Housing Prices in Ames, Iowa

Adrian Garrido

Objective

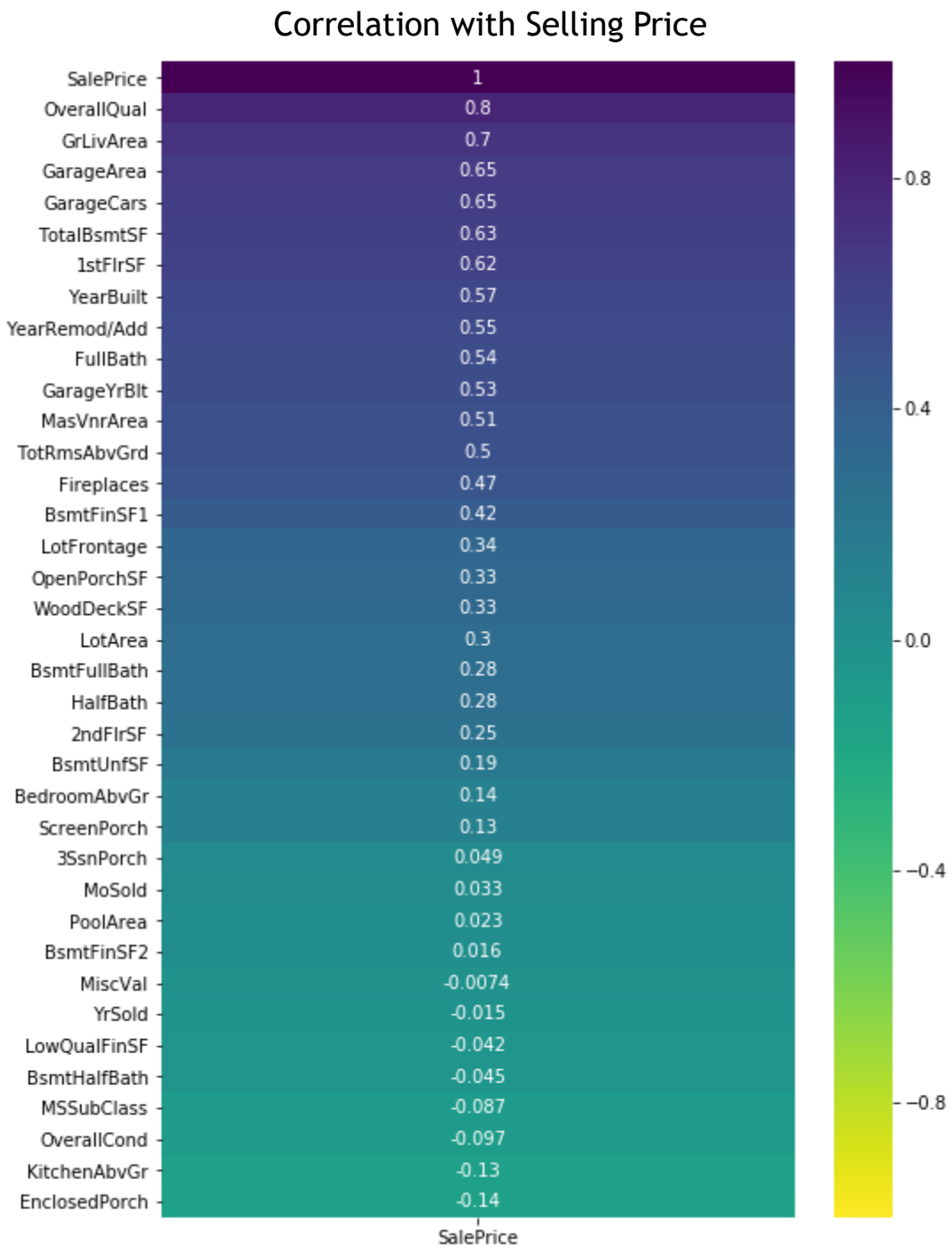
Creating a model that is able to predict selling price of a property as *accurately* as possible from the test dataset

