

# Sprzedaż gier na rynku tradycyjnym do 2016 r.

Jarosław Cierpich  
Arkadiusz Kasprzak

Zbiór danych:

<https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings>



# Plan prezentacji

1. Omówienie zbioru danych.
2. Wstępna obróbka zbioru danych.
3. Charakterystyka danych.
4. Elementy nietypowe dla każdej cechy osobno.
5. Klasteryzacja.
6. Klasyfikacja.
7. Podsumowanie i wnioski.



# Omówienie zbioru danych

- Zbiór dotyczy danych rynkowych gier na tradycyjne platformy (PC, konsole, handheldy)
- Dane zawarte w zbiorze dotyczą gier do 2016 roku, wszystkie dane dotyczące ilości są sumą od wydania gry, aż do sporządzenia zbioru danych
- Zbiór danych przed wstępną obróbką zawiera **16720** unikalnych wpisów
- Dla atrybutów podjętych analizie w.w. wpisy posiadają około **24000** brakujących wartości.
- **Główne atrybuty**, które zostały poddane analizie, to: oceny użytkowników, oceny krytyków, sprzedaż globalna, sprzedaż na rynku amerykańskim, sprzedaż na rynku europejskim, sprzedaż na rynku japońskim, rok wydania



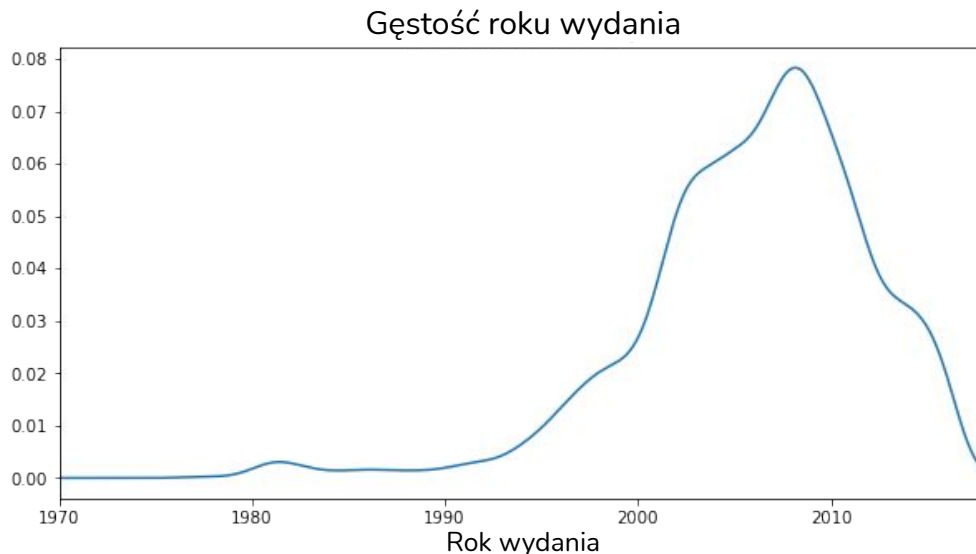
# Wstępna obróbka zbioru danych

- Atrybuty które były **ważne** z punktu widzenia analizy oraz posiadały niewiele brakujących elementów zostały uzupełnione **ręcznie** (np.: rok wydania)
- Ręcznie uzupełniono takie atrybuty jak **ocena użytkowników**, **ocena krytyków** dla gier z dużą ilością sprzedanych kopii
- Ze względu na dużą ilość brakujących informacji dla **słabo sprzedających** się gier (poniżej 200 tys. sprzedanych kopii) usunięto te, dla których brakowało danych o ocenach użytkowników lub ocenach krytyków
- Brakujące wartości dla **ocen użytkowników** oraz **ocen krytyków** uzupełniono korzystając z metody **k najbliższych sąsiadów**. W sumie uzupełniono ponad **6000** brakujących danych
- Po wszystkich w.w. zabiegach liczność zbioru zmieniła się z 16720 do 10451, natomiast parametry zbioru oraz najważniejsze informacje zostały zachowane.



# Charakterystyka danych - rok wydania

- Zbiór zawiera gry głównie z okresu 1995 - 2015. Występują gry starsze, tzw. gry ikoniczne
- Nie wszystkie gry z okresu 2015-2016 znajdują się w bazie, natomiast tendencja spadkowa co do ilości gier na platformy tradycyjne, jest zgodna z rzeczywistością

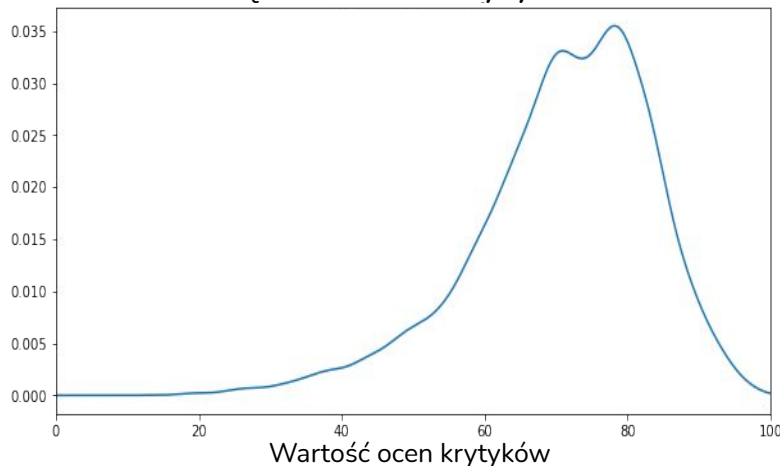




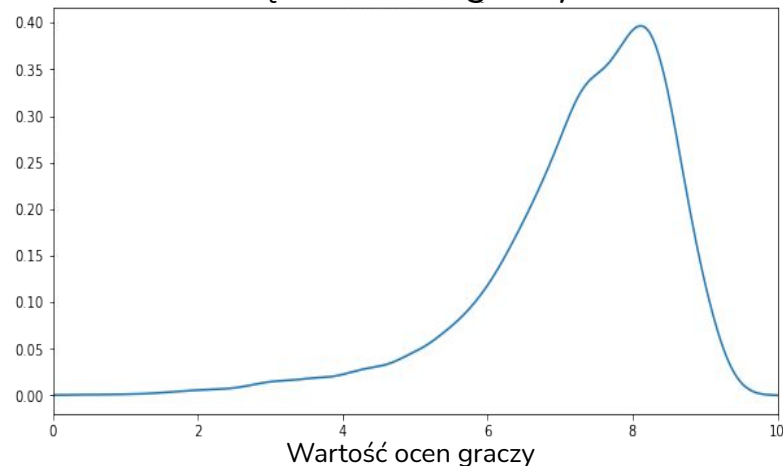
# Charakterystyka danych - oceny

- Zarówno oceny użytkowników, jak i krytyków przeważają w górnej części zakresu ocen
- Oceny użytkowników częściej są skrajnie dobre (ocena powyżej 8/10), natomiast krytycy zwracają większą uwagę na szczegóły, przez które przyznają oceny dobre (60-70/100)

