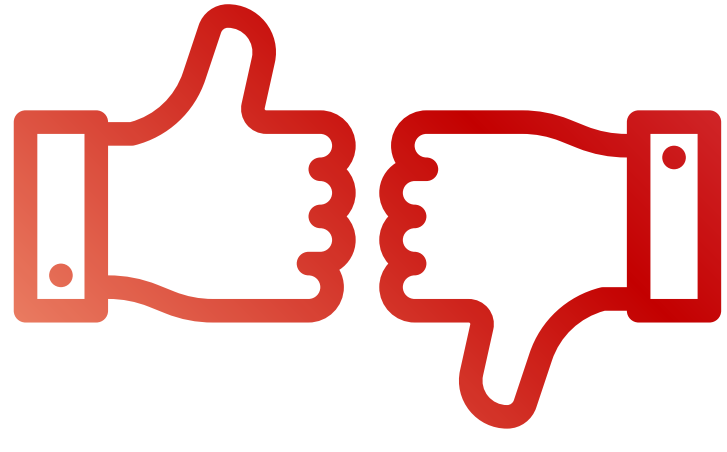




Komentarze jako źródło wiedzy



Autorzy:

Aleksandra Janczewska 256244@student.pwr.edu.pl

Arina Golubeva 251509@student.pwr.edu.pl

Piotr Szuba 252807@student.pwr.edu.pl

Opiekun projektu: Dr hab. inż. Tomasz Kajdanowicz



Politechnika Wroclawska



Katedra Sztucznej Inteligencji

Opis problemu

YouTube jest jednym z najbardziej popularnych mediów, gdzie autorzy udostępniają filmy o różnych tematach i zawartościach oraz podążają za sukcesem, który określa się liczbą wyświetleń i subskrybentów.

Korzystając z dostępnych informacji oraz komentarzy pod filmami postanowiono przeprowadzić badanie na temat zależności pomiędzy różnymi czynnikami a liczbą wyświetleń filmów oraz sprawdzić jakie informacje zawierają komentarze.



Metodyka

Zbrane dane zostały przygotowane do przetworzenia przez modele. Stosując model Twitter Sentiment PL (base) dostępny na stronie huggingface.com uzyskano etykiety wydzźwięku komentarzy na podstawie których została przeprowadzona analiza kontrowersji oraz inne analizy.

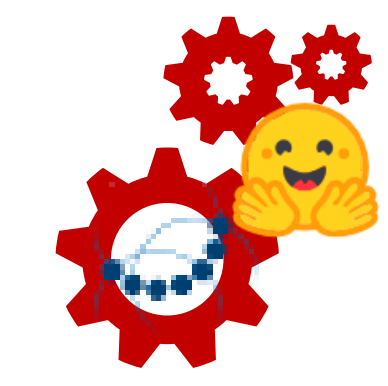
$$\text{kontrowersja} = \frac{\min(\text{Pozytywne}, \text{Negatywne})}{\max(\text{Pozytywne}, \text{Negatywne})} \times \frac{1}{\text{Neutralne}}$$

Oprócz kontrowersji zostały przeprowadzone również takie badania, jak:

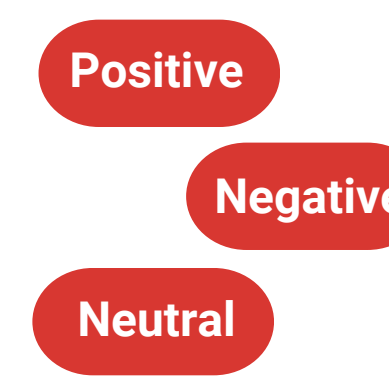
- analiza sentymentu,
- modelowanie aspektowe (za pomocą modelu aspectEmo),
- grupowanie komentujących i sprawdzenie czy kanał ma fanów.



Zebra-
nie
danych



herbert-base model
& AspectEmo



Analizy



Wnioski i
podsumowanie

Dane

Dane zostały zebrane z wykorzystaniem własnego narzędzia do pobierania danych z wybranych kanałów na YouTube.

Dane zostały pozyskane z 6 kanałów o różnej liczbie subskrypcji.

