Access to retail and home-goods in Detroit, MI

A project using skills learned in the IBM Data Science Professional Certificate program on Coursera.

Defining the issue and stakeholders:

One of the issues that some non-profit groups in Detroit address is the presence of food deserts. The wiki page for 'food desert' (https://en.wikipedia.org/wiki/Food_desert) provides an excellent and detailed description of this concept while the opening paragraph offers a succinct definition: "*A food desert is an area that has limited access to affordable and nutritious food,in contrast with an area with higher access to supermarkets or vegetable shops with fresh foods, which is called a food oasis.*" The issue of food deserts has already received a large amount of attention, research, and data mapping and so I will not be looking at that issue but by way of the food desert topic, I will transition to a similar issue that has received less attention: The lack of access to other retail.

In my particular situation I was a Detroit homeowner of a 100 year old house that required quite a few supplies for maintenance and upkeep. In addition, I am a gardener and this hobby also requires supplies. In both cases, there were minimal local options and a typical supply trip involved driving a minimum of 16 miles round-trip and a minimum of 45 minutes drive-time depending on traffic conditions.

With this experience as the motivator, the purpose of this project is to map location based access to hardware, homegoods, and bigbox retail shops that provide needed supplies for home maintenance and daily living. After the initial mapping, use the k-means method to group similar neighborhoods with a final outcome of providing an initial assessment of where there may be a need for non-profits to explore how to provide these supplies to under-served communities.

Description of the Desired Data and Data Collection:

Use the Four Square API to retrieve neighborhood specific data using desired venue categories as defined in the Foursquare Venue Category Hierarchy. (https://developer.foursquare.com/docs/build-with-foursquare/categories/) The api request will use the 'search?' end point and include the 'categoryId=' option.

Collecting geo-coded information:

1. The city of Detroit has identified neighborhoods and provides an online map. https://detroitmi.gov/webapp/interactive-district-map
2. Using this map, I cross referenced two other websites: https://www.zipmap.net/Michigan/Wayne_County/ and https://www.latlong.net/. These two sites provide a zipcode map of Detroit and the latitude and longitude of any location by way of dropping a pin.
3. The result of step 2 is a csv file of zipcode, neighborhood name, latitude and longitude. https://docs.google.com/spreadsheets/d/e/2PACX-1vQnWbnmwrC1bM1MLRu3SgFc_9UJZzBuo0lZV34YvkeZGuPhdrDb_3AP-BDp3d-6-lSKAu0m5rJaSTH6/pub?gid=590577091&single=true&output=csv

Venue Categories:
Big Box Store, 52f2ab2ebcbc57f1066b8b42 (to include Home Depot, Lowes, and similar)
Department Store, 4bf58dd8d48988d1f6941735 (to include Target and similar)
Discount Store, 52dea92d3cf9994f4e043dbb (to include Dollar stores)
Furniture / Home Store, 4bf58dd8d48988d1f8941735
Garden Center, 4eb1c0253b7b52c0e1adc2e9
Hardware Store, 4bf58dd8d48988d112951735
Kitchen Supply Store, 58daa1558bbb0b01f18ec1b4
Pharmacy, 4bf58dd8d48988d10f951735
Supermarket, 52f2ab2ebcbc57f1066b8b46 (to include Meijer and Walmart)
Warehouse Store, 52e816a6bcbc57f1066b7a54 (to include Costco and Sams Club)

Outcome: Use k-means clustering to visualize neighborhoods clustered by retail access.

Methodology:
1. Using the data created above, retrieve venue data for all Detroit neighborhoods.
2. The data presented two issues. One was duplicates due to overlapping results from closely situated neighborhoods. The second was poorly matched venues. For example, there are many plumbing supply shops assigned to the hardware store category that are not hardware stores.
3. Cleaned the data to remove duplicates and poorly matched venues. The first issue was resolved using Python while the second required manual review.
4. Created a map that plotted the neighborhoods and venues.
5. The first set resulted in 80 venues and a map with many empty areas. To address this problem, I repeated the search twice, with a radius of 1km and 1.5km. The third data-set resulted in 238 venues and a map with decent coverage over the whole city.
6. Transformed the data into a ranked list of venues per neighborhood.
7. Used k-means testing to cluster the neighborhoods into similar groups. I performed the test with k=3,4,5,6, and 7 with k=6 providing the best cluster set.
8. Mapped the clusters.
9. One clear result of this testing is that dollar stores, pharmacies, and hardware stores constitute the majority of retail access and so I created a data-set and map that excluded these three.

Results:
1. The city of Detroit has identified 200 neighborhoods.
2. Foursquare returns 238 venues within 1.5km of the geo-coded center of these 200 neighborhoods.
3. After grouping there are 92 neighborhoods with venues (46%) and 108 with none. (54%)
4. Venue Count: 113 dollar stores (47.5%), 80 pharmacies (33.6%), 30 hardware (12.6%), 8 big box (3.4%), 4 department (1.7%), 3 garden (1.3%).
5. The big three of dollar, hardware, and pharmacy stores account for 93.7% of venues.
5. Excluding the big three: 25 venues, 20 neighborhoods (10%).

Summary Discussion: The residents of Detroit are faced with an extreme lack of access to retail shops offering needed supplies for home maintenance and daily living.

Detroit is a large city with an approximate population of 672,00 and yet there are only 8 big box stores and 4 large department stores serving the entire city. From a total of 200 neighborhoods, 108 have no available retail, representing 54% of all neighborhoods. Of the 46% that do have retail access, 94% is provided by national dollar store chains Family Dollar, Dollar General, and Dollar Tree; pharmacy chains CVS, Rite-Aid, and Walgreens; hardware franchises True Value, ACE, and ACO; and a few other smaller chains or independent shops.

This market situation is significant in that it represents limited product selection, lower quality product, and markedly increased pricing. It is outside the scope of this project to provide a detailed analysis of these limitations but if someone wishes to get a sense of the problem, making a visit to one each of a dollar store chain, pharmacy chain, and local hardware store will be sufficient.

Conclusion: The purpose of this project was to research retail access in Detroit with the outcome being a decision as to whether there was a need for help from non-profit organizations. The result is a clear positive. The city of Detroit would benefit from alternative forms of access to supplies for home maintenance and daily living. Of particular interest is the 54% of neighborhoods with no access. One non-profit who has successfully implemented a localized supply system is the Detroit Agricultural Network through their Keep Growing Detroit project. https://detroitagriculture.net/ Further study of this group's work and a meeting with the organizers is a recommended next step for any non-profit wishing to develop a similar supply chain for other products.