Austin Neighborhood Restaurant Preferences

Contents

Introduction 2
Data 2
Methodology 4
Results 6
Discussion 8
Conclusion 9
References 9

Figure 1: Dataframe of neighborhoods 2
Figure 2: Map displaying locations of neighborhoods 3
Figure 3: Restaurant categories in each neighborhood 3
Figure 4: Catplot of restaurant categories 4
Figure 5: Catplot of number of restaurants in each neighborhood 5
Figure 6: Number of restaurants per category for each neighborhood 5
Figure 7: Optimizing k value 6
Figure 8: Clusters of neighborhoods with similar restaurant categories (Cluster 0=red, 1=purple, 2=blue, 3=green, 4=orange) 7

Introduction

A local meal delivery service is trying to determine how to customize their offerings and stock their local distribution points by neighborhood. They want to understand which products may sell better in a certain neighborhood and adjust their inventory at the distribution point responsible for that area. To do so, they would be interested in seeing the types of restaurants in each neighborhood and see if there are any groups of neighborhoods that have similar preferences.

Data

The geospatial data used for this research comes from city of Austin Open Data Portal [1]. A file containing the boundaries of all the neighborhoods within the city limits can be found on a page of the portal under Boundaries: Austin Neighborhood Planning Areas [2]. The boundary data was then converted to longitude and latitude coordinates using MyGeoData Cloud [3]. Once loaded into a dataframe, unnecessary columns were dropped, and columns were renamed, leaving only the name, latitude, and longitude of each of the 65 neighborhoods.

	neighborhood	latitude	longitude
0	Allandale	30.340218	-97.746431
1	Barton Hills	30.252465	-97.790368
2	Bouldin Creek	30.251534	-97.755814
3	Brentwood	30.329413	-97.731981
4	Central East Austin	30.270324	-97.726068
...	...	...	...
60	Westgate	30.224844	-97.797951
61	Windsor Park	30.308905	-97.690085
62	Windsor Road	30.298798	-97.752740
63	Wooten	30.360237	-97.723255
64	Zilker	30.255026	-97.768613

65 rows × 3 columns

Figure 1: Dataframe of neighborhoods

These coordinates for each neighborhood were validated by plotting them on a map.

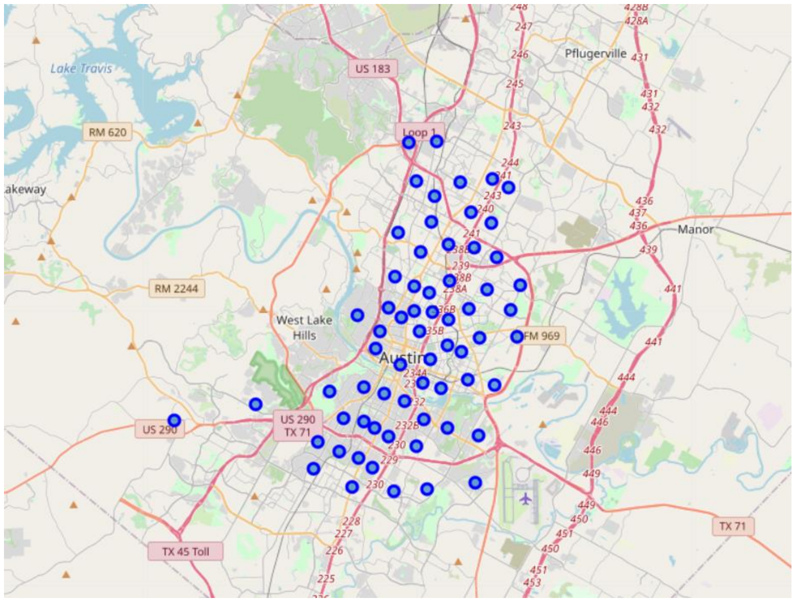

Figure 2: Map displaying locations of neighborhoods

The restaurant data comes from the Foursquare API [4]. To gather the needed data, we will run query the Venue Search endpoint [5] and specify the *Food* category ID (4d4b7105d754a06374d81259) with a radius of approximately 1 mile (1600m) of each neighborhood coordinate to limit our results to just restaurants in each area. This will return results such as the restaurant list found at https://foursquare.com/v/pinthouse-pizza/562e6fc3498e291ae03ab4be which would be categorized as a pizza place. The data in the response will be parsed just for the category and it will be associated with the neighborhood. Ambiguous restaurant categories such as 'Food Truck', 'Restaurant', 'Food', 'Food Stand', and 'Cafeteria' will be removed from the dataset. This left 3323 restaurants across all of the neighborhoods.

