Applied Data Science Capstone

A Survey of Hospital Density in Wisconsin

Metz, Chelsie
11-4-2019

1. Introduction

1.1 Background

When doing initial testing, healthcare manufacturers often choose to do test runs of their equipment at hospitals which are close by. This allows the manufacturer to quickly address any issues that may arise.

However, it is equally important to obtain feedback from users in the field at this stage. To accomplish this, equipment is tested at several local hospitals at a time to maximize feedback.

1.2 Problem

Location data can be used to determine which hospitals that a Wisconsin-based manufacturer should initially test at. This report will explain how to determine where to test at by weighing both the density of hospitals in the region and their distance from the manufacturer.

1.3 Interest

Medical manufacturers would be interested in this analysis so that they can determine the best local hospitals to test their equipment at.

However, the process described in this report could be used by any manufacturer to determine proper testing locations. For example, a lathe manufacturer could use this process to instead search for local machine shops to test their equipment at instead of searching for hospitals.

2. Data Acquisition

2.1 Data Sourcing

An arbitrary address in Milwaukee, Wisconsin was chosen to represent the medical manufacturer. Using the geopy library, the coordinates of this location were obtained for use with the Foursquare API. Using Foursquare, a search query was run to find the fifty closest hospitals to the manufacturer's coordinates.

2.2 Data Cleaning

The data returned was loaded into a DataFrame for increased readability. However, the size of the data returned was larger than what was necessary for the later analysis, so all columns were dropped from the DataFrame except for each hospital's longitude and latitude information.

3. Methodology

3.1 Heat Mapping

A map centered around the manufacturer was created using the Folium library. All hospitals found in the search were added to the map as blue markers using their coordinate information. A single red marker was added to represent the manufacturers location.

This allows the distribution of hospitals within the area to be easily seen, but does not yet provide enough information to determine where the majority of hospitals are located.

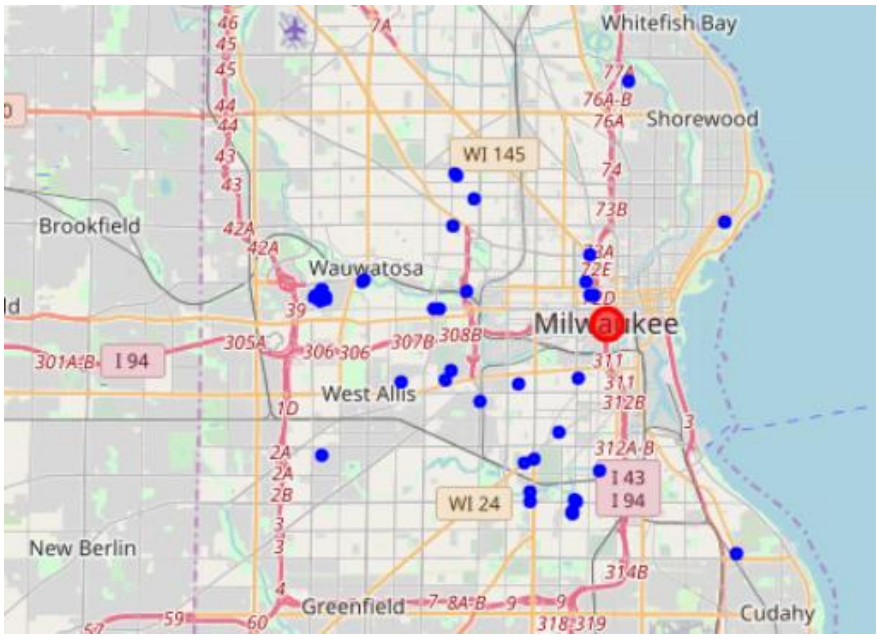

Figure 1: A heat map of hospitals in Milwaukee, WI in relation to the manufacturer location. The red circle represents the manufacturer's location. Blue circles represent hospital locations.

3.2 K-Means Clustering

To determine where the largest density of hospitals was located, K-means clustering was used. A group size of five was selected based upon the total of fifty hospitals being clustered. This split the hospitals into five separate groups based upon where they were located in proximity to each other. Each group was then assigned a color to make differentiating and plotting the groups easier.

