

Fulfillment performance and bottleneck analysis

Olist Brazilian E-commerce dataset (2016-2018)

Executive summary

Fulfillment efficiency represents one of the most critical determinants of customers satisfaction, operational effectiveness, and delivery reliability within modern e-commerce ecosystems. As digital marketplaces scale the ability to diagnose performance constraints and identify systemic inefficiencies becomes increasingly essential for maintaining competitive service levels.

This analysis investigates fulfillment dynamics within the Olist Brazilian marketplace dataset (2016-2018) with the objective of diagnosing delivery delays, identifying operational bottlenecks, and evaluating the primary drivers influencing late deliveries. Rather than focusing on isolated transactional events, the study adopts a systems-level perspective, examining how performance variability propagates through the fulfillment pipeline.

Through the application of SQL-engineered operational metrics and interactive Tableau visual analytics, the analysis reveals a highly asymmetric fulfillment structure in which last-mile delivery duration emerges as the dominant contributor to total cycle time. Furthermore, downstream delivery dynamics demonstrate a direct and measurable relationship with delay severity, confirming last-mile logistics as the primary driver of late fulfillment outcomes. Contrary to common assumptions, order complexity exhibits limited explanatory power in predicting delays, while aggregate delay behavior indicates a pronounced early-fulfillment bias across the network.

Collectively, these findings highlight the central role of downstream logistics variability in shaping fulfillment performance and underscore the disproportionate leverage associated with last-mile optimization strategies.

Business Context

Marketplace-based e-commerce platforms operate through a sequence of interconnected fulfillment stages, each contributing to the overall delivery experience. These stages typically include order approval workflows, seller processing operations, carrier handoff logistics, and last-mile delivery networks. Performance disruption at any stage has the potential to cascade through the pipeline, ultimately affecting delivery timeliness and customer satisfaction.

Delivery delays within such systems may originate from upstream administrative processes, internal operational inefficiencies, or downstream logistics variability. Identifying the locus of delay generation is therefore fundamental to designing effective intervention strategies.

The core business question: Which stage of the fulfillment pipeline contributes most significantly to delivery delays? Addressing this question requires moving beyond surface-level metrics and instead examining the structural dynamics governing fulfillment duration and variability.

Data engineering & metric design

To enable stage-level diagnostics, raw transactional records were transformed using PostgreSQL to construct analytical features representing fulfillment performance and delay behavior. This transformation layer converts event timestamps into interpretable operational metrics, allowing for systemic performance evaluation.

Key engineered metrics include approval time, seller processing duration, last-mile delivery time, total fulfillment cycle time, delivery delay (days), and late delivery flags. Additional features capturing order complexity and financial attributes provide further contractual dimensions for analysis.

By engineering metrics that reflect stage-specific durations, the analysis shifts from descriptive reporting towards diagnostic modeling, identification of bottleneck structures and delay propagation mechanisms.

Analytical Findings

Last-mile Delivery Dominates fulfillment Duration

Decomposition of the fulfillment cycle time reveals a highly asymmetric distribution of operational duration across pipeline stages. Comparative analysis of approval, processing, and delivery intervals indicates that last mile delivery accounts for the largest share of total fulfillment duration.

Analytical interpretation

This structural imbalance suggests that systemic fulfillment expansion is primarily driven by downstream logistics rather than internal approval or seller processing constraints. Upstream stages exhibit relatively stable and constrained durations, while delivery intervals introduce greater temporal variability.

Operational insight

From an operational strategy perspective, this finding indicates that optimization efforts targeting approval workflows or seller processing operations would likely yield marginal performance

improvements relative to interventions addressing delivery network efficiency. The dominant leverage point within the system resides within last-mile logistics dynamics.

Last-mile Duration directly drive delivery delays

Analysis of delay severity in relation to delivery duration reveals a strong positive relationship between last-mile transit time and delivery delay magnitude.

Graph interpretation – Last mile Duration vs delivery delay

The scatterplot visualizes individual orders across last-mile duration and capped delivery delay. The observed upward-sloping distribution indicates that extended last-mile durations are systematically associated with larger delivery delays.



