Hot City, Heated Calls: Understanding How Urban Features Affect Quality of Life Under Different Heat Conditions Using New York City's 311 and SHAP

1. INTRODUCTION

1.1. Research Background and Objective

This study investigates the impact of extreme heat on urban Quality of Life (QoL) in New York City using 311 service requests as a behavioral proxy. Specifically, we ask: how do environmental, socioeconomic, and urban morphology factors influence 311 report rates during extreme versus normal heat weeks? While we hypothesize that complaint rates will align with literature linking rising temperature to discomfort, our primary objective is to reveal the drivers of this behavior. To do this, we model weekly complaint rates using Ordinary Least Squares (OLS) and Random Forest (RF) regression, utilizing SHapley Additive exPlanations (SHAP) values to interpret the non-linear influence of these factors under different heat regimes.

1.2. Research Gap

While literature links rising temperatures to 311 requests (Harlan et al., 2006; Hsu et al., 2021) and identifies urban drivers of call patterns (Uejio et al., 2010), few studies examine how these drivers shift during extreme heat thresholds. Also, traditional OLS models fail to capture non-linear interactions, whereas predictive machine learning (ML) models often lack interpretability (Kontokosta & Tull, 2017). This study addresses these gaps by using SHAP to interpret ML models, revealing how socioeconomic and environmental drivers distinctively perform under extreme versus normal heat regimes (Lundberg & Lee, 2017).

2. DATA AND METHODS

2.1. Study Area and Period

The study area is based in New York City with spatial resolution at the census tract level, with these observations during summer 2025, defined as the beginning of June through August 23rd at a weekly temporal resolution. The last week of August was not used due to recent weather data only recorded up to the 24th, therefore not providing a whole week, so it was removed from the study.

2.2. Data Preparation

Heat Data

The subsequent removal of August's last week provided a total of 12 weeks in summer 2025, where extreme heat weeks were defined as at least two extreme heat days within a week with a temperature cutoff threshold at 93°F using the John F. Kennedy (JFK) weather station located at Philadelphia's international airport. This threshold was determined according to a climatological baseline from 1981 through 2010 daily max temperature with a 95th percentile, and this split the observations into two needed regimes: 17 extreme heat days and 71 normal heat days, providing 5 extreme heat weeks and 7 normal heat weeks. Data was directly downloaded from the National Oceanic and Atmospheric Administration (NOAA).

Socioeconomic Data

Socioeconomic data was derived from the United States Census, specifically the most recent 5-year American Community Survey (ACS) in 2023. Python's pyCensus module provided easy access to filter the data down to main investigative, derived variables included percent bachelor’s or more, percent renters, percent limited English, median household income, poverty rate, and percent non-white.

Urban Environmental Data

Environmental urban data were derived from Landsat & LULC raster calculations and OSM water data, specifically scenes within the same study timeline, with the manipulation and computation done through

ArcGIS Pro and Python. However, land-cover land-use (LULC) data was a static raster from 2024. These data included percent tree canopy, percent impervious surface, average building height (AH), building density (BD), normalized difference vegetation index (NDVI), water coverage ratio (WCR).

Urban Built and Spatial Data

Building data came from NYC open data of building footprint shp file with height field. Spatial data included deriving spatial features from Python's osmnx module to calculate points-of-interest (POI) density utilizing a 500-meter buffer and mean Euclidean distance to the nearest subway of census tract centroids.

POIs were determined as everyday main amenities, shops, leisure, and public transport categories in OpenStreet Maps (OSM) yielding 21,309 points, and are as follows: library, community_centre, social_facility, bus_station, bar, restaurant, fast_food, toilets, hospital, clinic, pharmacy, convenience, supermarket, alcohol, deli, park, station.

While many of these environmental and urban forms may be multicollinear, the goal is striving for interpretation rather than maximizing prediction, and this helps to explore the different properties of a city.

