



“Homebuying Powered by Data Science”

Alice Zhang

[in/xuyuan-zhang](https://www.linkedin.com/in/xuyuan-zhang)

[github.com/AliceXuyuan](https://github.com/AliceXuyuan)

Joe Lu

[in/joezlu](https://www.linkedin.com/in/joezlu)

[github.com/jzl4](https://github.com/jzl4)

Jordan Runge

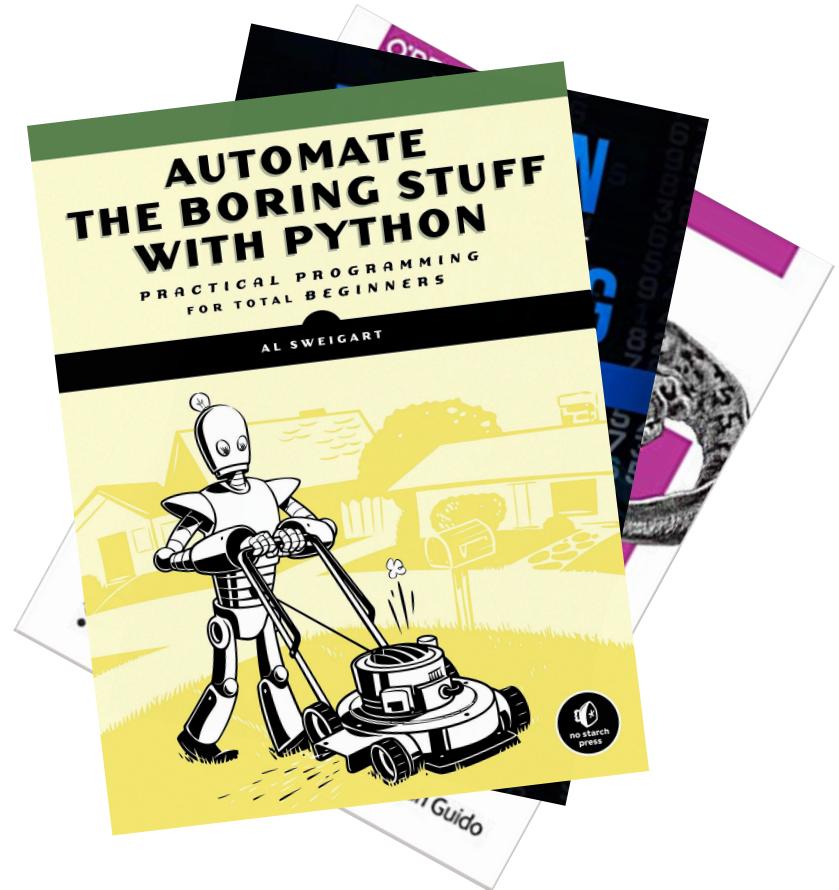
[in/jordanrunge](https://www.linkedin.com/in/jordanrunge)

[github.com/jordanrunge](https://github.com/jordanrunge)

Yunmei Zhang

[in/yunmeizhang](https://www.linkedin.com/in/yunmeizhang)

[github.com/zhangym1256](https://github.com/zhangym1256)



