EXECUTIVE SUMMARY

SCOPE OF WORK

This report details the results of Local Logic's analysis of Client X's single family rental (SFR) portfolio of roughly 3,500 assets in a dense, major urban market. The Client wished to understand the impact of location on sales price because the transaction price at acquisition and divestment directly impacts the Client's pro forma model for assessing new opportunities and the bigger picture of managing its existing portfolio over time.

To this end, Local Logic analyzed the impact of hundreds of location characteristics on sale price for single-family detached houses in the Client's identified market. The end result is a model, as well as this accompanying report, that identifies and explains the key drivers of value for home sale price. The model leverages both asset characteristics provided by the Client and location characteristics from Local Logic's ecosystem of insights.

TOP DRIVERS OF VALUE

From the hundreds of candidate Local Logic location features considered in our modeling effort, the top five most important drivers of value for single-family home sale prices in this market were, in order of importance:

Feature	Percentage sale price increase per feature unit increase
% Commute by transit (3 miles radius)	0.90%
% Population with post-secondary degree (0.5 mile radius)	0.56%
Vibrancy: difference in location score versus city score (Point 0 - 100 scale)	0.31%
Elementary Schools Score (Point 0 - 100 scale)	-0.09%
Street distance to nearest Starbucks (log - miles)	-3.27%

The percent price increase in the tables above represents the percent change in sales price per feature unit increase, all else held constant. For example, for each 1-percentage-point increase in the population with a post-secondary degree within a 0.5-mile radius surrounding the property, the property's sales price is predicted to increase by 0.56%, all else equal. Similarly, a property that would sell for $300,000 in a neighborhood where 30% of the population has a post-secondary degree would thus be predicted to sell for $16,800 more if 40% of the population had a post-secondary degree, all else equal.

The complete list of important features for sales prices can be found in the table on pages 4-5 below.

MODEL PERFORMANCE

Local Logic's location data proved to be highly valuable in explaining the variance in sales prices in the Client's portfolio. Specifically, Local Logic data improves the explanatory power of sales prices by 16 percentage points compared to models that only include basic property features such as property size (square footage, number of bedrooms and bathrooms, number of floors), and other commonly used attributes such as building age and date sold. The final model explained 74% of the total variation in sales prices. By uncovering what features drive property value, our model enables investors to incorporate important insights into their processes for assessing the investability of a property.

DATA DRIVEN DECISIONS

To allow the Client to use the findings of the model on an on-going basis after this initial analysis work, Local Logic built an interactive Sales Price Exploration tool that makes the model and underlying data available for any address in the market and can be directly applied by the Client in real time to answer investment decisions such as:

- **Explaining how location and other factors influence sales price**. Understanding the drivers of sales price is a key factor when estimating important metrics, such as capitalization rate.
- **Adjusting the price of actual comps based on differences in location data.** Comps are useful for triangulating the potential sales price of a property, but comps never have identical characteristics to a target property. Using the model provided here alongside property and location attributes for comps, one can estimate the difference in price between the target property and the comps considered.

