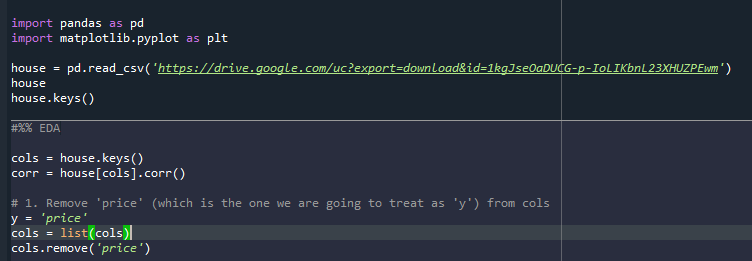
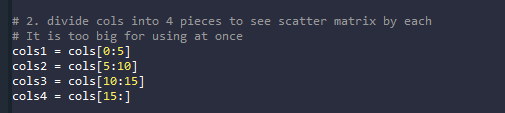
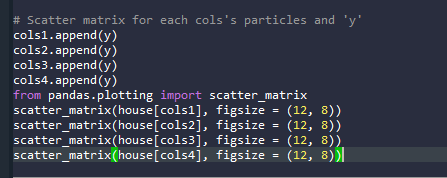
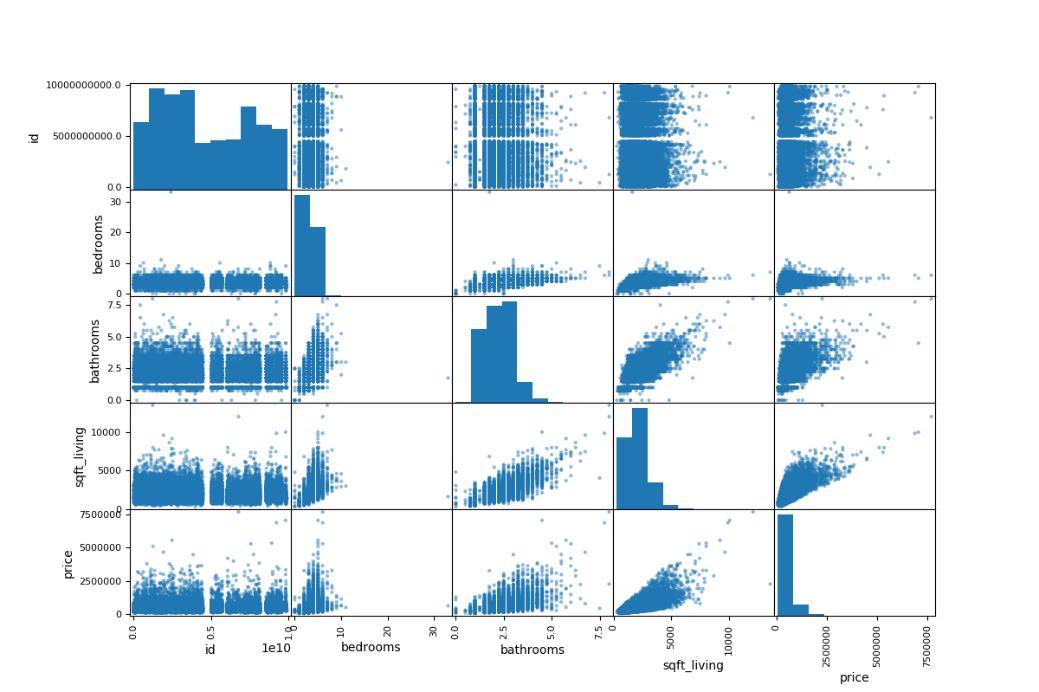
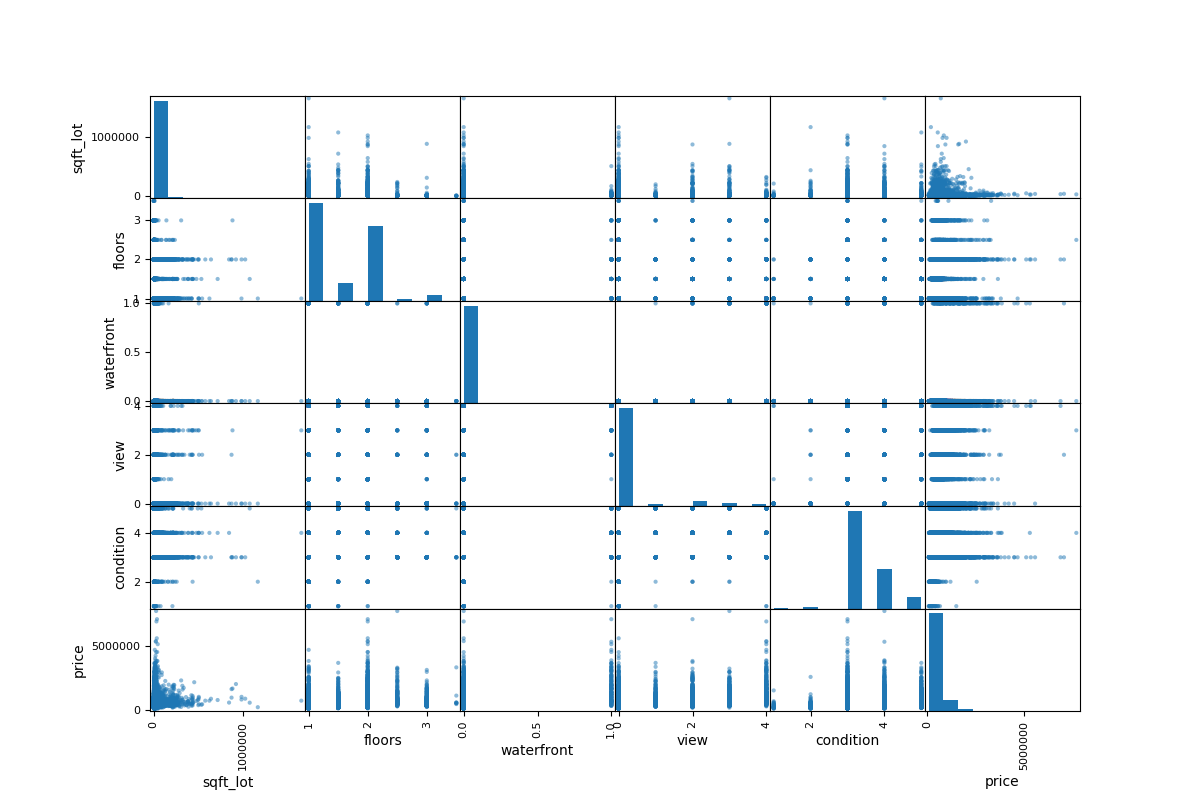
