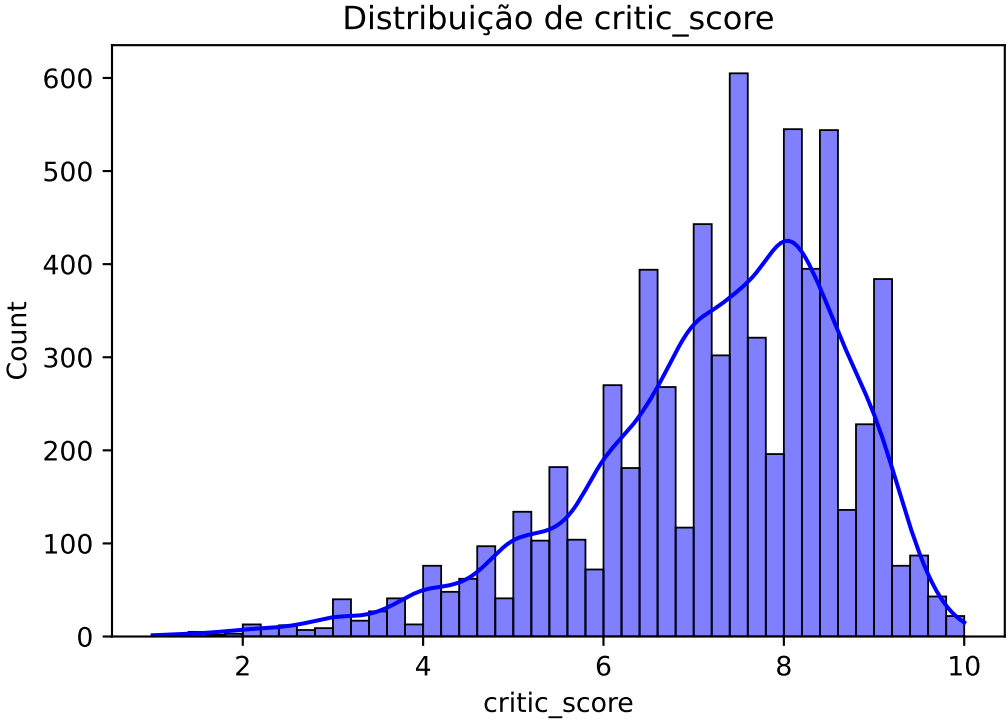
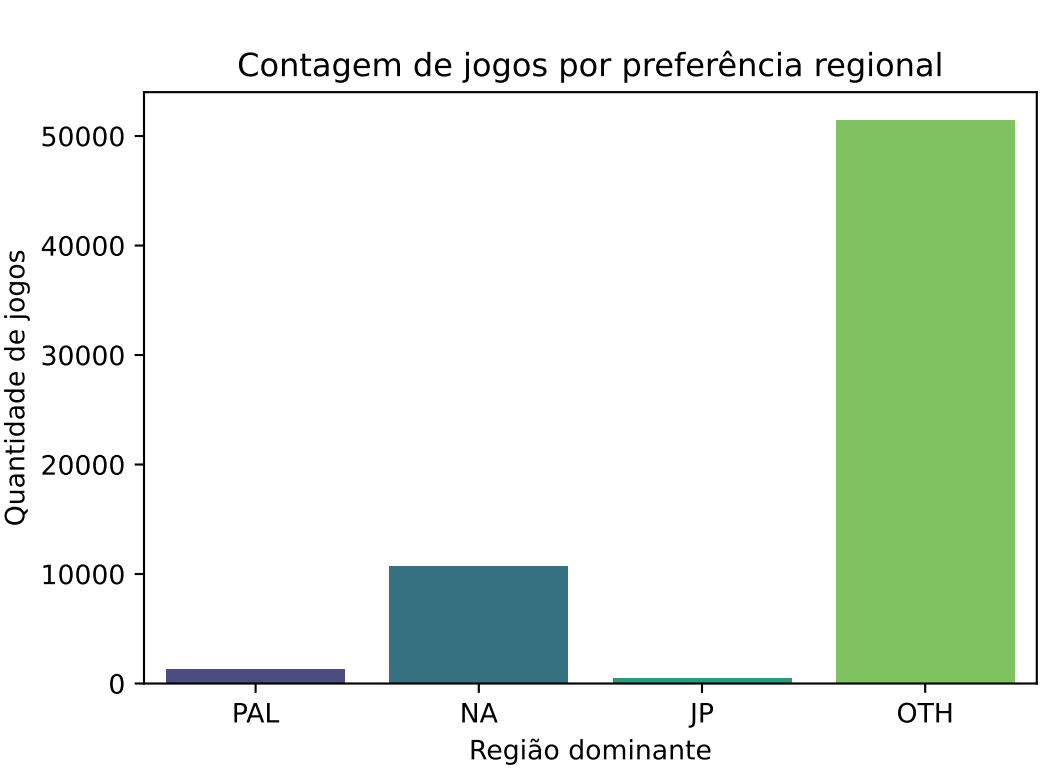
Saída: Leitura e Info do Dataset

```
# SEÇÃO 1: Importando bibliotecas e lendo o CSV
Carregando arquivo CSV...

===== Primeiras linhas do dataset =====
                                          img                         title console  ... other_sales release_date last_update
0  /games/boxart/full_6510540AmericaFrontccc.jpg           Grand Theft Auto V     PS3  ...        3.12   2013-09-17         NaN
1  /games/boxart/full_5563178AmericaFrontccc.jpg           Grand Theft Auto V     PS4  ...        3.02   2014-11-18  2018-01-03
2                    /games/boxart/827563ccc.jpg  Grand Theft Auto: Vice City     PS2  ...        1.78   2002-10-28         NaN
3  /games/boxart/full_9218923AmericaFrontccc.jpg           Grand Theft Auto V    X360  ...        1.42   2013-09-17         NaN
4  /games/boxart/full_4990510AmericaFrontccc.jpg    Call of Duty: Black Ops 3     PS4  ...        2.44   2015-11-06  2018-01-14

[5 rows x 14 columns]

===== Informações gerais =====
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 64016 entries, 0 to 64015
Data columns (total 14 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   img           64016 non-null  object
 1   title         64016 non-null  object
 2   console       64016 non-null  object
 3   genre         64016 non-null  object
 4   publisher     64016 non-null  object
 5   developer     63999 non-null  object
 6   critic_score  6678 non-null   float64
 7   total_sales   18922 non-null  float64
 8   na_sales      12637 non-null  float64
 9   jp_sales      6726 non-null   float64
 10  pal_sales     12824 non-null  float64
 11  other_sales   15128 non-null  float64
 12  release_date  56965 non-null  object
 13  last_update   17879 non-null  object
dtypes: float64(6), object(8)
memory usage: 6.8+ MB
None

===== Quantidade de linhas e colunas =====
(64016, 14)

===== Verificando valores nulos =====
img                 0
title               0
console             0
genre               0
publisher           0
developer          17
critic_score    57338
total_sales     45094
na_sales        51379
jp_sales        57290
pal_sales       51192
other_sales     48888
release_date     7051
last_update     46137
dtype: int64

Gerando histograma de 'critic_score'...
```