	Neighborhood	Restaurant Category
0	Allandale	New American Restaurant
1	Allandale	Pizza Place
4	Allandale	Thai Restaurant
5	Allandale	Mexican Restaurant
6	Allandale	Asian Restaurant
...	...	...
3901	Zilker	Café
3906	Zilker	Thai Restaurant
3907	Zilker	Snack Place
3909	Zilker	Italian Restaurant
3911	Zilker	Hot Dog Joint

3323 rows × 2 columns

Figure 3: Restaurant categories in each neighborhood

Methodology

To analyze the data Python 3 was used in a Jupyter notebook. The pandas library was used to collect and clean the data and the folium library was used for all mapping. Since the main question is which groups of neighborhoods are similar to each other, an unstructured clustering algorithm can be used to perform this analysis. The scikit learn library was used for the machine learning algorithm and the seaborn library was used for visualization.

After the restaurant category data was loaded and cleaned, some exploratory data analysis can be done. A count of the restaurant categories found that of the 3323 restaurants in the dataset, they are split over 79 different categories. A catplot of the categories with over 50 restaurants shows that the greatest number being Mexican restaurants of which there are more than double the next greatest category, Sandwich places. It is expected that the differences between clusters of neighborhood restaurant preferences will focus on a combination of these restaurant categories.

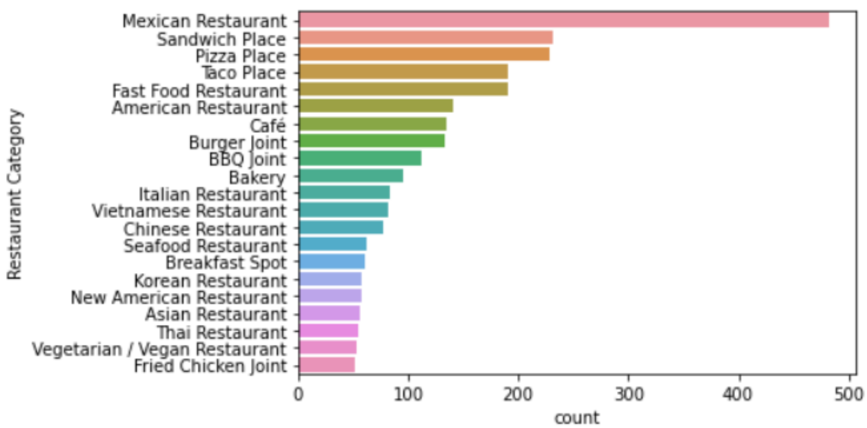

Figure 4: Catplot of restaurant categories

Another observation is that neighborhoods also vary in how many restaurants they have. This should be taken into account so that a low number of restaurants in a neighborhood will not skew the results.

