

Analiza zbioru “Netflix”

1. Cel pracy

Netflix ma dużą bazę pozycji. Celem pracy jest zbadanie szeroko rozumianego obecnego stanu kinematografii (popularność gatunków, czasu trwania filmów i seriali), w tym zbadanie charakterystycznych cech pozycji dostępnych na netflixie, a dodatkowo przyjrzenie się aspektom takim jak dostępność pozycji w skali świata oraz zbadanie jakości dostępnych filmów i seriali.

2. Opis zbioru danych

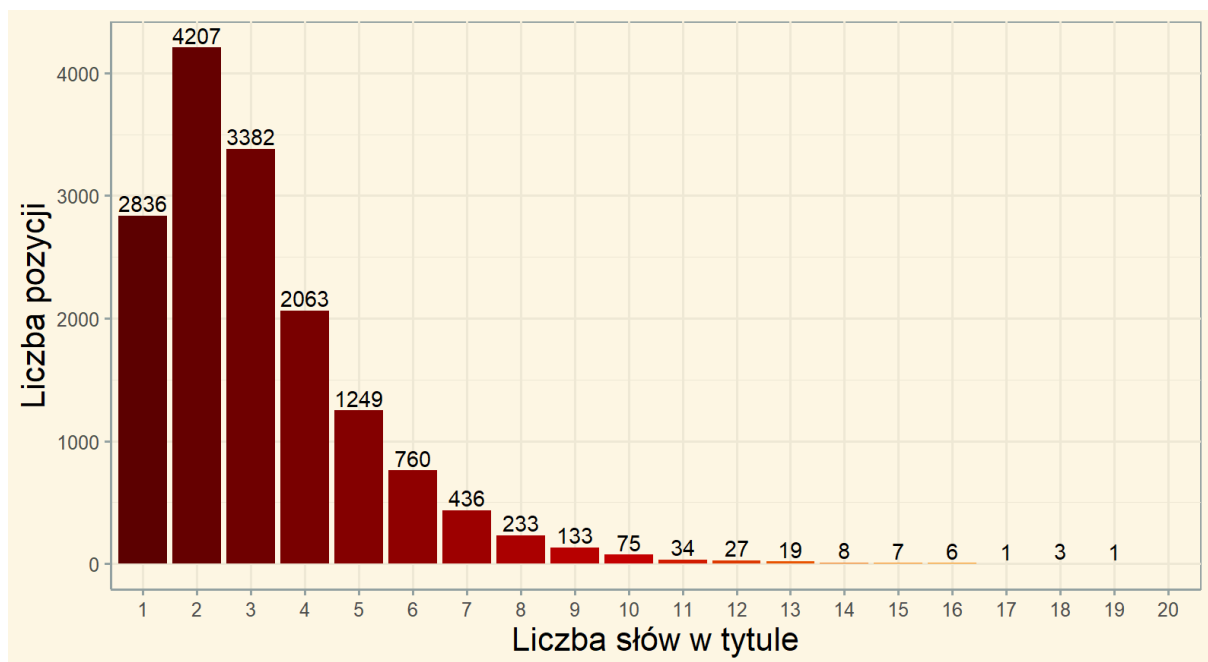
Zbiór danych zawiera 15 480 obserwacji i 11 zmiennych. Dane pochodzą z okresu od 14.04.2015 do 04.03.2021. Zawierają informację z serwisu Netflix oraz z IMDb. Zbiór został pobrany ze strony kaggle.pl (zbiór “*Latest Netflix data with 26+ joined attributes*”) z którego zostały wybrane zmienne uznane za istotne.

Zbiór zawiera następujące zmienne :

1. Tytuł: Podaje nazwę filmu/serialu
2. Serial/Film: Wskazuje, czy ta pozycja to serial telewizyjny czy film
3. Gatunki: Opisuje gatunek filmu/serialu
4. Tagi: Lista tagów w określająca film/serial
5. Języki: Wskazuje, w jakim języku film/serial jest dostępny
6. Dostęp w krajach: Podaje, w jakim kraju ten film/serial jest dostępny
7. Czas trwania: Wskazuje czas trwania filmu lub odcinka serialu
8. Liczba otrzymanych nagród: Mówi nam, ile nagród otrzymał ten film/serial
9. Liczba nominacji do nagród: Mówi nam, do ilu nagród ten film/serial został nominowany
10. Data premiery na Netflixie: Data premiery na Netflixie
11. Ocena IMDb: Ocena przyznana filmowi/serialowi przez IMDb

