

VISUALITZACIÓ DE DADES

GRAU EN ENGINYERIA DE DADES

CURS 2024-2025

PORTFOLI

PER QUÈ ÉS BO UN JOC?

1. RESUM

Aquest projecte té com a objectiu desenvolupar un dashboard interactiu per analitzar el mercat de videojocs a Steam, proporcionant informació clau per a empreses desenvolupadores que necessiten suport en la presa de decisions basades en dades. S'utilitzen tres grans conjunts de dades que recullen informació sobre ingressos, vendes globals i activitat dels jugadors. L'informe inclou una anàlisi exploratòria per validar cinc hipòtesis relacionades amb preus, vendes, valoracions, preferències regionals i retenció de jugadors. El dashboard es dissenya en Power BI i consta de tres pàgines temàtiques que mostren la intel·ligència de mercat a Steam, tendències globals i l'evolució temporal de l'activitat dels jugadors. Es permeten filtres per tipus de publicador, regió, gènere i període, facilitant una anàlisi detallada i personalitzada. L'objectiu és aportar valor al sector dels videojocs, combinant dades històriques i actuals per millorar l'estratègia comercial i productiva.

2. INTRODUCCIÓ - MOTIVACIÓ DEL TREBALL

El sector dels videojocs s'ha consolidat com una de les indústries més rellevants dins l'entreteniment digital, amb un creixement constant any rere any. Steam destaca com la plataforma més gran per a la distribució de jocs de PC, connectant milers de jocs amb milions de jugadors arreu del món. No obstant això, moltes petites i mitjanes empreses desenvolupadores troben dificultats per competir en aquest mercat saturat, especialment per la manca d'accés a eines que permetin extreure coneixement útil a partir de les dades disponibles.

Tot i l'abundància d'informació sobre preus, ingressos, ressenyes i hàbits dels jugadors, aquesta no es presenta de forma integrada ni fàcilment interpretable. Això complica la presa de decisions en àmbits com el llançament, la segmentació de mercat, o l'estratègia comercial. Per aquest motiu, es detecta una necessitat creixent d'eines interactives, accessibles i basades en dades que permetin una millor comprensió del mercat.

Davant d'aquesta realitat, el projecte proposa el desenvolupament d'un dashboard en Power BI que permeti analitzar i visualitzar dades clau de Steam. L'objectiu és facilitar l'exploració de tendències, preferències i patrons de comportament dels usuaris, oferint suport estratègic a les empreses desenvolupadores i aportant valor tangible a la indústria dels videojocs.

3. HIPÒTESIS

Abans de construir el dashboard i analitzar les dades, s'han definit cinc hipòtesis inicials per guiar l'estudi. Aquestes es basen en tendències conegudes, experiència prèvia i preocupacions comunes del sector. L'objectiu és enfocar l'anàlisi en aspectes clau del mercat de videojocs a Steam:

1. **Els jocs desenvolupats per grans empreses (AAA) tenen un preu mitjà més elevat que els jocs indie o AA, però això no garanteix necessàriament un nombre de còpies venudes superior.**
Aquesta hipòtesi busca explorar la relació entre pressupost de desenvolupament i rendiment comercial. Ens preguntem si els jocs més cars tenen més èxit o si els jocs independents aconsegueixen destacar per si mateixos.
2. **Els jocs millor valorats pels usuaris obtenen més vendes que aquells amb valoracions baixes.**
Partim de la intuïció que les bones ressenyes poden afavorir la visibilitat i confiança del consumidor, però volem validar si realment hi ha una correlació directa entre puntuacions i rendiment comercial.
3. **Existeixen diferències notables en les preferències de gènere segons la regió geogràfica, amb una distribució desigual del consum entre Europa, Nord-Amèrica, Japó i altres regions.**
Aquesta hipòtesi planteja la possibilitat que factors culturals influeixin en el tipus de videojocs més populars segons el territori, fet que podria tenir implicacions en la localització i màrqueting.
4. **Els videojocs amb una estratègia de preu ajustat aconsegueixen una retenció de jugadors superior (mesurada per temps mitjà de joc).**
Ens interessa determinar si existeix una relació entre el cost d'adquisició d'un joc i el compromís a llarg termini dels usuaris, ja que això podria indicar la rendibilitat de models de preu més accessibles.
5. **L'evolució del nombre de jugadors actius al llarg dels anys es veu influenciada per factors com ara expansions, descomptes, o tendències socials específiques.**
Aquesta hipòtesi pretén entendre millor el cicle de vida d'un videojoc i quins esdeveniments poden generar

pics de popularitat o bé provocar la seva caiguda.

