

TEAM MEMBERS

Jane Martha

James Njoroge

Leon Maina

Daniel Wahome

Magdalene Ondimu

Faith Mwenda



- Understanding Your Home's Value:
- Insights and Recommendations Agenda:
- Introduction: Understanding the Factors Influencing Your Home's Value Model Performance:
- How Reliable Are Our Predictions? Key Predictors:
- What Influences Your Home's Price? Model Assumptions: Ensuring Accuracy Over Time Recommendations:
- Maximizing Your Home's Value Next Steps: Enhancing Our Understanding Further





EXECUTIVE SUMMARY

- 1. Introduction: Understanding Your Home's Value Your home's value is influenced by various factors such as size, location, condition, and amenities. Our analysis aims to uncover the key drivers of house prices to help you make informed decisions.
- 2. Multiple linear regression modeling is used in this study to provide homeowners with useful information as they navigate the challenging real estate market. As market analysts specializing in real estate, our objective is to discern the intricate relationships among diverse property attributes and their impact on the selling price of properties. By exploring the King County House Sales data collection, we intend to provide homeowners with enlightening guidance on how to raise the estimated value of their homes by making thoughtful improvements



EXECUTIVE SUMMARY

Key Predictors: What Influences Your Home's Price? Larger Living Spaces: Properties with larger living spaces tend to have higher prices. Grade: Homes with a grade of 8 (Good) also command higher prices. Visualization: Scatter plot showing the relationship between living space and price, with different grades highlighted.

Model Assumptions: Ensuring Accuracy Over Time Our models meet important assumptions such as normality of residuals, homoscedasticity, and linearity. Regular monitoring of model diagnostics ensures ongoing accuracy. Visualization: Residual plots demonstrating the model assumptions.

Recommendations: Maximizing Your Home's Value Prioritize Upgrades: Consider investing in upgrades that increase living space or improve the overall grade of your property. Regular Maintenance: Maintain your home in good condition to preserve its value over time. Visualization: Bar chart showing the impact of different upgrades on home value.



PROBLEM STATEMENT

The main problem for homeowners is the failure in to make strategic decisions in investing for the proper renovation of their rental homes. The homes end up depreciating hence losing customers. The homeowners lack knowledge on how to be responsible owners and hence end up suffering a great loss.

- Key Question
- What are Assessment criteria, and policies for renovations
- What factors are involved in renovations
- How to optimize returns for homeowners

PROBLEM STATEMENT

- The study aims to decipher the complex interactions between different property attributes and their impact on home sale price by utilizing multiple linear regression models.
- The ultimate objective is to furnish market analysis with practical insights, augmenting their capacity to proficiently counsel homeowners in making critically better decision with the ultimate change in real estate



DATA UNDERSTANDING

- The dataset originate from King county house sale dataset
- The rows and columns contain unique features.
- The Key Features:
- No of floors, waterfront, view, square footage, grade, condition.

DATA SET

```
In [2]: #Loading data set for Analysis
    kc = pd.read_csv(r"C:\Users\HP\Documents\Flatiron\Assignments\Phase 2 Project\kc_house_data.csv")
    kc.head(5)
```

\cap	n±.	[2]	١.
v	uL	[4]	٠

id	date	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	 grade	sqft_above	sqft_basement	yr_built	yr_r
0 7129300520	10/13/2014	221900	3	1.00	1180	5650	1.0	NaN	NONE	 7 Average	1180	0	1955	
1 6414100192	12/9/2014	538000	3	2.25	2570	7242	2.0	NO	NONE	 7 Average	2170	400	1951	
2 5631500400	2/25/2015	180000	2	1.00	770	10000	1.0	NO	NONE	 6 Low Average	770	0	1933	
3 2487200875	12/9/2014	604000	4	3.00	1960	5000	1.0	NO	NONE	 7 Average	1050	910	1965	
4 1954400510	2/18/2015	510000	3	2.00	1680	8080	1.0	NO	NONE	 8 Good	1680	0	1987	

5 rows × 21 columns

DATA SET

t_living	sqft_lot	floors	waterfront	view	 grade	sqft_above	sqft_basement	yr_built	yr_renovated	zipcode	lat	long	sqft_living15	sqft_lot15
1180	5650	1.0	NaN	NONE	 7 Average	1180	0	1955	0.0	98178	47.5112	-122.257	1340	5650
2570	7242	2.0	NO	NONE	 7 Average	2170	400	1951	1991.0	98125	47.7210	-122.319	1690	7639
770	10000	1.0	NO	NONE	 6 Low Average	770	0	1933	NaN	98028	47.7379	-122.233	2720	8062
1960	5000	1.0	NO	NONE	 7 Average	1050	910	1965	0.0	98136	47.5208	-122.393	1360	5000
1680	8080	1.0	NO	NONE	 8 Good	1680	0	1987	0.0	98074	47.6168	-122.045	1800	7503

