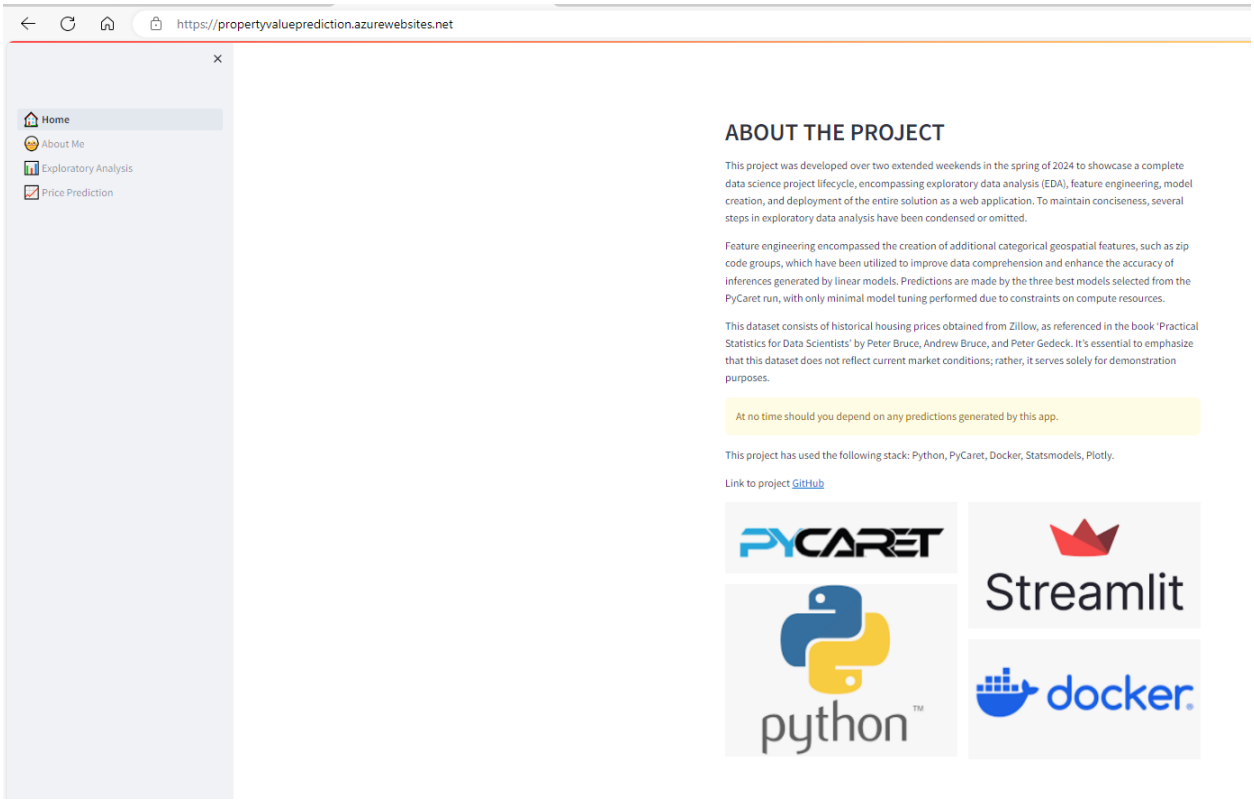
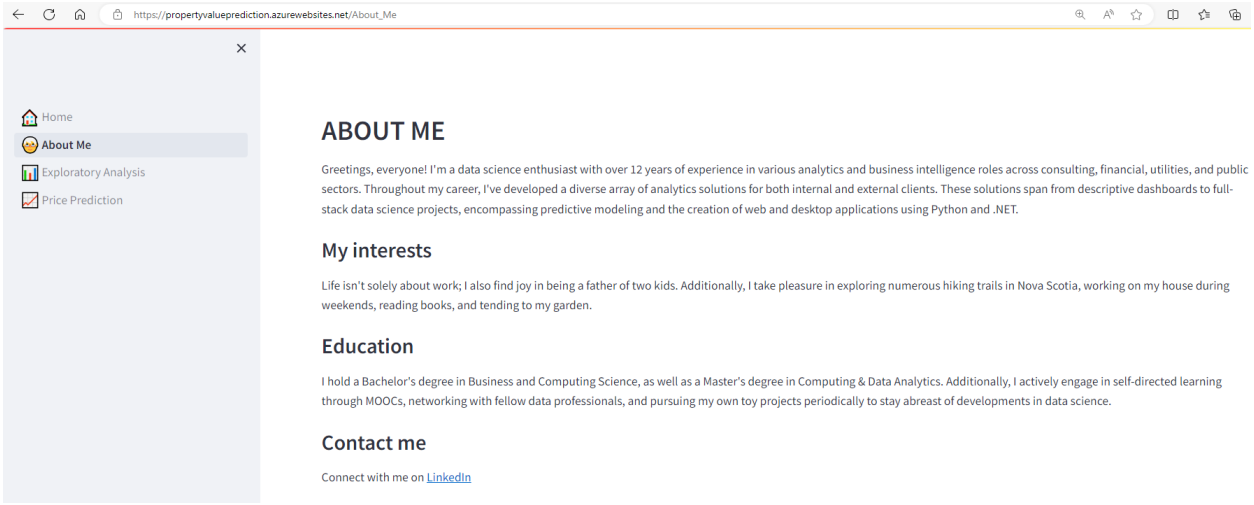

https://propertyvalueprediction.azurewebsites.net
Home
About Me
Exploratory Analysis
Price Prediction
ABOUT THE PROJECT
This project was developed over two extended weekends in the spring of 2024 to showcase a complete data science project lifecycle, encompassing exploratory data analysis (EDA), feature engineering, model creation, and deployment of the entire solution as a web application. To maintain conciseness, several steps in exploratory data analysis have been condensed or omitted.
Feature engineering encompassed the creation of additional categorical geospatial features, such as zip code groups, which have been utilized to improve data comprehension and enhance the accuracy of inferences generated by linear models. Predictions are made by the three best models selected from the PyCaret run, with only minimal model tuning performed due to constraints on compute resources.
This dataset consists of historical housing prices obtained from Zillow, as referenced in the book 'Practical Statistics for Data Scientists' by Peter Bruce, Andrew Bruce, and Peter Gedeck. It's essential to emphasize that this dataset does not reflect current market conditions; rather, it serves solely for demonstration purposes.
At no time should you depend on any predictions generated by this app.
This project has used the following stack: Python, PyCaret, Docker, Statsmodels, Plotly.
Link to project GitHub
PYCARET
Streamlit
python
docker


https://propertyvalueprediction.azurewebsites.net/About_Me
Home
About Me
Exploratory Analysis
Price Prediction
ABOUT ME
Greetings, everyone! I'm a data science enthusiast with over 12 years of experience in various analytics and business intelligence roles across consulting, financial, utilities, and public sectors. Throughout my career, I've developed a diverse array of analytics solutions for both internal and external clients. These solutions span from descriptive dashboards to full-stack data science projects, encompassing predictive modeling and the creation of web and desktop applications using Python and .NET.
My interests
Life isn't solely about work; I also find joy in being a father of two kids. Additionally, I take pleasure in exploring numerous hiking trails in Nova Scotia, working on my house during weekends, reading books, and tending to my garden.
Education
I hold a Bachelor's degree in Business and Computing Science, as well as a Master's degree in Computing & Data Analytics. Additionally, I actively engage in self-directed learning through MOOCs, networking with fellow data professionals, and pursuing my own toy projects periodically to stay abreast of developments in data science.
Contact me
Connect with me on LinkedIn

https://propertyvalueprediction.azurewebsites.net/Exploratory_Analysis

Home
About Me
Exploratory Analysis
Price Prediction

Exploratory data analysis

To select and engineer relevant features for the sales price prediction model, I've conducted an extensive exploratory data analysis (EDA) to identify the most important data attributes, interrelations between variables, and anomalies. This EDA phase encompassed fitting a Linear Regression model to illustrate the significance of various variables and their impact on the outcome variable. Additionally, the Linear Regression model acts as a reliable benchmark for the more advanced models employed in predicting property sales prices within the Price Prediction section.

