

- Reem Emad, Sec: 1, B.N.: 33, ID: 9190071
 - Dataset
 - Questions and Answers
 - 1. What is the most and least frequent type of lots' configurations?
 - 2. What is the relation between lot config and sale price? what is the average of sale price for each type of lot config?
 - 3. How much the sale price vary from different types of sale's condition?
 - 4. What is the sale condition and sale type that gets the highest price?
 - 5. How many houses have pools?
 - 6. What is the area for each pool quality?
 - 7. How much does the difference in having pools or not and quality of pools and the pools' area affect the house sale price?
 - 8. Is the dataset balanced or unbalanced?

Reem Emad, Sec: 1, B.N.: 33, ID: 9190071

Dataset

Ask a home buyer to describe their dream house, and they probably won't begin with the height of the basement ceiling or the proximity to an east-west railroad. But this dataset proves that much more influences price negotiations than the number of bedrooms or a white-picket fence. You can find this dataset on [Kaggle](#).

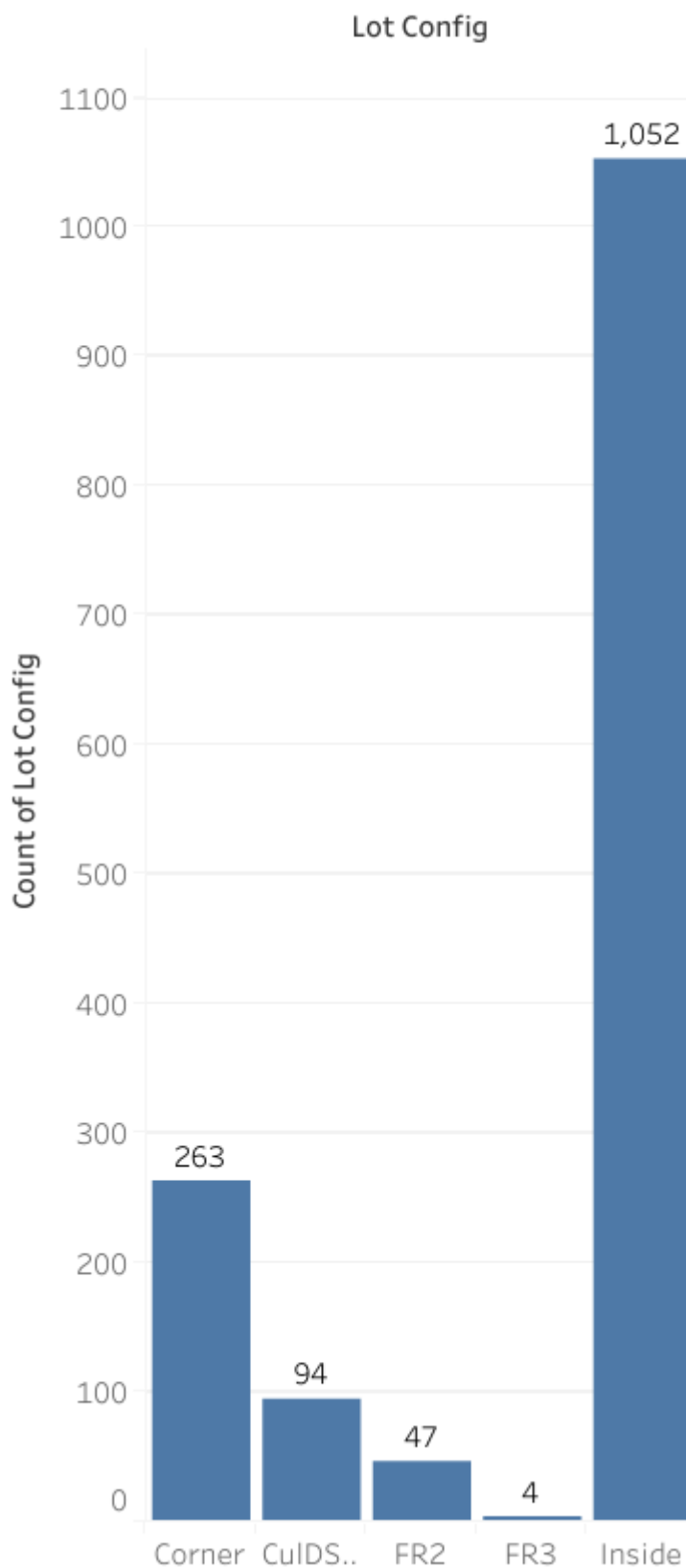
In the following questions, I'm trying to know how much some features affect each other and affects the sale price of the house with those features using different visualizing graphs and charts in Tableau.

