature importance through the Random Forest model underscores the predictive prowess of attributes such as nearest school rank, floor area, and postcode. These attributes, alongside others, significantly contribute to the model's accuracy in predicting housing prices.

With predictive performance at the forefront, the Gradient Boosting Regression emerges as the preferred model among those evaluated. This is evident through its lower Root Mean Squared Error (RMSE) and higher R-squared (R2) value, signifying its superior ability to encapsulate underlying patterns and relationships more adeptly. By adeptly capturing intricate property dynamics, the Random Forest model equips stakeholders with a potent tool for informed decision-making.

## 5.2 Recommendation

Derived from the findings, practical recommendations are offered to key stakeholders in the real estate landscape. Developers are advised to integrate attributes like nearest school rank and floor area into property designs to align properties with market preferences and gain a competitive edge. Prospective buyers and sellers can leverage insights from the analysis to make informed decisions based on preferences and financial considerations. Investors can capitalize on the quantified relationships presented in the study for informed investment choices.

Future research endeavours could consider integrating supplementary data sources, including economic indicators and demographic trends, to enhance the predictive accuracy of models. Furthermore, exploring temporal dynamics and evolving market trends would yield a deeper understanding of the complex housing landscape. By embracing these avenues, stakeholders can navigate the intricate real estate sector more strategically, utilizing insights to foster informed decision-making and fortify their positions in this dynamic domain.

## 5.3 Future scope

The study focusing on the impact of location and property attributes on housing prices in Perth holds promising prospects in the realms of both academia and real-world applications. As urbanization and population growth continue to shape the housing market, understanding the dynamics that drive property prices becomes increasingly valuable. Here are some potential future scopes for this research:

**Policy Formulation and Urban Planning:** The findings of this study can play a pivotal role in informing urban planning and policy formulation. Local governments and urban planners can utilize the insights to make informed decisions regarding zoning regulations, infrastructure development, and housing affordability initiatives. By strategically identifying areas of high growth potential, policy interventions can be designed to create sustainable and equitable urban landscapes.

**Real Estate Investment Strategies**: Real estate investors and developers stand to benefit significantly from the outcomes of this research. The study’s insights can guide investment decisions by highlighting areas with the most potential for value appreciation based on location and property attributes. This could lead to more efficient allocation of resources and reduced investment risk.

**Predictive Analytics and Machine Learning:** The use of advanced predictive analytics and machine learning models can be employed to forecast housing price trends based on the identified factors. Future researchers can refine and develop more accurate models that consider additional variables, thus enhancing the predictive power and reliability of such analyses.

**Socioeconomic Implications:** Exploring the impact of housing prices on socioeconomic factors, such as income distribution and gentrification, presents an intriguing avenue for future research. Understanding how housing prices affect different segments of the population can help policymakers address issues of housing affordability and social equity.

**Temporal Analysis:** This study’s framework can be extended to include a temporal dimension, analysing how the relationship between location, property attributes, and housing prices evolves over time. This longitudinal perspective would provide valuable insights into market trends and cyclical patterns.

**Comparative Studies:** Conducting similar analyses across different cities or regions allows for comparative assessments of the impact of location and property attributes on housing prices. Such cross-regional studies could uncover insights into the role of local versus global factors in shaping property markets.

**Data Integration:** As data availability and integration techniques improve, future research could incorporate a broader range of data sources, including social media sentiment, environmental factors, and mobility patterns. Integrating these diverse datasets could lead to a more comprehensive understanding of housing price dynamics.

**Behavioural Economics:** Incorporating behavioural economics principles could shed light on the psychological and emotional factors that influence buyers’ decisions. Understanding the role of perceptions, sentiments, and cognitive biases could enhance the accuracy of predictive models. In conclusion, the exploration of the impact of location and property attributes on housing prices in Perth lays the foundation for a multitude of future research directions. By delving deeper into these avenues, researchers can contribute to both theoretical advancements in the field of urban economics and practical solutions for the challenges faced by policymakers, investors, and communities in the evolving housing market landscape.

