

---

---

# Analysing Melbourne Housing Market

Coursera Capstone Project  
Applied Data Science

---

by Zin Myint Naung

---

---

# Objectives

- Understand Melbourne housing price in detail
- Explore different suburbs with their average housing prices
- Find out suitable places for potential buyer based on their preferred nearby venues
- To equip real estate agent with valuable knowledge about housing prices
- To help property builders to be able to estimate turnover before developing a residential or commercial venue
- For better decision making while buying a house

# Approach

- Understand business problem
- Data preparation
  - Define our data requirement
  - Collect required data
  - Wrangle data for analytical processing
- Define methodology
- Analyse data
- Gather findings
- Presentations
- Conclusion (Recommendation for further analysis)

# Business Problem

- **Problem** - Being one of the most attractive liveable cities, yet still difficult to find relevant properties prices of all suburb
- **Solution** - To develop a data science application which can provide valuable housing information for potential buyers/sellers

# Target Audiences

- Real estate agents (a.k.a Property brokers)
- Property Developers
- Potential buyers or sellers

# Data preparation

## Data Requirement

- Suburb (List of neighbourhood)
- Housing prices for each neighbourhood
- Latitude
- Longitude
- Nearby Venues (to analyse price variation based on nearby venues)

# Data set



From Kaggle -

[https://www.kaggle.com/anthonypino/melbourne-housing-market#Melbourne\\_Housing\\_FULL.csv](https://www.kaggle.com/anthonypino/melbourne-housing-market#Melbourne_Housing_FULL.csv)

- Total of 21 columns with sold prices from previous years for each suburb
- Included latitude and longitude columns
- Total of 34,857 records

# Data Wrangling

The following steps were taken during data wrangling stage:-

- Removed rows where price value is null
- Drop columns that are not necessary for our purpose
- Calculate the *mean* value of housing price for each suburb

