

# Spotify by Attributes



**Un'analisi degli attributi delle canzoni su Spotify**

**Basata su dataset di kaggle.com**

**Di Nicholas Chanine**

# Cos'è Spotify?

Spotify è una piattaforma di streaming musicale creata da Daniel Ek e Martin Lorentzon che ha trasformato l'accesso alla musica, soppiantando le piattaforme di streaming illegali come Napster in favore di un modello che coinvolge direttamente gli artisti offrendo loro un pagamento in base a quanti ascolti riceve ciascun brano.

Lanciata ufficialmente nel 2008, Spotify ha reso la musica disponibile on-demand, offrendo agli utenti uno sconfinato catalogo di brani in streaming gratuito o tramite abbonamenti premium.

Grazie alla sua facilità d'uso e efficacia degli algoritmi di raccomandazione, Spotify ha rapidamente guadagnato popolarità a livello globale, ridefinendo il modo in cui le persone ascoltano e scoprono la musica.

# Il dataset

È stata usata come base un dataset, “114000 Spotify Songs” (<https://www.kaggle.com/datasets/priyamchoksi/spotify-dataset-114k-songs>), aggiornato fino al 2023, che è stato poi integrato con una parte molto ristretta di un altro dataset più recente, “Spotify Song Attributes” (<https://www.kaggle.com/datasets/byomokeshsenapati/spotify-song-attributes>), per poter effettuare analisi anche su un nuovo genere musicale (il cosiddetto “phonk”).

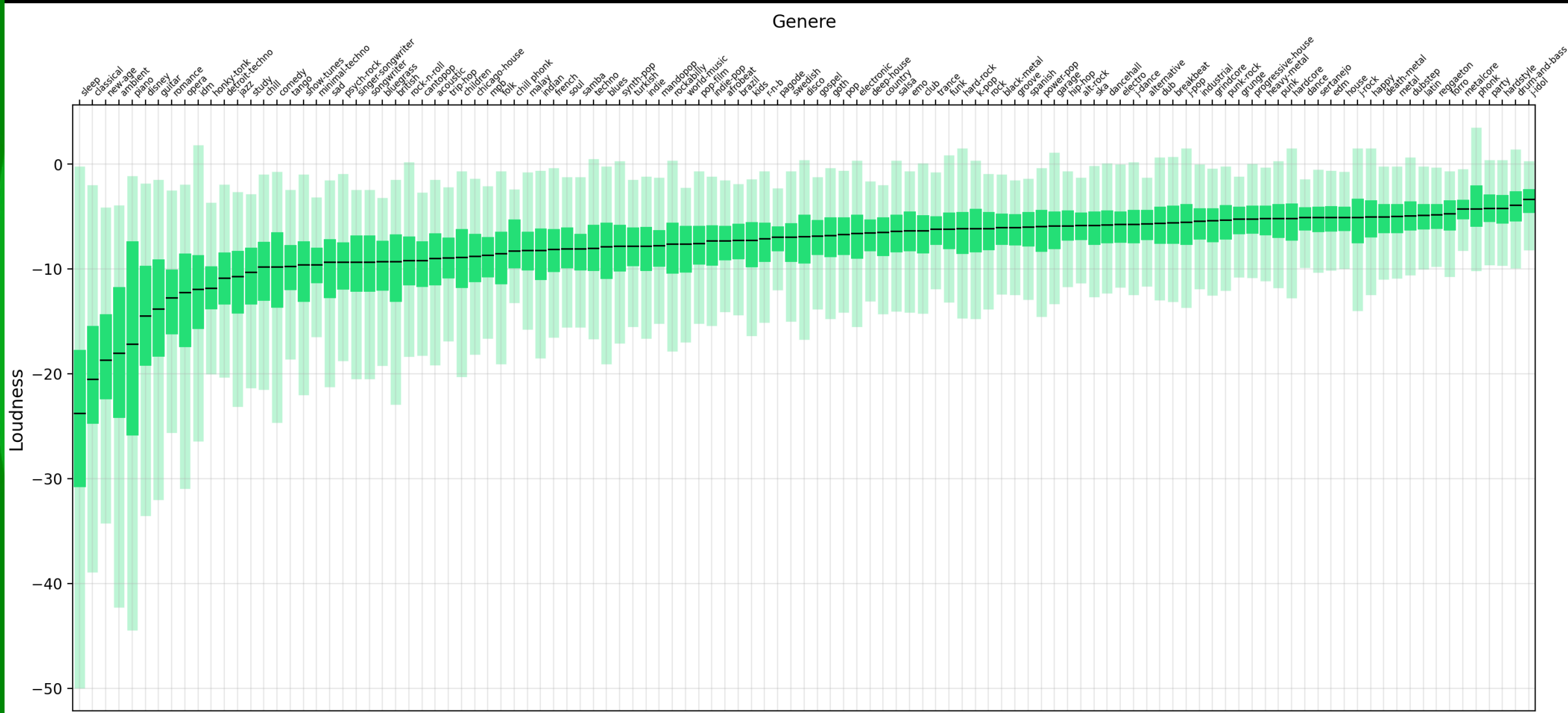
È giusto menzionare che questo secondo dataset è stato **modificato** per ridurre il numero di sottogeneri da 5 a 2: semplicemente, alcuni generi erano rappresentati da poche tracce o addirittura una sola, rendendo impossibile effettuare alcuna considerazione su questi gruppi.

Anche con questi accorgimenti, questi due sottogeneri contengono molte meno canzoni (circa 100 ciascuno) rispetto a un qualsiasi genere dell'altro dataset (più di 200 canzoni ciascuno), rendendo *alcune considerazioni possibilmente meno accurate*.

L'immagine di sfondo proviene da Google Immagini.

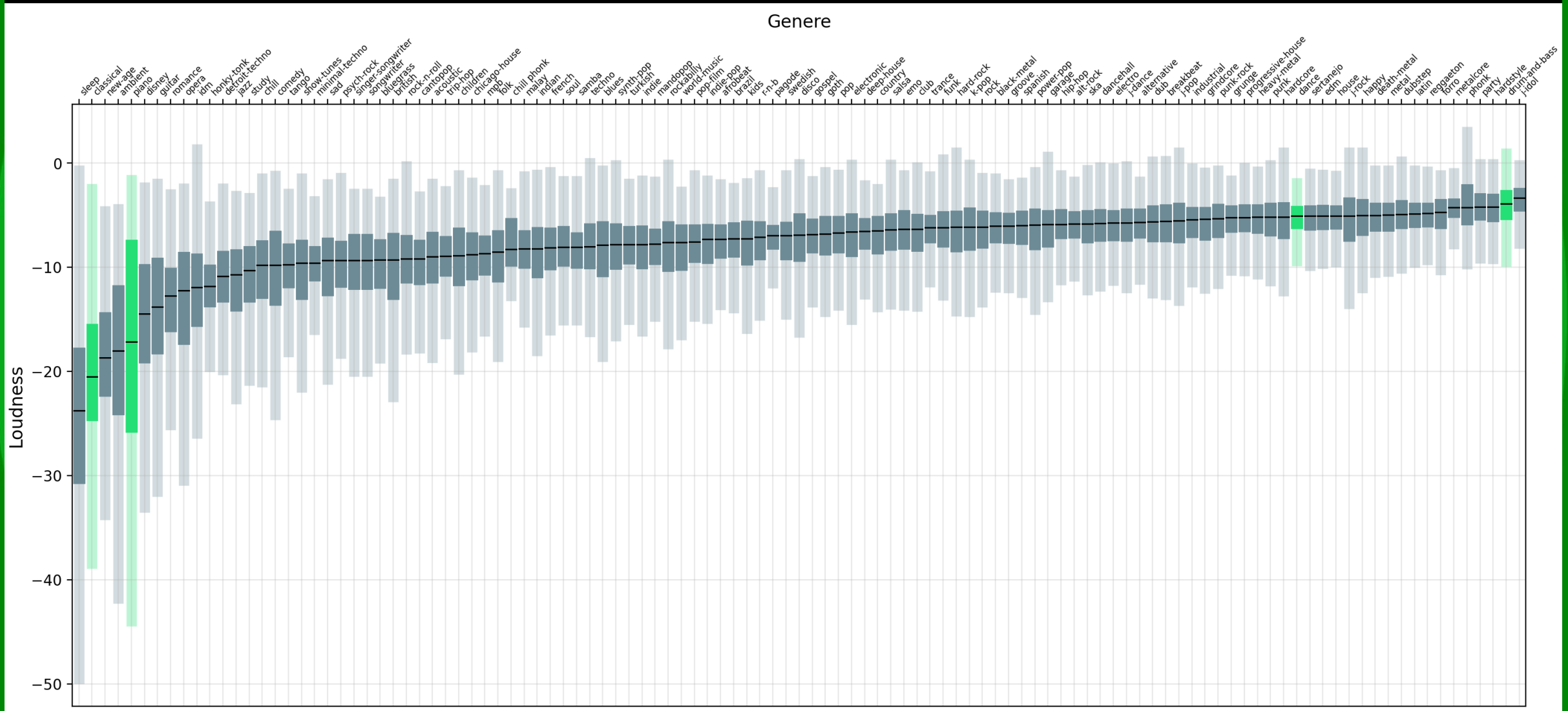
# Loudness per genere

Le canzoni con i **volumi più elevati** appartengono ai **generi di musica più pesanti** come phonk, metal e EDM, mentre quelle con loudness media più bassa sono sulla linea di musica classica o colonne sonore, il che riflette l'intensità generalmente minore di questi generi.



