# Applied Data Science with Real Estate Industry

*MAST30034 Group 43*
*Tutor: Lucas Fern*

THE UNIVERSITY OF MELBOURNE

# Overview

1. ***Introduction***
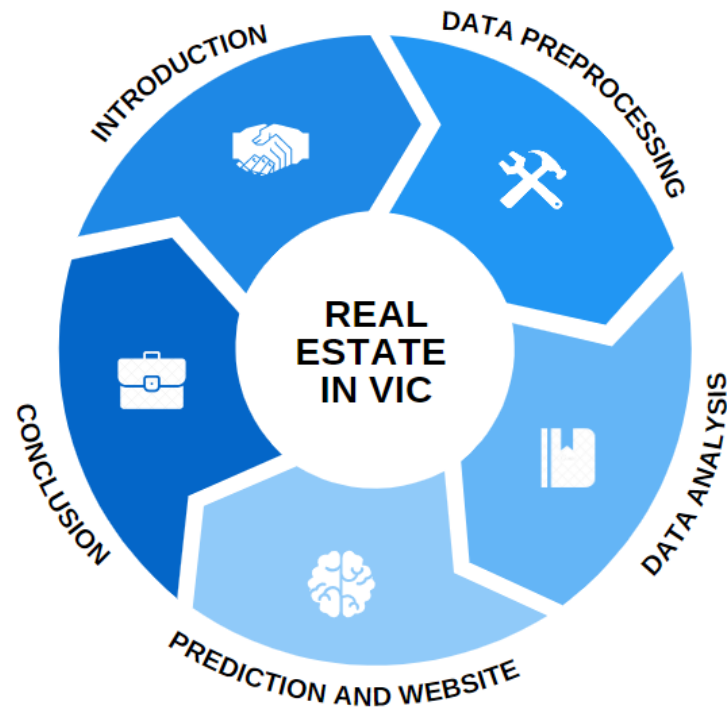   a. Our Team
   b. Our Project Goal
   c. Our Project Timeline
2. ***Approach***
   a. Data Downloading
   b. Data Preprocessing & Analysis
   c. Modelling & Website
3. ***Conclusion***
   a. Limitations
   b. Improvements & Recommendations

# Our team

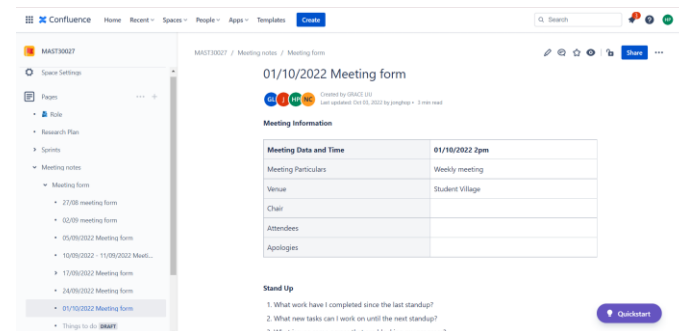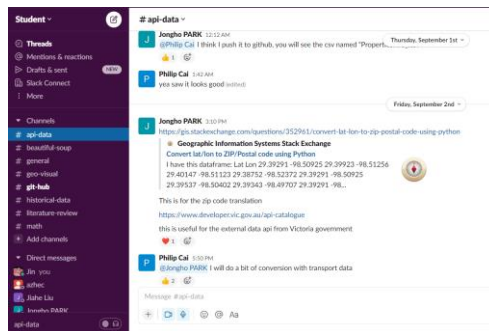| Jiahe Liu (Grace) | Anzhe Cai (Philip) | Jongho Park (John) | Hyunjin Park (Jin) | Nuo Chen (Katherine) |
|---|---|---|---|---|
| Team 2: Livability Ranking System | Team 1: Rental Price Prediction | Team 2: Livability Ranking System | Team 1: Rental Price Prediction | Team 1: Rental Price Prediction |