Questions and Answers

1. What is the most and least frequent type of lots' configurations?

Generating the following histogram for the lot config column: [lot Config](#)

Sheet 1

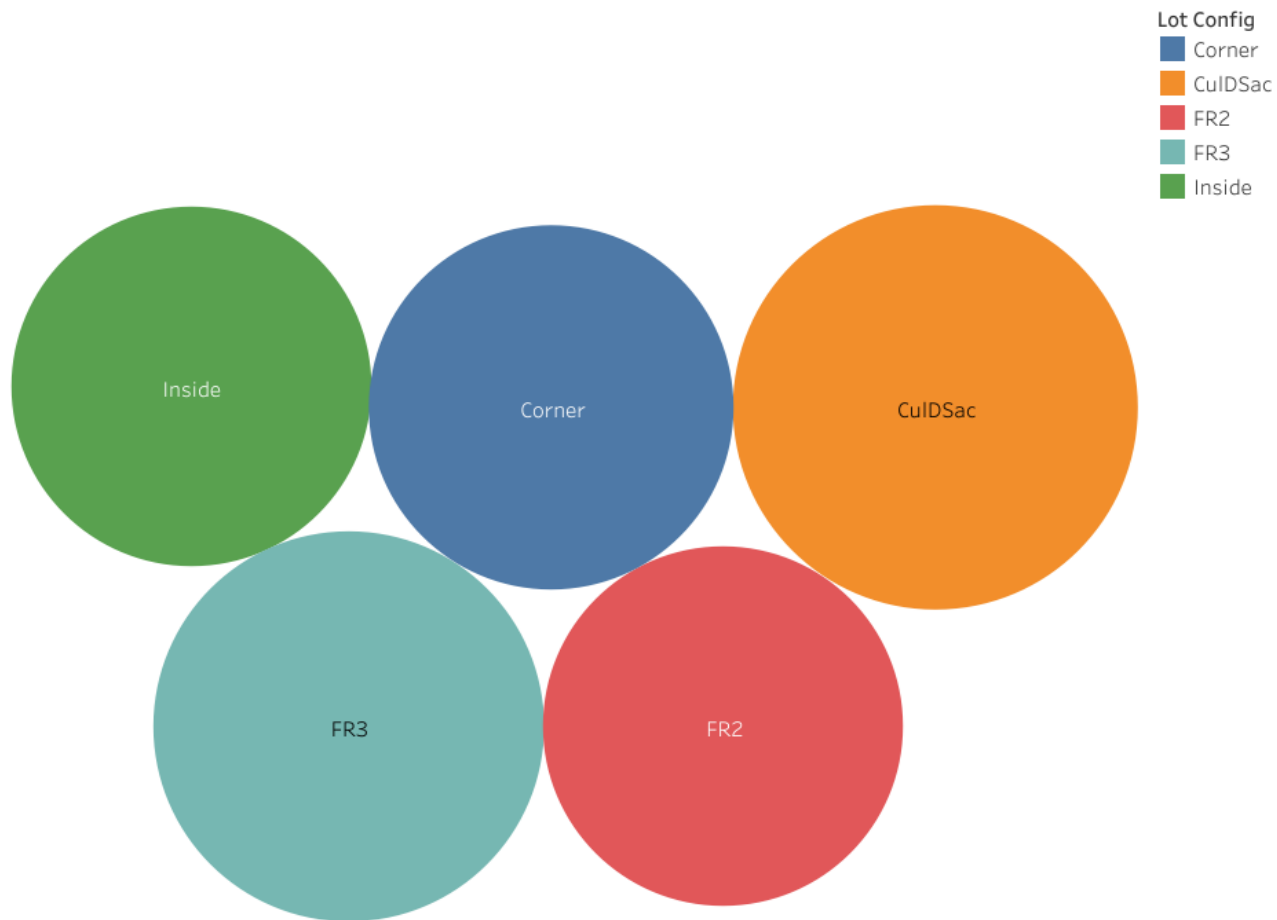


The height of the bar represents the number of occurrences of each type of lot config. It shows that the maximum frequent house lot config is "inside" with occurrences of number 1052, and the least frequent house lot config is "FR3" with occurrences of number 4.

2. What is the relation between lot config and sale price? what is the average of sale price for each type of lot config?

From the packed bubbles applied on lot config and sale price columns in this graph: [lot config & sale price](#)

