Recommendations for Property Valuation for Kings County Real Estate

SHIMNAZ FATHIMA

Outline of the Presentation

- Project Overview
- Methodology of the project
- Results of the analysis
- Inferences

Project Overview

o This project involves analysing sales data for Kings County properties.

o Multi-linear regression analysis was performed on this data

o Conclusions were derived from this analysis to provide Kings County Real Estate Agency.

Methodology

o Multi-linear regression was performed over 4 iterations.

o Initially the outliers were eliminated, and data was well understood.

o Check for linear regression theory assumptions was done.

o Check for multicollinearity was done and dummy variables were introduced.

o Log transformation and feature scaling was done to improve the model

o Key features were identified.

o Model validation was finally performed.

Results of the Analysis

o 1st Iteration


```
                        OLS Regression Results
==============================================================================
Dep. Variable:                  price   R-squared:                       0.563
Model:                            OLS   Adj. R-squared:                  0.562
Method:                 Least Squares   F-statistic:                     449.0
Date:                Sat, 02 Dec 2023   Prob (F-statistic):               0.00
Time:                        08:59:39   Log-Likelihood:                -64908.
No. Observations:                4889   AIC:                         1.298e+05
Df Residuals:                    4874   BIC:                         1.299e+05
Df Model:                          14
Covariance Type:            nonrobust
==============================================================================
                  coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      4.654e+06   1.85e+05     25.159      0.000    4.29e+06    5.02e+06
