Zillow

Data Analytics Team

Predictive Modeling
Presentation: **Aimes, Iowa**

Why Are We Here?

To Predict Home Sale Prices!

Stronger Predictions = Stronger User Confidence = Increased User Traffic = **Increased Ad Revenue!**

What are we working with?

Kaggle Datasets!

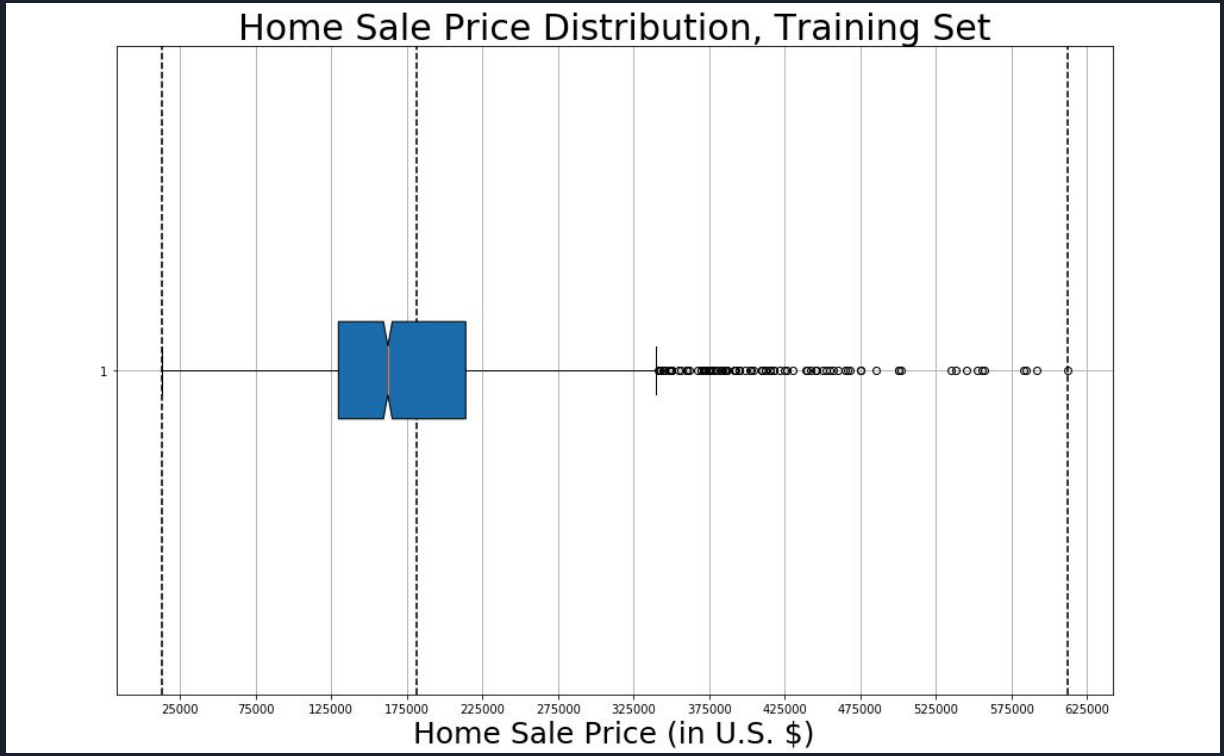
- Training Set (+2000 listings)
- Test Set (~880 listings)
- 80 Different Features included!

Some Numbers:

- **Average Home Sale Price:** just over **$180,000**
- **Ranged from: ~$13,000** to **~$610,000**
- **IQR: ~$130,000** to **~$215,000**

Sale Price Distribution, Visualized:

Sale Price Distribution, Visualized: ctd..



Model Building Process:

Step 1: **Visualize** the **Data (EDA)**

Step 2: **Clean & Format** the **Data**

Step 3: **Calculate** Our **Baseline Metric** (RMSE)

Step 4: **Build Our Model!**

Step 5: **Change Features, Parameters,** & **try again!**

Quick Correlation Recognition:

1. **Overall Quality** (ranked 1-10)
2. **Above Ground Living Area**
3. **Size of Garage** (in Sq. Ft.)
4. **Size of Garage** (in Car Capacity)
5. **Total Basement Area** (in Sq. Ft.)

Results:

Baseline Prediction Score:

RMSE of ~$84,000

Strongest Model Prediction Score:

RMSE of ~$25,500

Moving Forward..

We can **improve** our model!

Obstacle: Kaggle Submission Limit/Time

With more **time**, we could:

1. Try **different combinations** of **Features**
2. Try **different parameter settings**
3. Work with our **null values differently**

Recommendations:

We **humbly** recommend..

More Time!

And **also request**..

Datasets from other cities, states, regions, etc.