# Communication & Task Management Tools
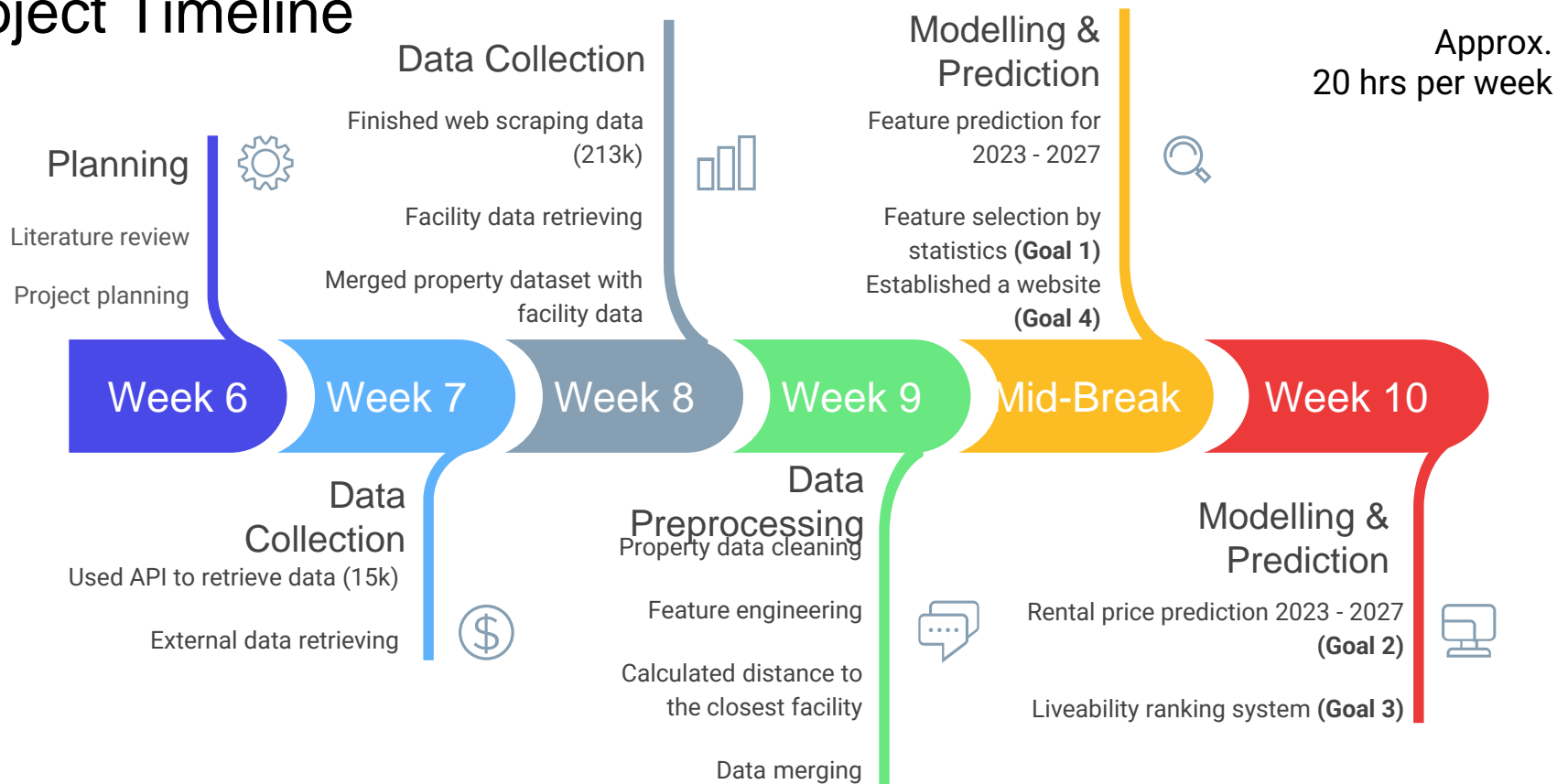
> **Slack, Confluence, GitHub**

# Project Goal

1. **Influential factors** on rental price

2. Suburbs with the highest rental price

   **growth rate** (2023-2027)

3. Most **liveable and affordable** suburbs

   & properties

4. Client-oriented **website** https://smart-

   renting-vic.herokuapp.com/
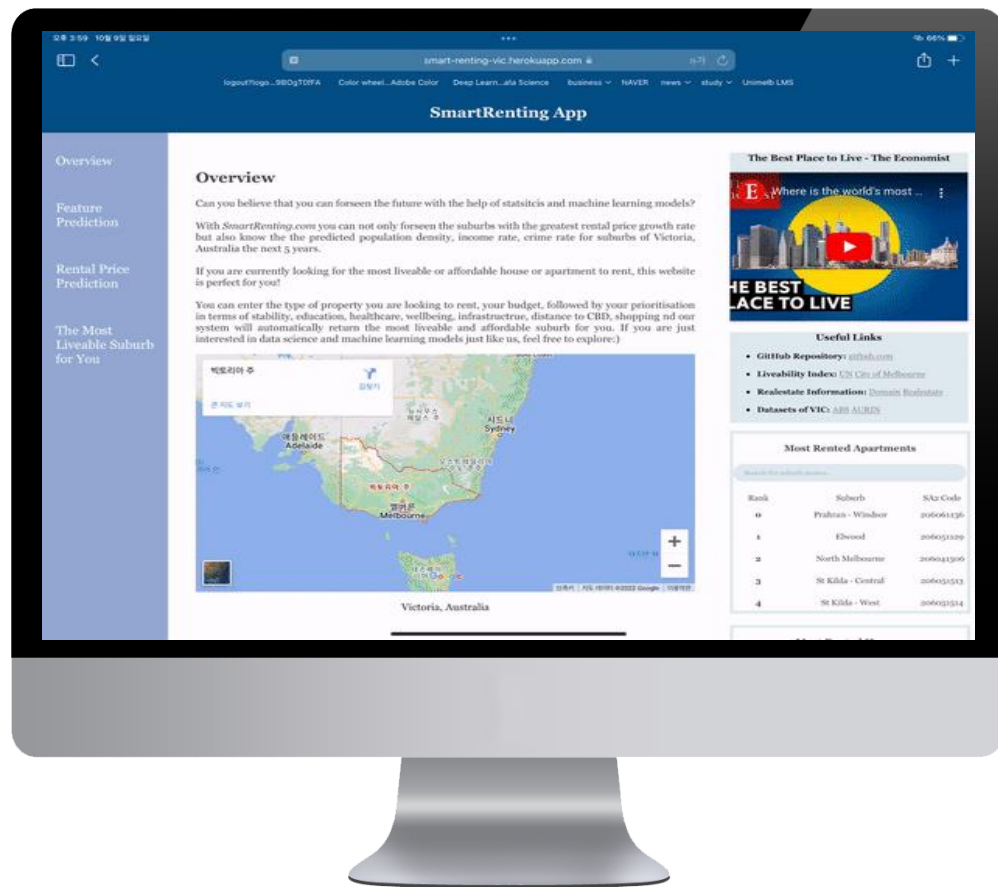
# Approach - Rental Data Downloading & Filtering

