

# Exploring the Relationship Between Parking Meter Density and Citation Rates in San Francisco.\*

Ethan Kusnadi

December 10, 2025

This project investigates how parking meter density is associated with citation rates in San Francisco using data published by the SFMTA. We fit a multiple linear regression model that includes meter density, cap color, and supervisor district as predictors. This model shows a slight positive association between meter density and citation rates, but it has limited explanatory power, suggesting that other factors likely play a larger role in determining citation activity. Future work could incorporate more predictors or explore alternative approaches such as weighted least squares.

## 1 Introduction

Parking meters are an important part of American city infrastructure, serving as a form of regulation to manage street parking. Because this system relies on drivers paying for their allotted time at a meter, rules are only enforced when parking monitors patrol a given area. As a result, inconsistent enforcement or differences in neighborhood traffic patterns can cause citation rates to vary across a city. This raises the question of which characteristics influence citation activity.

Previous studies have examined illegal parking behavior by relating drivers' dwell times to the probability of receiving a citation (Nourinejad, Gandomi, and Roorda 2020). Other work has explored how parking enforcement varies across neighborhoods, finding that citation rates tend to be higher in economically vulnerable and historically marginalized areas (Brazil, Vang, and Abdelnur 2024). However, there has been little research on how the design of the city itself, such as the density of parking meters, is associated with the citation rate. In this project,

---

\*Project repository available at: <https://github.com/ekusnadi/Math-261A-Project-2>.

we investigate the question: How is parking meter density associated with citation rates per meter in San Francisco, after accounting for meter rate and supervisor district?

Parking meter and citation data from the San Francisco Municipal Transportation Agency (SFMTA) were used to fit a multiple regression model to predict the log-transformed citation rate per meter using R (San Francisco Municipal Transportation Agency 2025a, 2025b). The analysis suggests a slight positive relationship between parking meter density and citation rate, but the model explains only a small percentage of the variance in the response, and other factors may contribute more to citation rates.

The remainder of this paper is structured as follows: Section 2 introduces the dataset and describes the data cleaning process, Section 3 explains the methods and models used in the analysis, Section 4 presents the results, and Section 5 discusses possible further explorations.

## 2 Data

The data used for this project are sourced from the DataSF Open Data Portal and are provided by the San Francisco Municipal Transportation Agency (SFMTA), a government agency. The raw data can be downloaded from the online portal as CSV files.

The Parking Meters dataset is created by the SFMTA and maintained through the SFpark data warehouse (Parking Space Inventory) (San Francisco Municipal Transportation Agency 2025a). This dataset contains the locations and key attributes of parking meters in San Francisco, with the most recent public update released in November 2025. Meter shop staff update the dataset periodically to reflect new regulations and operational changes. We filtered for active and temporarily inactive meters only, leaving 30,678 meters in the dataset. The variables used in this analysis include:

- **BLOCKFACE\_ID:** the ID of a blockface, which is the side of a street where a meter is located.
- **ACTIVE\_METER\_FLAG:** indicates whether the meter is active.
- **CAP\_COLOR:** meter rate zone, coded by color:
  - Black – Motorcycle parking
  - Brown – Tour bus parking
  - Green – Short-term parking
  - Grey – General metered parking
  - Red – Six-wheeled commercial vehicle parking
  - Yellow – Commercial vehicle parking
- **STREET\_NAME:** the name of the street.
- **STREET\_NUM:** the approximate street number of the meter.
- **supervisor\_district:** the City Supervisor district in which the meter is located, with 11 districts in total.

The second dataset is the SFMTA – Parking Citations & Fines dataset, which is updated daily and includes violation, location, and vehicle information (San Francisco Municipal Transportation Agency 2025b). For this project, we restrict the data to 2025 parking meter violations, corresponding to violation codes TRC7.2.23A, TRC7.2.23B, TRC7.2.30C, and TRC7.2.30D. As of December 6, 2025, the filtered dataset contains 224,550 parking meter citations. For this analysis, we use the **Citation Location** field, which provides the street address where the citation was issued.

The observational unit in this analysis is the blockface. Because there is human error involved in recording individual parking meters and street addresses, we aggregate meters by blockface and record how many citations were associated with meters on that blockface. Using a Python script, we constructed a blockface-level dataset by cleaning and merging the parking meter and citation data (Python Software Foundation 2025). Citation street addresses are parsed into street numbers and names. A citation is assigned to a blockface if its street number falls within the meter-defined number range for that blockface and matches the street-number parity. We then count the number of citations corresponding to each blockface and calculate the citation rate per meter. For attributes such as cap color and supervisor district, we take the class majority as the blockface attribute.

