A New Way to Compare Major Markets

An Analysis of City Similarity Using Business Type Concentration
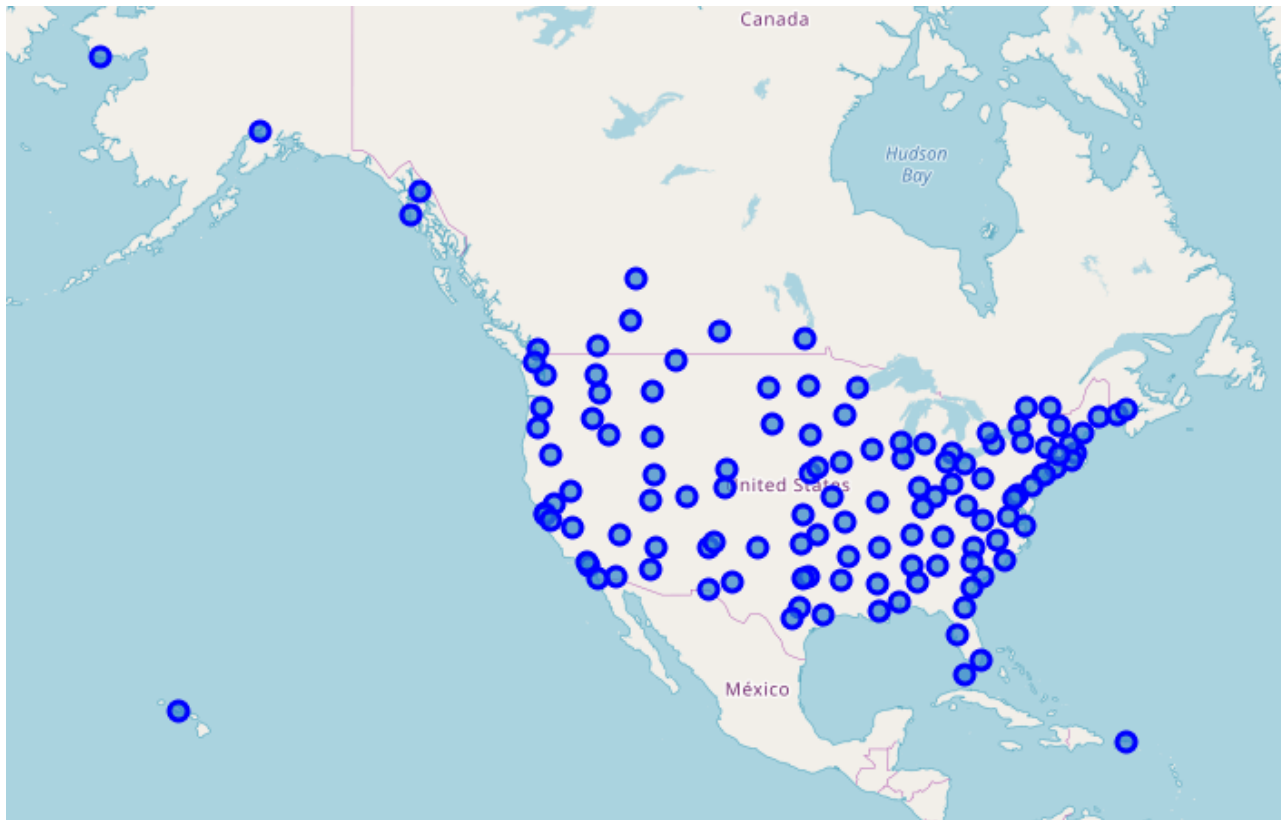
Introduction

- Problem: Businesses are not localized and want to choose markets to expand
  - Need to understand markets which they succeed and fail
    - Pursue markets where success is likely
    - Improve in market opportunities where success has not been achieved
  - Value in clustering city to reflect market similarity
    - Many factors impact a city; Each useful for analysis
    - One measurement not commonly used: similarity in businesses and venues which make up the city

- Goal: Compare similar cities based on the concentration of venue/business types in cities
  - Form clusters of major U.S. and Canadian markets to reflect similar venue/business distribution
  - Examine whether these business type composition similarities reflect geography and population characteristics

