

IBM Applied Data Science Capstone Project Report

By: Suhas Yadlapati

1. Introduction:

1.1 Background

Over the past few years, America has become a country that has been more focused on health than it has ever been before. A spotlight has been put on eating green, organic, and fresh. The younger generation more and more commonly are jumping on trends such as vegetarianism and veganism. There is no better time to capitalize on the health “craze” market than now. The best way to do this is through Vegetarian and Vegan restaurants.

1.2 Problem

However, it does not make sense to place a Vegetarian or Vegan restaurant anywhere. Unlike most other fast-food chains, Vegetarian or Vegan restaurants have a particular set of people who would like to dine in those types of restaurants. So, through this project, I plan to determine which locations in my home city of Austin would be best suited for adding a Vegetarian or Vegan restaurant.

1.3 Interest

I am sure the results that come out of this project would be of very high interest to investors and entrepreneurs that plan on venturing into this new market of healthy eating.

2. Data:

2.1 List of Neighborhoods

The first piece of data that would be needed to complete this project would be the list of neighborhoods found in the city of Austin. This data can be found through Wikipedia and can be implemented in my project by scraping the data from the Wikipedia page using BeautifulSoup packages.

2.2 Coordinates

Next, the coordinates of each neighborhood found through scraping the Wikipedia page will have to be collected. This can be done by using Python’s Geocoder package.

2.3 Venues

Finally, data from Foursquare API will be used to find venues near each of the Austin neighborhoods. This data will also be used to determine which neighborhoods have Vegan or Vegetarian restaurants.

3.Methodology:

3.1 Retrieve list of neighborhoods

To begin the project a list of neighborhoods in the general area of Austin, Texas would be needed. This will be collected through the scraping of a Wikipedia page using the beautifulsoup package as mentioned in the paragraph above.

3.2 Retrieve list of geographical coordinates

The next step will be to determine the geographical coordinates of each neighborhood found in the city of Austin. This will be done using the Geocoder package included in python. Once the longitude and latitude for each neighborhood are found the data will be placed into a pandas dataframe. To visualize the collection of coordinates the folium package will be used to portray the different neighborhoods as dots on a map of Austin.

3.3 Analyzation of Venues

Once the coordinates are obtained the next step will involve the use of the Foursquare API. We will first use the Foursquare API to find the top 100 venues in a 2000 meter radius of our many neighborhoods in Austin using the collected coordinates. Following, Foursquare will return data on each venue and we will extract information such as venue name, venue category, and venue coordinates. Once provided the data, we will calculate how many venues there are in each neighborhood and the number of unique venue categories. Then, we will calculate the occurrence of each venue category for each neighborhood. For this project, we will focus on the occurrence of Vegan/Vegetarian venues in each neighborhood.

3.4 Use of k-means clustering

Finally, we will cluster the data using the k-means clustering technique. The k-means clustering technique is considered one of the simplest data models. However k-means clustering is greatly used for clustering in many data science applications. To begin, will cluster the Austin neighborhoods into three clusters based on the frequency of Vegan/Vegetarian venues in each neighborhood. The results from this process allow us to determine which neighborhoods in Austin have a higher concentration of Vegan/Vegetarian venues and which neighborhoods in Austin have a lower concentration of Vegan/Vegetarian venues. This information should be able to help find an ideal place to place a Vegan/Vegetarian venue..

4.Results:

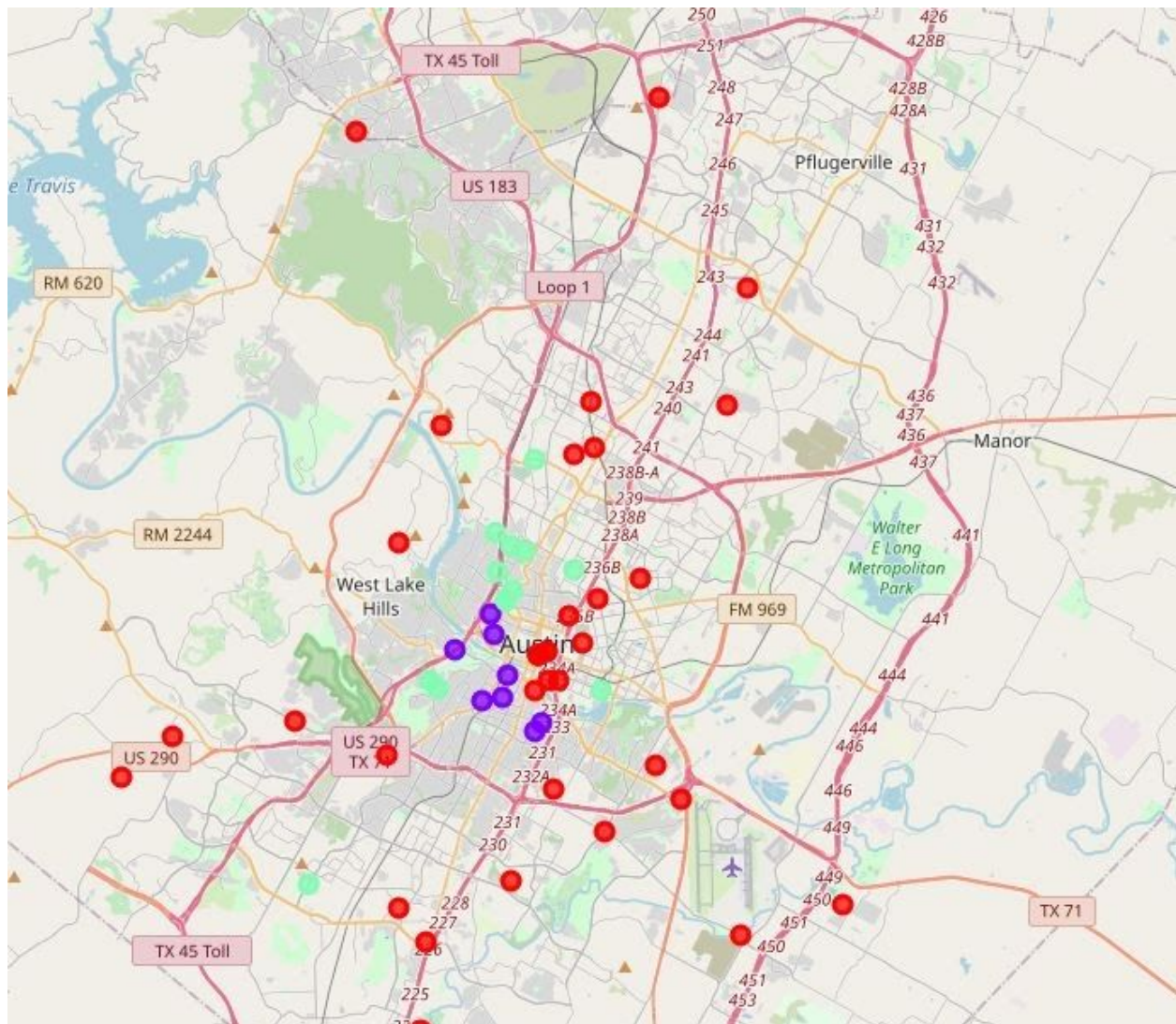
The results from the k-means clustering was able to show us how we can categorize Austin neighborhoods based on the frequency of Vegan/Vegetarian venues:

- Cluster 0:Neighborhoods with no Vegan/Vegetarian

- Cluster 1: Neighborhoods with the highest concentration of Vegan/Vegetarian venues
- Cluster 2: Neighborhoods with a moderate concentration of Vegan/Vegetarian venues

The results of the clustering are visualized in the map below:

- Cluster 0: Red
- Cluster 1: Purple
- Cluster 2: Green



5. Discussion

5.1 Overview

Based on the k-means clustering map, it seems most of the neighborhoods with a Vegan/Vegetarian venue are a bit west to the center of Austin. This consists of cluster 1 with the highest concentration of Vegan/Vegetarian venues and cluster 2 with a moderate concentration of Vegan/Vegetarian venues.

5.2 Cluster 0

The neighborhoods contained in the category of no Vegan/Vegetarian venues can be found all across Austin. Opening a Vegan/Vegetarian in those areas seems to have a great opportunity for one main reason. That is the lack of competition in those areas from other Vegan/Vegetarian venues would be second to none. However, I would suggest doing a bit more research on the demographics of the neighborhoods in cluster 1 to be sure they contain the correct customer base for Vegan/Vegetarian venues.

5.3 Cluster 1

In the case of opening restaurants in the neighborhoods in cluster one, it would not be such a great idea. Opening a Vegan/Vegetarian venue in the area with the highest concentration of Vegan/Vegetarian venues would bring a large amount of competition. This problem of competition is oftentimes hard for restaurants to overcome and require a large amount of marketing to draw in customers from already well-established Vegan/Vegetarian venues in the Area.

5.4 Cluster 2

The best cluster of neighborhoods for placing a Vegan/Vegetarian venue would probably be cluster 2. Firstly, cluster 2 neighborhoods would not have as much completion as cluster 1 neighborhoods. Also, due to the fact that they are a few other vegan restaurants in the neighborhoods of cluster 2, it can be determined that the people living in those neighborhoods already have an interest in Vegan/Vegetarian food.

6. Conclusion

The neighborhoods in cluster 2 are likely the locations to open a Vegan/Vegetarian venue in the city of Austin, Texas. The findings of this project will help investors and entrepreneurs interested in opening a Vegan/Vegetarian venue take advantage of high potential areas while dismissing overcrowded areas and those without an ideal customer base.