Cleaning and EDA

- Using the dictionary to interpret and change null values.
- Making assumptions

Logic vs. Numbers

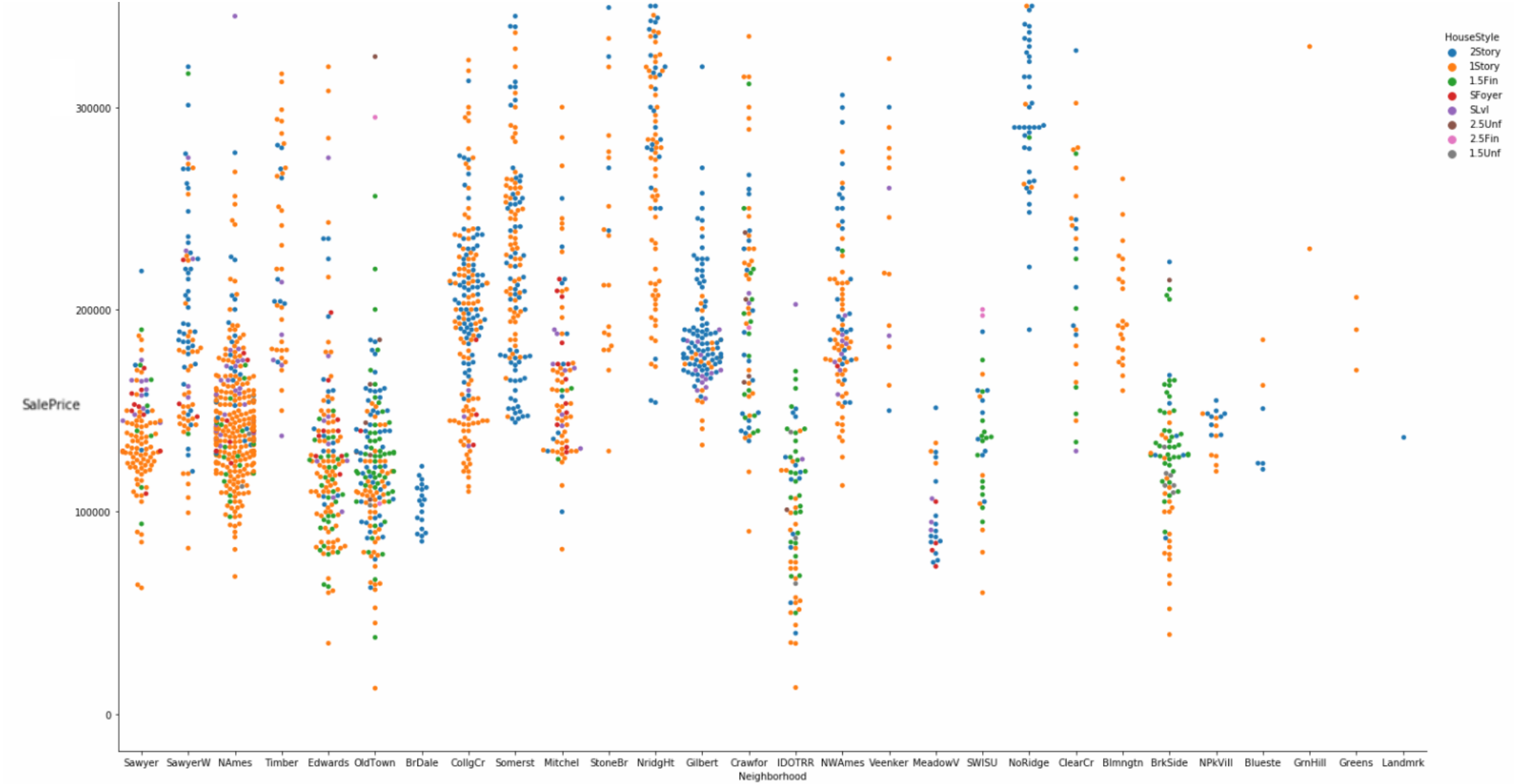
- What things do we ***believe*** to affect property prices?
- What things have the highest correlation with the price?



