

Optimal Location for restaurant in Austin, TX

(Capstone Project - The Battle of Neighborhoods)

Venkat Jakka

December 2020

1 Introduction

1.1 Background

There's an influx of tech companies moving to Austin. Lower costs, laid-back lifestyle continue to draw tech companies to Austin, Texas. According to the Austin Chamber of Commerce, 58 major companies relocated to the Austin area in 2019 alone – not including tech giants such as Apple, Amazon, and Google, who opened new offices in the region. Tech companies aren't the only ones who are flocking to Austin, either. Nearly 100 other companies in various sectors have announced that they are moving to the area or expanding their local operations in the coming year. Tens of thousands of well-paying new jobs are on their way to Austin, with more being announced every day.

1.2 Business Problem

All those jobs are going to require smart, motivated, skilled workers to fill them. And those workers need places to live and restaurants or food joints to eat. The objective of this capstone project is to find the most suitable location for an entrepreneur to open a new Italian restaurant in Austin, Texas. By using data science and machine learning methods such as clustering, this project will recommend a best suitable location to open a new Italian restaurant. As with any business, restaurant in a location is of utmost importance, so we will take several things into consideration and suggest an optimal location.

1.3 Interest

Due to the sudden influx of new software, hi-tech, auto and other significant industries, there will be lot population moving into Austin and its neighborhoods creating a potential opportunity for new businesses. Entrepreneurs who are moving into Austin would be very interested in scouting new locations for their establishments.

2 Data acquisition and cleaning

2.1 Data sources

Due to lack to readily available structured data, I scraped relevant data from Wikipedia page that lists Austin neighborhoods. Latitude and Longitude of these neighborhoods were obtained from Geocoder package. Venue data related to these neighborhoods was obtained using Foursquare API.

2.2 Data cleaning

Data scraped from Wikipedia page contained significant irrelevant data which needs to be filtered out to retrieve names of Austin neighborhoods. Adding to that there were duplicate neighborhood names which were removed and few with special characters which cleaned. The final dataset included 94 Austin neighborhoods.

Additionally, several neighborhoods had to be removed from the cleansed data frame since geolocator either unable to find geographical coordinates (latitude and longitude) or the neighborhoods were very far from Austin and cannot be considered as Austin neighborhood.

3 Methodology

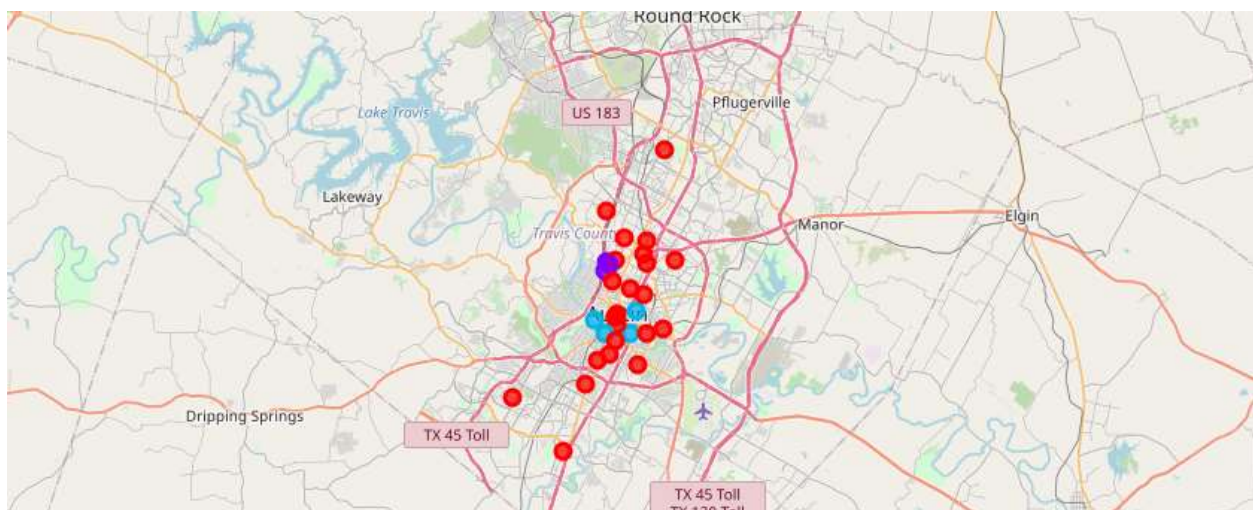
In this project the first step will be to collect data on the neighborhoods of Austin from Wikipedia . Since the data is not available preformatted, it must be scraped from Wiki webpage. The location coordinates of each neighborhood will then be obtained with the help of GeoPy Nominatim geolocator and appended to the neighborhood data. Using this data, a folium map of Austin neighborhoods will be created.

