IMPACT OF PRISONS ON THEIR SURROUNDING VENUES

by

Mg. Facundo N.Paris

April, 2020

IBM Data Science Course, Coursera

Table of Contents

List of Figures

Chapter 1

Introduction

A lot of factors come into play in determining the success of a venue. To mention but a few, technology, capacity and layout are important, but the region in which your venue is situated is by far the biggest deciding factor for how many consumers will arrive at your doorstep. In an article written by the real estate adviser Steve McLinden, he stated that "tangible influences such as heavy traffic, eyesore buildings, odors, railroads, foreclosures, prisons and psychiatric hospital institutions can and will adversely affect demand and value"[1].

In this work we will focus on prisons and how they affect their surroundings. Prison can affect the emotions of a neighborhood. In an article written by McGee, he describes the three possible reactions to being near prisons: "fear of harm from the inmates, economic anxiety, and civic pride (McGee 1981:110)." Another writer finds that these and similar objections to a prison in town derive from resident's fears of community change and loss of preferred lifestyles (Carlson 1988). Conversely, local supporters of prison siting tend to emphasize the jobs and economic benefits an institution would bring, while discounting the likelihood of any of the negatives identified above (Pagel 1988).

Until recently, most of these debates have been carried out with little substantive information to support or refute either view. The need for some basis on which to project

socioeconomic effects has been further reinforced by the presence in many states of siting guidelines requiring community support. The final and perhaps most important factor in stimulating research on prison effects has been the growth in new prison construction [2]. The United States prison population rate of roughly 700 inmates per 100,000 people is the second-highest of 222 countries tracked by the Institute for Criminal Policy Research. Nearly one out of every 100 American adults is behind bars. Private correctional facilities were a $4.8 billion industry in 2014, with profits of $629 million, according to market research firm IBISWorld [3]

The study is centered in the state of Alabama. For years, Alabama has practiced unnecessary and excessively long mandatory sentencing laws, resulting in a rising prison population. It operates the nation's most crowded prison system. In 2015 it housed more than 24,000 inmates in a system designed for 13,318. In 2019 the U.S. Department of Justice found conditions in Alabama prisons to be unsafe and unconstitutional, as result of a long civil rights investigation prompted by numerous deaths from violence in Alabama lockups [4].

1.1 Business problem

The main goal of this article is to describe one of the socioeconomic impacts of prisons: the number of venues in their surroundings. The focus will be to find and describe the influence radius of the prisons in the state of Alabama. Finally, we want to cluster the prisons by the similarities on their influences. This report is directed towards governments who want to better understand the behaviour of society, and also towards helping investors make more informed decisions on whether to start a venue or not near a facility. This is the capstone project in the IBM Data Science Course , hosted by Coursera.

Chapter 2

Data

The target state is Alabama but this type of analysis can be done in any state. For this project we need two sets of data. The first consists on a table with the names, population, and GPS locations of the prisons in Alabama. This information can be found in "https://www.prisonersofthecensus.org". This web page has done an incredible job recompiling data of this nature. They state that "The way the Census Bureau counts people in prison creates significant problems for democracy and for our nation's future. It leads to a dramatic distortion of representation at local and state levels, and creates an inaccurate picture of community populations for research and planning purposes", so they came up with their own data set.

The second data set will be the GPS location of the venues near each facility. For this we use Foursquare API.

First, we define a radius of influence of two kilometers. Then, for each prison we calculate the distance from the prison to the venues inside the influence radius. The distance is a continuous variable, so for the sake of simplicity we will categorize the distances in ranges. For example, a prison could have 10 venues in the first 500 meters, 35 venues from 500 to 1000 meters, 23 from 1000 to 1500 and 12 from 1500 to 2000 meters. Finally, we will use machine learning techniques to cluster similar prisons.

Chapter 3

Methodology

All the steps to process and explore the data, and the results will be on this link, so I suggest the interested reader open it: LIIIIIIIIIIIIIIIIINKKKKK

In the web page "https://www.prisonersofthecensus.org/data/kml.html" you can download an ".kml" file for each state in the USA. The beautifulsoup library enables us to handle this type of extension, so we can apply the knowledge of the previous labs. As an example, figure 3.1 shows the beginning of the file, where we can see the information for one facility, the Autauga Metro Jail. The $\langle Placemark \rangle$ divides the information of each facility. Inside two $\langle Placemark \rangle$ we can find the name, description, Correctional Population, and the GPS coordinates of each facility.

```xml
<?xml version="1.0" encoding="UTF-8"?>
        <kml xmlns="http://www.opengis.net/kml/2.2">
        <Document>
<Placemark>
        <name>Autauga Metro Jail</name>
        <description><div>Facility type: Local</div><div>Geoid: <a href="http://
www.prisonersofthecensus.org/data/2010blocks/010010202002008/">010010202002008</
a></div><div>County: Autauga County</div><div>2010 Correctional Population: 181</
div></description>
        <styleUrl>http://www.prisonersofthecensus.org/locator2010/map_styles.kml?
v=10#localPoint</styleUrl>
        <Point>
        <coordinates>-86.473973,32.465298</coordinates>
        </Point>
</Placemark>
```

