Tema: Analiza Pieței pentru Industria Jocurilor Video

Realizat de: Gogu Dan, gr. IA-211

Abstract

Acest articol își propune să prezică numărul de vânzări posibile pentru un anumit gen de joc. Setul de date a fost colectat de pe platforma internet cunoscută sub numele de Kaggle.com și a fost generat de vgchartz.com. Pentru folosirea setului de date, s-a folosit mediul de dezvoltare RStudio IDE și limbajul de programare R pentru curățarea, analiza și reprezentarea datelor. Algoritmul de învățare automată utilizat în acest proiect este regresia liniară. Având în vedere cunoștințele despre vânzările de jocuri video, ar fi fascinant să aflăm care sunt factorii necesari care fac ca un gen joc să se vândă mai bine decât alte genuri. De asemenea, intenționez să prezint rezultatele acestui studiu într-un mod intuitiv prin vizualizarea lor cu ajutorul ggplot2 în R. În cadrul acestui proiect, vom lua vânzările globale ca variabilă de răspuns și mă voi axa pe realizarea de predicții prin analiza celorlalte variabile din setul de date referitoare la genul jocului. Rezultatele vor ajuta companiile de producție de jocuri să înțeleagă factorii necesari generării unui joc de succes în ceea ce privește publicitatea.

Introducere

În era digitală în care ne desfășurăm viețile, industria jocurilor video a devenit o forță impresionantă, marcând o schimbare semnificativă în modul în care ne distrăm și recreăm. Cu milioane de jucători în întreaga lume și o diversitate impresionantă de titluri, analiza și înțelegerea tendințelor din spatele acestui univers dinamic au devenit cruciale. În acest context, voi prezenta un studiu detaliat al unui dataset, ce adună informații valoroase despre vânzările jocurilor video. Acest dataset nu este doar o simplă colecție de cifre; este o fereastră către comportamentul consumatorilor, preferințele lor în evoluție și influențele culturale care modelează această industrie în continuă expansiune. În cele ce urmează, voi explora aspecte precum topurile de vânzări, tendințele geografice și evoluția pieței, pentru a oferi o perspectivă asupra universului captivant al jocurilor video și modului în care acesta interacționează cu societatea noastră în schimbare accelerată.

Scopul acestui studiu este de a extrage informații importante din dataset-ul vast referitor la vânzările jocurilor video. Prin analiza detaliată a acestor date, ne propunem să identificăm modelele emergente, să înțelegem comportamentul consumatorilor și să anticipăm direcțiile viitoare ale acestei industrii dinamice. Acest demers are ca obiectiv principal furnizarea informațiilor esențiale pentru dezvoltatorii de jocuri și alți profesioniști din domeniul industriei jocurilor video. În plus, acest studiu își propune să ofere o înțelegere mai profundă a modului în care jocurile video influențează culturile locale și globale, precum și modul în care persoanele din diferite regiuni apreciază aceste jocuri. Prin elucidarea acestor aspecte, ne propunem să contribuim la luarea deciziilor informate și la promovarea unei creșteri sustenabile și echitabile în cadrul acestei industrii deosebit de dinamice.

