Insights from Failed Orders - Business Report

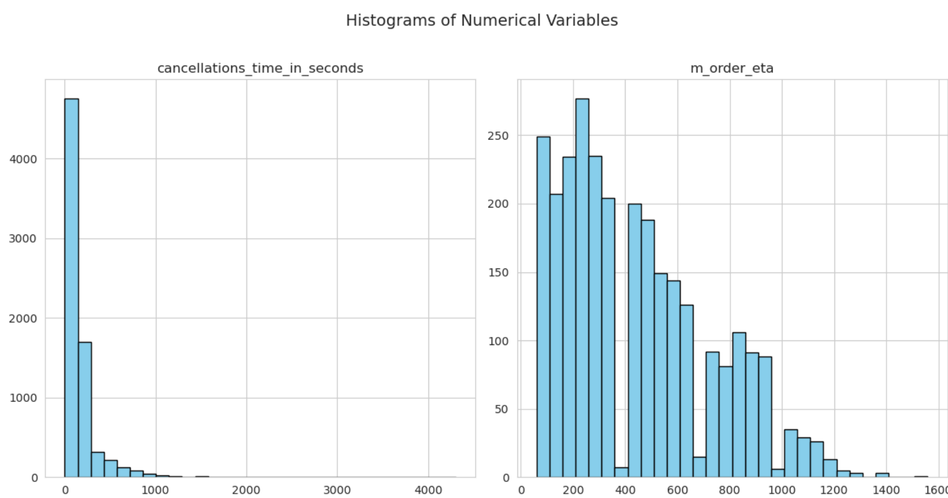
EDA

Missing Values Pattern (Summary)

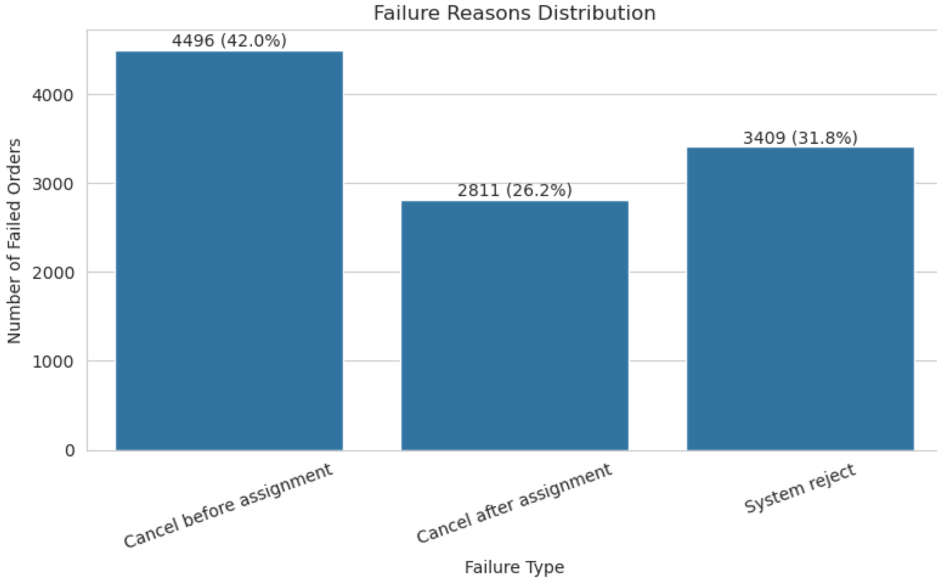
The missing-value pattern reflects the system's business logic:

`m_order_eta` is missing only for system-reject orders (status = 9).
System-cancelled orders never reach driver assignment, so no ETA is generated.

`cancellations_time_in_seconds` is missing only for customer-cancel orders (status = 4).
Cancellation time is recorded only when the customer actively cancels.
Therefore, we draw a conclusion that Missing ETA indicates system cancellation, while Missing cancellation time means customer cancellation. This is expected behavior, not a data-quality issue.



