

Analyzing Access to Health Care Services

For: Applied Data Science Capstone

Table of Contents

Introduction	3
Data	3
US Census Data	3
Foursquare API Data	5
Methodology.....	6
Analysis.....	7
Correlation Analysis	7
Correlation Analysis Results	7
Multiple Linear Regression	8
Multiple Linear Regression Results	8
Decision Tree Classification	9
Decision Tree Classification Results	9
Discussion	10
Conclusion.....	10
References.....	11

Introduction

Access to health care services is defined as "the timely use of personal health services to achieve the best health outcomes." (IOM, 1993). Having or not having access to health care services can have a great impact on a person's physical and mental health, and consequently affect a person's quality of life. More broadly it has an effect on the public health of a given society.

Access to health care services requires being able to find health care services within a specific geographic area, having insurance coverage, finding a good health care provider and other factors. This project will focus on studying select disparate geographic regions in the United States to discover similarities and differences as it relates to being able to access health care services. Specifically, the project will focus on answering the questions: Are there certain regions that have better access to health care services than others? What are the shared characteristics of these regions that make them have better or worse access to health care services? For example, anecdotal evidence appears to indicate low income neighborhoods often have fewer health care providers, hospitals and other health care resources than areas with higher income. This project will use available data to attempt to answer these questions more definitively.

The answer to these questions could provide valuable insights to public health officials. The hope is this information would be useful in guiding policy related to expanding health care services to areas that really need it.

Data

This project will use the following data to attempt to answer the questions defined in the introductory section:

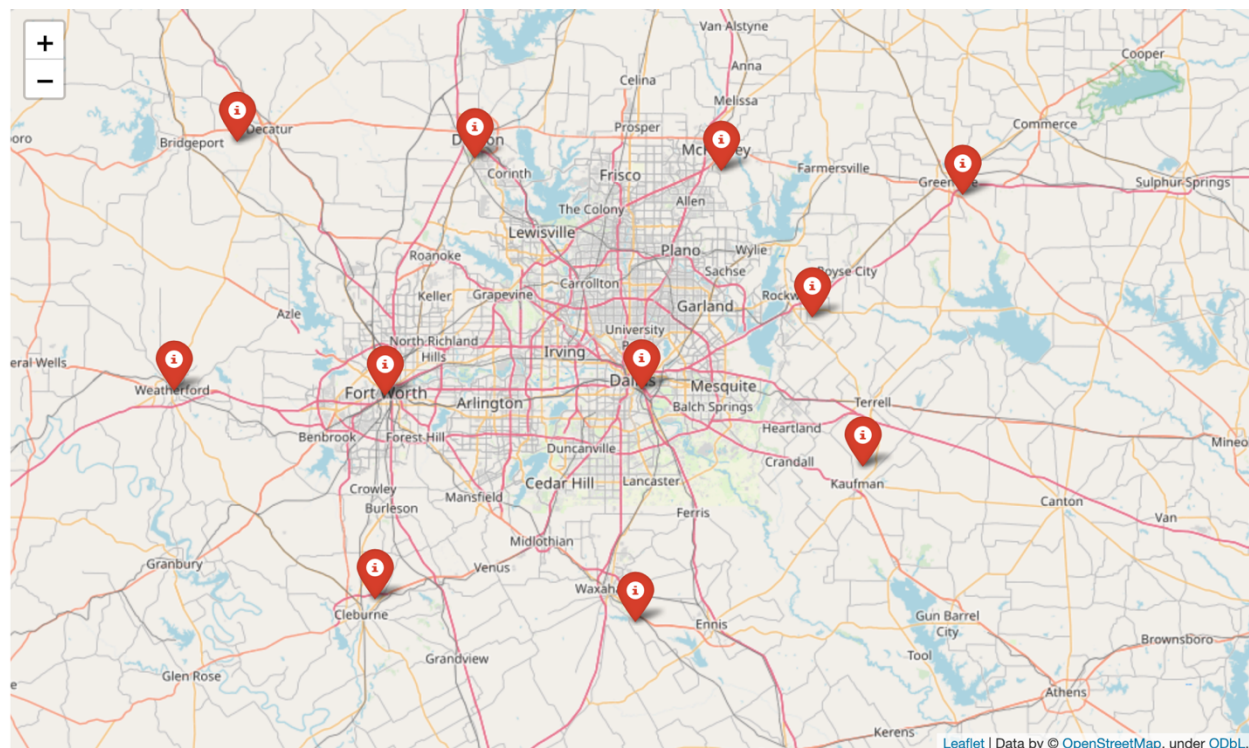
- Foursquare API used to obtain health care service provider information for a given geographic location. Specifically, the API will be queried for locations that match specific categories including: Hospitals, Doctor's Offices, Pharmacies, Dentist offices, Urgent care clinics and other medical facilities.
- US Census Bureau population demographics and economic data for a given location. This data will be used to provide additional information on the geographic regions being studied. Data to be used includes household income, population, race, ethnicity, age and other population demographics.

