Virtual Appraisal Tool

A simple Linear Regression model for use in virtual appraisals.

E-service Appraisal Update's

We have been asked to develop a simple Linear algorithm whose inputs can be obtained in a 2 minute online survey.

- **What Survey?** This year we launched a new web service to provide customers with automated appraisals, and document leads for our acquisitions teams.
- **Why?** Our current digital appraisal tool's participation rates are abysmal. Our UIX team has indicated that the current survey is likely to long, and asks to many probing questions, to get the kind of participation that we need.
- **What to do?** Our previous model was optimized for predictability, and so it's not surprising that UIX's suspicions proved to be true in our most recent round of customer surveys. In order   to increase participation, we needed to reduce the complexity of the model, and select for predictive variables that customers would share freely without to much hesitation, or that is easily obtainable in most internet listings.

How did we do? Let's take a look.

Table of Contents

So Many Questions

Our in-house predictive model uses 200 variables derived from 70 unique features. That's 70 questions we were expecting our users to answer in what is supposed to be a quick survey.

In addition , consider the needless cost of the intensive computation of using our enterprise level prediction software for uncommitted clientele? Our acquisitions team does a full follow up on all of our internet leads, so more information can be obtained then , when commitments from clients are more likely already in place!

Why Waste Resources ??

01
Feature Selection

Correlation Map

Let's Take a Look

Mean by Feature

Central_Air	Lot_Area	SalePrice	Sq_ft	Has_Garage	Large_Lot
0	8647.163121	100575.078014	1892.964539	0.687943	0.092199
1	9995.394226	187407.570604	2591.019948	0.963255	0.221522

Large_Lot	Lot_Area	Central_Air	SalePrice	Sq_ft	Has_Garage
0	8176.560521	0.920546	165525.296710	2399.452514	0.933582
1	16294.337931	0.970115	240301.735632	3074.213793	0.983908

Has_Garage	Lot_Area	Central_Air	SalePrice	Sq_ft	Large_Lot
0	7677.166667	0.614035	106065.991228	1994.412281	0.061404
1	10033.788302	0.949793	185870.075052	2575.278468	0.221532