Metode

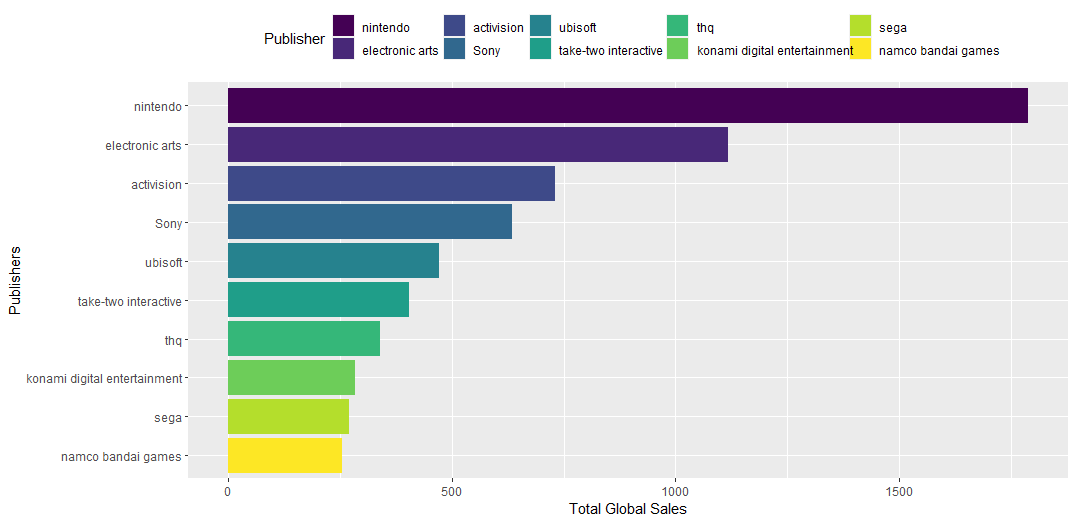
Datele au fost colectate în format Comma Separated Values de pe un site-ul VGChartz ce conține o bază de date a tuturor jocurilor existente, fiind selectate anume jocurile din 1988 până în 2016. Setul de date conține 16,719 rânduri cu 16 variabile. Pentru a realiza analize și preziceri bazate pe dataset-ul privind vânzările jocurilor video, am implementat diverse metode și tehnici de prelucrare a datelor în mediul de programare R. Iată câteva dintre aceste metode:

* **Explorarea Datelor (Exploratory Data Analysis - EDA):** Am început prin a efectua o analiză exploratorie a datelor pentru a înțelege structura setului de date, distribuția variabilelor și a identifica eventuale anomalii sau lacune. Folosind pachete precum dplyr și tidyverse, am realizat sumare statistice, histogramme și diagrame de dispersie pentru a obține o imagine cuprinzătoare a datelor noastre.
* **Curățarea Datelor (Data Cleaning):** Am aplicat tehnici de curățare a datelor pentru a gestiona missing values, duplicatelor și pentru a asigura consistența datelor. Funcții din pachetele tidyr și dplyr au fost folosite pentru a realiza aceste operațiuni.
* **Tehnici de Machine Learning**: Pentru a realiza preziceri cu privire la vânzările jocurilor video, am implementat modele de machine learning. Am utilizat, de exemplu, algoritmi de regresie pentru a anticipa vânzările viitoare pe baza datelor istorice. Pachete precum caret, randomForest sau glm au fost esențiale în implementarea acestor modele.
* **Validarea Modelului**: Am aplicat tehnici de validare a modelului pentru a asigura performanța și generalizarea corectă a predicțiilor. Validarea încrucișată (cross-validation) și măsurile precum R-squared sau Mean Squared Error au fost utilizate pentru a evalua performanța modelelor noastre.
* **Vizualizarea Rezultatelor**: Pentru a prezenta rezultatele într-un mod accesibil și informativ, am folosit biblioteci precum ggplot2 pentru a crea grafice și vizualizări relevante.

Prin aceste metode și tehnici implementate în R, am fost capabili să obținem nu doar o înțelegere profundă a datelor, ci și să realizăm preziceri semnificative cu privire la vânzările jocurilor video, oferind astfel o contribuție valoroasă la înțelegerea dinamicilor acestei industrii.

