Kaggle - Zillow's Home Value Prediction (Zestimate)

🖉 Book	
⊙ Category	Data Science Projects
🖉 Cover	
🗓 Created	@Mar 11, 2021
☰ Description	Can you improve the algorithm that changed the world of real estate?
☑ Featured	☐
☰ Tags	

- **What should be in the presentation?**

  1. A clear explanation of the problem and why its important.

  2. A summary of the data cleaning and exploration, including visualizations

  3. How I created a baseline

  4. My logic for selecting models to test, tuning models that I selected, and measuring the efficacy, e.g. why did I choose to measure the model using RMSE over R^2?

  5. The results I got for each model, also include a visualization.

  6. My approach to training the final model and making predictions.

  7. A summary of the project: results, findings, and areas that could be improved or explored in the future.

Zillow House Estimate Project

Zillow is a leading real estate and rental marketplace dedicated to empowering consumers with data, they serve the full lifecycle of owning and living in a home: buying, selling, renting, financing, remodeling, and more.

Zillow estimates the value of their homes and calls it "Zestimate".

From Zillow:

> " *"Zestimates" are estimated home values based on 7.5 million statistical and machine learning models that analyze hundreds of data points on each property."*

In 5/24/2017 Zillow has initiated a Kaggle competition in order to improve their Zestimate model.

This is my take on their competition.

The Problem

How much is a house really worth? answering this question can be a difficult task for ordinary people trying to purchase their first house.

A house is worth whatever people will pay for it and it depends on the market, the quality of the house, the location, and so on.

Unfortunately, it also depends on who you're asking, whether you're asking a lender, an agent, or a county tax assessor.

In other words, knowing how to calculate and estimate a home's value independently can have huge benefits to consumers who are trying to get by and buy a house. It can bring *control* and *power* in negotiations if consumers would know how much the house that they are interested in is actually worth.

The Data

In order to estimate Zillow's houses, in this competition, we were given a dataset of 2919 records of houses where each house is defined by 81 features.

These features are mostly indications of wealth and house quality, such as: having a pool or not, the square-foot area of the street around the house, overall condition, year built, and so on.

```
df.sample(5)
```

Sample Data

Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	PoolArea	PoolQC
1324	30	RL	50.0	5330	Pave	NaN	Reg	HLS	AllPub	...	0	NaN
539	20	RL	NaN	11553	Pave	NaN	IR1	Lvl	AllPub	...	0	NaN
1384	30	RL	NaN	25339	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN
1290	60	RL	86.0	11065	Pave	NaN	IR1	Lvl	AllPub	...	0	NaN
1053	60	RL	100.0	9500	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN

5 rows × 81 columns