

Recommend most suitable neighbourhood for homebuyers

Bijay Kumar Gupta

29 July 2020

1. Introduction

a. Background

During different phases of life, a person sees many changes and accordingly they move around different locations within a city or to a different city. People move around based on their requirement of change in work location, to live in a neighbourhood which is better suited to their current and future needs or change in living preference. When someone goes from school to job or changes job location or family members needs change e.g. kids going to specific schools or family member needing some assistance or change in interest around their living space – they need to decide to move to a neighbourhood which would fit their new requirements. It takes a lot of time and effort in identifying the right location to move in.

b. Problem

When facing this situation of changing location it is quite a time consuming task to identify the best location that fits the requirement. Once a decision is made then reversing is often impossible. So, it is very important to make the best choice of location. Location of restaurants, gyms, schools, offices, hospitals, play-area, parks, bars/pubs etc. are some of the examples of facilities which one has to keep in mind while making the decision.

c. Service/Interest

We can offer a service to identify the best location for a home buyer/renter for them. Based on the places of interest within the vicinity of home, neighbourhoods can be classified in different categories and therefore applicable to that class of people. This kind of service would be of interest to home buyers/renter.

2. Data

a. Sources

I have decided to work on “New Orleans, Louisiana”. We will need list of all the neighbourhoods of the city along with their latitude and longitude. This data can be found on the Wikipedia page ([link](#)). Next, we need data of venues in a neighbourhood and their names & categories. We can use “Foursquare” API for this purpose. We need to pass the latitude and longitude data of a neighbourhood to the API and it provides the information of all the places around that location within a specified radius.

b. Description

From the source we get a list of 72 neighbourhoods in ‘New Orleans’ along with their latitude and longitude. Data is in tabular format with 3 columns as shown below in figure1.

	Neighborhood	Longitude	Latitude
0	U.S. NAVAL BASE	-90.026093	29.946085
1	ALGIERS POINT	-90.051606	29.952462
2	WHITNEY	-90.042357	29.947200
3	AUDUBON	-90.121450	29.932994
4	OLD AURORA	-90.000000	29.924440

Figure1. Neighbourhoods in New Orleans

Foursquare gives data on venues around a location. I have used venue names and category for the purpose of assignment, there are many other fields available. A snapshot of data from Foursquare API is shown in figure2, it displays few places with names and categories within the vicinity of 500 meters of two of the neighbourhood.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	U.S. NAVAL BASE	29.946085	-90.026093	Richards Bar	29.943784	-90.022193	Bar
1	U.S. NAVAL BASE	29.946085	-90.026093	The Mighty Mississippi	29.949695	-90.023710	Boat or Ferry
2	ALGIERS POINT	29.952462	-90.051606	Congregation Coffee Roasters	29.951918	-90.053395	Coffee Shop
3	ALGIERS POINT	29.952462	-90.051606	Tout de Suite Café	29.952121	-90.051090	Café
4	ALGIERS POINT	29.952462	-90.051606	The Crown & Anchor	29.951416	-90.054220	Bar

Figure2. Places around neighbourhoods “US NAVAL BASE” & “ALGIERS POINT”

c. How will this data help in providing the service

Using data of venues in a neighbourhood we can classify a neighbourhood. With this approach all the neighbourhoods in the city can be categorized, which will be used to help people make the selection of neighbourhood.

d. Data Cleaning/Feature selection

We checked if there was any missing or unusual value in the data that consists of neighbourhood names and their latitude and longitude. Below figure shows there was no null in the scrapped data from wiki.

```
: print(df_wiki['Neighborhood'].isnull().sum(),df_wiki['Latitude'].isnull().sum(),df_wiki['Longitude'].isnull().sum())
orleans_df = df_wiki.copy()
0 0 0
```

Figure3. Check null values for scrapped data

We get a json file from Foursquare API. All information about a venue is in item 'keys', we extracted values of item. From this we further extracted values of venue name, category, latitude and longitude from 'keys' item for our use. In total there were 2703 venues were read from Foursquare for all 72 neighbourhoods, while pulling data we restricted to maximum of 100 venues per neighbourhood and to return venues within 1000 meters. However, we got less than 100 venues per neighbourhood because Foursquare has only these many venues.

Both the data frames are merged as shown in Figure2.

3. Exploratory Data Analysis

a. Plotting all neighbourhoods

Below is a plot of all the neighbourhoods with blue circle in city of New Orleans.



Figure4. Map all neighbourhoods in New Orleans

b. Venues around neighbourhood

There are total 2703 venues for all 72 neighbourhoods. Out of which there are 276 unique venues by category. Below is snapshot of count of venues returned for each neighbourhood.

There are 276 unique categories.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
ALGIERS POINT	27	27	27	27	27	27
AUDUBON	28	28	28	28	28	28
B. W. COOPER	31	31	31	31	31	31
BAYOU ST. JOHN	69	69	69	69	69	69
BEHRMAN	16	16	16	16	16	16
BLACK PEARL	56	56	56	56	56	56
BROADMOOR	53	53	53	53	53	53
BYWATER	36	36	36	36	36	36
CENTRAL BUSINESS DISTRICT	100	100	100	100	100	100
CENTRAL CITY	41	41	41	41	41	41
CITY PARK	20	20	20	20	20	20
DESIRE AREA	2	2	2	2	2	2
DILLARD	12	12	12	12	12	12
DIXON	37	37	37	37	37	37
EAST CARROLLTON	92	92	92	92	92	92

Figure5. List of neighbourhoods with count of venues

c. Realign and transform data

After we have pulled data from Foursquare, we need to convert it in a shape that can we used further. We realign data with rows of neighbourhoods and columns of venues, so that we can make comparison based on venues for each neighbourhood. We have grouped rows by neighbourhood. Each data point corresponding to a neighbourhood and venue is calculated as mean of frequency of occurrence of category. Figure is a snapshot of the final data frame that will be used for modelling.

