NATIONAL ECONOMICS UNIVERSITY

FACULTY OF ECONOMICAL MATHEMATICS

GROUP ASSIGNMENT – FINAL PROJECT

Data preparation, visualization, and storytelling with Uber TLC NYC 2019-2025 data

Group: 02

Members: Lê Ngọc Bích

Hàn Chí Kiên

Khổng Gia Minh

Lê Nguyễn Tuệ Minh

Nguyễn Tống Nguyên Vũ

Instructor: PhD. Nguyễn Tuấn Long

Class: DSEB 65A

Ha Noi, December 2025

Table of Contents

PART 1: DATA STORY

Operational data from 2023–2025 indicates that the Uber NYC market has reached a state of saturation. To break through current growth ceilings, this report shifts focus from scale expansion to depth optimization, proposing a strategy built on two core pillars:

1. **Financial Optimization (Pricing & Risk Strategy):** Transition from passive supply-demand pricing to active behavioral and risk-based pricing, aiming to maximize revenue yield per trip.

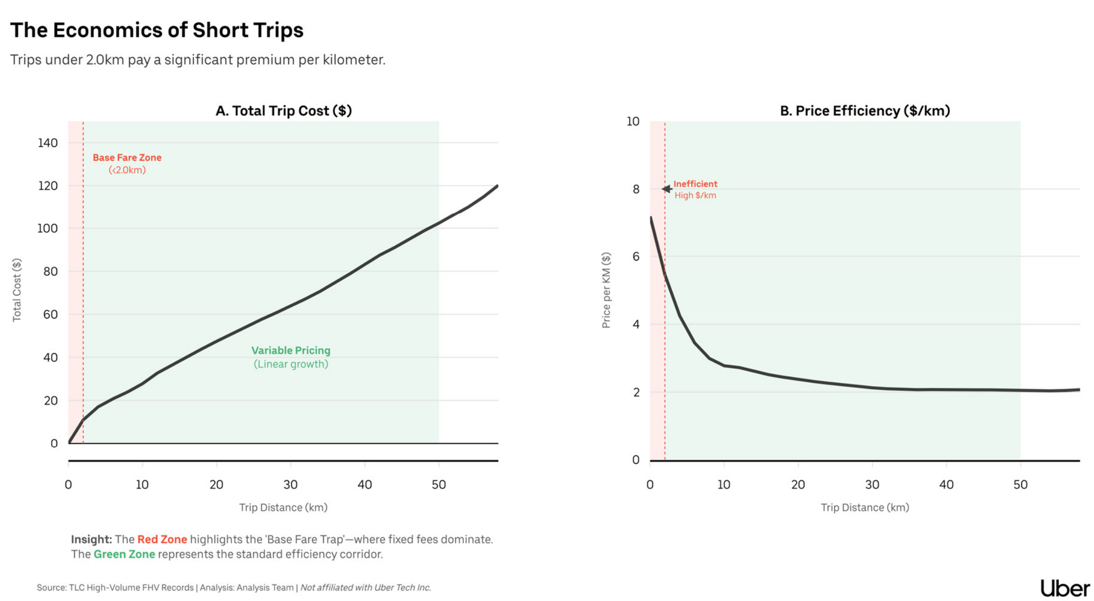
2. **Operational Optimization (Operations & Supply Strategy):** Analyze the impact of infrastructure (bridges, tunnels) and weather (rain, wind) to execute precise supply dispatching, thereby minimizing wait times and operational friction.

PROBLEM 1: PRICING OPTIMIZATION AND RISK MANAGEMENT STRATEGY

I. PRICING OPTIMIZATION STRATEGY

Uber's pricing strategy in NYC reflects a high degree of refinement, focusing on accurately reflecting the supply-demand tension across space, time, and trip characteristics to maximize revenue efficiency.

1. Non-Linear Pricing Structure and the Superior Value of Short Trips



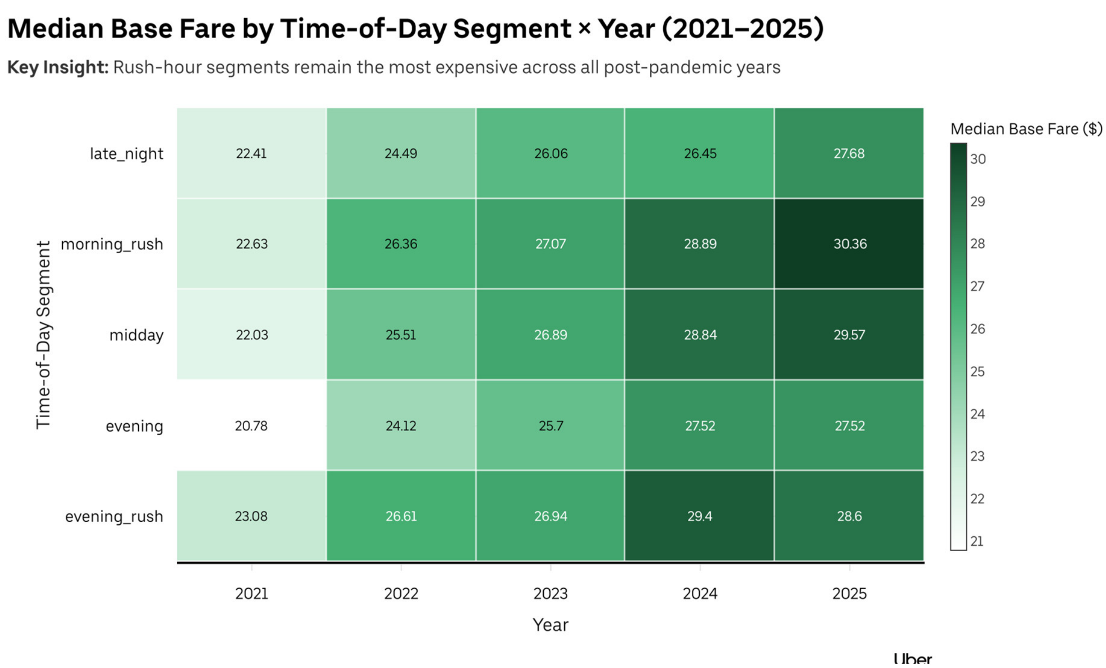
The pricing structure analysis identified a critical breakpoint in distance-based pricing at ~2 km through a segmented regression model. This segmentation shows a large disparity in revenue per kilometer:

- Median Price/km for short trips (0–2 km) is ~7.52 USD/km.
- Median Price/km for long trips (>2 km) is ~3.80 USD/km.

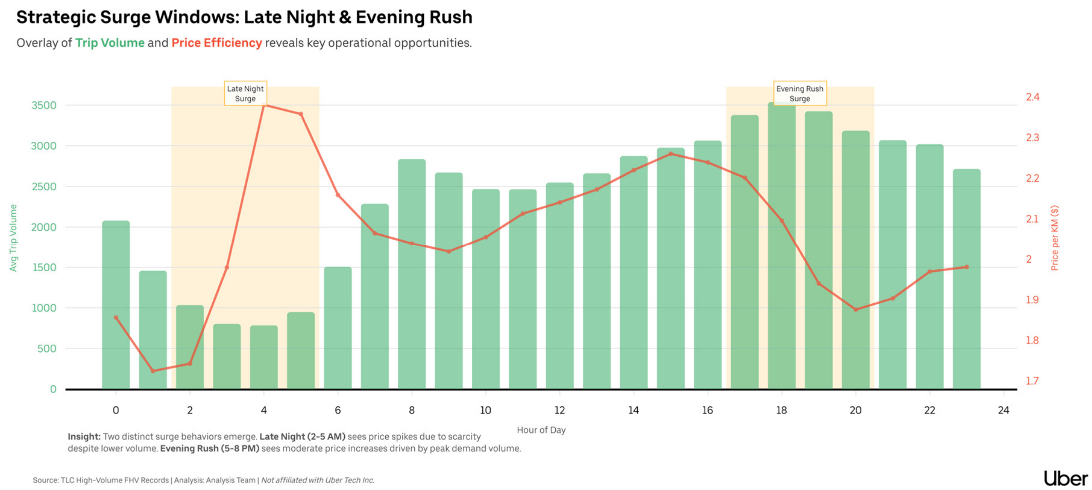
This creates a **Short Trip Premium** of up to +97.8% for trips under 2 km. The insight is that these short trips yield the highest revenue per kilometer in the system, nearly double that of long trips. The reason lies in the fare structure in NYC (Base Fare + Time-based Fee + Distance); in high-traffic and low-speed areas (e.g., Manhattan), the time-based fee increases in proportion, thereby increasing Cost/km. Users in NYC are willing to pay this high price for 1–2 km trips to avoid inconveniences like a crowded Subway or bad weather conditions.

Business Decision: Continue to optimally adjust the pricing structure for short trips (<2 km) by maintaining or slightly increasing the Base Fare/Minimum Fare, based on market evidence showing a real **willingness-to-pay** up to ~7.5 USD/km."

2. Pricing by Time and Demand (Time-of-Day & Surge)



Based on price data from 2021–2025, the structure of the Median Base Fare remains stable and consistent across years. Traditional peak periods such as the Morning Rush and Evening Rush consistently show the highest Median Base Fare (ranging from $26–$30 depending on the year), while Late Night is structurally the lowest-priced time segment. However, a deeper analysis of Surge Pricing reveals a significant difference in price sensitivity between these time segments.



Dynamic pricing analysis across time-of-day reveals two periods with completely different operational dynamics, each requiring its own pricing strategy: Evening Rush (4–8 PM) and Late Night (3–6 AM).

The Evening Rush (4–8 PM) is the time window with the highest trip volume, reaching +39% above the daily average. However, the price premium (surge) during this period remains very low—only +2.8% above the baseline. This imbalance is explained by rider behavior in NYC during late afternoon: although demand for going home is extremely high and traffic congestion is severe, riders still have many viable alternatives (such as the Subway and other public transit systems operating at full capacity). As a result, customers during Rush Hour exhibit high price sensitivity—they tolerate mild surge but reject aggressive surge levels. Data also shows that the Surge Score reaches its highest values between 3–7 PM (ranging from 1.22 to 1.35), indicating that this is the peak period for supply–demand tension.

In contrast, the Late Night window (3–6 AM) registers low volume (~35% of the daily average) but records the highest price premium of the entire day (+8.6%). The reason is that late at night, public transportation (Subway, buses) operates with reduced service or shuts down entirely, leaving riders with almost no substitutes. In this environment, mobility demand becomes urgent, resulting in extremely low price elasticity. Riders are willing to pay significantly higher prices to secure a trip, and increasing surge does not materially reduce volume.

Business Decisions:

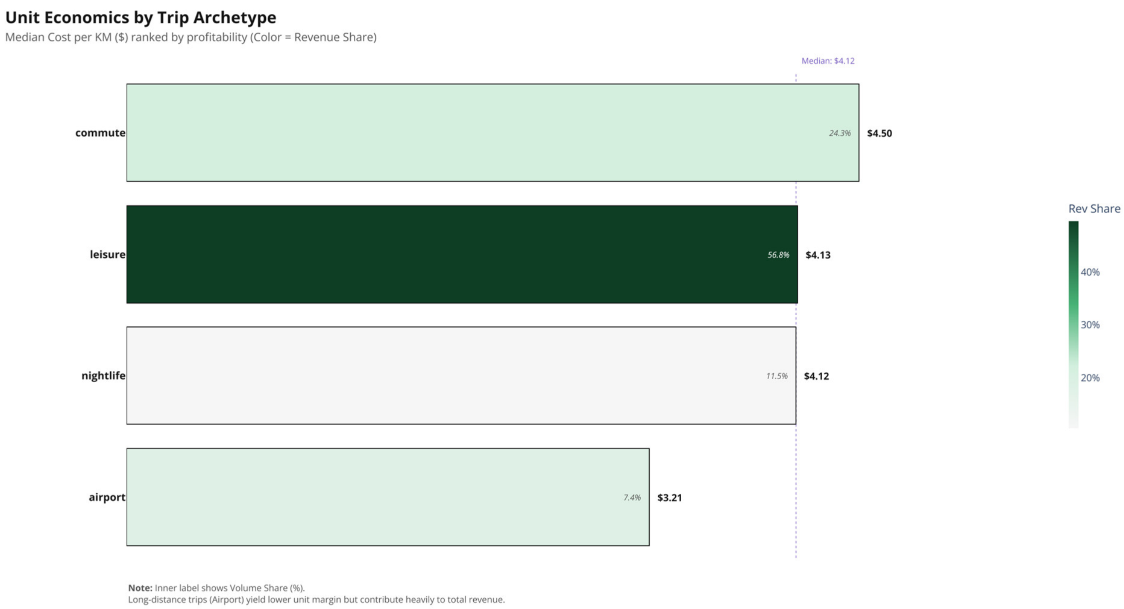
- Increase Surge during 4–8 PM:

Given the very high trip volume and stable demand, the platform should moderately increase surge during this period from the current +2.8% to around ~+5%. This adjustment is considered demand-safe due to the large market volume while also improving revenue and incentivizing more drivers to operate during peak hours.

- Apply Stronger Surge for 3–6 AM:

Based on insights about low price elasticity and high variance (uncertainty in supply/demand), the platform should adopt higher and more flexible surge levels for this time window to maximize revenue from urgent late-night trips.

3. Optimization by Trip Archetype



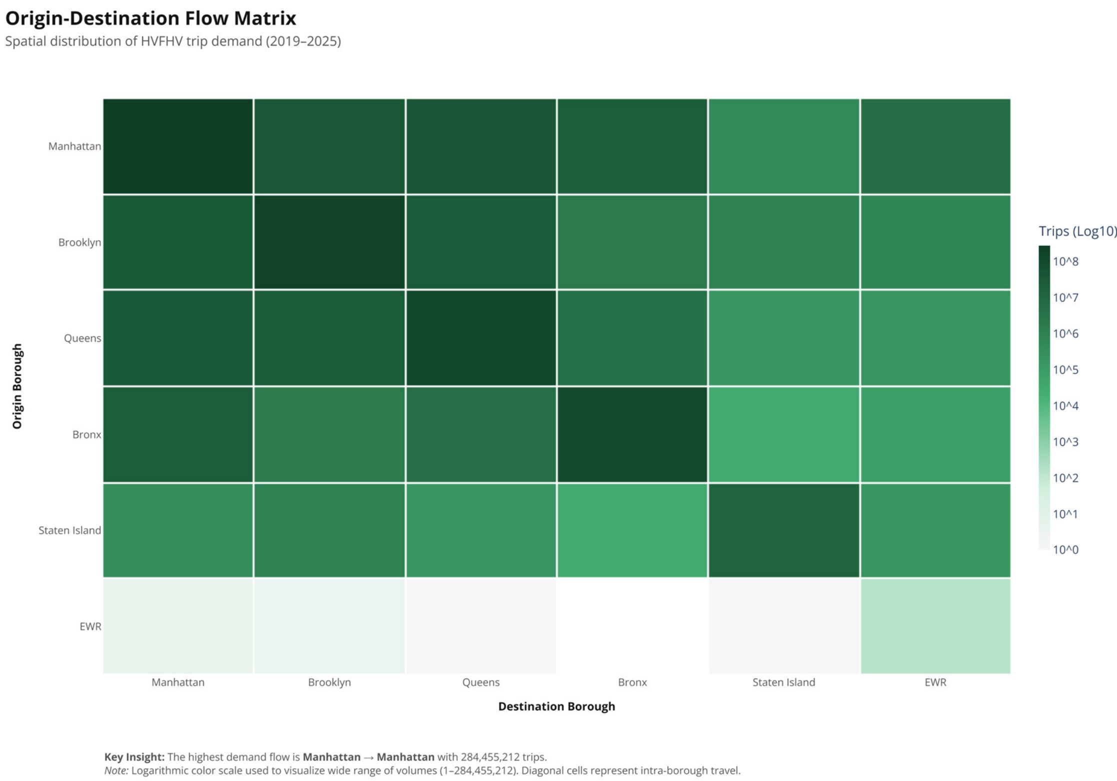
The Cost/km analysis by trip archetype shows a clear difference in revenue performance per unit of distance: Commute (~$4.50/km) and Leisure (~$4.13/km) are the two highest-value segments. Nightlife follows with ~$4.12/km, while Airport has the lowest Cost/km, at only ~$3.21/km.

The Leisure segment is the main operating engine of the system, accounting for 56.8% Volume and generating 49.6% Revenue. When combined with Commute, these two segments contribute a total of ~70% of revenue. Commute has a high Cost/km because these trips are primarily concentrated in Manhattan, where high traffic increases the time-based cost (Cost/min). Conversely, the Airport segment has a low value per km because it is dominated by fixed price regulations (TLC flat fees), which limit the ability to dynamically optimize pricing.

Business Decision: The platform should focus on optimizing Commute & Leisure through Marketing strategies and driver positioning based on time slices. For the Airport segment, the current pricing should be maintained to avoid losing market share to competitors (like yellow cabs) who already apply a flat fare.

4. Pricing by Geography and Weather:

4.1. Geographic Analysis and Intra-borough Trends



Geographically, although Manhattan dominates the number of trips, a deeper analysis shows that 74.8% of the total nearly 1 billion trips analyzed are intra-borough (within the borough). Among these, the three highest-traffic routes are all internal within their respective boroughs, led by the Manhattan → Manhattan route (~284 million trips), followed by Brooklyn → Brooklyn and Queens → Queens.

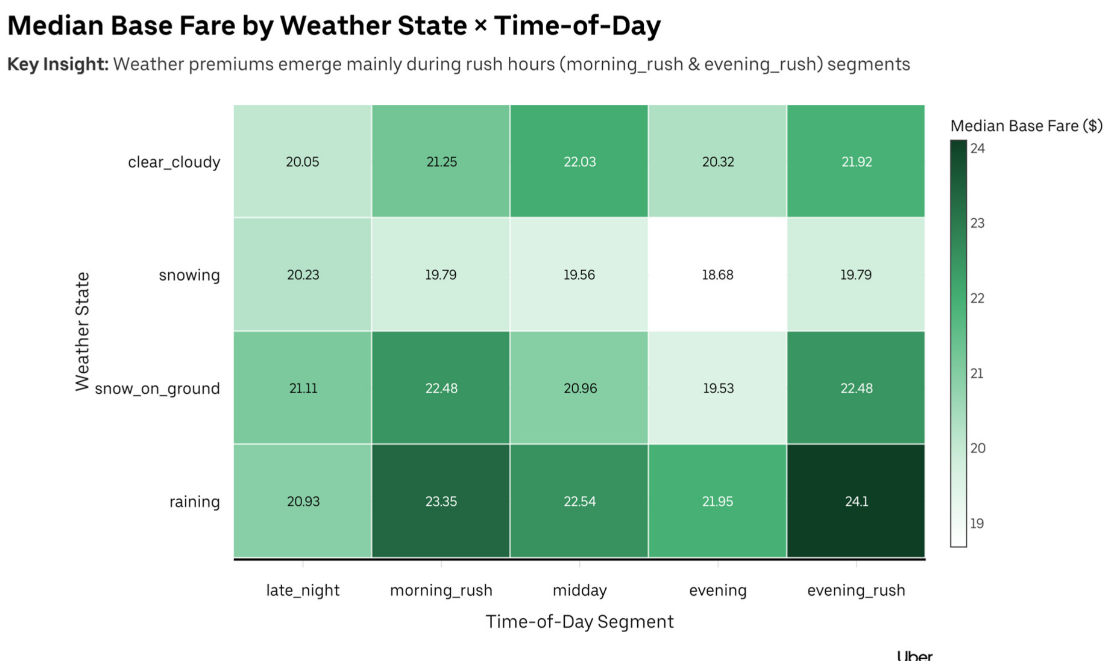
This finding confirms that Uber primarily operates as an intra-regional urban transport system in NYC. This reflects the extremely high demand in Manhattan due to population and commercial density, and Uber's growing role in serving areas in the periphery like Brooklyn and Queens that lack efficient public transit connections.

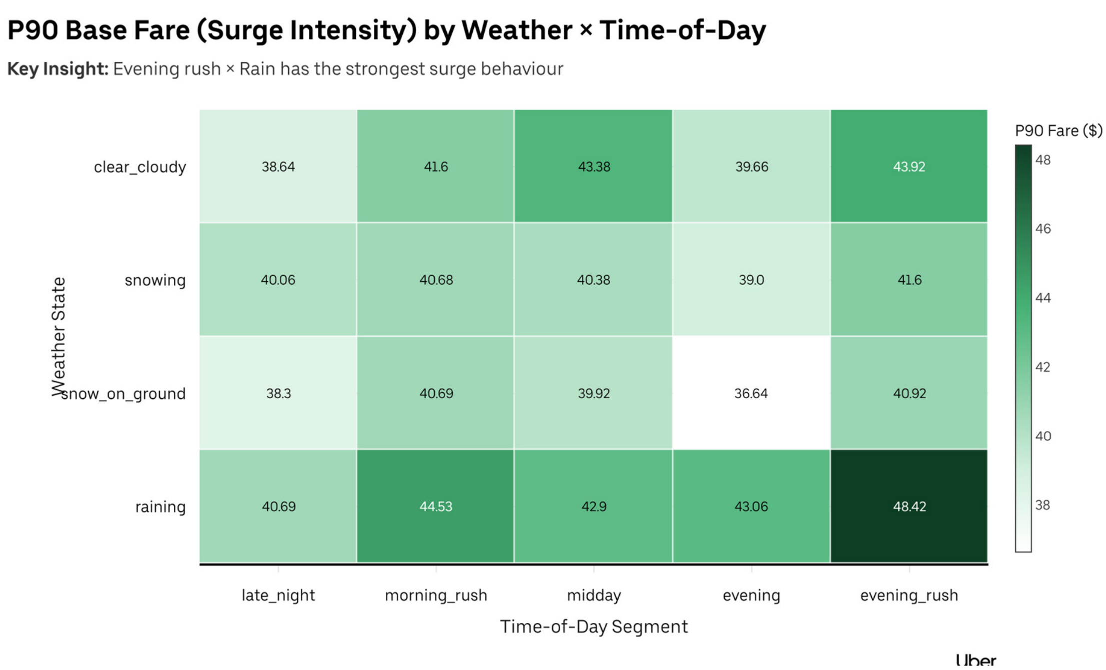
Business Decision: Continue to maintain strong resources in Manhattan (driver allocation, flexible pricing) while expanding services in peripheral areas by implementing local promotions or partnerships with transit-deprived residential areas.

4.2. The Non-linear Impact of Weather

An analysis of weather-state data shows that weather has a direct and consistent impact on trip prices, but Rain and Snow produce two completely different underlying mechanisms:

4.2.1. Rain — A Supply Shock



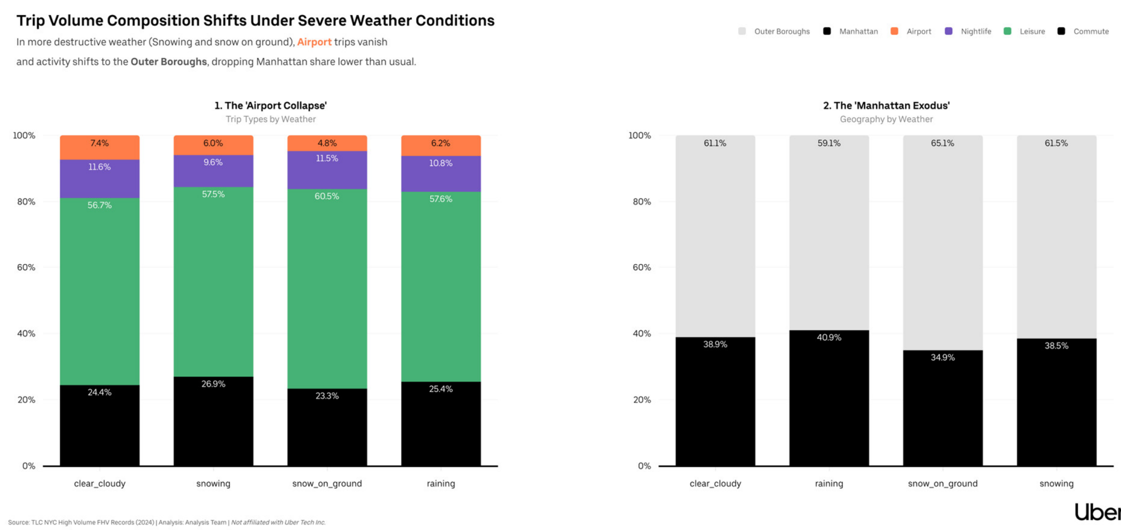


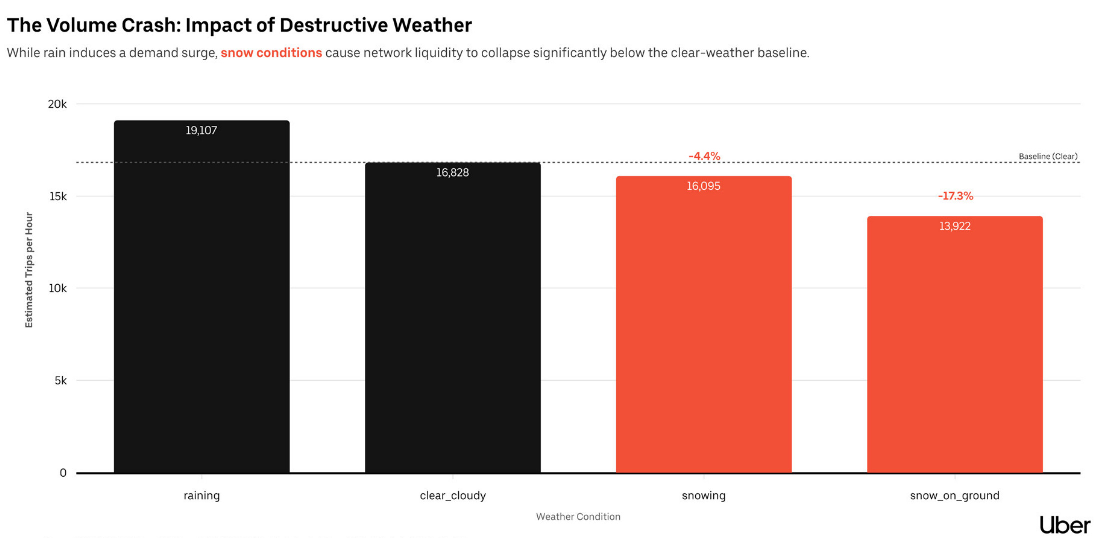
Rain creates a clear and consistent price-increasing effect (Median Fare rises across most time segments; for example, Evening Rush reaches $24.1). The P90 fare also increases noticeably.

Mechanism: When it rains, visibility decreases, roads become slippery, travel speed drops, and drivers' ability to accept additional trips becomes more limited. This makes supply less elastic, while demand tends to rise as riders want to avoid walking or being exposed to bad weather. This supply–demand shift pushes both the median fare and the P90 fare higher, especially during time segments that already have strong demand.

Business Decision: When Bad Weather overlaps with the Surge Window, a soft-cap surge strategy should be applied to prevent prices from rising too sharply and to maintain a reasonable rider experience.

