Color Code: A Machine Learning Study of Urban Facade Colors and Property Attributes

John Masataka Xianfeng Jiang[1], Skylar Siqi Zhang[2] (Alphabetical Order)

[1]Graduate School of Architecture, Planning and Preservation, Columbia University, Avery Hall, 1172 Amsterdam Ave, New York, NY 10027; e-mail: xj2329@columbia.edu
[2]Graduate School of Architecture, Planning and Preservation, Columbia University, Avery Hall, 1172 Amsterdam Ave, New York, NY 10027; e-mail: sz3304@columbia.edu

ABSTRACT

This study explores the relationship between urban building characteristics and visual aesthetics, as expressed through facade color, using a machine learning and geospatial analysis framework applied to New York City's Bronx borough. By computing the circular mean hue from segmented Google Street View images and linking them to parcel-level data from PLUTO, we examined how variables such as land use, zoning, building class, assessed value, floor area ratio (FAR), and construction year correlate with facade hue. Multiple supervised learning models, including Logistic Regression, Decision Tree, Random Forest, and K-Nearest Neighbors (KNN), were evaluated across datasets generated with varying spatial offsets. The KNN model trained on the 30-meter offset dataset demonstrated the highest predictive performance (accuracy = 62.2%) and was selected for downstream analysis. Statistical techniques, including Pearson correlation and ANOVA, revealed weak but statistically significant associations between facade hue and selected continuous and categorical variables. Notably, *BuiltFAR* and *LandUse* showed the most robust associations, suggesting that urban density and programmatic use may influence aesthetic patterns. This study contributes to the growing field of computational urban aesthetics and provides a foundation for further research into the spatial logic of visual experience in cities.

INTRODUCTION

Urban environments are visually rich, yet color remains an understudied dimension in urban analytics. While previous research has focused extensively on land use, spatial density, and architectural form, the color of building facades has rarely been treated as a dataset. However, façade color carries more than aesthetic significance, it may reflect cultural preferences, zoning policies, building materials, and maintenance levels. In diverse and dense neighborhoods such as the Bronx in New York City, variations in facade color may encode latent urban patterns worthy of investigation.

Recent advances in street view imagery and computer vision have enabled new methods for analyzing visual elements of the city. Current studies have demonstrated that facade color can influence real estate value and user preference (Chen et al., 2023), affect visual comfort (Wang et al., 2024), or align with street function and design trends (Zhai et al., 2023). However, many of these works emphasize perceptual or thermal aspects rather than exploring how facade color might correlate with measurable urban form and property data.

This study focuses on hue, the most interpretable component of color, representing its type, such as red, blue, or yellow, rather than its intensity. Hue serves as a stable and meaningful visual metric in architectural analysis: it can distinguish material types, and it is less affected by lighting variations, shadows, or time of day. Unlike saturation and brightness, which fluctuate with exposure and camera settings, hue remains relatively consistent in outdoor imagery, making it especially reliable for large-scale, street–view–based façade analysis.

Using segmented street view images and parcel-level data, this study investigates whether façade hue correlates with key urban variables. Through machine learning classification, statistical correlation analysis, and geospatial visualization, we aim to reveal how visual cues—specifically, building color—relate to patterns in urban form and policy.

DATA SOURCES

This study utilizes three primary data sources. First, Google Street View imagery was collected through the Google Maps API. Second, parcel-level data was obtained from the New York City PLUTO dataset (February 2025, 25v1), which includes variables such as AssessTot, BuiltFAR, YearBuilt, BldgClass, LandUse, and ZoneDist (NYC Department of City Planning, 2025). Finally, a derived dataset of façade hue values was created through semantic segmentation and HSV color extraction from the street view images, producing a mean_hue metric aligned with each tax lot.