Data

- Gathered 128 major U.S. and Canadian cities
  - Ultimately used 126 for analysis
- Recorded precise latitude and longitude
- Foursquare data to determine venue composition of each city
- Used Excel to create table also including population for final analysis

Cities Recorded in the Data

Methodology

- K-means clustering
  - Unsupervised Learning – no target variable for classification
  - Group similar cities in terms of business saturation by business type
- Used k=10 means for 126 cities
  - Desire to obtain several major clusters
    - Hope to avoid too large or too small clusters

Results

- 10 Clusters
- Size of Clusters Varied
- 4 Single-City Clusters
  - 3 were among the 4 smallest cities in the dataset
    - Outliers
- 5 major clusters (above 14% of the 126 cities)
  - No cluster larger than 26%
  - Major clusters together include 96% of cities in dataset

Clusters by Size

Cluster	
1	18
2	19
3	26
4	1
5	33
6	1
7	4
8	22
9	1
10	1

Clusters and Popular Business/Venue Types for Each City Included

Cluster #1

City	Latitude	Longitude	ClusterLabels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
Birmingham, Ala.	33.653330	-86.808890	0	Fast Food Restaurant	Pizza Place	Mexican Restaurant	Gas Station
Carlsbad, N.M.	32.411940	-104.236390	0	Pizza Place	Mexican Restaurant	Hotel	Burger Joint
Cheyenne, Wyo.	41.145560	-104.801940	0	Fast Food Restaurant	Mexican Restaurant	Clothing Store	Sandwich Place
Chicago, Ill.	41.836940	-87.684720	0	Mexican Restaurant	Sandwich Place	Taco Place	Italian Restaurant
El Centro, Calif.	32.800000	-115.567000	0	Fast Food Restaurant	Pizza Place	Mexican Restaurant	Coffee Shop
Fresno, Calif.	36.750000	-119.767000	0	Mexican Restaurant	Grocery Store	Taco Place	Coffee Shop
Grand Junction, Colo.	39.067000	-108.567000	0	Mexican Restaurant	Pizza Place	Coffee Shop	Fast Food Restaurant
Havre, Mont.	48.550000	-109.683000	0	Fast Food Restaurant	Food	Pizza Place	Sandwich Place
Helena, Mont.	46.595806	-112.027031	0	American Restaurant	Fast Food Restaurant	Sandwich Place	Coffee Shop
Hot Springs, Ark.	34.497220	-93.055280	0	Fast Food Restaurant	Hotel	Pizza Place	Mexican Restaurant
Klamath Falls, Ore.	42.225000	-121.781670	0	Pizza Place	Coffee Shop	Café	Mexican Restaurant
Lewiston, Idaho	46.410000	-117.020000	0	Fast Food Restaurant	Pizza Place	Pharmacy	Taco Place
Lincoln, Neb.	40.808890	-96.678890	0	Mexican Restaurant	Convenience Store	Park	Fast Food Restaurant
Montgomery, Ala.	32.361670	-86.279170	0	Fast Food Restaurant	Sandwich Place	Pizza Place	Fried Chicken Joint
Moose Jaw, Sask., Can.	50.393330	-105.551940	0	Fast Food Restaurant	Pizza Place	Pharmacy	Coffee Shop
Nelson, B.C., Can.	49.500000	-117.283330	0	Coffee Shop	Fast Food Restaurant	Restaurant	Grocery Store
Pierre, S.D.	44.368060	-100.336390	0	Fast Food Restaurant	Pizza Place	Bar	Hotel
Santa Fe, N.M.	35.667222	-105.964444	0	Mexican Restaurant	Fast Food Restaurant	Grocery Store	Café

