



COVID19 SYMPTOMS ANALYSIS USING TWITTER DATA

An end-to-end data science project

Naïla EL HAOUARI

09/12/2020

Presentation of this session

Giving you technical tools and basic understanding of how they work on a concrete example: a Twitter analysis of COVID19 symptoms.

→ We want to analyze the reception of the global pandemic on social media, and more specifically, COVID19 symptoms people might mention on Twitter.

Overview of this session	
Tweets collection	How to use the Twitter API and code example
Data preparation	Preprocessing and filtering of textual data
Data visualization	Interactive graphs using plotly
Modelization	Building a machine learning classifier

All ressources and notebooks:

https://github.com/Naila-elh/CRI_BigDataCourse_2020

TWITTER API



Requirements: create a developer's account on <https://developer.twitter.com/en>

Ressources:

- Official API documentation: <https://developer.twitter.com/en/docs/twitter-api>
- Python package *tweepy*: <http://docs.tweepy.org/en/latest/index.html>

Some requests examples:

- Get user timeline: get historical data, up to the 3,200 last tweets of a user (ref: [API](#) / [tweepy](#))
- Streaming: get real-time sampled data (1%) + filter (ref: [API](#) / [tweepy](#))
- Lookup user: get information – screen name, description, profile picture – on a user (ref: [API](#) / [tweepy](#))
- Followers ids: get a user's followers (ref: [API](#) / [tweepy](#))

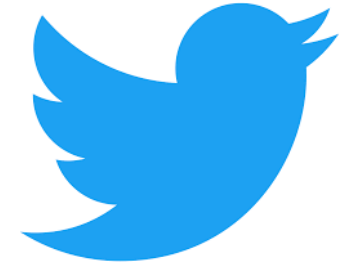
TWITTER API



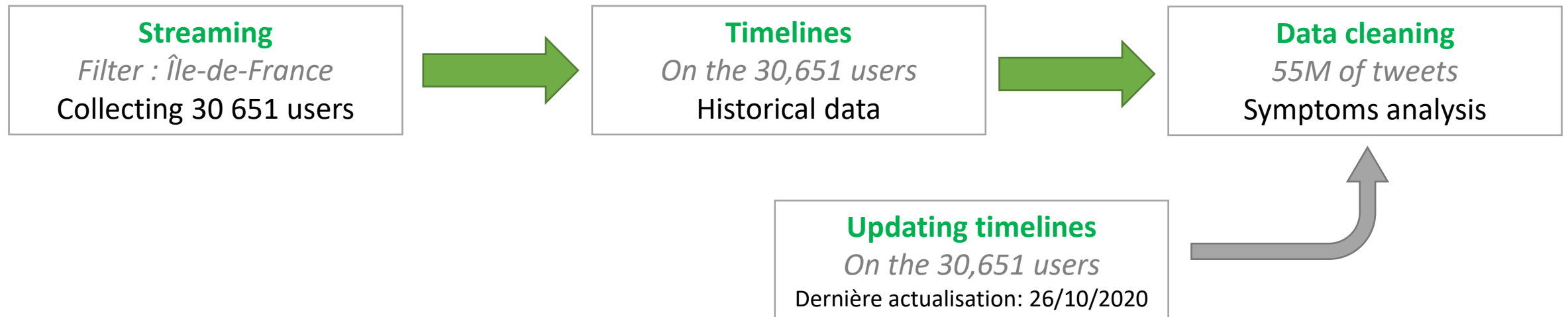
→ Tweets as **json** objects

```
{
  "created_at" : "Thu Apr 06 15:24:15 +0000 2017" ,
  "id_str" : "850006245121695744" ,
  "text" : "1\ Today we\u2019re sharing our vision for the future of the Twitter API platform!\nhttps://t.co/XweGngmxlP"
  "user" : {
    "id" : 2244994945 ,
    "name" : "Twitter Dev" ,
    "screen_name" : "TwitterDev" ,
    "location" : "Internet" ,
    "url" : "https://dev.twitter.com/" ,
    "description" : "Your official source for Twitter Platform news, updates & events. Need technical help? Visit https://t.co/XweGngmxlP"
  } ,
  "place" : {
  } ,
  "entities" : {
    "hashtags" : [
    ] ,
    "urls" : [
      {
        "url" : "https://t.co/XweGngmxlP" ,
        "unwound" : {
          "url" : "https://cards.twitter.com/cards/18ce53wgo4h/3xo1c" ,
          "title" : "Building the Future of the Twitter API Platform"
        }
      }
    ]
  } ,
  "user_mentions" : [
  ]
}
```

TWITTER API



How we collected the tweets



SOME CODE EXAMPLE: [streaming api](#) ; [get timeline api](#)

DATA ANALYSIS

Aim: getting the tweets of people talking about their symptoms.

Preprocessing:

- Remove retweets (half of the dataset)
- Remove mentions (@mention) and url ([url])
- Dictionary of symptoms

Symptoms:

- Cough (toux, toussé)
- Fever (fièvre)
- Sore throat (maux de gorge, mal à la gorge)
- Headache (mal de tête, mal à la tête, mal de crâne)
- Breathing difficulties (difficultés à respirer, difficultés respiratoires, mal à respirer)
- Loss of taste and smell (perte du goût, perte de l'odorat)
- Symptom (symptôme)

HANDS ON! [Notebook 3 - Data analysis and visualization](#)

CITIZEN SCIENCE

Issue: many **false positives** in our tweets mentioning symptoms



Symptom before

« @mention J'ai eu beaucoup de symptômes.. fin Février debut Mars, fin Mars (essoufflement quoi que je fasse) je suis donc allée chez le medecin, "c'est le stress" 🤔 je me reconnais dans ce Tread..' »

NOT a symptom

« Jcrois j'ai trop rigoler ojd j'ai mal à la gorge »

General news

« En cas de fièvre, privilégiez le paracétamol. En effet, les anti-inflammatoires comme l'ibuprofène, de part leur mécanisme d'action, peuvent aggraver l'infection ! 🦠 En cas de doute, demandez conseil à un professionnel de #santé 🧑🏻‍⚕️📺🧑🏻‍⚕️📺🧑🏻‍⚕️ #COVID—19 [url] »