Research Question

- How has post-recession Detroit fared?
- Is it a success story, a story of revitalization and renaissance? Or is the revitalization story a bit more complex, and should we be perhaps a bit more ambivalent about the triumphalist account of the city?
- To add a bit more, can we find distinctions between wealthier and poorer neighborhoods and what might those distinctions be? How can they be addressed?
- Specific response variables include mean income, total buildings permitted, crime and homicides, auctions (homes sold at auction), and Improve Detroit Tickets

Data Cleaning Process

- Import and clean datasets
- Organize them all into neighborhood form
- Using the geopandas library, I converted all the latitude and longitude coordinates into geographic points
- I then used a dictionary of neighborhoods in the city of Detroit (207 in total) to locate these specific points in neighborhoods, and included a neighborhood variable in each dataframe.

Datasets Used

- Libraries--the names and coordinates of all public libraries in Detroit.
- Detroit Demolitions --a dataset from January 1, 2014 to the present, which lists the names, sites, and prices of demolished buildings.
- Business Licenses—a dataset of all business licenses issued since 2015. It does not include the type of business.
- Auction Sales—The Detroit Land Bank buys and sells vacant properties to online bidders in an auction. The dataset begins in June 2014.
- Building Permits—issued by the City of Detroit Buildings, Safety, Engineering, and Environmental Department. These range from 2010 to the present, and include new building permits as well as alterations. Building activity primarily reflects a changing neighborhood.
- Annual Inspections—Inspections of commercial buildings by the city of Detroit inspectors. These inspections range from 2015 to the present, and are scheduled for all commercial properties annually.
- Fire Stations—a dataset of all the fire stations and their locations in Detroit and their coordinates.
- Blight—this dataset shows blight violations that have been issued property owners who have violated City of Detroit ordinances that govern how property owners maintain the exterior of their property. The dataset starts in 2004 and goes into the present.
- Childcare –a dataset of all childcare providers in Detroit.
- DDOT Bus stops—bus stops for the buses operated by the Detroit Department of Transportation. The data comes from August 2016.

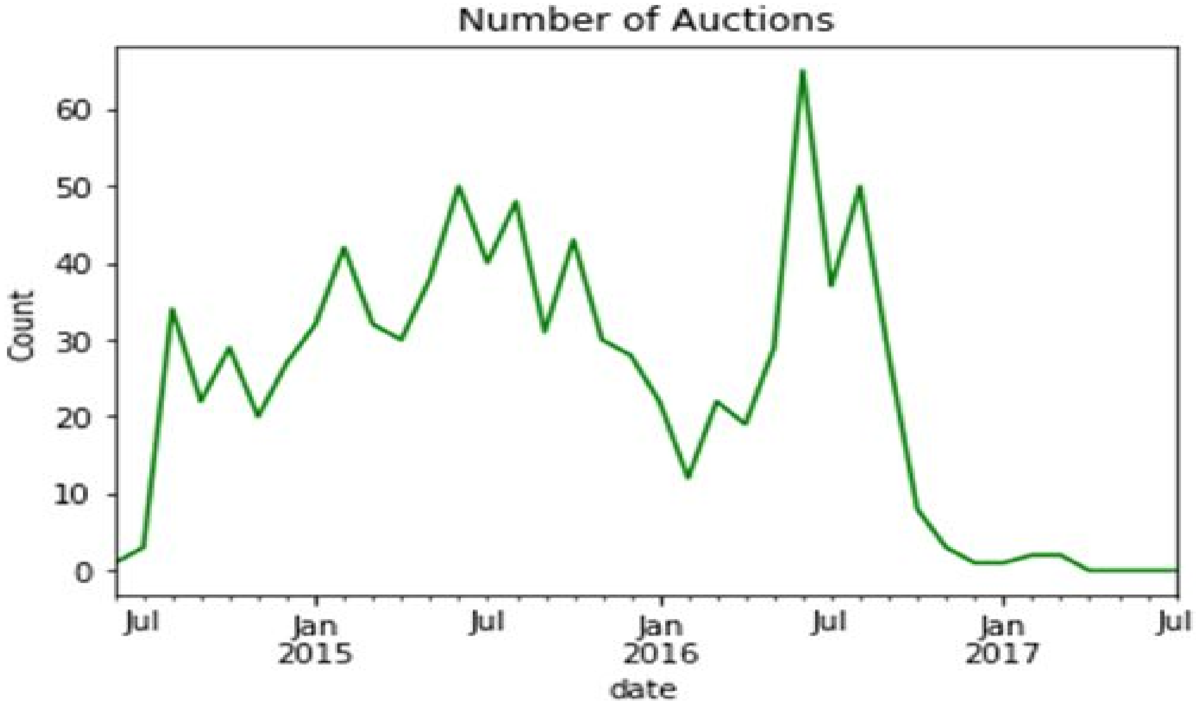
Datasets Used (continued)