Below, you'll find some of the results from this EDA grouped by topic. Click on each section to view the details.

- GEOGRAPHY BY ZIP CODE
- OUTLIERS
- SALES PRICES
- CUMULATIVE DENSITY
- FEATURE CORRELATIONS
- RESIDUAL ANALYSIS
- SALES TRENDS
- OTHER FEATURE RELATIONSHIPS
- GEOSPACIAL FEATURE ENGINEERING
- LINEAR REGRESSION ANALYSIS
- CONCLUSION AND NEXT STEPS

https://propertyvalueprediction.azurewebsites.net/Exploratory_Analysis

Home
About Me
Exploratory Analysis
Price Prediction

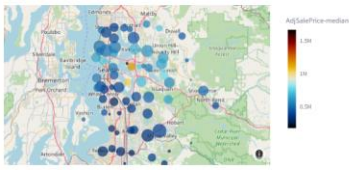
Exploratory data analysis

To select and engineer relevant features for the sales price prediction model, I've conducted an extensive exploratory data analysis (EDA) to identify the most important data attributes, interrelations between variables, and anomalies. This EDA phase encompassed fitting a Linear Regression model to illustrate the significance of various variables and their impact on the outcome variable. Additionally, the Linear Regression model acts as a reliable benchmark for the more advanced models employed in predicting property sales prices within the Price Prediction section.

Below, you'll find some of the results from this EDA grouped by topic. Click on each section to view the details.

GEOGRAPHY BY ZIP CODE

The dataset comprises 80 zip codes, with the two priciest ones situated in Mercer Island and Bellevue. In these areas, the median property prices were up to three times higher than those outside the city core. Conversely, less expensive zip codes are found outside the city core, where the majority of sales and purchase transactions occurred. This suggests that location and other geographical characteristics could serve as strong predictors of housing prices. The exploration of location features will be conducted later as part of the exploratory data analysis (EDA) process.



https://propertyvalueprediction.azurewebsites.net/Exploratory_Analysis

Home
About Me
Exploratory Analysis
Price Prediction

Exploratory data analysis

To select and engineer relevant features for the sales price prediction model, I've conducted an extensive exploratory data analysis (EDA) to identify the most important data attributes, interrelations between variables, and anomalies. This EDA phase encompassed fitting a Linear Regression model to illustrate the significance of various variables and their impact on the outcome variable. Additionally, the Linear Regression model acts as a reliable benchmark for the more advanced models employed in predicting property sales prices within the Price Prediction section.

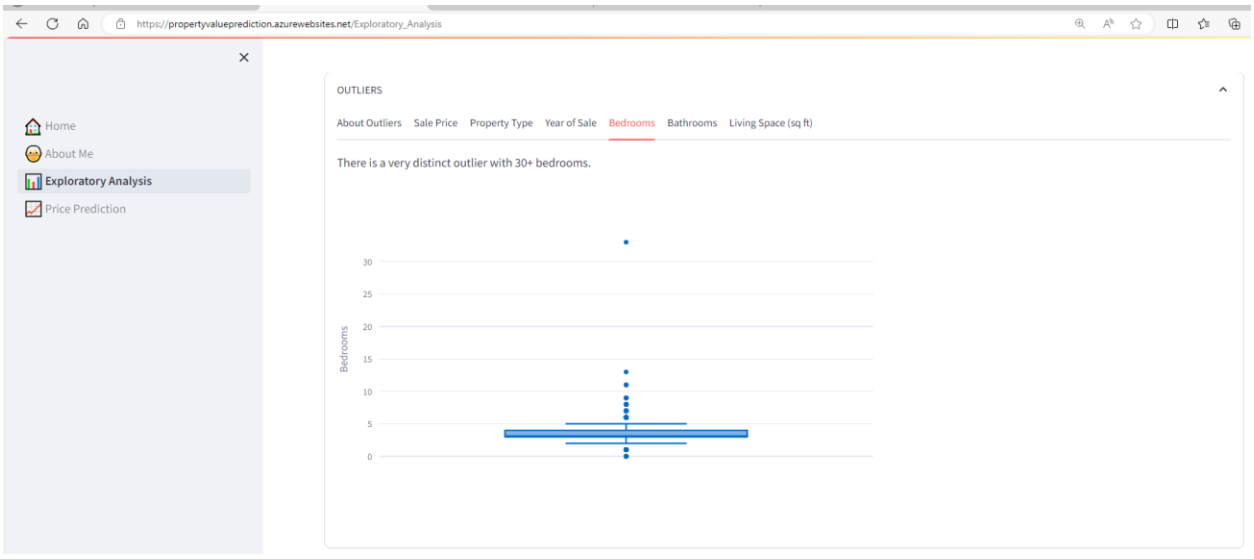
Below, you'll find some of the results from this EDA grouped by topic. Click on each section to view the details.