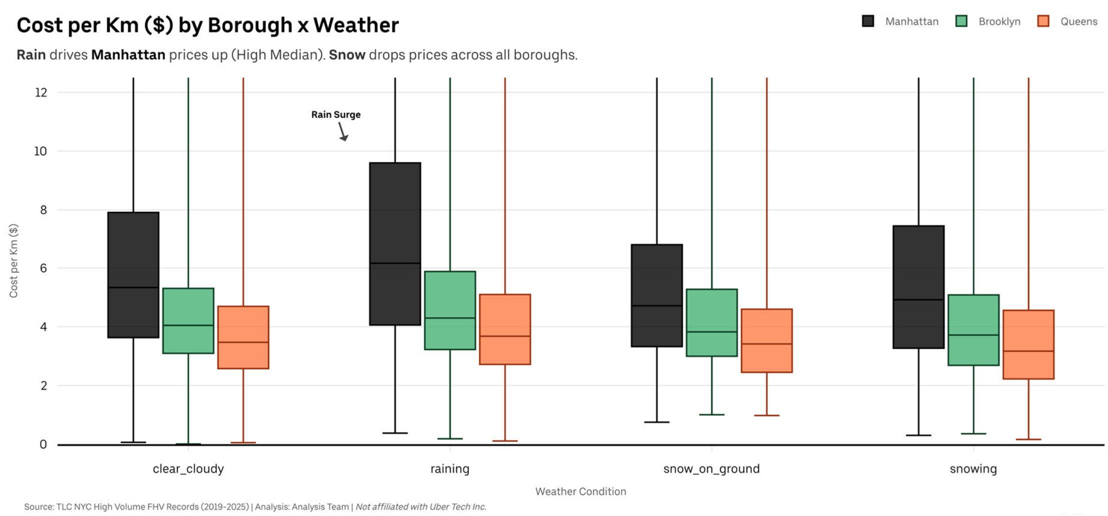
4.2.2. Snow — Demand Destruction

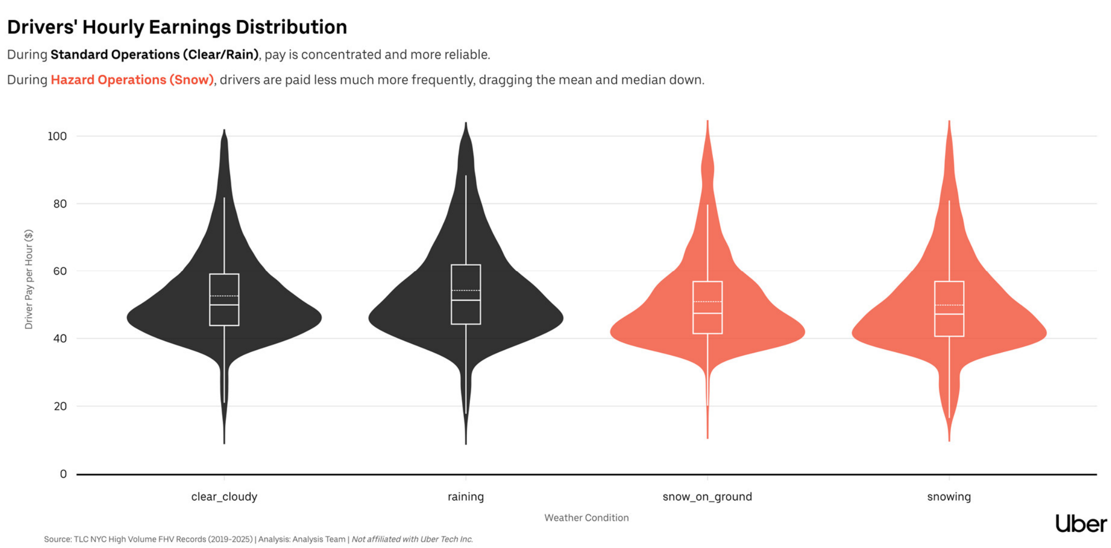




- **Disappearance of “Whales” (Composition Shift):** The decline in Median Fare is due to the loss of high-value trips: **Airport trips decrease by ~35%** (from 7.4% to 4.8% market share) and Volume Premium in Manhattan contracts. Activity shifts to lower-priced Outer Boroughs.

- **"Hunker Down" Effect (Volume Crash):** Snow triggers a "Cancellation" (stay-at-home) mentality instead of "Urgent Demand." Trip Volume **decreases by ~27%** compared to rainy days, causing the actual average speed to **increase** due to empty roads.





Failure of Unit Economics: The surge mechanism fails due to a lack of congestion or surplus demand. Cost-per-km in Manhattan decreases from ~$6.00 (Rain) → $5.00 (Snow). Drivers suffer a **Net Wage Loss**, earning ~$4/hr less than on rainy days due to the absence of high-value trips.

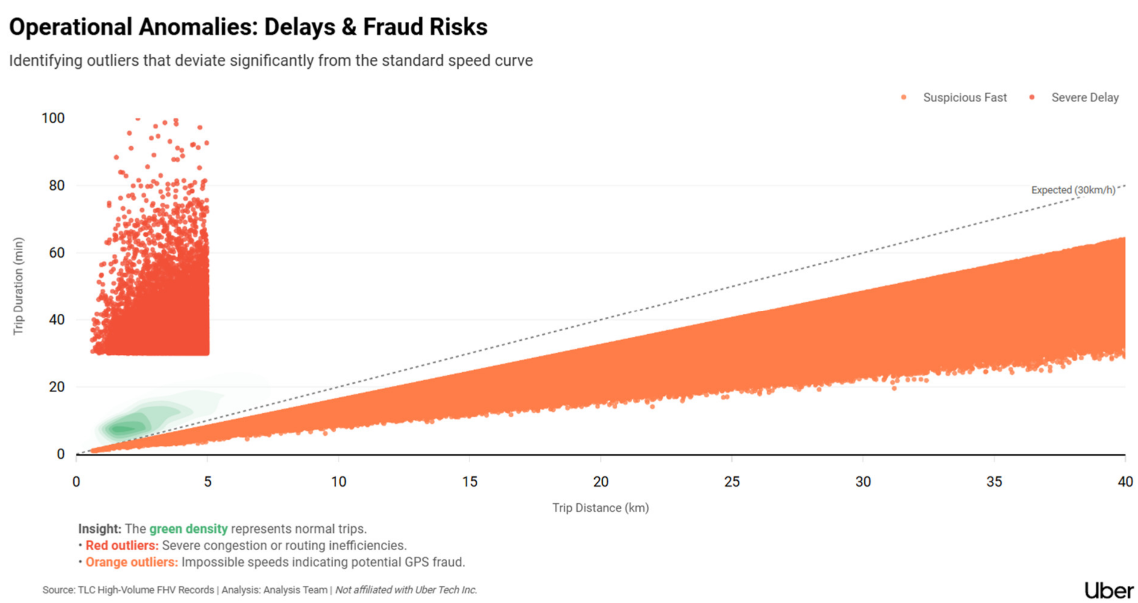
II. ANOMALY DETECTION & RISK MANAGEMENT

This section focuses on identifying operational risks, cost anomalies, and understanding user and driver behavior to maintain service quality and optimize the system.

1. Severe congestion and Operational risk

Analysis of anomalies related to travel speed and cost has highlighted severe breakdowns in service quality within the system.

1.1. Severe congestion ("Slower than walking")



Deep exploratory analysis of platform operational data has isolated a critical anomaly within the network: a cluster of over 31,000 trips - representing approximately 0.32% of total volume - operating at a mean velocity below 7 km/h. In the context of urban physics, this velocity is marginally superior to pedestrian speeds (5-6 km/h) yet significantly inferior to micromobility alternatives. These trips display a severe distortion in unit economics: despite traversing distances of under 5 kilometers, vehicle occupancy durations exceed 30 minutes due to acute infrastructure saturation. Consequently, the dynamic pricing algorithm, which accounts for time-based driver compensation, elevates the average fare to approximately $36.32. This results in a unit cost of roughly $9.83 per kilometer for the consumer - a price point that is economically unjustifiable for a service level that is functionally static. This represents not merely a localized inefficiency, but a fundamental failure in market resource allocation.

The persistence of this cluster stems from a negative interaction between physical infrastructure constraints and algorithmic pricing logic. Exogenously, the saturation of the Central Business District (CBD) in Manhattan has eroded the ride-hailing value proposition of "speed" and "convenience", reducing traffic flow to near-zero velocities during peak intervals. Endogenously, the current fare structure operates on a linear time-dependency model designed to protect driver earnings; however, in gridlock scenarios, this creates a negative feedback loop where service degradation (slower speeds) paradoxically results in higher costs for the consumer. From a consumer behavior perspective, this destroys consumer surplus and violates the implicit contract of the on-demand economy.

These 31,000 trips should not be categorized as revenue capture, but rather as "technical debt" in Customer Experience (CX). In the specific context of New York City, where the substitute goods - specifically the MTA Subway with a fixed marginal cost of $2.90 and a high-density Citi Bike network - are readily available, the elasticity of demand regarding service quality is profound. When consumers are forced to pay a premium approximately 12 times higher than mass transit for a velocity slower than walking, the probability of permanent churn is maximized. Furthermore, under the regulatory framework of the NYC Taxi & Limousine Commission (TLC) regarding minimum driver pay and utilization rates, these low-velocity trips degrade the aggregate network efficiency, indirectly inflating the operational costs the platform must absorb to maintain driver liquidity.

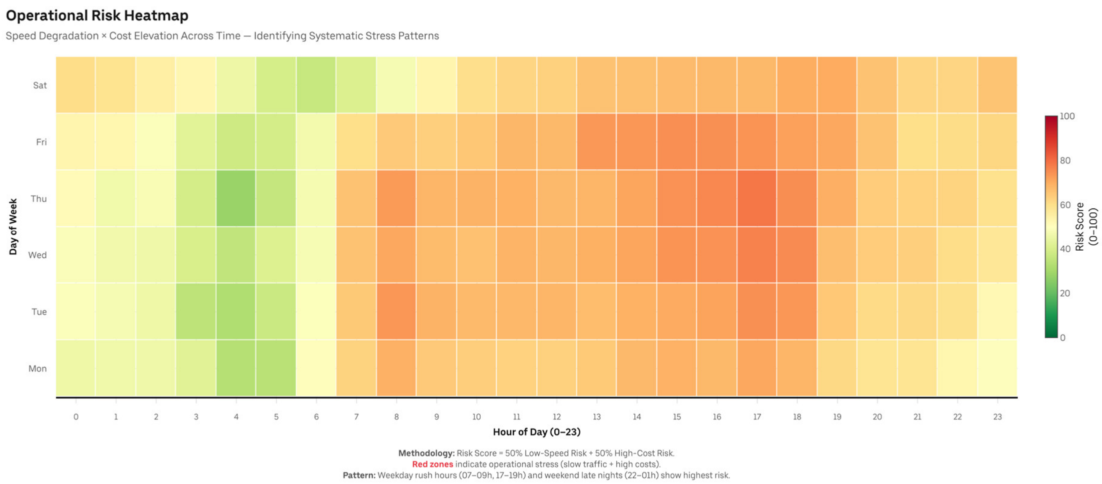
Business Decisions:

To rectify this structural inefficiency, the platform must transition from maximizing revenue per trip to optimizing for aggregate network velocity. This requires an algorithmic redirection of demand. When predictive models indicate an estimated velocity below 8 km/h, the user interface should actively deprioritize high-friction car travel in favor of micromobility integration. By guiding demand toward Lime or Citi Bike alternatives for short-distance vectors in the core zone, the platform effectively "cannibalizes" inefficient short-term revenue to preserve long-term customer retention and brand equity.

Operational efficiency can be reclaimed through the implementation of "Congestion-Smart Pickups." By leveraging traffic flow theory, the dispatch algorithm should enforce virtual pickup points (Virtual Stops) displaced by 1–2 city blocks from high-friction zones. Requiring the passenger to perform a short pedestrian ingress to a more fluid arterial road reduces deadhead time and ingress/egress friction. This optimization directly improves the Utilization Rate—a key compliance metric for TLC regulations—while simultaneously reducing the total trip duration and effective cost for the rider.

Mitigating the reputational risk associated with the "Efficiency Paradox" requires a shift toward radical transparency. Implementing pre-booking warnings that contextualize the estimated travel time against alternative modes (e.g., "Car: 35 mins vs. Walking: 20 mins") establishes the platform as a "Trusted Mobility Advisor" rather than a beneficiary of congestion. This expectation management strategy neutralizes the negative sentiment associated with inevitable urban gridlock, thereby protecting the platform's Net Promoter Score (NPS) in a high-stakes competitive environment.

1.2. Highest operational Risk time window



The identification of the 17:00–19:00 weekday window as the apex of operational risk highlights a fundamental structural flaw in the current ride-hailing model: the inverse value proposition. During these peak intervals, the platform exhibits a negative correlation between price and utility. The data indicates a composite risk score exceeding 73–76, characterized by a simultaneous spike in unit cost (~$6.16/km) and a collapse in service velocity (~16 km/h). This phenomenon is not merely a supply-demand imbalance but represents a failure of the algorithmic pricing mechanism to account for service utility. When surge pricing activates solely based on demand volume during gridlock, it creates a feedback loop of dissatisfaction - the user pays a premium for a service that is functionally inferior to alternative modes of transport, such as the subway or walking. This creates a critical vulnerability where high-value users are most likely to permanently churn due to buyer's remorse.

Business Decisions:

Recalibrating the surge pricing model is essential to ensure fairness during peak congestion. Recalibrate surge pricing to a utility-weighted model: Cap the distance multiplier but keep the time variable for driver pay in high-congestion zones (risk score > 75). This sustains supply without overcharging riders for slow service.

Implementing predictive congestion alerts is key to managing user expectations and promoting transparency. Implement predictive congestion alerts: Use radical transparency by warning users about high travel-time variance (e.g., "30-50% longer") before booking in high-risk zones. This manages expectations and reduces customer regret.

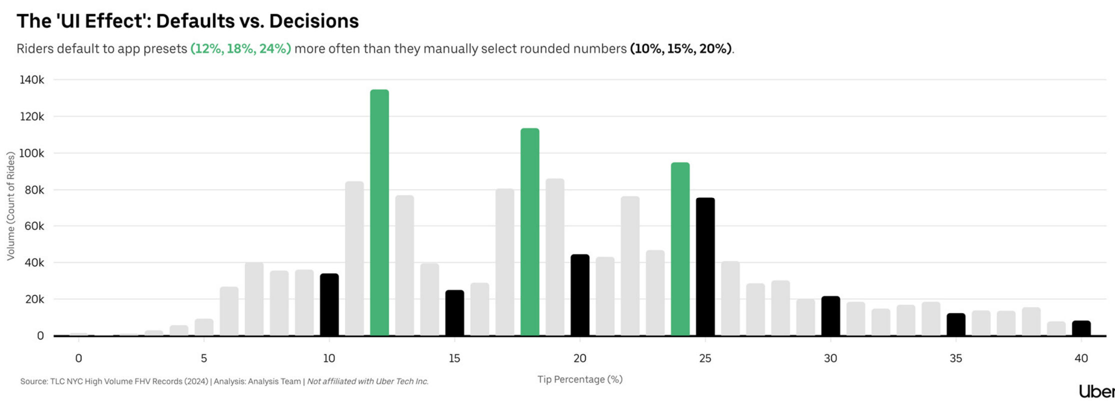
An active modal diversion strategy must be used as a critical retention mechanism in high-friction zones. Use active modal diversion for retention: In high-risk areas, prioritize alternative transport like micromobility or public transit for short trips (< 3km). This improves network efficiency and prevents user frustration.

Introducing a slow-speed credit mechanism serves as a powerful loyalty instrument for high-tier users. Introduce a slow-speed credit mechanism: Offer proactive post-trip credits (e.g., $5) to high-tier users if the average trip speed falls below a critical threshold (e.g., 8 km/h). This leverages the service recovery paradox to build loyalty.

2. Management of Passenger and Driver Behavior

2.1. Tipping behavior

a) The "UI" Effect

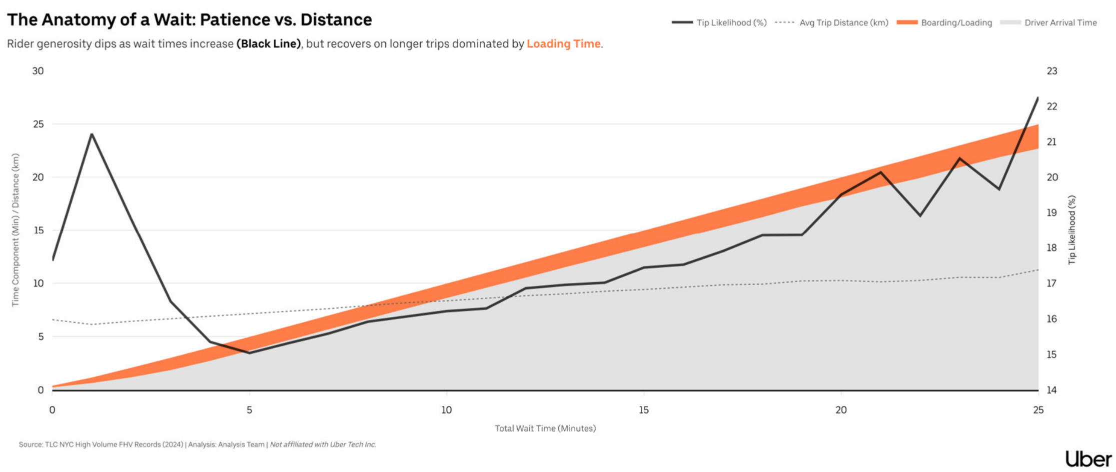


The observed concentration of tipping data around pre-set options illustrates the profound impact of choice architecture on driver compensation within the platform. This behavioral pattern confirms that for the vast majority of riders, the act of tipping is driven less by a calculated valuation of service quality and more by the cognitive convenience of accepting the default suggestion presented by the interface. In the context of the New York City market, where maintaining high driver retention is critical for network reliability, the reliance on a static set of tip options represents a significant inefficiency that artificially suppresses potential driver earnings. By treating all trips equally in the payment interface, the current system fails to capture the full consumer surplus that passengers might be willing to offer during premium or long-distance experiences.

Business Decisions:

To optimize this dynamic, the platform should transition from static defaults to an algorithmic choice architecture that presents context-aware gratuity options. Rather than displaying the same fixed dollar amounts for every journey, the interface should utilize predictive modeling to adjust the suggested values based on trip parameters such as duration, distance, and total fare. For instance, anchoring the default options to percentages for high-value airport trips while utilizing flat dollar amounts for short crosstown rides can subtly shift the distribution curve upward. This strategy effectively leverages the existing user bias toward defaults to maximize driver revenue without altering base fares, thereby strengthening the supply side of the marketplace through improved total earnings potential.

b) The Anatomy of a Wait

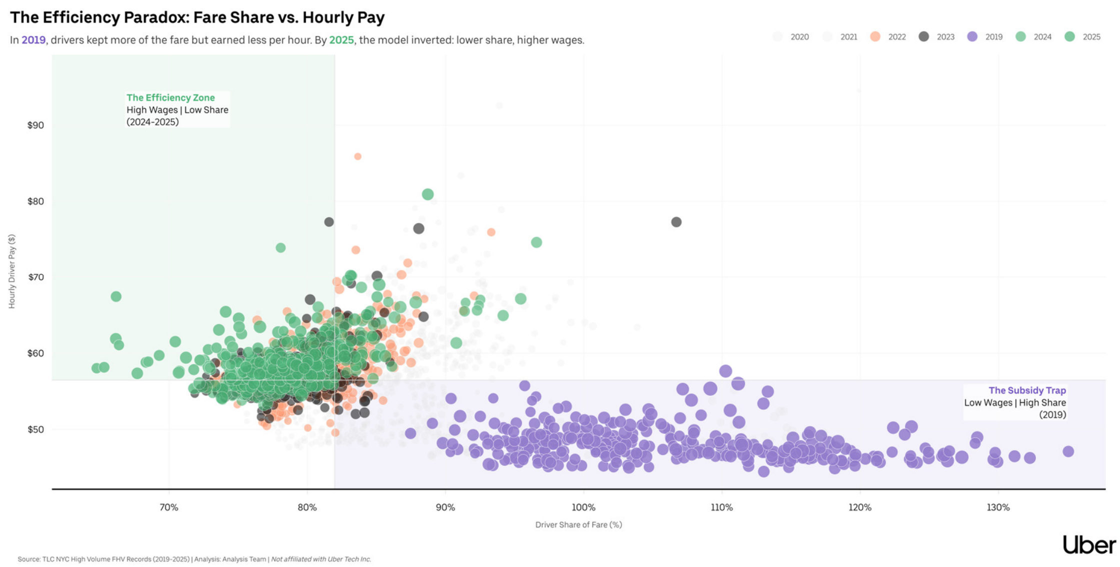


The empirical evidence regarding gratuity distribution reveals a distinct non-linear correlation between wait times and rider generosity, effectively described as a "Desperation U-Curve." The data highlights two statistically significant clusters of elevated tipping behavior: the immediate fulfillment window of zero to two minutes and the extreme latency tail exceeding twenty minutes. While the financial incentivization within the zero-to-two-minute interval aligns with rational market expectations for hyper-efficiency, the resurgence of high-value tips after delays surpassing twenty minutes indicates a fundamental psychological shift. In these high-latency scenarios, the rider's internal valuation mechanism transitions from a transactional assessment of speed to an emotional response of relief, viewing the eventual arrival of the driver as a rescue from market failure rather than a standard service provision.

Business Decisions:

To operationalize this insight, the platform must evolve its interface to mirror this emotional journey. Rather than issuing generic apologies for delays exceeding the twenty-minute threshold, the user experience should be re-engineered to explicitly visualize the driver's extraordinary logistical effort to bridge the supply gap. By reframing the narrative of the delay to highlight the specific exertion of the worker - such as the distance traveled to the pickup or the navigation of adverse conditions - the platform can legitimize the rider's inclination toward "rescue tipping." This strategy effectively converts operational inefficiencies into opportunities for maximizing driver revenue by aligning the payment architecture with the rider’s shift from frustration to gratitude.

2.2. Wage dynamics across subsidy eras



The empirical divergence between declining revenue-share percentages and rising driver hourly income during the 2023-2025 period validates the hypothesis that network utilization, rather than the static division of fare revenue, acts as the primary determinant of supplier welfare in the mature gig economy. This decoupling indicates that the platform has successfully transitioned from a high-subsidy growth model to a high-efficiency mature model where the algorithmic reduction of non-revenue cruising time generates significantly more value for the driver than a generous split of a sporadic trip. Within the specific regulatory framework of New York City, where the Taxi and Limousine Commission mandates minimum pay floors inextricably linked to utilization rates, this data demonstrates that the minimization of idle time has effectively shielded driver earnings from the compression of take rates, allowing drivers to capture a smaller slice of a significantly larger and more frequent economic pie.

Business Decisions:

To operationalize this insight, the platform must fundamentally restructure its driver communications and interface design to align with this utilization-first reality. As the current driver dashboard often emphasizes per-trip details which invites counterproductive scrutiny of the platform take rate, a strategic pivot would involve redesigning the interface to prominently visualize earnings per online hour as the north star metric, effectively reframing the value proposition from a transactional partnership to a shift-based employment proxy. Furthermore, the dispatch algorithm should be recalibrated to prioritize trip chaining—securing the subsequent pickup before the current drop-off is complete—above all other variables for drivers entering the core Manhattan zone, ensuring that the high-efficiency loop sustaining these elevated wages is protected against the friction of congestion.

PROBLEM 2: UBER NYC OPERATIONAL OPTIMIZATION AND SUPPLY STRATEGY

Timeline: 2019 – 2025 | **Data Source:** NYC TLC High Volume FHV Records

I. INTRODUCTION

Report Purpose: Providing a comprehensive strategic overview of Uber's operational ecosystem in New York City. It connects the evolution of core system mechanics (2019–2025) with a granular diagnosis of current bottlenecks (2023–2025) to propose actionable strategies for supply optimization and pricing efficiency.

Data Scope:

Macro Analysis (2019–2025): Assessment of market stability and Demand Recovery Trends (DRT).

Micro Diagnosis (2023–2025): Detailed investigation into current operational friction, specifically traffic congestion, geographic barriers, and weather sensitivity.

Technical Datasets:

- tlc_sample_processed (2019-2025)
- agg_network_monthly.parquet
- agg_timeline_hourly.parquet.

II. OPERATIONAL MECHANISMS AND MACROECONOMIC TRENDS

1. Dispatch Algorithm

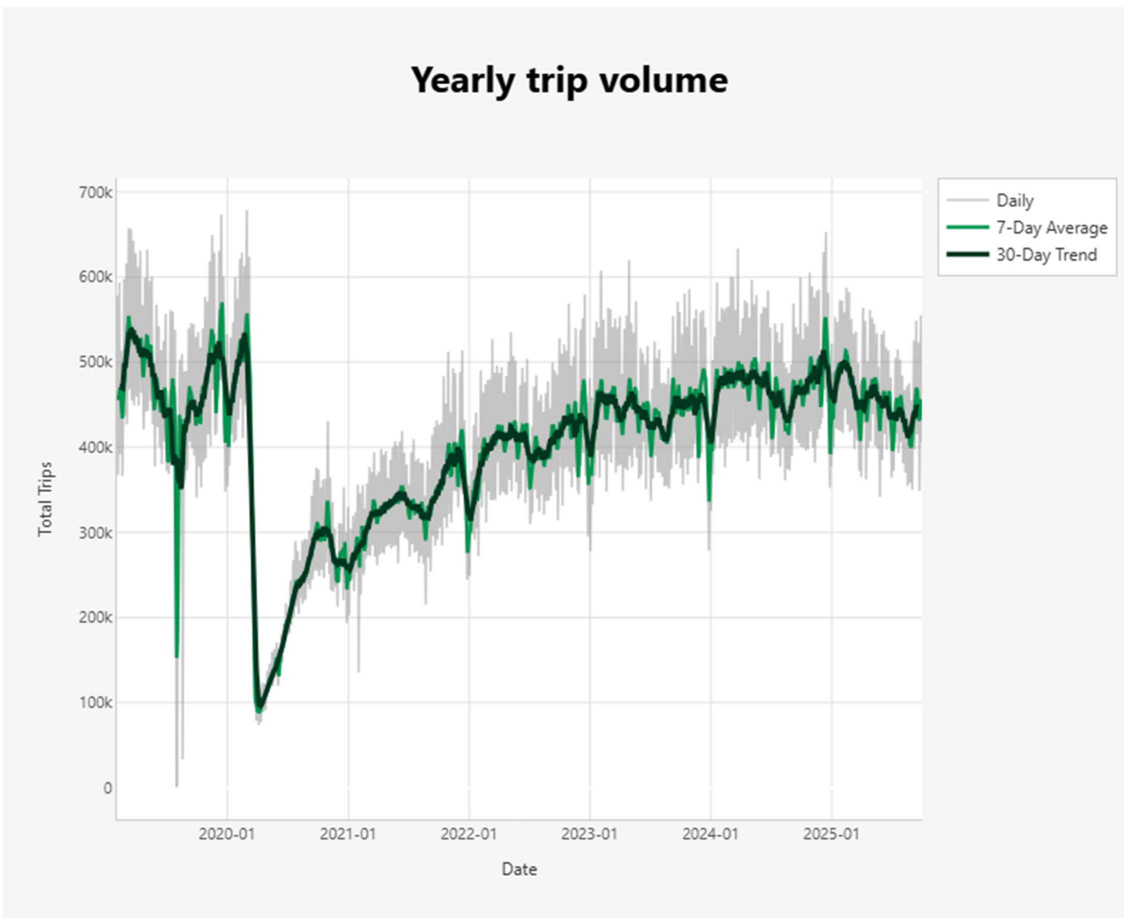
Uber's current Marketplace Engine represents a fundamental shift away from the traditional logic of "*the nearest driver picks up the nearest rider*" (Greedy Matching) toward a globally optimized dispatching system. This transformation is built on three core pillars:

First, Uber now uses Batched Matching. Instead of assigning a car the instant a request appears, the system groups multiple incoming ride requests within a short time window (typically 2–5 seconds) and constructs a matching graph. The algorithm then solves a network-wide optimization problem (such as a *Min-Cost Flow* formulation) to minimize total system cost. This approach reduces the average wait time for riders and allocates drivers more efficiently during peak demand periods.