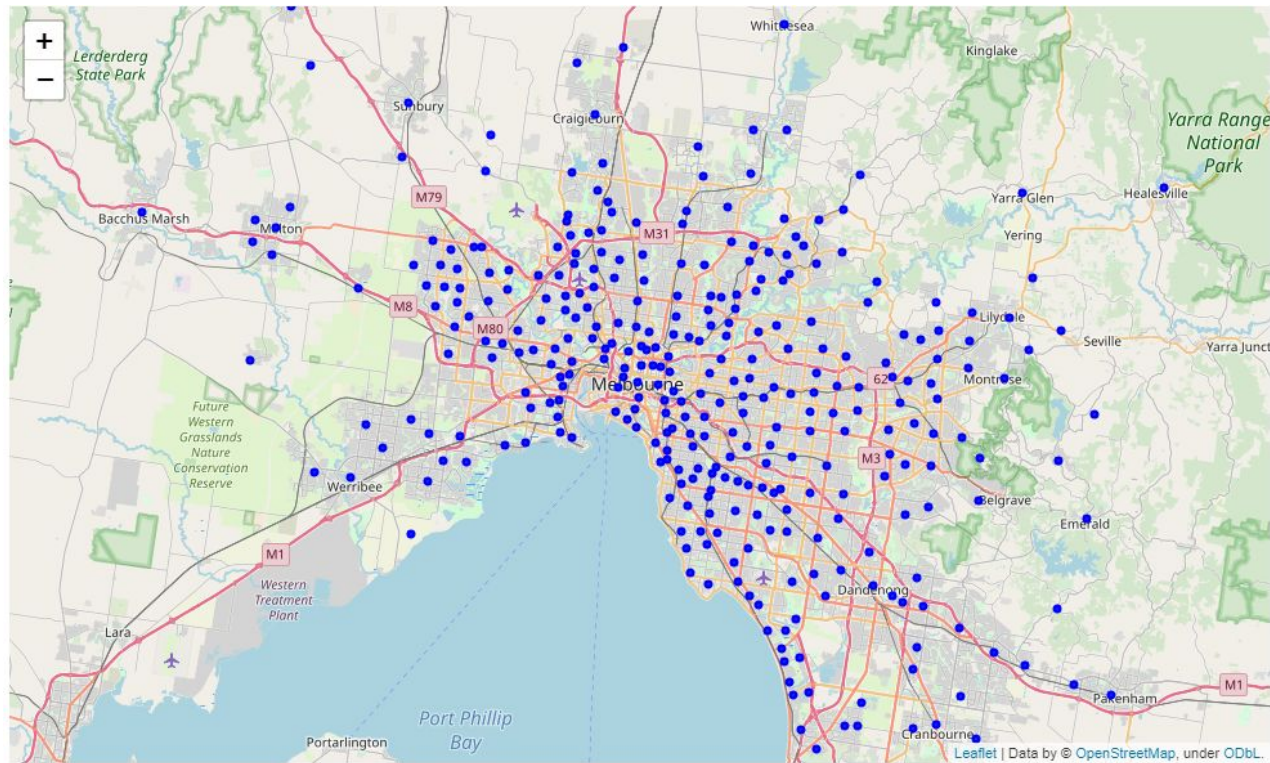
# Dataframe

- Pandas dataframe for further analysis should look like below:-
  - Suburb
  - Postcode
  - AvgPrice
  - Latitude
  - Longitude

	Suburb	Postcode	AvgPrice	Lattitude	Longitude
0	Abbotsford	3067.00	1096603.90	-37.80	145.00
1	Airport West	3042.00	780529.42	-37.72	144.88
2	Albert Park	3206.00	1983664.71	-37.84	144.95
3	Alphington	3078.00	1441155.56	-37.78	145.03
4	Altona	3018.00	872917.93	-37.87	144.82



