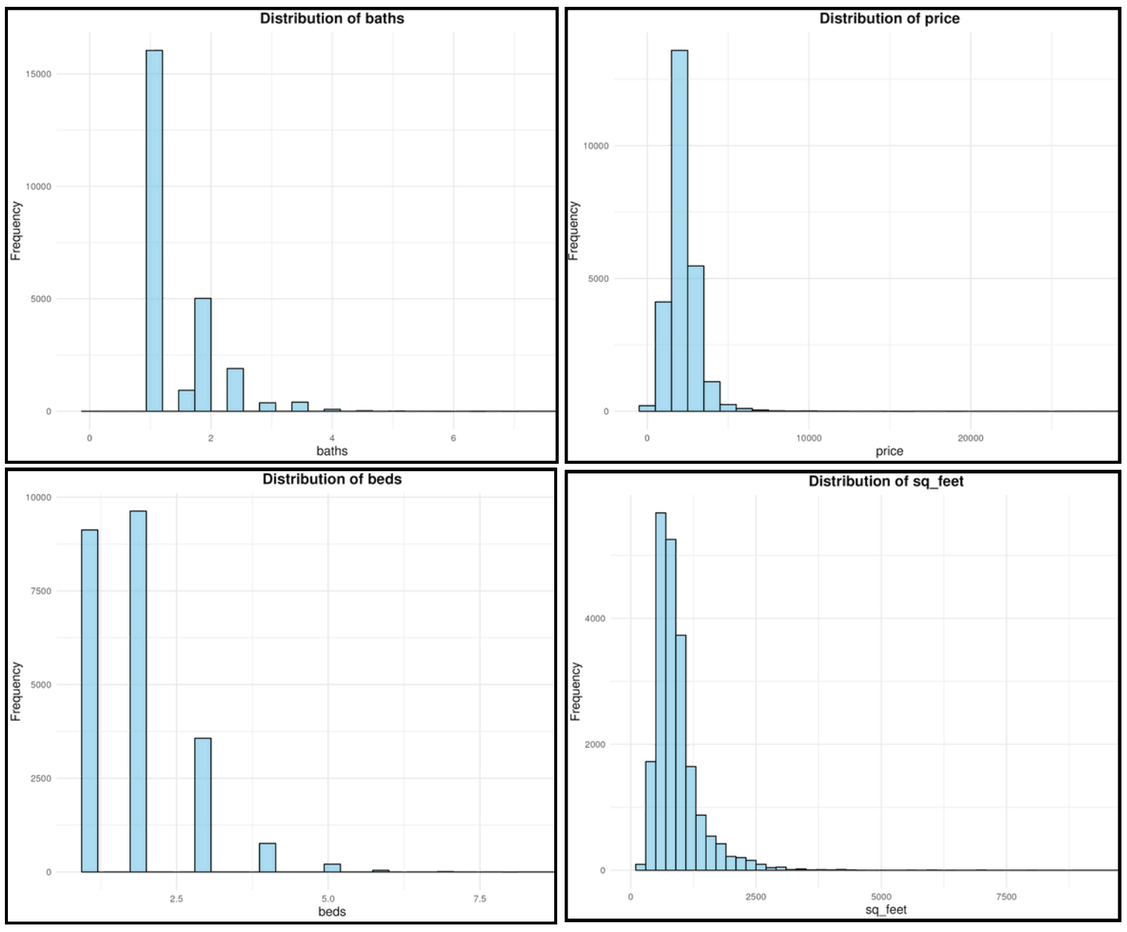
WHAT FACTORS INFLUENCE TORONTO'S RENTAL PRICES OVER THE DECADES?

An analysis of property listings to examine the impact of property type, size, and location on rents.

