IBM – Coursera
Data Science Specialization

Capstone Project Report

Welcome to Melbourne

Deepak Gupta – 2019

Table of Contents

Introduction

Melbourne has been voted the most liveable city in the World for seven years in a row, which attract a lot of families or individuals to migrate to Melbourne. The question / problem comes in deciding where to step foot in Melbourne, which suburb to live in. For many young families the priority is to provide best possible education to their children at an affordable price. In addition to good schools, the presence of other basic amenities is also important (Transport, Hospitals, Restaurants and others)

In this project the objective is to select top 10 Public Primary and Secondary Schools. List all the amenities for the corresponding suburbs where the schools are located with the median rent in the area so that the user be able to decide a suburb to settle down.

Data

In order to carry out the investigation and determining the most suitable Suburb, we need access to certain information. This information can be retrieved from the Data available at various sites and consolidated and evaluated to reach to the conclusion. Data required for our project is as follows:

1. Melbourne Postcode along with Geocoordinates

A list of all Melbourne suburbs along with their Geo – coordinates is required so that the suburbs can be visualised, and list will be used to iterate through the Foursquare website to obtain the list of venues around each Suburb.

a.https://www.costlessquotes.com.au/postcode_tool/postcode_list_VIC.php – used to obtain list of postcodes and suburbs

	Postcode	District	City/ Town/ Suburb
0	3000	Melbourne	Melbourne
1	3001	Melbourne	Melbourne
2	3002	Melbourne	East Melbourne
3	3003	Melbourne	West Melbourne
4	3004	Melbourne	St Kilda Road Melbourne

b. Geo-Coordinates for Victoria were assembled via various sites as there was no single source for them which was available to use. The information was consolidated in a csv file and loaded on to the system.

	Postcode	Suburb	State	Lat	Long
0	221	Barton	ACT	-35.20	149.10
1	800	Darwin	NT	-12.80	130.96
3	804	Parap	NT	-12.43	130.84
4	810	Alawa	NT	-12.38	130.88
5	810	Brinkin	NT	-12.38	130.88

2. List of Top Public / Private Schools in Melbourne

a. https://bettereducation.com.au/school - used to retrieve the list of top 10 Primary Schools in Melbourne both in Public and Private sector so the people can decide depending on their financial circumstances.

	School	Postcode	State Overall Score	Total Enrolments	Sector
0	Presbyterian Ladies' College	3125	100	1412	Non-government
1	St Andrews Christian College	3152	100	581	Non-government
2	Burwood East Primary School	3151	100	274	Government
3	Huntingtower School	3149	100	686	Non-government
4	Haileybury College	3173	100	3754	Non-government

3. Gather Neighbourhood information from Foursquare

a. http://foursquare.com – used to retrieve the list of venues for each suburb so that people can review and assess which suburb would be a good option and fulfil their requirements in terms of food, café and other facilities.

Methodology

The methodology used for this project is as follows:

1. Data is extracted from websites using Web Scraping technique using Beautiful Soup library available in Pandas

2. Data is also loaded via csv files for the Geo-Coordinates for the Suburbs of Melbourne

3. Data is cleaned by dropping with missing longitude or latitude values for the Geo-Coordinates file.

4. Once the Data frame is constructed with all Melbourne suburbs along with their Geo-Coordinates, its used to extract information from the Foursquare website to get the information for all the venues for each suburb.

5. The venues information are sorted to list top 10 Venues for each suburb and transform the information into a dataframe.