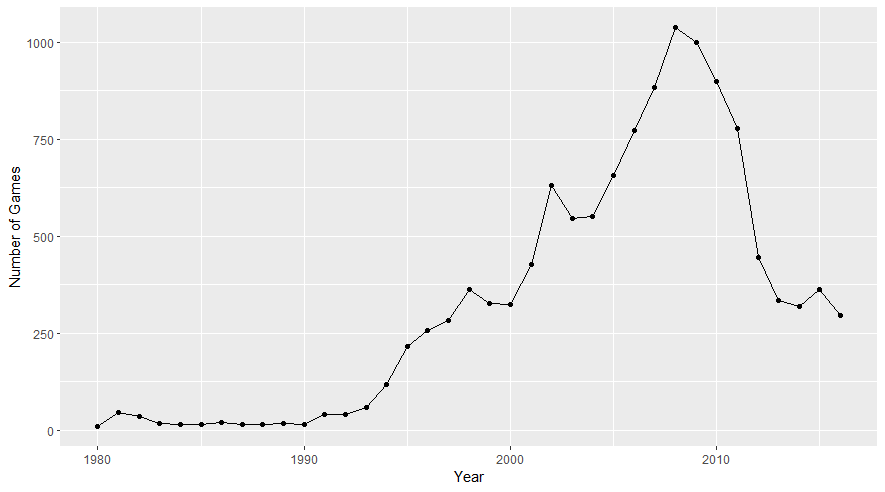
Rezultate

Primul pas a fost eliminarea observațiilor cu nume de jocuri lipsă și modificarea numelor subsidiarelor editorilor la numele companiilor-mamă. De exemplu, am schimbat editorii "Sony Computer Entertainment", "Sony Computer Entertainment America", "Sony Computer Entertainment Europe", "Sony Music Entertainment" și "Sony Online Entertainment" la "Sony". Începând cu începutul anilor 1980, au început să apară companii de editare. În setul nostru de date, există în total 582 de edituri, printre care se află companii precum ar fi: Sony, Nintendo, Electronic Arts, Ubisoft și altele. (Figura 1)



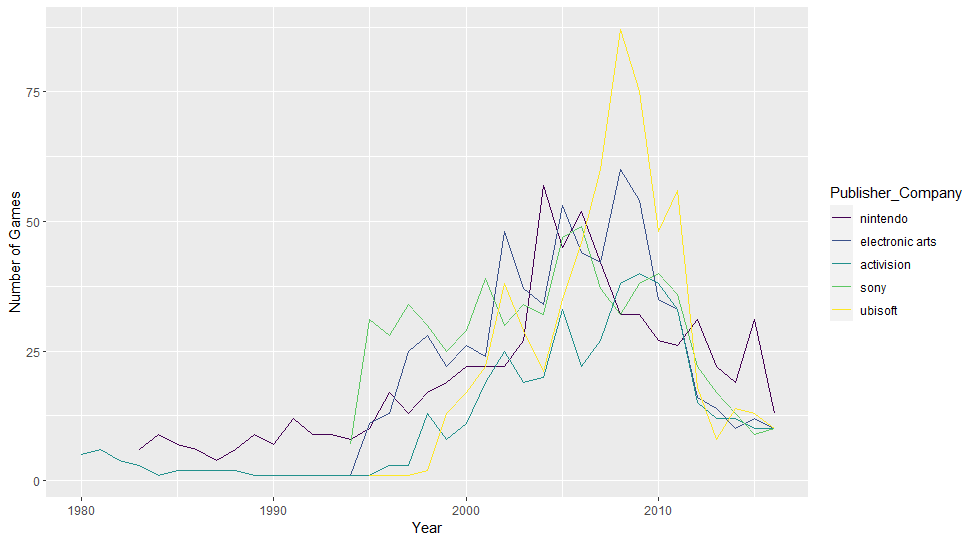
**Figura 1**. Vânzările globale ale primelor 10 companii.

Am dorit să aflu când industria jocurilor a experimentat o dezvoltare drastică. Bazându-mă pe numărul de jocuri lansate în fiecare an, am constatat că competiția a devenit intensă la sfârșitul anilor 2000. În acea perioadă, au fost lansate multe jocuri noi. (Figura 2)