bedrooms      -1.232e+04   2831.520     -4.352      0.000   -1.79e+04   -6772.395
bathrooms      2.201e+04   4556.691      4.830      0.000    1.31e+04    3.09e+04
sqft_living      45.5252      4.076     11.168      0.000      37.534      53.516
sqft_lot          0.0584      0.384      0.152      0.879      -0.695       0.812
floors         7.623e+04   6186.661     12.321      0.000    6.41e+04    8.84e+04
waterfront     1.384e+05   3.35e+04      4.126      0.000    7.26e+04    2.04e+05
view           2.354e+04   2965.739      7.937      0.000    1.77e+04    2.94e+04
condition      2.19e+04    3283.019      6.671      0.000    1.55e+04    2.83e+04
grade          9.873e+04   3607.760     27.365      0.000    9.17e+04    1.06e+05
sqft_above       30.5714      5.591      5.468      0.000      19.610      41.533
sqft_basement    14.9538      6.322      2.365      0.018       2.559      27.349
yr_built      -2713.8090     95.362    -28.458      0.000   -2900.761   -2526.858
yr_renovated     12.2441      5.359      2.285      0.022       1.737      22.751
sqft_living15    77.3434      5.881     13.152      0.000      65.815      88.872
sqft_lot15       -3.3516      0.531     -6.315      0.000      -4.392      -2.311
==============================================================================
Omnibus:                      146.730   Durbin-Watson:                   2.007
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              217.362
Skew:                           0.304   Prob(JB):                     6.31e-48
Kurtosis:                       3.835   Cond. No.                     6.08e+16
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 3.48e-22. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```