Figure 3.1: Information inside '.kml' file

In the section "Prisons location" of the link, we show the code used to convert this

information to a pandas data frame. In Figure 3.2 we show the first five elements of the dataframe.

	Name	Population	Longitud	Latitud
0	Autauga Metro Jail	181	-86.473973	32.465298
1	Baldwin County Jail	486	-87.774451	30.885110
2	Daphne City Jail	38	-87.892566	30.595724
3	Loxley Community Work Center	505	-87.778375	30.608661
4	Fairhope City Jail	17	-87.902480	30.524892

Figure 3.2: Pandas data frame of prisons and locations

With the describe() method we see that we have 107 facilities after cleaning, with an average population of 327, minimum population of 1, and maximum of 2399. The population number is important because a facility with a small number of inmates will be small, and may be unnoticed by the people on that neighborhood, introducing a bias in the study. It is possible that big facilities are constructed farther apart from cities, but if there is any, it will not go unnoticed. In Figure 3.3 we show a box plot of the population of all the facilities.
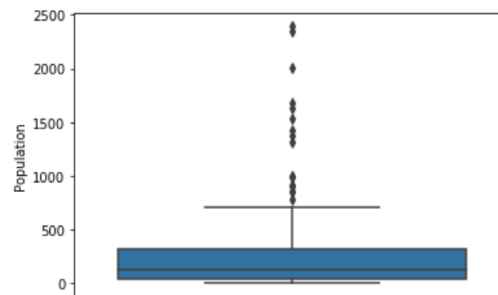

Figure 3.3: Population in the facilities

For this study we will keep only the prisons with populations larger than 600 inmates. With this restriction our dataset consists of 18 prisons. Using geopy and folium libraries , we show all the facilities in the state of Alabama where the population is bigger than 600 inmates in Figure 3.4.
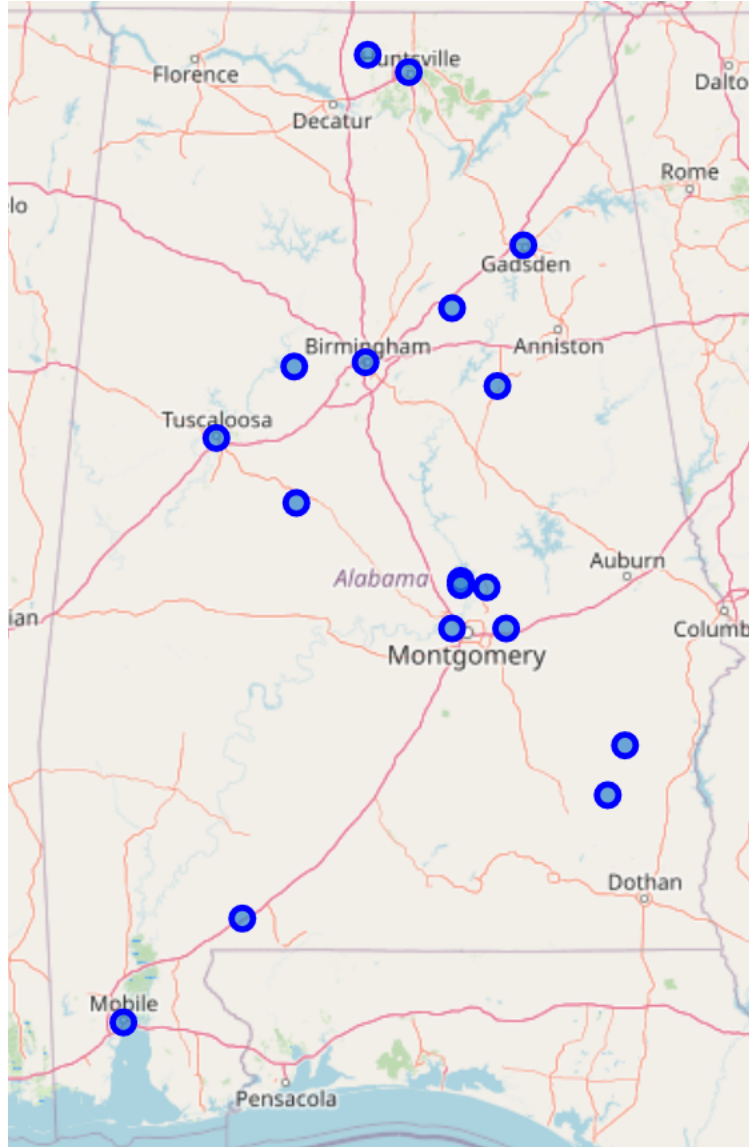
Figure 3.4: Prisons map