- Police—locations of police stations in Detroit.
- Schools—locations of schools in Detroit.
- SMART Bus stops—bus stops in Detroit from Suburban Mobility Authority for Regional Transportation. SMART runs buses between Macomb, Monroe, Oakland, and Wayne Counties, and primarily provides the main non-private car mechanism for people to travel to downtown Detroit.
- Traffic signs—locations of all stop, signal, and yield traffic signs in the greater Detroit area. Fewer traffic signs might be a proxy for more density, or perhaps less government attention to a neighborhood.
- Improve Detroit—this dataset comes from a mobile app, which allows users to report quality of life issues like potholes, running water, and damaged street signs to City Hall, along with photos of the problem.[1] The log of issues starts in December 2014, and runs to the present.
- Parcel Point Ownership—a dataset on the ownership history of the various parcels of property in Detroit. The data goes all the way back to 1912, but is only fairly complete beginning around 1968. FThis dataset comes from the Assessor's Office, and includes data on the assessed value of the property, the taxable value of the property, size, height, and features of property, and its sales history.
- Census data—I drew together vital demographic, income, housing, health, ethnic and racial composition, and other data from the 2015 American Community Survey.· Crime data—I looked at crime stats from 2011-2014 and 2016. I drew on two datasets, which only listed major crimes, and used this as a proxy for the broader crime rate.
- Homicide and crime data—I pulled in data on homicides from a variety of different sources, datasets from 2014, 2015, and 2016, as well as the broader crime rate dataset (2011-2014, 2016).

