



FINDING THE PERFECT HDB FLAT IN SINGAPORE

NAME: GENG ZHAOXIN

EMAIL: dorigeng.lg@gmail.com

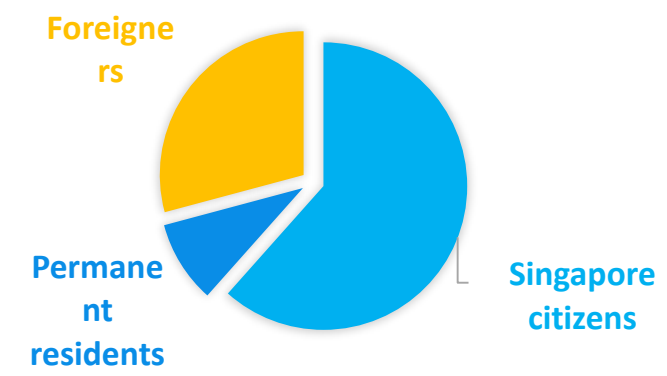
DATE: 8th April 2019



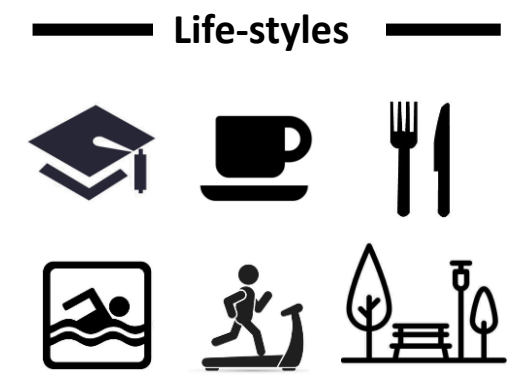
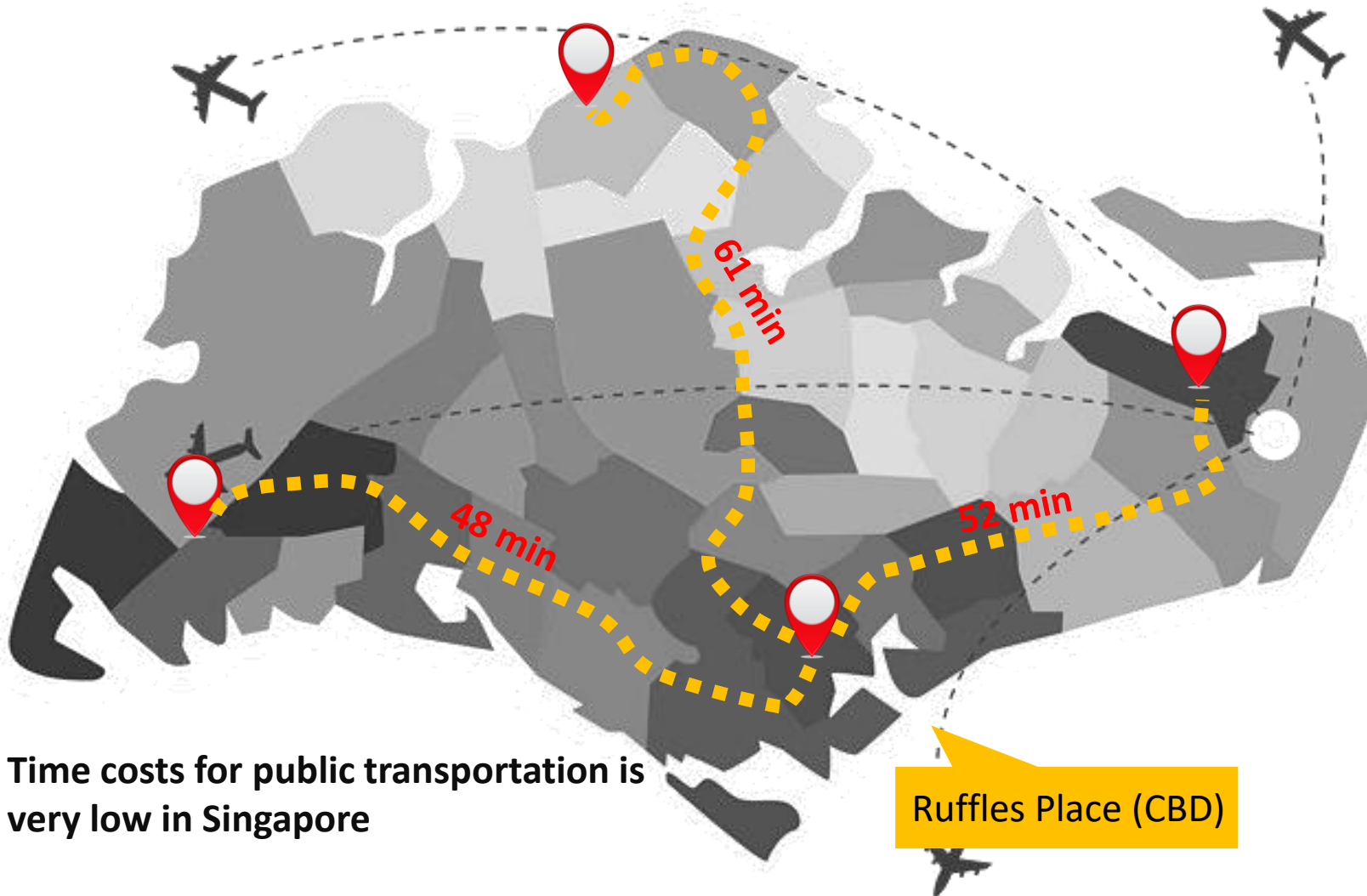
Southeast Asia



