



# PROJET FINAL DE FORMATION DATA ANALYST

Etude prospective pour déterminer les ingrédients-clés pour  
développer un jeux vidéo à succès

# INTRODUCTION ET PROBLÉMATIQUE

Au sein d'un studio de développement indépendant

## LE CONTEXTE :

- Données issues de deux sociétés :
  - Twitch, plateforme de streaming live détenue par Amazon
  - Steam, plateforme de vente de jeux vidéos possédée par la société Valve

## La problématique :

- Quels sont les types de jeux les plus joués ou les plus streamés ?
- Peut-on considérer que les types de jeux populaires aujourd'hui le seront demain?

# DONNÉES ET PRÉPARATION

L'importance de savoir relier deux bases venant de sociétés différentes

# PRÉPARATION DU FICHER TWITCH

- Suppression des colonnes inutiles pour ne conserver que :
  - Le nom du jeu ;
  - Les mois et année ;
  - Les heures visionnées et diffusées.
- Mise en format des noms de jeux : passage intégrale en majuscules pour faire le lien avec le fichier Steam.
- Ajout d'une colonne qui concatène le mois et l'année pour faciliter la création de graphiques chronologiques.

Game ▾	Month ▾	Year ▾	Hours_watched ▾	Hours_streamed ▾	Month&Year ▾
HALF LIFE 2: SURVIVOR	1	2016	400106	195	vendredi 1 janvier 2016
CRASH BANDICOOT	1	2016	399886	501	vendredi 1 janvier 2016

# PRÉPARATION DU FICHIER STEAM

- Travaux sur les colonnes catégories, genres, tags pour déterminer meilleure clé : le tag n'étant constitué que de 3 éléments (contre 7 pour le genre et une quinzaine pour la catégorie), c'est cet élément qui a été retenu.
- Suppression des colonnes inutiles (code identifiant le jeu, la disponibilité de l'anglais, l'âge requis, le prix, le temps médian de jeu et les catégories et genres puisque nous avons préféré le tag).
- Mise en format des noms de jeux : passage intégrale en majuscules puis suppression des doublons pour faire le lien avec le fichier Twitch.
- Ajout d'une colonne pour calculer le temps moyen de jeu en heures (la colonne originelle étant exprimée en minutes).

name	release_date	developer	publisher	steamspy_tags	achievements	positive_ratings	negative_ratings	average_playtime	owners	TempsMoyenHeures
VINCERE TOTUS ASTRUM	lundi 20 février 2017	Gamesare	Clickteam	Strategy	0	1	0	0	0-20000	0
BLESSED ONES: THE MAGIC WOLVES	vendredi 4 décembre 2015	Wolf Ray Studios	Wolf Ray Studios	Early Access;RPG;Indie	0	1	0	0	0-20000	0
CAMPGROUNDS: THE ENDORUS EXPEDITION COLLECTOR'S EDITION	jeudi 3 mars 2016	Jumb-O-Fun	Big Fish Games	Casual;Simulation;Strategy	0	1	0	0	0-20000	0

# TRAVAUX ANNEXES

- Split des tags :
  - Grâce au séparateur présent dans le champ « steamspy\_tag », nous avons pu séparer les 3 valeurs de tag.
  - Puis avec un pivot de colonne, nous avons créé une table de référence pour associer les 3 valeurs à chaque clé « steamspy\_tag ».
- Création d'une base de jeux communs :
  - Avec une fusion de requête basée sur le nom du jeu qui ne conserve que les éléments communs, nous avons créé une base de 614 jeux.
  - Avec l'ajout d'un flag « True », cela nous permet de n'afficher que les jeux en commun dans nos travaux.

steamspy_tags	Attribut	Valeur
Indie;Story Rich;First-Person	tag-1	Indie
Strategy;Indie;RTS	tag-2	Indie

name.steam	Twitch_game_data.Game	Commun
COUNTER-STRIKE	COUNTER-STRIKE	True
COUNTER-STRIKE: GLOBAL OFFENSIVE	COUNTER-STRIKE: GLOBAL OFFENSIVE	True
DOTA 2	DOTA 2	True

# ANALYSE EXPLORATOIRE

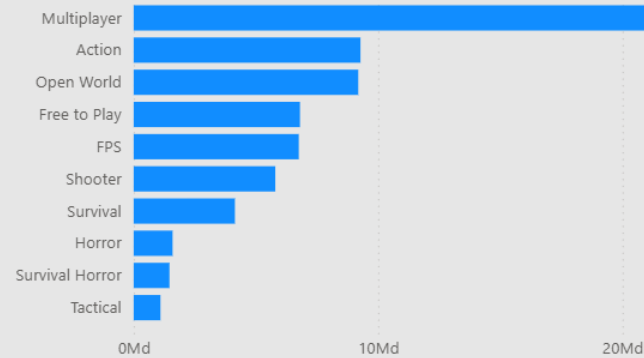
Des tags généraux aux cas particuliers



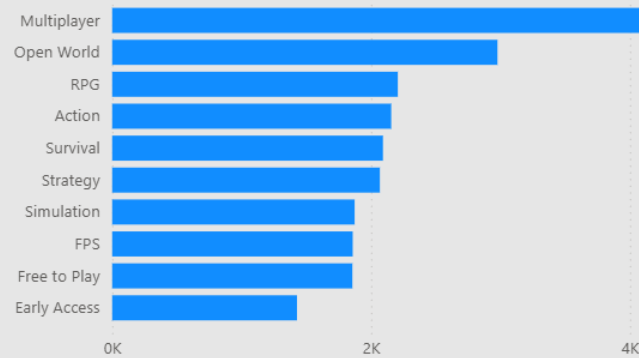
# Analyse de différents Tops 10

## Tops 10 de popularité des tags

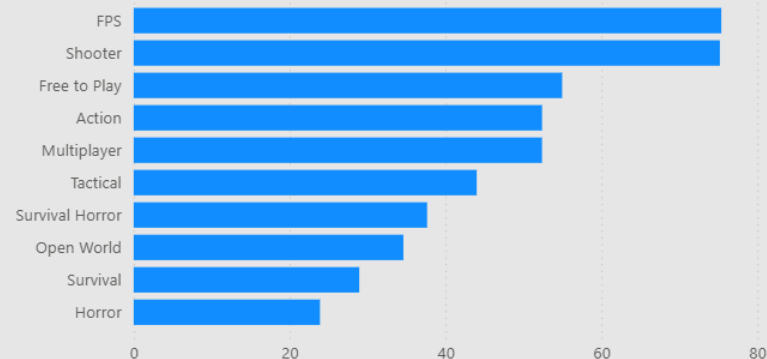
Top 10 des tags les plus regardés sur Twitch (en heures)



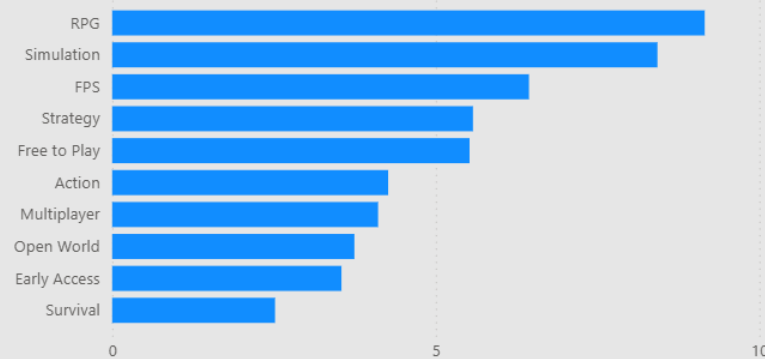
Top 10 du temps de jeu moyen Steam par tags (en heures)



Top 10 des tags avec le plus d'Achievements en moyenne



Top 10 des tags les plus appréciés (ratio avis positifs/négatifs)



- On remarque que, bien que les classements ne soient pas strictement identiques, on retrouve facilement les mêmes tags en tête.

