

Help Inform Tax Policy

Jiaxin Tang, Chunxi Wang , Lingshan Yang

Department of Electrical & Computer Engineering, College of Engineering, Boston University

BOSTON
UNIVERSITY

ABSTRACT

- ❖ The Revere Mayor's Office is looking for a deeper assessment of various data sets to better understand the number and percentage of households that are owner-occupied. They are not confident that their official records are accurate and believe there maybe a larger number of homes that are not owner-occupied. The goal of the project is to fully understand the levels of owner-occupied homes and implement a residential tax exemption as an important policy for providing tax relief for homeowners.

DATA

- ❖ Assessor's Database: Includes every property in the city, and who owns it; 9872 samples with 12 types of features (mailing address, owner name, parcel ID, etc)
- ❖ Residential Database: Contains a list of 36,943 residents 18 years and older, merged from both Revere's Voter Registration and Census lists
- ❖ Mailing-Bedroom-Unit Database: Consists of 15110 samples with the actual addresses and all property owners along with the number of units and bedrooms
- ❖ Zillow API Database: Contains features like, latitude, longitude, zipcode, home size, property size, tax value, estimate amount, estimate price range, etc according to each address.
- ❖ Historical Value Database: Contains 11564 distinct properties' values from 2007 to 2018

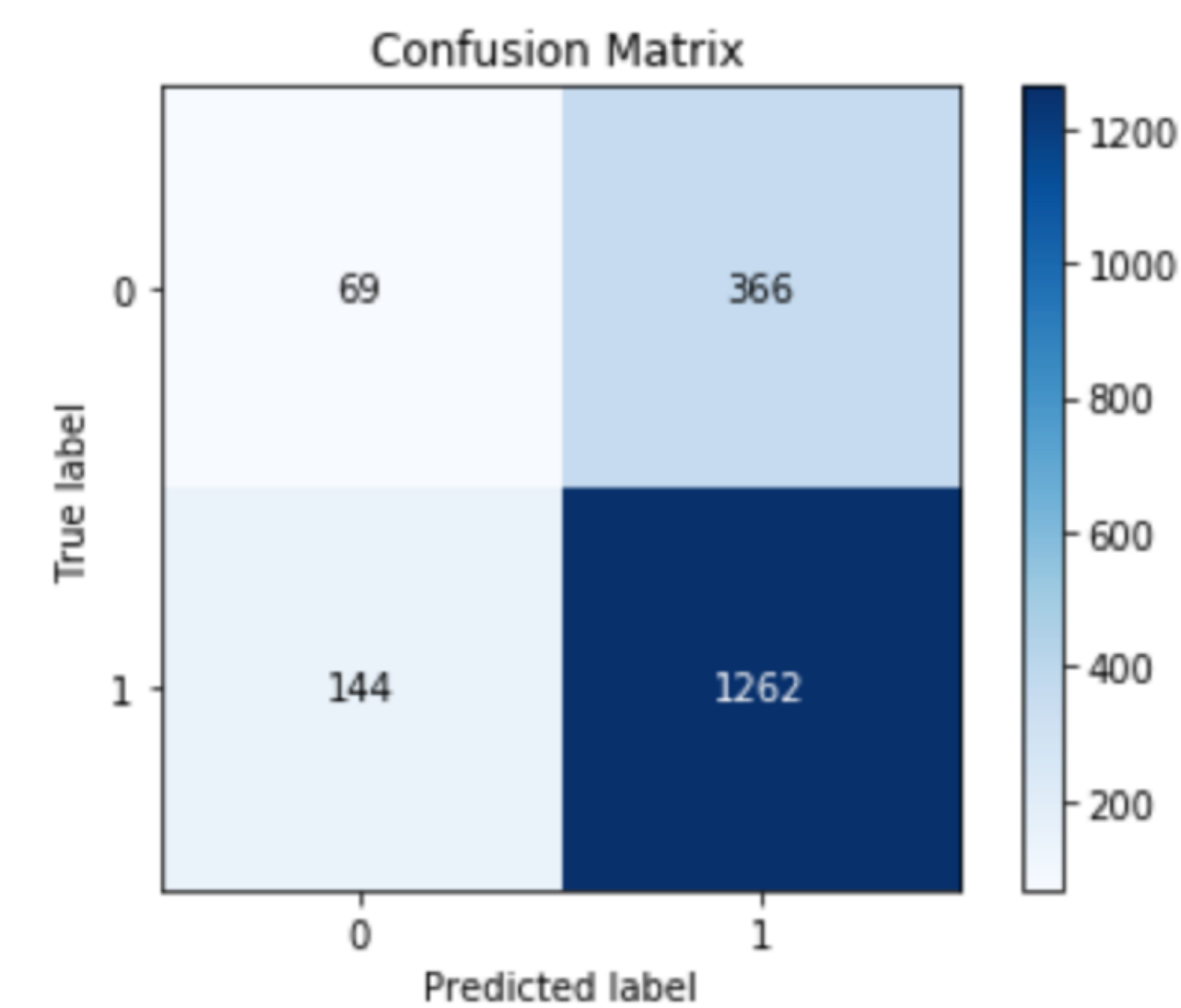
Methodologies

- ❖ Name Match and Address Match: Preprocess Assessor's Database and Residential Database by cleaning and concatenating the data to achieve confirmed owner-occupied house addresses.
- ❖ Zillow API: Utilize merged addresses to get features from Zillow for machine learning model
- ❖ Google Map API: Obtain latitude and longitude based on address. Visualize all the properties on real map.
- ❖ Machine Learning Model: Utilize SVM, LogisticRegression and RandomForestClassifier to train the model in order to determine whether the house is owner-occupied or not.
- ❖ Computation of Residential Tax Exemption Threshold: $\text{threshold} = (-\text{exemption} * \text{nr}) / (\text{tr} - \text{nr})$

RESULTS

- ❖ Based on the Name Match, we can conclude that 82.3% of owners have only one property while based on Address Match, the owner-occupied percent is 73.69%.
- ❖ Utilize RandomForestClassifier and oversample techniques to train the model to predict whether the house is owner-occupied. The accuracy is 72.3%
- ❖ With neighborhood boundaries of Revere and the house type information, we calculate the neighborhood-by-neighborhood owner occupied rate based on different house types.