Key Predictors of Toronto's Rental Market Trends

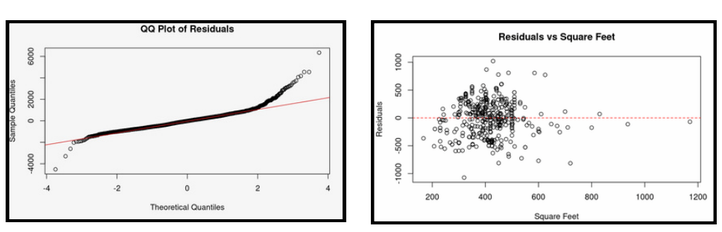


Reliability

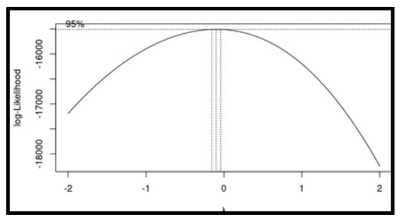
Category	Discussion
Strengths	• Large dataset: contains 25,000 listings, capturing a wide range of rental properties across Canada • Diverse representation: Includes multiple property types (apartments, townhouses, etc.) and geographic locations • Popularity and Credibility: Hosted on Kaggle, ensuring visibility and potential scrutiny of the public
Limitations	• Sampling bias: may exclude informal or offline rental arrangements, potentially overrepresenting high-end properties • Temporal snapshot: reflects the data only from June 2024, lacking seasonal or longitudinal context
Implications for Analysis	• Use the dataset to analyze rental trends and price determinants but acknowledge the sampling and temporal biases • Combine with external sources for validation (e.g., census or government statistics)

Analysis

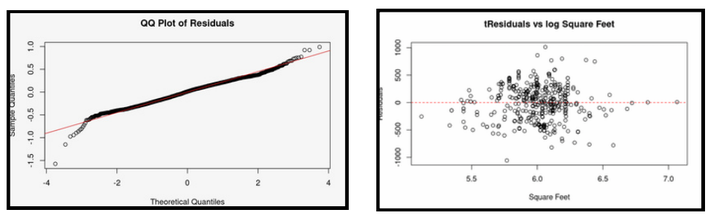
1. Initial fitted model (violation of assumptions)



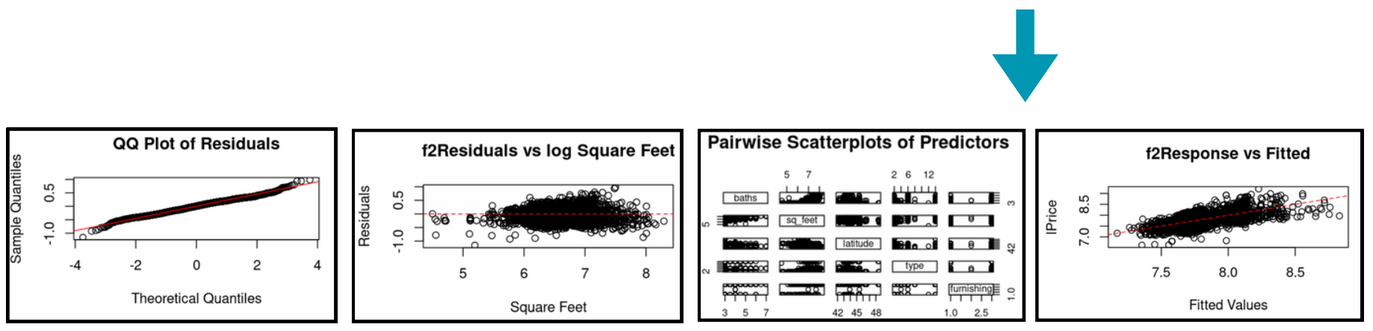
2. Box cox transformation applied



3. Improved plots (with the second tmodel)



After more transformations are applied



Conclusions

Comparison of Model Performance Metrics (R^2 Adjusted, AIC, BIC) Across Candidate Models

AIC/BIC/R^2 Values for Models			
Model_Type	R2_adj	AIC	BIC
Model 1	0.3162	−367.1218	−347.2458
Model 2	0.3669	−795.2110	−768.7097
Model 3	0.4099	−1,180.0070	−1,107.1280
Model 4	0.4218	−1,292.8700	−1,213.3660
Model 5	**0.4243**	**−1,314.6100**	**−1,221.8550**
Model 6	0.4242	−1,313.4780	−1,214.0980

Model 5 is highlighted as the optimal model, demonstrating the highest adjusted R^2 and the lowest AIC and BIC values.

The assumptions of multiple linear regression, including linearity, independence, homoscedasticity, and normality of residuals, were evaluated and appear to be reasonably met. While some limitations may exist in the residual analysis, the model selection process ensures that the final model is the best possible choice given the available data. Furthermore, all predictors in the model exhibit a strong linear relationship with the response variable, supporting the robustness and interpretability of the results.

The table summarizes the estimated effects of predictors on rental prices, highlighting significant influences such as square footage, number of bathrooms, and property type.
Negative coefficients for some property types reflect lower rental prices relative to the baseline category.

