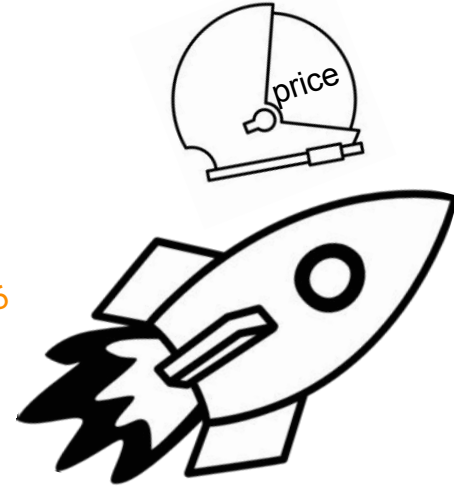

Flatiron School - Data Science: Module 1 Project

A guide by Gabe Vazquez

Understanding our data (21)

id date bedrooms sqft_living_15 yr_built sqft_lot_15
sqft_living bathrooms lat zipcode yr_renovated long
view waterfront floors grade
conditio sqft_basement sqft_abov
e sqft_basement





Hypothesis

→ **Bedrooms**

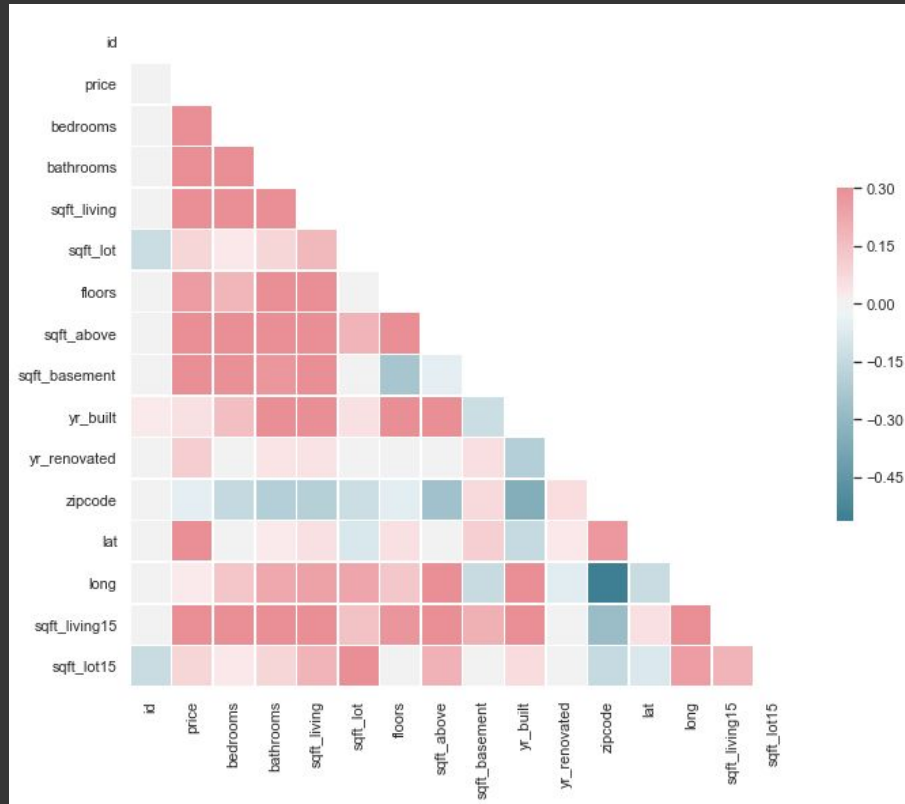
More bedrooms will equal a higher price

→ **Bathrooms**

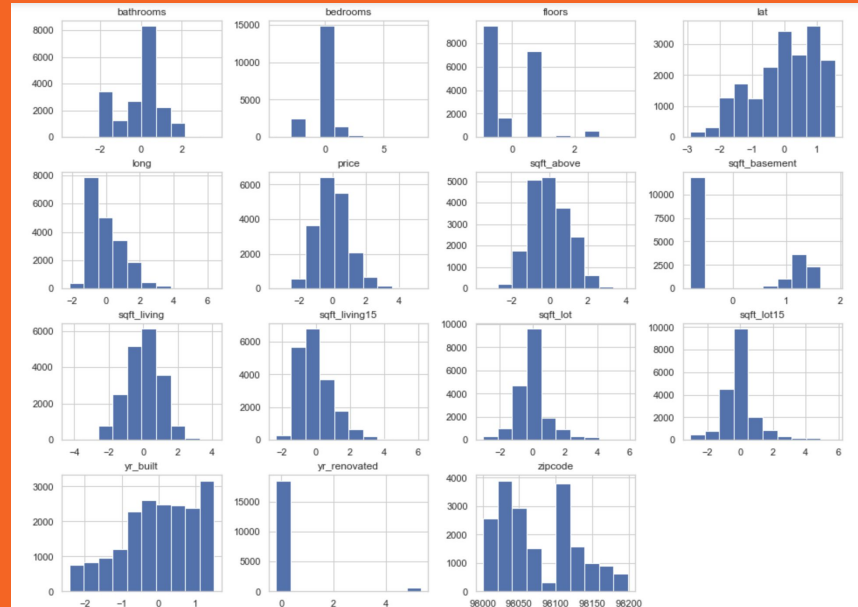
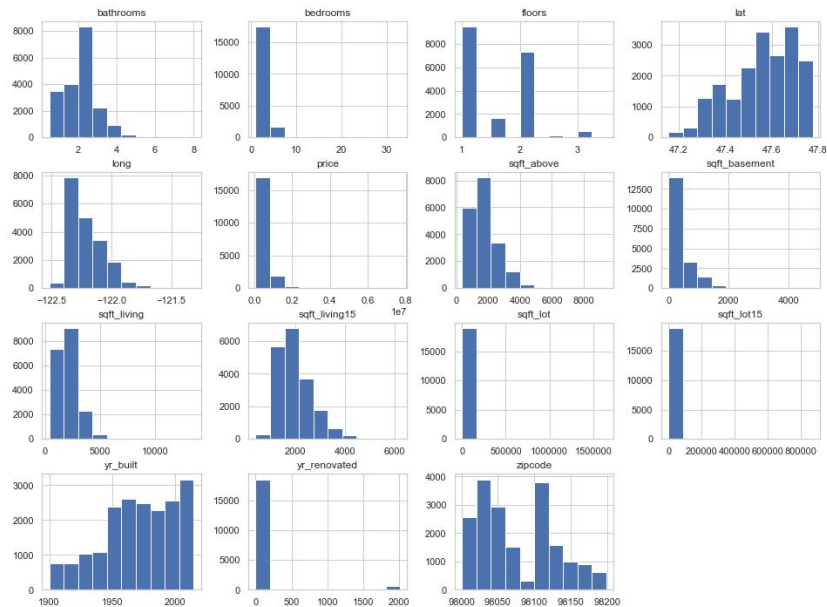
More bathrooms will increase the price

