**Terro's Real estate Agency Business Report**

**1.Comprehensive Descriptive Statistics Overview**

**Crime Rate (per capita crime rate by town)**

The mean crime rate of 4.87 indicates the central tendency of the data. It represents the average crime rate in the areas under consideration.

The standard deviation of 2.92 measures the spread or dispersion of the crime rate values around the mean. A higher standard deviation suggests greater variability in crime rates across the areas.

The skewness of 0.021 suggests a nearly symmetrical distribution, but the slight right skew indicates that there may be a few areas with higher crime rates.

**Age (proportion of houses built prior to 1940 in percentage terms)**

the proportion of houses built prior to 1940 in the dataset has a mean of 68.57%, with a standard deviation of 28.14%, suggesting some variability around the mean. The skewness of -0.59 indicates a slight leftward skew in the distribution. The range of 97.1% shows the spread of age percentages, ranging from the minimum to the maximum values in the dataset.

**INDUSTRY (proportion of non-retail business acres per town in percentage terms)**

The average town has approximately 11.13% of its acreage devoted to non-retail business. The variability around this average is moderate, with a standard deviation of 6.86%. The distribution is slightly skewed to the right, indicating the presence of towns with higher percentages of non-retail business acres.

**NOX (nitric oxides concentration parts per 10 million)**

The average concentration of nitric oxides is 0.554 parts per 10 million. The variability around this average is relatively low, with a standard deviation of 0.115. The distribution is slightly skewed to the right, indicating the presence of areas with higher NOX concentrations.

**AVG ROOM (average number of rooms per house)**

The average number of rooms per house is 6.28. The variability around this average is relatively low, with a standard deviation of 0.70.

The positive skewness of 0.40 indicates a rightward skew in the distribution. This suggests that there may be a tail on the right side of the distribution, with a few houses having a higher number of rooms.

**DISTANCE (from highway in miles)**

The average distance of houses from the highway is 9.5 miles. The variability around this average is relatively high, with a standard deviation of 8.70 miles. The distribution is noticeably skewed to the right, indicating the presence of houses with greater distances from the highway.

**TAX (full-value property-tax rate per $10,000)**

The average property-tax rate per $10,000 is 408.23. The variability around this average is relatively high, with a standard deviation of 168.5. The distribution is slightly skewed to the right, indicating the presence of areas with higher property-tax rates. The range of 524 shows the spread of property-tax rates, ranging from the minimum to the maximum values in the dataset

**PTRATIO (pupil-teacher ratio by town)**

The average pupil-teacher ratio by town is 18.45. The variability around this average is relatively low, with a standard deviation of 2.164. The distribution is slightly skewed to the left, indicating the presence of towns with lower pupil-teacher ratios.

