

BRESSOT Arnaud 11505990  
CANETE Valentin 11502374  
GUEDET Adrien 11516525  
LOUM Mouhamed 11514727

Cahier d’avancement

**Analyse de consommation des vidéos pour adulte**  
**Groupe 7**

**Table des matières**

Du 14 au 21 Novembre.....2  
Du 22 Novembre au 2 Décembre.....2  
Du 3 au 6 Décembre.....3  
Du 7 au 12 Décembre.....3  
Du 13 au 24 Décembre.....4  
Du 27 Décembre au 3 Janvier.....4  
Du 3 au 8 Janvier.....4  
Du 10 au 17 Janvier.....5  
Annexe 1 – Graphes.....6  
Annexe 2 – Contributions.....8

## **Du 14 au 21 Novembre**

- Sujet trouvé en regardant les Pornhub Insight, des données sur les visiteurs du site. Nous voulions avoir des données temporelles avec les recherches, les pays et les profils des utilisateurs.
- Recherche d'API pour accéder à ces données aux données des sites Pornhub, Redtube, xVideo et Xhamster
- Recherche d'API sur les sites pour adulte en général et notamment PORN JSON
- Envoi de mail à Pornhub, xVideo et xHamster.

## **Du 22 Novembre au 2 Décembre**

- Sans réponse des sites après 2-3 semaines, nous avons décidé de renoncer à leur aide et de créer notre propre jeu de données à partir de Pornhub.
- Réalisation de notre propre créateur de données
  - Collecte d'un nombre de X pages de vidéos
    - URL, titre, nombre de vue, catégories, type de production et rating
  - Mise à jour de ces données
    - Actualisation du nombre de vue et du rating avec une date
- Mise en œuvre en utilisant un serveur personnel
  - Mise en place des scripts
  - Création des CRONs
  - Correction de bugs
  - Vérification de la collecte

## Du 3 au 6 Décembre

- Réflexion autour des questions qu'on souhaitait aborder
- Choix de 4 graphes à montrer (Annexe 1) :
  - Evolution des catégories
  - Evolution des vues
  - Evolution du nombre de vue en fonction des jours de la semaine
  - Pertinence des informations par rapport au nombre de vue
- Mise en place du document de cadrage

## Du 7 au 12 Décembre

- Début de la transformation des données pour correspondre aux diagrammes
- Pour la catégorie en fonction du temps :
  - Choix de compter les vues d'une vidéo dans chaque catégorie à laquelle elle appartient
  - Transformation en dictionnaire, sous cette forme :
    - {date1 : {categorie1 : nbVue1, categorie2: nbVue2...}, date2 : {...}...}
  - Recherche d'une représentation intéressante vers [ceci](#)
- Pour l'évolution du nombre de vue :
  - Nettoyage des données pour régler le bug des vidéos supprimées (et propagation)
  - Recherche d'une représentation pertinente
- Pour les jours de grande influence :
  - Décompte des vues en retirant les vues du jour d'avant
  - Transformation en dictionnaire, sous la forme :
    - {Monday : nbVue1, Tuesday, nbVue2..}
  - Tentative de prise en compte des fuseaux horaires → Echech

## Du 13 au 24 Décembre

- Réflexion autour de la pertinence de l'évolution des vues
- Réflexion autour des fluctuations du nombre de vue :
  - A partir d'une moyenne :
    - Moyenne = 10
    - Jour 1 = 11 → + 1
    - Jour 2 = 7 → - 3
  - A partir de la veille :
    - Jour 1 = 10
    - Jour 2 = 11 → + 1
    - Jour 3 = 7 → - 4
- Changement du système de compte en utilisant le nombre de vue par jour au lieu du nombre de vue cumulé

## Du 27 Décembre au 3 Janvier

- Réunion de finition pour terminer les graphes
- Distribution du travail pour finir :
  - Arnaud : Cahier
  - Adrien : Visualisation des 3 premiers graphes
  - Valentin : Mise en pratique du coté énergie
  - Mouhamed : Réflexion autour de la pertinence

