Property Price Analysis-Brisbane Southern Suburbs 2019

Frank He

09 September 2019

1. Introduction

1.1 Background

My friend JV has been promoted to a manager role recently, while his income grows and the Bank interest rate is at historical low, he is looking for an investment property. As his best friend, I'm trying to help him with my data science skills.

We are targeting at Southern Suburbs in Brisbane, QLD, Australia, not only because the rapid population growth in this area, the large investments in infrastructures and potential bid for 2032 Olympic Games, but also JV wants to have it close to where he lives so he could manage the property and tenants himself.

1.2 Problem

Of all the Southern suburbs in Brisbane, we would like to find out which ones are most undervalued and worth investing.

1.3 Interest

JV, his family and I are interested to see if data science can help with making investment decisions.

2. Data acquisition and cleaning

2.1 Data Sources

Data were sourced from below:

Brisbane Suburb List: https://www.brisbane.qld.gov.au/about-council/council-information-and-rates/brisbane-suburbs

Brisbane Region Classification: https://en.wikipedia.org/wiki/List_of_Brisbane_suburbs

Geospatial Data: http://www.corra.com.au/australian-postcode-location-data/

Brisbane House Information by Suburbs: https://homesales.com.au/location/brisbane-qld/

Venues Information : https://api.foursquare.com/v2/venues/

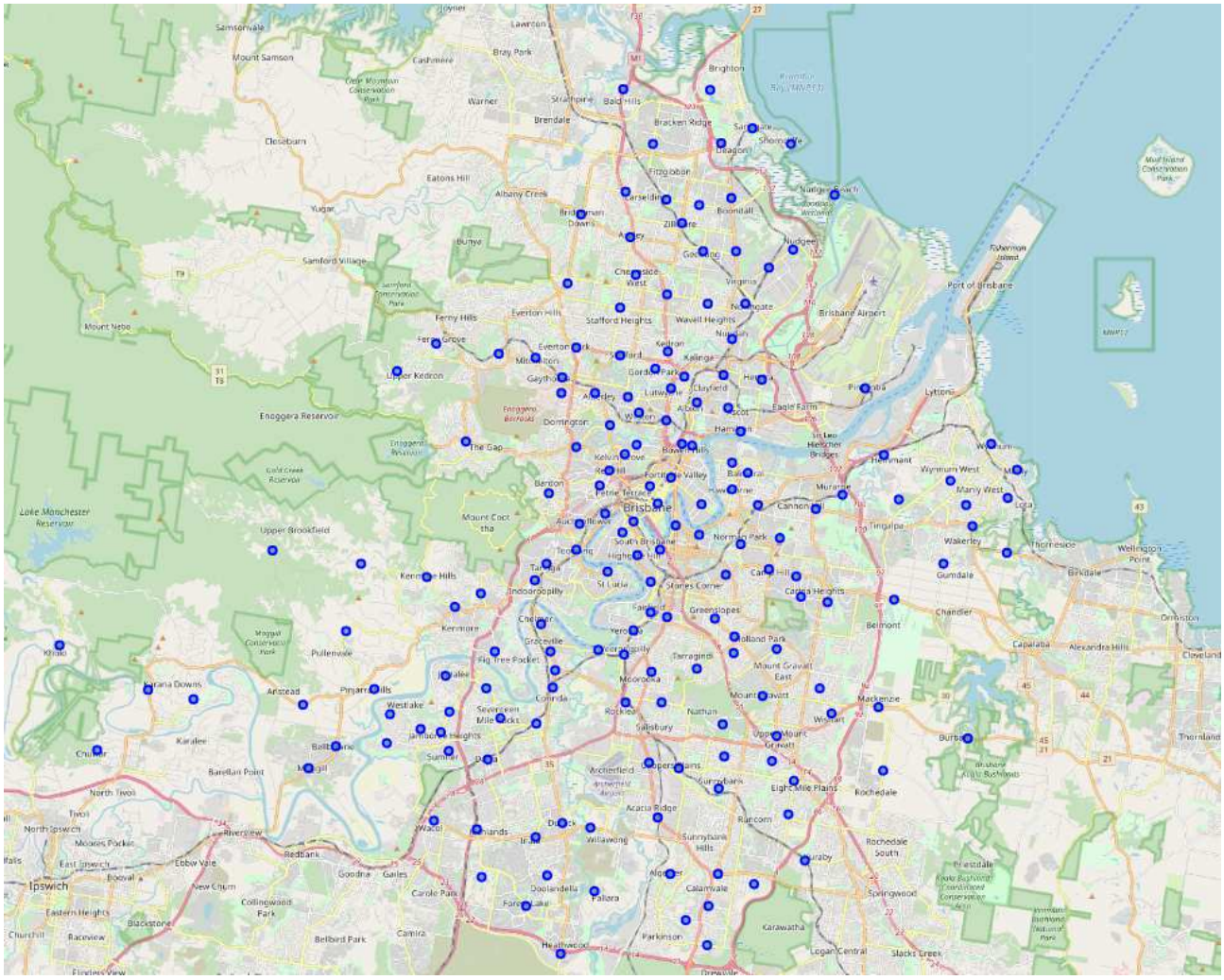
We will collect Brisbane Suburb list, Regions Classifications, Geospatial data, House information such as Median House Price 2019, Population and income, and also venues information from foursquare API.

