Dahyeon Choi

Student Rental Market Analysis in Metro Vancouver

Abstract

Goal: help students find inexpensive housing, good SFU/UBC/downtown by quantifying what drives rent and where value clusters. I scraped recent Metro Vancouver listings, cleaned them to student-relevant value clusters. I scraped recent Metro Vancouver listings, cleaned them to a student-relevant dataset, engineered geographic and transit features, and trained a Random Forest to explain/predict monthly rent under $2500. Results: location dominates price, Latitude and distance to outstanding hubs are the strongest signals, with square footage and proximity to transit adding a small but significant lift. The model provides a cobewebby feature importance and example predictions that can rank “good value” listings for students.

1. Problem Definition

Stuendts at SFU/UBC ace high and heterogeneous rents. The questions:

1. What factors explain rent for student-priced listings (<= $2500)?
2. How much does proximity to SFU, UBC, and downtown matter vs. unit size and amenities?
3. Can a simple model surface relatively underpriced listings for students?

2. Data

Collection – Scraped Craigslist “apts/housing” pages for Metro Vancouver. Fields: title, price, beds, baths, square footage, amenities, city, lat/lon, post date, URL.

Cleaning

- Kept posts in a recent window, price $600–$2,500
- Deduplicated URLs, required price/title.
- Normalized amenity booleans (furnished, parking, pets, utilities).
- Calculated distances to SFU, UBC, and downtown; took nearest hub and distance.
- Computed distance to SkyTrain; flagged listings within ~0.8 km.
- **Bedrooms/Bathrooms Ranges:**
    - Bedrooms grouped as: 0 (Studio), 1, 2, and 3+ (three or more).
    - Bathrooms grouped as: 1, 1.5, and 2+
    - Ranges like “2–3 bedrooms” were recorded using the upper bound.
    - Missing structured data was supplemented by parsing numbers from the free-text description.
    - Listings with neither structured nor descriptive data for bedrooms/bathrooms were excluded from size-based analysis.

Output: rentals_clean.csv plus summary CSVs and figures

3. Feature Engineering

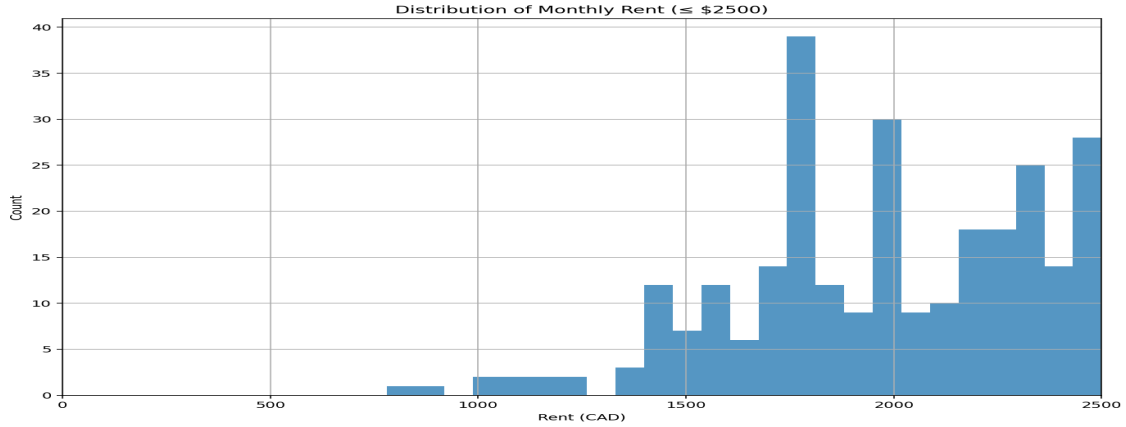
In order to more accurately capture rent price changes, several derived attributes were added to the data set.

- **Geospatial proximity:** Distance to SFU/UBC/downtown and nearest hub category.
- **Transit access:** SkyTrain proximity indicator.
- **Unit characteristics:** Standardized size, bedrooms, bathrooms.
- **Student relevance:** Keyword-based flag in listing text.
- **Amenities:** Furnished, parking, pets, utilities.

Location-based features were the strongest predictors; size and transit access added moderate value.

