**Sqft\_living:**

In this case we analyzed that prices are uprising as with according to the Sqft\_living following the existence of number of houses as described by per Sqft\_living area. So from this analysis it is clear that as Sqft\_living is increasing the number of houses area decreasing so its states that the houses with large Sqft\_living area are less.

|  |  |  |
| --- | --- | --- |
| Cases | Sqft\_living | Number of house |
| 1 | 201 – 467, 468 - 734 | 12, 252 |
| 2 | 735 – 1000, 1001 - 1267 | 1230, 2242 |
| 3 | 1268 – 1534, 1535 - 1801 | 2926, 3021 |
| 4 | 1802 – 2068, 2069 - 2334 | 2594, 2379 |
| 5 | 2335 – 2601, 2602 - 2868 | 1904, 1392 |
| 6 | 2869 – 3135, 3136 - 3402 | 1069, 817 |
| 7 | 3403 – 3668, 3669 - 3935 | 529, 394 |
| 8 | 3936 - 4202 | 253 |
| 9 | 4203 – 4469, 4470 - 4736 | 202 - 122 |
| 10 | 4737 – 5002, 5003 - 5269 | 64- 47 |
| 11 | 5537 – 5803, 5804 - 6070 | 32 - 22 |
| 12 | 6071 – 6603 | 12 |
| 13 | 6604 – 6870, 7137 - 7404 | 6 - 9 |
| 14 | 7405 - 7937 | 4 |
| 15 | 7938 - 13274+ | 1,2,3 |

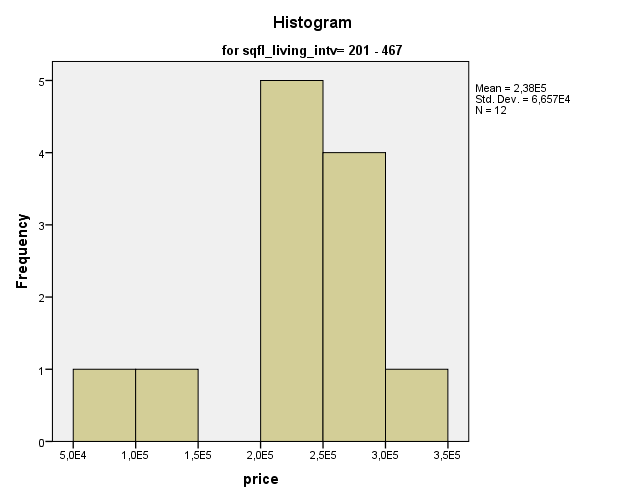
**Statistical analysis:**

Below in the table is given the information about Sqft\_living and number of house accordingly we collected the statistical analysis with the help of dependent factors that is prices.

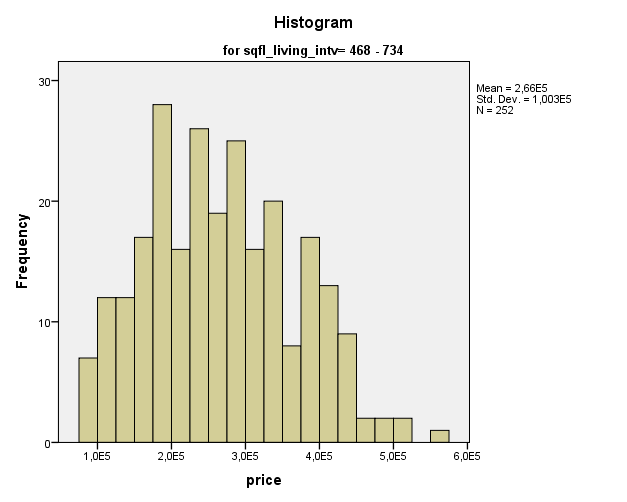
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 201-467 | 468 - 734 | 1001 -1267 | 2069-2334 | 3669-3935 | 5003 - 5269 | 7938 - 8204 |
| Mean | 2,37670833333E+005 | 2,65585706349E+005 | 3,42275008029E+005 | 5,18199588903E+005 | 1,02834012690E+006 | 1,84942361702E+006 | 4,58693333333E+006 |
| Median | 2,46000000000E+005 | 2,58750000000E+005 | 3,25000000000E+005 | 4,85000000000E+005 | 9,18602000000E+005 | 1,68000000000E+006 | 5,11080000000E+006 |
| Variance | 4431618390,152 | 10053167039,826 | 17953490582,742 | 40229538639,468 | 213244941190,101 | 618665439780,111 | 1256452213333,333 |
| Std Deviation | 6,657040175747E+004 | 1,002654827936E+005 | 1,339906361756E+005 | 2,005730257025E+005 | 4,617845181360E+005 | 7,865528842869E+005 | 1,120915792258E+006 |
| Minimum | 8,000000000E+004 | 7,500000000E+004 | 9,500000000E+004 | 1,942500000E+005 | 3,600000000E+005 | 4,520000000E+005 | 3,300000000E+006 |
| Maximum | 3,250000000E+005 | 5,700000000E+005 | 9,400000000E+005 | 1,565000000E+006 | 3,100000000E+006 | 3,650000000E+006 | 5,350000000E+006 |
| Range | 2,450000000E+005 | 4,950000000E+005 | 8,450000000E+005 | 1,370750000E+006 | 2,740000000E+006 | 3,198000000E+006 | 2,050000000E+006 |

