

aiHackCovid:

*an exploratory ML analysis of Covid-19 epidemiological
and socio-political data.*

27 June 2021

Team 2 Presentation



Athanasis Trantafyllou

Artificial Intelligence student,
Rijksuniversiteit Groningen



Bharath Raj

Artificial Intelligence student,
Rijksuniversiteit Groningen



Rossella Damiano

Biomedical engineering
student, Politecnico of Milan
(Italy)

GOALS:



Explore the policies adopted by the different European countries, trying to highlight similarities and patterns among them



Develop a predictive model able to predict the number of deaths on N-days in future

BBC

Covid-19: UK in early stages of third wave - scientist

On Monday, the UK reported more than 3,000 new Covid infections for ... Prof Gupta told BBC Radio 4's Today programme the UK was already in a third wave ... We know this has been a difficult pandemic, a dynamic situation.



DATA: exploratory analysis

From the **OxCGRt dataset**:

- C2 index (workplace closing)
 - C3 index (cancellation of public events)
 - C7 index (restrictions on internal movement)
-
- H4 index (emergency investment in healthcare)

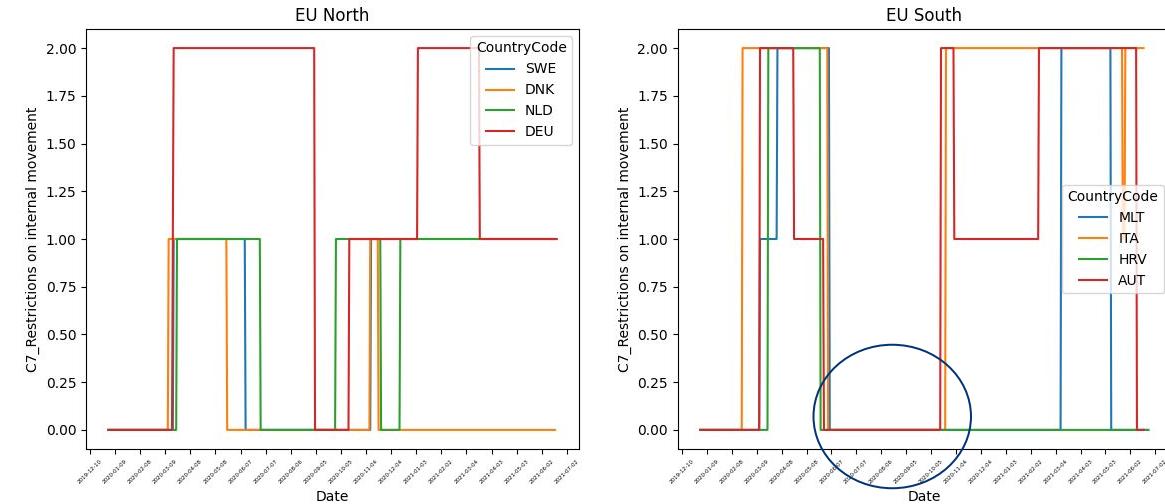


To look at how countries have managed choices with an high social impact



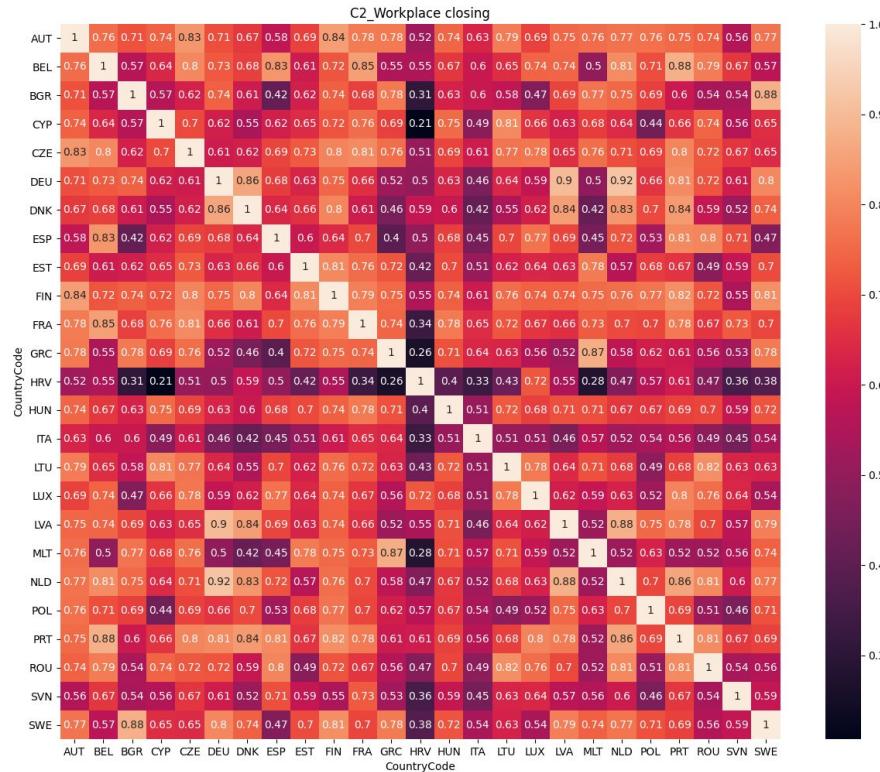
To look at how countries are providing resources to healthcare

RESULTS: C7 index



- North European countries (excepting for Germany) have not adopted very stringent restrictions on internal movements.
- South European countries have adopted prolonged and maximal level of restrictions, with the exception of summer months.

RESULTS: C2 index

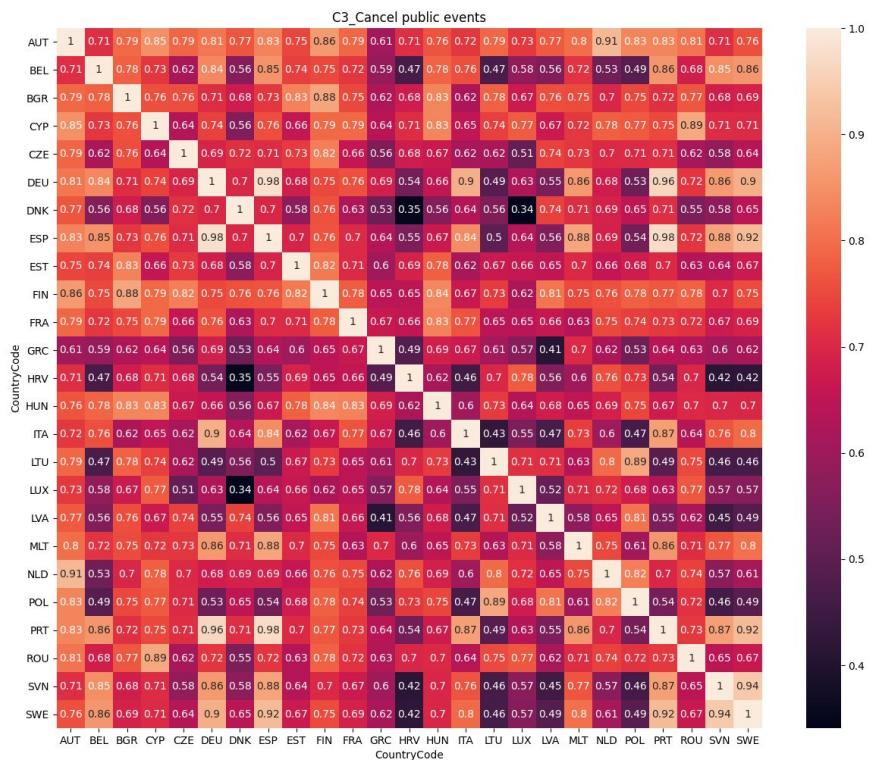


➤ Strongest and longer lasting workplace closures have been adopted in Germany, Italy, Greece and France.

BUT HOW MUCH ARE THESE MEASURES SIMILAR TO EACH OTHER?

➤ The strongest correlation is between Netherlands and Germany (0.92). The weakest one is between Croatia and Cyprus (0.21).