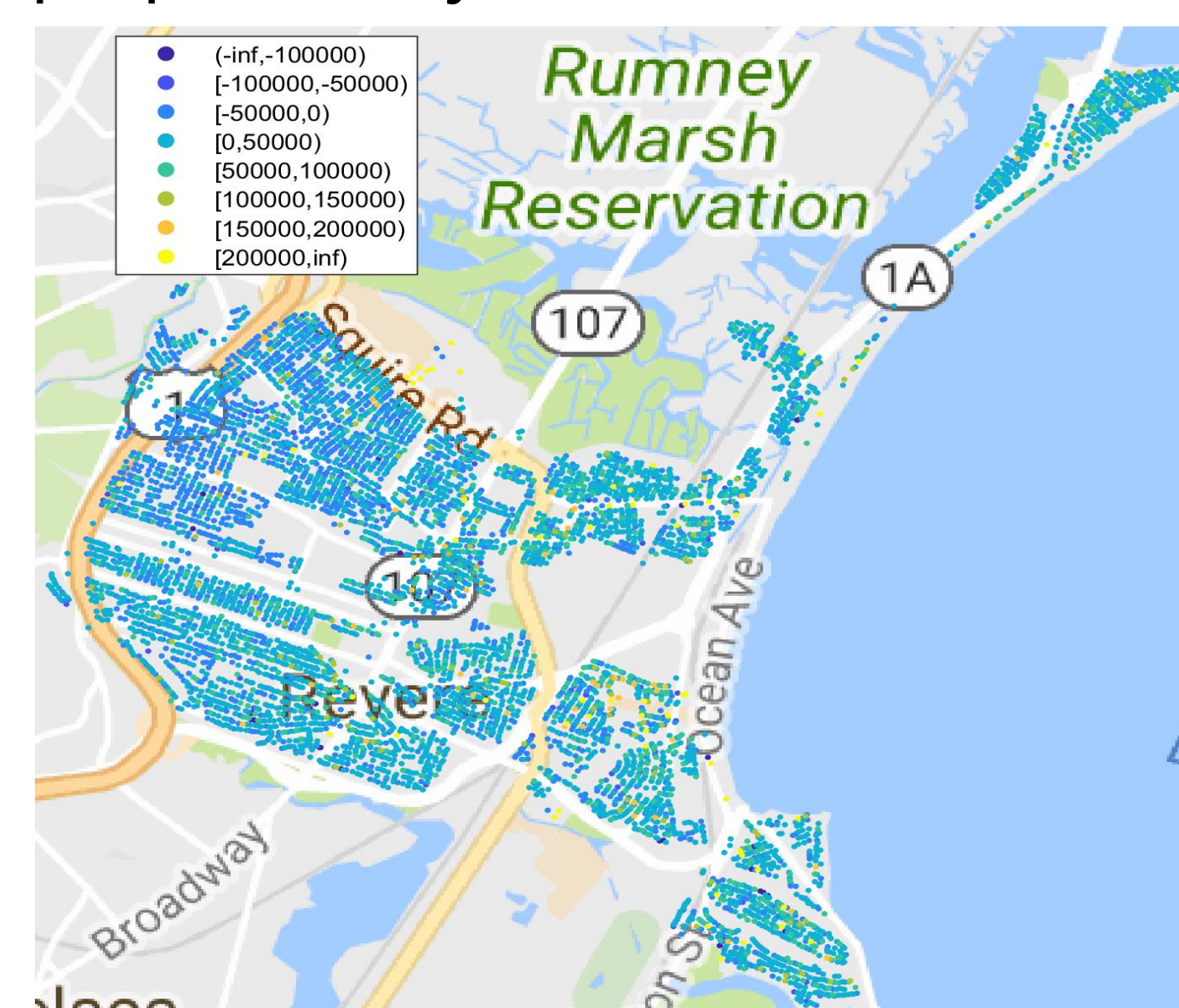
Accuracy: 0.7229766431287343



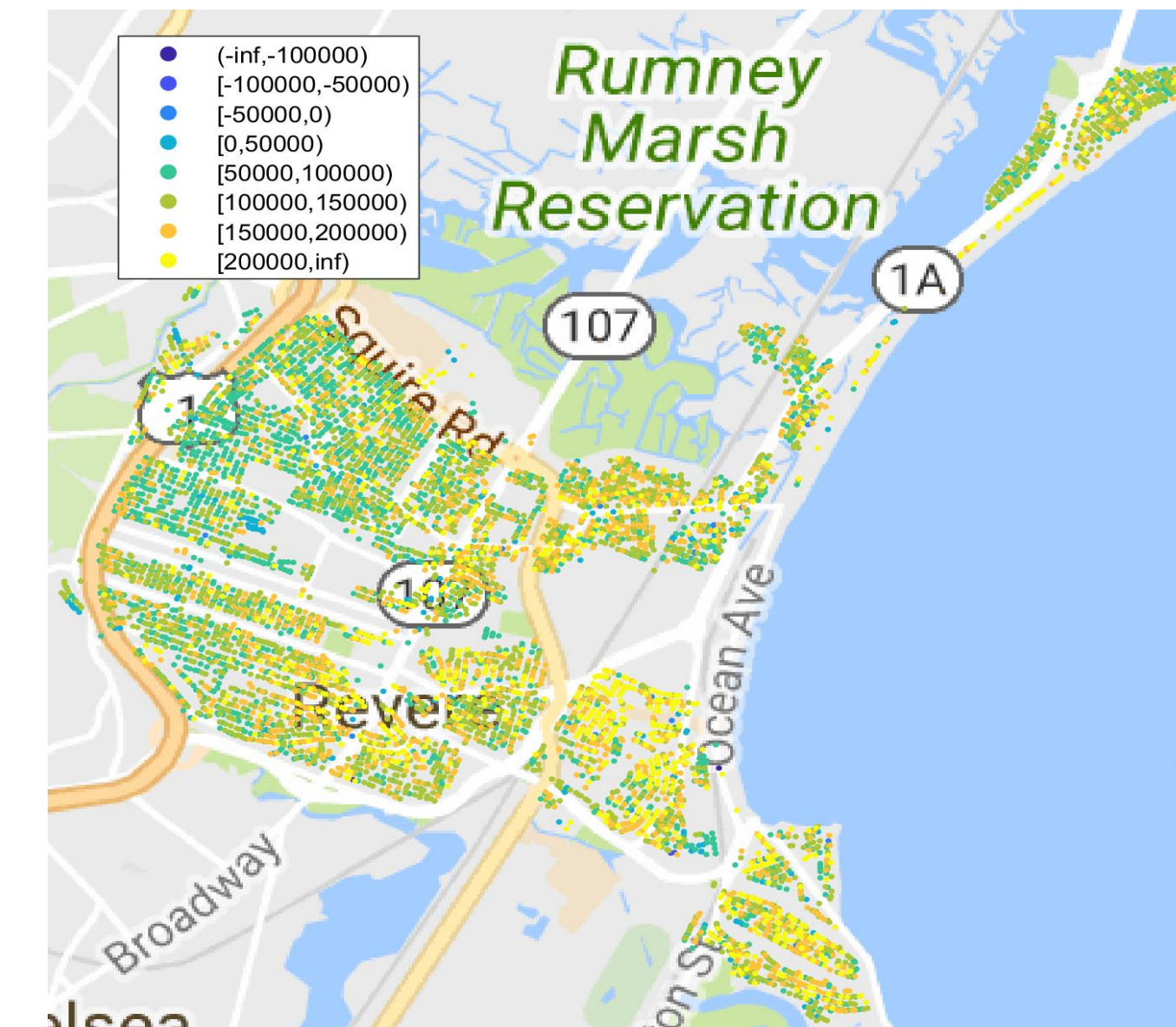
Neighborhood	ONE FAM	TWO FAM	CONDO	OTHERS
0 Point of Pines	0.852632	0.75	1	0.727273
1 West Revere	0.77763	0.766784	0.882353	0.909091
2 Beachmont	0.823899	0.682028	0.882353	0.714286
3 Shirley Ave	0.787565	0.658163	0.909091	0.477273
4 Oak Island	0.821705	0.638889	1	0.444444
5 Upper Revere St	0.760456	0.672131	0.636364	0.571429
6 Park Ave & Hills	0.821483	0.74502	0.906977	0.666667
7 Lower Revere St	0.828283	0.61194	0.9	0.631579
8 Library	0.795732	0.705645	0.8	0.566038
9 North Revere	0.744	0.76087	0.833333	0.333333
10 Riverside	0.784091	0.666667	0.833333	0
11 Revere Beach	0.6	0.833333	1	0.307692
12 Northgate-Squi	0	1	1	NaN

