IBM Applied Data Science Capstone

Food Trucks – Opening a New Business in Austin, Texas

By: Kevin R.

January 2021

Introduction

Austin is currently one of the fastest growing cities by population in the United States. Due to the combination of blue-collar and high-tech jobs, Austin's population is growing by roughly 3.5% each year[1]. With this growth, we can see that Austin is becoming more diverse not only in its population but in its style of living; a multicultural, but a multi-faceted style. Aside from being one of the fastest growing cities in the United States, Austin prides itself in being one of the largest food truck cities in the country (some even consider it the food truck capital of the country). As Austin continues to grow, we can expect more food trucks to enter the market, but the question is where should they park their truck?

Business Problem

The purpose of this analysis is to determine where should a new food truck in the city of Austin, Texas park itself. I will be using clustering and other data science techniques to create a visual suggestion of where the food truck should go in order to reduce competition.

Audience

Those that would be most interested in this analysis are people who are either considering opening a food truck in Austin or are considering moving from one area of Austin to another. As food trucks become more popularized, people will slowly drift to the relatively lower-cost alternative of opening a restaurant.

[1] https://www.macrotrends.net/cities/22926/austin/population

Data

The data that is needed is the following:

- List of neighborhoods in Austin
- Latitude and longitude of the neighborhoods
- Food truck data

In order to obtain this data, we must use the following sources:

- Wikipedia page outlining all neighborhoods within Austin (https://en.wikipedia.org/wiki/List_of_Austin_neighborhoods)
- Beautifulsoup and Geopy packages
- Foursquare API

Data Methodology

Using the Wikipedia page, we are able to get a list of all neighborhoods within Austin. Beautifulsoup will allow us to scrape the page for the necessary information. As of now, the webpage isn't perfect so we will need to ignore certain tables that are repetitive, while only looking at the information in the bullet points. Due to the way the Wikipedia page is set up, we had to manually put a stopper so that it stops at the last neighborhood. It's not ideal, but it was necessary. Once we get the unique names of the neighborhoods, we will use the Geopy package to get the coordinates of each neighborhood. Wikipedia has listed some neighborhoods that aren't recognized by Geopy, so I added an exception where if it doesn't recognize the neighborhood, it just skips it and removes it from the analysis. These neighborhoods are typically unincorporated zones that are locally known, but not official. For each neighborhood, the Geopy package gives us the center coordinates that we will then use with the FourSquare API.

Once we have the coordinates of each neighborhood, we can use the FourSquare API to get venue data in each of the neighborhoods. To accomplish this, we start looping the coordinates we received from Geopy. FourSquare sends us the response in a JSON format, so we need to clean the data by only retrieving the fields that we want such as: Name, Latitude, Longitude, and Category Name. Once we have all the venue information for each neighborhood, we need to group the FourSquare results by neighborhood and then get the frequency of category type