Second, Uber has removed physical distance (kilometers) as a decision metric and replaced it entirely with ETA (Estimated Time of Arrival). ETA is a more accurate measure of operational cost because it incorporates real-world conditions: traffic density, one-way street constraints, turning restrictions, direction of movement, and the feasibility of U-turns. This shift enables the platform to optimize *actual travel effort*, rather than relying on straight-line distance.

Finally, Uber has moved from a reactive stance to proactive Supply Shaping. Instead of letting drivers roam freely and randomly, the system uses economic instruments such as Surge Pricing and Demand Forecasting to create financial incentives that nudge drivers toward areas where demand is expected to spike. This strategy reduces idle time and ensures that supply is pre-positioned and ready before demand surges occur.

2. Overall Analysis



Causes of the Abnormal Early-Year Decline and the Impact of 2019 Regulatory Changes

Historical data shows that the sharp drop in trip volume and revenue for Uber/Taxi services in NYC during Q1 is not driven solely by seasonality. The downturn reflects a combined effect: natural seasonal lows and a series of significant regulatory shocks introduced in 2019.

A. Seasonal and Cyclical Factors

These are the natural forces that make January–February the annual "low season" for the transportation industry:

Post-Holiday Slump:

- Tourism: According to NYC & Company, tourist arrivals fall sharply once the November–December holiday peak ends.

- Events: This period lacks major concerts, sports events, and nightlife activities, reducing non-essential travel demand.

Resident Behavior & Weather:

- Domestic mobility: Data from TLC and MTA confirms that commuting and discretionary travel (shopping, dining, celebrations) remain subdued in the first weeks of the new year.

- Weather barriers: Low temperatures and winter storms discourage people from going out, pulling down total travel activity across the city (per NYC DOT patterns).

B. Policy and Operational Shocks in 2019

2019 introduced several regulatory measures that directly impacted trip numbers and operational performance:

1. Congestion Surcharge – Effective February 2019

- Rule: An extra $2.75 per Uber/Lyft trip entering Manhattan below 96th Street.

- Impact: A sudden increase in fares suppressed demand, especially short inner-Manhattan trips. The decline is clearly visible in Q1–Q2 2019.

2. Driver Minimum Wage – Effective January 2019

- Rule: TLC mandated a minimum earnings floor of roughly $17.22/hour after expenses for app-based drivers.

- Impact: This increased cost pressure on Uber, pushing the company to optimize driver efficiency, reduce idle time, and adjust fare structures.

3. Supply Restrictions (Vehicle Cap & Cruising Cap)

- Rule: NYC extended the freeze on new FHV licenses and introduced caps on the share of time vehicles can cruise without passengers in Manhattan.

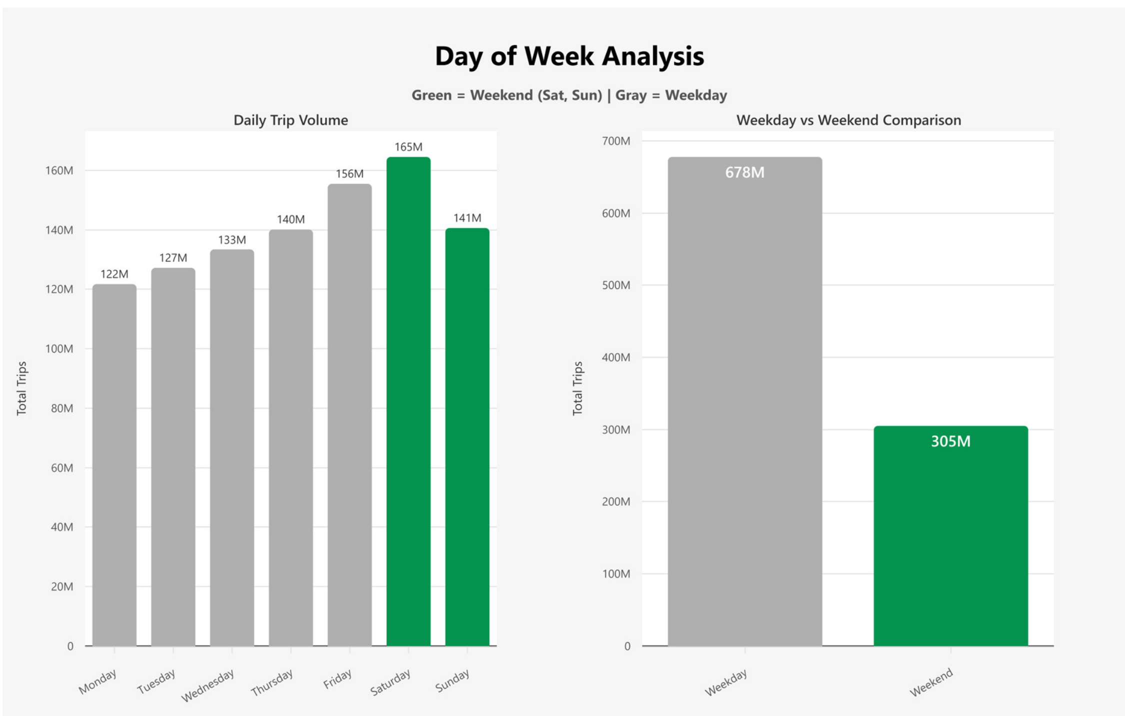
- Impact: Uber could no longer grow its fleet aggressively as in 2017–2018, forcing a shift away from rapid expansion toward controlled, sustainable growth.

4. Operational and Reputational Risks

- Driver strikes: Worldwide driver walkouts in March and May 2019 protested compensation and working conditions.

- Legal and regulatory setbacks: The high-profile loss of Uber's operating license in London (Nov 2019) created negative market sentiment and affected global brand perception.

3. Demand analysis



The chart illustrates total Uber trip volume across each day of the week over a five-year period. Weekdays are shown in gray, while Saturday and Sunday are highlighted in green to emphasize the segment targeted for deeper analysis.

Daily Trip Volume Patterns

Across five years of data, weekday trip volumes remain stable and follow a predictable commuter rhythm:

- Monday → Friday: 122M → 156M, rising gradually through the workweek. This reflects consistent demand driven by work, education, and routine travel patterns.

Weekend behavior diverges sharply from weekday structure:

- Saturday: 165M (highest of all days)

- Sunday: 141M
  The Saturday peak shows that weekend mobility is driven by nightlife, tourism, social activities, and discretionary travel—an entirely different demand engine. Sunday softens as the city transitions back toward weekday routines.

Weekday vs. Weekend Demand Split

Aggregating trips into two groups makes the contrast unmistakable:

- Total Weekday Trips: 678M

- Total Weekend Trips: 305M

Although weekdays dominate in total volume due to more calendar days, the per-day pattern tells the true story:

- Saturday alone exceeds every weekday, including Friday, revealing a structurally stronger weekend demand profile.

Interpretation for Uber NYC

The weekend segment emerges as a strategic opportunity, not just an extension of weekday behavior. Weekend mobility reflects:

- entertainment and nightlife flows,

- tourism surges,

- irregular long-distance movements,

- group trips and event-driven spikes.

These patterns are more volatile and less predictable than weekday commuting, making them highly relevant for decisions around:

- supply-demand balancing,

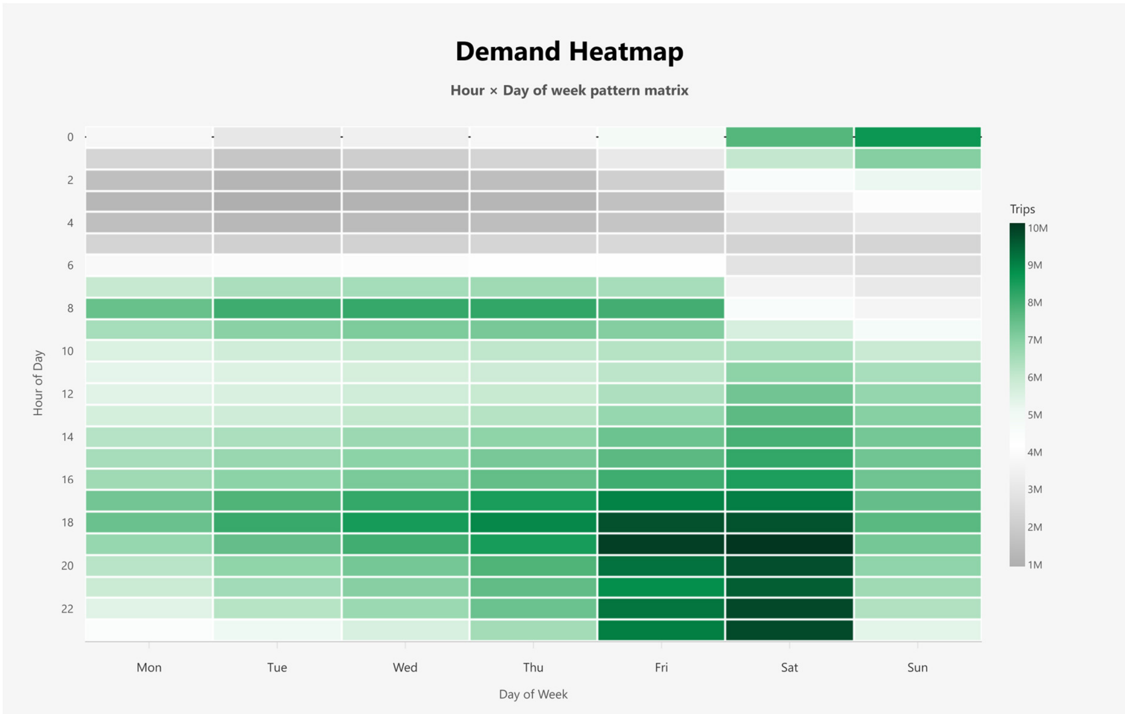
- pricing and surge strategy,

- driver deployment and incentives,

- revenue optimization.

By highlighting Saturday and Sunday in green, the visualization signals that weekend demand is the zone where Uber can unlock the most meaningful insights for operational and strategic planning.

4. Demand Heatmap Analysis



Strategic Interpretation

The heatmap reveals the hour-by-hour demand structure across the entire week, showing not only which days are strong but exactly when the most profitable windows occur.

- The central takeaway is unmistakable: Friday night and Saturday night form the single most intense demand zone of the entire week.

These periods mark New York's shift from commuter-driven mobility to nightlife-driven mobility, characterized by:

- restaurant traffic,
- bar and club movement,social events,
- tourism-heavy evening flows,
- late-night long-distance trips (high fare elasticity).

For Uber, these two blocks concentrate the operational conditions that matter most:

- supply shortages,
- surge multipliers,
- high-value trip patterns.

The heatmap therefore supports the strategic claim that weekend evening operations require targeted driver incentives, dynamic pricing design, and precise supply planning.

Key Insight From the Visual Pattern

Several dominant visual patterns structure the heatmap:

A. Moderate but steady weekday morning demand (7–9 AM)

- This aligns with standard commuter behavior: predictable, stable, but not the highest-revenue segment.

B. Afternoon–evening demand builds across all days

- Activity rises into dinner hours and post-work social time, forming a natural upward slope in demand.

C. The true peak is isolated to two windows:

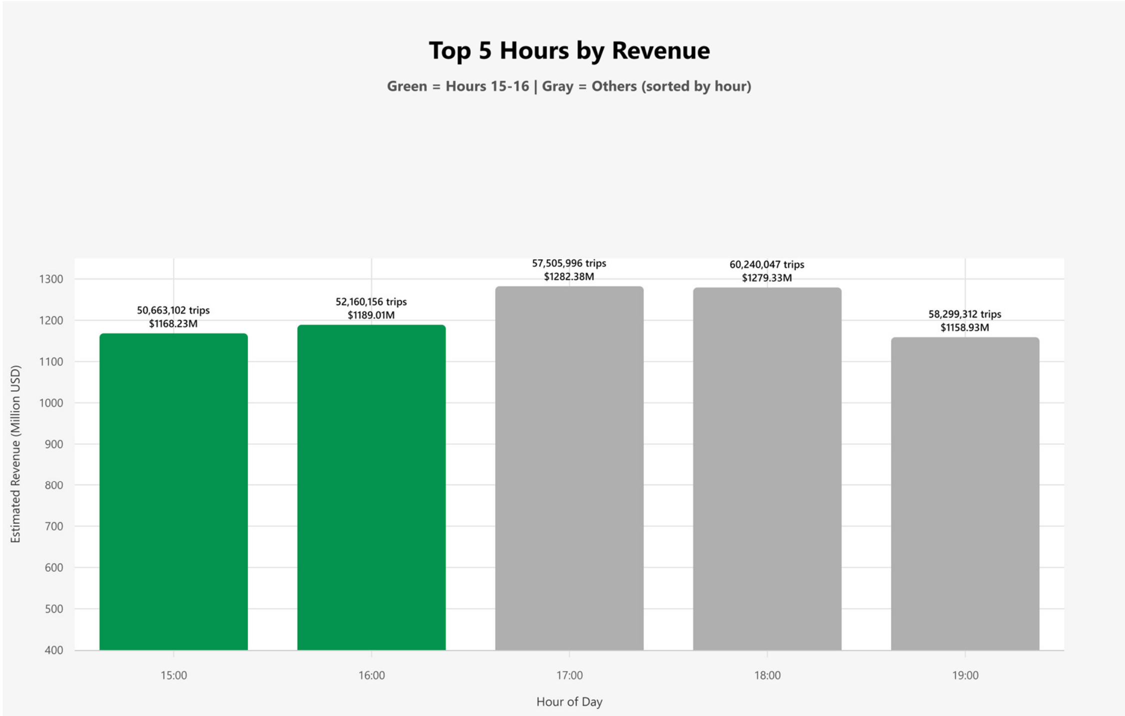
- Friday 18:00–23:00
- Saturday 18:00–24:00

These are the darkest cells on the heatmap and represent:

- the highest hourly trip totals in the entire 5-year dataset,
- the most intense mobility concentration,
- the strongest and most sustained surge-pricing opportunities.

In essence, Friday and Saturday nights are Uber's primary competitive battleground for maximizing earnings, and the heatmap visually makes this reality impossible to miss.

5. Revenue Analysis



Analysis of the Two Charts

The first chart exposes a clear analytical contradiction in Uber's hourly revenue structure: 15:00 and 16:00 generate nearly the same total revenue as the traditional commuter peaks at 17:00–18:00, despite handling substantially fewer trips.

- At 15:00, Uber completes 50.66M trips but earns $1,168.23M in revenue.
- At 16:00, Uber handled 52.16M trips for $1,189.01M.
- Meanwhile, at 17:00, trips rise to 57.51M but revenue increases only modestly to $1,282.38M.
- At 18:00, volume hits 60.24M—the highest among the peak hours—yet revenue remains close, at $1,279.33M, barely above the 15–16h block.

This comparison highlights a structural anomaly:

- Revenue is not proportional to the number of trips.
- Trip volume alone cannot explain why mid-afternoon hours perform like peak commuter hours.
- A higher-value trip component must be disproportionately concentrated at 15–16h.

The second chart offers the missing piece of the puzzle by showing the distribution of airport trips—one of the most profitable trip categories due to their longer distances, higher fares, and frequent surcharges.

- At 15:00, Uber records 4.61M airport trips (9.11%), the highest share of any hour in the top revenue group.

- At 16:00, airport trips reach 4.29M (8.23%), still significantly elevated. After 16:00, however, airport activity declines sharply:

- 17:00 → 6.56%

- 8:00 → 5.40%

- 19:00 → 4.82%

This reveals a consistent behavioral rule:

- The earlier the afternoon, the stronger the surge in airport-driven demand.

- Airport trips deliver far higher revenue per ride, which compensates for the lower total trip count at 15–16h.

- As a result, mid-afternoon hours achieve revenue parity with commuter peaks, even though they are not commuter-heavy hours.

Together, the two charts construct a coherent causal narrative:

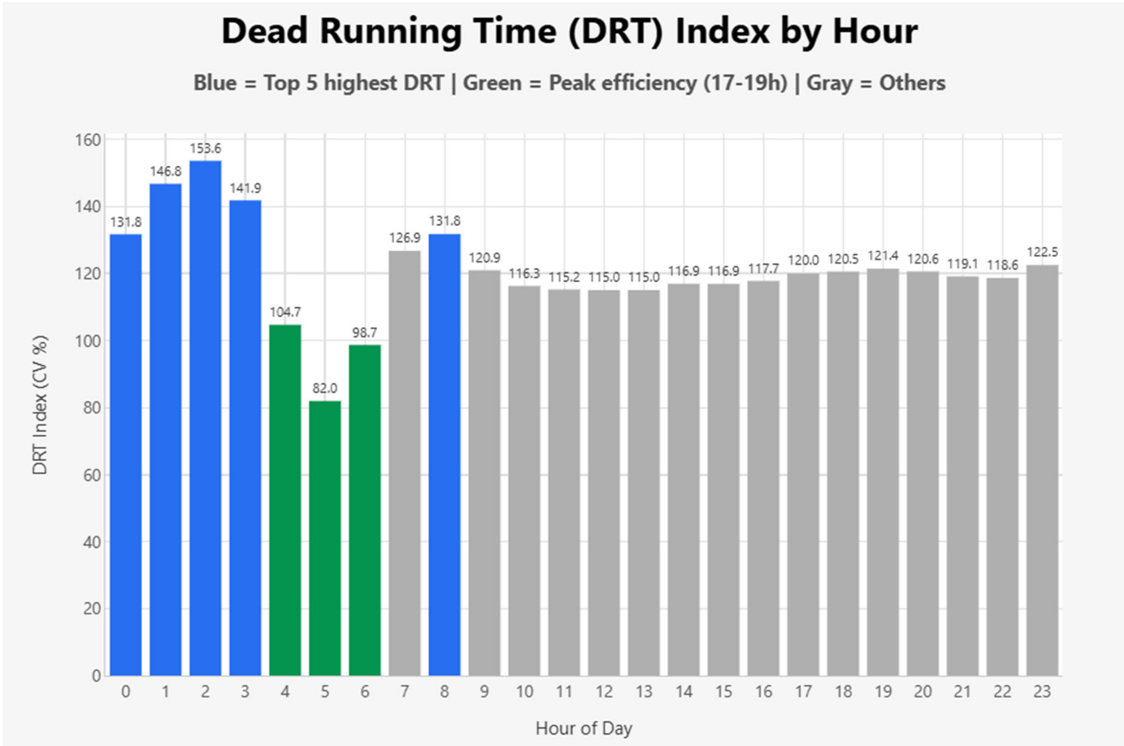
- Chart 1 introduces the anomaly: mid-level trip volume unexpectedly produces peak-level revenue.

- Chart 2 provides the causal mechanism: airport-heavy demand during 15–16h drives revenue upward more efficiently than commuter volume.

- When combined, the two visuals transform what could appear to be a statistical irregularity into a clear behavioral explanation, showing precisely why 15–16h operate like revenue peaks despite being structurally different from the commuter hours of 17–18h.

This pairing not only identifies the anomaly but also uncovers the operational insight: airport mix, not trip count, is the dominant revenue driver in the mid-afternoon window.

6. Demand Recovery Index (DRT)



Concept & Chart Analysis

Why CV Can Represent DRT (Dead Running Time)

In this chart, DRT is expressed using the coefficient of variation (CV%), because CV captures how unstable the hourly trip volume is. The logic is simple:

- High CV = high volatility in demand
  When the number of ride requests swings sharply from one day to another within the same hour, drivers cannot predict flow or chain trips efficiently. Some days that hour is busy, some days it collapses.

- Volatility → idle time → higher DRT
  Whenever demand drops suddenly or fluctuates irregularly, drivers end up waiting longer between trips. That waiting time is exactly what DRT measures.

- Low CV = consistent demand → low DRT
  If an hour shows stable trip volume across months/years, drivers rarely idle. They finish one ride and get matched again smoothly.

So CV acts as a statistical proxy for the same operational behavior that DRT reflects in real life: *Unstable demand creates dead running; stable demand reduces it.*

Instability Zone: Hours 0–3 and Hour 8 (High CV → High DRT)

Hours 0–3: The most chaotic period of the day

From midnight to 3AM, the chart shows the **highest DRT values**:

- 0h → 131.8
- 1h → 146.8
- 2h → 153.6 (peak)
- 3h → 141.9

What does this mean?

- Demand after midnight becomes **extremely irregular** — nightlife, events, bar closings, weather spikes, or nothing at all.

- Rider volume can swing from very high to nearly zero depending on the day.

- Drivers keep circulating because supply stays relatively constant, but demand collapses unpredictably.

This mismatch produces huge volatility, so CV spikes → DRT spikes.
This 0–3h window is the most unstable and least efficient part of the daily cycle.

Hour 8: A secondary instability bump

Hour 8 also jumps to 131.8, unusually high compared to surrounding hours.

Reason:

- At 8AM, many drivers start their shift at once (supply surge).

- But demand increases more smoothly — not a sharp spike.

- The sudden supply shock makes the system temporarily imbalanced.

Result: high CV, hence high DRT, despite daytime conditions normally being stable.

This hour behaves like a "mini early-morning chaos" inside the daytime block.

Most Stable Efficiency Zone: Hours 4–6 (Lowest CV → Lowest DRT)

You requested exactly this: highlight 4h, 5h, 6h as the most stable.

And the chart confirms it:

- 4h → 104.7
- 5h → 82.0 (lowest of the entire 24h)
- 6h → 98.7

Why are these hours stable?

- The late-night randomness is fading out.
- Dawn traffic patterns (early commuters, airport rides, logistics) begin to form.
- Driver and rider flows become predictable and consistent across days.

This block shows the lowest CV values → meaning demand is smooth and balanced. For drivers, this creates:

- short waiting time
- continuous trip chaining
- minimal dead running

This 4–6AM window is the most stable and efficient pre-peak zone in the 24-hour cycle.

Summary

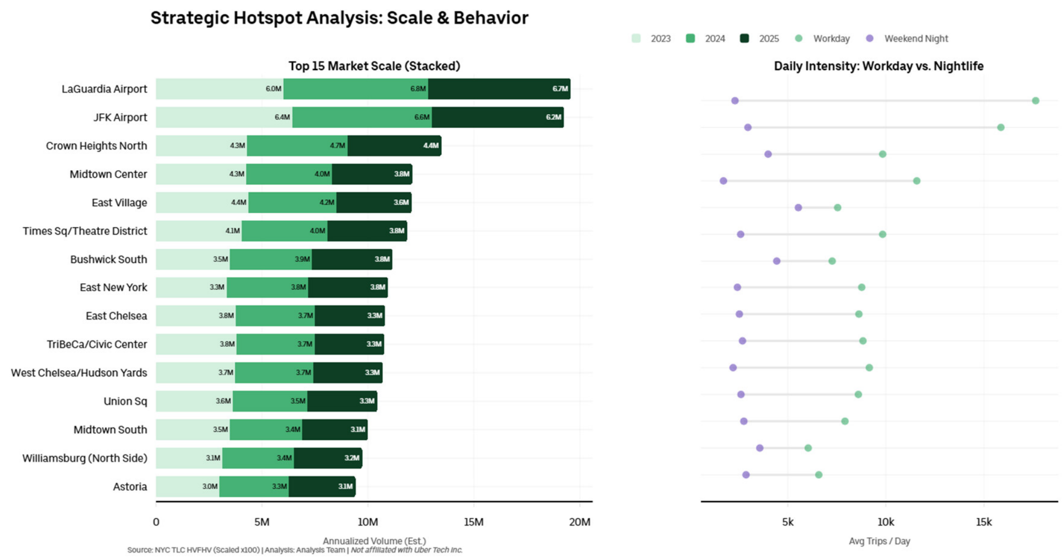
- CV represents DRT because demand volatility directly drives idle time.
- Hours 0–3 and 8 show the highest instability, leading to the worst DRT.
- Hours 4–6 form the most stable, predictable, and efficient early-day block.

III. CURRENT MARKET ANALYSIS

To ensure precision in strategic planning for 2025, this diagnostic section restricts the data scope to the 2023–2025 period.

- Noise Elimination: Excluding data from the pandemic and early recovery phases (2020–2022) is a necessary step to filter out anomalous fluctuations that are no longer representative of the current market.
- Structural Focus: Data from 2023 onwards accurately reflects structural behavioral trends within the "New Normal"—where user mobility habits have stabilized. This provides the most relevant reference point for optimizing current supply.

1. Growth Hotspots



Context & Forecast:

Following the recovery (2023) and stabilization (2024) phases, the 2025 forecast indicates a slight market deceleration in core zones driven by the implementation of the CBD Congestion Fee.

The Pivot: Current hotspots are no longer just areas of high demand; they are signals for Uber to shift from "broad coverage" to "targeted infrastructure investment" (e.g., smart signage, dedicated pickup lanes) to reduce operational friction.

Based on behavioral data, we segment the market into 4 strategic clusters to optimize supply allocation:

JFK & LaGuardia Airports

- Characteristics: Inelastic Demand. Although representing only ~8% of the data, these are "super-nodes" with continuous, dual-direction flows. Customers travel regardless of price.

- Strategy: Reliability First (24/7 Supply)

  These are defensive strongholds. Insufficient supply here results in immediate loss of market share and High Ticket Value revenue to Yellow Taxis and Lyft.

"The Utility Core": Midtown, Penn Station

Characteristics (Workday Logic): Compulsory/Utility Demand.

- Behavior: Dominated by office hours (Mon-Fri). Flows are highly directional (AM:Inbound → PM: Outbound) and predictable.

- Weekend: Becomes a "Ghost Town" with significantly reduced demand.

Strategy: Maximize Reliability & Efficiency, Focus supply during weekdays.

- Action: Deploy Uber Shuttle / Commuter Pass products to aggregate fixed demand and lower costs for frequent commuters.

  Scale down supply on weekend nights to avoid resource waste.

"The Leisure Hubs": East Village, Williamsburg

Characteristics (Weekend Night Logic): Leisure Demand.

