AI&ML TEAM 1

Beyond the Meter: AI-Powered Solutions for NYC's Congestion Surcharge

Afiqah, Alexander, Anirudh, Keuntae

Client: NYC Taxi and Limousine Commission (TLC)

Agenda

This project leverages AI-driven prediction models to redesign congestion surcharge policies, aiming for smarter traffic management and equity across NYC boroughs.

TRANSFORMING TAXI POLICY THROUGH AI

NYC
Taxi & Limousine Commission

Objective

Implement AI-driven, dynamic congestion surcharge strategies to balance traffic management, policy equity, and revenue sustainability in NYC's evolving transportation ecosystem.

Why it matters?

1. NYC traffic congestion has risen post-pandemic.
2. Surcharges fund critical Metropolitan Transportation Authority (MTA) services (~$400 million projected annually).

Project highlights

1. Built predictive classification and regression models using 2025 Yellow Taxi TLC trip data.
2. Designed dynamic, zone- and time-sensitive surcharge policies.
3. Achieved a +0.90% projected revenue uplift with improved fairness across boroughs.

GUIDING QUESTIONS BEHIND OUR ANALYTICAL FRAMEWORK

Optimization Challenge

How can congestion surcharge structures be redesigned to minimize peak-hour congestion while ensuring equitable passenger treatment across socioeconomic zones?

Predictive Modeling Goals

Can AI accurately forecast when surcharges should be triggered, and what surcharge amount would maximize fairness and revenue simultaneously?

Dynamic Pricing Exploration

Should surcharge levels vary by time of day and by traffic congestion intensity in different zones to better reflect real demand and usage patterns?

INTELLIGENT DATA INTEGRATION FOR ROBUST MODELING

Primary Data

NYC TLC Yellow Taxi Trip Records (January 2025, 3.4M trips).

TLC Taxi Zone Lookup Table (LocationID to Neighborhood Mapping).

Cleaning Strategy

Removed trips with invalid or missing congestion surcharge data.

Filtered out unrealistic trips (e.g., trip durations ≤0 or >120 minutes).

Feature Engineering

Derived trip_duration, pickup_hour, trip_distance and binary has_surcharge.

Merged spatial attributes for geospatial equity analysis.

FOUNDATIONAL INSIGHTS INTO SURCHARGE DYNAMICS

Temporal Analysis

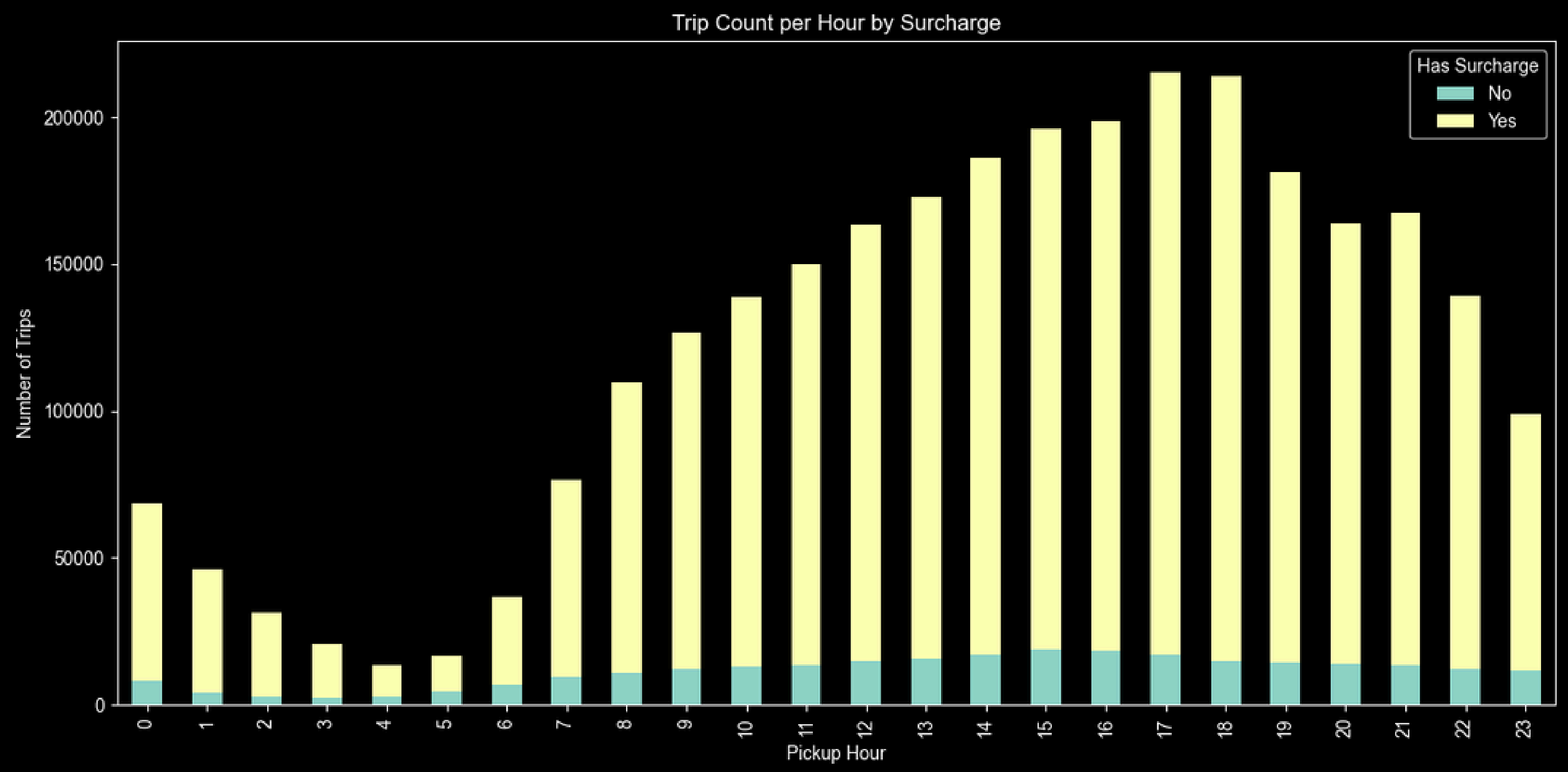
Surcharge mostly applied from 8 AM to 11 PM, peaking around 5–6 PM, aligning with NYC's commuter rush periods.

