

Title: Predicting “high property growth areas” in Sydney metropolitan area (MPA)

Introduction

Needless to say real estate properties are solid investments. Many investors both individuals and institutions resort to property investment because of their security and long term solid growth potential. In Australia historically house prices has shown a continued capital gain about 10% per annum for last 70 years. Recently the capital gain has dropped substantially but it is still most of the other investment vehicles. When it comes to buying property, all goes down to hence, the best strategy to buy in a good location. What is good location in terms of property price growth? Population growth is the main engine that drives capital growth of property in an area but other factors count as well. In this project an attempt is made to find a better way to locate a good area for property investment.

Background

All over the world residential and real estate properties are popular engines of investment due to their long terms growth potential and high security. Australia is not an exception, so property investment had been an important activity for fund manager, superannuation managers, and property trust managers and as well as individuals when it comes to retirement plans. This report is about the vehicles of property growth mainly location and how to choose them wisely from thousands of viable locations. Though growth of property in an area depends on many factors but it mainly turned on by increase in population in an area and presence of different amenities or venues.



Figure: Growth of Australian house price 1926-2016 (source: <https://www.livewiremarkets.com>)

Problem: Given a choice of different areas how to choose the best performing area in terms of property growth. This is a report about finding Local Government Areas (LGA) Sydney metropolitan area (MPA) with high potential of growth in terms of property price. All experts would agree on one single point: “population growth being the most important vehicle of consistent and permanent house price growth”. All other factors such as employment rate, communication network different venues are also very important factors but they are secondary to population growth. This report is based on the analysis of two

factors (ii) population growth and (ii) “foresquare api” venues. This projects aim to predict the house growth from 5 LGAs and select the best ones.

Validity of Hypothesis: Why house price increase with the increase of population? Or is it a valid statement at all? Besides population growth house price in general depend on many factors such as employment opportunity, communication & infrastructure, weather, presence of school, restaurant, hospitals, park etc. But on a critical analysis it can be logically deduced that people tend to go for living into those areas where such factors are already preferable. So if a place has high population growth then it provides strong evidence that that area has all those factors.

Business Value: This report has a great business value to those who invest in properties. The notebook template accompanying this report can serve as a template that can be customized further by interested parties and used for predicting long terms property price of any place of the world.

Data Acquisition

Population projection data for this project has been obtained from this source:
<https://www.planning.nsw.gov.au/-/media/Files/DPE/Other/Research-and-demography/Population-Projections/2019-NSW-Population-Projections-ASGS-2019-LGA.xlsx>

This database has population projections (2016-2041) data of different LGAS of NSW Australia. Here is a sample data

	SydneySuburb	2016	2021	2026	2031	2036	2041
4	ASGS 2019 LGA	2016.0	2021.0	2026.0	2031.0	2036.0	2041.0
5	Albury (C)	52171.0	54374.0	56490.0	58159.0	59440.0	60383.0
6	Armidale Regional (A)	30313.0	32736.0	35226.0	36486.0	37439.0	38124.0
7	Ballina (A)	42993.0	44237.0	45364.0	46248.0	46835.0	47092.0
8	Balranald (A)	2330.0	2437.0	2552.0	2764.0	2975.0	3185.0

- 1) We need location of the LGA. Location data of Sydney MPA was obtained from this source:
<https://www.latlong.net/category/towns-14-55.html>
<https://www.latlong.net/category/towns-14-55-2.html>
<https://www.latlong.net/category/cities-14-15.html>

Sample Data:

	SydneySuburb	Latitude	Longitude
0	Blacktown	-41.161388	146.234451
1	Liverpool	-35.708057	150.174438
2	Parramatta	-28.643057	153.615005
3	Sydney	-33.861111	121.891945
4	Camden	-31.948999	115.827003
0	Blacktown	-33.942173	151.101456
1	Liverpool	-33.751064	150.852631
2	Parramatta	-30.090109	145.937546
3	Sydney	-29.464411	149.845108
4	Camden	-28.085995	153.422348

Methodology

5 LGAs of Sydney MPA was chosen for this report namely: Parramatta, Black town, Liverpool, Camden and Sydney. Attempt has been to find the best 3 of the most potential LGAs in terms of property investment.

The following steps are followed to that aim:

- 1) Got population projection data from <https://www.planning.nsw.gov.au/> in its 2019-NSW-Population-Projections-ASGS-2019-LGA.xlsx". Data was injected into a data frame for analysis.

```
data = pd.read_excel(link, sheet_name='LGA population projections')
```

Once in a data frame initial analysis was done on that such as the size of the data base, what information it contains etc. Analysis reveals that, there are population data of 131 local government areas (LGA) of NSW. LGA in Australia is third tier in the administrative structure (Country, State, LGA) and can be a city, council, or a shire. An LGA is quite large in area and as well as population.