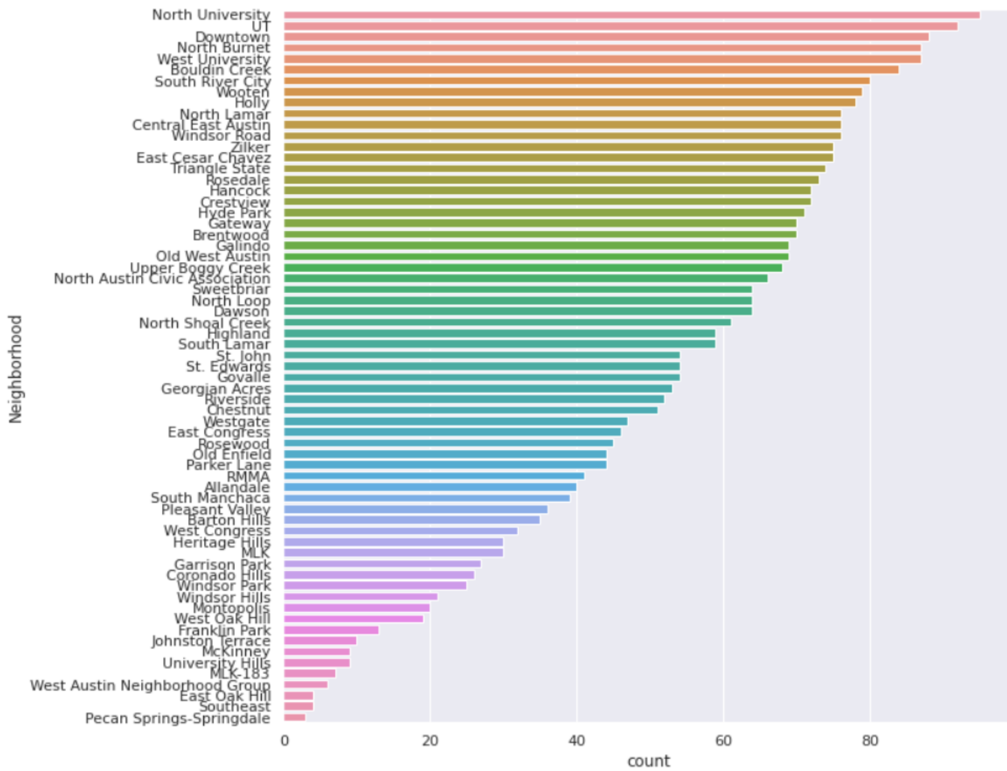

Figure 5: Catplot of number of restaurants in each neighborhood

The restaurant categories can then be grouped by neighborhood, one hot encoded, and summed to produce the total number of restaurants in each category for each neighborhood.

	Neighborhood	African Restaurant	American Restaurant	Argentinian Restaurant	Asian Restaurant	BBQ Joint	Bagel Shop	Bakery	Bistro	Brazilian Restaurant	...	Sushi Restaurant	Szechuan Restaurant	Taco Place	Taiwanese Restaurant
0	Allandale	0	1	0	3	1	0	4	0	0	...	0	0	2	0
1	Barton Hills	1	1	0	1	0	0	1	0	0	...	0	0	5	0
2	Bouldin Creek	0	4	0	0	5	0	1	0	0	...	1	0	4	0
3	Brentwood	0	1	0	1	2	1	3	0	0	...	1	0	4	0
4	Central East Austin	0	3	0	1	5	1	0	0	0	...	0	0	4	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
60	Windsor Hills	0	0	0	0	0	0	1	0	0	...	0	0	2	0
61	Windsor Park	0	0	0	0	0	0	0	0	0	...	0	0	3	0
62	Windsor Road	0	4	0	2	3	1	2	0	0	...	2	0	4	0
63	Wooten	0	2	0	1	1	0	2	0	0	...	1	0	7	0
64	Zilker	1	3	0	1	5	0	6	0	0	...	0	0	2	0

65 rows × 80 columns

Figure 6: Number of restaurants per category for each neighborhood

With this data in this format, a k-means clustering algorithm from the scikit-learn package can be used to group the neighborhoods by restaurant preferences. To determine the optimal k value, the elbow method was used after testing k values up to 10. After analysis, a k value of 5 was chosen and the clustering algorithm was run.

