

Forecasting Real Estate Prices in Victoria

MAST30034 Applied Data Science

Team 59

Terry Chow
Zhiqian Lai
Manish Khilari
Jiajia Guo
Arjun Rajkumar

Scope & Objectives

Scope -

The rental market in Victoria

Objective Questions -

- What are the most important internal and external predictors of rental prices?
- What are the top 10 suburbs with the highest predicted growth rate?
- What are the most liveable and affordable suburbs given the selected livability metrics?



Scope & Objectives (continued)

Objectives

- To reduce price uncertainty in the Victorian rental market
- To provide suburb recommendations to renters and investors seeking liveability and a return on their investment



Group Management & Timeline

- Task Tracking:

Trello

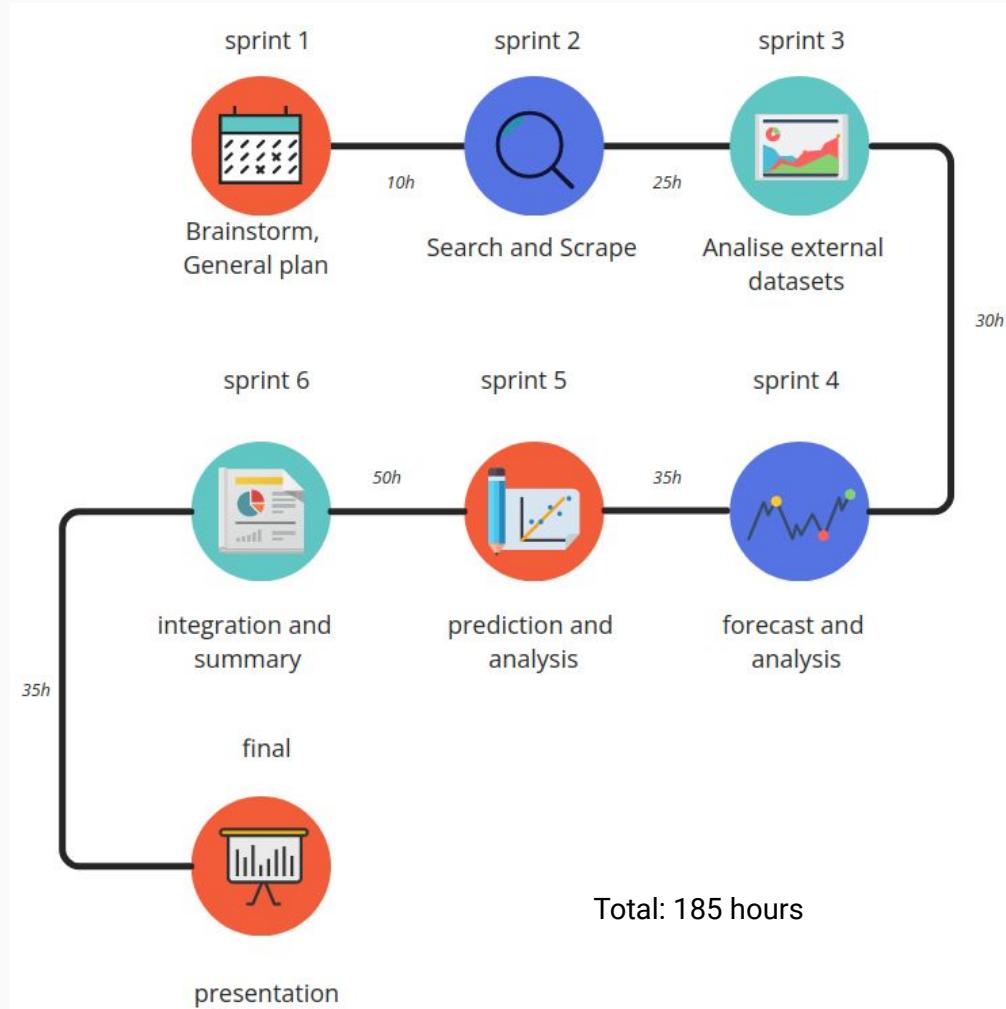


- Conferencing & Standups:

Video meeting through zoom



- Communications held via discord



Assumptions and Limitations

API Access

- API keys were stored offline to prevent unauthorised use
- No API keys were pushed online

Information Security and Privacy

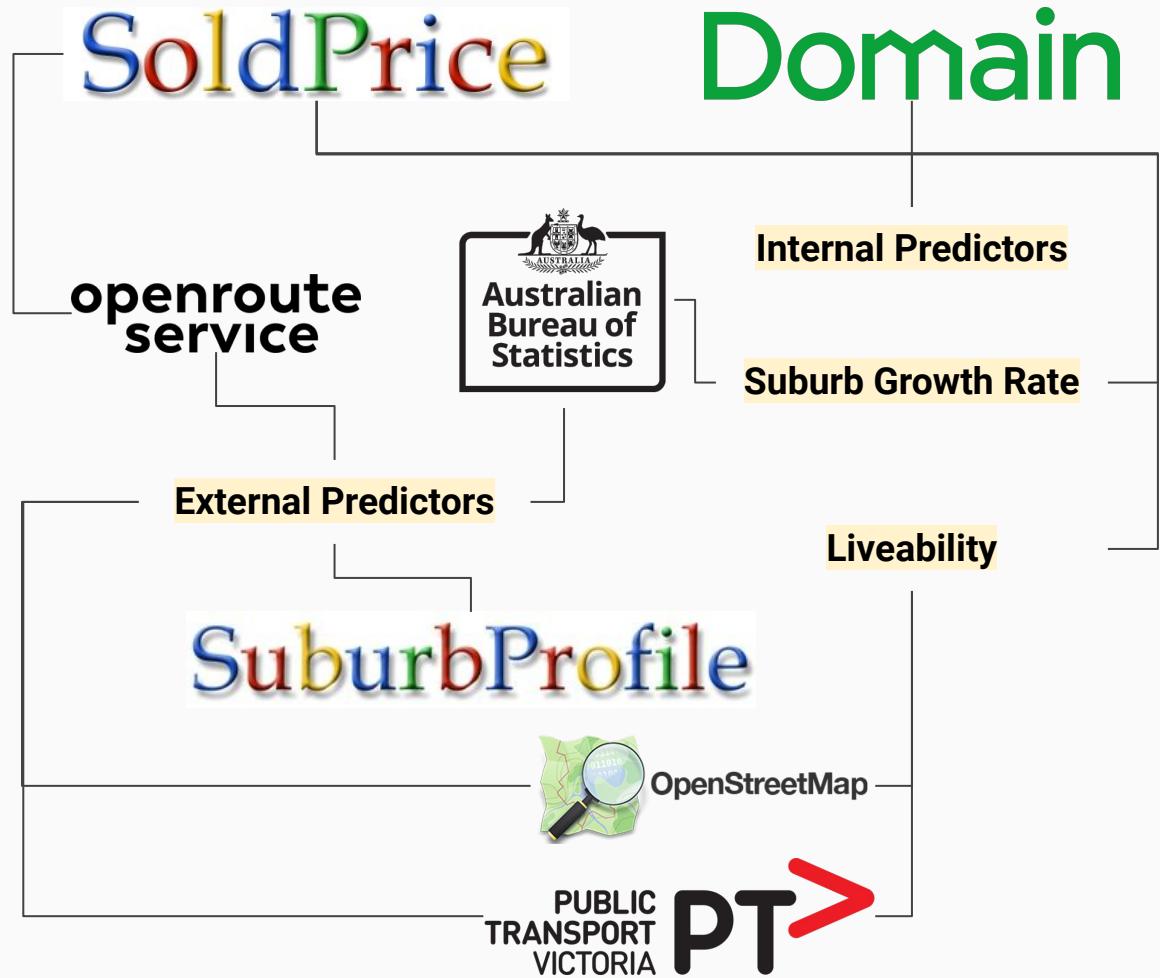
- Raw data was automatically downloaded from the relevant web locations
- No data was stored internally
- Code was stored on a private online repository

The screenshot shows the Australian Bureau of Statistics (ABS) website. At the top left is the ABS logo. The main navigation bar includes links for Home, Statistics, People, Education, and Data Access. Below the navigation, a breadcrumb trail shows Home > Statistics > People > Education > Data Access. The main content area has a title 'On this page' followed by a list of sections: Key statistics, COVID-19 in this publication, Students, Retention rates, Aboriginal and Torres Strait Islander students, Schools, Staff, Data downloads, Methodology, and Media releases. To the right of the sections is a table titled 'Available Data Properties'. The table has columns for 'Property', 'Basic', 'Detailed', and 'Full'. It lists various address components and other properties like id and relativeScore, each marked with a checkmark in all three columns.

