

Année 2021

RAPPORT DE PROJET - ISOC 631

PROJET TWITTER

“ API Twitter cartographie ”

Projet réalisé par

Clément MASSIT
Zied GOBJI

Projet encadré par

Kavé SALAMATIAN
Sorana CIMPAN



POLYTECH[®]
ANNECY-CHAMBÉRY

Remerciements

Tout d'abord, nous tenons à remercier et témoigner notre reconnaissance aux personnes suivantes, pour leur soutien et aide sur l'ensemble du projet malgré le contexte et les moyens de communication restreints.

Kavé SALAMATIAN, Sorana CIMPAN pour l'ensemble des cours et des explications durant notre projet, l'ensemble de ces précieux conseils nous ont apportés de nombreuses compétences et connaissances pour la réalisation de ce projet

Liens importants

L'ensemble des liens suivants sont constitutifs de notre projet, du point de vue développement des fonctionnalités et de la gestion de projets dans la globalité.

Lien vers le projet GitHub : <https://github.com/clement-massit/TWITTER>

TABLE DES MATIÈRES

Remerciements	2
Liens importants	2
TABLE DES FIGURES	4
INTRODUCTION	5
BESOINS ET OBJECTIFS DU PROJET	6
CONTEXTE	6
MOTIVATION	6
OBJECTIFS ET CONTRAINTES	6
GESTION DU PROJET	7
L'ÉQUIPE	7
RÉPARTITION DES TÂCHES	7
GESTION DE LA COMMUNICATION	8
DÉVELOPPEMENT TECHNIQUE	9
LA STRATÉGIE	9
La stratégie prévisionnelle	9
LES SOLUTIONS	10
L'architecture du dossier	10
Coeur du programme	10
Stockage des tweets	12
UTILISATION	14
Authentification	14
Streaming	14
Stockage	15
Analyse des tweets	17
Cartographie	17
BILAN	20
Bilans personnels	20
Bilan global	20

TABLE DES FIGURES

Figure 1 - Répartition des tâches effectuées par Clément MASSIT

Figure 2 - Répartition des tâches effectuées par Zied GOBJI

Figure 3 - Architecture du dossier principal

Figure 4 - Architecture du dossier 'twitter_streaming'

Figure 5 - Contenu du fichier contenant nos tokens.

Figure 6 - Script SQL utilisé pour créer la base de données.

Figure 7 - Base de données des tweets récupérés.

Figure 8 - Authentification à l'API Twitter.

Figure 9 - Déclenchement du streaming.

Figure 10 - Affichage de quelques informations relatives aux tweets.

Figure 11 - Base de données des informations liées aux polygones.

Figure 12 - Récupération des mots les plus communs.

Figure 13 et 14 - Maps représentatifs des différents tweets, dans les différentes villes, avec les mots les plus utilisés dans ces villes.

I. INTRODUCTION

Dans le cadre de notre 1ère année du cycle ingénieurs en Informatique Données et Usages à Polytech Annecy, il nous a été proposé de mettre en œuvre nos connaissances techniques avec un projet sur l'extraction de tweets de Twitter. Cette extraction aura pour objectif primaire l'extraction des données contenues dans les tweets pour, dans un second temps, les analyser.

Soucieux de développer nos connaissances en informatique, en cartographie, et plus particulièrement en data science, notre groupe composé de Clément MASSIT et Zied GOBJI, a saisi l'opportunité d'exploiter cet intérêt commun afin de mener à bien ce projet.

II. BESOINS ET OBJECTIFS DU PROJET

1. CONTEXTE

Nous avons, lors d'un projet antérieur à celui-ci, utilisé la méthode dite de scraping pour récupérer des informations relatives aux reviews laissées sur le site TripAdvisor, nous récupérons reviews sans utiliser l'API de TripAdvisor. Dans le cas de Twitter, nous avons accès à l'API nous permettant de récupérer des tweets, donc de faire des tweets une source d'informations exploitables importante. Aborder la récupération de données via l'utilisation d'API constitue donc une méthode différente pour nous, d'où notre intérêt. De plus, l'extraction de données est très utilisée notamment dans le marketing ou plus généralement dans le codage d'applications, la quantité de données disponibles est énorme et si certaines peuvent être considérées comme inutiles, d'autres comme par exemple les informations liées à la géolocalisation peuvent s'avérer très valuable.

2. MOTIVATION

L'objectif principal de ce projet concerne la collecte d'informations, en effet la récupération, l'exploitation et le traitement d'un grand nombre de données sont les principaux enjeux du domaine de la data science. Il est particulièrement intéressant de s'appuyer sur les différentes méthodes de collecte d'informations. C'est une nouveauté pour nous deux, donc l'expérience et l'apprentissage de ces méthodes ne peuvent qu'être exponentielles durant ce projet.

