CAPSTONE PROJECT: SINGAPORE PRIVATE PROPERTY ANALYSIS
(PRICE, LOCATIONS, NEARBY VENUES, PRIMARY SCHOOLS)

Yunjia Zeng

Background

- Singapore, the city state and island country in Southeast Asia
- One of the original Four Asian Tigers
- High GDP per capita ➔ highest stand of living in Asia-Pacific
- Forward-thinking policies: world-class infrastructures
- Famous for airports, housing, safety, and advanced infocomm networks
- Singapore as one of the most attractive cities to live in Asia

Business Problem

- Singapore's private properties highly sought-after by Singaporeans and foreign investors
- High prices of its private properties, not a trivial decision to make
- Too many factors to be considered in the decision making process
- Data analyses particularly useful in decision-making processes
- This project aims to help in decision-making process of buying Singapore properties

Target Audience

- Target audience: buyers of Singapore's private properties and property agents
- Potential home buyers: make more informed decisions
- Property agents: provide better recommendations to their customers

Objectives

- Better understanding of the available properties in market
- Cross-comparison of the properties
- Analyze near-by venues that the future property owners could enjoy
- Special bonus section for home buyers with preschool kids
- Primary schools nearby properties sorted out and analyzed

Data

- Based on the definition of the business problem, essential data for the analysis include:
  - Private properties for sale in the market
  - Recorded sale prices of the properties
  - Locations of the properties
  - Total units for each properties
  - Venues nearby the properties
  - Primary schools that are located within 1km of the properties

Data

◦ Due to the data size and availability constraints, the analyses will focus on the properties directly sold by all developers in a six-month time frame (from 2019 October to 2020 March).

◦ Following data sources will be utilized to generate the required data:

* The private property sale records from Singapore URA website[1]

* The location data of each property extracted from **geopy.geocoders**

* A number of venues with their categories within a specified distance (1km) of every single property obtained via **Foursquare API**

* All the primary schools within a distance of 1km of every single property obtained via **Foursquare API**

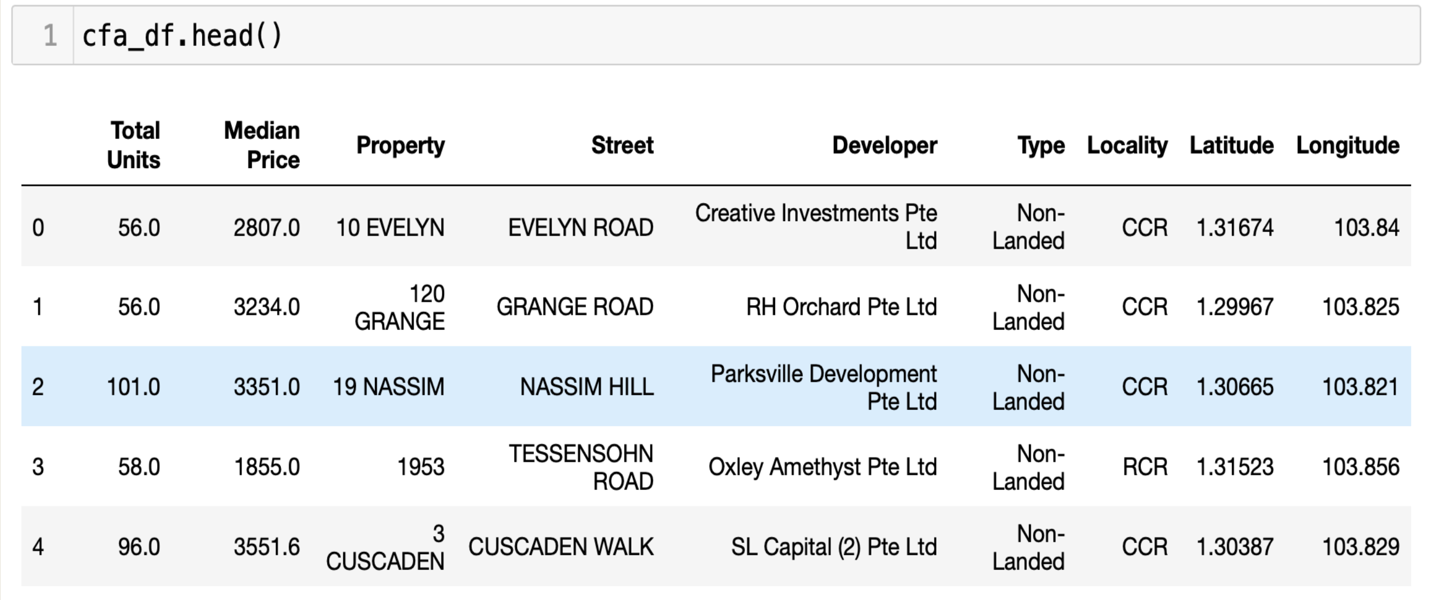
[1] https://www.ura.gov.sg/realEstateIIWeb/price/search.action

Private Property Sale Data

- The Singapore private property sale data from 2019 October till 202 March was downloaded from URA website.
- For the purpose of this analysis, only the columns with property name, street address, developer information, type and locality of the properties, total units of the property, median price of the property are kept.
- The median price is calculated as the average over the six-month period. The dataframe to be used is ‘cfa_df’. The sale data are based on the properties sold directly from developers.

Private Property Location Data

- The location data with latitude and longitude for every property is retrieved from geocoder, and also included in the 'cfa_df' dataframe.

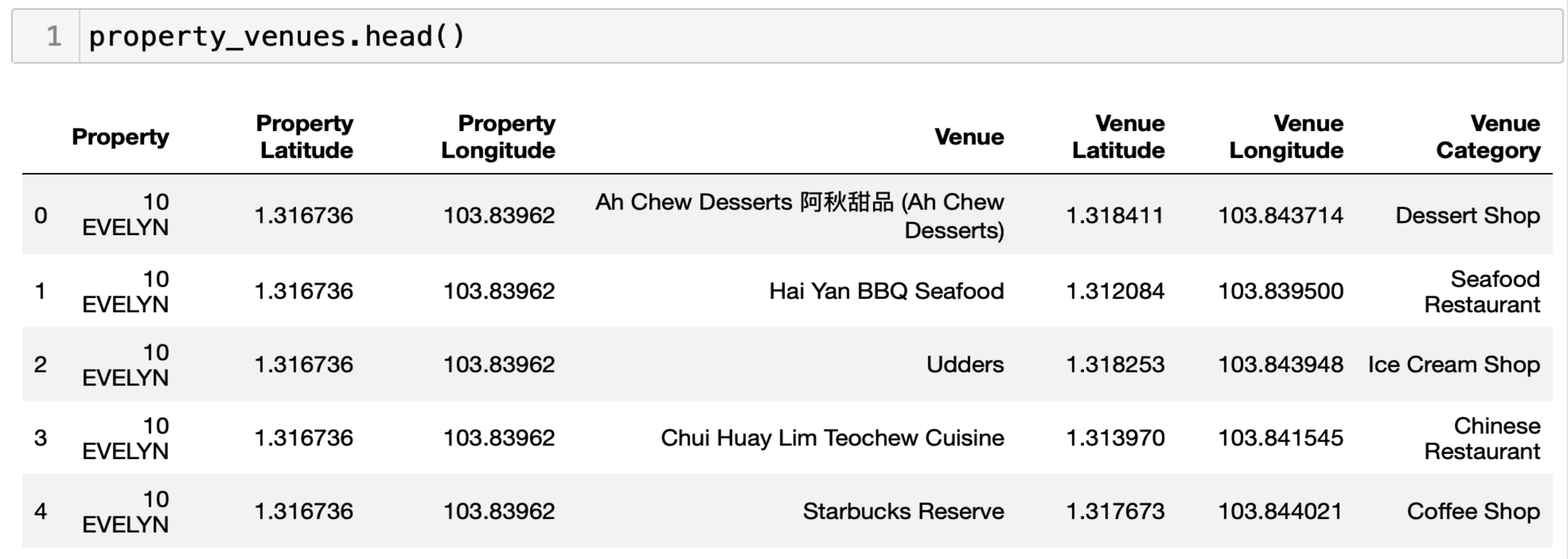


```
cfa_df.head()
```

	Total Units	Median Price	Property	Street	Developer	Type	Locality	Latitude	Longitude
0	56.0	2807.0	10 EVELYN	EVELYN ROAD	Creative Investments Pte Ltd	Non-Landed	CCR	1.31674	103.84
1	56.0	3234.0	120 GRANGE	GRANGE ROAD	RH Orchard Pte Ltd	Non-Landed	CCR	1.29967	103.825
2	101.0	3351.0	19 NASSIM	NASSIM HILL	Parksville Development Pte Ltd	Non-Landed	CCR	1.30665	103.821
3	58.0	1855.0	1953	TESSENSOHN ROAD	Oxley Amethyst Pte Ltd	Non-Landed	RCR	1.31523	103.856
4	96.0	3551.6	3 CUSCADEN	CUSCADEN WALK	SL Capital (2) Pte Ltd	Non-Landed	CCR	1.30387	103.829

Private Property Nearby Venue Data

- The top 100 venues within a distance of 1 kilometer from the property are retrieved from Foursquare.
- The ‘property_venue’ dataframe includes the venue name, venue location, and venue category besides the property information for each property.

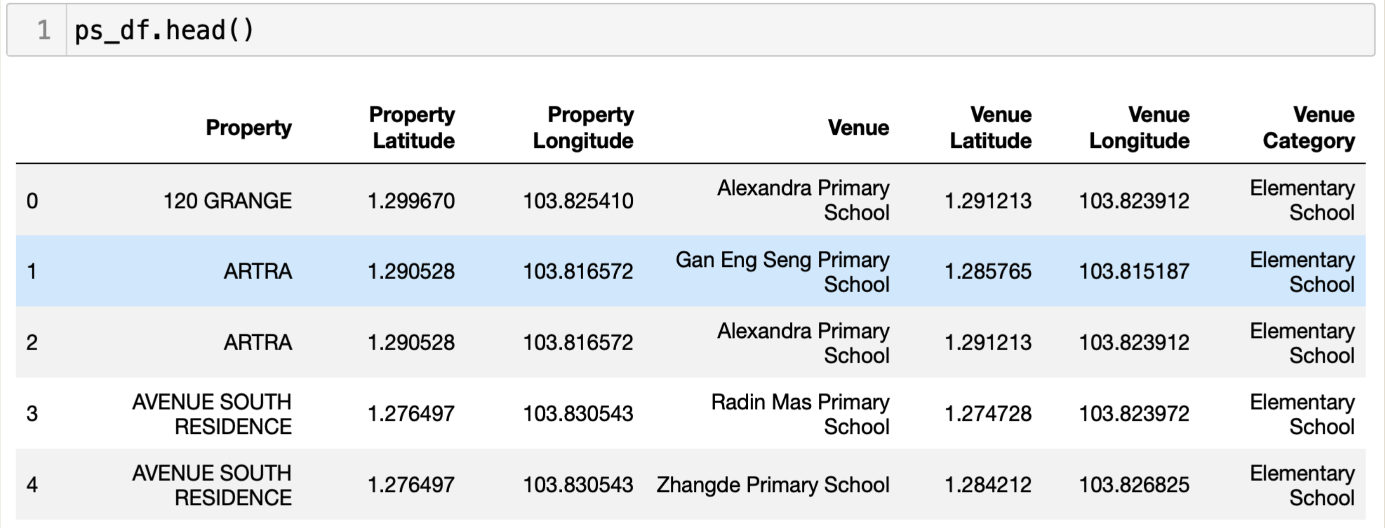


```
property_venues.head()
```

	Property	Property Latitude	Property Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	10 EVELYN	1.316736	103.83962	Ah Chew Desserts 阿秋甜品 (Ah Chew Desserts)	1.318411	103.843714	Dessert Shop
1	10 EVELYN	1.316736	103.83962	Hai Yan BBQ Seafood	1.312084	103.839500	Seafood Restaurant
2	10 EVELYN	1.316736	103.83962	Udders	1.318253	103.843948	Ice Cream Shop
3	10 EVELYN	1.316736	103.83962	Chui Huay Lim Teochew Cuisine	1.313970	103.841545	Chinese Restaurant
4	10 EVELYN	1.316736	103.83962	Starbucks Reserve	1.317673	103.844021	Coffee Shop

Private Property Nearby School Data

- To retrieve the nearby school information, the ‘property_venue’ dataframe was first explored.
- But it turns out very few schools are included in the top 100 venues for the properties.
- Another query was made via Foursquare to retrieve the top 100 schools within a distance of 1 kilometer from the properties.
- Among the retrieved schools, primary schools are picked and stored in the dataframe ‘ps_df’.