Cluster #2

City	Latitude	Longitude	ClusterLabels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
Austin, Tex.	30.267000	-97.733000	1	Food Truck	Hotel	Coffee Shop	Bar
Cincinnati, Ohio	39.100000	-84.517000	1	Bar	American Restaurant	Hotel	Sandwich Place
Cleveland, Ohio	41.482220	-81.669720	1	Pub	Bar	Lounge	Sushi Restaurant
Columbus, Ohio	39.983000	-82.983000	1	Bar	Pizza Place	Café	American Restaurant
El Paso, Tex.	31.759206	-106.490175	1	Bar	Coffee Shop	Mexican Restaurant	Fast Food Restaurant
Fargo, N.D.	46.877220	-96.789440	1	Coffee Shop	Bar	American Restaurant	Brewery
Flagstaff, Ariz.	35.199170	-111.631110	1	Coffee Shop	Brewery	American Restaurant	Mexican Restaurant
Jacksonville, Fla.	30.336940	-81.661390	1	Sandwich Place	Bar	Coffee Shop	Brewery
Juneau, Alaska	58.300323	-134.417639	1	Seafood Restaurant	Coffee Shop	Bar	Gift Shop
Knoxville, Tenn.	35.961700	-83.923200	1	Bar	American Restaurant	Mexican Restaurant	Hotel
Las Vegas, Nev.	36.175000	-115.136390	1	Bar	Mexican Restaurant	Gastropub	American Restaurant
Nashville, Tenn.	36.166670	-86.783330	1	Bar	Hotel	Restaurant	Music Venue
Oakland, Calif.	37.804440	-122.270830	1	Coffee Shop	Bar	Mexican Restaurant	Beer Garden
Oklahoma City, Okla.	35.482220	-97.535000	1	Bar	Pizza Place	Coffee Shop	Burger Joint
Omaha, Neb.	41.250000	-96.000000	1	Coffee Shop	Bar	Pizza Place	American Restaurant
Phoenix, Ariz.	33.450000	-112.067000	1	Coffee Shop	Hotel	Bar	Art Gallery
Shreveport, La.	32.514720	-93.747220	1	Bar	Hotel	Casino	American Restaurant
Spokane, Wash.	47.658890	-117.425000	1	Bar	Pizza Place	Coffee Shop	American Restaurant

Cluster #3

City	Latitude	Longitude	ClusterLabels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
Albuquerque, N.M.	35.110830	-106.610000	2	Hotel	Mexican Restaurant	Furniture / Home Store	Burger Joint
Atlanta, Ga.	33.755000	-84.390000	2	Hotel	American Restaurant	Aquarium	Coffee Shop
Baker, Ore.	44.779000	-117.834440	2	Coffee Shop	Hotel	Pizza Place	Fast Food Restaurant
Baltimore, Md.	39.283000	-76.617000	2	Hotel	American Restaurant	Aquarium	Seafood Restaurant
Charleston, S.C.	32.783330	-79.933330	2	Hotel	Southern / Soul Food Restaurant	Coffee Shop	Seafood Restaurant
Charlotte, N.C.	35.227220	-80.843060	2	Pizza Place	Hotel	Steakhouse	American Restaurant
Columbia, S.C.	34.000560	-81.034720	2	American Restaurant	Bar	Hotel	Coffee Shop
Dallas, Tex.	32.779170	-96.808890	2	Hotel	Steakhouse	Bar	Coffee Shop
Detroit, Mich.	42.331390	-83.045830	2	American Restaurant	Hotel	Coffee Shop	Steakhouse
Duluth, Minn.	46.788939	-92.098194	2	Hotel	Pizza Place	Brewery	Coffee Shop
Fort Worth, Tex.	32.750000	-97.333000	2	American Restaurant	Hotel	Bar	Mexican Restaurant
Idaho Falls, Idaho	43.500000	-112.033000	2	Hotel	Fast Food Restaurant	Mexican Restaurant	American Restaurant
Indianapolis, Ind.	39.768610	-86.158060	2	American Restaurant	Hotel	Steakhouse	Pizza Place
Jackson, Miss.	32.298890	-90.184720	2	Hotel	Sandwich Place	Bar	American Restaurant
Long Beach, Calif.	33.768330	-118.195560	2	Hotel	American Restaurant	Coffee Shop	Seafood Restaurant
New Orleans, La.	29.950000	-90.080000	2	Hotel	Cajun / Creole Restaurant	Cocktail Bar	Seafood Restaurant
Pittsburgh, Pa.	40.439720	-79.976390	2	Bar	Coffee Shop	American Restaurant	
Portland, Ore.	45.520000	-122.681940	2	Hotel	Coffee Shop	Bookstore	Sandwich Place
St. Louis, Mo.	38.627220	-90.197780	2	Hotel	Bar	Italian Restaurant	American Restaurant
San Antonio, Tex.	29.417000	-98.500000	2	Hotel	Mexican Restaurant	Theater	Plaza
San Diego, Calif.	32.715000	-117.162500	2	Hotel	Mexican Restaurant	Bar	Italian Restaurant
Seattle, Wash.	47.609720	-122.333060	2	Hotel	Coffee Shop	Seafood Restaurant	Sandwich Place
Sioux Falls, S.D.	43.536390	-96.731670	2	American Restaurant	New American Restaurant	Mexican Restaurant	Hotel
Sitka, Alaska	57.051561	-135.338642	2	Coffee Shop	Hotel	Trail	Zoo
Virginia Beach, Va.	36.850600	-75.977900	2	Beach	Seafood Restaurant	American Restaurant	Hotel
Wichita, Kan.	37.688890	-97.336110	2	American Restaurant	Sandwich Place	Hotel	Bar

