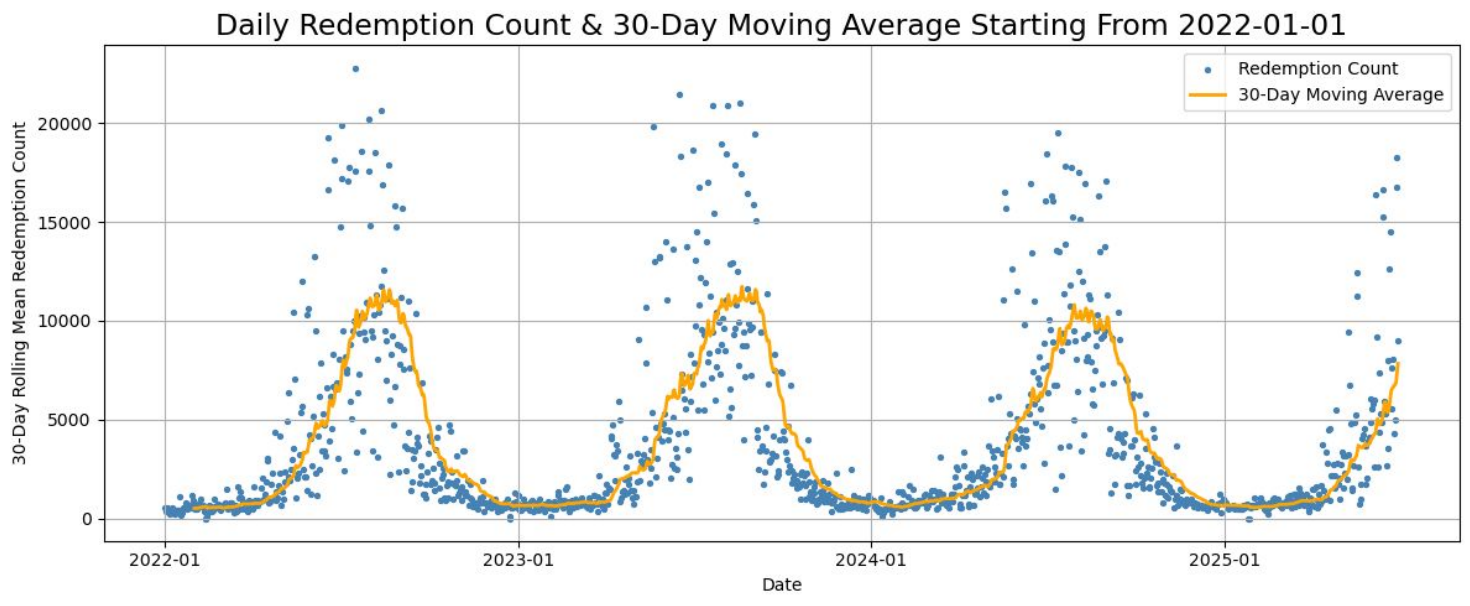
Forecasting Ferry Ticket Sales

Time Series Forecasting | LightGBM vs Prophet (Meta)
Toronto Islands ferry ticket redemptions

Reaz Elias - Data Scientist

Daily Redemption Count (Tickets scanned at ferry gates)



- ❑ Redemption Count: Tickets scanned at ferry gates (actual usage) from 2022-2025
- ❑ 30-day moving average shows trends from 2022-2025
- ❑ Summer peaks dominate ferry usage
- ❑ Winter lull with minimal ridership
- ❑ Sudden peaks/Spikes on some days

Reaz Elias - Data Scientist

Model 1: LightGBM (Gradient Boosting Trees)

