Boston Area Research Initiative ● Northeastern & Harvard Universities
E-Mail: bari@northeastern.edu ● Web: www.bostonarearesearchinitiative.net

Documentation for Cross-sectional Property Assessment Database 2022

Overview

This document describes the structure and organization of the City of Boston Assessing Department's centralized database for property-specific data for all uniquely identifiable properties in the city (n = 180,626) for the year 2022. Boston's Assessing Department is responsible for determining ownership and physical characteristics for all properties in the city in order to ensure fair assessment of both taxable and non-taxable property in Boston. This dataset can be used to analyze valuation, structure, land use, and other details for all properties. The data are released by the Assessing Department annually as part of the City of Boston's open data initiative through *data.boston.gov*. The data are then processed by the Boston Area Research Initiative (BARI), during which additional variables are introduced to facilitate informed analysis and other aggregate measures are generated describing the properties within neighborhoods.

The main dataset *(PADCross.Record.YEAR.csv)* is the base file derived from the tax assessor's annual release through data.boston.gov but is curated by BARI to contain a handful of additional variables to facilitate informed analysis.

We also offer aggregated datasets that track change in assessment and use over time at the tract and block group levels for the 2010 and 2020 Census Geographies, such as the *PADCross.CT10.YEAR.csv* file. This file contains aggregate measures, also known as ecometrics, that describe neighborhoods at the 2010 census tract level. These variables are available in a spreadsheet format (*.csv*) and as mappable shapefiles (*.shp*).

Before 2016, BARI has released these datasets using the name "Tax Assessor's Database" rather than "Property Assessment Database". We have changed names in order to match the names used by the City of Boston.

Table of Contents

1. Summary of Record-level Property Assessment Data (*PADCross.Record.YEAR.csv)* 3
1.1. Description of Variables 3
1.1.1. Identifying Characteristics 4
1.1.2. Property and Building Characteristics 4
1.1.3. Geographical information 10
2. Summary of Aggregate Measures (e.g. PADCross.CT.YEAR.csv & corresponding shapefiles) 11
2.1. Description of Variables 12
3. Appendix 14
3.1. Appendix A: Property Types (PTYPE) Codes 14
3.2. Appendix B: Codes for Land Use 15
3.3. Appendix C: Unit Imputation 16

1. Summary of Record-level Property Assessment Data (*PADCross.Record.YEAR.csv)*

The City of Boston's Assessing Department is responsible for determining accurate values for all properties in the city. To this end the Department maintains property ownership and value information to ensure fair assessment of both taxable and non- taxable property in Boston. Assessing records are compiled and reviewed annually to reflect changes to properties as a result of new construction, remodeling, and changes in ownership. The data contained within describes the property-specific address, ownership, type, structure, class, and valuation data. Upon annual review and re-assessment, property-specific data is updated and changes in calculated values are adjusted to reflect the most up to date status for each property. Property taxes (as indicated within this dataset as "GROSS_TAX") are also adjusted annually to reflect the annual taxation rates for residential and commercial properties.

The tax rate is the amount a taxpayer owes for each one thousand dollars of property value in a given year. The tax rate for a given fiscal year appears on the third quarter tax bill, which is typically issued in late December Commercial tax rates are calculated in the same manner.[1]

This dataset is a modified version of the original Tax Assessor's Data Set, including all variables included in the original data as well as others introduced by BARI to facilitate informed analysis. Overall, the dataset provides insight into the current physical makeup and history of properties in Boston.

1.1. Description of Variables

Property assessment variables are split into three categories: identifying characteristics, property and building characteristics, and geographical information. Identifying characteristics include variables regarding the basic identity and attributes of the address. Building characteristics include information on the physical attributes of the building containing the property. Geographical information provides further detail on the location of the property and the other geographies that contain it.