# References

Allegrante, J.P. and Sleet, D.A., 2021. Investing in public health infrastructure to address the complexities of homelessness. *International Journal of Environmental Research and Public Health*, *18*(16), p.8887.

Bangura, M. and Lee, C.L., 2020. House price diffusion of housing submarkets in Greater Sydney. *Housing Studies*, *35*(6), pp.1110-1141.

Bayliss, K., Van Waeyenberge, E. and Bowles, B.O., 2023. Private equity and the regulation of financialised infrastructure: the case of Macquarie in Britain’s water and energy networks. *New Political Economy*, *28*(2), pp.155-172.

Cloyne, J., Huber, K., Ilzetzki, E. and Kleven, H., 2019. The effect of house prices on household borrowing: A new approach. *American Economic Review*, *109*(6), pp.2104-2136.

Costello, G., Leishman, C., Rowley, S. and Watkins, C., 2019. Drivers of spatial change in urban housing submarkets. *The Geographical Journal*, *185*(4), pp.432-446.

Debelle, G., 2019, October. Housing and the Economy. In *CFA Societies Australia Investment Conference*.

Diamond, R. and McQuade, T., 2019. Who wants affordable housing in their backyard? An equilibrium analysis of low-income property development. *Journal of Political Economy*, *127*(3), pp.1063-1117.

Feng, X. and Humphreys, B., 2018. Assessing the economic impact of sports facilities on residential property values: A spatial hedonic approach. *Journal of Sports Economics*, *19*(2), pp.188-210.

Ficek, R.E., 2018. Infrastructure and colonial difference in Puerto Rico after Hurricane María. *Transforming Anthropology*, *26*(2), pp.102-117.

Gharahighehi, A., Pliakos, K. and Vens, C., 2021. Recommender systems in the real estate market—a survey. *Applied Sciences*, *11*(16), p.7502.

Ibraeva, A., de Almeida Correia, G.H., Silva, C. and Antunes, A.P., 2020. Transit-oriented development: A review of research achievements and challenges. *Transportation Research Part A: Policy and Practice*, *132*, pp.110-130.

Kang, Y., Zhang, F., Peng, W., Gao, S., Rao, J., Duarte, F. and Ratti, C., 2021. Understanding house price appreciation using multi-source big geo-data and machine learning. *Land Use Policy*, *111*, p.104919.

Shearer, C., 2000. The CRISP-DM model: the new blueprint for data mining. Journal of

data warehousing, 5(4), pp.13-22.

Karahan, F. and Rhee, S., 2019. Geographic reallocation and unemployment during the Great Recession: The role of the housing bust. *Journal of Economic Dynamics and Control*, *100*, pp.47-69.

Kendall, R. and Tulip, P., 2018. The effect of zoning on housing prices. *Reserve Bank of Australia Research Discussion Paper*, (2018-03).

Li, J., Hu, Y. and Liu, C., 2020. Exploring the influence of an urban water system on housing prices: Case study of zhengzhou. *Buildings*, *10*(3), p.44.

Lieber, J.K., 2022. A hedonic pricing model in Helsinki, Finland: exploring the impacts of green infrastructure on apartment listing prices.

Liu, S., Zhang, F. and Wu, F., 2022. Contrasting migrants’ sense of belonging to the city in selected peri-urban neighbourhoods in Beijing. *Cities*, *120*, p.103499.

Lowies, B., Squires, G., Rossini, P. and McGreal, S., 2022. Locked-out: generational inequalities of housing tenure and housing type. *Property Management*, *40*(4), pp.510-526.

Ma, L., Reed, R. and Liang, J., 2019. Separating owner-occupier and investor demands for housing in the Australian states. *Journal of Property Investment & Finance*, *37*(2), pp.215-232.

Massey, D.S. and Tannen, J., 2018. Suburbanization and segregation in the United States: 1970–2010. *Ethnic and racial studies*, *41*(9), pp.1594-1611.