UNDERSTANDING DRIVERS OF VALUE

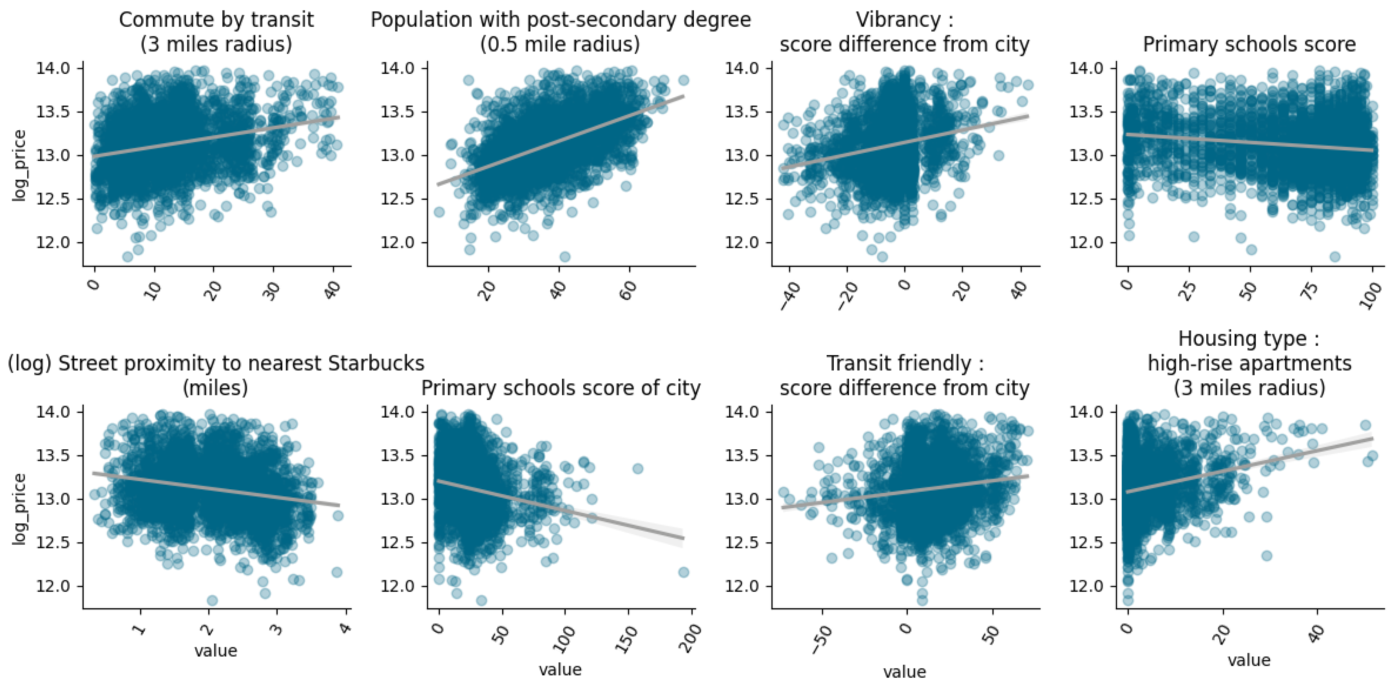
At the request of the Client, Local Logic maximized model interpretability over accuracy. To this end, a linear regression model was chosen for this portfolio analysis over other models that showed higher accuracy but that lacked the interpretability offered by the linear model.

Among the hundreds of candidate location features, Local Logic retained 17 in the model owing to their importance in helping improve model accuracy. Of these, some stood out as more important than others.

In particular, the model indicates that:

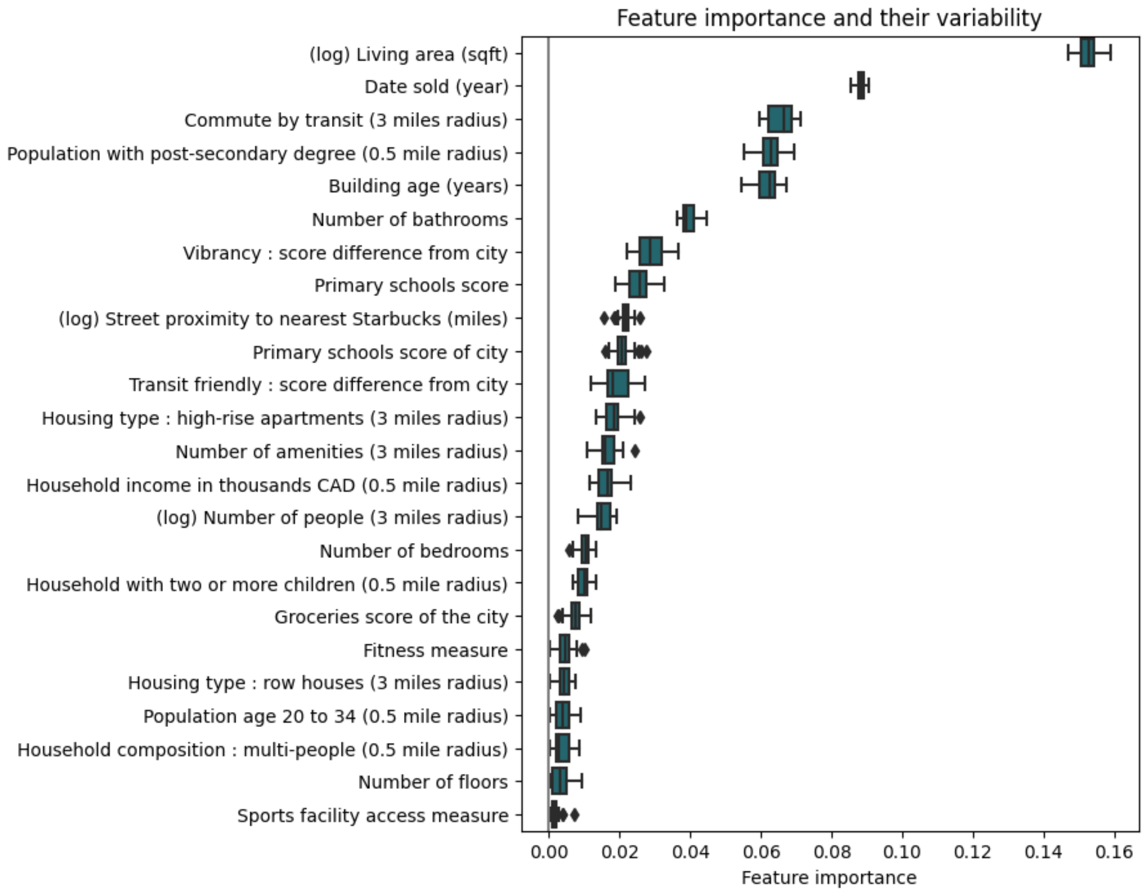
- Sale prices are higher where people commute by transit;
- Sale prices are higher where a larger fraction of the nearby population has a post-secondary degree;
- Sale prices are higher for homes with a vibrancy score higher than that of the surrounding city; and
- Sale prices are higher for homes with higher primary school scores.

The graphs below show the relationship between sale price and the top eight location features used in the model.



It is worth noting that some location features are highly correlated with one another – for example, nearby levels of income and education are highly correlated. In cases like this, the model prioritizes the feature with the largest impact on model performance (in this case population with post-secondary degree), while other correlated features appear further down the list, given that they explain the smaller portion of the price that the first variable alone could not.

The graph below offers a visual representation of the importance of the features identified by the model as having the most impact on sales price. Feature importance refers to the relative impact that feature has on the prediction being asked of the model. Unsurprisingly, the square footage of the living area of an asset is by far the most important determinant of sale price, yet a number of location features show up near the top of the list. In the graph, the horizontal bars also indicate the variability of the importance of the features during cross validation – the low variability gives us confidence in the robustness of the model.



The table below shows the features of the model listed in order of importance. As noted in the Executive Summary, the percent price increase in this represents the percent change in sales price per feature unit increase, all else held constant.

Feature	Percentage sale price increase per feature unit increase	Feature Importance
(log) Living area (sqft)	44.52%	0.152587
Date sold (day)	0.06%	0.088226
Commute by transit (3 miles radius)	0.90%	0.065665
Population with post-secondary degree (0.5 mile radius)	0.56%	0.06273
Building age (years)	-0.25%	0.061665
Number of bathrooms	6.51%	0.039424
Vibrancy: difference in location score vs city score	0.31%	0.028671
Primary schools score	-0.09%	0.025641
(log) Street proximity to nearest Starbucks (miles)	-3.27%	0.021395