4.1 RECOLLIDA I SELECCIÓ DE DADES

Per a la realització d'aquest projecte s'han utilitzat tres conjunts de dades públics del sector dels videojocs per obtenir una visió global del mercat. Aquests inclouen informació sobre vendes, popularitat i comportament dels jugadors, permetent fer anàlisis comparatives, segmentacions i correlacions entre variables rellevants. Els datasets seleccionats són els següents:

1. Steam_2024_bestRevenue_1500.csv:

El conjunt de dades **Steam 2024 Top 1500 Games by Revenue** (<https://www.kaggle.com/code/devraai/steams-top-1500-games-by-revenue-analysis-2024/notebook>) recull informació detallada sobre els 1500 videojocs amb majors ingressos a la plataforma Steam durant l'any 2024. Aquest dataset ha estat elaborat a partir de dades oficials de Steam, compilades i estructurades per facilitar-ne l'anàlisi econòmica i comercial.

Aquest conjunt de dades permet identificar quins jocs han generat més ingressos, en quines plataformes i amb quines característiques, proporcionant una base sòlida per estudiar el mercat actual de videojocs digitals. Disposa de **1500 files** i **10 columnes d'atributs**.

- **name:** Nom del videojoc. Cada fila representa un títol diferent.
- **releaseDate:** Data de llançament del joc. Per a analitzar si la data de llançament del videojoc afecta les seves vendes.
- **copiesSold:** Nombre total d'unitats venudes. Reflecteix l'abast del joc en termes d'acceptació pel públic.
- **price:** Preu inicial de venda al públic (en dòlars). Pot influir directament en els ingressos totals.
- **revenue:** Ingressos generats per les vendes del joc. Aquesta és la mètrica clau del dataset, utilitzada per ordenar els 1500 jocs.
- **avgPlaytime:** Temps mitjà que els jugadors han dedicat al joc (en hores). Pot indicar el nivell d'immersió o qualitat percebuda.
- **reviewScore:** Puntuació mitjana basada en valoracions d'usuaris i crítics. Permet veure si la qualitat percebuda té impacte en les vendes.
- **publisherClass:** Classificació de l'editor: AAA, AA o Indie. Serveix per analitzar quin tipus d'empresa genera més ingressos.
- **publishers:** Nom de l'empresa editorial responsable del llançament. Pot ajudar a identificar marques amb èxit sostingut.
- **developers:** Nom del desenvolupador o equip que ha creat el joc. S'utilitza per analitzar correlacions entre desenvolupadors i ingressos.

2. Vgsales.csv:

El conjunt de dades **Video Game Sales** (<https://www.kaggle.com/datasets/dandanjia/vgsales-csv>) proporciona informació històrica sobre les vendes globals de videojocs des de l'any 1980 fins aproximadament el 2016. Aquest dataset, àmpliament utilitzat per a anàlisis de mercat i estudis de tendències, inclou informació rellevant sobre els jocs més venuts en diverses plataformes i regions del món.

El dataset compta amb **16.598 files** i **11 columnes d'atributs**, cadascuna aportant informació clau per analitzar l'impacte comercial dels videojocs en diferents mercats i períodes de temps.

- **Name:** Nom del videojoc. És l'identificador principal per a cada fila i permet distingir entre diferents títols.
- **Platform:** Plataforma en què es va publicar el videojoc (com per exemple: PS4, Xbox 360, PC, etc.). Aquest atribut és útil per analitzar les vendes segons el suport tecnològic.
- **Year:** Any de llançament del videojoc. Serveix per estudiar l'evolució del mercat i identificar períodes de més producció o èxit comercial.
- **Genre:** Gènere del videojoc (Acció, Esport, Estratègia, etc.). Aquest atribut permet fer anàlisis comparatives entre gèneres pel que fa a vendes i popularitat.
- **Publisher:** Nom de l'empresa que ha publicat el videojoc. Pot ajudar a identificar quines empreses dominen el mercat o tenen més èxits comercials.
- **NA_Sales:** Vendes del videojoc en milions d'unitats a Amèrica del Nord (North America).

- **EU_Sales:** Vendes en milions d'unitats a Europa.
- **JP_Sales:** Vendes en milions d'unitats al Japó.
- **Other_Sales:** Vendes en milions d'unitats a la resta del món (altres regions no especificades).
- **Global_Sales:** Vendes totals globals del videojoc, resultat de la suma de totes les regions. És una mètrica clau per identificar els títols més exitosos a nivell mundial.
- **Rank:** Posició del videojoc en el rànquing de vendes globals. Serveix per ordenar els jocs segons el seu èxit comercial.

3. Valve_Player_Data.csv:

El conjunt de dades **Steam Player Data** (<https://www.kaggle.com/code/ryanlmcmillan/steam-player-data/input>) proporciona informació detallada sobre l'activitat dels jugadors als 100 jocs més populars de la plataforma Steam durant el període comprès entre 2012 i 2021. Aquest dataset ha estat recopilat a partir de dades de SteamCharts i posteriorment processat per facilitar-ne l'anàlisi.

El dataset conté dades mensuals sobre el nombre mitjà de jugadors per als jocs més destacats de Steam. Aquesta informació permet analitzar l'evolució de la popularitat dels jocs al llarg del temps i identificar tendències en el comportament dels jugadors. El dataset compta amb 5271 files i 8 columnes d'atributs.