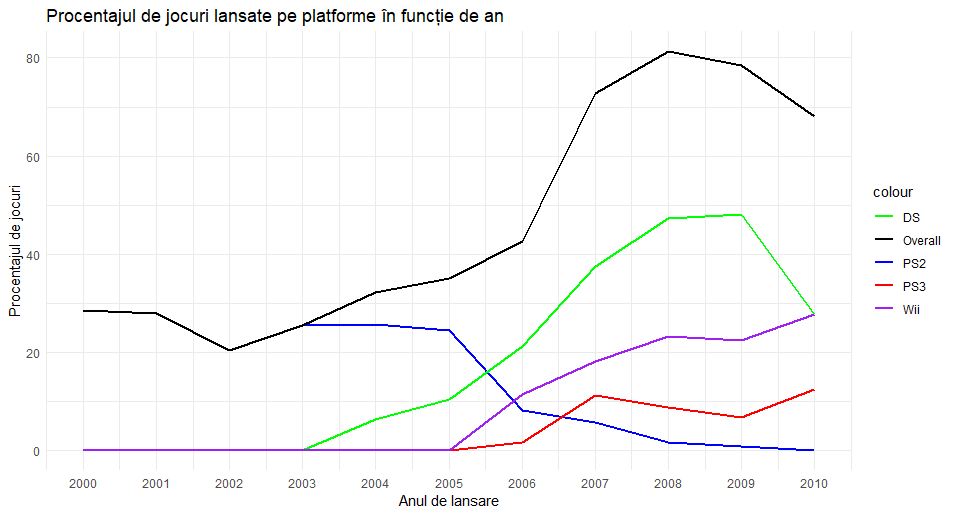
**Figura 2**. Numărul de jocuri lansate în diferiți ani.

Pentru primii 5 editori, Ubisoft a înregistrat o creștere drastică în ultimii ani ai decadelor 2000, deși este clasată doar pe locul 5 în funcție de vânzările totale. Electronic Arts a avut, de asemenea, o creștere semnificativă în acea perioadă. Ambii editori au fost relativ tineri comparativ cu alți editori de top, lansând primele lor jocuri în 1997 conform datelor disponibile. De asemenea, am observat că Nintendo și Sony au înregistrat o creștere semnificativă în mijlocul anilor 2000 (Figura 3). Prin cercetări online, am aflat că Nintendo și Sony sunt doi dintre cei trei producători de platforme de top (cel de-al treilea este Microsoft cu seria Xbox). În total, am 17 platforme în datele mele. Bazându-ne pe numărul de jocuri lansate pe fiecare platformă, 9 dintre primele 10 platforme sunt dezvoltate de Nintendo, Sony și Microsoft. Este important să menționăm că industria jocurilor video a evoluat rapid în această perioadă, iar succesul editorilor poate fi atribuit diverselor factori, cum ar fi lansarea de titluri populare, adaptarea la noile tehnologii și inovații în modul în care jocurile sunt dezvoltate, promovate și distribuite. Cercetarea online și analiza mai detaliată a lansărilor de jocuri și strategiilor de piață ale acestor editori ar putea oferi o perspectivă mai clară asupra motivelor succesului lor și a schimbărilor din industrie în acele perioade.



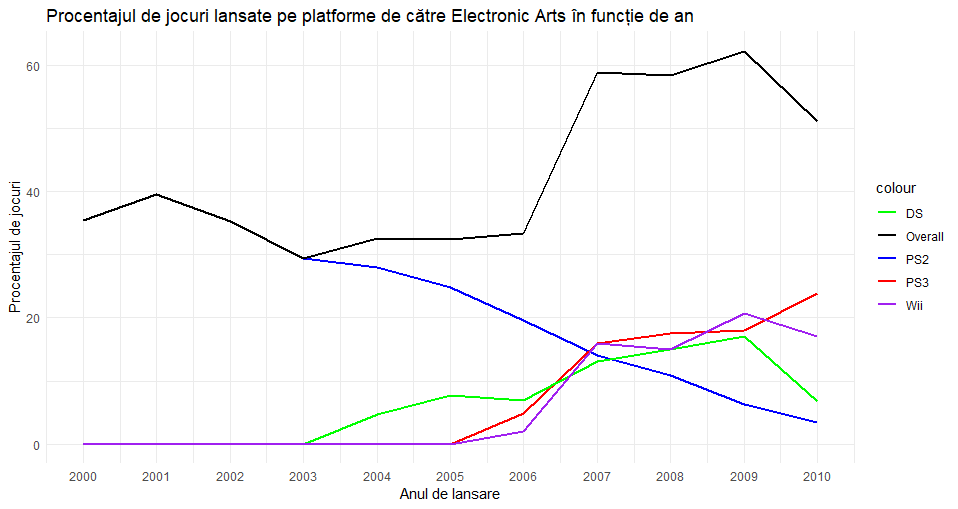
**Figura 3**. Numărul de jocuri vândute în fiecare an ale celor 5 companii.

