



DC housing market price

Final Project – 4/6/2019

Prediction (Y)

Redfin Estimate for 1248 Haskell St

\$928,957

+\$41K over list price of \$888K

Redfin Estimate based on recent home sales. [View estimate history.](#)



OPEN **1248 Haskell St**
Berkeley, CA 94702
Status: Active

\$888,000 **Price** **2** **Beds** **1** **Bath** **1,068 Sq. Ft.**
\$831 / Sq. Ft.
Redfin Estimate: \$928,957 On Redfin: 13 hours

Overview Property Details Property History Schools Tour Insights Public Facts Redfin Estimate Neighborhood Similar Homes

Go Tour This Home

THURSDAY 4 APR FRIDAY 5 APR SATURDAY 6 APR

Schedule Tour

It's free, with no obligation — cancel anytime

Ask a Question (510) 343-7365

Redfin Refund: \$3,277

Savings when you buy with a Redfin Agent

Start an Offer Live Chat

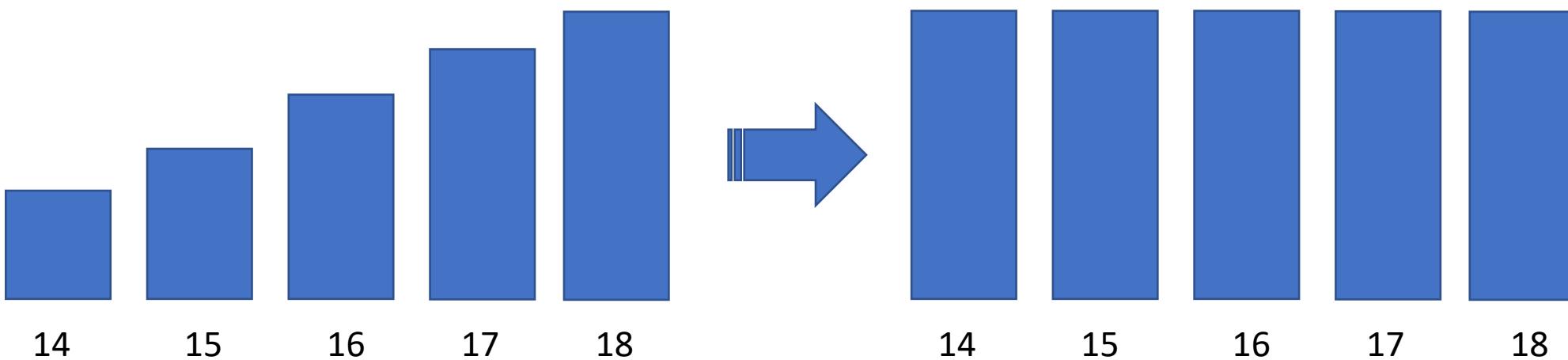
D.C. Residential Properties

Residential Properties in Washington D.C.