2.3. OLS Regression Model

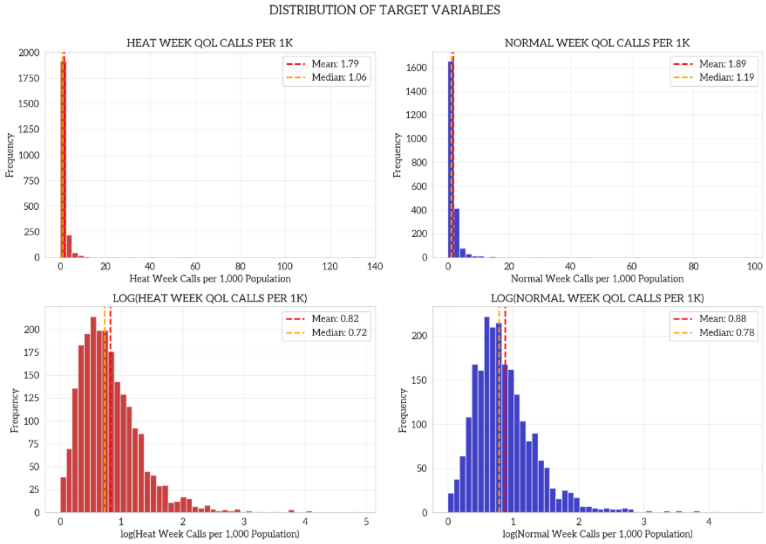
OLS regression served as an interpretable baseline to model linear associations between urban features and QoL complaints. We estimated separate cross-sectional models for extreme and normal heat weeks (N = 2,225 tract-weeks). Features were introduced in three blocks: Urban Environmental (NDVI, percent tree canopy, percent impervious surface, and WCR), Socioeconomic (median income, poverty rate, percent renters, percent limited English, percent bachelor's or more, and percent non-white), and Built / Spatial (AH, BD, distance to the nearest subway station, and 500-meter buffer POI density). While OLS allows for transparent coefficient comparison across heat regimes, we anticipated low R^2 values given the noisy, behavioral nature of 311 data, motivating the use of non-linear ML models.

2.4. ML Model and SHAP

To complement the OLS framework, we used a non-linear RF model to capture the complex, threshold-based interactions in heat stress. RF was selected for its robustness against multicollinearity and stability on moderate-sized datasets. Models were trained for both heat regimes using an 80 / 20 train-test split and three-fold Grid Search cross-validation. To interpret the results, we used SHAP, which simplifies predictions into additive contributions, allowing us to quantify global feature importance and visualize non-linear relationships across extreme versus normal heat conditions.

