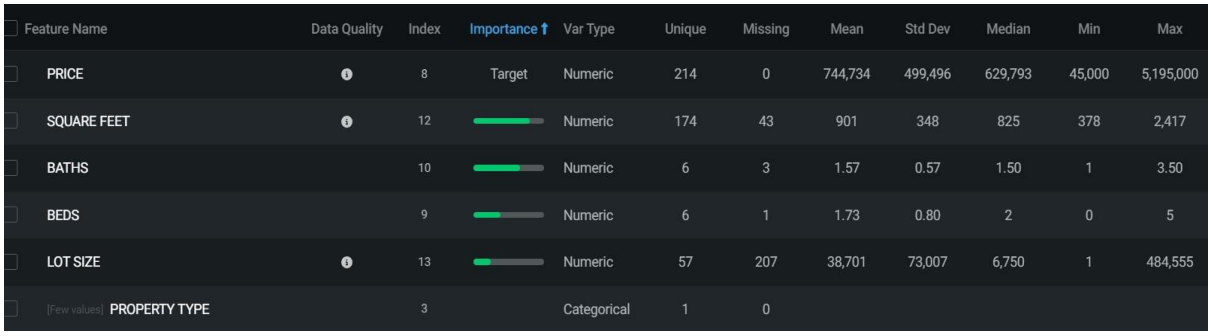
Business Case

The real estate market in New York is dynamic and complex, characterized by its high variability in property values across different neighborhoods. To refine our investment strategies and provide accurate valuations for our clients, we propose conducting a comprehensive regression analysis on a dataset comprising listings of at least 200 real estate properties in New York. This analysis aims to identify the factors most significantly impacting property values and forecast market trends, enhancing our decision-making process and competitive advantage in the real estate market.

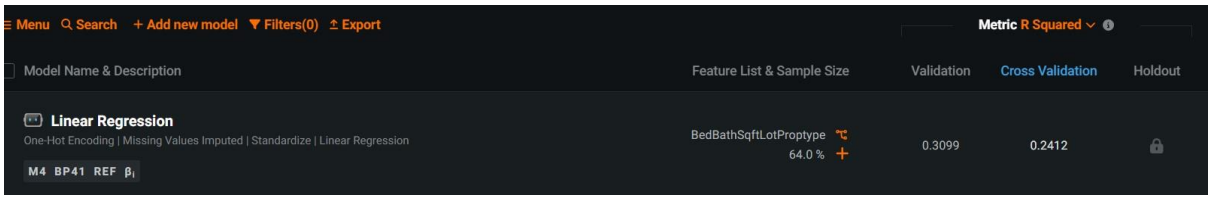
The data for condo listing is obtained from Redfin. A total of 350 listings were downloaded.

Q1. Data Pre-processing



Feature Name	Data Quality	Index	Importance	Var Type	Unique	Missing	Mean	Std Dev	Median	Min	Max
PRICE		8	Target	Numeric	214	0	744,734	499,496	629,793	45,000	5,195,000
SQUARE FEET		12		Numeric	174	43	901	348	825	378	2,417
BATHS		10		Numeric	6	3	1.57	0.57	1.50	1	3.50
BEDS		9		Numeric	6	1	1.73	0.80	2	0	5
LOT SIZE		13		Numeric	57	207	38,701	73,007	6,750	1	484,555
[Few values] PROPERTY TYPE		3		Categorical	1	0					