Showing **Mean Sale Price**
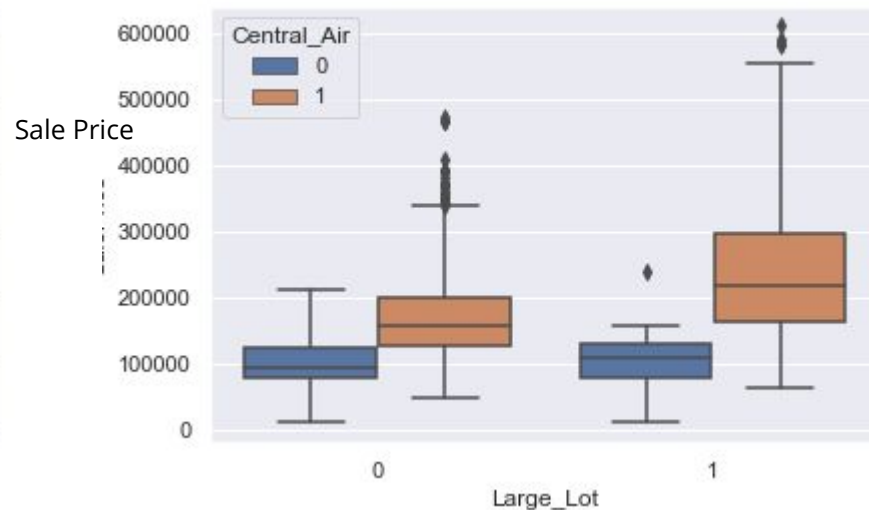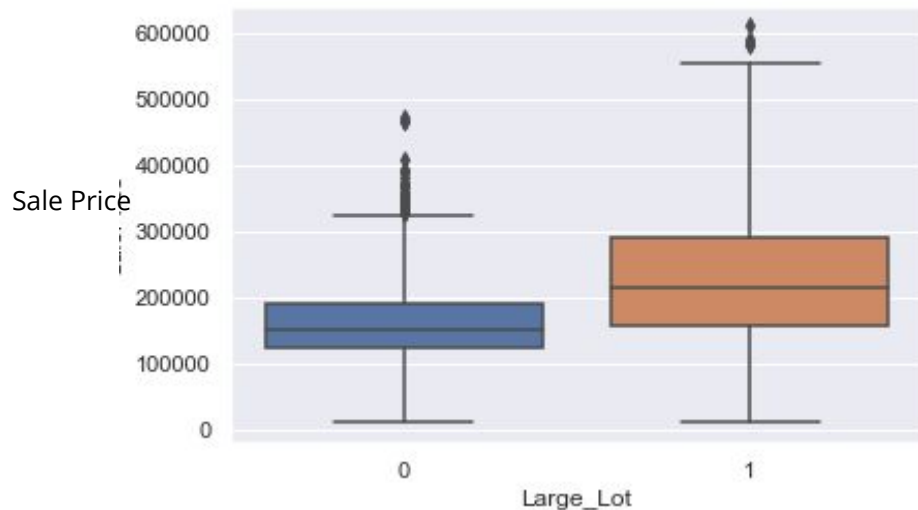To be significantly influence by the presence or absence of selected features.
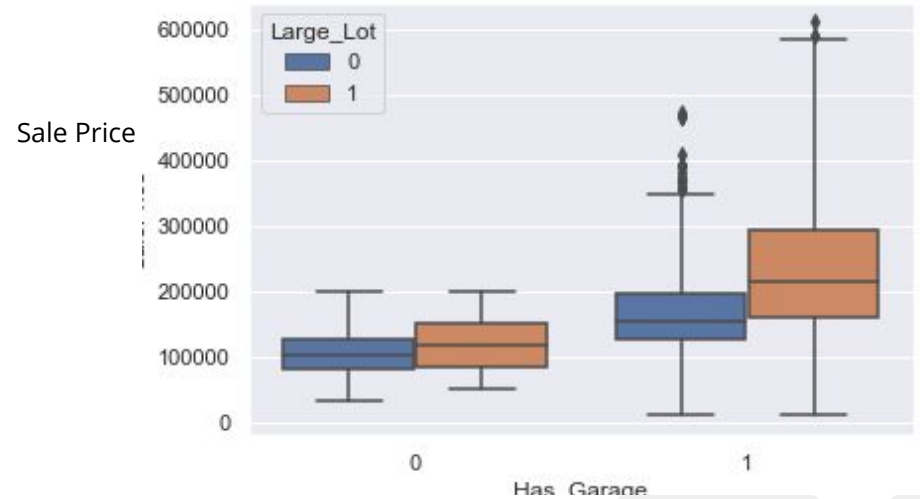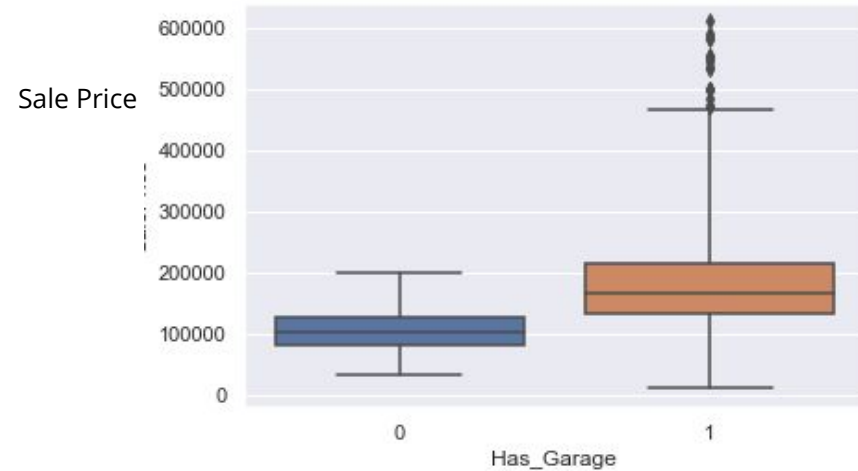
02
Visualizing Relationships

Large Lot & Central Air

A Large Lot is classified as having a lot area greater than the 75th percentile (12,000sq_ft)
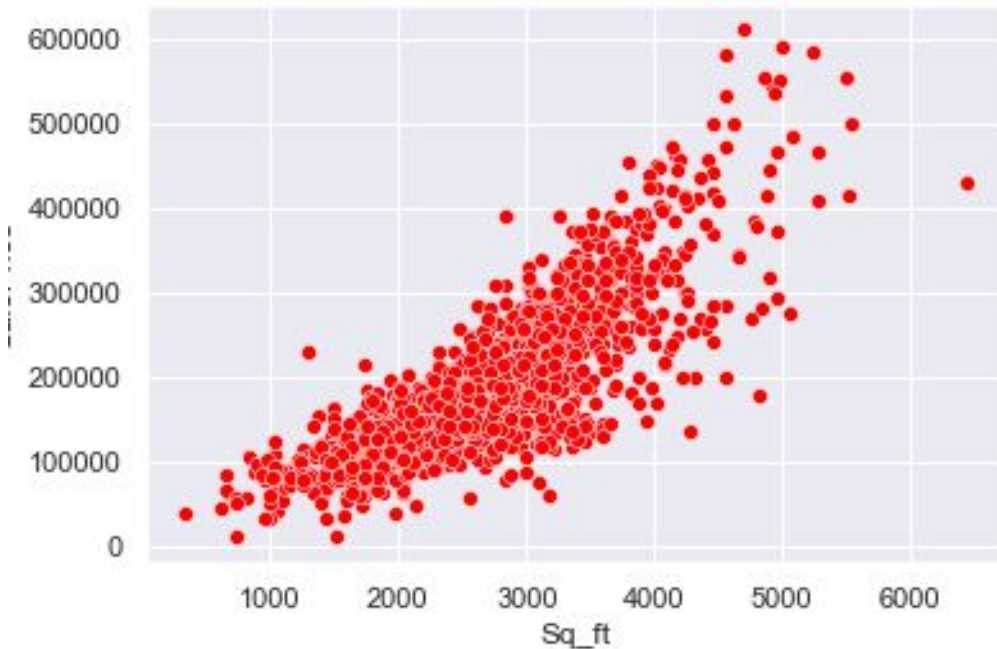
Garage

Having a Garage has an independent effect on Sale Price.

