

FUNDAÇÃO GETULIO VARGAS

ANALISE EXPLORATÓRIA DE DADOS E VISUALIZAÇÃO

LUIS ANTONIO ALVES

Rio de Janeiro - RJ
Abril 2021

HISTÓRICO DOS DADOS, ORIGEM E MÉTODOS APLICADOS

Escolhi os dados abertos da Receita Federal, onde a base está disponível. Feita uma análise prévia e faço um relato das minhas conclusões:

- 1) No momento que fiz o acesso as base a era composta por 20 arquivos com tamanho variando o tamanho entre 234,285 kb e 396.120 kb (sempre compactados). Após descompactar o primeiro arquivo veio com o nome (K3241.K03200DV.D00703.L00001), com o tamanho de 5.287.555 kb e tipo texto com 4.508.291 linhas e 1.200 posições de largura e não existe cabeçalho. Não contendo separação por espaços ou caracteres, mas no site também está disponível um arquivo LAYOUT_DADOS_ABERTOS_CNPJ.pdf, que contém descrições de toda estrutura do arquivo.
- 2) Restringir os dados a cidade de Barueri, município da Região Metropolitana de São Paulo, no estado de São Paulo.