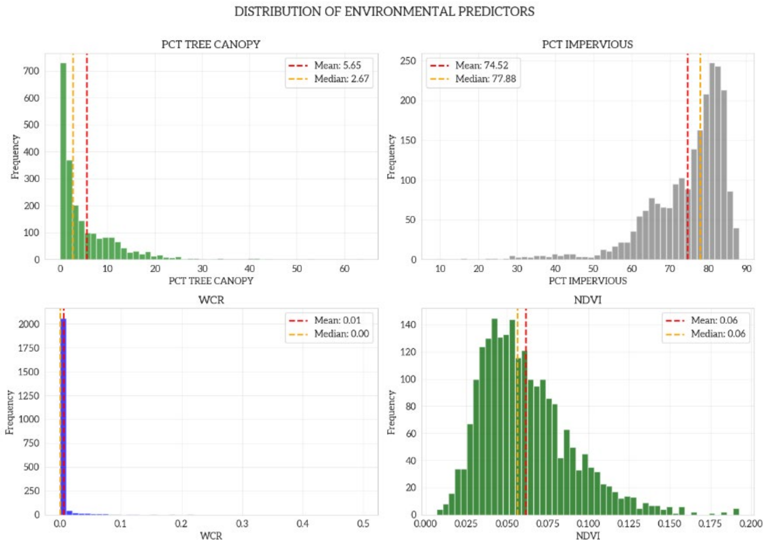
3. RESULTS

3.1. Exploratory Data Analysis

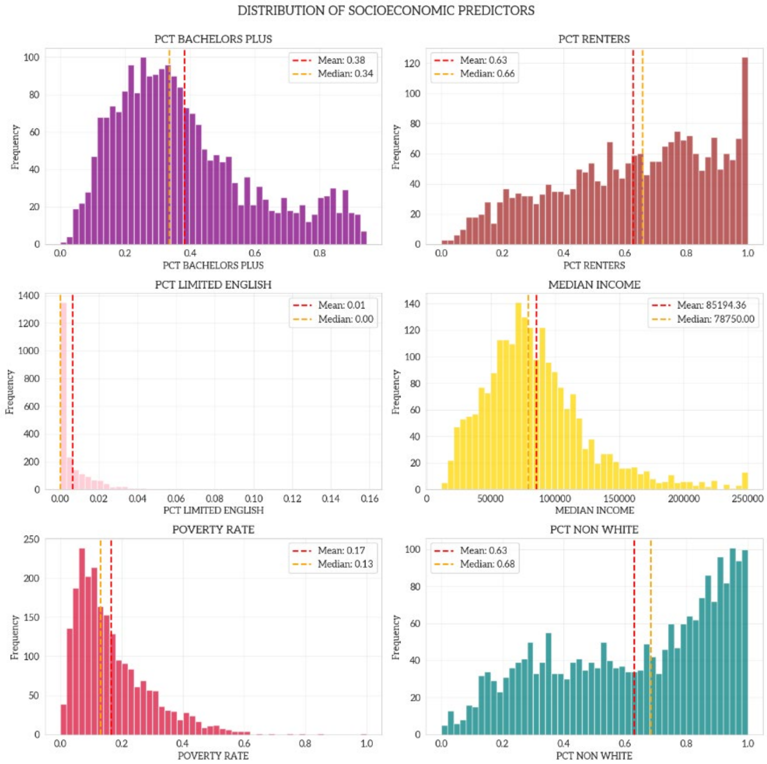


Target Variable Histogram

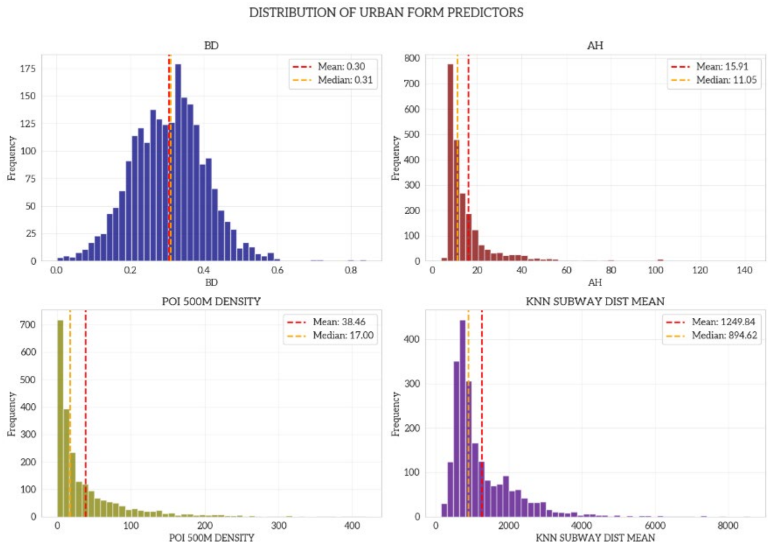
The great overlap between distributions suggests most tracts maintain relatively stable complaint rates regardless of heat conditions. However, some tracts do show meaningful elevation during heat weeks, highlighting importance of identifying characteristics that predict heat sensitivity rather than assuming universal heat response.