Podstawowe informacje o zbiorze danych:

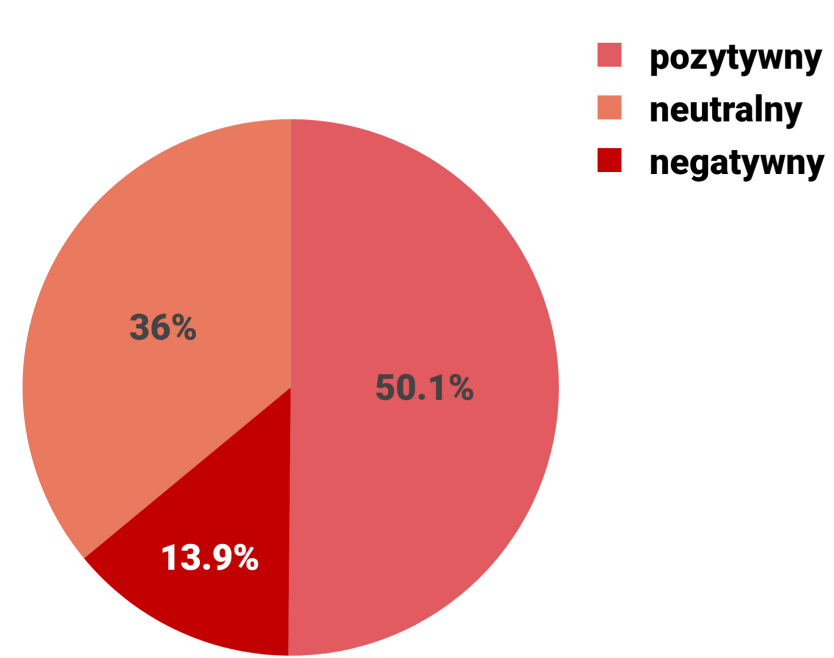
- 961 096 komentarzy i liczba polubień każdego z nich,
- 3917 filmów i liczba ich wyświetleń,
- dane zebrane od początku istnienia danego kanału do 26.11.2023.



