EM1404-2 project assignment (ay 2021-22)

General information and data source

The assignment will be focused on exploiting the COVID-19 Vaccine Data Repository made available by the European Centre for Disease Prevention and Control (ECDC) that can be found here:
https://www.ecdc.europa.eu/en/publications-data/data-covid-19-vaccination-eu-eea
The repository provides historical data about the vaccination (number of vaccine doses distributed by manufacturers, number of first, second and unspecified doses administered) collected during the last year among the countries of the EU/EEA.

Data about the spread of COVID-19 and vaccinations is adopted as the basis for many analysis and visualization websites around the world. As a trivial example, relating vaccination data with the diffusion of COVID-19 helps in understanding the efficiency of the vaccine, but many other factors need to be considered for an accurate analysis, such as local events (schools openings, holidays, etc.) government actions and restrictions, weather and temperature, just to mention a few. Further analyses could be done, relating the progress of the vaccination campaign to other factors like economic indicators, features of the healthcare systems, government political parties and so on.

For this assignment, each group is asked to use the ECDC data, together with other data sources considered useful (e.g. COVID-19 Data made available by the CSSE at Johns Hopkins University at https://github.com/CSSEGISandData/COVID-19), to carry on an analysis that can be meaningful to answer a few key questions, spanning from the effect of the vaccine and other factors to the spread of COVID-19 to the correlation of the efficiency of the vaccination campaign with economical and political factors.

Each group should be autonomous in choosing the extent of the analysis to conduct, the data sources to use, the statistical operation to perform, the visualization to present and, in the end, the conclusions that can be drawn from the analysis.

Process

In order to fulfill the assignment, each group should decide:

- The phenomenon to study. There is no predefined choice, you can analyze anything you find interesting. The goal is to study the effect of the vaccination on the spread of the virus and other aspect of the real world or, vice-versa, if and how parameters of the real world (either practical, political or physical) affect the vaccination;
- Which data sources, in addition to the data from ECDC, to use. How to integrate them together, which action should be taken to make them compatible and comparable;
- Which kind of statistical analysis to perform and which kind of visualization to adopt to explain the data.

Tools

The group can use any tool available in the Python ecosystem, including any third library. Actually, the group is encouraged in finding new tools and learning how to use them, since this will be an important part of their journey in the world of data analytics.

Presentation

Each group will be asked to present their work in a 30 minutes long presentation during the exam session. Each presentation will discuss:

- The scope and goals of the study;
- The results obtained and the observation made;
- The final thesis substantiated by the data, together with some explanatory hypothesis.

This will be done with the support of any kind of material the group feels userful, including slides, notebook, small videos, etc. Each member of the group should take part in the presentation. After the presentation individual oral examinations will be performed by the students of the group. For students that do not have already passed the first part, an additional oral session could be required once they pass it.
The group can ask to give the **presentation at any time** before the end of the last session of this academic year (2021-22) by sending an email to luca.cosmo@unive.it to agree on a date. If the presentation is given outside one of the designed exam periods, the final mark will be registered in the next period.
The group has to **submit by email all the material** useful for the evaluation of the project (including commented code and presentation slides) at least **three days before** the agreed presentation day.

Evaluation

The evaluation criteria will be the following:

- up to 10 points for methodological and technical quality of the study and relevance of the topic studied;
- up to 10 points for proper use of Python tools for data analytics
- up to 10 points for quality of the presentation and of the final explanation of the phenomenon observed

Laude will be assigned to outstanding works. The individual oral examination can raise or lower the grading of the assignment. The final score of the whole EM1404 exam will be an average of the first and second part.