DIVYE BAJAJ
NSUT'26
LINKEDIN: @divyebajaj

PROJECT :

ANALYSING USER LOGIN BEHAVIOUR FOR AN OTT STREAMING PLATFORM

This project analyzes user behavior on an OTT (Over-the-Top) streaming platform using a dummy dataset modeled on realistic usage patterns. The platform offers subscription-based access to digital content including movies, series, and live programming across mobile, desktop, and smart TV devices.
The dataset covers viewer activity from multiple countries — including India, the US, Canada, and parts of Europe — and includes login logs, session performance, and promotional campaign records.
Since the dataset is historical, a fixed reference date (March 31, 2024) was used for churn and retention analysis instead of the current system date.

Timeframe

January 2023 – March 2024 (15 months)

Tools Used

- SQL (Google BigQuery)
- Google Looker Studio
- Google Sheets / Excel
- MS Word / PowerPoint

Skills Applied

- Root Cause Analysis (RCA)
- Window Functions & CTEs
- Cohort & Segmentation Analysis
- DAU/MAU, Churn, and Retention Metrics
- Dashboard Development & Storytelling

OTTs in Market

NETFLIX

ALTT

SONY liv

prime

ZEE5

This project is made on dummy data and only made for educational purpose.

Following are the tables made using generative AI for sole purpose of practicing SQL queries, business skills and data analytics:

Viewers

- viewer_id **PK**: Unique identifier for each user
- signup_date: Date the viewer joined
- country: Country/region
- age_group: Viewer age segment (e.g. 18–25)

Logins

- viewer_id **FK**: Foreign key from Viewers
- login_date: Date the user accessed the OTT platform
- device_type: Smart TV, Mobile, Web, etc.
- session_score: Engagement score from viewing activity
- login_channel: Google, Apple, Email, Mobile OTP

Dataset CSVs Links:

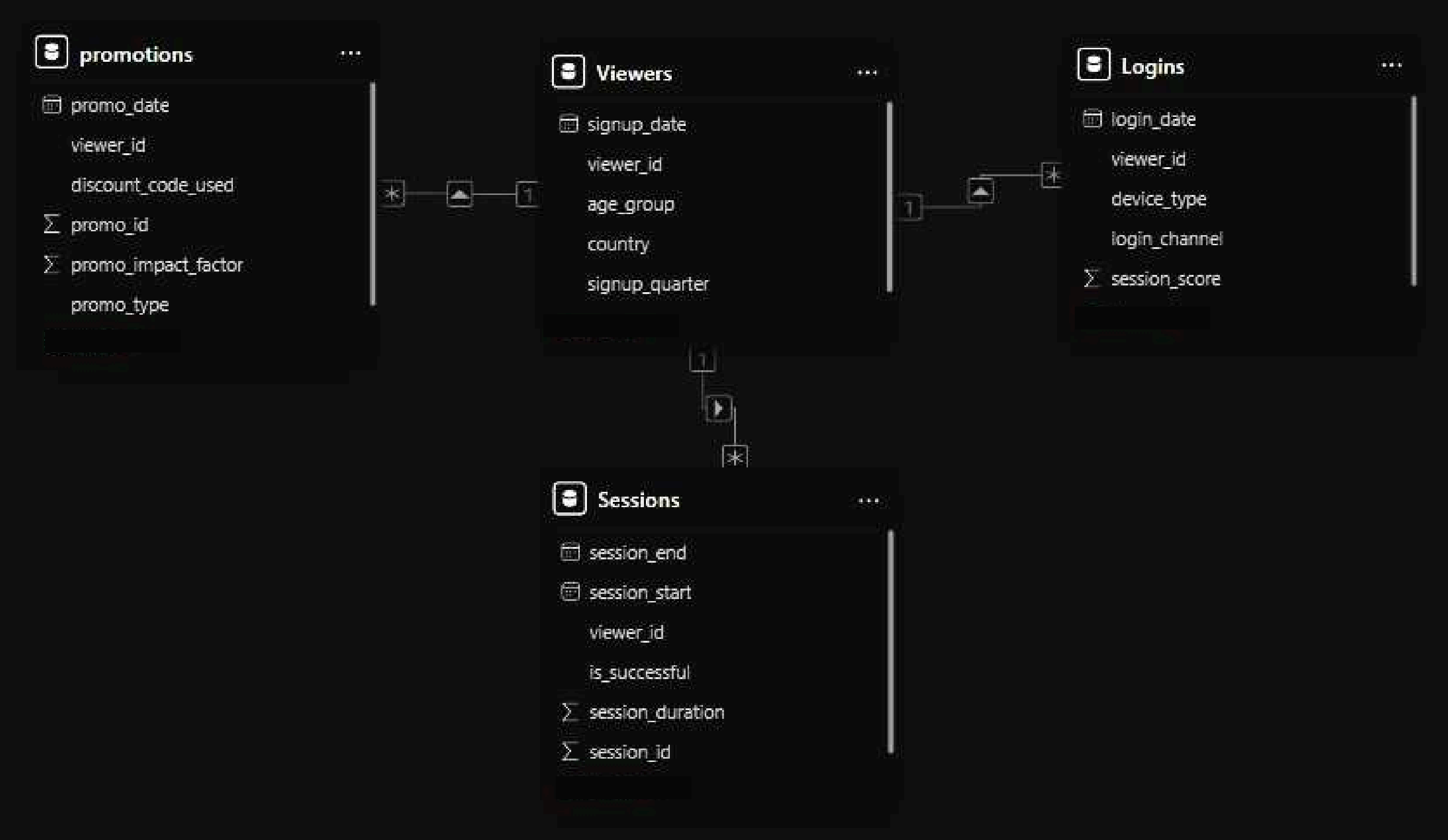
- **Viewers:** *Link*
- **Logins:** *Link*
- **Sessions:** *Link*
- **Promotions:** *Link*

