

### King County House Sales Analysis

#### DATA

- **Source of Data**
  - [House Sales in King County, USA](#)
  - This data set was downloaded from the open-source platform Kaggle.com
- **Data Content**
  - The data set consists of house sales in King County, Washington from May 2014 – May 2015. Along with prices, it consists of information on different home features such as square footage, sale ID, and date of sale.
- **Data Profile**
  - There are 21 columns and 21613 rows. After the data quality check there is 22 columns and 20939 rows.
    - Dropped “id” column since it will not be used for analysis
    - Split “date” into ‘year’, ‘month’, and ‘day’
  - There are no duplicates or null values in this dataset
    - Removed 674 outliers that were skewing the data

Variable	Description	Time Variant/Invariant	Data Type
ID	Unique ID for each home sold	Invariant	Qualitative
date	Date of home sale	Variant	Qualitative
price	Price of each home sold	Invariant	Quantitative
bedrooms	Number of bedrooms	Invariant	Quantitative
bathrooms	Number of bathrooms, where .5 indicates a bathroom with a toilet but no shower	Invariant	Quantitative
sqft_living	Square footage of the homes interior space	Invariant	Quantitative
sqft_lot	Square footage of the land space	Invariant	Quantitative
floors	Number of floors	Invariant	Quantitative
waterfront	Variable to indicate whether the property is on the waterfront	Invariant	Qualitative
view	An index of 0-4 of how good the view is	Invariant	Quantitative
condition	An index from 1-5 on the condition of the home	Invariant	Quantitative

grade	An index from 1-13 based on the home's quality of building construction and design	Invariant	Quantitative
sqft_above	The square footage of the interior spacing above ground level	Invariant	Quantitative
sqft_basement	Square footage of the interior space below ground level	Invariant	Quantitative
yr_built	Year home was built	Invariant	Quantitative
yr_renovated	Year home was last renovated	Invariant	Quantitative
zipcode	What zipcode the home is in	Invariant	Quantitative
lat	The latitude of the home	Invariant	Quantitative
long	The longitude of the home	Invariant	Quantitative
sqft_living15	Average size of interior housing living space for the closest 15 houses, in square feet	Invariant	Quantitative
sqft_lot15	Average size of land space for the closest 15 houses, in square feet	Invariant	Quantitative

- **Data Limitations & Ethics**

- Some potential data limitations:

- Data Completeness: The dataset may not include all house sales in King County, WA. It might only include a sample or subset of the total sales, which could introduce sampling bias and limit the generalizability of the findings.
    - Data Accuracy: The accuracy of the data could be compromised by errors, omissions, or discrepancies in recording the information. For example, there might be incorrect entries for variables such as price, square footage, or property characteristics.
    - Data Reliability: The dataset's reliability depends on the source from which it was obtained. If the data is sourced from multiple agents or agencies, inconsistencies or variations in reporting standards and methodologies might exist.
    - Data Bias: The dataset might exhibit inherent biases, such as underrepresentation of certain neighborhoods or property types, which could affect the accuracy and representativeness of the findings.
    - External Factors: The dataset might not capture external factors that could influence housing prices, such as changes in economic conditions, zoning regulations, or major infrastructural developments.

- There is no PII in this dataset so there are no ethical concerns

- **Questions**

- How does the price of houses vary based on the number of bedrooms and bathrooms?
- Is there a relationship between the size of the living area (sqft\_living) and the price of houses?
- Are houses with a waterfront view (waterfront variable) priced differently compared to those without a waterfront view?
- What is the distribution of house prices across different zip codes in King County, WA?
- Is there a correlation between the grade of a property (grade variable) and its selling price?
- Is there a correlation between the age of a property (yr\_built) and its selling price?
- Are houses that have been renovated (yr\_renovated) priced differently compared to those that haven't undergone renovations?
- How does the size of the lot (sqft\_lot) relate to the price of houses?