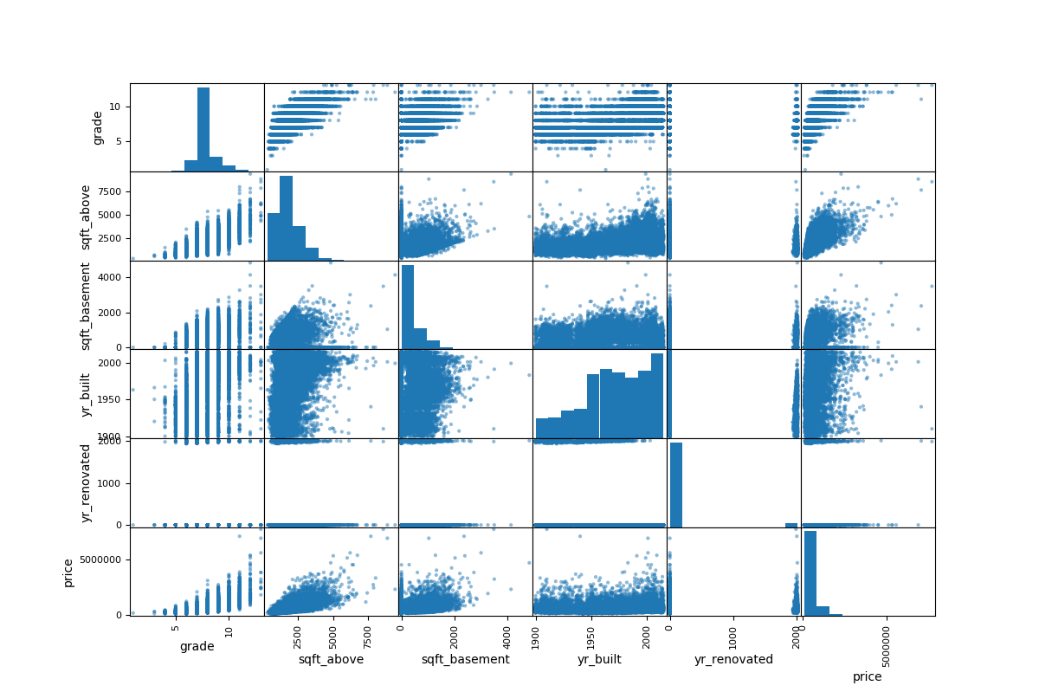
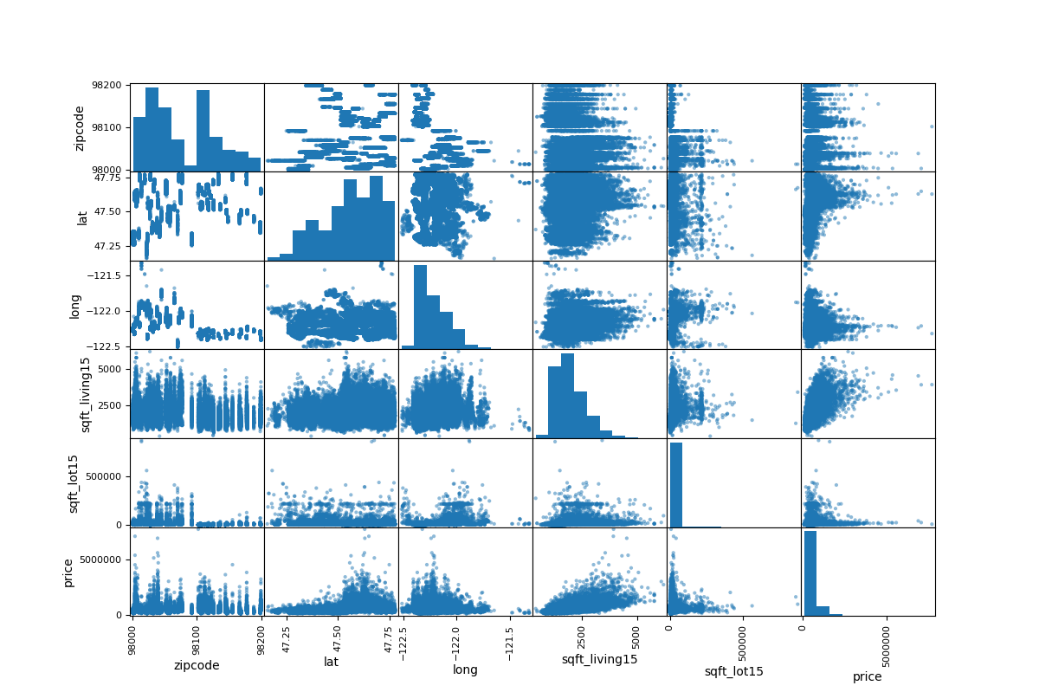
Assignment 02