The centroids of each group were also obtained from the clusters. These centroids represent the center location of the group which all other locations are clustered around. This can be used to determine the average distance from the manufacturer to the center of the cluster.

4. Results

4.1 Cluster Heat Map

Another heat map was then created, this time showing the area hospitals by colored cluster. Additionally, the centroids of each cluster were also added to the map, depicted by a dot with a black outline.

This heat map makes determining where hospital clusters are located in relation to the manufacturer much easier. By looking at this map, the selection of hospitals to visit can already be narrowed down as it is now readily apparent that two of the clusters contain many more hospitals than the others.

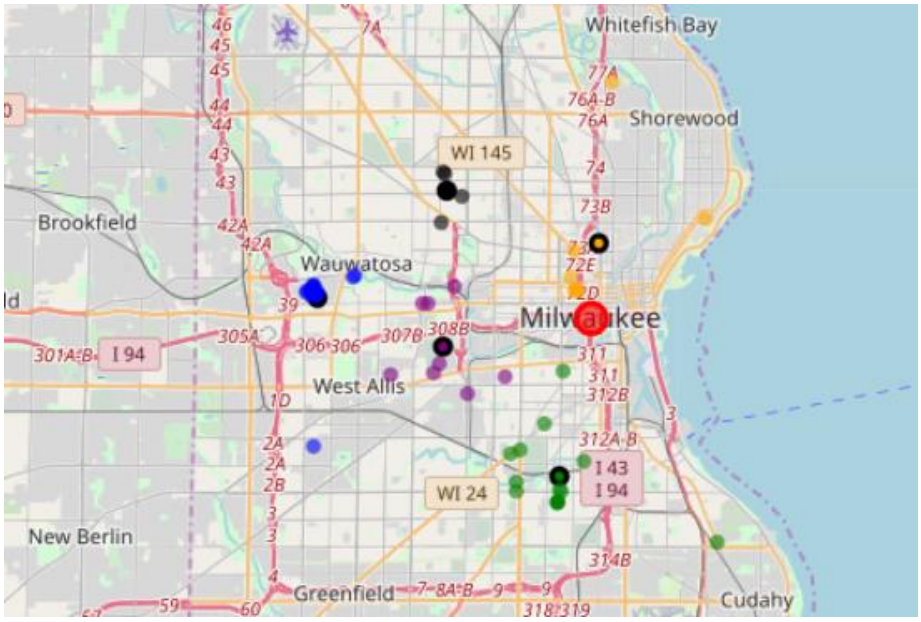

Figure 2: A heat map of Milwaukee hospitals when divided into five clusters. The red circle represents the manufacturer. Circles of other colors represent hospitals of a certain cluster. Circles with black outlines represent the centroid of their color's cluster.

4.2 Distance from the Manufacturer

While it is now more apparent where the majority of the hospitals are located, it is important to also look at their location relative to the manufacturer. To accomplish this, the geopy library was once again used to calculate the distance between the location of each cluster's centroid and the manufacturer.

Group	Number of Hospitals	Distance from Manufacturer
Black	4	5.69 km
Blue	17	8.16 km
Green	14	4.80 km
Orange	7	2.23 km
Purple	8	4.50 km

Table 1: The number of hospitals in each cluster and the distance from the manufacturer to the cluster's centroid.

5. Discussion

By first looking at the number of hospitals per cluster, this decision can be narrowed down to the Blue and Green clusters which contain significantly more hospitals than the remaining clusters. Of the two, the Green cluster is much closer, with the Blue cluster being almost twice as far from the manufacturer as it. For this reason, the manufacture should choose to test at the hospitals contained in the Green cluster.

6. Conclusion

By using K-means clustering, a selection of test sites located near a manufacturer can easily be determined. This method could be used again to determine test locations at a different manufacturer or over a larger amount of hospitals.

However, there is always the possibility that some of these sites may not be suitable for testing. For example, a manufacturer which develops ophthalmoscopes, a device for testing vision, will likely not find much value in testing at a hospital specialized in spinal surgery. In the future, it would be beneficial to narrow down the selection of sites further by specific test requirements.