Understanding the Investment Landscape in Detroit Using Foursquare Data

Coursera Capstone Project

2/6/2021

Chris Olson

Introduction

Over the past several years, the city of Detroit has continued to see investment in a variety of forms. These investments come after decades of population decline in the city, deindustrialization in what had been one of largest manufacturing centers in the US, and the city's bankruptcy in 2013. The city is looking forward to a brighter future however with the Little Caesars Arena opening in 2017 and the Q line opening in the same year. These major investments and others will hopefully spur continued investment in the city as more companies move from the suburbs of Detroit into the city proper. While the city has historically benefited from the clustering of the auto industry and continues to do so to some extent, the city must look to other industries to fuel its economic growth. The affordability of the city, and large metropolitan region provides an opportunity for businesses and amenities to move into the city. Understanding where different categories of businesses currently exist in the city will enable potential investors to either locate in a section of the city where competition may be low, or they may wish to locate in a section of the city where their business may benefit from business clustering. For groups and individuals concerned about social justice, understanding the current distribution of venues and their type in the city can provide valuable information about the unequal investment that the city has seen. This information may be used to fine tune city policy on a more granular level thus distributing investment more evenly in the city.

Overview of the Data

The primary data source for this project is Foursquare and the data queried from the Foursquare API. The names of the neighborhoods and the longitude and latitudes for the neighborhoods were acquired through shapefiles from the city of Detroit and then converted to longitude and latitude. Foursquare data does not capture every business per se but rather it captures venues. A venue is not the same as a business, rather is a place that someone visiting the city of Detroit may wish to go. Examples of venues which are not business would be parks or museums which are captured in the data because the data is about where people using the foursquare app want to go. Crucially, this venue data included the name of the venue, its location and which category it belongs to. In considering how this data relates back to the problems outlined in the introduction, for the business types which are captured in the data, coffee shops for example, they can use this data to either cluster with other coffee shops or they may wish to locate in a neighborhood they identify as underserved by coffee shops. Businesses looking to relocate to the city may wish to know what venues/amenities are available in the area in which they may locate as this will aid in attracting and retaining employees. Notably, in considering where to place its second headquarters, Amazon evaluated the amenities of the surrounding area as a criterion when making its decision. The location and distribution of certain types of venues throughout the city have social justice implications as large portions of the city may be lacking many venues and venue types relative to other parts of the city. The implications of this will vary based on the particular venue, we might expect that museums are clustered together in a central district but given the economic benefits that accompany a museum and other urban planning considerations, we might wish to see them more spread out amongst the neighborhoods of a city.

Methodology

Once the data was queried, I used the existing categorization provided by Foursquare in order to categorize each venue. I grouped the data by neighborhood and performed analysis on the data in two different ways. To show larger and broader tends in the city of Detroit, I simply counted how many venues there were per neighborhood. This simple analysis itself was able to produce visible segmentation of the city between the core areas which have received more interest and investment as compared to other peripheral areas which have received none. A large number of neighborhoods had only a handful of venues and 14 of the 208 neighborhoods had no venues at all.

To create a more granular set of data, I used one hot encoding based on the category assigned to the venue by Foursquare. I then converted this encoded data to a pivot table using neighborhood as an index which gave me the total counts of each venue type for each neighborhood including which venue types were not present in each neighborhood represented by zeros in the pivot table. This is as granular as the data would allow me to go and presents a both a wholistic look at neighborhood investment conditions since all types of venues can be seen. This pivot table also shows a sector specific look since the table could be used to count the number of restaurants in a neighborhood or even the number of restaurants of a certain cuisine in an area.

