HOUSING SALES PRICES VS VENUES DATA ANALYSIS OF SYDNEY

SHANSHAN JIN | 18-07-2020

AGENDA

- Purpose and Methodology
- Data Source and Preprocessing
- Exploring the result
- Conclusion

PURPOSE AND METHODOLOGY

- Background
  - Sydney is a good city for living but the housing price is high and growing fast. It’s a big issue to make good decision of buying a house.
  - Invest in Sydney also need support of venue density information.
- Purpose:
  - To help make better decision of buying a house or start a venue

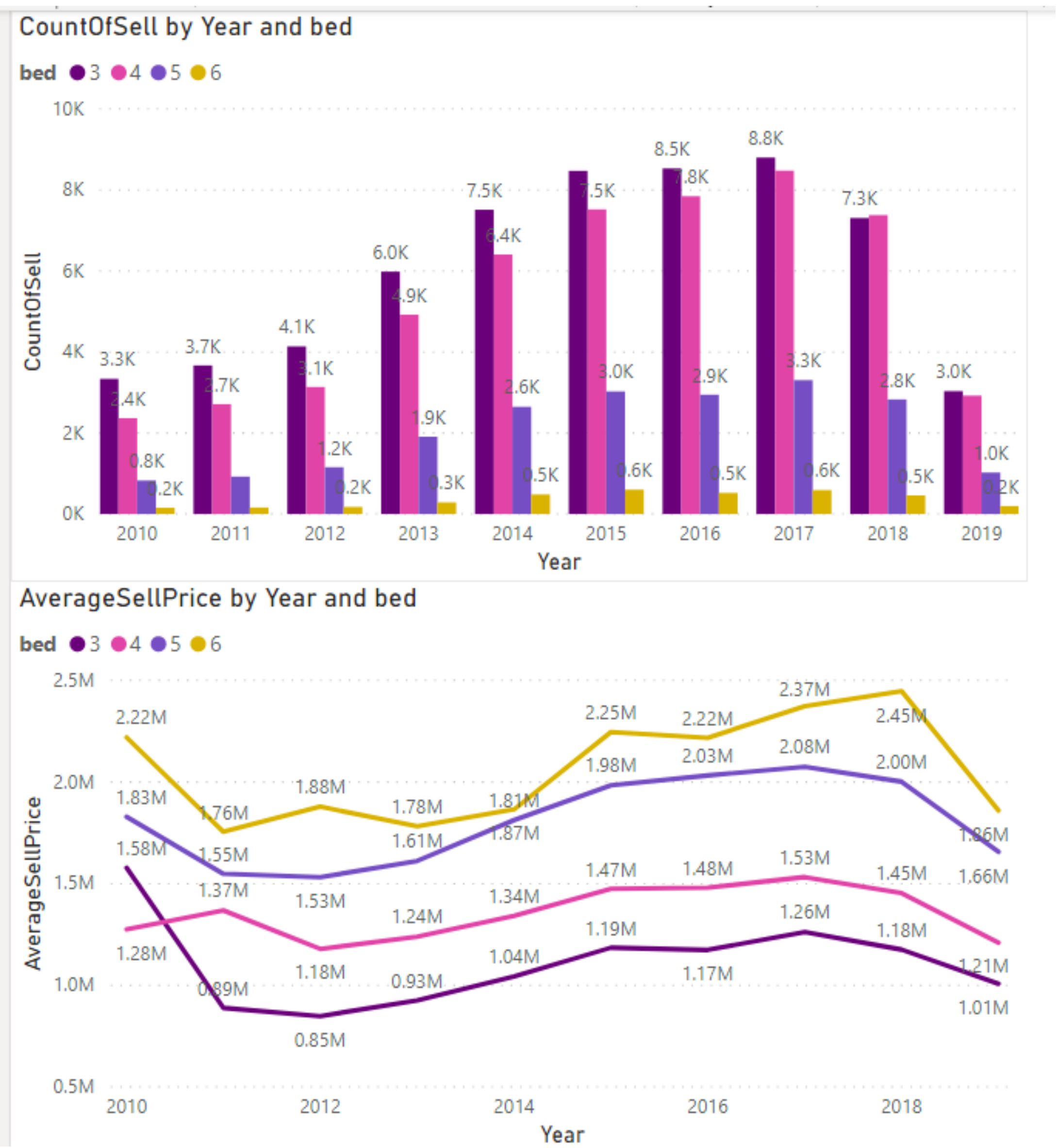
- Methodology :
  - Get data through scraping, Foursquare API and so on.
  - Using K-Means machine learning algorithm to cluster suburbs into groups
  - Using Folium to put suburbs in a map.