Sony a lansat PS2 și PS3 în anii 2000 și 2006, în timp ce Nintendo a lansat DS și Wii în 2004 și 2006. PS2, PS3, DS și Wii sunt enumerate ca cele mai bune 4 platforme de până acum. Întorcându-mă la întrebările pe care le aveam, principalul motiv pentru întârzierea dezvoltării editorilor tineri ar putea fi faptul că Sony și Nintendo au lansat noi platforme și au lansat o mulțime de jocuri noi pe acestea. În decurs de 3 ani, alți editori au urmat trendul și au dezvoltat mai multe jocuri noi pe aceste platforme, generând o explozie în industrie. Acum putem verifica ipoteza noastră. Rezultatele următoare pentru Ubisoft și Electronic Arts arată numărul de jocuri lansate pe PS2, PS3, DS și Wii, respectiv, în perioada 2000-2010, urmate de procentajele lor din totalul jocurilor. Ultima coloană "Overall" se referă la procentajul de jocuri pe cele patru platforme în total. Între 2007 și 2009, Ubisoft a lansat peste 70% dintre jocurile pe aceste patru platforme noi. (Figura 4)



**Figura 4**. Procentajul de jocuri lansate pe diferse platforme în functie de an al companiei Ubisoft.

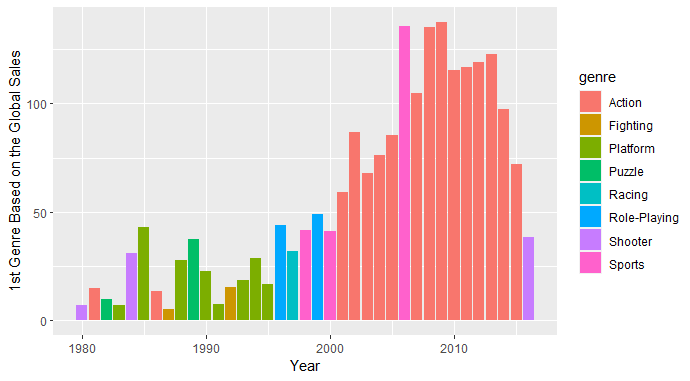
De asemenea, Electronic Arts a lansat aproximativ 60% dintre jocuri pe acestea. (Figura 5)



**Figura 5**. Procentajul de jocuri lansate pe diferse platforme în functie de an al companiei Electionic Arts.

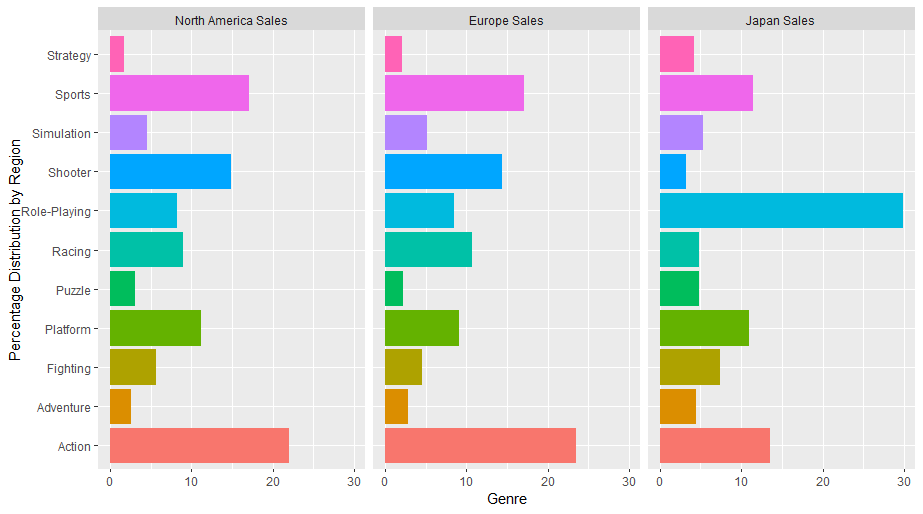
Acum, vom investiga cele mai bune 4 edituri și vom analiza cele mai bune 3 genuri ale acestora, precum și cele mai populare / profitabile 10 jocuri ale fiecăreia. Wii Sports este un success imens pentru Nintendo. Jocurile clasice “Mario” ocupă 5 dintre primele 10 cele mai populare jocuri (Figura 8). Cât despre Activision, jocurile de tip Shooter generează aproape jumătate din vânzările totale. Cele mai bune 10 jocuri ale companiei sunt toate de tip Shooter, iar 9 dintre acestea fac parte din seria "Call of Duty" (Figura 9). Pentru Sony, jocurile de tip Racing, generează 20.1% din vânzări, anume din seria de jocuri "Gran Turismo" (Figura 10). Genul "Acțiune" este cel mai popular pentru Ubisoft datorită seriei "Assassin's Creed" (Figura 11).

