



Univerzitet u Zenici
Politehnički fakultet
Softversko inženjerstvo



Analiza podataka o video igrama i Google trendovima: Istraživanje popularnih platformi, žanrova i geografskih razlika

Rudarenje podataka

Studenti: Adin Jahić, Merjem Bajramović
Br. Indexa : II-80, II-77

Profesor: Doc. dr. sc. Denis Čeke
Saradnik: mr. Fuad Hajdarević

Zenica, juni 2023. godine

Sadržaj

Uvod	3
1. Opis skupa podataka.....	3
1.1. Primjena algoritama rudarenja nad korištenim skupom podataka.....	4
2. Analiza podataka	5
2.1. Analiza platformi za igranje	5
2.2. Analiza žanrova.....	8
2.3. Analiza izdavača.....	10
3. Klasterizacija	11
3.1. K-Means klasterizacija korištenjem metode lakta.....	12
3.2. K-Means klasterizacija korištenjem metode koeficijenta siluete	16
3.2.1. Uvid u karakteristike klastera kroz sažete statistike.....	17
3.2.2. Proučavanje sažetih statističkih podataka za kategoričke varijable unutar klastera	18
4. Poređenje rezultata klasterizacije sa Google Trends podacima.....	20
4.1. Analiza prosječne prodaje top 5 žanrova sa Google Trends pretraživanjima.....	20
4.2. Analiza prosječne prodaje top 5 platformi sa Google Trends pretraživanjima	22
Zaključak.....	24
Reference.....	25

Uvod

Analiza podataka o video igrama postaje sve važniji alat u razumijevanju dinamike industrije video igara. Sa stalnim rastom i razvojem gaming pejzaža, korištenje tehnika rudarenja podataka kako bi se dobili uvidi u popularne platforme, žanrove i geografske preference postaje imperativno. Ovaj rad ima za cilj istražiti široke mogućnosti koje pruža rudarenje podataka u otkrivanju vrijednih obrazaca i trendova unutar sfere video igara.

U posljednjim godinama, dostupnost ogromne količine podataka povezanih s video igrama otvorila je nove pravce za istraživanje i analizu. Ovaj rad koristi sveobuhvatan set podataka preuzet sa *Kaggle*, poznate platforme za razmjenu skupova podataka i resursa za proučavanje podataka. Navedeni skup podataka obuhvata prodaju video igara iz različitih regija i uključuje vrijedne attribute kao što su žanrovi igara, platforme, izdavači i podaci o prodaji. Koristeći ove podatke, moguće je dublje pogledati u složenost industrije video igara u pogledu preferenci potrošača i tržišnih trendova.

Za provođenje analize podataka, korišten je programski jezik R[1] unutar okruženja Jupyter Notebook. R nudi širok spektar statističkih i vizualizacijskih paketa, što ga čini prikladnim alatom za istraživanje i tumačenje složenih setova podatak. Kombinacijom R-a i Jupyter Notebook-a, integrirani su kôd, analiza i vizualizacije, olakšavajući sveobuhvatno istraživanje korištenih podataka.

Jedan od ključnih ciljeva ovog rada je identifikacija najpopularnijih žanrova igara, platformi i izdavača u različitim regijama i globalno. Za postizanje navedenog, korištene su tehnike rudarenja podataka, uključujući algoritam *K-Means* klasterizacije. Algoritmi za grupisanje su posebno korisni u eksploratornoj analizi podataka, omogućavajući grupisanje sličnih tačaka podataka i identifikaciju obrazaca i veza koje možda nisu odmah očigledne. *K-Means* klasterizacija, posebno, particionira podatke u predefrirani broj klastera na osnovu mjera sličnosti.

Određivanje optimalnog broja klastera je kritičan korak u procesu K-Means klasterizacije. U ovom radu je korišteno nekoliko tehnika za određivanje odgovarajućeg broja klastera za analizu. Ove tehnike uključuju *Elbow Method*, koja ocjenjuje varijancu objašnjenu svakim dodatnim klasterom, i *Silhouette Coefficient*, koji mjeri kompaktnost i separaciju klastera. Iskorištavajući ove tehnike, moguće je uspostaviti pouzdano i smisleno rješenje za grupisanje podataka za analizu.

Pored seta podataka o prodaji video igara, kako bi se dodatno obogatila analiza, uključeni su podaci iz Google Trends[2]. Google Trends pruža vrijedne uvide u popularnost i interesovanje za pretragu određenih ključnih riječi tokom vremena. Istražujući trendove pretrage za popularne igre unutar definiranih vremenskih perioda, moguće je otkriti bilo korelacije između prodajnih brojki i javnog interesa. Ova dodatna dimenzija omogućava donošenje sveobuhvatnijih zaključaka.

U nastavku će biti predstavljena zapažanja koje proizlaze iz analize prethodno spomenutog skupa podataka.

1. Opis skupa podataka

Skup podataka "Video Game Sales"[3] obuhvata informacije o više od 16.500 video igara, te pruža vrijedne uvide u prodaju video igara i obuhvata različite informacije vezane za ovu industriju. Sastoji se od nekoliko kolona (*Slika 1*) koje predstavljaju različite attribute video igara i njihovu prodaju u različitim regijama:

- **Rank** – označava rangiranje određene igre na osnovu prodaje, pri čemu niži rang ukazuje na veću prodaju.
- **Name** – predstavlja naslov svake video igre.
- **Platform** – specificira gejming platformu ili konzolu na kojoj je igra objavljena, kao što su PlayStation, Xbox, PC,...
- **Year** – označava godinu u kojoj je igra objavljena. Ova informacija je ključna za analizu trendova i razumijevanje razvoja industrije video igara tokom vremena.
- **Genre** – kategorizuje igre u različite žanrove ili kategorije, kao što su akcija, sport, itd. Ova klasifikacija omogućava proučavanje popularnih žanrova i njihovih prodajnih rezultata.
- **Publisher** – pruža ime kompanije ili organizacije odgovorne za izdavanje i distribuciju igre. Pruža uvide u različite subjekte uključene u proizvodnju i marketing video igara.
- Kolone **NA_Sales**, **EU_Sales**, **JP_Sales**, **Other_Sales** i **Global_Sales** predstavljaju prodajne rezultate svake igre u milionima za različite regije: Sjeverna Amerika, Evropa, Japan, druge regije i globalno, redom. Ove kolone pružaju sveobuhvatan prikaz popularnosti igre i njenog komercijalnog uspjeha na različitim tržištima.

Skup podataka uključuje igre objavljene u periodu od 1980. do 2016. godine. Ova široka vremenska pokrivenost omogućava analizu trendova i promjena u industriji video igara tokom četiri decenije. Također pruža mogućnost proučavanja evolucije preferenci potrošača, popularnih platformi i žanrova igara tokom vremena.

Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
1	Wii Sports	Wii	2006	Sports	Nintendo	41.49	29.02	3.77	8.46	82.74
2	Super Mario Bros.	NES	1985	Platform	Nintendo	29.08	3.58	6.81	0.77	40.24
3	Mario Kart Wii	Wii	2008	Racing	Nintendo	15.85	12.88	3.79	3.31	35.82
4	Wii Sports Resort	Wii	2009	Sports	Nintendo	15.75	11.01	3.28	2.96	33
5	Pokemon Red/Pokemon Blue	GB	1996	Role-Playing	Nintendo	11.27	8.89	10.22	1	31.37
6	Tetris	GB	1989	Puzzle	Nintendo	23.2	2.26	4.22	0.58	30.26
7	New Super Mario Bros.	DS	2006	Platform	Nintendo	11.38	9.23	6.5	2.9	30.01

Slika 1 Prvih sedam redova skupa podataka "Video Game Sales"

1.1. Primjena algoritama rudarenja nad korištenim skupom podataka

Algoritam K-sredina (K-Means) je popularna tehnika nenadziranog učenja koja se koristi za klasterizaciju tačaka podataka u odvojene grupe na osnovu njihove sličnosti. U kontekstu skupa podataka o prodaji video igara, algoritam K-sredina može pomoći u identifikaciji obrazaca i grupisanju sličnih igara na osnovu njihovih prodajnih rezultata.

Kako bi se primijenio algoritam K-sredina, potrebno je odabrati atribut iz skupa podataka koji se žele uzeti u obzir prilikom klasterizacije. Na primjer, može se fokusirati na kolone kao što su "Global_Sales", "Genre" i "Platform" kako bi se obuhvatile informacije o prodajnom uspjehu, žanru i platformi igara.

Potrebno je odrediti broj klastera (k) koje algoritam treba da identifikuje. Određivanje optimalne vrijednosti k je ključno kako bi se dobili smisleni i tumačljivi klasteri. Prilikom određivanja optimalne vrijednosti k, do izražaja dolaze tehnike poput metode lakta (*Elbow Method*) i koeficijenta siluete (*Silhouette Coefficient*).

Metoda lakta uključuje pokretanje algoritma K-sredina za niz vrijednosti k i izračunavanje sume kvadrata udaljenosti između tačaka podataka i njihovih dodijeljenih centroida klastera. Prikazivanjem

sume kvadrata udaljenosti u odnosu na broj klastera, moguće je uočiti grafik koji nalikuje obliku lakta. "Tačka lakta" (*elbow point*) predstavlja vrijednost k kod koje dodatni klasteri neznatno poboljšavaju kvalitet klasterovanja. Ova tačka se može smatrati optimalnim brojem klastera za korišten skup podataka.

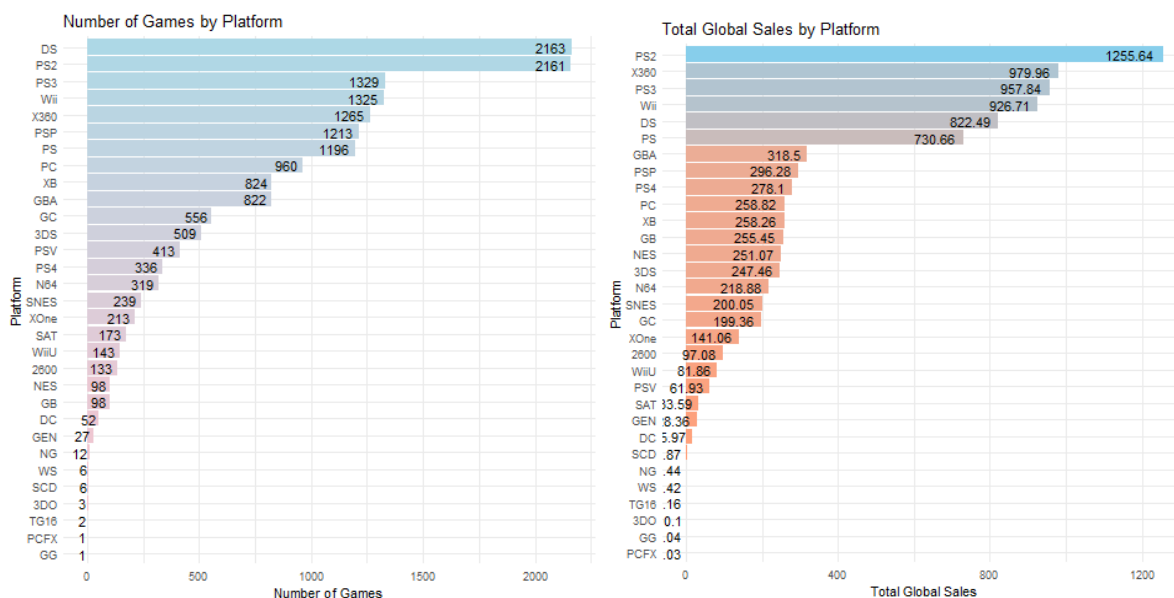
S druge strane, koeficijent siluete mjeri kompaktnost i razdvajanje klastera. On kvantifikuje koliko dobro svaka tačka podataka odgovara svom dodijeljenom klasteru u odnosu na susjedne klastere. Viši koeficijent siluete ukazuje na bolji kvalitet klasterizacije. Izračunavanjem koeficijenta siluete za različite vrijednosti k, moguće je identifikovati vrijednost koja daje najviši koeficijent, čime se određuje optimalan broj klastera.

