Investment without Displacement:

An industry-driven approach to ethical investing

Pei Dong, Joel Perez Ferrer, Sean Nassimiha, Xinyi Xu

As a young professional in the New York City metropolitan area, one could ask themselves: "How do I invest into a - not too expensive - area of this city without contributing to the overall gentrification, and therefore displacement and eventually total exclusion of the native population?".

Millennials have been concerned with the issue of ethical and sustainable investment for a long time now, contributing for example to the explosion of ESG funds in the financial landscape.

To help answer this question, thus fostering a current of positive, sustainable, and *healthy* investments for the city of NY, we decided to study more in depth what defines "Investment without Displacement", what are the common factors of areas that endured gentrification in the past, and what suggestions could we have for the future of investing in NY.

Gentrification is not a novel concept. The earliest gentrification can be traced back to the 3rd century, where large villas were replacing small shops in ancient Rome and Roman Britain. Whilst there are certainly beneficial parts of gentrification, in terms of reduction in crime, better social environment, improved business opportunities… Concerns in increased local services cost, homelessness, loss of social diversity, and especially population displacement, were raised.

We studied the US Census[1] and data from 2009 to 2018 in the New York-Newark-Jersey City, NY-NJ-PA Metropolitan Statistical Area. Through the insights of the data, we identify a list of potential tracts that may absorb investment with the least amount of population displacement, when appropriate policies apply to. We then analyse the commonalities between these areas, and try to measure what defines a sustainable and ethical investment for the future.

1 Introduction and Executive Summary

We studied the gentrification phenomenon in the New York-Newark-Jersey City, NY-NJ-PA Metropolitan Statistical Area from 2009 to 2018. In particular, we fully classified whether a tract is undertaking ongoing gentrification, advanced gentrification or any of the other possible stages of gentrification (more details below).

Subsequently, we built a completely new category of classification: "investment without displacement", which has resulted into a new level of analysis, allowing us to compare between

[1] Census Bureau's 2009-2018 American Community Survey: https://www.census.gov/programs-surveys/acs/.

tracts that have gone through gentrification and displacement, and those that were at risk of displacement but eventually were able to avoid it.

We then focused our analysis on the industries present within each tract by looking at the percentage of people working in the various sectors. By doing so, we found that the tracts whereas people are primarily working in the *Arts* and *Construction* industries are most stable under external investment, avoiding the most severe forms of gentrification and therefore avoiding population displacement. On the other hand, *Manufacturing* showed to be the predominant industry in tracts that suffered from displacement. This could give an important metric for sustainable investment decisions in the future.

2 Classification

Identifying and quantifying displacement is hard, various existing methodologies were compared by Easton et. al[2], and different metrics were raised by scholars with different focus. It is important to distinguish gentrification and displacement[3].

We consulted the classification methodology developed in the Urban Displacement Project[4] conducted by Chapple et. al from the New York University, and the two-step test studied by US Governing[5].

2.1 Definition of terms

We set 2009 as our *base year*, and 2018 as our *end year* unless otherwise stated. Only tracts with at least 500 residents (in the relevant years) are considered. Tracts with a coefficient of variation of more than 30% on some of the following variables are flagged as unreliable and not taken into consideration: population, housing units, median rent, median house value, median income, college count and renter count. Tracts with missing individual variables are not considered.

The definition of terms we studied is as following:

Term	**Definition**
Low Income	A tract (in a given year) that percentage of low-income households is higher than the regional median.
Moderate to High Income	A tract (in a given year) that percentage of low-income households is lower than or equal to the regional median.

[2] Easton, S., Lees, L., Hubbard, P. and Tate, N., 2020. Measuring and mapping displacement: The problem of quantification in the battle against gentrification. Urban studies, 57(2), pp.286-306.

[3] Zuk, M., Bierbaum, A.H., Chapple, K., Gorska, K. and Loukaitou-Sideris, A., 2018. Gentrification, displacement, and the role of public investment. Journal of Planning Literature, 33(1), pp.31-44.

[4] https://www.urbandisplacement.org/maps/ny

[5] https://www.governing.com/gov-data/gentrification-report-methodology.html

Vulnerable (to gentrification)	A tract (in a given year) that median household value is smaller than 80% of regional median, and satisfies *three* of the following: - is a low-income tract, - percentage of college-educated population is lower than the regional median, - percentage of renters is higher than regional median, and - percentage of non-white races is higher than regional median.
Demographic Change	A tract (form base year to end year) that: - percentage change in college educated population is higher than the regional median, and - percentage change in median household income is higher than the regional median.
Hot Market	A tract (form base year to end year) that: - change in median real rent is higher than the regional median, or - change in median household value is higher than the regional median.
Gentrification	A tract that is vulnerable in the base year, and from base year to end year: - is a hot market, and - there is a demographic change.
Displacement (low income)	A tract (from base year to end year) that: - some gentrification signs can be found, and - there is an absolute loss in low-income households.

2.2 Typology classification

These are the main typologies that were employed in our investigation, though the complete set of definitions and classifications can be found in the reference as well as in the appendix. Besides the typologies introduced in the previous research we introduced an additional type in our method. We only consider tracts with at least 500 residents in the base year. Our study mainly focuses on the tracts classified as OG, AG and IWD, see appendix for full classifications.

Typology	**Typology Criteria**
Ongoing Gentrification (OG)	A tract that is: - a low income tract in the end year, and - gentrified from the base year to the end year.
Not losing Low-Income Households	A tract that is: - a low income tract in the end year, and - not classified as OG.

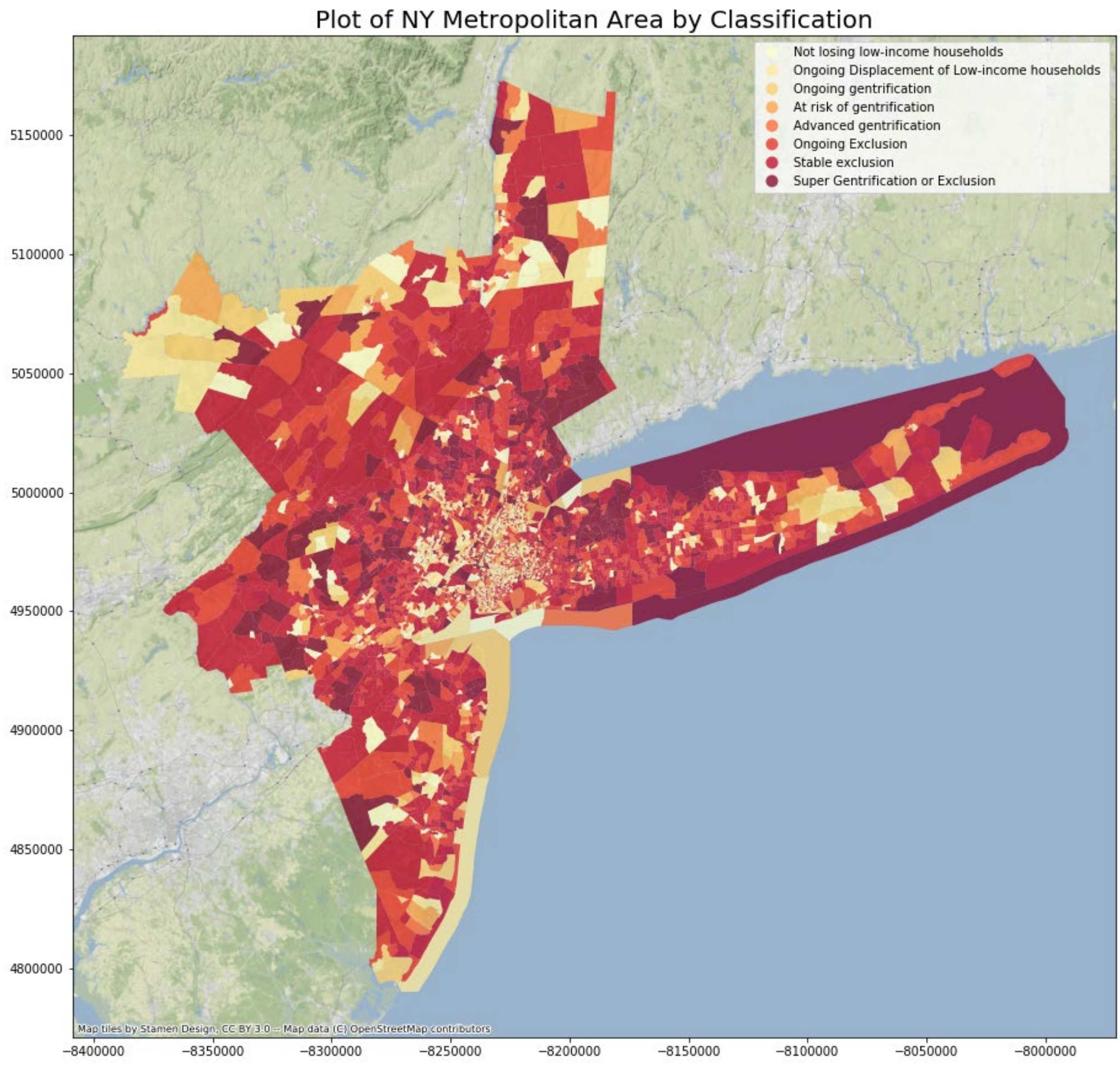
Advanced Gentrification (AG)	A tract that is: - a moderate to high income tract in the end year, and - gentrified in from the base year to the end year.
Invested without displacement (IWD)	A tract that is: - vulnerable in the base year, - not vulnerable in the end year, - a hot market from base to end year, - no demographic change, and not losing low-income households from base year to end year.

