



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

CENTRO DE TECNOLOGIA

ESCOLA POLITÉCNICA

Uso de Planilhas Eletrônicas para CTA de dados

MAE015 - Tóp Eng. Dados A 2024.2

Rio de Janeiro, 2024

Igor Soares Colonna – 119174972

Pedro Alves de Araújo – 123706163

Matheus Delduque Lopes da Silva – 122062207

Mikaela Rikberg Alves – 119106107

Lucas Barcelos de Sousa de Abreu – 123439994

I. Introdução

O presente relatório tem como objetivo estudar e aplicar as técnicas de carga, tratamento e análise de datasets utilizando a ferramenta MS-Excel. A análise de dados desempenha um papel fundamental em diversos campos, como meteorologia e política, sendo crucial para o suporte à tomada de decisões. Neste estudo, exploramos o uso de MS-Excel, uma ferramenta amplamente disponível e versátil, para realizar a manipulação e análise de dados de forma eficiente.

Para alcançar esse objetivo, o grupo foi encarregado de desenvolver dois estudos distintos, com diferentes níveis de complexidade, a partir de bases de dados disponíveis em diretórios de dados públicos.

O primeiro estudo se concentra em um dataset com dados relacionados ao eleitorado brasileiro, os estudos foram direcionados apenas para dois estados brasileiros: **Acre (AC) e Amapá (AP)**. Os dados foram obtidos no site “Eleitorado Atual” do TSE,

O segundo estudo se concentra em um dataset com informações meteorológicas, os estudos foram direcionados apenas para **uma estação nos anos de 2023 e 2024**. Os dados foram obtidos no site do Instituto Nacional de Meteorologia.

O primeiro estudo se concentra em um dataset com grau de dificuldade moderado, permitindo uma familiarização com as operações básicas de tratamento e análise de dados, como filtros, ordenação, uso de fórmulas e criação de gráficos. Já o segundo estudo, mais avançado, envolve um dataset mais complexo, exigindo técnicas adicionais como o uso de tabelas dinâmicas, análise de correlação entre variáveis e exploração de grandes volumes de dados.

II. Análise dos dados

A. Funções e Operações Utilizadas

Conversão de Arquivo

A conversão dos arquivos de **.csv** para **.xlsb** foi feita para reduzir o tamanho total dos datasets, o que proporcionou uma economia significativa de espaço em disco, especialmente em arquivos com muitas linhas e colunas. O formato **.xlsb** compacta os dados de maneira mais eficiente.

Filtros e Seleção de Dados

Foram utilizados **filtros** no Excel para selecionar apenas as colunas de interesse e eliminar as que não eram relevantes para a análise. As colunas e linhas selecionadas incluíram informações como gênero, faixa etária, estado civil e grau de escolaridade.

Remoção de Dados Inválidos

Linhas contendo valores inválidos foram excluídas, reduzindo ainda mais o tamanho do arquivo e garantindo a integridade dos dados para as análises.

Cálculos com Funções Estatísticas

Utilizamos funções como **SOMASE** e **MÉDIASE** para realizar somas condicionais e calcular as estatísticas dos eleitores com base em critérios específicos. As funções **SUBTOTAL** foram usadas para contar e somar os valores visíveis após a aplicação de filtros.

B. Dados obtidos dos datasets:

1. Datasets Eleitorais:

Perfil Eleitor Acre (AC)

- **Tamanho Inicial:** 72.630 KB
- **Mudança para formato .xlsb:** 5499 KB

- **Linhas Iniciais:** 350803
- **Colunas Iniciais:** 30
- **Células Iniciais:** 10524090

Redução de Dados:

- Após a seleção dos dados de interesse (linhas e colunas relevantes), o tamanho do arquivo foi reduzido para **3316 KB**.
- **Linhas com dados de interesse:** 350804
- **Colunas com dados de interesse:** 10
- **Células com dados de interesse:** 3508040

Exclusão de Linhas Inválidas:

- Após a exclusão de linhas com dados inválidos, o arquivo foi reduzido para **3307 KB**.

