Analysis of Housing Prices in Relation to Infrastructure and Landmarks in a Melbourne Suburb

---

Introduction

Objective

The objective of my analysis was to understand how the price of housing is impacted by the presence of various kinds of infrastructure and landmarks in the city of Melbourne. I believe that understanding the relationship between the two will help improve the quality of life in underdeveloped suburbs by applying actionable insights obtained from this analysis in city planning.

Data Description

For the purpose of this analysis I have used the the following three datasets

Dataset 1: Landmarks and places of interest, including schools, theatres, health services, sports facilities, places of worship, galleries and museums.
Dataset Link

*Dataset 2:*House Prices by Small Area - Sale Year
Dataset Link

Dataset 3: Small Areas for Census of Land Use and Employment (CLUE)
Dataset Link

I have the used aforementioned dataset to create my own merged updated dataset which holds information on the data on the prices of housing in various different suburbs and also the information on the various infrastructural amenities for each of those suburbs.

Data Sources

The data I have used is obtained form CoM Open Data Playground.

Data Attributes

The key attributes of the data used in this analysis include:

- Property Prices: The selling prices of houses and apartments in the suburb.
- Infrastructure: Availability and proximity of roads, co-ordinates, public transport, schools, hospitals, and shopping centers.
- Landmarks: Proximity to parks, recreational facilities, historical sites, and other notable landmarks.

Methodology

I started my analysis by scouring the CoM Open Data Playground to find datasets relevant to my chosen use case. After some digging, I managed to get hold of the following datasets:

Dataset 1: Landmarks and places of interest, including schools, theatres, health services, sports facilities, places of worship, galleries and museums.
Dataset Link

	theme	sub_theme	feature_name	co_ordinates
0	Transport	Railway Station	Flemington Bridge Railway Station	{'lon': 144.939277838304, 'lat': -37.788164588...
1	Mixed Use	Retail/Office/Carpark	Council House 2 (CH2)	{'lon': 144.966638432727, 'lat': -37.814259143...
2	Leisure/Recreation	Informal Outdoor Facility (Park/Garden/Reserve)	Carlton Gardens South	{'lon': 144.971266479841, 'lat': -37.806068457...
3	Place of Worship	Church	Wesley Church	{'lon': 144.968168215633, 'lat': -37.810157644...
4	Place of Worship	Church	St Augustines Church	{'lon': 144.954862000132, 'lat': -37.816974135...

*Dataset 2:*House Prices by Small Area - Sale Year
Dataset Link

	sale_year	small_area	type	median_price	transaction_count
0	2000	Kensington	Residential Apartment	232500.0	46
1	2000	North Melbourne	Residential Apartment	172500.0	196
2	2000	West Melbourne (Residential)	Residential Apartment	240300.0	66
3	2001	Carlton	Residential Apartment	273000.0	366
4	2001	Docklands	Residential Apartment	525375.0	766

Dataset 3: Small Areas for Census of Land Use and Employment (CLUE)
Dataset Link

	geo_point_2d	geo_shape	featurenam	shape_area	shape_len
0	{'lon': 144.94506274103145, 'lat': -37.7984489...	{'type': 'Feature', 'geometry': {'coordinates'...	North Melbourne	2408377.21789	7546.64919141
1	{'lon': 144.9221537804208, 'lat': -37.80962130...	{'type': 'Feature', 'geometry': {'coordinates'...	West Melbourne (Industrial)	5917883.21599	11793.960449
2	{'lon': 144.91223395712774, 'lat': -37.8318317...	{'type': 'Feature', 'geometry': {'coordinates'...	Port Melbourne	5470092.78821	15117.7233976
3	{'lon': 144.9416850851487, 'lat': -37.81851829...	{'type': 'Feature', 'geometry': {'coordinates'...	Docklands	2856028.73238	9588.24778517
4	{'lon': 144.966376105915, 'lat': -37.796011677...	{'type': 'Feature', 'geometry': {'coordinates'...	Carlton	2724820.83121	9914.977304

After working with each dataset individually, I gained insights into their pros and cons. Then, I focused on merging them appropriately. I began by combining Dataset 3 (CLUE) with Dataset 1 (landmarks). This merge was crucial because Dataset 1 contained detailed info about various landmarks in Melbourne, like their nature, theme, and coordinates. However, figuring out the suburbs to which these landmarks belonged was a bit tricky at first. But I sorted it out when I found suburb names and coordinates within Dataset 3, which allowed me to add a suburb column to the landmark dataset. With this integration, I could create better visualizations and manipulate the dataset more effectively.

