Periodically Data Modeling
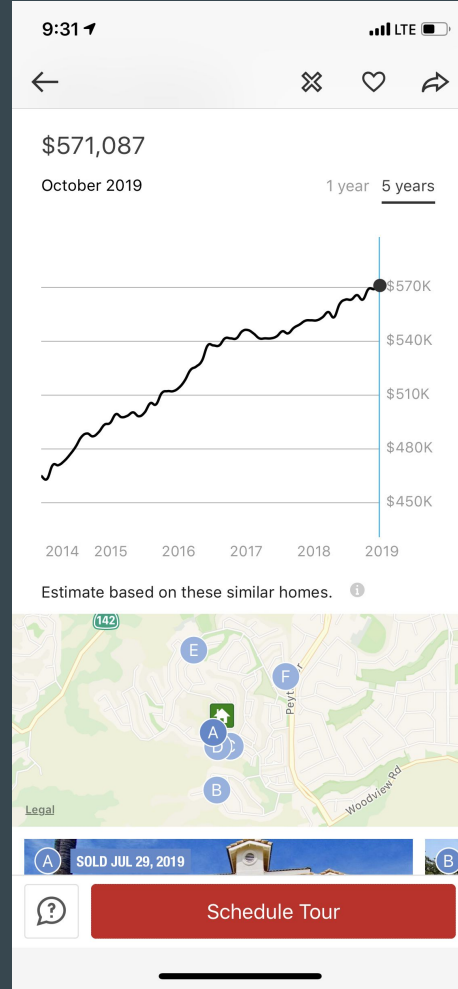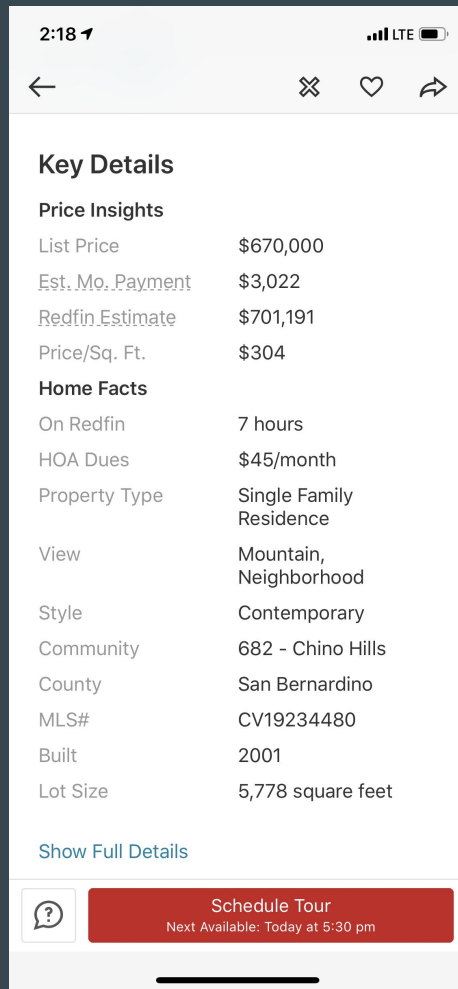
●●●

Echo Zhao

Problems We Are Solving

As a Real Estate App Company, price factor are important for:

- Content Listing as first impression
- Price predictor for investors
- Increase customer reliability on the app.
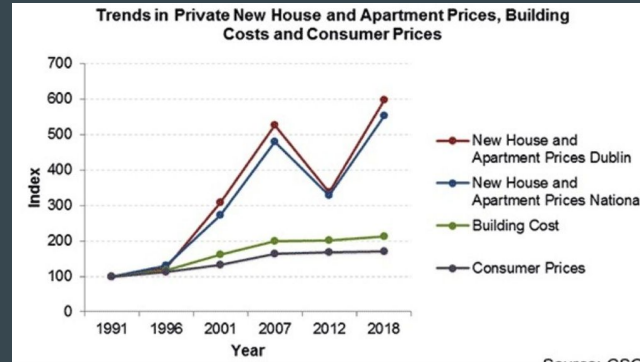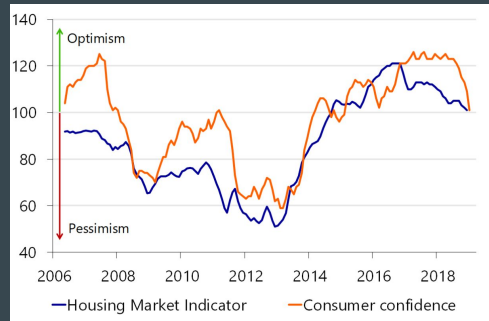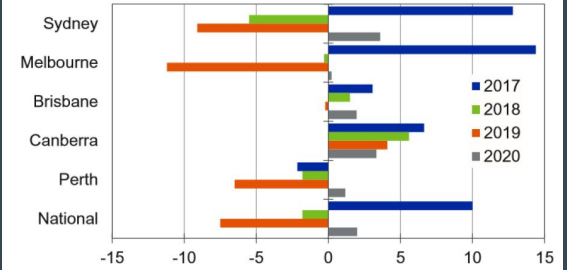
Why Do We Need To Update?

