NYU | Center for Urban Science + Progress

Exploring Gentrification and Displacement Through User-Generated Data

Kent Pan | Tiffany Patafio | Manrique Vargas
Jiawen Wan | Gabriel Yin

Agenda

- **Motivation**
- **Data & Methodology**
- **Modelling & Results**
- **Findings & Conclusions**

Why explore gentrification?

How can we know when an area is gentrifying?

What exactly is gentrification?

Urban renewal

Increased local economic development, employment

More investment in local infrastructure & services

Increased rental & home costs

Displacement of long-term residents

Decline in racial diversity & equity

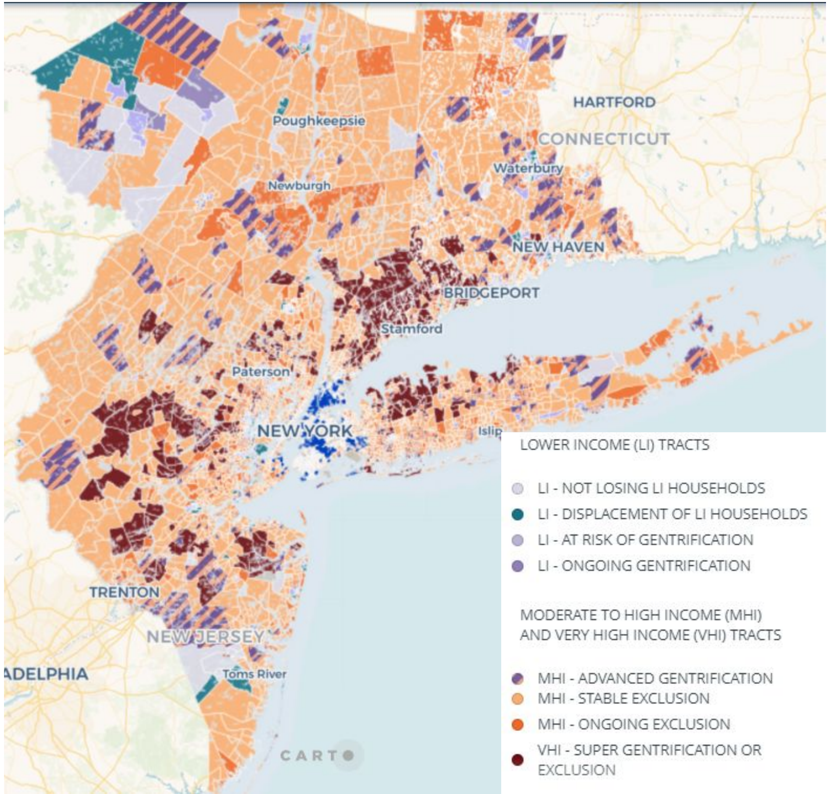
Cultural changes

Less community cohesion

Urban Displacement Project (UDP)

- Typologies using *census data* from 1990-2016
- Early warning tool - where is gentrification happening?
- Are there ways we can *better* understand these phenomena?

URBAN DISPLACEMENT PROJECT
UNIVERSITY OF CALIFORNIA BERKELEY



Problem Statement

Can near real-time, “user-generated” sources of geographic data from online platforms be used to understand gentrification?

Can online data sources capture another “layer” of human activity that is missed by traditional sources?

Can we leverage these non-traditional sources to better understand how human activity relates to gentrification and displacement?

Literature Review

Census data has been used in many studies of gentrification from prediction of neighborhood change (Reades, 2018) to real estate.

Steif (2016) used historical census data to predict 2011 home prices and in a separate study found house price change regressions to have the highest explanatory power.

Twitter data, for example, was used to understand mobility and visitor patterns across different types of neighborhoods in San Francisco. (Chapple, 2018)

Cerron et al (2018) used Foursquare data to study activities in urban spaces and, particularly, how they change across neighborhoods.

Data & Methodology

Data Sources

United States Census Bureau

- 1990-2016
- US wide
- Original UDP features
- Additional Census & ACS features
- Household Income; age, race and demographics; unemployment density

twitter

- 2012-2015
- Entire study region
- Raw tweet metadata aggregated to census tract
- # of resident vs visitor tweets, time series of tweet counts throughout day, etc.

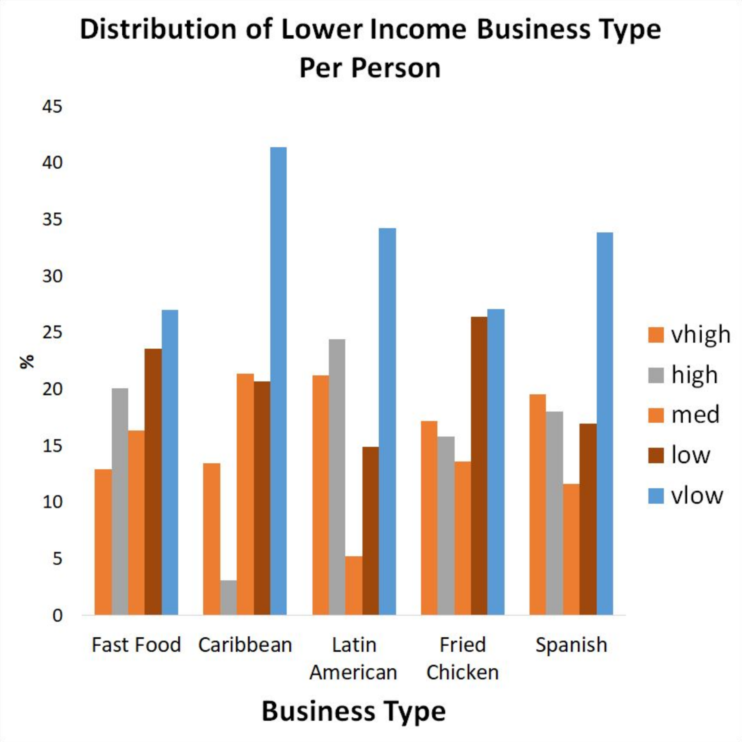
FOURSQUARE

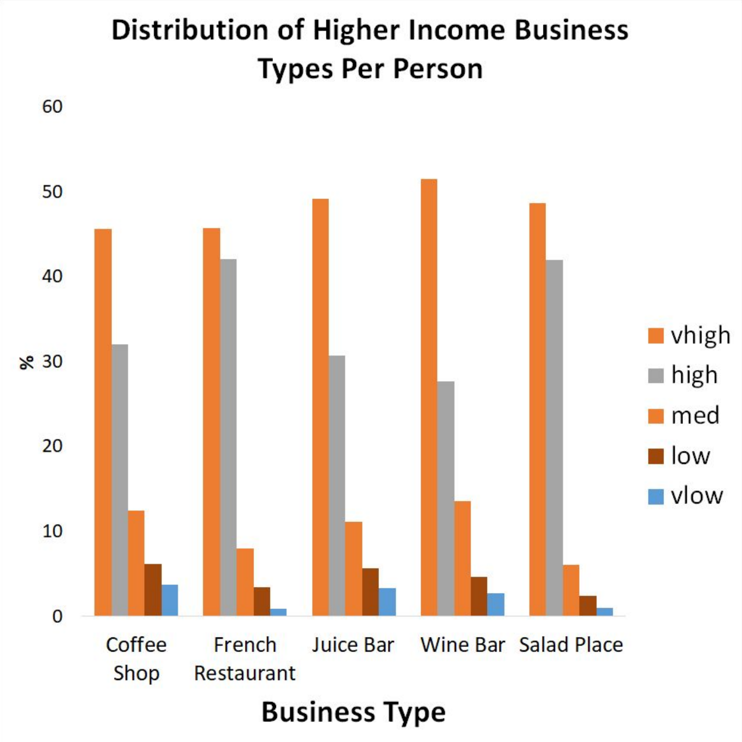
- 2009-2018
- NYC area only
- Sampled dataset (30%)
- Point data on individual businesses
- # of Check-ins, Class/Category of business, etc.