- Behavior: Explodes between 10 PM - 3 AM on Fridays and Saturdays. Customers are time-sensitive (need a car fast) rather than price-sensitive.

- Economics: High willingness to pay Surge Pricing and Tips.

Strategy: Yield Maximization.

- Shape supply towards these zones on weekend nights to capture high margins.

- Prioritize Availability algorithms to minimize wait times (reducing friction for intoxicated/impatient riders).

"The Stabilizers": Crown Heights, Bushwick

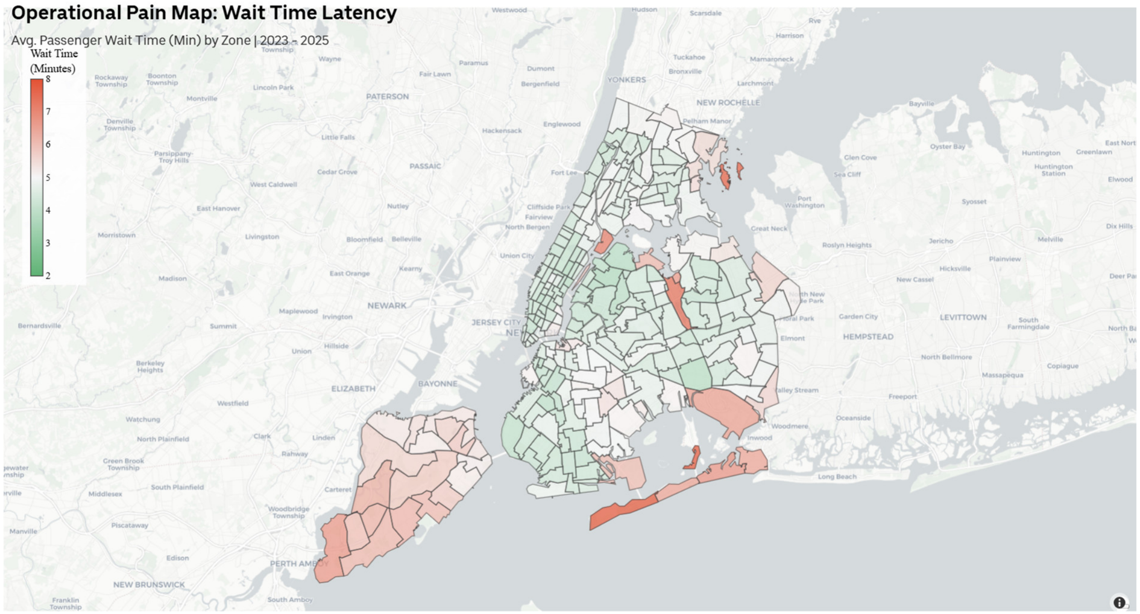
Characteristics: Resilient Demand.

- Consistent usage throughout the week (commuting on weekdays, local travel on weekends).

Strategy: Base Load Maintenance.

- Use these zones as a Supply Buffer on off-peak days (Mondays, Sundays) to help drivers maintain a baseline income when the core zones cool down.

2. Wait Time Analysis



Based on the wait time heatmap, the market is divided into three distinct operational zones requiring specific intervention strategies.

High Liquidity Zones

- Areas: Green zones (Most Regions in Manhattan and the Surroundings).
- Status: Efficient. Supply meets demand immediately.
- Analysis: This area has driver density. Low wait times indicate high liquidity.

Note: Low wait times do not guarantee fast trip speeds. We must cross-reference with Traffic Speed metrics to ensure vehicles are not stuck in congestion immediately after pickup.

Physical Congestion Zones

- Areas: Large orange zones (JFK, LaGuardia Airports).
- Status: Process Bottlenecks.
- Analysis: High wait times here are not caused by a lack of vehicles (supply is abundant). The root causes are complex terminal pickup procedures and sudden demand surges when flights land, creating local bottlenecks.

Structural & Geographic Friction Zones

These are "red alert" areas where the standard operational model fails due to external factors.

A. Staten Island (Economic Structure Issue):

- Status: Record high wait times despite the large land area.
- Cause: Low demand density. Drivers act rationally by avoiding this area to prevent idle time. When a request occurs, the system must dispatch a driver from far away.

B. The Rockaways (Geographic Isolation):

- Status: The "Dead End" effect.
- Cause: The area is isolated by Jamaica Bay with only two entry points. Drivers fear "Deadhead" risks (driving back empty) and toll costs reduce their profit margins.

C. Corona Park (Navigational Friction):

- Status: High pickup latency.
- Cause: The park is too vast without a clear grid system. GPS pin drift causes riders and drivers to lose time trying to locate each other.

TARGETED SOLUTIONS

We propose shifting from generic solutions to tailored mechanisms for each friction type:

Friction Type	Target Area	Strategic Solution	Mechanism
Economic (Low Density)	Staten Island	Uber Reserve	Shift from On-demand to Scheduled rides. This gives the system 30-60 minutes to dispatch a remote driver efficiently without making the rider wait.
Geographic (Tolls/Isolation)	The Rockaways	Return Toll Reimbursement	Automatically add a surcharge to the fare if the trip forces a driver across a toll bridge with a high risk of an empty return trip.

Navigational (Vast Parks)	Corona Park	Smart Meeting Points	Disable random GPS pins. The app only displays fixed "Blue Dots" (main gates/intersections), forcing riders to walk to a feasible pickup spot.
Crowds (Events)	Stadiums	Uber Zone (PIN Code)	Implement a FIFO queue. Riders receive a 6-digit PIN and take the first available car in the line, eliminating the need to find a specific license plate.

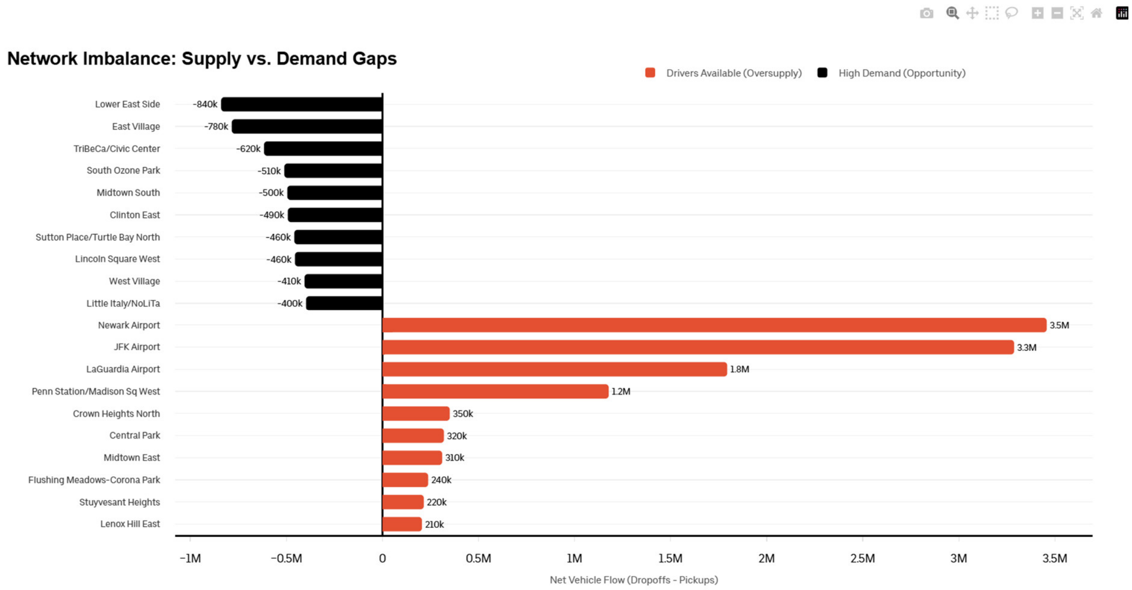
IV. DEEP ANALYSIS OF BOTTLENECKS

1. Dead Mileage

Dead Mileage (Zero-Passenger Miles) represents the most critical inefficiency in the network. It occurs when a driver travels empty from a drop-off point to a new pickup location.

- Operational Cost: Increases fuel consumption and vehicle depreciation without revenue.
- System Friction: Reduces overall fleet availability, triggering unnecessary price surges in adjacent neighborhoods.

We diagnose this by calculating Net Flow Imbalance (Drop-offs minus Pickups) to identify where the network is "leaking" supply.



Excess Drop-offs: JFK, LaGuardia, Penn Station

Data Signal: Massive inflow of vehicles, but significantly lower outflow of active trips.

Root Cause (Modal Split Asymmetry):

- Inbound (To Airport): Passengers prioritize Uber for convenience (luggage handling).
- Outbound (From Airport): Passengers shift to substitutes like the AirTrain, Bus, or the dedicated Yellow Cab queues (cheaper/no wait time).

The "Deadhead" Trap: Drivers dropping off at JFK/LGA face a dilemma: join a long queue in the "Wait Lot" or drive empty back to Queens/Brooklyn. Both options destroy hourly earnings.

Excess Pickups: Lower East Side, East Village

Data Signal: High request volume but low local vehicle turnover (insufficient drop-offs to recycle into new supply).

Operational Result: The system must dispatch cars from distant zones, increasing global ETA and dead mileage across the city.

STRATEGIC SOLUTION: HIGH-CAPACITY CONSOLIDATION

To solve the Airport problems, we must shift from individual unit optimization to high-capacity transit logic.

The Pivot: Uber Shuttle (Hybrid Transit)

- Status (2024-2025): Expansion of direct shuttle routes connecting Manhattan hubs to LaGuardia (LGA) and major event venues.

- Operational Mechanism:

    - Consolidation: Replaces ~50 individual UberX pickups with a single high-capacity vehicle.

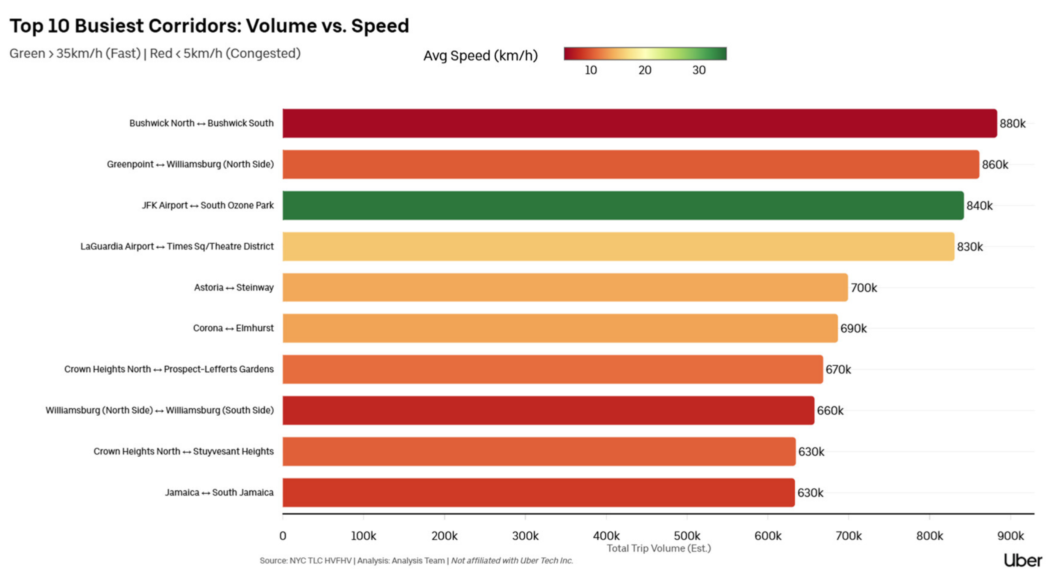
    - Friction Reduction: Eliminates the chaotic pickup at terminals. A single vehicle loads at a dedicated zone, drastically reducing dwell time and curb congestion.

- Strategic Value: Converts the "Dead Mileage" problem into a "Line Haul" efficiency model, capturing the price-sensitive travelers who previously defected to the AirTrain.

2. Traffic Corridors

(Analysis of high-volume routes to identify where the current vehicle-based model is inefficient.)

We define a "Corridor" as a high-frequency pair of Pickup and Dropoff zones. By overlaying speed data onto volume data, we identify "Red Corridors" - routes carrying massive passenger volume but suffering from critically low speeds.



A. Brooklyn Cluster: Intra-Borough Connectivity Gap

The Conflict: Routes between Bushwick North and Bushwick South/Williamsburg are consistently congested (High Volume / Low Speed).

Data Insight:

- Infrastructure Limit: The NYC Subway system is radial (Manhattan-centric), making cross-town travel within Brooklyn difficult.

- Behavioral Mismatch: Young residents use Uber for very short trips (1-3 km) for leisure, clogging narrow one-way streets.

Strategic Solution: Micromobility Integration. This is a prime market for e-bikes and scooters. Integrating Lime/Citi Bike or promoting Uber Moto for trips under 3km will offload traffic from the streets.

B. Queens Cluster: First-Mile Feeder Friction

The Conflict: The corridor from Jamaica to South Jamaica is severely gridlocked.

Data Insight:

- Transit Desert: South Jamaica has high density but no subway access. Residents rely on Uber to reach the Jamaica transit hub.

- Bottleneck: Individual Uber drop-offs at the station entrance compete with buses, causing local paralysis.

Strategic Solution: Aggregation & Smart Zoning.

- Shift to Uber Shuttle or Uber Pool to consolidate these individual trips.

- Coordinate with the MTA to establish "Smart Drop-off Zones" 1-2 blocks away from the main entrance to improve flow.

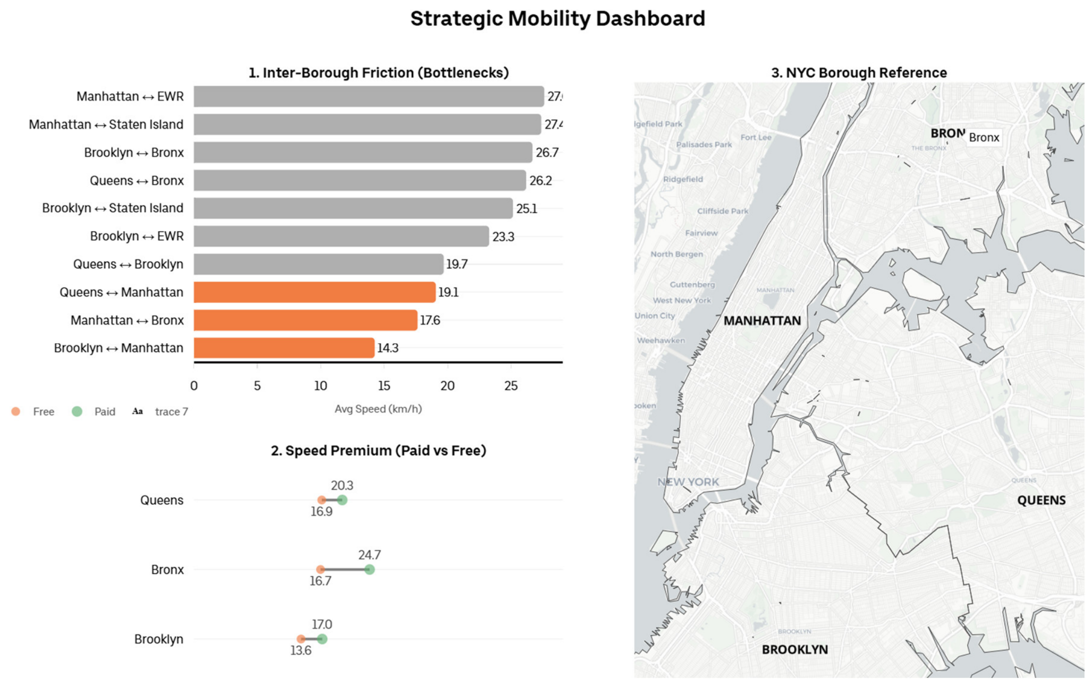
Conclusion:

In these peripheral zones, Uber acts as a bus service. The optimization strategy is not to add more cars, but to facilitate a Modal Shift: Two-wheelers for Brooklyn and Shuttles for Queens.

3. Border Friction

(Investigation into cross-borough connectivity and the economic value of paid infrastructure.)

This analysis utilizes two key datasets to determine if geographical bottlenecks (bridges/tunnels) are the driver of network inefficiency and whether paid routes offer a viable solution.



The data confirms low average speed from Bronx, Queens and Brooklyn to Manhattan. The reason for this might be these regions rely heavily on limited bridge/tunnel access to reach Manhattan. These entry points act as funnels, causing severe congestion that drags down the city-wide average speed.

Conclusion: The physical infrastructure connecting the boroughs is the primary constraint on network velocity, creating a structural ceiling on how fast the fleet can move.

Given that border crossings are the bottleneck, we analyze whether paying for premium infrastructure (Tolls) successfully bypasses this congestion. We calculate the Return on Investment (ROI) of tolls by comparing speed differentials between Paid and Free routes.

A. High Efficiency Zone (Manhattan → Bronx)

- Verdict: "High ROI / Value for Money"
- Data: Paid routes (e.g., Henry Hudson Bridge) are ~8 km/h faster (+50%) than free alternatives.
- Insight: Tolled infrastructure here effectively bypasses local street congestion.
- Action: Default to Paid Route. The app should explicitly highlight the time savings (e.g., "Pay $X to save 20 mins") to encourage conversion.

B. Saturation Zone (Manhattan → Queens)

- Verdict: "Low Marginal Benefit"
- Data: The speed difference is negligible (~3 km/h)
- Insight: Capacity Saturation. The Queens-Midtown Tunnel (Paid) is often just as congested as the Queensboro Bridge (Free). The infrastructure has reached its physical limit; paying extra yields minimal time savings.
- Action: Prioritize Saver Mode. Algorithms should not nudge users toward paid routes unless the time saving exceeds a 5-minute threshold to maintain price competitiveness.

C. Structural Failure Zone (Manhattan → Brooklyn)

- Verdict: "Systemic Gridlock"
- Data: Both options fail. Absolute speeds are critically low for both Free (~14 km/h) and Paid (~18 km/h) routes.
- Insight: Whether using the Battery Tunnel or Brooklyn Bridge, vehicles are trapped in city-wide congestion. Financial instruments (tolls) cannot solve this volume issue.
- Action: Manage Expectations.

  Implement Pre-Surge logic to balance supply.

  Increase ETA Buffer Time to prevent cancellations due to delays.

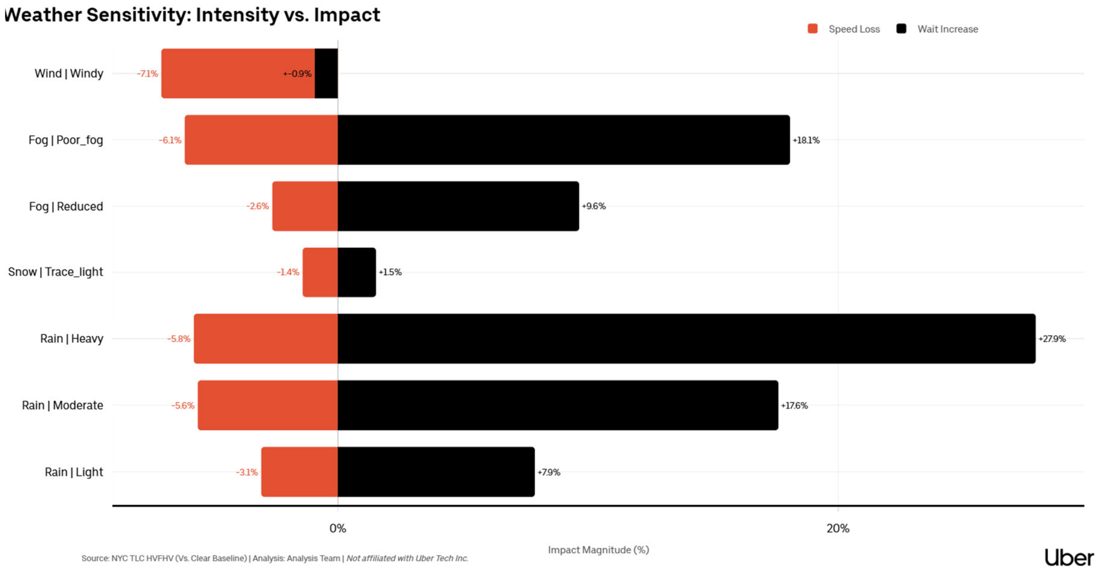
  During peak hours, suggest Modal Shifts (Subway/Moto) near bridge entries.

4. Weather Impact

(Diagnosis of how environmental factors degrade network performance)

Methodology:

We established a "Clear Baseline" (0%) using the median speed of ideal days (No Rain, No Snow, Low Wind, Clear Visibility). We then measured the percentage deviation in Speed and Wait Times for specific weather conditions.



Rain: The Supply Shock

- Data: Speed remains relatively stable (-6%), but Wait Times surge (+28%).
- Demand increases as pedestrians switch to cars, while supply decreases as drivers log off to avoid hassle or risk.

Wind: The Speed Trap

- Data: Shows the sharpest drop in Speed (-7.1%) as vehicles slow down on bridges for safety, yet Wait Times remain stable (-0.9%).
- Insight: The market remains in equilibrium. Strong winds slow the fleet down but do not cause a shortage of drivers.

Fog: The Double Hazard

- Data: Simultaneous drop in Speed (-6.1%) and spike in Wait Times (+18.1%).
- Insight: This represents the worst overall user experience condition, degrading both safety and availability.

Snow: Negligible Impact

- Insight: Impact is minimal in the dataset, likely due to recorded instances being light snow or rapid melting.

STRATEGIC RECOMMENDATIONS

Based on the diagnosis, we propose distinct protocols for Rain and Wind events.

For Rain (Retention Strategy)

- Action: Implement Pre-Surge Incentives.

- Mechanism: Utilize weather forecasting to trigger a Rain Bonus notification to drivers 30 minutes before precipitation starts. The goal is to prevent the initial log-off wave and maintain supply density.

For Wind (Expectation Strategy)

- Action: Dynamic ETA Adjustment

- Mechanism: Automatically add a 7-10% buffer to estimated arrival times when wind speeds exceed 20mph, particularly for routes crossing major bridges. This ensures the app sets realistic expectations for customers.

PART 2: ANALYSIS REPORT

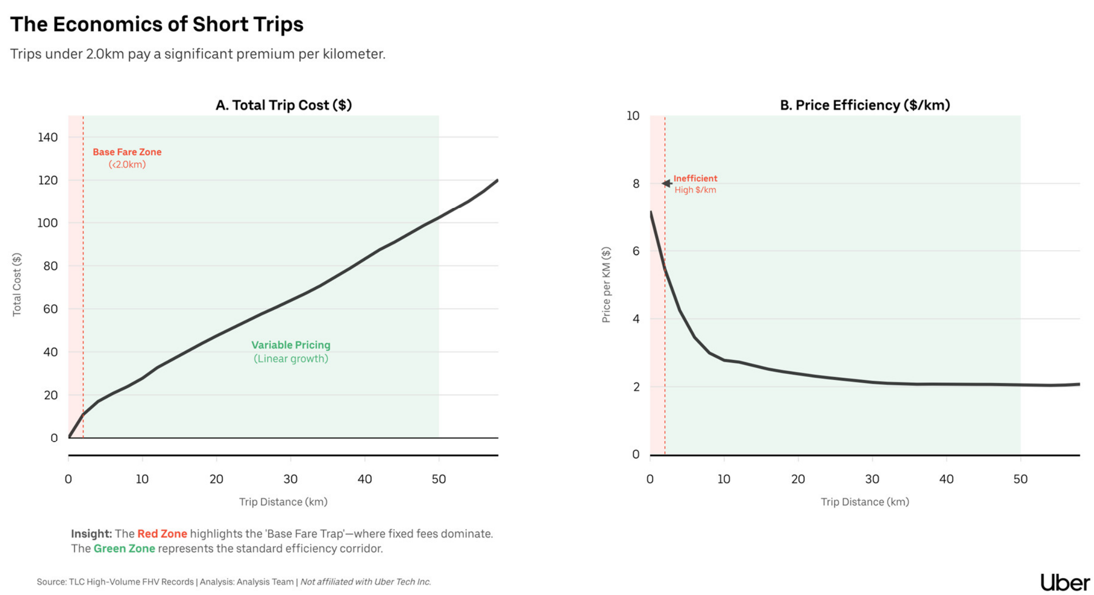
TECHNICAL ANALYSIS & VISUALIZATION METHODOLOGY

To transform complex operational data into actionable insights, this report moves beyond default charting templates to apply the design philosophy of "Explanatory Data Visualization." Grounded in the core principles of Storytelling with Data (Cole Nussbaumer Knaflic), we have developed an "Uber High-Definition" visualization system with a singular ultimate goal: Minimizing Cognitive Load for decision-makers. This section outlines the technical rationale behind the visualizations—ranging from eliminating Clutter, leveraging Preattentive Attributes (strategic color usage), to establishing a clear Narrative Flow - all standardized through the custom uber_style.py library.

PROBLEM 1:

1. Non-Linear Pricing Structure and the Superior Value of Short Trips

Chart: The Economics of Short Trips



Strategic Narrative Role

- Context & Goal: This visualization serves as the foundational "Explanatory" piece in the pricing analysis. It moves beyond simply showing that prices vary (Exploratory) to explaining why specific trips are inefficient.

- Big Message: Its strategic purpose is to prove the existence of a "Base Fare Trap" - a structural anomaly where short trips (<2.5 km) incur a disproportionately high unit cost due to fixed fees. This establishes the business case for potential product differentiation (e.g., distinct "Short Trip" incentives) or rider communication strategies.

Chart Selection Rationale

- Why this chart (Dual-Panel Trend Line)?
    - The analysis requires showing two distinct but related concepts simultaneously: the Absolute Cost (Panel A) and the Relative Efficiency (Panel B).
    - Panel A (Total Cost vs. Distance): A standard line chart is used here to establish the baseline expectation: "Cost increases with distance."
    - Panel B (Price per KM vs. Distance): This second view is critical. It transforms the data to reveal the *rate of change*. Without this panel, the "premium" on short trips is hidden in the intercept of the first chart.
- Why not other charts?
    - Scatter Plot: Rejected because displaying thousands of raw data points creates visual noise ("overplotting") that obscures the underlying pricing formula. The audience needs to see the *rule*, not the *noise*.
    - Bar Chart: Rejected because "Distance" is a continuous variable. Binning it into bars would lose the fidelity of the curve, specifically the sharp exponential decay in unit price at the start.
    - Conclusion: The Smoothed Trend Line is the optimal choice because it abstracts away the variance of individual trips to reveal the clear, mathematical structure of the pricing model.

