

Praca projektowa nr 3

Przetwarzanie danych ustrukturyzowanych

Norbert Olkowski Dawid Poławski Jakub Muszyński

Wydział Matematyki i Nauk Informacyjnych, Politechnika Warszawska

31 maja 2021

- ❶ Wstęp
- ❷ Zapytania
- ❸ Kwestie techniczne
- ❹ Zapytanie 1
- ❺ Zapytanie 2
- ❻ Zapytanie 3
- ❼ Zapytanie 4
- ❽ Podsumowanie

W realizacji projektu wykorzystaliśmy dane zebrane z serwisów:

- *Arqade*
- *Chess*
- *Board & Card Games*

Na ich podstawie przeprowadziliśmy analizę danych, mającą na celu odpowiedzenie na poniższe pytania.

Zapytania

- 1 Jak zmieniała się popularność gier na przestrzeni miesięcy?
- 2 Wybór 15 najaktywniejszych użytkowników forum *Arqade*.
- 3 Jak serial *Gambit Królowej* wpłynął na popularyzację szachów?
- 4 Porównanie popularności wybranych serwisów: *Arqade*, *Chess*, *Board & Card Games*.

Wykorzystane pakiety:

- data.table
- dplyr
- stringi
- tidyr
- ggplot2
- gganimate

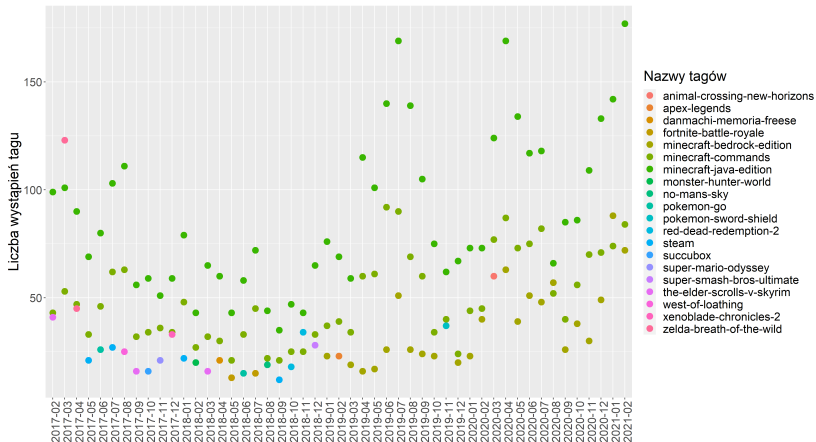
Sposób importu danych

- konwersja XML -> CSV przy pomocy funkcji udostępnionej na stronie prof. Gągolewskiego

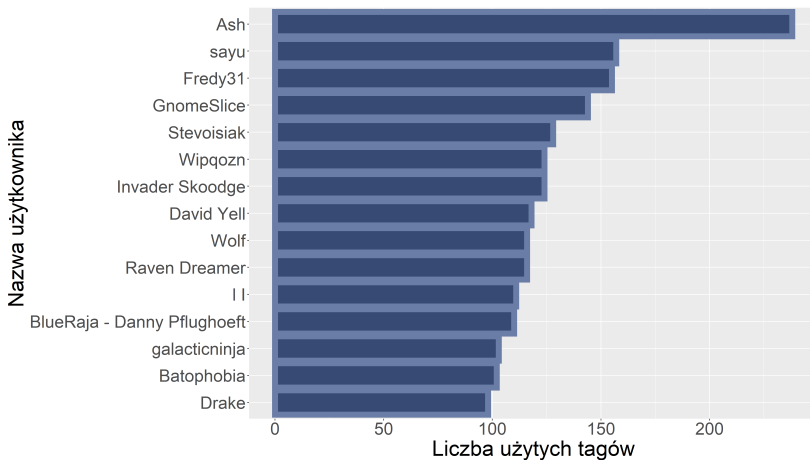
Wykorzystane struktury danych

Zapytanie 1

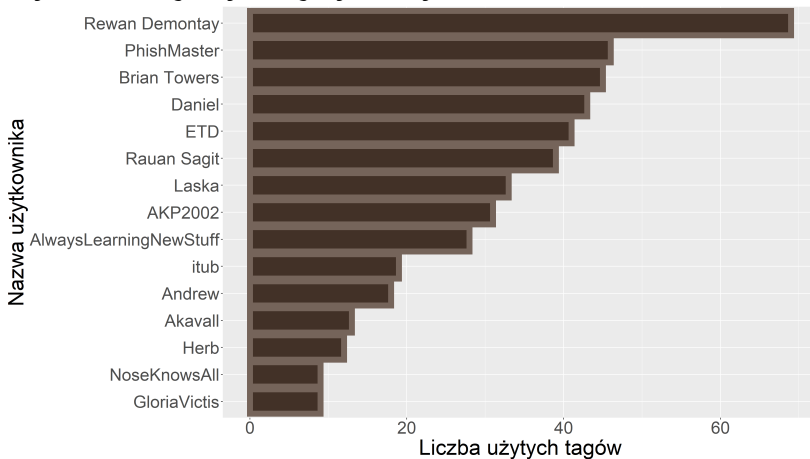
Jak zmieniała się popularność gier na przestrzeni miesięcy?



