



Prediction of Under Valued Properties in Singapore

BUSINESS PROBLEM / IDEA

An investment firm is exploring options to invest within Singapore due to its central location in Asia. Singapore is one of the largest metropolises in the world where over 5,638,700 (2018 estimate as per Wikipedia) people live and second dense country in the world with 7,804/km² (20,212.3/sq mi) population density. Investor did not apply any area restriction so the analysis is applied on all 28 districts of Singapore for predicting most attractive investment options. The overall purpose is to predict underpriced properties based on multiple parameters and avoid over-priced venues.

The property prices based on three major factors are considered:

- 1) Type property tenure
- 2) Distance from city central area, and
- 3) Proximity with major shopping/restaurant areas.

Following systematic approach will be applied to achieve desired results:

- Gather above mentioned three factor and current housing prices
- Derive housing prices formulae based on these three factors
- Shortlist top twenty undervalued locations (outliers) for further investigation and top twenty overvalued ones for rejection.

DATA

Based on business problems, following data sources are used to achieve desired outcomes:

- Recent property price in various neighbors using webscraping from websites
- Estates' locations and driving distances from city central area using GoogleMap API
- Proximity with major shopping areas using FourSquare API

Dated: 12th October 2019

Student: Ghazi Salah Uddin

DATA GATHERING

As first step, condominium-sale transactions data is collected using Web scrapping from <https://www.squarefoot.com.sg/latest-transactions/sale/residential/condominium>, which provides latest transactions. In total **704** condominium-sale transactions, records were collected from this source as by default it provides only last one month transactions details.

Table1: The sample of web scrapped data

Index	Date	District	Project Name	Address	PropertyType	Tenure	Area(sq ft)	Typeof Area	Price(\$psf)	Price(\$)
0	06-Oct-19	21	MAYFAIR MODERN		Condominium	99 Yrs FROM 2018	1,044	Strata	2,165	2,261,000
1	06-Oct-19	9	HAUS ON HANDY		Condominium	99 Yrs FROM 2018	743	Strata	2,865	2,128,000
2	06-Oct-19	18	TREASURE AT TAMPINES		Condominium	99 Yrs FROM 2018	1,033	Strata	1,328	1,372,000
3	06-Oct-19	18	TREASURE AT TAMPINES		Condominium	99 Yrs FROM 2018	1,238	Strata	1,282	1,587,000
4	06-Oct-19	18	TREASURE AT TAMPINES		Condominium	99 Yrs FROM 2018	915	Strata	1,271	1,163,000
5	06-Oct-19	18	TREASURE AT TAMPINES		Condominium	99 Yrs FROM 2018	592	Strata	1,372	812,000
6	05-Oct-19	28	PARC BOTANNIA		Condominium	99 Yrs FROM 2016	980	Strata	1,401	1,372,000
7	05-Oct-19	13	THE TRE VER		Condominium	99 Yrs FROM 2018	495	Strata	1,686	835,000
8	05-Oct-19	15	AMBER PARK		Condominium	Freehold	743	Strata	2,449	1,819,260
9	05-Oct-19	21	MAYFAIR MODERN		Condominium	99 Yrs FROM 2018	624	Strata	2,162	1,350,000

DATA CLEANING

Various data cleaning and compiling relevant data is performed as various steps:

- The transactions were grouped by each condominium (Project Name) and average per square foot (PSF) were calculated. After grouping, a total of **275** condominium records were found.
- Redundant columns such as Type of Area and Property Type were removed.
- Using the Google Map API, longitude and latitude information of each condominium were gathered and added into the table.
- Distance from central area were computed and added to the table.
- Then, locations are mapped on the map with prices highest to lowest are color coded (Dark-red, Red, Orange, Blue, Green, and Grey respectively) for visualizing.

Table2: Consolidate data table after data cleaning and collection.

