Food Claim Process

Company Background

Vivendo is a fast-food chain in Brazil with over 200 outlets. Customers often file claims for compensation from the company due to food poisoning. These claims are processed by the legal team, which has offices in four locations. The legal team wants to improve how long it takes to reply to customers and close claims. The head of the legal department has requested a report on how each location differs in the time it takes to close claims.

Business Question:

The legal team at Vivendo aims to improve their response time and close claims more efficiently. In order to achieve this, the head of the legal department has requested a report that compares the time it takes to close claims across each of their four locations. The following business questions will be answered through an analysis of the provided food claims dataset:

1) Which location has the highest number of claims? Are the number of claims balanced across all locations?
2) What is the distribution of the time it takes to close all claims?
3) How does the time to close claims vary by location?

Dataset:

The food claims dataset is available for download at the provided link here, with each row representing a single claim. However, before proceeding with any analysis, the dataset needs to be carefully validated and cleaned based on the information provided in the table below:

Column Name	Criteria
claim_id	Nominal. The unique identifier of the claim. Missing values are not possible due to the database structure.
time_to_close	Discrete. The number of days to close the claim. Any positive value. Replace missing values with the overall median time to close.
claim_amount	Continuous. The initial claim requested in the currency of Brazil, rounded to 2 decimal places. Replace missing values with the overall median claim amount.
amount_paid	Continuous. Final amount paid. In the currency of Brazil. Rounded to 2 decimal places. Replace missing values with the overall median amount paid.
location	Nominal. Location of the claim, one of "RECIFE", "SAO LUIS", "FORTALEZA", or "NATAL". Remove missing values.
individuals_on_claim	Discrete. Number of individuals on this claim. Minimum 1 person. Replace missing value with 0.
linked_cases	Nominal. Whether this claim is linked to other cases. Either TRUE or FALSE. Replace missing values with FALSE.
cause	Nominal. Cause of the food poisoning. One of "vegetable", "meat" or "unknown". Replace missing values with 'unknown'.

Data Cleaning and Validation:

The Food Claims dataset comprises 2000 rows and 8 columns, and no duplicates were found. However, the dataset also contained inconsistencies that required addressing before conducting any analysis.

claim_id: contains integer values ranging from 1 to 2000 with no missing values and requiring no further cleaning.

time_to_close: contains integer values ranging from 76 to 518 with no missing values.

claim_amount: contains continuous values ranging from 1637.94 to 76106.8. To ensure consistency, some values that were not rounded to 2 decimal places were rounded, and the 'R$' character was removed from all values.

amount_paid: contains continuous values ranging from 1516.72 to 52498.75. We observed 36 missing values in this variable, which we replaced with the overall median value of 20105.7 to ensure consistency. Additionally, we rounded some values that were not rounded to 2 decimal places.

location: contains nominal values as expected, with location of claims including 'RECIFE', 'FORTALEZA', 'SAO LUIS', and 'NATAL'. There were no missing values found in this column.

Individuals_on_claim: contains integer values ranging from 1 to 15. There were no missing values found in this column.

linked_cases: contains nominal values as expected, specifically 'TRUE' or 'FALSE'. However, there were 26 missing values which were replaced with 'FALSE'.