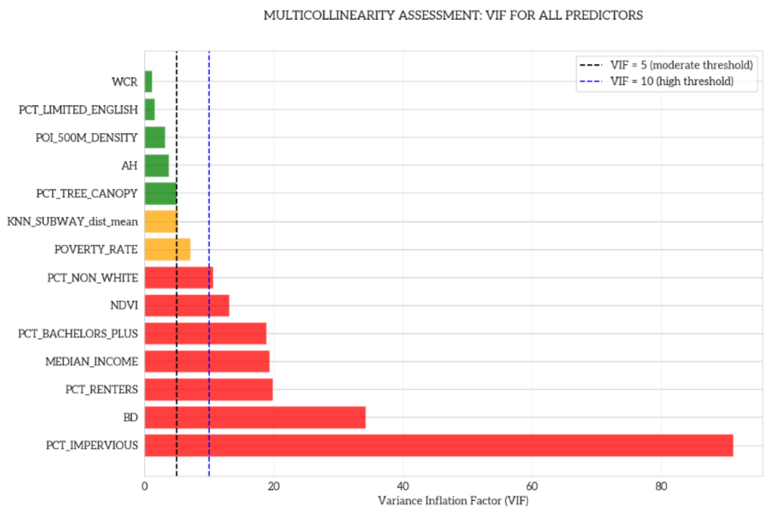
Environmental Predictors Histogram



Socioeconomic Predictors Histogram

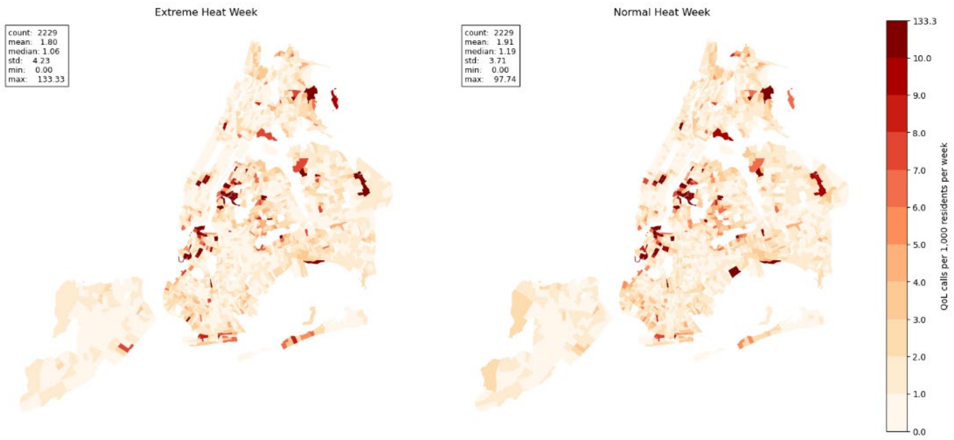


Urban Form Predictors Histogram

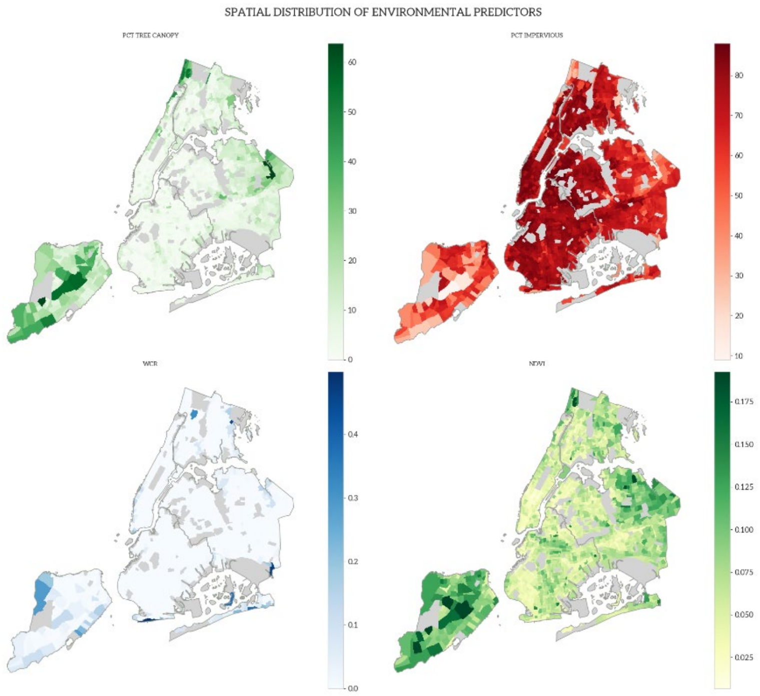


VIF Horizontal Bar Plot

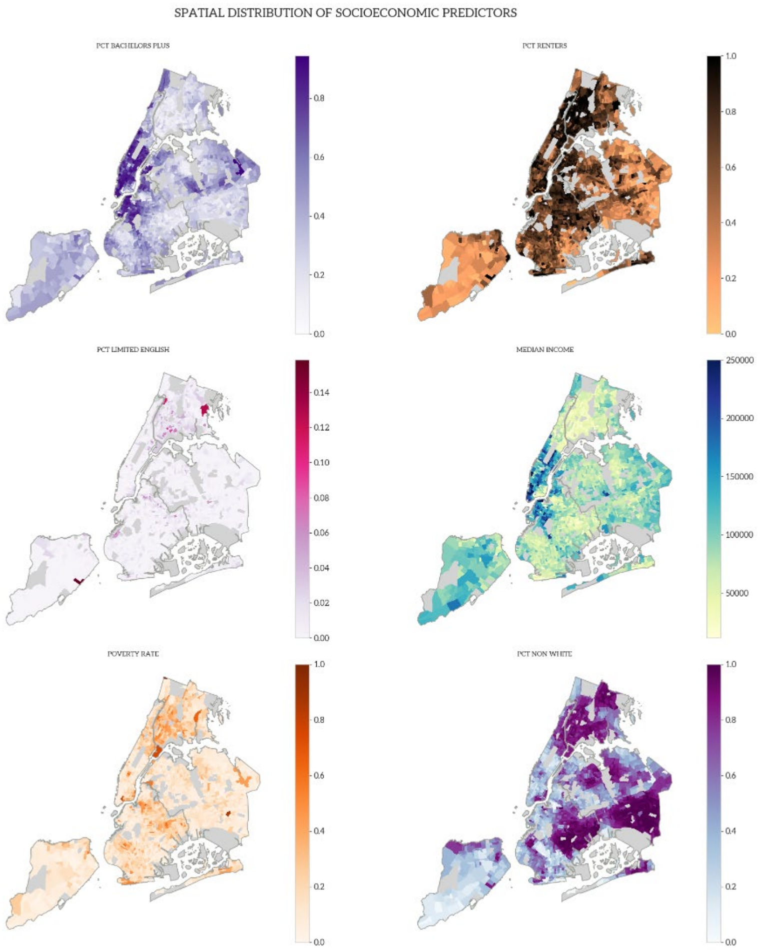
There are seven variables critical VIC violations at values over ten, two variables with moderate VIF violations at values between five and ten, and five variables with acceptable VIF values below five. With this, it seems the variables cluster strongly and reveal that socioeconomic and environmental inequality is not composed of independent factors, but rather tightly bundled mechanisms that drive 311 reporting behavior. This means that OLS, while providing an interpretable baseline, is not ideal for capturing all these variables, reinforcing the need for models robust against multicollinearity like RF.



