Using Linear Regression to Develop a Sustainable Housing Community

By Salome Grasland

Outline

- ★ Business Problem
- ★ Data
- ★ Data Engineering
- ★ Results
- ★ Recommendations

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
- ★ Homes range in price from $78,000 to $7,700,000

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

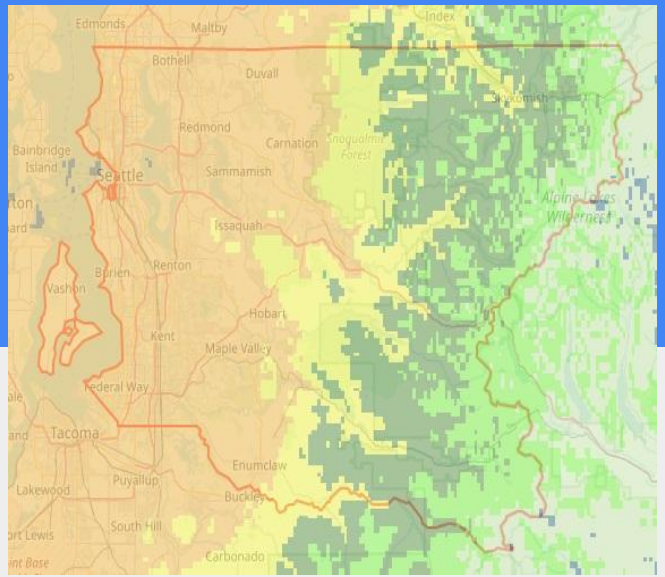
Data Engineering

three columns were engineered for our data set

‘zones’

Steady wants the homes to be built in a temperate area that lends itself to outdoor living.

(A map of the USDA hardiness zones of King County, Seattle)



- **We used the USDA Hardiness Zone map of King County and limited the dataset to homes in Zone 8b (the red portion of the map)– King County’s most temperate zone**
- **This new dataset contains homes only in Zone 8b, has 16 columns, 12,216 entries, and homes ranging in price from $78,000 to $7,700,000**

Data Engineering Continued

three columns were engineered for our data set

‘sqft_lawn’ & ‘sqft_lawn15’

Steady wants to know how lawn size affects the value of the homes they’re building.

- ‘sqft_lawn’ & ‘sqft_lawn15’ were engineered based on values taken from the sqft_living and sqft_lot columns
- The purpose of engineering these columns is to focus on our clients specific issue— building homes that have usable outdoor space

Results

Feature	Value
sqft_living	$263.19
sqft_living15	$113.04
sqft_lawn	$-1.05
bedrooms	$-47,017USD
view_EXCELLENT	$330,266USD
view_NONE	$-67,619USD
grade_Excellent	$595,692USD
grade_Luxury	$1,139,040USD

★ **Using a linear regression model we analyzed the dataset to find which features best predicted a home's price**

★ **The features in the left column of the chart were found to be best indicator of a home's price**

Features that had positive impact on price
square footage of living area, an excellent view, and excellent home quality (aka grade)

Feature that had a negative impact on price
Square footage of lawn, number of bedrooms, and having no view

Recommendations

$263

Build Larger Homes

Large homes are recommended because each additional square foot the home has adds $263.19 to its value

We want to warn Steady that each additional square foot of lawn does decrease the home's value by $1.05.

$113

A Community of Larger Homes

Our analysis shows that when the average size of the 15 nearest homes goes up a square foot it increases the group of homes value by $113.04

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

$595,692

Excellent views and excellent quality (grade)

-An excellent views increases the value of the home by $330,266
-An excellent grade increases the homes value by $595, 692

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

Thank You!

Email: salome.grasland@ncf.edu
GitHub: @SSGrasland
LinkedIn: www.linkedin.com/in/sam-grasland-02b99955