



Gaming Engagement Analysis

Press Start

Authors: Cayke Felipe dos Anjos and James Warsing

Games inspire creative thinking to fuel adventures

The gaming industry is a multi billion dollar market that allies innovative writing and storytelling with technological advancement in order to create a product that can entertain customers for tens to hundreds of hours. Because of the constant increase in complexity, releasing a competitive game nowadays can cost hundreds of millions of dollars.

Recent successes from this industry are:

- **GTA V** - with a **budget of \$265 million**, GTA V broke-even in just a few hours after release and has now sold approximately **\$8 billion**
- **Elden Ring** - the successor of Dark Souls, known for being extremely hard, costed **\$200 million** but has sold over **25 million** copies so far.

But also some huge failures:

- **Immortals of Aveum** - this first person shooting game **costed** around **\$125 million** and only sold **\$2 million**. The studio workers ended up being laid off or furloughed.



Business Problem

The company is expanding its portfolio by investing in a new game studio. Launching a new game in today's competitive entertainment industry requires a solid understanding of what drives game success and attracts audiences.

Our project aims to analyze a gamer engagement dataset. By using data analysis techniques and statistical modelling, we seek to predict the best features that correlate with high player engagement.

We are looking to answer the following questions:

What are the top features that correlate with gamer engagement?

How different are the audiences and their engagement choices?

How can we create conditions for higher engagement?



The Engagement Problem

Why engagement?

Engagement is a complicated measurement that estimates how involved in the game and with its community the player is. Along with playing the game for long hours, the player engages with the game when they are part of a community in the game (clan, guild or party) or outside of the game (Reddit, forums or meetups). Having a live community is important because they bring more gamers to the community.

Suicide Squad: Kill the Justice League was a game that had a \$200 millions loss.

- **Player counts** are **very low**, peaking at **14k** but now going around approximately **100~200 players**;
- **Not releasing exciting extra content** to keep players coming back;

If there is no good content to engage, then there is no incentive to keep playing.



Data Understanding

We used an [Online Gaming Behavior Dataset](#) from Kaggle to analyze and make models to help us determine the best ways to increase gamer engagement. This dataset includes 40,000 rows of information such as Age, Gender, Genre, Gaming Duration, and much more.



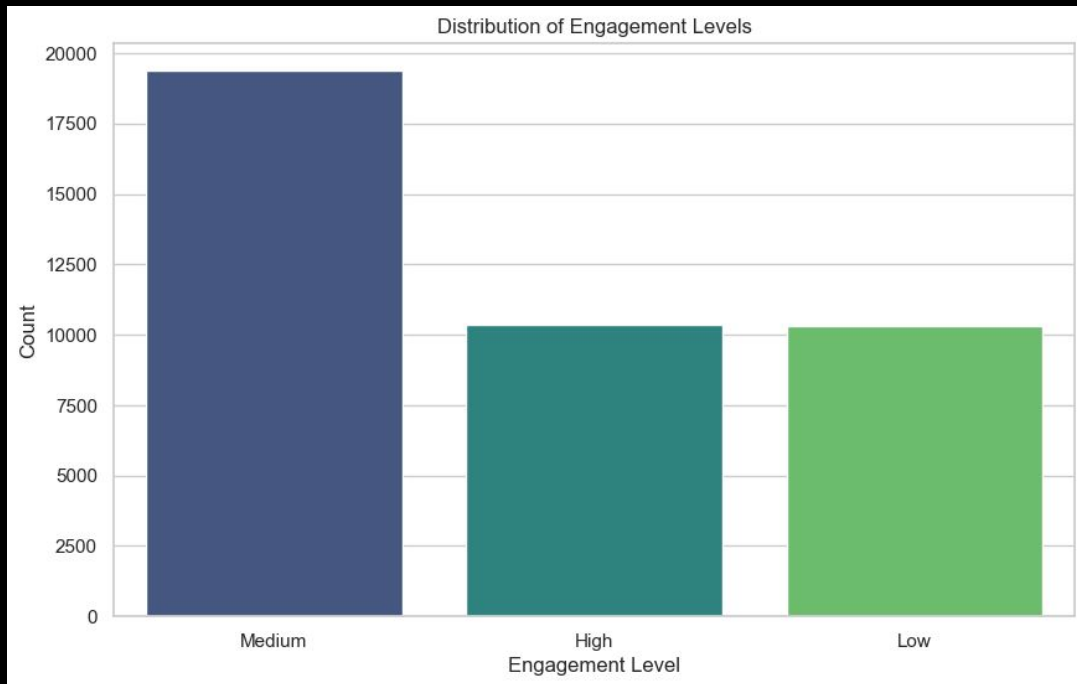
	PlayerID	Age	Gender	Location	Genre	PlayTimeHours	InGamePurchases	Difficulty	SessionsPerWeek	AvgSessionDurationMinutes	Level	AchievementsUnlocked	Engagement
40029	49029	32	Male	USA	Strategy	20.619662	0	Easy	4	75	85	14	Medium
40030	49030	44	Female	Other	Simulation	13.539280	0	Hard	19	114	71	27	High
40031	49031	15	Female	USA	RPG	0.240057	1	Easy	10	176	29	1	High
40032	49032	34	Male	USA	Sports	14.017818	1	Medium	3	128	70	10	Medium
40033	49033	19	Male	USA	Sports	10.083804	0	Easy	13	84	72	39	Medium

