

Applied Data Science Capstone Project

The Battle of Neighborhoods

Recommending Suitable Location Purchase a Property in Melbourne

Jia Xuan Tan (July 2020)

Table of contents

- Introduction / Business Problem
- The Data
- Methodology
- Result
- Discussion
- Conclusion

1. Introduction / Business Problem Section

1.1 Background

Melbourne is a diverse city and is a truly magnificent city in which to live, work and study. In 2017, Melbourne was once again ranked the world's most livable city by the Economist Intelligence Unit's (EIU) Global Liveability Index since the index began in 2002. In 2016 Melbourne achieved perfect scores in healthcare, education and infrastructure while it outranked Sydney in the areas of stability, and culture & environment.

With property prices falling for the last two months, and the economy will spiral further if the Covid-19 outbreak continues to spread, some experts are saying now isn't the right time to snap up a new home.

However, according to hotspotting.com.au managing director Terry Ryder, he believes that it's a great time to buy real estate in many locations around Australia now, particularly regional centers and the smaller capital cities, and especially for first-home buyers. Metropole Property Strategists CEO Michael Yardney also said "any time" could be either the worst time or the best time for you to buy property.

It really depends on your own goals, budget, timeline, risk profile and circumstances as to whether 2020 is a good time to buy.

1.2 Business Problem

In this case, a machine learning tool would be able to assist homebuyers in Melbourne to make effective decisions. As a result, the goal of this project is to develop machine learning algorithms to provide support to homebuyers in Melbourne to purchase a suitable and profitable real estate in this uncertain economic situation?

To start, we will be clustering Melbourne suburbs in order to recommend venues. We will recommend suitable venues according to various factors need such as :

- Housing Prices in Melbourne
- Demographics of Melbourne
- Nearby Venues/Facilities

1.3 Target Audience

The objective is to locate and recommend to interested home buyers / property investors which suburb of Melbourne will be the best choice to invest in. The users also expects to understand the rationale of the recommendations made.

This would interest anyone who wants to purchase properties in Melbourne.

2. The Data

For the below analysis we will get data from Kaggle as given below:

- Melbourne Housing Sales Price: https://www.kaggle.com/anthonypino/melbourne-housing-market?select=Melbourne_housing_FULL.csv

To explore and target recommended locations across different venues according to the presence of amenities and facilities, we will access data through FourSquare API interface and arrange them as a dataframe for visualization.

- Nearby Facilities/Venues

By merging data on Melbourne housing prices by suburb and data on amenities and facilities surrounding such properties from FourSquare API interface, we will be able to recommend profitable locations to invest in a property.

3. Methodology

3.1 Setting up the environment

```
Collecting package metadata (current_repodata.json): done
Solving environment: done
```

```
# All requested packages already installed.
```

```
Collecting package metadata (current_repodata.json): done
Solving environment: done
```

```
# All requested packages already installed.
```

```
Folium installed
Libraries imported.
```

3.2 Explore and Understand Data

Read the dataset that we collected from Kaggle "Melbourne Housing Market" into a pandas data frame and display the first five rows of it as follows:

[2]:

	Suburb	Address	Rooms	Type	Price	Method	SellerG	Date	Distance	Postcode	...	Bathroom	Car	Landsize	Bu
0	Abbotsford	68 Studley St	2	h	NaN	SS	Jellis	03-09-16	2.5	3067.0	...	1.0	1.0	126.0	
1	Abbotsford	85 Turner St	2	h	1480000.0	S	Biggin	03-12-16	2.5	3067.0	...	1.0	1.0	202.0	
2	Abbotsford	25 Bloomburg St	2	h	1035000.0	S	Biggin	04-02-16	2.5	3067.0	...	1.0	0.0	156.0	
3	Abbotsford	18/659 Victoria St	3	u	NaN	VB	Rounds	04-02-16	2.5	3067.0	...	2.0	1.0	0.0	
4	Abbotsford	5 Charles St	3	h	1465000.0	SP	Biggin	04-03-17	2.5	3067.0	...	2.0	0.0	134.0	

5 rows × 21 columns

< >

5 rows × 21 columns

(34857, 21)

Our dataset consists of over 300000 rows and 21 columns. We will now prepare and preprocess data accordingly.

