Projeto Power BI – Análise Municípios, Estados e Regiões

O Projeto de Análise de Dados usando ferramenta Power BI deverá conter análises sobre a situação dos Municípios, Estados e Regiões brasileiras, tanto em termos populacionais quando em categorização segundo o índice IDH (Índice de Desenvolvimento Humano), em seus 3 tipos de Medição (Renda, Educação e Longevidade), além do índice médio (IDHM). **ATENÇÃO:** ao contrário do que pode parecer num primeiro momento, o IDHM não é a média simples dos outros 3 índices, por isso ele existe como um valor separado.

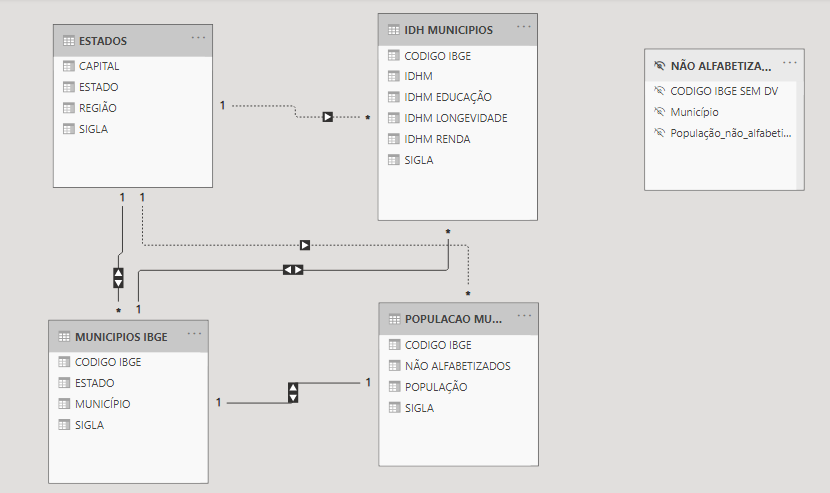
Os dados a serem usados estão disponíveis no arquivo ZIP, e são assim definidos:

* **MUNICIPIOS\_IBGE.xlsx** -> Contem a relação de Municípios brasileiros, com um código padronizado e o Estado ao qual o Município pertence.
* **IDH\_Municipios.xlsx** -> Contém os valores de IDH (Índice de Desenvolvimento Humano) de cada Município brasileiro, segundo dados coletados de relatório PNUD Brasil de 2010.
* **População\_Municipios.xlsx** -> Contém o número de habitantes em cada Município brasileiro.
* **Nao\_Alfabetizados\_MUNICIPIO.csv** -> Contém o número de não alfabetizados em cada Município brasileiro.
* **Estados** -> Os dados dos Estados e as respectivas regiões brasileiras deverão ser obtidos no site <https://www.estadosecapitaisdobrasil.com/>

Algumas observações sobre os dados:

* Como os dados são de origem distintas, pode acontecer de nem todos existirem em todos os lugares. Por exemplo, como a relação de Municípios é mais atual que a com os dados de IDH, pode acontecer de algum município não ter essa informação. O mesmo pode ocorrer com os dados de População e Não Alfabetizados. Nesses casos, a relação de MUNCIPIOS\_IBGE deve prevalecer, ficando “nulos” os dados faltantes para um determinado Município.
* Os dados devem ser “limpos” e “organizados” da forma que os alunos acharem melhor, porém algumas premissas deverão ser seguidas:
  + Na planilha com População\_Municipios.xlsx, os estados do Pará e Bahia aparecem como “Pará Pará” e “Bahia Bahia”, e devem ser ajustados, pois sem esse ajuste não é possível “juntar” os dados depois.
  + Os dados de População e Não Alfabetizados deverão, ao final da carga, ficarem na mesma tabela de dados.
* Independentemente da forma que os alunos fizerem o processo, ao final da “limpeza” / “organização” dos dados, as tabelas deverão estar como a imagem abaixo, inclusive com os relacionamentos ajustados para se comportarem como exemplificado na imagem.
* Após a carga de dados, os alunos poderão adicionar (e provavelmente deverão) novas colunas, conforme julgarem necessário. Os exemplos abaixo são apenas o resultado da “limpeza” / ”organização” dos dados.
* Também, caso seja necessário, os alunos poderão criar novas tabelas e/ou relacionamentos, bem como alterar os relacionamentos aqui abaixo descritos, desde que essa alteração seja para tornar possível criar uma medida ou análise.