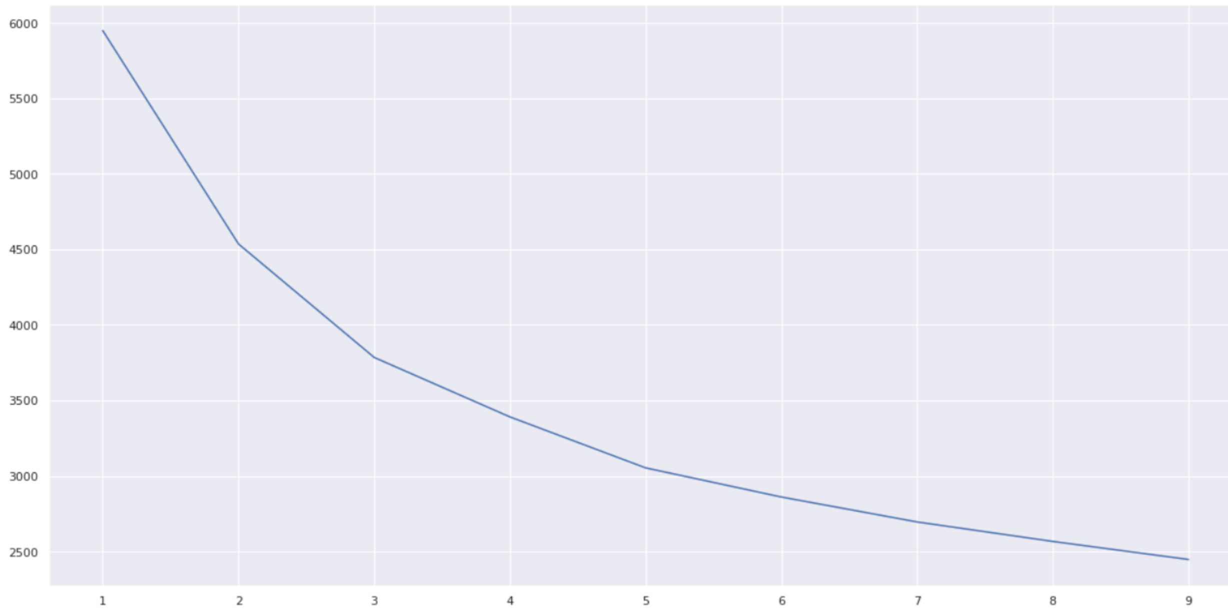

Figure 7: Optimizing k value

Results

The clustering algorithm group the neighborhoods into 5 clusters. By adding the clusters to the markers on the map, it is easier to visualize.

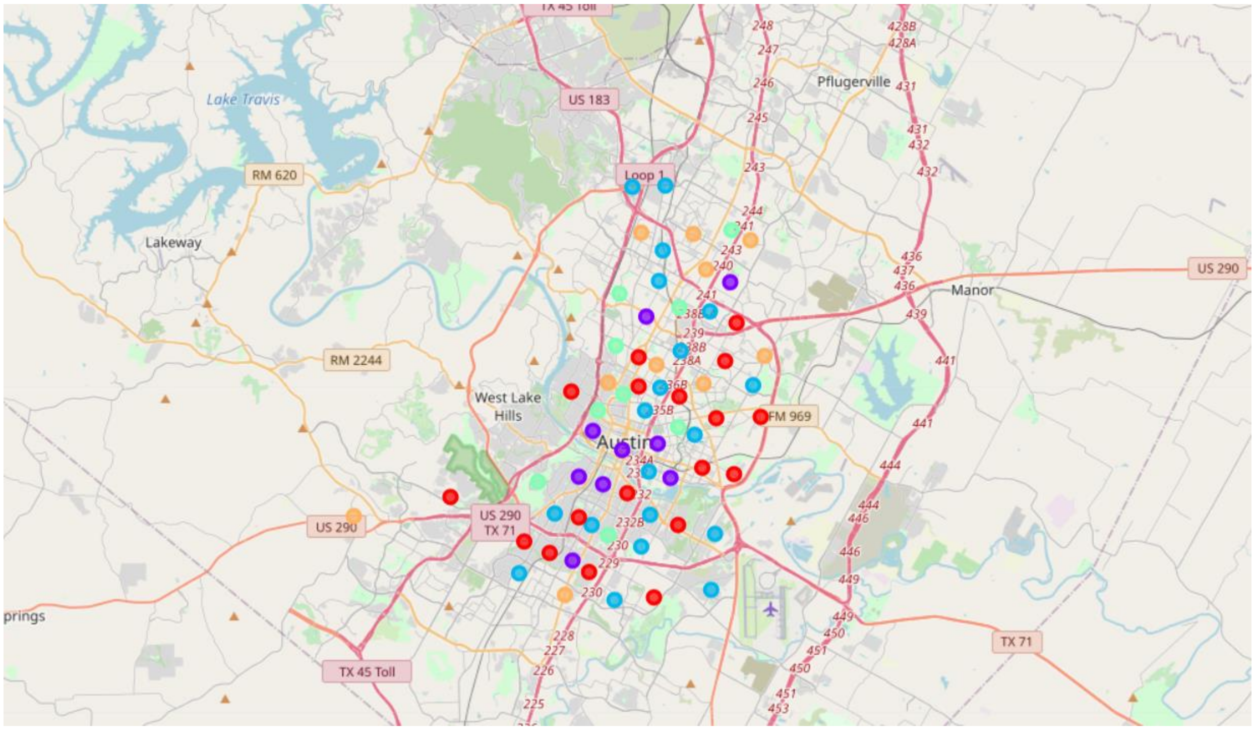

Figure 8: Clusters of neighborhoods with similar restaurant categories (Cluster 0=red, 1=purple, 2=blue, 3=green, 4=orange)

To examine what is similar within each cluster, additional analysis can be performed to extract the restaurant categories which had the highest average number in each cluster group.