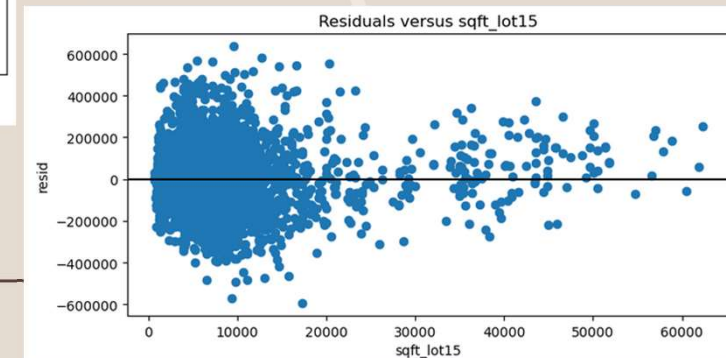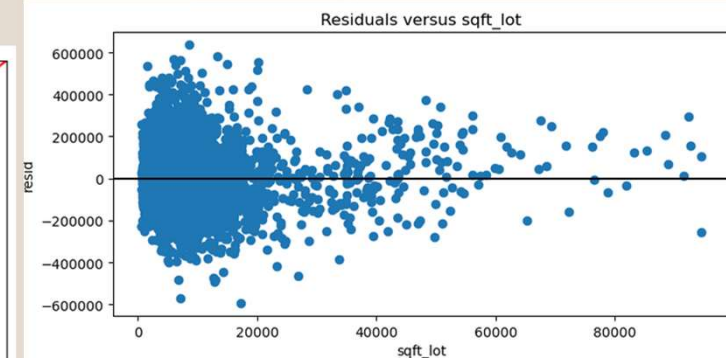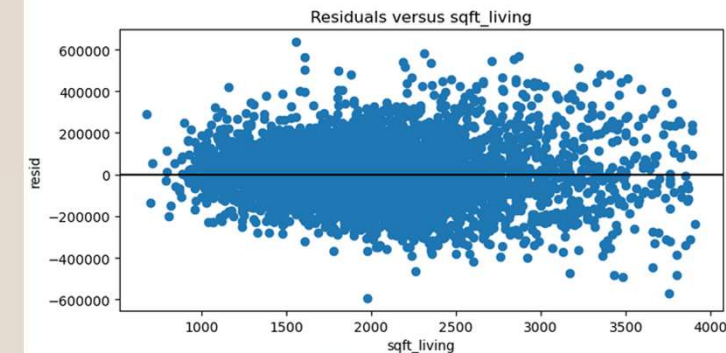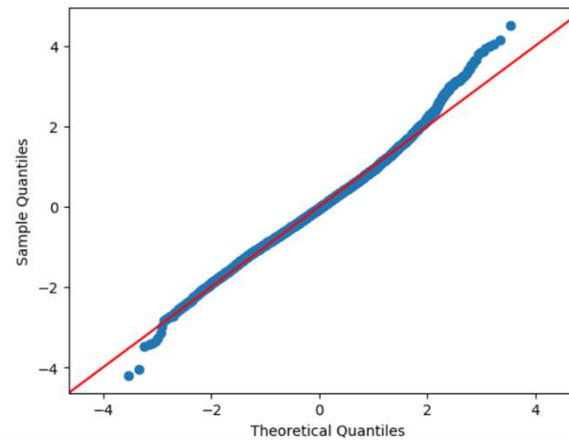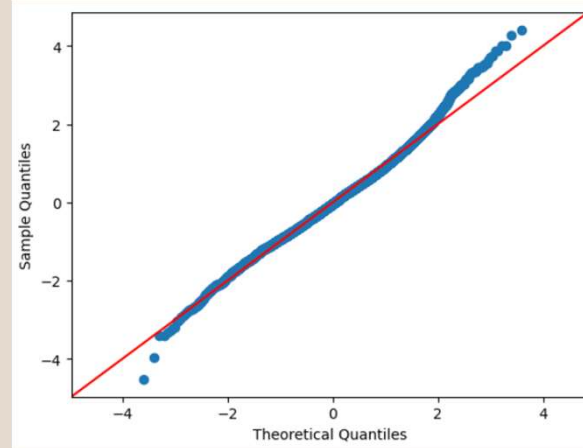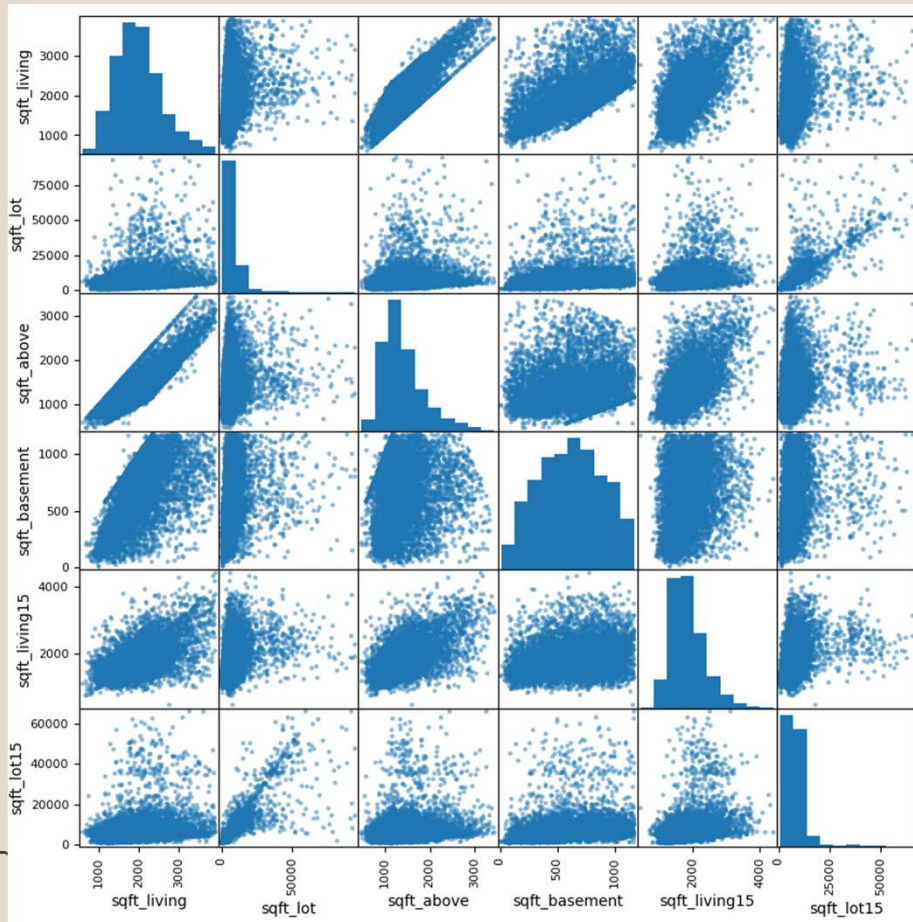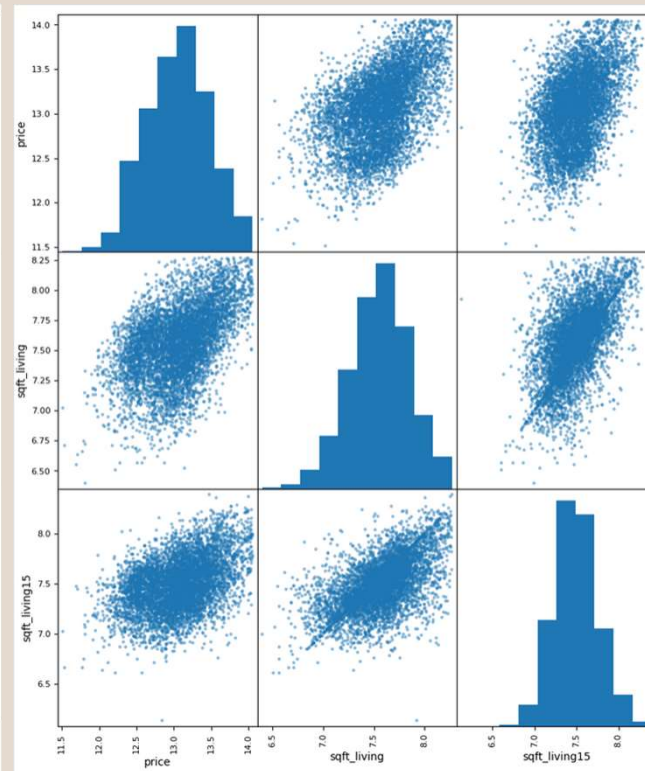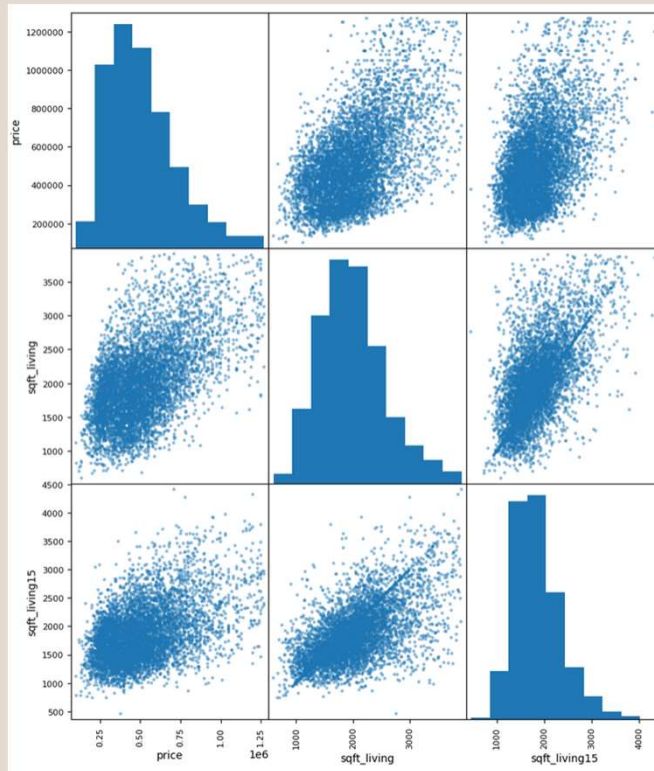
Results of the Analysis

o 2nd Iteration

Results of the Analysis

o 3ʳᵈ Iteration

Results of the Analysis

o 4th Iteration

Some of the key features identified:

- Sqft_living
- Grade (particularly on the higher scale)
- Sqft_living15
- No of Floors
- Condition

OLS Regression Results

Dep. Variable:	price	R-squared:	0.532
Model:	OLS	Adj. R-squared:	0.530
Method:	Least Squares	F-statistic:	217.2
Date:	Sat, 02 Dec 2023	Prob (F-statistic):	0.00
Time:	09:28:40	Log-Likelihood:	-6148.3
No. Observations:	5950	AIC:	1.236e+04
Df Residuals:	5918	BIC:	1.257e+04
Df Model:	31		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-0.6147	0.108	-5.684	0.000	-0.827	-0.403
sqft_living	0.1697	0.013	13.061	0.000	0.144	0.195
sqft_living15	0.1804	0.012	14.700	0.000	0.156	0.204
flr_2	0.2261	0.033	6.780	0.000	0.161	0.291
flr_3	0.5771	0.090	6.447	0.000	0.402	0.753
wtr_1	0.3390	0.168	2.013	0.044	0.009	0.669
view_1	0.1791	0.058	3.102	0.002	0.066	0.292
view_2	0.1525	0.040	3.804	0.000	0.074	0.231
view_3	0.1730	0.062	2.782	0.005	0.051	0.295
view_4	0.5154	0.094	5.455	0.000	0.330	0.701
cond_3	0.3594	0.101	3.567	0.000	0.162	0.557
cond_4	0.4533	0.101	4.480	0.000	0.255	0.652

cond_5	0.6196	0.104	5.958	0.000	0.416	0.824
grad_5	-0.4832	0.174	-2.770	0.006	-0.825	-0.141
grad_7	0.6551	0.039	16.780	0.000	0.579	0.732
grad_8	1.1078	0.045	24.778	0.000	1.020	1.195
grad_9	1.5666	0.058	26.943	0.000	1.453	1.681
grad_10	1.8306	0.081	22.498	0.000	1.671	1.990
grad_11	2.2129	0.215	10.286	0.000	1.791	2.635
bath_1	-0.0969	0.023	-4.255	0.000	-0.142	-0.052
y_reno_1970	-0.4240	0.184	-2.304	0.021	-0.785	-0.063
y_reno_2000	0.2353	0.089	2.655	0.008	0.062	0.409
y_reno_2010	0.1803	0.092	1.965	0.050	0.000	0.360