## Du 3 au 8 Janvier

- Pistes sur la pertinence des vidéos
- Prise en compte du coût des vidéo
  - Dans la collecte
  - Avec un nouveau graphe
- Travail sur la présentation

## Du 10 au 17 Janvier

- Prise en compte des modifications à apporter :
  - Mise en place de sliders
  - Remise en contexte
  - Réorganisation des graphes
- Spécification du navigateur
  - Popup sur les navigateurs autre que Firefox
- Changement du dendrogramme
  - Transformation en catégorie des vidéos supprimées
- Préparation du rendu
  - Readme anglais/français
  - Pull requests
- Dernières modifications
  - Graphe de consommation d'énergies des vidéos supprimées
  - Fix de bugs graphiques

## Annexe 1 – Graphes

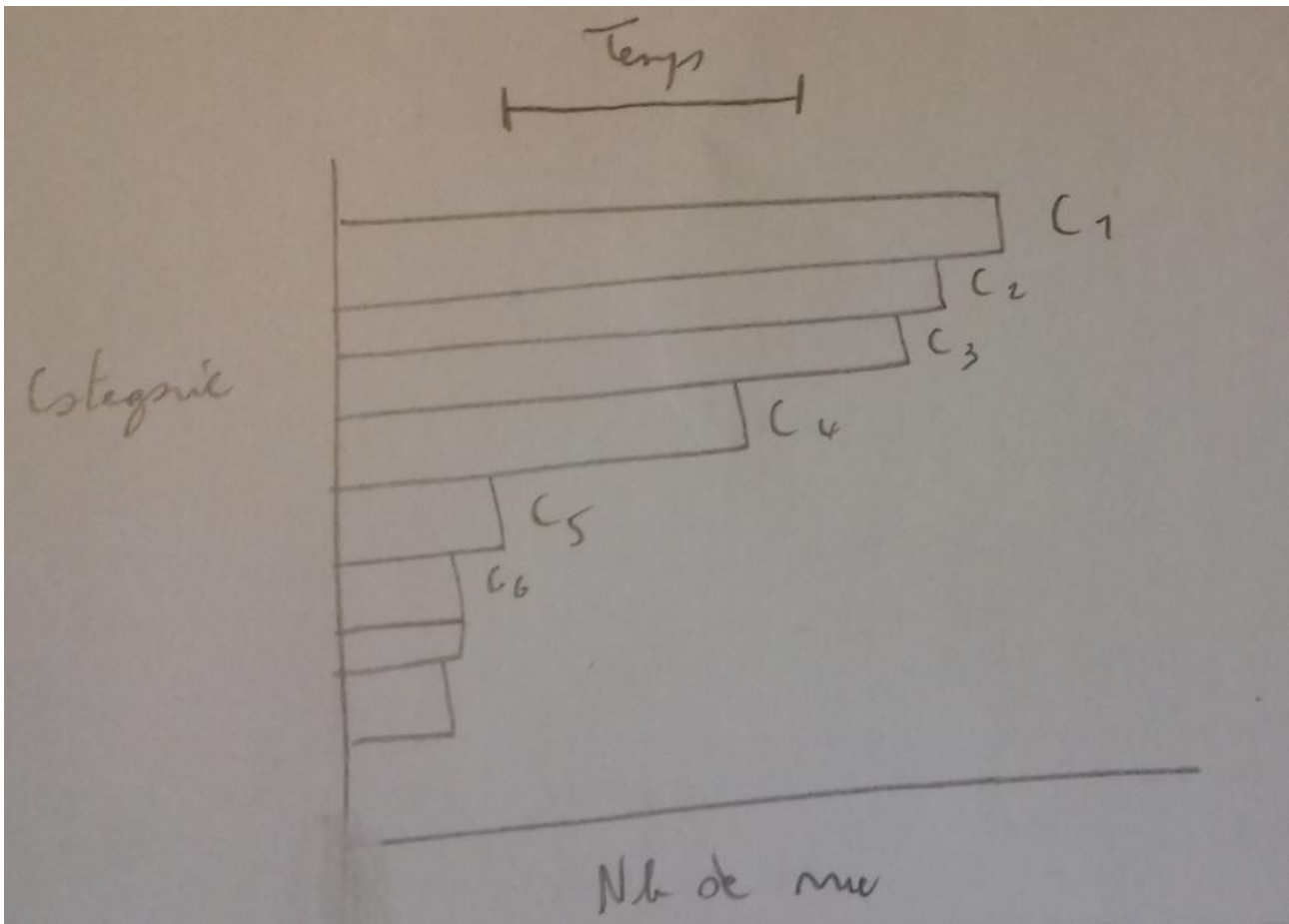


Figure 1: Classement des catégories en fonction des date

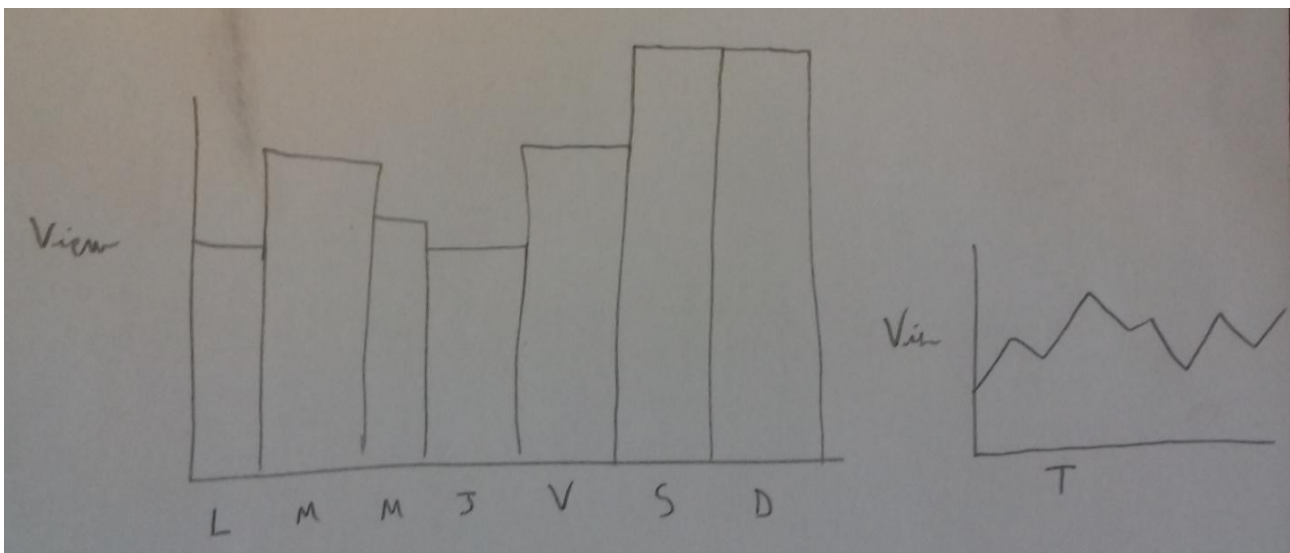


Figure 2: Evolution du nombre de vue en fonction du temps et des jours de la semaine