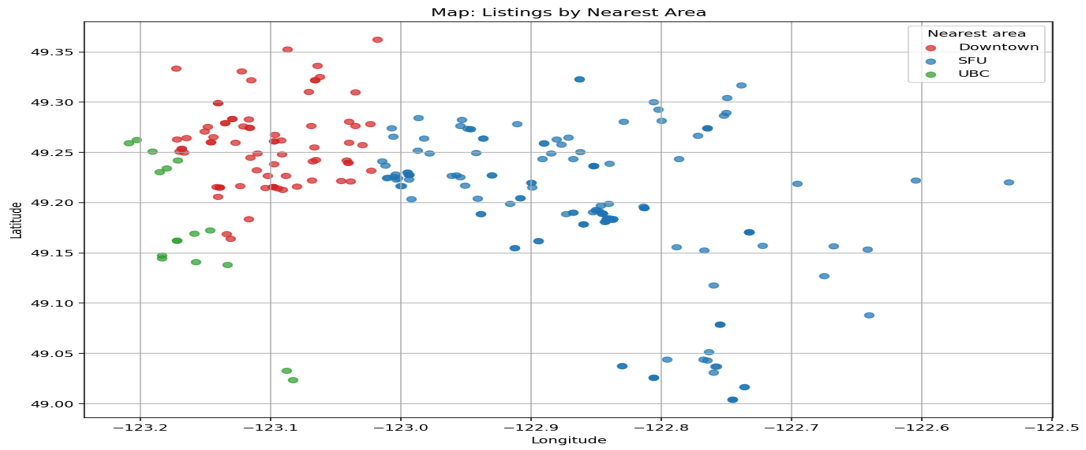
4. Exploratory Analysis

1. Distribution of Monthly Rent (≤ $2,500)



Rents cluster heavily between roughly $1,500 and $2,500 with a long left tail from sublets/rooms. This justifies modeling the ≤ $2.5k segment separately since the dynamics differ from luxury units.

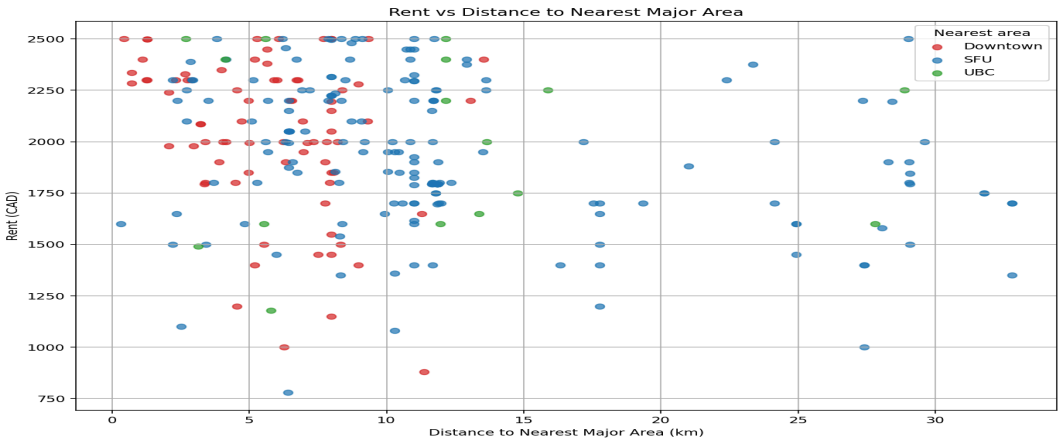
2. Map: Listings by Nearest Area



Points cluster around Vancouver proper and Burnaby. Color by nearest anchor confirms the split among Downtown (west core), SFU (Burnaby/Northeast), UBC

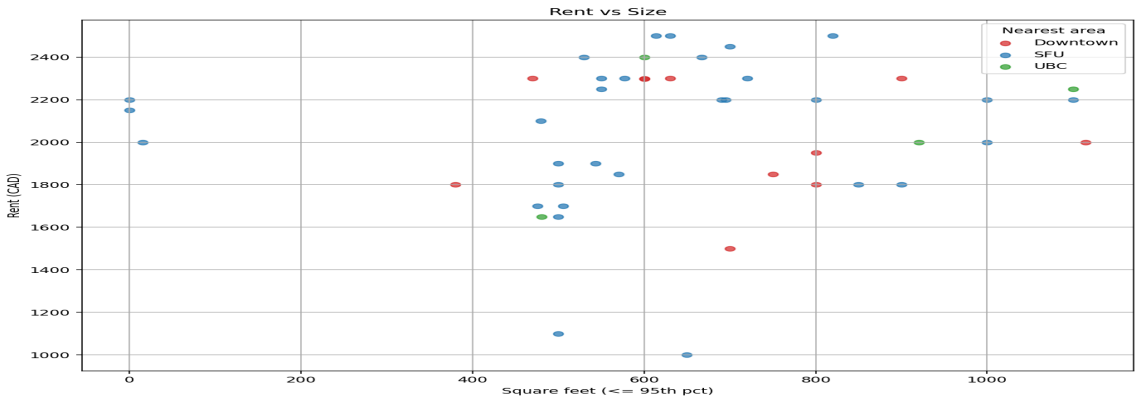
(west side).