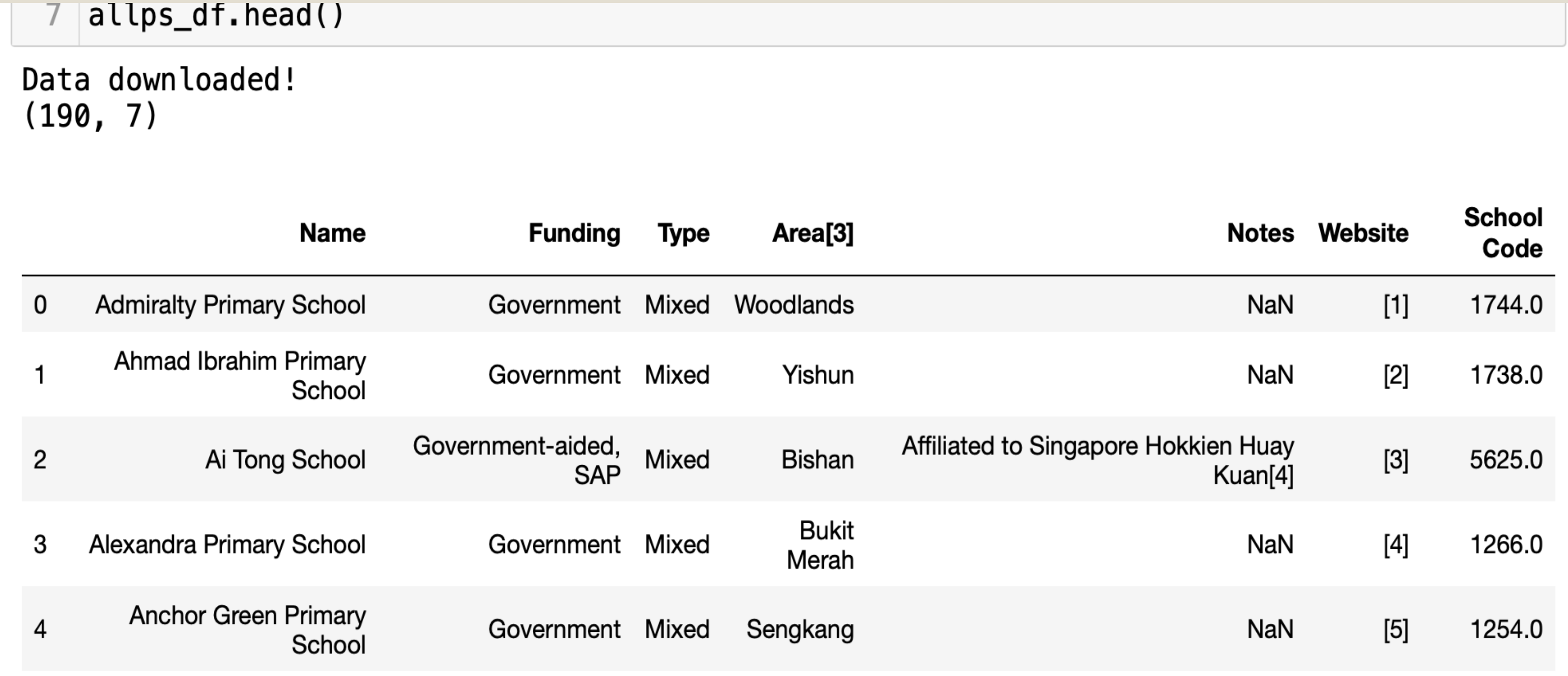


```
ps_df.head()
```

	Property	Property Latitude	Property Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	120 GRANGE	1.299670	103.825410	Alexandra Primary School	1.291213	103.823912	Elementary School
1	ARTRA	1.290528	103.816572	Gan Eng Seng Primary School	1.285765	103.815187	Elementary School
2	ARTRA	1.290528	103.816572	Alexandra Primary School	1.291213	103.823912	Elementary School
3	AVENUE SOUTH RESIDENCE	1.276497	103.830543	Radin Mas Primary School	1.274728	103.823972	Elementary School
4	AVENUE SOUTH RESIDENCE	1.276497	103.830543	Zhangde Primary School	1.284212	103.826825	Elementary School

All School Information

- The list of all the Singapore primary school information were retrieved from the website[2] and stored in ‘allps_df’ dataframe



```
allps_df.head()
```

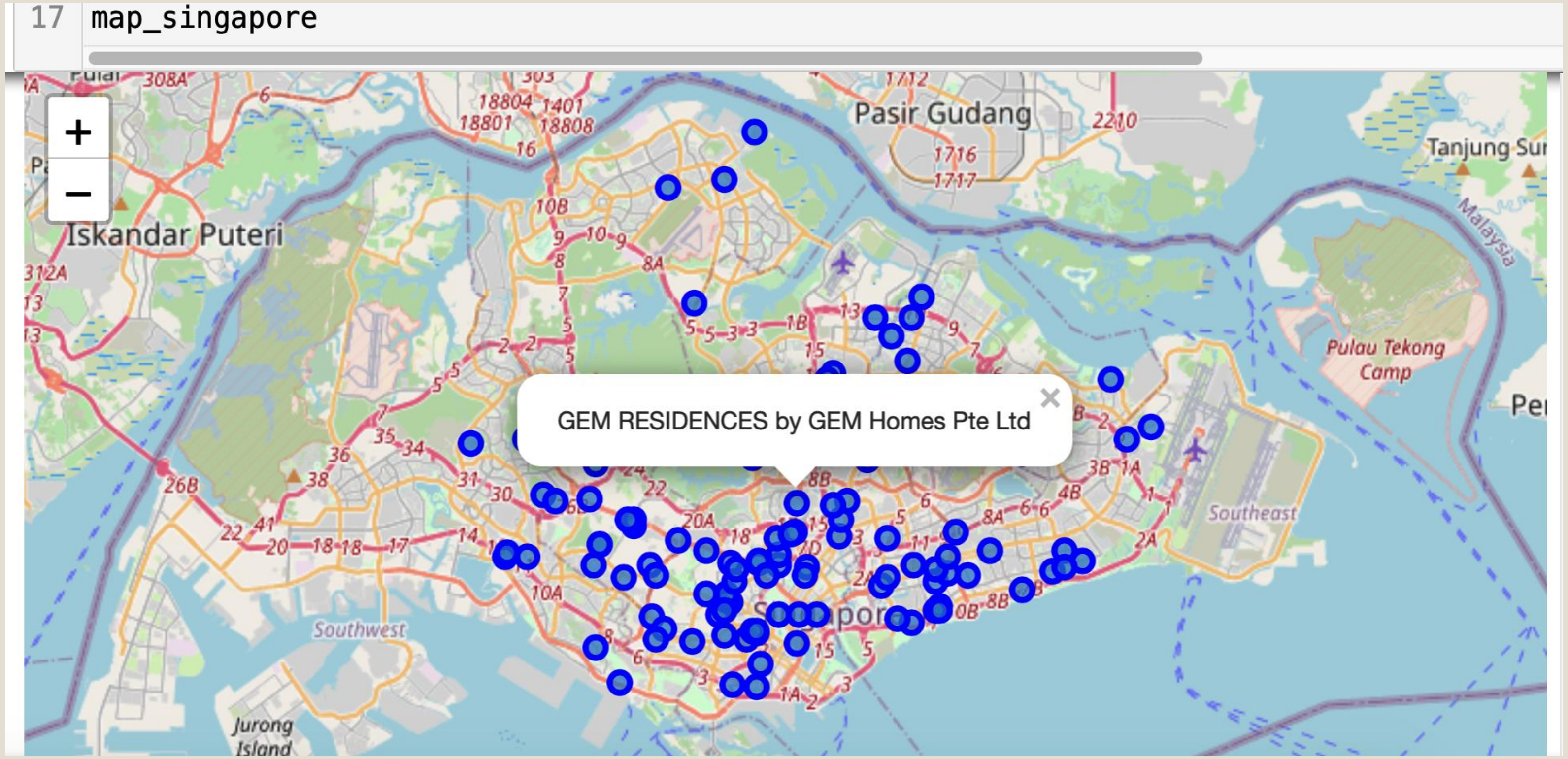
```
Data downloaded!
(190, 7)
```

	Name	Funding	Type	Area[3]	Notes	Website	School Code
0	Admiralty Primary School	Government	Mixed	Woodlands	NaN	[1]	1744.0
1	Ahmad Ibrahim Primary School	Government	Mixed	Yishun	NaN	[2]	1738.0
2	Ai Tong School	Government-aided, SAP	Mixed	Bishan	Affiliated to Singapore Hokkien Huay Kuan[4]	[3]	5625.0
3	Alexandra Primary School	Government	Mixed	Bukit Merah	NaN	[4]	1266.0
4	Anchor Green Primary School	Government	Mixed	Sengkang	NaN	[5]	1254.0

[2] https://en.wikipedia.org/wiki/List_of_primary_schools_in_Singapore

Initial visualization on Map

- All the properties considered are plotted as markers on the Singapore map
- Each pop-up shows on property and its developer



```
map_singapore
```