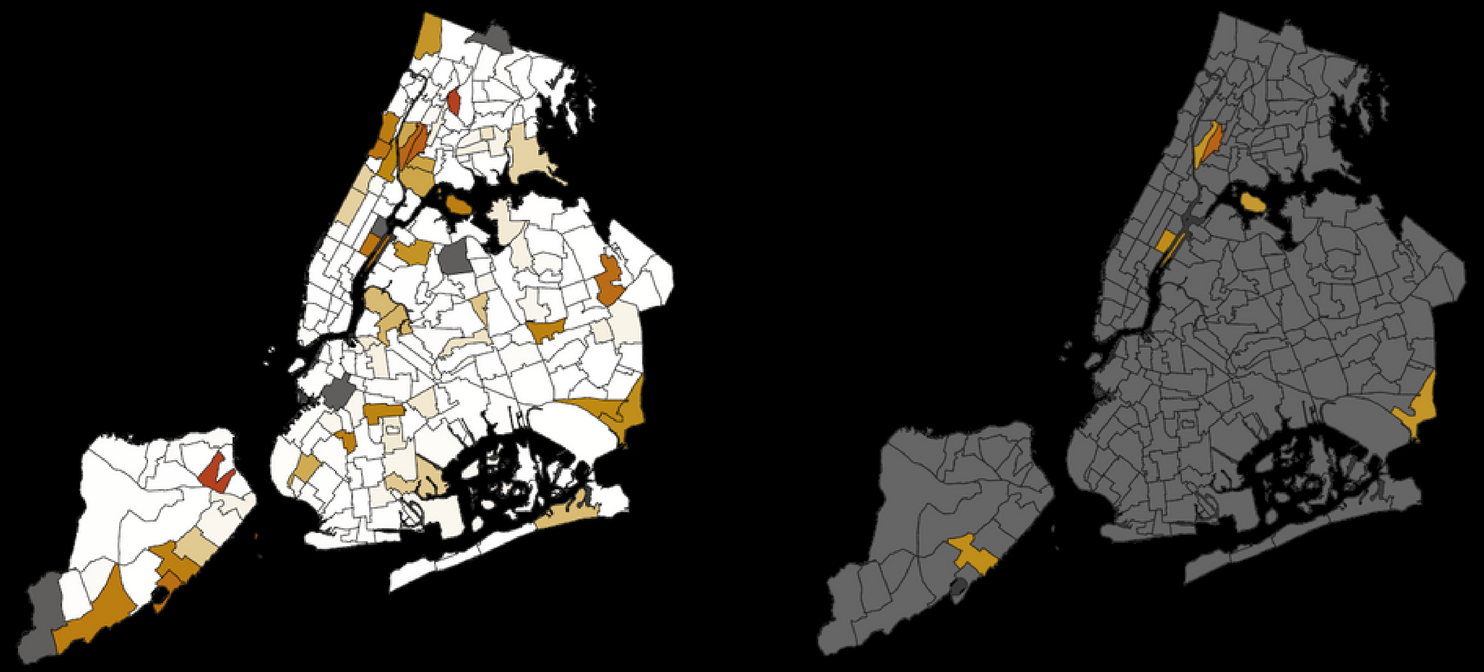
Spatial Analysis

Midtown and Uptown zones (Penn Station, Central Park South) had >90% surcharge application rates.

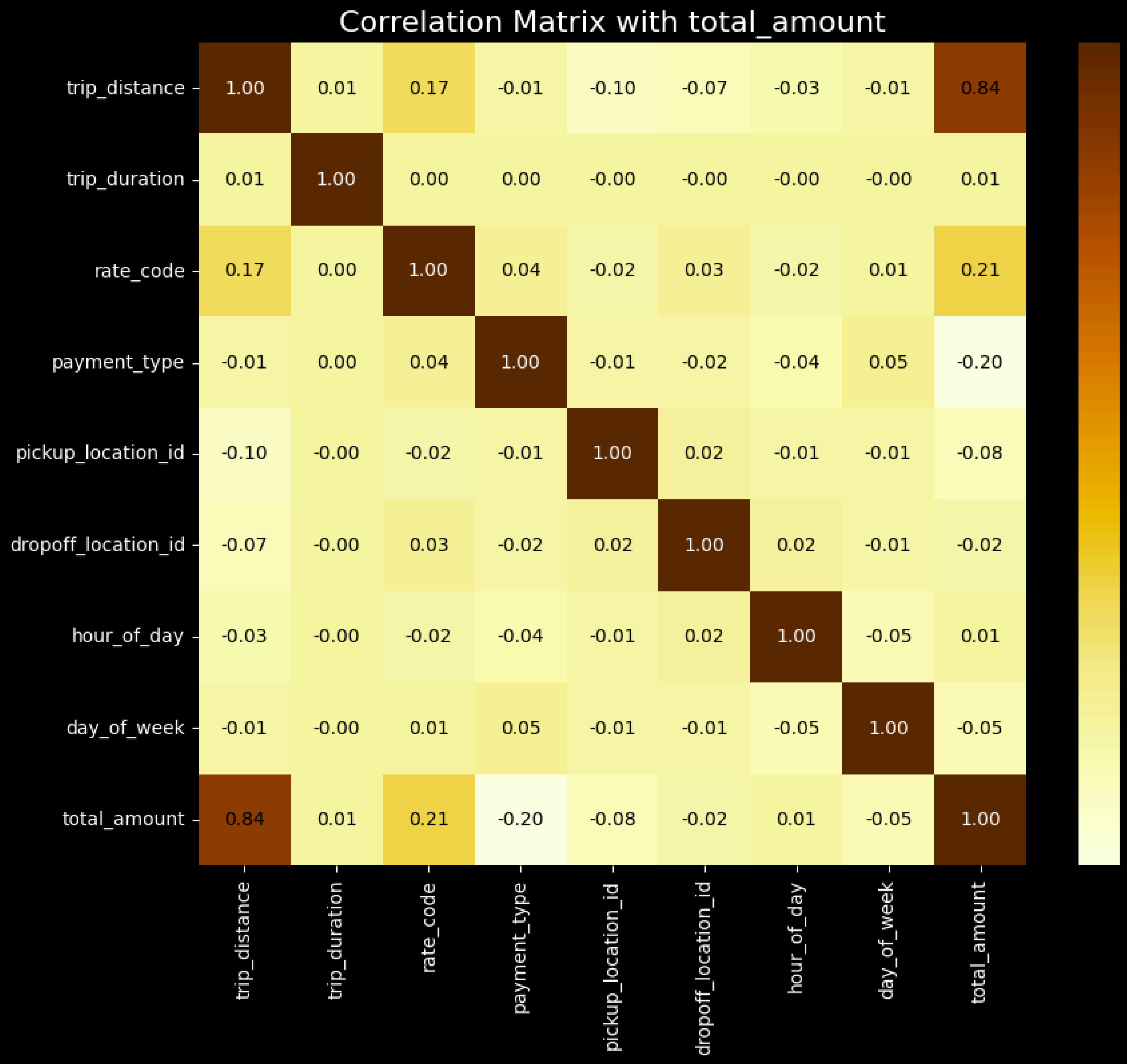
Equity Concern

Peripheral zones (e.g., Bronx, Eastern Queens) show inconsistent surcharge application despite similar congestion contexts.





Temporal and spatial biases suggest the need for a more targeted and adaptive surcharge framework.