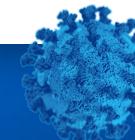
RESULTS: C3 index



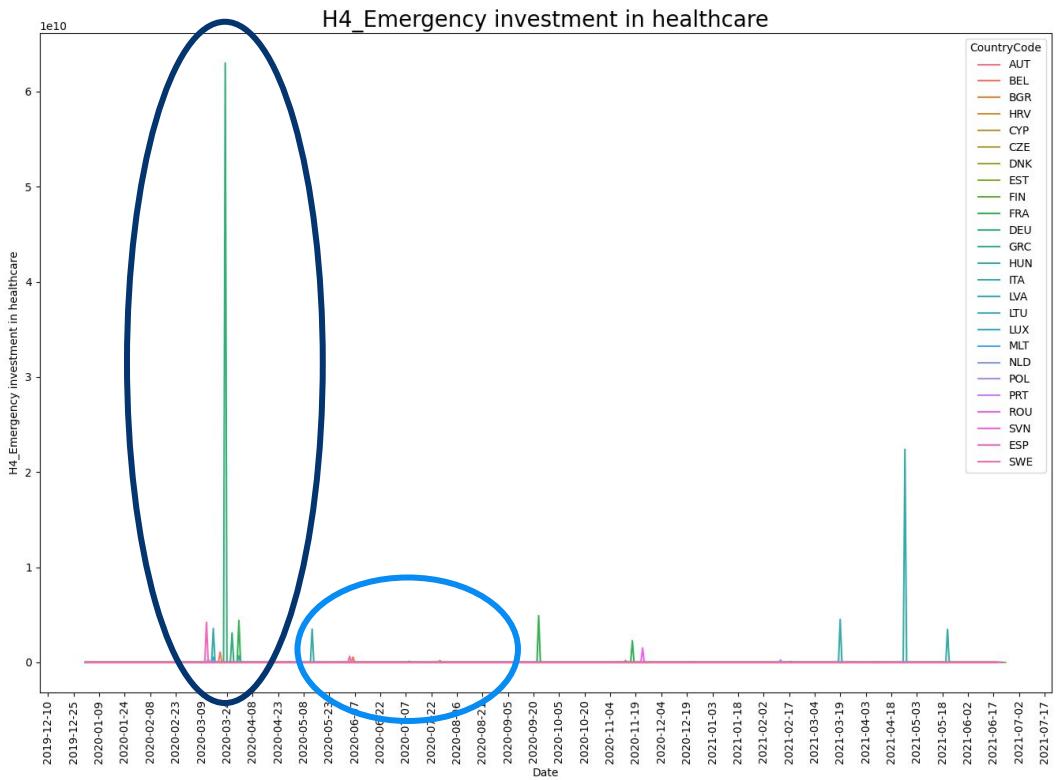
Only Netherlands, Croatia and Greece had removed restrictions during the summer months of the last year.



The majority of the countries maintained cancellation of public events during the whole pandemic period.