2.2 Data cleaning

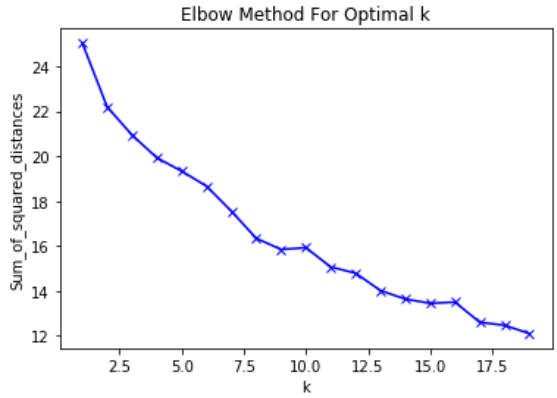
Data download and scraped using Beautiful Soup were combined into one table. The overall data quality is very good, as they are from the official sources. We have excluded 3 suburbs what has population less than 51, they are the industrial suburbs. Left 175 suburbs in total for analysis.

3. Exploratory Data Analysis

3.1 Total 175 Brisbane Suburbs are included in this analysis.

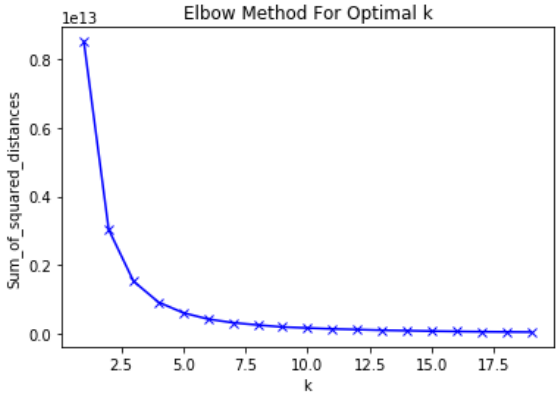


3.2 Next, unsupervised Machine Learning Algorithms: K-means clustering has been used to cluster the suburbs by venues. (Using Elbow Graph to find the best K).



