IBM Data Science Certificate Capstone Project report: Battle of Texas city living for young adults

By

Oladapo Akinyede

October 30, 2019

Introduction

The Objective of this project is to determine the best city that provides conveniences for young adults in Texas. The major cities being considered are Austin, Dallas, Houston and San Antonio. The criteria to be considered are venues for restaurants, activities like yoga, parks and other outdoor activities. The analysis will provide data of venues within the cities, that can best inform young adults, based on their personal preferences the city of choice to reside.

Data and Methodology

The data to be used is the coordinates from the center of downtown and the venues to be considered are data from Foursquare. A 3.1 mile (5Km) radius from the center of downtown will be considered for the analysis. The four clusters will look at type, location venues and accessibility of these conveniences being considered for young adults. A report will analyze using metrics and plots to compare and discuss pro and cons of these four cities being considered. The data analysis involves looking at data that are relevant to venues of interest.

The distance data will be obtained from a table input, since are only analyzing four geographical coordinates.

The data set for the geographical coordinates is below;

TX_City	Latitude	Longitude
Austin	30.26715	-97.743057
Downtown Houston	29.75187	-95.327438
Downtown Dallas	32.77667	-96.796989
Downtown San Antonio	29.42412	-98.493629

The venues for the four cities in Texas will be obtained from Foursquare. During the data analysis stage, graphical plots and data reduction will be used. Venues that are not relevant to daily living will be removed. A good example of venue to be removed from the dataset is the Hotel. The data set can then be analyzed for types of venues available within these cities and numbers of option available for the venues of interest. The geographical spread of these venues are then further analyzed to provide more insight for potential young adults that are looking to relocate to a major metropolitan city in Texas. A sample of the dataframe capture from Foursquare used for the analysis is below:

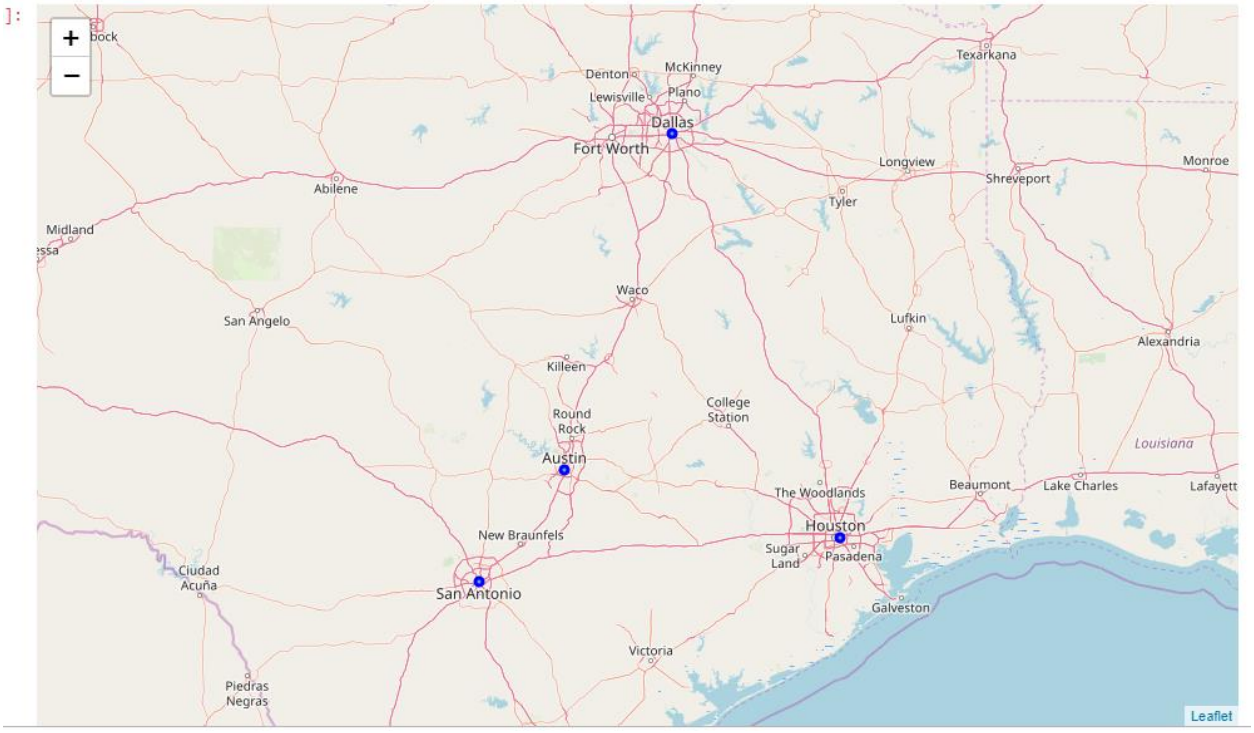
```
nbh_venues=nbh_venues[nbh_venues['Venue Category'] !='Hotel'] #Look through column to remove hotels
nbh_venues.head()
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Venue Type
0	Austin	30.267153	-97.743057	JuiceLand	30.266037	-97.742623	Juice Bar	Juice Bar
1	Austin	30.267153	-97.743057	The Roosevelt Room	30.267842	-97.746242	Bar	Bar
2	Austin	30.267153	-97.743057	Houndstooth Coffee	30.266194	-97.743025	Coffee Shop	Coffee Shop
3	Austin	30.267153	-97.743057	Eddie V's Prime Seafood	30.266339	-97.740504	Seafood Restaurant	Seafood Restaurant
4	Austin	30.267153	-97.743057	Alamo Drafthouse Cinema	30.267460	-97.739550	Movie Theater	Movie Theater

The dataset consist of the neighborhood venue name and geographical location/ coordinates. There were 100 locations captured for each city/ Neighborhood. The grouping of the location count is below;

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Austin	100	100	100	100	100	100
Downtown Dallas	100	100	100	100	100	100
Downtown Houston	100	100	100	100	100	100
Downtown San Antonio	100	100	100	100	100	100

The plot of the cities are as follows;



Results

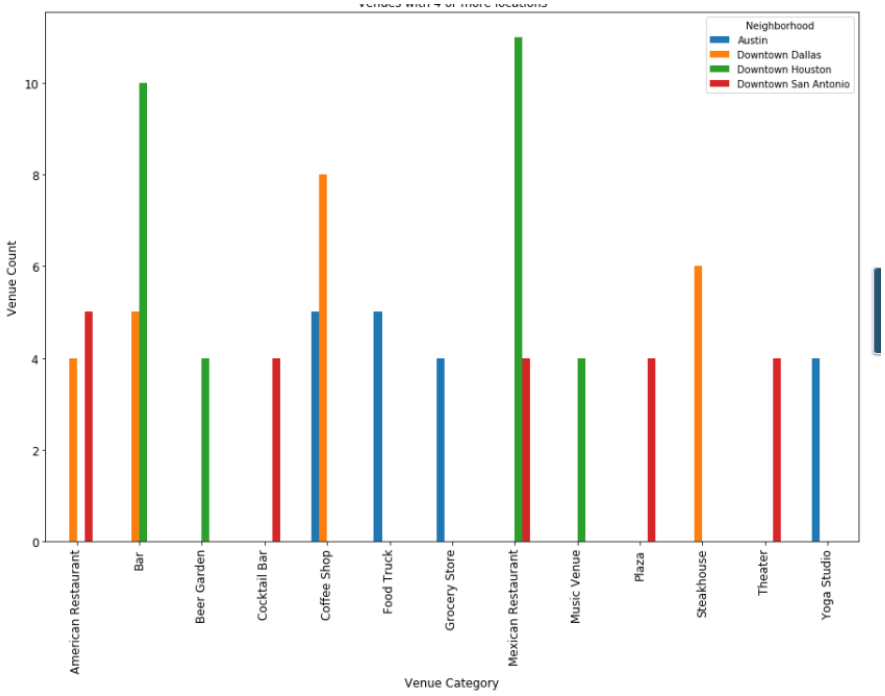
The venues data was the analyzed to obtain the number of venues within 3.1 miles of each city/ neighborhood of interest. The venues count data showed that hotels have the highest number of venues within the city. Since this is not a typical location of interest for a day to day living, the hotels were deleted from the dataset as part of the data cleaning process.

The data was further filtered into 2 dataset with venues counts of greater than 2 and greater than 3 locations. The data was rendered in a pivot format for easy review and plotting. The dataset for the 3 or more location is shown below;

Neighborhood	Austin	Downtown Dallas	Downtown Houston	Downtown San Antonio
Venue Category				
American Restaurant	0.0	4.0	0.0	5.0
Bar	0.0	5.0	10.0	0.0
