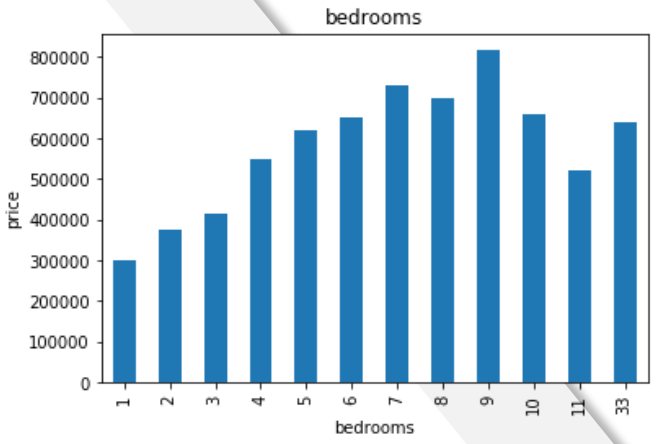
King County Housing PRICE PREDICTION

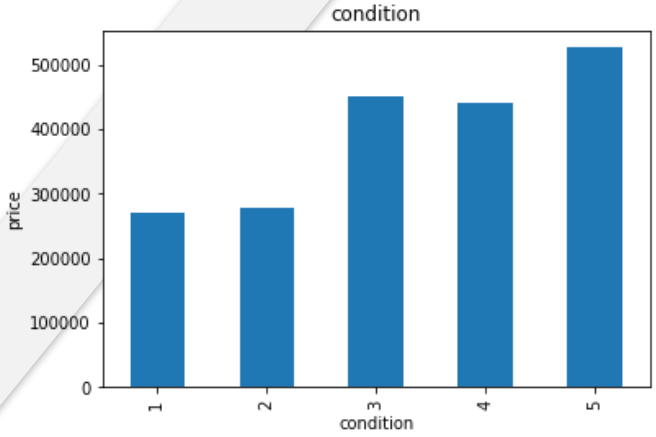
A Linear Regression Project

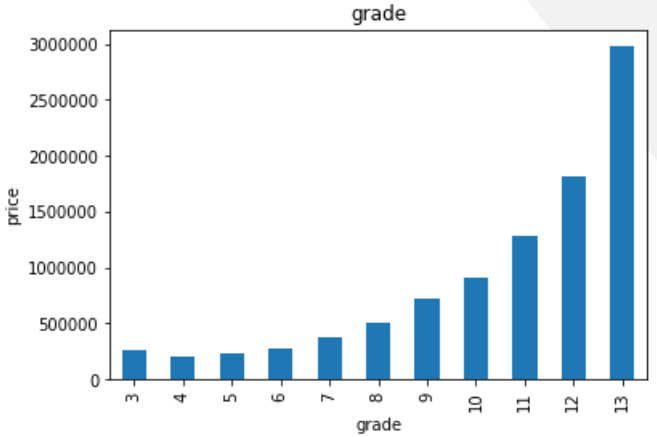
- KISHOR SHANKARANARAYAN

Motivation & Objective

Assisting Builders with predicting sales price and factors influencing the sales price