Les columnes inclouen:

- **Month_Year:** Indica el mes i l'any al qual corresponen les dades registrades. És clau per analitzar l'evolució temporal de l'activitat dels jugadors.
- **Avg_players:** Reflecteix el nombre mitjà de jugadors que han jugat al videojoc durant el mes especificat. Aquesta mètrica permet observar el nivell d'interès sostingut per part de la comunitat.
- **Gain:** Representa el canvi absolut (positiu o negatiu) en el nombre mitjà de jugadors respecte al mes anterior. Útil per detectar increments o pèrdues en la base d'usuaris d'un joc.
- **Percent_Gain:** Mostra el percentatge de variació en el nombre mitjà de jugadors respecte al mes anterior. Proporciona una visió relativa del creixement o disminució en popularitat.
- **Peak_Player:** Indica el nombre màxim de jugadors simultanis que ha tingut el joc durant el mes. Aquesta mètrica és rellevant per identificar moments puntuals d'activitat màxima (per exemple, després d'una actualització o llançament de contingut).
- **URL:** Conté l'enllaç d'origen de les dades, generalment provinent de SteamCharts, que serveix com a referència i font de verificació.
- **Date:** Inclou la data completa (any, mes i dia) en què es van recollir o publicar les dades. Pot ser útil per processos de validació o ordenació temporal precisa.
- **Game_Name:** És el nom del videojoc al qual corresponen les dades. Permet identificar a quin títol es refereixen les mètriques recollides en cada fila.

4.2 PROCESSAMENT I NETEJA DE DADES

Un cop seleccionades les dades, s'ha realitzat un procés de neteja i transformació per assegurar-ne la qualitat i coherència. Aquestes tasques han estat essencials per garantir la fiabilitat de les anàlisis i conclusions posteriors. A continuació, s'expliquen totes les tasques específiques que s'han dut a terme en aquesta fase:

- **Detecció i tractament de valors nuls o duplicats.** S'han inspeccionat tots els conjunts de dades per identificar valors nuls o buits en camps essencials dels datasets. En aquests casos, s'han eliminat les files afectades. També s'han detectat i eliminat registres duplicats, especialment en datasets temporals on la combinació de data i joc es repetia per error.
- **Estandardització de formats.** S'han uniformitzat tots els formats de data al patró DD/MM/YYYY per tal de facilitar l'ordenació cronològica i les agregacions mensuals o anuals. També s'han ajustat valors percentuals, assegurant que el camp **Percent_Gain** estigués expressat de manera coherent com a valor decimal.
- **Eliminació d'una fila considerada irrellevant:** En el procés de filtratge, també s'han eliminat una fila que no aportava valor analític per a l'objectiu del projecte. Aquest cas concret és el cas del joc *Pixel Noir*, l'únic representant d'una categoria de publisher (hobbyist) que apareixia una sola vegada en tot el conjunt de dades (1 entre 1500 registres). Aquesta dada no permet establir cap patró o estratègia representativa, i per tant, s'ha considerat irrellevant per a l'anàlisi de tendències generals.

4.3 ANÀLISI EXPLORATÒRIA I VALIDACIÓ D'HIPÒTESIS

Un cop completat el procés de neteja i preparació de les dades, s'ha dut a terme una anàlisi exploratòria amb l'objectiu de comprendre millor el comportament de les variables clau i contrastar les hipòtesis inicialment formulades. Aquesta fase ha estat essencial per identificar patrons, correlacions i tendències que poguessin guiar les visualitzacions del

dashboard.

Per cadascuna de les cinc hipòtesis plantejades, es va definir una estratègia analítica específica, seleccionant les mètriques més representatives i els tipus de visualitzacions més adequats per facilitar-ne la interpretació:

- **Hipòtesi 1:** Relació entre el preu mitjà i el volum de vendes segons la tipologia de publicador (AAA, AA, Indie, Hobbyist)
 - o Mètriques: Preu mitjà del videojoc, nombre total de còpies venudes, classificació del publicador.
 - o Visualitzacions considerades: Gràfics de barres agrupades, diagrames de dispersió amb codificació per color segons la classe de publicador.
- **Hipòtesi 2:** Relació entre la valoració mitjana dels usuaris i el volum de vendes
 - o Mètriques: Puntuació mitjana d'usuaris (review score), nombre de còpies venudes.
 - o Visualitzacions considerades: Gràfics de dispersió, gràfics de bombolles (codificats per grandària de vendes).
- **Hipòtesi 3:** Preferències de gènere per part dels usuaris segons la regió geogràfica
 - o Mètriques: Gènere del videojoc, regió (Europa, Japó, Nord-Amèrica, Altres), còpies venudes.
 - o Visualitzacions considerades: Gràfics de barres apilades, gràfics de sectors i heatmaps per creuar gènere i regió.
- **Hipòtesi 4:** Relació entre el preu i el temps mitjà de joc (retenció)
 - o Mètriques: Preu del videojoc, temps mitjà de joc, classificació del publicador.
 - o Visualitzacions considerades: Diagrames de dispersió, codificats per categoria de publicador per detectar estratègies comercials diferenciades.
- **Hipòtesi 5:** Evolució temporal del nombre de jugadors i l'impacte d'esdeveniments específics
 - o Mètriques: Nombre mitjà i màxim de jugadors actius, guany/pèrdua de jugadors per període, dates de llançament o actualitzacions importants.
 - o Visualitzacions considerades: Gràfics de línies temporals, gràfics d'àrees i barres amb codificació temporal.