NYC OpenData

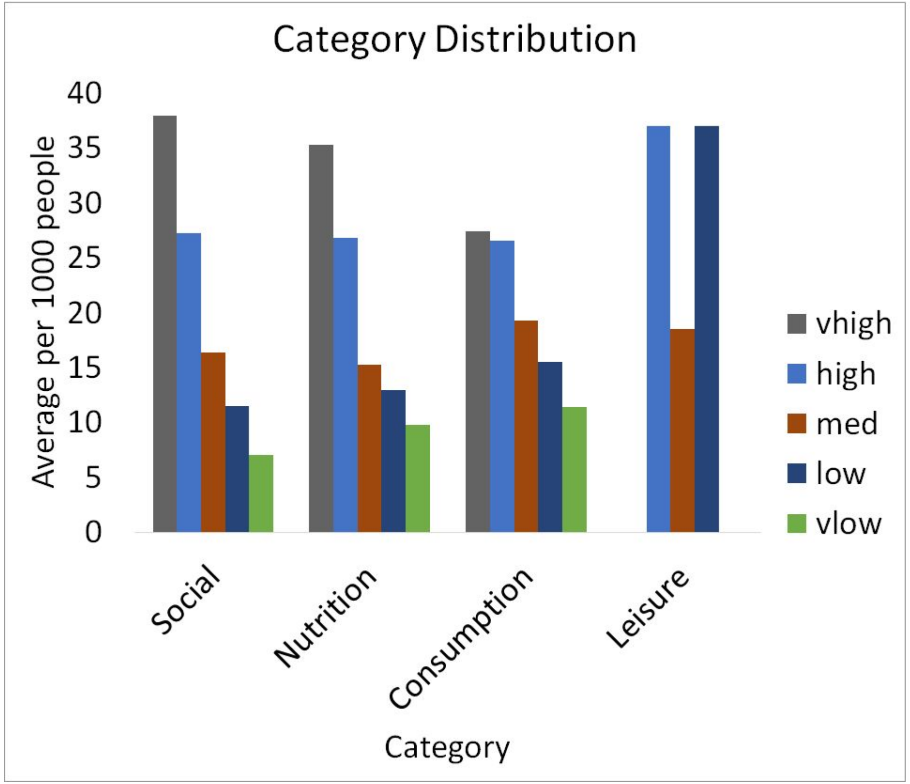
Zillow

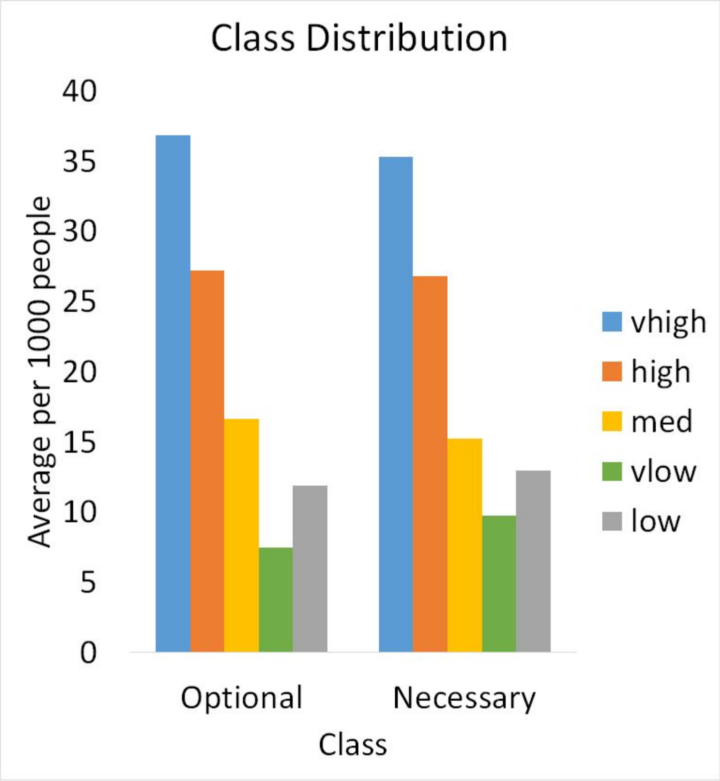
Data Exploration: Foursquare





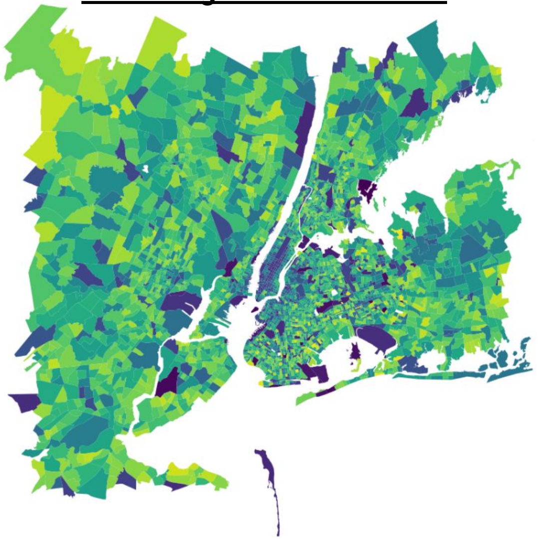
Data Exploration: Foursquare





Data Exploration: Twitter

Percentage of Local Tweets



Trends in Visitor and Local tweets:

- Manhattan averages about 20% local tweets
- Further from the city tracts reach almost 97%

