



## **PYTHON PARA CIÊNCIAS DE DADOS**

### **PROF.: Leandro Lessa**

### **Objetivos de Ensino**

Exercitar os seguintes conceitos trabalhados na disciplina:

1. Coleta de dados.
2. Analisar e realizar tratamento de dados.
3. Analisar resultados obtidos.
4. Conhecimento teórico ministrado nas aulas.

### **Enunciado**

Análise do Desenvolvimento Humano e Econômico Global com Base no Dataset Gapminder

Este estudo tem como objetivo analisar o desenvolvimento humano e econômico de diferentes países ao longo das décadas, utilizando a base de dados Gapminder. A base contém informações consolidadas sobre indicadores essenciais como expectativa de vida, produto interno bruto per capita (PIB per capita) e população, abrangendo diversos países e continentes ao longo dos anos.

A análise desta base de dados é fundamental para:

- Compreender as tendências globais de desenvolvimento humano: Observar como a expectativa de vida e o PIB per capita evoluíram ao longo do tempo em diferentes partes do mundo.

- Identificar padrões regionais e históricos: Analisar as disparidades entre continentes e países, bem como a evolução desses indicadores ao longo de cinco décadas.
- Explorar relações entre indicadores socioeconômicos: Investigar correlações entre riqueza média (PIB per capita), expectativa de vida e tamanho populacional, buscando entender como esses fatores se influenciam mutuamente.
- Visualizar o progresso e as desigualdades globais: Utilizar gráficos e mapas para destacar avanços significativos e também persistências de desigualdades sociais e econômicas.

Este trabalho se propõe, portanto, a realizar uma Análise Exploratória de Dados (EDA) que auxilie na compreensão dos fatores que moldam o bem-estar das populações ao redor do mundo, com base em evidências numéricas e visuais extraídas do dataset Gapminder.

## Descrição da Base de Dados

O conjunto de dados Gapminder é uma coleção rica de indicadores socioeconômicos e de saúde de diversos países ao redor do mundo, ao longo de várias décadas. Esse dataset foi desenvolvido a partir do trabalho da Fundação Gapminder, uma organização sem fins lucrativos criada pelo médico e estatístico sueco Hans Rosling, com o objetivo de promover uma visão baseada em dados sobre o desenvolvimento global.

O dataset apresenta dados como:

- Expectativa de vida (lifeExp)
- Produto Interno Bruto per capita (gdpPercap)
- População total (pop)
- Ano de referência (year)
- Continente (continent)
- País (country)

Essas informações são coletadas em intervalos de cinco anos, desde 1952 até 2007, abrangendo mais de 140 países.

### lifeExp – Expectativa de Vida

Definição: Representa a expectativa média de vida ao nascer, ou seja, quantos anos uma pessoa nascida em determinado país e ano pode esperar viver.

Unidade: Anos (pode ter casas decimais porque é uma média estatística da população).

Exemplo: Se  $\text{lifeExp} = 31.997$ , significa que, em média, as pessoas viviam aproximadamente 32 anos naquele país e período. Isso pode refletir uma realidade de:

- pobreza extrema
- conflitos
- doenças ou ausência de acesso a serviços de saúde.

### gdpPercap – PIB per Capita

É o Produto Interno Bruto (PIB) per capita, ou seja, a riqueza média gerada por pessoa em um país.

Unidade: Dólares internacionais (ajustado pela paridade do poder de compra - PPC).

Exemplo: Se  $\text{gdpPercap} = 14560.2$ , isso quer dizer que, em média, cada pessoa naquele país “representa” uma produção anual de aproximadamente US\$ 14.560,20.

## ATENÇÃO PARA TRATAMENTO DE DADOS

Avaliem se será necessário realizar tratamento de dados ausentes no dataset disponibilizado.

### Instruções para correção de dados ausentes

- Exclua todos os dados ausentes
- Elimine dados duplicados

## Análise do Desenvolvimento Humano e Econômico Global com Base no Dataset Gapminder

### Objetivo da Atividade

Esta atividade tem como objetivo realizar no **mínimo 15 análises de dados**, utilizando visualizações, tabelas e representações gráficas para investigar o desenvolvimento humano e econômico ao redor do mundo. As análises devem ser feitas com base no dataset Gapminder, que contempla dados históricos de:

- Expectativa de vida
- PIB per capita
- População
- Continente e país

O objetivo é obter uma visão abrangente sobre padrões globais de crescimento, desigualdade, evolução populacional e qualidade de vida.