# Per curiosità...

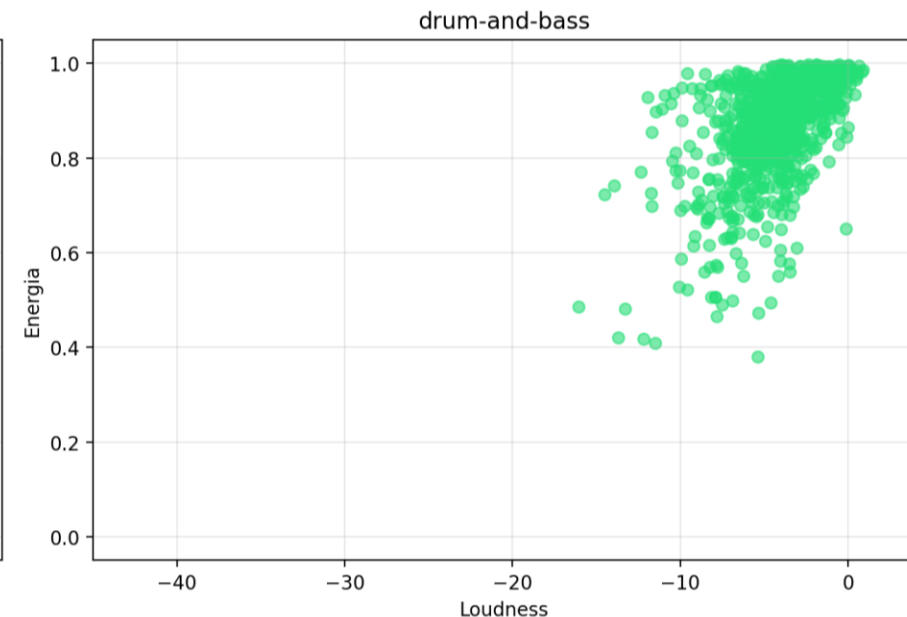
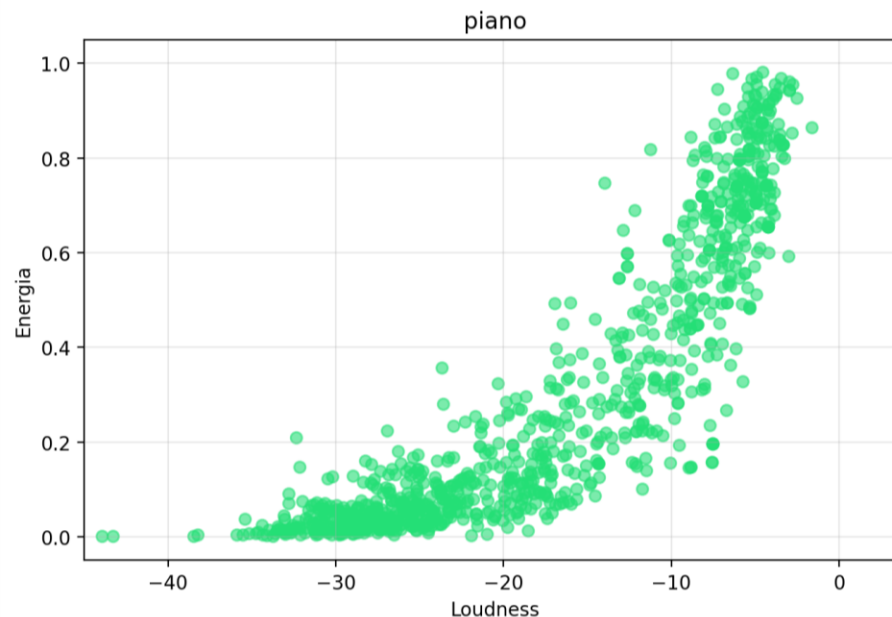
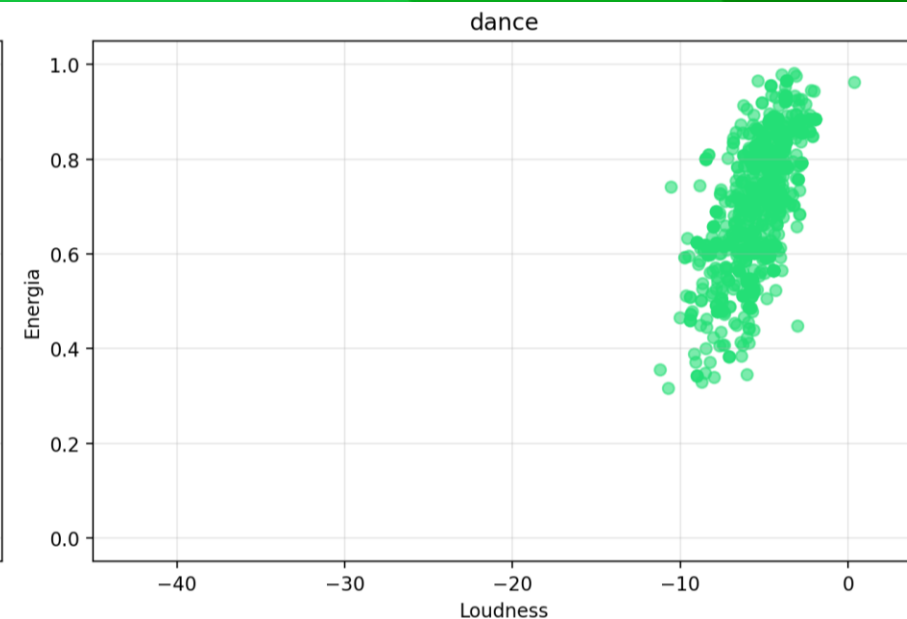
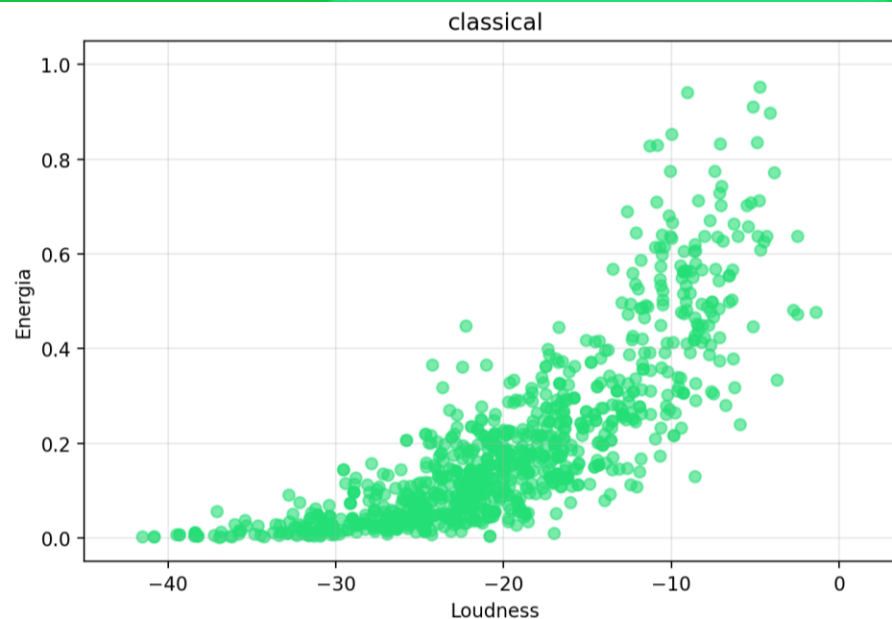
Alcuni generi hanno **varianza estrema**: verifichiamo se questa varianza è dovuta più a scarsa continuità di volumi tra un brano e l'altro a parità di intensità o semplicemente a un mood diverso (se il brano è tranquillo o aggressivo).



# Energia VS Loudness

È presente una **forte correlazione** tra energia e loudness.

Da questo grafico è resa evidente la differenza di filosofia sulla qualità del suono tra la musica classica, dove le registrazioni vengono pubblicate senza trattamento per preservare **l'integrità della registrazione**, e la musica elettronica, dove è presente molta **attenzione a mantenere volumi simili** tra una traccia e l'altra a parità di intensità, ed è molto sentita la “**loudness war**”, ovvero la sfida tra produttori a creare canzoni con i volumi più alti possibili.



