

IMDB FILMS



Wstęp

Celem niniejszej analizy jest zbadanie danych dotyczących najlepszych filmów dostępnych w serwisie IMDB, aby zrozumieć jakie czynniki mogą wpływać na wysokie oceny filmów. IMDB, jako jedna z najpopularniejszych baz danych filmowych na świecie, gromadzi informacje na temat tysięcy filmów, w tym szczegóły dotyczące ich fabuły, reżyserów, aktorów, gatunków oraz ocen wystawianych przez użytkowników.

W niniejszej analizie skupiamy się na identyfikacji kluczowych zmiennych, które mogą wpływać na oceny filmów. W tym celu wykonano szereg kroków, obejmujących:

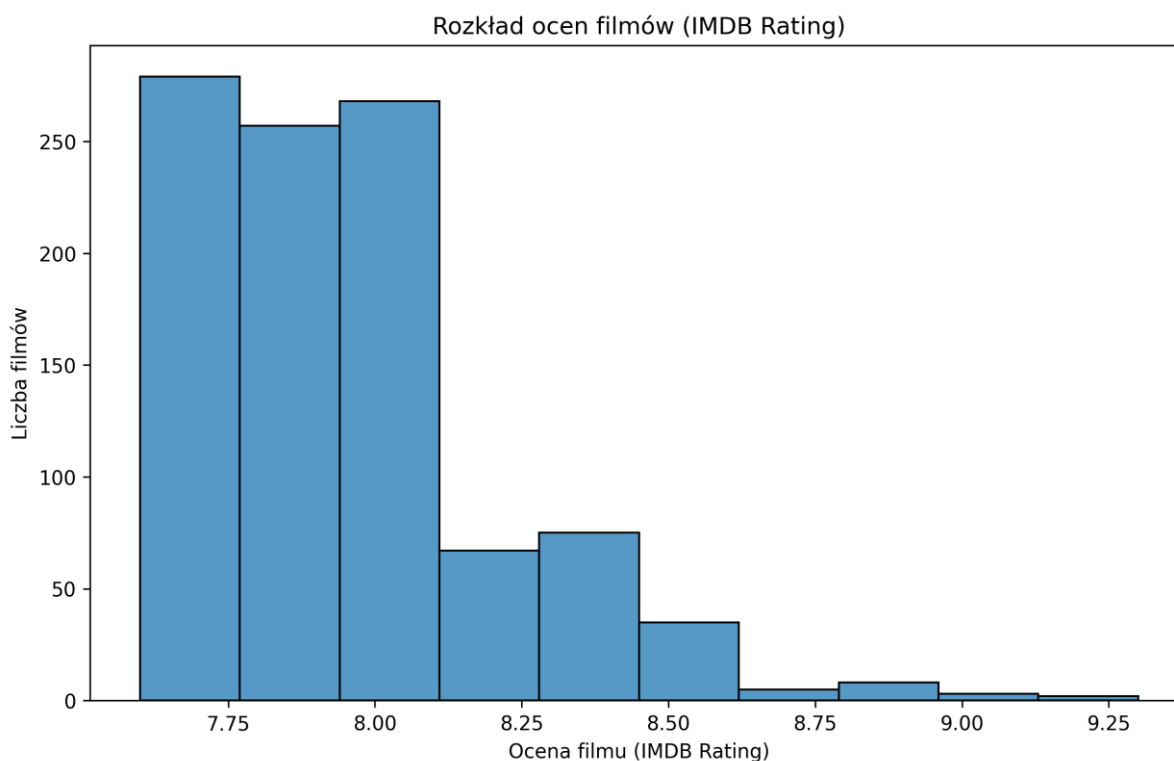
- 1. Przygotowanie danych:** Proces ten obejmował przetwarzanie danych, co pozwoliło na usunięcie niepoprawnych rekordów oraz dostosowanie ich do odpowiedniego formatu. Zastosowano metodę imputacji danych k-nearest neighbors. Użyto narzędzi do analizy danych w Pythonie, takich jak Pandas i Matplotlib, aby ułatwić te operacje.
- 2. Wizualizacja kluczowych zmiennych:** Wizualizacja danych odgrywa kluczową rolę w analizie, umożliwiając lepsze zrozumienie złożonych relacji między różnymi zmiennymi. W tej części analizy wykorzystano biblioteki Matplotlib i Seaborn do stworzenia wykresów, które ilustrują rozkład ocen filmów w zależności od ich gatunku, reżyserów oraz obsady.
- 3. Analiza wpływu reżyserów, aktorów i gatunku filmowego na ocenę filmu:** Kluczowym aspektem analizy jest ocena, w jaki sposób różne czynniki, takie jak reżyser, aktorzy oraz przynależność filmów do określonych gatunków, mogą wpływać na otrzymywane oceny. Badania te mają na celu zidentyfikowanie trendów oraz wyciągnięcie wniosków, które mogą pomóc w lepszym zrozumieniu dynamiki sukcesu filmowego.