Sessions

- session_id **PK**: Unique session ID
- viewer_id **FK**: Foreign key from Viewers
- session_start: Timestamp of session start
- session_end: Timestamp of session end
- session_duration: Duration of session in minutes
- is_successful: Whether the stream worked without buffering

Promotions

- promo_id **PK**: Unique promotion ID
- viewer_id **FK**: Viewer who received the promo
- promo_type: Free Trial, Referral, Cashback, etc.
- promo_date: When the promotion was triggered
- discount_code_used: Whether any discount/coupon was applied



TABLES

3

1. Identify Viewers Who Have Not Logged In Within the Last 3 Months

Link

What It's Asking:

- Return the list of viewer IDs who haven't logged into the OTT platform since January 1, 2024 (3 Months before last date of dataset).

Objective:

- Find inactive users based on their last login timestamp.

Approach:

- Use MAX(login_date) to determine each user's most recent login.
- Filter for users whose last_login is earlier than 3 months before the reference date '2024-03-31'.

Key SQL Concepts:

- MAX(), GROUP BY, INTERVAL, Subquery, Date Filtering

2. Daily Active Users (DAU) Trend

Link

What It's Asking:

- Track how many unique viewers log in each day to understand usage fluctuations over time.

Objective:

- Measure daily platform engagement by counting distinct users per day.

Approach:

- Group login records by login_date.
- Count the number of unique viewer_ids for each date.

Key SQL Concepts:

- COUNT(DISTINCT), GROUP BY, ORDER BY, Time-Series Aggregation

3. Monthly Active Users (MAU) Trend
Link

What It's Asking:

- Show how many unique users were active each month to identify broader engagement trends.

Objective:

- Understand user activity on a monthly scale and spot long-term shifts in platform usage.

Approach:

- Extract the year and month from login_date using FORMAT_DATE('%Y-%m').
- Count distinct viewer_ids per month to calculate MAU.

Key SQL Concepts:

- FORMAT_DATE(), COUNT(DISTINCT), GROUP BY, Monthly Aggregation

4. Who Watches Content at Least 3 Times a Week Link

What It's Asking:

- Identify high-engagement viewers who log in on 3 or more distinct days within any week.

Objective:

- Segment consistently active users for potential targeting or reward-based campaigns.

Approach:

- Extract week and year from login_date and group logins accordingly.
- Count unique login days and filter for users with 3 or more per week.

Key SQL Concepts:

- EXTRACT(), COUNT(DISTINCT), GROUP BY, HAVING

5. How Does Average Session Duration Vary by Age Group? *Link*

What It's Asking:

- Analyze if viewers from different age groups spend more or less time per session.

Objective:

- Understand content engagement patterns across age demographics.