# Data Science



# Art of Reality



Data Abundance

Computing Power and Tools

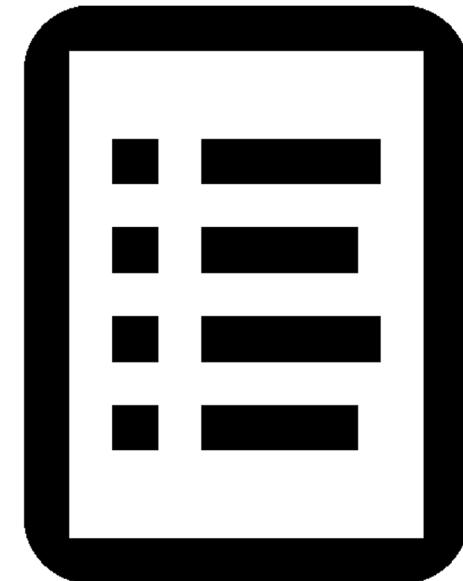
\$35 Trillion Market



**1460** Homes



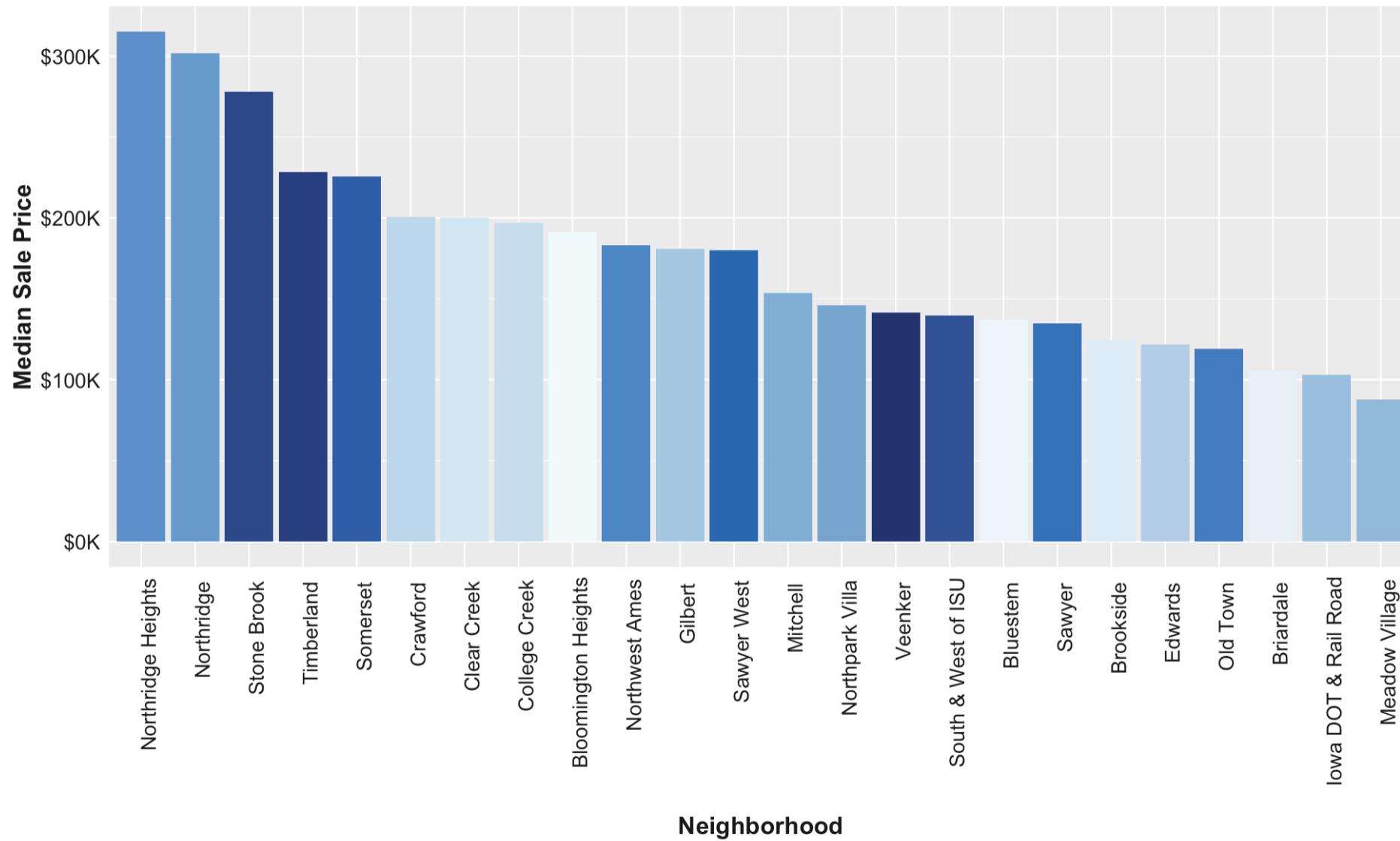
**80** Features



Cool. But how can we make sense of all this data?