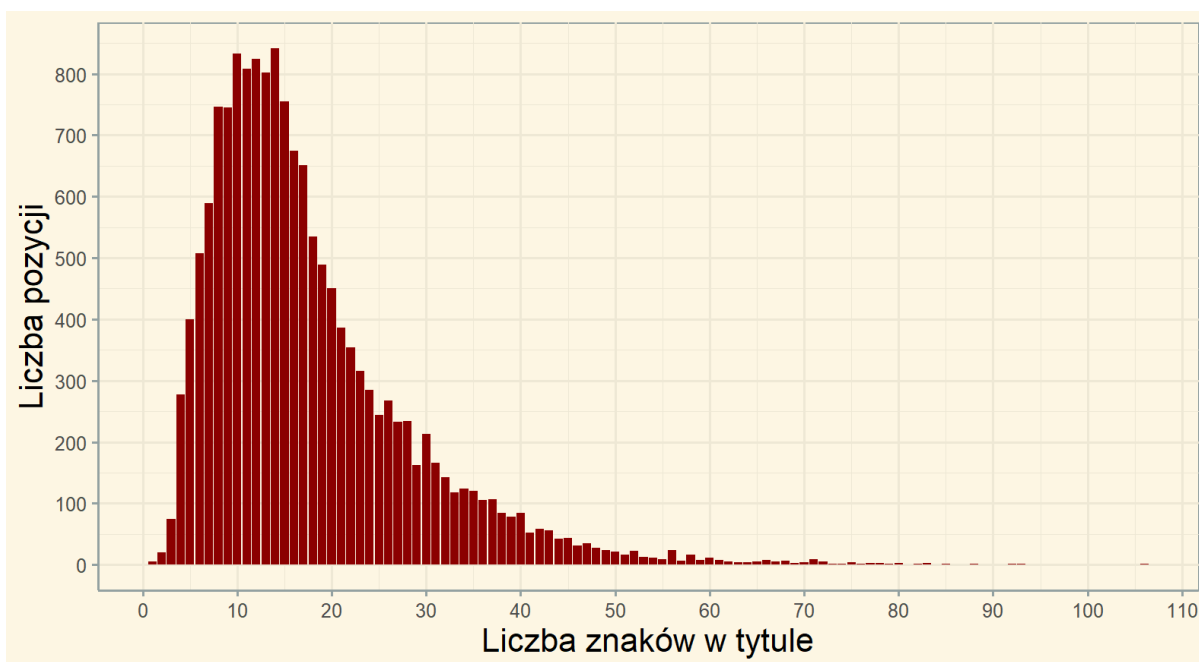
3. Analiza danych

3.1. TYTUŁ



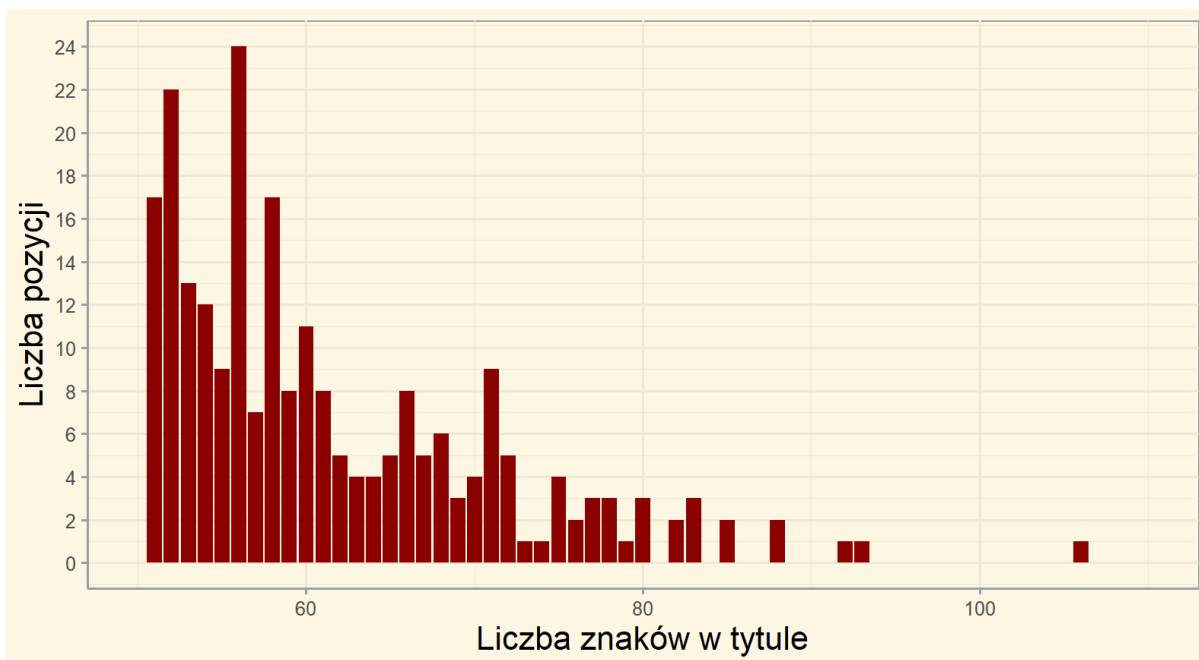
Wykres 1. Liczba słów w tytułach a liczba pozycji

Widać, że przeważają tytuły krótkie. Większość filmów stanowią filmy o liczbie słów od 1 do 4. Najwięcej występuje filmów dwuwyrazowych. Występują nawet filmy, które mają powyżej 12 słów.



Wykres 2a. Liczba znaków w tytule a liczba pozycji

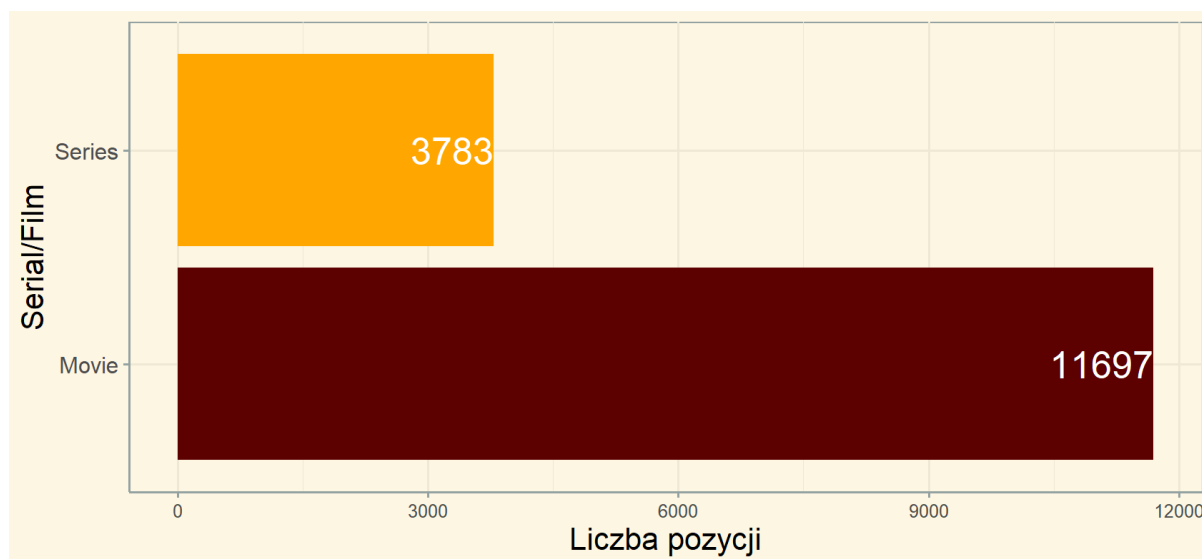
Tytuły zawierają najczęściej od 10 do 15 znaków. Wraz z większą ilością znaków tendencja jest wyraźnie spadkowa. W celu przyjrzenia się ile pozycji występuje przy liczbie znaków 50 zostało zrobione przybliżenie.



Wykres 2b. Liczba znaków w tytule a liczba pozycji

Widać, że pozycje o dłuższych tytułach występują znacznie rzadziej i nie przekraczają 120 znaków.

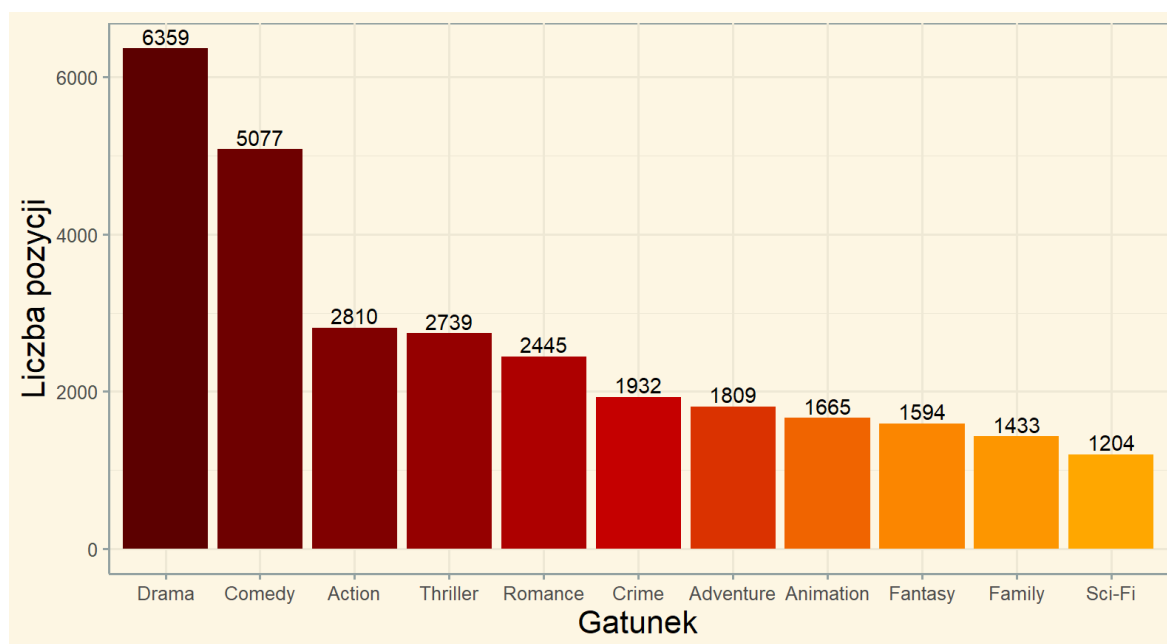
3.2. LICZEBNOŚĆ POZYCJI



Wykres 3. Liczba pozycji serialu i filmu

Widać, że mimo rosnącej popularności seriali w ostatnich czasach, filmy są zdecydowanie popularniejsze od seriali - występują 3 razy częściej.