Time, location, policies change make model inaccurate over time

Data Given:
Ames housing 2006-2010
What we can tell:

- Price range $12,789 to $611,657
- Average: $181,469
- Median: $162,500

Missing Data: 26 columns
Top 5:
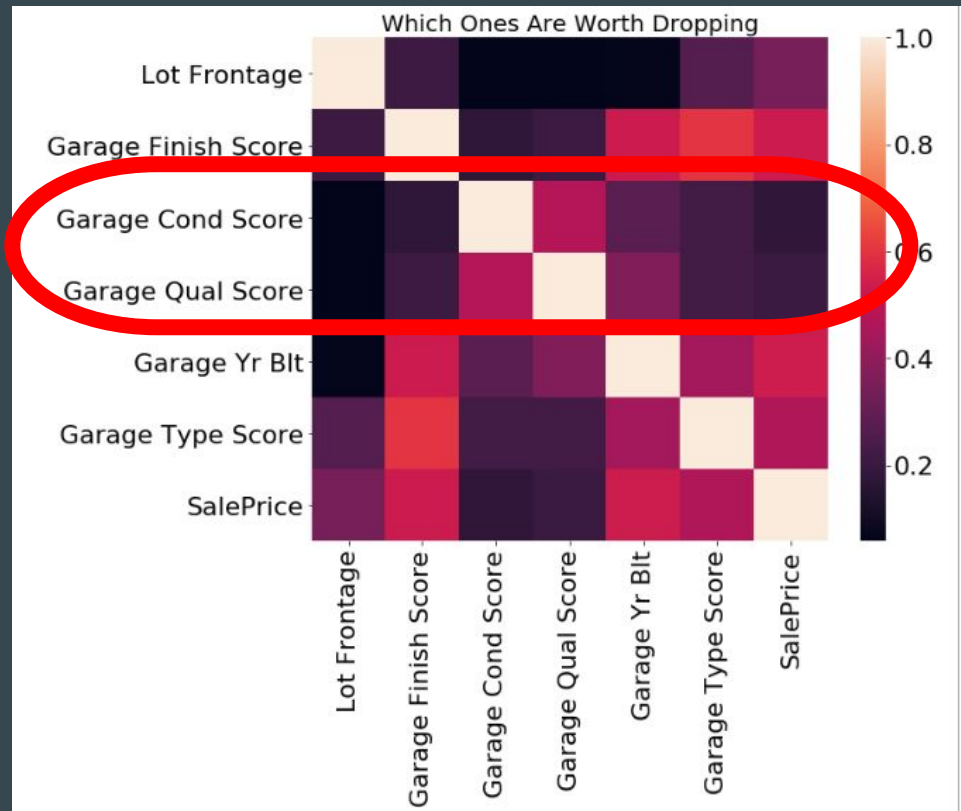Pool QC 2042
Misc Feature 1986
Alley 1911
Fence 1651
Fireplace Qu   1000
Total Data: 2051

