

# Gaming Engagement - Machine Learning

Arthur de Sá Tenório

## Features

Feature	Tipo	Observações
PlayerID	<u>Numérico</u> (Inteiro)	Não representa informação relevante para o problema
Age	<u>Numérico</u> (Inteiro)	Feature pouco informativa, mas pode ser útil em algumas faixas de idade
Gender	<u>Categórico</u> (Male, Female)	Feature não informativa
Location	<u>Categórico</u> (USA, Europe, Asia, Other)	Feature com pouquíssimo potencial distintivo ( <i>não foi útil durante testes</i> )
GameGenre	<u>Categórico</u> (Strategy, Sports, Action, RPG, Simulation)	Feature com pouco poder de distinção de forma isolada, mas pode ser útil em alguns casos ( <i>não foi útil durante testes</i> )
PlayTimeHours	<u>Numérico</u> (Real)	Feature com pouco poder de distinção de forma isolada, mas pode ser útil em alguns casos ( <i>não foi útil durante testes</i> )
InGamePurchases	<u>Booleano</u> (0, 1)	Feature não informativa
GameDifficulty	<u>Categórico</u> (Easy, Medium, Hard)	Feature com pouquíssimo potencial distintivo ( <i>pouco útil durante testes</i> )
SessionsPerWeek	<u>Numérico</u> (Inteiro)	Feature MUITO informativa
AvgSessionDurationMinutes	<u>Numérico</u> (Inteiro)	Feature MUITO informativa
PlayerLevel	<u>Numérico</u> (Inteiro)	Feature informativa ( <i>não tão informativa quanto esperado, mas ainda útil</i> )
AchievementsUnlocked	<u>Numérico</u> (Inteiro)	Feature informativa

		( <i>não tão informativa quanto esperado, mas ainda útil</i> )
EngagementLevel	<u>Categórico</u> (Low, Medium, High)	Variável alvo

