

Where Do We Build

Data Science Specialization Final Capstone Project

DavidL

September/October 2019

FoodForAll (FFA) - Wants to Cater to the Dallas/Fort Worth Area

- ▶ Want to include cities in Dallas County and Tarrant County - the two biggest counties in North Texas

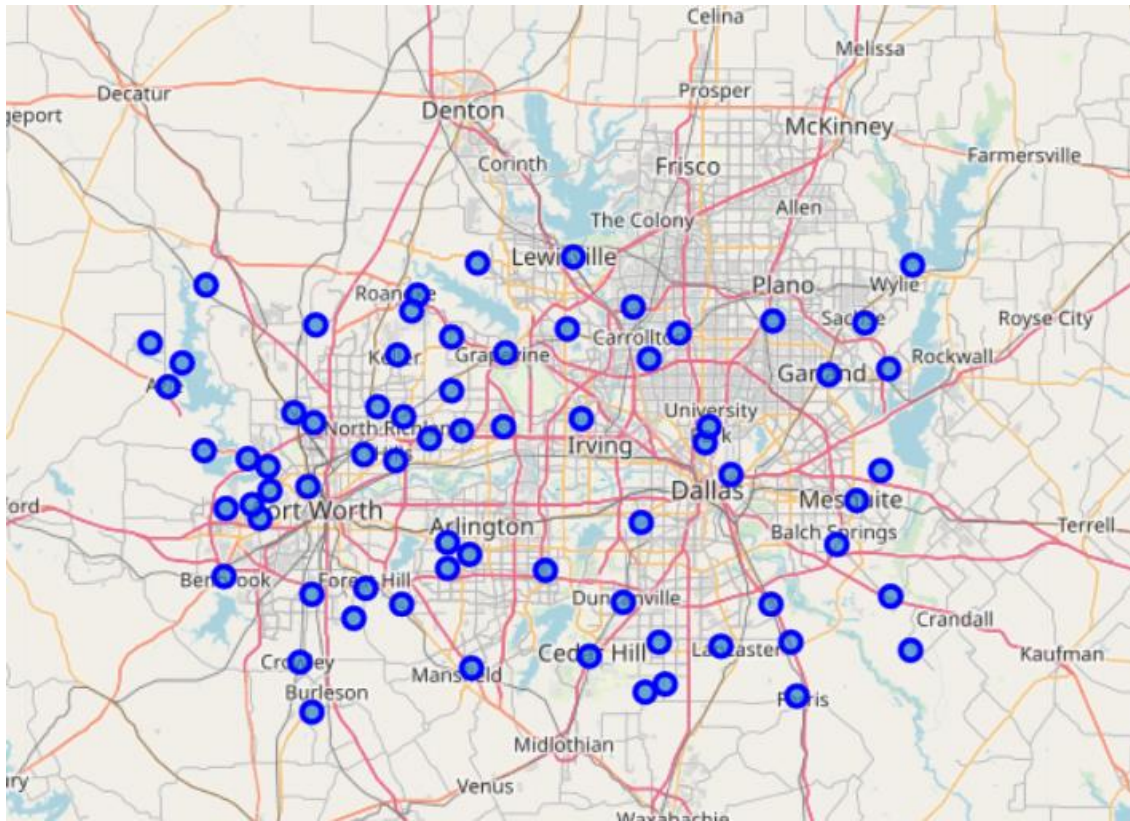
Data Sources

- ▶ Dallas County website: Dallas County city listing
<https://www.dallascounty.org/about-us/cities/>
- ▶ Tarrant County website: Tarrant County city listing
<http://access.tarrantcounty.com/en/county/about-tarrant/incorporated-areas.html>
- ▶ North Central Texas Council of Governments(NCTCOG) - census and location data for cities in the North Central Texas area
<http://data-nctcogis.opendata.arcgis.com/>
 - ▶ Data used is the 'City Limits (2010 Census)' data set
- ▶ FourSquare API Location Data: used to find features within a certain radius of each city center
<https://api.foursquare.com/v2/venues/explore>

Data Cleanup/Processing

- ▶ City Data
 - ▶ Screen Scrape of city data from County Websites
 - ▶ Combine data and remove duplicates and create one dataframe (some cities have parts in both counties)
- ▶ Location Data
 - ▶ Combine City data with the county location data for each city
 - ▶ Use the Foursquare APIs to create a frame with the top 100 features in each location radius
- ▶ Client Preferences
 - ▶ Create a dataframe to store client preferences
 - ▶ Use this to compute the recommendation

Map of the DFW Cities Researched



Descriptive Info for Features

- ▶ More than 6500 features returned
- ▶ 268 Unique features

```
dfw_venues.describe(include='all')
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
count	6550	6550.000000	6550.000000	6550	6550.000000	6550.000000	6550
unique	70	NaN	NaN	2144	NaN	NaN	268
top	Mesquite	NaN	NaN	Starbucks	NaN	NaN	Mexican Restaurant
freq	124	NaN	NaN	151	NaN	NaN	434
mean	NaN	32.767496	-97.065896	NaN	32.767768	-97.061953	NaN
std	NaN	0.493278	0.343883	NaN	0.492716	0.336235	NaN
min	NaN	26.402434	-98.980938	NaN	26.364866	-99.040992	NaN
25%	NaN	32.675537	-97.336874	NaN	32.679178	-97.315589	NaN
50%	NaN	32.812873	-97.135076	NaN	32.811789	-97.115544	NaN
75%	NaN	32.922750	-96.854388	NaN	32.913021	-96.827364	NaN
max	NaN	33.666497	-95.480086	NaN	33.690441	-95.397695	NaN

Results

Neighborhood	Rating
Newark	3.71
Fort Worth	3.27
Azle	2.98
Pelican Bay	2.98
Dallas	2.85

Future Possible Refinements

- ▶ Incorporate population of each of the cities to assist with segmentation. Without that data in this analysis, we don't know if the suggested areas could support another business - this would require some additional work.
- ▶ Increase Weighted Rankings for features: In some cases, it looks like some cities were ranked higher due to a higher quantity of mid-tier possible customers (per the client's preference survey).
- ▶ Look at ways to group Client's feature ratings: With the quantity of features in the data set, better ways to group the different types of features could help improve the recommendation score.
- ▶ Work with client to narrow down more of the demographic
- ▶ Increase Search Radius: To be more accurate, we should look at businesses using a wider radius