Rochester Minnesota Clustering Analysis

I. Introduction

Rochester, Minnesota USA is a city in the southeaster part of the state of Minnesota in Olmsted County. It is primarily known as the home of the world renown Mayo Clinic which serves over 1 million patients per year and conducts crucial medical research and training.

In 2013 the State Government of Minnesota initiated a "Destination Medical Center" initiative intended to invest over 5 billion dollars into infrastructure and expansion in Olmsted County to further support the Economic impacts of the Mayo Clinic. In 20 years' time the area's population is expected to have grown between 50 and 100%.

The increased population and increased patient volume expected presents key business problems across all sectors of the regional economy. All types of businesses from lodging, housing, business services, restaurants, and so on will need to be opened to handle the expansion. Cluster analysis may help to identify which types are missing in each natural cluster.

II. Data

The data for this project needs to come from three sources.

- There is a online database of zip codes and population data housed at https://www.zip-codes.com/county/mn-olmsted.asp This gives a breakdown of the county and metro area.
- The "GeoPy" package in Python will allow us to append latitude and longitude data to our zip code and population data from the prior source
- Finally, with the Foursquare API we can pull popular venues of interest for each of the latitude and longitude values from the DataFrame

III. Methodology

The data will be collected, merged, and cleaned within pandas DataFrames.

The primary analytic tool in use for this project will be k-means clustering. The k-means method generates k clusters based on a chosen distance metric for the relevant dataset. I will attempt various sizes of clusters such as 3, 4, or 5 groupings to see which presents the most feasible set of clustering for the county and city data.

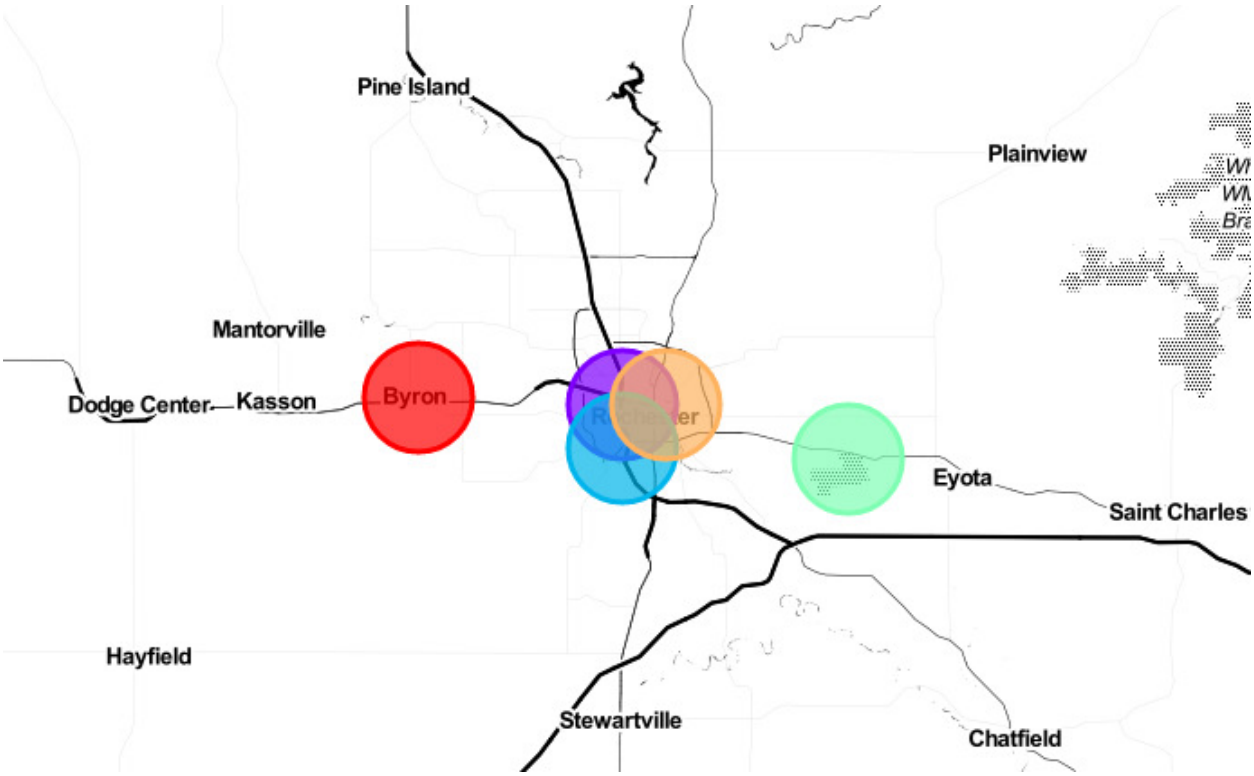
Once the clustering is completed the top venues from Foursquare can be studied to understand what is currently available in each cluster and what appears to be missing. This would allow an entrepreneur to determine which type of business would be best suited to each location.

IV. Results

The following venue types were found in each neighborhood grouping

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Byron	Business Service	Steakhouse	Convenience Store	Ice Cream Shop	Bar	Restaurant	Music Venue	Supplement Shop	Cosmetics Shop	Coffee Shop
1	Dover	Business Service	Bar	Rental Service	Post Office	Supplement Shop	Department Store	Cosmetics Shop	Convenience Store	Coffee Shop	Clothing Store
2	Eyota	Chinese Restaurant	Park	Supplement Shop	Business Service	Department Store	Cosmetics Shop	Convenience Store	Coffee Shop	Clothing Store	Campground
3	Northeast	Baseball Stadium	Baseball Field	Playground	Supplement Shop	Business Service	Department Store	Cosmetics Shop	Convenience Store	Coffee Shop	Clothing Store
4	Northwest	Coffee Shop	Hotel	Grocery Store	Italian Restaurant	BBQ Joint	Bakery	Breakfast Spot	Building	Chinese Restaurant	Storage Facility
5	Southeast	Campground	Park	Supplement Shop	Business Service	Department Store	Cosmetics Shop	Convenience Store	Coffee Shop	Clothing Store	Chinese Restaurant
6	Southwest	Coffee Shop	Hotel	Supplement Shop	Organic Grocery	Clothing Store	Convenience Store	Cosmetics Shop	Department Store	Discount Store	Fast Food Restaurant

The clustering result was as follows



V. Discussion

The clustering matched intuitive knowledge of the Rochester Area. Largely the results were as expected, though it was somewhat surprising to see the suburban neighborhoods to the Northwest of the City Limits were not clustered with Northwest Rochester. They appear to be relatively homogenous without detailed analysis.

VI. Conclusion

1. Cluster Number 1 encompassing Northwest Rochester, Byron, and Dover is lacking in activities such as campgrounds or sports facilities compared to the others. These are likely good business options.
2. Cluster Number 2 encompassing Southwest Rochester seems to have a balance of all types, any well-run business would be appropriate.
3. Cluster Number 3 encompassing Eyota seems to lack restaurant venues, this may be a good option in this area.
4. Cluster Number 4 encompassing Southeast Rochester is primarily services and shops. Lodging or Restaurants may be optimal.
5. Cluster Number 5 encompassing Northeast Rochester seems to be underdeveloped from a business perspective. Any options would likely be feasible