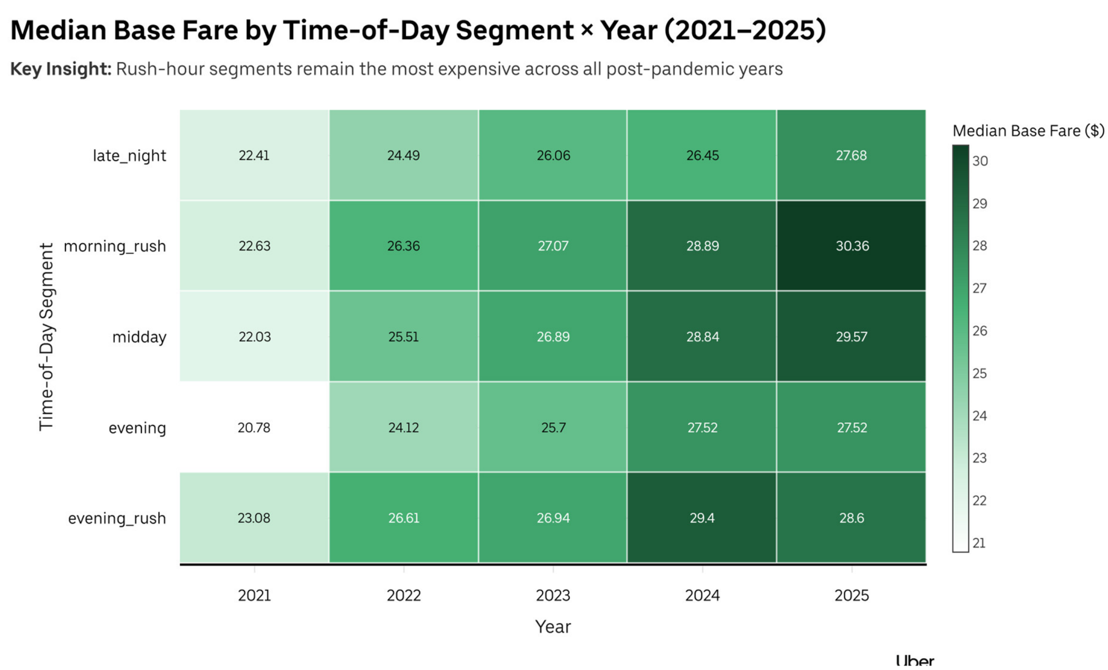
Applied Visualization Mechanics

- Clutter Elimination (Cognitive Load Reduction):
    - Abstraction: The most significant design move was the removal of all raw scatter points. By stripping away the "data ink" of individual rides, the chart forces the eye to focus solely on the trend (the Signal) rather than the variance (the Noise).
    - Direct Labeling: The legend was removed. Instead, labels like "Base Fare Zone" and "Inefficient" are placed directly on the canvas (Gestalt Principle of Proximity), eliminating the need for the eye to scan back and forth.

- Preattentive Attributes (Focus):
    - Semantic Coloring: Color is used strictly for meaning, not decoration.
        - Red (Warning): Highlights the "Short Trip" zone to signal high unit cost/inefficiency.
        - Green (Safe): Highlights the "Standard" zone to signal efficiency.
        - Black (Signal): The trend line is thick and black to ensure it is the primary layer of the visual hierarchy.
    - Enclosure: The background shading (Red/Green rectangles) leverages the Gestalt Principle of Enclosure to group continuous data into two distinct business segments ("Base Fare Zone" vs. "Variable Pricing"), making the abstract math immediately tangible.

2. Pricing by Time and Demand (Time-of-Day & Surge)

Chart: Median Base Fare by Time-of-Day over 2021-2025:



Strategic Narrative Role

- Context & Goal: This heatmap establishes the baseline pricing structure that underpins all later surge and weather analysis. It identifies whether the time-of-day hierarchy—the core driver of Uber's demand cycles—remains stable across post-pandemic years.

- Big Message: The visualization shows that rush-hour segments (morning_rush & evening_rush) consistently produce the highest median fares every year. Despite year-over-year inflation, the relative ranking of time segments does not change. This confirms that time-of-day is the fundamental organizing principle of base fares, and this structure persists across years.

Chart Selection Rationale

- Why this chart (Heatmap)?
    - The analysis compares a categorical × categorical matrix:
        - Time-of-Day Segment
        - Year
    - A heatmap allows patterns of relative intensity to appear immediately through color. It exposes:
        - Consistent high-fare segments
        - Stability of the pricing hierarchy across years
        - How inflation raises all values without reshaping the structure

- Why not other charts?
    - Line Chart: Shows trends but loses the cross-segment matrix structure.
    - Grouped Bar Chart: Too many bars, high cognitive load.
    - Small Multiples: Breaks the unified pattern, making comparison harder.

  The heatmap keeps the entire structure visible in a single, pattern-first view.

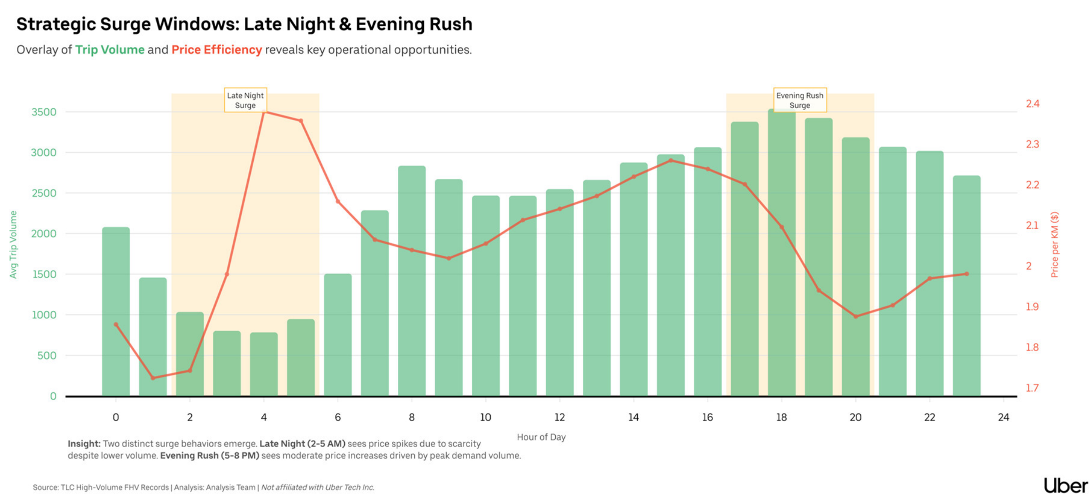
Applied Visualization Mechanics

- Clutter Elimination (Cognitive Load Reduction):
    - Removed cell borders and heavy gridlines
    - Simplified axes
- Preattentive Attributes (Focus):
    - Single sequential green scale communicates fare intensity
    - Darker cells highlight high-fare time segments
    - Lighter cells indicate baseline levels

- Visual Hierarchy & Ordering:
  - Time-of-day arranged by operational flow
  - Years ordered chronologically
  - Centered numeric labels allow quick verification
- Readability & Contrast:
  - Automatic black/white text ensures legibility
  - Expanded left margin prevents truncation

These design choices produce a clear, interpretable visual that reveals a stable, robust pricing structure across years.

Chart: Strategic Surge Windows: Late Night & Evening Rush



Strategic Narrative Role

- Context & Goal: This visualization serves a diagnostic and operational purpose within the broader report. It is not merely displaying data but identifying specific "opportunity windows" for intervention.
- Big Message: The chart is designed to communicate to operations teams that surge pricing is driven by two distinct mechanisms: scarcity (Late Night) versus sheer demand volume (Evening Rush). This distinction is critical for tailoring driver incentive strategies—monetary bonuses for night shifts vs. volume-based quests for rush hours.

Chart Selection Rationale

- Why this chart (Dual-Axis Combo Chart)?
    - The analysis requires the simultaneous comparison of two variables with fundamentally different units and scales: Trip Volume (Count) and Price Efficiency ($/km), over a shared temporal domain (24 hours).
    - Bar Chart (Volume): Bars are chosen for volume to represent the "weight" or magnitude of demand. They visually anchor the chart to the bottom, providing a stable context for the price fluctuations.
    - Line Chart (Price): A line is used for price to emphasize the *trend* and *volatility* across time. The line "floats" above the bars, allowing the viewer to instantly spot divergences—specifically where the red line spikes (high price) while the green bars remain low (low volume), indicating a scarcity-driven surge.
- Why not other charts?
    - Two Separate Charts: Side-by-side charts would make it cognitively difficult to align the peaks and troughs of price against volume hour-by-hour.
    - Scatter Plot: A scatter plot would lose the intuitive temporal sequence (0h to 23h) crucial for operational scheduling.

Applied Visualization Mechanics

- Clutter Elimination (Cognitive Load Reduction):
    - Gridline Hygiene: The secondary Y-axis gridlines were removed, leaving only the primary axis grid. This prevents the "grid prison" effect and avoids the visual confusion of misaligned grid lines, keeping the background clean.
    - Tick Optimization: X-axis ticks are spaced every 2 hours to reduce label density while maintaining readability.
- Preattentive Attributes (Focus):
    - Semantic Coloring: Color is used to encode meaning rather than just category.
        - Green (Volume): Associations with "Go" or positive business activity (Uber Green).
        - Red (Price): Associates high cost/premium with a "Hot" or "Alert" signal.

- Enclosure: The yellow background shading uses the Gestalt Principle of Enclosure to visually group specific hours into "Surge Windows." This pre-attentive attribute highlights the *actionable* timeframes immediately, separating them from the rest of the day.

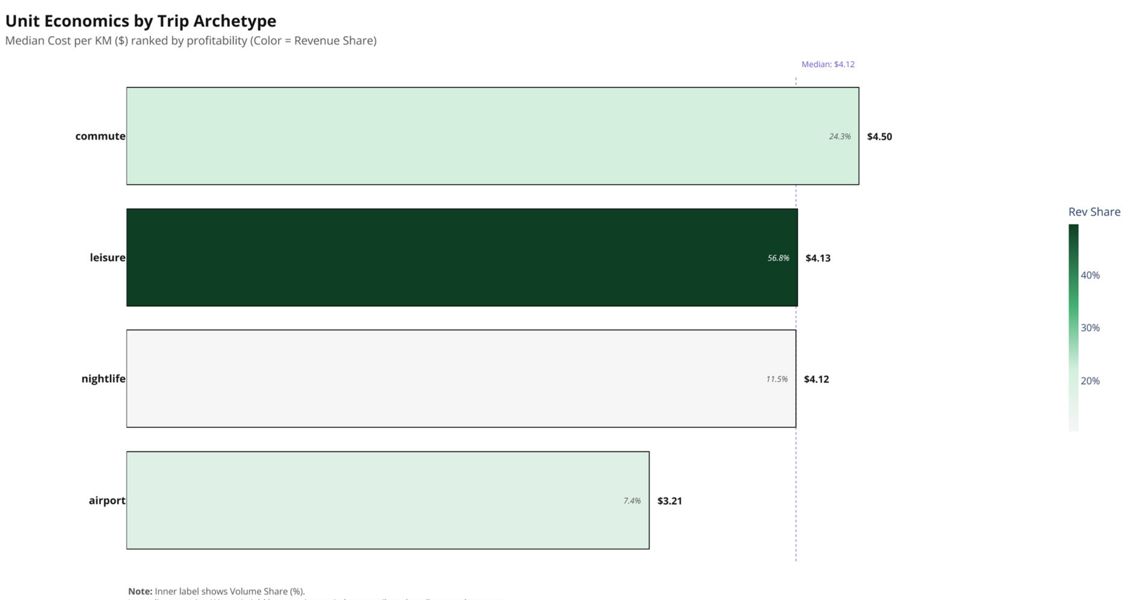
- Visual Hierarchy:

  - Text Integration: The legend is effectively replaced by colored text in the subtitle and direct labeling. This integrates the key into the narrative flow, reducing the need for the eye to scan back and forth to decipher the chart.

  - Action Title: The title "Strategic Surge Windows" immediately tells the user *what* to look for, rather than a generic description like "Volume vs. Price by Hour."

3. Optimization by Trip Archetype

Chart: Unit Economics by Archetype



Strategic Narrative Role

- Context & Goal: This graph serves a crucial Explanatory function within the report. It moves beyond a simple listing of costs to reveal the strategic trade-off between Unit Efficiency (Cost per KM) and Financial Value (Revenue Share).

- Big Message: The visualization is designed to communicate a specific business insight: "High unit prices do not equate to high business value." It highlights that while Short/Micro Trips are expensive on a per-unit basis (high pricing power), the bulk of the platform's revenue actually comes from Airport and Standard City trips (indicated by color intensity), guiding decisions on where to focus driver incentives vs. pricing premiums.

Chart Selection Rationale

- Why this chart (Multivariate Horizontal Bar)?
  - Categorical Ranking: A Horizontal Bar Chart is the standard and most effective choice for comparing and ranking categorical data with long labels (e.g., "Long Commute", "Standard City"). Vertical bars would force label rotation, increasing cognitive load.
  - Multidimensional Encoding: This specific design allows for the simultaneous comparison of three distinct metrics without clutter:
    - Length = Cost Efficiency (The primary sorting metric).
    - Color = Revenue Impact (The secondary "value" metric).
    - Label = Volume (Contextual detail).
- Why not other charts?
  - Scatter Plot: While a scatter plot could show Cost vs. Revenue, it fails to provide the clear, ordered ranking that a bar chart offers. Stakeholders need to see "Who is the most expensive?" immediately.
  - Grouped Bar Chart: Using separate bars for Cost and Revenue would double the ink and make the correlation harder to spot. The current design integrates them into a single visual object.

Applied Visualization Mechanics

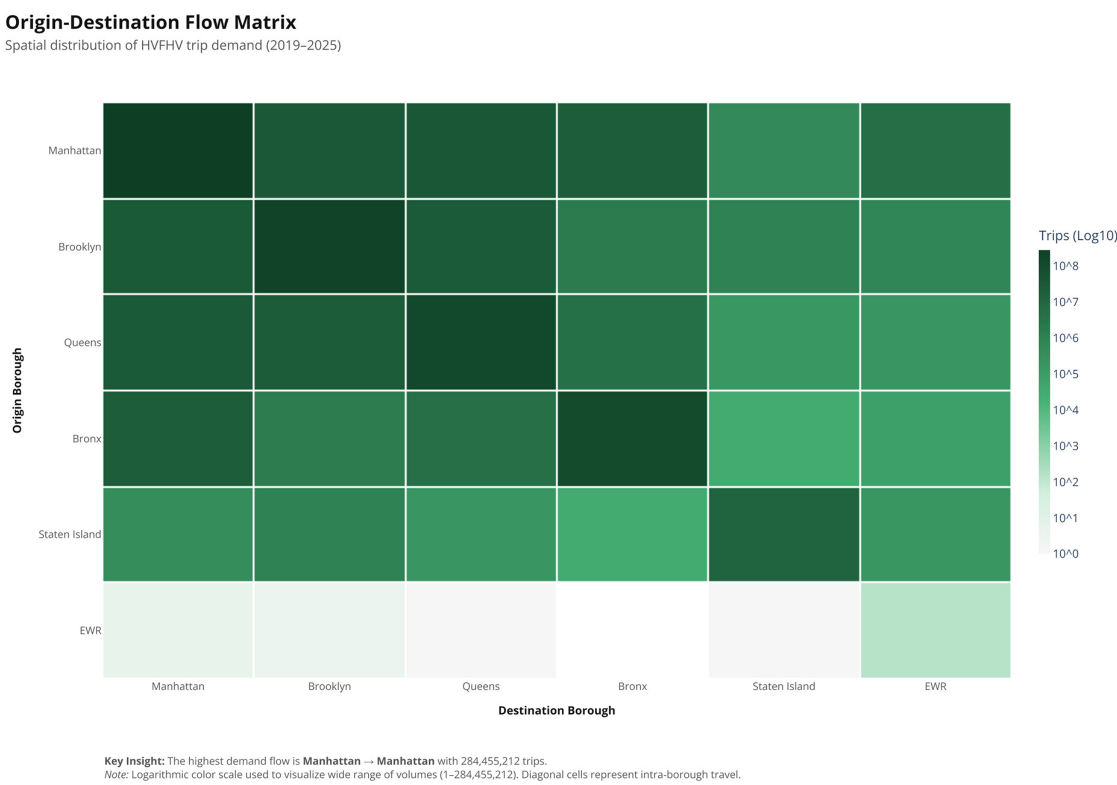
- Clutter Elimination (Cognitive Load Reduction):
  - Axis Removal: The X-axis line and ticks were completely removed. Following SWD principles, Direct Labeling places the specific dollar values (e.g., $8.20) at the end of each bar, eliminating the need for the eye to travel back and forth to an axis to estimate values.
  - Gridline Removal: Redundant gridlines were stripped away to leave a clean canvas where the data itself forms the structure.

- Preattentive Attributes (Focus & Hierarchy):
    - Semantic Coloring: The UBER_GREEN sequential scale is used pre-attentively. Darker green instantly draws the eye to high-revenue segments (Airport/Standard City), creating a visual hierarchy based on financial importance rather than just bar length.
    - Reference Anchoring: A vertical dotted line marks the "Global Median" ($3.99). This acts as a visual anchor, instantly segmenting the archetypes into "Premium Priced" (Right of line) and "Economy Priced" (Left of line), aiding rapid interpretation.

4. Pricing by Geography and Weather

4.1. Geographic Analysis and Intra-borough Trends

Chart: Origin-Destination Flow Matrix



Strategic Narrative Role

- Context & Goal: This visualization serves as the primary Exploratory overview of the network topology. It establishes the "ground truth" of where demand is concentrated before diving into temporal or pricing specifics.

- Big Message: The strategic purpose is to reveal the structural inequality of the network. It highlights that while the network is theoretically fully connected ($N \times N$), the actual demand follows a Power Law distribution where specific corridors (e.g., Manhattan $\rightarrow$ Manhattan) dominate the entire system, necessitating distinct operational strategies for "Core" vs. "Peripheral" routes.

Chart Selection Rationale

- Why this chart (Heatmap/Matrix Diagram)?
    - Data Density: The dataset represents a dense, multidimensional matrix (Origin $\times$ Destination). A heatmap is the most efficient way to display quantitative values for every possible permutation without overlap.
    - Pattern Recognition: It utilizes the Gestalt Principle of Proximity. By arranging boroughs on orthogonal axes, the viewer can instantly detect macro-patterns (e.g., the "diagonal" representing intra-borough travel) and outliers (high-demand inter-borough flows) without tracing complex lines.
- Why not other charts?
    - Sankey Diagram: Rejected because a fully connected Sankey with high variance in flow volume results in "visual spaghetti"—a tangled mess of lines that obscures the secondary routes.
    - Chord Diagram: Rejected because circular layouts often make it harder to compare magnitudes across non-adjacent nodes compared to the linear grid of a heatmap.

Applied Visualization Mechanics

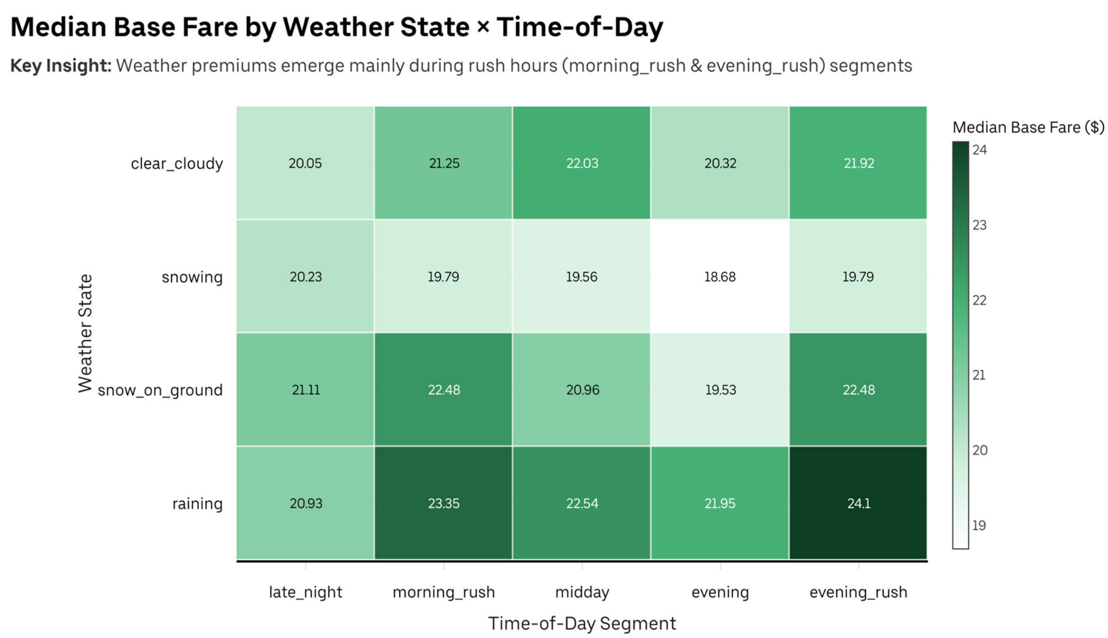
- Clutter Elimination (Cognitive Load Reduction):
    - Gridline Removal: Standard axis gridlines were removed. Instead, the white space between cells (xgap, ygap) creates a natural grid, defining boundaries without adding non-data ink.
    - Axis Optimization: Ticks were removed, and the Y-axis was reversed to match the intuitive "reading" order of a matrix (Top-Left to Bottom-Right).
- Preattentive Attributes (Focus & Hierarchy):
    - Logarithmic Scaling (log_10): This is the critical design mechanic. A linear color scale would render 90% of the chart white due to the massive outlier of Manhattan intra-trips. The Log transformation compresses the dynamic range, allowing the color saturation to reveal the structure of secondary flows (the "Long Tail") that would otherwise be invisible.

- Semantic Coloring: The UBER_GREEN sequential scale is used. Darker green intuitively signals "High Volume/Density," leveraging color intensity to encode magnitude pre-attentively.
- Accessibility: While the *visual* is logarithmic, the *interaction* (tooltip) displays the absolute integer value, ensuring data integrity is maintained for detailed inspection.

4.2. Nonlinear Impact of Weather

4.2.1. Raining – Supply Shock:

Chart: Weather Premiums by Time-of-Day (Median Fare Heatmap)



Strategic Narrative Role

- Context & Goal: This visualization plays an explanatory role in the pricing story. It is used to diagnose where weather actually matters for baseline pricing, not just to show that "rain is expensive."
- Big Message: The chart communicates that weather premiums are segment-specific, not system-wide. Clear/cloudy, raining, snowing, and snow_on_ground do not uniformly lift prices – they create noticeable premiums mainly in evening and evening_rush, while morning and midday remain far less sensitive. This distinction is important for strategy: weather-based pricing levers should be focused on congested evening periods, not across all time segments.

Chart Selection Rationale

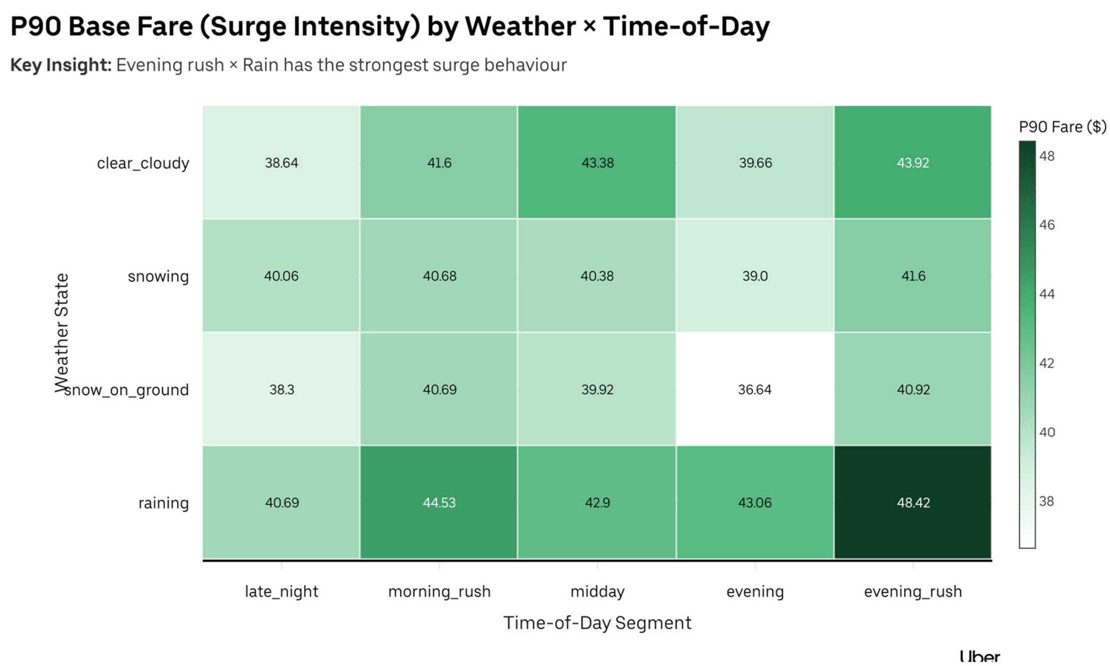
- Why this chart (Heatmap)?
    - The core analysis is a categorical × categorical comparison: Weather State × Time-of-Day Segment, with an intensity metric (Median Base Fare).
    - A heatmap allows the viewer to see both dimensions simultaneously, using color intensity to reveal clusters where weather and time-of-day interact to create higher fares.
    - It is optimized for pattern recognition: the viewer can instantly identify "hot rows + hot columns" instead of reading dozens of bars.
- Why not other charts?
    - Grouped Bar Charts: Would require separate clusters for each weather–time pair; this explodes into many bars and makes it hard to see the overall structure or cross-row patterns.
    - Multiple Line Charts / Small Multiples: Fragment the matrix into separate plots, forcing the viewer to mentally stitch patterns back together. The interaction effect becomes harder to perceive.
    - Single-number summaries (tables): Lose the spatial structure and visual gradient that make the weather–time interaction obvious.

Applied Visualization Mechanics

- Clutter Elimination (Cognitive Load Reduction):
    - Inner cell borders and heavy gridlines are removed, leaving a clean data matrix.
    - The default vertical y-axis title is replaced with a single, clear annotation, avoiding rotated text and label collisions.
    - Axis labels are minimal and readable; no redundant legends are used.
- Preattentive Attributes (Focus):
    - A single sequential green scale encodes fare intensity. Darker cells immediately flag weather–time combinations with higher median fares (e.g., raining × evening_rush).
    - Numbers are placed in the center of each cell, with automatic black/white contrast depending on background shade, ensuring legibility without extra visual noise.

- Visual Hierarchy & Ordering:
    - Weather states are ordered from more benign to more severe, and time-of-day segments follow the operational cycle, making the gradient of "how bad + when" intuitive to scan.
    - The title and subtitle play an action role: they tell the reader the key insight ("weather premiums emerge mainly during rush segments") before they begin decoding numbers, framing interpretation up front.

Chart: Weather-Driven Surge Intensity (P90 Fare Heatmap)



Strategic Narrative Role

- Context & Goal: This visualization serves as a risk and surge-diagnostics tool within the pricing report. Its goal is to show where the tail of the fare distribution (P90) spikes, not just the typical price.
- Big Message: The chart communicates that high-intensity surge is concentrated in a small set of conditions, especially raining × evening_rush, while other weather–time combinations remain relatively moderate. This supports the strategic distinction between "normal weather premium" and "true stress conditions" that require special surge and supply management.

Chart Selection Rationale

- Why this chart (Heatmap)?

- As with the median heatmap, the data is inherently a matrix: Weather State × Time-of-Day Segment, but now encoded with P90 Base Fare as a proxy for surge intensity.