3.3 Data preparation and preprocessing

Now, prepare the dataset for modeling process, opting for the most suitable machine learning algorithm.

	Suburb	Address	Rooms	Type	Price	Method	SellerG	Date	Distance	Postcode	...	Bathroom	Car	Landsize	Bu
1	Abbotsford	85 Turner St	2	h	1480000.0	S	Biggin	03-12-16	2.5	3067.0	...	1.0	1.0	202.0	
2	Abbotsford	25 Bloomburg St	2	h	1035000.0	S	Biggin	04-02-16	2.5	3067.0	...	1.0	0.0	156.0	
4	Abbotsford	5 Charles St	3	h	1465000.0	SP	Biggin	04-03-17	2.5	3067.0	...	2.0	0.0	134.0	
5	Abbotsford	40 Federation La	3	h	850000.0	PI	Biggin	04-03-17	2.5	3067.0	...	2.0	1.0	94.0	
6	Abbotsford	55a Park St	4	h	1600000.0	VB	Nelson	04-06-16	2.5	3067.0	...	1.0	2.0	120.0	

5 rows × 21 columns

Next, we drop unnecessary columns like Address, Rooms, Type, Method, SellerG, Date...

	Suburb	Address	Rooms	Price	Date	Postcode	Lattitude	Longitude
1	Abbotsford	85 Turner St	2	1480000.0	03-12-16	3067.0	-37.7996	144.9984
2	Abbotsford	25 Bloomburg St	2	1035000.0	04-02-16	3067.0	-37.8079	144.9934
4	Abbotsford	5 Charles St	3	1465000.0	04-03-17	3067.0	-37.8093	144.9944
5	Abbotsford	40 Federation La	3	850000.0	04-03-17	3067.0	-37.7969	144.9969
6	Abbotsford	55a Park St	4	1600000.0	04-06-16	3067.0	-37.8072	144.9941

Calculate the average prices of houses by location (show the first 5 rows):

	Suburb	Avg_Price
0	Abbotsford	1.096604e+06
1	Aberfeldie	1.354793e+06
2	Airport West	7.805294e+05
3	Albanvale	5.360556e+05
4	Albert Park	1.983665e+06

