**Projeto A3 para a UC de Inteligência Artificial.**

**Variáveis/Colunas**

* Price - preço dos diamante | float
* Carat - peso em quilates dos diamantes | float
* cut - qualidade do corte (Regular, Bom, Muito Bom, Premium, Ideal) | String
* color - cor do diamante, de J (pior) a D (melhor) | String
* clarity - clareza do diamante uma medida de quão claro é o diamante (I1 (pior), SI2, SI1, VS2, VS1, VVS2, VVS1, IF (melhor)) | String
* x - comprimento em mm | float
* y - largura em mm | float
* z - profundidade em mm | float
* depth - profundidade total em porcentagem | float
* table - do topo do diamante em relação ao ponto mais largo | float

Escolhemos o preço do diamante como o principal foco de estudo porque entender o preço nos ajuda a descobrir o que torna um diamante caro ou barato. Isso é útil para pessoas que fazem ou vendem joias, para aquelas que querem investir em diamantes e até mesmo para quem gosta de estudar essas pedras preciosas. Através disso, podemos aprender o que as pessoas valorizam em um diamante e como isso afeta os negócios e os preços no mercado.

**Explicação variáveis**

As variáveis que precisarão ser transformadas são "carat" (quilates), devido às diferentes escalas de valores, exigindo padronização para garantir que todas tenham o mesmo peso ao serem inseridas na etapa de machine learning. Além disso, também devemos transformar as variáveis "cut", “color” e “clarity”, por serem Strings, considerando que a maioria dos algoritmos apenas aceita valores numéricos para as colunas. Para isso, é necessário aplicar a codificação adequada para garantir sua utilização nos algoritmos de machine learning. Nesse caso, nossa variável-alvo é o preço do diamante, que é o indicador principal para avaliar um diamante.

Escolhemos esta base de dados devido à sua importância econômica e comercial. Compreendendo os fatores que influenciam o preço do diamante nos segmentos da indústria, consumo e investimentos, podemos abordar a falta de especificações no comércio de diamantes, que não atende a uma ampla gama de público, incluindo a classe média e a classe média baixa. Os dados nos permitirão explorar essa curiosidade em relação a esse nicho de mercado.