| Raccolta e analisi dei requisiti per un progetto di analisi dati |
| --- |
| Netflix e IMDB Dati |

| Sergio Fusar Imperatore | Giuseppe Schillaci |
| --- | --- |

| Epicode Day2 Exe 2 |
| --- |

# Obiettivi del progetto

Si vuole realizzare un’analisi per la Casa produttrice(HBO) che vuole lanciare un nuovo film e/o TV show su Netflix:

* Creare un nuovo film o tv show,
* In che paese crearlo,
* Quando lanciarlo,
* Che genere,
* Quali attori scegliere,
* Quale regista scegliere,
* Durata ideale del film,
* Durata ideale del tv show.

## Attori coinvolti

* Management dell’azienda di produzione cinematografica
* Azienda di consulenza
  + Team Data
    - Data Analyst
    - Data Engineer

## Descrizione del caso di studio

Netflix è una piattaforma di streaming popolare, che offre un vasto catalogo di film e serie, show TV. Lo studio di questo dataset ha come obiettivo quello di guidare la Casa Produttrice (es HBO) cliente per la valutazione di un nuovo film o serie TV in base alle preferenze del pubblico in termini di visualizzazioni e in termini di votazioni su IMDB.

Analizzeremo dei film e tv show maggiormente visti il:

* Il GENERE maggiormente scelto,
* Il CAST maggiormente scelto,
* Il REGISTA maggiormente scelto,
* La Media e varianza della durata dei film
* L’Anno di uscita
* Il VOTO IMDB
* Il paese dove è stato rilasciato
* Il paese dove è stato creato

## Analisi e raccolta dei requisiti

Per fornire le giuste informazioni al management della società il team di data analysis ha pensato di utilizzare dei dataset di Netflix con tutti i film presenti sulla sua piattaforma e il dataset di IMDB con le votazioni di tutti i film.

Dopo aver scaricato i dati di Netflix in formato csv e in formato txt abbiamo unito i dati in unico file csv. Successivamente abbiamo scaricato il dataset di IMBD in formato tsv che ha tutti i rating di tutti i film al mondo.

Dopo aver scaricato tutti i dati, ci siamo accorti nel dataset di Netflix c’è un id che è diverso dall’id dell IMDB quindi abbiamo preso la colonna “TITLE” di Netflix dataset e IMDB dataset

le abbiamo incrociate e tramite uno script di python abbiamo preso i soli dati di ratings dei film e Tv Show che sono presenti su Netflix.

**Identificatore dei dataset “TITLE” del film e o tv show**

**Team necessario**

Data Engineer

Progetta e crea la base di dati

Data Analyst

Analizza i dati tramite interrogazioni della base di dati e crea dei report

## Metadati

| Dataset | Tipologia licenza | formato | link fonte | livello file | descrizione |
| --- | --- | --- | --- | --- | --- |
| Dataset 1 | [CC0: Public Domain](https://creativecommons.org/publicdomain/zero/1.0/) | csv | https://www.kaggle.com/datasets/ariyoomotade/netflix-data-cleaning-analysis-and-visualization | 2 | tutti film e tv show presenti sulla piattaforma Netflix |
| Dataset 2 | [CC0: Public Domain](https://creativecommons.org/publicdomain/zero/1.0/) | txt | https://www.kaggle.com/datasets/shivamb/netflix-shows | 2 | visualizzazioni totali |
| Dataset 3 | [CC0: Public Domain](https://creativecommons.org/publicdomain/zero/1.0/) | tsv | <https://www.kaggle.com/datasets/ashirwadsangwan/imdb-dataset> | 2 | votazione di tutti i titoli su IMDB |

## Strutturazione dei requisiti in gruppi di frasi omogenee

Per i Film sono identificati da un codice numerico univoco, il titolo, il tipo, il regista, il cast, il paese, data di aggiunta alla piattaforma Netflix, il genere del film, la descrizione del film o tv show

Per i TV Show sono identificati da un codice numerico univoco, il titolo, il tipo, il regista, il cast, il paese, data di aggiunta alla piattaforma Netflix, il genere del film, la descrizione del film o tv show

