

Videogames stats analysis

Analisi esplorativa sulle statistiche di Videogiochi con focus sui molteplici aspetti che li riguardano.

Obiettivi:

- Analizzare i vari aspetti riguardanti i videogiochi come: piattaforma, developer, genere, anno di lancio, le valutazioni ecc., con lo scopo di offrire una ampia visione sulle possibili tendenze di mercato. (Metacritic)
- Determinare l'eventuale relazione tra la valutazione di un gioco e il suo successo sul mercato, misurato in ore di gioco effettuate. (Metacritic e Steam)

Dataset utilizzati:

- [Metacritic Rating dataset su Kaggle per videogiochi multiplatforma](#)

Questo dataset contiene diverse valutazioni e feedback degli utenti su diversi videogiochi multiplatforma, si prenderanno in esame solo i dati provenienti da videogiochi giocati sulla piattaforma PC Steam. Tali informazioni vengono utilizzate per calcolare il punteggio di valutazione normalizzato per ciascun videogioco in questa analisi.

- [Steam Game Played by Hours DataSet su Kaggle](#)

Questo dataset contiene una lista di videogiochi su Steam insieme al numero di ore di gioco effettuate da diversi utenti. Queste informazioni vengono utilizzate per calcolare il punteggio di gioco normalizzato per ogni videogioco in questa analisi.

Strumenti utilizzati: Per effettuare l'analisi si utilizzerà MySQL come editor per la pulizia, l'elaborazione e l'analisi dei dati. Per la presentazione dei risultati verrà utilizzato Power BI Desktop.

Fasi del progetto:

1. **Importazione dei Dataset:** i dataset iniziali verranno importati in MySQL in cui verranno create le relative tabelle.
2. **Pulizia, unione dati ed analisi:** utilizzando MySQL si eseguiranno le operazioni necessarie per pulire i dati ed unire i due dataset. Si inizierà, ad esempio, con la pulizia di valori nulli e l'eliminazione di dati superflui che non interessano all'analisi in oggetto; Si potrebbe calcolare la media delle ore di gioco per ciascun videogioco, i videogiochi più giocati in base al numero di ore totali, ed i videogiochi con le migliori e peggiori valutazioni.
3. **Esportazione dei risultati:** i risultati dell'analisi saranno esportati in Power BI Desktop per creare una dashboard interattiva.
4. **Creazione dashboard:** utilizzando Power BI Desktop si creeranno visualizzazioni interattive, tabelle e grafici per presentare i risultati dell'analisi svolta in modo chiaro ed efficace.

Fase 1: Importazione dei Dataset:

Per la prima fase del progetto si utilizzerà MySQL. In questa fase si andranno ad importare i due dataset, verranno quindi create le due relative tabelle contenenti i dati sulle quali verranno eseguite successivamente le analisi.

Esempio di creazione di una tabella:

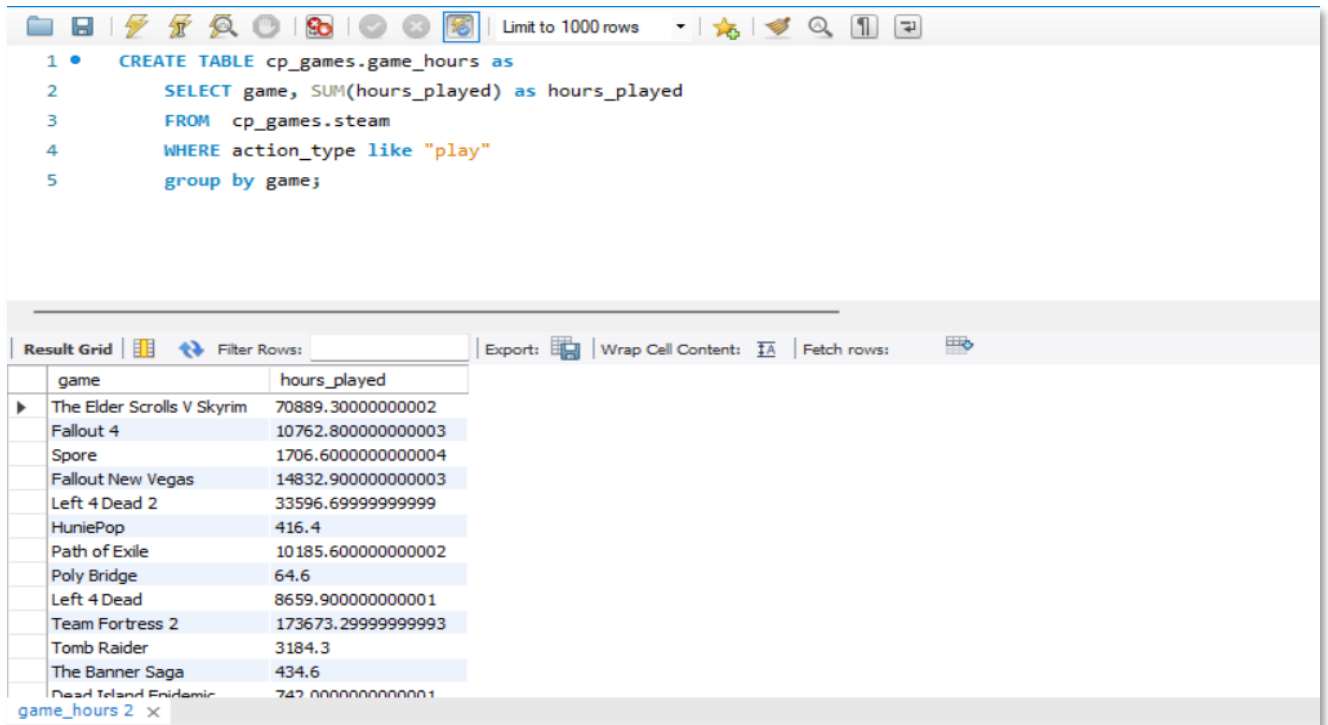
DDL for cp_games.metacritic

```
1 CREATE TABLE `metacritic` (  
2   `game` text,  
3   `platform` text,  
4   `developer` text,  
5   `genre` text,  
6   `number_players` text,  
7   `rating` text,  
8   `release_date` text,  
9   `positive_critics` int DEFAULT NULL,  
10  `neutral_critics` int DEFAULT NULL,  
11  `negative_critics` int DEFAULT NULL,  
12  `positive_users` int DEFAULT NULL,  
13  `neutral_users` int DEFAULT NULL,  
14  `negative_users` int DEFAULT NULL,  
15  `metascore` int DEFAULT NULL,  
16  `user_score` int DEFAULT NULL  
17 ) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci
```

Fase 2: Pulizia, unione dati ed analisi:

In questa fase, per iniziare si andranno a pulire i dati importati dai valori nulli, da quelli ridondanti e da quelli superflui e non interessanti per l'analisi.

Esempio di pulizia dei dati:



```
1 • CREATE TABLE cp_games.game_hours as
2     SELECT game, SUM(hours_played) as hours_played
3     FROM cp_games.steam
4     WHERE action_type like "play"
5     group by game;
```

game	hours_played
The Elder Scrolls V Skyrim	70889.300000000002
Fallout 4	10762.800000000003
Spore	1706.6000000000004
Fallout New Vegas	14832.900000000003
Left 4 Dead 2	33596.699999999999
HuniePop	416.4
Path of Exile	10185.600000000002
Poly Bridge	64.6
Left 4 Dead	8659.900000000001
Team Fortress 2	173673.29999999993
Tomb Raider	3184.3
The Banner Saga	434.6
Dead Island Epidemic	742.0000000000001

In questo caso ho eliminato i dati ridondanti tenendo conto solo quelli che avevano nella colonna *action* il valore "play", facendo inoltre la somma delle ore giocate per ogni gioco dai vari *user_id* presenti nel dataset, salvando il tutto in una nuova tabella utile per le future analisi.

