

Unbanking in NYC

Sarah Aita¹

¹New York University

July 16, 2021

Abstract

The purpose of this project is to explore whether there is a correlation pattern between the number of banks and banking status, as well as to find the driving factors behind the banking status of New York City residents. The analysis concluded that there is a weak correlation between the availability of banks in a neighborhood and the ratio of unbanked households. The increase in the ratio of unbanked households is mainly related to two socio-economic features: poverty level and unemployment.

<Sarah Aita, saraaita, soa238>

Introduction:

In 2017, more than 7.8% of New York State's population did not have any bank accounts, compared to a national rate of 6.5%. This rate tends to be even higher for New York city. In 2013, the unbanked households' rate in New York City was 11.7%, compared to a state rate of 8.5% and a national rate of 7.7% at the time. The highest unbanking rate is found in the Bronx, where it exceeds 20% ([FDIC, 2019](#)).

In 2015, [the Urban Institute published a report](#) that focused on identifying neighborhoods that have a high ratio of unbanked populations. Furthermore, the report investigates how neighborhood characteristics, such as income, poverty and demographics might relate to the banking status of its residents. This research project builds up on the [Urban Institute's report](#) and attempts to identify which socio-economic feature if a neighborhood affects residents' banking status.

Data:

This analysis uses [data from the Urban Institute](#), which contains the ratio of unbanked and foreign born population, unemployment, poverty, and median income for each neighborhood. To calculate the number of banks in each neighborhood, I used the [Department of Consumer Affairs \(DCA\) list of operating businesses](#). I selected all banks on the list, and joined each bank with their corresponding neighborhood. This allowed me to count the banks in each neighborhood and compare it with the [Urban Institute's dataset](#).

One underlying limitation in the available data is the geographic unit used. The only publicly available data about banking status was classified by Public Use Microdata Areas (PUMA). New York City contains 55 PUMAs, each of which contains a number of census tracts. Combining census tracts results in loss of information about the individual socio-economic characteristics of each census tract. More granular information could be insightful for this analysis and would increase the model's accuracy and precision.

The other limitation lies in [DCA's dataset](#). In order to identify the PUMA in which each bank is located, the latitude and longitude of each bank must be given in the data set. Some values in the data set did not have the geographic location and had to be dropped.

Methodology:

After processing the data, I analyzed the correlation between all variables to identify relationship patterns. Then, I selected the independent variables with the strongest correlation coefficients, and ran a regression analysis against the dependent variable, which is the ratio of unbanked population.

The dependent variables in this model are strongly co-variate, and it would have been a better option to go with more complex models such ridge regression. However, the linear regression model had an r-squared of 0.859, which means that the linear regression model is probably a good fit. It is noteworthy, that regression analysis was also used in the Urban Institute's analysis to create their prediction model.

Conclusions:

The correlation analysis showed that there is a strong positive correlation between banking status on one hand, and unemployment on the other. Similarly, there is a strong negative correlation between banking status and median income. These findings are in-line with Urban Institute's results. The Urban Institute's report identifies the Bronx as the borough with the highest unbanked population, highest poverty and unemployment level and lowest median incomes.

There is a weak correlation between banking status on one hand, and the availability of banks and percentage of immigrants in a neighborhood on the other. When comparing boroughs, the Bronx has one of the lowest number of banks, and a relatively high percentage of immigrants. So, I expected that these two variables to be strongly correlated to unbanking.

The regression model showed that the effect of unemployment rate is insignificant. With a 95% level of certainty and controlling for the number of banks and ratio of foreign born residents, the model concluded the following:

- The average ratio of unbanked households in a neighborhood increases by 0.85 per 1% increase in unemployment ratio
- The average ratio of unbanked households in a neighborhood increases by 0.53 per 1% increase in poverty ratio.

Future work:

Further research using more detailed information about the unbanked population is needed to be able to generalize the results. I recommend doing further analysis using more granular geographic units such as census tracts. Moreover, I recommend adding time as a new feature in this analysis. More insights can be gained by creating timeseries models to track the change of unbanking levels with socio-economic events.

