INSIGHTS FROM VIDEO GAME PLAYER ENGAGEMENT

Machine Learning I, Winter 2025

Presented by:
Vaishnavi Kokadwar
Ulka Khobragade
Mahender Reddy Pokala
Kunal Mody
Ultra Partihuttakorn

3/9/25

Problem Statement

Background

- Video game industry expected to surpass **$250B** by 2025, with over **3B** players globally
- Revenue and retention are being driven by in-game purchases and personalized recommendations

Target Identification and Model Building

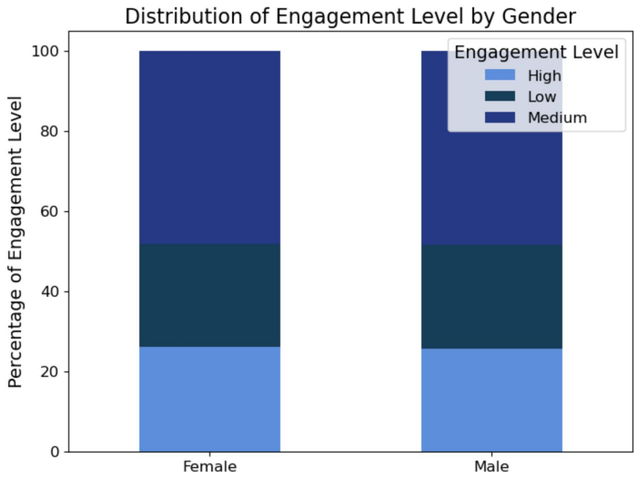
- **2** key business problems
- Develop a model to predict <u>player engagement level</u>, providing actionable insights for game developers.
- Develop a model to predict the combination of <u>in-game purchases</u> and <u>player engagement level</u> to deliver personalized recommendations and maximize revenue.

Model Evaluation and Business Value

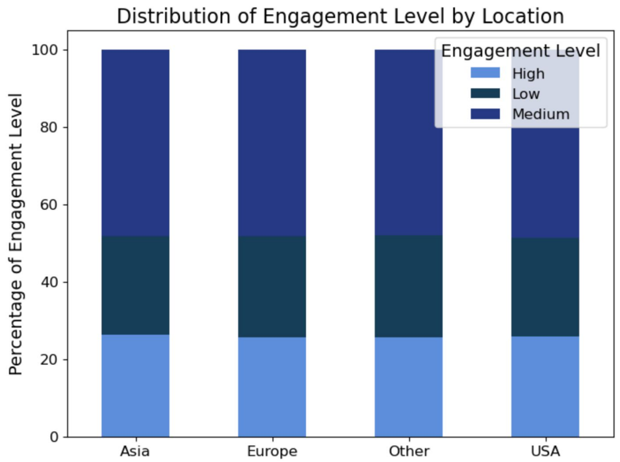
- Goal is to offer developers a comprehensive understanding of consumer preferences
- Tailored game genre and difficulty suggestions are used to increase overall user base.

