WickyNzioka /
house_sales_northwestern_county

<> **Code**   Pull requests   Actions   Projects   Wiki   Security   Insights   Settings

View license

☆ 0 stars   565 forks   ⊙ 0 watching   1 Branch   0 Tags   Activity

🌐 Public repository · Forked from **learn-co-curriculum/dsc-phase-2-project-v2-3**

main   1 Branch   0 Tags   Go to file   t   Go to file   +   Add file   Code   ...

This branch is **1 commit ahead of** `learn-co-curriculum/dsc-phase-2-project-v2-3:main` .

Contribute   Sync fork

WickyNzioka upload_assignment   1 minute ago   ...   🕘

📁 data	revisions for v2.3	3 years ago
📄 .canvas	sync .canvas	3 years ago
📄 .gitignore	Initial commit	4 years ago
📄 CONTRIBUTING.md	Initial commit	4 years ago
📄 Group 5 Phase 2 project present...	upload_assignment	1 minute ago
📄 Group 5 Project V2 - Jupyter Not...	upload_assignment	1 minute ago
📄 LICENSE.md	Initial commit	4 years ago
📄 README.md	upload_assignment	1 minute ago
📄 halfway-there.gif	Initial commit	4 years ago

📖 README   ⚖ License

Project Overview

This project develops a predictive model to estimate house sale prices in King County, Washington, using various features such as:

- Number of bedrooms and bathrooms
- Square footage of living space and lot
- Location attributes (waterfront, view)
- House condition, grade, year built and year of renovation The model aims to assist potential buyers in understanding and predicting house prices within the region.

Business Problem

This project tackled the issue of uncertainty in house prices in King County. For real estate agencies, it offers a solution to inaccurate listings, extended time on the market, and poor market predictions. The model provides data-driven price estimations, buyer engagement, and valuable market insights. For individual buyers and sellers, it eliminates information asymmetry, empowers informed decisions, reduces negotiation risks, and ultimately prevents financial losses. In essence, this project empowers both sides of the real estate market with data-driven tools for smarter pricing.

The Data

Data Source: King County Assessor's Office

Target Variable: Sale price of the house (what we are trying to predict) Features: House Attributes

id: Unique identifier for each house **date**: Date the house was sold **bedrooms**: Number of bedrooms **bathrooms**: Number of bathrooms **sqft_living**: Square footage of living space **sqft_lot**: Square footage of the lot **floors**: Number of floors **waterfront**: Boolean indicating waterfront proximity (various categories) **view**: Quality of view from the house (various categories) **condition**: Overall condition of the house **grade**: Overall grade of the house **sqft_above**: Square footage excluding basement **sqft_basement**: Square footage of the basement **yr_built**: Year the house was built **yr_renovated**: Year the house was renovated **zipcode**: ZIP code of the house

Location Information:

lat: Latitude coordinate **long**: Longitude coordinate

Neighborhood Features:

sqft_living15: Average square footage of living space for nearest 15 neighbors **sqft_lot15**: Average square footage of the lot size for nearest 15 neighbors

Feature Engineering

- We created new features to enhance the analysis:
  - **house_age**: Calculated as (current year - yr_built) to capture the age of the house.
  - **price_per_sqft**: Calculated by dividing price by sqft_living, providing better view of price based on the living space.
  - **season**: Extracted from the **date** feature to capture potential seasonal variations in house prices.
- We dropped some of the columns that we did not use for the analysis, and only worked with - **date**, **price**, **bedrooms**, **bathrooms**, **sqft_living**, **sqft_lot**, **waterfront**, **condition**, **grade**, **yr_built**,**yr_renovated**, **renovated**,**season**, **grade_no**
- We addressed the outliers by using the z_score. For the bedrooms feature, we addressed the outlier by filtering out the house data containing this outlier 33 bedrooms.

Model Training and Evaluation

From the correlation table we find: -The strongest positive correlation (0.7) is between price and square footage of living area, indicating that larger homes tend to have higher prices. -The second strongest positive correlation (0.67) is between price and the grade_no, suggesting that the higher the grade the higher the price. -The third strongest positive correlation (0.32) is between price and number of bedrooms, implying that more bedrooms in a home also contribute to its price, but to a less extent than bathrooms. -The weakest positive correlation (0.05) is between price and year built, meaning that newer homes have slightly higher prices than older homes, but the effect is negligible. -Price and Year Renovated is only (0.12). Reasons could be: - Renovations might cater to personal preferences rather than broad market appeal. - Recent renovations might indicate potential underlying issues with the property.

Modelling

To build a regression model we started with the baseline model which is a simple linear regression using **sqft_living**. Based on an results from OLS Summary, model Performance:

R-squared: 0.493 (explains 49.3% of price variation) Prob (F-statistic):0.00 (model is statistically significant) Coefficients: sqft_living: 280.8688 This means that for every additional square foot of living space, the predicted house price is expected to increase by $280.87

- For improved accuracy we added more fetaures; **grade_no**, **bathrooms**, **renovated**, **house_age** The model explains 60.7% of price variation, indicating a better fit than the previous model. Higher grades, more square footage, more bathrooms, renovations, and older houses generally lead to higher prices.
- We then added more features; **waterfront, floors and sqft_lot**. The model explains 64.6% variation being the best fit.

Conclusions

Key factors influencing house prices are living space, number of bedrooms and bathrooms. Pn the other hand, renovations do not heavily influence the prices. Average demand for houses through the seasons does not really vary, though there is a noticeable decrease during winter season.

Recommendations

- Emphasize living space, bathrooms and the square footage as key selling points to attract families and those seeking comfort.
- Consider targeted renovations that are only needed as there is no major impact on the value of your property.
- lot size does not necessarily impact the propert value, therefore the focus should be on the house size and not

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package