### Outras observações

- Foi tentado *Frequency Encoding* para GameGenre e Location, porém ele não funciona muito bem com o classificador Naive Bayes.
- Foi tentado *One Hot Encoding* para GameGenre e Location, porém as features mais confundem a classificação do que ajudam.
- Pouquíssima diferença ao utilizar GameDifficulty. Apenas confundiu um pouco mais a classificação para alto nível de engajamento.
- PlayTimeHours gerou um pouco mais de confusão durante a classificação.
- Aparentemente, o conjunto de features marcados em negrito é o melhor em questão de precisão (matriz de confusão e relatório de classificação abaixo).

```
Matriz de confusão:
[[2154  869   74]
 [ 186 5536   91]
 [   73  593 2435]]

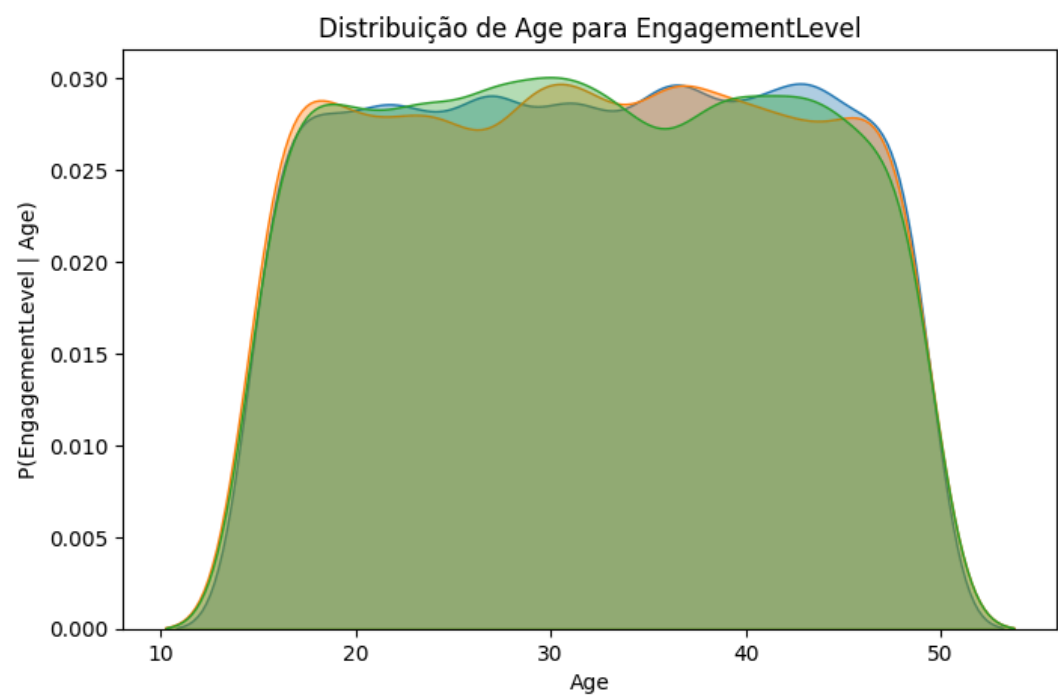
Relatório de classificação:
      precision    recall  f1-score   support

     0       0.89      0.70      0.78      3097
     1       0.79      0.95      0.86      5813
     2       0.94      0.79      0.85      3101

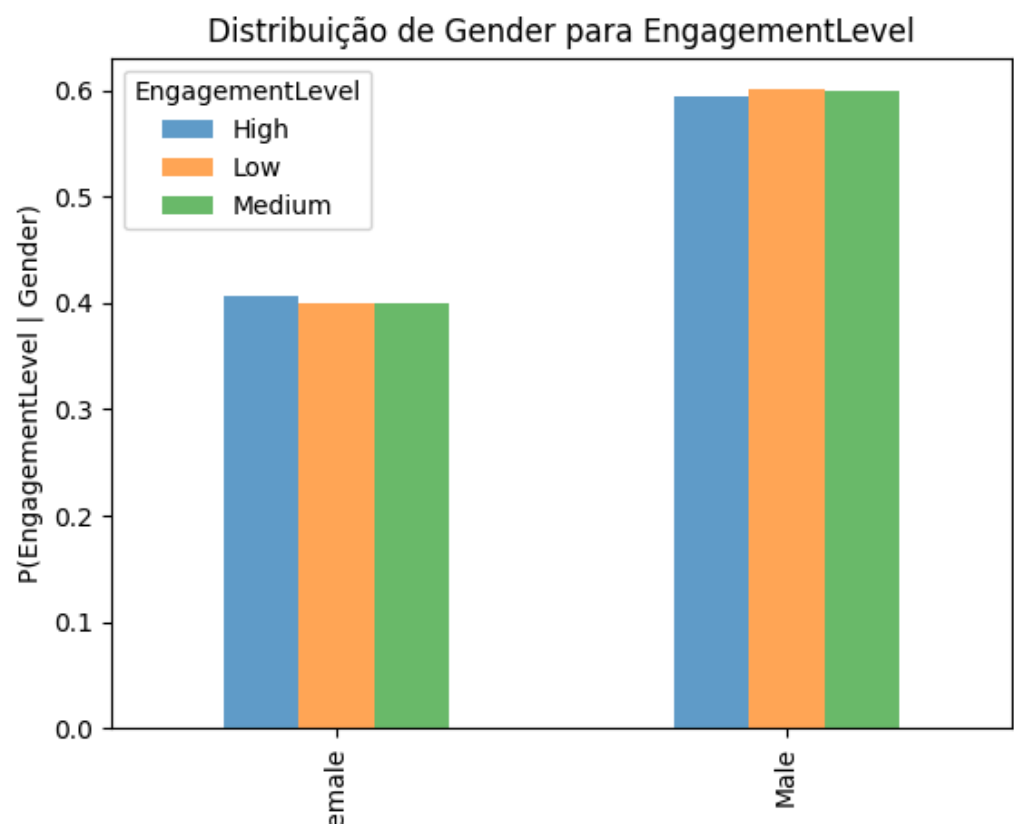
 accuracy          0.84      12011
  macro avg       0.87      0.81      0.83      12011
 weighted avg     0.85      0.84      0.84      12011
```

