Prediction of Under Valued Properties in Singapore

By Asadullah Khan Niazi

Dated – 25th Feb 2020

BUSINESS PROBLEM / IDEA

An investment firm is exploring options to invest in property 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 will be 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 will be considered:

1) Type of property (lease or owned) & lease tenure
2) Distance from city central area, and
3) Proximity with major shopping/restaurant areas.

Following step-by-step 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 will be used to achieve desired outcomes:

- Recent property price in various neighbors using webscraping from websites https://www.squarefoot.com.sg/latest-transactions/sale/residential/condominium
    - Reading the data
    - Webscrapping it from website
    - Cleaning the Webscrapped data

- Estates' locations and driving distances from city central area using GoogleMap API
    - Using the GoogleMAP API (after creating an account)
    - Calculate distances of the properties from the central locations and adding the information in the data frame

- Proximity with major shopping areas, Restaurants using FourSquare API
    - Using the FourSquare API, finding out the no. of important locations (e.g. Restaurants, Shopping Malls) in the vicinity of properties and adding this data in the data frame

- Understanding all the data using linear regressions
- Building the Model and get the necessary property categories

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 **795** condominium-sale transactions, records were collected from this source as by default is provides only last one month transactions details – it showed transactions from 1st Jan 2020.

Table 1: Sample of Web-scrapped data

	Date	District	Project Name	Address	PropertyType	Tenure	Area(sq ft)	Typeof Area	Price($psf)	Price($)
0	16 Feb 2020	28	PARC BOTANNIA		Condominium	99 Yrs FROM 2016	667	Strata	1,405	937,860
1	16 Feb 2020	20	JADESCAPE		Condominium	99 Yrs FROM 2018	527	Strata	1,751	923,700
2	16 Feb 2020	18	TREASURE AT TAMPINES		Condominium	99 Yrs FROM 2018	581	Strata	1,442	838,112
3	16 Feb 2020	18	TREASURE AT TAMPINES		Condominium	99 Yrs FROM 2018	1,012	Strata	1,329	1,345,112
4	16 Feb 2020	18	TREASURE AT TAMPINES		Condominium	99 Yrs FROM 2018	592	Strata	1,416	838,112
5	16 Feb 2020	18	TREASURE AT TAMPINES		Condominium	99 Yrs FROM 2018	1,335	Strata	1,353	1,806,112
6	16 Feb 2020	18	TREASURE AT TAMPINES		Condominium	99 Yrs FROM 2018	592	Strata	1,363	807,112
7	16 Feb 2020	18	TREASURE AT TAMPINES		Condominium	99 Yrs FROM 2018	581	Strata	1,452	844,112
8	16 Feb 2020	28	PARC BOTANNIA		Condominium	99 Yrs FROM 2016	980	Strata	1,436	1,406,790
9	15 Feb 2020	18	THE TAPESTRY		Condominium	99 Yrs FROM 2017	441	Strata	1,529	674,860

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 **277** 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.

Table 2: Consolidated Data Table with consolidation and collection

	Project Name	Tenure	Price($psf)	Longitude	Latitude	DistanceWithCentral(km)
0	WOODHAVEN	99 Yrs FROM 2011	1015.0	103.785410	1.431580	17.991419
1	WOODGROVE CONDOMINIUM	99 Yrs FROM 1997	673.5	103.783327	1.430415	17.970155
2	WEST BAY CONDOMINIUM	99 Yrs FROM 1991	801.0	103.766799	1.300242	9.595547
3	WATERFRONT KEY	99 Yrs FROM 2007	1066.0	103.932551	1.338767	10.944877
4	WATERFRONT ISLE	99 Yrs FROM 2009	1254.5	103.929216	1.340098	10.726281
5	WATERBANK AT DAKOTA	99 Yrs FROM 2009	1730.0	103.889742	1.307924	5.089083
6	WATER PLACE	99 Yrs FROM 1998	1182.5	103.873747	1.298078	3.003435
7	VIZ AT HOLLAND	Freehold	1684.0	103.803880	1.309969	6.061709
8	VILLA MARINA	99 Yrs FROM 1995	1001.0	103.928008	1.311607	9.104174
9	VARSITY PARK CONDOMINIUM	99 Yrs FROM 2004	929.0	103.768503	1.295721	9.321353