**Graphical analysis:**

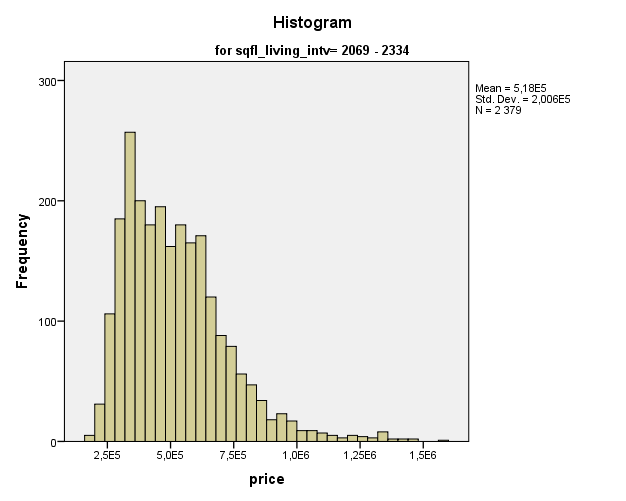
Below is the graphical distribution of Sqft\_living with the intervals as mentioned and it is distributed according to the prices following Sqft\_living. So on vertical we have the frequency distribution and prices on x-axis.



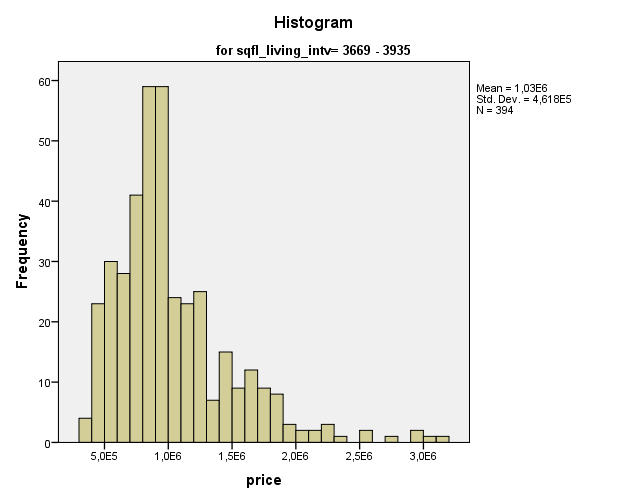
Below Histogram infers that as Sqft\_living is increasing so as the prices distribution is also increasing the houses with intervals of 468-734 so the houses are with high prices in between with large Sqft\_living area approx. 252 houses are in this range