Estatísticas Gerais:

- **Soma total de eleitores:** 612003
- **Soma de eleitores filtrados** (homens, 35 a 39 anos, casados, com nível médio completo): 2684

Perfil Eleitor Amapá (AP)

- **Tamanho Inicial:** 72.630 KB
- **Mudança para formato .xlsb:** 5499 KB
- **Linhas Iniciais:** 298258
- **Colunas Iniciais:** 30
- **Células Iniciais:** 8947740

Redução de Dados:

- Após a seleção dos dados de interesse, o tamanho do arquivo foi reduzido para **4024 KB**.
- **Linhas com dados de interesse:** 298258
- **Colunas com dados de interesse:** 10
- **Células com dados de interesse:** 2982580

Exclusão de Linhas Inválidas:

- Após a exclusão de linhas com dados inválidos, o arquivo foi reduzido para **3951 KB**.

Estatísticas Gerais:

- **Soma total de eleitores:** 570845
- **Soma de eleitores filtrados** (homens, 35 a 39 anos, casados, com nível médio completo): 2012

2. Datasets INMETRO:

Dados Meteorológicos Novo Horizonte 2023

Tamanho inicial: 794 KB

Linhas iniciais: 8760

Colunas iniciais: 19

Células iniciais: 166.440

Tamanho com dados de interesse: 606 KB

Mudando tipo de arquivo para .xlsb: 605 KB

Linhas com dados de interesse: 8760

Colunas com dados de interesse: 11

Células com dados de interesse: 96.360

Tamanho sem linhas com dados inválidos:
605 KB

Redução de Dados:

- Após a seleção dos dados de interesse, o tamanho do arquivo foi reduzido para **606 KB**.
- **Linhas com dados de interesse:** 8760
- **Colunas com dados de interesse:** 11
- **Células com dados de interesse:** 96.360

Exclusão de Linhas Inválidas:

- Após a exclusão de linhas com dados inválidos, o arquivo foi reduzido para **606 KB**.

Estatísticas Gerais:

- **Média da temperatura:** em 2023(26,5C°).

- **Meses analisados:** janeiro, fevereiro, março, abril, maio, junho.

Dados Meteorológicos Novo Horizonte 2024

Tamanho inicial: 493 KB

Linhas iniciais: 5857

Colunas iniciais: 19

Células iniciais: 111.283

Redução de Dados:

- Após a seleção dos dados de interesse, o tamanho do arquivo foi reduzido para **330 KB**.
- **Linhas com dados de interesse:** 5857
- **Colunas com dados de interesse:** 11
- **Células com dados de interesse:** 64.427

Exclusão de Linhas Inválidas:

- Após a exclusão de linhas com dados inválidos, o arquivo foi reduzido para **330 KB**.

Estatísticas Gerais:

- **Média da temperatura:** média em 2024(28C°).
- **Meses analisados:** janeiro, fevereiro, março, abril, maio, junho.

C. Dificuldades Encontradas

- **Compatibilidade de Funções:** A função **FILTRO** não estava disponível em certas versões do Excel, o que nos levou a usar métodos alternativos como a criação de colunas auxiliares para calcular medianas e outras estatísticas condicionais.
- **Redução de Tamanho:** Embora o formato **.xlsb** tenha sido eficiente para compactar os arquivos, ainda foi necessário realizar ajustes adicionais na seleção dos dados para atingir a redução desejada.

- **Suplementos e Extensões Utilizados:** O **Suplemento de Análise de Dados** foi habilitado no Excel para realizar algumas operações estatísticas avançadas, como cálculos de **desvio padrão** e **regressões**. Utilizamos também atalhos e ferramentas básicas do Excel, como o uso de filtros, soma condicional e Tabelas Dinâmicas.