| | Initial number of data | Dropping duplicates & non weekly-rent data | Filtering for residential property data | Number of outlier removed |
|---|---|---|---|---|
| API Retrieving | 15k | 15k → 14.6k (2.6% removed) | 345k → 282k (18.2% removed) | 282k ->**231k** (18% removed; 1-3% removed per year) |
| Web Scraping | 572k | 572k → 339k (40.6% removed) | | |
| Total | **587k** | 587k → 345k (41.2% removed) | | |

Assumption for **residence type**: rural rentings, holiday rentings are NOT residential property rentings.

Assumption for **residence type**: properties with no property type label are residential properties

- Internal Features - Number of beds, bathrooms, car spaces and residence type.
- Target response - Weekly rent

# Approach - External Data Collection

**Liveability Influential Factor**

| Facilities & Index | Corresponding livability factor |
|---|---|
| Primary/secondary schools | Education |
| Parks, Shopping malls | Culture and environment |
| Hospitals | Healthcare |
| Police stations, Train stations | Stability and Infrastructure |

Assumption for **facilities**: we will have the SAME facility dataset as 2022 (schools, hospitals etc.) in the next 5 years

Assumption for **SA2:** we will have the SAME SA2 index in the next 5 years as SA2 (2021)

- Retrieved through API url in GeoJson format
- Cumulative dataset from 2013 - 2022 (Registered dates)
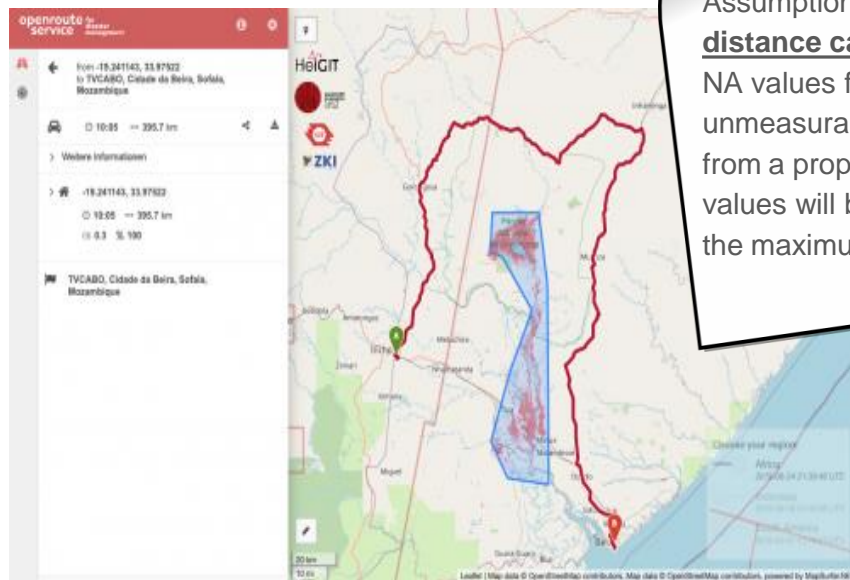
# Approach - Retrieving Distance with Open Route

Purpose:

● Measure liveability factors

Methods:

● Iterative by each SA2 suburb
● The distance to facilities (Police station, Hospital, etc..) and CBD
● Select the closest distance



Assumption for **distance calculation**: NA values for unmeasurable distances from a property. those values will be filled by the maximum distance.

# Approach - External Data Collection

**Economic Influential Factor**

| External Datasets | Year | Unique to |
|---|---|---|
| Total income | 2012 - 2019 | SA2 (2016) & Year |
| Crime Cases | 2013 - 2022 | Year & Postcode |
| GDP | 1970 - 2020 | Year |
| Saving rate | 1970 - 2020 | Year |
| Population | 2001 - 2021 | SA2 (2021) & Year |

- All external attributes are downloaded from URLs.
- Selected Range of Year: 1970 - 2022
- Missing years were filled by predictions

**Avg weekly rent for Victoria (2013-2022)**
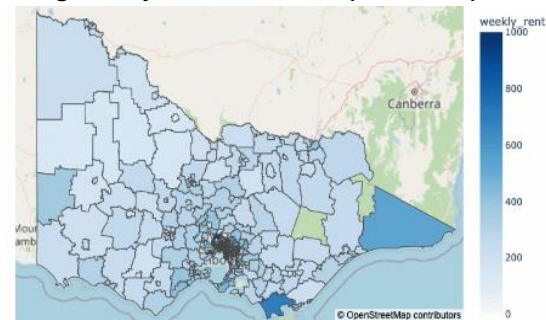


**Avg weekly rent for Melbourne(2013-2022)**
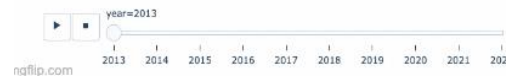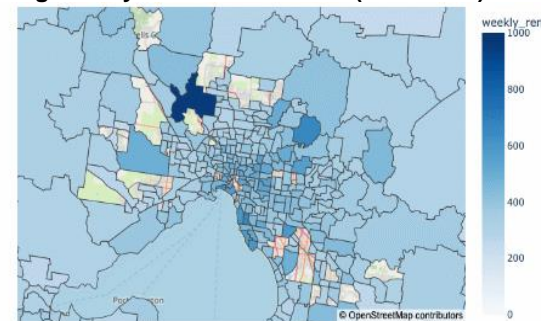
# Approach - Data Preprocessing