Gęstość ocen krytyków



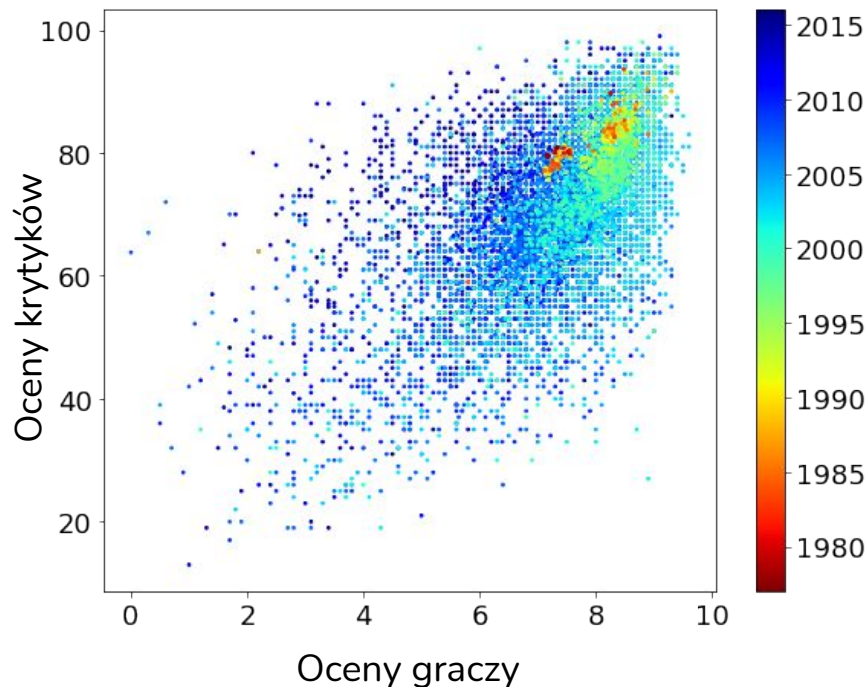
Gęstość ocen graczy





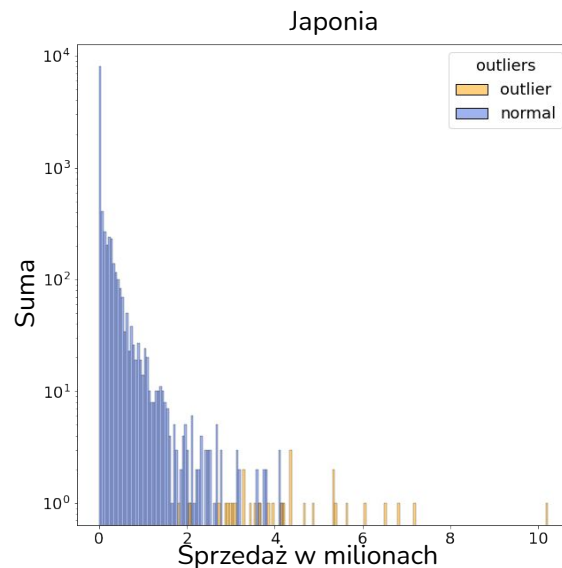
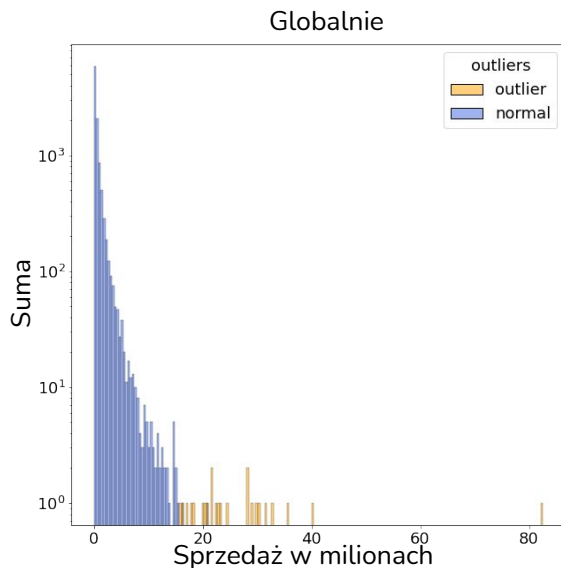
# Charakterystyka danych - oceny

- Na wykresie widać tendencję użytkowników do dawania skrajnie dobrych ocen grą, które krytycy ocenili jedynie dobrze
- W latach 2005-2015 powstało wiele gier-średniaków.
- Wśród najlepiej ocenianych gier znajdują się głównie te z okresu 1995-2005
- Starsze gry rzadko są słabo oceniane (częściowo wynika to z charakterystyki zbioru)