- A heatmap is ideal for showing where the extremes live: darker cells immediately reveal the hotspots of operational risk (very high fares) without forcing the viewer to scan a cluttered set of bars or lines.

- Using the same structure and layout as the median heatmap allows a direct mental comparison between "typical price" and "tail risk" under the same categorical grid.

- Why not other charts?

  - Bar / Column Charts: With many combinations, bars become dense and difficult to compare across both dimensions. Identifying extreme pockets (e.g., only some cells in a row) is harder.

  - Boxplots by category: Good for distribution, but poor at communicating a 2D interaction; we would need many separate panels.

  - Scatter Plot: Would lose categorical alignment and make it harder to answer "which specific weather × time block is risky?"

Applied Visualization Mechanics

- Clutter Elimination (Cognitive Load Reduction):

  - The visual grammar mirrors the median heatmap: no cell borders, minimal gridlines, and simplified axes to keep attention on the surge pattern.

  - Titles and subtitles explicitly state that this chart is about "P90 / surge intensity", differentiating it from the median version and reducing interpretive ambiguity.

- Preattentive Attributes (Focus):

  - The same sequential green scale is reused so that darker ⇒ more extreme, but here the viewer is primed to interpret dark cells as surge risk zones.

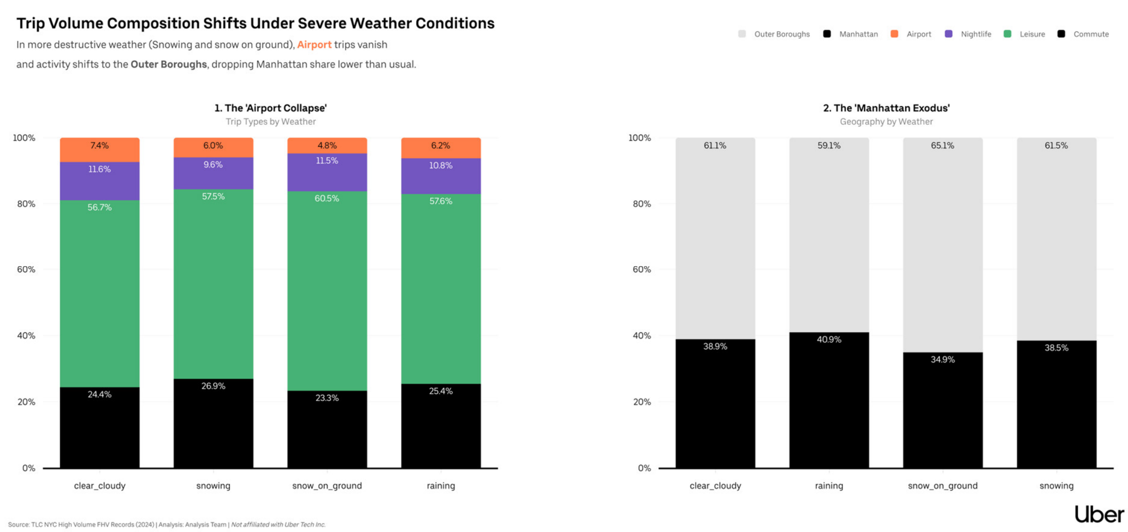
  - The contrast-aware text labels enable the viewer to read exact P90 values only where needed, while relying on color for broad pattern recognition.

- Visual Hierarchy & Ordering:

- Weather states and time segments are ordered identically to the median heatmap, so the viewer can mentally compare both charts cell-by-cell without relearning the layout.
- The emphasis is on cluster recognition: the darkest band in raining × evening_rush visually pops as the key risk area, aligning directly with the narrative that weather amplifies surge most strongly in the evening rush window.

4.2.2. Snow – Demand Destruction

Chart: Composition Shift Under Snow Weather



Strategic Narrative Role

- These dual 100% stacked bar charts act as evidence for a core demand-side claim: snow does not simply lower total trip volume — it reorganizes demand. The visual clarifies which trip types collapse most sharply when snow appears and how the spatial distribution of activity shifts away from Manhattan toward the Outer Boroughs.
- Big Message: Snow triggers a structural composition collapse, not a uniform decline. Airport trips experience the steepest contraction. Nightlife and leisure shrink materially as discretionary mobility disappears. Commute grows in share only because it becomes the last remaining stable category. Geographically, Manhattan loses relative weight under snow while Outer Borough share rises, revealing a weather-driven redistribution of movement.

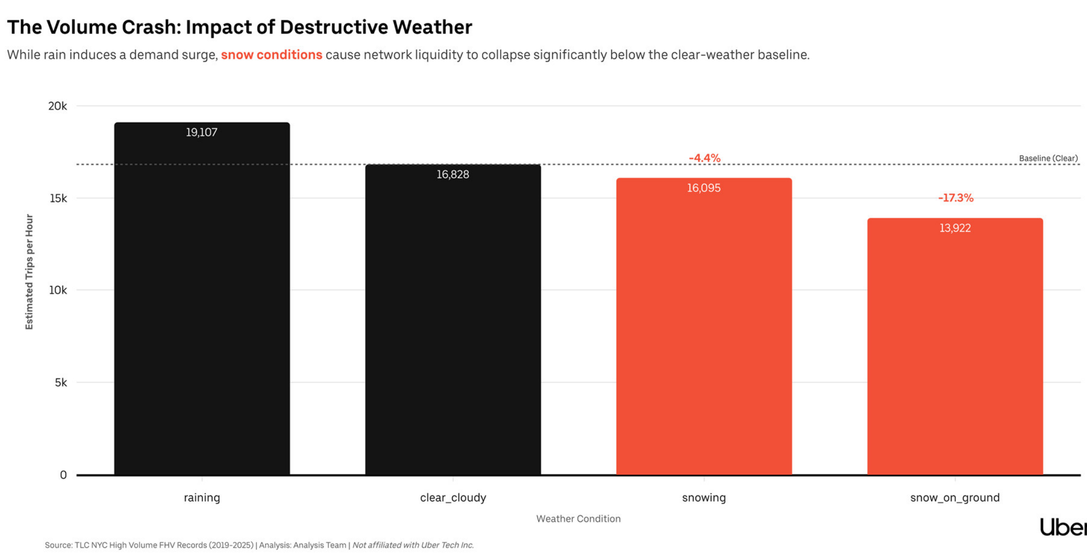
Chart Selection Rationale

- Dual 100% stacked bars are chosen because the analytical question concerns relative composition, not absolute counts. This format makes snow-driven distortions immediately visible and allows a clean side-by-side comparison of behavioral categories (Panel 1) and geography (Panel 2). Alternatives such as regular bars, line charts, pies, or heatmaps fail to show multi-category share shifts as clearly or comparably. The selected form directly supports the narrative of disproportional category-level disruption.

Applied Visualization Mechanics

- A fixed 0–100% frame reduces cognitive load and lets the viewer track share movement across weather states. Light gridlines and direct percentage labeling minimize friction. High-saturation colors highlight categories that collapse under snow, while core mobility categories use calmer tones to position them as what remains when discretionary travel disappears.
- The left-to-right flow mirrors the analytical story: first the behavioral collapse, then the spatial redistribution. Color semantics remain consistent across panels, reinforcing pattern recognition — particularly the airport contraction that defines the Airport Collapse phenomenon.

Chart: The Volume Crash Under Snow Weather



Strategic Narrative Role

- This chart provides quantitative confirmation of how destructive weather distorts total trip liquidity. It differentiates rain-driven demand surges from snow-driven network failures, showing that snowfall pushes hourly trip volume below the clear-weather baseline. The visual supports the broader claim that snow is not merely disruptive but economically contractionary for the mobility network.
- Big Message: Rain increases overall trip activity, but snow causes a volume crash. Clear/cloudy conditions serve as the baseline, yet snowing drops volume by roughly 4–5%, and snow_on_ground deepens the decline to more than 17%. The

severity of snow accumulation—not just snowfall itself—drives a measurable collapse in network liquidity.

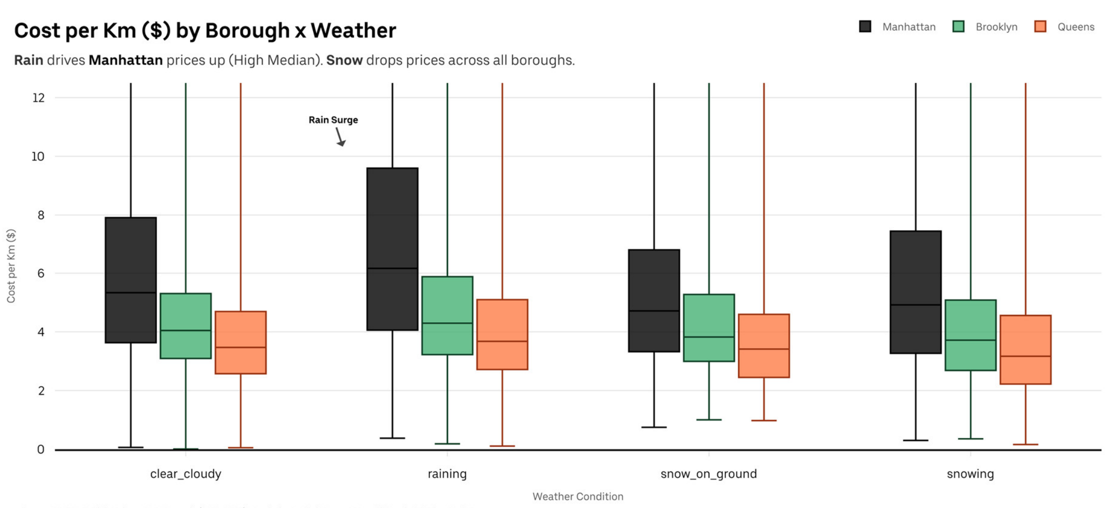
Chart Selection Rationale

A grouped bar chart is used because the analysis focuses on absolute trip volume, not composition. The height of each bar provides an intuitive read of how weather shifts total activity relative to the clear-weather reference line. Alternative chart types fail to communicate this as clearly: line charts imply continuity that weather categories do not have; 100% bars hide magnitude; pie charts cannot show relative deltas; heatmaps lose scale precision. Bars paired with a horizontal baseline make the drop under snow visually explicit.

Applied Visualization Mechanics

The dotted baseline anchors viewer interpretation, turning each deviation into an immediately readable gain or loss. High-contrast black bars emphasize normal and rain conditions, while saturated red highlights the destructive impact of snow. Percentage annotations above snowfall bars reduce cognitive load by surfacing the magnitude of decline without requiring mental calculation. The uncluttered vertical axis and wide spacing reinforce the narrative of a weather-driven volume crash.

Chart: The Price War — Cost ($/km) by Borough × Weather (Boxplot Panel)



Strategic Narrative Role

- This visualization explains how weather-driven inefficiencies translate into unequal pricing pressure across boroughs. It extends the story from demand disruption to price-performance deterioration, showing that adverse weather amplifies existing structural differences between Manhattan, Brooklyn, and Queens.

- Context & Goal: The chart is designed to compare how cost-per-km distributions shift under rain and snow. Because each borough has a unique mix of trip lengths, congestion patterns, and network constraints, weather impacts their price efficiency unevenly. The goal is to surface these asymmetries clearly and statistically.
- Big Message: Adverse weather produces a borough-specific price stress pattern. Manhattan maintains the highest medians and widest spreads across all weather conditions, reflecting congestion and short-trip density. Queens becomes increasingly inefficient under rain and snow, showing elevated tails driven by longer-distance trips slowed by poor road conditions. Brooklyn remains comparatively stable but still shows upward spikes under snow, indicating system-wide friction.

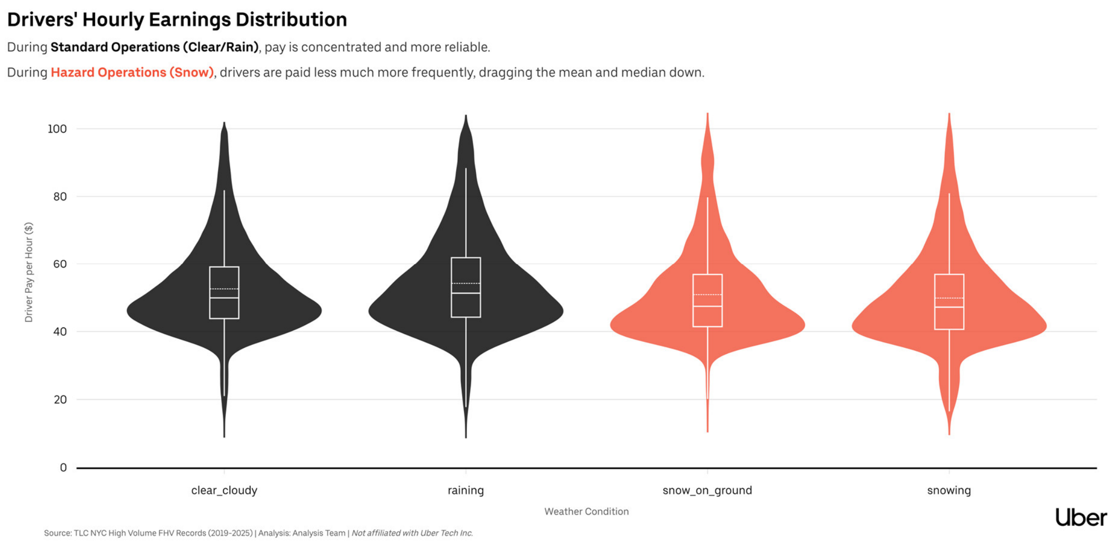
Chart Selection Rationale:

Grouped boxplots are chosen because the analysis requires comparing entire distributions, not point estimates. Cost-per-km is skewed and highly variable during bad weather; neither means nor medians can capture volatility, outliers, or tail behavior. Boxplots make spread, skew, and extreme values visible and comparable across boroughs and weather categories. Alternatives—bars, line charts, densities, scatterplots—fail either by hiding distribution shape or by overcomplicating multi-category comparisons. Boxplots provide the clearest statistical and comparative structure for this question.

Applied Visualization Mechanics:

Gridlines are kept minimal to reduce distraction, allowing the eye to focus on box shapes and whisker behavior. The boxes remain narrow to minimize ink while preserving distribution detail. Even spacing across weather states avoids crowding and keeps categorical comparisons readable. Distinct semantic colors differentiate boroughs, supporting immediate recognition and reducing cognitive switching. Snow conditions naturally draw attention through wider spreads and longer whiskers, reinforcing the narrative of weather-driven inefficiency. The left-to-right ordering follows the escalation of weather severity, while consistent color encoding across categories supports rapid borough-level comparisons.

Chart: Drivers’ Hourly Pay ($/hr) by Weather (Violin Plot)



Strategic Narrative Role

- This visualization delivers the supply-side perspective in the weather impact analysis. Prior visuals showed how rain and snow distort demand and pricing; this chart answers the operational question for drivers: “How does severe weather translate into actual hourly earnings?” It clarifies whether weather volatility creates opportunity, risk, or both.
- Context & Goal: The goal is to compare how reliably drivers earn under different weather states by examining the full distribution of hourly pay — not just averages. This helps explain driver behavior during hazard conditions, especially why some drivers stay online while others log off when snow hits.
- Big Message: The violins reveal that hourly earnings remain broadly centered around similar median levels in clear and rainy conditions, indicating stable operational performance. Under snow, however, earnings become more volatile:
  - The upper tail stretches higher, reflecting occasional high-paying trips.
  - The distribution widens, signaling inconsistent trip flow and longer idle time.
  - The median shifts slightly downward, showing that most drivers earn less even though rare spikes occur.

This supports the narrative that snowfall introduces instability — a mix of high-reward outliers and lower baseline earnings due to reduced trip turnover.

Chart Selection Rationale

- A violin plot with an embedded boxplot is selected because driver earnings are non-normal, heavy-tailed, and highly sensitive to weather conditions. The chart reveals density, central tendency, and extremes simultaneously — all essential for interpreting volatility.

- Bar charts hide spread; standalone boxplots miss distribution shape; line charts imply continuity that weather categories do not have. The violin-box hybrid is the most accurate representation of pay dynamics under weather stress.

Applied Visualization Mechanics

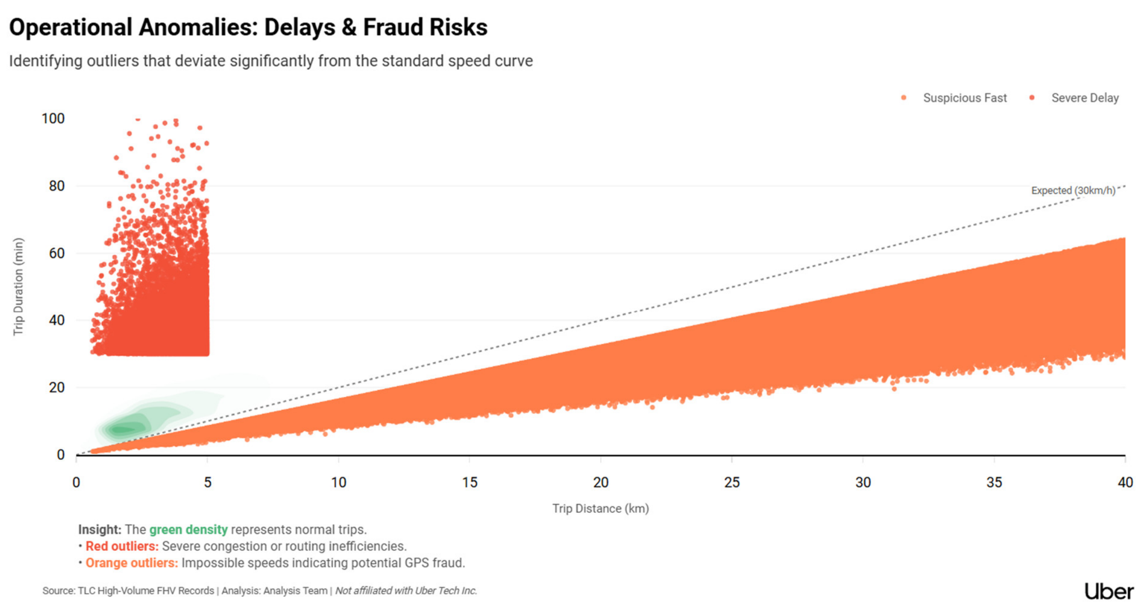
- Visual clutter is minimized to emphasize distribution shape and boxplot structure. Symmetrical violins allow quick weather-to-weather comparison. Consistent color encoding separates standard operations (clear/rain) from hazard operations (snow), drawing attention to the instability that snow introduces. Narrow boxplots highlight median shifts and interquartile compression/expansion.
- Weather categories are arranged progressively from clear to snow, reinforcing the escalation of variance. The broader violins in snow states naturally pull viewer focus toward the volatility that defines the hazard condition narrative.

II. Anomaly detection & Risk management

1. Severe Congestion and Operational Risk

1.1. Severe Congestion ("Slower than Walking")

Chart: Operational Anomalies (Delays & Fraud Risks)



Strategic Narrative Role

- Context & Goal: This visualization serves as the primary Diagnostic tool for the Risk Management section of the report. Its goal is to move beyond reporting *average* performance (which hides issues) to explicitly identifying edge cases that threaten platform reliability and integrity.

- Big Message: The strategic purpose is to enable "Management by Exception." It visually separates the 95% of "Normal" trips (Green) from the 5% of "Critical" trips (Red/Orange), forcing stakeholders to focus on two distinct operational threats: Severe Inefficiency (Congestion) and Potential Fraud (Impossible Speeds).

Chart Selection Rationale

- Why this chart (Hybrid Density-Scatter Plot)?
  - Variable Relationship: A Scatter Plot is the only effective way to visualize the correlation between two continuous variables—*Distance* and *Duration*—to detect deviations from the expected linear relationship (Speed).
  - Handling Big Data: A standard scatter plot with 60,000 points results in severe overplotting, making it impossible to distinguish between 10 points and 1,000 points in a dense area.
  - The Solution: A Hybrid Approach is selected to solve this (Lesson 11):
    - Density Contour (Heatmap): Used for the "Mass Market" data to show *where* the volume is without visual noise.
    - Scatter Points: Reserved *only* for the outliers. This allows individual inspection of the problem cases while summarizing the healthy majority.

Applied Visualization Mechanics

- Clutter Elimination (Cognitive Load Reduction):
  - Trace Removal: The redundant legend entry ("Trace 1") for the reference line was explicitly removed, cleaning up the visual interface.
  - Axis Simplicity: Gridlines are restricted to the Y-axis and kept light (GRAY_100) to aid in estimating duration magnitude without creating a "grid prison."
- Preattentive Attributes (Focus & Hierarchy):
  - Semantic Coloring: The color palette serves a functional taxonomy:
    - Green (Ground): Represents the "Safe Zone." The use of a gradient/contour reinforces it as a statistical distribution (Context).
    - Red (Signal): Triggers an immediate "Warning" for Severe Delays. These points lie *above* the expected slope.

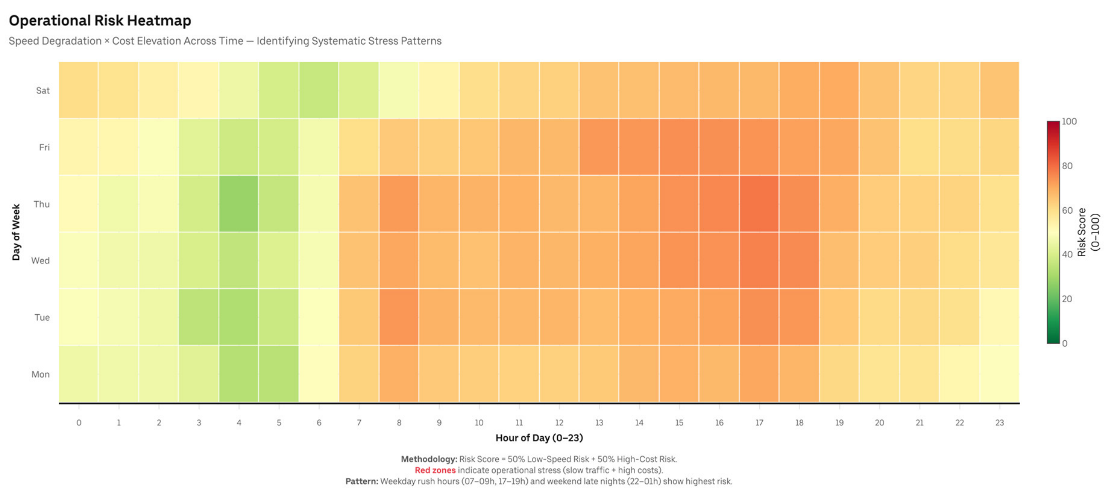
- Orange (Signal): Differentiates "Fraud Risk" from delays. These points lie *below* the physical limit line (impossible speeds).

  - Visual Hierarchy: A clear Figure-Ground relationship is established. The green density fades into the background, pushing the vivid Red and Orange markers to the foreground, directing the executive's eye immediately to the exceptions.

  - Direct Labeling: The "Expected Speed (30km/h)" line is labeled directly on the chart (Gestalt Principle of Proximity), removing the need for a legend lookup.

1.2. Highest Operational Risk Time Window

Chart: Operational Risk Heatmap



Strategic Narrative Role

- Context & Goal: This visualization acts as the Operational Compass for the report. It transitions the analysis from investigating *what* happened (e.g., delays, costs) to pinpointing *when* the system is most vulnerable.

- Big Message: The strategic purpose is to reveal Systematic Stress Patterns. By identifying specific time windows (e.g., Friday evenings, Monday mornings) where high costs and low speeds intersect, it enables operations managers to move from reactive fire-fighting to proactive scheduling of incentives and supply adjustments.

Chart Selection Rationale

- Why this chart (Heatmap)?

  - Multi-Dimensional Density: The analysis requires visualizing the intersection of two temporal dimensions (Day of Week vs. Hour of Day). A

Heatmap is the only effective way to display this 7 * 24 = 168 point matrix without the visual clutter of 7 overlapping lines or 168 separate bars.

  - Pattern Recognition: It leverages the Gestalt Principle of Continuity, allowing the eye to instantly detect "hotspots" (clusters of red) and "cool zones" (clusters of green) across the grid, facilitating rapid temporal diagnosis.

- Data Transformation:

  - Composite Metric: Instead of plotting raw speed or cost, the chart uses a constructed "Risk Score" (50% Speed Degradation + 50% Cost Elevation). This analytical abstraction simplifies decision-making by providing a single, normalized "Health Index" (0–100) for the platform.

Applied Visualization Mechanics

- Preattentive Attributes (Focus & Hierarchy):

  - Semantic Coloring: The Red-Yellow-Green reversed diverging color scale is used.

    - Red: Intuitively signals "High Risk/Danger," drawing immediate attention to stress periods.

    - Green: Signals "Optimal Operation," pushing low-risk periods to the background.

  - Colorbar Context: The legend is positioned unobtrusively to the right, clearly labeled "Risk Score (0–100)" to provide quantitative backing to the qualitative color perception.

- Clutter Elimination:

  - Gridline Management: Axis gridlines are minimized (GRAY_200), serving only as subtle guides to separate hours/days without breaking the visual flow of the heat blocks.

  - Axis Optimization: Ticks are aligned to hour integers, and the Y-axis uses mapped labels (Mon-Sun) rather than raw numbers, reducing cognitive load.

- Narrative Integration:

  - Action Title: "Operational Risk Heatmap" defines the chart's utility immediately. The subtitle explains the methodology ("Speed Degradation × Cost Elevation"), building trust in the metric.

- Annotated Insight: The caption decodes the visual pattern ("Weekday rush hours... weekend late nights"), ensuring the stakeholder walks away with the correct operational conclusion.

