



FELLOWSHIP PROGRAM ASSIGNMENT COVER SHEET

THIS FORM MUST BE AT THE FRONT OF EACH ASSIGNMENT

CANDIDATES MUST KEEP A COPY OF THEIR ASSIGNMENT

Candidate to complete the following section (and update details in header and footer):

CANDIDATE NUMBER: 222xxx	COURSE: Data Analytics Applications
DATE DUE: Friday, 23 September 2022 at 12:00pm (AEST)	

PLAGIARISM

By submitting your assignment, you are implicitly stating that the work is your own.

Remember that an important aspect of being a professional actuary is to always act with integrity.

Committing plagiarism by copying another person's work or not properly referencing other sources used in your assignment is a breach of the Integrity principle under the Actuaries Institute's Code of Conduct.

Be aware that your assignment may be vetted using Turnitin.



Question 1: Domain knowledge

Word count excluding headings, charts and references: 1490

Prior to building my model for predicting property sales prices, I took steps to acquire knowledge about the real estate industry with a focus on understanding the key drivers of property prices. Following this, I assessed how well these drivers were captured in my dataset.

I applied **two main strategies** to obtain this knowledge:

- Conducting **initial research** using research papers and real-estate websites
- **‘Walking in the shoes’** of a prospective property buyer by looking at property marketplaces such as Domain, and conducting market research.

I identified five key drivers of property sales prices, which will be discussed in more detail below.

- Location
- Property type
- Property features
- Property age and condition
- Supply and demand

1.1 Location

Impact on property sales prices

A key feature of residential property is its immovability which forms the basis of location being an important determinant of property prices (Coffee et al. 2013).

The location of a property strongly influences its sales price as buyers are willing to pay a price to access its associated desirable features (or avoid undesirable features). In particular:

- Location determines the property’s accessibility and proximity to amenities such as shopping centres, schools, recreational facilities and public transport. Most buyers are willing to pay more for the convenience of easier access to these local amenities, which increases the property value (Musa & Yusoff 2015).
- Property price generally increases with proximity to natural amenities such as parks, beaches and water views (Phibbs & Gurran 2021).
- Property price generally decreases in areas of lower socioeconomic status or with exposure to negative environmental externalities such as flood risk, air pollution, noise from traffic and high crime rates (Treg 2010).

I confirmed through my **market research** on Australian properties that property sales prices vary substantially by state (**Figure 1.1**) and by suburb (**Figure 1.2**).

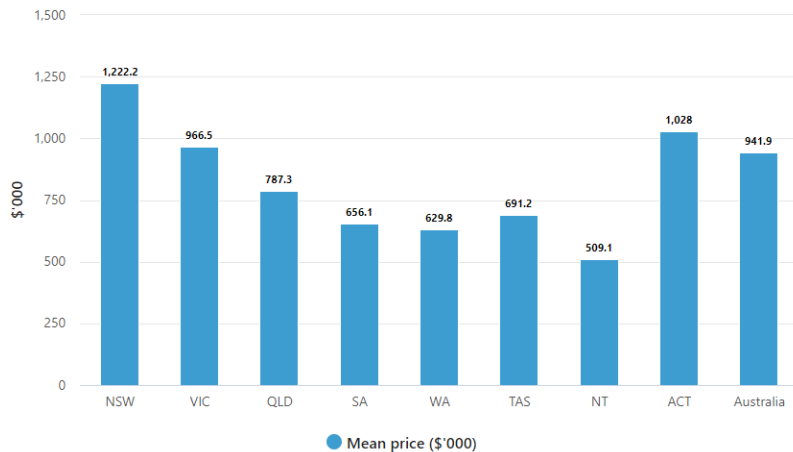


Figure 1.1 Average property price across Australian states and territories (March quarter 2022), **source:** ABS 2022