# Elementy nietypowe - sprzedaż

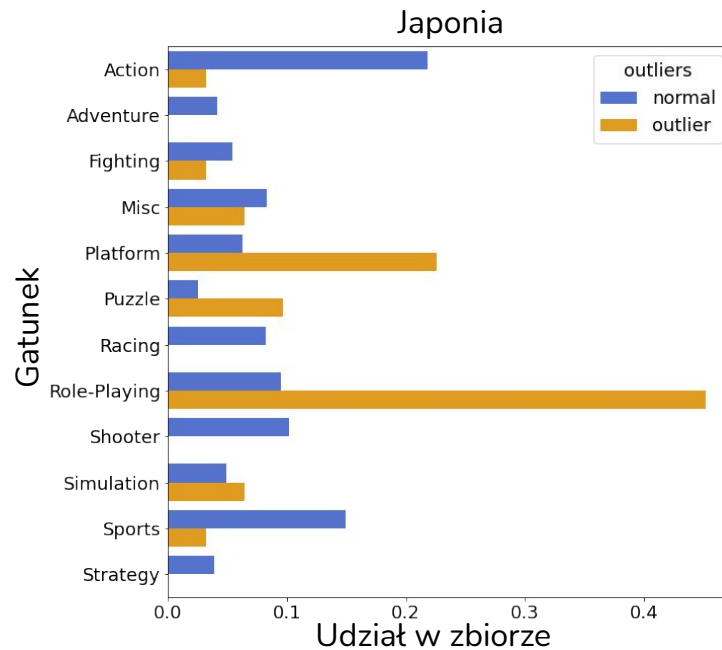
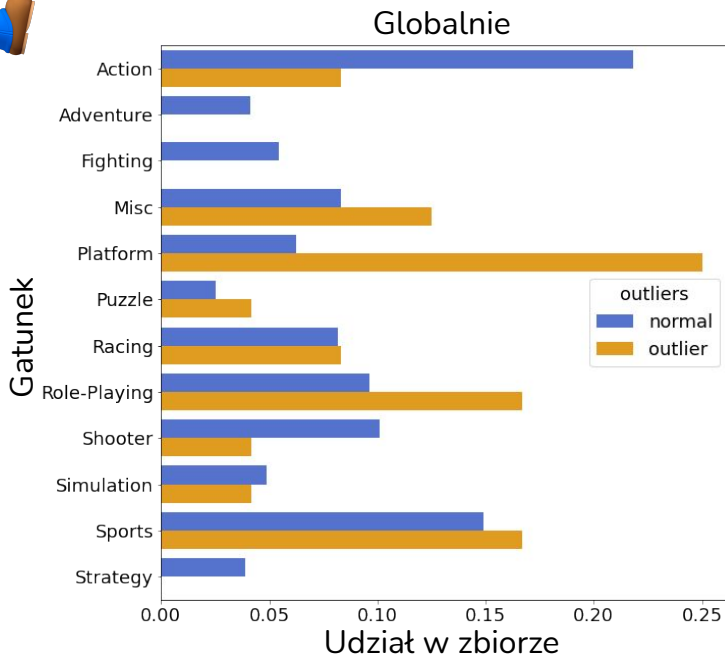
- Wykorzystano algorytm k-najbliższych sąsiadów
- Wyjątkowe są jedynie elementy o wysokiej wartości sprzedaży
- Na rynku japońskim więcej jest elementów o wartości bliskiej zero oraz więcej jest elementów o bardzo wysokiej sprzedaży





# Elementy nietypowe - sprzedaż

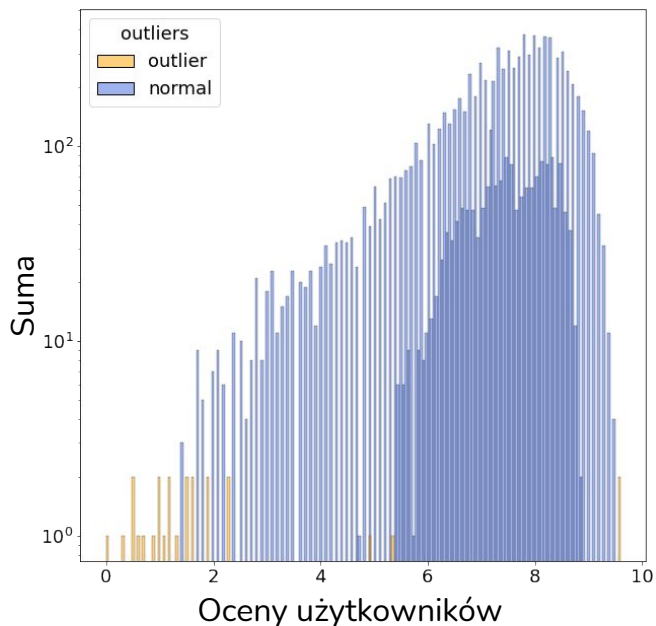
- Wykresy przedstawiają procentowy udział gatunku względem wszystkich elementów nietypowych/typowych
- Widoczne są bardzo duże różnice między rynkiem globalnym, a japońskim





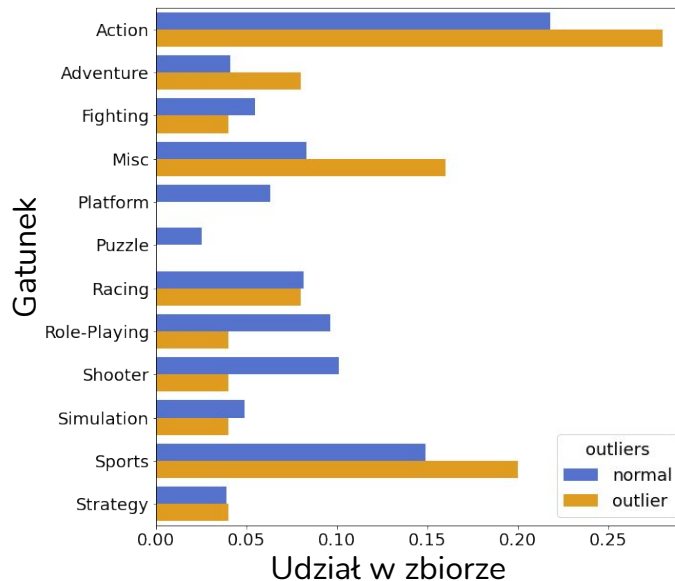
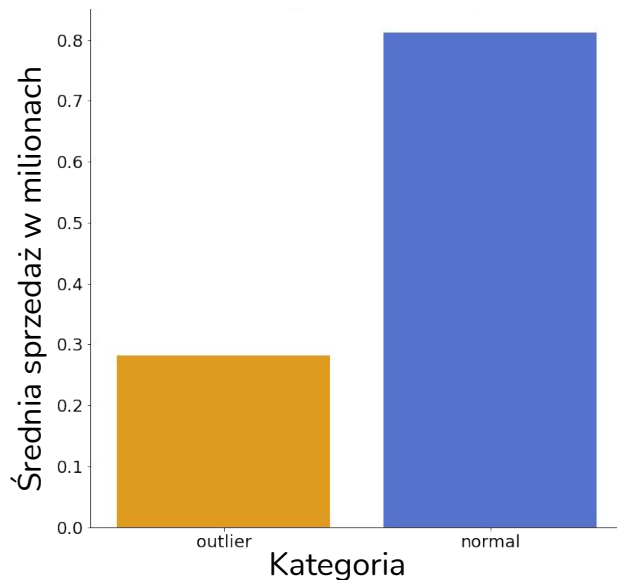
# Elementy nietypowe - oceny

- Elementami nietypowymi dla ocen użytkowników są głównie gry o bardzo niskich wynikach



# Elementy nietypowe - oceny

- Widać, że niska ocena przekłada się na sprzedaż, natomiast nie tak znacząco jak mogłoby się wydawać
- Nie ma tendencji co do gatunku gry.