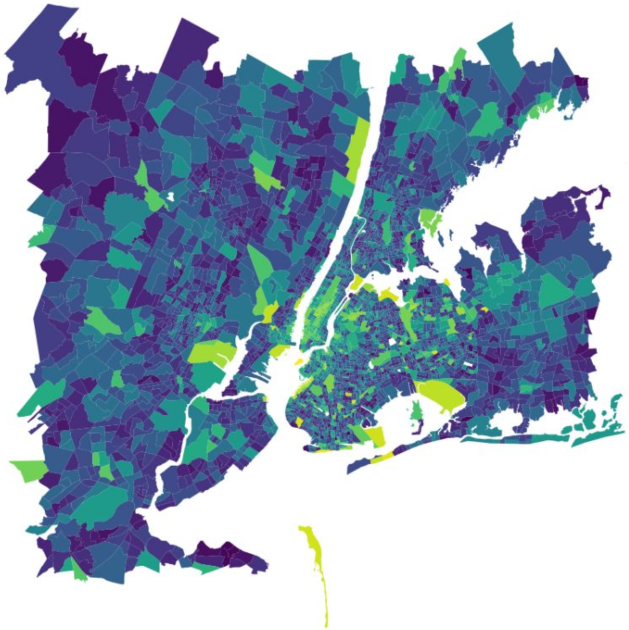
Creating neighborhood profiles:

- Dominated by Local (0, 2)
- High visitor, low local (1)
- High neighbor, low visitor (4)

Methodology: Twitter Feature Engineering

- Twitter Activity Patterns
  - Local, Visitor, Neighbor
  - Weekend vs Weekday
  - Day vs Night
- Understanding Visitor Demographics:
  - % of visitors from High, Medium, Low income census tracts
  - % of visitors from High, Medium, Low average percent of college education tracts
- Twitter values over Time
  - Annual number of Tweets (2012-2015)

Percentage of Visitor Tweets

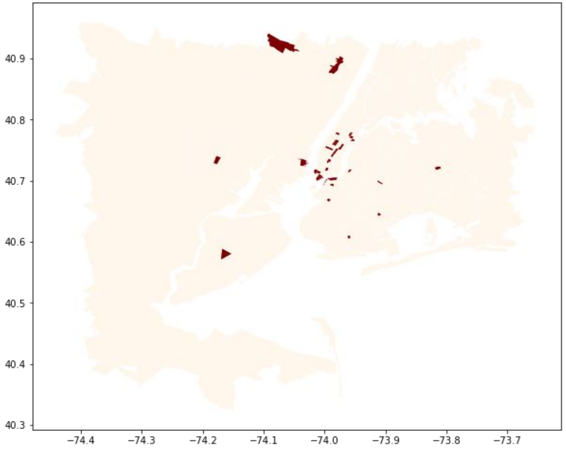


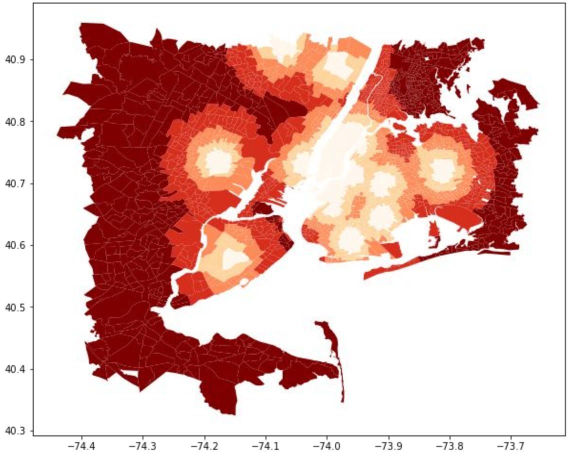
Methodology: Spatial Feature Engineering

Distance Calculations

With location such an important factor in our dataset, we looked at distances to local hotspots identified during exploration

E.g. Distance to Juice Bars, High Visitor Locations, Medium/High Income tracts

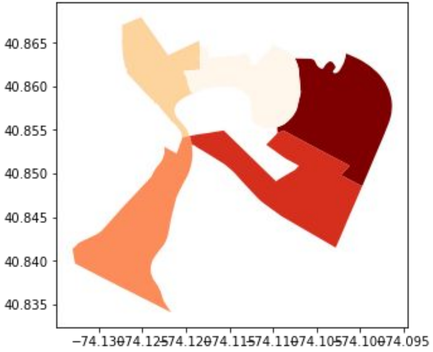


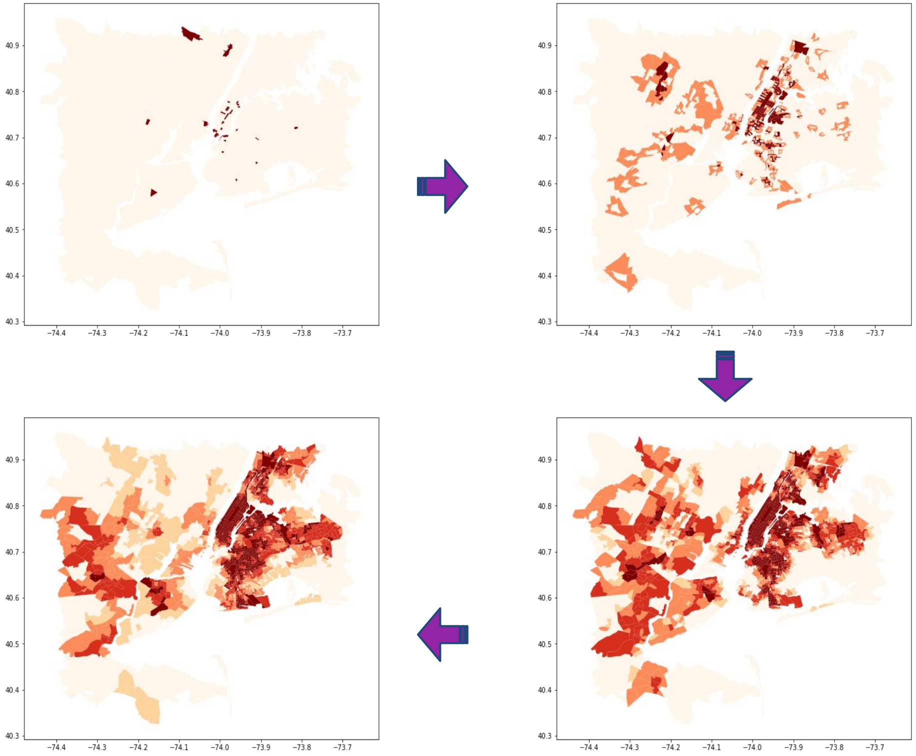


Methodology: Foursquare Feature Engineering

Nearest Neighbor Analysis

Use Spatial Weights to learn from the neighbouring Census Tracts.
e.g Sum of neighbouring Juice Bars