5. DISSENY I DESENVOLUPAMENT DEL DASHBOARD

Per al desenvolupament del dashboard s'ha utilitzat Power BI, una eina de visualització de dades potent, que permet construir informes dinàmics i interactius de manera eficient. Aquesta plataforma ha estat seleccionada per la seva capacitat d'integrar diferents fonts de dades, aplicar filtres avançats i oferir una experiència d'usuari intuïtiva.

El disseny del dashboard s'ha guiat per tres objectius fonamentals:

- **Interactivitat:** Permetre a l'usuari explorar les dades segons les seves necessitats específiques, a través de filtres i seleccions visuals que modifiquen el contingut en temps real.
- **Claredat:** Garantir una presentació neta i estructurada de la informació, prioritzant la comprensió immediata de les dades i la detecció de patrons rellevants.
- **Filtratge personalitzat:** Oferir una àmplia gamma de filtres globals i contextuals per tal que cada usuari pugui analitzar les dades segons criteris com l'any, la regió, el gènere del joc o un videojoc concret.

Per tal de facilitar la navegació i estructurar el contingut de manera lògica, el dashboard s'ha organitzat en tres pàgines temàtiques, cadascuna centrada en un àmbit específic d'anàlisi:

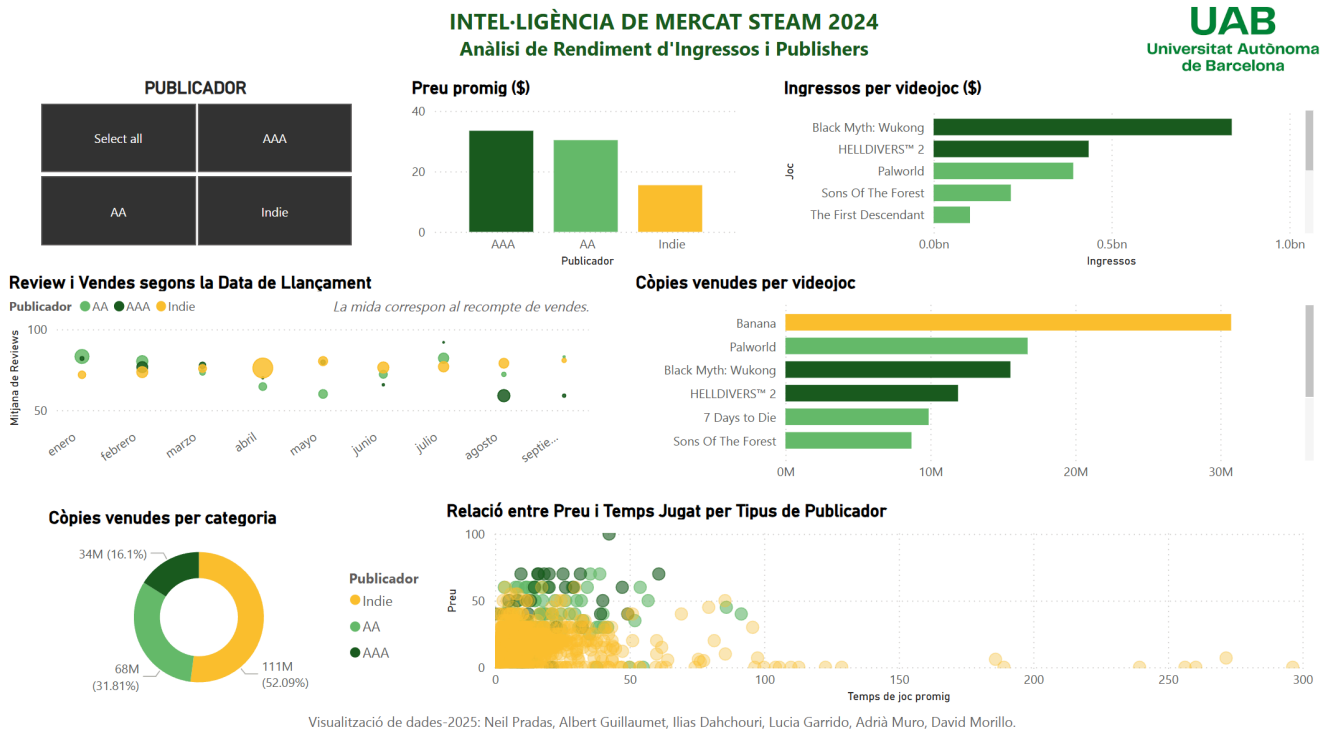
- **Pàgina 1:** Intel·ligència de Mercat Steam 2024: Anàlisi de Rendiment d'Ingressos i Publishers

Aquesta primera pàgina del dashboard ofereix una visió general del mercat actual de videojocs a la plataforma Steam, centrant-se en les dades més recents disponibles per a l'any 2024. L'objectiu és analitzar el rendiment econòmic i l'impacte dels diferents tipus de publicadors.