Nakon određivanja optimalne vrijednosti k, moguće je pokrenuti algoritam K-sredina s izračunatim optimalnim brojem klastera. Algoritam će dodijeliti svaku igru jednom od identifikovanih klastera na osnovu njihove sličnosti u pogledu, u ovom primjeru, prodajnih rezultata, žanra i platforme. Ovo rješenje klasterizacije može pružiti vrijedne uvide u različite grupe igara unutar skupa podataka i njihove prodajne obrasce.

2. Analiza podataka

2.1. Analiza platformi za igranje

Kako bi se dobio uvid u popularnost platformi za igre, obavljena je analiza korištenjem dostupnog skupa podataka. Cilj je identificirati platforme koje imaju značajno prisustvo u industriji video igara na temelju broja proizvedenih igara. Koristeći jednostavnu metodu brojanja u R-u, izračunat je broj igara po platformi (*Slika 2*). Ovo omogućava kvantificiranje proizvodnje igara za svaku platformu i određivanje njihove relativne popularnosti.



Slika 2 Broj igara po platformi (lijevo), ukupna prodaja po platformi (desno)

Analizom podataka o ukupnoj prodaji (*Slika 2*) moguće je primijetiti da Playstation 2 (PS2) predvodi sve ostale platforme, slijede ga Xbox 360 (X360), Playstation 3 (PS3), Wii i DS, što ilustruje ogromnu popularnost ovih platformi na globalnom nivou.

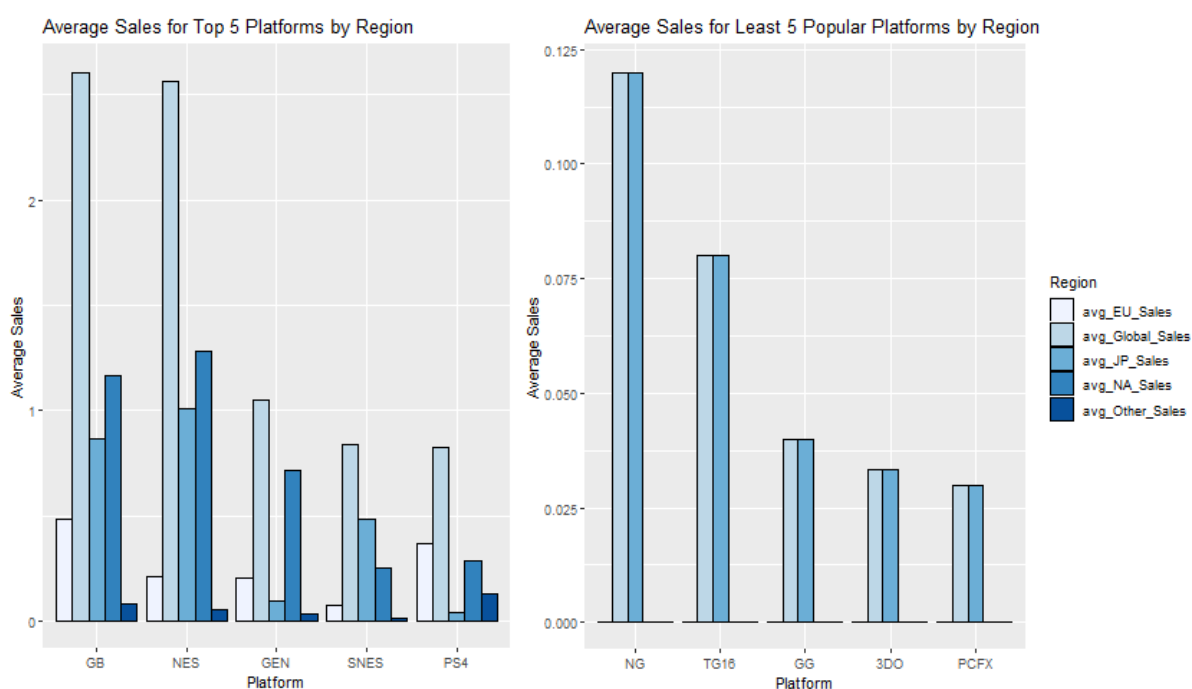
Međutim, prelaskom na analizu prosječne prodaje po regijama (*Slika 3*), redoslijed se značajno mijenja. Na primjer, prilikom analize prosječne prodaje u Sjevernoj Americi (NA), može se vidjeti da Game Boy (GB) i Nintendo Entertainment System (NES) predvode, iako se ne nalaze među prvih 5 po ukupnoj globalnoj prodaji. Ovo može biti rezultat ogromne popularnosti ovih platformi tokom njihovog vrhunca na tržištu Sjeverne Amerike.

U Evropi (EU), Playstation 4 (PS4) ima najveću prosječnu prodaju, što ukazuje na snažnu preferencu tržišta za novije Playstation konzole u Evropi. Xbox One (XOne) predvodi prosječnu prodaju u 'Ostalim' regijama, što može ukazivati na raznolik interes za platforme za igranje u tim regijama.

Japan (JP) predstavlja potpuno drugačiju sliku, gdje DS, prenosna platforma, ima najveću prosječnu prodaju. Ova zaznanja mogu odražavati jedinstvene potrošačke preference u Japanu za prenosne igre, zbog faktora kao što su dugi vremenski periodi putovanja. Međutim, važno je napomenuti da 3DO, GG, PCFX i NG, uprkos tome što se nalaze među zadnjih 5 po ukupnoj prodaji, imaju veću prosječnu prodaju u Japanu nego u nekim drugim regijama, što ukazuje na to da ove platforme imaju specifičnu, ali posvećenu korisničku bazu na Japanskom tržištu.

Suprotno tome, zanimljivo je primijetiti da platforme sa najvećom ukupnom prodajom (PS2, X360, PS3, Wii, DS) ne moraju nužno imati najveću prosječnu prodaju (*Slika 2*) u svim regijama. Iako ove platforme imaju široku globalnu privlačnost, ne dominiraju na svakom pojedinačnom tržištu.

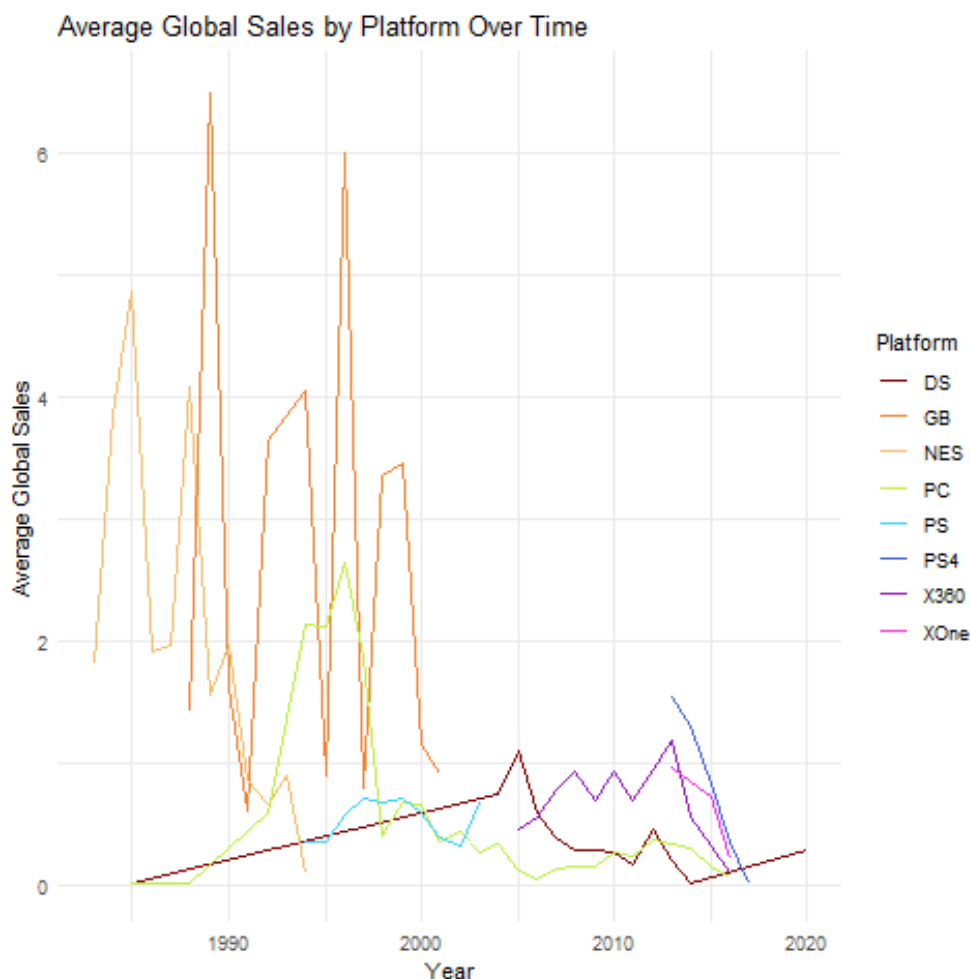
Na kraju, gledajući globalnu prosječnu prodaju, jasno je da GB, NES, Genesis (GEN) i WiiU, platforme koje se ne nalaze među prvih 5 po ukupnoj prodaji, imaju snažne brojke prosječne prodaje. To sugerira da su ove platforme imale periode visoke prodaje, čak i ako nisu uspjele održati te brojke na dugoročnom planu.



