Question 1

To build up the distribution of failure reasons, we classify the reasons according to column 'is_driver_assigned_key' and 'order_status_key'. And then we have four failure reasons as: 1) Client cancelled after driver assigned 2) System rejected after driver assigned 3) Client cancelled before driver assigned 4) System rejected before driver assigned
An analysis of 10,716 data entries reveals that 'Client cancellation before driver assignment' accounts for the highest number of failures at 4,496, followed by 'System rejection before driver assignment' at 3,406, and 'Client cancellation after driver assignment' at 2,811. There were only 3 instances of 'System rejection after driver assignment.'
This conclusion is reasonable since client and system tends to reject orders if no driver assigned.

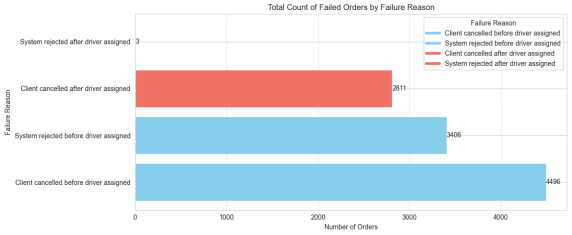
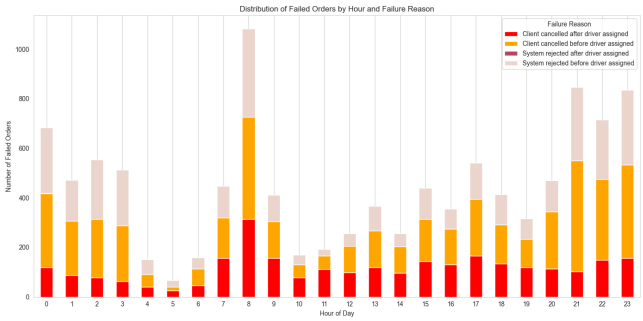


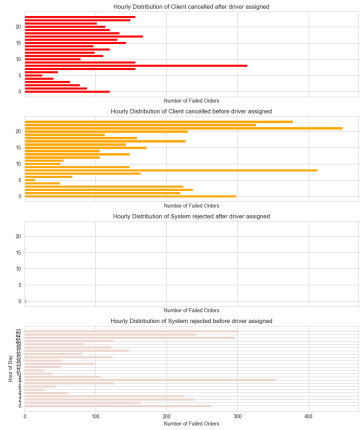
Figure 1: total failed orders by reason

Question 2

The stacked and unstacked bar graphs provide a visual representation of the failed orders by hour across four different failure reasons. From the data, it is evident that the peak hours for failed orders, particularly those cancelled by the client before driver assignment, occur during the late evening hours, with a significant spike at 21 hours. This trend could be attributed to higher demand and subsequent cancellation rates as clients make late evening travel plans. Conversely, system rejections before driver assignment exhibit a similar pattern but with less pronounced peaks, possibly indicating automated system checks that occur after a surge in bookings.



(a) Stacked Graph of Hourly Failed Orders

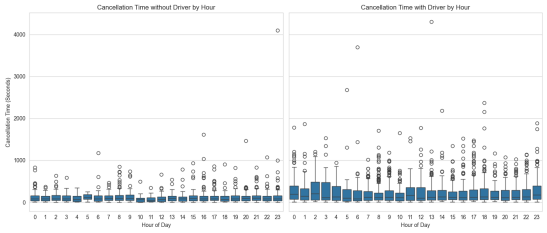


(b) Unstacked Graph of Hourly Failed Orders

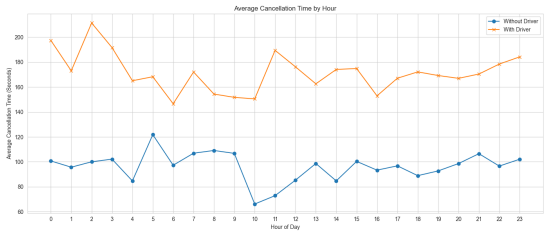
Figure 2: Hourly Distribution of Failed Orders

Question 3

The examination of the cancellation times through box plots revealed a widespread presence of outliers. Upon their removal, a more accurate trend in cancellation times was discerned. The resulting line graph indicates that the average time to cancellation without a driver remains fairly stable across all hours, suggesting consistent customer decision-making patterns throughout the day. Meanwhile, the average time to cancellation with a driver shows distinct peaks during the early and late hours, potentially reflecting a higher variability due to factors such as traffic conditions or driver availability at those times.