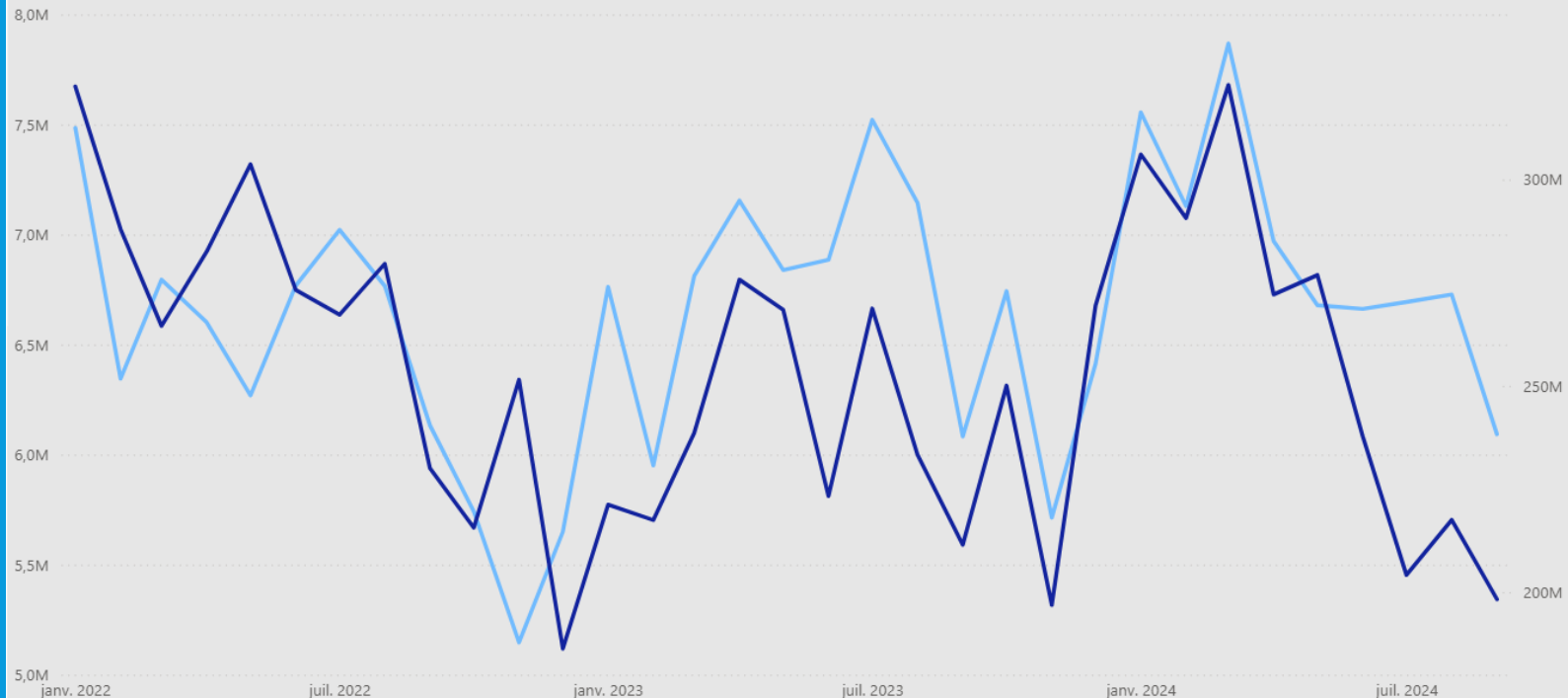
# Focus sur l'évolution du Multijoueur

## Evolution de l'audience Twitch sur le tag\_:

### Multiplayer

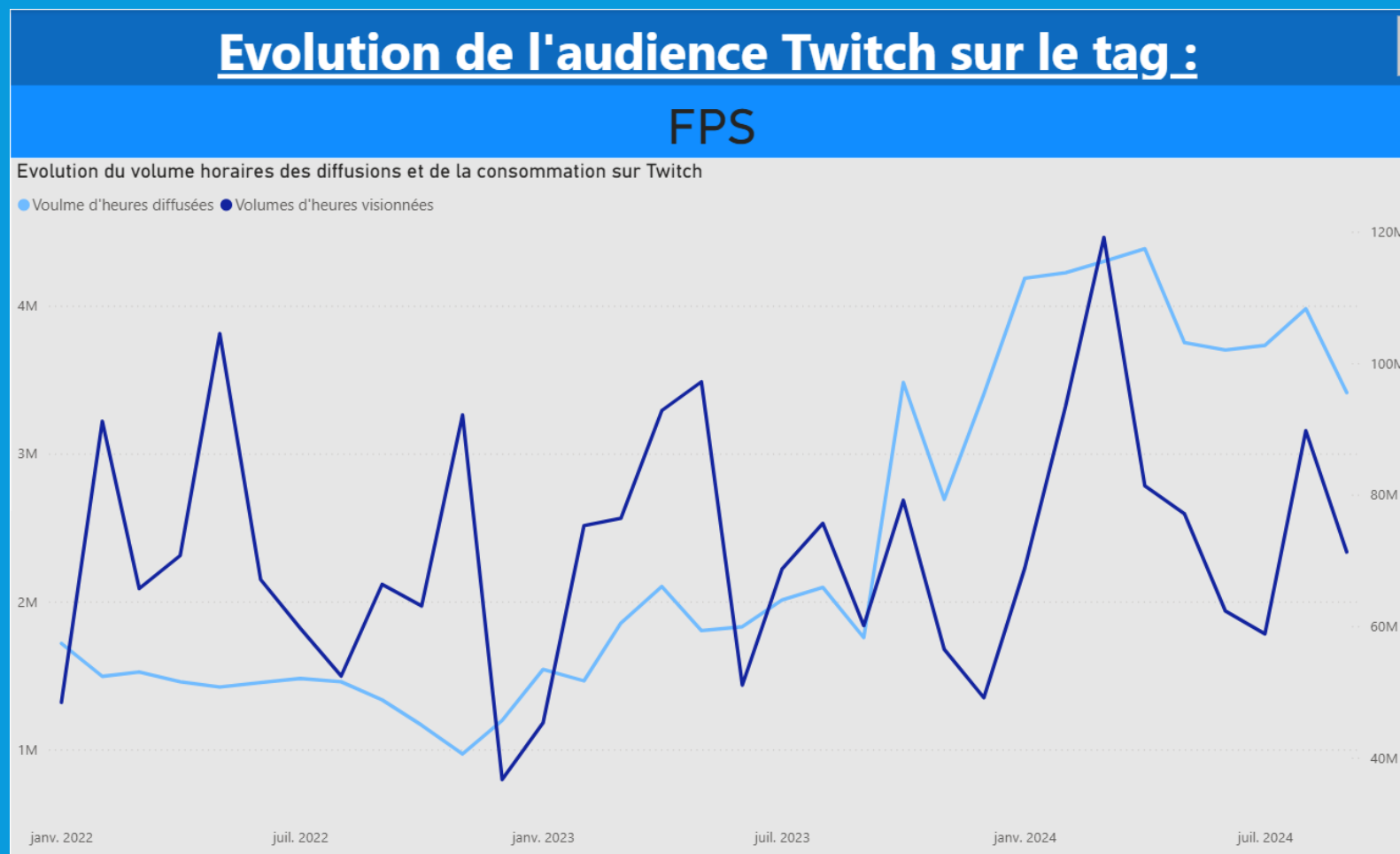
Evolution du volume horaires des diffusions et de la consommation sur Twitch