Next, I crafted various visualizations to gain a deeper understanding of the datasets. After that, I turned my attention to the housing price dataset, where I built visualizations to track price changes over the years. During this process, I found some missing median prices, which I fixed using imputation techniques to maintain data integrity. Additionally, I merged the two datasets based on suburbs, so I could analyze them together.

	sale_year	suburb	type	median_price	transaction_count	number_of_features	theme_Community Use	theme_Education Centre	theme_Health Services	theme_Industrial	...	theme_Place Of Assembly	theme_Place of Worship	theme_Purpose Built	theme_Residential Accommodation	theme_Retail	theme_Specialist Residential Accommodation	theme_
0	2000	Kensington	Residential Apartment	232500.0	46	13.0	0.0	1.0	0.0	0.0	...	0.0	1.0	0.0	0.0	0.0	0.0	
1	2000	North Melbourne	Residential Apartment	172500.0	196	14.0	1.0	1.0	1.0	0.0	...	0.0	1.0	0.0	0.0	0.0	0.0	
2	2000	West Melbourne (Residential)	Residential Apartment	240300.0	66	7.0	0.0	0.0	0.0	0.0	...	1.0	1.0	0.0	0.0	1.0	0.0	
3	2001	Carlton	Residential Apartment	273000.0	366	24.0	1.0	1.0	1.0	0.0	...	1.0	1.0	0.0	0.0	0.0	0.0	
4	2001	Docklands	Residential Apartment	525375.0	766	20.0	0.0	1.0	0.0	0.0	...	1.0	0.0	1.0	0.0	0.0	1.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
341	2016	Docklands	Residential Apartment	591000.0	371	20.0	0.0	1.0	0.0	0.0	...	1.0	0.0	1.0	0.0	0.0	1.0	
342	2016	Kensington	House/Townhouse	818750.0	172	13.0	0.0	1.0	0.0	0.0	...	0.0	1.0	0.0	0.0	0.0	0.0	
343	2016	Kensington	Residential Apartment	430000.0	128	13.0	0.0	1.0	0.0	0.0	...	0.0	1.0	0.0	0.0	0.0	0.0	
344	2016	Melbourne (Remainder)	Residential Apartment	582500.0	80	15.0	1.0	1.0	0.0	0.0	...	0.0	0.0	0.0	1.0	0.0	0.0	
345	2016	West Melbourne (Residential)	House/Townhouse	900000.0	26	7.0	0.0	0.0	0.0	0.0	...	1.0	1.0	0.0	0.0	1.0	0.0	

340 rows × 23 columns

Despite using the Random Forest Regressor model to predict prices, the accuracy wasn't great. So, I shifted my focus to understanding the main factors driving price variations.

To figure out these factors, I conducted a feature importance analysis. Using the Random Forest Regressor model, I aimed to quantify the contribution of each feature to predictive accuracy. Here's how I did it:

- Feature Importance Revelation: I used the feature_importances_ attribute of the Random Forest Regressor model to uncover the importance scores assigned to each feature.
- Feature Names Extraction: I extracted feature names from the training dataset using the X_train.columns attribute.
- Sorting by Impact: I sorted feature importance scores in descending order to identify the most influential features impacting house prices.

Conclusion

By conducting this analysis, I have gained insights into various aspects. I was able to draw insights regarding the nature of amenities in each suburb and their distribution throughout the city. Although I employed the RFR model to make predictions, the results were not as great as expected, possibly due to dataset limitations. If I had information on when landmarks were constructed, I could implement it to understand how prices vary year after year in relation to infrastructural development.

Upon running the feature importance analysis, I gained an understanding of the factors from my dataset and how they would justify the price growth impact. From the results of the feature importance analysis, I believe that a few insights can be drawn from this dataset.

1. The impact of transaction count on prices seems straightforward; scarcity of units available may drive prices higher. I feel that further investigations should be conducted with a larger dataset.

