Home-flipping in King County, WA

Isaac Barrera,

Evan Johnson,

Michael Lee

Agenda

- Stakeholders
- Data & Methods
- Findings
- Conclusion
- Future Research

Business Understanding

Stakeholder - house-flipping company in the King County area

Goal for home-flippers ⟶ cookie-cutter model

The Key Qualities

- Location
- Construction Quality
- Amount of Space

Data & Methods

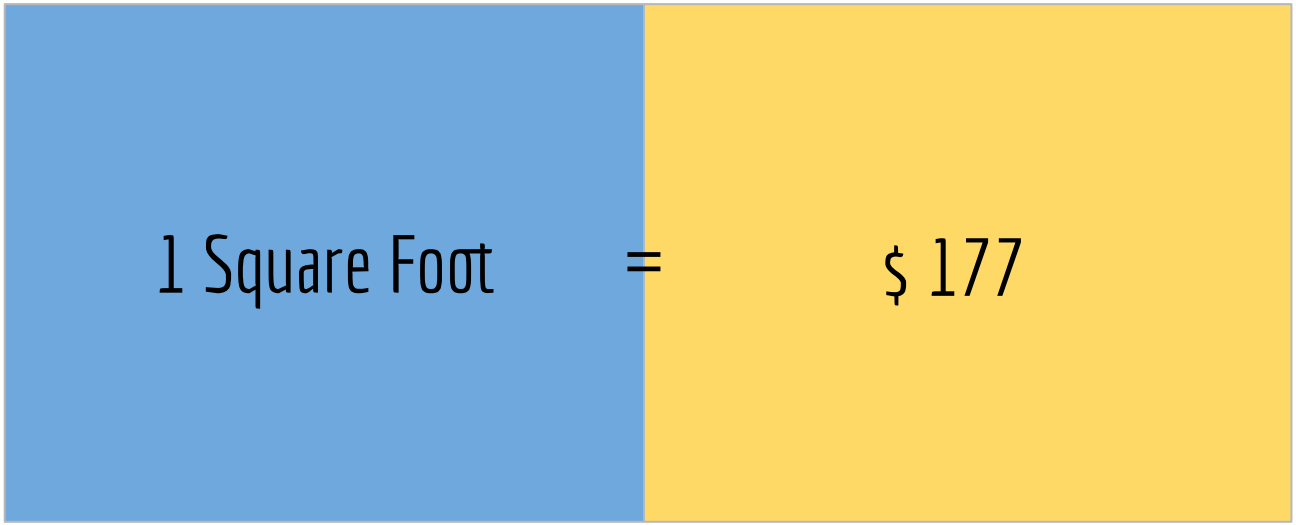
Data

Houses sold in King County between 2014 and 2015

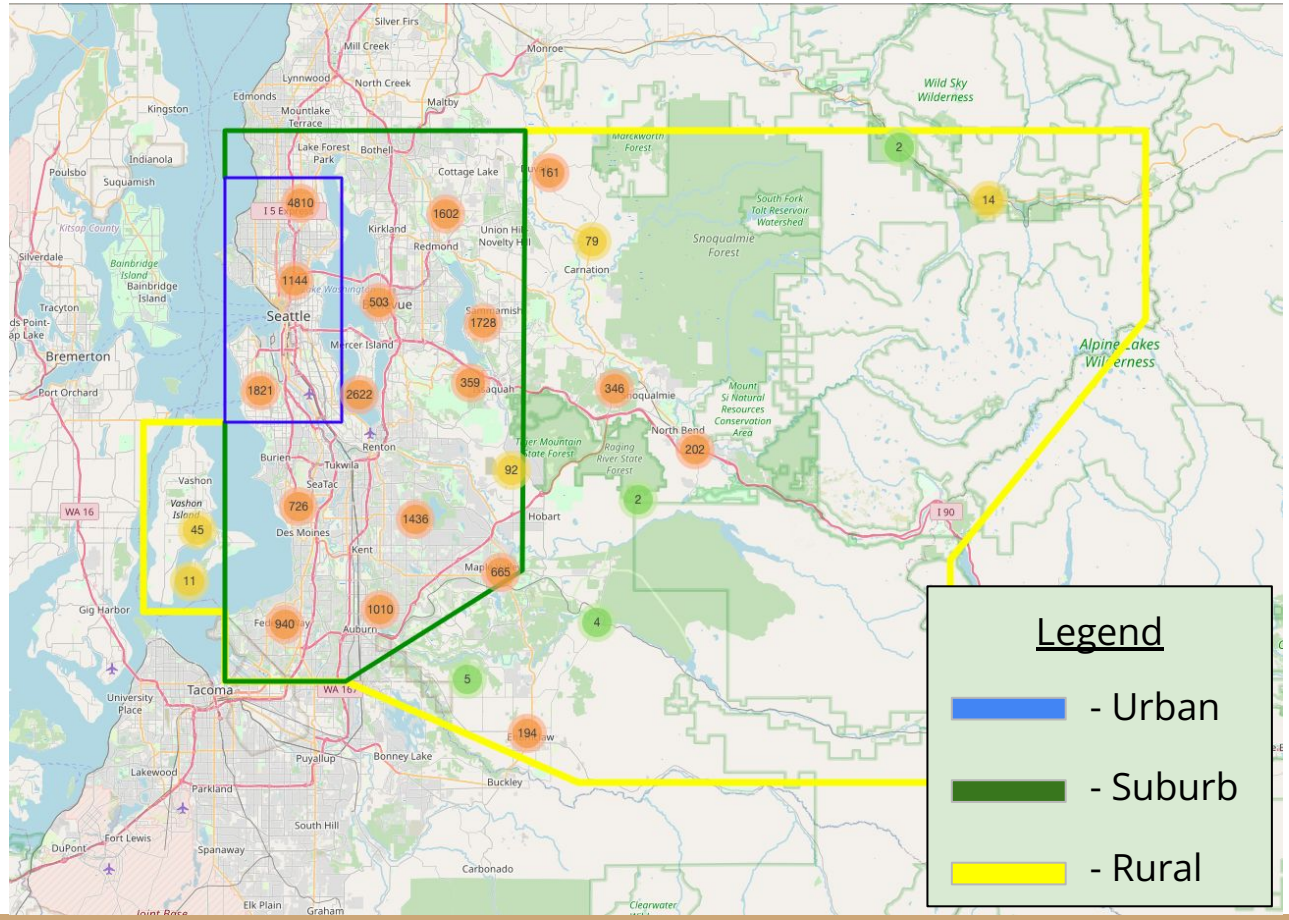
Engineered Features based on Cookie Cutter Model