Property	Basic	Detailed	Full
address	✓	✓	✓
addressComponents	✓	✓	✓
addressComponents.postCode	✓	✓	✓
addressComponents.state	✓	✓	✓
addressComponents.streetName	✓	✓	✓
addressComponents.streetNumber	✓	✓	✓
addressComponents.streetType	✓	✓	✓
addressComponents.streetTypeLong	✓	✓	✓
addressComponents.suburb	✓	✓	✓
addressComponents.unitNumber	✓	✓	✓
id	✓	✓	✓
relativeScore	✓	✓	✓

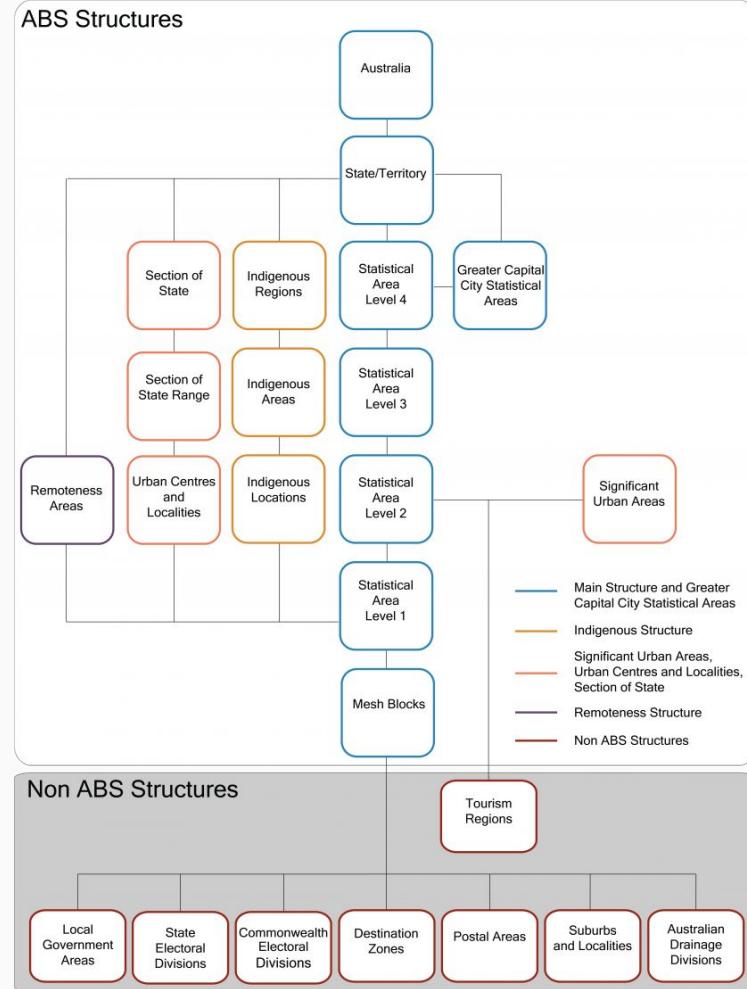
Data Pipeline Overview

- Real Estate Data
 - House Speaking same
 - Domain
- PTV Stations
- ABS
 - Population
 - Schools
 - LGA & Suburb Shapefiles
- OpenStreetMap Features



Data Retrieval (Manish)

- **ABS Statistical Geography**
 - Non overlapping States > SA4 > SA3 > SA2 > SA1 > Mesh Blocks
 - Suburbs however may overlap many SAs
- **Suburb to LGA / SA2 mapping**
 - **Constraint:** Suburb was not an ABS structure
 - This blocked ABS data joins on suburb
 - Resolved by mapping suburb to mesh block
 - Mesh block were then mapped to LGA and SA2
- **Domain**
 - Offered reputable and extensive data
 - Abandoned due to an API rate limit of 500 calls per day
 - This meant a dataset of size 50 000 would take 100 days to accumulate



Data Retrieval (Terry)

- House Speakingsame
 - Real Estate Data
 - 80K+ Properties Scrapped
 - Historic Rent & Sold Values since 2000
 - Suburb Rankings
 - Average Rent & Sold Value
 - Schools & Education
 - Median Income/Week
- OpenRouteService
 - Driving Distance & Duration to CBD from every property
- ABS
 - Suburb Shapefiles

Suburbs Scraped in order of Rent Value

Melbourne Rent Price - Ranking

View this report in [Sydney](#) [Adelaide](#) [Brisbane](#) [Perth](#) [Canberra](#)

Rank	Suburb	Value
1	Mount Buller	\$38,333
2	Falls Creek	\$25,500
3	Flinders	\$1,535
4	Sawmill Settlement	\$1,462
5	Portsea	\$1,350
6	Neerim South	\$1,333
7	Toorak	\$1,315
8	Balaclava	\$1,302
9	Seaspray	\$1,301
10	Buln Buln	\$1,254
11	Brighton	\$1,239
12	Kooyong	\$1,190

Parkville, VIC Search

[+ Set search filters and sorting](#)

A 2/189 Park Drive
Sold \$800,000 in Sep 2022
Last Sold \$207,000 in Jun 1990
Rent \$440pw in May 2017
Unit: 2 1 1
Land size: 582 sqm
Agent: Sam Abboud
[Map](#) | [Nearby](#)

[More Photos](#)

B 2/19-21 Park Drive
Sold \$817,000 in Sep 2022
Last Sold \$200,000 in Jul 2000
Rent \$370pw in Aug 2009
Unit: 2 1 1
Land size: 412 sqm
[Map](#) | [Nearby](#)

[More Photos](#)

C 2/19 Park Drive
Sold \$817,000 in Sep 2022
Last Sold \$200,000 in Jul 2000
Rent \$430pw in Nov 2015
Apartment: 2 1 1
Land size: 412 sqm | Building size: 75 sqm
Agent: Peter Varella

[More Photos](#)

Data Retrieval (Arjun)



ABS - <https://www.abs.gov.au/statistics>

Datasets:

- Regional population by age & sex : `age_sex`
- Deaths, Australia : `pop_deaths`
- Schools : `education`

Regional population by age and sex

Statistics about the population by age and sex for Australia's capital cities and regions

Reference period 2021

Released 30/08/2022

Next release 28/09/2023

▼ Previous releases

Key statistics

- The median age for capital cities (37.1 years) was younger than the rest of Australia (41.8).
- The youngest capital was Darwin with a median age of 34.5 years, while Adelaide was the oldest (39.3).
- Darwin was the only capital with more males than females.

Schools

Data on students, staff, schools, rates and ratios for government and non-government schools, for all Australian states and territories

Reference period 2021

Released 23/02/2022

Next release Unknown

▼ Previous releases

Key statistics

In Australia in 2021:

- 4,030,717 students were enrolled in 9,581 schools.
- the Year 7/8 to 12 full-time apparent retention rate was 83.1%.
- the average student to teaching staff ratio for all schools was 13.3 students to one teacher.
- the annual growth rate for school enrolments was 0.6% (23,743 students), the lowest growth rate since 2008.

Population Analysis and Forecasting

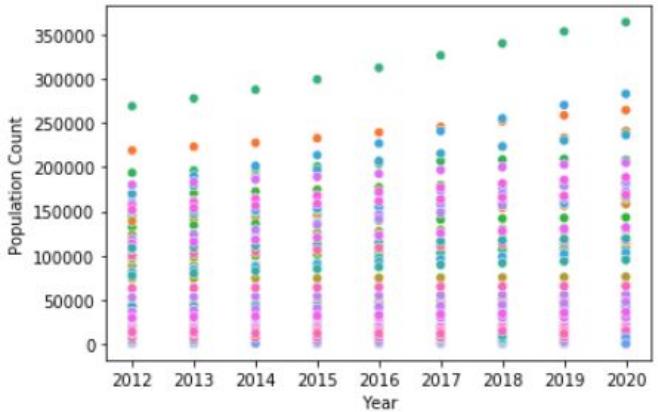


Figure 1 - Variance in population distribution across Victorian LGAs

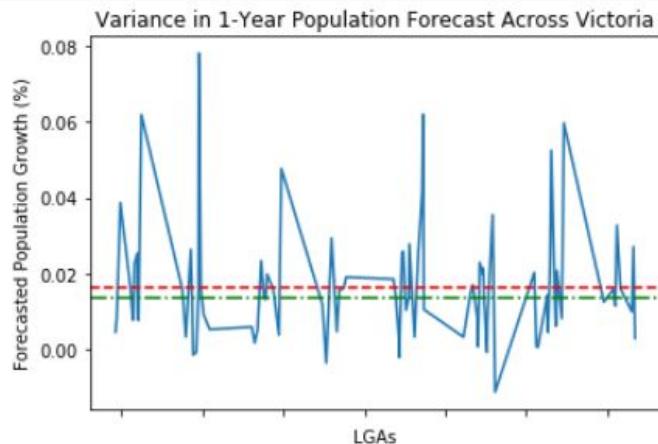


Figure 2 - Distribution of 1-year population forecast across Victorian LGAs

- Large variance in population across LGAs
- Explainable by difference in area, but could also be explained by intrinsic factors such as liveability & wealth