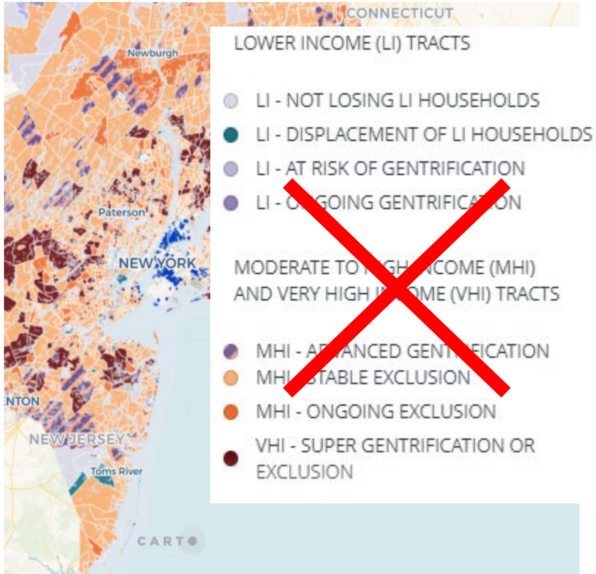


Methodology: Typologies

FOURSQUARE

twitter

Zillow

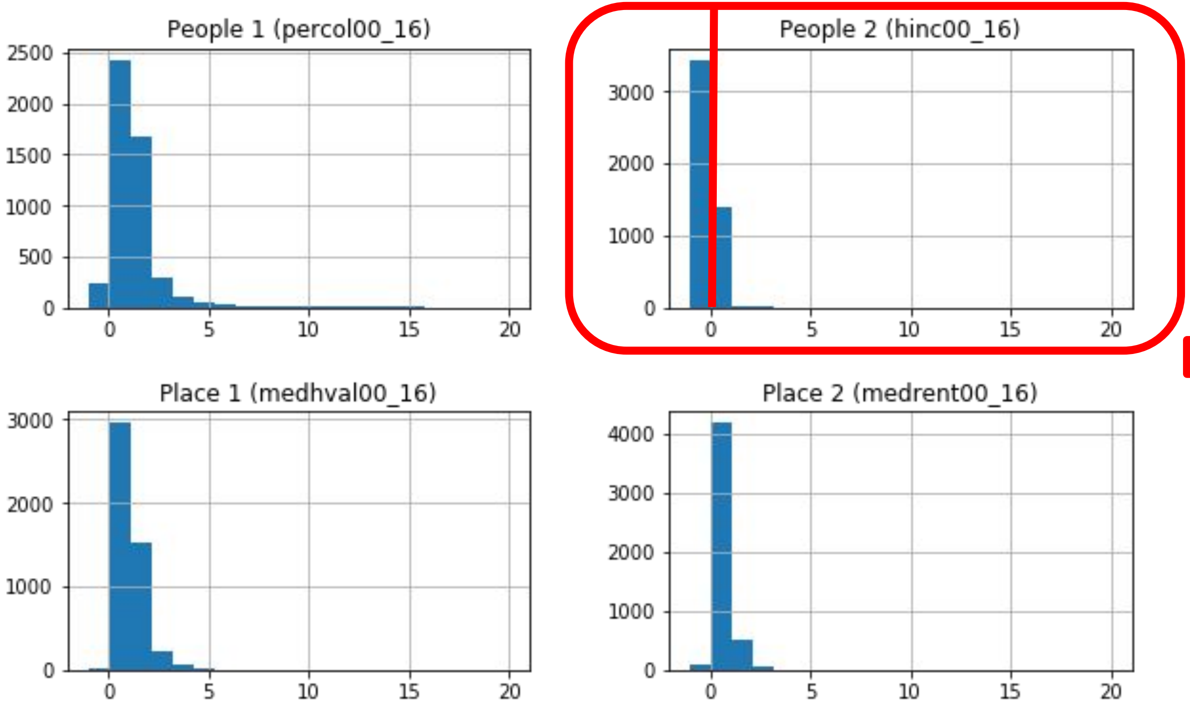


URBAN DISPLACEMENT PROJECT

Methodology: Typologies

People	*New high-income, higher-education populations*	*2000-2016* • % change in college education population • % change in household income
Place	*Land & housing markets - real estate investment and development*	*2000-2016* • % change in median home value • % change in median rental cost
Hybrid	*Combines elements from the two separate ideas (UDP)*	• Ongoing and advanced gentrification (1990-2016) • Ongoing and advanced gentrification (2000-2016) • “Super gentrification”

Methodology: Typologies

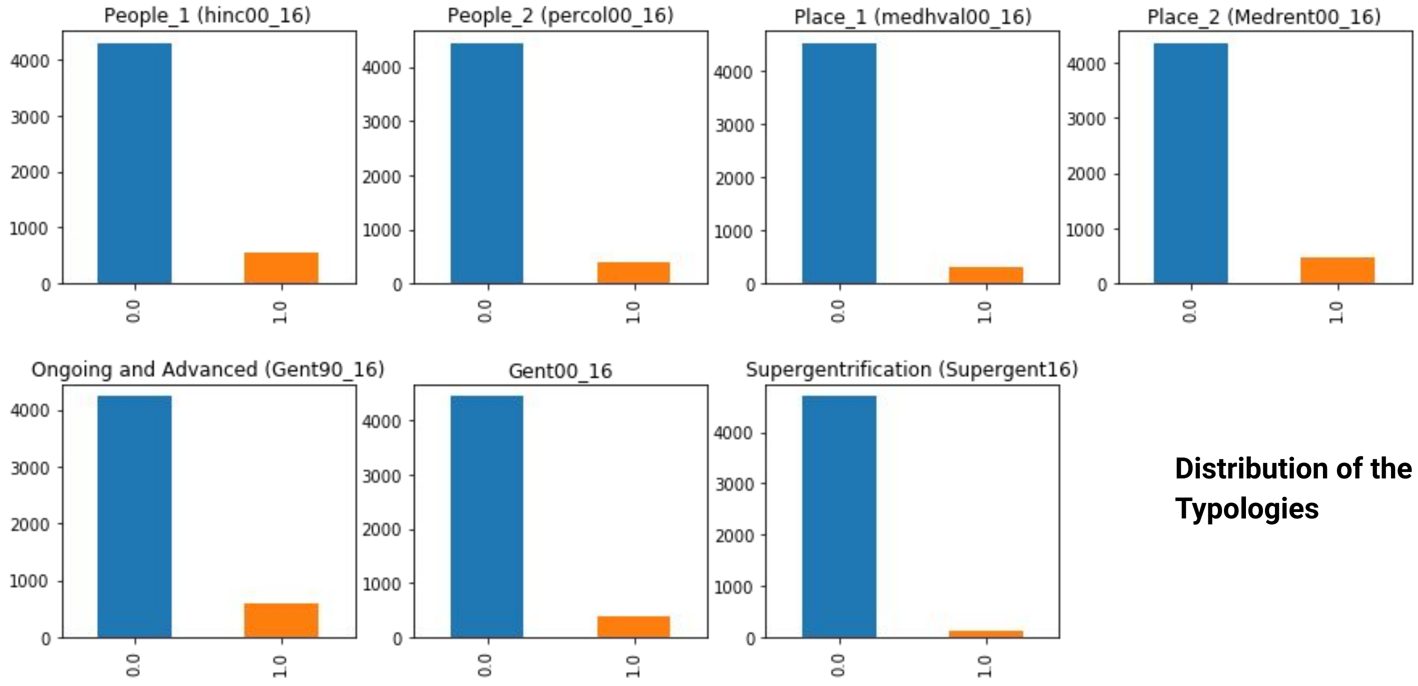


- People 1 (percol00_16)
  - Mean: 118%
  - Stdev: 118%
- People 2 (hinc00_16)
  - **Mean: - 1.8%**
  - Stdev: 23.2%

