



PROJETO EM CIÊNCIA DE DADOS

SUMÁRIO

SEMESTRE	2025/2
PROJETO	Além do Financeiro: quais são os privilégios da classe política brasileira?
COMPONENTES DO GRUPO	Felipe Augusto Batista Mendes dos Santos Guilherme Couto de Castro Gustavo Henrique da Silva Thiago Farias dos Santos

Breve descrição do problema

Avaliar a quantidade de presenças, participações e faltas justificadas de cada deputado, aplicar diferentes filtros que aproximem o regime de trabalho praticado na Câmara com o regime CLT nestes dados e calcular o ganho efetivo por dia trabalhado de cada deputado, explorando a desigualdade entre tabalhadores e a classe política do Brasil.

Breve descrição da solução proposta

Demonstração gráfica através de Dashboards comparativos, expondo os privilégios assegurados aos deputados federais brasileiros para além dos ganhos financeiros, além de exibir e explorar outliers.

Fases da Metodologia CRISP-DM

Compreensão do negócio:	<ul style="list-style-type: none">- Definir objetivo: analisar presenças, participações e faltas dos deputados.- Relacionar dados com parâmetros do regime CLT.- Formular hipóteses: diferença entre ganhos por dia trabalhado de deputados e trabalhadores CLT.	100%
Compreensão dos dados:	<ul style="list-style-type: none">- Coletar dados a respeito dos mandatos dos deputados da	0%

	<p>câmara através de fontes oficiais do governo.</p> <ul style="list-style-type: none">- Identificar variáveis relevantes: número de sessões realizadas, presenças, ausências (justificadas e não justificadas), participações em votações e remuneração total.- Entender as diferentes justificativas de ausência que um deputado federal pode alegar e se elas se aplicam de alguma forma àquelas aceitas pelo regimento da CLT.	
Modelagem:	<ul style="list-style-type: none">- Definir métricas de cálculo (ganho/dia, proporção de presença/ausência).- Aplicar filtros que simulam o regimento da CLT.- Criar cenários comparativos.	0%
Avaliação:	<ul style="list-style-type: none">- Verificar se os resultados confirmam as hipóteses.- Analisar se, após aplicação de filtros da CLT, muitos deputados ultrapassariam limites de faltas permitidos.- Comparar os resultados com o objetivo inicial.	0%
Aplicação:	<ul style="list-style-type: none">- Apresentar resultados em gráficos e tabelas comparativas.- Discutir o impacto social e político dos achados- Elaborar conclusões sobre a desigualdade trabalhista no país	0%

Resumo do que foi concluído até o momento

O grupo foi capaz de, até então, formular um problema de pesquisa capaz de ser respondido, além de terem sido formuladas hipóteses a partir de breves observações nos dados fornecidos pela Câmara dos Deputados. Também foram calculadas algumas das métricas pretendidas em uma amostra ínfima de dados de deputados selecionados aleatoriamente e, a partir destes cálculos, foi observada a grande possibilidade de que o problema de pesquisa seja respondido.

RELATÓRIO

1. Compreensão do Negócio

A pesquisa almeja avaliar a assiduidade dos deputados federais brasileiros, levando em consideração presenças, participações em votações e faltas (justificadas ou não). Serão determinados filtros baseados nas normas e regimentos da Consolidação das Leis de Trabalho (CLT), considerando apenas os artigos e definições válidos até o presente momento. Os filtros terão como objetivo remover alguns dos privilégios garantidos aos agentes políticos, igualando os deveres trabalhistas de um deputado aos de um trabalhador civil.

Background

Inspirado na constante revolta popular do povo civil brasileiro perante as práticas de inúmeras figuras políticas ao longo da história democrática do país, o tema busca explorar o quão desigual é a realidade dos agentes políticos da federação, não se restringindo somente aos ganhos financeiros que conforme o Decreto Legislativo 172/2022, em vigor desde o dia 1º de fevereiro de 2025, totalizam R\$ 46.366,19 brutos mensais, as métricas principais dirão respeito ao trabalho de cada um dos deputados, avaliando quantos dos que exercem mandatos no presente momento comparecem às sessões legislativas da câmara e quantos destes participam efetivamente de votações realizadas em dias de trabalho, filtrando deputados que vão a sessões, mas não exercem seus deveres como representantes daqueles que os elegem.

