# Projeto 1 – Conclusão

O Projeto 1 de Ciência dos Dados teve como objetivo fazer uma análise exploratória com os recursos, qualitativos e/ou quantitativos, vistos nas primeiras aulas dessa disciplina.

Para isso, primeiramente, foi escolhido além do IDH outras duas variáveis no Gapminder, Fertilidade e Consumo de Energia Total. Com as variáveis escolhidas foi feita a pergunta que o Projeto 1 tem como objetivo responde-la, que é “A Fertilidade junto ao Consumo de Energia Total estão associados ao IDH? Qual deles está mais fortemente associado?.

Para responder essa pergunta foi necessário abrir um documento no Jupyter Notebook e carregar as tabelas das variáveis acima. Para facilitar na programação e na organização dos dados foi mudado os índices, renomeados e juntadas as três tabelas em uma única. Após a junção percebeu-se que algumas linhas estavam sem dados, portanto tirou-se os espaços vazios, em seguida, salvou essa tabela em um arquivo CSV e Excel.

Feito isso, foi escolhido um ano para uma análise mais profunda, primeiramente foi o de 2008. O primeiro recurso para a análise foi um gráfico que engloba as 3 variáveis, sendo o eixo x a Fertilidade, o y o IDH, e o Consumo de Eletricidade ficou sendo a área das bolinhas demarcadas. Esse gráfico mostrou que as bolinhas maiores, ou seja, as que apresentavam maior Consumo de Energia estavam predominantes no países com o IDH mais alto, e nesses países observava também um menor número de Fertilidade. Visto isso, partiu-se para os gráficos com duas variáveis sendo tendo em vista analisar o IDH. O gráfico de Fertilidade e IDH teve uma reta de correlação muito boa, em que estava visível a forte associação entre eles. Já o de Consumo de Energia com o IDH não apresentou um gráfico bom, pois a reta não estava seguindo a maioria dos pontos, assim plotou-se um Boxplot para ver os outliers da reta. Esses outliers foram eliminados, e mesmo assim a reta não estava boa. Após essas análises gráficos foi solicitada uma tabela de correlações entre as variáveis, concluindo que a Fertilidade estava fortemente ligada com o IDH, o mesmo não foi visto com o Consumo de Energia, que apresentou pouca correlação com o índice. O mesmo foi feito para o ano de 2000, e chegou-se a mesma conclusão.

Por fim, como a correlação do IDH com a Fertilidade foi alta, criou-se um último gráfico em que visou a evolução dessa variável a esse índice ao longo dos anos 2005 a 2009, e percebeu-se que pouco mudou em relação a eles, ou seja, a associação entre essas variáveis sempre foi alta e estável.