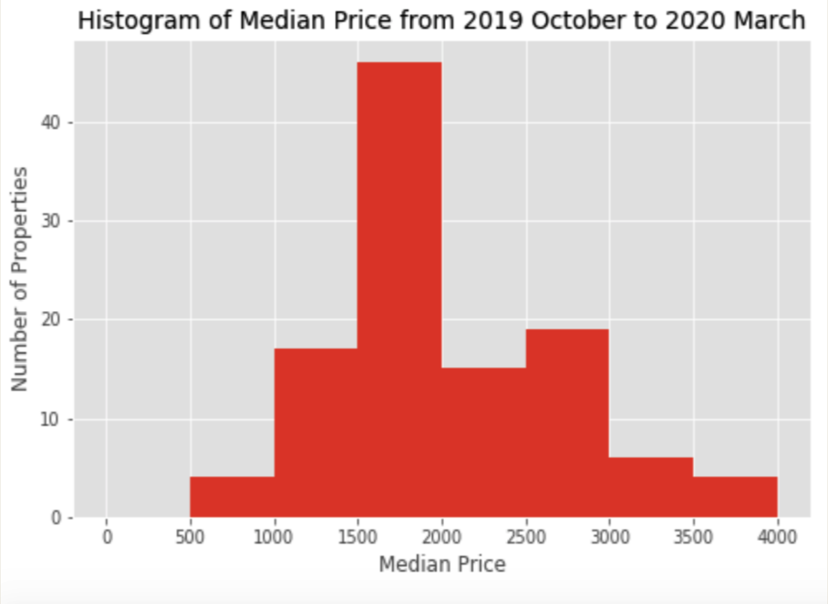
GEM RESIDENCES by GEM Homes Pte Ltd

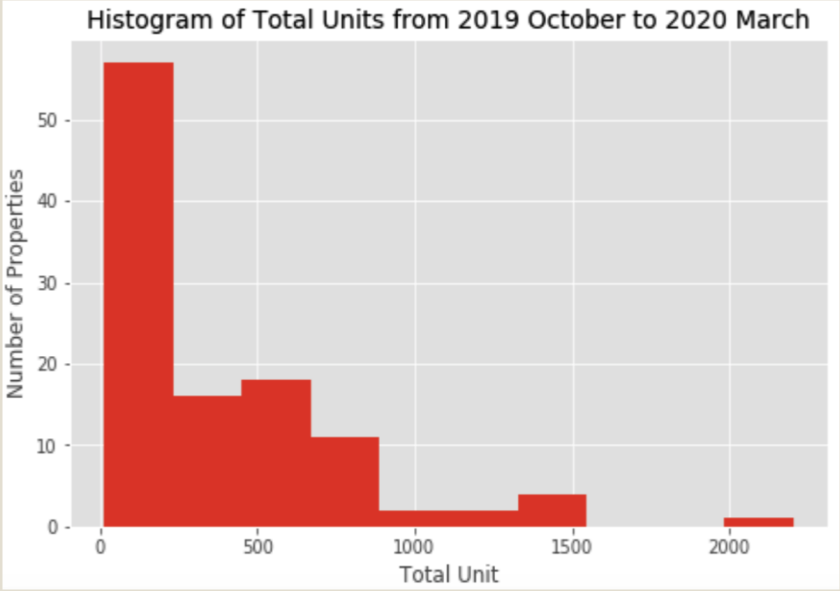
Methodology

- Exploratory analysis will first be performed on the data
- Different types of diagrams are used to visualize the property data
- The dataframes are also cross checked to find any abnormal entry to correct
- K-means is used to cluster all the properties considered by the types and density of their nearby venues
- The generated different clustered are examined in details to find their individual characteristics
- The school information is then apply to sort out properties that are suitable for different customers

Exploratory Data Analysis

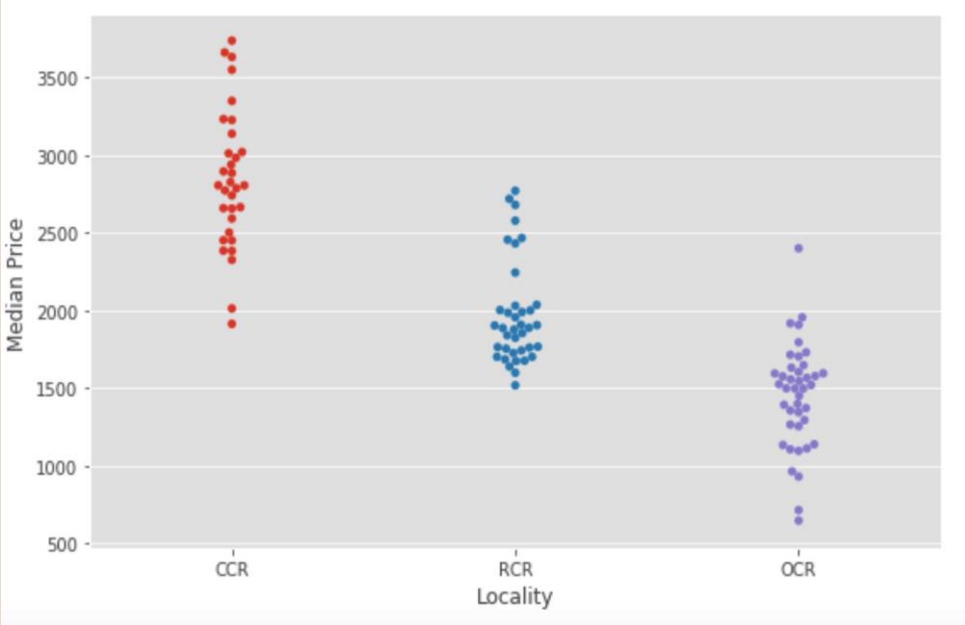
- It is found that the median price mostly ranges from 1500 to 2000 SGD per square feet, while majority of the total units in properties are less than 500.

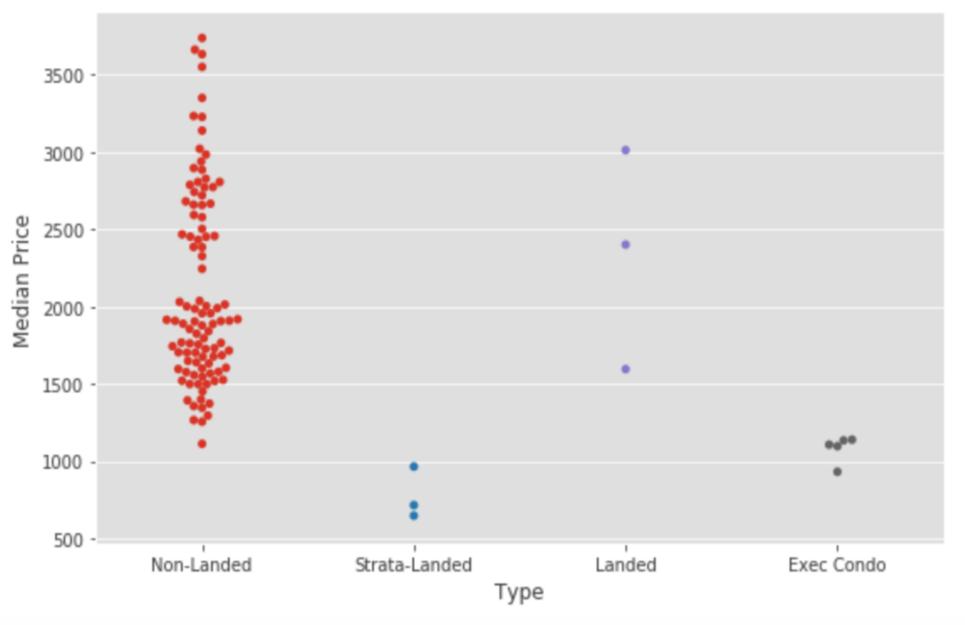




Exploratory Data Analysis

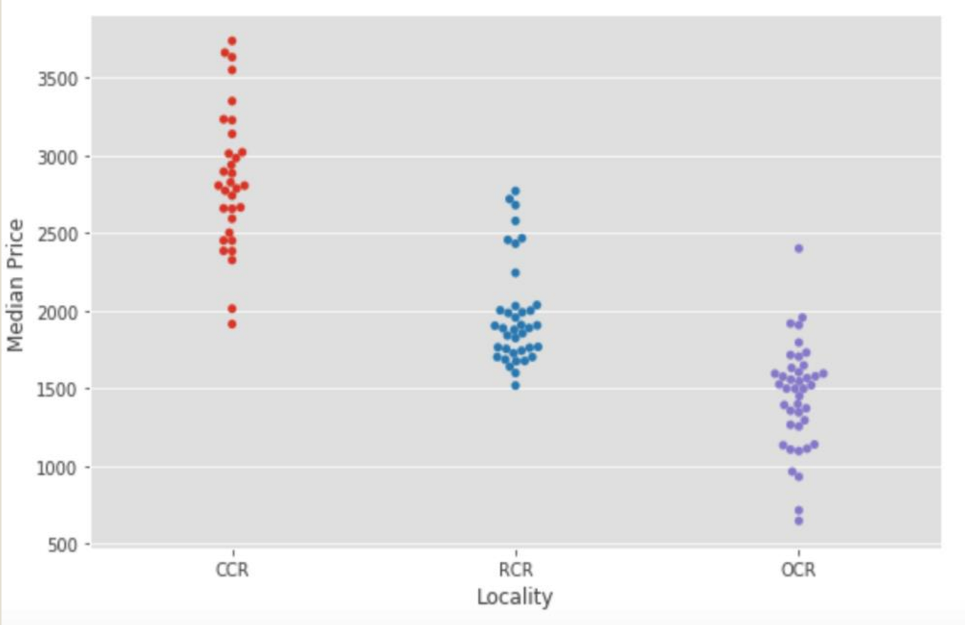
- Swarm plots for the category data: locality and type of the properties
- To visualize their distribution together with median price
- Such plots adjust the points along the categorical axis and prevent them from overlapping.

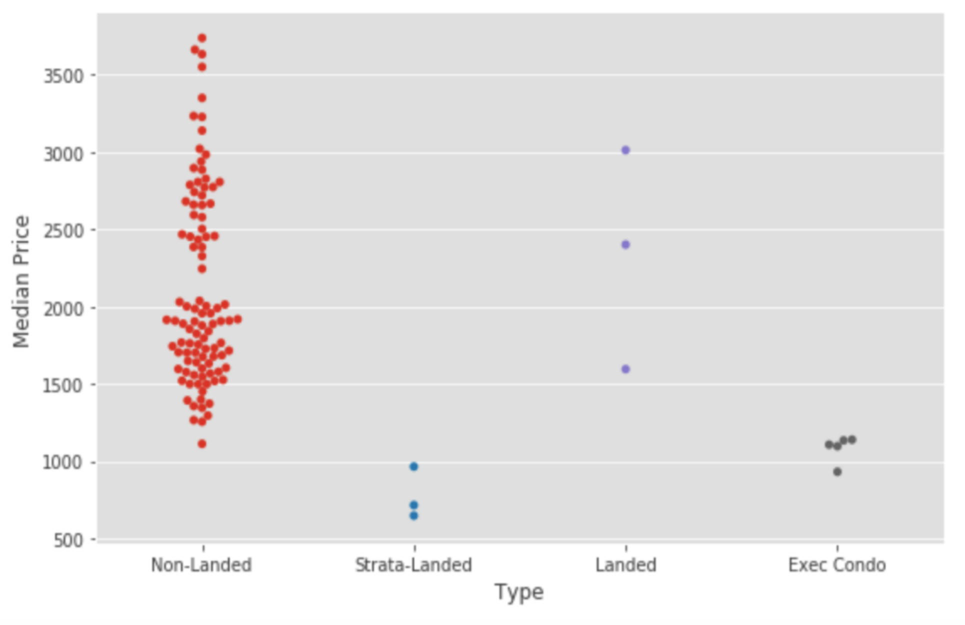




Exploratory Data Analysis

- Median price varies significantly across different locality type, with CCR type has the highest property price, followed by RCR, and the OCR being the category with the lowest price
- Most of the properties are of the type ‘Non-landed’, which also has the widest spread of price range