## Rental Dataset Cleaning

| Internal Features | Method | Data type |
|---|---|---|
| Residence Types | Keep House/Apartments, classify undefined to the two types | String |
| Bedroom number | String convert to floats | Float number |
| Bathroom number | String convert to floats | Float number |
| Car space number | String convert to floats | Float number |
| Listed price | Regex to extract weekly rents | Float number |

Assumption for **Residence Types**: the majority of property consists of house and apartment, thus classify unknown to these two

# Approach - Data Preprocessing

AHURi Australian Housing and Urban Research Institute

Australian Bureau of Statistics

## External Feature Engineering

| Economic Features | Method | Type |
|---|---|---|
| Income per person per SA2 (2016) | Total income / The number of persons (earners) | Float number |
| Population density per SA2 (2021) | The number of people / SA2 (2021) area size | Float number |
| Crime cases | Original data (offence count) | Integer |
| GDP | Original data | Integer |
| Saving rate | Original data | Float number |

Assumption for **income**: later year will be used, i.e 2012-2013 will be assumed as 2013

# Approach - Granularity

Friendly for general public's knowledge (suburb name)


Location

One property belongs to
**_one suburb_** & **_one postcode_**

One suburb may contain
**_many properties_** & **_many postcodes_**

One suburb has a
**_unique SA2 code_**

**Property** —— Coordinates —— **Suburb (SA2 Name)** —— **SA2 Code**
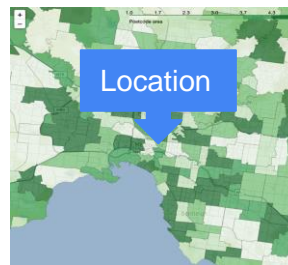
Friendly for statistical analysis

Address
nbeds
nbaths
ncarpark
Residence type
(SA2 code)
(Postcode)

Coordinates

**Postcode** —— **External Feature**


Location

Crime cases
Income rate
Population density
Facilities etc.
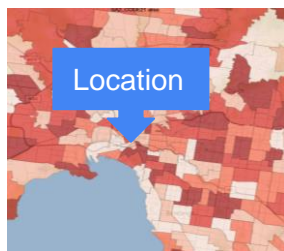
# Approach - Feature selections with Different Models

- **Ordinary Least Square**

  Forward selection with lowest AIC

- **XGboost**

  Forward selection with lowest Mean Squared Error (MSE)

- **Random Forest Feature selection**

  Forward selection with lowest Mean Squared Error (MSE)

*" What are the most important **internal and external features** in predicting **rental prices**? "*

| Method (Forward Selection) | Internal | External | | | | |
|---|---|---|---|---|---|---|
| | # bedrooms<br><br># bathrooms<br><br># carparks<br><br>Residence type | Distance to: park, CBD, train station, primary school, hospital | Distance to: police station, Shopping mall | Distance to secondary school | Income per person<br><br>Crime cases<br><br>Saving rate<br><br>Population density | GDP |
| OLS - AIC | ✅ | ✅ | ✅ | ❌ | ✅ | ✅ |
| XGboost - MSE | ✅ | ✅ | ❌ | ✅ | ✅ | ✅ |
| Random Forest - MSE | ✅ | ✅ | ✅ | ✅ | ✅ | ✅ |

# Approach - Which Model to use?

- Training & testing dataset: 2013-2022
- ***Rental price ~ internal factors and external factors***

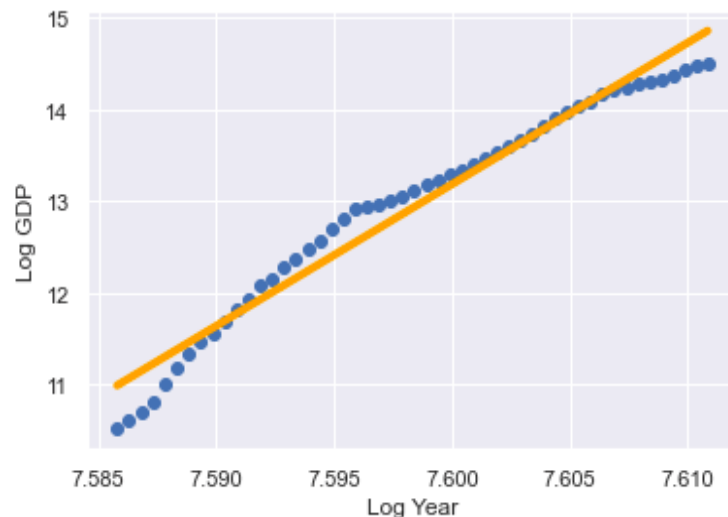| Model | OLS Regression | Random Forest Regressor | XGboost Regressor |
|---|---|---|---|
| **Training R²** | 0.998 | 0.955 | 0.742 |
| **Testing R²** | 0.544 | 0.727 | 0.697 |

# Approach - Predicted External Feature (2023 - 2027)

- **Linear Regression Models:**

  - Income per person ~ SA2 + year

  - Population density ~ SA2 + year

  - Crime cases ~ postcode + year

  - log(GDP) ~ log(year)