US Census Data

For the purposes of this project, counties in the Dallas–Fort Worth–Arlington, TX Metropolitan Statistical Area (DFW) will be studied. This area includes 11 counties and in 2018 had a population of 7,539,711 according to estimates by the U.S. Census Bureau. The 11 Counties include: Collin County, Dallas County, Denton County, Ellis County, Hunt County, Kaufman

County, Rockwall County, Johnson County, Parker County, Tarrant County and Wise County. See map below.

Dallas-Fort Worth-Arlington Counties



The following data was obtained from data.gov and is a result of the American Community Survey (ACS) an ongoing survey by the U.S. Census Bureau.

- 2018 Population demographics: This dataset includes the total population, sex, age, race and ethnicity data for the eleven counties.
- 2018 Household income: This dataset includes the estimated median household income in 2018 for the eleven counties
- 2018 Economic characteristics: This dataset represents the economic characteristics of the counties in the study and includes: unemployment rate
- This data was obtained from data.gov and is a result of the American Community Survey (ACS) an ongoing survey by the U.S. Census Bureau

U.S. Census Bureau Counties Data

	Collin	Dallas	Johnson	Rockwall	Tarrant	Ellis	Denton	Kaufman	Hunt	Parker	Wise
Total Population	1,005,146	2,637,772	171,361	100,657	2,084,931	179,436	859,064	128,622	96,493	138,371	68,305
Male	49.2%	49.4%	49.9%	48.6%	48.9%	49.3%	49.2%	48.6%	49.6%	49.6%	50.0%
Female	50.8%	50.6%	50.1%	51.4%	51.1%	50.7%	50.8%	51.4%	50.4%	50.4%	50.0%
Under 18	25.9%	26.1%	25.9%	27.0%	26.3%	26.6%	24.7%	27.8%	23.9%	24.7%	24.7%
18 and Over	74.1%	73.9%	74.1%	73.0%	73.7%	73.4%	75.3%	72.2%	76.1%	75.3%	75.3%
62 and Over	14.0%	13.6%	16.9%	15.4%	14.3%	15.5%	12.9%	15.9%	20.2%	19.3%	19.5%
White	67.8%	60.6%	90.1%	84.4%	65.4%	81.4%	73.6%	81.7%	80.8%	93.6%	91.1%
Black	10.1%	22.8%	3.7%	6.6%	16.6%	10.9%	10.2%	12.3%	8.7%	1.6%	2.5%
American & Alaskan Native	0.5%	0.5%	0.3%	0.5%	0.8%	1.3%	0.5%	0.3%	1.7%	1.2%	0.5%
Asian	15.9%	6.4%	1.2%	2.1%	5.6%	0.7%	9.3%	1.4%	1.6%	0.8%	0.0%
Hawaiian	0.0%	0.0%	0.0%	0.0%	0.2%	0.0%	0.1%	0.2%	0.0%	0.2%	0.1%
Other Race	2.6%	7.4%	2.5%	3.1%	8.0%	1.9%	2.5%	1.2%	6.5%	0.4%	4.0%
Hispanic All Races	15.4%	40.5%	22.0%	18.0%	29.2%	26.6%	19.5%	22.4%	17.0%	12.8%	19.7%
Median Household Income	\$94,192	\$56,854	\$62,066	\$94,767	\$64,874	\$72,213	\$83,376	\$65,390	\$51,928	\$74,625	\$61,709
Unemployment Rate	3.6%	4.3%	3.6%	5.3%	4.7%	3.3%	4.0%	6.9%	5.0%	4.8%	3.5%
No Health Insurance	11.3%	21.7%	16.5%	12.5%	17.2%	15.9%	11.1%	15.3%	18.5%	16.7%	17.0%
Families Below Poverty	4.5%	10.8%	8.4%	5.6%	9.3%	6.4%	4.3%	9.8%	8.0%	4.5%	5.8%

