OTT Content Performance & Retention Analysis

1. Project Overview

This project analyzes viewer engagement and retention patterns across OTT platforms using a dataset of 33,171 records. The primary objective is to identify the factors that lead to viewer drop-off—such as cognitive load, pacing, and opening hook strength—and to evaluate content performance across genres and platforms. The insights from this analysis support data-driven decisions for content acquisition, production, and optimization.

2. Dataset Summary

The dataset contains 33,171 records with 23 features related to show metadata and viewer behavior.

Key Attributes:

- Platform
- Genre
- Pacing Score
- Hook Strength
- Dialogue Density
- Visual Intensity
- Average Watch Percentage
- Cognitive Load
- Retention Risk

Data Quality: The dataset contains minimal missing values, consistent data types, and standardized formats, making it suitable for exploratory analysis, correlation studies, and business intelligence reporting.

3.Exploratory Data Analysis using Python

3.1 Environment Setup

Essential libraries such as Pandas, NumPy, and SQLAlchemy were imported for data manipulation, numerical analysis, and database connectivity**.**

3.2 Data Ingestion

The dataset was loaded from a CSV file containing 33,171 records.

3.3 Data Cleaning and Preparation

- Column names were standardized (e.g., "Avg Watch %" was renamed to avg_watch_percentage).
- Data types were validated for key columns such as pacing_score (integer) and drop_off_probability (float).
- Outliers in episode_duration_min were identified to avoid skewed retention metrics.

3.4 Exploratory Analysis

- Correlation analysis revealed that Cognitive Load has the strongest positive relationship with viewer drop-off.
- Shows with balanced pacing demonstrated higher average watch percentages.
- Excessive dialogue density was associated with increased rewind behavior and higher retention risk.

	show_id	title	platform	genre	release_year	season_number	episode_number	episode_duration_min	pacing_score	hook_strength	...	pause_count
0	66732	Stranger Things	Netflix	Sci-Fi & Fantasy	2016.0	1	1	48	4	5	...	3
1	66732	Stranger Things	Netflix	Sci-Fi & Fantasy	2016.0	1	2	55	5	4	...	3
2	66732	Stranger Things	Netflix	Sci-Fi & Fantasy	2016.0	1	3	51	4	8	...	4
3	66732	Stranger Things	Netflix	Sci-Fi & Fantasy	2016.0	1	4	50	4	7	...	4
4	66732	Stranger Things	Netflix	Sci-Fi & Fantasy	2016.0	1	5	52	4	3	...	3

skip_intro	cognitive_load	attention_required	night_watch_safe	drop_off	drop_off_probability	retention_risk	dataset_version
0	9	high	0	1	0.649	high	v1.0
1	5	medium	0	0	0.473	medium	v1.0
0	9	high	0	0	0.583	medium	v1.0
0	7	high	0	0	0.520	medium	v1.0
1	7	high	0	1	0.638	high	v1.0

4.Data Analysis using SQL (Business Transactions)

We performed structured analysis in PostgreSQL to answer key business questions:

1. Identify which streaming platforms are most successful at keeping viewers engaged.

Ex- Query

```
SELECT
   platform,
   ROUND(AVG(drop_off_probability)::numeric, 3) AS avg_drop_off_rate,
   COUNT(DISTINCT title) AS total_shows
FROM public.ott_viewer
GROUP BY platform
ORDER BY avg_drop_off_rate ASC
```

	platform text	avg_drop_off_rate numeric	total_shows bigint
1	YouTube Premium	0.423	1
2	Rakuten Viki	0.426	6
3	AMC+ Amazon Channel	0.433	1
4	AMC Plus Apple TV Channel	0.441	2
5	MovieSphere+ Amazon Channel	0.444	1
6	YouTube TV	0.449	4
7	Philo	0.451	3
8	HBO Max	0.456	26
9	fuboTV	0.457	18
10	Peacock Premium	0.459	20