Feature	Percentage sale price increase per feature unit increase	Feature Importance
Primary schools score of city	0.14%	0.021014
Transit: difference in location score vs city score	-0.12%	0.018955
Housing type : high-rise apartments (3 miles radius)	0.34%	0.018277
Number of amenities (3 miles radius)	0.01%	0.016541
Household income in thousands $ (0.5 mile radius)	0.07%	0.016163
(log) Number of people (3 miles radius)	-1.69%	0.014991
Number of bedrooms	1.13%	0.010148
Household with two or more children (0.5 mile radius)	0.13%	0.009846
Groceries: city score	-0.05%	0.007435
Fitness measure	0.02%	0.004583
Housing type : row houses (3 miles radius)	0.18%	0.004278
Population age 20 to 34 (0.5 mile radius)	0.12%	0.003962
Household composition : multi-people (0.5 mile radius)	0.26%	0.003114
Number of floors	-0.46%	0.002723
Sports facility access measure	0.01%	0.001326

As shown in the scatter plot below, which visualizes model error on predictions versus ground truth, the model performs reasonably well in all price ranges. The model's performance could be improved further with additional information on the assets within the Client's portfolio, such as lot size, quality of the building, presence of parking spot, recent renovations, etc.

LOCATION DATA VALUE

This sample report explores the drivers of single family home sale prices for properties in a single urban market. As expected, location characteristics play a major role in the pricing of a home.

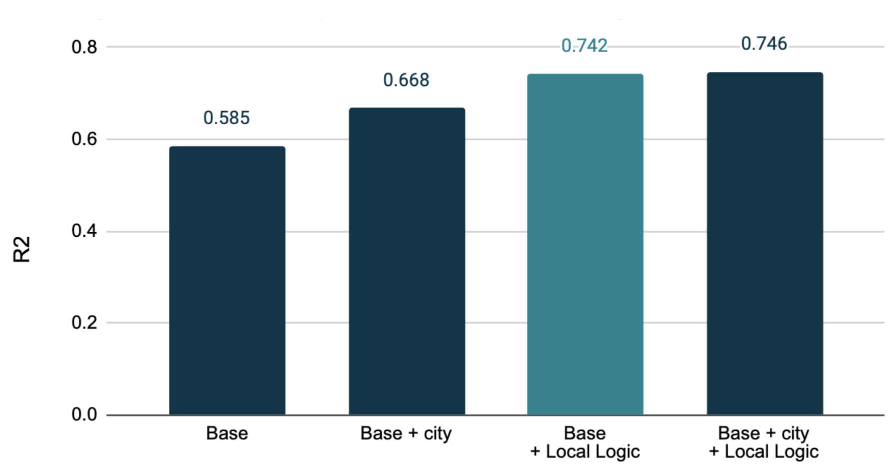
The Client provided Local Logic with a large dataset that includes several property-related features such as living area and year built, as well as geographic features such as municipal district.

Local Logic applied a hierarchical modeling approach to examine the relative importance of a property's base features, city name, and Local Logic data in explaining sale prices:

- **Base:** living area square footage, number of bathrooms and bedrooms, date sold, building age and number of floors.
- **Base + city:** the base features enriched with features indicating municipal district (here referred to as 'city'). Given the large number of municipal districts in the chosen city, we only encoded the ones that contained at least 50 properties in the training set.
- **Base + Local Logic:** the base features enriched with Local Logic data (19 location features selected from our hundreds of candidate features based on their importance for the model prediction task).
- **Base + city + Local Logic:** the combination of all of the above.

Local Logic compared the performance of the four linear regression models as measured with R^2 on a test dataset (see Fig 4 below). R^2 can be interpreted in terms of how well the regression model explains the observed data it has never seen before (i.e., data in the test set). For example, an R^2 of 0.7 reveals that 70% of the variability observed in sale price is explained by the model.