Findings

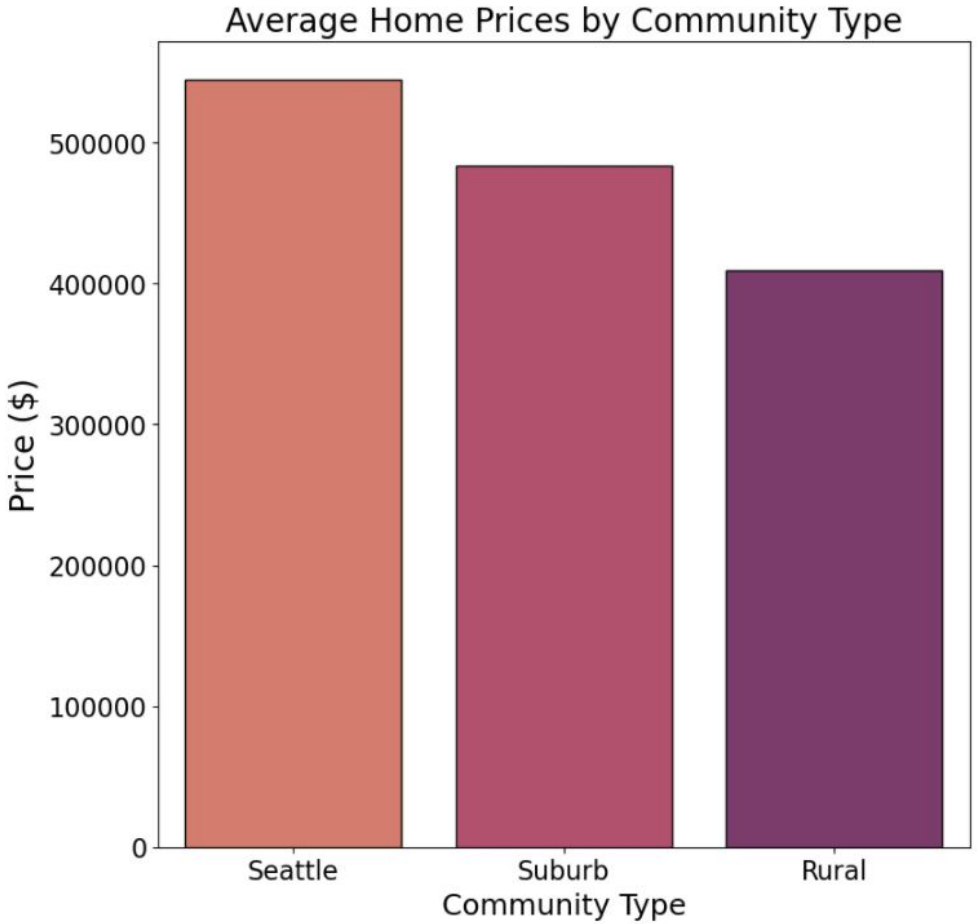
Amount of space ⟶ Livable Space showed most impact on home price



Location → Urban, Suburban, and Rural



Community Type is a Significant Input for Home Price Variability



Conclusion

Location is an important factor in determining project budget

Location Features	Average Impact to Home Price
Suburb	-$155,636
Rural	-$227,240

Focus on livable space and construction quality

Renovation Features	Average Impact to Home Price
Livable Space	$177 ↑
Construction Grade	$71,959 ↑

Future Research

Possible Research in the future

- Increase timeframe of dataset
- Explore other relative neighborhood metrics

Thank you!

