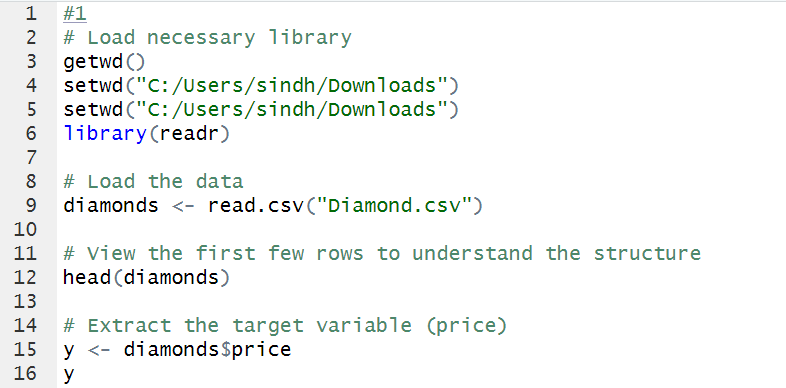
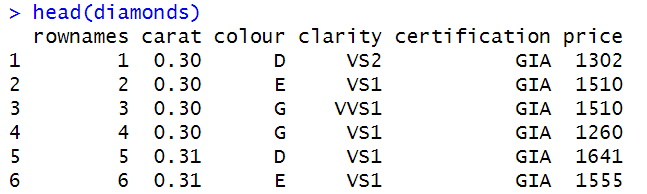
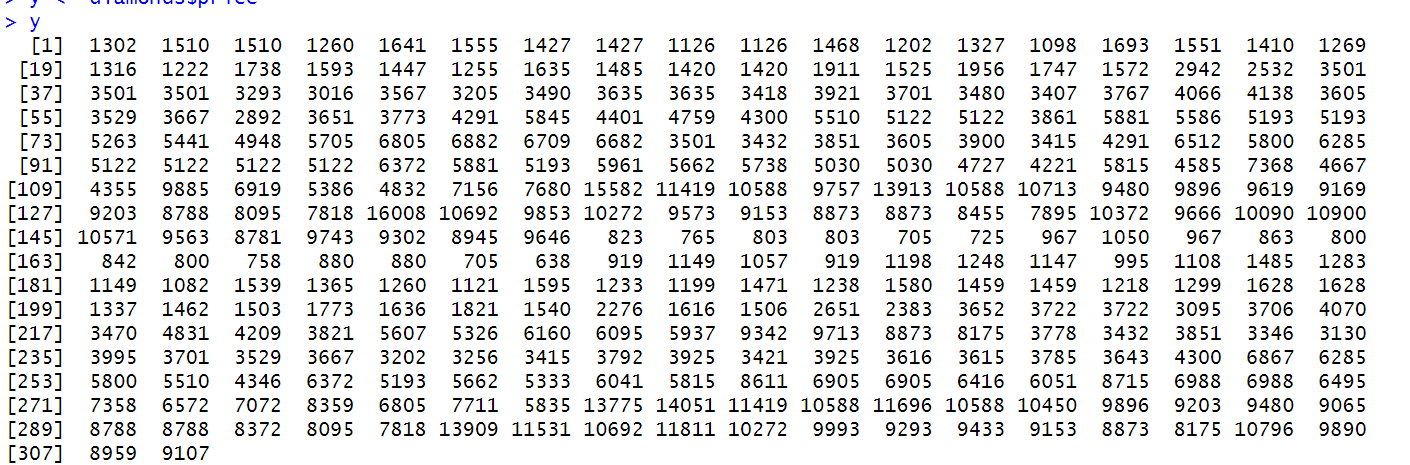
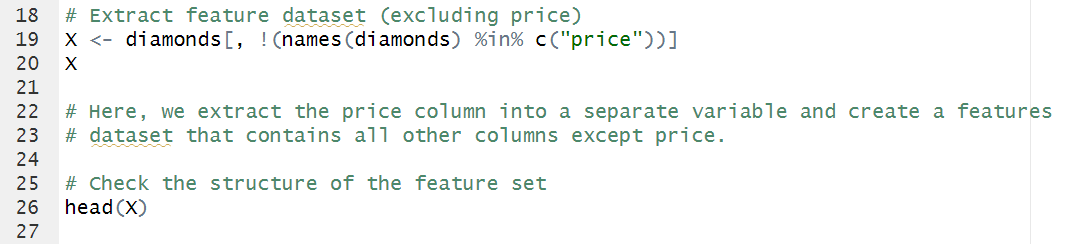
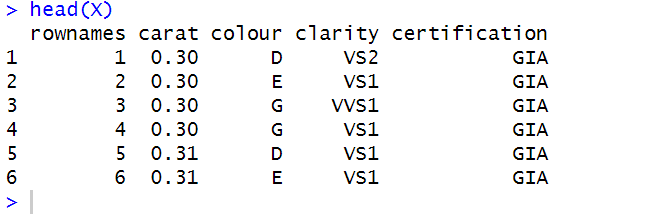
5303 PROJ 4

