PROJECT TITLE:
THE COVID-19 DATA VISUALIZATION AND STORY TELLING

Analytics Academy Course (Semester 2022-July)

Alexandre Santana

João Dias

Paulo Sousa

Data set and Project Objectives

Data Set Description:

Filename: covid-data.csv
This is a master dataset of Covid-19 (from 24/2/2020 to 5/3/2022) which involves the global level data related to Covid-19 for all countries.
This dataset has been provided by "Our World in Data".

Objectives:

• Applying the data cleaning and feature engineering methods to make the dataset prepared for data analysis
• Performing exploratory analysis on datasets and identifying the most influential features
• Analysing the data and creating the insightful information and communicate them through the data visualization graphs and dashboards.
• The following tools were used in this project:
1. Excel
2. Python
3. Tableau Desktop
4. PowerPoint

Analytics Academy

Intro 3D (Covid Worldwide)

Generated with Excel

Asking Questions

Q1 - How have Covid cases and deaths evolved in the world, and which continents have reached the highest numbers?

Q2 - Which are the Top 10 countries with the most cases and most deaths per million inhabitants?

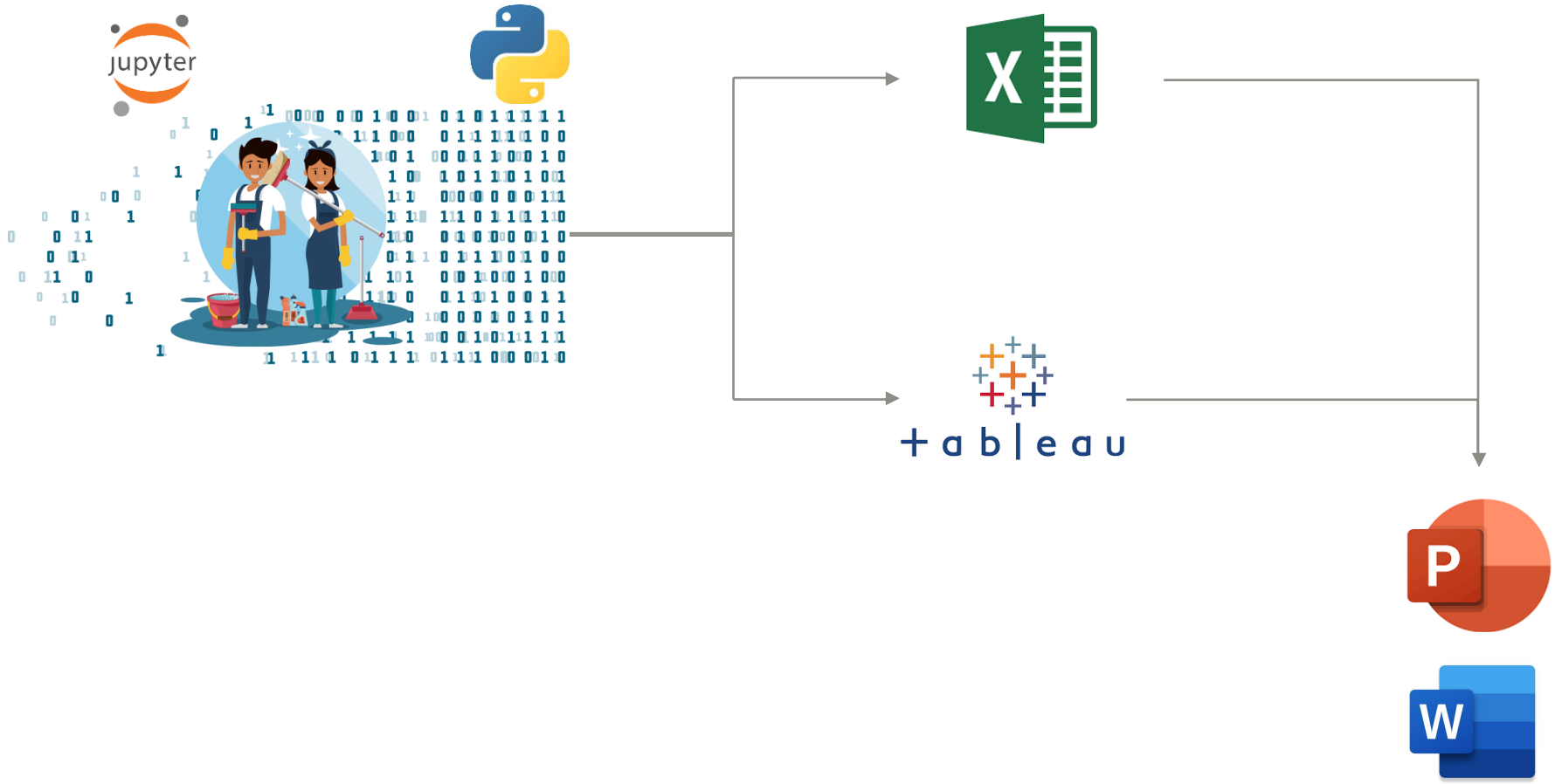
Q3 - How did the evolution of positive cases affect the number of patients in hospitals in Europe (EU)?