DATA SOURCE AND PREPROCESSING

- Data Sources:
  - Sydney property prices from 2000 to 2019 --- from Kaggle dataset.
  - Sdata of Great Sydney --- corra.com.au
  - The list of Opportunity Classes --- education.nsw.gov.au
  - Get the venues data of given Borough of Sydney --- Foursquare API

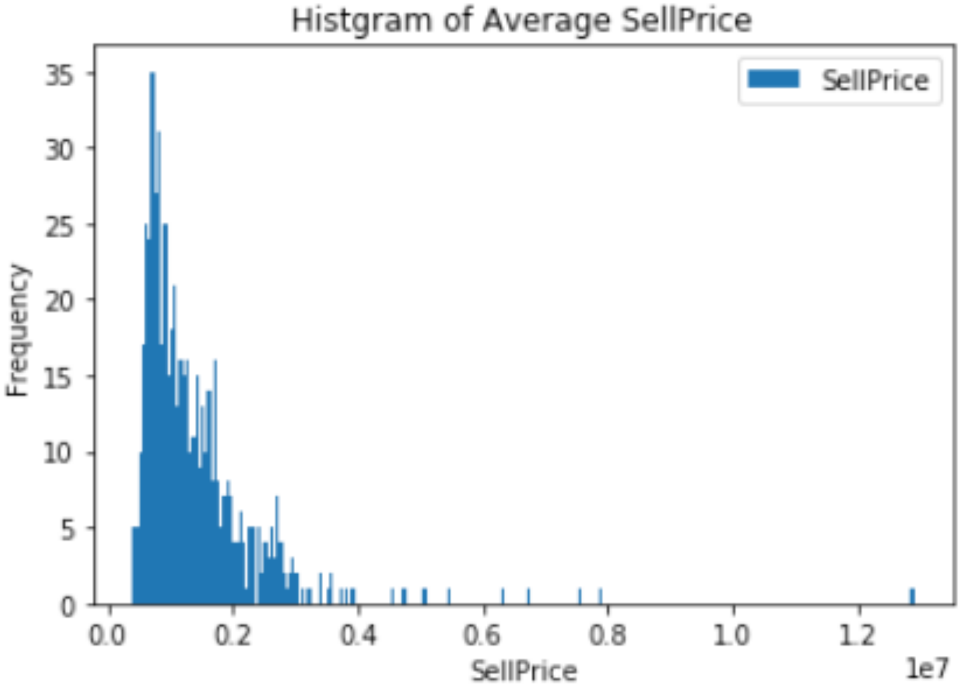
DATA SOURCE AND PREPROCESSING

- Data Selection
  - As can be seen in the plot. The number of bedrooms lead to differences in price, but the changing trends are similar.
  - So I decided not to use the bed column in this report.
  - Only keep the year, suburb and sell price column.



