Cloudy With A Chance of BIXI: A Demand Forecasting Machine Learning Model

Calvin Yip

Executive Summary

BIXI Montréal operates one of the largest bike-sharing systems in North America, where mismatches between bike supply and demand directly impact user satisfaction, missed trips, and operational costs. This project develops an **hourly, station-level demand forecasting model** to support more proactive rebalancing and short-term capacity planning.

Using **BIXI open trip data**, **station geolocation**, and **weather data from the Meteostat API**, historical trips were aggregated to a station–hour level. A feature engineering pipeline captured **daily and weekly usage cycles**, **short-term demand persistence**, **spatial effects**, and **weather conditions**, while carefully avoiding temporal leakage.

Multiple models were evaluated, including a naïve baseline, Random Forest, and Histogram-based Gradient Boosting (HGB). While Random Forest and HGB achieved similar predictive accuracy, **HGB was selected as the final model** due to its **comparable performance and significantly lower training time**, making it more suitable for operational use.

The final model achieves an average error of **approximately two trips per station per hour**, a level of accuracy that is operationally meaningful for bike-sharing systems. **SHAP-based interpretability** reveals that demand is primarily driven by **recent usage history and time-of-day patterns**, with **location and weather acting as secondary modifiers**.

This project demonstrates an **end-to-end applied analytics workflow** — from data integration and feature engineering to model selection, interpretability, and business translation — and provides a practical foundation for supporting real-world operational decision-making in urban mobility systems.

Project Overview

BIXI Montreal operates one of the largest bike-sharing systems in North-America, with thousands of bikes and stations distributed across the city. Data shows BIXI ridership has hit an all time high during 2024 (source). Undoubtedly, BIXI is an integral part of Montreal's public system and part of Montrealer's identity. With this in mind, a recurring challenge is balancing supply and demand throughout stations. Some stations experience shortages while others accumulate unused bikes. This affects user satisfaction, missed trips, additional operational costs, and the organization's contribution to a sustainable urban mobility.

By leveraging trip history data, weather, and temporal patterns, our team aims to uncover insights into usage behavior and provide a predictive model that supports BIXI's operational planning.

The objective of this project is to develop a machine learning model capable of forecasting **hourly bike demand at the station level** using historical usage data, temporal features, spatial information, and weather conditions. The focus is on short-term demand prediction that can support operational decision-making rather than long-term strategic forecasting.

Using historical BIXI trip data combined with weather and calendar information, this study applies tree-based machine learning models to capture non-linear relationships, temporal persistence, and cyclical demand patterns. Model performance is evaluated using standard regression metrics, interpretability techniques are employed to understand key demand drivers, and results are translated into practical operational insights for a bike-sharing operator.

Data & Feature Engineering

Data Sources

Three complementary data sources were used to capture demand, operational context, and environmental conditions affecting bike-sharing usage.

BIXI Open Data (source)

The primary dataset consist of historical BIXI trip records obtained from BIXI's open data portal. These records include information on trip start times and originating stations. The raw trip-level data were aggregated to the station-hour level, where each observation represents the total number of trips starting from a given station during a specific hour.

BIXI's station information Data (source)

Station information data were used to provide contextual information about station locations and characteristics. This dataset includes geographic attributes such as latitude and longitude. Rather than using station identifiers as categorical variables, geographic coordinates were used as continuous features to allow the model to learn neighbourhood-level demand patterns while maintaining generalizability.

Weather Data (Meteostat API)

Weather conditions were retrieved using the Meteostat API and merged with the station-hour dataset based on timestamp alignment. Weather variables include temperature, precipitation, and wind speed. These features capture environment factors known to influence bike usage

Feature Engineering

A structured feature engineering pipeline was applied to transform the raw datasets into a modeling-ready format while avoiding data leakage

Temporal features:

To capture the cyclical nature of urban mobility, hour-of-day and day-of-week were transformed using sinusoidal functions (sin and cos transformations). This ensures the model recognizes that hour 23:00 and hour 00:00 are chronologically adjacent, preventing the "numerical jump" that occurs with raw integer encoding.

Lag and rolling demand features:

To model short term demand persistence, lagged demand features (e.g. demand in the previous hour) and rolling averages over multiple horizons (e.g. 3 hour and 24 hour windows) were engineered. These features enable the model to learn momentum and recent usage trends

Weather features:

Feels_like_temperature and a rain indicator were included to reflect the impact of environmental conditions on rider behaviour

Spatial features:

Station latitude and longitude were incorporated as continuous variables to capture geographic demand differences without relying on the station name.