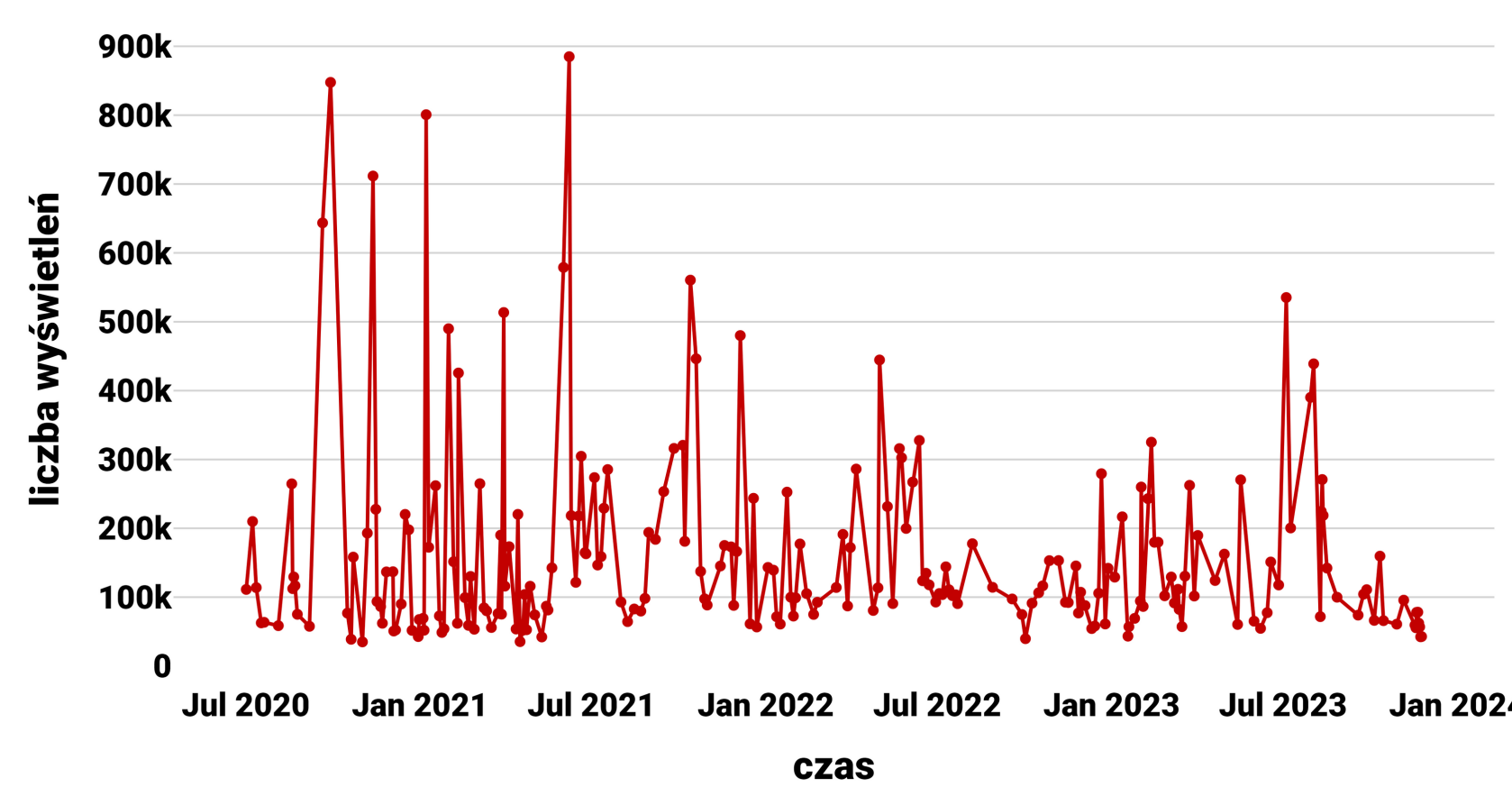
Analiza i wnioski

Analiza danych i statystyki dla przykładowego kanału

Rozkład wydzźwięku komentarzy



Liczba wyświetleń zmieniająca się w czasie

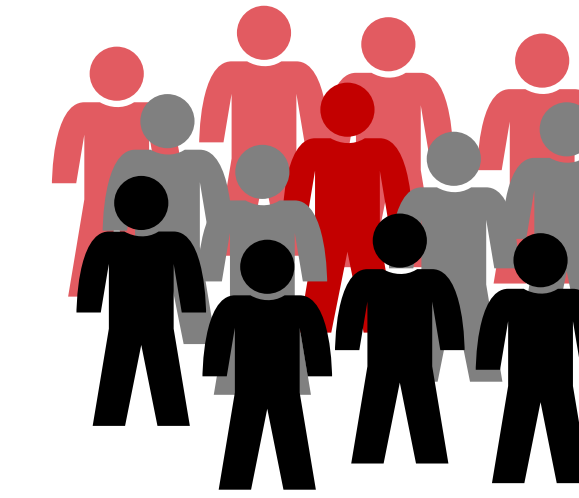


Wpływ sentymentu komentarzy na liczbę wyświetleń

Wszystkich komentujących (w obrębie jednego kanału) podzielono na 4 grupy (pozytywnych, negatywnych, neutralnych i zróżnicowanych) na podstawie sentymentu ich komentarzy. Zauważono, że korelacja pomiędzy liczbą wyświetleń filmu, a wydzźwiękiem komentarzy liczona osobno dla każdej grupy jest większa niż liczona dla wszystkich komentujących.