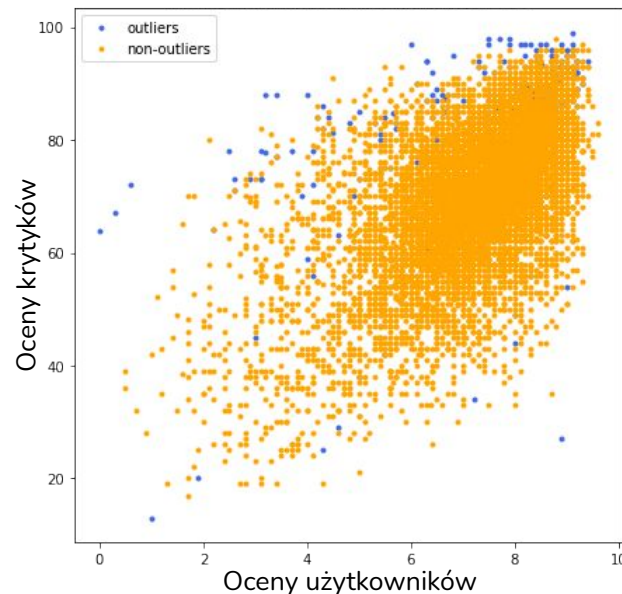
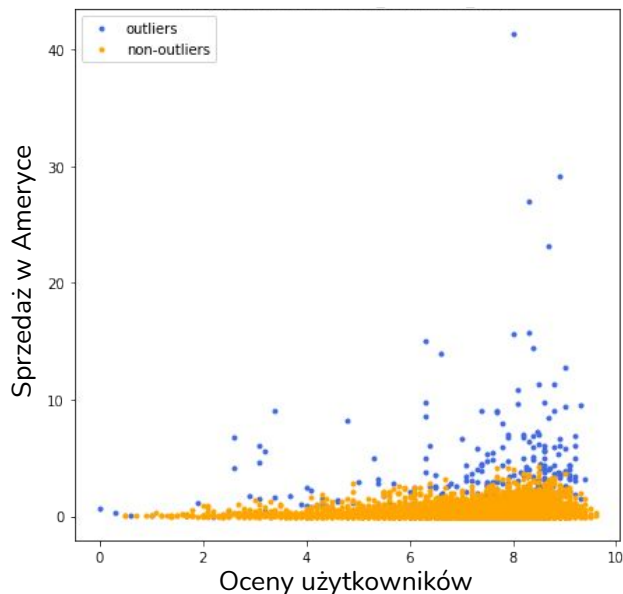
# Dane 6-wymiarowe

Do badań wykorzystane zostały następujące cechy:

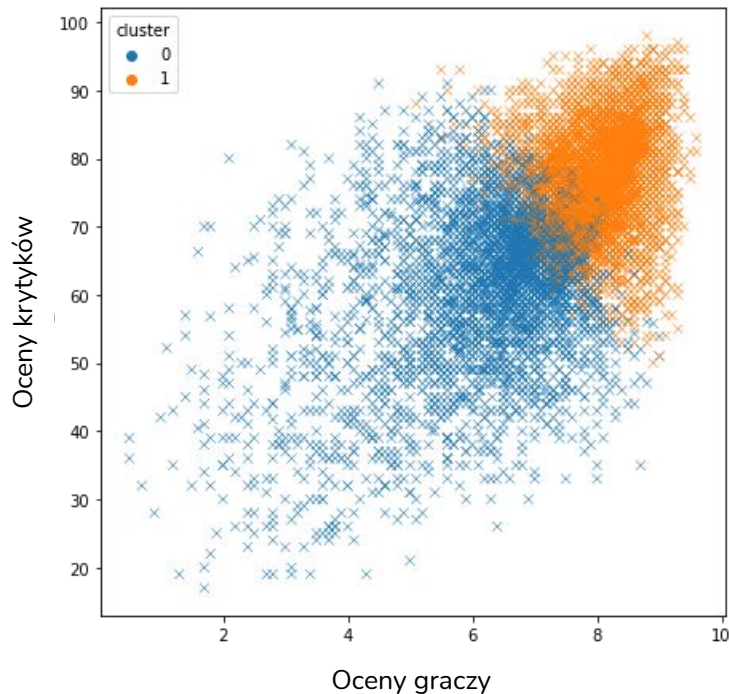
- oceny przyznawane przez krytyków (*Critic\_Score*)
- oceny przyznawane przez użytkowników (*User\_Score*)
- sprzedaż na rynku amerykańskim (*NA\_Sales*)
- sprzedaż na rynku europejskim (*EU\_Sales*)
- sprzedaż na rynku japońskim (*JP\_Sales*)
- rok wydania gry (*Year\_of\_Release*)

# Dane 6-wymiarowe: elementy nietypowe

- Przeważają elementy o bardzo wysokich wynikach sprzedaży i ocenach
- Występuje również kilka elementów nietypowych o bardzo niskich ocenach