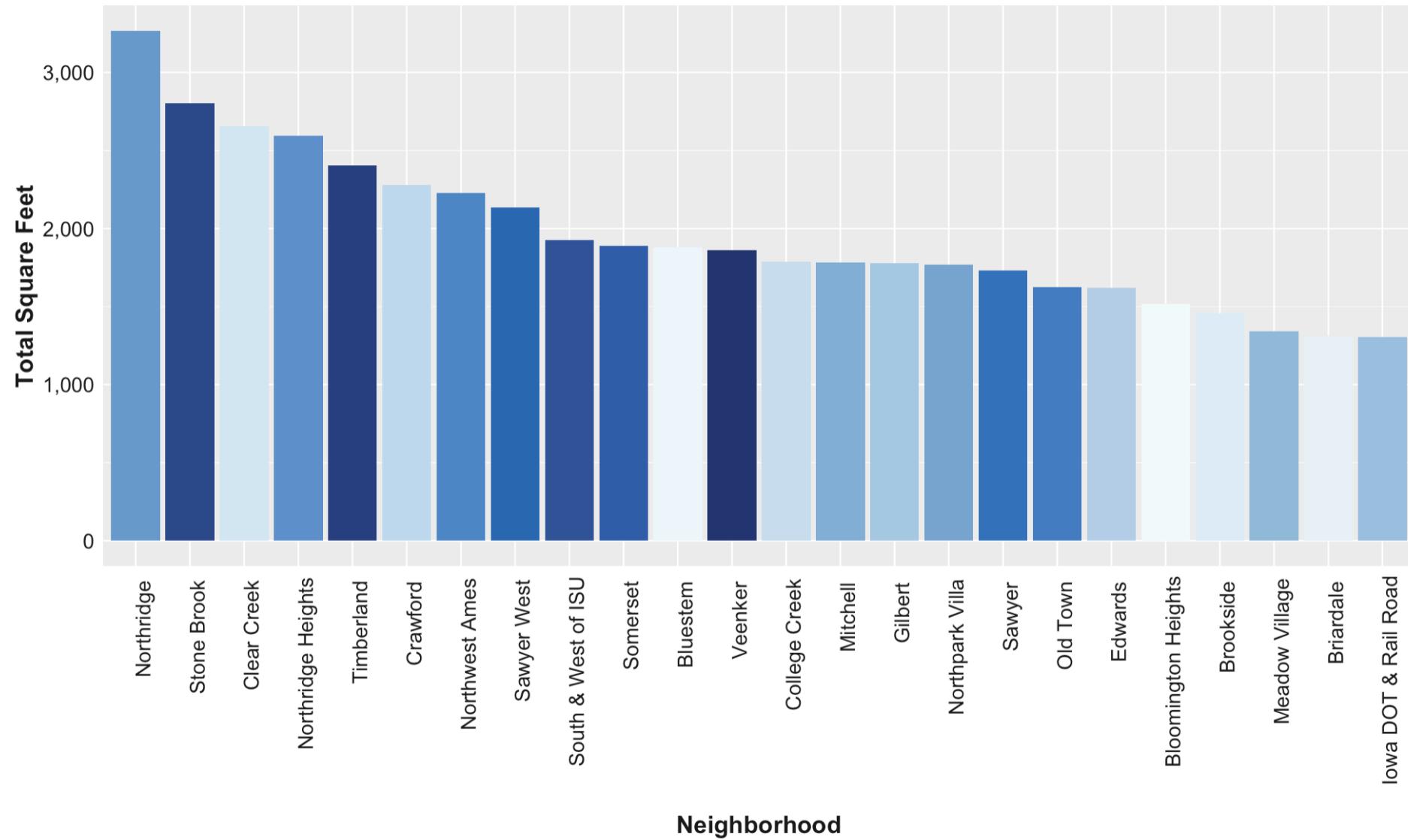
## Median Home Price by Neighborhood

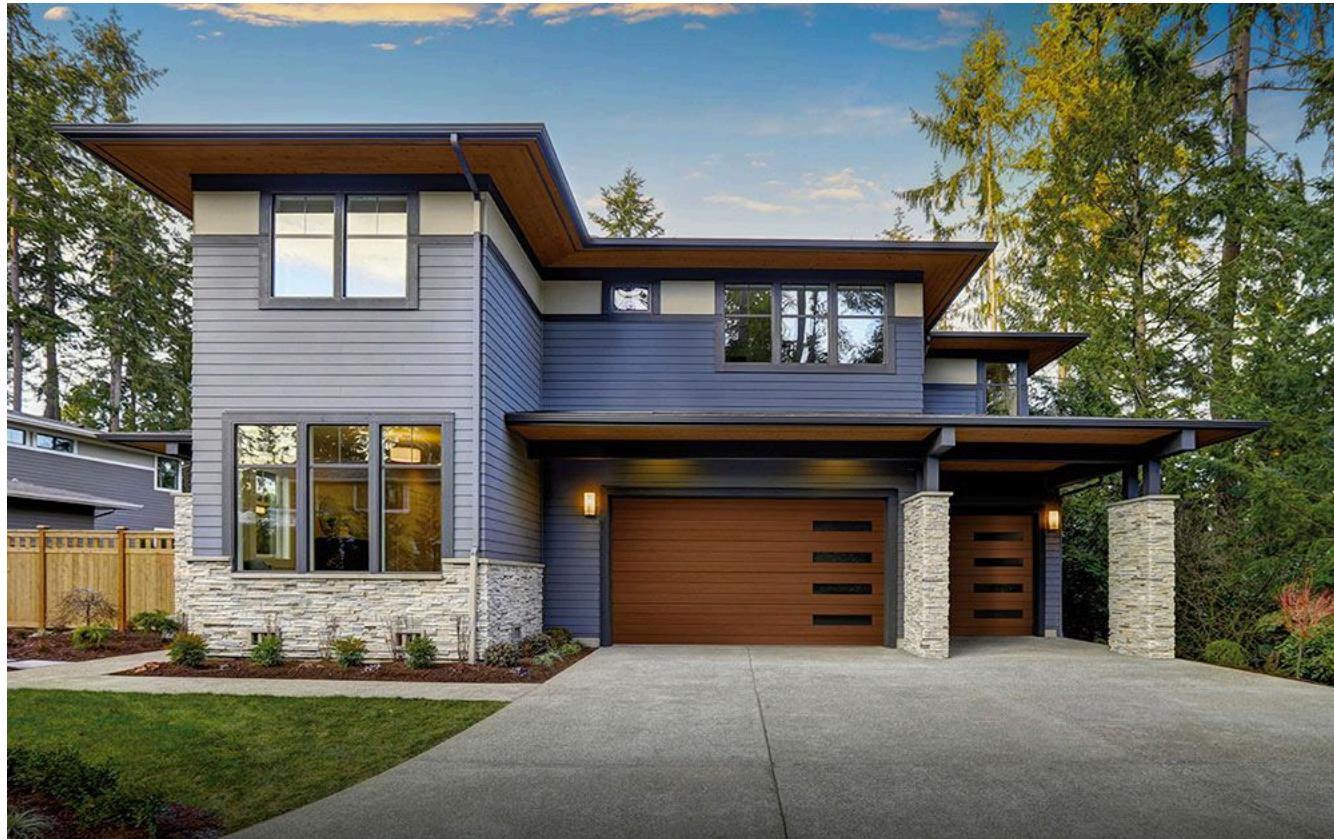
Homes sold between 2006-2010



## Median Square Feet by Neighborhood

Homes sold between 2006-2010



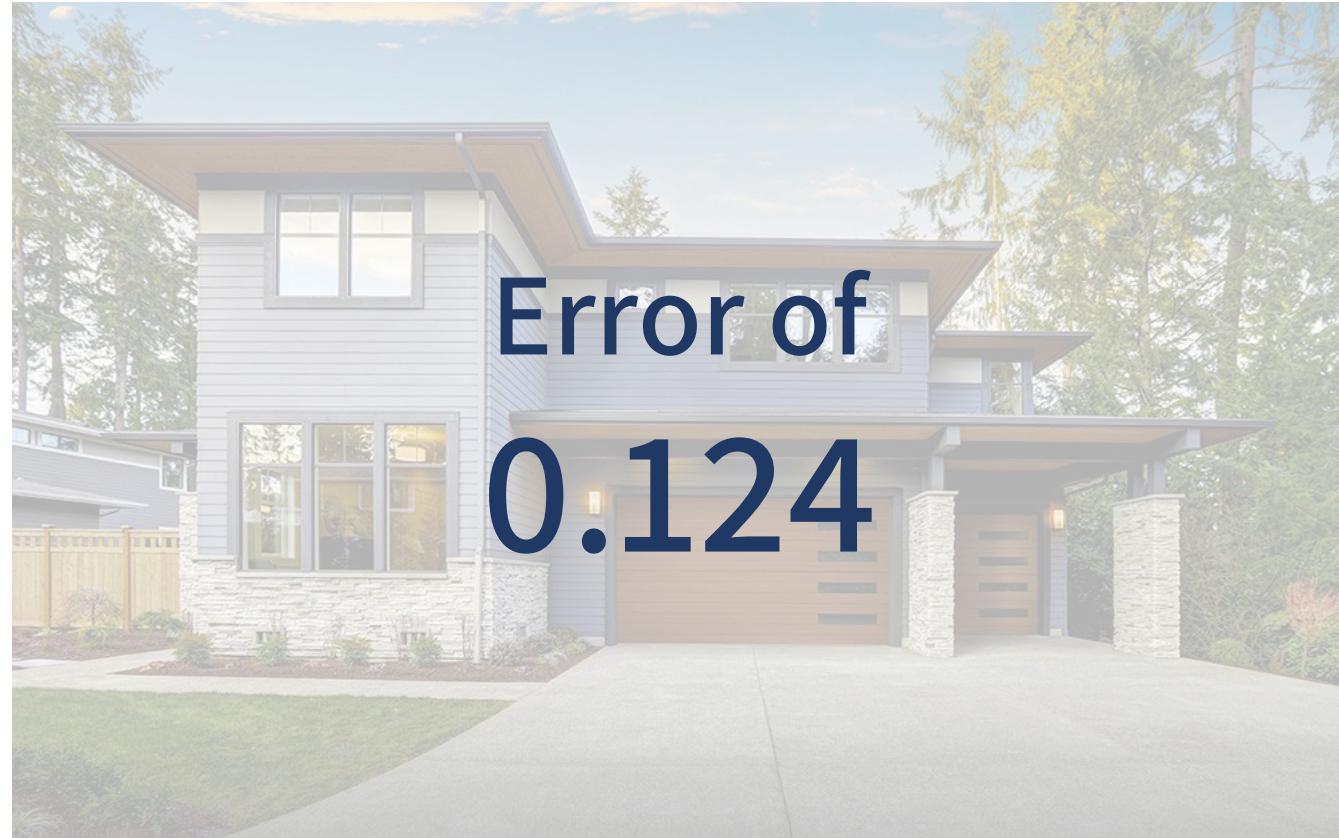


Interior		
Features	How 'big'?	How 'good'?
Kitchen		1
Bathrooms	1	
Bedrooms		
Basement	1	2
Living Areas	1	
Utilities		1
Misc Inside	1	1
	4	5

Exterior		
Features	How 'big'?	How 'good'?
Exterior		2
Garage	1	1
Lot	2	
Roof		
Outdoor Areas	1	
Misc Outside	4	3

Other	
Features	How 'good'?
Location	2
House 'Meta'	2
Sale Context	1
	5

21 Most Important Features



How do we know we're right?

