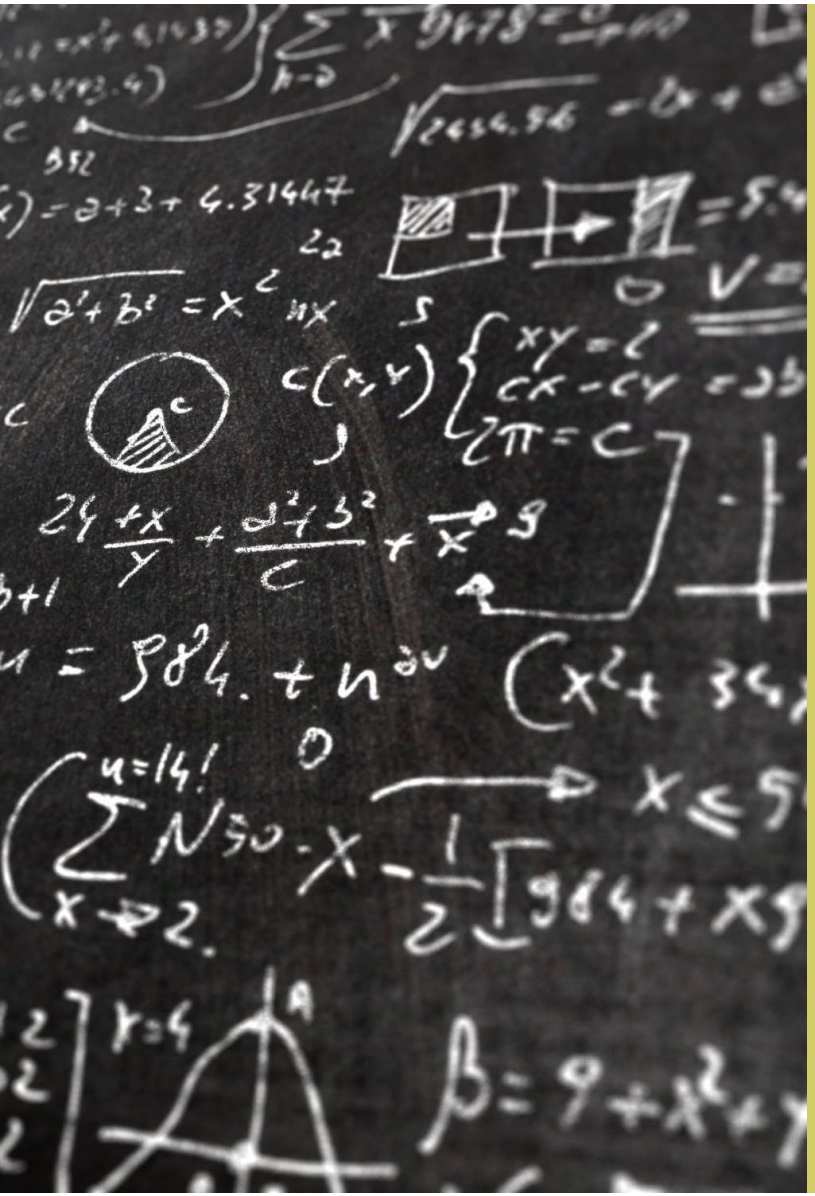


THE REAL STATE COMPANY

Linear Regression in King County, WA
House Sales Dataset

Rio Kinslow

PROJECT PROMPT

- Use multiple linear regression modeling to **analyze house sales** in King County, Washington.

PROBLEM STATEMENT:

HOW CAN WE MAXIMIZE PROFIT?

SEARCH OF ACTIONABLE INSIGHTS

UNDERSTANDING OF HOUSE MARKET

BUSINESS VALUE:

Price estimation

Feature selection and impact

Latest house price trends

METHODOLOGY

BUSINESS UNDERSTANDING

DATA UNDERSTANDING

DATA PREPARATION

MODELLING

RESULTS

METHODOLOGY:

STEP-BY-STEP

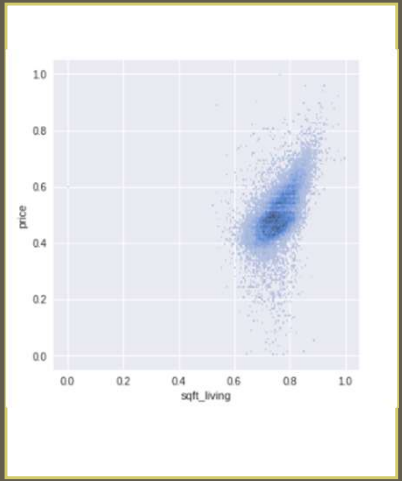
Business Understanding in obtaining and reading it using the Pandas package