3. OBJECTIFS ET CONTRAINTES

L'objectif principal de ce projet est de réussir à récupérer des tweets géolocalisés de Twitter ainsi que toutes les données associées afin de les traiter et d'épingler ces tweets sur une carte OpenStreetMap hébergée en local sur une page web.

Les principales contraintes ont été liées à l'API de Streaming de Twitter, particulièrement de la rate limite de l'API, et du fait que très peu de tweets possédaient de données liées à la localisation. De plus, le format des tweets est très complet, il faut alors

III. GESTION DU PROJET

1. L'ÉQUIPE

Zied GOBJI : Anciennement préparatoire au lycée Aristide Briand d'Evreux (27) parcours PCSI-PC, actuellement en 3ème année d'école d'ingénieurs spécialité Informatique Données & Usages à Polytech Annecy.

Clément MASSIT : Issu d'une classe préparatoire aux grandes écoles dans l'Isère (38) à Voiron au lycée Ferdinand Buisson en filière PTSI-PT, actuellement en 3ème année d'école d'ingénieurs spécialité Informatique Données & Usages à Polytech Annecy-Chambéry.

2. RÉPARTITION DES TÂCHES

Clément MASSIT

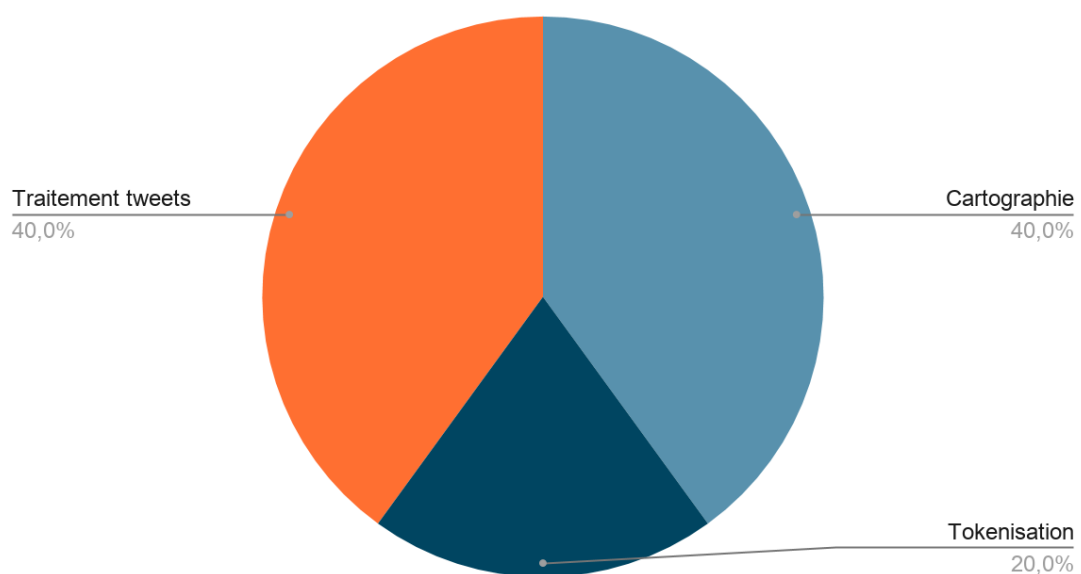


Fig.1 Répartition des tâches effectuées par Clément MASSIT, le pourcentage est indicateur du temps passé sur la tâche et de sa difficulté.