Let's get (a bit) technical.

1 Clean Data

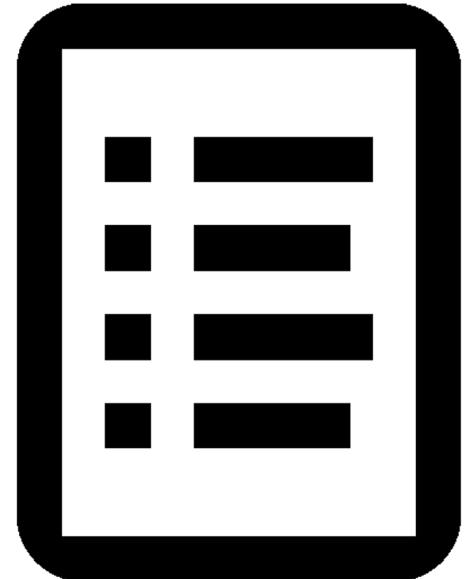
2 Select Data

3 Create Models

4 Assist Homebuyers

Linear  
Tree-based

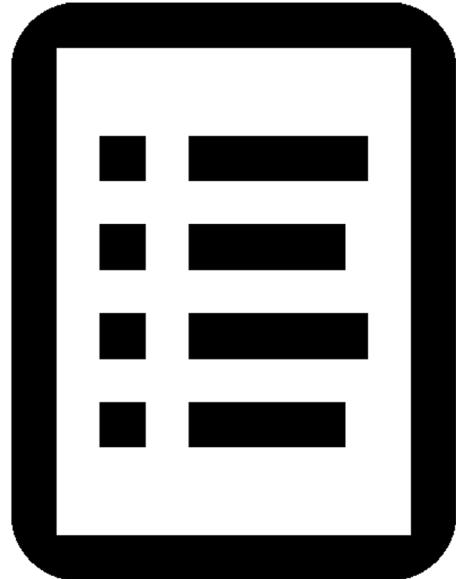
80 Features



6% Missing

Variable	Method
Quality of Basement	<u>Missing values are meaningful</u> because there is no basement
Zoning	Look relevant feature (neighborhood), impute <u>most frequent value</u>
Area of Mason Veneer	Look at feature relevant feature (exterior type), <u>impute median value</u>

