



BATTLE OF NEIGHBORHOODS

Saint Louis Missouri

By

John Simpson



PROBLEM STATEMENT

- Where in the city of Saint Louis would be the best place to open a new restaurant?
- How would I maximize any advantage I may have by opening an ethnic restaurant?
- How can I ensure lower competition for customers?
- Understanding the relationship between different area of the city.

OBJECTIVES

- Web scrap the names of community areas and zip codes around the greater Saint Louis metropolitan area. (Beautiful Soup Library)
- Collect the geospatial coordinates of the center of all studied communities (Geopy Library)
- Collecting Neighborhood's top venues using the Foursquare API
- Analyzing the venue data using the k-mean clustering algorithm (sklearn)
- Determining the density of food venues and the diversity of food venues in each community and representing them graphically.

WEB SCRAP AND GEOSPATIAL DATA

1. Communities and Postal Codes

<https://www.bestplaces.net/find/zip.aspx?st=mo&msa=41180>

2. Geospatial coordinates with the ArcGis API through Goepy library

3. Visualization of the data via Folium Library

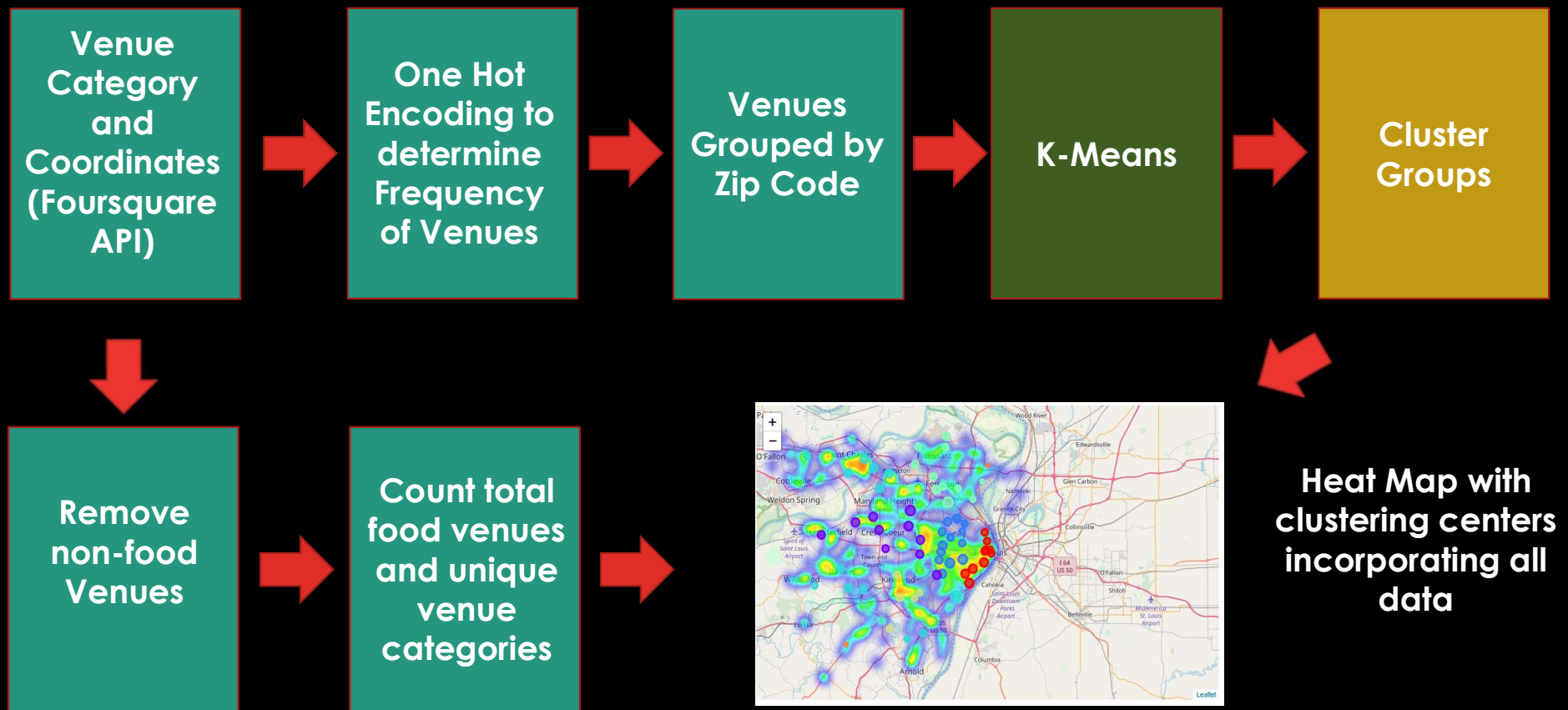
	ZipCode	AreaName
0	63005	Chesterfield
1	63010	Arnold
2	63011	Ballwin
3	63012	Barnhart