	Project Name	Tenure	Price(\$psf)	Longitude	Latitude	DistanceWithCentral(km)
0	WESTCOVE CONDOMINIUM	99 Yrs FROM 1995	875.0	103.767842	1.299290	9.461205
1	WEST BAY CONDOMINIUM	99 Yrs FROM 1991	811.0	103.766799	1.300242	9.595547
2	WATERFRONT WAVES	99 Yrs FROM 2007	986.0	103.931683	1.339192	10.892026
3	WATERFRONT KEY	99 Yrs FROM 2007	1203.4	103.932551	1.338767	10.944877
4	WATERFRONT ISLE	99 Yrs FROM 2009	1212.0	103.929216	1.340098	10.726281
5	WATERFRONT GOLD	99 Yrs FROM 2009	1192.0	103.930007	1.339721	10.773228
6	WATERBANK AT DAKOTA	99 Yrs FROM 2009	1650.0	103.889377	1.308146	5.068365
7	WATER PLACE	99 Yrs FROM 1998	1332.0	103.873747	1.298078	3.003435
8	VOLARI	Freehold	2054.0	103.834276	1.316262	4.137248
9	VIZ AT HOLLAND	Freehold	1314.0	103.803880	1.309969	6.061709

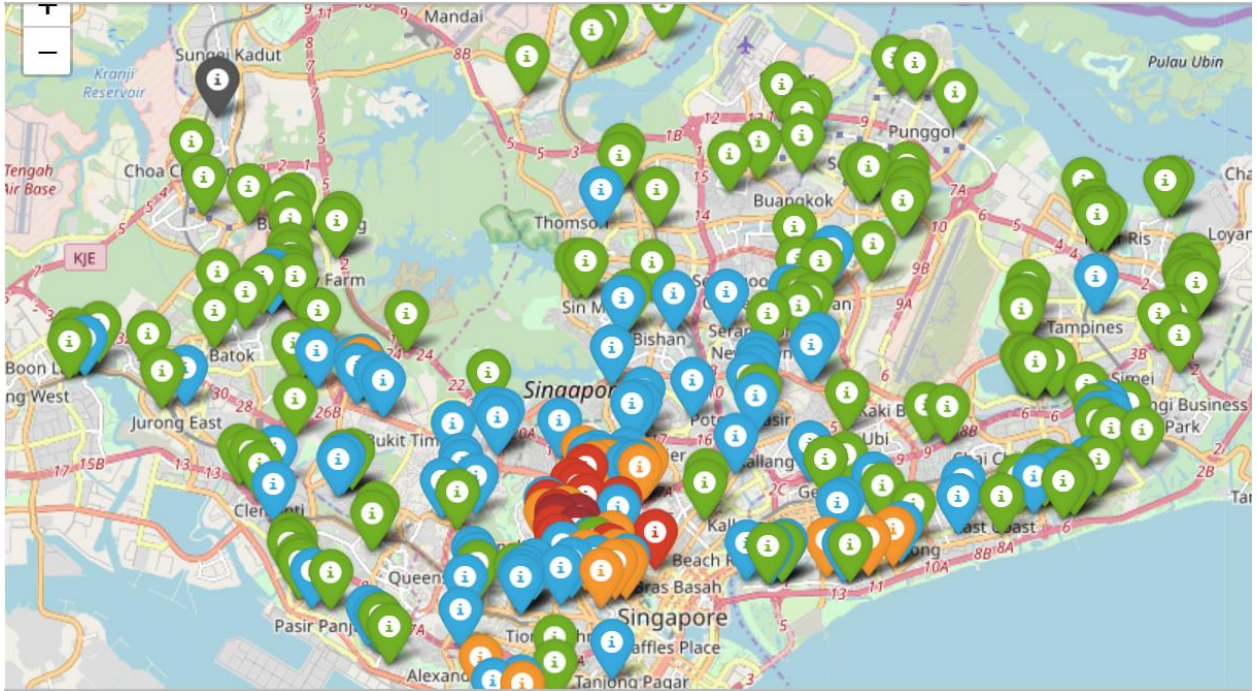


Figure 1. Location map with color-coded psf information

Following Histogram shows PSF price distribution

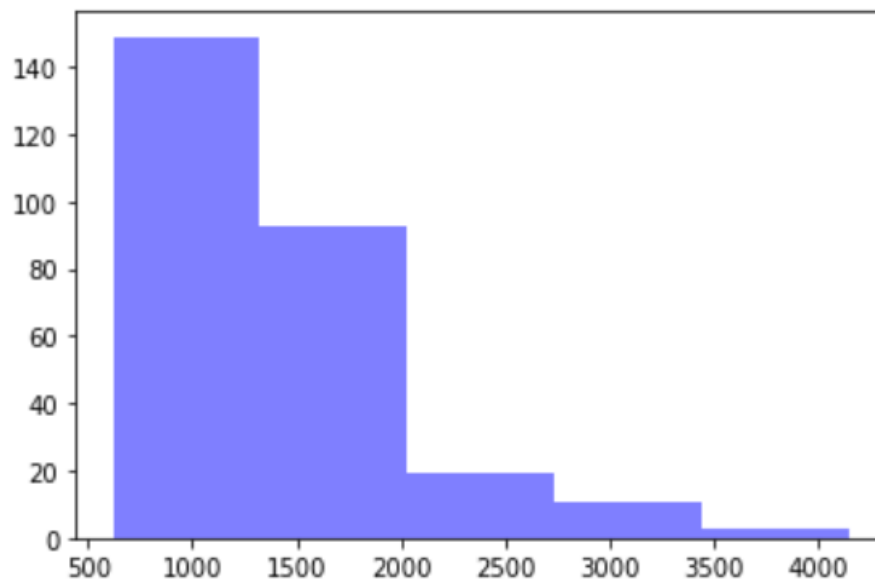


Figure 2. PSF Price Distribution