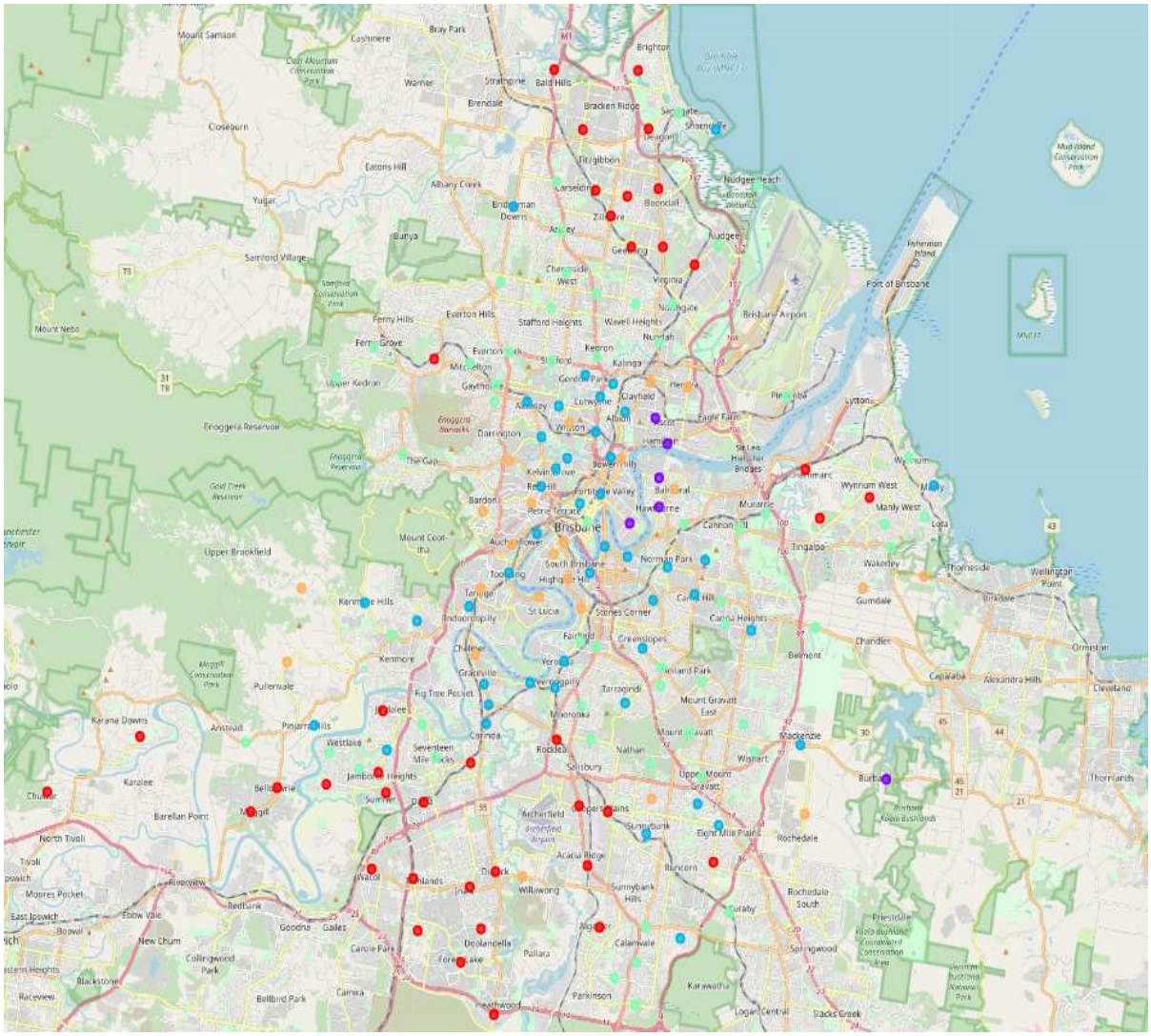
The algorithm is not working too well as sum of error is still trending downwards even when K=20.

let's add in pop, income and house price.

