

## Project Information

### Data Source

[House Sales in King County, Washington on Kaggle](#)

This dataset contains house sale prices in King County, Washington, for transactions between May 2014 and May 2015. It includes details such as property size, number of bedrooms and bathrooms, and various other home attributes. The dataset was sourced externally from Slideshare and made available by a creator on Kaggle. While the specific data collection methods weren't disclosed, the data does not appear to be AI-generated and is assumed to be real.

### Variables:

Column Name	Description	Variable Type
Id	Unique ID for each home sold	Categorical
Date	Date of the home sale	Categorical
Price	Price of each home	Quantitative
Bedrooms	Number of bedrooms	Quantitative
Bathrooms	Number of bathrooms, where .5 accounts for a room with a toilet but no shower	Quantitative
Sqft_living	Square footage of the apartments interior living space	Quantitative
Sqft_lot	Square footage of the land space	Quantitative
Floors	Number of floors	Quantitative
Waterfront	A dummy variable for whether the apartment was overlooking the waterfront or not	Categorical
View	An index from 0-4 of how good the view of the property was	Quantitative
Condition	An index from 1 to 5 on the condition of the apartment	Quantitative
Grade	An index from 1 to 13, where 1-3 falls short of building construction and design, 7 has an average level of construction and design, and 11-13 have a high quality level of construction and design.	Quantitative
Sqft_above	The square footage of the interior housing space that is above ground level	Quantitative

Sqft_basement	The square footage of the interior housing space that is below ground level	Quantitative
Yr_build	The year the house was initially built	Categorical
Yr_renovated	The year of the house's last renovation	Categorical
Zipcode	What zipcode the house is in	Categorical
Lat	Latitude	
Long	Longitude	
Sqft_living15	The square footage of interior housing space for the nearest 15 neighbors	Quantitative
Sqft_lot15	The square footage of the land lots of the nearest 15 neighbors	Quantitative

I chose this dataset because I grew up near King County in Washington state, so it feels personal to me. I'm also interested in predictive analysis, and using this dataset to forecast future home prices sounds exciting. Since the data is from 2014-2015, it would be interesting to compare the predictions with today's housing prices to see how accurate the model turns out.

### Data Profile

Consistency checks done to ensure there are no duplicates, mixed-type data, missing values, or outliers that are unreasonable given the context of this data set. The following changes were made.

- Dropped 'id' column – identifier column not necessary for analysis.
- Updated data type for 'date' column to datetime format.
- Dropped record containing 33 bedrooms. All remaining values are reasonable.

[Data Profile\\_King County House Sales](#)

### Limitations

The data is from 2014-2015 which means that insights gleaned from an analysis on this data may not reflect current housing price trends. Additionally, with limited geographical coverage, it restricts the ability to generalize findings to other regions.

## **Ethical Considerations**

Although the dataset is shared under a CC0 Public Domain license, granting permission for its use and distribution, there is limited information regarding the original collection and handling of the data. This raises concerns about whether informed consent was obtained from individuals whose data is included. While the open license allows for its ethical use in research and analysis, it leaves unanswered questions about the dataset's origin.

Potential biases, such as collection bias, may exist because the dataset's scope and methodology are unclear. Without transparency, it's difficult to know if certain factors were overemphasized or omitted based on the collector's assumptions. For instance, if the dataset predominantly includes homes sold within a specific price range or focuses on single-family homes but excludes other types like condos, it might not fully represent the housing market. Such biases could distort the conclusions drawn from the analysis.

## **Key Questions:**

- What factors have the strongest correlation with house prices?
- How does the number of bedrooms or bathrooms influence the sale price of homes?
- Which neighborhoods in King County have the highest average home price?
- Does the year of renovation significantly impact property values in King County?
- How does the proximity to the waterfront impact property values in King County?
- Is there a difference in pricing trends between homes with high 'view' scores versus those with lower scores?
- What economic or environmental factors might influence future housing prices in King County beyond what is captured in this dataset?
- How might changes in housing policy or interest rates affect the predictions based on this dataset?