Cluster #4

City	Latitude	Longitude	ClusterLabels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
Toledo, Ohio	41.66556	-83.57528	3	Discount Store	Intersection	Fast Food Restaurant	Art Museum

Cluster #5

City	Latitude	Longitude	ClusterLabels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
Albany, N.Y.	42.652900	-73.757220	4	Pub	Café	Coffee Shop	Pizza Place
Anchorage, Alaska	61.217000	-149.900000	4	Seafood Restaurant	Park	Coffee Shop	Clothing Store
Boise, Idaho	43.617000	-116.200000	4	Coffee Shop	Pizza Place	Hotel	Brewery
Boston, Mass.	42.358060	-71.063610	4	Italian Restaurant	Park	Pizza Place	Bakery
Calgary, Alta., Can.	51.050000	-114.067000	4	Coffee Shop	Steakhouse	Hotel	Restaurant
Des Moines, Iowa	41.590830	-93.620830	4	Coffee Shop	Italian Restaurant	Hotel	Bar
Edmonton, Alb., Can.	53.533000	-113.500000	4	Café	Bar	Coffee Shop	Italian Restaurant
Eugene, Ore.	44.051940	-123.086670	4	Brewery	Coffee Shop	Pizza Place	Thai Restaurant
Grand Rapids, Mich.	42.961110	-85.655660	4	Coffee Shop	American Restaurant	Brewery	Café
Honolulu, Hawaii	21.300000	-157.817000	4	Japanese Restaurant	Coffee Shop	Bakery	Dessert Shop
Houston, Tex.	29.762780	-95.383060	4	Park	Trail	Coffee Shop	Pizza Place
Kansas City, Mo.	39.099720	-94.578330	4	Coffee Shop	American Restaurant	Music Venue	Brewery
Kingston, Ont., Can.	44.233000	-76.500000	4	Pub	Coffee Shop	Café	Bar
Los Angeles, Calif.	34.050000	-118.250000	4	Coffee Shop	Bar	French Restaurant	Italian Restaurant
Minneapolis, Minn.	44.983000	-93.267000	4	Coffee Shop	Park	Theater	Music Venue
Newark, N.J.	40.720000	-74.170000	4	Portuguese Restaurant	Brazilian Restaurant	BBQ Joint	Lounge
New Haven, Conn.	41.310000	-72.923610	4	Pizza Place	Coffee Shop	American Restaurant	Italian Restaurant
Ottawa, Ont., Can.	45.424720	-75.695000	4	Coffee Shop	Hotel	Restaurant	Mexican Restaurant
Philadelphia, Pa.	39.952780	-75.163610	4	Coffee Shop	Bar	Italian Restaurant	Wine Bar
Portland, Maine	43.667000	-70.267000	4	Coffee Shop	Brewery	American Restaurant	Bar
Providence, R.I.	41.823610	-71.422220	4	Italian Restaurant	Pizza Place	Bar	American Restaurant
Raleigh, N.C.	35.767000	-78.633000	4	Italian Restaurant	Cocktail Bar	Music Venue	Coffee Shop
Reno, Nev.	39.527220	-119.821940	4	Bar	Pub	Coffee Shop	Breakfast Spot
Sacramento, Calif.	38.555560	-121.468890	4	Coffee Shop	Mexican Restaurant	Vietnamese Restaurant	American Restaurant
St. John, N.B., Can.	45.280560	-66.076110	4	Coffee Shop	Park	Grocery Store	Hotel
Salt Lake City, Utah	40.750000	-111.883000	4	Coffee Shop	Bar	Thai Restaurant	Vegetarian / Vegan Restaurant
San Francisco, Calif.	37.783000	-122.417000	4	Coffee Shop	Sushi Restaurant	Marijuana Dispensary	Gym / Fitness Center
San Jose, Calif.	37.333333	-121.900000	4	Mexican Restaurant	Bar	Coffee Shop	Cocktail Bar
Syracuse, N.Y.	43.048940	-76.144440	4	Coffee Shop	Bakery	Italian Restaurant	Pizza Place
Vancouver, B.C., Can.	49.250000	-123.100000	4	Coffee Shop	Vietnamese Restaurant	Arts & Crafts Store	Indian Restaurant
Victoria, B.C., Can.	48.428610	-123.365560	4	Coffee Shop	Restaurant	Breakfast Spot	Vegetarian / Vegan Restaurant
Washington, D.C.	38.906188	-77.017263	4	Cocktail Bar	Coffee Shop	Bar	Italian Restaurant
Winnipeg, Man., Can.	49.899440	-97.139170	4	Coffee Shop	Asian Restaurant	Café	Hotel

