

## 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 deste ano é explorar o pertencimento a organizações internacionais (OCDE, OTAN, MERCOSUL, etc). Uma análise para ser considerada completa deverá incluir 2 eixos:

- Apresentar as diferenças entre países que estão nestas associações versus países que não estão
- Encontrar diferenças de evolução de indicadores no tempo entre países que estão versus países que estão fora.

É interessante procurar organizações que deixem países comparáveis de fora, ou para as quais exista registro de países que passaram a fazer parte da organização no período coberto pelo Gapminder.

Exemplos de pergunta:

- Existe aumento da evolução do IDH de países que passam a fazer parte da OCDE?
- Países que entram para a OTAN passam a ter aumento ou diminuição do gasto militar?

- Um país X que entrou para a OCDE passou a melhorar sua posição relativa no indicador de PIB per capita em relação a países comparáveis que não faziam parte da OCDE?

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 organizações internacionais em:

[https://en.wikipedia.org/wiki/List\\_of\\_intergovernmental\\_organizations](https://en.wikipedia.org/wiki/List_of_intergovernmental_organizations)

## **Tarefas:**

Primeiro escolha a organização internacional que vai estudar. Recomendamos que tenha pelo menos duas opções iniciais para realizar testes e no final de definir pela que proporcionar as análises mais interessantes.

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

A seguir, uma vez escolhidos a organização internacional e as variáveis, formule suas hipóteses deixando claro o mecanismo.

Exemplos de hipóteses:

- Após se juntar à OTAN, um país passará a gastar menos com orçamento militar porque passa a contar com o apoio de aliados em caso de conflito, deixando de contar só com seus recursos.
- Após se juntar à OTAN, um país passa a gastar mais com orçamento militar pois passa a ser obrigado pela organização a se comprometer com investimentos mínimos em defesa.

- Países que se juntam ao MERCOSUL têm aumento de exportações porque há aumento dos mercados consumidores.
- Países que se juntam à União Europeia têm maior índice de industrialização pois recebem mais investimento estrangeiro.

Note que muitas vezes após separar o ano escolhido, você deve então juntar as variáveis em um DataFrame e salvar um arquivo que contém os dados dos anos escolhido.

Monte uma coluna que especifica se cada um dos países pertence ou não à organização de interesse.

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 explicando a natureza da organização,
  - b. Explicar a pergunta de pesquisa, esclarecendo a hipótese a ser testada
  - c. Análises e explicação dos resultados,
  - d. 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.
6. Você deve entrar com um ou mais notebooks “sujos” contendo tudo o que tentou fazer e a montagem das bases de dados, e uma versão mais limpa que já lê o arquivo resultado dos joins e apresenta a análise de forma clara.

## **CRONOGRAMA:**

### **Turmas A, B e C**

<b>DATA</b>	<b>Finalização:</b>
20/08	APS1: Entrega parcial contendo a leitura e junção dos dados, limpeza da base, pergunta formulada contendo hipótese e mecanismo. Coloque seu nome no Jupyter Notebook para facilitar a organização. Via Blackboard.
<b>29/08</b>	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)
<b>Elaborar análises exploratórias de dados (univariadas e multivariadas), utilizando ferramentas estatísticas e computacionais adequadas.</b>	<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>Explica a hipótese a ser avaliada, o mecanismo e a relaciona aos dados</p> <p>Constrói uma base de dados com variáveis pertinentes ao objetivo definido, sejam essas do tipo qualitativo ou quantitativo.</p> <p>Interpreta corretamente o que as ocorrências conjuntas significam em função dos objetivos da pesquisa.</p> <p>Analisa de forma básica variáveis de países que pertencem à organização versus os que não pertencem. Por exemplo gráficos de barra ou setor comparando países membros x não, ou plotado evolução temporal de indicadores de forma simples</p> <p>É capaz de filtrá-las, cruzá-las e processá-las para embasar sua análise.</p> <p>Interpreta os dados e fornece conclusão</p>	<p>Explica a hipótese com riqueza de detalhes, incluindo o provável mecanismo e a relação entre os dados. Cita referências (artigos científicos ou no mínimo notícias) que embasam a hipótese</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>Tem cuidados adicionais para comparar nações que pertencem à organização versus nações similares que não pertencem.</p> <p>Comunica os resultados com bastante clareza.</p>	<p>Executa os comportamentos da rubrica B de maneira excepcional.</p> <p>Usa indicadores relativos ou derivados: ex.: colocar valores em proporção do GDP, usar posição percentil de países para análises, evolução relativa do GDP, etc.</p> <p>Incorpora aspecto histórico, demonstrando a evolução da análise temporalmente.</p> <p>Expõe os resultados obtidos de maneira excepcional.</p>