

FINDING A GOOD LOCATION FOR GOURMET PIZZA PLACES IN CUYAHOGA COUNTY

IBM Data Science Professional Capstone Project

April 2019

OBJECTIVE

- Basilico Pizzeria, a local gourmet pizza chain in the Cleveland, Ohio area, wants to open 2 new restaurants.
- Need to choose cities without a lot of competition and where the population and median income are high enough to support the business.
- The restaurant business is very competitive, so the owners want to base their 2 new ventures on some solid data.

DATA ACQUISITION AND CLEANING

- Data for latitude and longitude for all US cities was downloaded from: <https://simplemaps.com/data/us-cities>. Removed the non-Ohio locations and uploaded the CSV. See header below.

city	county	lat	lng	population
Dunkinsville	Adams	38.8559	-83.4713	NaN
Cherry Fork	Adams	38.8862	-83.6146	155
Peebles	Adams	38.9463	-83.4094	1753
West Union	Adams	38.7917	-83.5441	3074
Manchester	Adams	38.6902	-83.6053	1978

- Obtained Ohio median income data from https://www.cleveland.com/datacentral/2016/12/median_family_income_household.html and converted it to a CSV and uploaded it. See header below.

city	median income
New Albany	\$214,850
Indian Hill	\$211,795
Pepper Pike	\$186,765
Powell	\$146,442
Dublin	\$144,005

DATA ACQUISITION AND CLEANING (CON'T)

- Next, merge the dataframes and filter dataframe to include only cities in Cuyahoga County. See header below.

city	lat	lng	population	median income
University Heights	41.4948	-81.5348	13126.0	84594.0
Lakewood	41.4824	-81.8008	50279.0	70448.0
Richmond Heights	41.5589	-81.5029	10421.0	63029.0
Shaker Heights	41.4744	-81.5496	27448.0	107938.0
Brecksville	41.3079	-81.6193	13425.0	121974.0

- To the right is a map of our dataset.



DATA ANALYSIS

- Let's begin by looking at the summary statistics.

	lat	lng	population	median_income
count	38.0000	38.0000	38	38
mean	41.4375	-81.6611	67026	80272
std	0.0731	0.1576	277729	30601
min	41.3079	-81.9315	6198	30466
25%	41.3848	-81.8115	12780	58253
50%	41.4356	-81.6289	18711	76733
75%	41.4844	-81.5228	27791	102063
max	41.5904	-81.4401	1731637	186765

- We will use the population and median income statistics as a factor in our decision making.
- Particularly we will be looking for population and median income near the 75 percentile.

DATA ANALYSIS (CON'T)

- Remove population and median income before clustering.
- Below is the header of the dataset we will use for clustering

city	county	lat	lng
University Heights	Cuyahoga	41.4948	-81.5348
Lakewood	Cuyahoga	41.4824	-81.8008
Richmond Heights	Cuyahoga	41.5589	-81.5029
Shaker Heights	Cuyahoga	41.4744	-81.5496
Brecksville	Cuyahoga	41.3079	-81.6193

DATA ANALYSIS (CON'T)

- Next we need to look at the competition.
- We'll use Foursquare to get venues in our cities.
- Our query returned 417 venues and 129 unique categories. Below is the header.

city	Venue	Venue Latitude	Venue Longitude	Venue Category
University Heights	Bialy's Bagels	41.4979	-81.5360	Bagel Shop
University Heights	Geraci's Restaurant	41.4982	-81.5368	Restaurant
University Heights	HomeGoods	41.4985	-81.5363	Furniture / Home Store
University Heights	Peter's Tailor Shop	41.4983	-81.5361	Tailor Shop
University Heights	Quagmire's	41.4934	-81.5402	Bar

DATA ANALYSIS (CON'T)

- Next let's filter the data frame to look at only pizza places. Below is the result.

