Predicting Citi Bike Usage in a Post-Congestion Pricing Era: A Multi-Model Urban Forecasting Framework

Venakata Sai Vardhan Bora
Department of Artificial Intelligence, Yeshiva University
NY, USA
vbora1@mail.yu.edu

Prudhvi Raj Maharana
Department of Artificial Intelligence, Yeshiva University
NY, USA
pmaharan@mail.yu.edu

Praveen Kumar Govind Reddy
Department of Artificial Intelligence, Yeshiva University
NY, USA
pgovindr@mail.yu.edu

Andria Grace
Department of Artificial Intelligence, Yeshiva University
NY, USA
alnu@mail.yu.edu

***Abstract*—On January 5, 2025, New York City became the first major U.S. city to implement congestion pricing, charging vehicles a toll to enter Manhattan's central business district. While designed to reduce traffic and emissions, this policy is also shifting commuter behavior—prompting increased reliance on micromobility options like Citi Bike. In response, this study presents a data-driven framework to forecast daily Citi Bike demand at the station level by integrating historical usage data (2021–2024), daily weather records, and detailed urban contextual features.**

We constructed spatial profiles for each station by quantifying nearby transit hubs, retail zones, residential buildings, and civic spaces within a 500-meter radius using open geospatial data. These features were combined with temporal variables (e.g., lagged demand, weekday/weekend flags) to train and evaluate sixteen models, ranging from classical time series to deep learning. Deep learning models—especially sequence-based architectures—achieved the highest accuracy, with the top model reaching a mean absolute percentage error of 8.7%. Ensemble methods like Random Forest and XGBoost offered competitive performance and interpretability. We also propose extending the framework with a binary policy indicator and a difference-in-differences approach to assess the impact of congestion pricing. This work offers actionable insights for sustainable urban mobility planning. This work focuses on daily station-level demand forecasting, with current implementation evaluated on a representative Manhattan station, and the approach scalable across the broader Citi Bike network.

***Index Terms*—Citi Bike, demand forecasting, time series prediction, urban mobility, deep learning, machine learning, congestion pricing, urban features, spatial modeling**

I. Introduction

A. Motivation: Urban Mobility and the Role of Congestion Pricing in NYC

Urban centers across the globe are facing mounting transportation challenges, including severe traffic congestion, deteriorating air quality, rising carbon emissions, and inequitable access to mobility services [1], [2]. These pressures are intensifying with population growth, urban sprawl, and climate change. To confront these issues, many cities are enacting bold, data-informed policies designed to reshape commuting behavior and promote sustainable, multimodal transportation systems. Among such strategies, congestion pricing has emerged as a high-impact solution aimed at internalizing the external costs of car travel.

In a landmark move, New York City introduced North America's first congestion pricing initiative on January 5, 2025, targeting Manhattan's central business district. The policy imposes tolls on vehicles entering the zone during peak hours, with goals that extend beyond traffic mitigation: reducing greenhouse gas emissions, reallocating road space, and generating revenue for public transit infrastructure improvements. Early projections estimate a significant decline in vehicle volume and a redirection of funds—potentially over $500 million annually—toward enhancing subway, bus, and alternative transportation services.

While the benefits to traditional transit systems are anticipated, the policy's ripple effects on other modes—particularly micromobility platforms like Citi Bike—remain less understood. Congestion pricing may trigger widespread behavioral changes, encouraging commuters to seek cost-effective, flexible, and environmentally friendly alternatives to car travel. As shared bike systems become increasingly embedded in urban mobility ecosystems, the ability to accurately forecast demand shifts in response to evolving policies, seasonal weather, and infrastructural change becomes crucial. Such forecasting not only ensures operational efficiency but also strengthens the city's capacity to manage a multimodal future that is equitable, responsive, and sustainable.

B. Why Forecasting Citi Bike Demand Matters

Citi Bike is the largest bike-sharing system in the United States [3], [4] and a critical component of New York City's broader transportation network. With more than a thousand stations operating across boroughs, demand fluctuates daily, influenced by a multitude of interdependent variables including weather conditions, nearby land use, transit accessibility, seasonal patterns, and special events. Accurately forecasting this demand is essential for both operational efficiency and long-term planning. It enables real-time decisions such as timely bicycle redistribution, station maintenance, and resource allocation. Strategically, forecasting supports decisions around network expansion, public policy evaluation, and integration with mass transit systems.