EDA: Distribution of Engagement Level across categorical variables

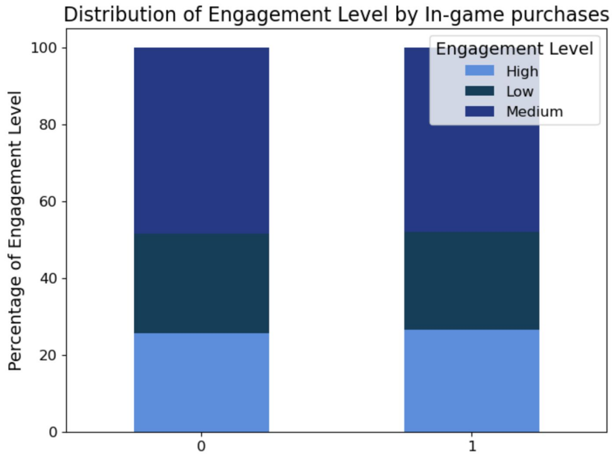
Gender



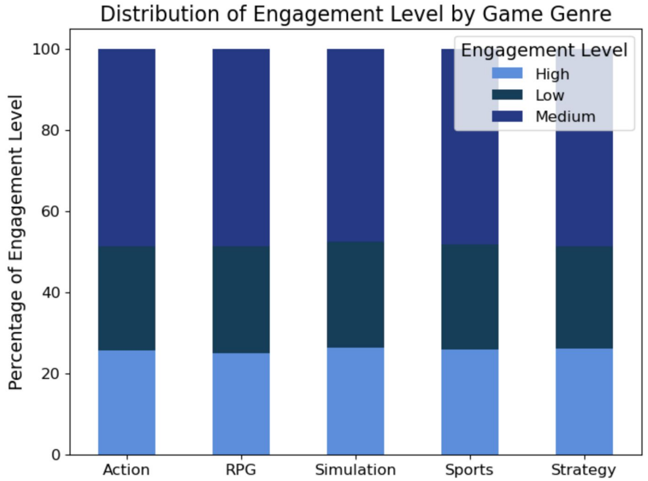
Location



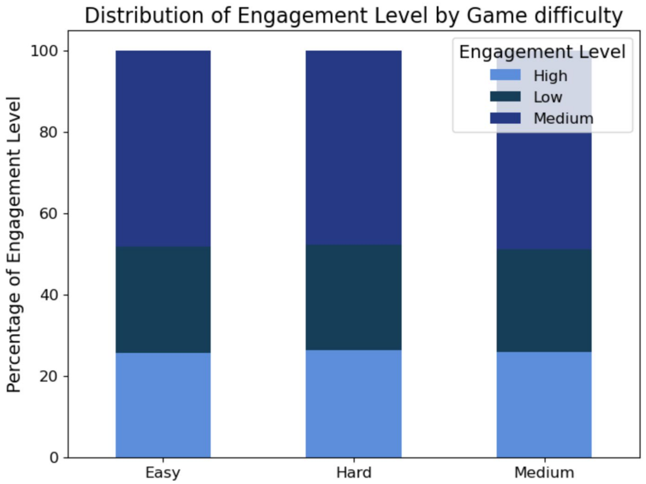
In-Game purchases



Game Genre

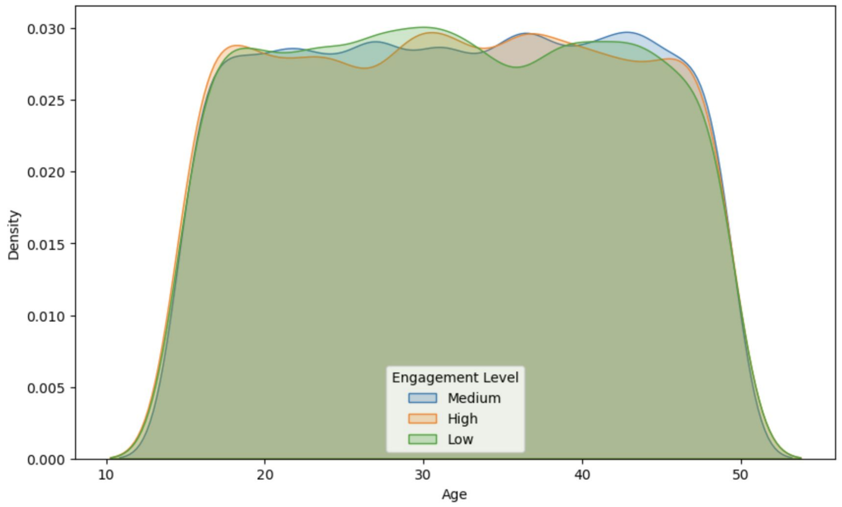


Game Difficulty

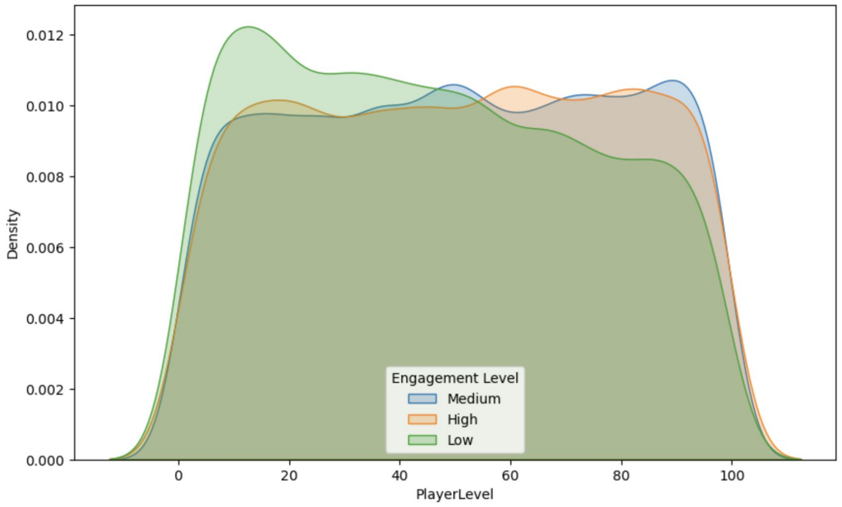


EDA: Density Plots of Numerical Features, faceted by Engagement Level

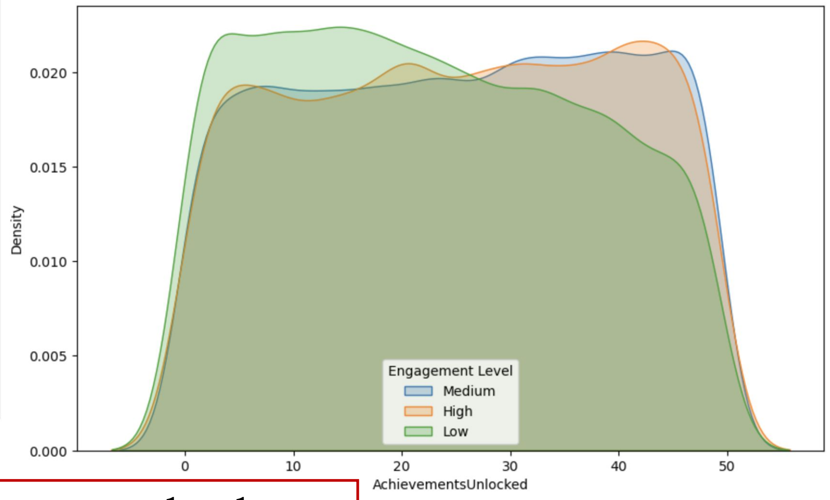
Age distribution by Engagement level



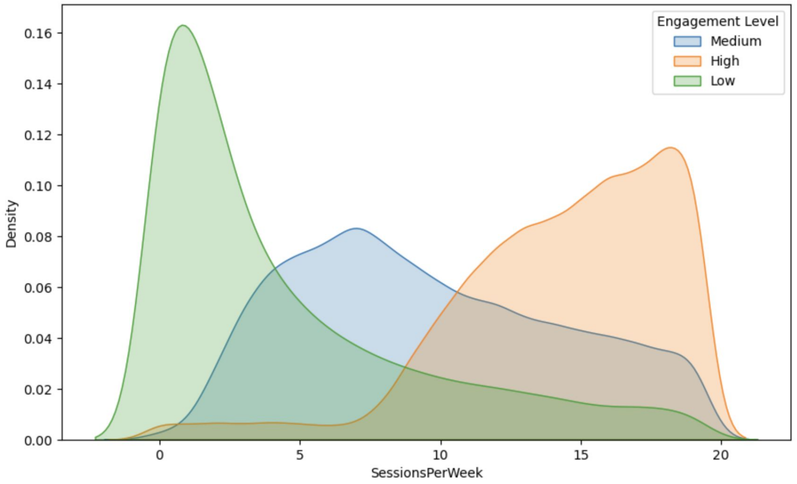
Player level distribution by Engagement level

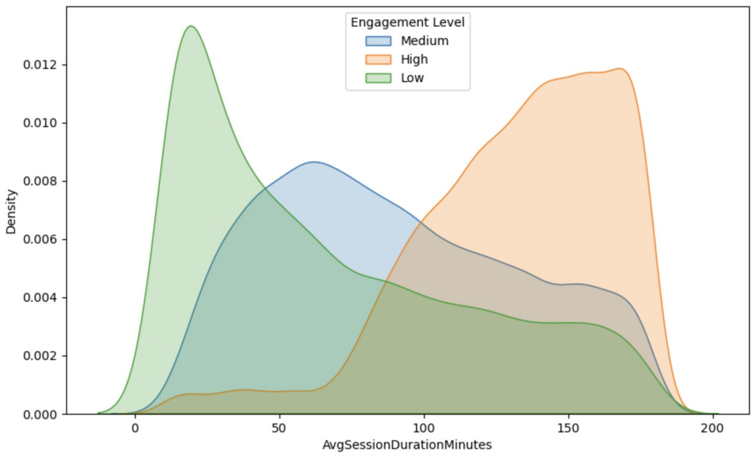


Achievement unlock distribution by Engagement level



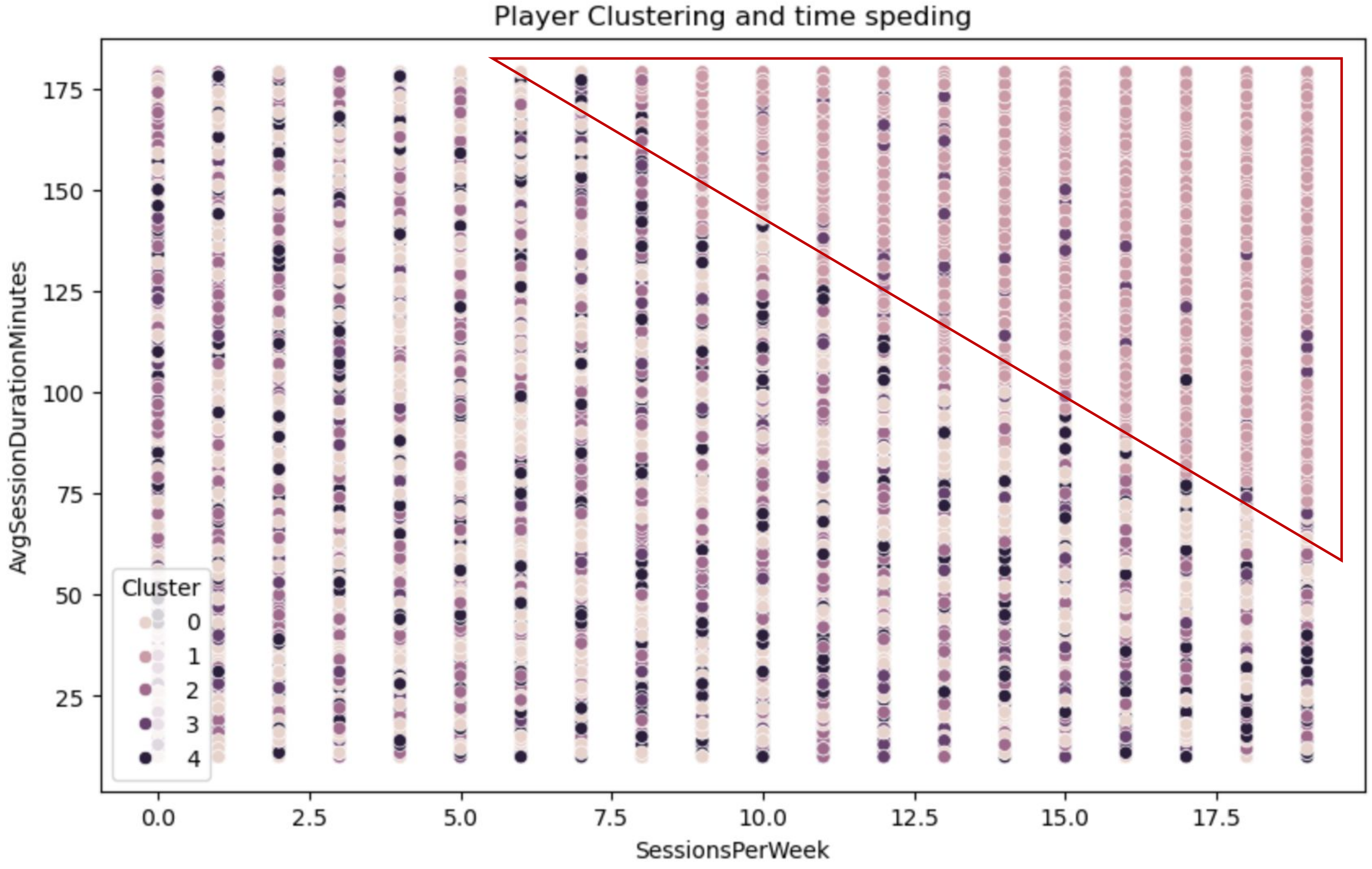
Session per week and Avg Session Duration distribution by Engagement level



