

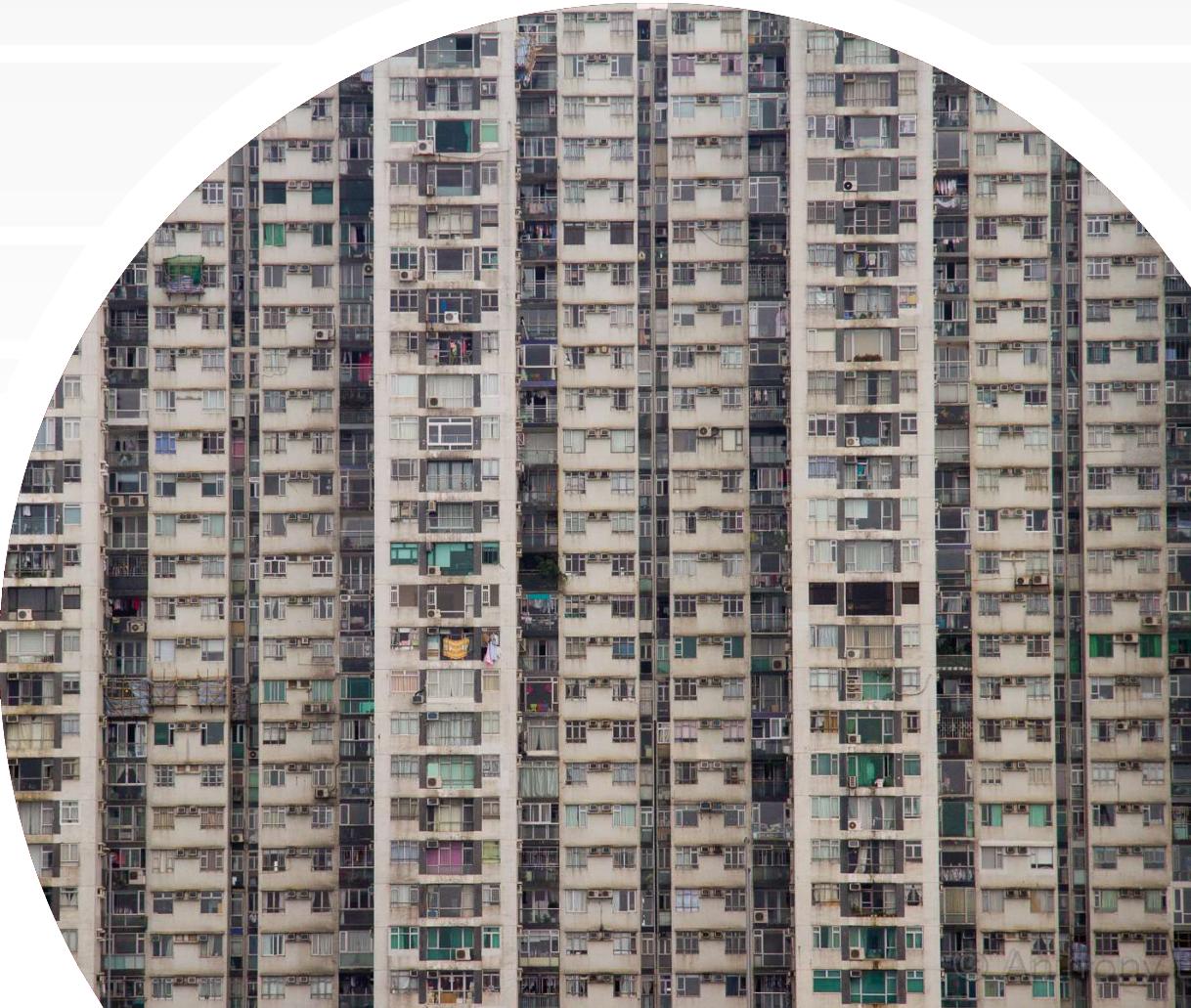
► Applied Data Science Capstone

Selection of Your Dream House in Hong Kong



Table of Contents

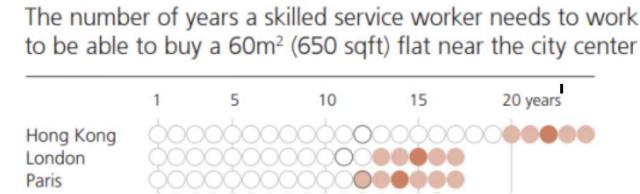
- Introduction**
- Data**
- Methodology**
- Results**
- Discussion and Conclusion**
- Reference**