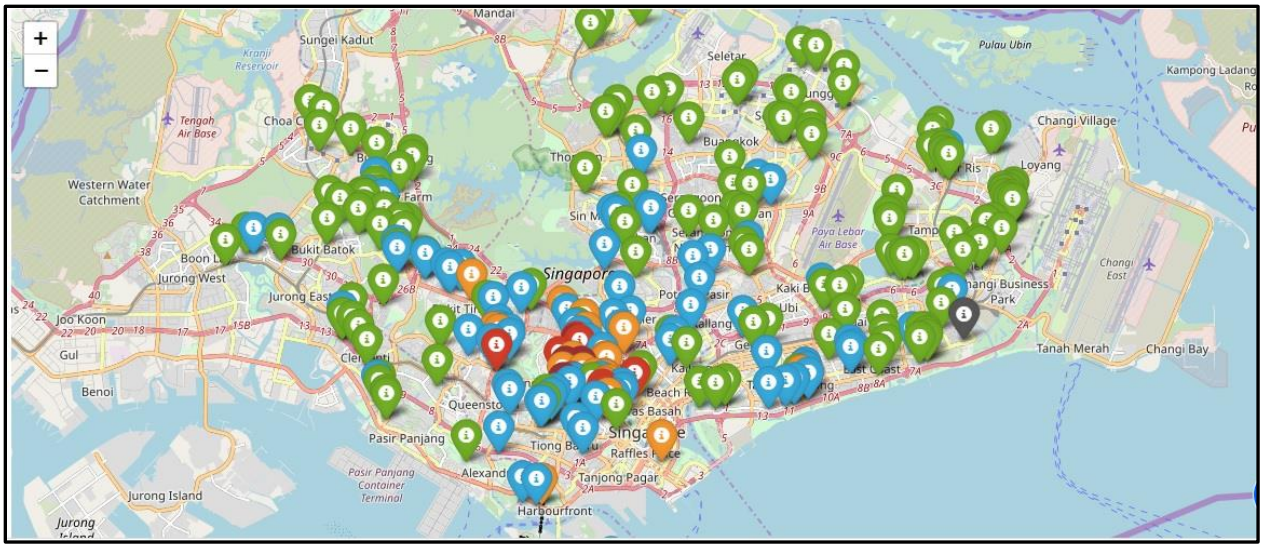

Figure 1: Location Map - Prices highest to lowest are color coded (Dark-red, Red, Orange, Blue, Green, and Grey respectively

Following Histogram shows the Price per Square Foot distribution

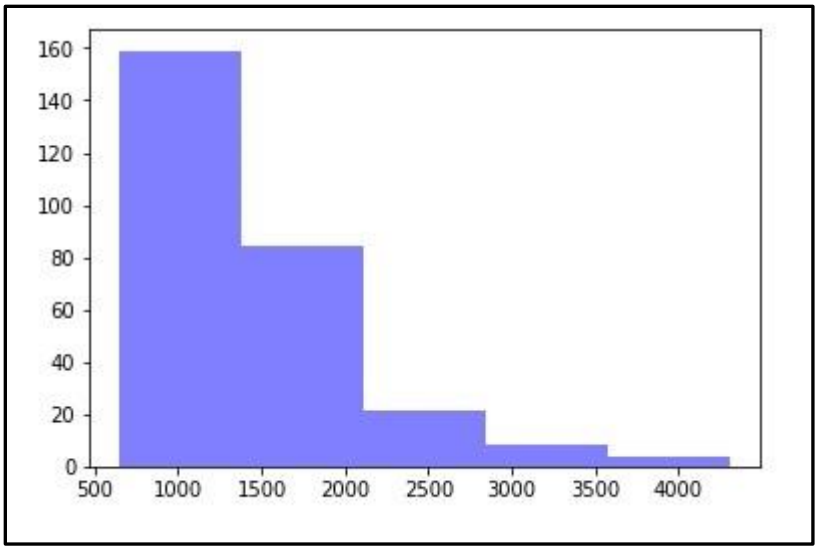

Figure 2: Price per Square foot distribution

ANALYSIS

As shown in the location map, it is difficult to analyze over 300 properties with price/square foot ranging from $500 to $4,000. Therefore, data science methodology is used to build a price model.
Before building the model lets study price correlation with each parameters: distance from central, having low lease year (newer condominium), and accessibility to shopping malls & restaurants.