As shown in the chart below, Local Logic's location data dramatically improves model performance when predicting sales price. Combining all the data available (base + city + Local Logic data) resulted in the best performing model, even if only a small improvement over the base + Local Logic model. Since the results are more interpretable without the municipal districts included, and the loss in performance of removing these features is small, the results in this report focus on the linear model built on the base + Local Logic data only.



The above graph visualizes a comparison of the performance of 4 linear models using different data subsets. The model using both base and Local Logic features explains 74% of sale price, a 16 percentage point improvement over just base features. While the inclusion of municipal district features also improves model performance over base features, it does so to a lesser extent than Local Logic features.

USE CASES

COMPARING PROPERTIES

Below is an example of how the model can be used to compare the drivers of sales price of two very similar prospective properties. Both properties have the same amount of living space and an identical number of bedrooms and bathrooms. However, *Prospective Property A* is located in an area where the population is slightly more educated, uses transit more for commuting, and has a better Vibrancy Score compared to *Prospective Property B*. The Sales Price Exploration tool can be used to investigate the factors that influence sales prices and assess the impact of location on driving sales prices up or down. Such insights are key factors to understand the value of a prospective property, as well as estimate important metrics like capitalization rate.

	Property attributes	
Feature	**Prospective Property A**	**Prospective Property B**
Living area (sqft)	1145	1145
Baths	2	2
Beds	3	3
Commute by transit (3 miles radius)	8.28%	6.78%
Population with post-secondary degree (0.5 mile radius)	51.0%	48.0%
Vibrancy: difference in location score vs city score	-6.72	–9.22
Predicted sales price	**$374,775**	**$360,703**

The above assessment indicates that *Prospective Property A* is predicted to have a higher sales price than *Prospective Property B*, despite having only small, but key, differences in their location features. Specifically, *Prospective Property A* is expected to sell for $14,072 more than *Prospective Property B*, solely because of its slightly more favorable location characteristics. As such, Local Logic's granular location features can explain why two seemingly identical properties in very similar locations differ in value.

Note that, in this example, to make things clearer, we have simplified by including only a handful of features. The full model developed by Local Logic contains 24 features (18 from Local Logic, 6 from the Client). The Client is directed to use the full model when comparing target properties.

ASSESSING INVESTMENT OPPORTUNITIES

Perhaps the most powerful way to use the sales price prediction model is in conjunction with actual comps. Comps are useful for triangulating the potential sales price of a property, but they never have characteristics identical to those of a target property. Using Local Logic's model, it is possible to estimate the difference in sale price between a target property and selected comps based on the difference in values observed across the target property and those comps.

Below is an example using real properties from the Client's dataset to estimate the sales price of a target property given one comparable property. (It is possible to compare as many as desired.) In this example, both properties are single-story and have 2 bathrooms, 3 bedrooms, and a similar amount of living space (roughly 0.4% difference). While the properties are situated only 0.3 miles from each other, subtle but important differences are observable in their location attributes. Many investors assume they 'control for location' by selecting comps within the immediate vicinity of a prospective property, but this table makes clear that location attributes vary, sometimes in important ways, even in close proximity.

Feature	Target property attributes	Comp attributes	Difference	Expected price change
Living area Square feet	1145	1150	-5	-$1,420
Date sold Date	2023-01-20	2022-12-14	37	$15,238
Commute by transit (3 miles radius) Percentage (0 - 100	9.12	8.68	0.44	$2,896
Population with post-secondary degree (0.5 mile radius) Percentage (0 - 100	51.03	49.34	1.69	$6,963
Building age Years	10.5	13.5	-3.0	$5,599
Vibrancy : difference location vs city Score difference (-100 to 100)	-2.6	-3.2	0.6	$1,365
Primary schools score Score (0 - 100)	72.34	75.68	-3.34	$2,215
Street distance to nearest Starbucks Miles	6.16	6.63	-0.47	$1,571
Household income in thousands USD (0.5 mile radius) Percentage (0 - 100	111.93	106.98	4.95	$2,428
Household with two or more children (0.5 mile radius) Percentage (0 - 100	33.7	32.0	1.7	$1,616
Housing type : row houses (3 miles radius) Percentage (0 - 100	2.46	1.34	1.12	$1,463
Population age 20 to 34 (0.5 mile radius) Percentage (0 - 100	15.24	13.65	1.59	$1,415
Household composition : multi-people (0.5 mile radius) Percentage (0 - 100	2.45	1.85	0.60	$1,099
Predicted difference in sale price				**$41,542**