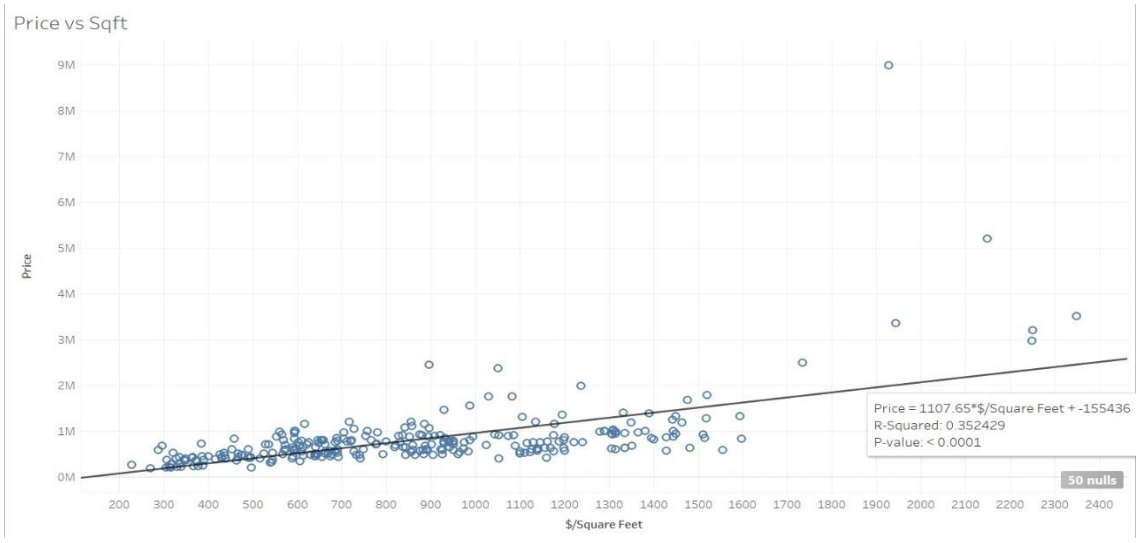
Q2. Multiple Linear Regression Model

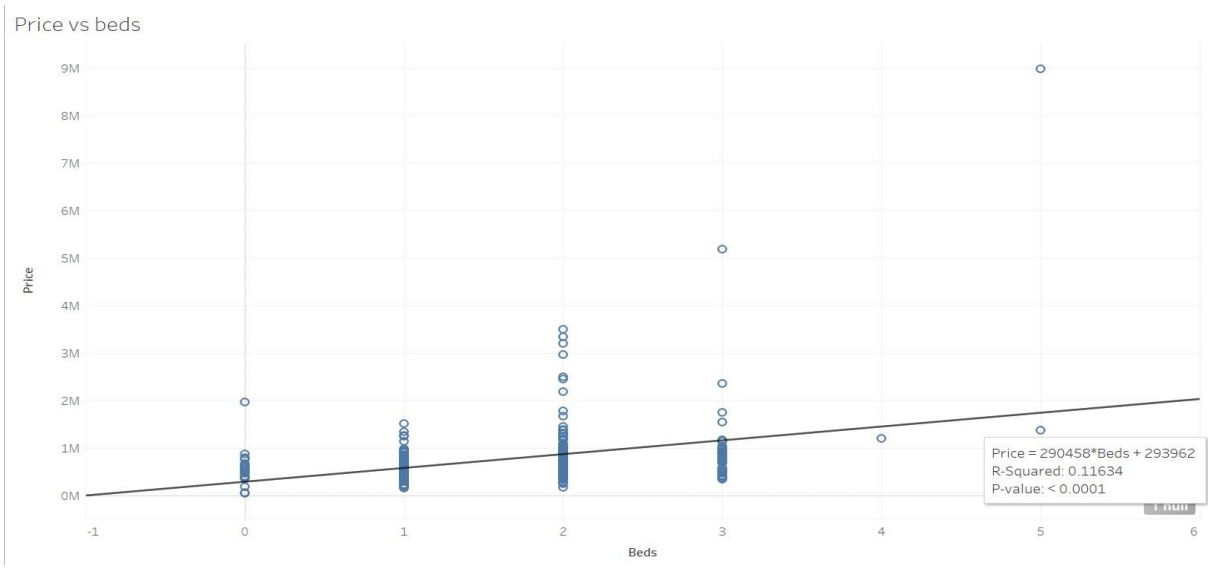


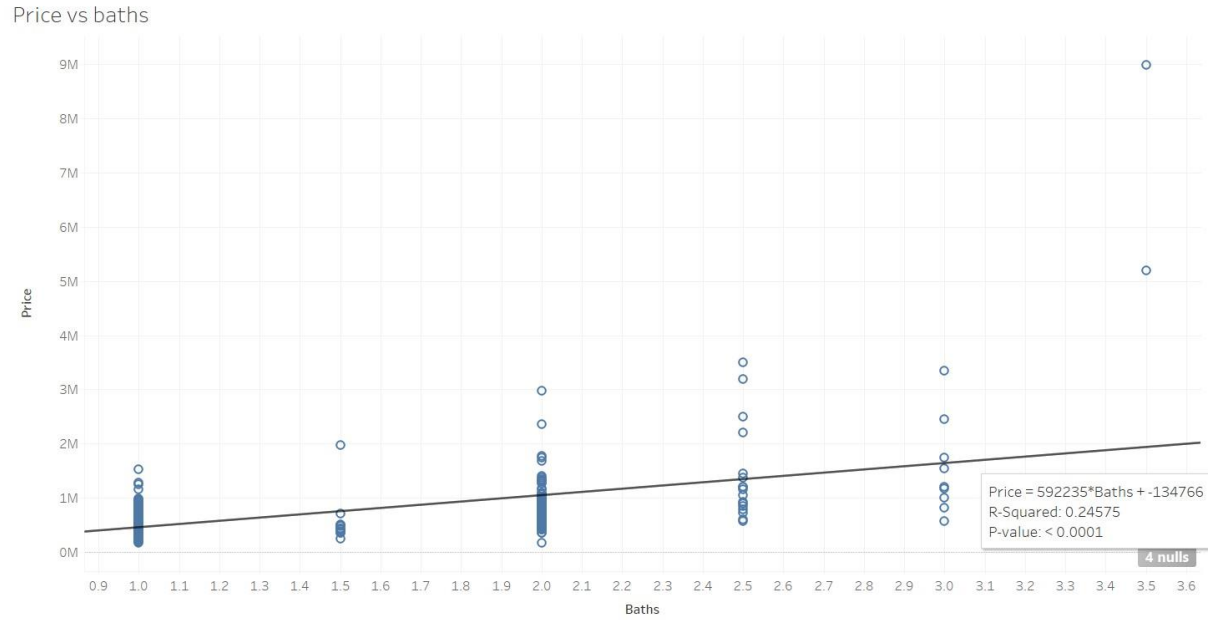
R2	0.24
MAE	$2.6M
RSME	$4.02M
MAPE	41.8%

An R^2 of 0.24 suggests that approximately 24% of the variance in the dependent variable can be explained by the model. This implies that the model might not be capturing a lot of the variability in the data, indicating a relatively weak fit. Further evident by MAE of $2.6M and RSME of $4.02M. Additionally, A MAPE of 41.8% indicates that the model's predictions are, on average, 41.8% away from the actual value.

Q3. Simple Linear Regression Models







R2 values for simple linear regression

Sq ft	0.35
Beds	0.11
Baths	0.24

Q4. Best Predictor

Based on the R2 values, sq footage is the best predictor of asking price in New York condo market.