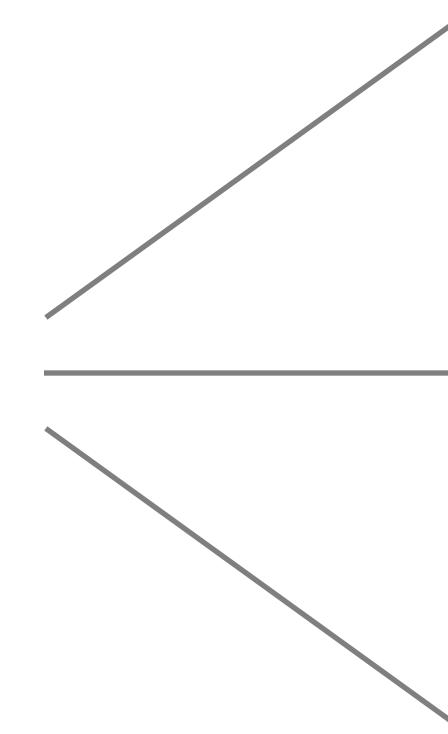
80 Features



33 Continuous

14 Ordinal

33 Categorical



1  
Clean Data

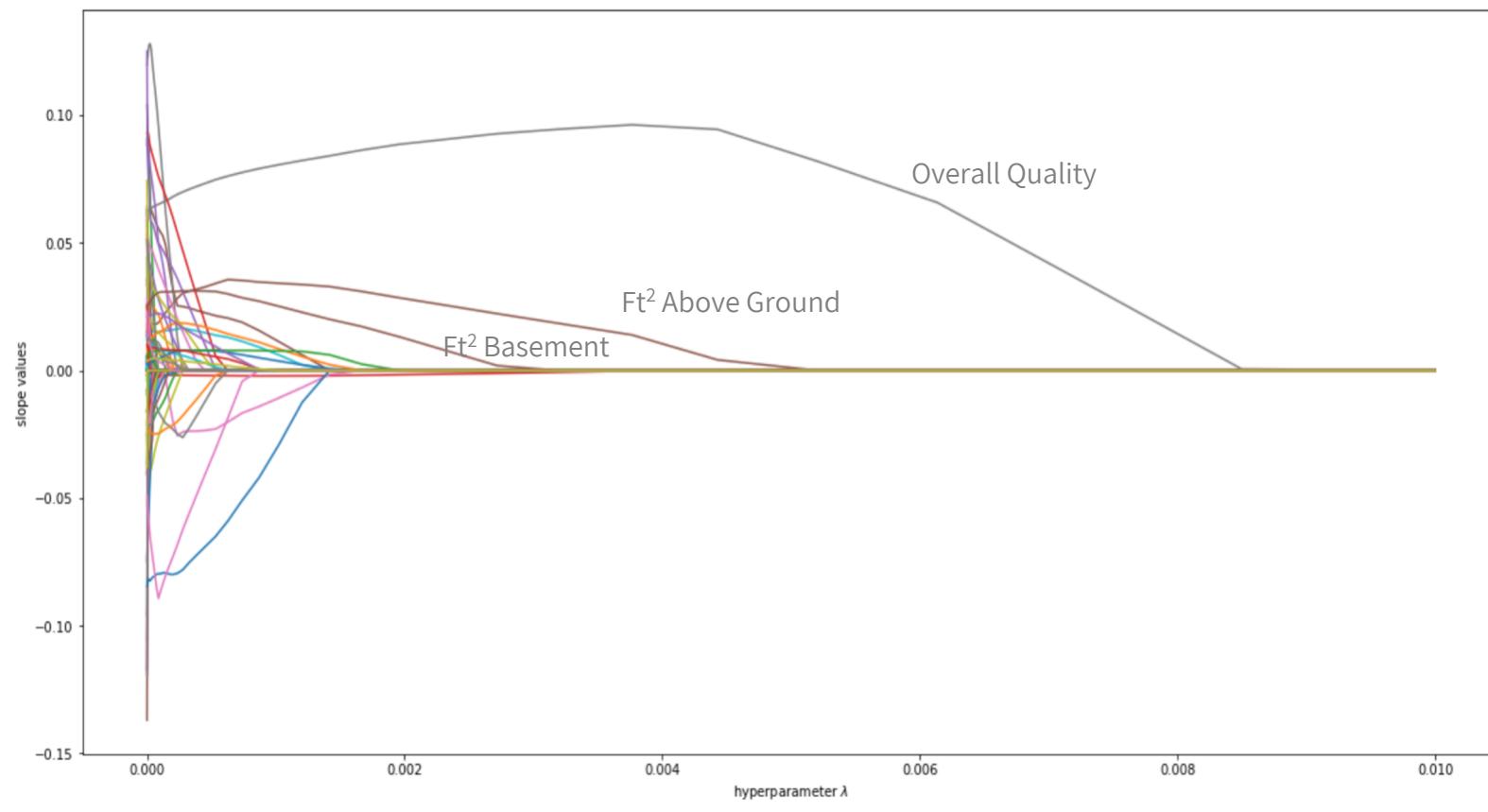
2  
Select Data

3  
Create Models

4  
Assist Homebuyers







1. Overall Quality
2. Ft<sup>2</sup> Above Ground
3. Ft<sup>2</sup> Basement
4. Total Bathrooms
5. Ft<sup>2</sup> Garage
6. Age of House
7. Kitchen Quality
8. Fireplace Quality
9. Ft<sup>2</sup> Lot
10. Height of Basement
11. Exterior Quality
12. Number of Fireplaces