city	Venue	Venue Latitude	Venue Longitude	Venue Category
Strongsville	Pizza Hut	41.3121	-81.8346	Pizza Place
Lyndhurst	Tasty Pizza	41.5199	-81.4956	Pizza Place
Lyndhurst	Vicinato Pizzeria	41.5194	-81.4969	Pizza Place
Garfield Heights	Papa John's Pizza	41.4164	-81.6059	Pizza Place
Garfield Heights	Pizza Pan	41.4214	-81.6091	Pizza Place
North Olmsted	Little Caesars Pizza	41.4180	-81.9192	Pizza Place
Berea	Marco's Pizza	41.3716	-81.8689	Pizza Place
Berea	Romeo's Pizza	41.3725	-81.8680	Pizza Place
Parma Heights	Pizza Hut	41.3867	-81.7690	Pizza Place
Parma Heights	Godmother's Pizza & Subs	41.3856	-81.7611	Pizza Place
East Cleveland	Happy's Pizza	41.5292	-81.5826	Pizza Place
North Royalton	Marco's Pizza	41.3135	-81.7492	Pizza Place
Solon	Domino's	41.3846	-81.4394	Pizza Place
Solon	Papa John's Pizza	41.3871	-81.4415	Pizza Place
Solon	Pizza Hut	41.3854	-81.4364	Pizza Place
Solon	Antonio's Pizza Solon	41.3840	-81.4362	Pizza Place
Fairview Park	Domino's	41.4460	-81.8550	Pizza Place

DATA ANALYSIS (CON'T)

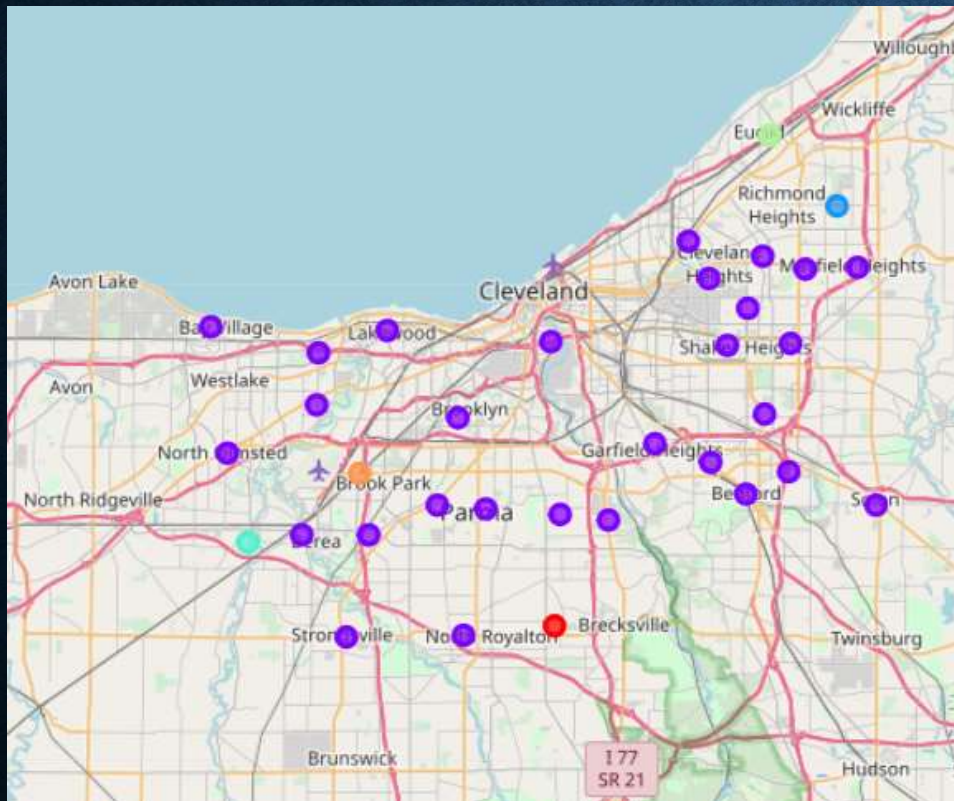
- Next, let's get the counts by city with the groupby function. Below is the result.

city	counts
Berea	2
East Cleveland	1
Fairview Park	1
Garfield Heights	2
Lyndhurst	2
North Olmsted	1
North Royalton	1
Parma Heights	2
Solon	4
Strongsville	1

- I'm surprised there aren't more pizza places listed.
- I'm sure they have a lot of chain and carryout pizza places that aren't reported as much on Foursquare, but that's not our market anyway.
- This does tell us we don't want to open up our pizza place in Solon!

