
PROJET IC05

Les tendances sur TikTok



Table des matières

Introduction	2
I. Récupération des données	3
I.1. Chemin de pensée pour la récupération des données	3
I.1.1 Première ébauche en R	3
I.1.2. Questionnement sur le point de départ	4
I.1.3. Données utiles	5
I.2. Aspiration des données	6
I.2.1. Travail sur les tendances	6
I.2.2. Travail sur les utilisateurs	9
II. Statistiques descriptives	10
II.1. Hypothèses préliminaires	10
II.2. Analyse des résultats	10
II.2.1. Engagement	10
II.2.2. Cycle de vie	14
II.2.3. Audience	17
Conclusion	21

Introduction

La plateforme TikTok s'est rapidement imposée comme l'un des médias sociaux les plus influents de notre époque. Au cœur de son succès réside le concept de "viralité", où les vidéos se propagent rapidement, atteignant des audiences massives en un laps de temps record. Directement matérialisé par la plateforme par une mise en avant des sujets tendances dans la barre de recherche, le phénomène de viralité sur TikTok est complexe et multifactoriel.

Initialement orientée vers l'importance des audios et musiques dans le succès des vidéos sur TikTok, notre étude s'est par la suite recentrée sur le rôle de l'onglet tendance dans la viralité des contenus. Pourquoi les vidéos sont-elles poussées en tendances par l'algorithme ?

Dans ce compte rendu de projet, nous expliquerons l'évolution de notre questionnement sur la plateforme en abordant la récupération des données nécessaires à l'étude. Nous tenterons ensuite d'analyser ces données et leur pertinence concernant la viralité des contenus sur TikTok.

Vous pouvez retrouver l'ensemble de nos travaux sur le GitHub du projet. Le rôle des programmes et leur utilisation y est décrite. Les courbes et graphiques finalement non utilisés pour l'analyse de données y sont également présents : https://github.com/4l3x4ndre/IC05_projet_A23.

I. Récupération des données

I.1. Chemin de pensée pour la récupération des données

I.1.1 Première ébauche en R

En partant de l'idée initiale d'analyse de la viralité des tendances audios, nous avions établi une procédure de récupération de données en R.

Pour trouver les audios d'origines à analyser, nous partions de la page "PourToi" (<https://tiktok.com/foryou>). Nous récupérions le nom de l'audio de chaque vidéo ainsi que le lien vers la page qui regroupe toutes les vidéos employant cet audio.

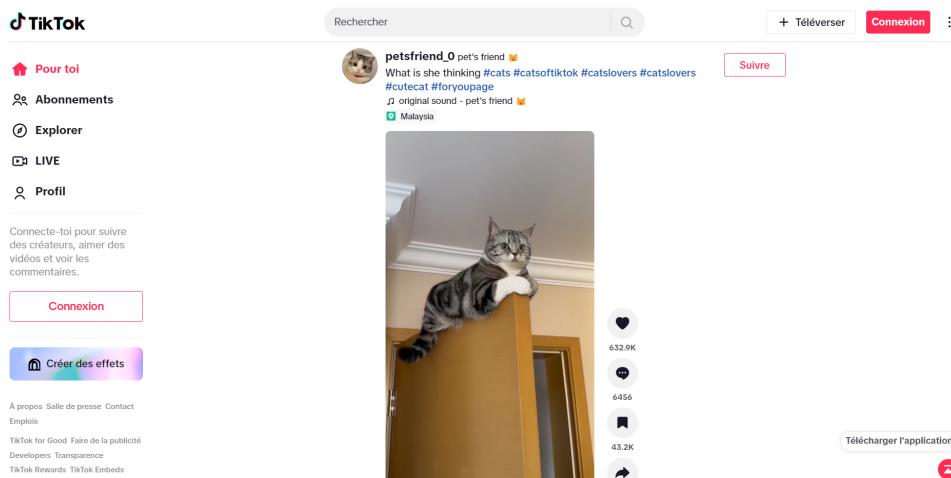


Figure I.1.1.1 : page PourToi

En parcourant ces pages "musiques", on trouve sur chacune d'entre elles le nombre de vidéos employant cet audio et toutes les vidéos. On récupère alors les liens vers chaque vidéo.

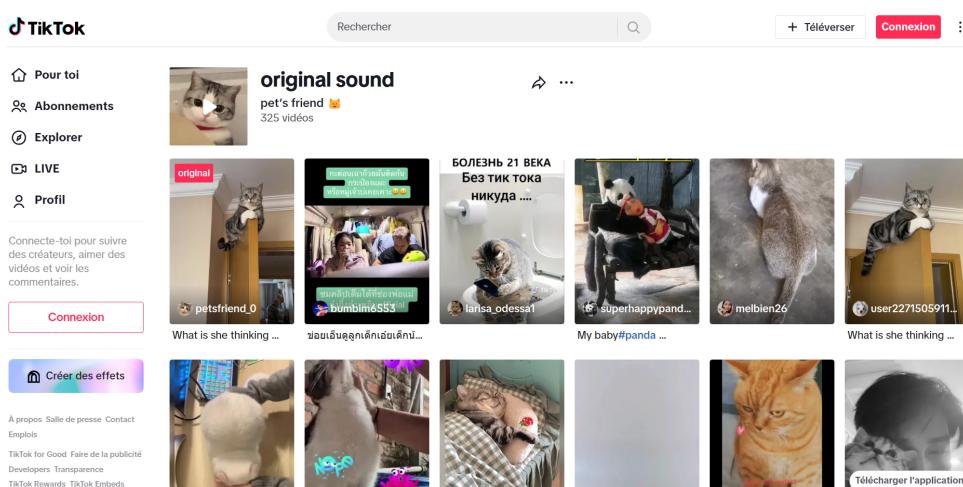


Figure I.1.1.2 : page musique

En parcourant ces pages “vidéos”, on récupère le nom d’utilisateur qui a produit la vidéo ainsi que le nombre de j’aimes, partages, signets et commentaires de celle-ci.

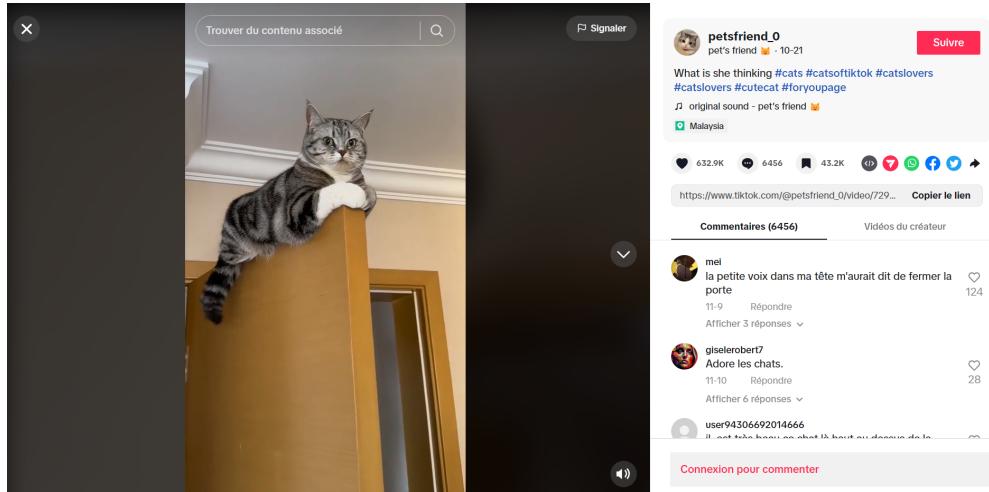


Figure I.1.1.3 : page vidéo

I.1.2. Questionnement sur le point de départ

Après une rapide analyse du contenu récupéré, on remarque le manque de pertinence de nos données. En effet, selon la page PourToi générée, les audios récupérés cumulent parfois très peu de vidéos ou bien celles-ci sont datées. Les résultats posent aussi la question de la géographie des tendances avec l’apparition de vidéos provenant de tous les continents.

Pour se recentrer sur les tendances, qui plus est occidentales, on décide de changer notre point de départ. En partant de la barre de recherche de la page d’accueil, nous avons accès à un certain nombre de recommandations de tendances. On abandonne donc l’analyse des audios au profit des tendances mises en avant par TikTok.