- **Quadratic Regression Model:**

  - Saving rate ~ year^2



| SA2 | year | income_pp | popu_dens | gdp | saving rate | crime cases |
|-----|------|-----------|-----------|-----|-------------|-------------|
| 201011001 | 2023 | 68806.0 | 630.0 | 3567759 | 13.58 | 7754 |

# Approach - Constant features (2022, 2023-2027)

| SA2 | Residence Type | nbed | nbath | ncar | distance _cbd | distance_ park | etc |
|-----|---------------|------|-------|------|--------------|---------------|-----|
| 215011390 | 0 | 2 | 1 | 0 | 25 | 14 | … |
| 215011390 | 0 | 2 | 1 | 1 | 28 | 12 | … |
| 215011390 | 1 | 3 | 1 | 0 | 26 | 7 | … |
| 215011390 | 0 | 3 | 1 | 1 | 20 | 11 | … |
| … | … | … | … | … | … | … | … |

Where Residence Type 0 = House, Residence Type 1 = Apartment

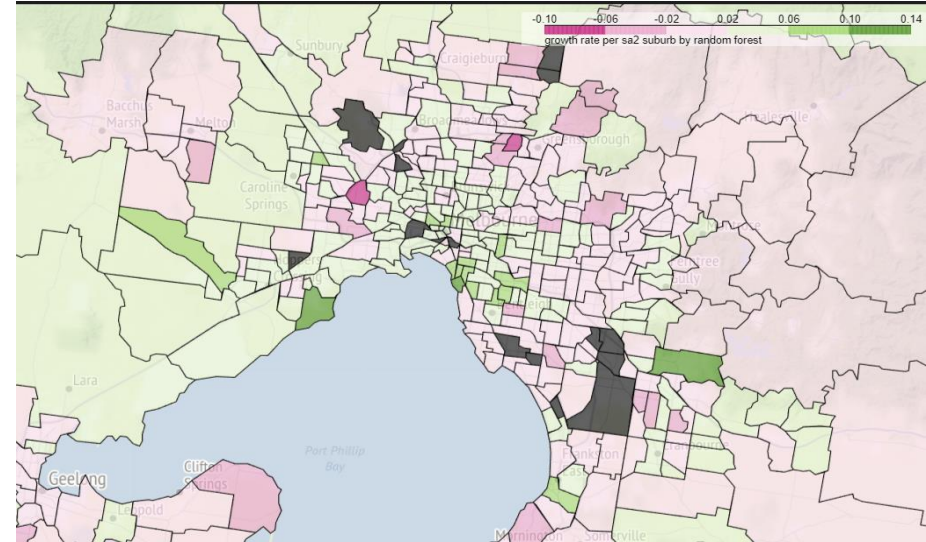# Visualisation - Average Growth Rates by SA2 (2023-2027)
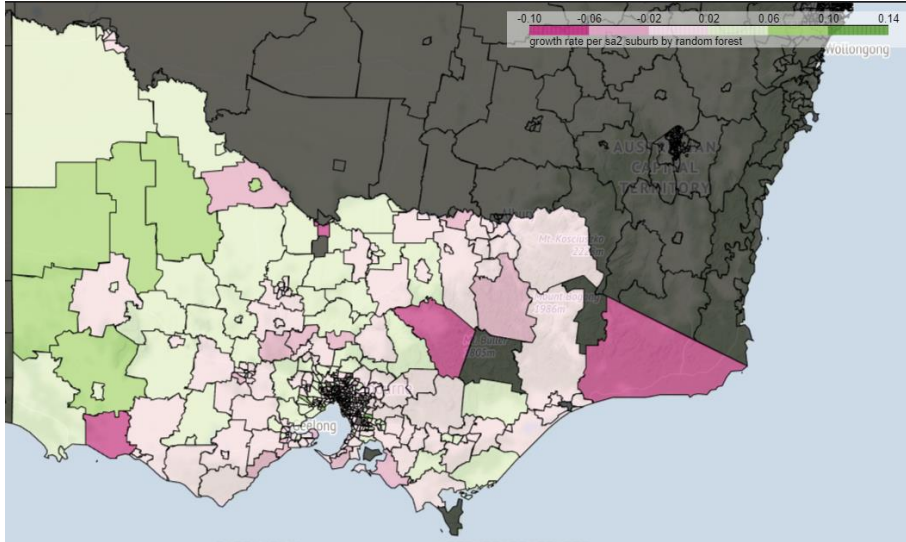


> **Avg growth rate in Victoria (2023-2027)**

**Dark Green: Positive growth rate**
**Light green, pink: Mild growth rate**
**Dark Pink: Negative growth rate**

> **Avg growth rate in Melbourne (2023-2027)**

**Dark Green: Positive growth rate**
**Light green, pink: Mild growth rate**
**Dark Pink: Negative growth rate**

# Approach - Random Forest Rg. & top 10 growing suburbs

*" What are the top 10 suburbs (SA2) with the highest predicted growth rate? "*

| | | | | |
|---|---|---|---|---|
| **1st** | Narre Warren (North): 14.75% | | **6th** | Hawthorn (East): 8.55% |
| **2nd** | Point Cook (East): 14.16% | | **7th** | Kerang: 8.37% |
| **3rd** | Elwood: 10.16% | | **8th** | St Kilda (Central): 7.70% |
| **4th** | Yarriambiack: 9.96% | | **9th** | St Kilda (West): 7.69% |
| **5th** | Nhill Region: 8.92% | | **10th** | Frankston: 7.49% |