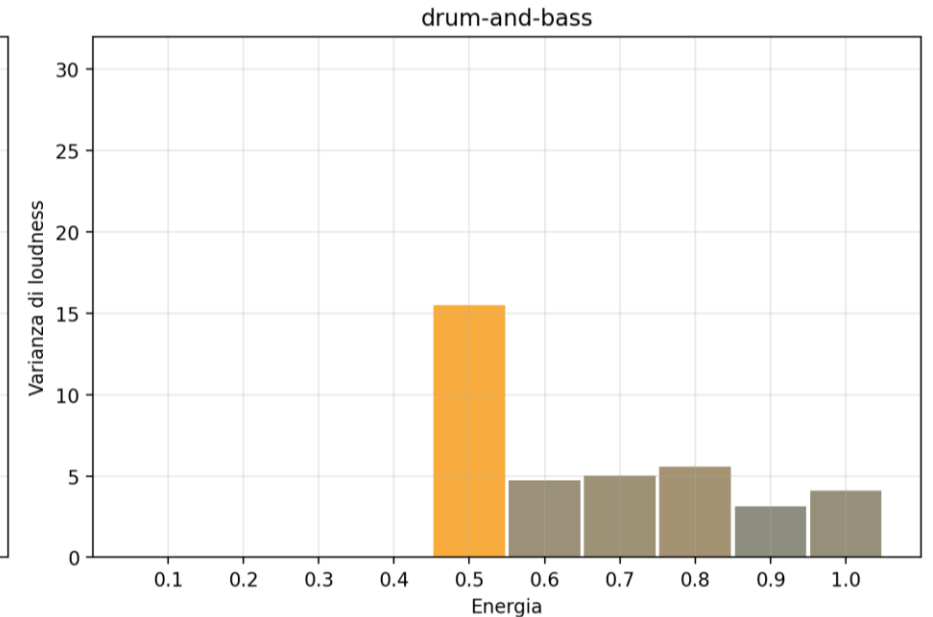
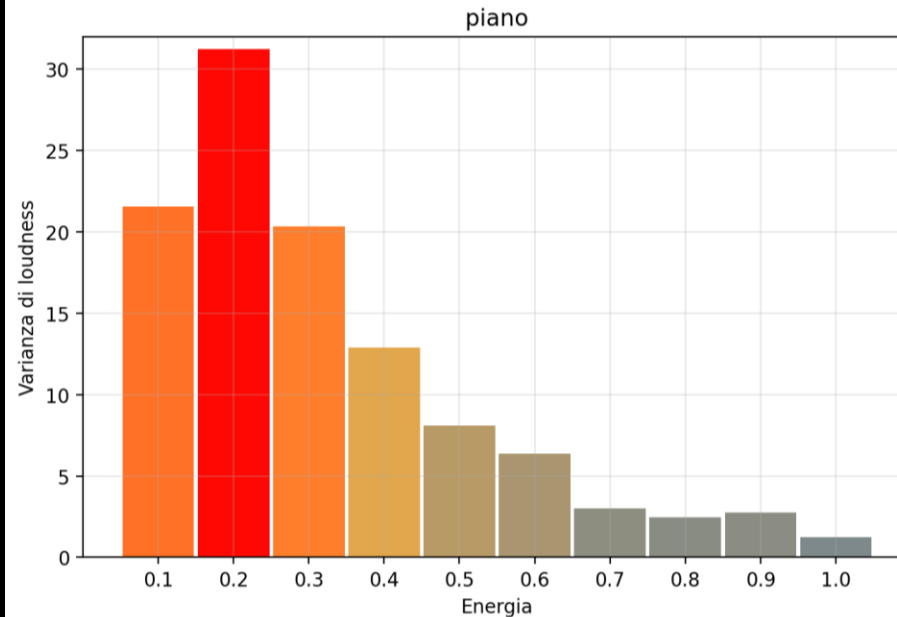
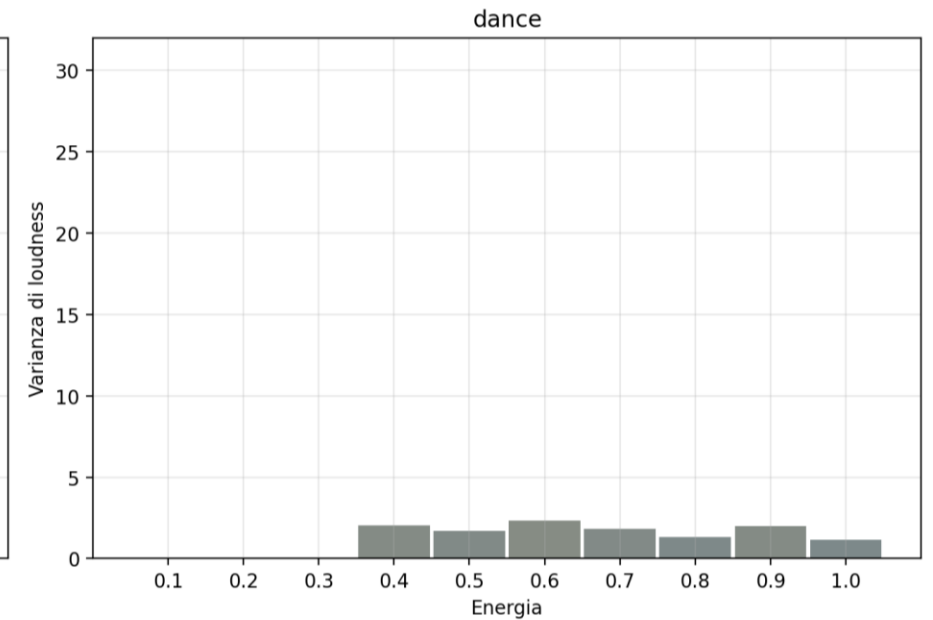
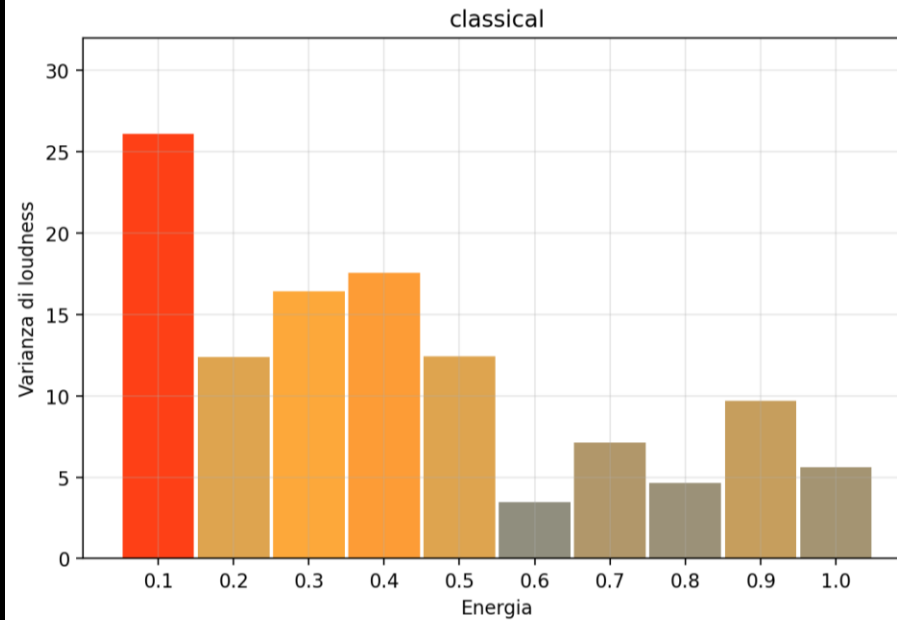
# Varianza di Loudness per intervallo di energia

Più è bassa la varianza per un dato intervallo di energia, maggiore è la **continuità dei volumi** tra un brano e l'altro. Nella musica classica e al pianoforte non ci si cura di questa **elevata disparità**, mentre nell'elettronica sì.

Ma esiste una motivazione: i DJ hanno bisogno di far suonare bene i **passaggi da un brano all'altro**, e la coerenza tra i volumi aiuta.

Spotify si avvale dell'attributo loudness proprio per **regolare il volume di ciascun brano** in modo da mantenerlo consistente tra un brano e l'altro, lavorando come una sorta di DJ digitale.

Notiamo che la **varianza è più bassa a intensità più elevate**, in quanto più è alta la loudness più ci si avvicina al **limite fisico** del volume raggiungibile in un file audio, e questo limite aiuta a rendere i livelli più uniformi.



## Cosa intende Spotify per popolarità?

Spotify classifica la popolarità di un brano attraverso un **indice da 0 a 100** denominato “popularity”. Questo punteggio viene attribuito in base a diverse metriche, tra cui il numero di ascolti e la data di pubblicazione.

