**Terminology for column headers** on my Project on Knime, based on Covid cases in our State (RS, Brazil):

I began my data collecting (from a COVID-19 panel made available by our Health Secretary) with a selection of geographical data and 7 variables to describe a profile of each COVID-19 case. This is the first block of nodes. I wished to perform, in my analysis, a clustering of COVID-19.

A) I have four levels of geographical aggregation of the 497 cities in our State, i.e., Health administration of our State is divided (according to geographical proximity, here on the second block of nodes) into:

- **MUNICIPIO**=**Municipios**=**NomeMunic**=**Munic** 🡪 city names or municipalities.

- 19 regional coordination headquarters, on the column “**CRS/Municipio Sede**”;

- 30 Health Regions (grouping arbitrary numbers of cities, according to geographical proximity), on the column “**RegSaud**”.

- After some time collecting data about COVID-19, our Health Secretary decided to regroup some Health Regions into 21 COVID Regions, here in the column “**Reg-Covid**”, also by geographical proximity.

1) **GENERO** = Gender of patient.

2) **Fx\_Etar** = Age Group of patient.

3) **CRITERIO**=diagnostic criteria for Covid cases (and registers, with 1 case per line).

4) **DATA\_CONFIRM** = DiagnosticDate of COVID positive cases.

5) **EVOLUCAO**=Evolution (Recuperado=Recovered; Obito=Death) 🡪 cases severity.

6) **HOSPITALIZACAO** = Hospitalization (Sim=Yes/Nao=No)" 🡪 cases severity.

7) **SRAG** – SARS 🡪 cases severity.

I intended to make a **Time Series** with all cases diagnosed in the period of my research.

B) I also downloaded (from another Secretary (DEE-RS)) a file with gender and group ages of population in each city (in 36 columns), here on the third block of nodes.

C) on the fourth block of nodes I worked with information of a sample of respondents to my survey (made on Google Forms). Straightly in Knime I normalized or labeled those answers according to a crescent economic influence (“0” or “1” for binary categorical variables; “0” to “7” for different options on questions using ordinal variables:

c.1) I got 26 columns for the profile of survey respondents and their businesses. **Idade** = Age; **TmpForm** = TimeSinceGraduation; **Genero** = Gender; **Formac** = Academic career; **AtuaPublPriv** = professional sector; **RotinaClinGest** = everyday activities; a sequence of 8 columns for **professional background**; **Gest+Covid** and **NMeses** for coincidence between pandemic and management activities; the same 4 geographical columns above mentioned (+2 columns for size (**NCadeiras**) and district location (**LocalCons**) of business); **AtendPart**, **%Partic**, **%Conv**, **%Cooper** for cash inflow methods.

c.2) I also adopted 8 variables as economic metrics. (**FuncMonth** = If business was working in each month; **DemanMonth** = customers interests in make business in each month; **EfetMonth** = sales made in each month; **OcupCadMonth** = time spent with individual clients in each month; **CustFXMonth** = fixed costs in each month; **CustVarMonth** = variable costs in each month; **FatBrtMonth** = gross invoices in each month; **LucratMonth** = profitability in each month). But Google Forms answered me back with one different colum by month (total of sixteen months of my research interval) for each metric, which generated (8\*16=) 128 columns (while I expected just 8).