

Zillow Regression Analysis to Inform Purchase Decisions

**March 10, 2023
By Andrew Levinton**

Overview

- Business Problem Discussion
- Data used to conduct study
- Preliminary (Baseline Model) - Flaws, and plans for improvement
- Methodology for improvement of data
- Model - Showing the addition of categorical variables
- Presentation, Interpretation, and Recommendations from Final model
- Business questions model can answer
- Plan for Future work

The Data - KC Housing Dataset - Link Below

<https://info.kingcounty.gov/assessor/DataDownload/default.aspx>

Columns from dataset

The full list of columns with descriptions from the data can be located in the readme file of the repository.

Length

- 30,155 Data Points
- After Nulls, outliers, and data cleaning approximately 28,004 data points remain.
- ~7% of the data is removed from data cleaning as a result.

Data Timeline

- All house ages are within the years of 1900-2022.
- All house sales in the dataset are in the years of 2021-2022.

Business Problem:

- Zillow is looking to find ways to manage its inventory to curb future costs and understand how to improve pricing.
- Zillow has decided to hire a consulting data scientist to give recommendations on how to enter and behave within the target market.



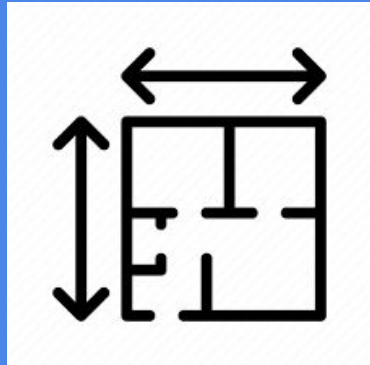
Business **Understanding**

- Zillow seeks to focus on the real estate market of King County in Seattle.
- Before looking for inventory, Zillow needs to understand how to determine the opportunity cost.



Features from Base Model - Numerical Data only

Square Footage of Home,
Basement, etc.



Location



Age








Renovations, Condition



Model Improvements - Add in relevant features

Whats the heat source?

TYPES OF HEAT SOURCES FOR HEATING SYSTEMS:

 FURNACES	 BOILERS	 HEAT PUMPS	 ACTIVE SOLAR HEATING	 ELECTRIC HEATING
A furnace heats air and uses a blower motor and air ducts to deliver warm air throughout the house.	A boiler heats water to provide hot water or steam for heating. Heat is distributed through a series of pipes.	A heat pump pulls heat from the surrounding air to warm the house. Heat also flows from the house cooling.	The sun heats a liquid or air in a solar collector to provide solar energy for space heating and hot water. Solar also has uses for home cooling.	Sometimes called electric resistance heating, electric heating is a process in which electricity is converted into heat.
EFFICIENCY 80-95%, depending on the system's age	EFFICIENCY 80-95%, depending on the system's age	EFFICIENCY 14-300%	EFFICIENCY Not applicable	EFFICIENCY 90-100%
LIFE EXPECTANCY 15-20 years	LIFE EXPECTANCY 15-20 years	LIFE EXPECTANCY 15 years	LIFE EXPECTANCY 20+ years	LIFE EXPECTANCY 20+ years
FUEL Natural gas, propane, heating oil, electricity	FUEL Natural gas, propane, heating oil, electricity	FUEL Electricity, geothermal energy	FUEL Solar energy	FUEL Electricity
PRO Inexpensive	PRO Easier to practice zone heating and cooling than if you live in your home.	PRO Can provide both your heating and cooling needs if you live in a moderate climate.	PRO Can reduce your fuel bill.	PRO Inexpensive equipment cost and high efficiency.
CON The system can be loud.	CON Expensive to install and require a serious commitment to a serious temperature management system.	CON Expensive to install.	CON Can require a special heating source, which makes the heat pump more costly.	CON Can be expensive to convert an existing system.

ENERGY.GOV

Is it on a waterfront?



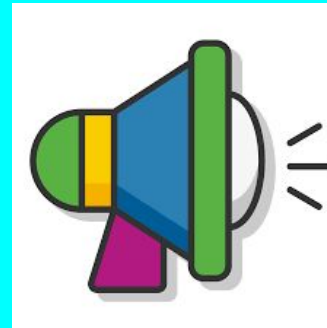
Does the house have a view?



Is it near a greenbelt?



Does the house have a nuisance?



Additional Model Improvements

Location location location!

