



FINDING THE BEST LOCALITY BASED ON INTEREST

Abstract

Day to day life, people migrate from one location to another. In This project, we will see how data can help then find a place of their interest.

THIRUKUMARAN VASANTHA THANANJAYAN

20-Nov-2020

Contents

1. Introduction	2
1.1 Background.....	2
1.2 Problem	2
1.3 Interest.....	2
2. Data	2
3. Methodology	2
4. Result.....	5
5. Discussion.....	6
6. Conclusion	6
7. Reference	6

1. Introduction

1.1 Background

People migrate from one location to another. It is part of our life. Most people would have migrated from one location to another at least once in their life. It may be due to job change, looking for the school for children or moving to another rented house.

1.2 Problem

It will be challenging to find the place of your interest. Everyone will have different demand and interest. In this article we will see how we use publicly available data to find solution.

1.3 Interest

Anyone migrating from one location to another will be interested in understanding the place they are moving to.

2. Data

Based on the problem we are looking for area of interest. This information, we can get it from Foursquare [\[1\]](#) APIs.

And other information we need is, how do we split the location. There are multiple ways we can do this:

- Divide the area into small grids
- Divide by county
- Divide by zip code

I went for the last option, based on zip code. For this, I needed the zip code details and found www.unitedstateszipcodes.org [\[2\]](#). They have a free database (for non-commercial use) with all USA zip codes and their location co-ordinates.

3. Methodology

In my work, I used SQL databases like MySQL, MariaDB. However, for this project, I tried the python dataframes for the first time. It was a new experience.

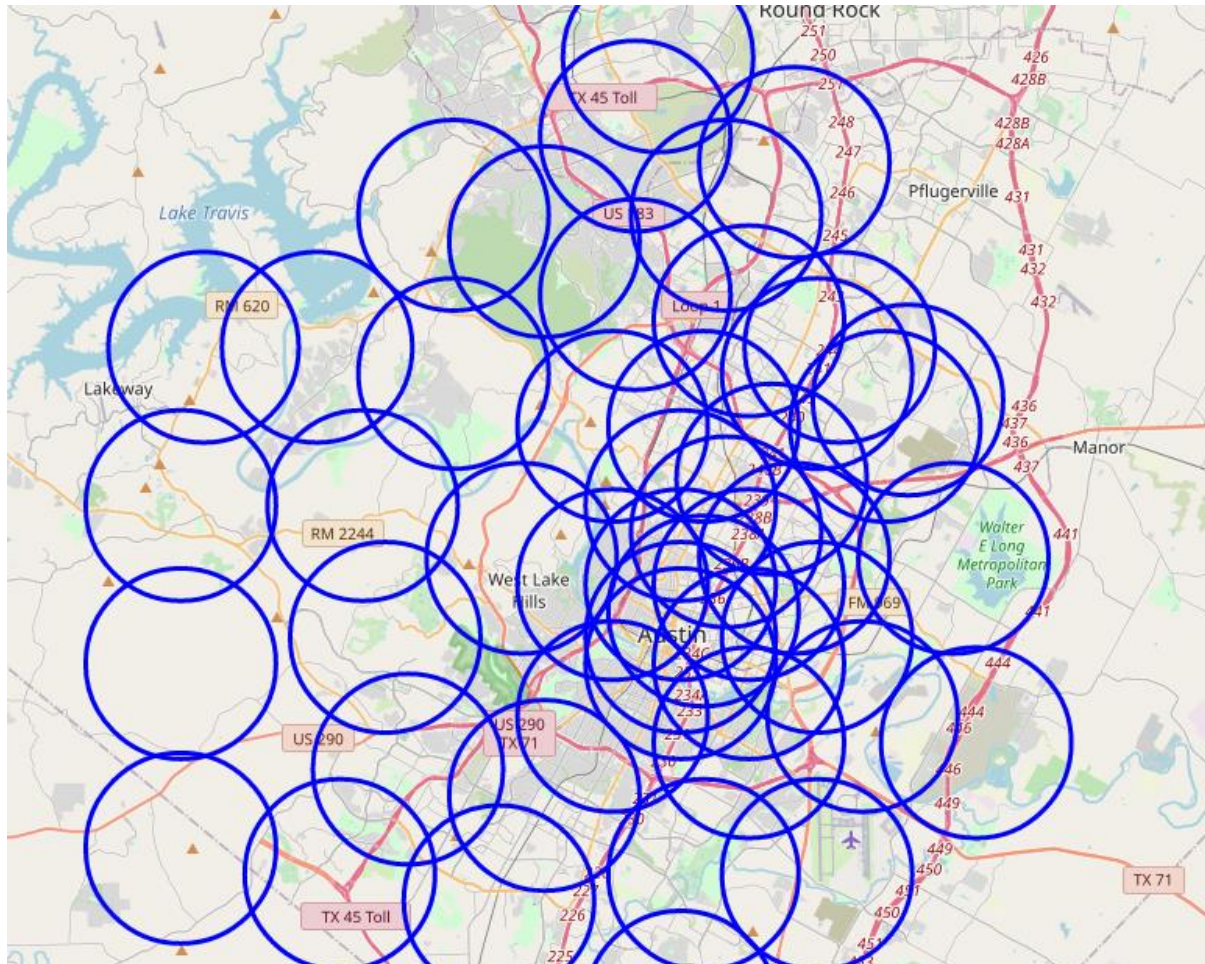
I filtered only zip code, city, state and co-ordinate information.

	zip	primary_city	state	latitude	longitude
0	501	Holtsville	NY	40.81	-73.04
1	544	Holtsville	NY	40.81	-73.04
2	601	Adjuntas	PR	18.16	-66.72
3	602	Aguada	PR	18.38	-67.18
4	603	Aguadilla	PR	18.43	-67.15

Here are the user inputs I have considered: (New) Location, interested venues.