different house type owner-occupied rate by neighborhood in Revere

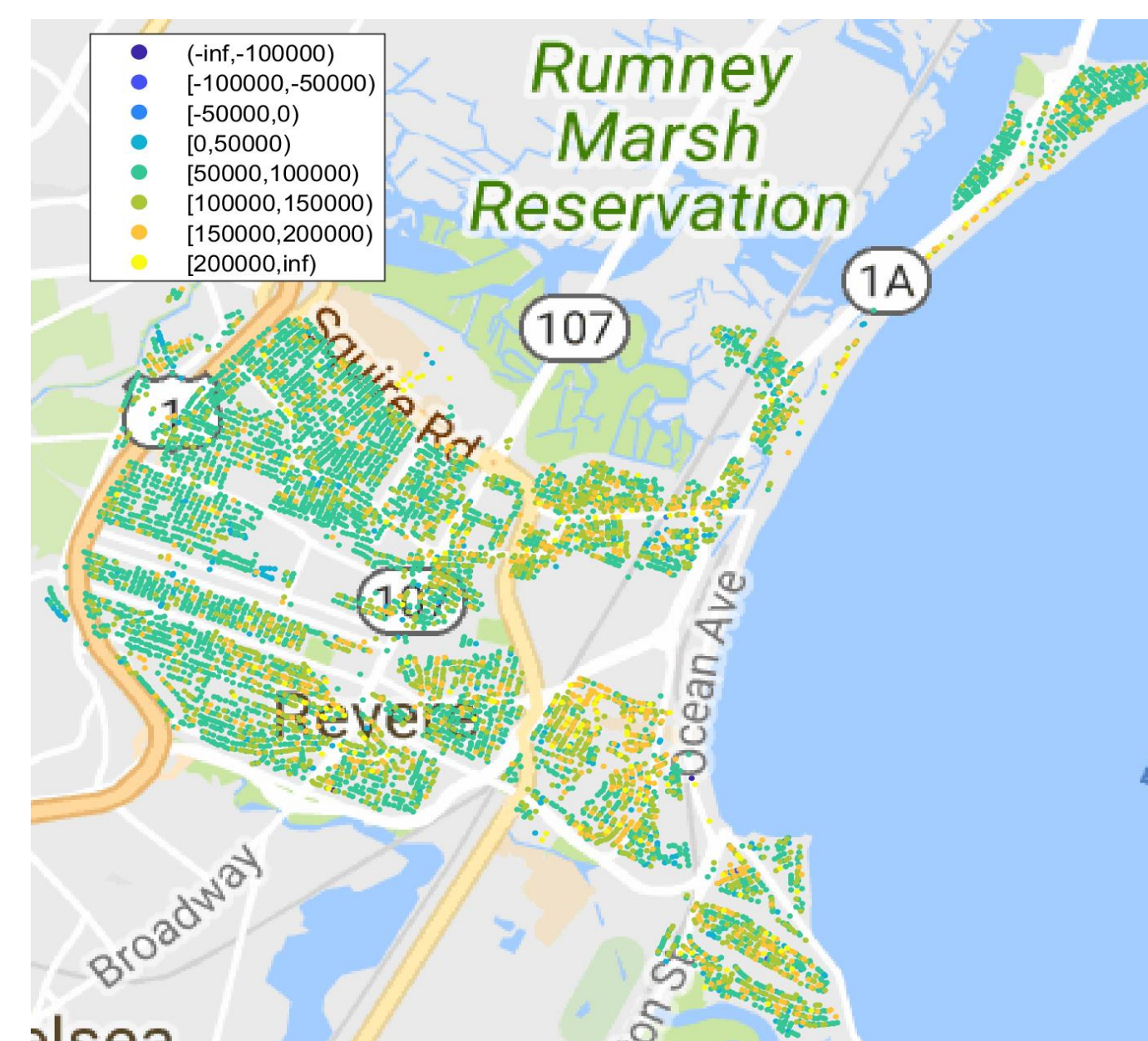
- ❖ Use Google Map API to get real location of properties and colorize properties by different color to reveal value increases.