3.3. GATUNKI



Wykres 3. Popularność gatunków

Najpopularniejszym gatunkiem jest dramat (6 359), następnie komedia (5 077). Kolejne gatunki takie jak filmy akcji czy thrillery występują już w podobnym do siebie rzędzie wielkości, znacznie rzadziej w stosunku do dramatu i komedii.

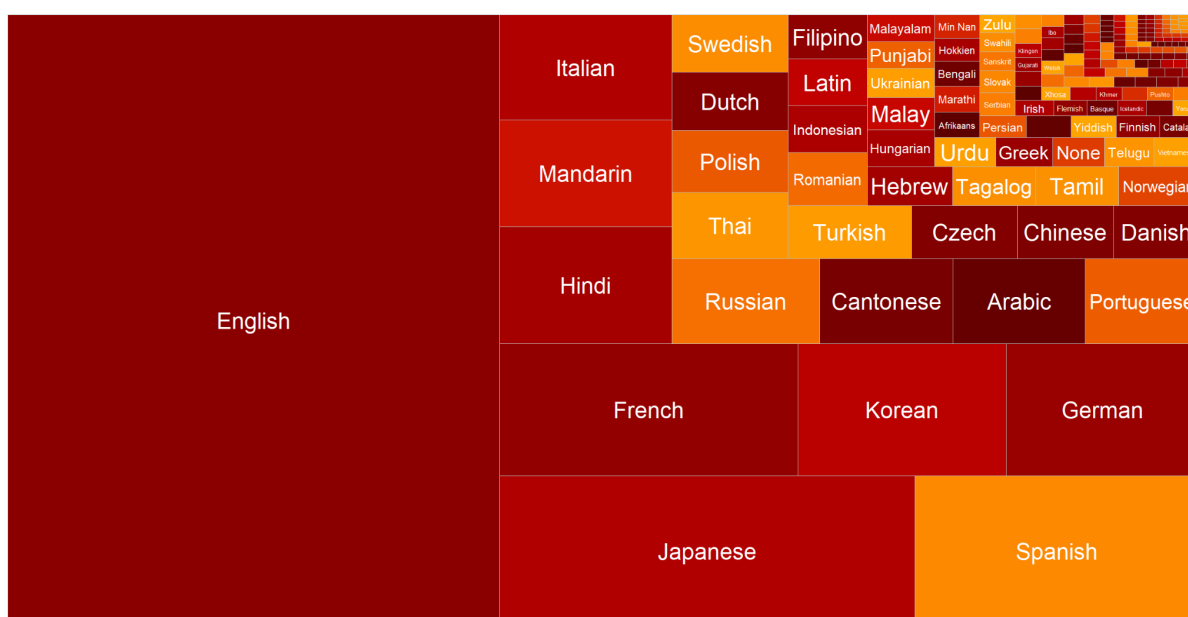
3.4. TAGI



Wykres 4. Częstotliwość występowania słów w Tagach

Na Wordcloud zostały zilustrowane najczęściej występujące słowa w Tagach - słowach kluczach, które pomagają identyfikować filmy i serie. Widać, że zdecydowanie najczęściej występuje '&' co się może wiązać z popularnością tytułów o dwóch słowach. Następnie wybijają się 'Anime', której popularność widać bardzo w występowaniu języków w bazie Netflix.

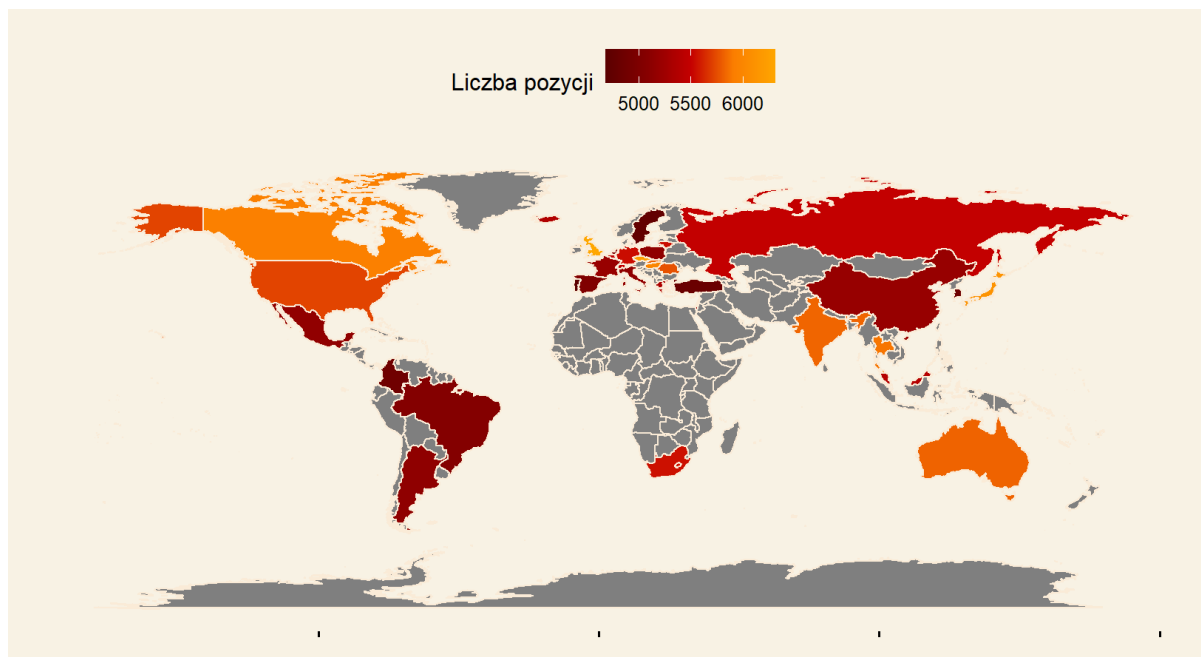
3.5. JĘZYKI



Wykres 5. Dostępność języków filmów i seriali

W tym zbiorze najbardziej dostępnym językiem wśród filmów i seriali jest angielski, następnie japoński - może to wynikać z anime, której popularność było widać w tagach.