Categorical Columns

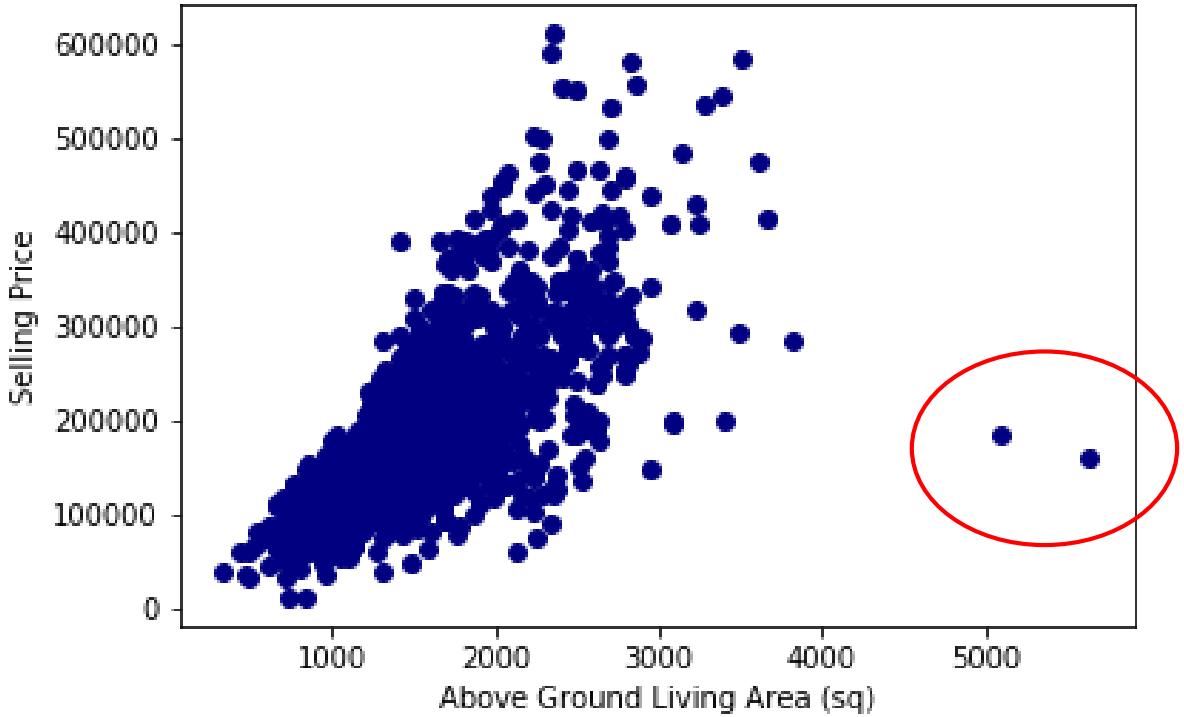
- Assigned a numerical value to each possible outcome on categorical columns
- Example for Pool Quality column:
  - Excellent = 4
  - Good = 3
  - Average/Typical = 2
  - Fair = 1
  - No Pool = 0

