

## PROJETO EM CIÊNCIA DE DADOS

### SUMÁRIO

SEMESTRE	2024/2
PROJETO	Influência das Despesas e Receitas Partidárias no Total de Votos Obtidos no Rio Grande do Sul
COMPONENTES DO GRUPO	Bernardo Kautz Gabriel de Pinho das Neves Rodrigues Gustavo Alves Collioni

#### Breve descrição do problema

Baseando-se em séries temporais, as despesas e receitas partidárias exercem influência sobre o total de votos obtidos no Rio Grande do Sul?

#### Breve descrição da solução proposta

- 1) **Exploração inicial:** reunir as bases de dados referentes às votações por partido político nos municípios gaúchos (2012-2024) e às prestações de contas disponíveis (2018-2024), corrigindo dados inconsistentes e filtrando aqueles de interesse.
- 2) **Fitting:** utilizar métodos estatísticos e visualizações para identificar tendências, comparando variáveis para revelar informações significativas.
- 3) **Interpretação e validação:** verificar se os resultados alcançados respondem à pergunta inicial e garantir a consistência frente aos dados, assegurando análises robustas e reproduzíveis; se possível, estabelecer o indicador  $\frac{\text{Despesas Anuais}}{\text{Votos Recebidos}}$  e prognósticos.
- 4) **Finalização:** desenvolver um relatório detalhado, incluindo gráficos dinâmicos e tabelas interativas, para sugerir recomendações embasadas.

#### Fases da Metodologia CRISP-DM

- 1) **Compreensão dos Dados:** realiza-se uma análise exploratória dos dados coletados, visando identificar padrões, anomalias e a qualidade geral do conjunto. Concluído.
- 2) **Preparação dos Dados:** busca-se efetuar atividades como limpeza, tratamento de valores ausentes, transformação de variáveis e criação de novas *features*. Concluído.
- 3) **Autocrítica:** examina-se os resultados obtidos, comparando-os com os objetivos do projeto e identificando oportunidades de melhoria. Concluído.

## Resumo do que foi concluído até o momento

Na primeira entrega, foram realizadas a compreensão e a integração. Em seguida, deu-se início às atividades de limpeza, criação de gráficos e desenvolvimento das principais ideias a serem testadas. Para finalizar, delineou-se uma conclusão dirigida pelo indicador  $\frac{\text{Despesas Anuais}}{\text{Votos Recebidos}}$ .

## Autocrítica

A experiência adquirida revelou as restrições inerentes à manipulação de dados eleitorais, especialmente considerando o elevado grau de sigilo. O nível de aderência à metodologia CRISP-DM parece ser satisfatório. A avaliação do progresso foi extremamente positiva, com todos os membros trabalhando de forma colaborativa. O escopo foi atingido com êxito, porém com um questionamento em aberto. Nota: 9,0.

# RELATÓRIO

## 1. Compreensão dos Dados

### Coleta dos dados

As bases de dados do Tribunal Superior Eleitoral (TSE) mostram-se organizadas e acessíveis, providenciadas no portal de transparência em formato .csv.

### Descrição dos dados

Somente as variáveis usadas foram separadas, descartando as demais. Caso outras sejam abrangidas, serão adicionadas.

Despesas Partidárias, RS (2018-2024)	
Campo	Significado
AA_EXERCICIO	Ano de referência da eleição para geração do arquivo
CD_MUNICIPIO	Código do município de abrangência do prestador de contas
NM_MUNICIPIO	Descrição do município de abrangência do prestador de contas
NR_CNPJ_PRESTADOR_CONTA	CNPJ do prestador de contas
SG_PARTIDO	Sigla do partido político prestador de contas
DS_TP_FORNECEDOR	Descrição do tipo de fornecedor informada pelo prestador de contas em relação à despesa, podendo assumir os valores: 'Pessoa Física' ou 'Pessoa Jurídica'
NR_CPF_CNPJ_FORNECEDOR	CPF/CNPJ do fornecedor de bens e/ou serviços informada pelo prestador de contas em relação à despesa
DT_PAGAMENTO	Data do pagamento da despesa informada pelo prestador de contas em relação à despesa.
VR_GASTO	Valor da despesa em reais (R\$), informada pelo prestador de contas em relação à despesa

