



Analisi delle caratteristiche dei Film e del loro successo economico: Un'indagine esplorativa

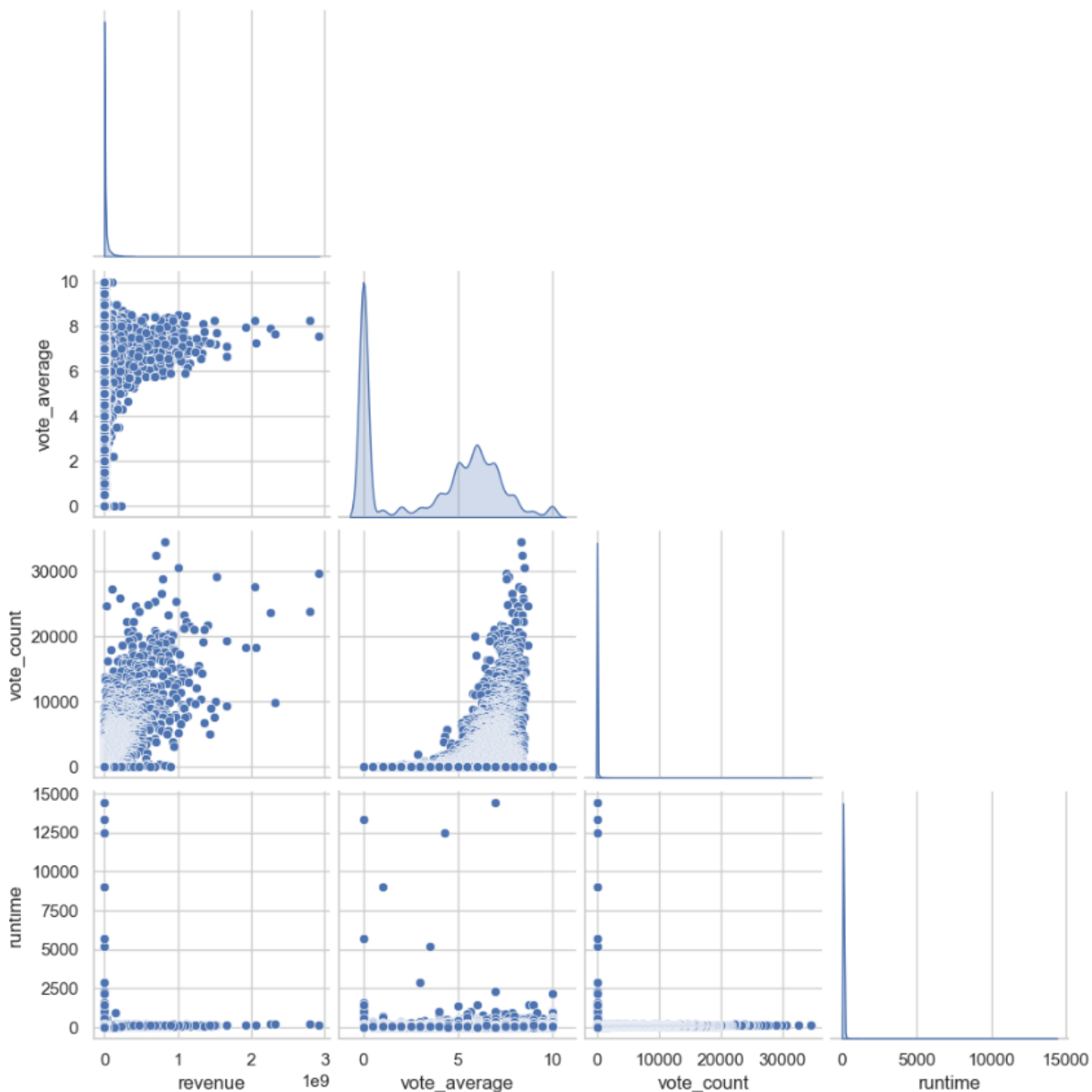
Corso Python & Machine Learning

Simone Buglione

Analisi preliminare

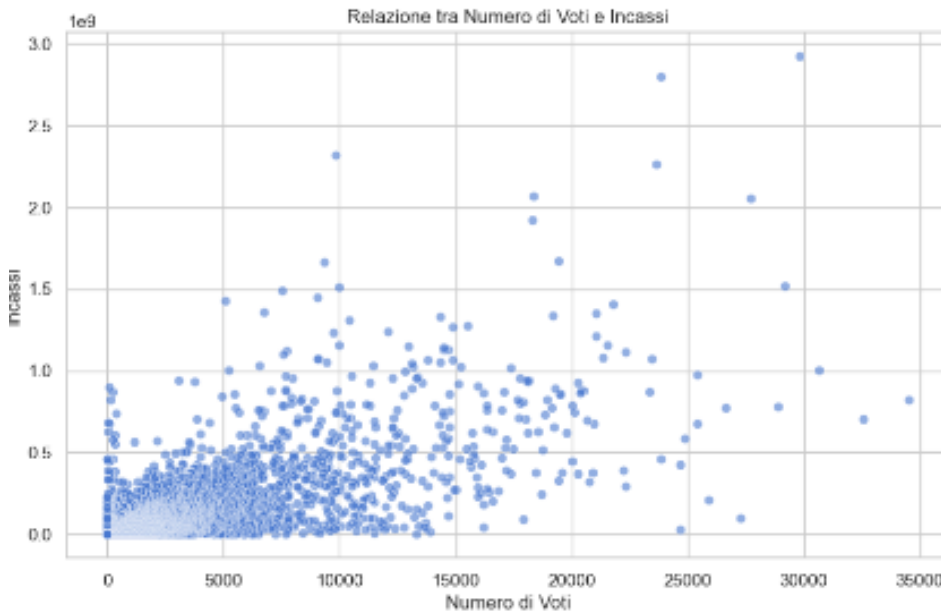
PairPlot iniziale

- ✚ Dopo aver selezionato le colonne più interessanti, è possibile notare una relazione tra revenue e vote_count, emerge che maggiore è il numero di voti, e maggiore tende ad essere il revenue. Quindi i voti con maggiori voti generano spesso guadagni elevati.
- ✚ Non c'è una relazione forte tra voti medi e numero voti, quindi un punteggio alto di voti non sempre è correlato ad un numero elevato di voti, potrebbero esserci film di nicchia con voti alti ma pochi spettatori.
- ✚ I picchi a ridosso dello 0 sono probabilmente valori mancanti o errati, che non sono necessari per l'analisi e quindi vanno rimossi.
- ✚ Non c'è una relazione forte tra revenue(incassi) e durata del film(runtime).