Să vedem care este genul cel mai popular pentru fiecare an, bazat pe Vânzările Globale Totale și Numărul de Jocuri lansate. Am combinat cele două metrici deoarece Vânzările Globale Totale din setul nostru de date sunt cumulative și nu cunoaștem exact numărul de vânzări pentru fiecare an, așa că am folosit numărul de jocuri lansate pe gen pentru a susține rezultatele. Începând cu secolul al XXI-lea, jocurile de acțiune se mențin pe primul loc în topul genurilor de jocuri video timp de peste 10 ani. Într-un cluster, observăm o bară înaltă pentru "Sport" în 2006 (Figura 6). Acest lucru se datorează faptului că jocul Wii Sports a fost lansat în 2006 și vânzările globale totale au atins nivelul remarcabil de 82,5 milioane.



**Figura 6**. Cel mai vândut tip de joc pentru fiecare an.

Atunci când explorăm industria jocurilor, este important să știm dacă popularitatea diferitelor genuri are modele similare între regiuni. Așadar, am calculat și comparat distribuția procentuală a genurilor pentru America de Nord, Europa și Japonia. Modelele pentru America de Nord și Europa sunt foarte similare. Ambele au genurile Acțiune, Sport și Shooter ca cele mai populare 3 genuri. Piața din Japonia este foarte diferită de acestea. Între cele 11 genuri, jocurile Roleplay au cea mai mare pondere în vânzări, de peste două ori mai mare decât vânzările jocurilor de acțiune, care se clasează pe locul al doilea. În mod surprinzător, genul Shooter, al treilea cel mai popular gen în America de Nord și Europa, se află pe ultimul loc în Japonia (Figura 7).



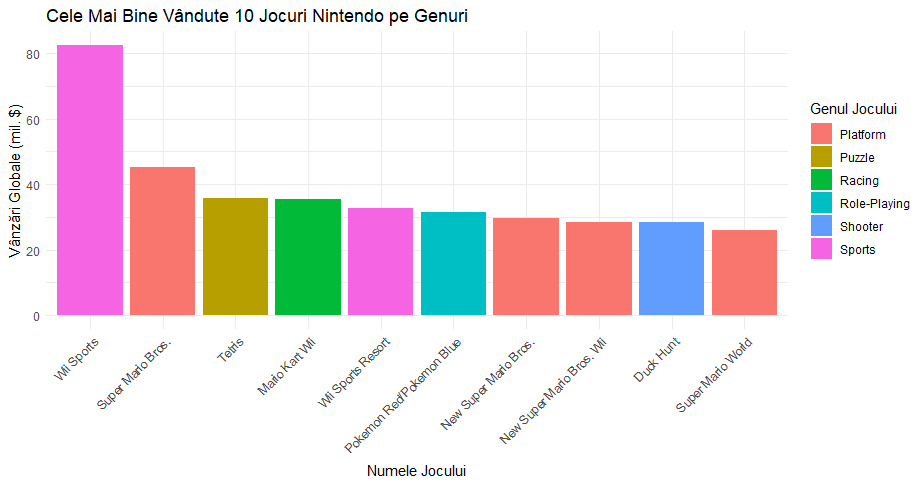
**Figura 7**. Procentul vânzărilor unui tip de joc în diferite regiuni.