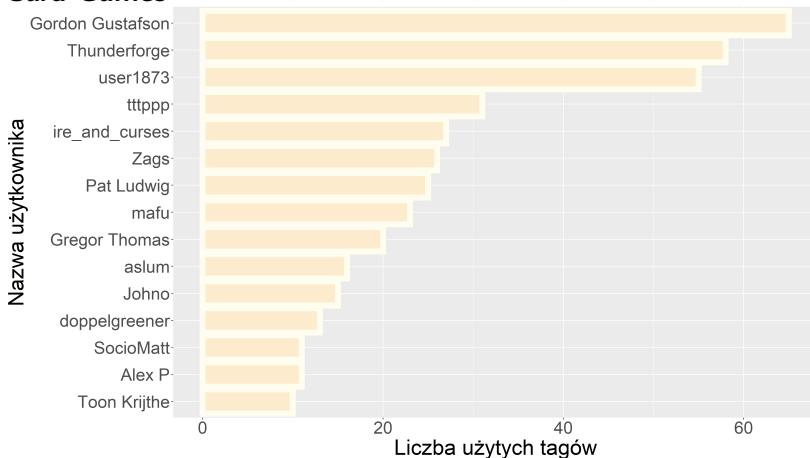
Wybór 15 najaktywniejszych użytkowników forum *Arqade*



Wybór 15 najaktywniejszych użytkowników forum *Chess*

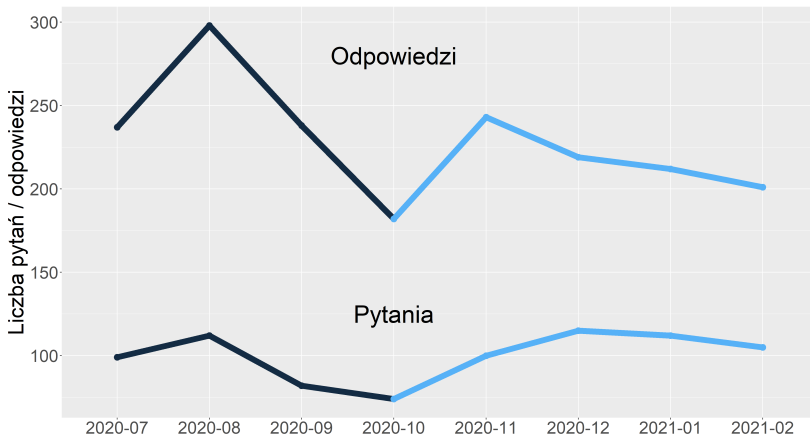


Wybór 15 najaktywniejszych użytkowników forum *Board & Card Games*



Zapytanie 3

Jak serial *Gambit Królowej*¹ wpłynął na popularyzację szachów?



¹Serial Netflix'a (2020), reż. Scott Frank, Allan Scott

Zapytanie 4

Wzory wykorzystywane do obliczenia wartości trendu

$$\gamma(x) = \begin{cases} \frac{ViewCount}{Score} & \text{jeśli } Score \neq 0 \\ \frac{ViewCount}{Score+1} & \text{jeśli } Score = 0 \end{cases}$$

$$PunktyTrendu(x) = \begin{cases} Score & \text{jeśli } x \text{ jest odpowiedzią} \\ \gamma(x) & \text{jeśli } x \text{ nie jest odpowiedzią} \end{cases}$$

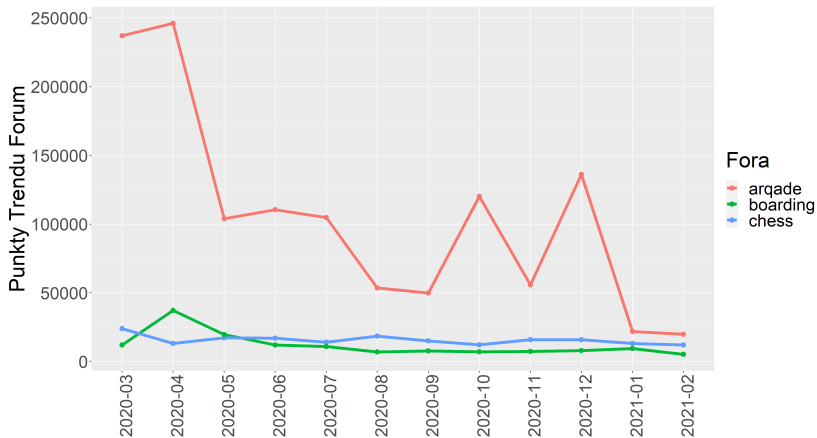
$$PunktyTrenduForum(y) = \sum_{i=w_1}^{w_n} PunktyTrendu(i)$$

gdzie:

- w_i to i -ty wiersz ramki Posts z serwisu y ;
- $i \in [n]$;
- n - liczba wierszy ramki Posts z serwisu y .

Zapytanie 4

Porównanie popularności wybranych serwisów: *Arqade*, *Chess*, *Board & Card Games*



Podsumowanie

- ❶ Po przeanalizowaniu danych wnioskujemy, że największe zainteresowanie na przestrzeni czterech ostatnich lat osiągnęła gra *Minecraft*. Ponadto wzrosło zainteresowanie grą *Fortnite*.
- ❷ Liczba tagów, którą użył *Ash* jest około 3,5 raza większa niż pozostali najaktywniejsi użytkownicy w wybranych forach.
 - najaktywniejszym użytkownikiem forum *Arqade* był *Ash*
 - najaktywniejszym użytkownikiem forum *Chess* był *Rewan Demontay*
 - najaktywniejszym użytkownikiem forum *Board & Card Games* był *Cordon Gustafson*
- ❸ Serial *Gambit Królowej* wpłynął pozytywnie na wzrost zapytań na forum oraz pozytywnie wzrost odpowiedzi, po premierze filmu na *Netflixie*.
- ❹ Na podstawie zastosowanego wskaźnika do porównania popularności forów wnioskujemy, że bezapelacyjnym zwycięzcą jest forum *Arquade*.

*As for elegance, R is refined, tasteful, and beautiful. When I grow up, I want to marry R. - **Andy Bunn***