Proseguendo nell'elaborazione si andrà a calcolare lo *score* per ogni gioco al fine di ottenere una sorta di classifica. Per calcolare lo score, partendo dai dati relativi alle valutazioni presenti nella tabella *metacritic*, assegneremo i punti come segue:

- +1 per ogni voto positivo da parte di critica e user
- +0.5 per ogni voto neutrale da parte di critica e user
- -1 per ogni voto negativo da parte di critica e user

Il risultato ottenuto verrà quindi inserito nella nuova tabella creata.

Calcolo dello score per ogni gioco:

```
1 CREATE TABLE game_score as
2 SELECT game, (o.positive_critics + o.neutral_critics + o.negative_critics + o.positive_users + o.neutral_users + o.negative_users) as score
3 FROM
4 (SELECT game, positive_critics, neutral_critics*0.5 as neutral_critics, negative_critics*-1 as negative_critics, positive_users, neutral_users*0.5 as neutral_users, negative_users*-1 as negative_users
5 FROM cp_games.game_critics) as o
6 ORDER BY score DESC;
```

Result Grid

game	score
Portal 2	1786.0
The Elder Scrolls V: Skyrim	1358.0
The Witcher 3: Wild Hunt	1277.0
The Witcher 2: Assassins of Kings	1214.5
BioShock Infinite	1174.5
Battlefield 3	946.0
Undertale	786.0
Guild Wars 2	632.5
Tomb Raider	563.0
Deus Ex: Human Revolution	550.0
Terraria	522.5
Dishonored	510.5
Torchlight II	491.0

Si prosegue con la join delle due tabelle ottenute:

```
1 SELECT distinct gs.game, gs.score, gh.hours_played
2 FROM cp_games.game_score as gs
3 INNER JOIN cp_games.game_hours as gh
4 ON gs.game = gh.game
5 ORDER BY gs.score DESC;
```

Result Grid

game	score	hours_played
Portal 2	1786.0	9117.099999999997
BioShock Infinite	1174.5	3689.1000000000013
Undertale	786.0	444.6000000000001
Tomb Raider	563.0	3184.3
Terraria	522.5	29951.7999999999985
Dishonored	510.5	2842.2
Torchlight II	491.0	6891.9
Far Cry 3	478.5	5076.000000000002
Borderlands 2	424.5	22667.8999999999994
The Secret World	424.0	358.69999999999993
Pillars of Eternity	392.0	1523.2
Bastion	332.0	651.90000000000003
Path of Exile	292.0	10185.6000000000007

Fase 3: Esportazione dei dati:

A questo punto, avendo già diversi dati da poter visualizzare, andiamo a connettere il DB creato in MySQL e salvato sul nostro server locale, con Power BI, così da poter avere sempre aggiornati in tempo reale i dati derivanti dalle possibili future operazioni svolte in MySQL.

