

Pyladies São Paulo Open Data Day - Analisando Dados do ENEM



Angélica Cristina Cruz Custódio
Thanuci Silva
Março/24

Um pouco sobre nós

Angélica



Uma pessoa que gosta de café, livros e de escutar/contar/viver boas histórias.

Gosto de aplicar a beleza da matemática no mundo real e ver o que acontece :)



Estudei estatística e computação. Trabalho como Cientista de Dados na [Pier](#) - seguradora digital e como Professora na [Pós-graduação em Data Science e Decisão](#) no Insper.



★ Um livro:

Persépolis - Marjane Satrapi

★ Uma comida:

Bolo de milho da minha mãe com cafézinho

Um pouco sobre nós

Thanuci

Generalista em todos os sentidos.

É Bióloga e Cientista de Dados. É pesquisadora e trabalha no mundo corporativo.

Escuta, Rap, Rock, Sertanejo e Funk. Gosta de todos.

Estudei Licenciatura em Ciências Biológicas.

Trabalho como Especialista de Ensino e Aprendizagem no Insper e sou professora de alguns programas de Data Science na mesma instituição.



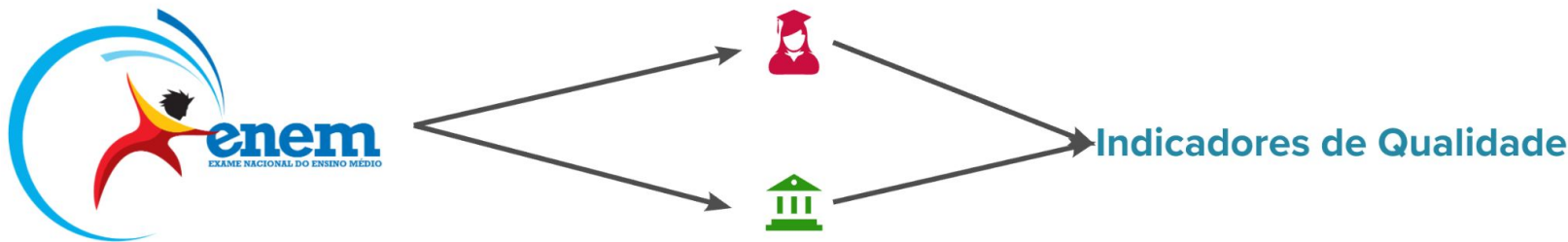
Um livro:

Indomável - Glennon Doyle

Uma comida: pão de queijo

Dados Enem

Importância dos dados educacionais públicos



Auxiliar a tomada de decisões em diversos contextos:

- Avalia a educação básica
- Suporta políticas públicas
- Programas de ensino (básico e superior)
- Projetos e ações educativas
- Impacta diretamente a educação superior por ser usado como nota de entrada em um grande número de Instituições de Ensino Superior (IES).

O Exame Nacional do Ensino Médio (ENEM)



Primeira aplicação: 1998

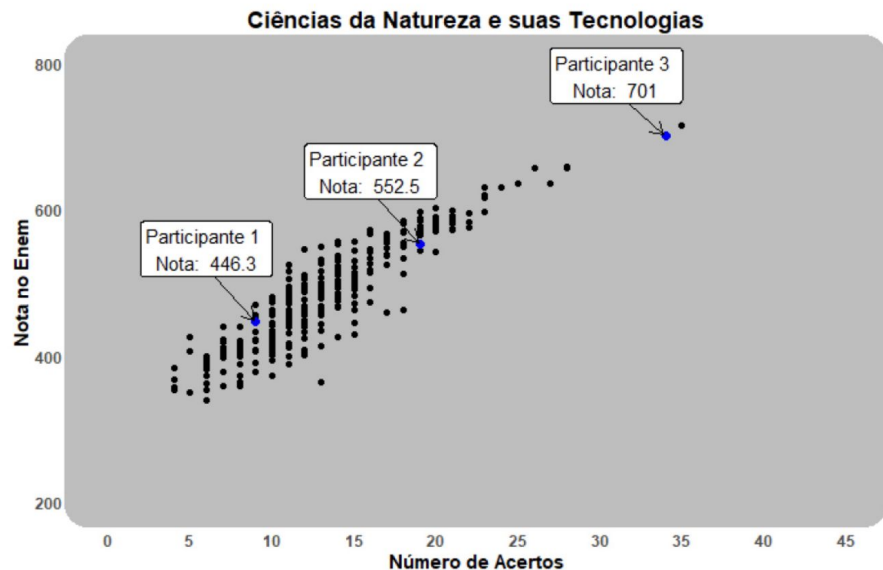
Composição da prova (180 questões):