Analysis Methods

We used multiple machine learning models such as logistic regression and decision trees to predict what had the largest impact on the player's engagement

Here we look at the distribution of engagement. Most of the data shows that generally players seem to have a medium engagement level when playing games.

We aim to find why players have higher engagement and capitalize on that.



Analysis Methods

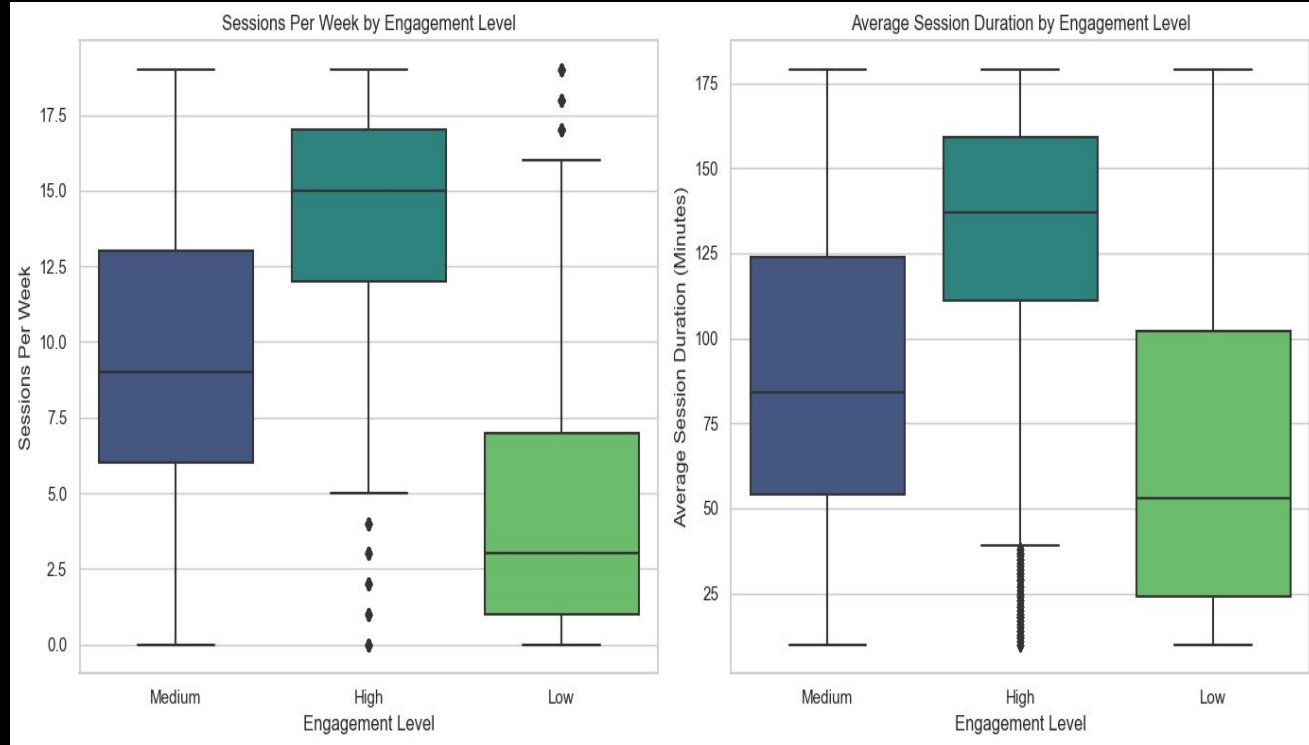
This heatmap shows that playtime has the highest correlation to our target, engagement.