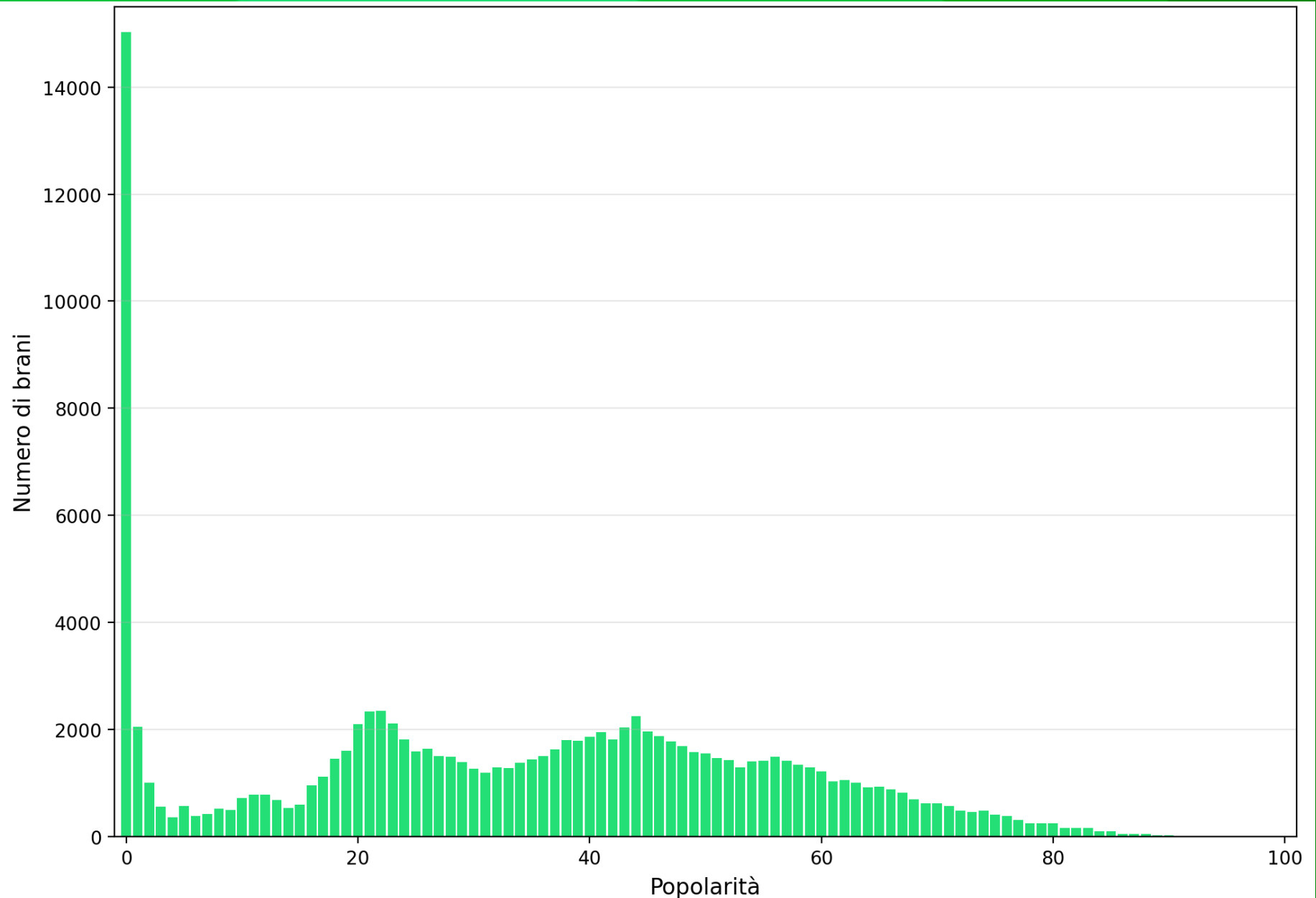
Una canzone con popolarità **25** potrebbe appartenere a un **genere di nicchia**, ma con un **forte seguito**; una canzone con punteggio vicino a **100** solitamente è musica che gode di un **appeal generale** come la musica pop, o musica diventata **virale** grazie ai social media.



## Numero di brani per popolarità

La maggior parte dei brani ha popolarità pari a 0; solo relativamente **pochi brani** ricevono numeri consistenti di ascolti.

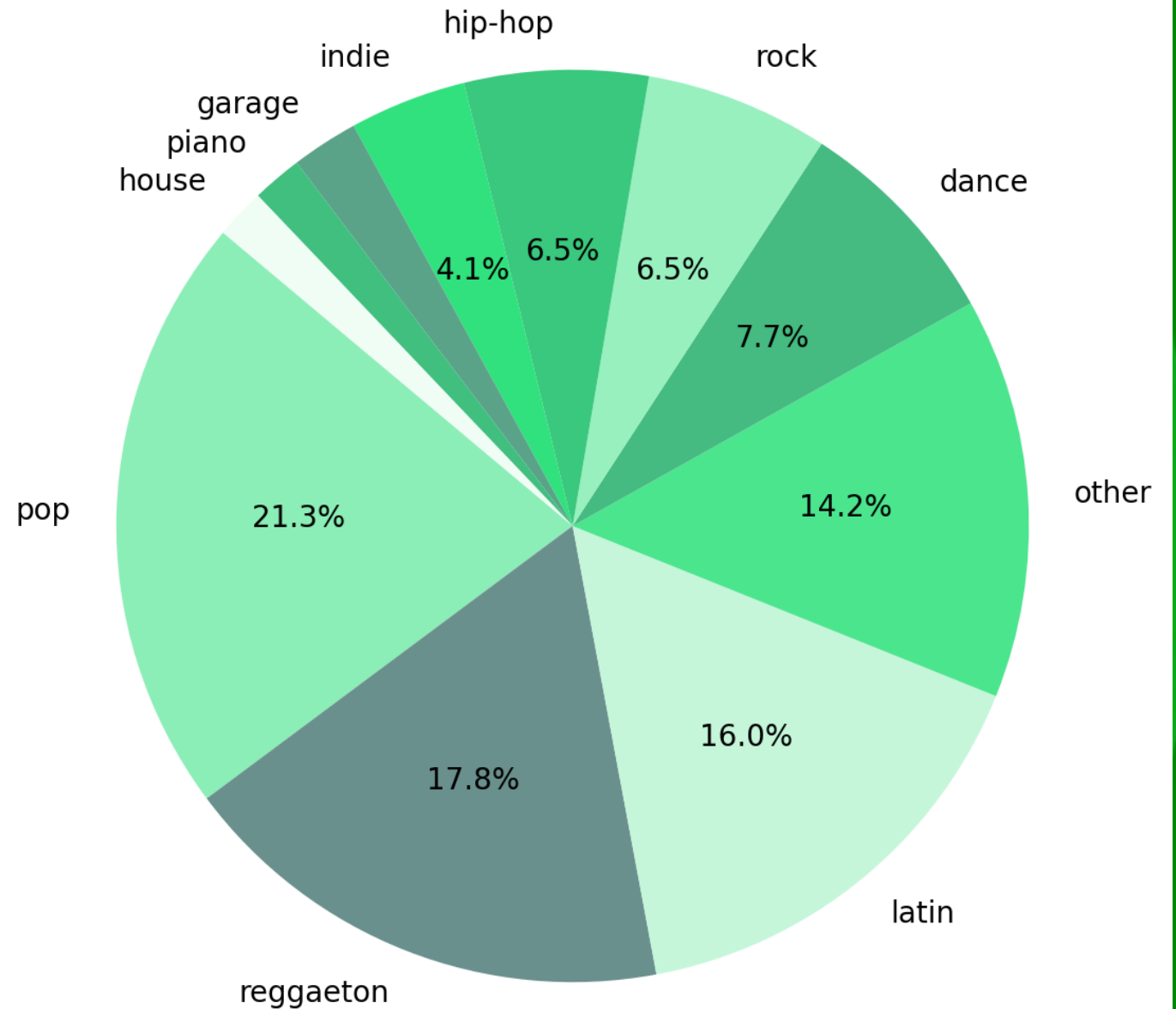
La scarsità di brani tra i livelli di popolarità 10 e 20 sembra indicare un intervento dell'algoritmo di Spotify nel promuovere o declassare i brani presenti in questo intervallo.



## Quali sono i generi delle canzoni più popolari?

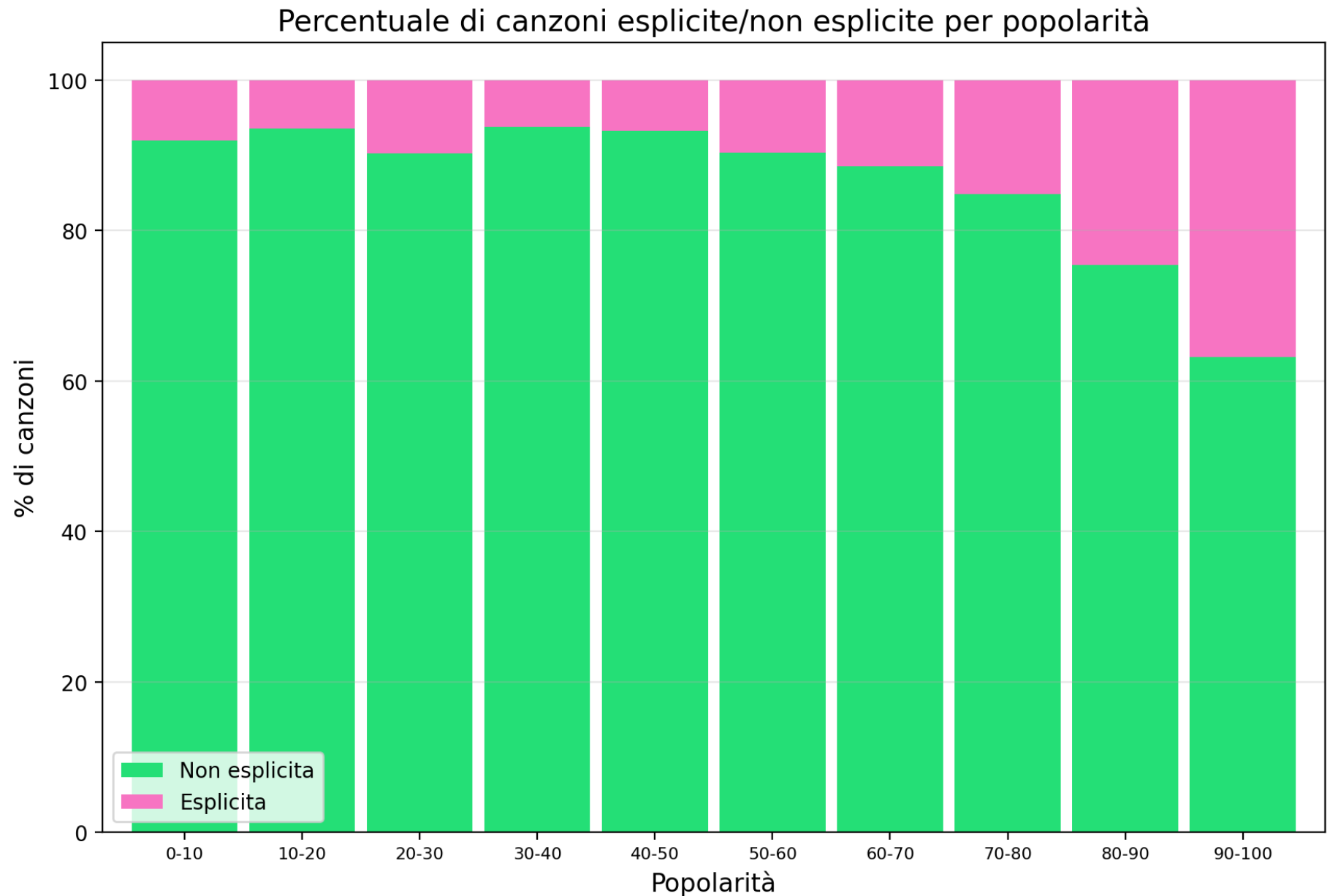
Notiamo che il **pop** e il **reggaeton**, insieme alla **musica latino-americana** in generale, compongono **più del 50%** delle canzoni più popolari (con l'attributo "popularity" *arbitrariamente scelto come superiore a 88*).