Task 1 — Failure Reasons Distribution



Cancel-before-assignment represents the largest share of failed orders, accounting for 42% of all cancellations, while cancel-after-assignment contributes only 26.2%.

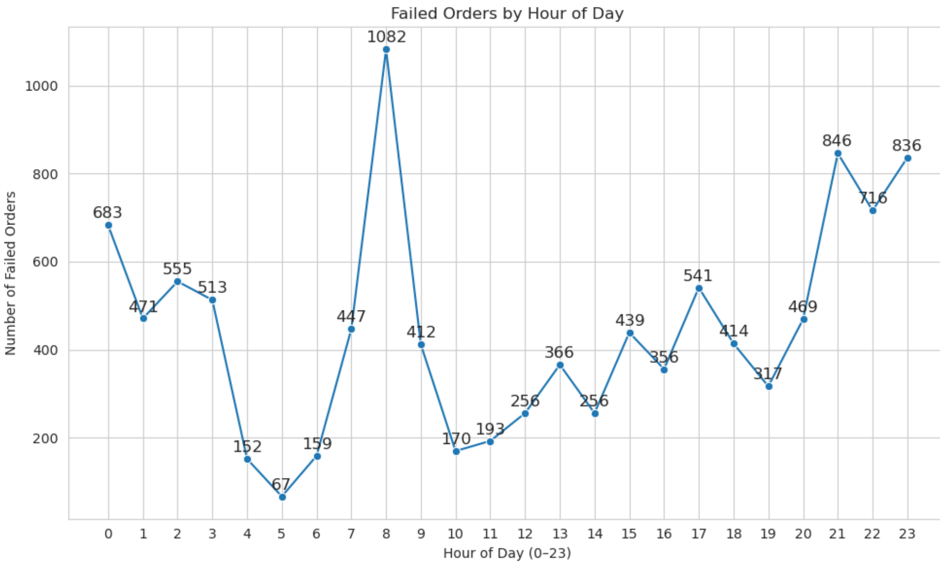
This pattern suggests that many customers cancel very shortly after submitting the order—before the system has time to assign a driver. One possible explanation is that these users realize they need to modify their order details (e.g., pickup location, timing) and therefore cancel immediately.

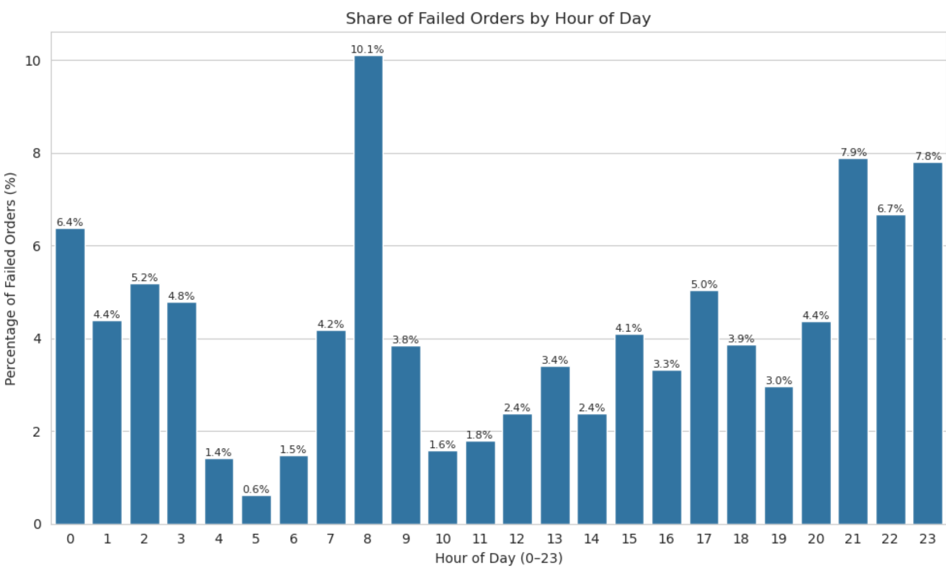
A useful next step would be to investigate whether these customers tend to place a new order shortly after cancelling. If this behavior is common, the product team could consider adding an "Edit Order" feature to reduce unnecessary cancellations and improve user experience.