	ZipCode	AreaName	Latitude	Longitude
0	63005	Chesterfield	38.656650	-90.586180
1	63010	Arnold	38.436944	-90.366567
2	63011	Ballwin	38.600194	-90.542303
3	63012	Barnhart	38.336075	-90.402166



ANALYSIS WORK FLOW

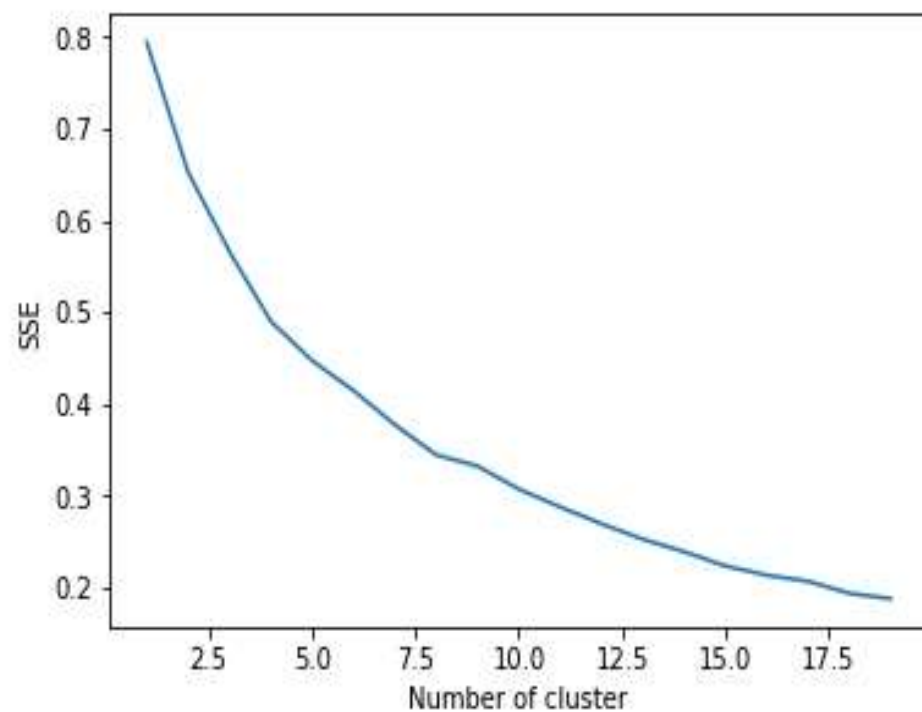


K MEANS OPTIMIZATION

Two primary methods to determine the optimal number of clusters:

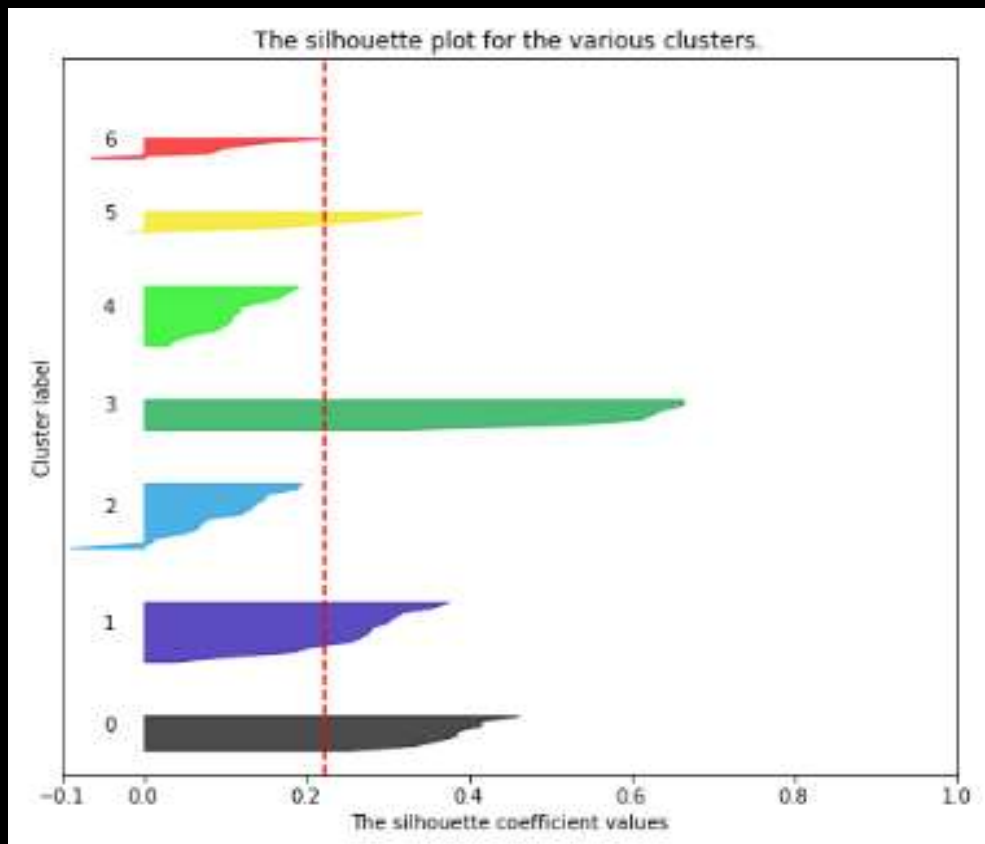
1. Elbow Plot: Plot the sum of squared residuals versus the number of clusters after generating the model. The point of the change in slope indicates the optimal number.
2. Silhouette score: Silhouette coefficient is calculated from the mean intra-cluster distance and the mean nearest cluster distance. A value of 1 indicates a good assignment, 0 a questionable assignment, and a score of -1 indicates possible misassignment.

ELBOW PLOT



Elbow plot method is inconclusive. Best value may be 7 clusters.