Fig : Image showing number of LGAs in different States of Australia

(source:https://en.wikipedia.org/wiki/Local_government_in_Australia#Local_governments_by_type_and_state)

Local government area types	NSW	Vic	Qld	WA	SA	Tas	NT	Total
Boroughs		1						1
Cities	28	33	7	29	21	6	2	126
Councils	28				15			43
District councils					25			25
Municipalities	6					23	3	32
Regional councils	8				4			12
Rural cities		6			1			7
Shires	58	39	27	104			3	231
Towns			1	8	2			11
Aboriginal councils					5			5
Aboriginal shires			12					12
Region			31				9	40
Sub-total	128	79	78	141	73	29	17	545
Unincorporated	2	10			1		5	18
Total	130	89	78	141	74	29	22	563

Our data frame has 131 rows which matches with the number of LGAs in NSW

```
1]: df_SydSub.shape
```

```
1]: (131, 7)
```

Selected areas of interest for this project are: Parramatta (84 SQ KM), Blacktown (246 sq KM) , Liverpool (305sqKM), Camden (201SQKM) and Sydney (25 SQKM) are all LGAs. They are filtered out from the population projection data.

```
[6]: df_popTrend = pd.merge(df_SydSub, df_selected_Subrub, on='SydneySuburb',how='inner')
df_popTrend.head(10)
```

```
[6]:
```

	SydneySuburb	2016	2021	2026	2031	2036	2041
0	Blacktown	164534.0	198945.0	220879.0	229397.0	228720.0	234613.0
1	Camden	18114.0	17272.0	16459.0	15580.0	14635.0	13641.0
2	Liverpool	37694.0	43768.0	48429.0	48497.0	48218.0	49362.0
3	Parramatta	265468.0	269596.0	275246.0	281802.0	288431.0	296625.0
4	Sydney	20558.0	20188.0	19746.0	19169.0	18453.0	17619.0

- 2) Arranging Data: In order to filter the areas in accordance to where the population is projected to be of highest increase, the following analysis was performed:

```
[7]: #get population increase prediction % increase from 2016-2041
df_popTrend['%Increase/Decrease'] = round((((df_popTrend['2041']-df_popTrend['2016'])/df_popTrend['2016'])*100),2)

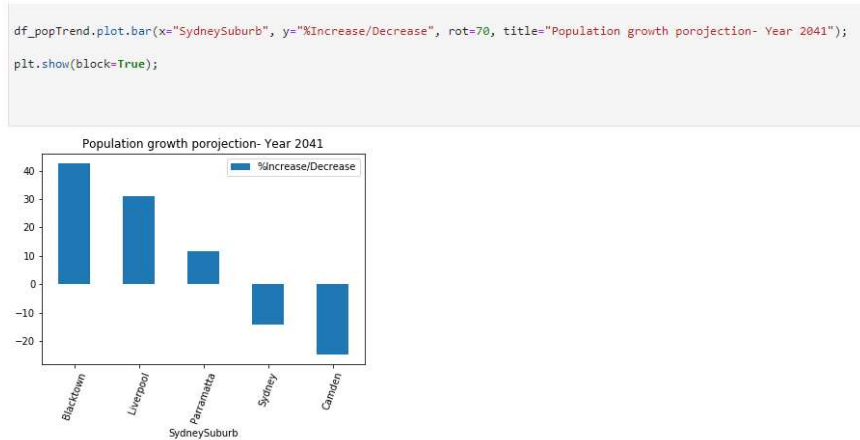
#arrange them in descending order
df_popTrend = df_popTrend.sort_values("%Increase/Decrease", ascending = False)

[8]: df_popTrend.head(10)
```

	SydneySuburb	2016	2021	2026	2031	2036	2041	%Increase/Decrease
0	Blacktown	164534.0	198945.0	220879.0	229397.0	228720.0	234613.0	42.59
2	Liverpool	37694.0	43768.0	48429.0	48497.0	48218.0	49362.0	30.95
3	Parramatta	265468.0	269596.0	275246.0	281802.0	288431.0	296625.0	11.74
4	Sydney	20558.0	20188.0	19746.0	19169.0	18453.0	17619.0	-14.30
1	Camden	18114.0	17272.0	16459.0	15580.0	14635.0	13641.0	-24.69

Fig: Projected population growth of selected LGAs of NSW 2016-2041

A bar chart was plotted for a better visualization about the population projection



3) **Getting location data:** In order to analyze the property investment potential of an area besides population there are other important factors hence analyze of an area in terms of certain criteria:

- Communication network
- Employment opportunity
- Educational Facilities
- Common amenities
- Recreational facilities such as park etc.

In order to collect those information, venue data was collected by using foresquare api. Venue around 5 KM from the central point of the LGA was collected and arranged as a data frame along with other important data.

