Analiza i wizualizacja danych

**Projekt**

**Prowadzący: dr Jarosław Olejniczak**

**Autor: Piotr Świtakowski**

**Grupa: WCY18IJ4S1**

# Opis zbioru danych

Zbiór danych **Top Games on Google Play Store** pobrany ze strony:

<https://www.kaggle.com/dhruvildave/top-play-store-games>

Zbiór danych **zawiera informacje na temat najpopularniejszych gier na platformę Android** (ze sklepu Google Play Store). Dokładniej mówiąc, zbiór danych przestawia top100 z każdej kategorii gier dostępnej na Sklepie Play – oceny jakie otrzymały od użytkowników, ich ceny oraz liczbę instalacji.

Dane zostały zebrane 9 kwietnia 2021 roku.

Zbiór zawiera 1730 obserwacji. Znajdują się w nim następujące kolumny:

* **rank** – pozycja gry w danej kategorii (top1 – top100)
* **title** – tytuł gry
* **total ratings** – całkowita liczba ocen użytkowników danej gry
* **installs** – średnia ilość instalacji danej gry
* **average rating** – średnia ocena użytkowników (w skali od 1 do 5)
* **growth (30 days)** – Procentowy wzrost średniej oceny gry na przestrzeni ostatnich 30 dni
* **growth (60 days)** - Procentowy wzrost średniej oceny gry na przestrzeni ostatnich 60 dni
* **price** – cena gry (w dolarach)
* **category** – kategoria gry
* **5 star ratings** – ilość ocen 5 (najlepsza możliwa ocena), jaką otrzymała dana gra
* **4 star ratings** – ilość ocen 4 jaką otrzymała dana gra
* **3 star ratings** – ilość ocen 3 jaką otrzymała dana gra
* **2 star ratings** – ilość ocen 2 jaką otrzymała dana gra
* **1 star ratings** – ilość ocen 1 (najgorsza możliwa ocena), jaką otrzymała dana gra
* **paid –** określa czy gra jest płatna bądź darmowa

# Modyfikacja zbioru danych

W zbiorze nie było brakujących wartości, nie musiałem usuwać bądź modyfikować żadnych wierszy.

Następnie sprawdziłem, czy w zbiorze wystąpiły jakieś mocno wystające wartości – również brak.

Kolumna **installs** zawierała dane w formacie tekstowym, np. 100.M (100 milionów instalacji) – przekształciłem te dane do postaci liczbowej, np. ‘100.M’ na ‘100 000 000’

Kolumna **paid** zawierała wartości w postaci ‘True’ lub ‘False’ – zamieniłem te wartości na liczbowe, tj. ‘True’ na ‘1’ i ‘False’ na ‘0’.

Kolumna **category**: każda kategoria gier była niepotrzebnie poprzedzona słowem „GAME”. Pozbyłem się tego.