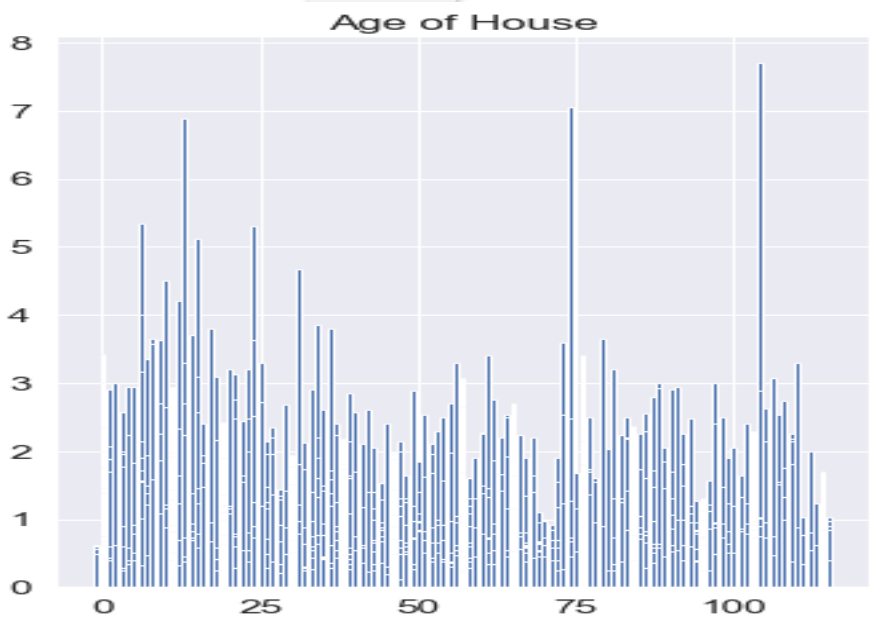


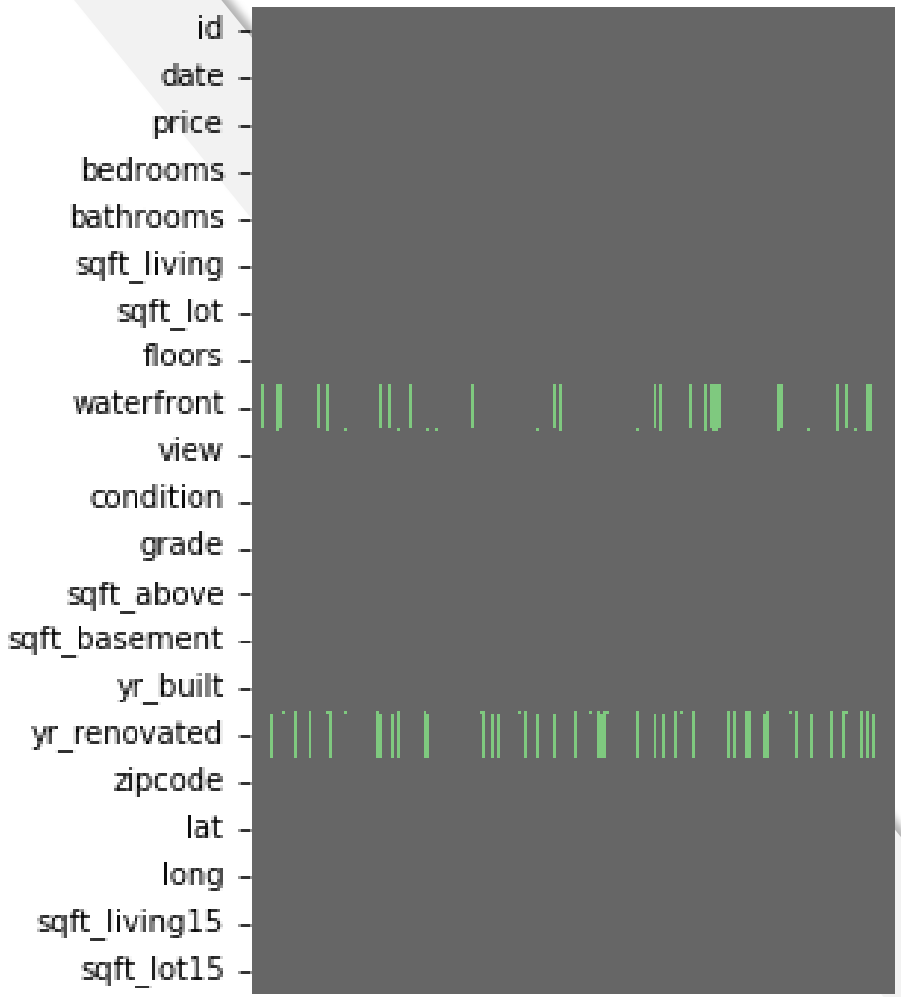




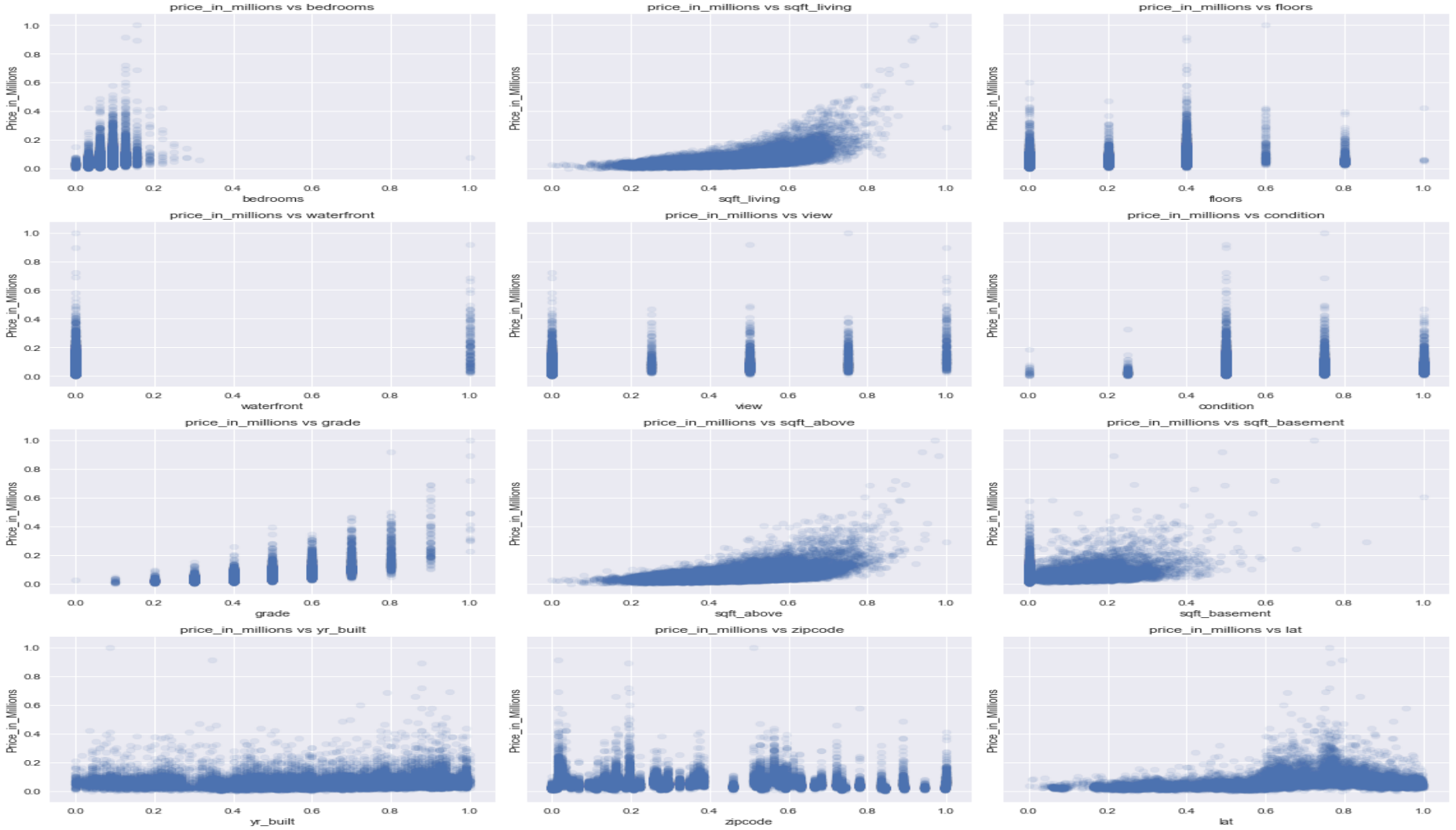
Quick Data Overview

- **Price :Target Prediction**
- **Price ranging: 78K-770k**
- **Property Age :0Y - 114Y**
- **21 Columns & 21597 rows**