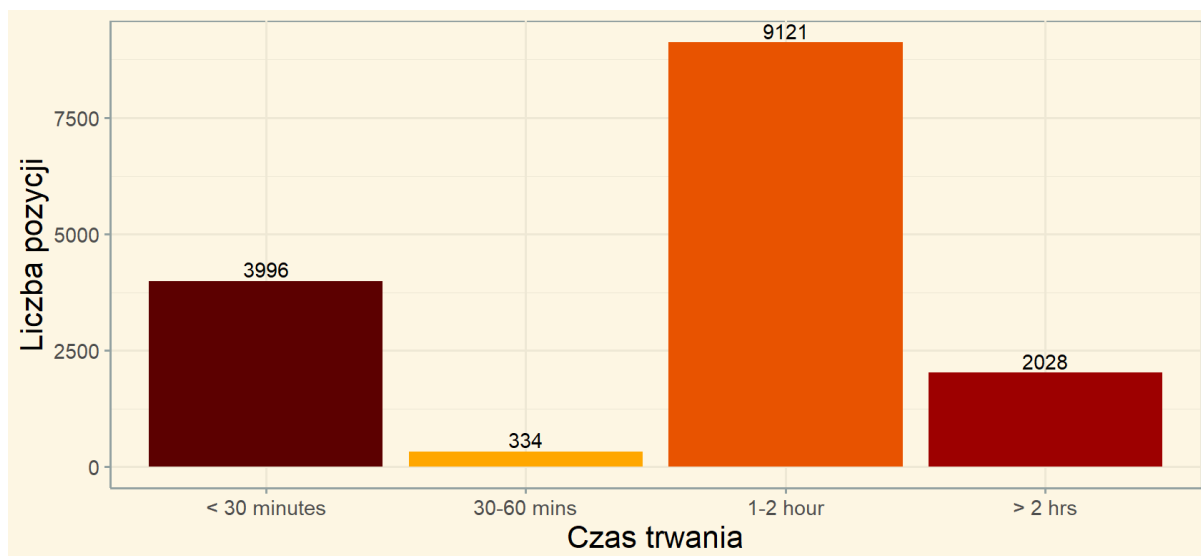
3.6. DOSTĘPNOŚĆ NA ŚWIECIE



Wykres 6. Dostępność na świecie

Widać, że najwięcej dostępnych pozycji jest w Japonii, USA, Południowej Korei, a najmniej w Ameryce Południowej, Szwecji. Na szaro zaznaczone są kraje dla których danych brak w tym zbiorze.

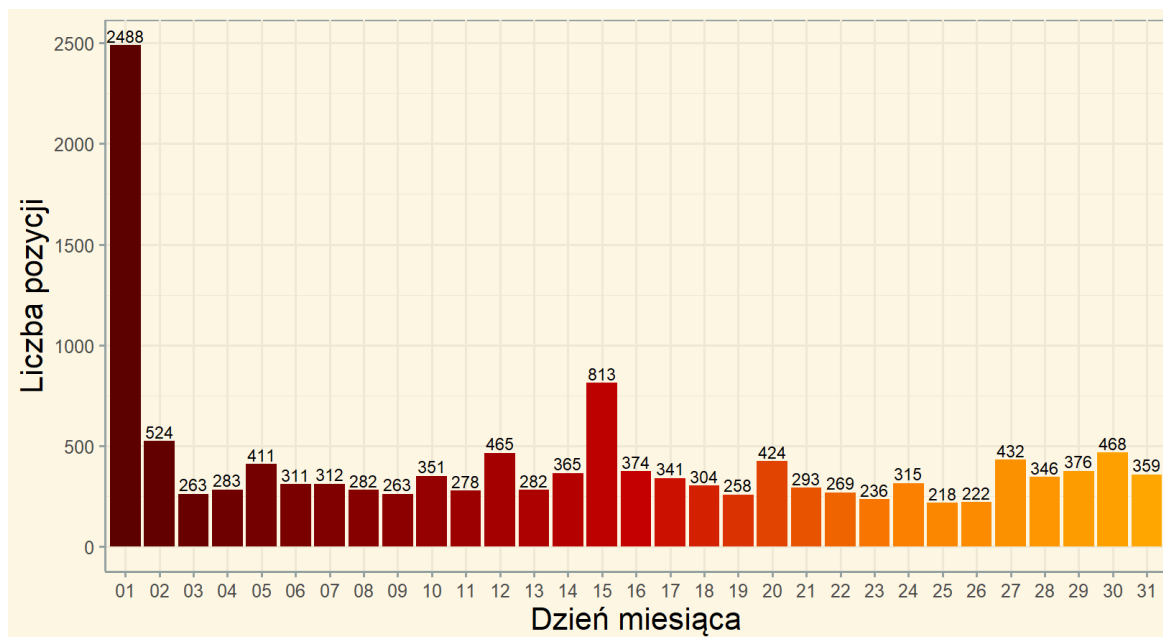
3.7. CZAS TRWANIA



Wykres 7. Czas trwania a liczba pozycji

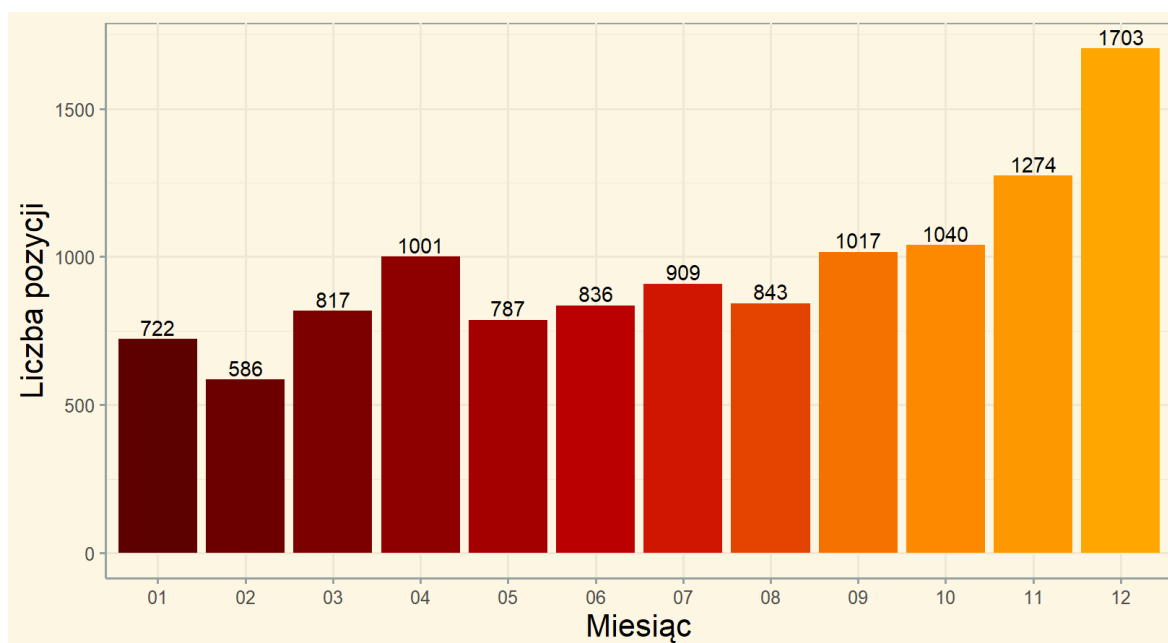
Zdecydowana większość pozycji trwa 1 - 2h, co jest typowe dla filmów. Znajduje to odzwierciedlenie w liczbie pozycji filmów w tej bazie. Liczba pozycji dla czasów w przedziale charakterystycznym dla seriali występują w mniejszej ilości.