# Visualizing Housing Price of Melbourne

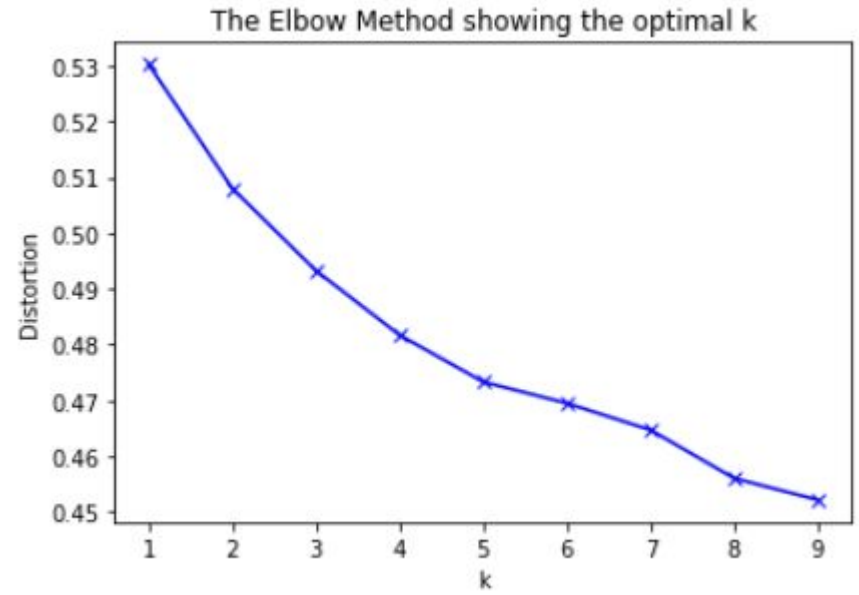


# Nearby venues data using Foursquare API

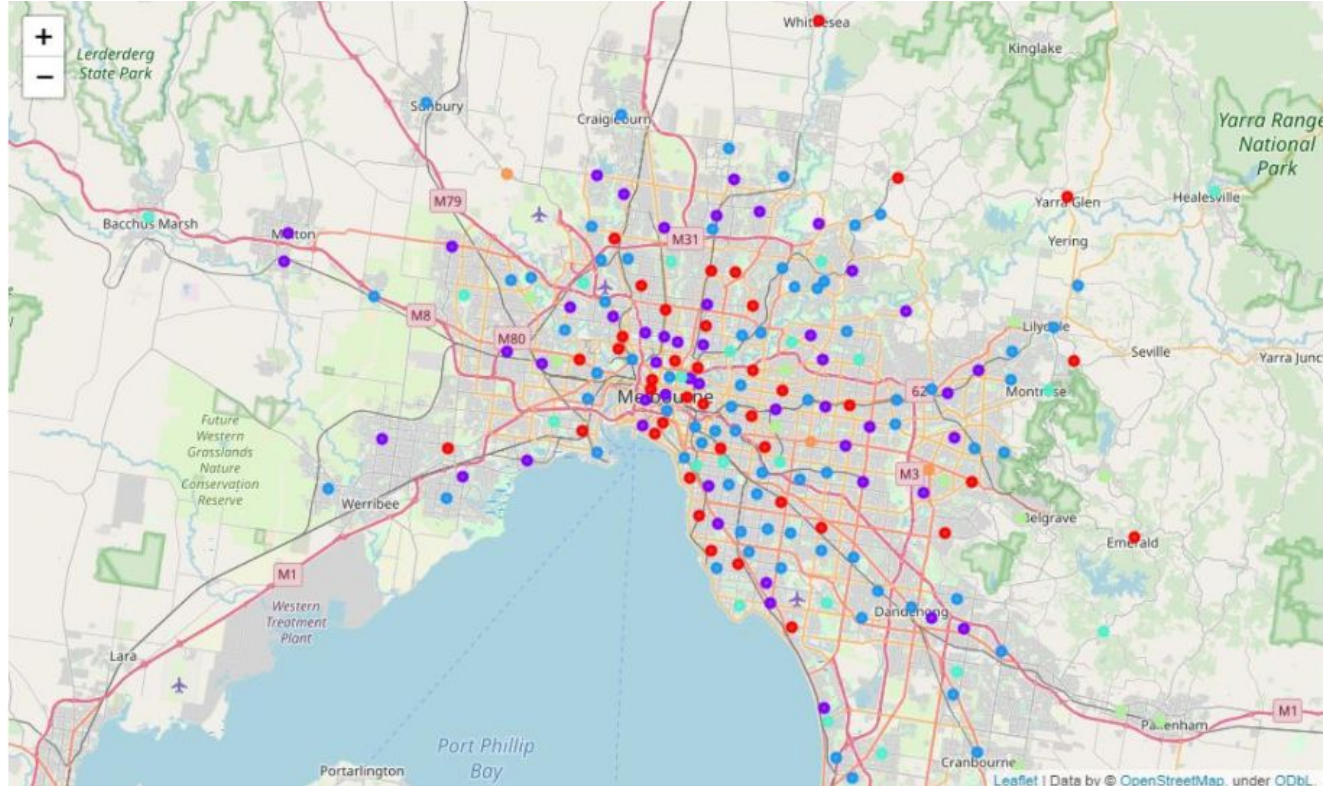
	Suburb	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Abbotsford	Pub	Café	Convenience Store	Thrift / Vintage Store	Pizza Place	Rock Climbing Spot	Coffee Shop	Bus Line	Burger Joint	Gay B
1	Aberfeldie	Café	Grocery Store	Zoo Exhibit	Food & Drink Shop	Fish & Chips Shop	Fish Market	Flea Market	Flower Shop	Food	Food Court
2	Airport West	Italian Restaurant	Pharmacy	Grocery Store	Food & Drink Shop	Fish & Chips Shop	Fish Market	Flea Market	Flower Shop	Food	Food Court
3	Albanvale	Rental Car Location	Farmers Market	Filipino Restaurant	Fish & Chips Shop	Fish Market	Flea Market	Flower Shop	Food	Food & Drink Shop	Food Court
4	Albert Park	Café	Italian Restaurant	Seafood Restaurant	Deli / Bodega	Australian Restaurant	Snack Place	Breakfast Spot	Middle Eastern Restaurant	Bookstore	Greek Resta