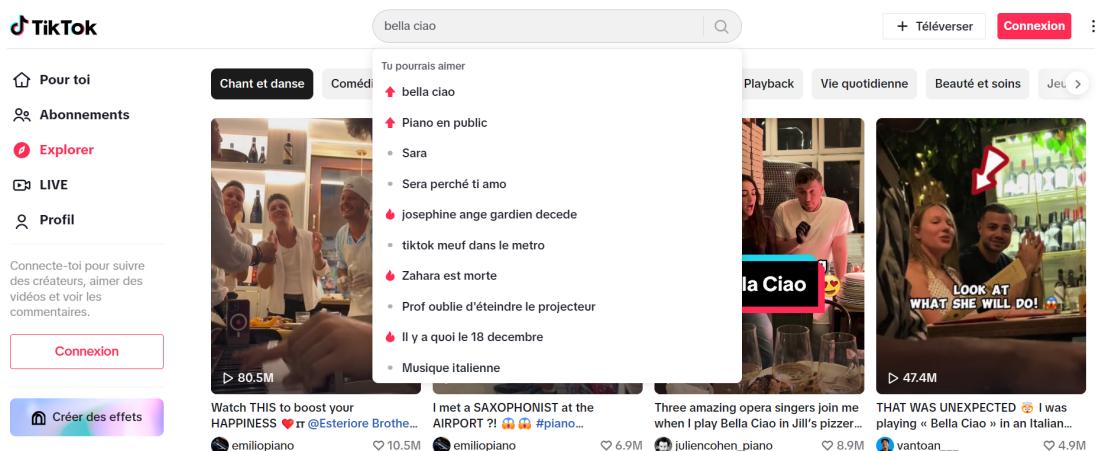


Figure I.1.2.1 : barre de recherche et tendances

Ce changement de direction conduit à une reformulation de notre sujet d’étude initial. On cherchera désormais à identifier sur quels critères les tendances sont créées.

I.1.3. Données utiles

Sur la base de ce nouveau questionnement, nous avons identifié un certain nombre de nouvelles données utiles à récupérer. En plus du nombre de j'aimes, commentaires, signets et partages de chaque vidéo, nous récupérerons désormais les hashtags associés aux vidéos, indicateur essentiel du sujet desdites vidéos.



Figure I.1.3.1 : Interface des données

Il sera également utile de constituer une base de données jumelle à partir de la page PourToi pour comparer les résultats. Les points d'entrées étant drastiquement différents, la comparaison est osée mais pas impossible. Les vidéos tendances sont des vidéos récemment émises (même si des exceptions peuvent exister) tandis que les vidéos de l'échantillon PourToi en sont à des instants différents de leur cycle de vie, mélangées entre vidéos inconnues et vidéos populaires, parfois même issues de tendances. Le fil PourToi n'est donc pas un vivier de vidéos neutre et idéal mais il est le fil classique de la consommation de contenu sur TikTok.

Nous effectuerons également un travail sur les commentaires et les utilisateurs ayant posté ces commentaires. En accédant à leur profil utilisateur, nous identifierons leur centres d'intérêt à partir de leur page "à aimé" publique.

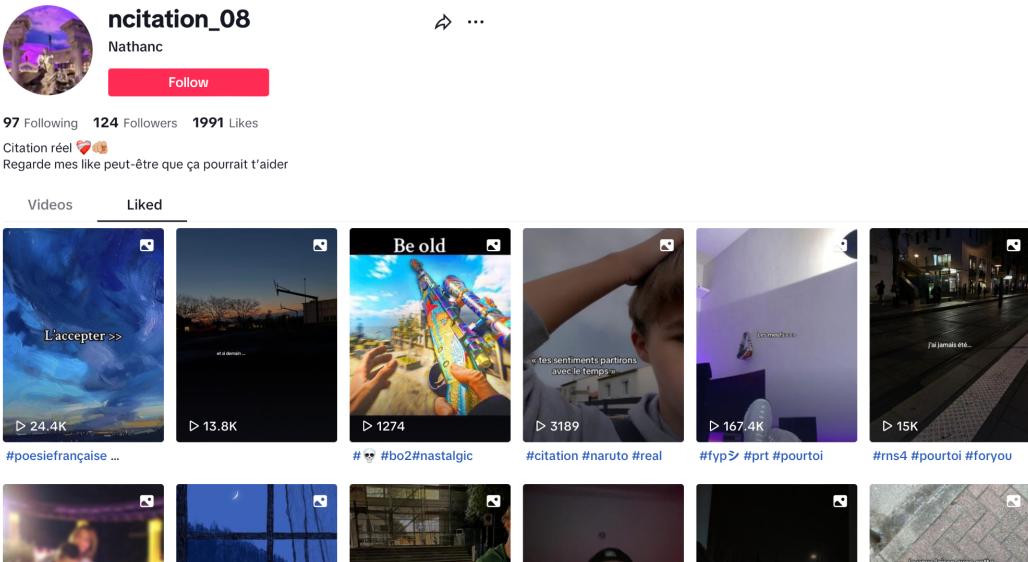


Figure I.1.3.2 : page “à aimé”

Enfin, le nombre de vues est difficilement accessible, celui-ci apparaissant sur le compte de l’utilisateur mais pas sur la vidéo. Nous nous contenterons des j’aimes comme métrique de base.

I.2. Aspiration des données

I.2.1. Travail sur les tendances

L’aspiration des données a été réalisée avec *RSelenium*, reposant sur *Docker*. Le container Firefox pour Selenium¹ a été utilisé par Selenium pour accéder aux pages web. Dans cette partie seront détaillées les processus ayant permis de récupérer des données et leurs obstacles.

Que ce soit sur la récupération des données PourToi ou des données tendances, le flux de navigation est le même : il faut descendre le plus en bas possible de la page, puis parcourir chacun des blocs vidéos pour récupérer les données directement lisibles (les 4 métriques : j’aimes, nombre de commentaires, nombre de signets et nombre de partages), le nom d’utilisateur et le lien de son profil ainsi que les hashtags et le lien de la vidéo. Pour les tendances, il est nécessaire de cliquer sur la vidéo pour pouvoir afficher et récupérer les quatre métriques, cachées sur les pages tendances.

Un programme de récupération des commentaires a également été implémenté pour les vidéos tendances.

¹ <https://hub.docker.com/layers/selenium/standalone-firefox-debug/latest/images/sha256-3dc866181894650e1a41c552ef33db4582a7b7e418fc8e25da5be30617a8fc01?context=explore>

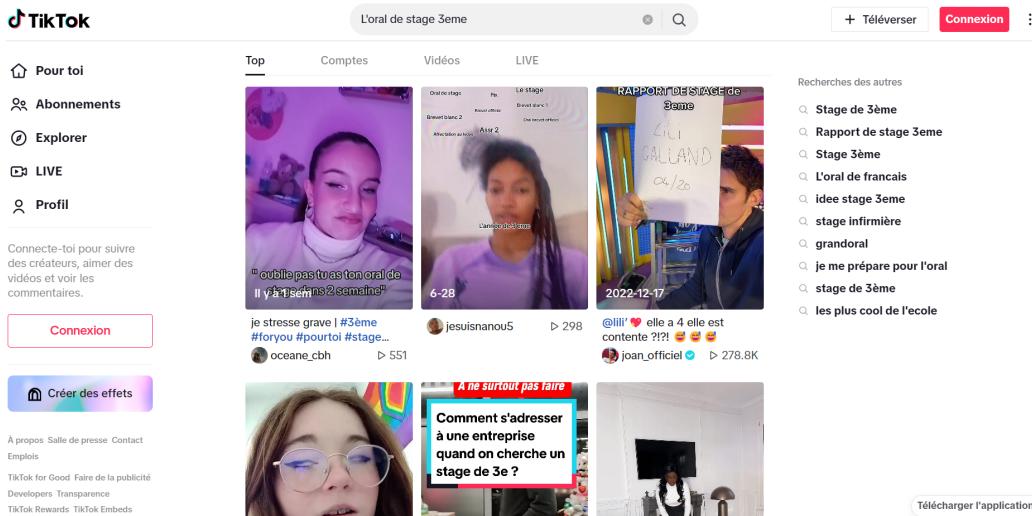


Figure I.2.1.1 : page tendance

La difficulté de la récupération de données sur TikTok réside dans l'utilisation d'un pilote de navigateur web (WebDriver). Un tel dispositif est nécessaire, car les pages TikTok sont dynamiques : leurs contenus ne s'affichent qu'une fois la page chargée dans un navigateur. Une technologie étant capable de contrôler un navigateur web est donc nécessaire. Cet outil permet de reproduire de façon automatique et itérative les actions d'un utilisateur sur le web. Dans ce projet, l'outil a été utilisé pour naviguer sur chacune des pages d'intérêts, d'attendre que la page se charge et enfin d'ajouter les données à la base de données.