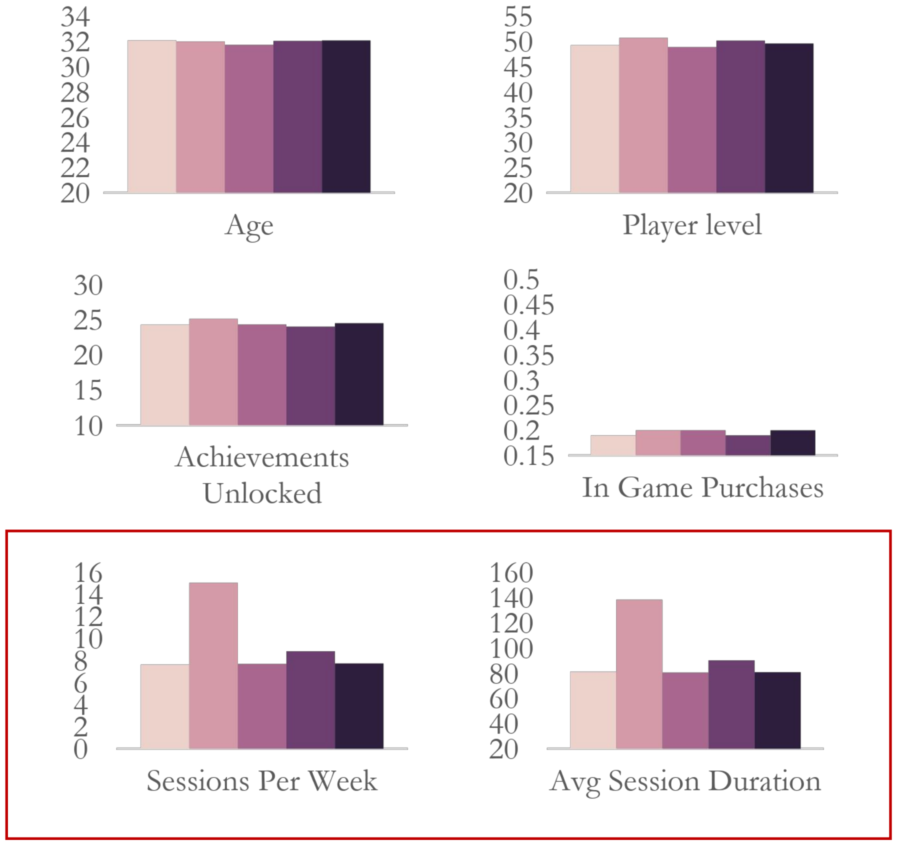


EDA: K-Means Clustering Results (TLDR; ineffective)

