**Covid-19 in Mexico analysis**

Choosing the topic for analysis was one of the first challenges we encountered during the project. We started by trying to find subjects that were related to the team profile. Most of the subjects we were discussing were related to economic data (GDP, unemployment rates, etc.) and we were reviewing the data from international agencies. While we were approaching to determine the topic, most of the countries we wanted to compare did not have the same years length in data, so it was not the best option for analysis.

In the start, we chose a topic related to the BRICS and Mexico economic situation. After reviewing the data, we realized that it did not have enough data for the analysis, so this option was discarded. In that moment, the strategy changed to look for databases that accomplished the requirements. This process took us around 2 days to finally agree with a recent topic we wanted and were interested to, which leaded us to Covid-19 in Mexico. Being honest, at first, we were not keen on the Covid-19 subject but in the end, it turned out to be not just interesting but also enlightening.

Later we started reviewing the data to visualize the objective and scope of the analysis. For that we found several catalogues and definitions of the columns integrated in the database which helped us a lot. Next, we started cleaning up the data were we had to use visual studio code to remove some accents and special characters because after trying to do it while reading the csv file with diverse encodings, the accents and special characters remained in the database. During this process, we also had to change uppercases and lowercases in some texts, review a second time accents and special characters, and change them, replace values in database with the catalogues data and at last check for missing data.

At last, we had a main database to start the analysis which we segment in subtopics for research:

* Distribution and evolution of the disease.
* Age and Gender risk mortality.
* Pre-existing diseases associated with COVID-19.
* Correlation between death and diseases.
* Death probability.
* Cases by states, municipalities, hospitals, and mortality.

The subtopics were distributed in the team for making the analysis and visualizations. The results of those analysis are the following:

* From January to May, in Mexico there are around 26,000 cases that had a positive result in the tests of Covid-19, which means that 1 of every 4 people that had test have contracted the virus despite from the contingency plan implemented by the Mexican government. From these positive cases the distribution by gender is 58% men and 42% women.
* Furthermore, we observed that the majority of the cases are between the age of 30 and 60 years old, which is normal given that most of the people between those ages have to work and are more exposed than the ones staying at home.
* Sadly 10% of the Covid-19 cases end up being fatal and more than 80% of the deaths are between the age of 40 and 80 years old. This results also confirm what other countries were warning about the age risk, there is more probability of dying because of the virus for older people.
* Also, we find that age and pre-existence conditions increase the probability of dying of a person. Specially diabetes and obesity could increase the probability of death in up to 4% (considering all infected population). In addition, we observed that all pre-existing conditions had a correlation with the possibility that a person could die from Covid-19. Nevertheless, the correlations are very low (highest is around 0.15) which means that a person with those conditions can still recover from the disease.
* We confirmed that the positive cases over time had an exponential growth (we already knew this same behavior from countries who were hit by the virus before Mexico). On the other hand, the number of decease cases have a logarithmic growth which is an expected result from the model that the Mexican government chose to implement called the Sentinel Model during the phases of contingency.
* As we could see from other countries cases, given the density of the population, the number of positive cases also varied in Mexico per states. Nonetheless each state has implemented a different strategy so the results may vary.
* After illustrating the cases on maps, the results were as expected, both the density and the implemented plans had important repercussions on the spread and distribution of the virus. The state with most positive cases is Mexico City with 8,158, its followed by the State of Mexico with 3,210 and Baja California with 2,016. Also, those three states have the higher number of deaths in the country. As a matter of fact, that does not mean that those states have the highest mortality rate. The states with the highest mortality rate are Colima with 51.72%, Oaxaca with 32.97% and Zacatecas with 30.28%.