Cependant, l'obstacle essentiel à ce type d'outil n'a pas pu être contré : les captcha doivent être résolus à la main pendant la récupération, un captcha apparaissant toutes les 5 à 10 vidéos environ (Figure I.2.1.2).

De plus, il n'a pas été possible de se connecter en passant par *Selenium*. Seules les fonctionnalités de base de la plateforme ont pu être utilisées. L'algorithme de recommandation ne peut donc pas se baser sur l'historique de navigation.

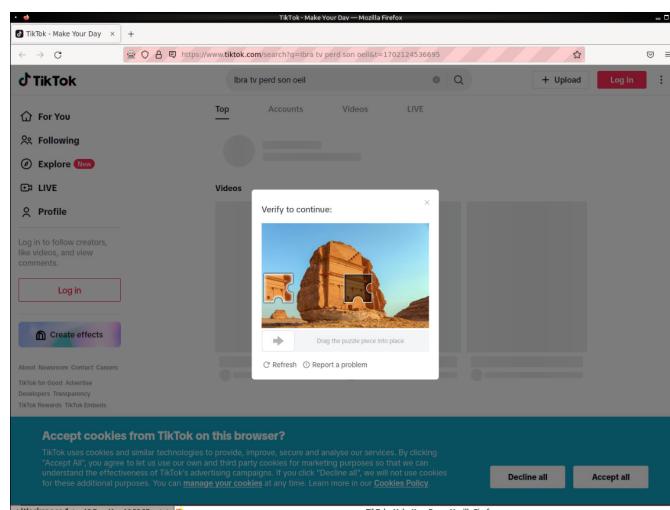


Figure I.2.1.2 : la page web ne se charge pas tant que le captcha n'est pas résolu.
Cette capture d'écran est issue du navigateur lancé sur Docker. On retrouve la couleur rouge dans la barre URL indiquant le contrôle du navigateur par un script

Dernier point technique, la récupération est basée sur le chemin *xpath* des classes des éléments web de TikTok. Une fois la donnée identifiée, elle est ajoutée à l'objet *data frame* qui est écrit en mémoire à la fin du programme. Toutefois, si elle ne l'est pas, l'outil *Selenium* lève une exception et arrête le processus. Or, les aléas dûs aux temps de chargements, aux données absentes ou vidéos indisponibles rendent l'accès aux données parfois impossible. Les programmes sont donc recouverts de structure *tryCatch* pour empêcher Selenium d'arrêter le programme si la donnée cherchée est absente.

La suite de cette sous-partie traite des particularités de la recherche des données tendances.

Une page tendance TikTok permet de visualiser six vidéos lorsque l'utilisateur n'est pas connecté. Par conséquent, seules ces vidéos seront analysées pour chacune des quatorze tendances observées. La structure web des pages vidéos de la plateforme ne rend que partiel la récupération des commentaires. En effet, les mentions/réponses entre utilisateurs dans l'espace commentaire, représentées par des "@" ne peuvent pas être récupérées. Seul le texte l'est.

De plus, TikTok semble détecter l'utilisation d'un outil automatisant la navigation, puisqu'après le début de notre projet, les descriptions/hashtags se sont retrouvés englobés dans un élément web parent, affichés seulement lors d'un clique sur le bouton "More" au dessus de chaque vidéo pour voir la description. Le clique peut être réalisé par Selenium en détectant l'élément web correspondant à chaque itération de vidéo, mais dans certains cas la récupération de ce bouton peut échouer. Les commentaires étant de surcroît peu lisibles algorithmiquement, les données commentaires n'ont pas été utilisées.

Les données récupérées pour chaque vidéo tendance sont donc :

Nom d'utilisateur	Lien utilisateur	Lien vidéo	Commentaires	Hashtags	Date de publication
-------------------	------------------	------------	--------------	----------	---------------------

Puis, à chaque fois que le programme est lancé (par exemple chaque jour), de nouvelles colonnes sont ajoutées correspondant aux 4 métriques du jour x.

Nombre de j'aimes	Nombre de commentaires	Nombre de signets	Nombre de partage
-------------------	------------------------	-------------------	-------------------

Pour les vidéos tendances, il y a donc 6 colonnes fixes et 4 ajoutées à chaque récupération.

Le procédé est le même pour les vidéos PourToi mis à part pour les urls des vidéos et des utilisateurs qui n'ont pas été récupérés. Les quatres métriques sont donc représentées une fois dans ce jeu de données.

Les 14 tendances analysées sont :

GTA6	Candice poils	Cyril hanouna chute	Cargo mouillé	Skin gratuit lego fortnite	L'oral de stage 3eme	Le prÃ©sident Macron humilié à Dubai	pubg tiktok videos
Cristiano Ronaldo	Try not to laugh level hard 100000000000	Gaza pnl	Boutique aujourd'hui 9 décembre	Savannah et Justriadh se marient	french riots	Ibra tv perd son oeil	

Chaque tendance a sa base de données qui est composée à minima des 10 colonnes décrites plus haut. Il y a donc 14 bases de données tendances. Elles sont ensuite regroupées dans une base de données générale contenant toutes les tendances². Cette base contient le nom des tendances et, pour chaque jour de mesure des métriques, une colonne apparaît pour la moyenne de chaque métrique ainsi qu'une colonne pour le maximum et le minimum.

De même, les données de la page PourToi sont enregistrées dans une table³ contenant les quatres métriques, les hashtags et les commentaires. Ces vidéos ne sont parcourues qu'une seule fois lors de leur découverte. L'évolution des métriques n'est ainsi pas enregistrée.

I.2.2. Travail sur les utilisateurs

Les mêmes techniques de scrapping, à savoir R, Sélénum et Docker sont utilisées pour obtenir des données et métadonnées liées aux utilisateurs de la plateforme.

Concernant le scraping des commentaires et leur évolution temporelle ; A partir d'une vidéo nous sommes capables de récupérer les commentaires de cette dernière et d'utiliser les métadonnées de ces commentaires, à savoir la date, pour dresser un histogramme de l'évolution des commentaires en fonction du temps.

Pour réaliser cela, il est nécessaire de réaliser un scroll infini, celui-ci se base sur la longueur de la page. Le programme tente de scroller, si du nouveau contenu se charge la longueur de la page augmente, il continue donc de scroller. Si la longueur ne change pas, il s'arrête, le scroll est fini.

Ensuite, nous récupérons l'indicateur temporel du commentaire, ce dernier prenant plusieurs formats (Exemple : x days ago, 12-03, x h ago), il faut donc convertir ces dates pour pouvoir afficher l'histogramme.

Concernant l'accès aux pages à aimé des utilisateurs ; Certains utilisateurs choisissent de laisser leurs vidéos aimées publiques, ce qui nous permet de récupérer les métadonnées de ces dernières, comme les titres/hashtags ou les auteurs de ces vidéos. Cela nous permettra par exemple d'étudier les centres d'intérêts des utilisateurs ayant commenté une vidéo.

Concernant la réalisation de nuages de mots à partir de titres/hashtags ; Pour réaliser nos nuages de mots il est nécessaires de réaliser un filtrage des mots clés obtenus pour obtenir un résultat important, nous retirons donc les stopwords français et les mots clés comme #foryou ou #fy, qui sont ajoutés sur presque chaque vidéo par les utilisateurs pour espérer augmenter la viralité.

² Fichier *trends-total.csv* du dossier *data*

³ Fichier *foryou.csv* du dossier *data*

II. Statistiques descriptives

II.1. Hypothèses préliminaires

Avant la récupération des données, nous avions établi quelques hypothèses sur les liens existants entre ces données et la mise en avant des contenus sur la plateforme.

Concernant l'engagement des utilisateurs à travers les métriques classiques, à savoir les j'aimes, partages, signets et commentaires, nos hypothèses étaient les suivantes :

- Le taux d'engagement sur les vidéos tendances est supérieur à celui des vidéos du fil classique.
- La temporalité de l'engagement compte pour la mise en avant en tendances d'un contenu par TikTok.
- Les vidéos tendances sont caractérisées par un pic d'engagement au début de leur cycle de vie.