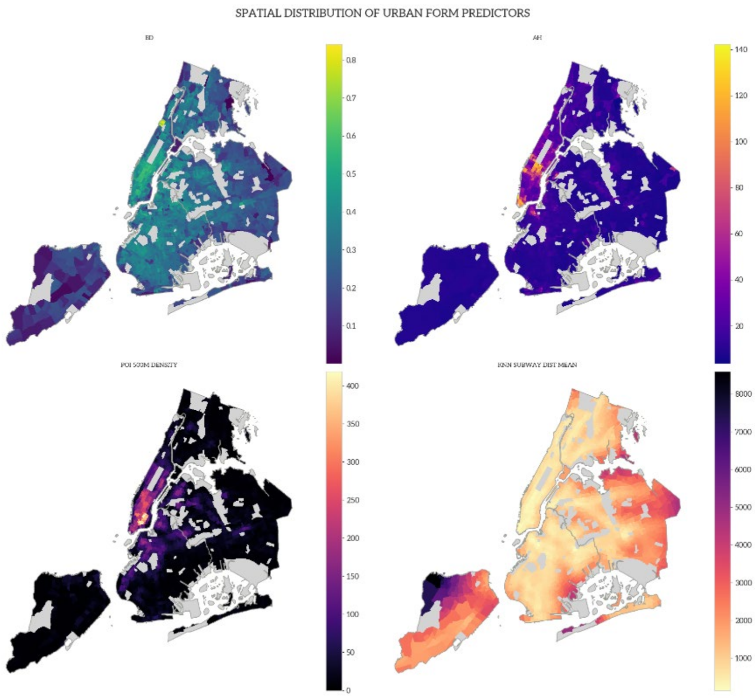
Target Variable Map



Environmental Predictors Map



Socioeconomic Predictors Map



Urban Form Predictors Map

The outliers significantly wash out the other tracts in NYC, many 311 super users are in the Queens Borough in the Long Island City area right around LaGuardia Community College. This general area looks like it has a higher non-white and higher poverty rate. Most notably, the percent change map lacks spatial autocorrelation, but visually the impact must be influenced by some scattered tracts with low baseline activity that make it look fragmented. This could indicate some kind of stress that's unique to the tracts and their mechanism shifts.

