Trip Distance and Duration Patterns in Austin's Shared Micromobility System

Donghyun Kim
University of Michigan

Abstract

Shared micromobility systems have become an important component of urban transportation, yet bicycles and electric scooters may serve distinct roles within these systems. Using trip-level data from Austin's shared micromobility program for the year 2019, this study examines differences in trip distance and trip duration between bicycles and scooters under comparable temporal and spatial conditions. Two descriptive research questions are addressed: whether scooter trips are systematically shorter than bicycle trips after controlling for time and location, and whether trip duration differs by vehicle type once trip distance is taken into account.

After preprocessing more than 5.4 million trips and trimming extreme outliers, log-linear regression models were used to account for strong right skewness and heteroscedasticity in both distance and duration. Regression models controlled for hour of day, day of week, and starting council district. Model selection was guided by the Bayesian Information Criterion (BIC), with robustness checks conducted using alternative specifications including Box–Cox transformations, weighted least squares, and robust regression.

The results show a large and robust difference in trip distance by vehicle type: scooter trips were approximately 40% shorter than bicycle trips under similar temporal and spatial conditions. In contrast, differences in trip duration were minimal once distance was controlled for. Allowing for an interaction between vehicle type and trip distance slightly improved model fit, but predicted durations for scooters and bicycles were nearly identical for typical trip lengths.

These findings suggest that scooters and bicycles play complementary roles within Austin's micromobility system. Scooters are primarily used for short, localized trips, while bicycles support longer travel, yet both modes offer similar travel-time performance for a given distance. Although the analysis is descriptive and not causal, the results have practical implications for fleet composition, rebalancing strategies, and infrastructure planning in cities that support shared micromobility.

Keywords: shared micromobility, linear regression, log transformation

Trip Distance and Duration Patterns in Austin's Shared Micromobility System

Shared micromobility refers to short-term rentals of bicycles, e-bikes, and electric scooters that allow users to pick up and drop off vehicles across an urban area. In the United States and Canada, use of shared micromobility systems has grown rapidly over the past decade. According to the National Association of City Transportation Officials (NACTO), approximately 130 million trips were taken on shared bicycles and scooters in 2022, representing a 40% increase since 2018 and a 35-fold increase since 2010 (NACTO, 2023). More recent reports indicate that ridership continued to increase in 2023, and shared micromobility is now characterized as a popular and growing component of urban transportation systems that often replaces short car trips and extends the reach of public transit (NACTO, 2024).

Austin is one of the U.S. cities that has explicitly incorporated micromobility into its long-term transportation planning. The *Austin Strategic Mobility Plan* identifies a high reliance on single-occupancy vehicle travel and outlines a long-term vision for a more multimodal transportation system that supports walking, bicycling, and transit (City of Austin, 2019). Within this policy framework, shared micromobility is positioned not only as a convenience for individual riders but also as a tool for addressing congestion, reducing emissions, and improving access to dense urban destinations.

For practitioners, the distinction between scooters and bicycles has direct operational implications. Operators must decide what proportion of their fleet should consist of scooters versus bicycles, how pricing and incentives should differ by mode, and where and when to reposition vehicles to meet demand. City transportation agencies and infrastructure planners face related questions. If scooters are primarily used for short, dense urban hops while bicycles are used for longer, commuting-oriented trips, then the optimal design of protected lanes, speed limits, and parking corrals will differ by mode. Conversely, if the two modes exhibit similar distance and duration patterns once time and location are accounted for, this would suggest that scooters and bicycles function as closer substitutes than complements within the overall mobility system.

A key challenge is that simple averages can be misleading. Scooters may appear to have shorter trips than bicycles simply because they are used more often during busy daytime hours or in compact downtown districts, where all trips tend to be shorter. Likewise, bicycles may show longer average durations merely because people ride them more often on weekends or for recreational purposes. To understand the intrinsic roles of each mode, scooters and bicycles must be compared under similar conditions—trips that occur at the same times of day, on the same days of the week, and in similar parts of the city. When studying trip duration, it is also necessary to control for distance, because longer trips naturally take more time regardless of vehicle type.

This project therefore focuses on two related research questions. First, after controlling for time of day, day of week, and council district, are scooter trips systematically shorter than bicycle trips in Austin's shared micromobility system? Second, after controlling for trip distance and the same temporal and spatial factors, do scooters and bicycles differ systematically in trip duration? These questions are descriptive rather than causal: the goal is to summarize conditional patterns in typical trip distance and duration by vehicle type, not to estimate the causal effect of vehicle choice on travel behavior. Nevertheless, such conditional patterns can inform practical decisions about fleet composition, rebalancing strategies, and city-level investments in lanes and parking infrastructure for shared scooters and bicycles.

