Housing Prices - Conclusions

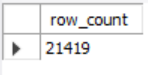
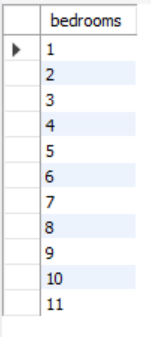
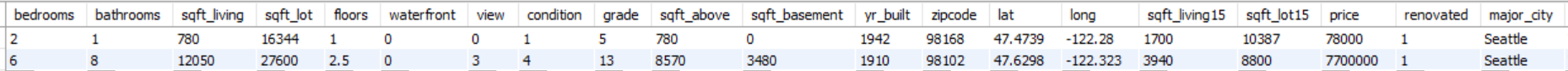
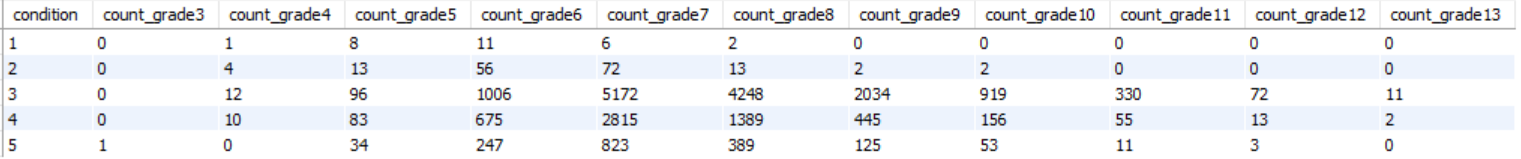
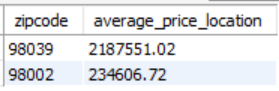
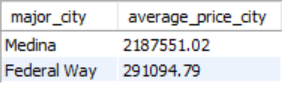
## Data Cleaning:

* Did the column headers naming convention change: No
* Number of NaNs: 0
* How many duplicated rows were there: 177 duplicates rows - we kept only the rows with the most recent date
* Rows that presented data inconsistencies were dropped, such as this property that had 33 bedrooms and less than 2 bathrooms
* Calculated columns were added:
  + ‘renovated’, that informs if the property was or not renovated and dropped column yr\_renovated due to missing data;
  + ‘major\_city’, calculated through the zipcode
  + ‘state’, calculated through the zipcode

Note: While calculating the ‘major\_city’ and ‘state’ we were able to conclude that all our database properties are located in the United States and in the Washington State.

## SQL Questions:

After the data cleaning activity we were able to conclude:

* How many properties does our database have: 21419 rows (properties)
  + 
* The properties can have:
  + from 1 to 11 bedrooms
    - 
  + from 0.5 to 8 bathrooms
    - 1 -- full bathroom (sink, toilet, tub AND shower)
    - 1.25 -- one full bathroom (sink, toilet, tub AND shower) and 1/4 bathroom (toilet)
    - 1.5 -- one full bathroom (sink, toilet, tub AND shower) and one half bathroom (toilet and sink)
    - 1.75 -- one full bathroom (sink, toilet, tub AND shower) and 3/4 bathroom (sink, toilet tub OR shower)
  + from 1 to 3.5 floors
  + a minimun grade of 3 and maximun grade of 13
* What are the cheapeast and most expensive house in our database: both are located in Seattle and cost '78000' and '7700000', respectevely:
  + 
* What is the average price of all properties: 541856
* How much does a house without a waterfront and with cost on average: '533263.10' and '1662524.18' (higher than the avergae of all properties), respectevely
* What is the correlation between grade and condition: columns grade and condition appear to have a positve correlation, as the higher values are centered or concentrated along the diagonal. When the grade increases the condition thends to increase and when the grade decreases the condition tends to decrease, and vice versa.
  + 
* The properties in our database are available in how many locations and cities: 70 different locations (zipcodes) and 24 different major cities:
  + 
  + 
* How many properties in our database have been renovated: 20917
  + 
* What are the zipcodes (locations) with the highest and lowest average of prices: '98039' and '98002', respectevely
  + 
* What are the major cities with the highest and lowest average of prices: Medina and Federal Way, respectevely
  + 
* How many houses were graded with the highest grade possible: 13
* What is the average price for the properties with the highest grade possible: '3710769'
* How many houses are there with the best condition possible: '1686'
* What is the average price for the properties with the best condition possible: '613274'
* How many properties have the lowest grading: 0
* How many houses are there with the worst condition possible: 28
* What is the average price for the properties with the worst condition possible: '349480'
* How many properties have the highest grade and best condition possible: 0
* What is the average condition for all properties: 3
* What is the percentage of properties with condition higher than the average: 34%
* What is the average cost for the properties with conditions higher than the average: '543236'
* How much higher than the total average is that average price in percentage: 0.3%