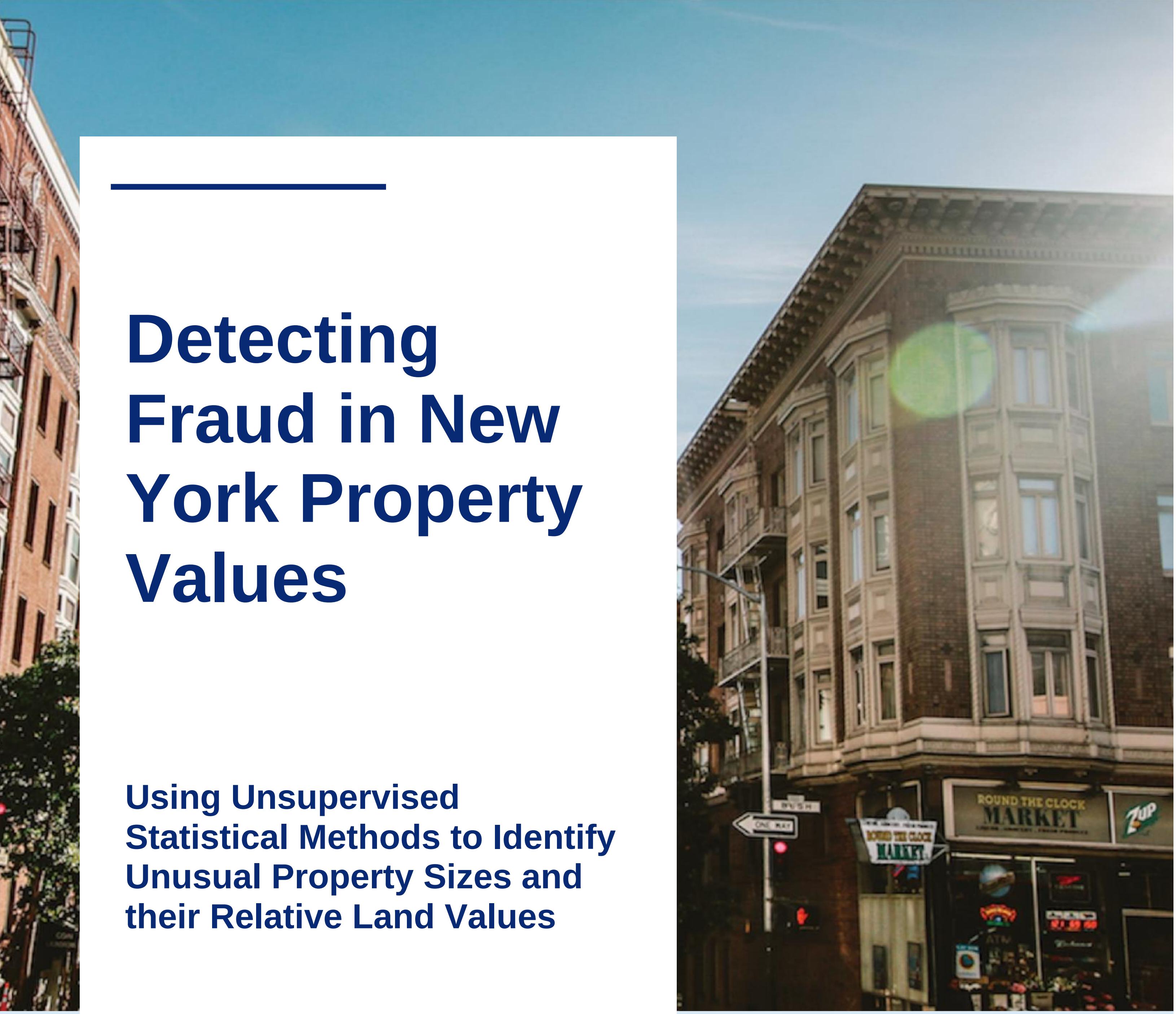


Detecting Fraud in New York Property Values

Using Unsupervised Statistical Methods to Identify Unusual Property Sizes and their Relative Land Values