References and Links:

**** Link to project's python notebook :** https://github.com/saraaita/PUI2018_-soa238/blob/master/00_Project/What%20socio-economic%20feature%20contribute%20to%20unbanking%25

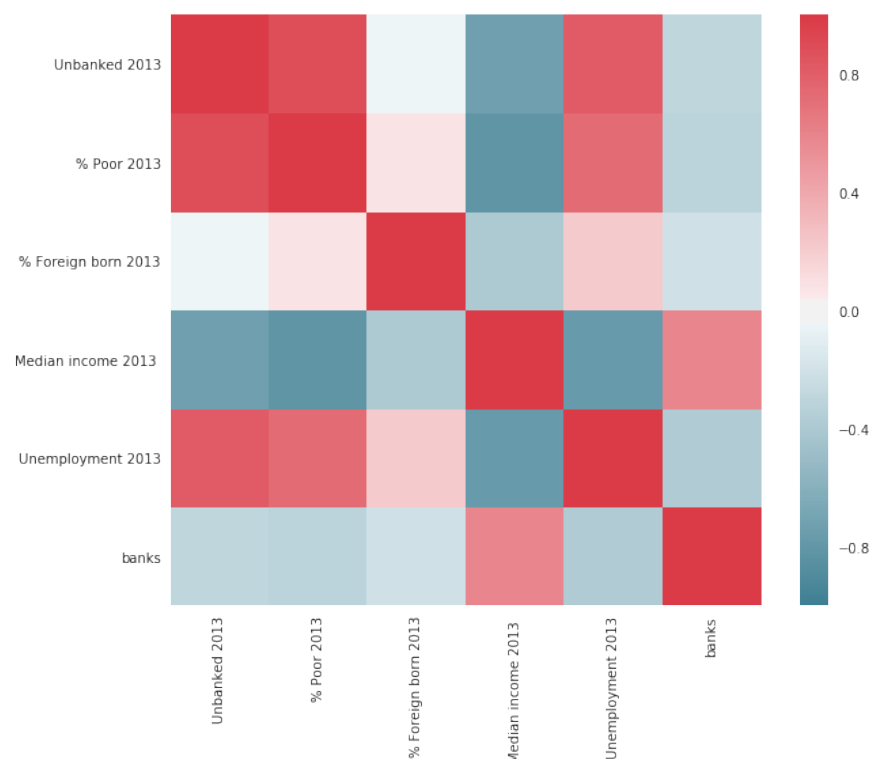


Figure 1: Fig.1 shows a correlation matrix between all features. There is a weak correlation between the availability of banks in a neighborhood and the ratio of unbanked households. The increase in the ratio of unbanked households is mainly related to two socio-economic features: poverty level and unemployment.

%20(1).ipynb

1. Federal Deposit Insurance Corporation (FDIC). 2019. "How America Banks: Household Use of Banking and Financial Services." FDIC. Retrieved from: <https://www.fdic.gov/householdsurvey/>
2. NYC Department of Consumer Affairs (DCA). "Legally operating businesses." NYC Open Data. Retrieved: <https://data.cityofnewyork.us/Business/Legally-Operating-Businesses/w7w3-xahh/data>
3. Ratcliffe, Caroline, Signe-Mary McKernan, Emma Kalish, Steven Martin, Timothy Meko, and Benjamin Chartoff. 2015. "Where Are the Unbanked and Underbanked in New York City?" Washington, DC: Urban Institute. Retrieved from: <https://www.urban.org/sites/default/files/publication/71511/2000430-Where-Are-the-Unbanked-and-Underbanked-in-New-York-City.pdf>
1. Ratcliffe, Caroline, Signe-Mary McKernan, Emma Kalish, Steven Martin, Timothy Meko, and Benjamin Chartoff. 2015. "Data Visualization: Where Are the Unbanked and Underbanked in New York City?" Washington, DC: Urban Institute. Retrieved from: <https://www.urban.org/interactive-map-where-are-unbanked-and-underbanked-new-york-city>