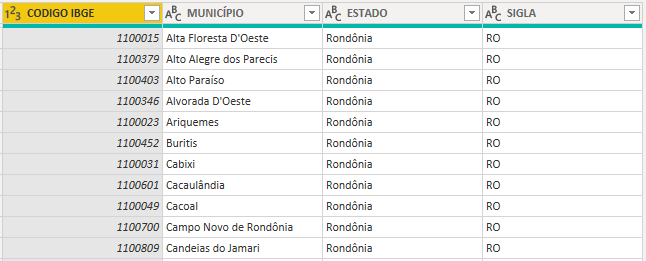
**TABELAS / COLUNAS e RELACIONAMENTOS após carga de dados:**



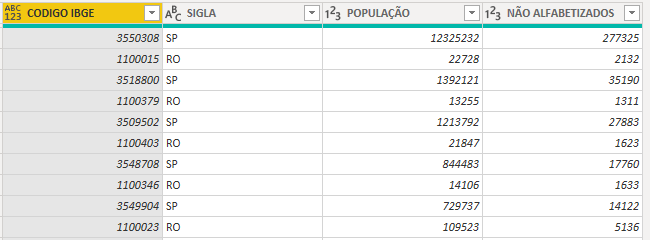
**Dados de ESTADO:**

****

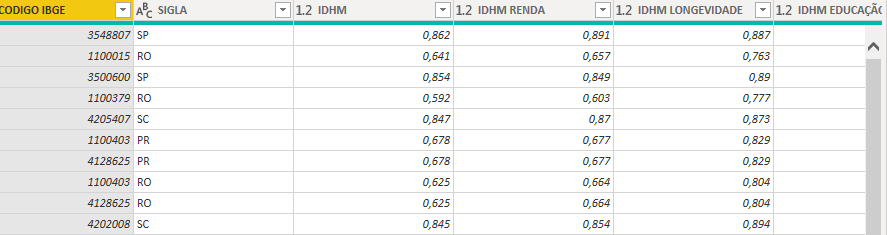
**Dados de MUNICIPIO IBGE:**

****

**Dados de POPULAÇÃO MUNICIPIOS:**

****

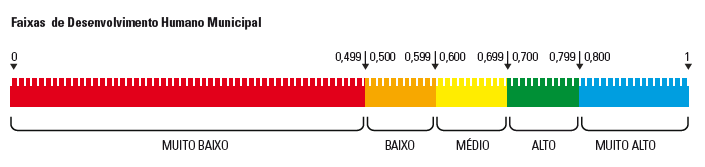
**Dados IDH MUNICIPIOS:**

****

**ANÁLISES**

Os grupos serão os responsáveis por definir quais serão as análises a serem demonstradas, quais serão as “medidas” necessárias para as análises escolhidas, bem como a melhor forma de apresentação de uma determinada análise. Porém, os seguintes itens deverão ser contemplados:

* Possuir Segmentação de dados por todas as Dimensões. Identificar as dimensões faz parte da definição.
* Ter ao menos uma visualização de dados em Mapa.
* Ter ao menos uma visualização de dados em gráfico de barras.
* Ter ao menos uma visualização de dados em gráfico de Pizza.
* Ter ao menos uma visualização de dados em Matriz. Não usar Matriz como uma tabela simples.
* Análise sobre População deverá conter dados sobre alfabetizados, mesmo esse dado não existindo diretamente nos dados carregados. Mas, por lógica, ele pode ser obtido facilmente.
* Ao menos uma análise de IDH precisa seguir o critério do relatório PNUD, onde:
  + **IDH MUITO BAIXO** -> valores entre 0,000 e 0,499
  + **IDH BAIXO** ->valores entre 0,500 e 0,599
  + **IDH MÉDIO** -> valores entre 0,600 e 0,699
  + **IDH ALTO** -> valores entre 0,700 e 0,799
  + **IDH MUITO ALTO** -> valores entre 0,800 e 1,000



**Avaliação**

* A avaliação do trabalho será baseada nos seguintes critérios:
* Assertividade na carga dos dados (limpeza e organização), conforme as especificações descritas acima.
* Quantidade e relevância das Medidas criadas. Quanto mais medidas, melhor.
* Quantidade e relevância das Análises cridas. Quanto mais análises, melhor.
* Criatividade na escolha das análises, bem como na escolha da visualização das análises.
* Os trabalhos deverão ser entregues até dia 23/11. No dia 27/11, a critério do professor, algum grupo poderá ser escolhido para apresentação aos demais ou, ainda, o professor poderá usar a aula para sanar quaisquer dúvidas sobre o trabalho, antes de fechar a avaliação final. Aula do dia 04/11 fica reservada para eventuais dúvidas e fechamento das notas.