Zied GOBJI

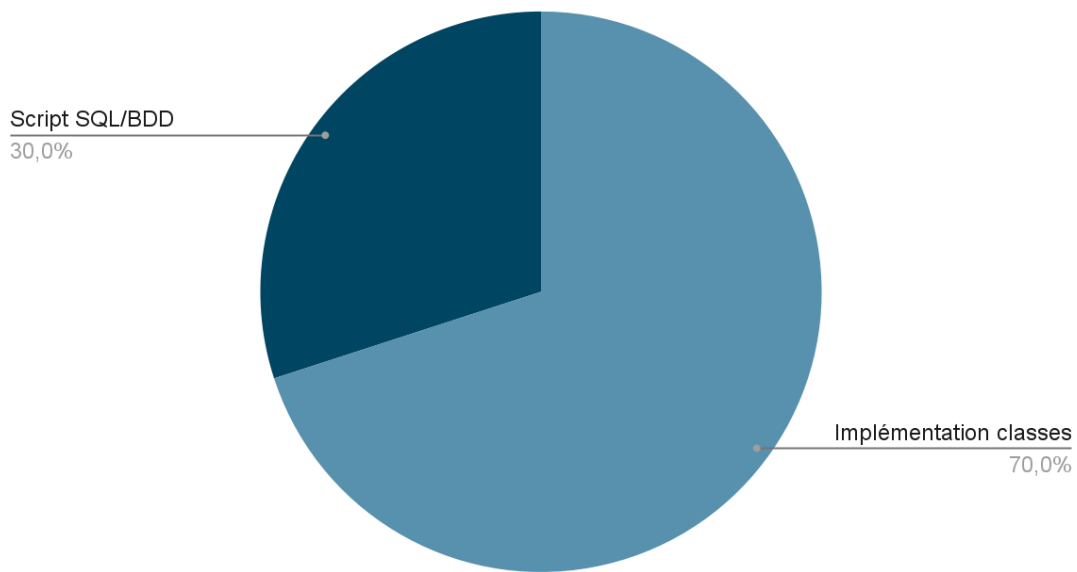


Fig.2 Répartition des tâches effectuées par Zied GOBJI, le pourcentage est indicateur du temps passé sur la tâche et de sa difficulté.

3. GESTION DE LA COMMUNICATION

Dans le contexte sanitaire actuel avec l'ensemble des restrictions de déplacement et de regroupement, nous avons été dans l'obligation de suivre ce projet en distanciel. Nous avons donc mis en œuvre des outils de communication pour pouvoir échanger entre les membres du groupe de projet et les clients (ici les professeurs).

Pour communiquer d'une manière professionnelle avec les clients nous avons utilisé l'outil **Microsoft Teams** sous forme de réunion privée.

Pour communiquer entre les membres du groupe de projet et partager nos interrogations nous avons utilisé la plateforme de discussion **Discord**.

IV. DÉVELOPPEMENT TECHNIQUE

1. LA STRATÉGIE

a. La stratégie prévisionnelle

Pour récupérer les tweets, Twitter met à la disposition des utilisateurs possédant un compte développeur, il faut ensuite créer une application et attendre la réponse de Twitter. Une fois que nous avons notre compte développeur de validé, Twitter met à notre disposition 2 APIs : l'API de recherche et l'API de streaming. L'API de recherche est une API qui fonctionne selon un modèle de requête/réponse entre le client et le serveur où le client envoie une requête et le serveur y répond. L'API de streaming fonctionne selon un modèle d'écoute, on envoie une requête au serveur pour une certaine information, le serveur y répond mais l'échange ne se termine pas tout de suite, la requête reste active le temps d'une mise à jour. C'est donc vers l'API de streaming que nous allons nous tourner afin d'avoir une map qui se met à jour en temps réel.

Une fois les tweets récupérés, nous les avons traités et stockés dans une base de données phpmyadmin en localhost. Stocker ces données facilite la cartographie sur OpenStreetMap et le traitement sémantique par tokenisation.

Pour créer une map facilement à l'aide des coordonnées GPS des tweets, nous avons utilisé la librairie Folium dont nous développerons le fonctionnement dans la section suivante.

De manière analogue, nous avons utilisé la bibliothèque Natural Language Toolkit pour le procédé de tokenisation, dont le fonctionnement sera également détaillé par la suite.

2. LES SOLUTIONS

a. L'architecture du dossier

Voici comment est organisé notre dossier :