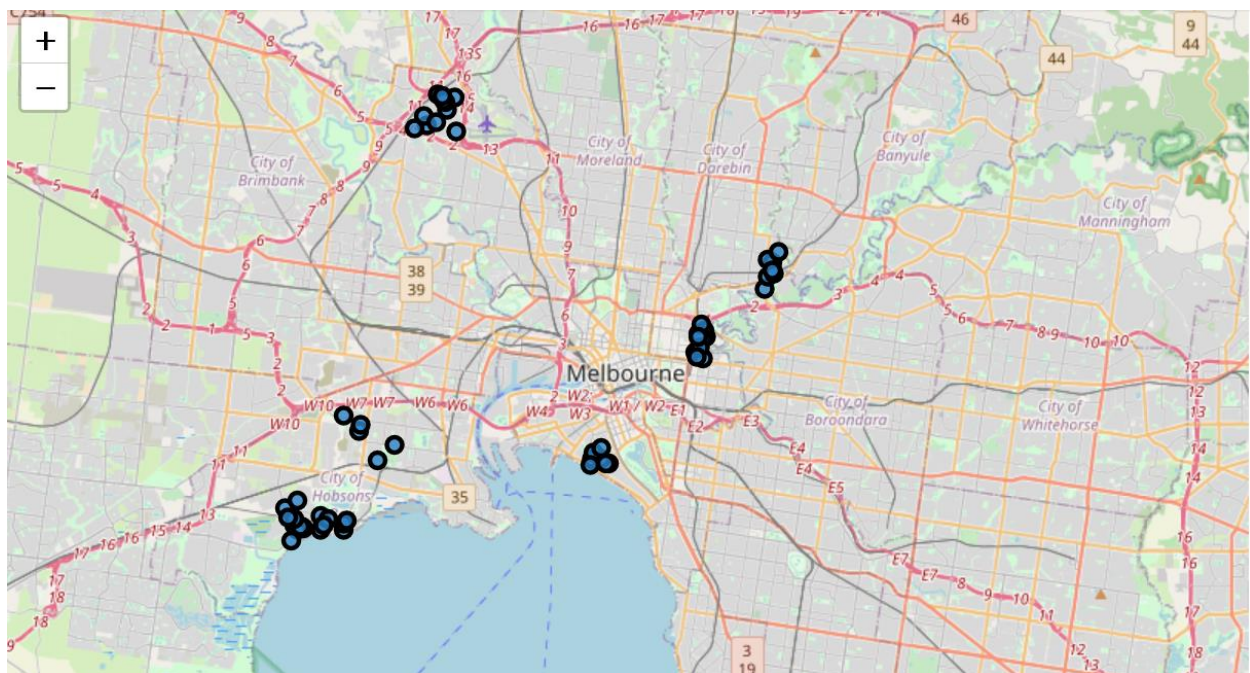
Next, set the budget that we are willing to spend on buying a house (let say \$1 Million to \$1.5 million):

	Suburb	Avg_Price
0	Abbotsford	1.096604e+06
1	Aberfeldie	1.354793e+06
6	Alphington	1.441156e+06
12	Ascot Vale	1.100420e+06
14	Ashwood	1.220920e+06
...
327	Wildwood	1.030000e+06
329	Williamstown	1.368712e+06
331	Windsor	1.055295e+06
333	Wonga Park	1.357500e+06
337	Yarraville	1.012868e+06

96 rows × 2 columns

There are 96 locations where the houses fit in the budget of \$1 Million to \$1.5 million.

Now, we create a map of Melbourne:



3.5 Modelling

After exploring the dataset, we can use the clustering method to analyze property. We will use k-means clustering to analyze as it is simple and efficient in terms of computational cost, and it is highly flexible to account for evolutions in property market in Melbourne.

Using foursquare API, we will analyze neighborhoods to recommend locations where home buyers can make a property investment. We will then recommend profitable venues according to amenities and essential facilities surrounding such venues schools, restaurants, hospitals & grocery stores.