Oltre a questi generi super-popolari, rimangono popolari il **rap**, il **rock** e la **musica dance**.



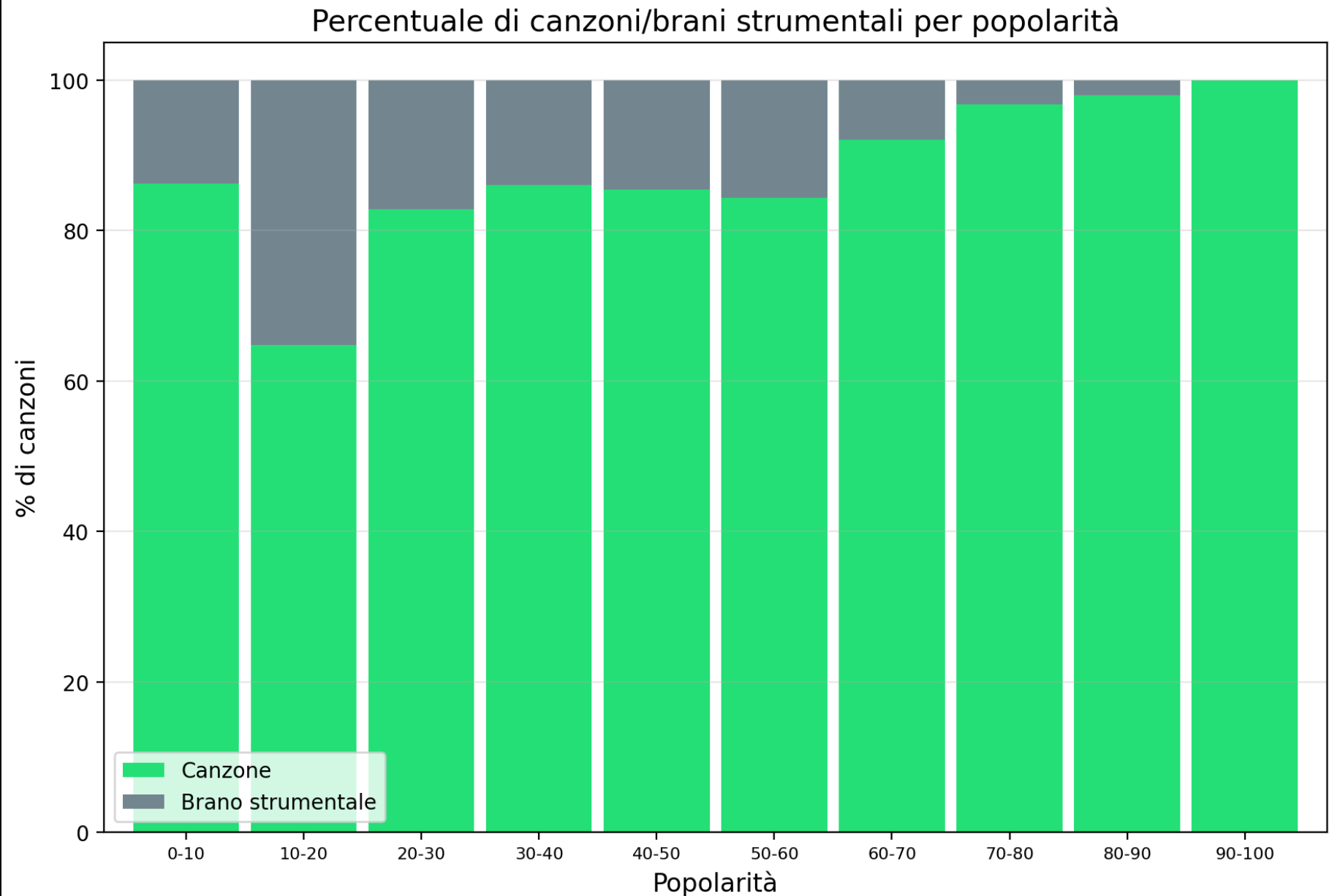
## Le canzoni esplicite sono più popolari? 1

Sebbene in media siano sotto il 10%, le canzoni più popolari in assoluto tendono ad essere **esplicite più spesso**: per popolarità maggiore di 90 sono quasi il **40%**.



## I brani strumentali sono popolari?

Come ci si può aspettare, i brani strumentali non sono popolari come le canzoni, anche se sono comunque rappresentati fino all'intervallo 80-90.

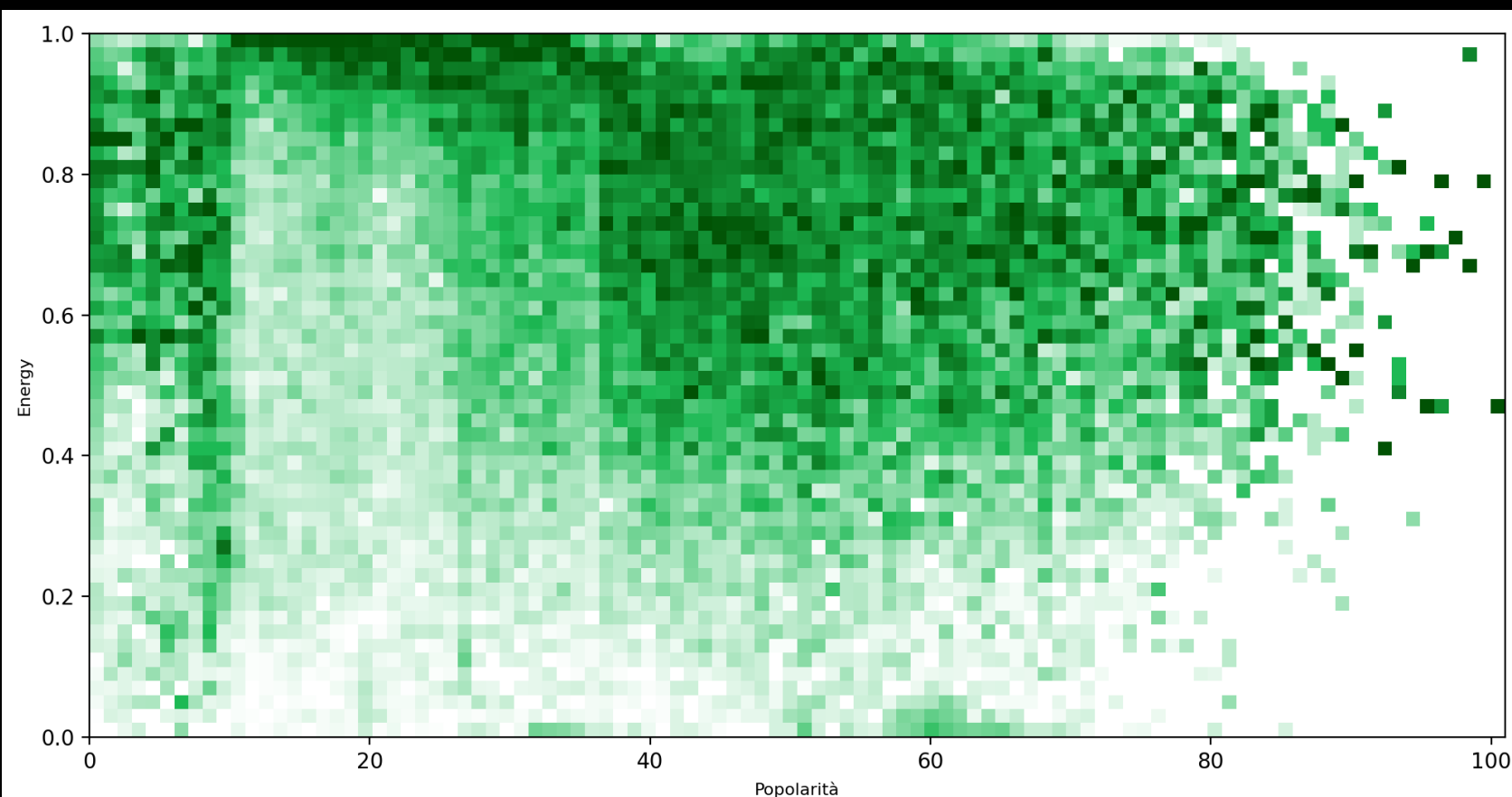


# Esiste un legame tra energia e popolarità?

Notiamo che in media l'energia dei brani è piuttosto elevata. Al di sotto di 10, dove si concentra la maggioranza dei brani, i brani non sono stati ancora riconosciuti dall'algoritmo e/o non provengono da artisti affermati o case discografiche famose.