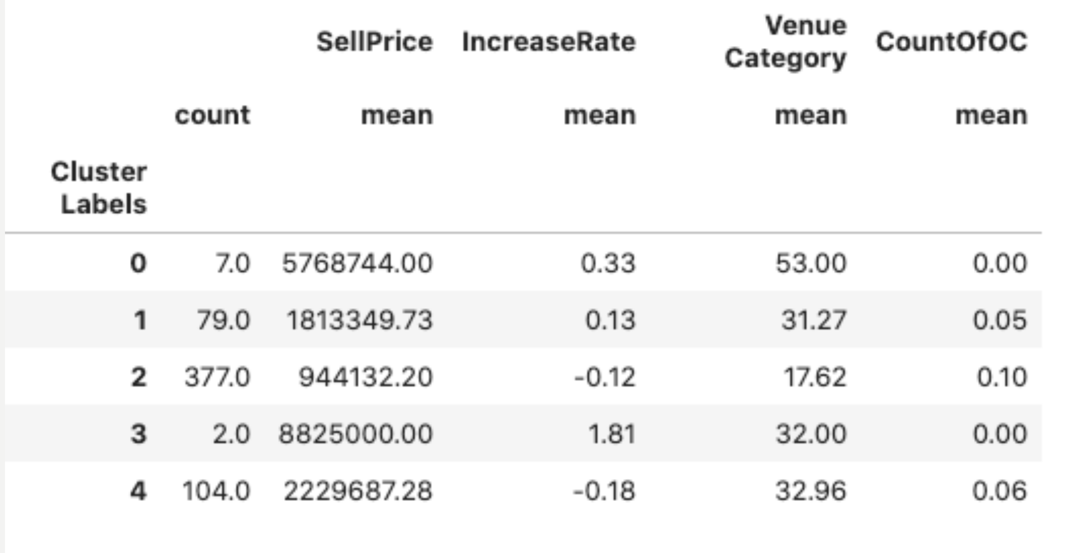
DATA SOURCE AND PREPROCESSING

- Data Selection
  - considering that the latest price contains more information indicating the current situation. I calculated average sell price and yoy change rate by year and only selected Year 2019 for analyzing use.
  - Using histogram I found that the housing price is close to normal distribution.
  - So I normalized the sell price and yoy change rate using scikit learn preprocessing scale method.

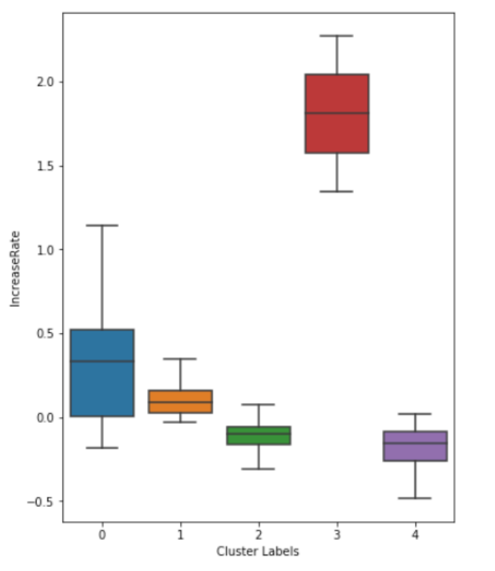


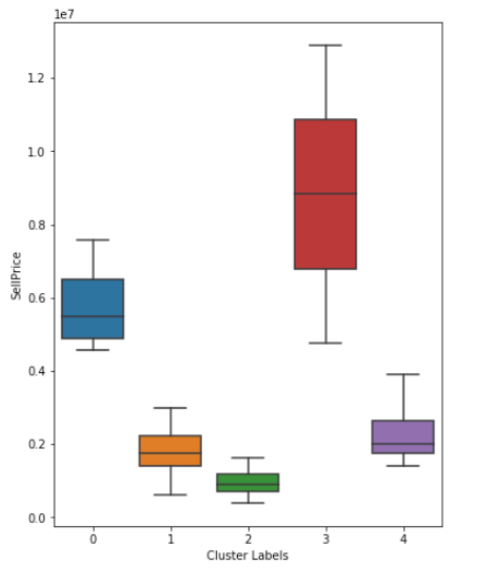
EXPLORING THE RESULT

- General result
  - considering that the latest price contains more information indicating the current situation. I calculated average sell price and yoy change rate by year and only selected Year 2019 for analyzing use.
  - Using histogram I found that the housing price is close to normal distribution.
  - So I normalized the sell price and yoy change rate using scikit learn preprocessing scale method.



