Data Science Capstone Project

An Assessment of the Potential to Open a Shopping Mall

in Perth, Western Australia

1. INTRODUCTION

1.1 Perth is the fastest growing city in Australia. At the same time, it is also the most remote city in the world being some 3000 miles from the next city of similar size, Adelaide. This provides a potentially unique market for businesses. The growth in population means that the demand for goods and services in the city and its surrounding areas continues to increase. However, the city must be self-sufficient in meeting this demand, as distances preclude the cost-effective resupply and movement of goods and services from other nearby sources.

1.2 This unique situation, provides a potentially lucrative opportunity to establish a business providing essential goods and services to new and expanding communities. New communities in the Perth region are being developed to a similar plan, where residential zones are supported by retail centres and industrial parks. The centralization of retail outlets, the relatively fixed customer base and the need to provide a

wide-range of goods and services to fully service all the community's needs, suggests a retail solution based around a shopping mall construct.

Example Perth Suburb Development Plan

1.3 A shopping mall is more than just a place to buy groceries. It is also a place where relaxation and recreational activities can also be conducted by individuals, families and communities. These activities can involve motoring, banking, dining, fashion, beauty and health spas, movies, indoor sports and many more. Shopping malls are a potential win-win opportunity for both customers and retailers. Customers can find everything they need under one roof, while retailers have a mass market and the opportunity to distribute and market their products and services efficiently and effectively. For property developers, the building of shopping malls allows them to earn consistent rental income and achieve economies of scale in terms of operational and management costs. However, the building of shopping malls is not a new concept and there are many companies looking for opportunities to build more. There are many factors involved in the selecting where a new mall should be built; however, as with many property developments schemes one the main criteria for success is 'location, location, location'.

2. BUSINESS PROBLEM

2.1 **Project Objective.** The objective of this capstone project is to analyse and select the best locations in the city of Perth, Western Australia to open a new shopping mall. Using data science methodology and machine learning techniques like clustering, this project aims to provide the information to answer the question: In the city of Perth, Western Australia, if a property developer is looking to open a new shopping mall, where would you recommend to do it?

2.2 **Target Audience.** This capstone project will be useful to property developers looking to open a new mall in the city of Perth, Western Australia, investors seeking a new business opportunity, or retailers looking to expand. This project is timely as the city is forecasting a levelling off in the rate of its growth and the time available to establish new businesses may be limited. A fall in the population could even result in an over-provision of retail centres and the rationalization of even the existing business footprint in the city.

3. DATA SECTION

3.1 **Data Requirements**. To resolve the problem, we will need the following data:

- List of neighbourhoods in Greater Perth. This defines the scope of this project which includes the City of Perth and its suburbs within the region of Western Australia.
- Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and also to get the venue data.
- Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighbourhoods to analyse the scope of existing retail coverage.
- Population data for each neighbourhood, to help identify whether the clusters can provide sufficient business demand to make the investment in a shopping centre profitable.

3.2 **Data sources and extraction methods.** Data in respect of the City of Perth has been drawn from several sources. The main ones being: https://en.wikipedia.org/wiki/List_of_Perth_suburbs and https://quickstats.censusdata.abs.gov.au/census_services/getproduct/census/2016/quickstat/SSC51218

and http://www.corra.com.au/australian-postcode-location-data/. These data sources give a complete listing of the current suburbs of Perth with census information relating to 2016 and the respective latitudes and longitudes. We will use web scraping techniques, where appropriate, to extract the data from these pages, with the help of Python requests and Beautifulsoup packages. Then we will get the geographical coordinates of the neighbourhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighbourhoods. After that, we will use Foursquare API to get the venue data for those neighbourhoods. Foursquare has one of the largest databases of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the Shopping Mall category in order to help us to solve the business problem put forward. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).

4. METHODOLOGY

4.1 **Data Collection.** The first step of the analysis is to collate the relevant data.