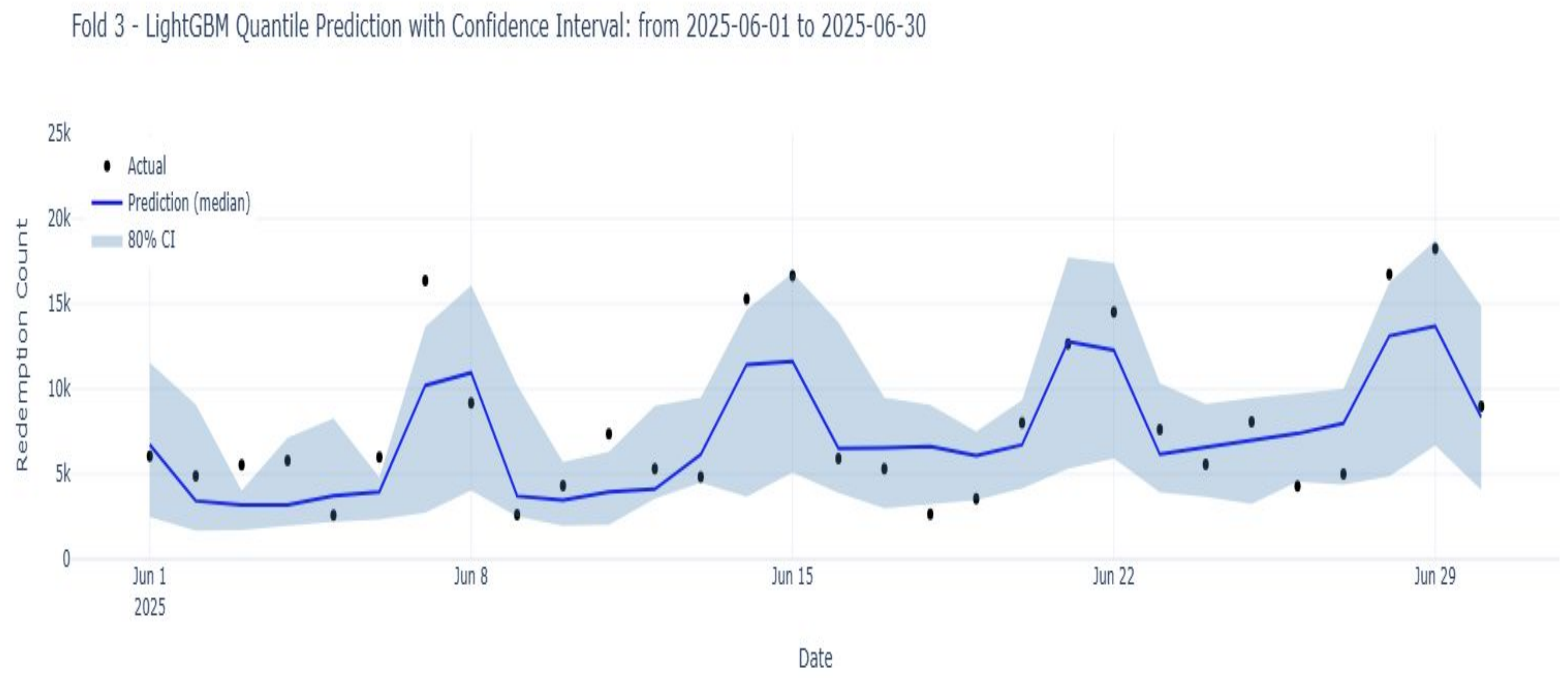
- **Gradient Boosting** uses decision trees as weak learners and builds them sequentially, with each new tree trained to correct the residuals (errors) of the previous one by minimizing a loss function (e.g., mean squared error) using gradient descent.
- **LightGBM** (**Light Gradient Boosting Machine**) is a faster and more memory-efficient version of gradient boosting. Like XGBoost, it builds decision trees one after another, each trying to fix the mistakes of the previous one.
- **Grows trees leaf-wise instead of level-wise:** XGBoost grows trees level by level (more balanced), while LightGBM grows them leaf by leaf, choosing the split that reduces error the most - even if it makes the tree deeper on one side. This often gives better accuracy.
- **Faster training on large datasets:** LightGBM is optimized for speed. It uses techniques like histogram-based binning, which groups similar values into buckets to speed up computation.
- Handles **large-scale data efficiently:** It uses less memory and works well even when dataset is big or has many features.