3.2. OLS Model Results

3.2.1 Normal Heat Model

```
                              OLS Regression Results
==================================================================================
Dep. Variable:     normalweek_calls_per_1k   R-squared:                       0.084
Model:                                 OLS   Adj. R-squared:                  0.078
Method:                      Least Squares   F-statistic:                     14.22
Date:                     Wed, 03 Dec 2025   Prob (F-statistic):           3.78e-33
Time:                             20:44:47   Log-Likelihood:                -1379.4
No. Observations:                     2192   AIC:                             2789.
Df Residuals:                         2177   BIC:                             2874.
Df Model:                               14
Covariance Type:                 nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
const                    0.8620      0.010     88.589      0.000       0.843       0.881
PCT_TREE_CANOPY          0.0532      0.026      2.033      0.042       0.002       0.105
PCT_IMPERVIOUS           0.1464      0.032      4.565      0.000       0.084       0.209
WCR                      0.0254      0.012      2.104      0.035       0.002       0.049
NDVI                    -0.0681      0.016     -4.347      0.000      -0.099      -0.037
PCT_BACHELORS_PLUS      -0.0105      0.021     -0.509      0.611      -0.051       0.030
PCT_RENTERS             -0.0087      0.017     -0.521      0.602      -0.041       0.024
PCT_LIMITED_ENGLISH     -0.0006      0.011     -0.056      0.956      -0.022       0.021
MEDIAN_INCOME            0.0309      0.019      1.638      0.102      -0.006       0.068
POVERTY_RATE            -0.0694      0.015     -4.563      0.000      -0.099      -0.040
PCT_NON_WHITE           -0.0305      0.013     -2.310      0.021      -0.056      -0.005
BD                      -0.0671      0.017     -3.835      0.000      -0.101      -0.033
AH                      -0.0529      0.013     -4.101      0.000      -0.078      -0.028
POI_500M_DENSITY        -0.0256      0.014     -1.819      0.069      -0.053       0.002
KNN_SUBWAY_dist_mean    -0.0225      0.013     -1.684      0.092      -0.049       0.004
==============================================================================
Omnibus:                      683.326   Durbin-Watson:                   1.591
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             2766.578
Skew:                           1.470   Prob(JB):                         0.00
Kurtosis:                       7.652   Cond. No.                         8.67
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

OLS Normal Heat Week Results

The normal heat OLS model was statistically significant but had a low R^2 of 0.084, explaining approximately 8% of spatial variation. This aligns with the complexity of 311 behavior and suggests the presence of non-linear relationships better suited for machine learning. Significant linear predictors with $p < 0.05$ included PCT_TREE_CANOPY, PCT_IMPERVIOUS, NDVI, POVERTY_RATE, PCT_NON_WHITE, BD, and AH.

3.2.2 Extreme Heat Model

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     heatweek_calls_per_1k   R-squared:                       0.088
Model:                               OLS   Adj. R-squared:                  0.082
Method:                    Least Squares   F-statistic:                     14.94
Date:                   Wed, 03 Dec 2025   Prob (F-statistic):           4.79e-35
Time:                           20:44:21   Log-Likelihood:                -1453.7
No. Observations:                   2192   AIC:                             2937.
Df Residuals:                       2177   BIC:                             3023.
Df Model:                             14
Covariance Type:               nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
