

# L'Industrie du Jeux-Vidéo (1980-2023)

## I. Introduction

### Motivation de la thématique choisie

L'industrie du jeu vidéo a connu une transformation radicale au cours des quarante dernières années, passant d'un marché de niche à une industrie culturelle dominante. Cette évolution soulève des questions fondamentales sur les dynamiques qui régissent le marché actuel, notamment l'équilibre entre la course au volume et l'exigence de qualité. Pour répondre à ces interrogations, ce rapport s'appuie sur une base de données historique couvrant la période 1980-2024. En croisant les volumes de sorties, les notes critiques via le Metascore, les temps de jeu et les stratégies des studios, l'analyse tente de déchiffrer les grandes tendances structurelles du secteur ainsi que l'évolution des goûts des joueurs.

A travers ce projet nous avons cherché en quoi l'industrie vidéoludique a-t-elle évolué entre la course au volume et l'exigence de qualité, et quelles dynamiques structurelles (genres, saisonnalité, acteurs clés) régissent aujourd'hui ce marché ?

### Liste des sites web choisis

La collecte des données a été réalisée à partir de la plateforme de référence RAWG.io, choisie pour sa fiabilité concernant les titres, les dates de sortie, les notes de presse et les durées de jeu estimées par la communauté.

### Légalité et éthique

Ce processus d'extraction a été mené dans un cadre éthique et légal strict, respectant les fichiers robots.txt des sites cibles et veillant à ne pas saturer les serveurs par des requêtes trop fréquentes. De plus, les données ont été exploitées uniquement à des fins pédagogiques, en prenant soin de citer les sources pour respecter la propriété intellectuelle.

## II. Analyse, Traitement et Visualisation

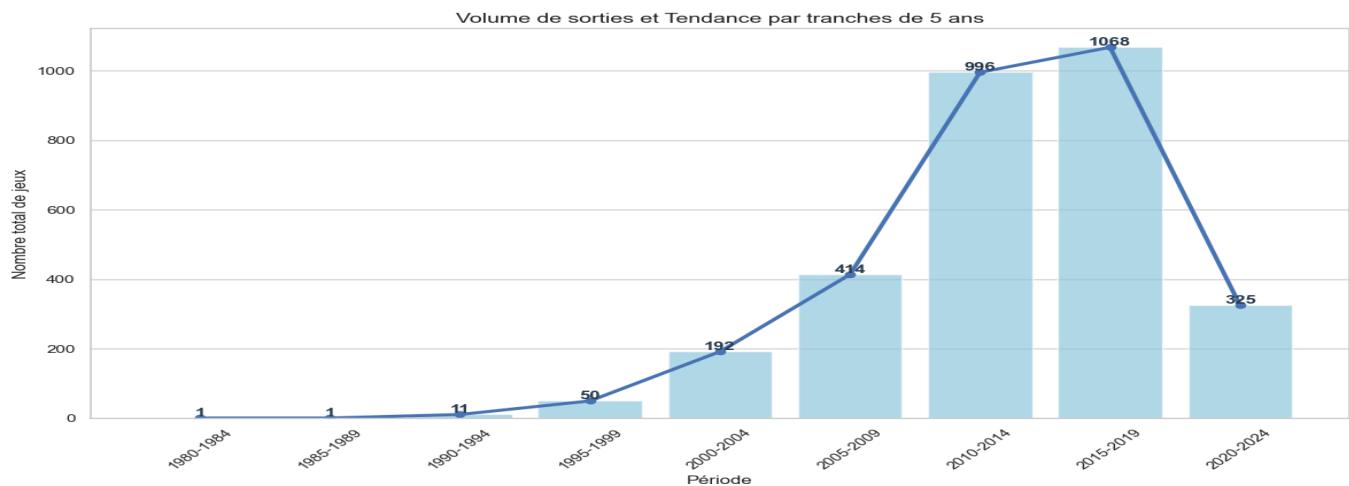
### Analyse et extraction

Sur le plan technique, le traitement des données a suivi un protocole rigoureux d'extraction, de nettoyage et d'exportation en trois phases. Un script Python (Selenium + BeautifulSoup) a d'abord parcouru automatiquement rawg.io pour collecter les URLs de jeux, puis chaque jeu a été visité individuellement pour extraire 8 champs d'information avec application d'un filtre strict (rejet de tout jeu incomplet à 100%), et enfin les données validées ont été standardisées (normalisation des dates, traduction en français) et exportées en format Excel avec génération d'un rapport d'extraction détaillé. Des sauvegardes automatiques tous les 200 jeux garantissent la préservation des données en cas d'interruption. Le processus a débuté par l'extraction des variables clés telles que le titre, le Metascore, l'éditeur et le développeur.

