



We will go through data exploration to identify most important features and empirically study how the various factors influence the house prices.



This dataset contains house sale prices for King County, which includes Seattle, and includes 21,613 rows and 21 columns such as price, grade, year built among other things.

PROCESS DATA

D1

Explore Dataset

O2

Pre-processing Data

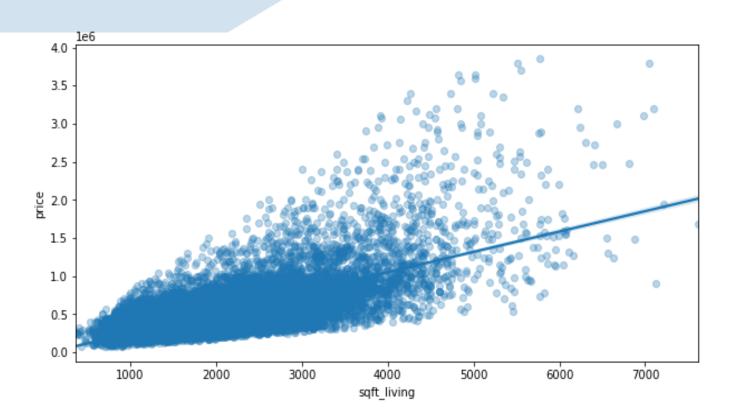
O3

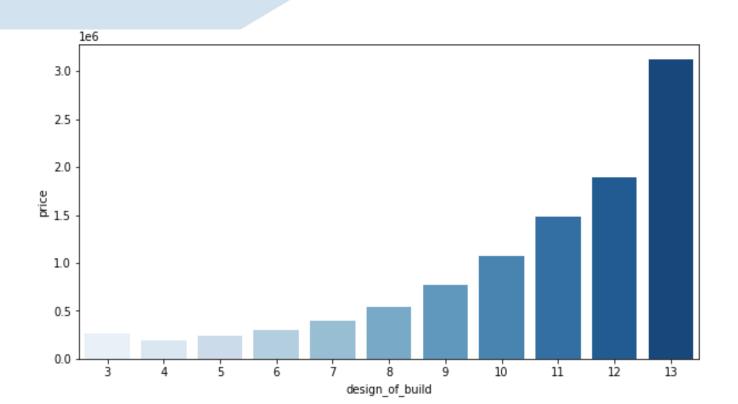
Analysis and Visualization

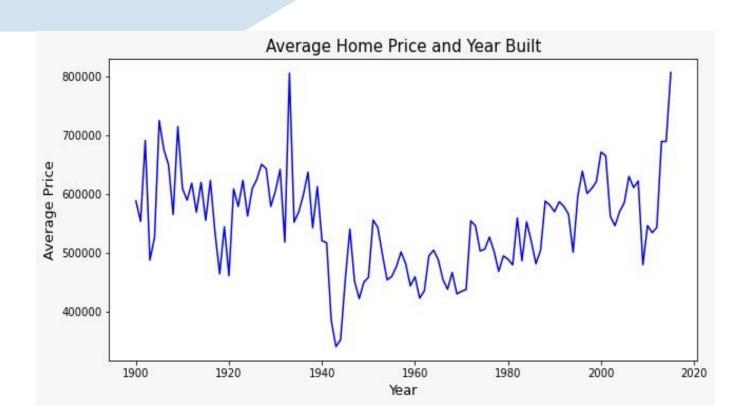
Modeling

04

index	1	0.032	0.012	0.11	0.046	-0.009	0.082	0.075	-0.045	0.0044	0.027	1.00
price	0.032	1	0.3	0.52	0.7	0.4	0.68	0.61	0.32	0.31	0.6	- 0.75
bedrooms	0.012	0.3	1	0.51	0.57	0.079	0.36	0.47	0.3	-0.004	0.39	- 0.50
bathrooms	0.11	0.52	0.51	1	0.76	0.19	0.67	0.68	0.29	0.029	0.57	
sqft_living	0.046	0.7	0.57	0.76	1	0.28	0.77	0.88	0.43	0.058	0.76	- 0.25
view	-0.009	0.4	0.079	0.19	0.28	1	0.25	0.16	0.28	0.0065	0.28	- 0.00
grade	0.082	0.68	0.36	0.67	0.77	0.25	1	0.76	0.17	0.12	0.72	0.25
sqft_above	0.075	0.61	0.47	0.68	0.88	0.16	0.76	1	-0.056	0.0063	0.74	0.20
sqft_basement	-0.045	0.32	0.3	0.29	0.43	0.28	0.17	-0.056	1	0.11	0.2	0.50
lat	0.0044	0.31	-0.004	0.029	0.058	0.0065	0.12	0.0063	0.11	1	0.054	0.75
sqft_living15	0.027	0.6	0.39	0.57	0.76	0.28	0.72	0.74	0.2	0.054	1	1.00
	index	price	bedrooms	bathrooms	sqft_living	view	grade	sqft_above	sqft_basement	lat	sqft_living15	1.00











LinearRegression

R2= 0.6976360812207014 Rmse = 189121.2697053251



Lasso

R2= 0.6976361846682814 Rmse = 189121.2986867042



Ridge

R2= 0.6976361039405464 Rmse = 189121.26976381202



04

Polynomial

R2=0.7878648040299494 Rmse = 154307.90303156493



GradientBoosting Regressor Model

R2= 0.8530941308569066 Rmse = 74986.42924142526

Test Modeling

GradientBoostingRegressor Model

R2= 0.8618223385055157 Rmse= 128155.87399722042



THANKS...