Receitas Partidárias, RS (2018-2024)	
Campo	Significado
CD_MUNICIPIO	Código do município de abrangência do prestador de contas
NM_MUNICIPIO	Nome do município de abrangência do prestador de contas
NR_CNPJ_PRESTADOR_CONTA	CNPJ do prestador de contas
SG_PARTIDO	Sigla do partido político prestador de contas
DS_TP_ORIGEM_DOACAO	Descrição do tipo de origem da doação (recurso) informada pelo prestador de contas em relação à doação, podendo assumir os valores: 'Origem não identificada', 'Cotas do Fundo Partidário', 'Candidatos', 'Pessoas Físicas', 'Pessoas Jurídicas' ou Partidos políticos'
NR_CPF_CNPJ_DOADOR	CPF/CNPJ do doador declarado à Justiça Eleitoral
CD_MUNICIPIO_DOADOR	Código do município do doado
NM_MUNICIPIO_DOADOR	Nome do município do doador
DS_TP_ESPECIE_RECURSO	Espécie do recurso, podendo assumir os valores: 'Aviso de crédito', 'Cartão de crédito', 'Cheque', 'Em espécie', 'Outros títulos de crédito' ou 'Transf. Eletrônica/Débito/Boleto'
VR_RECEITA	Valor da doação em reais (R\$)

\* Não há uma coluna que inclua o ano de apuração.

Votação por Município/Zona Eleitoral, RS (2012-2024)	
Campo	Significado
ANO_ELEICAO	Ano de referência da eleição para geração do arquivo
NR_TURNO	Número do turno da eleição
DS_ELEICAO	Descrição da eleição
NM_MUNICIPIO	Nome do município onde ocorreu a eleição
DS_CARGO	Descrição do cargo de candidatura
SG_PARTIDO	Sigla do partido político do candidato
QT_VOTOS_NOMINAIS_VALIDOS	Quantidade de votos nominais totalizados

### Análise exploratória dos dados

Os conjuntos de dados foram cruzados conforme seus respectivos anos, a fim de compreender a extensão dos aportes financeiros recebidos pelos partidos políticos, suas despesas e, por último, a soma de votos nominais válidos, categorizando-os segundo uma variedade critérios, como fornecedor de bens e/ou serviços mais relevante, espécie mais comum dos pagamento e cargo de candidatura mais impulsionado.

Os valores processados atendem ao padrão esperado, demonstrando uma organização adequada às necessidades da pesquisa, com poucas exceções de linhas incorretas.

## Verificação de qualidade dos dados

A abordagem priorizou garantir a consistência dos valores de acordo com as nomenclaturas das variáveis. Por exemplo, poucas linhas apresentaram “votos nominais válidos” iguais a -3, algo que é considerado contraintuitivo.

## 2. Preparação dos Dados

### Integração de dados

#### **Despesas Partidárias, RS (2018-2024):**

Realizou-se a criação do conjunto de dados “despesa\_geral\_RS” por meio da união das tabelas “despesa\_anual\_2017\_RS”, “despesa\_anual\_2018\_RS”, “despesa\_anual\_2019\_RS”, “despesa\_anual\_2020\_RS”, “despesa\_anual\_2021\_RS”, “despesa\_anual\_2022\_RS”, “despesa\_anual\_2023\_RS” e “despesa\_anual\_2024\_RS”.

#### **Receitas Partidárias, RS (2018-2024):**

Realizou-se a criação do conjunto de dados “receita\_geral\_RS” por meio da união das tabelas “receita\_anual\_2017\_RS”, “receita\_anual\_2018\_RS”, “receita\_anual\_2019\_RS”, “receita\_anual\_2020\_RS”, “receita\_anual\_2021\_RS”, “receita\_anual\_2022\_RS”, “receita\_anual\_2023\_RS” e “receita\_anual\_2024\_RS”.

#### **Votação por Município/Zona Eleitoral, RS (2012-2024):**

Realizou-se a criação do conjunto de dados “votacao\_total\_RS” por meio da união das tabelas “votacao\_partido\_munzona\_2012\_RS”, “votacao\_partido\_munzona\_2014\_RS”, “votacao\_partido\_munzona\_2016\_RS”, “votacao\_partido\_munzona\_2018\_RS”, “votacao\_partido\_munzona\_2020\_RS”, “votacao\_partido\_munzona\_2022\_RS” e “votacao\_partido\_munzona\_2024\_RS”.

