Para o nosso segundo projeto dentro do curso vamos focar na criação de diferentes visualizações que apresentam o relacionamento entre os dados. Vamos utilizar os recursos que aprendemos até aqui para responder aos questionamentos relativos ao [**IDEB**](https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/ideb) **(Índice de Desenvolvimento da Educação Básica)**.

Teremos acesso a um conjunto de dados que compila os indicadores do **IDEB**, divididos por região, que é fruto das **notas das provas** do [SAEB](https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/saeb) (Sistema de Avaliação da Educação Básica) realizadas entre 2005 a 2021 e as **taxas de aprovação** divulgadas pelo [Censo Escolar](https://www.gov.br/inep/pt-br) do mesmo período.

Para compreender o storytelling do projeto, esses dados descrevem o desempenho dos(as) estudantes em testes de proficiência de Língua Portuguesa e Matemática, além da taxa de aprovação escolar. O IDEB é medido em uma escala de 0 a 10 e é uma importante ferramenta para acompanhar o desenvolvimento da educação básica no país, permitindo a avaliação da efetividade das políticas públicas na área e a identificação das regiões que necessitam de maior atenção.

Este projeto nos seguirá nesta aula e, para que possamos trabalhar com os dados, precisamos compreender as informações que temos acesso. Sendo assim, vamos verificar os dados que possuímos em cada coluna do dataset ideb\_reg\_2005\_2021.csv:

* **ano**: ano de representação dos dados
* **regiao**: região onde a prova foi realizada
* **rede**: rede onde os dados foram compilados (estadual ou privada)
* **ensino**: etapa de ensino dos aplicantes (fundamental ou medio)
* **anos\_escolares**: fase do ensino dos aplicantes (EFAI - Ensino Fundamental Anos Iniciais , EFAF - Ensino Fundamental Anos Finais, EM - Ensino Médio)
* **taxa\_aprovacao**: taxa de aprovação dos aplicantes da região na fase
* **indicador\_rendimento**: indicador de rendimento baseado na taxa de aprovação dos aplicantes
* **nota\_saeb\_matematica**: nota de proficiência em matemática não padronizada (0-500) no exame do SAEB ou Prova Brasil
* **nota\_saeb\_lingua\_portuguesa**: nota de proficiência em língua portuguesa não padronizada (0-500) no exame do SAEB ou Prova Brasil
* **nota\_saeb\_media\_padronizada**: média da proficiência em língua portuguesa e matemática, padronizada para um indicador entre 0 e 10
* **ideb**: índice de desenvolvimento da educação básica resultante do produto entre a nota média padronizada (N) e o indicador de rendimento (P)

Os questionamentos que vamos responder durante a exploração dos dados são os seguintes:

1. Qual a relação entre o IDEB e o indicador de rendimento das escolas no ensino médio? É positiva, negativa ou não estão correlacionadas?
2. Qual a relação entre as notas das disciplinas de língua portuguesa e matemática no SAEB por região no Ensino Fundamental Anos Iniciais? Conseguimos traçar algum paralelo entre elas?
3. Qual seria a relação entre a taxa de aprovação e a média da nota do SAEB no Ensino Fundamental Anos Finais? Conseguimos notar como as duas se distribuem em relação ao IDEB e as regiões?

Essas questões serão nossas principais formas de conduzir nosso curso daqui em diante. Iremos responder aos poucos a cada uma delas. Conto contigo nessa caminhada! Bons estudos e até o próximo vídeo.