**Análise do Impacto de Fatores Socioeconômicos na Expectativa de Vida**

Problem Statement:

Compreender quais fatores socioeconômicos têm maior influência na expectativa de vida de diferentes países, permitindo que formuladores de políticas públicas e organizações de saúde global focalizem seus esforços nos aspectos mais impactantes para melhorar a qualidade de vida da população.

Objetivo:

O objetivo principal deste projeto é inferir as relações entre diversos indicadores socioeconômicos e a expectativa de vida, identificando quais fatores têm maior peso nessa métrica de saúde pública.

Base de Dados:

Para este projeto, utilizaremos o conjunto de dados "Life Expectancy (WHO)" disponível publicamente no Kaggle. Este dataset contém informações de 193 países, coletadas pela Organização Mundial da Saúde (OMS) e pelas Nações Unidas de 2000 a 2015.

Link para o dataset: https://www.kaggle.com/kumarajarshi/life-expectancy-who

Variáveis principais:

- Variável dependente: Expectativa de vida

- Variáveis independentes potenciais:

- PIB per capita

- Escolaridade

- Percentual de gastos com saúde do PIB

- Taxa de mortalidade adulta

- Taxa de mortalidade infantil

- Índice de HIV

- Índice de obesidade em adultos

Etapas do Projeto:

1. Coleta e Preparação dos Dados:

- Download do dataset

- Limpeza e tratamento de dados ausentes

- Análise exploratória inicial

2. Análise Exploratória de Dados (EDA):

- Visualização da distribuição das variáveis

- Análise de correlações entre as variáveis

- Identificação de outliers e padrões interessantes

3. Pré-processamento:

- Normalização ou padronização das variáveis, se necessário

- Codificação de variáveis categóricas, se aplicável

- Divisão dos dados em conjuntos de treinamento e teste

4. Modelagem:

- Construção do modelo de regressão linear múltipla

- Avaliação dos pressupostos do modelo (linearidade, homocedasticidade, independência e normalidade dos resíduos)

- Análise dos coeficientes e sua significância estatística

5. Interpretação dos Resultados:

- Análise do R² e R² ajustado

- Interpretação dos coeficientes em termos de impacto na expectativa de vida

- Testes de hipóteses para a significância dos coeficientes

6. Validação do Modelo:

- Análise de resíduos

- Testes para multicolinearidade (VIF - Variance Inflation Factor)

- Validação cruzada, se aplicável

7. Conclusões e Recomendações:

- Resumo dos fatores mais influentes na expectativa de vida

- Discussão sobre as implicações para políticas públicas e saúde global

- Sugestões para futuras pesquisas ou análises

Aplicação de Negócio:

Este projeto tem aplicações claras para organizações internacionais de saúde, governos e ONGs que trabalham com desenvolvimento humano e saúde pública. Os resultados podem guiar:

1. Alocação de recursos em programas de saúde pública

2. Formulação de políticas focadas nos fatores mais impactantes para a expectativa de vida

3. Definição de metas e indicadores para programas de desenvolvimento internacional

4. Priorização de investimentos em áreas como educação, saúde e desenvolvimento econômico

Ao focar no aspecto inferencial, este projeto permitirá não apenas identificar correlações, mas também entender a magnitude e a direção do impacto de cada fator na expectativa de vida, fornecendo insights valiosos para tomadores de decisão em saúde pública e desenvolvimento humano.

Gostaria que eu elaborasse mais algum aspecto específico deste projeto ou que iniciássemos a implementação de alguma etapa?