# Gett Failed Orders Analysis - Documentation

## Project Overview

Objective: Investigate why certain Gett orders failed by analyzing timestamps, cancellations, ETA, and spatial patterns.

Data Sources:  
- data\_orders.csv: Order-specific metadata including cancellation reasons, timestamps, coordinates, ETA, and driver assignment status.  
- data\_offers.csv: Mapping of offers sent to drivers for each order.

## Key Tasks & Interpretations

## 1. Distribution of Orders by Failure Reason

Orders were classified into three main categories based on status codes:  
- Cancelled before driver assigned  
- Cancelled after driver assigned  
- Rejected by system

A bar chart was created showing that "Cancelled before driver assigned" was the most frequent reason. This indicates that users often cancel before a match is found, likely due to long ETAs or app response delays.

Insight: Improving perceived wait time and app feedback could reduce premature cancellations.

## 2. Failed Orders by Hour and Reason

Orders were grouped by the hour of day and analyzed based on failure reason. A stacked histogram revealed:  
- Rejections spike during early morning and late evening.  
- Cancellations before assignment peak during off-peak hours.

Interpretation: These trends could reflect driver shortages or system strain during specific hours.

## 3. Time to Cancellation by Driver Assignment and Hour

Using cancellation time (in seconds), the average time before canceling was plotted:  
- Users wait longer to cancel if a driver was assigned.  
- Sharp peaks were visible at morning and late-night hours.

Conclusion: Assignment increases user patience. High wait times despite driver assignment could reflect geographic or traffic inefficiencies.

## 4. Distribution of ETA by Hour

Boxplots were created showing the distribution of ETA (estimated time of arrival) by hour of day.  
- ETAs are higher in early mornings and late evenings.  
- Greater ETA variance occurs during peak traffic hours.

Interpretation: ETAs affect user behavior directly. Interventions to optimize route efficiency or ETA accuracy could help reduce cancellations.

## 5. Bonus: H3 Hexagon Map of Failed Orders

Using H3 spatial indexing (resolution 8), failed orders were clustered into hexagonal bins and plotted using Folium.  
- The top hexes that contain 80% of failed orders were identified.  
- These were mapped by frequency using red marker intensities.

Conclusion: A small number of geographic areas account for the majority of failed orders.

## Final Summary

This analysis highlights multiple levers that can reduce ride failure rates at Gett:  
- Enhance app responsiveness and early feedback  
- Predict high-failure time windows and adjust supply  
- Address high-failure geographic hotspots  
- Improve ETA estimation and reduce uncertainty

## Tools Used

- Python (Pandas, Seaborn, Matplotlib)  
- H3, Folium for geospatial mapping  
- Google Colab for interactive analysis

Prepared by: Suddha Abhitej  
Project Repository: https://github.com/Suddha-Abhitej/gett-failed-orders-analysis