y_built_1930	-0.2148	0.053	-4.016	0.000	-0.320	-0.110
y_built_1940	-0.2238	0.036	-6.168	0.000	-0.295	-0.153
y_built_1950	-0.5800	0.034	-16.831	0.000	-0.648	-0.512
y_built_1960	-0.9415	0.034	-27.663	0.000	-1.008	-0.875
y_built_1970	-1.0399	0.034	-30.718	0.000	-1.106	-0.974
y_built_1980	-1.0431	0.039	-26.771	0.000	-1.120	-0.967
y_built_1990	-1.2063	0.053	-22.699	0.000	-1.310	-1.102
y_built_2000	-0.9202	0.047	-19.680	0.000	-1.012	-0.828
y_built_2010	-0.8407	0.060	-13.973	0.000	-0.959	-0.723

Omnibus:	66.968	Durbin-Watson:	2.027
Prob(Omnibus):	0.000	Jarque-Bera (JB):	68.972
Skew:	-0.262	Prob(JB):	1.05e-15
Kurtosis:	3.056	Cond. No.	37.4

Results of the Analysis

o Model Validation

Data set divided into:

- Training data 12471 datapoints

- Testing Data 4157 data points

- Train Mean Squared Error: 0.103365

- Test Mean Squared Error: 0.103383

Inferences

The following inferences can be derived from the results:

o Based on the coefficients, the features with high impact on the price are sqft_living, floors, waterfront, condition, grade and renovation.

o For targeting high price properties:

- The number of floors of the house can influence the price. The log of the price will increase by 0.23-0.578 by having 2-3 floors. Which means, there will be an increase of price by 1.26-1.78 times

- Having a waterfront property can improve the price by 1.4 times the value of the property.

- Better the condition and grade (in terms of the Kings County Grading system), better will be the price of the property.

Inferences

o Recommendations that could be provided to sellers looking to improve the value of the house:

- Doing a renovation can increase the value of the house by around 1.25 times the value of the house.

- Properties with just one bathroom depreciated the value of the house by 0.91 times. Increasing the number of bathrooms while doing the renovations would be recommended.