

Análise de Dados do Enem 2021

Como trabalho final do curso Big Data Science da Semantix, fomos convidados a fazer algum tipo de análise dos dados no conjunto de arquivos fornecidos pelo INP e encontrados em <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enem> .

Este trabalho de conclusão visa reforçar o aprendizado. Em especial, esse é meu primeiro projeto em análise de dados e estou bastante satisfeito com a possibilidade de começar a trilhar esse caminho de análise das coisas que envolvem e impactam de alguma forma as nossas vidas.

O trabalho realizado

Após ter uma ideia inicial de que tipo de bibliotecas poderiam ser necessárias para o trabalho e importá-las, começamos o trabalho através da carga do arquivo `MICRODADOS_ENEM_2021.csv` . Aqui, tivemos logo o primeiro desafio pois a leitura do arquivo através `read_csv` apresentava erro. Com algumas pesquisas, acertamos o encoding do comando e prosseguimos tomando contato com o glossário de dados e realizando algumas manipulações para diminuir o tamanho do data frame. Com um data frame menor, realizamos algumas explorações iniciais. Nela verificamos algumas correlações entre as notas dos candidatos, utilizando o comando `corr` e visualizando o resultado num heat map, usando a biblioteca `seaborn`. Há uma boa correlação entre as notas de Linguagens e Ciências Humanas, onde o score foi de 0,8. Em seguida, notamos uma correlação de 0,7 entre as notas das demais provas, exceto a nota de redação em que houve um coeficiente de correlação de 0,5 com todas as notas das demais provas.

Em seguida, verificamos a distribuição dos candidatos com relação à cor da pele e o tipo de escola que estudou.

Por fim, estudamos a relação entre a nota média, coluna que adicionamos ao data frame e a renda familiar do candidato. As tabelas e o gráfico produzido com o `matplotlib` mostrou a tendência de que o candidato obteve uma nota melhor, quanto maior era a renda familiar declarada.

Conclusão

Muitos códigos foram tentados e em alguns casos, foram dados passos além do que talvez fosse necessário, no entanto, devido à pouca experiência, esses passos adicionais foram necessários. Com o passar do tempo e com treinamento, espero que possa realizar tarefas como essa em bem menos tempo e com menos linhas de código.