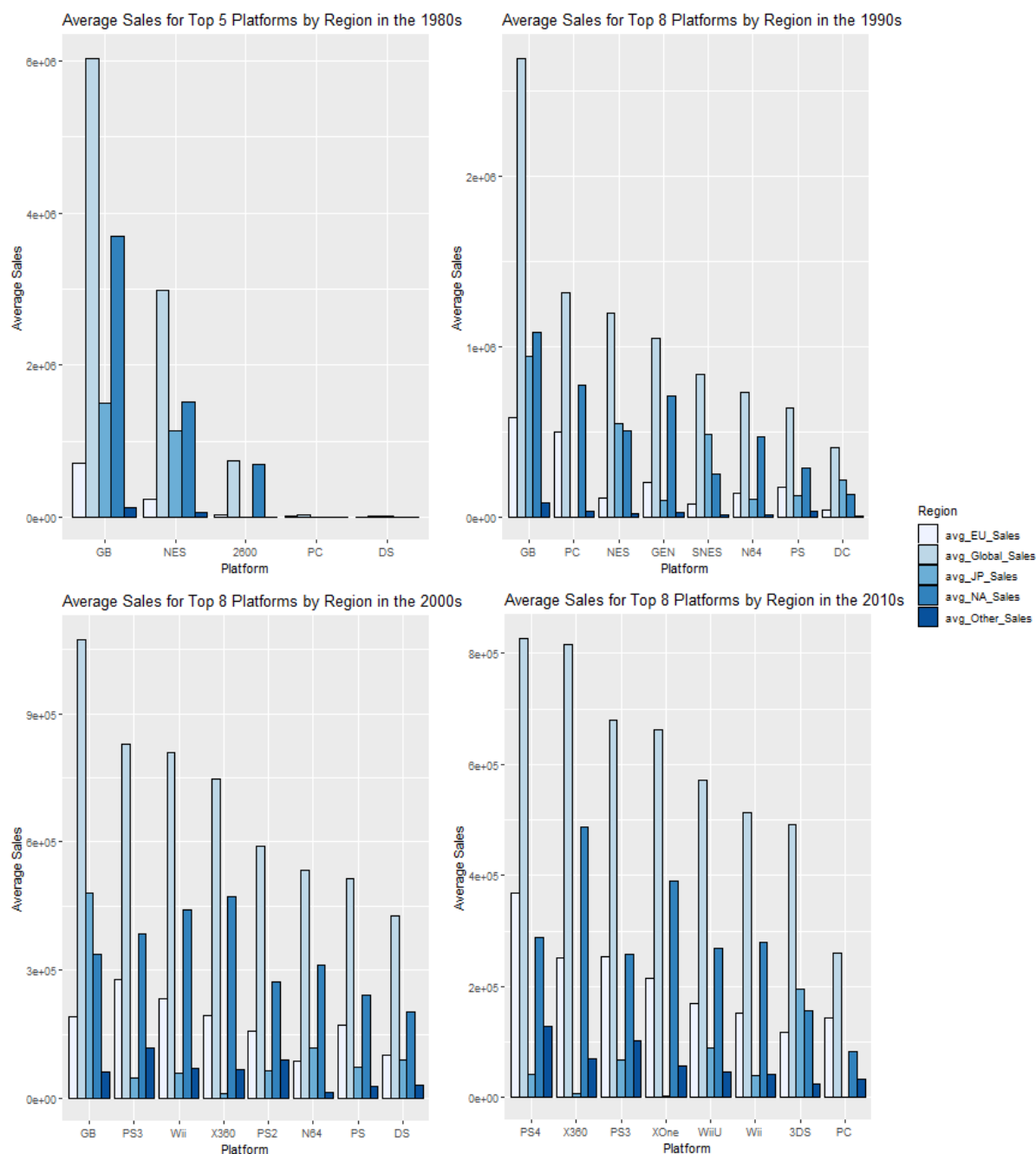
Slika 3 Analiza prosječne prodaje prvih (lijevo) i posljednjih (desno) 5 platformi za igranje po regijama

Detaljnijim pregledom prosječne prodaje najpopularnijih 5 platformi kroz decenije (*Slike 4 i 5*) – 1980, 1990, 2000 i 2010, može se zapaziti evolucija, koja se ogleda ne samo u promjenama preferenci platformi za igranje, već i u sve većem geografskom proširenju tržišta. Proučavajući ove promjene, mogu se izvući vrijedni uvidi koji opisuju razvoj industrije video igara od njenih ranih faza do savremenog globalnog dosega:

- **Regionalna dominacija:** U 1980-ima i 1990-ima primjetno je da su neke platforme, poput Game Boy-a (GB) i Nintendo Entertainment System-a (NES), bile pretežno uspješne u Sjevernoj Americi i Japanu (*Slika 5*). Međutim, u 2000-ima i 2010-ima trend je počeo da se mijenja. Na primjer, platforme PS3 i PS4 su se izvanredno dobro kotirale u Sjevernoj Americi, Evropi i drugim regionima, što pokazuje više globalizirano tržište video igara.
- **Dugovječnost platformi:** Nekoliko platformi pokazuje trajnu prisutnost kroz više decenija. Na primjer, PC je neprestano imao snažnu prodaju od 1980-ih (*Slika 4, zelena*), s postepenim rastom u Evropi i Sjevernoj Americi, dok su druge platforme, poput GB-a i NES-a, postepeno izbljedjele nakon početnog uspjeha.
- **Uspon novih platformi:** Svaka decenija donosi dominantne platforme koje imaju značajan utjecaj na tržište. U 1990-ima se pojavio PlayStation (PS), 2000-e su obilježile uspon PS2 i PS3, dok su u 2010-im uspjeh ostvarile PS4 i Xbox One (XOne) platforme.
- **Jedinstveno Japansko tržište:** Japan često predstavlja drugačije popularne platforme u odnosu na Sjevernu Ameriku i Evropu. Na primjer, platforma DS je ostvarila značajnu prodaju u Japanu u 1980-ima prije nego što je stekla popularnost u drugim regionima (*Slika 5*).



Slika 4 Prikaz globalnih prodaja za 8 top platformi kroz vrijeme



Slika 5 Grafički prikaz prosječne prodaje za top 5 platformi za regije kroz decenije

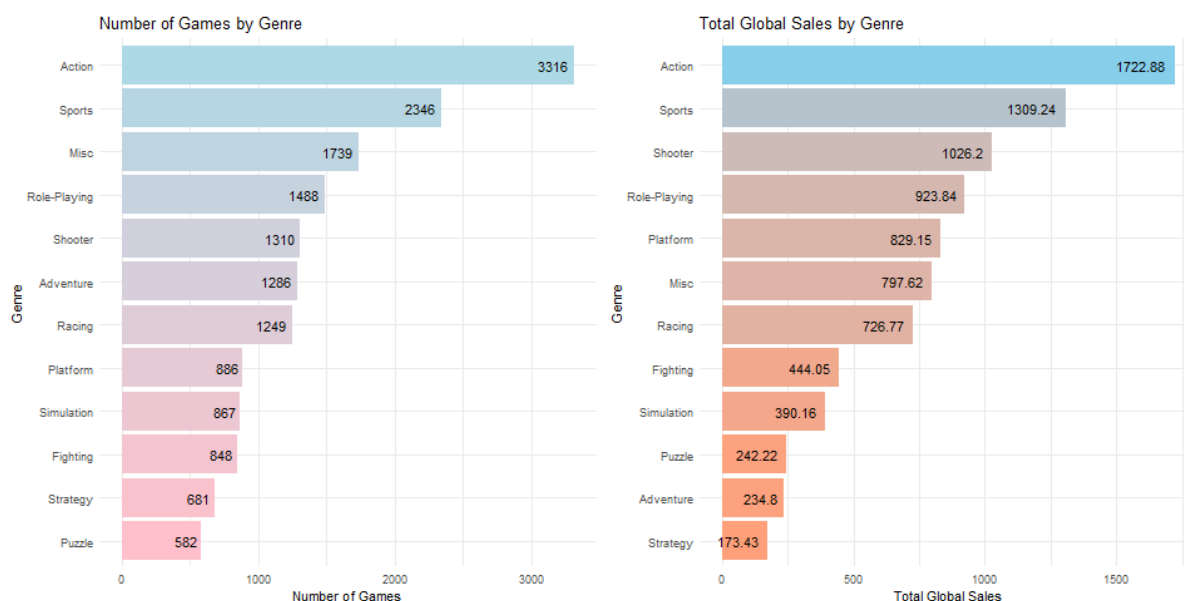
2.2. Analiza žanrova

Veza između broja proizvedenih igara unutar određenog žanra i globalne prodaje tog žanra (Slika 6) pruža zanimljive uvide u potrošačko ponašanje i trendove industrije. Iako bi se na prvi pogled očekivao direktan odnos između ova dva aspekta (tj. žanr sa većim brojem igara bi prirodno imao veću prodaju zbog većeg izbora za potrošače) podaci ukazuju na suptilniju stvarnost.

Uzimajući u obzir slučaj žanra akcija, koji ima najveći broj proizvedenih igara, a akoder vodi u ukupnoj globalnoj prodaji, što ukazuje da veliki obim proizvodnje igara može rezultirati većom prodajom. To može biti rezultat različitih faktora kao što su širi apel potrošača, svestranost u dizajnu igara ili veća vjerovatnoća stvaranja hit igre upravo zbog velikog obima proizvodnje.

Međutim, posmatrajući žanrove sportskih igara i pucačina, primjetan je suprotan trend. Uprkos tome što je žanr sportskih igara proizveo značajno više igara od žanra pucačina, upravo je žanr pucačina ostvario veću prosječnu globalnu prodaju. Ova razlika između obima proizvodnje i uspjeha u prodaji može biti utjecana brojnim faktorima. Na primjer, publika za pucačke igre, iako možda manja, može biti posvećenija i time vjerovatnije kupovati nove naslove. Alternativno, moguće je da igre ovog žanra često imaju veću "igrivost" ili duži vijek trajanja, što dovodi do manje učestalih, ali značajnijih kupovina.

Ove razlike u trendovima naglašavaju da proizvodnja i prodaja videoigara nisu isključivo igra brojeva. Faktori kao što su karakteristike žanra, potrošačke preferencije, kvaliteta igara i zasićenje tržišta svi igraju važnu ulogu. Stoga, iako obim proizvodnje unutar nekog žanra može dati neku naznaku o popularnosti, to ne predviđa nužno njen komercijalni uspjeh. Strateške odluke u razvoju igara i marketingu moraju uzeti u obzir ove složene dinamike kako bi se maksimizirao prodajni potencijal.



Slika 6 Odnos broja igara po žanru (lijevo), odnos ukupne globalne prodaje po žanru (desno)

Analiza prosječne prodaje žanrova u zavisnosti od regije (Slika 7) pruža uvide u geografske preference. Žanr platformskih igara izaziva značajno interesovanje širom svijeta, nadmašujući druge žanrove po prosječnoj prodaji. Posebno, popularnost ovog žanra u svim regijama potvrđuje njegovu vodeću poziciju na globalnom tržištu videoigara.

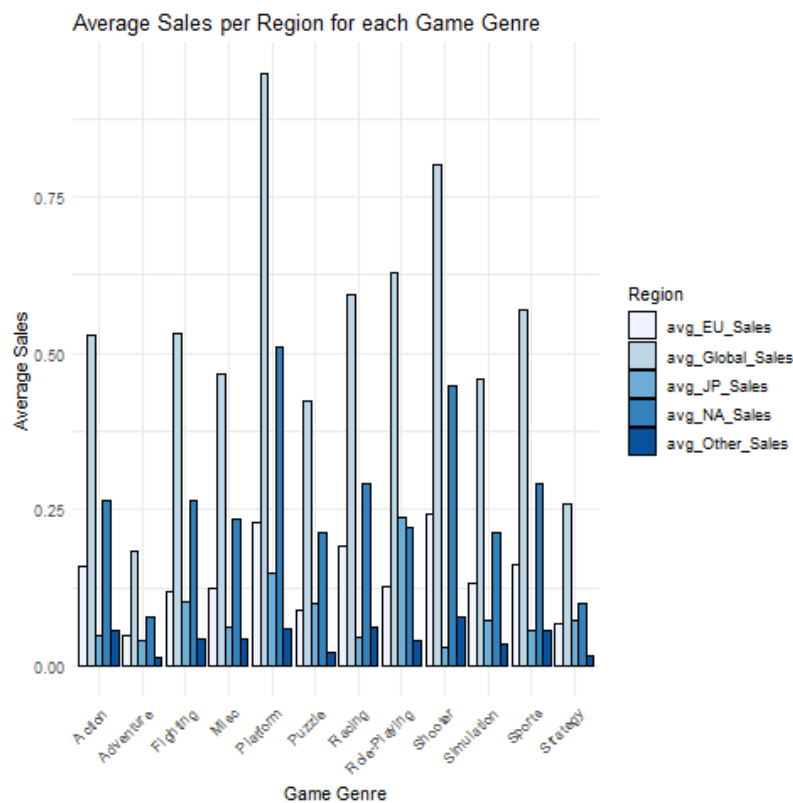
S druge strane, žanr Role-Playing se ističe na Japanskom tržištu, gdje nadmašuje druge žanrove u popularnosti. Ova jedinstvena preferenca može ukazivati na kulturnu bliskost regiona prema uronjenim igračkim iskustvima s naglaskom na priču.