● Volume d'heures diffusées ● Volumes d'heures visionnées



- Premier des tops 10 Twitch et Steam
- L'activité n'est pas stable mais elle reste forte autour de 200 à 300 millions d'heures visionnées par mois.
- Le volume d'heures visionnées semble proportionnel à celui des heures diffusées.

# Focus sur l'évolution du FPS



- Premier du top 10 Achievements
- Ce tag génère entre 40 et 120 millions d'heures visionnées par mois.
- L'activité n'est pas stable et en plus, elle semble réagir peu à l'offre : depuis le second semestre, le volume diffusé a triplé mais le volume consommé n'a pas suivi.

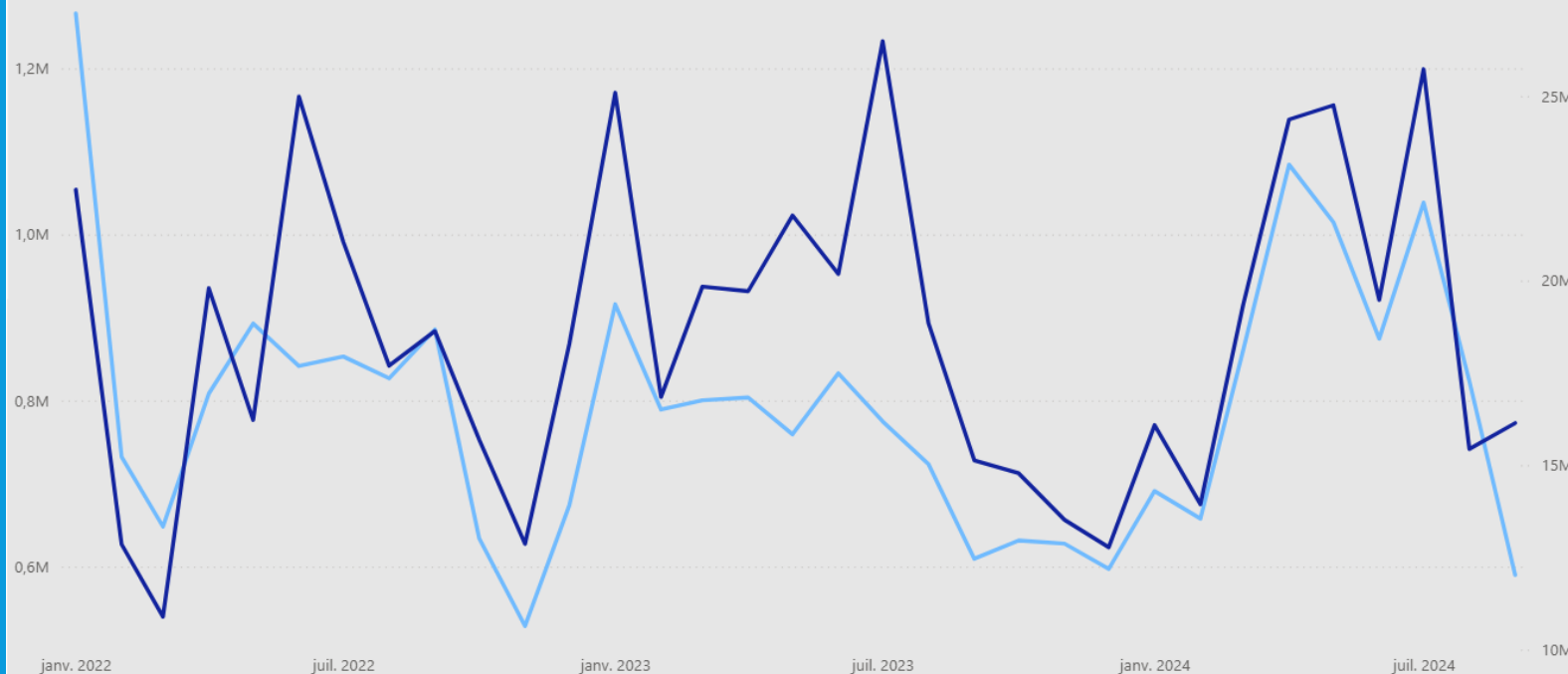
# Focus sur l'évolution du RPG

## Evolution de l'audience Twitch sur le tag\_:

### RPG

Evolution du volume horaires des diffusions et de la consommation sur Twitch

● Volume d'heures diffusées ● Volumes d'heures visionnées

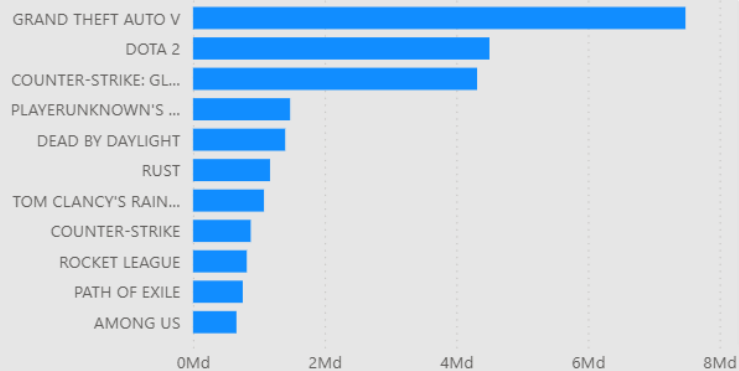


- Premier du top 10 du ratio avis positifs/négatifs.
- Ce tag génère entre 15 et 25 millions d'heures visionnées par mois.
- L'activité n'est pas stable mais il semblerait que la consommation réponde facile à la diffusion voir augmente parfois spontanément