Concernant l'audience des vidéos :

- Les vidéos tendances sont consommées par plusieurs communautés.
- Il y a correspondance entre les centres d'intérêts des utilisateurs engagés et les thèmes des vidéos dans le fil PourToi, là où l'audience est plus diverse concernant les vidéos tendances.

II.2. Analyse des résultats

II.2.1. Engagement

Au niveau des chiffres, on obtient :

- Un nombre moyen de j'aimes de 144 919 sur l'échantillon tendances contre 1 189 892 pour l'échantillon PourToi.
- Un nombre moyen de commentaires de 1142,75 sur l'échantillon tendances contre 7 940 pour l'échantillon PourToi.
- Un nombre moyen de partages de 2 890 sur l'échantillon tendances contre 20 480 pour l'échantillon PourToi.
- Un nombre moyen de signets de 16 608 sur l'échantillon tendances contre 109 801 pour l'échantillon PourToi.

Si les rapports d'échelle sont déséquilibrés avec des chiffres bien supérieurs du côté de l'échantillon PourToi, il sera intéressant de remettre en perspective ces chiffres par rapport à la temporalité des interactions dans la partie suivante. Les vidéos PourToi sont des vidéos dont la date de publication peut être bien antérieure à la date de collecte tandis que les vidéos tendances sont très récentes, pour la plupart.

Analysons plus en détail ces chiffres. La figure ci-dessous fait une représentation du nombre de j'aimes par tendance dans des boîtes à moustaches. On se concentre ici sur les j'aimes mais le comportement des autres métriques est équivalent. Le premier encadré présente un premier échantillon de tendances. Le second constitue l'échantillon enrichi par de nouvelles tendances quelques jours plus tard.

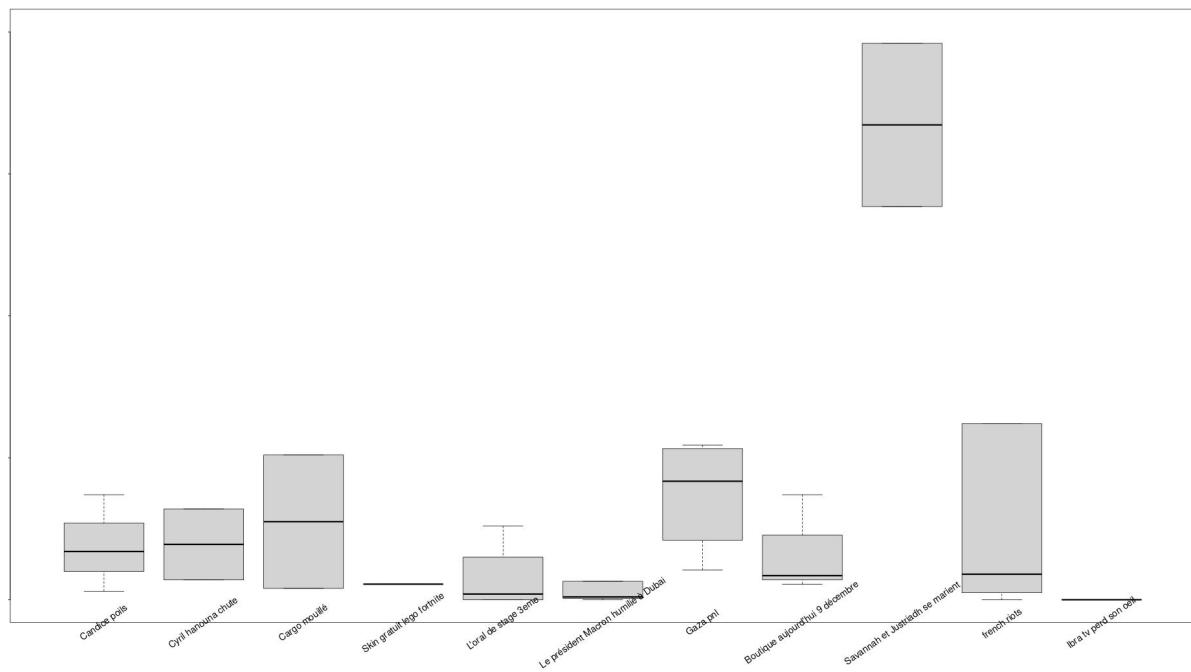


Figure II.2.1.1 : Boîte à moustache du nombre de j'aimes (premier échantillon)

En regardant chaque tendance du premier encadré, la plupart des tendances ont un jeu de données relativement dispersées avec une boîte d'étendue interquartile élancée, la médiane est centrée. Les moustaches sont presque accolées à cette boîte. On constate donc la mise en tendance d'un groupe homogène de vidéos du point de vue du sujet et des statistiques.

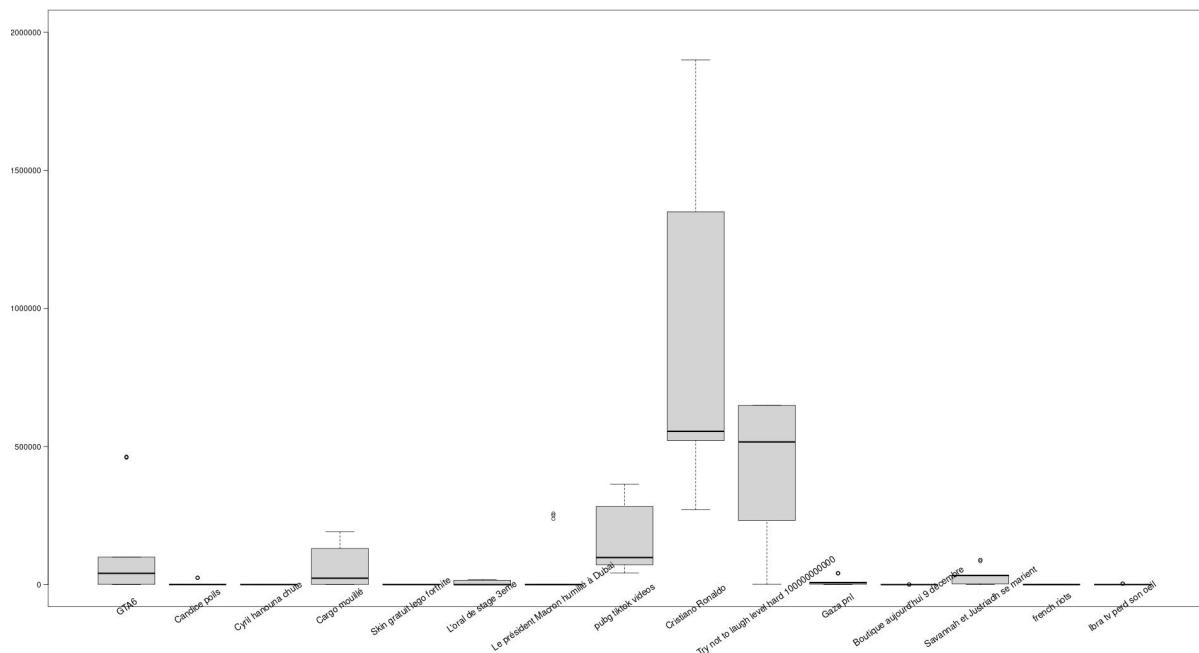


Figure II.2.1.2 : Boîte à moustache du nombre de j'aimes (deuxième échantillon)

Dans le second encadré (si on regarde sans l'écrasement du graphique), quelques jours plus tard, la médiane descend vers le fond de la boîte d'étendue. Si 50% des observations se situent en dessous de la médiane, située elle même en bas de la boîte d'étendue qui comprend 50% des observations totales, cela signifie que très peu d'observations dépassent ce ventre mou.

On a donc deux principales agglomérations de vidéos. Une moitié basse inférieure à la médiane accumulant peu de j'aimes relativement à la tendance. Une autre majorité dans la boîte d'étendue. Enfin, on a l'apparition d'une ou plusieurs valeurs aberrantes représentées par les points au-delà de la moustache supérieure. Il s'agit souvent d'un seul point, une vidéo qui se détache et récolte l'engagement du public.

Il est intéressant de relever le cas de la tendance sur Cristiano Ronaldo qui sort du lot. C'est ici une tendance un peu particulière, ne relevant pas d'un fait d'actualité. Celle-ci cumule un nombre d'interactions bien supérieur aux autres tendances et ses vidéos ne sont pas toutes nouvelles. Nous en discuterons dans la partie sur le cycle de vie des vidéos.