RESULTS: H4 index



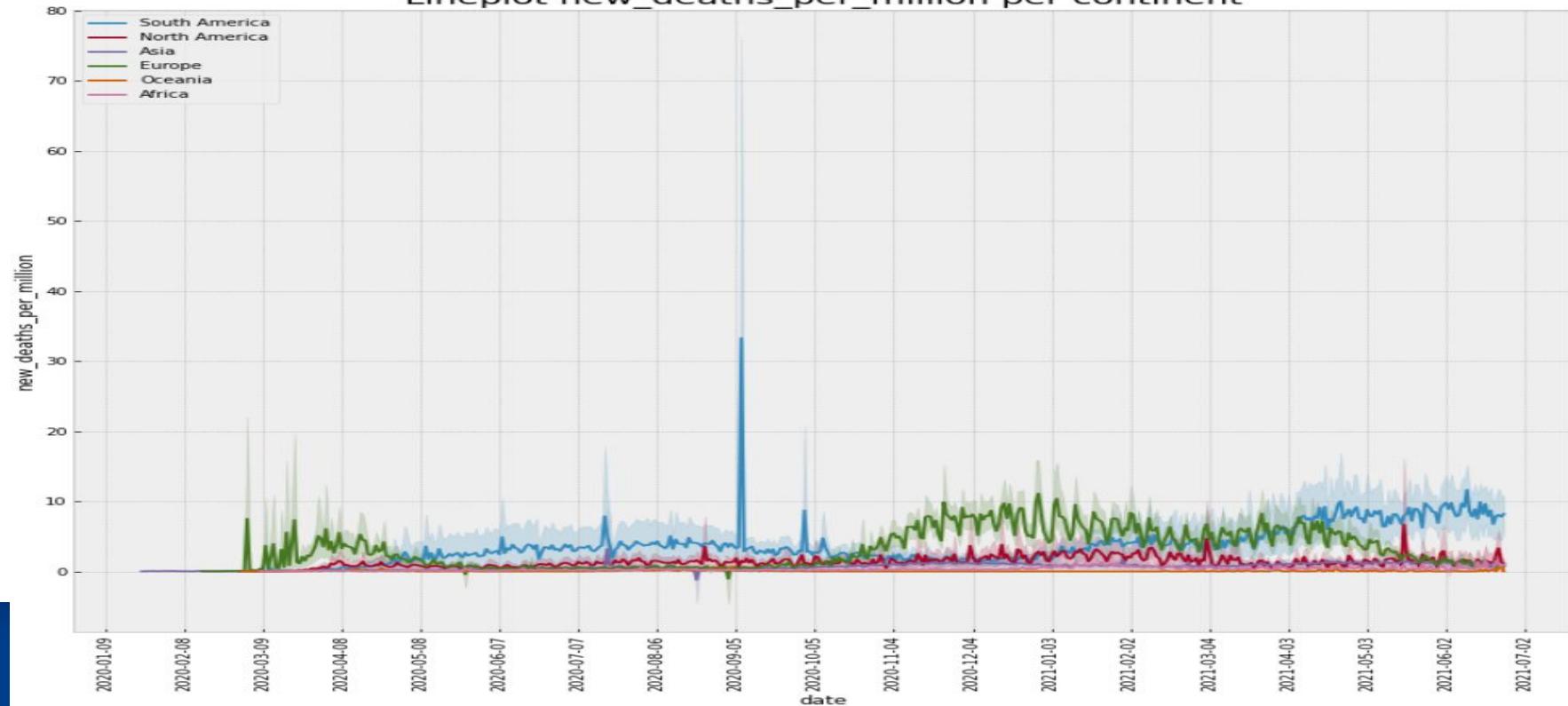
- Not all the countries have done emergency investment in healthcare.
- Higher investments were done during the first pandemic wave.
- During the summer period, when COVID19 cases were decreasing, almost no investment in any country was done.

DATA: predictive model

➤ Owid Covid-19 dataset

Observing the global scenario

Lineplot new_deaths_per_million per continent



DATA: predictive model

- Owid Covid-19 dataset
- **Predictive model** characterized by:

Inputs variables

Stringency score	It is the current stringency score.
Vaccination ratio	Total people vaccinated (at least one dose)/Total population in 2020
Reproduction rate	It is the current reproduction rate. If it is <1, the pandemic has subsided towards an endemic.
Positive rate	It is the share of COVID19 that are positive, giving as a rolling 7-day average

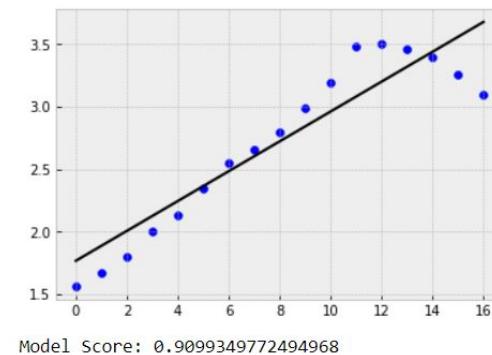
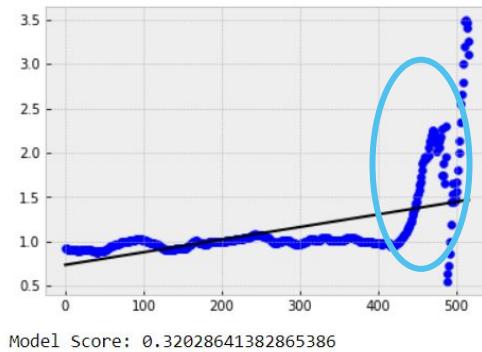
Predictable variable:

Nº of deaths per million	
---------------------------------	--

METHODS: preprocessing

1. Problem of **missing data** into the dataset:
 - put equal to the n° of deaths until 22-01-2020
 - put equal to 0 the n° of persons vaccinated before the starting of vaccinations
 - put equal to 0 the positive rate until 01-01-2020
2. **Normalization** of data
3. Handling the **reproduction rate**:

Regression model fit over the entire dataset.



