The Battle of Neighborhoods - Melbourne

Introduction

Background
This is the capstone project of IBM Data Science Professional Certificate.
In this project I will use Foursquare location data to explore neighborhoods in Melbourne and solve a problem.

Problem Statement
I will explore the neighborhoods in Melbourne to answer the question: "Where is the Best location for a licensed restaurant in Melbourne".

Who would be interested?
Melbourne is the food capital of Australia. There is an immense variety of restaurants and opportunity for a business owner if they find the right location. A project such as this is of direct interest and benefits to new restaurateurs in Melbourne.

Data

Data Origin
My main data sources are from
1.'Corra.com.au', which has the Melbourne suburbs and their latitude and longitude coordinates
2.Venues data from Foursquare

Data Description
Once formatted into a dataframe, the data is as below:

	Neighborhood	Latitude	Longitude
0	MELBOURNE	-38.365017	144.765920
1	EAST MELBOURNE	-37.816640	144.987811
2	WEST MELBOURNE	-37.806255	144.941123
3	MELBOURNE	-37.837324	144.976335
4	WORLD TRADE CENTRE	-37.822262	144.954856

Feature Selection
In this project, I will get data of recommended venues inside 1000 meters radius of every neighborhood, calculate the top10 most common venues by its category as features. Plus, the

11th feature is if there is a bus/metro station nearby.

Methodology

Exploratory Data Analysis

The venues data from Foursquare is then used to create a dataFrame :
One-hot encoding is used to get mean of each category for every neighborhood:
Frequency of categories for each neighborhood is obtained and the top venues differentiated as follows:

	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
0	MELBOURNE	-38.365017	144.765920	0.0	Home Service	Café	Light Rail Station	Australian Restaurant	Hotel	Sandwich Place	Steakhouse	Asian Restaurant
1	EAST MELBOURNE	-37.816640	144.987811	0.0	Café	Thai Restaurant	Indian Restaurant	Museum	Light Rail Station	Sushi Restaurant	Gym	Park
2	WEST MELBOURNE	-37.806255	144.941123	0.0	Train Station	Café	Platform	Bar	Bus Stop	Theater	Fabric Shop	Falafel Restaurant
3	MELBOURNE	-37.837324	144.976335	0.0	Home Service	Café	Light Rail Station	Australian Restaurant	Hotel	Sandwich Place	Steakhouse	Asian Restaurant
4	WORLD TRADE CENTRE	-37.822262	144.954856	0.0	Café	Bar	Hotel	Japanese Restaurant	Coffee Shop	Gym	French Restaurant	Italian Restaurant

Algorithm Selection

k-means is used to cluster the neighborhoods and find a pattern in them. The reason for this is that we need to understand how the data is organized in an unlabelled dataset. Clustering is split into 5 clusters.

Results

Folium library is used to create a map with clustered markers
All 5 clusters are sizable but do what specific traits

Cluster 1

```
melbourne_merged.loc[melbourne_merged['Cluster Labels'] == 0, melbourne_merged.columns[[1] + list(range(5, melbourne_merged.
shape[1]))]]
```

	Latitude	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Static
0	-38.365017	Café	Light Rail Station	Australian Restaurant	Hotel	Sandwich Place	Steakhouse	Asian Restaurant	Pizza Place	Vietnamese Restaurant	Yes
1	-37.816640	Thai Restaurant	Indian Restaurant	Museum	Light Rail Station	Sushi Restaurant	Gym	Park	Steakhouse	Hardware Store	Yes
2	-37.806255	Café	Platform	Bar	Bus Stop	Theater	Fabric Shop	Falafel Restaurant	Farm	Farmers Market	Yes
3	-37.837324	Café	Light Rail Station	Australian Restaurant	Hotel	Sandwich Place	Steakhouse	Asian Restaurant	Pizza Place	Vietnamese Restaurant	Yes
4	-37.822262	Bar	Hotel	Japanese Restaurant	Coffee Shop	Gym	French Restaurant	Italian Restaurant	Fried Chicken Joint	Cocktail Bar	No
5	-37.823258	Bar	Grocery Store	Hotel	Italian Restaurant	Theater	Art Gallery	Performing Arts Venue	Australian Restaurant	Steakhouse	Yes
6	-37.814719	Restaurant	Coffee Shop	Shopping Mall	Australian Restaurant	Pizza Place	Bar	Pub	Sandwich Place	Thai Restaurant	No

Cluster 2

```
melbourne_merged.loc[melbourne_merged['Cluster Labels'] == 1, melbourne_merged.columns[[1] + list(range(5, melbourne_merged.
shape[1]))]]
```

	Latitude	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Station
78	-37.700040	Tennis Court	Basketball Court	Food Service	Zoo Exhibit	Fabric Shop	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	No
109	-37.700648	Zoo Exhibit	Iraqi Restaurant	Event Space	Fabric Shop	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Field	No
180	-37.744219	Park	Country Dance Club	Event Space	Fabric Shop	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Field	No
184	-37.739262	Tennis Court	Event Service	Event Space	Fabric Shop	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Field	No
193	-37.690585	Convenience Store	Zoo Exhibit	Fish Market	Fabric Shop	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Field	No
197	-37.730758	Fast Food Restaurant	Zoo Exhibit	Iraqi Restaurant	Event Space	Fabric Shop	Falafel Restaurant	Farm	Farmers Market	Field	No
217	-37.796947	Zoo Exhibit	Iraqi Restaurant	Event Space	Fabric Shop	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Field	No
275	-37.805715	Bar	Zoo Exhibit	French Restaurant	Fabric Shop	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Field	No