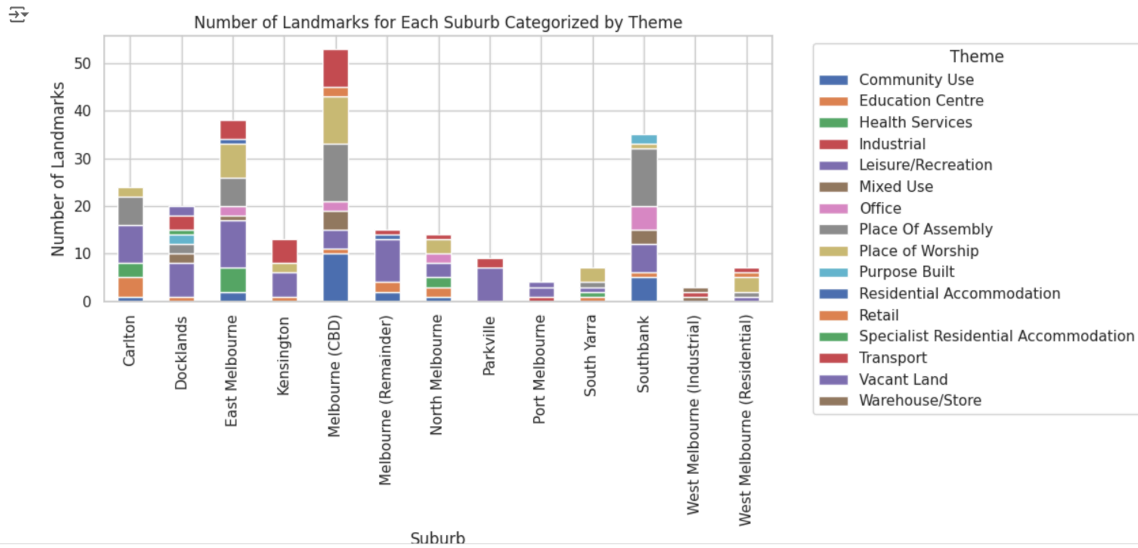
2. Purpose-built real estate affecting housing prices in a suburb indicates that the presence or absence of specialized properties tailored to specific needs or demographics influences the overall property market. This impact can stem from various factors, including the alignment of property offerings with market demand, provision of tailored amenities and infrastructure, influence on neighborhood character, perceived investment potential, and dynamics of supply and demand. Purpose-built properties contribute to shaping the identity of a suburb, attracting buyers, investors, and driving up property prices due to increased demand and limited supply. Consequently, purpose-built real estate serves as a significant determinant of housing market dynamics within a given area.

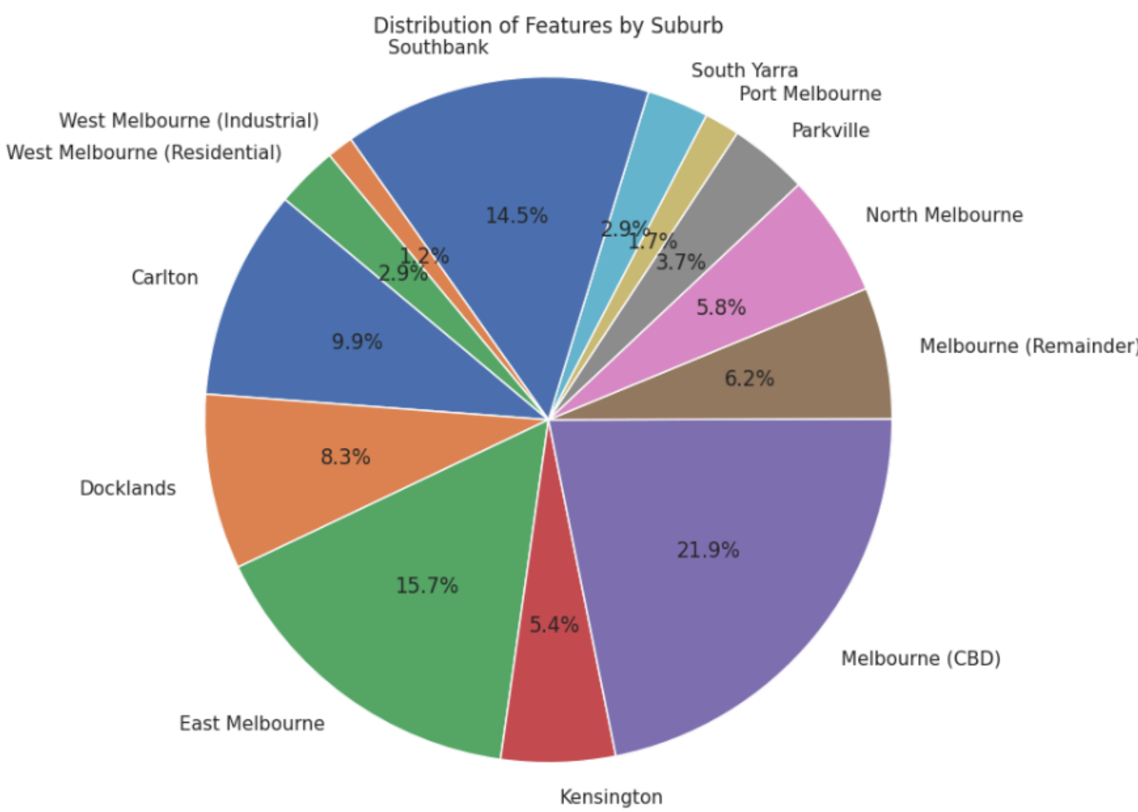
3. The impact of the sale year on median prices is pretty straightforward, as financial market conditions definitely influence house purchasing ability.