Introduction



- Hong Kong housing rated the most expensive in the world[1]
- Buying a property of your own is the most important decision
- Hong Kong though a very small city, there are 18 geographically and administratively divided districts, each has its own characteristics
- Currently NO specific info available that can reference the homebuyer's wish list e.g. preferred facilities available, price etc. to the district's characteristics
- How can we help these **homebuyer who would like to buy their dream house in Hong Kong?** Essentially, we want to help them by answering the following question:

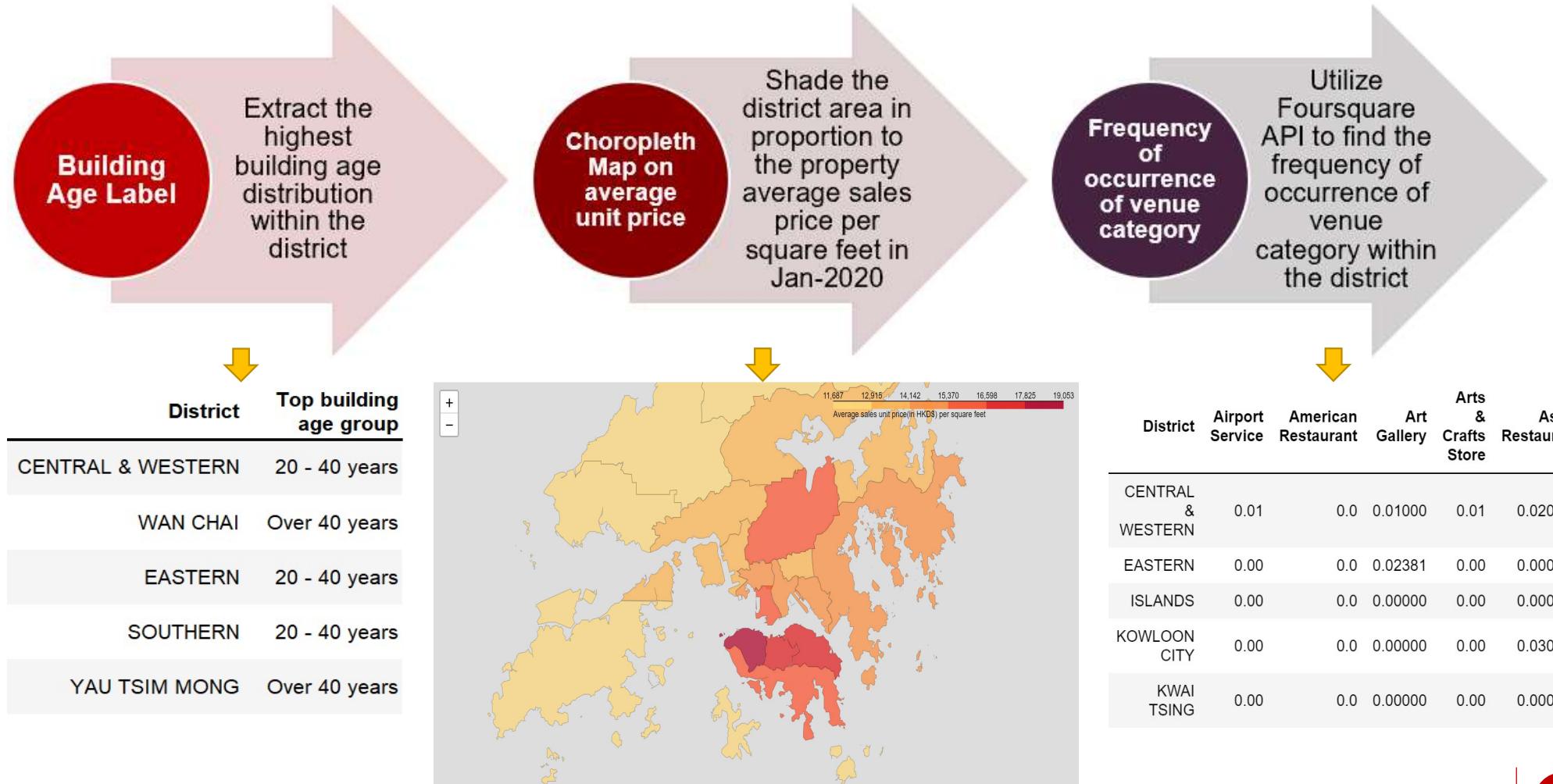


“ Which district area’s characteristics meet most of my criteria of my dream house? ”