METHODOLOGY

This study develops a geospatial and visual analytics pipeline to examine the relationship between building façade hue and urban parcel attributes in the Bronx, New York City. The methodology consists of following main stages:

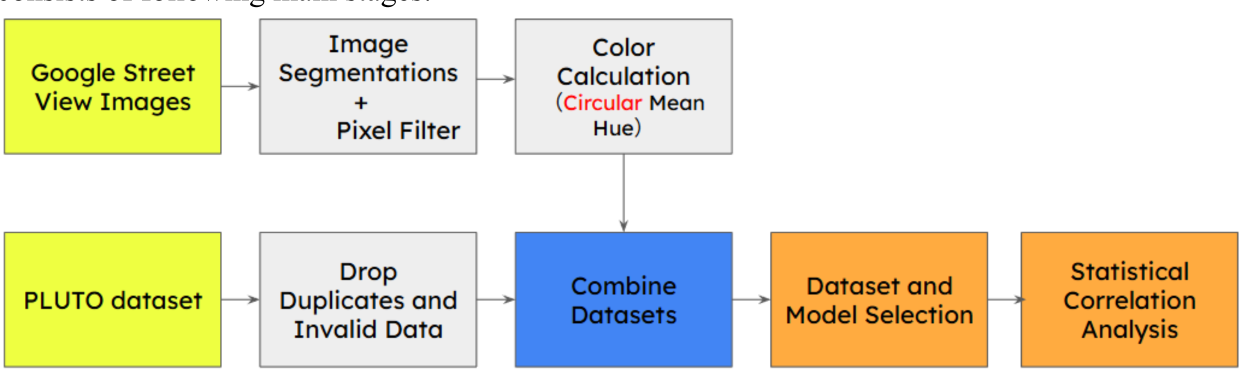

Figure 1. Workflow

1. Street View Image Collection: Street-level imagery was acquired, targeting geographic coordinates within the Bronx. For each location, four directional views (facing north, east, south, and west) were created to ensure complete coverage of surrounding façades. These images serve as the primary visual input for the subsequent segmentation and hue extraction steps. To maximize alignment with building parcels, geographic points were sampled based on parcel centroids and later refined through spatial offset experimentation.

Figure 2. Street View Images Collected Using Google Maps Street View API.

2. Facade Segmentation via Deep Learning: To isolate building façades from background elements in Google Street View (GSV) images, we applied a semantic segmentation model—SegFormer-B0, pretrained on the ADE20K dataset. The model was implemented using Hugging Face Transformers and classifies each pixel in an image according to 150 semantic categories. For each GSV image, we identified pixels labeled as "building" (label ID 1) and retained only those regions. All other regions were masked as transparent. The output was a set of façade-only images in RGBA format, allowing for cleaner and more accurate color extraction in the next step. This deep learning-based preprocessing step ensures that hue values are calculated from façade surfaces exclusively, excluding sky, vegetation, people, and other visual noise.

Figure 3. Building Façade Segmentation from the Street View Images.

3. Hue Extraction from Segmented Images: The RGBA images from the previous step were processed using OpenCV to extract hue information. Each image was converted to HSV color space, and pixels were filtered by alpha transparency, saturation, and brightness thresholds to remove low-quality or non-building areas. Hue values were converted to degrees (0–360°), and

the circular mean of hue was computed for each image. Four directional images (north, east, south, west) per location were averaged to produce a single *mean_hue* value per street view ID (*svid*).

```
FOR each street_view_id in dataset:
    hues = []

    FOR direction in [north, east, south, west]:
        image = load_RGBA_image(svid, direction)
        hsv = convert_to_HSV(image)
        mask = valid_pixels(alpha, saturation, value)
        hue_values = hsv.hue[mask] * 2  // degrees

        IF hue_values not empty:
            mean_hue = circular_mean(hue_values)
            add to hues

    IF length(hues) == 4:
        final_mean = circular_mean(hues)
        write to CSV as svid's mean_hue
```

Figure 4. Hue Extraction Algorithm (Pseudocode).

4. Data Cleaning: To ensure consistency and validity in subsequent analyses, the dataset was filtered to retain only parcels located in the Bronx (Borough == 'BX') with non-missing and plausible values across all key variables. These include AssessTot (assessed property value), BuiltFAR (floor area ratio), YearBuilt (greater than 250 to exclude placeholder entries), BldgClass, LandUse, ZoneDist, as well as the extracted mean_hue from Google Street View images. This step removed incomplete or erroneous records and ensured the analytical dataset consisted of high-quality observations suitable for correlation and classification modeling.