Exploratory Data Analysis Stage



OLS: Sqft_living & Sqft_living15

Dep. Variable:	price_in_millions	**R-squared:**	0.374
Model:	OLS	**Adj. R-squared:**	0.374
Method:	Least Squares	**F-statistic:**	1.292e+04
Date:	Mon, 02 Nov 2020	**Prob (F-statistic):**	0.00
Time:	07:55:02	**Log-Likelihood:**	39911.
No. Observations:	21597	**AIC:**	-7.982e+04
Df Residuals:	21595	**BIC:**	-7.980e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-0.0532	0.001	-51.399	0.000	-0.055	-0.051
sqft_living	0.2503	0.002	113.670	0.000	0.246	0.255

Omnibus:	19384.826	**Durbin-Watson:**	1.978
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1655521.719
Skew:	3.977	**Prob(JB):**	0.00
Kurtosis:	45.148	**Cond. No.**	10.3

Dep. Variable:	price_in_millions	**R-squared:**	0.296
Model:	OLS	**Adj. R-squared:**	0.296
Method:	Least Squares	**F-statistic:**	9069.
Date:	Mon, 02 Nov 2020	**Prob (F-statistic):**	0.00
Time:	07:55:23	**Log-Likelihood:**	38633.
No. Observations:	21597	**AIC:**	-7.726e+04
Df Residuals:	21595	**BIC:**	-7.725e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-0.0635	0.001	-47.640	0.000	-0.066	-0.061
sqft_living15	0.2197	0.002	95.230	0.000	0.215	0.224

Omnibus:	20314.294	**Durbin-Watson:**	1.976
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1914774.434
Skew:	4.272	**Prob(JB):**	0.00
Kurtosis:	48.330	**Cond. No.**	11.1

OLS: Grade & View

Dep. Variable:	price_in_millions	**R-squared:**	0.155
Model:	OLS	**Adj. R-squared:**	0.155
Method:	Least Squares	**F-statistic:**	3951.
Date:	Mon, 02 Nov 2020	**Prob (F-statistic):**	0.00
Time:	07:55:05	**Log-Likelihood:**	36661.
No. Observations:	21597	**AIC:**	-7.332e+04
Df Residuals:	21595	**BIC:**	-7.330e+04
Df Model:	1		
Covariance Type:	nonrobust		