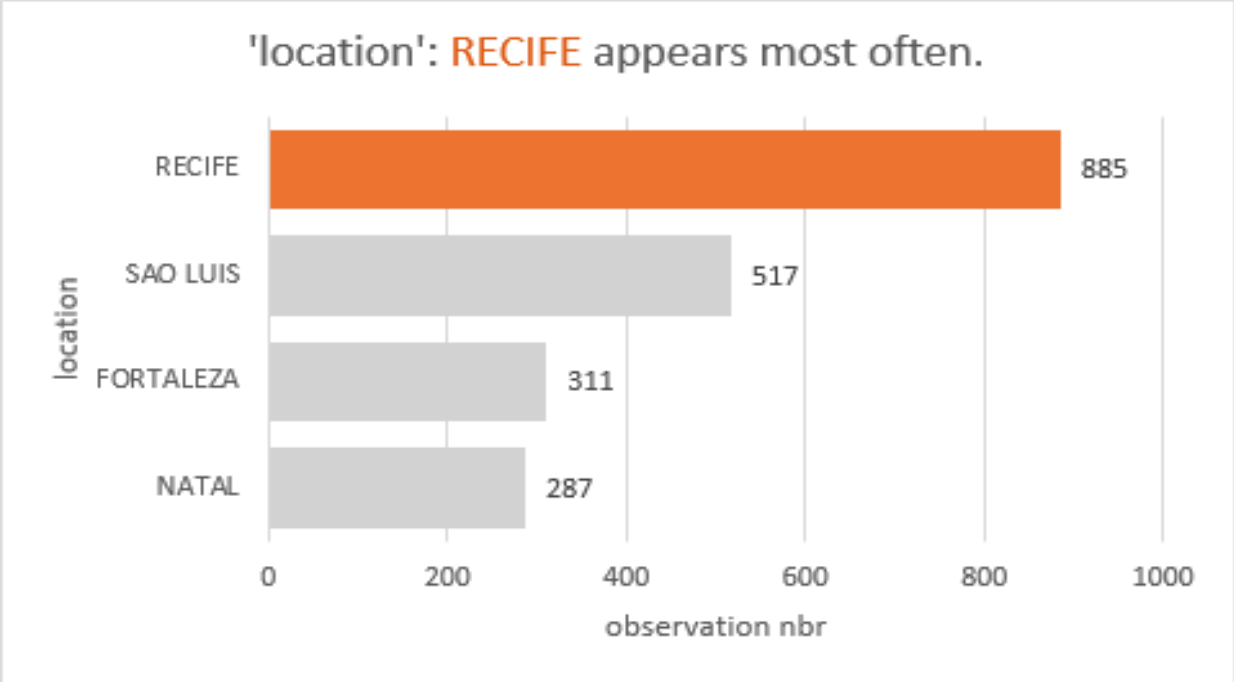
cause: contains nominal values which include "meat", "vegetable", and "unknown". There were 16 rows with the value “VEGETABLES”, which I corrected to “vegetable”. There were no missing values found in this column.

Exploratory Data Analysis:

1) Which location has the highest number of claims? Are the number of claims balanced across all locations?

Upon examining the provided data, it is evident that the legal team at Vivendo has offices in four locations across Brazil, including Recife, Sao Luis, Fortaleza, and Natal. The analysis of the count of claims by location, depicted in a bar chart, reveals that the RECIFE location has the highest number of claims with 885, followed by SAO LUIS with 517, FORTALEZA with 311, and NATAL with 287. It is worth noting that the number of claims is not balanced across all locations, with RECIFE having almost three times the number of claims as NATAL.