FEBRUARY 23, 2021

Executive Summary

Unsupervised Methods for Identifying Property Fraud

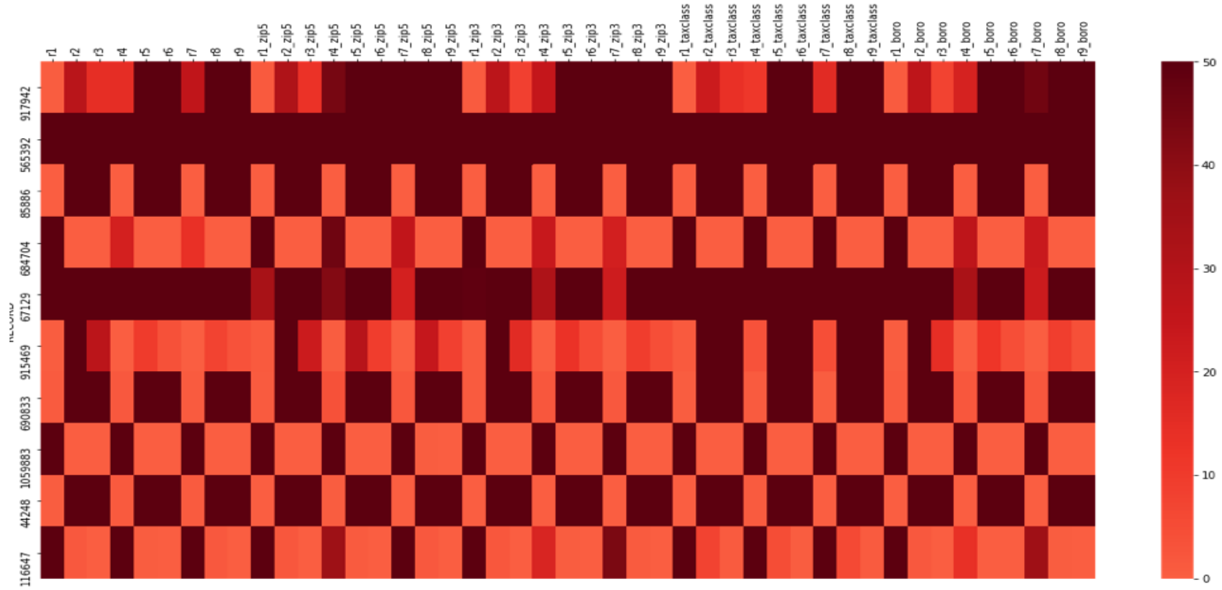
The IRS estimates that tax fraud costs the US government as much as $450 bullion annually. With an estimated 122 million households in the US, it is incredibly difficult for tax investigators to efficiently identify cases of potential real estate and mortgage fraud. This project focuses on building an unsupervised model for finding and filtering for properties based on unusual property sizes and values.

Over one million property records in New York City were analyzed and ordered based on an averaged rank of two different scores. The model uses forty-five input variables that compare the ratio of the land values to the property building and lot sizes. Missing values were replaced by imputing the averages based on the tax class of the property. The variables were standardized using z-scaling and the dimensionality of these variables was reduced using Principal Component Analysis (PCA) to eight key features.

The first scoring method uses the Minkowsky distance between the scaled variables and the origin, with the higher the distance, the more "unusual" the record. The second method uses an autoencoder, an unsupervised method that compresses the data and learns how to reconstruct the data, and creates a score based on the Minkowsky distance of the autoencoder error (the difference between the input and outputs). These two scores were then converted into ranks, based on other scores in the dataset, and averaged to create a single scaled score. The returned dataset contains the top 100 records that were flagged as anomalies, which would be given to a property investigator.

Analysis of the Top 10 Unusual Records

Records were classified as unusual based on the 45 engineered ratio variables, with the heatmap below displaying the specific values that contributed to the model flagging this record as unusual.



Heat Map for Top 10 Unusual Records

Record: 917942

Address: 154-68 BROOKVILLE BOULEVARD, 11422

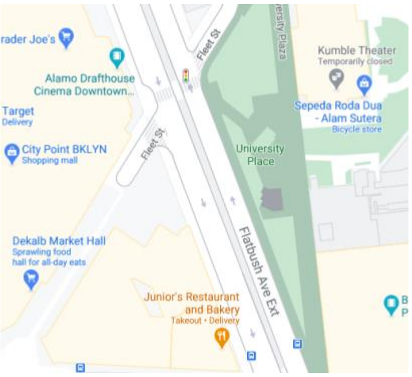
Description: This record had unusually high land values with AVLAND ($1.7 billion) and AVTOT ($4.6 billion). After investigation of the property through google maps, no building appears at the given address, the nearest building being a Holiday Inn across the street.