Analiza została przeprowadzona przy użyciu Pythona, co umożliwiło nie tylko automatyzację procesu, ale również wykorzystanie zaawansowanych narzędzi analitycznych. W rezultacie, raport ten ma na celu przedstawienie najciekawszych zależności i wzorców, które mogą być istotne dla producentów filmowych, analityków oraz pasjonatów kina.

W raporcie tym nie będę skupiał się na dokładnym omówieniu procesu czyszczenia oraz formatowania danych aby nie zaburzyło to czytelności. Aspekty te będą szczegółowo opisane w kodzie.

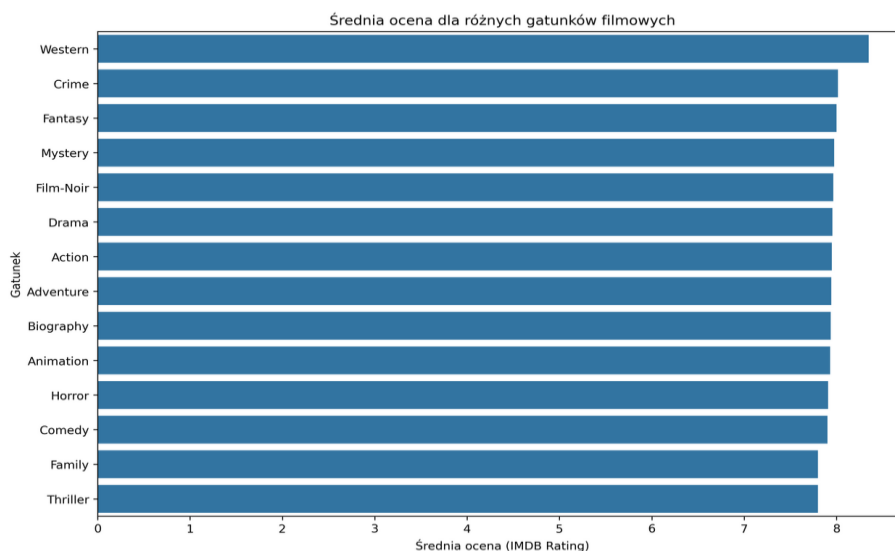
Wizualizacja danych

Rozkład ocen filmów

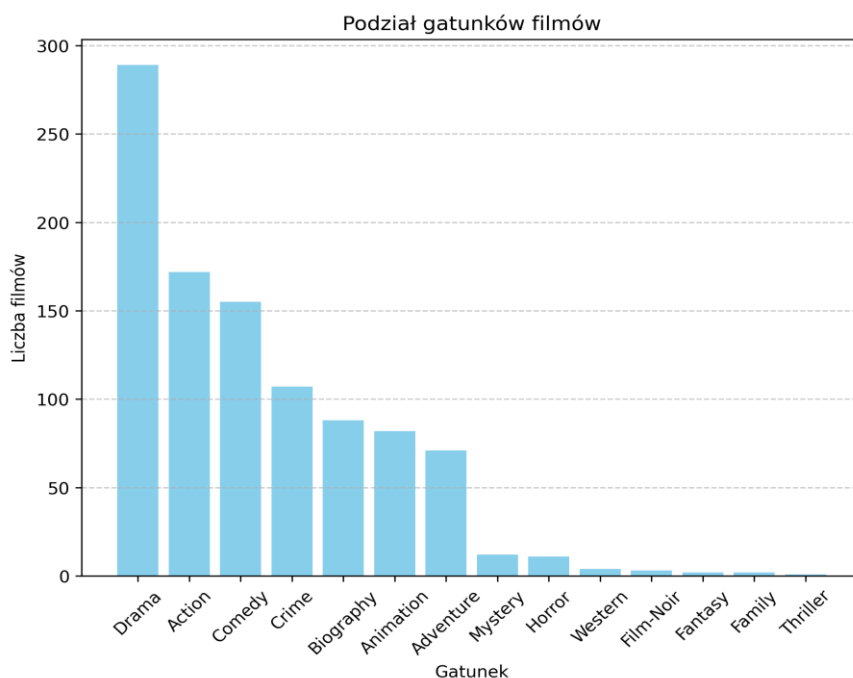


Choć istnieje wiele produkcji, które zdobywają uznanie krytyków i publiczności, większość z nich oscyluje wokół ocen 8 lub niższych, nawet wśród 1000 najlepiej ocenianych filmów. Filmy najczęściej oceniane są przez wielu różnych krytyków o różnorodnym guście filmowym, stąd osiągnięcie średniej ocen bliższej 10 jest nie lada wyzwaniem.

