Data Analyst – Challenge 1

For the data analysis in this module, we considered these three Excel spreadsheets:


muestraDatosCovid .csv


201128 Descriptores_.xlsx


201128 Catalogos.xlsx

After this, I considered using the pivot tables that you will find in the other sheets of the attached document below this text.


Resolución Challenge.xlsx

In the spreadsheet, you can observe how:

1. Two relational columns were established to obtain the sex "male/female" and to obtain the value of the names of the corresponding states to increase clarity about the data.

2. Subsequently, correlations were made between the added columns and the value of both the mode and the median by state. On the "Age Pivot Table" sheet, a pivot table was generated that classified the states by their name, along with the average age per state, the average mode, and median per state. This is because Excel does not allow us to work with modes and medians directly in pivot tables.

3. Later, a difference in days was made between the dates of symptom presentation and the date of notification of having COVID.

4. To finish the challenge, the process from step 2 was repeated, now establishing relationships between days and sex, and classifying in the same way the mean, median, and mode in another sheet ("Days Pivot Table") in a pivot table.

Relevant notes and conclussions.

It is highly likely that the most repeated value (mode) being 0 is due to the fact that in different facilities (especially government ones), the test is ordered when doctors notice the existence of these symptoms. However, there are rare occasions where patients (and even doctors) report the onset of symptoms prior to the admission date for COVID.