

IBM Data Science Capstone Project

By Jonathan de Steuben

Introduction:

In this project, we will identify areas within Washington, D.C. that would benefit from the placement of various medical facilities. The target audiences for this report will be local health officials, local elected officials, and healthcare providers.

While there are existing medical facilities in Washington, D.C., this report will aim to locate areas within the city that have a smaller concentration as well as areas that could be considered underserved. We will determine what type of facility would best fit the area and make an appropriate recommendation.

Data:

Based on the description of the problem, this report will leverage open data provided by different sources. The following factors will influence the recommendations that will be made:

- Demographic data for Washington, D.C. wards
- Number of medical facilities within each ward
- Type of medical facilities

For this project, we will examine neighborhoods at the ward level. The following sources will be used to generate our data and subsequent recommendations:

- [Urban Institute](#) - ward demographic data
- [Washington, D.C. Open Data](#) - geoJSON files for visualization
- [Foursquare API](#) - location data for medical facilities

1. Neighborhood Data:

We will use the coordinates of neighborhoods to find medical facilities within a specific range. We'll use the neighborhood coordinates to visualize some of the data. After this, we will use these coordinates to find facilities in targeted areas.

The data will work in tandem with data at the ward level. Washington, D.C. doesn't collect demographic data at the neighborhood level, so our recommendation will be contingent on information we get from each ward.

2. Demographic Data:

Now that we've loaded our neighborhood coordinate data, we'll import files containing information accessed from the Urban Institute. The data includes information on population size and income.

Using this data, we will look at wards that have a higher percentage of poor residents.

With the demographics data loaded, we can already note some major differences between the wards, including the percentage of poor persons and the average family income. We'll plot these and examine them further.

For our next step, we'll use the neighborhood data as the input for a call to Foursquare's Places API. We'll import our keys and create a function that will give us a DataFrame of results for analysis. Hidden cells, and more information about using the foursquare app and what we're doing with venue categories.

There are many categories available for us to search for using Foursquare's Places API. For the purposes of this project, we will only need to search for locations associated with a medical service or healthcare provider.

Methodology

As stated in the introduction of the project, our goal will be to examine areas in Washington, D.C. wards that may have shortages of medical care due to their location. Our goal is to determine the best possible location for the placement of various medical facilities based on their absence in these areas.

We began by mapping our neighborhoods and running our DataFrame through Foursquare's Places API to obtain the total number of medical facilities in the District. Now that we have our results, we will use the data to determine shortage areas and make the recommendation for new or additional facilities. This is where we'll apply demographic data to search for areas with the highest need.

Analysis

We'll begin by visualizing the demographic data by ward to guide our decision on additional medical facilities. After we examine these areas, we'll hone in on areas that may be considered underserved. We'll use choropleth maps with Folium to visualize the wards. From there, we'll determine a suitable location for increased medical services given what presently exists.

We can see from the resulting heatmap that there are a few neighborhoods where medical facilities are out of proximity:

- **Benning Ridge**
- **Marshall Heights**
- **Deanwood**
- **Kenilworth**
- **Lincoln Heights**
- **Greenway**
- **Capitol View**

This concludes our analysis section.

Results and Discussion

Based on our analysis, we've determined a group of neighborhoods that would be the best choices to add increased medical services and facilities. Given the demographic information that we used in the analysis stage, we determined that these neighborhoods are lower income and have higher percentages of unemployed persons, which we believe should be given higher priority based on their need.

By the end of the project, we were able to pinpoint these neighborhoods. We can present this new information to different stakeholders as a public interest project that can benefit underserved areas in the city that don't have any medical facilities to serve their populations.

Conclusion

In this project, our goal was to use the data science skills learned in Coursera's IBM Data Science course to find areas in Washington D.C. to find areas that would benefit from additional medical facilities. Using open data, previous course material, data visualization packages, and other resources, we were able to apply a solution to our problem using data.