Real mean income has decreased 1.8% during 16 years, while rent and median housing value has increased 95%

- Place 1 (medhval00_16)
  - Mean: 117%
  - Stdev: 109%
- Place 2 (medrent00_16)
  - Mean: 74%
  - Stdev: 42%

Methodology: Typologies



Distribution of the 7 Target Binary Typologies

Methodology: Modelling

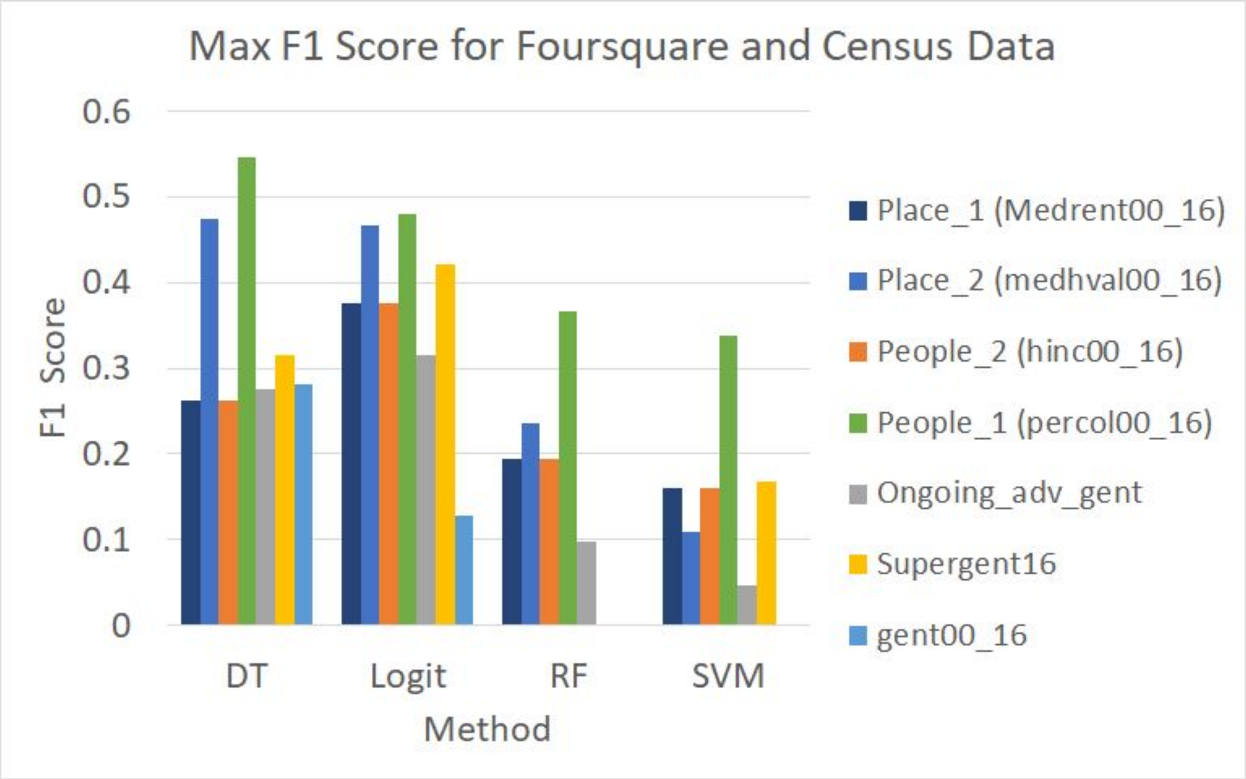
Each of the 7 binary variables was modelled for four feature sets using the following model types:

- Random Forest
- Decision Tree
- Logistic Classification
- Support Vector Machines

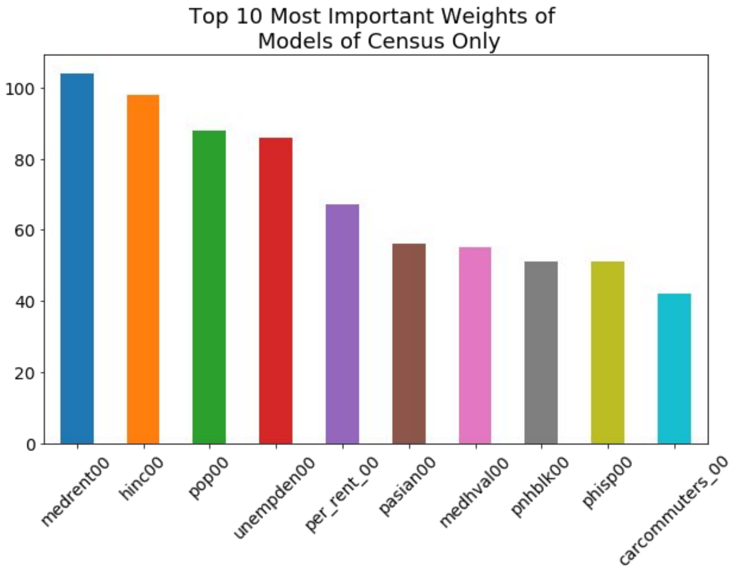
Model Group	Target Variables
Census data only	7 binary markers of gentrification
Census data + Foursquare	
Census data + Twitter	
Combined Model (Census, Foursquare, and Twitter)	