## Glossario dei termini

| Termini | descrizione | sinonimo | collegamento |
| --- | --- | --- | --- |
| Movie | film di minimo 60 min | Film | Regista,Attori |
| Tv Show | Sono dei film di almeno 45 min ripetuti per due o più episodi | Serie Tv | Regista,Attori |
| Netflix | Piattaforma streaming per film e serie tv | Piattaforma | Director, country |
| Rating | Votazione in base a preferenza da x utenti | Votazione | tvshow,film,categoria,attori,registi |
| IMDB | Dataset di tutti i ratings di tutti i film nel mondo |  | tvshow,film,categoria,attori,registi |
| Genere | Tipologia di film | Categoria | regista, film, tvshow |
| Cast | Attori presenti nel film |  | attori, registiIn In base i |
| Listed\_in | categoria appartenente | Categoria | film, tvshow |
| votes | numero voti |  | film, tvshow |

## Tipologia di dati

Incrociando i dati di netflix con i dati delle view e i dati del rating su IMDB capire quali film sono piaciuti maggiormente dagli utenti

| Attribuito | Descrizione | Tipo di dato |  |
| --- | --- | --- | --- |
| show\_id | id numerico identificativo del film | alfanumerico |  |
| type | scelta tra Movie, Tv Show | categorico | Movie, Tv Show |
| title | titolo del film | testo |  |
| director | direttore del film | testo |  |
| cast | attori presenti | testo |  |
| country | paese di orgine del film | categorico | United States, Ghana, Burkina Faso, United Kingdom, Germany, Ethiopia… |
| data\_added | data di aggiunta del film sul portale | data |  |
| release\_year | data di rilascio del film | data |  |
| rating | voto | alfanumerico |  |
| duration | numerico nel caso dei film, nel caso dei tvshow è il numero delle stagioni | numerico |  |
| listed\_in | sezione di netflix dove è situato il film | testo | Dramas, Independent Movies, International Movies,Crime TV Shows, Docuseries, International TV Shows,Children & Family Movies, Comedies |
| description | descrizione breve del film | testo |  |
| Title\_id | id del film | alfanumerico |  |
| Ratings | Votazione media da 1 a 10 | numerico |  |
| Votes | Numero di voti | numerico |  |

## 

Anteprima dataset file **Netflix.csv**

| show\_id | type | title | director | cast | country | data\_added |
| --- | --- | --- | --- | --- | --- | --- |
| s1 | Movie | Dick Johnson Is Dead | Kirsten Johnson |  | United States | September 25, 2021 |
| s2 | TV Show | Blood & Water |  | Ama Qamata, Khosi Ngema, Gail Mabalane, Thabang Molaba, Dillon Windvogel, Natasha Thahane, Arno Greeff, Xolile Tshabalala, Getmore Sithole, Cindy Mahlangu, Ryle De Morny, Greteli Fincham, Sello Maake Ka-Ncube, Odwa Gwanya, Mekaila Mathys, Sandi Schultz, Duane Williams, Shamilla Miller, Patrick Mofokeng | South Africa | September 24, 2021 |
| s3 | TV Show | Ganglands | Julien Leclercq | Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabiha Akkari, Sofia Lesaffre, Salim Kechiouche, Noureddine Farihi, Geert Van Rampelberg, Bakary Diombera |  | September 24, 2021 |
| s4 | TV Show | Jailbirds New Orleans |  |  |  | September 24, 2021 |
| s5 | TV Show | Kota Factory |  | Mayur More, Jitendra Kumar, Ranjan Raj, Alam Khan, Ahsaas Channa, Revathi Pillai, Urvi Singh, Arun Kumar | India | September 24, 2021 |
| s6 | TV Show | Midnight Mass | Mike Flanagan | Kate Siegel, Zach Gilford, Hamish Linklater, Henry Thomas, Kristin Lehman, Samantha Sloyan, Igby Rigney, Rahul Kohli, Annarah Cymone, Annabeth Gish, Alex Essoe, Rahul Abburi, Matt Biedel, Michael Trucco, Crystal Balint, Louis Oliver |  | September 24, 2021 |
| s7 | Movie | My Little Pony: A New Generation | Robert Cullen, José Luis Ucha | Vanessa Hudgens, Kimiko Glenn, James Marsden, Sofia Carson, Liza Koshy, Ken Jeong, Elizabeth Perkins, Jane Krakowski, Michael McKean, Phil LaMarr |  | September 24, 2021 |

