



Applied Data Science Capstone Project

The Battle of Neighborhoods

Recommending Suitable Location Purchase a Property in Melbourne

Table of contents

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

Background

- Melbourne was once again ranked the world's most livable city by the Economist Intelligence Unit's (EIU) Global Livability Index since the index began in 2002.
- 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](https://www.hotspotting.com.au) managing director Terry Ryder, he believes that it's a great time to buy real estate especially for first-home buyers.

Business Problem

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

The Data

For the 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.

Methodology

Explore and Understand Data

Extracting data from Kaggle

[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



Calculate the average price by suburb
and set the budget that we are willing to
spend on buying a house.

(let say \$1 Million to \$1.5 million)

The table shows a list of 96 locations that have
houses within the price of \$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		

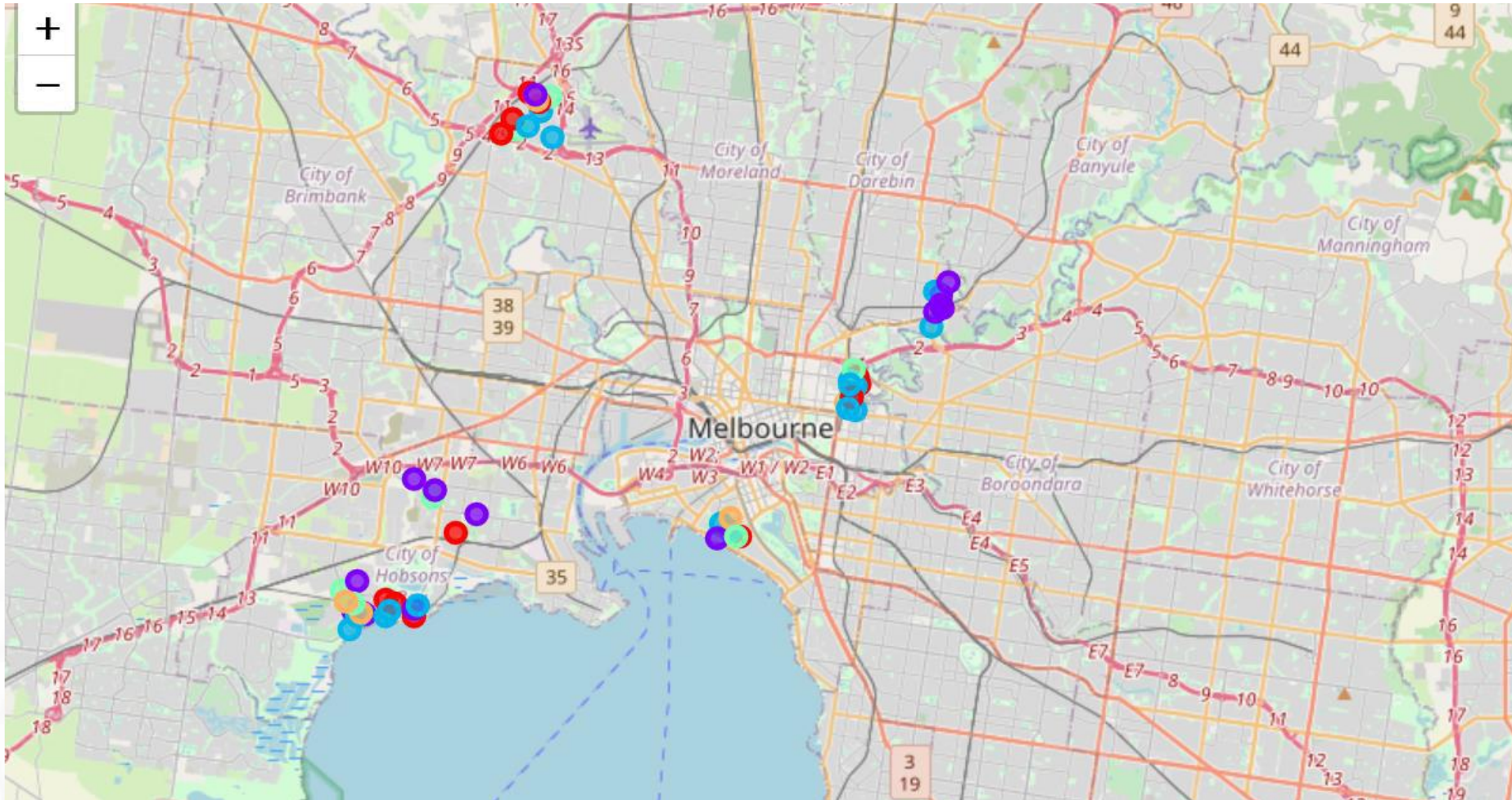
After exploring the dataset, we can use the clustering method to analyze property.

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, and 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.

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

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.

	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, from the table, we can conclude that Niddrie is the most ideal suburb to invest a property in for the given criteria.

	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

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. certain clusters such as 3, 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...