# DESCRIZIONE GENERALE DEGLI STEP PROGETTUALI

Innanzitutto, ho utilizzato Python per estrarre il dataset direttamente dall'API di Gamalytic. Il codice Python che ho scritto ha permesso di richiedere i dati e trasformarli in un formato tabellare, pronto per ulteriori analisi. Ogni riga del dataset corrisponde a un videogioco e contiene varie informazioni come il nome del gioco, il prezzo, il numero di copie vendute, e molte altre.

Il codice Python è visibile nel file ***EstrazioneDatiDaWeb.ipynb***.

Il dataset estratto tramite Python è stato esportato in un file Excel, denominato “**DatasetPartenza.xlsx**”.

Tra i campi più interessanti troviamo il campo “tags”, che contiene una lista di tag associati a ciascun gioco, come "3D", "Multiplayer", "Magic", e così via. Ogni gioco può avere anche più di 10 tag e questi possono essere ripetuti tra i vari giochi. I tag rappresentano o un genere specifico del gioco oppure un suo attributo.

Il file “**Tags\_details**” contiene una lista di tutti i possibili tags, con la relativa classificazione in termini di attributo o genere.

Il campo “genres” del dataset di partenza associa ad ogni videogioco una lista di “macro-generi”, che non sono stati utilizzati per le analisi, in quanto i generi indicati per ogni videogioco dal campo “tags” risultavano avere un livello di dettaglio superiore, trattandosi di “sottogeneri”.

Una volta estratto il dataset principale, l'ho importato in Power Query Editor per fare delle trasformazioni. Una delle principali operazioni è stata la creazione di campi booleani. Ho creato le colonne personalizzate “3D”, “Multiplayer” e “FirstPerson”, che assumono il valore "Sì" se il gioco ha quel tag nella propria lista, altrimenti "No". Questi campi sono stati fondamentali per effettuare il filtraggio delle informazioni nelle dashboard.

A questo punto ho creato la prima dashboard in PowerBI, per un’analisi globale, con filtri basati su date e sui campi booleani appena creati. I campi principali oggetto dell’analisi sono:

* **numero di giochi rilasciati**;
* **copie vendute**;
* **fatturato**;
* **numero medio di ore di gioco**;
* **rating**;

Per rendere possibile la visualizzazione di tutte queste informazioni in una sola dashboard è stata necessaria la creazione di misure apposite che adattassero il grafico al campo selezionato, (per la struttura del dataset non avrebbe avuto senso studiare l’andamento nel tempo delle copie vendute né del fatturato, in quanto le date di acquisto dei videogiochi non sono note, motivo per cui questi due campi risultano vuoti nel grafico temporale).

Per poter analizzare i generi in modo approfondito, ho creato un dataset specifico a partire dal dataset principale e dalla tabella sui tags. Proprio da quest’ultima ho selezionato i tag che rappresentano generi e li ho trasformati in nuove colonne nel dataset principale. Ogni colonna rappresenta un genere. Successivamente, ho riempito le celle in modo tale che ogni cella contenesse un valore numerico se il videogioco di quella riga aveva quel genere nella lista dei tag, altrimenti riportava un errore.

Questo dataset è memorizzato all’interno del file Excel “**Genres.xlsx**”.

Per semplificare il dataset ed eliminare i campi ridondanti, ho eseguito l'unpivoting in Power Query Editor ed ho eliminato tutte le righe che contenevano errori, cioè quei giochi che non avevano il genere associato. Questo ha trasformato le varie colonne dei generi in righe, creando un dataset con due colonne: una per l’ID del gioco e una per il genere. Ogni riga rappresenta un’associazione gioco-genere. Ad esempio, il gioco 0 ha nel campo tags questi 5 generi

Dopo questa pulizia, ho eseguito una left-outer join tra questo dataset con ID e generi e il dataset principale. In questo modo, ho ottenuto una versione finale in cui ogni videogioco aveva tante righe quante erano i generi associati, con una colonna che riportava, per ogni riga, il genere corrispondente, e tutte le altre informazioni derivanti dal dataset principale.

Ho quindi importato il nuovo dataset in PowerBI e ho creato una dashboard specifica per l’analisi dei generi. In modo simile a quella creata per l’analisi globale, anche questa dashboard permette di filtrare e analizzare le performance dei videogiochi ponendo però il focus sul loro genere.

Dopo aver completato l’analisi dei generi, ho ripetuto lo stesso procedimento anche per gli attributi dei giochi. Ho creato un dataset simile su Excel, l’ho caricato sulla PQE e ne ho fatto l’unpivoting delle colonne degli attributi. Ho unito anche questo al dataset principale per creare una terza dashboard che permette l’analisi degli attributi, proprio come fatto per i generi.

Tale dataset è memorizzato nel file Excel “**Attribute.xlsx**”.

# LINK DI ACCESSO AL DRIVE CONTENENTE I FILE DEL PROGETTO

A causa delle dimensioni considerevoli dell’elaborato progettuale, è stato creato un drive accessibile a chiunque clicchi sul seguente link:

<https://drive.google.com/drive/folders/1ZL4lvMBu8qn2oWZjjhaEGFloU_DXLLku?usp=sharing>