

RACIAL DETERMINATION IN NYC SCHOOL ENROLLMENT PROFILE

Camilo Navarrete (ch.navarrete.lineros@gmail.com)



RACIAL DETERMINATION IN NYC SCHOOL ENROLLMENT PROFILE

Camilo Navarrete (ch.navarrete.lineros@gmail.com)

The USA society racial segregation problem through its history. One of the most diverse ethnic areas in USA is the New York City, in consequence it reflects the result of public policies regarding this issue. After more than half a century of the civil rights movements and the UNESCO implemented policies, we still can talk describe schools based on the race and poverty index. According to the present analysis we can profile schools as: "Black schools with high indexes of poverty", "Hispanic schools with high indexes of poverty" and "mixed white-Asian schools with medium indexed of poverty". Despite the discouraging scenario, this could be explained for different causes such as the migratory waves to USA since the 70s. Also, we can argue the Asian and Hispanic differences in the family income and cultural background as a cause of this profiling. Also, the Black segregation may reflect consolidated urban areas with an historical background. In conclusion a racial determination in NYC school enrollment does not reflect necessary the failure in the public policies, but it's more a snapshot of the historical moment.

Keyword: *racial segregation, public school, public policies*

Disclaimer: The USA Census Bureau [thub.com/IBM/skillsnetwork/wiki/Watson-Studio-Setup](https://www.thub.com/IBM/skillsnetwork/wiki/Watson-Studio-Setup) uses the term "Black", "Asian", "White", "Hispanic". Also, the official data from the NYC uses the same term. It's far from the author the intention to make any discriminatory statement using these words.

Contenido

Introduction	3
Methodology	3
Data source and problem	3
Data Quality Assessment	3
Feature Engineering.....	3
Models and Performance	3
Results	4
Discussion	4
Conclusion	5
References	6

Introduction

The education as a human activity faces a challenge to keep a balance between the new challenges of the moment and keep the cultural bases such as language, traditions, among others. For this reason, the educational political policies could shape the society. The last century has a plethora of examples including racial segregation, religious rivalries, totalitarian nationalism, etc.

The educational public policies have been strongly determined by the recent history of the last century, the effect of multilateral international agreements and new research works in philosophy and psychology. In consequence, social and political pressure emerges and make people take decisions that have a high economic cost and may affect a complete generation of school graduates. For this reason, is necessary to determinate the scenario, identify the problem in order to evaluate among multiple alternatives.

Among the most challenging problems a society have to face, we can identify the need to give education for a non-homogenous population. The source of this diversity could be religious, ethnic, a political affiliation, etc. What is more complex could be all together at the same time. This situation is quite common in the economic centers, we can find classic examples such as capital of the Roman Empire, the Hellenic Alexandria of Egypt, the Paris of fin-de-siècle or the current New York City for the United States of America. We will focus in the later one as a paradigmatic example for the case.

One of world known characteristic of New York City (NYC) is the ancestry origin diversity. For this reason, the enclaves like the little Italy, the China town, the Jewish neighborhood in Brooklyn, the Black Harlem, etc., are portrait in many of the popular culture movies, TV shows, comics and are easily identify as a synonym of NYC. But this arise a question, is there something known as a NYC identity? How the NYC diversity is compatible with general political policies? In order to address these questions, we will make a diagnosis of the current composition in the NYC public schools and then discuss facing the historical data.

Methodology

Data source and problem

To answer the question "what is the composition in the NYC public schools?" we will determinate a period and some features. We will use the school-level student demographic and enrollment data from 2014-15

through 2018-19 provided by the NYC Department of Education. Then the approach will be to find among all the features some of them comparable and identify patterns to get a profile among the different schools.

Data Quality Assessment

There are some features in the DataFrame such as "Grades" that are not in all the schools. For this reason, is better to do not consider them for this analysis. Other characteristic is the duplicate information of the students as raw number or percentages. Since the percentage allow us to have some standardization in the data, we will continue only with it.

Finally, we can observe there are data from five years. Maybe they are not comparable, and we are mixing different things. For this reason, the first thing will be to make an analysis in comparing each of the selected feature per year to evaluate if we can consider all as a big group. For this research we will consider each point even if we are considering the same school from different years putting aside the problem of the pseudo replicates.

Then we can analyze some features regarding the Business Intelligence problem, in this case 'English Language Learners' is a variable to consider with caution, because a high value would indicate a public school of migrant students and make a bias over the distribution. Maybe the racial segregation is forced for causes of adequation desired for the community. In the same way 'Disabilities' it's not a variable dependent to any economic index or ethnicity. In both cases it's better to stablish a range for the query based on a previous histogram.

Feature Engineering

The first thing in this case was dealing with the 'Year' feature, to convert this one to a data type from a composed string name. Then drop the null data if there any.