Objetivos de negócio e critérios de sucesso

Tem-se como objetivos principais mensurar a assiduidade dos deputados federais atualmente com mandato e avaliar o comprometimento destes com seu papel dentro do regime democrático brasileiro, explorando casos de deputados que possam por quaisquer motivos se abster de uma parte significativa de votações. Para que sejam efetivamente cumpridos tais objetivos, serão mensurados alguns dados, tais quais:

- Percentual de presença de cada deputado;
- Percentual de participações em votações de cada deputado;

- Relação do percentual de presença pelo percentual votações de cada deputado;
- Número de deputados que ultrapassariam os limites de faltas permitidas pelo regime CLT.

Para além das métricas anteriores, também será calculado o salário efetivo de cada deputado após a aplicação de filtros da CLT, considerando como faltas todas aquelas justificadas que excedam o limite de faltas justificadas por ano delimitado pelo que determina a CLT, a partir do número obtido de presenças conforme os parâmetros estabelecidos será calculado o salário com desconto de faltas deste deputado. O mesmo será feito para o número de votações participadas por cada deputado, relacionando salário com número de votações.

Tendo em vista diferenciar deputados com alto impacto externo, para além do impacto político dentro da Câmara dos Deputados, também será calculado o número de dias que um deputado esteve em compromissos relacionados ao mandato (reuniões, comissões, agendas externas, etc.), mas que não impactarão em nenhuma das métricas estabelecidas anteriormente.

Inventário de recursos

Para que seja respondido o problema, serão coletados, filtrados por categorias de interesse e, caso necessário, tratados os dados fornecidos oficialmente pelo Estado brasileiro a respeito das legislaturas de cada um dos deputados. Para maior segurança na obtenção e armazenamento dos dados, serão utilizadas as APIs fornecidas pelo próprio governo.

Requisitos, suposições e restrições

Para a realização do projeto, têm-se os seguintes requisitos: coleta e organização dos dados oficiais de presenças, participações em votações e faltas dos deputados federais; Aplicar filtros e métricas baseadas no regimento da Consolidação das Leis Trabalhistas (CLT) visando simular a frequência requerida de um trabalhador formal; Calcular indicadores de desempenho, como o salário efetivo; Produzir relatórios e demonstrações gráficas que evidenciem a desigualdade entre agentes políticos e agentes civis, assim como demonstrar casos extremos nos dados.

São assumidas as seguintes suposições a respeito da natureza dos dados e para a metodologia utilizada: Os dados fornecidos oficialmente pelo governo são corretos, completos e confiáveis; Somente dias de sessões deliberativas e votações são considerados dias de trabalho, desconsiderando dias em que deputados estão em outras atividades políticas ligadas ao mandato, como agendas externas; Para as justificativas, somente licenças por interesse particular são consideradas faltas; A simulação do regime CLT é válida para comparação, mesmo sabendo que deputados têm funções e jornadas distintas de trabalhadores formais.

Têm-se como restrições do método aplicado: Somente dados públicos oficiais são utilizados, nenhum outro dado para além destes é tomado como parâmetro; Foco em métricas quantitativas (presenças, faltas, salário, participações, etc.), sem análise qualitativa do exercício de cada mandato ou dos projetos votados; Dada a natureza pouco explícita das licenças por interesse particular e à ausência de justificativas formais equivalente às de doença, morte de familiar e licença maternidade da CLT, não é possível avaliar os casos em que deputados se ausentam pelos motivos citados, o mesmo se aplica para afastamentos judiciais não registrados oficialmente.

Terminologia

Para que não existam ambiguidades, nomenclaturas ou conceitos desconhecidos ao leitor, dispõe-se os conceitos necessários abaixo:

- Deputado Federal: Parlamentar eleito para atuar no Congresso Nacional, responsável por criar e votar leis de abrangência nacional.
- Sessão Deliberativa: Reunião oficial da Câmara dos Deputados onde são debatidas e votadas propostas legislativas.
- Presença em Sessão: Registro de que o deputado participou de uma sessão deliberativa.
- Falta Justificada: Ausência em uma sessão ou votação registrada oficialmente, com justificativa aceita (ex.: licença por interesse particular ou missão oficial).
- Missão Oficial: Atividade relacionada ao mandato do deputado, como viagens nacionais ou internacionais, participação em comissões ou eventos representando o país.
- Licença por Interesse Particular: Direito do deputado de se ausentar de sessões e votações por motivos pessoais, sem perda de remuneração.
- Ganho por Dia Trabalhado: Indicador calculado dividindo-se a remuneração total bruta do deputado pelo número de presenças em sessões deliberativas.
- Ganho por Votação Participada: Indicador calculado dividindo-se a remuneração total bruta do deputado pelo número de votações participadas.
- CLT: Consolidação das Leis de Trabalho, conjunto de normas que rege os direitos e deveres de trabalhadores formais no Brasil.

