Identifying Food and Healthcare Deserts and their Impact

Brad Mills

July 8, 2020

1. Introduction

1.1 Background

The concept of food deserts is a growing problem that has been highlighted by the United States Department of Agriculture (USDA) and can be defined as geographic areas where a resident's access to affordable, healthy food options, especially fresh fruits and vegetables, is limited or nonexistent due to the absence of grocery stores within convenient travelling distance.[1] Similarly, medical or healthcare deserts are described as regions where there is inadequate access to one or more kinds of medical services.[2] Beyond the inconvenience factor of not having readily available access to basic human needs, the absence of healthy food and/or healthcare can have a broad negative impact on a neighborhood and those living in it.

1.2 Problem/Idea

Identifying food deserts and healthcare deserts by using location data such as what is available through the Foursquare API should be achievable with a little creativity and elbow grease, aka data wrangling. This neighborhood data could further be linked to educational data from a different source to explore whether it is indicative of other issues, e.g. lower graduation rates, so that potential problems can be identified early and addressed through corrective interventions.

1.3 Interests

Local governments, grocers, healthcare institutions, and educators should all be interested in the timely identification of food and healthcare deserts within their communities. Local governments could use the information to get out ahead of urban planning and seek to recruit new businesses to their area. Grocery stores could use the data to look for new areas of possible expansion, as could healthcare providers on the medical side. And if a correlation can be seen between food or healthcare deserts and low graduation rates, school administrators and educators would be interested in early detection of problem areas and the clustering of similar neighborhoods within their school district to help identify students that might need additional assistance due to the other challenges that often accompany living in a food and healthcare desert.

2. Data

2.1 Data sources

The source of location data for this capstone project will be calls to the Foursquare API. Based on an initial review of the approximately 330 unique Venue categories observed

in the NYC and Toronto lab examples, the following categories should prove useful in identifying food deserts:

- Discount Store*
- Deli / Bodega*
- Supermarket
- Organic Grocery
- Grocery Store
- Market
- Farmers Market
- Health Food Store
- Cheese Shop
- Fish Market
- Convenience Store*

The categories footnoted with '*' above will require a bit more investigation. In small neighborhoods these are often a source of food, but not necessarily the healthy kind, so I may in the end come up with some kind of formula or ratio that would actually penalize for their presence.

To identify healthcare deserts, again based on initial research, the presence or absence of the following categories could be used:

- Pharmacy
- Optical Shop
- Doctor's Office
- Drugstore
- Medical Center

And to link to a second data source to evaluate the potential correlation between food and healthcare deserts to school graduation rates, I will be using one or both of the following categories:

- High School
- School

In the assessment of graduation rates and their correlation to food and healthcare deserts, I will be concentrating on Indiana High Schools located primarily in the Indianapolis metropolitan area. Graduation rate data for 2019 is currently available in Excel spreadsheet format on the Indiana Department of Education's website.[3]

2.2 Data cleaning

While graduation data and statistics for Indiana and the Indianapolis area High Schools is available online, initial inspection indicates that, as usual, it will take a bit of data wrangling to get it into a usable format. Use of the pandas library and methods should help, but some specific items to attend to include:

- The Excel spreadsheet contains multiple tabs, two of which relate specifically to graduation rates at the individual school-level.
- Of the two sheets related to school-level graduation rates, one waives the need for students to pass exit exams and potentially hides some of the challenges of coming from a disadvantaged area, so I'll need to decide whether to use one or both of the graduation rates.
- The sheet that distinguishes between NonWaiver students (students that passed an exit requirements exam) is not particularly well formatted for parsing, e.g. column headers appear at different levels and not in the first row, so I'll need to leverage pandas read options or other methods.
- Some schools are of such a small size that their calculations are not statistically significant, so those rows will need to be filtered out.
- I'll be focusing on Indianapolis area schools for the purpose of mapping and clustering, so will need to filter on a list of desired school corporations or within a specific distance from downtown Indianapolis.
- When measuring Venues within a certain radius of a central neighborhood point, which in this case will be based on school addresses, it may be necessary to factor in urban vs rural locations. In the city, at least in Indianapolis where public transportation exists but is less than ideal, reasonable travel may be by foot or bus and somewhat more restricted in distance than in the outer regions of the city where car transportation is the norm. I will likely need to infer community type and add an indicator column for this as I pre-process the data.

3. References

1. https://foodispower.org/access-health/food-deserts/
2. https://en.wikipedia.org/wiki/Medical_deserts_in_the_United_States
3. https://www.doe.in.gov/sites/default/files/accountability/2019-state-grad-rate-data-20191231.xlsx