On cherche maintenant à comparer cet échantillon aux vidéos du fil PourToi. Nous avons tracé pour chaque vidéo le nombre de commentaires, partages et signets en fonction du nombre de j'aimes. On retrouve ainsi en rouge l'échantillon des vidéos PourToi et en bleu l'échantillon des vidéos tendances. Chaque point correspond à une vidéo. Pour chacune des catégories, on observe quelque chose de linéaire pour la section tendances. Cela indique une propension importante des utilisateurs à diffuser et interagir avec le contenu autant qu'ils ont tendance à aimer. A l'inverse, la répartition est chaotique en ce qui concerne les vidéos PourToi.

Pour les partages, un nombre important de j'aimes n'implique donc pas nécessairement un nombre de partage proportionnel dans le cas des vidéos PourToi. La plupart des points rouges se trouvent à droite de la droite bleue, ce qui indique un taux de partage plus important sur les vidéos tendance. Pour un même nombre de j'aimes, plus d'utilisateurs partagent les vidéos tendances que les vidéos PourToi.

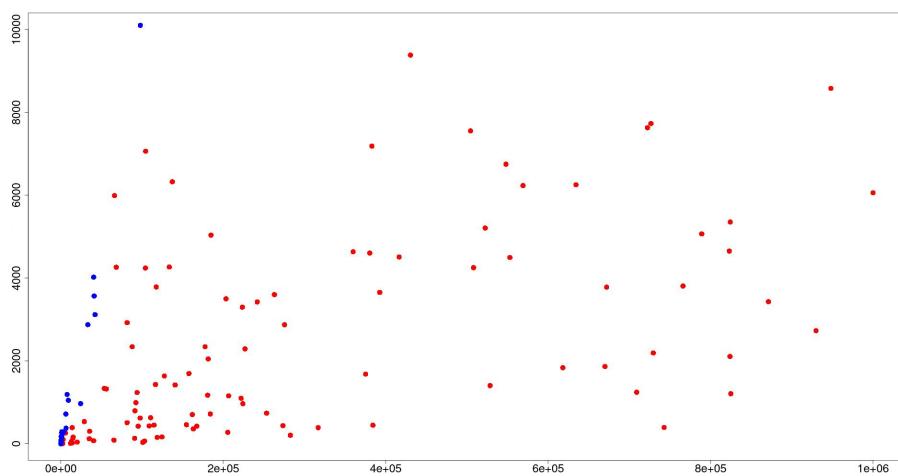


Figure II.2.1.3 : Nombre de partages en fonction de nombre de j'aime

En revanche, en ce qui concerne les signets la tendance est inversée, le nuage de points rouge est à gauche de la droite bleue. C'est assez logique puisque la fonction signet correspond à la sauvegarde du contenu dans un dossier. Les utilisateurs sont plus à même de sauvegarder des contenus mis en

avant pour eux dans leur page PourToi plutôt que de sauvegarder des contenus tendances très éphémères liés à l'actualité. On observe d'ailleurs que la répartition des points rouges est beaucoup moins chaotique pour cette catégorie.

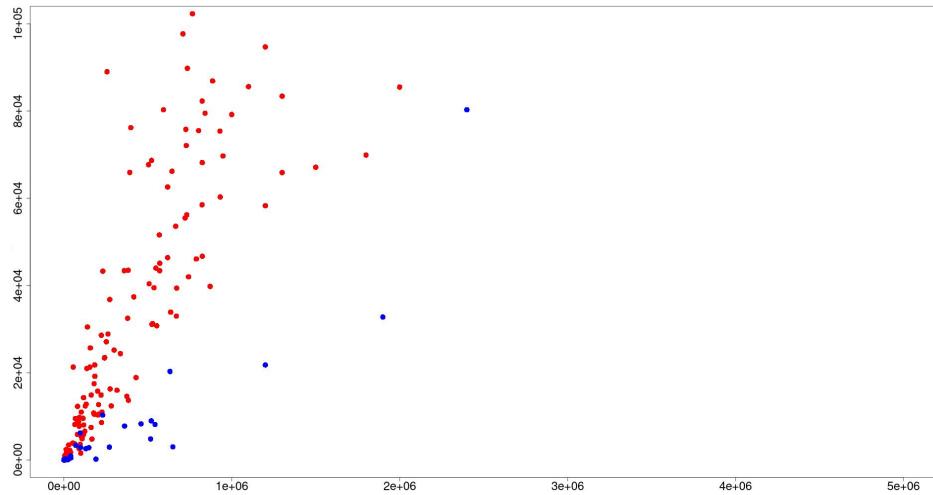


Figure II.2.1.4 : Nombre de signets en fonction de nombre de j'aimes

Pour la variation du nombre de commentaires en fonction du nombre de j'aimes, la progression semble encore linéaire pour le jeu de données tendances. Cette fois-ci en revanche la droite sectionne en deux le nuage de points PourToi. Un utilisateur ne commenterait pas nécessairement plus un contenu tendance qu'un contenu PourToi qu'il a aimé.

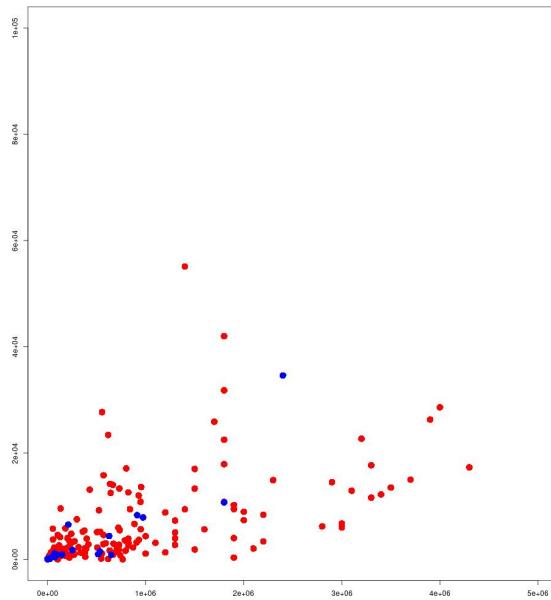


Figure II.2.1.5 : Nombre de commentaires en fonction de nombre de j'aime

Finalement ces résultats ne semblent pas indiquer un taux d'engagement général plus élevé pour les contenus en tendance que pour les contenus PourToi si on considère l'engagement dans la globalité des métriques considérées. Le partage est donc la donnée qui semble influencer la mise en avant en tendance ou non d'un contenu. Il semble caractériser au mieux la viralité d'une vidéo, ce qui est cohérent puisque la notion de viralité correspond à une diffusion.

II.2.2. Cycle de vie

Au-delà des statistiques, la temporalité de l'engagement nous semblait compter dans l'appréciation que fait TikTok d'un sujet tendance ou non. Pour étudier le cycle de vie d'une vidéo, les seules données datées récupérables sont les commentaires. En étudiant leur répartition temporelle, on peut mesurer l'engagement sur la durée.

Nous avons choisi d'étudier la temporalité de quatres types de vidéos différentes:

- Une tendance passée liée à un sujet d'actualité, les émeutes de Paris datant de mars 2023.
- Une vidéo qui est montée en tendance mais avec une thématique reliée à un sujet particulier, ici une vidéo sur le footballeur Cristiano Ronaldo.
- Deux vidéos trouvées sur une page PourToi (sans être connecté) et qui ne sont pas en tendances à l'instant de notre étude. Leurs thèmes sont assez universels et peuvent potentiellement toucher une grande audience ; une vidéo attendrissante de panda et une vidéo de danse sur une musique des années 90.

