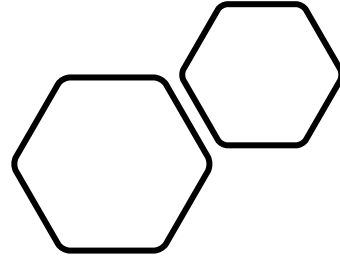


Capstone Project: The Battle of Neighbourhoods



An analysis of neighbourhoods near
Melbourne CBD using data science.

Phang Mun Yun

May 31, 2021

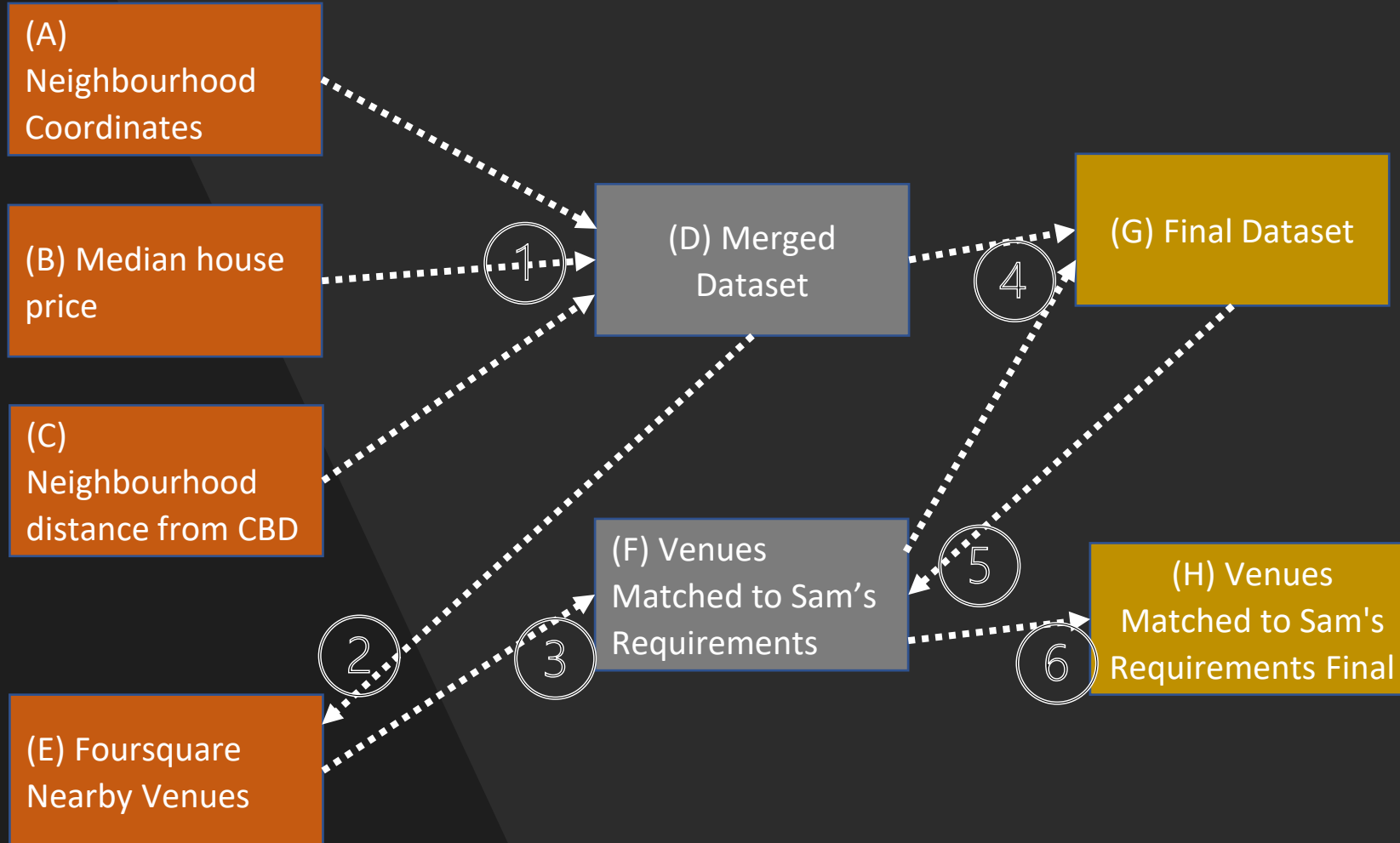
Buying your first home is stressful...

- Sam needs help to find a suitable neighbourhood in Melbourne (Victoria, Australia)
- There more than 300 neighbourhoods in the Greater Melbourne region alone!
- Sam's requirements:
 - Within 25 KMs from the Central Business District (CBD)
 - Budget of A\$ 1 million
 - Nearby venues:
 1. Gym
 2. Park
 3. Grocery store
 4. Japanese restaurant
 5. Train station
- Which neighbourhood should he focus on his in-depth research?