[1] Tax rate calculation information published by the City of Boston Assessing Department through the department's website: http://www.cityofboston.gov/assessing/taxrates.asp

1.1.1. Identifying Characteristics

- *PID* is the 10-digit property identification number, unique to each property. The first two digits indicate the Ward, digits 3 thru 7 are the parcel, and digits 8 thru 10 are the sub-parcel.
- *CM_ID* is the 10-digit parcel number of the main condo building parcel. All condo units in each building are related to this number.
- *GIS_ID* is another 10-digit property identification number. It is the unique identifier for the land the property is in (this is slightly different than *LAND_PARCEL_ID*, however, as the latter combines some land parcels into one).
- *ST_NUM* is the street number of the property.
- *ST_NAME* is the street name of the property.
- *UNIT_NUM* is the specific unit number within a multi-unit building.
    - Note: Tabulated from the City of Boston's Street and Address Management (SAM) system. For cases without unit data in SAM values are imputed[2] (see Appendix D for methodology).
- *ZIPCODE* is the zip code of the property.

1.1.2. Property and Building Characteristics

- *PTYPE* is the land usage Property Type of the property. Property Types are classified according to State Class Codes, which are three digit codes. PTYPES between 001 and 299 are Residential properties. PTYPES between 300 and 399 are Commercial properties. PTYPES between 400 and 465 are Industrial properties. PTYPES greater than 900 are Exempt properties, meaning they are fully (or partially) exempt from property taxes. See Appendix A for a full list of PTYPES.
- *LU* is the Land Use type for the property. Codes for land use can be found in Appendix B.

[2] 55% of the properties had data indicating units within the property. We imputed the number of units for the remaining parcels. For certain land uses, we assumed the number of units based on the definition of the land use itself including R1-R3, commercial lots, residential lots, and condo parking and condo main. The number of units of the remaining properties were imputed using regression-based imputation, leveraging data from assessments including the total assessed value for the property, land, and building and the total gross floor area and living area.

- *OWN_OCC* is a one-character code that indicates if an owner receives a residential exemption for the property. "Y" indicates that the owner claims to live within the property (a.k.a. the property is "owner-occupied") and a "N" indicates the opposite.
- *OWNER* is the primary owner of the property, as of the date of the prior calendar year.
- *MAIL_ADDRESSEE* is the name associated with the street mailing address of the owner, if it is not the owner.
- *MAIL_ADDRESS* is the street mailing address of the owner, to which the property tax bill is mailed.
- *MAIL_CITY* is the city and state to which the property tax bill is mailed.
- *MAIL_ZIPCODE* is the zip code of the property where the tax bill is mailed.
- *LAND_VALUE* is the assessed value of the land.
- *BLDG_VALUE* is the total assessed value for the building on the property.
- *TOTAL_VALUE* is the total assessed value for the property. It is a summation of the assessed values of the land and building.
- *GROSS_TAX* is the amount billed to the owner as property excise tax. It is based on the total assessed value multiplied by the tax rate. Tax rates are adjusted each year for Residential and Commercial property types.
- *LAND_SF* is the total size of the property in square feet. This is also known as the lot size.
- *YR_BUILT* is the year in which the property was built. The original dataset held many properties whose year of construction was listed as zero. It was fixed by updating the YR_BUILT variable, which now contains a "NA" value where it previously showed a "0".
- *YR_REMODEL* is the year in which the property was last remodeled. For some properties the year of its most recent remodel was listed as zero. It was fixed by updating the YR_REMODEL variable, which now contains a "NA" value where it previously showed a "0".
- *GROSS_AREA* is the gross floor area for commercial properties