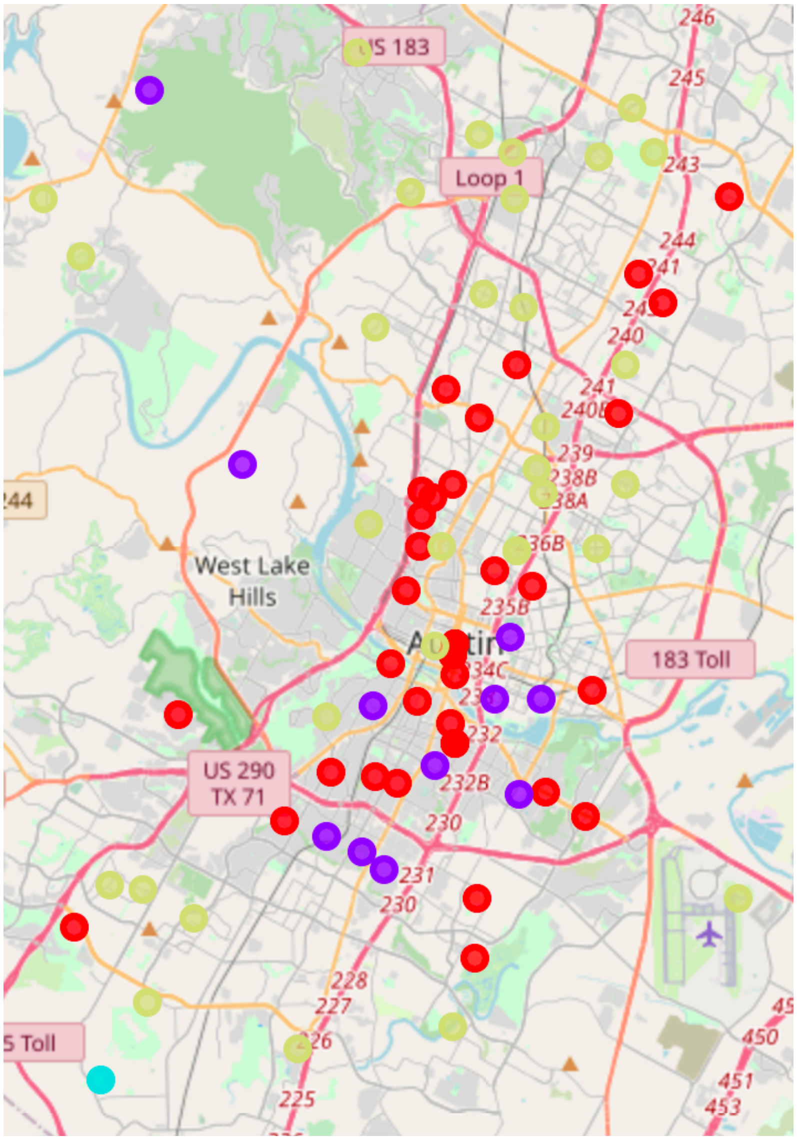
(Category Name). This will give us a simple breakdown such as: 10% of all venues in X neighborhood belong to Y category. This will allow us to easily cluster the results and see where we can segment the neighborhoods. To continue, we need to filter the dataframe so that it only contains “Food Truck” as the category (Category Name).

To determine the concentration of Food Trucks, we will use K-Means Clustering which is a form of unsupervised machine learning. The idea behind this method is to give us clusters where the results are as close as possible while minimizing the number of clusters. At first, the cluster count was 3 but as we finished the analysis, we increased it to 4 clusters as there are certain anomalies in the data. Once we gave each neighborhood a cluster number, we were able to impose the data onto a Folium Map which allowed us to see the result and determine which cluster would be better suited for a new Food Truck.

Results

As previously mentioned, the cluster count was increased to 4 due to an anomaly where only one neighborhood had an excess of food trucks due to not having many other venues in the area.

- Cluster 0: Red circles with low Food Truck count
- Cluster 1: Purple circles with moderate Food Truck count
- Cluster 2: Blue circle with only one Food Truck count (anomaly)
- Cluster 3: Yellow circles with low Food Truck count (residential areas)



Discussion

As shown in the Results section, some of the greatest potential areas are shown to be near the downtown area of Austin. The yellow circles (Cluster 3) should be considered “no-go” even though there isn’t much competition as residential areas typically have considerably less foot-traffic. The blue circle (Cluster 2) is the anomaly as that neighborhood consists of roughly 16% Food Trucks as venues, which is considerably higher than any other neighborhood. Another interesting observation is that some of the best areas to bring a new Food Truck is located within the downtown area of the city, which makes us think: why is this? One of the limitations of this analysis is that it does not take into consideration the local laws and ordinances that would either allow or prohibit a Food Truck from operating in the area. Being cautious, I would instead recommend for new Food Truck owners to move into the purple circles (Cluster 1) as the numbers are very close while comparing them to the red circles (Cluster 0) – but, again, this is solely on the instinct that there must be a valid legal reason as to why there aren't as many Food Trucks as possible in downtown Austin. Though we just recommended where Food Trucks should park, let's not forget that Food Trucks are mobile and can change areas when they want to.

Another form of limitation this analysis had was the lack of socioeconomic information such as: mean income and population density per neighborhood. With this information, further analysis can be done to extrapolate where a Food Truck might be able to park itself for maximum gain, or perhaps to help them create a pricing strategy. Due to the nature of this project, we used FourSquare's free API, meaning that we could not use the fullest features without incurring additional costs.

Conclusion

Throughout this analysis, we have explored the business problem, determined what type of data we need and where to get it, extracted and cleaned the data, used K-Means Clustering to cluster the information into 4 groups, and finally imposed the information onto a Folium Map to more easily give recommendations to prospective Food Truck owners. Ultimately, as mentioned before, the ideal area would be Cluster 1 but let us not forget that they can easily move into Cluster 0 due to the nature of the business (mobile).