Analizând rezultatele regresiei liniare aplicate setului de date privind vânzările jocurilor video, putem observa că modelul prezintă o precizie moderată. Erorile evaluative indică o deviere relativ mică a predicțiilor față de valorile reale, evidențiate de Root Mean Square Error (RMSE) de 1.29367. Cu toate acestea, coeficientul de determinare (R-squared) de 0.1678688 sugerează că doar aproximativ 16.8% din variabilitatea datelor poate fi explicată de model, indicând o performanță limitată în explicarea relației dintre variabilele independente și vânzările de jocuri video (Figura 12). Mean Absolute Error (MAE) de 1.018265 reflectă o medie a discrepanțelor absolute dintre predicții și observații. În concluzie, deși modelul pare să ofere o aproximare moderată, există o nevoie potențială de îmbunătățire a preciziei pentru a dezvălui relații mai profunde în datele de vânzare a jocurilor video.

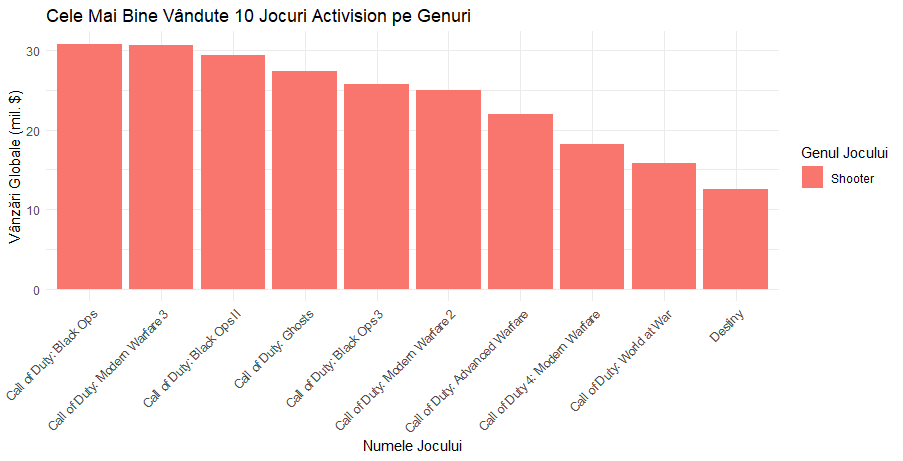
Codul programului realizat în R, este disponibil la [r-code/code.r at main · dGxxxx/r-code (github.com)](https://github.com/dGxxxx/r-code/blob/main/code.r).

Materiale Suplimentare

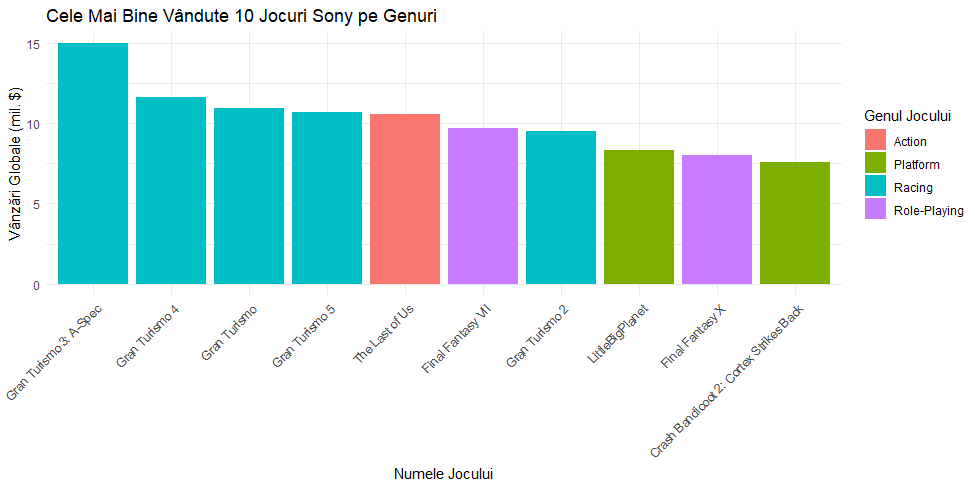
**Anexa 1.** Figuri adăugătoare folosite pentru verificarea ipotezelor.