Foursquare API Data

Foursquare is a mobile application which provides a search service to find places of interest locally. Foursquare also makes the database of these places available through it's API.

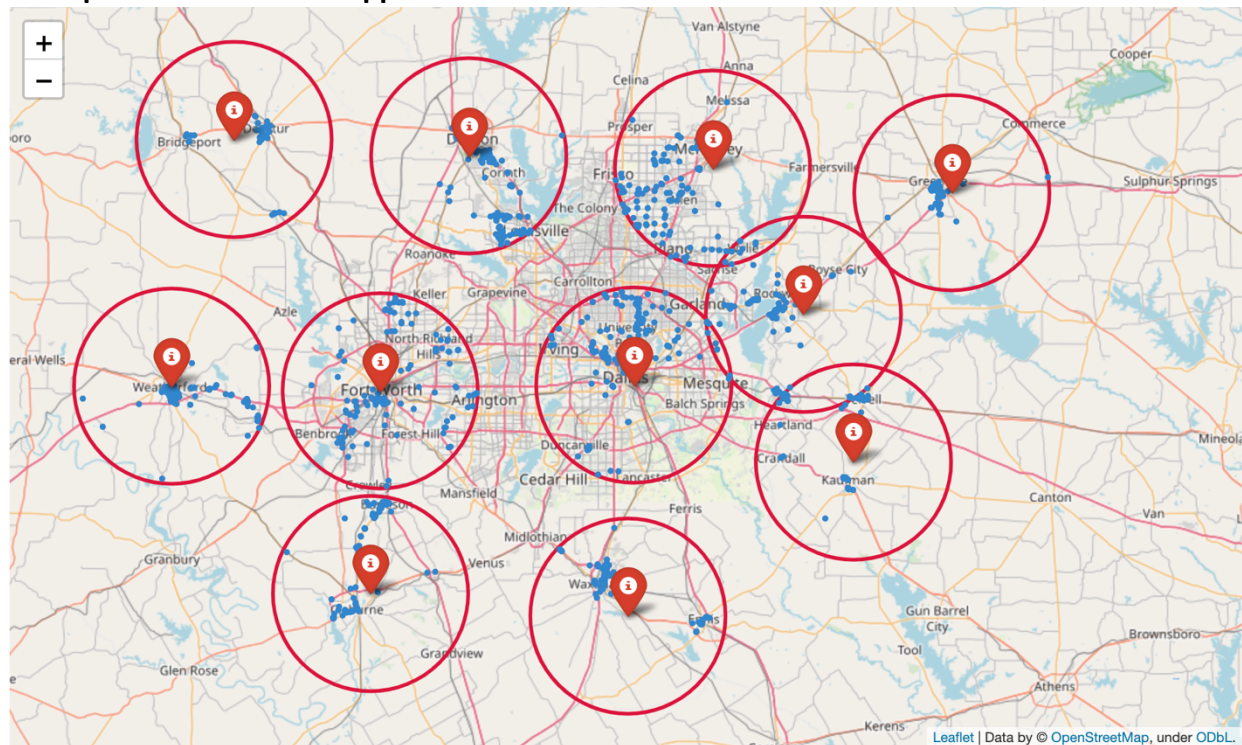
For the purposes of this study, the Foursquare places API was used to retrieve data regarding different categories of health care service providers within a defined geographic area. To do this:

1. The geographic coordinates of the counties were retrieved using a geocoding library.
2. The Foursquare places API requires a radius to be defined when retrieving places of interest. An area with a radius of 25,000 meters was defined around the center of the counties. This had the least amount of overlap between the counties while covering most of the area of study.
3. For each county, the Foursquare places API was used to retrieve places tagged with a category of medical centers which included hospitals, doctor's offices, pharmacies, dentist offices, urgent care clinics and other medical facilities. The dataset was cleaned to remove non health care related categories.

Foursquare Places API Data

County	Dentist's Office	Doctor's Office	Emergency Room	Eye Doctor	Hospital	Maternity Clinic	Medical Center	Pharmacy	Physical Therapist	Rehab Center	Spa	Urgent Care Center	Total Medical Centers
Collin	12	45	3	1	6	0	22	7	0	1	0	4	101
Dallas	17	63	9	5	37	0	52	2	0	0	2	3	190
Denton	17	70	3	3	14	0	28	3	6	1	0	2	147
Ellis	23	62	1	2	4	1	23	0	3	1	0	1	121
Hunt	16	29	0	0	4	0	24	0	1	1	0	0	75
Johnson	21	32	2	2	5	1	14	1	2	0	0	1	81
Kaufman	24	34	2	1	3	1	11	0	1	4	0	0	81
Parker	21	61	0	3	6	1	25	0	6	2	0	1	126
Rockwall	23	19	3	2	3	0	26	0	2	3	0	3	84
Tarrant	9	36	4	4	18	0	38	0	3	0	0	2	114
Wise	5	21	1	0	4	1	15	1	3	0	0	1	52

Foursquare Places Data Mapped



Methodology

The objective of this study was to answer the questions: Are there certain regions that have better access to health care services than others? What are the shared characteristics of these regions that make them have better or worse access to health care services?

For the purposes of this study, the measure of access to health care services was defined as the ratio of population in a defined region to total health care services available in the area. The total health care services in the region is the sum of all different categories of medical centers gathered through the Foursquare API. A low ratio of people to medical centers indicates higher access to care, meaning the population has more options available for care. This ratio served as the target variable for the study. For further clarity the formula is shown below.

The ratio of people to medical centers = Total Population in County / Sum of Medical Centers in county

Data from the US Census Bureau provided population economic and demographic characteristics by county. The Foursquare API was then used to obtain data on different categories of health care services within a defined radius of 25,000 meters. These two datasets were then combined to provide independent variable to be analyzed against the target variable defined above.

The tools selected to understand the relationship between the ratio of people to medical centers and the economic and demographic characteristics of a county were correlation analysis, multiple linear regression and decision trees.

Analysis

Correlation Analysis

The Pearson correlation coefficient, a measure of the linear correlation between two variables, was calculated for the ratio of people to medical centers and the selected economic and demographic characteristics of the counties. The results are shown below:

Correlation to PeopleToMedicalCenters	
PeopleToMedicalCenters	1.000000
PercBlack	0.758471
PercOther	0.682679
PercHispanicAll	0.608594
PercAsian	0.597868
PercFamiliesBelowPoverty	0.381150
PercFemale	0.229905
PercNoHealthInsurance	0.201154
PercUnder18	0.138546
PercHawaiian	0.125324
2018MedianHouseholdIncome	-0.046710
UnemploymentRate	-0.118880
Perc18andOver	-0.138546
PercAmericanNative	-0.195468
PercMale	-0.229905
Perc62andOver	-0.626003
PercWhite	-0.873660

Correlation Analysis Results

Race appears to have a strong correlation with the target variable, the ratio of people to medical centers. As the percent of Black, Hispanic, Asian and people of other races increases so does the ratio of people to medical centers. Conversely as the percent of white people increases the ratio of people to medical centers decreases.

