

DS 3000 project database information

To create a database, run the `boston_housing_import.py` function in the import database folder with an updated path and the years you wish to include (2004-2023)

The database has two tables. PID in the Assessment table is a Foreign Key of PID in the Building table. With all years included, the database takes up around 189mb.

Building table:

PID {Primary Key}: Unique 10-digit parcel number. First 2 digits [are the ward](#), digits 3 to 7 are the parcel, and digits 8 to 10 are the sub-parcel.

st_num: street number (360)

st_name: street name (huntington)

st_name_suf: street suffix (av)

zipcode: A **four-digit** code of the parcel's location. The first digit (0 in new england) was cut off as a leading zero— I'm working on fixing this bug but it seems to be uniform so we can always add it back in when doing analysis in the interim.

Assessment table:

PID {Partial Primary Key, Foreign key}: Unique 10-digit parcel number. First 2 digits [are the ward](#), digits 3 to 7 are the parcel, and digits 8 to 10 are the sub-parcel. Foreign key of PID in building table.

year {Partial Primary Key}: The year the assessment was completed in

LV: the value of the land under the building (USD)

BV: the value of the building itself (USD. Market value is LV+BV)

lu: the land use of the property. Key is below

land_sf: Amount of square feet of land

gross_area: sum of all areas on all floors of a building

Num_floor: number of floors.

Example query:

This query would return the average market value per year for Northeastern's zip code (2115).

```
SELECT year, AVG(market_value) AS average_market_value
FROM (
  SELECT a.year, (a.bv + a.lv) AS market_value
  FROM assessments a
  INNER JOIN buildings b ON a.PID = b.PID
  WHERE b.zipcode = '2115'
) AS subquery
GROUP BY year;
```

Land Use (lu) property key

Commercial/Industrial:

AH: Agricultural/Horticultural

C: Commercial

I: Industrial

CC: Commercial condominium

Residential:

R1: Residential 1-family

R2: Residential 2-family

R3: Residential 3-family

R4: Residential 4 or more family

CD: Residential condominium unit

A: Residential 7 or more units

RC: Mixed use (residential and commercial)

Other/no value/land:

CM: Condominium main (no value)

CL: Commercial land (can have building value, not really useful though)

RL: Residential land (can have building value, not really useful though)

CP: Condo parking (no land value)

EA: Tax-exempt (121A - redevelopment corporations)

E: Tax-exempt (governments, universities, etc)