Sheet 2

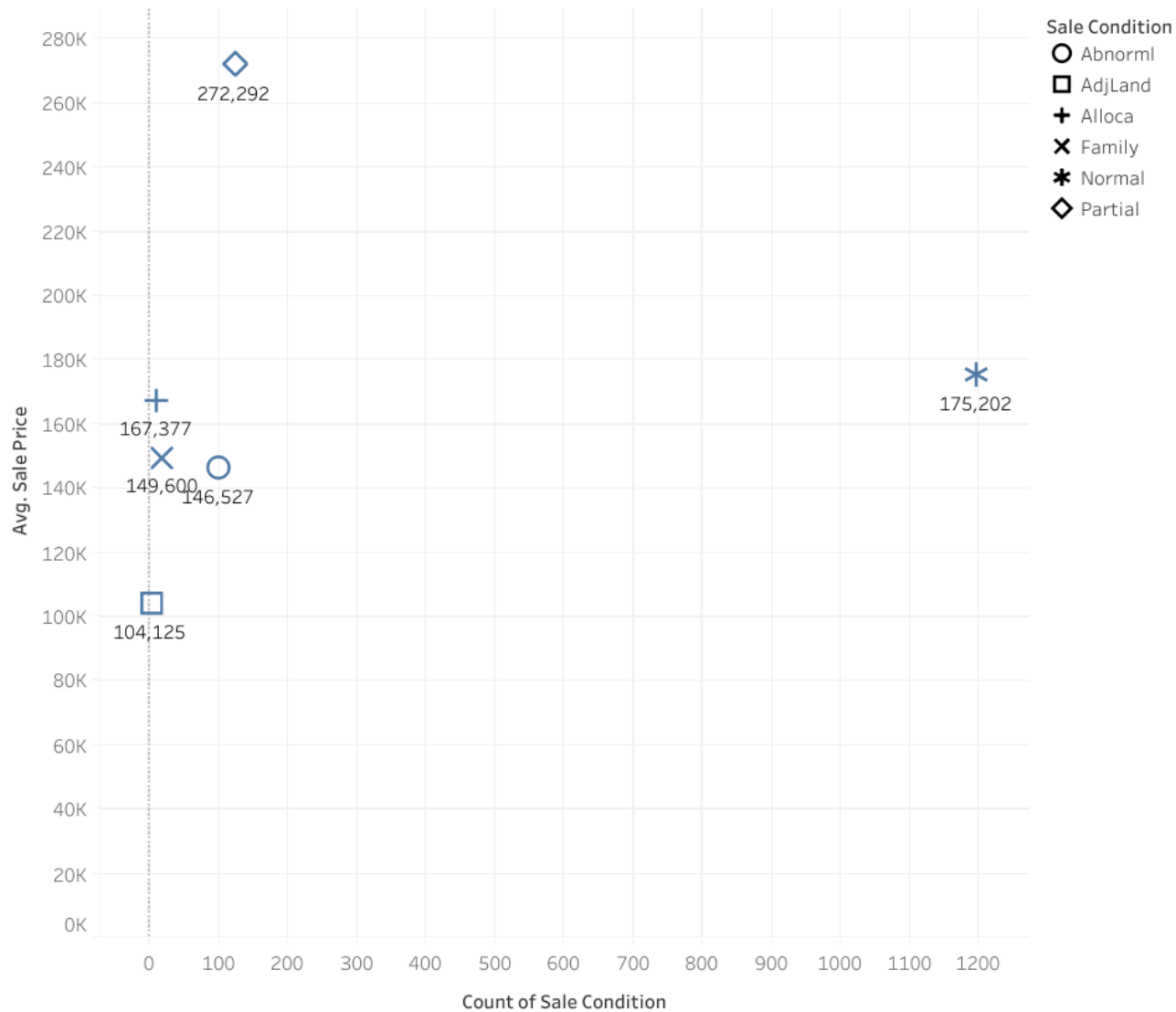


The circles' area represents the average sale price of each lot config type. It shows that all lot configuration houses have nearly same average sale price, this tells that the lot configuration doesn't have a great impact on the sale price because however the high difference in cuoccurrences of each type, they all have same average sale price. We