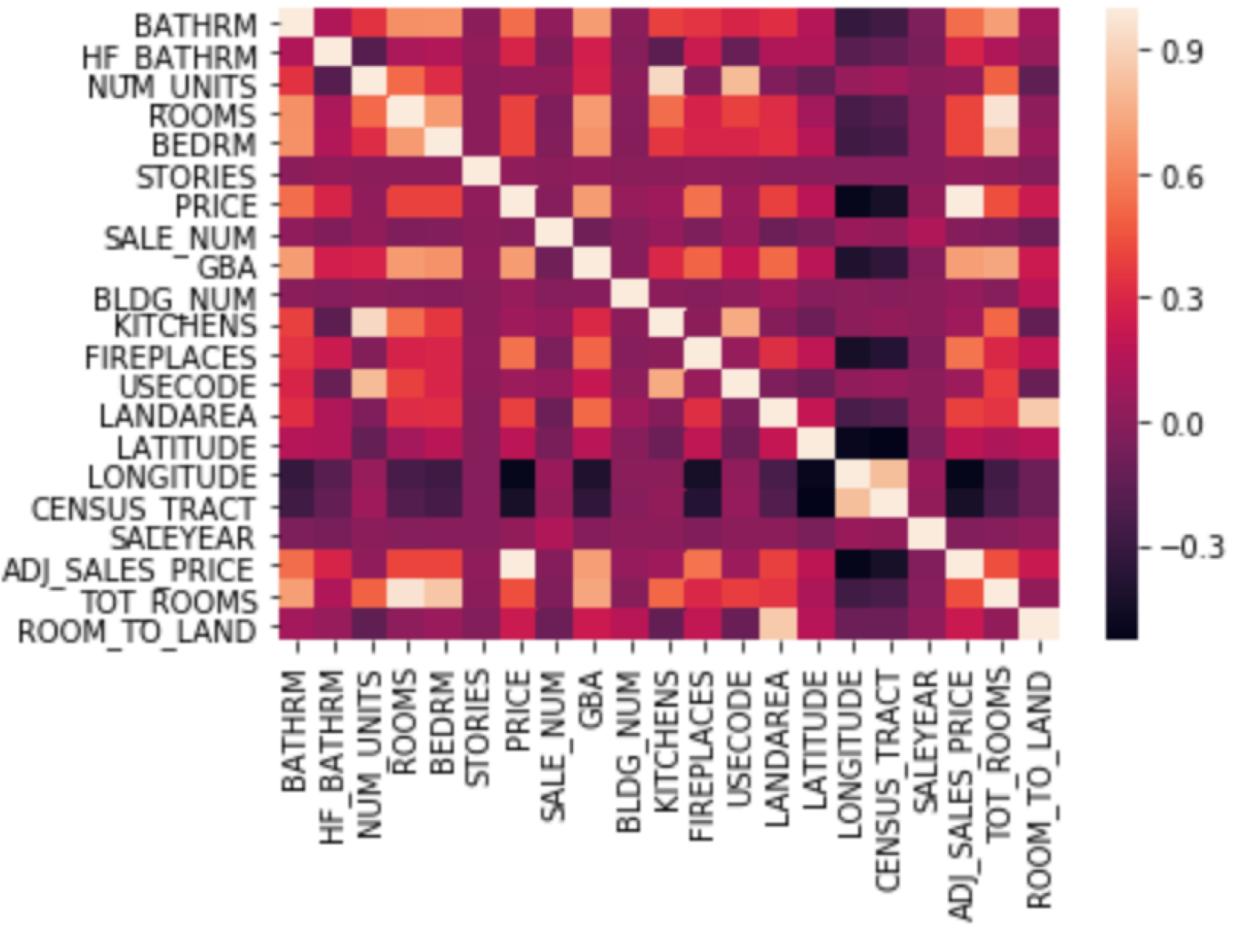
Kaggle.com

- 41 columns
 - Numerical: rooms, bathrooms, longitude, ...
 - Categorical: post code*, type of heating, type of room, ...
- 160K rows across from 1982 to 2018