6. Now using the kmeans as the clustering method, the suburbs are divided into 5 clusters according to the venues.

7. The clusters are depicted on the map, to emphasise the results using the map as visualisation tool.

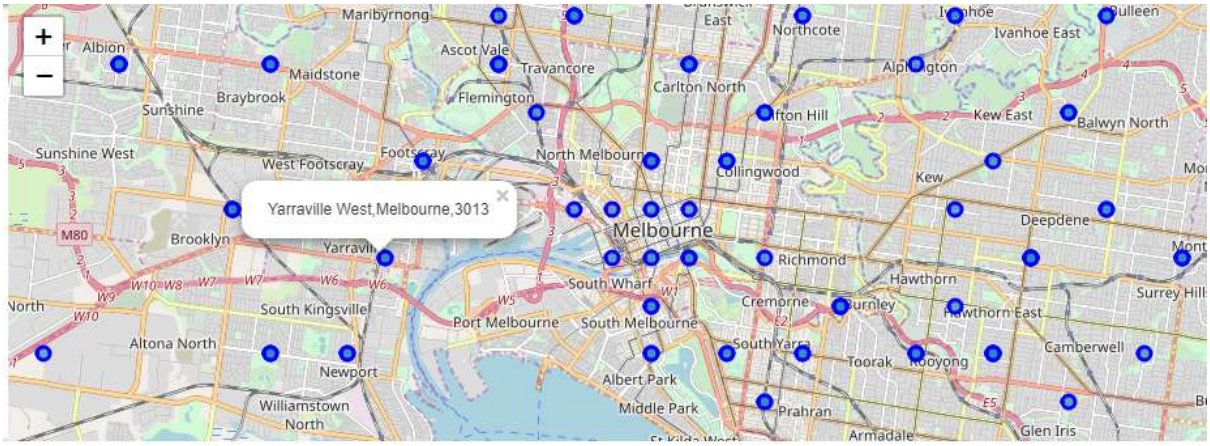
8. The list of Top Primary schools are obtained from the website again using Web Scraping technique and Beautiful Soup Library available in Pandas

9. Finally the list is segregated into Top 10 Primary Government Schools and Top 10 Primary Private Schools in Melbourne

Results

The following results were achieved and are listed below with relevant snapshots from the Jupyter Notebook

1. Visual Depiction of Melbourne Suburbs



	Postcode	District	Suburb	State	Lat	Long
2863	3977	Melbourne	Skye	VIC	-38.21	145.32
2914	8004	Melbourne	St Kilda Road	VIC	-37.84	144.98
2915	8006	Melbourne	Abeckett Street	VIC	-37.81	144.96
2917	8009	Melbourne	Flinders Lane	VIC	-37.82	144.96
2918	8010	Melbourne	Law Courts	VIC	-38.19	146.29

2. Venues information from Foursquare

a. Json Format

```
results = requests.get(url).json()
results
```

```
4]: {'meta': {'code': 200, 'requestId': '5c5e39176a60712d31ca1e70'},
     'response': {'groups': [{'items': [{'reasons': {'count': 0,
           'items': [{'reasonName': 'globalInteractionReason',
             'summary': 'This spot is popular',
             'type': 'general'}]},
         'referralId': 'e-0-4d9e49d4a4675481aa278be6-0',
         'venue': {'categories': [{'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/food/coffeeshop_',
            'suffix': '.png'},
           'id': '4bf58dd8d48988d1e0931735',
           'name': 'Coffee Shop',
           'pluralName': 'Coffee Shops',
           'primary': True,
           'shortName': 'Coffee Shop'}],
          'id': '4d9e49d4a4675481aa278be6',
          'location': {'address': '8 Exploration Ln.',
           'cc': 'AU',
           'city': 'Melbourne',
           'country': 'Australia',
           'crossStreet': 'at Little Lonsdale St.',
```

b. DataFrame with Venues for all Suburbs

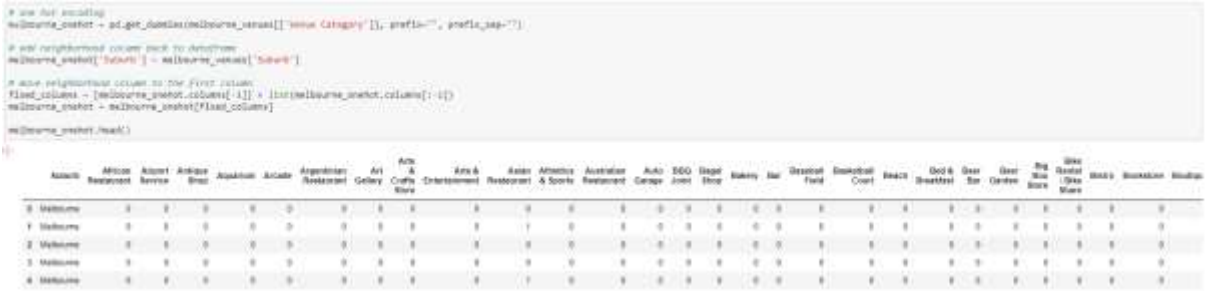
```
In [20]: print(melbourne_venues.shape)
         melbourne_venues.head()
```