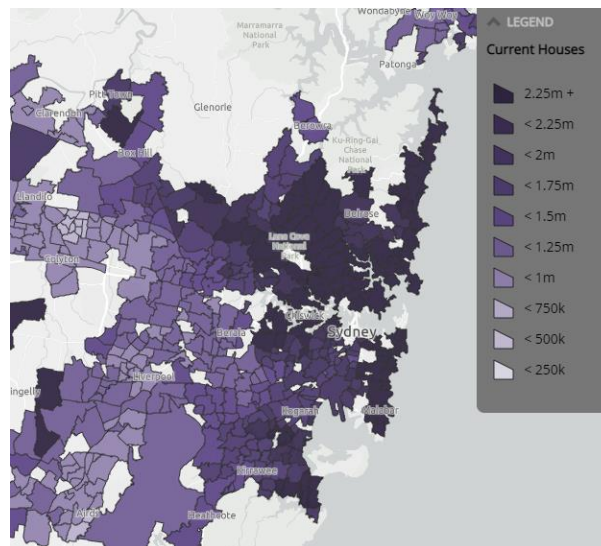


Figure 1.2 Average house price in Sydney by suburb (July 2022), **source:** CoreLogic 2022

How well it is captured by the dataset

The dataset contains three variables which relate to property location – address, suburb and postcode. Together, these fields identify the exact location of each property and therefore, location is **well captured** by the dataset.

Some limitations include:

- The dataset does not include state, however this can be identified from postcode.
- The dataset does not explicitly contain information relating to locational attributes such as distance to the central business district or crime rates, which also impact price. These are sometimes mentioned in the ad heading or ad body but data cleaning and processing is required to extract this information, and it is not mentioned for all properties. However, the address field is granular enough that these additional location-related attributes can be derived with the support of external data sources if needed.



Sources

ABS 2022, *Total Value of Dwellings*, accessed 6 August 2022,
<<https://www.abs.gov.au/statistics/economy/price-indexes-and-inflation/total-value-dwellings/latest-release>>

Coffee, NT, Lockwood, T, Hugo G, Paquet, C, Howard, NJ & Daniel, M 2013, *Relative residential property value as a socio-economic status indicator for health research*, accessed 7 August 2022,
<<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3637187/>>

CoreLogic 2022, *Mapping the Market*, accessed 7 August 2022,
<<https://www.corelogic.com.au/our-data/mapping-market>>

Musa, U & Yusoff, WZW 2015, *Impact of Location and Dwelling Characteristics on Residential Property Prices/ Values: A Critical Review of Literature*, accessed 7 August 2022,
<<https://www.managementjournal.info/index.php/IJAME/article/download/409/349>>

Phibbs, P & Gurran, N 2021, *The role and significance of planning in the determination of house prices in Australia: Recent policy debates*, accessed 7 August 2022,
<<https://journals.sagepub.com/doi/full/10.1177/0308518X21988942>>

Treg, C 2010, *A Multilevel Property Hedonic Approach to Valuing Parks and Open Space*, accessed 6 August 2022,
<<https://scholarworks.uvm.edu/cgi/viewcontent.cgi?article=1229&context=graddis>>



1.2 Property type

Impact on property sales prices

Property type has a significant impact on property sales prices. For example, in the Australian real estate market, there are five main residential property types:

- Stand-alone houses
- Semi-detached houses, e.g. duplexes
- Terrace houses
- Townhouses
- Apartments or Units

Sales prices vary by property type for many reasons, and I have listed some below:

- **Right to land ownership:** The price of standalone houses includes the value of the land and dwelling value (Home Loan Experts 2019), however other property types such as apartments and townhouses are usually strata-titled meaning that land ownership is shared by multiple people (Hadgelias 2019). The inclusion of land ownership on an individual title increases property sales prices as land typically appreciates in value (Butkovich 2019).
- **Size:** Each property type has their own typical size range. Houses are typically the largest, followed by semi-detached houses, townhouses and apartments. Property prices also tend to follow this order.
- **Flexibility to renovate:** Stand-alone houses are owned on a single title, and therefore there is more flexibility for renovations, extensions or a rebuild which enhances the property value (Hadgelias 2019). Conversely, apartments and townhouses usually belong to a body corporate which limits the potential for capital gains through renovations.