Nethercote, M., 2020. Build-to-Rent and the financialization of rental housing: future research directions. *Housing studies*, *35*(5), pp.839-874.

Nozhati, S., Rosenheim, N., Ellingwood, B.R., Mahmoud, H. and Perez, M., 2019. Probabilistic framework for evaluating food security of households in the aftermath of a disaster. *Structure and Infrastructure Engineering*, *15*(8), pp.1060-1074.

Shi, S., Rahman, A., and Wang, B.Z., 2020. Australian housing market booms: Fundamentals or speculation? *Economic Record*, *96*(315), pp.381-401.

Su, S., He, S., Sun, C., Zhang, H., Hu, L. and Kang, M., 2021. Do landscape amenities impact private housing rental prices? A hierarchical hedonic modelling approach based on semantic and sentimental analysis of online housing advertisements across five Chinese megacities. *Urban Forestry & Urban Greening*, *58*, p.126968.

Wei, C., Fu, M., Wang, L., Yang, H., Tang, F. and Xiong, Y., 2022. The research development of hedonic price model-based real estate appraisal in the era of big data. *Land*, *11*(3), p.334.

Wimark, T., Haandrikman, K. and Nielsen, M.M., 2019. Migrant labour market integration: The association between initial settlement and subsequent employment and income among migrants. *Geografiska Annaler: Series B, Human Geography*, *101*(2), pp.118-137.

Wittowsky, D., Hoekveld, J., Welsch, J. and Steier, M., 2020. Residential housing prices: impact of housing characteristics, accessibility and neighbouring apartments–a case study of Dortmund, Germany. *Urban, Planning and Transport Research*, *8*(1), pp.44-70.

Wu, C., Ren, F., Hu, W. and Du, Q., 2019. Multiscale geographically and temporally weighted regression: Exploring the spatiotemporal determinants of housing prices. *International Journal of Geographical Information Science*, *33*(3), pp.489-511.

Yang, L., Chen, Y., Xu, N., Zhao, R., Chau, K.W. and Hong, S., 2020. Place-varying impacts of urban rail transit on property prices in Shenzhen, China: Insights for value capture. *Sustainable Cities and Society*, *58*, p.102140.

Yang, L., Chu, X., Gou, Z., Yang, H., Lu, Y. and Huang, W., 2020. Accessibility and proximity effects of bus rapid transit on housing prices: Heterogeneity across price quantiles and space. *Journal of Transport Geography*, *88*, p.102850.

Yang, L., Wang, B., Zhou, J. and Wang, X., 2018. Walking accessibility and property prices. *Transportation research part D: transport and environment*, *62*, pp.551-562.

# Appendices

## Appendix A: Glossary of Terms

|  |  |
| --- | --- |
| AUD | Australian Dollar |
| GR | Geographic Reallocation |
| BRT | Bus Rapid Transit |
| STN | Station |
| SCH | School |
| CBD | Central Business District |
| RMSE | Root Mean Squared Error |
| R2 | R-squared |
| M2 | Meter Square |
| URL | Uniform Resource Locator |
| EDA | Exploratory Data Analysis |
| IDE | Integrated Development Environment |
| CRISP-DM | Cross-Industry Standard Process for Data Mining |
| SVM | Support Vector Machines |

## Appendix B: Import necessary library for the task

import pandas as pd

import numpy as np

#data visualization libaries

import matplotlib.pyplot as plt

import seaborn as sns

import plotly.express as px

import plotly.graph\_objects as go

from plotly.subplots import make\_subplots

from urllib.request import urlopen

import json

#model building

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import mean\_squared\_error, r2\_score

from sklearn.ensemble import RandomForestRegressor

from sklearn.model\_selection import train\_test\_split, GridSearchCV

from sklearn.model\_selection import RandomizedSearchCV

from sklearn.metrics import classification\_report

from sklearn.ensemble import GradientBoostingRegressor

from sklearn.feature\_selection import SelectFromModel

from sklearn.preprocessing import StandardScaler