**1. Handling Missing Values :**

**Bathrooms and Bedrooms :** The missing values in the 'bathrooms' and 'bedrooms' columns were filled based on the category of the listing. If it's a rental property, 'bathrooms' were filled with 1 and 'bedrooms' with 1; if it's an apartment, 'bathrooms' were filled with 2 and 'bedrooms' with 2; for other housing types, 'bathrooms' were filled with 3 and 'bedrooms' with 3.

**Amenities :** Mode values for 'amenities' were calculated separately for each category ('rent', 'apartment', 'housing'). Missing values in 'amenities' were then filled based on the mode of the respective category.

**Pets Allowed :** Missing values were filled with 'None'.

**City Name** **:** Missing values were filled with the mode (most frequent value) of the column.

**State** **:** Missing values were filled with the mode of the column.

**Address** **:** Addresses were filled based on the mode addresses for each city. If there's only one address for a city, that address is used; otherwise, the mode address for the city is used. If mode address isn't available, a default value ('unknown') is used.

**Longitude and Latitude** **:** Missing values in 'longitude' and 'latitude' columns were filled with the mean of the respective column. Additionally, negative values in the 'longitude' column were made positive using the absolute function (**abs()**), ensuring all longitude values are positive.

**2. Handling Categorical Data :**

**Label Encoding :** Categorical columns were identified, and label encoding was applied using `LabelEncoder()` from `sklearn.preprocessing`. This transforms categorical values into numerical labels, which is necessary for machine learning algorithms to operate on them.

These preprocessing techniques ensure that the dataset is clean, filled with meaningful values, and properly formatted for training machine learning models. Additionally, the techniques applied maintain the integrity and relevance of the data for accurate modeling and predictions.

**4. Feature Scaling:**

To handle highly varying magnitudes or values or units, normalization was performed features which have a relatively small range of values ('category', 'amenities', 'bathrooms','bedrooms','has\_photo','pets\_allowed','price','square\_feet','address','cityname','state','latitude','longitude','source') and standardization on the features with a wide range of values ('title', 'body') to avoid weighing greater values, higher and consider smaller values as the lower values, regardless of the unit of the values