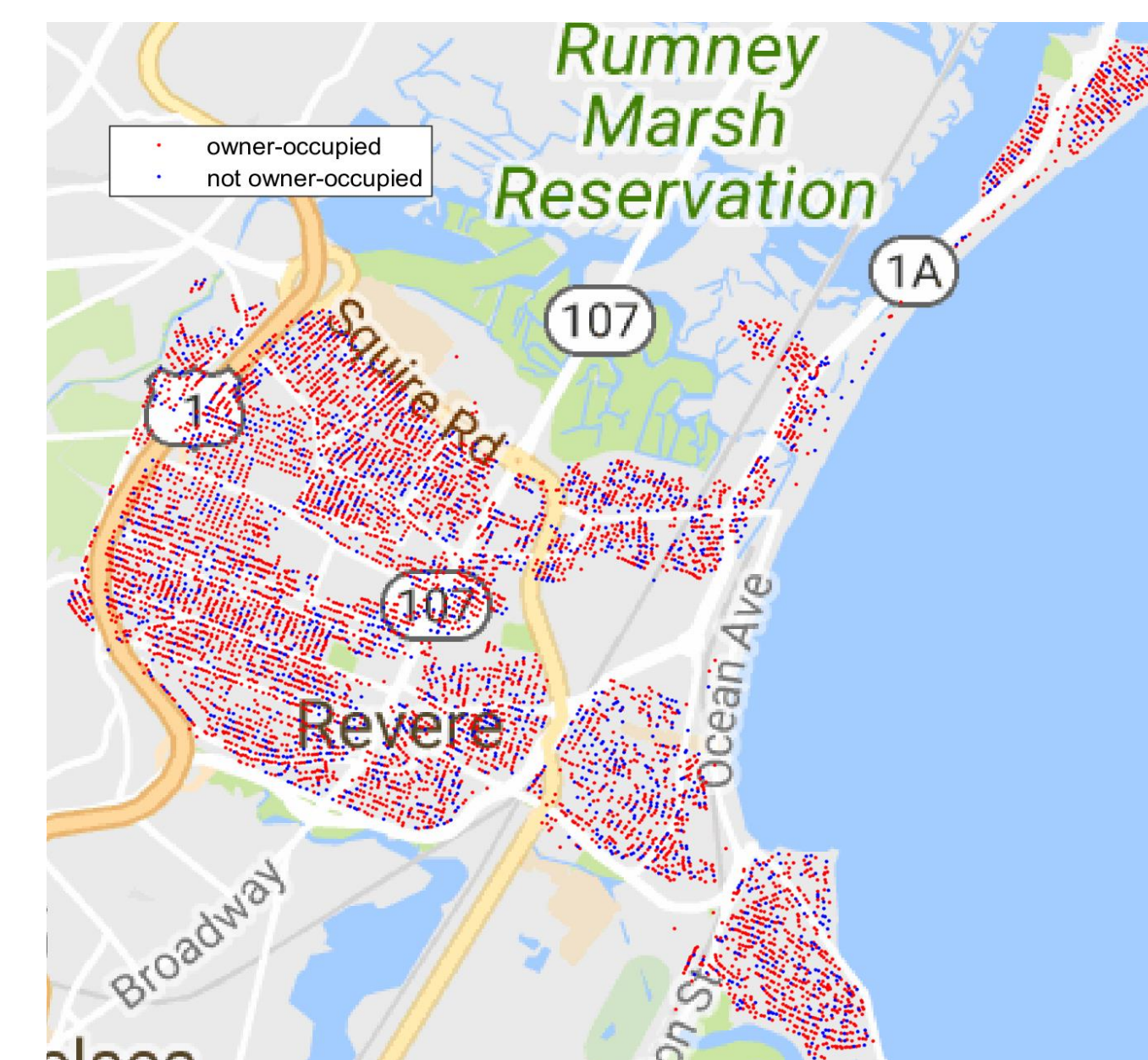
Property Value Increase Map based on the past 10 years (from 2008 to 2018)