Note: as already stated, the classification "Investment Without Displacement" has been constructed entirely by us, as a consequence of the absence of a metric that could describe the successful stories of the NYC tracts - those tracts that were able to cope with gentrification and did not result in population displacement. By adopting this new definition, we were able to effectively compare the IWD areas to the gentrified areas, and understand what are the key differences that allowed IWD areas to avoid being negatively impacted.

2.3 Data Exposition

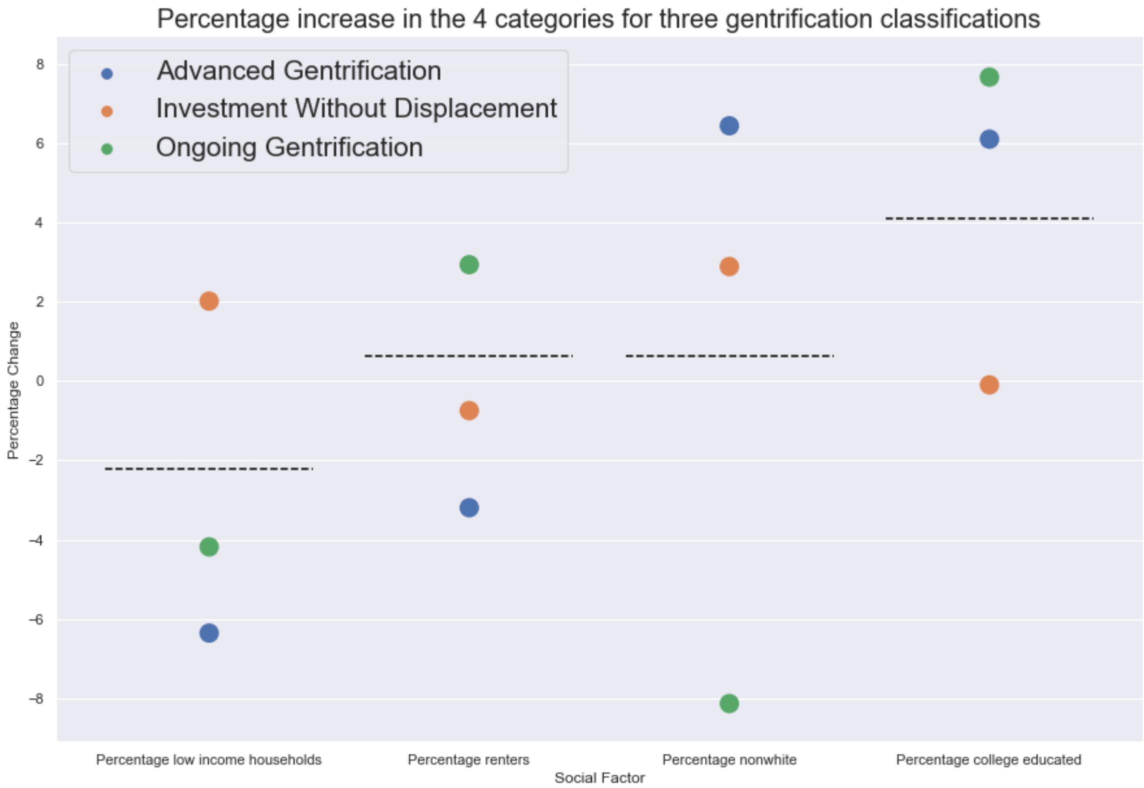
2.3.1 Demographic Changes by Classification Type

We studied the gentrification that happened between 2009 to 2018 in NY-NJ-PA Metropolitan Statistical Area, the full classification is in *Plot of NY Metropolitan Area by Classification*.



Among these tracts, 49 of them are OG, 13 of them are AG, and 66 are IWD.

As a sanity check, we decided to determine the percentage changes in four different population factors: *'Percentage of Low Income Households'*, *'Percentage of Renters'*, *'Percentage of Non-white population'*, *'Percentage of college educated'* in the AG, OG and IWD tracts between 2009 and 2018. The figure below shows the median of these percentage changes in tracts under each typology. (Full percentage change distribution is attached in the appendix.)

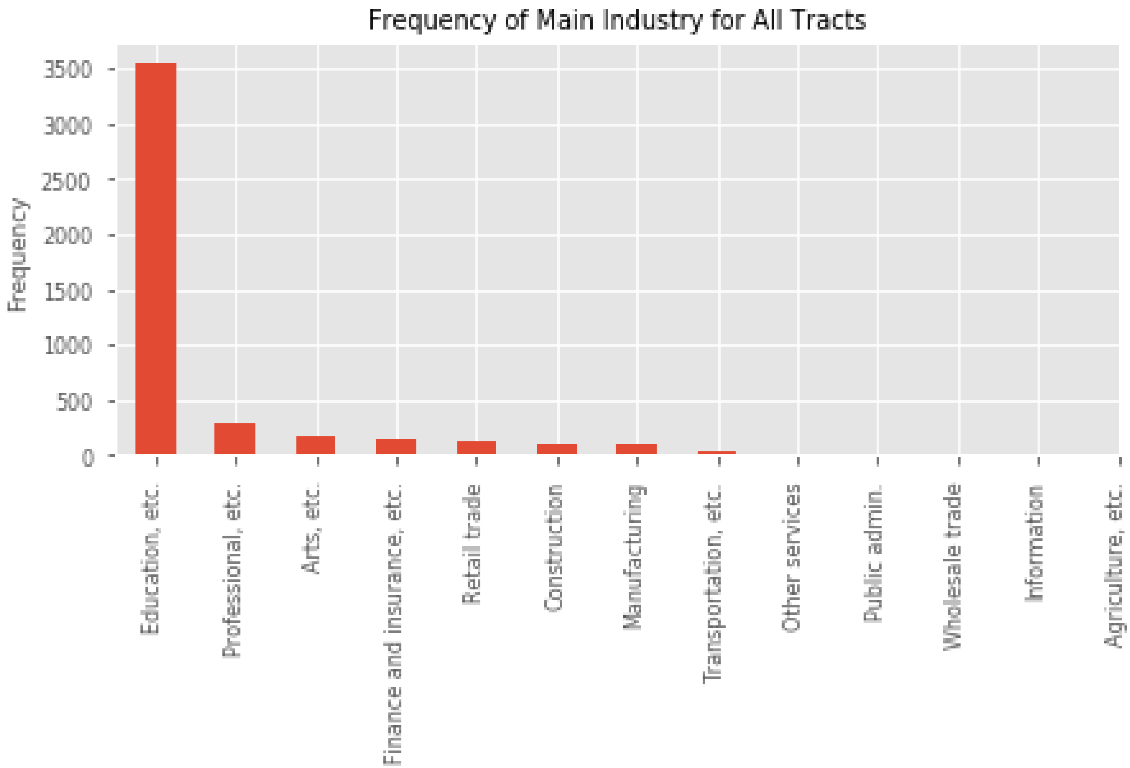


By looking at this plot, it is clear that areas where IWD happened did not go through the same changes as OG and AG areas for the percentages of low income households and of college educated people. This means that whereas AG and OG neighbourhoods saw a strong increase in average income, linked to a new influx of rich families, areas of IWD were able to avoid displacement. The dashed line represents the average change over all tracts. It is important to understand however that the gentrification categorisation itself was based on these four factors, and therefore this plot should be used only as a means of understanding the *magnitude* of these changes.