need to combine this feature with other features for better correlation with sale price because there is no direct correlation between lot config and sale price of houses.

3. How much the sale price vary from different types of sale's condition?

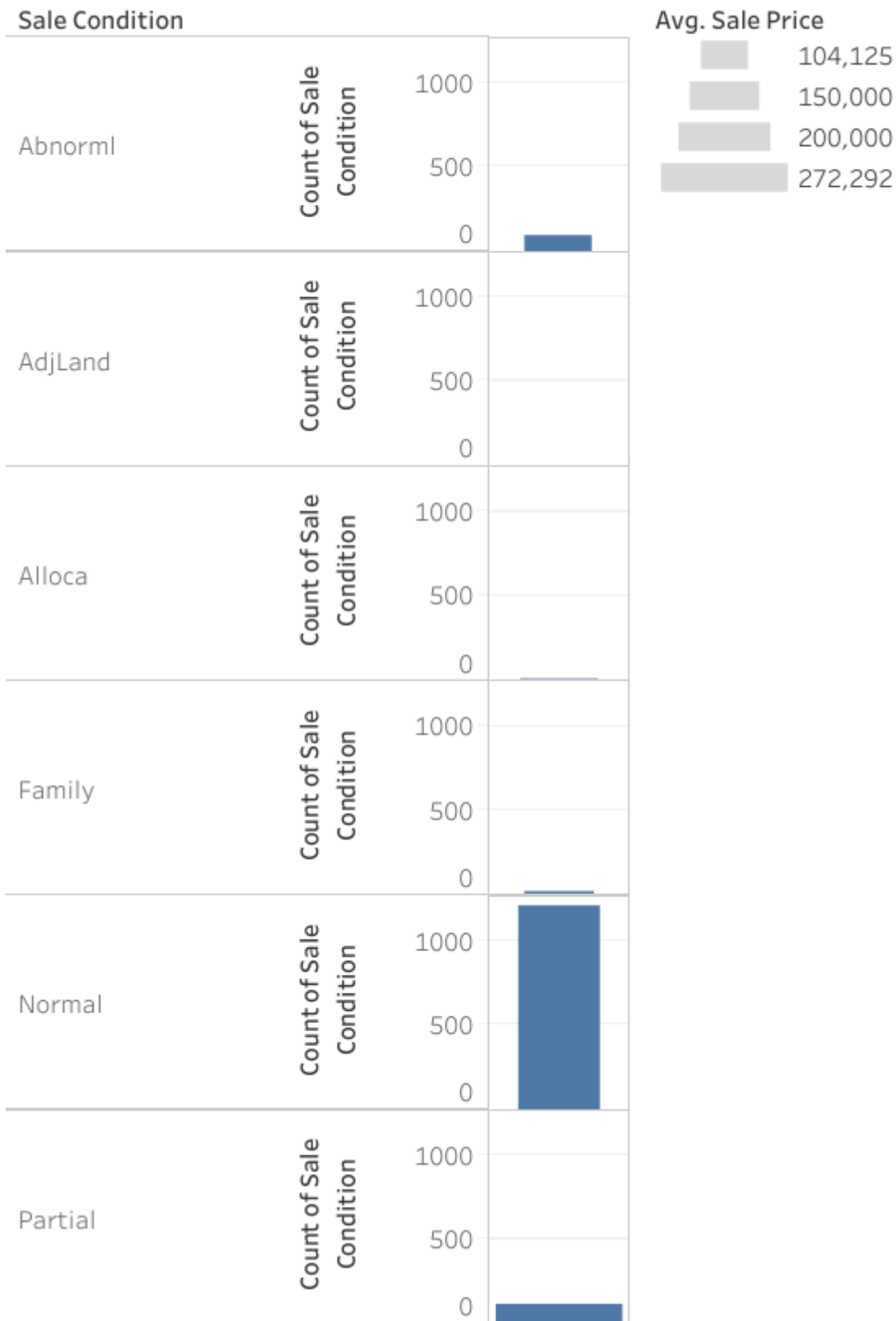
From the following scatter plot applied on sale condition and sale price columns: [sale condition & sale price](#)