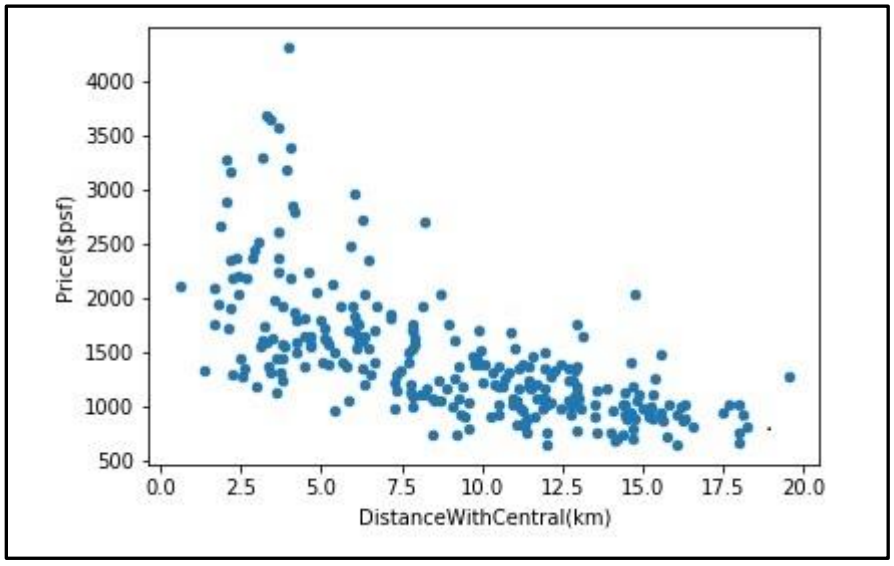

Figure 3: Distance from Central Singapore vs. Price

There is a strong negative correlation between the price and the distance from central area. Let's look at having Years from when the lease started contribution to the price.

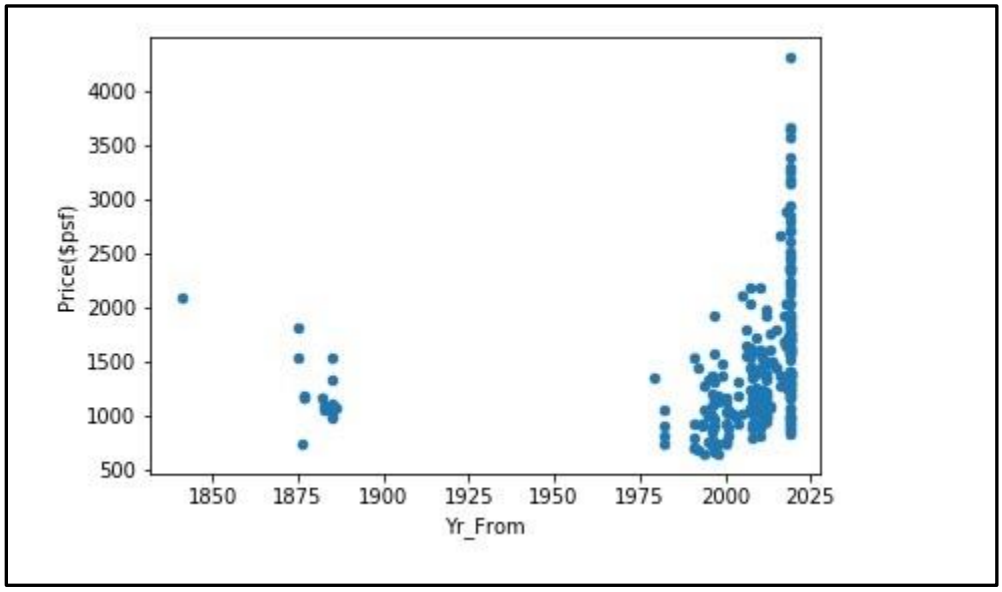

Figure 4: Years from when the lease started Vs. Price

From this data, there is a weak correlation that later the lease of a property started, more will be the price. Let investigate the last parameter, accessibility to restaurants and shopping malls.