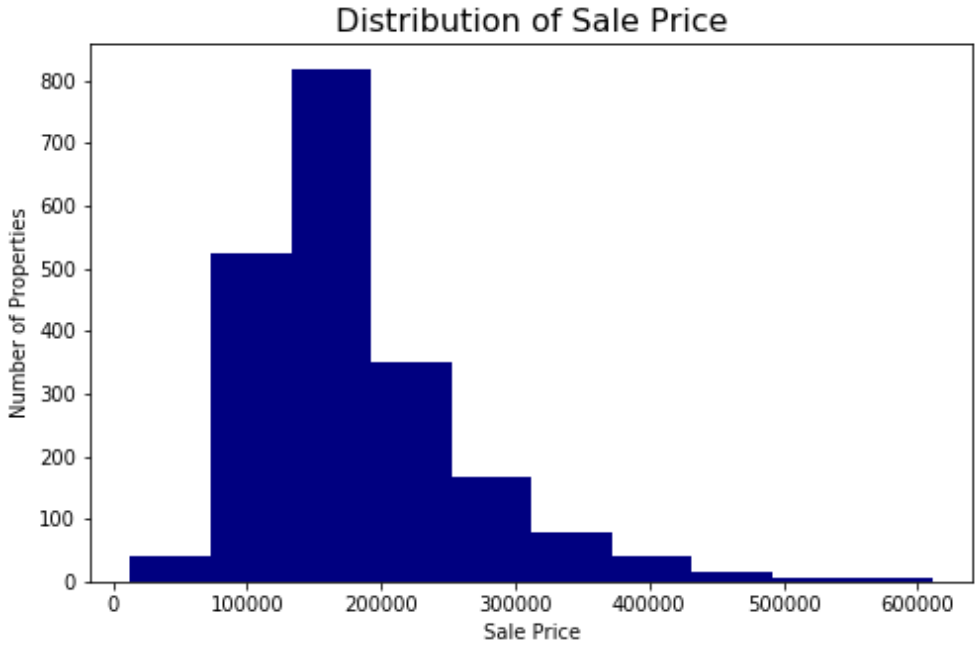
Neighborhood Distributions

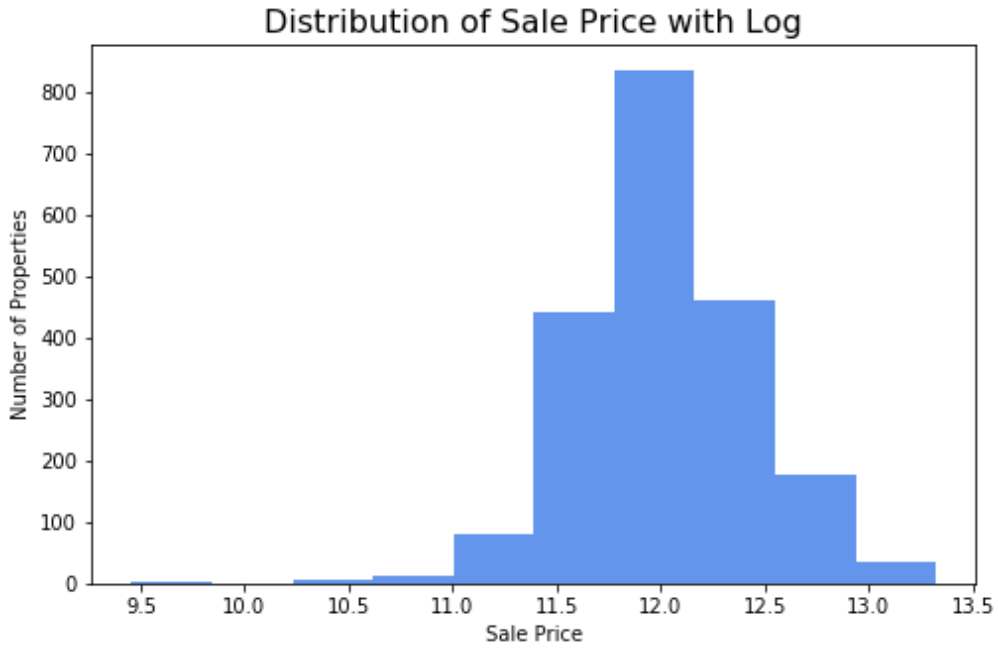


Outliers

- Getting rid off outliers. Why or why not?







In [ ]:
```
for c in train._get_numeric_data():
    if train[c].skew() > 0.90:
        train[c] = np.log(train[c] + 1)
```

Making distributions more normal

Applying the natural logarithm

Train Test Split?

Conclusion and Results

Key components of my model:

- Proper Data Cleaning
- Log Transformations
- Feature Selection
- Lasso Model

#	Team Name	Score	Entries	Last
1	DTrichter	18687.68913	44	10m
2	minion_of_boom	19059.30654	10	1d
3	Joey Romness	19075.81471	55	2h
4	Bruno Santos	19293.42942	23	4h
5	Laura Luo	19403.63562	5	3h
6	Johannes Huessy	19631.00167	30	2d
7	Stephen Tse	19812.98562	29	4h
8	Andrew Picart	19822.16606	10	1h
9	Melissa	19837.30056	25	7h
10	Tonya	19937.16207	18	5h
11	Ari Mello	20014.70264	7	1d
12	Nate Gunawan	20014.70264	49	7h
13	AQQU	20023.82126	12	10h
14	Chris Birch	20056.23922	9	2d
15	Tony	20414.09433	11	2h
16	Boom Devahastin Na Ayudhya	20627.33437	2	9d
17	Marielle Marcus	20672.46260	35	3h
18	Adrian	21149.52374	34	2h

Your Best Entry ↑