

VCODFI 2A

SAE-SID

Analyse des tendances d'achats des jeux videos en ligne



Diego CASAS BARCENAS Matteo CAI et Leo UNITE

20/01/2024

I. Présentation du projet	2
Contexte	2
Problématique.....	2
Objectifs	2
II. Modélisation du SID	2
Choix du processus métier à modéliser	2
Définition de la granularité	2
Choix des dimensions.....	3
Identification des faits (Mesures).....	3
III. Intégration de données (Processus ETL)	4
Description du processus d'extraction	4
Transformation des données.....	4
Gestion des erreurs et des données manquantes.....	5
Création de colonnes personnalisées et conditionnelles	5
Analyse de la qualité des données	5
Création et utilisation de tables de temps	5
Création des mesures(fait) avec DAX.....	6
Description des modèles de données utilisés.	6
IV. Restitution	8
Reporting:.....	8
Video Games.....	8
Tableaux de bords	9
1. Video Games.....	9
2. Steam.....	10
3. Epic Games	12

I. Présentation du projet

Contexte

Les données liées aux jeux vidéo présentent une opportunité unique pour comprendre l'industrie à travers plusieurs perspectives : les ventes globales, les avis des joueurs, et la popularité dans le domaine de l'eSport. Cette analyse vise à créer une vision globale pour les décideurs dans ce secteur dynamique.

Problématique

Quels sont les plus grands consommateurs de produits en ligne ?

Objectifs

- Fournir une vision complète de l'industrie des jeux vidéo à travers des analyses basées sur des données fiables.
- Identifier les tendances clés et les facteurs de succès des jeux vidéo populaires.
- Présenter les résultats à l'aide de tableaux de bord interactifs sur Power BI.
- Assurer une gestion rigoureuse du projet en stockant l'ensemble des fichiers et livrables dans un [répertoire GitHub](#) structuré et daté, accompagné d'un [fichier README](#) détaillant les étapes du projet.

II. Modélisation du SID

Choix du processus métier à modéliser

L'analyse des performances des jeux vidéo se concentre sur l'évaluation des ventes, en prenant en compte plusieurs critères tels que les régions, les plateformes, les genres et les années de sortie. L'objectif est de comprendre comment ces facteurs influencent les résultats commerciaux des jeux.

Définition de la granularité

Chaque ligne du fait correspond à un jeu vidéo avec ses ventes globales et régionales (Amérique du Nord, Europe, Japon, autres régions).

Choix des dimensions

Processus	Genre	Date	Prix	Score	les éditeurs	Plateforme
Vente de jeu	☐—	☐—	☐	☐	☐—	☐—
Notation du jeu	☐—	☐—	☐	☐	☐—	☐—

Nous avons choisi d'utiliser plusieurs dimensions pour structurer les données de manière efficace.

- La dimension **dim_publishers** est utilisée pour représenter les éditeurs de jeux, fournissant ainsi des informations sur les entités responsables de la publication des titres.
- La dimension **dim_type_jeux** permet de décrire les genres de jeux, tels que les jeux d'action, de rôle, de simulation, etc., afin de mieux catégoriser les différents types de jeux vidéo.
- La dimension **dim_type_platform** est dédiée aux plateformes de jeu, telles que les consoles, PC ou mobiles, permettant ainsi de spécifier les environnements dans lesquels les jeux peuvent être joués.
- La dimension **dim_annee** capture les années de sortie des jeux, offrant ainsi une perspective chronologique.

Enfin, la **TableDate** est conçue pour ajouter des informations temporelles détaillées, telles que l'année, le trimestre et le mois, afin de faciliter l'analyse des données selon des périodes spécifiques.

