Analysis of opening a new shopping centre in Sydney

Introduction

A shopping centre is a multipurpose place that contributes significantly in liveability of a region. A variety of stores, restaurants, entertaining places, cinemas and sport complex are examples of places in a shopping centre that motivates people go shopping. In addition to shopping, entertaining activities are of prime importance in these places. Therefore, building a new shopping mall in a region could have a significant impact on people lives in the region.

Probably the main question for market study would be the place that shopping mall is aimed to be built. One of the major factors is the number of neighbouring shopping centres in the area and also in the city.

Current study reveals specific areas in Sydney which are potentially suitable for opening a new shopping centre. The study uses data analysis techniques together with clustering machine learning to suggest new places for this purpose.

Data Source

Data of Sydney suburbs is web scraped from Wikipedia. The link to data of Sydney suburbs is https://en.wikipedia.org/wiki/Category:Suburbs_of_Sydney [1].The Geographical data for location of each suburb is obtained from Geocoder [2] Geopy and [3] Python libraries. Data for venue of shopping centres in Sydney is obtained from Fouresquare [4].

Methodology

Data acquisition and cleaning

List of neighbourhoods

Data downloaded or scraped from multiple sources and combined in a table for further processing. First, data for Sydney suburbs were scraped from Wikipedia web page [1] (https://en.wikipedia.org/wiki/Category:Suburbs_of_Sydney). There

were 200 neighbourhood in this web page. However, there was not a table for data in this web page. So, initially list of neighbourhoods were scraped from this page by using Beautifulsoup package. After scraping the list of suburbs three steps were followed to create data frame. I) Creating a list to store neighbourhood data, II) appending the data into the list and III) create a new data frame from the list.

Geographical data for neighbours

The Wikipedia web page does not have geographical data for locations in neighbourhoods. Therefore, the next step was acquiring longitude and latitude of each location using Geocoder Python package. This includes following steps: defining a function to get coordinates, calling the function to get the coordinates, store in a new list using list comprehension, creating temporary data frame to populate the coordinates into Latitude and Longitude and finally merging the coordinates into the original dataframe.

A map from dataframe was created using Folium package and is shown in Figure 1.

