

Customer Service (CS) Insight Analyst

Uncovering hidden patterns in customer service data

Etraveli Group selection process

Candidate: Sabrina F. Tigik

December 31st 2024

Uppsala, Sweden

Objective

Provide data-based insights on sales and customer service data to help enhance customer satisfaction and operational efficiency, fomenting healthy business growth for the Etraveli group.

Summary of the datasets: **orders**

- **orders.parquet**:
 - Contains **6 300 206** unique order IDs.
 - From the 22 columns, 16 are categorical data objects related to airline tickets transactions.
 - Within the categorical data, I chose for further analysis:
order_id, brand, partner, device_type,
booking_system_source_type, cancel_reason,
and change_reason.
 - In the remaining columns, two could be quantified and were considered in the analysis: **is_cancelled** and **is_changed**.

Summary of the datasets: **errands**

- **errands.parquet**:
 - It is a subset of the `orders` dataset.
 - Contains **2 885 296** unique errand IDs, which are associated with one or more order IDs.
 - From the 8 columns, 6 are categorical data objects related to customer service.
 - Within the categorical data, I chose for further analysis:
`order_number`, `errand_category`, `errand_type`,
`errand_action`, and `errand_channel`.
 - The only columns that could be quantified was:
`is_test_errand`.

Methodology: pre-processing the data

I started with loading and pre-processing the data:

- The `order_number` column in the `errands` dataframe needed to be converted to base-10 format to be cross-referenced with the `order_id` columns in the `orders` dataframe.
- Once converted, I added an `order_id` column with the values to the `errands` dataframe and used it to merge `orders` into `errands`.
- After that, I had three dataframes: `orders`, `errands` and `errands_in_orders`.

Methodology: `is_test_errand` column

- The `is_test_errand` column in the errands dataset was not mentioned in the Case Study file with the instructions for the study.
- `is_test_errand` has 0 and 1 values, probably intended to be boolean, but the values were read as int64.
- After careful consideration, I decided to remove the 204 404 orders flagged as test.

The decision was based on the fact that the instructions do not mention that column. Therefore, I could not make a sound judgment about its results, as the test errands could be test calls to evaluate customer service.

Methodology: data analysis approach

- After extensive data exploration, I decided to base my analysis on one crucial parameter for the client: **time**.
- The clients' time is measured as the number of errands per order ID.
- More errands indicate more time spent contacting customer service, which, even if not true, can be perceived as a lack of effort by the company to solve customers' problems.
- Consequently, a long series of customer-company interactions can potentially deteriorate the company's image to the client.

Number of errands per order

To test how common is for an order to have more than one errand:

- The basic statistics of the errands_in_orders dataframe are shown in the table:

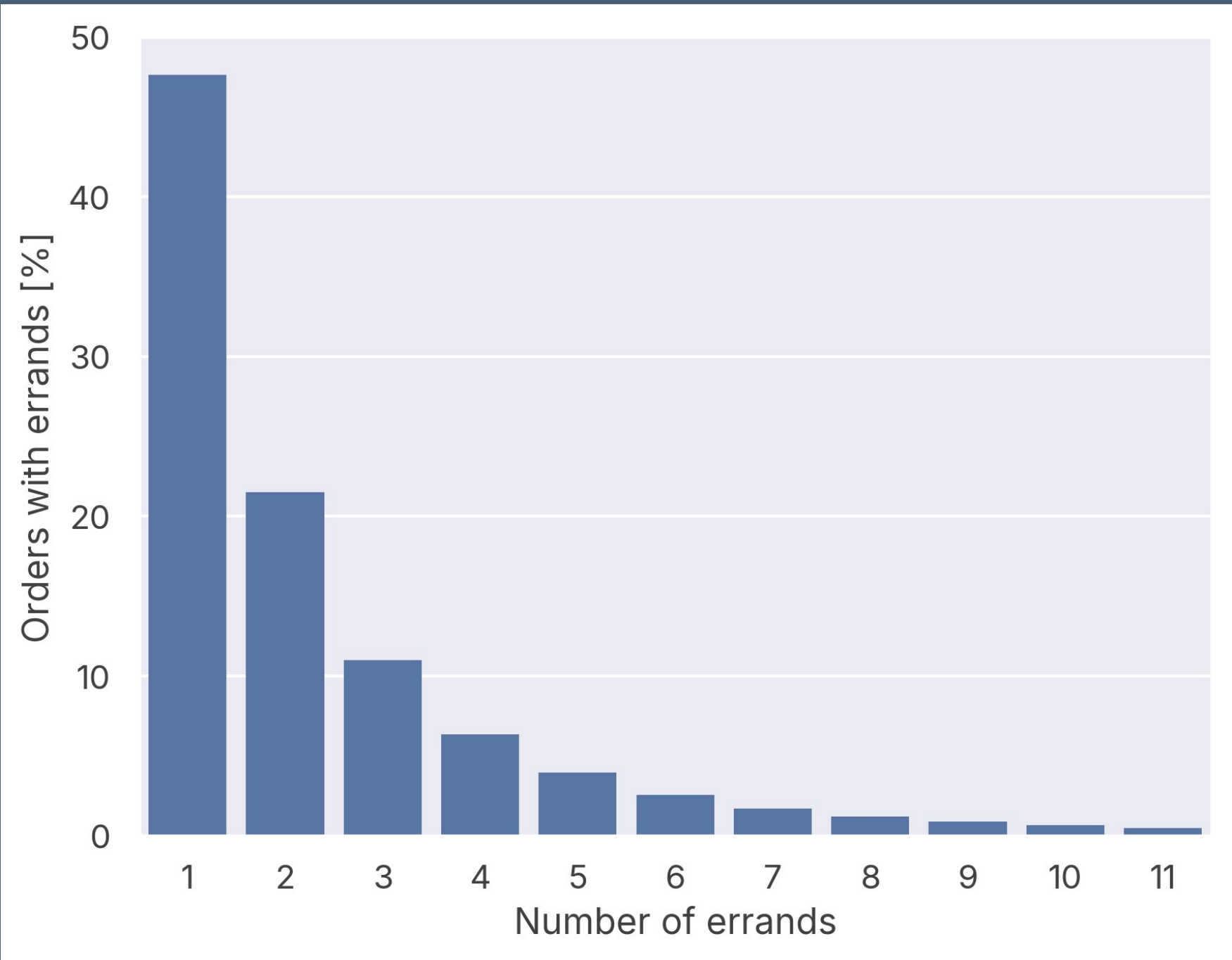
Max val	Min val	Mean	Std dev
234	1	2.62	3.14

- Based on the mean and the standard deviation, one can conclude that most of the orders have few errands.
- However, the maximum value indicates that some orders have many errands.
- To avoid bias towards high values, I excluded order IDs with errand counts above the 9th percentile.

Number of errands per order

- The table shows that the orders dataset has 1 024 728 orders with one or multiple errands.
- Counting the number of errands per order we obtain 108 different values, which I call “bins”.
- The bulk dataset is distributed in 11 bins, containing 1 004 150 orders, corresponding to ~90% of the sample.
- The box plot shows the distribution of the bulk dataset where:
 - 47,6% of orders have one errand, indicating that more than **half of the clients have to contact customer service more than once.**

	Number of bins	Number of orders	Percentage
Bulk dataset	11	1 004 150	89.8%
Outliers	97	20 578	10.2%
Total	108	1 024 728	100%



Percentage of errands with errands, per number of errands.

Testing correlations

The errand rate can be compared with information about the orders and customer-company interaction.

- From the previously selected columns, mentioned in slides 3 and 4, I kept only the `errand_category`.

In the next slides I use heat maps to evaluate how the categories within this data is related to the number of errands per order.