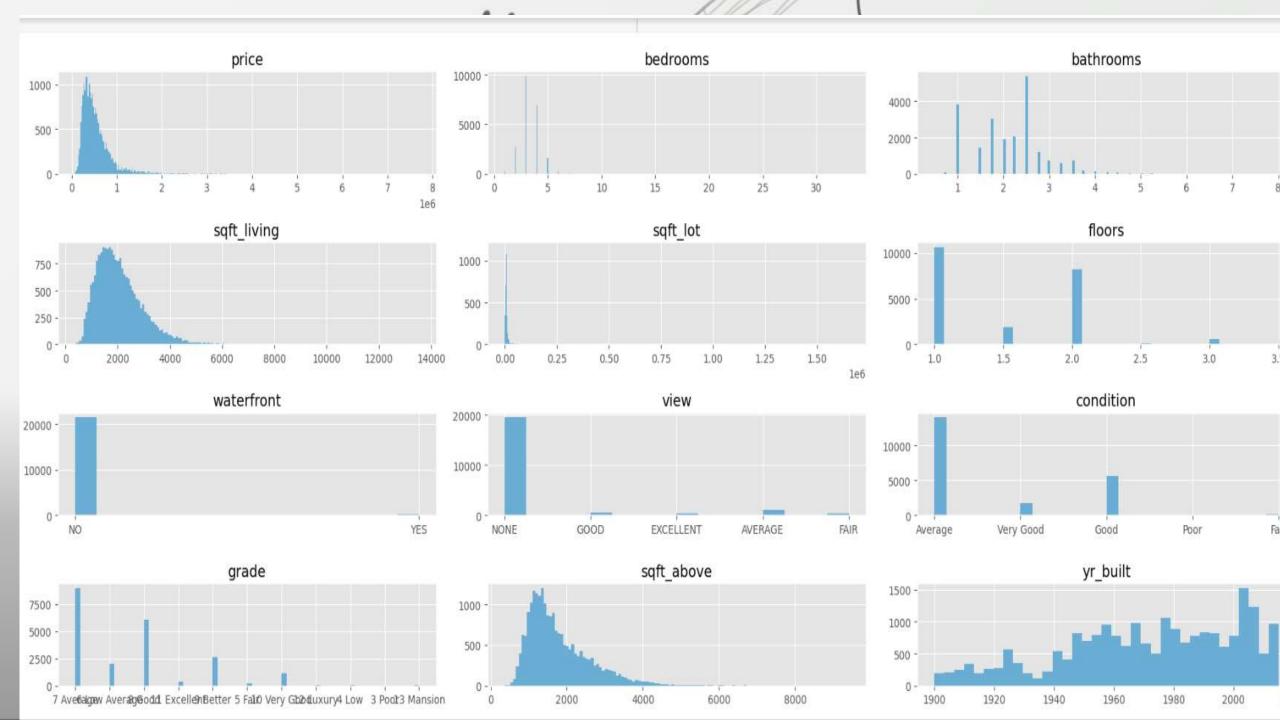
DATA UNDERSTANDING

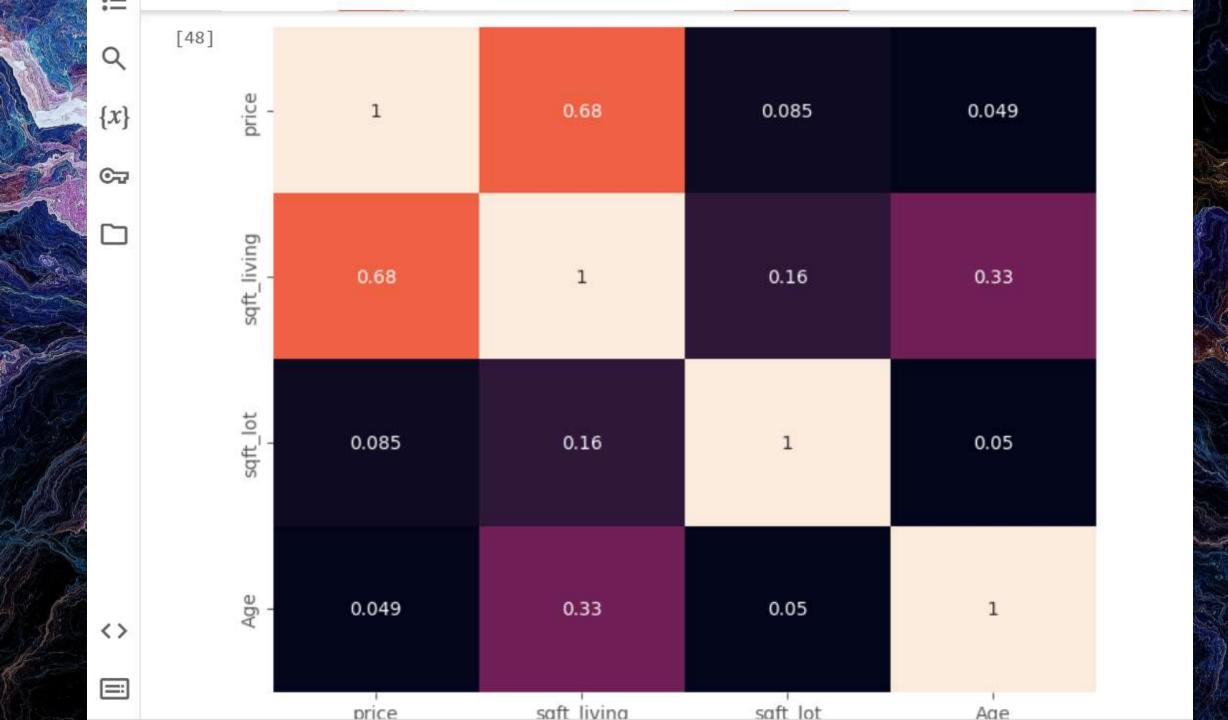
#Data understanding and exploration

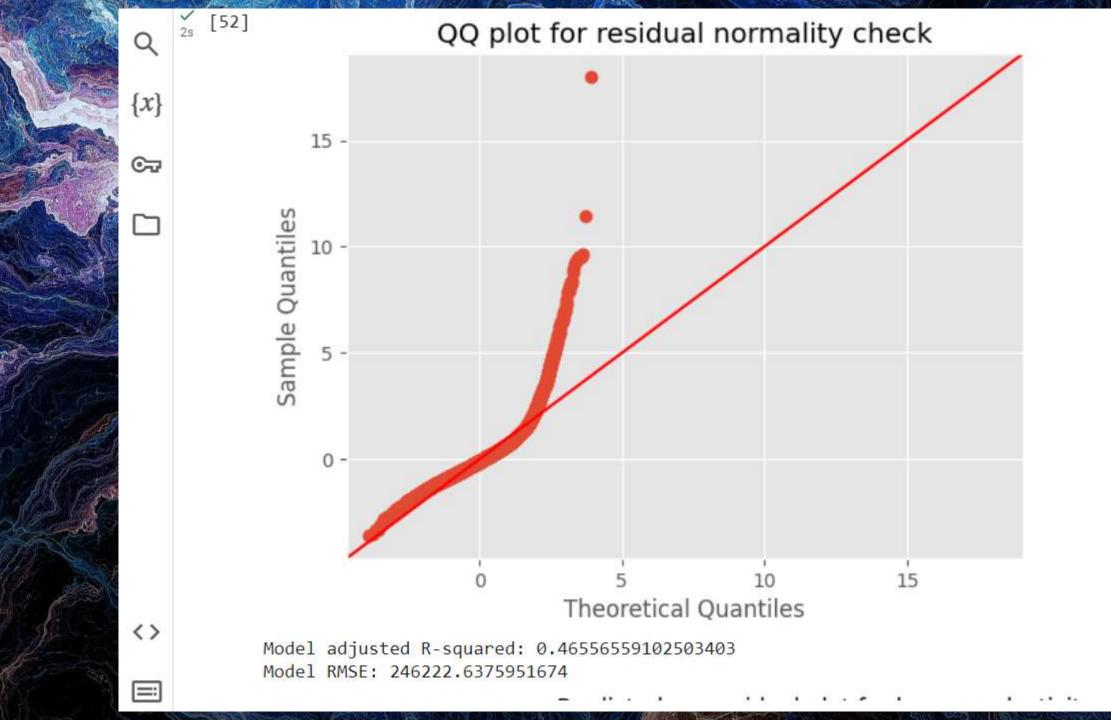
```
kc.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21597 entries, 0 to 21596
Data columns (total 21 columns):
    Column
                   Non-Null Count
                                  Dtype
   id
                                  int64
 0
                   21597 non-null
 1
   date
                   21597 non-null object
   price
                   21597 non-null
                                  int64
   bedrooms
                   21597 non-null
                                  int64
    bathrooms
                   21597 non-null float64
                   21597 non-null
   saft living
                                  int64
   sqft lot
                   21597 non-null
                                  int64
   floors
                   21597 non-null float64
   waterfront
                   19221 non-null
                                  object
  view
                   21534 non-null
                                  object
   condition
                                  object
                   21597 non-null
 11
   grade
                   21597 non-null
                                  object
                   21597 non-null
                                  int64
 12 sqft above
 13
    sqft_basement
                  21597 non-null
                                  object
 14 yr built
                   21597 non-null
                                  int64
 15
   yr renovated
                   17755 non-null
                                  float64
 16
   zipcode
                   21597 non-null
                                  int64
    lat
                   21597 non-null float64
 17
                   21597 non-null float64
 18
    long
```