- Ciências Humanas e suas Tecnologias (História, Geografia, Filosofia e Sociologia);
- Ciências da Natureza e suas Tecnologias (Química, Física e Biologia);
- Linguagens, Códigos e suas Tecnologias (Língua Portuguesa, Literatura, Língua Estrangeira - Inglês ou Espanhol, Artes, Educação Física e Tecnologias da Informação e Comunicação);
- Matemática e suas Tecnologias (Matemática).
- Redação

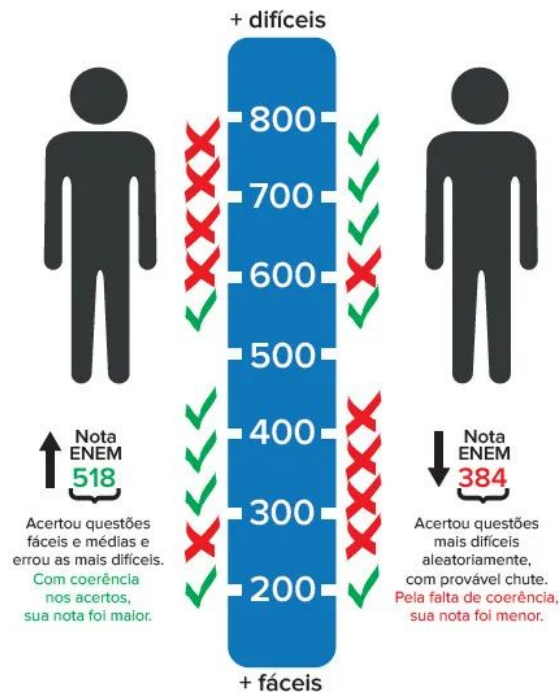
O Exame Nacional do Ensino Médio (ENEM)

Teoria de Resposta ao Item

[Guia do Participante - Entenda sua nota](#)



Exemplos de resultados calculados pela TRI



A base de dados

Acesso Microdados

```
[53]: dfMicroDados.head()
```

```
[53]:
```

	NU_INSCRICAO	NU_ANO	TP_FAIXA_ETARIA	TP_SEXO	TP_ESTADO_CIVIL	TP_COR_RACA	TP_NACIONALIDADE	TP_ST_CONCLUSAO	TP_ANO_CONCLUIU	TP_ESCOLA	...	Q016	Q017	Q01
0	210053865474	2021	5	F	1	1	1	1	3	1 ...	A	A		
1	210052384164	2021	12	M	1	1	1	1	11	1 ...	A	A		
2	210052589243	2021	13	F	3	1	1	1	15	1 ...	B	A		
3	210052128335	2021	3	M	1	3	1	2	0	2 ...	A	A		
4	210051353021	2021	2	F	1	3	1	2	0	2 ...	B	A		

5 rows × 76 columns

```
[54]: dfMicroDados.shape
```

```
[54]: (3389832, 76)
```

Arquivos utilizados:

- *MICRODADOS_ENEM_2021.csv*
- *Dicionário_Microdados_Enem_2021.xlsx*

Gerando valor a partir da base de dados

Qual é a pergunta que desejamos responder?

- O ano em que o aluno concluiu o ensino médio influencia na nota final?
- O grau de escolaridade da mãe do aluno influencia na nota final?
- Tipo de Escola que o aluno estudou?