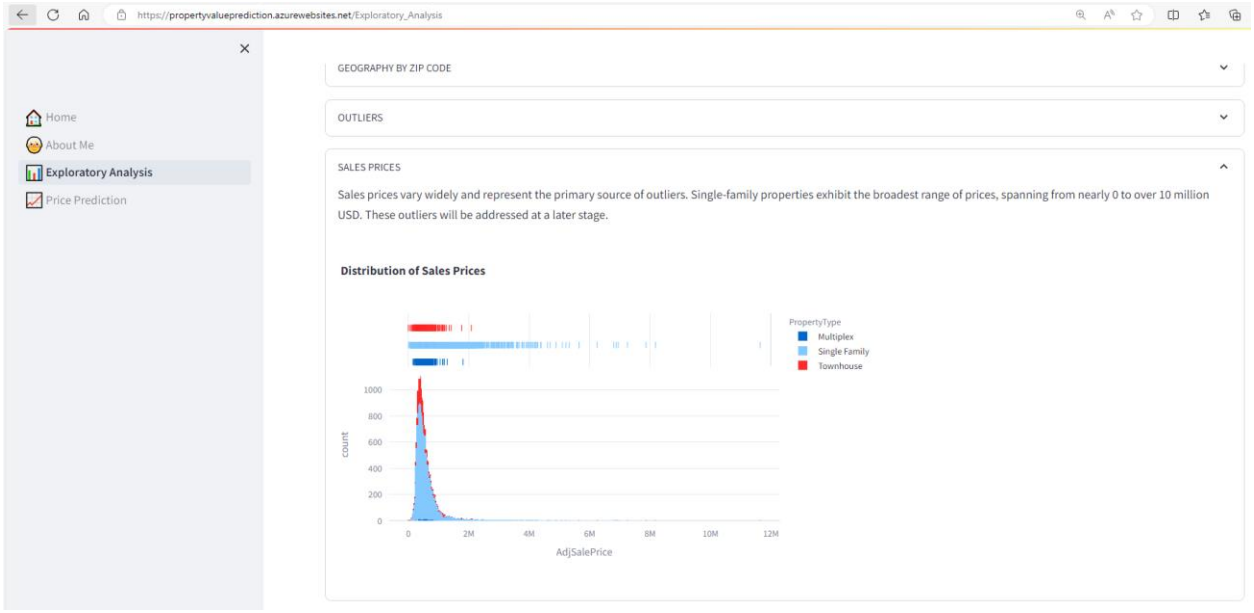
- GEOGRAPHY BY ZIP CODE

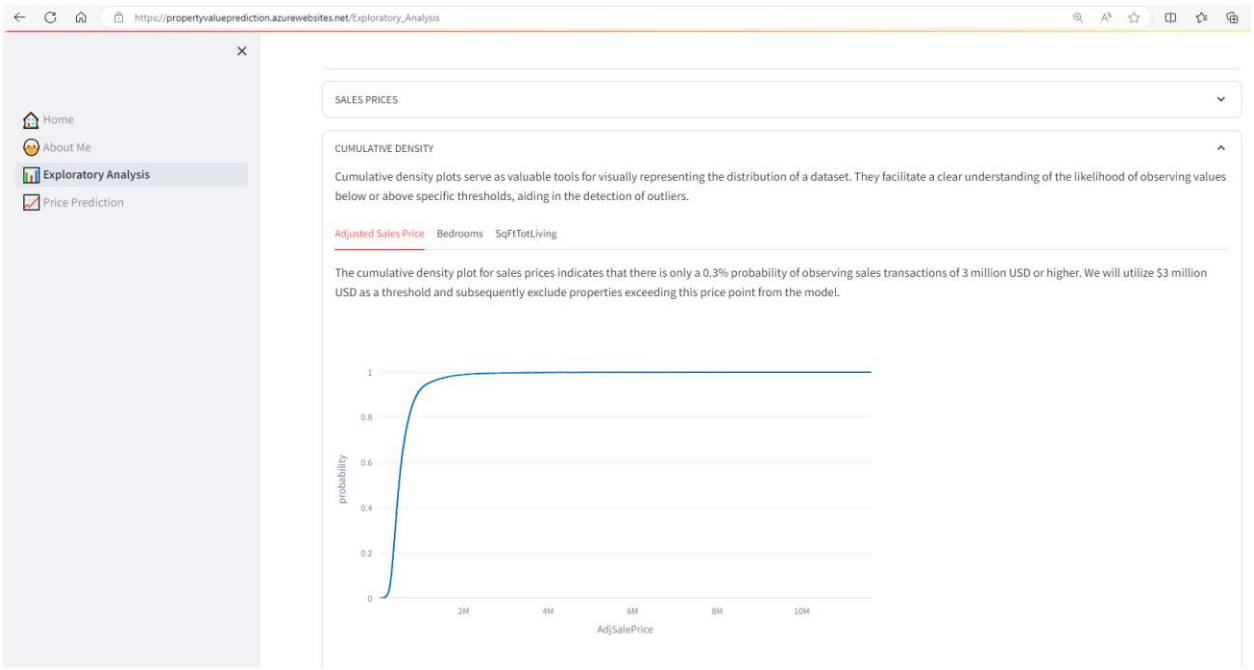
OUTLIERS

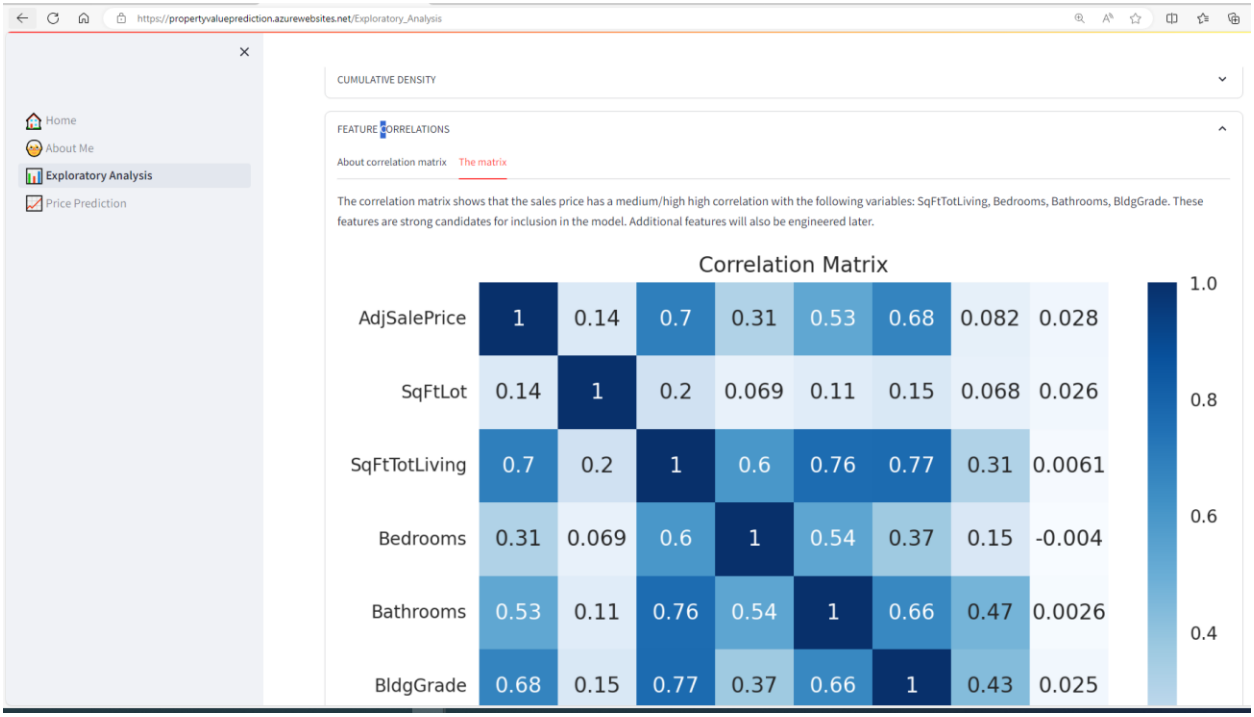
About Outliers | Sale Price | Property Type | Year of Sale | Bedrooms | Bathrooms | Living Space (sq ft)

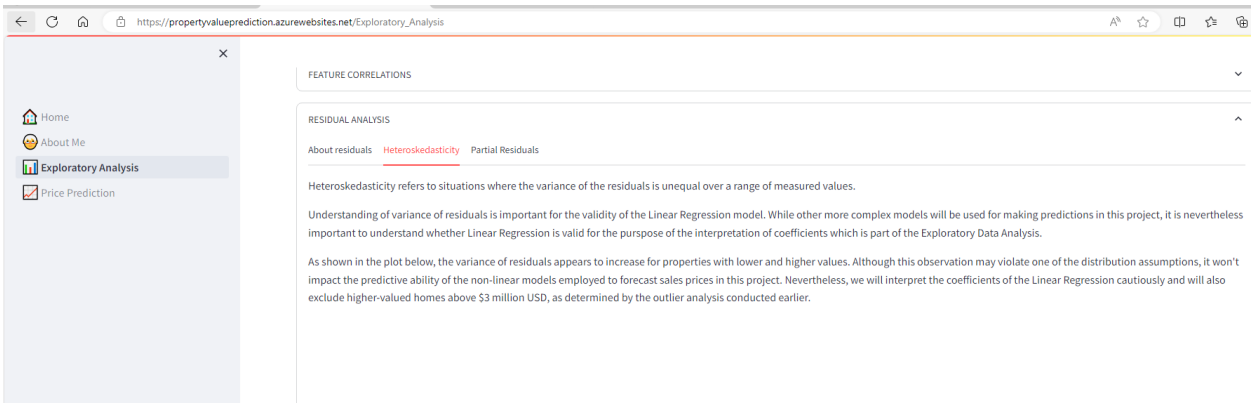
There seem to be outlier properties with exceptionally high and low sales prices. In the upcoming tabs, we'll delve further into these outliers and contemplate excluding them later if it will enhance our model.