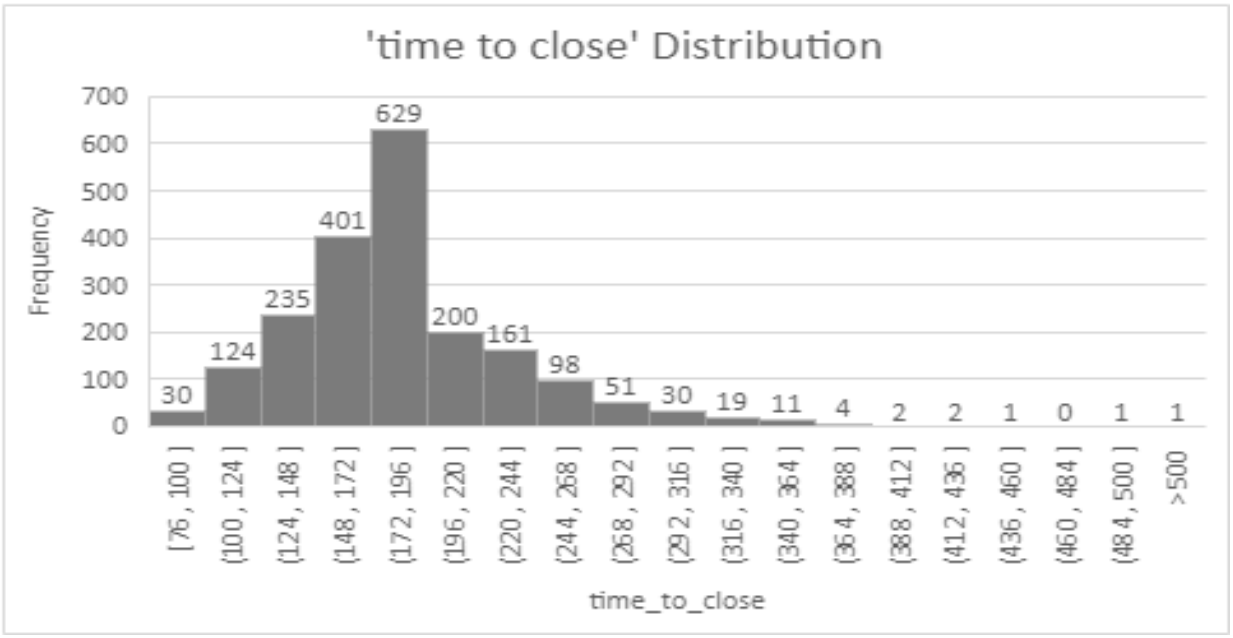
To address the time , we need to further investigate the time it takes to close claims by location. It is possible that the location with the highest number of claims also has the longest time to close them, which would require more resources and attention from the legal team.



2) What is the distribution of the time it takes to close all claims?

We can see that the time it takes to close all claims varies widely, ranging from 76 to 518 days. The most frequently occurring time-to-close values seem to be between 170 and 190 days with an average of 186 days, but there are also several peaks in other ranges, such as 160-165 days and 180-185 days.

It's worth noting that there are a few outliers with very high time-to-close values, such as 453 and 518 days. However, since there is no context provided, it is unclear whether these outliers are a normal part of the data or whether they represent exceptional cases.

