

DESENVOLVIDO POR: André Carvalho, Carlos Alberto Braga, Fernanda de Souza Borges, Letícia Araújo Costa, Pedro Mello Antunes

INTRODUÇÃO

Esta Nota Técnica tem como objetivo consolidar a análise e a documentação de conjuntos de dados relacionados à arrecadação tributária, considerando o uso dessas informações em projetos de aprendizado profundo (Deep Learning). O estudo foi elaborado a partir de dois documentos principais, contendo tanto informações técnicas sobre modelagem e organização de dados quanto uma visão geral das arrecadações tributárias federais e estaduais nos anos de 2021, 2022 e 2023.

Os dados analisados incluem séries históricas anuais de arrecadação federal e dados desagregados por estado, permitindo uma visão integrada da evolução das receitas tributárias e da sua distribuição geográfica. A organização desses conjuntos de dados possibilitou análises comparativas entre diferentes anos e estados, destacando padrões relevantes e tendências nas arrecadações.

A motivação para esta análise é a necessidade crescente de uso de tecnologias avançadas, como aprendizado profundo, para extrair insights mais ricos a partir de grandes volumes de dados fiscais. Esse tipo de abordagem não apenas permite maior eficiência na análise de dados, mas também auxilia na tomada de decisões estratégicas por investidores públicos e na formulação de políticas tributárias baseadas em evidências.

Além disso, este estudo considera a importância da padronização e da clareza na apresentação dos dados, alinhando-se às boas práticas de gestão de informações fiscais. A transposição e organização dos conjuntos de dados, assim como a aplicação de métodos de tratamento de dados, garantem maior integridade e confiabilidade dos resultados apresentados.

Por fim, uma introdução de técnicas de visualização de dados, como gráficos de barras e comparações temporais, reforça o entendimento das informações apresentadas e destaca a relevância de práticas modernas de análise de dados para a administração pública e para estudos econômicos. Este documento, portanto, oferece uma base sólida para a análise tributária, considerando tanto os desafios técnicos quanto as oportunidades oferecidas pela modernização dos sistemas de informação e análise fiscal.

DADOS UTILIZADOS

Para a análise e desenvolvimento desta Nota Técnica, foram utilizados quatro conjuntos de dados principais que abrangem informações relacionadas à arrecadação tributária federal e estadual, organizados conforme descrito a seguir:

1.serie_historica_2021

- **Descrição:** Este conjunto de dados contém uma série histórica de arrecadação das receitas federais para o ano de 2021. Os dados incluem informações incluídas sobre diferentes categorias de tributos, como impostos diretos e indiretos, contribuições previdenciárias e receitas administradas por outros órgãos.
- **Tratamento aplicado:**
 - Os dados originais foram transpostos, reorganizando as linhas e colunas para uma

melhor estruturação.

- A primeira linha do arquivo foi utilizada como referência para renomear as colunas, garantindo consistência nos nomes das variáveis.

2.serie_historica_2022

- **Descrição:** Conjunto de dados que registra a arrecadação federal do ano de 2022, com a mesma estrutura e categorias de dados do ano anterior.
- **Tratamento aplicado:**
 - Assim como no conjunto de dados de 2021, os dados foram transpostos para adequação à análise.
 - Os títulos das colunas foram atualizados com base na primeira linha do arquivo, garantindo uniformidade nos campos.

3.serie_historica_2023

- **Descrição:** Este conjunto de dados apresenta os dados referentes à arrecadação federal do ano de 2023, com informações detalhadas sobre diferentes tributos e contribuições.
- **Tratamento aplicado:**
 - A transposição dos dados foi realizada para padronizar a estrutura e alinhar com os conjuntos de dados de anos anteriores.
 - As colunas foram renomeadas de acordo com os cabeçalhos encontrados na primeira linha do arquivo.

