

House Sales in King County, USA

Introduction

This dataset, sourced from King County, Washington, provides detailed information on house sales between May 2014 and May 2015. With Seattle as the major city in the region, the dataset offers valuable insights for real estate market analysis and serves as an excellent foundation for predictive modeling, particularly in regression-based models. By examining various features of homes sold, such as size, location, and amenities, it is possible to model housing prices and explore factors influencing real estate values in this area.

Data Source and Collection

The dataset is publicly available on Kaggle, contributed by the user "harlfoxem." It was likely collected from King County's public records, which track real estate transactions, house characteristics, and geographic data for tax assessment purposes. Additional sources, such as King County Assessor's Office records, have verified the variables and their definitions.

[House Sales in King County, USA](#)

About Dataset

This dataset consists of 21 variables and approximately 20,000 records of house sales in King County. Each record corresponds to an individual property sale, with features capturing physical, structural, and geographic details.

Data Content and Variables:

id: Unique identifier for each home sold.

date: Date of the home sale.

price: Sale price of the home.

bedrooms: Number of bedrooms.

bathrooms: Number of bathrooms (0.5 denotes a half-bath).

sqft_living: Interior living space area in square feet.

sqft_lot: Land space in square feet.

floors: Number of floors in the house.

waterfront: Binary variable indicating if the property has a waterfront view.

view: Rating from 0 to 4 for property view quality.

condition: Index from 1 to 5 representing the property's condition.

grade: Construction and design quality index, from 1 to 13.

sqft_above: Square footage of interior living space above ground.

sqft_basement: Square footage of basement space.

yr_built: Year the house was built.

yr_renovated: Year the house was last renovated.

zipcode: Zip code location.

lat and **long:** Latitude and longitude coordinates.

sqft_living15: Interior living space in square feet for the 15 nearest neighbors.

sqft_lot15: Lot space in square feet for the 15 nearest neighbors.

Benefits of using this dataset:

This dataset is ideal for developing predictive models, especially in the context of real estate price estimation. Its comprehensive nature allows for detailed regression analyses, which can help predict home prices based on specific property characteristics. Additionally, the dataset is valuable for conducting a market analysis to identify trends and key factors influencing property values in King County. The variety of features available—including structural characteristics, size, and location—offers a well-rounded view of the real estate market. Geographic data, such as latitude, longitude, and zip code, further enhance its utility, allowing for spatial analysis and enabling researchers to explore how location impacts housing prices.

Data Limitations:

While the dataset is rich in information, it does have several limitations. The time frame covered is narrow, only spanning sales from May 2014 to May 2015, which restricts its use for analyzing long-term trends. Moreover, its scope is limited to King County, Washington, meaning that findings may not be generalizable to other regions or markets. Certain data points are also incomplete; for instance, some homes lack information on the year of renovation, or condition ratings may be inconsistent. Additionally, factors external to King County, such as economic trends or housing demand in other areas, are not accounted for, potentially limiting the robustness of predictions for different market conditions.

Questions to explore with the analysis:

1. Which factors are most influential in determining house prices?
2. How do geographic factors, such as proximity to water or specific zip codes, affect housing prices?
3. Does the condition or grade of a property correlate more strongly with price?
4. How does square footage (both living space and lot size) impact property value?
5. Is there a premium on newer homes or recently renovated homes compared to older structures?
6. How does the number of bedrooms and bathrooms influence the overall price?