

DATA MINING PROJECT

COVID-19 and Vaccination

Görkem Deniz SEYHAN 217CS2013

1) DATA

- I will use Data on COVID-19 Vaccination in The EU/EEA. This dataset represent which country using which vaccine and number of vaccinations per week. There is a .csv formatted version of this data in the following link:
<https://www.kaggle.com/arashnic/data-on-covid19-vaccination-in-the-eueea>
- I will use a similar data set as an additional source.
https://www.kaggle.com/gpreda/covid-world-vaccination-progress?select=country_vaccinations.csv
- I will also use covid-19 data besides vaccine data. Death and case numbers are data that will be useful for me.
https://www.kaggle.com/josephassaker/covid19-global-dataset?select=worldometer_coronavirus_daily_data.csv

And

<https://www.kaggle.com/tunguz/the-our-world-in-data-covid-vaccination-data>

2) Research Question and Hypothesis

- Which vaccine is more preferred?
- Which EU country has the highest covid-19 population?
- Which EU country has the most vaccinated population?
- Which EU country has the lowest case / death rate?
- Hypothesis: As vaccination increases, the number of deaths decreases.

3) Experiments

- I will find out which vaccines are used in which country and how much.
- I will find out how many people each vaccine has been given in total.
- I will find the sum of deaths and cases for each country.
- I will look for a relationship between the number of vaccinations and the number of deaths.
- I will estimate the vaccination deadline for the countries.
- I will estimate the number of cases in the future.