3.8. DATA PREMIERY NA NETFLIXIE



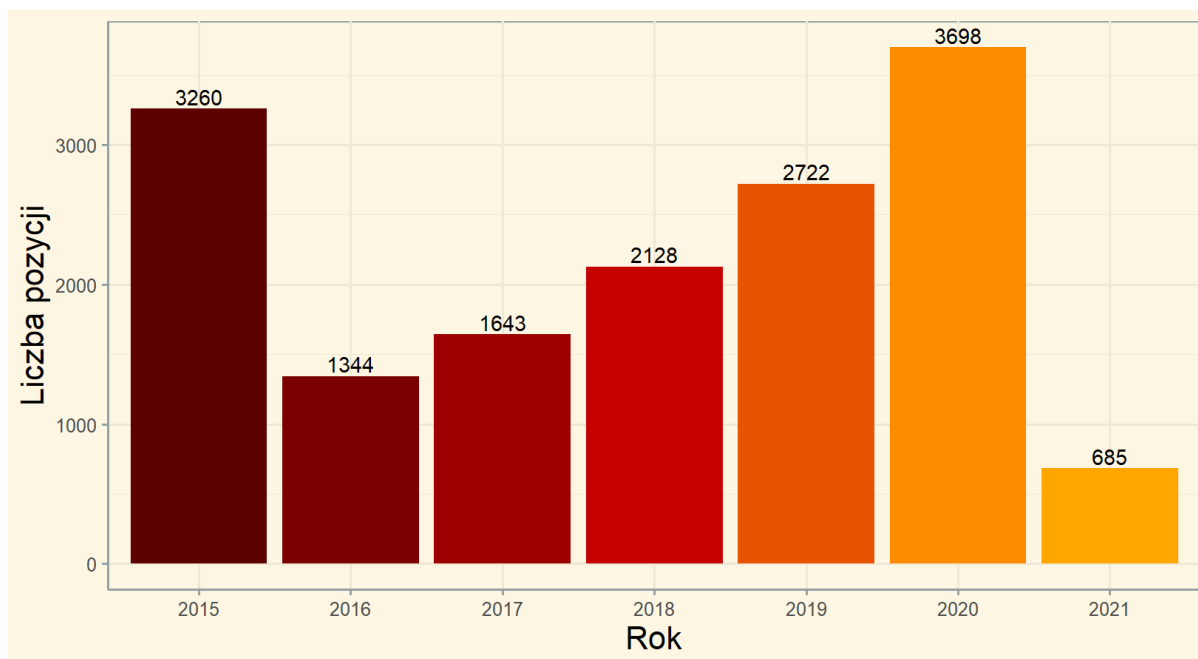
Wykres 8. Ilość premier w poszczególnych dniach miesiąca

Na wykresie zostały zaprezentowane dane dla przedziału czasu od 01.05.2015 do 28.02.2021. Można zauważyć, że na Netflixie premiery są planowane pierwszego dnia miesiąca, co może wynikać z celów promocyjnych. Kolejna wyższa wartość to połowa miesiąca - 15 dzień. Dla pozostałych dat rząd wielkości jest podobny.



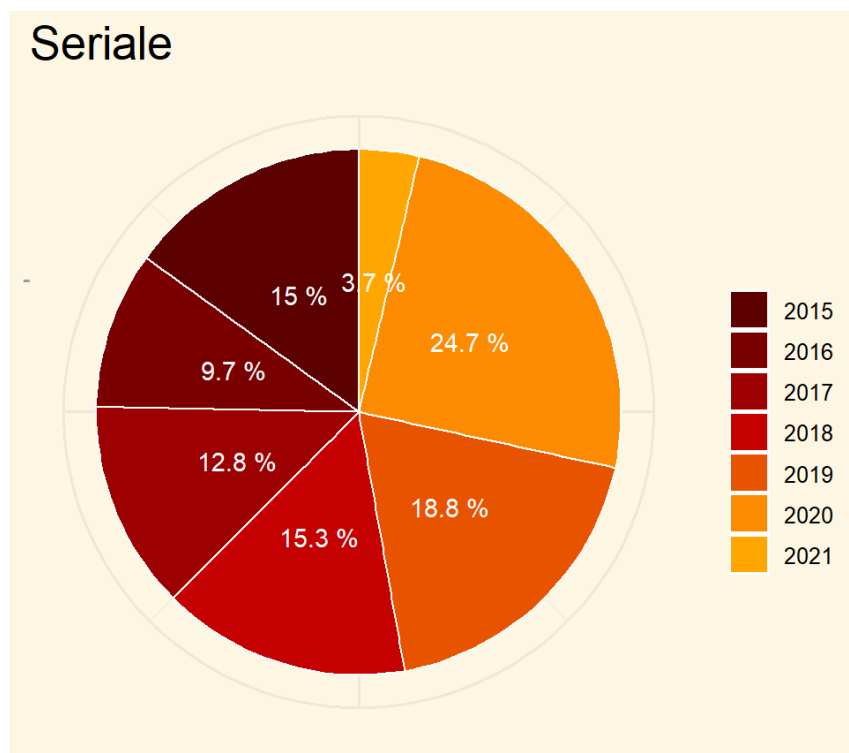
Wykres 9. Ilość premier pozycji w miesiącu

Na podanym wykresie zostały zaprezentowane dane dla dat od 01.01.2016 do 31.12.2020. Od września do grudnia liczba wypuszczanych pozycji rośnie. Widać, że najwięcej pozycji jest wypuszczanych w grudniu. Może to wynikać z tego, że poza klasycznymi pozycjami, w grudniu również wypuszczane są filmy o tematyce świątecznej.



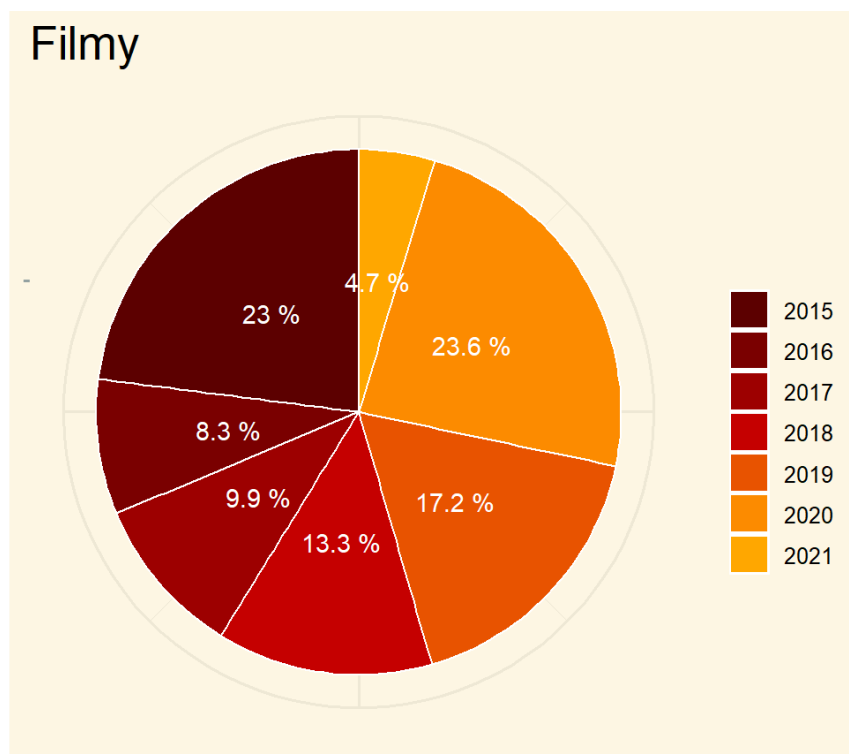
Wykres 10. Łączna ilość premier w latach