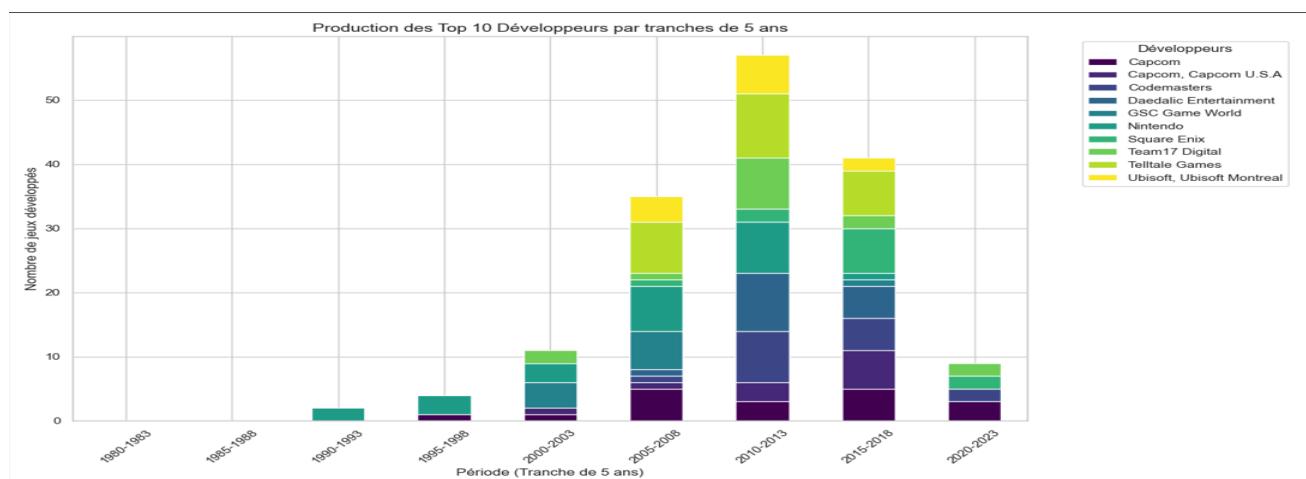
## Traitement et export

La base de données brute a ensuite nécessité un nettoyage approfondi pour harmoniser les colonnes, notamment en traduisant les dates et les genres en version française. Une étape cruciale du traitement sous Python a consisté à gérer les valeurs manquantes pour éviter de fausser les calculs de moyennes. Finalement, les données nettoyées ont été exportées au format CSV avec un encodage latin1 afin de préserver les caractères spéciaux français.