Identification des faits (Mesures)

Les faits principaux que nous avons identifiés pour évaluer les ventes des jeux vidéo sont les suivants :

Global_Sales, qui représente les ventes totales des jeux au niveau mondial, nous offrant une vision d'ensemble des performances des titres à l'échelle internationale. Ensuite, nous avons les **NA_Sales**, qui correspondent aux ventes réalisées spécifiquement en Amérique du Nord, un marché clé pour de nombreux jeux. De même, les **EU_Sales** nous permettent de suivre les ventes en Europe, tandis que les **JP_Sales** nous donnent un aperçu des ventes au Japon, une région importante, notamment pour les jeux développés localement. Enfin, les **Other_Sales** regroupent toutes les ventes réalisées dans d'autres régions du monde, permettant de compléter l'analyse en tenant compte des marchés émergents ou moins présents sur la scène internationale.

III. Intégration de données (Processus ETL)

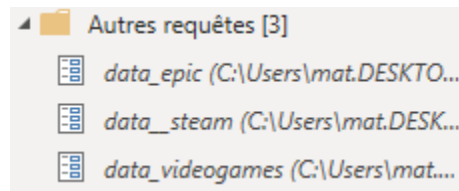
Description du processus d'extraction

Les données utilisées pour la réalisation du projet proviennent de plusieurs fichiers CSV. Ces fichiers contiennent des informations essentielles sur les jeux vidéo et leurs performances commerciales. Les fichiers sont les suivants :

- Steam 2024 best revenu
- Vidéo Games 1500
- Epic Games Store-Video Games 2024

Premièrement, nous avons ouvert l'application et cliqué sur l'onglet "**Accueil**", puis sélectionné "**Obtenir les données**". Ensuite, nous avons choisi l'option "**Texte/CSV**" pour importer nos fichiers de données au format CSV. Après avoir sélectionné les fichiers, nous avons défini un **chemin d'accès** (path) spécifique où nous avons enregistré notre travail. Cela nous permet de collaborer facilement en équipe et de travailler depuis différents postes, tout en garantissant que tous les membres de l'équipe accèdent aux mêmes fichiers de données, quelle que soit leur localisation ou le poste utilisé.

Ex :



Transformation des données

Pour préparer les données à l'analyse, plusieurs étapes de nettoyage et de transformation ont été réalisées pour garantir leur cohérence et leur pertinence :

- Suppression des colonnes non pertinentes.
- Modification des types de données pour une cohérence optimale.

Les types de données ont été modifiés pour garantir une meilleure cohérence et faciliter leur utilisation dans le cadre de l'analyse.

Tout d'abord, les points (".") utilisés comme séparateurs décimaux dans certains fichiers, pratique courante aux États-Unis, ont été remplacés par des virgules (",") afin de se conformer aux conventions européennes. Cette modification permet de garantir que les valeurs numériques soient correctement interprétées et traitées. De plus, certaines données initialement enregistrées sous forme numérique, comme les dates ou les années, ont été converties en format **Date**.

Certaines données, initialement sous forme numérique, représentaient des dates ou des années. Ces informations ont été converties en format **Date** pour faciliter leur manipulation et permettre des analyses temporelles, comme les tendances sur l'année, les trimestres ou les mois. En particulier, les dates au format américain (MM/DD/YYYY) ont été converties au format européen (DD/MM/YYYY), afin d'uniformiser les données.

Tout d'abord, nous avons dupliqué les tables et supprimé les colonnes inutiles. Ensuite, nous avons éliminé les doublons et ajouté un index pour définir les clés primaires. Enfin, pour établir les clés primaires dans les tables, nous avons utilisé la fusion de colonnes des bases initiales afin de relier les données entre elles.

Justification : Amélioration de la lisibilité et adaptation aux objectifs d'analyse.

Gestion des erreurs et des données manquantes

- Suppression des doublons.

Création de colonnes personnalisées et conditionnelles

- Création de la colonne Prix en Euro.

Analyse de la qualité des données

Analyser et rapporter sur la qualité des données après transformations, en utilisant des outils comme le profil de colonne pour montrer la distribution des données et identifier des anomalies potentielles.

Création et utilisation de tables de temps

Lors des manipulations dans power query et dans PowerBI, les données fournies contenaient déjà des dates. Donc nous n'avons pas eu besoin de créer de nouvelle table de temps.