ANALYSIS

As shown in the location map, it is difficult to analyse over 300 properties with psf price ranging from S\$500 to S\$4,000. Therefore, data science methodology is used to build a price model.

Before building the model let study price correlation with each parameters: distance from central, having low lease year, newer condominium, and accessibility to restaurants.

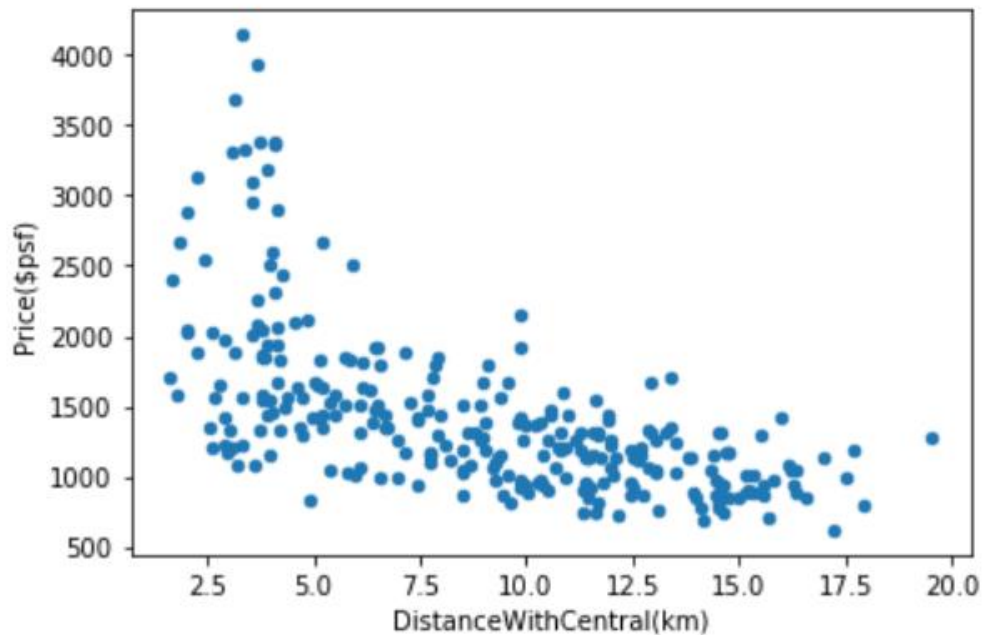


Figure 3. Distance from central vs price

There are strong correlation between the price and the distance from central area. Let look at having low lease year parameter contribution to the price.

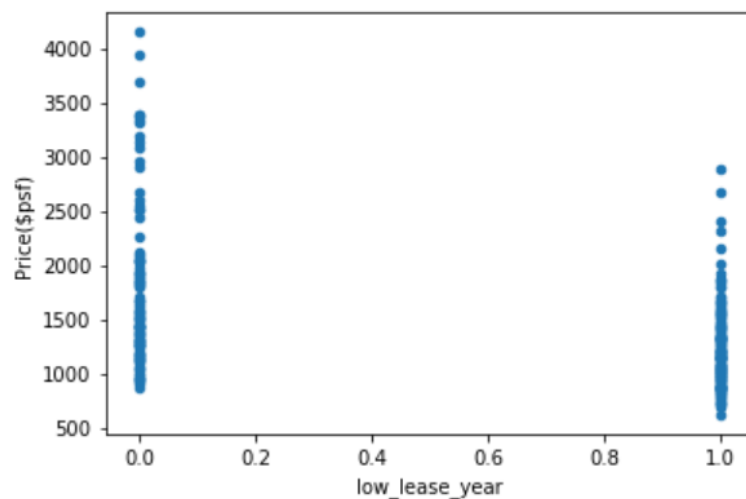


Figure 4. Low lease vs price

According to the data, there is a visual limit for how much maximum people are willing to pay for low lease property type. The data makes sense as well. The question is how exactly it contributes towards price. Let investigate the last parameter, accessibility to restaurants.