III. Resultados

Análise dos resultados estatísticos eleitorais:

Ao comparar as estatísticas dos datasets dos estados do Acre e Amapá, observamos que, no Amapá, a faixa etária mais comum entre os eleitores é de 25 a 29 anos, enquanto no Acre, a faixa etária predominante é de 21 a 24 anos. Em ambos os estados, a maioria da população eleitoral é solteira, destacando-se como o estado civil mais frequente entre os eleitores. No que diz respeito ao nível de escolaridade, a maior parte dos eleitores em ambos os estados possui ensino médio incompleto, refletindo uma distribuição educacional semelhante entre eles. Além disso, a faixa etária de 21 a 24 anos apresenta uma grande concentração de eleitores tanto no Acre quanto no Amapá, reforçando a relevância desse grupo nas eleições. Por fim, a maior parte dos eleitores está concentrada nas capitais dos dois estados, Rio Branco (Acre) e Macapá (Amapá), evidenciando a importância dessas cidades nos resultados eleitorais.

Proposta para comitê de campanha eleitoral:

Com base nos dados analisados, podemos identificar alguns pontos relevantes que podem auxiliar na estratégia de campanha de um candidato. No estado do Acre, o município com o menor número de eleitores é Santa Rosa do Purus, enquanto no Amapá, o município com o menor número de eleitores é Pracuúba. Apesar de serem localidades com poucos eleitores, esses municípios podem ser alvos de campanhas específicas, uma vez que, em áreas com menor densidade populacional, cada voto pode ter um impacto significativo nos resultados.

Por outro lado, ao excluir as capitais, os municípios com o maior número de eleitores são Sena Madureira no Acre e Santana no Amapá. Essas cidades, com uma alta concentração de eleitores, devem ser priorizadas nas ações de campanha. Estratégias mais intensivas, como comícios e ações presenciais, podem garantir uma maior visibilidade do candidato e um impacto direto no eleitorado, maximizando o potencial de votos.

Ao observar a variação de eleitores por município e gênero, vemos que municípios menores, como Santa Rosa do Purus, têm uma quantidade muito reduzida de eleitores, enquanto as capitais, Rio Branco e Macapá, apresentam

uma concentração significativa. Além disso, a distribuição de eleitores por gênero é, em sua maioria, equilibrada, mas alguns municípios podem apresentar uma predominância de um dos gêneros, o que pode influenciar na abordagem da campanha em certas localidades.

Analisando as características predominantes dos eleitores, identificamos que a maioria, em ambos os estados, está na faixa etária de 21 a 24 anos. Isso sugere que os jovens adultos representam uma parte importante do eleitorado, sendo recomendável que a campanha foque em temas como educação, emprego e tecnologia, que são mais atrativos para esse público. Além disso, a maior parte dos eleitores é solteira, o que sugere que questões relacionadas à juventude e independência podem ter grande repercussão.

Outro ponto relevante é o nível de escolaridade dos eleitores. A maioria dos eleitores em ambos os estados possui ensino médio incompleto, o que indica que propostas voltadas para a qualificação profissional e melhoria no acesso à educação podem ter um impacto significativo.

Em resumo, é recomendável que a campanha se concentre em municípios com alta concentração de eleitores, como Sena Madureira no Acre e Santana no Amapá, enquanto também considere as características demográficas predominantes. Propostas que abordem temas como oportunidades educacionais, inserção no mercado de trabalho e independência para jovens adultos podem ser estratégicas para atingir a maioria do eleitorado.

Análise dos resultados estatísticos meteorológicos:

A cidade de Novo Horizonte, localizada no estado de São Paulo, apresentou um aumento significativo na média das temperaturas ao comparar os períodos de janeiro a junho dos anos de 2023 e 2024. A análise revelou uma elevação de 1,5°C nesse intervalo. Esse aumento pode ser atribuído a diversos fatores, como a intensificação da poluição, queimadas recorrentes e a ação antrópica em geral, que contribuem para a degradação ambiental e afetam diretamente o clima local.

Proposta Ambiental para Redução dos Efeitos do Aquecimento em Novo Horizonte

1. Criação de Corredores Verdes e Expansão de Áreas Arborizadas

- **Objetivo:** Ampliar a cobertura vegetal urbana, plantando árvores em áreas públicas e criando corredores verdes ao longo de avenidas e parques. Isso ajudará a diminuir a temperatura local por meio da sombra e da evapotranspiração das plantas, além de melhorar a qualidade do ar.