Średnie oceny dla różnych gatunków filmowych

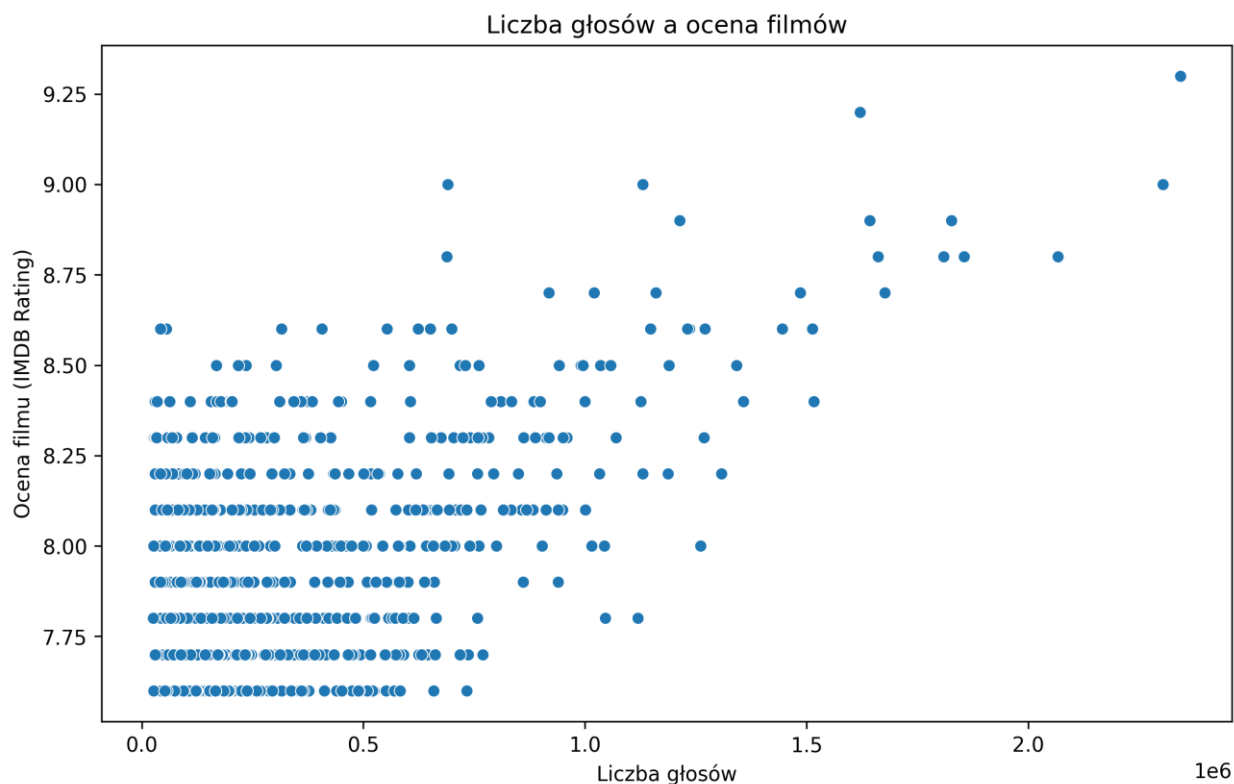


Wydawać by się mogło, że wpływ gatunku na średnią ocenę w IMDB Rating będzie znaczący i bardzo wyraźnie zauważalny. Okazuje się, że poza westernem, który znacząco wyprzedził konkurencję, średnie oceny nie różnią się aż tak bardzo. Po analizie okazało się jednak, że niektóre z gatunków filmowych występują o wiele częściej w naszym zbiorze w porównaniu do innych. To powoduje że wyniki nie mogą być do końca reprezentatywne. Z tego powodu lepszym sposobem na wybranie gatunku filmu na wieczór jest