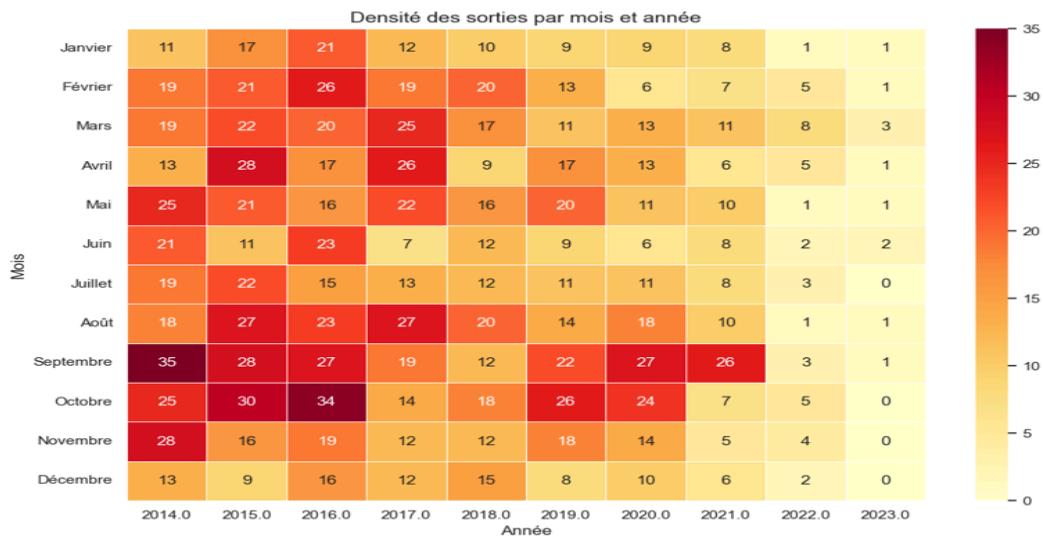
## Visualisation et Interprétation



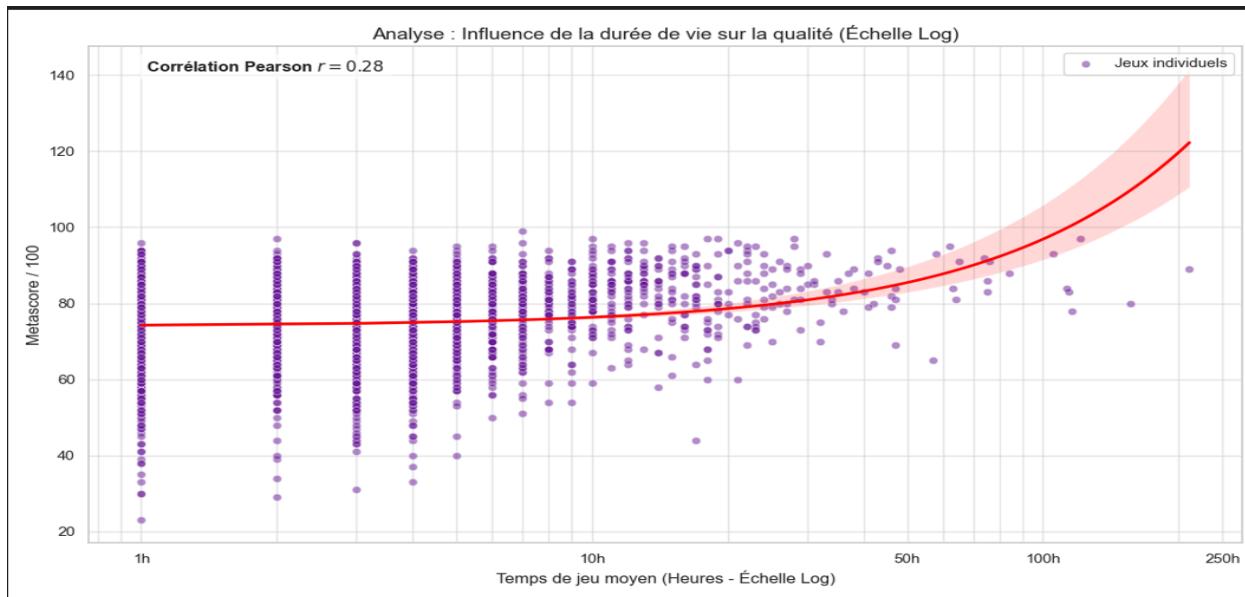
L'analyse temporelle du volume de sorties met en évidence une croissance soutenue de l'industrie. Après des débuts relativement modestes durant les années 1980 et 1990, le nombre de publications a augmenté progressivement avant de connaître une accélération spectaculaire sur la période 2013-2019, durant laquelle plus de 1 000 jeux ont été publiés en seulement cinq ans. Si une baisse apparente est observée sur la période récente de 2020 à 2024, celle-ci doit être interprétée avec prudence car elle résulte vraisemblablement de l'incomplétude actuelle de la base de données plutôt que d'un réel ralentissement économique du secteur.