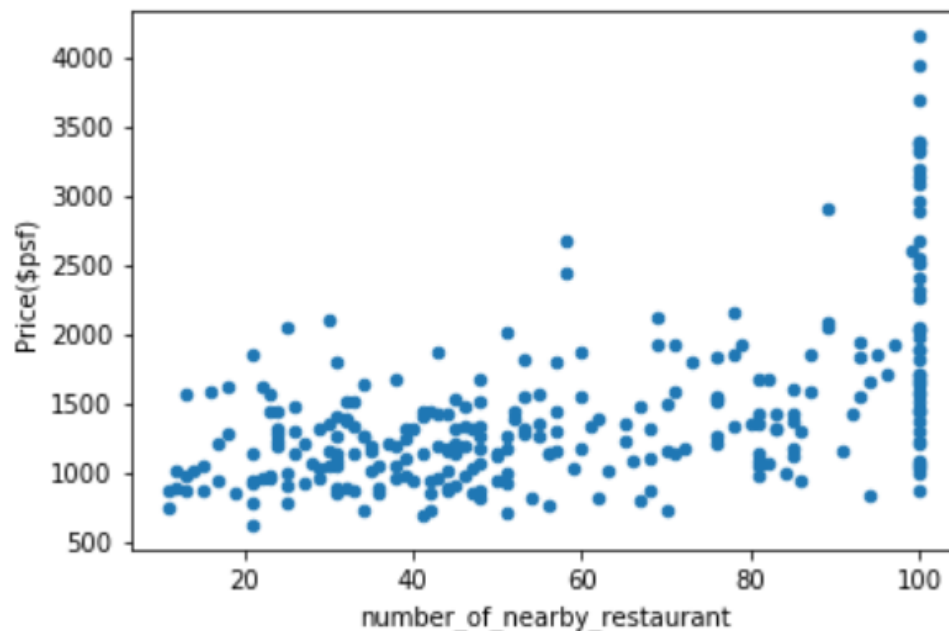


Figure 5. Number of nearby restaurants vs price

There is slight contribution of the accessibility to restaurants towards price but not obvious. It could be because most people staying in condo tend to own private cars. It is very clear that property prices are evaluated based on multiple factors.

To build the price model, we need to consider multiple parameters at the same time. After input parameters are normalized, multi-variable linear regression was used to build the price model. Coefficients of the model are

$[-243.61350285 \ -201.67559929 \ 26.75175797 \ -80.42373764 \ 138.9311806]$

RESULT AND DISCUSSION

According to the coefficients, Distance from Central (-243.61) is the highest negative contributor and number of nearby restaurants (138.93) is the highest positive contributor. However, the distance parameter contributes about double than the restaurant parameter. Freehold type (-80.42) is more preference than lease hold type (-201.67). The number of remaining lease year (26.75) positively affect the price as well.

With the model, top twenties undervalued properties will be shortlisted for further investigations.