- *RES_FLOOR* is the number of levels in the structure that is located on the property.
- *STRUCTURE_CLASS* is the structural classification of commercial buildings. Classes include: *A* for Steel Structure, *B* for Reinforced Concrete, *C* for Brick & Concrete, *D* for Wood/Frame, *E* for Metal, and *R* for Residential.
- *ROOF_STRUCTURE* is the roof type for structures. Types include: F for Flat, G for Gable, H for Hip, L for Gambrel, M for Mansard, O for Other, and S for Shed.
- *ROOF_COVER* is the exterior finish type for the roof of the properties. Types include**:** A for Asphalt Shingl, C for Composition, R for Rubber Roof, S for Slate,T for Tile, O for Other and W for Wood Shingle.
- *BED_RMS* is the total number of bedrooms in a structure.
- *FULL_BTH* is the total number of Full Bathrooms in a structure. A full bath is also known as a four-piece bath, which includes a shower, a tub, a sink, and a toilet.
- *HLF_BTH* is the number of Half Bathrooms in a structure. A half bath typically includes a sink and a toilet, also known as a powder-room.
- *BTHRM_STYLE1* is the bath style of the first bathroom in a structure. Types include: L for Luxury, M for Modern, N for No Remodeling, S for Semi-Modern.
- *BTHRM_STYLE2* is the bath style of the second bathroom in a structure. Types are above.
- *BTHRM_STYLE3* is the bath style of the third bathroom in a structure. Types are above.
- *KITCHEN_TYPE* is the type of kitchen in a condominium unit. Types include: F for Full Eat In, 1F - 1Full Eat In Kitchens, 2F - 2 Full Eat In Kitchens, 3F - 3 Full Eat In Kitchens, 4F - 4 Full Eat In Kitchens, 0F - 0 Full Eat In Kitchens, N for None, O for One-Person, and P for Pullman.
- *KITCHEN_STYLE1* is the kitchen style of the first kitchen in a structure. Types include: L for Luxury, M for Modern, N for No Remodeling, S for Semi-Modern.

- *KITCHEN_STYLE2* is the kitchen style of the second kitchen in a structure. Types are same as *KITCHEN_STYLE1*.
- *KITCHEN_STYLE3* is the kitchen style of the third kitchen in a structure. Types are same as *KITCHEN_STYLE1*.
- *HEAT_TYPE* is the type of heating in a structure. Types include: N for None, S for Space Heater, W for Hot Water, E for Electric, P for Heat Pump, F for Forced Air, and O for Other.
- *AC_TYPE* indicates if the structure has air conditioning. Types include: C for Central Air Conditioning, D for Ductless Air Conditioning, "Y" for Yes and N for None.
- *FIRE_PLACE* is the total number of fireplaces in a structure.
- *EXT_COND* is the exterior condition of a structure. Types include: A for Average, E for Excellent, F for Fair, G for Good, and P for Poor.
- *OVERALL_COND* is the overall condition for a structure. Types include: VG for Very Good, US for Unsound, A for Average, EX and E for Excellent, F for Fair, G for Good, P for Poor, VP for Very Poor, AVG for Default or Average.
- *INT_COND* is the interior condition of a structure. Types include: A for Average, E for Excellent, F for Fair, G for Good, and P for Poor.
- *PROP_VIEW* is the view for a structure. Types include**:** A for Average, E for Excellent, F for Fair, G for Good, P for Poor, and S for Special.
- *NUM_BLDGS* is the number of buildings in property.
- *BLDG_TYPE* is the building style for properties. The styles are: BL for Bi-Level, BW for Bungalow, CL for Colonial, CN for Contemporary, CP for Cape, CV for Conventional, DK for Decker, DX for Duplex, L for Tri-Level, Oth for Other, RE for Row End, RM for Row Middle, RN for Ranch, RR for Raised Ranch, SL for Split Level, TF for Two-Family Stack, TD for Tudor, SD for Semi Detached, and VT for Victorian.
- *RES_UNITS* is the number of residential units in a property.
- *COM_UNITS* is the number of commercial units in a property.

