ESERCIZIO 1

L’argomento che ho preso in considerazione per questo esercizio è “Produzione e lettura dei libri in Italia nell’anno 2021.

Con 90.195 opere librarie pubblicate, il 2021 è caratterizzato da una forte ripresa della produzione editoriale, non solo nel confronto con il primo anno segnato dal Covid-19 ma anche rispetto alla situazione precedente (+8,2% sul 2019). Il maggior incremento riguarda l’editoria per bambini e ragazzi: +17,7% le opere pubblicate e +5,6& le copie stampate.

Come primo esempio di dataset strutturato riporto:



In questo dataset gli attributi sono:

* INDICATORE;
* VALORE ASSOLUTO;
* VARIAZIONE PERCENTUALE RISPETTO AL 2020;
* EDITORI E PRODUZIONE DI LIBRI;
* TIPOLOGIA OPERE;
* LETTURA DI LIBRI (persone di 6 anni e più).

Mentre gli identificativi sono:

* Opere pubblicate;
* Copie stampate (in migliaia);
* Editori attivi (attività principale);
* Prime edizioni;
* Edizioni successive;
* Ristampe;
* Lettori/lettrici;
* Lettori/lettrici di libri cartacei;
* Lettori/lettrici di e-book/libri on line;
* Persone che ascoltano audiolibri.

Un altro esempio di dataset strutturato che mostra i titoli pubblicati e la tiratura per genere nel 2021 in valori percentuali:

Immagine che contiene testo, schermata, numero, Carattere

Descrizione generata automaticamente

In questo dataset gli attributi sono:

* Opere di genere varia adulti;
* Opere di genere per bambini e ragazzi;
* Opere di genere scolastico.

Mentre gli identificativi sono:

* % tiratura;
* % numero titoli pubblicati.

L’ultimo esempio di dataset strutturato che ho trovato è:

Immagine che contiene testo, schermata, numero

Descrizione generata automaticamente

Questo dataset mostra i canali di commercializzazione diversificati per editore, in questo caso gli attributi sono:

* Micro editori;
* Piccoli editori;
* Medi editori;
* Grandi editori;
* Totale.

Mentre gli identificativi sono:

* Nessuna;
* Fino al 10%;
* Dall’11% al 25%;
* Dal 26% al 50%;
* Dal 51% al 75%;
* Dal 75% al 90%;
* Dal 91% al 99%.

Come esempi di dataset non strutturati ho pensato di prendere queste due immagini:

Immagine che contiene testo, libro, Pubblicazione, Libro di self-help

Descrizione generata automaticamente

Su questa prima immagine l’analisi che svolgerei è ad esempio per ogni titolo di libro andare a vedere le copie vendute per ogni regione italiana e andare a fare una media, in base sempre alle copie vendute, del genere letterario “preferito” in quella regione.

Un altro dataset non strutturato potrebbe essere quest’altra immagine:

Immagine che contiene scena, libro, interno, biblioteca

Descrizione generata automaticamente

Si tratta della biblioteca Nazionale francese, in questo caso andrei ad analizzare gli ingressi di persone residenti in Francia per fascia di età e di genere, il tempo di permanenza e magari il genere di libri presi in prestito per andare a studiare le abitudini di lettura francesi.

Un’altra analisi possibile potrebbe essere sulla fascia oraria degli ingressi in biblioteca per studiare le abitudini giornaliere e lo spazio che ha l’hobby della lettura nella vita della popolazione francese.