Cluster #6

City	Latitude	Longitude	ClusterLabels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
Eastport, Maine	44.91361	-67.00389	5	Seafood Restaurant	Food	State / Provincial Park	Bakery

Cluster #7

City	Latitude	Longitude	ClusterLabels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
Key West, Fla.	24.559720	-81.783610	6	Hotel	Cuban Restaurant	Resort	Bed & Breakfast
Montreal, Que., Can.	45.508890	-73.561670	6	Café	French Restaurant	Hotel	Restaurant
New York, N.Y.	40.661000	-73.944000	6	Caribbean Restaurant	Café	Bakery	Cocktail Bar
San Juan, P.R.	18.451522	-66.069481	6	Caribbean Restaurant	Hotel	Italian Restaurant	Restaurant

Cluster #8

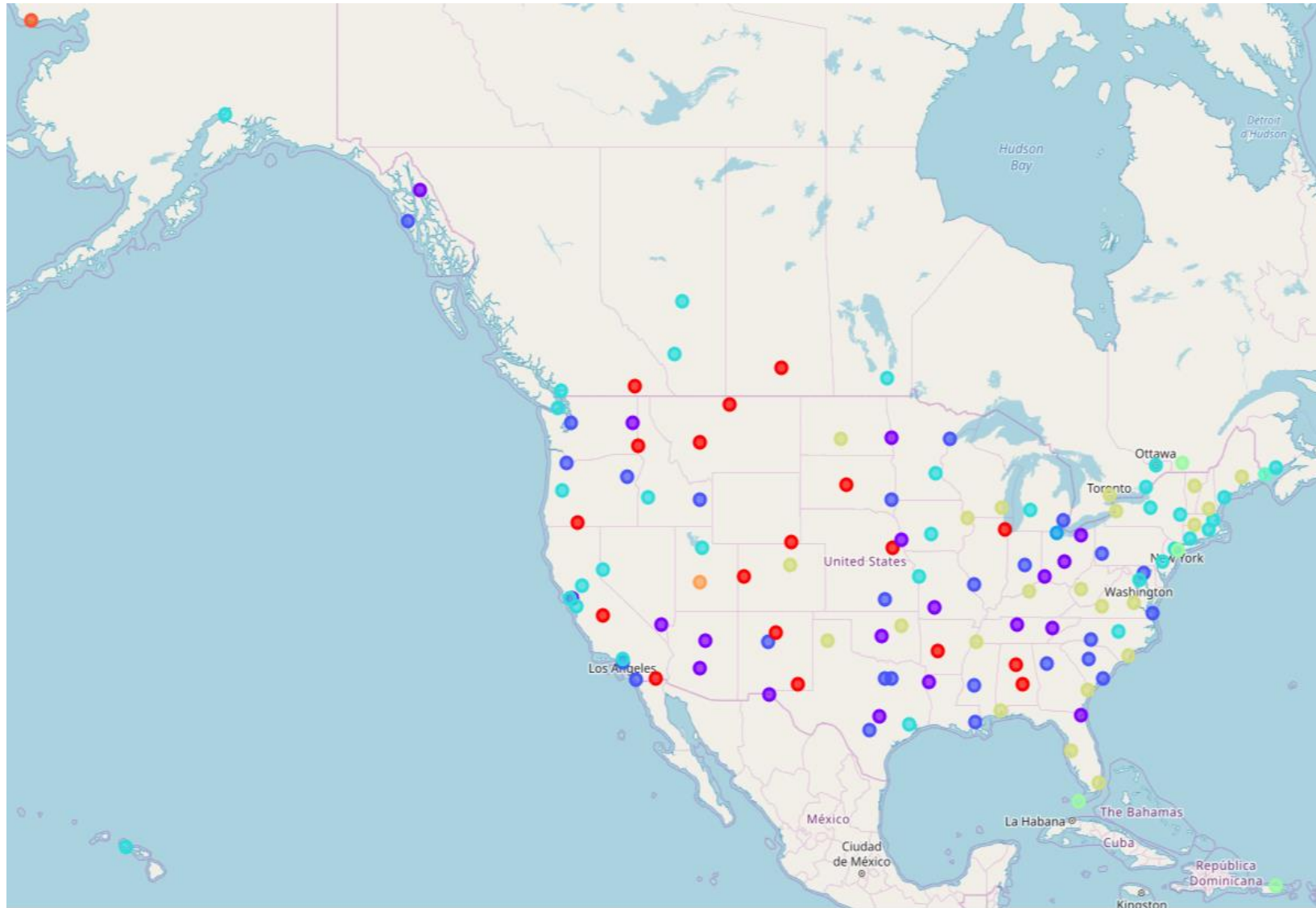| City	Latitude	Longitude	ClusterLabels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
Amarillo, Tex.	35.19917	-101.845280	7	Sandwich Place	Restaurant	American Restaurant	Bank
Bangor, Maine	44.80000	-68.800000	7	Rental Car Location	Pizza Place	Hotel	Airport Terminal
Bismarck, N.D.	46.81333	-100.778890	7	Clothing Store	Pizza Place	Coffee Shop	Sandwich Place
Buffalo, N.Y.	42.90472	-78.849440	7	Discount Store	Bar	Intersection	Gay Bar
Charleston, W. Va.	38.34722	-81.633330	7	Pizza Place	Bar	Discount Store	American Restaurant
Denver, Colo.	39.76185	-104.881105	7	Coffee Shop	Pool	American Restaurant	Sandwich Place
Dubuque, Iowa	42.50000	-90.690000	7	Pizza Place	Bar	Mexican Restaurant	Coffee Shop