Na wykresie jest zaprezentowana liczba pozycji (filmów, jak i seriali) wypuszczanych przez Netflix dla poszczególnych lat. Wyraźny spadek występujący w 2021 jest ze względu na to, że posiadane dane nie obejmują całego roku - są one jedynie do połowy marca. Między 2015, a 2016 widoczny jest nagły spadek - prawdopodobnie wiąże się to z tym, że w 2015 serwis zaczął streamingować w nowych krajach, więc w tym roku pojawiło się wiele nowych pozycji dostępnych dla tych krajów.



Wykres 11. Premiery seriali w podziale na lata

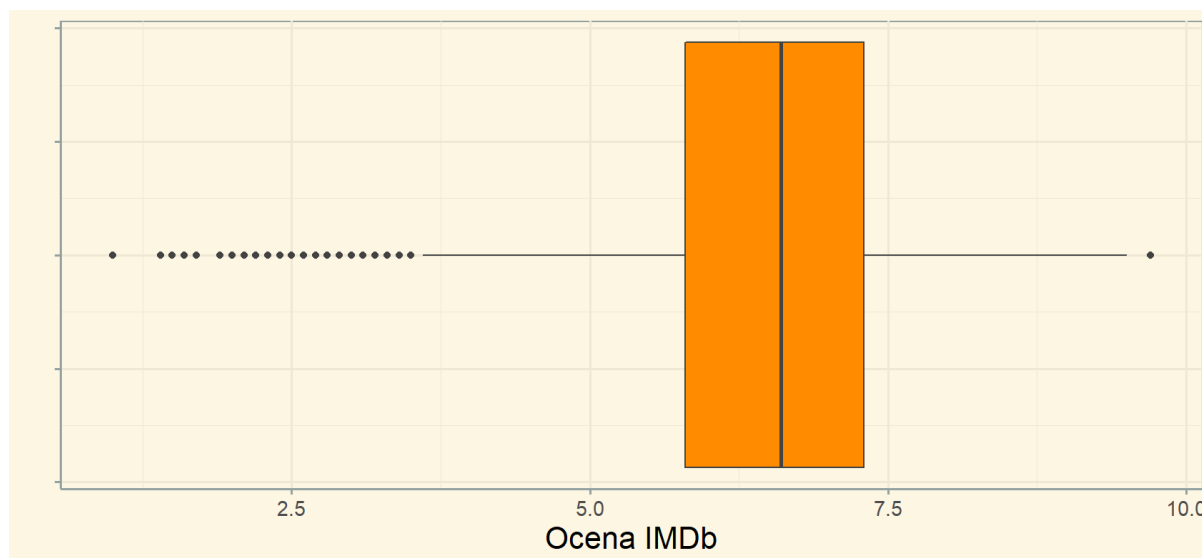
Na wykresie widać, że seriale od 2018 roku stają się coraz bardziej popularne. Rok 2021 nie jest reprezentatywny, dane są tylko do początku marca.



Wykres 12. Premiery filmów w podziale na lata

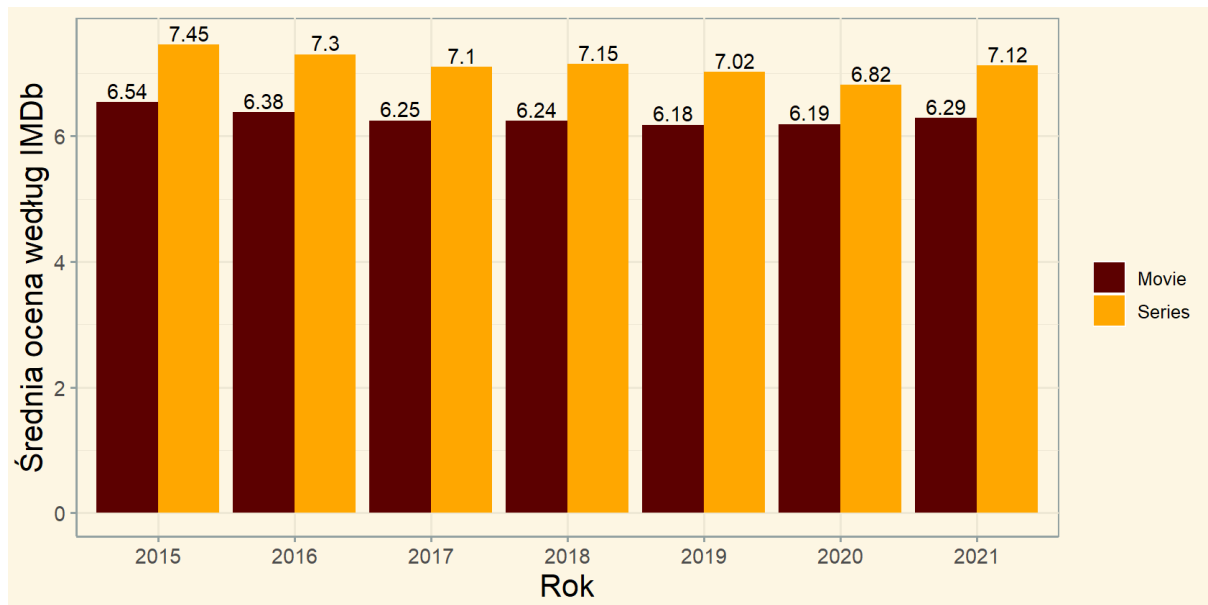
Widać wyraźny spadek w latach 2017, 2018 w stosunku do roku 2015. Jednakże w latach 2019 i 2020 trend jest rosnący i w 2020 jest na poziomie roku 2015. Rok 2021 nie jest reprezentatywny, dane są tylko do początku marca.