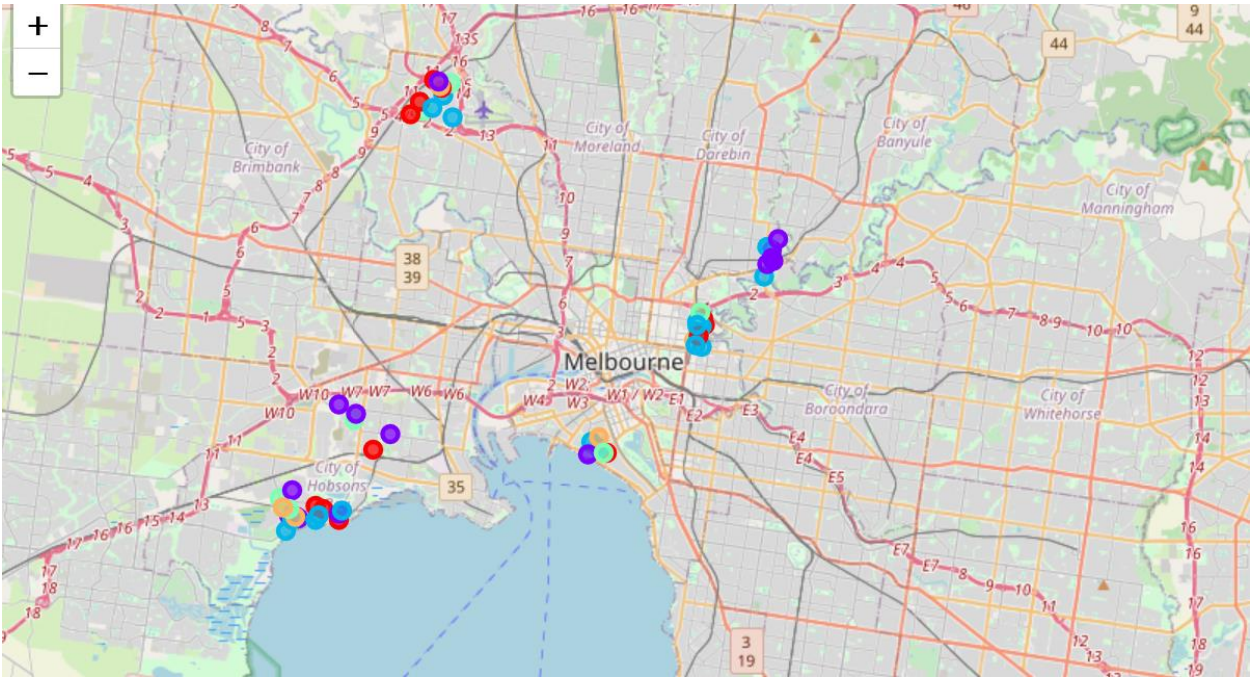
	Suburb	Suburb Latitude	Suburb Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Aberfeldie	-37.7996	144.9984	Retreat Hotel	-37.801126	144.997548	Pub
1	Aberfeldie	-37.7996	144.9984	Rita's Cafeteria	-37.799978	144.994047	Pizza Place
2	Aberfeldie	-37.7996	144.9984	Yarra Hotel	-37.800361	144.996311	Pub
3	Aberfeldie	-37.7996	144.9984	Lentil As Anything	-37.802724	145.003507	Vegetarian / Vegan Restaurant
4	Aberfeldie	-37.7996	144.9984	Lulie St Tavern	-37.799914	144.994818	Dive Bar
...
1365	Yarraville	-37.8327	144.8451	Ferguson Plarre Bakehouse	-37.827991	144.847502	Coffee Shop
1366	Yarraville	-37.8327	144.8451	Takechiho	-37.828310	144.848455	Sushi Restaurant
1367	Yarraville	-37.8327	144.8451	Aldi	-37.827834	144.847558	Supermarket
1368	Yarraville	-37.8327	144.8451	EB Games	-37.827546	144.847611	Video Game Store
1369	Yarraville	-37.8327	144.8451	Noodle Box	-37.827628	144.848191	Asian Restaurant

1370 rows × 7 columns

We find the top 10 venues/facilities nearby each housing location

	Suburb	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Aberfeldie	Café	Pub	Farmers Market	Dive Bar	Chinese Restaurant	Rock Climbing Spot	Coffee Shop	Record Shop	Cultural Center	Pizza Place
1	Alphington	Vietnamese Restaurant	Café	Thai Restaurant	Korean Restaurant	Pub	Chinese Restaurant	Asian Restaurant	Park	Vegetarian / Vegan Restaurant	Bakery
2	Ashwood	Vietnamese Restaurant	Café	Thai Restaurant	Korean Restaurant	Grocery Store	Pub	Bakery	Brewery	Vegetarian / Vegan Restaurant	Chinese Restaurant
3	Bentleigh	Café	Pub	Pizza Place	Farmers Market	Music Venue	Rock Climbing Spot	Coffee Shop	Record Shop	Cultural Center	Park
4	Bentleigh East	Vietnamese Restaurant	Café	Thai Restaurant	Pub	Korean Restaurant	Bar	Brewery	Asian Restaurant	Bakery	Vegetarian / Vegan Restaurant

After our inspection of venues/facilities nearby the most profitable property in Melbourne, we can start by clustering properties by venues/facilities nearby.