- *EXT_FINISHED* is the exterior finish type for condominium buildings. Types include: A for Asbestos, B for Brick & Stone, C for Cement Board, F for Frame/Clapboard, G for Glass, K for Concrete, M for Vinyl, O for Other, P for Asphalt, S for Stucco, U for Aluminum Sliding, V for Brick/Stone veneer, 10 for Aluminium/Vinyl, 12 for Concrete Block, 09 for Wood Siding, 01 for Brick, 06 for Metal/Glass, 03 for Poured Concrete, 04 for Precast Concrete, 02 for Stone, 08 for Stucco, 15 for Corrug Siding, 11 for Metal Siding, 05 for Concrete and Glass, 13 for Br Sill/Sash, 07 for Stone/Marble, 14 for Hollow Tile and W for Wood Shake.
- *NUM_PARKING* is the number of parking spaces associated with a property.
- *CORNER_UNIT* indicates if a condo unit is in the corner of a building. Y indicates Yes and N indicates No.
- *LIVING_AREA* is the total living area for residential properties.
- *AV_BLDG_PER_SF* expresses the assessed value of a property's building, divided by its gross floor area in square feet.
- *AV_LAND_PER_SF* is the assessed value of a property's land, divided by the total property area in square feet.
- *SIMPLIFIED_LU* The Tax Assessor's dataset includes 17 different types of land use, including 6 different codes for residential use at varying densities (single floor houses, two-floor, etc). The many different classifications for similar uses can result in "artificially" high scores when used to calculate the diversity in land use distribution for a given area. SIMPLIFIED_LU reflects land use according to a simplified classification, with eight core uses: Residential, Commercial, Condo, Mixed Residential/Commercial, Agricultural, Industrial, Tax Exempt, and Tax Exempt by the Boston Redevelopment Authority (applies to properties that are undergoing renovation projects).
- *COOL_SCORE* represents residential cooling types in form of energy efficiency score. Each residential cooling system type is allocated a numeric score based on its energy efficiency performance.
- *AGE_SCORE* This variable represents residential unit energy efficiency based on the age of building. The scores were allocated on the assumption

that older buildings are more energy inefficient.

- *HEAT_SCORE* represents the residential heating system types in form of energy efficiency score. Each residential heating system type is allocated a numeric score based on its energy efficiency performance.
- *EE_SCORE* is an aggregate variable that combines the HEAT_SCORE, COOL_SCORE and AGE_SCORE in weighted sum [EE_SCORE = AGE_SCORE+0.75*HEAT_SCORE+0.75*COOL_SCORE]. It indicates the property specific composite energy efficiency index. This variable is only defined for properties with land usage R1, R2, or R3.
    - *Note:* This variable takes into account three other variables which are otherwise not included in visualization data but necessary to calculate energy efficiency scores.
- *BLDG_AGE* is the current year minus the year in which the building was most recently remodeled or the year in which it was first built if it was never remodeled.
- *LU_DESC* is the description of the type of the property. Refer the appendix for the various Land Use Description types.
- *CD_FLOOR* is the number of commercial levels in the structure that is located on the property.
- *INT_WALL* is the condition of the wall in the interior of the structure. Types include**:** N for Normal, S for Substandard, G for Good and E for Elaborate
- *KITCHEN* is the number of Kitchens in a structure.
- *TT_RMS* is the total number of rooms in a structure.
- *BDRM_COND* is the condition of the Bedroom in a structure**. Types include:**

  Average, Good, Fair, Poor, Excellent.
- *HEAT_FUEL* is the type of fuel used for heating in a structure. Types include: Y for Self Contained, I for Indiv. Cntrl, N for None and C for

Common.

- *UNIT_N* is an estimation of the number of units within the property.
- *UNIT_N_ORIG* is a more conservative but less complete estimate of the number of units within the property, based solely on the number of units appearing in the SAM data (left blank for properties without unit-level information in the SAM data).