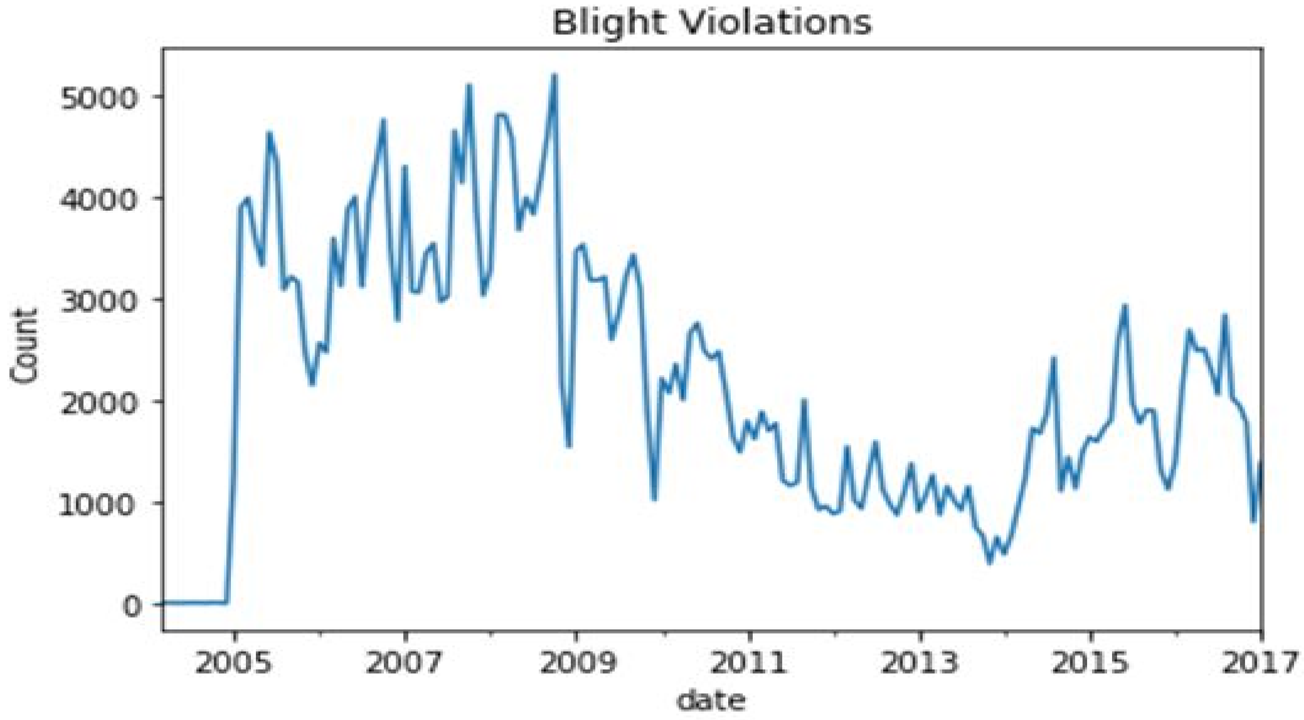
Data Exploration Process

- Group datasets by neighborhood
- Create visualizations to explore interesting features of the various datasets
- Find useful summary features (averages, counts, yearly changes, yearly totals, etc.)
- Check to see if neighborhoods are missing from the dataset, fill missing neighborhoods with zero or NaN values
- Combine into a dataset with 207 observations, save it to the large scale dataset for model building
- For census data, I mapped the census tracts onto neighborhoods
- Not going to describe in full detail all the features I used, as this presentation would be 100 slides, but will talk about a few interesting ones.

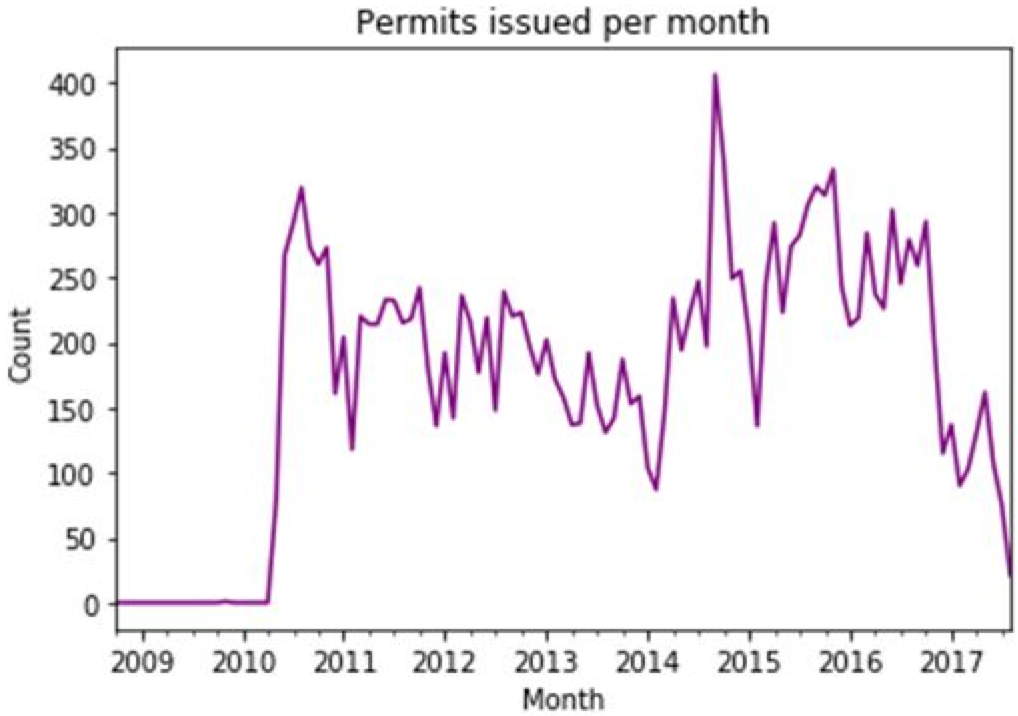
Auctions remained stable, with slight uptick in early 2016



Blight Violation frequency changed pre and post Recession



Total permits issued per month spiked in late 2014



Data exploration summary

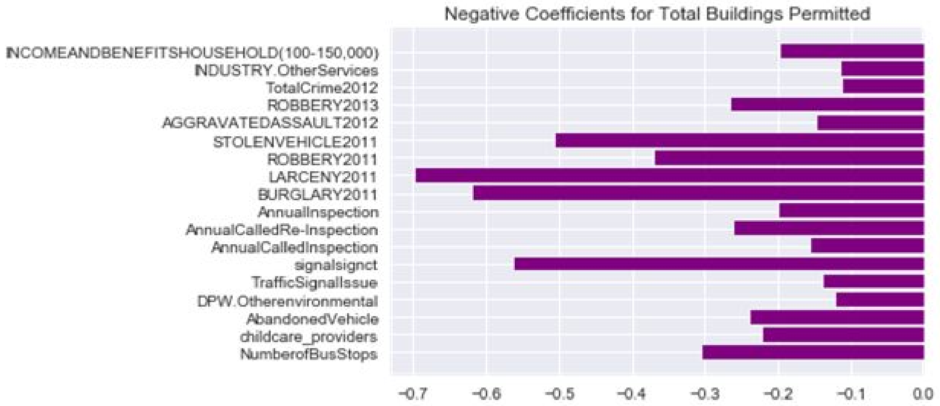
- Chose to use auction as a response variable
- Based on division in blight violations, made them into two separate categories, pre and post 2010
- Relative stability of permits (with one major spike) means that it could also serve as a response variable
- To these, added mean income, crime/homicide, and Improve Detroit Tickets and their wait time