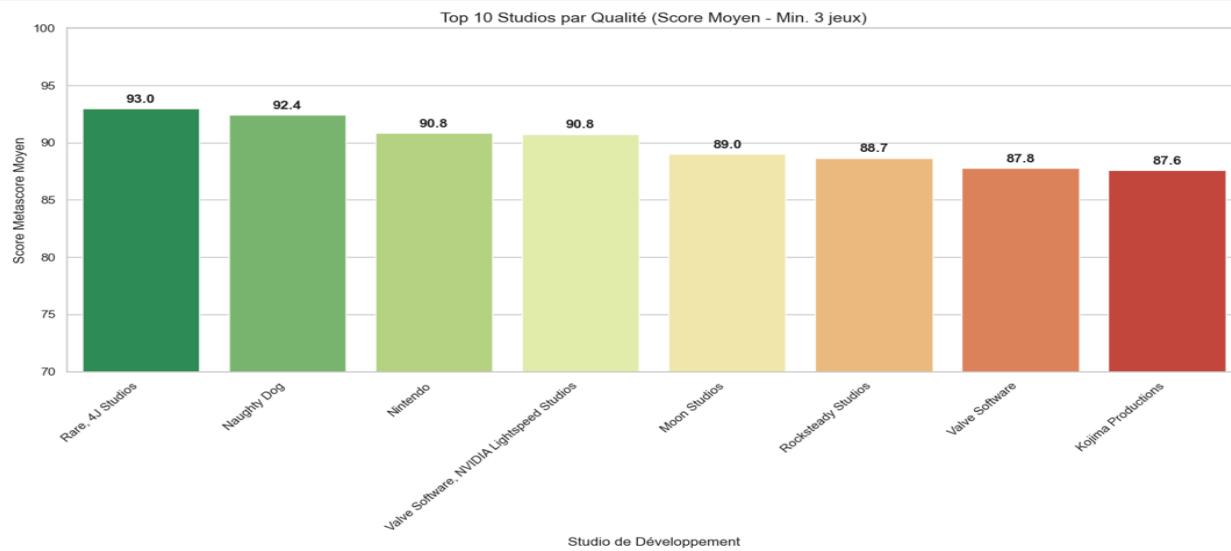
Parallèlement à l'évolution globale du volume, l'étude de la production des dix développeurs les plus prolifiques révèle des cycles d'activité bien définis. Jusqu'au début des années 2000, la production de ces studios restait limitée et dispersée. Une montée en puissance s'observe dès 2005, culminant avec un pic d'activité sur la période 2010-2013, marquée par une forte contribution d'acteurs majeurs comme Ubisoft, Telltale Games, Nintendo et Square Enix. La diminution du nombre de jeux développés par ces acteurs après 2015 s'explique par des évolutions individuelles, telles que la fermeture définitive de Telltale Games en 2018 ou la variabilité des cycles de production, plutôt que par une tendance sectorielle uniforme.



En ce qui concerne la répartition temporelle des sorties, l'analyse par heatmap met en lumière une saisonnalité commerciale très marquée. Les mois de septembre et octobre concentrent systématiquement les densités de lancements les plus élevées, reflétant une stratégie classique de positionnement à la rentrée pour anticiper la période des fêtes de fin d'année. À l'inverse, les données pour les années 2022 et 2023 présentent des valeurs majoritairement faibles ou nulles, ce qui confirme le délai nécessaire pour l'évaluation définitive des jeux et l'enregistrement complet des informations dans la base de données après leur sortie.