Sheet 3



This is another way for same question: [sale condition & sale price](#)

Sheet 3 (another way)

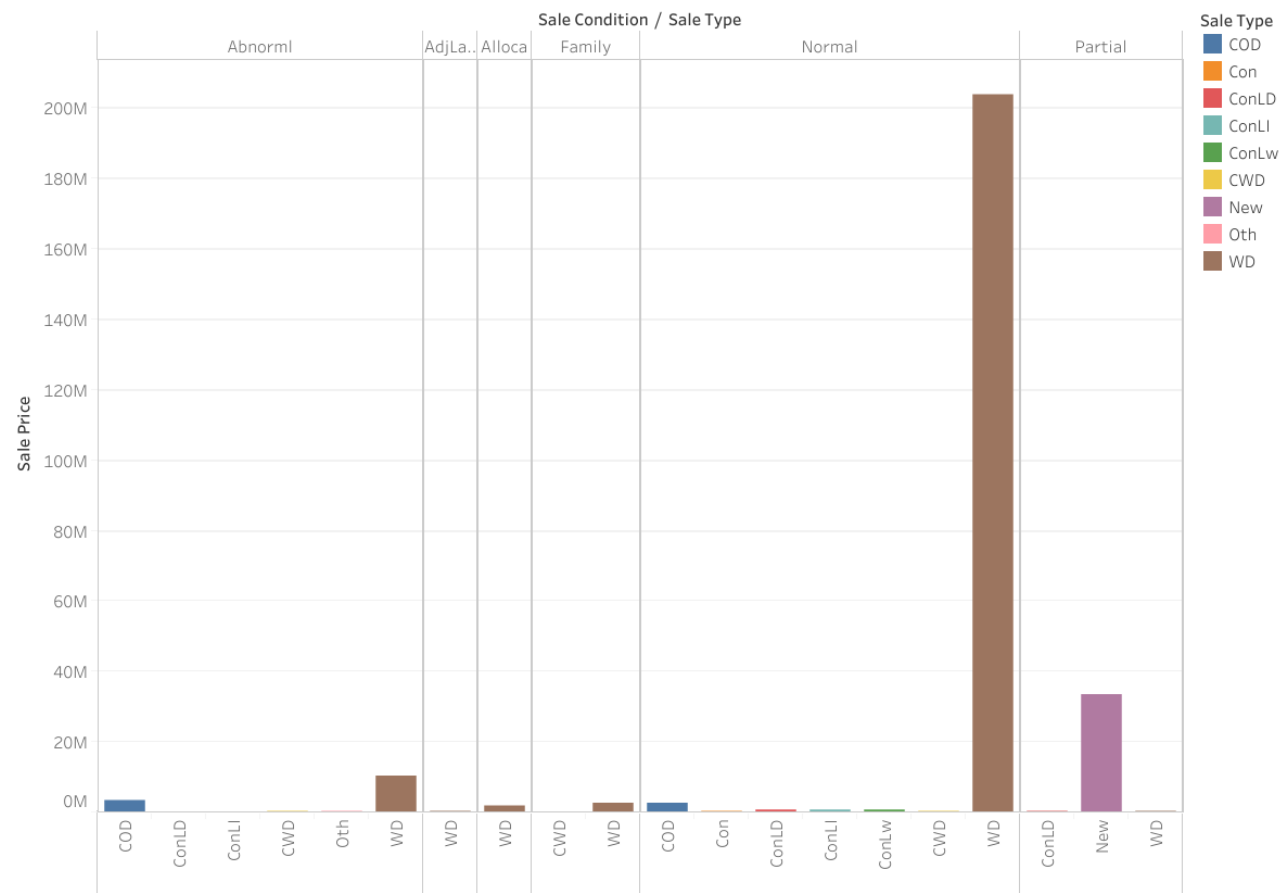


It shows that the most frequent sale condition is the normal one and it has the moderate average of sale price, this is meaningful. The highest average sale price is for the houses that are sold at partial complete of their building(partial) this is surprising but it may have other conditions that make the average price to be that high. The adjoining land purchase(adjLand) has both the least count of sale condition and least average sale price.