Modelling & Results

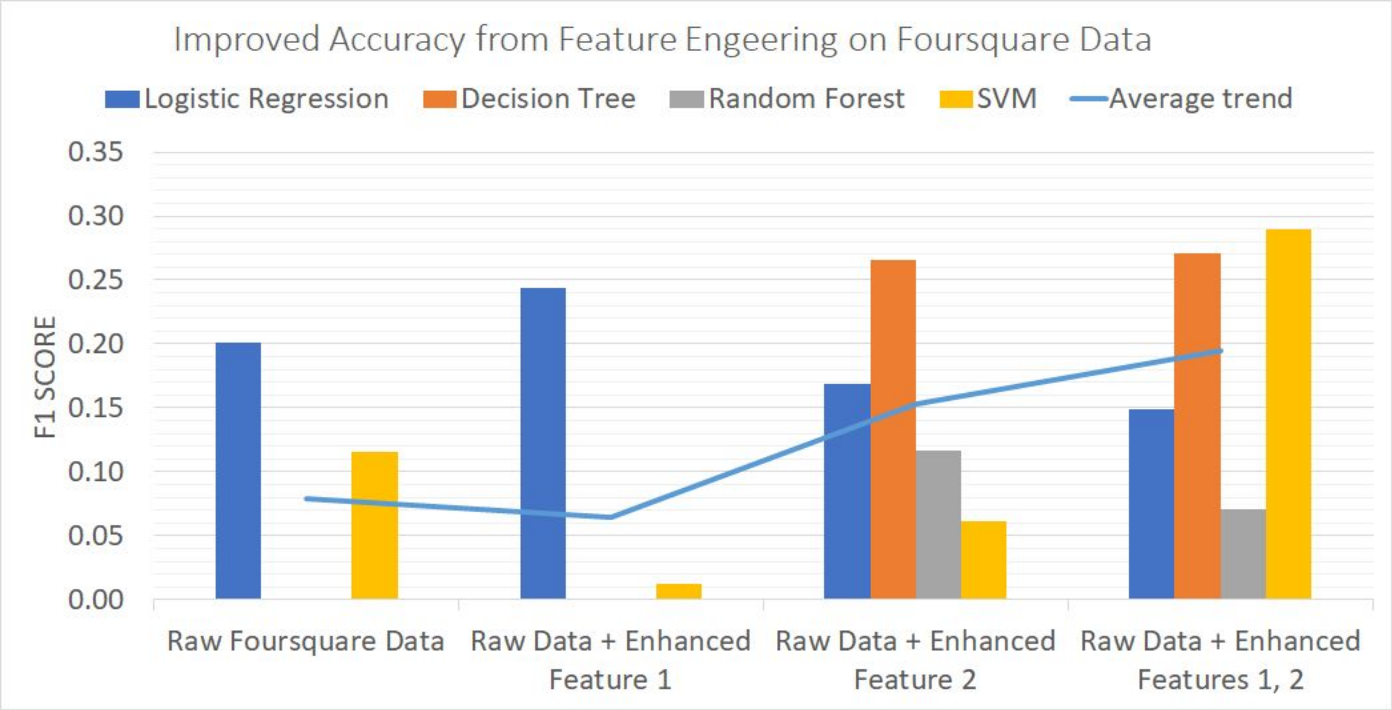
Results: Census Findings



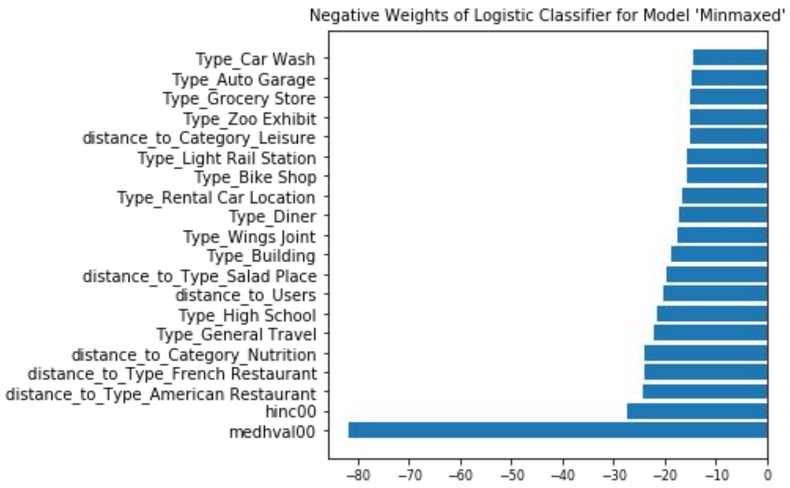
Results: Census Findings

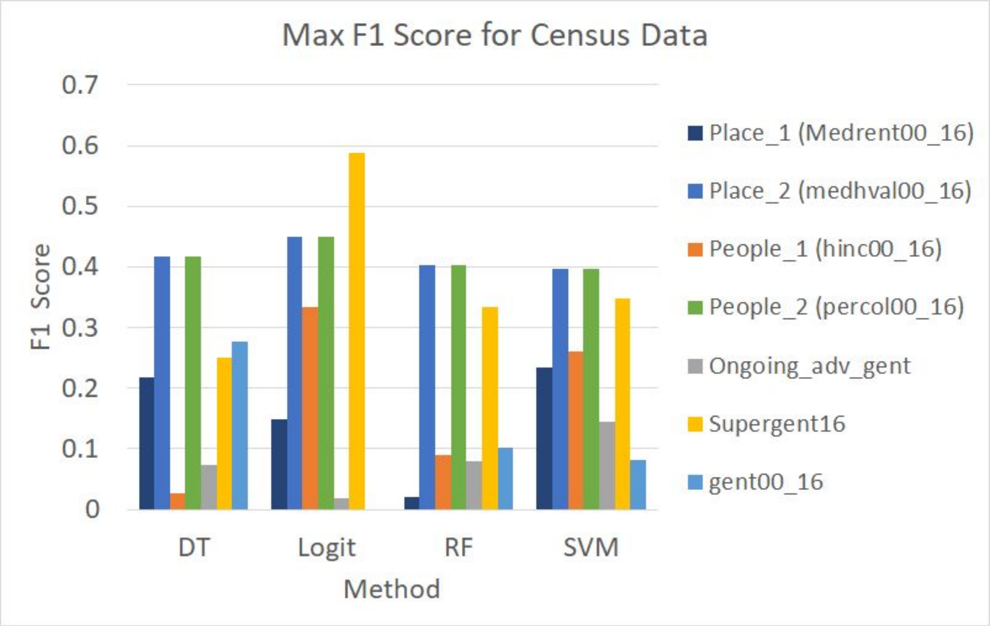


Results: Foursquare Findings

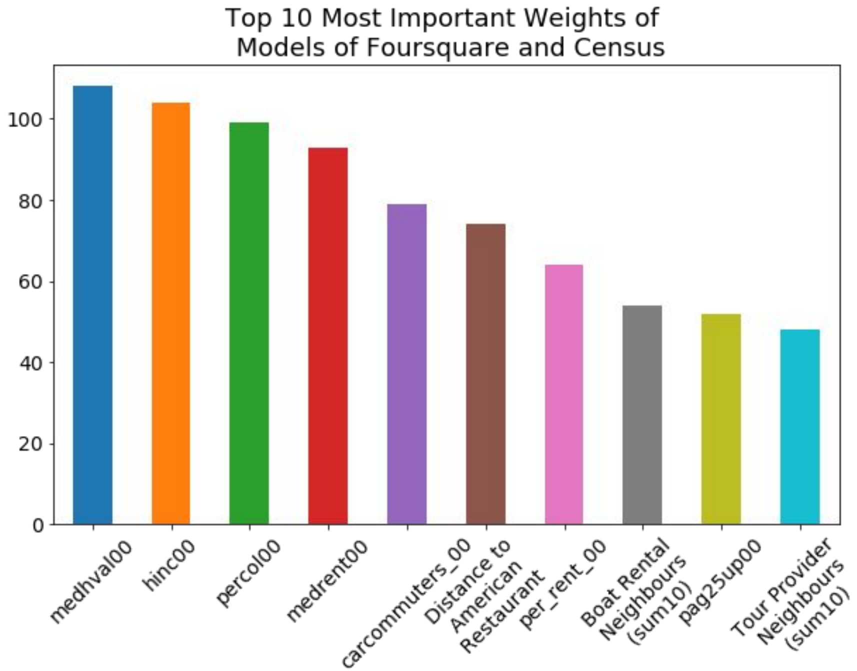


Results: Foursquare Findings





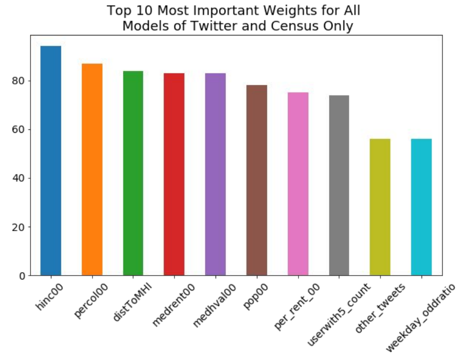
Results: Foursquare Findings

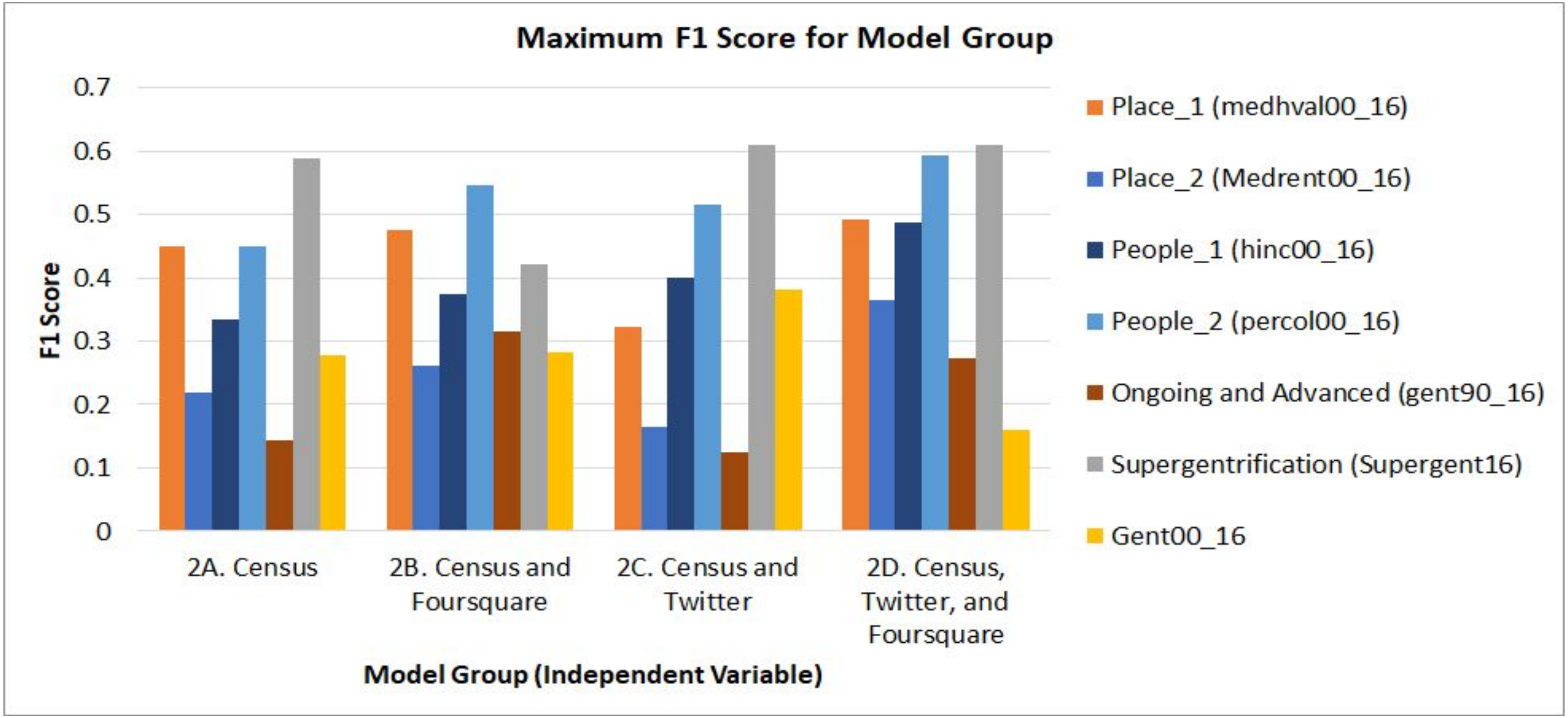


Results: Twitter Findings

- Twitter features consistently outperformed Census data alone
  - Supergentrification outperformed all other binary variables across feature sets
  - % Change in Household income improved most with the inclusion of Twitter data and % Change Median rent benefited most from our derived Twitter features
- Overall, Twitter data proved most helpful in improving in modelling people gentrification metrics
