

---

# Interstate infection interaction

An analytic look at CoViD-19 interactions in  
the Euregio Maas-Rhine area

Florian Schweitzer  
18.10.2021

---

# CRISP-DM - Project Understanding

## Project

Euregio Maas-Rhine area with 4M people  
(Aachen, Liege, Belgian-Limburg, Dutch-Limburg, Eastern Belgium)

State/county based look at infection rates  
and development

The people of the EMR have close knit interactions for

- Business
- Leisure



<https://euregio-mr.info/>

# CRISP-DM - Project Understanding

---

## Situation Assessment

Ex-Post analysis

Rules dictated from elsewhere (State/EU)

## Hypothesis:

Travel spreads CoViD-19 and we can show a correlation between certain impactful factors and infection spike events

- ~~Essential travel~~
- Borders open/closed (Non-essential work commute)
- Leisure commute

## Resources

- Infection data
- Regulation/holiday timelines

# CRISP-DM - Project Understanding

---

## **Goals of the Data Mining**

Assess the impact of different factors to state and EMR

Find the parameters under which the hypothesis for a correlation can be proven/rejected

If the hypothesis can be proven, make a classification of the infection spike events and attempt a prediction of future trends

# CRISP-DM - Project Understanding

---

## **Project Plan**

Construct infection timelines

- State based/EMR based
- Use/Create a measure that is independent from baseline

Classify infection events

- Create factor timelines
- Check correlation/Silhouette
- Find/Create a feature for dependency

Apply predictions

- Get a prediction of change in infections based on the focussed factors

# CRISP-DM - Data Description

## Initial Data Description

The CoViD-19 data provided by the ECDC contains daily values of the 14-day notification rate categorized by country and county.

Can the source of a spike be identified amongst the focus-factors?

Can a prediction be made for expected infection rate change?

The dataset used is provided in UTF-8 encoded JSON format and additionally contains the sources of the respective data. Germany's and the Netherlands data was reported via TESSy, Belgian data was retrieved from their website.

The date is a String in YYYY-MM-dd format  
The 14 day rate is a floating point number  
The other values are categorical

```
{  
  "country" : "Germany",  
  "region_name" : "Städteregion Aachen",  
  "nuts_code" : "DEA2D",  
  "date" : "2021-10-10",  
  "rate_14_day_per_100k" : 83.4790476566624,  
  "source" : "TESSy COVID-19, subnational daily data"  
},
```

# CRISP-DM - Data Description

---

## Initial Data Description continued...

Data for holidays and vacations can be acquired from multiple places and be aggregated from holidays and vacations.  
Weekends need to be added to get accurate start/end days.

## Roadmap

Until Dec 6th I aim to have the

- Data Understanding and Preparation done
- Modeling Technique chosen based on the findings

After Dec 6th

- Implementation and evaluation the of model
- Evaluation of results

# Sources

---

- Homepage der EMR  
<https://euregio-mr.info/>
- Border situation/regulations  
<https://itemcrossborderportal.maastrichtuniversity.nl>
- Border regulations  
<https://ec.europa.eu/>
- RKI CoViD-19 Datahub  
<https://npgeo-corona-npgeo-de.hub.arcgis.com/>
- ECDC  
<https://www.ecdc.europa.eu/>
- Holidays  
<https://holidaycalendar.com/>