SILHOUETTE SCORE

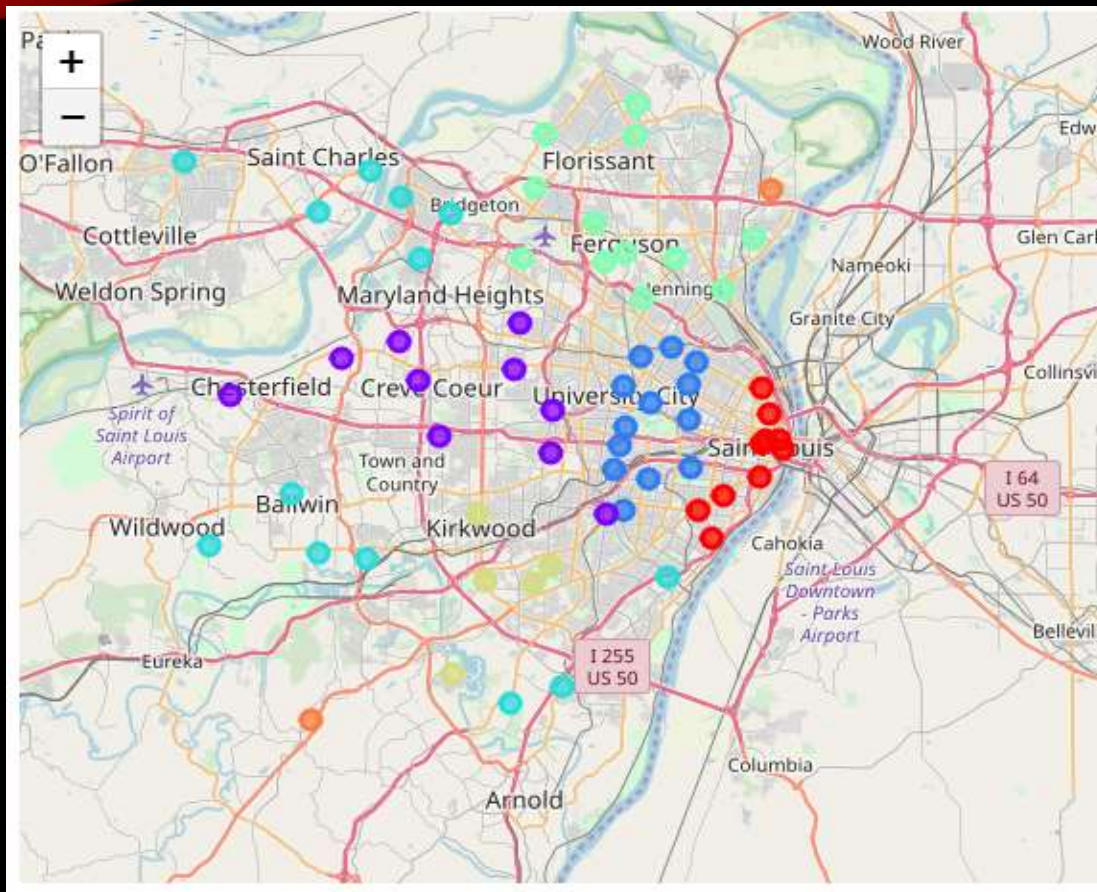


```
For 2 Clusters the average silhouette_score is : 0.16465279456333792
For 3 Clusters the average silhouette_score is : 0.16979980586217155
For 4 Clusters the average silhouette_score is : 0.20910853132701365
For 5 Clusters the average silhouette_score is : 0.2159659500739034
For 6 Clusters the average silhouette_score is : 0.20647028017579427
For 7 Clusters the average silhouette_score is : 0.22325384875601562
For 8 Clusters the average silhouette_score is : 0.21801845100625375
For 9 Clusters the average silhouette_score is : 0.21502686043137162
For 10 Clusters the average silhouette_score is : 0.22741873189010187
```