After merging, 163,754 citations (72.9%) are matched to a blockface. Unmatched citations may correspond to new or recently deactivated meters not yet updated in the Parking Meters dataset. They may also come from meters having inaccurate or missing street-number records or from citation locations with data entry errors. These unmatched observations may introduce bias if citation recording errors are correlated with blockface characteristics.

In total, the final blockface-level dataset contains 2,926 blockfaces. The dataset includes, for each blockface:

- **meter\_count**: the number of meters on the blockface. Because density is defined at the blockface level, **meter\_count** serves as the measure of meter density and is used interchangeably with it throughout the analysis.
- **citation\_count**: the number of citations whose location matched the blockface's street-number range and parity.
- **citation\_rate\_per\_meter**: **citation\_count** divided by **meter\_count**.
- **log\_rate**:  $\log(\text{citation\_rate\_per\_meter} + 1)$ .
- **cap\_color\_majority**: cap color most common across meters on the blockface.
- **supervisor\_district**: assigned using the majority of meters on the blockface.

The distribution of citation rates per meter is heavily right-skewed, with most blockfaces receiving relatively few citations per meter and a small number of blockfaces receiving much higher rates, seen in Figure 1. This skewness violates the linear regression assumption of approximately normally distributed residuals. To address this, we choose to model the natural log-transformed citation rate instead, using  $\log(\text{citation\_rate\_per\_meter} + 1)$  as the response variable, where adding 1 allows blockfaces with zero citations to be included while avoiding

undefined values. This transformation produces a distribution that is much closer to normal, making it more suitable as the response variable in our regression model.

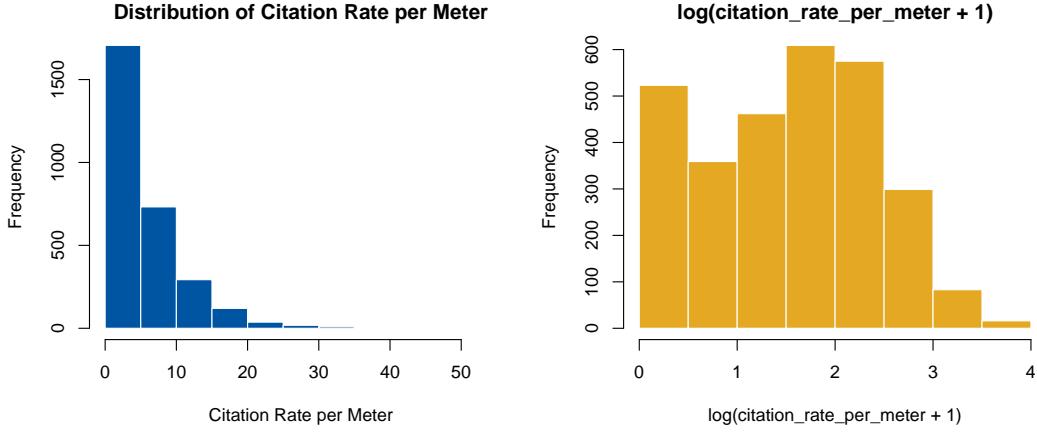


Figure 1: Distributions of Citation Rate per Meter (left) and  $\log(\text{citation\_rate\_per\_meter} + 1)$  (right).

The distribution of meter counts reveals a heavy right-skew, shown in Figure 2. To combat this, we plan to train a model that uses  $\log(\text{meter\_count})$  as a predictor to better capture variation in meter density. In this scenario, we do not need to add 1 to `meter_count` as a blockface with no meters will not be in the dataset.