Data understanding using the Pandas and NumPy packages

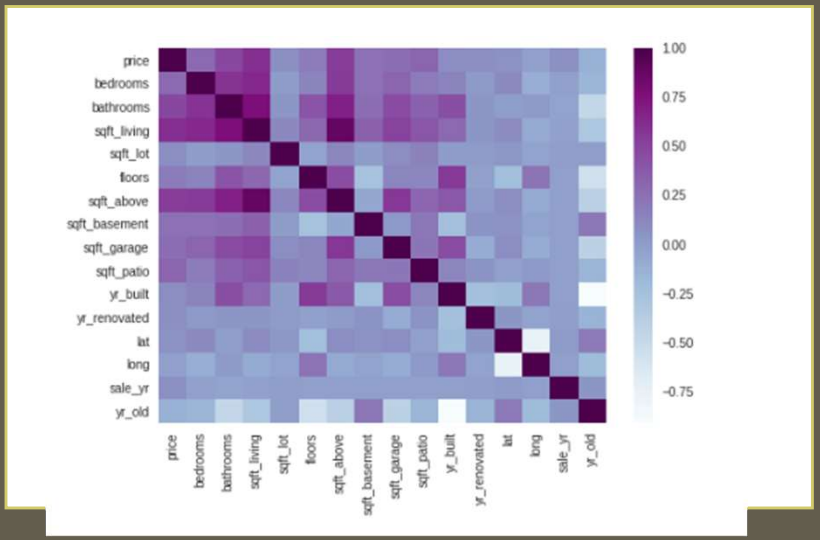
Data Exploration using Matplotlib and Seaborn

Model data using linear regression

Interpret results



price	1.000
bedrooms	0.289
bathrooms	0.480
sqft_living	0.609
sqft_lot	0.087
floors	0.181
sqft_above	0.539
sqft_basement	0.245
sqft_garage	0.264
sqft_patio	0.314
yr_built	0.096
yr_renovated	0.085
lat	0.063
long	-0.022
sale_yr	0.074
yr_old	-0.127



RECOMMENDATION #1

LIVING SPACE (SQFT)

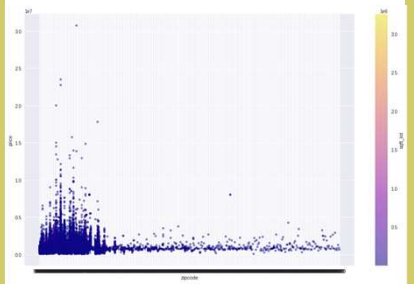
- Living space square footage is the single most important feature
- Larger homes mandate a higher asking price
- Attempt to sell homes with a great deal of square footage

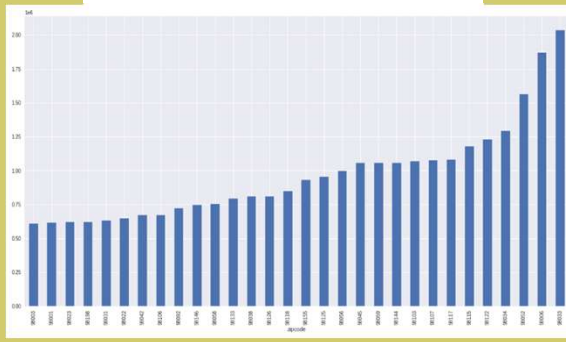
RECOMMENDATION #2:

ZIPCODE (LOCATION)

- Location of homes is very important
- House prices seem to be clustered in terms of zipcode
- Certain zipcodes can encompass many factors and variables

sqft_living	229.0350
zipcode_98003	3579.3787
zipcode_98006	8.907e+05
zipcode_98022	4072.6313
zipcode_98023	-1.106e+04
zipcode_98031	4.293e+04
zipcode_98033	1.142e+06
zipcode_98034	6.081e+05
zipcode_98038	1.015e+05
zipcode_98042	4471.8438

