



Projeto

Pratique e solidifique as habilidades necessárias para o processo de **Importação e Arrumação** dos dados.

Visão Geral

Estudar por conta própria não é tarefa fácil, por isto criei este projeto para que você tenha a oportunidade de você aplicar o que foi visto em aula.

Este projeto consiste em um problema comum do cotidiano de quem trabalha com dados. Em quase todas as situações, espera-se que os dados que você vai receber ou obter estarão desarrumados, por isso quero exemplificar isto com os *dados do IBGE sobre pessoas de 10 anos ou mais de idade que utilizaram a Internet, no período de referência dos últimos três meses, por grupos de anos de estudo, segundo as Grandes Regiões, as Unidades da Federação e as Regiões Metropolitanas em 2015*, como segue abaixo:

Fontes:

<https://seriesestatisticas.ibge.gov.br/series.aspx?no=7&op=0&vcodigo=FDT0957&t=pessoas-desocupadas-grupos-anos-estudo>

https://downloads.ibge.gov.br/downloads_estatisticas.htm

Seu trabalho será arrumar estes dados, identificando antes qual formato ele está e só depois escolher a técnica ideal para transformá-lo na forma padrão vista nas aulas.

Objetivo

- Desenvolver a habilidade de Importação e Arrumação dos dados tão presente no dia a dia do analista de dados.

O que será cobrado neste projeto

- Habilidade de importação de arquivos .xlsx
- Habilidade de trabalhar com tibbles
- Construção de partições de um mesmo banco de dados
- Capacidade de decidir qual técnica usar para arrumar os dados
- Capacidade de criar um arquivo final com base nas partições

Tempo total estimado

- 60 minutos

Etapas

1. Faça a importação do banco de dados para o R (Tabela 1.1.28.1 e 1.1.28.3)
2. Transforme-os em tibbles
3. Crie duas partições: uma para a quantidade absoluta e outra para as porcentagens
4. Nomear corretamente as variáveis de cada partição
5. Escolha corretamente qual função usar para arrumar os dados em cada partição
6. Junte os dados arrumados para criar apenas um tibble final que seja algo como:

```
# A tibble: 252 x 4
  Local      `Anos de estudo` Quantidade `Porcentagem (%)`
  <fct>      <chr>          <chr>      <chr>
1 Brasil    < 4 anos      5068.075    5
2 Norte    < 4 anos      348.574     5.2
3 Rondônia < 4 anos      42.42       6
4 Acre      < 4 anos      18.042      6.3
5 Amazonas < 4 anos      73.231      4.6
6 Roraima   < 4 anos      13.017      4.9
7 Pará      < 4 anos      162.343     5.5
8 Região Metropolitana de Belém < 4 anos      54.131      4.4
9 Amapá     < 4 anos      14.188      4.2
10 Tocantins < 4 anos      25.333      4.5
# ... with 242 more rows
```

7. Por fim, envie o código que você desenvolveu para chegar a este tibble final.