15146314 Yang, Seunghyuck

1. EDA
   1. Remove 'price' (which is the one we are going to treat as 'y') from cols  
      
   2. Divide cols into 4 pieces to see scatter matrix by each  
      (Computer is not affordable with holding the process at once)  
      
   3. Scatter matrix for each cols's particles and 'y'  
        
      







Result for C:

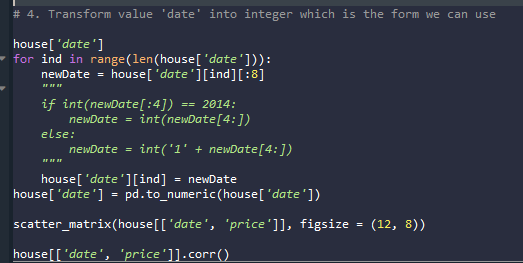
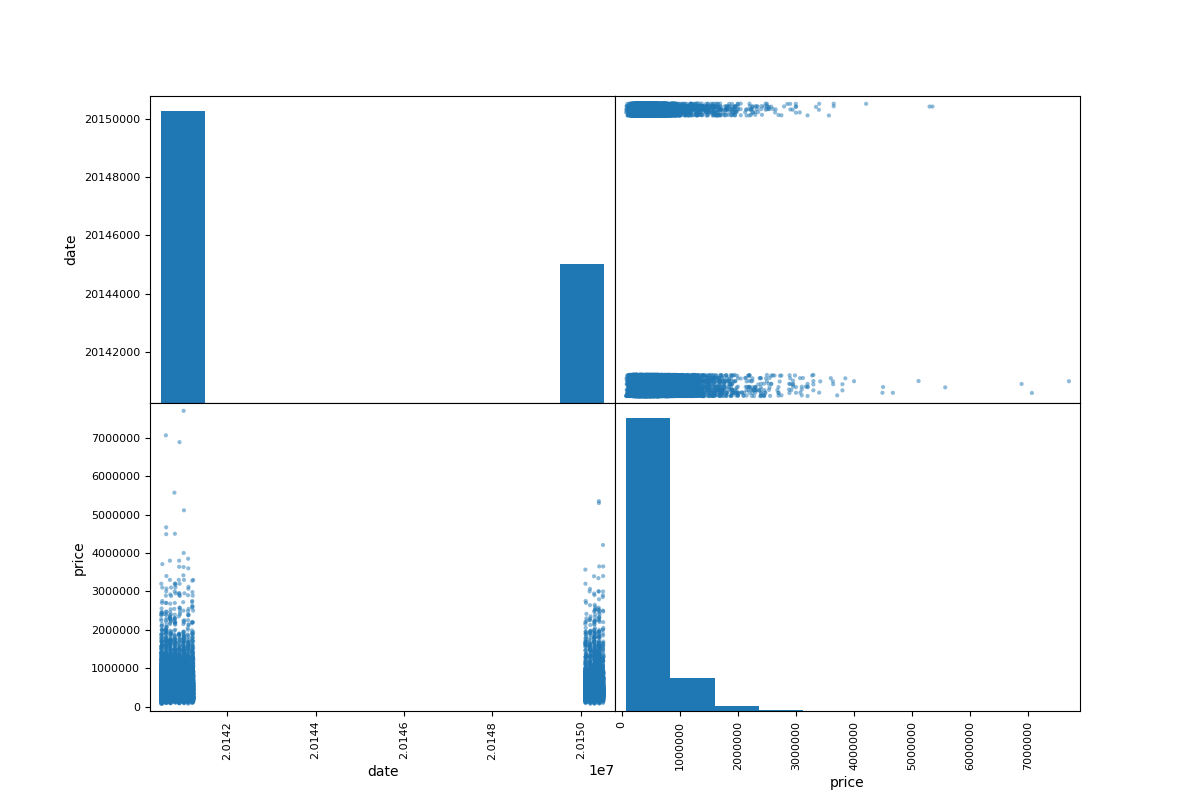
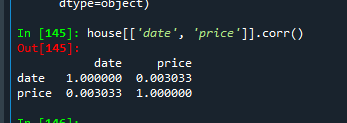
Distinguish by tier:

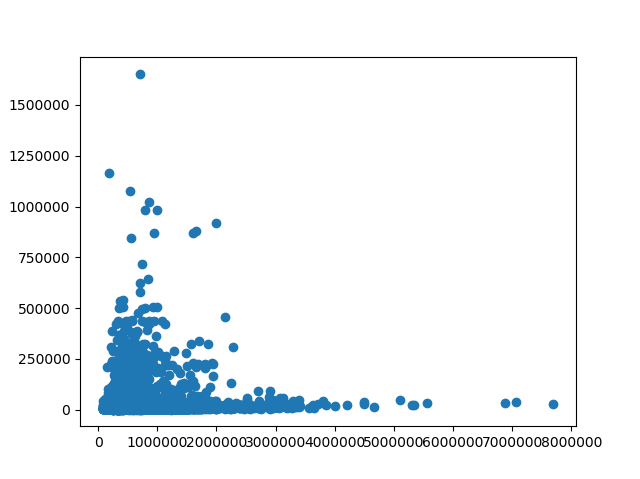
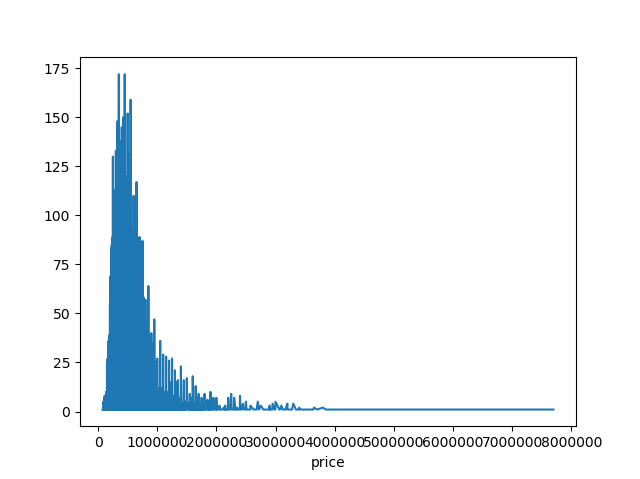
* + - 1. Bathrooms / sqft\_living / grade / sqft\_above / sqft\_basement / sqft\_living15 / sqft\_lot15(inverse) / sqft\_lot(inverse)
      2. Bedrooms / condition / lat / long
      3. Id / floors / waterfront / view / yr\_built / yr\_renovated / zipcode

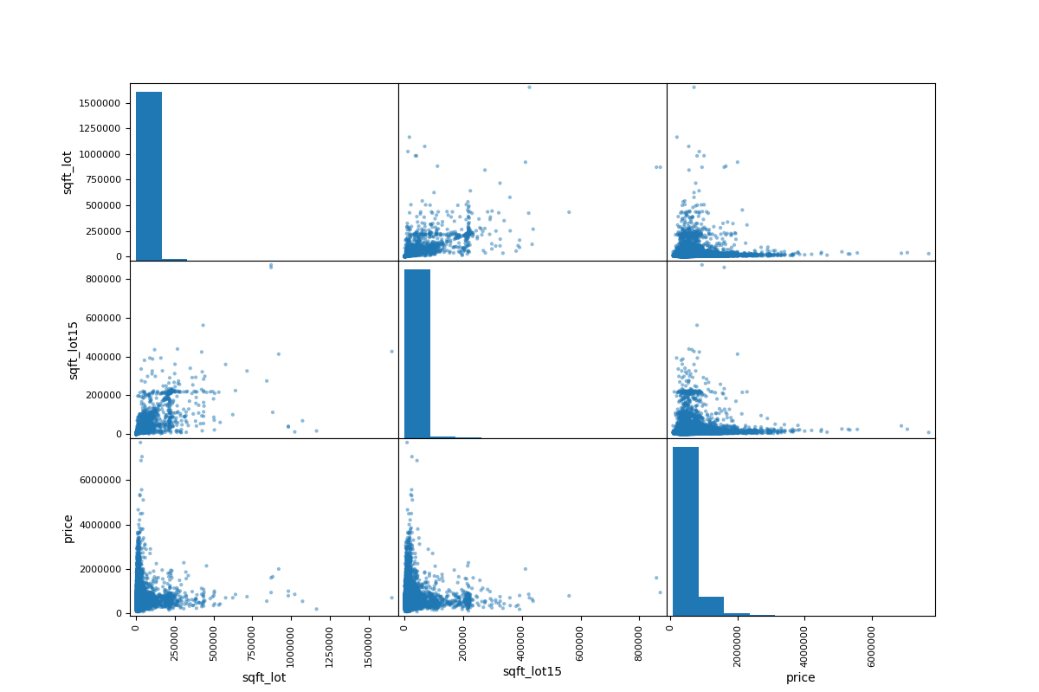
Criterion: Whether they seem correlated with price with bare eyes?