To confirm the effect of property type on sales price, I conducted **market research** on Domain across many suburbs, with Chatswood included as an example in **Figure 1.3**. It is clear that houses have a higher median price than units, even when they have the same number of bedrooms.

Market trends

View median property prices in Chatswood to get a better understanding of local market trends.

BEDROOMS	TYPE	MEDIAN PRICE	AVG DAYS ON MARKET	CLEARANCE RATE	SOLD THIS YEAR	
2	House	-	-	-	7	+
3	House	\$2.765m	63 days	68%	41	+
4	House	\$3.31m	57 days	58%	30	+
5	House	\$3.638m	106 days	81%	30	+
1	Unit	\$786k	55 days	43%	62	+
2	Unit	\$1.2m	45 days	53%	110	+
3	Unit	\$1.998m	94 days	66%	35	+

* Data based on sales within the last 12 months

Figure 1.3 Impact of property type on median property price in Chatswood, **source:** Domain 2022



How well it is captured by the dataset

The dataset captures property type **very well**, as it is explicitly defined in the field 'PropertyType'. The categories are reasonably granular, and have separated out the five residential property types listed above. This means we will be able to use property type to predict price.

The dataset also includes non-residential property types such as warehouse, livestock and cropping. This suggests that although property type is well captured in the dataset, some of them are not relevant for modelling as the app will mainly be targeted to residential property buyers.

Sources

Butkovich, D 2019, *House vs. apartment: Four key things you should consider before purchasing your first home*, accessed 7 August 2022, < <https://www.domain.com.au/advice/house-vs-apartment-four-key-things-you-should-consider-before-purchasing-your-first-home-807988/>>

Domain 2022, *Chatswood NSW 2067*, accessed 22 September 2022, <<https://www.domain.com.au/suburb-profile/chatswood-nsw-2067>>

Hadgelias, G 2019, *Townhouse vs house vs apartment: which is a better investment?*, accessed 7 August 2022, < <https://raywhitepaddington.com.au/news/townhouse-vs-house-vs-apartment-which-is-a-better-investment>>

Home Loan Experts 2019, *Units Vs Houses: Which is a better investment?*, accessed 7 August 2022, <<https://www.homeloanexperts.com.au/property-investment-tools/units-vs-houses-investment/>>



1.3 Property features

Impact on property sales prices

Structural attributes such as the number of bedrooms and other rooms, number of bathrooms, number of car spaces, size and number of floors are important features that affect price as they impact the usability of the property space. (Musa & Yusoff 2015). Other desirable attributes such as fences, gardens, swimming pools and outdoor entertainment areas will also increase property prices (Musa & Yusoff 2015).

To confirm the effect of property features on sales price, I conducted **market research** on Domain across many suburbs, with Glen Waverley included as an example in **Figure 1.4**. It is clear that for both houses and units, the median price increases with the number of bedrooms.

Market trends

View median property prices in Glen Waverley to get a better understanding of local market trends.

BEDROOMS	TYPE	MEDIAN PRICE	AVG DAYS ON MARKET	CLEARANCE RATE	SOLD THIS YEAR	
2	House	\$1.164m	145 days	40%	26	+
3	House	\$1.501m	32 days	80%	188	+
4	House	\$1.564m	40 days	73%	226	+
5	House	\$2.05m	62 days	65%	88	+
1	Unit	\$514k	-	-	18	+
2	Unit	\$703k	102 days	73%	90	+
3	Unit	\$946k	45 days	74%	52	+

* Data based on sales within the last 12 months

Figure 1.4 Impact of number of bedrooms on median property price in Glen Waverley, **source:** Domain 2022

How well it is captured by the dataset

This driver is **reasonably well** captured by the dataset through the following fields:

- Beds
- Bathrooms
- Car spaces
- Ad heading
- Ad body

The most fundamental property features – number of bedrooms, bathrooms and car spaces – are very well captured as they are explicitly recorded in separate fields and are filled in for majority of the dataset.