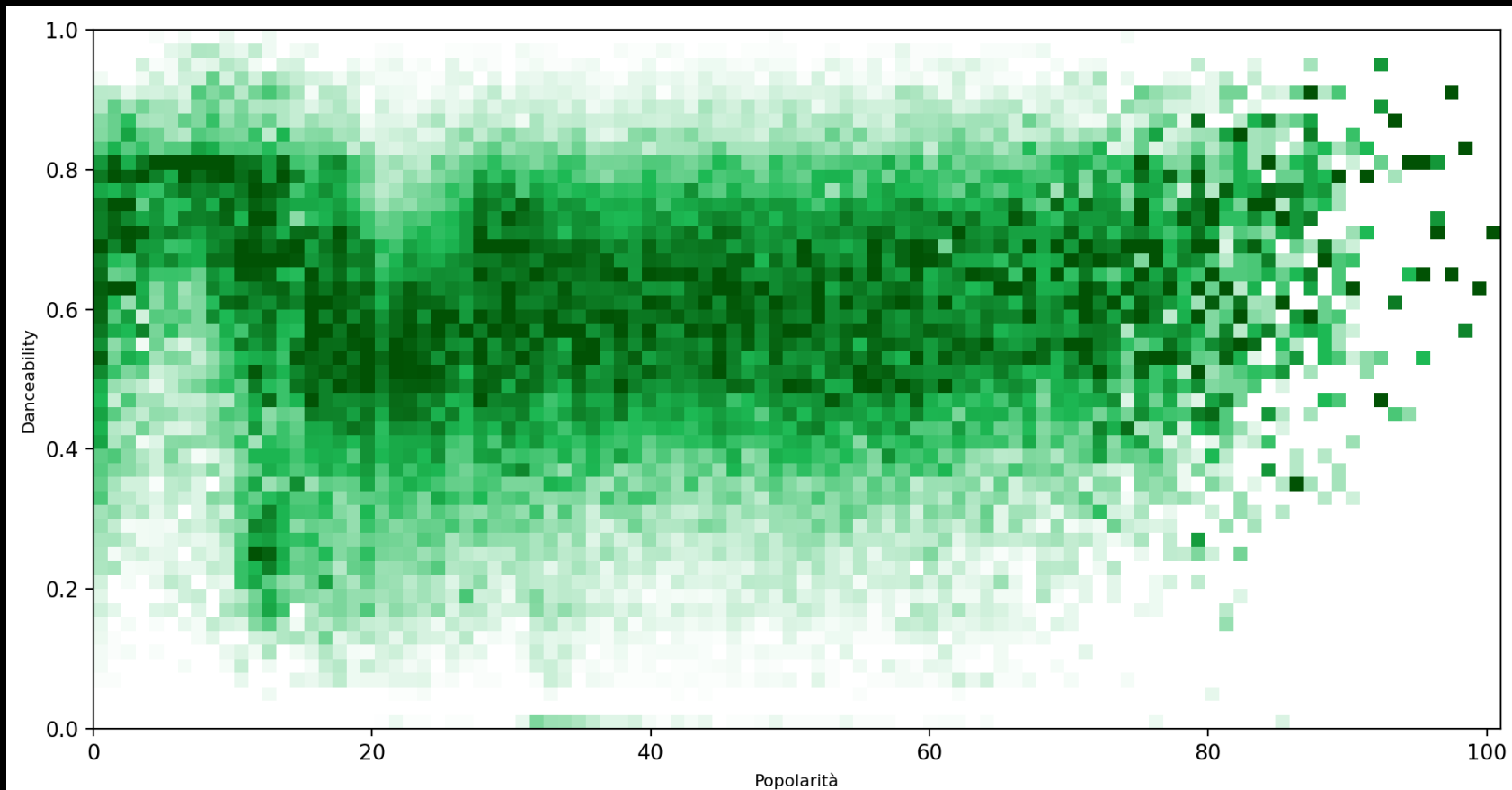
Tra i 10 e i 35 si trovano tracce ad **altissima intensità**; la discontinuità oltre i 35 sembra suggerire un **intervento da parte dell'algoritmo** di Spotify nel selezionare brani, ovvero l'inserimento in playlist editoriali. Con popolarità oltre 80 si nota una leggera diminuzione dell'energia.

**Nota:** il numero di brani diminuisce notevolmente all'aumentare della popolarità: un quadrato verde chiaro a sinistra potrebbe contenere molte più tracce che un quadrato nero a destra.



# Esiste un legame tra ballabilità e popolarità?

Le stesse fasce si possono all'incirca ritrovare per l'aspetto di ballabilità, dove nel lato sinistro del grafico l'andamento è più discontinuo. In questo caso, possiamo notare una **correlazione tra ballabilità e popolarità** maggiore di 70: **le tracce più popolari in assoluto sono più ballabili.**



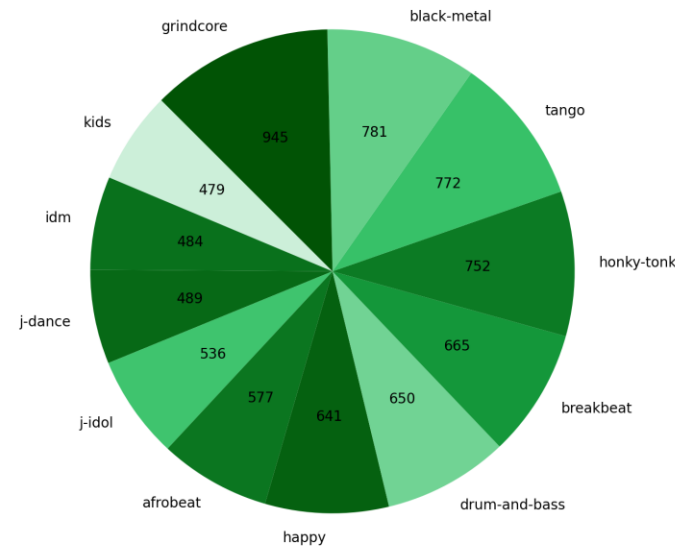
**Nota:** il numero di brani diminuisce notevolmente all'aumentare della popolarità: un quadrato verde chiaro a sinistra potrebbe contenere molte più tracce che un quadrato nero a destra.

# Esiste un legame tra genere e popolarità? 1

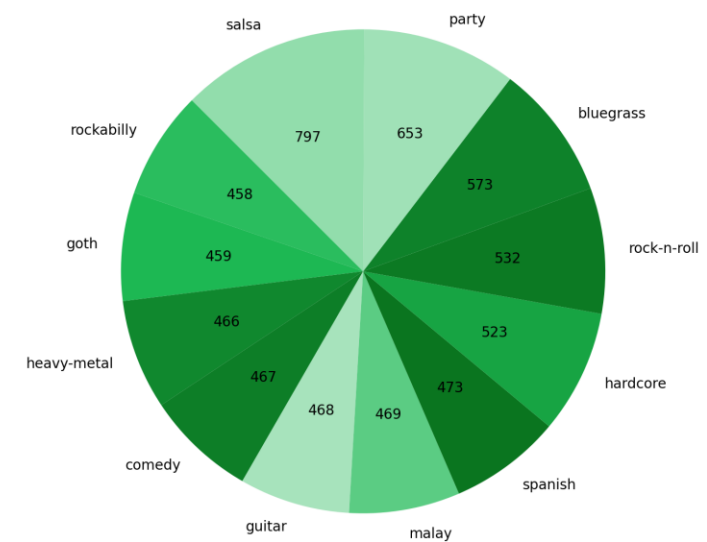
Effettuiamo un'analisi dei generi all'interno delle fasce di popolarità che abbiamo individuato nei grafici precedenti, *dividendo le due fasce 10-35 e 35-100 lungo la metà di ciascuna*, e selezionando i 12 generi più rappresentati.

Notiamo che i generi più rappresentati in ciascuna fascia sono molto vari. Analizziamo più attentamente.