Q4 - In Europe (EU), what is the correlation between GDP Per Capita vs New Cases and also Population Density vs Cases per Million?

Q5 - What is the influence of vaccination on new cases and deaths in Portugal?

Q6 - How was the evolution of new cases and deaths in Portugal, and how can we predict the future?

Data Analysis Tools Flux



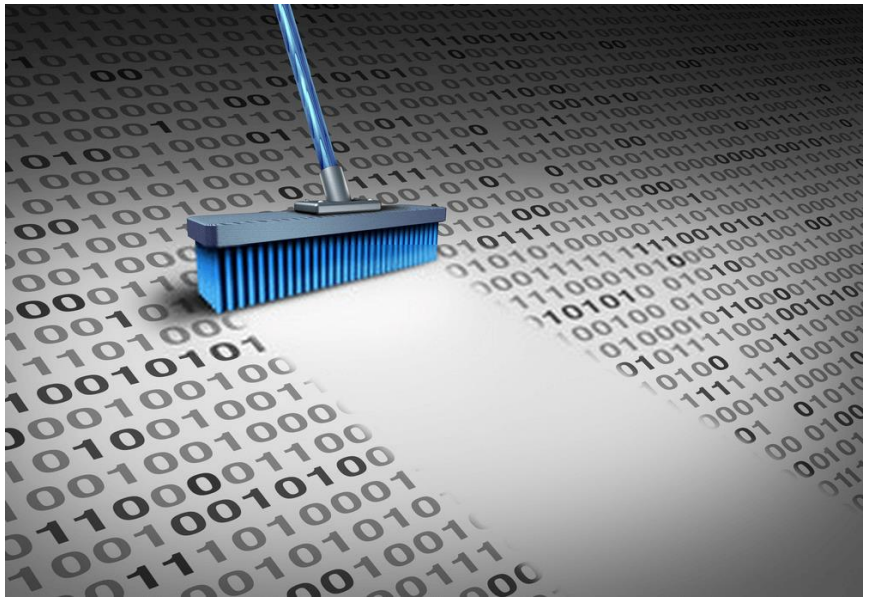
Preparing and Cleaning Data

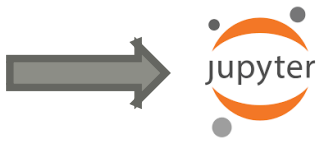
In order to prepare and clean the data, we first loaded the data file (covid-data.csv) into Jupyter Notebook.

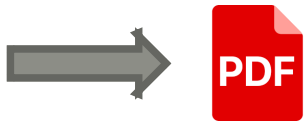
Then we isolated the columns containing the data necessary to answer the defined questions.