4.arrecadacao_estado

- **Descrição:** Conjunto de dados que contém informações sobre a arrecadação tributária desagregada por estado (UF), permitindo uma análise detalhada da distribuição geográfica das receitas.
- **Observações:**
 - Este arquivo inclui informações tanto de arrecadação previdenciária quanto de tributos específicos, categorizados por tipo de imposto.
 - As colunas numéricas foram convertidas para o tipo **float**, o que permitiu cálculos precisos e análises estatísticas consistentes.
 - Apenas os dados relativos aos anos de 2021, 2022 e 2023 foram considerados, excluindo-se os períodos anteriores ou irrelevantes.

Relações Entre os Datasets

O vínculo principal entre os conjuntos de dados de arrecadação federal (séries históricas) e o conjunto de dados de arrecadação estadual foi estabelecido por meio da variável "Ano". Este relacionamento possibilitou análises integradas, incluindo:

- Comparações entre a arrecadação total nacional e a arrecadação individual por estado em diferentes anos;
 - Identificação de estados que se destacam, positivos ou negativos, em relação à média nacional de arrecadação;
 - Estudo do impacto relativo aos diferentes tipos de tributos na arrecadação estadual e federal.
-

Dicionário de Dados

Os conjuntos de dados apresentam categorias desenvolvidas para cada tipo de tributo e contribuição, organizadas da seguinte forma:

1. Impostos Federais Diretos e Indiretos:

- Exemplos: IPI (Imposto sobre Produtos Industrializados), IRPF (Imposto de Renda Pessoa Física), IRPJ (Imposto de Renda Pessoa Jurídica), IOF (Imposto sobre Operações Financeiras), entre outros.
- Cada tributo é segmentado em subcategorias, como IPI sobre fumo, bebidas e automóveis.

2. Contribuições Previdenciárias:

- Inclui receitas relacionadas à segurança social, com separações entre contribuições administradas pela Receita Federal e aquelas administradas por outros órgãos.

3. Receitas Administradas por Outros Órgãos:

- São consideradas receitas não tributárias ou específicas, como as receitas de fundos especiais e programas de recuperação fiscal (REFIS e PAES).

4. Dados por Estado:

- Para cada Unidade Federativa (UF), estão incluídas informações planejadas sobre a arrecadação mensal, permitindo análises temporais e regionais.
-

Com essas informações, os conjuntos de dados fornecem uma base robusta para análises quantitativas, como a identificação de padrões regionais, tendências ao longo do tempo e impactos fiscais de diferentes categorias de tributos.

METODOLOGIA

A metodologia adotada para esta análise foi estruturada em várias etapas, contemplando o tratamento dos dados, organização, modelagem, análise estatística e visualização, de modo a garantir uma abordagem consistente e replicável. Abaixo, são detalhadas as etapas realizadas:

1. Organização e Estruturação dos Conjuntos de Dados

● Transposição e Reformatação:

Os conjuntos de dados de séries históricas ([serie_historica_2021](#), [serie_historica_2022](#) e [serie_historica_2023](#)) foram inicialmente transpostos para adequar os dados ao formato necessário para análise. Essa etapa foi essencial, pois os dados originais foram estruturados com os anos e categorias de arrecadação desejada em linhas, sendo reorganizados para que essas informações passassem a compor colunas.

- Após a transposição, a primeira linha de cada conjunto de dados, contendo os nomes das variáveis, foi utilizada para renomear as colunas, garantindo uniformidade e clareza nos campos.
- Dados redundantes ou inconsistentes foram excluídos durante essa etapa.

● Filtragem Temporal:

Apenas dados dos anos de 2021, 2022 e 2023 foram considerados. Para o conjunto de dados [arrecadacao_estado](#), uma filtragem incluída colunas numéricas representando diferentes tipos de tributos, e as variáveis categóricas como "Ano", "Mês" e "UF" (Unidade

Federativa) foram mantidas para análise descritiva e relacional.

2. Tratamento dos Dados

- **Conversão de Tipos de Dados:**

- Foi desenvolvida uma função personalizada para converter os valores nas células dos conjuntos de dados em números do tipo `float`, garantindo que qualquer valor não numérico fosse substituído por valores nulos (`NaN`) ou um padrão adequado.
- No conjunto de dados `arrecadacao_estado`, todas as colunas numéricas, exceto as variáveis categóricas "Mês" e "UF", foram convertidas usando a função `pd.to_numeric`, com tratamento para erros.