With the advent of congestion pricing, the forecasting challenge has become more urgent and complex. This major policy shift could trigger meaningful changes in travel behavior, making it vital to understand how Citi Bike usage responds to new incentives and constraints. Addressing this requires predictive models that are both accurate and sensitive to spatiotemporal and policy contexts. These considerations strongly motivated the development of our integrated, data-driven forecasting framework.

C. Research Questions

This study is designed to explore several key research questions that sit at the intersection of predictive modeling, urban analytics, and policy analysis: How accurately can daily Citi Bike demand at the station level be forecast using a combination of historical usage data, weather variables, and static urban contextual features?

Which modeling approaches—including classical time series, machine learning, and deep learning techniques—perform best in this spatiotemporal forecasting task?

How do urban features such as proximity to transit, commercial activity, and residential density influence demand, and can these be leveraged to improve model performance?

Can a forecasting framework be designed to support future integration of policy-related variables, such as congestion pricing indicators, to assess real-world impact?

D. Overview of Approach and Key Findings

To address the questions above, we assembled a comprehensive station-level dataset by integrating four years of Citi Bike trip records (2021–2024), daily weather data, and detailed urban context features obtained from OpenStreetMap. Each station's surroundings were profiled by counting the number of relevant urban amenities—transit stops, commercial zones, residential areas, and public institutions—within a 500-meter buffer. These static spatial features were combined with time-based variables (e.g., weekday/weekend flags, lagged booking counts, holidays) and environmental factors (e.g., temperature) to train a diverse suite of models.

We implemented and rigorously evaluated sixteen forecasting techniques across four methodological categories: (1) classical time series models (e.g., ARIMA, SARIMA), (2) regularized regression methods (e.g., ElasticNet), (3) tree-based ensemble learners (e.g., Random Forest, XGBoost), and (4) deep learning models (e.g., GRU, CNN-LSTM, Transformer). Each model was trained using pre-2025 data and assessed on holdout sets using standard metrics like MAE, RMSE, and MAPE.

Our evaluation revealed that deep learning models, particularly sequence-aware architectures, delivered the highest prediction accuracy, with a top-performing model achieving a MAPE of 8.7%. Ensemble methods, especially Random Forest and XGBoost, also performed strongly and offered interpretable insights into feature importance. Building on this validated forecasting pipeline, we propose an extension using a binary indicator for post-congestion pricing periods and a Difference-in-Differences (DiD) framework to facilitate causal policy impact assessment.

II. Related Works

A. Time Series, Machine Learning, and Deep Learning for Bike-Sharing Forecasting

Forecasting demand in bike-sharing systems has been a widely explored topic, particularly in large urban environments such as New York City [5], [6], [12]. Early studies focused on classical time series models, including ARIMA, SARIMA, and Holt-Winters exponential smoothing, which performed reasonably well for short-term predictions but lacked the ability to incorporate exogenous variables like weather and events. These models often treated each station in isolation and assumed linear relationships, limiting their ability to adapt to complex urban dynamics.

As computational power and data availability improved, machine learning (ML) models such as Random Forest, XGBoost, and ElasticNet became popular for demand forecasting. These models allowed the integration of additional features like temperature, precipitation, and holiday effects, and captured non-linear relationships more effectively. More recently, deep learning approaches, particularly Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, have been applied to bike-share data. These models are adept at capturing long-term dependencies and seasonality in time series data and have shown superior performance in many urban forecasting applications.

B. Forecasting with Urban Context and Spatial Features

Beyond temporal and weather factors, recent research has emphasized the importance of spatial and built environment features in understanding bike-sharing behavior [7], [13], [17]. Tang et al. (2021) analyzed the impact of urban POIs such as parks, retail outlets, and transit stations on bike-share demand across several U.S. cities. Their findings confirmed that the built environment significantly influences ridership patterns. Lin et al. proposed a graph-based neural network model to forecast Citi Bike usage, incorporating spatial dependencies between stations, while Sankaran et al. developed a system dynamics model that blended ML with behavioral insights.

These studies underscore that urban context matters—stations surrounded by transit hubs, residential buildings, or commercial districts experience different usage patterns than those in isolated or purely residential areas. However, many prior works relied on generic GIS datasets or categorical zone indicators rather than station-level, fine-grained contextual data.

C. Comparison to Prior Work

While prior models have incorporated elements of spatial or contextual awareness, few combine fine-grained station-level urban features [11] with a broad benchmarking of traditional, ensemble, and deep learning models. Moreover, very few studies have evaluated how these forecasting systems could be adapted to assess the effects of urban policy interventions, such as congestion pricing.