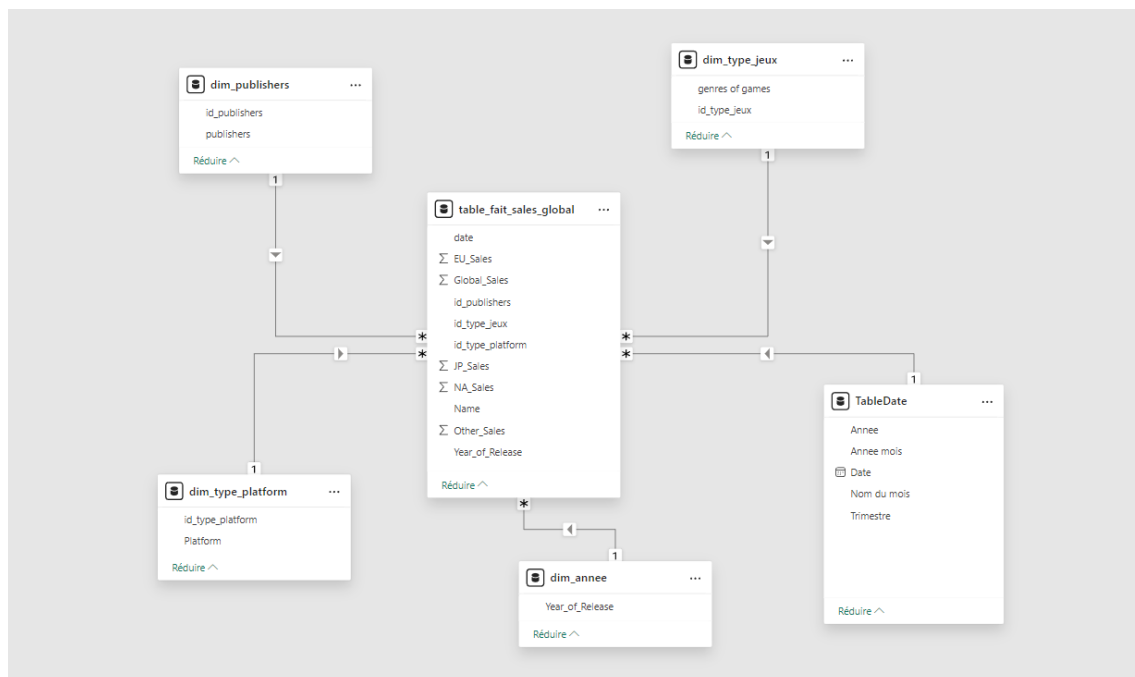
Création des mesures(fait) avec DAX

Nous avons utilisé une requête DAX pour créer une table calendrier dans notre modèle de données. Cette table comprend des colonnes essentielles pour l'analyse temporelle, telles que :

- Année
- Trimestre
- Mois

```
1 TableDate =  
2  
3 VAR DebutDate = DATE(1995,1,1)  
4 VAR FinDate = DATE(2020,12,31)  
5 RETURN  
6 ADDCOLUMNS(CALENDAR(DebutDate,FinDate),  
7 | | | "Annee", FORMAT([Date],"YYYY"),  
8 | | | "Trimestre" , "T"&FORMAT([Date],"Q"),  
9 | | | "Annee mois", FORMAT([Date],"YYYYMM"),  
10 | | | "Nom du mois", FORMAT([Date],"MMMM YY")  
11 )
```

Description des modèles de données utilisés.



Dans notre projet, nous avons choisi d'utiliser un modèle en étoile en raison de sa simplicité et de son efficacité pour les analyses multidimensionnelles.

Ce modèle comprend une table centrale de faits, représentant les ventes des jeux vidéo, et plusieurs dimensions qui décrivent les caractéristiques des données.

Initialement, nous avons envisagé de nous concentrer sur la table Steam, en analysant ses revenus. Cependant, nous avons rapidement constaté que les jeux communs entre les différentes bases de données étaient trop peu nombreux, rendant cette approche moins pertinente.

Nous avons donc décidé de nous concentrer sur la table Vidéo Games, qui nous a permis de mieux capturer les données relatives aux ventes dans différentes régions et de réaliser une analyse plus complète.