Beer Garden	0.0	0.0	4.0	0.0
Cocktail Bar	0.0	0.0	0.0	4.0
Coffee Shop	5.0	7.0	0.0	0.0
Food Truck	5.0	0.0	0.0	0.0
Mexican Restaurant	0.0	0.0	9.0	4.0
Music Venue	0.0	0.0	4.0	0.0
Plaza	0.0	0.0	0.0	4.0
Steakhouse	0.0	7.0	0.0	0.0
Theater	0.0	0.0	0.0	4.0
Yoga Studio	4.0	0.0	0.0	0.0

From the data set we can see certain venues that are unique to specific neighborhood, while some are common to these neighborhoods. For example, there are unique venues for Food Trucks and Yoga studios in Austin, Steakhouse in Dallas, Bee garden and Music Venue in Houston and Theater is San Antonio. This does not tell the whole picture since if we increase the dataset to greater than 2 venues, the unique locations might not exists for these venues, however for those young adults who highly prioritize these venues, this data set will provide insight for their selection. A good example is for someone who uniquely like the night life of beer gardens and music, Houston will be a very good choice for such a young adult.

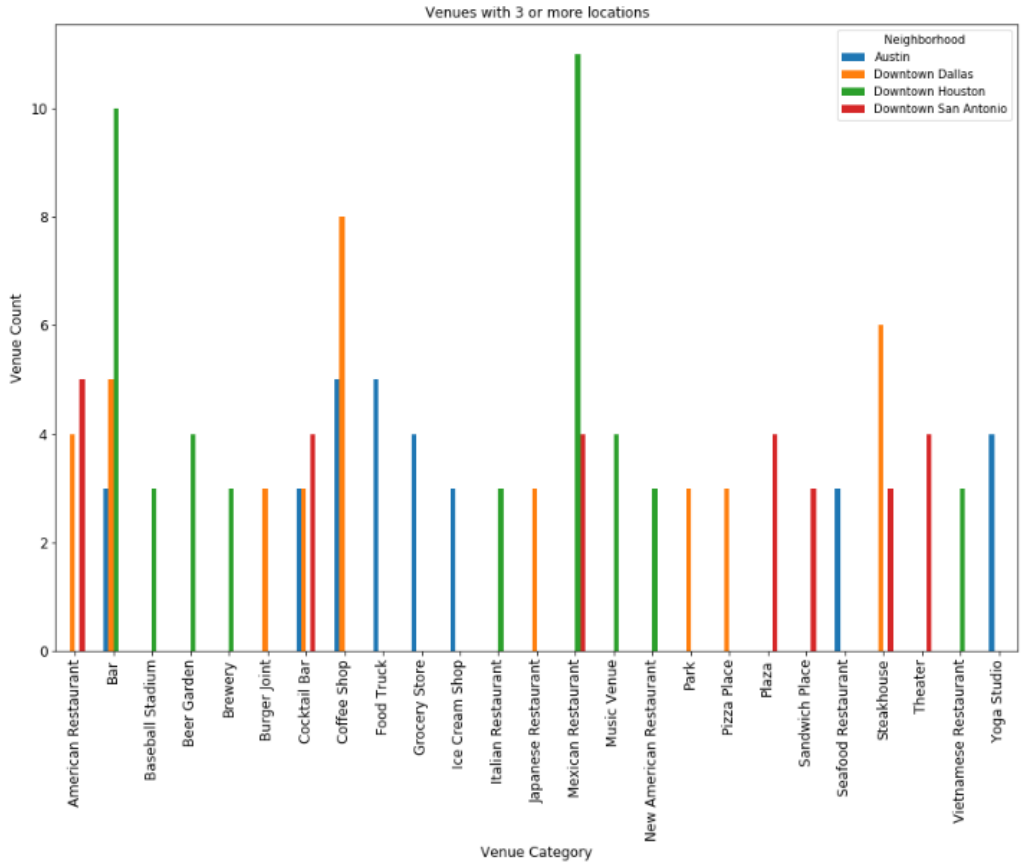
A plot of the pivot data showing a better visual of the table above is shown below;



Increasing the data set to 3 or more venues will produce the table below;

Neighborhood Venue Category	Austin	Downtown Dallas	Downtown Houston	Downtown San Antonio
American Restaurant	0.0	4.0	0.0	5.0
Bar	3.0	5.0	10.0	0.0
Baseball Stadium	0.0	0.0	3.0	0.0
Beer Garden	0.0	0.0	4.0	0.0
Brewery	0.0	0.0	3.0	0.0
Burger Joint	0.0	3.0	0.0	0.0
Cocktail Bar	3.0	3.0	0.0	4.0
Coffee Shop	5.0	8.0	0.0	0.0
Food Truck	5.0	0.0	0.0	0.0
Grocery Store	4.0	0.0	0.0	0.0
Ice Cream Shop	3.0	0.0	0.0	0.0
Italian Restaurant	0.0	0.0	3.0	0.0
Japanese Restaurant	0.0	3.0	0.0	0.0
Mexican Restaurant	0.0	0.0	11.0	4.0
Music Venue	0.0	0.0	4.0	0.0
New American Restaurant	0.0	0.0	3.0	0.0
Park	0.0	3.0	0.0	0.0
Pizza Place	0.0	3.0	0.0	0.0
Plaza	0.0	0.0	0.0	4.0
Sandwich Place	0.0	0.0	0.0	3.0
Seafood Restaurant	3.0	0.0	0.0	0.0
Steakhouse	0.0	6.0	0.0	3.0
Theater	0.0	0.0	0.0	4.0