Žanrovi poput akcije i borbenih igara zadržavaju široku privlačnost na različitim područjima. Akcijske igre osvajaju značajan tržišni udio u Sjevernoj Americi i Evropi, dok borbene igre pokazuju prilično ravnomjernu geografsku distribuciju popularnosti.

Pucačine, iako popularne širom svijeta, duguju svoj uspjeh pretežno tržištima Sjeverne Amerike i Evrope. Njihova znatno manja prodaja u Japanu i drugim regionima ukazuje na regionalno koncentrisanu popularnost. Strateške igre, uprkos nižoj globalnoj prodaji, pokazuju skromnu popularnost u Sjevernoj Americi, gdje dominiraju prodajom u odnosu na druge regione.

Zaključno, regionalne preference se značajno razlikuju u pogledu različitih žanrova videoigara. Dok žanrovi poput platformskih, akcionih i borbenih igara uživaju široku globalnu privlačnost, drugi poput Role-Playing i pucačkih igara manifestiraju više lokalizovanu popularnost u regijama poput Japana,

Sjeverne Amerike i Evrope.

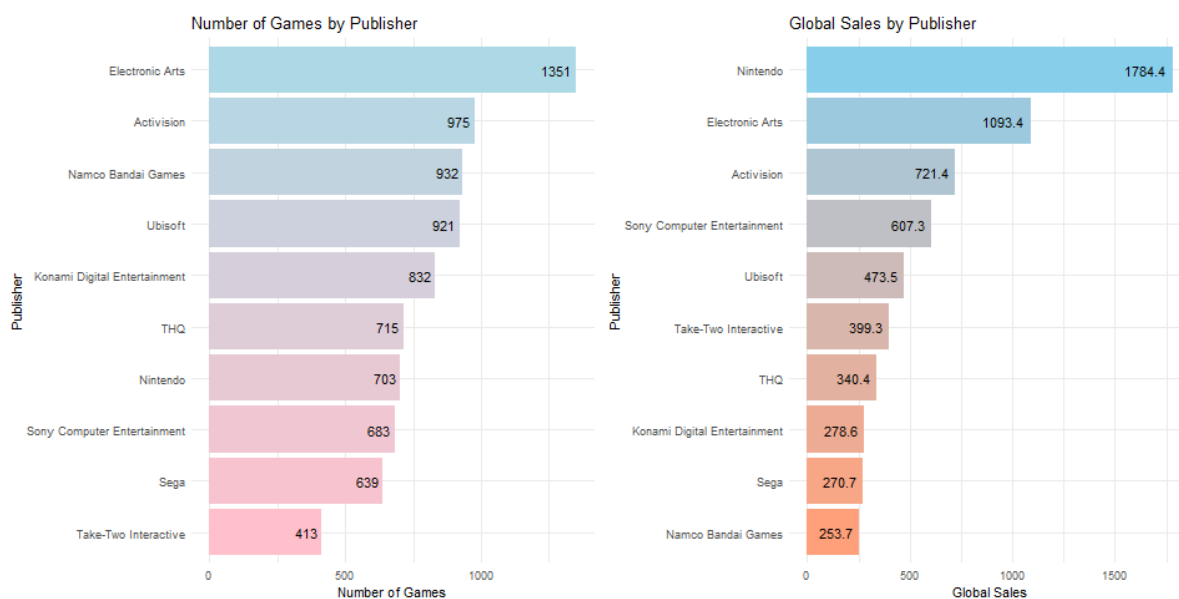


Slika 7 Prosječna prodaja žanrova po regiji

2.3. Analiza izdavača

Kroz analizu izdavača je napravljen fokus na 10 vodećih izdavača po broju proizvedenih igara kao i prosječnoj prodaji (Slika 8). Zaključci koji se mogu izvući iz ovih podataka:

- **Electronic Arts je vodeći izdavač po broju proizvedenih igara:** Electronic Arts je objavio znatno više igara od bilo kojeg drugog izdavača na listi, ukupno 1351 igra. Ovo ukazuje na to da Electronic Arts ima veliki kapacitet za proizvodnju i raznovrstan portfolij igara.
- **Nintendo prednjači po globalnoj prodaji:** Iako je sedmi po broju objavljenih igara (703 igre), Nintendo predvodi ljestvicu po ukupnoj globalnoj prodaji. Ovo implicira da su Nintendoove igre općenito komercijalno uspješnije u usporedbi s drugim izdavačima. Drugim riječima, Nintendo ima veći prosječni broj prodanih primjeraka po igri.
- **Veći broj igara ne znači nužno veću prodaju:** Kao što je već prethodno navedeno, Electronic Arts ima najveći broj igara, ali je drugi po prodaji. S druge strane, Take-Two Interactive je deseti po broju igara, ali šesti po prodaji, što implicira da komercijalni uspjeh izdavača ne ovisi samo o količini proizvedenih igara, već i o faktorima poput kvalitete igara, marketinških strategija, ugleda marke itd.
- **Neki izdavači se fokusiraju na kvalitet umjesto na kvantitet:** Pojedini izdavači koji imaju relativno manje igara (Take-Two Interactive – deseti po broju igara) ostvaruju značajnu prodaju (Take-Two Interactive – šesti po globalnoj prodaji), što sugerira da su njihove igre uspješnije u prosjeku.
- **Ubisoft i Activision imaju uravnotežen pristup:** Obje ove izdavačke kuće imaju visok broj igara i visoku globalnu prodaju, što ukazuje na uravnotežen pristup između kvantiteta i kvaliteta.



Slika 8 Odnos broja igara po izdavaču (lijevo), odnos ukupne globalne prodaje po izdavaču (desno) za top 10 izdavača

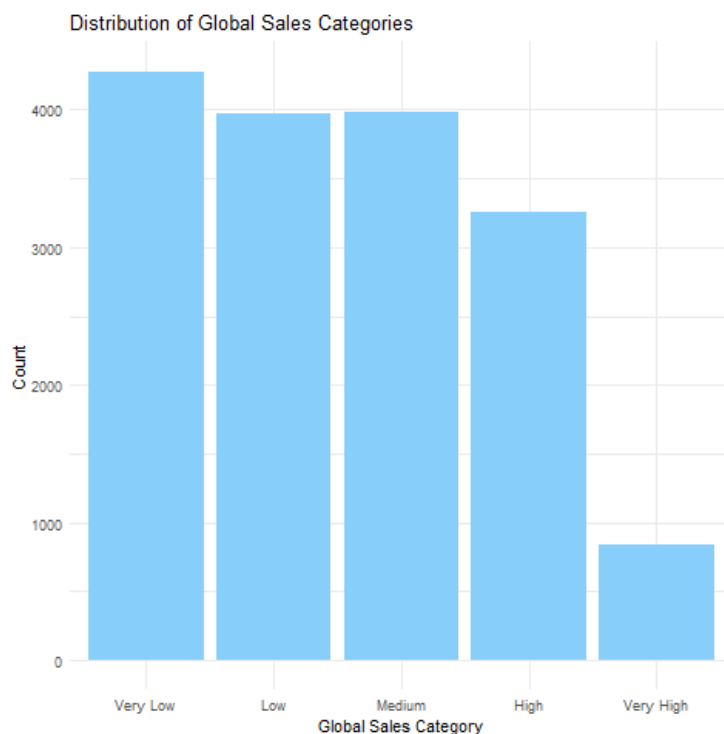
3. Klasterizacija

Prije početka procesa klasterizacije obavljena je segmentaciju globalne prodaje u različite kategorije koje se kreću od "Veoma niske" do "Veoma visoke". Ovo se postiže diskretizacijom podataka o globalnim prodajama u kategorije definisane određenim pragovima (Slika 10). Ova kategorizacija olakšava jasnije razumijevanje raspodjele prodaje u skupu podataka. Raspodjela se vizualizira kroz histogram koji pruža trenutni prikaz broja igara unutar svake kategorije prodaje (Slika 9). Segmentacija globalne prodaje na različite kategorije može biti izuzetno korisna iz nekoliko razloga:

- **Pojednostavljena analiza:** Kontinuirani podaci, posebno kada obuhvataju širok raspon, mogu biti teški za analizu i tumačenje. Segmentacijom tih podataka u kategorije, stvara se jednostavna diskretna varijabla koja se može lakše razumjeti.
- **Identifikacija trendova:** Segmentacija podataka o prodaji olakšava identifikaciju trendova i uzoraka. Na primjer, moguće je utvrditi koje su kategorije prodaje najčešće i identifikovati karakteristike igara koje pripadaju tim kategorijama.
- **Vizualizacija podataka:** Kategorizovani podaci se lakše vizualiziraju i tumače u grafičkom formatu. Pri radu s velikim skupom podataka, kategorizovani podaci se mogu predstaviti na jednostavan način, što olakšava komunikaciju rezultata.
- **Alokacija resursa:** Identifikacijom kategorija koje generišu veću prodaju, resursi poput marketinškog budžeta, razvojnog napora itd. mogu se efikasnije rasporediti.
- **Strateško planiranje:** Razumijevanjem kojim segmentima ide dobro, donosioci odluka mogu bolje definirati ciljna tržišta i razviti proizvode prilagođene tim kategorijama. Također je moguće identifikovati slabije performirajuće kategorije i razviti strategije za njihovo poboljšanje.
- **Prediktivno modeliranje:** Kategorijske varijable često poboljšavaju performanse modela mašinskog učenja. Diskretizacija kontinuiranih varijabli može pomoći modelima prilikom opažanja važnih obrazaca u podacima, posebno kada je veza između ulaznih i izlaznih varijabli

nelinearna.

Sveukupno gledano, segmentacija globalne prodaje u kategorije pomaže pretvoriti kompleksne podatke u korisne uvide koje se mogu primijeniti u praksi.



Very Low	Low	Medium	High	Very High
4277	3973	3985	3256	836

Slika 9 Graf (gore) i tabela (dole) distribucije globalne kategorije prodaje

	Category	Min	Max
1	Very Low	-Inf	0.06
2	Low	0.06	0.17
3	Medium	0.17	0.47
4	High	0.47	2.00
5	Very High	2.00	Inf

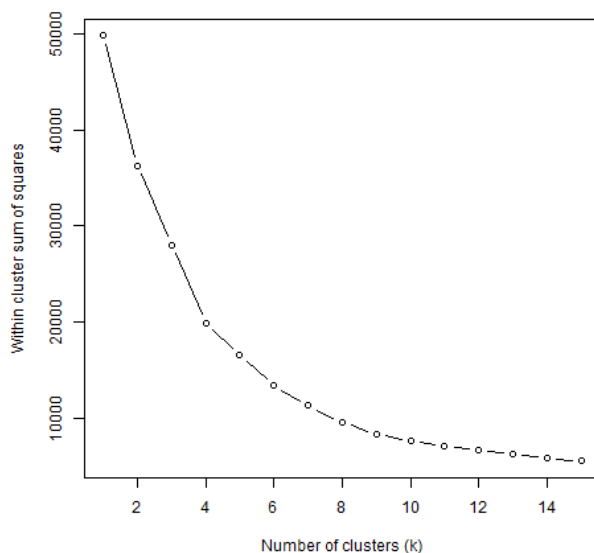
Slika 10 Prikaz raspona prodaje za svaku od kategorija globalne prodaje

Analizom distribucije globalne prodaje video igara moguće je primijetiti da je broj igara koje ostvaruju visoku prodaju manji u odnosu na broj igara sa niskom ili prosječnom prodajom. To ukazuje na relativnu rijetkost uspješnih igara u odnosu na manje uspješne. Posebno je zanimljivo uočiti da je broj igara sa izuzetno visokom prodajom naglo smanjen, što sugerira da su izuzetno uspješne igre iznimno rijetke.

