



Università degli Studi di Bari Aldo Moro

## Dataset LFM-1b\_artist

Emanuele Fontana

Tirocinio tesi triennale in Informatica  
Anno accademico 2023/2024

LFM-1b\_artist è un dataset contenente informazioni riguardanti le interazioni tra utenti e artisti musicali. Questo dataset è stato partizionato seguendo due modalità diverse:

- **Random:** Le interazioni sono state selezionate in modo randomico. Non vi è alcuna garanzia che ogni utente sia presente
- **Stratificazione:** Per ogni utente è stato selezionato il numero di interazioni indicato. Si è fatto in modo che ogni utente sia presente

Il dataset inoltre è stato partizionato con i seguenti rapporti:

- **75%:** Sono state selezionate il 75% delle interazioni dal dataset originale
- **50%:** Dal dataset partizionato al 75% sono state selezionate circa il 66,7% delle interazioni, in modo tale da ottenere il 50% delle interazioni dal dataset originale
- **25%:** Dal dataset partizionato al 50% sono state selezionate circa il 50% delle interazioni, in modo tale da ottenere il 25% delle interazioni dal dataset originale

Ecco la tabella con le informazioni riguardanti il dataset:

Partizione	Modalità	Utenti	Item
100%	/	120322	3123496
75%	Random	120303	2627356
50%	Random	120274	2058958
25%	Random	120166	1353238
75%	Stratificazione	120322	2627027
50%	Stratificazione	120322	2058132
25%	Stratificazione	120322	1352544

Tabella 1: Statistiche dataset LFM-1b\_artist

---

**Random75** Come possiamo vedere qui sono contenuti il 99.98% degli utenti originali e l'84.11% degli item originali.

**Random50** Come possiamo vedere qui sono contenuti il 99.96% degli utenti originali e il 65.91% degli item originali.

**Random25** Come possiamo vedere qui sono contenuti il 99.87% degli utenti originali e il 43.42% degli item originali.

**Stratificazione75** Come possiamo vedere qui sono contenuti il 100% degli utenti originali e l'84.10% degli item originali.

**Stratificazione50** Come possiamo vedere qui sono contenuti il 100% degli utenti originali e il 65.89% degli item originali.

**Stratificazione25** Come possiamo vedere qui sono contenuti il 100% degli utenti originali e il 43.30% degli item originali.