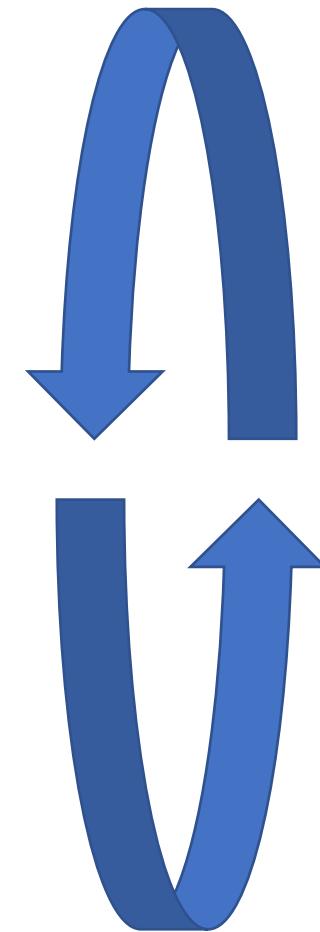
5.9% increase year over year



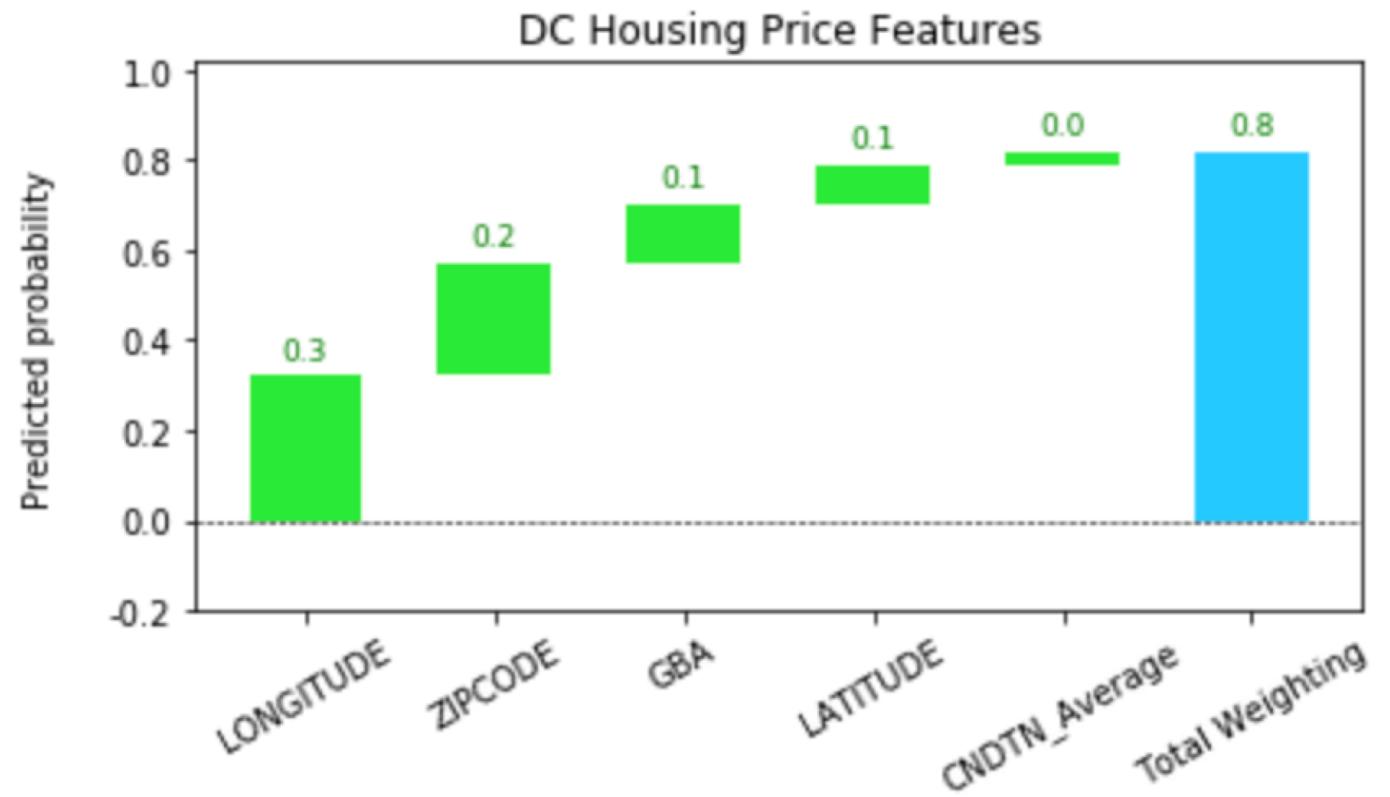
A heat map is a graphical representation of data where the individual values contained in a matrix are represented as colors. "Heat map" is a newer term but shading matrices have existed for over a century.



- Dummies
 - E.g. Post Code
- Feature engineering
 - E.g. ROOMS + BATHROOMS / square feet
- rf_model.feature_importances_
 - E.g. removing anything with less than 0.09 (less than 9%)
- Removing categorical data with too many categories
 - .unique() .nunique()
 - e.g. CENSUS_BLOCK with 3K categories
- Removing features with similar information

