Recommending Restaurant Location

Magnus Billups

April 2, 2019

1. Introduction

A prospective business owner wants to open a new Asian Restaurant in Austin, Texas. The owner needs to know what part of Austin would be ideal for opening their new location.

1.1 Problem

Opening a business in the proper location can be pivotal to success, if you open in the wrong location you won't get the right amount of customer traffic to stay in business no matter how good your restaurant is. In this report we will be analyzing which part of Austin would be best based on similarity of popular restaurants in that area*. We will be looking at metrics such as venue locations, popularity, and similarity of food served.

*Normally I would also look at the cost of property in that area. This would give a more realistic idea of the risk factor for opening a business. However, due to the lack of available pricing data I am ignoring this aspect.

1.3 Interest

Any business owner would be interested in knowing the best spot to open their new venue. It helps to reduce risk of dealing with competitors and/or not having enough customer traffic.

2. Data Sources, Cleaning, and Features

To perform our analysis I used venue data pulled from Foursquare's Places API. To get a reasonable amount of data points I pulled data from the search and explore endpoints. Both of these endpoints return data on venues that are within a specified radius of the central location provided.

The search API returns data on venues near a given location. The data returned can be tuned using the category parameter, and I used this parameter to narrow the results to only match venues categorized as Asian Restaurants.

The explore API returns data containing recommended venues with a specified radius of a given location. As with the search API the data returned can be tuned using a query parameter, this parameter was also set to only match venues categorized as Asian Restaurants.

2.1 Data Cleaning

Most of the data that was provided by the search results was delivery information or redundant location data. The data provided by the explore results was more concise but still included redundant or unneeded venue information. After reviewing both data result sets I was able to narrow down the necessary information to create a combined result set shown in figure 1 below.

Fig. 1

	name	categories	latitude	longitude	postalCode
0	Broth & Basil	[Vietnamese Restaurant]	30.468596	-97.595634	78660
1	Sushi Nini	[Sushi Restaurant]	30.479040	-97.671817	78664
2	Saigon El Vendeur	[Vietnamese Restaurant]	30.265481	-97.744288	78701
3	Koriente Restaurant	[Korean Restaurant]	30.267108	-97.736029	78701
4	P.F. Chang's	[Chinese Restaurant]	30.263782	-97.741413	78701

2.3 Feature Selection

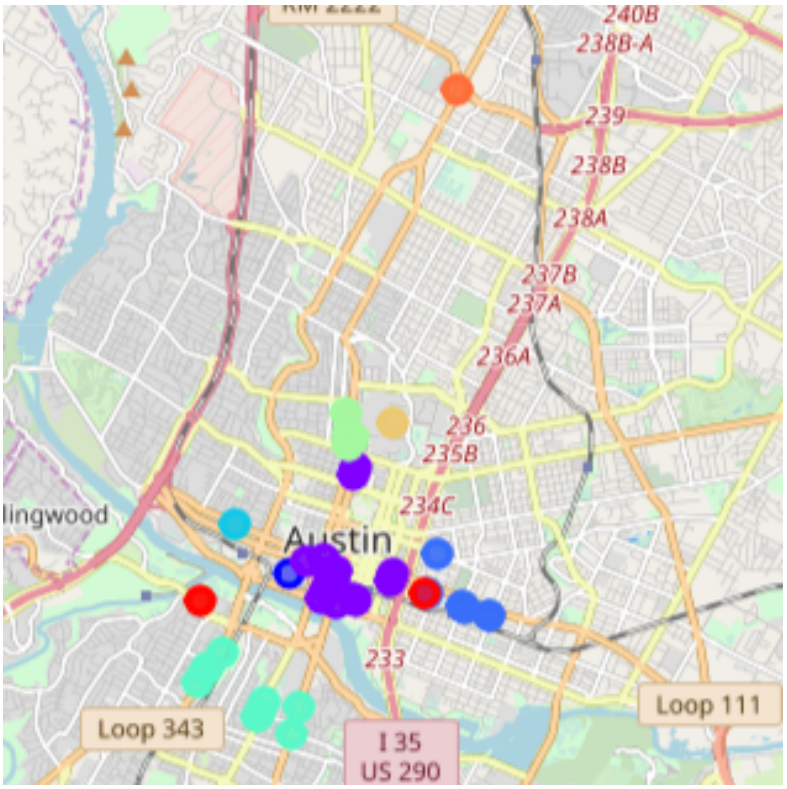
In order to give an accurate estimate of the best location I needed to be able to visualize the data, so I decided to focus on features that would lend to that. The data provided by Foursquares APIs included both the latitude and the longitude of every venue it returned. This was used for creating a map to visualize the density of competition in the region.