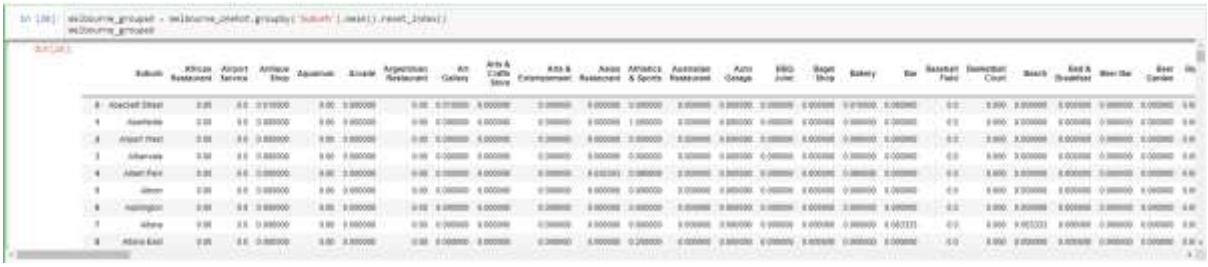
```
(6344, 7)
```

click to scroll output; double click to hide

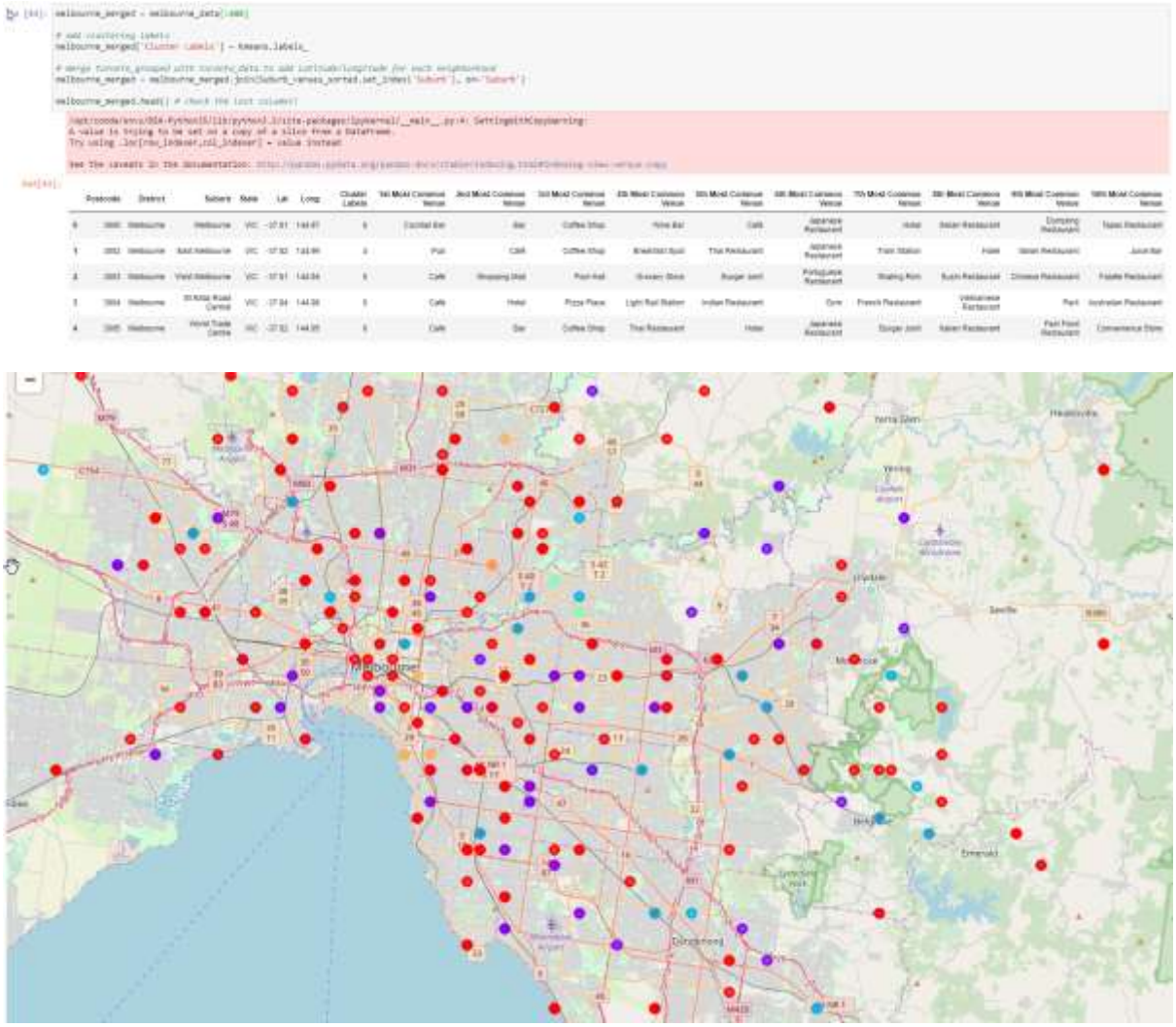
Out[20]:

	Suburb	Suburb Latitude	Suburb Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Melbourne	-37.81	144.97	The League of Honest Coffee	-37.809279	144.968483	Coffee Shop
1	Melbourne	-37.81	144.97	Gingerboy	-37.811223	144.970982	Asian Restaurant
2	Melbourne	-37.81	144.97	Traveller	-37.811391	144.971232	Coffee Shop
3	Melbourne	-37.81	144.97	Longrain Restaurant & Bar	-37.810733	144.971147	Thai Restaurant
4	Melbourne	-37.81	144.97	Rice Papr Scrs	-37.811298	144.971311	Asian Restaurant





3. Suburbs Clustering using kmeans method



3. Top10 Primary Government & Non-Government Schools

```
df_Schools.head()
```

	School	Postcode	State Overall Score	Total Enrolments	Sector
0	Presbyterian Ladies' College	3125	100	1412	Non-government
1	St Andrews Christian College	3152	100	581	Non-government
2	Burwood East Primary School	3151	100	274	Government
3	Huntingtower School	3149	100	686	Non-government
4	Haileybury College	3173	100	3754	Non-government

```
## Merging Postcode and Schools Dataframe

school_data = pd.merge(melbourne_data,school_data, on='Postcode')
school_data.head()

## Converting State Overall Score Column to numeric so that it can be used to Sort the Dataframe

school_data['State Overall Score'] = pd.to_numeric(school_data['State Overall Score'])
school_data = school_data.sort_values(by='State Overall Score',ascending=False)
school_data.head()
```

']:

	Postcode	District	Suburb	State	Lat	Long	School	State Overall Score	Total Enrolments	Sector
678	3152	Melbourne	Knox City Centre	VIC	-37.87	145.24	St Andrews Christian College	100	581	Non-government
326	3103	Melbourne	Balwyn	VIC	-37.81	145.08	Fintona Girls' School	100	456	Non-government
365	3109	Melbourne	Tunstall Square Po	VIC	-37.81	145.19	Beverley Hills Primary School	100	473	Government
664	3150	Melbourne	Wheelers Hill	VIC	-37.88	145.17	Glendal Primary School	100	873	Government
786	3173	Melbourne	Keysborough	VIC	-37.99	145.15	Haileybury College	100	3754	Non-government

```
## Creating a Dataframe with Top 10 Government Primary Schools

Gov_Schools = school_data.loc[school_data['Sector'] == 'Government']
Gov_Schools.reset_index(drop=True,inplace=True)
Gov_Schools = Gov_Schools[:10]
Gov_Schools.head(100)
```