Distribuição de critic_score
Count
0
100
200
300
400
500
600
2
4
6
8
10
critic_score

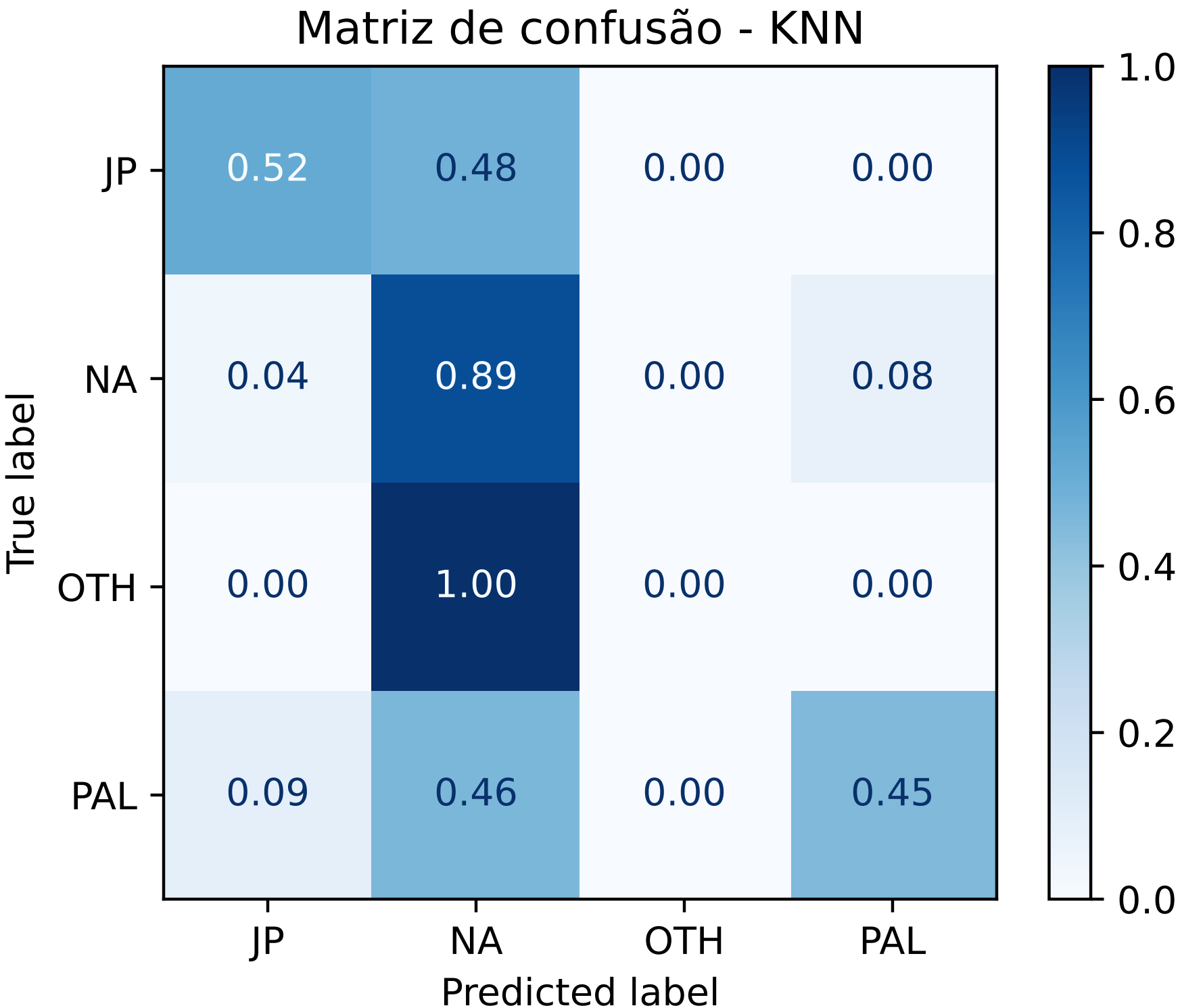
Saída: Region Preference

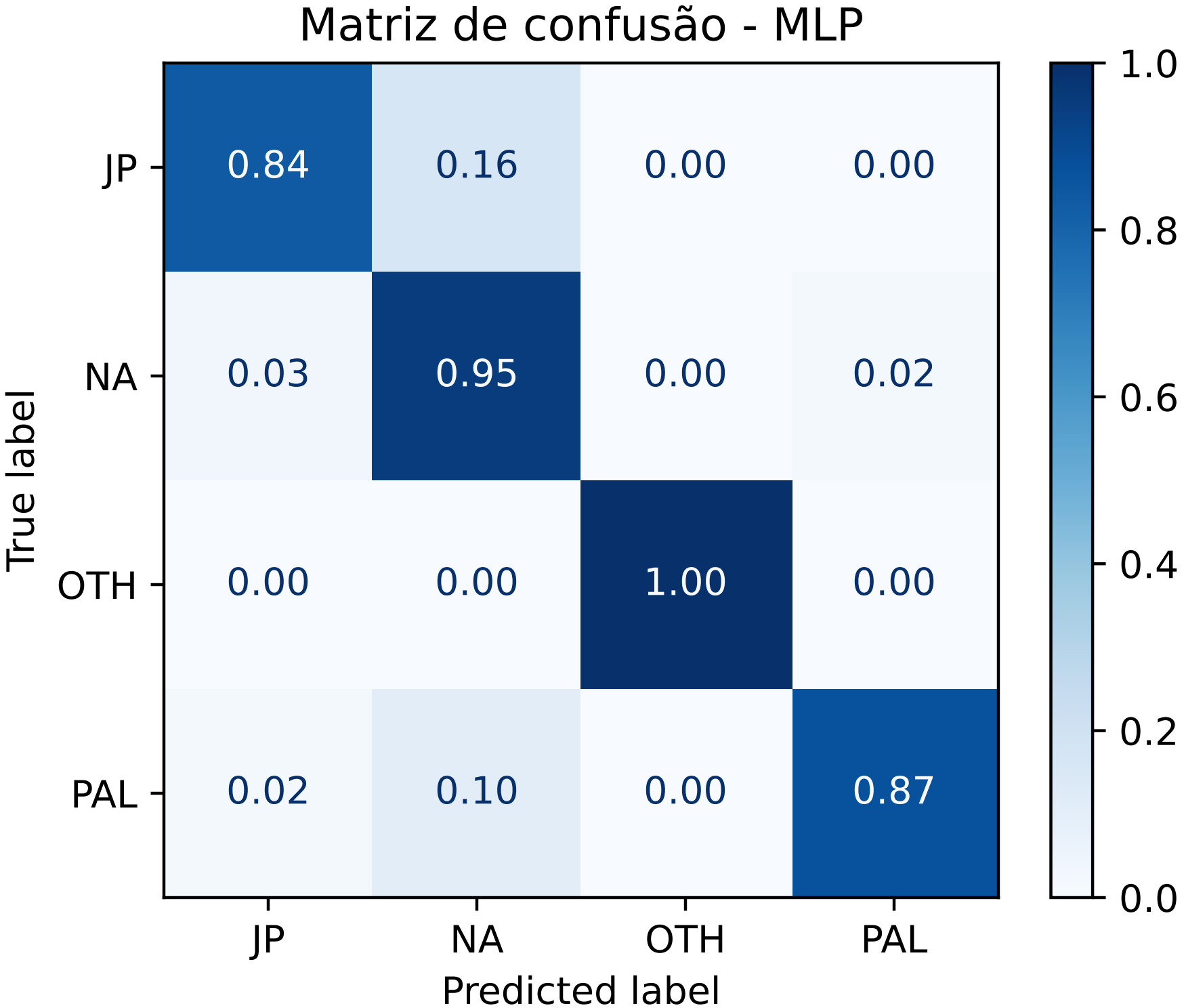
```
SEÇÃO 3: Criando a coluna 'region_preference'

===== Distribuição da classe (region_preference) =====
region_preference
OTH    51437
NA     10755
PAL     1329
JP       495
Name: count, dtype: int64
```