Method

Data Source and Study Design

Trip-level records from Austin's shared micromobility system obtained from the Austin Open Data Portal ("Shared Micromobility Vehicle Trips") were analyzed. The public dataset contains all trips taken between 2018 and 2022. For the present study, attention was restricted to trips that occurred in calendar year 2019. Using the pre-pandemic year avoids confounding by COVID-19–related disruptions and allows the results to be interpreted as representing a typical baseline year of demand and operations.

The unit of analysis is a single completed trip. Two continuous outcomes were analyzed: trip distance (meters) and trip duration (seconds). The primary exposure of interest is vehicle type (bicycle vs. scooter). Additional covariates included hour of day, day of week, and the council district where the trip started, which capture temporal and spatial variation in demand and infrastructure. The study design is observational and retrospective; all analyses are descriptive and associational rather than causal.

Sample Selection and Preprocessing

Raw trip-level data were imported into R and cleaned prior to analysis. The dataset was restricted to trips recorded in 2019 and retained only bicycle and scooter trips. Observations with missing or non-positive values for trip distance or trip duration were removed. The final analytic dataset included vehicle type, trip distance, trip duration, hour of day, day of week, and starting council district. To ensure appropriate modeling, hour, day of week, and council district were coded as categorical factors so that each level could have its own mean, avoiding implausible linear trends that would arise if these nominal variables were treated as numeric. Accordingly, these predictors were converted to factors: vehicle type (bicycle vs. scooter), hour of day (24 levels), day of week (7 levels), and starting council district (10 levels).

Because the raw distributions of distance and duration were highly right-skewed, outlier trimming was applied using the $3 \times IQR$ rule, a standard criterion for identifying extreme observations in heavy-tailed distributions. The $3 \times$ IQR threshold was used rather than the more common $1.5 \times$ IQR rule to remove only clearly implausible extreme values while retaining the substantial natural variability characteristic of micromobility trip data. Values less than $Q_1 - 3 \times IQR$ or greater than $Q_3 + 3 \times IQR$ were removed, with negative lower bounds truncated at zero because distance and duration cannot be negative. For distance, the quartiles were 611 m and 2,019 m ($IQR = 1{,}408$ m), giving an upper bound of 6,243 m; for duration, the quartiles were 251 s and 749 s ($IQR = 498$ s), giving an upper bound of 2,243 s. After trimming, the analytic dataset contained 5,486,004 trips. Within this dataset, the mean distance was 1,383 m (median 1,078 m), and the mean duration was 526 s (median 400 s).

All subsequent exploratory analysis and modeling used this trimmed dataset. Subsamples of 100,000 trips were used for visualizations to maintain readability, while all inferential models used the full dataset (see Appendix A).

Measures and Exploratory Patterns

Our primary substantive interest lies in differences between bicycle and scooter trips under comparable conditions. Hour of day, day of week, and starting district were included as control variables to account for systematic temporal and spatial variation in travel behavior. Exploratory boxplots showed that both distance and duration varied substantially by hour, day, and district. For example, trips tended to be longer during evening hours and in certain districts. These patterns indicated that comparing bicycles and scooters without adjusting for time and location would conflate mode differences with contextual differences in when and where people ride. Therefore, all regression models included hour, day, and district as categorical controls.

Exploratory histograms of the preprocessed data showed that, although extreme values had been removed and the distributions were somewhat less skewed, both distance and duration remained strongly right-skewed. In addition, residual-versus-fitted plots from an ordinary least squares model fit on the raw scale exhibited clear heteroscedasticity, with residual spread increasing at larger fitted values. These features indicated violations of linear model assumptions on the raw scale. Specific exploratory results are presented in Appendix A.

Modeling Strategy

To address the strong right skew observed in the raw distance and duration measures and to mitigate heteroscedasticity apparent in raw-scale modeling diagnostics, both outcomes were transformed using natural logarithms of trip distance and trip duration. Histograms and boxplots of the transformed outcomes showed reduced skewness and compression of the upper tails, motivating the use of log-scale models for subsequent analyses (see Figure 1).