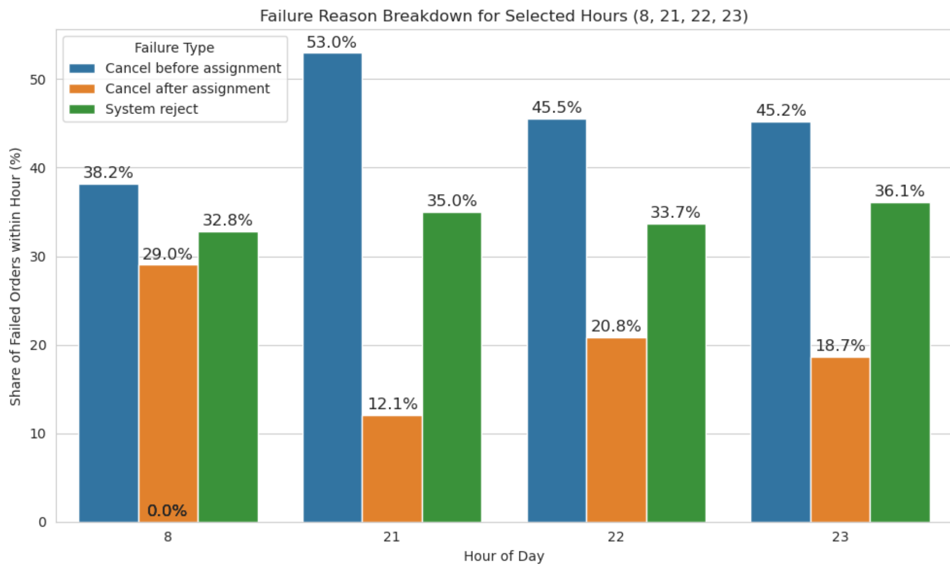
On the other hand, cancellations after driver assignment occur less frequently. This implies that the speed of driver assignment may play a critical role in reducing cancellations.

Enhancing matching efficiency—such as improving driver availability in high-demand areas or optimizing dispatch algorithms—could reduce the waiting time before assignment and ultimately lower the overall cancellation rate.

Task 2 — Failed Orders by Hour







We observe that the cancellation share at 8 AM is noticeably higher than at other hours. A plausible explanation is that 8 AM corresponds to the morning commute peak, during which users are rushing to work and therefore have lower patience and reduced tolerance for waiting times. Passenger demand is also elevated at this hour, meaning the system may not be able to assign drivers quickly enough.

Moreover, the pressure of not being late leads commuters to make decisions rapidly and switch to alternative transportation options such as metro or buses. As a result, even when a driver has already been assigned, users may still choose to cancel the order if the estimated arrival time seems too long or uncertain.

To address the unusually high cancellation rate observed at 8 AM, we suggest:

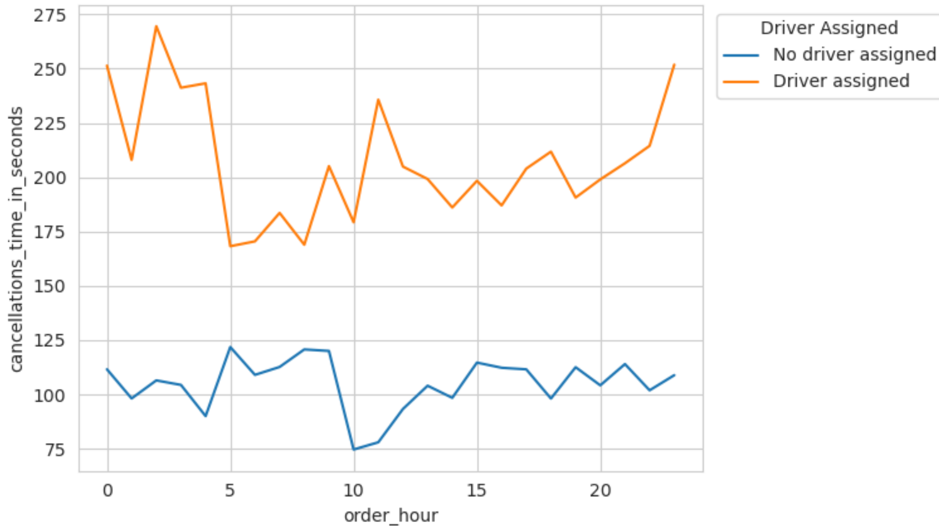
1. Improve ETA accuracy during peak hours.
Because commuters at 8 AM are highly sensitive to even small increases in waiting time, providing a more accurate and reliable ETA estimation can help reduce cancellations. Enhancing the ETA model specifically for peak-hour traffic conditions—by incorporating real-time congestion data, historical morning patterns, and driver density—can lower user uncertainty and prevent users from abandoning their rides after assignment.