Small Questions

 # 1 What trip factors (e.g., distance, duration) best predict fare amounts?

 # 2 How can congestion surcharges be optimized to reduce congestion and ensure fairness?

 # 3 Can AI-driven dynamic pricing achieve policy goals effectively?

MACHINE LEARNING TO PREDICT SURCHARGE APPLICATION

Model Comparison

LOGISTIC REGRESSION	RANDOM FOREST CLASSIFIER	GRADIENT BOOSTING CLASSIFIER
Accuracy: 91% ROC AUC: 0.913 Precision for Surcharge Class: 0.982 Recall for Surcharge Class: 0.914	Accuracy: 85% ROC AUC: 0.860 Precision for Surcharge Class: 0.969 Recall for Surcharge Class: 0.868	Accuracy: 87% ROC AUC: 0.863 Precision for Surcharge Class: 0.972 Recall for Surcharge Class: 0.879
Best model for policy explainability and regulatory reporting	**Good model for feature importance exploration (especially via SHAP values)**	**Slower training compared to Logistic Regression but better at capturing minor zone-level effects**

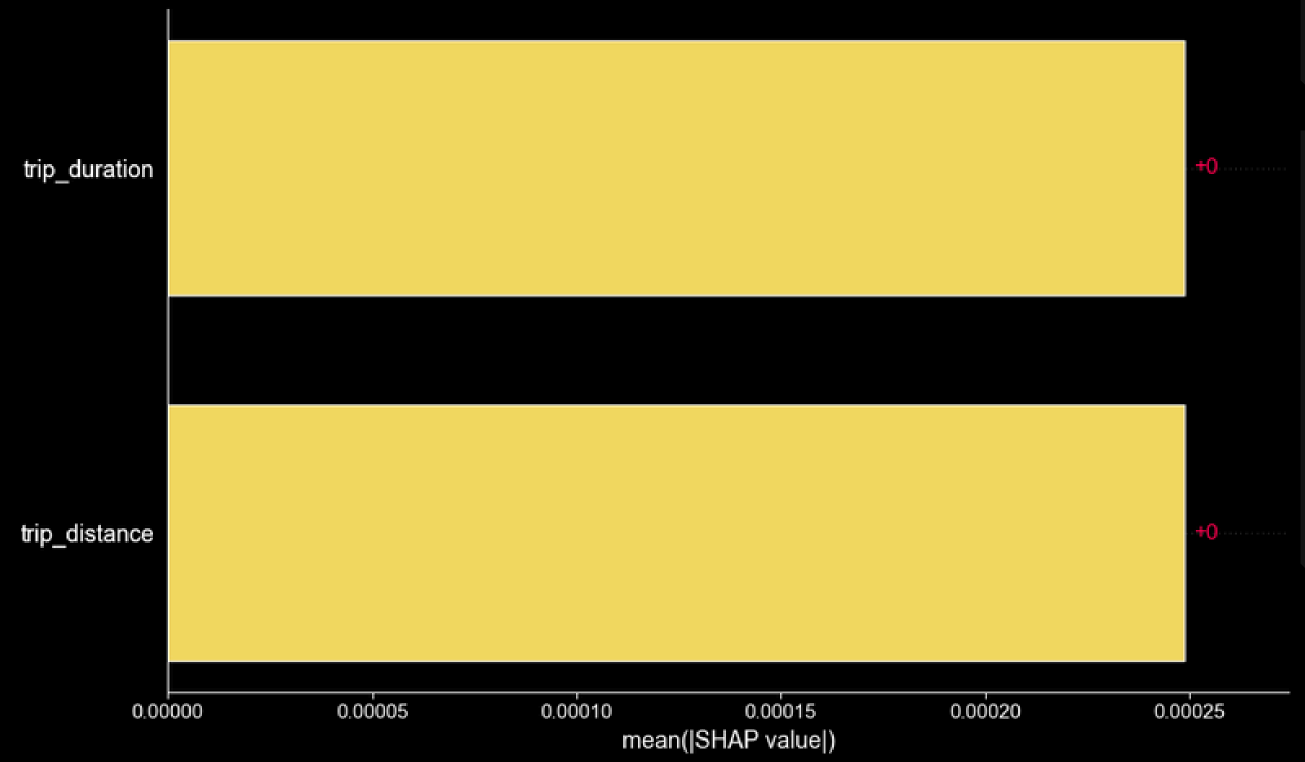
MACHINE LEARNING TO PREDICT SURCHARGE APPLICATION

Model Comparison

Feature Importance (Logistic Regression Coefficients)

Top predictive factors:

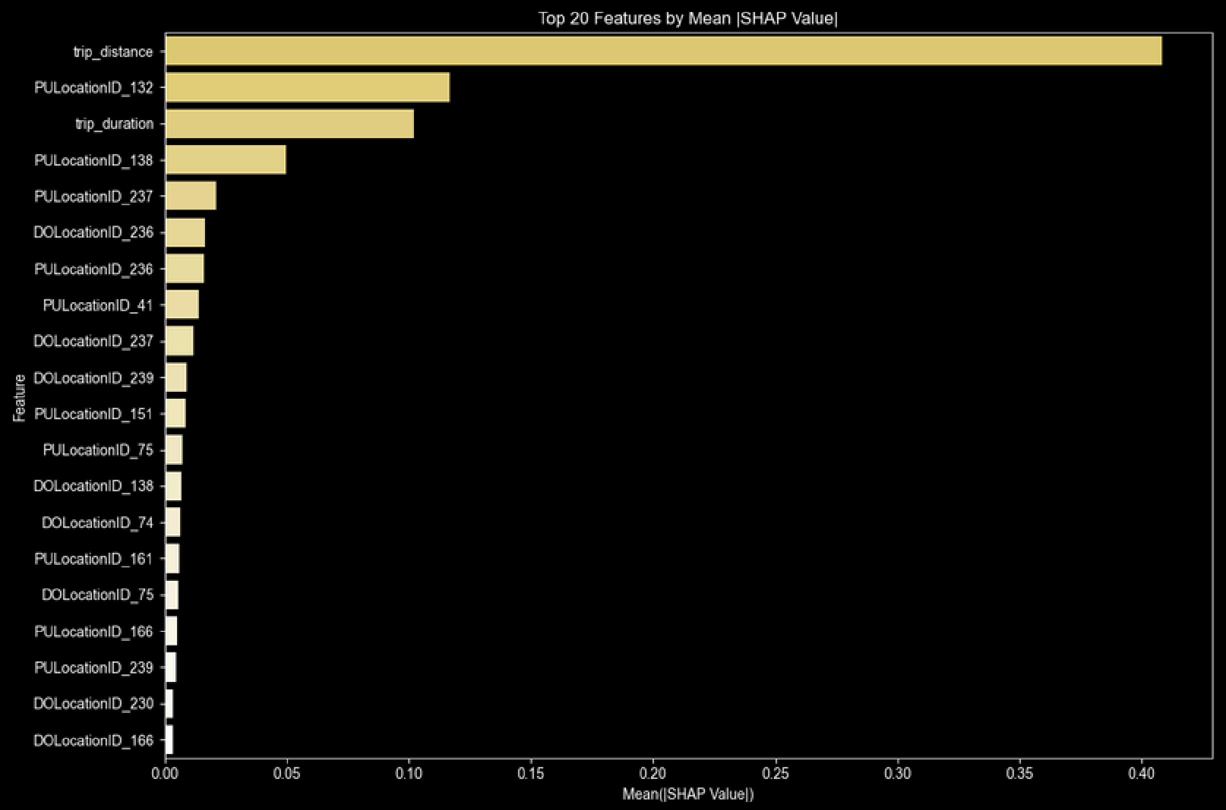
1. Trip Distance (Positive correlation)
2. Pickup Hour (Higher surcharge probability during daytime/peak hours)
3. Specific Pickup Zones (e.g., Zone 132 - JFK Airport)

