Finding the Ideal Neighborhood for New Yorkers Moving to Austin, Texas

Jeremy Candelas

Introduction

Background

For my capstone project, I would like to evaluate a problem which is particularly relevant to me. For one summer, I interned in New York City for IBM's marketing department. While living in New York, I spent most of my time in Gramercy Park and the Flatiron District which were close to where I was living at the time. The office where I worked itself was in-between Union Square and Washington Square Park, in Greenwich Village.

While I enjoyed the New York City neighborhoods I stayed in quite a lot, I did not decide to move to New York City. Instead, I had asked to work at the offices in Austin, Texas as I am originally from the area. Since starting at the Austin office nearly a year ago, I have seen the office experience significant growth. Likewise, my team/organization at IBM has continued to grow, although predominately in the NYC offices where they are quickly running out of room.

Problem

At this point, it is hard for visiting employees to find a seat, and even harder to find a quiet meeting room in the NYC office. To maintain its current levels of growth, either IBM will need to find new office space, or hire more workers in cities like Austin where there is still plenty of room to grow. The question then is, how to convince a New Yorker to move to Austin?

One possible solution would be to identify neighborhoods in Austin which are much like New York City. While IBM Marketing employees at the NYC offices certainly live in all different parts of New York (and even other states such as New Jersey), I will base my project off of a few assumptions due to time limitations for research in regard to this course.

The first assumption is that these new hires are interns being extended a return offer under the condition they will work in the Austin office. Second, we will assume that they spent most of their time in the three aforementioned neighborhoods (Gramercy Park, the Flatiron District, and Greenwich Village) just as I did when I was an intern, and as the majority of my intern class did.

With these assumptions in mind, I will answer the question: Should IBM begin to place more people in their Austin office, which neighborhoods would their new hire class find to be be most similar to the three aforementioned neighborhoods in New York City?

Data to Be Used

Sources of Data

My first source of data will the list of neighborhoods in Austin, sourced from Wikipedia. From there, I will collect the GPS coordinates of each neighborhood listen, again, from Wikipedia. While the vast majority of the Austin neighborhoods listed on Wikipedia have GPS coordinates on their page, there are a few that do not. These mostly appear to be "unofficial" neighborhoods, or subsections of existing neighborhoods. For example, the Chinatown "neighborhood" is merely a street in Austin, and so it would seem acceptable to drop such neighborhoods from my data for the sake of simplicity and data cleansing.

Once this information is obtained, I will then convert the GPS coordinates into Latitude and Longitude values in order that I will be able to use them in conjunction with the FourSquare API. From the API, I plan to obtain a list of the most popular venues within each neighborhood in Austin.

I will compare this data to the data which we obtained in the Segmenting and Clustering Neighborhoods in NYC lab, that we did in the capstone course. This data will be brought in for analysis against the data I have obtained for Austin's neighborhoods.

Analysis

After receiving the venue information, I will then segment and cluster the neighborhoods of Austin (there are nearly 60) to identify what they have access to and are known for. With this information clear, I will then compare my cluster to the clusters generated in our prior analysis of neighborhoods in New York City.

Using these key data points, we will be able to identify the defining characteristics of the NYC neighborhoods I've mentioned and how they compare with the Austin neighborhoods. With this information at hand, we would then be able to identify just how similar they are to the New York City neighborhoods and pinpoint a cluster of neighborhoods that are sufficiently similar to that of New York City.

Methodology

Data is obtained from Wikipedia concerning the neighborhoods in Austin and their GPS coordinates, which are then converted to their Latitude and Longitude values using a function that was suggested by a user, *fraxel*, on the website StackOverflow.com. Wikipedia identifies some 58 neighborhoods in Austin, of which there are nine without GPS coordinates on their

Wikipedia pages. As such, this information is dropped from the dataframe, leaving 49 neighborhoods to be evaluated.