Data Sourcing

1. Neighbourhood coordinates from [Corra](#)
2. Neighbourhood distance from CBD from [Myboot](#)
3. Nearby venues from [Foursquare](#)
4. Median house price from [Victoria State Government](#)

Data Cleaning



- Each dataset is cleaned so that it can merge with other datasets
- Unmerge values are removed

Outcome

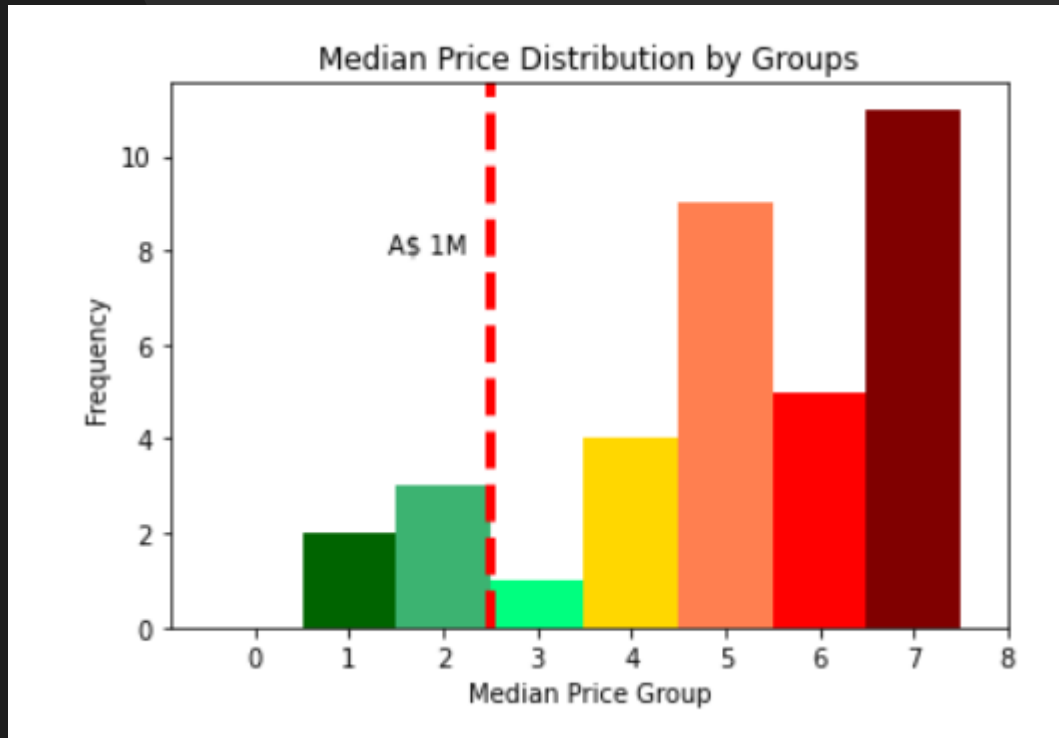
(G) Final Dataset

- 35 neighbourhoods

(H) Venues Matched

- 403 venues

Findings – Median Price Distribution

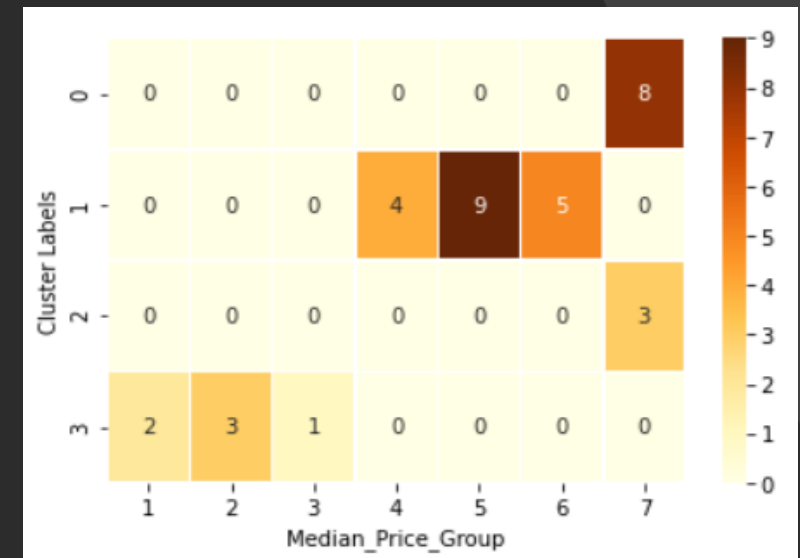


- Only a minority ($5/35 = 15\%$ = 1 out of 7) of the neighbourhoods shortlisted are within Sam's budget.
- There are 2 neighbourhood groups that might be too expensive for Sam, that is neighbourhoods in Group 6 & 7.
- Nevertheless, we will not remove these datapoints as they are median prices only, and there might be properties priced around Sam's budget, when he continues his research on the listing websites.

Findings – Cluster Analysis

Cluster Labels	Grocery Store	Gym	Japanese Restaurant	Park	Train/Tram Station	km	lat	lon	Median_Price	Median_Price_Group
3	4.000000	2.000000	1.666667	1.666667	1.500000	12.200000	-37.827170	145.059342	9.179333e+05	1.833333
1	3.444444	1.888889	2.111111	2.000000	2.277778	8.500000	-37.821884	145.027613	1.305606e+06	5.055556
0	3.000000	1.000000	1.750000	2.250000	3.000000	7.337500	-37.838766	145.018720	1.853500e+06	7.000000
2	3.000000	1.666667	2.333333	2.333333	3.666667	5.366667	-37.848668	145.015247	2.532500e+06	7.000000

- Based on the table above, we can observe that as km from CBD is greater, the median price decreases.
- Cluster 1 looks interesting, as its distribution of shortlisted venues are greater than 1.8 across all categories, suggesting a more balanced cluster than the others.
- I would recommend him to start with cluster 1 & 3 first, as these neighbourhoods are within his budget (around median price).



Conclusion

- The purpose of this project is to prepare a shortlist of neighborhoods for Sam to consider based on his needs.
- I was able to provide a more refined list of neighborhoods that I feel Sam would like.
 - From over 300 neighborhoods near Melbourne to less than 25 neighborhood recommendations, Sam will be able to find his dream home very soon!
- Some potential improvement data points that could increase the value of this project:
 1. Friends and family location by neighborhood data
 2. Available homes for sale & price data
 3. Crime statistics by neighborhood data
 4. Neighborhood demographics data