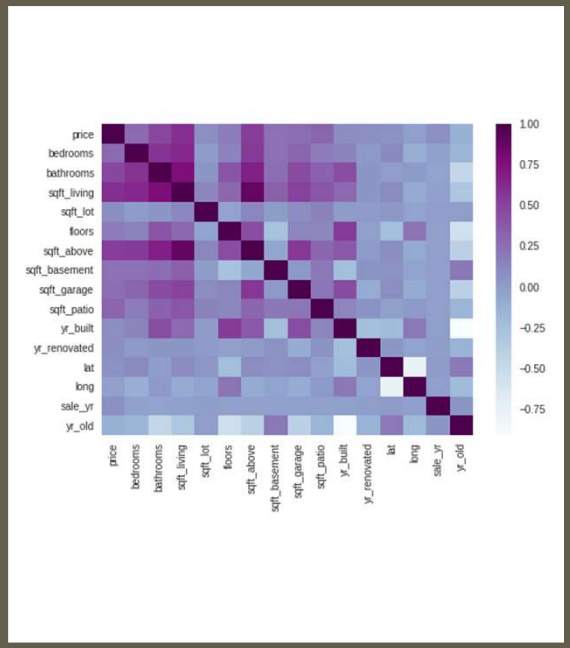


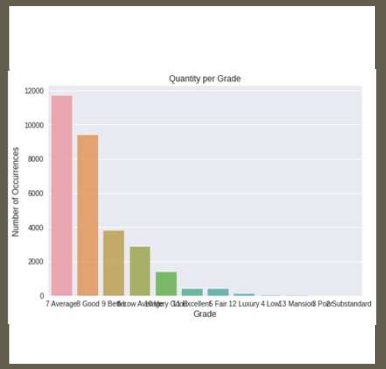


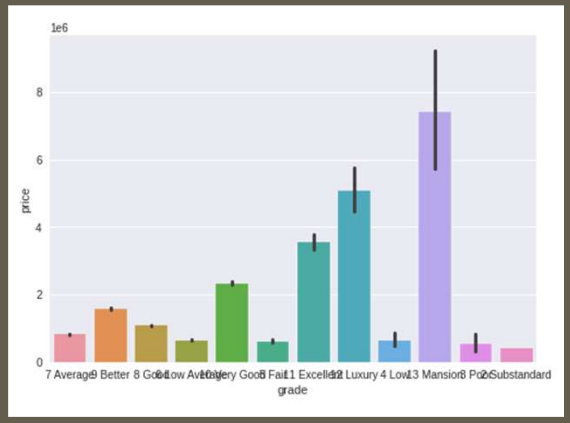
RECOMMENDATION #3:

GRADE

- King County, WA grading system very influential
- The higher the grade, the higher price that the house commands
- The grading distribution appears to follow a normal curve







	price
price	1.000
bedrooms	0.289
bathrooms	0.480
sqft_living	0.609
sqft_lot	0.087
floors	0.181
sqft_above	0.539
sqft_basement	0.245
sqft_garage	0.264
sqft_patio	0.314
yr_built	0.096
yr_renovated	0.085
lat	0.063
long	-0.022
sale_yr	0.074
yr_old	-0.127

OLS Regression Results

Dep. Variable:	price	R-squared:	0.531
Model:	OLS	Adj. R-squared:	0.530
Method:	Least Squares	F-statistic:	483.9
Date:	Fri, 30 Sep 2022	Prob (F-statistic):	0.00
Time:	20:30:53	Log-Likelihood:	-2.5349e+05
No. Observations:	17570	AIC:	5.071e+05
Df Residuals:	17528	BIC:	5.074e+05
Df Model:	41		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-2.675e+08	1.42e+07	-18.821	0.000	-2.95e+08	-2.4e+08

OLS Regression Results

Dep. Variable:	price	R-squared:	0.689
Model:	OLS	Adj. R-squared:	0.688
Method:	Least Squares	F-statistic:	942.3
Date:	Fri, 30 Sep 2022	Prob (F-statistic):	0.00
Time:	21:08:30	Log-Likelihood:	-2.4500e+05
No. Observations:	17498	AIC:	4.901e+05
Df Residuals:	17456	BIC:	4.904e+05
Df Model:	41		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-2.651e+08	9.31e+06	-28.479	0.000	-2.83e+08	-2.47e+08

FINAL MULTI-VARIATE LINEAR REGRESSION MODEL:

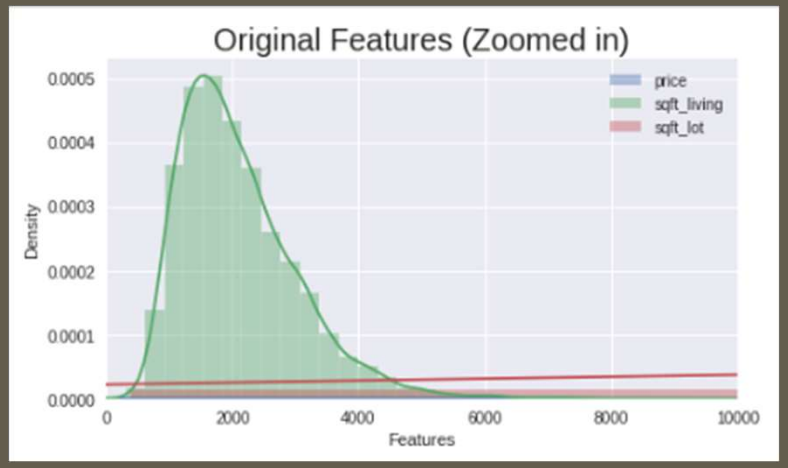
- (top)