	Neighborhood	Yoga Studio	Accessories Store	African Restaurant	American Restaurant	Antique Shop	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	...
0	ALGIERS POINT	0.000000	0.000	0.0000	0.000000	0.000000	0.0	0.000000	0.000000	0.037037	...
1	AUDUBON	0.000000	0.000	0.0000	0.000000	0.000000	0.0	0.035714	0.000000	0.000000	...
2	B. W. COOPER	0.000000	0.000	0.0000	0.000000	0.000000	0.0	0.032258	0.000000	0.000000	...
3	BAYOU ST. JOHN	0.000000	0.000	0.0000	0.028988	0.000000	0.0	0.000000	0.000000	0.000000	...
4	BEHRMAN	0.000000	0.000	0.0000	0.000000	0.000000	0.0	0.000000	0.062500	0.000000	...
5	BLACK PEARL	0.000000	0.000	0.0000	0.000000	0.000000	0.0	0.017857	0.000000	0.000000	...
6	BROADMOOR	0.000000	0.000	0.0000	0.000000	0.000000	0.0	0.000000	0.000000	0.018888	...
7	BYWATER	0.000000	0.000	0.0000	0.000000	0.027778	0.0	0.027778	0.000000	0.027778	...

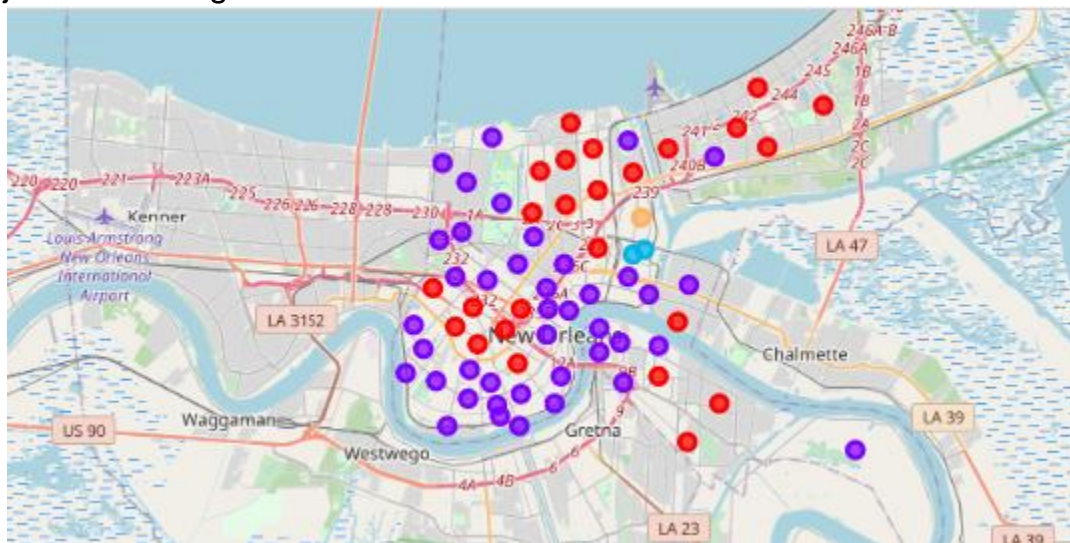
Figure6. List of neighbourhoods with average occurrence of venues

4. Modelling

We have to cluster each neighbourhood based on the venue category. We are using k-means clustering method to classify all neighbourhoods. We have used 5 clusters, i.e. classify all neighbourhoods into 5 cluster. K-means clustering is an unsupervised clustering methods. Since we do not have any cluster pre defined we have to use un-supervised method.

5. Result visualization

After running k-means clustering we have got the classified all neighbourhoods in 5 clusters. To visualize all neighbourhoods we have joined the neighbourhood data frame with classification values.



6. Discussion

After running the k-means clustering machine learning algorithm we have segregated all neighbourhoods in 5 different categories. I am naming all five categories as follows, based on venues in the vicinity:

- "Family", few examples of venues around are – restaurants, café, bars, stores, malls.
- "Youngster" few venues around are – coffee shops, hotels, sandwich place, bars, café, college buildings, dance studio etc.
- "Leisure", few venues around are places to hangout
- "Country" few places are farms and farmers market
- "Open" because of lot of open spaces like parks and farms.

7. Conclusion

In this project I have classified all neighbourhoods in New Orleans based on the venues in the vicinity. I found that majority of neighbourhoods can be classified into 2 categories. Other 3 categories have few neighbourhood and therefore are very exclusive. The two categories of neighbourhood are named "" and "". A person who is

moving to a different location within New Orleans can use this report to make a choice of appropriate neighbourhood. For example, pick the neighbourhood which is in same category of their current neighbourhood, which will place them in different location but similar neighbourhood.