# Approach -
# Livability Scoring and
# Ranking

*" What are the most liveable and affordable suburbs according to your chosen metrics? "*

https://smart-renting-vic.herokuapp.com/



| Criteria | Preference |
|---|---|
| Preferred residence type | Please choose your preferred residence type: Apartment ⌄ |
| Weekly rent budget | Please choose your preferred weekly rent budget: under $400 per week ⌄ |
| Stability - distance to police station: 10 | Not important(1) ━━━━━○ Extremely important(10) |
| Stability - number of crime cases: 10 | Not important(1) ━━━━━○ Extremely important(10) |
| Education - distance to primary school : 1 | Not important(1) ○━━━━━ Extremely important(10) |
| Education - distance to secondary school: 1 | Not important(1) ○━━━━━ Extremely important(10) |
| Infrastructure - distance to train stations: 6 | Not important(1) ━━━○━━ Extremely important(10) |
| Healthcare - distance to hospitals: 4 | Not important(1) ━━○━━━ Extremely important(10) |
| Wellbeing - distance to parks: 4 | Not important(1) ━━○━━━ Extremely important(10) |
| Distance to City: 10 | Not important(1) ━━━━━○ Extremely important(10) |
| Shopping - distance to shopping malls: 6 | Not important(1) ━━━○━━ Extremely important(10) |

# Approach - Livability Scoring and Ranking

*" What are the most liveable and affordable suburbs according to your chosen metrics? "*

https://smart-renting-vic.herokuapp.com/



The top 5 recommended suburbs for you to rent at are (with liveability score out of 100):

1. Yarram, 70.37
2. Loddon, 70.32
3. Corangamite - North, 69.3
4. Wattle Glen - Diamond Creek, 68.38
5. East Melbourne, 68.35

The top 5 recommended properties for you to rent are (with liveability score out of 100):

1. 13/657 BRUNSWICK STREET NORTH, FITZROY NORTH , Apartment, $360.0 per week, 77.71
2. 20/104 THE AVENUE, PARKVILLE , Apartment, $340.0 per week, 76.42
3. 605/150 BRUNSWICK STREET, FITZROY , Apartment, $344.0 per week, 76.31
4. 5/64 STEPHEN STREET, YARRAVILLE , Apartment, $315.0 per week, 75.39
5. 5/81 BELLAIRS AVENUE, YARRAVILLE , Apartment, $305.0 per week, 75.12

averaged livability score per sa2

19   27   36   44   53   62   70

| Property Name | Distance to CBD (km) | Score for 'Distance to CBD" | Crime Cases (cases per year) | Rank for "Crime Cases" | Score for "Crime Cases" | Distance to the closest train station(km) | Rank for "Distance to the closest train station" | Score for "Distance to the closest train station" |
|---|---|---|---|---|---|---|---|---|
| Victoria One | 0.73 | **3/3** | 14323 | 1 | **1/3** | 0.55 | 3 | **3/3** |
| Student Village | 1.44 | **2/3** | 2227 | 2 | **2/3** | 0.63 | 2 | **2/3** |
| Sky One | 19.99 | **1/3** | 1804 | 3 | **3/3** | 0.72 | 1 | **1/3** |

Victoria One:  ⅓ x 1 + ⅓ x ⅓ + ⅓ x 1 = 0.7778 (livability score 77.78)

https://smart-renting-vic.herokuapp.com/

Student Village  ⅓ x ⅔ + ⅓ x ⅔ + ⅓ x ⅔ = 0.6667 (livability score 66.67)