- **Benefícios:** Redução do efeito de ilhas de calor, purificação do ar e promoção da biodiversidade urbana.

2. Adoção de Energias Renováveis

- **Objetivo:** Incentivar o uso de energias renováveis, como solar e eólica, tanto para residências quanto para indústrias, por meio de incentivos fiscais e subsídios para instalação de sistemas fotovoltaicos e turbinas eólicas de pequeno porte.
- **Benefícios:** Redução da dependência de combustíveis fósseis, contribuindo para a diminuição das emissões de gases de efeito estufa.

3. Educação Ambiental e Conscientização da População

- **Objetivo:** Desenvolver programas educativos nas escolas e campanhas de conscientização para a comunidade, focando na importância da preservação ambiental, redução do uso de plásticos, reciclagem e economia de energia.
- **Benefícios:** A longo prazo, uma população mais consciente contribuirá ativamente para a mitigação dos efeitos das mudanças climáticas, adotando hábitos de consumo mais sustentáveis.

Conclusão

Através da otimização dos arquivos e aplicação de filtros, conseguimos uma análise focada em dados relevantes, reduzindo consideravelmente o tamanho dos arquivos e extraíndo informações precisas sobre o perfil eleitoral e dados meteorológicos. As dificuldades encontradas foram superadas utilizando alternativas disponíveis no Excel para calcular as estatísticas necessárias.

Repositório git contendo os arquivos xlsb.

<https://github.com/mikasa42/Uso-de-Planilhas-Para-Tratamento-de-Dados>

IV. Apêndices

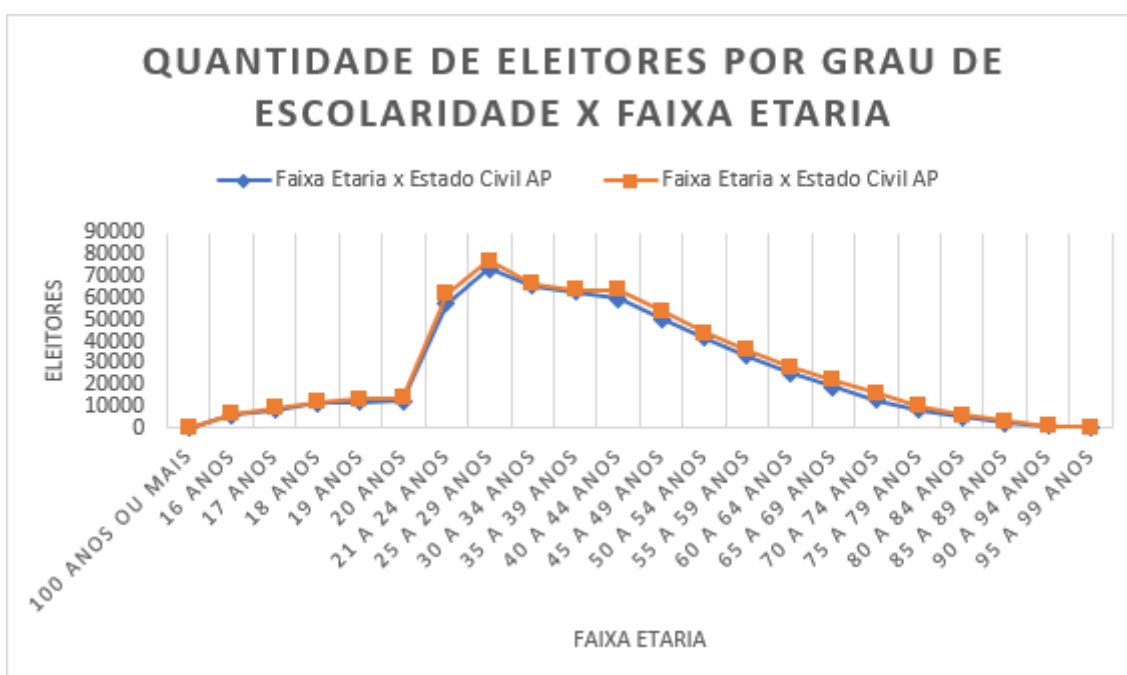
DS_GENERO (Vários itens)