LightGBM with **quantile regression**

In this project, I've used:

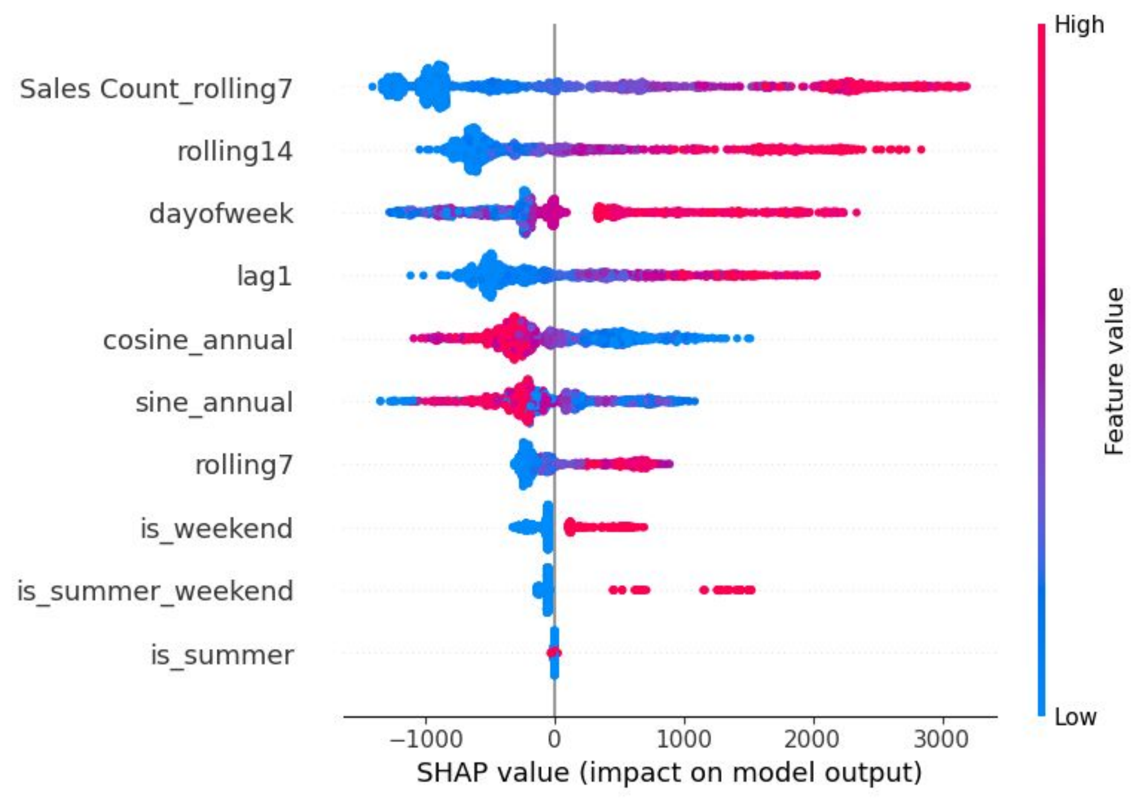
- ❏ **Advanced Feature Engineering:** Flags for weekends, holidays, and seasonal patterns. Engineered lag features and rolling window statistics to capture recent trends.
- ❏ **TimeSeriesSplit Cross-Validation:** Ensured model evaluation respects temporal order, avoiding data leakage and mimicking real-world forecasting setup.
- ❏ **Hyperparameter Tuning with Optuna:** Automatically searched for optimal learning rate, tree depth, and number of leaves using Optuna's efficient optimization.
- ❏ **Quantile Regression for Confidence Intervals:** Trained three models to predict the 10th, 50th (median), and 90th percentiles, giving a 80% prediction interval to quantify forecast uncertainty.

LightGBM Quantile Prediction: Jun 1- 30, 2025 (MAE: 2179)



Reaz Elias - Data Scientist

Feature Importance: What Drives the Median Prediction?