En complément, nous avons choisi des dimensions essentielles pour affiner notre analyse : la dimension des types de jeux (pour étudier les différents genres de jeux vidéo), la dimension des plateformes (pour comprendre sur quelles plateformes les jeux sont les plus populaires), la dimension des éditeurs (pour analyser la répartition des ventes par éditeur), et enfin, une table calendrier pour organiser les données temporelles, nous permettant de suivre les ventes sur différentes périodes (années, mois, trimestres).

IV. Restitution

Reporting:

Les visualisations interactives ont été conçues dans PowerBI pour répondre aux objectifs identifiés.

Exemples :

Video Games

- **Analyse des ventes par genre et région :**

- Des treemaps affichent la répartition des ventes (Global, EU, JP, NA) selon les genres de jeux vidéo.
- Ces visualisations permettent d'identifier les genres les plus populaires dans chaque région et d'établir des tendances de consommation spécifiques.

- **Top développeurs par nombre de ventes et répartition régionale :**

- Un graphique en barres met en avant les 10 meilleurs développeurs en termes de ventes globales, en détaillant les parts de marché par région (Japon, Amérique du Nord, Europe).
- Cet élément permet d'évaluer la performance des studios en fonction de leur implantation géographique.

- **Évaluation des jeux selon les scores critiques :**

- Un graphique en anneau présente les moyennes des scores attribués aux jeux par les critiques, facilitant la comparaison des titres les plus performants selon les éditeurs.
- Cela aide à mettre en évidence les franchises les plus appréciées par le public et les critiques.

- **Interaction et filtres :**

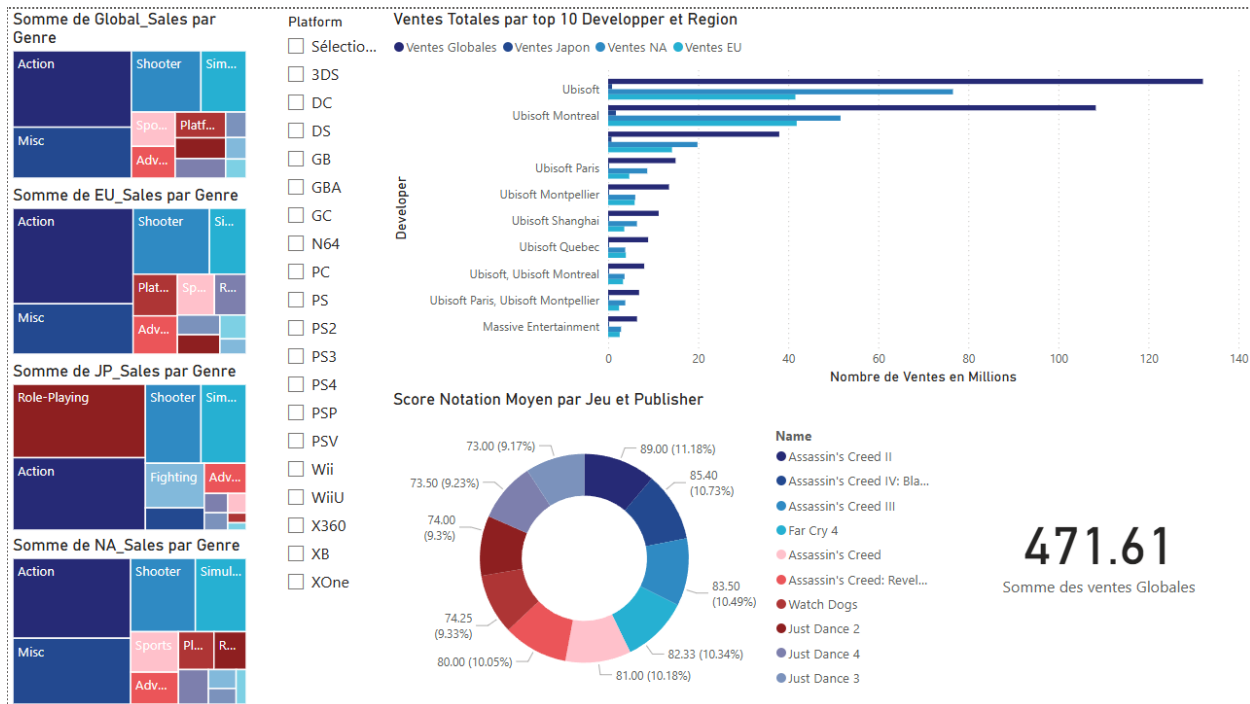
L'utilisation de filtres dynamiques (par plateforme, éditeur, et genre) permet d'explorer les données de manière personnalisée, en fonction des besoins spécifiques des utilisateurs.