  - The single most important feature in predicting the % Change in Income typology was % of High Educated Visitors
- Important features included those around visitor tweets, weekend/weekday tweet volumes

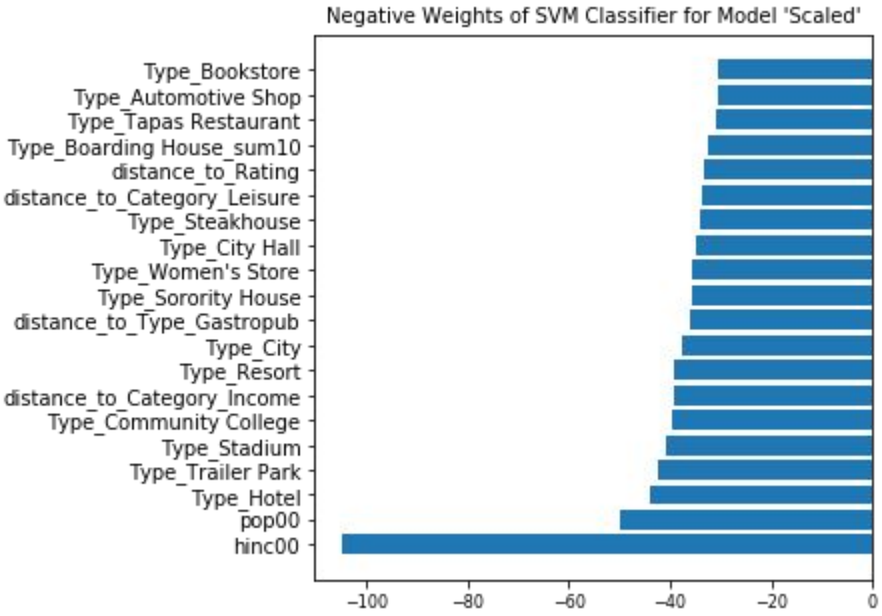
	Featureset				
	Census Only	Census + Twitter Raw	All + Distance Features	All + Derived Twitter Features	Grand Total
Gentrification (90-16)	0.0000	0.1439	0.1878	0.0000	0.0829
% Change College Educated	0.2857	0.3008	0.3129	0.5213	0.3552
% Change Housenhold Income	0.0124	0.2816	0.1961	0.4592	0.2373
% Change Med Home Value	0.2883	0.4154	0.2178	0.2105	0.2830
% Change Med Rent	0.0000	0.0233	0.0400	0.2901	0.0883
Gentrification (00-16)	0.3883	0.0000	0.3802	0.0000	0.1921
Supergentrification (00-16)	0.6341	0.6111	0.7143	0.7027	0.6656
Grand Total	0.2298	0.2537	0.2927	0.3120	0.2721