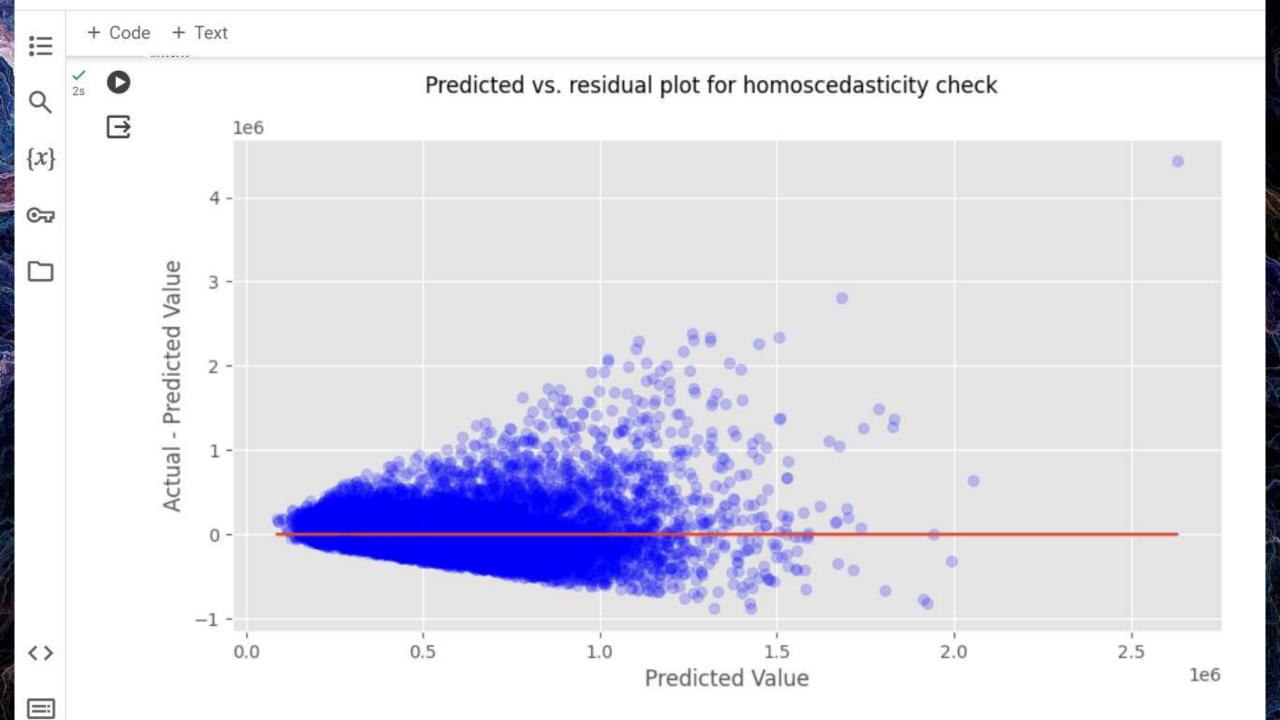
DATA UNDERSTANDING CONT'D

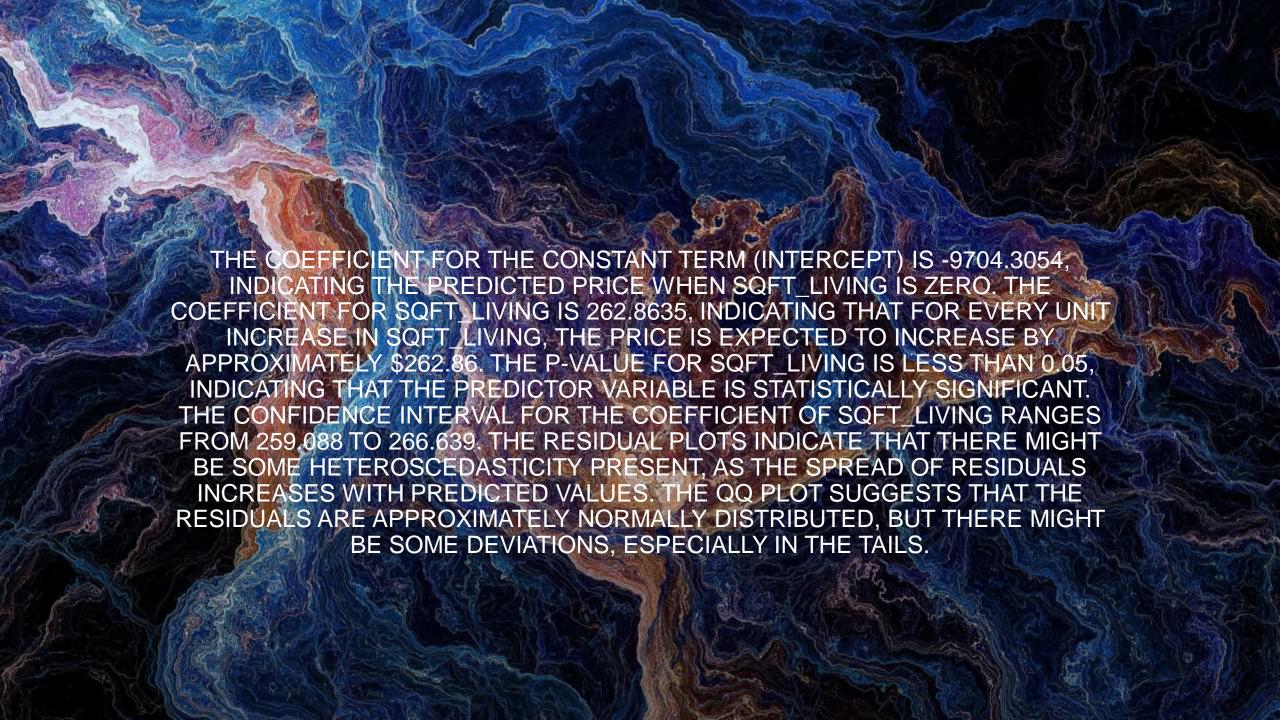
```
18 long 21597 non-null float64
19 sqft_living15 21597 non-null int64
20 sqft_lot15 21597 non-null int64
dtypes: float64(5), int64(10), object(6)
memory usage: 3.5+ MB
```