| release\_year | duration | listed\_in | description |
| --- | --- | --- | --- |
| 2020 | 90 min | Documentaries | As her father nears the end of his life, filmmaker Kirsten Johnson stages his death in inventive and comical ways to help them both face the inevitable. |
| 2021 | 2 Seasons | International TV Shows, TV Dramas, TV Mysteries | After crossing paths at a party, a Cape Town teen sets out to prove whether a private-school swimming star is her sister who was abducted at birth. |
| 2021 | 1 Season | Crime TV Shows, International TV Shows, TV Action & Adventure | To protect his family from a powerful drug lord, skilled thief Mehdi and his expert team of robbers are pulled into a violent and deadly turf war. |
| 2021 | 1 Season | Docuseries, Reality TV | Feuds, flirtations and toilet talk go down among the incarcerated women at the Orleans Justice Center in New Orleans on this gritty reality series. |
| 2021 | 2 Seasons | International TV Shows, Romantic TV Shows, TV Comedies | In a city of coaching centers known to train India’s finest collegiate minds, an earnest but unexceptional student and his friends navigate campus life. |
| 2021 | 1 Season | TV Dramas, TV Horror, TV Mysteries | The arrival of a charismatic young priest brings glorious miracles, ominous mysteries and renewed religious fervor to a dying town desperate to believe. |
| 2021 | 91 min | Children & Family Movies | Equestria's divided. But a bright-eyed hero believes Earth Ponies, Pegasi and Unicorns should be pals — and, hoof to heart, she’s determined to prove it. |

Anteprima Tabella Visualizzazioni Netflix\_views.txt

| show\_id | views |
| --- | --- |
| s1 | 7595 |
| s2 | 4442 |
| s3 | 7926 |
| s4 | 4499 |
| s5 | 3689 |
| s6 | 4778 |
| s7 | 6878 |

Anteprima Dataset **IMDB.tsv**

| title\_id | title | rating | votes |
| --- | --- | --- | --- |
| tt0000001 | Dick Johnson Is Dead | 5.7 | 1911 |
| tt0000002 | Blood & Water | 5.8 | 257 |
| tt0000003 | Ganglands | 6.5 | 1716 |
| tt0000004 | Jailbirds New Orleans | 5.6 | 169 |
| tt0000005 | Kota Factory | 6.2 | 2532 |
| tt0000006 | Midnight Mass | 5.1 | 173 |
| tt0000007 | My Little Pony: A New Generation | 5.4 | 790 |

## Supporto alle decisioni aziendali

Quale film è stato maggiormente visto?

Quale genere è meglio scegliere?

Del film maggiormente visto quale attore è più presente?

Quale regista è più presente nei film con votazione IMDB più alta?

Quale attore(cast) è più presente nei film con votazione IMDB più alta?

I film piu’ visti sono i film con il rating IMDB maggiore?

In quale paese sono stati prodotti i film più visti

In che anno sono stati prodotti i film più’ visti

In che anno sono stati prodotti i film con votazione IMDB maggiore?

## Outcome atteso

In base ai dati raccolti e analizzati, conoscendo le preferenze degli utenti della piattaforma possiamo consigliare un nuovo film o tv show per l’anno 2023.

Dei film già prodotti quali sponsorizzare tramite una campagna di marketing.

Se inserire nella piattaforma di un paese (es ITA) i film presenti solamente nella piattaforma di un altro paese(es USA).

Dai dati si evince che converrebbe produrre o un film di tipo Horror di almeno 90 Minuti e proporlo nei paesi USA, Spagna, Italia, Portogallo oppure

un Tv show di almeno 3 stagioni di tipo comedy in paesi come Germania, Francia, Olanda, Canda.