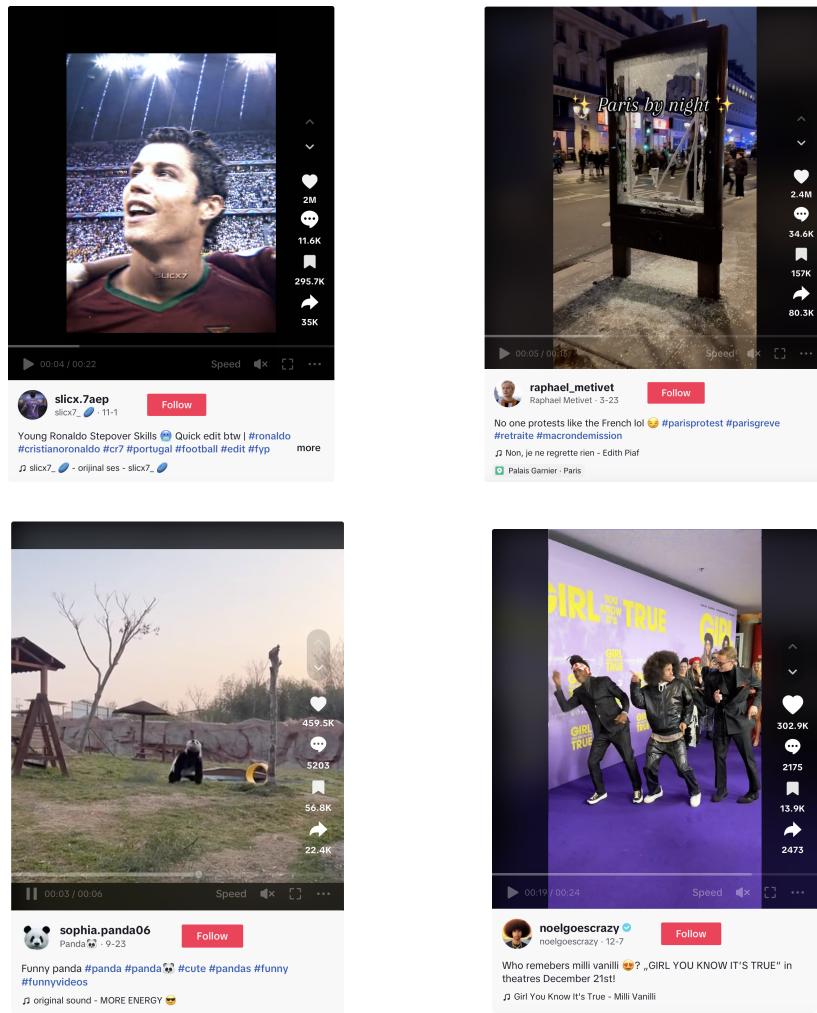


Figure II.2.2.1 : Vidéos choisies pour notre étude

Émeutes Paris

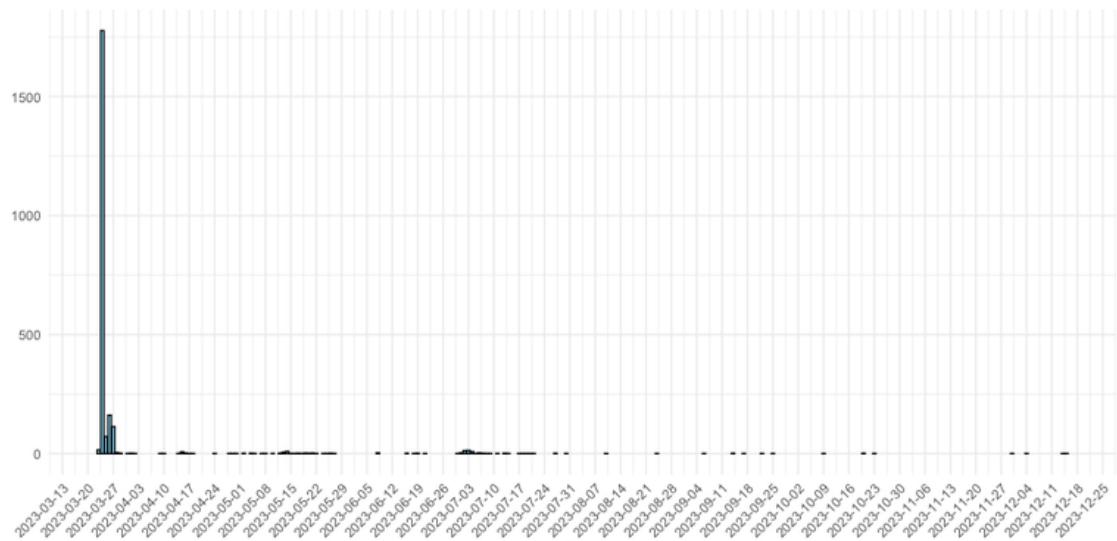


Figure II.2.2.2 : Nombre de commentaires en fonction du temps (Émeutes)

Sur la vidéo liées aux émeutes de Paris, on constate qu'il y a un fort taux d'engagement lors de sa semaine de publication, mais le nombre de commentaires postés par jour diminue fortement et devient quasi nul. Ce phénomène est explicable par la nature du contenu qui est fortement liée à l'actualité et qui ne suscite plus d'intérêt dans les semaines qui suivent.

On constate qu'il y a quelques nouveaux commentaires qui apparaissent, notamment en été, il s'agit principalement de comptes étrangers, amusés par la situation en France, ou qui s'inquiètent pour leurs vacances.

Cristiano Ronaldo

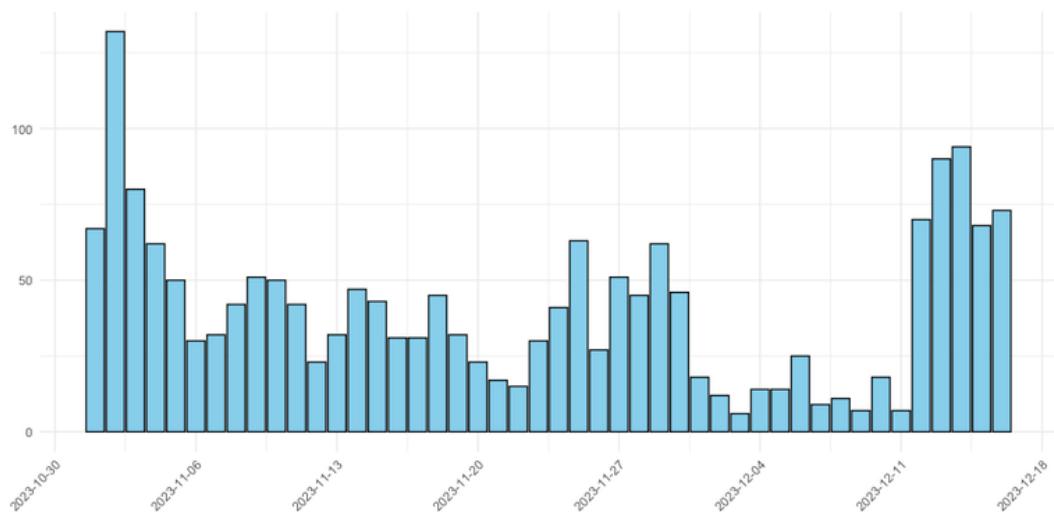


Figure II.2.2.3 : Nombre de commentaires en fonction du temps (Ronaldo)

Si on se penche sur une tendance comme celle de Cristiano Ronaldo, on observe encore un pic important lors de son apparition, mais cette fois-ci on retrouve une certaine constance dans la fréquence et le nombre des commentaires par la suite et ce sur plusieurs mois.

Cela peut notamment s'expliquer par le sujet de la tendance, qui n'est pas liée à une actualité et mais plutôt reliée à une thématique particulière, ici une personnalité importante du football. On peut supposer que l'engagement de cette vidéo est entretenu par une communauté d'utilisateurs actifs, nous essayerons de prouver ceci dans une dernière partie.

Finalement, on peut à nouveau expliquer les pics d'engagement qui poussent en tendance le sujet grâce à l'actualité. En octobre, avec les révélations sur la potentielle fin de carrière de Ronaldo et en mi-décembre, à cause du mercato. On voit d'ailleurs en décembre un regain d'intérêt d'autant plus important que ce pic est précédé d'une baisse du nombre de commentaires.

Vidéos PourToi

Sur une vidéo de la page PourToi, comme la vidéo de danse, on observe une répartition relativement constante des commentaires sur la durée.