Dane niegrupowane

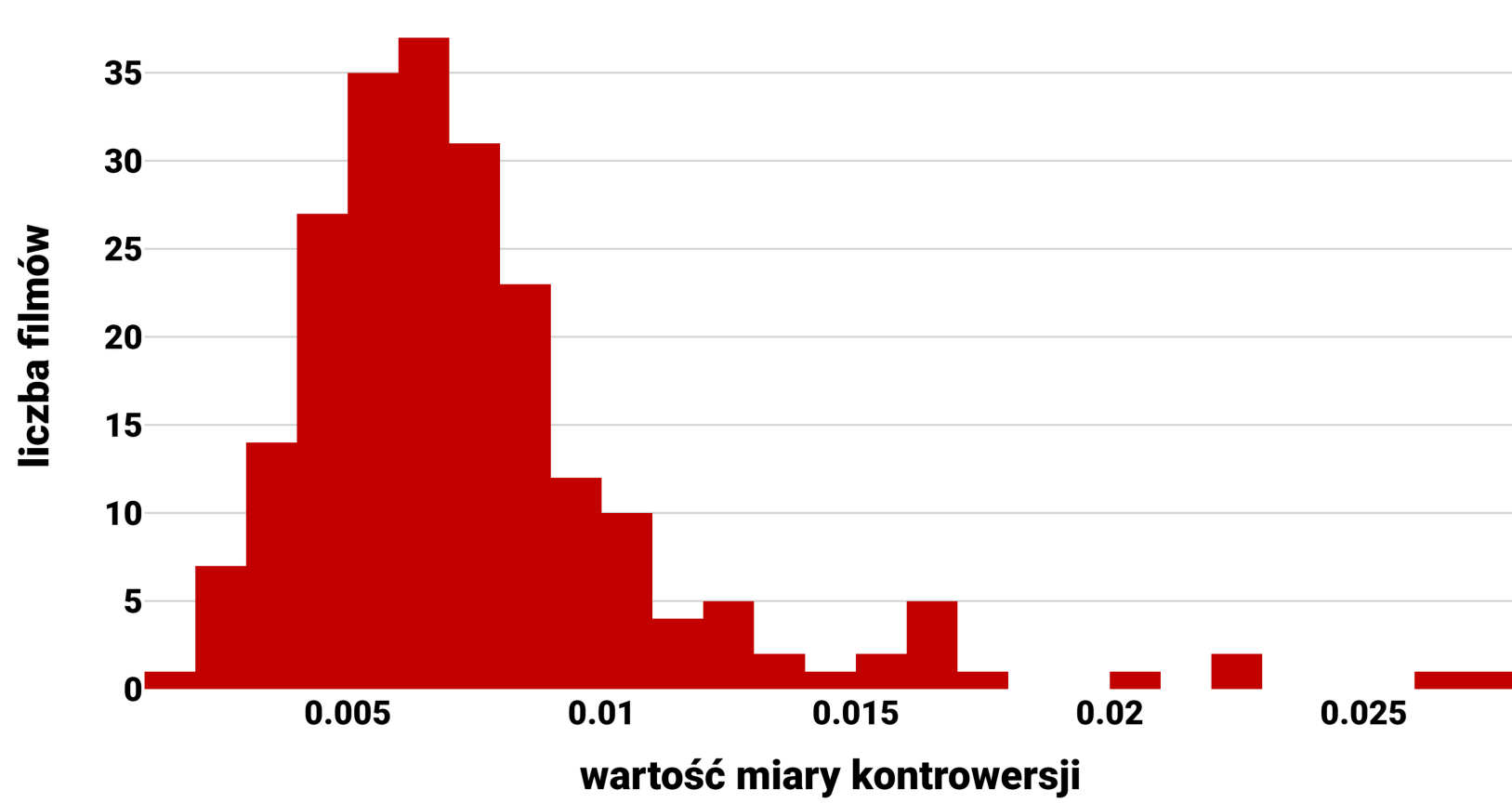
Korelacja pomiędzy liczbą wyświetleń danego filmu, a wydzźwiękiem komentarzy w zależności od kanału waha się od **-0.08** do **0.25**.