4. What is the sale condition and sale type that gets the highest price?

From the following side-by-side bars applied on sale condition, sale type and (sum)sale price columns: [sale cond, type and price](#)

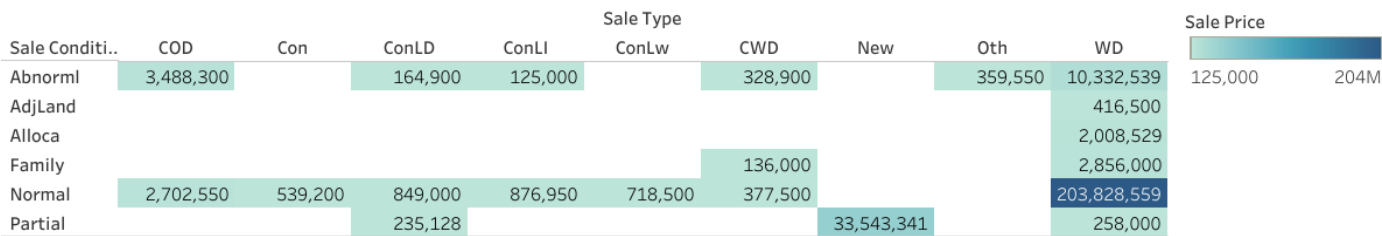
Sheet 8



The highest price is from the sale condition: normal, and sale type: WD (**Warranty Deed - Conventional**)

This is another way for same question: [sale cond, type and price](#)

Sheet 8 (2)

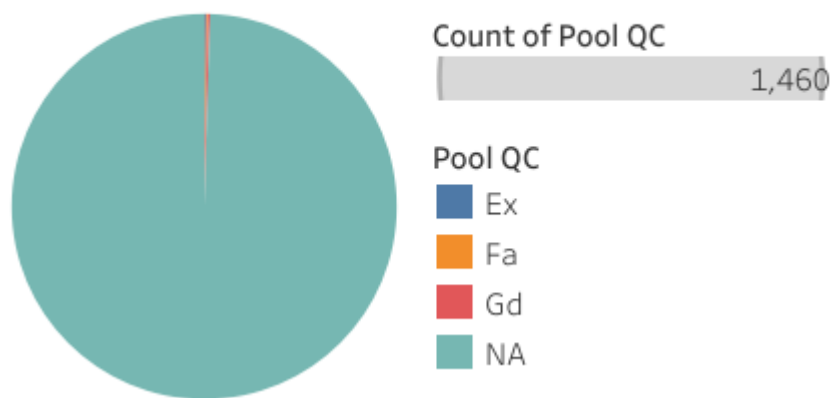


As the color gets darker, the value of price sale increases.