For this example, I have taken Austin,TX and “Gym, Park, Indian Restaurant, Bar” as interested venues.

First, I filtered the zip code for Austin and visualized the zips with the area of diameter 4km using python folium library.



This showed good amount of coverage. It has high density in the centre of the town. For this problem that is ok.

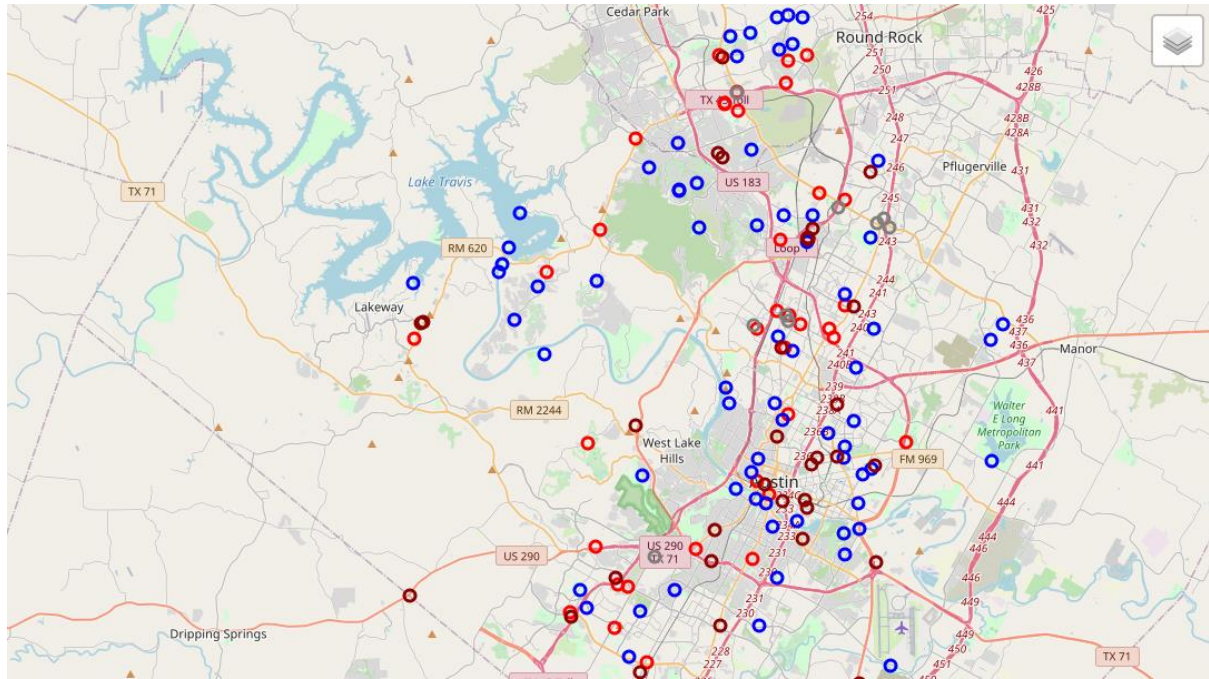
Next, I collected all the nearby places using Foursquare APIs and converted into a table.

	zip	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	73301	30.26	-97.74	Ladybird Lake Hike & Bike Trail (Rainey St.)	30.258917	-97.740925	Trail
1	73301	30.26	-97.74	Kimpton Hotel Van Zandt	30.260078	-97.739147	Hotel
2	73301	30.26	-97.74	Craft Pride	30.257896	-97.738862	Beer Garden
3	73301	30.26	-97.74	Rainey Street Outdoor Food Trucks	30.258436	-97.738968	Food Truck
4	73301	30.26	-97.74	Fairmont Austin	30.262074	-97.738261	Hotel