spojrzenie jak dużo poszczególnych gatunków znalazło się w zbiorze 1000 najlepszych filmów według IMDB.



Króluje tutaj dramat, film akcji oraz komedia. Faktycznie, myśląc o klasykach kina bierzemy pod uwagę filmy głównie z tych kategorii.

Szukanie perełek

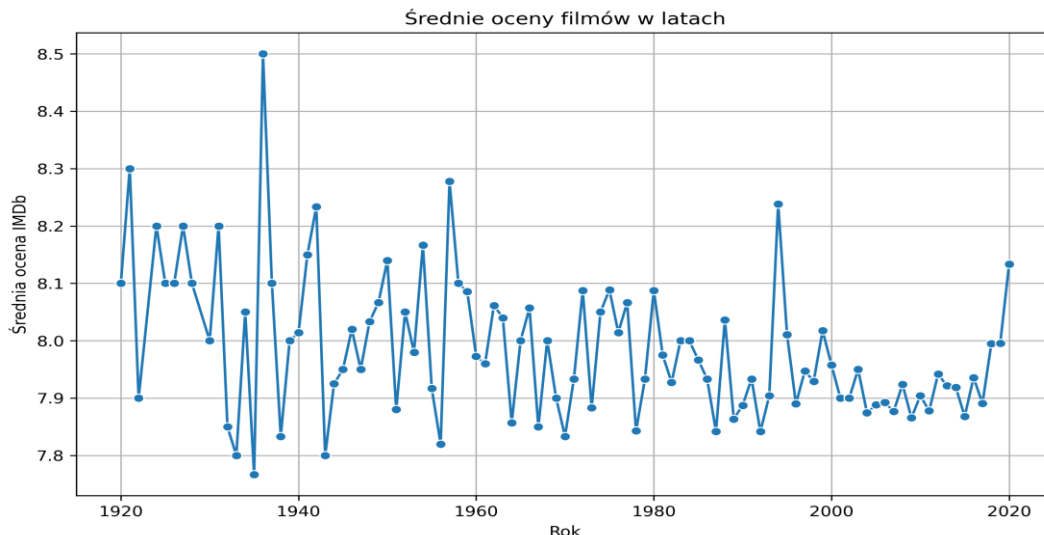


Jak widać występuje pozytywna zależność między liczbą oddanych głosów na danych film a jego oceną. Nie jest to jednak zaskakujące. W trakcie analizy stwierdziłem, że ciekawym będzie znalezienie filmów “perełek”, które uzyskały wysokie oceny równocześnie nie uzyskując tak dużo głosów.