# Best k value for K-Means algorithm

- As shown in the graph, the best value for k is 6

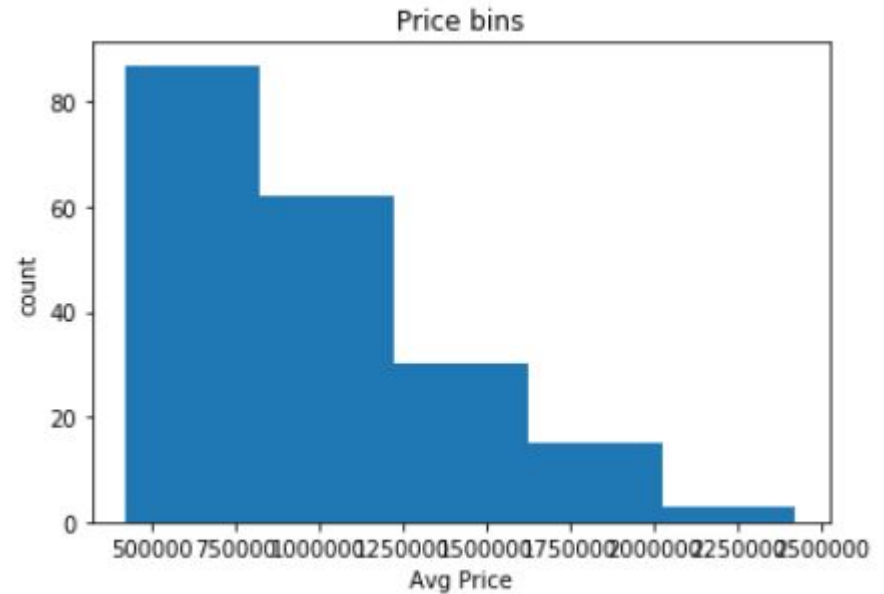


# Cluster map of Melbourne's nearby venues



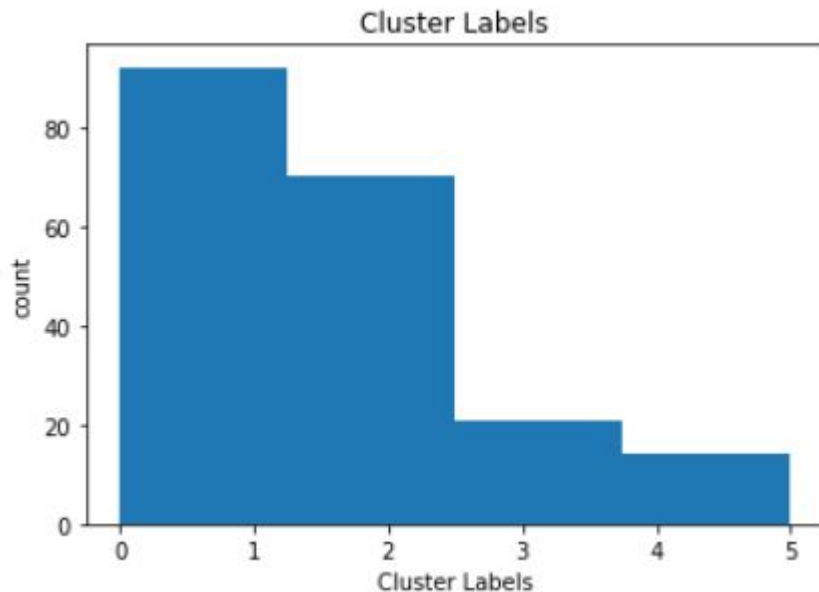
# Binning average housing price

- Five bins will be used as follows:-
  1. Low Level : < 800,000
  2. Below Average : 800,000 – 1,200,000
  3. Average : 1,200,000 – 1,600,000
  4. Above Average : 1,600,000 – 2,000,000
  5. High Level : > 2,000,000