Sky One:  ⅓ x ⅓ + ⅓ x 1 + ⅓ x ⅓ =0.5556 (livability score 55.56) (⅓ refers to the weight of each criterion)
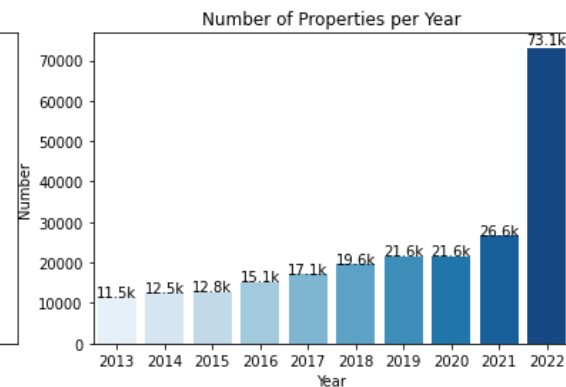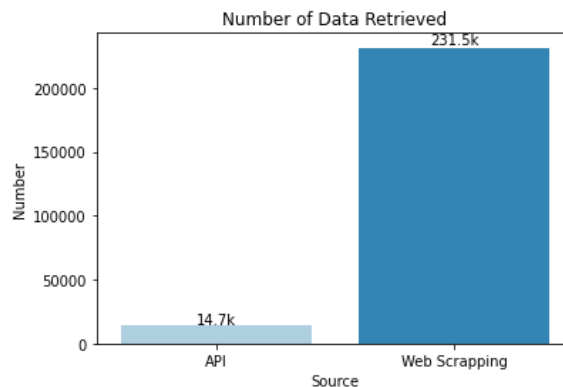
# Business Application - Industry *"Needs Pyramid"*
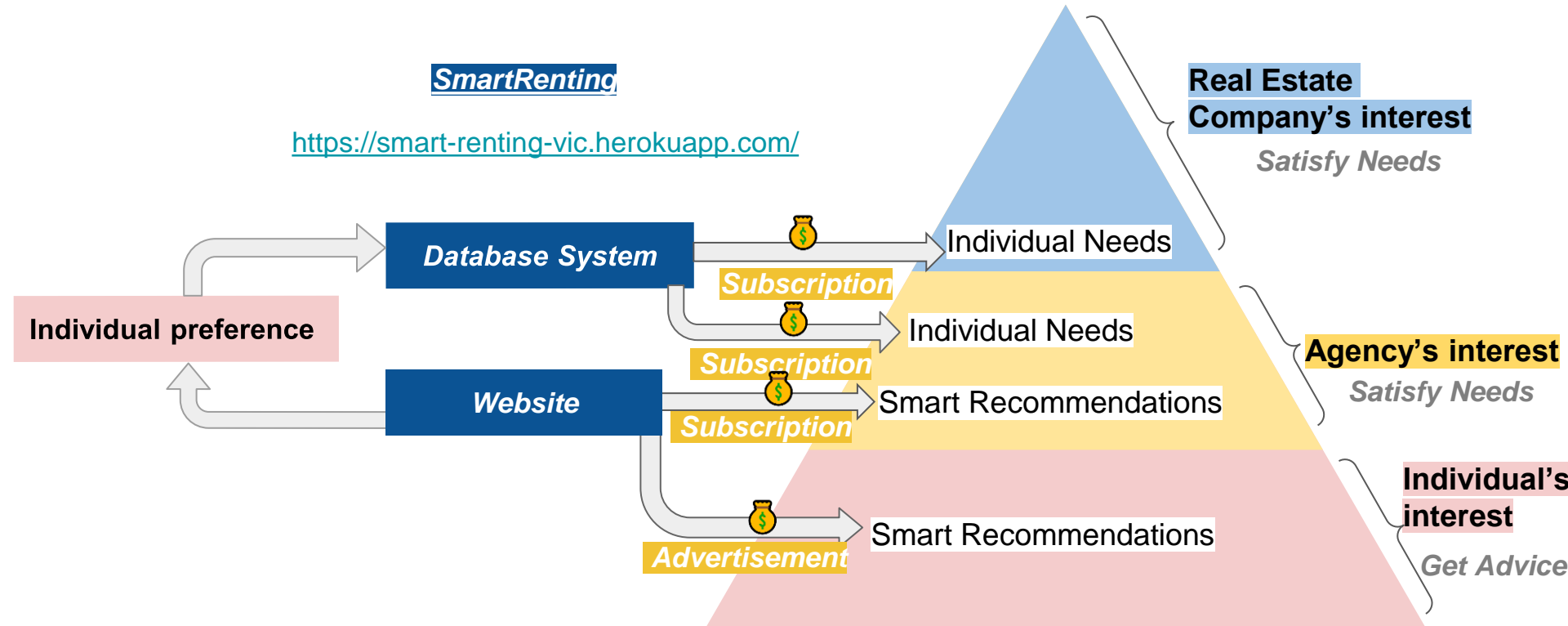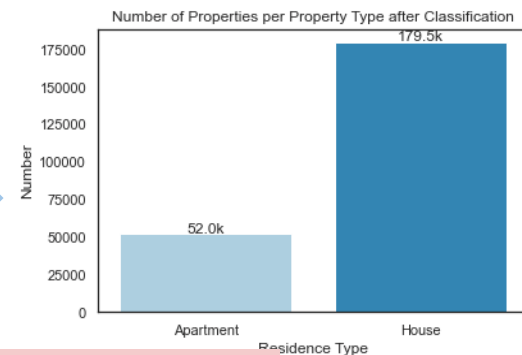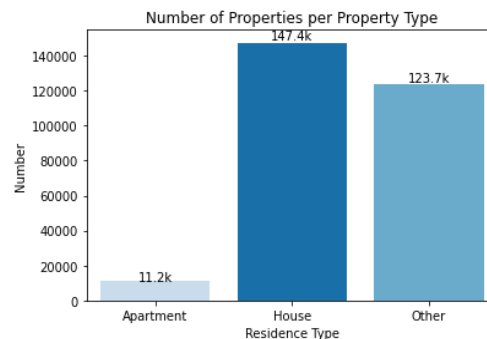
# Limitations