Brookville Blvd
Holiday Inn New York Jfk Airport Area
3.9 ★ (949)
3-star hotel
way Blvd
y Blvd
Broo

Record: 565392

Address: FLATBUSH AVENUE

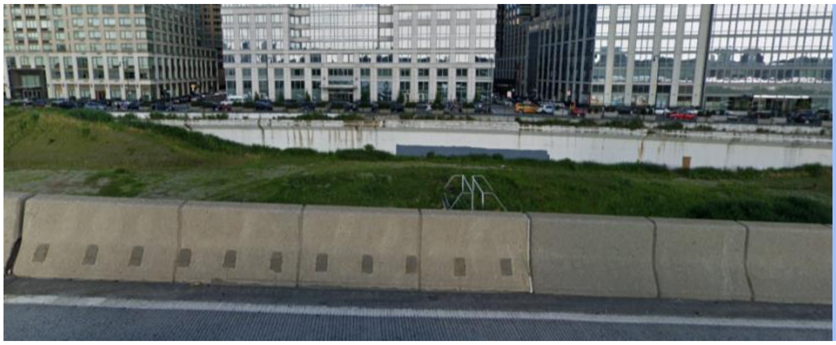
Description: This is a government-owned property with high land values: FULLVAL ($4.3 billion), AVLAND ($1.95 billion), and AVTOT ($1.95 billion). Without a specific address given, a precise location cannot be pinpointed. However, University Park appears to be a likely match based on the missing building lot fields.

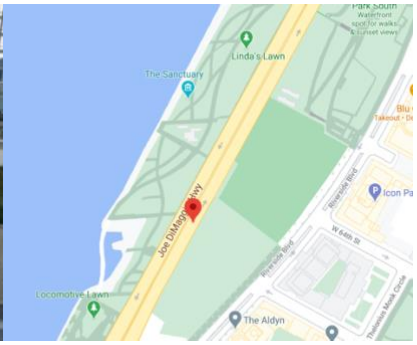


Record: 85886

Address: JOE DIMAGGIO HIGHWAY

Description: Although the lot size is large compared to other records on file, the the building sizes, BLDFRONT (8) and BLDDEPTH (8), are small given the property's high FULLVAL ($70.2 million). The listed owner is marked as Parks and Recreation, with the park identified below being a likely property match.





Record: 684704

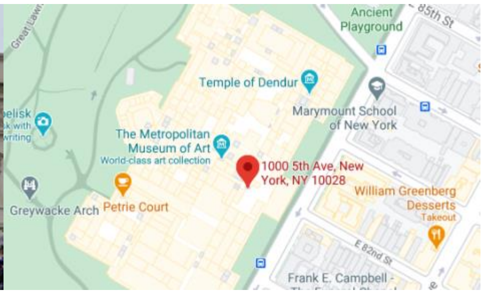
Address: 69 Street

Description: This property, owned by W Rufert, is missing a significant number of fields and was flagged due to a small lot depth compared to the average land value from a similar tax class. The lack of information makes it difficult to pinpoint an exact location.