Models and Performance

The proposed approach after the Exploratory Data Analysis and the feature engineering data will be to reduce the dimension of the data frame with Principal Component Analysis (PCA) and then using the t-Distributed Stochastic Neighbor Embedding (TSNE).

After obtaining a two-dimension Data Frame a plot will be used as a guide to have a first insight in the possible clusters. But, to have a robust support to the decision there will be implemented the K-Means based on

centroids determination, the Spectral clustering and the DBSCAN.

The original data will be used to a Deep Learning training using autoencoder to reduce the noise. Then the resulting Data Frame will go to the same Dimension Reduction and Clustering to decide the better result.

Results

Based on the results of the Deep Learning the Reduced Dimension following a PCA->TNSE allows to obtain a good fit for the Spectral Clustering algorithm. The results in **Fig 1.** allow us to distinguish three clear groups of similar size where the % Black and % Hispanic students make two clear profiles in the public schools, both with high economic needs. On the other hand, a more diverse group with a high, but not dominant, % White students with medium economic needs.

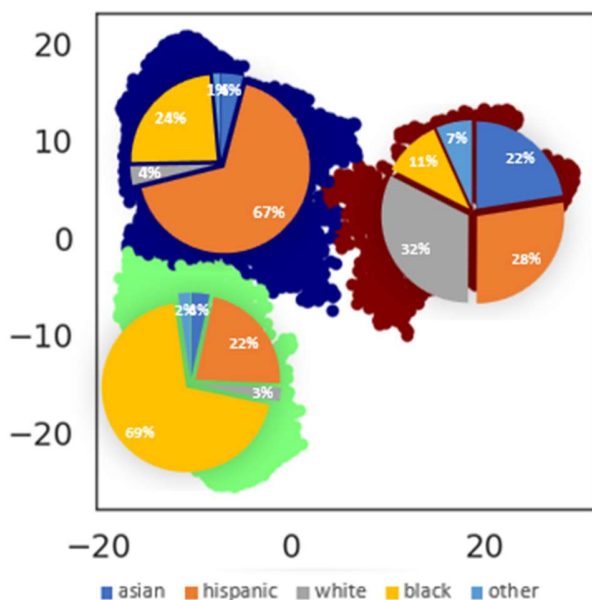


Fig 1. Clustering for the public-school students of NYC from 2013-2018

The NYC 2010 Census indicates a racial composition of 44.6% White, 25.1% Black, 27,5 % Hispanic and 11,8% Asian. According to this the more diverse cluster is closer to the demographic data. One interesting result is the mutual exclusion between dominant Hispanic schools and Black dominant schools.

Discussion

The public-school profile of the NYC students indicates a natural clustering scenario mainly based on the race. The deep learning helped us to reduce the noise in the analysis and establish three clusters. This is in agreement with the historical data of NYC. Despite the short story as a big economic center, the city has received different waves of migrants. It's quite normal that the ones who arrives tried to find contact networks of jobs in the new land based on the relateness, for this reason the self-forming neighborhoods are present. The reason of NYC as a constant center of foreigners it out of topic of this research.

The cluster 1 indicates schools with high percentage of Hispanic, English language learners with high economic needs. This corresponds with a recent migration of Mexicans show below. This profile of new arrival citizens with low incomes could be related with the Irish migration about two centuries ago or the Italians at the beginning the last century. This may indicate that more than a racial segregation the ethnic profile in some self-formed neighbors could be one characteristic of the migration as a phenomena per se.

In the same line of the last part of the analysis, we can argue that "% White" it's quite ambiguous since a mixed racial person Irish and Polish ancestry classifies just as white and not as Multiracial, despite the differences in both cultures. We can argue the same of the Black Hispanic called "mulato" in Latin-American, or the cultural differences of a 4th generation Black resident and a migrant from Africa. For this reason, we should complement the analysis with a geographical distribution of the racial groups.

The geographic distribution data obtained from Mae Francke job clearly shows:

1. mutual exclusion of White, Hispanic and Black population.
2. multiple ethnic centers of the same "race classification"

Based on the recent history of USA we can relate the Black segregation since laws like Interracial marriage has been fully legal in all U.S. states since the 1967 Supreme Court decision that deemed anti-miscegenation state laws unconstitutional (via the 14th Amendment adopted in