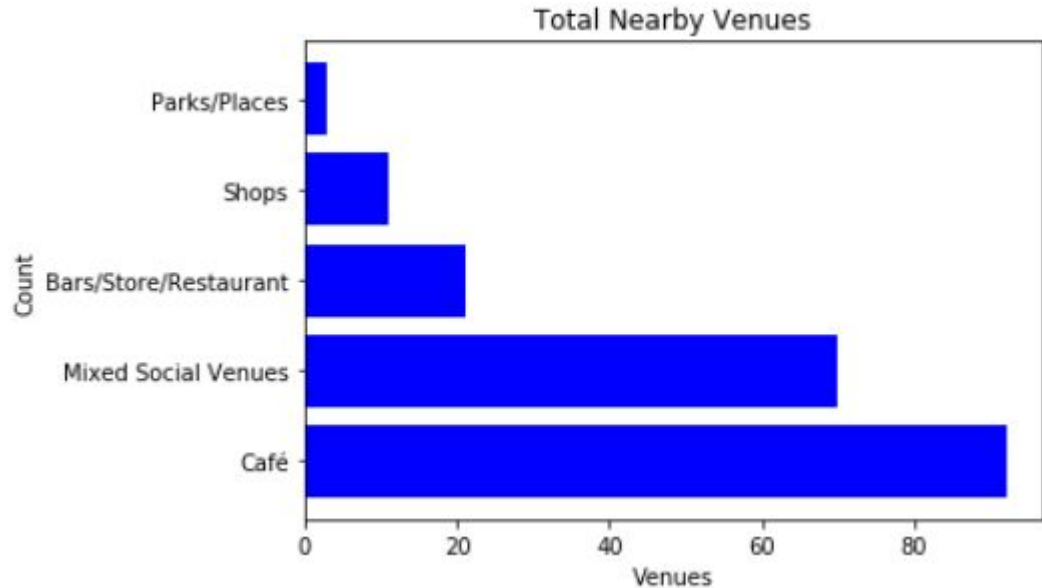
# Binning for cluster labels

- Four bins will be used for five different venues category as follows:-
  1. Cafe
  2. Mixed Social Venues
  3. Bars, Store and Restaurant
  4. Shops
  5. Parks and places



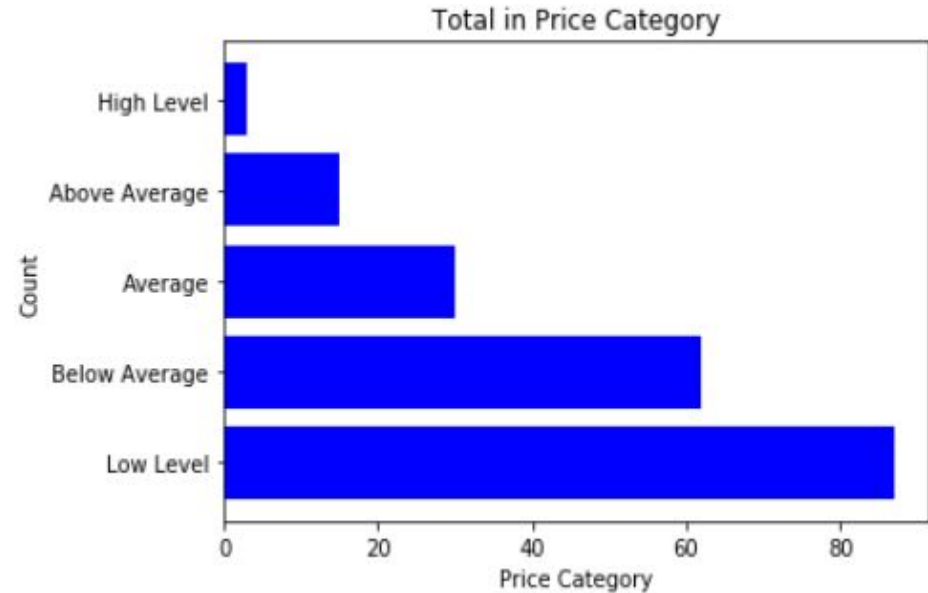
# Result & Findings

- most frequent venues are:-
  - Cafe
  - Mixed Social Venues
  - Bars, Store & Restaurant
- least frequent venues are:-
  - Shops
  - Parks & other places