Data

Data	Provider	Usage
1. GeoJSON file for the 18 districts[2]	Esri China (HK) Ltd	To visualize the region of space for the 18 districts
2. Coordinates of the 18 districts[3]	Hong Kong GIS Resource	To pass the coordinate to the Foursquare API
3. Stock distribution by the age of private domestic[4]	Centraline Property Agency Limited	One of the consideration factor for homebuyer; to be used as labeling on the map
4. Latest average sales price per square feet per district (Jan-2020)[4]	Centraline Property Agency Limited	Important data for homebuyer ;to be used to visualize on the map
5. Venue data in the district[5]	Foursquare API	To be used for clustering

Methodology



Methodology



Utilize the frequency occurrence of venue category data to cluster the districts into group

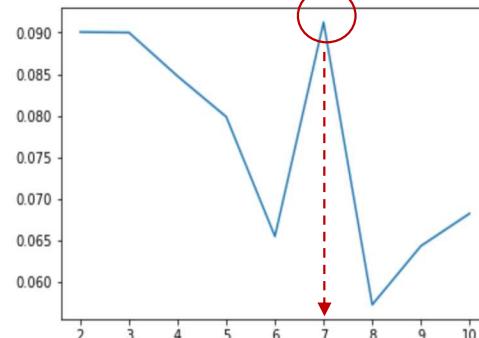


Utilize Silhouette Score to determine the k value

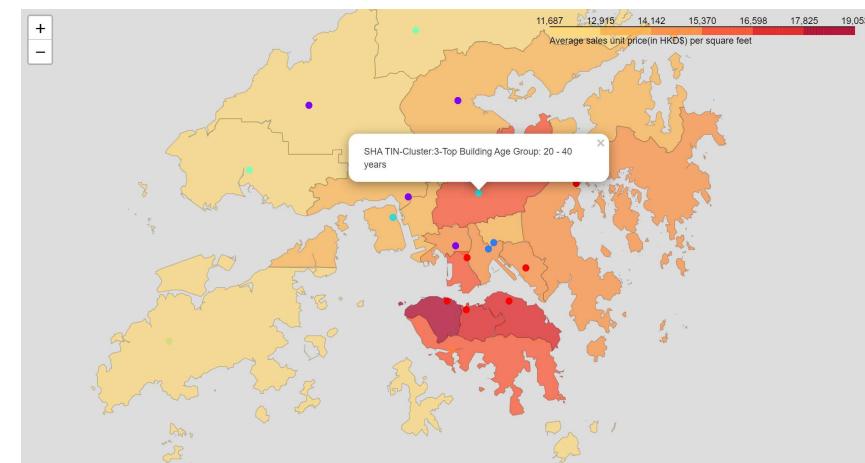


Choropleth map + Cluster points + Building age label

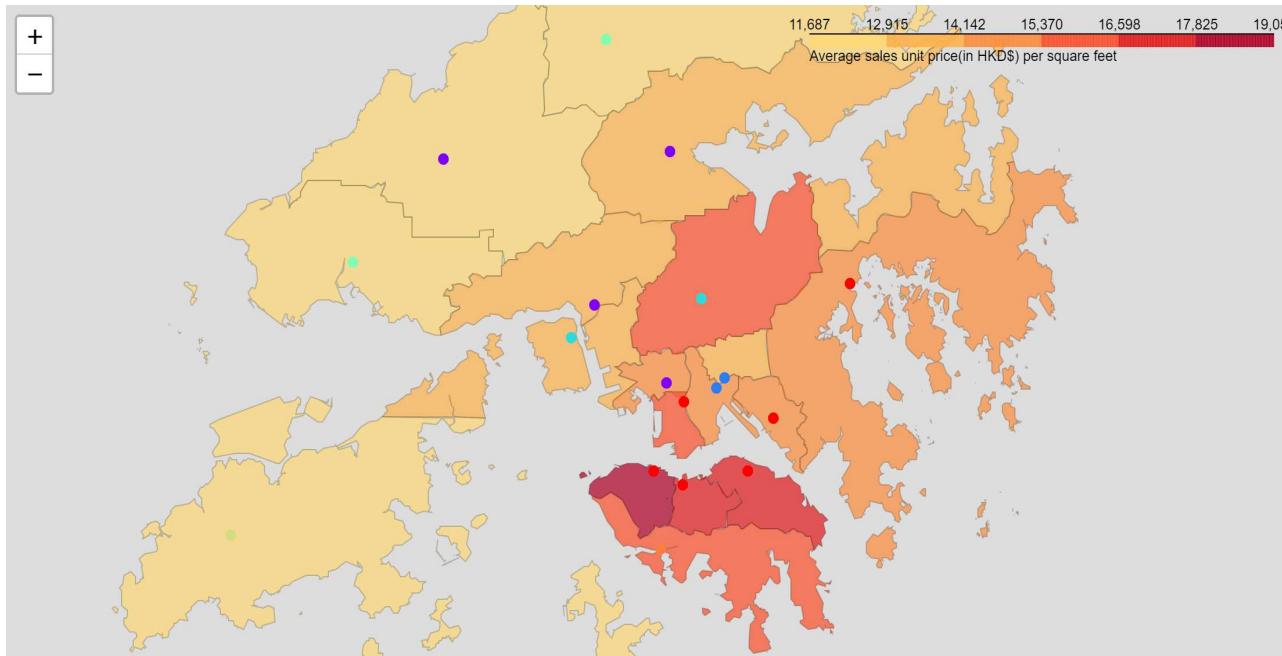
District	Lat	Long	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
YAU TSIM MONG	22.321389	114.1725	0	Coffee Shop	Noodle House	Cha Chaan Teng
KWUN TONG	22.313333	114.225833	0	Chinese Restaurant	Coffee Shop	Café
SAI KUNG	22.381389	114.270556	0	Seafood Restaurant	Café	Thai Restaurant
CENTRAL & WESTERN	22.286667	114.155	0	Coffee Shop	Japanese Restaurant	Chinese Restaurant
WAN CHAI	22.279722	114.171667	0	Coffee Shop	Hotel	Café