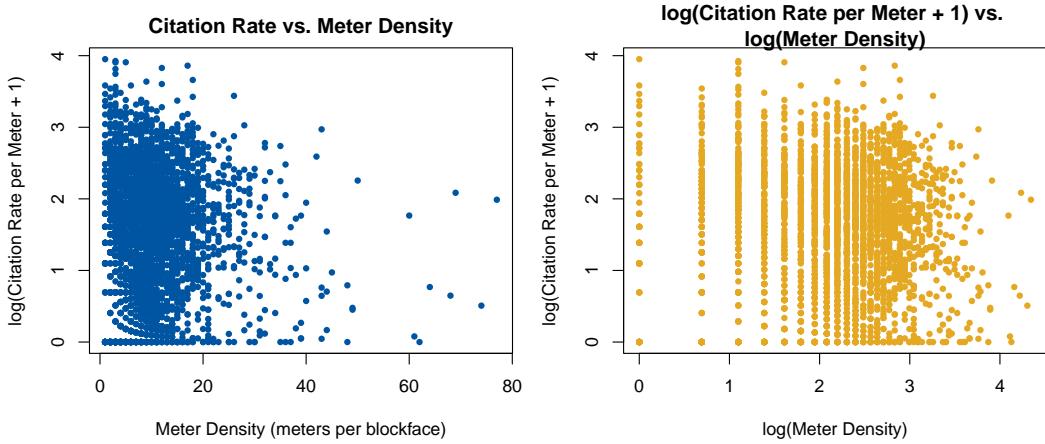


Figure 2: Scatterplots of log citation rate versus meter density on the original (left) and log scale (right).

Figure 3 shows the distribution of the two categorical predictors used in the regression model. Most blockfaces contain meters with a grey cap color, which represents general metered parking. This imbalance implies that coefficient estimates for smaller classes may be less accurate. The

number of metered blockfaces also varies significantly across districts. Because blockfaces are spatially clustered, it is important to account for location-based differences in the analysis.

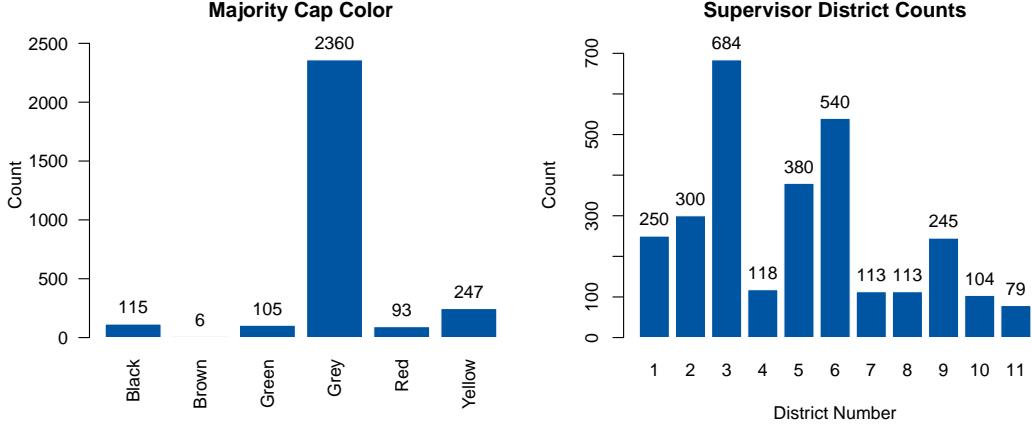


Figure 3: Distributions of Cap Colors Count (left) and Supervisor Districts Count (right).

### 3 Methods

To model the citation rate per meter, we use meter density (meter count) as the main predictor and also include the categorical variables cap color and supervisor district. Meter density is a reasonable predictor because blockfaces with more meters may exhibit higher turnover and potentially more violations, although it is not obvious whether citation rates per meter should increase or decrease as density rises. Cap color represents the meter rate zone, providing information about pricing and typical duration of stay. Supervisor district is included to account for potential location-based differences such as traffic or enforcement patterns.

We perform multiple linear regression, fitting two models with the `lm()` function from the R programming language (R Core Team 2025). The first model we explore is given by the following equation:

$$\begin{aligned} \log(\text{citation\_rate\_per\_meter}_i + 1) = & \beta_0 + \beta_1 \text{meter\_count}_i \\ & + \sum_{j=1}^5 \gamma_j C_{ij} + \sum_{k=1}^{10} \delta_k D_{ik} + \varepsilon_i, \\ & \text{for } i = 1, \dots, n. \end{aligned}$$

In this model, the predictors are meter density and the categorical variables cap color  $C$  and supervisor district  $D$ . The intercept  $\beta_0$  corresponds to the expected value of  $\log(\text{citation\_rate\_per\_meter} + 1)$  for a blockface in the reference categories (Black cap color

and District 1) when the meter density is equal to 0. Because blockfaces with zero meters are not meaningful in this context, the intercept is not interpretable but serves as a baseline to measure the effects of the predictors. The parameter  $\beta_1$  represents the expected change in  $\log(\text{citation\_rate\_per\_meter} + 1)$  associated with one additional meter on a blockface, holding the cap color and district constant. For the categorical predictors, the coefficients  $\gamma_j$  and  $\delta_k$  represent the expected differences in  $\log(\text{citation\_rate\_per\_meter} + 1)$  between each cap color or district and their reference groups, holding all other variables constant. The random error term  $\varepsilon_i$  captures the unexplained variability in the response.