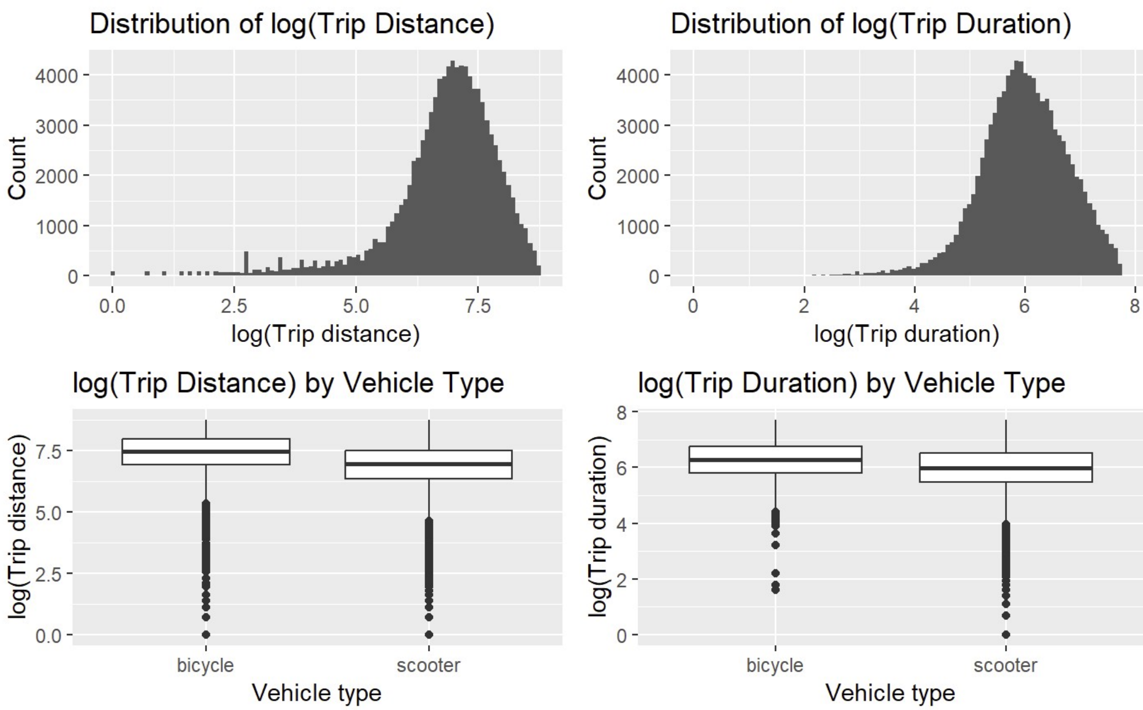

Figure 1
Distributions of log-transformed trip distance and duration, and boxplots by vehicle type.

Primary analyses were conducted using ordinary least squares (OLS) regression on the log-transformed outcomes. To assess the robustness of results to departures from standard linear model assumptions, alternative specifications were additionally considered, including Box–Cox transformations, weighted least squares (WLS), and robust Huber regression. These alternative models were used to evaluate sensitivity to heteroscedasticity and heavy-tailed residuals, as well as to examine whether the estimated effect of vehicle type remained stable across modeling choices.

Model adequacy was evaluated using residual-versus-fitted plots and normal Q–Q plots as qualitative diagnostics for linearity, heteroscedasticity, and departures from normality. Given the extremely large sample size and the inclusion of multiple categorical predictors (hour, day, and district), these diagnostics were interpreted as checks for major violations of modeling assumptions rather than as strict tests of residual normality. Generalized variance inflation factors (GVIFs) were also computed to assess multicollinearity and examined studentized residuals, leverage values, and Cook's distances to identify potentially influential observations.

For both Question 1 and Question 2, model selection among competing specifications was guided by the Bayesian Information Criterion (BIC). Given the extremely large sample size and the inclusion of numerous categorical predictors (hour, day, and district), BIC was used to discourage overfitting and to prioritize more less complex, interpretable models. Compared with AIC, BIC imposes a stronger penalty for model complexity and is particularly appropriate in very large samples. Interaction terms were additionally considered for Question 2 because it is plausible that the scooter–bicycle difference in trip duration varies with trip length. For the selected interaction model, partial residual plots for log distance were examined to assess the adequacy of the assumed linear functional form. All models were estimated using the full analytic sample of 5.49 million trips.

Results

Question 1: Effect of Vehicle Type on Trip Distance

Trip distance was modeled as a function of vehicle type and the temporal and spatial control variables:

$$\log(Distance_i) = \beta_0 + \beta_1 I\{Scooter_i\} + \boldsymbol{\gamma}^\top \mathbf{X}_i + \varepsilon_i$$

where the indicator $I\{Scooter_i\}$ equals 1 for scooter trips and 0 for bicycle trips, which serve as the reference category. The vector $\mathbf{X}_i$ consists of dummy variables for hour of day, day of week, and starting council district, capturing systematic temporal and spatial variation in trip behavior. The coefficient vector $\boldsymbol{\gamma}$ represents the associated effects of these control variables. The error term ε_i captures unexplained variation in log-transformed trip distance.

