Abstract

5000 individual data sets of information on real estate properties were investigated beginning with initial cleaning. This involved changing any values of “None” to “0” to maintain consistent data type, rounding down any non-whole values of bathrooms since it is impossible to have a fraction of a bathroom (depending on definition) so it was rounded down because if the given value has some accuracy it must have at least this many bathrooms. The entries with a sqrt\_ft value of 0 were removed because (assuming this represents the area of the property) a property cannot have an area of 0 making the entry erroneous.