CS591
Final Project Report
Xi Zhang

https://data.cityofboston.gov/Permitting/Property-Assessment-2015/yv8c-t43q

https://data.cityofboston.gov/dataset/Property-Assessment-2014/qz7u-kb7x

Those two dataset are property-related. The first dataset is about property tax that City of Boston collected in 2015. The second dataset is property tax that City of Boston collected in 2014.

Here is the dataset's field I will use for this project:

Data field name	Meaning	Value Type	Density	Missing Value	Example	Irregularity
ST_NUM	Street number	Number	Medium	Yes	32	No
ST_NAME	Street name	Text	High	No	Chestnut	Yes
ZIPCODE	Zip code	Number	Medium	No	02215	Yes
Full_address	Full address	Number+ text	High	No	1000 Centre ST 1	Yes
AV_TOTAL	Total property value	Number	Medium	No	414,300	Yes
LAND_SF	Land square feet	Number	Medium	Yes	4,096	Yes
YR_BUILT	Year built	Number	Medium	Yes	1920	Yes
R_TOTAL_RMS	Total rooms (House)	Number	Low	No	5	Yes
R_BDRMS	Bedroom number (House)	Number	Low	No	3	Yes
R_FULL_BTH	Full bathroom number(House)	Number	Low	No	2	Yes

R_HALF_BTH	Half bathroom number(House)	Number	Low	No	1	Yes
R_KITCH	Kitchen number (House)	number	Low	No	1	No
U_NUM_PARK	Parking lot number (Condo)	Number	Low	No	2	No
U_BDRMS	Bedroom number (Condo)	Number	Low	No	3	Yes
U_FULL_BTH	Full bathroom number(Condo)	Number	Low	No	2	Yes
U_HALF_BTH	Half bathroom number(Condo)	Number	Low	No	1	Yes
U_TOT_RMS	Total room number (Condo)	Number	Low	No	5	No

By using those two datasets I want to verify this hypothesis: In same area (same zip code) if a property has more bathroom, it will increase the property's value the most. Bathroom is the most dominant part than other parts like garage, bedroom's number, living room area and etc.

1st step: sort all the data by its zip code, then look at the house/condo's total room number, living room number and etc. In this circumstance, check if there is the more bathrooms it has, the more value it is. For example, one more bathroom increases the property value by 10%.

2nd step: sort all the data by its total room number, to check the location's factor to its value. For example, closer to downtown Boston by one area block (Find a downtown Boston area (one zip code) as standard), the area will increase by 5%.

.

.

.

Get results of two year data.

In this way, I will analysis all the factors (Living room's quantity, bathroom quantity, garage quantity, location, kitchen quantity and etc. (All list in the table above)) in the dataset will affect

to its property value. And verify whether my hypothesis that bathroom affects the property value the most is true. If it is not true, the result will show which factor will be the most. Beside, get the general property value change percentage as standard, to calibrate my results.