The second step will be to explore each of neighborhoods and their venues using Foursquare location data. The venues of the neighborhoods will be analyzed in detail and patterns will be discovered. This discovery of patterns will be carried out by grouping the neighborhoods using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and it is highly suited for this project as well. Following this, each cluster will be examined and a decision will be made regarding which cluster fits our need. The factor that will determine this is the frequency of occurrence of restaurants and other food venues within the cluster.

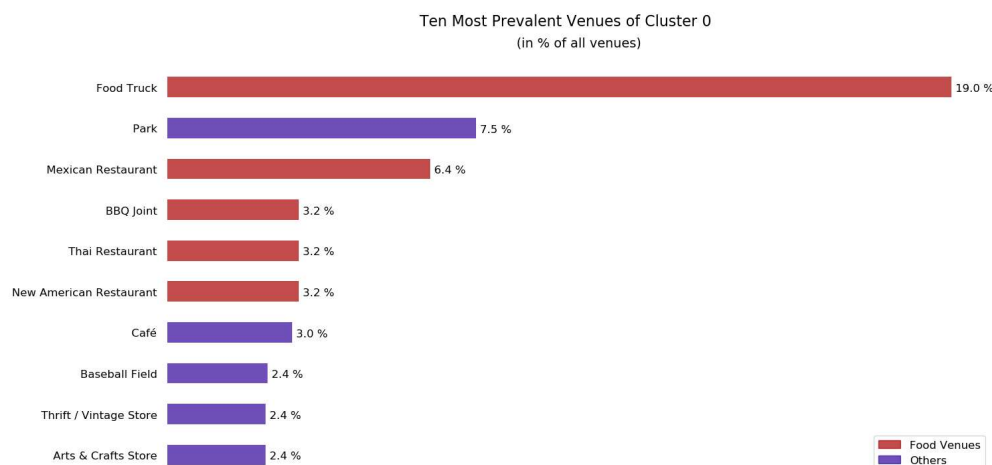
Once a cluster is picked, the neighborhoods in that cluster will be investigated with regards to the number of Italian restaurants in its vicinity. The results of the analysis will highlight potential neighborhoods where an Italian restaurant may be opened based on geographical location and proximity to competitors.

4 Results

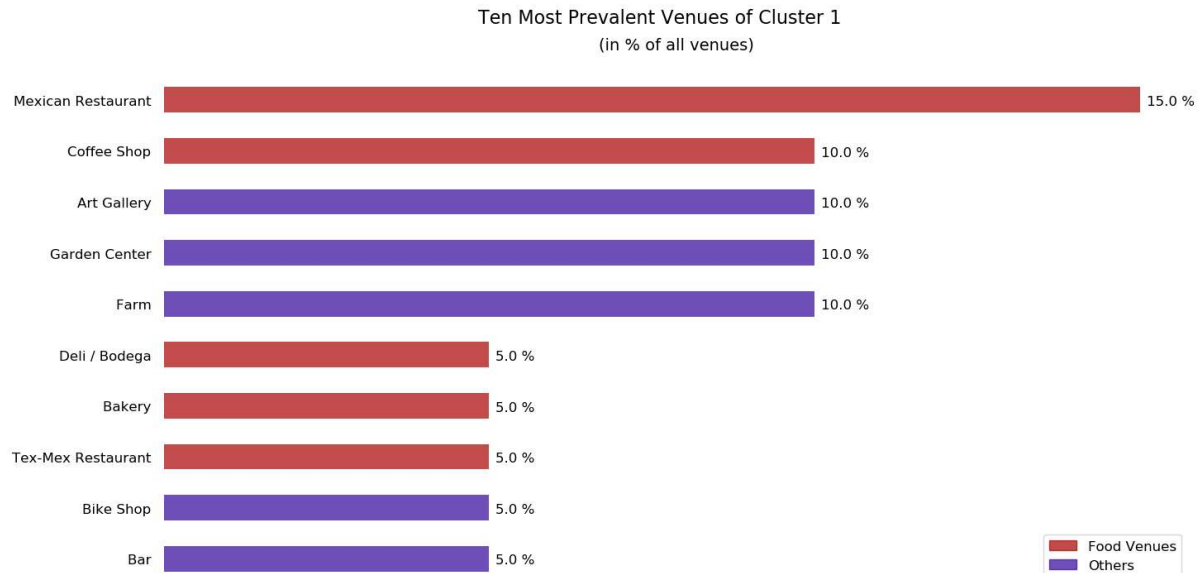
Using K-means clustering technique similar data points have been grouped together forming clusters. These clusters are further analyzed based on their ten most prevalent venues to determine they present an optimal location for a new Italian restaurant.