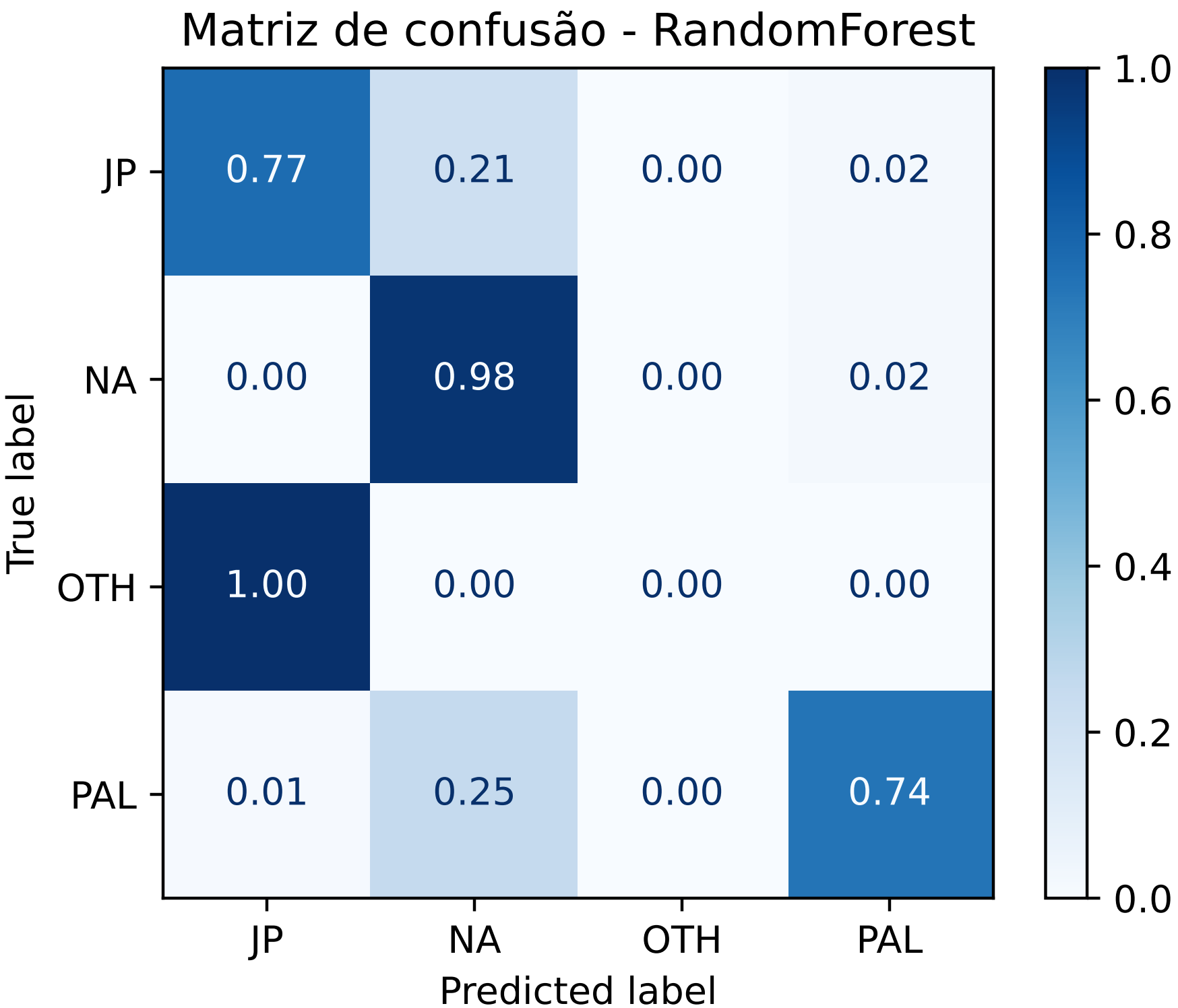

Contagem de jogos por preferência regional
Quantidade de jogos
0
10000
20000
30000
40000
50000
PAL
NA
JP
OTH
Região dominante

Saída: Pré-processamento

```
SEÇÃO 4: Pré-processamento

**************************************************************************

62806 linhas foram removidas por conterem valores nulos nas colunas necessárias.


Linhas que sobraram no dataset: 1210


**************************************************************************

   console  genre  publisher  developer  critic_score  na_sales  jp_sales  pal_sales  other_sales region_preference
0       12      0         58        287           9.4      6.37      0.99       9.85         3.12                PAL
1       13      0         58        287           9.7      6.06      0.60       9.71         3.02                PAL
2       11      0         58        287           9.6      8.41      0.47       5.49         1.78                 NA
4       13     14          7        360           8.1      6.18      0.41       6.05         2.44                 NA
5       20     14          7        165           8.7      9.07      0.13       4.29         1.33                 NA
<class 'pandas.core.frame.DataFrame'>
Index: 1210 entries, 0 to 15044
Data columns (total 10 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   console            1210 non-null   int64
 1   genre              1210 non-null   int64
 2   publisher          1210 non-null   int64
 3   developer          1210 non-null   int64
 4   critic_score       1210 non-null   float64
 5   na_sales           1210 non-null   float64
 6   jp_sales           1210 non-null   float64
 7   pal_sales          1210 non-null   float64
 8   other_sales        1210 non-null   float64
 9   region_preference  1210 non-null   object
dtypes: float64(5), int64(4), object(1)
memory usage: 104.0+ KB
None
===== Formato dos dados =====
X_train: (847, 9), y_train: (847,)
X_test : (363, 9), y_test : (363,)

Classes no treino:
region_preference
NA     550
PAL    182
JP     111
OTH      4
Name: count, dtype: int64

Classes no teste:
region_preference
NA     219
PAL     87
JP      56
OTH      1
Name: count, dtype: int64
```

