

Microdados do ENADE e Censo Superior

Para cursos de graduação na área de Computação no Nordeste.

A análise deve ser feita a nível de **CURSO**.

O ENADE, por coletar dados individuais dos estudantes, dará o perfil detalhado, e o Censo Superior dará os dados de infraestrutura e características do curso/IES.

Observações gerais do ENADE (2021, 2017 e 2014).

- Dividir categorias administrativas entre Públicas e Privadas;
- As áreas de enquadramento do ENADE são os respectivos códigos:

Código da Área	Significado	Sugestão de Sigla	2021	2017	2014
72	Tecnologia em Análise e Desenvolvimento de Sistemas	ADS	✓	✓	✓
79	Tecnologia em Redes de Computadores	TRC	✓	✓	✓
4004	Ciência da Computação	CC	✓	✓	✓
4005	Ciência da Computação (Licenciatura)	LCBCC	✓	✓	✓
4006	Sistemas de Informação	SI	✓	✓	✓
6409	Tecnologia em Gestão da Tecnologia da Informação	GTI	✓	✓	✗
5809	Engenharia de Computação	EC	✗	✓	✓

- Como a Região Nordeste é composta por 9 estados, na hora de extrair devemos extrair esses estados.
- As colunas de ANO_FIM_EM e ANO_IN_GRAD devem ser avaliadas com mais critério, porque tem alunos que informaram números randômicos e que não condizem (terminar o EM em 1500, iniciar a graduação em 1600) etc.
- O tipo de presença a ser considerado deve ser APENAS do código 555, presente e com resultado válido.

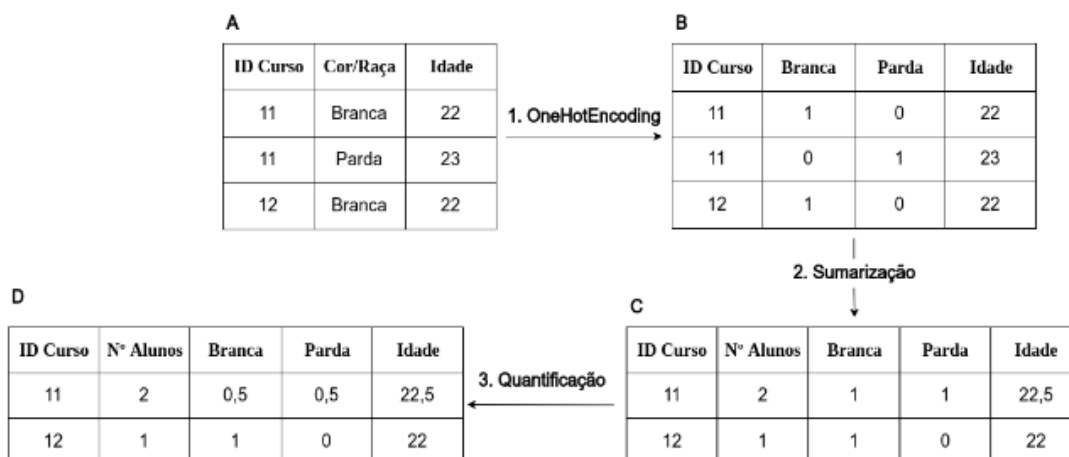
Com isso, vamos obter apenas os cursos com resultados válidos, e descartar uma grande parte do conjunto que possuem códigos de cursos que não fazem parte da região Nordeste e principalmente, de computação.

Questão	Temática	Respostas	Sugestão
01	Estado civil	5 opções	Manter.
02	Cor/Raça	6 opções	Manter.
03	Nacionalidade	3 opções	Dropar coluna
04	Escolarização do Pai	6 opções	Manter.
05	Escolarização da Mãe	6 opções	Manter.
08	Renda total da Família	7 opções	Manter.
09	Situação Financeira	6 opções	?
11	Tipo de Financiamento	11 opções	?
12	Auxílio Permanência	6 opções	2 opções.
13	Bolsa Acadêmica	6 opções	2 opções.
15	Ação afirmativa/Inclusão	6 opções de sim e não	2 opções.
17	Tipo de Escola EM	6 opções	Manter,
21	Curso superior na família	Sim/Não	Manter.

Mesclagem das informações do CURSO (quando forem sumarizadas) com as informações de IES/CURSO).

A ideia é que, passado um processo de sumarização, as informações sejam mescladas. A priori, temos cada linha sendo um aluno, o que torna a dimensionalidade grotescamente gigante.

Ano	Código do Curso	Estado Civil	...
2021	1000	A	...
...
2021	1000	D	...
2021	1505	A	...



1. Sumarizar e quantificar atributos socioeconômicos?
2. Como tratar os missings em GTI em 2014 e EC em 2021, já que não estavam presentes?
3. Para perfil socioeconômico convém excluir itens pedagógicos (nota, turno, IES). Se quiser analisar desempenho versus perfil, mantenha a nota geral (NT_GER) para pós-análise?
4. Codificação de variáveis ordinais e padronização z-score.
5. Para definir a unidade de análise:
 - a. Estudantes – clusters de perfis individuais;
 - b. Curso/IES – transformar **várias linhas-aluno** em **uma única linha-curso/IES**, resumo a estatísticas únicas (média/proporção).
 - c. Se eu quiser comparar a evolução temporal dos perfis, é mais fácil manter o nível de estudante e depois clusterizar **cluster x ano x curso**?

Possíveis análises:

Código do Curso	Estado Civil	Interpretação
Socioeconômicas (renda, escolaridade, tipo de escola, status, etc)	Clusters de alta renda, pais com ensino superior etc; Clusters de renda média, pais ensino médio, escola privada etc; Clusters de baixa renda etc etc	Descobrir perfis socioeconômicos. A média de NT_GER so é calculada para enxergar diferenças de desempenho.
...	...	O algoritmo mistura desempenho e na hora de formar os grupos não consegue explicar se NT_GER explica o cluster ou foi a causa do agrupamento.
1000	D	Não diz nada, cria faixas de notas e não diz nada sobre escola etc.