Regression model fit from 500 days back from 23-06-2021.

References to [1], [2], [3].

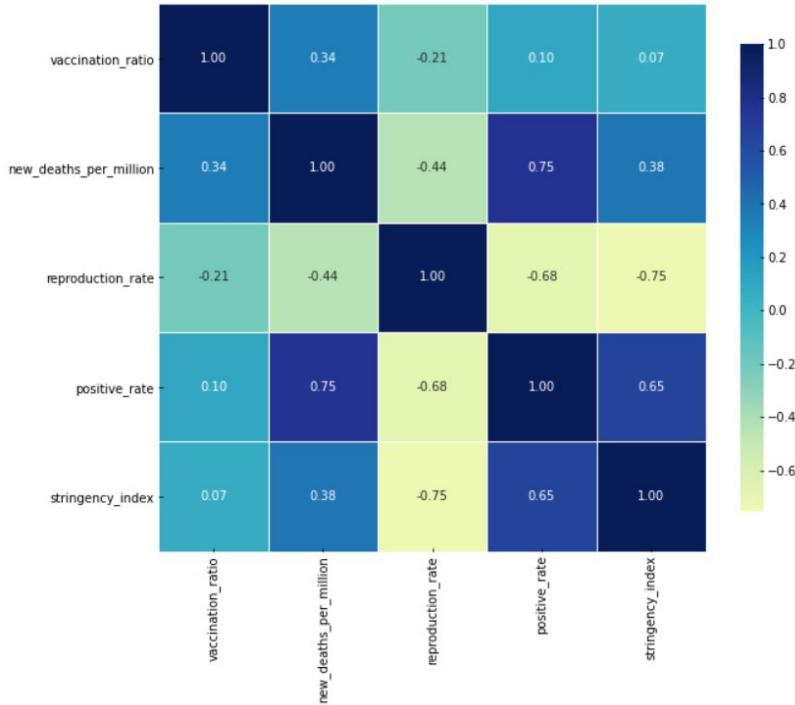
METHODS: exploratory analysis

Scaled Data Overview



Scaled data graph which shows how the different features evolve through time.

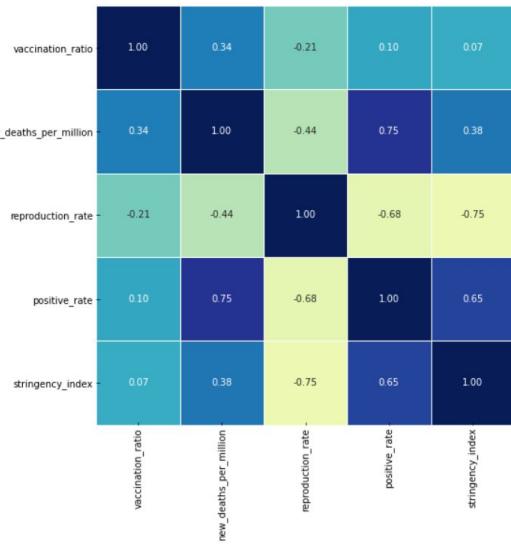
METHODS: exploratory analysis



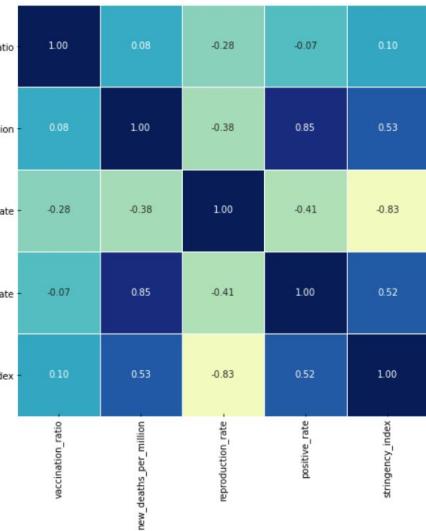
- Strong positive correlation between the positive rate and the new deaths per million (0.75).
- Strong negative correlation between the reproduction rate and the stringency index. (-0.75).
- Strong positive correlation between the stringency index and the positive rate (0.65).
- Strong negative correlation between the reproduction rate and the positive rate (-0.68).