This initially involved web-scraping data from two open source data sites to provide a list of 355 Perth

[53]:

	Suburb	Borough	Population	Latitude	Longitude
0	Alexander Heights	Wanneroo	7821	-31.835333	115.849098
1	Alfred Cove	Melville	2192	-32.040061	115.825994
2	Anketell	Kwinana	234	-32.223657	115.835308
3	Applecross	Melville	5693	-32.024363	115.837525
4	Ardross	Melville	3516	-32.024363	115.837525

Suburbs, which form the boundary for our area of analysis. This dataset was reduced in size for all Suburbs for which Latitude and Longitude data was not available. Using this data set, Foursquare was used to identify the venues within 30km of Perth City centre. This large radius was used to gain the widest possible view of the available data across the entire of the Perth Area, taking account of the dispersed nature of the

geography and population. Some 3231 venues of interest were identified from Foursquare, providing 7 data fields.

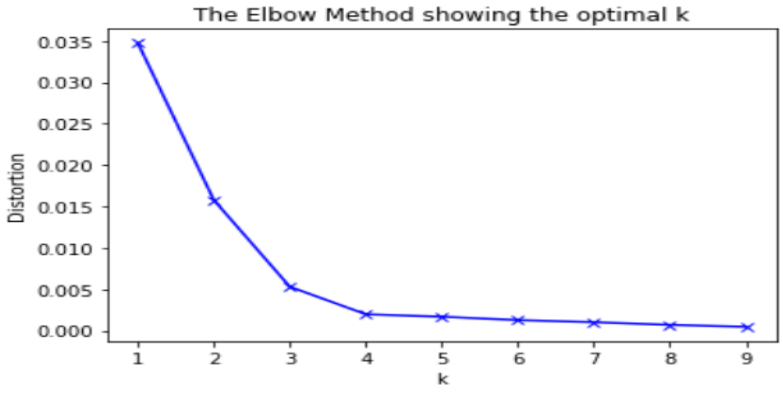
```
print(perthvenues.shape)
perthvenues.head(10)
```

```
(3231, 7)
```

	Suburb	Suburb Latitude	Suburb Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Alexander Heights	-31.835333	115.849098	Domino's Pizza	-31.830509	115.853342	Pizza Place
1	Alexander Heights	-31.835333	115.849098	Alexander Heights Shopping Centre	-31.829769	115.853448	Shopping Mall
2	Alexander Heights	-31.835333	115.849098	Red Rooster	-31.830807	115.853138	Fast Food Restaurant
3	Alexander Heights	-31.835333	115.849098	Cradle Props	-31.840754	115.849288	Photography Studio
4	Alexander Heights	-31.835333	115.849098	Video Ezy - Alexander Heights	-31.830601	115.853329	Video Store
5	Alexander Heights	-31.835333	115.849098	Paloma Sk8 Park	-31.829711	115.851570	Skate Park
6	Alexander Heights	-31.835333	115.849098	Liquorland	-31.830309	115.853543	Liquor Store
7	Alexander Heights	-31.835333	115.849098	Coles	-31.830137	115.854053	Supermarket
8	Alexander Heights	-31.835333	115.849098	Suria Cafe	-31.838613	115.839627	Asian Restaurant
9	Alfred Cove	-32.040061	115.825994	Melville Aquatic Centre	-32.039364	115.830950	Water Park

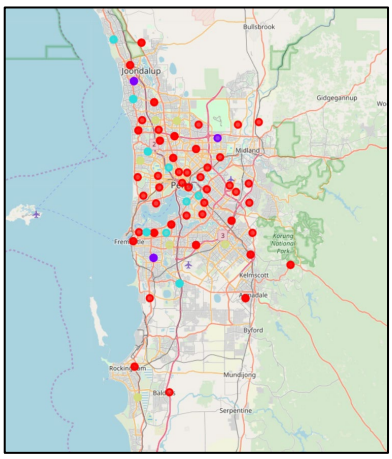
We used the Venue Category to further rationalize the dataset to include only those entries for Suburbs in dataset. This resulted in 244 Suburbs containing a total of 204 different Venue Categories. This was the primary dataset on which our analysis was conducted.

4.2 **Data Exploration.** From this dataset we identified that 73 Suburbs possess one or more Shopping Malls. We then analysed the distribution of these malls across the Suburbs and clustered them using K-means to identify those Suburbs which have the same level of coverage in terms of Shopping Malls. Before conducting this, we performed the Elbow Method to identify the optimal number of clusters. The result showed an optimal number of clusters to be 4.