All features were aligned chronologically to ensure that only information available prior to the prediction time was used. The final modeling dataset was stored as a parquet file and treated as immutable during the modeling phase to ensure reproducibility and consistency across experiments.

All preprocessing and feature engineering steps were implemented using reproducible pipelines, with fixed train–validation–test splits and version-controlled datasets to ensure consistency across experiments.

Modeling Approach

Problem Formulation

This forecasting task is framed as a supervised regression problem, where the target variable is the total number of trips originated from a station during a given hour. The goal is to predict this value using historical demand patterns, temporal structure, weather conditions, and spatial context

Train-Validation-Test Strategy

Since bike demand show strong temporal dependence, the dataset was split using a time-aware approach rather than random sampling. Observations were ordered chronologically and divided into training and validation, and test sets to reflect real world forecasting conditions and prevent information leakage from the future.

Baseline model

A naïve baseline model was used to predict the global mean demand. This model established a minimum performance threshold. This baseline provides a reference point for assessing whether more complex machine learning models offer meaningful improvements

Machine Learning Models

Two tree-based models were explored:

Random Forest Regressor:

Random Forest was selected for its ability to model non-linear relationships, handle mixed feature types, and provide robust performance with minimal preprocessing. It also supports model interpretability through SHAP analysis.

Histogram-based Gradient Boosting Regressor:

Histogram-based Gradient Boosting was evaluated as a more computationally efficient alternative. By discretizing continuous features into bins, this method significantly reduces training time and memory usage while maintaining strong predictive performance.

Both models were trained using the same features set and evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R^2, and training time. Model interpretability was conducted using SHAP values derived from the Random Forest model to understand feature importance and directional effects

Technical Note on Interpretability

While SHAP values were calculated using the Random Forest model for computational ease, the insights are considered representative of the final HGB model. Both are tree-based ensembles that converged on similar performance metrics, indicating they are leveraging the same underlying feature hierarchies to make predictions.

Model Performance Comparison

Model	MAE	RMSE	R^2	Training Time
Baseline (Global Mean)	3.22	4.6	-0.0011	0.0s
Random Forest [RF]	2.05	3.15	0.645	8 min and 41s
HGB Regressor (Selected)	**2.04**	**3.14**	**0.648**	**35.6s**

While an R^2 of **0.648** suggests there is still unexplained variance, this is a strong result for station-level hourly forecasting. Urban demand is subject to inherent "stochastic noise"—individual decisions that no model can perfectly predict. By reducing the Mean Absolute Error (MAE) by **36%** compared to the baseline, the HGB model successfully captures the primary drivers of Montreal's transit rhythm.

Model Interpretability & Insights (SHAP Analysis)

SHAP analysis was conducted using the Random Forest model to understand the key drivers of hourly BIXI demand. The results highlight the relative importance and directional effects of temporal, spatial, weather, and calendar features.

Hourly BIXI demand is primarily driven by recent usage and strong daily cycles, with spatial location and weather providing secondary adjustments and calendar effects playing a limited role.

Demand Inertia variables (rolling_24h, lag_1h, rolling_3h) are by far the most influential features. High recent usage consistently increases predicted demand, while low recent activity strongly suppresses it. This indicates strong temporal persistence in bike usage patterns: stations that are busy tend to remain busy in the near future.

Hour-of-day features (hour, hour_sin, hour_cos) rank among the most important non-lag predictors. The non-linear SHAP patterns show daily rhythms, capturing the morning and the evening peak demand associated with commuting hours. This shows that the model successfully learned the cyclical daily usage patterns rather than relying on the raw hour values alone.

Latitude and **longitude** show moderate SHAP importance, indicating that geographic location influences demand levels. These features capture neighbourhood-level differences (e.g. downtown vs. residential areas) and acts as a proxy for station characteristics, without relying on the station name itself.

Weather variables such as **temperature**, **precipitation**, and **rain indicators** have smaller SHAP values compared to temporal features. Higher temperatures generally increase demand, while precipitation and rain slightly reduce it. Weather effects do not override strong temporal and historical demand patterns.

Calendar-related features (day_of_week, is_weekend, month) contribute modestly to predictions. While weekday/weekend differences exist, their impact is smaller than demand inertia and cyclical daily usage patterns. This suggests that hourly usage patterns are more strongly driven by routine behavior than by the broader calendar structure.

Model Diagnostics & Error Analysis

The diagnostic analysis confirms that the model generalizes well across typical demand levels, with larger errors primarily concentrated in rare high-demand scenarios

To assess the reliability and limitations of the final model, we examined prediction errors on the test set using an actual vs predicted plots and the residual distribution.