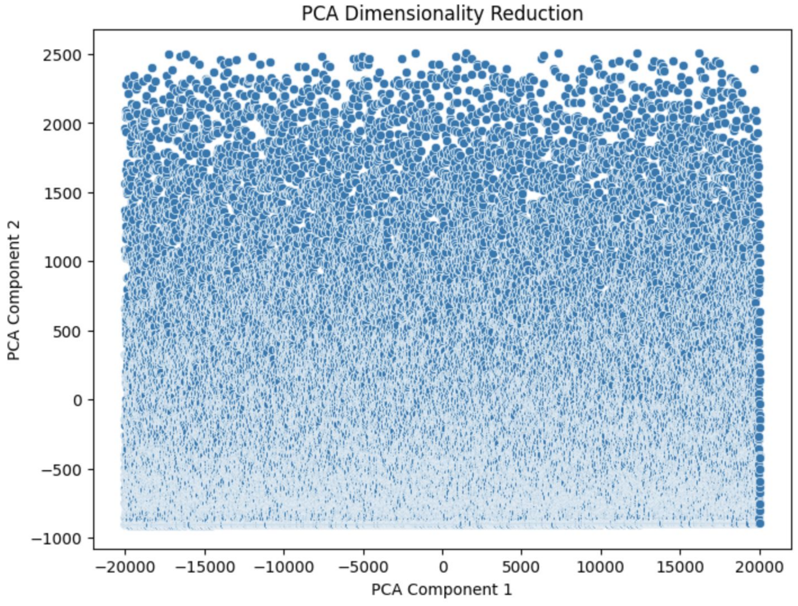
K-mean clustering

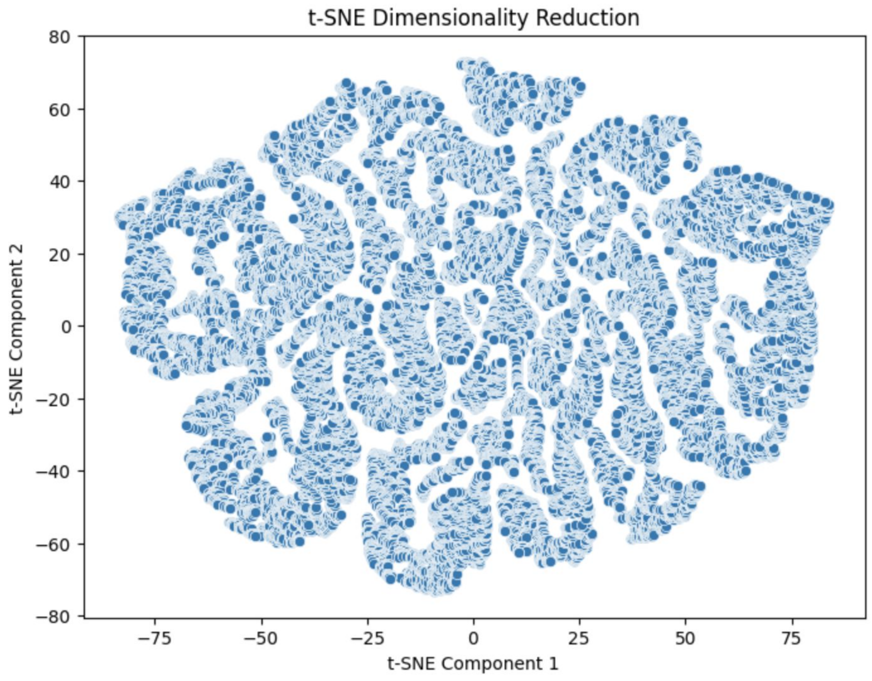


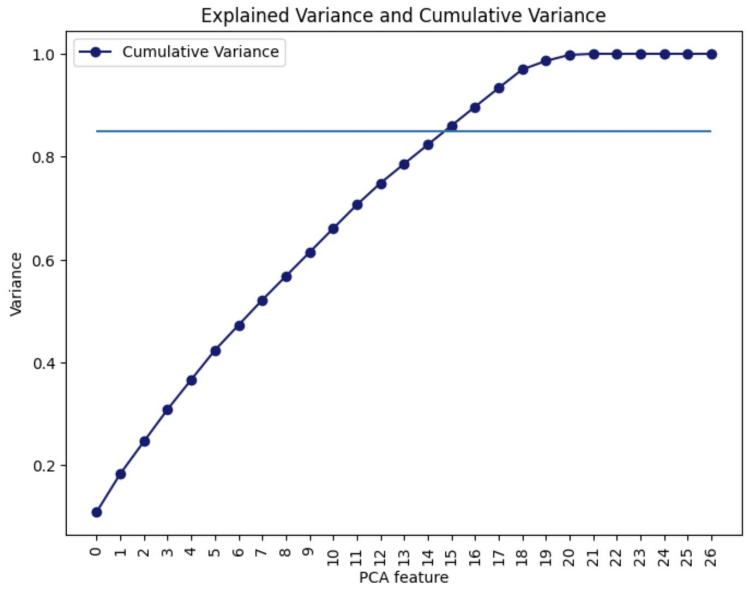
Average numerical variables by clusters



EDA: Dimension Reduction



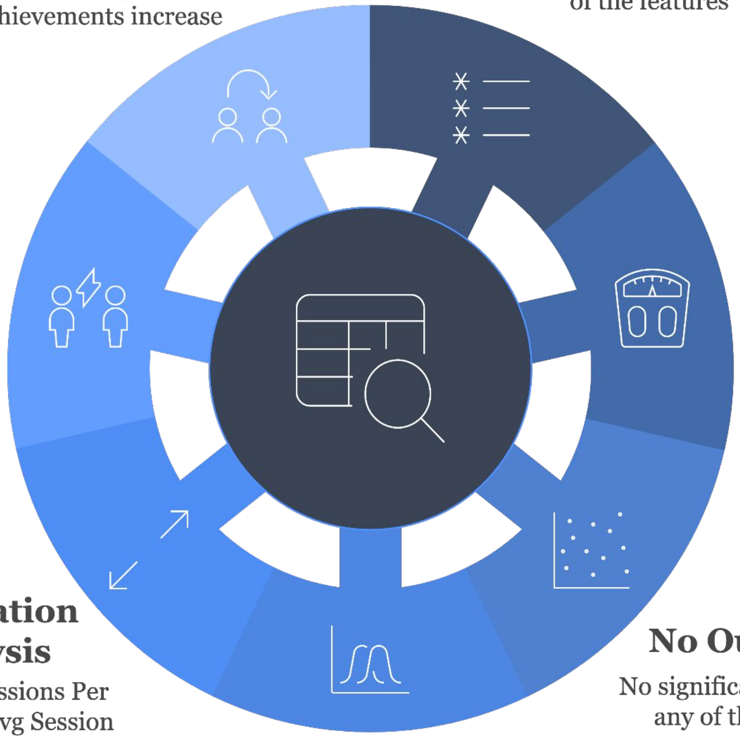




Key Insights

- **15** components make up 85% variance
- No discernable neighborhood features observed

EDA: Key Insights



Category Shift
Movement from Low to Medium categories as Achievements increase

No Missing Values
No missing values in any of the features

Contradictory Relationship
PlayTimeHours and WeeklySessionHours relationship is misleading

Data Imbalance
“InGamePurchase” feature (20%: 1 and 80%:0)

Correlation Analysis
High for Sessions Per Week and Avg Session Duration Minutes

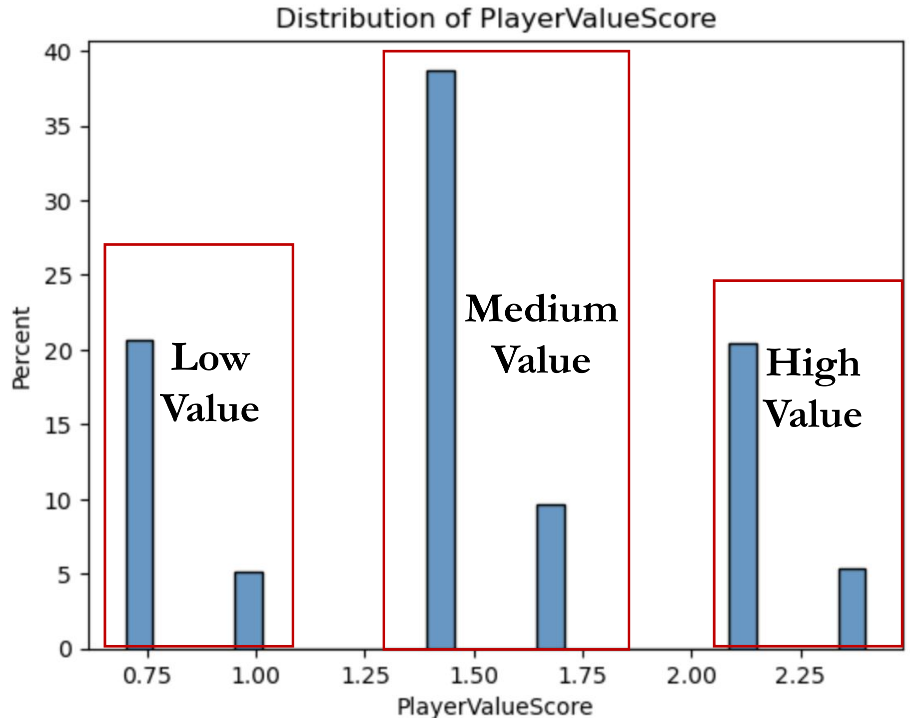
No Outliers
No significant outliers in any of the features

Uniform Distribution
Observed for most of the numeric variables

Feature Engineering: Target Variable + Feature Creation, Standardization

Customize Target Variable

Player Value Score = (0.7*Engagement Level) + (0.3*In-game purchases)



Generate "Value segment"

```
ValueSegment
Medium Value    19374
High Value      10336
Low Value       10324
```

Generate HourPerWeek as a proxy of time spending variable

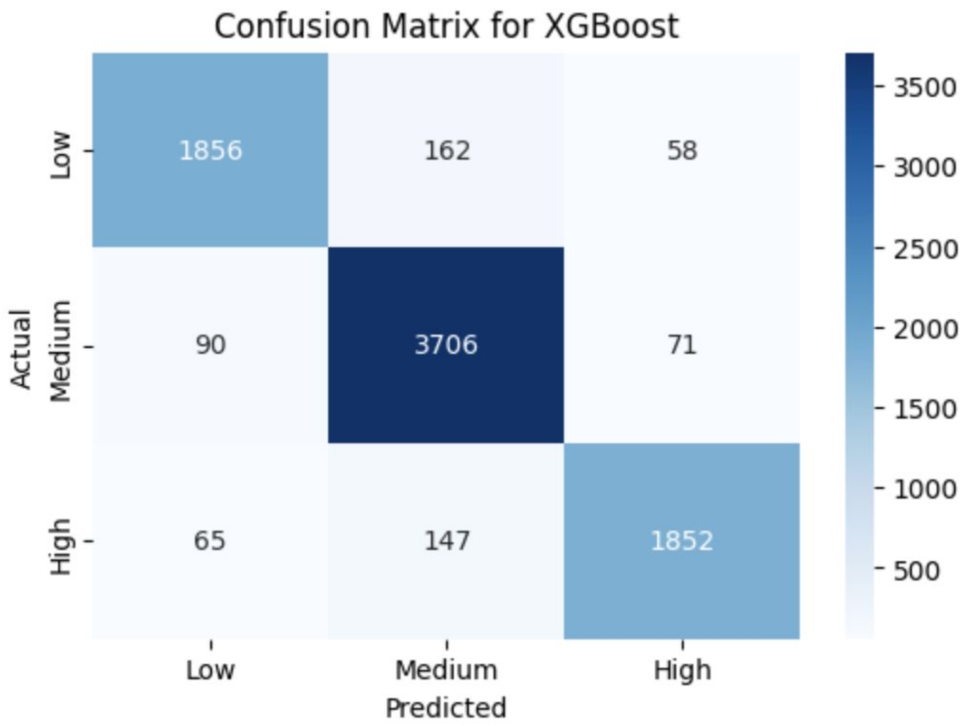
HourPerWeek = (AvgSessionDurationMinutes * SessionsPerWeek)/60