3.9. OCENA IMDB



Wykres 13. Przedział Ocen IMDB

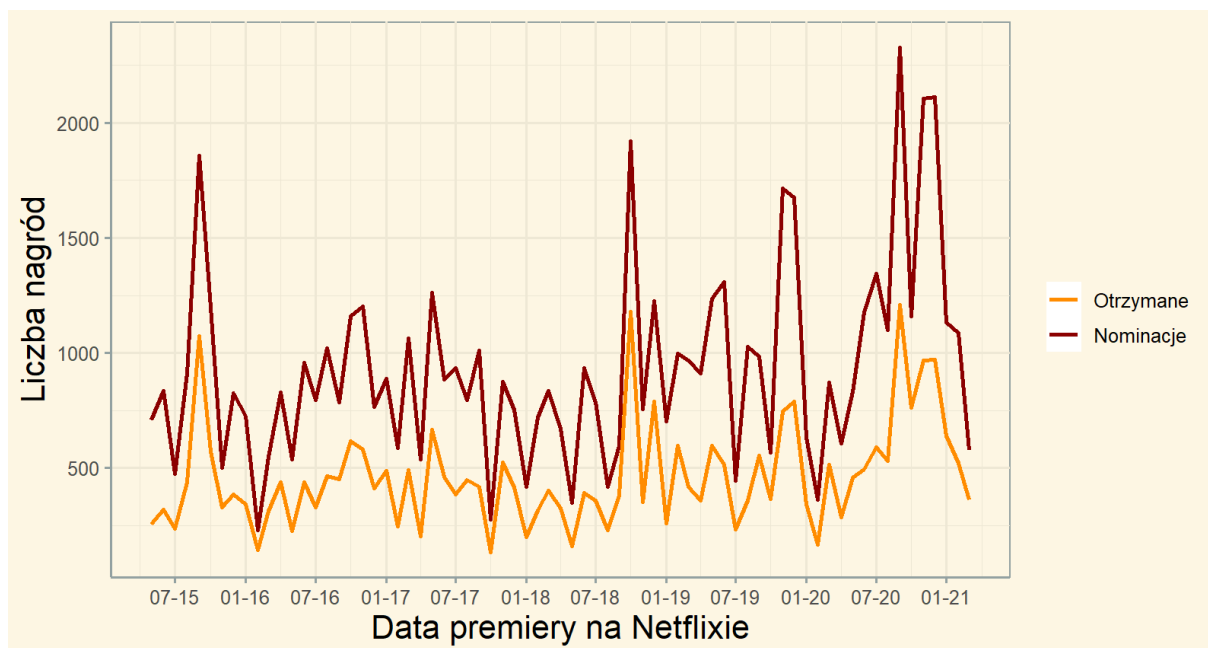
Na wykresie widać, że pojedyncze niewiele pozycje zbliża się do 10 punktów. Tylko połowa filmów ma ocenę powyżej 7. Pojawiają się filmy, które mają ocenę poniżej 2.5.



Wykres 14. Średnia ocen w podziale na lata dla poszczególnych filmów i seriali

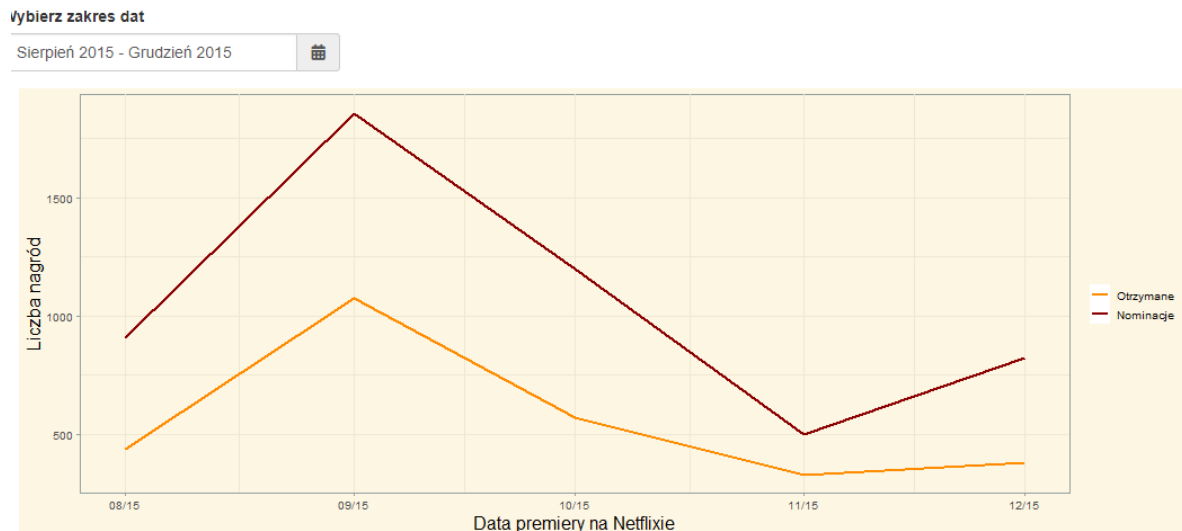
Na przestrzeni lat widać, że pozycje które są dostępne na netflixie oceniane są na podobnym poziomie, w tym samym rzędzie wielkości. Przy czym seriale dla każdego poszczególnego roku były oceniane wyżej niż film.

3.10. NAGRODY



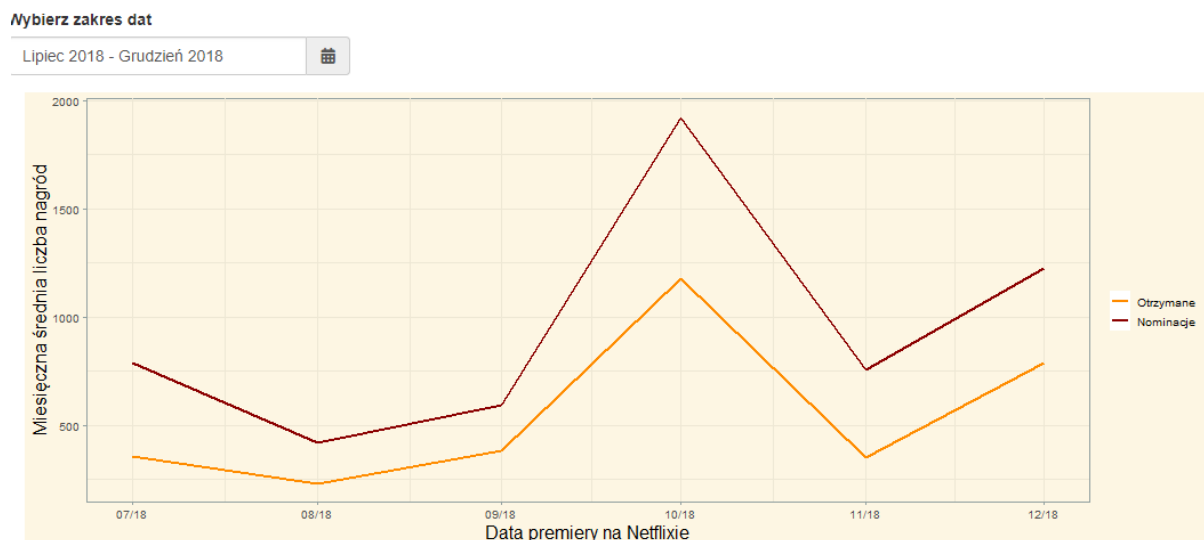
Wykres 15. Liczba nagród otrzymanych i nominowanych a data premiery na Netflixie

Na wykresie zostały zaprezentowane nominacje nagród, wraz z otrzymanymi nagrodami. Największa ilość nominowanych, jak i otrzymanych nagród przypada na okres premier wskazanych na poniższych wykresach (wykresy zostały wykonane za pomocą biblioteki Shiny):



Wykres 15a. Liczba nagród otrzymanych i nominowanych a data premiery na Netflixie

We wrześniu 2015 otrzymanych było około 1080 nagród a nominowanych około 1850.

