INTRODUZIONE AI DATI

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 e cerca su internet i potenziali dataset utilizzabili. Puoi per esempio cercare sul sito kaggle.com, o yahoo finance, eurostat, our world in data. Per ogni dataset descrivi le analisi che svolgeresti su di esso e l’obiettivo di queste analisi, oltre alle figure che professionali di cui pensi di necessitare.

(Riflettere sul tipo di informazioni che vogliamo ricavare. Ex. Vendita di auto, potremmo voler sapere: il numero di auto vendute al mese; la regione o la provincia da cui provengono le persone che comprano le auto…) N.B. non scegliere dataset oltre i 10 Mb, vanno bene i Kb di dati.

DATASET SCELTO SU KAGGLE: ***Netflix Movies and Shows***

Da questo dataset potremmo ricavare le seguenti informazioni:

* Film più votati
* Top 10 film
* Generi più graditi in base ai voti
* Paesi che producono maggior numero di film
* Vedere correlazione tra gradimento e durata degli show
* Anni in cui sono stati rilasciati più film
* Quante serie sono state cancellate dopo la prima stagione
* Vedere se ci sono più film o show sulla piattaforma

Figure professionali di cui necessitiamo:

* **Data analyst**: si occupa della data cleaning, selezione dei dati rilevanti e incrociando i dati rilevanti produce grafici. Ex. Elimina un campo non rilevante per l’analisi;
* **Data scientist**: fa delle predizioni in base ai trend emersi dall’analisi. Ex. Analizza i grafici inseriti nelle dashboard e predice possibili andamenti del settore cinematografico;
* **Data journalist**: si occupa del report in cui espone il risultato dell’analisi. Ex. Scrive un articolo su una serie/ un genere analizzando i grafici emersi.

CONSEGNA 2: scegliere un altro dataset relativo all’argomento del precedente e capire come connetterlo al primo.

DATASET N.2 ***Disney+ Movies and Shows***

Per effettuare il confronto scegliamo dei campi comuni, ad esempio titolo, tipo, anno di rilascio, durata, genere, produzione per paese, numero di stagioni, voto su imdb. Ad esempio, attraverso la variabile titolo possiamo vedere se un film o uno show è presente su entrambe le piattaforme; possiamo confrontare i generi per vedere quali sono i più graditi sulle piattaforme; possiamo vedere quale piattaforma ha la media voti più alta e quindi presumere quale delle due fornisce prodotti di qualità maggiore.