Louisville, Ky.	38.22533	-85.741700	7	Bar	Pizza Place	Coffee Shop	Sandwich Place
Manchester, N.H.	42.99083	-71.463610	7	Café	American Restaurant	Pizza Place	Donut Shop
Memphis, Tenn.	35.11750	-89.971110	7	Discount Store	Bar	Café	Convenience Store
Miami, Fla.	25.77528	-80.208890	7	Smoke Shop	Seafood Restaurant	Mexican Restaurant	Bar
Milwaukee, Wis.	43.05000	-87.950000	7	Bar	American Restaurant	Sandwich Place	Grocery Store
Mobile, Ala.	30.69444	-88.043060	7	Intersection	Seafood Restaurant	American Restaurant	Southern / Soul Food Restaurant
Montpelier, Vt.	44.26000	-72.575280	7	Gas Station	Convenience Store	Hotel	Thai Restaurant
Richmond, Va.	37.53300	-77.467000	7	Park	American Restaurant	Pizza Place	Coffee Shop
Roanoke, Va.	37.27083	-79.941670	7	Coffee Shop	American Restaurant	Sandwich Place	Park
Savannah, Ga.	32.01700	-81.117000	7	Department Store	Furniture / Home Store	Hotel	Fast Food Restaurant
Springfield, Mass.	42.10139	-72.590280	7	Donut Shop	Sandwich Place	Gas Station	American Restaurant
Tampa, Fla.	27.96806	-82.476390	7	Cuban Restaurant	Park	Coffee Shop	Spanish Restaurant
Toronto, Ont., Can.	43.74167	-79.373330	7	Coffee Shop	Park	Shopping Mall	Sandwich Place
Tulsa, Okla.	36.13139	-95.937220	7	Sandwich Place	Fast Food Restaurant	Burger Joint	Pizza Place
Wilmington, N.C.	34.22333	-77.912220	7	Fast Food Restaurant	Shoe Store	Department Store	Clothing Store