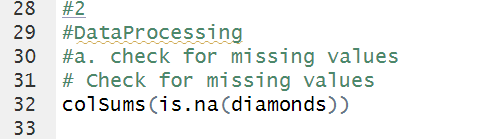


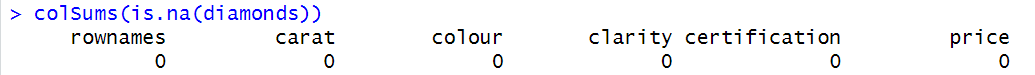


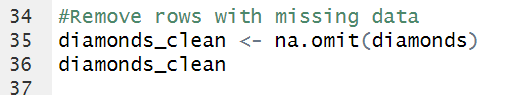


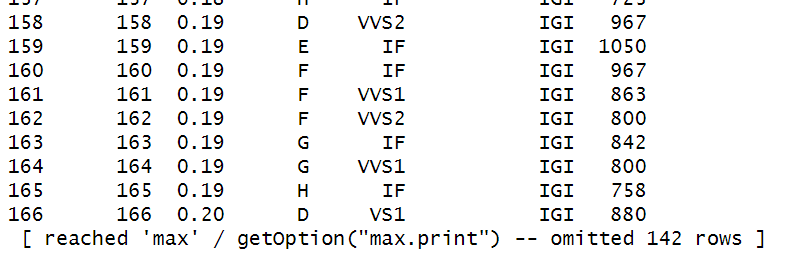


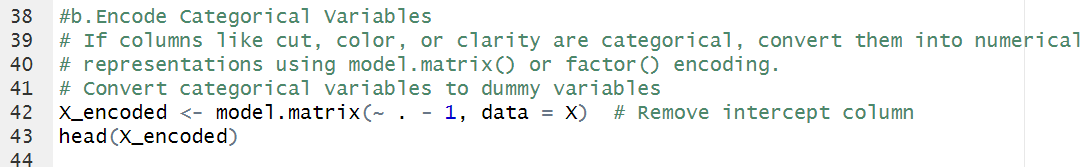


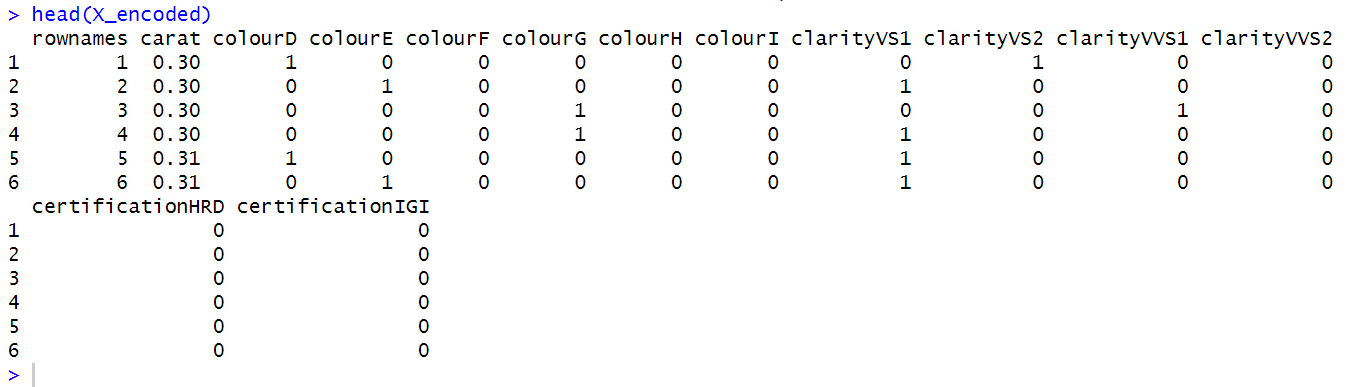


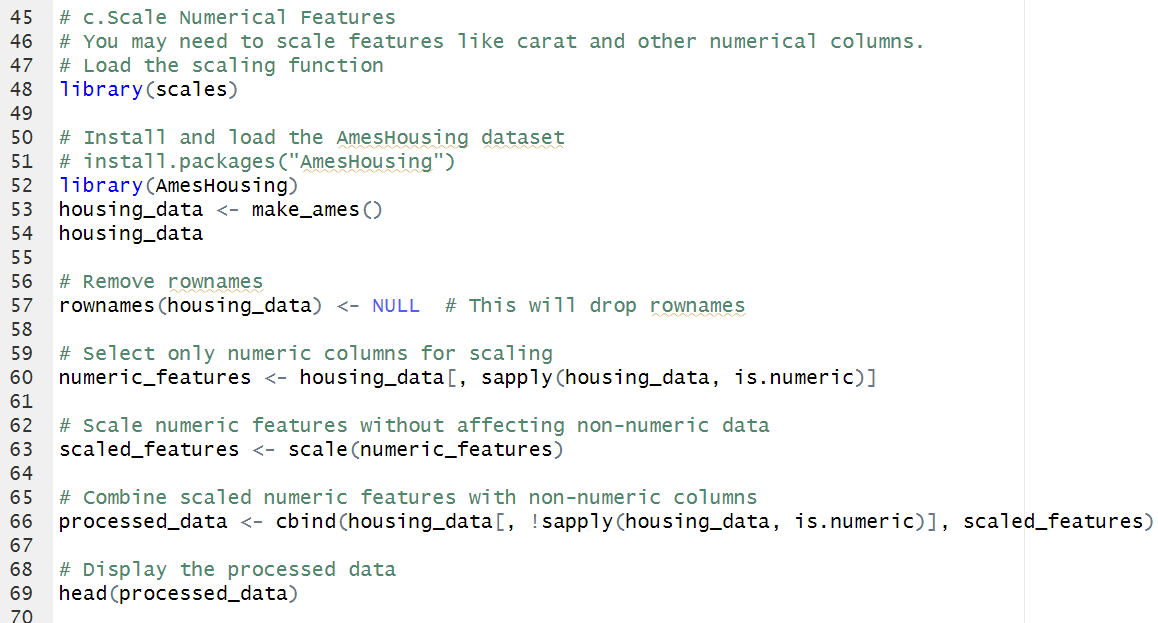


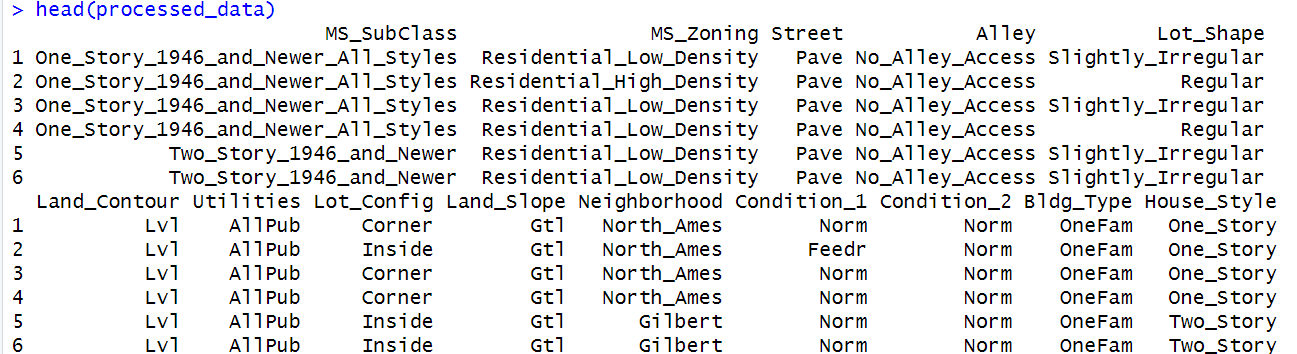


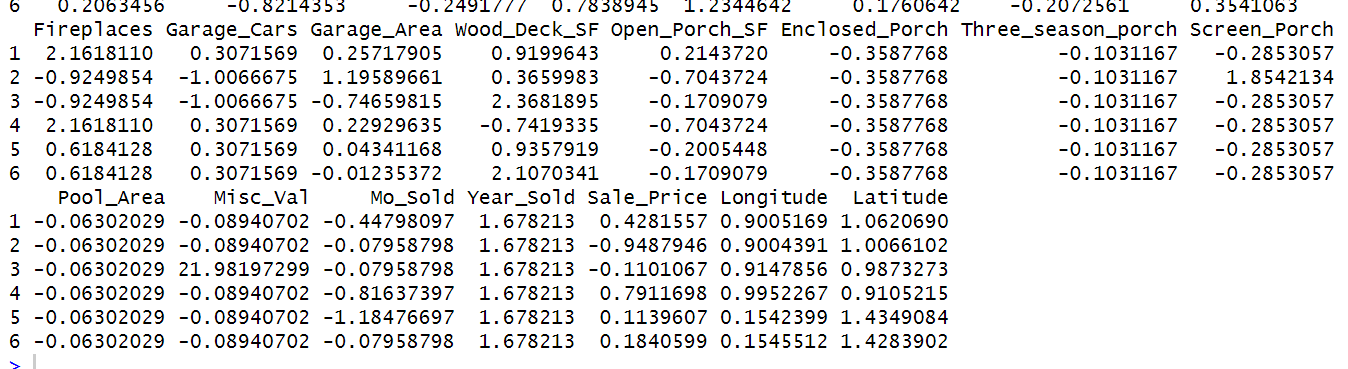


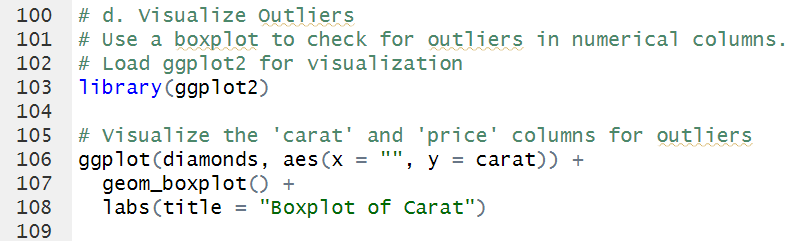