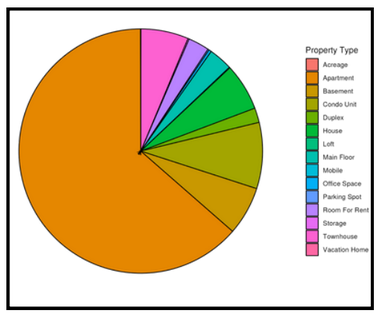
term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	504.46465	501.18718	1.0065394	3.142000e-01	−478.05813	1486.98743
lsqft	626.99115	27.98827	22.4019298	1.727159e-106	572.12321	681.85909
baths	746.42599	20.29649	36.7761180	3.297102e-265	706.63694	786.21504
latitude	−56.65139	10.32743	−5.4855262	4.304593e-08	−76.89719	−36.40559
typeBasement	−699.87526	56.21756	−12.4494058	4.126054e-35	−810.08366	−589.66686
typeCondo Unit	−218.09449	128.08373	−1.7027494	8.867088e-02	−469.18866	32.99968
typeDuplex	−869.04412	427.79617	−2.0314444	4.225742e-02	−1707.69184	−30.39640
typeHouse	−1040.66517	65.67006	−15.8468735	2.372070e-55	−1169.40417	−911.92618
typeMain Floor	−95.82867	302.98198	−0.3162850	7.517980e-01	−689.79177	498.13444
typeRoom For Rent	−298.65125	609.42589	−0.4900534	6.241154e-01	−1493.36422	896.06171
typeTownhouse	−298.93840	48.32102	−6.1865081	6.590439e-10	−393.66649	−204.21030
furnishingNegotiable	−811.88702	437.12387	−1.8573386	6.331586e-02	−1668.82069	45.04664
furnishingUnfurnished	−521.79691	61.13648	−8.5349524	1.787975e-17	−641.64830	−401.94552

Table 1. Model Summary

Predictor	Estimate	Standard.Error	t.Statistic	P.Value
(Intercept)	6.148263	0.075692	81.227	< 2e-16
lsqft	0.170397	0.013538	12.587	< 2e-16
baths	0.217056	0.007469	29.060	< 2e-16
beds	0.062927	0.006242	10.081	< 2e-16
typeBasement	−0.286957	0.019942	−14.389	< 2e-16
typeCondo Unit	−0.017580	0.045494	−0.386	0.6992
typeDuplex	−0.356938	0.151935	−2.349	0.0188
typeHouse	−0.368825	0.023372	−15.781	< 2e-16
typeMain Floor	−0.070962	0.107706	−0.659	0.5100
typeRoom For Rent	−0.304863	0.216870	−1.406	0.1599
typeTownhouse	−0.089852	0.017225	−5.216	1.89e-07
furnishingNegotiable	−0.235509	0.155277	−1.517	0.1294
furnishingUnfurnished	−0.108050	0.021622	−4.997	6.00e-07

Table 2. T-test summary for the final model coefficients.

Distribution of Rental Property Types



F-Test Summary for Final Model	
Statistic	Value
Residual Std. Error	0.02749
Multiple R-Squared	0.4215
Adjusted R-Squared	0.4203
F-Statistic	337.5
P-Value	< 2.2e-16

Table 3. F-test summary evaluates the overall fit of the final model, showcasing its explanatory power through adjusted R-squared and residual statistics.

Suggested Improvements

- <u>Outlier Treatment</u>
  - Address outliers through robust regression techniques or by capping/extending the range
- <u>Missing Data</u>
  - Consider multiple imputation or sensitivity analysis to evaluate potential bias from missing data
- <u>Model Refinement</u>
  - Explore alternative models or interactions to better account for variance and clustering

Limitations of Analysis

- <u>Model Violations</u>
  - Normality and Constant Variance: Residuals vs. fitted plot shows fanning patterns, indicating heteroscedasticity
- <u>Clustering</u>
  - Residuals vs. beds plot reveals some clustering
- <u>Normality Deviations</u>
  - The normal Q-Q plot exhibits notable deviations at the lower end of the distribution
- <u>Transformations</u>
  - Both prices and square feet were log-transformed to address skewness, but this complicates interpretability of results
- <u>Missing Data</u>
  - Approximately 10% of the data for key predictors was missing and removed under the assumption of no bias
  - However, this deletion may introduce bias
- <u>Outliers</u>
  - Outliers are evident in residuals vs. fitted and residuals vs. predictors plots (e.g., square feet, beds, and baths)