# Gráficos de verossimilhança

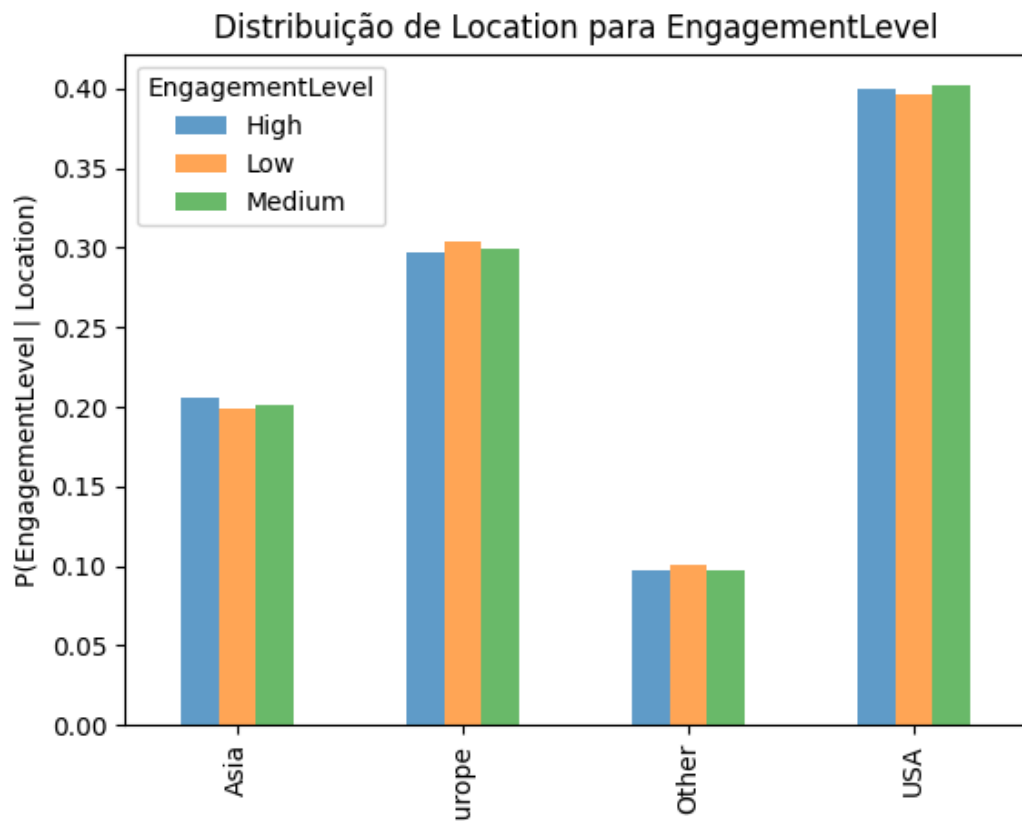
## Age



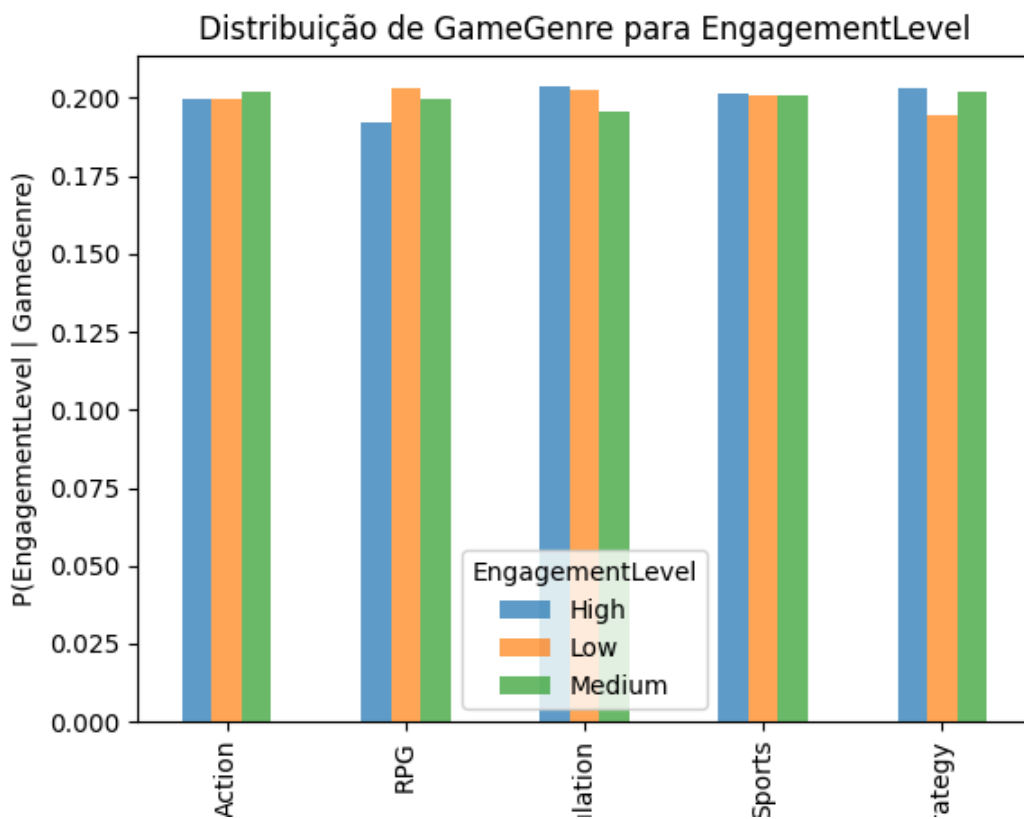