→ **Sqft**

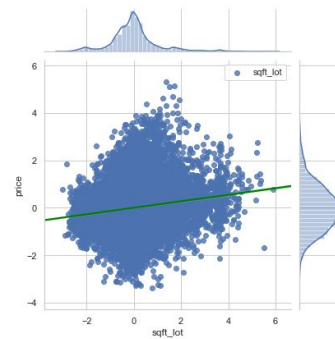
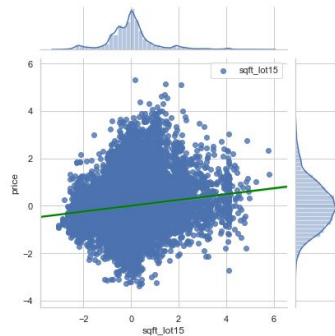
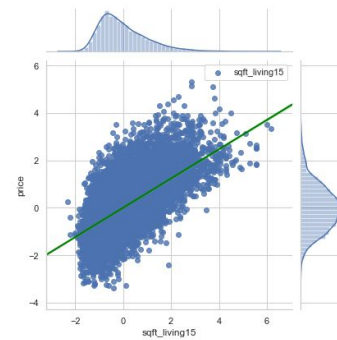
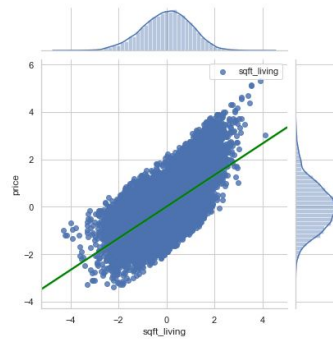
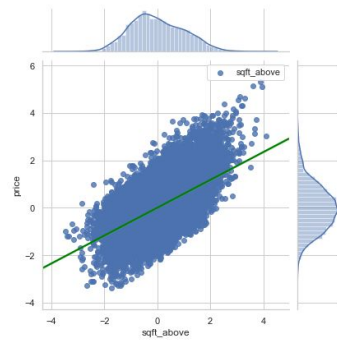
A larger home will cost more



Apples to Apples



—



	0	1	2	3	4
0	ind_var	r_squared	intercept	slope	p-value
1	bedrooms	0.118341	1.5439e-16	0.344007	0
2	bathrooms	0.270736	1.5439e-16	0.520323	0
3	sqft_living	0.445292	5.6101e-15	0.667302	0
4	sqft_lot	0.0189547	1.5439e-16	0.137676	9.85637e-82
5	floors	0.0984952	1.5439e-16	0.313839	0
6	sqft_above	0.346138	1.5439e-16	0.588335	0
7	sqft_basement	0.0520208	1.5439e-16	0.228081	1.29308e-224
8	yr_built	0.00671185	1.5439e-16	0.0819259	6.63749e-30
9	yr_renovated	0.0102159	1.5439e-16	0.101074	1.06478e-44

```
Index(['grade_10', 'grade_11', 'grade_12', 'grade_13', 'grade_3', 'grade_4',  
      'grade_5', 'grade_6', 'grade_7', 'grade_8'],  
      dtype='object')
```

OLS Regression Results

```
=====
```

Dep. Variable:	price	R-squared:	0.499
Model:	OLS	Adj. R-squared:	0.499
Method:	Least Squares		
Date:	Thu, 24 Jan 2019	Prob (F-statistic):	0.00
Time:	08:06:09	Log-Likelihood:	-20573.
No. Observations:	19164	AIC:	4.117e+04
Df Residuals:	19153	BIC:	4.125e+04
Df Model:	10		
Covariance Type:	nonrobust		

```
=====
```


Cross - Validation

```
array([-0.23649182, -0.26042306, -0.24438096, -0.25134974, -0.23174412,  
       -0.24540708, -0.24754896, -0.25010873, -0.25031747, -0.20286663])
```



Best Predictors

- Grade
- Sqft
- Bedrooms
- Bathrooms
- Sqft