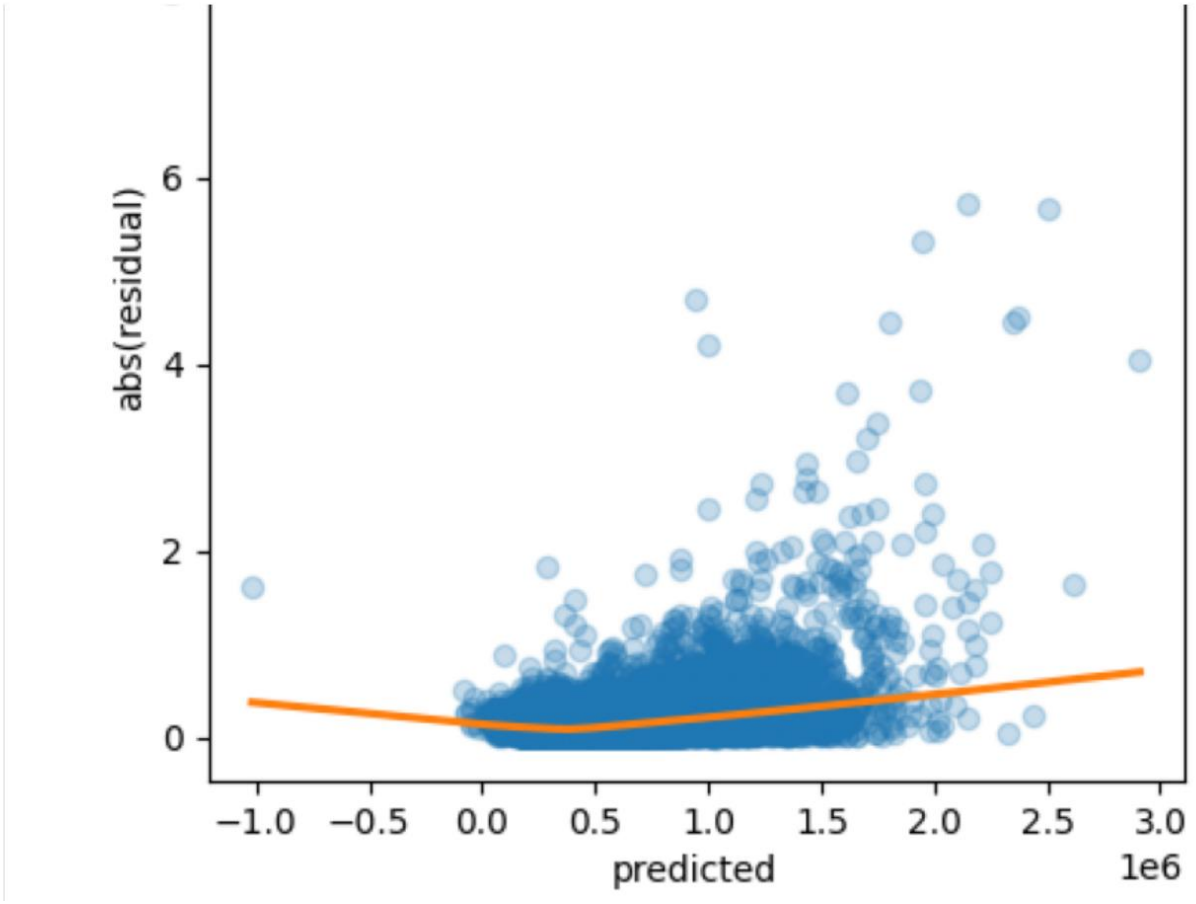

https://propertyvalueprediction.azurewebsites.net/Exploratory_Analysis
Home
About Me
Exploratory Analysis
Price Prediction
OUTLIERS
About Outliers
Sale Price
Property Type
Year of Sale
Bedrooms
Bathrooms
Living Space (sq ft)
There is a very distinct outlier with 30+ bedrooms.
Bedrooms
30
25
20
15
10
5
0


https://propertyvalueprediction.azurewebsites.net/Exploratory_Analysis
Home
About Me
Exploratory Analysis
Price Prediction
GEOGRAPHY BY ZIP CODE
OUTLIERS
SALES PRICES
Sales prices vary widely and represent the primary source of outliers. Single-family properties exhibit the broadest range of prices, spanning from nearly 0 to over 10 million USD. These outliers will be addressed at a later stage.
Distribution of Sales Prices
PropertyType
Multiplex
Single Family
Townhouse
count
1000
800
600
400
200
0
0
2M
4M
6M
8M
10M
12M
AdjSalePrice


https://propertyvalueprediction.azurewebsites.net/Exploratory_Analysis
Home
About Me
Exploratory Analysis
Price Prediction
SALES PRICES
CUMULATIVE DENSITY
Cumulative density plots serve as valuable tools for visually representing the distribution of a dataset. They facilitate a clear understanding of the likelihood of observing values below or above specific thresholds, aiding in the detection of outliers.
Adjusted Sales Price
Bedrooms
SqFtTotLiving
The cumulative density plot for sales prices indicates that there is only a 0.3% probability of observing sales transactions of 3 million USD or higher. We will utilize $3 million USD as a threshold and subsequently exclude properties exceeding this price point from the model.
probability
AdjSalePrice


https://propertyvalueprediction.azurewebsites.net/Exploratory_Analysis
Home
About Me
Exploratory Analysis
Price Prediction
CUMULATIVE DENSITY
FEATURE CORRELATIONS
About correlation matrix
The matrix
The correlation matrix shows that the sales price has a medium/high high correlation with the following variables: SqFtTotLiving, Bedrooms, Bathrooms, BldgGrade. These features are strong candidates for inclusion in the model. Additional features will also be engineered later.
Correlation Matrix
AdjSalePrice 1 0.14 0.7 0.31 0.53 0.68 0.082 0.028
SqFtLot 0.14 1 0.2 0.069 0.11 0.15 0.068 0.026
SqFtTotLiving 0.7 0.2 1 0.6 0.76 0.77 0.31 0.0061
Bedrooms 0.31 0.069 0.6 1 0.54 0.37 0.15 -0.004
Bathrooms 0.53 0.11 0.76 0.54 1 0.66 0.47 0.0026
BldgGrade 0.68 0.15 0.77 0.37 0.66 1 0.43 0.025
1.0
0.8
0.6
0.4


https://propertyvalueprediction.azurewebsites.net/Exploratory_Analysis
Home
About Me
Exploratory Analysis
Price Prediction
FEATURE CORRELATIONS
RESIDUAL ANALYSIS
About residuals
Heteroskedasticity
Partial Residuals
Heteroskedasticity refers to situations where the variance of the residuals is unequal over a range of measured values.
Understanding of variance of residuals is important for the validity of the Linear Regression model. While other more complex models will be used for making predictions in this project, it is nevertheless important to understand whether Linear Regression is valid for the purspose of the interpretation of coefficients which is part of the Exploratory Data Analysis.
As shown in the plot below, the variance of residuals appears to increase for properties with lower and higher values. Although this observation may violate one of the distribution assumptions, it won't impact the predictive ability of the non-linear models employed to forecast sales prices in this project. Nevertheless, we will interpret the coefficients of the Linear Regression cautiously and will also exclude higher-valued homes above $3 million USD, as determined by the outlier analysis conducted earlier.


abs(residual)
6
4
2
0
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
predicted
1e6

https://propertyvalueprediction.azurewebsites.net/Exploratory_Analysis

Home
About Me
Exploratory Analysis
Price Prediction

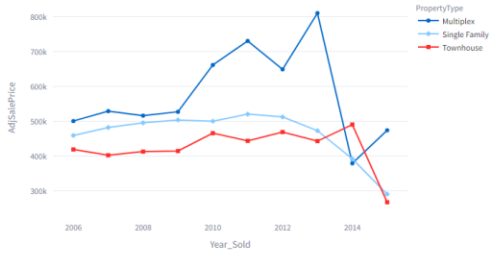
FEATURE CORRELATIONS

RESIDUAL ANALYSIS

SALES TRENDS

As seen in the plots below, the median sales price for all property types in the County decreased from its peak. However, it remains uncertain whether this decline reflects the true economic environment or stems from limitations within the dataset. Additionally, it's important to refrain from drawing inferences from any models in this project beyond the date ranges covered by the dataset.