**LSTAT (% lower status of the population)**

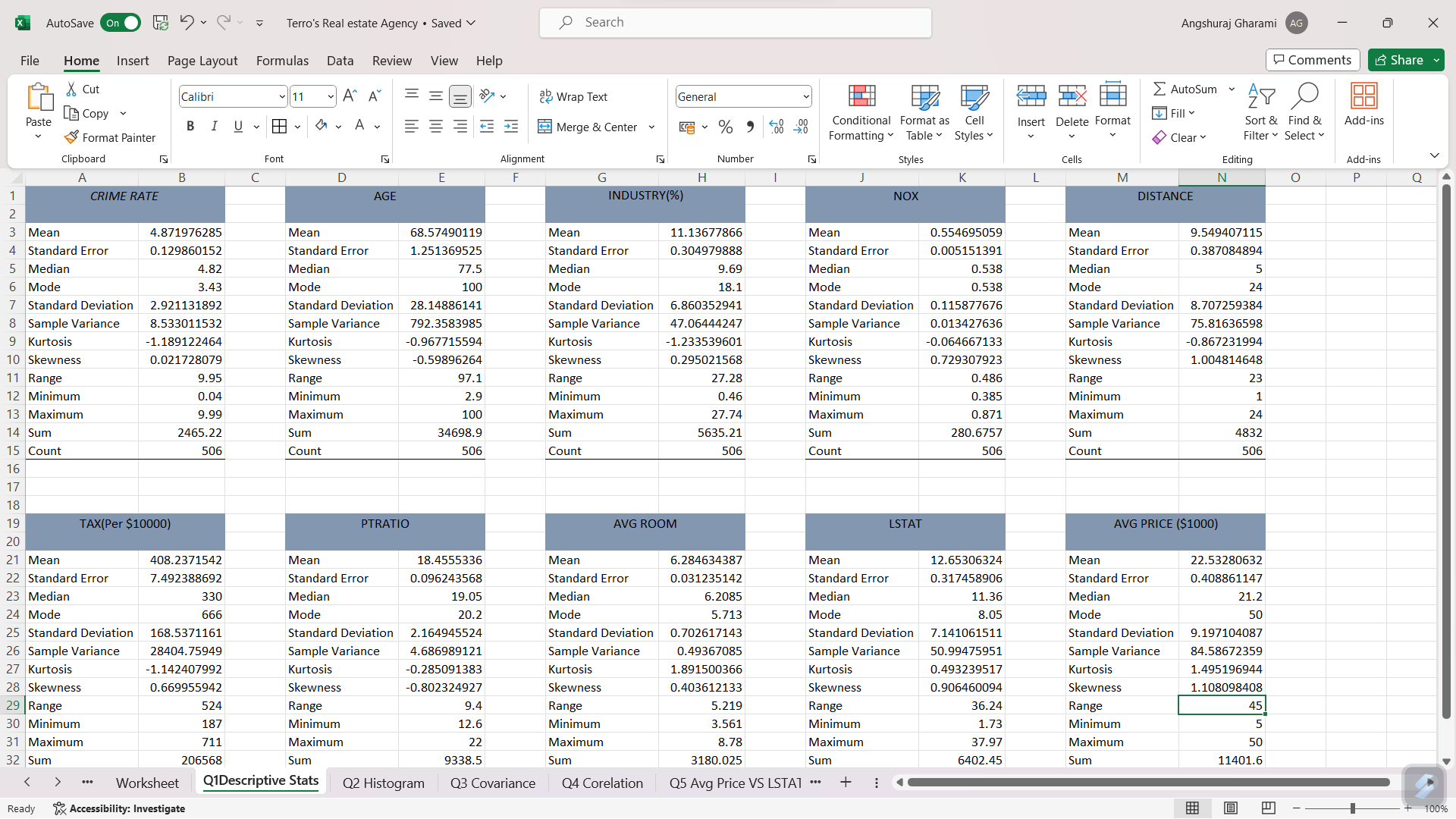
The average % lower status of the population is 12.63. The variability around this average is relatively high, with a standard deviation of 7.14. The distribution is noticeably skewed to the right, indicating the presence of areas with higher percentages of lower-status population. The range of 36.24 shows the spread of % lower status, ranging from the minimum to the maximum values in the dataset.

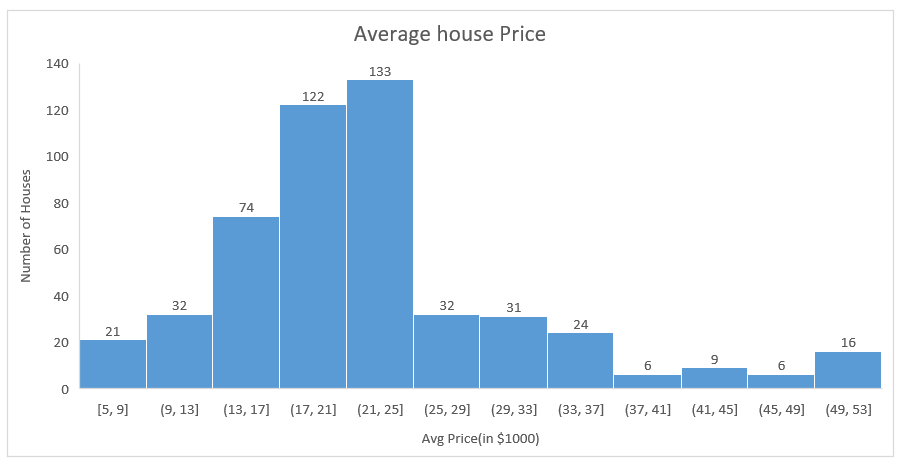
**AVG PRICE (Average value of houses in $1000)**