## Entregável

A entrega deverá ser feita por meio de um Jupyter Notebook (.ipynb), contendo:

- Código-fonte documentado
- Visualizações com interpretação
- Anotações em Markdown explicando decisões e insights

## Estrutura Recomendada do Notebook

### **1. Título e Introdução**

- Apresente o tema do projeto
- Descreva brevemente os objetivos da análise
- Explique o contexto da base de dados Gapminder
- Liste as etapas que serão realizadas no notebook

### **2. Coleta, Limpeza e Análise Exploratória (AED)**

Objetivo: Preparar os dados, tratando inconsistências e identificando padrões iniciais.

- Markdown: Documentar cada passo, justificativas e observações.
- Código: Carregamento, limpeza e transformações.
- Visualizações: Gráficos e tabelas com insights exploratórios.

### **3. Funções de Visualização e Análise Gráfica**

Objetivo: Criar funções reutilizáveis para gerar gráficos e apoiar a análise.

- Markdown: Explicações sobre as funções e os tipos de gráficos utilizados.
- Código: Funções dinâmicas com parâmetros flexíveis.
- Exemplos: Geração de gráficos específicos e interpretação dos resultados.

#### **4. Interpretação das Visualizações**

Objetivo: Interpretar criticamente os dados com base nos gráficos e tabelas.

- Markdown: Explicação de cada visualização e justificativa da sua escolha.
- Código: Análises adicionais, se necessário (cálculos, agrupamentos, etc.).

#### **5. Síntese dos Principais Insights**

Objetivo: Resumir os principais achados da análise.

- Resumo Markdown: Destacar padrões, tendências, anomalias e correlações encontradas.
- Aplicações Práticas: Possíveis implicações para políticas públicas, desenvolvimento social e econômico, etc.

#### **6. Conclusão e Reflexão**

Objetivo: Refletir sobre o processo de análise e o aprendizado obtido.

- Quais foram os maiores desafios?
- O que foi aprendido?
- Como essa experiência ajudou no desenvolvimento de habilidades analíticas?

#### **Atenção à Estrutura Recomendada**

É fundamental que o aluno siga a estrutura sugerida do notebook, abordando todos os 6 itens descritos na seção de entrega.

O não cumprimento dessas etapas poderá resultar em penalizações na nota final. Cada parte da estrutura foi pensada para garantir um trabalho completo, organizado e com maior aproveitamento analítico.

Para alcançar uma avaliação positiva, assegure-se de seguir as recomendações e desenvolver cada seção de forma clara, coerente e bem documentada.

## **BÔNUS – Itens Opcionais que Valem Pontos Extras**

### **Modularização de Código em Python**

- Recomendamos a utilização de modularização (funções bem definidas, código limpo e organizado).
- Embora não seja obrigatório, será avaliado com prestígio e poderá aumentar sua nota final.

### **Integração com Outras Bases de Dados**

Alunos que incluírem outras fontes de dados relevantes para enriquecer a análise receberão bônus na nota final.

- A base complementar deve ser:
- Indicada claramente no notebook (nome, link ou referência)
- Compartilhada junto com o trabalho
- Utilizada de forma inteligente e relevante, gerando insights adicionais coerentes com o estudo

Exemplo: Dados sobre educação, meio ambiente, IDH, ou políticas públicas que complementem a leitura dos dados do Gapminder.

## Penalizações

### Proibição do Uso de IA Generativa

- O uso de IA para geração automática de código ou análise é estritamente proibido.
- O trabalho deve ser autoral, refletindo o entendimento e raciocínio analítico do aluno.
- Detecção de uso de IA poderá resultar em penalização parcial ou total da nota.

### Ferramentas Sugeridas

- Pandas: Manipulação e análise de dados
- Matplotlib / Seaborn: Visualizações estáticas e interativas
- NumPy: Cálculos estatísticos
- Markdown: Estruturação textual e explicativa

### Dicas para a Documentação

- Utilize cabeçalhos e listas para organizar o conteúdo
- Mantenha explicações claras e concisas
- Comente todo o código de forma objetiva, especialmente trechos menos intuitivos

Boa análise e mãos à obra!