1.1.3. Geographical information

- *X* is the longitude of the property.
    - o This is derived from the City of Boston's *Parcels 2017* shapefile.
- *Y* is the latitude of the property.
    - o This is derived from the City of Boston's *Parcels 2017* shapefile.
- *LAND_PARCEL_ID* is the unique ID of the land parcel containing the property. For more information on this ID and the geography to which it corresponds, see BARI's Geographical Infrastructure 2017.
- *TLID* is the identifier for the segment of road containing the property.
    - o This is found by subsetting the 2013 TIGER lines street segments to only those that match the street name of the property, and then finding the one that is geographically closest to the property.
- *Blk_ID_10* is the 2010 Census Block ID number.
    - o This is found by spatially overlaying the longitude and latitude of the property onto the Census Blocks shapefile.
- *BG_ID_10* is the 2010 Census Group ID number.
- *CT_ID_10* is the 2010 Census Tract ID number.
- *Blk_ID_20,* the 2020 Census Block ID number.
    - o This is found by spatially overlaying the longitude and latitude of the property onto the Census Blocks shapefile.
- *BG_ID_20*, the 2020 Census Group ID number.
- *CT_ID_20*, the 2020 Census Tract ID number.

2. **Summary of Aggregate Measures** *(e.g., PADCross.CT10.YEAR.csv & corresponding shapefiles)*

Neighborhood-level datasets were created that describe aggregate features of neighborhood properties. Aggregations are made at the census tract or block group level for the 2010 and 2020 Census Geographies. CT10 indicates that it utilizes 2010 Census tract levels for aggregation. Aggregate measures are provided in both standard format (.csv) and as mappable shape files (.shp). Truncated variable names for the latter format are included in parentheses following the original variable names. Variable names for shapefiles are in parentheses.

2.1. **Description of Variables**

- *CT_ID_10* is the 2010 Census Tract ID number.
- *EE_SCORE.res (EESR)* is the average energy efficiency index (*EE_SCORE* above) for residential properties in the area. For more information on how the energy efficiency index is calculated, see the documentation above.
- *AV_LAND_PER_SF.res (ALPSFR)* is the average assessed value of a property's land, divided by the total property area in square feet (*TOTAL_PER_SF* above) for all residential properties in the area.
- *AV_LAND_PER_SF.nonres (ALPSFN)* is the average assessed value of a property's land, divided by the total property area in square feet (*TOTAL_PER_SF* above) for all non-residential properties in the area.
- *AV_BLDG_PER_SF.res (ABPSFR)* is the mean assessed value of a property's building, divided by its gross floor area in square feet (*AV_BLDG_PER_SF* above) for all residential properties in the area.
- *AV_BLDG_PER_SF.nonres (ABPSFN)* is the mean assessed value of a property's building, divided by its gross floor area in square feet (*AV_BLDG_PER_SF* above) for all non- residential properties in the area.
- *YR_BUILT_REMOD.res (YBRR)* is the mean value of the latest year remodeled or the year built for all residential properties in the area.
- *YR_BUILT_REMOD.nonres (YBRN)* is the mean value of the latest year remodeled or the year built for all non-residential properties in the area.
- *DEC_BUILT_REMOD.res (DBRR)* is the modal value of the latest decade it was remodeled in or the decade it was built in for all residential properties in the area.

- *DEC_BUILT_REMOD.nonres (DBRN)* is the modal value of the latest decade it was remodeled in or the decade it was built in for all non-residential properties in the area.
- *nbhdval (random)* represents the residuals extracted from a multilevel linear regression model that calculated the unique effect each census tract had on property values when controlling for the lot size, gross floor area, total living area, number of floors, age, and land-usage of buildings located within it. The top 4% of property values (*TOTAL_VALUE*) were dropped and the outcome variable was log-transformed to prevent any impacts on outliers on model results (final values were un-transformed for interpretability). Census tracts that have fewer records than 7 were dropped from the model. Values can be interpreted as the estimated value of a single-family residential building with average size and age in a census tract, though the interpretation of the positive (negative) effect on values is extensible to all land uses. Higher values are indicative of a more positive effect of the census tract and lower values are indicative of a negative effect. (Not included for census block groups owing to small within-geography sample sizes.)