Analytical interpretation

This relationship provides direct evidence that delays scale proportionally with downstream transit duration. The widening variability band observed at higher delivery durations indicates increasing volatility as transit time expands, reflecting nonlinear risk accumulation within extended logistics networks.

Operational insight

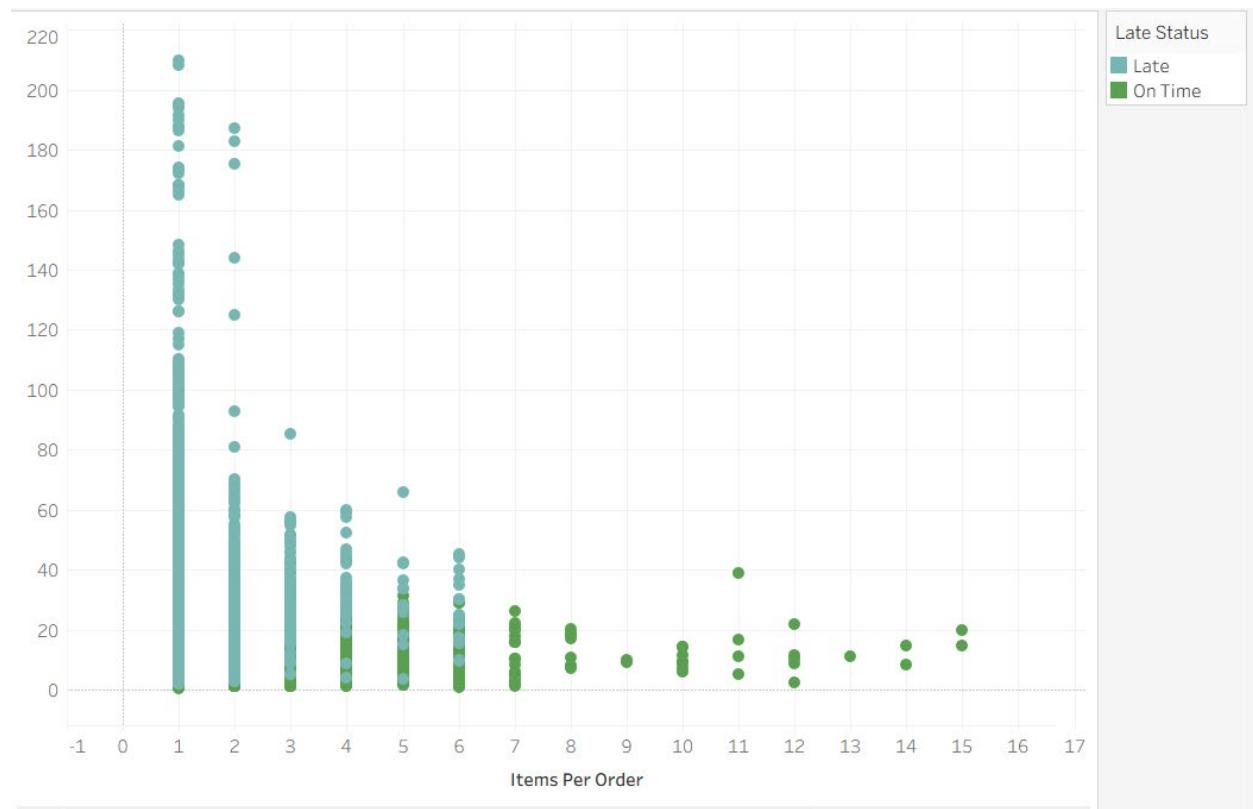
Rather than delays emerging randomly across fulfillment stages, lateness appears structurally linked to last-mile dynamics. This confirms last-mile delivery as the dominant delay driver and validates the bottleneck hypothesis established in the cycle time decomposition analysis.

Order Complexity is not the primary Delay Driver

Contrary to intuitive expectations, increased order complexity does not consistently result in longer fulfillment cycle times.