Visualitzacions:

- **Filtre de tipus de publicador:** Permet filtrar la informació segons la classificació dels publishers (AAA, AA, Indie) per facilitar comparacions específiques. Aquesta eina de filtratge ajuda a personalitzar l'anàlisi segons el segment d'interès i a contrastar comportaments de mercat entre categories, clau per validar la Hipòtesi 1 i 4. Cada tipus de publicador es representa amb un color diferent, fet que permet una classificació visual clara dels resultats en tots els gràfics de la pàgina.

- **Preu mitjà per tipus de publicador:** Un gràfic de barres vertical mostra com varia el preu mitjà dels videojocs segons la categoria del publicador, destacant diferències en l'estratègia de posicionament de preus. Aquesta visualització permet detectar si les grans productores fixen preus més alts que els estudis independents, reforçant l'anàlisi de la relació entre pressupost i estratègia comercial (Hipòtesi 1).
- **Ingressos per videojoc:** Amb un gràfic de barres horitzontals es visualitzen els jocs amb majors ingressos, mostrant l'impacte econòmic de cada títol en el mercat de Steam. Aquest gràfic ajuda a identificar quins jocs dominen el mercat en termes econòmics i a avaluar si els jocs amb preus més elevats aconsegueixen també més ingressos globals, complementant Hipòtesis 1 i 2.
- **Còpies venudes per videojoc:** Un diagrama de barres horitzontals veure quins videojocs han tingut més èxit en termes de vendes absolutes. Permet detectar jocs destacats en popularitat i contraposar-los amb el seu preu o valoració, clau per analitzar la correlació entre qualitat, cost i èxit comercial (Hipòtesis 1 i 2).
- **Evolució mensual del volum de vendes i la valoració mitjana:** Un gràfic de bombolles representa les còpies venudes i la puntuació mitjana dels usuaris per mes, aportant informació sobre la relació entre qualitat percebuda i rendiment comercial al llarg del temps. Aquest gràfic permet observar patrons temporals entre bones valoracions i vendes, facilitant l'avaluació de si una alta puntuació realment es tradueix en èxit comercial sostingut (Hipòtesi 2).
- **Distribució de vendes per categoria de publicador:** Un diagrama de sectors mostra la quota de mercat de cada segment editorial en termes de còpies venudes totals. Ajuda a visualitzar quin tipus de publicadors dominen en volum de vendes i si els editors més petits aconsegueixen competir amb els AAA, essencial per contrastar Hipòtesi 1.
- **Relació entre preu i temps mitjà de joc per tipus de publicador:** Un gràfic de dispersió correlaciona el preu dels videojocs amb la durada mitjana de joc, diferenciant per tipus de publicador. Aquesta anàlisi permet observar estratègies diferenciades entre empreses i valorar el rendiment del producte en relació amb l'experiència de l'usuari. És clau per validar la Hipòtesi 4, ja que mostra si un preu més ajustat pot correlacionar-se amb una major retenció de jugadors o si hi ha una relació inversa.



• Pàgina 2: Tendències Globals de Videojocs i Insights Regionals

Aquesta pàgina del dashboard ofereix una anàlisi històrica del mercat de videojocs entre els anys 1980 i 2016, amb especial èmfasi en les diferències regionals i l'evolució de la indústria al llarg del temps.

Visualitzacions:

- **Top 10 de videojocs més venuts:** Es mostra el rànquing dels jocs amb més vendes, desglossades per regió (Europa, Japó, Nord-Amèrica i altres), per identificar patrons de consum i èxits globals. Aquest gràfic permet

detectar si l'èxit d'un joc és global o concentrat en una regió específica, ajudant a analitzar l'impacte cultural en el consum de videojocs (Hipòtesi 3) i els títols que han marcat tendència (Hipòtesi 5). A tota la pàgina s'estableix un codi de colors basat en la regió del videojoc. Això permet una ràpida identificació visual de les diferències regionals i facilita la comparació entre mercats.