- **Verificação e Correção de Dados Faltantes:**

- Uma análise detalhada foi realizada para identificar valores ausentes (`NaN`) em cada conjunto de dados.
 - Medidas foram tomadas para preencher valores faltantes de forma consistente, utilizando a média ou mediana, dependendo da categoria, para evitar distorções nos resultados.
-

3. Modelagem e Integração dos Dados

- **Relacionamento entre conjuntos de dados:**

- A integração entre os conjuntos de dados foi realizada utilizando a coluna "Ano", que serviu como chave para vincular os dados de arrecadação por estado com as séries históricas federais.
- Essa integração incluiu análises comparativas entre a arrecadação total nacional e a arrecadação específica de cada estado, além de possibilitar a análise de padrões regionais e temporais.

- **Criação de Novas Variáveis:**

- No dataset `arrecadacao_estado`, foi implementada a variável "Soma Total", que agrega os valores de arrecadação de todas as colunas numéricas para cada combinação de "Ano" e "UF".
 - Essa variável foi essencial para resumir os dados e facilitar a análise de tendências.
-

4. Análise de Dados

- **Exploração Descritiva:**

- Foi realizada uma análise inicial dos dados, incluindo estatísticas descritivas como médias, medianas, totais acumulados e variações ao longo do tempo.
- Comparações entre estados e anos foram realizadas para identificar outliers e padrões relevantes.

- **Separação por Categorias de Impostos:**

- Os tributos foram divididos entre diretos e indiretos, e os totais de arrecadação de cada categoria foram calculados.
- Essa divisão facilitou a compreensão de como cada tipo de imposto contribui para a arrecadação geral em diferentes anos e regiões.

5. Visualização dos Dados

- **Construção de Gráficos e Relatórios Visuais:**

- Gráficos de barras foram gerados para comparar a arrecadação total entre os anos de 2021, 2022 e 2023. Esses gráficos evidenciaram tendências gerais, como crescimento ou redução das receitas fiscais.
- A arrecadação por estado foi representada graficamente para destacar as diferenças regionais, permitindo análises previstas sobre a contribuição individual de cada Unidade Federativa.
- Um gráfico específico foi elaborado para comparar os valores arrecadados de impostos diretos e indiretos, ajudando a entender a dinâmica fiscal do país.

- **Análise Mensal:**

- Um gráfico categorizado por meses foi gerado utilizando a biblioteca Seaborn, permitindo observar a sazonalidade das arrecadações. Essa análise revelou padrões relacionados a picos de arrecadação em determinados períodos do ano, como janeiro e meses trimestrais.
-

6. Críticas e Avaliações

- A metodologia incluiu uma etapa de avaliação crítica dos resultados obtidos, buscando interpretar os padrões de arrecadação observados e correlacioná-los com fatores econômicos e administrativos.
 - Foram identificados gargalos no processo de coleta e tratamento de dados, bem como sugestões para melhorias, como a implementação de automatizações no fluxo de dados e maior integração de tecnologias avançadas para análise preditiva.
-

Nessas etapas, você garantiu a integridade dos dados analisados, permitindo a obtenção de insights significativos para a formulação de políticas públicas e a melhoria dos processos de gestão tributária.

RESULTADOS

A análise realizada com os conjuntos de dados fornecidos gerou uma série de resultados que destacam tendências importantes na arrecadação tributária nacional e estadual, além de revelar padrões específicos relacionados a diferentes tipos de tributos e comportamentos regionais e temporais. Estes resultados são organizados em detalhes a seguir:

1. Evolução das Receitas Federais (2021-2023)

- A arrecadação total federal apresentou um crescimento consistente ao longo do período desenvolvido, com variações anuais influenciadas por fatores econômicos, políticas tributárias e condições macroeconômicas, como inflação, câmbio e preço de commodities.
- Os valores arrecadados em 2021 ainda refletem os efeitos da retomada econômica pós-pandemia, com um leve crescimento em comparação com 2020. Já os anos de 2022 e

2023 mostraram aumentos mais expressivos devido à normalização de atividades econômicas e ao fortalecimento de setores industriais e de serviços.