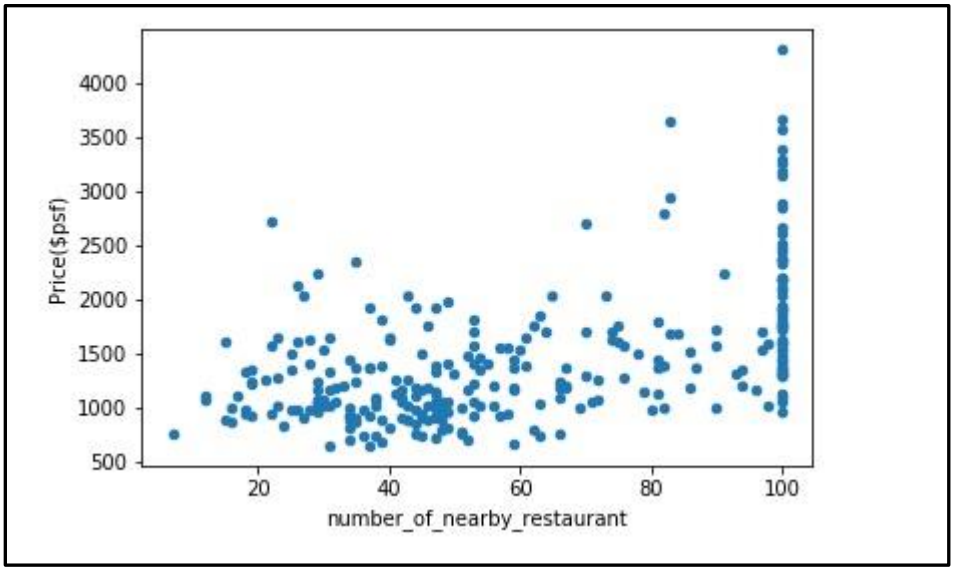

Figure 5: No. of nearby restaurants/shopping locations vs. Price

There is medium level 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:

- Coefficients: [[-277.30641897 -35.88619065 32.72750304 114.03858073 91.32257268]]
- Absolute Mean Error is 21.40%

RESULTS AND DISCUSSION

Model coefficients of DistanceWithCentral(km), low_lease_year, Yr_From, freehold, number_of_nearby_restaurant are -277.30641897, -35.88619065, 32.72750304, 114.03858073, 91.32257268 respectively.

According to the coefficients, Distance from Central (-277.30641897) is the highest negative contributor and Freehold type (114.03858073) (which means that the property is not on lease by owned by an individual or party) is the highest positive contributor with number of nearby restaurants NOT far behind (91.32257268). However, the distance from central parameter contributes more than double than the freehold type and almost thrice as the nearby restaurants parameter in magnitude. Lease hold type (-35.88619065) (meaning it is a lease property) negatively affects the price while the number of remaining lease years (32.72750304) positively affect the price i.e. the latest the lease start year is; pricier the property and vice versa.

With the model, top twenties undervalued properties will be shortlisted for further investigations. The formula (y_hat - y)/y gives the depiction of under-valued and over-valued properties for example a positive value will be an overvalued property.

