Group Delivery

18/01/2021 to 25/01/2021

Group C

Roberto Molleda (@Rmolleda)

Alfonso García (@alfonsogarcia-git)

Leonardo Sánchez (@leosanchezsoler)

Data Science

The Bridge

Index

General Vision

This project aims to **shed light** on the current state of the coronavirus pandemic. Since records from the spread and evolution of the disease during the last months are available, it's been possible to **make a report** focused on five different countries around the globe (**Mexico, Brazil, Iran, Netherlands** and **Spain**).

Our World In Data (**OWID**) dataset has been the **main source** for this delivery, but there were also others that provided other variables of potential interest for deeper investigations in the future.

Goals

The **main goal** was to deploy all the accumulated expertise in EDA achieved from the past weeks in order to go further than required.

That is the reason why, after just one week, the team was able to have a **public website** with daily updated info. Working and developing a **OOP structure** that created plots just by introducing a few parameters was successfully accomplished too.

Specifications

Software

Python in different versions has been the **main software** used for this task.

Hardware

Almost **every computer** can run this program, even though some devices may last more than expected.

For instance, here is the list of devices used by the members of the group:

@Rmolleda

- MSI:
    - Processor → Intel(R) Core(™) i5-4200H Cpu @ 2.80 GHz
    - Memory → 8.00 GB

@alfonsogarcia-git

- MacbookPro:
    - Processor → 2.9 Ghz Dual-Core i7
    - Memory → 8 GB 1600 MHz DDR3
    - Intel Graphics 4000 1536 MB

@leosanchezsoler

- MacbookAir:
    - Processor → 1.6 GHz Dual-Core Intel Core i5
    - Memory → 4 GB 1600 MHz DDR3
    - Graphics → Intel HD Graphics 6000 1536 MB

Requirements

To execute the program, **all tools and datasets** are inside the program, in order to make things as easy as possible for the user.

In the case of the **api**, it is mandatory to ask for the token it in this **url**.

Steps

1. Research the context

In this particular case,

2. Get Data

As mentioned before, the data was obtained from this **url**, which is the result of merging five different datasets with variables coming from transnational institutions (**United Nations**, **World Bank**, **European Centre for Disease Prevention and Control**…), universities (**Johns Hopkins University**) and geographical and temporal coverage collected by **OWID.**

3. Data Wrangling

For this project, the Data Wrangling phase did not take much time. OWID dataset was quite **tidy and organized.** After a glimpse of the data, there weren't strange patterns or significant rare values.

Despite the fact that some variables were fully composed of NaN values, they weren't removed in order to track records of them in the near future.

4. Data Mining/ Clean Data

After a closer look, some patterns were found in the data.

For instance, in **Spain** there **weren't records** of new cases or deaths **during the weekend.** In that particular case, NaN values of those days were **replaced** with the mean of that week.

Others

For this project the group decided to take a step further and share the results in a **website** to broaden the research to a larger audience.

The project can also be downloaded from our **GitHub repository**.

Sources

Data Source

- *OWID dataset*

  - https://ourworldindata.org/coronavirus-source-data

- *OWID dataset documentation*

  - https://github.com/owid/covid-19-data/blob/master/public/data/README.md

Solve code doubts

- Stack Overflow

  - https://stackoverflow.com/

- Plotly documentation

  - https://plotly.com/python/

- Flask documentation

  - https://flask.palletsprojects.com/en/1.1.x/

- Pandas documentation

  - https://pandas.pydata.org/

Organization tools

- Trello

  - https://trello.com/

- Creately

  - https://creately.com/

- Google Drive

    - https://drive.google.com/drive/

- Discord

    - https://discord.com/

Git repository

- GitHub
    - https://github.com/

Conclusions

After achieving so many unexpected goals in a very little amount of time, it was possible to draw some conclusions both on the project itself and the inspected data.

Code conclusions

The issue of tackling so many different areas of coding has been a major force against our deadlines. In spite of all the difficulties that came across the project, as a team, we could reverse the situation and even encountered unexpected (but satisfying results).

This project taught the members of the team how synergies are, by far, the most powerful source for developing fast projects successfully.

Data conclusions

From the inspected data the team concluded that there's a direct correlation between the total deaths and the total cases. There's also a linnear correlation between icu patients and the number of new cases.

The team expected to achieve some remarkable insights from the stringency data, but it will require further investigations and a deeper research on this topic.

It is obvious that data is not uniform depending on the country, so it is very hard to find