7 Clusters is again generally supported, however it is questionable.

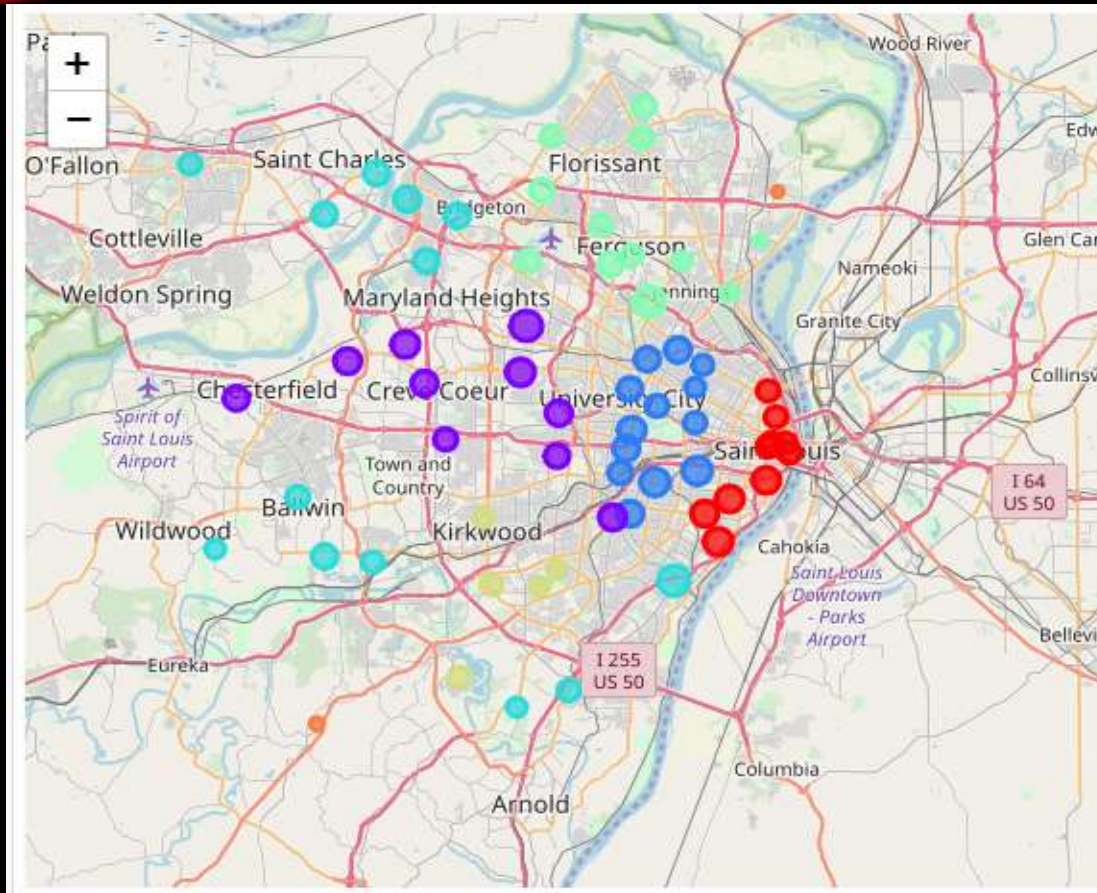
Considering both methods, 7 clusters will be considered optimal.

VISUALIZATION OF ANALYSIS



Clusters appear to fall along known geographic and socioeconomic boundaries.