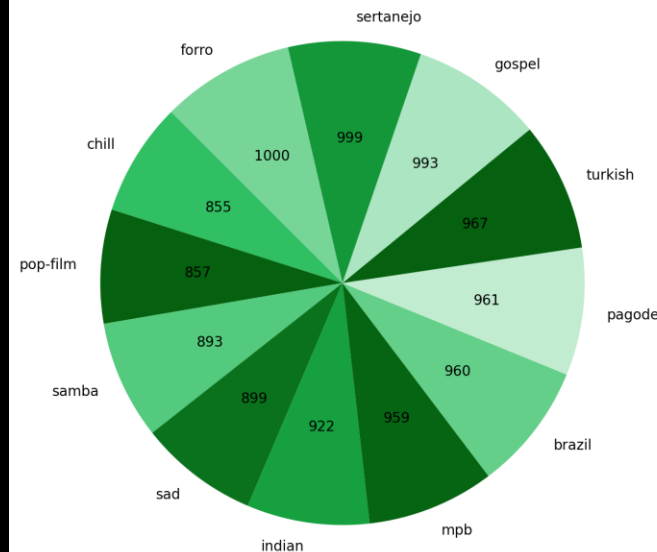
Popolarità 10-22



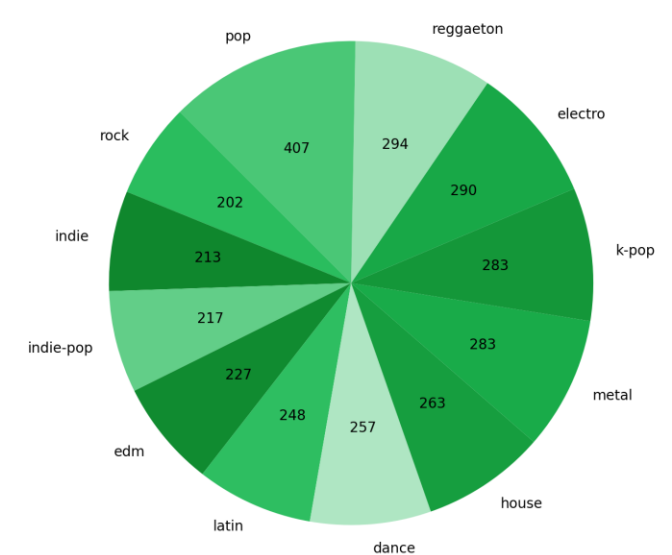
Popolarità 22-35



Popolarità 35-67

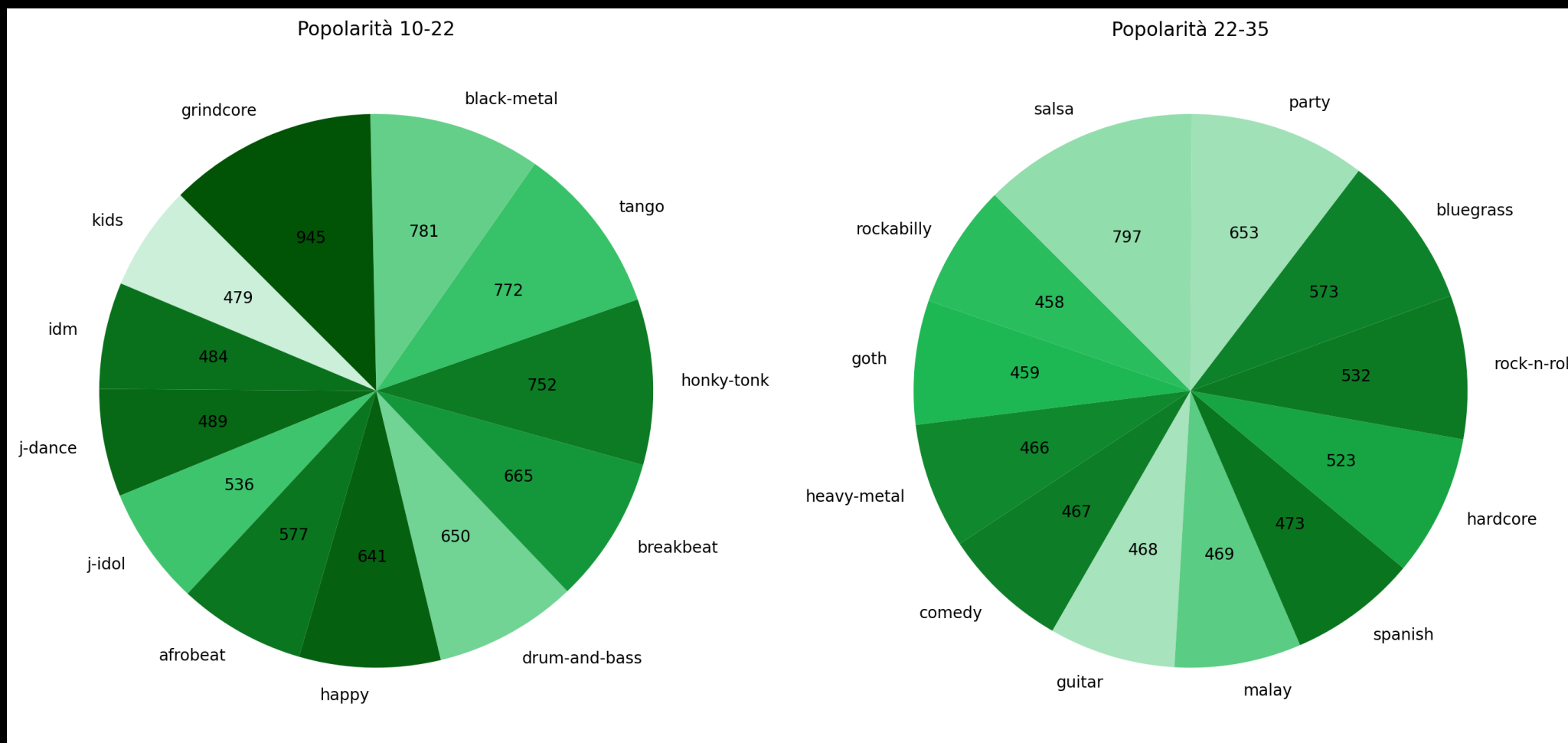


Popolarità 67-100



## Esiste un legame tra genere e popolarità? 2

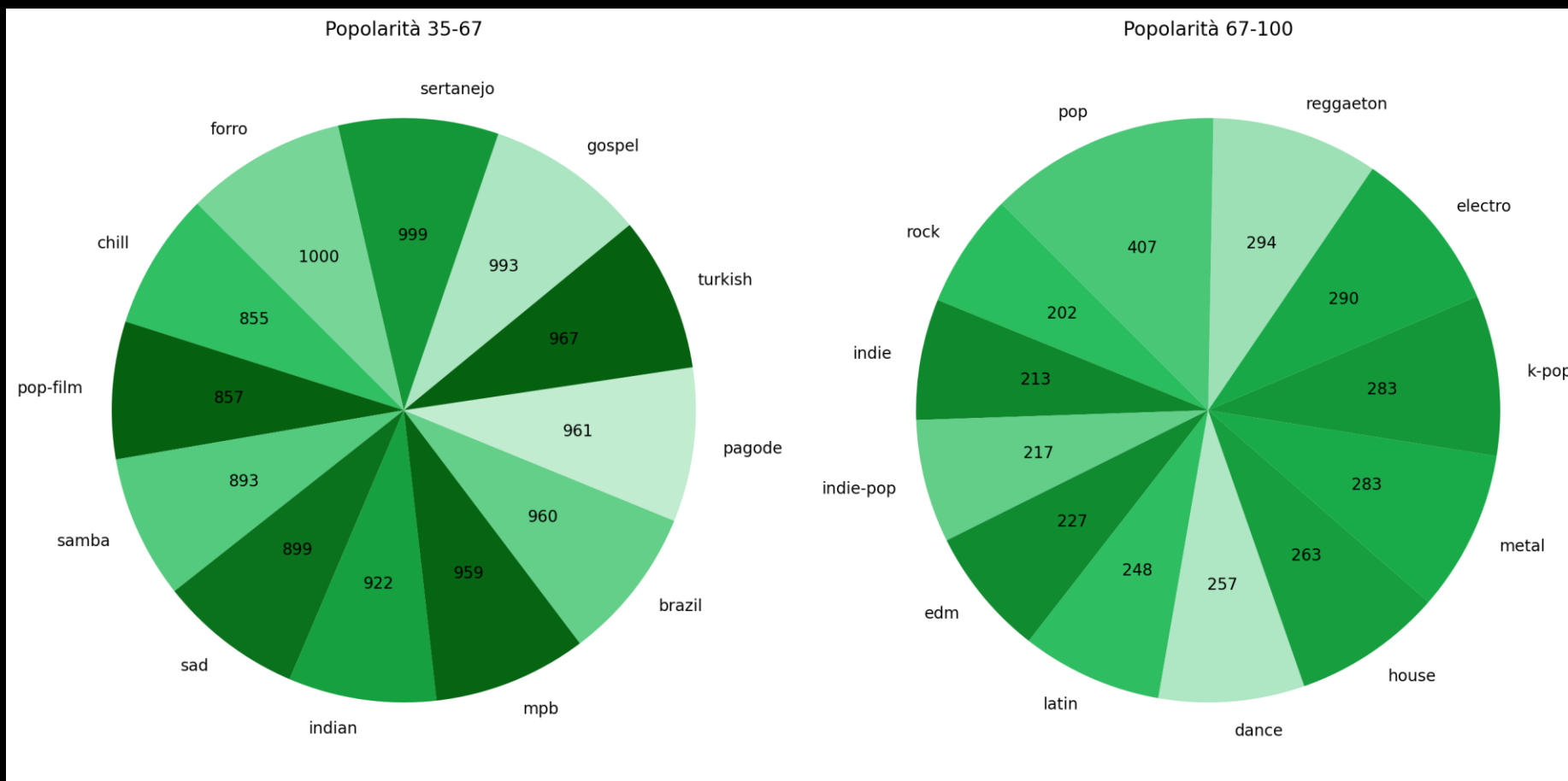
Tra 10 e 35, i generi prevalenti sembrano essere **derivati del metal e dell'EDM**.