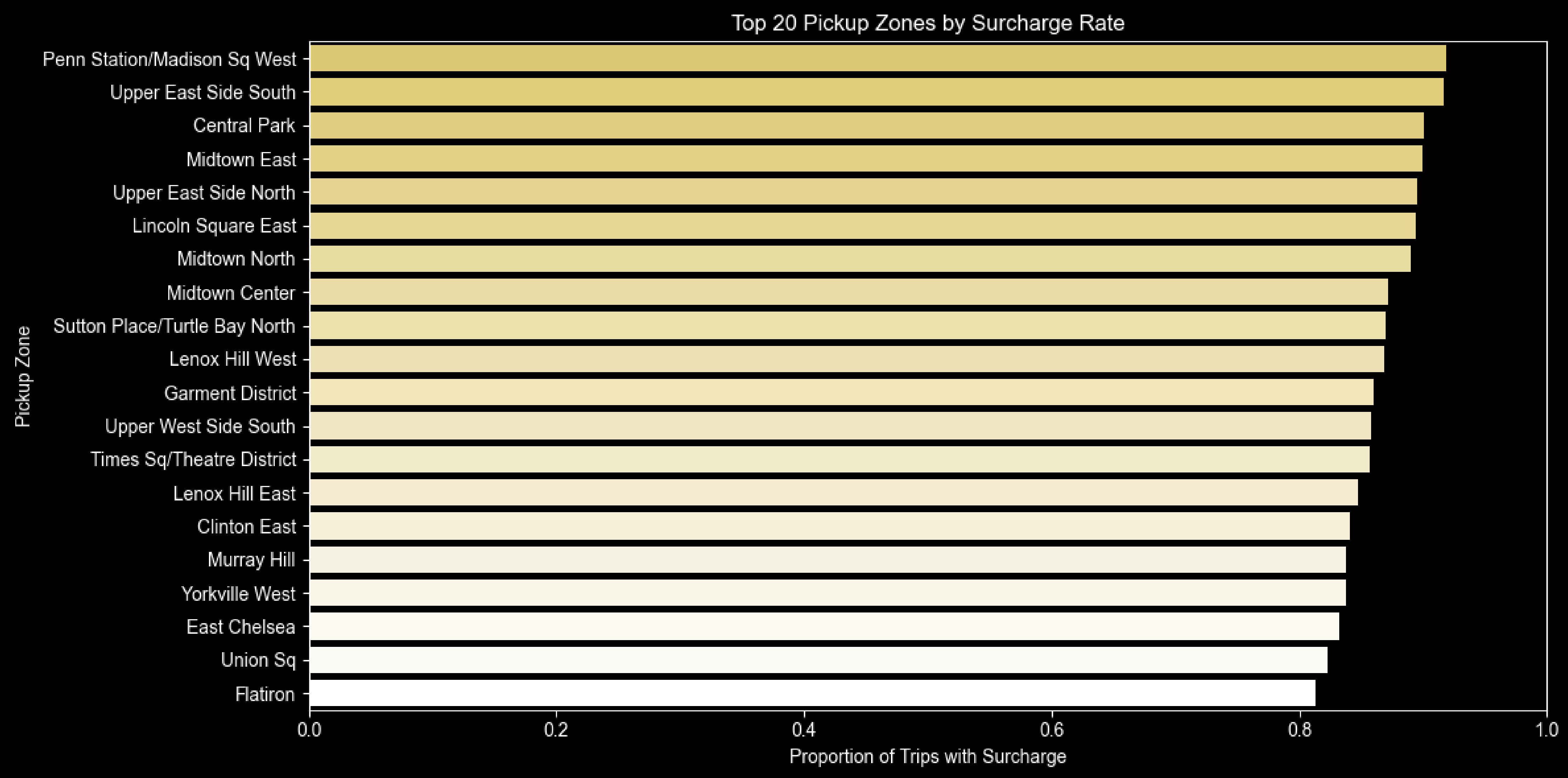


SHAP Analysis (Random Forest Model)

SHAP values revealed the true "impact ranking" of features:

1. Trip Distance: Most influential, consistently increasing surcharge probability as distance grows.
2. Trip Duration: Important but less dominant than distance.
3. Pickup Location: Certain zones increase surcharge likelihood e.g. 132 (JFK Airport), 236 (Upper Easat Side North), 161 (Midtown Center).
4. Passenger Count and Day of Week had marginal impact compared to trip geometry factors.





Most high-surcharge zones are in Midtown and Uptown Manhattan—areas with heavy traffic, tourism, and commuter flow.

Zones like Penn Station, Central Park, and Upper East Side show surcharge rates above 90%, indicating consistent congestion.

This pattern suggests that surcharges are effectively concentrated in NYC's busiest core zones, aligning with traffic management goals.

Small Questions

 # 1 What trip factors (e.g., distance, duration) best predict fare amounts?

 # 2 How can congestion surcharges be optimized to reduce congestion and ensure fairness?

 # 3 Can AI-driven dynamic pricing achieve policy goals effectively?

FORECASTING SURCHARGE MAGNITUDES FOR FAIRNESS

With models that are well-suited for structured data like taxi trips, where both linear trends and complex interactions may exist.

	MAE	R^2
Linear Regression	**0.307**	**0.218**
Ridge Regression	**0.292**	**0.314**
Random Forest	**0.329**	**0.215**
Gradient Boosting	**0.338**	**0.248**

To predict the congestion surcharge amount, we compared four different regression models:

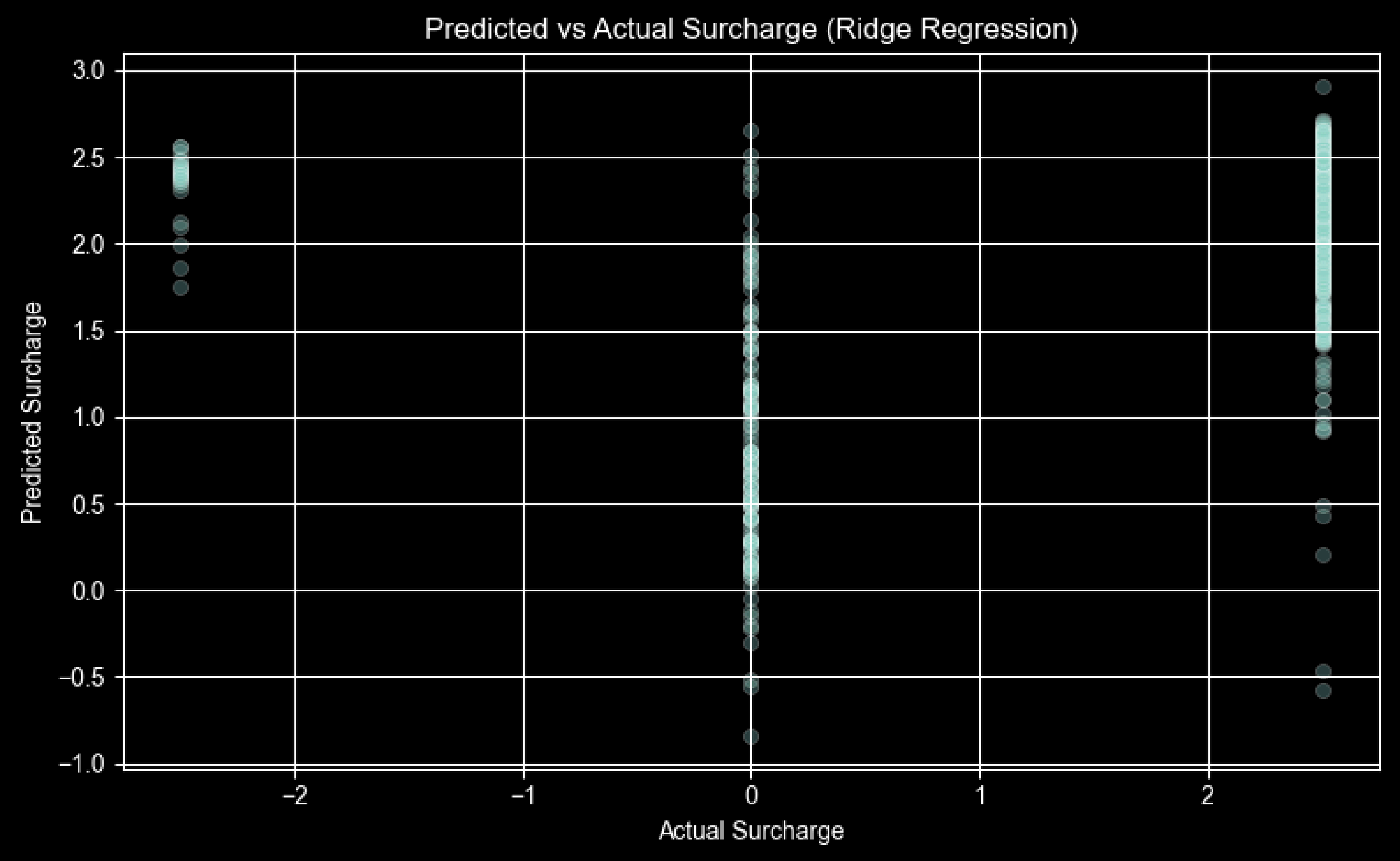
Linear Regression: Simple and easy to interpret
Ridge Regression: Handles many features better by reducing overfitting
Random Forest: Captures complex, nonlinear patterns in trip data
Gradient Boosting: Focuses on improving hard-to-predict cases

Ridge Regression performed the best, with the lowest MAE and highest R^2

90% of trips have fixed surcharges ($2.50 or $0), limiting model variance.

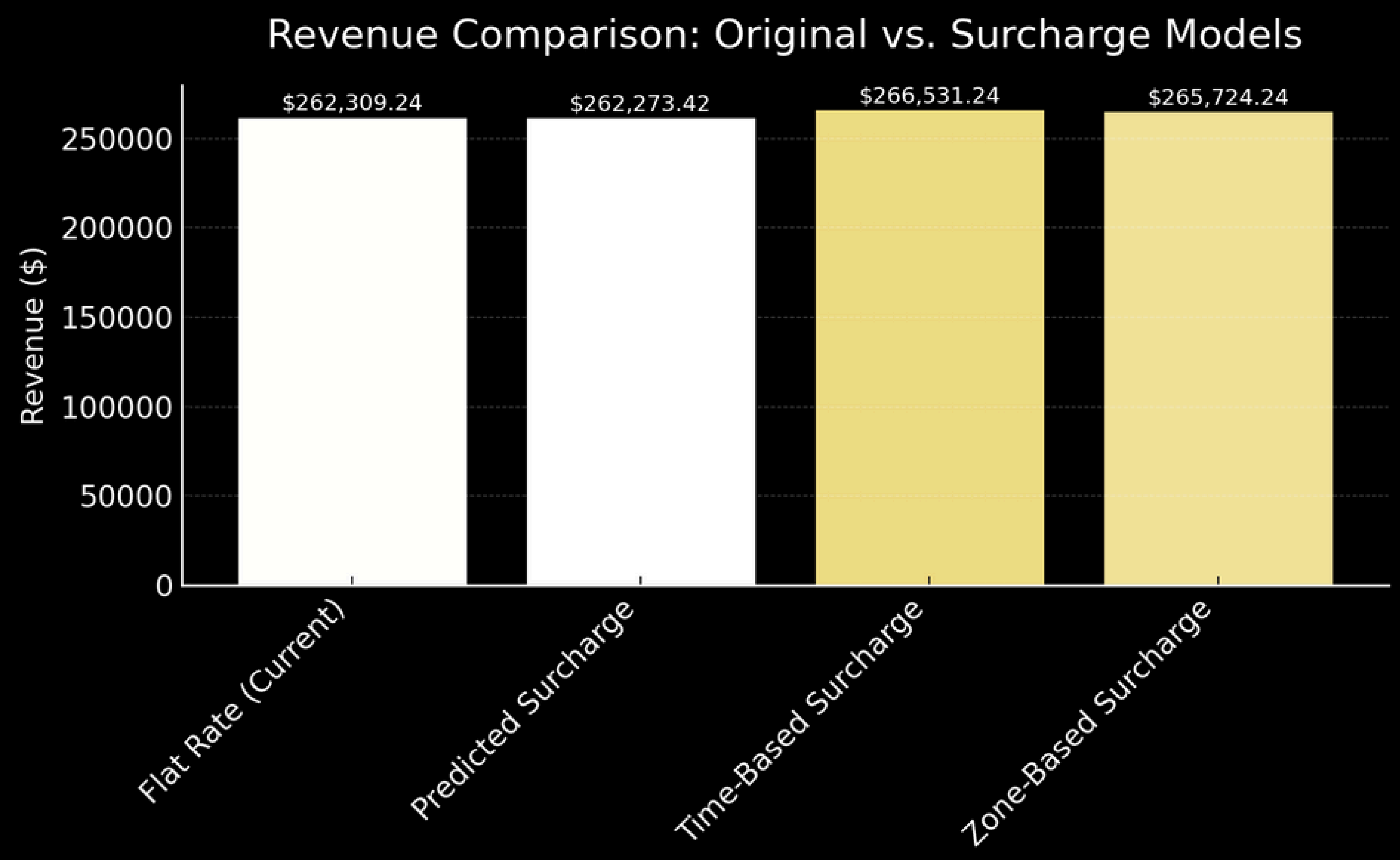
Policy Impact

Longer trips and pickups from congestion-heavy zones had the highest likelihood of surcharges.



SIMULATING FUTURE POLICY SCENARIOS BASED ON AI MODELS

POLICY	REVENUE	CHANGE ($)	CHANGE (%)	KEY TAKEAWAY
Flat Rate (Current)	$262,309.24	—	—	Stable, but lacks flexibility.
Predicted Surcharge	$262,273.42	-$35.82	-0.01%	Negligible change in city revenue (−0.01%), showing the limits of naive dynamic models.
Time-Based Surcharge	$266,531.24	+$4,222.00	+1.61%	Effective for boosting peak-hour revenue by applying higher surcharges during 7–9 AM and 5–7 PM.
Zone-Based Surcharge	$265,724.24	+$3,415.00	+1.30%	Balances revenue with targeted congestion by selectively applying premiums in congested hubs (Zones 132, 236, 161).