# Dane 6-wymiarowe - 2 klastry



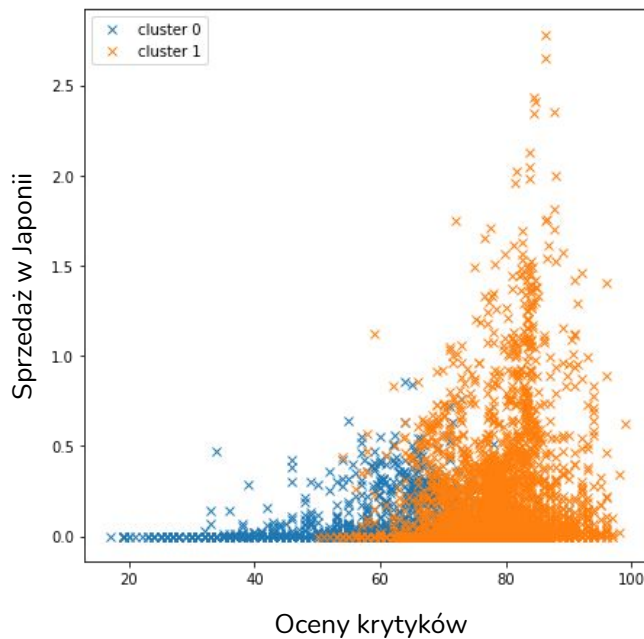
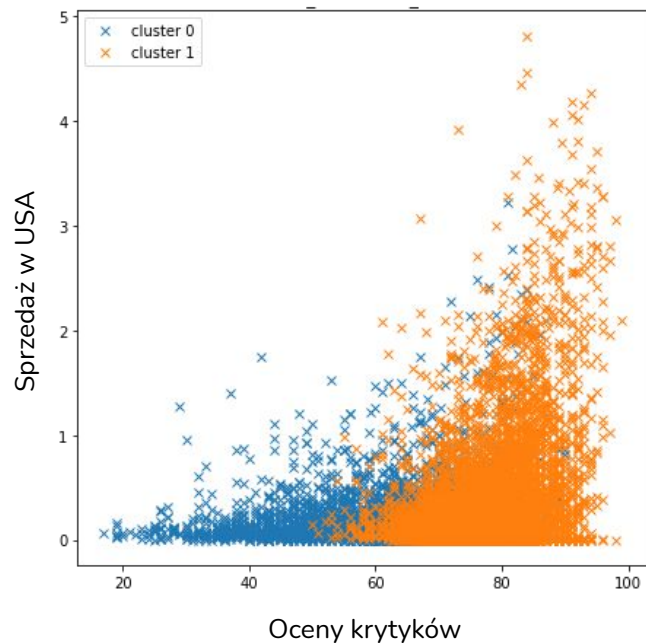
Dane podzielone zostały na 2 klastry:

- klaster **0**, oznaczony kolorem **niebieskim**, o liczności **4310**
- klaster **1**, oznaczony kolorem **pomarańczowym**, o liczności **5829**

Klaster **pomarańczowy** charakteryzuje się **wysokimi** ocenami.



# Dane 6-wymiarowe - 2 klastry



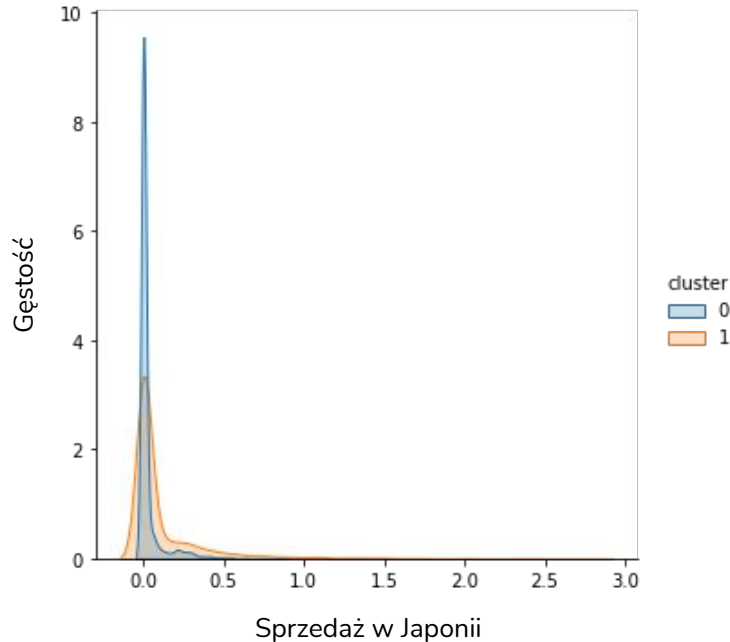
Klaster **niebieski** - **niska** sprzedaż (szczególnie Japonia).

Klaster **pomarańczowy** - **zróżnicowana** sprzedaż.

Dobre oceny nie są zatem warunkiem wystarczającym dobrej sprzedaży.

Rynek europejski podobny do amerykańskiego, dlatego pominięty.

## Dane 6-wymiarowe - 2 klastry

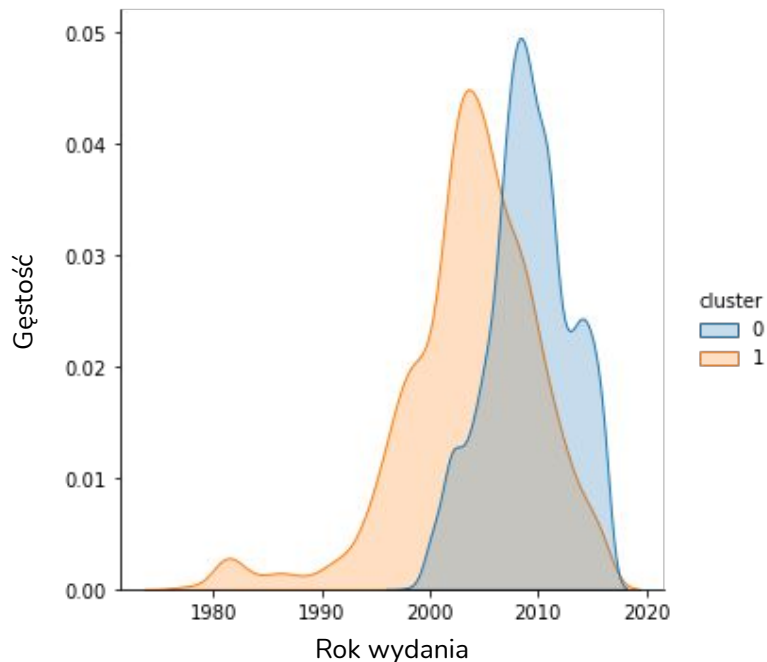


Oba klastry zawierają gry o niskiej sprzedaży, klaster pomarańczowy ponadto te sprzedające się dobrze (ogon estymatora).

Japonia - bardzo wyraźna dominacja gier o słabej i zerowej sprzedaży (gry niewydane, różnice kulturowe).



