

PROJETO 1

ANÁLISE EXPLORATÓRIA DE DADOS

GAPMINDER WORLD

Vocês fizeram uma exploração guiada nas aulas 1 e 2 usando dados de um site chamado Gapminder (<https://www.gapminder.org/tools/>). Este site coleta dados de diversas organizações como OECD, World Bank, UN, etc. Além de conseguir visualizar um gráfico interativo no website, é possível realizar o download dos dados.

Para cada indicador oferecido é possível baixar um arquivo que contém a informação daquele país ao longo dos anos em que há dados disponíveis

Neste projeto você deverá realizar uma análise exploratória sobre os dados disponíveis no Gapminder. Para tanto você deverá acessar o Gapminder para verificar quais indicadores estão disponíveis.

O tema central do projeto é formular uma pergunta sobre o IDH (Índice de Desenvolvimento Humano – HDI em inglês). O IDH é calculado usando 3 fatores:

- Expectativa de Vida
- Anos médios de estudo e anos esperados de escolaridade
- Renda (GDP ou PIB) per capita

Exemplos de pergunta:

- Além das 3 variáveis que entram no cálculo do IDH, qual outra (sem ser derivada direta das 3) tem mais correlação com o IDH nos países?
- Um certo país A melhorou seu IDH bem mais que o país B. Que variável está diferente nos dois instantes de tempo e poderia estar associada a esta melhora?
- Nas redes sociais, frequentemente se propaga que um IDH está fortemente associado com liberdade econômica. Será que é mesmo?

Para descobrir quais variáveis estão disponíveis recomendamos usar a ferramenta do Gapminder ([https://www.gapminder.org/tools/#\\$chart-type=bubbles](https://www.gapminder.org/tools/#$chart-type=bubbles)). Também recomendamos ver os vários vídeos disponíveis na página do Gapminder.

Leia mais sobre IDH em:

https://pt.wikipedia.org/wiki/%C3%8Dndice_de_Desenvolvimento_Humano

Tarefas:

Você deve acessar o site do Gapminder e verificar quais indicadores estão disponíveis no banco de dados. Deverá então escolher 3 indicadores (HDI e mais dois a sua escolha) e formular uma pergunta que deve ser respondida através da análise exploratória.

Devido à formulação do IDH, você não pode escolher nenhum dos indicadores que compõem diretamente o seu valor. Neste caso as correlações seriam bastante óbvias.

Você vai precisar escolher um ano base para sua investigação. Cada variável de interesse estará em um arquivo, portanto precisará baixar em <https://www.gapminder.org/data/> os dados do HDI e dos dois outros indicadores que escolheu. Escolha situar sua análise em anos em que os dados que escolheu estejam disponíveis para uma boa quantidade de países.

Após separar o ano escolhido, você deve então juntar as 3 variáveis em um DataFrame e salvar um arquivo que contém os dados do ano escolhido.

Inicie a sua análise aplicando o ferramental aprendido em aula sempre tentando coletar informações que lhe ajudarão a responder a pergunta formulada. Não se esqueça que nem toda ferramenta é adequada para todos os tipos de variáveis.

Conclua escrevendo detalhadamente o que os dados estão lhe explicando.

Caso tenha sucesso na análise exploratória, você pode então verificar como a sua conclusão evolui ao longo do tempo, repetindo a análise, porém utilizando outros anos como base.

REGRAS:

1. O Projeto 1 é estritamente INDIVIDUAL.
2. Elaborar um relatório que contenha as análises que contribuíram para as conclusões de acordo com seu objetivo inicial. Use o **arquivo layout** disponibilizado na pasta Projeto1 do github:
3. O relatório deve ser feito em IPython Notebook (com células de texto formatadas em Markdown) e deve conter, pelo menos, as seguintes seções:
 - a. Introdução,
 - b. Análises e explicação dos resultados,
 - c. Conclusão.

A estrutura do documento deve ser clara e de fácil compreensão da linha de raciocínio. Nesse caso, o notebook não deve haver excesso de impressões não discutidas de variáveis e de *dataframe*.

Aconselhamos fazer uma análise geral e, após finalizada, salve com outro nome, limpe seu IPython Notebook apenas com os resultados relevantes e melhore seu texto.

4. Seu projeto deve ser adicionado no seu github dentro de uma pasta chamada Projeto1.
5. O arquivo DEVE ser com extensão .ipynb

CRONOGRAMA:

Turmas A, B e C

DATA	Finalização:
15/08	APS1: Entrega parcial contendo a leitura e junção dos dados, limpeza da base e pergunta formulada. Coloque seu nome no Jupyter Notebook para facilitar a organização. Via Blackboard.
23/08	Fazer git push em seu Github até 23:59 do Projeto 1 finalizado.

RUBRICS DE AVALIAÇÃO DO OBJETIVO DE APRENDIZADO

Objetivo de aprendizado	Insatisfatório (I)	Em desenvolvimento (D)	Essencial (C)	Proficiente (B)	Avançado (A)
Elaborar análises exploratórias de dados (univariadas e multivariadas), utilizando ferramentas estatísticas e computacionais adequadas.	<p>Não consegue trabalhar com bases de dados de forma proficiente.</p> <p>Apresenta problemas com arquivos, formatos de arquivos ou não tem habilidades básicas de filtragem de dados.</p>	<p>Consegue abrir e filtrar dados de forma básica, mas não consegue conectar os dados à prática de análise.</p> <p>O notebook é praticamente uma coleção de comandos do pandas e gráficos, com pouca análise e interpretação.</p> <p>Tem dificuldade para lidar com valores inválidos e não consegue eliminá-los, corrompendo a análise.</p>	<p>Descreve com riqueza de detalhes o objetivo que deseja avaliar nesse projeto de acordo com a vertente escolhida.</p> <p>Constrói uma base de dados com variáveis pertinentes ao objetivo definido, sejam essas do tipo qualitativo ou quantitativo.</p> <p>É capaz de filtrá-las, cruzá-las e processá-las para embasar sua análise.</p>	<p>Executa os comportamentos da rubrica C de maneira excepcional.</p> <p>Analisa ocorrências conjuntas com recursos adequados ao tipo de variável (crosstab, gráficos de barras, de setor, histograma e/ou boxplot), respeitando o tipo de variável.</p> <p>Interpreta corretamente o que as ocorrências conjuntas significam em função dos objetivos da pesquisa.</p> <p>Comunica os resultados com bastante clareza.</p>	<p>Executa os comportamentos da rubrica B de maneira excepcional.</p> <p>Incorpora aspecto histórico, demonstrando a evolução da análise temporalmente.</p> <p>Expõe os resultados obtidos de maneira excepcional.</p>