Boston Property Assessment Data Audit 2014-2017

Tyler Brown

2017-10-13

Here's the counts of what's blank or not.

Table 1: Count of blanks and NAs

variables	num_blanks	num_nans	field_types	percent_nan
X1	NA	NA	integer	NA
PID	NA	NA	character	NA
CM_ID	NA	376374	character	55.98
ST_NUM	NA	43726	character	6.50
ST_NAME	NA	NA	character	NA
ST_NAME_SUF	NA	8009	character	1.19
UNIT_NUM	NA	424631	character	63.16
ZIPCODE	NA	6	character	0.00
PTYPE	NA	6	integer	0.00
LU	NA	NA	character	NA
OWN_OCC	NA	NA	character	NA
OWNER	NA	NA	character	NA
OWNER_MAIL_ADDRESS	NA	340109	character	50.59
OWNER_MAIL_CS	NA	340109	character	50.59
OWNER_MAIL_ZIPCODE	NA	2	character	0.00
AV_LAND	NA	4	integer	0.00
AV_BLDG	NA	NA	integer	NA
AV_TOTAL	NA	NA	integer	NA
GROSS_TAX	NA	NA	integer	NA
LAND_SF	NA	12119	integer	1.80
YR_BUILT	NA	9773	integer	1.45
YR_REMOD	NA	114947	integer	17.10
GROSS_AREA	NA	12669	integer	1.88
LIVING_AREA	NA	12662	integer	1.88
NUM_FLOORS	NA	63602	numeric	9.46
STRUCTURE_CLASS	NA	336522	character	50.05
R_BLDG_STYL	NA	420720	character	62.58
R_ROOF_TYP	NA	420727	character	62.58
R_EXT_FIN	NA	420727	character	62.58
R_TOTAL_RMS	NA	319522	integer	47.53
R_BDRMS	NA	319514	integer	47.52
R_FULL_BTH	NA	319512	integer	47.52
R_HALF_BTH	NA	319512	integer	47.52
R_KITCH	NA	319512	integer	47.52
R_HEAT_TYP	NA	420717	character	62.58
R_AC	NA	420725	character	62.58
R_FPLACE	NA	319522	integer	47.53
S_NUM_BLDG	NA	497445	integer	73.99
S_BLDG_STYL	NA	636258	character	94.64
S_UNIT_RES	NA	480660	integer	71.49
S_UNIT_COM	NA	480654	integer	71.49
S_UNIT_RC	NA	480652	integer	71.49

variables	num_blanks	num_nans	field_types	percent_nan
S_EXT_FIN	NA	636258	character	94.64
U_BASE_FLOOR	NA	431059	integer	64.12
U_CORNER	NA	431061	character	64.12
U_ORIENT	NA	431052	character	64.11
U_TOT_RMS	NA	431067	integer	64.12
U_BDRMS	NA	431054	integer	64.11
U_FULL_BTH	NA	431052	integer	64.11
U_HALF_BTH	NA	431060	integer	64.12
U_KIT_TYPE	NA	552854	character	82.23
U_HEAT_TYP	NA	431056	character	64.12
U_AC	NA	431052	character	64.11
U_FPLACE	NA	431061	integer	64.12
Latitude	NA	238267	numeric	35.44
Longitude	NA	238267	numeric	35.44
U_NUM_PARK	NA	490465	character	72.95
Year	NA	NA	integer	NA
MAIL_ADDRESSEE	NA	481657	character	71.64
R_BTH_STYLE	NA	546632	character	81.31
R_BTH_STYLE2	NA	563727	character	83.85
R_BTH_STYLE3	NA	619901	character	92.20
R_KITCH_STYLE	NA	546632	character	81.31
R_KITCH_STYLE2	NA	606581	character	90.22
R_KITCH_STYLE3	NA	643226	character	95.67
R_EXT_CND	NA	546632	character	81.31
R_OVRALL_CND	NA	546632	character	81.31
R_INT_CND	NA	546632	character	81.31
R_INT_FIN	NA	546632	character	81.31
S_EXT_CND	NA	653720	character	97.23
U_BTH_STYLE	NA	550513	character	81.88
U_BTH_STYLE2	NA	626465	character	93.18
U_BTH_STYLE3	NA	660481	character	98.24
U_KITCH_TYPE	NA	550513	character	81.88
U_KITCH_STYLE	NA	550513	character	81.88
U_INT_FIN	NA	550513	character	81.88
U_INT_CND	NA	550513	character	81.88
U_VIEW	NA	550513	character	81.88
MAIL_ADDRESS	NA	332207	character	49.41
MAIL_CS	NA	332206	character	49.41
R_VIEW	NA	546632	character	81.31