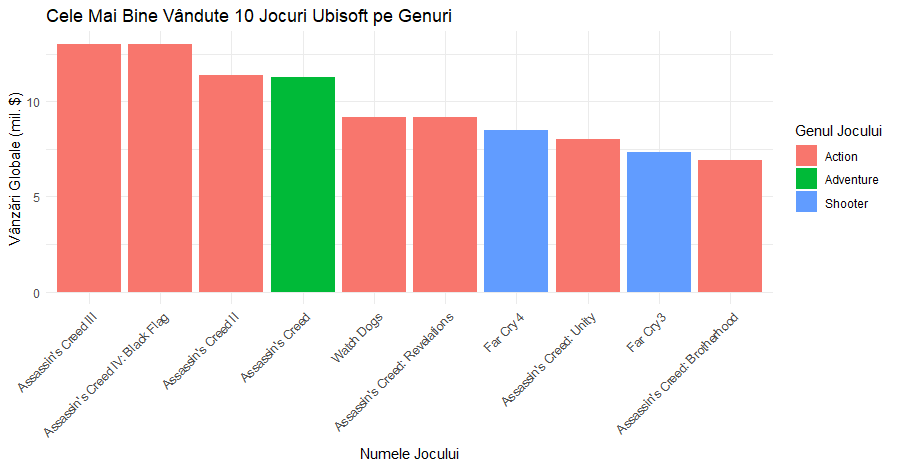
**Figura 8**. Top 10 jocuri vândute de către Nintendo și genul acestora.



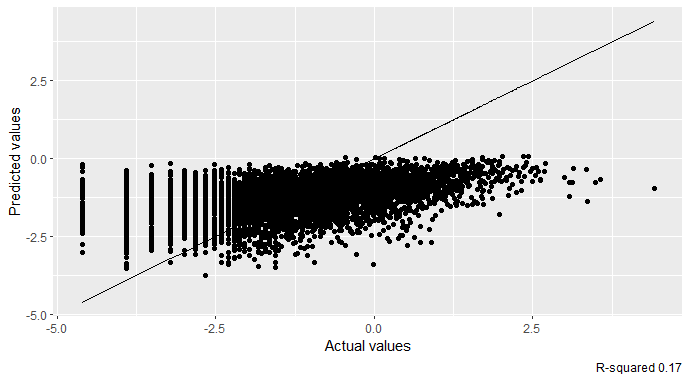
**Figura 9**. Top 10 jocuri vândute de către Nintendo și genul acestora.



**Figura 10**. Top 10 jocuri vândute de către Sony și genul acestora.



**Figura 11**. Top 10 jocuri vândute de către Ubisoft și genul acestora.



**Figura 12**. Valori prezise și valorile adevarate.

**Anexa 2.** Variabilele folosite pentru preziceri și analiza setului de date:

* Coloanele: **na\_sales, eu\_sales, jp\_sales, other\_sales și global\_sales**, ne pot ajuta să anticipăm tendințele de popularitate a genurilor de jocuri. Cuantificând vânzările în fiecare regiune, putem identifica genurile de jocuri care se bucură de popularitate într-o anumită zonă.
* Coloanele: **critic\_score, critic\_count, user\_score, user\_count**, ne pot ajuta să verificăm dacă există o corelație puternică între aceste scoruri și vânzările într-o anumită regiune, dezvoltatorii pot să-și direcționeze resursele către dezvoltarea și promovarea jocurilor care primesc recenzii mai bune.

Bibliografie

1. R: The R Project for Statistical Computing [Internet]. [citat 12 decembrie 2023]. Disponibil la: https://www.r-project.org/
2. Tidyverse [Internet]. [citat 12 decembrie 2023]. Disponibil la: https://www.tidyverse.org/
3. A Grammar of Data Manipulation [Internet]. [citat 12 decembrie 2023]. Disponibil la: https://dplyr.tidyverse.org/
4. Tidy Messy Data [Internet]. [citat 12 decembrie 2023]. Disponibil la: https://tidyr.tidyverse.org/