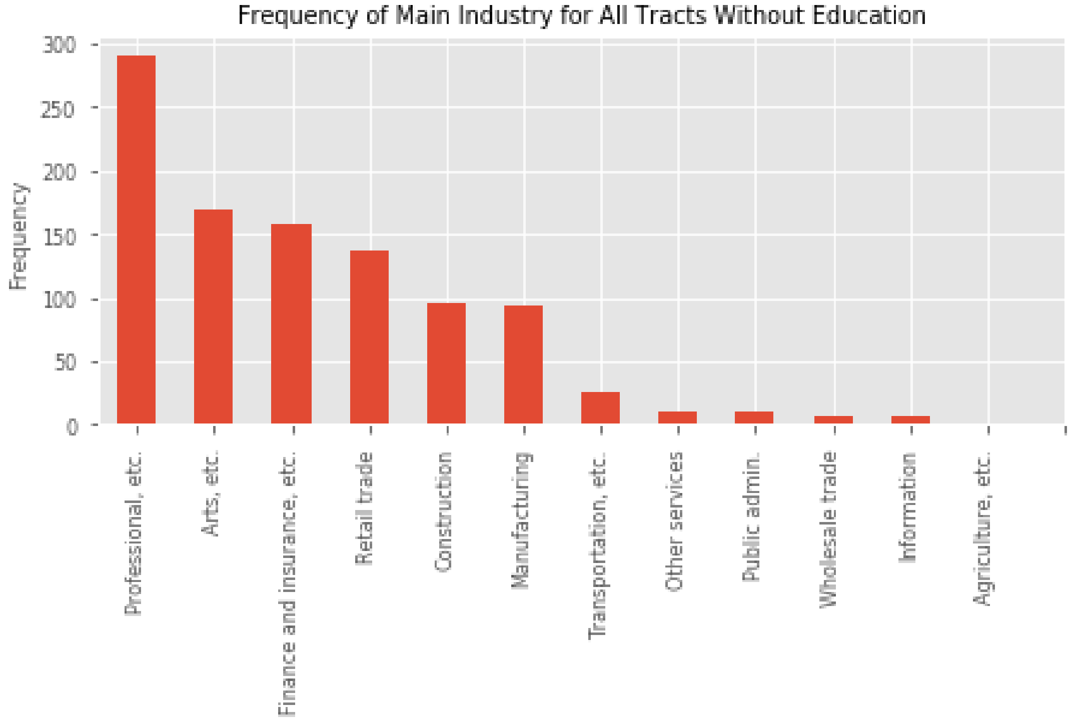
After the classification, new features were introduced from the census to understand what factors would influence and determine the future gentrification type of a specific tract . Our initial hypothesis was that the industry distribution of the workers was a major factor since it captured the characteristics of the sources of income from working class populations that mainly rely on their salary (as opposed to passive investments) to sustain their households.

2.3.2 Industry Distribution by Classification Type

As the centre focus of our exploration, we explored the extent to which the industry affected the outcome of gentrification. That is, over all tracts that are OG, AG or IWD, how the dominated industries (a tract is dominated by some industry if most of its working population are working in that industry) are distributed.



Since 76.49% of the tracts are dominated by *Educational services, and health care and social assistance* and this industry seems to have a fairly consistent distribution (i.e. for most tracts it takes 20-30% of the working population as seen in the distribution of people working in an industry in section 5.2 of the appendix), this industry was removed in order to explore the distribution across the remaining industries. This should help our analysis on the basis that, excluding *Educational services, etc.,* the remainder of the tracts are much more widely distributed in Main Industry.



A plot of the studied tracts highlighted according to the main industry (excluding education) can be found in the appendix.

3 Analysis

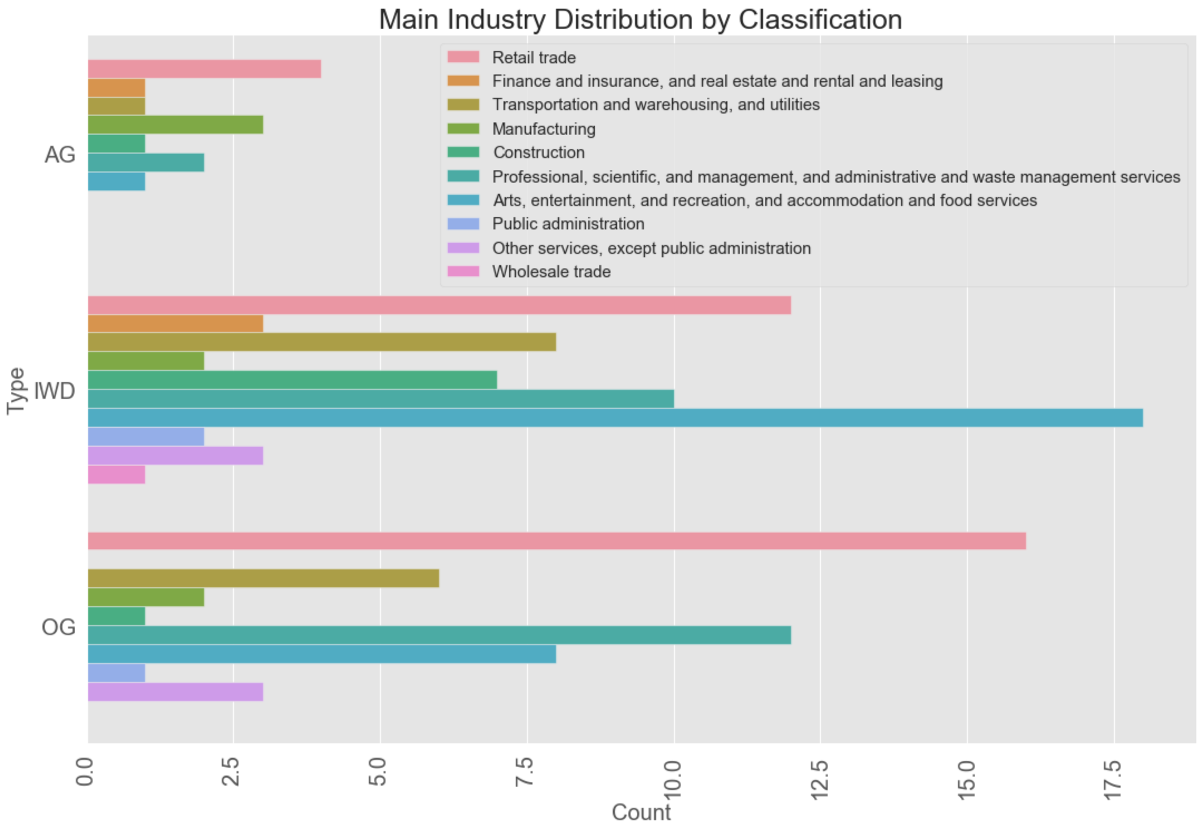
By applying our methodology and comparing the tracts with gentrification features (from 2009 to 2018) which however are showing the least population displacement, we spotted interesting shared features. We selected tracts with a potential to be gentrified with least risk on population displacement based on those features.

We executed simple linear regression on the various tracts using the percentages of population working in each of the defined industries. By doing so, we are able to make predictions on the *vulnerable* tracts present in 2018, giving information on the likelihood of those tracts to absorb investment without being displaced, and therefore creating a first rough guideline on ethical and sustainable investment.