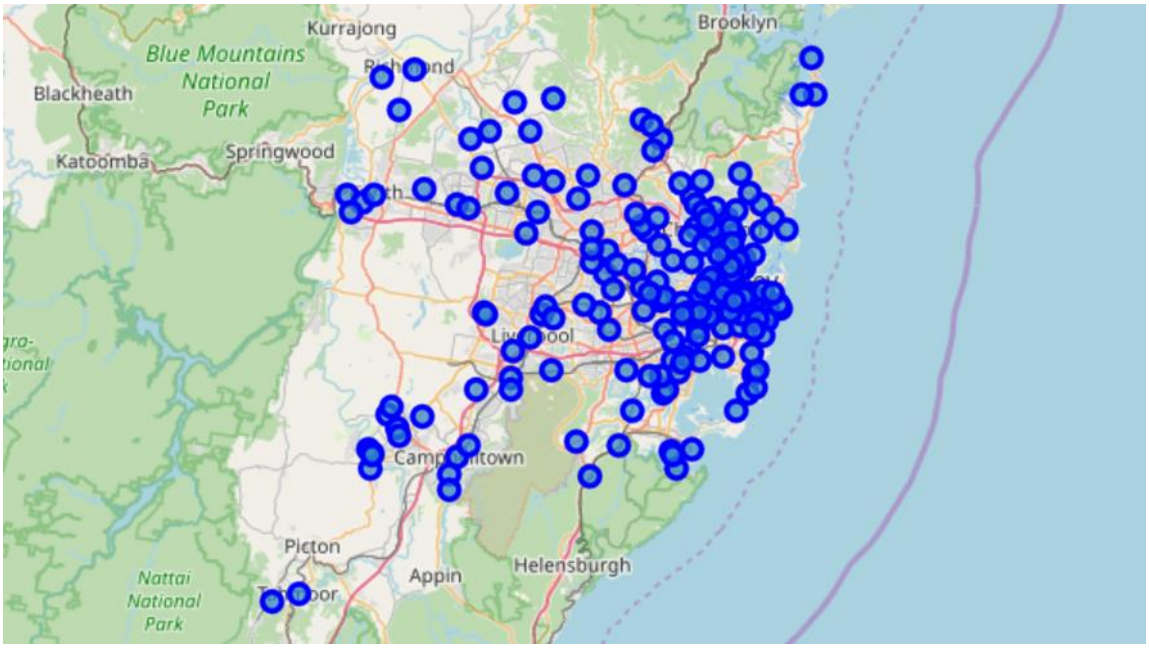

Figure 1. A map of 200 neighbourhood locations in Sydney created by Folium package.

Venue data for neighbours

Foursquare is a popular web service that provides comprehensive data for more than 150 million locations around the globe. The source for venue data was chosen to be Foursquare. A developer app was created in Foursquare web site and through its API, venue data for each neighbourhood was stored in a new data frame which was later merged to original data frame. The first five rows of the resulting data frame is shown in Table 1.

Table 1. List of neighbourhoods in Sydney with geographical and venue data. First column was scraped from Wikipedia, 2nd and 3rd columns were got from Geocoder and last four columns were obtained from Foursquare.

	Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	Agnes Banks, New South Wales	-33.61445	150.71083	Wog Mobile	-33.619594	150.706412	Rental Car Location
1	Agnes Banks, New South Wales	-33.61445	150.71083	Yarramundi Reserve	-33.613377	150.698378	Nature Preserve
2	Agnes Banks, New South Wales	-33.61445	150.71083	D & V Turf Supplies Pty Ltd	-33.623196	150.702574	Other Repair Shop
3	Agnes Banks, New South Wales	-33.61445	150.71083	Navua Reserve	-33.608786	150.696020	Park
4	Agnes Banks, New South Wales	-33.61445	150.71083	Trees Adventure	-33.612809	150.692359	Rock Climbing Spot

Data Analysis

Initially the number of unique categories which can be curated from all the returned venues were determined. Thereafter, list of categories were check to contain “Shopping Mall”.

Analysing Neighbours

Neighbours were analysed by grouping the rows and taking the mean of frequency of occurrence for each venue category. This step is also helpful for doing clustering. Now, “Shopping Mall” data is filtered and we can start clustering for this category. Table 2. shows the mean of frequency of occurrence of each venue categories.

Table 2. Mean of frequency of occurrence of each venue category

	Neighborhoods	ATM	Accessories Store	Afghan Restaurant	Airport	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Arepa Restaurant	...	Warehouse Store	Water Park	Whisky Bar	Wine Bar	Wine Shop
0	Agnes Banks, New South Wales	0.0	0.00	0.0	0.0	0.00	0.0	0.0	0.00	0.0	...	0.00	0.0	0.00	0.00	0.0
1	Alexandria, New South Wales	0.0	0.00	0.0	0.0	0.00	0.0	0.0	0.00	0.0	...	0.00	0.0	0.00	0.00	0.0
2	Allambie Heights, New South Wales	0.0	0.00	0.0	0.0	0.00	0.0	0.0	0.02	0.0	...	0.00	0.0	0.00	0.00	0.0
3	Annandale, New South Wales	0.0	0.00	0.0	0.0	0.00	0.0	0.0	0.00	0.0	...	0.00	0.0	0.00	0.01	0.0
4	Appin, New South Wales	0.0	0.00	0.0	0.0	0.00	0.0	0.0	0.00	0.0	...	0.00	0.0	0.00	0.00	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
195	Summer Hill, New South Wales	0.0	0.00	0.0	0.0	0.00	0.0	0.0	0.00	0.0	...	0.00	0.0	0.00	0.00	0.0
196	Surry Hills, New South Wales	0.0	0.00	0.0	0.0	0.00	0.0	0.0	0.00	0.0	...	0.00	0.0	0.01	0.02	0.0
197	Sydney Olympic Park	0.0	0.01	0.0	0.0	0.00	0.0	0.0	0.01	0.0	...	0.01	0.0	0.00	0.00	0.0
198	Tahmoor, New South Wales	0.0	0.00	0.0	0.0	0.00	0.0	0.0	0.00	0.0	...	0.00	0.0	0.00	0.00	0.0
199	Tempe, New South Wales	0.0	0.00	0.0	0.0	0.02	0.0	0.0	0.00	0.0	...	0.00	0.0	0.00	0.00	0.0