const                    0.8027      0.010     79.747      0.000       0.783       0.822
PCT_TREE_CANOPY          0.0240      0.027      0.885      0.376      -0.029       0.077
PCT_IMPERVIOUS           0.1241      0.033      3.740      0.000       0.059       0.189
WCR                      0.0244      0.012      1.957      0.051   -5.37e-05       0.049
NDVI                    -0.0801      0.016     -4.943      0.000      -0.112      -0.048
PCT_BACHELORS_PLUS      -0.0154      0.021     -0.720      0.471      -0.057       0.027
PCT_RENTERS             -0.0202      0.017     -1.172      0.242      -0.054       0.014
PCT_LIMITED_ENGLISH      0.0009      0.011      0.082      0.935      -0.021       0.023
MEDIAN_INCOME            0.0299      0.020      1.531      0.126      -0.008       0.068
POVERTY_RATE            -0.0699      0.016     -4.443      0.000      -0.101      -0.039
PCT_NON_WHITE           -0.0399      0.014     -2.922      0.004      -0.067      -0.013
BD                      -0.0950      0.018     -5.250      0.000      -0.130      -0.060
AH                      -0.0457      0.013     -3.426      0.001      -0.072      -0.020
POI_500M_DENSITY        -0.0175      0.015     -1.201      0.230      -0.046       0.011
KNN_SUBWAY_dist_mean    -0.0104      0.014     -0.748      0.454      -0.038       0.017
==============================================================================
Omnibus:                      690.899   Durbin-Watson:                   1.582
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             2744.943
Skew:                           1.495   Prob(JB):                         0.00
Kurtosis:                       7.594   Cond. No.                         8.67
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

OLS Extreme Heat Week Results

The extreme heat OLS model was highly significant but showed limited explanatory power at $R^2 = 0.088$, suggesting that reporting behavior is likely driven by non-linear structures. Six features had statistically significant linear associations ($p<0.05$), which were PCT_IMPERVIOUS, NDVI, POVERTY_RATE, PCT_NON_WHITE, BD, and AH.

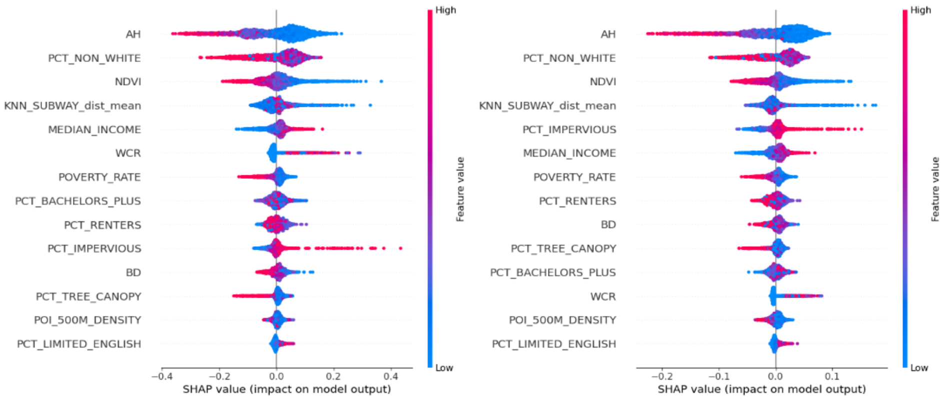
3.2.3 OLS Comparison

Across both the extreme-heat-week and normal-heat-week OLS models, only about half the urban features exhibit statistically significant linear associations with QoL 311 report density, and both models possessed low explanatory power ($R^2 < 0.10$). This consistency indicates that the linear models capture only a small portion of the urban environmental, socioeconomic, and built indicators driving 311 reporting. Taken together, these results reinforce the limitations of linear OLS for this problem. They emphasize the need for ML approaches that can better capture non-linear effects, enabling a more precise comparison of how urban features influence 311 reporting differently under extreme heat versus normal heat conditions.

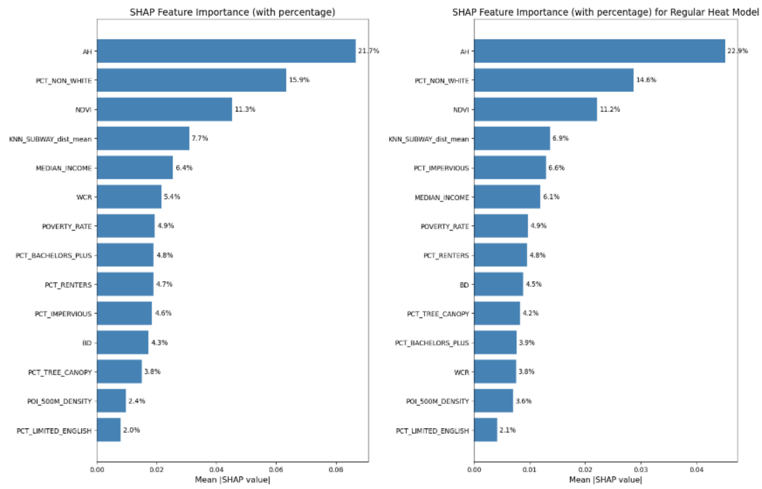
3.3. ML and SHAP Results

3.3.1 ML Model Result

RF substantially outperformed the OLS baseline, validating the importance of non-linear effects. The regular heat model achieved a test R^2 of 0.274 versus OLS' 0.08, while the extreme heat model reached 0.2461. This slight performance drop suggests extreme heat behaviors are more variable. SHAP analysis revealed consistent dominant predictors across both regimes: average building height, percent non-white, and NDVI, as well as revealing complex non-linear contributions from secondary predictors like subway accessibility and income.



SHAP Importance Beeswarm Plots



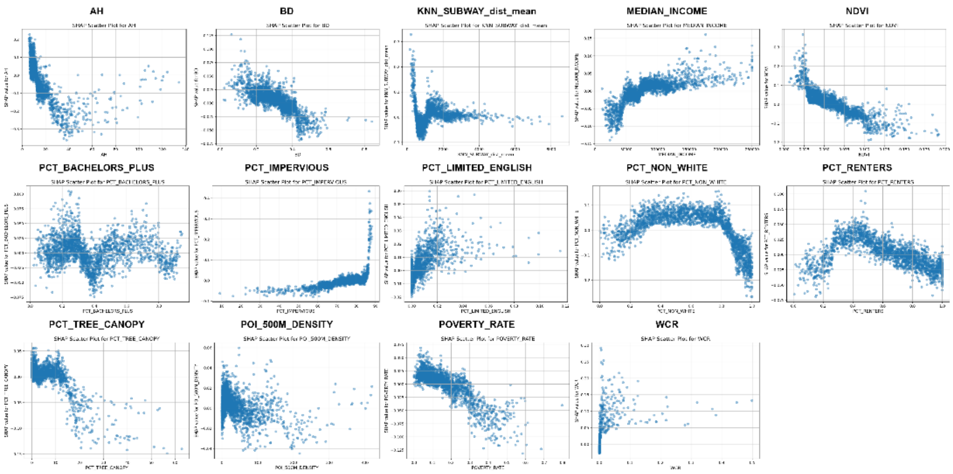
SHAP Importance Bar Charts

3.3.2 Extreme Heat vs Normal Heat

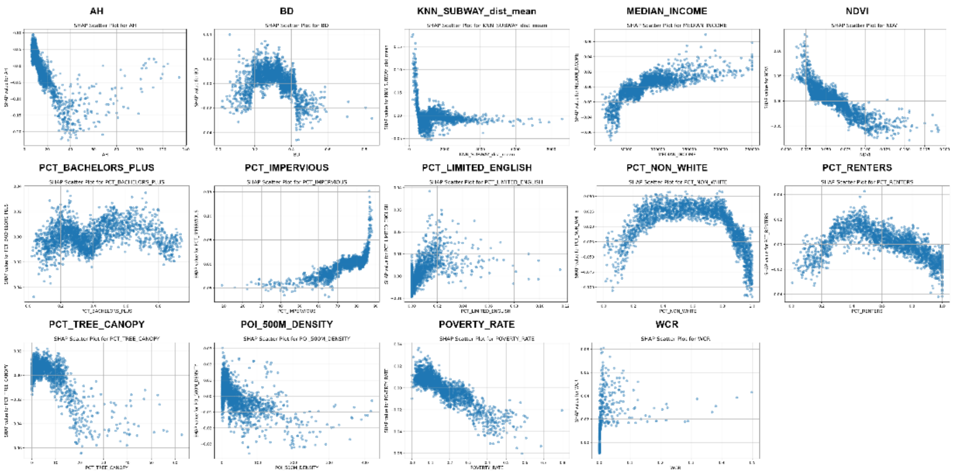