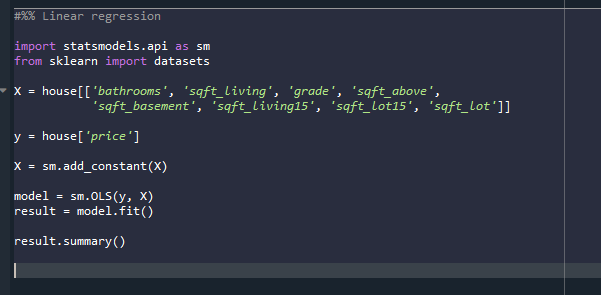
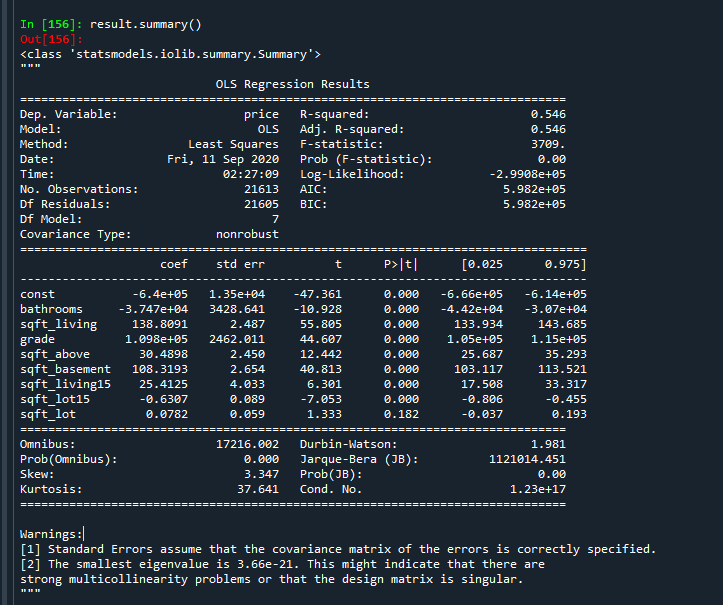
Problem: Since date value is not integer, it cannot be used for original statues.

Solution: Transform it to Integer.

* 1. Transform value 'date' into integer which is the form we can use  
       
       
     By the result, it is proved that date has no correlation with price, which is the result that indicates date to tier 3.

1. Preprocessing
   1. Transform data
      1. Sqft\_lot, sqft\_lot15 -> failed to transform them into form which has linear correlation with price.  
         I can’t decide whether they are related or not, so I decide to see both results.  
          



* 1. Create Train Dataset
     1. Variables: Bathrooms / sqft\_living / grade / sqft\_above / sqft\_basement / sqft\_living15 / (+ sqft\_lot15 / sqft\_lot)
     2. Train and see result  
        1. With sqft\_lot15 / sqft\_lot  
          
          
          
        2. Without sqft\_lot15 / sqft\_lot  
        