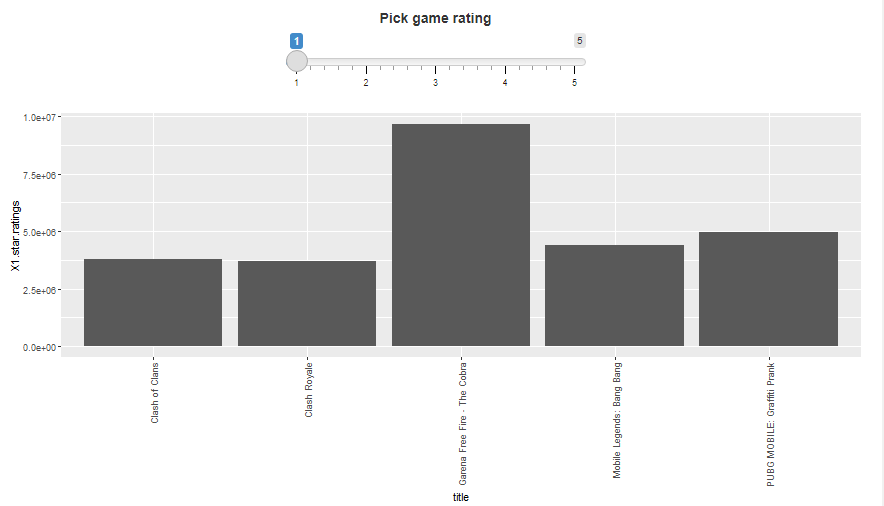
# Wizualizacja danych

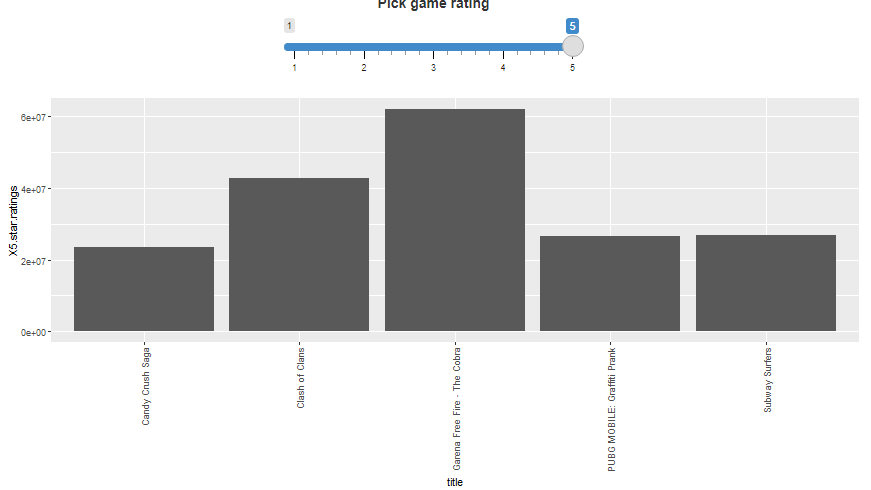
## Word Cloud (wykres interaktywny)



Pierwszym stworzonym wykresem jest word cloud zawierający nazwy najpopularniejszych gier. Im większa czcionka, tym bardziej popularna jest gra. Widocznie odstaje od reszty jeden tytuł – „Garena Free Fire – The Cobra”. Oznacza to, że w badanym okresie była to zdecydowanie najbardziej popularna gra dostępna na platformę Android. Dodatkowym aspektem pokazanym na wykresie jest kolor czcionki – im jaśniejszy, tym więcej pobrań miała aplikacja. Jak nie trudno się domyślić, ilość pobrań idzie w parze z popularnością.