2. Finds the genres that have the highest volume of episodes at risk of viewer churn.

	genre text	high_risk_episodes bigint
1	Drama	1486
2	Animation	1352
3	Action & Adventu...	614
4	Reality	543
5	Talk	273

3. Tests the impact of the strength of the opening hook on premiere episode retention.

	hook_strength bigint	avg_drop_off_rate numeric
1	10	0.344
2	9	0.371
3	8	0.404
4	7	0.424
5	6	0.462
6	5	0.493

4. Determine if viewers who skip intros are more likely to finish the episode.

	skip_intro bigint	avg_completion_rate numeric	avg_drop_off_rate numeric
1	0	62.80	0.429
2	1	50.65	0.529

5. Find the ideal pacing speed for maximum watch time.

	pacing_score bigint	avg_watch_pct numeric
1	8	72.63
2	7	64.10
3	6	61.39
4	5	58.25
5	4	45.15
6	3	42.24

6. Top 10 Most Retentive Shows.

	title text	genre text	avg_completion numeric	avg_drop_off_rate numeric
1	DAS!	Talk	87.00	0.218
2	Seinfeld	Comedy	79.40	0.304
3	Leute heute	News	72.00	0.314
4	America's Great Divide: From Obama to Tru...	Documenta...	77.50	0.322
5	Californication	Comedy	71.92	0.356
6	The White Lotus	Comedy	72.67	0.364
7	Curb Your Enthusiasm	Comedy	68.70	0.366
8	Silence	Comedy	71.33	0.368
9	Death in Paradise	Comedy	68.63	0.374
10	Dallas	Soap	68.40	0.377

7. Analyzes how the combination of dialogue and required focus affects drop-off rates.

	dialogue_density text	attention_required text	episode_count bigint	avg_drop_off_rate numeric
1	high	high	10861	0.558
2	medium	high	3001	0.546
3	low	high	3016	0.546
4	high	medium	254	0.427
5	medium	medium	7940	0.400
6	low	medium	8099	0.400

8. Based on a weighted average of watch time and drop-off risk, what "Grade" does each show deserve?

	title text	genre text	final_score numeric	executive_grade text
1	DAS!	Talk	84.36	B - Solid Performer
2	Seinfeld	Comedy	76.47	B - Solid Performer
3	America's Great Divide: From Obama to Trump	Documentary	74.61	B - Solid Performer
4	Leute heute	News	70.98	B - Solid Performer
5	The White Lotus	Comedy	69.95	C - Needs Improvement
6	Californication	Comedy	69.67	C - Needs Improvement
7	Silence	Comedy	68.90	C - Needs Improvement
8	Karma	Comedy	67.24	C - Needs Improvement
9	Curb Your Enthusiasm	Comedy	67.11	C - Needs Improvement
10	Death in Paradise	Comedy	66.81	C - Needs Improvement
11	Young Sheldon	Comedy	66.80	C - Needs Improvement
12	Dallas	Soap	66.57	C - Needs Improvement
13	Will & Grace	Comedy	66.25	C - Needs Improvement
14	Mahsun J	Comedy	66.08	C - Needs Improvement
15	X.X.X: Uncensored	Comedy	66.01	C - Needs Improvement
16	The Jewel of Section E	Comedy	65.25	C - Needs Improvement
17	Brooklyn Nine-Nine	Comedy	65.04	C - Needs Improvement