- 5 LGAs have the highest forecasted 1-year growth rate (above 5%)
- **Average & median** growth rate both just under 2%

Forecasting (continued)

- Most LGAs have a fairly static population forecast
- 6 LGAs predicted to see an increased rate of growth
- 2 LGAs with highest forecasted growth:
 - Colac-Otway
 - Wellington

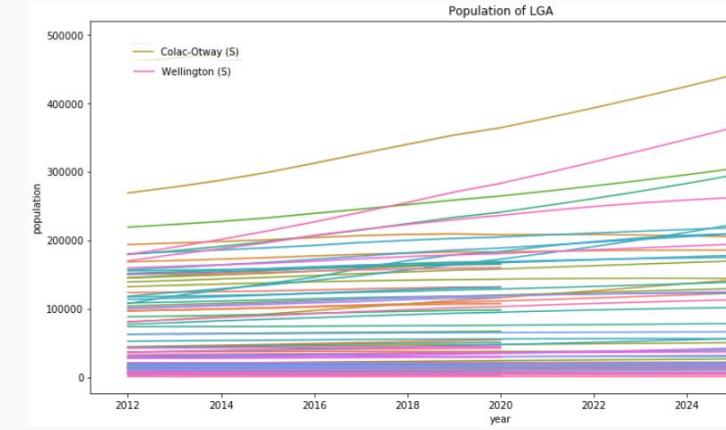


Figure 3 - Population over time with 5-year forecast for affluent LGAs



Figure 4 - Brown: Shire of Colac-Otway, pink: Shire of Wellington

- 2 LGAs with extremely promising population forecasts

- Both coastal cities
- Moderate distance from state capital
- Strong indicators of growth

Real Estate in Colac-Otway



Figure 3 - Economic indicators from Colac-Otway's tourism website

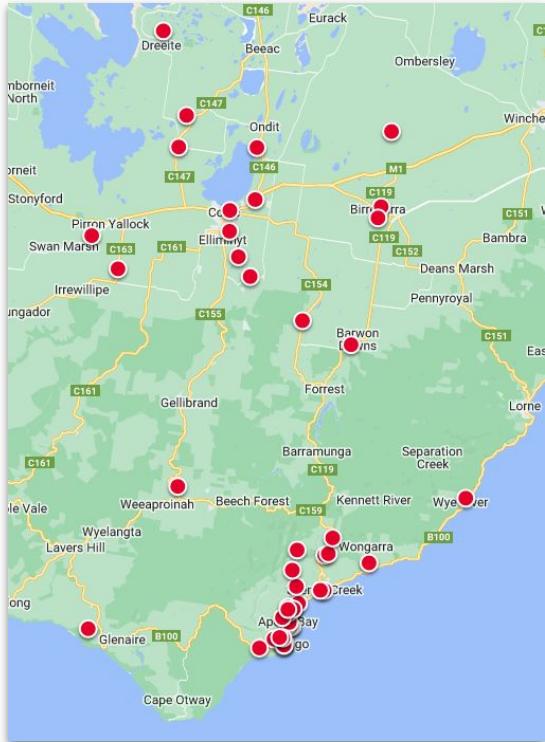


Figure 5 - realestate.com listings in Colac-Otway greater than \$1 million

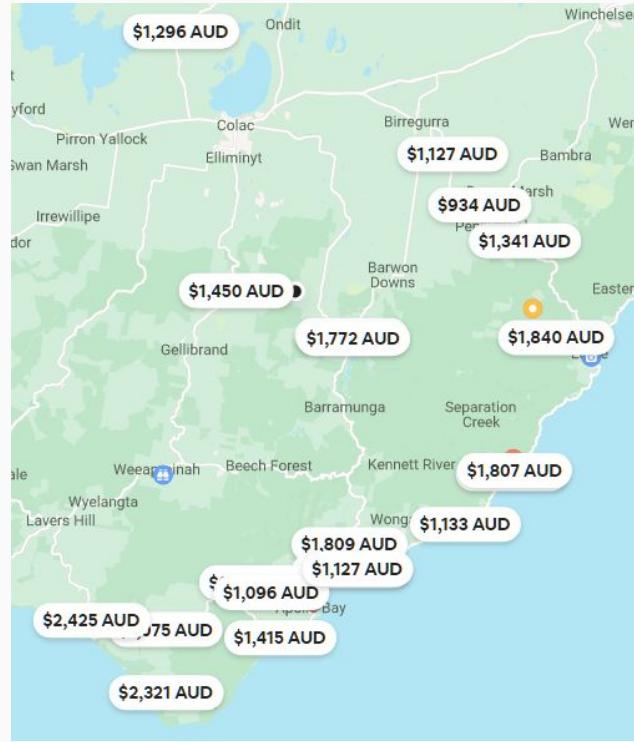


Figure 6 - 'Superhost' Airbnb listings in Colac-Otway

- 59 properties currently listed with a sale price over \$1 million
- Over 1000 homes listed on Airbnb

Data Retrieval (Zhiquan)

- PTV (Public Transport Victoria)
 - Contains 3 main types
 - More than millions of raw data
- Vic Crime statistics

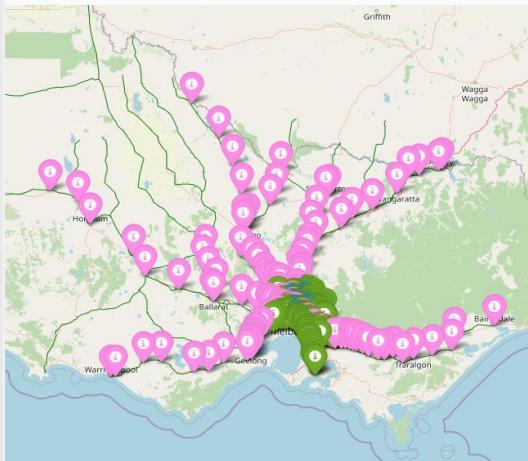


Figure 7 - Visualization for PTV tram and train



Figure 8 - Visualization for PTV bus

Latest crime data by area

Please note that movements in recorded crime data may be impacted by changes in legislation and operational police practice. Information is available about notable changes in the [Explanatory Notes](#), refer to this information when comparing data over time. Below highlights substantial changes that should be considered when analysing the data.

From March 2020 to December 2021 Victoria responded to the global COVID-19 pandemic through a range of health responses that included periods of restrictions on people's movements and daily activities. The health response was supported by the use of offences codes relating to breaches of Chief Health Officer (CHO) Directions the [Public Health and Wellbeing Act 2008](#). The restrictions on movement and introduction of breaches of CHO directions had a flow on effect for Victorian recorded crime, and this differed by crime type. For more information about the impacts of COVID-19 on crime please see the CSA paper [Police-recorded crime trends in Victoria during the COVID-19 pandemic: update to the end of December 2020](#). Data trends overtime should be interpreted with caution.

Downloadable excel files are available on the Download data webpage.

