Pre-processing Housing Price dataset

Consider that you are analysing a dataset that contains information from houses. You are asked to develop a system to predict the price of the house based on the available information. Before you move on to build a predictive model you should explore the data to better know its characteristics.

Consider the dataset in the file Housing.csv. Develop code for the following analysis and to answer the questions.

1. **Standardization**: Using scikit-learn's StandardScaler, standardize the 'price' and 'area' features. Then, create a scatter plot using matplotlib to visualize the relationship between these standardized features. How does the plot change compared to using the original values?

2. **Log Normalization**: Apply log transformation to the 'price' feature using numpy's log function. Create histograms of the original and log-transformed 'price' using seaborn. How does the distribution change? Think of the implications for further analysis.

3. **Outlier Handling**: Identify outliers in the 'price' column using the Interquartile Range (IQR) method. Create a box plot using seaborn to visualize the outliers. Then, use pandas to create a new dataframe without these outliers. How does this affect the mean and median of the 'price' feature?

4. **Encoding Categorical Variables**: Use pandas' get_dummies() function to one-hot encode the 'furnishingstatus' feature. Then, use scikit-learn's LabelEncoder to encode the 'mainroad', 'guestroom', 'basement', 'hotwaterheating', 'airconditioning', and 'prefarea' features. Discuss the pros and cons of each encoding method for this dataset.

5. **Feature Engineering**: Create a new feature 'price_per_sqft' by dividing 'price' by 'area'. Then, use seaborn to create a pair plot of 'price', 'area', and 'price_per_sqft'. What insights can you gain from this new feature?

6. **Feature Interaction**: Create an interaction feature between 'bedrooms' and 'bathrooms' by multiplying them. Use pandas to calculate the correlation between this new feature and 'price'. How does this correlation compare to the individual correlations of 'bedrooms' and 'bathrooms' with 'price'?

7. **Combining Techniques**: Standardize all numerical features, handle outliers in the 'price' feature, and encode categorical variables. Then, use seaborn's heatmap to visualize the correlation matrix of the processed features. Which features show the strongest correlations with 'price' after these transformations?