BIC was used to compare candidate specifications for modeling trip distance. As shown in Figure 2, the log-transformed ordinary least squares (OLS) model achieved the lowest BIC among all competing models, indicating the most favorable balance between model fit and complexity. Among the log-scale specifications (log-OLS, WLS, and Huber), BIC values were of similar magnitude, whereas the Box–Cox specification yielded a substantially higher BIC under this criterion.

Model <chr>	BIC <dbl>
OLS_raw	92487424
OLS_log	17005040
BoxCox_log	43809476
WLS_log	17019218
Huber_log	17099228

Figure 2
Bayesian Information Criterion (BIC) comparison across candidate trip distance models.

As an additional robustness check, the estimated scooter coefficients were compared across the log-scale specifications. As shown in Table 1, the scooter coefficient was nearly identical under log-OLS and weighted least squares and remained similar under Huber regression, indicating that the estimated scooter–bicycle difference in trip distance is not sensitive to the choice of fitting method on the log scale. Box–Cox specifications produced a similarly negative scooter effect, though coefficients are not directly comparable due to differences in scale (see Appendix B).

Table 1
Estimated scooter effect on trip distance across log-scale model specifications

Model	Scooter coefficient (log scale)
Log-OLS	-0.518
WLS	-0.519
Huber	-0.454

Based on both the BIC comparison and the interpretability of coefficients on the log scale, the log-OLS model was selected as the final specification for Question 1. The estimated scooter coefficient in the final model was $\hat{\beta}_1 = -0.518$ ($p < 0.001$). On the log scale, this coefficient represents a multiplicative difference. Converting to the original scale, $\exp(-0.518) \approx 0.596$ indicating that scooter trips are, on average, 40.4% shorter than bicycle trips when occurring at the same hour, on the same day, and in the same district (see Table 2).

Table 2
Estimated scooter effect on trip distance (log-OLS).

Term	Estimate	SE	$\exp(\hat{\beta})$	% change
Scooter (vs. bicycle)	-0.5183	0.00215	0.5955	-40.45

Residual-versus-fitted plots for the log-OLS model showed that log transformation reduced—but did not eliminate—heteroscedasticity. Variance increased slightly for longer trips, though the improvement over the raw-scale model was substantial. Normal Q–Q plots indicated approximate normality in the central region with minor tail deviations. These diagnostics suggest that the log-OLS model satisfies linearity and variance-stabilization assumptions reasonably well for descriptive inference.

GVIF values for all categorical predictors were near 1, indicating negligible multicollinearity. Cook's distances were uniformly small (< 0.02), even for high-leverage combinations of rare hour–district pairs, demonstrating that no single observation strongly influenced the estimated coefficients. Across WLS, Huber, and Box–Cox models, the estimated scooter effect remained negative and large, confirming that the conclusion is robust to alternative specifications. Diagnostic plots, coefficient comparisons, and supplementary assessments are provided in Appendix B.

Question 2: Effect of Vehicle Type on Trip Duration Conditional on Distance

Trip duration was modeled as a function of vehicle type, trip distance, and the same set of temporal and spatial control variables used in Question 1. Because raw trip duration exhibited substantial right skewness and heteroscedasticity even after outlier trimming, inference was conducted on the log scale. The baseline specification was given by

$$\log(Duration_i) = \alpha_0 + \alpha_1 I\{Scooter_i\} + \alpha_2 \log(Distance_i) + \boldsymbol{\delta}^\top \mathbf{X}_i + u_i$$

where X_i again collects dummy variables for hour, day, and district, δ is the corresponding coefficient vector, and u_i is an error term.

In the main-effects model, the estimated scooter coefficient was $\hat{\alpha}_1 = -0.028$ ($p < 0.001$). On the log scale, this corresponds to a duration ratio of $\exp(-0.028) \approx 0.972$, indicating that scooter trips were, on average, approximately 2.8% shorter in duration than bicycle trips when occurring at the same distance and under the same temporal and spatial conditions. However, although statistically significant given the large sample size, this difference is small in magnitude relative to the dominant effect of trip distance. For a typical trip length in the sample, the implied difference in expected duration is on the order of only a few seconds, rendering it negligible from an operational or rider-experience perspective.

Alternative functional forms, including Box–Cox transformation, weighted least squares, and Huber regression, were examined as robustness checks by comparing BIC values and the stability of key coefficient estimates. As in Question 1, the log-transformed duration model provided the most appropriate baseline specification.