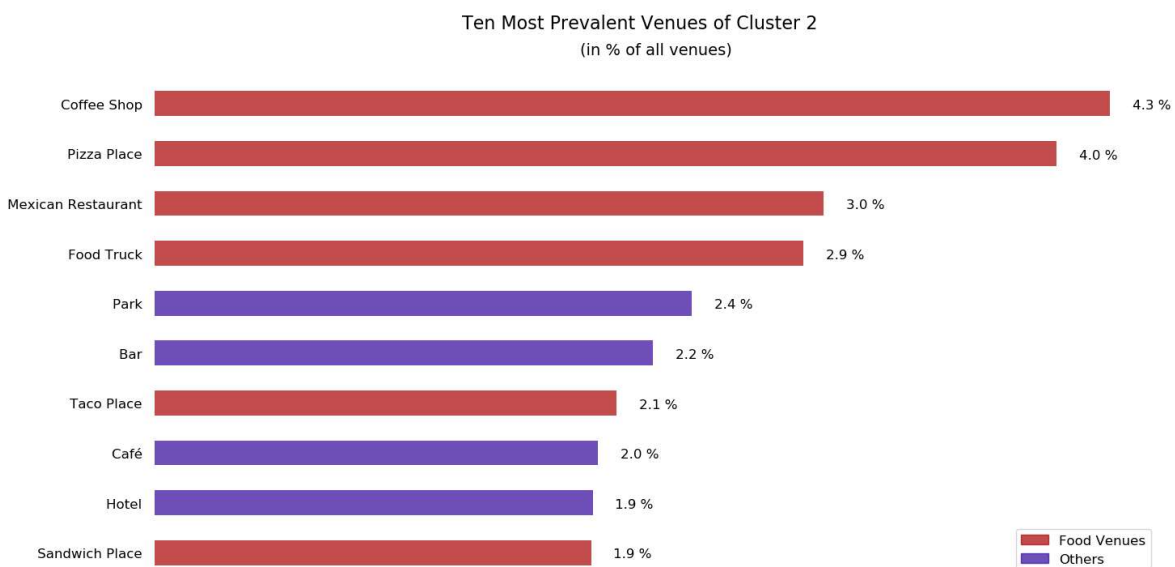
Cluster 0 has 5 food venues among the top 10, with Food Truck and Mexican Restaurants making up nearly 25% of all venues. This poses considerable competition for a new food establishment hence it is not a best option to explore further in terms of setting up a new restaurant.



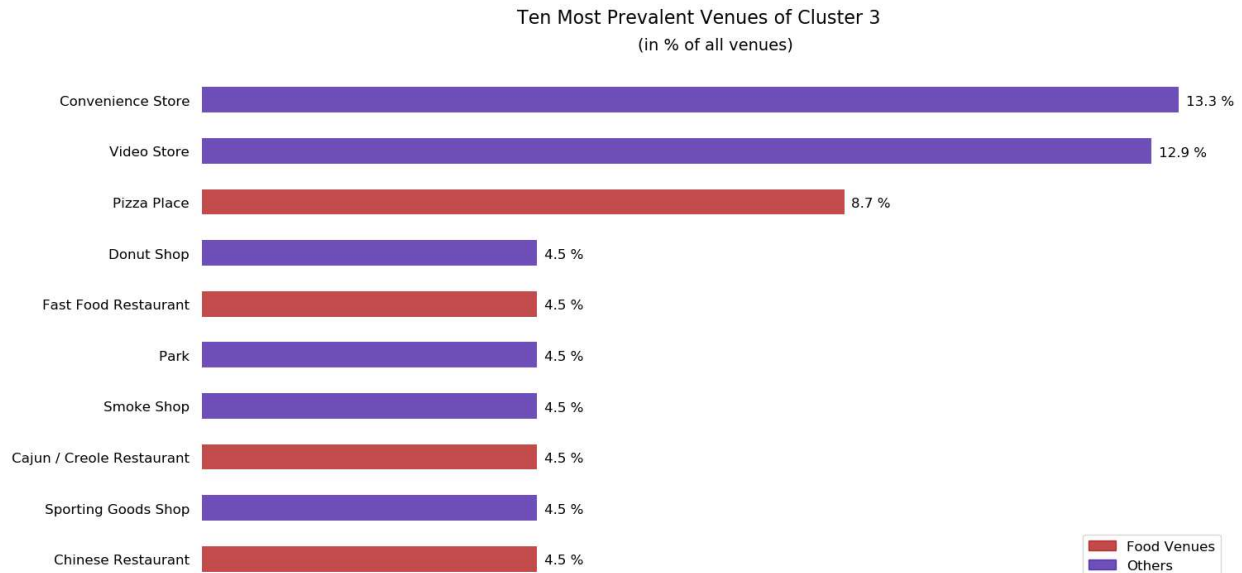
Cluster 1 has 5 food venues among the top 10, with Mexican Restaurants making up a huge majority (nearly 20%) of all venues. Adding to this other than 1 art gallery there are not enough venues to create required foot traffic for a new restaurant, so Cluster 1 is not the best option.