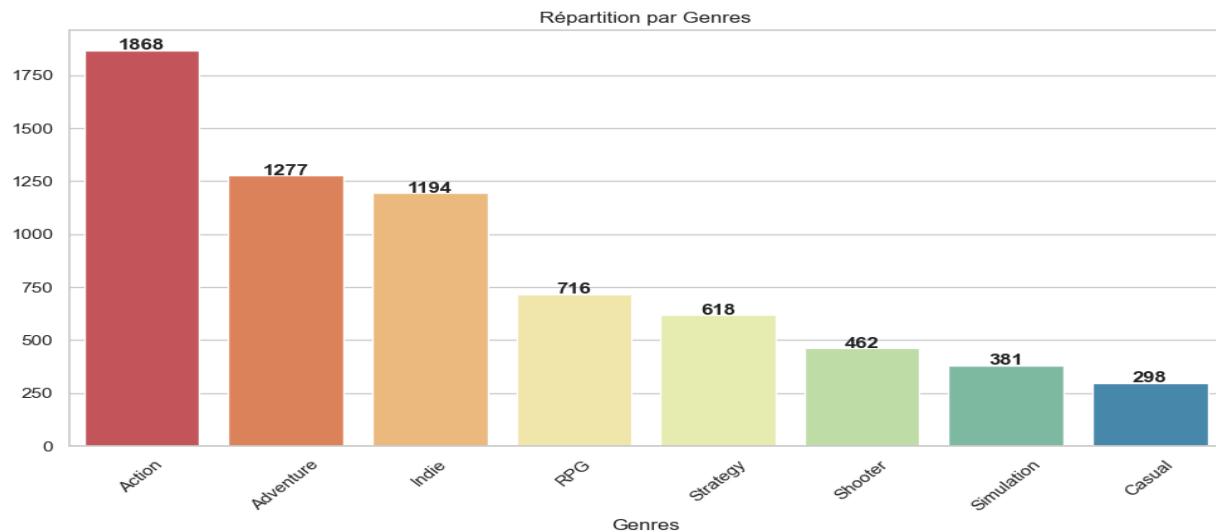


L'étude s'est également penchée sur la relation entre la durée de vie d'un jeu et sa réception critique pour déterminer si un temps de jeu plus long est synonyme de meilleure qualité. La courbe de tendance dégage une légère corrélation positive, indiquant que les jeux nécessitant un investissement en temps élevé, généralement au-delà de 20 à 30 heures, tendent à obtenir des scores Metacritic supérieurs. Cependant, avec un coefficient de corrélation de Pearson de 0,28, cette relation reste modérée et ne constitue pas une règle absolue. Une forte hétérogénéité est observée pour les jeux courts, d'une durée de une à dix heures, où certains titres sont excellents tandis que d'autres sont fortement critiqués.



Si l'on délaisse le volume pour se concentrer sur l'excellence critique, le classement des studios par qualité moyenne offre une perspective différente. Pour figurer dans ce palmarès, les studios devaient avoir publié au moins trois jeux, garantissant ainsi une moyenne représentative. Les résultats témoignent d'une forte exigence qualitative, l'ensemble des studios présents affichant une note moyenne supérieure à 87 sur 100. Le studio Rare / 4J Studios se distingue en tête avec une note moyenne de 93, suivi de très

près par Naughty Dog à 92,4 et Nintendo à 90,8, prouvant que ces acteurs privilégient la constance de l'excellence à la quantité



Concernant les genres, le marché est dominé par un trio de tête clair : l'Action prédomine largement avec 1 868 titres, suivi par l'Aventure avec 1 277 titres et les jeux Indés avec 1 194 titres

### **Retour d'Expérience et Responsabilité :**

#### **Analyse et Extraction de Données :**

Tasnim Alshawwa (10%) : J'ai été responsable de la dimension éthique de l'extraction en analysant le fichier **robots.txt** du site source.

Violette Grosjean (90%) : J'ai développé la partie extraction pour collecter automatiquement les informations des jeux vidéo depuis rawg.io, avec gestion des erreurs et filtrage des jeux incomplets. Des contrôles stricts et des sauvegardes régulières garantissent la robustesse du processus et évitent la perte de données comme le script dure un certain temps.

#### **Traitement et Export :**

Violette Grosjean (70%) : J'ai transformé les données brutes en tableau structuré, en traitant les dates et en réorganisant les colonnes pour produire un fichier Excel exploitable.

Tasnim Alshawwa (30%) : J'ai pris en charge la transformation des données brutes en indicateurs exploitables. Mon travail consistait à nettoyer la base comme gérer les NAs, à extraire les durées de jeu numériques via des expressions régulières et à convertir les dates en entiers. J'ai également segmenté les données par tranches de 5 ans et créé des tables croisées pour permettre la génération des graphiques appropriés comme les graphiques empilés et de la heatmap de saisonnalité.

#### **Visualisation et Interprétation :**

Tasnim Alshawwa (90%) : J'ai été responsable de la création des représentations graphiques pour rendre les données compréhensibles et visuelles. J'ai sélectionné des graphiques variés et adaptés à chaque problématique, tels que des nuages de points pour les corrélations, des barres empilées pour comparer la production des studios, et une heatmap pour analyser la saisonnalité des sorties. Concernant l'optimisation de la lisibilité, j'ai appliqué une échelle logarithmique sur l'axe des temps de jeu. Cela permet d'observer les jeux courts et les jeux massifs sur le même plan sans écraser les points. Ensuite, j'ai personnalisé les palettes de couleurs (ex: *Spectral*, *Magma*) et ajouté des étiquettes de données automatiques au-dessus des barres pour une lecture immédiate des chiffres clés, sans avoir à consulter le tableau brut.

Finalement, J'ai donné du sens aux visualisations produites, et analysé les tendances cachées derrière les graphiques pour en tirer des conclusions pertinentes.

Violette Grosjean (10%) : Nous nous sommes accordés pour savoir quelle variable pouvait être intéressante à visualiser et j'ai complété certains commentaires de graphiques.