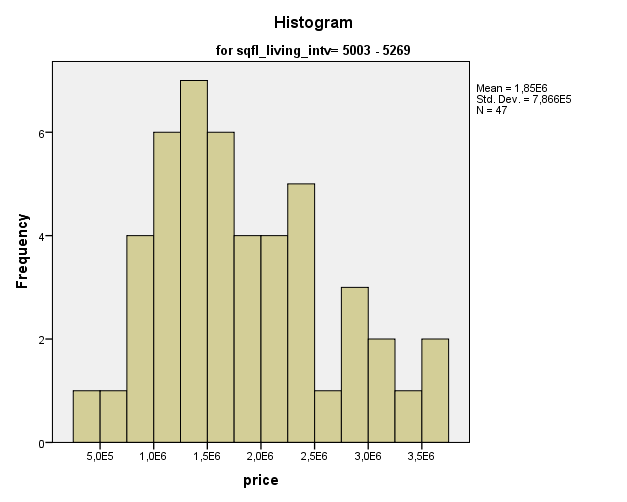
If we pay attention so as the Sqft\_living area is increasing so as prices from below histogram we can infer that frequency and number of houses are dependent on prices as distribution with Sqft\_living. And expensive houses are in range starting from 2200 000$ to 7 000 000$.



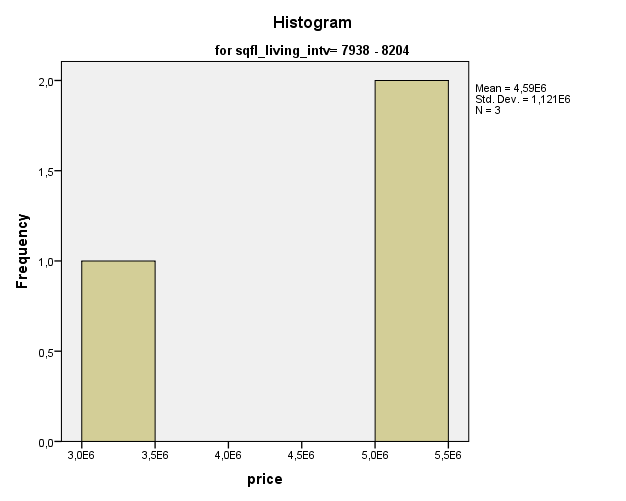
Below histogram illustrate that prices distribution from the intervals from 3669-3935, we can analyze that price distribution is increased at 2nd past where it is approx. 1 0000 000$ so frequency of expansive houses are at ~48-59%.



Below histogram illustrates that as intervals between Sqft\_living increasing the prices is also increasing but the number of houses are decreasing so it’s mean that large houses are in less number with high prices as median of 18 000 000$ as by frequency in between 6-7.

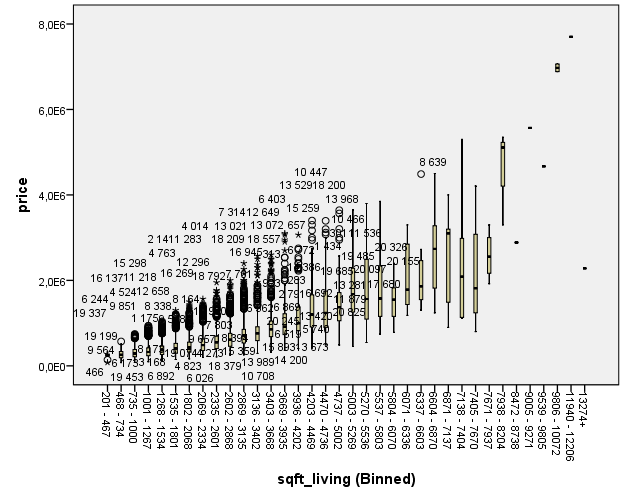


To narrow down our analysis in below histogram the prices of Sqft\_living area of 7938-8204 more than 55,000 000$ and frequency of houses are ~3.



**Outliers:**

With the help of outliers here we can easily analyze our standard errors in our experiment done on this factor analysis.



Above diagrams of (plotting) can illustrates that as Sqft\_living is increasing so as the prices but there are some outliers below which can show the error in the data or data anomaly. But they are not below zero. As Sqft\_living area is increasing the prices are increasing to as almost ~80000 000. So the conclusion is that houses with large Sqft\_living area are much more expensive and their frequency (number of houses) decreased.