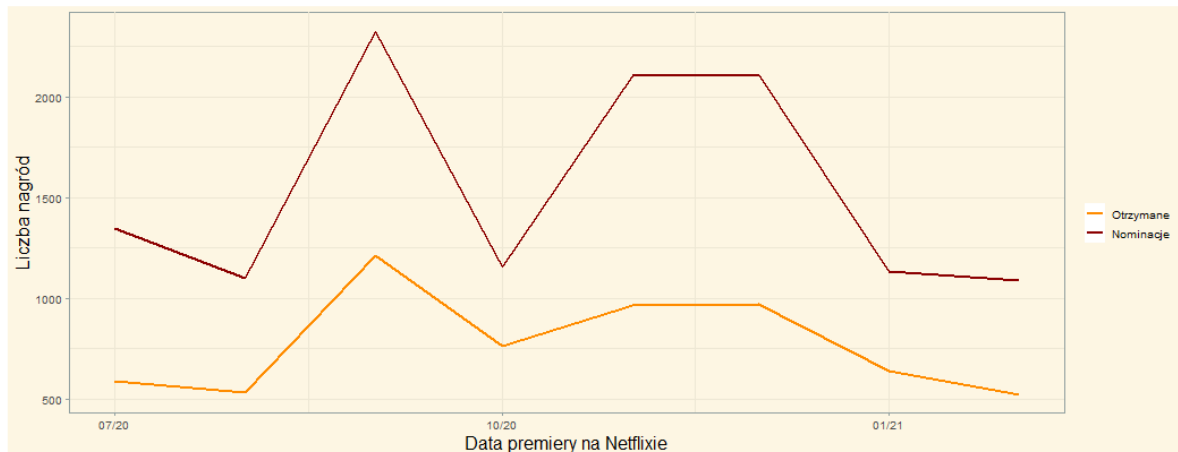


Wykres 15b. Liczba nagród otrzymanych i nominowanych a data premiery na Netflixie

W październiku 2018 średnio otrzymanych około 800, nominowanych około 1200.

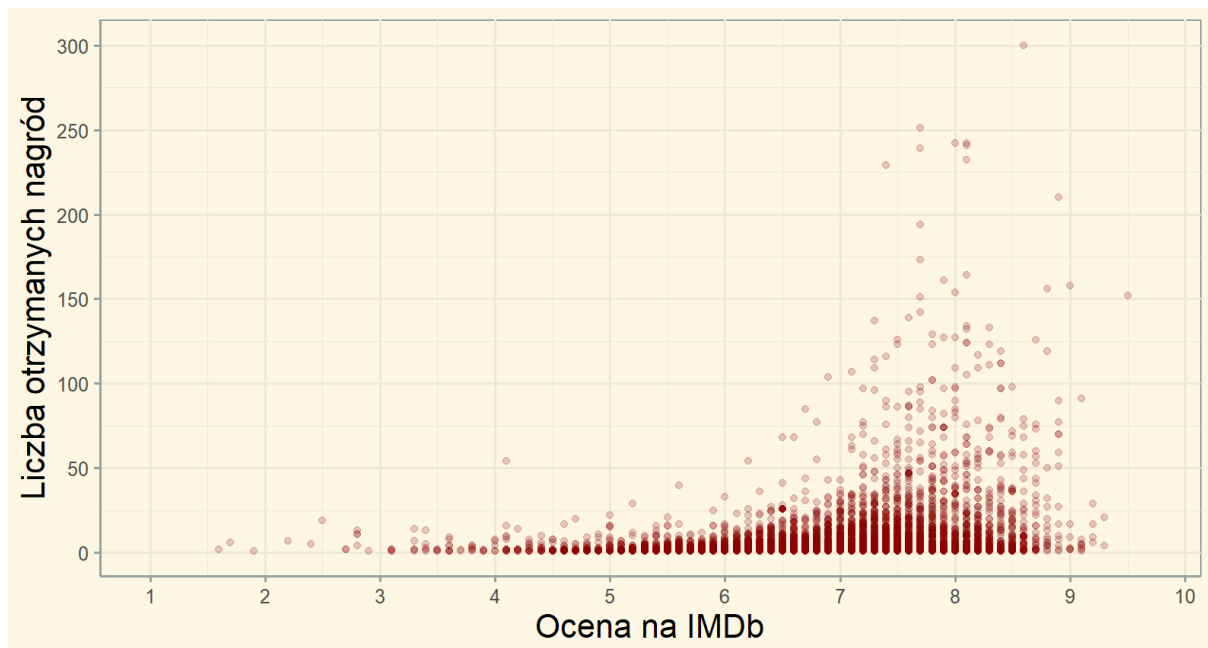
Wybierz zakres dat

Lipiec 2020 - Luty 2021



Wykres 15c. Liczba nagród otrzymanych i nominowanych a data premiery na Netflixie

We wrześniu 2020 otrzymanych nagród zostało około 1200, a nominowanych około 2300. Wpływ ma na to fakt, że w 2010 zostało najwięcej pozycji opublikowanych (co pokazuje wykres 10).



Wykres 16. Ocena IMDB a liczba otrzymanych nagród

Na wykresie została sprawdzona czy głosy użytkowników na serwisie IMDB odzwierciedlają ilość nagród otrzymanych przez filmy i seriale. Od oceny 7,5 widać znaczny wzrost ilości otrzymywanych nagród. Widać też, że pomimo tego, że część pozycji otrzymała niską ilość punktów np. 4-5, nagrody zostały im przyznane.

4. Podsumowanie

Większość pozycji dostępnych stanowią filmy. Przeciętny czas trwania pozycji wynosi między 1 a 2 h. Do najpopularniejszych gatunków należy dramat i komedia. Najbardziej popularnym językiem jest angielski. Drugim w kolejności jest język japoński, ze względu na Anime, co było też widać w analizie tagów. Najwięcej dostępnych pozycji (filmów i seriali) jest w Japonii, USA, Południowej Korei, a najmniej w Ameryce Południowej, Szwecji. Tytuły zawierają zazwyczaj między 1 a 4 słów i zawierają najczęściej od 10 do 15 znaków. Premier możemy spodziewać się pierwszego dnia miesiąca, ewentualnie 15 dnia miesiąca, oraz najwięcej ich jest w grudniu. Ocena pozycji dostępnych na Netflixie przez użytkowników IMdb utrzymuje się na przestrzeni lat na podobnym poziomie, przy czym seriale były oceniane zawsze wyżej niż filmy. Od oceny 7,5 widać znaczny wzrost ilości otrzymywanych nagród. Największa liczba nagród została przyznana filmom opublikowanym na netflixie we wrześniu 2020 roku. W tym roku też miała miejsce największa ilość premier.