Our approach builds on this foundation by:

- Integrating static, station-centric urban features using OpenStreetMap data within a 500-meter buffer;
- Combining these features with weather and time-based variables in a unified pipeline;
- Comparing 16 forecasting models across multiple ML and DL families;
- Proposing a policy-aware extension for difference-in-differences analysis using the same architecture.

D. Contribution and Novelty of This Study

This study contributes to the growing literature on bike-share demand forecasting in several key ways:

Integration of fine-grained urban contextual data: We systematically quantify the built environment around each station, capturing variations in residential density, transit access, and commercial activity at a granular level.

Model-agnostic benchmarking: We evaluate a comprehensive suite of forecasting methods—including ARIMA, ElasticNet, ensemble learning, and deep learning models—on the same dataset, allowing robust comparison.

Policy-readiness: Our architecture is designed to support binary policy indicators (e.g., post-congestion pricing) and can be extended using a difference-in-differences framework to evaluate real-world interventions.

Spatiotemporal fusion: By combining time series dynamics with spatial features, our models provide both accurate forecasts and insights into the urban conditions driving Citi Bike usage.

In doing so, this work bridges a gap between operational forecasting and policy evaluation in micromobility systems and lays the groundwork for real-time, context-aware decision-making.

III. Data and Feature Engineering

A. Datasets Utilized

This study combines three primary datasets to develop a robust station-level demand forecasting framework:

1) Citi Bike Trip Data (2021–2024): Daily trip data was sourced from citibikenyc, covering the period from January 2021 through December 2024. Each record includes the station identifier, date, and the number of bike trips that began at the station on that day.
2) Weather Data: Weather metrics, including daily average temperature, were collected from historical NOAA records for New York City. Temperature is known to significantly impact bike-share usage, with warmer days generally correlating with higher ridership.
3) Urban Feature Data from OpenStreetMap (OSM): We extracted detailed spatial attributes around each Citi Bike station using OSM's Overpass API. For each station, we computed the number of surrounding points of interest (POIs) within a 500-meter radius. These include counts of transit stops (subway, bus), libraries, colleges, parks, attractions, athletic facilities, supermarkets and convenience stores, malls and shopping centers, restaurants, cafes, bars, retail stores, hospitals, residential buildings, civic buildings, and event venues. This rich spatial context was used as a static urban profile input for each station, providing a granular view of the built environment.

B. Spatial Feature Extraction

To capture the effect of built environment on bike usage, we engineered a set of station-specific urban features. For each Citi Bike station, we calculated the number of amenities and infrastructures located within a 500-meter radius using latitude and longitude coordinates as the center point. The extracted features include:

- Transit access: Number of subway and bus stops
- Land use: Counts of residential buildings, commercial buildings, restaurants, cafes, malls, retail stores
- Public amenities: Libraries, colleges, hospitals, civic buildings, parks
- Attraction and leisure: Event venues, attractions, athletic facilities
- Accessibility: Supermarkets and convenience store

These variables were compiled into a unified static feature vector per station and served as key explanatory inputs in our modeling framework. Table illustrates a sample of the engineered urban context features for selected Citi Bike stations in Manhattan, showing the diversity of nearby amenities used to enrich the model's spatial awareness.

C. Temporal Feature Construction

Temporal features were engineered to account for seasonality and behavioral patterns:

- Day of week (0–6) to capture weekday vs. weekend trends
- Is weekend (binary indicator)
- Holiday indicator (if applicable, based on U.S. federal holiday calendar)

TABLE I
URBAN FEATURE VARIABLES AND DESCRIPTIONS (WITHIN 500M)

Feature Code	Description
ux_metro	Number of metro/subway stops
ux_bus	Number of bus stops
ux_parks	Number of parks
ux_libs	Number of public libraries
ux_edu	College/university facilities
ux_attract	Tourist and local attraction count
ux_sports	Athletic centers and sports venues
ux_groc	Grocery/convenience stores
ux_malls	Malls and shopping centers
ux_dine	Cafes, bars, and restaurants
ux_retail	Retail and merchandise stores
ux_med	Nearby hospitals and clinics
ux_resi	Residential building density
ux_civic	Civic/government buildings
ux_eventhub	Stadiums, arenas, concert halls

- Lag features: Recent demand trends were incorporated through lag variables such as booking count from the previous day and the same day one week earlier.

These temporal variables were critical in enabling models to learn regular weekly cycles and adjust for anomalous patterns around holidays or weekends.