Vietnamese Restaurant	0.0	0.0	3.0	0.0
Yoga Studio	4.0	0.0	0.0	0.0

The plot for the 3 or more locations is shown below;



From the plots, Houston seem to have added more unique venues which includes baseball stadium, Italian restaurant, American and Vietnamese restaurant. Austin Added Ice Cream shop, Seafood restaurant and Grocery store Dallas added a park as unique venue. Interestingly, no venue shared commonality between all neighborhoods except we lump Cocktail bar and bar into the same category.

To explore how these venues are popular we also listed the frequency of visit. The data is presented below for the top 10 visited venues;

Neighborhood	Austin	Downtown Dallas	Downtown Houston	Downtown San Antonio
1st Most Common Venue	Food Truck	Coffee Shop	Mexican Restaurant	American Restaurant
2nd Most Common Venue	Coffee Shop	Steakhouse	Bar	Theater
3rd Most Common Venue	Yoga Studio	Bar	Beer Garden	Plaza
4th Most Common Venue	Grocery Store	American Restaurant	Music Venue	Mexican Restaurant
5th Most Common Venue	Cocktail Bar	Park	Vietnamese Restaurant	Cocktail Bar
6th Most Common Venue	Ice Cream Shop	Japanese Restaurant	Brewery	Sandwich Place
7th Most Common Venue	Seafood Restaurant	Pizza Place	New American Restaurant	Steakhouse
8th Most Common Venue	Bar	Burger Joint	Baseball Stadium	Museum
9th Most Common Venue	Capitol Building	Cocktail Bar	Italian Restaurant	Restaurant
10th Most Common Venue	Burger Joint	Southern / Soul Food Restaurant	Beer Bar	Seafood Restaurant

Concluding Remarks

The analysis provides adequate insight for ant young adults based on their preferences looking at both number of venues available and the popularity of these venues. Both data perspectives corroborate each other with returning very similar results. The similarity of both results provides a basis for any young adult to look at the results and get very good insight on their preferred downtown neighborhood in the great state of Texas, USA.