# Documentatie

## Applicatie vereisten

1. De file moet automatisch via python gedownload en verder automatisch in een eigen database ingelezen worden.
2. Deze vier landen moeten worden meegenomen en de rest moet worden uitgesloten :
   1. Suriname,
   2. Guyana,
   3. Frans-Guyana &
   4. Brazilië.
3. De volgende kolommen moet niet worden meegenomen:
   1. icu\_patients
   2. icu\_patients\_per\_million
   3. hosp\_patients
   4. hosp\_patients\_per\_million
   5. weekly\_icu\_admissions
   6. weekly\_icu\_admissions\_per\_million
   7. weekly\_hosp\_admissions
   8. weekly\_hosp\_admissions\_per\_million
   9. new\_tests
   10. total\_tests
   11. total\_tests\_per\_thousand
   12. new\_tests\_per\_thousand
   13. new\_tests\_smoothed
   14. new\_tests\_smoothed\_per\_thousand
   15. positive\_rate
   16. tests\_per\_case
   17. tests\_units
4. De data moet dagelijks ingelezen worden
5. Er moet een dashboard van de data gemaakt worden in Power BI met de volgende info:
   1. Aantal vaccinaties (dagelijks)
   2. Aantal doden (dagelijks, maand)
   3. Aantal testen (dagelijks)
   4. Aantal positieve gevallen (dagelijks)

## Beperkingen

Tijdens het inlezen van de data uit <https://covid.ourworldindata.org/data/owid-covid-data.csv> zijn er enkele beperkingen opgetreden:

1. De data die van vier landen getrokken moet worden is incompleet. Van de 4 gevraagde landen (Suriname, Guyana, Frans-Guyana & Brazilië) komt Frans-Guyana niet voor in de bron
2. Van de bron data zijn er namelijk 66 kolommen (zie bijlagen) waarvan er uit deze 17 niet nodig zijn voor het dashboard en worden uitgesloten. Wat betekent dat er voor tenminste 39 kolommen een aparte date type gedefinieerd moet worden in de database.
3. Bij het creeren van een dashboard is niet aangegeven als de data uit de eigen database gelezen moet worden of rechtstreeks vanuit de github repository
4. Hoewel 'new\_tests' niet nodig was voor de initiële vereisten, is het nu nodig voor het maken van een dagelijks dashboard van het aantal tests

## Oplossing

De 4 landen criteria wordt meegnomen in de script. Als de gegevens van Frans-Guyana later beschikbaar komen, worden deze toegevoegd aan de database.

In plaats van het uitsluiten van de 17 genoemde kolommen, is besloten om alleen de kolommen mee te nemen die nodig zijn voor het dashboard, samen met extra kolommen om een onderscheid te maken tussen land en tijd. Omdat dit slechts een beperkt aantal kolommen betreft, heeft het toewijzen van de datatypen hieraan minder tijd gekost. Deze zijn:

|  |  |  |
| --- | --- | --- |
| Kolom | Omschrijving | Datatype |
| iso\_code | de 3 landen iso code | varchar |
| location | land | varchar |
| date | datum | date |
| new\_cases | nieuwe gevallen | integer |
| new\_deaths | aantal doden | integer |
| new\_vaccinations | aantal vaccinaties | integer |
| new\_tests | aantal testen | integer |

## Architectuur

* Database : mysql   
  Database naam : covid\_data
* Python versie 3.11
* Python Libraries pandas , mysql.connector

Installatie instructies

1. Copieer de script download.py naar gewenste locatie
2. Verander de database gegevens op regels 13-16
   * host
   * user
   * password
   * database

Deze zelfde gegevens zijn nodig voor de Power Bi Dashboard

1. Via de command prompt ga je in de folder waar de script zich bevindt
2. Run script met de volgende commando : python download.py
3. Volg de scherm instructies

## Bijlagen

1. Kolommen van <https://covid.ourworldindata.org/data/owid-covid-data.csv> :  
     
   1. iso\_code
   2. continent
   3. location
   4. date
   5. total\_cases
   6. new\_cases
   7. new\_cases\_smoothed
   8. total\_deaths
   9. new\_deaths
   10. new\_deaths\_smoothed
   11. total\_cases\_per\_million
   12. new\_cases\_per\_million
   13. new\_cases\_smoothed\_per\_million
   14. total\_deaths\_per\_million
   15. new\_deaths\_per\_million
   16. new\_deaths\_smoothed\_per\_million
   17. reproduction\_rate
   18. icu\_patients
   19. icu\_patients\_per\_million
   20. hosp\_patients
   21. hosp\_patients\_per\_million
   22. weekly\_icu\_admissions
   23. weekly\_icu\_admissions\_per\_million
   24. weekly\_hosp\_admissions
   25. weekly\_hosp\_admissions\_per\_million
   26. total\_tests
   27. new\_tests
   28. total\_tests\_per\_thousand
   29. new\_tests\_per\_thousand
   30. new\_tests\_smoothed
   31. new\_tests\_smoothed\_per\_thousand
   32. positive\_rate
   33. tests\_per\_case
   34. tests\_units total\_vaccinations
   35. people\_vaccinated
   36. people\_fully\_vaccinated
   37. total\_boosters
   38. new\_vaccinations
   39. new\_vaccinations\_smoothed
   40. total\_vaccinations\_per\_hundred
   41. people\_vaccinated\_per\_hundred
   42. people\_fully\_vaccinated\_per\_hundred
   43. total\_boosters\_per\_hundred
   44. new\_vaccinations\_smoothed\_per\_million
   45. new\_people\_vaccinated\_smoothed
   46. new\_people\_vaccinated\_smoothed\_per\_hundred
   47. stringency\_index
   48. population\_density
   49. median\_age
   50. aged\_65\_older
   51. aged\_70\_older
   52. gdp\_per\_capita
   53. extreme\_poverty
   54. cardiovasc\_death\_rate
   55. diabetes\_prevalence
   56. female\_smokers
   57. male\_smokers
   58. handwashing\_facilities
   59. hospital\_beds\_per\_thousand
   60. life\_expectancy
   61. human\_development\_index
   62. population
   63. excess\_mortality\_cumulative\_absolute
   64. excess\_mortality\_cumulative
   65. excess\_mortality
   66. excess\_mortality\_cumulative\_per\_million