Cluster #9

City	Latitude	Longitude	ClusterLabels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
Richfield, Utah	38.76583	-112.0875	8	Pizza Place	Steakhouse	Fast Food Restaurant	Sandwich Place

Cluster #10

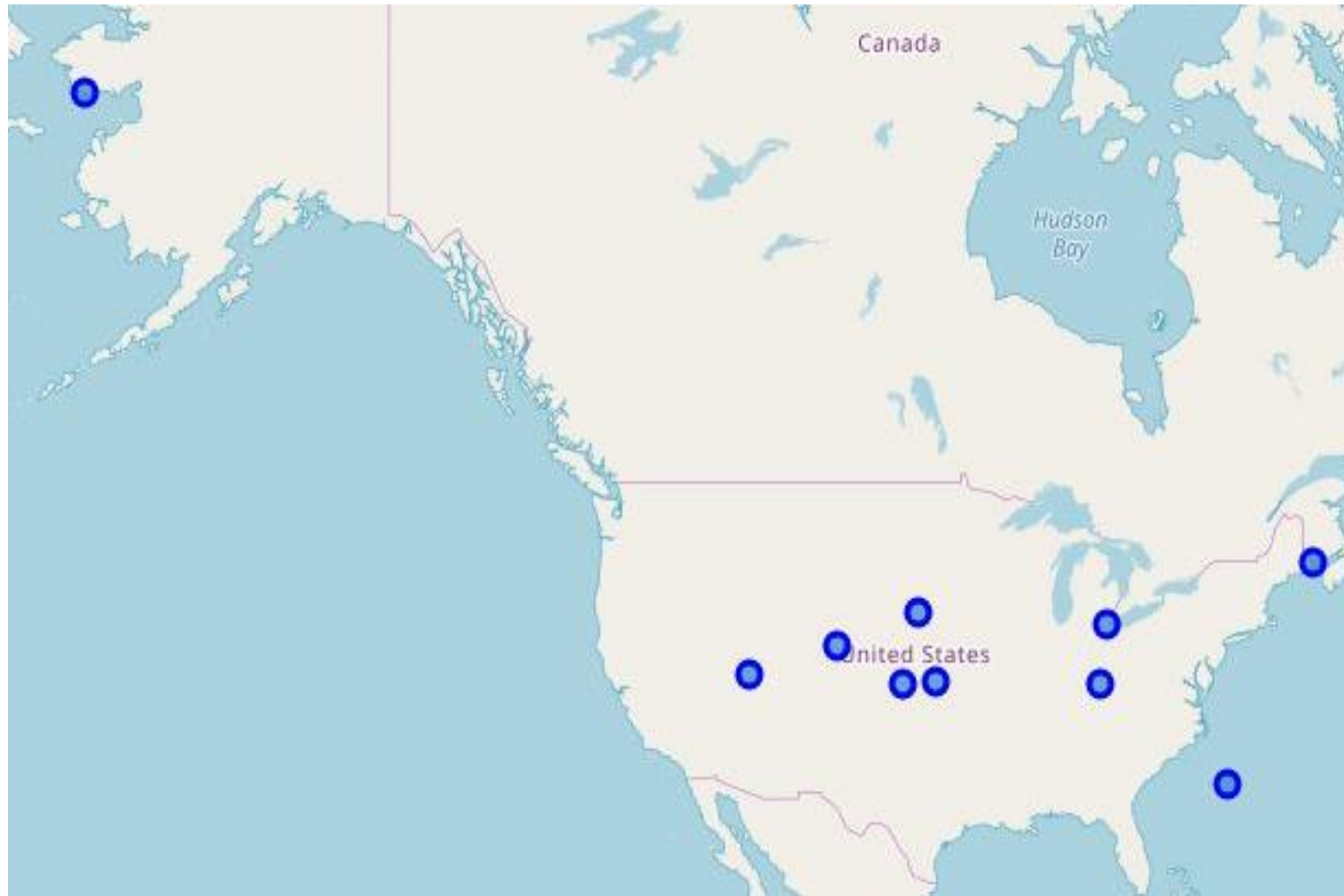| City	Latitude	Longitude	ClusterLabels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
Nome, Alaska	64.50389	-165.39944	9	Hotel	Grocery Store	Bakery	Restaurant