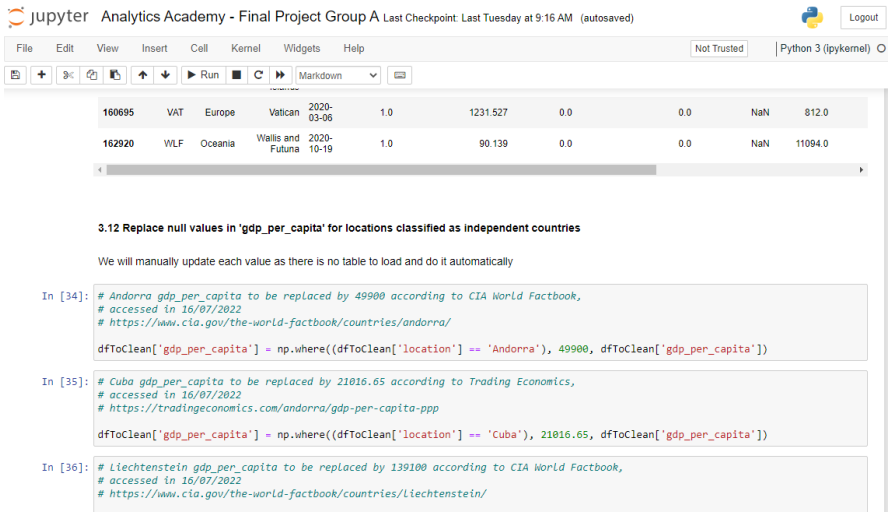
After, we proceeded with data cleaning using various techniques.

The entire process can be found in detail here:








jupyter Analytics Academy - Final Project Group A Last Checkpoint: Last Tuesday at 9:16 AM (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

Run Markdown

160695	VAT	Europe	Vatican	2020-03-06	1.0	1231.527	0.0	0.0	NaN	812.0
162920	WLF	Oceania	Wallis and Futuna	2020-10-19	1.0	90.139	0.0	0.0	NaN	11094.0

3.12 Replace null values in 'gdp_per_capita' for locations classified as independent countries

We will manually update each value as there is no table to load and do it automatically

In [34]:
```
# Andorra gdp_per_capita to be replaced by 49900 according to CIA World Factbook,
# accessed in 16/07/2022
# https://www.cia.gov/the-world-factbook/countries/andorra/

dfToClean['gdp_per_capita'] = np.where((dfToClean['location'] == 'Andorra'), 49900, dfToClean['gdp_per_capita'])
```

In [35]:
```
# Cuba gdp_per_capita to be replaced by 21016.65 according to Trading Economics,
# accessed in 16/07/2022
# https://tradingeconomics.com/andorra/gdp-per-capita-ppp

dfToClean['gdp_per_capita'] = np.where((dfToClean['location'] == 'Cuba'), 21016.65, dfToClean['gdp_per_capita'])
```

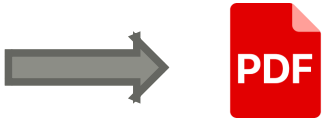
In [36]:
```
# Liechtenstein gdp_per_capita to be replaced by 139100 according to CIA World Factbook,
# accessed in 16/07/2022
# https://www.cia.gov/the-world-factbook/countries/liechtenstein/
```

Exploratory Analysis and Visualization

Targeting the interpretation of the available data and to reveal important aspects, we loaded the sets in Tableau.

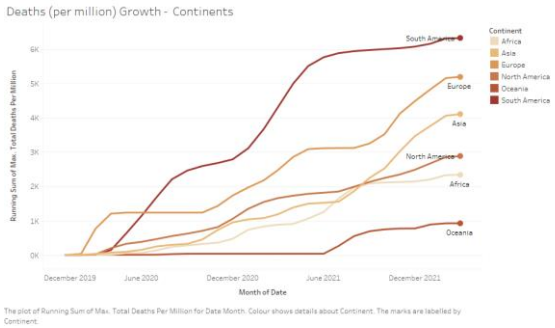
Chart types used:
- Bar charts
- Line charts
- Scatter Plots
- Heat Maps