Sales Price | Volume



https://propertyvalueprediction.azurewebsites.net/Exploratory_Analysis

Home
About Me
Exploratory Analysis
Price Prediction

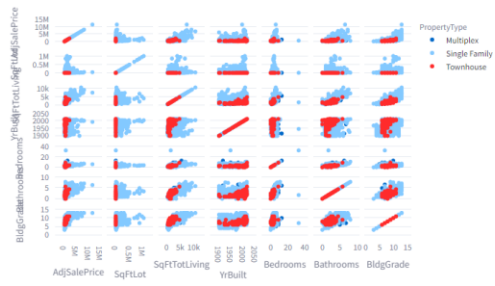
RESIDUAL ANALYSIS

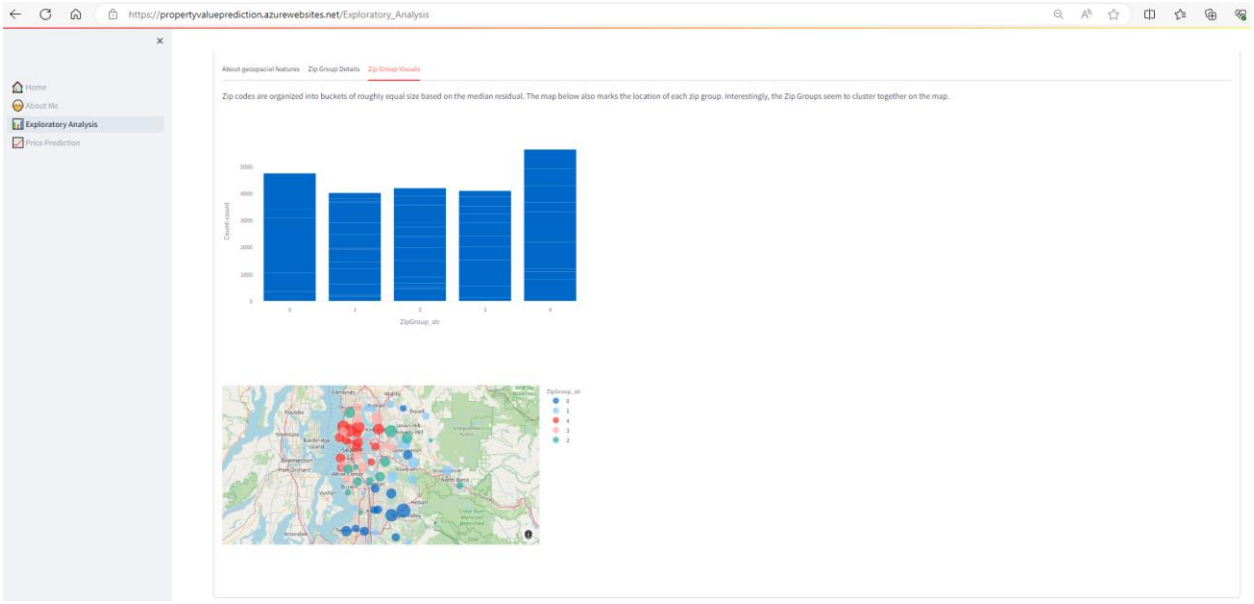
SALES TRENDS

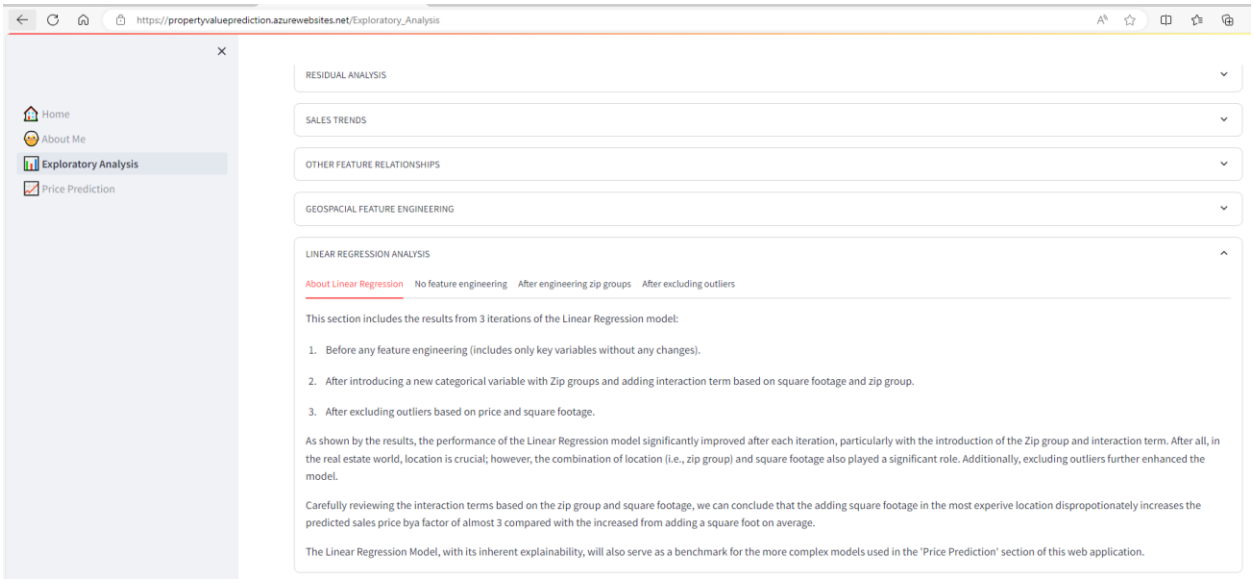
OTHER FEATURE RELATIONSHIPS

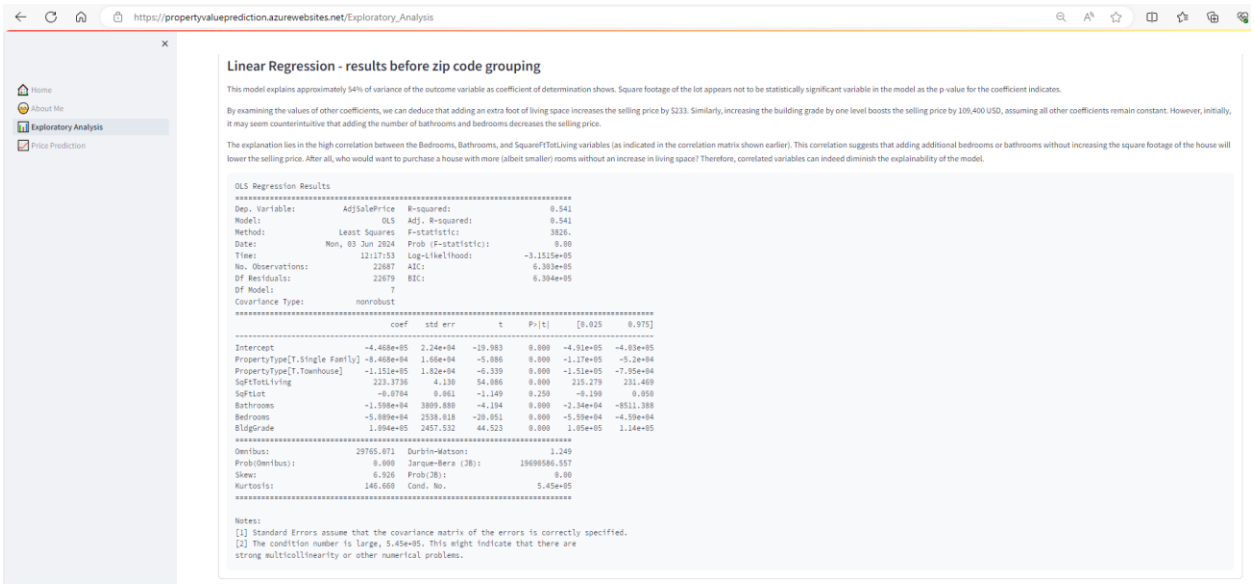
Before removing outliers | After removing outliers

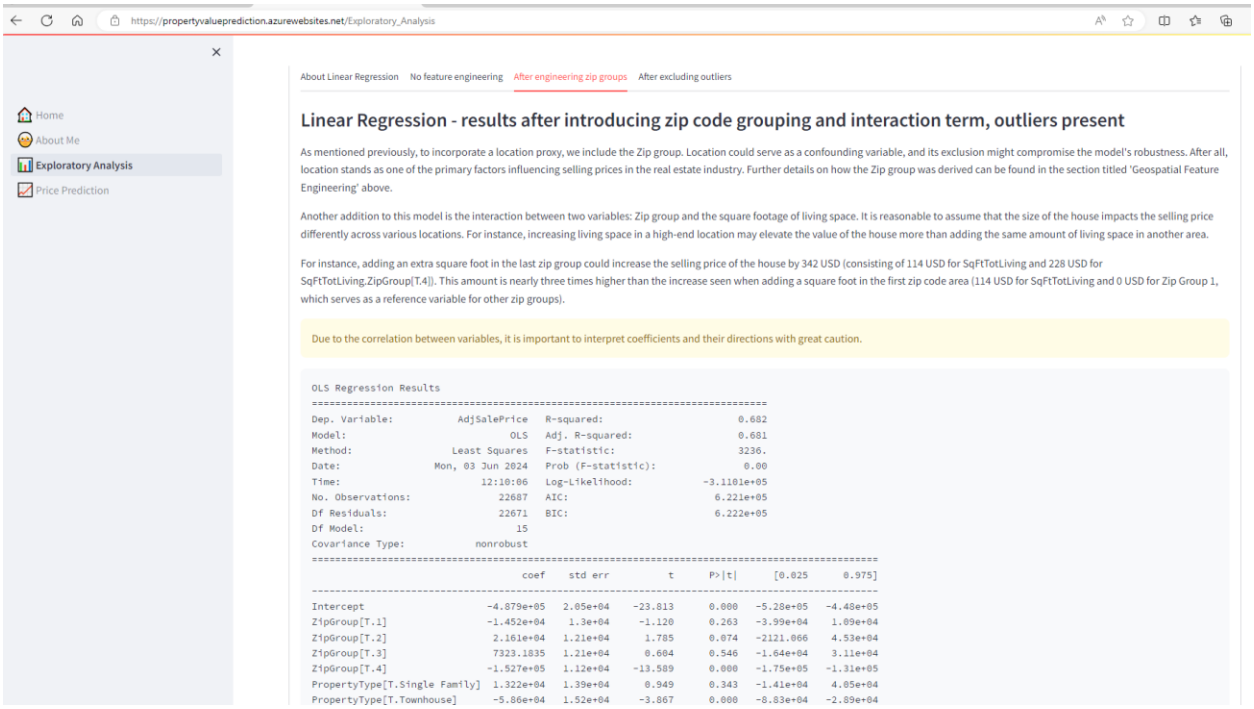
This pairplot represents the relationship between pairs of variables BEFORE OUTLIERS ARE REMOVED in the dataset and helps identify possible relationships between variables, and outliers in the data. It is clear that there are outliers in the data in terms of number of bedrooms, sales price and square footage. As noted earlier, these some of these outliers can affect robustness of our model. The diagonal of the plot shows the relationship of the variable with itself.




https://propertyvalueprediction.azurewebsites.net/Exploratory_Analysis
Home
About Me
Exploratory Analysis
Price Prediction
About geospacial features
Zip Group Details
Zip Group Visuals
Zip codes are organized into buckets of roughly equal size based on the median residual. The map below also marks the location of each zip group. Interestingly, the Zip Groups seem to cluster together on the map.
Count_count
ZipGroup_str


https://propertyvalueprediction.azurewebsites.net/Exploratory_Analysis
Home
About Me
Exploratory Analysis
Price Prediction
RESIDUAL ANALYSIS
SALES TRENDS
OTHER FEATURE RELATIONSHIPS
GEOSPACIAL FEATURE ENGINEERING
LINEAR REGRESSION ANALYSIS
About Linear Regression
No feature engineering
After engineering zip groups
After excluding outliers
This section includes the results from 3 iterations of the Linear Regression model:
1. Before any feature engineering (includes only key variables without any changes).
2. After introducing a new categorical variable with Zip groups and adding interaction term based on square footage and zip group.
3. After excluding outliers based on price and square footage.
As shown by the results, the performance of the Linear Regression model significantly improved after each iteration, particularly with the introduction of the Zip group and interaction term. After all, in the real estate world, location is crucial; however, the combination of location (i.e., zip group) and square footage also played a significant role. Additionally, excluding outliers further enhanced the model.
Carefully reviewing the interaction terms based on the zip group and square footage, we can conclude that the adding square footage in the most experive location dispropotionately increases the predicted sales price bya factor of almost 3 compared with the increased from adding a square foot on average.
The Linear Regression Model, with its inherent explainability, will also serve as a benchmark for the more complex models used in the 'Price Prediction' section of this web application.