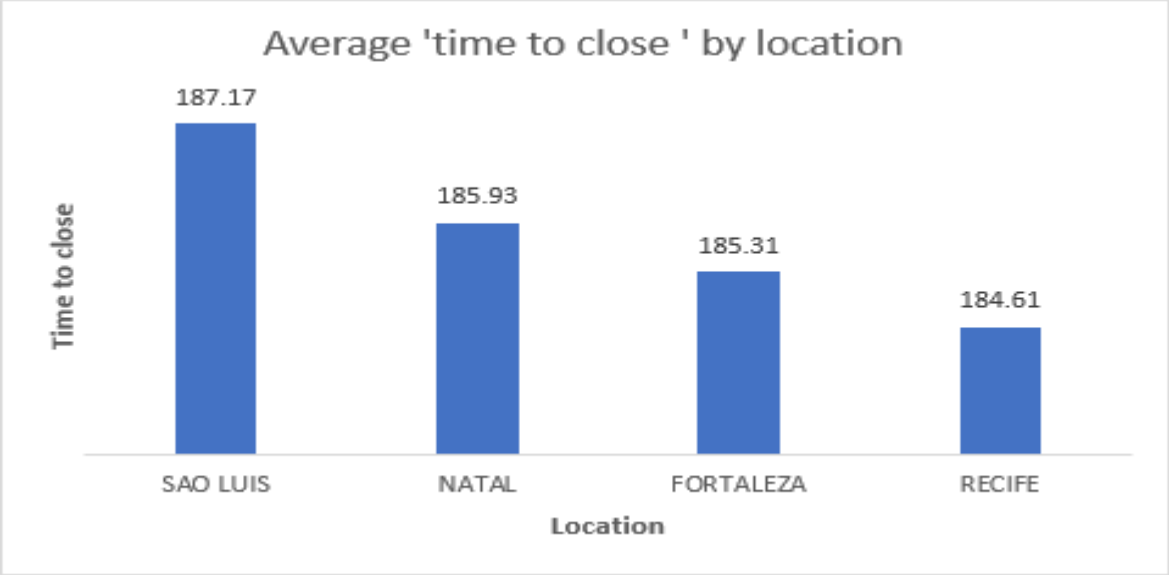


Despite the data being highly variable and lacking a distinct pattern or distribution, the histogram reveals a right-skewed distribution of the time it takes to close all claims, with a longer tail of values on the right side. This suggests that most claims are resolved quickly, but there are a few that take much longer. By utilizing the additional data provided, we can analyze the time it takes to close claims based on location and type of claim, which can yield valuable insights for the legal team.

3) How does the time to close claims vary by location?

a) Average time to close by location:

Although the average time it takes to close a claim varies slightly across the four locations (SAO LUIS, NATAL, FORTALEZA, and RECIFE), the differences are not significant. The average time ranges from 184.61 days in RECIFE to 187.17 days in SAO LUIS, with an overall difference of 3 days between all four locations. However, to draw more conclusive insights about the relationship between location and time to close, we would need additional information such as standard deviation or other measures of variability.



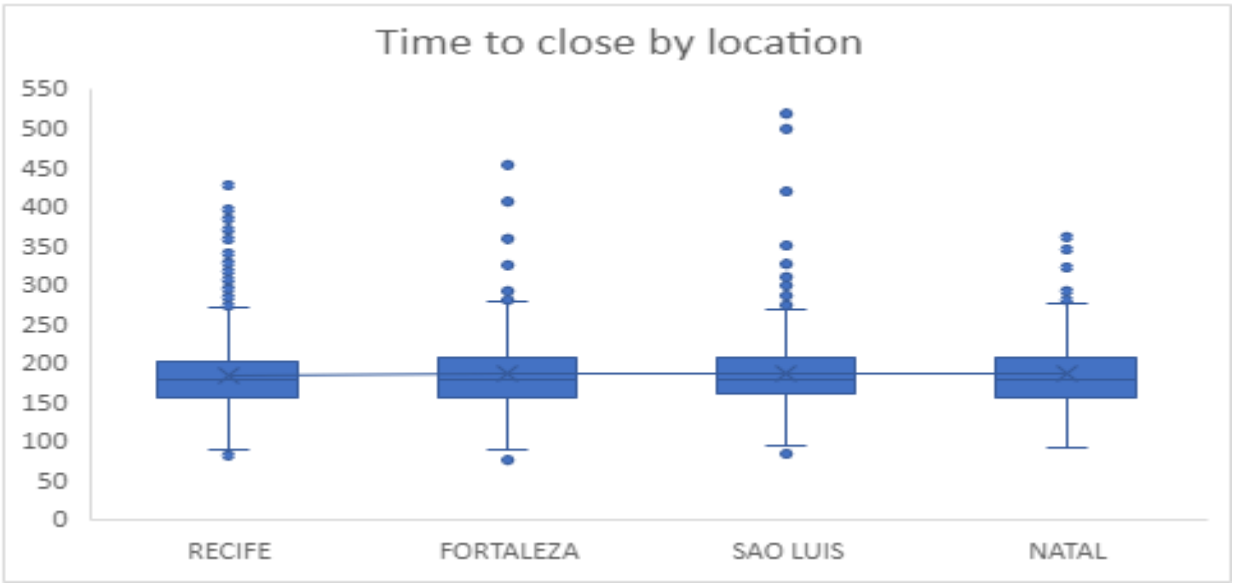
b) Measurement of variability:

Location	Standard Deviation	Median	Average
SAO LUIS	53.12609363	180	187.17
RECIFE	48.01325402	179	185.93
FORTALEZA	47.58299983	179	185.31
NATAL	47.05916019	178	184.61

we can see that the standard deviation of time to close is relatively similar across all four locations, ranging from 47 to 53. This suggests that the variability of time to close is consistent across these locations.