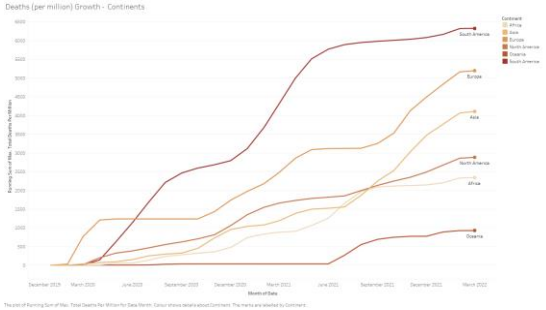
Tableau file with Dashboards and Story Telling can be found here:



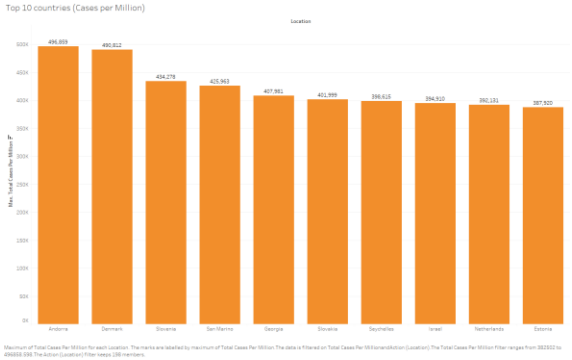
Answering Questions (1)

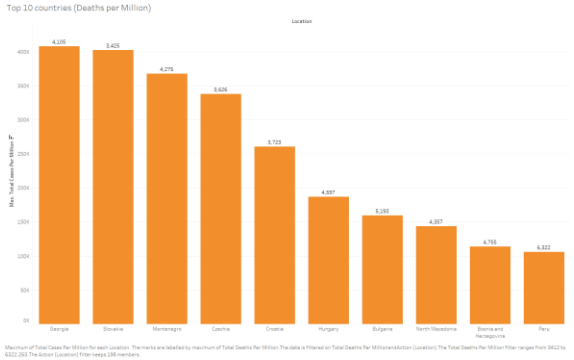
R1 - The evolution by continent of Covid-19 cases per million inhabitants and deaths represented by the trend graph, shows that Europe, Asia and Africa are on the list in cases and South America, Europe and Asia are at the top in terms of deaths.





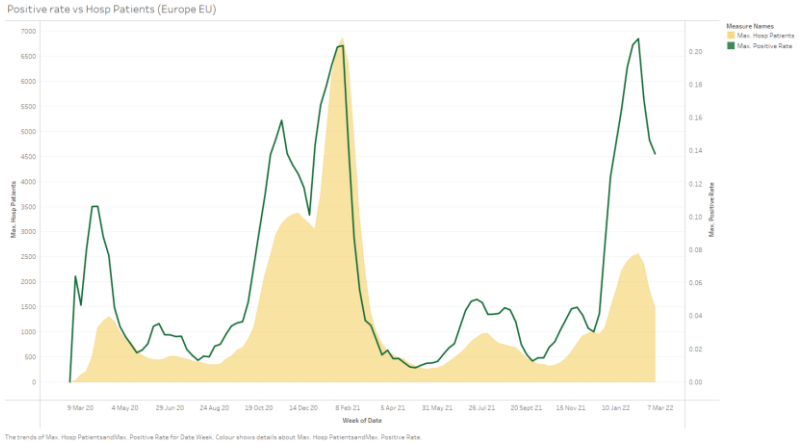
R2 - The top 10 countries, regarding both cases and deaths per million inhabitants, are represented using bar charts. Central and North European Countries are highlighted in this representation.