https://propertyvalueprediction.azurewebsites.net/Exploratory_Analysis
Home
About Me
Exploratory Analysis
Price Prediction
Linear Regression - results before zip code grouping
This model explains approximately 54% of variance of the outcome variable as coefficient of determination shows. Square footage of the lot appears not to be statistically significant variable in the model as the p-value for the coefficient indicates.
By examining the values of other coefficients, we can deduce that adding an extra foot of living space increases the selling price by $233. Similarly, increasing the building grade by one level boosts the selling price by 109,400 USD, assuming all other coefficients remain constant. However, initially, it may seem counterintuitive that adding the number of bathrooms and bedrooms decreases the selling price.
The explanation lies in the high correlation between the Bedrooms, Bathrooms, and SquareFtTotLiving variables (as indicated in the correlation matrix shown earlier). This correlation suggests that adding additional bedrooms or bathrooms without increasing the square footage of the house will lower the selling price. After all, who would want to purchase a house with more (albeit smaller) rooms without an increase in living space? Therefore, correlated variables can indeed diminish the explainability of the model.
OLS Regression Results
Dep. Variable: AdjSalePrice R-squared: 0.541
Model: OLS Adj. R-squared: 0.541
Method: Least Squares F-statistic: 3826.
Date: Mon, 03 Jun 2024 Prob (F-statistic): 0.00
Time: 12:17:53 Log-Likelihood: -3.1515e+05
No. Observations: 22687 AIC: 6.303e+05
Df Residuals: 22679 BIC: 6.304e+05
Df Model: 7
Covariance Type: nonrobust
coef std err t P>|t| [0.025 0.975]
Intercept -4.468e+05 2.24e+04 -19.983 0.000 -4.91e+05 -4.03e+05
PropertyType[T.Single Family] -8.468e+04 1.66e+04 -5.086 0.000 -1.17e+05 -5.2e+04
PropertyType[T.Townhouse] -1.151e+05 1.82e+04 -6.339 0.000 -1.51e+05 -7.95e+04
SqFtTotLiving 223.3736 4.130 54.086 0.000 215.279 231.469
SqFtLot -0.0704 0.061 -1.149 0.250 -0.190 0.050
Bathrooms -1.598e+04 3809.880 -4.194 0.000 -2.34e+04 -8511.388
Bedrooms -5.889e+04 2538.018 -20.051 0.000 -5.59e+04 -4.59e+04
BldgGrade 1.894e+05 2457.532 44.523 0.000 1.85e+05 1.14e+05
Omnibus: 29765.071 Durbin-Watson: 1.249
Prob(Omnibus): 0.000 Jarque-Bera (JB): 19690586.557
Skew: 6.926 Prob(JB): 0.00
Kurtosis: 146.660 Cond. No. 5.45e+05
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 5.45e+05. This might indicate that there are
strong multicollinearity or other numerical problems.


https://propertyvalueprediction.azurewebsites.net/Exploratory_Analysis
Home
About Me
Exploratory Analysis
Price Prediction
About Linear Regression
No feature engineering
After engineering zip groups
After excluding outliers
Linear Regression - results after introducing zip code grouping and interaction term, outliers present
As mentioned previously, to incorporate a location proxy, we include the Zip group. Location could serve as a confounding variable, and its exclusion might compromise the model's robustness. After all, location stands as one of the primary factors influencing selling prices in the real estate industry. Further details on how the Zip group was derived can be found in the section titled 'Geospatial Feature Engineering' above.
Another addition to this model is the interaction between two variables: Zip group and the square footage of living space. It is reasonable to assume that the size of the house impacts the selling price differently across various locations. For instance, increasing living space in a high-end location may elevate the value of the house more than adding the same amount of living space in another area.
For instance, adding an extra square foot in the last zip group could increase the selling price of the house by 342 USD (consisting of 114 USD for SqFtTotLiving and 228 USD for SqFtTotLiving.ZipGroup[T.4]). This amount is nearly three times higher than the increase seen when adding a square foot in the first zip code area (114 USD for SqFtTotLiving and 0 USD for Zip Group 1, which serves as a reference variable for other zip groups).
Due to the correlation between variables, it is important to interpret coefficients and their directions with great caution.
OLS Regression Results
Dep. Variable: AdjSalePrice R-squared: 0.682
Model: OLS Adj. R-squared: 0.681
Method: Least Squares F-statistic: 3236.
Date: Mon, 03 Jun 2024 Prob (F-statistic): 0.00
Time: 12:10:06 Log-Likelihood: -3.1101e+05
No. Observations: 22687 AIC: 6.221e+05
Df Residuals: 22671 BIC: 6.222e+05
Df Model: 15
Covariance Type: nonrobust
coef std err t P>|t| [0.025 0.975]
Intercept -4.879e+05 2.05e+04 -23.813 0.000 -5.28e+05 -4.48e+05
ZipGroup[T.1] -1.452e+04 1.3e+04 -1.120 0.263 -3.99e+04 1.09e+04
ZipGroup[T.2] 2.161e+04 1.21e+04 1.785 0.074 -2121.066 4.53e+04
ZipGroup[T.3] 7323.1835 1.21e+04 0.604 0.546 -1.64e+04 3.11e+04
ZipGroup[T.4] -1.527e+05 1.12e+04 -13.589 0.000 -1.75e+05 -1.31e+05
PropertyType[T.Single Family] 1.322e+04 1.39e+04 0.949 0.343 -1.41e+04 4.05e+04
PropertyType[T.Townhouse] -5.86e+04 1.52e+04 -3.867 0.000 -8.83e+04 -2.89e+04

https://propertyvalueprediction.azurewebsites.net/Exploratory_Analysis

Home
About Me
Exploratory Analysis
Price Prediction

Linear Regression - results after introducing zip code grouping and interaction term, outliers present

As mentioned previously, to incorporate a location proxy, we include the Zip group. Location could serve as a confounding variable, and its exclusion might compromise the model's robustness. After all, location stands as one of the primary factors influencing selling prices in the real estate industry. Further details on how the Zip group was derived can be found in the section titled 'Geospatial Feature Engineering' above.

Another addition to this model is the interaction between two variables: Zip group and the square footage of living space. It is reasonable to assume that the size of the house impacts the selling price differently across various locations. For instance, increasing living space in a high-end location may elevate the value of the house more than adding the same amount of living space in another area.

For instance, adding an extra square foot in the last zip group could increase the selling price of the house by 342 USD (consisting of 114 USD for SqFtTotLiving and 228 USD for SqFtTotLiving:ZipGroup[T.4]). This amount is nearly three times higher than the increase seen when adding a square foot in the first zip code area (114 USD for SqFtTotLiving and 0 USD for Zip Group 1, which serves as a reference variable for other zip groups).

Due to the correlation between variables, it is important to interpret coefficients and their directions with great caution.

OLS Regression Results

Dep. Variable:	AdjSalePrice	R-squared:	0.682
Model:	OLS	Adj. R-squared:	0.681
Method:	Least Squares	F-statistic:	3236.
Date:	Mon, 03 Jun 2024	Prob (F-statistic):	0.00
Time:	12:17:53	Log-Likelihood:	-3.1101e+05
No. Observations:	22687	AIC:	6.221e+05
Df Residuals:	22671	BIC:	6.222e+05
Df Model:	15		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-4.879e+05	2.05e+04	-23.813	0.000	-5.28e+05	-4.48e+05
ZipGroup[T.1]	-1.452e+04	1.3e+04	-1.120	0.263	-3.99e+04	1.09e+04
ZipGroup[T.2]	2.161e+04	1.21e+04	1.785	0.074	-2121.066	4.53e+04
ZipGroup[T.3]	7323.1835	1.21e+04	0.604	0.546	-1.64e+04	3.11e+04
ZipGroup[T.4]	-1.527e+05	1.12e+04	-13.589	0.000	-1.75e+05	-1.31e+05
PropertyType[T.Single Family]	1.322e+04	1.39e+04	0.949	0.343	-1.41e+04	4.05e+04
PropertyType[T.Townhouse]	-5.86e+04	1.52e+04	-3.867	0.000	-8.83e+04	-2.89e+04
SqFtTotLiving	114.4857	4.867	23.521	0.000	104.945	124.026
SqFtTotLiving:ZipGroup[T.1]	36.4813	5.488	6.648	0.000	25.725	47.237
SqFtTotLiving:ZipGroup[T.2]	41.2295	5.364	7.686	0.000	30.716	51.743
SqFtTotLiving:ZipGroup[T.3]	76.1137	5.584	13.630	0.000	65.168	87.059
SqFtTotLiving:ZipGroup[T.4]	228.1176	4.825	47.288	0.000	218.661	237.575
SqFtLot	0.7172	0.052	13.859	0.000	0.616	0.819
Bathrooms	-5118.4136	3205.231	-1.597	0.110	-1.14e+04	1164.060
Bedrooms	-4.154e+04	2123.239	-19.566	0.000	-4.57e+04	-3.74e+04
BldgGrade	1.054e+05	2069.829	50.929	0.000	1.01e+05	1.09e+05

Omnibus:	30737.778	Durbin-Watson:	1.583
Prob(Omnibus):	0.000	Jarque-Bera (JB):	33537624.445
Skew:	7.193	Prob(JB):	0.00
Kurtosis:	190.807	Cond. No.	5.88e+05

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 5.8e+05. This might indicate that there are

https://propertyvalueprediction.azurewebsites.net/Exploratory_Analysis

Home
About Me
Exploratory Analysis
Price Prediction

Linear Regression - results after introducing zip code grouping, outliers excluded

Excluding outliers in the selling price and number of bedrooms, as discussed earlier, further improves the performance of the model. This iteration of the Linear Regression model now explains more than 72% of the variance in the selling price.