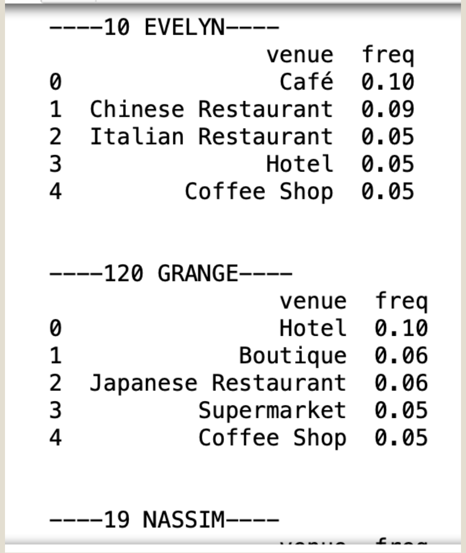


Geocoder Issue

- When exploring the data of venues nearby properties, it is first noted that there are in total 300 unique categories of venues.
- The venue data are then encoded via one-hot of the 300 unique categories, and the data are further grouped by every single property considered in the analysis.
- It was then figured out that the number of properties in the venue data is one less than the original property data.
- The missing property is found out, and it turns out that the geographical data of this missing property is wrongly obtained from geocoder with a negative values in its latitude.
- The correct geo data is generated, and the dataframes are updated accordingly.

Venue Data Preprocessing

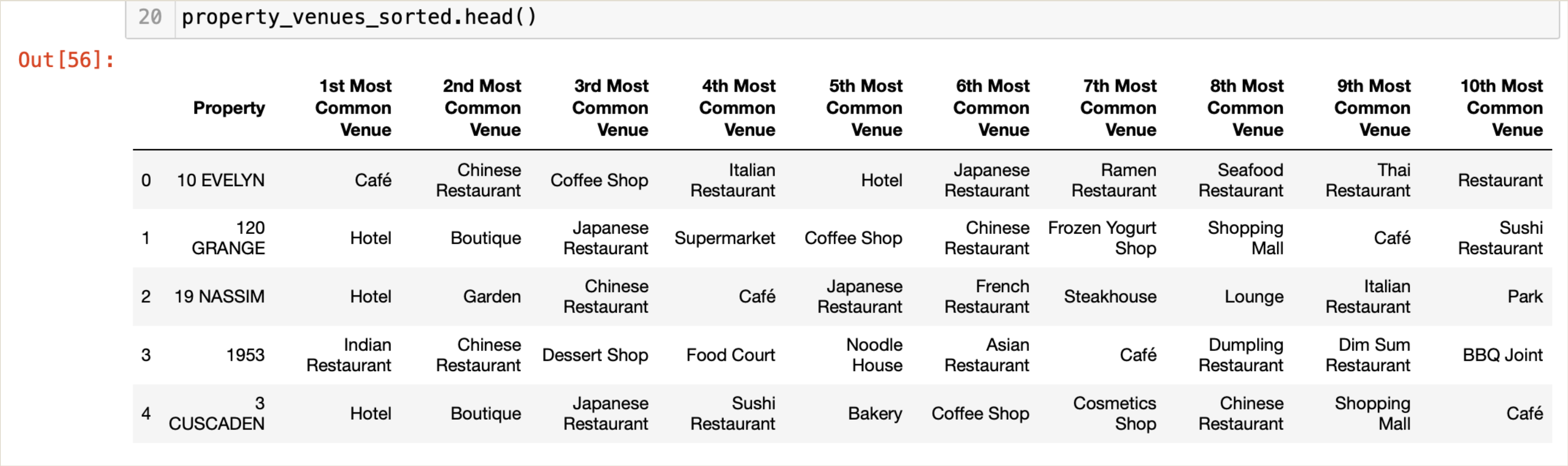
- The densities of each venue categories nearby the every single properties are then calculated
- And the top five venues of the properties are printed out in notebook
- In the next step, the top ten common venues of every property are summarized into the dataframe 'property_venues_sorted'.



```
----10 EVELYN----
                 venue  freq
0                 Café  0.10
1   Chinese Restaurant  0.09
2   Italian Restaurant  0.05
3                Hotel  0.05
4          Coffee Shop  0.05


----120 GRANGE----
                  venue  freq
0                 Hotel  0.10
1              Boutique  0.06
2   Japanese Restaurant  0.06
3           Supermarket  0.05
4           Coffee Shop  0.05


----19 NASSIM----
```



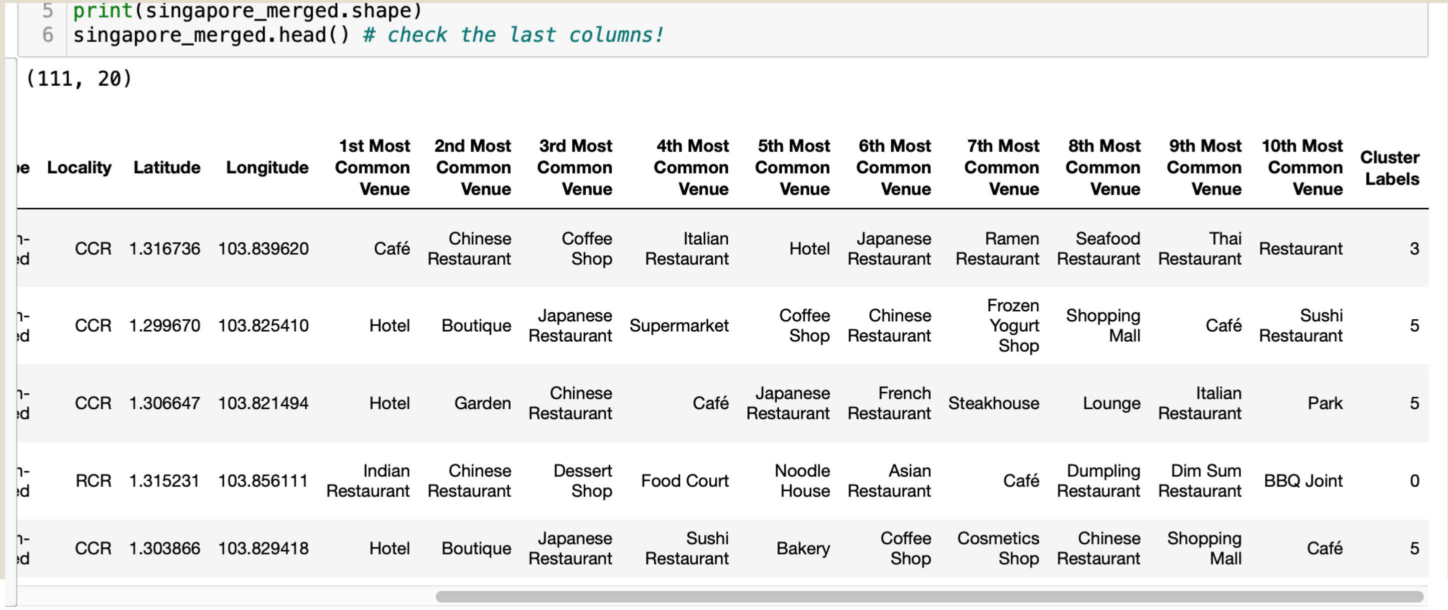
```
property_venues_sorted.head()
```

Out[56]:

	Property	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	10 EVELYN	Café	Chinese Restaurant	Coffee Shop	Italian Restaurant	Hotel	Japanese Restaurant	Ramen Restaurant	Seafood Restaurant	Thai Restaurant	Restaurant
1	120 GRANGE	Hotel	Boutique	Japanese Restaurant	Supermarket	Coffee Shop	Chinese Restaurant	Frozen Yogurt Shop	Shopping Mall	Café	Sushi Restaurant
2	19 NASSIM	Hotel	Garden	Chinese Restaurant	Café	Japanese Restaurant	French Restaurant	Steakhouse	Lounge	Italian Restaurant	Park
3	1953	Indian Restaurant	Chinese Restaurant	Dessert Shop	Food Court	Noodle House	Asian Restaurant	Café	Dumpling Restaurant	Dim Sum Restaurant	BBQ Joint
4	3 CUSCADEN	Hotel	Boutique	Japanese Restaurant	Sushi Restaurant	Bakery	Coffee Shop	Cosmetics Shop	Chinese Restaurant	Shopping Mall	Café

Clustering with Venues

- Use preprocessed venue data to cluster all the properties
- Grid search to figure out the best cluster value for the k-mean approach
- All the properties can be optimally categorized into seven clusters
- ‘singapore_merged’ dataframe contains all the information, including property information, top ten venue information, as well as the newly assigned cluster labels.



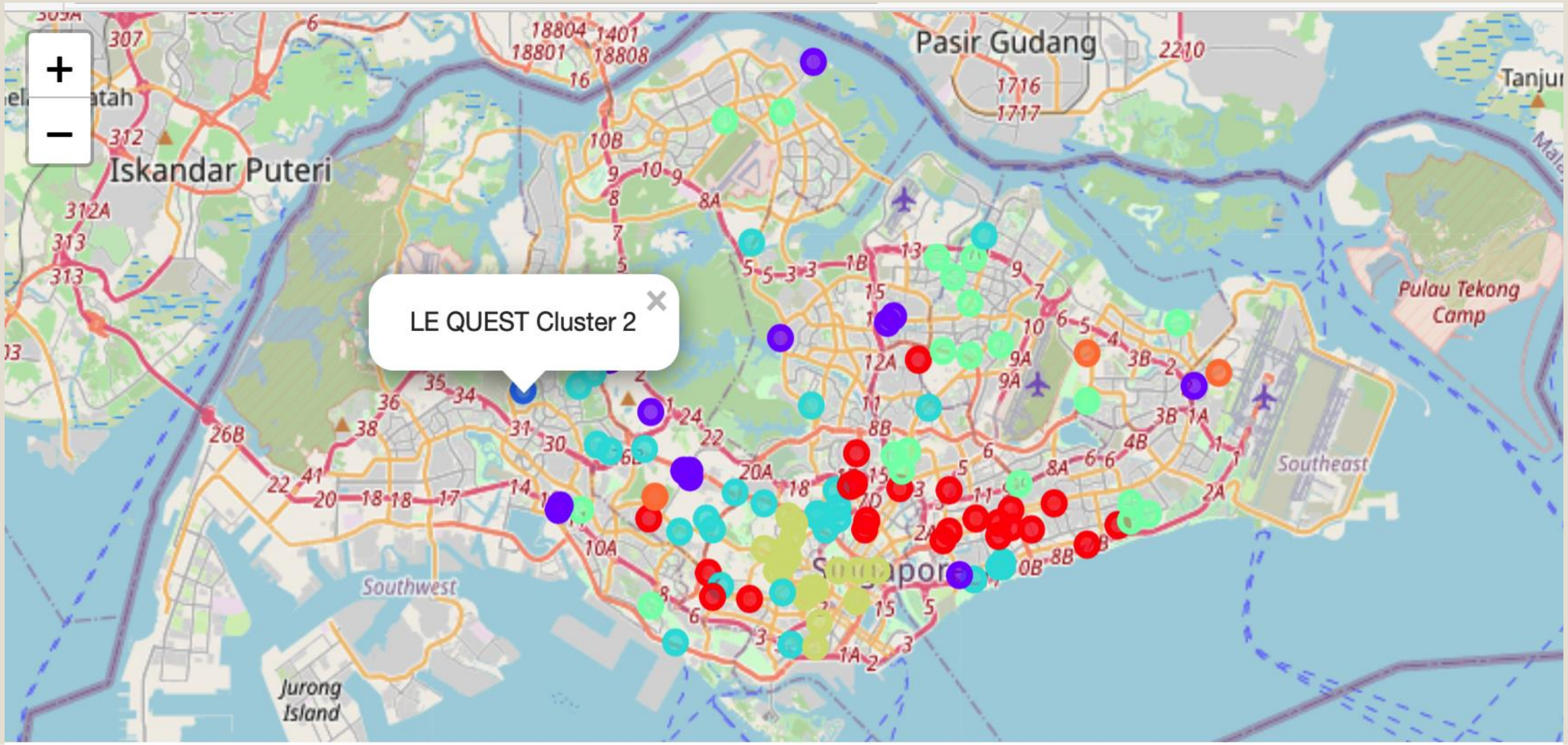
```python
print(singapore_merged.shape)
singapore_merged.head() # check the last columns!
```

```
(111, 20)
```

