Moving to Perth

Finding the Best Suburb to Rent a House

Introduction

- In this project we will try to find an optimal suburb to rent a house in. The report is targeted at a couple who have two dogs, love Chinese food, enjoy shopping and looking to move to the city of Perth, Western Australia.
- The couple has hired a data scientist to help and the criteria they have given to select a suburb are those suburbs that have the highest combined quantities of parks, Chinese restaurants and shopping malls.

Data

- A table containing all the suburbs in Perth was scraped from:

https://en.wikipedia.org/wiki/List_of_Perth_suburbs

- The table was converted to a dataframe which contained 355 rows and two columns.

	Suburb	Local government area
0	Alexander Heights	Wanneroo
1	Alfred Cove	Melville
2	Alkimos	Wanneroo
3	Anketell	Kwinana
4	Applecross	Melville
...	...	...
350	Wooroloo	Mundaring
351	Wungong	Armadale
352	Yanchep	Wanneroo
353	Yangebup	Cockburn
354	Yokine	Stirling

Data

- The Geocoder library was used to get latitude and longitude values of each suburb.
- The Foursquare API was used to create a dataframe containing the venue category of each suburb within 1km of its latitude and longitude value.

	Suburb	Suburb Latitude	Suburb Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Alexander Heights	-31.83101	115.85356	Domino's Pizza	-31.830509	115.853342	Pizza Place
1	Alexander Heights	-31.83101	115.85356	Coles	-31.830137	115.854053	Supermarket
2	Alexander Heights	-31.83101	115.85356	Alexander Heights Shopping Centre	-31.829769	115.853448	Shopping Mall
3	Alexander Heights	-31.83101	115.85356	Red Rooster	-31.830807	115.853138	Fast Food Restaurant
4	Alexander Heights	-31.83101	115.85356	Woolworths	-31.826764	115.852742	Supermarket

Data Wrangling

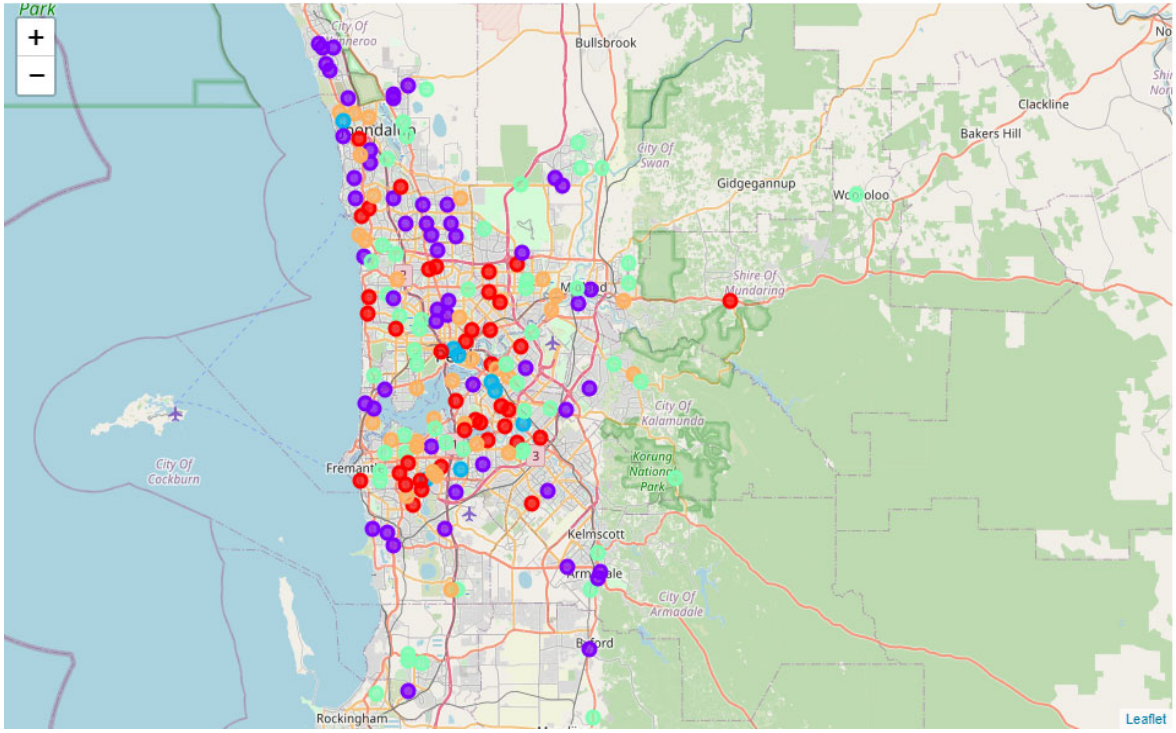
- The dataframe was then filtered to suburbs that contained either parks, Chinese restaurants or shopping malls
- One hot encoding was performed to vectorize the categorical data.
- Finally the dataframe was grouped by the sum of each suburb.