Graph Interpretation – Order Complexity vs Fulfillment time

The complexity-performance visualization indicates limited correlation between items per order and total cycle time. Fulfillment variability appears distributed independently of order size.



Analytical interpretation

While higher-complexity orders require additional coordination, the absence of a strong upward performance gradient suggests that operational inefficiencies are not strictly driven by item count. Variability concentration among low-complexity orders further challenges conventional assumptions regarding complexity-driven delays.

Operational insight

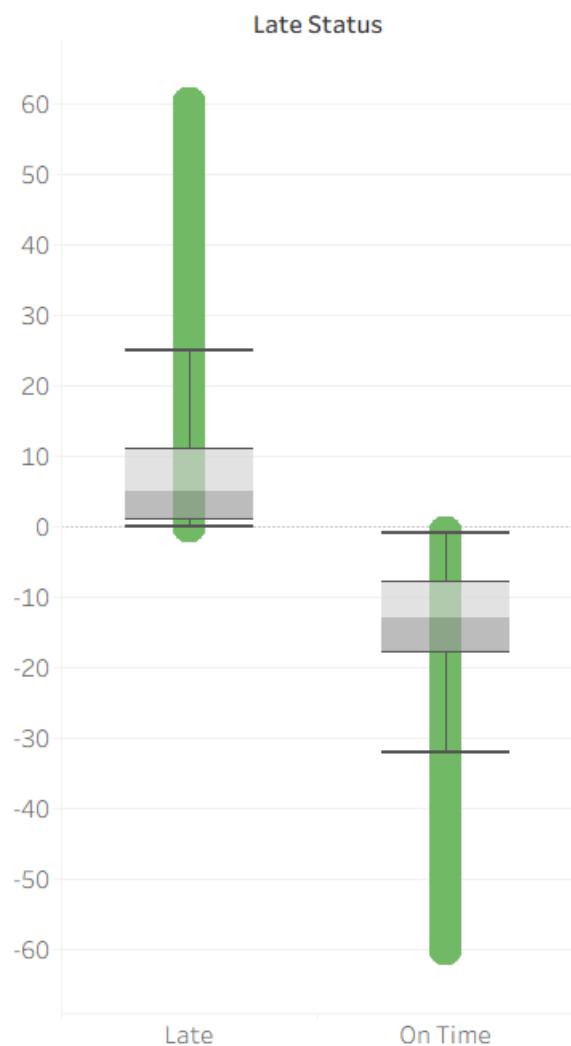
This finding implies that systemic delays are more likely attributable to downstream logistics variability rather than internal handling complexity. From a business intelligence perspective, this represents a critical reframing of delay attribution mechanisms.

Delivery Performance Exhibits Early Fulfillment Bias

Aggregate delay metrics indicate that deliveries are completed ahead of estimated delivery dates on average.

Graph interpretation – delivery delay distribution

The delay distribution boxplot illustrates a negatively centered distribution with right-skewed characteristics. While many deliveries occur early or near schedule, a minority of orders exhibit extreme lateness.



Analytical interpretation

This asymmetric distribution reflects a fulfillment network characterized by early-delivery clustering alongside volatile late exceptions. Delay behavior is therefore dominated by tail-risk events rather than uniform performance degradation.