Objetivos de mineração e critérios de sucesso

A mineração de dados para este projeto tem como objetivo extrair, organizar e analisar informações sobre presenças, participações em votações e faltas de deputados federais a partir de dados oficiais disponibilizados pela Câmara dos Deputados. As análises visam:

1. Calcular a presença de cada deputado;
2. Calcular a taxa de votações participadas por cada deputado;
3. Aplicar filtros baseados no regime CLT para simular limites de faltas e jornadas de trabalho;
4. Calcular o ganho efetivo por dia trabalhado e por votação participada por cada deputado;
5. Identificar padrões e evidências de desigualdade entre agentes políticos e agentes civis.

Para que haja sucesso na mineração dos dados, estabelecem-se os critérios:

- Dados corretamente coletados, tratados e organizados de forma consistente;
- Métricas calculadas de maneira precisa e reproduzível;
- Resultados que permitam comparar quantitativamente o desempenho dos deputados com parâmetros da CLT;
- Visualizações e relatórios claros, que evidenciem as conclusões desejadas e permitam confirmar ou refutar as hipóteses iniciais.

Avaliação inicial de técnicas e ferramentas

Serão majoritariamente utilizados recursos computacionais para coleta, filtragem, tratamento visualizações dos dados necessários para a realização da pesquisa, são eles, de forma geral:

- Linguagem de programação Python para coleta, filtragem e tratamento de dados;
- Application Programming Interface (API) fornecida oficialmente pelo governo;
- Power BI ou ferramenta similar para construção de visualizações gráficas dos dados obtidos.

2. Autocrítica

Até o presente momento pouco foi realizado para o desenvolvimento do projeto, isso se deve ao fato de que tardiamente foi decidido a linha a ser seguida na pesquisa. Diferentes ideias foram propostas desde o início do desenvolvimento e, por fim, foi escolhido o tema apresentado no tópico anterior, já que dentre os propostos o presente seria o menos sujeito a barreiras que não possibilitariam uma conclusão alinhada aos objetivos. Apesar do pouco desenvolvimento, o grupo se encontra satisfeito com o tema escolhido, já que será possível explorar uma problemática conhecida do país, a dos privilégios da vida política, nos aspectos financeiros e outros que regem a qualidade de vida de toda e qualquer pessoa, como o tempo livre para dedicar a outras atividades além do trabalho.

Após algumas análises baseadas nos dados disponíveis e, sabendo quais são, acredita-se que há uma aderência considerável à metodologia CRISP-DM. A metodologia além de possibilitar uma maior organização em etapas, seu aspecto recursivo possibilita o levantamento de novas perguntas conforme o andamento da coleta e análise de dados, possibilitando uma análise mais completa da situação política no Brasil.

A primeira etapa do projeto pôde fornecer ao grupo uma maior noção de como uma pesquisa na área da ciência de dados pode ser feita, de aspectos mais teóricos, como o da organização e adequação à metodologia CRISP-DM e entender como, a partir de um dataset, um problema pode ser explorado e, caso não possa ser respondido apropriadamente, adaptado em alguns aspectos almejando evitar o desenvolvimento de uma pesquisa a partir de um problema que não pode ser respondido com os dados e fontes disponíveis. No grupo foi determinada uma nota 8 para o projeto, isso se deve a alguns alinhamentos ainda não realizados que dizem respeito às restrições da pesquisa, serão elaboradas maneiras de determinar uma maneira mais justa para que seja feita a análise das faltas justificadas de deputados, se possível serão desenvolvidos métodos para identificar se um deputado falta por uma justificativa plausível, tal qual uma justificativa por doença ou procedimento cirúrgico debilitante, porém será necessária uma análise mais aprofundada dos dados disponíveis.