## Esiste un legame tra genere e popolarità? 3

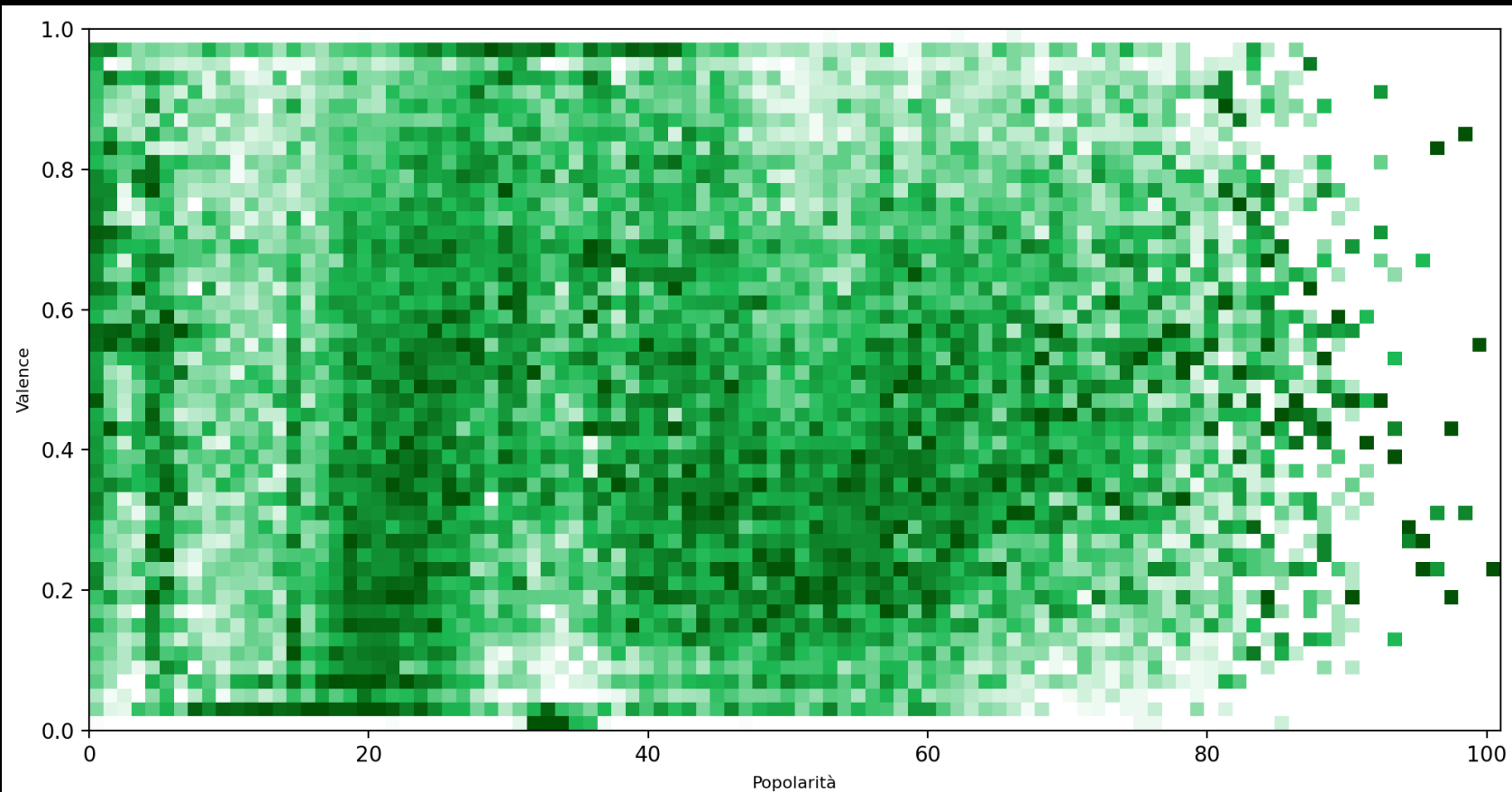
Tra 35 e 67, i generi principali sembrano **musica tipica** di diversi paesi, tra cui indiana, turca e brasiliana (5 generi su 12 sono di musica brasiliana). Tra 67 e 100, si ritrovano principalmente i generi più vicini al **pop** e alla musica **latino-americana**.



# Le canzoni allegre sono più popolari?

Le canzoni **tristi**, serie o “arrabbiate” sembrano essere **apprezzate di più** rispetto a quelle allegre, che rimangono comunque ben rappresentate tra le canzoni più popolari. Tuttavia, le canzoni più popolari in assoluto non sembrano trovarsi mai su nessuno degli estremi.

**Nota:** il numero di brani diminuisce notevolmente all'aumentare della popolarità: un quadrato verde chiaro a sinistra potrebbe contenere molte più tracce che un quadrato nero a destra.

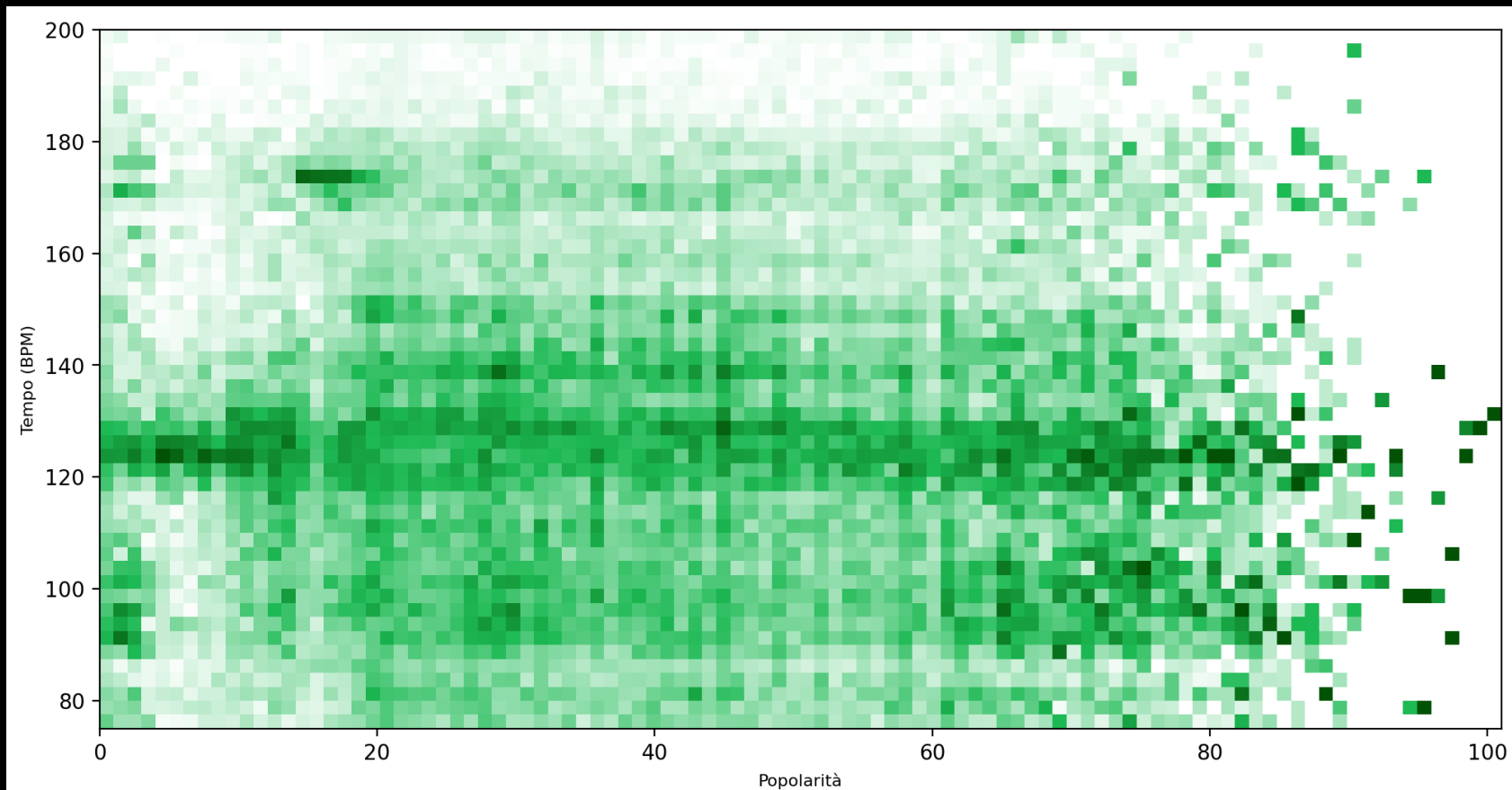


# Il tempo di un brano incide sulla sua popolarità?

Dall'intervallo 0-10 possiamo notare un'elevatissima densità di brani intorno ai 125 BPM\*, tipici della musica da club, tuttavia la popolarità di un brano non è fortemente correlata al tempo (anzi: l'abbondanza dei 125 BPM non ne rispecchia l'apprezzamento).

Tra le canzoni più popolari invece si nota una leggera **preferenza ai BPM tra 90 e 105, tipici del reggaeton**.

**Nota:** il numero di brani diminuisce notevolmente all'aumentare della popolarità: un quadrato verde chiaro a sinistra potrebbe contenere molte più tracce che un quadrato nero a destra.



\*: Battiti al minuto, misura del tempo di un brano

**Grazie per l'attenzione!**