Database MySQL

Server

127.0.0.1:3306

Database

cp_games

Opzioni avanzate

OK

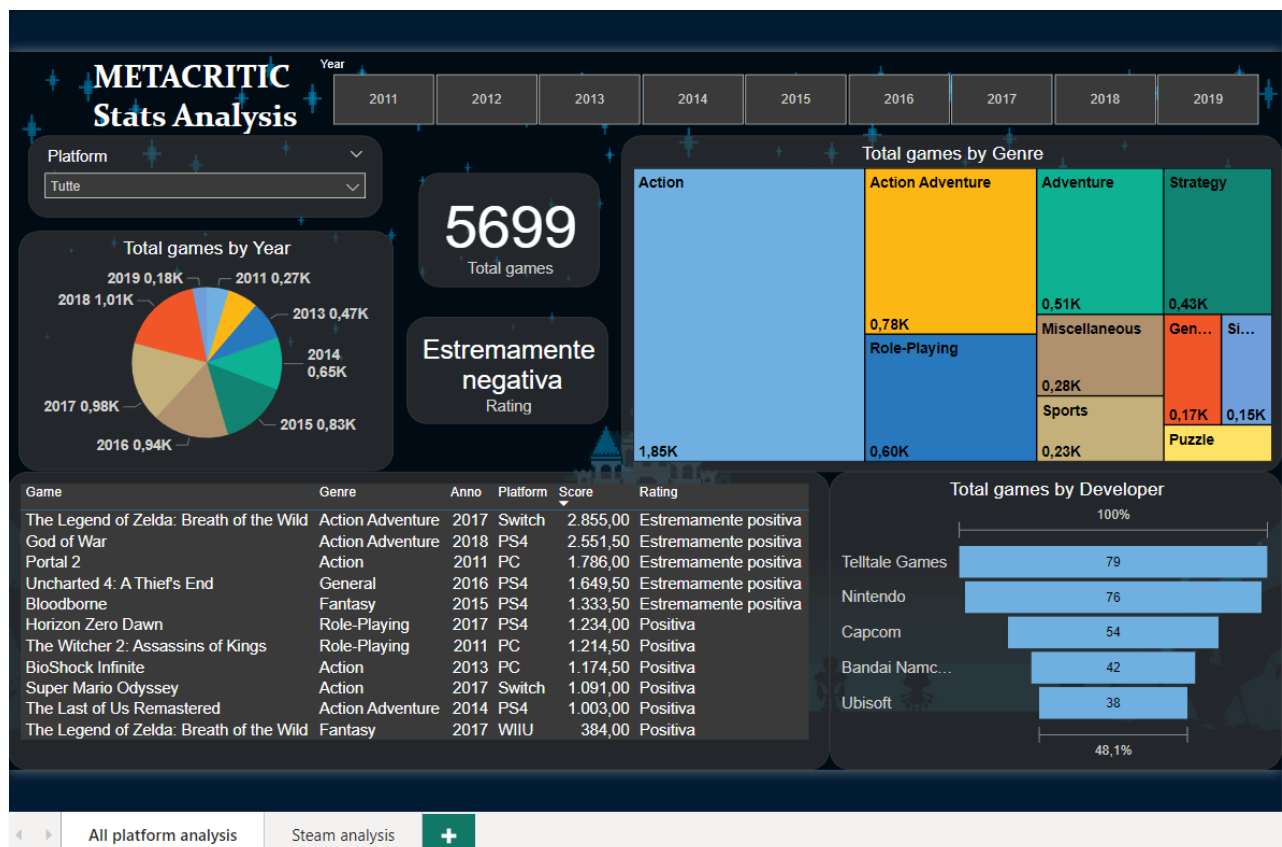
Annulla

Fase 4: Creazione della dashboard:

Con i dati ricavati dalle analisi svolte in MySQL cominciamo la creazione della dashboard interattiva.

Sulla prima pagina analizzeremo il dataset di Metacritic, applicheremo dei filtri che l'utente potrà utilizzare per avere un focus su argomenti di interesse, come quello per anno di pubblicazione del videogioco e la piattaforma di destinazione. Andremo ad inserire quindi grafici e schede che l'utente potrà analizzare.

È stato poi creato un cluster per rendere parlante il dato dello score che in base al punteggio viene definito: estremamente negativo, negativo, neutrale, positivo, estremamente positivo.



Nella seconda pagina invece, andremo ad analizzare più nello specifico i dati ricavati dalla join tra lo score ottenuto dai calcoli sulle valutazioni presenti nel dataset di Metacritic e il totale delle ore giocate per ogni videogioco ottenuto dal dataset di Steam.



In conclusione, si osserva che non c'è una correlazione vera e propria tra il videogioco con lo score migliore e quello con più ore giocate; questo dipende da vari fattori come:

- il genere del gioco: un gioco online competitivo richiederà molte più ore di gioco rispetto ad un gioco di avventura single player.
- L'anno di pubblicazione: con molta probabilità un gioco uscito da più tempo avrà più ore di gioco rispetto agli ultimi pubblicati.