]:

	Postcode	District	Suburb	State	Lat	Long	School	State Overall Score	Total Enrolments	Sector
0	3109	Melbourne	Tunstall Square Po	VIC	-37.81	145.19	Beverley Hills Primary School	100	473	Government
1	3150	Melbourne	Wheelers Hill	VIC	-37.88	145.17	Glendal Primary School	100	873	Government
2	3167	Melbourne	Oakleigh South	VIC	-37.93	145.10	Oakleigh South Primary School	100	1006	Government
3	3150	Melbourne	Glen Waverley	VIC	-37.88	145.17	Mount View Primary School	100	1083	Government
4	3150	Melbourne	Glen Waverley	VIC	-37.88	145.17	Glendal Primary School	100	873	Government
5	3150	Melbourne	Brandon Park	VIC	-37.88	145.17	Glendal Primary School	100	873	Government
6	3150	Melbourne	Wheelers Hill	VIC	-37.88	145.17	Mount View Primary School	100	1083	Government
7	3109	Melbourne	Tunstall Square Po	VIC	-37.81	145.19	Doncaster Gardens Primary School	100	681	Government
8	3150	Melbourne	Brandon Park	VIC	-37.88	145.17	Mount View Primary School	100	1083	Government
9	3109	Melbourne	The Pines	VIC	-37.81	145.19	Doncaster Gardens Primary School	100	681	Government

In [69]:
```
## Creating a Dataframe with Top 10 Non - Government Primary Schools

NonGov_Schools = school_data.loc[school_data['Sector'] == 'Non-government']
NonGov_Schools.reset_index(drop=True,inplace=True)
#NonGov_Schools = df_Gov_Schools[:10]
NonGov_Schools.head(100)
```

Out[69]:

	Postcode	District	Suburb	State	Lat	Long	School	State Overall Score	Total Enrolments	Sector
0	3152	Melbourne	Knox City Centre	VIC	-37.87	145.24	St Andrews Christian College	100	581	Non-government
1	3103	Melbourne	Balwyn	VIC	-37.81	145.08	Fintona Girls' School	100	456	Non-government
2	3173	Melbourne	Keysborough	VIC	-37.99	145.15	Haileybury College	100	3754	Non-government
3	3152	Melbourne	Wantirna South	VIC	-37.87	145.24	St Andrews Christian College	100	581	Non-government
4	3152	Melbourne	Wantirna South	VIC	-37.87	145.24	Waverley Christian College	100	1867	Non-government
5	3149	Melbourne	Mount Waverley	VIC	-37.88	145.13	Huntingtower School	100	686	Non-government
6	3152	Melbourne	Knox City Centre	VIC	-37.87	145.24	Waverley Christian College	100	1867	Non-government
7	3149	Melbourne	Pinewood	VIC	-37.88	145.13	Huntingtower School	100	686	Non-government
8	3122	Melbourne	Hawthorn West	VIC	-37.84	145.05	Scotch College	100	1883	Non-government
9	3122	Melbourne	Hawthorn West	VIC	-37.84	145.05	Erasmus School	100	100	Non-government
10	3122	Melbourne	Auburn South	VIC	-37.84	145.05	Erasmus School	100	100	Non-government

Discussion

As the initial problem / requirement was to enable a new migrant wanting to settle in Melbourne, decide which suburb would be a good option. The priority was the education so once according to personal financial circumstances the schools are shortlisted. Then the various clusters of suburbs can be reviewed which are created based on other facilities available nearby each suburb.

Conclusion

Finally, to conclude I would like to emphasis that having the right information is very critical in making informed decision. As new migrant is looking for as much information as possible, using the tool created would assist them in selecting a suburb primary on the basis of education and then shortlist on the one which has other facilities required as well.