### Limpeza dos dados

#### **Despesas Partidárias, RS (2018-2024):**

As colunas a seguir foram excluídas, devido à sua irrelevância em relação ao escopo do trabalho: “TP\_DESPESA”, “SG\_UF”, “NR\_ZONA”, “DS\_GASTO”, “DS\_FONTE\_DESPESA”, “SQ\_DESPESA”, “HH\_GERACAO”, “AA\_AIDF”, “VR\_DOCUMENTO”, “NR\_AIDF”, “NM\_PARTIDO”, “DS\_TP\_DOCUMENTO”, “NR\_DOCUMENTO”, “CD\_TP\_ESFERA\_PARTIDARIA”, “CD\_TP\_DOCUMENTO”, “DT\_GERACAO”, “DS\_TP\_ESFERA\_PARTIDARIA”.

Também, retiraram-se as observações “#NULO#” e, de “VR\_PAGAMENTO”, com valor igual a zero.

#### **Receitas Partidárias, RS (2018-2024):**

As colunas a seguir foram excluídas, devido à sua irrelevância em relação ao escopo do trabalho: “DT\_GERACAO”, “HH\_GERACAO”, “CD\_TP\_ESFERA\_PARTIDARIA”, “DS\_TP\_ESPERA\_PARTIDARIA”, “SG\_UF”, “NR\_ZONA”, “NR\_ZONA”, “CD\_TP\_NATUREZA\_RECURSO”, “DS\_TP\_NATUREZA\_RECURSO”,



"CD\_TP\_ESPECIE\_RECURSO", "NR\_RECIBO\_DOACAO", "NR\_DOCUMENTO", "CD\_TP\_FONTE\_RECURSO",  
"NM\_PARTIDO", "CD\_TP\_ESFERA\_PARTIDARIA\_DOADOR", "NR\_ZONA\_DOADOR",  
"SQ\_CANDIDATO\_DOADOR", "SQ\_CANDIDATO\_DOADOR", "SQ\_CANDIDATO\_DOADOR",  
"DS\_CANDIDATO\_CARGO\_DOADOR", "CD\_TP\_FONTE\_RECURSO", "DS\_TP\_FONTE\_RECURSO",  
"DT\_RECEITA", "NR\_CANDIDATO\_DOADOR", "CD\_CANDIDATO\_CARGO\_DOADOR",  
"DS\_TP\_ESFERA\_PARTIDARIA\_DOADOR", "SG\_UF\_DOADOR".

Também, retiraram-se as observações "#NULO#" e, de "VR\_RECEITA", com valor igual a zero.

#### Votação por Município/Zona Eleitoral, RS (2012-2024):

As colunas a seguir foram excluídas, devido à sua irrelevância em relação ao escopo do trabalho:  
"DT\_GERACAO", "HH\_GERACAO", "CD\_TIPO\_ELEICAO", "NM\_TIPO\_ELEICAO", "CD\_ELEICAO",  
"DT\_ELEICAO", "TP\_ABRANGENCIA", "NR\_SECAO", "CD\_CARGO", "NR\_LOCAL\_VOTACAO",  
"SQ\_CANDIDATO", "NM\_LOCAL\_VOTACAO", "DS\_LOCAL\_VOTACAO\_ENDERECO", "SG\_UF", "SG\_UE",  
"NM\_UE", "CD\_MUNICIPIO", "NR\_ZONA".

Criação de atributos e registros

#### Despesas Partidárias, RS (2018-2024):

(-).

#### Receitas Partidárias, RS (2018-2024):

Devido à ausência de uma coluna que contenha o ano de apuração, decidiu-se pela inclusão de uma para que o dataset pudesse ser criado.

#### Votação por Município/Zona Eleitoral, RS (2012-2024):

(-).

Descrição do dataset final

Os conjuntos de dados finais estão estruturados em séries temporais, com periodicidades anual sobre "Despesas Partidárias, RS (2018-2024)" e "Receitas Partidárias, RS (2018-2024)" e bianual para "Votação por Município/Zona Eleitoral, RS (2012-2024)".

Para incluir as corridas eleitorais, foram definidas faixas que asseguram que as rubricas de anos sem eleições sejam levadas em conta.

Corridas Eleitorais, RS	
Faixa	Anos
1	2017-2018
2	2019-2020
3	2021-2022
4	2023-2024

Com o intuito de responder à questão central, optou-se por ignorar o segundo turno entre as contagens em decorrência da não participação de todos os grupos partidários e cargos eleitorais em “votação\_total\_RS” (ex.: vereadores não concorrem duplamente, somente no primeiro domingo). Sendo assim, através do indicador  $\frac{\text{Despesas Anuais}}{\text{Votos Recebidos}}$  e da verificação da paridade entre arrecadações e desembolsos, tornou-se possível responder, conforme os dados obtidos: “Baseando-se em séries temporais, as despesas e receitas partidárias exercem influência sobre o total de votos obtidos no Rio Grande do Sul?”.