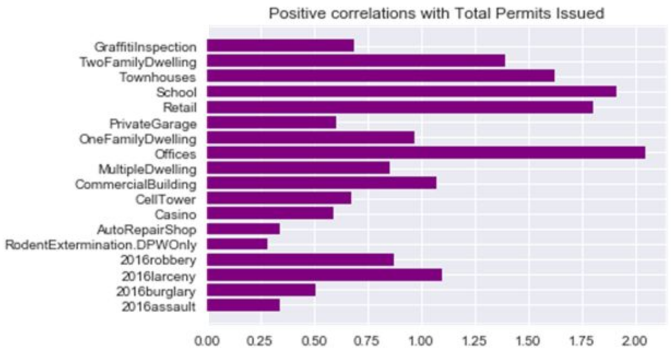
Modeling Process

- Dropped strings and indexes
- Replaced NaNs with 0
- Built functions to implement models
- Tested out Lasso, Ridge, Decision Tree Regression, Support Vector Regression, Random Forest, and Gradient Boosted Regression
- Used feature selection to improve models
- Picked the best model from each of these
- Lasso and Gradient Boosted Regression performed the best

Total Buildings Permitted

- Lasso model
- Total Buildings Permitted Mean: 236.9
- Best alpha: 1.0
- Number of features used: 313
- Root Mean Squared Error on train data: 1.9195287277293813
- Root Mean Squared Error on test data: 43.04753879564554
- Positive coefficients come from different types of buildings constructed
- Negative coefficients primarily come from crime, as well as there being richer people in the neighborhood.



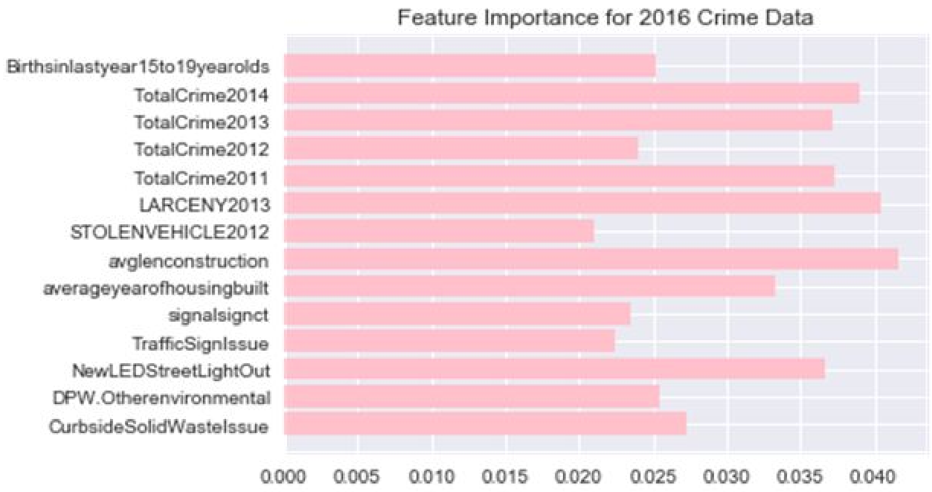


Mean Household Income

Gradient Boosted Regression Model,
Mean Household Income: 42321.783
Median Household Income: 39926.0
Root Mean Squared Error on train data: 0.29
Root Mean Squared Error on test data: 1193.188
Pretty much just proxies for income