3.1 Main Industry for Investment without Displacement

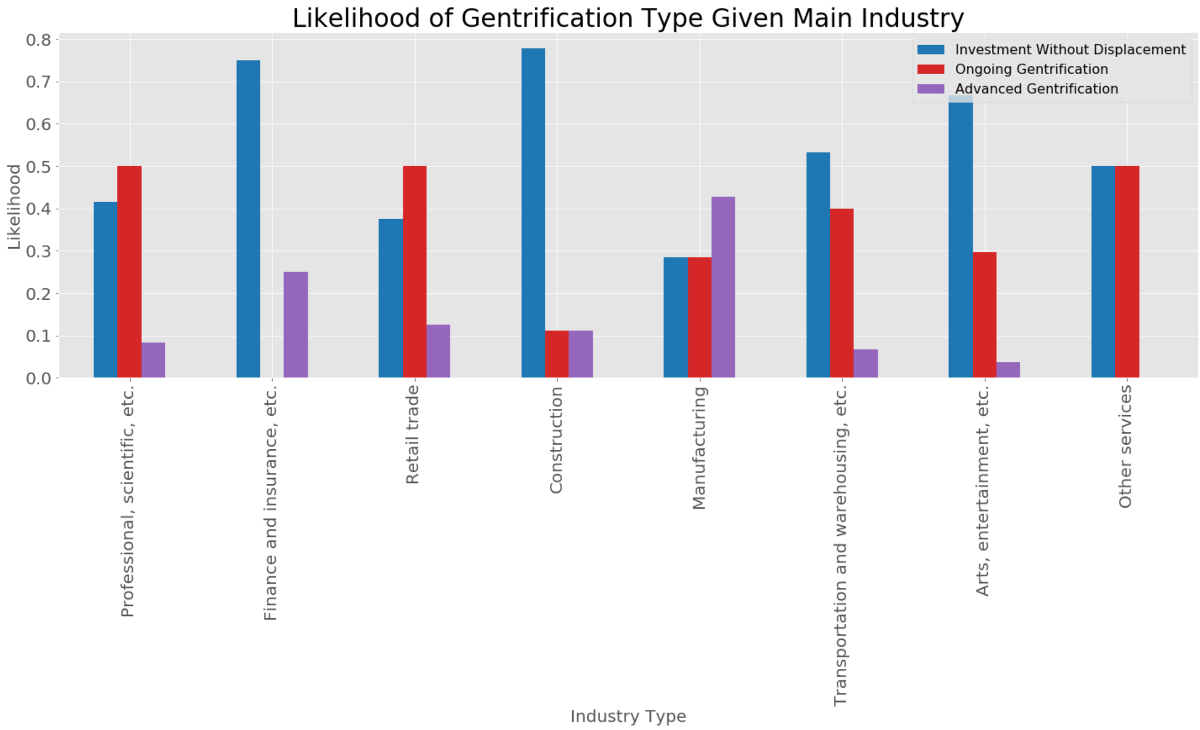
We decided to plot the main industries of each tract by focusing on the subdivision between AG, IWD and OG. This plot is useful to have a first insight into the distribution of the industries among the different gentrification categories.

We decided to plot the Main Industry Distribution (excl. Education) by Classification in order to understand how impactul different industries were and how much they contributed towards the gentrification type.



This figure reveals some key aspects regarding the impact that every industry has on the gentrification status of the tracts. Many of the tracts going through AG or OG show a high presence of *Retail Trade*, or *Professional, Scientific, etc*. main industries, whereas tracts where IWD happened show a higher number of *Arts, Entertainment, etc.* professionals.

One should notice that some of the industries are very poorly represented as *main* industries among the various tracts (wholesale trade professionals are not going to be as prevalent as retail trade professionals), and therefore specific distribution will not be statistically significant. The plot is therefore highly skewed towards more common industries, in such a way that the impact of common industries might seem higher than the impact of rarer ones. This issue has been dealt through the subsequent plot, where the likelihood of any classification type given the main industry was plotted. In this plot, any industry with less than 3 corresponding tracts was removed since the sample size was not considered sufficient.



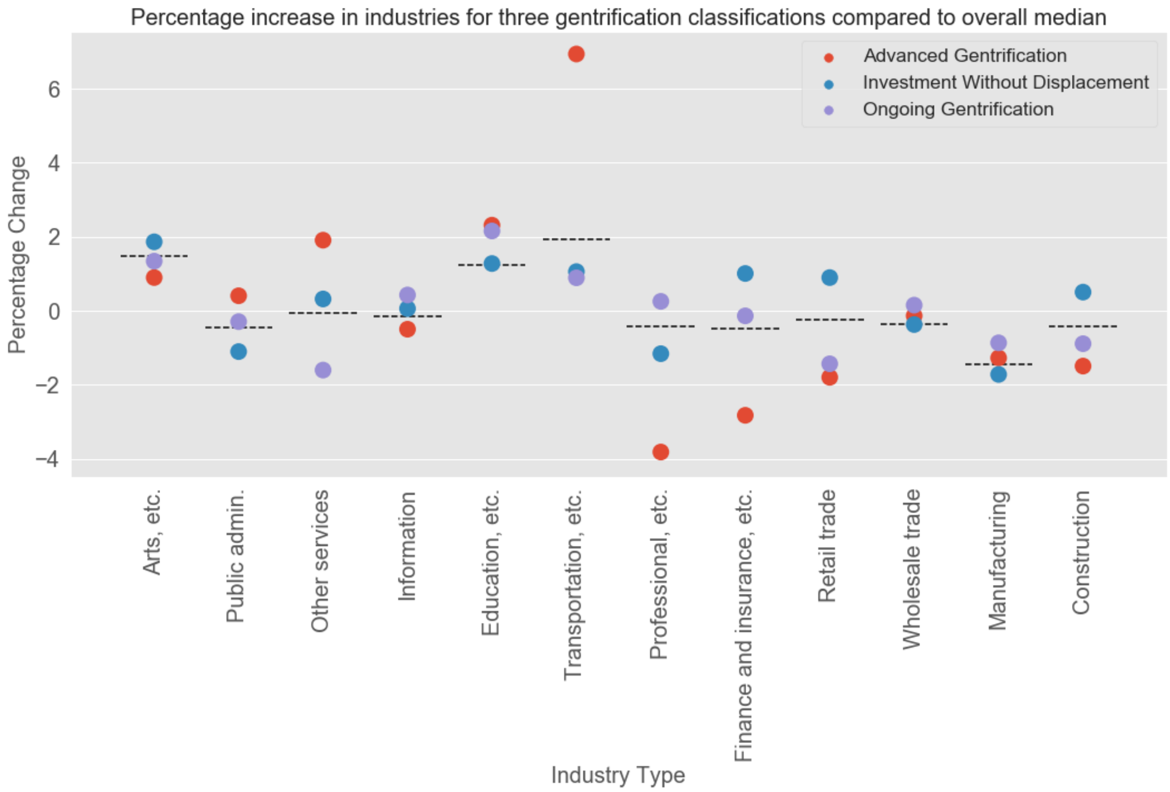
A clear pattern can be noticed. On one hand, *Construction* and *Arts, Entertainment, etc.* as well *Finance and insurance, etc.* industries strongly resist displacement upon first signs of gentrification, being the most stable when investment is injected into the tract. On the other hand, *Manufacturing*, and *Retail Trade*, are highly liable to gentrification. While these industries need not behave uniformly and are extremely complex and reliant on a wide variety of conditions to explain their growth, we formulated a few hypothesis as to why this might be the case:

- *Finance and insurance* are highly robust industries and a crucial backbone of New York's economy, a city that is objectively the largest financial hub in the United States and one of the most important in the world.
- *Construction* is the industry that often materialises these investments (building more valuable houses, renovations and such), therefore it seems intuitive that workers in such a field would benefit from more valuable houses being built around them.
- *Retail trade* is at risk if richer people move into the neighbourhood, as these individuals might not be satisfied by those existing family owned shops that could be replaced easily by their preferred alternatives.
- *Manufacturing* is also at risk given the increasing land value: factories are highly relied on their production plants. If the land rent rises, the fixed cost of operating a factory might rise to the point where the factory owner chooses to relocate the factory to another city, or even country to keep or improve profits. Also, richer people tend to live in eco-friendly areas, so the production lines that may harm the environment will be relocated. As a result, those manufacturers no longer need to live there.

3.2 Change in Industry Distribution by Type

While making these analyses, we have solely considered the main industries for each tract. This meant that most of the data regarding the percentage of population working in each industry within every tract was discarded. To remedy this issue, the following analysis will be based on the full industry distribution, and therefore will not only consider the main industry.

The percentage change of the population working in each industry for every tract between 2009 and 2018 was plotted below for the three gentrification classifications. To effectively read this graph, a percentage change close to the geographical-area average is a sign of lower displacement, whereas a higher deviation from the average is a sign of strong displacement, possibly caused by gentrification.



