Entrega 1 - Álgebra Linear, Vetores e Geometria Analítica

# Descrição da Entrega

Tema: Representação e Manipulação de Dados  
  
Formato de Entrega: Código-Fonte, Relatório com Capturas de Tela e Explicação Matemática.  
  
Objetivo: Estruturar um conjunto de dados relevante para um problema prático e representá-lo usando vetores e matrizes.  
  
Atividades:  
- Escolher ou coletar um conjunto de dados relevante para uma aplicação de IA.  
- Representar esses dados na forma de vetores e matrizes, utilizando Python e NumPy.  
- Implementar operações básicas de adição, multiplicação e/ou transposição de matrizes ou outra operação utilizada em sala.

# Estrutura do Relatório

Este relatório contém explicações detalhadas sobre cada célula de código implementada no notebook, além de exemplos práticos com vetores e matrizes em Python. O relatório está organizado da seguinte forma:

* Importação de Bibliotecas
* Definição de Vetores e Matrizes
* Operações Matemáticas
* Agrupamento por Categoria
* Cálculo de Coeficientes de Regressão
* Centralização e Correlação pelo Cosseno
* Projeção Linear
* Conclusão

# Explicação das Células de Código

## 1. Importação de Bibliotecas

Nesta parte do código, foi feita a importação da biblioteca NumPy, que é essencial para a manipulação de arrays, vetores e matrizes em Python. O NumPy permite realizar operações matemáticas de forma otimizada.

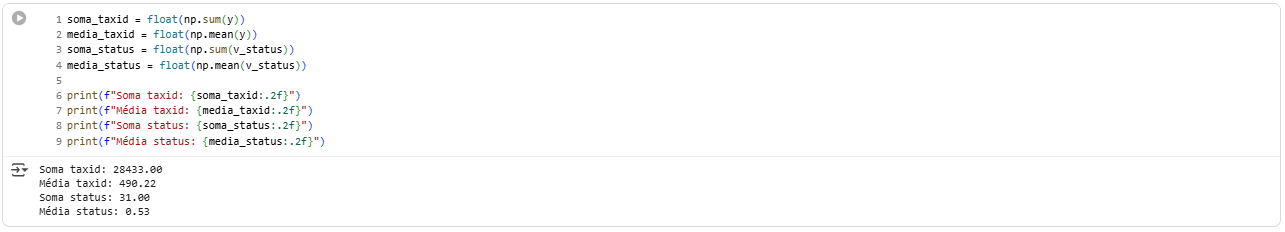
## 

## 2. Definição de Vetores e Matrizes

Aqui foram criados vetores e matrizes representando os dados escolhidos. Essas estruturas são fundamentais para representar informações em IA e aplicações matemáticas. Os vetores são listas unidimensionais, enquanto as matrizes são representadas como tabelas bidimensionais.



## 3. Operações Matemáticas

Foram implementadas operações como:  
- Adição de matrizes: soma elemento a elemento.  
- Multiplicação de matrizes: combina informações de duas matrizes em uma nova matriz.  
- Transposição: troca linhas por colunas.

**4. Agrupamento por Categoria**

O código converte a coluna taxid em um vetor numérico e identifica as categorias únicas da coluna gender. Em seguida, cria uma matriz binária (one-hot) para representar cada gênero. Usando multiplicação matricial, calcula a soma de taxid para cada categoria de gênero. Por fim, organiza os resultados em um DataFrame e exibe o total de vendas por gênero de forma ordenada.



## **5. Cálculo de Coeficientes de Regressão**

O código calcula os coeficientes de uma regressão linear. Primeiro, monta os produtos matriciais XᵀX e Xᵀy para usar no método da eliminação de Gauss, obtendo beta\_gauss. Caso haja erro, o cálculo é ignorado. Em seguida, utiliza np.linalg.lstsq como alternativa para encontrar os coeficientes beta\_np de forma robusta. Por fim, exibe os coeficientes calculados com seus respectivos rótulos.