Other less prominent but important variables were:

Percent of people 62 years and older. As the percent of people 62 years and older increases the ratio of people to medical centers decreases.

Percent of families below the poverty level. As the percent of families below poverty increases so does the ratio of people to medical centers.

Percent of people with no health insurance. As the percent of people with no health insurance increases so does the ratio of people to medical centers.

Percent of males and females. As the percent of males increases the ratio of people to medical centers decreases. Conversely as the percent of females increases so does the ratio of people to medical centers.

Multiple Linear Regression

Multiple linear regression was performed to attempt to understand the relationship between the target variable, The ratio of people to medical centers and the selected economic and demographic characteristics of the counties. The resulting coefficients are shown below.

	Coefficient
PercNoHealthInsurance	2820.645487
PercHawaiian	2321.243105
UnemploymentRate	2197.381495
PercUnder18	1885.544236
PercFemale	1195.724673
2018MedianHouseholdIncome	-0.449435
PercHispanicAll	-155.923479
PercOther	-693.850425
PercAmericanNative	-923.354811
PercMale	-1195.724673
PercAsian	-1404.617663
Perc18andOver	-1885.544236
PercFamiliesBelowPoverty	-2561.726281
Perc62andOver	-2878.634517
PercWhite	-2918.651966
PercBlack	-4419.315507

Multiple Linear Regression Results

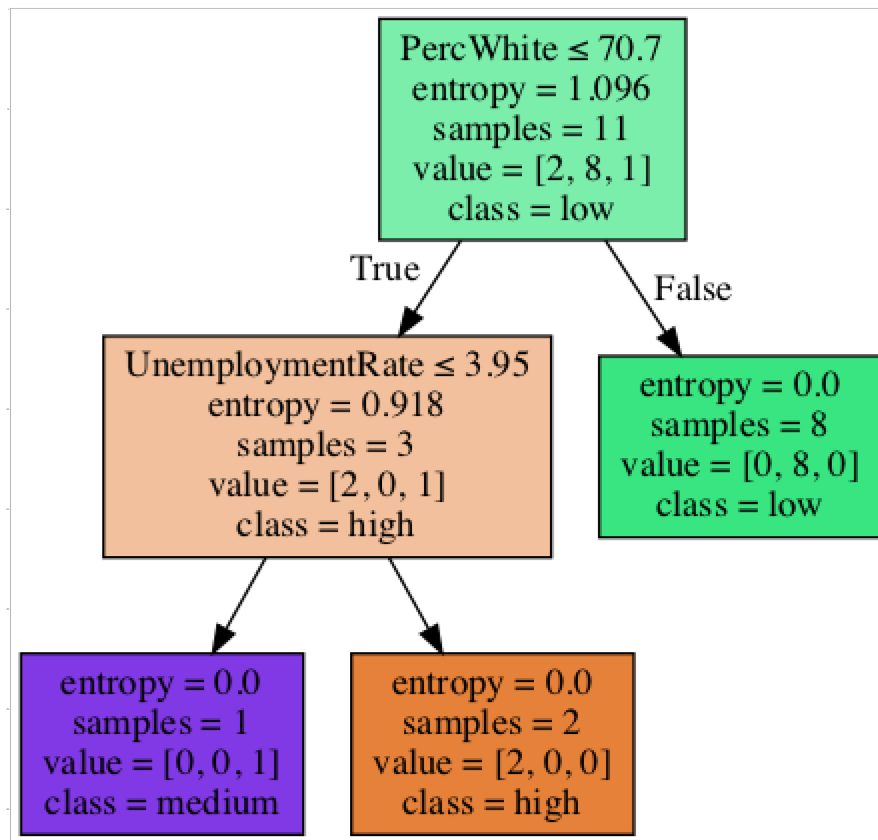
The resulting coefficient values should be interpreted to mean for every unit increase in an independent variable, there's an increase or decrease in the value of the independent variable.