5. Spatial Integration with Parcel Data: The hue data were spatially joined to New York City's PLUTO dataset. To align imagery with parcels, we tested multiple offset distances (6m, 10m, 12m, 15m, 20m, and 30m). Each offset version linked geotagged hue points with the nearest polygonal parcel boundary, producing multiple datasets for comparison.

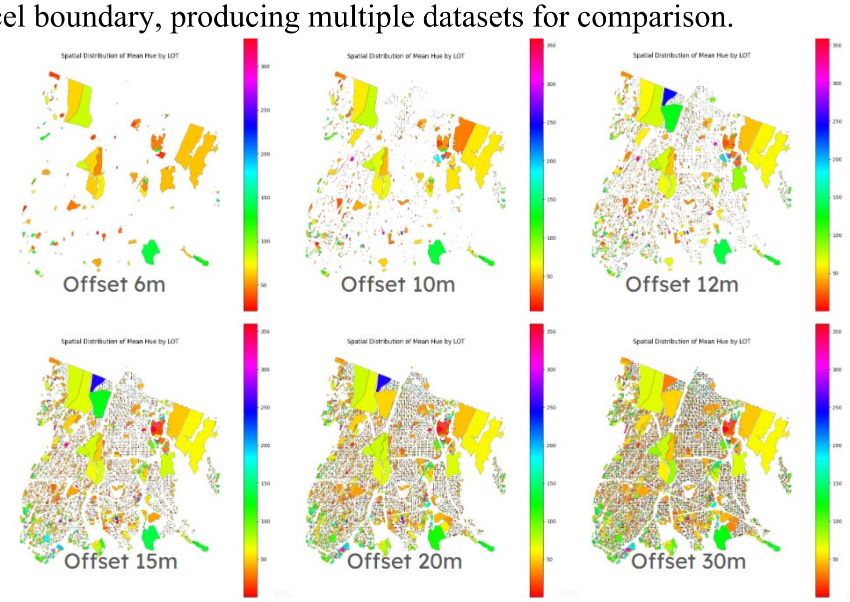

Figure 5. Tested Dataset in Different Offset Distances.

6. Dataset and Model Selection: To determine which spatial offset configuration best captured the relationship between façade hue and parcel attributes, we trained four classification models (Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbors) to predict whether a parcel's mean_hue fell above or below the median (i.e., warm or cool). The highest-performing model was KNN on the 30-meter offset dataset, which achieved 62.2% accuracy and was selected for downstream statistical analysis.

7. Statistical Correlation Analysis: With the best-performing dataset selected, we conducted statistical analysis to explore the relationships between mean_hue and urban variables. For continuous variables (*AssessTot*, *BuiltFAR*, *YearBuilt*), we used Pearson correlation to assess linear relationships. For categorical variables (*BldgClass*, *LandUse*, *ZoneDist*), we applied one-way ANOVA to evaluate whether group means differed significantly. Visualizations including scatterplots, boxplots, and choropleth maps supported both descriptive and inferential interpretation.

RESULT

This section presents the results of model performance evaluation and statistical analysis linking façade hue to urban parcel attributes. After selecting the optimal dataset configuration based on classification accuracy, we conducted both correlation and variance analysis to quantify the relationships between façade color and urban form.

1. Dataset Selection: Model Evaluation and Dataset Selection: Among all tested spatial configurations, the 30-meter offset dataset yielded the highest classification performance when using K-Nearest Neighbors (KNN), achieving an accuracy of 62.2% (Table 1). This dataset was used for subsequent statistical analysis, as it best captured the relationship between building façade hue and urban attributes.

Dataset	Logistic Regression	Decision Tree	Random Forest	KNN
Offset 12m	0.604	0.588	0.619	0.595
Offset 15m	0.576	0.591	0.597	0.593
Offset 20m	0.589	0.604	0.615	0.610
Offset 30m	0.575	0.592	0.598	0.622

Table 1. Dataset Selection Result.