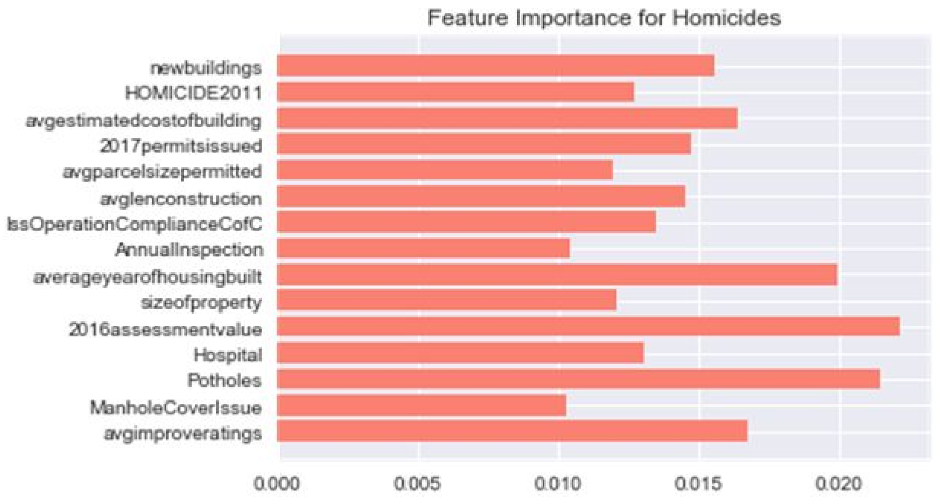
2016 Crime Data

- Gradient Boosted Regression Model
- Crime Mean: 150.1352657004831
- Root Mean Squared Error on train data: 2.8323436921143834
- Root Mean Squared Error on test data: 35.45128560571354
- Crime in 2016 is highly influenced by past crime data
- Teenage pregnancy and year of housing construction is also significant here
- Series of issues related to Improve Detroit or other citations and complaints that predict crime.



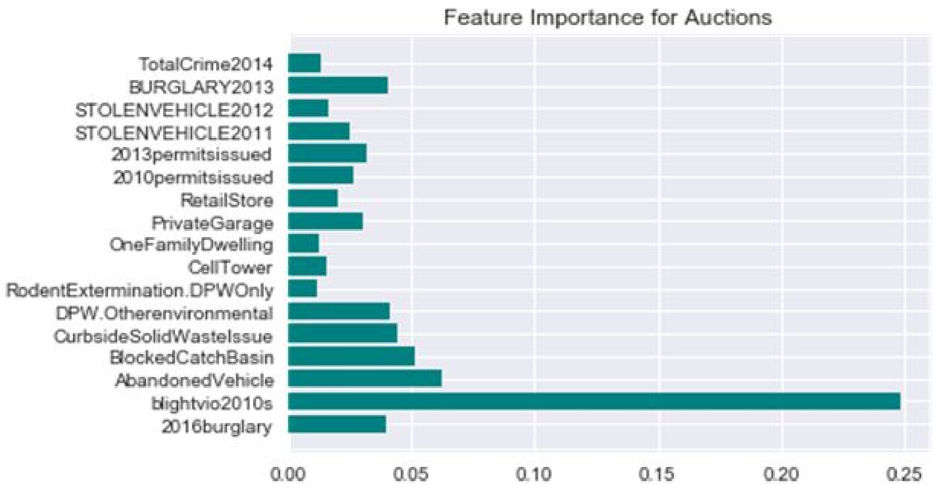
Homicides

- Gradient Boosted Regression
- Average Homicides per Neighborhood: 2.855
- Root Mean Squared Error on train data: 0.005769891832668183
- Root Mean Squared Error on test data: 5.783756023762857
- Perplexingly, this has little to do with crime rates per se
- Tracks things like assessment value, average year of housing built, average length of construction and average parcel size permitted
- Lots of features seem to be proxies for rich/poor neighborhoods