## Dane 6-wymiarowe - 2 klastry



Średnia w klastrze pomarańczowym: **2004**

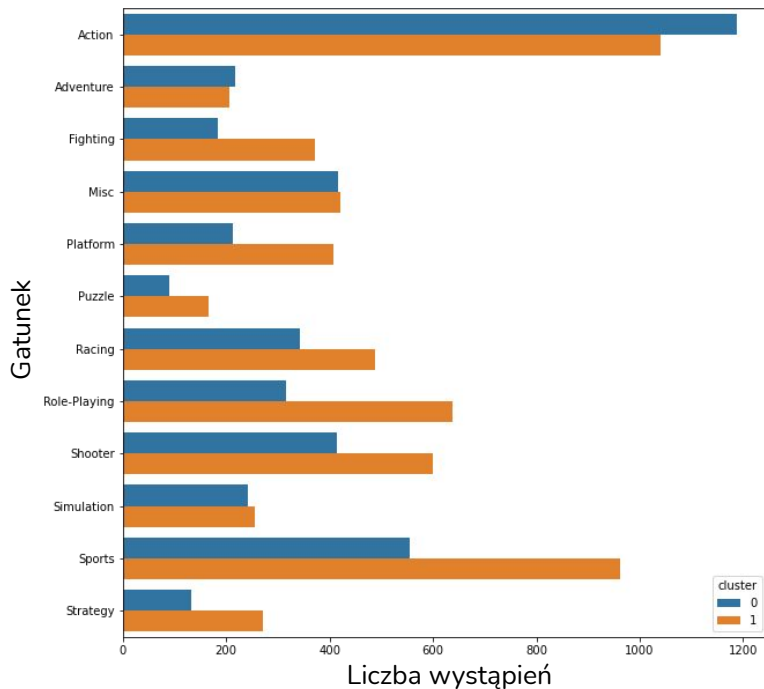
Średnia w klastrze niebieskim: **2009**

Klaster pomarańczowy - wysokie oceny

Wniosek: gry wydawane około 2004 roku  
zwykle odnosiły większy sukces.

Można wskazać kilka potencjalnych powodów  
takiego stanu rzeczy.

## Dane 6-wymiarowe - 2 klastry

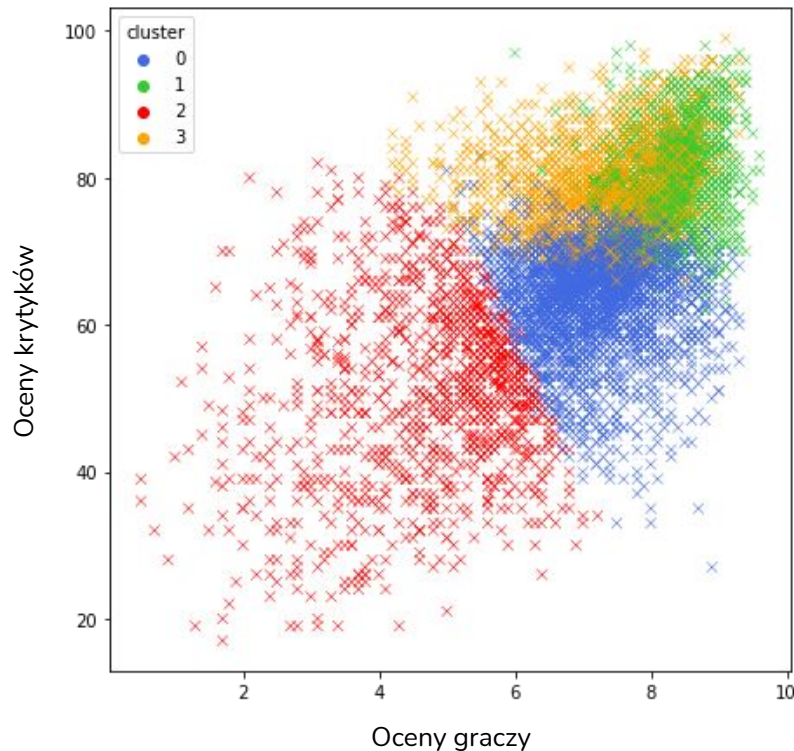


Różnice w **proporcjach** między poszczególnymi gatunkami.

Większość gier **akcji** znajduje się w klastrze **niebieskim** (niska sprzedaż i oceny)

Wyróżniają się gatunki **niszowe**, np.: gry sportowe, strategiczne.

# Dane 6-wymiarowe: 4 klastry



Podobna analiza przeprowadzona została dla 4 klastrów (metoda k-średnich).

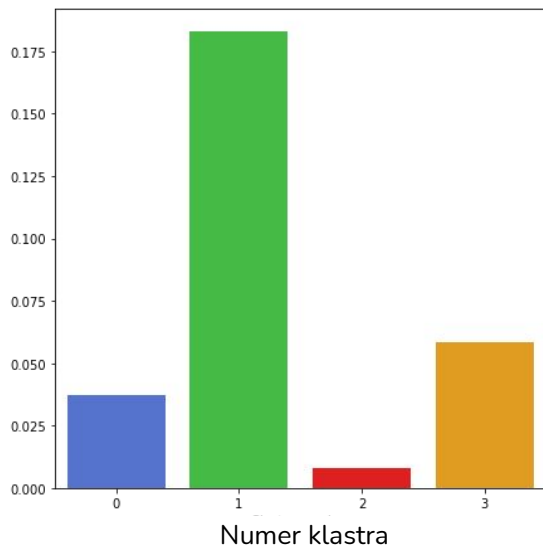
Zaprezentowane zostaną jedynie najważniejsze wnioski.

Intuicyjna interpretacja klastrów (oceny użytkowników i krytyków).

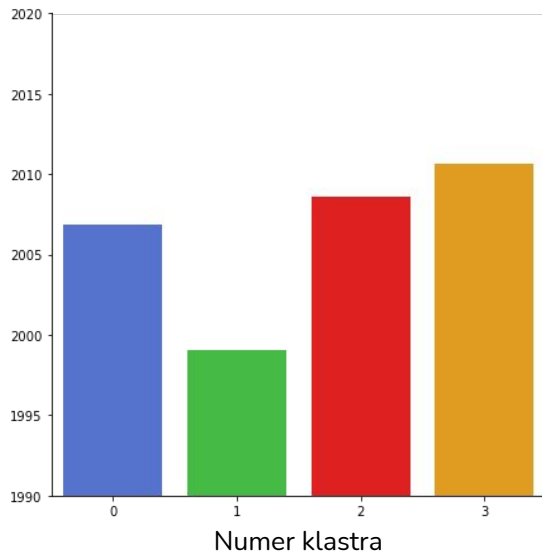


## Dane 6-wymiarowe: 4 klastry

Średnia sprzedaż w Japonii



Średni rok wydania



Średnia sprzedaż gier z klastra **zielonego** (bardzo wysokie wyniki zarówno od krytyków, jak i użytkowników) w Japonii.