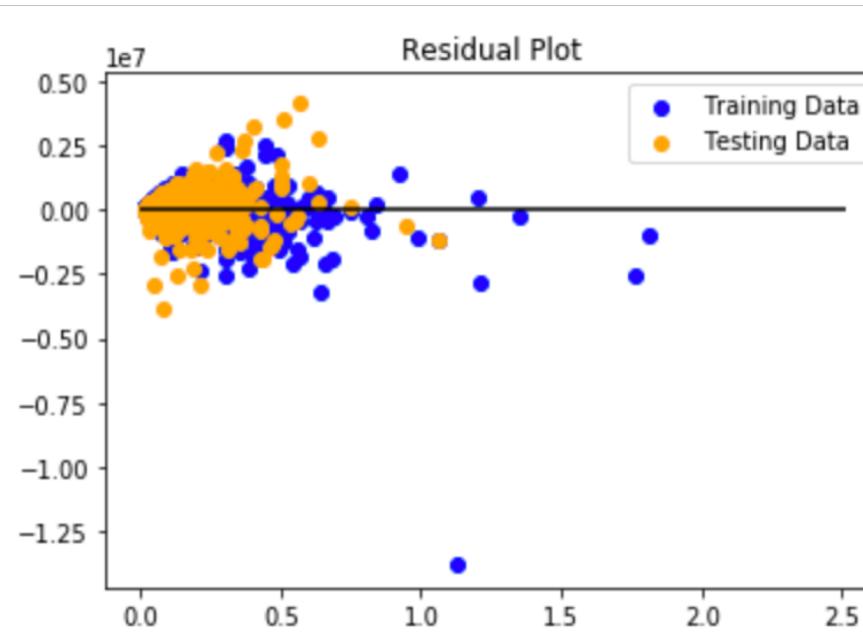


Gross floor **area** (GFA) in real estate is the total **floorarea** inside the **building** envelope, including the external walls, and excluding the roof. Definitions of GFA, including which **areas** are to be counted towards it and which **areas** aren't, vary around the world.

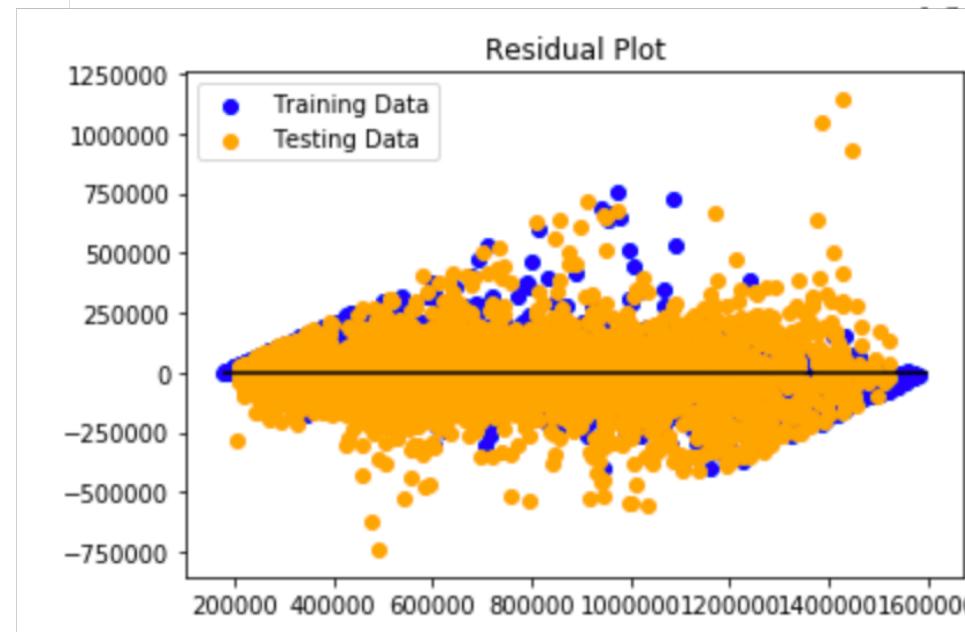


In applied statistics, a partial residual plot is a graphical technique that attempts to show the relationship between a given independent variable and the response variable given that other independent variables are also in the model.

Before Filtering Prices



After Filtering Prices



Tuning the hyper-parameters of an estimator

```
rf_model = RandomForestRegressor(max_depth= 10)
```

5%	10%	15%	20%	25%	30%
77%	97%	92%	95%	93%	93%
67%	85%	81	83%	84%	80%
10	5	11	12	9	13

Interesting findings

2014	2015	2016	2017	2018
68%	76%	75%	85%	57%

Conclusion

The r-squared for testing is: 85%
Predicted Price Mean = \$ 711,225.32
Mean Squared Error = \$ 161,599.93