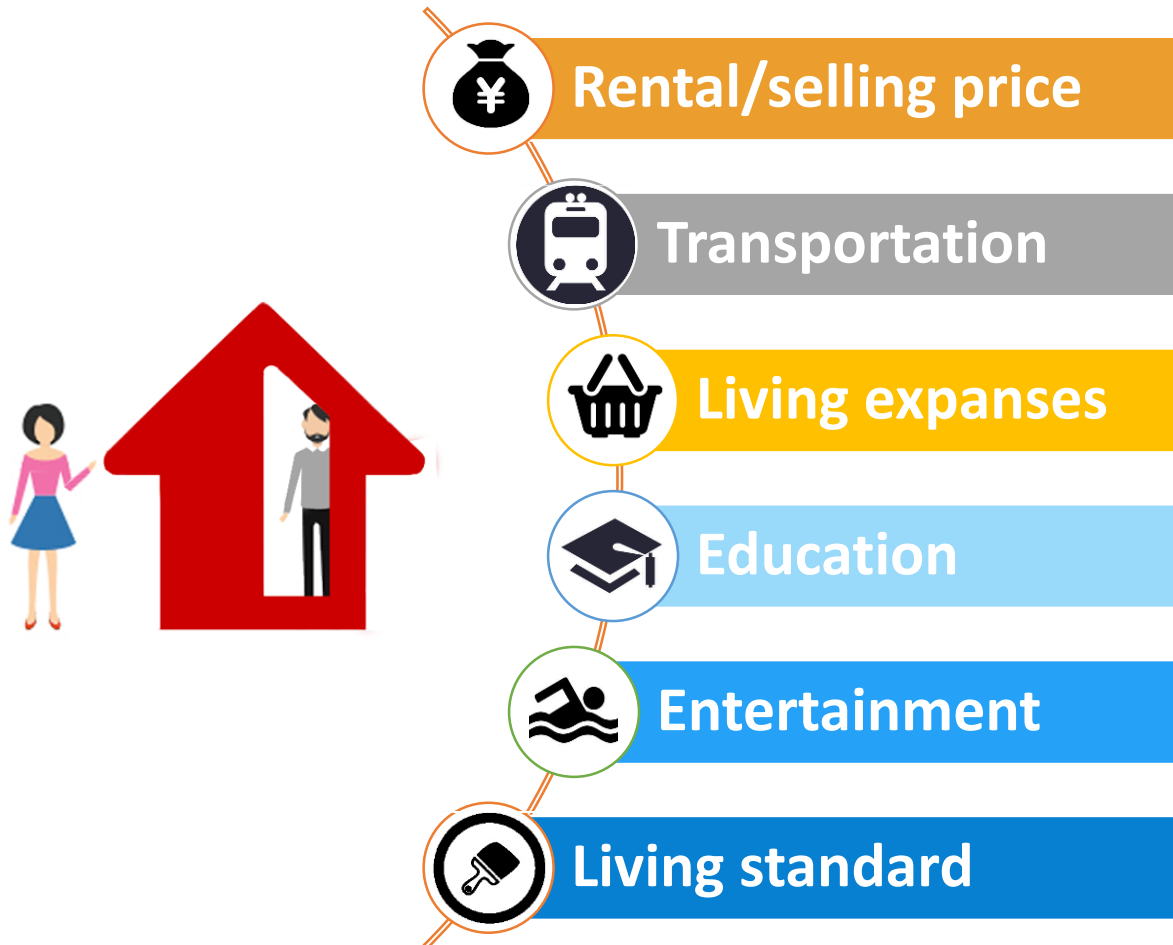
SINGAPORE POPULATION



Where to live?