Table 3: Top 20 Under-valued properties

Project Name	Tenure	Price($psf)	Longitude	Latitude	DistanceWith Central(km)	low_ lease_ye	freeh old	Yr_F rom	number_of _nearby_re staurant	PriceDiffe rentPerce nt
KANDIS RESIDENCE	99 Yrs FROM 2016	1273	103.838741	1.459187	19.531974	1	0	2016	23	-0.543079
MAYFAIR MODERN	99 Yrs FROM 2018	2029.5	103.734722	1.346672	14.761231	1	0	2018	43	-0.535187
LE NOUVEL ARDMORE	Freehold	4306.25	103.829215	1.311597	4.005001	0	1	2019	100	-0.510599
BISHOPSGATE RESIDENCES	Freehold	3651	103.825232	1.299448	3.424045	0	1	2019	83	-0.428092
TWENTYONE ANGULLIA PARK	Freehold	3674	103.83083	1.304666	3.305964	0	1	2019	100	-0.414631
3 ORCHARD BY-THE-PARK	Freehold	3580	103.826118	1.304599	3.682151	0	1	2019	100	-0.405749
LAKEVILLE	99 Yrs FROM 2013	1484	103.72664	1.346977	15.572989	1	0	2013	52	-0.381794
NOUVEL 18	Freehold	3392	103.829465	1.31241	4.059658	0	1	2019	100	-0.379685
MIDWOOD	99 Yrs FROM 2018	1655	103.765443	1.364632	13.144406	1	0	2018	40	-0.375586
LEEDON GREEN	Freehold	2726.756757	103.804507	1.314915	6.290768	0	1	2019	22	-0.371949
ROYALGREEN	Freehold	2706.25	103.795918	1.332199	8.220992	0	1	2019	70	-0.353514
VAN HOLLAND	Freehold	2953.357143	103.80388	1.309969	6.061709	0	1	2019	83	-0.348139
GRAMERCY PARK	Freehold	3300.5	103.827918	1.299761	3.194765	0	1	2019	100	-0.346308
THE JOVELL	99 Yrs FROM 2018	1256.333333	103.966916	1.359462	15.396096	1	0	2018	21	-0.337317
ARDMORE PARK	Freehold	3189	103.830531	1.311381	3.897525	0	1	2019	100	-0.337059
HAUS ON HANDY	99 Yrs FROM 2018	2887	103.847845	1.301161	2.044613	1	0	2018	100	-0.337033
THE AVENIR	Freehold	3267.222222	103.838101	1.295892	2.04695	0	1	2019	100	-0.317958
DUCHESS RESIDENCES	999 Yrs FROM 1875	1818	103.802963	1.325245	7.124541	1	0	1875	39	-0.308896
COCO PALMS	99 Yrs FROM 2008	1404	103.947399	1.373163	14.621129	1	0	2008	53	-0.306012
MAYFAIR GARDENS	99 Yrs FROM 2018	2030.333333	103.792571	1.335304	8.727387	1	0	2018	73	-0.303804

Table 4: Top 20 over-valued properties

	Project Name	Tenure	Price($psf)	Longitude	Latitude	DistanceWithCentral(km)	low_lease_year	freehold	Yr_From	number_of_nearby_rest	PriceDifferentPercent
Project Name											
RIVER PLACE	RIVER PLACE	99 Yrs FROM 1995	1340	103.842234	1.291098	1.348985	1	0	1995	100	0.442948
KING'S MANSION	KING'S MANSION	Freehold	1372	103.899716	1.302121	5.788309	0	1	2019	100	0.455813
WATER PLACE	WATER PLACE	99 Yrs FROM 1998	1182.5	103.873747	1.298078	3.003435	1	0	1998	66	0.45777
THE STELLAR	THE STELLAR	Freehold	1080	103.768856	1.294583	9.264252	0	1	2019	38	0.463899
THE AMERY	THE AMERY	Freehold	1295	103.909789	1.31335	7.323512	0	1	2019	100	0.469178
OASIS GARDEN	OASIS GARDEN	Freehold	1214	103.884226	1.340153	7.305193	0	1	2019	67	0.479684
BEDOK COURT	BEDOK COURT	99 Yrs FROM 1982	752	103.945494	1.323451	11.39731	1	0	1982	45	0.490611
GROSVENOR VIEW	GROSVENOR VIEW	Freehold	1065	103.913006	1.328804	8.533013	0	1	2019	32	0.508579
CLYDESVIEW	CLYDESVIEW	Freehold	1245	103.819804	1.295596	3.771811	0	1	2019	29	0.518755
ISLAND VIEW	ISLAND VIEW	Freehold	1196	103.794194	1.281075	6.359018	0	1	2019	59	0.529034
SIMSVILLE	SIMSVILLE	99 Yrs FROM 1994	1050	103.890125	1.318296	5.822595	1	0	1994	100	0.577443
KERRISDALE	KERRISDALE	99 Yrs FROM 1998	1137.75	103.859324	1.314626	3.609284	1	0	1998	100	0.579469
KEW GREEN	KEW GREEN	99 Yrs FROM 1994	651	103.952732	1.319786	11.99897	1	0	1994	31	0.613575
WEST BAY CONDOMINIUM	WEST BAY CONDOMINIUM	99 Yrs FROM 1991	801	103.766799	1.300242	9.595547	1	0	1991	62	0.618797
PALM HAVEN	PALM HAVEN	999 Yrs FROM 1876	743	103.879151	1.361632	9.230711	1	0	1876	63	0.622471
NIM GARDENS	NIM GARDENS	Freehold	836	103.864931	1.382051	11.057114	0	1	2019	24	0.704255
SOPHIA RESIDENCE	SOPHIA RESIDENCE	Freehold	1298	103.849827	1.303108	2.229493	0	1	2019	100	0.708097
CENTRAL GROVE	CENTRAL GROVE	99 Yrs FROM 1997	968	103.885773	1.317336	5.397597	1	0	1997	100	0.741329
CENTRAL VIEW	CENTRAL VIEW	99 Yrs FROM 1998	740	103.879727	1.354018	8.466156	1	0	1998	36	0.741763
CASABELLA	CASABELLA	Freehold	991	103.801019	1.32538	7.298846	0	1	2019	48	0.750658