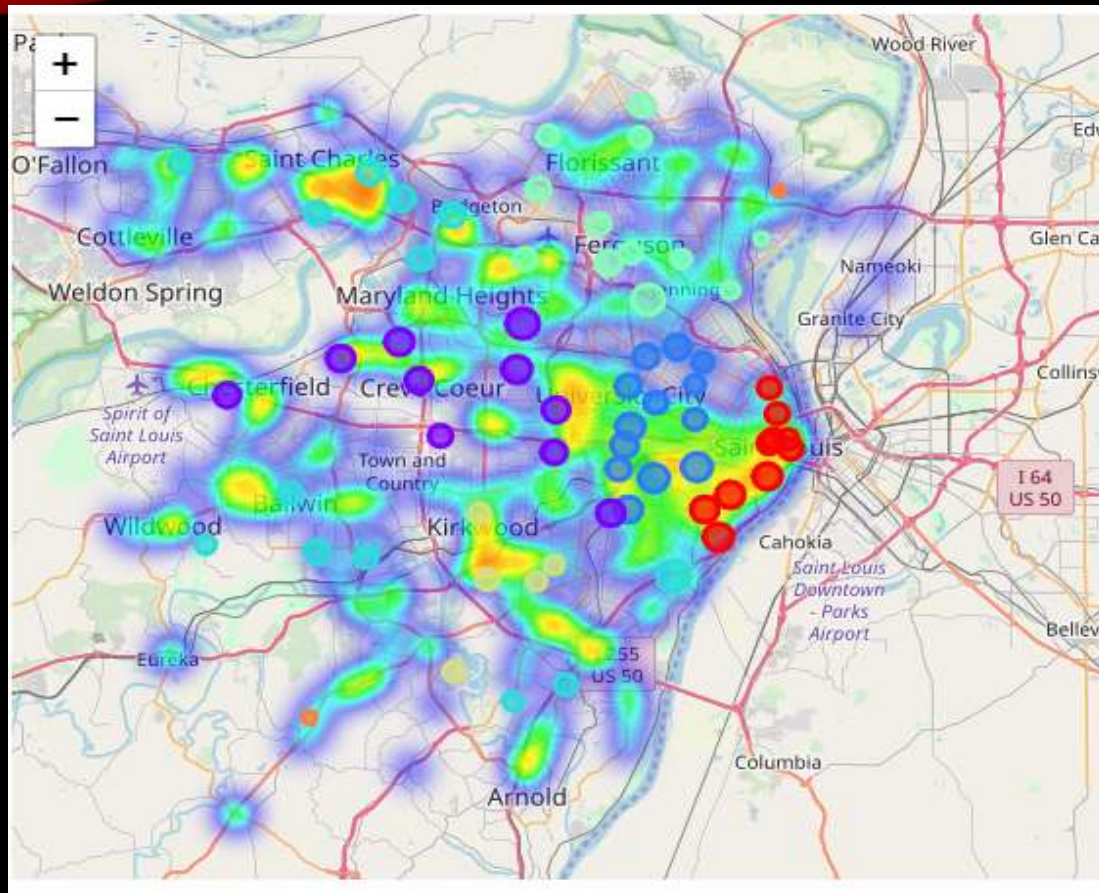
INCORPORATION OF DIVERSITY SCORE



Size of cluster marker increased or decreased relative to the mean diversity score for all zip codes.

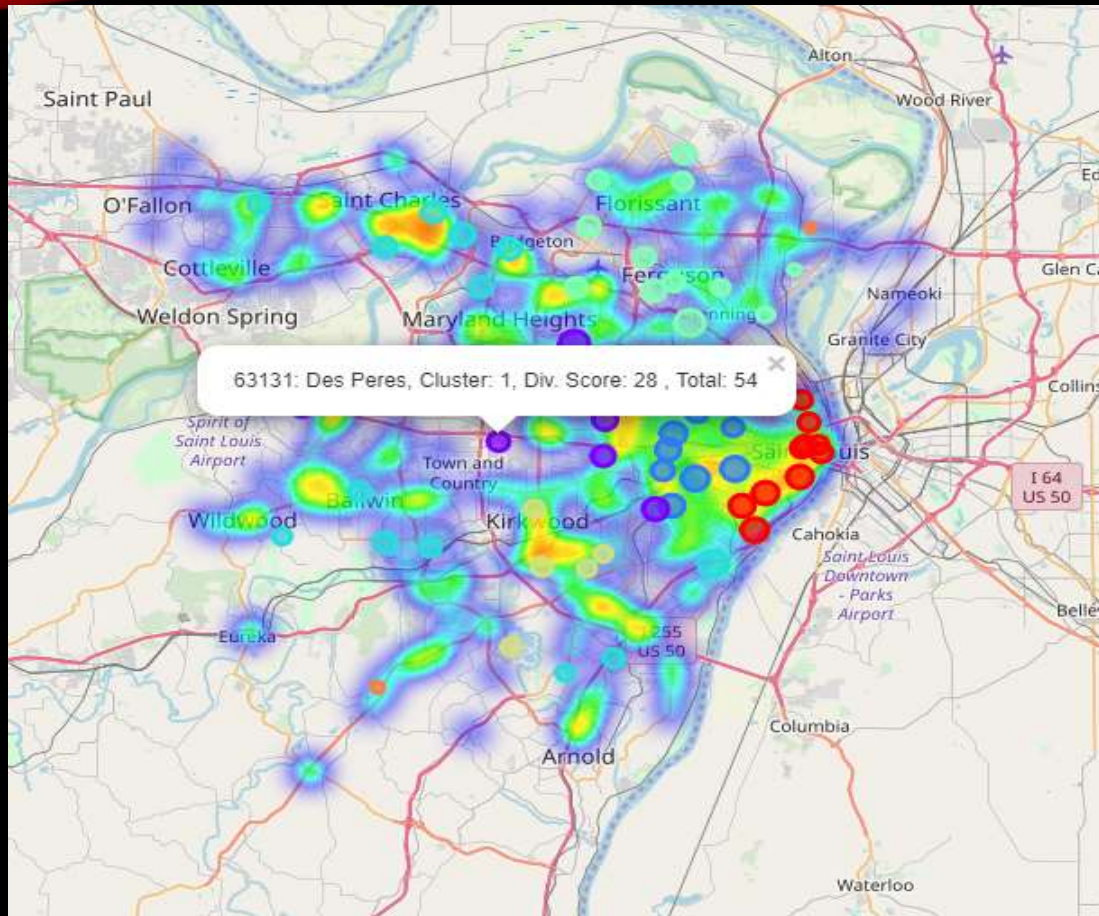
Now to incorporate the locations and density of food venues in the form of a heat map.