Key features of property



Explanation	Variable Name
Selling price per square inch	psf_sale
Distance to the nearest MRT exit	dis_MRT
Distance to the nearest Fairprice supermarket	dis_fairprice
Numbers of primary schools within 1km	dis_school
Distance to the nearest swimming complex	dis_pool
The year when the HDB block is completed	year_completed

Where we get the data?



Data resources

HDB blocks data

<https://data.gov.sg/>

Pirce data

<https://www.srx.com.sg/>

Venue data

MRT data



Fairprice data

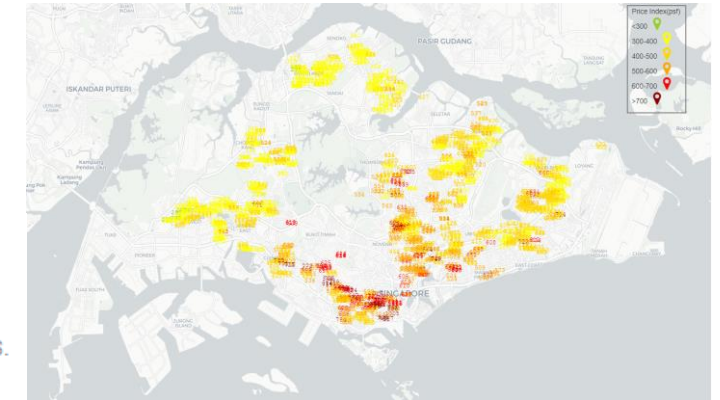
Primary school data

Public swimming complex data

HDB Property Information

Housing and Development Board / 04 Apr 2019

HDB property information contains the location of existing HDB blocks, highest floor level, year of completion, type of building and number of HDB flats (breakdown by flat type) per block etc.



LTA MRT Station Exit

Land Transport Authority / 07 Nov 2018

The layer contains the locations of MRT station exits.



<https://www.fairprice.com.sg/>

PRIMARY SCHOOLS BY
PLANNING AREA IN YEAR 2019

<https://www.moe.gov.sg/admissions/primary-one-registration/information-on-primary-schools/listing-by-planning-area>



Ministry of Education
SINGAPORE



<https://www.myactivesg.com/Facilities/Swimming-Pools>



FOURSQUARE API

What we get from the data?

1. A dataset searching for preferred HDB blocks

Filter

```
] import pandas as pd
df = pd.read_csv('hdb_after_cluster_k11_hie.csv')
df0000 = pd.read_csv('hdb-property-information.csv')

]: df.head()

]:
```

	year_completed	dis_MRT	prim_no	psf_sale	dis_fairprice	dis_pool	cluster_lable	blk_no	street	latitude	longitude
0	2009	199.468223	1.0	1005.1875	220.887653	2301.414936	10	1A	CANTONMENT RD	1.277830	103.840953
1	2014	628.886166	2.0	1075.0000	426.195557	1145.402321	5	10A	BOON TIONG RD	1.286793	103.833135
2	1977	1031.568882	1.0	287.6250	1048.577178	919.340838	0	36	MARSILING DR	1.442024	103.775188
3	1980	1031.568882	1.0	289.2000	1048.577178	919.340838	0	201	MARSILING DR	1.442024	103.775188
4	1989	951.765621	5.0	290.2500	1283.229845	879.546854	0	812	JURONG WEST ST 81	1.346300	103.695902

```
]:
```

filter example

```
df = df[(df['dis_MRT'] <= 350) & (df['prim_no'] >= 3) & (df['year_completed'] >= 2000) & (df['dis_fairprice'] < 500)]
```

```
]: df.head(1)
```

```
]:
```