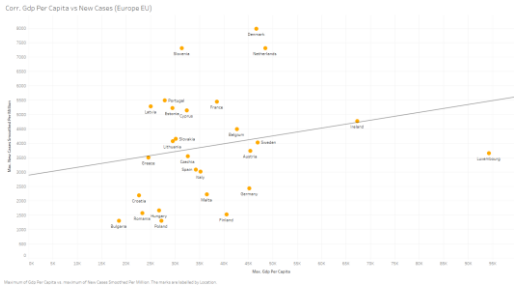


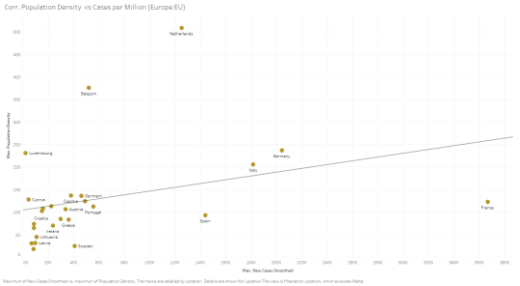
Answering Questions (2)

R3 - The line and area graph for the European Union Countries shows that the peaks of positive cases are coincident with the peaks of hospitalized patients, justifying the overload of these services during the times when there were more cases.



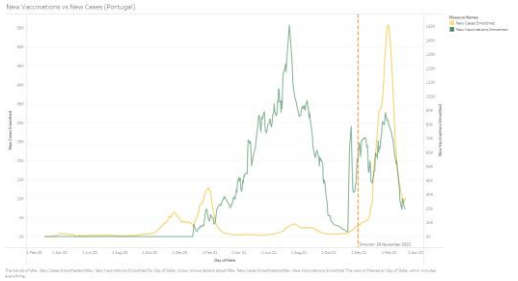
R4 - Checking the correlation (GDP per capita vs New Cases) and (Population Density vs New Cases) using scatter plot together with linear regression, we achieve on both cases as a result, a low R-Squared (near zero) and a high P-Value (>0.05) meaning the correlation is weak and evidence is not strong enough to suggest an effect exists in the population.
In conclusion, both GDP per capita and Population Density in the European Union, are not factors influencing Covid-19 new cases raise.



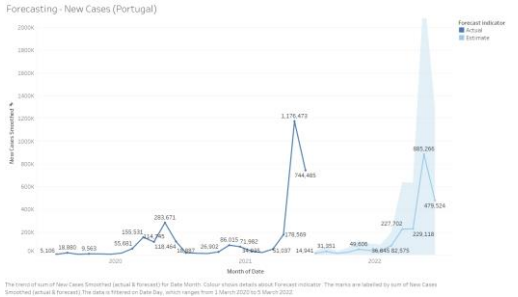


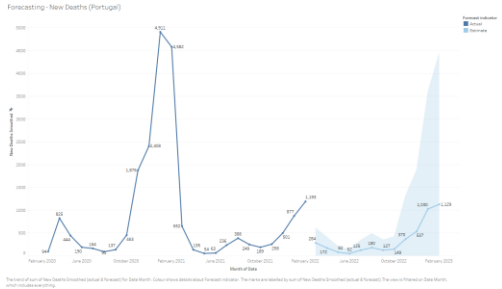
Answering Questions (3)

R5 - Accessing the relation between New Vaccinations and New Cases in Portugal using a line chart, it is possible to observe that although the first vaccination peak is coincident with a time of few hight new cases, the same is not true after the appearance of the Omicron variant. On the other hand, comparing the evolution of vaccination with the emergence of new deaths, it becomes evident that the increase in vaccination had a great influence on controlling the number of deaths.



R6 – Using run charts for both new cases and new deaths in Portugal, adding the Forecast function from Tableau, it is possible to predict the future for 1 year. For both new cases and new deaths, we can predict a sharp decline in March 2022, with a new rise at the end of the year around October and November, culminating in identical numbers of deaths in February 2022. Regarding cases, this rise but did not reach the pre-forecast values.





Project Report

Final Project – Covid 19 Data Analysis

Group A

Alexandre Santana
João Tiago Dias
Paulo Sousa

Online Program in Big Data Talent Discovery and Upskilling

Data Analytics and Data Science

Instructor:

Vala Rohani

July 2022

Analytics Academy of Data Corner in collaboration with Polytechnique Institute of Setubal
Sponsored by Santander Portuguese Universities

Thank you for watching!