1. **Missing Values Handling**:
   * Replaced missing values in the amenities column with 'Not Specified'.
   * Replaced missing values in the pets\_allowed column with 'Unknown' and standardized the values to allowed categories ("None", "Cats", "Dogs", or "Cats,Dogs").
   * Replaced missing values in bedrooms and bathrooms columns with their respective medians.
   * Dropped rows where critical columns (latitude, longitude, price, price\_display) had missing values.
   * Filled missing values in cityname and state columns with 'Unknown'.
2. **Column Removal**:
   * Dropped the address column, possibly due to irrelevance or high missing value ratio.
3. **Outlier Detection and Removal**:
   * Used the Interquartile Range (IQR) method to detect and remove outliers from numerical columns, including bathrooms, bedrooms, price, square\_feet, latitude, and longitude.
4. **Data Type Conversion**:
   * Converted price\_display to a numeric column by removing non-numeric characters (e.g., $).
   * Ensured numeric data consistency for bedrooms and bathrooms.
5. **Data Standardization and Transformation**:
   * Standardized numeric data using StandardScaler for PCA.
   * Conducted PCA with a specified number of components to identify key patterns in the data.
6. **Date Conversion**:
   * Converted timestamps in the time column to datetime format for trend analysis.
7. **Final Cleaning**:
   * Removed entries with logically inconsistent values (e.g., square\_feet <= 0 or price <= 0).
   * Exported the cleaned dataset to a CSV file named apartment\_cleaned\_data.csv.

These steps ensure the dataset is clean, consistent, and ready for further analysis or modeling.