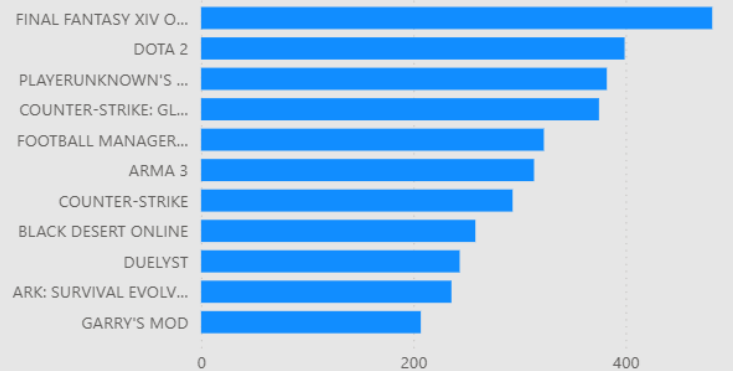
# Analyse par jeu

## Classements des jeux les plus populaires

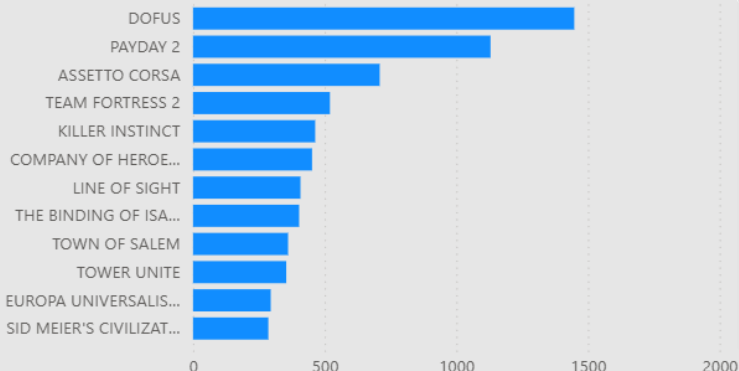
Nombre d'heures de visionnage sur Twitch par jeu



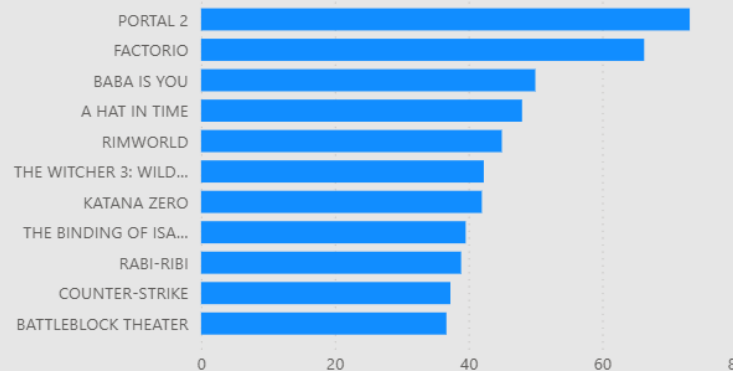
Temps moyen de jeu sur Steam (en heures)