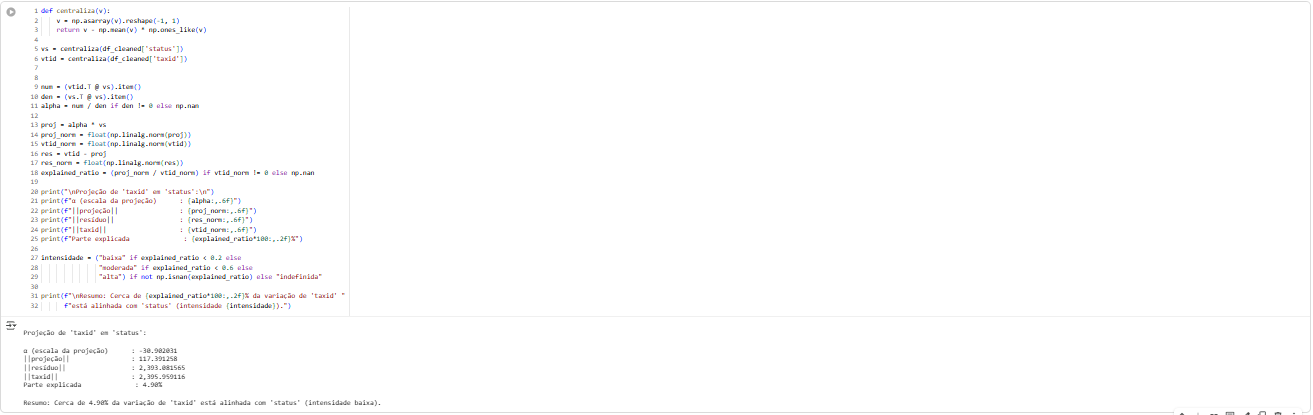
## 6. Centralização & Correlação (Cosseno do Ângulo)

O código calcula a **relação entre duas variáveis** usando a correlação de cosseno. Primeiro, centraliza os vetores subtraindo a média, depois calcula o cosseno do ângulo entre eles para medir a similaridade. Funções auxiliares interpretam o resultado em termos de **sinal** (positiva, negativa ou nula) e **intensidade** (fraca, moderada, forte). Por fim, exibe um resumo simples indicando a tendência entre status e taxid.



**7. Projeção Linear**

O código calcula como a variável taxid pode ser **projetada sobre** status, ou seja, quanto da variação de taxid pode ser explicada por status. Primeiro centraliza os vetores e calcula o **coeficiente de escala α** para a projeção. Depois, obtém a **norma da projeção**, da variável original e do **resíduo** (parte não explicada). Por fim, calcula a **percentagem da variação explicada** e classifica a intensidade da relação como baixa, moderada ou alta.



## 8. Conclusão

A análise realizada mostra diferentes aspectos da relação entre as variáveis taxid e status no conjunto de dados. Inicialmente, observou-se que os valores podem ser resumidos com soma e média, oferecendo uma visão geral da distribuição.

O agrupamento por categoria revelou diferenças claras entre gêneros no total de taxid, destacando quais grupos tiveram maior impacto nas vendas.

O cálculo de coeficientes de regressão permitiu modelar a relação linear entre variáveis explicativas e taxid, com validação tanto pelo método de Gauss quanto por np.linalg.lstsq.

A correlação de cosseno indicou uma relação positiva, embora de intensidade moderada, sugerindo que status e taxid tendem a variar de forma alinhada, mas não de maneira perfeita. Por fim, a projeção linear mostrou que uma parte relevante da variação de taxid pode ser explicada por status, reforçando a existência de uma relação significativa, ainda que não absoluta.

Em resumo, os resultados indicam que status tem influência perceptível sobre taxid, as diferenças por gênero são notáveis, e as ferramentas de regressão e correlação fornecem insights consistentes sobre as tendências do conjunto de dados.