	index	wikiUrl	URL	Latitude	Longitude
0	Allandale	https://en.wikipedia.org/wiki/Allandale,_Austi...	https://en.wikipedia.org/wiki/Allandale,_Austi...	30.341944444444444	-97.74388888888889
1	Anderson Mill, Texas	https://en.wikipedia.org/wiki/Anderson_Mill,_T...	https://en.wikipedia.org/wiki/Anderson_Mill,_T...	30.455	-97.80916666666667
2	Barton Hills	https://en.wikipedia.org/wiki/Barton_Hills,_Au...	https://en.wikipedia.org/wiki/Barton_Hills,_Au...	30.258055555555554	-97.78111111111112
3	Bouldin Creek	https://en.wikipedia.org/wiki/Bouldin_Creek,_A...	https://en.wikipedia.org/wiki/Bouldin_Creek,_A...	30.249444444444446	-97.75527777777778
4	Bremond Block Historic District (Austin, Texas)	https://en.wikipedia.org/wiki/Bremond_Block_Hi...	https://en.wikipedia.org/wiki/Bremond_Block_Hi...	30.270555555555553	-97.74638888888889
5	Brentwood	https://en.wikipedia.org/wiki/Brentwood,_Austi...	https://en.wikipedia.org/wiki/Brentwood,_Austi...	30.34	-97.73083333333334
6	Bryker Woods	https://en.wikipedia.org/wiki/Bryker_Woods,_Au...	https://en.wikipedia.org/wiki/Bryker_Woods,_Au...	30.304722222222225	-97.75166666666667
7	Canyon Creek	https://en.wikipedia.org/wiki/Canyon_Creek,_Au...	https://en.wikipedia.org/wiki/Canyon_Creek,_Au...	32.39111111111111	-97.73944444444444
8	Chinatown, Austin	https://en.wikipedia.org/wiki/Chinatown,_Austin	NaN	NaN	NaN
9	Circle C Ranch	https://en.wikipedia.org/wiki/Circle_C_Ranch	https://en.wikipedia.org/wiki/Circle_C_Ranch	30.1875	-97.89083333333333
10	Clarksville Historic District (Austin, Texas)	https://en.wikipedia.org/wiki/Clarksville_Hist...	https://en.wikipedia.org/wiki/Clarksville_Hist...	30.280833333333334	-97.76222222222222
11	Congress Avenue Historic District	https://en.wikipedia.org/wiki/Congress_Avenue_...	https://en.wikipedia.org/wiki/Congress_Avenue_...	30.267777777777777	-97.7425
12	Copperfield	https://en.wikipedia.org/wiki/Copperfield,_Aus...	NaN	NaN	NaN
13	Crestview	https://en.wikipedia.org/wiki/Crestview,_Austi...	https://en.wikipedia.org/wiki/Crestview,_Austi...	30.34	-97.73083333333334
14	Dawson	https://en.wikipedia.org/wiki/Dawson,_Austin,_...	NaN	NaN	NaN
15	Dove Springs	https://en.wikipedia.org/wiki/Dove_Springs,_Au...	NaN	NaN	NaN

Figure 1: Dataframe of Neighborhoods and their Latitude and Longitude Values

Using the Folium library, I then plot these neighborhoods against a map of Austin, Texas as shown below.

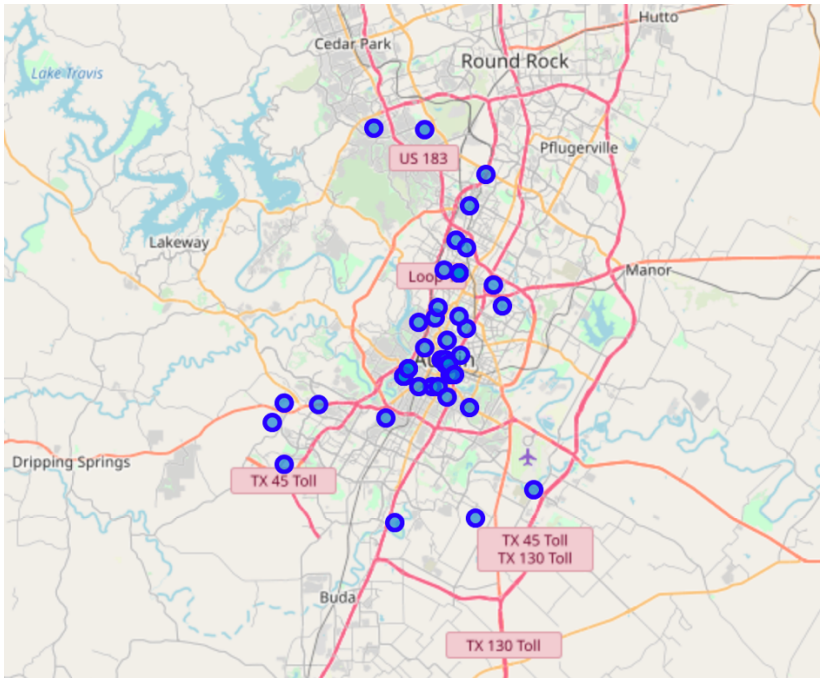

Figure 2: A Folium Map of the Neighborhoods in Austin, Texas

At this point, a connection to the FourSquare API is established. The venues for each neighborhood are requested and stored in a dataframe. A sample call is made to the Allandale neighborhood to ensure the information is being received correctly. In this example of the df.head(), we are shown only the first five rows of results of the dataframe.

```
In [18]: print(atx_venues.shape)
         atx_venues.head()

         (1004, 7)
```

Out[18]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Allandale	30.341944444444444	-97.74388888888889	Yard Bar	30.342881	-97.738871	Bar
1	Allandale	30.341944444444444	-97.74388888888889	Taqueria Arandas No. 3	30.341555	-97.739067	Mexican Restaurant
2	Allandale	30.341944444444444	-97.74388888888889	Hot Rod Coffee Trailer	30.341761	-97.738946	Food Truck
3	Allandale	30.341944444444444	-97.74388888888889	U-Haul Moving & Storage at Burnet Rd & Fm 2222	30.342257	-97.738979	Storage Facility
4	Allandale	30.341944444444444	-97.74388888888889	Marisco Grill	30.340751	-97.739250	Mexican Restaurant

Figure 3: The head of the venues in Allandale, their latitude/longitude, and each venue category

Satisfied with this result, we then identify the number of venue categories that were pulled from the API, to which we see it is 202 unique venue categories.

