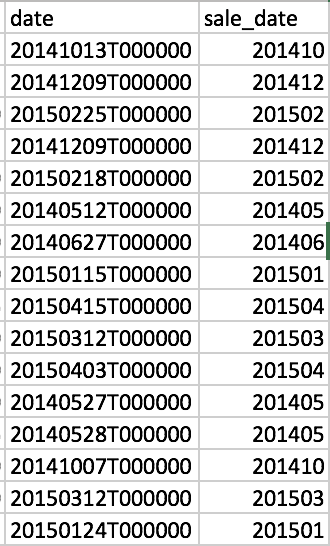
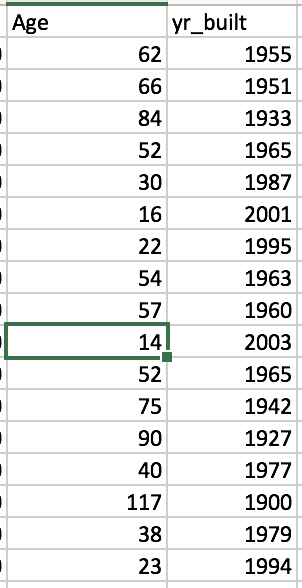
1. Modified the ‘date’ field, only keep the year and month. Ignore ‘date’ field



1. Create a ‘age’ field, use current year - yr\_built, in the model, only use ‘age’ field



1. Exclude ‘long’ and ‘lat’ field, which indicate the longitude and latitude of the house. Because all the house in the dataset are located in the same county. Given the geographical proximity, the difference in position can be ignored.
2. Exclude ‘zipcode’ field .
3. Create a dummy variable for ‘yr\_renovated’. ( yr\_renovated=0 if yr\_renovated ==0, =1 otherwise)
4. Create a new variable named ‘Expensive’ according to the ‘Price’ variable ( Expensive=1 if price>average\_price, =0 otherwise) .
5. Exclude ‘sqrt\_living’ field, which is consisted of ‘sqrt\_above’ and ‘sqrt\_basement’.