[Download data](#)

Screenshot of www.crimestatistics.vic.gov.au

Data Retrieval (Jiajia)

Maps

- Google map
- Cinematreasures
- Mapquest

Bank loan & Financial data

- Customer targeted
- Incomplete dataset / mostly been processed
- Limited time period

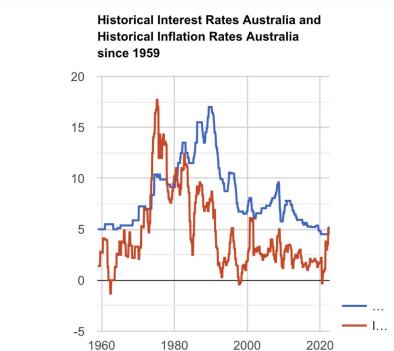


Figure 9: Historic Australian rates

Movie Theaters in Victoria

Showing 1 - 30 of 135 open movie theaters

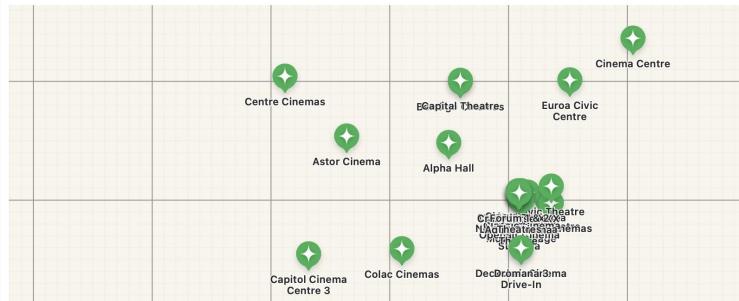


Figure 10: Theaters in VIC through Cinematreasures

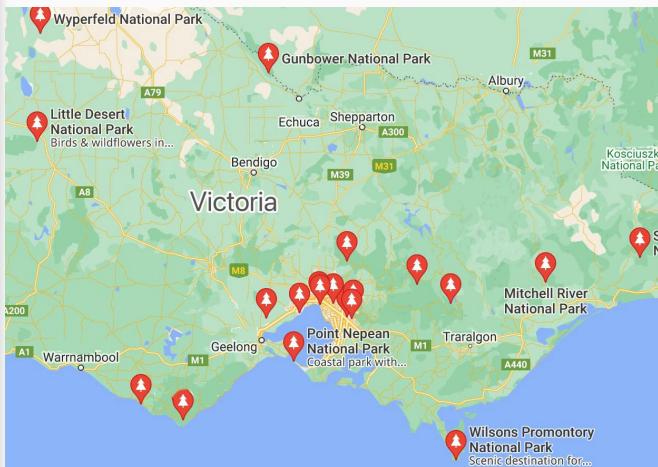


Figure 11: Parks in VIC through GoogleMap

Historic home loan rate averages (1959 - 2022)

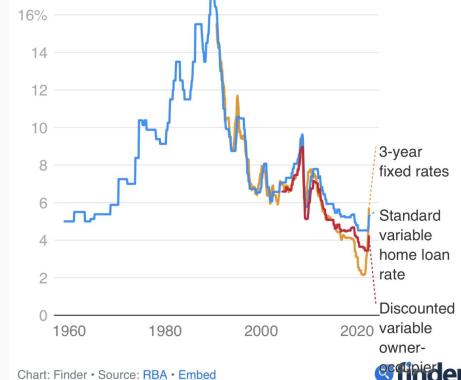


Figure 12: Home loan rate activity

Approach

Granularity: Suburb

Upsides

- Suburb is found in many location based open datasets
- Suburb is a more interpretable breakdown of area; LGAs & SA2s are exclusively used by the Government

Downsides

- Suburb is not considered an Australian Bureau of Statistics (ABS) Statistical Area
- There was high variance in property values within individual suburbs



Figure 13: Victorian suburb boundaries

Reproducibility

Raw Data Reproducibility

- Raw data is automatically downloaded from the relevant web locations
- No raw data is stored internally

Code Reproducibility

- Code and results are reproducible given requirements are met
 - Valid API keys
 - Python requirements
- All requirements and dependencies are listed

✓ Latest release

↓ Data download

Schools

Data on students, staff, schools, rates and ratios for government and non-government schools, for all Australian states and territories

Reference period 2021

Released 23/02/2022

Next release Unknown

▼ Previous releases

Key statistics

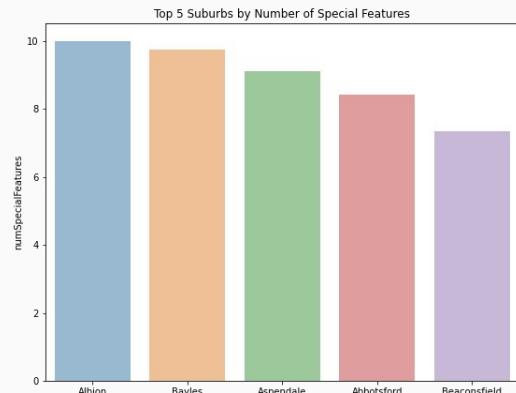
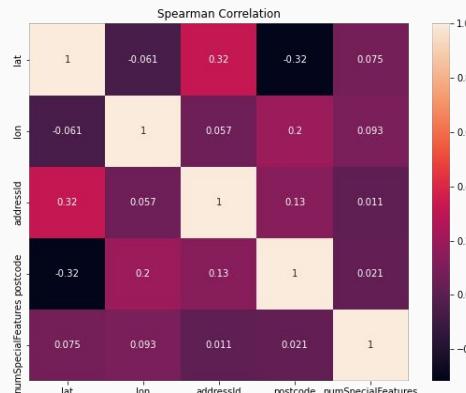
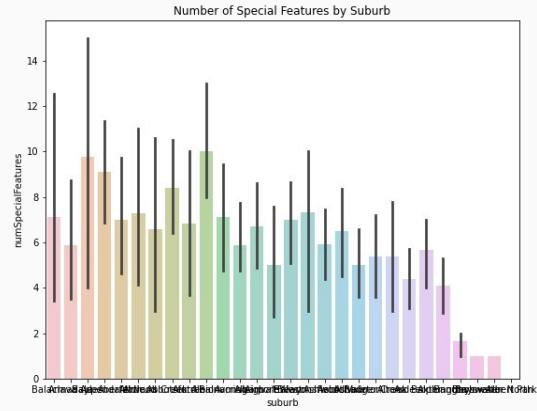
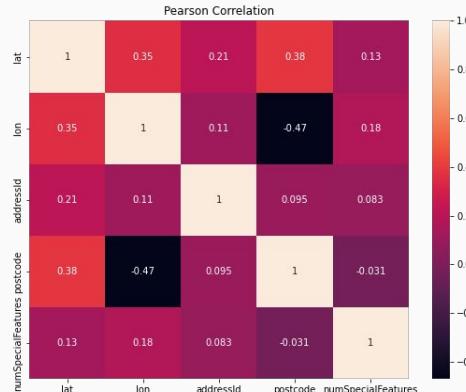
In Australia in 2021:

- 4,030,717 students were enrolled in 9,581 schools.
- the Year 7/8 to 12 full-time apparent retention rate was 83.1%.
- the average student to teaching staff ratio for all schools was 13.3 students to one teacher.
- the annual growth rate for school enrolments was 0.6% (23,743 students), the lowest growth rate since 2008.

Internal Predictors

Features of the property itself

- Land size
 - Number of
 - Car spaces
 - Bedrooms
 - Bathrooms
 - Number of special features
(listed exclusively by Domain)
 - Balcony
 - Ensuite
 - Laundry
 - Lock Up Garage
 - Fireplace
 - Pool
 - Tennis Court