DATA ANALYSIS (CON'T)

- Below are the resulting dataframe and map.



city	Cluster	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
University Heights	0	Restaurant	Bar	Furniture / Home Store
Lakewood	0	Sandwich Place	Salon / Barbershop	Mobile Phone Shop
Shaker Heights	4	Light Rail Station	Women's Store	Football Stadium
Bedford Heights	8	Home Service	Rest Area	Intersection
Strongsville	0	American Restaurant	Shoe Store	Bank
Cleveland	0	Pub	Molecular Gastronomy Restaurant	Bus Station
Lyndhurst	0	American Restaurant	Pizza Place	Shipping Store
Cleveland Heights	0	ATM	Pool	Park
Parma	0	Convenience Store	Sandwich Place	Golf Course
Seven Hills	7	Baseball Field	Pool	Women's Store
Middleburg Heights	0	Automotive Shop	American Restaurant	Community Center
Mayfield Heights	0	Sandwich Place	Women's Store	Men's Store
South Euclid	0	Ice Cream Shop	Bar	Flower Shop
Euclid	6	Pet Store	Women's Store	Convenience Store
Maple Heights	0	Sandwich Place	Convenience Store	Deli / Bodega
Garfield Heights	0	Pharmacy	Fast Food Restaurant	Pizza Place
North Olmsted	0	American Restaurant	Sandwich Place	Fast Food Restaurant
Olmsted Falls	2	Flower Shop	Women's Store	Hot Dog Joint
Bay Village	10	Soccer Field	Gym / Fitness Center	Italian Restaurant
Brooklyn	10	Park	Football Stadium	Soccer Field
Berea	0	Pizza Place	Ice Cream Shop	Sandwich Place
Brook Park	1	Beach	Convenience Store	Deli / Bodega
Broadview Heights	5	Gym	Women's Store	Food Truck
Bedford	0	Bar	Hot Dog Joint	Café
Beachwood	12	Intersection	Bakery	Women's Store
Parma Heights	0	Pizza Place	Bar	Chinese Restaurant
East Cleveland	0	Mobile Phone Shop	Mattress Store	American Restaurant
Warrensville Heights	0	Gym / Fitness Center	Discount Store	Fast Food Restaurant
North Royalton	9	Pool	Gym / Fitness Center	Pizza Place
Rocky River	10	Gym / Fitness Center	Pub	Playground
Highland Heights	3	Concert Hall	Hot Dog Joint	Cosmetics Shop
Independence	11	Theater	Sandwich Place	Art Gallery
Solon	0	Pizza Place	Grocery Store	Sandwich Place
Fairview Park	0	Ice Cream Shop	Italian Restaurant	Donut Shop

DATA ANALYSIS (CON'T)

- We have an interesting finding.
- Category 0 has the most members it's members appear to have more pizza places and other restaurants.
- Remove category 0 and merge with median income data.
- This is the dataset from which we will pick the 2 new locations.

city	population	median_income	Cluster	1st	2nd	3rd
Brook Park	18691	60571	1	Beach	Convenience Store	Deli / Bodega
Olmsted Falls	8890	79699	2	Flower Shop	Women's Store	Hot Dog Joint
Highland Heights	8397	114792	3	Concert Hall	Hot Dog Joint	Cosmetics Shop
Shaker Heights	27448	107938	4	Light Rail Station	Women's Store	Football Stadium
Broadview Heights	19197	99286	5	Gym	Women's Store	Food Truck
Euclid	47360	49668	6	Pet Store	Women's Store	Convenience Store
Seven Hills	11655	92235	7	Baseball Field	Pool	Women's Store
Bedford Heights	10555	51591	8	Home Service	Rest Area	Intersection
North Royalton	30247	90771	9	Pool	Gym / Fitness Center	Pizza Place
Bay Village	15328	109673	10	Soccer Field	Gym / Fitness Center	Italian Restaurant
Brooklyn	10828	54293	10	Park	Football Stadium	Soccer Field
Rocky River	20264	100334	10	Gym / Fitness Center	Pub	Playground
Independence	7114	105600	11	Theater	Sandwich Place	Art Gallery
Beachwood	11707	115536	12	Intersection	Bakery	Women's Store

CONCLUSION

- Based on population, median income, and the amount of competition, it looks like Rocky River and Shaker Heights are good locations for our 2 gourmet pizza places.
- This is based on selecting the 2 cities with over \$100k median income and then the highest population among those.
- Both population and median incomes are well above the mean for Cuyahoga county and close to the 75 percentile.

	lat	lng	population	median income
count	38.0000	38.0000	38	38
mean	41.4375	-81.6611	67026	80272
std	0.0731	0.1576	277729	30601
min	41.3079	-81.9315	6198	30466
25%	41.3848	-81.8115	12780	58253
50%	41.4356	-81.6289	18711	76733
75%	41.4844	-81.5228	27791	102063
max	41.5904	-81.4401	1731637	186765