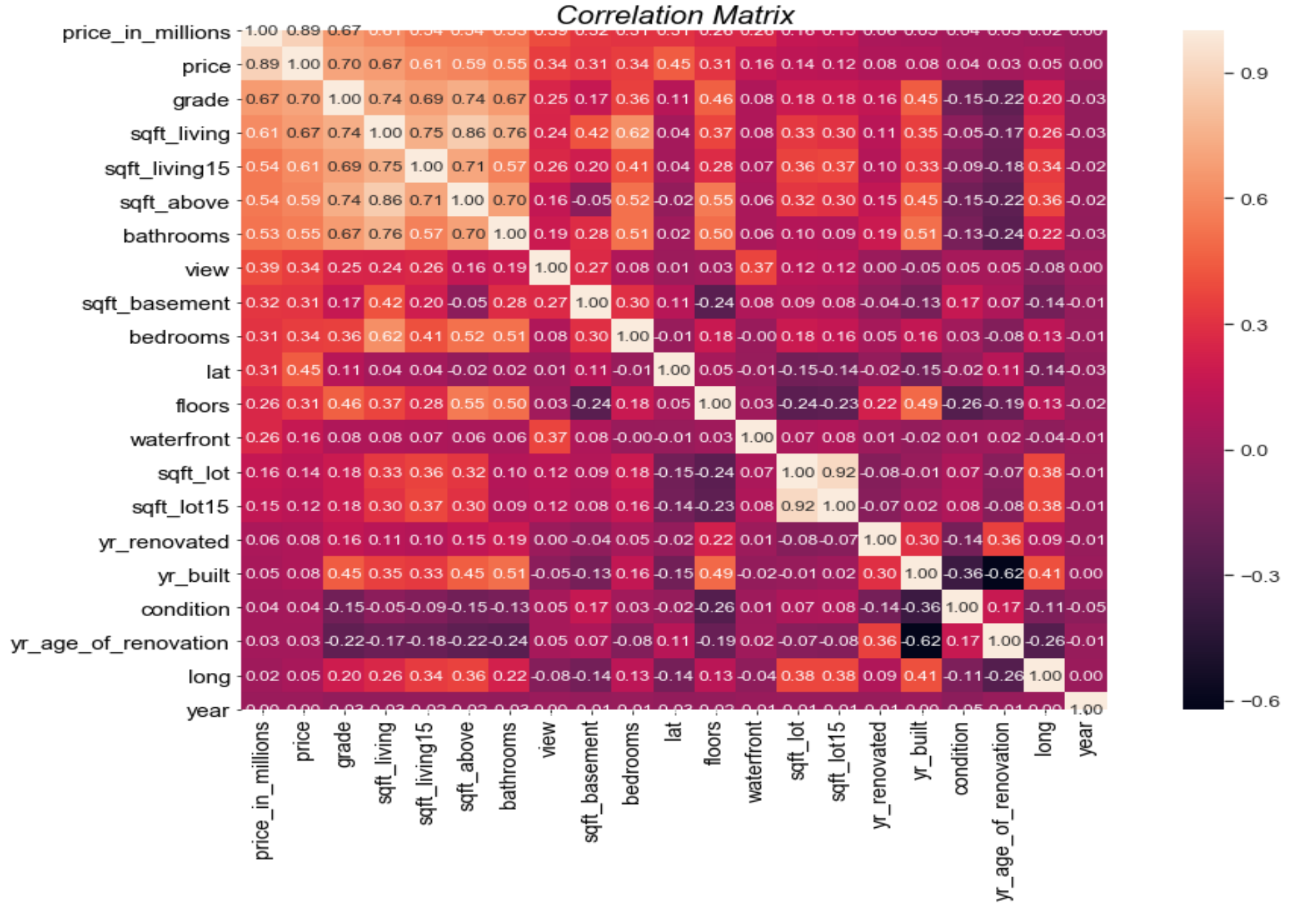
	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	0.0549	0.000	174.036	0.000	0.054	0.055
view	0.0991	0.002	62.854	0.000	0.096	0.102

Omnibus:	18082.913	**Durbin-Watson:**	1.961
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1077709.968
Skew:	3.669	**Prob(JB):**	0.00
Kurtosis:	36.820	**Cond. No.**	5.25

Dep. Variable:	price_in_millions	**R-squared:**	0.446
Model:	OLS	**Adj. R-squared:**	0.446
Method:	Least Squares	**F-statistic:**	1.740e+04
Date:	Mon, 02 Nov 2020	**Prob (F-statistic):**	0.00
Time:	07:55:10	**Log-Likelihood:**	41228.
No. Observations:	21597	**AIC:**	-8.245e+04
Df Residuals:	21595	**BIC:**	-8.244e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-0.0672	0.001	-67.209	0.000	-0.069	-0.065
grade	0.2744	0.002	131.895	0.000	0.270	0.278

Omnibus:	19879.964	**Durbin-Watson:**	1.968
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	2043898.709
Skew:	4.081	**Prob(JB):**	0.00
Kurtosis:	49.954	**Cond. No.**	10.4