Sq_ft and Sale Price

The only non-engineered numerical feature used in our final model has a strong linear relationship with Sale Price.

Sale Price

03

Model Results

Final Model

42k

MSE
Mean Error in Model Predictions

70%

Train Score
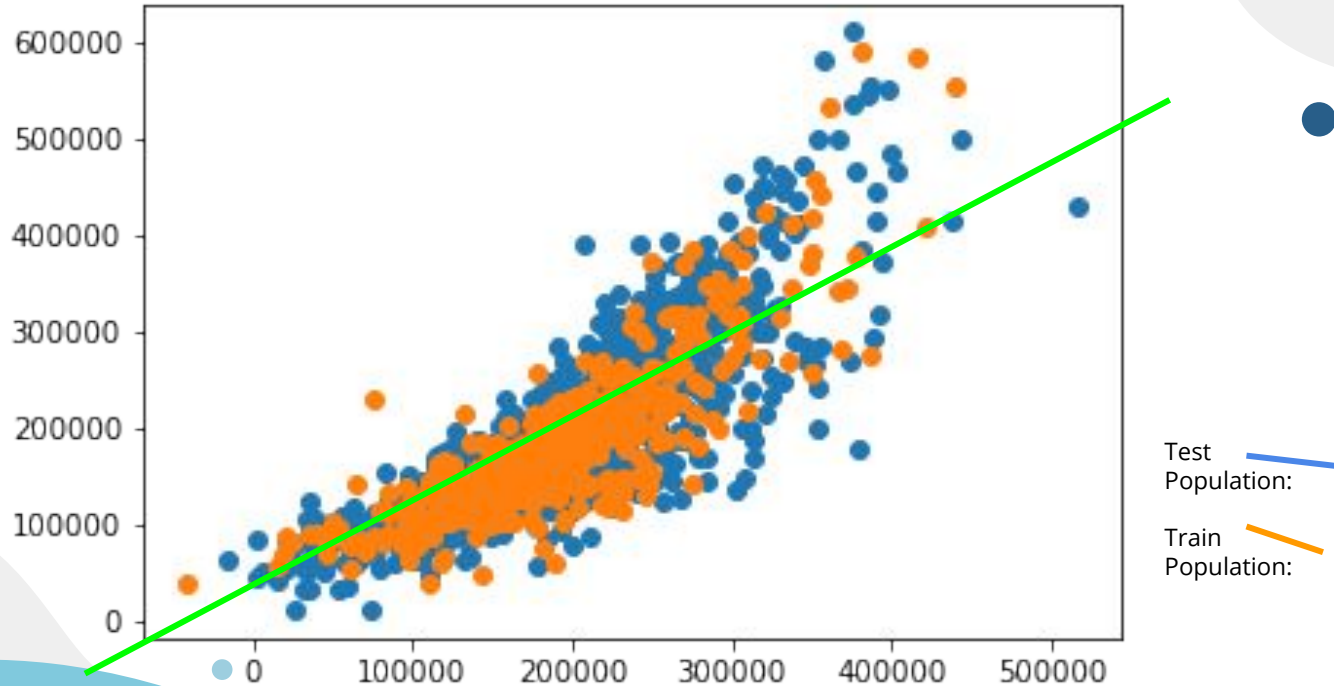More features are needed to improve our model.

70%

Test Score
Our model is just as good at generalizing to new data as it is on unseen data!

	Id	Lot_Area	Central_Air	Sq_ft	Has_Garage	Large_Lot	Finished_Basement
0	2658	9142	0	2948	1	0	0
1	2718	9662	1	3934	1	0	0
...	...	...	...	...	...	...	...
876	1672	9000	1	1728	1	0	1
877	1939	8400	1	1848	1	0	1

03
Model Results

Unintended Products!

Test Population:

Train Population:

04

Next Steps

Team Effort

UIX

- Determine the optimal number of survey questions to maximize participation and data collection.

- Provide ongoing feedback to Analytics team's on survey performance and focus panel results.

Analytics

- Continue to identify the strongest predictive variables in isolation to minimize the number of survey questions needed.

- Stay up to date with UIX recommendations for survey length , and keep UIX up to date on participation metrics.

Marketing

- Strong focus on driving participation in our lead-generating E-services in the upcoming ad campaign.