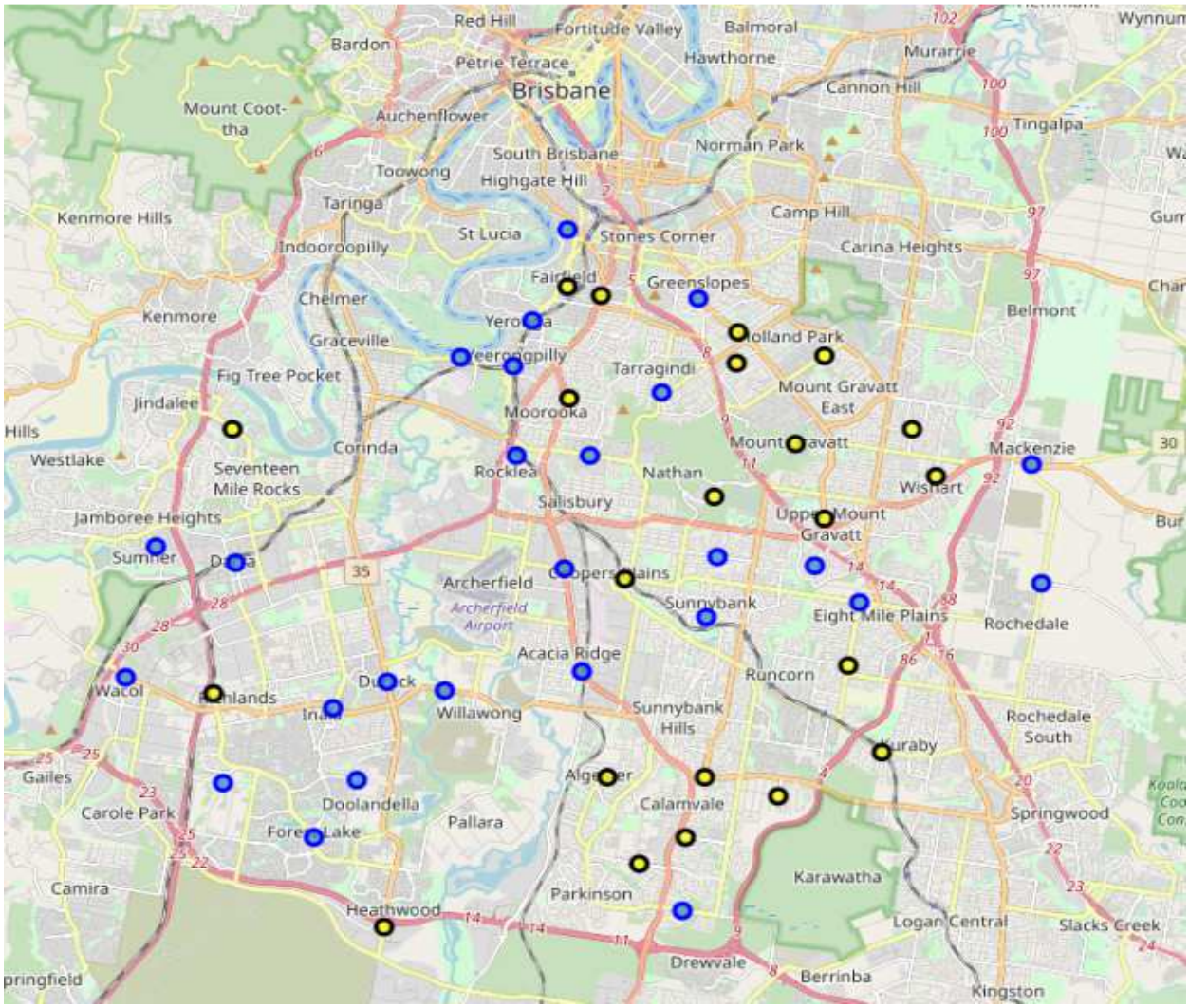


Now the Elbow Graph shows much clear results, K=5 seems to be the reasonable choice.

The 175 suburbs are clustered into 5 as shown below:


Mount Samson
Bray Park
Cashmere
Brighton
Strathpine
Bald Hills
Brendale
Bracken Ridge
Closeburn
Eatons Hill
Albany Creek
Tingalpa
Bridgeman Downs
Zillmere
Boondall
Nudgee
Nudgee Beach
Geebung
Virginia
Fort of Brisbane
Samford Village
Burpa
Everton Hills
Stafford Heights
Wavell Heights
Nundah
Brisbane Airport
Mount Nebo
Ferny Hills
Ferny Grove
Keperra
Kalinga
Gaythorne
Clayfield
Albion
Eagle Farm
Enoggera Reservoir
Darrington
Wilston
Hamilton
The Gap
Kelvin Grove
Bowen Hills
Bardon
Petrie Terrace
Fortitude Valley
Brisbane
Hawthorne
Murarrie
Cannon Hill
Lytton
Wynnum
Wynnum West
Manly West
Tingalpa
Wakerley
Thorneside
Wellington Point
Upper Brookfield
Mount Coot-tha
Auchenflower
South Brisbane
Norman Park
Toowong
Highgate Hill
Kenmore Hills
Indooroopilly
St Lucia
Stones Corner
Camp Hill
Gumdale
Chandler
Belmont
Capalaba
Alexandra Hills
Cleveland
Ormiston
Fairfield
Greenslopes
Holland Park
Kenmore
Chelmer
Graceville
Yeronga
Tarragindi
Pullenvale
Fig Tree Pocket
Mount Gravatt East
Karana Downs
Anstead
Pinjarra Hills
Westlake
Corinda
Rocklea
Nathan
Mackenzie
Wishart
Seventeen Mile Rocks
Salisbury
Upper Mount Gravatt
Karalee
Jamboree Heights
Sumner
Archerfield
Eight Mile Plains
Rochedale
Belivah
Barellan Point
North Tivoli
Acacia Ridge
Runcorn
Tivoli
Moores Pocket
Redbank
Willawong
Sunnybank Hills
Rochedale South
East Ipswich
Ebbw Vale
Goodna
Gailes
Algester
Priestdale
New Chum
Carole Park
Doolandella
Pallara
Calamvale
Springwood
Collingwood Park
Forestlake
Blackstone
Raceview
Bellbird Park
Camira
Parkinson
Karawatha
Logan Central
Slacks Creek