Now we use the Foursquare API. First, you have to create an account to do the requests. Once you have the necessary credentials, you are ready to request the information you need. In the section "Let's create a function that returns the venues in a prison's surroundings" of the link, we define a function named "getNearbyVenues" that returns the venues in the surroundings of a prison. Then, we use the geopy distance library to calculate the distance between a prison and the venues nearest to it. Then, we obtain a dataframe such as the one shown in figure 3.5. Each row in this table corresponds to a venue near the facility

indicated in the first column(Name). We establish the limit of venues at 300.

	Name	Prison Latitude	Prison Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Distance
0	Bibb Correctional Facility	32.925245	-87.164289	Piggly Wiggly	32.937650	-87.165119	Convenience Store	1377.942930
1	Bibb Correctional Facility	32.925245	-87.164289	Bjs Restaurant	32.924740	-87.181952	American Restaurant	1653.008469
2	Draper / Staton / Elmore prisons	32.589624	-86.332650	Scarlett Oak Dr	32.601624	-86.344780	Athletics & Sports	1751.429296
3	Draper / Staton / Elmore prisons	32.589624	-86.332650	The Rogers Farm	32.589363	-86.351538	Scenic Lookout	1773.509217
4	Draper / Staton / Elmore prisons	32.572940	-86.327885	The Spillway	32.568231	-86.342427	River	1461.929914

Figure 3.5: Data frame

Chapter 4

Results and discussion

In Figure 4.1 we show a map of the prisons and corresponding venues. We put the limit of venues at 300, but when we use ".groupby('Name').count()" we see that the max number of venues near a facility is less or equal to 100. Also some prisons don't have any venues at all in the 2 kilometers radius.

Let's take the Madison County Main Detention facility as an example. In figure 4.2 we look a little bit closer at this facility. As we said earlier, distance is a continuous variable, so we divide the 2 kilometer (2000 meters) radius in 10 intervals of 200 meters each. We then classify the distance between the prison and a particular venue according to which interval it is in. With all the venues near this prison, we have the following histogram, see figure 4.3.
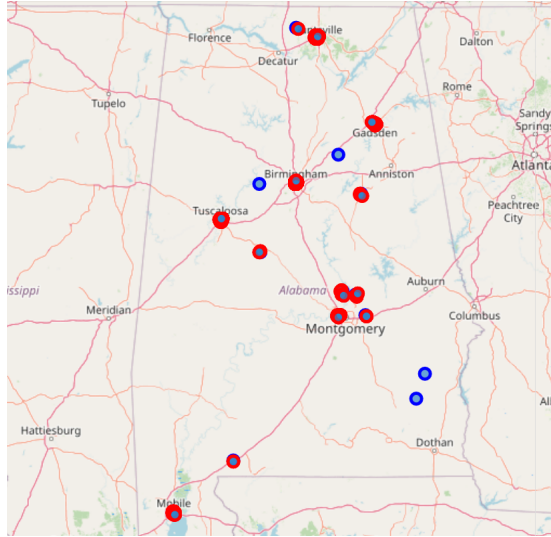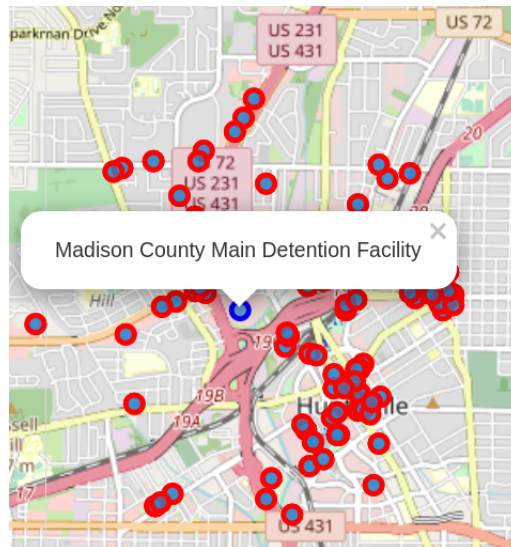
Figure 4.1: Map of prisons and venues