The mean average value of houses is $22.53 (in $1000). The standard deviation of $9.19K measures the spread or variability of the average house values. With a standard deviation of $9.19, the house values exhibit a notable degree of variability around the mean.

The positive skewness of 1.10 indicates a rightward skew in the distribution. This suggests that there may be a tail on the right side of the distribution, with some areas having higher average house values. The range of $45K provides the difference between the maximum and minimum values in the dataset. This indicates the full extent of variability in average house values.

More detailed statistics and information can be found in the Excel workbook, specifically in the sheet labelled 'Descriptive Statistics’.



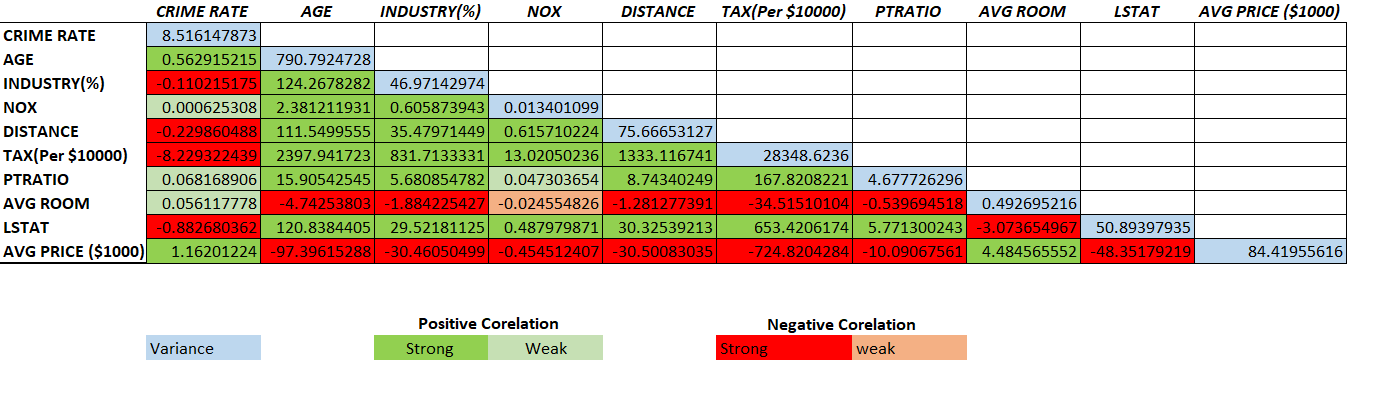


Upon analysing the histogram of average house prices, it is noteworthy that the highest number of houses, out of the total 506, falls within the price range of 21-25K.

Moreover 50% of the houses in the locality have average price ranging from 17K to 25K.