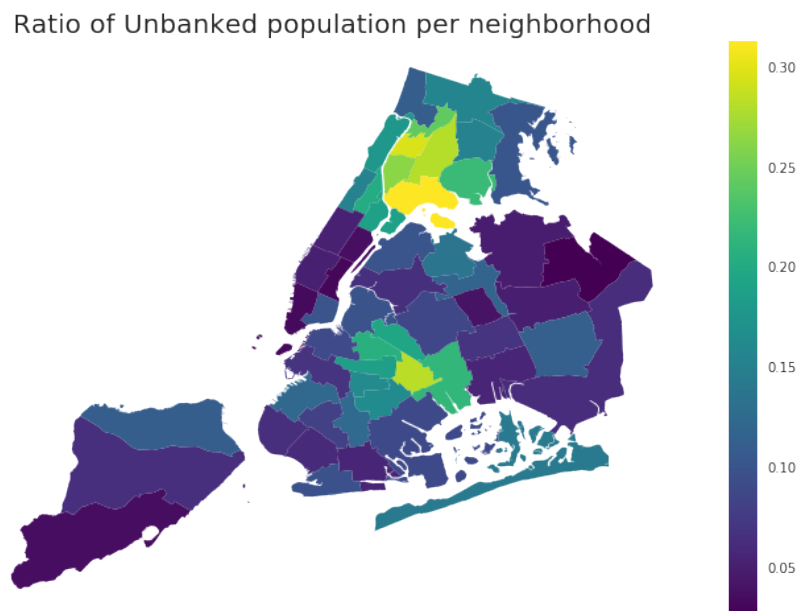


Figure 2: Fig. 2 shows the ratio of unbanked populations per neighborhood. The highest unbanking ratio is found in the Bronx, followed by Brooklyn, and the lowest is in the east side of Manhattan and south Staten Island.

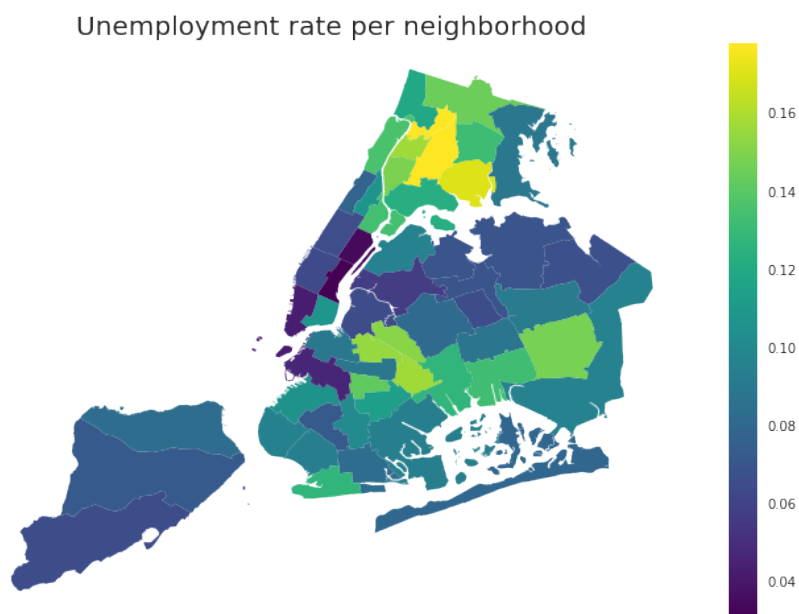


Figure 3: Fig.3 shows the unemployment rate in each neighborhood. The highest unemployment rate is found in the Bronx and the lowest is in the east side of Manhattan. By comparing this map to the map of unbanked households in Fig.2 , it is clear that there is a correlation.

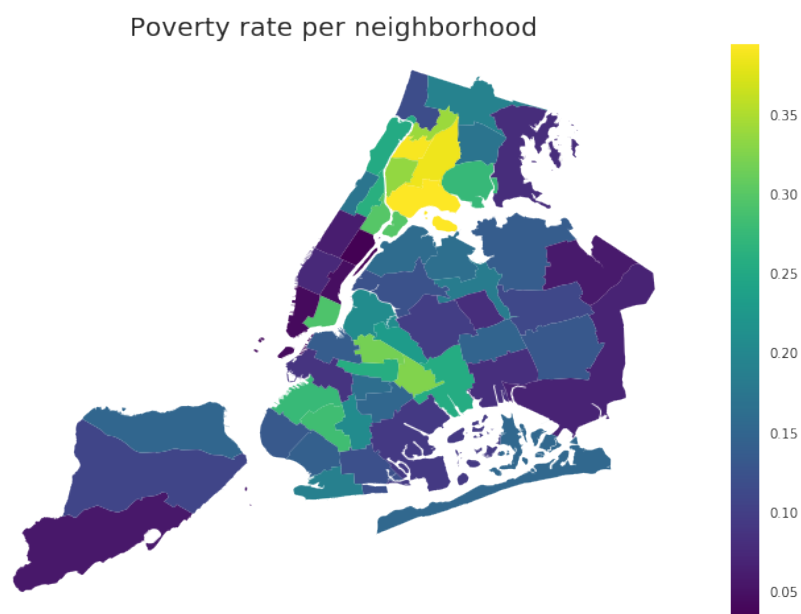


Figure 4: Fig.4 shows the poverty rate in each neighborhood. The highest rate is found in the Bronx and the lowest is in Manhattan and Staten Island. Similar to the unemployment map, there is clear correlation with fig. 2