2. Introduce peak-hour driver incentives between 7–9 AM.
The elevated cancellation rate at 8 AM is strongly tied to an imbalance between high passenger demand and limited driver availability. Implementing time-specific incentives can increase the number of active drivers during the morning rush, improving assignment speed and reducing wait times. A stronger supply presence during this window directly addresses one of the root causes of early-stage cancellations.

3. Streamline early-stage matching to assign drivers more quickly.
Given the urgency of commuters' travel needs, even minor delays in driver assignment significantly increase the likelihood of cancellation. Optimizing the initial matching logic—such as reducing early filtering constraints or allowing a faster assignment of the nearest available driver—can help ensure that users receive a match before deciding to switch to alternative

transport modes. Faster assignment is especially critical at 8 AM, when user patience is at its lowest.

Task 3 — Average Time to Cancellation (With vs Without Driver Assignment)



Insights from Cancellation Time Analysis

We observe several notable patterns from the cancellation-time behavior:

1. Customers with a driver assigned wait significantly longer before cancelling—typically 3 to 4 minutes, compared to only 1 to 2 minutes for orders without a driver assignment.
   This suggests that once users see a confirmed driver, their commitment increases and they tolerate longer wait times.

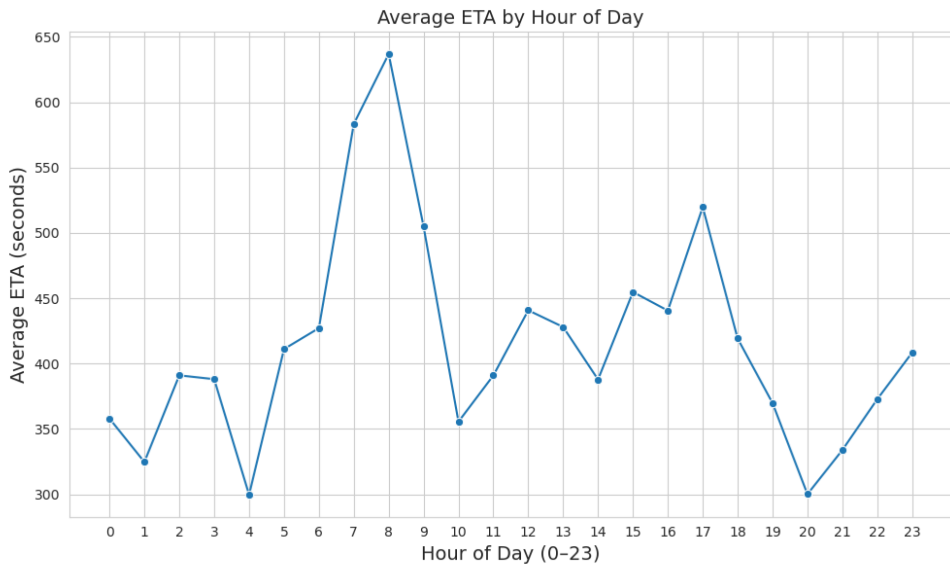
2. During late-night hours (23:00–04:00), customers show even higher willingness to wait when a driver is assigned.
   A likely explanation is the limited availability of alternative transportation options at night, making users more patient even when ETAs fluctuate.