3.3 We use Multiple Linear Regression model to predict Southern Brisbane Suburbs within the 5 clusters. Trained with Inner, Northern, Eastern and Western data. The undervalued outcome for Southern suburbs showing as below, we have plotted the undervalued with blue and the overvalued ones with yellow.



Each cluster will have different Intercept and Coefficients for its own multiple linear regression model. For example, Clusters showing below:

```
X1 = train1[['Population','Avg Income']]
yT1 = train1[['Median House Price']]

# with sklearn
regr = linear_model.LinearRegression()
regr.fit(X1, yT1)

print('Intercept: \n', regr.intercept_)
print('Coefficients: \n', regr.coef_)
```

```
Intercept:
 [289877.48466496]
Coefficients:
 [[  3.7661257  246.79097348]]
```

```
X2 = train2[['Population','Avg Income']]
yT2 = train2[['Median House Price']]

# with sklearn
regr2 = linear_model.LinearRegression()
regr2.fit(X2, yT2)

print('Intercept: \n', regr2.intercept_)
print('Coefficients: \n', regr2.coef_)
```

```
Intercept:
 [2481108.16780428]
Coefficients:
 [[   20.37662016 -1143.72083074]]
```

```
X3 = train3[['Population','Avg Income']]
yT3 = train3[['Median House Price']]

# with sklearn
regr3 = linear_model.LinearRegression()
regr3.fit(X3, yT3)

print('Intercept: \n', regr3.intercept_)
print('Coefficients: \n', regr3.coef_)
```

```
Intercept:
 [674755.93498253]
Coefficients:
 [[  1.25150843 178.42155356]]
```

```
X4 = train4[['Population','Avg Income']]
yT4 = train4[['Median House Price']]

# with sklearn
regr4 = linear_model.LinearRegression()
regr4.fit(X4, yT4)

print('Intercept: \n', regr4.intercept_)
print('Coefficients: \n', regr4.coef_)
```

```
Intercept:
 [326035.1216507]
Coefficients:
 [[-2.78387375e-01  3.68834627e+02]]
```

```
X5 = train5[['Population','Avg Income']]
yT5 = train5[['Median House Price']]

# with sklearn
regr5 = linear_model.LinearRegression()
regr5.fit(X5, yT5)

print('Intercept: \n', regr5.intercept_)
print('Coefficients: \n', regr5.coef_)
```

```
Intercept:
 [994865.18154198]
Coefficients:
 [[-3.48677542 87.21801653]]
```

3.4 Detailed prediction for all Southern Brisbane Suburbs

Undervalued Suburbs:

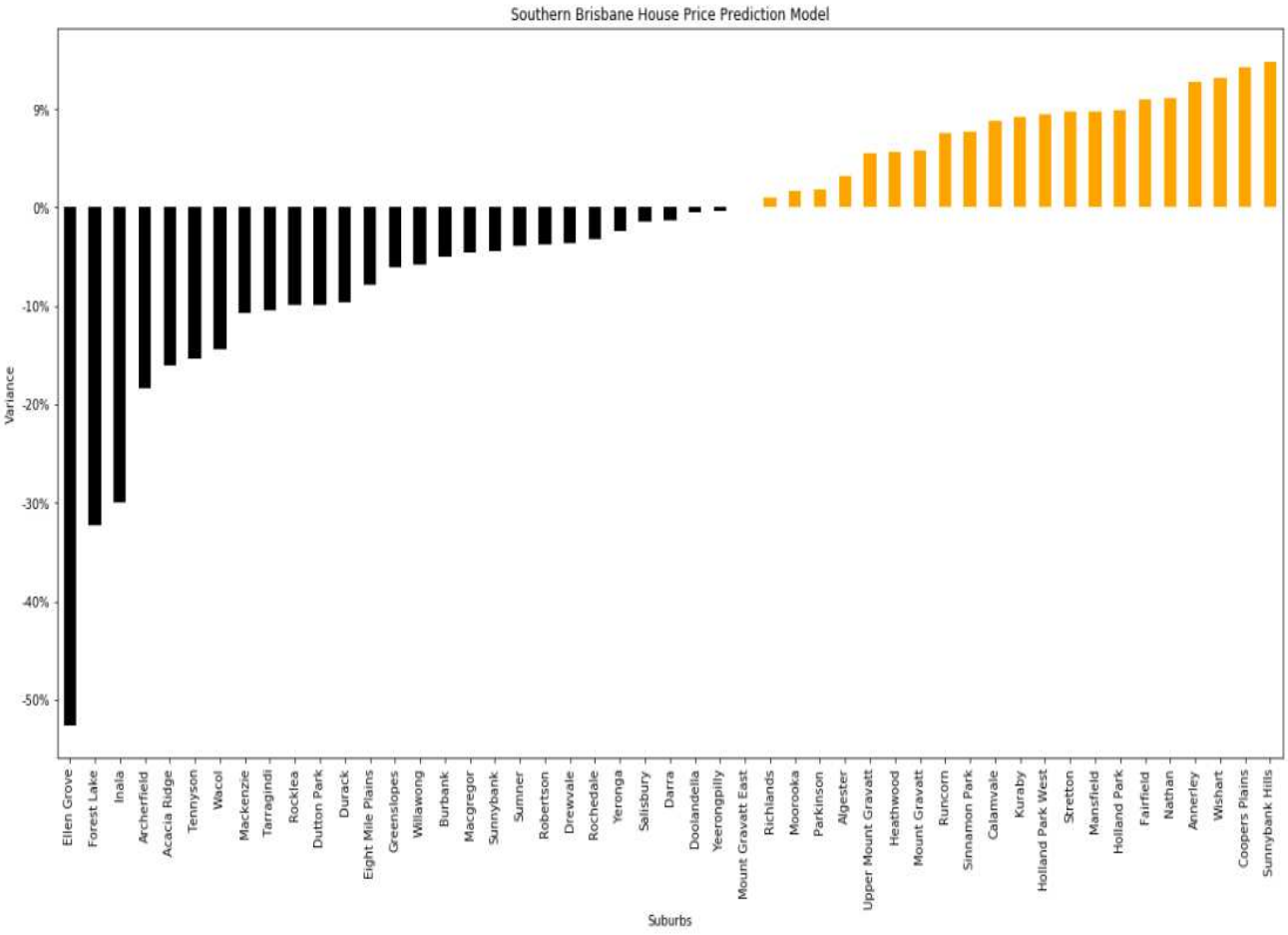
Suburbs	Median House Price	Predicted	Variance	lat	lon	Population	Avg Income
Ellen Grove	287993	439,325	-52.55%	-27.6127	152.951	2527	567
Forest Lake	441430	583,862	-32.27%	-27.624	152.97	22426	849
Inala	358343	465,473	-29.90%	-27.5973	152.974	13795	501
Archerfield	389395	460,850	-18.35%	-27.5684	153.024	510	685
Acacia Ridge	403843	468,793	-16.08%	-27.5896	153.028	6944	619
Tennyson	773504	892,613	-15.40%	-27.5245	153.002	859	1215
Wacol	399069	456,739	-14.45%	-27.5908	152.93	2957	631
Mackenzie	767975	850,134	-10.70%	-27.5468	153.125	1845	970
Tarragindi	776961	858,511	-10.50%	-27.5317	153.045	9964	960
Rocklea	423229	465,137	-9.90%	-27.5447	153.014	1255	691
Dutton Park	961298	1,055,847	-9.84%	-27.498	153.025	1471	758
Durack	430047	471,581	-9.66%	-27.5917	152.986	6177	642
Eight Mile Plains	776984	838,340	-7.90%	-27.5752	153.088	13378	823
Greenslopes	802293	850,514	-6.01%	-27.5121	153.053	8564	925
Willawong	994213	1,051,320	-5.74%	-27.5934	152.998	193	655
Burbank	1333318	1,399,442	-4.96%	-27.5589	153.164	1137	966
Macgregor	769838	804,845	-4.55%	-27.5676	153.078	5576	690
Sunnybank	773968	807,992	-4.40%	-27.5781	153.055	8090	690
Sumner	492134	511,062	-3.85%	-27.5637	152.936	540	888
Robertson	1003527	1,040,692	-3.70%	-27.5657	153.057	4867	720
Drewvale	612356	634,390	-3.60%	-27.6391	153.05	3943	839
Rochedale	1029620	1,062,493	-3.19%	-27.5712	153.127	1091	819
Yeronga	833031	852,796	-2.37%	-27.5168	153.017	5540	959
Salisbury	605137	614,242	-1.50%	-27.5448	153.03	6096	786
Darra	459754	465,486	-1.25%	-27.567	152.953	3838	653
Doolandella	483738	486,167	-0.50%	-27.612	152.979	3104	748
Yeerongpilly	836792	839,424	-0.31%	-27.5263	153.013	1984	909