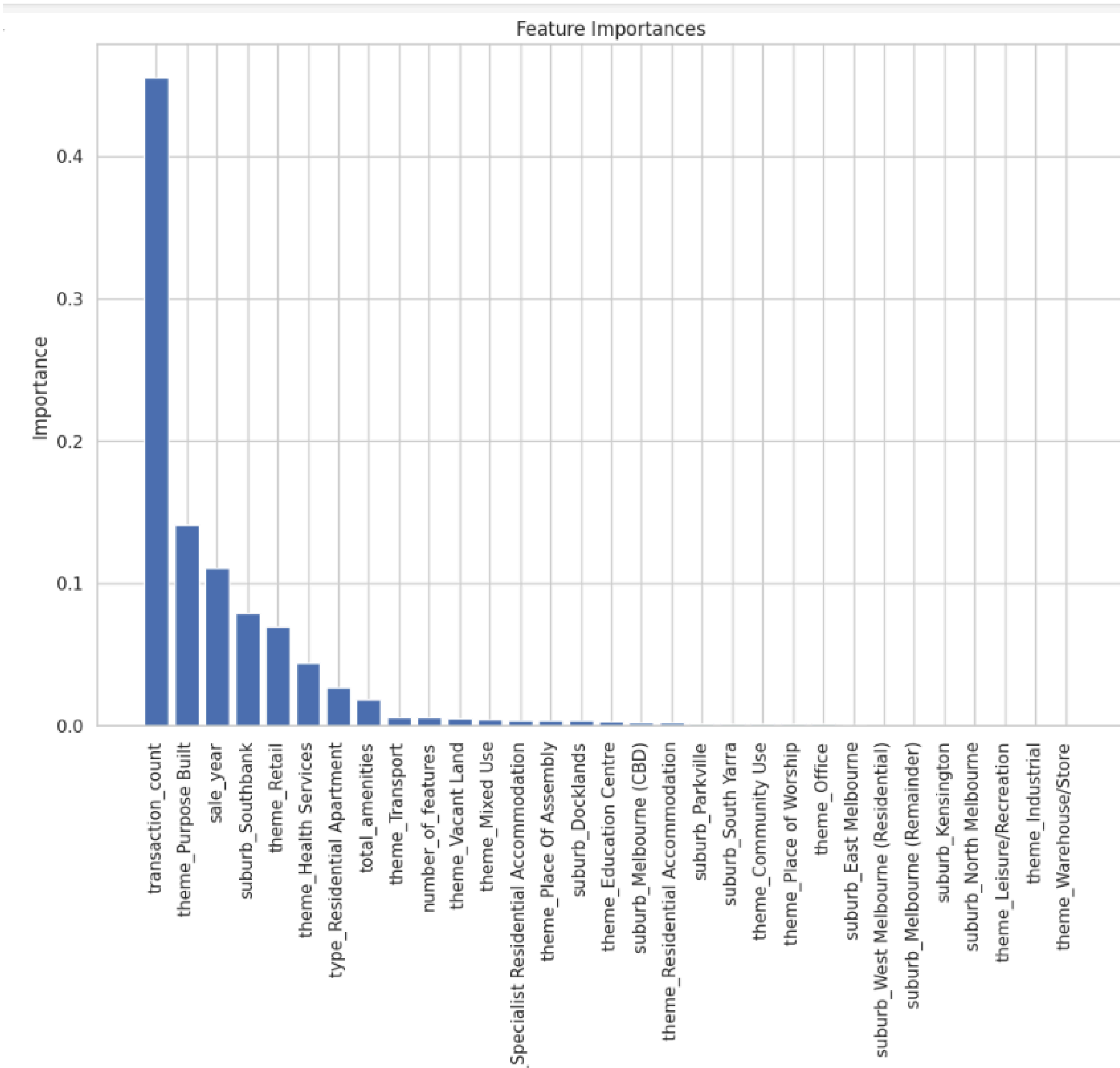
4. While all other factors seem like they could drive up the price, the entry of whether a house or apartment is in Southbank impacting the price is quite interesting, as I feel that the presence of numerous tourist spots in Southbank is what drives the price, making it premium real estate.