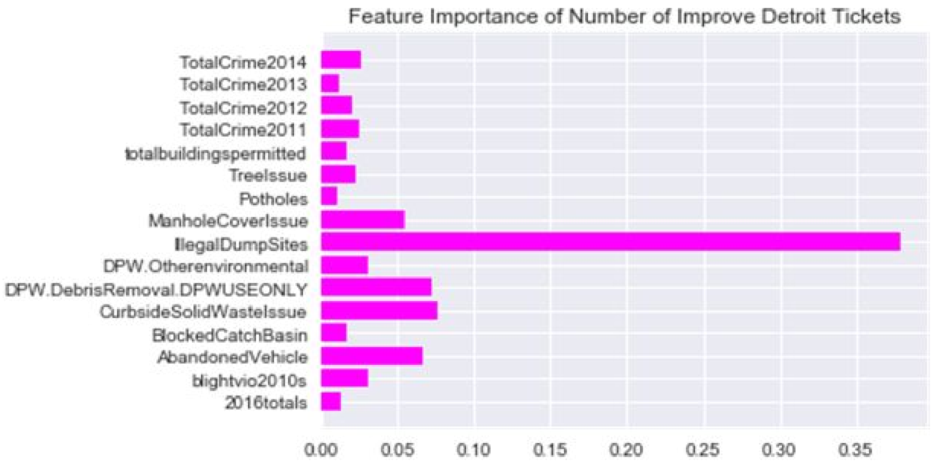
Total Auctions

- Random Forest Regression Model
- Mean Auctions per Neighborhood: 4.4975845
- Root Mean Squared Error on train data: 4.355330534999539
- Root Mean Squared Error on test data: 8.54984426243721
- Exceptionally strong correlation between blight violations and auctions.
- Should be worrisome to policy makers, as blight violations might be causing auctions.



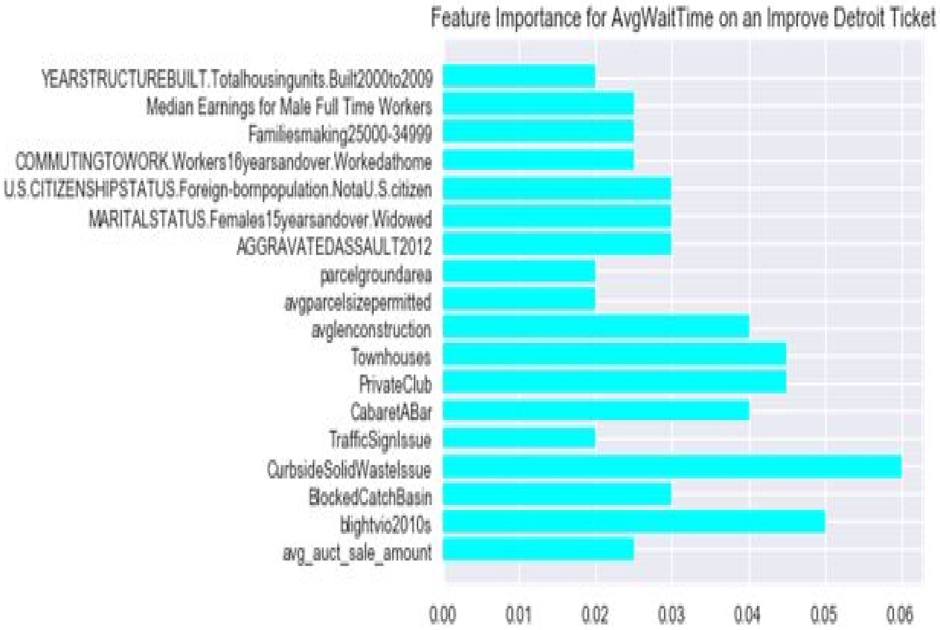
Improve Detroit Tickets

- Random Forest Regression
- Average Number of Improve Detroit Tickets: 884.6763285024155
- Root Mean Squared Error on train data: 474.6401565768995
- Root Mean Squared Error on test data: 145.6483962647053
- High connection to Illegal Dump Sites
- Crime data is also a predictor, perhaps a correlation with higher crime neighborhoods
- Interesting that there's no income divide here



Improve Detroit Ticket Average Wait Times

- Gradient Boosted Regression
- Average Wait Time on Improve Detroit Ticket: 18 days
- Root Mean Squared Error on train data: 2.864260201625337
- Root Mean Squared Error on test data: 8.50389998120845
- First, depends on the issues that are reported
- Second, wait time seems to correlate with features like wait time, such as avgparcelsize, parcelgroundarea, and avgauctsaleamount



Conclusions

- **Blight violations need reform, they have far too strong an influence on auctions.**
- **Auctions are also connected to crime data and other issues such as Blocked Catch Basin, Abandoned Vehicle, and Rodent Extermination.**
- **Some types of buildings seem to correlate strongly with more building permits issued.**
- **Specific types of crime and issues seems to have outsized predictive power.**
- **Homicides seem to be predicted by different features than crime more generally.**
- **Illegal Dump Sites have the highest feature importance for Improve Detroit Tickets.**
- **Improve Detroit tickets appear to be distributed fairly equitably, and not concentrate in neighborhoods based on income.**
- **Curbside Solid Waste and Blight Violations are the strongest predictors of average wait time.**
- **Average wait time generally seems to correlate with income.**