From the map, we can visualize that there 5 different clusters.

Below are the top 10 most common venues for each cluster.

Cluster 0:

	Avg_Price	1sb Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	1.354793e+06	Café	Pub	Farmers Market	Dive Bar	Chinese Restaurant	Rock Climbing Spot	Coffee Shop	Record Shop	Cultural Center	Pizza Place
29	1.349966e+06	Café	Pub	Pizza Place	Farmers Market	Music Venue	Rock Climbing Spot	Coffee Shop	Record Shop	Cultural Center	Park
57	1.308933e+06	Vietnamese Restaurant	Café	Thai Restaurant	Pub	Vegetarian / Vegan Restaurant	Grocery Store	Korean Restaurant	Asian Restaurant	Bakery	Gastropub
81	1.336622e+06	Gym / Fitness Center	Grocery Store	Park	Women's Store	Football Stadium	Electronics Store	Farm	Farmers Market	Fast Food Restaurant	Fish & Chips Shop
108	1.303871e+06	Café	Gym / Fitness Center	Fast Food Restaurant	Shopping Mall	Women's Store	Football Stadium	Electronics Store	Farm	Farmers Market	Fish & Chips Shop

Cluster 1:

	Avg_Price	1sb Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
	1.031231e+06	Department Store	Supermarket	Fast Food Restaurant	Donut Shop	Portuguese Restaurant	Paper / Office Supplies Store	Optical Shop	Coffee Shop	Sandwich Place	Café
	1.024457e+06	Café	Beach	Italian Restaurant	Light Rail Station	Hotel Bar	Farmers Market	Pier	Middle Eastern Restaurant	Deli / Bodega	Fishing Spot
	1.000880e+06	Café	Beach	Light Rail Station	Hotel Bar	Italian Restaurant	Park	Farmers Market	Fishing Spot	Middle Eastern Restaurant	Breakfast Spot
	1.001500e+06	Market	Train Station	Liquor Store	Café	Thai Restaurant	Convenience Store	Park	Fast Food Restaurant	Farmers Market	Gym / Fitness Center
	1.030773e+06	Café	Thai Restaurant	Liquor Store	Gym / Fitness Center	Indian Restaurant	Fish Market	Train Station	Greek Restaurant	Farmers Market	Convenience Store

Cluster 2:

	Avg_Price	1sb Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
30	1.140140e+06	Vietnamese Restaurant	Café	Thai Restaurant	Pub	Korean Restaurant	Bar	Brewery	Asian Restaurant	Bakery	Vegetarian / Vegan Restaurant
35	1.133078e+06	Café	Pub	Brewery	Pizza Place	Hotel Bar	Gastropub	Park	Music Venue	Japanese Restaurant	Gym
36	1.101050e+06	Café	Pub	Brewery	Pizza Place	Farmers Market	Hotel Bar	Garden	Park	Music Venue	Japanese Restaurant
51	1.176425e+06	Café	Pub	Brewery	Pizza Place	Grocery Store	Greek Restaurant	Gay Bar	Gastropub	Indian Restaurant	Football Stadium
58	1.115553e+06	Café	Pub	Brewery	Pizza Place	Hotel Bar	Gay Bar	Record Shop	Japanese Restaurant	Indian Restaurant	Gym

Cluster 3:

	Avg_Price	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
14	1.220920e+06	Vietnamese Restaurant	Café	Thai Restaurant	Korean Restaurant	Grocery Store	Pub	Bakery	Brewery	Vegetarian / Vegan Restaurant	Chinese Restaurant
54	1.222446e+06	Café	Pub	Convenience Store	Record Shop	Furniture / Home Store	Dive Bar	Rock Climbing Spot	Coffee Shop	Scenic Lookout	Football Stadium
72	1.262471e+06	Sports Club	Gym / Fitness Center	Grocery Store	Park	Business Service	Women's Store	Electronics Store	Farm	Farmers Market	Fast Food Restaurant
79	1.226199e+06	Department Store	Fast Food Restaurant	Supermarket	Donut Shop	Café	Food Truck	Food Court	Moving Target	Light Rail Station	Shopping Mall
140	1.186375e+06	Café	Beach	Italian Restaurant	Park	Convenience Store	Light Rail Station	Hotel Bar	Deli / Bodega	Pier	Fishing Spot

Cluster 4:

	Avg_Price	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
	1.441156e+06	Vietnamese Restaurant	Café	Thai Restaurant	Korean Restaurant	Pub	Chinese Restaurant	Asian Restaurant	Park	Vegetarian / Vegan Restaurant	Bakery
	1.415362e+06	Department Store	Donut Shop	Supermarket	Fast Food Restaurant	Optical Shop	Shopping Mall	Sandwich Place	Electronics Store	Coffee Shop	Big Box Store
	1.393527e+06	Café	Italian Restaurant	Park	Convenience Store	Asian Restaurant	Gastropub	Vegetarian / Vegan Restaurant	Seafood Restaurant	Fast Food Restaurant	Light Rail Station
	1.443683e+06	Pizza Place	Convenience Store	Train Station	Park	Thai Restaurant	Café	Shopping Mall	Food Court	Donut Shop	Electronics Store
	1.395016e+06	Discount Store	Pizza Place	Convenience Store	Train Station	Thai Restaurant	Café	Shopping Mall	Food Court	Donut Shop	Electronics Store

3.6 Assigning weights to some of the categories that potential homebuyers want to consider (user input)

Weights ranging from 1 to 4, 4 being the most important category that homebuyers consider:

	Suburb	Suburb Latitude	Suburb Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	weights
0	Aberfeldie	-37.7996	144.9984	Retreat Hotel	-37.801126	144.997548	Pub	3.5
2	Aberfeldie	-37.7996	144.9984	Yarra Hotel	-37.800361	144.996311	Pub	3.5
7	Aberfeldie	-37.7996	144.9984	The Park Hotel	-37.802769	144.997029	Pub	3.5
21	Aberfeldie	-37.7996	144.9984	Mavis the Grocer	-37.803110	144.997020	Convenience Store	3.0
23	Aberfeldie	-37.7996	144.9984	Abbotsford IGA	-37.800114	144.995684	Grocery Store	4.0
...
1350	Warrandyte	-37.8451	144.8529	Aldi	-37.844628	144.845189	Supermarket	2.5
1354	Whealers Hill	-37.8345	144.8444	Woolworths	-37.835145	144.846803	Supermarket	2.5
1360	Yarraville	-37.8327	144.8451	Woolworths	-37.835145	144.846803	Supermarket	2.5
1362	Yarraville	-37.8327	144.8451	Coles	-37.827696	144.847944	Grocery Store	4.0
1367	Yarraville	-37.8327	144.8451	Aldi	-37.827834	144.847558	Supermarket	2.5

225 rows × 8 columns

Copying only the relevant columns like suburbs and weights to group all the venues by suburb and calculating the means for each suburb and merging the table for which we calculated the means of weights suburbs to the actual table that we got from Kaggle and normalize the data. (first 5 rows):

	Suburb	Avg_Price	weights
0	Aberfeldie	0.798974	0.785714
1	Alphington	0.994283	0.656250
2	Ashwood	0.496221	0.750000
3	Bentleigh	0.788057	0.785714
4	Bentleigh East	0.313535	0.694444

Calculating the differences of normalized columns to determine the suburb that has maximum difference and conclude that the suburb would be the best fit (lowest price with most desired categories chosen by potential homebuyers:

	Suburb	Avg_Price	weights	difference
36	Niddrie	0.004633	1.00	0.995367
37	North Warrandyte	0.109683	1.00	0.890317
31	Mitcham	0.210631	1.00	0.789369
13	Caulfield North	0.247393	1.00	0.752607
43	Oakleigh South	0.009575	0.75	0.740425

4. Result & Discussions

From the analysis, we may analyze our results according to the five clusters we have produced. Even though, all clusters could praise an optimal range of facilities and amenities, we have found two main patterns. The first pattern we are referring to, i.e. Clusters 3 and 4, may target home buyers who prefers to stay close to grocery stores, markets or convenient stores. Instead, the second pattern we are referring to, i.e. Clusters 0, 1 and 2, may target individuals who love pubs, cafes, and sports.

Secondly, given the nearby venues/facilities that potential homebuyers might consider when choosing a location, Niddrie is the cheapest suburb to consider, with favorable nearby venues/facilities like Grocery Store, Restaurant, Bus Stop, Train Station, Convenience Store, Pub, Supermarket, Shopping Mall, Food Court and Gym / Fitness Center. This is followed by North Warrandyte, Mitcham, Caulfield North...

5. Conclusion

To sum up, Melbourne is a diverse city and is a truly magnificent city in which to live, work and study. In 2017, Melbourne was once again ranked the world's most liveable city by the Economist Intelligence Unit's (EIU) Global Liveability Index since the index began in 2002. In 2016 Melbourne achieved perfect scores in healthcare, education and infrastructure while it outranked Sydney in the areas of stability, and culture & environment.

With property prices falling for the last two months, and the economy will spiral further if the Covid-19 outbreak continues to spread, some experts are saying now isn't the right time to snap up a new home.

However, according to hotspotting.com.au managing director Terry Ryder, he believes that it's a great time to buy real estate in many locations around Australia now, particularly regional centres and the smaller capital cities, and especially for first-home buyers. Metropole Property Strategists CEO Michael Yardney also said "any time" could be either the worst time or the best time for you to buy property.

It really depends on your own goals, budget, timeline, risk profile and circumstances as to whether 2020 is a good time to buy.

In this case, a machine learning tool would be able to assist homebuyers in Melbourne to make effective decisions. As a result, the goal of this project is to develop machine learning algorithms to provide support to homebuyers in Melbourne to purchase a suitable and profitable real estate in this uncertain economic situation?

To solve this business problem, we clustered Melbourne neighborhoods in order to recommend venues and the current average price of real estate where homebuyers can make a property investment. We recommended profitable venues according to amenities and essential facilities surrounding such venues schools, restaurants, hospitals & grocery stores.

First, we gathered Melbourne housing data on Kaggle from https://www.kaggle.com/anthonyypino/melbourne-housing-market?select=Melbourne_housing_FULL.csv. Moreover, to explore and target recommended locations across different venues according to the presence of amenities and essential facilities, we accessed data through FourSquare API interface and arranged them as a data frame for visualization. By merging data on Melbourne properties and the relative sales price data from Kaggle and data on amenities and essential facilities surrounding such properties from FourSquare API interface, we were able to recommend profitable property investments.

Secondly, in the methodology section, we used the k-means clustering technique as it is simple and efficient in terms of computational cost, is highly flexible to account for evolutions in property market in Melbourne.

Finally, we concluded two main perspectives. First, we analyzed our results according to the five clusters we produced. While Clusters 3 and 4 may target home buyers who prefer convenience in buying household items, Clusters 0, 1 and 2 may target individuals who love pubs, cafes and sports. Secondly, homebuyers can input their preferred nearby venues/facilities and find out the most suitable location within their budget using the machine learning algorithm that we developed. In this example, Niddrie is the most suitable location with all the preferred nearby venues/facilities that is within our budget.