D. Data Preprocessing

Prior to modeling, the data underwent several preprocessing steps to ensure quality and consistency:

- Missing value handling: Days with missing booking counts or weather data were dropped or imputed using rolling averages
- Outlier detection: Days with zero bookings during peak seasons (e.g., summer weekdays) were flagged and reviewed for possible station outages or data errors. In extreme cases, they were excluded from training.
- Normalization: Continuous variables such as booking count and temperature were scaled using Min-Max normalization for deep learning models, facilitating faster convergence and numerical stability.
- Log Transformation: To address skewness in booking count distributions, a $\log(1 + x)$ transformation was applied prior to scaling, helping the model better handle variations in magnitude between low-demand and high-demand days.

Together, these data engineering steps enabled the integration of diverse sources—temporal, spatial, and environmental—into a unified, clean dataset suitable for high-performance forecasting.

IV. METHODOLOGY

A. Forecasting Problem Setup

The primary objective of this study is to forecast daily Citi Bike demand at the individual station level across Manhattan, with a specific focus on capturing short-term usage patterns that reflect both historical trends and external contextual influences. The forecasting task involves predicting the number of bike rentals that will originate from a given station on the following day, using a combination of past booking counts, meteorological data, and static urban spatial features. This enables fine-grained demand estimation that supports real-time operational decisions and long-term infrastructure planning.

We frame this as a univariate time series forecasting problem augmented with exogenous covariates, where the target variable is the daily number of trips starting at each station. Exogenous inputs include weather metrics (e.g., average temperature), calendar indicators (e.g., weekends), and fixed station-level features derived from geospatial data (e.g., number of nearby subway stops or commercial venues). This formulation allows the models to learn complex temporal dependencies while incorporating influential external signals that drive urban mobility behavior.

To maintain modeling clarity and computational feasibility, we focus our initial evaluation on one representative Citi Bike station (Station ID: 4818.03) situated in a high-demand, transit-dense region of central Manhattan. This station was selected for its consistent usage pattern and representative urban context. However, the framework we developed is fully generalizable and can be scaled to forecast demand across the entire Citi Bike network by retraining on other station datasets using the same architecture and feature design.

B. Model Categories and Forecasting Techniques

We benchmarked sixteen models across three major categories: classical time series methods, machine learning (ML), and deep learning (DL) models.

TABLE II
SUMMARY OF MODEL FAMILIES AND CONFIGURATIONS

Category	Models Implemented
Time Series	ARIMA, SARIMA, Holt-Winters, BSTS
Machine Learning	ElasticNet, Random Forest, XGBoost, LightGBM, CatBoost
Deep Learning	GRU, CNN-LSTM, Transformer, DeepAR

C. Deep Learning Architecture Details

Our best-performing models were based on Recurrent Neural Networks (RNNs), specifically GRU and CNN-LSTM architectures. Below is an overview of their configurations:

- **CNN-LSTM Hybrid**
  - **Input:** Sliding window of 30 previous days
  - **Features:** Scaled booking count, temperature, is_weekend
  - **Architecture:**
    - 1D Convolutional layer (kernel size = 3, filters = 64)
    - LSTM layer (units = 50)
    - Dropout (rate = 0.2)
    - Dense output layer (1 unit for next-day prediction)

- **GRU Model**
  - Similar to LSTM but uses Gated Recurrent Units (GRU)
  - Fewer parameters and faster convergence on small datasets
  - Captured temporal trends effectively, especially around seasonal transitions

Both models were trained using the Adam optimizer, early stopping based on validation loss, and scaled inputs (Min-Max normalization after log transformation).

D. Assumptions and Design Choices

Key design assumptions in this study include:

- Station Independence: Each station is modeled independently for simplicity and interpretability.
- Static Urban Features: Urban contextual features (e.g., nearby transit stops, retail stores) are assumed to be stable during the 2021–2024 period.
- Univariate Forecasting: Only the station's own demand is forecasted (no joint modeling with other stations).
- Window Size = 30 Days: Chosen to capture at least one full weekly cycle and short-term trends.

We also chose to focus on point forecasts, though probabilistic forecasting was briefly explored via DeepAR.

V. Experiments and Evaluation

A. Experimental Setup

To evaluate forecasting performance, we focused on daily demand for a representative station in central Manhattan (Station ID: 4818.03). The dataset spans four years (January 2021 to December 2024), and was chronologically split into three subsets:

- Training set: 70% of the data (2021–early 2023)
- Validation set: 15% (mid-2023)
- Test set: 15% (late 2023 to 2024)

A rolling window approach was used for time series modeling, with a look-back window of 30 days to predict the next day's demand. All features, including booking count, temperature, and is_weekend indicators, were scaled using Min-Max normalization. Lag features were included for ML models, while DL models received time-series inputs as sequences.

