

## RELATÓRIO DA BASE DE DADOS SOCIODEMOGRÁFICOS - 24/04/23

**NOME COMPLETO:** Pode ser usado como ID, mas na prática deve ser removido.

**1 - Sexo:** Já está binarizado, logo não precisa modificar.

**2- Idade:** Não é necessário modificar, dado que é um valor numérico, mas é interessante normalizar o valor.

**3 - Raça:** Trata-se de uma variável categórica que vai de 1 a 5, então é interessante usar codificação one-hot.

**4 - Situação Civil:** Variável categórica que varia entre 1 e 3, e então é interessante usar one-hot.

**5 - Local de origem:** É uma variável categórica que varia de 1 a 93, sendo 1 Sobral. Acredito ser melhor criar uma coluna com indicando se é de Sobral ou não.

**6 - Local de Residência:** Segue o mesmo princípio do local de origem. Nesse caso, seria interessante, dependendo do nível de dificuldade, criar uma nova coluna com a distância da cidade de origem e Sobral.

**7 - Frequência que viaja para sua cidade:** Variável que varia de 0 a 5, onde 0 significa que a pessoa é de Sobral e 5 semestralmente. Como se trata de algo hierárquico, não preciso modificar.

**8 - Atualmente reside com:** Variável categórica, que deve usar one-hot.

**9 - Qual a sua religião?:** Variável categórica que deve usar one-hot.

**10 - Qual o curso de graduação cursa na UVA:** Variável categórica que tem muitas categorias, de 1 a 19. Acho interessante agrupar pelas áreas determinadas no texto:

- Ciências Biológicas, Zootecnia e Enfermagem - Área de Ciências Biológicas e da Saúde
- Ciências Contábeis, Administração e Sup.Const. Edifício - Área de Negócios e Administração
- Letras, Pedagogia, Educação Física, História, Filosofia, Direito, Geografia, Ciências Sociais e Química - Área de Humanidades e Ciências Sociais
- Matemática, Engenharia Civil, Computação e Física - Área de Exatas e Tecnologia.

**11 - Em que semestre você está matriculado:** Variável numérica que representa o semestre da pessoa. Então mantém como está.

**12 - Quantas disciplinas e/ou módulos está cursando no semestre em questão?:** Trata-se de uma variável numérica que representa o número de disciplinas. Então mantém como está.

**13 - Já reprovou alguma disciplina/ módulo?:** Variável já binarizada, sim ou não, logo mantém como está.

**14 - Em termos gerais, você considera seu curso estressante?:** Variável binarizada, sim ou não, logo mantém como está.

**14. 1 - Indique os principais fatores de estresse:** Repete a coluna anterior, logo deve ser removida.

- **Principais estressores (1 a 4):** Está em one-hot, então mantém como está.

**15 - Já se afastou dos estudos por algum motivo?:** Variável binarizada, mantém como está.

**15. 1 - Qual motivo do afastamento:** Repete a coluna anterior, logo pode ser removida.

- **Motivo do afastamento (1 a 4) :** Está em one-hot, então mantém como está.

**16 - Em que turno você está frequentando a maior parte das disciplinas?:** Variável categórica com 4 valores, então é interessante ser codificada para one-hot.

**17 - Qual o transporte você utiliza para chegar à Universidade:** Variável categórica que deve usar one-hot.

**18 - Atividades profissionais:** Variável binarizada, mantém como está.

**19 - Atividades acadêmicas (Estágio):** Variável binarizada, mantém como está.

**20 - Quanto a Bolsas ou Auxílio Universitário:** Variável binarizada, mantém como está.

**20. 1 - Caso receba, especifique qual ou quais:** Coluna repete os valores da anterior, então deve ser removida.

- **Tipos de auxílio:** Já está em one-hot para cada um dos tipos de auxílio, logo não precisa alterar.

**21 - Desenvolve algum tipo de atividade física de forma regular?:** Variável binarizada, mantém como está.

**21. 1- Se sim, qual a frequência (por semana)?:** Como se trata de quantidade e é uma variável numérica que varia entre 0 e 7, mantém como está.

**22 - Horas de Lazer:** Como se trata de quantidade e é uma variável numérica, mantém como está.

**23 - Quais as suas principais atividades de Lazer:** A coluna apresenta apenas um valor, então pode ser removida.

- **Tipos de atividade de lazer:** Já está em one-hot, logo mantém como está.

**24 - Faz uso contínuo de algum medicamento para tratamento da saúde:** Variável binarizada, mantém como está.

**24.1 - Qual medicamento?:** Coluna repete os valores da anterior, então deve ser removida.

- **Tipos de medicamentos:** Já está em one-hot, logo mantém como está.

**25 - Fez uso de alguma substância psicoativa durante os últimos três meses?:** Variável binarizada, mantém como está.

**25.1 - Qual substância psicoativa?:** Coluna repete os valores da anterior, então deve ser removida.

- **Tipos de substâncias psicoativas:** Já está em one-hot, logo mantém como está. No entanto, acho interessante criar uma outra coluna sendo ela a soma dos tipos de substância psicoativa.

**25. 2 - Durante os três últimos meses, com que frequência você utilizou essa(s) substância(s):** Variável que se refere à quantidade e é numérica, então mantém como está.

**26 - Em geral, qual a quantidade aproximada de horas de sono diárias?:** Variável numérica referente a quantas horas foram dormidas. Talvez seja interessante agrupar segundo o agrupamento feito no texto: 2 a 5 Horas, 6 a 7 Horas, 8 a 12 Horas e fazer one-hot encoding.