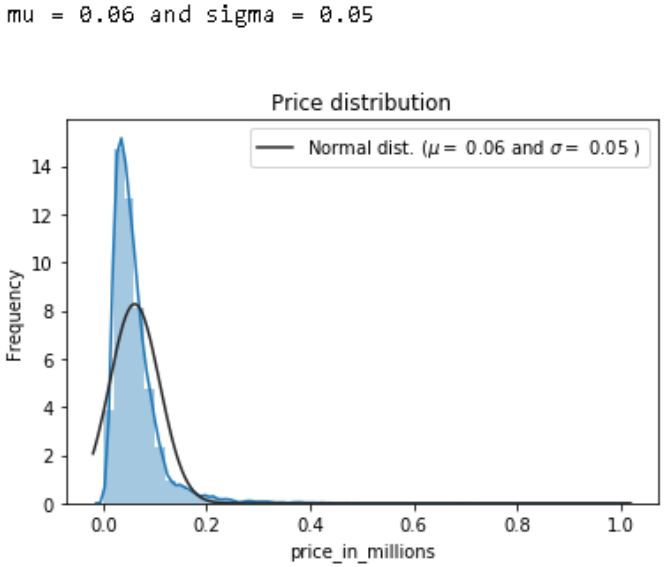

Correlation Matrix
price_in_millions
price
grade
sqft_living
sqft_living15
sqft_above
bathrooms
view
sqft_basement
bedrooms
lat
floors
waterfront
sqft_lot
sqft_lot15
yr_renovated
yr_built
condition
yr_age_of_renovation
long
year
0.9
0.6
0.3
0.0
−0.3
−0.6

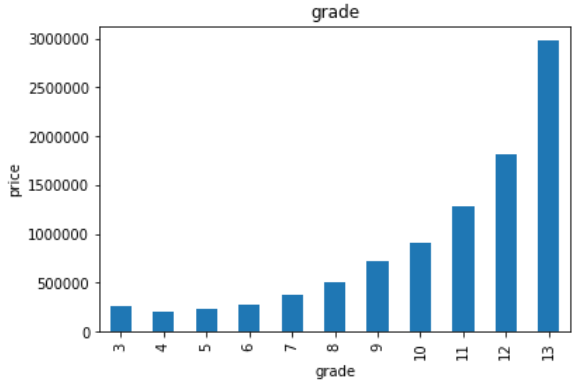
Final Model Score

LinearRegression	LinearRegression
r2_score 0.598	r2_score 0.623
RMSE = 0.03	nRMSE = 0.029
MSE = 0.001	MSE = 0.001
MAE = 0.018	MAE = 0.016
R2 = 0.598	R2 = 0.623
Adjusted R2 = 0.5971605568445475	Adjusted R2 = 0.6186657440919736
Mean Absolute Percentage Error = 37.051221507742724 %	Mean Absolute Percentage Error = 34.603584041739346 %
RandomForestRegressor	**RandomForestRegressor**
r2_score 0.746	r2_score 0.743
RMSE = 0.024	nRMSE = 0.024
MSE = 0.001	MSE = 0.001
MAE = 0.013	MAE = 0.012
R2 = 0.746	R2 = 0.743
Adjusted R2 = 0.7454696055684454	Adjusted R2 = 0.7400453480945284
Mean Absolute Percentage Error = 24.458483906399238 %	Mean Absolute Percentage Error = 24.19407842637895 %