Property Value Increase Map based on the past 7 years (from 2011 to 2018)



Property Value Increase Map based on the past 3 years (from 2015 to 2018)



Property Location Map with Owner-Occupied Feature

- ❖ Test our formula based on the data in 2015 and find out lower owner-occupied rate will generate higher exemption threshold.
- ❖ Utilize the owner-occupied rate information to predict the impact of residential exemption based on different neighborhoods. And help the government to make decision on the tax policy on different district.

Value of owner-occupied home	Tax bill w/o exemption	Tax bill with exemption	Change in tax bill*
\$30,000.00	\$433.50	\$7.29	(\$426.21)
\$140,000.00	\$2,023.00	\$1,733.19	(\$289.81)
\$200,000.00	\$2,674.59	\$2,674.59	(\$215.41)
\$260,000.00	\$3,757.00	\$3,615.99	(\$141.01)
\$320,000.00	\$4,624.00	\$4,557.39	(\$66.61)
\$373,718.62	\$5,400.23	\$5,400.23	\$0.00
\$430,000.00	\$6,213.50	\$6,283.29	\$69.79
\$490,000.00	\$7,080.50	\$7,224.69	\$144.19
\$550,000.00	\$7,947.50	\$8,166.09	\$218.59
\$610,000.00	\$8,814.50	\$9,107.49	\$292.99
\$670,000.00	\$9,681.50	\$10,048.89	\$367.39
\$730,000.00	\$10,548.50	\$10,990.29	\$441.79

Impact of Residential Exemption Using real owner occupied rate in West Revere

CONCLUSIONS

- ❖ Generate a new list of 8807 distinct samples with owner-occupied attribute with 12 features in each sample, around 75.9% are owner occupied.
- ❖ Work out a sheet of neighborhood-by-neighborhood owner occupied rate based on different house types.
- ❖ Use Zillow API and Google Map API to get the value of house based on location and also compare the latitude and longitude of the addresses that not match with official ones.
- ❖ Calculate the impact of residential exemption by the given formulas.
- ❖ Generate straightforward figure to show the impact of residential exemption using estimated numbers. An utilize the real Revere neighborhood data to calculate the results.
- ❖ Evaluate the impact on past development of proximity to amenities from the visualization results.
- ❖ Use the features such as latitude and longitude, house value, to build a machine learning model to decide the likelihood that a house is owner-occupied or not.

FUTURE WORK

- ❖ For address match part, we will uniform the format of names and addresses in a better way to elevate the quality of the calculation of owner-occupied percent.
- ❖ For machine learning part, we will improve accuracy by selecting appropriate feature, increasing the size of data set to the whole Boston area and training a more generalized model to interpret results in a reasonable way.