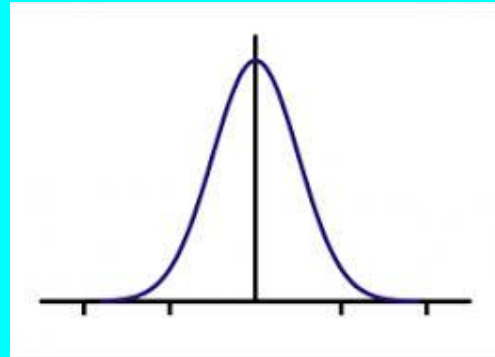
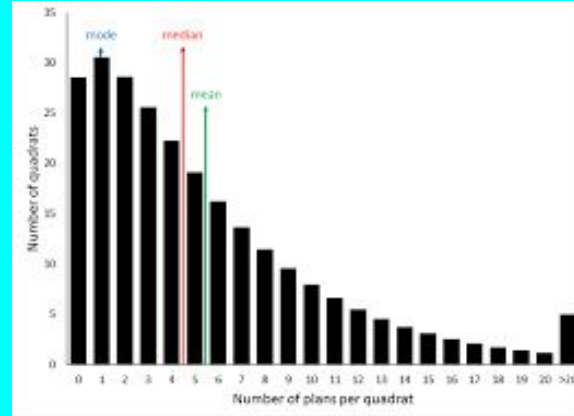
Lake Sammamish



Lake Washington



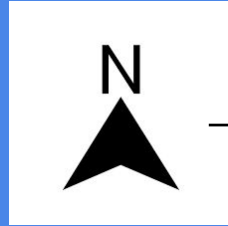
Unskewing the data: Separating the signal from the noise!



Conclusion and Interpretation of Final Model - Positive Impacts

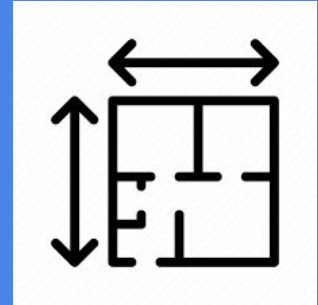
Most Positive Impacts on Price

- **Latitude - Houses Further North**
- **Water proximity - Lake Sammamish**
houses near this lake tend to have much higher prices than other houses.



Other Positive Impacts

- **Grade, condition - Higher Grade, better condition**
- **Square footage of the house apart from basement**
- **The number of bathrooms**
- **Square footage of the basement**
- **Size of the view from the house**



Conclusion and Interpretation of Final Model - Negative Impacts

Negative Impacts on the Price

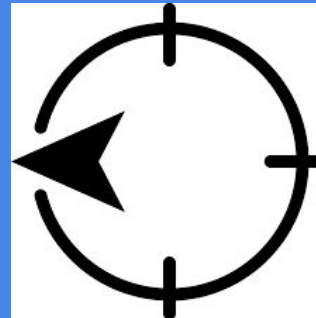
- **Nuisances**



- **Older houses in need of renovations and lower grade/quality tend to have lower prices.**



- **Houses further West**
- **Houses near Lake Washington**
(to be revisited later)



Recommendations

- Look at properties that are near Lake Sammamish or that are further north that also is accompanied with a waterfront.
- Try and buy older homes in the aforementioned areas as older homes tend to be cheaper in terms of price.
- Ensure the grade and condition are of high quality through either pre-assessed parameters or renovations
- Adding bathrooms can raise the price for resell value.
- Avoid buying houses near nuisances and further West as it may result in "holding the bag" scenarios leading to longer times held with inventory.

Questions Model Can Answer

```
graph TD; A[Questions Model Can Answer] --> B[Should the House be on a Waterfront?]; A --> C[How Far North Should the House Be?]; B --> D[How old?]; B --> E[What renovations?]; C --> F[How Far West?]; C --> G[What condition is good to buy or sell?];
```

Should the House be on
a Waterfront?

How old?

What renovations?

How Far North Should
the House Be?

How Far West?

What condition is
good to buy or sell?

Future Work

- In the future work, it is worth revisiting the value of the homes on the remaining waterfronts and seeing if there is any statistical significance. More exploration is needed but was not ready to be presented at this time.
- The views that are highlighted in the `column_names.md` documentation can be explored and onehotencoded and could be a potential candidate feature.
- Outliers of the dataset should be further explored.
- A look at school Districts was initiated to see if the rating of the school district was correlated with housing prices. Data was scraped from the GreatSchools API and will be further explored.