5. How many houses have pools?

From the following pie chart applied on poolQC(Pool quality): [Frequent PoolQC](#)

Sheet 5

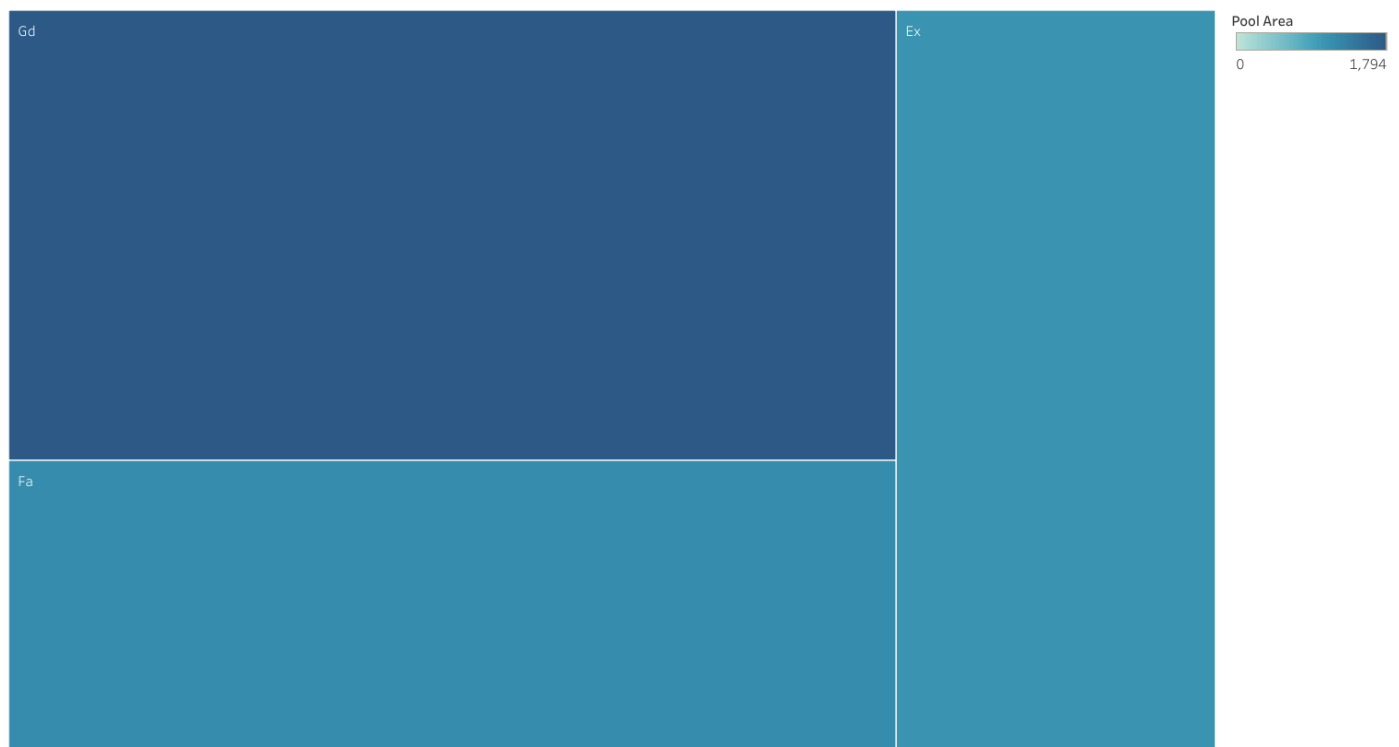


It shows that most of the houses (about 98%) don't have pools.