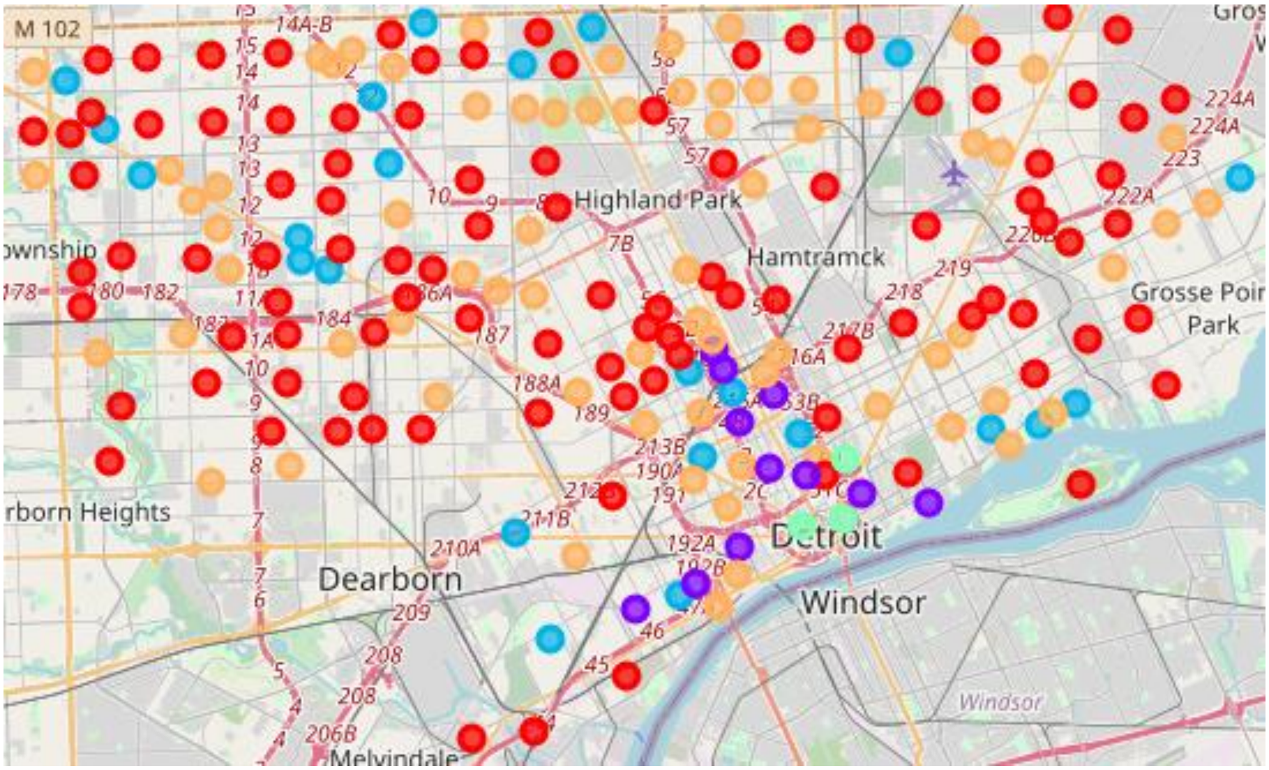
To segment the city into clusters, I used the K-means algorithm since this allowed control over how many segments could be produced. While I attempted to use the DBSCAN algorithm in order to make inferences about similar economic/investment conditions which may be occurring in the city, the algorithm failed to produce meaningful segmentation. With the higher-level venue data, the parameters of the algorithm had to be set very high such that the algorithm lost meaning in order to produce a small enough number of clusters in order to make inferences. With the more granular pivot table data, the algorithm either assigned the entire city to a single cluster or assigned too many segments to be meaningful.

While using K-means, I segmented the city into 5 sections as this produced understandable groupings, however neighborhoods could possibly be grouped into fewer segments without losing meaning given most of the city lacks venues. Using K-means did, however, produce one notable anomaly. Since the algorithm has a randomized start, there was one neighborhood, Corktown, which was initially listed as its own cluster but when the algorithm was rerun it was re-segmented to no longer be its own cluster. In rerunning the algorithm, I did not encounter this result again, but I do believe it might be considered meaningful given the attention Corktown has received with the historic Michigan Central Station currently being restored by the Ford Motor Company.

Additionally, in an attempt to create some form of timeseries I queried Foursquare data from 2018 and 2021, there were no differences in the results, however, as only a few new venues were added to the entire city.

<u>Results</u>

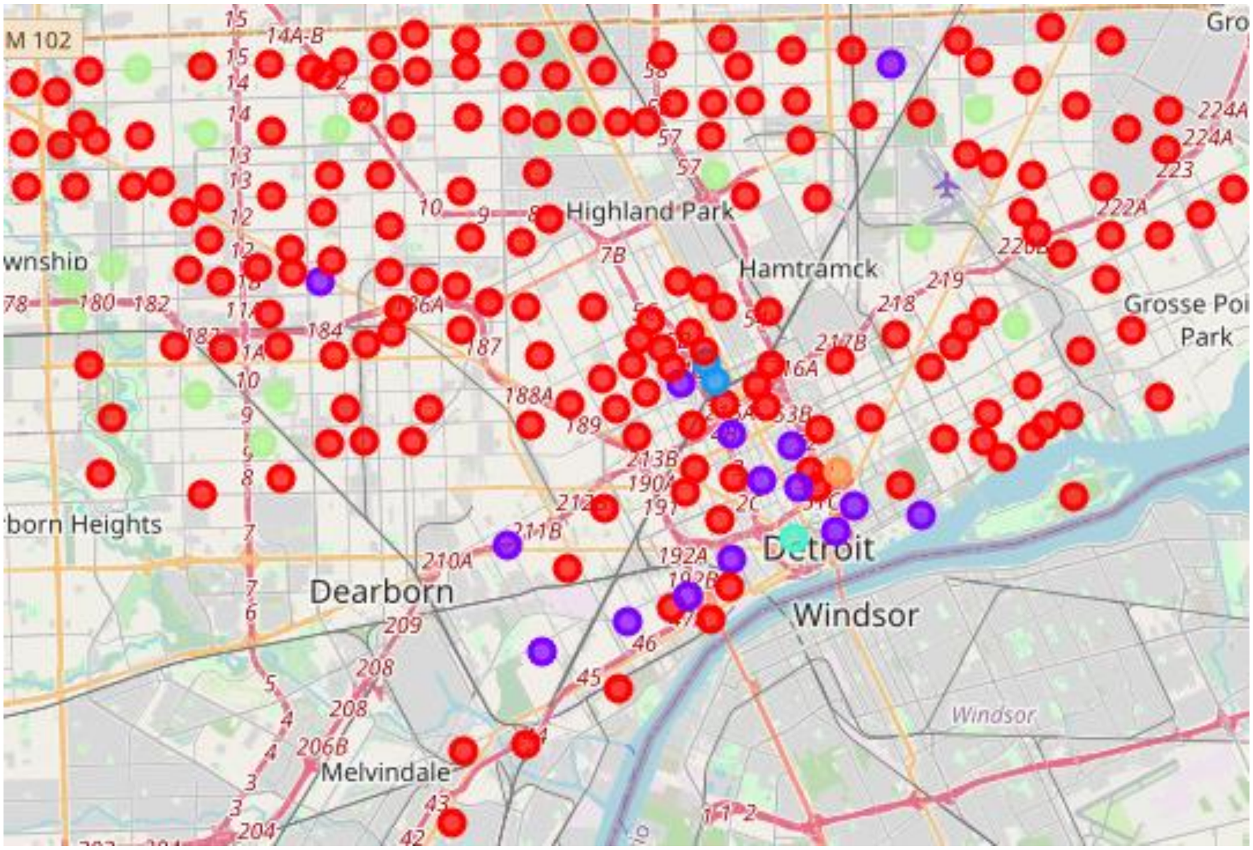
Maps produced by the K-means segmentation show a city that is highly segmented with the core downtown area of the city of Detroit holding most of the venues according to the Foursquare data with areas surrounding this core holding more venues than the periphery of the city.