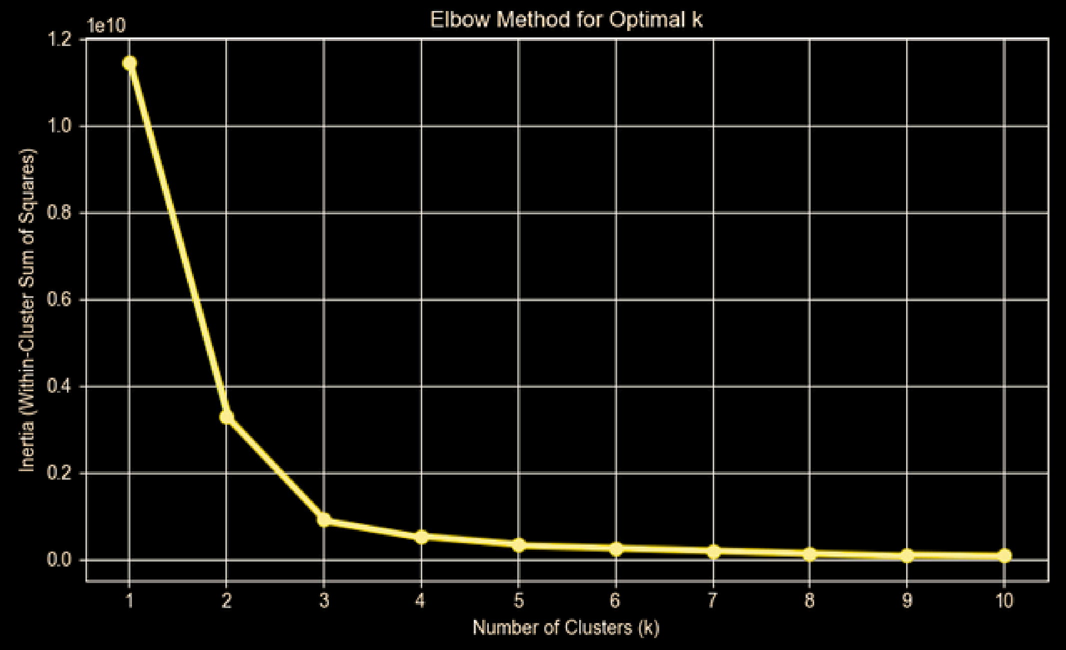
Dynamic adjustments by time and zone are financially and operationally superior.

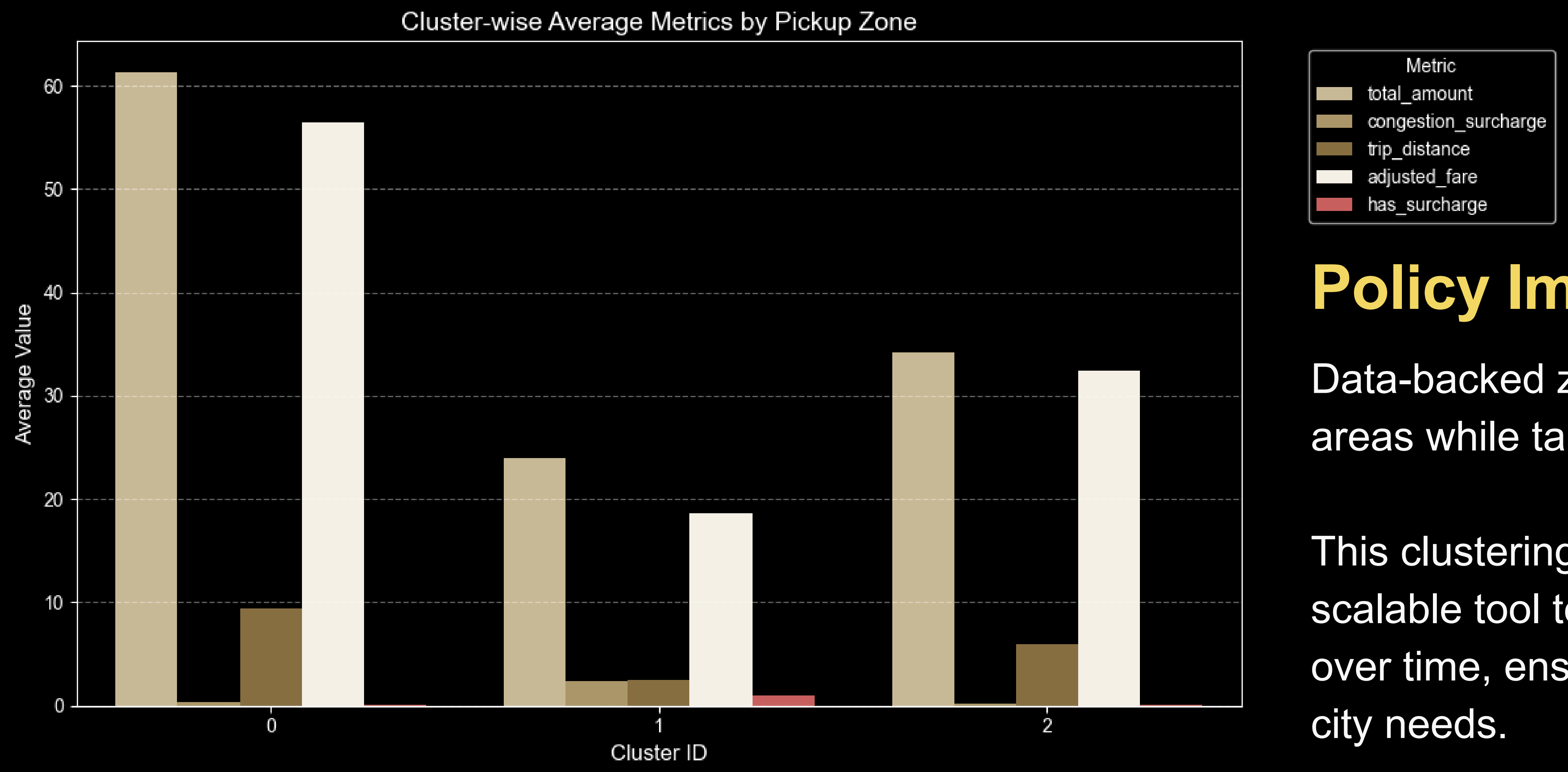
SMARTER ZONE POLICIES THROUGH CLUSTERING ANALYTICS

K-Means clustering on zone-level averages for fare, trip distance, and surcharge prevalence.

Identified Groups:

- Cluster 0: Long Distance, Premium Zones (e.g., Airports) → Fixed surcharge.
- Cluster 1: Local Trips, High Surcharge Zones → Reduced/waived off-peak surcharge.
- Cluster 2: Mixed Commuter Zones → Time-variable surcharge (higher peak, lower off-peak).





Optimal K Value

In order to design even smarter, more targeted policies.

Policy Impact

Data-backed zoning prevents overcharging residential areas while targeting congestion hubs.

This clustering framework gives policymakers a powerful, scalable tool to dynamically adapt the surcharge system over time, ensuring it stays fair, efficient, and responsive to city needs.