# Result & Findings

- frequent prices category are:-
  - Low Level
  - Below Average
  - Average
- less frequent prices category are:-
  - Above Average
  - High Level





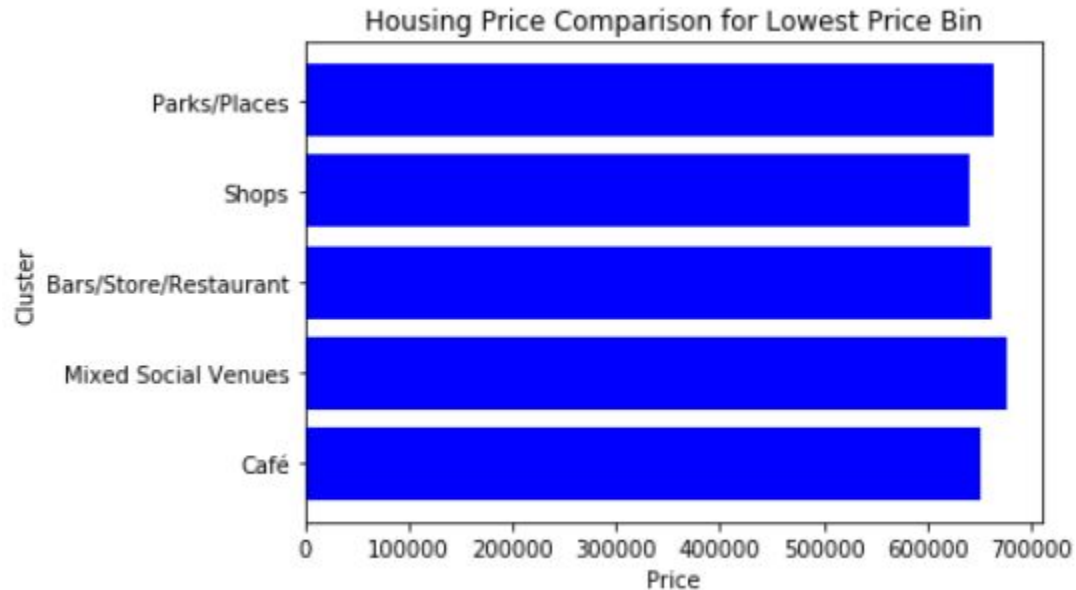
# Summary

- cheapest housing price locations are near either Cafe or Shops where by most expensive housing price locations are near Parks or Mixed Social Venues

	Café	Mixed Social Venues	Bars/Store/Restaurant	Shops	Parks/Places
Low Level	650443.98	676052.28	662906.31	640540.97	665000.00
Below Average	1017134.57	962482.67	984402.94	885100.00	1078272.75
Average	1393744.33	1390169.05	1369113.25	nan	1308933.03
Above Average	1801127.71	1765751.65	1747225.83	1858816.27	nan
High Level	2183091.25	2423333.33	nan	nan	nan

# Summary

- Lowest housing price inside “Low Level” price bin is near Shops



# Conclusion

- Analyse housing price data for Melbourne
- Fetch nearby venues data using Foursquare API to analyse price variation based on different localities
- Cluster the data set using unsupervised machine learning method, namely k-means
- Find out lowest and highest housing price by comparing each price category with nearby venues data

Niá:wen  
நீன்றி  
masi chok  
唔該  
T'ooyaksiy' nisim  
Graciās  
謝謝  
Thank You  
màhsì'  
Merci  
Obrigado  
Miigwech  
Se:kenh:  
ありがとうございました