METHODS: exploratory analysis

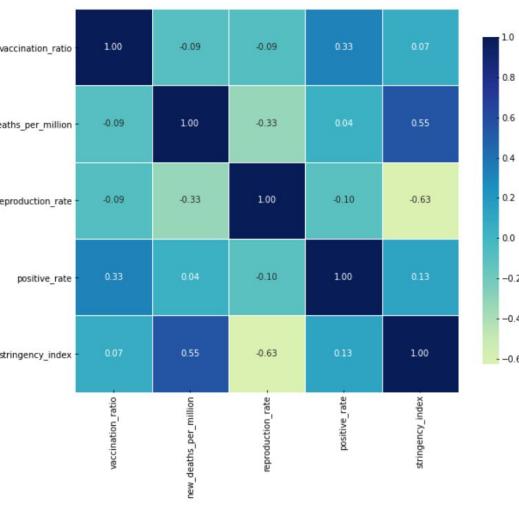
Correlation Matrices



a) Global



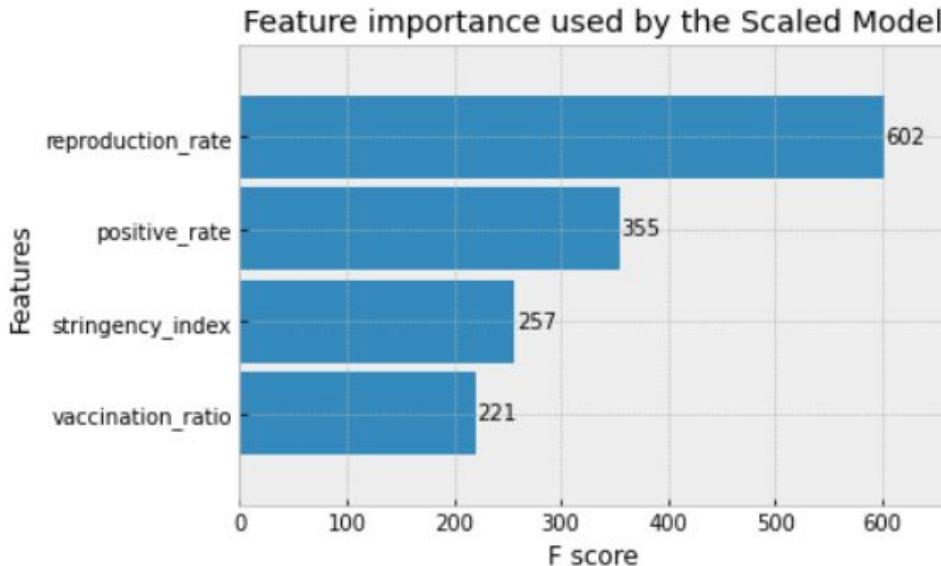
b) EU



c) NL

RESULTS: predictive model

We use the XGBOOST tree based regressor model for predictions



The **reproduction rate** plays a significant role in determining the number of new deaths per million.

THIS CAN'T BE APPRECIATED BY SIMPLY LOOKING AT THE DATA!



RESULTS: predictive model

Prediction of model which used scaled data



Global

METHODS: Synthetic data projection

Vaccination ratio: Approach- linear regression model.

Reason - This has to keep increasing because of all the vaccination drives happening in the world. So the max value for vaccination ratio is 1

Stringency index: Approach - ARIMA / linear regression model

Reason - This can be projected using a regression model as we see that it keeps decreasing gradually over a period of time.