Map of 10 Means Clustering Results

Discussion

- ► **Examination of Clusters**
  - ► Geography
  - ► Population
- ► **Unusual Clusters**
  - ► 4 Single-City
    - ► Abnormal cities – 3 of 4 smallest in dataset
  - ► 1 Cluster of 4 Cities
    - ► Popular Tourism Cities
    - ► Less Geographic Diversity
    - ► Unusual Population Variation
      - ► Contained New York City
        - ► Outlier among small group of data (4 cities)

Geography – Centers of Each Cluster

Geography – Standard Deviations in Latitudes and Longitudes of Clusters

- ▶ 4 Not Computed (Single City)

- ▶ 5 Major Clusters and 1 Small Cluster

- ▶ Minimum Standard Deviation Within Clusters

  - ▶ 5.7 Latitude (≈393 miles North/South)

  - ▶ 6.42 Longitude (≈340 miles East/West)

    - ▶ Small Sample (East Coast, tourism heavy cities)

    - ▶ Among major clusters: 10.04 Longitude (≈532 miles East/West)

- ▶ Substantial geographic diversity in clusters evident

  - ▶ Similarities in business composition and venue type saturation due to factors beyond geographic proximity

Cluster	Pop	Lat	Long
1	627969.10	6.18	11.39
2	397486.81	6.85	15.44
3	422432.05	6.55	16.70
4	nan	nan	nan
5	786382.41	6.96	24.86
6	nan	nan	nan
7	3987689.99	12.86	6.42
8	573813.13	5.70	10.04
9	nan	nan	nan
10	nan	nan	nan

Standard Deviation within Each Cluster

Population

- Focus on 5 major clusters
  - Enough data to analyze good sample of city sizes
- Similar Means
- Large Standard Deviations
- Clustering by business/venue type composition goes beyond population

Cluster	Pop
1	244071.33
2	501913.63
3	523851.77
4	278508.00
5	623328.55
6	1219.00
7	2656107.75
8	382253.91
9	7723.00
10	3797.00

Mean Population within Each Cluster

Cluster	Pop
1	627969.10
2	397486.81
3	422432.05
4	nan
5	786382.41
6	nan
7	3987689.99
8	573813.13
9	nan
10	nan

Standard Deviation City Population within Each Cluster

Conclusion

- 5 major clusters can be used to assess similarity of cities in terms of business composition in the city/metro area

- Clusters could be associated with commonly used similarity metrics (geographic proximity and population similarity), but also are not heavily linked to either

  - Holistic analysis would be best

- Clustering by business/venue composition another tool in market analysis

  - Useful to combine with other tools for best and most rigorous analysis of North American markets

References

▶ Aklson, A. (n.d.). Applied Data Science Capstone. Retrieved from https://www.coursera.org/learn/applied-data-science-capstone/home/welcome

▶ Latitude and Longitude of U.S. and Canadian Cities. (n.d.). Retrieved from https://www.infoplease.com/world/united-states-geography/latitude-and-longitude-us-and-canadian-cities

▶ Rosenberg, M. (2018, September 28). How to Figure Out the Distance Between Degrees of Latitude and Longitude. Retrieved from https://www.thoughtco.com/degree-of-latitude-and-longitude-distance-4070616