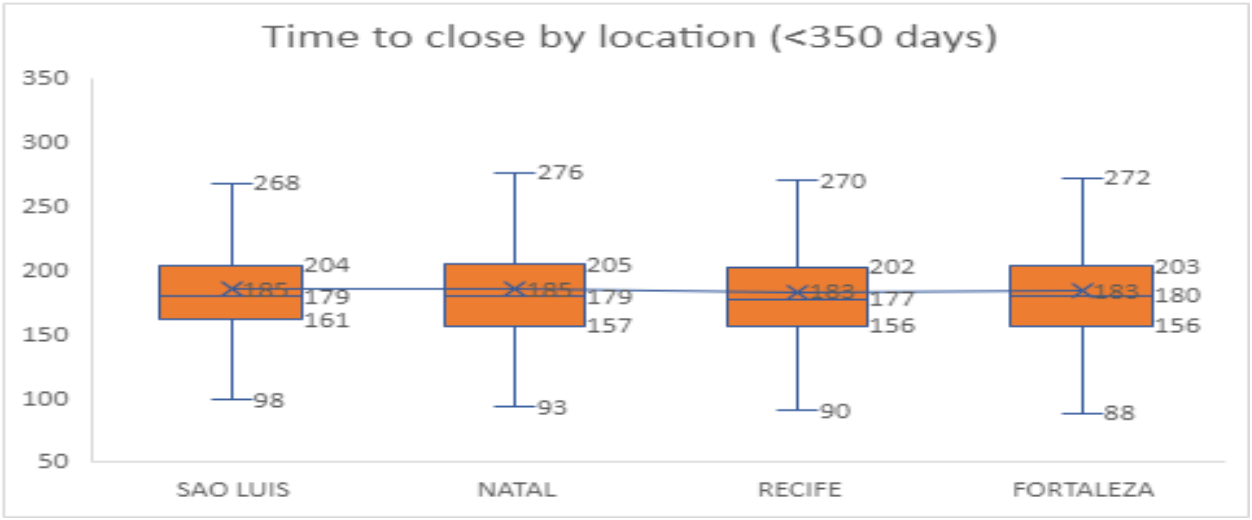
In terms of the median and average of time to close, we can see that there is also relatively little variation across these locations. The median time to close ranges from 178 to 180, while the average time to close ranges from 184.61 to 187.17.

We can visualize this relationship between time to close and location using a box plot, which shows the distribution of time to close for each location. This can help us see any potential outliers or differences in the spread of the data.



The box plot analysis shows that the distribution of time to close claims is similar across all locations, except for Recife which had a lower median, upper and lower quartiles compared to the other locations. Sao Luis had the largest range of time to close, as indicated by the largest standard deviation, which could be attributed to two extreme outliers (claims 518 and 499) that had high claim amounts and involved multiple individuals, leading to longer times to close. In contrast, the other locations had a more limited range of time to close claims.

Based on the first box plot and the previous histogram, it can be observed that the majority of the time to close values are below 350 days. Therefore, by excluding outliers and only considering cases that were closed within 350 days, the second box plot shows that Recife continues to outperform the other locations. Recife's median, upper, and lower quartiles remain lower than or equal to those of the other locations.



Conclusion and Recommendations:

Based on our analysis, we can conclude that Recife location outperformed the other locations in terms of time to close claims, even after removing outliers. The distribution of time to close claims is mostly under 350 days. Sao Luis had the widest range of time to close, likely due to extreme outliers. However, we found that the number of claims is not balanced across all locations, with Recife having a significantly higher number of claims than the other locations.

To improve the response time and close claims more efficiently, the legal team at Vivendo may need to why Recife outperforms the other locations in terms of closing claims in a timely manner. They should also investigate factors such as workload distribution across locations, staff training, and communication between locations. Additionally, further analysis could be done to explore the root causes of outliers, such as examining the type of claims, the complexity of cases, or potential bottlenecks in the process. Overall, a more detailed investigation of the underlying factors that affect the time to close claims could lead to better strategies for improving the efficiency of the process.