The silhouette [6] ranges from -1 to +1, where a high value indicates the object is well matched to its own cluster and poorly matched to neighboring clusters. Hence, k values of 7 was used as shown in above graph of silhouette value against k value.



Results



- K value of 7 was chosen as it is with the highest silhouette score
- Clusters are named according to the top venues available in the cluster
- Cluster points visualized on the choropleth helps homebuyer to assess the characteristics as well as property pricing and building age (through popup label) within the district

Cluster #	Color	Cluster Name	Features (top venues)
0	Red	Commercial and residential mix area with Coffee shops/Café	Coffee shops/Café/Hotel
1	Purple	Local foods area	Noodle house/Cha Chaan Teng/Chinese restaurant
2	Blue	Non-local foods area	Thai restaurant/ Pizza Place / Dessert shop
3	Cyan	Massive bus transportation network area	Bus station/Bus stop
4	Light Green	Recreation facilities area	Athletics&sports/Coffee Shop/Shopping Mall/Zoo
5	Light Yellow	Tourist area	Dessert shop/Cable car
6	Yellow	Others	Sushi restaurant/Thai restaurant/Jewelry store/Furniture store

Discussion & Conclusion

- Though Hong Kong being a very small city, the k-means clustering has grouped districts into 7 clusters each with its own characteristics e.g. cluster #4 with most the recreation facilities available. Based on the results, below in yellow are the recommendations.
- This should help to answer homebuyer's question > Which district area's characteristics meet most of my criteria of my dream house?

Recommendation:

Inadequate venue data

- North, Southern and Wong Tai Sin Districts were with less than 20 number of venues returned from Foursquare API
- Low number of available data points made the clustering results of these districts in question
- More data points gather from other channel e.g. Places API from Google Maps Platform can be included in future study to increase the reliability of the clustering results

Other data

- Other types of data which interest the homebuyer e.g. number of high ranking schools within the districts can be included in future study to enrich the info available to the homebuyer

Reference

[1]: **UBS**

<https://hongkongbusiness.hk/residential-property/news/hong-kong-faces-worlds-greatest-housing-bubble-risk-ubs>

[2]: **Hong Kong 18 districts GeoJSON file**

http://opendata.esrichina.hk/datasets/eea8ff2f12b145f7b33c4eef4f045513_0?geometry=113.136%2C21.852%2C114.992%2C22.298

[3]: **Hong Kong 18 districts coordinates**

<http://www.hkgisa.org.hk/hong-kong-gis-resources>

[4]: **Hong Kong property statistics Jan-2020 from one of the largest property agency - Centaline Property Agency Limited**

http://hkdata.centanet.com/BigData/hong-kong?field=T_POP&sort=default

[5]: **Foursquare**

<https://foursquare.com/city-guide>

[6]: **Silhouette Score**

[https://en.wikipedia.org/wiki/Silhouette_\(clustering\)](https://en.wikipedia.org/wiki/Silhouette_(clustering))

[7]: **Google Map Platform**

<https://developers.google.com/places/web-service/intro>



THANK YOU