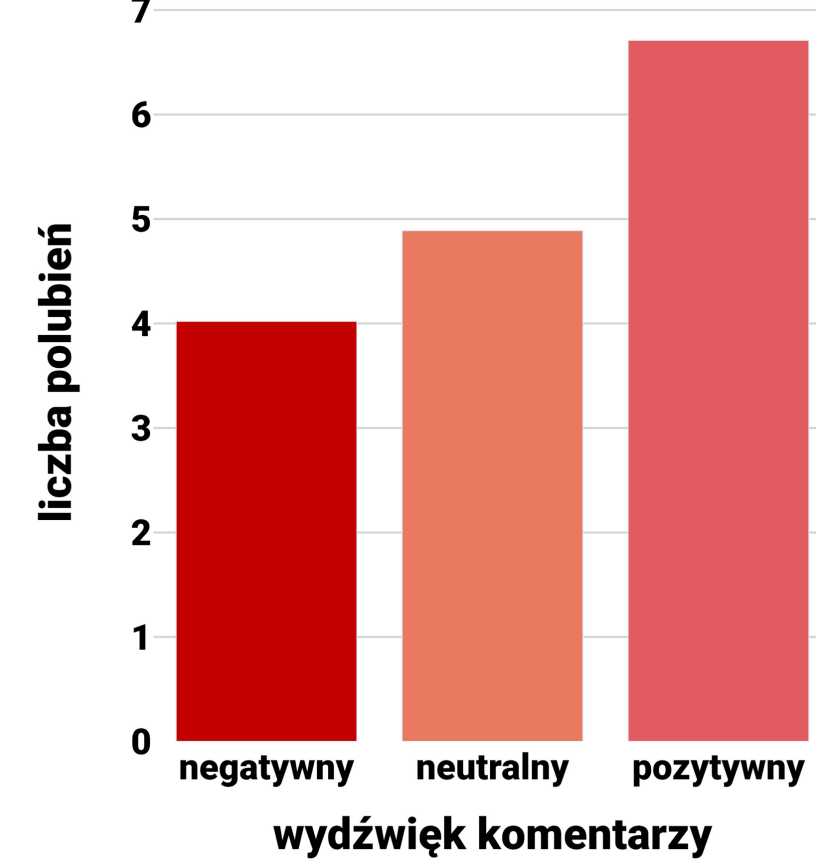
Dane pogrupowane

Korelacja pomiędzy liczbą wyświetleń danego filmu, a wydzźwiękiem komentarzy w zależności od grupy i kanału waha się od **-0.37** do **0.92**.