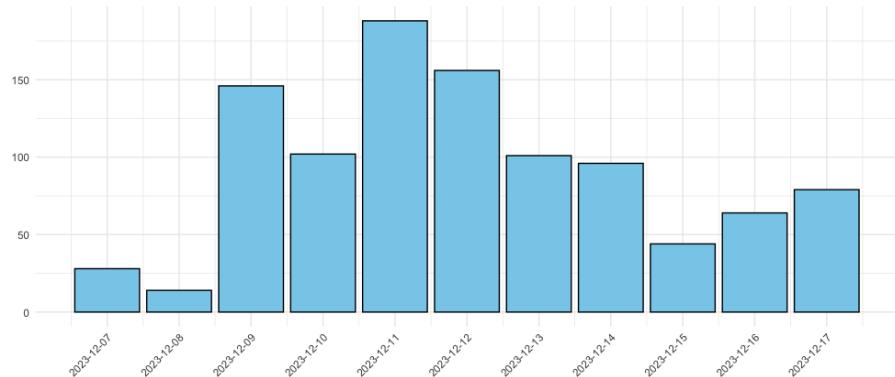


Figure II.2.2.4 : Nombre de commentaires en fonction du temps (Dance)

Ce comportement n'est pas forcément vérifiable pour toutes les vidéos PourToi. La vidéo suivante traitant des pandas récolte un fort engagement au début de son cycle de vie puis une forte baisse de l'engagement. Il est possible que cette vidéo ait été en tendance.

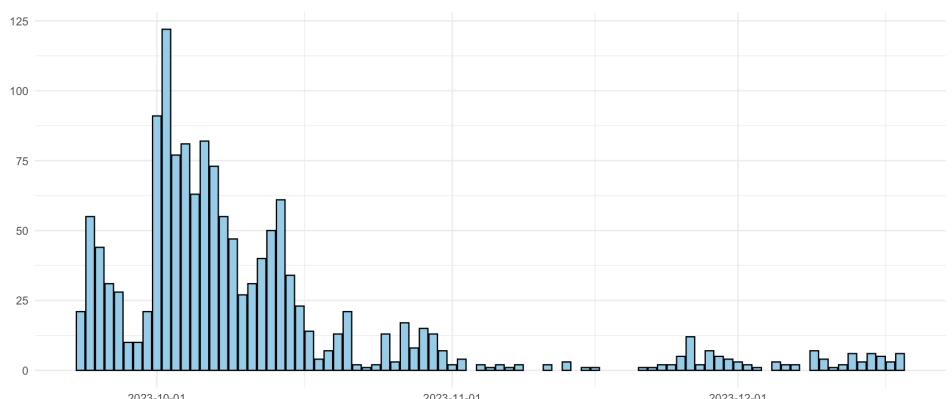


Figure II.2.2.5 : Nombre de commentaires en fonction du temps (Panda)

Pour autant, la vidéo continue de générer des commentaires, de manière relativement constante. On suppose que ceci est dû à la variété d'utilisateurs à qui cette vidéo est présentée.

II.2.3. Audience

Enfin, on cherche à connaître la portée de ces tendances, sont-elles mises en avant pour des profils variés ou n'intéressent-elles que les utilisateurs concernés. Pour étudier l'audience des tendances, on cherche à voir si les utilisateurs qui interagissent par le biais des commentaires ont des centres d'intérêts similaires au sujet de la tendance en question. Pour cela on identifie leurs centres d'intérêt à partir des vidéos qu'ils ont aimé. On compare les titres et hashtags de ces vidéos aimées à ceux de la tendance commentée.

Premièrement, au niveau des hashtags des deux échantillons, on observe une grande différence. Les hashtags de l'échantillon tendance sont répartis en fonction de leur popularité. Au moment de la génération du nuage de mots, la tendance la plus populaire, recommandée en premier dans la barre de recherche était celle de gta. On retrouve mis en avant les hashtags #gta6, #rockstargames mais aussi #jeuxvideo. Du côté de l'échantillon pour toi, les hashtags sont très divers mais d'importances équivalentes : chaque vidéo inscrit des hashtags différents relatif à leur thème. Seuls les hashtags #foryou, #fyp, #foryoupage, #edit ou encore #funny sont mis en avant, des hashtags communs.

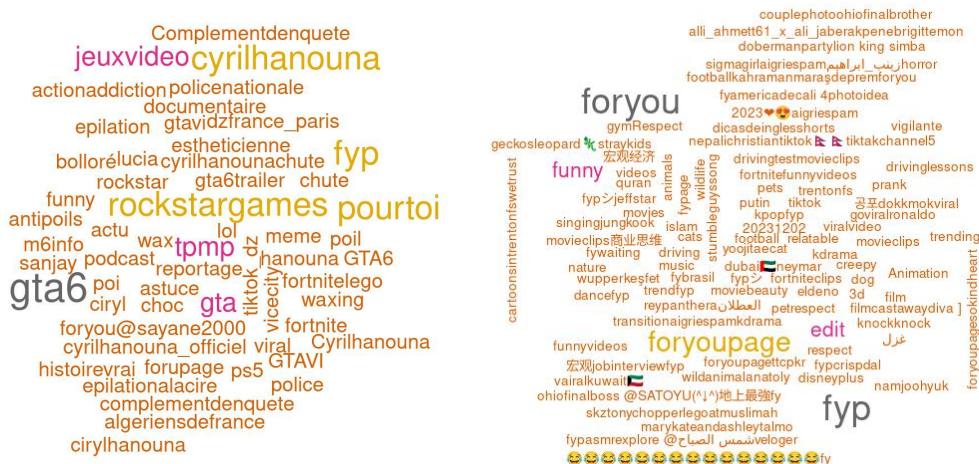


Figure II.2.3.1 : Nuage de mots tendances vs PourToi

On cherche maintenant à étudier les centres d'intérêts des utilisateurs ayant commenté les vidéos étudiées précédemment. Nous étudions donc les sujets les plus aimés par les utilisateurs ayant mis leurs j'aimes en public grâce à des nuages de mots.

Émeutes Paris

Pour une vidéo tendance comme celle des émeutes, on retrouve un large panel d'utilisateurs aux centres d'intérêts variés. Les hashtags de la vidéo en question ne correspondent pas aux sujets de prédilection des utilisateurs en commentaire.

La figure suivante représente les hashtags retrouvés dans les vidéos aimées par les utilisateurs ayant commenté la vidéo :

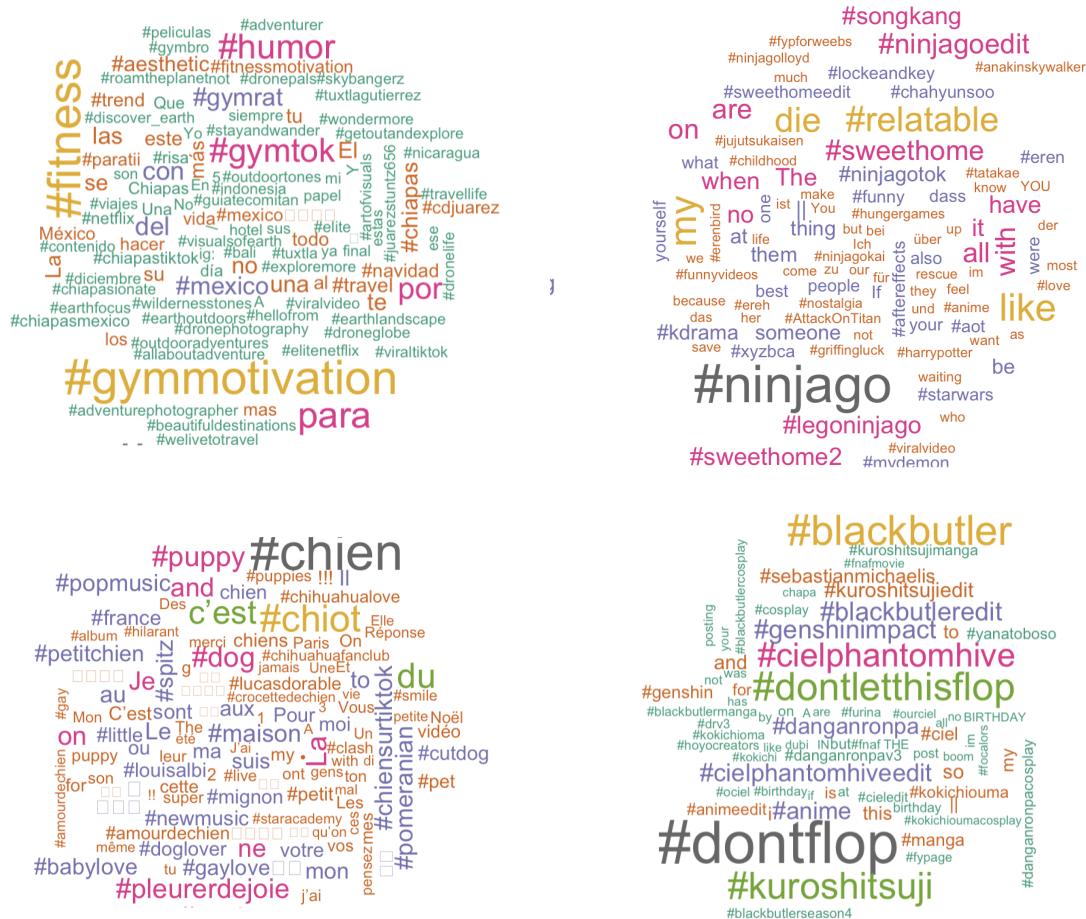


Figure II.2.3.2 : Centres d'intérêts d'utilisateurs ayant liké la vidéo émeutes

Cristiano Ronaldo

En revanche, pour une vidéo au profil comme celle de Ronaldo, on observe que les utilisateurs ont des centres d'intérêts directement liés au thème de la vidéo.