B. Model Evaluation Metrics

Model performance was assessed using three standard error metrics:

- **Mean Absolute Error (MAE):**

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

- **Root Mean Squared Error (RMSE):**

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

- **Mean Absolute Percentage Error (MAPE):**

$$\text{MAPE} = \frac{100}{n}\sum_{i=1}^{n} \left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

These metrics were chosen to capture both scale-dependent (MAE, RMSE) and scale-independent (MAPE) error characteristics.

C. Model Comparison and Results

All sixteen models were trained and evaluated using the same splits and inputs. The best overall performance was achieved by a CNN-LSTM model, with a test-set MAPE of 8.7%, outperforming classical and machine learning baselines.

Model	**MAE**	**RMSE**	**MAPE (%)**
ARIMA	5.82	6.91	13.8
SARIMA	5.24	6.42	11.9
Holt-Winters	5.30	6.38	12.1
ElasticNet	4.78	5.94	15.2
Random Forest	3.94	4.70	10.3
XGBoost	3.89	4.63	9.9
LightGBM	3.95	4.68	10.1
CatBoost	3.92	4.65	9.8
GRU	3.63	4.27	9.2
CNN-LSTM (best)	**3.42**	**4.06**	**8.7**
Transformer	4.51	5.58	14.5
DeepAR	4.89	6.12	16.0
BSTS	5.45	6.45	13.4
VAR	5.76	6.81	14.0
SARIMAX	5.11	6.29	11.5
ARIMAX	5.34	6.51	12.6

TABLE III
COMPARISON OF FORECASTING MODELS BASED ON MAE, RMSE, AND MAPE

D. Visualization of Model Predictions

Visual comparisons were created to illustrate actual vs. predicted booking counts for selected models on the test set. The CNN-LSTM model closely tracked real demand, even capturing short-term peaks and drops associated with cold weather or weekends.

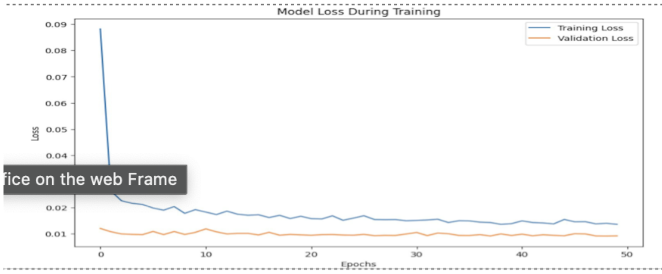

Fig. 2. Loss

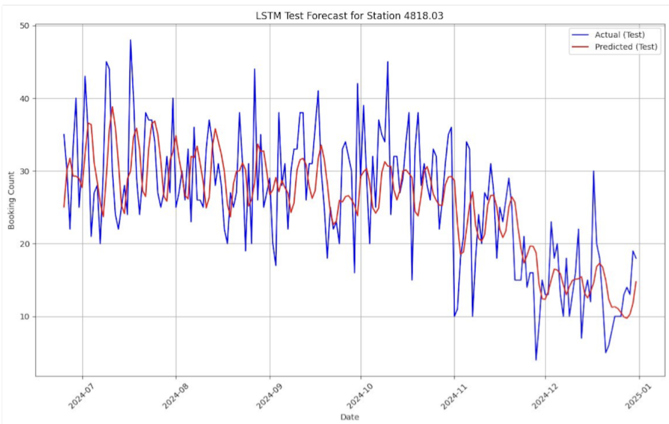

Fig. 1. CNN-LSTM Predictions vs Actual Values (Test Set)

These plots highlight the advantage of deep learning models in adapting to non-linear, seasonal, and behavioral shifts, particularly when external features such as temperature and weekend indicators are included [9].

E. Observations and Insights

1) Deep learning models outperformed others, with CNN-LSTM achieving the lowest forecast error.
2) Tree-based ensembles (Random Forest, XGBoost) offered strong performance and valuable interpretability via feature importance rankings.
3) Classical time series models, though interpretable, lagged in accuracy—especially during periods of irregular demand.
4) MAPE variation across models revealed their sensitivity to low-volume days, particularly during winter months.

VI. Results and Discussion

A. Top-Performing Models

The results from our evaluation indicate that deep learning models, particularly the CNN-LSTM, consistently outperformed all other approaches. The CNN-LSTM model achieved a mean absolute percentage error (MAPE) of 8.7 percent, the lowest among all sixteen models evaluated. It demonstrated superior ability to track both long-term trends and short-term fluctuations in Citi Bike demand, adapting well to variations driven by weather, weekday/weekend cycles, and other latent patterns.