Other features such as “garden” and “swimming pool” were found in the ad heading and/ or ad body through a quick text search. However, it is more difficult to extract these as rigorous data cleaning and text processing would be required, and a property may also have features that are not mentioned at all in the ad heading or body. Therefore, these other features are not captured well in the dataset.

Sources

Domain 2022, *Glen Waverley VIC 3150*, accessed 22 September 2022,
<<https://www.domain.com.au/suburb-profile/glen-waverley-vic-3150>>

Musa, U & Yusoff, WZW 2015, *Impact of Location and Dwelling Characteristics on Residential Property Prices/ Values: A Critical Review of Literature*, accessed 7 August 2022,
<<https://www.managementjournal.info/index.php/IJAME/article/download/409/349>>



1.4 Property age and condition

Impact on property sales prices

All else equal, older properties tend to sell for lower prices due to physical wear and tear and potential functional obsolescence arising from missing or outdated home features (Shultz 2018). Their lower sales price may also be due to higher annual maintenance and repair costs in addition to less energy efficient heating and cooling systems (Treg 2010).

However, the effect of property age on sales price is not completely straightforward as it also depends on its condition. For example, the rate of physical depreciation varies for different properties depending on the quality of construction and materials used (Shultz 2018). It may also be partially offset by periodic maintenance and renovations that keep the property in good condition and help retain its value (Shultz 2018).

A property's age may also incorporate a "vintage effect" whereby a particular desirable characteristic such as building quality is correlated with the year in which it was built (Goodman & Thibodeau 1995). These vintage effects generally have a positive impact on property price.

How well it is captured by the dataset

This driver is **not well captured** by the dataset as it does not contain any variables that explicitly relate to age or condition.

Age and condition are occasionally referred to in the ad heading and ad body through descriptions such as "new" and "vintage", however these sorts of descriptions can be subjective and are not mentioned consistently across enough properties to provide meaningful value.

Sources

Goodman, A & Thibodeau TG, 1995, *Age-Related Heteroskedasticity in Hedonic House Price Equations*, accessed 13 August 2022, <https://www.researchgate.net/profile/Allen-Goodman/publication/241145211_Age-Related_Heteroskedasticity_in_Hedonic_House_Price_Equations/links/02e7e5319e307939cc000000/Age-Related-Heteroskedasticity-in-Hedonic-House-Price-Equations.pdf?origin=publication_detail>

Shultz, S 2018, *Housing Depreciation Revisited: Hedonic Price Modelling Versus Assessor Estimates*, accessed 13 August 2022, <https://www.unomaha.edu/college-of-business-administration/center-for-real-estate-and-asset-management/research/shultz_2018_jhr_housing_depreciation.pdf>

Treg, C 2010, *A Multilevel Property Hedonic Approach to Valuing Parks and Open Space*, accessed 6 August 2022, <<https://scholarworks.uvm.edu/cgi/viewcontent.cgi?article=1229&context=graddis>>



1.5 Supply and demand

Impact on property sales prices

Property prices are largely influenced by demand and supply factors. According to economic theory, prices will rise when demand is higher relative to supply. Demand for property is driven by both owner-occupiers and investors.

- In the short-term, demand is largely influenced by interest rates. All else equal, lower interest rates result in reduced mortgage repayments which means that prospective buyers will be able to borrow more at a given repayment-to-income ratio (Rahman 2008). This generally results in an increase in property demand.
- In the long-term, demand is more driven by population growth and underlying average household size. Increased demand results from high population growth and reduced average household size (Kohler & Merwe 2015).