3. Around 10:00 AM, customers without a driver assignment cancel much faster.
   Several factors may contribute to this behavior:
   - Lower urgency: Unlike the morning commute, users at 10 AM may have more flexible schedules.
   - Lower demand density: Requests may come from more dispersed locations, slowing assignment and prompting early cancellation.
   - More alternatives available: Public transportation or walking are more viable in mid-morning.
   - Perceived inefficiency: Users may interpret slow assignment as “no drivers nearby,” leading to quicker cancellations.

Overall Conclusion

Across all hours, customers consistently wait longer when a driver is assigned, whereas orders without driver assignment tend to be cancelled early.
This highlights the importance of accelerating the initial driver-assignment process.
Even if reassignment occurs later, reducing the waiting gap before the first assignment can meaningfully decrease the overall cancellation rate.

Task 4 — Average ETA Distribution by Hour



According to the line chart, ETA is highest around 7–8 AM and again around 5 PM. These peaks correspond to the morning and evening rush hours, during which travel demand increases sharply. As a result, driver supply becomes insufficient, matching takes longer, and customers experience extended waiting times before a driver is assigned.

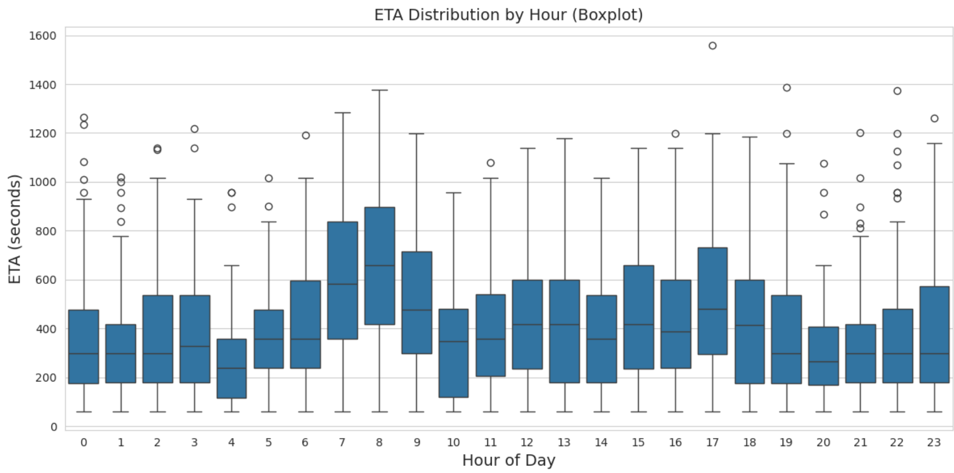
To mitigate long ETAs during these peak periods, several strategies can be considered:

1. Increase driver supply through time-based incentives (e.g., morning and evening peak-hour bonuses) to encourage more drivers to be online when demand surges.

2. Improve demand forecasting to pre-position drivers in high-volume areas before the rush begins.

3. Enhance the matching algorithm to prioritize faster assignments during rush hours, such as relaxing matching constraints or expanding the matching radius temporarily.

4. Refine ETA prediction models to incorporate real-time traffic and congestion patterns, helping set more accurate user expectations and reduce perceived delays.



A key observation from the ETA boxplot is the presence of long whiskers in several hours of the day.
Long whiskers indicate that a portion of orders experienced significantly higher ETA than the typical range, sometimes far beyond the interquartile spread. This suggests considerable variability and instability in driver assignment efficiency during these hours.