## Top 5 gier z największą ilością każdej z ocen (wykres interaktywny)



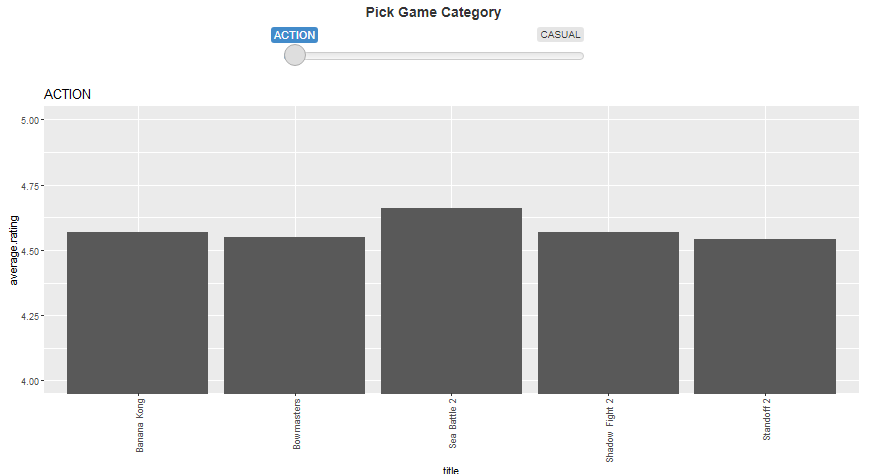


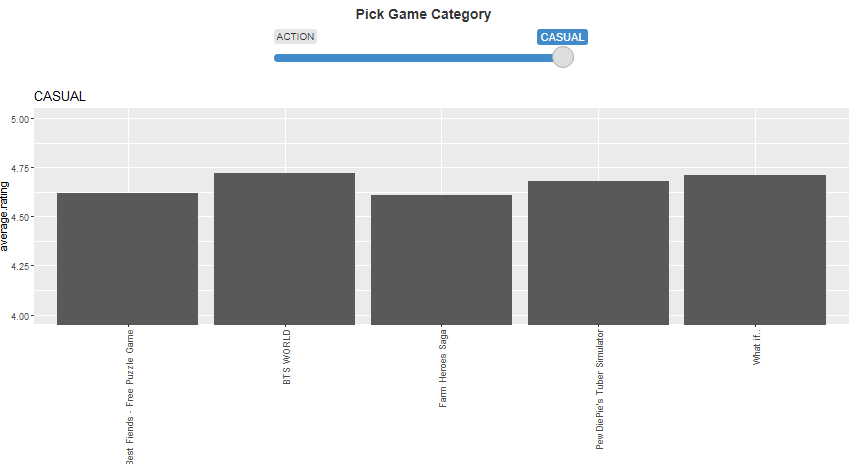
Wykres ukazujący top5 gier z największą ilością każdej z możliwych ocen. Na wykresie interaktywnym możemy wybrać dowolną ocenę, a zobaczymy top5 gier z największą ilością wybranej oceny.

W kategorii najlepszej oceny (5) wygrywa „Garena Free Fire – The Cobra”, czyli najpopularniejsza gra w całym zestawieniu jest równocześnie najczęściej oceniana jako 5.

Co ciekawe, w przeciwnym zestawieniu, czyli najwięcej najgorszych ocen, dominuje ten sam tytuł. Można z tego wywnioskować, że wraz z ilością ocen rośnie szansa na otrzymanie najgorszej możliwej oceny. Mimo tak dużej ilości ocen negatywnych, tej grze wciąż udało się utrzymać top1 pod względem popularności.

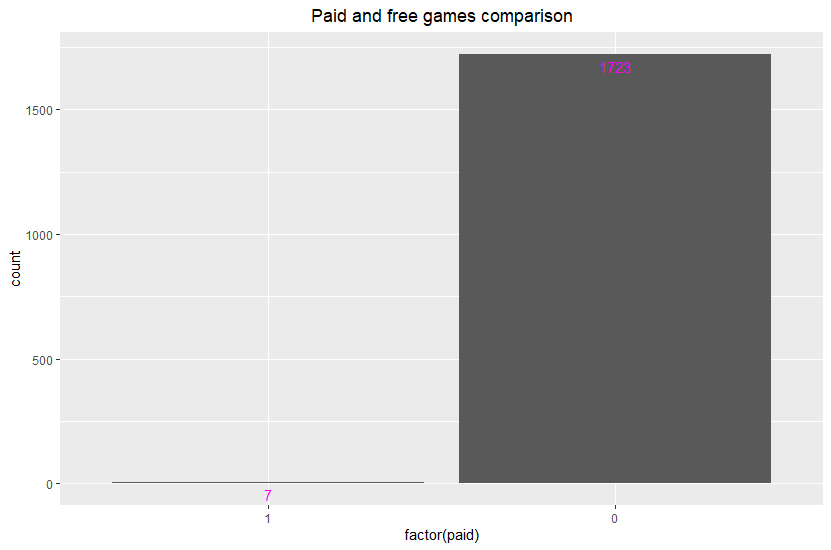
## Porównanie top5 gier z każdej kategorii pod względem średniej oceny (wykres interaktywny)





Wykres interaktywny przedstawia top5 gier z pięciu wybranych kategorii pod względem ich średniej oceny. Możemy zauważyć, że oceny wszystkich przedstawionych na wykresach gier mieszczą się w przedziale <4.0,5.0>, jednak żadna z topowych gier nie przekracza oceny 4,8. Przy tak dużej ilości ocen, ciężko jest to osiągnąć.