Which Ones Are Worth Dropping

Dropped the remaining 5% rows

Columns with category
Gd:5,
TA:4,
Ex:3,
Fa:2
Po':1

Correlation between Factors
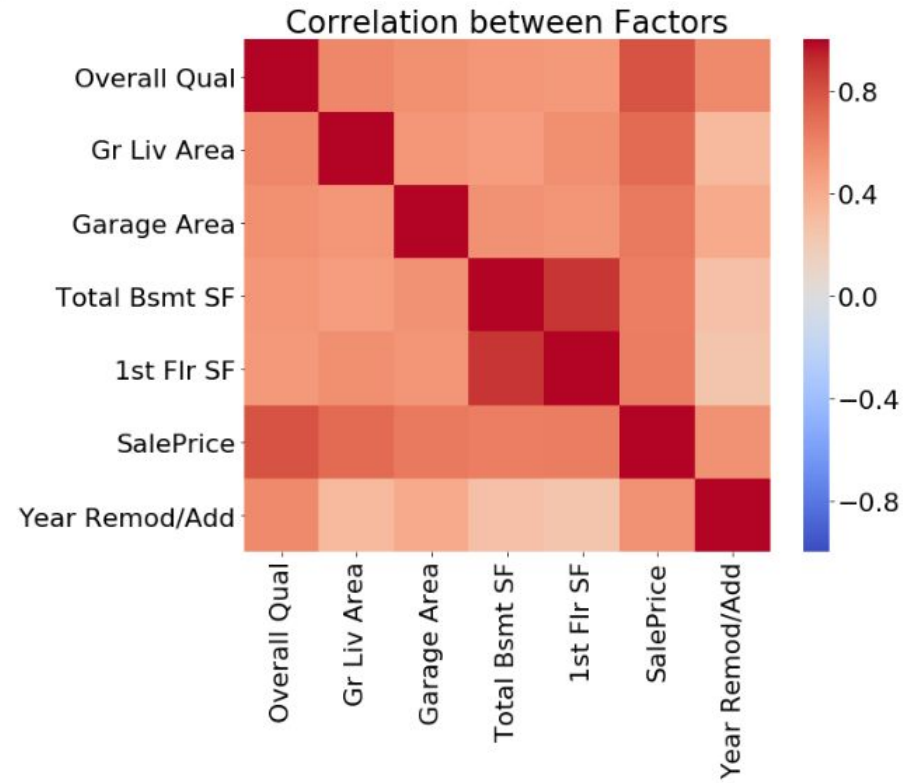
Feature Engineering (Polynomial and Lasso)

	Predictor Variables	Coef_weight	absolute
143	Overall Qual Gr Liv Area	18267.51	18267.51
468	Total Bsmt SF Neighborhood Score	14519.41	14519.41
139	Overall Qual Total Bsmt SF	12731.17	12731.17
596	Gr Liv Area Sale Type Score	6738.40	6738.40
331	BsmtFin SF 1^2	6562.00	6562.00
184	Overall Cond Gr Liv Area	6308.71	6308.71
472	Total Bsmt SF Bsmt Qual Score	-6114.02	6114.02
215	Year Built Year Remod/Add	5503.59	5503.59
330	Mas Vnr Area Bsmt Qual Score	-4308.70	4308.70
154	Overall Qual Garage Area	3844.34	3844.34
127	Lot Area Heating QC Score	3644.12	3644.12
594	Gr Liv Area Neighborhood Score	3483.76	3483.76
849	Garage Area Screen Porch	3301.35	3301.35
153	Overall Qual Garage Cars	2735.12	2735.12
53	MS SubClass Bsmt Unf SF	-2676.88	2676.88
151	Overall Qual Fireplaces	2455.66	2455.66
302	Mas Vnr Area Bsmt Full Bath	2377.78	2377.78
363	BsmtFin SF 1 Neighborhood Score	2264.64	2264.64
174	Overall Cond Year Built	1951.98	1951.98
778	TotRms AbvGrd Sale Type Score	1777.09	1777.09

Model Scores

- Sample Score(Cross_val):0.89
- Model Score(train) R2: 0.91
- Test score(holdout)R2: 0.90
- Residual Mean: 35