Map showing the results of K-means applied to the venue counts data of the city

Color	Avg Number of Venues
Green	47
Purple	23.1
Blue	9.6
Orange	5
Red	1.7

As can be seen from the map and the corresponding table which lists the average number of venues per color coded cluster, most of the venues in the city are concentrated downtown. This is what we could expect to find in most any city, however when we consider the city of Detroit, we must keep in must the geographic size of the city. The city of Detroit is 142.9 square miles, making it the 64th largest city by area in the United States and it could be reasonably expected for a city of this size that there would more than one node where there would be a high density of venues.



Map produced by pivot table data, please note that the colors do not convey the same meaning as the pervious map as the colors are assigned based on the K-means cluster label. In one iteration of the algorithm Corktown (now in purple) was listed as its own label.

Purple	
Venue Type	Average Venues per Cluster Neighborhood
Pizza Place	.87
Bar	.87
American Restaurant	.8

Blue	
Venue Type	Average Venues per Cluster Neighborhood
Plaza	1.5
Chinese Restaurant	1
Hotel	1

Light Blue/Downtown	
Venue Type	Average Venues per Cluster Neighborhood
Coffee Shop	7
Lounge	5
Park	4

Orange/Eastern Market	
Venue Type	Average Number of Venues per Cluster Neighborhood
Farmers Market	6
Coffee Shop	3
Food	2

Red	
Venue Type	Average Venues per Cluster Neighborhood
Liquor Store	.18
Intersection	.13
Discount Store	.12

Tables above show the 3 most frequent venues types per venue cluster

Looking at the more granular pivot table data, we can see that venues, and as we are attempting to infer investment, is even more concentrated in the downtown area. We can see that the downtown area contains a high number of venues while the areas surrounding it have markedly less but still higher than the rest of the city. The rest of the city by contrast we know lacks very many venues as seen in the previous map and the tables show that one of these venues is intersections, which while Foursquare captures as a category, it seems unlikely this is a destination which draws people. The green neighborhoods have no venues of any kind and were thus manually assigned a cluster because they were not picked up by the pivot table or second running of the K-means algorithm.

Discussion

The Foursquare data would tell us that Detroit is a highly segmented city with most of its investments/venues and amenities located in the downtown while the periphery and majority of the geographic area of the city lack these venues. Indeed, we see this pattern in the news as well. We should be cautious in extrapolating this data too far however as more information is needed to assess whether the low venue counts in certain neighborhoods may be attributable to some other factors such as Foursquare's data collection. We can see on the respective maps that the Bell Isle park, famously designed by Frederick Law Olmsted, there is only one venue, the park itself. A visit to this park however would reveal that it contains numerous venues such as a botanical garden, one of the nation's oldest aquariums and several others. There are likely other examples of where the Foursquare data may not be as accurate or detailed as we would like it to be. Nevertheless, the data corroborates what we might have expected to find and news media reporting on the city.

Conclusion

Despite potential weaknesses and limitations in the data, it is still telling of what exists in the city. We can safely say in interpreting this data for the various stakeholders, businesses looking to relocate to Detroit will want to locate in downtown areas of the city where there are more venues. Alternatively, these businesses can look at up and coming areas which surround the downtown such as Corktown which is continuing to see more investment and redevelopment. Community groups will seek a more even distribution of venues thought the city to the benefit of all Detroiters. While forecasts are difficult to make, this data and understanding of the current landscape of the city give us some idea of where the city is going.