*Scegli un argomento di tuo interesse, riporta su un testo quali informazioni ti piacerebbe derivare cercando su internet i potenziali dataset strutturati (1) e non strutturati (2) utilizzabili (3 del primo, 2 del secondo).*

*Per ogni dataset strutturato individua gli attributi e l’identificativo, per ogni dataset non strutturato descrivi le analisi che svolgeresti su di esso e l’obbiettivo di queste analisi.*

Argomento: **Le 5 Serie tv più viste su netflix**



1)

* Nel primo dataset creiamo una tabella con:

- l’ID,

-titolo del film.

* Nel secondo avendo già l’id associato al film inseriamo come attributi (oltre l’id):

-ID

-numero ore di visione totali,

-paesi con numero di ore di visione maggiore.

* Nel terzo come attributi mettiamo:

-ID

-genere,

-numero di episodi,

-paese di produzione.

2)

* Dai trailer su youtube possiamo estrapolare il numero di visioni e di like.
* Dai commenti sia dei trailer che delle successive recensioni vediamo se ci sono più commenti positivi o negativi (sentiment analysis).

L’obbiettivo di queste analisi può essere verificare quale serie tv è riuscita a catturare l’attenzione maggiore prima dell’uscita del film e confrontarlo poi con il numero degli ascolti; dai commenti si può verificare invece il gradimento degli utenti ante e post il film, verificando se ci sono più commenti negativi e/o positivi prima e dopo, per capire se l’utente ha cambiato idea (opinione) sul film.