Record: 67129

Address: 1000 5 AVENUE, 10028

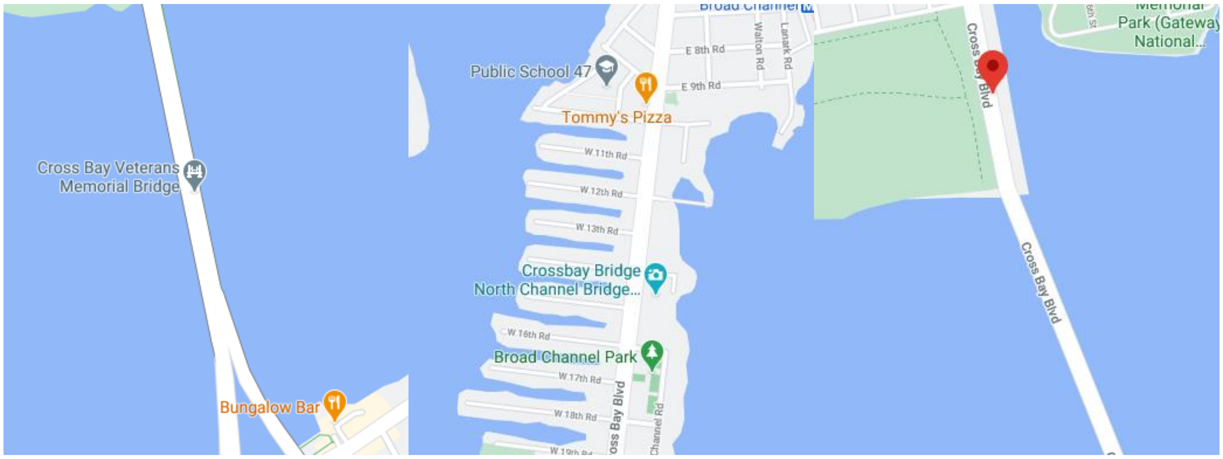
Description: This property's land values are huge compared to others in the dataset with a FULLVAL of $6.2 billion, AVLAND of $2.7 billion) and AVTOT of $2.7 billion. The large lot size and exact address helps us identify this record as the Metropolitan Museum of Art.


Ancient Playground
85th St
Temple of Dendur
Marymount School of New York
The Metropolitan Museum of Art
World-class art collection
1000 5th Ave, New York, NY 10028
William Greenberg Desserts
Takeout
Greywacke Arch
Petrie Court
E 82nd St
Frank E. Campbell

Record: 915469

Address: CROSS BAY BOULEVARD, 11414

Description: Compared to other building areas by a similar tax class, this record was flagged due to the large FULLVAL of $540 million. The given lot values appeared unusual as well, with dimensions of 999 by 999. However, this property is owned by the US government with many different bridges being likely candidates.



Record: 690833

Address: 83-98 FOREST PARKWAY, 11385

Description: Given the street address and the park owner, Parks and Recreation, the park located below appears to match this record. The small building dimensions to (20 x 20) to the large land values: FULLVAL ($242 million), AVLAND ($104 million), and AVTOT ($109 million) caused the model to flag this record.

83-98 Forest Pkwy, Flushing, NY 11385
Jackie Robinson Pkwy
Bethel Ave
Forest Park Dr
Forest Park Carousel Amusement V
George Seuffert Sr. Bandshell
Forest Park Golf Course
Airplane Field
Forest Park Dr
Oak Ridge
Forest Pkwy
Park Ln S
85th St
86th St
87th St

Jackie Robinson Pkwy
Forest Park Skate
Forest Park Carousel Amusement Village
83-98 Forest Pkwy, Flushing, NY 11385
George Seuffert Sr. Bandshell
Forest Park Dr
Forest Park Tennis C
Forest Park Golf Course
Park Dr

Record: 1059883

Address: SAGONA COURT

Description: This record is missing all the property value fields (FULLVAL, AVLAND, AVTOT) and the building area values (BLDFRONT, BLDDEPTH) and was flagged for having a small relative lot area compared to the average land value for the tax class. While an exact address wasn't given, this block contains 4 properties all with style/sizing to the property in the photo to the right (20 Sagona Court).

Sagona Ct
Sagona Ct
40
20
30
92
88

Record: 44248

Address: 2 MADISON AVENUE, 10010

Description: Another Parks and Recreation owned property, this record was flagged for having large land values: FULLVAL ($180 million), AVLAND ($77 million), AVTOT ($81 million) to a low building area: depth (20) and front (20). The lot values (610 front, 534 depth) make Madison Square Park a likely property match.

5th Ave
Madison Square Park
Urban oasis with Flatiron building views
Police Officer Moira Ann Smith Playground
Jemmy's Dog Run
Dog park with a section for small breeds
2 Madison Ave, New York, NY 10010
Eternal Light Flagstaff
Madison Ave
Sony/ATV Music Publis
Eleven Madison Takeout
Madison Ave

Record: 116647

Address: 1849 2 AVENUE, 10128

Description: Although the building volume for this property is large (standing at over 35 stories), this record was flagged for having high land values: FULLVAL ($161 million), AVLAND ($19 million), AVTOT ($72 million) compared to its lot size: LTDEPTH (75) and LTFRONT (25).

S
Cardtronics ATM
1849 2nd Ave, New York, NY 10128
Rite Aid
Drug store
235