Nombre d'"Achievements" par jeu

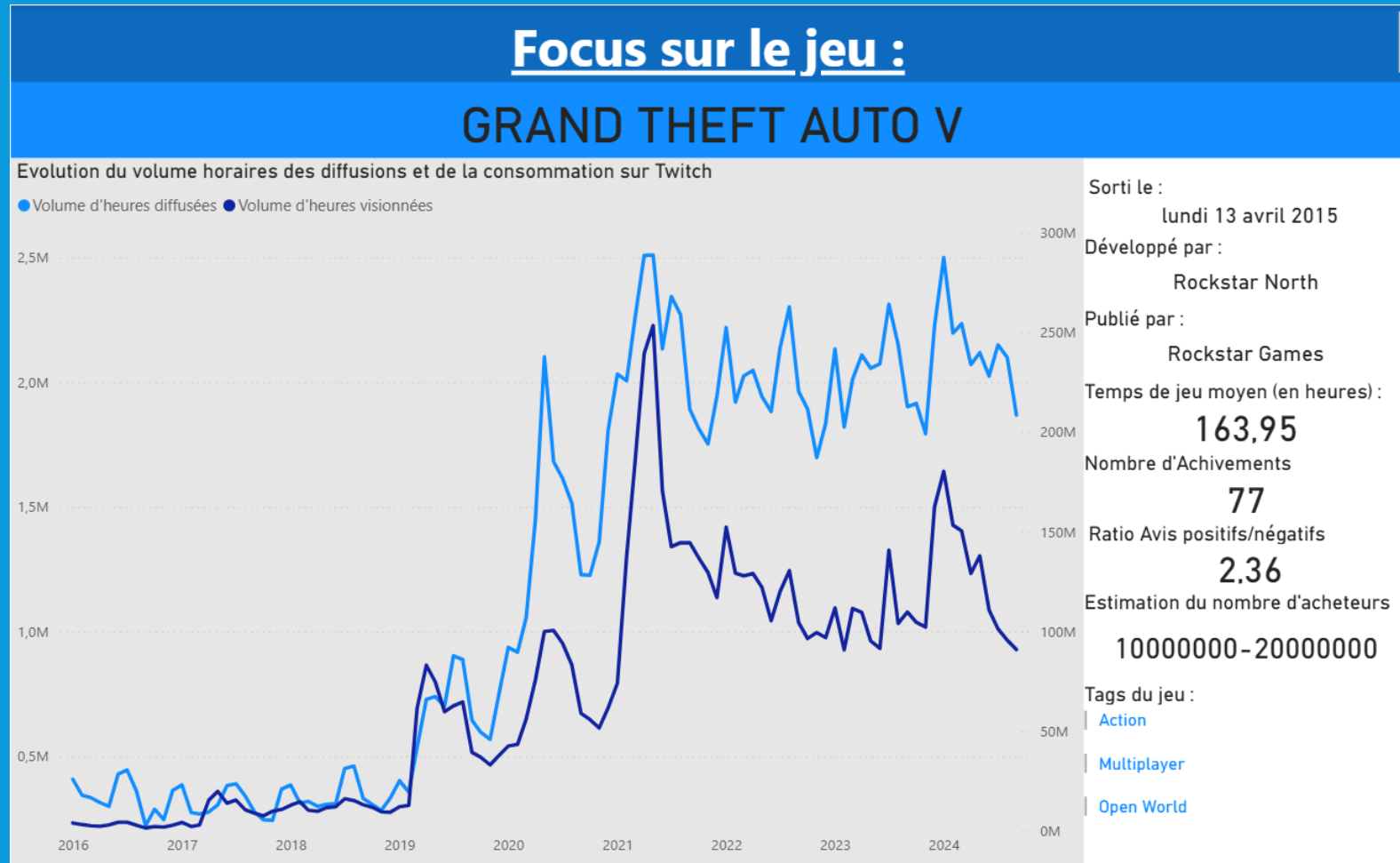


Ratio du nombre d'avis positifs sur le nombre d'avis négatifs



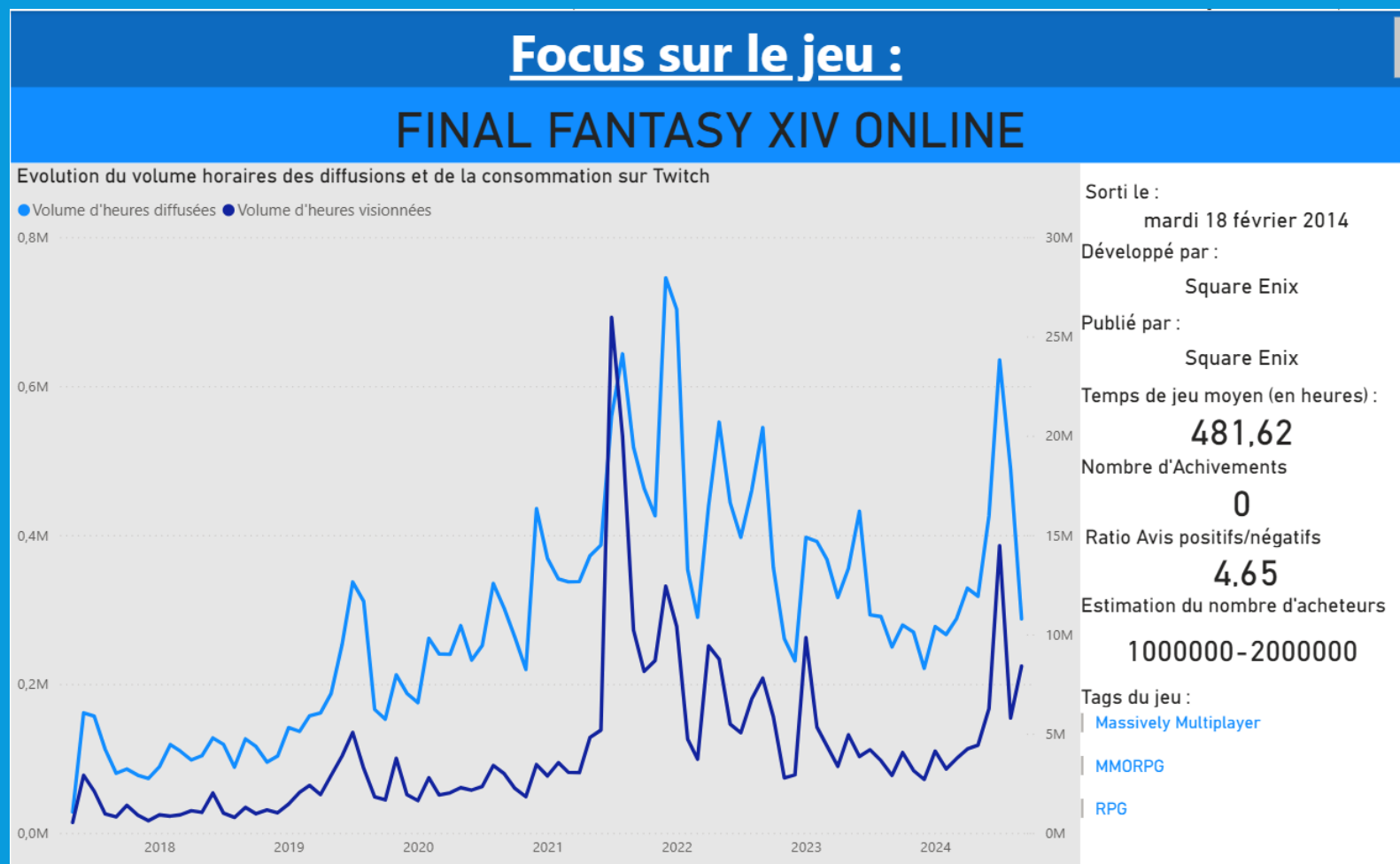
- Pas de corrélation aussi évidente entre nos 4 indicateurs qu'avec les tags mais on peut mettre cela sur le fait qu'il y a beaucoup de jeux à analyser.
- Afin d'éviter de s'éparpiller, nous allons nous concentrer sur les n°1 mais aussi les n°2 à 4 du temps de visionnage Twitch et du temps moyen de jeu Steam pour la suite de l'analyse.

# Le plus diffusé sur Twitch



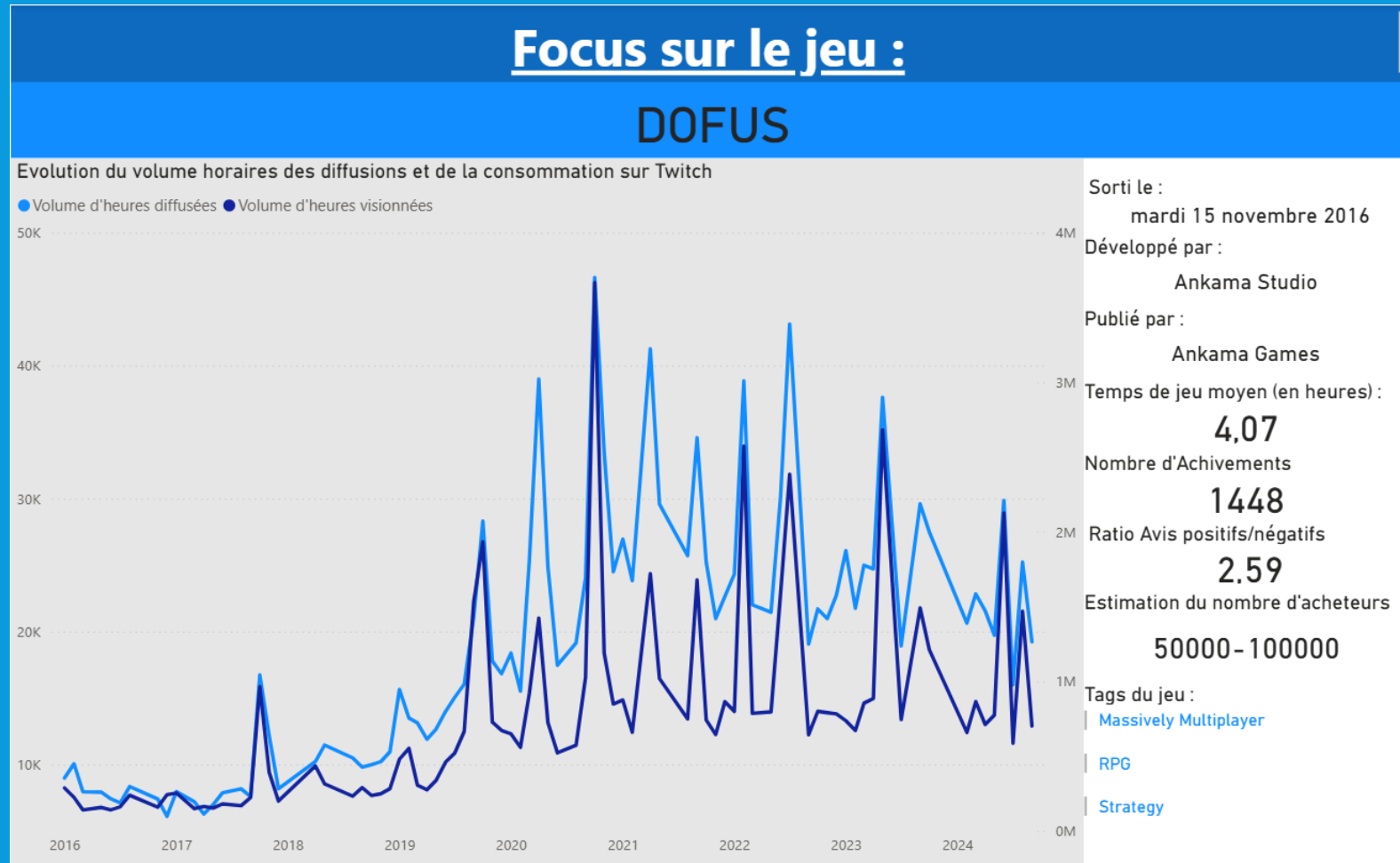
- Pic d'activité corrélé à la crise Covid-19
- Depuis 2022, le volume visionné ne suit pas le volume diffusé
- Les volumes de diffusions et de visionnage ne sont pas revenus à leurs niveaux d'avant la crise Covid-19