Table 3. Top 20 Under Valued Properties

Project Name	Project Name	Tenure	Price(\$psf)	Longitude	Latitude	DistanceWithCentral(km)	low_lease_year	freehold	Yr_From	number_of_nearby_restaurant	PriceDifferencePercent
KANDIS RESIDENCE	KANDIS RESIDENCE	99 Yrs FROM 2016	1278.857143	103.838741	1.459187	19.531974	1	0	2016	24	-0.553447
TWENTYONE ANGULLIA PARK	TWENTYONE ANGULLIA PARK	Freehold	4146	103.83083	1.304666	3.305955	0	1	2019	100	-0.491164
3 ORCHARD BY-THE-PARK	3 ORCHARD BY-THE-PARK	Freehold	3938	103.826118	1.304599	3.682151	0	1	2019	100	-0.469679
THE MARQ ON PATERSON HILL	THE MARQ ON PATERSON HILL	Freehold	3690	103.830257	1.301872	3.136714	0	1	2019	100	-0.425696
THE JOVELL	THE JOVELL	99 Yrs FROM 2018	1286.875	103.966701	1.362074	15.53681	1	0	2018	18	-0.402709
THE LAKEFRONT RESIDENCES	THE LAKEFRONT RESIDENCES	99 Yrs FROM 2010	1419	103.721147	1.343864	15.971925	1	0	2010	44	-0.390081
NOUVEL 18	NOUVEL 18	Freehold	3384	103.829453	1.312404	4.059904	0	1	2019	100	-0.389158
THE NASSIM	THE NASSIM	Freehold	3372	103.82388	1.307686	4.09427	0	1	2019	100	-0.38756
SYMPHONY SUITES	SYMPHONY SUITES	99 Yrs FROM 2014	1135	103.845801	1.436482	16.982261	1	0	2014	21	-0.384601
PARC ROSEWOOD	PARC ROSEWOOD	99 Yrs FROM 2011	1185	103.786742	1.429247	17.695395	1	0	2011	38	-0.375956
BISHOPSGATE RESIDENCES	BISHOPSGATE RESIDENCES	Freehold	3335	103.82586	1.299887	3.391567	0	1	2019	100	-0.368875
HILLTOPS	HILLTOPS	Freehold	3309	103.839904	1.308442	3.085894	0	1	2019	100	-0.358703
MAYFAIR MODERN	MAYFAIR MODERN	99 Yrs FROM 2018	2150.2	103.782804	1.339202	9.83515	1	0	2018	78	-0.355352
ARDMORE PARK	ARDMORE PARK	Freehold	3189	103.830589	1.311391	3.894594	0	1	2019	100	-0.348882
CUSCADEN RESERVE	CUSCADEN RESERVE	99 Years Leasehold	3392	103.825885	1.305391	3.758971	0	0	2019	100	-0.337042
DUCHESS RESIDENCES	DUCHESS RESIDENCES	999 Yrs FROM 1875	1877	103.802963	1.325245	7.124541	1	0	1875	43	-0.336368
MEYERHOUSE	MEYERHOUSE	Freehold	2673	103.895211	1.299551	5.21836	0	1	2019	58	-0.328583
HAUS ON HANDY	HAUS ON HANDY	99 Yrs FROM 2018	2877	103.847918	1.301023	2.028145	1	0	2018	100	-0.327388
THE CLAYMORE	THE CLAYMORE	Freehold	3090	103.832349	1.309124	3.576628	0	1	2019	100	-0.322214
8 SAINT THOMAS	8 SAINT THOMAS	Freehold	3140.25	103.837581	1.297821	2.239844	0	1	2019	100	-0.309038

Table 4. Top 20 Over Valued Properties

Project Name	Project Name	Tenure	Price(\$psf)	Longitude	Latitude	DistanceWithCentral(km)	low_lease_year	freehold	Yr_From	number_of_nearby_restaurant	PriceDifferencePercent
THE PEARL @ MOUNT FABER	THE PEARL @ MOUNT FABER	99 Yrs FROM 2002	1213	103.825626	1.272414	3.087742	1	0	2002	76	0.437333
SELETAR SPRINGS CONDOMINIUM	SELETAR SPRINGS CONDOMINIUM	99 Yrs FROM 1997	730	103.880424	1.388929	12.153288	1	0	1997	42	0.452201
VILLA MARINA	VILLA MARINA	99 Yrs FROM 1995	973	103.928762	1.313282	9.248241	1	0	1995	81	0.453831
GRANDEUR 8	GRANDEUR 8	99 Yrs FROM 2002	937.333333	103.848436	1.37691	10.389008	1	0	2002	86	0.473083
NOUVELLE PARK	NOUVELLE PARK	Freehold	942	103.878876	1.368738	9.964798	0	1	2019	40	0.526685
WESTCOVE CONDOMINIUM	WESTCOVE CONDOMINIUM	99 Yrs FROM 1995	875	103.767842	1.29929	9.461205	1	0	1995	68	0.529695
TANJONG RIA CONDOMINIUM	TANJONG RIA CONDOMINIUM	99 Yrs FROM 1993	1090	103.87593	1.29818	3.214245	1	0	1993	66	0.540775
HORIZON TOWERS	HORIZON TOWERS	99 Yrs FROM 1979	1211	103.834371	1.299345	2.610705	1	0	1979	100	0.54356
CHERRYHILL	CHERRYHILL	Freehold	1126	103.879992	1.351115	8.180725	0	1	2019	85	0.563604
DUNMAN VIEW	DUNMAN VIEW	99 Yrs FROM 1997	1061	103.898439	1.311275	6.107121	1	0	1997	100	0.590171
WEST BAY CONDOMINIUM	WEST BAY CONDOMINIUM	99 Yrs FROM 1991	811	103.766799	1.300242	9.595547	1	0	1991	62	0.600426
SUITES @ EAST COAST	SUITES @ EAST COAST	Freehold	929	103.933791	1.315858	9.873363	0	1	2019	51	0.611966
CENTRAL GROVE	CENTRAL GROVE	99 Yrs FROM 1997	1050	103.885773	1.317336	5.397597	1	0	1997	100	0.644962
SIMSVILLE	SIMSVILLE	99 Yrs FROM 1994	1032	103.890117	1.318294	5.821762	1	0	1994	100	0.648001
KERRISDALE	KERRISDALE	99 Yrs FROM 1998	1083	103.859324	1.314626	3.609284	1	0	1998	100	0.688798
TERESA VILLE	TERESA VILLE	Freehold	1168	103.82545	1.277691	2.93622	0	1	2019	72	0.705894
CITY SQUARE RESIDENCES	CITY SQUARE RESIDENCES	Freehold	1219	103.858816	1.311909	3.304211	0	1	2019	100	0.730707
LE CRESCENDO	LE CRESCENDO	Freehold	992	103.890345	1.332034	6.948355	0	1	2019	84	0.83995
MALVERN SPRINGS	MALVERN SPRINGS	Freehold	1000	103.905195	1.307356	6.574975	0	1	2019	100	0.925165
QUEENS	QUEENS	99 Yrs FROM 1998	833	103.808441	1.294263	4.929774	1	0	1998	94	1.070684