Approach:

- Join session data with viewer demographics using viewer_id.
- Group by age_group and calculate average session_duration.

Key SQL Concepts:

- JOIN, GROUP BY, AVG(), ORDER BY

6. How Does Average Session Duration Vary by Region? *Link*

What It's Asking:

- Compare how long users from different countries engage with content per session.

Objective:

- Identify regional differences in viewing behavior to guide localization or product performance tuning.

Approach:

- Join Sessions with Viewers to associate users with their countries.
- Group by country and calculate the average session duration.

Key SQL Concepts:

- JOIN, GROUP BY, AVG(), ORDER BY

SQL QUERIES

7. How Many Users Dropped Off After Their First Login? Link

What It's Asking:

- Find the number of users who logged in only once and never returned.

Objective:

- Identify immediate churn to evaluate onboarding effectiveness.

Approach:

- Group logins by viewer_id and count total logins per user.
- Filter users with only 1 login using HAVING COUNT = 1.

Key SQL Concepts:

- GROUP BY, HAVING, COUNT(), Subquery

8. What's the Average Time Between a User's Signup and Last Login (by Segment)? Link

What It's Asking:

- Measure how long users from different age groups stay active on the platform.

Objective:

- Evaluate user lifetime across age-based segments.

Approach:

- Calculate days between signup_date and latest login_date for each viewer.
- Group results by age_group and average the time difference.

Key SQL Concepts:

- DATE_DIFF, JOIN, AVG(), GROUP BY

9. Are Session Errors Linked to Churn? Link

What It's Asking:
- Determine if users who experienced failed sessions are more likely to stop using the platform.

Objective:
- Assess the impact of technical failures on user retention.

Approach:
- Split users into churned and active based on last login date.
- Compare session failure rates (is_successful = FALSE) between both groups.

Key SQL Concepts:
- CASE WHEN, JOIN, CTEs, DATE FILTERING

10. Does Login Channel Impact Long-Term Usage? Link

What It's Asking:
- Examine if users signing in via Google, Email, Apple, etc., show different usage or churn patterns.

Objective:
- Understand which login methods are linked to stronger user retention.

Approach:
- Calculate average login frequency per login_channel.
- Compare churn rates (inactivity > 3 months) across channels.

Key SQL Concepts:
- AVG(), MAX(), CASE WHEN, CTEs, JOIN

11. Which Signup Cohorts (e.g., Jan 2023) Have the Worst Retention? Link

What It's Asking:
- Identify which monthly signup groups had the lowest percentage of users still active after 30 days.

Objective:
- Spot weak-performing cohorts to evaluate onboarding or promo effectiveness.

Approach:
- Flag users who returned 30+ days after signup, grouped by signup month.
- Calculate retention % per cohort_month and rank lowest to highest.

Key SQL Concepts:
- DATE_DIFF, DATE_TRUNC, CASE WHEN, WINDOW FUNCTIONS, GROUP BY

12. Which Promotion Types Created Long-Term OTT Viewers? Link

What It's Asking:
- Determine which types of promotional campaigns led to users staying active for 90+ days.

Objective:
- Evaluate the effectiveness of different promo types in driving sustained platform engagement.

Approach:
- Measure active duration between first and last login per viewer.
- Count users active ≥ 90 days and calculate percentage by promo_type.

Key SQL Concepts:
- DATE_DIFF, CASE WHEN, JOIN, GROUP BY, ROUND()

SQL QUERIES

13. Do Promotion-Acquired Viewers Watch More Than Organic Ones? Link

What It's Asking:

- Compare how many sessions users from promotions complete vs those who joined organically.

Objective:

- Measure the engagement quality of promo campaigns compared to natural user acquisition.

Approach:

- Label users as either Promotion or Organic based on presence in promotions.
- Calculate average session count per acquisition type.

Key SQL Concepts:

- UNION, JOIN, COUNT(), AVG(), ROUND(), GROUP BY

14. How Long Do Viewers Stay Active After a Trial Offer? Link

What It's Asking:

- Calculate how many days users continue engaging with the platform after receiving a "Free Trial."

Objective:

- Measure trial campaign effectiveness by tracking post-trial activity duration.

Approach:

- Identify users who activated a Free Trial and their latest login.
- Compute number of days between trial start and last login.