7]:

	SydneySuburb	%Increase/Decrease	Latitude	Longitude	Arts & Entertainment	College & University	Event	Food	Nightlife Spot	Outdoors & Recreation	Professional & Other Places	Residence	Shop & Service	Travel & Transport
0	Blacktown	42.59	-33.7668	150.9053	5	4	0	39	4	8	43	4	51	9
1	Liverpool	30.95	-33.9286	150.9180	5	2	0	13	6	6	9	3	16	5
2	Parramatta	11.74	-33.8136	151.0034	5	7	2	86	13	20	50	5	67	21
3	Sydney	-14.30	-33.8688	151.2093	42	62	7	146	120	99	136	24	137	133
4	Camden	-24.69	-34.0544	150.6958	5	2	0	11	5	5	7	1	7	0

4) **Analysis of foresquare venue data:** The main focus of this project is on the theme that property price grows mainly with the growth of population, which is true but other factors are also responsible. Such factors are communication network, opportunities for employment, college/universities, shop and services. But there is no data to show the exact relationship between different factors the growth of house prices. Hence in the absence of such data, each important contributing factor was multiplied by a number with reasonable guess. The more

important the factor the higher the number. The following table depict as how the different factors had been multiplied

Table: Multiple (weight) with which different factors are multiplied

Factor	Multiple
Population growth (%)	10
Travel & Transport	5
College & University	4
Shop & Service	3
Outdoors & Recreation	2
Professional & Other Places	2

All other venue types are given a weightage of 1

Side Note: The weight values for different venue type are calculated guess. However, in order to make that more realistic; research is required to find out the relationship between growths of house price for different factors. For instance how house price would increase or decrease with the addition or deletion of a communication network, services etc. These sorts of research can be based upon proper regression analysis. Such analysis would be worthwhile for the topic of this project however for the purpose of this assignment this seems to be too much and out of scope.

After multiplying the weightage factor with the population growth and all the venues and summing them up on a “Total Points” column we get the following result.

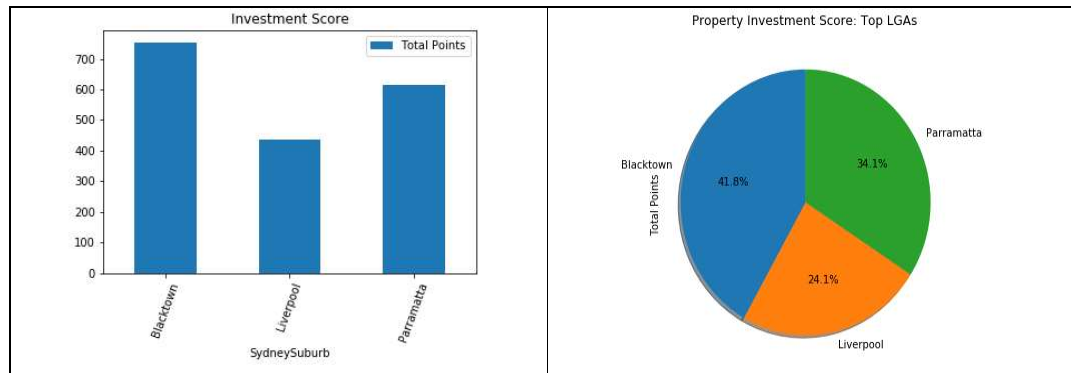
[84]:

	SydneySuburb	Total Points
0	Blacktown	754.9
1	Liverpool	434.5
2	Parramatta	616.4

[Please note: Only top 3 LGA in terms of population growth has been taken into account]

- Data Visualization:** Once the total points has been obtained, the data has been shown as graph and pie chart to depict which LGAs are winners in terms of potentials for future property price growth.

Fig: Bar chart and Pie chart showing the potential of 3 top LGAs for property investment



Result & Recommendations

The result is quite evident from the data visualization graphs. The most potential LGA of Sydney MPA is Black Town, followed by Parramatta and lastly Liverpool

```
]: int ('Conclusion: from amongst the 5 selected LGAs Black Town LGA seems to be the best choice for property investment followed by Parramatta')
```

Anybody interested to buy investment property in Sydney MPA can choose property from any of the three LGAs. Data analysis shows that Blacktown to be the winner but as this a long term prediction of 20 years many other factors may come into play at the mean time. Any of the three can be very good location for property investment from risk mitigation point of view.

Future direction

The methodology of this project can serve as a template for identifying potentialities of different area for property investment. But this can (and should) be modified in the following lines.

- Calculate the weightage factors of individual venues by doing regression analysis on the effects different venue types on property value.
- Venue type can be further filtered for such analysis to get a more accurate results.
- In this project population growth has only been taken into account without taking into account without comparing that with historical house price in that area. For more accurate predictions such compare can be included in any future modification of this methodology