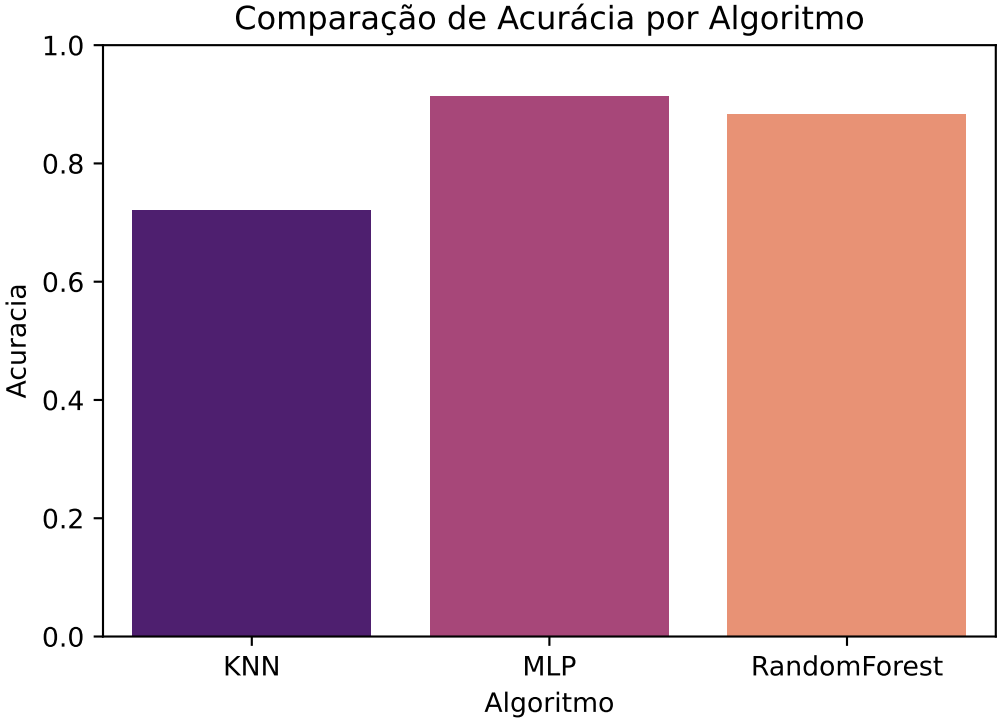

Matriz de confusão - KNN
JP
NA
OTH
PAL
True label
0.52
0.48
0.00
0.00
0.04
0.89
0.00
0.08
0.00
1.00
0.00
0.00
0.09
0.46
0.00
0.45
JP
NA
OTH
PAL
Predicted label
1.0
0.8
0.6
0.4
0.2
0.0


Matriz de confusão - MLP
JP
NA
OTH
PAL
True label
0.84
0.16
0.00
0.00
0.03
0.95
0.00
0.02
0.00
0.00
1.00
0.00
0.02
0.10
0.00
0.87
JP
NA
OTH
PAL
Predicted label
1.0
0.8
0.6
0.4
0.2
0.0


Matriz de confusão - RandomForest
JP
NA
OTH
PAL
True label
0.77
0.21
0.00
0.02
0.00
0.98
0.00
0.02
1.00
0.00
0.00
0.00
0.01
0.25
0.00
0.74
JP
NA
OTH
PAL
Predicted label
1.0
0.8
0.6
0.4
0.2
0.0

Saída: Resultados Finais

```
===== Comparativo final =====
      Algoritmo  Acuracia  Precisao    Recall
0           KNN  0.721763  0.713054  0.721763
1           MLP  0.914601  0.915172  0.914601
2  RandomForest  0.884298  0.887134  0.884298

=== KNN ===
Acurácia : 0.7218
Precisão : 0.7131
Recall   : 0.7218

=== MLP ===
Acurácia : 0.9146
Precisão : 0.9152
Recall   : 0.9146

=== RandomForest ===
Acurácia : 0.8843
Precisão : 0.8871
Recall   : 0.8843
```


Comparação de Acurácia por Algoritmo
1.0
0.8
0.6
0.4
0.2
0.0
Acuracia
KNN
MLP
RandomForest
Algoritmo