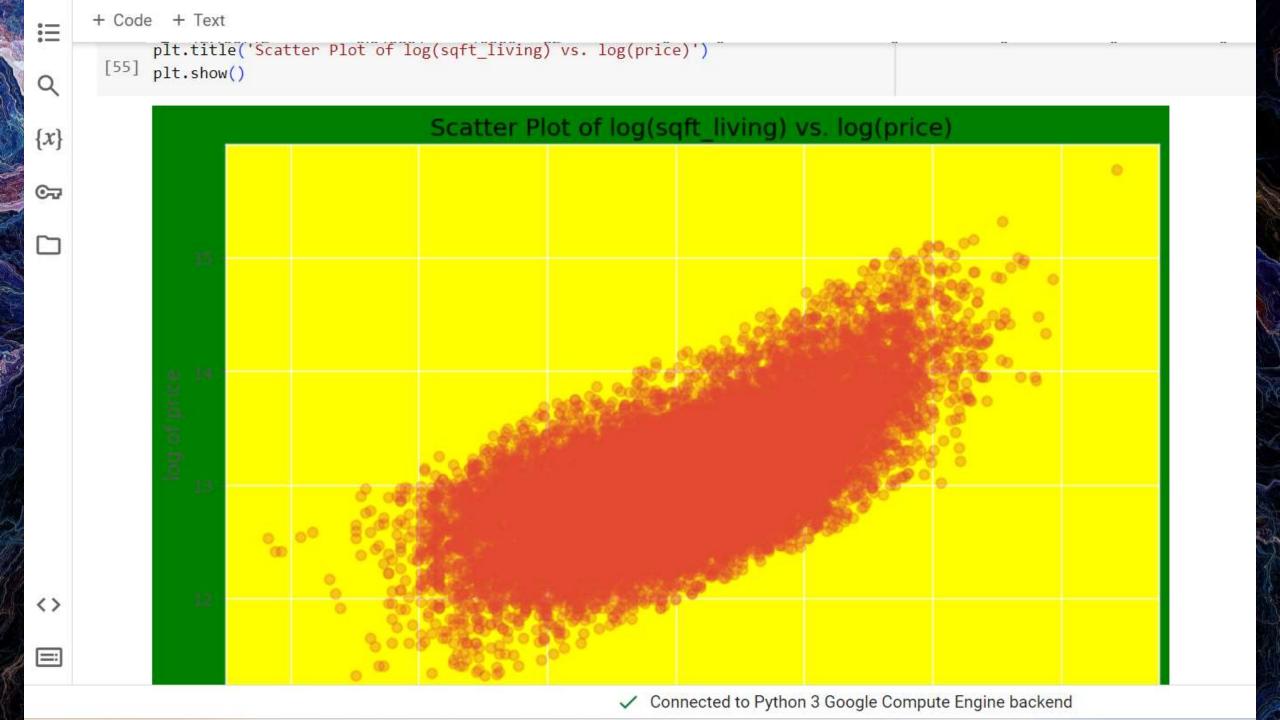


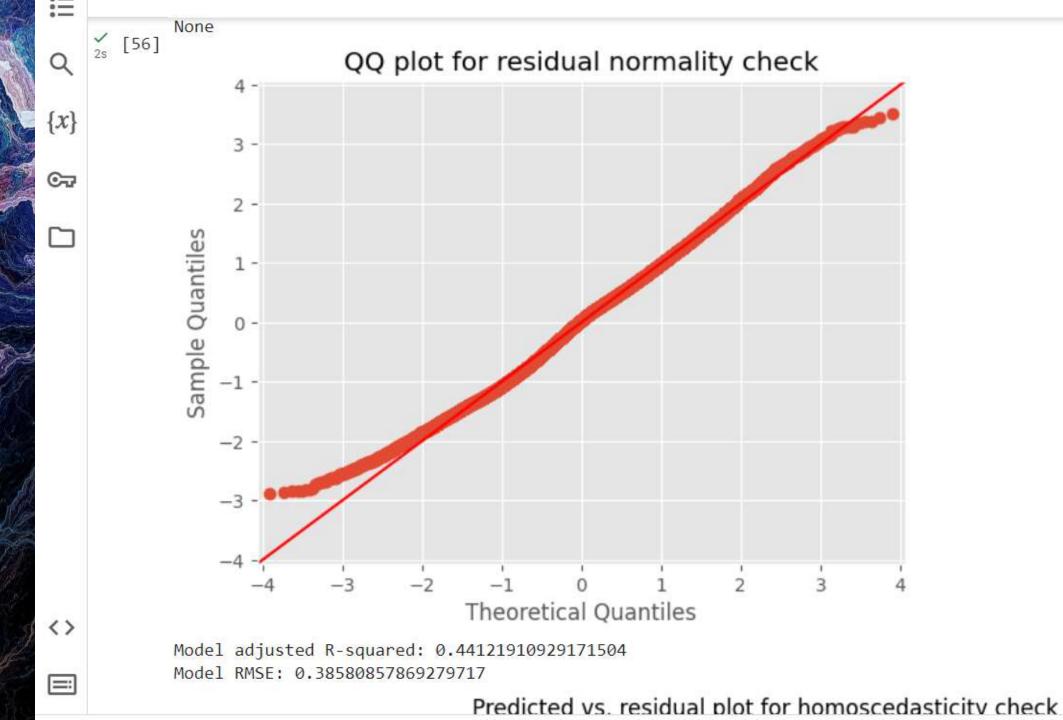


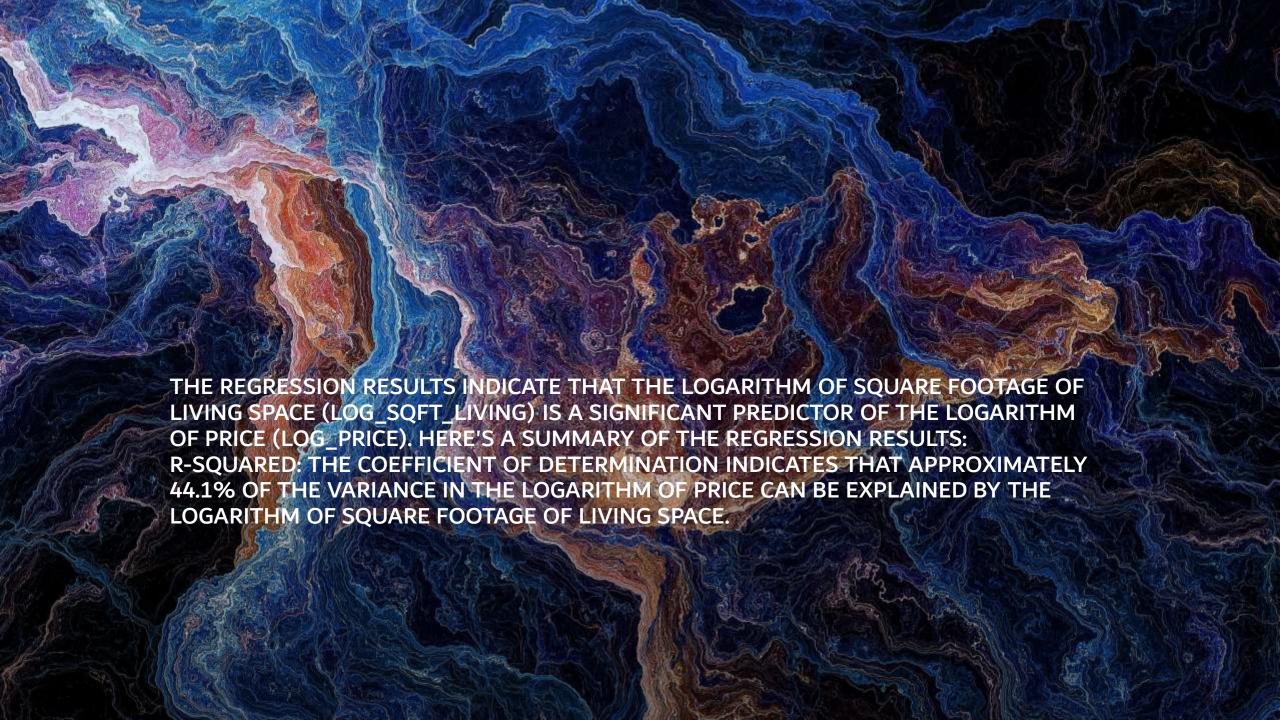


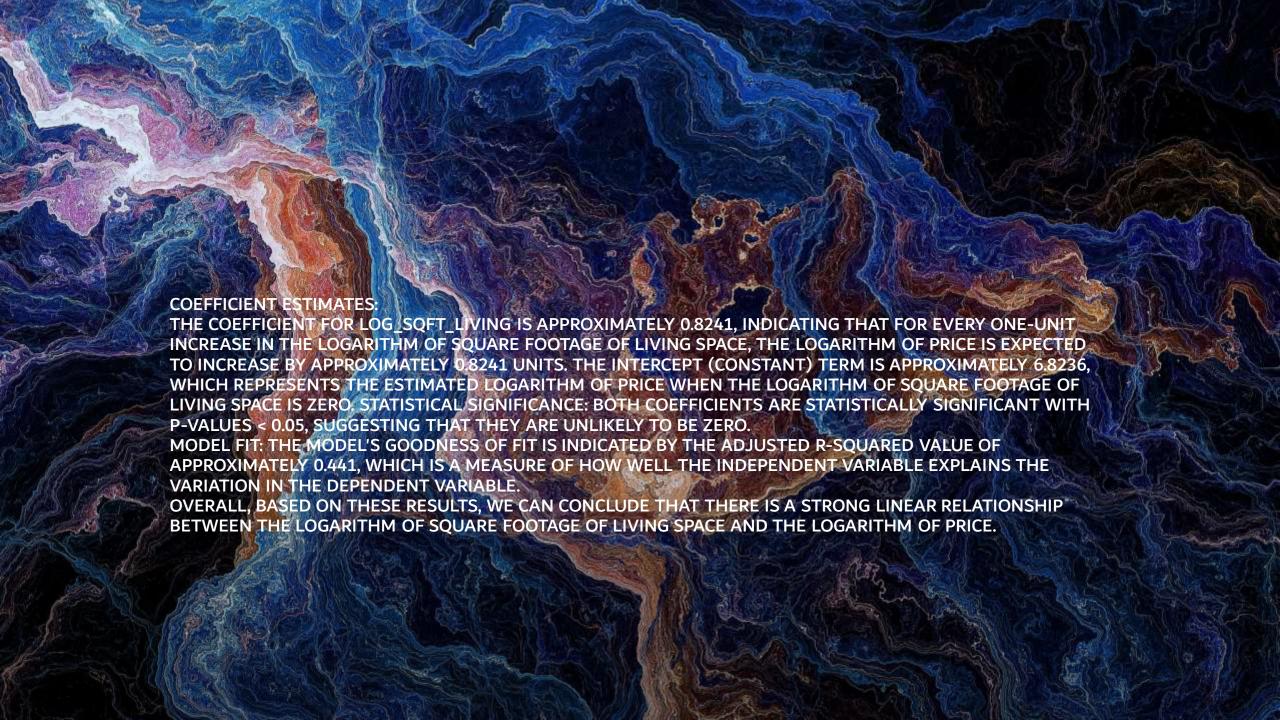


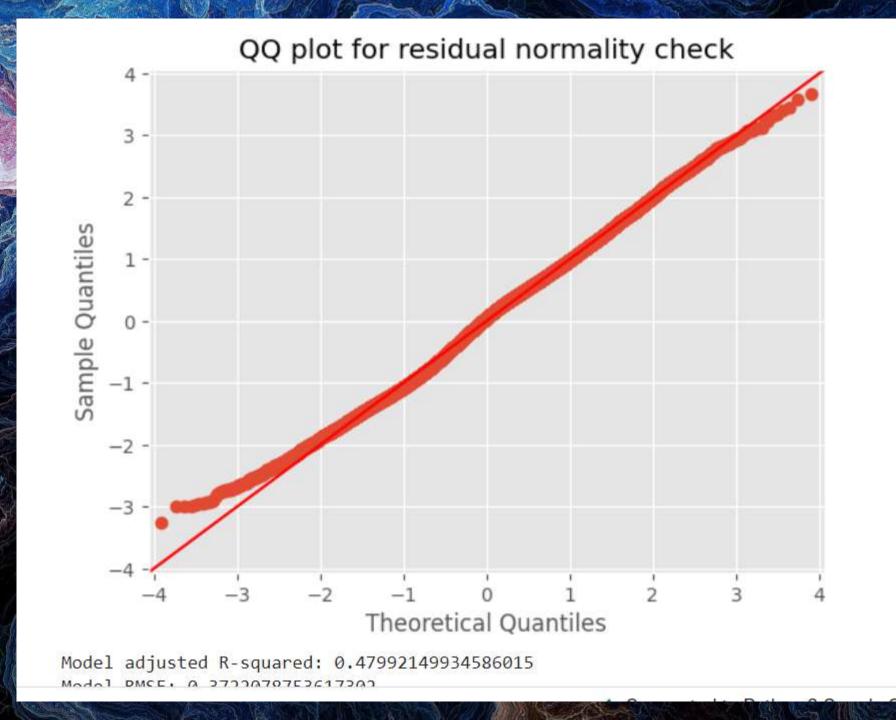


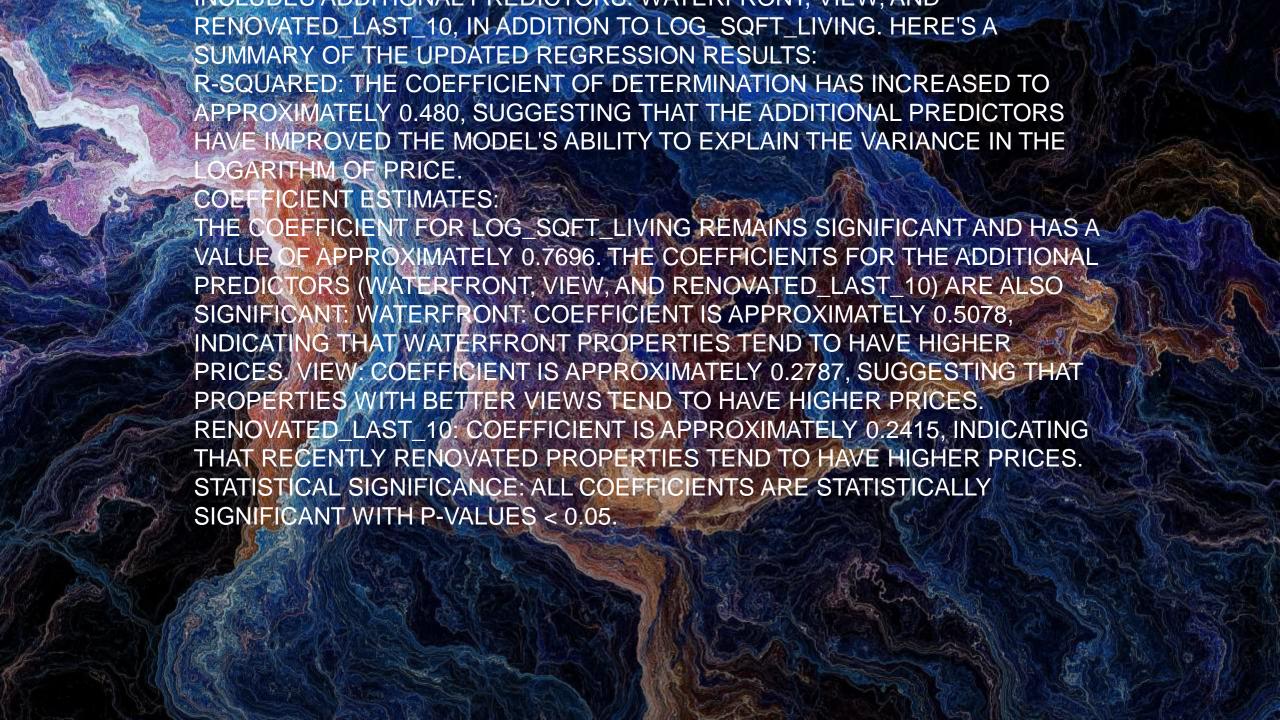


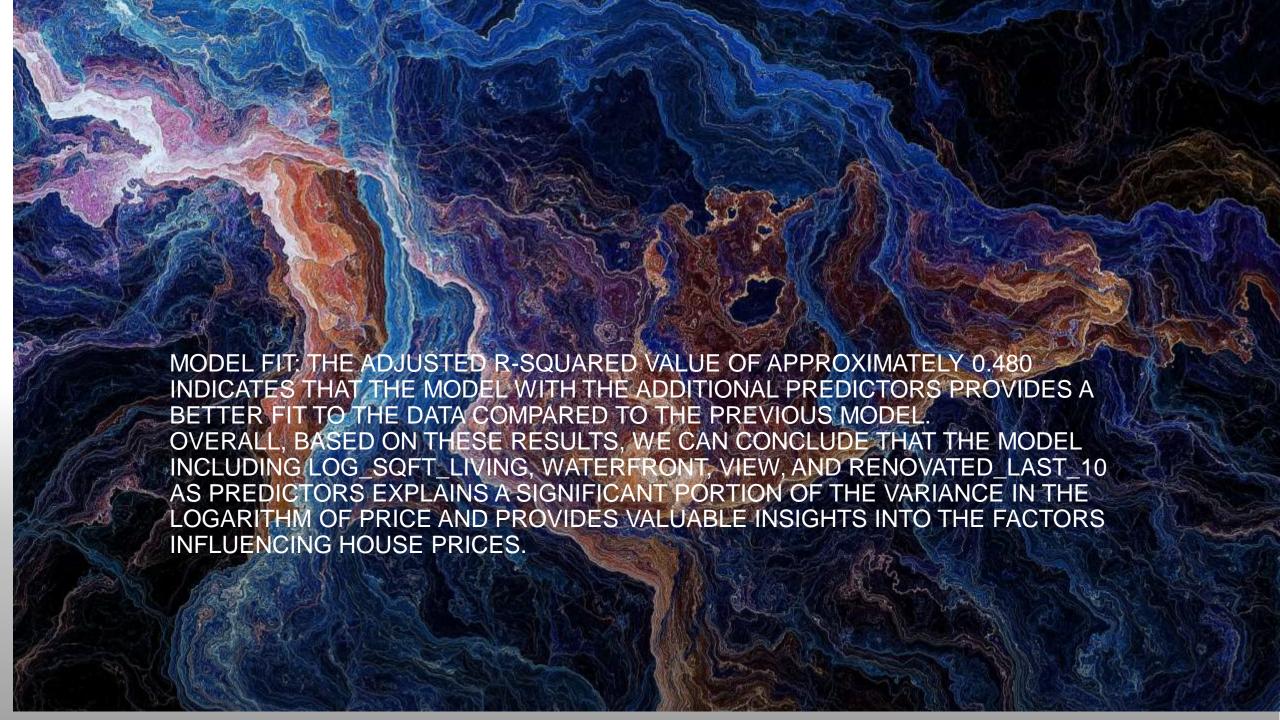












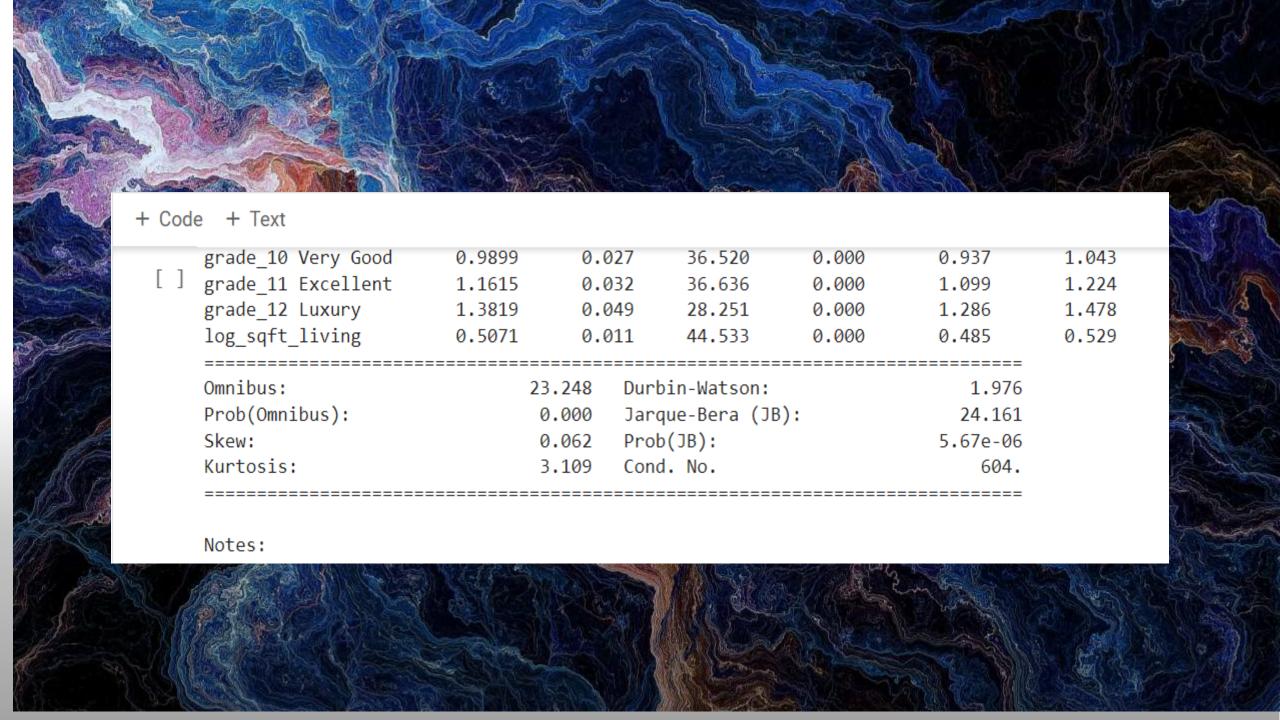
```
[62] # Defining target variable and predictors
Q
           y = kc_transformed.log_price
\{x\}
           X = kc_transformed[floors_dummies +
                              condition dummies +
C
                              grade_dummies +
                              ['view']+
                               'waterfront']+
                              ['renovated last 10']+
                              ['log_sqft_living']]
           model = reg_qq_sced(y, X)
                                    OLS Regression Results
           Dep. Variable:
                                    log price
                                               R-squared:
                                                                             0.590
                                               Adj. R-squared:
                                                                             0.590
           Model:
                                          OLS
           Method:
                                 Least Squares
                                               F-statistic:
                                                                             1709.
                              Thu, 28 Mar 2024
                                               Prob (F-statistic):
                                                                              0.00
           Date:
                                               Log-Likelihood:
                                                                           -6659.6
           Time:
                                     14:32:48
           No. Observations:
                                        21378
                                               AIC:
                                                                         1.336e+04
           Df Residuals:
                                        21359
                                               BIC:
                                                                         1.351e+04
           Df Model:
                                           18
           Covariance Type:
                                    nonrobust
<>
           ______
                                                                P> |t|
                                   coef
                                          std err
                                                                          [0.025
                                                                                    0.975]
0.000
                                                                          9.399
                                                                                     9.655
                                 9.5268
                                           0.065
                                                    145.617
           const

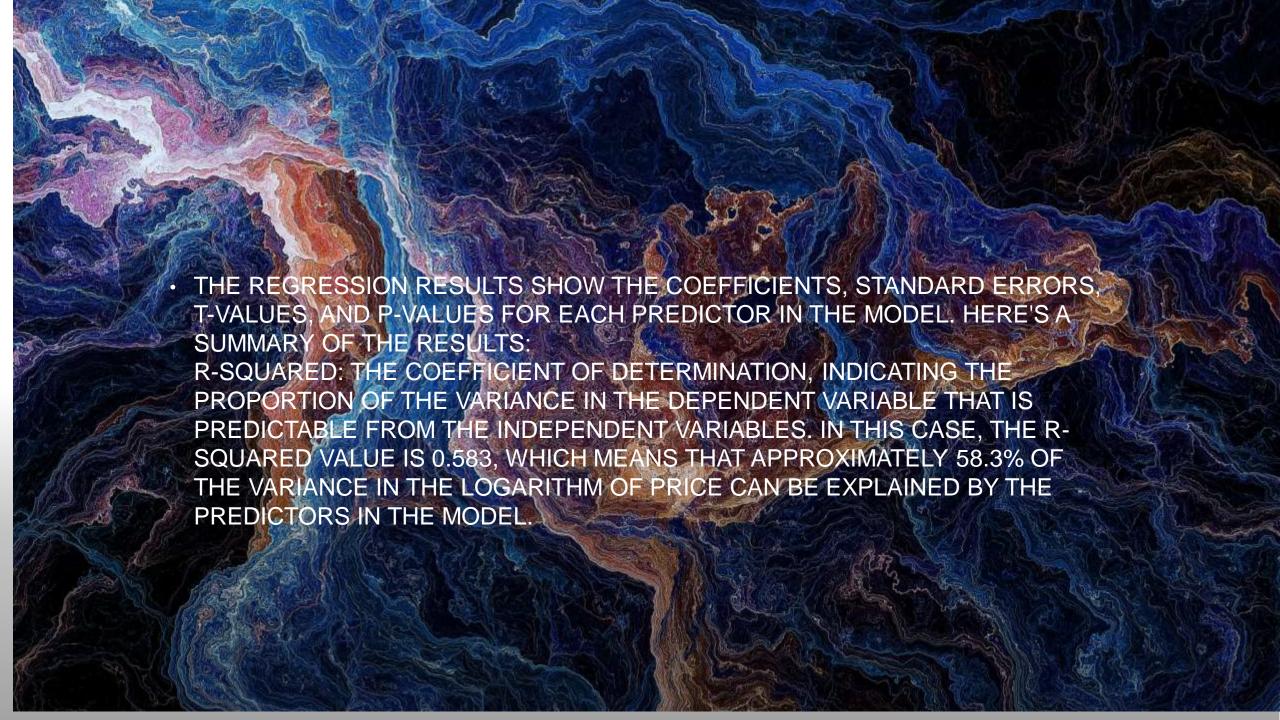
✓ Connected to Python 3 Google Compute Engine backend

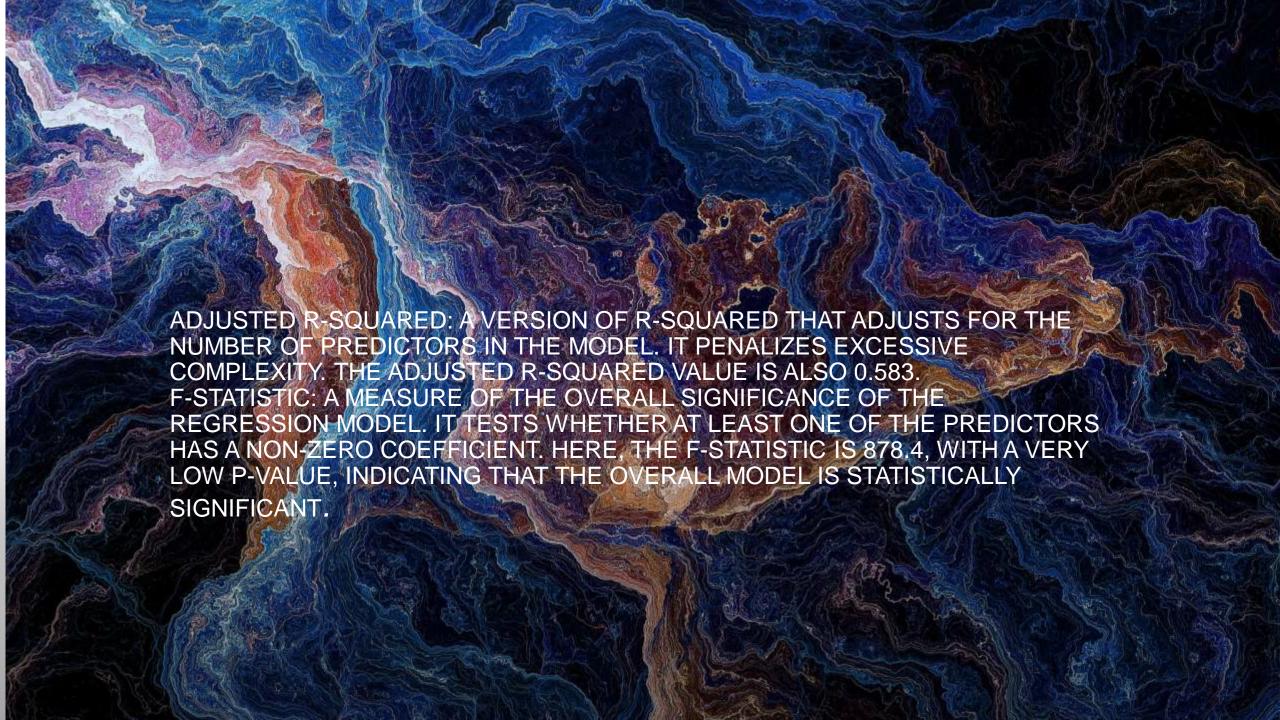
_____68°
                                                                                                                                  ^ 🧒 🙆 🛜 🗇 🗩
```