Due to the correlation between variables, it is important to interpret coefficients and their directions with great caution.

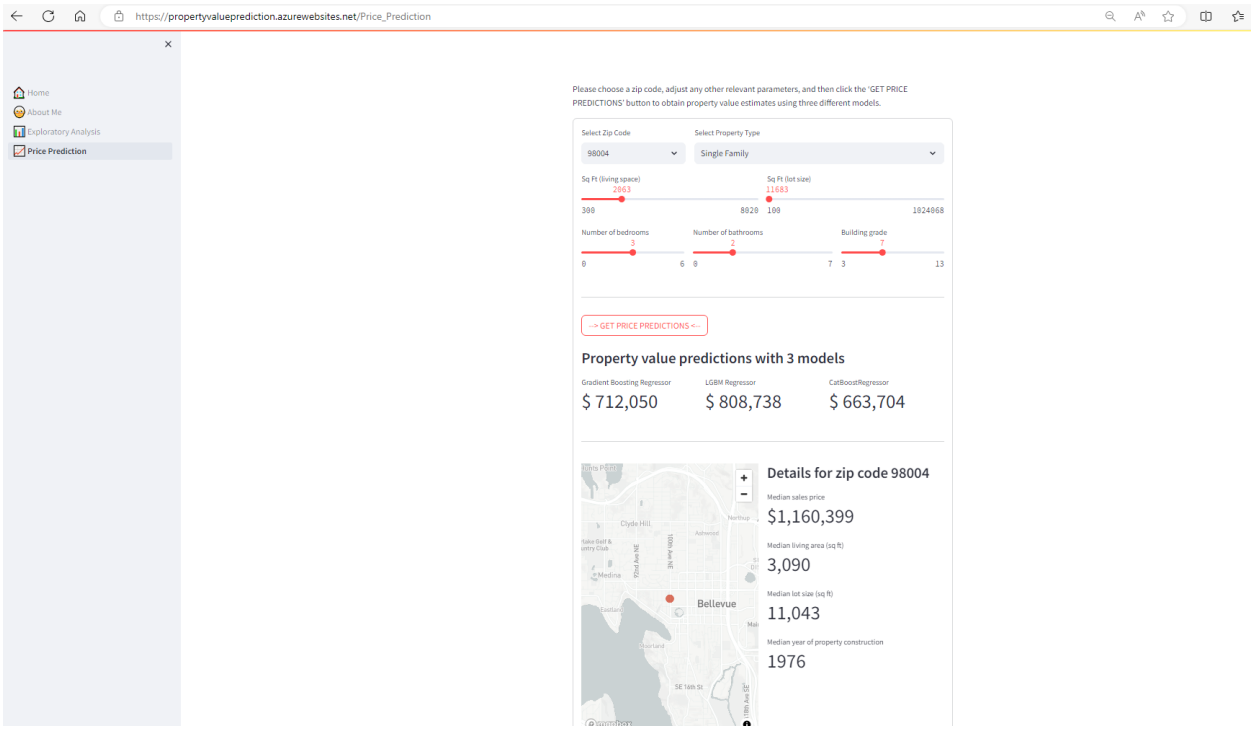
OLS Regression Results

Dep. Variable:	AdjSalePrice	R-squared:	0.722
Model:	OLS	Adj. R-squared:	0.722
Method:	Least Squares	F-statistic:	3985.
Date:	Mon, 03 Jun 2024	Prob (F-statistic):	0.00
Time:	12:17:53	Log-Likelihood:	-3.0303e+05
No. Observations:	22556	AIC:	6.061e+05
Df Residuals:	22540	BIC:	6.062e+05
Df Model:	15		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-4.813e+05	1.59e+04	-30.328	0.000	-5.12e+05	-4.5e+05
ZipGroup[T.1]	-8089.6155	9908.925	-0.808	0.419	-2.74e+04	1.14e+04
ZipGroup[T.2]	2.338e+04	9250.317	2.528	0.011	5250.922	4.15e+04
ZipGroup[T.3]	6.695e+04	9477.368	7.065	0.000	4.84e+04	8.55e+04
ZipGroup[T.4]	1.55e+04	8773.983	1.766	0.077	-1708.447	3.27e+04
PropertyType[T.Single Family]	3697.4086	1.08e+04	0.341	0.733	-1.75e+04	2.49e+04
PropertyType[T.Townhouse]	-8.931e+04	1.18e+04	-7.600	0.000	-1.12e+05	-6.63e+04
SqFtTotLiving	104.9431	3.742	28.045	0.000	97.609	112.278
SqFtTotLiving:ZipGroup[T.1]	34.2042	4.205	8.134	0.000	25.962	42.446
SqFtTotLiving:ZipGroup[T.2]	41.4991	4.108	10.102	0.000	33.447	49.551
SqFtTotLiving:ZipGroup[T.3]	43.3298	4.431	9.780	0.000	34.646	52.014
SqFtTotLiving:ZipGroup[T.4]	138.8485	3.817	36.377	0.000	131.367	146.330
SqFtLot	0.6241	0.040	15.794	0.000	0.547	0.702
Bathrooms	4987.1276	2470.292	2.019	0.044	145.185	9829.070
Bedrooms	-3.461e+04	1737.285	-19.932	0.000	-3.8e+04	-3.12e+04
BldgGrade	1.025e+05	1591.580	64.395	0.000	9.94e+04	1.06e+05

Omnibus:	9799.583	Durbin-Watson:	1.448
Prob(Omnibus):	0.000	Jarque-Bera (JB):	273497.892
Skew:	1.502	Prob(JB):	0.00
Kurtosis:	19.792	Cond. No.	5.88e+05

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 5.88e+05. This might indicate that there are
strong multicollinearity or other numerical problems.


https://propertyvalueprediction.azurewebsites.net/Price_Prediction
Home
About Me
Exploratory Analysis
Price Prediction
Please choose a zip code, adjust any other relevant parameters, and then click the 'GET PRICE PREDICTIONS' button to obtain property value estimates using three different models.
Select Zip Code
98004
Select Property Type
Single Family
Sq Ft (living space)
2663
300
8020
Sq Ft (lot size)
11683
100
1024068
Number of bedrooms
3
0
6
Number of bathrooms
2
0
7
Building grade
7
3
13
--> GET PRICE PREDICTIONS <--
Property value predictions with 3 models
Gradient Boosting Regressor
$ 712,050
LGBM Regressor
$ 808,738
CatBoostRegressor
$ 663,704
Bellevue
Details for zip code 98004
Median sales price
$1,160,399
Median living area (sq ft)
3,090
Median lot size (sq ft)
11,043
Median year of property construction
1976