Iron Hack
Data Analyze project

Mathieu Guarino

Goals

Use 2 datasets (API and CSV) to push data into cloud SQL database

Get merged data and do some stats with the data

1.1 API Data

From GHO API we get 2 dataframes

➔ **Country Dataframe**

- Country code (AGO)
- Country name (Angola)
- ParentDimension (REGION)
- Dimension (COUNTRY)
- ParentCode (AFR)
- ParentTitle (Africa).

➔ **Doctors Dataframe**

- SpatialDim (AGO)
- NumericValue (2146)

1.2 CSV File

We have one dataframe

➔ **Estimated_numbers**

- Country (Afghanistan)
- Year (2017)
- ….
- …
- No of deaths median
- No of case median
- …

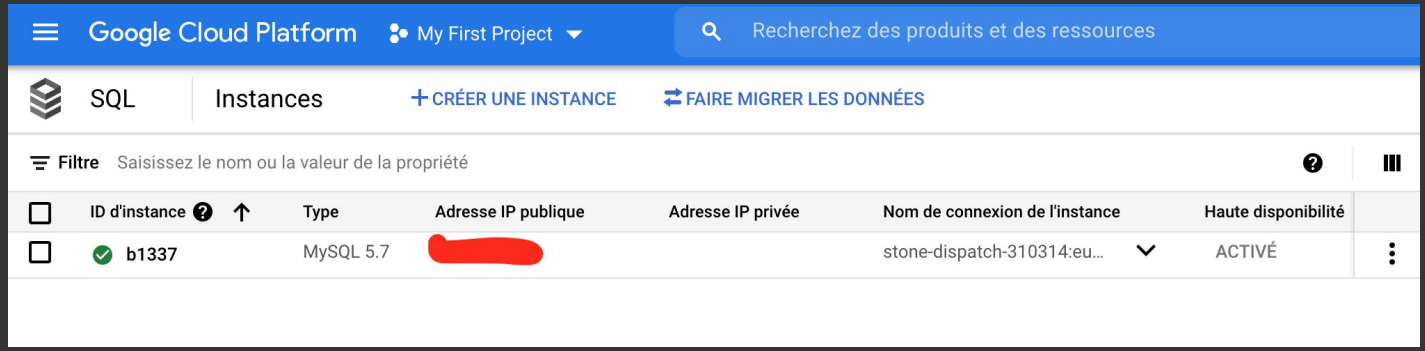
2. SQL

- ➔ **SQL Cloud connexion**
  Tutorial and documentation
- ➔ **Creating 2 Tables with the right data types**
- ➔ **Inserting data into tables**
- ➔ **Get and merge data from tables**

How to connect with cloud SQL ?

https://towardsdatascience.com/sql-on-the-cloud-with-python-c08a30807661

I first created an account on google and downloaded the 3 .pem files needed to connect with python. We also get the IP adress needed.
I created an account with a password to the database.

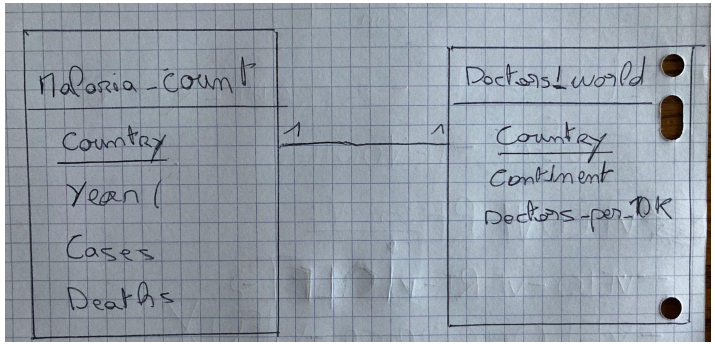


Connexion : SQL with Python

```python
import mysql.connector
from mysql.connector.constants import ClientFlag

config = {
    'user': 'root',
    'password': 'Password123',
    'host': '94.944.94.94',
    'client_flags': [ClientFlag.SSL],
    'ssl_ca': 'ssl/server-ca.pem',
    'ssl_cert': 'ssl/client-cert.pem',
    'ssl_key': 'ssl/client-key.pem'
}

# now we establish our connection
cnxn = mysql.connector.connect(**config)
```



Creating tables : What do we need ?

- Clean data
- Data types

The first table for doctors contains :

- Countries (VARCHAR)
- Continents (VARCHAR)
- Doctors_per_10k (FLOAT)

The second table for sick people contains :

- Countries (VARCHAR)
- Year (SMALLINT)
- Cases (BIGINT)
- Deaths (BIGINT)

```
CREATE TABLE malaria_count
(‘Country VARCHAR(255),’
‘Year SMALLINT,’
‘Cases BIGINT,’
‘Deaths BIGINT )’)
```

Hint

I had to make some tests before getting the right data format and data so I cleaned my table with :

'TRUNCATE TABLE malaria_count'

Inserting data : What do we need ?