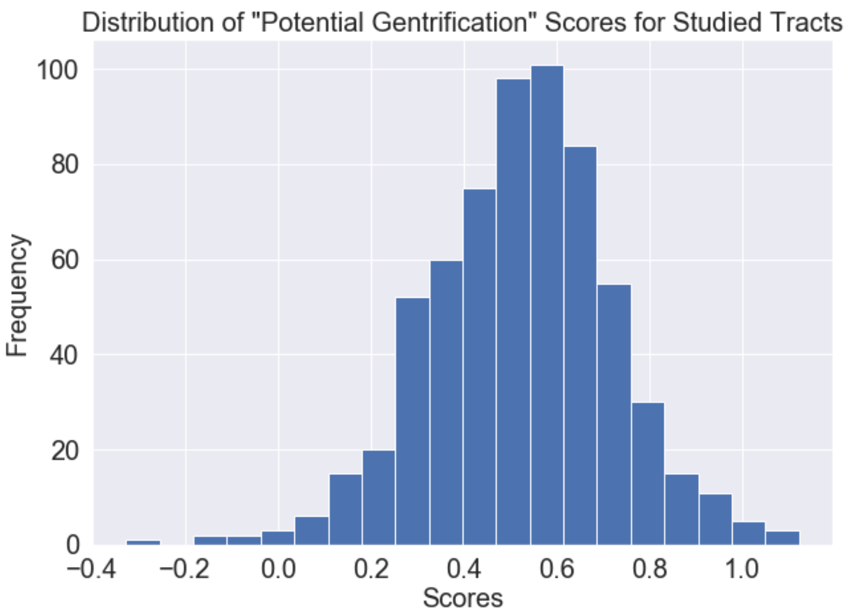
Areas of advanced gentrification saw a high increase of *Transportation* professionals, and a strong decrease in *Professional, scientific, and management, etc.* On the other hand, areas where investment without displacement took place have seen a fairly stable change in main industries, meaning that the population was most likely not displaced.

3.3 Linear Regression based on the distribution of Industry

In order to better quantify how much the industries contributed and how reliable they were, we decided to employ a machine learning approach. Given the very low number of datapoints in the dataset (of the order of low hundreds) we decided to use a Logistic Regression even though others were tried (such as the Naive Bayes method).

The Logistic Regression model was trained to classify whether a tract was likely to become gentrified (corresponding to label 0 if the type was Ongoing Gentrification or Advanced Gentrification) or not (corresponding to label 1 if the type was Investment Without Displacement). The model was only provided with the distribution of industries on the base year and the type that would eventually become.

Based on the model that was trained according to whether a tract became gentrified or not after investment, we made predictions upon tracts that were classified as vulnerable in the final year. These predictions signal how confidently we believe that a tract will become gentrified or not in the next decade based on the industry distribution of the population. We will also remove the tracts with small populations as these may have skewed distributions of industries.



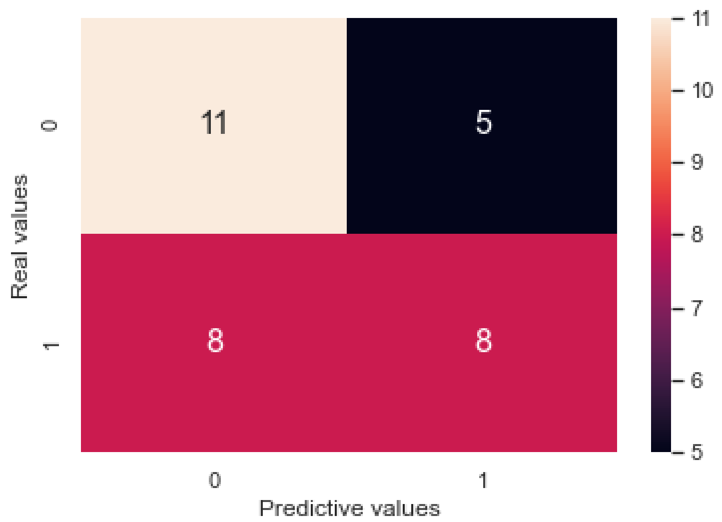
From here we can generate a sorted list of investments and classify them between *potential investment without displacement* if their predicted type is > 0.5 and *potential gentrification* if their predicted type is < 0.5.

The best and worst scorers for the distribution are shown in the table below.

Best Scorers		Worst Scorers	
GEOID	Score	GEOID	Score
36061000201	1.120874	36061008400	-0.327768
34013006600	1.070498	36061007600	-0.122147
36005040100	1.056847	34013016700	-0.111036
34013003100	1.044140	36047051100	-0.068386
34013018600	1.035507	34031175900	-0.063437

The complete predictions can be found in the attached code (2018predictionsfinal.csv).

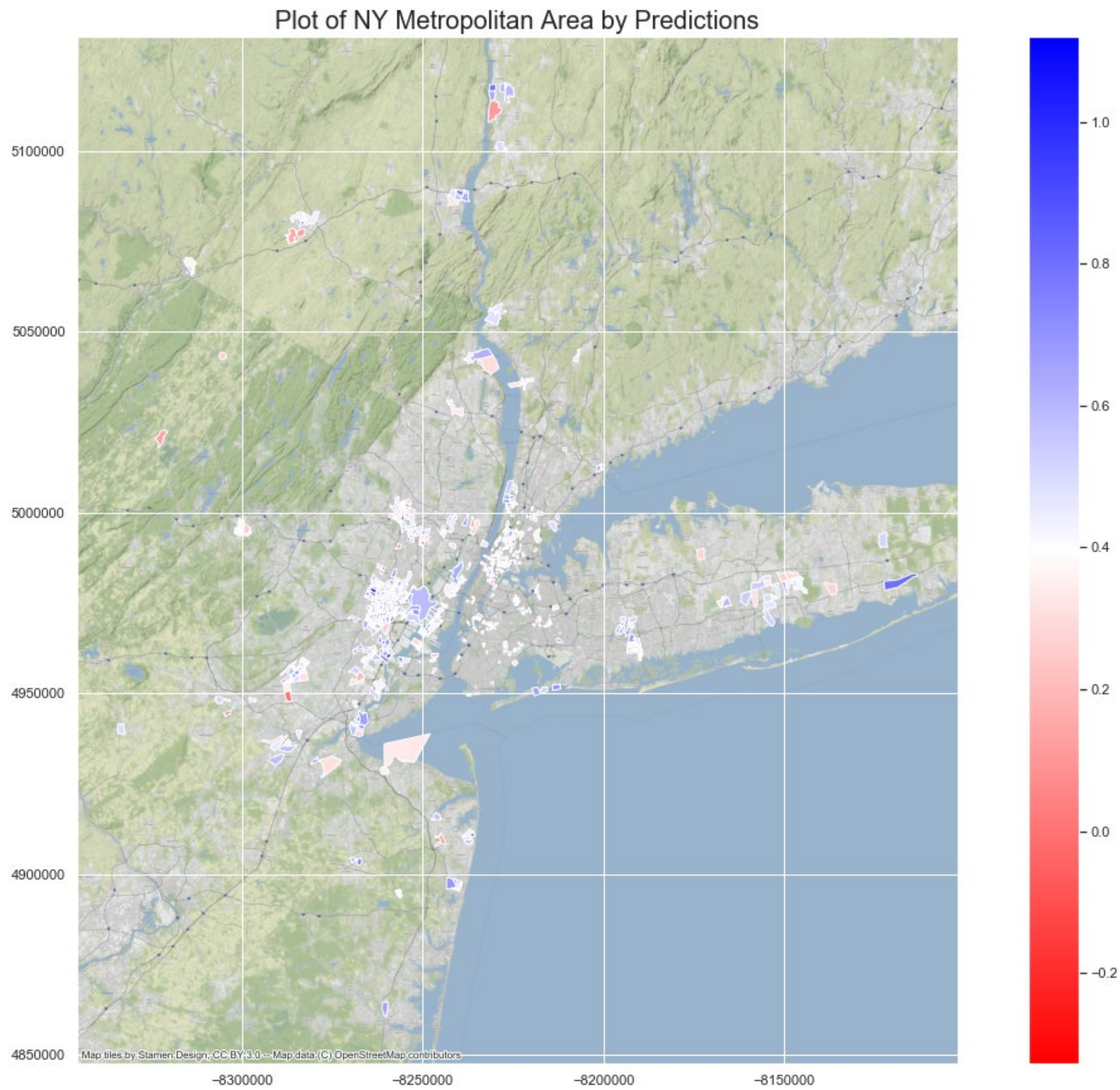
In order to understand how accurate our model is, we rounded any value above 0.5 to 1 (Investment Without Displacement) and any value below 0.5 to 0 (Advanced/Ongoing) Gentrification. We then divided our dataset in training and testing data and obtained an accuracy rating of 0.6953 for training data and of 0.59375 for testing data, as well as the following confusion matrix.