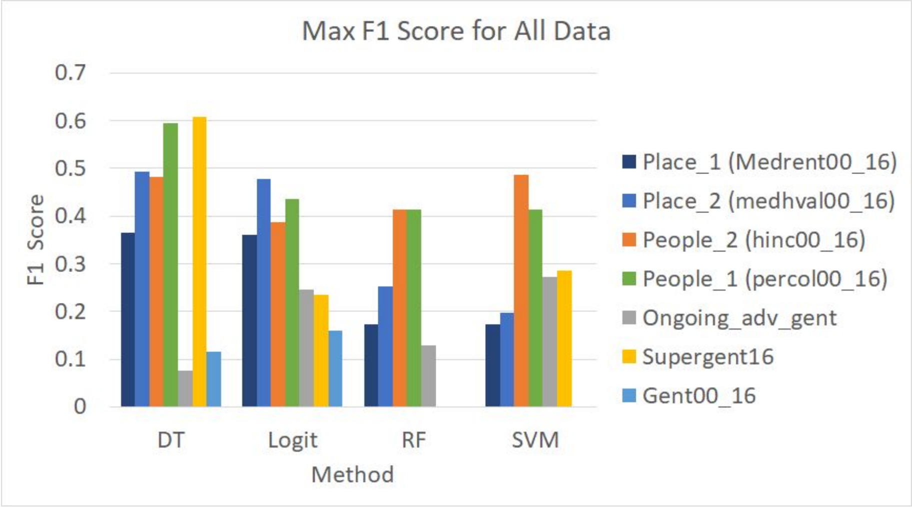




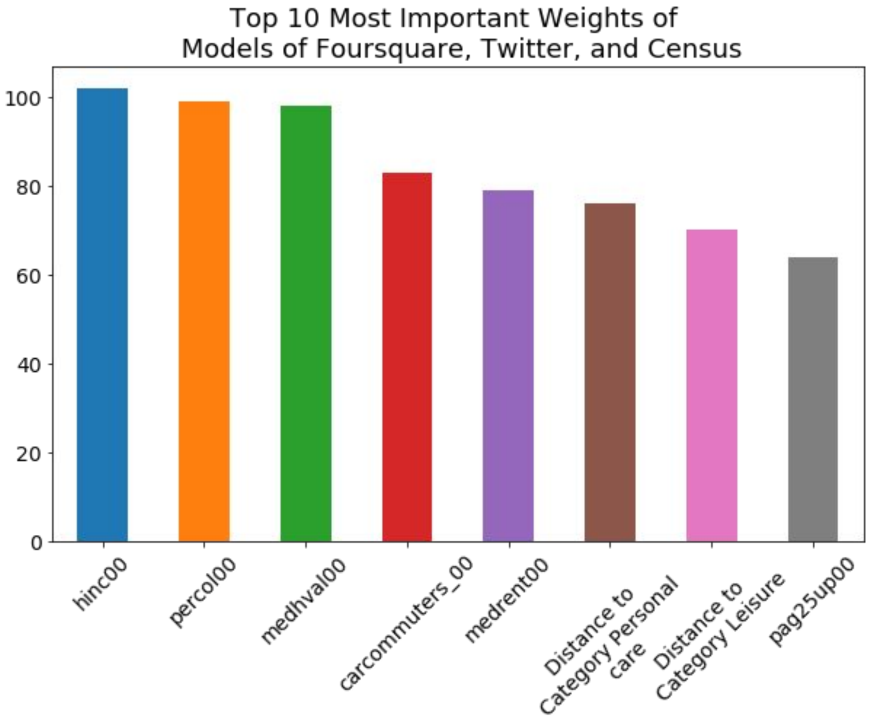
The F1 scores above represent the scores of our best performing models for each feature set and target variable combination.

Results: Combined Model Findings





Results: Foursquare, Census, and Twitter



Findings & Conclusions

Results: Interesting Findings

For Twitter data, visitor demographic features-- like the percentage of highly educated visitors-- were most important in modelling "people gentrification."

For the NY Metro area, rent has increased almost 75% since 2000 and wages have decreased 1%.

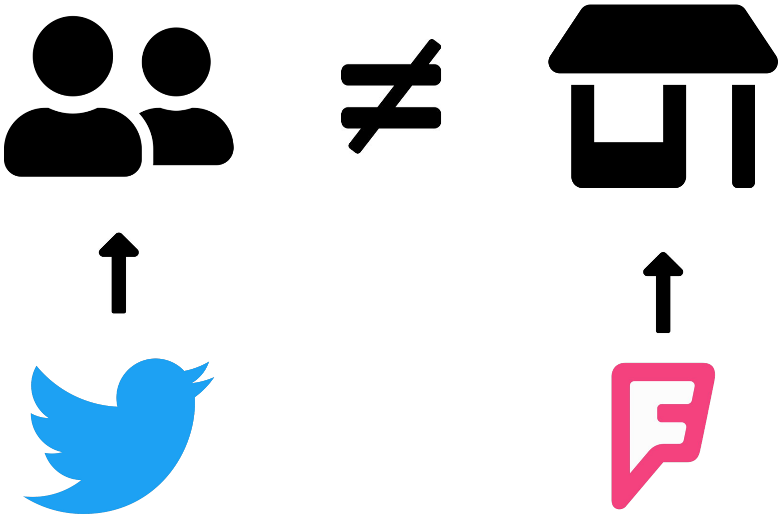
On average, low income neighborhoods have seen a 2% larger increase in rent and a 9% larger decrease in wages.

Improved Foursquare features like distance to boat rentals and to tour providers proved to be two important features.

Our highest Visitor Tweet Locations were: JFK, Penn Station/MSG, Times Square, and Central Park.

Conclusions

- More specific target metrics gave us better insight
- People and place gentrification behaved differently in our models
- As more data accumulates, we believe the social media data assessed will be a powerful source of insight for those studying gentrification



Where to learn more?

Project Website:
https://ace-gabriel.github.io/twitter_gentrification/

Sponsor Website:
https://www.urbandisplacement.org/about/dr-karen-chapple

Github:
https://github.com/mv1742/UDPNY
https://github.com/patafiot/Gentrification-Capstone

Tableau Exploratory Analysis:
https://public.tableau.com/views/TwitterDeepDive/Aboutthisworkbook?:embed=y&:display_count=yes&:origin=viz_share_link
https://public.tableau.com/profile/tiffany.patafio#!/vizhome/FoursquareDeepDive/WhatsintheData

Thank you!

Questions?