Small Questions

1 What trip factors (e.g., distance, duration) best predict fare amounts?

2 How can congestion surcharges be optimized to reduce congestion and ensure fairness?

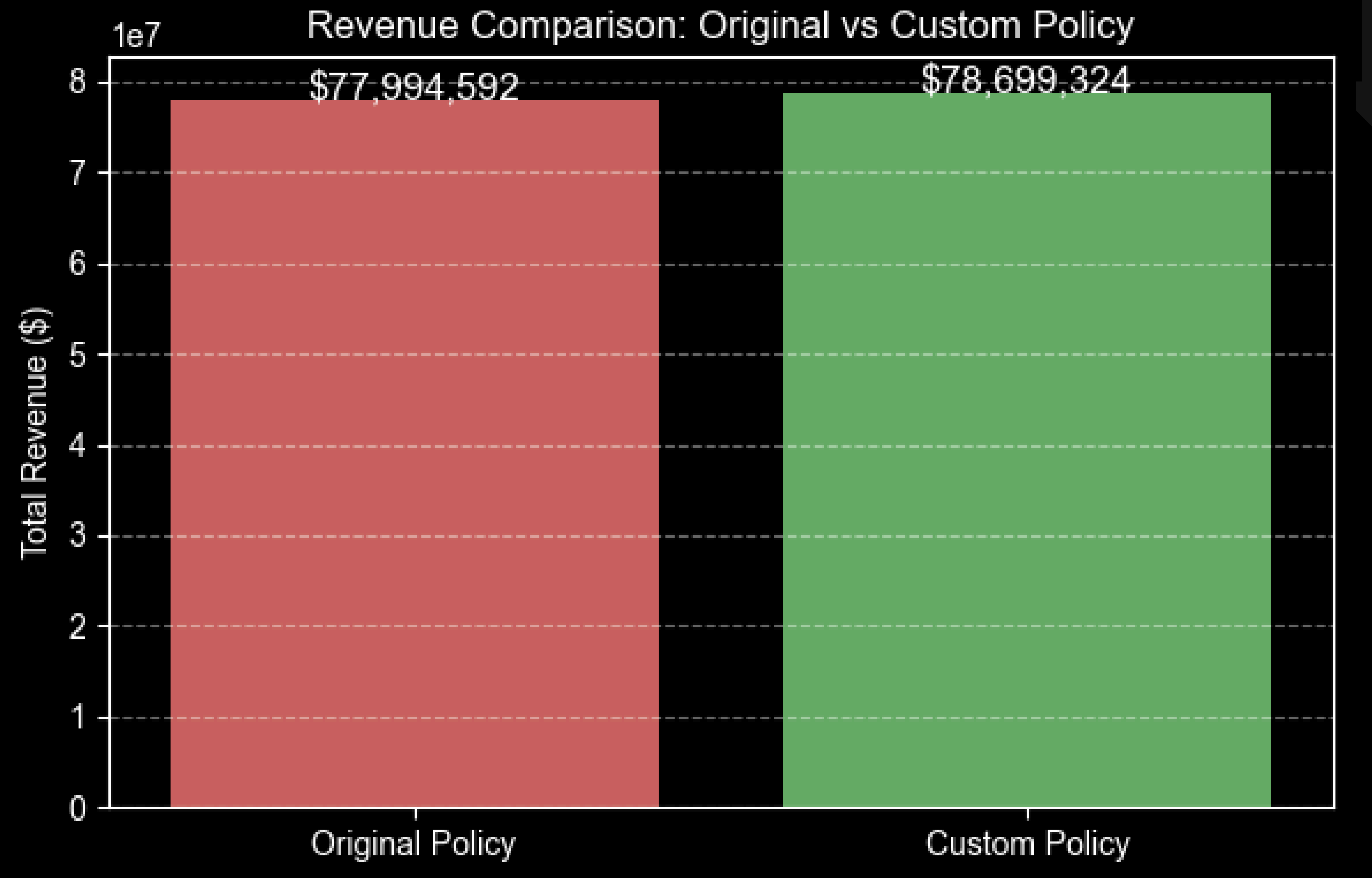
3 Can AI-driven dynamic pricing achieve policy goals effectively?

PROJECTED OUTCOMES FROM CLUSTER-BASED PRICING POLICIES

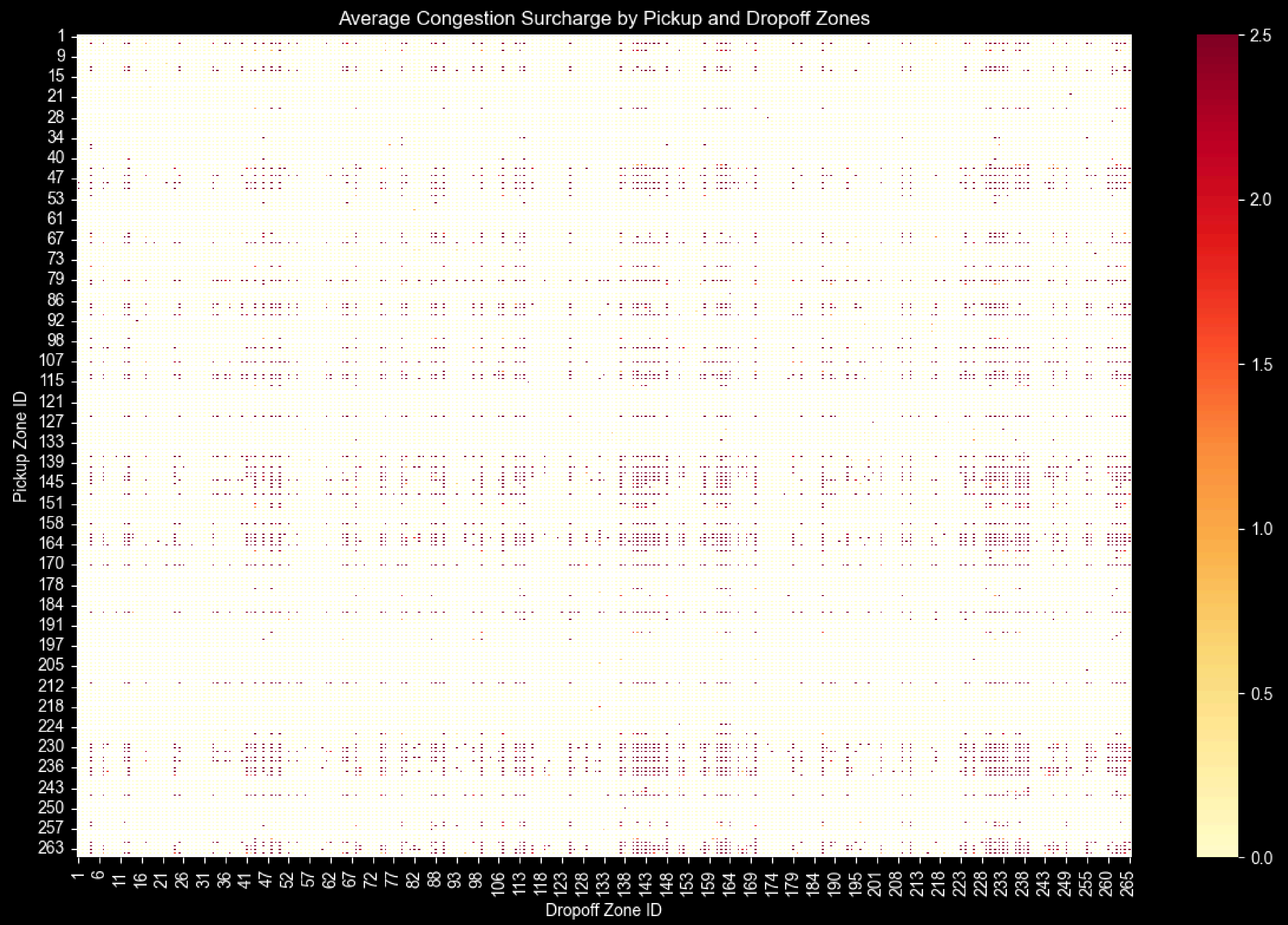
Dynamic Pricing for Real-Time Optimization

