**Context**

This analysis aims to determine if there are any meaningful correlations between our attributes. This analysis seeks to interpret the correlational relationships to gain insight into what attributes strongly affect other attributes and how that relationship is possible, and its implications if any.

**Data Preparation**

During data preparation, inconsistent data was detected. Data entries on the Charles River dummy variable are neither a one nor a zero. The inconsistent values are handled by rounding the decimal entries to obtain either a one or a zero. Any other inconsistencies will be treated as outliers and will promptly be removed from the data set. Another area for improvement with the data is missing data located in the median value of owner-occupied homes in $1000's. The missing data is handled by replacing the values with the mean value of the MEDV attribute. Outliers were discovered within the dataset; however, the outliers are a natural result of the study and do not consist of errors or accidental values. Therefore, the outliers must remain within the data set as they are a natural part of the study into the attributes that concern the housing values in the suburbs of Boston.

**Correlation Methods**

The analysis is a correlation analysis using correlation coefficients and matrices to identify attributes that are correlated to one another. A matrix is generated that consists of the person correlation coefficients. Once the correlation matrix is generated, a visualization is created to ease the difficulty of identifying correlated attributes. Developing a matrix without data reduction returns all attributes and their corresponding correlation values.

**Results**

Analyzing the correlation matrix visualization reveals that the concentration of nitric oxides has a strong positive correlation, a coefficient of 0.98 ,with the average number of rooms per home. This correlation is due to the increased number of appliances in larger homes. A larger home will require a greater quantity of appliances. As a result, the nitric oxides produced by the greater number of appliances will increase the nitric oxide concentration in the home.

There is a strong negative correlation, a coefficient of -0.74, between a home's age and the home's distance from Boston employment centers, and the home's access to radial highways. The correlation likely exists because development can be complex in more historic areas. Therefore, the development only occurs around older homes and results in those homes being farther from more developed areas where employment centers are likely to be found.

An additional correlation that is significant is the strong negative correlation, a coefficient of -0.67, between the percentage of the population that is lower status and the median value of owner-occupied homes. This relationship shows that homes located in areas where the percent of the population is lower status are greater than those with a lower median value.

Lastly, there is a strong negative correlation, a coefficient of -0.68, between the pupil-teacher ratio by town and the proportion of African Americans by city. This relationship clearly shows that the ratio between pupils and teachers decreases in areas where African Americans make up a more significant proportion of the population.

**Interpretation**

The correlation matrix has revealed information beneficial to understanding the housing market and societal issues. The correlation between nitric oxide concentration and the number of rooms in the home shows that homes need fewer appliances. Development in historical areas is complex, resulting in older homes being located further away from Boston employment centers. Development efforts could focus on historical home regions while following the protocol to preserve important historical sights.

The correlation between the lower-status population and a home's median value reveals an incentive to help the lower-status population into a higher status. By achieving that status change, the median value of homes in the area will increase. Thus, increasing the value of the houses in the area.

Lastly, in areas where African Americans are of more significant proportion, the pupil-teacher ratio is worse. This correlation reveals that the areas where the African American population is more significant are being put at a disadvantage as in those communities, there are more students per one teacher, which hinders student achievement. A tremendous effort needs to be made to improve the pupil-teacher ratio in those communities by introducing more teachers into the area or creating additional schools to lower the student population at the schools.

**Data Preparation**

Graphical user interface, text, application

Description automatically generated

**Correlation Matrix**

Graphical user interface, text, application, chat or text message

Description automatically generated

**Visualizations**

Timeline

Description automatically generated

Calendar

Description automatically generated