Isaac Barrera

IB-7

ibarr24@gmail.com

Evan Johnson

Ewjohn127

ewjohn24@gmail.com

Michael Lee

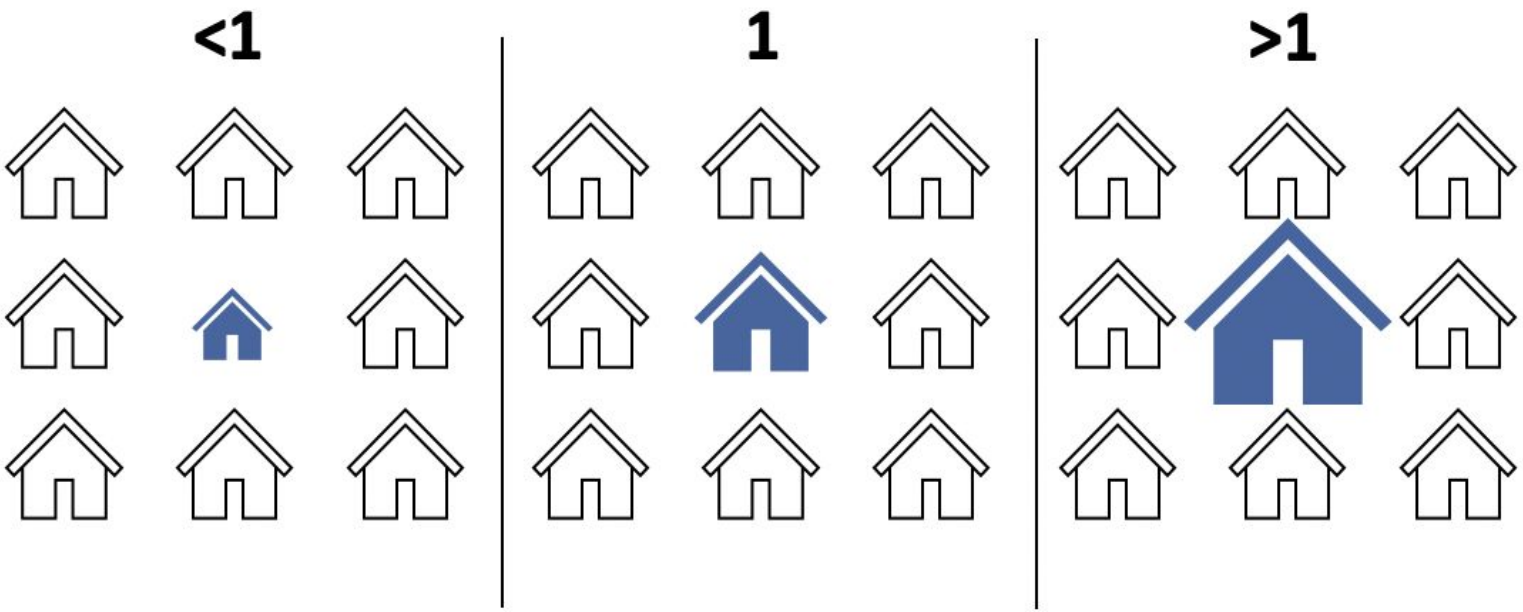
Baekho4245

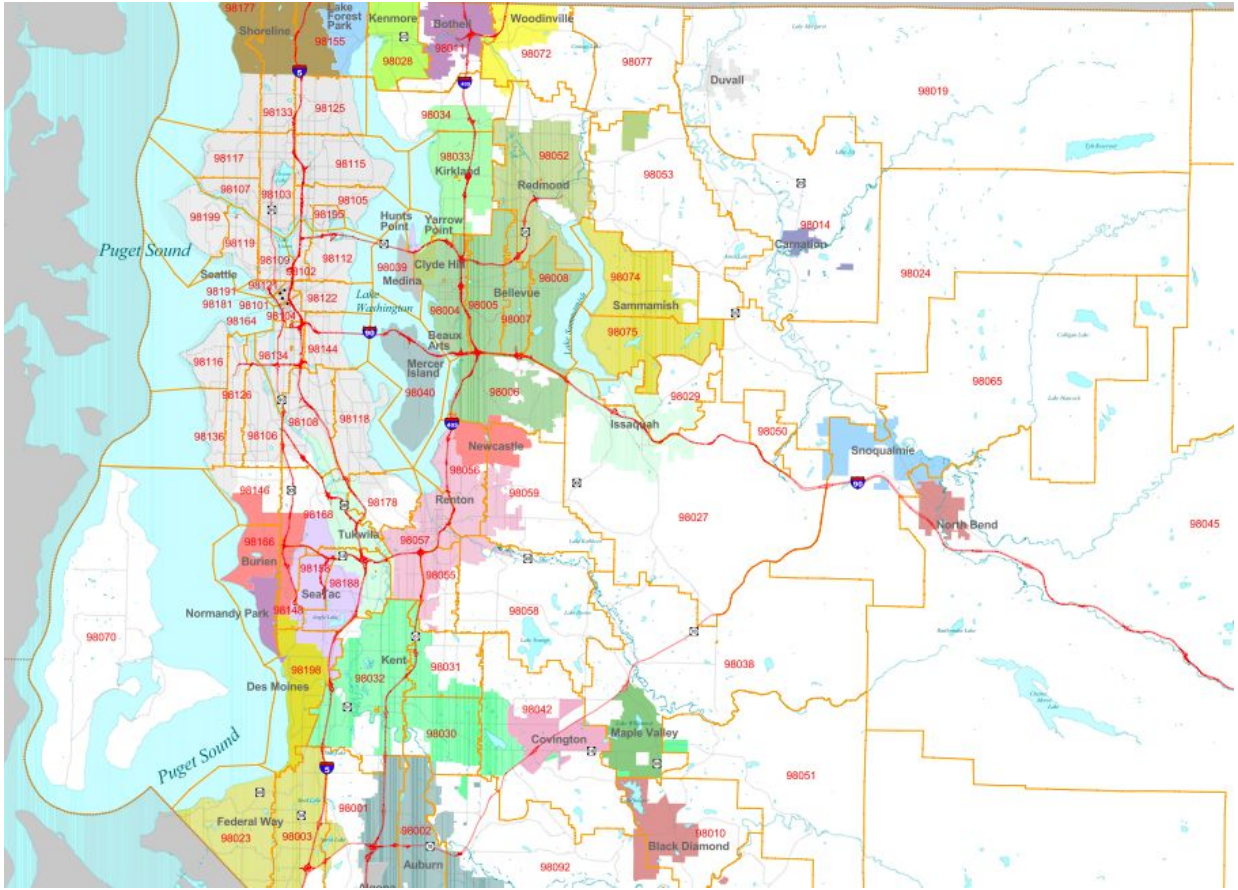
baekho5767@gmail.com

Appendix

Feature Engineering

Relative Living Area




Puget Sound
Puget Sound
Lake Washington
Shoreline
Lake Forest Park
Kenmore
Bothell
Woodinville
Duvall
Kirkland
Redmond
Hunts Point
Yarrow Point
Clyde Hill
Medina
Bellevue
Beaux Arts
Mercer Island
Seattle
Issaquah
Sammamish
Carnation
Snoqualmie
North Bend
Newcastle
Renton
Tukwila
Burien
SeaTac
Normandy Park
Des Moines
Kent
Federal Way
Auburn
Covington
Maple Valley
Black Diamond
98177
98155
98028
98011
98072
98077
98019
98133
98125
98034
98117
98115
98033
98052
98053
98107
98103
98105
98199
98195
98014
98119
98109
98112
98039
98008
98074
98024
98102
98191
98122
98005
98181
98101
98004
98007
98164
98075
98144
98116
98134
98065
98126
98040
98006
98029
98108
98118
98136
98106
98050
98146
98056
98059
98178
98168
98027
98045
98186
98057
98195
98055
98188
98148
98058
98070