Project and Initiative	
Original Policy Revenue: $77.99M Projected Revenue under Custom Policy: $78.70M	Potential for improved neighborhood equity Expected stronger congestion management during peak hours
Net Gain: +$704,732 (+0.90%)	

A simple three-cluster pricing model yields financial, operational, and social benefits.



HOW AI ILLUMINATES TAXI SYSTEM BEHAVIOR



Key Findings

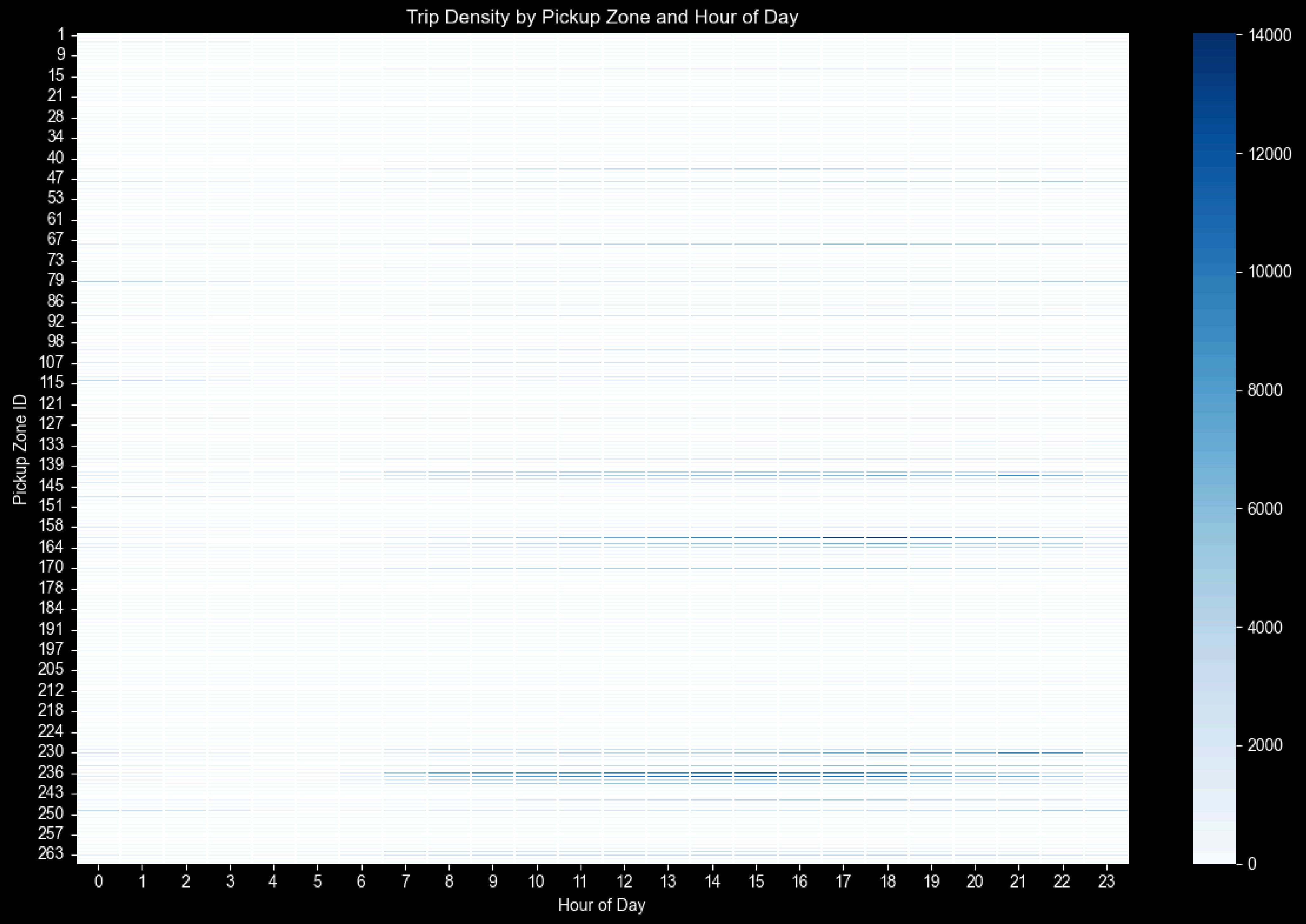
- Highest surcharges seen in Midtown–Midtown and airport trips.
- Some outer-borough and low-traffic trips also triggered surcharges.
- Indicates inconsistency across similar congestion contexts.

POLICY IMPLICATIONS

- Waive charges for low-congestion zone pairs (e.g., bottom-left of heatmap).
- Define policy based on congestion clusters, not just pickup points.
- Improve transparency by explaining where and why charges apply.

Surcharge concentration highest below 60th Street (core congestion zone)

HOW AI ILLUMINATES TAXI SYSTEM BEHAVIOR



Key Findings

- Peak pickups concentrated between 8–10 AM and 4–7 PM, matching NYC's commuting rush hours.
- Heavy pickup activity from zones near Midtown and Penn Station during peak times.
- Some outer-neighborhood residential zones showed off-peak spikes, hinting at flexible commuting patterns.

POLICY IMPLICATIONS

- Apply higher surcharges during commuting peaks to better manage congestion.
- Invest in targeted transit improvements (e.g., more subway service) in dense pickup areas.

Trip density peaks in Midtown during business hours

STRATEGIC RECOMMENDATIONS FOR NYC TLC

NYC Taxi & Limousine Commission

ADOPT DYNAMIC SURCHARGE PRICING

- Implement real-time surcharge adjustments based on trip pickup time, pickup zone congestion levels, and demand intensity.
- Example: Higher surcharges during peak commute hours (7–9 AM, 5–7 PM) and in hyper-congested areas like Midtown and Penn Station.

DEPLOY ZONE-CLUSTERED FARE STRUCTURES

Using machine learning clustering, customize surcharges by grouping zones into categories:

- Premium Zones: Consistent high fares → Fixed surcharge ($2.75).
- Residential Zones: Low income, local trips → Reduced or waived surcharges during off-peak hours.
- Commuter Zones: Time-variable surcharges to encourage off-peak travel.

EQUITY ENHANCEMENT MEASURES

- Incorporate income-adjusted exemptions or discounts for trips originating from historically disadvantaged boroughs (e.g., Bronx, Eastern Queens).
- Policy modeled after successful congestion pricing mitigations proposed in London and Stockholm.

Thank You

Google Collab Notebook