Nicolò Giovannacci

EPICODE – DATA ANALYST 02-23 PT

Esercitazione M1-D1
Introduzione al mondo dei dati

Consegna

L'esercizio è mirato a prendere confidenza con il concetto di dato e informazione.

Scegli un argomento di tuo interesse, riporta su un documento testuale quali informazioni ti piacerebbe derivare cercando su internet i potenziali dataset strutturati e non strutturati (almeno 3 del primo tipo, 2 del secondo) utilizzabili. 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'obiettivo di queste analisi.

Svolgimento

Scelgo come argomento di interesse il cinema. Si prendono in considerazione 3 tipi di dataset STRUTTURATI, da usare come base per una analisi sui ricavi registrati nelle sale cinematografiche nel periodo 2015-2020. Tale analisi non verrà eseguita, né ne saranno illustrati nel dettaglio gli obiettivi, in quanto non è questo lo scopo del lavoro, che si pone come unica finalità quella di prendere confidenza con il concetto di dato e informazione.

Di seguito si riportano i 3 dataset STRUTTURATI

Ricavi Dell'industria Cinematografica a Livello Globale		
ID	**Anno**	**Ricavi (miliardi di dollari)**
1	2015	38,3
2	2016	37,6
3	2017	39,4
4	2018	41,4
5	2019	41,7
6	2020	11,9

Numero Degli Iscritti Alle Principali Piattaforme Di Streaming			
ID	**Anno**	**Utenti Amazon Prime (milioni)**	**Utenti Netflix (milioni)**
1	2015	45	70
2	2016	60	89
3	2017	100	110
4	2018	125	139
5	2019	150	167
6	2020	200	203

Per l'ultimo dataset si riportano solo 6 righe della tabella per semplicità. La tabella riporta infatti tutti film usciti ogni anno, dal 2015 al 2020, con i rispettivi incassi.

Incassi Film			
ID	**Anno**	**Film**	**Incasso (milioni di dollari)**
1	2015	Arrival	203
2	2016	Get Out	255
3	2017	Bohemian Rhapsody	910
4	2018	Parasite	263
5	2019	The Father	28
6	2020	Tenet	365
…	…….	…….	…….

In ogni tabella, attraverso l'attributo "ID", si vuole poter identificare ogni riga in maniera univoca. Non sempre ciò è possibile attraverso le singole istanze degli altri attributi: nell'ultima tabella, ad esempio, potrebbero esserci film con lo stesso titolo, con lo stesso incasso o film usciti nello stesso anno; o ancora potremmo avere qualche dato mancante se ad esempio non riuscissimo a reperire l'incasso di qualche film di nicchia.

Per quanto concerne i dati NON STRUTTURATI si scelgono di riportare:

- I film per intero, sottoforma di file video
- Le locandine dei film

Il film oggetto di entrambi i dataset sono tutti i film usciti dal 2015 al 2020, coerentemente ai dataset strutturati.

Le analisi sui dataset NON STRUTTURATI sono eseguibili principalmente tramite intelligenza artificiale. Per quanto concerne i file video si potrebbe eseguire una analisi sul montaggio video dei film, per scoprire se, ad esempio, film con un montaggio più frenetico portano lo spettatore ad apprezzare di più pellicola, che quindi potrebbe anche per questo motivo vantare incassi più alti.

Per quanto riguarda le locandine dei film si potrebbe eseguire una analisi volta a classificare le locandine stesse in elaborate (ovvero ricche di soggetti/elementi raffigurati) o minimali, per scoprire quale delle due tipologie di locandina possa attrarre di più lo spettatore in sede di pubblicità e di come questo fattore possa contribuire ad un aumento degli incassi dei film nelle sale.