Ces visualisations offrent ainsi une compréhension approfondie des tendances du marché vidéoludique et permettent une prise de décision éclairée sur les stratégies à adopter.

Tableaux de bords

Des tableaux de bord dynamiques mettent en évidence :

1. Video Games



La répartition des ventes par genre et région

- Une analyse visuelle des genres de jeux les plus vendus, segmentée par marché (Global, Europe, Japon, Amérique du Nord).
- Cela permet d'identifier les préférences régionales et d'orienter les stratégies marketing en fonction des attentes locales.

L'analyse par plateforme de jeu

- Grâce aux filtres dynamiques, l'utilisateur peut explorer les ventes par plateforme spécifique (PS4, Xbox, Wii, etc.), permettant d'évaluer l'impact de chaque support sur le marché global.

2. Steam

Moyenne totale du prix des jeux (17,52 €)

- Ce chiffre représente le prix moyen des jeux répertoriés dans l'ensemble des données.

Nuage de points (Copies vendues vs. Prix)

- Un graphique en nuage de points montre la relation entre le prix des jeux et le nombre de copies vendues. La majorité des jeux se situent dans la tranche de prix inférieure à 50 €, avec des volumes de ventes variés.

Revenu moyen par éditeur (barres verticales)

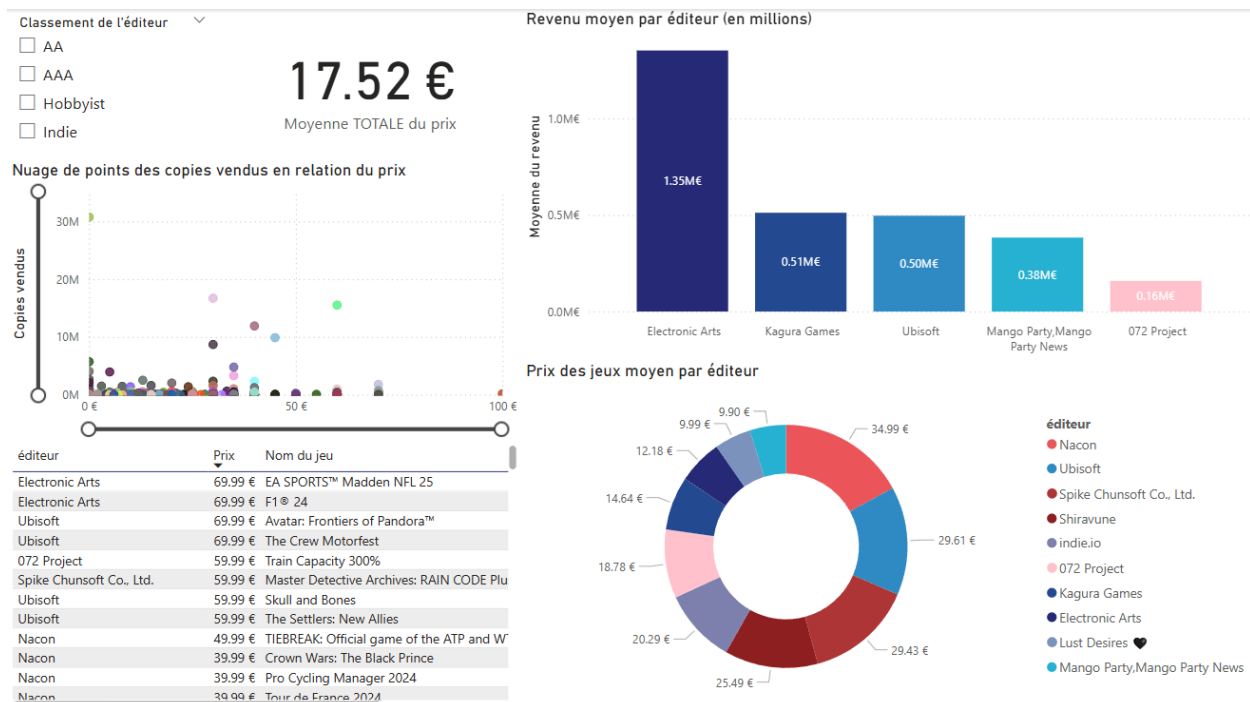
- Un graphique à barres indique les revenus moyens par éditeur en millions d'euros. Electronic Arts est en tête avec 1,35 M€, suivi de Kagura Games, Ubisoft et d'autres éditeurs ayant des revenus plus modestes.