5. Retail outlets, healthcare, residential apartments, and the total number of amenities are quite self-explanatory, I think.

6. It is noteworthy that the total_amenities column represents how many total landmarks exist, and number_of_features indicates how many unique kinds of landmarks exist. For example, if there are 7 bus stops and 2 train stations in a suburb, total amenities will show 9, while the number of features will show 2, as a bus stop is one kind of feature and a train station is another. Thus, the data shows that having more amenities is more important than having more distinct features.


Number of Landmarks for Each Suburb Categorized by Theme
Number of Landmarks
0
10
20
30
40
50
Carlton
Docklands
East Melbourne
Kensington
Melbourne (CBD)
Melbourne (Remainder)
North Melbourne
Parkville
Port Melbourne
South Yarra
Southbank
West Melbourne (Industrial)
West Melbourne (Residential)
Suburb
Theme
Community Use
Education Centre
Health Services
Industrial
Leisure/Recreation
Mixed Use
Office
Place Of Assembly
Place of Worship
Purpose Built
Residential Accommodation
Retail
Specialist Residential Accommodation
Transport
Vacant Land
Warehouse/Store


Distribution of Features by Suburb
Southbank
14.5%
South Yarra
2.9%
Port Melbourne
1.7%
Parkville
3.7%
North Melbourne
5.8%
Melbourne (Remainder)
6.2%
Melbourne (CBD)
21.9%
Kensington
5.4%
East Melbourne
15.7%
Docklands
8.3%
Carlton
9.9%
West Melbourne (Residential)
2.9%
West Melbourne (Industrial)
1.2%


Feature Importances
Importance
0.0
0.1
0.2
0.3
0.4
transaction_count
theme_Purpose Built
sale_year
suburb_Southbank
theme_Retail
theme_Health Services
type_Residential Apartment
total_amenities
theme_Transport
number_of_features
theme_Vacant Land
theme_Mixed Use
Specialist Residential Accommodation
theme_Place Of Assembly
suburb_Docklands
theme_Education Centre
suburb_Melbourne (CBD)
theme_Residential Accommodation
suburb_Parkville
suburb_South Yarra
theme_Community Use
theme_Place of Worship
theme_Office
suburb_East Melbourne
suburb_West Melbourne (Residential)
suburb_Melbourne (Remainder)
suburb_Kensington
suburb_North Melbourne
theme_Leisure/Recreation
theme_Industrial
theme_Warehouse/Store

```
Feature Importances in Percentage:
transaction_count: 45.52%
theme_Purpose Built: 14.18%
sale_year: 11.10%
suburb_Southbank: 7.90%
theme_Retail: 6.99%
theme_Health Services: 4.43%
type_Residential Apartment: 2.70%
total_amenities: 1.85%
theme_Transport: 0.64%
number_of_features: 0.63%
theme_Vacant Land: 0.56%
theme_Mixed Use: 0.51%
theme_Specialist Residential Accommodation: 0.43%
theme_Place Of Assembly: 0.39%
suburb_Docklands: 0.39%
theme_Education Centre: 0.30%
suburb_Melbourne (CBD): 0.26%
theme_Residential Accommodation: 0.21%
suburb_Parkville: 0.20%
suburb_South Yarra: 0.18%
theme_Community Use: 0.17%
theme_Place of Worship: 0.17%
theme_Office: 0.16%
suburb_East Melbourne: 0.07%
suburb_West Melbourne (Residential): 0.05%
suburb_Melbourne (Remainder): 0.01%
suburb_Kensington: 0.01%
suburb_North Melbourne: 0.01%
theme_Leisure/Recreation: 0.00%
theme_Industrial: 0.00%
theme_Warehouse/Store: 0.00%
```

Code: https://github.com/vaibhavideo/MOP-Codes