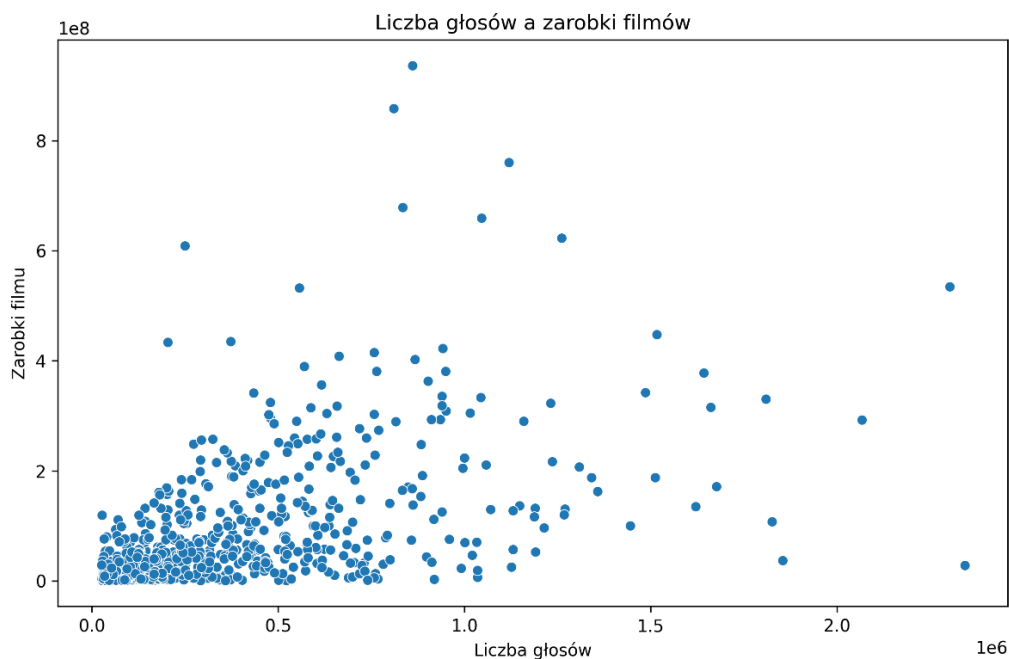
- Hamilton
- Soorarai Pottru
- Seppuku
- Shichinin no samurai
- It's a Wonderful Life
- Nuovo Cinema Paradiso
- Hotaru no haka
- Once Upon a Time in the West
- Modern Times
- City Lights

Może więc lepiej obejrzeć filmy z tej listy zamiast po raz kolejny oglądać Ojca Chrzestnego. Obejrzymy coś dobrego, a równocześnie nie na tyle popularnego aby wszyscy o danej produkcji słyszeli. Może będzie to dobry początek do ciekawej dyskusji w gronie znajomych.