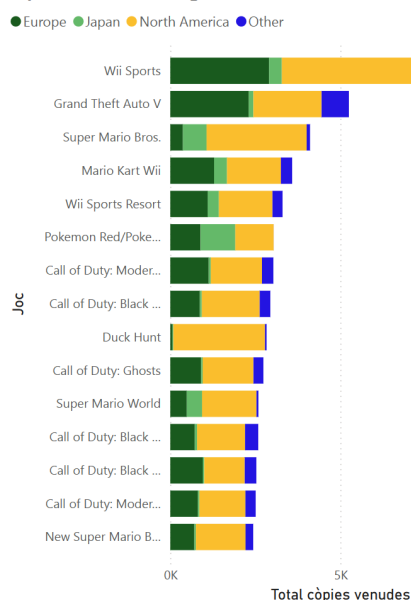
- **Evolució temporal de les vendes per regió:** Un gràfic de línies mostra la tendència de les vendes anuals en cada regió, posant de manifest moments d'expansió o retrocés segons el context geogràfic. És essencial per validar la Hipòtesi 5, ja que permet veure l'impacte d'esdeveniments específics (nous llançaments, modes o crisis) en el comportament dels jugadors al llarg del temps i per regió.
- **Distribució de vendes per regió:** Un gràfic de sectors resumeix la quota de mercat regional, amb Nord-Amèrica com a principal contribuent, seguida d'Europa i Japó. Aporta una visió ràpida de quines regions tenen més pes en el mercat global, permetent focalitzar estratègies de màrqueting i localització (Hipòtesi 3).
- **Vendes per gènere i regió:** Un gràfic de barres que combina informació sobre preferències de gènere de videojocs amb la seva distribució regional, mostrant, per exemple, l'èxit del gènere d'acció a totes les regions o la popularitat específica d'altres gèneres en mercats concrets. És clau per validar la Hipòtesi 3, ja que mostra clarament les diferències culturals en gustos i pot ajudar a identificar oportunitats de localització de continguts segons regió.

TENDÈNCIES GLOBALES DE VIDEOJOC I INSIGHTS REGIONALS

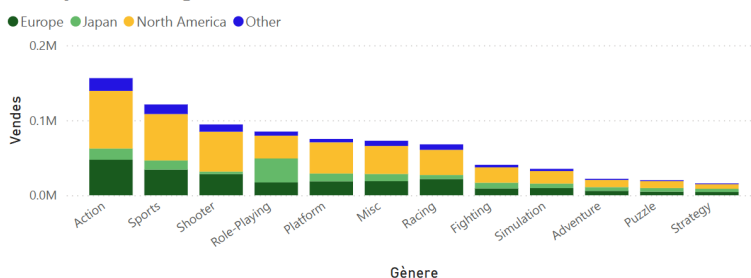
Anàlisi Històrica del Mercat (1980-2016)

UAB
Universitat Autònoma
de Barcelona

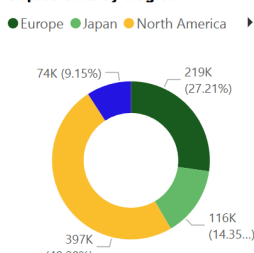
Top 15 most sold videogames



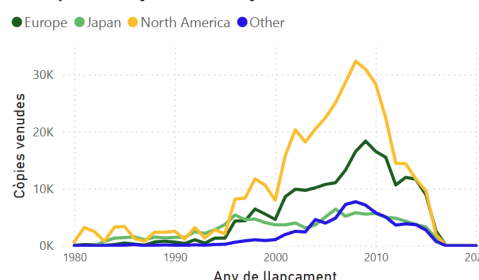
Sales by Genre and Region



Copies sold by Region



Sales per country and release year



Visualització de dades-2025: Neil Pradas, Albert Guillaumet, Ilias Dahchouri, Lucia Garrido, Adrià Muro, David Morillo.