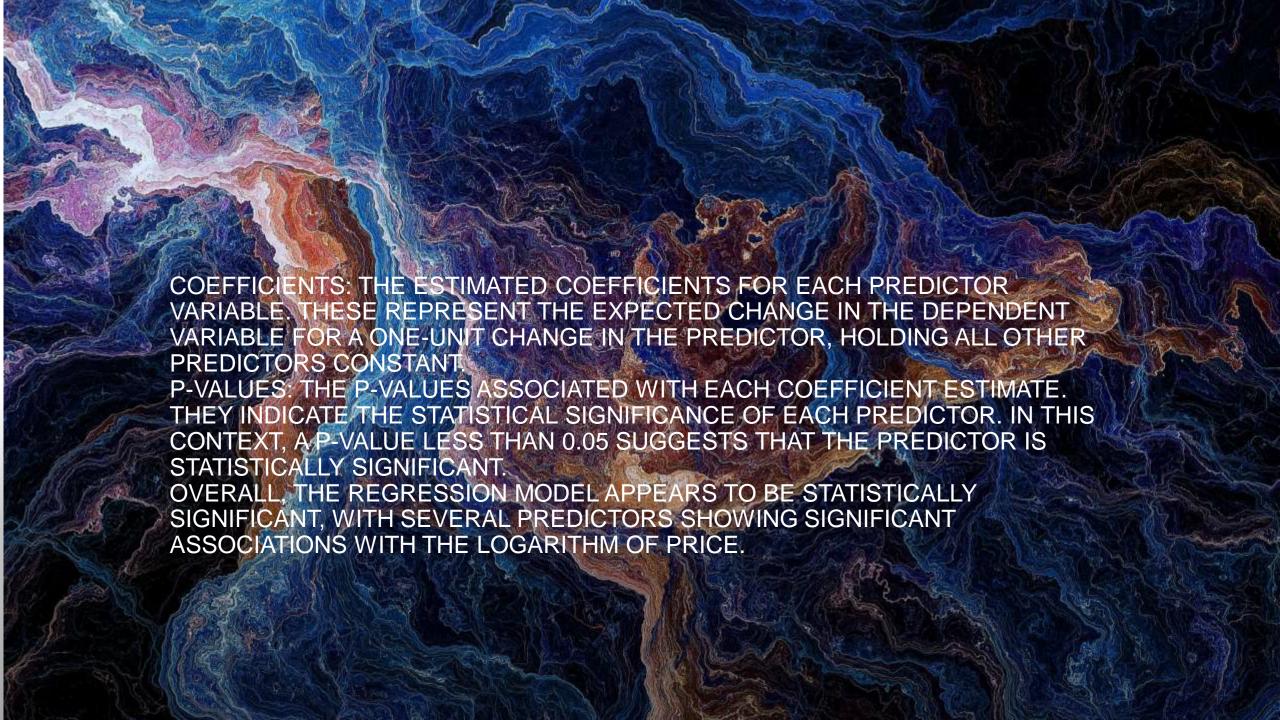
		coef	std err	t	P> t	[0.025	0.975]	
	const	8.8143	0.093	95.087	0.000	8.633	8.996	
	floors_1.5	0.1731	0.008	20.419	0.000	0.157	0.190	3
	floors_2.0	-0.0271	0.007	-4.061	0.000	-0.040	-0.014	The same of the sa
90	floors_2.5	0.1376	0.028	4.955	0.000	0.083	0.192	500
3	floors 3.0	0.1154	0.015	7.663	0.000	0.086	0.145	1 V
	bedrooms_2	-0.0525	0.027	-1.971	0.049	-0.105	-0.000	180
	bedrooms_3	-0.2178	0.027	-8.145	0.000	-0.270	-0.165	W.
	bedrooms_4	-0.2409	0.027	-8.798	0.000	-0.295	-0.187	
	bedrooms_5	-0.2386	0.029	-8.295	0.000	-0.295	-0.182	Wind.
	bedrooms 6	-0.2437	0.035	-6.981	0.000	-0.312	-0.175	
	bathrooms_1.0	-0.0238	0.046	-0.521	0.602	-0.113	0.066	SIN
O	bathrooms_1.5	-0.0738	0.047	-1.584	0.113	-0.165	0.018	18
	bathrooms_1.75	-0.0394	0.046	-0.850	0.395	-0.130	0.051	
	bathrooms_2.0	-0.0467	0.047	-1.004	0.315	-0.138	0.044	1000
<>	bathrooms_2.25	-0.0610	0.047	-1.302	0.193	-0.153	0.031	N. C.
	bathrooms_2.5	-0.1134	0.047	-2.427	0.015	-0.205	-0.022	
⊞	bathrooms_2.75	-0.0405	0.047	-0.853	0.394	-0.134	0.053	
	bathrooms 3.0	-0.0168	0.048	-0.350	0.727	-0.111	0.078	1 m

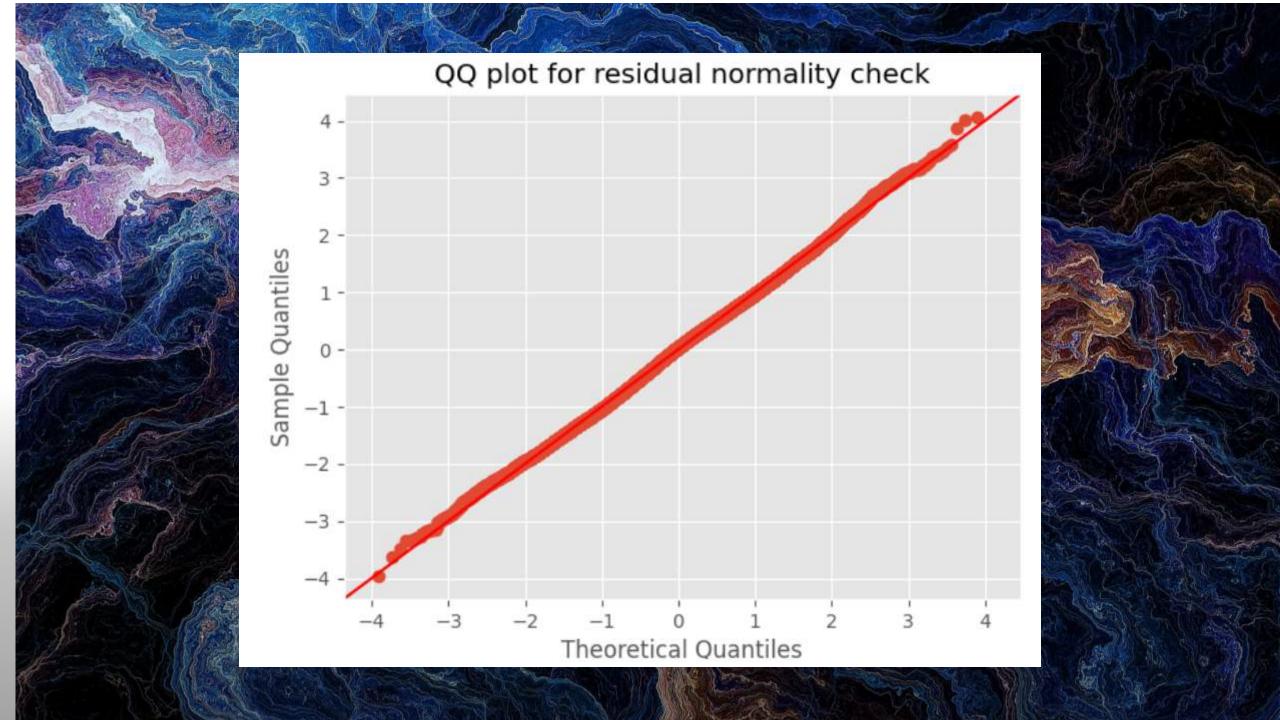
		coef	std err	t	P> t	[0.025	0.975]	
	const	8.8143	0.093	95.087	0.000	8.633	8.996	
	floors_1.5	0.1731	0.008	20.419	0.000	0.157	0.190	3
	floors_2.0	-0.0271	0.007	-4.061	0.000	-0.040	-0.014	The same of the sa
90	floors_2.5	0.1376	0.028	4.955	0.000	0.083	0.192	SW
3	floors 3.0	0.1154	0.015	7.663	0.000	0.086	0.145	1 V
	bedrooms_2	-0.0525	0.027	-1.971	0.049	-0.105	-0.000	180
	bedrooms_3	-0.2178	0.027	-8.145	0.000	-0.270	-0.165	W.
	bedrooms_4	-0.2409	0.027	-8.798	0.000	-0.295	-0.187	
	bedrooms_5	-0.2386	0.029	-8.295	0.000	-0.295	-0.182	Wind.
	bedrooms 6	-0.2437	0.035	-6.981	0.000	-0.312	-0.175	
	bathrooms_1.0	-0.0238	0.046	-0.521	0.602	-0.113	0.066	SIN
O	bathrooms_1.5	-0.0738	0.047	-1.584	0.113	-0.165	0.018	18
	bathrooms_1.75	-0.0394	0.046	-0.850	0.395	-0.130	0.051	
	bathrooms_2.0	-0.0467	0.047	-1.004	0.315	-0.138	0.044	1000
<>	bathrooms_2.25	-0.0610	0.047	-1.302	0.193	-0.153	0.031	N. C.
	bathrooms_2.5	-0.1134	0.047	-2.427	0.015	-0.205	-0.022	
⊞	bathrooms_2.75	-0.0405	0.047	-0.853	0.394	-0.134	0.053	
	bathrooms 3.0	-0.0168	0.048	-0.350	0.727	-0.111	0.078	