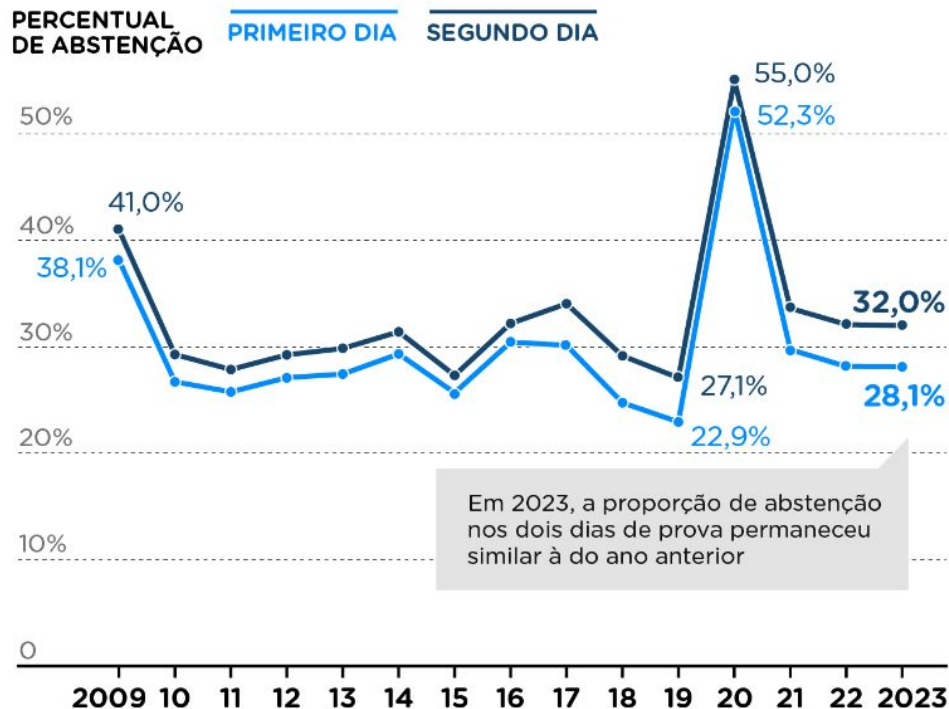


Aplicação Dados Enem

Atenção aos dados faltantes

Taxa de abstenção em cada dia de prova do Enem

% sobre o total de inscritos, de 2009 a 2023

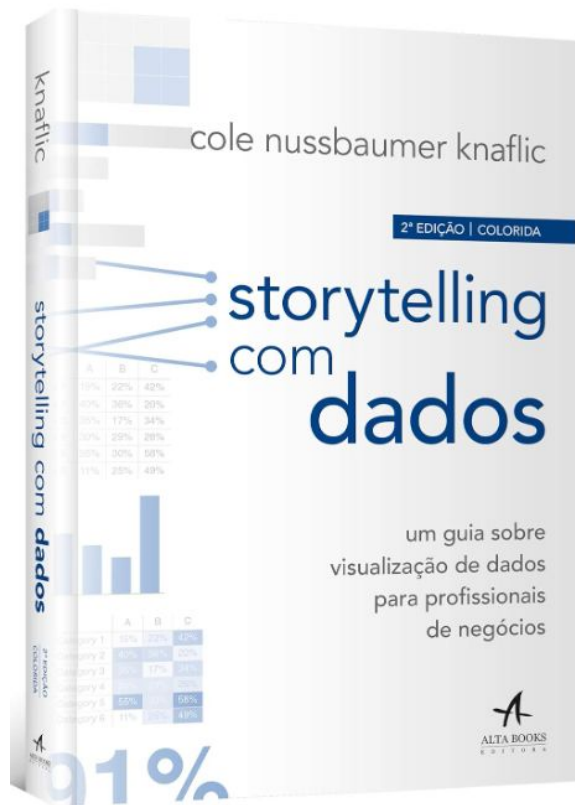


Já ouviu falar que a falta de informação pode ser uma informação?



Vamos para o python?

Referências



Desvendando a Arte de Contar Histórias com Dados

Livro sobre análise de dados e como contar histórias com as nossas análises!

Referências

- Dutra, J. F., Firmino Júnior, J. B. & Fernandes, D. Y. S. **Fatores que podem interferir no desempenho de estudantes no ENEM: uma revisão sistemática da literatura.** Revista Brasileira de Informática na Educação, 31, 323-351. DOI: 10.5753/rbie.2023.3087.
- Stearns, B., Rangel, F., Firmino, F., Rangel, F., & Oliveira, J. (2017). **Prevendo Desempenho dos Candidatos do ENEM Através de Dados Socioeconômicos.** In Anais do XXXVI Concurso de Trabalhos de Iniciação Científica da SBC. Porto Alegre: SBC.
- Silva, V. A. A. da, Moreno, L. L. O., Gonçalves, L. B., Soares, S. S. R. F., & Júnior, R. R. S. (2020). Identificação de Desigualdades Sociais a partir do desempenho dos alunos do Ensino Médio no ENEM 2019 utilizando Mineração de Dados. Anais do Simpósio Brasileiro de Informática na

Obrigada!

[Linkedin Angélica](#)

[Linkedin Thanuci](#)