Unaza Ali

Data Source

Sourcing Open Data

Dataset Name: King County House Sales

Source: Kaggle – [House Sales Dataset](#)

Description:
This dataset contains sales data for houses in King County, Washington (which includes Seattle). It includes home characteristics such as square footage, number of bedrooms and bathrooms, lot size, date of sale, renovation year, and geographic details like latitude and longitude. The dataset includes over 21,000 rows and 21 columns.

Why I Chose This Dataset:
I selected this dataset because it meets all the project brief requirements and is easy to work with. It includes multiple continuous and categorical variables, time-based data (sale dates), and geographic information (zip codes, latitude, longitude). Housing market analysis is also a practical and relevant topic with real-world impact, making it a strong choice for demonstrating key data analytics skills.

Data Profile

Cleaning & Preparation Steps:

- Dropped the id column (not needed for analysis).
- Converted date to proper datetime format for time-based analysis.
- Converted zip code to string to treat it as a categorical variable.
- Created a new column was_renovated to flag if a house was renovated.
- Checked for duplicate rows: **0 found**.
- Checked for missing values: **0 found**.

Variable Overview:

- **Continuous Variables**: price, sqft_living, sqft_lot, bathrooms, sqft_above, sqft_basement
- **Categorical Variables**: zip code, condition, view, grade, floors, waterfront
- **Geographic Variables**: zip code, lat, long
- **Time-Based Variable**: date

Limitations:

- Some columns like yr_renovated contain many zeros, which may limit insight.
- Zipcode-level location data may not fully capture neighborhood-level differences.

Ethical Considerations:

- Dataset contains no personal or sensitive information.
- Regional housing data may reflect market or economic biases in pricing.

Questions to Explore

1. **What factors most strongly influence house prices in King County?**
2. **How do house prices vary across different zip codes?**
3. **Do renovated houses sell for more than unrenovated ones?**
4. **Is there a seasonal trend in house sales?**