3.1. K-Means klasterizacija korištenjem metode lakta

Prije početka klasterizacije izdvojene su kolone žanr (Genre), platforma (Platform) i globalna prodaja (Global_Sales), pri čemu se kategoričke varijable žanr i platforma konvertuju u numerički format. Podaci se zatim skaliraju radi efikasnog izračunavanja, a eventualne nedostajuće vrijednosti se izostavljaju.

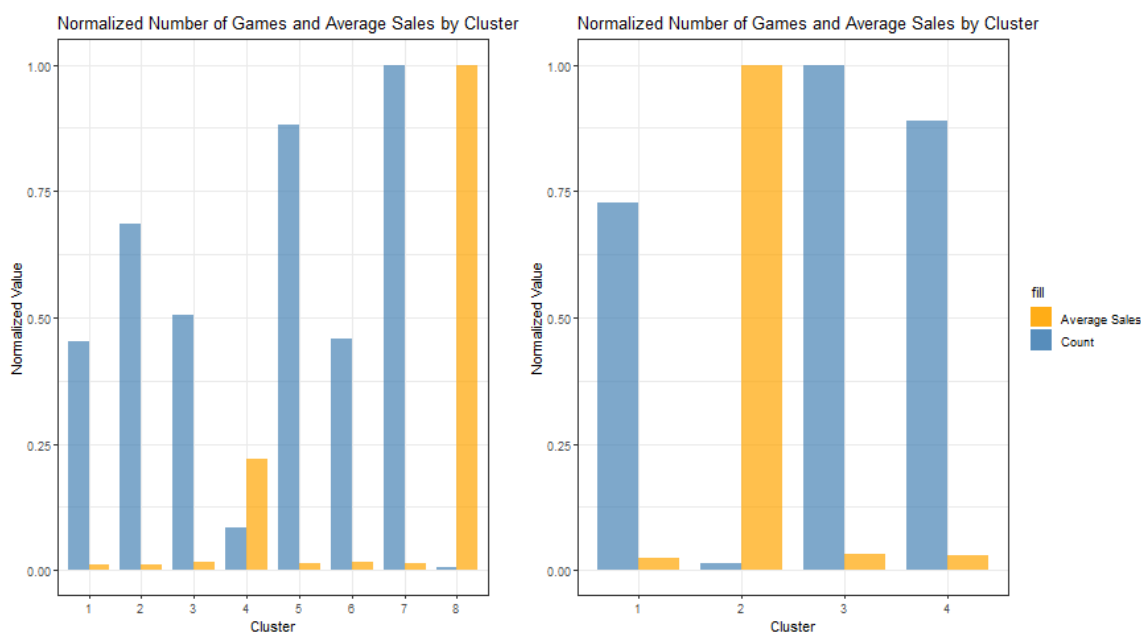
Nakon izdvajanje željenih kolona određuje se optimalan broj klastera za K-sredina algoritam. Ovaj algoritam ima za cilj podijeliti podatke u 'k' klastera, pri čemu svaka tačka podataka pripada klasteru čija je srednja vrijednost najbliža. Za različite vrijednosti 'k' (od 1 do 15) se računa suma kvadrata unutar klastera (*Within Cluster Sum of Squares - WSS*), a za odabir optimalnog broja klastera koristi se metoda lakta (*Elbow Method*). Tačka lakta predstavlja tačku u kojoj dodavanje više klastera neznatno poboljšava objašnjenost varijanse unutar klastera, što dovodi do izbora 8 klastera za ovaj skup podataka (*Slika 11*).



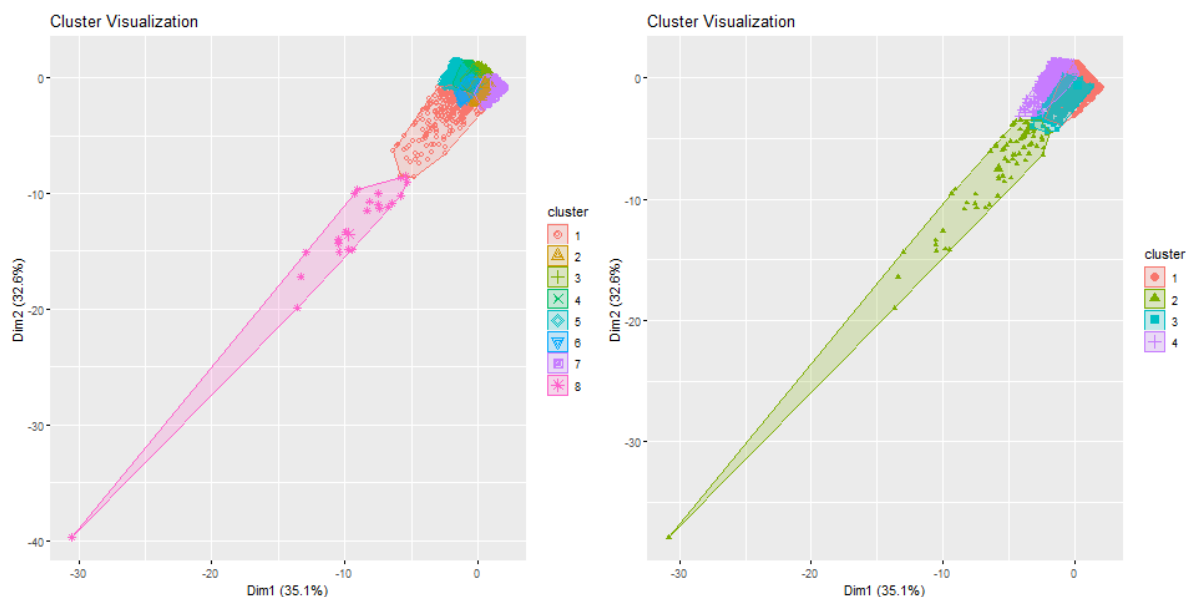
Slika 11 Prikaz grafa sume kvadrata unutar klastera u odnosu na broj klastera

Nakon određivanja broja klastera, nad podacima se pokreće algoritam k-sredina klasterizacije sa postavljenim 'k' na 8, a dobijene dodjele klastera se dodaju nazad u originalni skup podataka.

Konačno, kôd generiše sažetak karakteristika svakog klastera – broj igara u svakom klasteru i prosječnu globalnu prodaju (*Slika 12*). Ovi klasteri predstavljaju grupisanja igara koje su slične po žanru, platformi i globalnoj prodaji. Igre unutar istog klastera imaju više zajedničkih karakteristika međusobno nego sa igrama u drugim klasterima.



Slika 12 Odnos broja igara i prosječne prodaje po klasteru za 8 (lijevo) i 4 (desno) klastera



Slika 13 Vizualizacija klastera za klasterizaciju sa 8 (lijevo) i 4 (desno) klastera

U tabeli (Slika 14) su prikazane karakteristike svakog klastera:

- **cluster:** 8 klastera označenih brojevima od 1 do 8.
- **count:** Broj igara u svakom klasteru. Na primjer, klaster 1 sadrži 1842 igre, dok klaster 8 sadrži samo 23 igre.
- **avg_global_sales:** Prosječna globalna prodaja (u milionima) igara u svakom klasteru. Prosječna globalna prodaja igara u klasteru 1 iznosi 0,326 miliona jedinica, dok je u klasteru 8 značajno veća i iznosi 28,0 miliona jedinica.

Na temelju ovog sažetka, moguće je izvući zanimljive zaključke. Klasteri 4 i 8 imaju znatno veću prosječnu prodaju u usporedbi s drugim klasterima, što sugerira da ovi klasteri možda predstavljaju vrlo uspješne igre. Međutim, također imaju najmanje igara (339 i 23 igre, redom), što ukazuje na to da su takve uspješne igre relativno rijetke. Klasteri 2, 5 i 7 imaju najviše igara, ali im je prosječna prodaja bliža nižem kraju, što sugerira da ovi klasteri možda predstavljaju igre s prosječnom prodajom.

cluster	count	avg_global_sales
<int>	<int>	<dbl>
1	1842	0.326
2	2794	0.334
3	2059	0.443
4	339	6.16
5	3597	0.356
6	1866	0.423
7	4078	0.411
8	23	28.0

cluster	count	avg_global_sales
<int>	<int>	<dbl>
1	4514	0.384
2	77	16.6
3	6212	0.516
4	5524	0.471

Slika 14 Sažetak karakteristika svakog od klastera za klasterizaciju sa 8 (gore) i 4 (dole) klastera

Obzirom da su u procesu istraživanja trendova na tržištu video igara bitne igre koje se ističu među ostalima, moguće je napraviti klasterizaciju korištenjem manjeg broja klastera. U ovom slučaju, uzet je $k=4$ obzirom da ova vrijednost predstavlja tačku u kojoj kriva prestaje da naglo opada.

Na temelju rezultata klasterizacije k -sredina sa 4 i 8 klastera, mogu se uočiti sljedeća zapažanja:

- **Veličina klastera:** Korištenje 4 klastera, rezultira relativno velikim klasterima, s izuzetkom jednog manjeg klastera koji sadrži 77 igara. Međutim, primjenom 8 klastera, raspodjela igara po klasterima je uravnoteženija, s primjetno manjim klasterom od 23 igre i nešto manjim klasterom od 339 igara. Ovo ukazuje na to da rješenje sa 8 klastera pruža detaljniju grupiranost igara na temelju njihovih karakteristika.
- **Varijacija prodaje:** Analizirajući prosječne globalne prodaje za svaki klaster, primjećuje se da sa 4 klastera većina igara s prosječnom globalnom prodajom manjom od 1 grupisana je u velikim klasterima, dok su igre s izrazito višim prodajama grupisane u manji klaster. Međutim, sa 8 klastera, igre su jasnije segmentirane na temelju njihovih prodaja. Klasteri 4 i 8 se izdvajaju po većim prosječnim globalnim prodajama, što ukazuje na bolje razdvajanje igara s izuzetnim prodajnim rezultatima.
- **Izuzeci u prodaji:** Model s 8 klastera vrši razdvajanje igara s izrazito visokim globalnim prodajama (prosječna prodaja od 6.16 i 28.0) u odvojene klastera (klasteri 4 i 8). Ova separacija nije tako izražena u modelu s 4 klastera, gdje postoji jedan klaster (klaster 2) s prosječnom prodajom od 16.6 koji je relativno manji u usporedbi s drugim klasterima, što implicira da su iznimno uspješne igre smještene u ovaj odvojeni klaster prilikom podjele na 4 klastera.