Actual vs. Predicted demand

Looking at an actual vs predicted scatter plot, we can see the relationship between the predicted and observed demand values. Most observations cluster closely around the diagonal line, indicating that the model produces accurate predictions across the majority of demand levels

At higher demand levels, predictions tend to slightly underestimate extreme peaks. This is expected in demand forecasting problems where sometimes rare surges are more difficult to predict and models are naturally biased toward the central range of observed values.

Overall, the plot suggests that the model captures the general structure of demand well, with not systematic breakdown across low or moderate demand levels.

Residual Distribution

The residual distribution is centered close to zero, indicating that the model does not exhibit systematic bias toward over or under predictions. Most prediction errors fall within a narrow range, indicating that they are small and stable. There are a small number of large residuals in the tails of the distribution which reflects the extreme demand situations. These outliers are consistent with the sudden spikes caused by unobserved factors such as events or abrupt behavioural changes, which are not explicitly modeled.

Temporal Backtesting

To simulate real-world deployment, a temporal walk-forward backtesting approach was employed. Unlike standard cross-validation, this method respects the arrow of time by training the model on historical data (2024) and evaluating it on a strictly subsequent, unseen period (2025).

To prevent data leakage, all dynamic features—including lagged demand and rolling averages—were computed using only information available prior to each specific prediction timestamp. This ensures the model is not "peeking" into the future to make its forecasts.

The backtest yielded the following results:

- **MAE:** ~1.97
- **RMSE: ~**3.12
- **R^2:** ~ 0.65

The fact that these metrics closely align with (and even slightly exceed) the initial test-set performance confirms the model's robustness. It demonstrates that the patterns learned from 2024—such as commuting cycles and weather sensitivities—remain stable and highly predictive for the 2025 season, providing confidence for its use in live operational environments.

Business Interpretation & Practical Implications

With an average error of roughly two trips per station per hour, the model provides sufficiently accurate forecasts to support short-term rebalancing and capacity planning decisions in a bike-sharing system.

The final model achieves a Mean Absolute Error (MAE) of approximately 2 trips per station per hour on the test set. In practice, this means that for a given station and hour, the predicted demand typically deviates from actual usage by only two bike trips

For a bike sharing system operating at an hourly resolution, this level of accuracy is operationally meaningful. Many stations experience demand in the range of 10–40 trips per hour during peak periods, making an error of two trips small relative to total usage. This enables the model to provide reliable short-term forecasts that can support decision-making rather than merely descriptive analysis.

Operational Use Cases

This model’s prediction can support several operational and planning applications for bike sharing operator:

1. Rebalancing and fleet management

Hourly demand forecasts allow operators to anticipate stations that are likely to experience shortages or surpluses, enabling more proactive rebalancing decisions rather than reactive responses.

2. Capacity planning

Station with consistent high predicted demand can be identified as candidates for capacity expansion, while low demand stations may require fewer docks or bikes

3. Staffing and logistics optimization

Accurate demand forecasts help allocate rebalancing resources more efficiently, particularly during peak commuting hours when demand variability is highest.

Strategic Insights

Other than operation usage, the model provides strategic insights into usage behavior. Since demand is strongly influence by recent demand and intraday cycles suggests that bike usage follow a stable routine pattern, while weather and calendar effects play a secondary role. This suggests that operational strategies should primarily focus on time-based patterns and historical usage trends, using weather forecasts as a supplementary adjustment rather than a primary driver.

Limitations & Future Work

Limitations

Although the model show strong predictive performance, it has some limitations that should be acknowledged.

First, the model relies primarily on historical demand, temporal pattern, and weather variables. While these features capture the dominant structure of bike-sharing demand, external factors such as festivals, road closures, public transit disruptions or strikes are not explicitly modeled. These events can lead to sudden demand spikes and difficult to predict using historical patterns alone.

Second, the model does not account for operational feedback, such as bike rebalancing actions or temporary station outages. Operations will occasionally directly affect the bike availability – ignoring these interventions may limit the model's accuracy during peak period of heavy operational activity.

Third, spatial effects are represented using latitude and longitude, which captures the broad neighbourhood-level differences. However, it does not fully reflect local infrastructure such as proximity to metro stations, bike lanes, or points of interests.

Finally, the model was conducted at an hourly aggregation level, which smooths short-term fluctuations in demand. While it is appropriate for operational planning, more granular temporal dynamics (example 15-minutes interval) are not captured.

Future Work

First, include event-based and contextual data such as public events, holidays or transit disruptions could improve predictions during abnormal demand periods.

Second, spatial modeling could be refined to include stations into functional zones (e.g. downtown, residential, campus areas) or by engineering features based on distance to transit hubs and major points of interests.