Through one-hot encoding, we prepare our neighborhoods to be analyzed based upon these venue identifications. We then group the venue classifications in order to identify the top 10 most popular venues for each location.

Results

Through K-Means clustering, we use a range of 1-10 to identify the best number of clusters to use in our analysis. We see that this number is five and proceed our analysis with those five clusters. The clusters are generated as such:

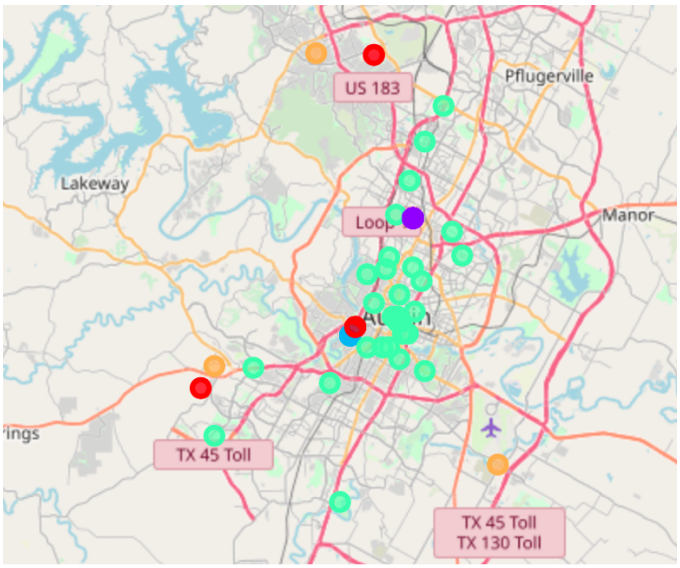

Figure 4: A Folium Map of the K-Means Clusters (n=5)

We find that:

- Cluster 0 is defined by parks and trails, and seems to have high access to public transport with bus stops.
- Cluster 1 is defined by trails, pools, parks and yoga. It does not seem to have many bus or train stops.
- Cluster 2 is defined by the outdoors as well, and appears to be a scenic location--again though, little public transport.
- Cluster 3 is the largest cluster, and it seems to be clustered by its coffee shops, bars and restaurants.
- Cluster 4 is defined by parks and playgrounds.

The vast majority of these clusters do not have many options for public transport. By and large, Austin seems to be well-known for its outdoor spaces, something which sets it highly apart from NY as a whole. While NY does have some degree of outdoor spaces, it is limited and nowhere near the size of Austin's.

Data from the Manhattan Segmenting and Clustering assignment is brought in to compare against. From here, the top venues for the 5 clusters identified in the previous assignment are brought in and evaluated.

Discussion

Based upon the introduction to this report, I mention that the neighborhoods I most want to compare to are Gramercy, Flatiron and Greenwich Village. Gramercy is in cluster 1, Flatiron and Greenwich are both in 2. In addition to Gramercy, cluster 1 also contains Murray Hill, which is where my temporary home was technically-located, and the East Village where many of us would also go to hang out. And so, these two clusters are the ones I will compare to Austin's neighborhood clusters.

These are most well-defined by:

- Cluster 1: Coffee Shops, Bars, and a variety of regional cuisines.
- Cluster 2: Theatres, Gyms, and Italian Restaurants.

Comparing these to Austin, Manhattan-Cluster-1 is most similar to Austin-Cluster-3. Both have a large number of coffee shops, bars, and a variety of regional cuisines. However, a fan of New York's theatres in Manhattan-Cluster-2 may be more drawn to Austin-Cluster-0 which has a large number of theatres. While it does not have quite the same prevalence of gyms, it should be

noted that it is well known for its outdoor spaces, which could be beneficial for New Yorkers for whom fitness is very important.

Conclusion

It can be seen that while the two cities are vastly different, they do share in some common features. Both cities are technological hubs which have made them highly attractive to employers, and the offerings of the two cities have made them highly attractive to employees alike. Should IBM need to expand its presence in Austin to account for the lack of space in New York City, a number of IBM-ers should be well-pleased to find that they will be able to enjoy in many of the same highlights of New York City, especially those of the neighborhoods I mentioned previously—Gramercy, the Flatiron District, and Greenwich Village.

A New York-to-Austin transplant will be able to find themselves surrounded by a variety of regional/cultural cuisine, and an abundance of bars and coffee shops to keep themselves occupied. While the number of gyms is not as high in Austin, there is a beneficial tradeoff considering the number of outdoor spaces in the form of parks and trails for future Austinites to keep fit, and a number of theatres present for those who are fans of the arts. All-in-all, Austin would not be so difficult for a New Yorker to adjust to, and is a highly viable option for a company like IBM to promote to future employees to consider working in.