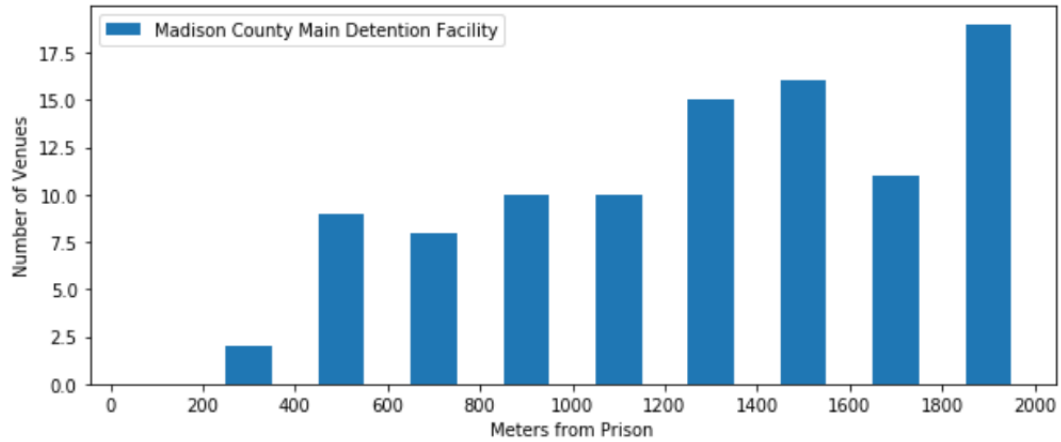Figure 4.2: Madison County Main Detention Facility

Figure 4.3: Histogram of Madison County Main Detention Facility

We see that in the first 200 meters there are no venues and, as we increase the distance from the facility, there is a growth in the number of venues. With this graph we clearly show an example of the negative impact of the prison on the number of venues in its surroundings.

If we extend this analysis to the rest of the prisons, we obtain Figure 4.4. Again, there is a clear pattern for all the prisons. In the first 200 meters there is a very small number of venues, less than 2, and the majority of the prisons present zero surrounding venues. Then the number starts to grow until we reach 1000 meters, after which the behaviour does not follow a clear pattern.
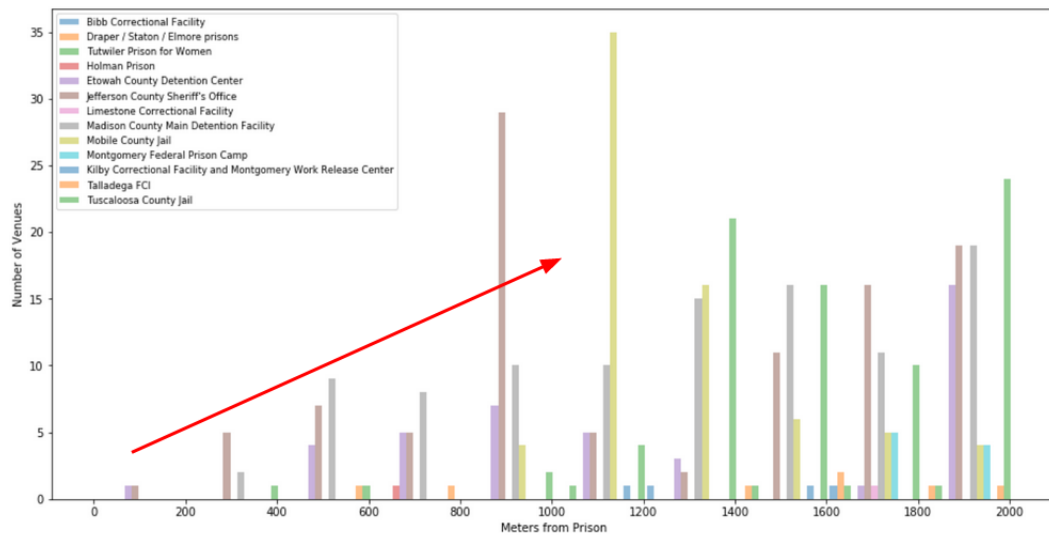
Figure 4.4: Histogram of all the prisons and the venues