e	Locality	Latitude	Longitude	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	Cluster Labels
n- d	CCR	1.316736	103.839620	Café	Chinese Restaurant	Coffee Shop	Italian Restaurant	Hotel	Japanese Restaurant	Ramen Restaurant	Seafood Restaurant	Thai Restaurant	Restaurant	3
n- d	CCR	1.299670	103.825410	Hotel	Boutique	Japanese Restaurant	Supermarket	Coffee Shop	Chinese Restaurant	Frozen Yogurt Shop	Shopping Mall	Café	Sushi Restaurant	5
n- d	CCR	1.306647	103.821494	Hotel	Garden	Chinese Restaurant	Café	Japanese Restaurant	French Restaurant	Steakhouse	Lounge	Italian Restaurant	Park	5
n- d	RCR	1.315231	103.856111	Indian Restaurant	Chinese Restaurant	Dessert Shop	Food Court	Noodle House	Asian Restaurant	Café	Dumpling Restaurant	Dim Sum Restaurant	BBQ Joint	0
n- d	CCR	1.303866	103.829418	Hotel	Boutique	Japanese Restaurant	Sushi Restaurant	Bakery	Coffee Shop	Cosmetics Shop	Chinese Restaurant	Shopping Mall	Café	5

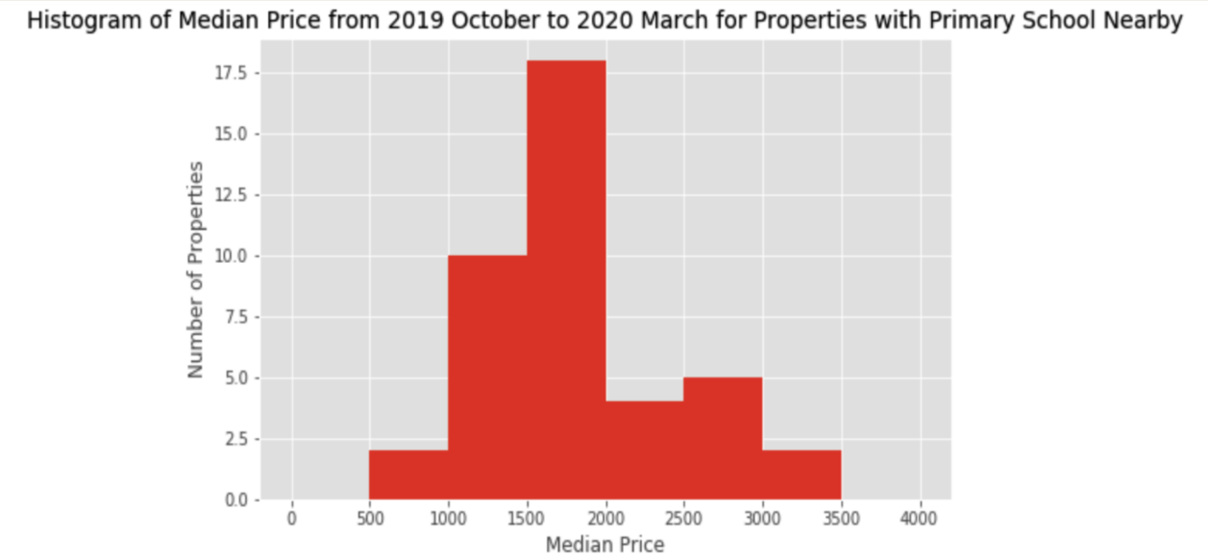
Clustering with Venues

- Visualize the clustered properties in Singapore map, with different colors of the markers representing different clusters



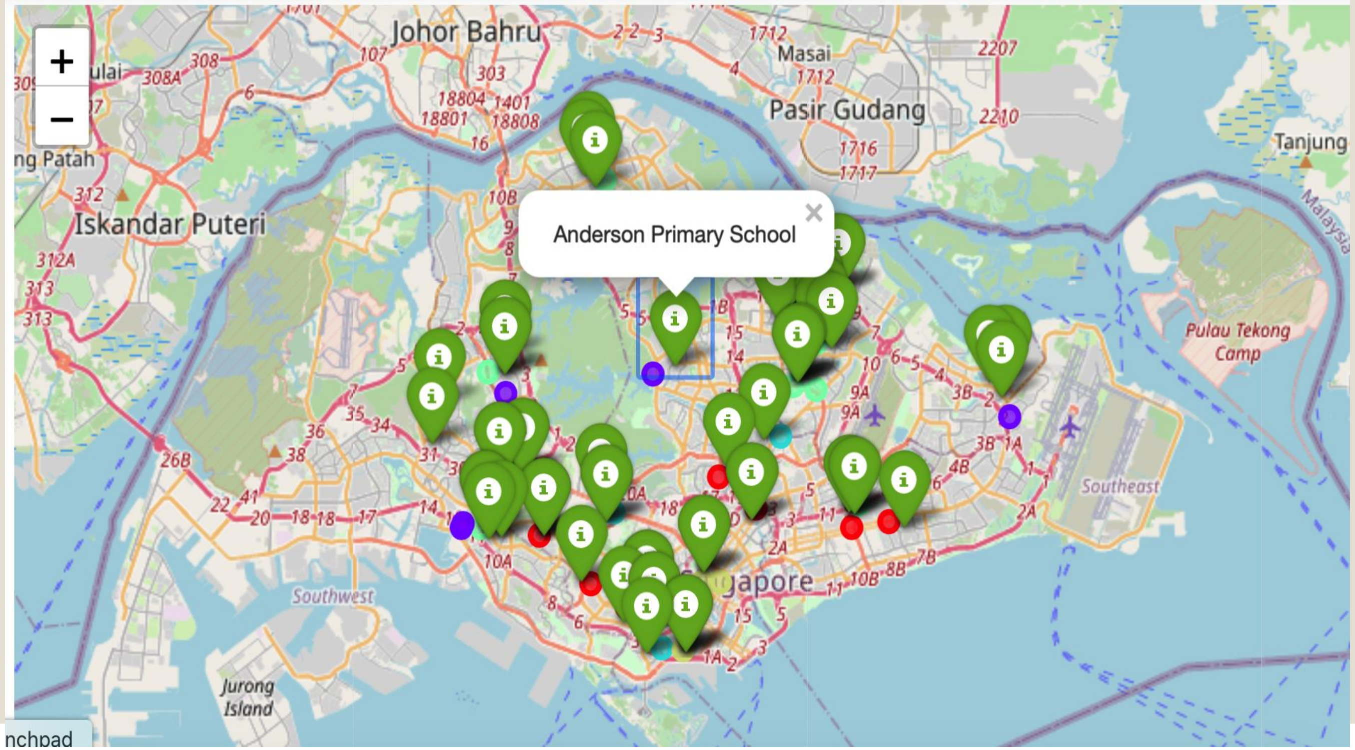
Primary School Sortout

- The primary school data are grouped by the nearby property to find out the 41 properties that have primary school within one kilometer distance
- The price distribution of these properties shows a similar distribution as that of the median price distribution of all the properties
- .Most of the properties has a median price between 1000 to 2000 SGD per square feet.
- The plot indicates that the primary school might not be very influential in determining the price of the properties.



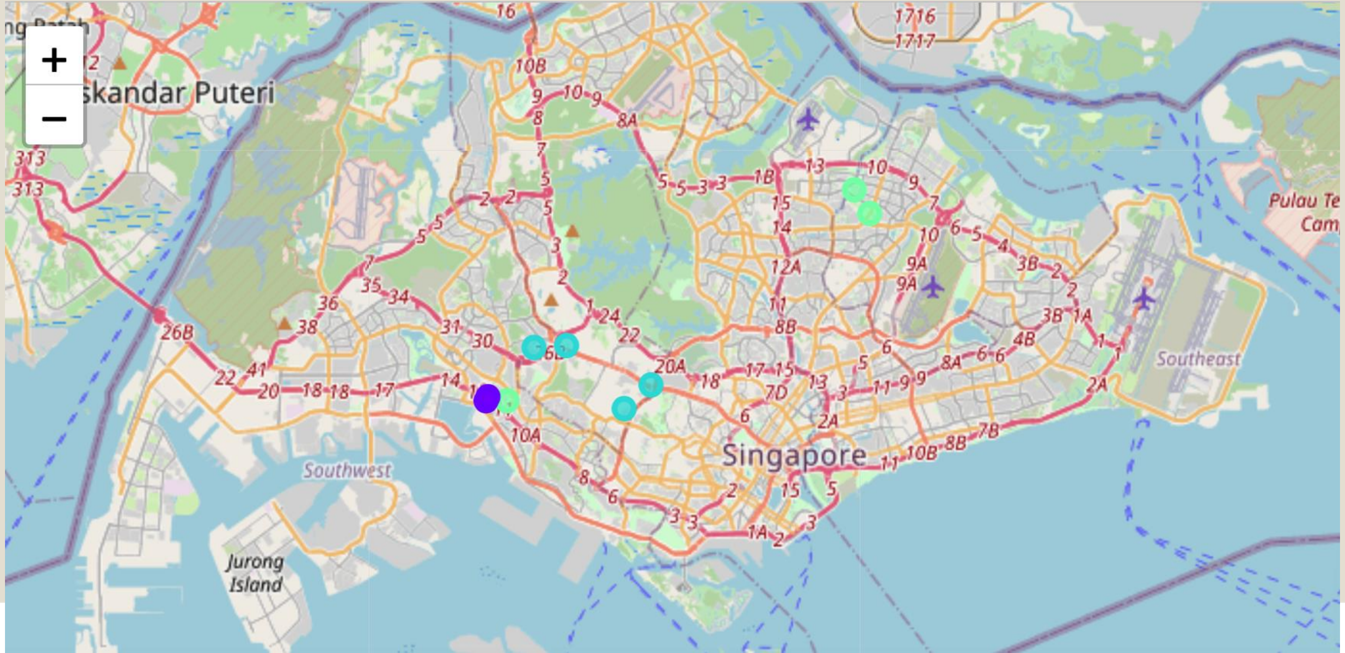
Primary School Sortout

- All these properties together with the nearby schools are plotted on Singapore map.
- The properties are represented as circle markers with colors representing the cluster labels, and the schools are in green icon markers



