**Título do Projeto: Análise dos Determinantes do Desempenho Acadêmico em Escolas Públicas**

Problem Statement: Identificar e quantificar os principais fatores que influenciam o desempenho acadêmico dos estudantes em escolas públicas, visando auxiliar gestores educacionais e formuladores de políticas públicas a desenvolverem estratégias mais eficazes para melhorar a qualidade da educação.

Objetivo: O objetivo principal é inferir as relações entre diversos fatores socioeconômicos, escolares e familiares e o desempenho acadêmico dos estudantes, determinando quais variáveis têm maior impacto nesse desempenho.

Base de Dados: Para este projeto, utilizaremos o conjunto de dados do SAEB (Sistema de Avaliação da Educação Básica) disponibilizado pelo INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira). Este dataset contém informações detalhadas sobre o desempenho dos estudantes e características das escolas, professores e alunos no Brasil.

Link para o dataset: <https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/saeb>

Variáveis principais:

* Variável dependente: Nota média dos alunos em Matemática e Língua Portuguesa
* Variáveis independentes potenciais:
  + Nível socioeconômico da escola
  + Formação dos professores
  + Infraestrutura da escola
  + Tamanho da turma
  + Localização da escola (urbana/rural)
  + Índice de complexidade de gestão da escola
  + Regularidade do corpo docente
  + Nível socioeconômico dos alunos
  + Escolaridade dos pais
  + Horas-aula diárias
  + Participação dos pais na vida escolar

Etapas do Projeto:

1. Coleta e Preparação dos Dados:
   * Download e importação dos microdados do SAEB
   * Limpeza e tratamento de dados ausentes ou inconsistentes
   * Agregação dos dados em nível de escola
2. Análise Exploratória de Dados (EDA):
   * Visualização da distribuição das notas médias
   * Análise de correlações entre as variáveis
   * Identificação de padrões e tendências nos dados
3. Pré-processamento:
   * Normalização ou padronização das variáveis numéricas
   * Codificação de variáveis categóricas
   * Tratamento de outliers, se necessário
4. Modelagem:
   * Construção do modelo de regressão linear múltipla
   * Verificação dos pressupostos do modelo
   * Análise inicial dos coeficientes e sua significância estatística
5. Seleção de Variáveis:
   * Utilização de métodos como Stepwise, Lasso, ou Ridge para selecionar as variáveis mais relevantes
   * Comparação de modelos com diferentes conjuntos de variáveis
6. Interpretação dos Resultados:
   * Análise do R² e R² ajustado
   * Interpretação dos coeficientes em termos de impacto no desempenho acadêmico
   * Testes de hipóteses para a significância dos coeficientes
7. Validação do Modelo:
   * Análise de resíduos
   * Testes para multicolinearidade
   * Validação cruzada para avaliar a estabilidade do modelo
8. Análise de Interações:
   * Investigação de possíveis interações entre variáveis (por exemplo, nível socioeconômico e infraestrutura da escola)
   * Inclusão de termos de interação significativos no modelo
9. Conclusões e Recomendações:
   * Resumo dos fatores mais influentes no desempenho acadêmico
   * Discussão sobre as implicações para políticas educacionais
   * Sugestões para estratégias de melhoria do desempenho escolar

Aplicação de Negócio: Este projeto tem aplicações práticas para diversos setores:

1. Secretarias de Educação: Podem utilizar os insights para desenvolver políticas educacionais mais eficazes e direcionadas.
2. Gestores Escolares: Podem focar em melhorias nos fatores mais impactantes para o desempenho dos alunos.
3. Formuladores de Políticas Públicas: Podem criar programas e alocar recursos de forma mais eficiente para melhorar a qualidade da educação.
4. ONGs e Fundações Educacionais: Podem direcionar seus esforços e investimentos para as áreas de maior impacto.
5. Pesquisadores em Educação: Podem utilizar os resultados como base para estudos mais aprofundados sobre fatores específicos.

Ao focar no aspecto inferencial, este projeto permitirá não apenas prever o desempenho acadêmico, mas entender profundamente quais fatores o influenciam e em que magnitude. Isso fornecerá insights valiosos para decisões estratégicas e operacionais no setor educacional, possibilitando intervenções mais eficazes para melhorar a qualidade da educação pública.