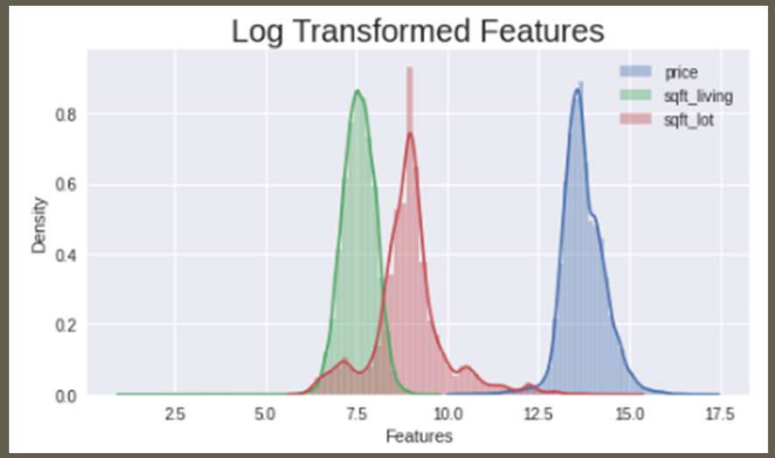
multi-variate linear regression model accounting for sqft_living, zipcode, and grade w/o significant features

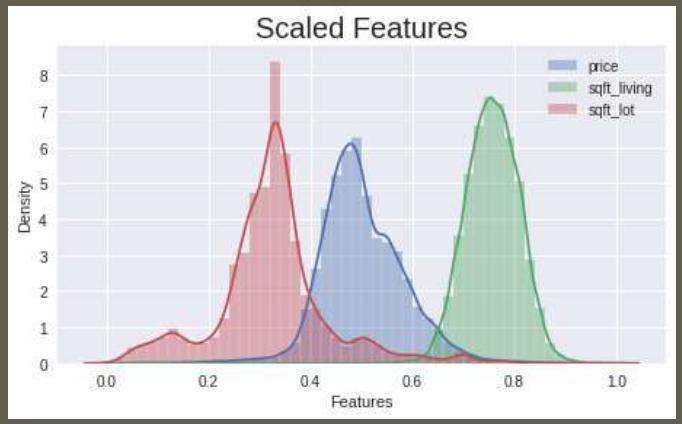
- (bottom)

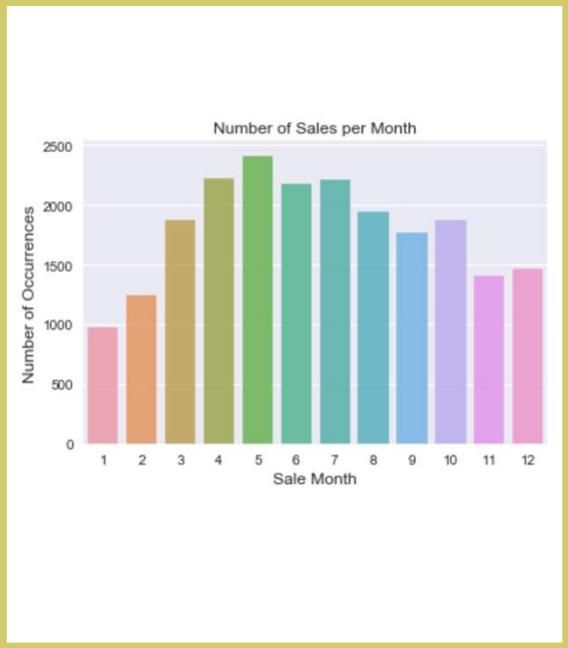
multi-variate linear regression model accounting for sqft_living, zipcode, and grade w/ significant features

COMPARISON FOR THE FINAL MODEL









INTERESTING INSIGHTS/ADDITIONAL RECOMMENDATIONS:

- There are increased sales during the late spring/early summer months
- If looking for an additional feature in a home, prioritize the number of bathrooms
- Upward trend in bathrooms vs. price

FUTURE WORK AND RESEARCH:

- Improve linear regression model
- Import and concatenate other datasets
- Simulate different scenarios
- Search for any confounding variables

THANK YOU
“ I AM OPEN FOR QUESTIONS, FEEDBACK & COMMENTS”

RIO KINSLOW CONTACT INFORMATION:

- Contact Number: 765-894-2149
- Email Address: rakinslow1980@gmail.com

LinkedIn: www.linkedin.com/in/rio-ann-kinslow