Overvalued Suburbs:

Suburbs	Median House Price	Predicted	Variance	lat	lon	Population	Avg Income
Sunnybank Hills	696700	593,181	14.86%	-27.6114	153.054	16830	737
Coopers Plains	558606	479,216	14.21%	-27.5703	153.037	4207	703
Wishart	728594	632,575	13.18%	-27.549	153.104	10460	839
Annerley	743990	649,116	12.75%	-27.5116	153.032	10664	884
Nathan	625304	556,168	11.06%	-27.5532	153.056	1396	625
Fairfield	728395	648,424	10.98%	-27.5099	153.025	2553	876
Holland Park	733388	660,965	9.88%	-27.5193	153.061	7848	914
Mansfield	680301	613,949	9.75%	-27.5394	153.099	8473	787
Stretton	917266	828,471	9.68%	-27.6154	153.07	4067	833
Holland Park West	741835	671,079	9.54%	-27.5255	153.061	5965	940
Kuraby	692961	629,634	9.14%	-27.6064	153.092	7776	829
Calamvale	682224	622,379	8.77%	-27.6239	153.05	15291	815
Sinnamon Park	737448	680,559	7.71%	-27.5392	152.953	6361	966
Runcorn	574068	530,937	7.51%	-27.5883	153.085	14074	762
Mount Gravatt	666457	627,947	5.78%	-27.5423	153.074	3237	821
Heathwood	576073	543,276	5.69%	-27.6424	152.985	1820	999
Upper Mount Gravatt	635271	600,566	5.46%	-27.5578	153.08	8852	751
Algester	531232	513,984	3.25%	-27.6113	153.034	8262	782
Parkinson	664635	652,012	1.90%	-27.6295	153.04	9538	891
Moorooka	646356	635,659	1.66%	-27.5328	153.025	9984	847
Richlands	479524	474,896	0.97%	-27.594	152.948	2077	718
Mount Gravatt East	639905	639,094	0.13%	-27.524	153.08	10891	857

4. Observations

The histogram on the right also shows that Ellen Grove would be the most undervalued suburbs in Southern Brisbane, with median market price of $280K and predicted price of $440K, undervalued by 52%, while Sunnybank Hills is the most overvalued Suburb, overvalued by 15%.



5. Conclusions

In this study, we have used the Unsupervised Machine Learning Algorithms: K-means clustering and Supervised Machine Learning Algorithms: Multiple linear regression to predict whether a suburb in Southern Brisbane is undervalued.

This analysis might have some limitations as there are many other factors other than venues, population, income that have impacts on House Prices. A more sophisticated model can be built incorporating data such as crime rates, employment rates, school rankings, etc...

However, my friend JV is happy enough with the initial analysis and planned his weekend house inspections around Ellen Grove and Forest Lake, let's wish him good luck.