Prix moyen par éditeur (diagramme circulaire)

- Un diagramme circulaire montre les prix moyens des jeux par éditeur. Les éditeurs comme Nacon et Ubisoft ont des prix moyens élevés (autour de 35 €), tandis que certains éditeurs affichent des prix moyens plus bas (environ 10 €).

Tableau des jeux individuels

- Une liste présente les jeux avec leur prix, leur éditeur, et leur nom. Les titres les plus coûteux sont associés à des éditeurs AAA comme Electronic Arts et Ubisoft.



La répartition du revenu en relation avec le prix du jeu et le classement de l'éditeur

- Cette analyse cherche à explorer la relation entre le nombre de ventes et les prix des jeux segmente par les principaux classements.
- Il nous permet de comprendre la relation entre le prix d'un jeu, son succès ou échec grâce au nombre de copies vendues et du revenu obtenu.
- Le but de cette analyse est de comprendre le comportement des utilisateurs et d'établir des possibles stratégies de ventes pour obtenir un plus grand bénéfice.

La répartition du revenu entre le top 10 d'éditeurs

- Cette analyse cherche à visualiser la distribution des revenus entre le top 10 des éditeurs qui ont le plus grand nombre de jeux dans leurs catalogues.
- Le but de cette analyse est de comprendre la relation entre le revenu moyen par éditeur et le prix moyen des jeux vendus sur leurs catalogues.

3. Epic Games

Nombre de jeux par caractéristiques (barres horizontales)

- Ce graphique montre le nombre de jeux répartis selon leurs caractéristiques.
- Les jeux Single Player sont les plus nombreux (233 titres).
- Les jeux avec support manette (Controller Support Single Player) et ceux combinant Single Player avec support manette sont également bien représentés (56 et 47 titres respectivement).
- Les autres caractéristiques comme Co-op ou Multiplayer ont un nombre de jeux plus réduit.

Top 10 des prix moyens par éditeur (diagramme circulaire)

- Les éditeurs sont classés selon le prix moyen de leurs jeux.
- Versus Evil a un prix moyen de 8,34 €, suivi de Devolver Digital (7,15 €) et Raw Fury (6,61 €).
- Les prix moyens les plus bas sont observés chez 1C Entertainment (3,67 €).

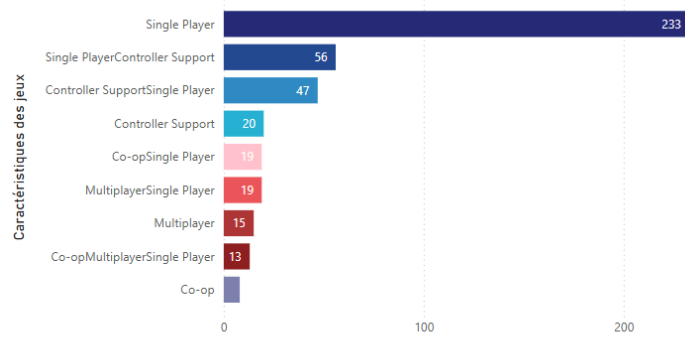
Tableau des jeux individuels avec leurs caractéristiques

- Une liste présente les jeux avec leur éditeur, leur prix, et leurs caractéristiques.
- Les jeux tels que Cardpocalypse ou Mothership affichent des prix situés entre 6,48 € et 10,65 €.
- Les titres Single Player avec support manette et les jeux Co-op sont aussi inclus dans cette répartition.