Errand rates within errands category

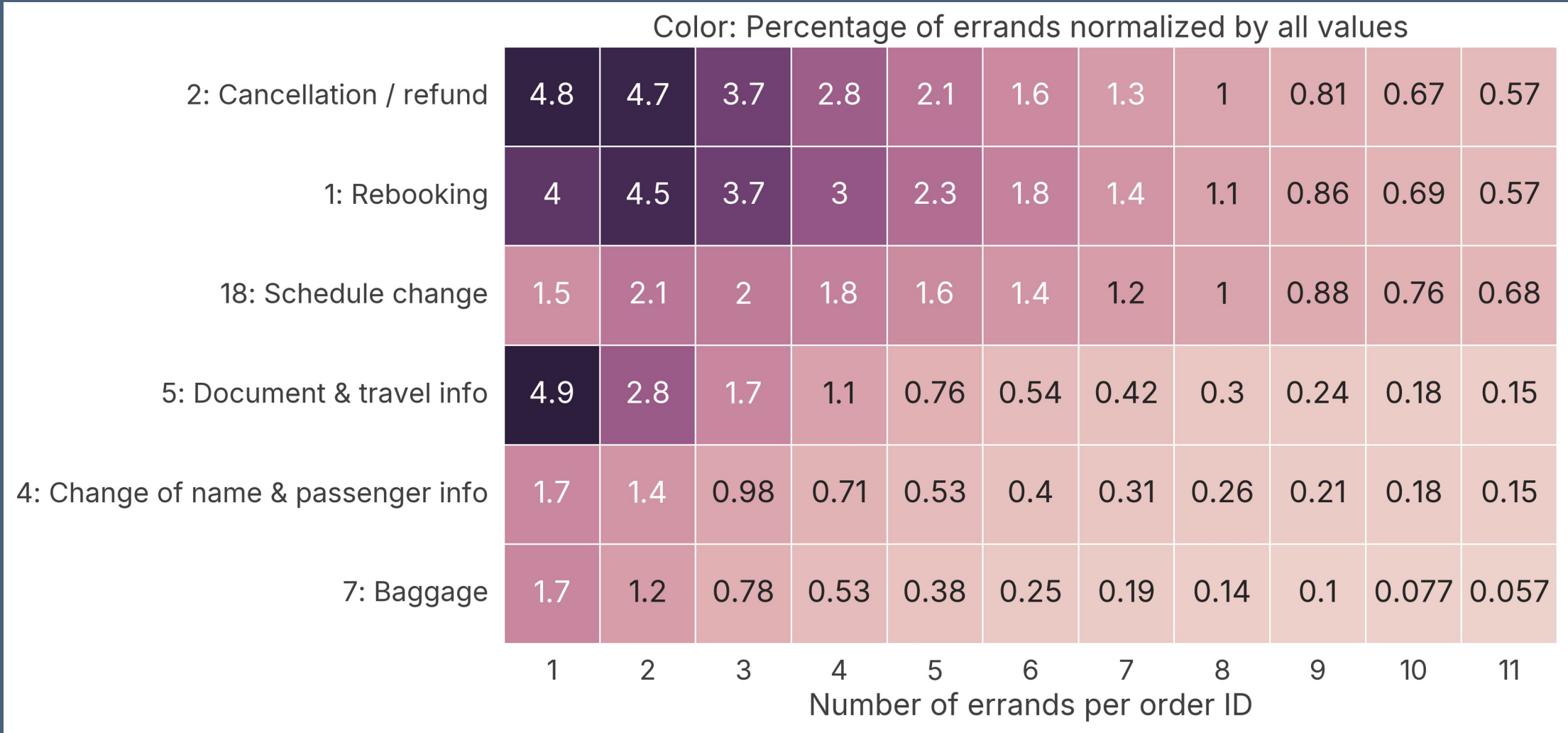
2: Cancellation / refund, 1: Rebooking and **18: Schedule change** show a tendency to extend the duration of customer service.

- These are complex issues that do not depend only on Etraveli’s actions.
- However, it is possible to reduce the impact of these issues on customer satisfaction and reduce the company’s negative perception by the clients.

Proposed action:

- Customer Service personnel training.
- Improved communication channels with booking systems and banks.

These actions can help to reduce the number of contacts per client and improve the efficiency of customer service operations, reducing costs and personnel time per client issue.



Percentage of orders with errands is distributed among the six most frequent errand categories.

Final remarks

- Data analysis shows that three errand categories are associated with extended customer service contacts.
- The issues are complex, and the solution I propose is based on a limited set of variables.
- My hypothesis is based on the fact that clients do not like to spend their time in customer service, and the company can use the extra time to tackle different issues.
- I noticed that `errand_type`, `errand_category`, `errand_action` and `errand_channel` are connected.
- Understanding how this chain of customer operations connects with the number of errands per order could improve this preliminary analysis and lead to better insights.
- Having data from customer opinion survey would also be very helpful to understand what actions could improve the customer experience when prolonged customer service interactions are unavoidable.

I learned a lot and had a lot of fun analyzing this data.

Thank you so much for the opportunity!