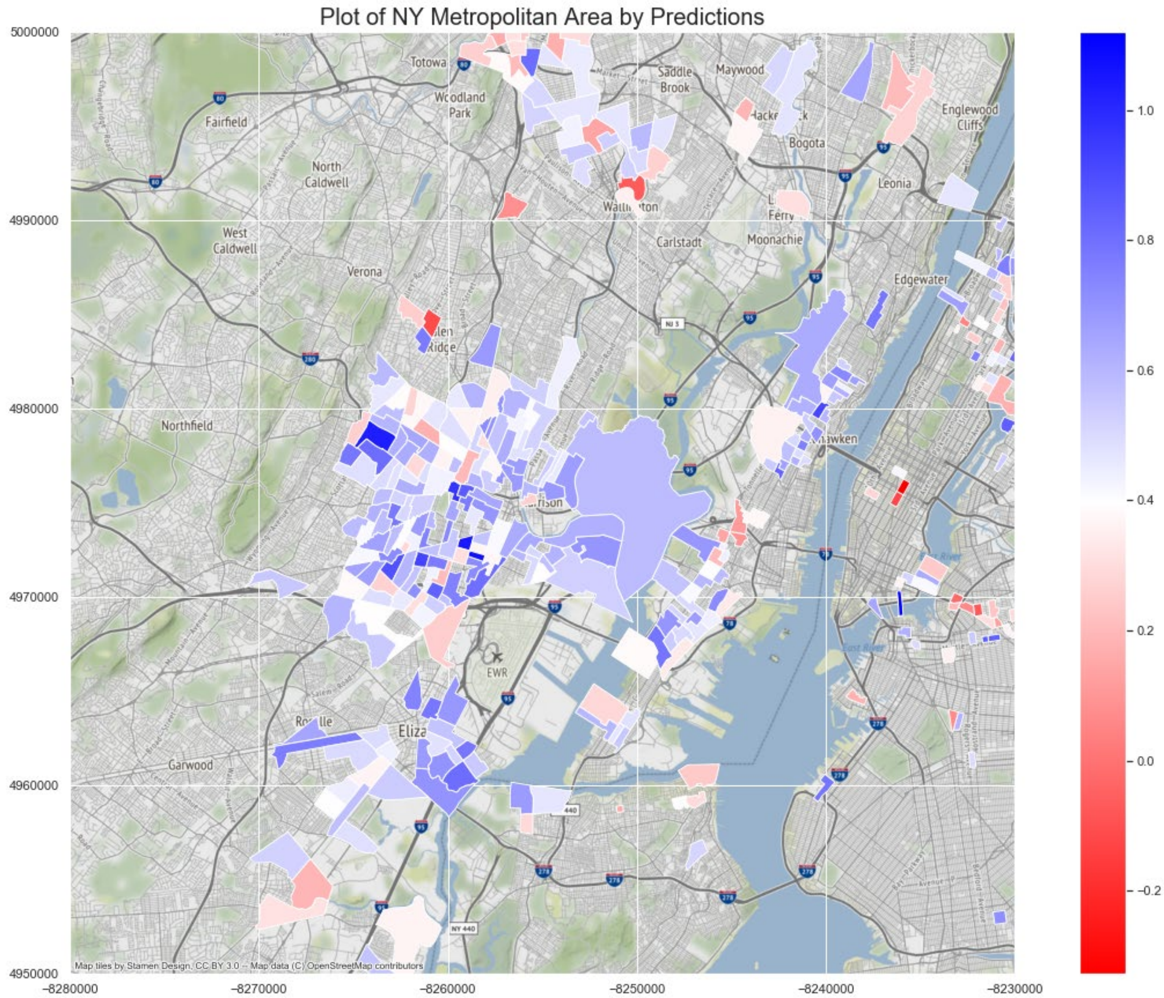
The complete summary of the results of the model can be found in the section 5.5 in the appendix. Some of the highlights are that it has an R-squared score of 0.166 and that for many of the industries the p value is much higher than 0.05 which means that those industries are not great predictors of the type. With this consideration in mind, we can look at the coefficient of each of the industry types to understand how positive or negative they are.

One obvious assumption that needs to be highlighted at this stage is that our model assumes that the industries that were good predictors in 2009-2018 will remain a good predictor in the following decade. Of course, this is not necessarily true as it could be influenced by many external factors out of our control (new industry regulations, major geopolitical events, recessions etc).

The studied tracts by prediction are mapped on the geographical area of the NY metropolitan area in the following plot, where the blue tracts are the ones we predict will fit into the IWD definition and vice-versa.



Zooming-in into the denser area:



Our classification and consequent prediction aims to help government authorities to identify and protect tracts that would suffer from gentrification and enable investors to pick out tracts that would be likely to adapt well to investment and would have lower probability of demographic displacement with the negative ramifications.

Needless to say these need not mean such tracts are good zones for overall investment as they are vulnerable to begin with. It would be the role of investors to predict the tracts that will become hot markets and are therefore great opportunities for investment.

4 Conclusion and Discussion

According to our investigation, the distribution of the industry where low income workers work seems to be a key factor in determining the fate that a neighborhood will suffer upon an increase of its wealth due to the influx of richer individuals moving in.

On one hand, *Construction* and *Arts, Entertainment, etc.* as well *Finance and insurance, etc.* industries strongly resist displacement upon first signs of gentrification, being the most stable when investment is injected into the tract. On the other hand, *Manufacturing*, and *Retail Trade*, are highly liable to gentrification.

Therefore, in order to mitigate the effects of gentrification in New York, there are multiple avenues concerning industry that could be explored. The direct way would be to encourage investors (public or private) to target areas which are not at as much risk whilst discouraging zones that are likely to suffer gentrification given their industry.

Another long-term alternative would be to incentivise industry transitions through public investment or otherwise to gentrification-resilient industries such as *Arts, Entertainment, etc*.

As for the limitations of our studies, we inserted many assumptions about the classification method that may not fully capture the effects of gentrification which still remain a topic of discussion in the social sciences.

Another concern is the small dataset that was used for our prediction model and the fact that it only contained information from one decade which may lead to our model not extrapolating accurately to later decades.

We believe that with a bigger sample size (more cities that are similar to New York), taking more variables into consideration and access to more decades we would be able to increase our accuracy and perhaps use a more sophisticated machine learning model.

