



Data Visualization part 3

Comparando Gráficos

Recuperando a variável dados

```
tmdb = pd.read_csv('../arquivos/tmdb_5000_movies.csv')
total_por_lingua = tmdb.original_language.value_counts() # mostra a contagem de cada língua
total_geral = total_por_lingua.sum() # soma todas as categorias
total_de_ingles = total_por_lingua.loc['en'] # localiza (soma), tudo que encontrar da categoria en
total_do_resto = total_geral - total_de_ingles # subtrai total - categoria en
dados = {
    'lingua': ['ingles', 'outros'],
    'total': [total_de_ingles, total_do_resto]
}
dados
pd.DataFrame(dados) # transforma o dicionário criado em em Data Frame com o pandas
dados = pd.DataFrame(dados)
```

Plotando e Comparando os Gráficos

```
sns.barplot(x='lingua', y='total', data=dados)
```

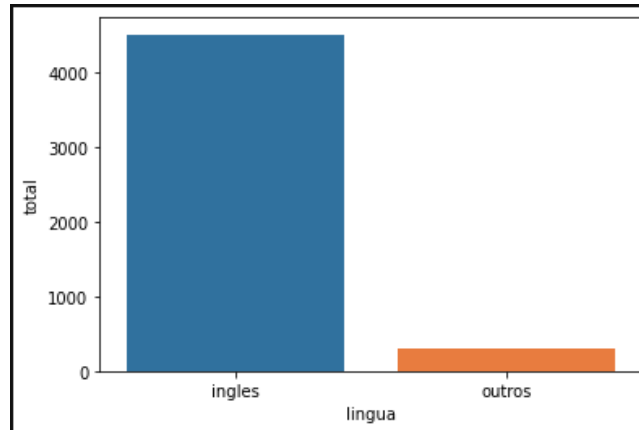


gráfico de barras com seaborn

```
plt.pie(dados.total, labels=dados.lingua)
```

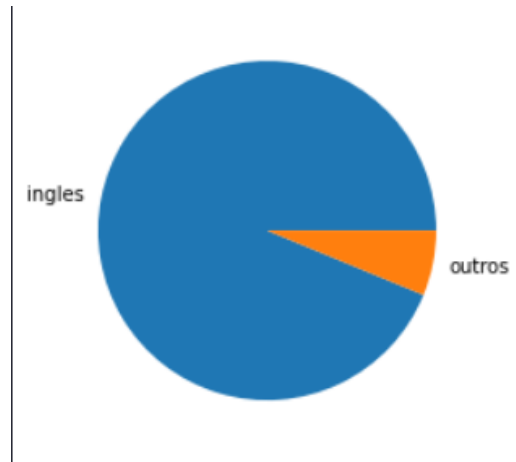


gráfico de pizza com matplotlib

- apesar dos dois gráficos mostrar que a língua inglesa é predominante, no gráfico de barras é muito mais evidente essa diferença, portanto nesse caso, é preferível usá-lo

Analizando as Outras Línguas

- para analisar as outras línguas devemos extrair o que não importa, nesse caso, a língua inglesa
- pegamos o data frame completo e usamos a query
- buscamos os filmes que a original_language não é inglês

```
tmbd.query("original_language != 'en'") # atenção ao uso correto das áspas
total_por_lingua_de_outros_filmes = tmbd.query("original_language != 'en'").original_language.value_counts()
# contando apenas as colunas com outra língua
total_por_lingua_de_outros_filmes
filmes_sem_lingua_original_em_ingles = tmbd.query("original_language != 'en'")
# variável grande, mas explícita
```

Plotando com Catplot

```
sns.catplot(x='original_language', kind='count', data=filmes_sem_lingua_original_em_ingles)
```

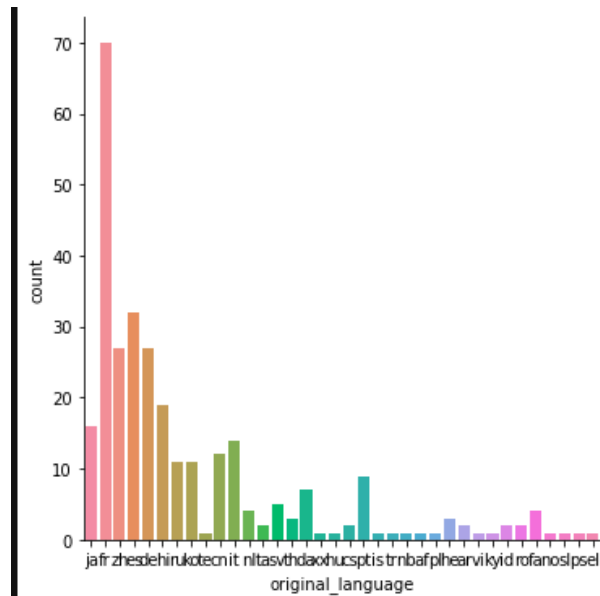


gráfico de barras do seaborn com filmes que não tem língua original em inglês