Grau de Escolaridade x Faixa Etaria AC	Quantidade Total
100 anos ou mais	177
16 anos	6774
17 anos	9645
18 anos	12289
19 anos	13445
20 anos	13962
21 a 24 anos	62056
25 a 29 anos	77072
30 a 34 anos	66325
35 a 39 anos	63728
40 a 44 anos	63828
45 a 49 anos	53887
50 a 54 anos	43790
55 a 59 anos	36161
60 a 64 anos	28423
65 a 69 anos	22099
70 a 74 anos	16340
75 a 79 anos	10372
80 a 84 anos	6430
85 a 89 anos	3332
90 a 94 anos	1362
95 a 99 anos	506
Total Geral	612003

DS_GENERO (Vários itens)

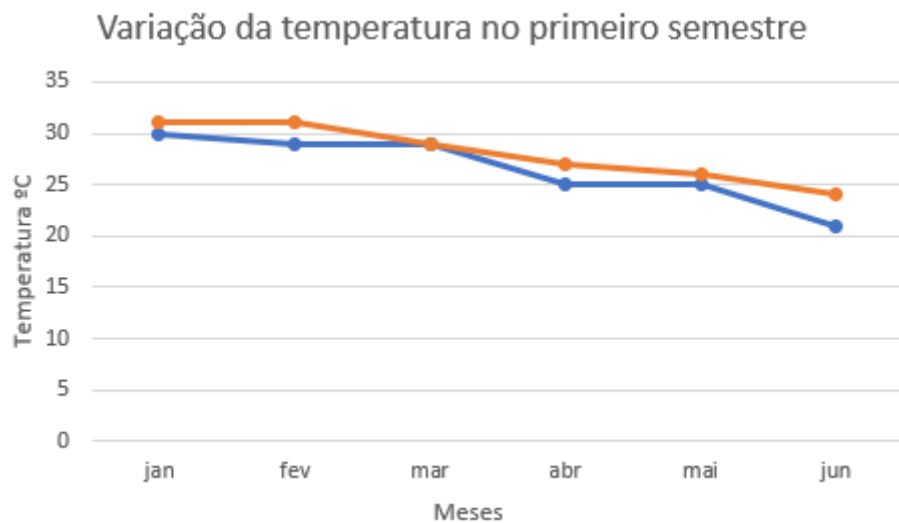
Faixa Etaria x Estado Civil AP	Quantidade Total
100 anos ou mais	122
16 anos	6248
17 anos	8468
18 anos	11536
19 anos	12102
20 anos	12521
21 a 24 anos	57257

25 a 29 anos	73504
30 a 34 anos	65782
35 a 39 anos	62870
40 a 44 anos	59608
45 a 49 anos	50464
50 a 54 anos	41863
55 a 59 anos	33176
60 a 64 anos	25716
65 a 69 anos	19272
70 a 74 anos	13057
75 a 79 anos	8360
80 a 84 anos	5275
85 a 89 anos	2411
90 a 94 anos	961
95 a 99 anos	272
Total Geral	570845



Rótulos de Linha	Máx. de TEMPERATURA MÁXIMA NA HORA ANT. (AUT) (°C)
<01/01/2023	
jan	30
fev	29
mar	29
abr	25
mai	25
jun	21
Total Geral	30

Rótulos de Linha	Máx. de TEMPERATURA MÁXIMA NA HORA ANT. (AUT) (°C)
<01/01/2024	
jan	31
fev	31
mar	29
abr	27
mai	26
jun	24
Total Geral	31



V. Bibliografia

[1]- INSTITUTO NACIONAL DE METEOROLOGIA (INMET). **Dados Históricos**. Disponível em: <https://portal.inmet.gov.br/dadoshistoricos>.

[2]- TRIBUNAL SUPERIOR ELEITORAL (TSE). **Eleitorado Atual - Dados Abertos**. Disponível em: https://dadosabertos.tse.jus.br/pt_PT/dataset/eleitorado-atual.