5 Appendix

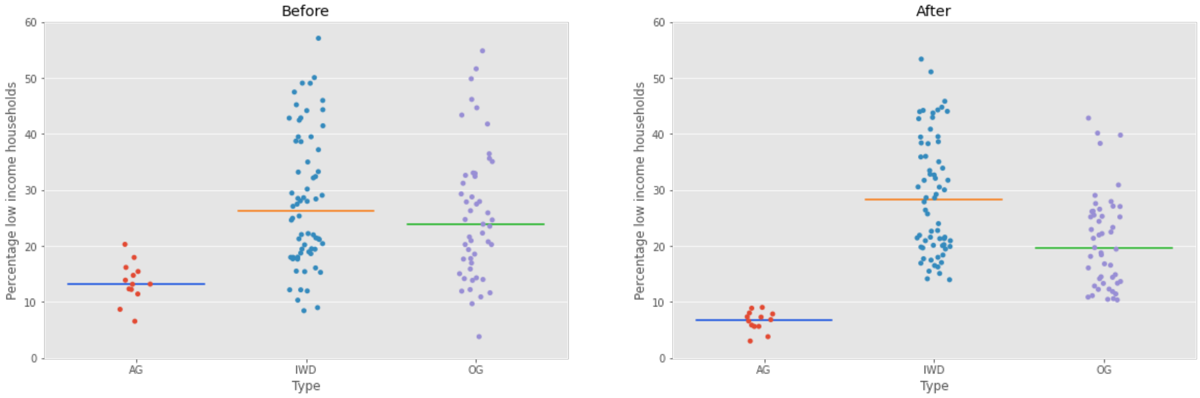
5.1 Full Typology Classifications

Typology	Typology Criteria
Ongoing Gentrification (OG)	A tract that is: - a low income tract in the end year, and - gentrified from the base year to the end year.
At Risk of Gentrification (at the end year)	A tract that is: - a low income and vulnerable tract in the end year, - a hot market from base year to end year, and - not currently undergoing displacement or OG.
Not losing Low-Income Households	A tract that is: - a low income tract in the end year, and - not classified as OG or at risk of gentrification.
Ongoing Displacement of Low-Income Households	A tract that: - is a low income tract in the end year, and - displacement happened from base year to end year.
Advanced Gentrification (AG)	A tract that is: - a moderate to high income tract in the end year, and - gentrified in from the base year to the end year.
Invested without displacement (IWD)	A tract that is: - vulnerable in the base year, - not vulnerable in the end year, - a hot market from base to end year, - no demographic change, and not losing low-income households from base year to end year.
Ongoing Exclusion	A tract that: - is a moderate-to-high-income tract in end year, - occurred absolute loss of low-income households from base year to end year, and - the percentage of low-income households in the end year is lower than the base year.
Stable Exclusion	A tract that is: - a moderate-to-high-income tract in end year, and - not classified as Ongoing Exclusion
Super Gentrification or Exclusion	A tract that: - median household income higher than 200% of

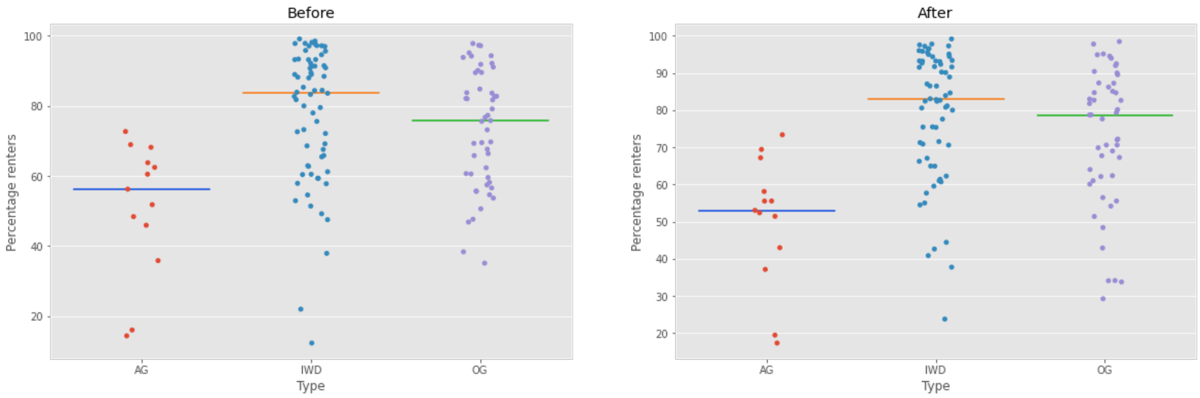
	regional median in the end year, and - shows indicators of Gentrification or Exclusion.

5.2 Percentage changes of four main social factors in AG, OG and IWD

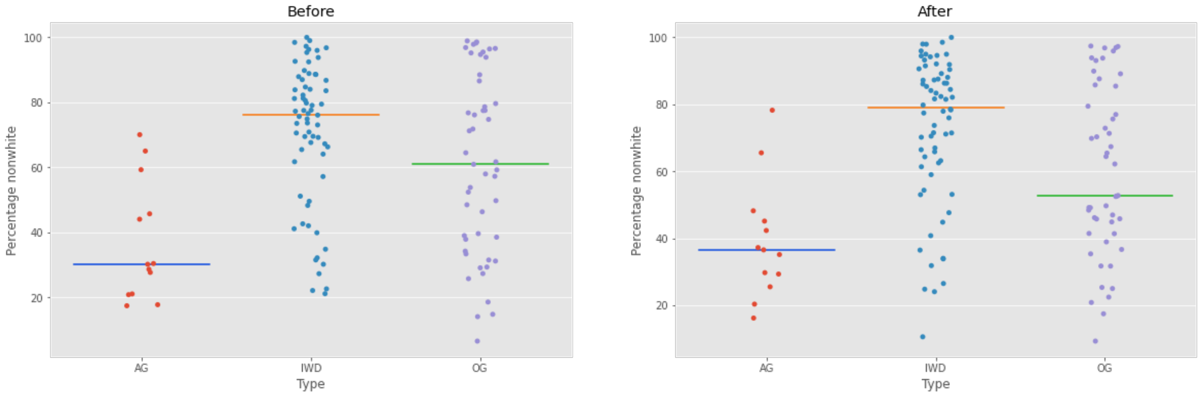
Here we present the precentage change from base year to the end year, in four different population factors: ‘Percentage of Low Income Households’, ‘Percentage of Renters’, ‘Percentage of Non-white population’, ‘Percentage of college educated’ in the AG, OG and IWD tracts. Here ‘before’ stands for the base year, and ‘after’ stands for the end year.



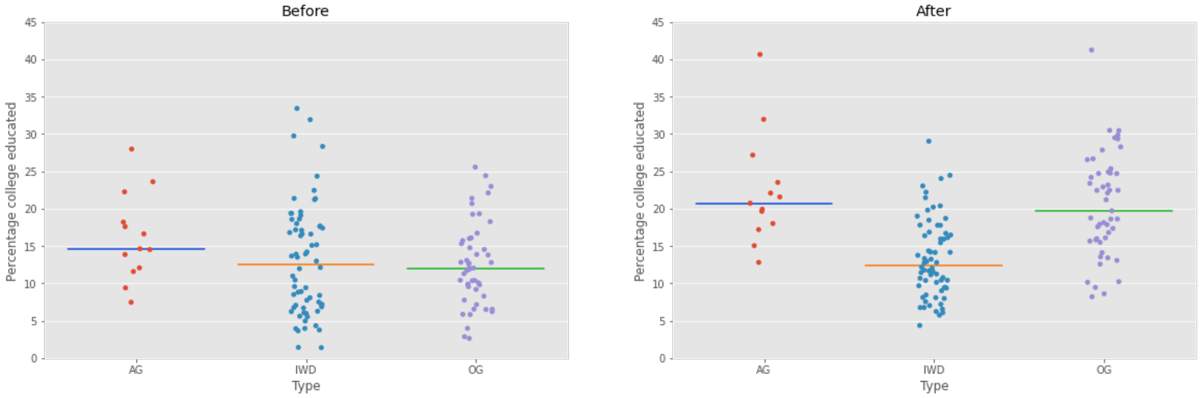
5.2.1 percentage change of low-income households



