The dataset consists of 98 rows and 8 columns.

First, I compared the data to its description in the data dictionary.

The column “Time to Close” presented one negative value. So, the negative value row was dropped, and the dataset remained with 97 rows.

The type of the column “Claim Amount” was text. To turn it into numeric, the values were updated (the numbers were extracted from the string and the column was turned into numeric).

There are six claims without Individuals (0). Even though there are no people associated with them the rows were kept.

According to the data dictionary, the blank values in the column Cause were replaced with ‘unknown’.

According to location, São Luis is the city with more claims, followed by Recife. Both cities concentrate 55.67% of total claims.

Gráfico, Gráfico de barras

Descrição gerada automaticamente

To analyze the variable time to close claims, I used a boxplot and a histogram.

The histogram shows the distribution of the claims. Most of them were closed in up to 1000 days.

Gráfico, Histograma

Descrição gerada automaticamente

The boxplot shows that 25% of the claims take up to 349 days to close, half of them take up to 639 days and 75% take up to 1143 days.

Gráfico, Gráfico de caixa estreita

Descrição gerada automaticamente

Tabela

Descrição gerada automaticamente

São Luís is the city in which it takes longer to close a claim. Observing the boxplot, there are no outliers for the city. All the other cities present outliers which make the mean higher when compared to the median.

Tabela

Descrição gerada automaticamente

Gráfico, Gráfico de caixa estreita

Descrição gerada automaticamente