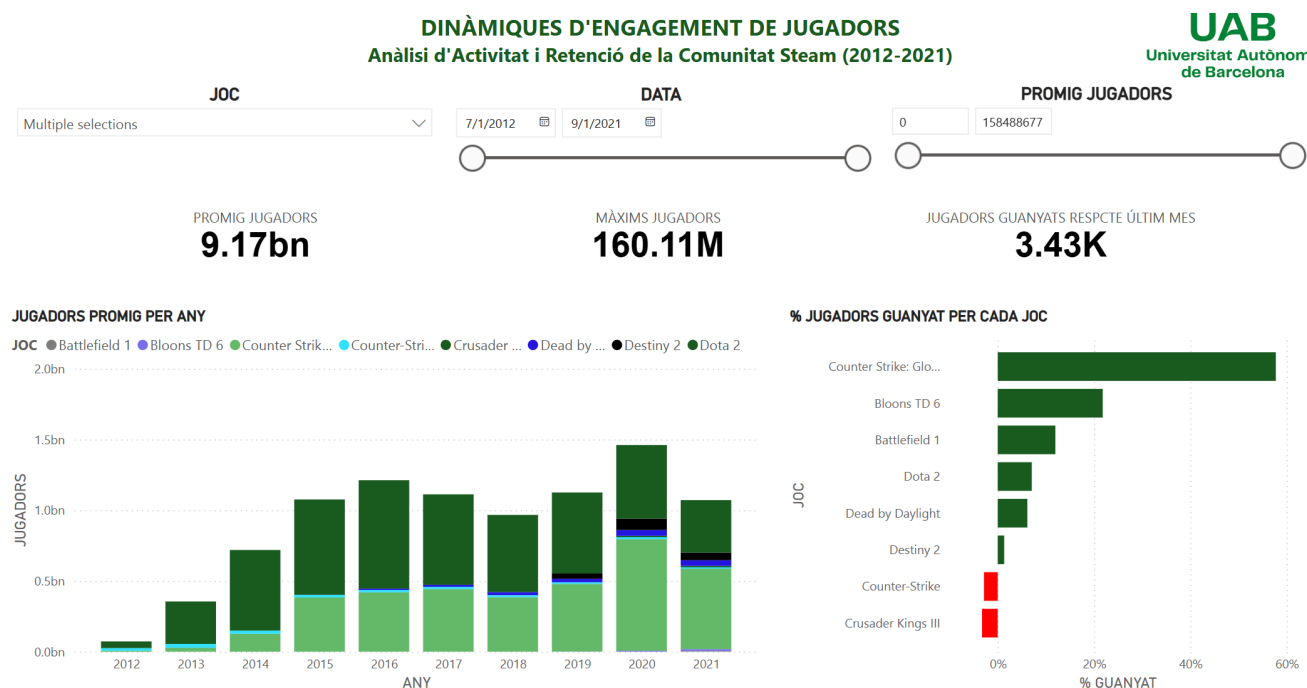
- **Pàgina 3: Dinàmiques d'Engagement de Jugadors: Activitat i Retenció a Steam (2012–2021)**

Aquesta pàgina se centra en l'anàlisi de l'activitat i la retenció de la comunitat de jugadors a la plataforma Steam, amb dades històriques que cobreixen el període 2012–2021. L'objectiu principal és entendre el comportament dels usuaris al llarg del temps i identificar quins videojocs han aconseguit mantenir o incrementar la seva base de jugadors.

Visualitzacions:

- **Indicadors globals:** Es mostren tres mètriques clau: el nombre mitjà de jugadors, el màxim nombre assolit i el total de jugadors guanyats respecte al mes anterior. Aquestes mètriques ofereixen una visió ràpida i comparativa de l'estat actual i passat de l'activitat dels jugadors, ajudant a detectar pics d'ús, tendències de creixement o signes de desinterès. Són útils per avaluar la salut general d'un videojoc o del mercat en un moment concret.
- **Filtre temporal i per videojoc:** L'usuari pot seleccionar intervals de dates i videojocs concrets per adaptar l'anàlisi al seu interès i comparar dinàmiques específiques de cada títol. Cada videojoc representat es mostra d'un color diferent. Permet fer una exploració personalitzada per comparar comportaments entre títols concrets o veure com ha evolucionat un sol joc en un període determinat, donant flexibilitat a l'anàlisi de la Hipòtesi 4.

- **Jugadors promig per any:** Un gràfic de barres apilades mostra com ha evolucionat la mitjana anual de jugadors per videojoc, posant en relleu els períodes de creixement o declivi d'activitat per cada títol analitzat. Ajuda a identificar quins jocs han mantingut una base de jugadors estable o han patit una davallada. També es poden detectar jocs que han anat guanyant popularitat a llarg termini, indicant una forta capacitat de retenció o una comunitat fidel.
- **Percentatge de jugadors guanyats per videojoc:** Un gràfic de barres horitzontals mostra quins jocs han guanyat o perdut més jugadors en relació amb el període anterior. Aquest gràfic és clau per analitzar la retenció i el creixement a curt termini. Permet detectar quins jocs tenen dinàmiques de fidelització o quins poden haver sofert una fuga de jugadors. La pèrdua de jugadors es representa amb color vermell i el guany amb color verd. És especialment útil per validar la Hipòtesi 4 amb dades objectives.



Visualització de dades-2025: Neil Pradas, Albert Guillaumet, Ilias Dahchouri, Lucia Garrido, Adrià Muro, David Morillo.

6. CONCLUSIONS

Després d'analitzar en profunditat les dades recopilades i visualitzades a través del dashboard creat amb Power BI, podem extreure conclusions valuoses sobre les dinàmiques del mercat de videojocs. A continuació, es presenten les conclusions obtingudes a partir de cada hipòtesi, basades en l'anàlisi de dades concretes i els resultats observats durant el desenvolupament del projecte.

- **Hipòtesi 1:** Els jocs desenvolupats per grans empreses (AAA) tenen un preu mitjà més elevat que els jocs indie o AA, però això no garanteix necessàriament un nombre de còpies venudes superior.