1 Clean Data

2 Select Data

3 Create Models

4 Assist Homebuyers



1

Clean Data

2

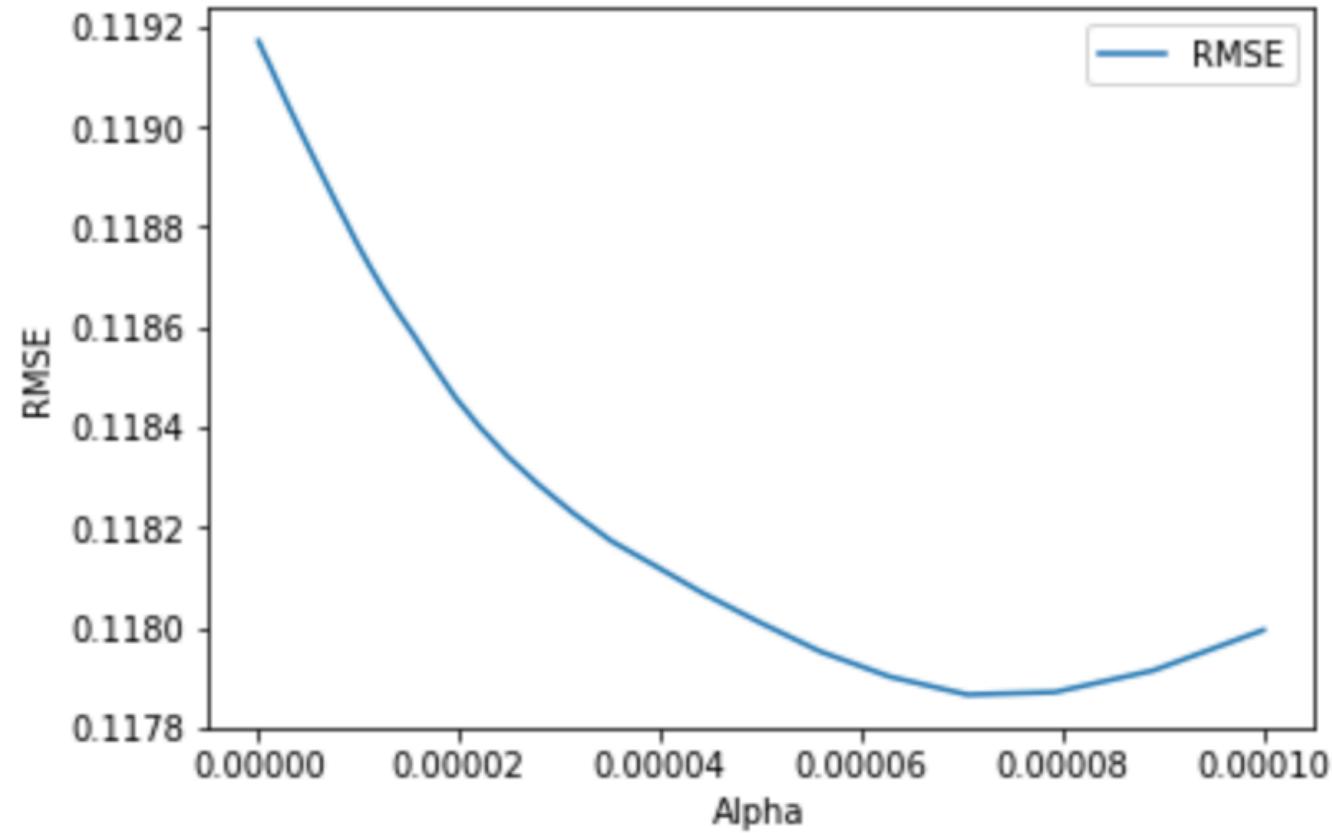
Select Data

3

Create Models

4

Assist Homebuyers



1 Clean Data

2 Select Data

3 Create Models

4 Assist Homebuyers



Model	Error (RMSE)
Lasso	0.13306

1 Clean Data

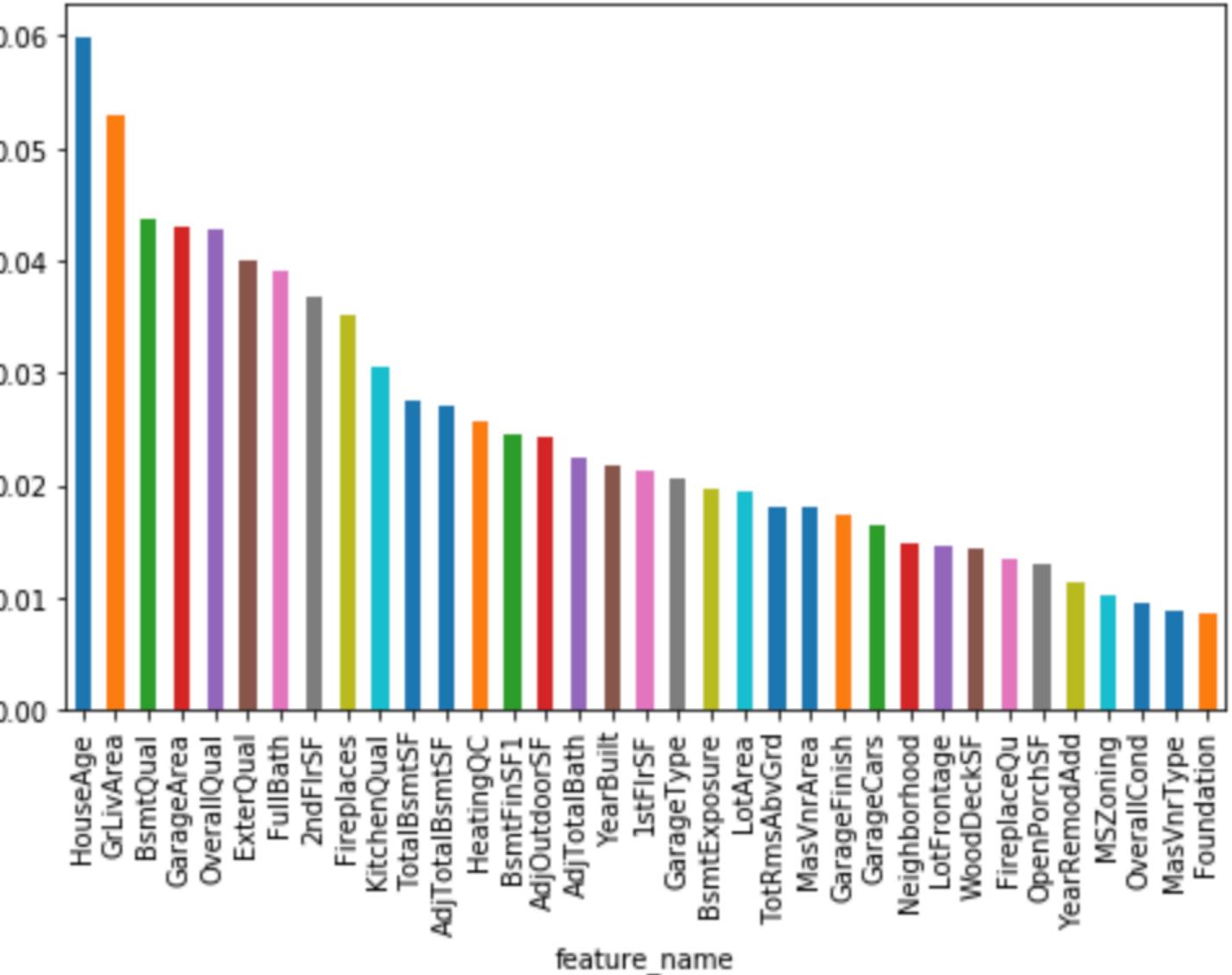
2 Select Data

3 Create Models

4 Assist Homebuyers



Feature Importance Plot of 500-Tree GBM



1 Clean Data

2 Select Data

3 Create Models

4 Assist Homebuyers



1 Clean Data

2 Select Data

3 Create Models

4 Assist Homebuyers



Model	Error (RMSE)
Lasso	0.13306
GBM	0.13501

1 Clean Data

2 Select Data

3 Create Models

4 Assist Homebuyers



1 Clean Data

2 Select Data

3 Create Models

4 Assist Homebuyers



Model	Error (RMSE)
Lasso	0.13306
GBM	0.13501
Bayesian Ridge	0.13101

1

Clean Data

2

Select Data

3

Create Models

4

Assist Homebuyers



And then we combined models.

1 Clean Data

2 Select Data

3 Create Models

4 Assist Homebuyers