## Gender



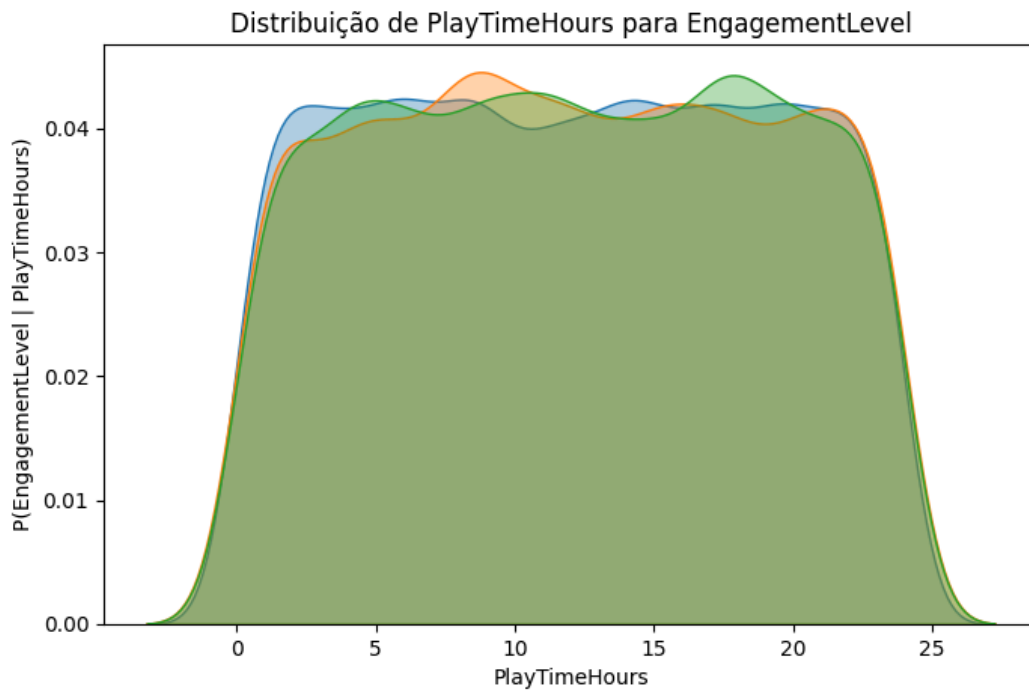
## Location



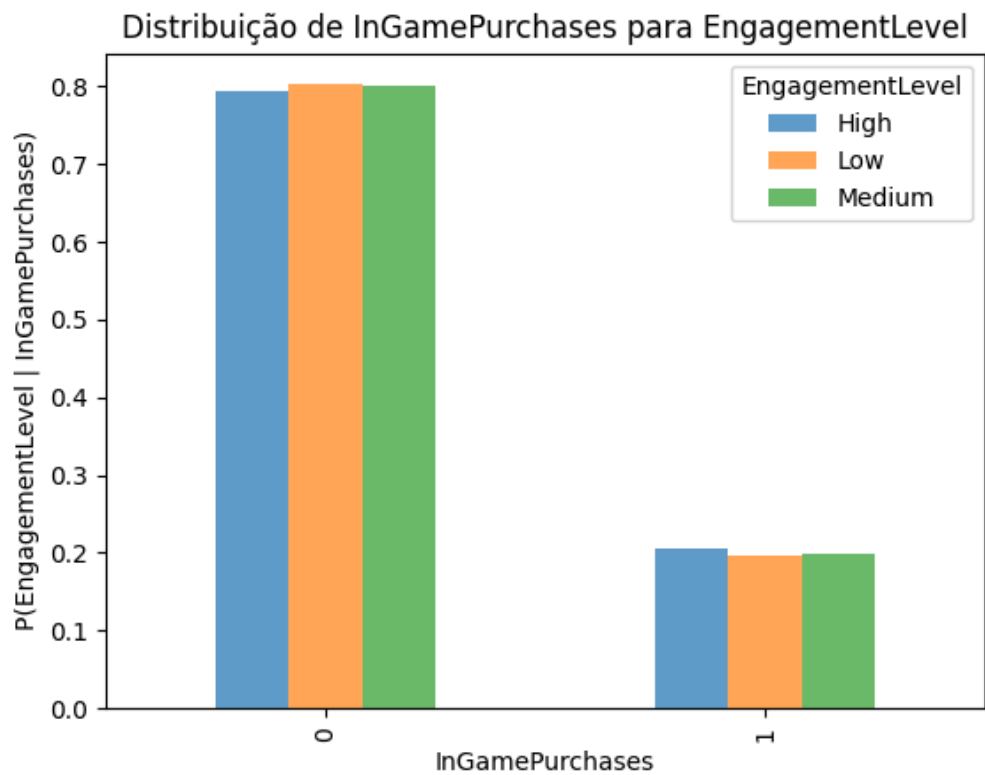