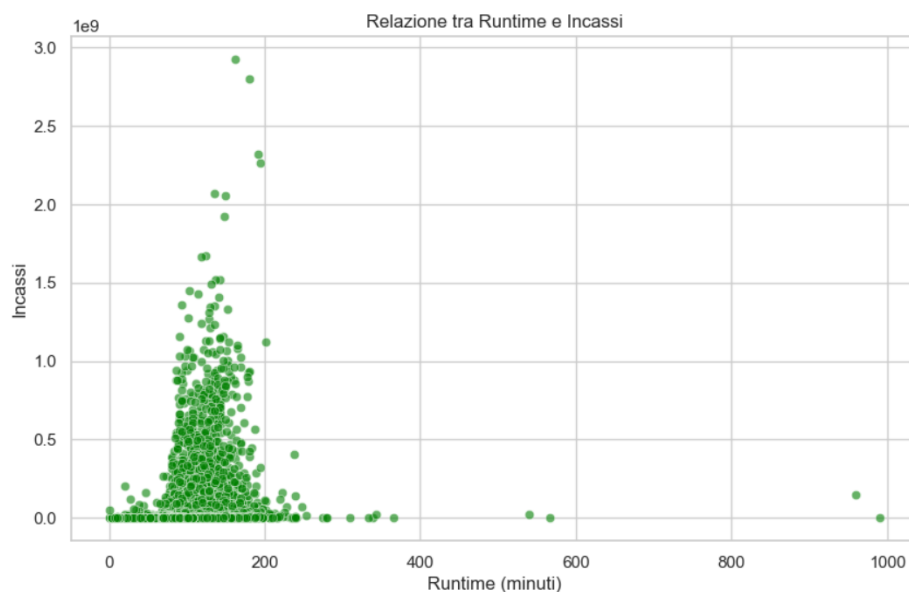
Relazione tra numero di voti e incassi

Dal grafico sottostante emerge che un numero maggiore di voti tende ad essere associato a incassi più alti, anche se c'è una grande variabilità e dispersione nei dati, quindi ci sono delle eccezioni. In generale emerge però la tendenza generale che conferma che un **film con più voti è anche un film che ha avuto incassi maggiori**.



Relazione tra Runtime e Incassi

La relazione tra la durata del film non sembra essere significativa, anche se ci sono film con durate estreme che ottengono incassi elevati, suggerendo quindi che la durata non è un fattore determinante per il successo economico.



Heatmap delle correlazioni tra le variabili

La heatmap delle correlazioni tra le variabili principali del dataset mostra che il numero di voti è fortemente correlato con gli incassi, confermando il grafico precedente, suggerendo che un maggiore coinvolgimento del pubblico tende a portare a maggiori guadagni. Le altre variabili, tendono invece ad avere correlazioni positive ma poco significative per l'analisi di riferimento.



Obiettivo e domande di ricerca

L'obiettivo di questa analisi è analizzare le caratteristiche dei film allo scopo di identificare i fattori che influenzano il loro successo economico. Nello specifico si vuole comprendere come variabili come il numero di voti, il voto medio, la durata del film e il genere influenzano gli incassi dei film.

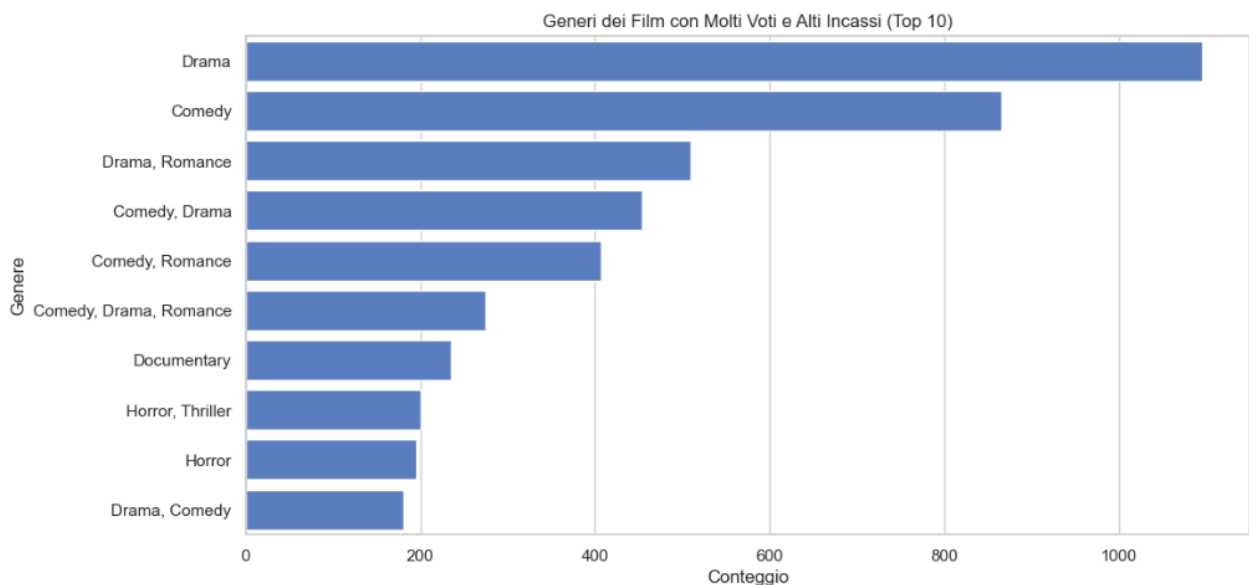
A seguito dell'analisi preliminare, le domande poste sono le seguenti:

1. Quali sono le caratteristiche comuni dei film con molti voti e alti incassi?
2. Come varia la distribuzione dei voti in base al genere del film?
3. Ci sono differenze significative tra i film con molti voti ma incassi bassi e quelli con molti voti e incassi alti?

