# HAPPY HOME HELPER

Dorothy Alexander and Will Cline

# SUMMARY

Our **best model** incorporates these columns:

bedrooms', 'bathrooms', 'sqft_living', 'sqft_lot', 'floors', 'grade', 'sqft_above', 'sqft_living15', 'sqft_lot15', 'age', 'months_ago_sold'

**One hot encoded** the column **zip code** in order to have all of the zip codes accounted for individually

Columns were **created** by **multiplying** other **top correlated** models:

'sqft_living&above', 'sqft_living&grade', 'sqft_living&living15', 'grade&sqft_above', 'bathrooms&sqft_living', 'sqft_above&sqft_living15', 'grade&sqft_living15'

**Standard scaler** applied on **all** the features and a **log function on price**.

**Outliers** beyond 3 STD were **cut out** of the data.

# BUSINESS PROBLEM

Predictive!!!!

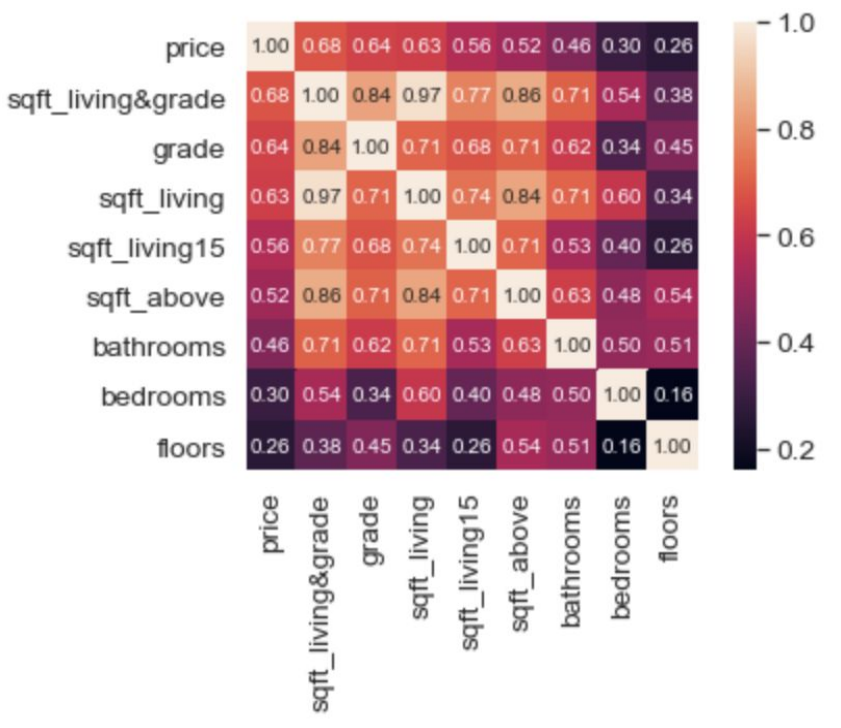
- Create a tool that helps get predictions on specific house prices
- Enter in information about their home and receive an output prediction
- Real estate companies would like to be able to use this for their clients
- We have been contacted by a wealthy investor who wants to get in the real estate market….



