

Data Collection and Preprocessing Phase

Visualizing Housing Market Trends: An Analysis of Sale Prices and Features using Tableau

Section	Description
Data Overview	The dataset contains Transformed housing data and 21,609 house sale records, including Property features such as Sales price, area, bedrooms, bathrooms, floors and location. There are a total of 31 columns, out of which Sale Price can be supposedly taken as a dependent variable. The other variables are different features, locations and date, etc. regarding the houses.
Data Cleaning	Missing values were handled, duplicates were removed, and data inconsistencies were corrected. Removed 11 duplicate records to ensure data integrity. Verified that no missing values exist.
Data Transformation	Applied filtering, sorting, pivoting, and created calculated fields such as Sales price (Average), price per square foot (SUM) and Age of houses (Average) and filters like Sales price (Top 25), Ages of houses (Top 10) and Measure Names. Converted multiple binary zip code group columns into a single categorical variable. Removed or adjusted potential outliers in 'Sale Price' and 'Lot Area' using statistical techniques.

Data Type Conversion	Corrected data types, ensuring numerical fields (e.g., price, area) are properly formatted. Ensured that numerical fields (e.g., Sale Price, Lot Area) are in the correct format. Converted categorical variables like Zipcode Groups into appropriate data types.
Column Splitting and Merging	Split combined address columns and merged relevant features for better analysis. Merged zip code binary columns into a single categorical field for better data structure.
Data Modeling	Established relationships between features and variables such as house size, number of floors, bathrooms, bedrooms, sale price and Age of houses (in years). Identified key relationships between features such as Sale Price, Lot Area, and House Features for further analysis in Tableau.
Save Processed Data	-