Operational insight

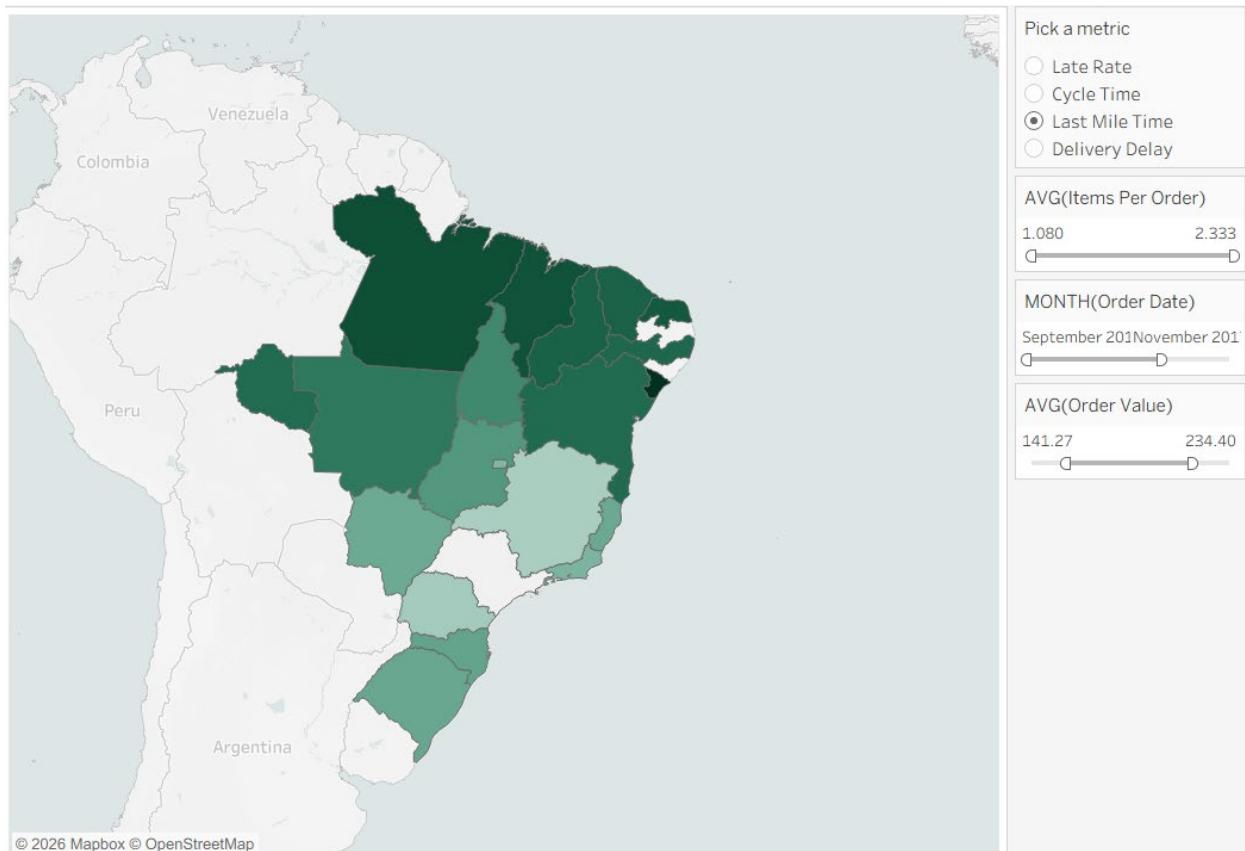
Such patterns are consistent with real-world logistics systems where variability increases with transit duration and performance volatility is driven by downstream distribution.

Geographic Variation in fulfillment performance

Geographic analysis reveals measurable variation in fulfillment performance across customer states.

Graph interpretation – Geographic Fulfillment Performance map

The map visualization highlights spatial variation across performance metrics. Certain regions demonstrate elevated late delivery rates or extended cycle times.



Analytical Interpretation

Spatial heterogeneity suggests that fulfillment performance is influenced by regional logistics network efficiency, carrier infrastructure, and delivery routing dynamics.

Operational insight

Geographic segmentation therefore represents a critical dimension for operational optimization strategies and performance intervention design.

Conclusion & Business Implications

This analysis demonstrates that fulfillment delays within the Olist marketplace dataset are not uniformly distributed across operational stages. Instead, delays emerge from a structurally asymmetric system in which downstream logistics, specifically last-mile delivery dynamics, dominate both cycle time expansion and delay severity.

Key business implications include the disproportionate leverage associated with last-mile optimization strategies, the limited explanatory power of order complexity metrics, and the importance of geographic segmentation in understanding fulfillment variability.

By adopting a systems-level analytical framework, this study illustrates how operational metrics, visual analytics, and bottleneck diagnostics can be combined to generate actionable business intelligence insights.