Primary School Sortout

- Primary schools that are affiliated with Hokkien Huay Kuan and all the special assistance plan (SAP) schools are sorted out in the dataframe ‘sps_df’
- Those properties that are within one kilometer distance from these schools are also sorted out in the dataframe ‘picked_merged’
- Visualize these nine properties on Singapore, they are belong to three clusters “2, 4, 5”



Results and Discussion on Property Sale Data Analysis

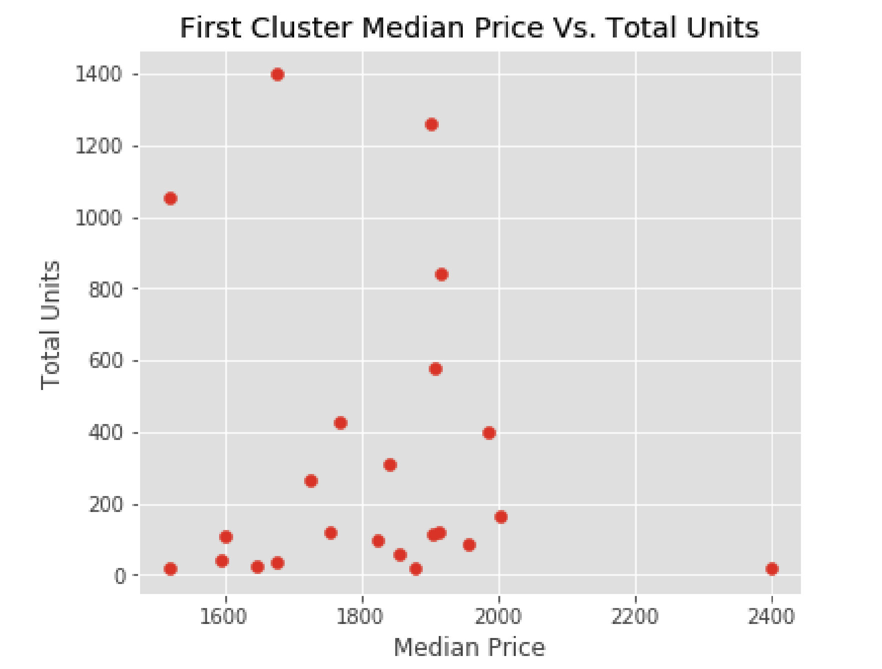
- Median price mostly ranges from 1500 to 2000 SGD per square feet
- Majority of the total units in properties are less than 500
- The median price varies significantly across different locality type
- CCR type has the highest property price, followed by RCR, and the OCR being the category with the lowest price
- As expected, location plays a significant role in pricing properties
- Three regions: Core Central Region (CCR), Rest of Central Region (RCR) and the Outside Central Region (OCR)
- CCR and RCR are the central regions of Singapore, and therefore the property prices for these two regions are higher than those in OCR
- CCR has properties with the highest price because of its premier location
- Non-landed properties: price ranges from 1000 to 4000 SGD per square feet
- Strate-landed properties: prices between 500 to 1000 SGD per square feet
- Landed properties: price above 1500 SGD per square feet
- Executive condo: price around 1000 SGD per square feet

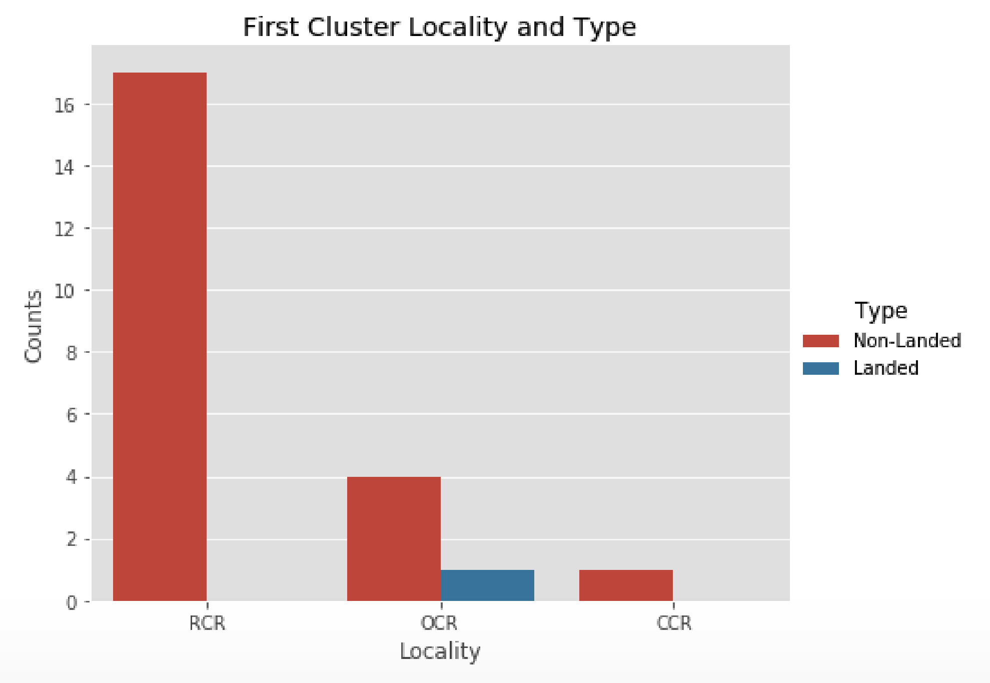
Results and Discussion on Property Clustering

- Analyze every single clusters of the properties
- The scatter plot of the property median price and total units
- The histograms of locality and property type
- The detailed listing of the venues in each clusters are printed out in the notebook.

Results and Discussion on Property Clustering: First Cluster

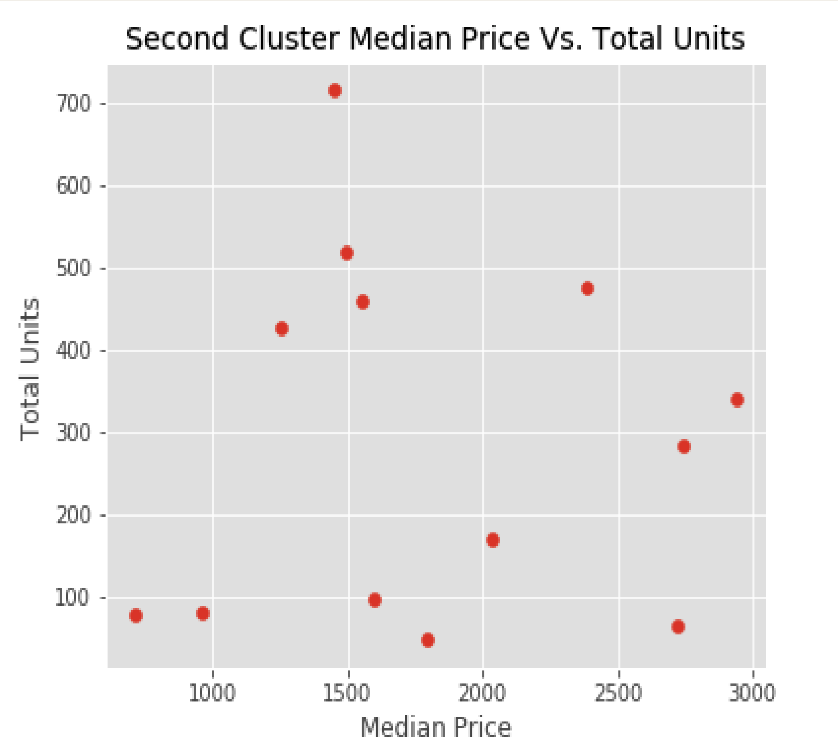
- Median price mainly from 1500 to 2000 SGD per square feet
- Mainly located in the RCR region
- Most common venues: food courts, various restaurants and cafes



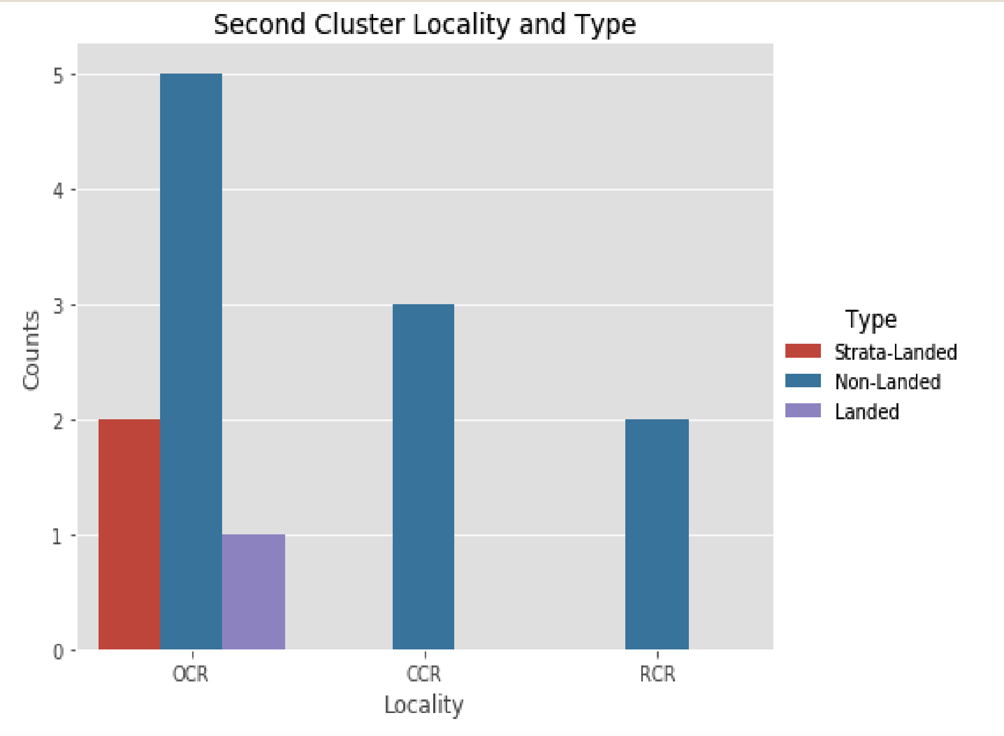


Results and Discussion on Property Clustering: Second Cluster

- A wider range of prices
- Three different property types and locations
- Most common venues: bus stations, cafes, restaurants, and recreation centers