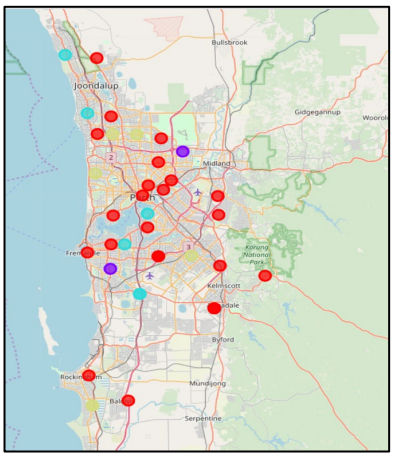


4.3 **Data Analysis**. The results of the clusters created using k-means were analysed in turn. These clusters were then plotted onto a folium map of the Greater Perth area, with different colours representing

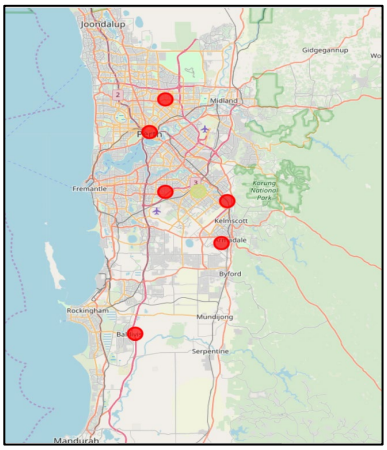
the concentration of Shopping Maps for each Suburb. Suburbs with a population of less than 5000 were not included in this analysis, as it was assumed that these Suburbs would provide a viable market for the investment required. The results of this analysis were plotted onto the Chloropleth maps, based on Stamen



Malls in Suburbs population > 5000



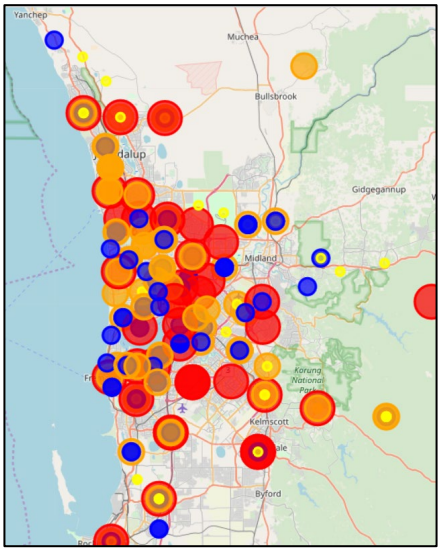
Malls in Suburbs population > 10000

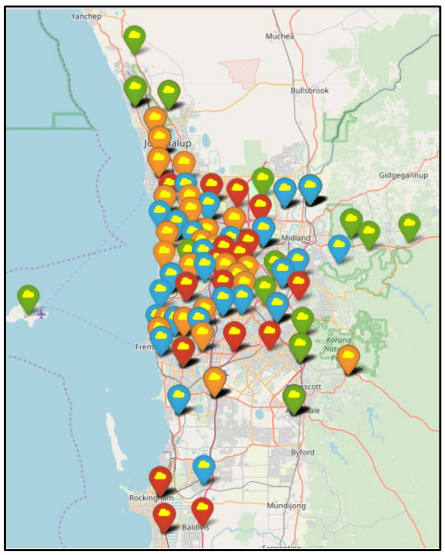


Malls in Suburbs population > 20000

Terrain tiles, as outlined below:

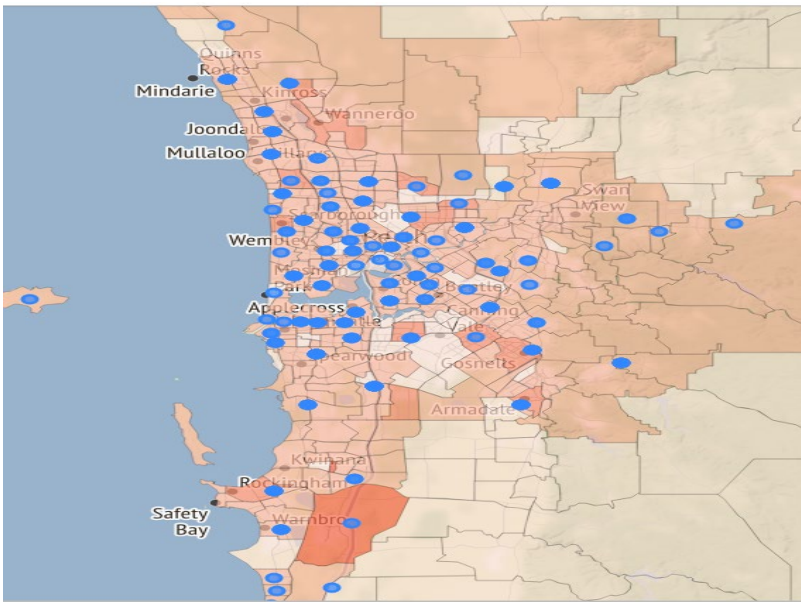
These maps did not provide the clarity required to draw firm conclusions, although they did seem to suggest some areas of Perth where Shopping Malls could be under-represented. At this point, it became clear that more investigation was required into the relationship between the density of population and the location of malls. Initially, Circle Markers and Icons were used to show population densities, with the





markers and icons being coloured coded and sized to represent different population densities. Again,

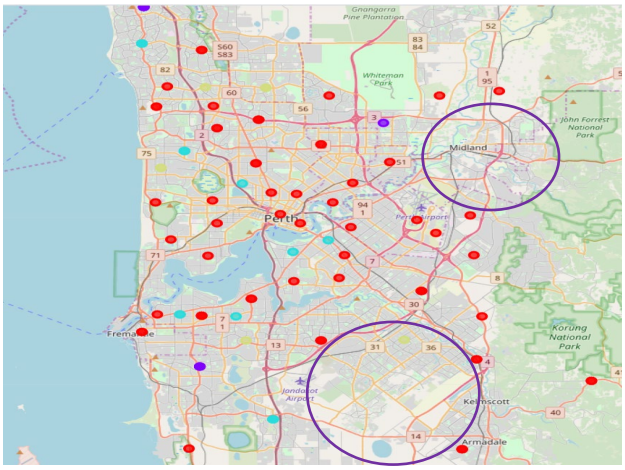
these maps did not provide the necessary clarity to draw definitive conclusions. It was clear another form of mapping was required. The solution was to develop a Heat Map based on the population of each



suburb, mapped against a json file of Western Australia, over which the locations of all the malls could be overlaid. The results were as follows:

5. RESULTS

5.1 **Initial Observations.** An initial analysis of the various visualizations, suggested that there could be potential Shopping Mall opportunities in the Eastern suburbs beyond Perth Airport and in the South Western suburbs in the vicinity of Jandakot Airport, as shown below.



However, it was clear from inconsistencies in the source data that further analysis would be required to verify the accuracy of these initial findings.

6. DISCUSSION

6.1. **Data Accuracy.** The analysis of the geography, population, venues and mapping of the Perth area highlighted a number of concerns over data accuracy and completeness. These involve:

- The lack of current and complete data sources for all the key data.
- 111 Suburbs were excluded from the analysis due to lack of data. This will include high-density population areas and shopping malls that are likely to affect the results of our analysis.

6.2 **Other Factors.** The data analysed in the project included only that which could be easily obtained from open sources. It was a narrow dataset, which did not include wider information that would be relevant to consider in potential investment decisions. These considerations could include information such as:

- Suitability of suburbs/postal codes as the basis to map the location of malls
- Size and coverage of existing mall footprint.
- Impact of other retail venue types on the location of malls
- Future Perth Local Authority development plans

7. CONCLUSION

7.1. This project has set the foundations for the identification of future retail development opportunities in Perth Western Australia. It has signposted areas where future development opportunities in Perth may exist, but our analysis is too incomplete to support effective decision-making at this time. However, this project has highlighted the limitations placed upon it through difficulties in gathering the required data and suggested how some of these limitations could be overcome. Next steps should be to conduct further analysis, taking into account the findings and recommendations of this project.