	year_completed	dis_MRT	prim_no	psf_sale	dis_fairprice	dis_pool	cluster_lable	blk_no	street	latitude	longitude
292	2000	194.583621	4.0	361.625	292.548722	3411.690167	9	351B	CANBERRA RD	1.450891	103.819987

```
]: print('There are ' + str(df.shape[0]) + ' block groups meet your need')
There are 75 block groups meet your need

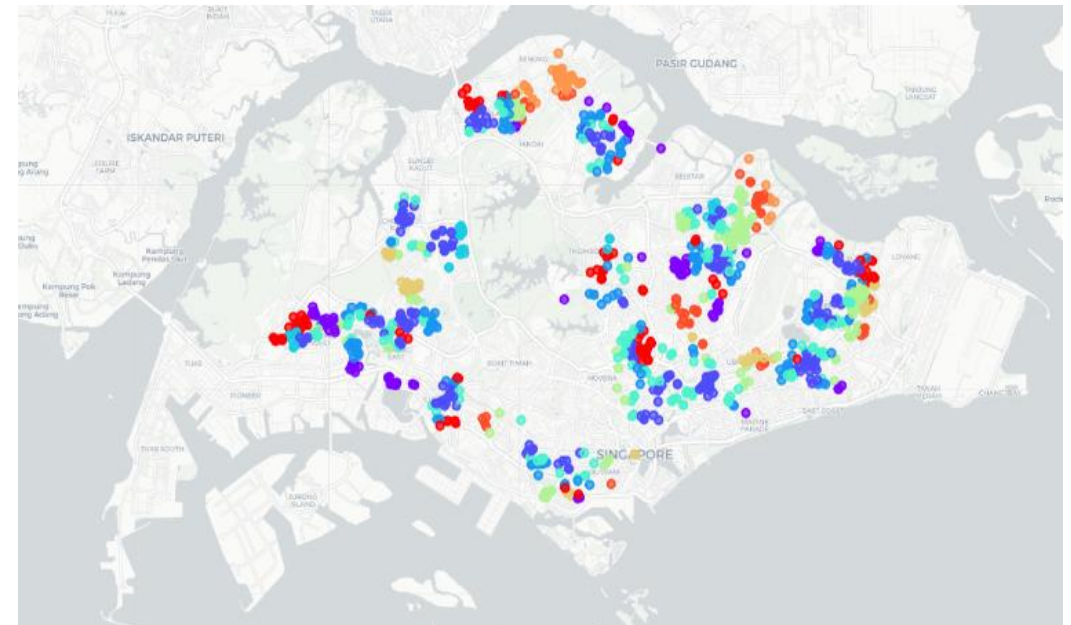
]: # check out similar blocks within the same group
i = 292
df0000[(df0000['year_completed'] == df.loc[i, 'year_completed']) & (df0000['street'] == df.loc[i, 'street']) & (df0000['residential'] == 'Y')]

]:
```

	blk_no	street	max_floor_lvl	year_completed	residential	commercial	market_hawker	miscellaneous	multistorey_carpark	precinct_pavilion	...	3room
5550	351B	CANBERRA RD	14	2000	Y	N	N	N	N	N	...	
5552	351C	CANBERRA RD	14	2000	Y	N	N	N	N	N	...	
5554	351D	CANBERRA RD	14	2000	Y	N	N	N	N	N	...	

An example of how to use HDB filter dataset

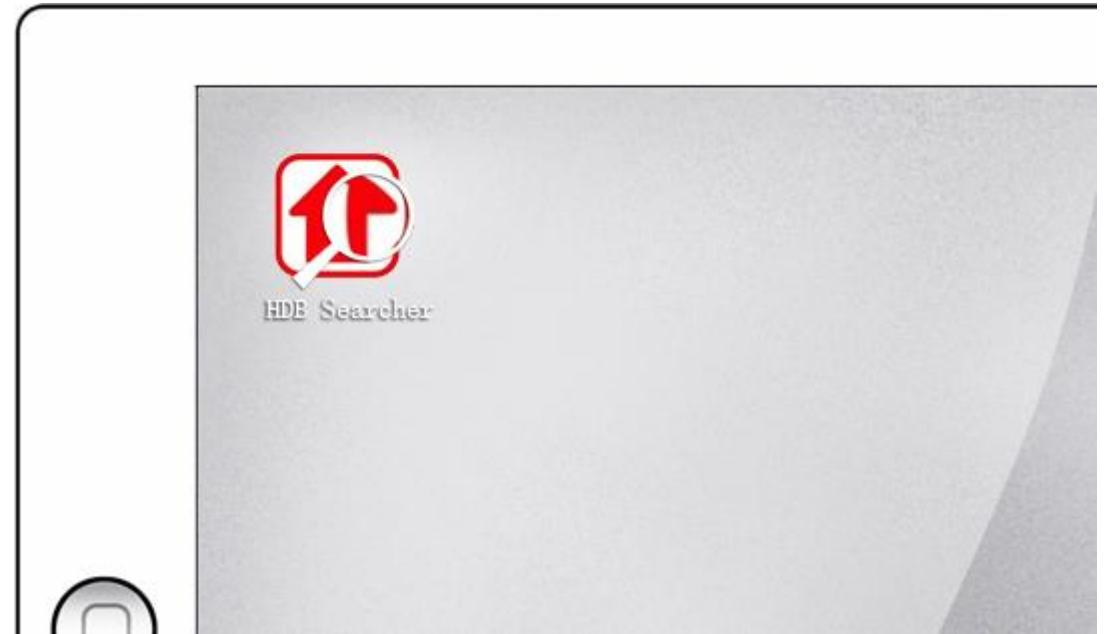
2. Categories based on clustering method



What will we do next?

- Create an app (by using python GUI package PyQt and the dataset we got)

- Extend to condos' market





Thanks

For Your Time