The second model we consider is given by the following equation:

$$\begin{aligned}\log(\text{citation\_rate\_per\_meter}_i + 1) = & \beta_0 + \beta_1 \log(\text{meter\_count}_i) \\ & + \sum_{j=1}^5 \gamma_j C_{ij} + \sum_{k=1}^{10} \delta_k D_{ik} + \varepsilon_i,\end{aligned}$$

for  $i = 1, \dots, n$ .

Again, we include the categorical variables cap color  $C$  and supervisor district  $D$ , whose coefficients have the same interpretation as in the first model. In this model, we use  $\log(\text{meter\_count})$  as the main predictor to address the strong right-skew in meter density, creating a log-log model. The intercept  $\beta_0$  now corresponds to the expected value of  $\log(\text{citation\_rate\_per\_meter} + 1)$  for a blockface in the reference categories when  $\text{meter\_count} = 1$  (since  $\log(1) = 0$ ), providing a meaningful baseline value for the model. The parameter  $\beta_1$  represents the expected percent change in the citation rate per meter associated with a 1% increase in the meter density, holding cap color and supervisor district constant. As before, the random error term  $\varepsilon_i$  captures the unexplained variability in the response.

The regression model assumes linearity, meaning that the response is a linear function of the predictors. After the log transformation of the citation rate per meter, the linearity assumption seems to be reasonably satisfied.

The linear regression model also assumes that the error terms are independent, have constant variance (homoscedasticity), have a mean of zero, and are approximately normally distributed. The independence assumption states that the observations should be unrelated. In this problem, nearby blockfaces may be spatially correlated, which could lead to correlated errors and represents a possible limitation. Because the linear regression model is estimated using the ordinary least squares method to minimize the sum of squared residuals, the residuals produced have a sample mean of zero by construction, satisfying the mean-of-zero assumption. The homoscedasticity and normality assumptions will be evaluated after fitting the model using diagnostic plots.

## 4 Results

After training the models, we evaluate the results. For Model 1, using meter count on a blockface as the main predictor, the estimated coefficients are  $\hat{\beta}_0 = 1.0630$  and  $\hat{\beta}_1 = 0.01045$ . This means that for each additional meter on the blockface,  $\log(\text{citation rate per meter} + 1)$  is expected to increase by 0.0104, holding cap color and district constant. A 0.0104 increase on the log scale corresponds to about a 1.05% increase in  $(\text{citation rate per meter} + 1)$ , indicating a small but positive association between meter density and citation rates. The  $R^2$  value of this model is low at 0.1302, which explains about 13% of the variation in the response, meaning that most of the variability is unexplained by this model and other factors are involved.

For Model 2, which uses  $\log(\text{meter\_count})$  as the main predictor, the estimated coefficients are  $\hat{\beta}_0 = 0.93374$  and  $\hat{\beta}_1 = 0.13392$ . The intercept  $\hat{\beta}_0$  implies that, for the baseline cap color and supervisor district, a blockface with exactly one meter is expected to have a citation rate per meter of approximately 1.54 on the original scale. Holding cap color and district constant,  $\hat{\beta}_1$  implies that a 1% increase in the meter count is associated with about a 0.13392% increase in  $(\text{citation rate per meter} + 1)$ . The  $R^2$  value of this model is 0.1343, which is slightly higher than Model 1 but does not represent a meaningful improvement in explanatory power.

For each of the models, we separate the data by cap color, then plot the fitted regression line for each cap color while holding the supervisor district fixed, as shown in Figure 4. We do the same for the supervisor districts while holding cap color constant, which produces similar output. For each plot, the fitted lines share a common slope but differ in their intercepts, reflecting the categorical effects in the model. These differences indicate that cap color and supervisor district shift the baseline level of citations across rate zones and locations.

We also plot the fitted values against the residuals for both models in Figure 5. This plot reveals a funnel-shaped pattern with the spread of residuals increasing at higher fitted values. This suggests heteroscedasticity, which violates the equal variance assumption of linear regression. This may be caused by the response being log-transformed from a non-negative quantity, which limits how negative residuals can be.