Table 1: Cluster 0 similarities

	Avg	%
Mexican Restaurant	2.888889	0.171053
Fast Food Restaurant	2.055556	0.121711
Pizza Place	1.555556	0.092105
Taco Place	1.000000	0.059211
Sandwich Place	0.944444	0.055921

Cluster 0 has a relatively low number of restaurants in each neighborhood but of the restaurants that are there, Mexican restaurants have the highest number at around 17%.

Table 2: Cluster 1 similarities

	Avg	%
Mexican Restaurant	10.333333	0.133813
Pizza Place	5.222222	0.067626
BBQ Joint	4.333333	0.056115
American Restaurant	3.666667	0.047482
Burger Joint	3.555556	0.046043

Cluster 1 has a fairly high number of restaurants in each neighborhood and is the only cluster with BBQ, American, and Burgers in the top 5.

Table 3: Cluster 2 similarities

	Avg	%
Mexican Restaurant	11.789474	0.207600
Taco Place	3.736842	0.065802
Pizza Place	3.684211	0.064875
Fast Food Restaurant	3.631579	0.063948
Sandwich Place	3.368421	0.059314

Cluster 2 has the highest number of Mexican restaurants in both absolute quantity and percentage averaging 11 Mexican restaurants covering over 20% of the overall number of restaurants in the cluster.

Table 4: Cluster 3 similarities

	Avg	%
Mexican Restaurant	5.444444	0.108889
Fast Food Restaurant	3.777778	0.075556
Sandwich Place	3.333333	0.066667
Pizza Place	3.111111	0.062222
Taco Place	2.888889	0.057778

Cluster 3 has the most diverse restaurants with Mexican as the highest category of restaurant at only 10%.

Table 5: Cluster 4 similarities

	Avg	%
Sandwich Place	9.3	0.116981
Mexican Restaurant	6.4	0.080503
Pizza Place	5.6	0.07044
Café	5.3	0.066667
Taco Place	4.6	0.057862

Cluster 4 is another diverse cluster and the only cluster which did not have Mexican as the highest. Rather, this cluster has more Sandwich places and has a high number of cafes.

Discussion

Looking at the data, we can observe that Austin has a large number of categories of restaurants. That being said, there are the most Mexican restaurants in the city followed by Sandwich places, Taco places, Pizza places, and Fast Food restaurants. It is not surprising that most of the clusters of neighborhoods also reflected this fact. Three of the clusters (cluster 0, 2, and 3) had these 5 restaurant categories in it’s top five. Only cluster 1 had BBQ, American, and Burgers in it’s top 5. Only cluster 4 had cafés in it’s top 5.

Geographically, only cluster 1 and cluster 4 seem to be primarily grouped together with cluster 1 primarily in the south of the city and cluster 4 primarily in the north of the city.

Further analysis can be done to correlate other important factors such as per-captia spending on food and preferences for ordering in meals by neighborhood which would provide additional information to determine how to best stock inventory.

Assuming these neighborhoods order in meals with the same preferences as local restaurants in their neighborhood, it is expected that stocking Mexican food, sandwiches, tacos, pizza, and types fast food across the whole city would be safe. In cluster 1 stocking BBQ, American food, and burgers may also be a good option.

Conclusion

K-means clustering allowed neighborhoods in Austin to be grouped together by restaurant categories.

References

[1] "Austin Open Data Portal," [Online]. Available: https://data.austintexas.gov/.

[2] "Boundaries: Austin Neighborhood Planning Areas," [Online]. Available: https://data.austintexas.gov/dataset/Boundaries-Austin-Neighborhood-Planning-Areas/nz5f-3t2e.

[3] "MyGeoData Cloud," [Online]. Available: https://mygeodata.cloud/.

[4] "Foursquare API," [Online]. Available: https://developer.foursquare.com/developer/). .

[5] "Venus Search endpoint," [Online]. Available: https://developer.foursquare.com/docs/api-reference/venues/search/.