Cluster 3

```
melbourne_merged.loc[melbourne_merged['Cluster Labels'] == 2, melbourne_merged.columns[[1] + list(range(5, melbourne_merged.
shape[1]))]]
```

	Latitude	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Station
9	-37.808769	Bakery	Pizza Place	Thai Restaurant	Park	Train Station	Gastropub	Gym	Supermarket	Juice Bar	Yes
11	-37.814624	Electronics Store	Food Truck	Zoo Exhibit	Fish Market	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Field	No
19	-37.842477	Grocery Store	Cooking School	Convenience Store	Ice Cream Shop	Thai Restaurant	Beer Garden	Pizza Place	Zoo Exhibit	Filipino Restaurant	No
22	-37.856902	Park	Restaurant	Fried Chicken Joint	Supermarket	Fish & Chips Shop	Seafood Restaurant	Bakery	Coffee Shop	Diner	No
28	-37.834094	Chinese Restaurant	Gourmet Shop	Zoo Exhibit	Fish & Chips Shop	Fabric Shop	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	No
67	-37.794333	Park	Fried Chicken Joint	Performing Arts Venue	Fish & Chips Shop	Gym	Pizza Place	Burger Joint	Bar	Ice Cream Shop	No
84	-37.765707	Supermarket	Bar	Train Station	Clothing Store	Coffee Shop	Seafood Restaurant	Performing Arts Venue	Shopping Mall	Pharmacy	Yes

Cluster 4

```
melbourne_merged.loc[melbourne_merged['Cluster Labels'] == 3, melbourne_merged.columns[[1] + list(range(5, melbourne_merged.
shape[1]))]]
```

	Latitude	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Station
31	-37.768904	Skating Rink	Gas Station	Bakery	Zoo Exhibit	Fish & Chips Shop	Event Space	Fabric Shop	Falafel Restaurant	Farm	Yes
35	-37.734277	Pub	Restaurant	Fish & Chips Shop	Zoo Exhibit	Fish Market	Event Space	Fabric Shop	Falafel Restaurant	Farm	No
74	-37.718965	Pub	Café	Pizza Place	Fish Market	Event Space	Fabric Shop	Falafel Restaurant	Farm	Farmers Market	No
87	-37.753802	Sports Club	Pub	Zoo Exhibit	Fish Market	Fabric Shop	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Yes
142	-37.529372	Café	Sandwich Place	Zoo Exhibit	Fish & Chips Shop	Event Space	Fabric Shop	Falafel Restaurant	Farm	Farmers Market	No
460	-38.099498	Pub	Train Station	Shopping Mall	Filipino Restaurant	Ethiopian Restaurant	Event Service	Event Space	Fabric Shop	Falafel Restaurant	Yes
481	-38.180291	Sports Bar	Pub	Gas Station	Fish Market	Event Space	Fabric Shop	Falafel Restaurant	Farm	Farmers Market	Yes
496	-38.142320	Pub	Restaurant	Hotel Bar	Zoo Exhibit	Filipino Restaurant	Event Service	Event Space	Fabric Shop	Falafel Restaurant	No

Cluster 5

```
melbourne_merged.loc[melbourne_merged['Cluster Labels'] == 4, melbourne_merged.columns[[1] + list(range(5, melbourne_merged.
shape[1]))]]
```

	Latitude	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Station
526	-38.293801	Zoo Exhibit	Country Dance Club	Fabric Shop	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Field	Filipino Restaurant	No
539	-38.748434	Campground	Zoo Exhibit	Fish Market	Fabric Shop	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Field	No
657	-38.346708	Zoo Exhibit	Country Dance Club	Fabric Shop	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Field	Filipino Restaurant	No
888	-37.962809	Lake	Zoo Exhibit	Event Service	Fabric Shop	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Field	No
986	-37.484940	Zoo Exhibit	Country Dance Club	Fabric Shop	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Field	Filipino Restaurant	No
1326	-37.286853	Zoo Exhibit	Country Dance Club	Fabric Shop	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Field	Filipino Restaurant	No
1359	-37.153711	Zoo Exhibit	Country Dance Club	Fabric Shop	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Field	Filipino Restaurant	No
1736	-36.803325	Zoo Exhibit	Country Dance Club	Fabric Shop	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Field	Filipino Restaurant	No
2096	-36.802505	Campground	Zoo Exhibit	Fish Market	Fabric Shop	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Field	No

Recommendation

Cluster 1 has a nice mix of cafes, restaurants and pubs.
Cluster 2 is largely characterised by tourist locations.
Cluster 3has a predominance of parks.
Cluster 4 is strongly related to the presence of pubs.
Cluster 5 is a mix of tourism and camping grounds.

Given the question we are traying to answer is a licensed restaurant. Cluster 4 seems the most applicable cluster to concentrate on.

Discussion

K means is the most applicable algorithm selection. Once of the factors to work through is the sheer amount of venues in the Melbourne area. This is both a good thing, but a challenging thing when trying to find the best location for a restaurant. Grouping all this information into 5 clusters goes along way to making sense of the data and business possibilities. What would be great to see at a later stage is opening hours. The licensed restaurant can only be opened at certain times in certain locations. A predominance of pubs tells us that the opening hours will be find in Cluster 4 location. Re running the algorithm though with some feature engineering on opening hours would assist in interpreting the best location and k means results.

Conclusion

To conclude, neighborhoods data was collected and formatted. Venues data using Foursquare was also obtained. Features were transformed with one hot encoding and the top venues differentiated. K means was selected as the best algorithm and 5 clusters shaped into the data. The resulting clusters indicated a preference for cluster 4 and given the amount of licencsed premised and the problem statement centering on the best location for a licenced premises in Melbourne.