5.2.2 percentage change of renters

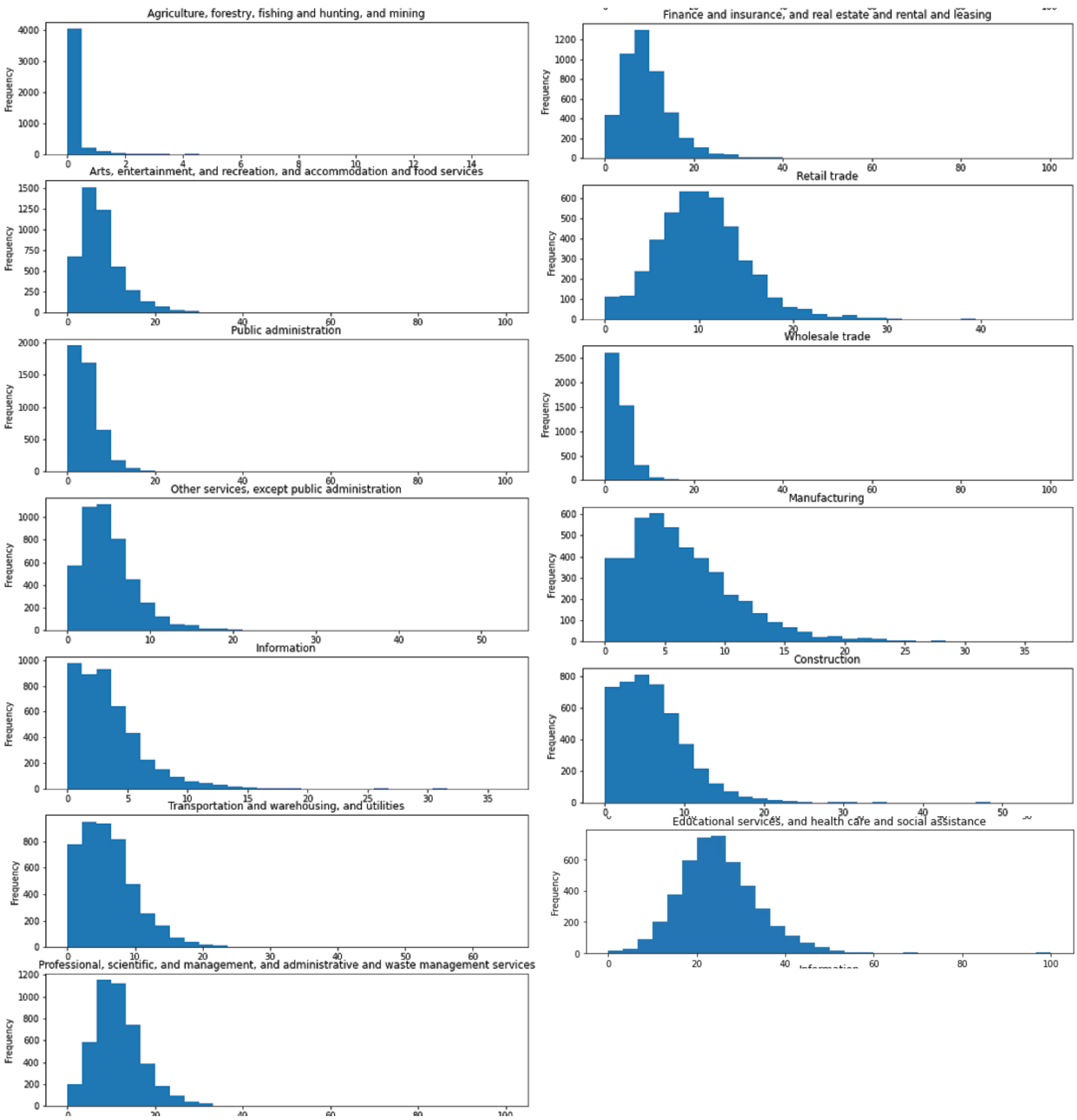


5.2.3 percentage change of non-white population

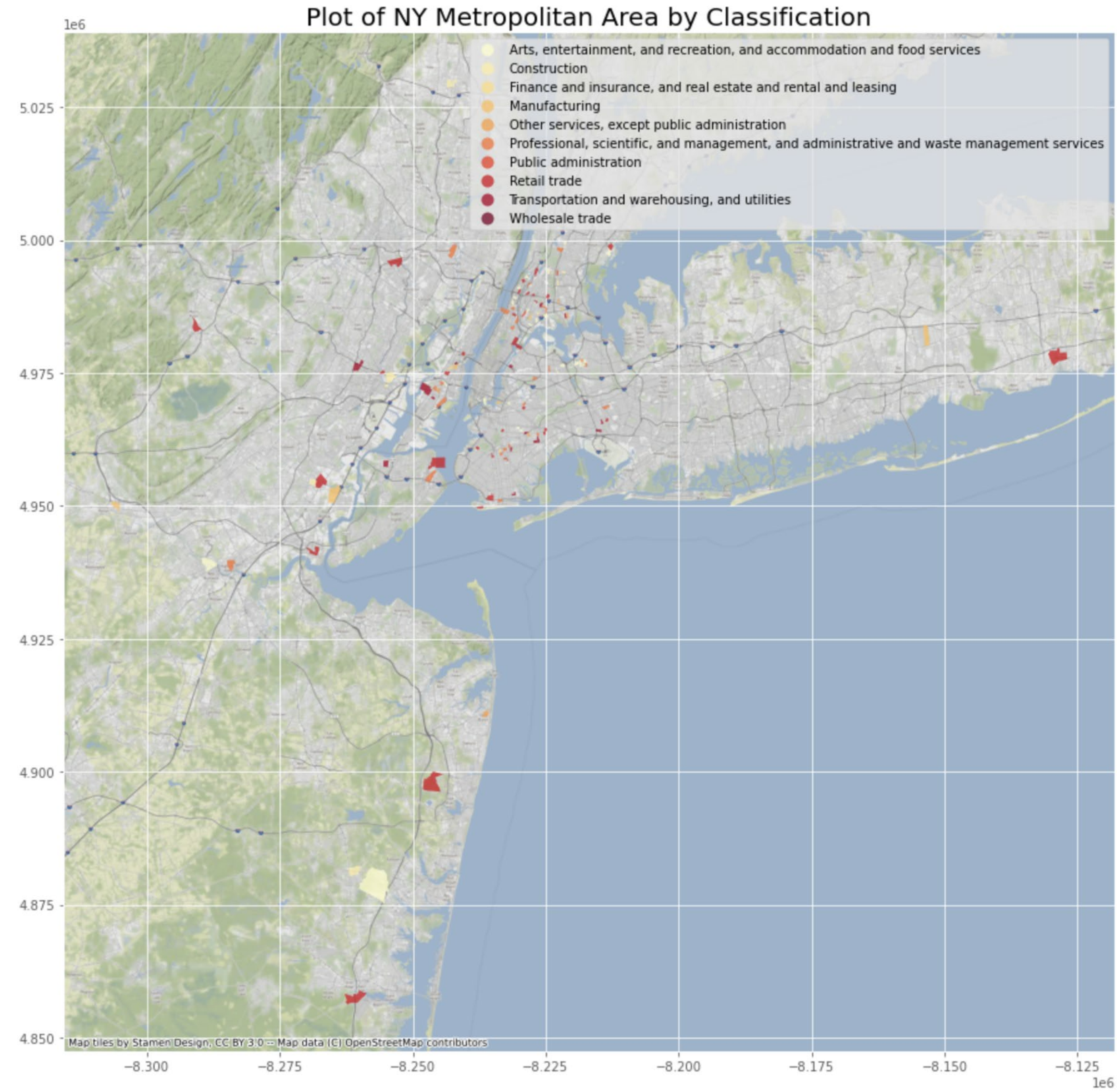


5.2.4 percentage change of college-educated population

5.3 Industry percentage distribution by industry



5.4 Plot of the OG, AG and IWD tracts highlighted according to the main industry



5.5 Summary of the OLS Regression for Gentrification Type based on industry distribution.

OLS Regression Results

Dep. Variable:	Type	**R-squared:**	0.166
Model:	OLS	**Adj. R-squared:**	0.079
Method:	Least Squares	**F-statistic:**	1.903
Date:	Sat, 24 Oct 2020	**Prob (F-statistic):**	0.0408
Time:	16:58:01	**Log-Likelihood:**	-81.246
No. Observations:	128	**AIC:**	188.5
Df Residuals:	115	**BIC:**	225.6
Df Model:	12		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	0.0003	0.001	0.393	0.695	-0.001	0.002
Agriculture, forestry, fishing and hunting, and mining	-0.0310	0.061	-0.507	0.613	-0.152	0.090
Arts, entertainment, and recreation, and accommodation and food services	0.0244	0.007	3.327	0.001	0.010	0.039
Public administration	0.0158	0.014	1.148	0.253	-0.011	0.043
Other services, except public administration	-0.0006	0.011	-0.060	0.952	-0.022	0.020
Information	-0.0296	0.020	-1.508	0.134	-0.068	0.009
Educational services, and health care and social assistance	0.0048	0.010	0.489	0.626	-0.015	0.024
Transportation and warehousing, and utilities	-0.0047	0.009	-0.496	0.621	-0.023	0.014
Professional, scientific, and management, and administrative and waste management services	0.0138	0.011	1.219	0.225	-0.009	0.036
Finance and insurance, and real estate and rental and leasing	0.0089	0.004	2.238	0.027	0.001	0.017
Retail trade	-0.0151	0.009	-1.634	0.105	-0.033	0.003
Wholesale trade	0.0351	0.018	1.913	0.058	-0.001	0.071
Manufacturing	-0.0067	0.010	-0.671	0.504	-0.027	0.013
Construction	0.0107	0.009	1.210	0.229	-0.007	0.028

Omnibus:	197.854	**Durbin-Watson:**	0.276
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	11.273
Skew:	-0.149	**Prob(JB):**	0.00357
Kurtosis:	1.577	**Cond. No.**	1.36e+17