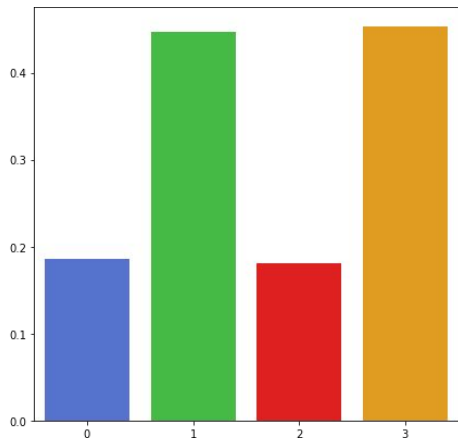
Średnio są to gry starsze (średnia: 1999).

W Japonii obserwujemy ponadto wyjątkowo **niską** sprzedaż gier o **słabych** ocenach (klastry czerwony).



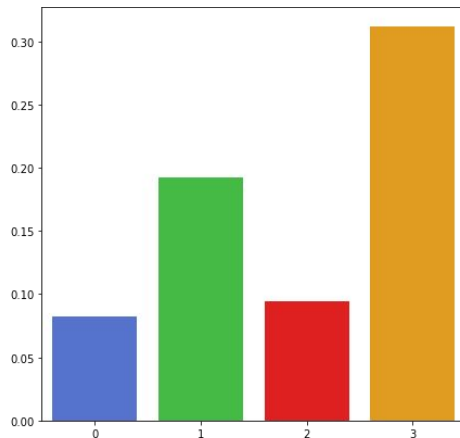
# Dane 6-wymiarowe: 4 klastry

Średnia sprzedaż w USA



Numer klastra

Średnia sprzedaż w Europie



Numer klastra

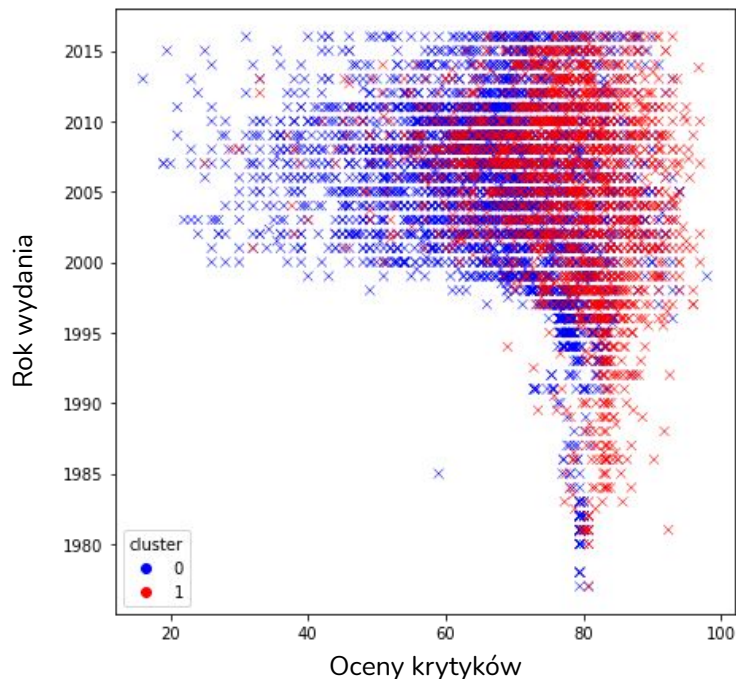
Tendencje zdecydowanie mniej widoczne na rynku amerykańskim oraz europejskim.

Nie występuje wyraźne rozróżnienie w sprzedaży gier ocenianych wysoko i bardzo wysoko.

Nie ma znacznego rozróżnienia w sprzedaży gier przeciętnych i bardzo słabych.

Brak wyraźnej zależności od roku wydania.

# Klasyfikacja



Klasyfikujemy: **globalną sprzedaż gier**  
na podstawie: **ocen krytyków i roku wydania**.

Dokonano agregacji sprzedaży gier wydanych na różne platformy - punkt widzenia wydawcy.

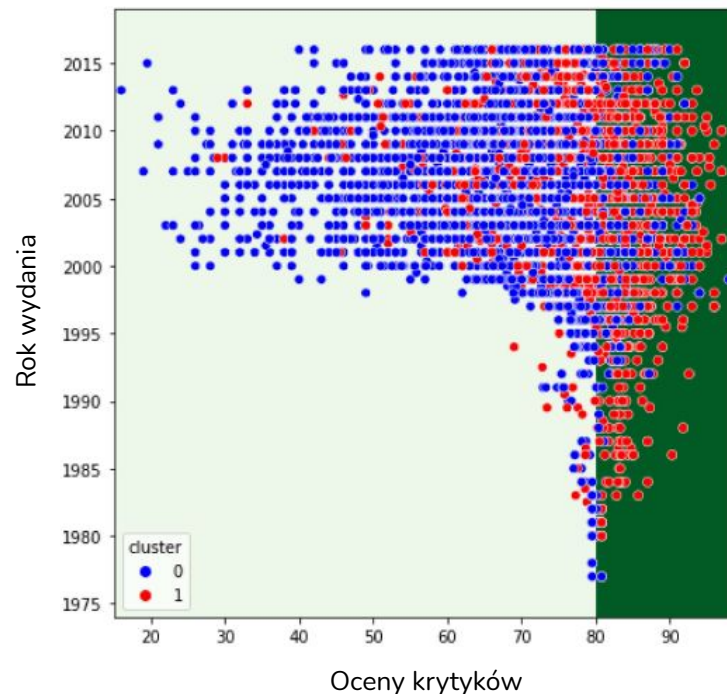
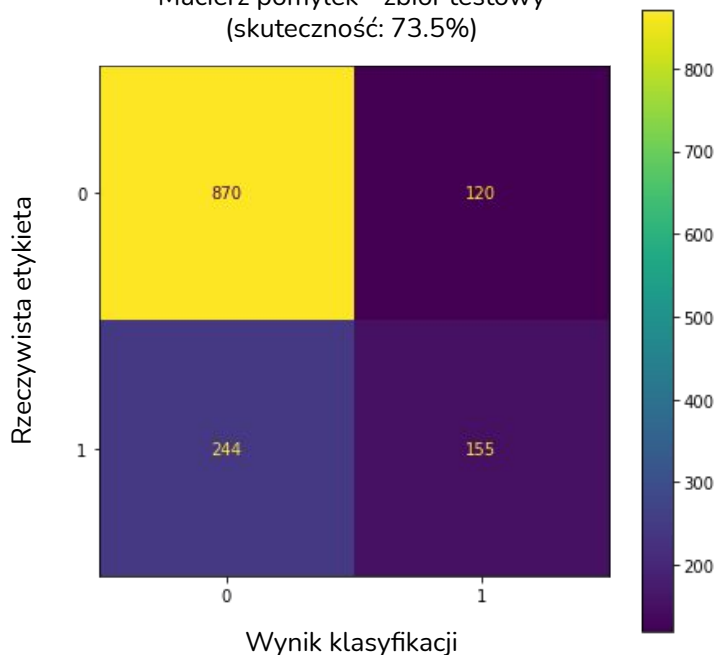
Globalną sprzedaż podzielono na **dwie klasy**:

- gry o **wysokiej** sprzedaży - powyżej 1 miliona (1993 elementy)
- gry o **niskiej i przeciętnej** sprzedaży - poniżej 1 miliona (4952 elementy)

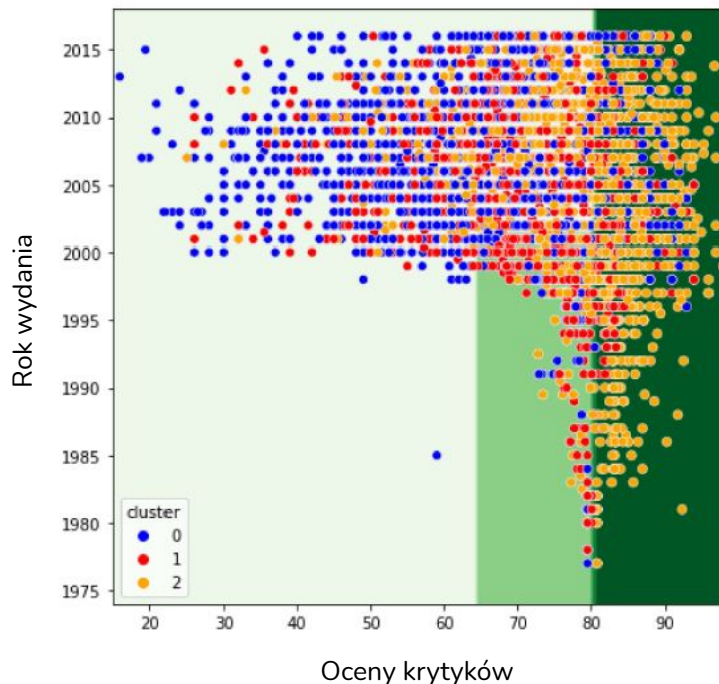
Podział zbioru na uczący i testowy (8:2), drzewa decyzyjne.

# Klasyfikacja - drzewo decyzyjne

Macierz pomyłek - zbiór testowy  
(skuteczność: 73.5%)



# Klasyfikacja - 3 klasy



Rozróżniamy dodatkowo sprzedaż słabą od przeciętnej.

Klasy:

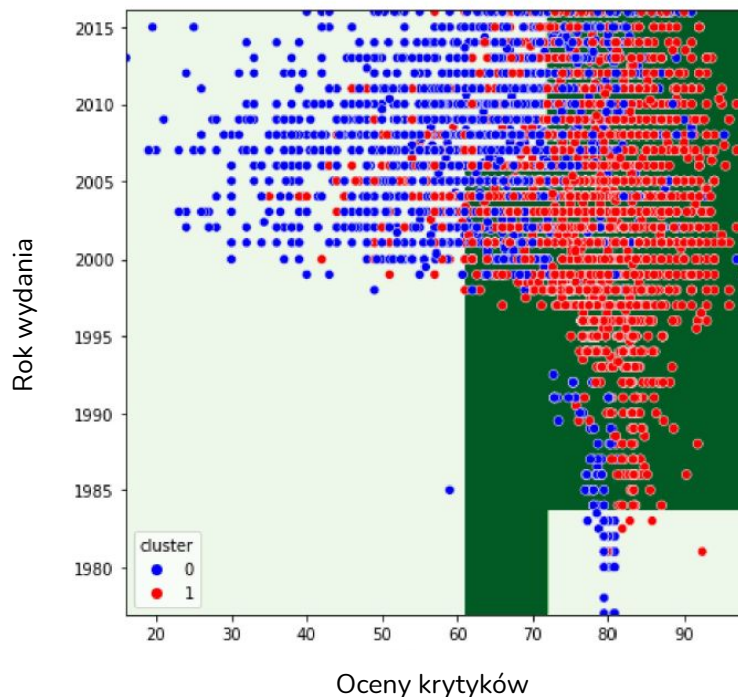
- gry o **wysokiej** sprzedaży - powyżej 1 miliona
- gry o **przeciętnej** sprzedaży - poniżej 1 miliona, powyżej 0.3 miliona
- gry o **słabej** sprzedaży - poniżej 0.3 miliona

Wynik dla drzewa decyzyjnego:

- 50.8% (zbiór testowy)
- 57.0% (zbiór uczący)



# Klasyfikacja



Podobne badanie, ale z **ocenami graczy**.

Klasy:

- gry **ciepło przyjęte** przez graczy (oceny większa od 7.5)
- gry o **przeciętnych i niskich** ocenach od graczy

Wynik dla drzewa decyzyjnego:

- 76.0% (zbiór testowy)
- 76.1% (zbiór uczący)



# Podsumowanie i wnioski

Analiza zbioru pozwala wyciągać ogólne, zgodne z doświadczeniami autorów, wnioski na temat rynku.

Z uwagi na bardzo ogólny charakter zbioru (brak informacji takich jak budżet czy szczegóły dotyczące rozgrywki), nie można wyciągać natomiast wniosków pozwalających dokładnie scharakteryzować cechy “gry idealnej” - tzn. analiza nie daje nam przepisu na pewny sukces.

Analiza dostarczyła natomiast wielu informacji na temat praktyk, które mogą negatywnie wpłynąć na wynik gry.

Brak jest niestety darmowych, ogólnodostępnych źródeł, zawierających bardziej szczegółowe informacje - analizowany zbiór był jednym z dwóch zbiorów dotyczących tematyki gier komputerowych, które udało się autorom znaleźć.



**Dziękujemy za uwagę.**