# Le plus joué sur Steam



- Pas d'achievments et c'est seulement le 12<sup>ème</sup> jeu le plus diffusé sur Twitch.
- Le jeu a connu plusieurs événements, notamment en 2021 (Fan Festival « Around the World », Commémoration des 8 ans du jeu solo,...) qui correspondent à certains pics d'activité Twitch.

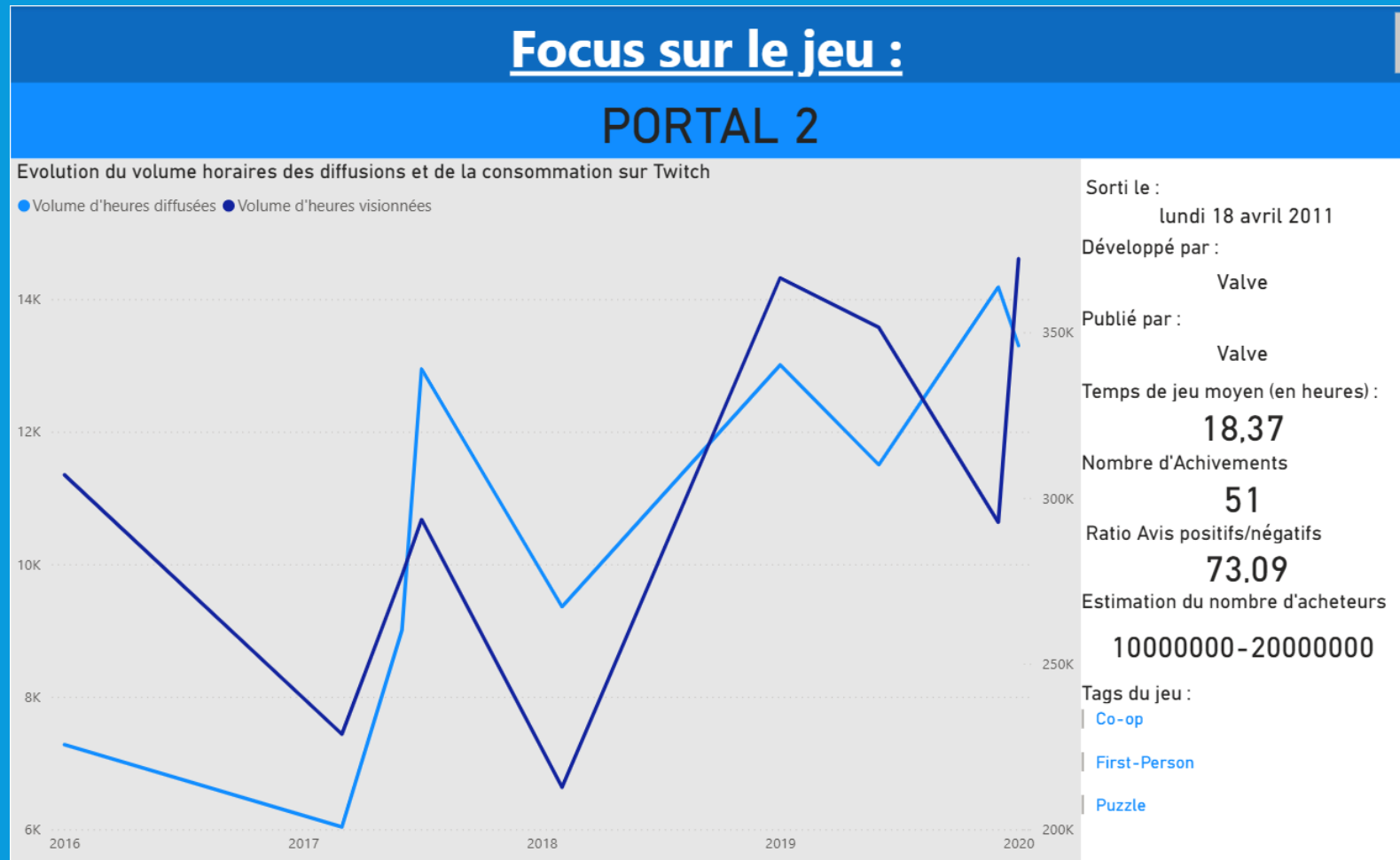
# Le plus difficile à finir



- Un record du nombre d'achivements mais une durée moyenne de jeu très faible (donc probablement peu de jeu qui les accomplissent)
- Il semblerait qu'il y ait beaucoup d'événements éphémères sur le jeu, cela pourrait expliquer les pics fréquents d'activité