Model	Error (RMSE)
Lasso	0.13306
GBM	0.13501
Bayesian Ridge	0.13101
Stacked 25% Lasso 40% GBM 35% Bayesian Ridge	0.12412

1

Clean Data

2

Select Data

3

Create Models

4

Assist Homebuyers



Amazing. How will this help homebuyers?

1

Clean Data

2

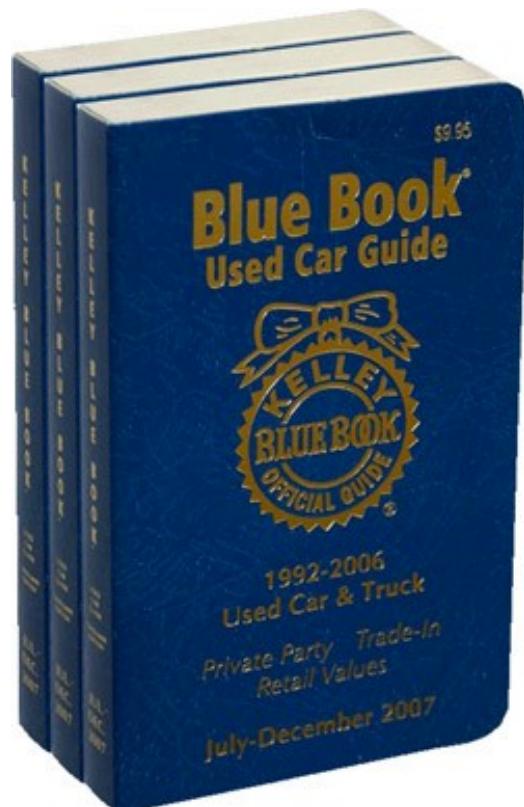
Select Data

3

Create Models

4

Assist Homebuyers



1

Clean Data

2

Select Data

3

Create Models

4

Assist Homebuyers



**Recommendation  
Engines**

**Market  
Forecasting  
Tools**



“Homebuying Powered by Data Science”

Alice Zhang

[in/xuyuan-zhang](https://www.linkedin.com/in/xuyuan-zhang)

[github.com/AliceXuyuan](https://github.com/AliceXuyuan)

Joe Lu

[in/joezlu](https://www.linkedin.com/in/joezlu)

[github.com/jzl4](https://github.com/jzl4)

Jordan Runge

[in/jordanrunge](https://www.linkedin.com/in/jordanrunge)

[github.com/jordanrunge](https://github.com/jordanrunge)

Yunmei Zhang

[in/yunmeizhang](https://www.linkedin.com/in/yunmeizhang)

[github.com/zhangym1256](https://github.com/zhangym1256)