Shahzeb khalid

**Descriptive Statistics: Numerical Measures**

**Home Equity Loans:**

The data set HMEQ reports characteristics and delinquency information for home equity loans. A home equity loan is a loan where the obligor uses the equity of his or her home as the underlying collateral.

**Descriptive Statistics:**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **LOAN** | **MORTDUE** | **VALUE** | **YOJ** | **DELINQ** | **CLAGE** | **NINQ** | **CLNO** | **DEBTINC** | **DEROG** |
| MEAN | 4944.04 | 67780.36 | 82567.98 | 7.56 | 0.36 | 151.86 | 0.57 | 18.79 | 33.51 | 0.23 |
| MEDIAN | 5100.0 | 68815.0 | 80027.0 | 5.0 | 0.0 | 138.65 | 0.0 | 17.0 | 34.45 | 0.0 |
| Stan.Dev | 1037.59 | 26825.01 | 32137.5 | 6.96 | 1.01 | 63.91 | 1.4 | 8.57 | 8.63 | 0.58 |
| Range | 4500.0 | 171975.0 | 230818.0 | 24.0 | 6.0 | 335.04 | 10.0 | 42.0 | 54.3 | 3.0 |
| Skewness | -1.17 | 0.62 | 1.43 | 0.99 | 3.25 | 0.75 | 4.96 | 0.78 | 0.6 | 2.66 |
| Variance | 1076592.88 | 719581365.77 | 1032819227.06 | 48.41 | 1.02 | 4084.08 | 1.95 | 73.51 | 74.39 | 0.34 |

             Low Quantile            High Quantile

Loan            4400.0                  5700.0

MORTDUE        52337.0                 79636.0

VALUE          63459.0                 97275.0

YOJ                2.0                    10.0

DEROG              0.0                     0.0

DELINQ             0.0                     0.0

CLAGE     108.07259081          86.42441602

NINQ               0.0                     1.0

CLNO              13.0                    24.0

DEBTINC   26.354889295         40.15052706

**Interpretation:**

**Value:** The mean of total value for home equity is $ 82567.98. This data shows the average value of a house that people are buying. On the other hand, the loans they are being approved is significantly lower than the value of a house. In order for people who are buying the houses around that value, they take in a long-term mortgage for the rest of the value. The average value that a delinquent credit score is approved of $67780.36.

**DEROG:** The median is the middle number in the data set. As for standard deviation indicates that the data points tend to be close to the mean. When we look at the median for it. This tells us the number of people who have the derogatory reports are zero, which is good, that they don’t have many loans. Also for standard deviation when you at derogatory reports, it shows how far apart the numbers are and it is indicating how the numbers are spread out from the average or the expected values. Since the low standard deviation indicates that the numbers are very close to the mean.

**YOJ:** It is used to represent the numbers of years spent at a current job. For us to calculate, the longevity of working year at present job, it would help us to use rage to find it. The range is used to tell us the difference between the largest and the smallest values. In this case, it would be able to tell us the basic details on a set of data spreading or variability which is 24 years at present job.

**Loan:** In order to have better representation, in skewness, it gives you a better measure on-base performance prediction. It is also a zero for its normal distribution and the symmetric data have a skewness of zero. Skewness values can be either positive or negative. As for us, it is useful to calculate loan and side of the scale it is high from, in our case, it is skewed towards negative. Which indicates that its skewed to negative. This tells us that the majority of the people who are getting the loan are on the right side of the scale. If we look back at mean where the higher number of people are getting the loans.

**NINQ:** The number of recent credit inquiries. By looking at variance it will help determine, how far the set of numbers is spread out. It can also describe the amount of random variable differs from its expected mean. In our case when we look at the variance which is 1.95. By compare we can see that the data is not as far from the set number of the mean, because variance treats all deviations from the mean the same no matter the direction.

**CLNO:**  This means the number of credit lines. I will be using quartiles in order to have a better look. Quartiles in statistics are values that divide your data into quarters. They are dived into four segments where the numbers fall on the number line. This will help to determine where most of the data lie. When the data is separated by segments, this will range from the lower half, median, and upper half. We will be observing that the 25% of data we can see is 13.0 and the rest of 75% of data is 24.0. This shows us that most of our data lie within the upper half. Hence there is a higher number of people who are asking for a credit line.

                          OUTLIERS

Loan Outlier          [1700.0, 1800.0]

Mortdue Outlier       [179725.0]

Value Outlier         [251962.0]

YOJ Outlier           []

DEROG Outlier         [3.0, 2.0, 2.0, 2.0, 2.0, 2.0, 2.0]

DELINQ Outlier        [6.0, 4.0, 4.0]

CLAGE Outlier         [350.43613755]

NINQ Outlier          [10.0, 9.0]

CLNO Outlier          [47.0, 47.0]

DEBTINC Outlier       [72.670433313]

**DELINQ: ­­­­­**The number of delinquent credit lines, this can occur when people don’t pay off the credit when they make the purchase. Using an outlier model, it will help with the observation with the numbers that lies outside the overall pattern of distribution. When we observe the data we have 6.0, 4.0, 4.0, we see that the majority of it outlier. Usually, the presence of an outlier indicates some sort of problem. This can be a case which does not fit the model under study or an error in measurement.

Correlation: CLAGE & DEBTINC

 [[ 1.         -0.12362713]

 [-0.12362713  1.        ]]

**CLAGE & DEBTINC:** I have chosen two variables as correlation, which are Clage: Age of oldest credit line in months and DEBTINC: Debt-to-income ratio. Correlation usually calculates the strength of the relationship between the relative movements of the two variables. The rage of it is bounded by 1.0 value basis or -1.0 or 1.0. also a value of -1.0 to 1.0 means there is a perfect negative relationship between two variable. It is greater then it is the other way around if 1.0 to -1.0. When we observe the data we have two negative correlation for both Clage and Debtinc.

Since what we have is neither -1.0 to 1.0 or 1.0 to -1.0, we can say that we have no correlation**.**