(a) Boxplot of Cancellation Time by Hour



(b) Average Cancellation Time by Hour, With and Without Driver

Figure 3: Cancellation Time Analysis

Question 4

The bar chart depicting the average Estimated Time of Arrival (ETA) by hour illustrates distinct trends in transportation dynamics over the course of a day. Notably, the ETA peaks during the early morning hours, suggesting less traffic or faster travel times due to lower demand. There is a noticeable dip in the mid-morning hours, which could be indicative of an increase in traffic as the day progresses. The ETA again rises towards the late morning and remains relatively high until the evening, aligning with typical rush hour patterns. The lowest ETA occurs late at night when roads are likely to be less congested, resulting in quicker transportation.

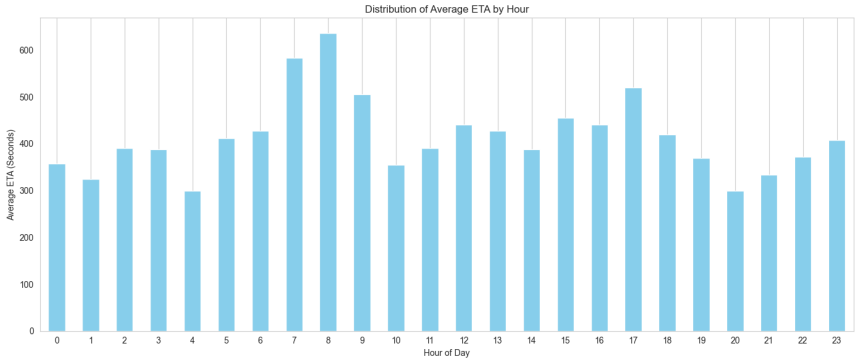


Figure 4: Distribution of Average ETA by Hour

Question 5

The hexagonal grid overlay on the map, generated using h3 and folium packages, enables an analysis of geospatial distribution of failed orders. The analysis revealed that a concentration of size 8 hexagons encapsulates 80% of all order failures, with the most intense areas depicted in darker shades indicating a higher density of failed orders. This distribution is especially prominent in urban centers, likely reflecting higher order volumes and consequently a greater number of failures. Such visualizations are instrumental for identifying hotspots that may benefit from targeted interventions to improve service reliability.

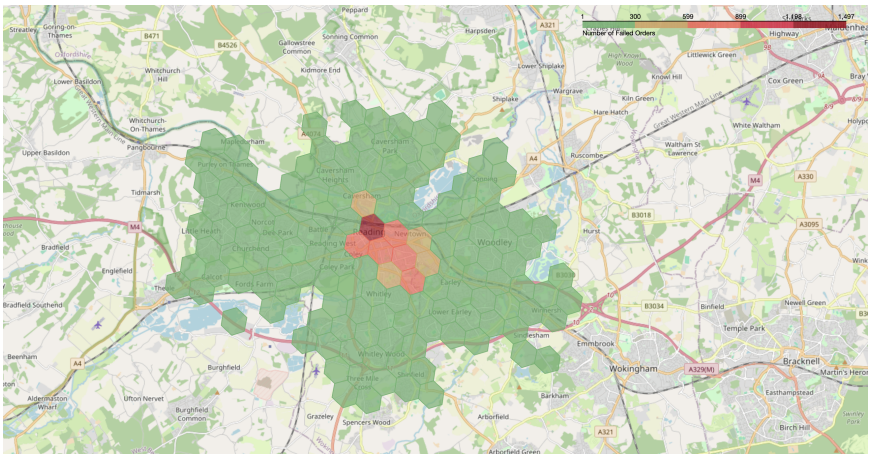


Figure 5: Geospatial Distribution of Failed Orders Across Hexagonal Grids