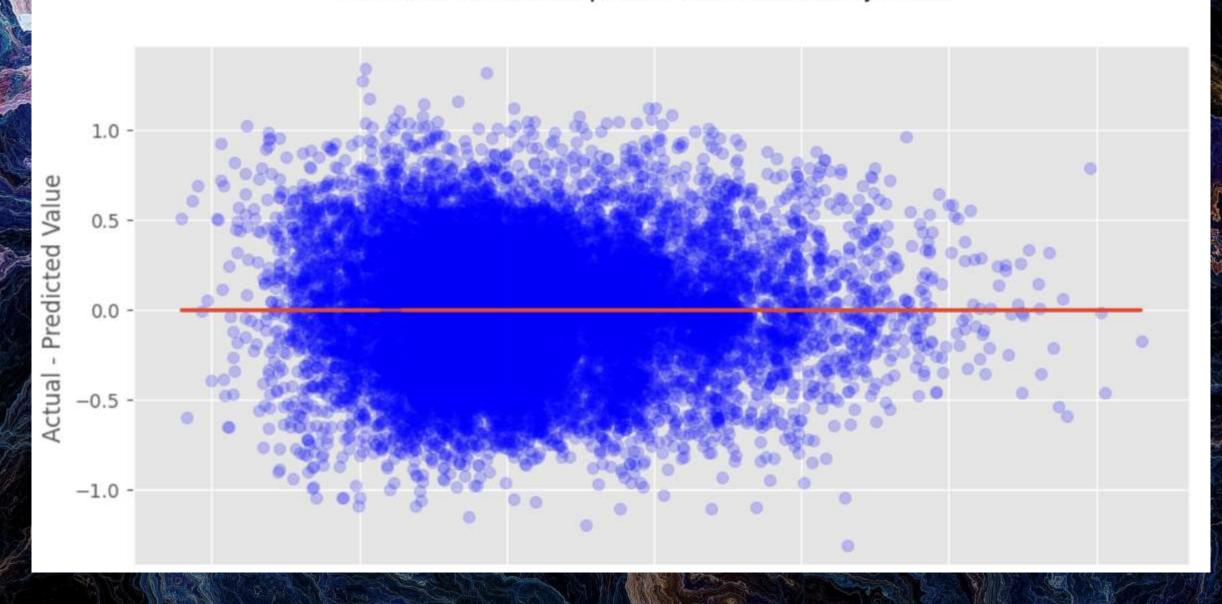


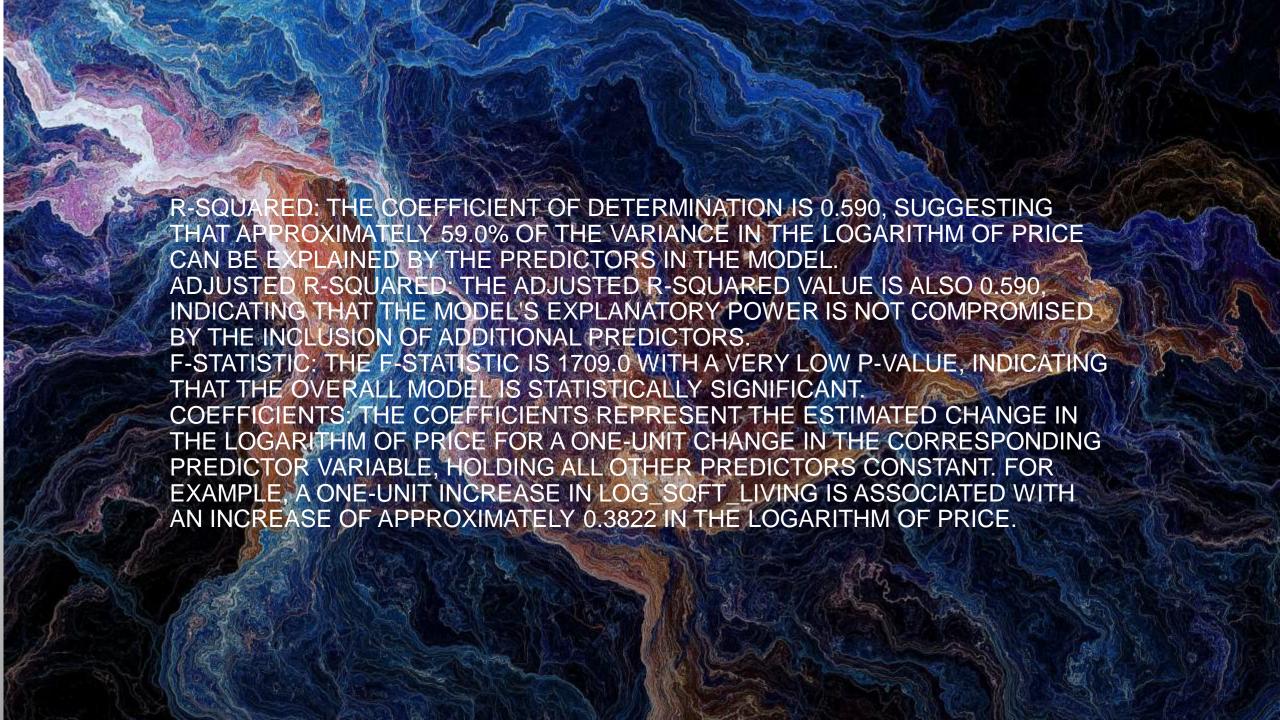


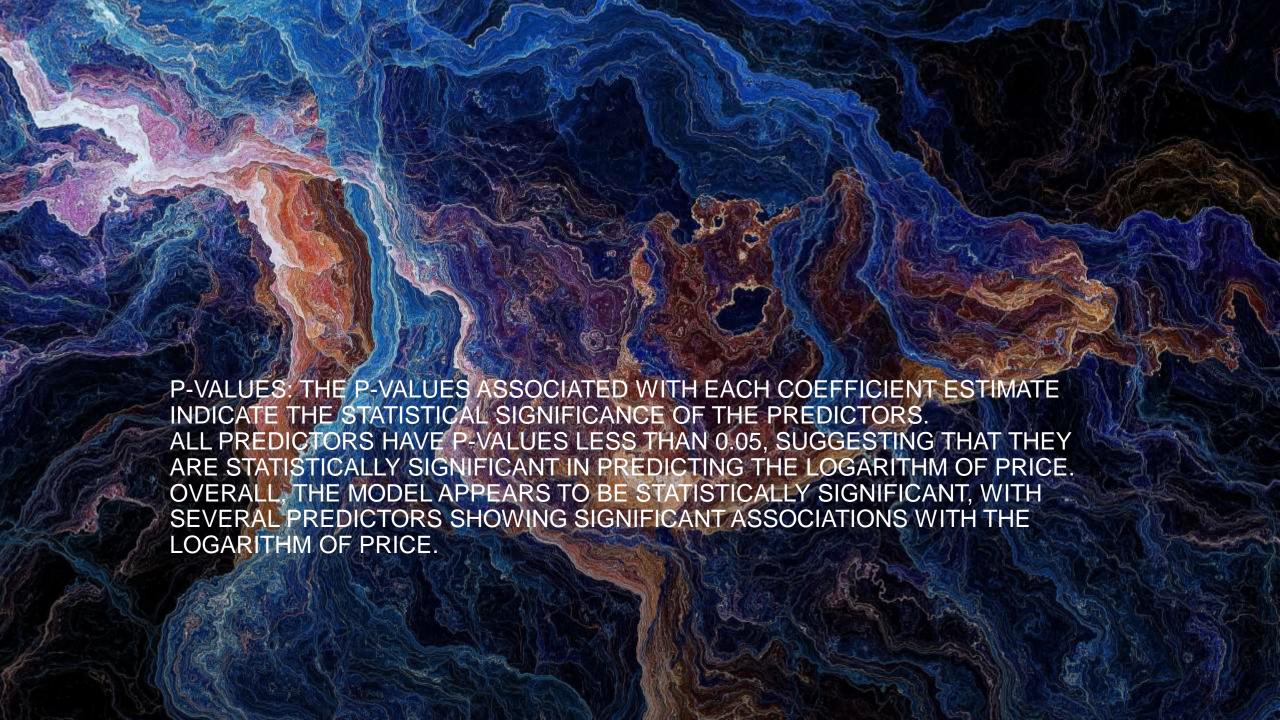




Predicted vs. residual plot for homoscedasticity check







OBSERVATIONS

OBSERVATIONS

- An observation is made based on market analysis that the property value type is the one that impact the correlation of renovation method selected.
- There is strong correlation between the price and the grade since the quality of the house determines the price in which the owner will spend on renovating

OBSERVATIONS

 Owners should give priority to the renovations that are more unique in order to be able to increase the property value.

POSITIVE EFFECTS OF RENOVATION

RENOVATION IMPACTS

- Owners estimating the duration in which were constructed and lastly renovated so as to know when to decide on renovation
- Renovating houses after a short duration makes it more attractive since customers will presume it as quality and durable.
- Paying attention to areas such floors, bathrooms, bedrooms, and other crucial areas makes it more

RENOVATIONS IMPACT

 The condition of the house gives it a proper estimate value.

WHAT TO GIVE PRIORTY

CONCLUSIONS

- Providing a large area for living space. This makes a customer feel better since there is adequate space instead of congestion.
- Proper maintenance. Ensuring that everything is in good condition such as no leakage, rotted wood, or rusted metals gives a convincing way for one to want to rent.
- Ensuring that the house structures are quality and in current trend.

RECOMMENDATIONS

- Be keen observation of the current trend
- Invest in proper and quality renovations
- Give priority to areas that will give you more profit
- Ensure frequent maintenance
- To avoid costs caused by customers' damage during their occupation in the houses, set extra charges for every damage by customers during the stay in.

CONCLUSION

 Owners of real estate should be ready to take risks when it comes to the interest of the customers. They should stay ahead and know the current trends in order to boost the quality of houses and meet customers expectations.

