Housing Dataset

The housing dataset provides comprehensive information on various attributes associated with residential properties, including price, number of bedrooms and bathrooms, square footage, location details, and other relevant features. The objective of this project is to conduct an in-depth analysis of the dataset to derive valuable insights for stakeholders in the real estate industry.

The dataset contains following information:

1. Date: The date when the property information was recorded.
2. Price: The price of the residential property.
3. Bedrooms: The number of bedrooms in the property.
4. Bathrooms: The number of bathrooms in the property.
5. Sqft_living: The total square footage of living space in the property.
6. Sqft_lot: The total square footage of the lot or land area associated with the property.
7. Floors: The number of floors in the property.
8. Waterfront: Indicates whether the property has a waterfront view (binary: 0 for no, 1 for yes).
9. View: An index from 0 to 4 representing the quality of the view from the property.
10. Condition: An index from 1 to 5 representing the overall condition of the property.
11. Sqft_above: The square footage of the interior space above the ground level.
12. Sqft_basement: The square footage of the basement space in the property.
13. Yr_built: The year when the property was built.
14. Yr_renovated: The year when the property was last renovated.
15. Street: The street address of the property.
16. City: The city where the property is located.
17. Statezip: The state and zip code of the property.
18. Country: The country where the property is located.

Tasks to be undertaken include:

1. **Data Cleaning and Preprocessing:**
    - Handle missing values: Identify and address missing values in the dataset through imputation or removal, ensuring data completeness.
    - Address inconsistencies: Detect and rectify any inconsistencies or anomalies in the data, such as erroneous entries or irregular formatting.

- Handle categorical variables: Encode categorical variables appropriately to facilitate analysis.

2. **Exploratory Data Analysis (EDA):**
    - Univariate Analysis: Explore distributions and summary statistics of individual variables such as price, square footage, and number of bedrooms and bathrooms.
    - Bivariate Analysis: Investigate relationships between pairs of variables, such as price vs. square footage, bedrooms vs. bathrooms, etc., using visualizations and statistical methods.
    - Multivariate Analysis: Examine interactions and dependencies among multiple variables, identifying correlations and patterns that may influence housing prices.
3. **Visualization:**
    - Utilize various visualization techniques (scatter plots, histograms, box plots, etc.) to visually represent the distribution and relationships within the dataset.
    - Create insightful visualizations to illustrate trends, outliers, and geographical patterns in housing prices and attributes.
4. **Feature Engineering:**
    - Extract or create new features that may enhance predictive models or provide additional insights into housing prices, such as age of the property, price per square foot, etc.
    - Perform dimensionality reduction if necessary to simplify the dataset while preserving important information.
5. **Analysis and Interpretation:**
    - Interpret findings from the data exploration and visualization, providing insights into factors influencing housing prices.
    - Identify key drivers of housing prices, such as location, property size, condition, and amenities, through statistical analysis and modeling.
    - Offer recommendations or suggestions based on the analysis, such as potential strategies for pricing, investment, or property development.
6. **Documentation:**
    - Document the entire analysis process, including data cleaning steps, exploratory analysis findings, and interpretation of results.
    - Present the analysis report in a clear and concise manner, with visualizations and insights that cater to both technical and non-technical audiences.