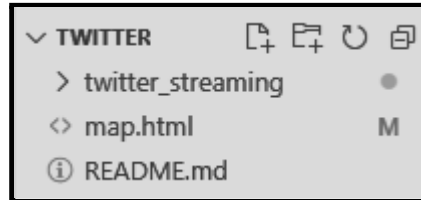


Fig.3 Architecture du dossier principal

Le dossier du code source est le folder **'twitter_streaming'**

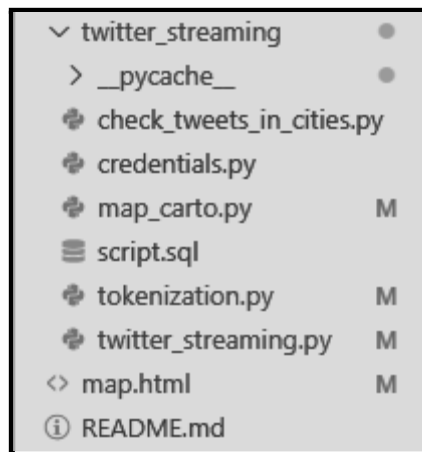


Fig.4 Architecture du dossier **'twitter_streaming'**

b. Coeur du programme

Nous avons créé un fichier credentials comprenant nos tokens par simple souci d'ergonomie, nous codions chacun de notre côté et il était peu pratique de changer à chaque fois de tokens.

```
# Variables qui contiennent les données dont on va se servir pour
accéder à l'API de Twitter

# Zied
ACCESS_TOKEN = '1369656545647157250-qL3Ka0yjh2MW1Gw3xk52sG9ZwBKmQ'
ACCESS_SECRET = 'fwqbhYHmGOTQbqkddPzSCUsZLRJ9yuMsLtTgJBUGU1Bz'
CONSUMER_KEY = 'T9YeTlmlwLyGGnV1Ar2UQsgDMH'
CONSUMER_SECRET = '4Cq74qqm47s6PSykLn3GP1qogCn0QqiQfCt0eYJGylvW5PaUr0'

# Clément
```

```
ACCESS_TOKEN = '1179115613966536706-43kJxODf465JpsC1LTxkpWGzoSjKv1'  
ACCESS_SECRET = 'h4dMsnVMMTcNLMua8ZqYJiQYrsIMloNkRWF5JUAq0KdMt '  
CONSUMER_KEY = 'quSMSDm2KMENSM8k5egwhh9Cp'  
CONSUMER_SECRET = 'X5PQ1JvkvBCzyWv2azJ2u3NNAWdZxXMx3vnUbN47Kx4IhwnvE '
```

Fig.5 Contenu du fichier contenant nos tokens.

Le fichier `check_tweets_in_cities` dans lequel est définie la fonction `check_tweets_in_city` qui prend en argument une ville et qui renvoie la liste des tweets qui ont pour ville la ville rentrée en paramètre, ainsi nous obtenons la liste des tweets d'une ville voulue.

Afin d'avoir un aspect visuel au projet, nous afficherons les tweets sur une carte à l'aide du module '**Folium**'. Pour les afficher, il suffit de les récupérer depuis la base de données puis de les afficher sur la carte à l'aide de 'Marker'. Tout le traitement cartographique est implémenté dans le fichier '**map_carto**'.

La librairie **NLTK** permet la tokenisation, en ajoutant manuellement les stop words on peut affiner la liste des mots que l'on récupère, les stop words agissent comme un filtre, ils ne seront pas conservés dans la liste des mots les plus fréquents (par ex: 'la', 'de', '.', '?...'), on peut ainsi définir la fonction `list_word_most_common` qui prend en paramètre une ville et qui renvoie la liste des mots les plus communs. Ainsi c'est le fichier '**tokenization**' qui procèdera à la tokenisation.

Le fichier '**twitter_streaming**' est le cœur du programme. Nous avons besoin de 3 grandes classes pour pouvoir lancer un stream :

```
class TwitterAuthenticator():  
class TwitterStreamer():  
class TwitterListener(StreamListener):
```

TwitterAuthenticator(): permet l'authentification avec les tokens respectifs

TwitterStreamer(): permet de lancer un stream avec la possibilité de filtrer les recherches.

TwitterListener(StreamListener): permet d'écouter les tweets reçus et d'en traiter les principales informations. Nous pouvons également traiter les différents types d'erreurs que nous pouvons rencontrer en streamant.

c. Stockage des tweets

Le fichier '**script.sql**' comprend simplement le script sql à insérer sur phpmyadmin pour générer toutes les tables et les colonnes nécessaires.

```
set NAMES 'UTF8';

DROP TABLE IF EXISTS tweets_streaming;
CREATE TABLE tweets_streaming (
  id INT PRIMARY KEY AUTO_INCREMENT,
  created_at varchar(100),
  user_name varchar(100),
  text_contenu varchar(255),
  latitude varchar(100),
  longitude varchar(100),
  place varchar(100),
  id_place varchar(100)
);

DROP TABLE IF EXISTS polygon;
CREATE TABLE polygon (
  id INT PRIMARY KEY AUTO_INCREMENT,
  id_place varchar(100),
  place varchar(100),
  lat_1 varchar(100),
  long_1 varchar(100),
  lat_2 varchar(100),
  long_2 varchar(100),
  lat_3 varchar(100),
  long_3 varchar(100),
  lat_4 varchar(100),
  long_4 varchar(100)
);
```

Fig.6 Script SQL utilisé pour créer la base de données.

Ce script SQL nous a permis de stocker nos données dans une base de données hébergée en local.