Confirmada parcialment. Les visualitzacions de la primera pàgina del dashboard mostren que els jocs AAA tenen, efectivament, un preu mitjà superior. No obstant això, no tots aconsegueixen destacar en volum de vendes. Alguns títols independents, tot i tenir preus més baixos, han obtingut un gran èxit comercial, fet que indica que el preu i la mida de l'empresa no són els únics determinants de l'èxit.

- **Hipòtesi 2:** Els jocs millor valorats pels usuaris obtenen més vendes que aquells amb valoracions baixes.

Confirmada. També, a partir de la primera pàgina, l'anàlisi de la relació entre valoració mitjana i volum de vendes mostra una tendència clara: els jocs amb puntuacions més altes solen vendre més. Tot i que hi ha excepcions puntuals, aquesta correlació és consistent i valida la influència de la qualitat percebuda en l'èxit comercial.

- **Hipòtesi 3:** Existeixen diferències notables en les preferències de gènere segons la regió geogràfica.

Confirmada. Aquesta hipòtesi es valida a la segona pàgina del dashboard. S'observen diferències notables en les

preferències de gènere per regió. Per exemple, el gènere d'acció és dominant globalment, però altres com el rol, l'esport o l'estratègia tenen presències molt variables segons el territori. Aquest resultat evidencia la importància de la localització i l'adaptació de continguts segons el mercat.

- **Hipòtesi 4:** Els videojocs amb una estratègia de preu ajustat aconseguixen una retenció de jugadors superior (mesurada per temps mitjà de joc).

No confirmada. A partir de les visualitzacions de la primera pàgina del dashboard s'ha pogut concloure que la relació entre preu i temps mitjà de joc no mostra una correlació clara. Alguns jocs barats tenen alts nivells de retenció, però també n'hi ha de cars amb una durada mitjana molt elevada. Per tant, la retenció sembla estar més relacionada amb la qualitat i l'experiència que no pas amb el cost.

- **Hipòtesi 5:** L'evolució del nombre de jugadors actius al llarg dels anys es veu influenciada per factors com ara expansions, descomptes, o tendències socials específiques.

Confirmada. La última pàgina del dashboard ha premès evidenciar que l'anàlisi temporal revela pics d'activitat coincidents amb llançaments de contingut addicional, promocions o altres esdeveniments rellevants. Aquest comportament demostra que accions específiques poden tenir un impacte directe en la participació dels usuaris i que la vida útil d'un videojoc pot allargar-se amb estratègies ben planificades.

7. POSSIBLES MILLORES

Tot i que el projecte ha proporcionat una visió valuosa sobre diversos factors que poden influir en l'èxit d'un videojoc, identifiquem algunes millores que podrien enriquir la visualització i aprofundir en les anàlisis:

- **Detall geogràfic més precís:** Un dels nostres objectius era representar les vendes per país mitjançant un mapa interactiu. No obstant això, els datasets disponibles només oferien informació per regions generals (Europa, Japó, Amèrica del Nord). Hauria estat molt enriquidor disposar d'un dataset amb vendes per país per poder visualitzar, mitjançant un mapa coroplètic interactiu, com varien les vendes de videojocs a escala mundial amb major granularitat.
- **Ubicació de les empreses desenvolupadores:** També vam considerar crear una visualització similar al mapa de vendes, però centrada en la localització geogràfica de les seues principals de les empreses desenvolupadores. La idea era analitzar si el fet que una empresa estigui situada en un determinat país o regió té un impacte positiu en les vendes dels seus jocs en aquella zona. Per fer-ho, s'hauria necessitat construir un nou dataset amb les coordenades geogràfiques de cada empresa i relacionar-lo amb les dades de vendes existents.
- **Anàlisi de les valoracions dels usuaris:** Una altra millora rellevant hauria estat incloure gràfiques detallades a partir de les *reviews* dels usuaris de Steam i de l'activitat a plataformes de streaming com Twitch.. Aquesta informació podria haver aportat insights sobre la qualitat percebuda dels jocs i la seva relació amb altres variables com les vendes, el preu o el gènere proporcionant informació valuosa pels desenvolupadors.
- **Model ML per predir rendiment abans del llançament:** Ens hagués agradat fer amb més dades una predicció respecte el rendiment d'un possible videojoc que es llançara en un futur mitjançant fonaments de machine learning.

Aquestes millores podrien contribuir a generar una anàlisi més rica i amb major profunditat geoespacial i qualitativa, obrint noves línies de recerca per a futures iteracions del projecte.

8. BIBLIOGRAFIA

- [1] Search | Kaggle. (s. f.). <https://www.kaggle.com/search>
- [2] Kfollis. (s. f.). Documentación de Power BI - Power BI. Microsoft Learn. <https://learn.microsoft.com/es-es/power-bi/>
- [3] Dashboard en Power BI. Ventas de videojuegos. – agustincastro.es. (2024, 7 maig). <https://agustincastro.es/index.php/2024/05/07/dashboard-en-power-bi-ventas-de-videojuegos/>