

PREDICT GAMING BEHAVIOR DATASET

Questo dataset raccoglie metriche e dati demografici completi relativi al comportamento dei giocatori negli ambienti di gioco online. Include variabili quali dati demografici dei giocatori, dettagli specifici sui giochi, metriche di coinvolgimento e una variabile target che riflette la fidelizzazione dei giocatori.

CARATTERISTICHE & POTENZIALI APPLICAZIONI

- **Modellizzazione predittiva** dei modelli di fidelizzazione e coinvolgimento dei giocatori.
- **Analisi dei fattori** che influenzano il comportamento dei giocatori e le prestazioni di gioco.
- **Ottimizzazione** del design del gioco, delle strategie di marketing e dei miglioramenti dell'esperienza dei giocatori.

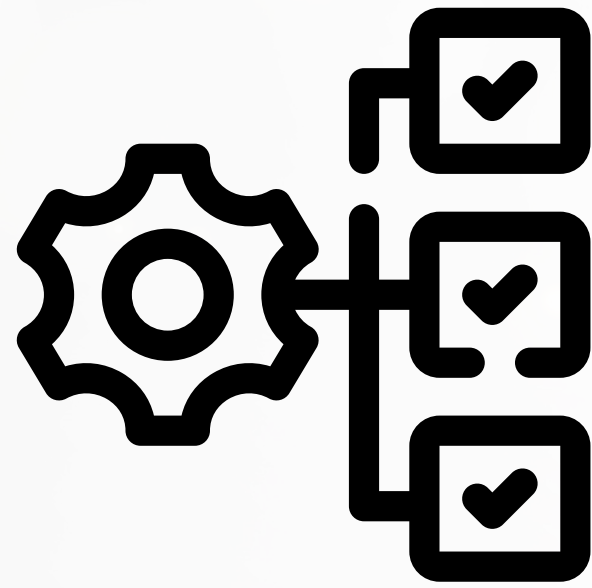
Utilizzo

Il Dataset è adatto per esplorare i modelli di comportamento nel gioco online, sviluppare modelli di apprendimento automatico per la previsione del coinvolgimento dei giocatori e condurre ricerche nell'ambito dell'analisi dei giochi.

- **PLAYERID:** UNIQUE IDENTIFIER FOR EACH PLAYER.
- **AGE:** AGE OF THE PLAYER.
- **GENDER:** GENDER OF THE PLAYER.
- **LOCATION:** GEOGRAPHIC LOCATION OF THE PLAYER.
- **GAMEGENRE:** GENRE OF THE GAME THE PLAYER IS ENGAGED IN.
- **PLAYTIMEHOURS:** AVERAGE HOURS SPENT PLAYING PER SESSION.
- **INGAMEPURCHASES:** INDICATES WHETHER THE PLAYER MAKES IN-GAME PURCHASES (0 = NO, 1 = YES).
- **GAMEDIFFICULTY:** DIFFICULTY LEVEL OF THE GAME.
- **SESSIONSPERWEEK:** NUMBER OF GAMING SESSIONS PER WEEK.
- **AVGSESSIONDURATIONMINUTES:** AVERAGE DURATION OF EACH GAMING SESSION IN MINUTES.
- **PLAYERLEVEL:** CURRENT LEVEL OF THE PLAYER IN THE GAME.
- **ACHIEVEMENTSUNLOCKED:** NUMBER OF ACHIEVEMENTS UNLOCKED BY THE PLAYER.
- **ENGAGEMENTLEVEL:** CATEGORIZED ENGAGEMENT LEVEL REFLECTING PLAYER RETENTION ('HIGH', 'MEDIUM', 'LOW').