Histogram poziomu kontrowersji komentarzy

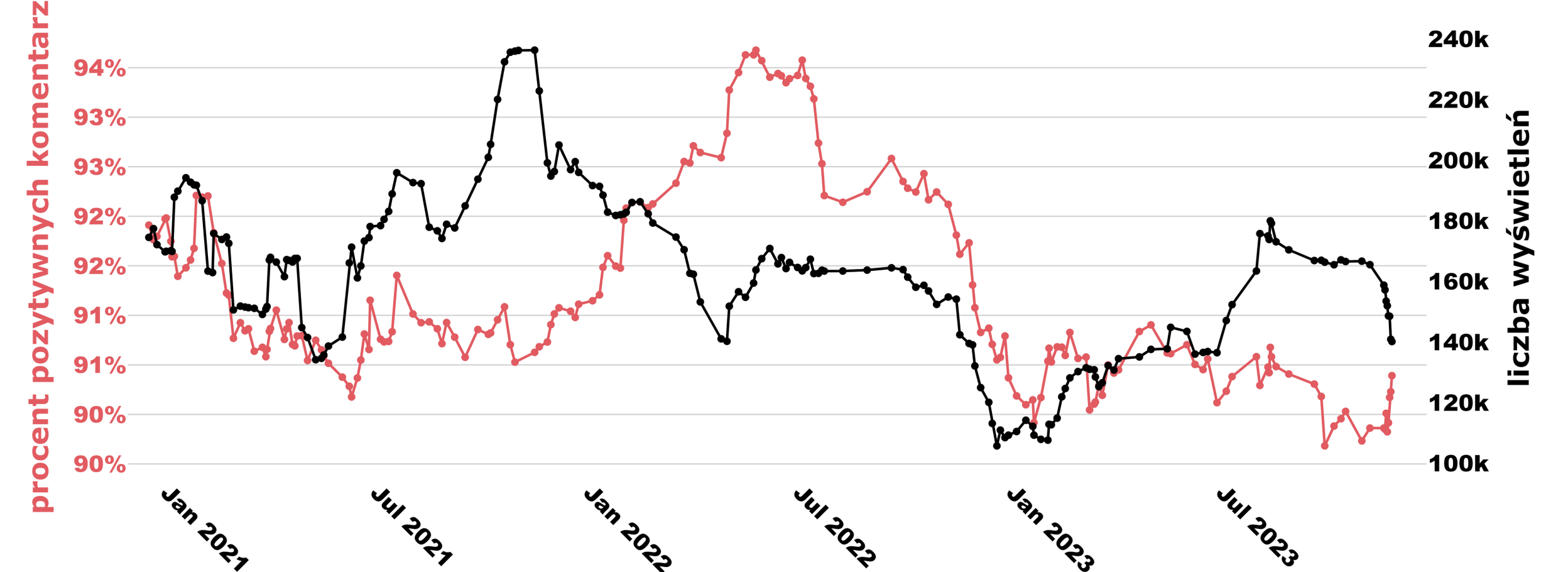


Średnia liczba polubień komentarzy

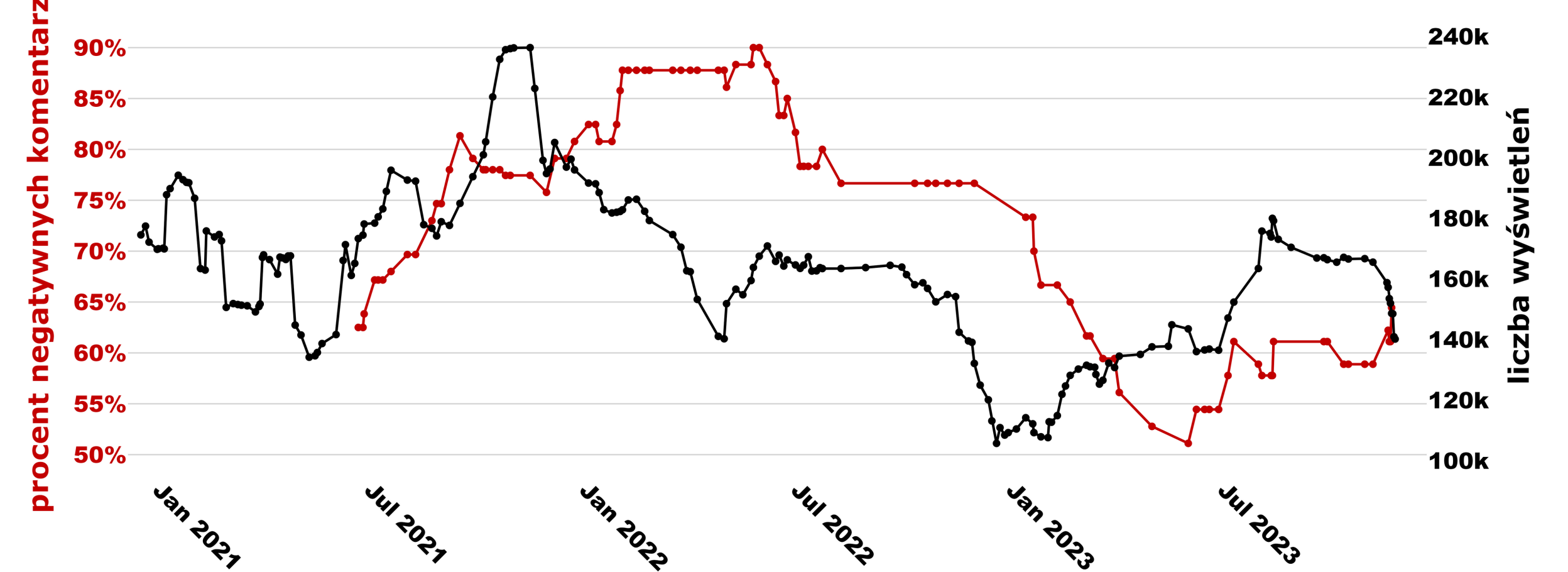


Zestawienie liczby wyświetleń z procentem komentarzy o określonym wydzźwięku

Procent pozytywnych komentarzy w grupie "pozytywnej"



Procent negatywnych komentarzy w grupie "negatywnej"



Wpływ kontrowersji w komentarzach na liczbę wyświetleń

Korelacja pomiędzy policzonym dla każdego filmu poziomem kontrowersji, a liczbą wyświetleń w zależności od kanału waha się w granicach od **-0.111** do **0.224**.

Modelowanie aspektowe

Dla każdego z kanałów przeprowadzono modelowanie aspektowe komentarzy. Zostały wybrane pozytywne i negatywne aspekty, o których pisali komentujący.

Negatywne

Pozytywne



Negatywne i pozytywne aspekty komentarzy grupy najczęściej komentujących osób, która skomentowała 10 najnowszych filmów na przykładowym kanale.

Ciekawe spostrzeżenia

- Rozkład wydzźwięku wszystkich komentarzy może być alternatywną miarą popularności filmu zastępującą liczbę polubień filmu.
- Dla filmów, które mają większą liczbę wyświetleń komentarze są bardziej negatywne niż pozytywne.
- Korelacja pomiędzy liczbą komentarzy o określonym wydzźwięku, a liczbą wyświetleń jest największa gdy brane są pod uwagę tylko komentarze o negatywnym wydzźwięku.
- Udało się znaleźć osobę która skomentowała 84% filmów na jednym z kanałów, co wskazuje na możliwość wykrycia największych fanów kanału.