2. Management of Passenger and Driver Behavior

2.1. Tipping behavior

Chart:

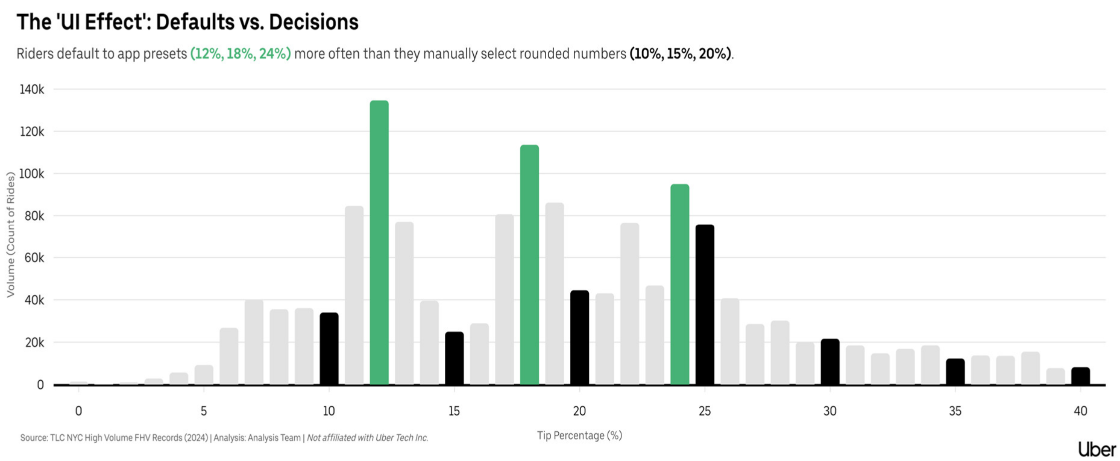
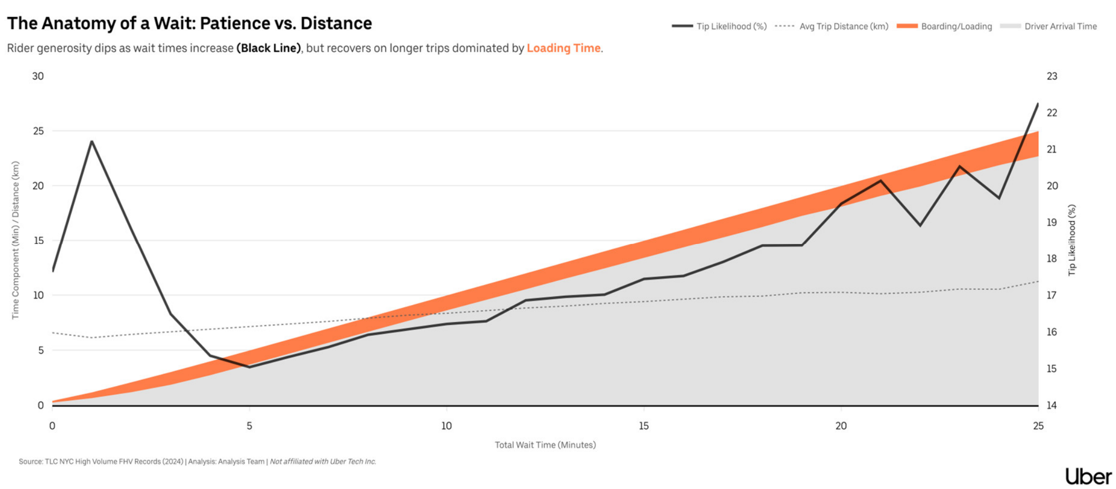
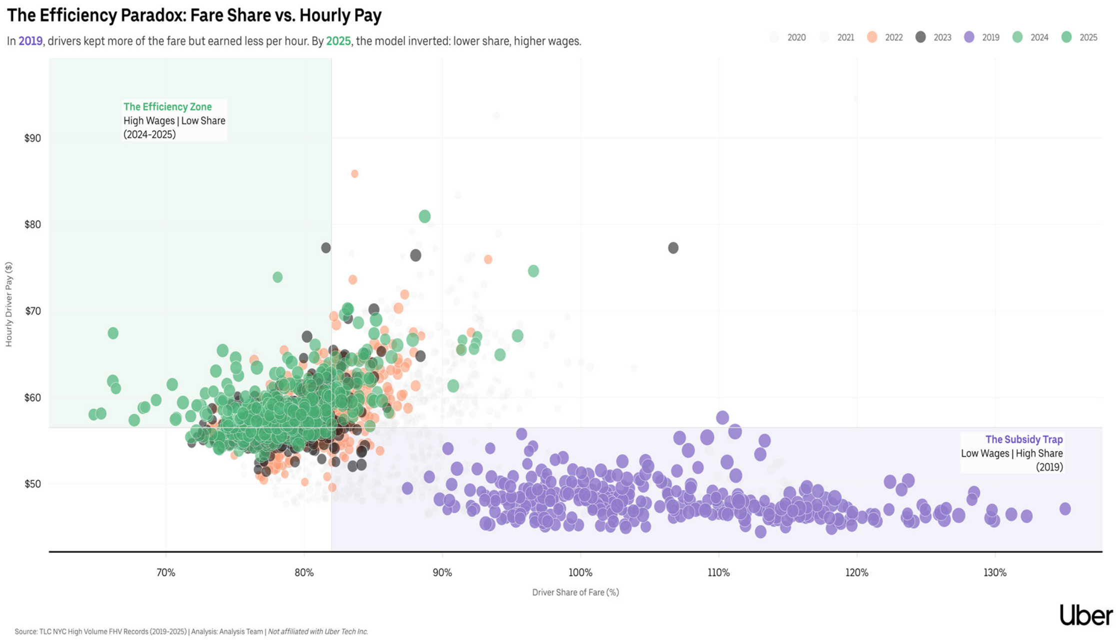


Chart:



2.2. Wage dynamics across subsidy eras

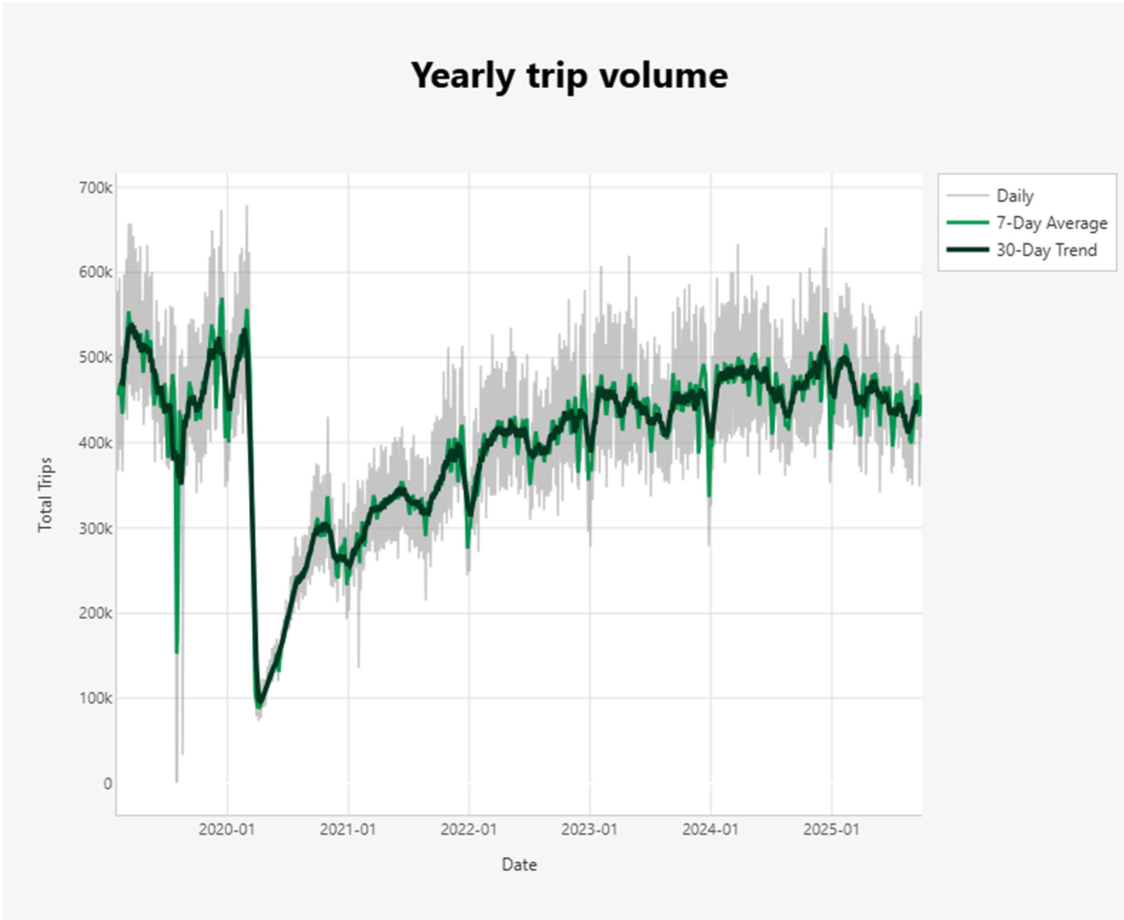
Chart:



PROBLEM 2

1. OPERATIONAL MECHANISMS AND MACROECONOMIC TRENDS

1.1. Overall Analysis



Strategic Narrative Role:

This chart defines the long-term demand storyline for NYC mobility. It does more than display daily volume—it illustrates how the system collapsed, recovered, and eventually stabilized across several years:

- By layering raw daily data with 7-day and 30-day smoothing curves, the chart clearly separates noise (high-frequency daily swings) from structure (macro-level trajectory).

- This hierarchy enables viewers to pinpoint when the system was healthy, when it was disrupted, and how recovery unfolded.

- The visual establishes the overarching timeline that frames all subsequent insights: airport surge patterns, peak-hour efficiency, seasonal variability, and macro mobility cycles.

- It anchors the narrative in a clear pre-COVID → COVID collapse → the collapse at the beginning of each ear → the recovery of uber .

Chart Selection Rationale:

Daily Line (Raw):

- A continuous line chart is necessary for day-by-day data because it preserves temporal flow.

- Bar or area charts would exaggerate magnitude or visually clutter the series, obscuring trend interpretation.

7-Day Moving Average:

- Added to smooth weekday–weekend oscillation while still reflecting short-term behavioral shifts.

- Monthly aggregation was rejected because it would hide reopening shocks and event-driven fluctuations.

30-Day Trend Line:

- Provides the structural backbone of the visualization, allowing multi-year recovery patterns to emerge.

- Regression trendlines and LOESS curves were avoided to prevent oversmoothing or introducing artificial shapes.

Together, these three layers allow the viewer to understand immediate volatility, short-term rhythm, and long-term trajectory within a single coherent view—something no alternative chart type can deliver without sacrificing temporal clarity.

Applied Visualization Mechanics:

Clutter Reduction:

- Unnecessary gridlines, borders, and heavy axis ticks were removed to maximize data-ink efficiency

- Only minimal reference scaffolding remains, ensuring the trend lines dominate visual attention.

Preattentive Attributes:

- A deliberate color hierarchy guides interpretation:

- Dark green → 30-day trend (primary structural signal)

- Medium green → 7-day average (secondary smoothing layer)

- Soft grey → daily volatility (background context)

- This pushes high-frequency noise backward and pulls the macro-structure forward instantly.

Layer Alignment:

- The 30-day line serves as the visual anchor at the top of the hierarchy.

- The 7-day average sits between the trend and raw data, forming a clean three-tier information stack.

- This alignment guides the viewer's eye from structure → pattern → noise, mirroring the intended analytical workflow and supporting intuitive interpretation.

1.2. Demand analysis

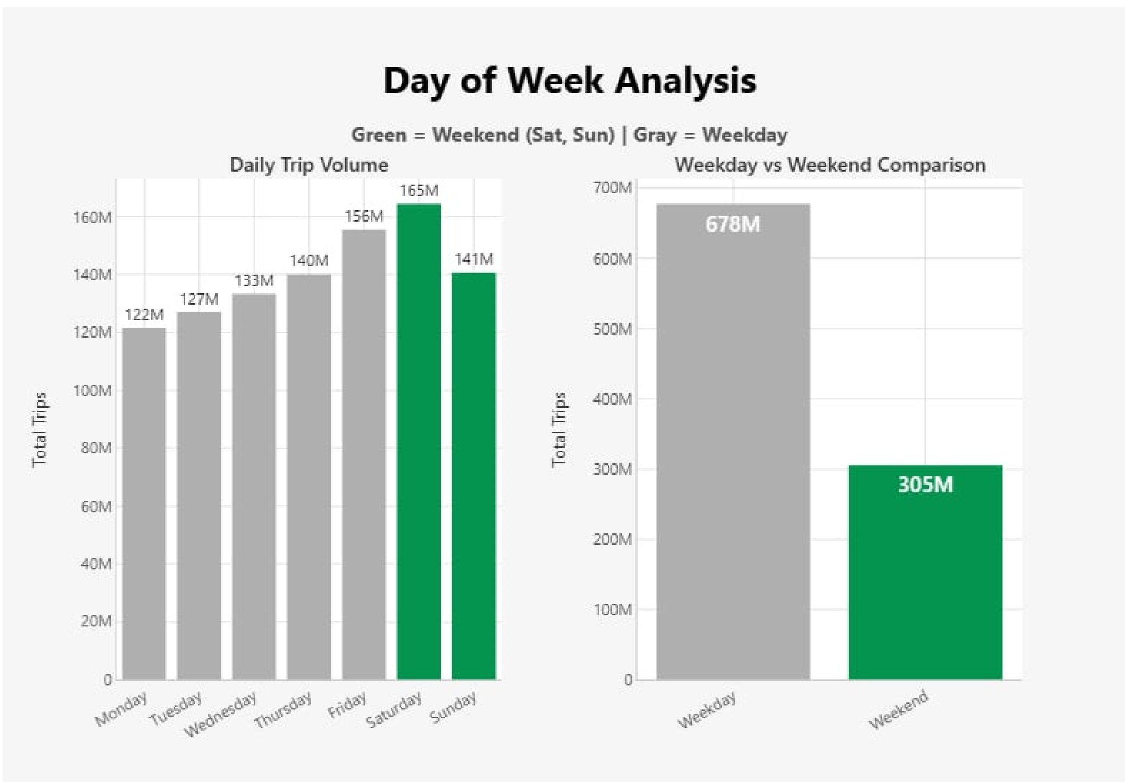


Chart Type: Dual Bar Comparison (Daily Bars + Aggregated Bars)

Strategic Narrative Role:

This visual defines the weekly Demand Pattern Framework. It reveals how rider activity splits between commuter-driven weekdays and leisure-driven weekends.

- The left panel shows that weekdays maintain consistent but moderate volumes (122M–156M), while weekends—especially Saturday at 165M—produce sharp behavioral spikes.

- The right panel condenses the pattern into a single contrast: 678M weekday trips vs. 305M weekend trips, making the structural difference unmistakable.

- The narrative takeaway is that weekday demand is routine and stable, whereas weekend demand is episodic and event-driven, requiring a behavioral rather than scale-based operational strategy.

Chart Selection Rationale:

Daily Bar Chart (Left):

- Bars were selected because weekday and weekend categories are discrete and non-continuous.

- A line chart would falsely imply temporal continuity, and a pie chart would distort magnitude comparison.

- Vertical bars support clean, intuitive comparison across all seven days.

Binary Bar Chart (Right):

- A simple two-bar comparison communicates the structural gap between weekday and weekend totals with maximum clarity.

- Stacked bars or tables would dilute the contrast and hide the binary relationship.

- This format foregrounds the key strategic point: weekday volume is more than double weekend volume, even though weekends include the highest individual-day peak.

Applied Visualization Mechanics:

- Color Encoding: Weekend trips use UBER_GREEN to instantly separate leisure-driven behavior from weekday commuter patterns shown in neutral gray, leveraging preattentive processing to direct attention to the behavioral segment.

- Clutter Reduction: Non-essential gridlines, heavy borders, and redundant tick labels were removed to keep focus on bar heights and numeric labels rather than chart scaffolding, increasing the data-ink ratio and improving readability.

- Label Emphasis: Large, centered labels like 165M and 678M provide immediate numerical comprehension without requiring axis reference, reducing cognitive load and supporting quicker decision-making.

- Structured Layout: Daily bars on the left reveal micro-behavior across the week, while aggregated bars on the right deliver the macro-summary. This left-to-right structure mirrors analytical reasoning—from granular observation to high-level conclusion—and ensures smooth narrative flow.

1.3. Demand Heatmap Analysis

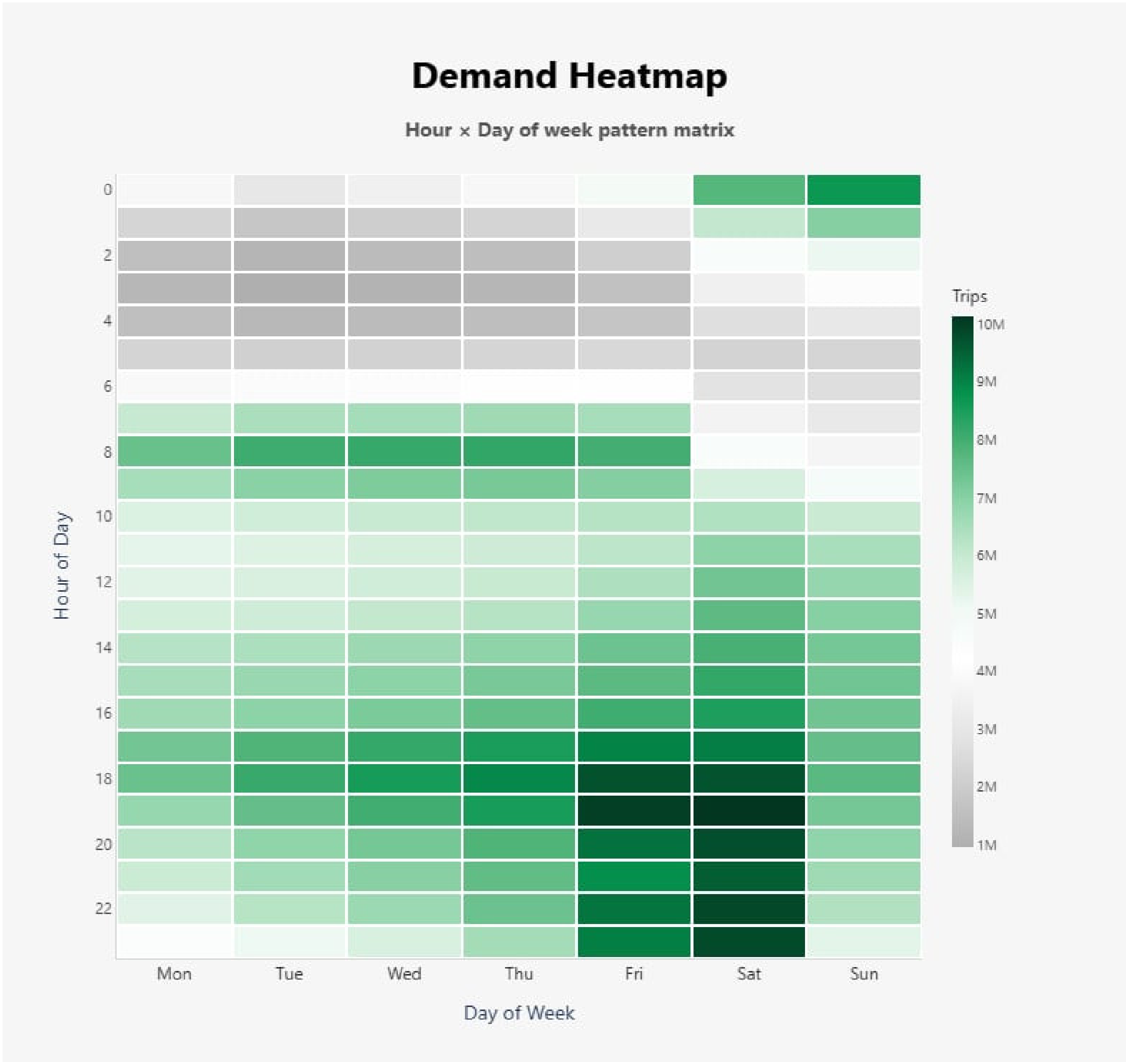


Chart Type: Day × Hour Demand Heatmap

Strategic Narrative Role:

The heatmap reinforces the report's central message: Uber demand varies meaningfully across both hours and days, and weekend evenings consistently emerge as the highest-value operating window.

- By displaying all 168 hour–day combinations, it establishes a complete behavioral map of rider activity.

- This comprehensive layout allows the viewer to instantly identify profitable demand clusters and understand the system's natural demand cycles.

Chart Selection Rationale:

A heatmap is the most effective chart type for comparing two categorical dimensions simultaneously (day × hour).

- Bar or line charts would either oversimplify the pattern or distort the temporal structure by forcing sequential interpretation.

- Heatmaps allow high-demand regions to stand out immediately, making the Friday/Saturday evening "hot zone" visually dominant.

- This format directly supports the communication objective: identify *when* demand peaks, not just *how much* it peaks.

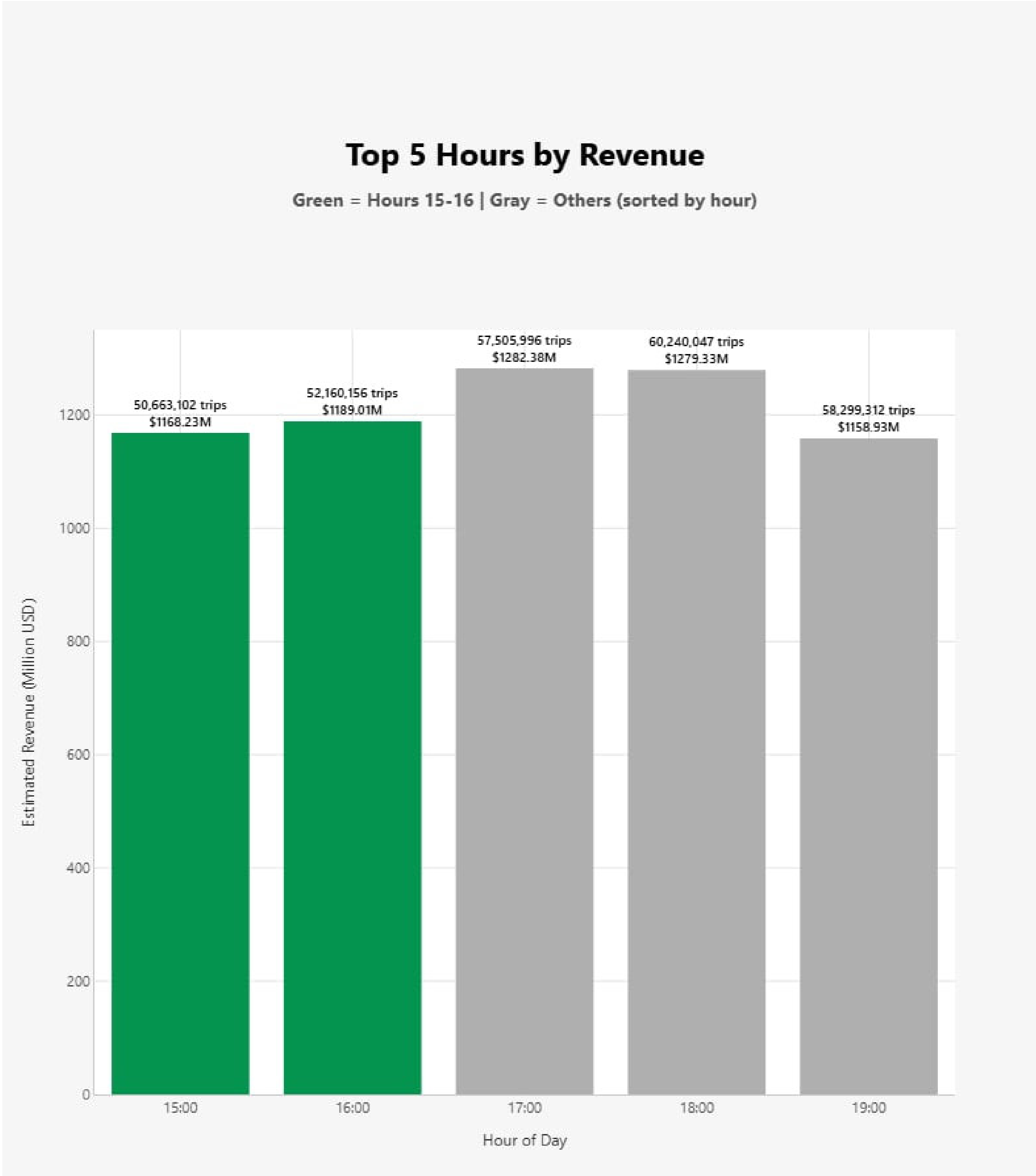
Applied Visualization Mechanics:

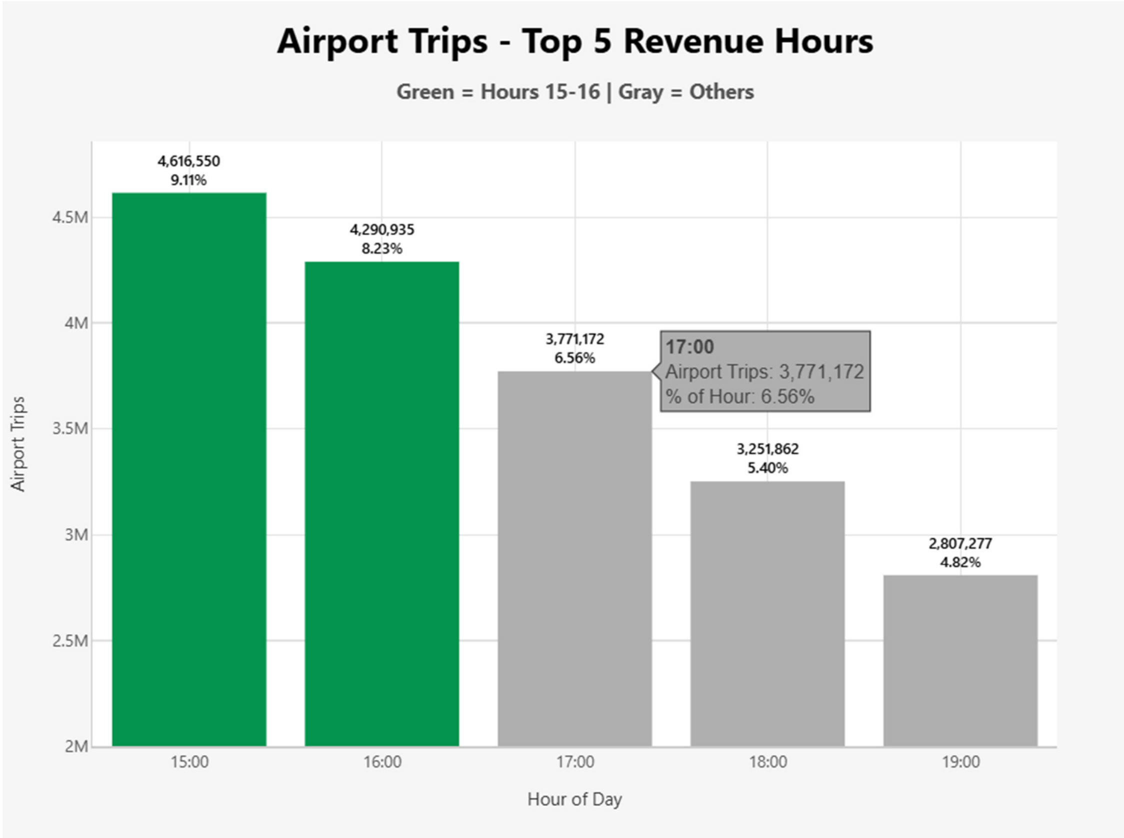
- Color Encoding (preattentive): Darker green tones automatically pull the viewer's attention to the highest-demand cells before conscious analysis begins, enabling rapid pattern detection.

- Clutter Reduction: Heavy gridlines, dark borders, and dense cell labels were removed to lower cognitive load and keep visual emphasis on the demand gradient rather than chart scaffolding.

- Intentional Layout: Days run left-to-right (Mon → Sun) and hours top-to-bottom (0h → 23h), naturally placing the peak demand cluster in the bottom-right corner—where viewers' gaze typically concludes—enhancing emphasis on the weekend evening block.