GEOGRAPHICAL VIEW OF OVER-VALUED AND UNDER-VALUED PROPERTIES

There are a couple of undervalued properties at the east end of Singapore. 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. This information can be further analyzed.

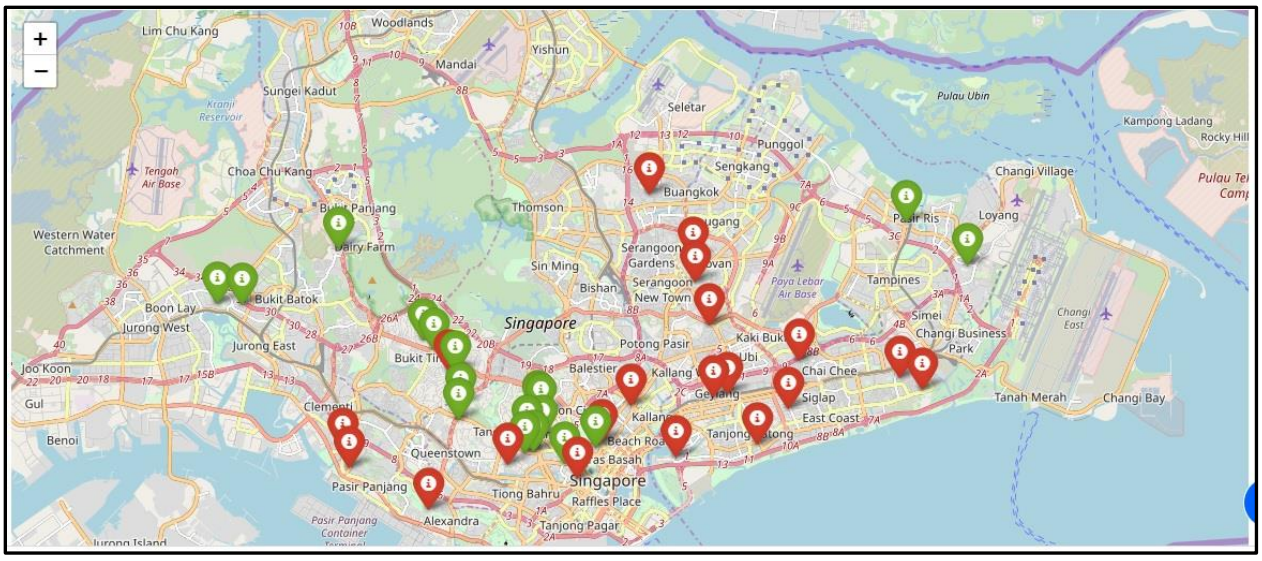

Figure 6: Under-Valued (Green) and Over-Valued (Red) Properties Map

References:

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

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