Standardization

$$X_{\text{scaled}} = \frac{X - \mu}{\sigma}$$

Model: Predicting Engagement Level

Best Performing Model:



Key Metrics:

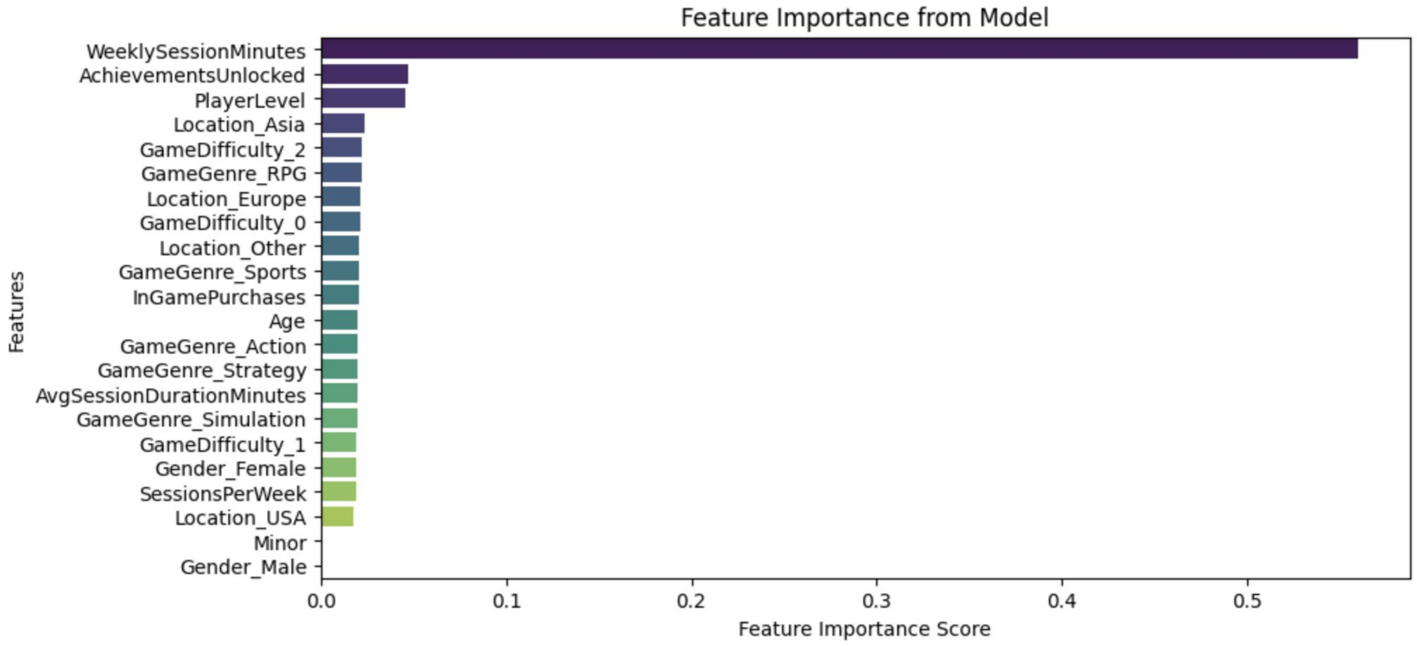
Overall accuracy: 93%

Test Set RMSE: 0.35

Train Set RMSE: 0.30

CV RMSE: 0.30

Feature Importance for XGBoost:

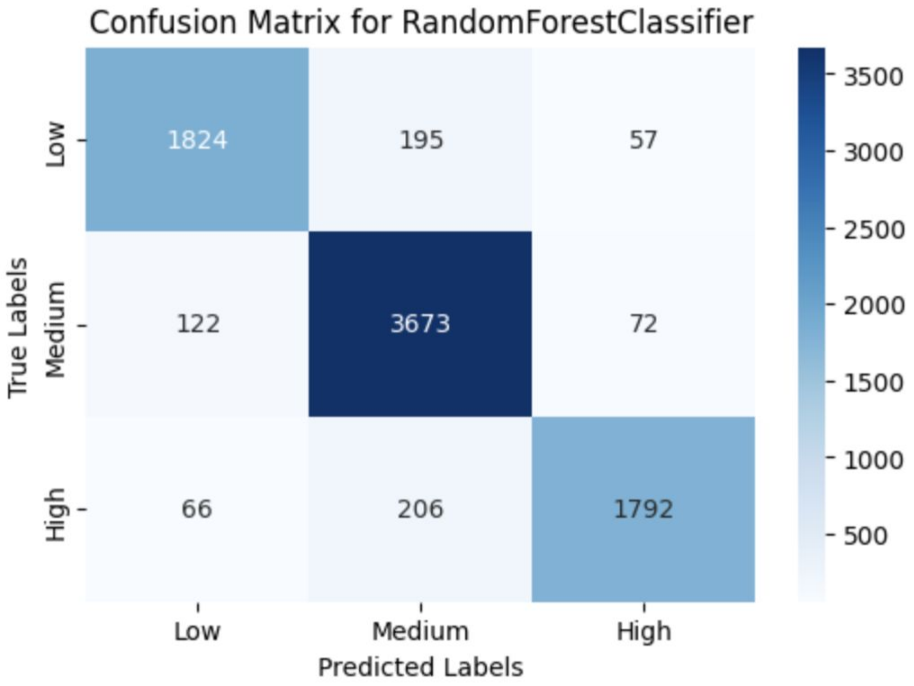


Key Insights:

- **Weekly Session Minutes** is the most important feature
- **Achievements Unlocked** and **Player Level** moderately important
- Maximum prediction accuracy observed for **Medium** Engagement Level

Model: Predicting Engagement Level with further optimization

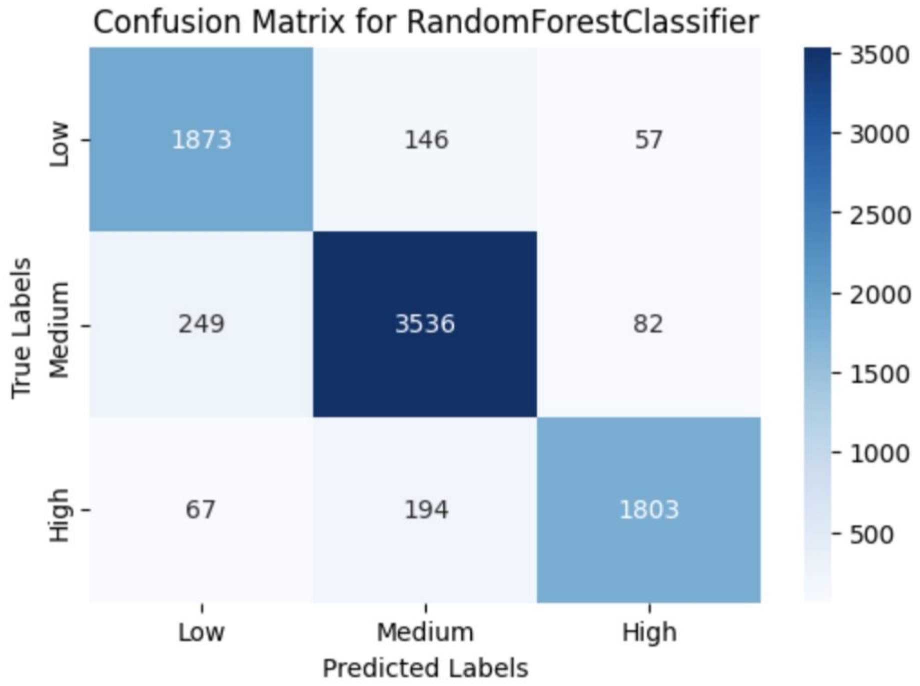
Using only t-SNE Reduced Data:



Key Metrics

91% : Overall accuracy: 90%
0.37 : Test Set RMSE: 0.38
0.35 : Train Set RMSE: 0.37
0.37 : CV RMSE: 0.38

Optimizing for Low Engagement Players:



Key Insights:

- Most of the accuracy maintained even with t-SNE reduced data
- Using class weights of 3 for "Low" and 1 for "Medium" and "High" Engagement levels improves labelling accuracy for Low category
- Higher accuracy for Low category (high churn risk) results in higher mislabeling for other 2 categories

Model: Predicting Customer Value Segments

In sample Statistic | Out-of-sample Statistic | Confusion matrix

Random Forest

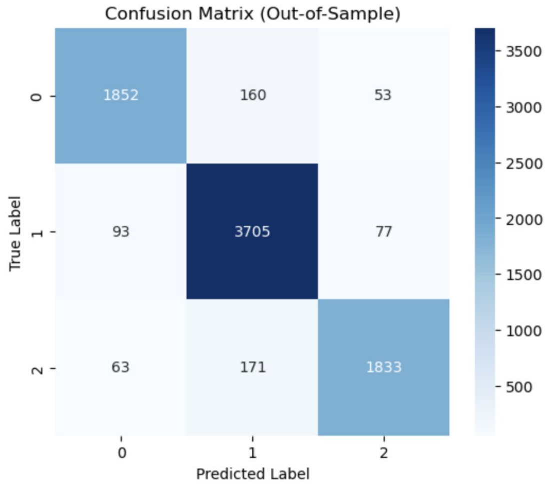
🚀 In-Sample Classification Report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	8259
1	1.00	1.00	1.00	15499
2	1.00	1.00	1.00	8269
accuracy			1.00	32027
macro avg	1.00	1.00	1.00	32027
weighted avg	1.00	1.00	1.00	32027

📊 Out-of-Sample Classification Report:

	precision	recall	f1-score	support
0	0.92	0.90	0.91	2065
1	0.92	0.96	0.94	3875
2	0.93	0.89	0.91	2067
accuracy			0.92	8007
macro avg	0.92	0.91	0.92	8007
weighted avg	0.92	0.92	0.92	8007

10 Fold Cross Validation: 92% Accuracy



XG Boost

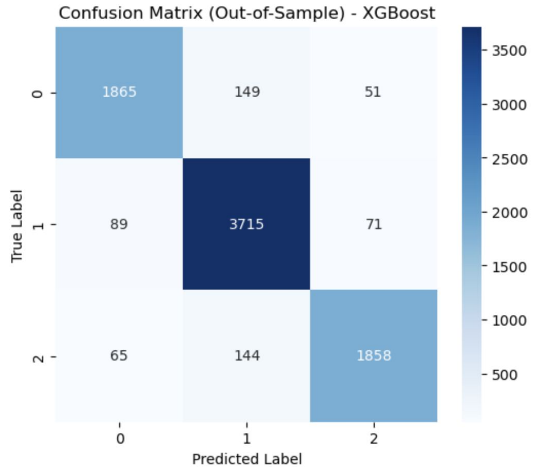
🚀 In-Sample Classification Report (XGBoost):

	precision	recall	f1-score	support
0	0.95	0.91	0.93	8259
1	0.94	0.97	0.96	15499
2	0.95	0.91	0.93	8269
accuracy			0.94	32027
macro avg	0.94	0.93	0.94	32027
weighted avg	0.94	0.94	0.94	32027

📊 Out-of-Sample Classification Report (XGBoost):

	precision	recall	f1-score	support
0	0.92	0.90	0.91	2065
1	0.93	0.96	0.94	3875
2	0.94	0.90	0.92	2067
accuracy			0.93	8007
macro avg	0.93	0.92	0.92	8007
weighted avg	0.93	0.93	0.93	8007

10 Fold Cross Validation: 93% Accuracy



Model selection: XGBoost is our best model
Feature importance: HourPerWeek is the most dominant feature.

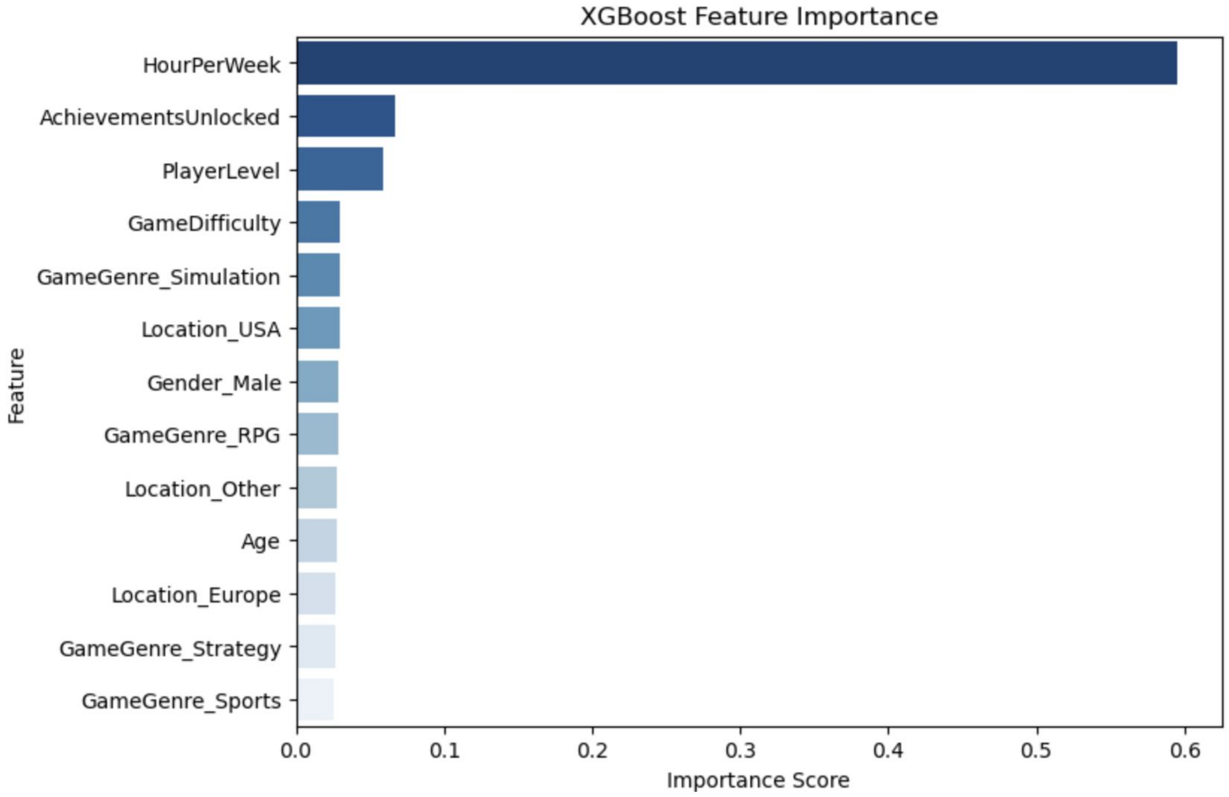
XGBoost is our best-performing model because:

(1) Accuracy: The model achieved the highest accuracy with a cross-validation score of 93%.

(2) No overfitting or underfitting: The model performs consistently across both the training and test datasets, indicating no signs of overfitting. Additionally, it maintains high accuracy and F1 scores in both datasets, demonstrating that it's also not underfitting.