98198
98032
98031
98038
98042
98030
98051
98091
98023
98003
98002
98010
98092

Model Results

Train Dataset OLS Regression Results

OLS Regression Results

Dep. Variable:	price	R-squared:	0.607
Model:	OLS	Adj. R-squared:	0.607
Method:	Least Squares	F-statistic:	4220.
Date:	Wed, 06 Oct 2021	Prob (F-statistic):	0.00
Time:	11:35:26	Log-Likelihood:	-2.1990e+05
No. Observations:	16424	AIC:	4.398e+05
Df Residuals:	16417	BIC:	4.399e+05
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	4.991e+05	1232.572	404.939	0.000	4.97e+05	5.02e+05
sqft_living	1.369e+05	2301.379	59.473	0.000	1.32e+05	1.41e+05
view	3.073e+04	1277.267	24.056	0.000	2.82e+04	3.32e+04
grade	7.811e+04	1883.941	41.460	0.000	7.44e+04	8.18e+04
relative_living_area	-3.705e+04	1590.031	-23.304	0.000	-4.02e+04	-3.39e+04
suburb	-7.642e+04	1445.739	-52.861	0.000	-7.93e+04	-7.36e+04
rural	-7.278e+04	1406.682	-51.739	0.000	-7.55e+04	-7e+04

Omnibus:	3288.957	Durbin-Watson:	2.022
Prob(Omnibus):	0.000	Jarque-Bera (JB):	10142.078
Skew:	1.031	Prob(JB):	0.00
Kurtosis:	6.250	Cond. No.	3.52