ABOUT

TEAM MEMBERS



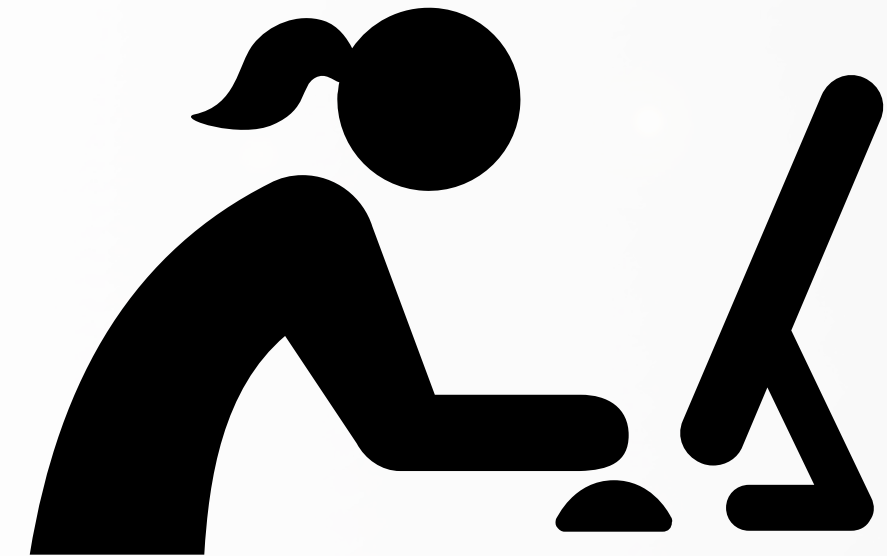
PROJECT MANAGER

Alessio



TECH LEAD

Francesco



DEVELOPERS

Gaia
Michela

IL PROGETTO



La prima query è stata pensata per analizzare in modo approfondito i profili dei giocatori, mettendo insieme diverse dimensioni dei dati disponibili. Per fare questo, abbiamo utilizzato una struttura con più CTE, cioè sotto-query organizzate con la clausola WITH, che ci permettono di suddividere logicamente le informazioni in quattro macro-aree distinte, ma tra loro complementari. Una volta definite queste quattro aree, le informazioni vengono messe in relazione tra loro attraverso il PlayerID, così da creare una visione integrata per ogni singolo utente. Questo ci consente di effettuare un'analisi trasversale che combina anagrafica, comportamento, risultati e spesa. L'obiettivo principale di questa query è cercare di individuare relazioni significative tra alcuni di questi fattori, in particolare tra:

- L'età del giocatore,
- La durata media delle sue sessioni,
- Il numero di achievement che ha sbloccato,
- E gli acquisti effettuati all'interno del gioco.

IL PROGETTO



La seconda query Abbiamo sviluppato una query che ci permette di analizzare i comportamenti dei nostri giocatori, mettendo in relazione due variabili fondamentali: il genere del gioco (cioè il tipo, come azione, strategia, ecc.) e il genere del giocatore (maschio o femmina). L'obiettivo è capire quanto siano difficili i giochi utilizzati e quanto coinvolgano i giocatori, e fare questo non solo in termini numerici, ma anche in modo da poter tradurre i risultati in etichette comprensibili e subito leggibili, come "basso", "medio" o "alto". Per farlo, abbiamo trasformato le etichette testuali presenti nel dataset (come "facile", "medio", "difficile" per la difficoltà del gioco, e "basso", "medio", "alto" per il coinvolgimento) in valori numerici. In questo modo possiamo calcolare delle medie per ciascun gruppo. A partire da queste medie, abbiamo poi riclassificato i risultati in categorie qualitative: ad esempio, se il livello medio di coinvolgimento è inferiore a 2, viene classificato come "basso"; se è tra 2 e 3, diventa "medio"; e se è pari o superiore a 3, viene definito "alto". Lo stesso principio è stato applicato alla difficoltà. Questo tipo di analisi ci permette, ad esempio, di capire se un certo tipo di gioco tende a generare più coinvolgimento tra le donne o tra gli uomini, oppure se i giochi più difficili coincidono con i livelli di engagement più alti.

QUERY 1

-- Selezione dati demografici dei giocatori

WITH Demographics AS (
SELECT PlayerID, Age, Gender, Location
FROM gaming_behavior),

-- Selezione informazioni comportamentali

Behavior AS (SELECT PlayerID, PlayTimeHours, AvgSessionDurationMinutes,
SessionsPerWeek
FROM gaming_behavior),

-- Selezione statistiche di gioco

Stats AS (SELECT PlayerID, PlayerLevel, AchievementsUnlocked
FROM gaming_behavior),

-- Selezione dati relativi al tasso di partecipazione e acquisti in game

Rate AS (SELECT PlayerID, InGamePurchases, EngagementLevel, GameGenre,
GameDifficulty
FROM gaming_behavior)