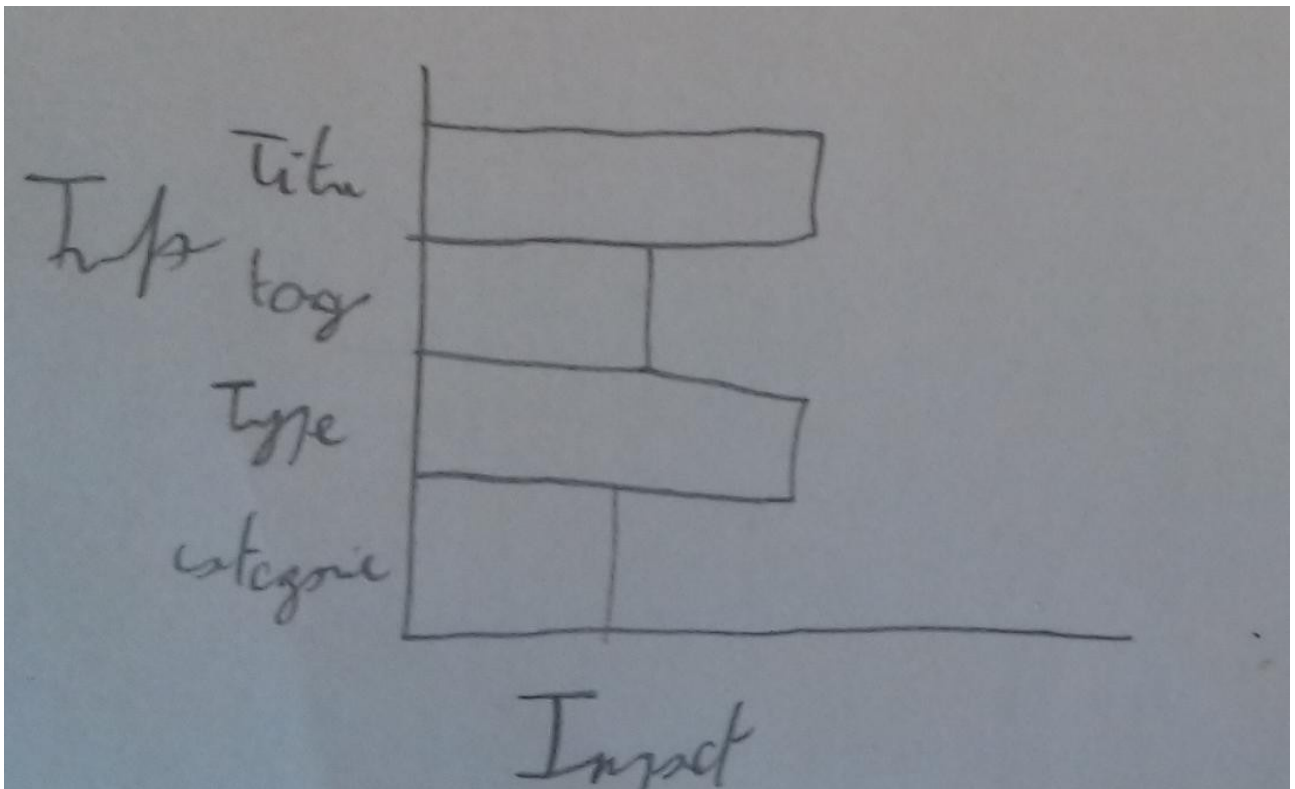


Figure 2: Figure 3: Graphe de pertinence des informations d'une vidéo

## Annexe 2 – Contributions

### Arnaud Bressot

- Organisation du projet
  - Programmation des réunions
  - Distribution des taches
  - Suivi de l'avancement
  - Documents administratifs
- Transformation des données
  - Création de fichiers exploitables
  - Calculs des coûts, des nombres de vues combinées
- Réalisation du graphe des vues par dates

### Valentin Canete

- Réalisation des scripts de collecte
  - Collectes des vues, des titres, des catégories
  - Enrichissement avec le coût et les tags
  - Réalisation du suivi en fonction du temps
- Réalisation du graphe de coût suivant la durée et le nombre de vue

### Adrien Guedet

- Réalisation du graphe du classement des catégories
  - Ajout des Sliders
  - Mise à jour dynamique du classement
  - Renommage des catégories family friendly
- Réalisation du graph du nombre de vue par jour de la semaine

### Mouhamed Loum

- Réalisation du Dendrogramme
- Réalisation du graphe de consommation d'énergie des vidéos supprimées