Posmatrajući reprezentaciju klastera (*Slika 13*) potrebno je uzeti u obzir:

- **Tumačljivost:** Prekomjeren broj klastera može rezultirati modelom koji je teško interpretirati. Kada se klasteri značajno preklapaju, postaje izazovno identificirati jedinstvene karakteristike koje svaki klaster predstavlja. S druge strane, nedovoljan broj klastera može zanemariti suptilne razlike u podacima, što rezultira pojednostavljenim prikazom. U slučaju korištenja 8 klastera, primjećuje se prisutnost dva izdvojena klastera, dok ostali klasteri pokazuju izrazito slične karakteristike u pogledu prosječne globalne prodaje. Ovo se podudara s prethodnom analizom (prikazanom na Slici), koja je otkrila da je broj igara s vrlo niskim do prosječnim prodajnim rezultatima iznimno sličan.
- **Veličina klastera:** Evaluacija broja tačaka (count) u svakom klasteru može biti dodatno korisna. U kontekstu modela od 8 klastera, klaster 8 pokazuje prisustvo svega 23 tačke, dok u modelu od 4 klastera, Klaster 2 uključuje 77 tačaka. Ove brojke značajno se razlikuju od ostalih klastera i također se odlikuju značajnom udaljenošću od ostatka klastera. Specifično, ti klasteri predstavljaju izuzetke u vidu najuspješnijih igara.
- **Homogenost i separacija:** U idealnim uvjetima, teži se homogenosti klastera (odnosno, sličnosti instanci unutar istog klastera) i dobroj separaciji (odnosno, različitosti instanci između različitih klastera). Analiza klastera dobivenih primjenom K -sredina algoritma s 4 i 8 klastera ukazuje na odsustvo preklapanja unutar skupine igara koje ostvaruju najvišu prosječnu prodaju, ističući njihovu razliku u odnosu na ostale igre.
- **Značajnost:** Relevantnost: Svaki klaster bi trebao suštinski predstavljati nešto značajno. U ovom primjeru, klaster označava grupu igara sličnih prodajnih rezultata. Ukoliko se klasteri previše približe jedan drugome u pogledu prosječnih globalnih prodaja, to sugerira nedostatak raznolikosti među klasterima. Kada se analizira velik broj klastera na temelju dostupnih podataka, može se zaključiti da postoji obilje igara koje postižu sličan uspjeh, dok se manji broj

igara izdvaja po iznimno visokim rezultatima.

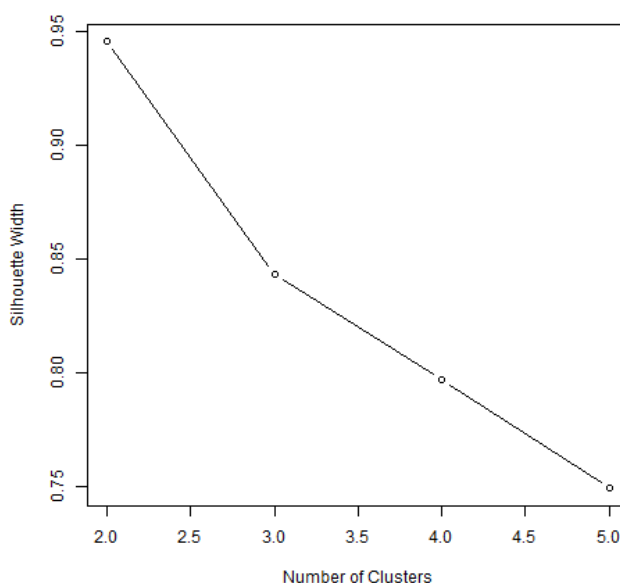
Kroz detaljniju analizu igara unutar ovih klastera, moguće je identificirati zajedničke karakteristike koje bi mogle objasniti njihovu prodaju. Navedene karakteristike obuhvataju specifičan žanr igre ili platformu za koji je igra namijenjena, što ima potencijalni utjecaj na popularnost tih igara. U nastavku će biti prikazana primjena K-sredina klasterizacije koja će pružiti daljnji uvid u specifičnosti igara unutar svakog klastera.

3.2. K-Means klasterizacija korištenjem metode koeficijenta siluete

U cilju analize klasterizacije korištenjem K-sredina algoritma o prodaji igara grupisanih na osnovu njihovih performansi u različitim regionima izdvojene su četiri kolone koje sadrže podatke o prodaji u četiri regiona: Sjeverna Amerika (NA_Sales), Evropa (EU_Sales), Japan (JP_Sales) i ostali region (Other_Sales).

Optimalan broj klastera se određuje primjenom metode siluete, koja predstavlja alat za procjenu kvalitete klasterizacije. Putem petlje se iterira kroz raspon potencijalnih veličina klastera (od 2 do 5) i računa prosječna širina siluete za svaki od njih. Širina siluete mjeri bliskost svake tačke unutar klastera sa tačkama u susjednim klasterima. Ova mjera se kreće u rasponu od -1 do 1, pri čemu visoka vrijednost ukazuje na to da je objekat dobro povezan sa svojim klasterom, ali loše povezan sa susjednim klasterima.

Nakon izračuna širine siluete za svaki potencijalni broj klastera, kreiran je grafik (*Slika 15*) kako bi se vizualizirale proračunate vrijednosti. Ovaj grafik omogućava identifikaciju broja klastera koji maksimizira prosječnu širinu siluete, što predstavlja optimalan balans između kohezije (sličnosti elemenata unutar klastera) i separacije (različitosti među različitim klasterima). Na temelju maksimalne širine siluete (2), utvrđen je optimalan broj klastera. S tim brojem klastera, provedena je ponovna klasterizacija K-sredina na podacima o prodaji.



Slika 15 Grafik vizualizacije prosječne širine siluete za svaki od klastera

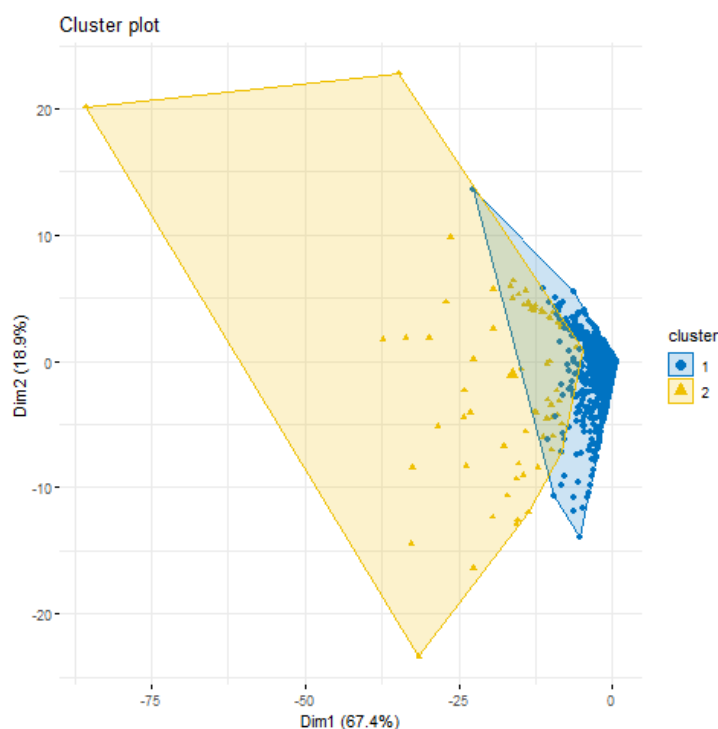
Klasterizacijom igara na osnovu njihovog prodajnog uspjeha u različitim regijama, moguće je tumačiti rezultate sažetka širine siluete na sljedeći način:

Klaster 1 (veličina: 16.521, prosječna širina siluete: 0,95): Ovaj klaster obuhvata veliki broj igara, a visoka prosječna širina siluete od 0,95 ukazuje na dobru separaciju i pravilno dodjeljivanje podataka

ovom klasteru. Ovi rezultati sugeriraju da igre unutar ovog klastera dijele sličan prodajni uspjeh u različitim regijama, što ih razlikuje od igara u drugim klasterima. Visoka prosječna širina siluete ukazuje na izražen obrazac ili preferencu za prodajni uspjeh u regijama za igre unutar ovog klastera.

Klaster 2 (veličina – 77, prosječna širina siluete – 0,19): Ovaj klaster obuhvata manji broj igara u usporedbi s klasterom 1. Prosječna širina siluete od 0,19 ukazuje na manju separaciju podataka unutar ovog klastera u usporedbi s klasterom 1. To ukazuje na veću varijabilnost ili preklapanje prodajnih rezultata igara u ovom klasteru u različitim regijama. Niža prosječna širina siluete ukazuje da klaster možda nema tako jasan obrazac u regionalnim preferencama prodaje u usporedbi s klasterom 1.

Na temelju širine siluete, klaster 1 predstavlja grupu igara s izraženim i jasnim obrascem prodajnog uspjeha u različitim regijama, dok klaster 2 predstavlja manju grupu igara s manje izraženim regionalnim obrascima prodaje.



Slika 16 Vizualizacija klastera dobivenih korištenjem metode širine siluete

Daljnje istraživanje i analiza karakteristika klastera je od velike važnosti. Ovo uključuje pregledanje specifičnih igara unutar svakog klastera i upoređivanje njihovog prodajnog uspjeha u različitim regijama. Ovakva analiza omogućava dublje uvide u obrasce i preference unutar svakog klastera te pomaže u razumijevanju rezultata klasteriranja u kontekstu ove analize.

3.2.1. Uvid u karakteristike klastera kroz sažete statistike

Nadovezujući se na prethodno provedenu klasterizaciju K-sredina podataka o prodaji igara koristeći širinu siluete, provedena je analiza karakteristika sadržaja klastera. U početnom koraku, dodijeljeni klasteri dobiveni iz K-sredina algoritma se dodaju kao nova varijabla u skupu podataka, što omogućava detaljan pregled sastavnih dijelova podataka unutar svakog pojedinog klastera.

Primenom funkcije "by" u programskom jeziku R izračunate su sažete statistike za svaki klaster. Ova funkcija, koja se može fleksibilno koristiti, primenjuje funkciju "summary" na različite nivoe faktora koji predstavljaju dodijeljene klasterne. Funkcija "summary" pruža statistički pregled za svaku promenljivu, uključujući minimum, maksimum, medianu, srednju vrednost, prvu i treću kvartil. Ovaj

sveobuhvatan prikaz olakšava duboko razumevanje distribucije prodaje igara u svakom klasteru i pomaže u identifikaciji karakteristika koje čine svaki klaster posebnim.

Korištenjem funkcije "*table*" za tabulaciju veličine svakog klastera, se generira frekvencijska distribucija klastera. Ove informacije otkrivaju broj igara dodijeljenih svakom klasteru i pružaju uvid u distribuciju podataka u klasterima.

Centroidi svakog klastera, koji su dostupni putem elementa "*centers*" modela K-sredina, su izračunati kao srednja vrijednost svih tačaka unutar odgovarajućeg klastera za svaku varijablu. Oni pružaju naznaku karakterističnih atributa svakog klastera i predstavljaju arhetipske igre unutar svakog klastera na osnovu prodajnog uspjeha.

Na kraju je izračunata ukupna suma kvadrata unutar klastera, što je komponenta modela K-sredina. Ova metrika kvantifikuje kompaktnost klastera, pri čemu niže vrijednosti ukazuju na bolju klasterizaciju, što implicira da su tačke unutar odgovarajućih klastera međusobno bliže.