Close behind were the Gated Recurrent Unit (GRU) and tree-based ensemble models, such as Random Forest, XGBoost, and CatBoost. These models also delivered high accuracy, with MAPE values under 10.5 percent. While they did not quite match the performance of deep learning models on high-variance days, they offered substantial interpretability through feature importance analysis.

B. Feature Importance and Model Interpretability

Using feature importance scores from Random Forest and XGBoost models, we identified the most influential predictors of station-level demand:

- Lagged demand values (previous day and previous week)
- Average temperature
- Weekend indicator
- Proximity to transit infrastructure (number of metro and bus stops)
- Residential and commercial density

To move beyond exploratory data analysis (EDA) [15], we employed model-based explanation techniques using feature importance scores derived from Random Forest and XGBoost models. These techniques quantify the contribution of each input variable to the model's predictions, providing interpretable insights into the drivers of Citi Bike demand. Features such as lagged demand, average temperature, weekend indicators, and proximity to subway or bus stops consistently ranked highest. This highlights the practical utility of urban features in both forecasting and policy evaluation contexts. In future work, advanced explanation methods like SHAP (SHapley Additive exPlanations) could be used to further enhance transparency. These results validate the value of integrating urban contextual features into forecasting models. For example, stations with high residential density and nearby subway access saw consistent weekday demand, while those near restaurants and event venues showed peaks during weekends and evenings. Such findings offer intuitive insights for operations and policy planning.

For instance, stations surrounded by a high number of restaurants and event venues tended to show elevated demand on weekends and evenings, while those near subway stops and residential clusters exhibited consistent weekday usage. This spatial heterogeneity, as captured in our urban features dataset, was instrumental in boosting model performance. Importantly, features like 'Restaurants, Cafes, and Bar_500m', 'Retail Stores_500m', and 'Attractions_500m' emerged as strong demand drivers during leisure periods, whereas 'metro_500m' and 'residential_500m' were dominant on workdays.

To further validate the relevance of urban contextual features, we computed the Pearson correlation coefficients between each spatial feature and the station-level booking counts. As shown in Figure, features such as proximity to metro stops, bus stops, attractions, and supermarkets exhibit moderate positive correlations with Citi Bike demand. This supports our

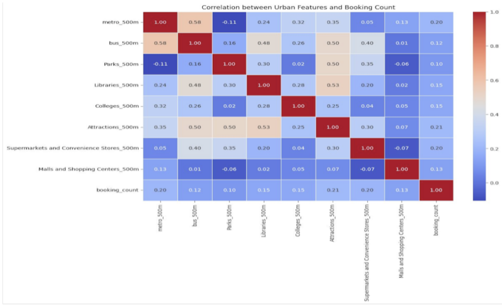

Fig. 3. Correlation between urban features and booking count. Stronger positive correlations are observed with proximity to metro stations, bus stops, and attractions.

model-based findings and highlights the urban elements most influential in driving usage patterns.

C. Accuracy vs Interpretability Trade-off

While deep learning models such as CNN-LSTM delivered the highest forecasting accuracy, they operate as black-box architectures, making them less interpretable for stakeholders without technical backgrounds. These models capture complex nonlinear relationships and long-range temporal dependencies, but their inner workings—such as hidden state transitions and weight updates—are often opaque to end-users, limiting their direct applicability in policy settings where transparency and justification are crucial.

In contrast, ensemble learning models like Random Forest and XGBoost, though slightly less accurate, offered clear interpretability through feature importance scores and decision path visualizations. These models provided actionable insights into which variables—such as lagged demand, weather patterns, or surrounding land use—most strongly influenced predictions. This transparency is especially valuable when communicating with planners, policymakers, or the general public who need to understand the rationale behind predictive trends.

This trade-off highlights a practical balance in urban analytics between operational forecasting, where accuracy and responsiveness are paramount, and policy evaluation, where explainability and accountability take precedence. In real-world deployments, both model families can be used in tandem—with deep learning models delivering high-precision forecasts for tactical planning and ensemble models offering interpretable diagnostics that support strategic decisions, stakeholder engagement, and trust-building in the modeling process.