So the longer players are playing the game, the more engaged they are



Analysis Methods

The boxplots show that more sessions played per week coupled with longer session durations result in higher overall engagement levels.



Choosing our Model



Baseline Model - Logistic Regression with All Features

2nd Model - Logistic Regression with Most Correlated Features

3rd Model - Decision Tree Classifier

4th Model - 2 Logistic Regression Models Based on First Split from Decision Tree Classifier (SessionsPerWeek)

5th Model - Decision Tree Classifier with Hyperparameters Tuned

Choosing our Model

Accuracy

Train
78%

Test
78%

Baseline Model - Logistic Regression with All Features

2nd Model - Logistic Regression with Most Correlated Features

77%

77%

3rd Model - Decision Tree Classifier

100%

83%

4th Model - 2 Logistic Regression Models Based on First Split from Decision Tree Classifier (SessionsPerWeek)

88%

86%

5th Model - Decision Tree Classifier with Hyperparameters Tuned

92%

90%

Business Recommendations

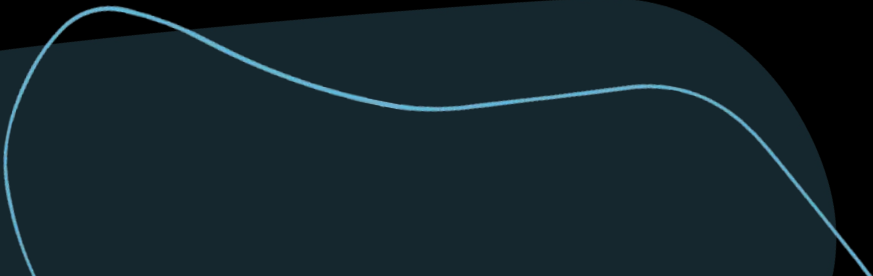
Use best model to estimate weighted average engagement for different groups based on features

Obtain difference between maximum and minimum average engagement to obtain average engagement increase and significance

In order to create an engaging game, we have 3 different recommendations for the **American market** in descending order of importance:

- **More sessions are better:** larger number of sessions per week can increase engagement by 1.26. We recommend creating **special events** during **different days of the week** (dungeons, missions, bosses) with special rewards.
- **Balance is everything:** players with **very large average session duration** engage poorer than gamers with **low duration**. In order to prevent a decline in this behavior we suggest **advising the gamer to take a break to stretch** after a **long time (~2h)** which also **increases the number of sessions**. Playing the game in **moderate amounts** per session can increase engagement by **0.7**.
- **Feeling of achievement and progress is important:** players with **higher levels and more achievements** engage **better**. The game should be able to **allow** the player to **level up faster initially** and **unlock achievements** that are meaningful. This can **increase the average engagement by 0.14** for **higher level players** and **0.1** for **more achievements**.

Next Steps

- Pull more datasets that have information about specific games and their cost so we can determine what games are profitable.
 - Sentiment analysis from social media and gaming forums to see what games are highly played and trending among players.
 - Obtain information about when games release to find the best time to release a game and when not to release a game. (e.g. Christmas/Near other big name releases)
- 
- A decorative blue wavy line that starts on the left side of the slide, curves upwards and then downwards, ending on the right side. It is positioned in the bottom left area of the slide.

Thanks For Playing

Cayke Felipe dos Anjos

✉ caykefelipe01@gmail.com

🐙 [@cayke-fda](#)

 [cayke-fda](#)

James Warsing

✉ warsingjt@gmail.com

🐙 [@james-warsing](#)

 [jameswarsing](#)