Comme énoncé précédemment, les tweets seront stockés dans une base de données hébergée en local sous cette forme :

	id	created_at	user_name	text_contenu	latitude	longitude	place	id_place
<input type="checkbox"/> Éditer Copier Supprimer	336	Thu Jun 03 14:24:48 +0000 2021	NS World Studio International	Plus que quelques places disponibles pour la maste...	43.5311	5.454	Aix-en-Provence	1dd0e4e359a6d2e5
<input type="checkbox"/> Éditer Copier Supprimer	280	Thu Jun 03 12:51:08 +0000 2021	Tourisme Corse Orientale - L'Orientale -	Vient de publier une photo à Aléria https://t.co/c...	42.11619	9.51325	Aléria	0fd76f62c0d9b5dd
<input type="checkbox"/> Éditer Copier Supprimer	325	Thu Jun 03 14:08:18 +0000 2021	Art Is An Ale Brewing	The * is out! Don't miss it! Nous serons ouverts ...	47.41168068	0.98651701	Amboise	330326ad0a0793e9
<input type="checkbox"/> Éditer Copier Supprimer	363	Thu Jun 03 21:13:04 +0000 2021	RootCat ☆ S. Pietri	23.23 #TODAY 🚩 #ROOTCAT #DAILY #DANK #VIDEO #MEME...	48.84848	1.4359	Anet	42e2ad87820b1986
<input type="checkbox"/> Éditer Copier Supprimer	271	Thu Jun 03 12:42:15 +0000 2021	Nicodem Déménagement Annecy	Ont est encore bien cachés au clos du buisson à An...	45.9167	6.15	Annecy-le-Vieux	50f64b50fa49c8a2
<input type="checkbox"/> Éditer Copier Supprimer	361	Thu Jun 03 21:07:15 +0000 2021	brandon	🔪🔪 à L'Arlatan https://t.co/jcYHhdu39	43.678577	4.627228	Arles	2074192aa1e2683d
<input type="checkbox"/> Éditer Copier Supprimer	316	Thu Jun 03 13:58:58 +0000 2021	Marcel et Maurice	Vient de publier une photo à Aurillac https://t.co...	44.92792	2.4449	Aurillac	3832a67fa4523976
<input type="checkbox"/> Éditer Copier Supprimer	366	Fri Jun 04 11:22:20 +0000 2021	Chantal Cabuy	I'm at Ermitage Saint Antoine in Barbizon, Île-de-...	48.445347	2.605782	Barbizon	53130a5642b8ce2d
<input type="checkbox"/> Éditer Copier Supprimer	368	Fri Jun 04 11:23:34 +0000 2021	Chantal Cabuy	I'm at Barbizon in Île-de-France https://t.co/D0E5...	48.44568469	2.60435465	Barbizon	53130a5642b8ce2d
<input type="checkbox"/> Éditer Copier Supprimer	259	Thu Jun 03 10:51:30 +0000 2021	Viladis	Zac Efron is now a man à Bayonne, L'aquitanie, Fra...	43.49317596	-1.47376334	Bayonne	1fc42cb070b8b349
<input type="checkbox"/> Éditer Copier Supprimer	237	Thu Jun 03 10:19:16 +0000 2021	Dakor beauty France	Venez réaliser vos prestations chez Dakor beauty F...	47.243095	6.0209346	Besançon	12eb9b254faf37a3
<input type="checkbox"/> Éditer Copier Supprimer	240	Thu Jun 03 10:22:26 +0000 2021	Sortir dans le sud	Vient de publier une photo à Trampoline Parc. Bezi...	43.34339	3.25352	Béziers	696efc0590746aae
<input type="checkbox"/> Éditer Copier Supprimer	392	Fri Jun 04 11:51:11 +0000 2021	Kyo'hon - Manga Café associatif	Encore un ticket gagnant ! Même s'il ne reste que ...	43.34519	3.21969	Béziers	696efc0590746aae
<input type="checkbox"/> Console de requêtes SQL Supprimer	341	Thu Jun 03 20:35:23 +0000 2021	Antonetta Santagata	Blauvac et ses #coquelicots @ Blauvac	44.03027	5.19851	Blauvac	10efee44fd7bc1b9

Fig.7 Base de données des tweets récupérés.

V. UTILISATION

1. Authentification

La première étape est l'authentification, notamment avec les tokens, d'où l'importance de les avoir récupérés en amont. On définit donc les constantes `CONSUMER_KEY`, `ACCESS_TOKEN`, `CONSUMER_SECRET` et `ACCESS_SECRET` auxquelles seront associées les différents tokens. Vient donc l'étape de l'authentification, notamment avec les méthodes `OAuthHandler()` et `API()` de `tweepy` qui permettront de se connecter à l'API de Twitter.

```
auth = tweepy.OAuthHandler(credentials.CONSUMER_KEY,
credentials.CONSUMER_SECRET)
auth.set_access_token(credentials.ACCESS_TOKEN, credentials.ACCESS_SECRET)

# Create the api to connect to twitter with your credentials
api = tweepy.API(auth,wait_on_rate_limit=True,
wait_on_rate_limit_notify=True,compression=True)
```

Fig.8 Authentification à l'API Twitter.