Here is the map of how outliers distribute. One cluster of under values properties is identified within center, while overvalued properties are found scattered across the Singapore rather than gathered in specific district.

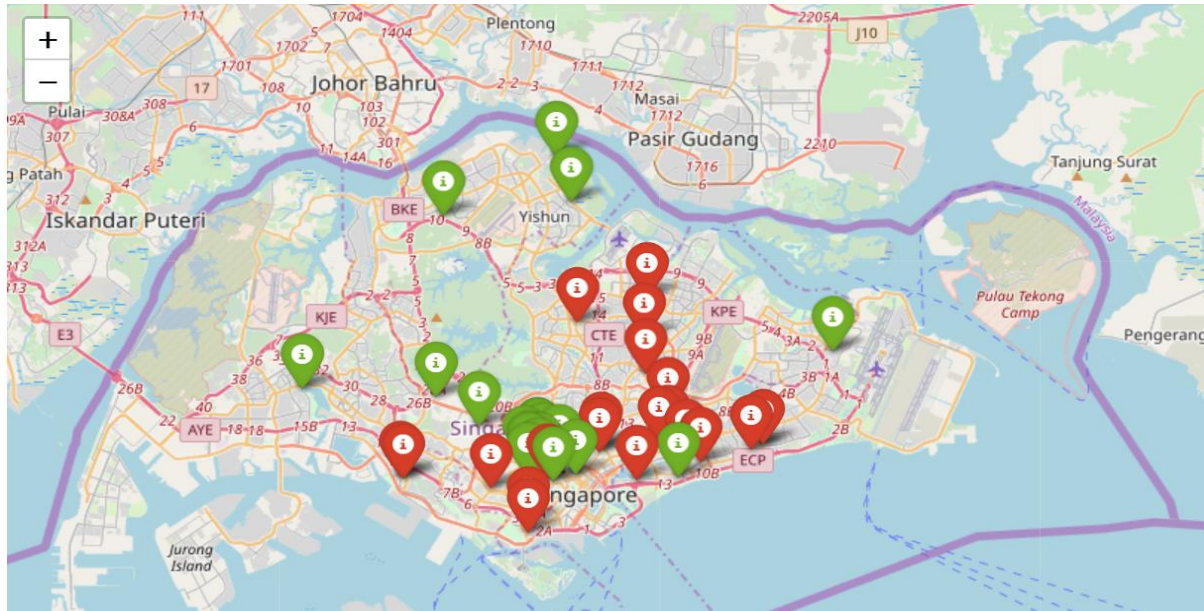


Figure 6. Under-Valued (Green) and Over-Valued (Red) Properties Map

References:

- [1] <https://www.squarefoot.com.sg/latest-transactions/sale/residential/condominium>
- [2] Google Maps API documentation
- [3] CloudSquare API documentation

Disclimer: This analysis is only from study perspective and should be analyzed alot before expecting any commercial benefits from this analysis.