Key SQL Concepts:

- DATE_DIFF, JOIN, WHERE, AVG(), MIN(), MAX()

15. Which Promotions Correlate with Smooth Technical Experiences?

Link

What It's Asking:

- Identify which promotional campaigns are associated with the highest percentage of successful (non-buffering) sessions.

Objective:

- Assess the technical quality of viewer experience linked to each promotion type.

Approach:

- Join promotions with session logs to match promo users with session outcomes.
- Calculate success rate per promo_type using is_successful.

Key SQL Concepts:

- JOIN, CASE WHEN, COUNT(), SUM(), ROUND(), GROUP BY

16. Peak Login Days Across Regions

Link

What It's Asking:

- Find out which day of the week users in each country are most active on the platform.

Objective:

- Help the business optimize release schedules, marketing pushes, or support coverage based on region-specific activity patterns.

Approach:

- Join login data with viewer demographics to get region info.
- Count logins per weekday and country, then rank activity levels.

Key SQL Concepts:

- JOIN, FORMAT_DATE('%A'), GROUP BY, COUNT(), ORDER BY

17. Are Some Regions Showing a Steep Engagement Decline? Link

What It's Asking:

- Identify countries where the number of active users is consistently dropping month over month.

Objective:

- Detect high-risk markets that may need product or marketing intervention.

Approach:

- Track monthly active users per country and compare with previous months.
- Count how often activity declined to measure trend severity.

Key SQL Concepts:

- LAG(), COUNTIF(), DATE FORMATTING, PARTITION BY, CASE WHEN

18. Do Returning Viewers Spend More Time Than New Viewers? Link

What It's Asking:

- Compare the average session duration between one-time users and those who come back.

Objective:

- Evaluate the value of retained users by measuring their session engagement.

Approach:

- Classify users based on whether they had only one session or multiple.
- Compare average session time for each group.

Key SQL Concepts:

- CASE WHEN, GROUP BY, AVG(), ROUND(), CTEs

VISUALIZATION

Platform:
Looker Studio (Google Data Studio)

What's Covered:
- Daily & Monthly Active Users (DAU/MAU trends)
- Viewer distribution by age, region, and acquisition type
- Peak login hours and days
- Retention and session quality insights
- Promotion type vs. churn or session success
- Country-wise engagement trend (map view)

Tools:
- SQL (BigQuery as backend)
- Looker Studio (for data storytelling)
- Custom filters for age group, region, and promo type

Live Dashboard Link:
- *View Dashboard in Looker Studio*

Visual PDF Report
- *Visualization PDF*

Business Context

The OTT platform observed fluctuations in user activity and engagement post-trial periods, and inconsistent session outcomes across channels and geographies. This case study explores root causes behind user drop-off and highlights segments with sustained value.

Real Data Highlights

- **Viewers:** 1,000
- **Total Sessions:** 13,971
- **Total Logins:** 15,964
- **Countries:** 10
- **DAU Trend:** Fluctuating below 40 users/day
- **MAU Trend:** Peaked ~800

Key Findings

Churn & Drop-off

- Significant churn: DAU never stabilizes beyond 40 daily users
- Most viewers log in only a few times over many months
- MAU peaked around mid-2024, then declined into 2025

Session Behavior

- Avg. session duration highest in 18–25 and 26–35
- Older age groups (56+) drop to ~15–30 mins

Regional Patterns

- Top regions: India, France, UK
- Lowest: Germany, Canada
- Session duration low in Germany & France

Technical Experience

- Mobile OTP = highest playback success
- Facebook/Email = lowest session success
- Playback issues → higher churn

Promotional Strategy

- Free Trials = high sessions, low retention
- Loyalty/Referral = better sustained engagement

CASE STUDY

Root Causes Identified

High Churn

Most users don't return after trial; no ongoing engagement strategy

Short Sessions

Older users and certain countries (Germany, France) disengaged

Technical Failures

Facebook/Email login correlated with lower session success

Unbalanced Campaigns

Over-reliance on Free Trials with little retention payof

Recommendations

1. Launch re-engagement campaigns targeting trial drop-offs within 14 days
2. Optimize login channels – fix technical issues with Facebook/Email
3. Segment retention messaging by age and country (e.g., 45+ viewers)
4. Shift marketing mix toward referral/ loyalty offers over Free Trials

THE END!

Made by:

Divye Bajaj

LinkedIn