23-3-2022

Heatmaps & Boxplots

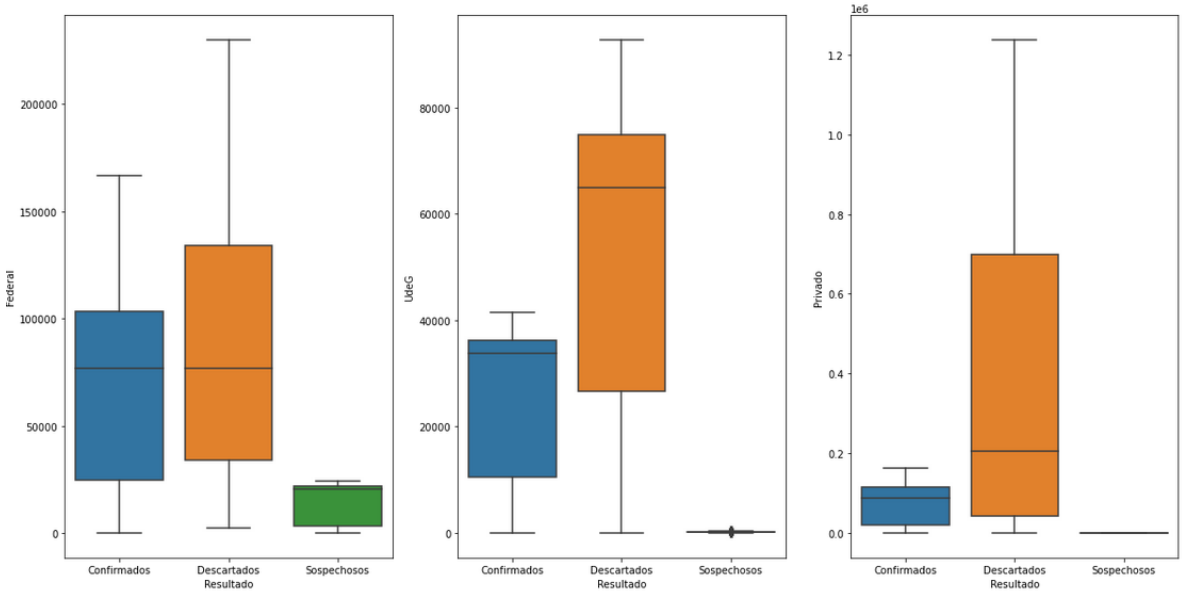
Edgar Asael Martínez

Tania Sayuri Guizado Hernández | A01640092
Ricardo Jiménez Ureña | A01636825
Carlos David Blanco | A01633323
Rafael Jiménez | A01637850
Carlos Estrada Ceballos | A01638214

DIGITAL TOOLS: THE ART OF ANALYTICS – TC1002S.100

Correlation

Boxplots



Answers to the questions:

- **Is there any variable that does not provide information?**
  Yes, the "Rapidas" variable.
- **If you had to eliminate variables, which ones would you eliminate and why?**
  I would eliminate the "Rapidas" column as the values it contains do not contribute to the analysis and are mostly *Na* values.
- **Are there variables with strange data?**
  Yes, variables such as the "Privado" and "Rapidas" columns contain non-numeric values within the test records (Na), which can cause problems when trying to analyze our variables.
- **If you compare the variables, are they all in similar ranges? Do you think this will affect the analysis?**
  Some variables have very varied ranges, this can affect if there are many outliers as it may cause an incorrect analysis of the data as they do not reflect the general behavior of the data affecting the final results.
- **Can you find groups that look alike? What are these groups?**
  We can observe that the most similar groups are the " Confirmados " and "Descartados" results tests by the "Federal" variable or column within our Boxplot analysis, this gives us an understanding that on a large percentage of days, the number of "Confirmados" and " Descartados" result tests were very similar.

Git: https://github.com/sayuriGui/AnaliticsChallenge/blob/main/heatmaps_boxplots.ipynb