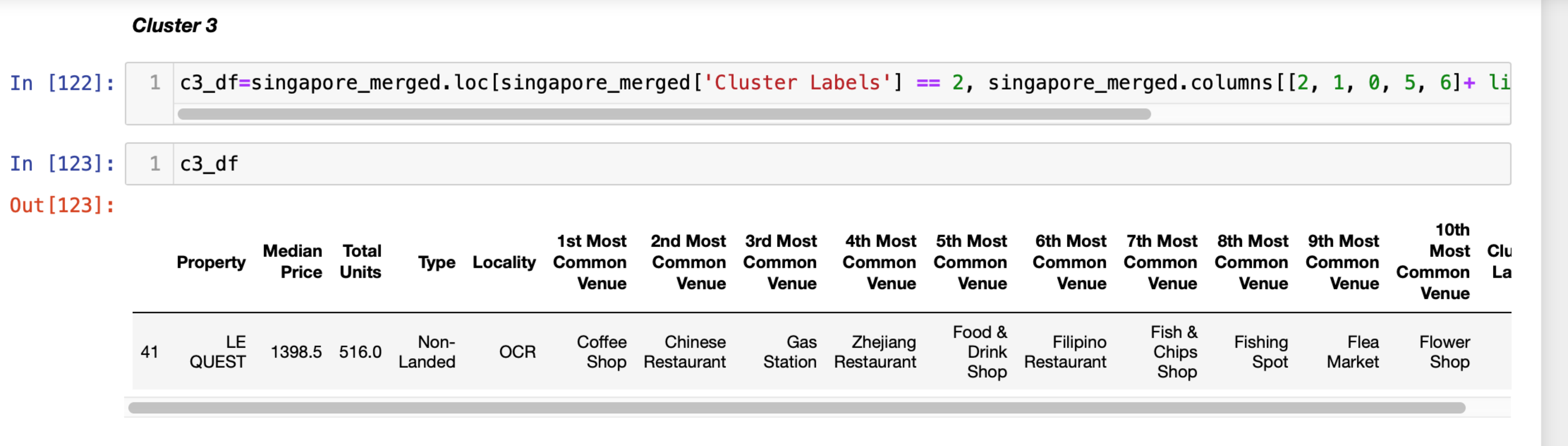
Second Cluster Median Price Vs. Total Units



Second Cluster Locality and Type

Results and Discussion on Property Clustering: Third Cluster

- Contains only one property in the west most region of Singapore
- Distinct location ➔ significantly different venue categories and densities
- Most common venues: coffee shop, restaurant, and gas station



Cluster 3

```
In [122]:  c3_df=singapore_merged.loc[singapore_merged['Cluster Labels'] == 2, singapore_merged.columns[[2, 1, 0, 5, 6]+ li
```

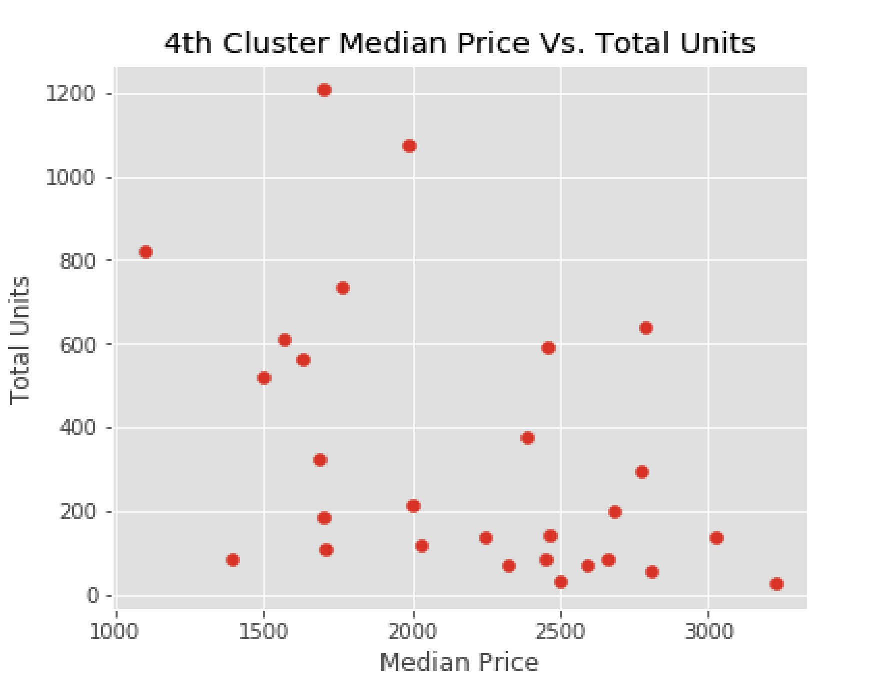
```
In [123]:  c3_df
```

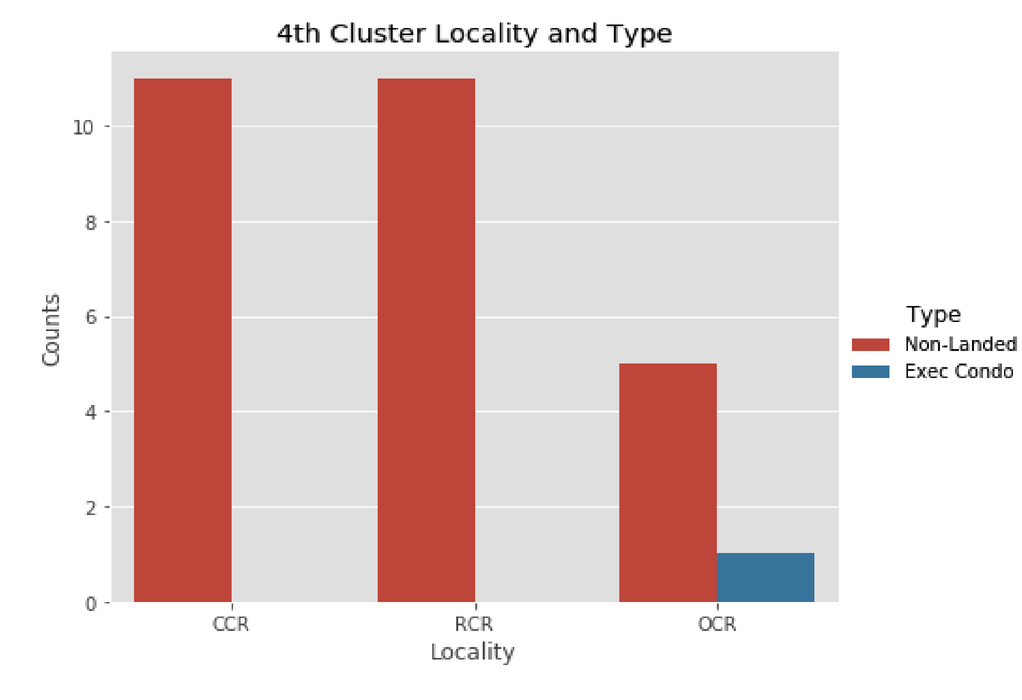
Out[123]:

	Property	Median Price	Total Units	Type	Locality	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	Clu La
41	LE QUEST	1398.5	516.0	Non-Landed	OCR	Coffee Shop	Chinese Restaurant	Gas Station	Zhejiang Restaurant	Food & Drink Shop	Filipino Restaurant	Fish & Chips Shop	Fishing Spot	Flea Market	Flower Shop	

Results and Discussion on Property Clustering: Fourth Cluster

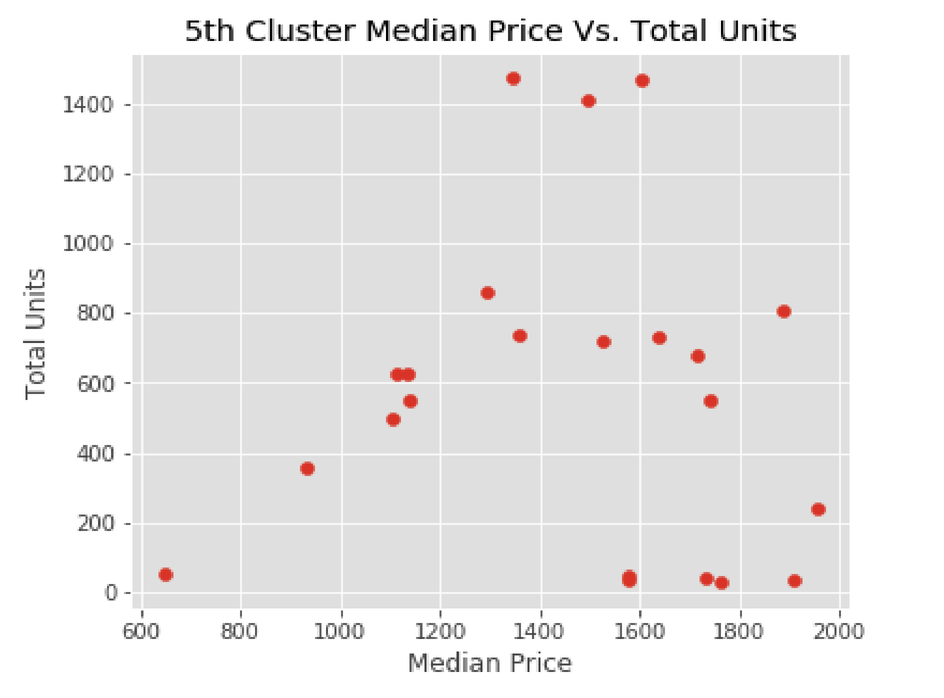
- More property prices between 2000 to 3000 SGD per square feet with less than 500 units
- Mainly located in CCR and RCR
- Most common venues: cafes, restaurants, bus stations, and hotels, and more varieties