2. Streaming

Ensuite l'objectif principal de ce projet est d'arriver à utiliser l'API de Streaming, ainsi une fonction principale nous permettra de streamer via un écouteur. En particulier on stream certains tweets en langue française et avec une localisation qui est passée en paramètre (pour la localisation on reste sur un polygone qui se rapproche d'un contour très grossier de la France).

```
def stream_tweets(self, fetched_tweets_filename, location):
    # Gère l'authentification et la connexion à l'API de streaming
    listener = TwitterListener(fetched_tweets_filename)
    auth = self.twitter_authenticator.authenticate_twitter_app()

    # On instancie un objet de la classe Stream, qui va nous
    permettre de récupérer les tweets
    stream = Stream(auth, listener)
    stream.filter(locations=location, languages=["fr"])
```

Ensuite pour lancer le stream c'est la fonction suivante qui va déclencher le streaming en continu des tweets.

```
def Stream_via_hash_tag_method():  
    ##### TWEET STREAM FROM LOCATION #####  
    twitter_streamer = TwitterStreamer()  
    twitter_streamer.stream_tweets(fetched_tweets_filename, France)
```

Fig.9 Déclenchement du streaming.

A la réception d'un tweet on pourra faire apparaître les informations du tweet via la fonction `on_data ()`, puis les ajouter directement dans la base de données.

Par exemple on peut afficher les informations qui nous intéressent de la façon suivante :

```
Thu Jun 03 14:08:18 +0000 2021  
Art Is An Ale Brewing  
The ☀ is out! Don't miss it! Nous serons ouverts jusqu'à ce que  
la pluie vienne #quickapero  
  
#hardseltzer... https://t.co/SeckTvqyXa  
47.41168068  
0.98651701  
Amboise  
330326ad0a0793e9
```

Fig.10 Affichage de quelques informations relatives aux tweets.

3. Stockage

On récupère les informations relatives aux tweets sous forme d'un dictionnaire, les informations sont directement envoyées par requête SQL dans notre base de données, base à partir de laquelle nous pourrions récupérer les données sous forme de liste pour mieux les analyser.

A la réception d'un tweet, on insère directement les informations du tweet dans la base de données et aussi les informations concernant les polygones des villes (une simple requête SQL pourra supprimer les doublons dans la base de données).

			id	id_place	place	lat_1	long_1	lat_2	long_2	lat_3	long_3	lat_4	long_4
<input type="checkbox"/>		Éditer		Copier		Supprimer	326	42e2ad87820b1986	Anet	48.830396	1.422482	48.871601	1.463966
<input type="checkbox"/>		Éditer		Copier		Supprimer	324	2074192aa1e2683d	Arles	43.326885	4.425886	43.760524	4.876824
<input type="checkbox"/>		Éditer		Copier		Supprimer	329	53130a5642b8ce2d	Barbizon	48.434204	2.576852	48.460375	2.630638
<input type="checkbox"/>		Éditer		Copier		Supprimer	355	696efc0590746aae	Béziers	43.297394	3.127343	43.39941	3.341688
<input type="checkbox"/>		Éditer		Copier		Supprimer	304	10efee44fd7bc1b9	Blauvac	44.00616	5.171496	44.06506	5.312034
<input type="checkbox"/>		Éditer		Copier		Supprimer	314	40ce1cd0eac02958	Bordeaux	44.811122	-0.638964	44.916354	-0.533496
<input type="checkbox"/>		Éditer		Copier		Supprimer	342	0653bb913c88c1ea	Brussels	50.796328	4.313989	50.913706	4.436947
<input type="checkbox"/>		Éditer		Copier		Supprimer	316	5810a26841e482c3	Caen	49.153091	-0.413428	49.216178	-0.33049
<input type="checkbox"/>		Éditer		Copier		Supprimer	328	00fbd6ddcbe3aea3	Camiers	50.539013	1.57638	50.582056	1.649683
<input type="checkbox"/>		Éditer		Copier		Supprimer	327	57452d804166c06a	Chaumont-en-Vexin	49.248693	1.837408	49.298107	1.928117
<input type="checkbox"/>		Éditer		Copier		Supprimer	341	5d9705592c59b2ee	Corbeil-Essonnes	48.569089	2.436386	48.625327	2.492299
<input type="checkbox"/>		Éditer		Copier		Supprimer	308	587fed6c2b94ea22	Décines-Charpieu	45.744313	4.932574	45.807631	4.985468
<input type="checkbox"/>		Éditer		Copier		Supprimer	305	3a8a2c667faaf9ba	Exeter	50.693458	-3.568747	50.748422	-3.461006
<input type="checkbox"/>		Éditer		Copier		Supprimer	337	09af7e209dc28f5f	Frontignan	43.4163	3.672493	43.484278	3.821754
<input type="checkbox"/>		Éditer		Copier		Supprimer	315	2cc7ac0cc51b1f11	L'Île-Saint-Denis	48.916148	2.290808	48.951836	2.341709
<input type="checkbox"/>		Éditer		Copier		Supprimer	339	2f6d5845d6bcac42	La Bresse	47.984605	6.824188	48.057943	7.016182
<input type="checkbox"/>		Éditer		Copier		Supprimer	336	2594965a1ecb7ca7	La Rochelle	46.133121	-1.241744	46.191	-1.111318
<input type="checkbox"/>		Éditer		Copier		Supprimer	306	58e4b2483e65d0ae	La Valette-du-Var	43.12316	5.952839	43.173244	6.025856

Fig. 11 Base de données des informations liées aux polygones.