200 rows × 332 columns

Table 3. shows the venue category for “Shopping Mall”. This data will be used for further analysing by K-means clustering method.

Table 3. Venue category for “Shopping Mall” category.

	Neighborhoods	Shopping Mall
0	Agnes Banks, New South Wales	0.00000
1	Alexandria, New South Wales	0.01000
2	Allambie Heights, New South Wales	0.02000
3	Annandale, New South Wales	0.00000
4	Appin, New South Wales	0.09434

k-means clustering

k-means clustering identifies k number of centroids, and then assign every data point to the nearest cluster. In clustering, centroids are kept as small as possible. This method is one most popular unsupervised machine learning methods and is considered to be an efficient method for the current project. Here, neighbourhoods are clustered into three groups based on the frequency of occurrence of shopping malls. The results show the density of shopping centres in various neighbourhoods. Consequently, it would be possible to predict which neighbour would be possible for opening new shopping centre.

Results

Table 4. shows results of clustering for different neighbours in Sydney.

Table 4. Cluster groups of shopping malls for neighbours in Sydney

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
0	Agnes Banks, New South Wales	0.000000	0	-33.61445	150.71083
112	Kurraba Point	0.000000	0	-33.84249	151.22256
113	Kyeemagh, New South Wales	0.000000	0	-33.94986	151.16380
114	La Perouse, New South Wales	0.000000	0	-33.98795	151.23110
116	Lane Cove, New South Wales	0.014493	0	-33.81347	151.17017
...	...	...	...	...	...
160	Peakhurst, New South Wales	0.030303	2	-33.96257	151.05809
161	Penrith, New South Wales	0.024096	2	-33.75374	150.69820
20	Bexley, New South Wales	0.028169	2	-33.95089	151.12555
37	Campbelltown, New South Wales	0.050847	2	-34.06408	150.81658
139	Middle Cove, New South Wales	0.027778	2	-33.79250	151.20731

200 rows × 5 columns

The results from k-means clustering show three groups of clusters. Data in each cluster have similar density of shopping centres. Accordingly, the following labelling applies to the clusters.

- Cluster 0: Neighbours with moderate number of shopping centres
- Cluster 1: Neighbours with low number of shopping centres
- Cluster 2: Neighbours with high number of shopping centres

Based on these results a map of clusters was created. This map is shown in Figure 2. In this figure, Cluster 0 is shown with red markers, Cluster 1 in blue and Cluster 2 in light green colour.