- Alguns tributos, como o Imposto sobre Produtos Industrializados (IPI) e o Imposto de Renda Retido na Fonte (IRRF), tiveram aumento significativo na arrecadação, refletindo tanto o aumento do consumo de bens como o reaquecimento do mercado de trabalho formal.
-

2. Distribuição Geográfica da Arrecadação por Estado

- As análises regionais indicaram diferenças significativas na arrecadação entre as Unidades Federativas (UFs). Estados como São Paulo, Minas Gerais e Rio de Janeiro continuam liderando em termos de arrecadação, devido à sua maior concentração de atividade econômica, industrial e populacional.
 - Alguns estados da Região Norte e Nordeste receberam arrecadação proporcionalmente inferior à média nacional, o que destaca a necessidade de políticas regionais para estimular o desenvolvimento econômico e melhorar a arrecadação tributária.
 - A arrecadação previdenciária, em particular, apresentou um padrão de concentração em estados com maior número de trabalhadores formais e maiores massas salariais, como São Paulo e Minas Gerais.
-

3. Comparação entre Impostos Diretos e Indiretos

- **Impostos Diretos:**
 - Representaram a maior parte da arrecadação total em todos os anos analisados. Isso inclui tributos como o Imposto de Renda (IR) e a Contribuição Social sobre o Lucro Líquido (CSLL).
 - Os impostos diretos refletem o desempenho das empresas e o rendimento das pessoas físicas, o que os torna indicadores sensíveis ao crescimento econômico.
- **Impostos Indiretos:**
 - Inclui tributos como o IPI e o Imposto sobre Operações Financeiras (IOF). Sua arrecadação apresentou variações mais sensíveis ao consumo e à atividade industrial.
 - Observamos um crescimento significativo na arrecadação de impostos indiretos nos anos de 2022 e 2023, impulsionado pelo aumento do consumo e da produção industrial.

A comparação gráfica evidenciou que os impostos indiretos têm uma contribuição mais estável, enquanto os impostos indiretos são mais suscetíveis a flutuações econômicas.

4. Tendências Mensais na Arrecadação

- A análise mensal revelou picos de arrecadação em períodos específicos do ano, como janeiro, que coincidem com o prazo de prazos para declarações fiscais, e meses relacionados ao fechamento de trimestres fiscais, como março, junho, setembro e dezembro.
- Esses picos estão diretamente associados ao calendário fiscal e à sazonalidade econômica, refletindo atividades específicas, como pagamento de tributos trimestrais e fechamento de contas de empresas.
- A análise também destacou a sazonalidade na arrecadação previdenciária, com aumentos em

períodos de ajustes anuais, como dezembro, quando as empresas contabilizam e cobram contribuições adicionais.

5. Análise da Arrecadação por Tipo de Tributo

- Os tributos que mais desenvolvem para a arrecadação total incluem:
 - **Imposto de Renda Pessoa Física (IRPF):** Com crescimento consistente devido à retomada do mercado de trabalho e ao aumento da formalização.
 - **Contribuições Previdenciárias:** Fortemente especificações em estados mais industrializados e com maior número de trabalhadores formais.
 - **IPI:** Flutuações graves devido à variação na produção industrial, especialmente nos setores de bebidas e automóveis.
 - **COFINS e PIS/PASEP:** Representaram uma parcela relevante da arrecadação, mostrando crescimento constante devido à sua influência sobre o faturamento das empresas.
 - Tributos menores, como o Imposto sobre Importação, também apresentaram crescimento devido à alta do dólar e ao aumento da importação de produtos em determinados períodos.
-