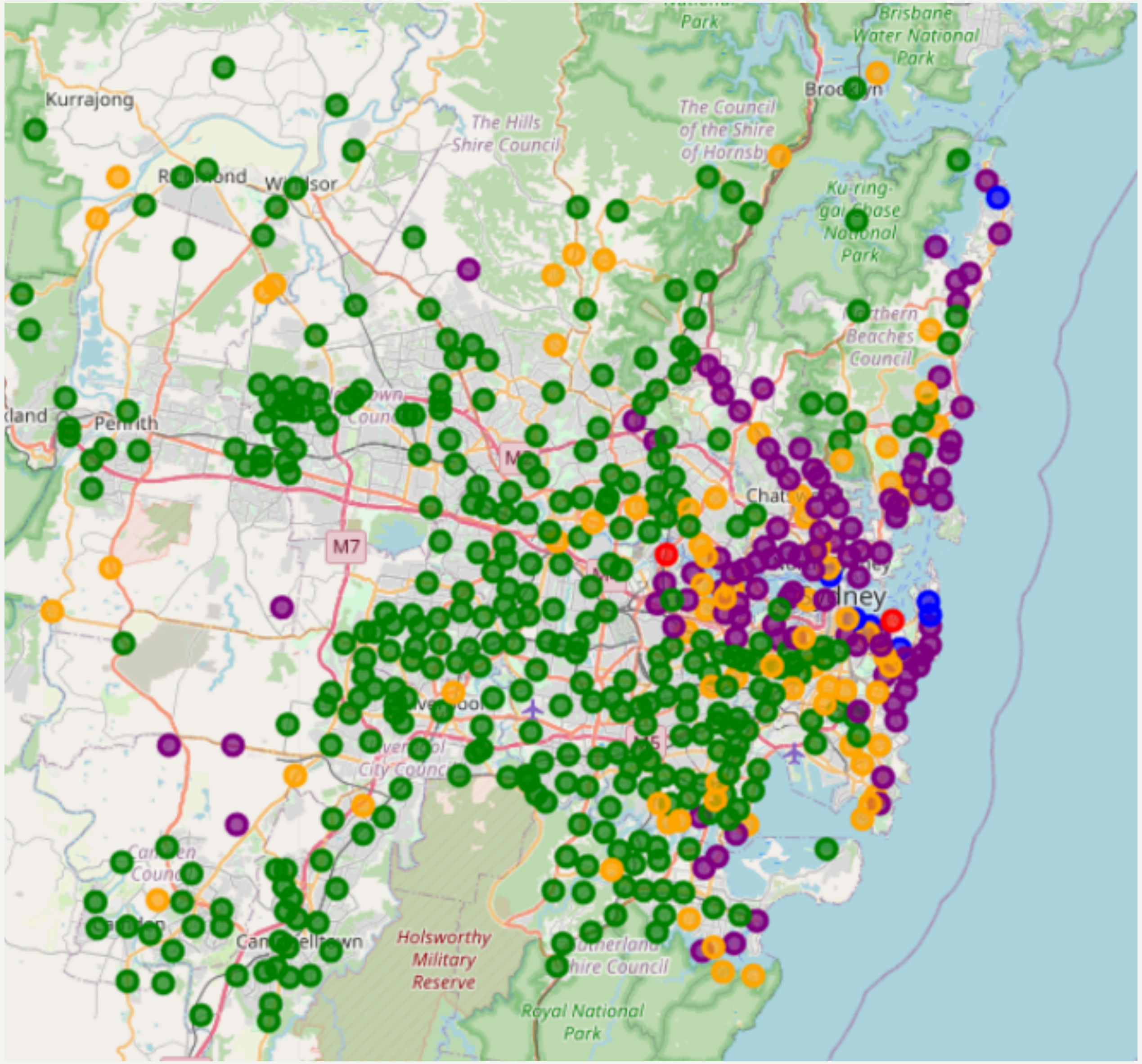
		SellPrice	IncreaseRate	Venue Category	CountOfOC
	count	mean	mean	mean	mean
Cluster Labels					
0	7.0	5768744.00	0.33	53.00	0.00
1	79.0	1813349.73	0.13	31.27	0.05
2	377.0	944132.20	-0.12	17.62	0.10
3	2.0	8825000.00	1.81	32.00	0.00
4	104.0	2229687.28	-0.18	32.96	0.06





EXPLORING THE RESULT

- Blue and Red --- cluster 0 and 3 .
  - The average sell price is the highest and so do the increase rate (over 5m).
  - All located near beaches with ocean or river view.
  - Very good venue density (53 within 1000m)
- Purple--- cluster 4.
  - Located near CBD of Sydney
  - the housing price is relatively high (2.23m)
  - The increase rate is the lowest (-18%)
  - Good venue density(33 within 1000m)
- Orange --- cluster 1.
  - Medium sell price (1.81m)
  - Positive increase rate (13%).
  - A good venue density (31 within 1000m)
  - Mostly locate not far away from CBD
- Green --- cluster 2.
  - The average sell price is the lowest (0.94m).
  - Low increase rate (-12%)
  - Least venues (18 within 1000m)
  - Far away from CBD.



CONCLUTION

- Personally, I feel this report is helpful in making decision of buying a house.
  - For me, suburbs in cluster is 1 is the most suitable.
  - It is convenient with good venue density and not far away from CBD.
  - It is not very expensive compared with other clusters except cluster 2.
  - It has a good increase rate for investment.
- People decided to buy a house or want to start a venue in Sydney can have a look at planforms providing such information as sell price, increase rate, venue density and categories. People can achieve better outcomes through their access to the platforms where such information is provided.

THANK YOU！