The variables with the most significant positive impact on the value of the target variable, the ratio of people to medical centers, are percent of people with no health insurance, percent of people of the Hawaiian race, percent unemployed. For example, a 1% increase in percent of people with no health insurance results in a 2820.645487 increase in the ratio of people to medical centers.

The variables with the most significant negative impact on the value of the target variable, the ratio of people to medical centers, are percent of black people, percent of white people, percent of people 62 and over and percent of families below poverty level. Meaning a unit increase results in a decrease in the ratio of people to medical centers. For example, a 1% increase in the percent of people 62 and over results in a 2878.634517 decrease in the ratio of people to medical centers.

Decision Tree Classification

A decision tree classification model was built to attempt to understand the relationship between the ratio of people to medical centers and the selected economic and demographic characteristics of the counties. The decision tree classification model requires the independent variable to be a discrete categorical variable. To meet this requirement the continuous variable the ratio of people to medical centers was transformed into three evenly spaced categories named High, Medium and Low. The resulting tree is shown in the visualization below.



Decision Tree Classification Results

The most prominent variable identified by the model was the percent of white people in the population. When a county population was made up of 70.7% or less of white people, the

population was later classified as having a high or medium ratio of people to medical centers. This was further broken down by the next most prominent variable, the unemployment rate. When a population had an unemployment rate of 3.95% or less, the population was classified as having a medium ratio of people to medical centers otherwise it was classified to have a high ratio.

Discussion

The results of correlation analysis and decision classification model appear to agree that the racial make-up of a county is strongly associated with higher access to health care services for people living in the county. Specifically, for the counties in the study, if a county is largely made up of people classified as white from the US Census Bureau data, the county has a lower ratio of people to medical centers which means higher access to health care services for its population. This information can be used by policy makers to further study the underlying causes and to design intervention efforts to provide higher access for these minority populations. Suggestions would be to look at the relationship between race, unemployment and income.

The results of correlation analysis and multiple linear regression analysis appear to agree that if a county has a larger population of people 62 and older, the county has a lower ratio of people to medical centers which means higher access to health care services for its population. This could be explained by the fact that this population is largely made up of retirees who have Medicare healthcare coverage and are able to afford health care services. This information could be used by policy makers to increase access to insurance coverage that seems to be indirectly associated with higher access to care.

The results of multiple linear regression analysis and decision classification model appear to agree that if a county has a higher unemployment rate, the county has a higher ratio of people to medical centers which means lower access to health care services for its population. It appears to support anecdotal evidence that people that are unemployed and consequently have lower income have lower access to care. Knowing this, policy makers can look for ways to assist people who are unemployed to be able to pay for health care services, which could as a result boost access to health care for a county's population.

Conclusion

In conclusion this limited study provided useful insights that could be used to guide policy changes to ensure everyone has improved access to health care services. The study revealed race, age and unemployment are the primary factors that affect a population's access to care. In answering the questions posed in the introductory section; Higher access to health care services was associated with counties with a large percent of white people, a large percent of people 62 years and older and lower unemployment rate.

Future studies could improve on this work by expanding the area of study to include a larger number of counties across the United States. The study could also be improved by using other sources to complement the Foursquare API places which in some cases appeared to be incomplete. Other research could also explore the underlying reasons for why these factors have an impact on access to health care services.

References

- Institute of Medicine, Committee on Monitoring Access to Personal Health Care Services. Access to health care in America. Washington, DC: National Academy Press; 1993. <https://www.ncbi.nlm.nih.gov/books/NBK235882/>