## GameGenre



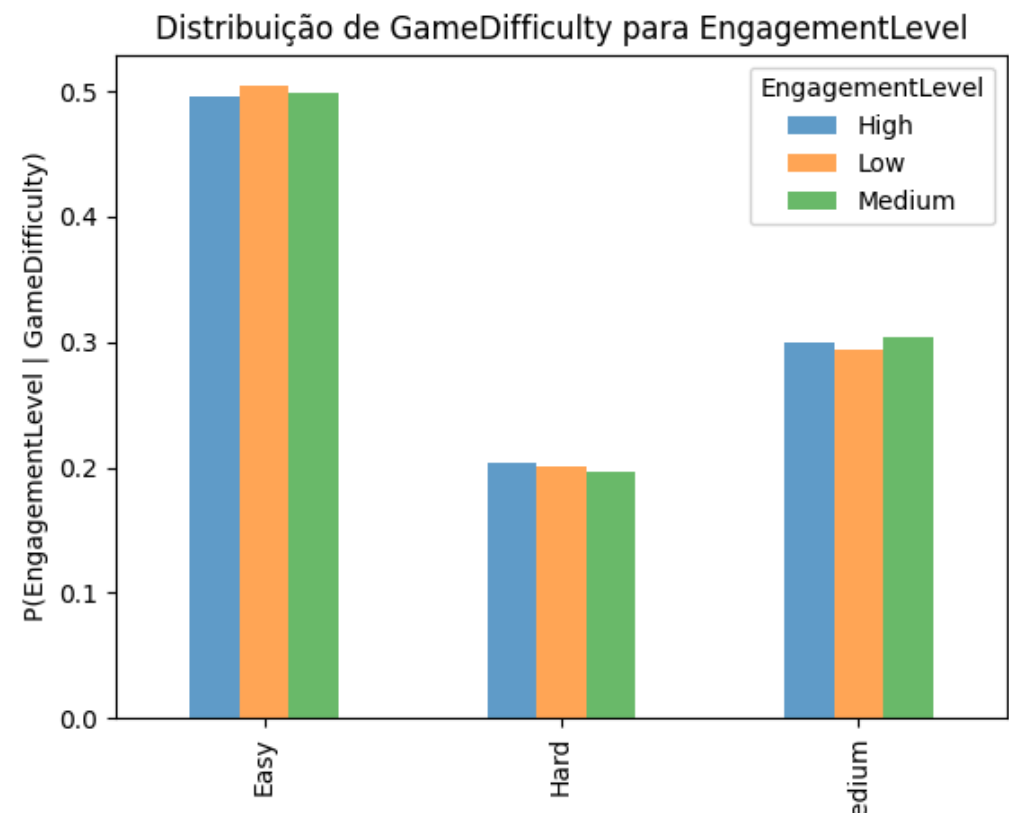
## PlayTimeHours



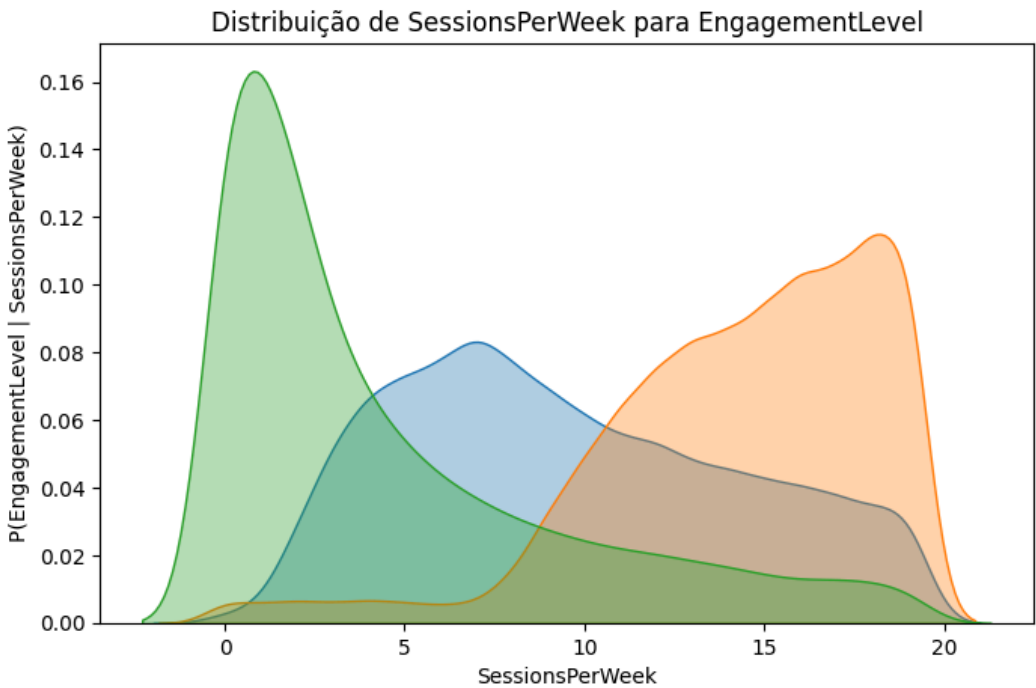
