

House Sales in King County USA

April 2023

Data

- Data Source: This data was downloaded from an open source, Kaggle. This data was publicly available and considered open-sourced data. Please access the link [here](#).
- Data Collection: This data was collected from Harlfoxem.
- Data Contents: This data contains house sale prices from all homes sold between May 2014 and May 2015.
- Data Profile: 21 columns with 21,437 rows
 - After cleaning, 16 columns with 21,614
 - Removal of several columns that were not needed for analysis (sqft_living15, sqft_living, grade, sqft_above and sqft_lot15)
 - 177 duplicate Ids were found and removed.
 - One outlier, 2402100895 contained 33 bedrooms which when looking at the sqft, price, etc, it seemed like an incorrect input. I imputed 3 since it is the average of bedrooms.

Index	Column	Description	Time Variant/ Invariant	Data Type
1	Id	Unique Id for each listing	Time Invariant	Quantitative
2	Date	Date listed	Variant	Quantitative
3	Price	Price that house is listed for	Time Invariant	Quantitative
4	Bedroom	Amount of bedrooms in the house	Time Invariant	Qualitative

5	Bathroom	Amount of bathrooms in the house	Time Invariant	Qualitative
6	sqft_lot	Square footage of of estate	Time Invariant	Quantitative
7	Floors	Number of floors within house	Time Invariant	Qualitative
8	Waterfront	Property is waterfront or not	Time Invariant	Qualitative
9	View	Property does/ doesn't have view	Time Invariant	Qualitative
10	Condition	Condition of the house, based on a 5 point scale	Time Invariant	Qualitative
11	sqft_basement	Square footage of basement	Time Invariant	Quantitative
12	yr_built	The year that the estate was built	Variant	Quantitative
13	yr_renovated	The year that the house was renovated.	Variant	Quantitative
14	Zip Code	Zip Code of the estate	Time Invariant	Quantitative
15	lat	Latitude of estate	Time Invariant	Quantitative
16	long	Longitude of the estate	Time Invariant	Quantitative

Limitations & Ethics

- 177 duplicates were found in the ID column and removed. The data also contains some columns that will not be needed. No missing data seems to be present and columns are clearly labeled.
- This data does not contain any sensitive or personal data. The data set does contain zip code, longitude, and latitude of the house but this is public information.

Questions to explore

- How does the renovation year impact the price?
 - No correlation has been found.
- Does the geographical location of the estate impact the price?
 - Yes, when estates are closer to Bellevue and Lake Washington the prices go up dramatically.
- What is the average sqft of houses within King Country?
 - AVG sqft_lot is 15135
- Are there certain variables that impact the price more?
 - Geographical location seems to have the largest impact on price however the variables all together make up the price of an estate.
- Does the year built impact the price in a negative/ positive way?
 - No correlation has been detected.
- What properties include waterfront/ view?
 - Only 10% of all estates for sale included a view. 99% of estates did not include a waterfront.
- Are certain zip codes pricier than others?
 - 98004, 98006, and 98052 have the highest cost of houses.

Hypothesis

- If there is more square footage, the price of the estate will go up.