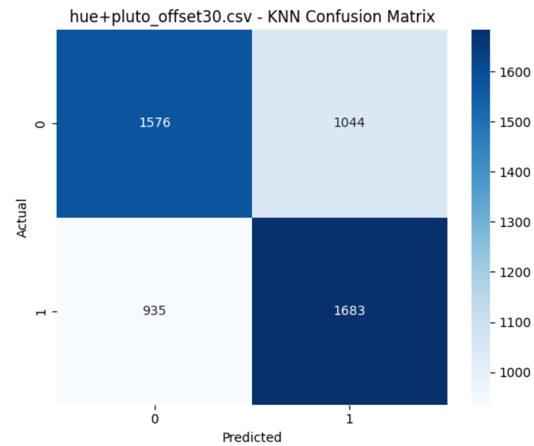

Figure 1. Confusion Matrix of KNN at 30m-offset Dataset.

2. Correlation with Continuous Variables: Pearson correlation analysis was conducted between *mean_hue* and three continuous variables: *AssessTot, BuiltFAR,* and *YearBuilt*. Results are summarized in Table 2. *AssessTot* and *BuiltFAR* showed statistically significant but very weak correlations with façade hue. No significant relationship was found for *YearBuilt*.

Variable	**Pearson r**	**p-value**
AssessTot	0.018	0.0031
BuiltFAR	–0.044	<0.001
YearBuilt	–0.002	0.7614

Table 2. Pearson Correlation Results.

3. Hue Variation Across Categorical Variables: One-way ANOVA tests revealed significant differences in mean_hue across all three categorical attributes: *BldgClass, LandUse, and ZoneDist* (Table 3). Among them, *LandUse* showed the largest effect size. Visualizations further confirmed systematic hue differences between residential, commercial, and industrial land uses.

Variable	**F-statistic**	**p-value**
BldgClass	6.478	<0.001
LandUse	46.203	<0.001
ZoneDist	15.029	<0.001

Table 3. ANOVA Summary.

DISCUSSION

This study set out to investigate whether building façade hue, as extracted from street view imagery, reflects underlying urban parcel characteristics. While the results do not indicate strong predictive power, they reveal subtle but consistent relationships between visual appearance and select urban variables.

The correlation analysis showed that assessed property value (*AssessTot*) had a small but statistically significant positive correlation with *mean_hue*. This suggests that properties with higher assessed value may be associated with slightly warmer or more saturated façade colors. One possible explanation is that higher-value properties are more likely to undergo regular renovation or aesthetic upgrades, which can affect façade coloration. Conversely, built floor area ratio (*BuiltFAR*) was negatively correlated with hue, implying that denser or more vertically developed parcels tend toward cooler tones. This may reflect material or regulatory constraints in high-density developments, where glass and concrete dominate the palette.

No meaningful relationship was observed between building age (*YearBuilt*) and hue. This finding suggests that façade color is not simply a function of building vintage and may be more influenced by maintenance cycles, owner preferences, or neighborhood aesthetics.

The ANOVA results provided stronger evidence for categorical distinctions. Façade hue varied significantly across land use types, zoning designations, and building classifications. In particular, *LandUse* had the largest F-statistic, indicating that programmatic function is a key factor shaping the visual expression of buildings. Commercial and mixed-use parcels exhibited warmer hues on average, while industrial and open space zones trended cooler. These patterns may be linked to branding, signage, or institutional design standards. Similarly, zoning codes, which regulate allowable use and built form, appear to have indirect effects on visual aesthetics.

Together, these findings support the hypothesis that façade color, while influenced by many factors, is not randomly distributed in the urban environment. Instead, it encodes aspects of land use and property regulation, offering a subtle yet interpretable layer of spatial information. Importantly, this study demonstrates that combining deep learning–based visual segmentation with parcel-level urban data allows for scalable, data-driven analysis of architectural appearance at the city scale.

LIMITATIONS AND FUTURE WORK

While this study offers a new framework for linking building façade hue to urban parcel attributes, several limitations should be acknowledged, pointing toward valuable directions for future research.

