

Analises:

DataSet: o DS é representado no arquivo combined_news_data.csv, que combina em média 1400 notícias por time de todos os times (20) da série A brasileira, sendo assim tendo em média 28.000 notícias.

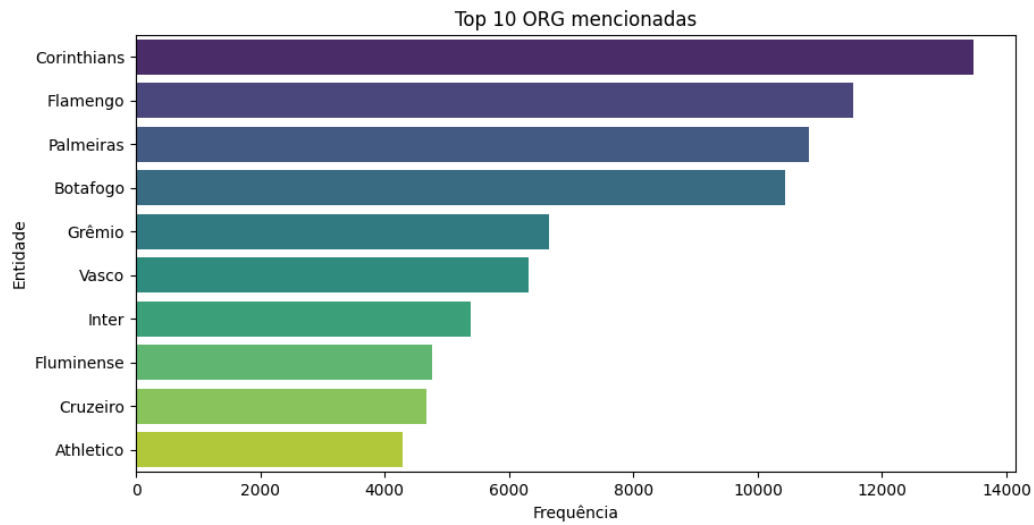
- 1) **Rodar NER na base completa, contabilizando a frequência de entidades nomeadas, incluindo uma interpretação dos resultados.**

Entities que tentei capturar:

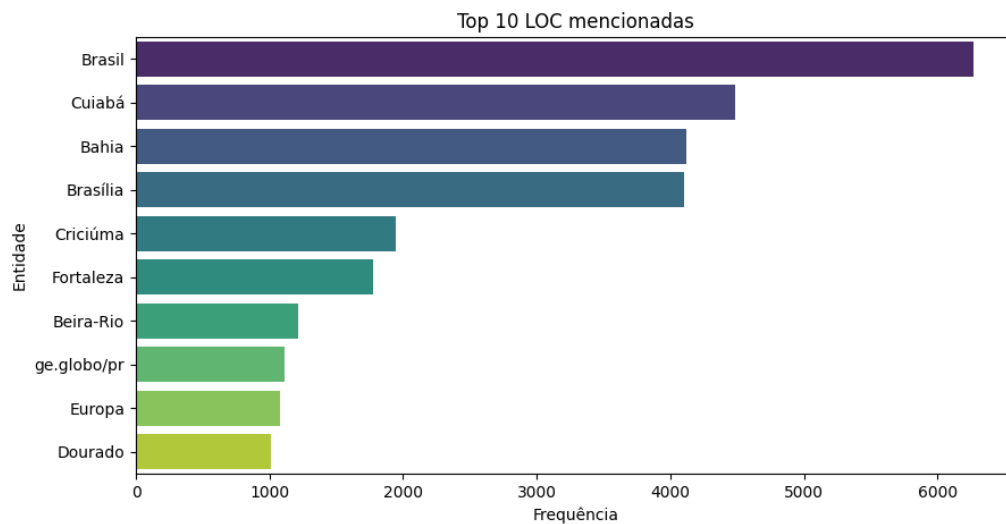
```
print("Rodando NER na base completa...")
entities = {
    "PERSON": [],
    "ORG": [],
    "LOC": [],
    "EVENT": [],
    "DATE": [],
    "TIME": [],
    "MONEY": [],
    "PERCENT": [],
    "QUANTITY": [],
    "ORDINAL": [],
    "CARDINAL": [],
    "GPE": [],
    "WORK_OF_ART": [],
    "LAW": [],
    "LANGUAGE": [],
    "PRODUCT": [],
    "NORP": [],
    "FAC": []
}
```

Somente duas entities trouxeram resultados ORG e LOC, isso deve acontecer pelo linguajar “futebolístico” que o DS tem...

ORG



A obtenção de Organizações foi perfeita, nenhum erro, todas as 10 são realmente instituições/organizações que no nosso contexto são os times



Das LOC's, teoricamente temos todos certos também, temos cidades, estádios, continentes, países. E o único fora disso seria o GE.globo, mas ele identifica pelo /pr.

```
Nenhuma entidade encontrada para EVENT.  
  
Nenhuma entidade encontrada para DATE.  
  
Nenhuma entidade encontrada para TIME.  
  
Nenhuma entidade encontrada para MONEY.  
  
Nenhuma entidade encontrada para PERCENT.  
  
Nenhuma entidade encontrada para QUANTITY.  
  
Nenhuma entidade encontrada para ORDINAL.  
  
Nenhuma entidade encontrada para CARDINAL.  
  
Nenhuma entidade encontrada para GPE.  
  
Nenhuma entidade encontrada para WORK_OF_ART.  
  
Nenhuma entidade encontrada para LAW.  
  
Nenhuma entidade encontrada para LANGUAGE.  
  
Nenhuma entidade encontrada para PRODUCT.  
  
Nenhuma entidade encontrada para NORP.  
  
Nenhuma entidade encontrada para FAC.  
  
Process finished with exit code 0
```

NO DS acontece isso, tentei parametrizar mas acredito que, novamente, é pelo linguajar..