1.4. Revenue Analysis


Top 5 Hours by Revenue
Green = Hours 15-16 | Gray = Others (sorted by hour)
50,663,102 trips
$1168.23M
52,160,156 trips
$1189.01M
57,505,996 trips
$1282.38M
60,240,047 trips
$1279.33M
58,299,312 trips
$1158.93M
Estimated Revenue (Million USD)
0
200
400
600
800
1000
1200
15:00
16:00
17:00
18:00
19:00
Hour of Day



(Chart Type: Dual Bar Comparison + Airport-Mix Contrast)

Strategic Narrative Role:

This visual defines the Revenue Driver Interpretation Framework. It moves beyond simply displaying which hours have the highest total trips (Scale) to revealing why certain hours earn peak-level revenue despite lower trip volume (Behavior). This distinction makes the strategic argument explicit: revenue at 15:00 and 16:00 is inflated not by high ride count but by airport-heavy, high-value trip composition. The chart proves that a "volume-based assumption" cannot fully explain revenue patterns—trip quality (airport share), not quantity, drives revenue parity with commuter peaks.

Chart Selection Rationale:

Bar Chart (Up – Revenue vs Trips):

- We selected a grouped bar chart to compare trip volume and revenue magnitude across discrete hours.
- A line chart would falsely imply temporal continuity, and a pie chart would distort relative height comparisons.
- Bars produce a clean, categorical scan that makes the "low volume, high revenue" anomaly immediately visible.

Bar Chart (Down – Airport Share by Hour):

This format creates a clear side-by-side comparison of airport-trip proportions.

- A grouped bar or table would introduce clutter and make proportional gaps harder to read.
- Showing airport share as standalone bars makes the causal explanation (airport mix drives revenue) instantly interpretable.

Applied Visualization Mechanics:

Clutter Reduction:

- Gridlines, borders, and redundant axis markings were removed to maximize the Data-Ink Ratio.
- Only essential scaffolding remains, ensuring viewers focus on the contrast between bars rather than chart decoration.

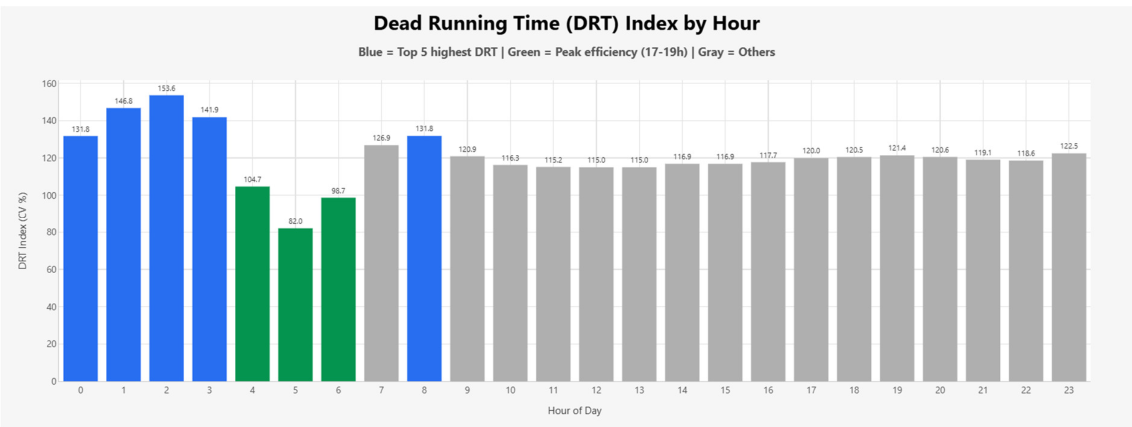
Preattentive Attributes:

- UBER_GREEN is used to highlight the key hours (15:00 and 16:00) that carry the strategic narrative.
- Neutral grey is applied to surrounding hours to push them into the perceptual background.
- This hierarchy guides the viewer's eye directly toward the behavioral insight: airport-heavy hours punch above their volume.

Alignment:

- Hours are left-aligned and ordered consistently so the viewer's gaze flows naturally from Scale (revenue/volume) to Behavior (airport mix).
- Green bars appear first in each chart region, anchoring visual focus and making the causal relationship visually linear: High Airport Share → High Revenue Despite Lower Trips.

1.5. Demand Recovery Time Index (DRT)



(Chart Type: Hourly Bar Comparison + CV Overlay)

Strategic Narrative Role:

This visual defines the Driver Efficiency Interpretation Framework. It clarifies not just *how long* drivers wait (DRT), but *why* certain hours create more or less idle time by pairing DRT with its statistical driver, CV (volatility).

- It shows that unstable early-morning demand produces extremely high DRT.
- It shows that midday hours stabilize but remain operationally inefficient.
- It highlights 17–19h as the most efficient period of the entire day, where drivers experience minimal—or near-zero—idle time.
  This visual grounds later business recommendations by demonstrating that the 17–19h block is the true efficiency peak for driver-side operations.

Chart Selection Rationale:

A bar chart is selected because the comparison spans 24 discrete hourly categories, requiring a format that supports clean magnitude comparison rather than trends. The viewer must quickly identify:

- which hours produce the highest DRT,
- which hours produce the lowest DRT,
- how these clusters relate to overall system behavior.

A line chart would falsely imply temporal continuity and distort the hour-by-hour categorical nature of DRT. A heatmap or dot plot would increase visual overhead without improving interpretability. Bars also allow direct placement of numeric DRT and CV labels, enabling instant dual-variable interpretation.

The bar chart is therefore the most efficient medium for revealing both peak inefficiency and peak efficiency at a glance.

Applied Visualization Mechanics:

Color Encoding:

- Blue marks the worst-performing hours with the highest DRT.
- Green highlights the target window (17–19h) with the strongest operational efficiency.
- Grey establishes a neutral baseline for remaining hours.
  This color hierarchy acts as a preattentive cue, enabling viewers to understand the operational structure before reading any text.

Clutter Reduction:

- Gridlines are minimized to avoid visual noise.
- Heavy borders and redundant ticks are removed to elevate bar contrast.
- Direct labeling for both DRT and CV eliminates dependency on axes.

Layout Structuring:

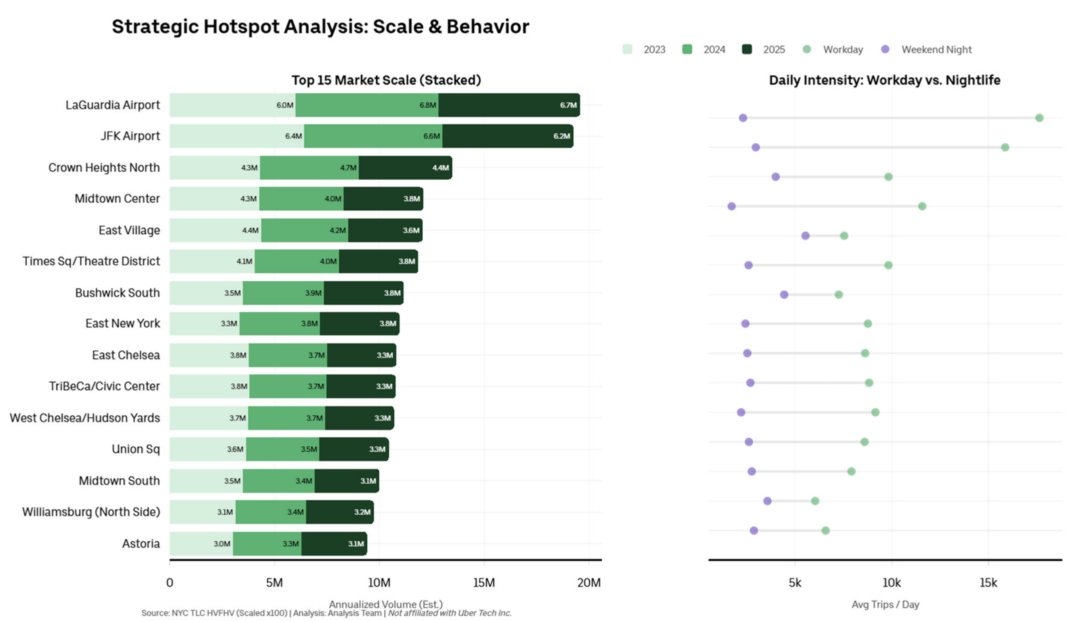
- Bars are spaced widely to keep each hour visually distinct.
- Tall bars for early-morning hours and short bars for evening hours create unmistakable visual contrast.
- Placing the green 17–19h bars toward the right visually frames the efficiency dip as the natural focal point of the chart.

Together, these mechanics ensure the viewer instantly understands:

- where inefficiency occurs,
- where the system performs best,
- and why 17–19h is emphasized as the operational sweet spot.

2. CURRENT MARKET ANALYSIS

2.1. Growth Hotspots



(Chart Type: Combined Stacked Bar + Dumbbell Plot)

Strategic Narrative Role: This graph defines the Supply Allocation Strategy. It moves beyond simply showing where the market volume is (Scale) to revealing when these markets are active (Behavior). This distinction provides evidence that a "one-size-fits-all" strategy cannot apply to both Airports (24/7 inelastic demand) and Leisure Hubs (Nightlife-driven).

Chart Selection Rationale:

- Stacked Bar (Left): We selected a Stacked Bar to visualize the Total Market Size over time. This allows for a clearer comparison of absolute magnitude between zones compared to a Pie Chart, which often makes comparing areas difficult.

- Dumbbell Plot (Right): This is the optimal choice for visualizing the Behavioral Gap. It highlights the disparity between "Workday" and "Weekend Night" demand on a single line. A Grouped Bar Chart here would create unnecessary clutter and make the gap harder to scan.

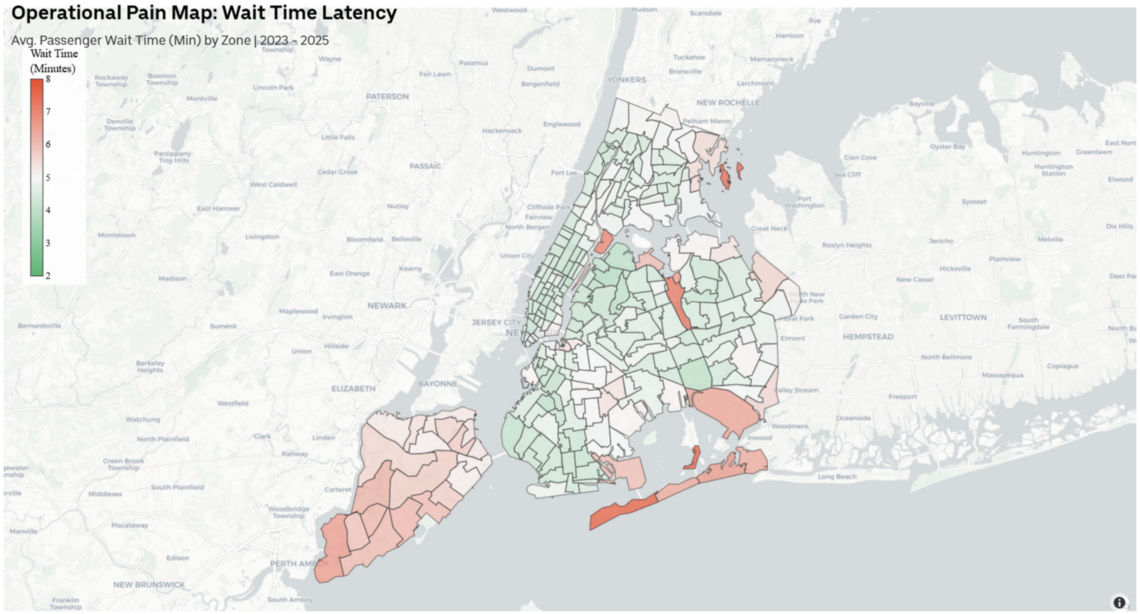
Applied Visualization Mechanics:

- Clutter Reduction: We stripped away non-essential vertical gridlines and chart borders to maximize the Data-Ink Ratio.

- Preattentive Attributes: We utilized UBER_GREEN (Workday) and UBER_PURPLE (Nightlife) to create immediate contrast. Less critical elements were pushed to the background using grey.

- Alignment: Zone names are left-aligned and centered, acting as a visual anchor that allows the eye to scan naturally from Scale (left) to Behavior (right)

2.2. Wait Time Analysis



(Chart Type: Choropleth Mapbox - Simplified Geometry)

Strategic Narrative Role:
This serves as a diagnostic map of Network Health. It identifies "High Friction Zones" where customers experience excessive wait times, signaling a high risk of churn to competitors or alternative transport modes.

Chart Selection Rationale:

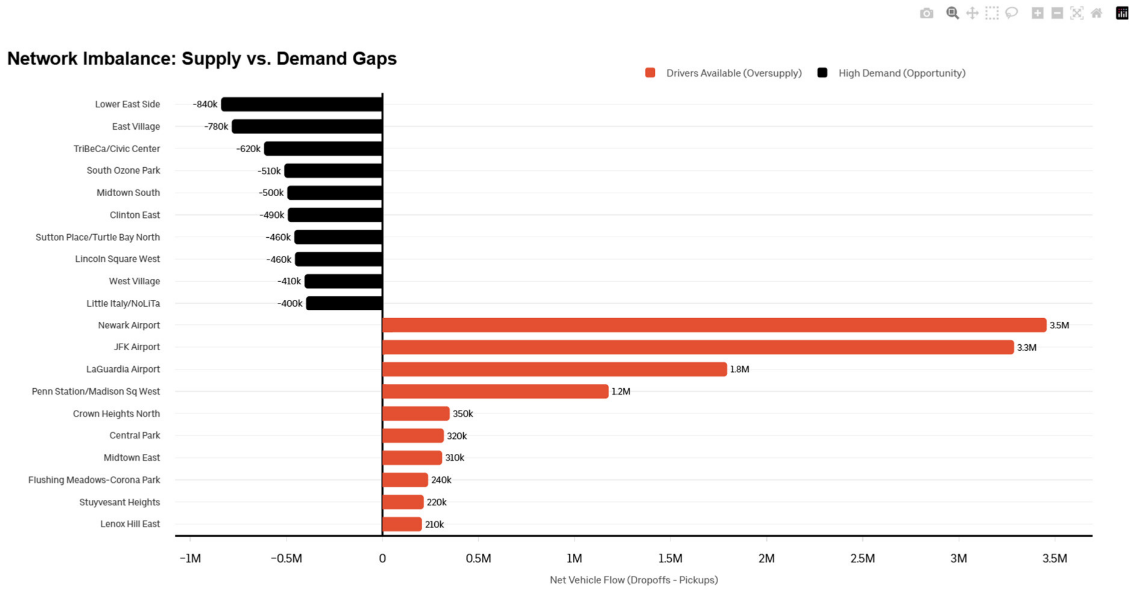
Choropleth Map: Because the data is geospatial, a map allows us to identify clustering of operational bottlenecks (e.g., the contiguous red zones around Airports or Staten Island). A table would fail to convey these spatial relationships effectively.

Applied Visualization Mechanics:

- Simplification: We intentionally simplified the shapefile geometry. Street-level detail is "noise"; the overall zone shape is the "signal." This reduces cognitive load for the executive audience

- Diverging Color Scale: We employed a diverging scale: UBER_GREEN (Fast) to White (Average) to UBER_RED (Slow). This leverages Preattentive Processing, forcing the eye to ignore the average (white) data and focus immediately on the problematic red outliers

2.3. Dead Mileage



(Chart Type: Diverging Bar Chart / Butterfly Chart)

Strategic Narrative Role: This graph exposes the greatest source of operational waste: Dead Mileage. It highlights the structural imbalance between where trips end (Drop-off) and where new demand originates (Pickup).

Chart Selection Rationale:

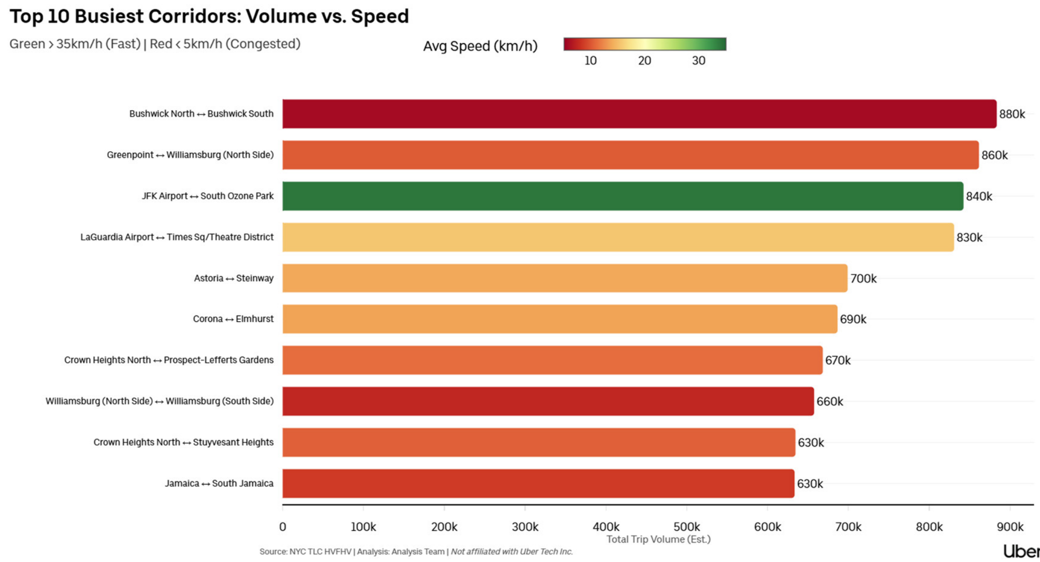
Diverging Bar Chart: This is the superior choice for visualizing Net Flow. The central axis represents equilibrium (0).

- Bars to the left (negative) visualize supply deficits.
- Bars to the right (positive) visualize supply surpluses.
- This structure is far more intuitive than dual bar charts for comparing offsets.

Applied Visualization Mechanics:

- Strategic Coloring: We used UBER_BLACK for "High Demand" (Opportunity—requiring focus) and UBER_RED for "Oversupply" (Risk—requiring alert).
- Labeling: Data labels are rounded (e.g., 2.5M) and placed directly next to the bars via the Proximity Principle, eliminating the need for the audience to scan back and forth to an axis.

2.4. Traffic Corridors



(Chart Type: Horizontal Bar Chart with Color Gradient)

Strategic Narrative Role:

This graph answers the question: "Are our highest-revenue 'backbone' routes operating efficiently?" It synthesizes two dimensions: Commercial Importance (Volume) and Operational Performance (Speed).

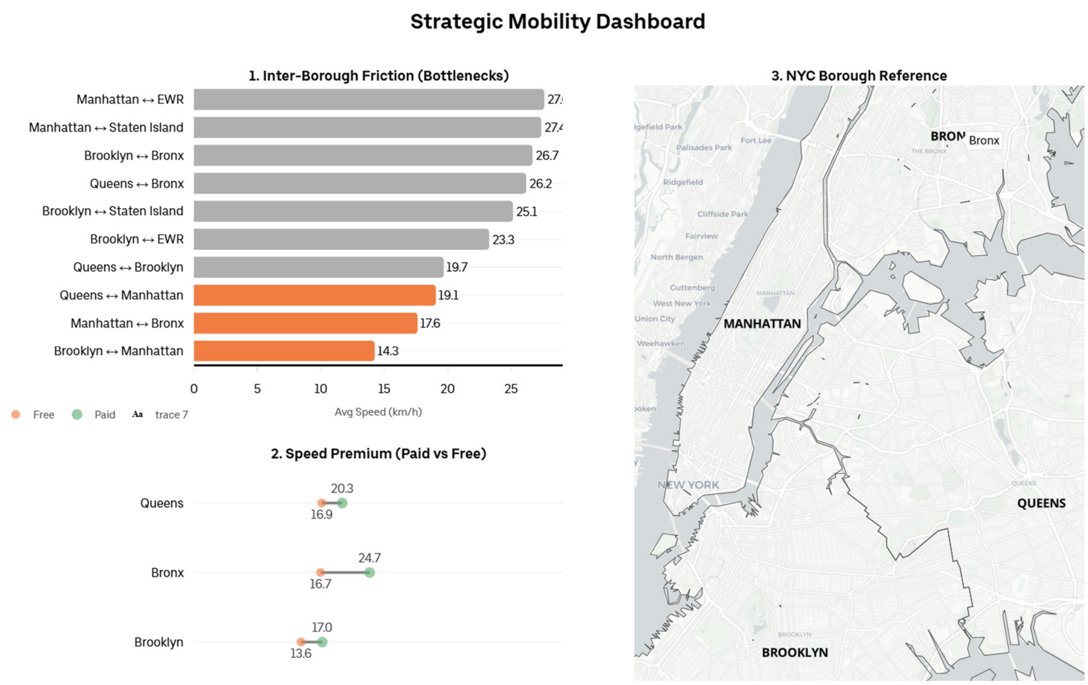
Chart Selection Rationale:

Horizontal Bar: Given the long labels required for corridor names (e.g., "LaGuardia Airport → Times Sq"), a horizontal bar chart is the only viable option to ensure text is legible and not diagonal or truncated.

Applied Visualization Mechanics:

- Visual Hierarchy: The data is sorted descending by Volume, prioritizing business impact
- Color Encoding: Instead of using color for Volume (which is already encoded by bar length), we used color to encode Speed. Dark red bars alert the viewer to high-volume routes that are suffering from severe congestion, effectively telling a multidimensional story in 2D space.

2.5. Border Friction



Strategic Narrative Role

Objective: This dashboard provides empirical evidence of Infrastructure Constraints. It shifts the strategic focus from attributing delays to "driver shortages" to acknowledging the physical limitations of the NYC transport network.

Narrative Flow:

- Problem Identification: Identifies inter-borough traffic flows (e.g., Brooklyn → Manhattan) that suffer from the lowest speeds network-wide.

- Solution Assessment: Evaluates whether paid infrastructure (Tolls) delivers a significant speed advantage (Cost-Benefit Analysis).

- Spatial Context: Provides immediate geographic localization of these bottlenecks to support operational decision-making.

Chart Selection Rationale

A. Bar Chart (Top Left): Inter-Borough Friction

- Selection Rationale: A Horizontal Bar Chart is the optimal choice for comparing and ranking categories with Long Category Labels (e.g., "Brooklyn → Manhattan")

- Why Preferred: Vertical bar charts would force text rotation or truncation, reducing legibility and increasing cognitive load for the audience.

B. Dumbbell Plot (Bottom Left): Speed Premium

- Selection Rationale: This is a specialized chart designed to compare the difference between two data points on a single axis (Free Speed vs. Paid Speed). It emphasizes the gap between the two states rather than just displaying absolute values.

- Why Preferred: Using a Grouped Bar Chart would require the audience to compare the heights of adjacent bars, creating visual clutter and making it harder to instantly perceive the differential compared to the clean lines of a Dumbbell plot.

C. Mapbox Map (Right): Borough Reference

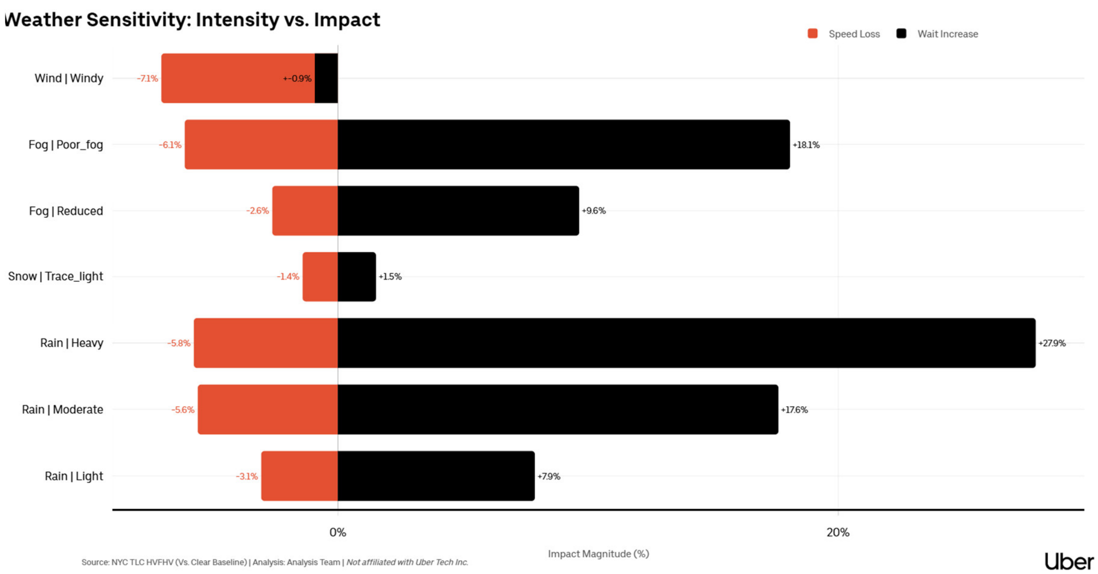
Selection Rationale: The map serves as a contextual anchor. For transportation data, visualizing physical boundaries (rivers, bridges) allows the viewer to immediately link the abstract metrics on the left with the physical terrain on the right.

Applied Visualization Mechanics:

- Clutter Reduction:
    - Gridlines Removed: The code sets showgrid=False (or minimizes visibility). Since values are provided via direct labeling next to bars and data points, gridlines become redundant non-data ink and are removed to improve the Data-Ink Ratio.
    - Geometric Simplification: Using simplify(tolerance=0.001) on the map removes unnecessary street-level detail. This cleans up the visual field and focuses attention on the overall shape of the boroughs (Gestalt Principle of Closure).
- Preattentive Attributes:
    - Strategic Color:
        - Bar Chart: A distinct color (e.g., UBER_ORANGE or Bold Black) is used exclusively for the Top 3 slowest routes, while the remaining routes are rendered in gray (GRAY_500). This immediately directs the eye to the most critical issues
        - Dumbbell Plot: UBER_GREEN is used for "Paid" (Positive/Faster) and UBER_ORANGE for "Free," maintaining consistent color semantics throughout the report.
    - Strategic Contrast: We intentionally bolded and colored only the "bottom 3" bars (representing the slowest/worst performing routes) while rendering the rest in muted gray. This leverages the Preattentive Attribute of Intensity/Color to create a sharp visual contrast.

Visual Hierarchy: Sorting the Bar Chart by speed creates a logical flow, allowing the audience to easily scan and process the information in an ordered manner

2.6. Weather Impact



(Chart Type: Tornado Chart / Paired Bar)

Strategic Narrative Role:

This graph decodes network breakage during adverse weather. It distinguishes between Safety Risks (Speed Loss) and Supply Risks (Wait Time Increase).

Chart Selection Rationale:

Tornado Chart: A variation of the bar chart that allows for the comparison of two distinct metrics with different units (% Speed Decrease vs. % Wait Increase) against a single categorical axis (Weather Conditions). This facilitates an immediate comparison of the "Double Hazard."

Applied Visualization Mechanics:

- Logical Ordering: Categories are sorted by weather severity (Light Rain → Heavy Rain → Storm), creating a linear narrative structure

- Color Semantics: We used UBER_RED for "Speed Loss" (Safety Hazard/Negative) and UBER_BLACK for "Wait Increase" (Operational Issue), utilizing color to convey meaning