First, the analysis is limited to the Bronx borough of New York City. Although the Bronx offers a diverse urban fabric, the findings may not generalize to other contexts with different architectural styles, zoning policies, or demographic compositions. Future studies could expand the geographic scope to include other boroughs or cities, allowing for cross-urban comparison and model generalization.

Second, the study relies solely on attributes available in the PLUTO dataset, such as land use, FAR, assessed value, and year built. These variables describe the physical and regulatory aspects of the built environment but do not include demographic data such as race, income, or education. Integrating census tract–level demographic data in future work could enable a more direct investigation of how façade color reflects socio-spatial inequalities.

Third, although the SegFormer model effectively isolates building regions for color analysis, its classification may still include noise or artifacts, particularly in visually complex street scenes. Future work could incorporate fine-tuned or façade-specific segmentation models, or even multi-class filtering (e.g., walls only), to improve precision in hue extraction.

Additionally, the study focuses exclusively on hue as the primary visual feature. While hue is a valuable and interpretable metric, other color dimensions such as saturation, brightness, or color diversity may offer further insight. Expanding the color analysis to include these metrics, or combining them into composite color typologies, could yield a more nuanced understanding of urban aesthetics.

Finally, while this study uses statistical correlation and simple classification models, future research could adopt more advanced machine learning approaches (e.g., interpretable deep models, spatial regressions) to capture non-linear relationships and spatial dependencies. Incorporating temporal data could also reveal how façade colors evolve over time due to gentrification, policy changes, or redevelopment.

REFERENCES

Abitbol, J. L., & Karsai, M. (2020). *Socioeconomic correlations of urban patterns inferred from aerial images: Interpreting activation maps of convolutional neural networks*. arXiv preprint arXiv:2004.04907. https://arxiv.org/abs/2004.04907

Azarnejad, A. (2017). *Impact of building façades' color on building and urban design* [Doctoral dissertation]. https://www.researchgate.net/publication/317318196

Chen, K., Lin, H., Shyr, O. F., & You, S. (2023). *What are the differences in urban citizens' preferences for the colour of condominium building facades?* Humanities and Social Sciences Communications, 10, Article 224. https://www.nature.com/articles/s41599-023-02372-9

Liu, Y., Zhang, X., Ding, J., Xi, Y., & Li, Y. (2023). *Knowledge-infused contrastive learning for urban imagery-based socioeconomic prediction*. arXiv preprint arXiv:2302.13094. https://arxiv.org/abs/2302.13094

Machicao, J. (2022). *A deep-learning method for the prediction of socio-economic indicators from street-view imagery using a case study from Brazil*. Data Science Journal, 21(1), 1–11. https://datascience.codata.org/articles/1356/

Wang, Z., Shen, M., & Huang, Y. (2024). *Exploring the impact of facade color elements on visual comfort in old residential buildings in Shanghai: Insights from eye-tracking technology*. Buildings, 14(6), 1758. https://www.mdpi.com/2075-5309/14/6/1758

Yong, X., & Zhou, X. (2024). *MuseCL: Predicting urban socioeconomic indicators via multi-semantic contrastive learning*. arXiv preprint arXiv:2407.09523. https://arxiv.org/abs/2407.09523

Zhai, Y., Gong, R., Huo, J., & Fan, B. (2023). *Building façade color distribution, color harmony and diversity in relation to street functions*. ISPRS International Journal of Geo-Information, 12(6), 224. https://www.mdpi.com/2220-9964/12/6/224

Zhang, G., Yi, J., Yuan, J., & Li, Y. (2023). *DAS: Efficient street view image sampling for urban prediction*. https://www.researchgate.net/publication/
Caparol. (n.d.). *Colour and material trends on the facade*.
https://www.caparol.de/en/design/inspiration/facade-design/colour-and-material-trends-on-the-facade
NYC Department of City Planning. (2025). *Primary Land Use Tax Lot Output (PLUTO)*.
Retrieved from https://www.nyc.gov/content/planning/pages/resources/datasets/mappluto-pluto-change#mappluto