# THE MAN THE INVESTOR THE LEGEND


# USING HEAT MAP TO CONFIRM CORRELATION

Heat map of cross correlations



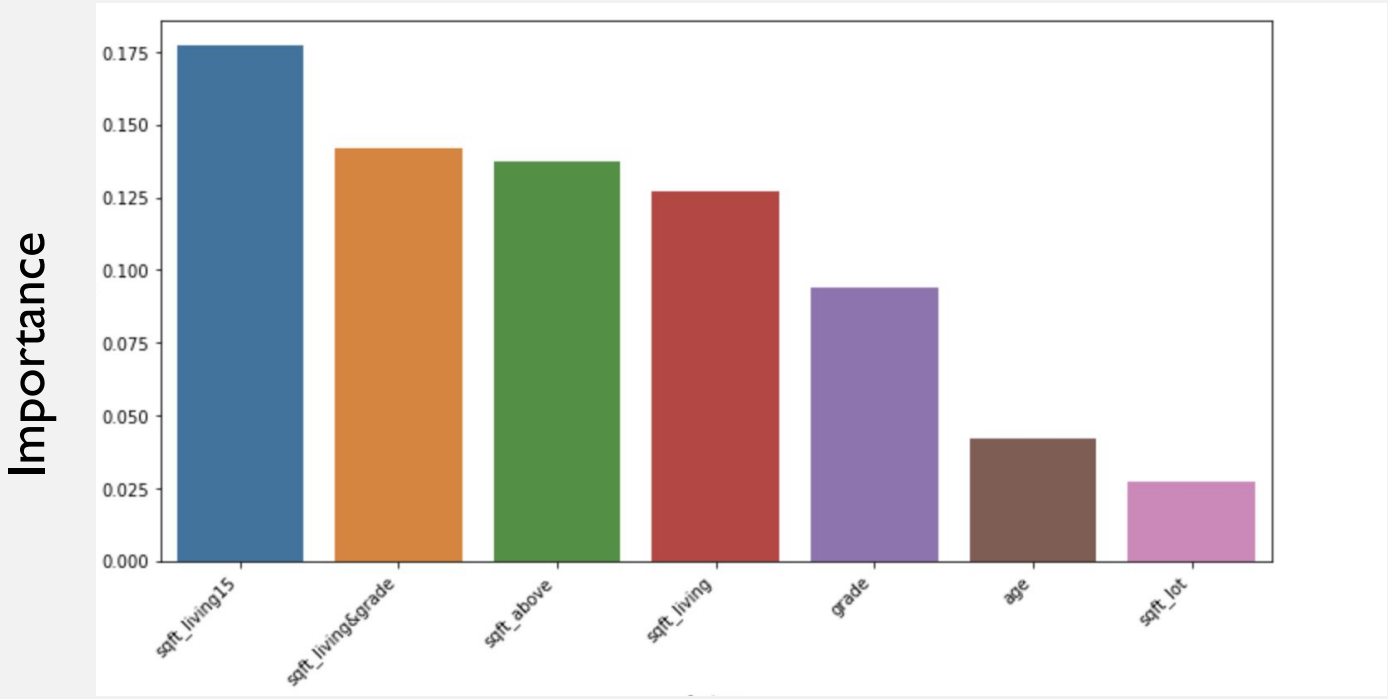

Top cross correlations

| | feature_a | feature_b | correlation |
|---|---|---|---|
| 0 | yr_built | age | 1.000000 |
| 1 | sqft_living | sqft_above | 0.876448 |
| 2 | sqft_living | grade | 0.762779 |
| 3 | sqft_living | sqft_living15 | 0.756402 |
| 4 | grade | sqft_above | 0.756073 |
| 5 | bathrooms | sqft_living | 0.755758 |
| 6 | sqft_above | sqft_living15 | 0.731767 |
| 7 | sqft_lot | sqft_lot15 | 0.718204 |
| 8 | grade | sqft_living15 | 0.713867 |
| 9 | price | sqft_living | 0.701917 |

Top correlations to price

| | price |
|---|---|
| **sqft_living&grade** | 0.680787 |
| **grade** | 0.640878 |
| **sqft_living&living15** | 0.634868 |
| **sqft_living** | 0.631505 |
| **grade&sqft_living15** | 0.628235 |
| **bathrooms&sqft_living** | 0.609536 |
| **sqft_living&above** | 0.586148 |
| **grade&sqft_above** | 0.584144 |
| **sqft_living15** | 0.560833 |
| **sqft_above&sqft_living15** | 0.560066 |