Test Dataset OLS Regression Results

OLS Regression Results

Dep. Variable:	price	R-squared:	0.582
Model:	OLS	Adj. R-squared:	0.581
Method:	Least Squares	F-statistic:	949.6
Date:	Wed, 06 Oct 2021	Prob (F-statistic):	0.00
Time:	11:35:26	Log-Likelihood:	-54949.
No. Observations:	4106	AIC:	1.099e+05
Df Residuals:	4099	BIC:	1.100e+05
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	4.958e+05	2452.368	202.160	0.000	4.91e+05	5.01e+05
sqft_living	1.315e+05	4539.602	28.971	0.000	1.23e+05	1.4e+05
view	3.012e+04	2525.386	11.926	0.000	2.52e+04	3.51e+04
grade	7.546e+04	3743.019	20.159	0.000	6.81e+04	8.28e+04
relative_living_area	-3.234e+04	3076.933	-10.510	0.000	-3.84e+04	-2.63e+04
suburb	-7.938e+04	2858.897	-27.765	0.000	-8.5e+04	-7.38e+04
rural	-7.338e+04	2749.503	-26.689	0.000	-7.88e+04	-6.8e+04

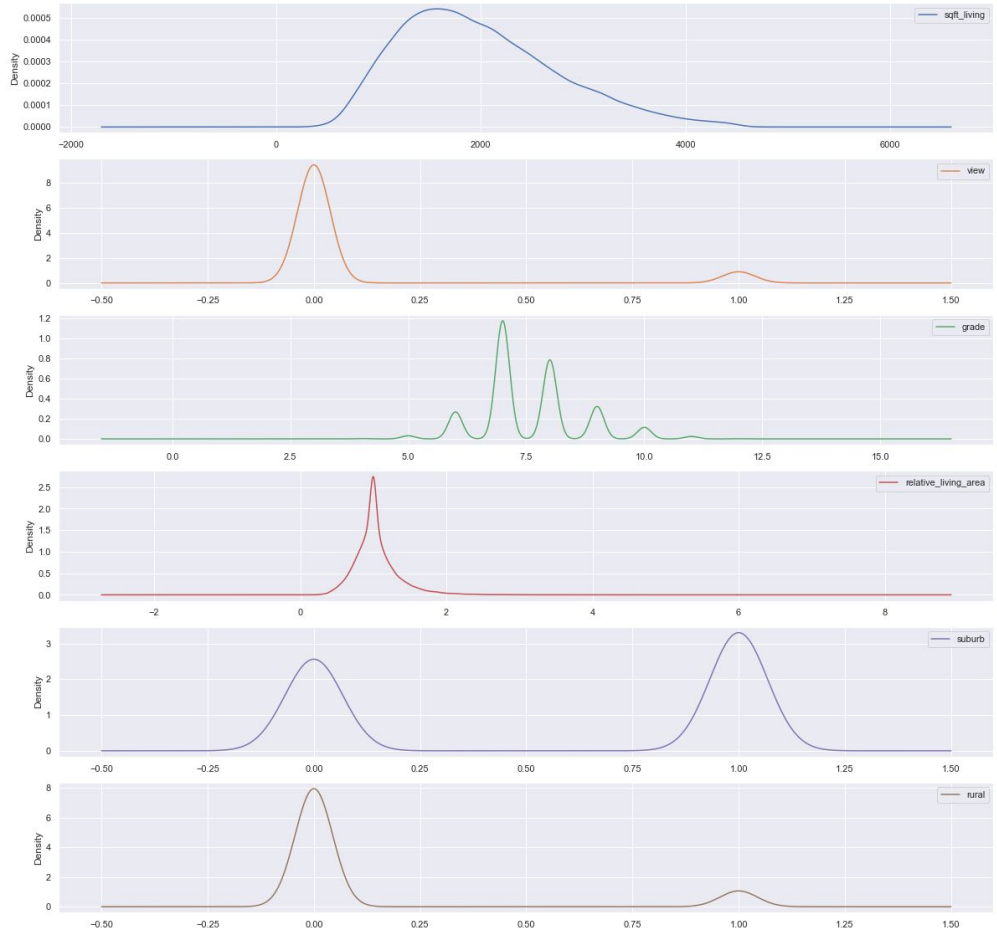
Omnibus:	840.604	Durbin-Watson:	2.033
Prob(Omnibus):	0.000	Jarque-Bera (JB):	2718.625
Skew:	1.028	Prob(JB):	0.00
Kurtosis:	6.415	Cond. No.	3.38