- **API**
  - Limited API call quotas

- **Data**
  - Imbalanced dataset
  - Unstructured Web-scraping data

- **Feature Classification**
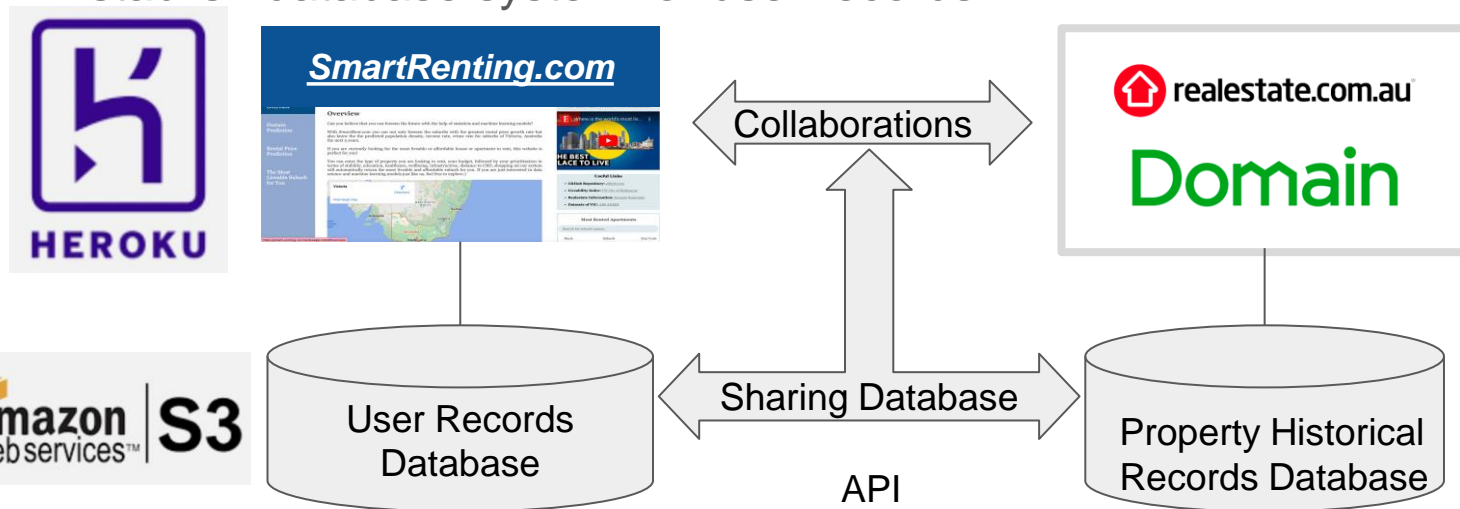  - Misclassification in residence type



Imbalanced dataset



Residence Type Classification

# Improvements & Recommendation

- Improve property data accuracy

- Collaborations with real estate companies

- Establish database system for user records

# Reference

Bezerra, A., Alves, G., Silva, I., Rosati, P., Endo, P. T., & Lynn, T. (2019). A preliminary exploration of uber data as an indicator of urban liveability. 2019 International Conference on Cyber Situational Awareness, Data Analytics And Assessment (Cyber SA). https://doi.org/10.1109/cybersa.2019.8899714

Burke, T. (1999). Private rental in australia. Terry Burke 1999, 1–18. https://www.semanticscholar.org/paper/Private-rental-in-Australia- Burke/34328916dff5c5115e808a8ed79bebf9806ae0ff

Editor. (2021, June 4). Asking rent vs effective rent explained. Office Spaces Singapore. Retrieved August 30, 2022, from https://officespaces.com.sg/what-is-the-difference- between-asking-rent-vs-effective-rent/

For Sale For Lease. (2018, January 31). How to determine property rental price. Retrieved August 23, 2022, from https://forsaleforlease.com.au/how-to/how-to-determine-property- rental-price/

Gampala, V., Sai, N. Y., & Bhavya, T. N. S. (2022). Real-estate price prediction system using machine learning. 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), 533–538. https://doi.org/10.1109/icaaic53929.2022

Higgs, C., Badland, H., Simons, K., Knibbs, L. D., & Giles-Corti, B. (2019). The urban liveability index: Developing a policy-relevant urban liveability composite measure and evaluating associations with transport mode choice. International Journal of Health Geographics, 18(1). https://doi.org/10.1186/s12942-019-0178-8

Hulse, K., Martin, C., James, A., & Stone, W. (2018). Private rental in transition: Institutional change, technology and innovation in australia. AHURI Final Report, 296. https://doi.org/10.18408/ahuri-5112101

Ooi, G. L., & Yuen, B. (2009). World cities: Achieving liveability and vibrancy: Achieving liveability and vibrancy. Wspc/Others.

Razaghi, T. (2022, August 18). If property prices are going down, why are rents going up? The Sydney Morning Herald. Retrieved August 23, 2022, from https://www.smh.com.au/property/news/if-property-prices-are-going-down-why-are- rents-going-up-20220816-p5badd.html

Tan, K. Y., & Woo, W. T. (2012). Ranking the liveability of the world's major cities. Van Haren Publishing.

Thome, J., Li, A., Sivaraman, V., & Bridge, C. (2014). Mobile crowdsourcing older people's opinions to enhance liveability in regional city centres. 2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP). https://doi.org/10.1109/issnip.2014.6827675

Victoria State Government. (2022, March). Rental report statistics – march quarter 2022. Homes Victoria, Victorian Government, Australia. https://www.dffh.vic.gov.au/publications/rental-report

Australian Bureau of Statistics. 2022. *Housing Occupancy and Costs, 2019-20 financial year*. [online] Available at: <https://www.abs.gov.au/statistics/people/housing/housing-occupancy-and-costs/latest-release> [Accessed 8 October 2022]. https://www.abs.gov.au/statistics/people/housing/housing-occupancy-and-costs/latest-release

Hulse, K., Martin, C., James, A. and Stone, W., 2018. *Private rental in transition: institutional change, technology and innovation in Australia | AHURI*. [online] Ahuri.edu.au. Available at: <https://www.ahuri.edu.au/research/final-reports/296> [Accessed 8 October 2022].