3. Appendix

3.1. Appendix A: Property Types (PTYPE) Codes

PROPERTY OCCUPANCY CODES

Code	DESCRIPTION
	Multiple Use Property
010	CONDO MULTI-USE
012	VACANT LAND
013	RES/COMMERCIAL USE
019	RES/EXEMPT USE
025	RES/COMM MIXED USE
026	RES/COMM MIXED USE
027	RES/COMM MIXED USE
031	CON MULTI-USE
	Residential Property
101	SINGLE FAM DWELLING
102	RESIDENTIAL CONDO
103	MOBILE HOME
104	TWO-FAM DWELLING
105	THREE-FAM DWELLING
106	ADD'L RES IMPROVEMENT
107	OTHER RESIDENTIAL
108	CONDO PARKING
109	MULTIPLE BUILDINGS
110	CONDO STORAGE
	Apartment Property
111	APT 4-6 UNITS
112	APT 7-30 UNITS
113	APT 31-99 UNITS
114	APT 100+ UNITS
115	CO-OP APARTMENT
116	RES PARKING GARAGE
117	DAY CARE USE
118	ELDERLY HOME
119	RES PARKING LOT
120	APARTMENT
121	ROOMING HOUSE
122	ROOMING HOUSE
123	RESIDENCE HALL
124	DORMITORY
125	APT SUBSIDIZED HOUSING
126	APT SUBSIDIZED HOUSING
127	APT SUBSIDIZED HOUSING
128	CONDO APARTMENT
129	RECTORY, CONVENT
130	VACANT LAND
131	VACANT LAND
132	VACANT LAND
202	VACANT LAND
211	VACANT LAND
	Commercial Property
300	HOTEL/LODGING
301	MOTEL
302	INN, RESORT, B & B
303	PRIV CITY CLUB
304	NURSING HOME
305	HOSPITAL
306	LABORATORY
307	VETERINARY HOSPITAL
309	MEDICAL CLINIC
310	LAUNDRY FACILITY
311	LAUNDROMAT
312	SELF STORAGE WAREHOUSE
313	LUMBER YARD STORAGE
314	TRUCK TERMINAL
315	ANCILLARY STORAGE
316	WAREHOUSE
317	WAREHOUSE
318	COLD STORAGE WAREHOUSE

Code	DESCRIPTION
	Commercial Property Cont.
319	RETAIL/OFFICE
320	RETAIL/WHLSALE/SERVICE
321	RETAIL STORE
322	RETAIL STORE
323	SHOPPING CENTER
324	SUPERMARKET
325	RETAIL STORE
326	RESTAURANT/SERVICE
327	RESTAURANT/LOUNGE
328	FAST FOOD RESTAURANT
329	BAR/TAVERN/PUB
330	SHOWROOM
331	AUTO SUPPLY / Service
332	REPAIR GARAGE
333	GAS STATION
334	SERVICE PLAZA RETAIL
335	CAR WASH
336	PARKING GARAGE
337	PARKING LOT
338	SUBTERRANEAN GARAGE
339	PARKING GARAGE
340	OFFICE
341	BANK BUILDING
342	MEDICAL OFFICE
343	OFFICE 1-2 STORY
344	OFFICE 3-9 STORY
345	OFFICE BUILDING
346	OFFICE BUILDING
347	OFFICE BUILDING
348	OFFICE BUILDING
350	POSTAL SERVICE
351	TRAINING FACILITY
352	DAYCARE USE
353	SOCIAL CLUB
354	MAUSOLEUM
355	FUNERAL HOME
356	COMM CONDO
357	RETAIL CONDO
358	OFFICE
359	CONDO PARKING
360	MUSEUM, GALLERY
361	NIGHT CLUB
362	MOVIE THEATER
363	DRIVE-IN THEATER
364	STAGE THEATER
365	AUDITORIUM / SPORT CTR
366	FIELDHOUSE /TRACK
367	RACE TRACK
368	FAIRGROUND, PARK
369	RETAIL USE
370	BOWLING ALLEY
371	ARENA ICE SKATING
372	ARENA ROLLER SKATING
373	SWIMMING POOL
374	HEALTH CLUB
375	TENNIS/ RACQUET CLUB
376	ATHLETIC FACILITY
377	RECREATION BUILDING
378	SCHOOL (TAXABLE)
379	CHURCH, SYNAGOGUE
380	GOLF COURSE
381	TENNIS COURT(S)
382	STABLE, KENNEL
383	SWIMMING POOL
384	BOAT HOUSE / MARINA
385	TAXABLE BLDG ONLY