(3) Explainability: The model provides clear interpretability, identifying time spent gaming (HourPerWeek) as the most influential feature for determining a player's Value Segment.

Feature importance from XGBoost model



Recommendation system: Maximizing customer value by tailored recommendations

New player

Gender: Male
Age: 25 years old
Location: USA

Re-iterate
"GameGenre" and "GameDifficulty"
in the model to predict ValueSegment

ValueSegment = f(Age, Gender, Location

GameGenre, GameDifficulty)

Suggest
"GameGenre" and "GameDiffuculty"
that maximize the ValueSegment

$$\max_{Genre,Difficulty} Value\ Segment$$

$$= \boxed{Sport\ game, Medium\ level}$$

Recommendation

Business Value

- Marketing Strategies Personalization
  Develop tailored marketing strategies for each engagement level category.
  For example, offer loyalty benefits to highly engaged players and provide greater playtime incentives to players with low engagement

- Reducing Churn Rate
  Focus on players with low engagement to prevent churn and maintain platform engagement.
  For example, recommend low-difficulty games that lead to more in-game achievements, thereby improving engagement and player satisfaction

- Effective Strategies must focus on encourage increased gameplay
  We found that player engagement is primarily determined by the time spent gaming, rather than age, location, or gender. Thus, the business should focus on strategies that encourage increased gameplay across all player segments, rather than targeting specific demographics.

- Recommendation system
  We recommend implementing a personalized recommendation system for game genre and difficulty, which will help convert new players into high-value customers.

Future Scope

Further Model Enhancements

Balance Feature Importances

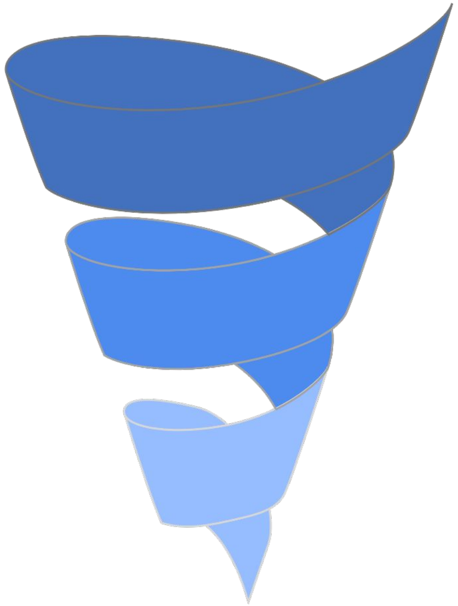
Add additional feature transformations or model engineering steps to ensure no over-dependence on one feature

Add Relevant Features

Incorporate additional features including temporal into Engagement Level Prediction

Customize Model & Metrics

Develop metrics to ensure sustainable results with new production data and prevent data drift



Additional methodologies for further insight

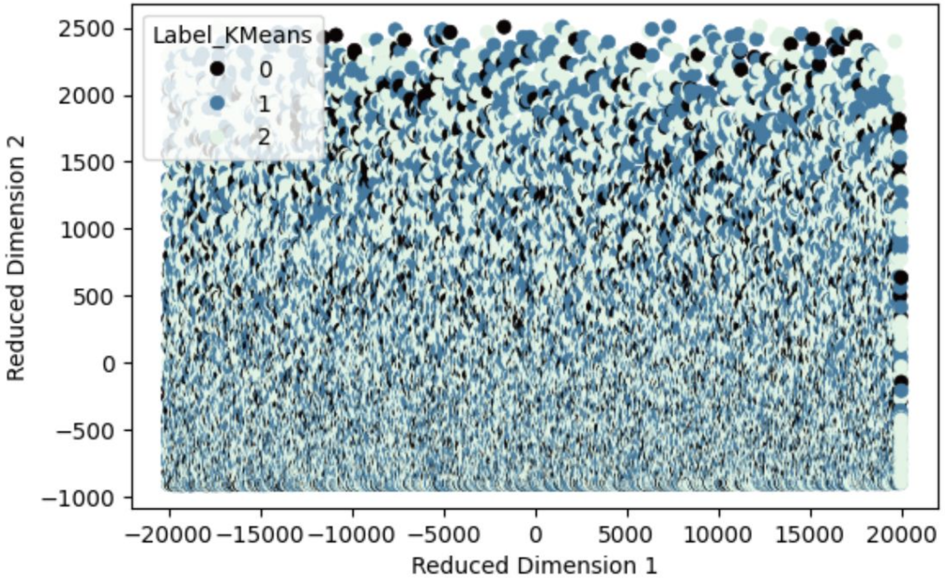
- <u>Graph Network Analysis</u>

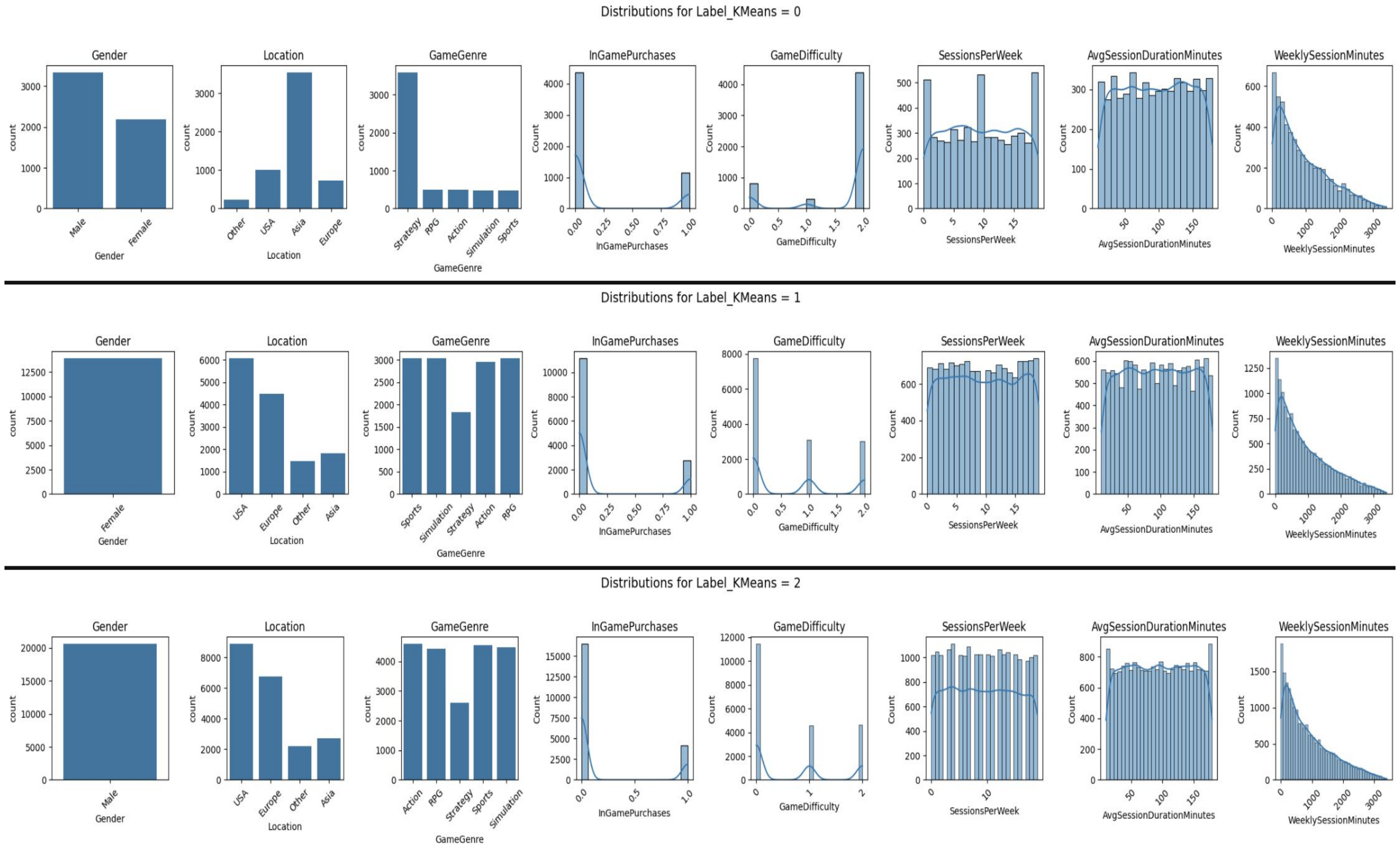
  Graph Network Analysis can be used to model connections between players. In this network, nodes represent players, while edges represent shared attributes such as preferred game genre, location, or age group. This analysis allows us to identify communities within the player base, enhancing community engagement and ultimately generating higher network value.

