

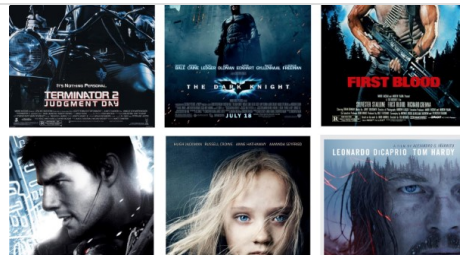
Tipos de dados

Mais fontes de dados

- THE MOVIE DATA BASE - tmdb 5000
 - conjunto de dados com 5000 informações, armazenado no kaggle

TMDB 5000 Movie Dataset
Metadata on ~5,000 movies from TMDb

<https://www.kaggle.com/tmdb/tmdb-movie-metadata>



Analisando dados do arquivo movies

- carregando o arquivo tmdb_5000_movies.csv

Coluna original_language

- dados do tipo categoria nominal
- não existe uma ordem de grandeza

- mesmo considerando uma ordem alfabética, não existe uma língua que vem antes da outra
- não existe ranking nem ordem, portanto é uma variável categórica

Variável Categórica Ordinal

- parecido com a nominal, mas existe uma ordem entre elas
- Ex: primeiro grau < segundo grau < terceiro grau
- apesar de existir uma ordem não é possível aplicar operações aritméticas entre elas

Coluna budget

- variável orçamento é do tipo quantitativa contínua
- quantitativa por que representa um valor
- contínua por que pode ser infinitos valores, inclusive casas decimais

Coluna vote_count

- quantidade de votos é uma variável do tipo quantitativa intervalar
- representa uma quantidade
- aumenta sempre no intervalo de 1 voto
- não existe meio voto
- podem existir outros intervalos
- nota do filme, por exemplo:
 - 0.5, 1, 1.5, ..., 5
 - intervalo de 0.5
 - as notas são de 0.5 a 5, aumentando sempre 0.5