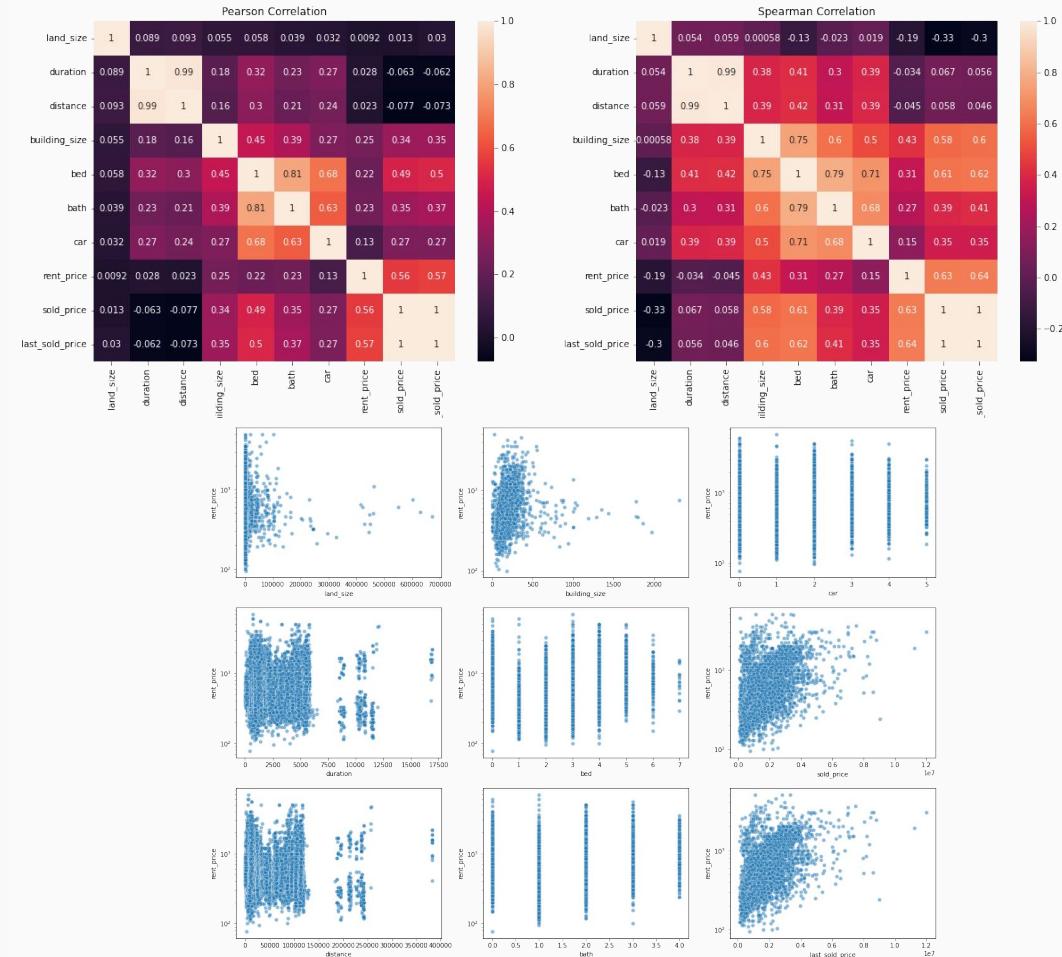


Figures 14-17: Feature analysis on Domain.com data

Internal Predictors

Features of the property itself

- Number of PTV stations within a specified radius r
 - Radius r set to 500m
- CBD access (from OpenRouteService)
 - Travel distance (m) to the CBD
 - Travel duration (s) to the CBD



Figures 18-20: Feature analysis on OpenRouteService data

External Features

External Features: The features of the wider context (SAL or LGA) which may influence the rent price.

Two types:

- Static (long term)
Eg. number of PTV stations
- Time series (dynamic)
Eg. population

Categorization method: Expectation of changes for next five years.



[Licensable: creative common license](#)

External Static Features

- PTV
- Occupations
- Household income
- School
- Property values
- Property selling data
- Housing Turnover
- Crime
- Occupations
- More ...

Visual Results

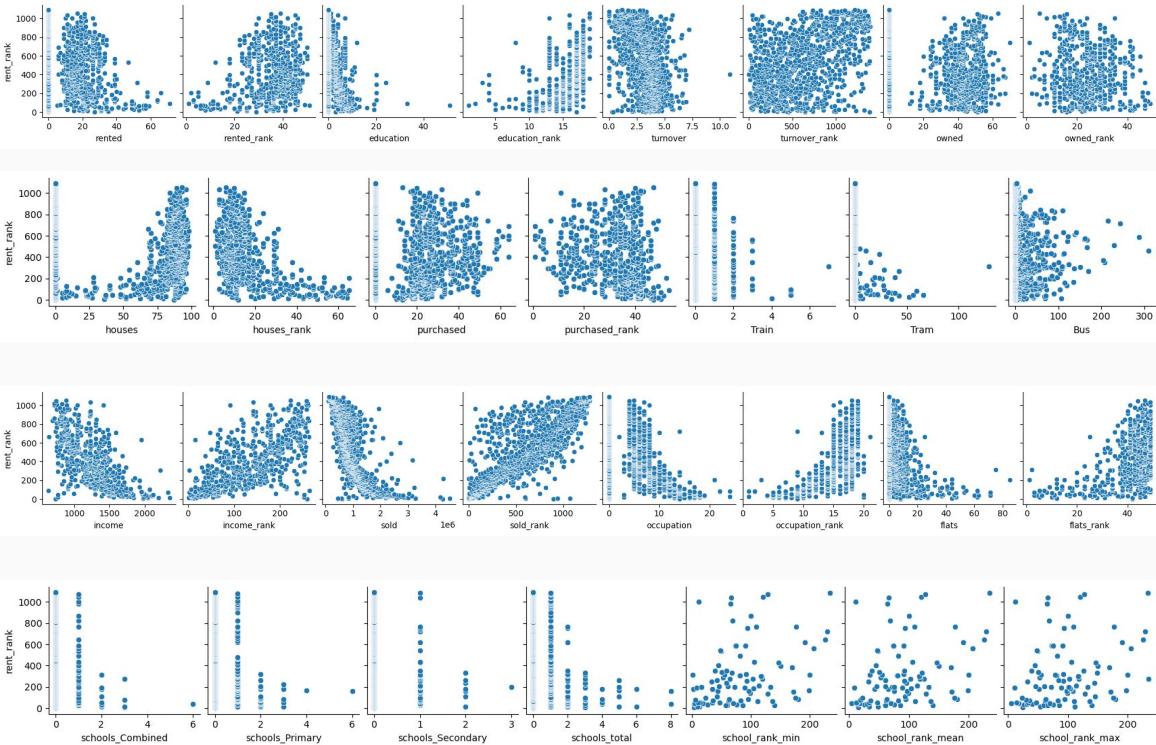


Figure 21 - 24: The scatter plots of features of external datasets (X s) vs rent rank (y).

External Static Features

Useful Features:

Strong:

- Pro occupations rate rank
- Median household income
- Property sold values

Moderate:

- Flats percentage
- House percentage rank
- Education rank
- Turnover rank
- Rented

Metric: Spearman Correlation Coefficient

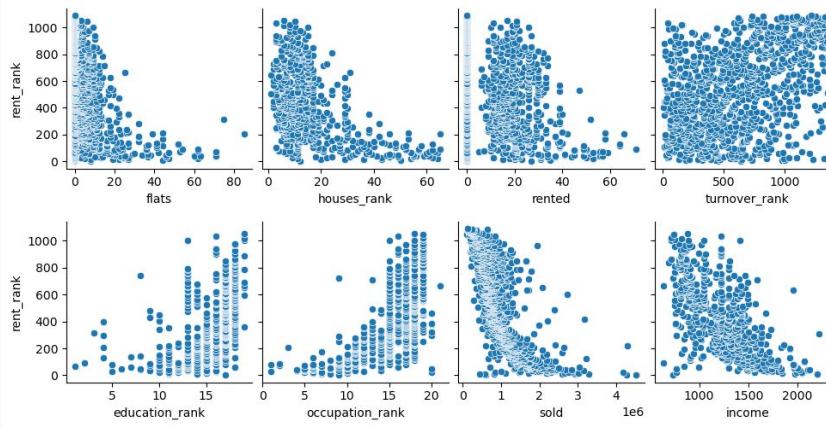
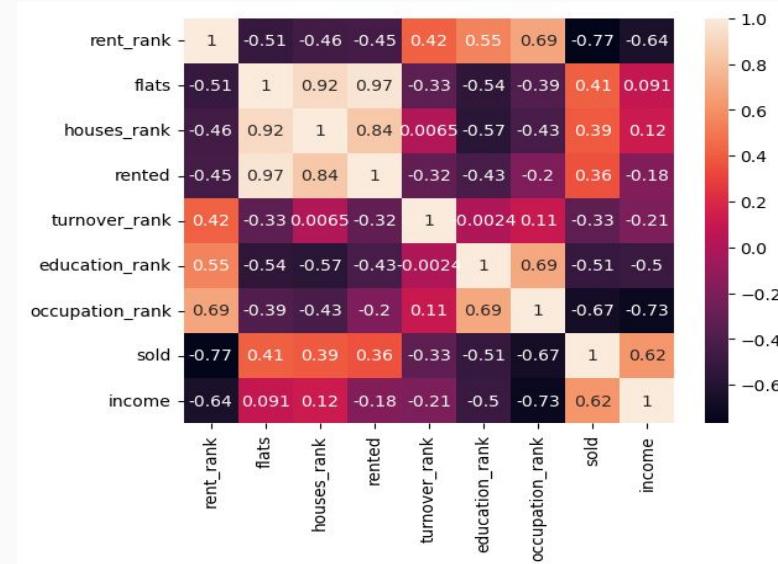


Figure 25: The scatter plots of selected features from external datasets (X s) vs rent rank (y).