	Suburb	Chinese Restaurant	Park	Shopping Mall
0	Alexander Heights	0	1	1
1	Alfred Cove	0	4	2
2	Alkimos	0	2	0
3	Applecross	0	3	0
4	Ardross	0	1	0

K-Means Clustering

- Quickly find varying combinations of the venue category data
- Find locations with the highest amount of all three venue categories combined
- Scikit-learn was used for the clustering of the data
- Five clusters were chosen for this problem as this would be sufficient to find suburbs with the highest combinations of all three categories

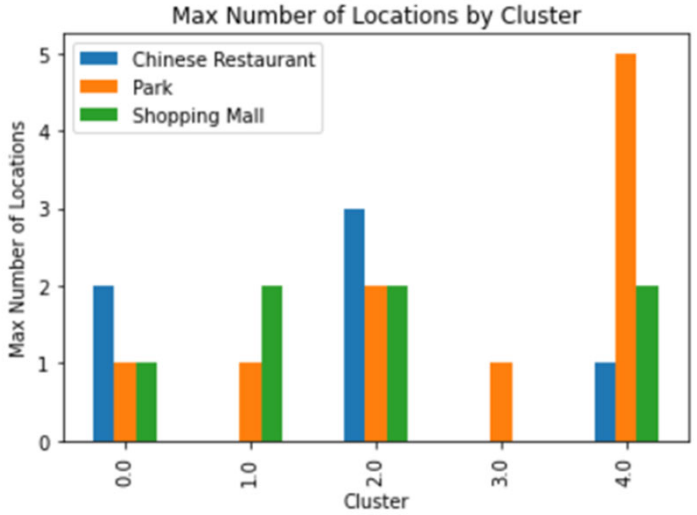
K-Means Clustering



Location of clusters – 0 -red, 1 – purple, 2 – blue, 3 – green and 4 - orange

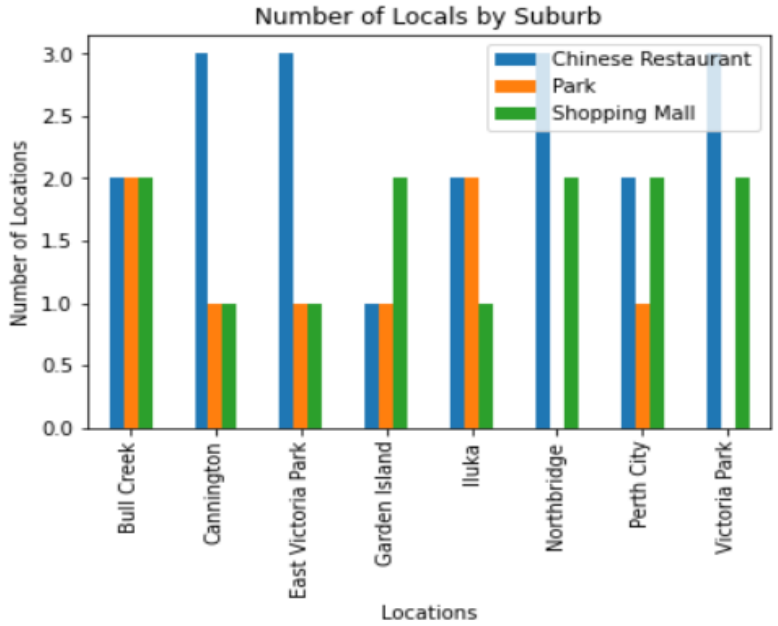
K-Means Clustering

- Cluster 2 - highest mean Chinese restaurants and shopping malls and highest max Chinese restaurants and shopping malls
- Cluster 4 – highest mean & max parks but low mean for shopping mall and Chinese restaurant
- Suburbs in **cluster 2** have the highest quantities of combined parks, Chinese restaurants and shopping malls