Generi dei film con molti voti e alti incassi

Dal grafico sottostante emerge che:

- ✚ **Generi Dominanti:** Il genere "Drama" è il più rappresentato tra i film con molti voti e alti incassi, seguito da "Comedy". Questo suggerisce che i film drammatici e comici tendono a coinvolgere maggiormente il pubblico, generando incassi elevati.
- ✚ **Generi Misti:** Generi come "Drama, Romance" e "Comedy, Drama" sono anche molto presenti. Questo indica che il pubblico preferisce film che combinano più generi, probabilmente perché attraggono un pubblico più ampio.
- ✚ **Documentari e Horror:** I generi "Documentary" e "Horror, Thriller" hanno una presenza significativa, ma minore rispetto ai primi. Questo potrebbe indicare che anche se questi generi hanno un pubblico specifico, non raggiungono lo stesso livello di incassi dei generi più popolari.

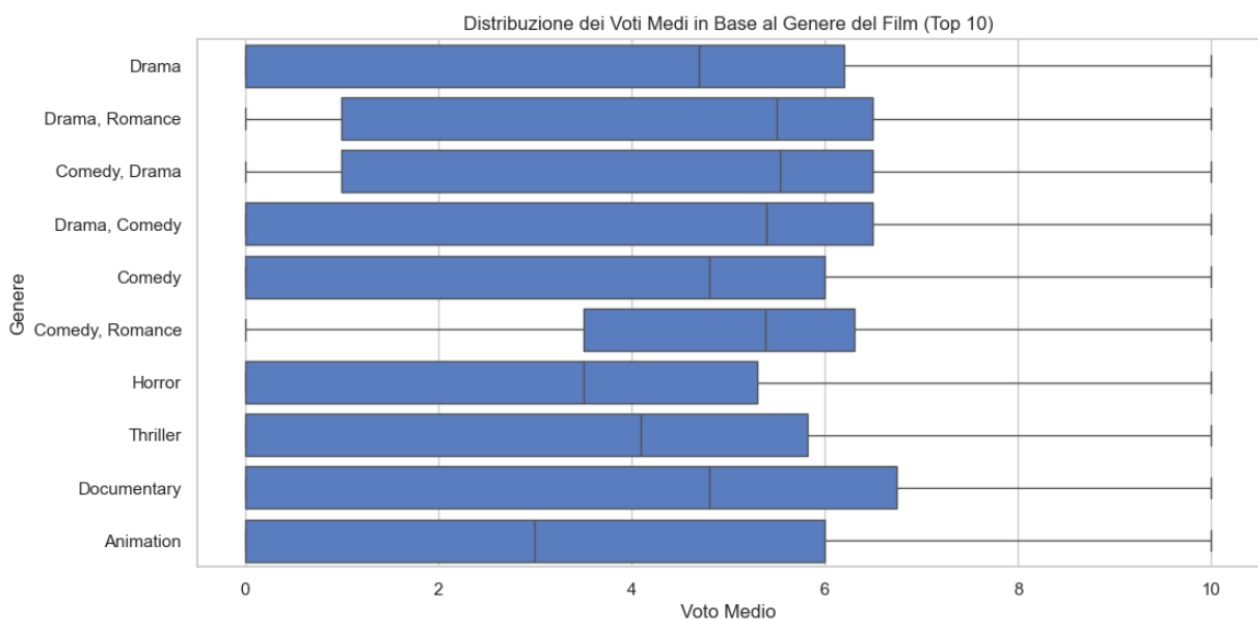


Distribuzione dei voti in base al genere del film

dal grafico che mostra la distribuzione dei voti medi (vote_average) per i primi 10 generi di film emerge che:

- ✚ Il genere **Drama** risulta essere il genere con la media di voti più alta, suggerendo che i film drammatici ricevono solitamente valutazioni positive dal pubblico.
- ✚ **Drama Romance e Comedy Drama:** potrebbero suggerire che le storie che mescolano elementi drammatici con il romanticismo o la commedia piacciono molto al pubblico.
- ✚ **Comedy e Comedy Romance:** Anche i film comici e le commedie romantiche hanno una buona media di voti, dimostrando che il pubblico apprezza l'umorismo e le storie d'amore leggere.
- ✚ **Documentary:** I documentari hanno una media di voti relativamente alta, indicando che sebbene possano avere un pubblico più ristretto, sono molto apprezzati dallo spettatore.
- ✚ **Horror, Thriller:** Questi generi hanno una media di voti più bassa rispetto agli altri, suggerendo che possono essere più divisivi o che il pubblico ha aspettative diverse su questo genere di film.

Tra i vari aspetti analizzati, quello più interessante che emerge è certamente quello **emotivo**, in quanto generi come quello romantico e comico ricevono valutazioni più alte, indicando che il **coinvolgimento emotivo** ha un impatto positivo sul pubblico.

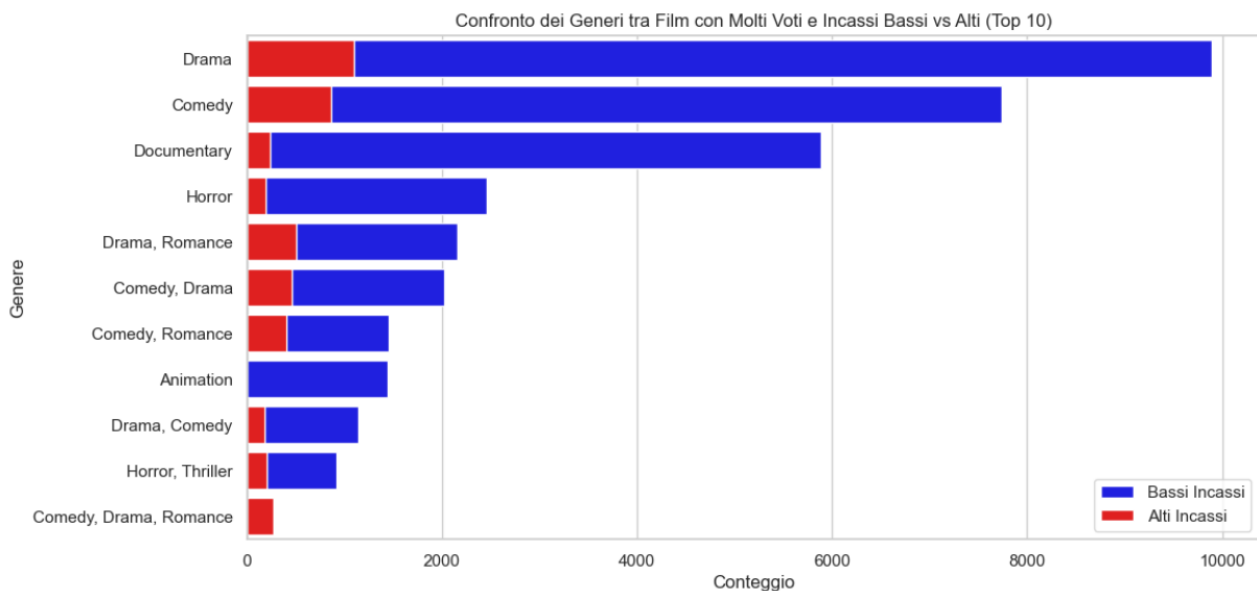


Confronto dei generi tra film con molti voti e incassi bassi vs alti

Dal grafico sottostante che confronta i generi cinematografici tra film con molti voti e incassi bassi (in blu) e film con incassi alti (in rosso) emerge che:

- ✚ **Drama e Comedy:** Questi generi hanno un numero significativamente maggiore di film con incassi bassi rispetto a quelli con incassi alti. Questo suggerisce che, sebbene siano generi popolari, **non sempre garantiscono alti incassi**.
- ✚ **Drama, Romance e Comedy, Drama:** Questi generi combinati mostrano una presenza sia di film con incassi bassi che alti, indicando che **possono attrarre un pubblico ampio e variegato**.
- ✚ **Documentary e Horror:** Questi generi tendono ad avere più film con incassi bassi, anche se ci sono alcuni film con incassi alti suggerendo che, sebbene abbiano un pubblico di nicchia, possono comunque ottenere successo.
- ✚ **Animation e Horror, Thriller:** Questi generi hanno una presenza minore in generale, ma sono comunque rappresentati sia nei film con incassi bassi che alti.

È interessante notare come generi popolari come drama e comedy non sempre garantiscono alti guadagni, suggerendo ancora una volta che aspetti quali ad esempio la qualità del film, del marketing ecc. giocano un ruolo fondamentale nel determinare il successo economico del genere.



Conclusioni

Dall'analisi esplorativa, emergono diversi aspetti interessanti riguardanti le caratteristiche dei film e il loro successo economico. La relazione tra il numero di voti e gli incassi è molto significativa: un numero maggiore di voti tende ad essere associato a incassi più alti, suggerendo che il **coinvolgimento del pubblico** è un fattore cruciale per il successo finanziario di un film. Inoltre, non emerge una relazione forte tra i voti medi e il numero di voti, indicando che **un alto punteggio non sempre corrisponde a un numero elevato di voti**. Questo potrebbe essere dovuto a film di nicchia con voti alti ma pochi spettatori. Inoltre, la **durata del film non sembra essere un fattore determinante per il successo economico**, anche se ci sono film con durate estreme che ottengono incassi elevati.

La heatmap delle correlazioni conferma che il **numero di voti è fortemente correlato con gli incassi**, suggerendo che un **maggiore coinvolgimento del pubblico tende a portare a maggiori guadagni**. Le altre variabili mostrano correlazioni positive ma meno significative. Analizzando i generi dei film con molti voti e alti incassi, emerge che i generi "Drama" e "Comedy" sono i più rappresentati, seguiti da generi combinati come "Drama, Romance" e "Comedy, Drama". Questo indica una **preferenza per film che combinano più generi**, probabilmente perché attraggono un pubblico più ampio. I documentari e i film horror, sebbene abbiano un pubblico specifico, non raggiungono lo stesso livello di incassi dei generi più popolari.

La distribuzione dei voti medi per i primi 10 generi di film mostra che i **film drammatici tendono a ricevere valutazioni positive dal pubblico**. Anche i film comici e le commedie romantiche hanno una buona media di voti, dimostrando che **il pubblico apprezza l'umorismo e le storie d'amore leggere**. I documentari, sebbene abbiano un pubblico più ristretto, sono molto apprezzati. I generi horror e thriller, invece, hanno una media di voti più bassa, suggerendo che possono essere più divisivi o che il pubblico ha aspettative diverse.

Infine, il confronto tra i generi cinematografici tra film con molti voti e incassi bassi e quelli con incassi alti rivela che **generi popolari come "Drama" e "Comedy" non sempre garantiscono alti incassi**. Questo suggerisce che altri fattori, come la **qualità del film e il marketing, giocano un ruolo cruciale nel determinare il successo economico**.

In conclusione, il **coinvolgimento del pubblico è un fattore chiave per il successo finanziario di un film**, ma la **qualità percepita e la durata del film non sono necessariamente indicatori di successo economico**. I generi drammatici e comici, specialmente quelli che combinano più generi, tendono a ricevere valutazioni positive e a generare incassi elevati. Altri fattori come **l'aspetto emotivo, la qualità del film, il marketing e il cast giocano un ruolo fondamentale nel determinare il successo economico**.