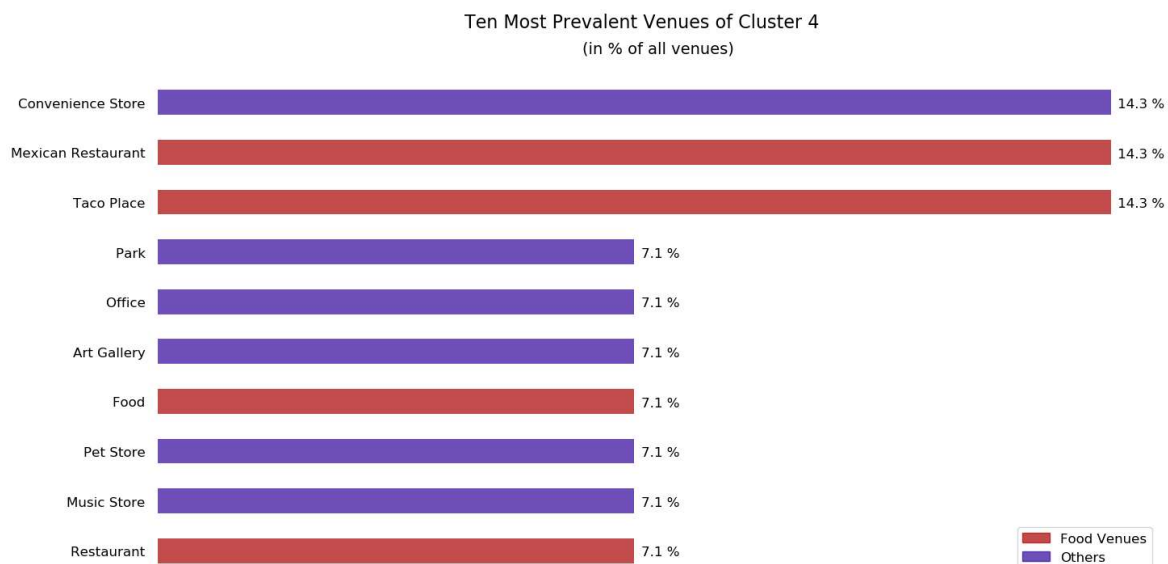
Cluster 2 has 6 food venues in the top 10 venues of Cluster 2 with only 5% full-service Restaurants which makes it viable. But this cluster has only 1 Park and no other office / business establishments to bring in foot traffic needed for a new restaurant. Analyzing the top venues, it appears the fast food and other food establishments in this cluster are around the park, so a new full-service Italian restaurant will not be viable in this cluster.



Cluster 3 has 4 food venues in the top 10 mostly fast food, but not many restaurants. There is a Park and shopping places which are favorable indicators of foot traffic, which can be optimal for new restaurant. Before we finalize on the cluster, let's look at the last cluster.



Cluster 4 has 4 food venues in the top 10 with only 2 of them qualified as full serving restaurants. Additionally, compared to cluster 3 there are a lot of public venues in this cluster - venues that see a lot of foot traffic such as parks, art gallery, office and department stores. The presence of these high traffic places indicates cluster 4 as optimal place for the new restaurant.



To conclude, cluster 4 that includes Central East Austin, Barton Creek and Bouldin Creek are optimal locations for a new Italian Restaurant. Due to the presence of Office venues, Park, Art Gallery and several shopping establishments, it will bring in significant foot traffic (customers) required for the new restaurant. Adding to this, there are no Italian restaurants in top 10 venue which suggests lack of increased competition.

5 Conclusion

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing the machine learning by utilizing k-means clustering and providing recommendation to the stakeholder. We used K-means clustering algorithm to group similar neighborhoods into clusters and further analyze the cluster to recommend a viable cluster of locations suitable to address the business problem – best location for a new Italian restaurant in Austin neighborhood.

6 References

List of Austin Neighborhoods

https://en.wikipedia.org/wiki/List_of_Austin_neighborhoods

Foursquare Developer Documentation

<https://developer.foursquare.com/docs>