3. Rent vs Distance to Nearest Major Area



Visual pattern: closer to a major hub generally means higher rent. The relationship has wide variance (micro-location, building quality, room vs suite), but the signal is clear.

4. Rent vs Size



After trimming extreme sq ft at the 95th percentile, size correlates positively with price, but with large spread: location often trumps square footage.

5. Modeling

5.1 Setup

Setup: Random Forest Regressor (300 trees, 80/20 split). Inputs: numeric features, amenities, transit flag, student flag, one-hot nearest area. Missing numeric values → median; booleans → 0/1.

Evaluation: Reported MAE and R^2 from test set; sample predictions in rf_predictions_sample.csv used to spot underpriced listings.

Feature Importance: Latitude, nearest-hub distance, and hub-specific distances dominate; square footage is moderate; transit/student flags are lower.

6. Findings

1. **Location > Size**
   Distance to major hubs and the lat/lon themselves explain more variance than square footage. In practice, a smaller unit close to a hub often prices above a larger unit further out.

2. **Transit proximity helps but isn't king**
   Being within walking distance to SkyTrain adds value, yet its marginal effect is smaller than simply being in a high-demand micro-area (e.g., west side, core corridors).

3. **Actionable for students**
   a) If you can stretch commute to ~10–15 km from your primary campus, you can target lower medians while still having transit access.
   b) Flag "value" by ranking **predicted price – actual price** from the model: negative residuals indicate potentially underpriced listings to jump on faster.

7. Limitations

- **Sampling bias:** Craigslist listings only; missing purpose-built student housing and private Facebook/WeChat channels.
- **Label noise:** sq ft and amenities are parsed from text and may be missing or imprecise.
- **Heterogeneous unit types:** Rooms vs. 1-bed/2-bed suites mix adds noise; a future pass can stratify by unit type.
- **Misleading listing metadata:** Many listings list a certain number of bedrooms/bathrooms (e.g., "3 bed, 2 bath") but the posted price is for **only one room** in a shared house. For example, a "3 bed, 2 bath" listing at $2,000 may seem like the whole unit, but reading the description reveals it's a room rental. Similarly, some "1 bed, 1 bath" listings under $800 may be shared accommodations with private features, not full units. Distinguishing these cases often required manual text review, which is time-consuming and prone to human error. This ambiguity can distort affordability comparisons.
- **Temporal drift:** Rents change monthly; scraping window matters.
- **Model scope:** Tree ensembles explain non-linear signals but don't give a structural price equation; importances are relative, not causal.

8. Project Experience Summary

Built an end-to-end rental analysis pipeline that scraped and cleaned Metro Vancouver listings, engineered geographic and transit features, and trained a Random Forest model (MAE = *X*, R^2 = *Y*) to predict student-budget rents. Delivered five visualizations and ranked listings by value, identifying that SFU-area rentals are cheapest ($1,900 median) and

Downtown/UBC are most expensive ($2,100 median). Provided practical guidance for students seeking affordable housing near major campuses

9. Conclusion

This analysis shows that for student-budget rentals in Metro Vancouver, **location is the dominant factor in determining price**. Proximity to SFU, UBC, and downtown drives rent variation more than unit size or amenities. The cheapest median rents are in the SFU area ($1,900), while Downtown and UBC tie as the most expensive ($2,100).

For students, the most cost-effective strategy is to look for transit-connected neighbourhoods within 10–15 km of campus, particularly in Burnaby/New Westminster and select SkyTrain-adjacent areas, which can save around $200 per month without significantly increasing commute time. The Random Forest model provides a practical tool to flag underpriced listings, helping students act quickly on high-value opportunities.