Comparing heat conditions reveals stability in the main drivers but shifts in secondary drivers. The top four predictors—AH, PCT_NON_WHITE, NDVI, and subway distance—remained consistent, reflecting persistent structural influences. However, WCR gained influence during extreme heat, likely due to the the increased value of cooling, water-based amenities. On the other hand, physical morphology metrics like imperviousness and building density have decreased importance, suggesting these factors matter less once critical heat thresholds are crossed.

3.3.3 Non-Linear Relationships

SHAP scatter plots revealed clear non-linear shapes for most features, with only NDVI and POVERTY_RATE showing linear trends. The top predictor, average building height, had a persistent U-shaped relationship, where medium heights correlated with the lowest complaint rates. On the other hand, PCT_NON_WHITE and PCT_RENTERS followed an inverted-U pattern. It should also be noted that building density shifted behavior as it resembled an inverted-U in normal heat but transitioned to a negative linear relationship under extreme heat.



Extreme Heat Week



Normal Heat Week

4. DISCUSSION

4.1 Result Interpretation & Discussion

Overall, the ML model outperformed OLS, although the slightly lower performance during extreme heat suggests added complexity in reporting behavior, and the top predictors (AH, PCT_NON_WHITE, NDVI, subway distance) remained consistent across regimes. However, water coverage ratio gained importance during extreme heat, while physical morphology features (imperviousness and building density) declined, which suggests certain built environment effects taper off at high temperatures.

SHAP analysis also revealed dominant non-linear mechanisms where NDVI and poverty rate showed linear negative trends, reflecting vegetation cooling and reporting bias in lower-income areas between the two. Then average building height showed a consistent U-shaped relationship, where medium-height buildings appeared to optimize shading and ventilation better than exposed low-rises or street-canyon high-rises. Building density also shifted from an inverted-U during normal heat to a negative linear trend during extreme heat, suggesting that density produces better adaptation or indoor retreat when heat stress peaks.

4.2 Limitations

This study has several limitations. First, the analysis focuses on a single summer season in 2025, which may restrict the temporal representativeness of the findings, so extending the study period to multiple years (e.g., 2021–2024) would allow for increasing the reliability and representation of our findings. Second, although 311-based outcomes are inherently difficult to fully explain, the RF R^2 of around 0.25 indicates that there is still substantial unexplained variance. Incorporating additional environmental, socio-economic, and spatial variables could further improve model performance and provide a more complete picture of the drivers of QoL-related 311 report density.

5. REFERENCES

Harlan, S. L., Brazel, A. J., Prashad, L., Stefanov, W. L., & Larsen, L. (2006). Neighborhood microclimates and vulnerability to heat stress. *Social Science & Medicine*, 63(11), 2847–2863.

Hsu, A., Sheriff, G., Chakraborty, T., & Manya, D. (2021). Disproportionate exposure to urban heat island intensity across major US cities. *Nature Communications*, 12(1).

Kontokosta, C. E., & Tull, C. (2017). A data-driven predictive model of city-scale energy use in buildings. *Applied Energy*, 197, 303–317.

Lundberg, S., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. NIPS'17: *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 9781510860964.

Uejio, C. K., Wilhelmi, O. V., Golden, J. S., Mills, D. M., Gulino, S. P., & Samenow, J. P. (2010). Intra-urban societal vulnerability to extreme heat: The role of heat exposure and the built environment, socioeconomics, and neighborhood stability. *Health & Place*, 17(2), 498–507.