Suburb Growth

- 10 suburbs were identified each using historic data and through forecasting
- Suburbs that appeared in the top 10 with both methods are:
 - Buln Buln
 - Maindample
 - Moorooduc
 - Sawmill Settlement
- All locations are heavy tourist hubs, with a focus on natural attractions
- Moorooduc's average rent was predicted to spike in 2022, however this was not observed
- Model may not generalise well, although there are differences between Moorooduc and the other suburbs to the right

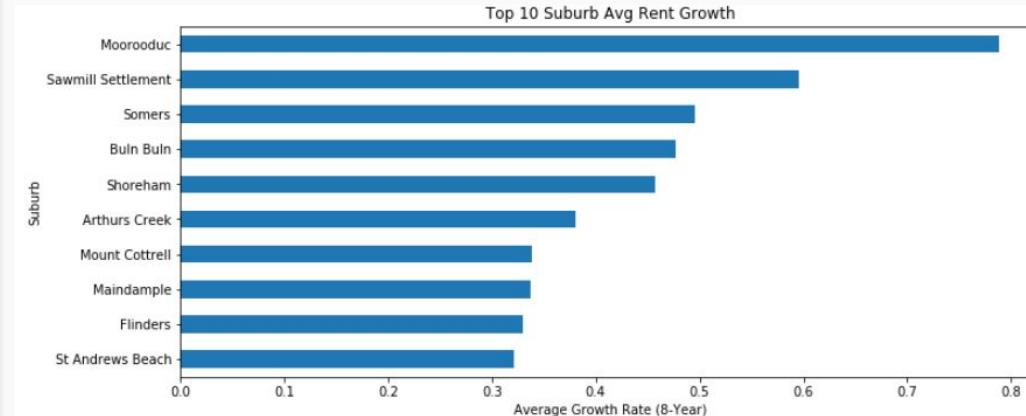


Figure 27: Victorian suburbs with highest rent growth over 8 years (historic)

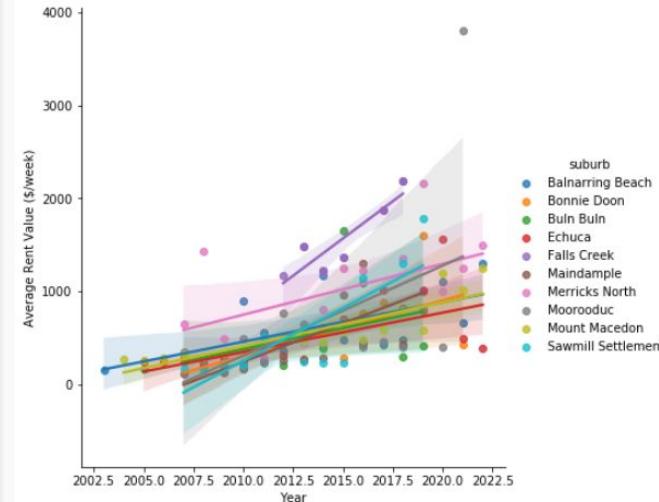


Figure 28: 3-year forecast of suburb growth (average rent value)

Liveability



What is Liveability?

SUSTAINABLE DEVELOPMENT GOALS

"The Liveability Index is a composite score based on measures related to aspects of liveability including **Social Infrastructure**, **Walkability**, **Public Transport**, **Public Open Space**, **Housing Affordability**, and **Local Employment**^[1]."

- [Australian Urban Observatory](#)

"The Liveability Index survey focuses on five key areas: **Belonging**, **Connection**, **Health and Fitness**, **Safety**, and **Environment and Sustainability**."

- [Stockland](#)

Liveability

Main Chosen Aspects:

-  Walkability
-  Public Transport
-  Education
-  Affordability
-  Employment

Data Availability



OpenStreetMap (JiaJia)

- Shopping Centres
- Theatres
- Parks



PTV (Mark)

- Bus Stations
- Train Stations
- Tram Stations

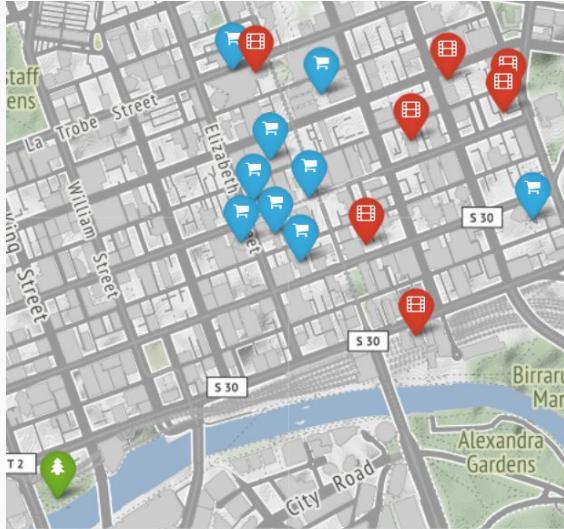
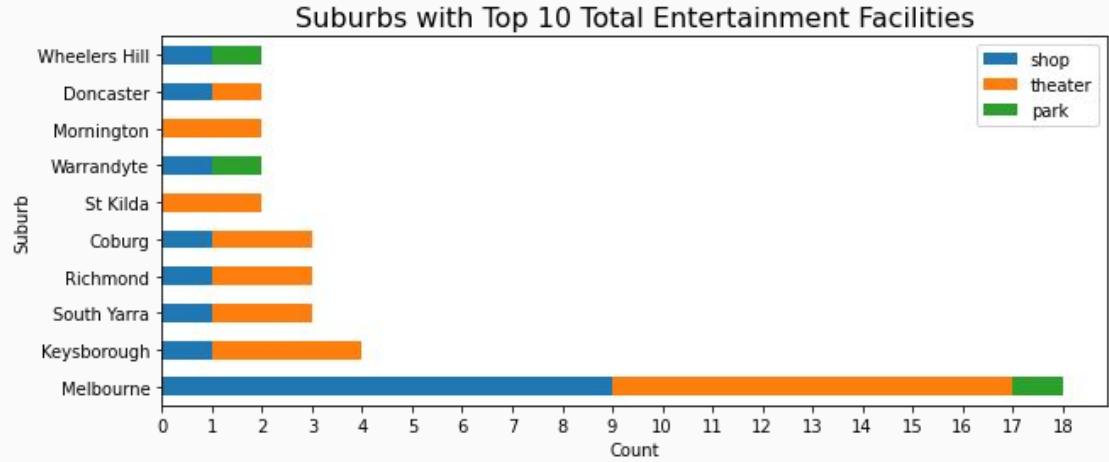
HouseSpeakingSame (Terry)

-  Affordability (Rent/Sold)
-  Tertiary Education Rankings
-  Professionals Rankings
-  Income Rankings
-  School Rankings

Walkability

Main Chosen Aspects:

-  **Walkability**
-  **Public Transport**
-  **Education**
-  **Affordability**
-  **Employment**