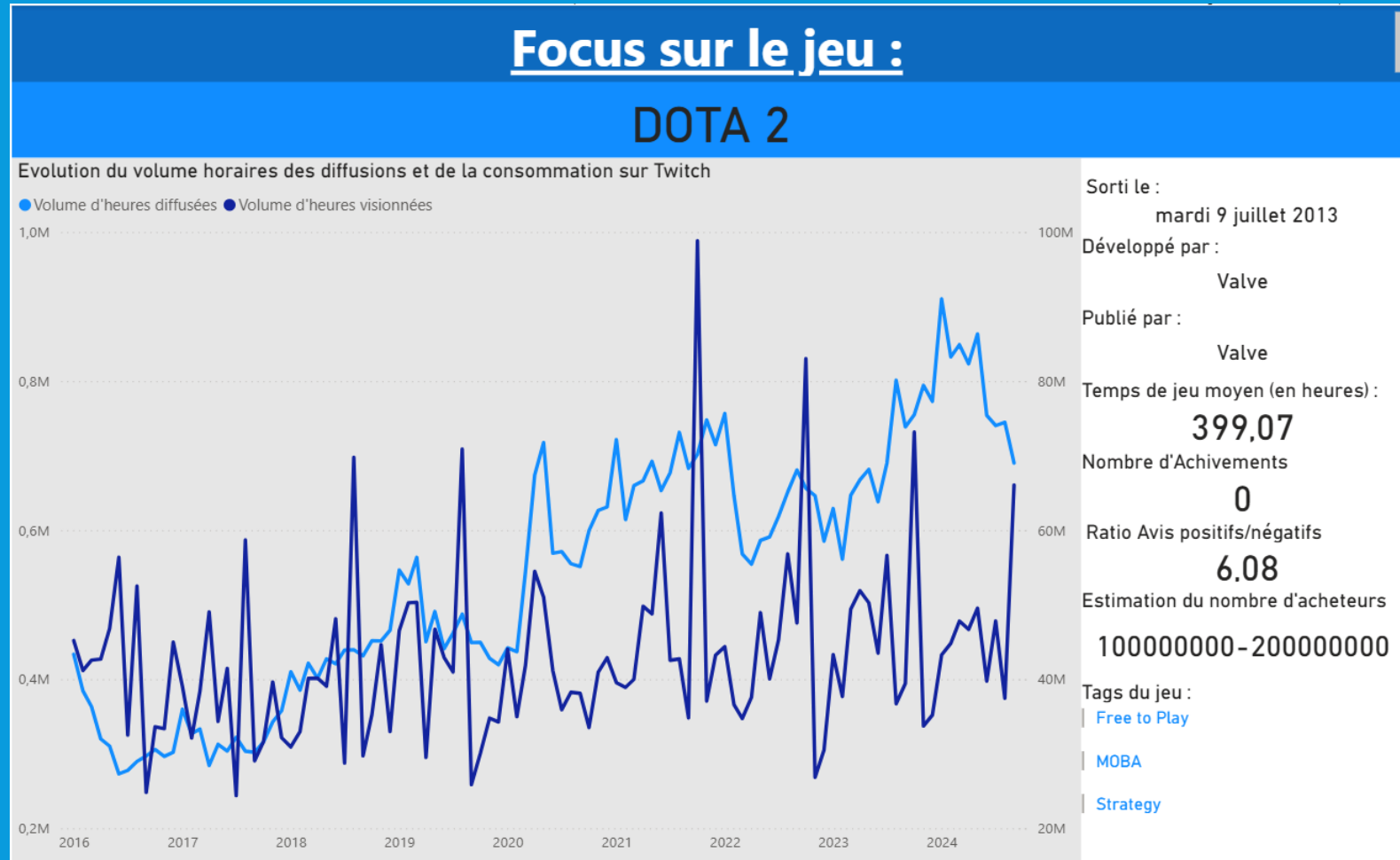


# Le plus aimé



- Pas de donnée sur une diffusion twitch à partir de 2021 malgré le record du ratio d'avis positifs/négatifs.
- S'agissant d'un jeu à énigme qui peut se finir en une dizaine d'heures, seule une petite communauté de speedrunner a continué à y jouer.

# Le numéro 1 de l'E-sport



- Référence de l'E-sport disponible en Free-to-play.
- Les prix sont parfois d'une valeur dépassant le million de dollars, ce qui explique les pics d'activité plus ou moins important et le fait que le jeu soit toujours autant joué.

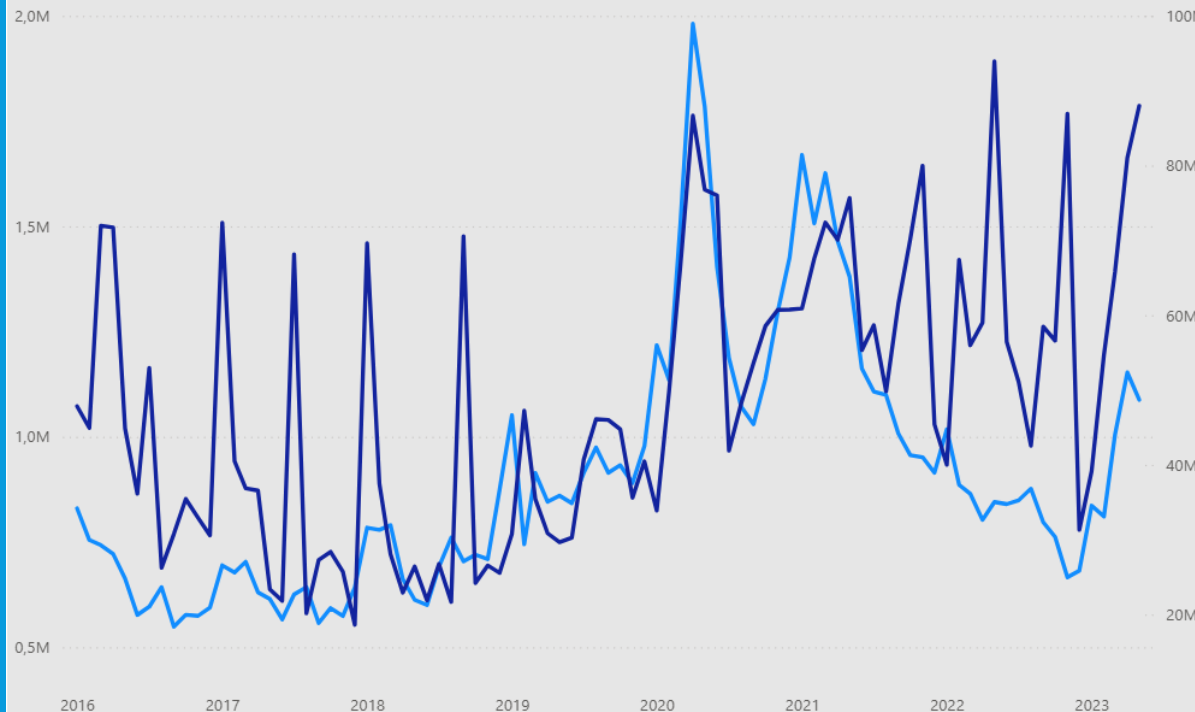
# La référence qui dure

## Focus sur le jeu :

### COUNTER-STRIKE: GLOBAL OFFENSIVE

Evolution du volume horaires des diffusions et de la consommation sur Twitch

● Volume d'heures diffusées ● Volume d'heures visionnées



Sorti le :

mardi 21 août 2012

Développé par :

Valve;Hidden Path Entertainment

Publié par :

Valve

Temps de jeu moyen (en heures) :

**374,90**

Nombre d'Achivements

**167**

Ratio Avis positifs/négatifs

**6.57**

Estimation du nombre d'acheteurs

**50000000 - 100000000**

Tags du jeu :