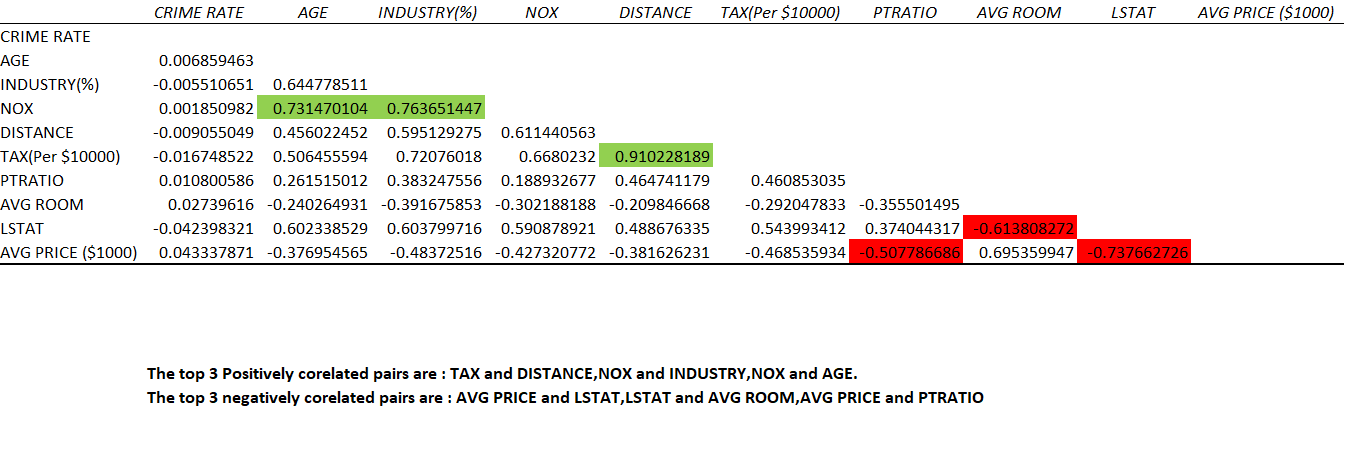
The histogram analysis reveals that approximately 6% of the houses in the locality are classified as expensive, with prices exceeding 40K

**Covariance Matrix**

****

After creating the covariance matrix, it is observed that variables with ‘Green fill’ exhibit a positive covariance, suggesting a tendency to increase or decrease together, while variables with ‘Red fill’ show a negative covariance, indicating an inverse relationship. The magnitude of the covariance values further emphasizes the strength of these associations. Higher magnitudes indicate stronger relationship and vice-versa.

As covariance values range from -∞ to +∞, making it hard to interpret, a more practical approach involves computing the correlation between variables. Correlation provides a standardized measure between -1 and 1, offering a clearer understanding of the strength and direction of relationships among the variables in the dataset.

**Corelation Matrix**

After constructing the correlation matrix, it is evident that certain variables exhibit significant correlations.

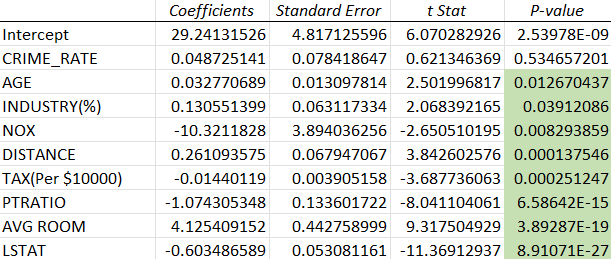
Notably, TAX and DISTANCE, NOX and INDUSTRY, NOX and AGE demonstrate a strong positive correlation, indicating a direct relationship.

On the other hand, AVG PRICE and LSTAT, LSTAT and AVG ROOM, AVG PRICE and PTRATIO shows a negative correlation, suggesting an inverse association.

These findings provide valuable insights into the interdependencies among key factors.

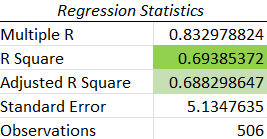
**Multi Linear Regression Model**

From the multi linear regression model we get the coefficients and p values



The p value of CRIME RATE is greater than 0.05, from this we can conclude that crime rate is not a significant predictor for average price.

NOX, TAX, PTRATIO and LSTAT have negative coefficient which means with increase in any of these variables will result in decrease in the average price of the house.



All independent variables together explain 69% of the variability for average price of the house.

**Conclusion**

As our analysis concludes, we gain a profound understanding of the factors influencing house prices in Boston.