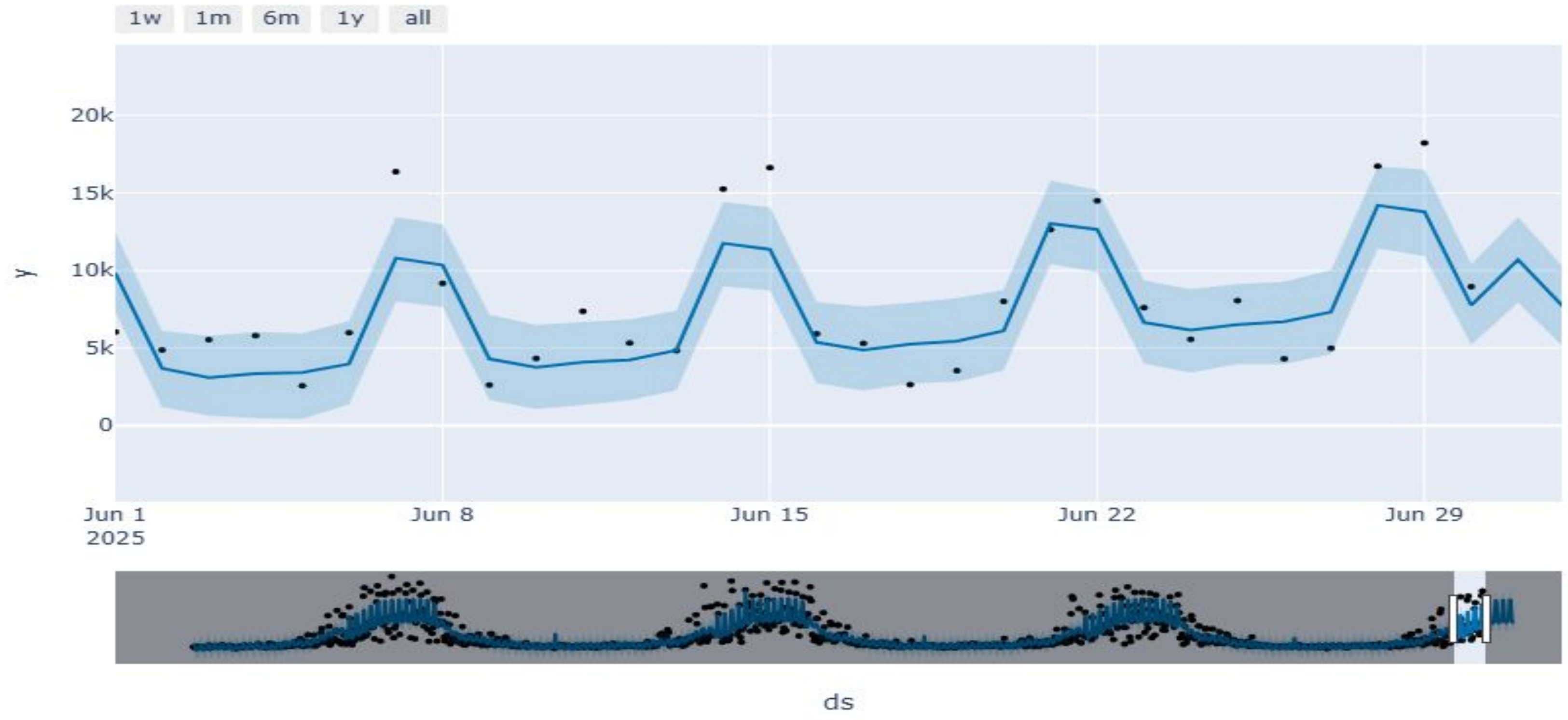
Reaz Elias - Data Scientist

Model 2: Prophet by Meta

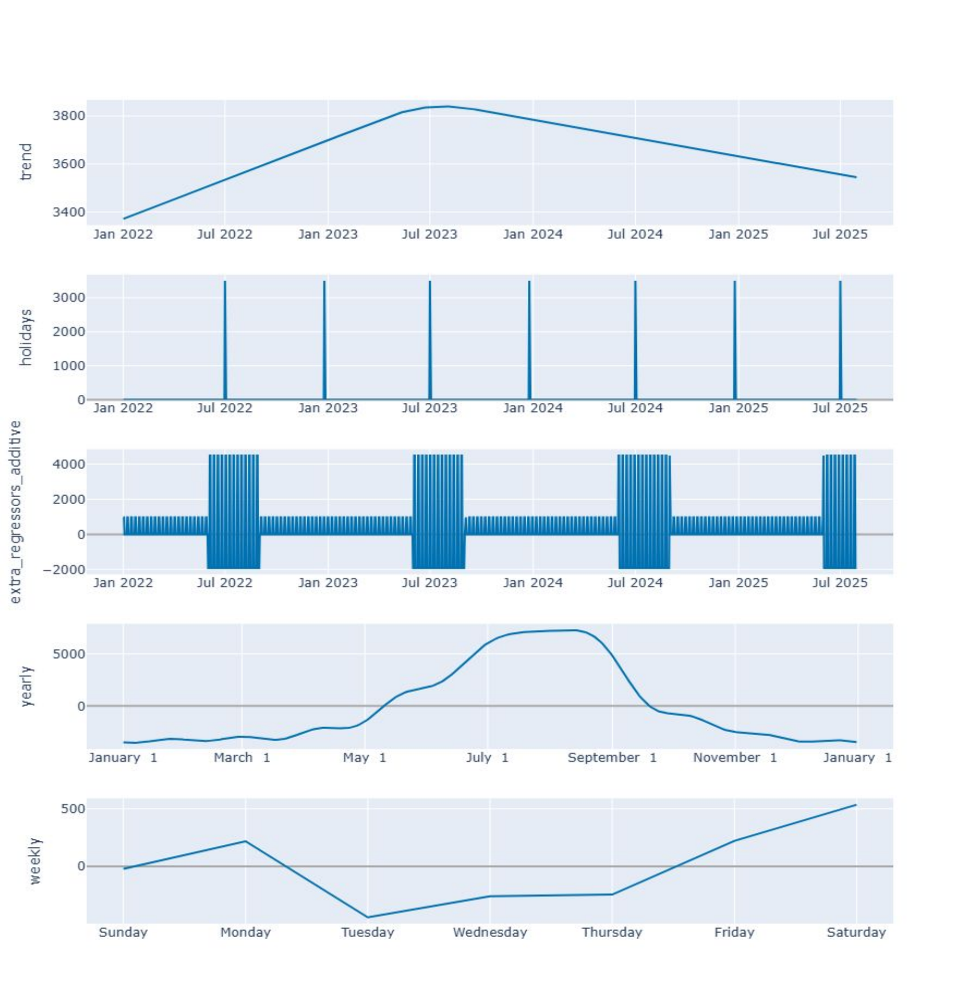
Prophet is an open-source time series forecasting tool developed by **Meta (Facebook)**. It's designed for **business analysts and data scientists** who need to produce accurate forecasts quickly and with easy interpretability.

- **Handles complex seasonality**: Automatically models **daily, weekly, and yearly patterns** in the data.
- **Built-in support for holidays**: Specify country-specific or custom holiday calendars, adjusting forecasts for expected spikes or dips.
- **Robust to missing data and outliers**: Handle gaps in data and irregular time series without requiring heavy preprocessing.
- **Human-friendly parameters**: Offers intuitive parameters for trend changes, holidays, and seasonality - great for explaining models to non-technical stakeholders.

Prediction using Prophet: Jun 1- 30, 2025 (MAE: 2017)



Trend and Seasonality Components from Prophet



Conclusion

	Prophet (by Meta)	LightGBM
Strength	Time series domain knowledge (trend, seasonality, holidays)	Flexible machine learning model for complex patterns
Interpretability	High - clear trend/seasonality components	Medium - explainable via SHAP values
Use Case	Captures global patterns and calendar effects	Works well with rich feature sets (e.g. weather, weekdays, promotions)
Integration	Excellent for incorporating **domain features** (e.g. holidays, weekends)	Great for modeling **interactions and nonlinearities** in tabular form.