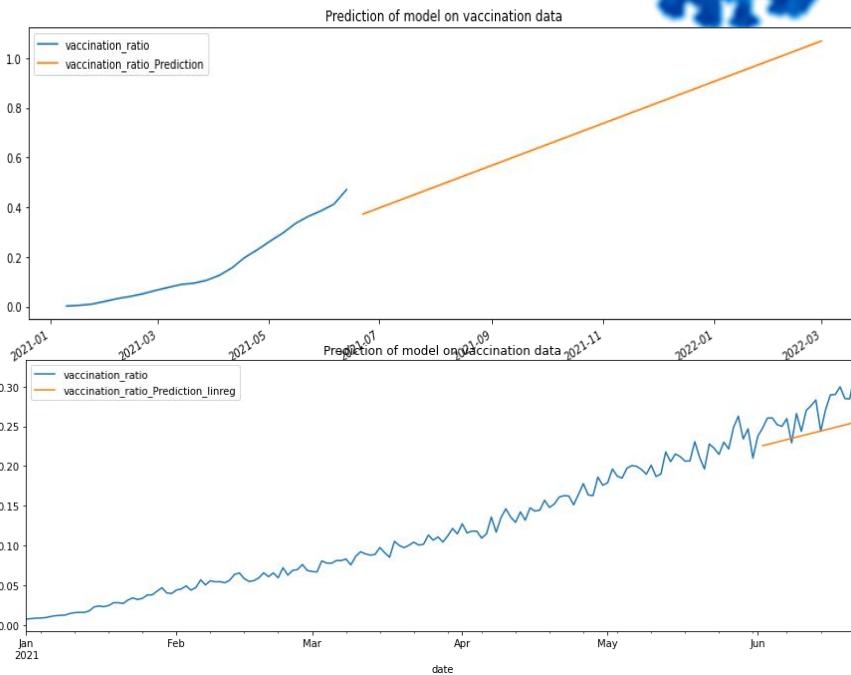
positive rate: Approach - ARIMA / RNN model

Reason - This has got the second highest F-score on the XGboost model and having a good estimate at future point of time(synthetic data) will directly impact our prediction on that future date

positive rate: Approach - ARIMA / RNN model

Reason - This has got the second highest F-score on the XGboost model and having a good estimate at future point of time(synthetic data) will directly impact our prediction on that future date

Projection for NL



Projection Globally

RESULTS: Synthetic data projection

Prediction of model which used unscaled data



- Simple models do not have a memory to keep track of time

What did not work!

- XGBOOST model gave us a baseline inference, now we can focus on more memory based architectures for predicting new_deaths_per_million
- LSTM network captures the data trend for a 30 day window, still for larger time windows the predictions are off.

WHAT'S NEXT?