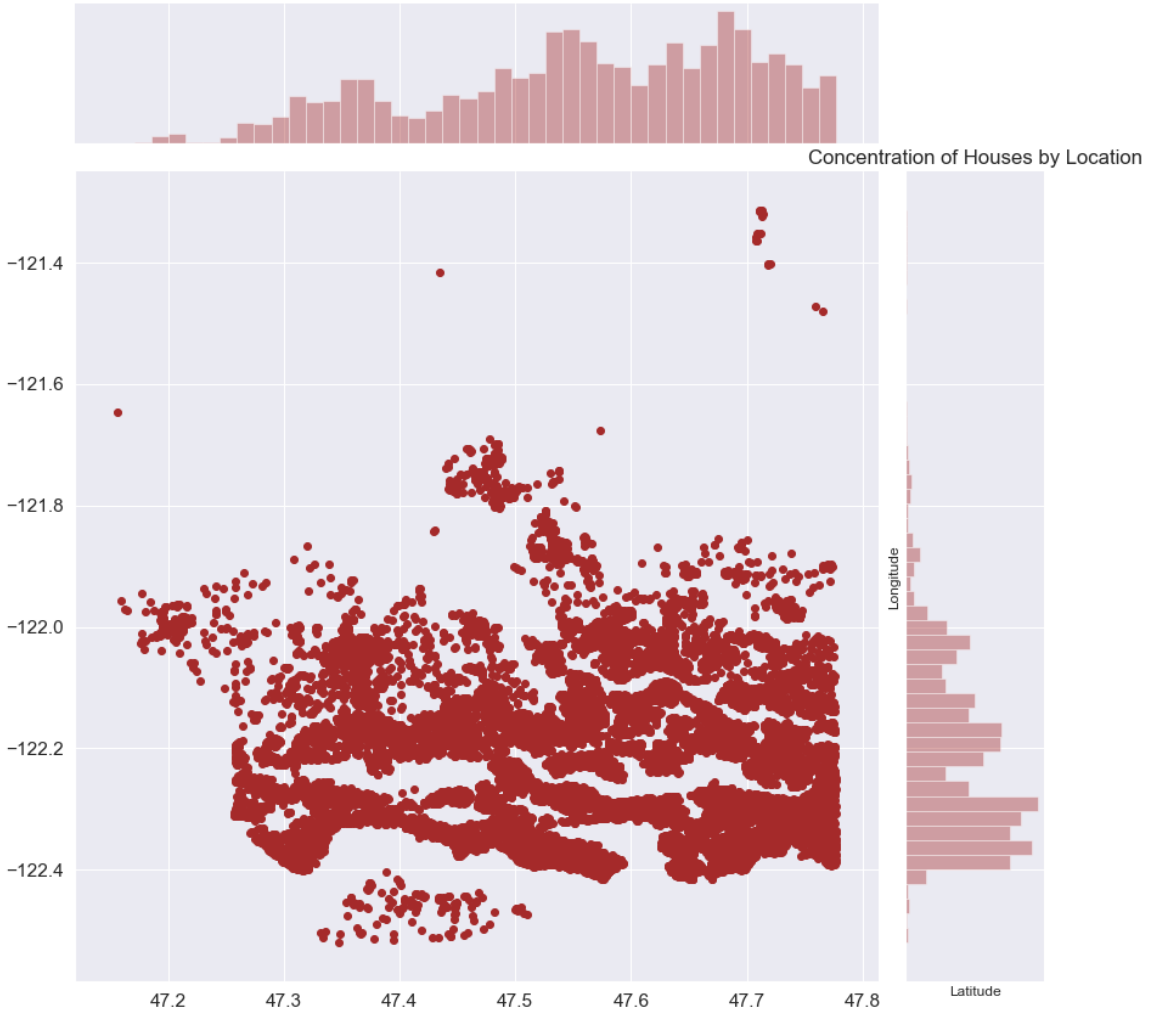
Recomendation

- There has to be property high in the grading system of Kings county
- The property has to be uniform with the neighbour
- Higher Square footage does not equate to higher prices or profit after surpassing certain size
- Price range of house between 100k and 200k sell the most





Next step : More research on Location



Thank you

- Github project link :
- https://github.com/ksis1st/Kings_County_Housing_Price_Prediction
- Linkedin :
- https://www.linkedin.com/in/kishor-shankaranarayan-bab2a311/