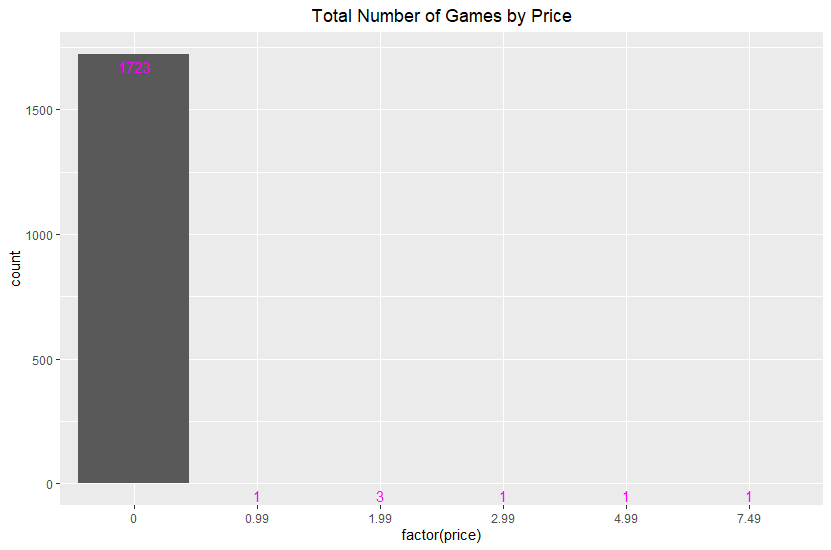
## Porównanie ilości darmowych oraz płatnych gier



Jak widzimy na wykresie, płatne gry stanowią bardzo mały odsetek najpopularniejszych gier na rynku. Świadczy to o tym, że:

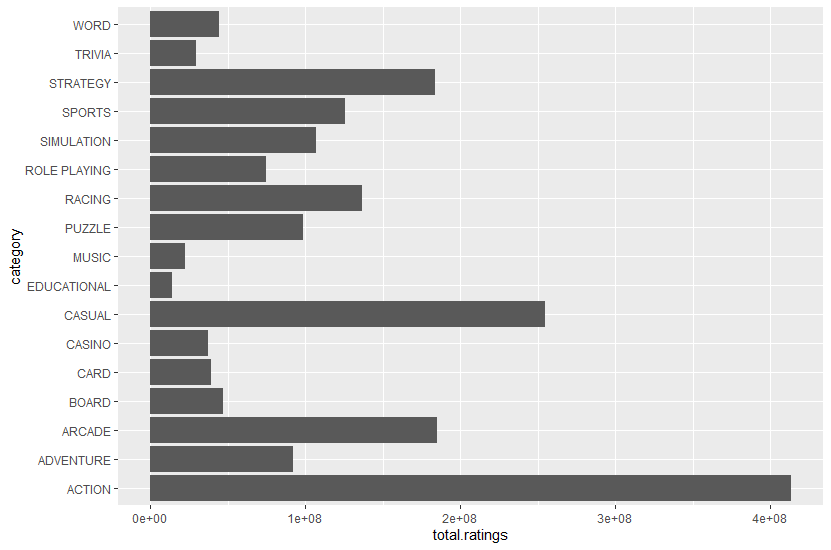
* Ludzie nie chcą płacić za gry mobilne
* Sporą część graczy mobilnych stanowią dzieci, wydatek na gry jest dla nich dużym problemem
* Deweloperzy wolą zarobek na reklamach podczas grania, niż bezpośrednią opłatę za możliwość pobrania gry

## Ilość gier w poszczególnych cenach



Jak widzimy, praktycznie wszystkie gry w naszym zestawieniu są darmowe. Jedynie kilka gier jest płatnych i nie są to szczególnie wygórowane ceny. Świadczy to o tym, że rynek gier mobilnych dominują darmowe gry, jedynie nieliczne płatne gry osiągają duży sukces.

## Ilość ocen w zależności od kategorii



Z wykresu możemy odczytać, że najczęściej ocenianą kategorią gier jest gra akcji, zaraz po niej gra casualowa. Z kolei najrzadziej ocenianymi kategoriami gier są gry edukacyjne oraz gry muzyczne. Może to wynikać z faktu, że gry mobilne zazwyczaj nie są grami „na dlużej” i użytkownicy chętniej wybierają gry akcji, aby lepiej się bawić podczas gry.

# Podsumowanie

Rynek gier mobilnych nie jest skomplikowany – darmowe gry dominują rynek. Przeciętny gracz mobilny nie jest nastawiony na dłuższą rozgrywkę i woli wybrać darmowy tytul, najczęściej z kategorii gry akcji bądź gry casualowe.

Biorąc pod uwagę rząd ilości ocen jakie otrzymują topowe gry, można śmialo stwierdzić, że rynek ten jest potężny i ciągle się rozwija.

Jeżeli ktoś chciałby zwiększyć szanse na sukces swojej gry mobilnej, powinien celować w wymienione wyżej kategorie.