## InGamePurchases



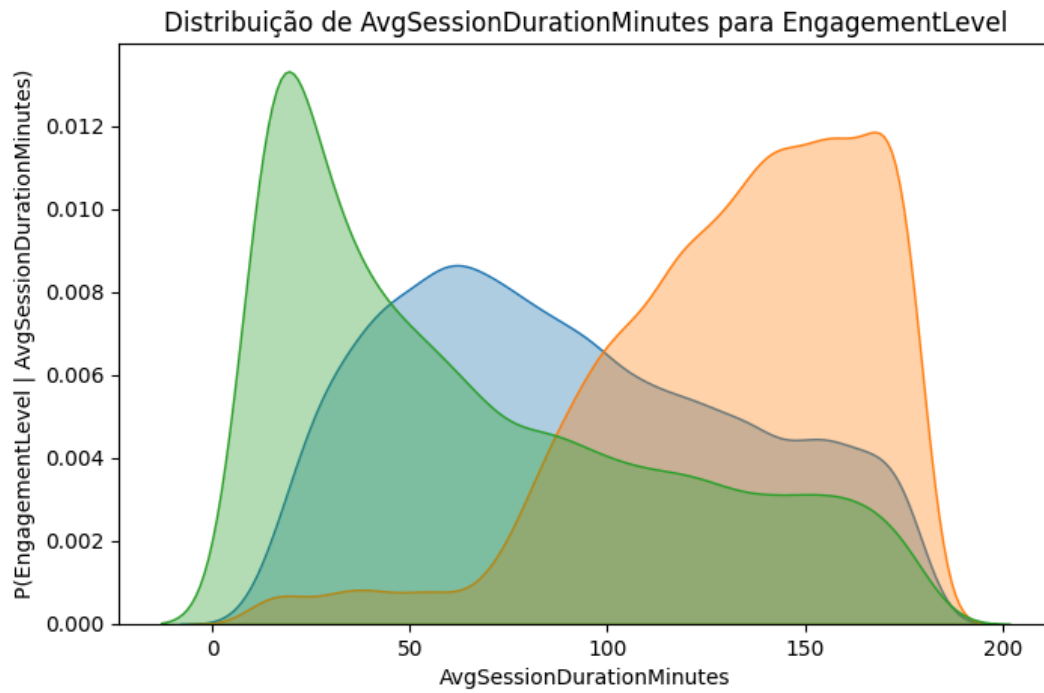
GameDifficulty



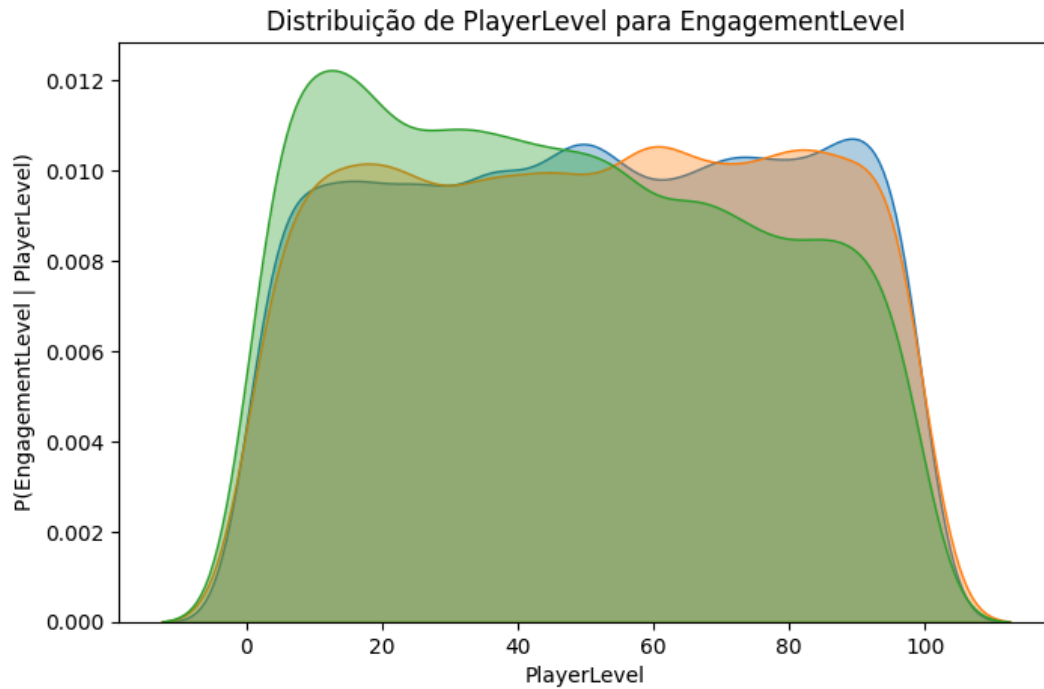
SessionsPerWeek



## ***AvgSessionDurationMinutes***



## ***PlayerLevel***



## ***AchievementsUnlocked***