Residual Mean

Predicted Values vs. Actual Values
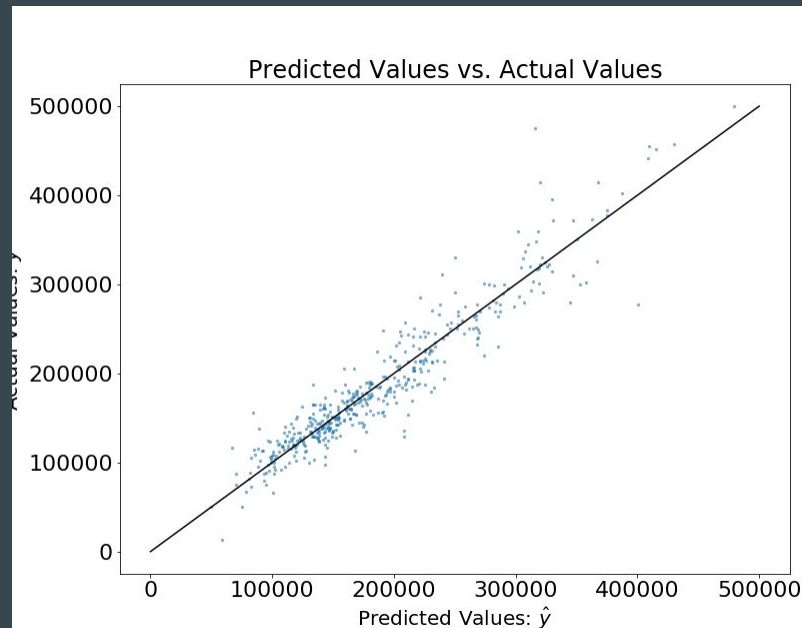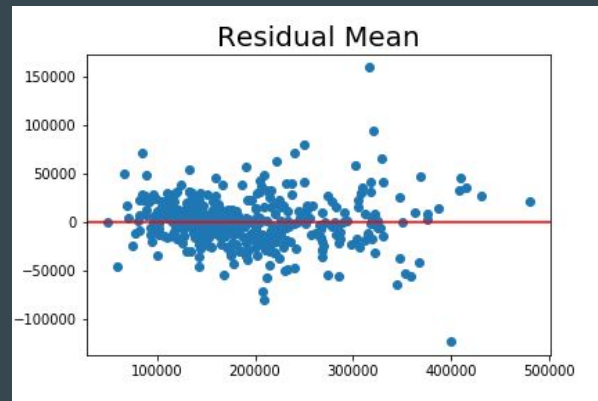
Proposed Price Determine Factors
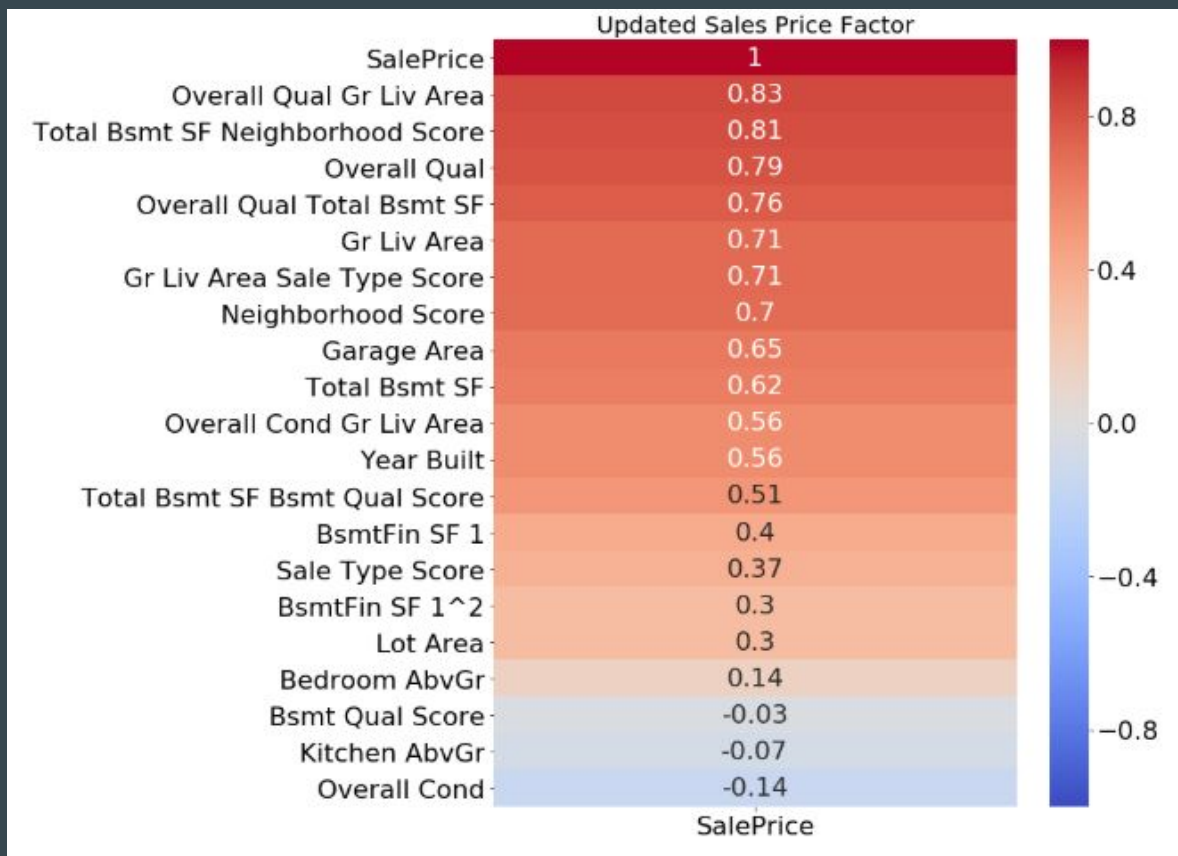
Summary

We can not tell Ames is affected by 2008 Recession
Housing is mainly based on housing quality, basement and neighbourhood
Combined features sometime bring more value to the house

Recommendation

- To better understand the housing insights, we need to breakdown Overall Quality in the given information
- For company, we can show features on the front page along with house pictures to gain user loyalty
- Use updated features to show users housing prediction
- Get partnership with remodel companies to bring company more profit