The categories feature was also returned by Foursquare and was used to get a more concise representation of what types of Asian restaurants are in the region and the density of each type. The postal code was similarly used to visualize where the highest density of competition is, as these will likely be areas to avoid due to too much competition.

3. Model

I decided that the best model to use for what I wanted was a Cluster. I used the postal codes provided in the data to create the cluster groups. I chose to do this because it creates clear distinctions within the data, and makes visualizing the data simple and understandable.

To visualize the data I create a map displaying each data point using a color system based on the data points postal code. As shown in the map clustering the data this way makes identifying things like venue density fairly easy. However, in order to capture the whole map some of the data points appear to overlap and can make accurate counts difficult to see.



To get a better representation of the data points I decided to use a bar graph as seen in figure 3 below.

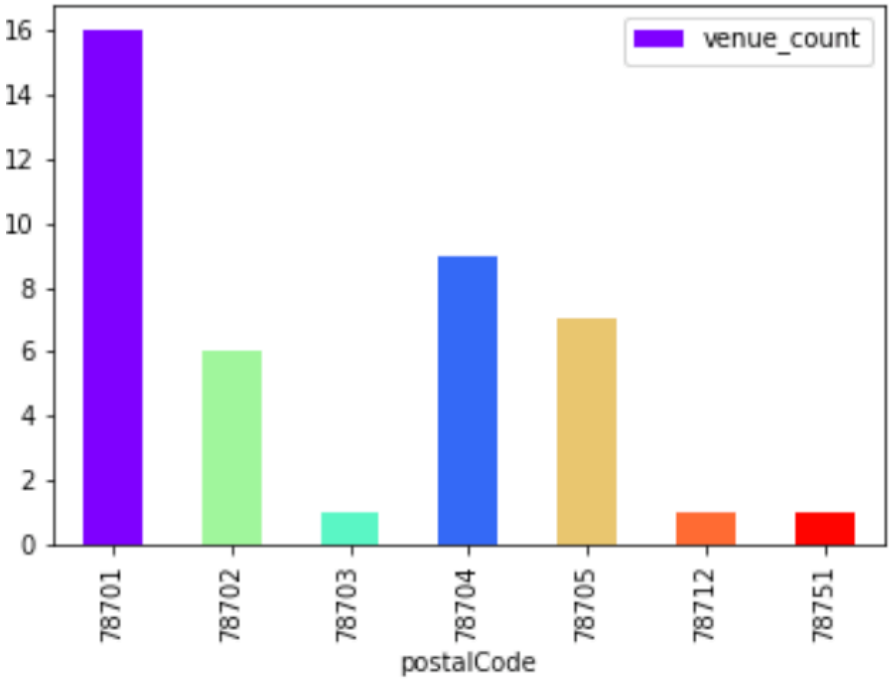


Fig. 3

4. Results

Based on the distribution of venues I would recommend that prospective business owner look into building his new restaurant in the postal code 78702. This area has existing restaurants in the area which means there is some demand for them. At the same time it doesn't have the highest concentration which means that there won't be as much competition.

5. Conclusions

While using only the postal code provides enough information to make a suggestion on where to physically place a new restaurant within an area. That suggestion is not as concrete as it could be, if it were backed by pricing data, and what venues were trending in thee area.