D. Limitations

Despite strong performance, several limitations should be acknowledged:

1) Censored Demand: The Citi Bike data used in this study [8], [18] captures observed usage, which may underrepresent true demand during peak hours or high-traffic periods when bikes are unavailable at a given station. This "censored demand" implies that the actual demand may be higher than recorded, especially at popular or understocked stations. As a result, our forecasts may be biased downward in such scenarios. Future work could incorporate station inventory logs, rebalancing operations, or alternative data sources (e.g., user app interactions) to better estimate and model unmet demand.
2) Station Independence: Our models treat each station independently. While this simplifies modeling and improves scalability, it neglects possible spatial correlations (e.g., overflow demand shifting to nearby stations).
3) Static Urban Features: Urban contextual features are treated as constant over time. However, Manhattan's infrastructure can evolve—new developments, station relocations, and street redesigns may change demand drivers over time.
4) No Direct Policy Variable: Though the framework supports the inclusion of a congestion pricing indicator, our current forecasts rely only on pre-policy data. This limits the ability to infer causal impact until sufficient post-policy data becomes available.

In summary, our findings confirm that combining temporal dynamics with spatial intelligence significantly improves bike-share demand forecasting. The CNN-LSTM model, enriched with urban and environmental inputs, offers a strong baseline for operational planning. Meanwhile, tree-based models enhance interpretability—crucial for planners and decision-makers. Future extensions will build upon this foundation to evaluate the causal effects of major transportation policies like congestion pricing.

VII. Policy Implication and Extension

A. Supporting Congestion Pricing Analysis with Forecasting Models

The congestion pricing program launched on January 5, 2025 in Manhattan has introduced a new dimension to urban transportation dynamics. While its impact on vehicle traffic and public transit has been discussed, its effect on micromobility systems like Citi Bike remains underexplored. Our demand forecasting framework offers a powerful baseline for monitoring and analyzing how Citi Bike usage changes in response to this policy.

Because our models are designed to incorporate temporal, environmental, and spatial variables, they can be readily extended to assess demand shifts that occur post-policy. The high forecasting accuracy (MAPE 8.7%) of the CNN-LSTM model establishes a reliable counterfactual for what Citi Bike usage would have been without the policy, thereby enabling robust impact evaluation.

B. Integration of Binary Policy Indicator

To make the model policy-aware, we propose introducing a binary congestion pricing feature into the existing input structure. This feature takes a value of:

- 0 for all dates prior to January 5, 2025
- 1 for all dates following the implementation of the policy

This binary indicator can be treated similarly to our existing is_weekend feature, allowing deep learning models like CNN-LSTM or GRU to learn the policy's influence over time. The design allows the model to learn new post-policy patterns, even without explicitly encoding policy details like toll amounts.

C. Difference-in-Differences (DiD) Framework Overview

In addition to enhancing our forecasting model, we propose a Difference-in-Differences (DiD) approach to quantify the causal effect of congestion pricing. DiD is a statistical method used to compare the change in outcomes over time between a group exposed to an intervention (treatment group) and one that is not (control group).

In our context:

- Treatment group: Citi Bike stations located inside the congestion pricing zone (CPZ)
- Control group: Stations located just outside the CPZ, which are not directly affected by the toll

By comparing the pre- and post-policy changes in demand across both groups, we can isolate the net effect of the congestion pricing policy on bike-share usage, controlling for seasonality and macro trends.

This extension would complement the LSTM model's predictive power by offering a causal lens on the observed changes, especially useful in a policy analysis setting.

D. Use Cases for Planners and Operators

The integration of congestion pricing indicators and the application of DiD methods make our forecasting system directly valuable to urban planners, policymakers, and Citi Bike operators:

- Resource Allocation: Anticipate surge in demand in key CPZ areas and pre-position bikes accordingly.
- Infrastructure Planning: Identify stations that may require expansion, new docks, or improved connectivity.
- Policy Evaluation: Provide empirical evidence of how congestion pricing influences mobility patterns, supporting transparency and iterative improvements.
- Scenario Simulation: Run what-if simulations to estimate the effect of similar policies in other parts of the city.

By embedding policy-sensitivity into our models, we move beyond forecasting and toward real-time policy monitoring, enabling New York City to adapt more effectively to the ripple effects of its bold mobility reforms.

VIII. CONCLUSION

This study presents a robust, data-driven forecasting framework to estimate daily Citi Bike demand at the station level across Manhattan. By integrating historical trip data, weather records, and granular urban contextual features, we constructed a comprehensive spatiotemporal dataset that supports high-accuracy modeling. We evaluated sixteen forecasting methods spanning classical time series models, machine learning algorithms, and deep learning architectures. Among them, deep learning models—particularly CNN-LSTM—achieved the best predictive performance, while ensemble models offered strong interpretability and operational insights.