For both models, the histogram of residuals shows a roughly normal distribution of residuals and the Q-Q plot shows the residuals lying close to the reference line, indicating that the normality assumption for residuals is satisfied (Figure 6).

## 5 Discussion

Model 2 shows a slightly improved fit over Model 1 and provides a more meaningful baseline due to the log-log specification. The log transformation of meter density also gives us a percent-change interpretation for the slope, which is more natural in this context because the effect of adding additional meters differs between low-density and high-density blockfaces. For these reasons, we chose Model 2 as our final model.

Model 1 Fitted Lines by Cap Color

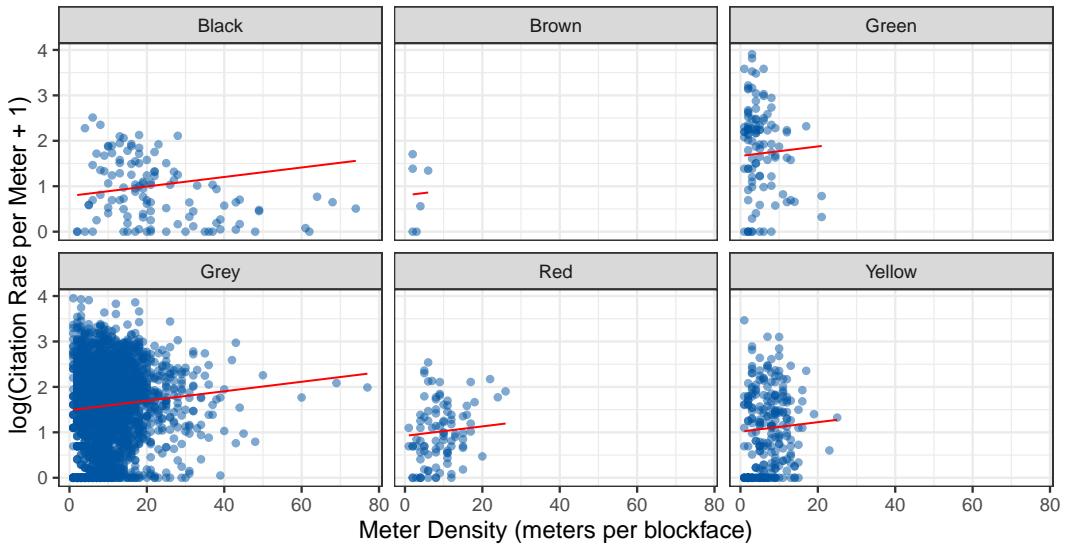


Figure 4: Model 1 - Fitted Lines by Cap Color for Supervisor District 3.

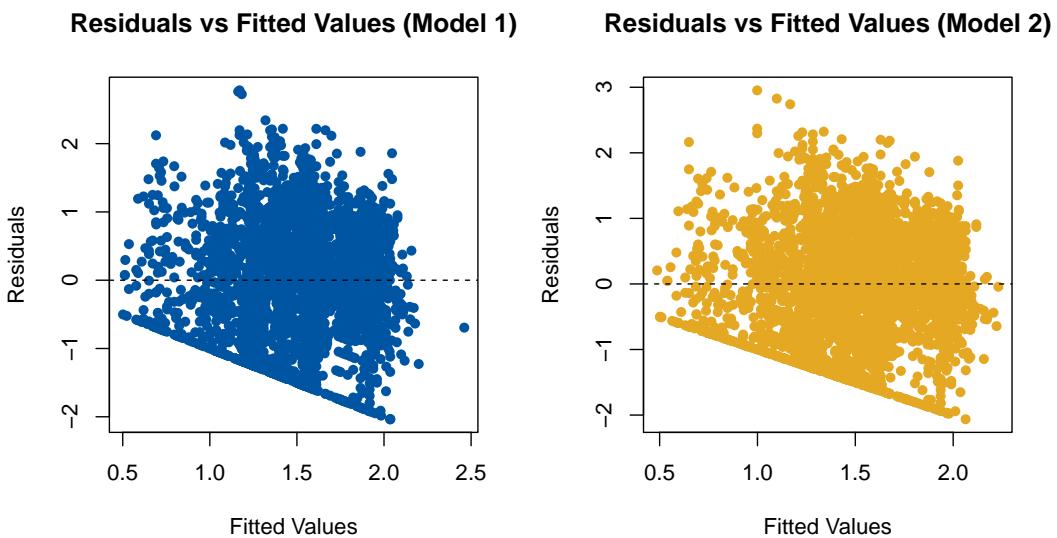


Figure 5: Residuals versus fitted values from both regression models.