Code	DESCRIPTION
	Commercial Property Cont.
386	CAMPGROUND FACILITY
387	PAY PARKING LOT
388	AIR RIGHTS PROPERTY
389	BLDG: CHAP 61 B LAND
390	VACANT LAND
391	VACANT LAND
392	VACANT LAND
393	VACANT LAND
394	UTILITY BLDG /SHED
395	AIR FREIGHT TERMINAL
396	HANGAR: STORAGE, MAINT
397	PASSENGER TERMINAL
398	AIRPORT TERMINAL
399	GREENHOUSE
	Industrial Property
400	OLD MANUFACTURING
401	WAREHOUSE
402	OFFICE/INDUSTRIAL USE
403	MANUFACTURING
404	LIGHT MFG / R & D
405	INDUSTRIAL LOFT
406	COMPUTER EQUIP BLDG
407	MACHINE SHOP
408	NEWSPAPER PLANT
410	MINING, QUARRYING
412	METAL PROCESSING
413	AUTO SALVAGE YARD
414	FOOD PROCESS PLANT
415	BOTTLING PLANT
416	CANNERY
417	DAIRY
420	TANKS: ABOVE GROUND
421	TANKS: UNDER GROUND
422	ELEC POWER PLANT
423	ELEC TRANS R O W
424	ELEC SUBSTATION
425	GAS MANUFACTR PLANT
426	GAS PIPELINE R O W
427	GAS STORAGE
428	GAS PRESSURE STATION
430	TELEPH EXCHG STATION
431	TELEPH RELAY TOWER
432	CABLE TV FACILITY
433	RADIO /TV TRANS FACIL
435	RADIO TV STUDIO
436	STUDIO/REMOTE
437	TELECOM EQUIPMENT
438	TELECOM EQUIPMENT
439	BANK ATM
440	VACANT LAND
441	VACANT LAND
442	VACANT LAND
445	RAILROAD PROP
446	UTILITY: WATER SEWER
450	INDUSTRIAL CONDO
465	COM BILLBOARD
	Exempt Ownership
900	U.S. GOVERNMENT
901	COMMONWEALTH OF MASS
902	CITY OF BOSTON
903	BOST REDEVELOP AUTH
904	PRIV SCHOOL /COLLEGE
905	CHARITABLE ORGANIZTN
906	RELIGIOUS ORGANIZATN