- Cleaned dataframes
- SQL Query

"INSERT INTO malaria_count (Country, Year, Cases, Deaths) VALUES (%s, %s, %s, %s)")

Get and merge data

We need to identify the common value in both tables, it is “Country” :

```
SELECT * FROM doctors_world INNER JOINT
malaria_count ON doctors_world.Country =
malaria_count.Country
```

Hint

We need to drop unused columns and rename them to have clean data to use

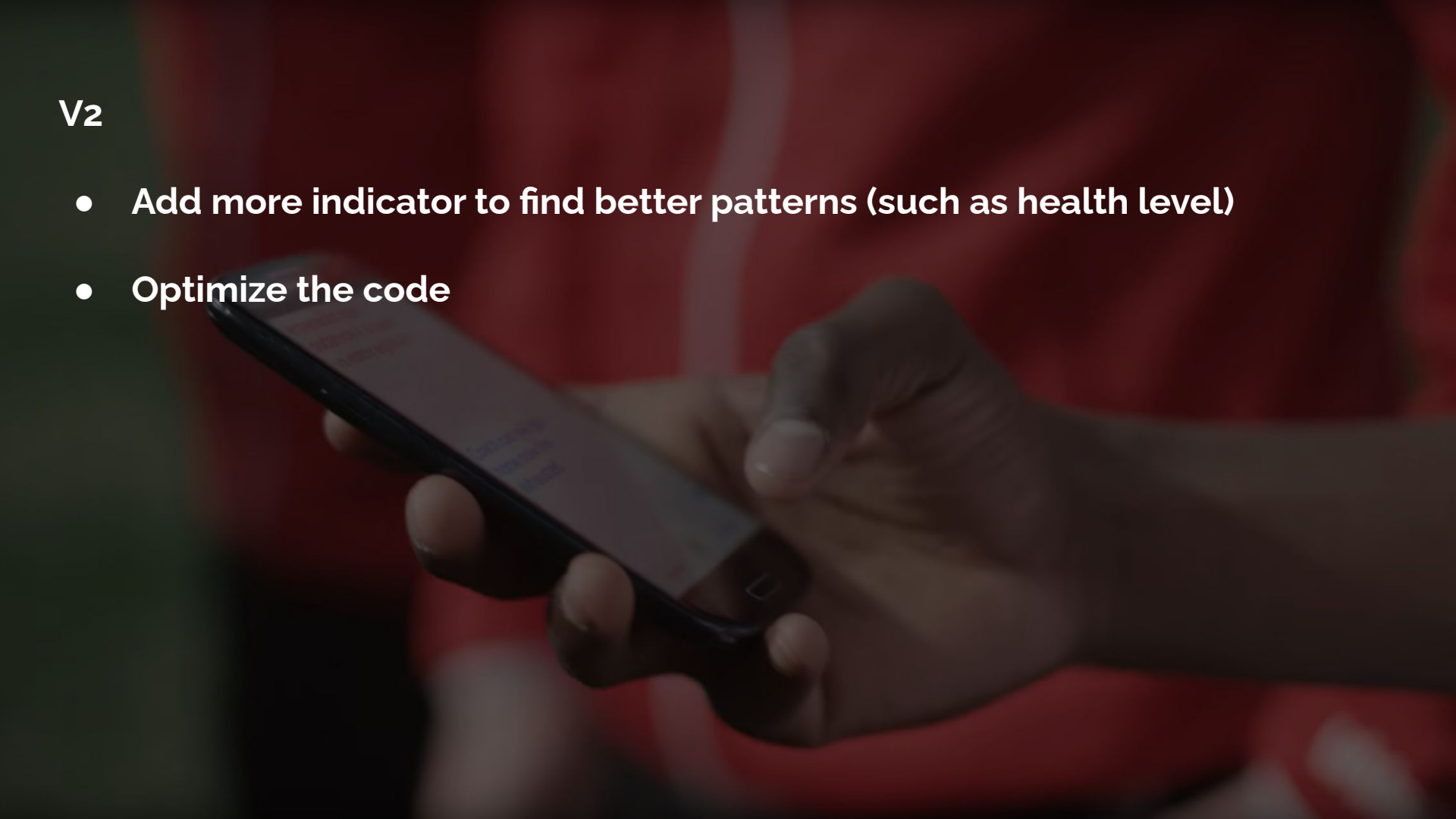
Histoire imaginée à titre d'illustration uniquement

3. Find an answer

- ➔ The number of doctors is not the only link with the sick people
  - ◆ Western Pacific and Africa has a low % of doctors but Western Pacific has a low amount of sick people

Challenges ?

- Find the right and usable source of data : it took me a lot of time !
- Find common data between the two source to find patterns
- Use an online SQL cloud
- I had not enough time to show the multiple indicators I have selected at the start and make deeper analyzes



V2

- Add more indicator to find better patterns (such as health level)
- Optimize the code