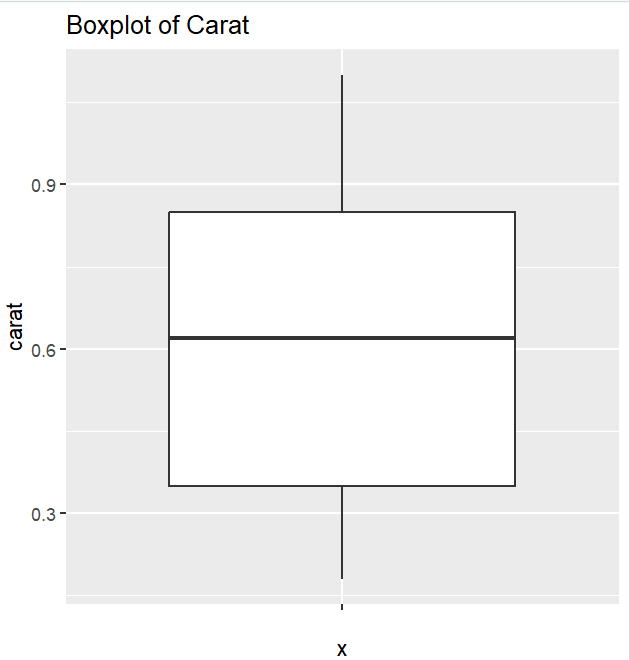


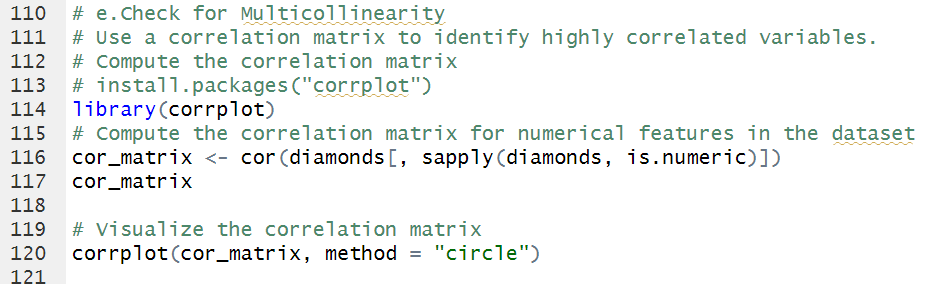


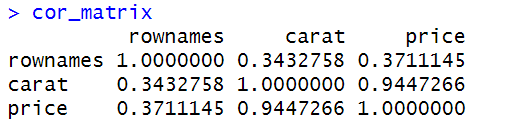


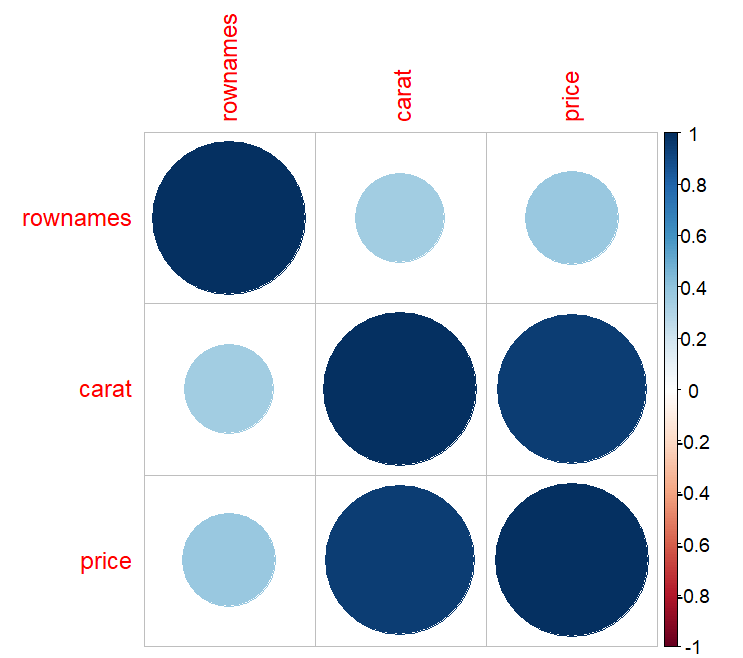


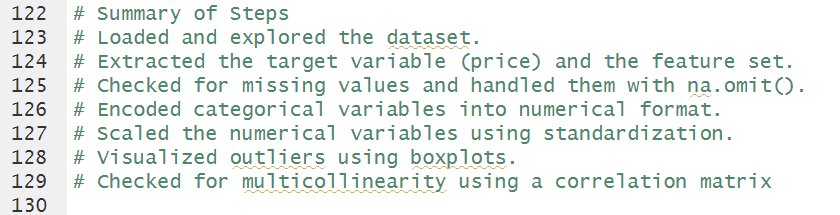












**Preprocessing Steps:**

1. **Check for Missing Values**
   * **Purpose**: Identify any missing values in the dataset to ensure data quality.
   * **Action**: Examine each column for missing entries, as they can affect model performance and may require imputation.
2. **Handle Categorical Variables**
   * **Purpose**: Convert categorical variables (colour, clarity, and certification) into a numerical format suitable for machine learning models.
   * **Action**: Use one-hot encoding to transform these categorical features into numerical columns, where each unique category becomes a separate binary column.
3. **Feature Scaling**
   * **Purpose**: Normalize numerical features to bring them to a similar scale, as this helps models that rely on distance-based calculations (e.g., linear regression).
   * **Action**: Apply standardization or normalization to the carat feature.
4. **Remove Irrelevant Columns**
   * **Purpose**: Remove columns that don't provide information for model training.
   * **Action**: Exclude the rownames column, as it only serves as an index and does not add any meaningful information.

