

Using Linear Regression to Develop a Sustainable Housing Community

By Salome Grasland

Outline

- ★ Business Problem
- ★ Data
- ★ Methods
- ★ Results
- ★ Conclusions

Business Problem



- ★ A result of the Covid pandemic is that people are now spending more time in their yards
- ★ The real estate development group, Steady, has approached us to help design a community of homes for outdoor living
- ★ Steady wishes to build the community of homes in the most temperate zone of King County, Seattle.

Data

- ★ Data is from the King County Housing Dataset from Kaggle.
- ★ Includes homes sold between May 2014 and May 2015, features 21 columns, and over 21,000 entries.
- ★ The dataset contains categorical and numerical columns, with data types of integers, objects, and floats.

Column Names and Descriptions for King County Data Set

id - Unique identifier for a house
date - Date house was sold
price - Sale price (prediction target)
bedrooms - Number of bedrooms
bathrooms - Number of bathrooms
sqft_living - Square footage of living space in the home
sqft_lot - Square footage of the lot
floors - Number of floors (levels) in house
waterfront - Whether the house is on a waterfront
view - Quality of view from house
condition - How good the overall condition of the house is. Related to maintenance of house. See the [King County Assessor Website](#) for further explanation of each condition code
grade - Overall grade of the house. Related to the construction and design of the house. See the [King County Assessor Website](#) for further explanation of each building grade code
sqft_above - Square footage of house apart from basement
sqft_basement - Square footage of the basement
yr_built - Year when house was built
yr_renovated - Year when house was renovated
zipcode - ZIP Code used by the United States Postal Service
lat - Latitude coordinate
long - Longitude coordinate
sqft_living15 - The square footage of interior housing living space for the nearest 15 neighbors
sqft_lot15 - The square footage of the land lots of the nearest 15 neighbors

Methods

Data Engineering

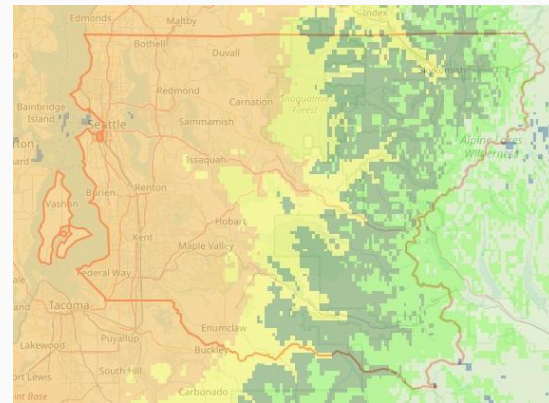
3 columns were constructed: hardness zones ('zones'), square footage of lawn as a percent of total lot size ('sqft_lawn_prc'), and square footage of lawn as a percent of total lot size for the nearest 15 homes ('sqft_lawn_prc15').

Modeling

For modeling OLS regression was used and two values were looked at in the summary: R-squared and p-value.

Validating the Model

The performance of our model it was cross-validated using a train test split using scikit learn



King County Hardiness Zone Map

Results

**Model 4 has the highest
r-squared value at 0.546**

'sqft_lawn_prc', 'sqft_living',
'sqft_living15', 'bedrooms',
'grade_Excellent',
'view_EXCELLENT',
'grade_Luxury', 'view_NONE'

Train
RMSE
0.38425

**The small
difference in values
means the model
generated has good
predictive abilities**

Test
RMSE
0.3854

**\$10,000 increase in
price**

\$5400 increase in
square foot of living

\$4200 increase in
square foot of living

does decrease the
value of the home

**More lawn space
does decrease the
value of the home**

What does this mean
for our client, Steady?

Conclusions

\$5400

Build Larger Homes

For each \$10,000 increase in price for the home we see a \$5,400 increase in value of the square foot of living



We also want to warn Steady that increasing lawn size does decrease home value.

\$4200

That all homes built be larger

For each \$10,000 increase in price for the 15 nearest homes we also see a \$4200 increase in square foot of living for the 15 nearest homes.



- We advise that all homes built be of consistent size
- All homes built should be large (~2000 square feet)

\$\$\$

Excellent views and excellent grades

Homes with excellent views and excellent grades in general increase the value of the home.



- custom design
- high quality cabinet work, wood trim
- and highest quality materials.

Thank You!

Email: salome.grasland@ncf.edu

GitHub: [@SSGrasland](https://github.com/SSGrasland)

LinkedIn: www.linkedin.com/in/sam-grasland-02b99955