Root Mean Squared Error

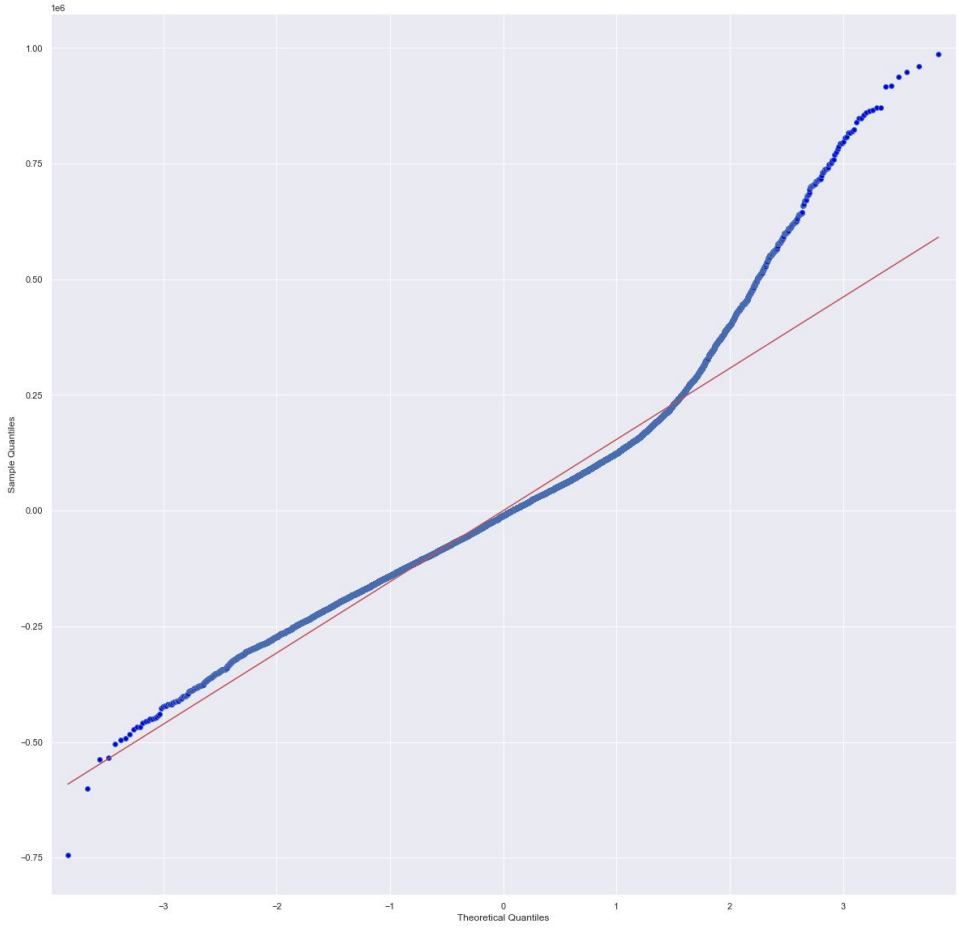
DATASET	RMSE
Train	$157,928.02
Test	$156,958.71
Difference	$969.31

Model Validation

Normalcy of Input Variables



Normalcy of Residuals



Variance of Residual Error