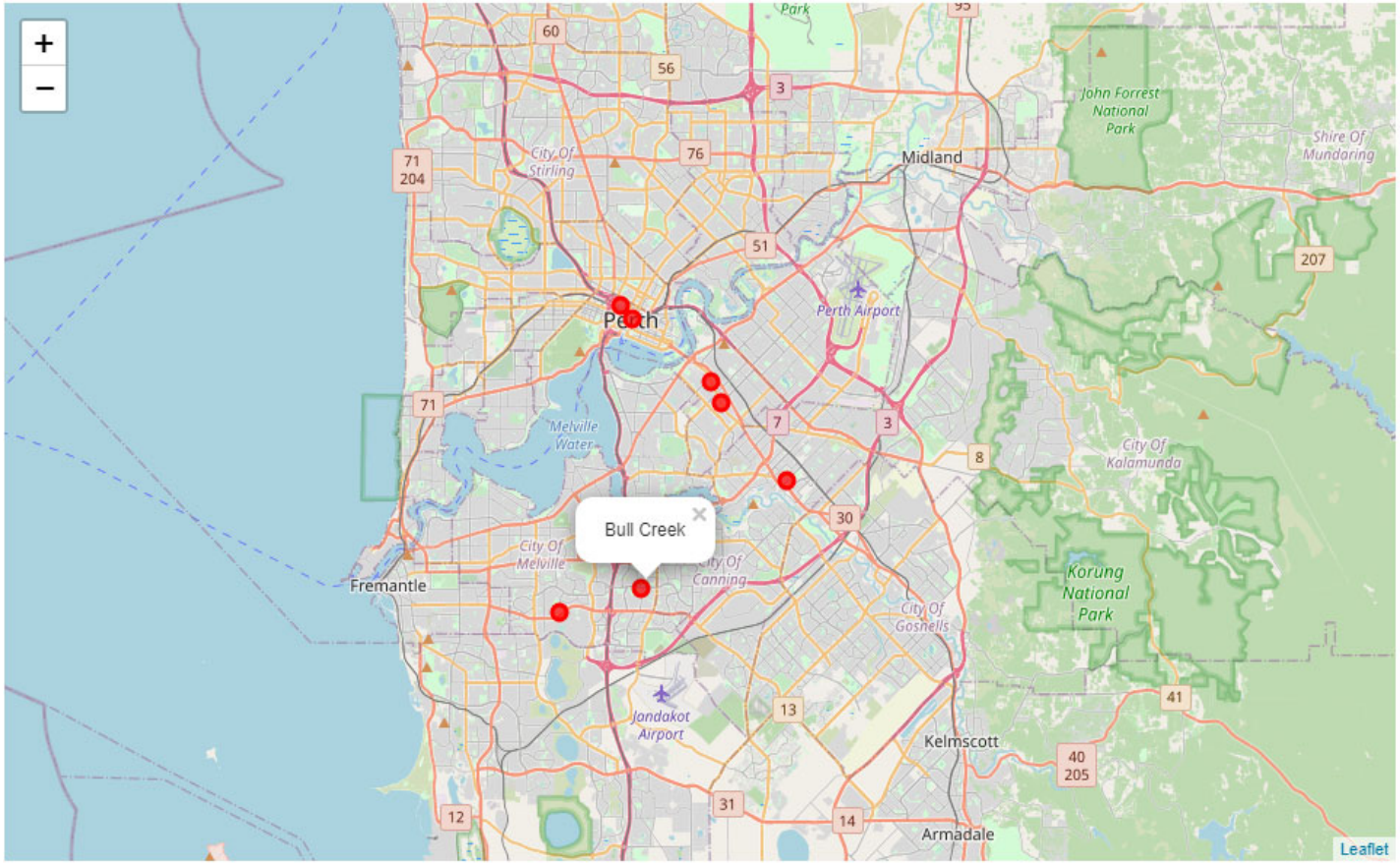


Cluster 2 Suburbs

- Bull Creek – highest number of parks and shopping malls and the second highest number of Chinese restaurants
- All excluding 2 suburbs are reasonable choices as they provide at least one option for each venue category
- Northbridge and Victoria Park are not recommended as they have no parks.



Bull Creek



Cluster 2 suburbs shown in red