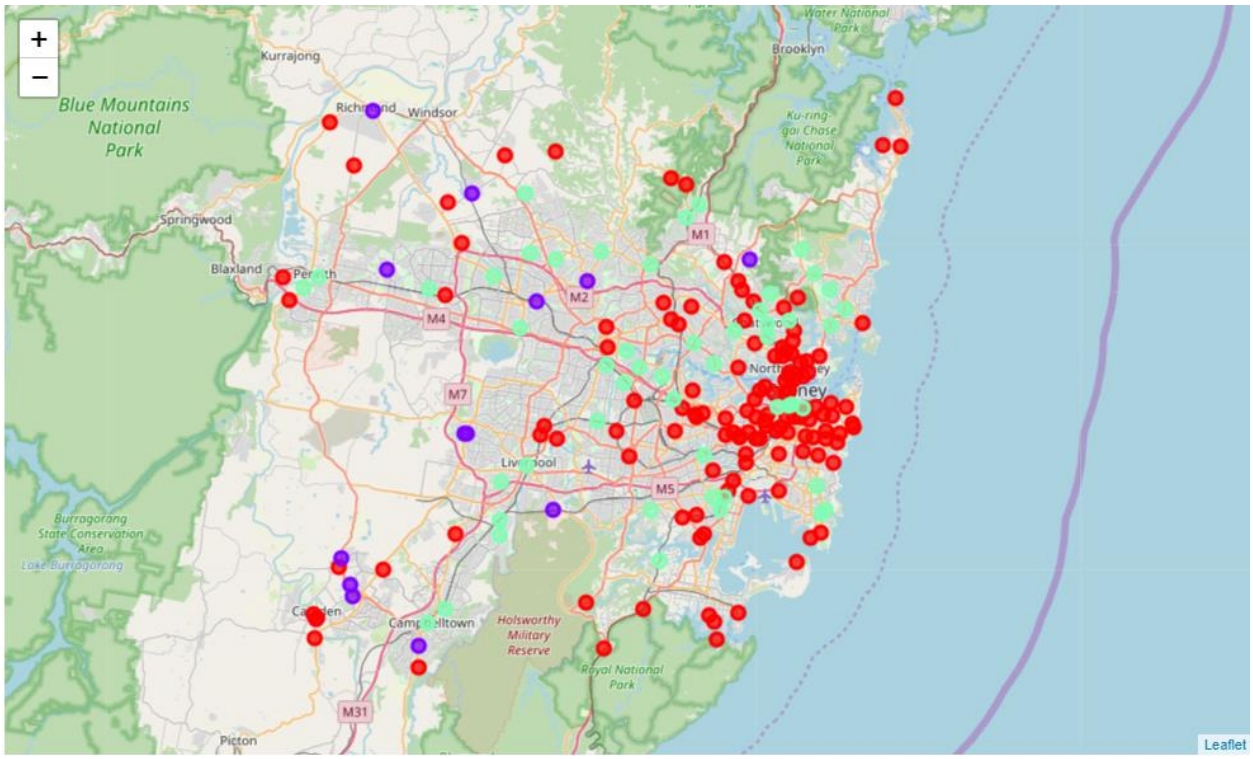

Figure 2. Map of clusters of shopping centres in Sydney. Data from [1-4].

As it can be seen from the figure, there are more shopping centres in the centre of Sydney. Red points show moderate number and green points show high concentration of malls in Sydney. From centre to west number of malls decreases and blue points are more which are evidence of potential for opening new shopping centres.

Discussion

Results of this study show that most of shopping malls are located in the central suburbs of Sydney. Surrounding suburbs have lower number of shopping malls which suggests that opening new shopping centres in these areas would have little or low competition. By a precise look at the map we find out that western part of Sydney has low number of malls.

Therefore, this study suggests that a new shopping mall in central suburbs will encounter high level of competition and it is recommended to open a new shopping mall far from the centre in western part of city. However, exact recommendation of new places for opening shopping malls includes other affecting parameters such as population, number of companies, types of industries and other factors which were out of focus of this study.

Conclusions

This project studies potential possibility of opening a new shopping centre in Sydney. Data for this study were obtained from different sources. Based on the results of this study, central parts of Sydney were found to be highly dense of shopping malls and as the distance from centre increases the competition for a new shopping mall decreases. Finally, it was suggested that western suburbs of Sydney have lower density of shopping malls and considered to have potential for opening a new mall.

References

1-https://en.wikipedia.org/wiki/Category:Suburbs_of_Sydney

2- https://pypi.org/project/geocoder/

3- https://pypi.org/project/geopy/

4- https://foursquare.com/