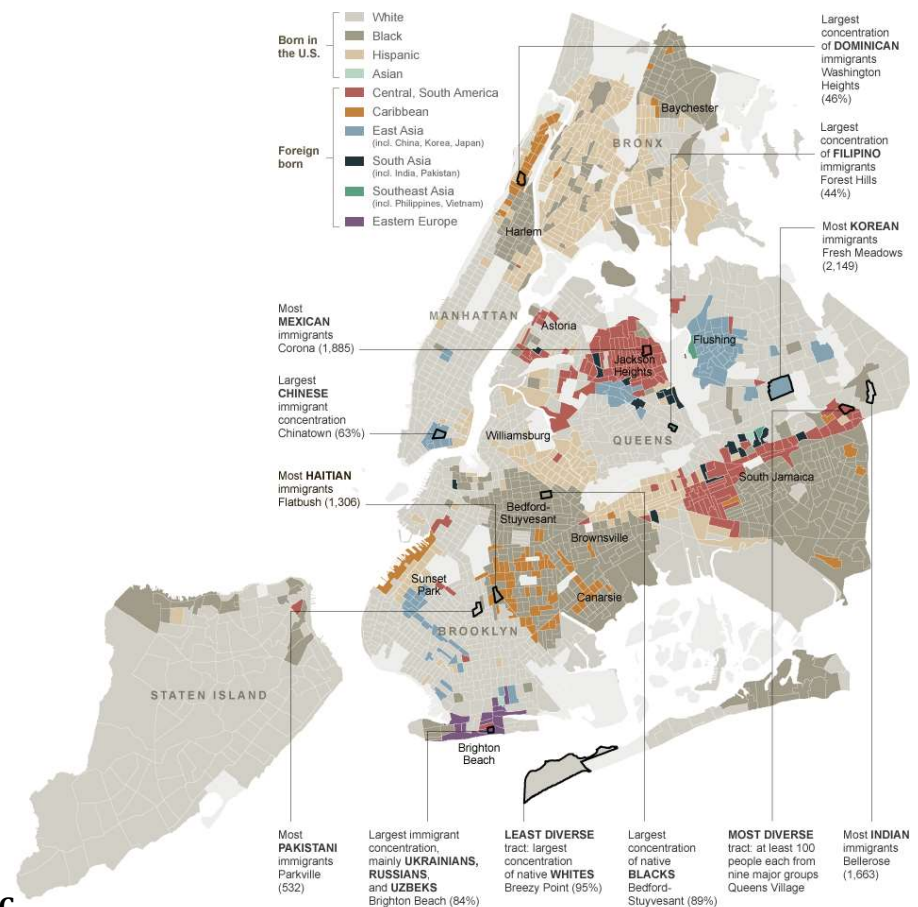


Fig 2. Ethnic enclaves in NYC

1868) with many states choosing to legalize interracial marriage at much earlier dates.

This is quite different from the "Hispanic" culture since it was common since the first law in the Modern Spain in 1514. In the same way we can argue the resistance of the Dominican migrants to self-identify as Black as an issue in the racial profile, or the classification problem of migrants from Uruguay and Argentina to be classified as "White". The term by itself has arisen many attempts to solve this like "white Hispanic", "non-white Hispanic", among others.

Despite the problems, we should see the natural distribution of the citizen in the map of **Fig 2.** as a cue to understand how is naturally understood. According to this, the racial profile 1. reflects the migratory phenomena 2. it may hide more clusters due the ambiguity of the racial classification.

I would suggest a better enrollment profile considering the culture (ancestry of the students or parents), how many generations has been the family in USA or NYC, income level and other items such as religious feature and some

classification of the conservative/liberal score of the family. This proposed approach would allow to distinguish between two groups of Asian such Indian vs Filipino. Or religious groups like orthodox Jews families from secular ones. A better profile in the student gives a better picture of the scenario to implement educational public policies and evaluate how effective would be the implementation of strategies to migrant integration to the USA society, stop the drug in the youngsters or the bullying.

Conclusion

"What is the composition in the NYC public schools? This research approach shows that there is a strong weight on the racial level. And we can identify at least three clusters: i) multi racial with medium economic needs ii) schools with high percentage of Black students and high economic needs iii) schools with high percentage of Hispanic students and high economic needs. This clusters reflect the geographic distribution of the ethnics in NYC caused by the migration. Nevertheless, I strongly suggest a profile including cultural features to correlate better with the population neighborhoods.

References

School-level student demographic and enrollment data from 2014-15 through 2018-19 Data Provided by NYC Department of Education <https://data.cityofnewyork.us/Education/2018-2019-School-Demographic-Snapshot/45j8-f6umm>

<https://www.6sqft.com/infographic-shows-387-years-of-immigration-patterns-in-nyc/>

<https://www.census.gov/topics/population/race/about.html#:~:text=Black%20or%20African%20American%20%E2%80%93%20A,tribal%20affiliation%20or%20community%20attachment.>

<https://archive.nytimes.com/www.nytimes.com/interactive/2011/01/23/nyregion/20110123-nyc-ethnic-neighborhoods-map.html>

<https://rpubs.com/maefrancke/raceinnyc>