Insights from Failed Orders

Gett, previously known as GetTaxi, is an Israeli-developed technology platform solely focused on corporate Ground Transportation Management (GTM). They have an application where clients can order taxis, and drivers can accept their rides (offers). At the moment, when the client clicks the *Order* button in the application, the matching system searches for the most relevant drivers and offers them the order. In this task, we would like to investigate some matching metrics for orders that did not complete successfully, i.e., the customer didn't end up getting a car.

Assignment

Please complete the following tasks.

1. Build up distribution of orders according to reasons for failure: cancellations before and after driver assignment, and reasons for order rejection. Analyse the resulting plot. Which category has the highest number of orders?
2. Plot the distribution of failed orders by hours. Is there a trend that certain hours have an abnormally high proportion of one category or another? What hours are the biggest fails? How can this be explained?
3. Plot the average time to cancellation with and without driver, by the hour. If there are any outliers in the data, it would be better to remove them. Can we draw any conclusions from this plot?
4. Plot the distribution of average ETA by hours. How can this plot be explained?
5. **BONUS: Hexagons**
   Using the `h3` and `folium` packages, calculate how many size 8 hexes contain 80% of all orders from the original data sets and visualise the hexes, colouring them by the number of fails on the map.

Data Description

We have two data sets: `data_orders` and `data_offers`, both being stored in a CSV format. The `data_orders` data set contains the following columns:

- `order_datetime` – time of the order
- `origin_longitude` – longitude of the order
- `origin_latitude` – latitude of the order
- `m_order_eta` – time before order arrival
- `order_gk` – order number
- `order_status_key` – status, an enumeration consisting of the following mapping:

This data project has been used as a take-home assignment in the recruitment process for the data science positions at Gett.

- o 4 – cancelled by client
- o 9 – cancelled by system, i.e., a reject
- `is_driver_assigned_key` – whether a driver has been assigned
- `cancellation_time_in_seconds` – how many seconds passed before cancellation

The `data_offers` data set is a simple map with 2 columns:

- `order_gk` – order number, associated with the same column from the orders data set
- `offer_id` – ID of an offer

This data project has been used as a take-home assignment in the recruitment process for the data science positions at Gett.