The impact of changes in demand on property sales prices depends on how supply changes in response. Typically, there is a lag in property market supply to respond to increases in demand given the length and complexity of property planning and construction (Kohler & Merwe 2015). When supply falls short of demand, price will usually increase until sufficient supply becomes available. Conversely, an oversupply of properties could also result in lower property prices.

How well it is captured by the dataset

This driver is **not well captured** by the dataset. This is because the dataset only contains properties sold in a three-month period (April to June 2021). This is a short period of time in the context of the real estate market environment and is not sufficient to observe material changes in supply and demand. There is some supply and demand information implicitly captured by the 'AdPrice' field as the advertised price (if available) will most likely reflect the market environment at that time.

Sources

Kohler, M & Merwe, M 2015, *Long-run Trends in Housing Price Growth*, accessed 13 August 2022, <<https://www.rba.gov.au/publications/bulletin/2015/sep/3.html>>

Rahman MM 2008, *Australian Housing Market: Causes and Effects of Rising Price*, accessed 13 August 2022, <https://eprints.usq.edu.au/4614/2/Rahman_2008.pdf>



Question 2: Deployment decisions

Word count excluding headings: 966

The following deployment decisions should be considered when planning the app roll-out:

2.1 Where the data will be sourced from

The web app needs a data source for property listings. We need to make decisions regarding **which real estate website(s) to use** and the **number of websites to source from**:

- The key advantage of using more websites is that it provides a more complete view of the listings available. However, it also comes with some complications:
 - We will need to collate the data from multiple sources. This requires effort to clean the data so that it is in one consistent format for the model to use. There also needs to be a way to remove duplicate entries if properties are listed on more than one website.
 - We will need to obtain permission to access data from more data providers. This is likely to involve more time and monetary cost.
- It is worth considering starting off with one major reputable real estate website that captures a sizeable portion of the market. We could expand to other data sources later once the app is more established.

2.2 How the data will be sourced and processed

Decisions need to be made regarding the **way that data will be obtained** by the web app and **how this data will be processed** using the model. This directly affects the end-user experience.

The two main options are:

- **On-demand processing:** every time a user makes a search, the app will retrieve all property listings from the user's desired postcodes from the real estate website(s) and immediately process it through the model to make predictions. It then determines the set of likely listings within the user's price range.
- **Batch processing:** the list of properties is obtained and processed periodically in batches. Every time the user makes a search, the app will obtain the relevant output from this pre-processed dataset.

Comparing the two options:

- On-demand processing will provide more up-to-date information.
 - For batch processing, the frequency of processing (e.g. daily) will need to be decided. If this is not done regularly enough, then users may miss out on some property listings which is not ideal. However, processing too frequently will unnecessarily use up resources to repeatedly obtain and process large datasets.
 - On-demand processing may have a poorer user experience (i.e. feel slower to the user) since more processing needs to be done every time the user makes a search.
-



- Batch processing may require more data to be obtained, as we need to obtain **all** listings in each batch and not just those which are in our users' interested postcodes. There may be cost implications depending on our arrangements with the data providers.

In both options above, the data would need to be obtained in a timely and efficient manner. This is likely to involve using an "API", which is a mechanism that allows computer software to transmit data in a structured and automated way.

2.3 How the model should be tested internally

Decisions need to be made regarding the types of testing to be performed in preparation for the deployment of the app. For example:

- We could test **how well the app handles a heavy load**. This is to prepare it for higher-than-expected usage when the app is rolled out, such as if many users are searching for properties at the same time. This is important to make sure that the user experience of the app is not compromised under a heavy load.
- We could test **whether data can be correctly retrieved from real estate websites**. This is particularly relevant if the web app sources from multiple real estate websites and in this case, we would also test whether the separate data sources can be merged together correctly.
- We could test **how secure the app is** and whether it has any security loopholes that might allow the public to gain unauthorised access to our systems. This includes gaining unauthorised access to the underlying predictive model that will be powering the app.

2.4 How the model will be deployed

The model deployment strategy should be decided.