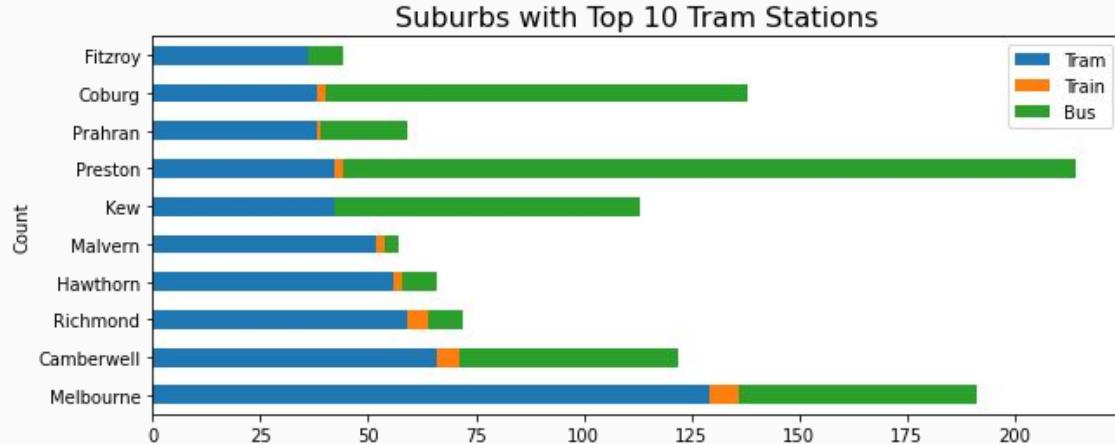
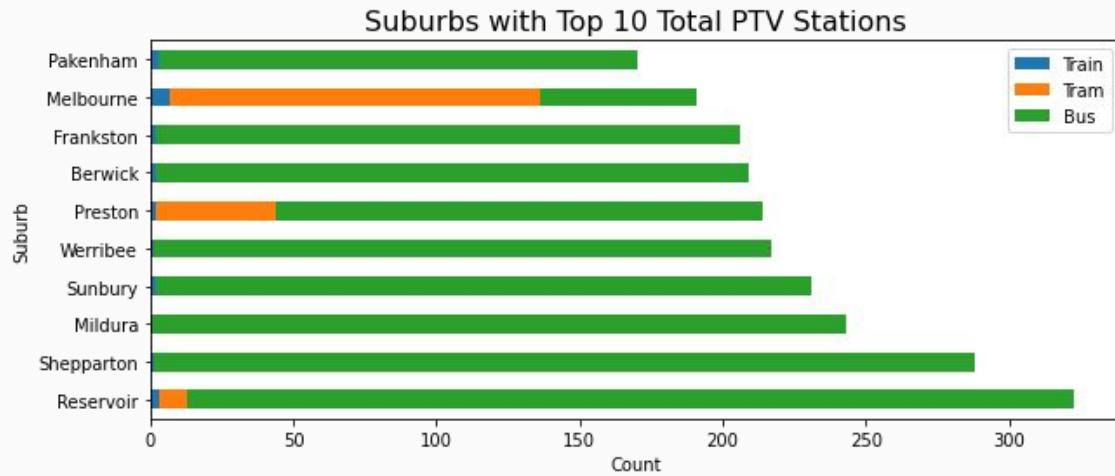
Theater
Shopping Mall
Park

The 3 Types of Entertainment Facilities in the Top Suburb, Melbourne.

Public Transport

Main Chosen Aspects:

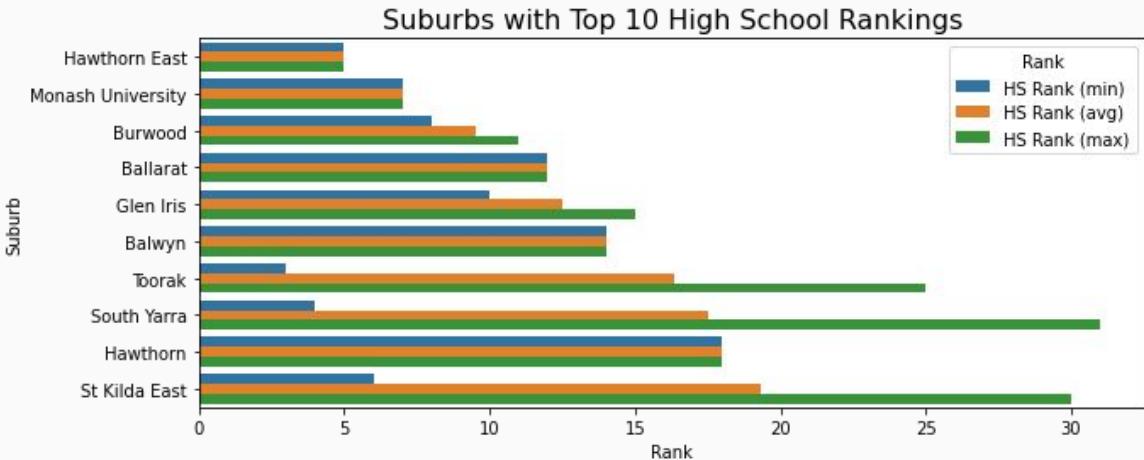
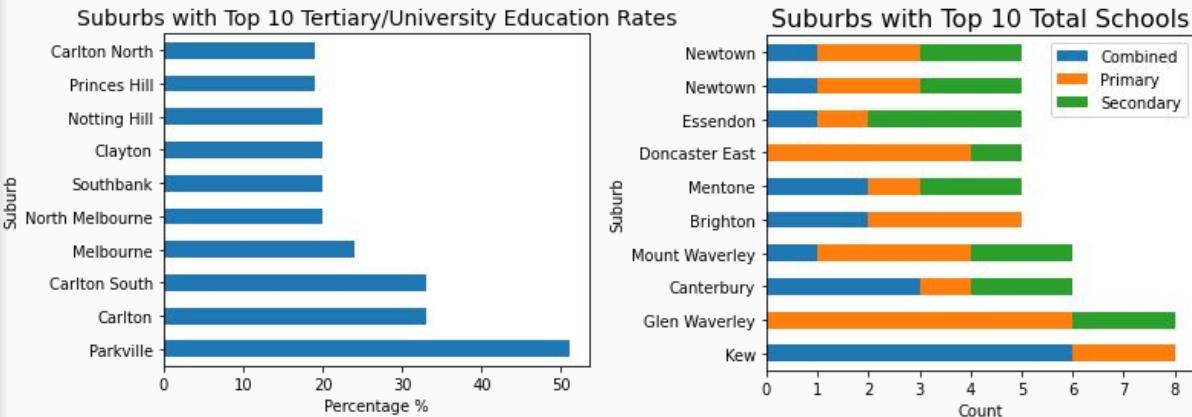
-  Walkability
-  Public Transport
-  Education
-  Affordability
-  Employment



Education

Main Chosen Aspects:

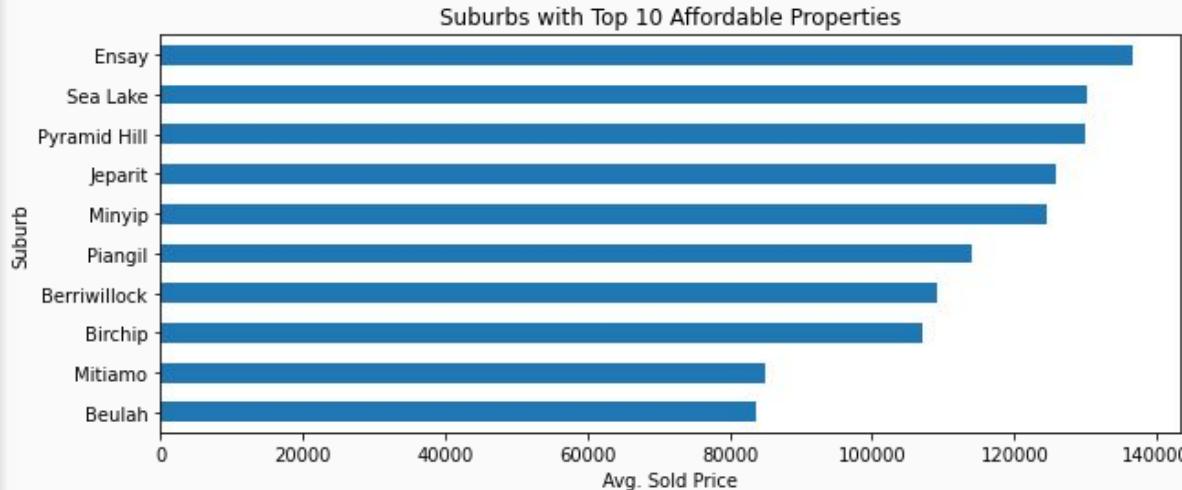
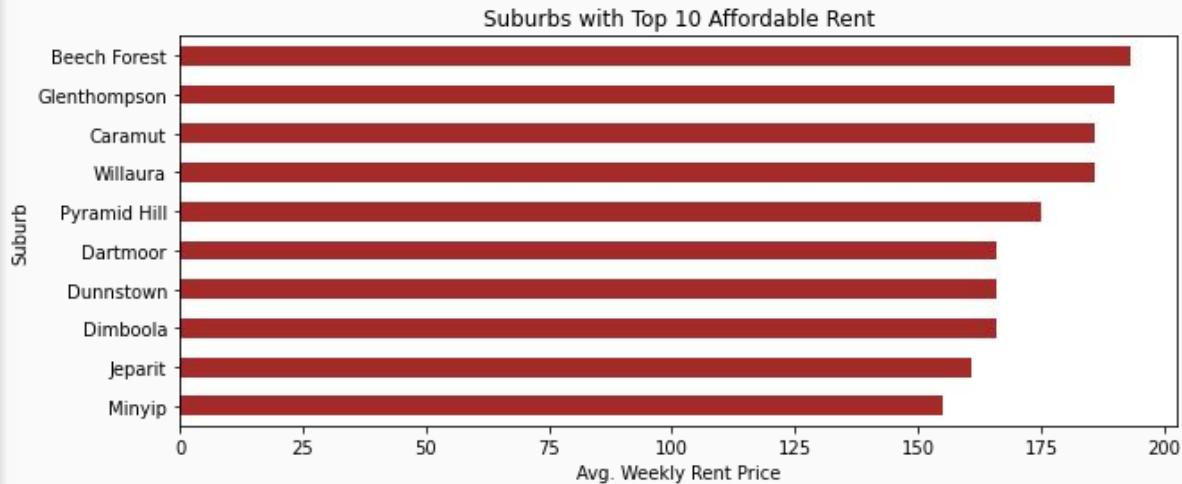
- Walkability
- Public Transport
- Education
- Affordability
- Employment