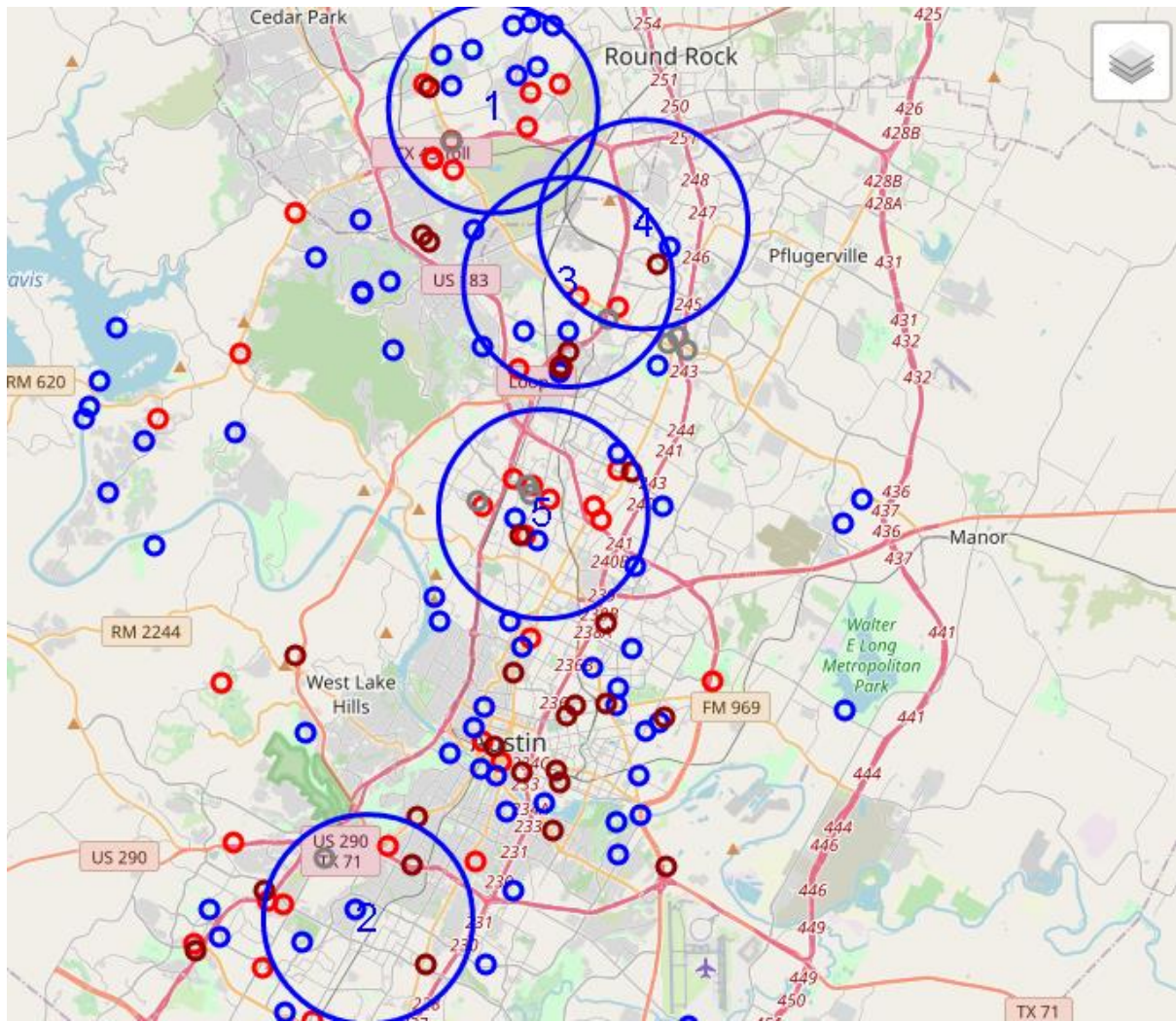
Then I filtered only the user’s interesting venues.

	zip	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	73301	30.26	-97.74	Lustre Pearl	30.260715	-97.738281	Bar
1	73301	30.26	-97.74	Auditorium Shores at Lady Bird Lake	30.259792	-97.749013	Park
2	73301	30.26	-97.74	RIDE Indoor Cycling	30.264560	-97.746643	Gym
3	73301	30.26	-97.74	Butler Park (formerly Town Lake Park)	30.261792	-97.754974	Park
4	73301	30.26	-97.74	Rustic Tap	30.269533	-97.749145	Bar

Again, utilized folium to visualize how these places are spread across.



Then grouped the area based on the coverage of all/most of the venues the user is interested in.



We found the area that the user can start looking before he/she moved to the new location.

4. Result

We found the area(s) which may match the user's preference. These are the zip codes the user can start with.

zip	Latitude	Longitude	Gym	Park	Indian Restaurant	Bar
78717	30.49	-97.75	True	True	True	True
78745	30.21	-97.80	True	True	True	True
78727	30.43	-97.72	True	True	True	True
78728	30.45	-97.69	True	True	True	True
78757	30.35	-97.73	True	True	True	True
78753	30.38	-97.67	True	True	True	True
78731	30.35	-97.77	True	True	True	True
78716	30.26	-97.74	True	True	False	True
78715	30.26	-97.74	True	True	False	True

5. Discussion

In this model, I mostly used the visualization methods to identify the solution. We can solve this used multiple ways as I discussed in the Data section. This show how we can utilize the data science to solve the day to day problems. One additional improvement we could do is, we can group some of the zip codes using DBSCAN algorithm. After seeing the result, I felt that I don't need to go for that option.

6. Conclusion

As more people are migrating to new places and this model will help them to identify the places they can start looking before they move in. Again, there is always place for improvements

7. Reference

[1] [Foursquare API](#)

[2] www.unitedstateszipcodes.org