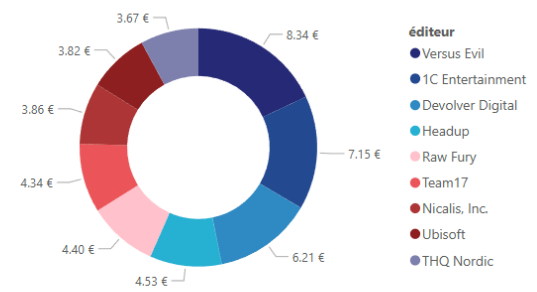
Prix moyen par caractéristiques des jeux (barres verticales)

- Les jeux Single Player ont un prix moyen de 4,98 €.
- Les jeux combinant Single Player et Controller Support sont légèrement moins chers (4,76 €).
- Les prix les plus bas concernent les jeux Co-op Multiplayer Single Player (3,00 €).

Nombre de jeux par caractéristiques



Top 10 des prix Moyens par éditeur

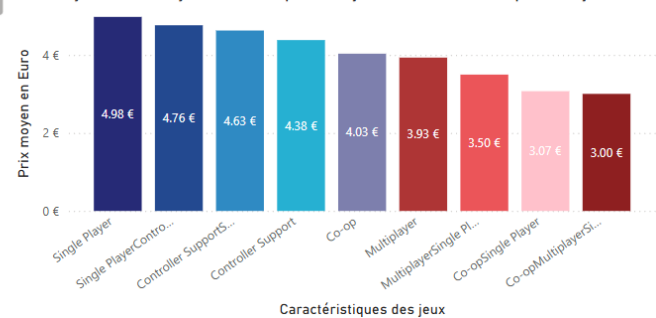


Caractéristiques des jeux

- ☐ Select all
- ☐ Controller Support
- ☐ Controller SupportSingle Player
- ☐ Co-op
- ☐ Co-opMultiplayerSingle Player
- ☐ Co-opSingle Player
- ☐ Multiplayer
- ☐ MultiplayerSingle Player
- ☐ Single Player
- ☐ Single PlayerController Support

éditeur	Prix en Euro	Nom du jeu
Versus Evil	10.65 €	Cardpocalypse
Versus Evil	6.48 €	Hitchhiker - A Mystery Game
Versus Evil	10.33 €	Mothergunship
Versus Evil	6.48 €	Sockventure
Versus Evil	10.33 €	The Hand of Merlin
Versus Evil	6.48 €	Tower of Guns
Versus Evil	7.80 €	UnMetal
Versus Evil	8.20 €	Yaga
Ubisoft	8.24 €	Anno 1404 History Edition
Ubisoft	5.49 €	Anno 1701 History Edition
Ubisoft	0.02 €	Anno 2070
Ubisoft	0.02 €	Anno 2205
Ubisoft	5.49 €	Assassin's Creed II
Ubisoft	0.03 €	Assassins Creed Odyssey

Prix moyen en Euro by Caractéristiques des jeux and Caractéristiques des jeux



La répartition de jeux en relation avec leurs caractéristiques

- Cette analyse cherche à explorer la relation entre le nombre de jeux et les caractéristiques.
- Il nous permet de comprendre la tendance des prix par rapport aux caractéristiques.
- Le but de cette analyse est de comprendre le rapport entre la variation des prix et la quantité des caractéristiques techniques si elles existent.

La répartition du prix entre le top 10 du nombre de jeux par éditeur

- Cette analyse cherche à visualiser la distribution des prix moyens entre le top 10 des caractéristiques les plus communes parmi la base de données.
- Le but de cette analyse est d'observer la distribution des prix moyens des jeux sur les éditeurs avec le plus grand nombre de jeux dans leurs catalogues pour mettre en évidence la relation prix - caractéristiques.