Na osnovu sažetih statističkih podataka, moguće je izvući sljedeće zaključke:

- Klaster 1 sadrži većinu igara (16.521), dok se samo 77 igara nalazi u Klasteru 2.
- Klaster 2 predstavlja igre s visokim prodajnim rezultatima u svim regijama. Prosječna prodaja u Sjevernoj Americi (NA_Sales) iznosi oko 8,29 miliona jedinica, u Europi (EU_Sales) oko 4,66 miliona jedinica, u Japanu (JP_Sales) oko 2,18 miliona jedinica i u ostalim regijama (Other_Sales) oko 1,40 miliona jedinica.
- Klaster 1 predstavlja igre s nižim prodajnim rezultatima. Prosječna prodaja u klasteru 1 je znatno niža u usporedbi s klasterom 2 u svim regijama. Prosječna prodaja u Sjevernoj Americi iznosi oko 0,23 miliona jedinica, u Europi oko 0,13 miliona jedinica, u Japanu oko 0,07 miliona jedinica, dok je prodaja u ostalim regijama oko 0,04 miliona jedinica.
- Ukupna suma kvadrata unutar klastera iznosi 10.442,81, što predstavlja mjeru kompaktnosti klastera. Bez usporedbe s drugim modelima ili skupovima podataka, teško je odrediti je li ova vrijednost visoka ili niska.
- Igre u klasteru 2 mogu se smatrati "hitovima" ili "blockbusterima" zbog znatno većih prodajnih rezultata u usporedbi s većinom igara u klasteru 1. Ove igre mogu biti naslovi poznatih franšiza ili izdavača. Igre u klasteru 1 vjerojatno predstavljaju širi raspon performansi, od prosječnih izvođača do slabije uspješnih igara.
- Ovi zaključci su grubo interpretirani na temelju sumarnih statističkih podataka. Detaljnije istraživanje igara unutar svakog klastera (njihov žanr, izdavač, itd.) može pružiti suptilnije uvide.

3.2.2. Proučavanje sažetih statističkih podataka za kategoričke varijable unutar klastera

Proučavanje sažetih statističkih podataka ili frekvencijskih brojanja za svaku kategoričku varijablu unutar svakog klastera ima izuzetnu vrijednost u razumijevanju popularnosti i trendova u video igrama na osnovu različitih karakteristika kao što su žanr, platforma i izdavač, posebno kada se kombinuje sa podacima o prodaji u različitim regijama.

Analiza klastera pomaže u kategorizaciji igara na osnovu njihovih prodajnih rezultata u različitim regijama. Proučavanjem broja igara u svakom klasteru i karakteristika klastera, može se utvrditi koje grupe igara ostvaruju veću prodaju. Igre u klasterima sa većom prodajom mogu se smatrati popularnijima.

Frekvencijska brojanja za žanr, platformu i izdavača unutar svakog klastera pružaju dodatni detaljniji

uvid (Slika 17). Na primjer, ako je određeni žanr dominantan u klasteru sa visokom prodajom, to može ukazivati da je taj žanr trenutno popularan među igračima. Slično, ako određena platforma dominira u klasteru sa snažnom prodajom, to ukazuje da su igre namijenjene za tu platformu u trendu.

Također, izdavač sa mnogo igara u klasterima sa visokom prodajom vjerovatno proizvodi igre koje se uklapaju u trenutne trendove tržišta. Ova vrsta analize također pomaže u identifikaciji promjena trendova tokom vremena.

Genre	Cluster		Platform	Cluster		Publisher	Cluster	
	1	2		1	2		1	2
Action	3244	9	3DS	495	5	Activision	954	12
Adventure	1275	1	DS	2122	11	Atari	346	1
Fighting	835	1	GBA	809	2	Capcom	376	0
Misc	1703	7	GC	542	0	Electronic Arts	1339	0
Platform	862	14	N64	313	3	Konami Digital Entertainment	823	0
Puzzle	568	3	PC	943	0	Namco Bandai Games	928	0
Racing	1218	8	PS	1187	2	Nintendo	653	43
Role-Playing	1461	10	PS2	2123	4	Sega	632	0
Shooter	1264	18	PS3	1296	8	Sony Computer Entertainment	678	4
Simulation	849	2	PS4	334	2	Square Enix	231	0
Sports	2300	4	PSP	1197	0	Take-Two Interactive	404	8
Strategy	671	0	PSV	412	0	Tecmo Koei	338	0
			Wii	1279	11	THQ	712	0
			X360	1221	14	Ubisoft	916	2
			XB	802	1	Warner Bros. Interactive Entertainment	217	0

Slika 17 Sažeti statistički podaci za kategoričke varijable žanr (lijevo), platformu (sredina) i izdavača (desno)

Za svaku od ovih analiza, kreira se frekvencijska tablica koja prikazuje koliko igara u svakom klasteru pripada određenoj kategoriji varijable. Ovo može pružiti uvid u vrste igara koje su prepoznatljive u svakom klasteru. Međutim, treba imati na umu da to ne znači nužno da je varijabla uzrokovala da igra pripada određenom klasteru - ovo su korelacije, a ne uzročni odnosi. Ako je broj platformi ili izdavača velik, frekvencijske tablice mogu biti teške za interpretaciju. U nastavku je razmotrena identifikacija žanrova, kao i 15 najboljih izdavača i platformi, nakon čega je izvršeno ograničavanje frekvencijskih tablica na ove kategorije (Slika 17).

Prvo se računaju frekvencijska brojanja za sve platforme i izdavače, a zatim identifikuje najboljih 15. Zatim ponovno računaju frekvencijska brojanja, ali samo za igre u kojima je platforma ili izdavač među najboljih 15, što pruža usmjereniji pregled raspodjele igara unutar svakog klastera za najčešće platforme i izdavače.

Tumačenje sažetih statističkih podataka:

- **Žanr:** Određeni žanrovi, kao što su platformeri i pucačine, su česti u klasteru 2. Većina žanrova pretežno se javlja u klasteru 1, što ukazuje da ovaj klaster sadrži više mainstream ili uobičajenih žanrova igara.
- **Top 15 platformi:** Postoje jasne razlike između klastera kada se radi o platformama. Na primjer, klaster 2 ima veći udio igara za platforme kao što su Wii, X360 i DS, koje su među najpopularnijim platformama. Mnoge platforme kao što su PC, GC, PSP i PSV imaju sve svoje igre smještene u klasteru 1.

- **Top 15 izdavača:** Kada se analiziraju izdavači, primjećuje se da je Nintendo dominantan izdavač igara u klasteru 2, što može ukazivati na specifičnu marketinšku strategiju ili određenu vrstu igara koja ih izdvaja od ostalih izdavača. Većina vodećih izdavača ima većinu svojih igara smještenih u klasteru .

Općenito, klaster 2 se može smatrati posebnom tržišnom nišom - s određenim žanrovima, platformama i izdavačima koji su prisutniji u odnosu na klaster 1. To može ukazivati na specijalizirane, jedinstvene ili izuzetno uspješne igre. S druge strane, klaster 1, kao veći klaster, obuhvaća širok spektar uobičajenih igara, koje uključuju različite žanrove, platforme i izdavače. Može predstavljati veću raznolikost prosječnih ili tipičnih igara u industriji.

4. Poređenje rezultata klasterizacije sa Google Trends podacima

Kako bi se postigla jasnija analiza trendova u video igrama, rezultati klasterizacije su upoređeni sa podacima iz Google Trends-a. Google Trends je alat koji omogućava razumijevanje interesa za određene pretraživačke pojmove tokom vremena, a ti pojmovi u ovom slučaju predstavljaju ključne riječi vezane za video igre. Ova usporedba se temelji na prirodi Google Trends-a kao odraza javnog interesa i ponašanja. Stoga, kombinovanje ovoga sa rezultatima klasterizacije k-sredina može pružiti jedinstvene uvide u to kako identificirani klasteri odgovaraju stvarnim trendovima i interesima tokom vremena. Također može otkriti kako su određene platforme za igre ili žanrovi stekli ili izgubili popularnost, te pružiti uvide o promjenama u trendovima u igračkoj industriji tokom godina.

Zanimljivo je primijetiti da se vremenski periodi obuhvaćeni korištenim skupom podataka o igrima i podacima iz Google Trends-a preklapaju. Dok skup podataka pruža pregled industrije od 1980. do 2016. godine, Google Trends omogućava proširenje ove analize do trenutnog vremena. Ovo proširenje omogućava vizualizaciju razvoja klastera identificiranih na osnovu skupa podataka u periodu od 2004. do 2016. godine, kao i njihov daljnji razvoj u godinama nakon 2016. godine, što pruža sveobuhvatnije razumijevanje trajektorije industrije video igara.

4.1. Analiza prosječne prodaje top 5 žanrova sa Google Trends pretraživanjima

U početnoj fazi analize vrši se priprema skupa podataka o igrima. Podaci su odabrani tako da se fokusira na godine 2004. i kasnije, kako bi se uskladili s podacima iz Google Trends-a. Nakon toga se koristi paket `dplyr` za izvršavanje operacija na podacima, posebno za određivanje prosječne globalne prodaje po žanru za svaku godinu.

U sljedećoj fazi analize vrši se identifikacija top 5 žanrova na osnovu ukupne prodaje. Prodaja ovih žanrova se zatim normalizuje kako bi se skalirale vrijednosti na uniformni opseg (u ovom slučaju od 0 do 100). Ovaj korak normalizacije je neophodan radi usklađivanja sa podacima iz Google Trends-a koji djeluju na skali od 0 do 100. Paralelno s tim, izvršava se ekstrakcija podataka iz Google Trends-a pomoću paketa `gtrendsR`. Ekstrakcija se fokusira na prethodno identificirane top 5 žanrova, pri čemu se bilježi metrika "*interest over time*" za svaki žanr i pohranjuje u listu. Nakon toga, datumi iz Google Trends-a se pretvaraju u godine, te se izračunava prosječni interes za pretraživanje za svaki žanr godišnje. Imena kolona podataka iz Google Trends-a se prilagođavaju datasetu prodaje kako bi se osigurala jednolikost u uporedivom datasetu.

Podaci o prodaji i podaci iz Google Trends-a su integrisani u sveobuhvatan dataframe nazvan `comparison_data`, pri čemu su podaci usklađeni po žanru i godini. Ovaj rezultirajući dataset sadrži normaliziranu prodaju i normalizirane pretrage iz Google Trends-a za svaki od top 5 žanrova, sortirane

po godinama. Ova integrirana skupina podataka pruža jasan i informativan prikaz evolucije javnog interesa za najpopularnije žanrove, kao što je prikazano putem podataka iz Google Trends-a, u skladu s njihovom prodajnom uspješnošću.

Rezultati poređenja prosječne prodaje igara (normalizirane) i interesa iz Google Trends-a (normaliziranog) od 2004. do 2016. (sa jednim unosom za 2017.) za top 5 žanrova (*Slika 18*):

Platformske igre:

Popularnost platformskih igara u smislu prodaje pokazuje primjetan vrhunac u 2009. i 2012. godini, što se ne podudara savršeno sa interesom u Google Trends-u koji doseže vrhunac 2009. godine, a zatim opada. Ova razlika može ukazivati da, uprkos opadajućem općem interesu (pretrage), žanr održava snažnu prodaju tokom nekih godina. Ovo može biti rezultat jezgre lojalnih korisnika koji redovno kupuju nove igre iz žanra, čak i ako je širi javni interes u opadanju.

Trkačke igre:

Žanr trkačkih igara pokazuje zanimljiv trend gdje podaci iz Google Trends-a ukazuju na konstantan interes tokom godina, često prelazeći 70. Međutim, podaci o prodaji se ne poklapaju s ovim visokim nivoom interesa, pri čemu je prosječna prodaja uglavnom oscilirala između 20-40 normalizirane prodaje. Moguće je da, iako postoji širi interes za žanr (kroz pretrage), to se ne odražava u visokim prodajama. Moguće je pretpostaviti da su trkačke igre dobro pokrivene kroz nekoliko dominantnih naslova, ili igrači tendiraju da se dugo zadržavaju na jednoj igri, što ograničava potencijal za visoku prodaju unutar žanra.