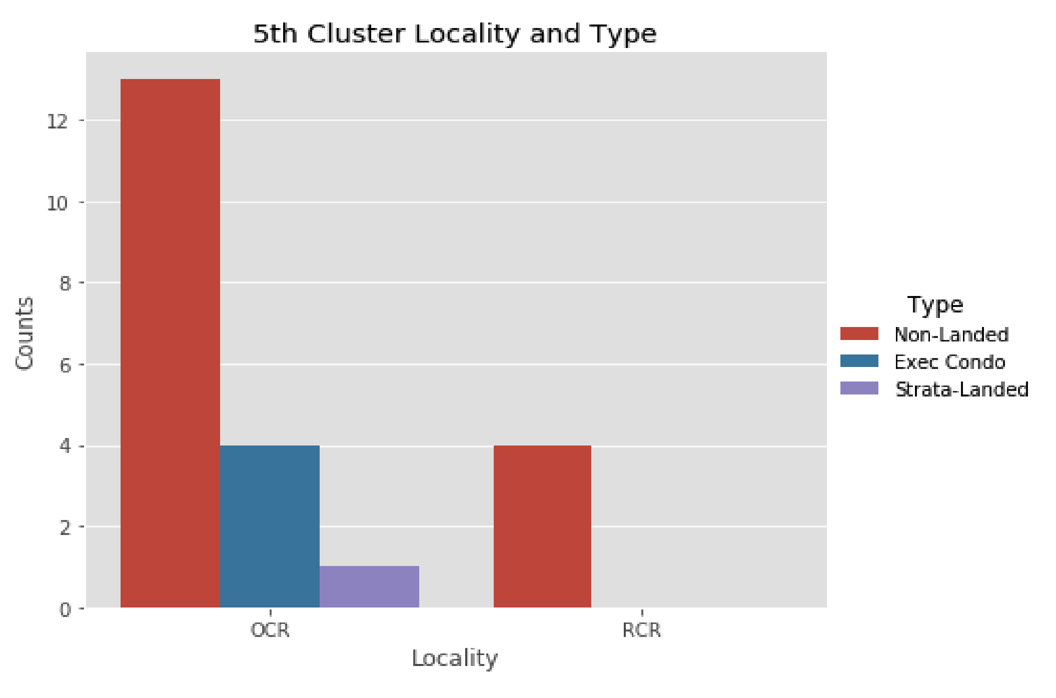




Results and Discussion on Property Clustering: Fifth Cluster

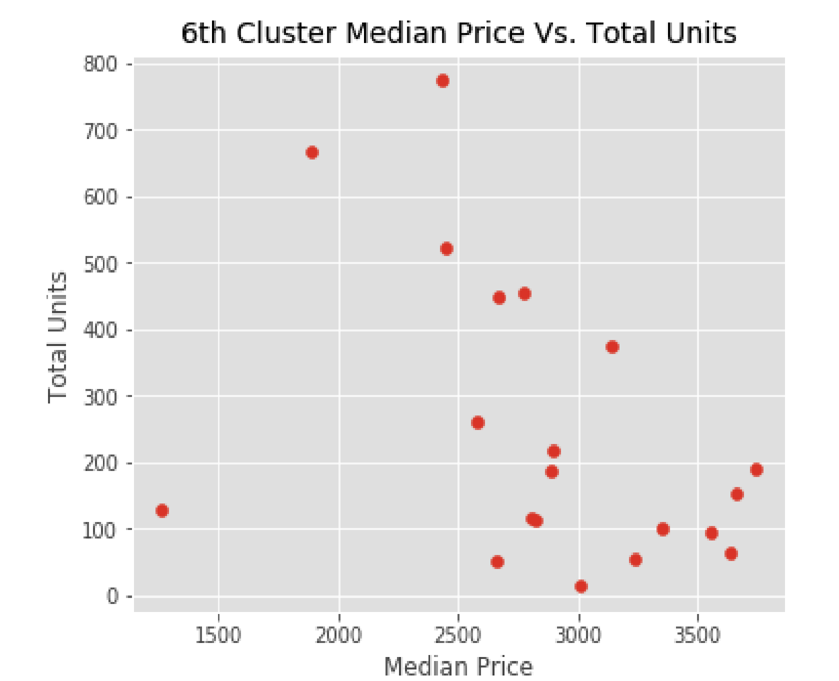
- All priced below 2000 SGD per square feet
- Mainly located in OCR
- Most common venues: coffee shops, food courts, bus stops and gyms

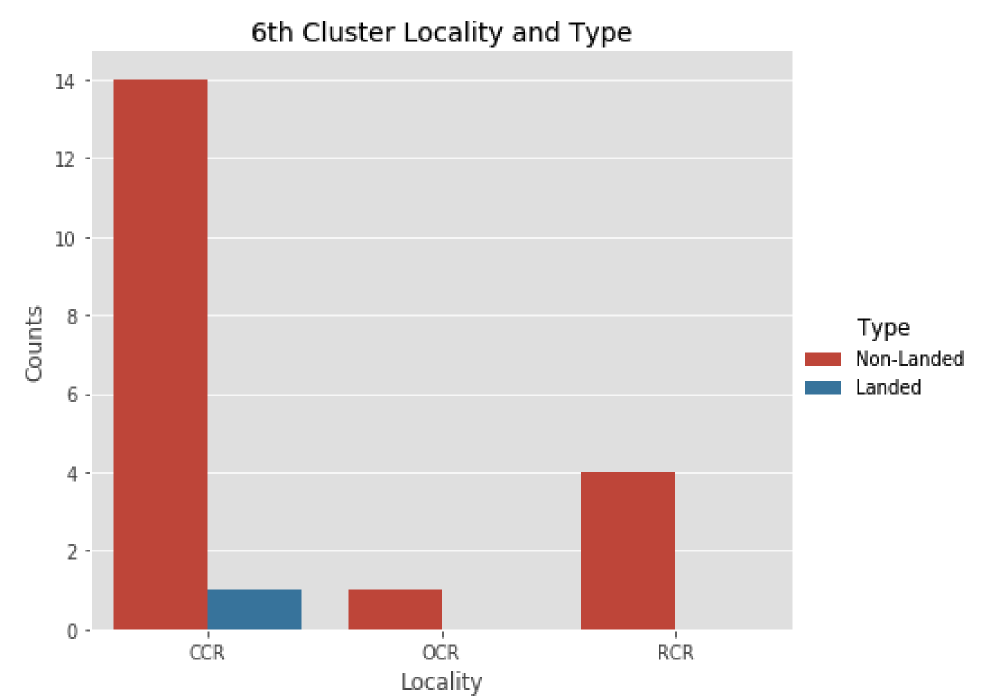




Results and Discussion on Property Clustering: Sixth Cluster

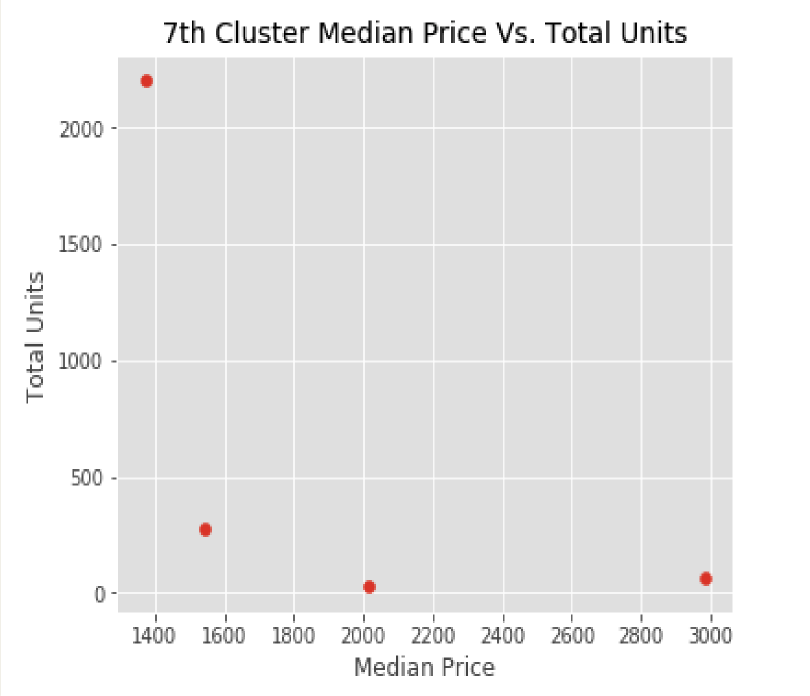
- Priced mainly above 2500 SGD per square feet
- Manly located in CCR and RCR
- Most common venues: hotels, boutiques, and Japanese restaurants.

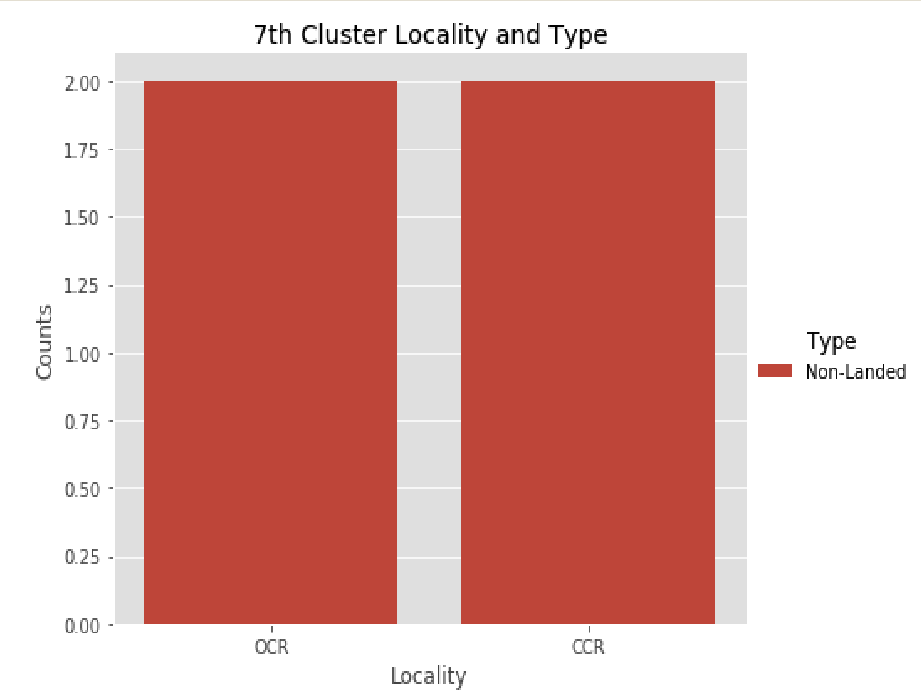




Results and Discussion on Property Clustering: Seventh Cluster

- Only four properties
- Most common venues: bus station, cafes, and restaurants
- Most distinct feature: bus station as the most common venue





More Discussion

- Although the clustering is performed based on the density and type of venues nearby properties, the resulted clusters also exhibit distinct characteristics in terms of location and price
- Particularly obvious in clusters 1, 4, 5, and 6:
- Cluster 1 is priced between SGD 1500 to 2000, and mainly located in RCR region
- Cluster 4 with more average prices between SGD 2000 to 3000 and include quite a significant amount of properties in CCR
- Cluster 5 all priced below SGD 2000, and mainly located in OCR regions
- Cluster 6 mostly have the median price above 2500 SGD, and they are mainly in central regions CCR and RCR
- The categories of nearby venues are also quite distinct, like in cluster 6, hotels, boutiques, and Japanese restaurant are most common, while in cluster 5, the most common types are coffee shops, food courts and bus stops.

More Discussion

- There are 41 properties that have primary school within one kilometer distance
- The price distribution in these properties exhibits a similar trend as that of the median price distribution of all the properties
- Moreover, among these properties, nine of them are located within 1 kilometer distance of the schools that are either affiliated or have special assistance plan
- These nine properties belong to cluster 2, 4, and 5
- Parents with preschool kids may prefer these properties taking into account their children’s admission to primary schools

Conclusion

- Studied Singapore properties in terms of price, total units, locality, type, nearby venues, and primary school
- Generated location data via Geocoder and obtain nearby venue information including primary schools through Foursquare
- Exploratory data analysis performed on the property data to generate an initial understanding of the Singapore properties
- Wrong data entry spotted and corrected
- Clustering of the properties performed based on the types and density of nearby venues
- K-means and grid search applied to find the best clustering results
- Each of the seven clusters examined in terms of detailed listing of common nearby venues and plots of property price, number of units, locality, and type
- Further analyze and discuss cluster characteristics
- Sort out various primary schools nearby properties
- To benefit investors and home buyers who intend to purchase Singapore properties, as well as property agents
  - More systematic view on the property information in Singapore
  - Narrow down their search based on the clusters as well as choosing life styles
  - Useful school information for parents with preschool kids