HEAT MAP



Heat map indicates the location and density of food venues in each zip code, assuming they are within the top 100 venues of an area.

RESULTS



The most interesting location with the core area of the city appears to be Des Peres within cluster one. It appears to have a low density of venues and diversity score for the cluster.

RESULTS

AreaName	Chesterfield	Chesterfield	Overland	Webster Groves	Ladue	Des Peres	Olivette	Creve Coeur	Brentwood	Saint Louis
1st Most Common Venue	Fast Food Restaurant	Sandwich Place	Convenience Store	Italian Restaurant	Italian Restaurant	Grocery Store	Italian Restaurant	Sandwich Place	American Restaurant	Sandwich Place
2nd Most Common Venue	Italian Restaurant	Grocery Store	Breakfast Spot	Pizza Place	Grocery Store	Italian Restaurant	Grocery Store	American Restaurant	Pizza Place	Italian Restaurant
3rd Most Common Venue	American Restaurant	American Restaurant	Grocery Store	Sandwich Place	American Restaurant	American Restaurant	American Restaurant	Italian Restaurant	Grocery Store	Grocery Store
4th Most Common Venue	Grocery Store	Park	American Restaurant	American Restaurant	Pizza Place	Park	Breakfast Spot	Grocery Store	Park	Bakery
5th Most Common Venue	Sandwich Place	Italian Restaurant	Sandwich Place	Mexican Restaurant	Chinese Restaurant	Coffee Shop	Sandwich Place	Burger Joint	Sandwich Place	American Restaurant
6th Most Common Venue	Hotel	Bakery	Indian Restaurant	Brewery	Sandwich Place	Restaurant	Seafood Restaurant	Ice Cream Shop	Coffee Shop	Burger Joint
7th Most Common Venue	Coffee Shop	Korean Restaurant	Pizza Place	Grocery Store	Hotel	Steakhouse	Pizza Place	Korean Restaurant	Seafood Restaurant	Restaurant
8th Most Common Venue	Clothing Store	Ice Cream Shop	Italian Restaurant	Bar	Coffee Shop	Sandwich Place	Chinese Restaurant	Bakery	Restaurant	Thai Restaurant
9th Most Common Venue	Pizza Place	Smoothie Shop	Park	Deli / Bodega	Gym	Golf Course	Gym	Park	Gym	Pizza Place
10th Most Common Venue	Steakhouse	Lingerie Store	Bar	Bakery	Park	Bakery	Steakhouse	Restaurant	Italian Restaurant	Coffee Shop
Diversity	31	33	39	33	32	28	35	32	30	34
All Ven	58	60	65	65	61	54	67	67	62	70

Examining the cluster, the city in general has fewer food venues and venue types in its top 10.

CONCLUSION

- Based on the analysis, the best location to open a restaurant, based on density and variety appears to be Des Peres.
 - Lower density of food venues.
 - Lower variety of food venues.
 - Fewer food venues in the top ten list for Des Peres compared to the rest of the cluster.
 - The inclusion of a park and golf course would indicate a high level of foot traffic.



THANK YOU!