As **variáveis categóricas** são um tipo de variável usado em Ciência de Dados para representar dados que podem ser classificados em diferentes categorias ou grupos. Por exemplo, se uma pessoa cientista de dados estiver analisando o desempenho acadêmico de estudantes em uma escola, ela pode usar variáveis categóricas para classificar o desempenho dessas pessoas em diferentes grupos, como "excelente", "bom" ou "regular". Isso permite identificar padrões ou tendências no desempenho.

Existem muitos exemplos de variáveis categóricas na Ciência de Dados. Algumas das mais comuns incluem a cor dos olhos, o tipo sanguíneo, a marca de um carro ou a escolaridade. Essas variáveis não possuem uma escala numérica e não podem ser medidas em termos de magnitude ou intensidade. Em vez disso, elas são usadas para classificar dados em diferentes grupos ou categorias.

Dentre os tipos de variáveis, temos:

* **Variáveis categóricas nominais**

São aquelas que não possuem uma ordem ou hierarquia específica entre as categorias. Por exemplo, se uma pessoa cientista de dados estiver analisando a preferência musical de um grupo de pessoas, ela pode usar variáveis categóricas nominais para classificar as pessoas em diferentes grupos, como: "rock", "jazz" ou "pop".

* **Variáveis categóricas ordinais**

Possuem uma ordem específica entre as categorias. Por exemplo, se a pessoa cientista de dados estiver analisando a escolaridade de um grupo de pessoas, ela pode usar esse tipo de variável para classificar as pessoas em diferentes grupos, como: "ensino fundamental completo", "ensino médio completo" ou "ensino superior completo" e assim por diante.

* **Variáveis categóricas binárias**

São um tipo especial de variável categórica que possui apenas duas categorias possíveis, por exemplo: sim/não, verdadeiro/falso ou presente/ausente. As variáveis categóricas binárias são úteis porque permitem que cientistas de dados analisem a distribuição de dados em apenas duas categorias possíveis.