6. What is the area for each pool quality?

From this treemap applied on pool quality and pool area : [PoolQC & PoolArea](#)

Sheet 6



Each square represents the area of the labeled pool quality on average. It shows that the Pools are in the moderate(good)(gd) have biggest area and rest of pools are distributed semi-equal in pool area Excellent(Ex) and Fair(Fa). So pool area won't a big difference for distinguishing between pools' quality.

7. How much does the difference in having pools or not and quality of pools and the pools' area affect the house sale price?

From the following heat map applied on poolQC, PoolArea and sale price columns: [No pool, PoolQC, PoolArea & sale price](#)

Sheet 7

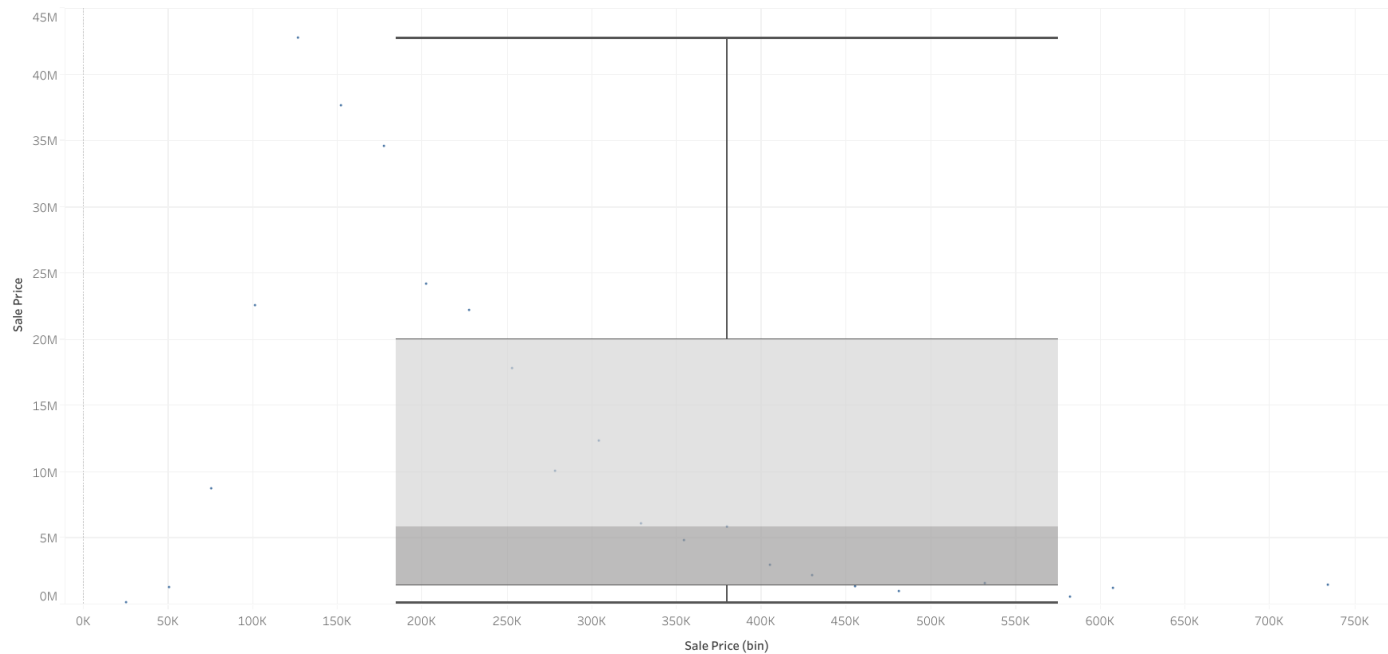


It shows that however the pool area of quality excellent doesn't have bigger area as the good pool but the price of houses having excellent pools is much higher than the sale price of houses with good quality pools. The houses without pools have least sale price on average, this is meaningful.

8. Is the dataset balanced or unbalanced?

From the box plot applied on sale price column: [sale price](#)

Sheet 10 (3)



We can see that the data is unbalanced, it is biased to the lower part(from 1M to 20M). It also shows that there is no outliers.

This Gantt shows the distribution of sale price in the dataset: [sale price](#)