### 3. Autocrítica

Durante todo o desenvolvimento das atividades, o grupo demonstrou compromisso com o enfrentamento do desafio recebido. A coordenação entre os membros facilitou a execução das tarefas e assegurou que todas as etapas fossem seguidas. Como consequência, foi viável chegar a um palpite, como apontado abaixo.

Em ordem, durante o intervalo abrangido, os partidos políticos mais populares foram: (i) Partido da Social-Democracia Brasileira (PSDB); (ii) Movimento Democrático Brasileiro (MDB); (iii) Partido dos Trabalhadores (PT); (iv) Progressistas (PP); e (v) Partido Liberal (PL). Dessa forma, o nível de eficiência de cada um é importante para compreender se a movimentação de recursos financeiros é decisiva, como levantado; o Gráfico 1, referente à elaboração em Power BI (na imagem, o recorte 2021-2022), serve como suporte.

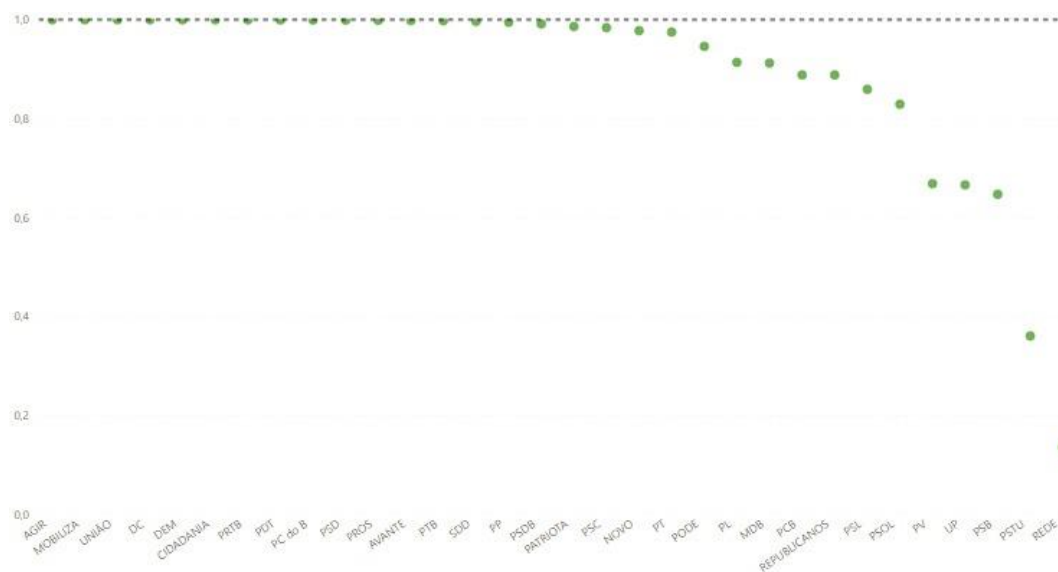
**Gráfico 1 – Razão de despesas anuais por votos recebidos**



Fonte: os autores a partir de TSE (2024).

Como evidenciado, partidos políticos com menor visibilidade possuem “eleitores cujo engajamento requer dispêndios mais significativos”, sustentando a afirmação de que a resposta correta à pergunta proposta deve ser negativa. No entanto, essa conclusão é válida apenas se a correlação entre “Despesas Partidárias, RS (2018-2024)” e “Receitas Partidárias, RS (2018-2024)” for, em sua maioria, próxima a 1, a fim de englobá-las inteiramente sem a obrigação de implementar outro indicador; o Gráfico 2, referente à elaboração em Power BI (na imagem, o acumulado), serve como suporte.

**Gráfico 2 – Despesas e receitas partidárias, correlação**



Fonte: os autores a partir de TSE (2024).

Em síntese, despesas e receitas partidárias não têm impacto sobre o total de votos conquistados no Rio Grande do Sul, considerando que, proporcionalmente, aqueles que possuem maior aceitação tendem a empregar menos capital financeiro por eleitor comprometido. Entre as potenciais contribuições futuras, destaca-se a investigação dos motivos pelos quais diferentes partidos políticos registram desembolsos superiores às suas arrecadações.

Por último, a fim de aprimorar a compreensão sobre a visualização, é exibido o diagrama inserido no Power BI associado ao trabalho, que parece caracterizar-se como um esquema do tipo constelação.