# FEATURE IMPORTANCE

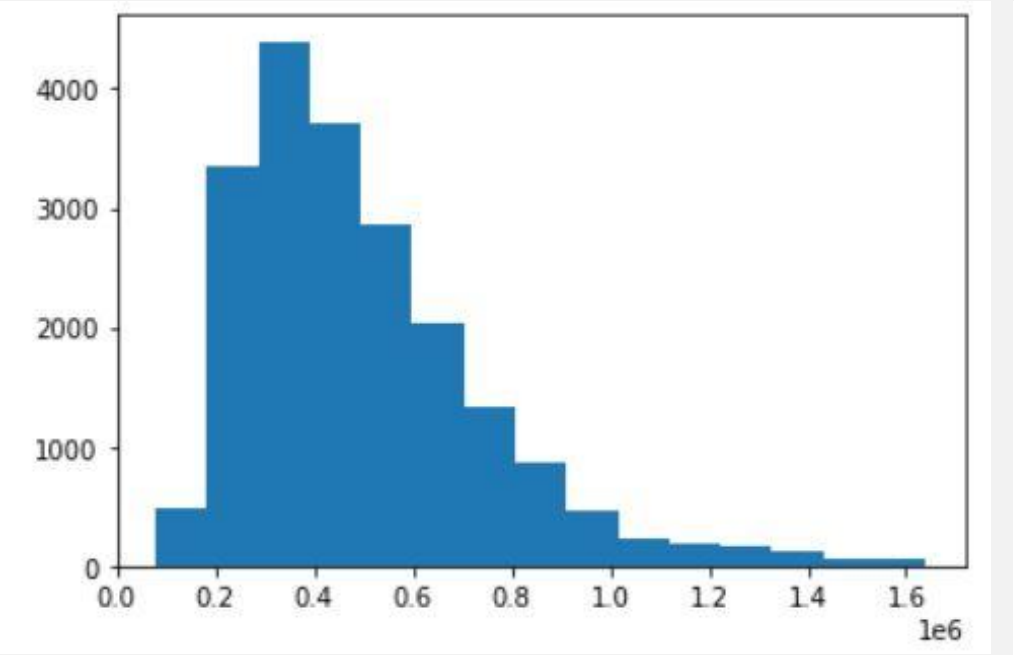


# Error(RMSE) of various models:

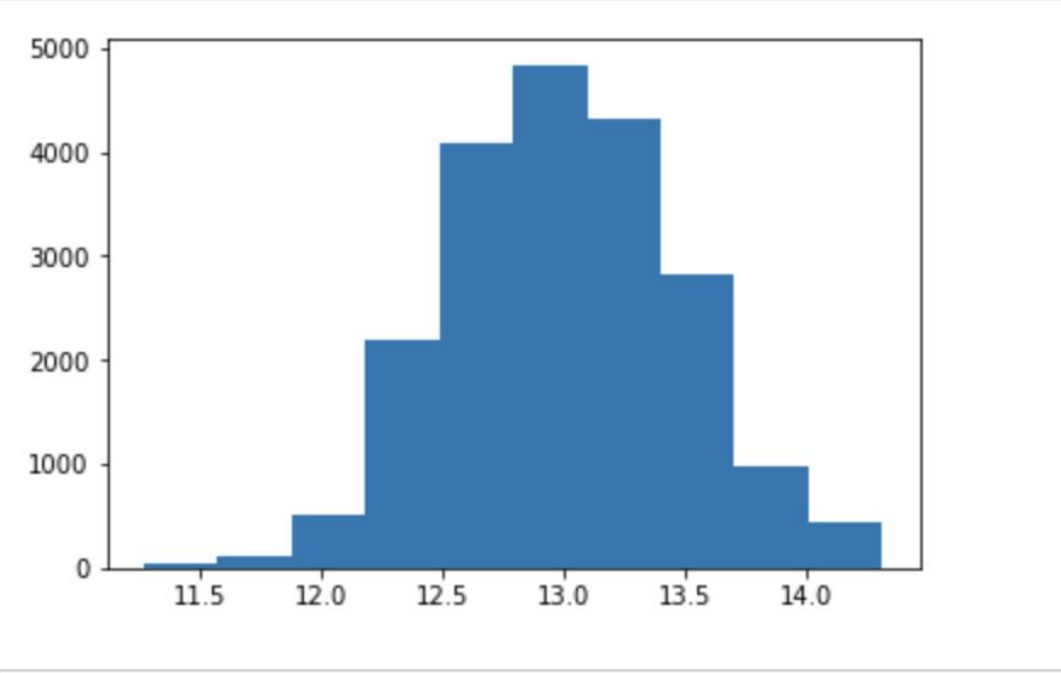
| Columns used in DF, | Training RMSE, | Test RMSE |
|---|---|---|
| Top3: | (179317.18, | 176639.52) |
| All Usable besides zip: | (156101.43, | 153769.06) |
| All Usable: | (105202.55, | 105940.13) |