Dans les hashtags des utilisateurs commentateurs, on retrouve principalement les mêmes hashtags que ceux de la vidéo ou des hashtags du même thème. Cela peut s'expliquer par le sujet particulier de la vidéo, la présence de fans fidèles dans le milieu de football.

Ce qui confirme notre hypothèse précédente, la vidéo de Ronaldo continue de vivre grâce à une communauté qui consomme de nombreuses vidéos autour de ce centre d'intérêt.



Figure II.2.3.3 : Centres d'intérêts d'utilisateurs ayant liké la vidéo sur Ronaldo

Vidéos PourToi

On effectue la même démarche sur la vidéo du panda du fil PourToi. Ici, on constate la présence des deux phénomènes précédents. D'une part, certains utilisateurs ont des centres d'intérêts très proches du sujet de la vidéo, notamment des amateurs d'animaux et de pandas. D'autre part, d'autres utilisateurs ont des centres d'intérêts totalement différents comme fortnite ou les animés.

On peut donc supposer que l'algorithme de la plateforme propose aux utilisateurs de nouveaux sujets pour sans cesse ajuster leurs préférences. Pour ce faire, il présente des vidéos très appréciées par la communauté reliée à ce sujet.

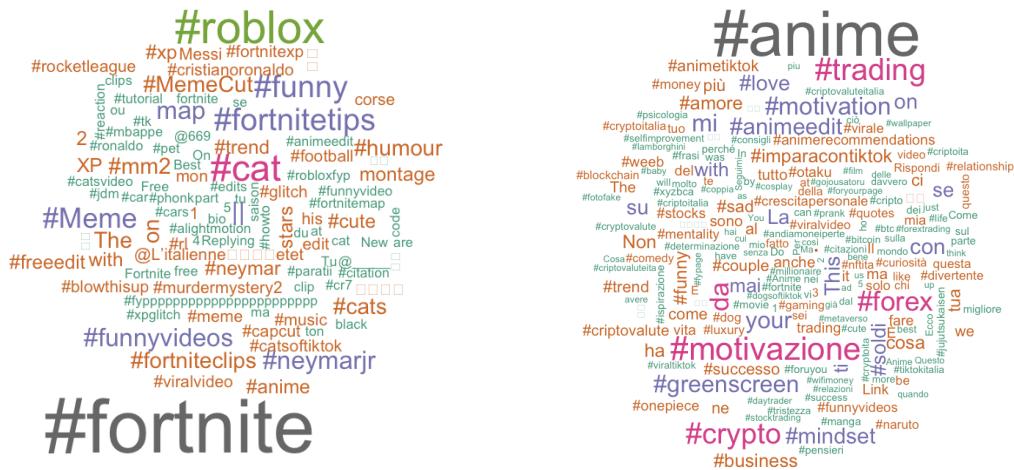
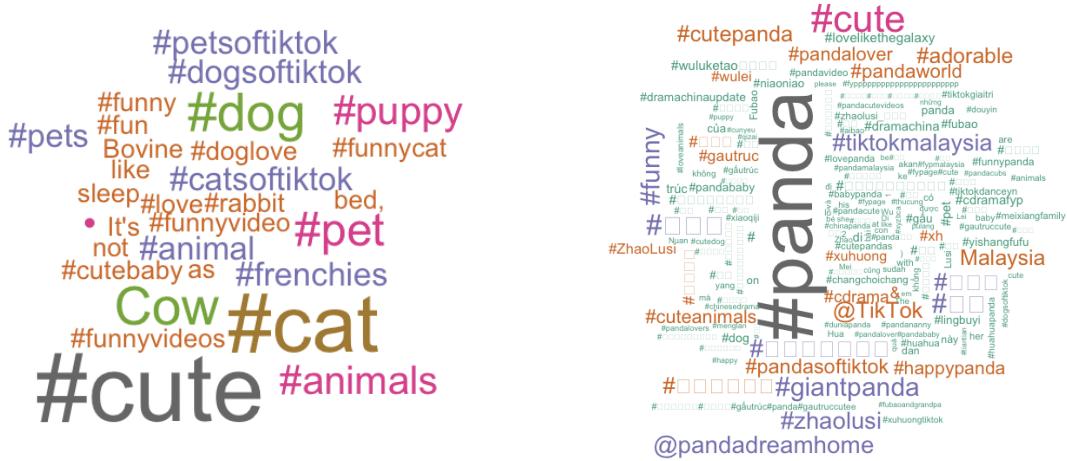


Figure II.2.3.4 : Centres d'intérêts d'utilisateurs ayant liké la vidéo panda

Conclusion

La récupération des données nous a permis d'explorer la structure web de la plateforme TikTok. Des obstacles restent infranchissables, ce qui rend le processus très long. Certaines vidéos sont parfois rendues indisponibles et perturbent le flux de récupération. Une métrique supplémentaire intéressante aurait été le nombre de vues, initialement non enregistré car indisponible sur les pages musiques. Sur les pages tendances, cette donnée apparaît. Un travail futur serait de l'inclure à la récupération et à l'analyse. De plus, il serait intéressant de développer une récupération des citations/mentions entre utilisateurs dans les commentaires.

L'analyse des données nous expose un comportement similaire entre les vidéos poussées en tendance par la plateforme : l'émergence d'un groupe de vidéos au sujet commun suivi d'une explosion de l'une de ces vidéos, une stagnation des autres et l'apparition de nouvelles vidéos en lien.

En comparant les statistiques des vidéos tendances à celles d'un échantillon PourToi, la métrique utile à la sélection des tendances par TikTok semble être le partage. Une vidéo avec un fort taux de partage (par rapport aux j'aimes) est plus à même d'être poussée en tendance. Cela recoupe la notion de viralité qui définit la diffusion rapide d'un contenu : une diffusion par la partage, rapide par un taux important. La fonction signet semble ignorée dans le choix des tendances puisque son taux est plus important chez l'échantillon PourToi. La représentation des commentaires en fonction des j'aimes ne nous donne pas de réponse concernant l'importance des commentaires mais l'étude de la temporalité des nous montre des pics d'engagement pour les vidéos tendances. Plus que la quantité, c'est la fréquence et l'intensité de l'engagement par les commentaires, mais aussi par les autres métriques non mesurables en temps, qui semble donc avoir un impact sur la mise en tendance des contenus. Si les tendances sont majoritairement des sujets d'actualité, le regain d'intérêt soudain pour un sujet peut le porter en tendances comme vu avec Cristiano Ronaldo. Enfin, l'audience des tendances est plus variée que celle des vidéos PourToi. Des utilisateurs sont donc poussés vers les contenus tendances indépendamment de leurs centres d'intérêt habituels. Pour résumer, nos hypothèses semblent se confirmer.

Cependant, il est important de nuancer les résultats obtenus sur plusieurs points. D'abord, la quantité de données analysée reste assez faible, avec seulement six vidéos par tendances, les statistiques descriptives sont à prendre avec des pincettes. Aussi, notre étude n'a pas pu observer l'évolution des métriques dans le temps et les vidéos tendances étant de nouvelles vidéos, les résultats numériques sont assez faibles et/ou variés suivant les tendances. Concernant l'audience, les utilisateurs ne donnant la plupart du temps pas un accès public aux vidéos qu'ils ont aimées, nous ne pouvons avoir la certitude que l'analyse des centres d'intérêts est représentative de tous les utilisateurs.

Pour conclure, il serait intéressant de poursuivre l'analyse temporelle. Les vidéos tendances sont ensuite reversées dans la masse de contenu de TikTok devenant à leur tour des vidéos PourToi, ont-elles des comportements différents de vidéos jamais passées par les tendances ?