Here, the Sale Price Exploration tool indicates that the target property should be priced $41,542 more than the comparable property. It also gives insight into which attributes are most responsible for the estimated price difference between the two properties. Amongst the most important drivers of price difference are: 1) the percentage of population with a post-secondary degree ($6,963); 2) the percentage of population that commutes by transit ($2,896); and 3) the household income of the surrounding area ($2,448).

Note that, in the example above, to simplify the exercise, we removed all rows with a contribution of less than $1,000. The Client is directed to use all features in this kind of exercise. In this case, including all features would have resulted in a predicted difference in sale price between the two properties of $45,740.75.

NEXT STEPS

With this model and analysis in hand, the Client will use the model in on-going investment decision making.

In order for the Client to use the model to generate a sale price prediction for a prospective property, the Client must have access to the data behind the 24 features used in the model, including the 6 property features for which it has direct access to the data, and the 18 location features, which will be provided through a separate agreement with Local Logic.

Local Logic aims to constantly improve and expand the set of location features available for clients to use for investment decisions and/or similar modeling exercises. This means that clients will have access to a growing number and diversity of signals about location as they opt to re-train their model in the future or as they opt to build new models to predict other performance variables down the line.

METHODOLOGY

To ensure a robust, interpretable and usable approach to modeling, Local Logic proceeded as follows:

1. Align with Client on Response Variable

The response variable is the variable we aim to predict in our modeling effort. It is sometimes also referred to as the ‘target variable’ or ‘output variable.’ Different modeling approaches and different input data will be necessary for different response variables. For example, the drivers of home sale prices might be different from the drivers of time on the market. This makes it very important for the client and Local Logic to align upfront on which response variable the client wishes to model. In this case, the client chose ‘sale price’ as the response variable.

2. Build a property dataset combining property and location attributes

Local Logic combines its set of location insights with additional property data provided by the Client to create a single unified dataset providing a deep view into the location and property attributes of each property in the portfolio.

3. Split the dataset into train and test sets to measure generalizability

Building a model on the entire dataset described above would allow us to achieve very high predictive performance for this dataset, but would significantly increase the risks that the model would not do well on data it has never seen before. In machine learning terminology, this increases the risk of overfitting the model on the portfolio, reducing its value for the client's stated goals: using the model to improve decisions about properties and markets it has never encountered before.

To overcome the risk of overfitting and to improve the generalizability of the model, we split the dataset into two using a randomized stratified sampling process. One portion (80%) of the data is used for training the model, while the other portion (20%) is hidden or reserved. Once a sufficient level of performance is achieved, the model is tested on the hidden test set to measure its capacity to generalize. The performance metrics we report are those based on the model's performance on the test set.

The Client's home sales dataset consisted of roughly 3,500 unique single family home properties. All assets were located in the same dense, urban market in Canada. After reserving 20% of the Client's data set for model evaluation and removing outliers and instances with insufficient data, we obtained a final training dataset of roughly 2,750 assets for our sales price prediction model.

4. Feature engineering and feature selection

Using the dataset consisting of location and property attributes, we select features that improve the model's performance, excluding features that offer no or very little help in this task.

5. Test model via cross validation

At this stage the model is tested by performing a cross validation in order to measure its performance on unseen data and ultimately to judge its capacity to generalize beyond the data it was trained on.

6. Iterate on steps 3 - 5 until model generalizes well

7. Measure the model's performance

Once the model is built, we measure its performance on the test set that was set aside in step 3.