Code	DESCRIPTION
	Exempt Ownership Cont.
907	EXEMPT 121A PROPERTY
908	BOSTON HOUSING AUTH
914	COMMONWEALTH OF MASS
923	COMMONWEALTH OF MASS
924	COMMONWEALTH OF MASS
925	COMMONWEALTH OF MASS
	Exempt Property Type
937	DORMITORY
941	AUDITORIUM / THEATER
942	CLASSROOM
943	SCIENCE LAB
944	DINING FACILITY/CAFETERIA
945	ACTIVITY/SOCIAL CENTER
946	RETAIL USE (EXEMPT)
947	ATHLETIC/SPORTS CTR
948	LAUNDRY FACILITY
949	STORAGE AREA
950	APARTMENT BLDG
951	DORMITORY
952	OFFICE/ADMINISTR BLDG
953	MEDICAL CLINIC
954	MEDICAL OFFICE
955	LABORATORY
956	MORGUE
957	MAINTENANCE/SERVICE AREA
958	REHAB/COVALES FACILITY
959	ASSISTED LIVING/ELDERLY
960	EXEMPT OFFICE CONDO
961	PARKING GARAGE
962	PARKING LOT
963	UTILITY/EQUIPMENT BUILDING
965	GOV'T OFFICE BLDG
966	MANUFACTURING
967	LOFT BUILDING/COMMERCIAL
968	WAREHOUSE
969	BOAT REPAIR/STORAGE
970	CHURCH, SYNAGOGUE
971	RECTORY, CONVENT
972	CORRECTIONAL BLDG
973	ADMINISTRATIVE BLDG
974	FIRE STATION
975	POLICE STATION
976	SCHOOL
977	COLLEGE
978	LIBRARY
979	HOSPITAL
980	WATER TREATMT PLANT
981	INCINERATION PLANT
982	ARMORY
983	CEMETERY
984	PUBLIC BEACH
985	OTHER EXEMPT BLDG
986	VACANT LAND
987	VACANT LAND
988	HOTEL/CONVENTION CENTER
989	PASSENGER TERMINAL
990	RETAIL CONDO
991	OFFICE CONDO
992	RESIDENTIAL CONDO
993	INDUSTRIAL CONDO
995	CONDO MAIN
999	PARTIAL EXEMPT ENTITY

3.2. **Appendix B: Codes for Land Use**

USE CODE	DESCRIPTION
A	Residential 7 or more units
AH	Agricultural/Horticultural
C	Commercial
CC	Commercial condominium
CD	Residential condominium unit
CL	Commercial land
CM	Condominium main (physical structure housing all related condo units with no assessed value)
CP	Condo parking
E	Tax-exempt
EA	Tax-exempt (121A)
I	Industrial
R1	Residential 1-family
R2	Residential 2-family
R3	Residential 3-family
R4	Residential 4 or more family
RC	Mixed use (res. and comm.)
RL	Residential land

Boston Area Research Initiative ● Northeastern & Harvard Universities
E-Mail: bari@northeastern.edu ● Web: www.bostonarearesearchinitiative.net

3.3. **Appendix C: Unit Imputation**

The City of Boston's Street and Address Management (SAM) system provides information about all properties in Boston, dividing properties into discrete units where possible. 89,440 properties (49.5%) in the assessment data set were separated into units. We tabulated the number of units for these, and then imputed the number of units for the remaining parcels using two methods.

For certain land uses, we assumed the number of units based on the definition of the land use itself. These included: Single-Family Residential, Two-Family Residential, and Three-Family Residential, which were set equal to the legal number of residential units associated with the designation (e.g., Two-Family Residential = 2 units); Commercial Lots, Residential Lots, and Condo Parking, which were set equal to zero units as they have no buildings on them; Condo Main, set equal to one unit as it is the lobby of a condo building. This accounted for 1,780 parcels that previously did not have unit data.

The remaining 5,553 properties (3%) were distributed across nine land uses: Residential 4, Apartments (residential with 7 or more units), Commercial Condominium, Commercial, Condominium, Exempt, Exempt (121A), Industrial and Residential-Commercial. We used regression-based imputation, leveraging data from assessments including the total assessed value for the property, land, and building (*AV_TOTAL*, *AV_LAND*, and *AV_BLDG*, respectively) and the total gross floor area and living area (*GROSS_AREA* and *LIVING_AREA*, respectively). The last two also had missing values, so we first imputed values for them based on the other variables. We then ran 9 separate generalized linear models, one for each land use, that used assessed value and area to predict the number of units for all cases for which this information was known. The parameters from these models were then used to estimate the number of units (rounded to the nearest whole number) for those properties for which this information on the number of units was missing.

The imputed values were assessed for outliers using Mahalanobis distance. There were no multivariate outliers. However, for 5 records, the prediction was above 700 units which is considered unrealistic after case-by-case assessment. The imputed values for these cases were removed (i.e., an NA value).