Our results confirm that combining temporal dynamics with built environment data substantially enhances forecast accuracy, with the best model reaching a MAPE of 8.7%. Feature analysis further revealed the importance of proximity to transit infrastructure, recent usage trends, and environmental conditions. These findings provide not only operational guidance for Citi Bike operators but also policy-relevant insights as cities implement mobility interventions like congestion pricing.

Looking ahead, we propose extending this framework with a binary policy indicator to capture post-congestion pricing effects and applying a difference-in-differences (DiD) method for causal impact assessment. Future work will focus on generalizing the model to all stations, incorporating dynamic spatial features, and integrating real-time data streams. By bridging demand forecasting and policy evaluation, this work sets the foundation for smarter, more adaptive urban mobility planning in the era of climate-conscious transport reform.

REFERENCES

[1] E. Fishman, S. Washington, and N. Haworth, "Bike Share: A Synthesis of the Literature," *Transport Reviews*, vol. 33, no. 2, pp. 148–165, 2013.

[2] S. A. Shaheen, S. Guzman, and H. Zhang, "Bikesharing in Europe, the Americas, and Asia: Past, Present, and Future," *Transportation Research Record*, vol. 2143, no. 1, pp. 159–167, 2010.

[3] O. O'Brien, J. Cheshire, and M. Batty, "Mining Bicycle Sharing Data for Insights into Sustainable Transport Systems," *Journal of Transport Geography*, vol. 34, pp. 262–273, 2014.

[4] G. McKenzie, "Spatiotemporal Comparative Analysis of Scooter-Share and Bike-Share Usage Patterns in Washington, D.C.," *Journal of Transport Geography*, vol. 78, pp. 19–28, 2019.

[5] D. Buck et al., "Are Bikeshare Users Different from Regular Cyclists? A First Look," *Transportation Research Record*, vol. 2387, no. 1, pp. 112–119, 2013.

[6] E. Martin and S. A. Shaheen, "Evaluating Public Transit Modal Shift Dynamics in Response to Bikesharing," *Journal of Transport Geography*, vol. 41, pp. 315–324, 2014.

[7] J. Zhao, W. Deng, and Y. Song, "Ridership and Effectiveness of Bike-Sharing in China," *Transport Policy*, vol. 35, pp. 253–264, 2014.

[8] A. Goodman and J. Cheshire, "Inequalities in the London Bicycle Sharing System Revisited," *Journal of Transport Geography*, vol. 41, pp. 272–279, 2014.

[9] A. Faghih-Imani and N. Eluru, "Analyzing Bike-Sharing System User Destination Choice: Chicago's Divvy," *Journal of Transport Geography*, vol. 44, pp. 53–64, 2015.

[10] R. Nair, E. Miller-Hooks, R. C. Hampshire, and A. Busic, "Large-Scale Vehicle Sharing Systems: Analysis of Vélib'," *International Journal of Sustainable Transportation*, vol. 7, no. 1, pp. 85–106, 2013.

[11] M. Ricci, "Bike Sharing: A Review of Evidence on Impacts and Implementation," *Research in Transportation Business & Management*, vol. 15, pp. 28–38, 2015.

[12] A. Faghih-Imani et al., "Hail a Cab or Ride a Bike? Travel Time Comparison in NYC," *Transportation Research Part A*, vol. 101, pp. 11–21, 2017.

[13] Y. Zhang et al., "Impact of Built Environment Factors on Public Bike Use: Zhongshan, China," *Journal of Transport Geography*, vol. 58, pp. 59–70, 2017.

[14] D. Fuller et al., "Impact Evaluation of BIXI Bike Share in Montreal," *American Journal of Public Health*, vol. 103, no. 3, pp. e85–e92, 2013.

[15] E. Murphy and J. Usher, "Role of Bicycle-Sharing in the City: Irish Experience," *International Journal of Sustainable Transportation*, vol. 9, no. 2, pp. 116–125, 2015.

[16] Y. Zhang and Z. Mi, "Environmental Benefits of Bike Sharing: Big Data Analysis," *Applied Energy*, vol. 220, pp. 296–301, 2018.

[17] A. A. Campbell et al., "Shared Bicycles and E-Bikes in Beijing: Choice Factors," *Transportation Research Part C*, vol. 67, pp. 399–414, 2016.

[18] W. Li, Y. Zhang, and W. Wang, "Bike-Sharing Systems' Impacts on Congestion in US Cities," *Journal of Transport Geography*, vol. 68, pp. 74–83, 2018.