Filmy na przestrzeni lat

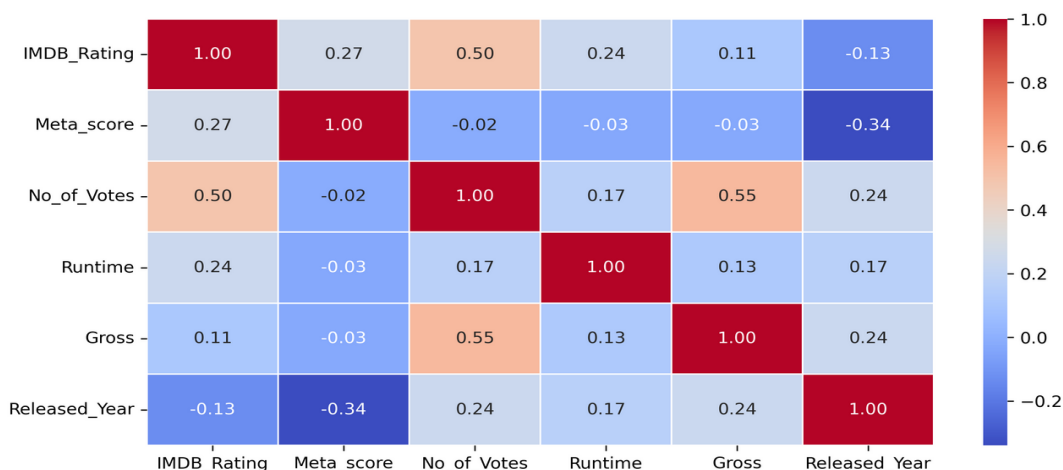


Widzimy delikatny trend spadkowy. Czy wynika to z tego, że filmy w przeszłości były lepsze? Być może tak. Może jednak do dawnych filmów przyciąga nas ich niepowtarzalny klimat. Współczesne produkcje często skupiają się na efektach specjalnych, co może odciągać uwagę od opowieści i postaci. To sprawia, że widzowie tęsknią za dawnym stylem, który często kładł większy nacisk na narrację i rozwój bohaterów. Na poniższym wykresie widzimy również jak mocno przychody filmu skorelowane są z ilością ocen, co może zachęcać reżyserów do tworzenia filmów przyciągających tłumy, a nie koniecznie o bardzo wysokiej jakości produkcji.



Wpływ na ocenę filmu

Macierz korelacji



W celu zobrazowania zależności między konkretnymi zmiennymi stworzyłem macierz korelacji. Czynnikiem najmocniej skorelowanym z oceną IMDB jest liczba głosów. W mniejszym stopniu jest to również wynik metascore, który obrazuje średnią ocenę uzyskaną przez film od różnych krytyków filmowych. Również długość filmu wykazuje słabą ale dodatnią zależność. Bardzo mocna zależność występuje między zarobkami a liczbą głosów, co jest logicznie uzasadnione. Macierz korelacji nie pozwala jednak w tak łatwy sposób poradzić sobie ze zmiennymi kategorycznymi.

Wpływ reżysera

Aby sprawdzić, którzy reżyserzy uzyskują najlepszą średnią ocen uśredniliśmy oceny filmów przez nich stworzonych. W ramach tej analizy rozpatrywaliśmy reżyserów którzy stworzyli więcej niż 3 filmy znajdujące się w czołowej stawce, aby wyłować tych najlepszych.

1. Frank Darabont 8.95
2. Lana Wachowski 8.70
3. Irvin Kershner 8.70
4. Masaki Kobayashi 8.60
5. Fernando Meirelles 8.60
6. George Lucas 8.60
7. Sudha Kongara 8.60
8. Thomas Kail 8.60
9. Roberto Benigni 8.60
10. Milos Forman 8.50

Wśród tej grupy znajdują się znane twarze takie jak George Lucas czy Frank Darabont czyli twórcy filmów "Skazani na Shawshank" oraz "Zielona Milla". Jak widać reżyserzy mają duży wpływ na oceny uzyskiwane przez film.

Wpływ aktorów

Stosując podobne podejście wobec aktorów uzyskujemy ranking średnich ocen uzyskanych przez filmy z udziałem konkretnego aktora. Przedstawmy listę 10 zwycięzców.

- | | |
|---------------------|-----|
| 1. Bob Gunton | 9.3 |
| 2. William Sadler | 9.3 |
| 3. Aaron Eckhart | 9.0 |
| 4. John Travolta | 8.9 |
| 5. Caroline Goodall | 8.9 |
| 6. Aldo Giuffrè | 8.8 |
| 7. Elliot Page | 8.8 |
| 8. Meat Loaf | 8.8 |
| 9. Sean Bean | 8.8 |
| 10. Gary Sinise | 8.8 |

Podsumowanie

Raport ten dostarcza wielu informacji o najczęstszych gatunkach filmowych wśród najlepiej ocenianych filmów. W trakcie analizy udało się wyłuskać nie tylko popularne filmy, ale także te mniej znane, które zyskały wysokie oceny krytyków i widzów, ale jeszcze nie zdobyły szerokiej popularności. To szczególnie interesujące, ponieważ może wskazywać na skryte perełki w kinie, które zasługują na większą uwagę i promocję. Dzięki tej analizie uzyskujemy wgląd w oceny filmów na przestrzeni lat. Z raportu można się również dowiedzieć o współzależnościach między oceną filmu a innymi zmiennymi. Warto również zwrócić uwagę na aktorów i reżyserów, którzy mają największy wpływ na późniejsze oceny filmów. Analizując dane, można zauważyć, że niektóre nazwiska przyciągają uwagę widzów i krytyków, co może przekładać się na wyższe oceny i zainteresowanie produkcją.