- The web app can be **released to all users at once**. This is generally undesirable because while it allows greater usage of the app from release date, any issues that were undetected in the testing phase will impact all users which may damage the reputation of the business.
- A better alternative is to **progressively roll-out the app to users**. This limits the impact of any issues to small group and can give stakeholders confidence if it is successful on the subset of users.
 - A decision needs to be made regarding which subset of users it will initially be rolled-out to. For example, if there are existing real-estate clients then they could be asked to trial and give feedback.
 - Once the app appears to be working as intended, it can be progressively rolled-out to the remainder of the users.

2.5 How the model will be monitored

It is important to decide on the metrics that will be used to monitor the performance of the model. These metrics can help inform whether the model is good enough to deploy to all users, or if further refinement can be made even after it is widely deployed. Some examples include:



- **Comparison of predicted versus actual sold price:** When properties listed on the web app are sold, their predicted sales price band can be compared to the actual sales price to evaluate how well the model is performing. This form of monitoring will need to take place over a long-term as properties are sold over time. In particular, it will be important to ensure the model remains accurate as market conditions change, given the model is underpinned by three months of data only.
 - **User feedback:** This can be requested to help identify if the users experience any issues with the app and if the model has identified more relevant properties than using advertised prices. This provides a quicker and more direct measurement of whether the model is meeting its objectives.
-



Question 6: Video summary

<https://www.youtube.com/watch?v=hzwI4UL1nqQ>



Question 7: Proposed refinement

Word count excluding headings and references: 497

A proposal has been made to refine the app by allowing users to enter any range for their budget rather than being restricted to a fixed price band.

Advantage 1: Relevant and tailored results

The proposed refinement will allow the app to provide **more relevant and tailored results** to users.

- Users may have a budget range that spans multiple price bands or is narrower than the pre-defined bands. In both cases, not all the properties displayed would be relevant to the user. It will be difficult to determine which properties fall within budget and which do not, given the app does not provide a more granular estimate of the price.
- With the proposed refinement, the user will be able to define a specific budget range that is unique to their circumstances. This ensures the app only presents properties within their actual budget.

Advantage 2: Future proof

The proposed refinement will make the web app more **future proof**.

- Fixed price bands are suitable in the property market and economic environment at the time they were defined. However, property prices tend to increase over time. For example, Australian home values have increased by 72% in the past decade (CoreLogic 2022). Therefore, these fixed bands are unlikely to remain suitable for future price levels.
- With the proposed refinement, the user can define the budget range that is relevant in the property market at the time. It will be easier for the app to reflect changes in inflation and general price levels.

Disadvantage 1: Delay in roll-out of app

The proposed refinement will **require time** to implement which will delay the roll-out of the app.

- Currently, the model has been built to estimate price bands only. If we implement the proposed refinement, we will need to rebuild the models so that they produce an estimate of the price itself, which is then used to decide whether the property belongs to the user's defined budget range.
 - This is not a trivial change to the model structure, so will require some time to rebuild and refine the model, which will result in a delay in the roll-out of the app.
-



Disadvantage 2: No indication of how confident we can be that a property will be within budget

The proposed refinement will require the model to produce an estimate of the price itself rather than just allocate the properties to price bands. This makes it more **difficult to know how confident we can be that a property will be within budget**.

- The current model produces probabilities of properties belonging to each fixed price band. This allows us to determine how confident we can be in the model's predictions. For example, we can be more confident if the model assigns an 80% probability of belonging to a price band, as opposed to 60%. These confidence measures can also be provided to the buyer to give them an idea of which properties are more or less likely to be within their budget.
- With the proposed refinement, the model will no longer be structured in a way that automatically provides these probabilities. This makes it much harder to calculate a level of confidence for a property being within the buyer's budget.

Sources

CoreLogic 2022, *The long game...30 years of housing values*, accessed 17 September 2022, <<https://www.corelogic.com.au/news-research/news/2022/the-long-game-30-years-of-housing-values>>
