

Project Title

Construction of Hospitals Dependence on State County Demographic Data

Introduction Section

This is a description of the problem and a discussion of the background. Here I will discuss the business problem and who would be interested in this project.

A group of potential investors are looking into the possibility of building new, for-profit hospitals in various counties in the Commonwealth of Pennsylvania.

These investors want to know, based on specific state demographic data, how they can arrive at an accurate decision before providing the funding for this type of venture.

The investors want to base their decision on county population, median household income and the number of already-existing hospitals in the same county.

The investors want to make sure that the county that they plan to build and open the new hospital has an adequate population as well as a population that has enough potential income to support the revenue streams of the hospital. The investors also want to avoid building new facilities where there might be too much competition which could eventually drive down revenue streams and profitability and ruin any chance of a return on their investment. Thus the need for this project.

Just a note: The type of Hospitals that the Investors are thinking of building are 'Urgent Care Centers'.

Data Section

This is a description of the data and how it will be used to solve the problem. This is where I describe the data that will be used to solve the problem and the source of the data.

The data used to help in the final investor decision-making process will come from two sources:

- a. The World Wide Web which has web sites that contain Pennsylvania state county demographic table data such as county population, median household income and latitude and longitude coordinates of each county.
- b. The Foursquare location data that will provide the number of hospitals in each of the Pennsylvania state counties. The county latitude and longitude coordinates from the web-based table data will be used as some of the parameters sent to the Foursquare API.

These two data sources will help the investors to decide if it will be worth the financial investment for them to proceed with the construction and opening of new hospitals in a particular county.

Median Household Income must be at a certain high level as well as a certain county population high level. In addition, the number of already-existing hospitals must not be *above* a certain level also.

Methodology Section

The initial exploratory data analysis that was done was to peruse the sources of the data that will be needed to produce the required outcome. Source data from an HTML table and CSV file from a web source was used. The data was combined together along with latitude and longitude coordinates. The Foursquare API was then used. I needed to get the Median Household Income, Population and the number of already-existing Urgent Care Centers in order to do this initial analysis. The Investors' criteria was to only consider building new for-profit Urgent Care Centers if ALL of the following was true about a particular County in Pennsylvania:

County Number of Existing Urgent Care Centers:	Must be Ten (10) or less
County Total Population:	Must be at least 30,000 people or more
County Median Household Income:	Must be at least \$40,000 or more

Inferential statistical testing was done and observed as follows (this was 'before' the final data frame was filtered per investor criteria):

The Lowest Urgent Care Center Count for a County was: 2
The Highest Urgent Care Center Count for a County was: 50
The Average Urgent Care Center Count was: 27
The Standard Deviation of Urgent Care Center Count was: 18

The Lowest Median Household Income for a County was: \$9,686
The Highest Median Household Income for a County was: \$84,741
The Average Median Household Income was: \$46,664
The Standard Deviation Median Household Income was: 9,686

The Lowest Population for a County was: 5,085
The Highest Population for a County was: 1,526,006
The Average Population for a County was: 189,588
The Standard Deviation Population for a County was: 268,339

Results Section

Applying the Investors' rigid criteria (as mentioned in the Methodology Section above) to the final data frame resulted in the following outcome:

County	Hospital Count	Medianhouseholdincome	Population
Elk	2	43,745	31,946
Lycoming	3	42,689	116,111
McKean	4	40,097	43,450
Tioga	4	40,338	41,981
Centre	5	47,016	153,990
Clarion	7	40,028	39,988
Blair	9	42,363	127,089
Huntingdon	10	41,700	45,913

As you can see, we have just a total of eight (8) Pennsylvania Counties that will undoubtedly meet the stringent criteria of the Investors who wished this research be done as accurately as possible. Hospital counts are all 10 or less, Income is at least \$40,000 and Population is at least 30,000. This definitely meets the investors' criteria requirements.

Discussion Section

Here are the results again. This section is where I discuss any observations I noted and any recommendations I can make based on the results.

County	Hospital Count	Medianhouseholdincome	Population
Elk	2	43,745	31,946
Lycoming	3	42,689	116,111
McKean	4	40,097	43,450
Tioga	4	40,338	41,981
Centre	5	47,016	153,990
Clarion	7	40,028	39,988
Blair	9	42,363	127,089
Huntingdon	10	41,700	45,913

Observations:

- From a separate web site, I found that the eight counties listed above are in predominately sparsely-populated areas of the state that are not near any highly densely-populated metropolitan areas of the state or near large cities.
- The already-existing Hospital Counts, for the most part, are in the single digits.
- The Median Household Incomes are in the \$40,000 range of values.
- The Populations whose values are in the 5-digit range account for about half the counties listed.
- The Populations whose values are in the 6-digit range account for about the other half of the counties listed.

It is my presumption that the Investors did well with their criteria requirements since:

- The existing hospital counts are so low that there does not appear to be a gross amount of competition.
- The Incomes appear to be sufficient to support a person's ability to pay for their urgent care medical expenses.
- The county populations appear to be enough to sustain a sufficient flow of visitors into the urgent care centers.

My recommendations:

- The investors could loosen their criteria to allow for a slightly lower median household income since health care is a service that every person needs at some point in their lives and with the absence of a large pool of competitors, they could build more than just one new urgent care center in any of these eight counties.
- The investors could apply this same research to other states in the country or even expand their horizons by considering building new for-profit urgent care centers in other countries.

Conclusion Section

In conclusion, in this research that I conducted, I acquired Pennsylvania demographic data to aid in the decision-making of a group of investors who wished to pool their funding with the ultimate goal of building new for-profit hospitals (urgent care centers) in various Pennsylvania counties.

The analysis involved strict criteria that was applied to county population, median household income and the existence of already-existing competitors.

By combining the data sources together and using Foursquare location data API and applying the investor criteria, I was able to create a list of potential counties that might be suitable for the goals of the investors.