# LOG FUNCTION = BEST SCALER

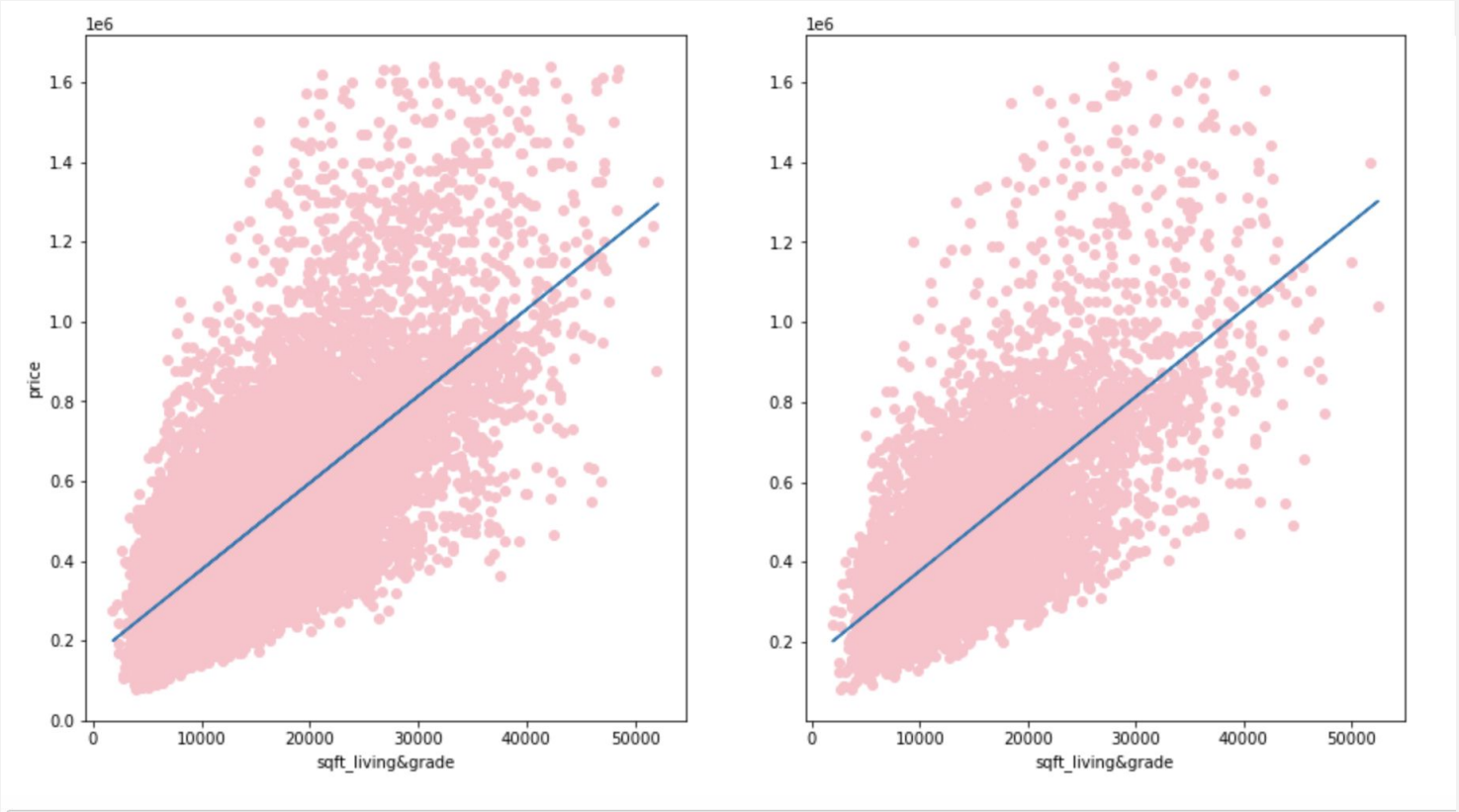
Price distribution without log function




Price distribution with log function



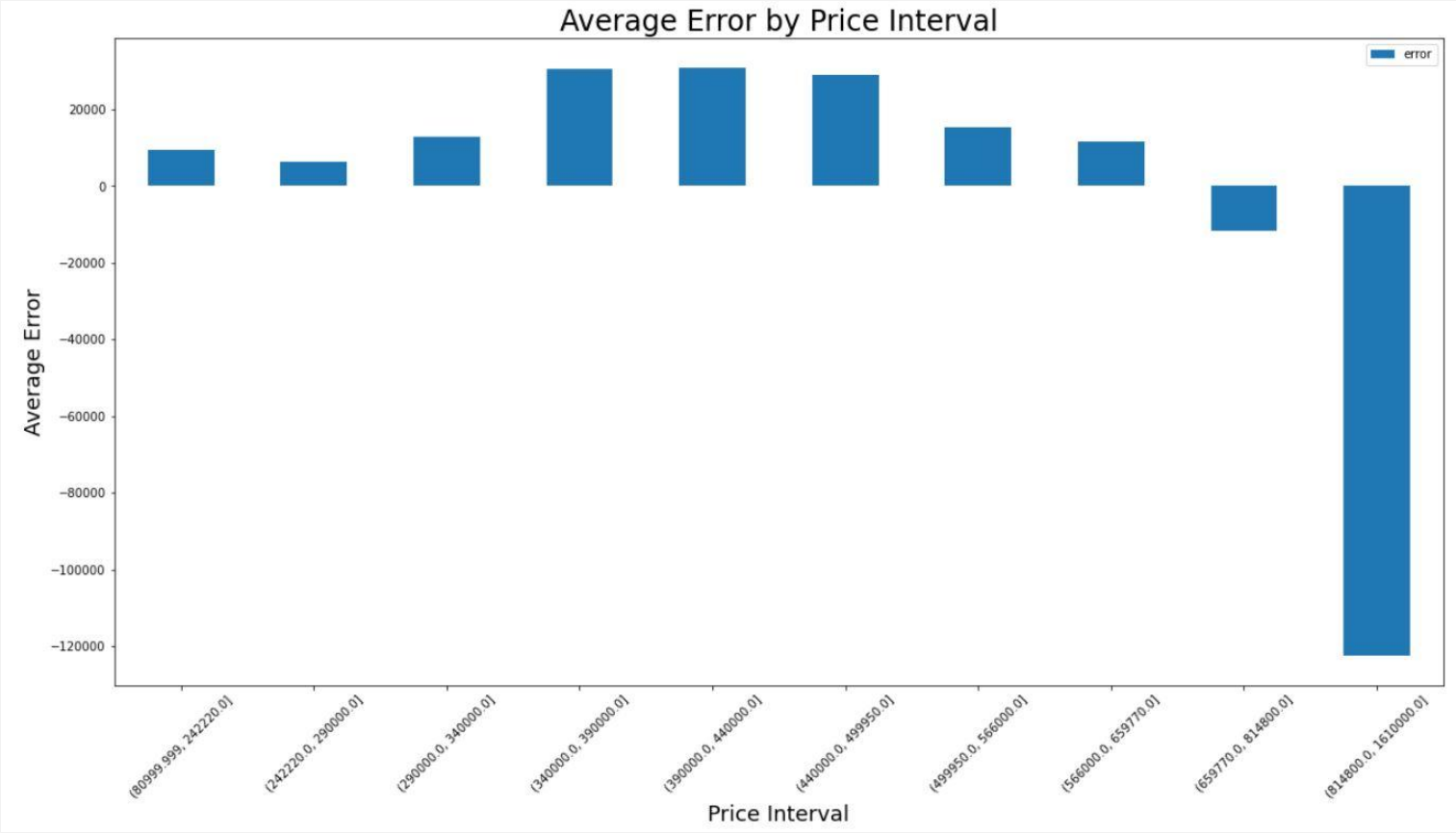
# Train, Test, Split: sqft_living&grade




TRAIN

TEST

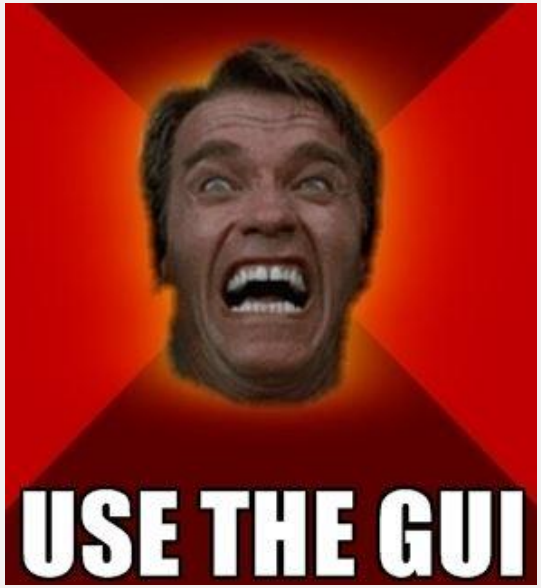
# Where was our model getting the most error?



# GUI

CHECK OUT HOW OUR MODEL PERFORMS!!!!

| | date | price | bedrooms | bathrooms | sqft_living | sqft_lot | floors | grade | sqft_above | yr_built | yr_renovated |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 10/13/2014 | 221900.0 | 3 | 1.0 | 1180 | 5650 | 1.0 | 7 | 1180 | 1955 | 0.0 |



# Data Key

- id - Unique ID for each home sold
- date - Date of the home sale
- price - Price of each home sold
- bedrooms - Number of bedrooms
- bathrooms - Number of bathrooms, where .5 accounts for a room with a toilet but no shower
- sqft_living - Square footage of the apartments interior living space
- sqft_lot - Square footage of the land space
- floors - Number of floors
- waterfront - A dummy variable for whether the apartment was overlooking the waterfront or not
- view - An index from 0 to 4 of how good the view of the property was
- condition - An index from 1 to 5 on the condition of the apartment,
- grade - An index from 1 to 13, where 1-3 falls short of building construction and design, 7 has an average level of construction and design, and 11-13 have a high quality level of construction and design.
- sqft_above - The square footage of the interior housing space that is above ground level
- sqft_basement - The square footage of the interior housing space that is below ground level
- yr_built - The year the house was initially built
- yr_renovated - The year of the house's last renovation
- zipcode - What zipcode area the house is in
- lat - Lattitude
- long - Longitude
- sqft_living15 - The square footage of interior housing living space for the nearest 15 neighbors
- sqft_lot15 - The square footage of the land lots of the nearest 15 neighbors

https://www.kaggle.com/danacassidy/keras-house-sales-prediction

# THANK YOU!

QUESTIONS?