Role-Playing igre (RPG):

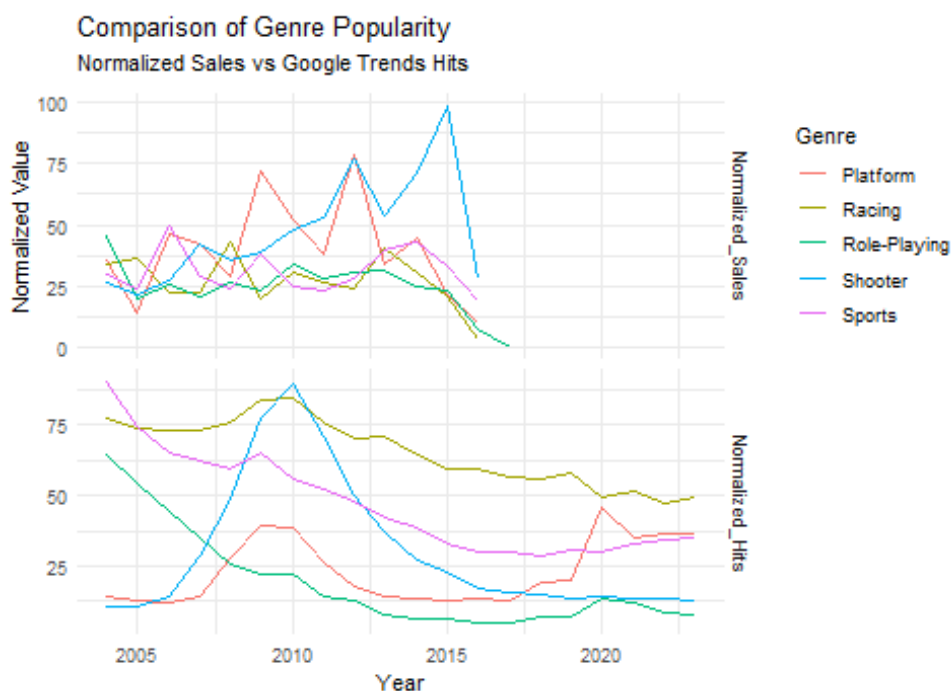
Žanr Role-Playing igara pokazuje trend pada prodaje, kao i interesa iz Google Trends-a tokom vremena, s posebno izraženim padom interesa prema Google Trends-u. Ovo sugerira da je žanr možda suočen s opadajućom popularnošću tokom ovog perioda. Činjenica da i prodaja i interes opadaju paralelno sugerira moguću povezanost između interesa iz Google Trends-a i prodaje igara za ovaj žanr.

Pucačine:

Žanr pucačkih igara pokazuje rastući trend u prodaji, sa vrhuncem u 2015. godini. Međutim, ovo se ne poklapa dobro s podacima iz Google Trends-a koji pokazuju najveći interes oko 2010. godine, a zatim opadanje. Zanimljivo je da su uprkos opadajućem trendu interesa, pucačke igre imale najveću prodaju 2015. godine. Ova razlika može biti rezultat uspješnih marketinških kampanja, objavljivanja visokoprofilnih igara ili raznih drugih faktora koji pokreću prodaju nezavisno od općeg interesa za pretragu.

Sportske igre:

Žanr Sportskih igara pokazuje relativno stabilan trend u prodaji, s blagim opadanjem tokom vremena. Međutim, podaci iz Google Trends-a ukazuju na opadanje interesa od 2004. do 2016. godine. Ovaj žanr često je dominiran godišnjim izdanjima popularnih franšiza, što može održati prodaju relativno stabilnom, uprkos opadajućem općem interesu.



Slika 18 Grafički prikaz poređenja prosječne prodaje igara (gore) i interesa iz Google Trends-a (dole) za top 5 žanrova

Ukratko, veza između podataka iz Google Trends-a i prodaje nije uvijek jasna ili direktna, što ukazuje na to da faktori koji nisu povezani s općim interesom javnosti (kao što su kvaliteta igara, marketinški napori i lojalnost brendu) mogu značajno utjecati na prodaju. Ipak, podaci iz Google Trends-a pružaju dodatnu perspektivu o popularnosti žanrova igara tokom vremena.

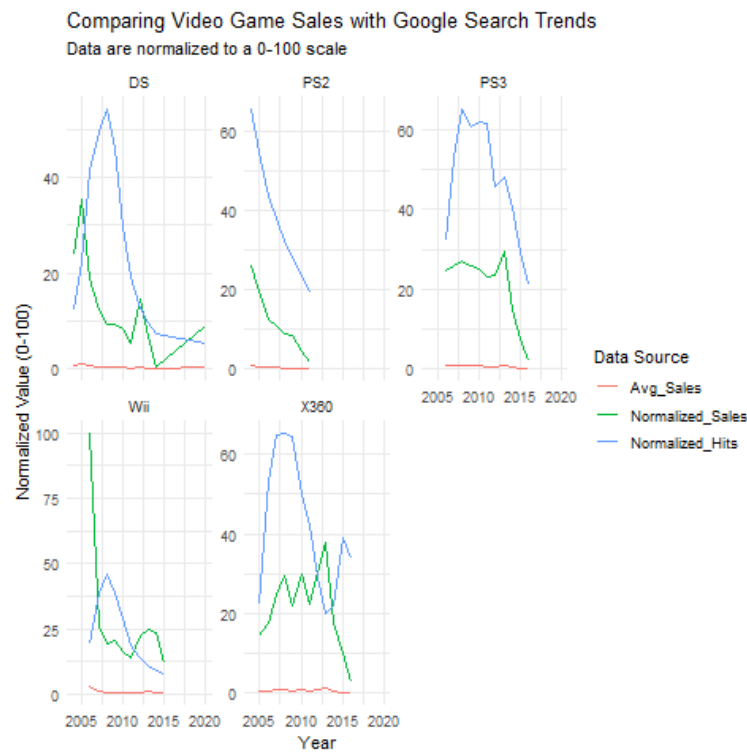
4.2. Analiza prosječne prodaje top 5 platformi sa Google Trends pretraživanjima

Analizom skupa podataka o globalnoj prodaji video igara, računa se prosječna prodaja za svaku platformu po godinama, nakon čega se dobiveni podaci normalizuju, a zatim identificirano pet najpopularnijih platformi. Kako bi se istražio odnos između prodaje i interesovanja na internetu, prikupljeni su i normalizovani Google Trends podaci za ove platforme (Slika 19).

Ovi konsolidovani podaci omogućavaju poređenje prodajnih rezultata platforme i njene popularnosti u pretragama na internetu, kako je praćeno putem Google Trends-a. Posmatranja pokazuju zanimljiva odstupanja između prodajnih podataka i podataka iz Google Trends-a, što ukazuje da popularnost pretraga na internetu nije uvijek usklađena sa prodajnim rezultatima.

- **Prodaje i trendovi:** Primjetno je da se prodaje (temeljene na globalnim podacima o prodaji) i trendovi (temeljeni na podacima iz Google Trends-a) ne podudaraju uvijek. Na primjer, za DS platformu u 2004. godini, prodaje su veće od trendova. Ovo može ukazivati na to da uspješnost prodaje platforme nije nužno povezana s njenom popularnosti u online pretraživanjima.
- **Popularnost platformi tijekom vremena:** Platforme imaju svoje vrhunce i padove tijekom godina. Uzimajući za primjer platformu DS; njen vrhunac popularnosti u pogledu prodaje i online pretraživanja se dogodio oko 2008. godine, nakon čega je počela opadati. To je prirodno jer se na tržištu pojavljuju novije platforme, dok starije postaju zastarjele.
- **Dominacija platformi:** Platforma Wii je 2006. godine imala najviše prodaje od 100. Unatoč tome, normalizirani trendovi za Wii u toj godini su relativno niski, iznosili su 19,2, što sugerira da prodaja i interes za pretraživanje ponekad ne moraju biti usklađeni.

- **Konstantne platforme:** Neke platforme poput PS3 su pokazale vrlo konstantan uspjeh tokom godina, kako u prodaji tako i u interesu za online pretraživanjima.
- **Prodaja nasuprot interesa za pretraživanje:** Neke platforme pokazuju veće prodaje u kasnijim godinama svog životnog ciklusa (kao što je X360 u 2013. godini), ali imaju manji interes za pretraživanje. Ovo može biti rezultat lojalnosti prema brendu ili snažne korisničke baze koja više ne mora pretraživati platformu jer je već informirana o njoj.



Slika 19 Grafički prikaz poređenja prosječne prodaje igara (zelena) i interesa iz Google Trends-a (plava) za top 5 platformi

Zaključak

Ovaj seminarski rad ističe odnose između različitih žanrova igara, platformi, izdavača i njihove prisutnosti na tržištu. Korištenjem metoda grupisanja i analize podataka u programskom jeziku R, identificirane su jasne razlike između različitih skupova igara. Rezultati istraživanja ukazuju na postojanje dvije glavne kategorije na tržištu industrije igara: iznimno popularne igre i ostale igre. Iz analize proizlazi da su među najpopularnijim igrama žanrovi platformera i pucačina, koji se često javljaju na platformama poput Wii, X360 i DS, a često su izdani od strane Nintendo-a, koji je istaknut kao jedan od vodećih izdavača na tržištu. S druge strane, skup ostalih igara obuhvata širok spektar mainstream ili uobičajenih igara različitih žanrova, platformi i izdavača.

Istraživanje je jasno pokazalo da Google Trends pruža uvid u dinamiku popularnosti žanrova igara tijekom vremena. Međutim, treba naglasiti da veza između podataka iz Google Trends-a i prodaje igara nije uvijek izravna, što ukazuje na to da isključivo javni interes ne određuje direktno uspjeh u prodaji. Drugi ključni faktori, kao što su kvaliteta igara, marketinški naponi i lojalnost prema brendu, imaju značajnu ulogu u vođenju prodaje. Važno je duboko razumjeti ove složene dinamike u industriji igara. Stoga se ne bi trebalo oslanjati se samo na opći javni interes, kako ga mjeri Google Trends, već bi se trebao pažljivo razmotriti izbor žanra igara, platformi i strategija izdavanja. Na taj način moguće je proizvoditi igre koje ne samo da odražavaju interese igrača, već također imaju potencijal za visoku razinu prodajnog uspjeha.

Važno je istaći da proučavanjem iste analize s podacima iz različitih perioda mogu se primijetiti promjene u strukturi klastera, žanrovima igara, platformama i izdavačima koji dominiraju u klasterima s visokom prodajom. Osim toga, ova analiza može se koristiti za predviđanje budućih trendova. Ako se određeni žanr ili platforma počinju češće pojavljivati u klasterima s visokom prodajom, to može ukazivati na novi trend. Ovakvi uvidi su od iznimne važnosti za razvojne timove i marketinške stručnjake kako bi mogli predvidjeti kretanje tržišta i strategijski planirati buduće projekte.

Reference

- [1] “R-4.3.0 for Windows.” *Download R-4.3.0 for Windows. The R-Project for Statistical Computing.*, cran.r-project.org/bin/windows/base/. (Pristupljeno 15. mart 2023.)
- [2] *Google Trends*, trends.google.com/trends/. (Pristupljeno 03. juni 2023.)
- [3] Smith, Gregory. “Video Game Sales.” *Kaggle*, 26 Oct. 2016, www.kaggle.com/datasets/regorut/videogamesales. (Pristupljeno 30. maj 2023.)