**Para saber mais: The Movie Database**

Neste curso, estamos utilizando uma base de dados bem conhecida no Kaggle: o [TMDB 5000 Movie Dataset](https://www.kaggle.com/datasets/tmdb/tmdb-movie-metadata). É um conjunto de dados com informações gerais sobre filmes para as pessoas que querem praticar a ciência de dados em um outro contexto no ambiente do Kaggle. Este conjunto de dados contém detalhes de quase 5.000 filmes, incluindo informações como título, gênero, elenco, equipe, receita bruta, orçamento e muito mais.

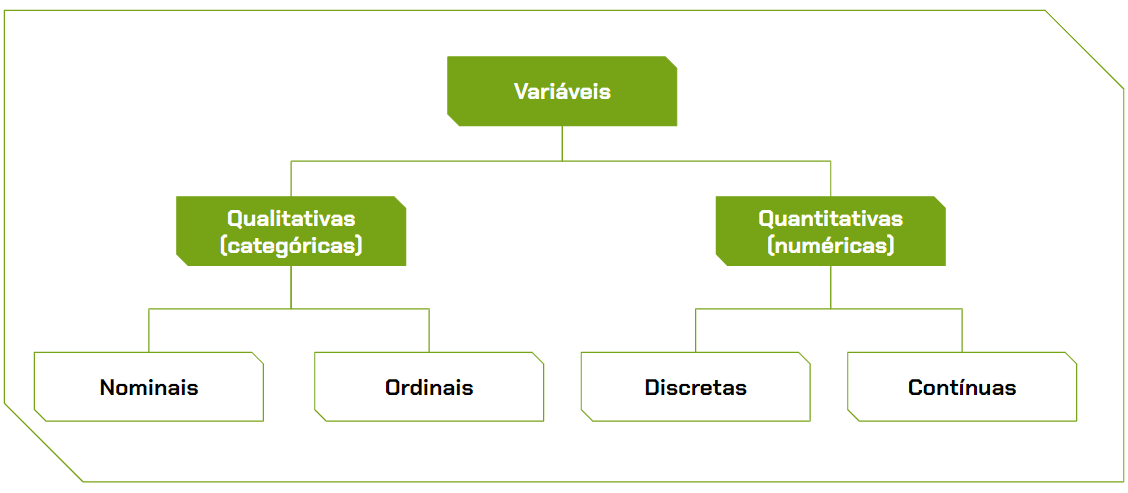
Cientistas, analistas de dados e entusiastas de cinema podem explorar esses dados para descobrir insights interessantes sobre tendências de filmes, preferências de público, desempenho financeiro de filmes e muito mais. Com a variedade de informações disponíveis, os usuários podem realizar a análise exploratória dos dados, identificar padrões dos gêneros mais populares, analisar a influência de atores e diretores no desempenho de filmes, criar modelos de recomendação de filmes e muito mais.

Além disso, o conjunto de dados também é uma ótima ferramenta de aprendizado permitindo a prática de habilidades de manipulação de dados, visualização e análise exploratória de dados em um contexto real.

**Para saber mais: classificando as variáveis**

Variáveis são características, atributos ou quantidades de uma amostra ou população que podem ser mensuradas, contadas ou categorizadas. Elas são fundamentais para descrever e entender os fenômenos em estudo. As variáveis podem ser classificadas basicamente em dois tipos: **qualitativas** (ou categóricas) e **quantitativas** (ou numéricas).

Na imagem abaixo, podemos observar os diferentes tipos de variáveis, seja por sua natureza, quanto pelas suas derivações:



### Variáveis Qualitativas

Também conhecidas como categóricas, representam características que não podem ser medidas numericamente, mas podem ser categorizadas, ou seja, representam uma classificação. Estas variáveis podem ser **nominais** ou **ordinais**.

Observação: a depender dos tipos dos dados, é possível que uma **variável qualitativa** possa ser representada por números. Por exemplo, uma coluna que representa uma característica em que 0 simboliza “Não” e 1 simboliza “Sim”. Os números aqui servem para resumir a representação dos dados e não devem ser utilizados para cálculo, por exemplo, de média.

#### Nominal

#### Uma variável qualitativa nominal é aquela que classifica os elementos em classes ou categorias em relação ao tipo do dado representado. Ela não possui uma relação de grandeza ou ordem dos dados

**Exemplos:** Nome de produtos, departamentos, métodos de pagamento ou variáveis geográficas, como, cidade, estado e região.

#### Ordinal

Uma variável qualitativa ordinal também é aquela que classifica os elementos em classes ou categorias, porém existe uma **ordem natural ou hierárquica** entre as categorias.

**Exemplos:** grau de escolaridade (ensino infantil, ensino fundamental, ensino médio), nível de satisfação do cliente (insatisfeito, neutro e satisfeito) e nível socioeconômico.

### Variáveis Quantitativas

Representam medidas numéricas ou quantidades observáveis que podem ser expressas em números. Elas são resultantes de **contagens** (conjunto finito de valores) ou **mensurações** (conjunto infinito de valores). Estas variáveis podem ser **discretas** ou **contínuas**.

#### Discreta

Uma variável quantitativa discreta é aquela que apresenta valores que podem ser contados e geralmente inteiros. **Exemplos:** quantidade de produto vendido, número de cômodos em uma casa e idade.

#### Contínua

Uma variável quantitativa contínua é aquela que apresenta valores que podem assumir qualquer valor dentro de um intervalo específico e geralmente são expressos como números reais.

**Exemplos:** peso de um produto, altura de uma pessoa, tempo e temperatura de um experimento.

Compreender as variáveis e onde elas se posicionam dentro dessas classificações facilita bastante na escolha das técnicas, visuais e análises para cada tipo de dado. Por isso, quando estiver mexendo com conjuntos de dados em seus projetos, analise o que cada coluna pode representar e como poderá te auxiliar na investigação por meio da ciência de dados.