Affordability

Main Chosen Aspects:

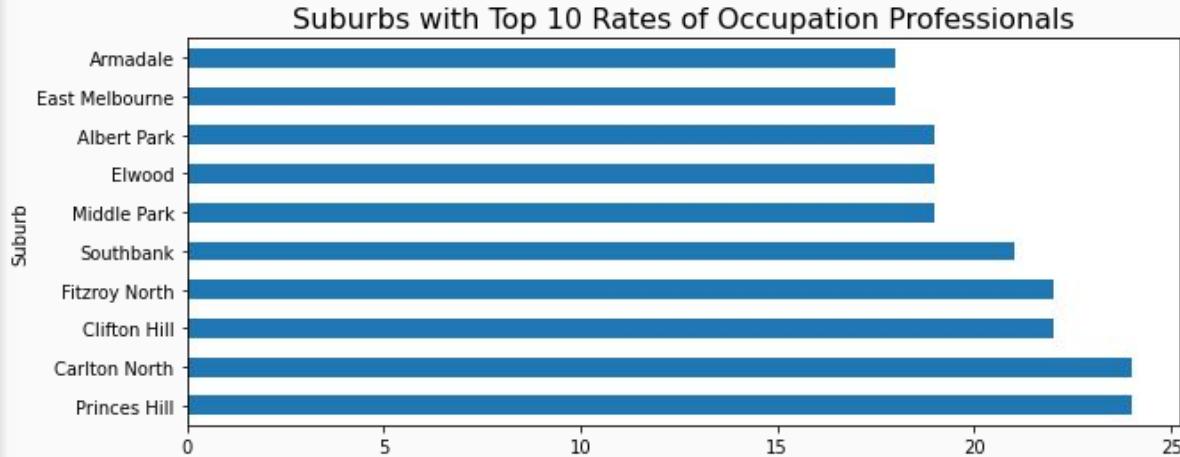
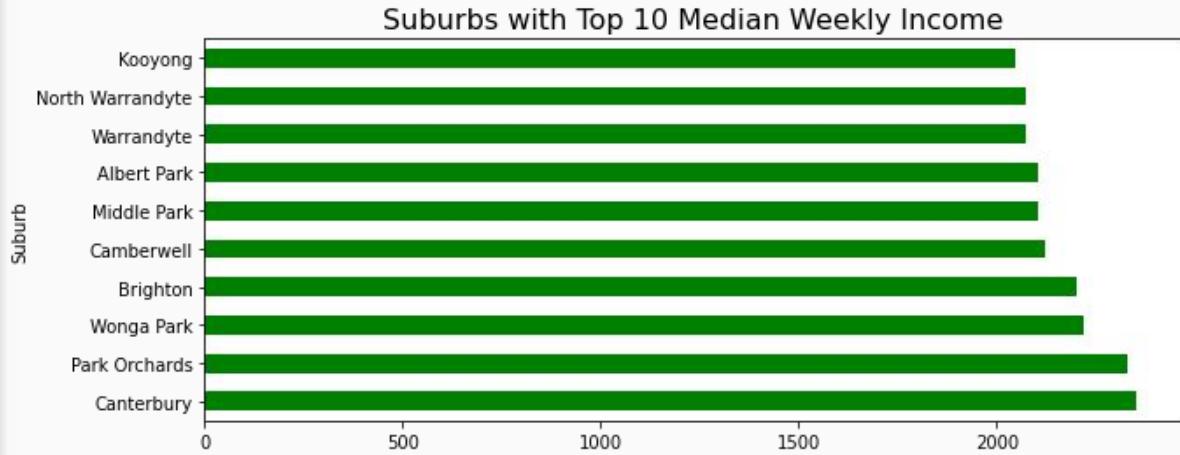
-  Walkability
-  Public Transport
-  Education
-  **Affordability**
-  Employment



Employment

Main Chosen Aspects:

-  Walkability
-  Public Transport
-  Education
-  Affordability
-  Employment



Liveability Index

Ranking by Averaging Attribute Percentiles (Equal Importance)

Suburb	Education	Income	Sold	Rent	Professionals	Schools	Train	Tram	Bus	Stations	Entertainment	Mean
Melbourne	Top 0.2%	Top 71.5%	Top 52.7%	Top 71.1%	Top 10.4%	Top 1.8%	Top 0.0%	Top 0.0%	Top 6.9%	Top 0.6%	Top 0.0%	Top 19.6%
Northcote	Top 4.8%	Top 12.5%	Top 91.3%	Top 91.3%	Top 3.8%	Top 3.8%	Top 0.2%	Top 1.0%	Top 4.1%	Top 2.6%	Top 2.8%	Top 19.8%
Richmond	Top 3.8%	Top 6.1%	Top 87.7%	Top 92.3%	Top 5.1%	Top 3.8%	Top 0.2%	Top 0.1%	Top 21.7%	Top 5.0%	Top 0.1%	Top 20.5%
Hawthorn	Top 0.7%	Top 9.6%	Top 98.3%	Top 95.6%	Top 0.9%	Top 13.1%	Top 2.8%	Top 0.2%	Top 21.7%	Top 5.9%	Top 5.1%	Top 23.1%
Essendon	Top 10.7%	Top 11.9%	Top 90.1%	Top 81.8%	Top 7.1%	Top 0.6%	Top 0.8%	Top 1.1%	Top 2.8%	Top 1.8%	Top 99.9%	Top 28.1%
Camberwell	Top 4.8%	Top 0.6%	Top 97.3%	Top 96.0%	Top 1.7%	Top 1.8%	Top 0.2%	Top 0.1%	Top 7.9%	Top 1.7%	Top 99.9%	Top 28.4%
Kew	Top 3.8%	Top 1.6%	Top 98.1%	Top 96.2%	Top 1.7%	Top 0.1%	Top 99.9%	Top 0.4%	Top 4.5%	Top 2.2%	Top 5.1%	Top 28.5%
Brunswick	Top 1.0%	Top 23.6%	Top 83.9%	Top 87.4%	Top 7.1%	Top 99.9%	Top 0.8%	Top 1.0%	Top 4.7%	Top 2.8%	Top 5.1%	Top 28.9%
Eltham	Top 14.9%	Top 3.3%	Top 81.5%	Top 82.0%	Top 7.1%	Top 3.8%	Top 12.5%	Top 99.9%	Top 7.3%	Top 8.1%	Top 5.1%	Top 29.6%
Belgrave	Top 20.4%	Top 21.8%	Top 59.9%	Top 74.7%	Top 10.4%	Top 3.8%	Top 12.5%	Top 99.9%	Top 11.3%	Top 12.0%	Top 5.1%	Top 30.2%

Ranking by Weighted Importance

[Interactive Dashboard](https://suburb-liveability.herokuapp.com/)
[\(https://suburb-liveability.herokuapp.com/\)](https://suburb-liveability.herokuapp.com/)

Conclusion and Future Direction

Using

- Real Estate Data
- PTV Stations
- Suburb Metrics
- ABS Statistics
- OpenStreetMap Features

We have answered the 3 questions

Most Important Internal Predictors

- Last Sold Price
- Building Size
- Standard Room Features
- Special Features

Most Important External Predictors

- Professional occupation rate
- Property sold value
- Median household income
- Population: Proven with a singular case study in Colac-Otway, but further analysis into other high-forecasted growth areas required

Suburb Future Growth

- Suburb was an appropriate granularity from the outset, however smaller areas led to smaller, more partitioned categories which impacted aggregates and analysis
- Only features from the ABS were used to forecast growth, so a future study could take other possibly relevant datasets/features into account

Liveability

- Top 3 : **Melbourne, Northcote & Richmond** by Equal Importance Metric
- [Interactive Dashboard](#) to allow ranking with subjective importance