We calculate the average value of near venues for all the prisons. Using regplot from seaborn library, we show a positive linear relationship between the meters from a facility and the number of venues, Figure 4.5. So, again, the presence of a prison impacts its surroundings in a negative way, reducing the number of venues in its vicinity. The number of venues grows as we go further away from the facilities.
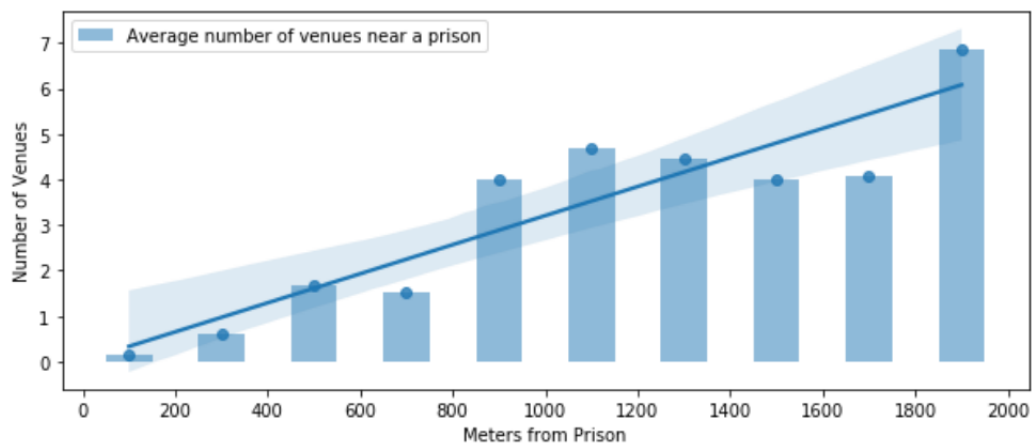

Figure 4.5: Average number of venues near a prison

4.0.1 Cluster of prisons

As the final step of this analysis we apply a machine learning technique to cluster the prisons based on the distance to the venues. We use K-Means and the number of clusters was set to 3. The Map is shown in Figure 4.6.

When we analyze each cluster, we see that the violet dots are the prisons with almost no venues near them, the red dot is the Mobile County Jail that has a medium amount of venues, concentrated in the range of 1000-1200m. The last cluster corresponds to the green dots, which includes Jefferson County Sheriff's Office, Madison County Main Detention Facility and the Tuscaloosa County Jail. These facilities are the ones with the most venues in their surroundings.

The top 5 biggest cities in the state of Alabama are: Birmingham ,Montgomery ,Mobile ,Huntsville, Tuscaloosa. There is a strong correlation between the city size and the number of venues near a facility. Green dots are in Birmingham,Huntsville and Tuscaloosa. Mobile is the red dot, and In Montgomery there are a lot of facilities but they are far away from the city.
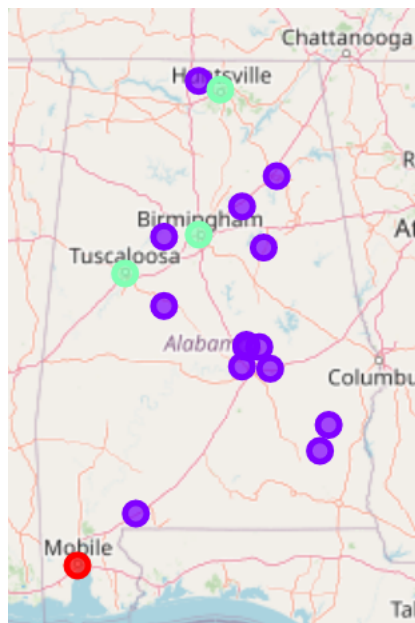

Figure 4.6: Cluster Map

As a business recommendation, we recommend investors try to stay at least 1000 meters away from a facility, depending on the city. Keep in mind that a cluster of shops attracts more shoppers, and in zones near prisons it is more likely that there won't be a lot of venues around.

Chapter 5

Conclusion

We describe the impact of a prison on the number of venues on their surroundings. We find that in the vicinity of a facility it is difficult to find a venue. There is also a clear growing pattern in the number of venues as we go farther and farther away from the facilities. There is a linear positive relationship between the average number of venues near a prison and the distance from it. Finally, we cluster the prisons by the similarities on their influences, finding that the biggest cities are the ones that have prisons with most venues around. We recommend the investors that try to stay at least 1000 meters away from a facility, but this depends on the city.

Bibliography

[1] Steve McLinden. https://www.bankrate.com/finance/real-estate/will-n earby-prison-affect-home-resale-value.aspx.

[2] Ph.D. Katherine A. Carlson. Prison impacts: a review of the reaserch. *Peninsula College, Port Angeles*.

[3] PBS. https://www.pbs.org/independentlens/blog/prison-economy-how-do-prisons-affect-the-places-we-live/.

[4] https://en.wikipedia.org/wiki/alabama_department_of_corrections.