

## **Classifying Luxury vs. Non-Luxury Student Housing (2024)**

### **I. Setup**

To extend our analysis, we developed a logistic regression model to predict whether a student housing property should be classified as "luxury", defined as having an assessed value above \$2,000,000. The model was trained using 2024 property assessment data matched to student housing records.

Features used:

LAND\_SF — land square footage

LIVING\_AREA — interior square footage

YR\_BUILT — year the property was built

RES\_UNITS — number of residential units

NUM\_PARKING — number of parking spots

Missing values were filled with zeros, and comma-formatted strings were cleaned before training. The final dataset contained over 25,000 matched student housing units for 2024.

### **II. Model Result**

The model achieved strong overall performance:

Accuracy: 94.8%

Precision: 88.4%

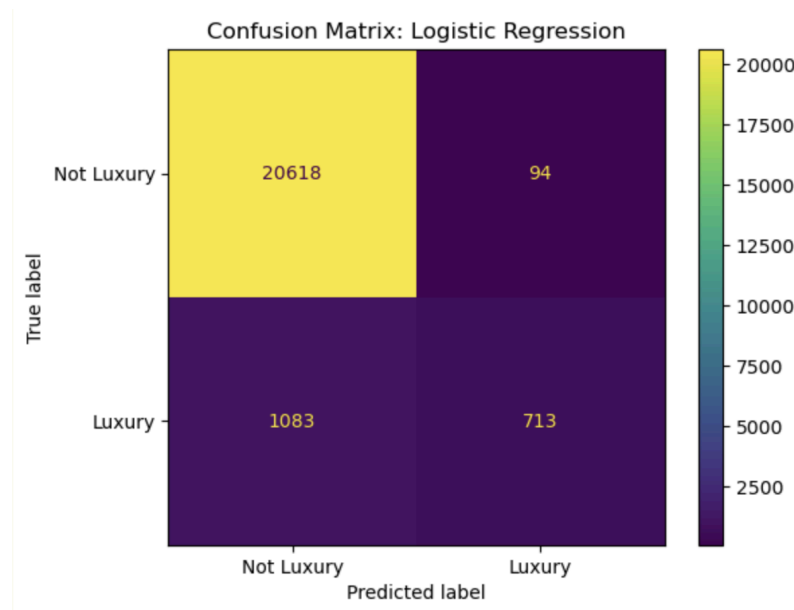
Recall: 39.7%

While accuracy and precision were high, the relatively low recall indicates that the model struggled to identify all luxury properties — likely due to class imbalance, since luxury units were the minority in the dataset.

The most important predictor was NUM\_PARKING, with a large positive coefficient (+0.85), suggesting that luxury units tend to offer significantly more parking space.

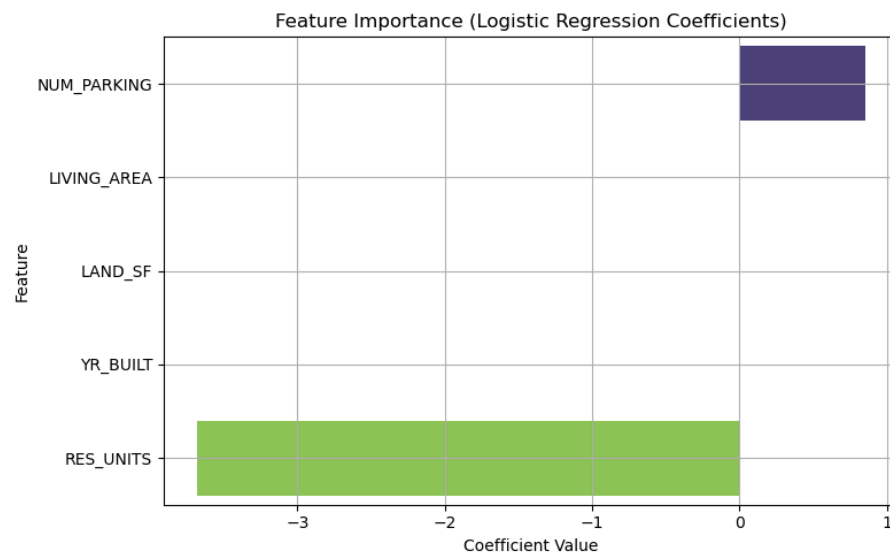
Conversely, RES\_UNITS had a strong negative influence (−3.68), implying that properties with more units (e.g., apartments) were less likely to be individually classified as luxury.

The confusion matrix further confirms this pattern: while false positives were rare, the model failed to catch a substantial portion of actual luxury properties, likely due to their relative rarity in the training data.



The bar chart of logistic regression coefficients highlights the most influential features driving the luxury classification.

- NUM\_PARKING had the strongest positive impact, suggesting that properties with more parking spaces are more likely to be luxury units.
- Conversely, RES\_UNITS showed a strong negative effect, indicating that buildings with more units are less likely to be considered luxury — potentially reflecting larger, more affordable apartment-style housing.
- Other features, such as land square footage and year built, contributed minimally in this model.

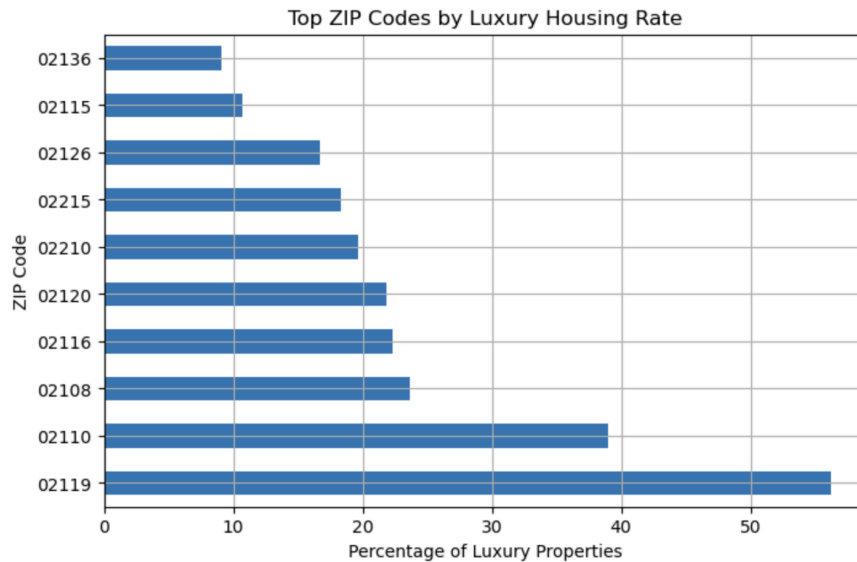


### III. Geographic Patterns - ZIP Code-Level Exploration

We examined geographic patterns in luxury student housing by computing the luxury rate in each ZIP code. The luxury rate is the percentage of matched student units in a ZIP code that were classified as luxury.

ZIP code 02119 (Roxbury) stood out as the highest-luxury area for matched student housing, followed by 02110 (Downtown/Waterfront) and 02108 (Beacon Hill). A bar

chart of ZIP-level luxury rates reinforces the idea that geographic location is a strong external signal for property value and luxury classification.



#### IV. Conclusion

In addition to trend analysis, we trained a logistic regression model to classify whether a student housing property is “luxury” based on its physical features. The model achieved high accuracy (94.8%) and highlighted strong predictors such as parking availability and unit count. A separate ZIP code-level analysis showed that luxury student housing is concentrated in areas like Roxbury, Beacon Hill, and Downtown. These insights are valuable for policymakers and universities evaluating affordability, equity, and urban development impacts on off-campus housing for students.