

# Can '311 Service Requests' data be used to accurately predict gentrification in New York City neighbourhoods?

Team 8 - Timur Khalikov, Nicolas Anguita, Rayan Hatout, Ishan Paidhungat

## Highlights

- Processed and cleaned NYC census and 311 call data.
- Explored and visualized the data to develop hypothesis
- Tested hypothesis that 311 call data could be used to predict gentrification in NYC.
- Concluded that using the classification models we did, the 311 call data cannot be used to accurately predict gentrification in New York.

## Background

Gentrification is a process with positive (e.g. increased diversity) and negative (e.g. forced exit of low-income residents) socioeconomic impacts on cities.

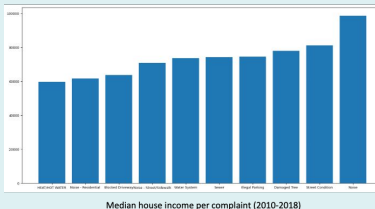
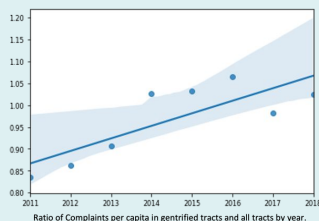
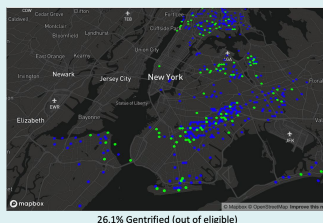
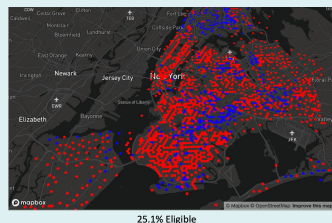
Anticipating this phenomenon would help the government plan in advance to reduce its negative effects. It would also help identify regions where property values are rising.

After some preliminary research, we hypothesized that the types and volumes of '311 Service' requests' may be a good indicator for gentrification.

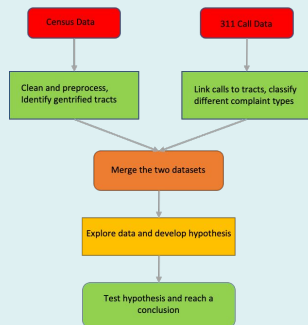
## Initial exploration

After computing which tracts are gentrified, we visualised them on a map. We also tested two hypotheses:

- 1) Number of complaints per capita increases faster than average in areas undergoing gentrification
- 2) Types of complaints is related to the tract's median household income



## Data

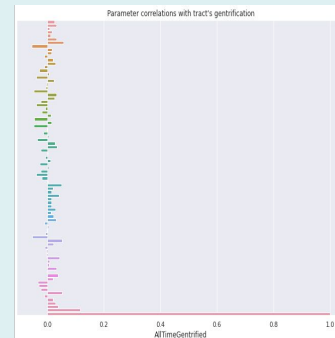


## Model

We computed the yearly changes for each type of complaints in every tract. We then considered the ratio of these changes to changes in the city as a whole, as there was a stable growth in total volume in 311 calls. Finally we considered the ratios of total changes in volumes of calls in each tract for each type over the given 8 year period. We expected all of these variables to increase with gentrification.

## Analysis

The results did not meet our initial expectations. The data suggests that there is negligible correlation between 311 call data and gentrification (shown in the graph below). The variable with the highest correlation was total complaints per capita with 0.12, which is almost negligible. Further Machine Learning techniques such as Random Forest, K Nearest Neighbours and SVC produced no positive results



## Conclusion

Our analysis suggests that, using the classification models we did, the 311 call data cannot be used to accurately predict gentrification in New York City.