FPS

Multiplayer

Shooter

- Autre référence de l'E-sport avec des prix allant jusqu'à 25 000 dollars (expliquant là aussi les différents pics d'activité)
- La première version de Counter-Strike est sortie en novembre 2000

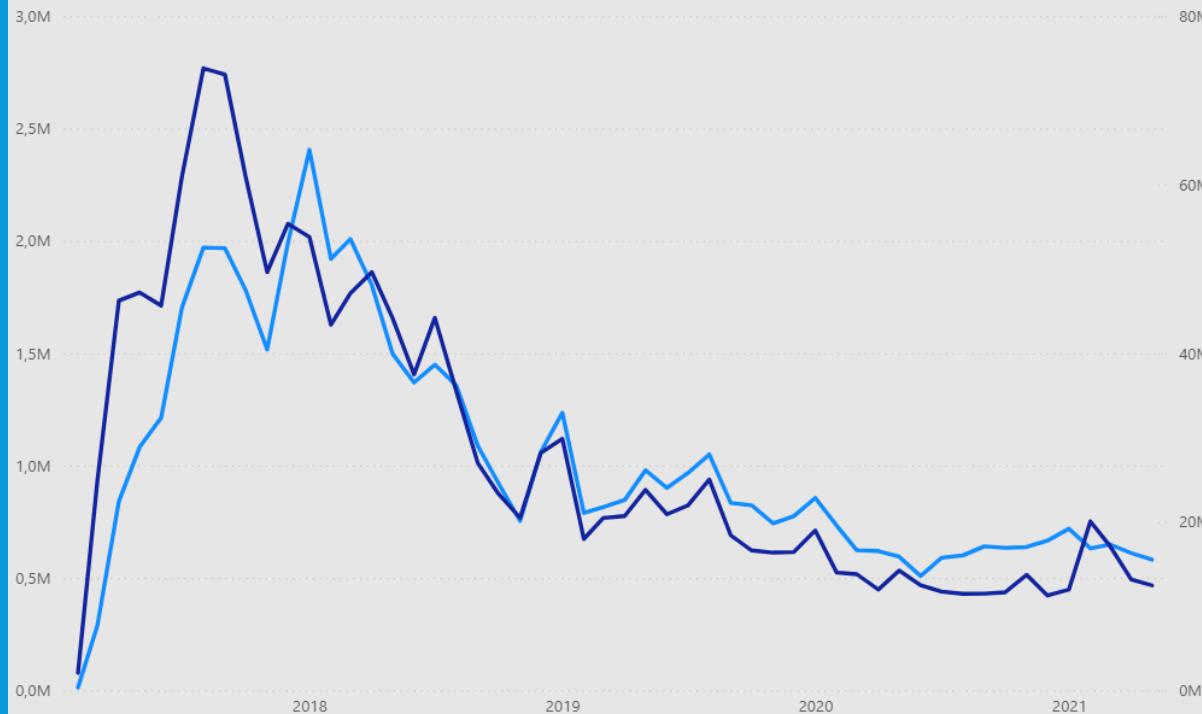
# Le succès tombé dans l'oubli

## Focus sur le jeu :

### PLAYERUNKNOWN'S BATTLEGROUNDS

Evolution du volume horaires des diffusions et de la consommation sur Twitch

● Volume d'heures diffusées ● Volume d'heures visionnées



Sorti le :  
jeudi 21 décembre 2017

Développé par :  
PUBG Corporation

Publié par :  
PUBG Corporation

Temps de jeu moyen (en heures) :  
**382,30**

Nombre d'Achivements  
**37**

Ratio Avis positifs/négatifs  
**1.02**

Estimation du nombre d'acheteurs  
**50000000 - 100000000**

Tags du jeu :

Multiplayer  
Shooter  
Survival

- Jeu très populaire à sa sortie mais aussi vivement critiqué (presque autant d'avis négatifs que positifs)
- Il y a toujours des événements en 2025 mais il semblerait que ceux-ci soient moins suivis (pas de données Twitch à partir de 2022)

# RECOMMANDATIONS BUSINESS

Une recommandation et des petits conseils pour aider

# MISER SUR LE MULTIJOUEUR EN PRIORITÉ

- Le multijoueur semble un élément incontournable pour générer du trafic (que ce soit en temps de jeu Steam ou de visibilité sur Twitch).
- Certaines méthodes peuvent expliquer ce succès comme les événements E-sportif (exemples de CSGO et Dota 2).
- Prenons aussi en compte que la rejouabilité du multijoueur aide à augmenter la durée de vie d'un jeu :
  - Les énigmes de Portal 2 sont finies (une fois résolu, elles perdent de leurs intérêts) ;
  - Les matchs de CSGO ou Dota 2, bien qu'étant très similaires (même carte, mêmes objectifs), varient dans leur déroulement (les joueurs peuvent tenter des tactiques différentes ou avoir des performances variables).

# LES PETITS PLUS QUI AIDENT

- Faire de la qualité pour durer :
  - Même si l'exemple de Portal 2 semble indiqué qu'être le meilleur jeu du monde, ne suffit pas à rester attractif, celui de PUBG semble montré qu'un échec critique est difficile à rattraper (son ratio avis positifs/négatifs est 6 fois inférieur à celui de CSGO ou Dota 2).
- Attention à l'over-achievements :
  - Le cas de Dofus semble indiquer que mettre beaucoup de choses dans un jeu ne suffit pas à le rendre populaire. Eviter de surcharger le jeu peut aussi permettre de dégager du temps pour peaufiner ce qui est déjà en place.
- Les autres genres récurrents (avec leurs avantages et inconvénients) :
  - GTA 5 est un monde ouvert qui mise sur l'action (Open-World) ;
  - FF XIV est un RPG qui demande du temps pour construire son personnage ;
  - Dota 2 est gratuit à l'achat mais propose un système d'abonnement a posteriori (Free-to-Play);
  - CSGO est un FPS immersif ;
  - ...

# SOURCES

- Steam Dataset // multiples fichiers : <https://www.kaggle.com/datasets/nikdavis/steam-store-games>
- Top Jeux 2016-2023 // Twitch\_game\_data.csv : <https://www.kaggle.com/datasets/rankirsh/evolution-of-top-games-on-twitch>
- Site officiel de Dota 2 : <https://www.dota2.com>
- Le calendrier des compétitions CSGO : <https://www.hltv.org/events>
- Les temps en speedrun de Portal 2 : [https://www.speedrun.com/portal\\_2](https://www.speedrun.com/portal_2)



# GLOSSAIRE

- Tag : caractéristique phare d'un jeu vidéo
- Streaming : diffusion d'un contenu vidéo en ligne
- Achievement : jalon déterminé par les développeurs pour déterminé l'avancement d'un joueur dans le jeu
- Speedrun : activité consistant à finir un jeu le plus vite possible
- E-sport : activité de compétition autour du jeu vidéo