-- Confronto tra i dati indicati per evidenziare la correlazione tra gli acquisti, la durata

-- media di gioco, l'età e gli achievement sbloccati

SELECT b.PlayerID, s.AchievementsUnlocked , b.AvgSessionDurationMinutes, d.Age,
r.IngamePurchases

FROM Behavior b

JOIN Stats s on b.PlayerID = s.PlayerID

JOIN Demographics d on b.PlayerID = d.PlayerID

JOIN Rate r on b.PlayerID = r.PlayerID

ORDER BY s.AchievementsUnlocked DESC, r.InGamePurchases DESC;

QUERY 2

1. **SELECT** GameGenre,
2. -- Selezione il genere del gioco (es: azione, strategia, ecc.) Gender,
3. -- Selezione il genere del giocatore (maschio o femmina)
4. -- Calcolo la difficoltà media, trasformando le etichette testuali in valori numerici
5. **AVG**(**CASE WHEN** GameDifficulty = 'Easy' **THEN** 1 **WHEN** GameDifficulty = 'Medium' **THEN** 2 **WHEN** GameDifficulty = 'Hard' **THEN** 3 **END**) **AS** Average_Difficulty,
6. -- Ottengo così la difficoltà media per ciascun gruppo
7. -- Calcolo anche il coinvolgimento medio, trasformando i livelli in numeri
8. **AVG**(**CASE WHEN** EngagementLevel = 'Low' **THEN** 1 **WHEN** EngagementLevel = 'Medium' **THEN** 2 **WHEN** EngagementLevel = 'High' **THEN** 3 **END**) **AS** Average_Engagement
9. -- Ottengo il coinvolgimento medio per genere e tipo di gioco
10. -- In base alla media di engagement, assegno un'etichetta descrittiva
11. **CASE WHEN** **AVG**(**CASE WHEN** EngagementLevel = 'Low' **THEN** 1 **WHEN** EngagementLevel = 'Medium' **THEN** 2 **WHEN** EngagementLevel = 'High' **THEN** 3 **END**) < 2 **THEN** 'Low' **WHEN** **AVG**(**CASE WHEN** EngagementLevel = 'Low' **THEN** 1 **WHEN** EngagementLevel = 'Medium' **THEN** 2 **WHEN** EngagementLevel = 'High' **THEN** 3 **END**) >= 2 **AND** **AVG**(**CASE WHEN** EngagementLevel = 'Low' **THEN** 1 **WHEN** EngagementLevel = 'Medium' **THEN** 2 **WHEN** EngagementLevel = 'High' **THEN** 3 **END**) < 3 **THEN** 'Medium' **ELSE** 'High' **END AS** Corresponding_engagement,
12. -- Mostro l'engagement medio in forma testuale
13. -- Faccio lo stesso con la difficoltà: traduco il valore medio in un'etichetta
14. **CASE WHEN** **AVG**(**CASE WHEN** GameDifficulty = 'Easy' **THEN** 1 **WHEN** GameDifficulty = 'Medium' **THEN** 2 **WHEN** GameDifficulty = 'Hard' **THEN** 3 **END**) < 2 **THEN** 'Easy' **WHEN** **AVG**(**CASE WHEN** GameDifficulty = 'Easy' **THEN** 1 **WHEN** GameDifficulty = 'Medium' **THEN** 2 **WHEN** GameDifficulty = 'Hard' **THEN** 3 **END**) >= 2 **AND** **AVG**(**CASE WHEN** GameDifficulty = 'Easy' **THEN** 1 **WHEN** GameDifficulty = 'Medium' **THEN** 2 **WHEN** GameDifficulty = 'Hard' **THEN** 3 **END**) < 3 **THEN** 'Medium' **ELSE** 'Hard' **END AS** Corresponding_difficulty
15. -- Mostro la difficoltà media in forma testuale **FROM** online_gaming_behavior_dataset
16. -- Utilizzo il dataset principale sul comportamento nel gaming online
17. **GROUP BY** GameGenre, Gender
18. -- Raggruppo i dati per genere di gioco e genere del giocatore **ORDER BY** GameGenre, Gender;
19. -- Ordino i risultati in modo leggibile, prima per genere di gioco poi per genere