Cette base de données regroupe uniquement les données des polygones afin de les manipuler lors de la cartographie des villes. On affichera un polygone représentant grossièrement la ville. Ainsi nous aurons besoin de 4 points géographiques (un point étant associé à une latitude et à une longitude) pour tracer ce polygone.

4. Analyse des tweets

Une fois que tous nos tweets sont dans la base de données, il faut les analyser. Plus particulièrement nous analysons le contenu même du texte du tweet. On passe donc par la librairie de tokenisation **NLTK**. En effet, le but de ce procédé sera d'afficher une liste de 10 mots les plus fréquents par les tweets d'une ville donnée. Pour ce faire, le fichier '**tokenization.py**' retourne cette liste.

```
def list_word_most_common(city):
    liste_tweets = check_tweets_in_city(city)
    caractere = ""
    liste_common_words = []
    for tweets in liste_tweets:
        caractere += tweets[3]

    txt = filtre_stopfr(word_tokenize(caractere, language='english'))

    fd = nltk.FreqDist(txt)
    most_common = fd.most_common()

    for i in range(0, 11):
        liste_common_words.append(most_common[i][0])

    return liste_common_words
```

Fig.12 Récupération des mots les plus communs.

5. Cartographie

Pour représenter les tweets sur une map et avoir une idée de la répartition des mots clés des différentes villes nous avons utilisé la librairie Folium, comme expliqué précédemment, cette librairie va nous permettre de créer une page web HTML avec une carte avec différents marqueurs dont la localisation dépendra des informations de longitude et de latitude obtenues dans les tweets et stockées dans la base de données.