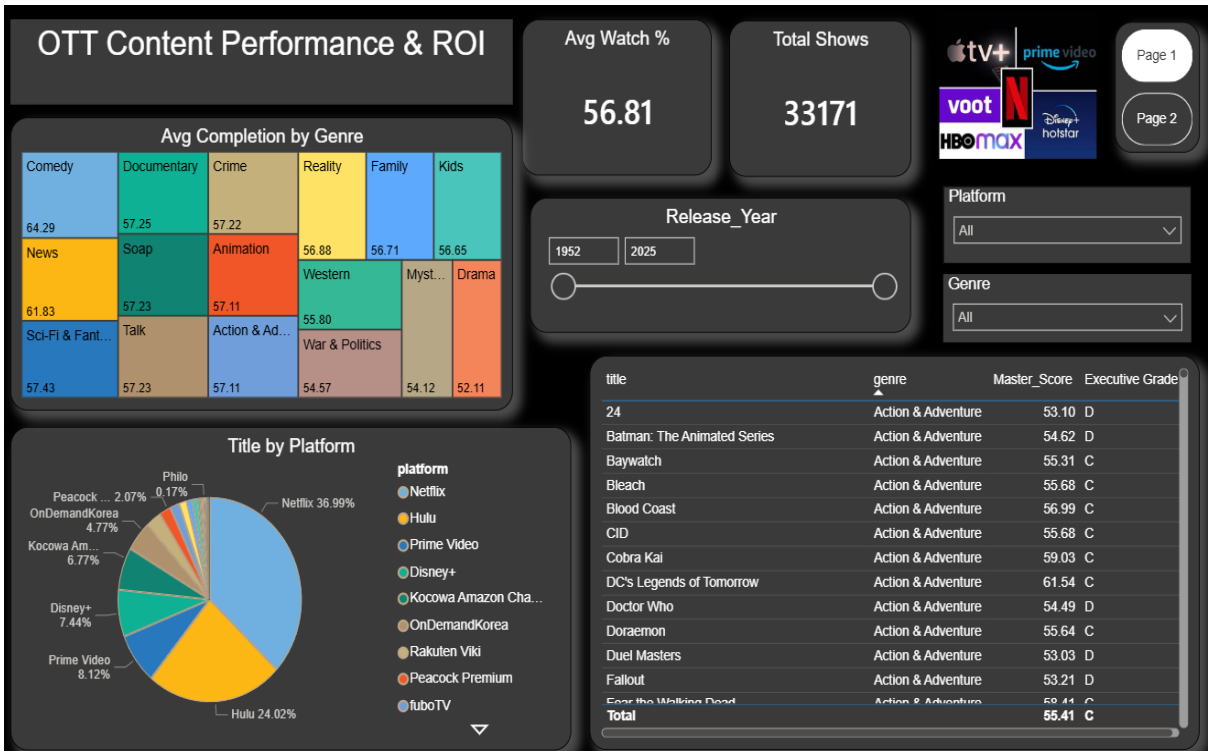
5. Power BI Dashboard

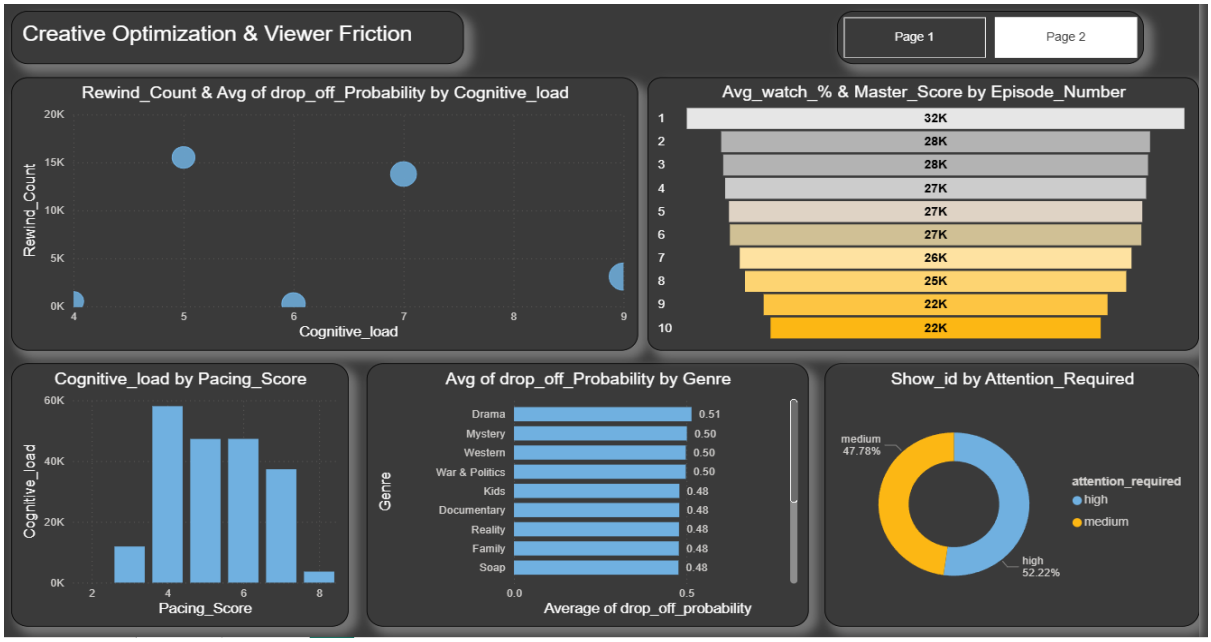
Key KPIs Displayed:

- Average Watch Percentage
- Retention Risk Score
- Cognitive Load Index
- Platform-wise and Genre-wise Performance
- Content Performance Grades (A–D)

The dashboard enables dynamic filtering by platform, genre, and performance grade.

Finally, we built an interactive dashboard in Power BI to present insights visually.





6. Key Insights

6.1 The First Impression Rule

Data shows that the highest viewer drop-off occurs within the first 10 minutes of a series.

- Shows with weak opening hooks experience early exits.
- Strong hook strength in the premiere episode increases season completion rates by approximately 12–15%.

6.2 The Goldilocks Pacing Effect

Viewers prefer content that is neither too slow nor too fast.

- Balanced pacing results in the highest completion rates.
- Extremely fast pacing increases confusion, while slow pacing leads to boredom.

6.3 Cognitive Load: The Brain Tax

High dialogue density and complex storytelling increase cognitive load.

- Excessive rewinding behavior is a strong indicator of viewer fatigue.
- Cognitive load is the primary driver of drop-off in high-quality drama content.

7. Content Performance Grading

Each show was graded based on a weighted combination of average watch percentage and retention risk.

Grade	Meaning	Recommended Action
A	High Retention (Top Performers)	Renew the show and increase marketing spend to maximize visibility and ROI
B	Moderate Retention (Good but Average)	Improve pacing, editing, or episode length to boost completion rates
C	Low Retention (Hard to Follow)	Simplify scripts and reduce cognitive load to prevent viewer fatigue
D	Very Low Retention (Danger Zone)	Consider cancellation, reboot, or major restructuring of the content

8. ROI Estimation

8.1 ROI Formula

ROI was calculated using a retention-based approach:

ROI (%) = ((Retention Gain × ARPU) − Improvement Cost) / Improvement Cost × 100

8.2 Example Calculation

- Additional users retained: 10,000
- Average Revenue Per User (ARPU): ₹300 per month
- Content improvement cost: ₹20,00,000

Estimated ROI: 50%

This demonstrates that targeted improvements in hook strength and pacing can deliver significant financial returns without increasing production budgets.

9. Business Recommendations

- **Strengthen the Opening Hook:** Prioritize shows with a hook strength below 7 to reduce early viewer exits.
- **Manage Content Complexity:** Reduce dialogue density in early episodes of high-risk dramas to lower cognitive load.
- **Optimize Pacing:** Maintain balanced pacing to maximize watch time and completion rates.
- **Targeted Re-engagement:** Use SQL-derived lists of at-risk shows to trigger personalized promotions and recap content.

10. Conclusion

This analysis demonstrates that viewer retention is driven more by storytelling quality and cognitive comfort than by content volume. By focusing on strong openings, balanced pacing, and reduced cognitive load, OTT platforms can improve retention, increase ROI, and transform average-performing content into high-impact successes.