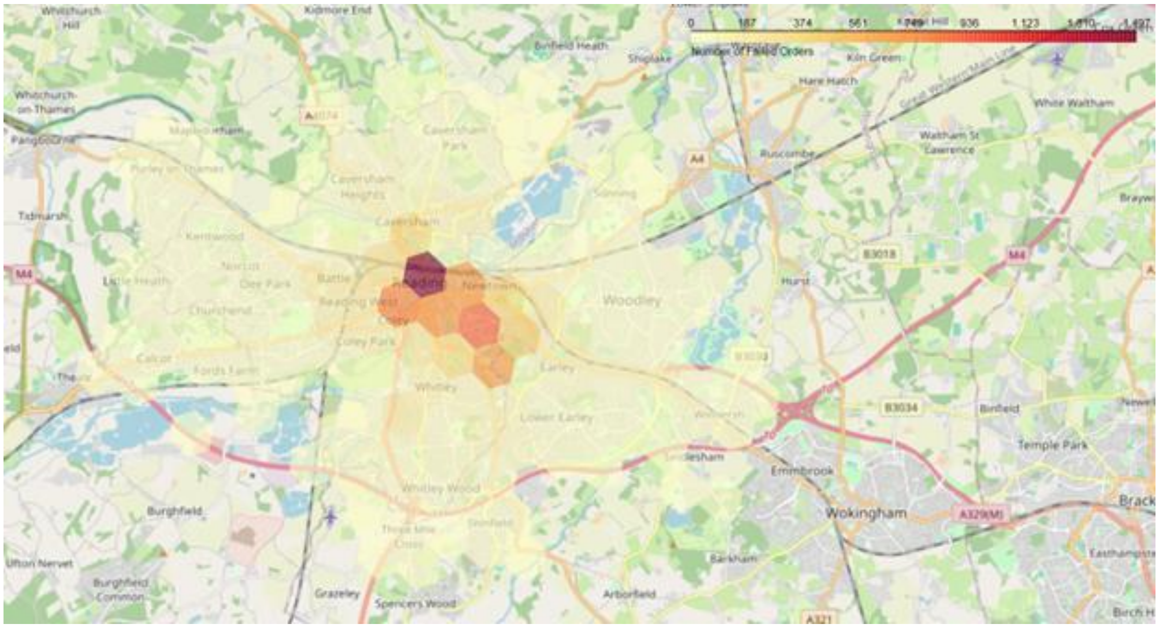
Possible Reasons for the Long-Whisker Pattern

- Supply–demand imbalance: During peak commuting hours or late at night, the number of active drivers may be insufficient to meet demand, forcing the system to assign drivers located much farther away.
- Traffic congestion: Heavy traffic in the morning and evening rush hours increases travel time for drivers to reach passengers, leading to extreme ETA values.
- Sparse nighttime driver distribution: Late-night periods often exhibit low driver density, causing large fluctuations in ETA when the nearest driver is far away.
- Unpredictable assignment delays: Inconsistencies in matching speed may cause some users to wait much longer before a driver is found.

Potential Solutions

- Strengthen driver supply during critical hours
  Implement peak-hour incentives or guaranteed minimum earnings for drivers during 7–9 AM, 17 PM, and late-night periods to reduce extreme ETA outliers.
- Improve pre-positioning using demand forecasting
  Predict high-demand areas and encourage drivers to move to those zones before spikes occur, reducing long-distance matches.
- Enhance ETA prediction models
  Incorporate real-time traffic, road conditions, and driver density to more accurately reflect potential delays and reduce user frustration caused by misaligned expectations.
- Optimize matching logic for high-variance hours
  Allow more flexible matching criteria (e.g., expanded radius) during peak times to avoid extreme wait times for certain passengers.

Task 5 — Hexagon Spatial Analysis



Interpretation

The spatial analysis using H3 hexagons shows that failed orders are highly concentrated in the central Reading area. A small cluster of hexes in the city center accounts for the largest share of failed orders, indicating that this region experiences both high demand and operational pressure during peak periods. This suggests that supply may not be sufficient to meet the concentrated demand, or that matching and ETA performance degrade under heavy load.

Operational Implications

- Increasing driver supply or providing targeted incentives in the city center may reduce the majority of failures.
- Improving matching efficiency or ETA estimation during peak hours in central Reading could significantly reduce cancellations.
- Monitoring peripheral zones with moderate failure intensity could help identify emerging hotspots.