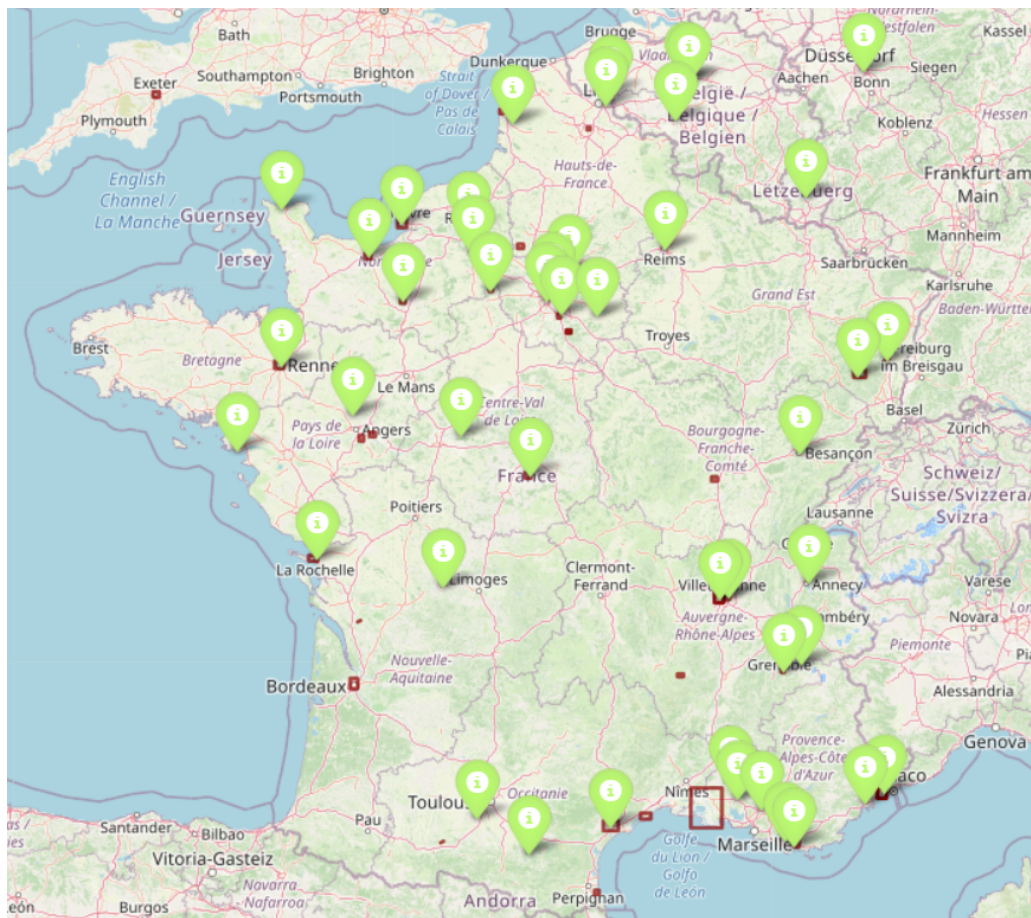


Fig.13 Maps représentatives des différents tweets, dans les différentes villes, avec les mots les plus communs dans ces villes.

Lorsque l'utilisateur clique sur un marqueur d'une ville, un popup s'ouvre avec la liste des 10 mots les plus utilisés dans la ville. Les screenshots suivants montrent l'exemple de la ville de Décines-Charpieu (69) et la ville de Toulon (83). Certains mots peuvent parfois paraître récurrents, cela est dû au nombre de tweets streamés. En effet beaucoup de tweet géolocalisés présentent le même contenu pour des villes différentes.

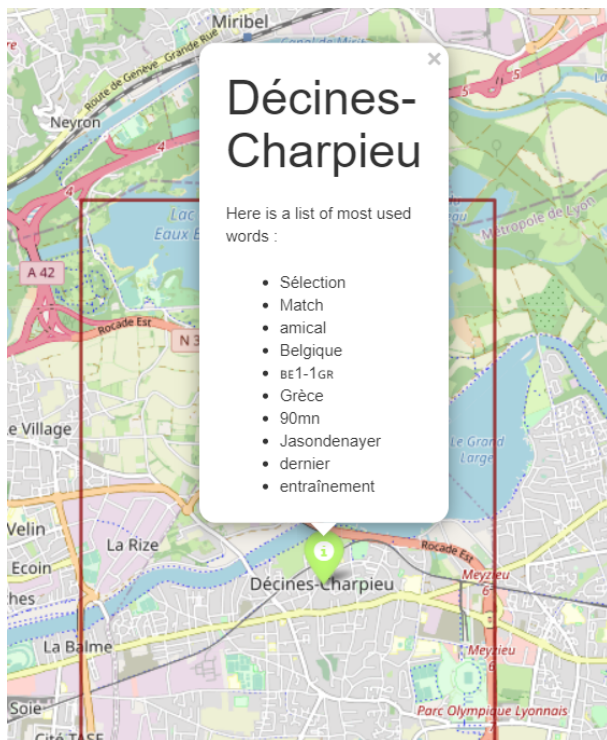


Fig.14 Maps représentatives des différents tweets, dans les différentes villes, avec les mots les plus communs dans ces villes.

VI. BILAN

1. Bilans personnels

“Mener un problème d’une telle ampleur a été pour moi particulièrement enrichissant en termes d’apprentissage et de curiosité. J’ai dû être amené à me familiariser avec certaines librairies qui m’ont appris tant de connaissances. Rassembler un grand nombre de tweets fut très intéressant j’ai pu apprendre quelque chose de très concret en utilisant une plateforme que j’utilise aussi à savoir Twitter. Ainsi le fait d’arriver à produire un visuel à l’aide de la technologie de cartographie rend le projet d’autant plus satisfaisant.” - Clément MASSIT

“Personnellement, ce sujet m’a donné du fil à retordre, j’ai passé énormément de temps à regarder des documentations et à modifier le code en conséquence, ce qui a rendu la compréhension globale assez floue au départ. Cependant, je pense ce projet très formateur d’un point de vue data mining et data analyse, même si cela n’en constitue que la surface, ce projet n’en a pas atténué mon intérêt” - Zied GOBJI.

2. Bilan global

D’un point de vue général, utiliser python afin d’avoir un rendu géographique de la position des tweets et faire du traitement sémantique a vraiment été formateur et a constitué une mise en œuvre des bases de la science des données en s’attaquant à 2 grands domaines : la récupération de données et leur analyse. D’une part, la récupération propre et claire des données fut assez longue à traiter étant donné le peu de tweets géolocalisés présents sur la plateforme. Nous pourrions passer des heures à streamer des centaines de tweets afin d’obtenir des résultats précis et révélateurs du travail mais la base que l’on possède reste favorable à l’étude. D’autre part, l’analyse a été plutôt aisée. Le principe de tokenisation et la cartographie des tweets par rapport à la France s’est fait de manière naturelle à adopter.

Pour conclure, une certaine satisfaction a été partagée par l’ensemble des membres du groupe en réalisant ce projet ainsi qu’en interprétant les résultats.