Thank You!

Appendix

Appendix 1: Clustering

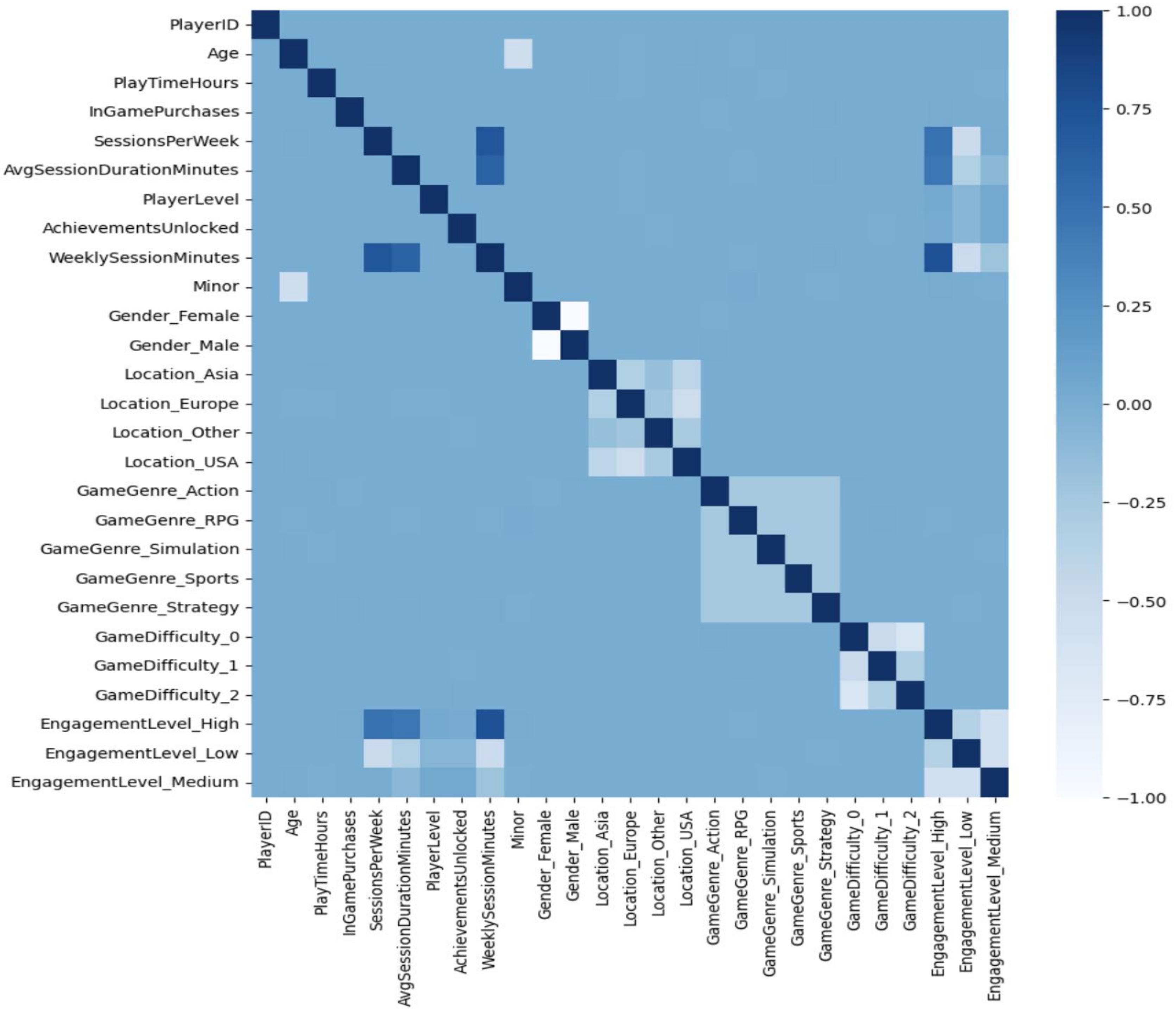




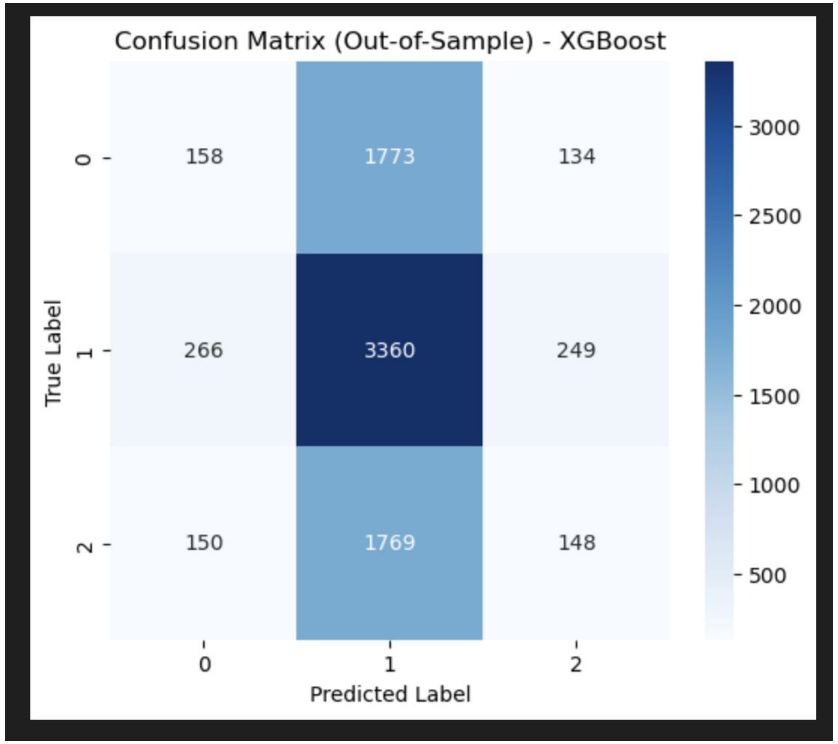
Key Insights

- Clustering using k=3 not useful
- Uniform distribution of many numerical features result in lack of natural clustering within data
- Clustering using only categorical and high-variance features gives better results

Appendix 2: Correlation Matrix



Appendix 3: Recommendation system model performance



🚀 In-Sample Classification Report (XGBoost):

	precision	recall	f1-score	support
0	0.69	0.21	0.32	8259
1	0.54	0.95	0.69	15499
2	0.70	0.20	0.31	8269
accuracy			0.57	32027
macro avg	0.64	0.45	0.44	32027
weighted avg	0.62	0.57	0.50	32027

📊 Out-of-Sample Classification Report (XGBoost):

	precision	recall	f1-score	support
0	0.28	0.08	0.12	2065
1	0.49	0.87	0.62	3875
2	0.28	0.07	0.11	2067
accuracy			0.46	8007
macro avg	0.35	0.34	0.29	8007
weighted avg	0.38	0.46	0.36	8007

Appendix 4: Shapley values