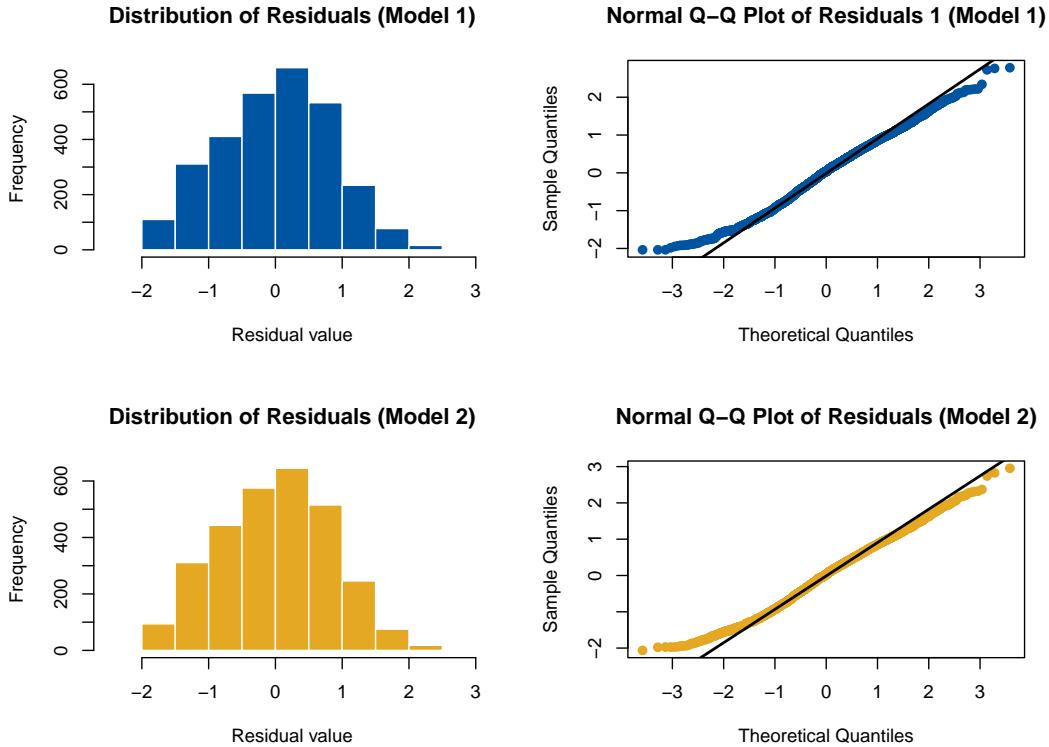


Figure 6: Histogram and Q-Q plot of residuals from both regression models.

The results indicate a small positive association between meter density and citation rate per meter. After controlling for cap color and supervisor district, blockfaces with more meters tend to have slightly higher citation rates per meter. However, the relatively low  $R^2$  values suggest that meter density explains only a small portion of the variation in citation rate, and additional factors not included in this dataset likely play an important role.

The residual diagnostics also reveal some heteroscedasticity, violating the constant variance assumption. While the log transformation slightly improves the model fit, these diagnostics suggest that the linear regression model does not fully capture the underlying relationship in citation rates and meter density.

In future analyses, we could explore using weighted least squares regression to address the heteroscedasticity observed in the residuals. Additional predictors such as patrol activity or neighborhood traffic patterns may also help capture more of the variability in citation rates and improve overall model performance.

## References

- Brazil, Noli, Brittany Vang, and Huda Abdelnur. 2024. “Neighborhood Inequality in Government Fines: The Case of Parking Tickets in 16 US Cities.” *Cities* 152: 105229.
- Nourinejad, Mehdi, Amir Gandomi, and Matthew J Roorda. 2020. “Illegal Parking and Optimal Enforcement Policies with Search Friction.” *Transportation Research Part E: Logistics and Transportation Review* 141: 102026.
- Python Software Foundation. 2025. *Python: A Programming Language*. <https://www.python.org/>.
- R Core Team. 2025. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.r-project.org/>.
- San Francisco Municipal Transportation Agency. 2025a. “Parking Meters.” <https://data.sfgov.org/Transportation/Parking-Meters/8vzz-qzz9>; DataSF.
- . 2025b. “SFMTA Parking Citations & Fines.” <https://data.sfgov.org/Transportation/SFMTA-Parking-Citations-Fines/ab4h-6ztd>; DataSF.