

## King County House Sales 2014-2015 by Josh Abrams

### Executive Summary

A real estate company is looking to buy homes in King County Washington. to people who are looking for accommodations in specific locals. They want to make the most of their investments by investing in the most lucrative areas of King County. In this analysis we will explore the geographical trends and look at a price prediction model that can allow hosts to have a better understanding of what income can be made from their property based on historical property data and demographics.

### Data Sources

This dataset contains house sale prices for King County, which includes Seattle. It includes homes sold between May 2014 and May 2015.

<https://www.kaggle.com/datasets/harlfoxem/housesalesprediction?resource=download>

Postal Zip Codes and Associated Cities:

<https://www.unitedstateszipcodes.org/>

(Note that cities were defined by postal zip code)

### Limitation and Ethics

This data is already anonymized, so I don't need to worry about identifying individuals.

The limitations of these studies include not knowing the cost of renovations not knowing taxes and other relevant info for the property and relying on government collected data.

### Data Cleaning and Data Consistency Checks

- Converted the price column to \$
- Converted the date column to more sensible format (Changed date format to remove "T00000" and put in standard American date format)
- Added city and State Columns from (<https://www.unitedstateszipcodes.org/>)
- Check for missing data (Some had bedrooms of 0 assuming this is a studio apartment or house)
- Check for duplicate data( No duplicates)
- Check for mixed data type (No mixed data types)

## Data Profile

Number of rows and columns in the cleaned & amended dataset: 21613 x 23 (original data is, 21613 x 21)

## Column Details

Variable	Description	Data Type	Time Variant
id	Identification	Qualitative, Ordinal	No
date	Date sold	Quantitative, Discrete	Yes
price	Sale price	Quantitative, Discrete	No
bedrooms	Number of bedrooms	Quantitative, Continuous	No
bathrooms	Number of bathrooms	Quantitative, Continuous	No
sqft_liv	Size of living area in square feet	Quantitative, Continuous	No
sqft_lot	Size of the lot in square feet	Quantitative, Continuous	No
floors	Number of floors	Quantitative, Continuous	No
waterfront	'1' if the property has a waterfront, '0' if not.	Qualitative, Ordinal	No
view	An index from 0 to 4 of how good the view of the property was	Qualitative, Ordinal	No
condition	Condition of the house, ranked from 1 to 5	Qualitative, Ordinal	Yes
grade	Classification by construction Quality which refers to the types of materials used and the Quality of workmanship. Buildings of better Quality (higher grade) cost more to build per unit of measure and command higher value. Additional information in: KingCounty	Qualitative, Ordinal	No
sqft_above	Square feet above ground	Quantitative, Continuous	No

sqft_basmt	Square feet below ground	Quantitative, Continuous	No
yr_built	Year built	Quantitative, Discrete	Yes
yr_renov	Year renovated. '0' if never renovated	Quantitative, Continuous	Yes
zipcode	5 digit zip code	Qualitative, Ordinal	No
City	City of sale based on zip code association	Quantitative, Discrete	No
State	State of Sale	Quantitative, Discrete	No
lat	Latitude	Quantitative, Discrete	No
long	Longitude	Quantitative, Discrete	No
sqft_liv15	Average size of interior housing living space for the closest 15 houses, in square feet	Quantitative, Continuous	No
sqft_lot15	Average size of land lots for the closest 15 houses, in square feet	Quantitative, Continuous	No

### Questions to Explore

Can we predict the price of a house in different regions?

Can we estimate the popularity of an area based on prior house sales

- What can we learn about different regions from the data what is undervalued and overvalued?