Because the difference in trip duration between scooters and bicycles may vary with trip length, interaction models were subsequently evaluated including vehicle type, log distance, hour of day, and starting council district. As shown in Figure 3, comparison of BIC values across candidate interaction models indicated that the vehicle type × log distance interaction achieved the lowest BIC. This specification was therefore selected as the final model.

Model <chr>	BIC <dbl>
main_log	8621600
Veh_dis_log	8618148
Veh_hour_log	8619729
Veh_district_log	8621300

Figure 3
Bayesian Information Criterion (BIC) comparison across candidate trip duration models.

In the final interaction model, the estimated scooter main effect and interaction coefficients were $\hat{\alpha}_1 = 0.429$ and $\hat{\alpha}_3 = -0.063$. The positive main effect reflects the scooter–bicycle difference at the reference value of log distance (i.e., log(Distance) = 0, corresponding to a distance of 1 meter), while the negative interaction term indicates that this difference decreases as trip distance increases. Because this reference value corresponds to an unrealistically short trip, the scooter main effect is not substantively meaningful on its own and must be interpreted jointly with the interaction term. Evaluating the combined scooter effect at the sample mean of log distance ($\bar{x} = 6.80$) yields $\Delta_{\text{scooter}}(\bar{x}) = \hat{\alpha}_1 + \hat{\alpha}_3\bar{x} \approx 0.003$, corresponding to a duration ratio of $\exp(0.003) \approx 1.003$ (see Table 3). Thus, for a typical trip distance, scooter and bicycle trips have nearly identical expected durations once distance and contextual factors are controlled for.

Table 3
Scooter effect on trip duration: main-effects vs. interaction model (log scale).

Model	$\hat{\alpha}_1$ (scooter)	$\hat{\alpha}_3$ (interaction)	$\bar{x}$	$\Delta_{\text{scooter}}(\bar{x})$
Main-effects (no interaction)	-0.0281	–	–	-0.0281
Interaction (vehicle × log(distance))	0.4290	-0.0626	6.8042	0.0030

Residual-versus-fitted plots indicated variance patterns similar to those observed in raw-scale models, with funnel-shaped spread persisting at larger fitted values, suggesting remaining heteroscedasticity. Partial residual plots for log distance supported an approximately linear relationship between log duration and log distance across most of the observed range. Normal Q–Q plots showed departures from normality in the tails, while the central portion of the distribution was closer to linearity. GVIFs for vehicle type, log distance, and their interaction ranged from approximately 5 to 9, remaining below the commonly used threshold of 10 for problematic multicollinearity. Cook's distances were uniformly small (< 0.02), indicating that no individual observation exerted undue influence on the estimated coefficients. Specific diagnostic plots, robustness checks, and additional evaluations are reported in Appendix C.

Conclusion

This study provides a descriptive comparison of bicycle and scooter trips in Austin's shared micromobility system, focusing on conditional differences in trip distance and duration. However, several limitations of this analysis should be noted. As an observational study, the analysis characterizes conditional patterns in trip distance and duration rather than isolating causal effects. Observed differences between scooter and bicycle trips may therefore reflect rider self-selection, trip purpose, or other unobserved contextual factors rather than inherent performance differences between vehicle types. In addition, the dataset lacks information on route characteristics such as elevation, traffic conditions, infrastructure quality, and weather, all of which may systematically affect travel distance and duration. From a modeling perspective, although log transformations and alternative specifications helped mitigate some violations of model assumptions, residual heteroscedasticity and departures from normality remained. Future work could address these issues using more flexible variance structures, such as generalized linear models, mixed-effects models with random slopes for time or location, or semiparametric approaches that allow for nonlinear distance effects.

Within these limitations, the results indicate clear differences in how shared scooters and bicycles are used within Austin's micromobility system. Scooter trips are substantially shorter than bicycle trips in terms of distance, even after controlling for time of day, day of week, and district, suggesting that the two modes serve complementary roles rather than acting as close substitutes. At the same time, once trip distance is controlled for, the expected duration of scooter and bicycle trips is nearly identical, indicating comparable travel-time performance for a given trip length. Taken together, the analysis shows that the primary distinction between scooters and bicycles in Austin's shared micromobility system lies in how far people ride them, not in how fast or efficiently they travel for a given distance.

References

City of Austin. (2019). *Austin Strategic Mobility Plan*. https://www.austintexas.gov/department/austin-strategic-mobility-plan

National Association of City Transportation Officials. (2023). *Shared micromobility in the U.S. and Canada: 2022*. https://nacto.org/publication/shared-micromobility-report-2022/

National Association of City Transportation Officials. (2024). *Shared micromobility report 2023*. https://nacto.org/publication/shared-micromobility-report-2023/