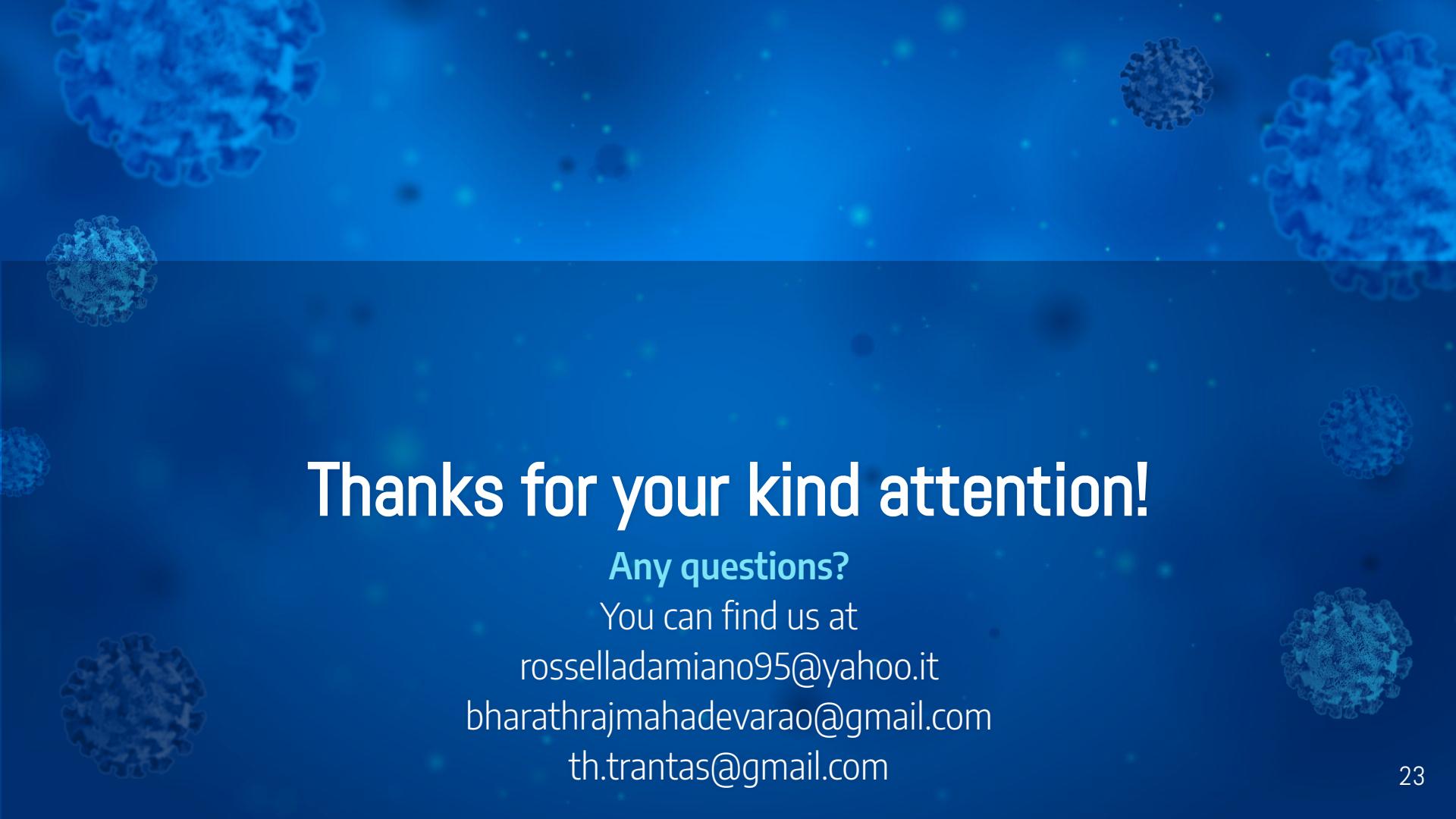
- Explore, quantitatively, the way the different policies adopted by the different countries have influenced over the spread of COVID 19 and over the deaths, with the aim to define the most **efficient strategy**.
- Refine the regression model when data about **vaccination effectiveness** in the population.
- Try better models for the projection.



REFERENCES:

- [1] Wu, J.T., Leung, K., Bushman, M., Kishore, N., Niehus, R., de Salazar, P.M., Cowling, B.J., Lipsitch, M., Leung, G.M.: Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat. Med.* 26, 1–5 (2020) (<https://www-nature-com.proxy-ub.rug.nl/articles/s41591-020-0822-7>)
- [2] Lai, C.C., Shih, T.P., Ko, W.C., Tang, H.J., Hsueh, P.R.: Severe acute respiratory syndrome coronavirus 2 (SARSCoV-2) and corona virus disease-2019 (COVID-19): the epidemic and the challenges. *Int. J. Antimicrob. Agents* 55(3), 105924 (2020) (<https://pubmed.ncbi.nlm.nih.gov/32081636/>)
- [3] Liu, Y., Gayle, A.A., Wilder-Smith, A., Rocklöv, J.: The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J. Travel Med.* 27(2), 1 (2020) (<https://pubmed.ncbi.nlm.nih.gov/32052846/>)



The background of the slide features a dark blue gradient with several stylized, light blue COVID-19 virus particles scattered across it, some in the foreground and others in the background.

Thanks for your kind attention!

Any questions?

You can find us at
rosselladamiano95@yahoo.it
bharathrajmahadevarao@gmail.com
th.trantas@gmail.com

Credits

Special thanks to all the people who made and released these awesome resources for free:

- ▶ Presentation template by [SlidesCarnival](#)
- ▶ Photographs by [Unsplash](#)
- ▶ [COVID-19 virus image](#) from CDC by Alissa Eckert & Dan Higgins · Public domain and thus free of any copyright restrictions