6. Padrões Regionais e Econômicos

- Os estados com economias mais planejadas incluíram maior arrecadação de impostos indiretos, enquanto os estados dependentes de atividades específicas, como agricultura ou mineração, tiveram maior arrecadação de equipamentos em tributos relacionados a exportações e atividades locais.
 - A análise regional também destacou o impacto dos incentivos fiscais, que podem ter limitado a arrecadação de alguns estados, ao mesmo tempo em que estimulam setores específicos da economia.
-

7. Visualização dos Resultados

- Gráficos de barras foram utilizados para ilustrar as tendências anuais de arrecadação, evidenciando o crescimento contínuo entre os anos de 2021 e 2023.
 - Comparações entre arrecadação estadual e federal foram apresentadas para destacar diferenças regionais, mostrando a relevância de estados como São Paulo, Minas Gerais e Rio de Janeiro.
 - Análises específicas sobre tributos, como impostos diretos e indiretos, foram visualizadas para mostrar sua contribuição relativa e as variações temporais.
-

Sintético

Os resultados indicam que a arrecadação tributária no Brasil reflete tanto as dinâmicas macroeconômicas quanto as especificidades regionais e setoriais. Essa análise oferece insights valiosos para gestores públicos, destacando áreas prioritárias para políticas públicas e ações externas ao equilíbrio fiscal e à eficiência tributária.

CONCLUSÕES

Uma análise detalhada dos conjuntos de dados relacionados à arrecadação tributária federal e estadual permitiu a identificação de padrões relevantes, desafios e oportunidades para a gestão fiscal e tributária no Brasil. As conclusões são apresentadas em tópicos para maior clareza e objetividade:

1. Crescimento Consistente da Arrecadação Federal

Os dados mostram um crescimento contínuo da arrecadação tributária federal entre 2021 e 2023. Isso reflete a retomada econômica pós-pandemia, marcada por:

- A normalização das atividades econômicas após o impacto inicial da COVID-19.
- O aumento do consumo e da produção industrial, especialmente nos setores como automóveis, bebidas e eletrodomésticos, que desenvolve para o incremento de impostos como IPI e ICMS.
- O crescimento da massa salarial e da formalização do trabalho, que elevou a arrecadação de contribuições previdenciárias e do Imposto de Renda Retido na Fonte (IRRF).

Este cenário demonstra a resiliência da economia brasileira e a eficácia de algumas políticas econômicas inovadoras no período, embora o ritmo de crescimento varie entre regiões e setores.

2. Disparidades Regionais na Arrecadação

As análises regionais destacaram desigualdades significativas entre as Unidades Federativas (UFs) em termos de arrecadação tributária. Estados como São Paulo, Minas Gerais e Rio de Janeiro concentraram a maior parte das receitas, devido a fatores como:

- Alta densidade populacional e industrial.
- Diversificação econômica e maior formalização das atividades econômicas.

Por outro lado, os estados das regiões Norte e Nordeste tiveram arrecadação proporcionalmente inferior à média nacional. Esse padrão reflete desafios estruturais, como menor industrialização, dependência de atividades agropecuárias e baixo nível de formalidade. Essa disparidade reforça a necessidade de políticas públicas regionais para estimular o desenvolvimento econômico, reduzir a desigualdade fiscal e promover maior equilíbrio na distribuição de receitas.

3. Contribuição dos Impostos Diretos e Indiretos

A análise evidenciou que:

- **Os impostos diretos**, como o Imposto de Renda (IRPF e IRPJ) e a Contribuição Social sobre o Lucro Líquido (CSLL), representam uma parcela significativa e estável da arrecadação total. Eles estão diretamente relacionados ao desempenho econômico, especialmente da classe trabalhadora e das empresas.
- **Os impostos indiretos**, como o IPI e o IOF, apresentaram maior volatilidade, influenciados por flutuações no consumo e na produção industrial.

A distinção entre essas duas categorias de tributos é crucial para entender como a arrecadação responde a diferentes fatores econômicos e para avaliar o impacto de políticas tributárias em setores específicos da economia.

4. Sazonalidade da Arrecadação

A análise mensal revelou padrões sazonais previsíveis, com picos de arrecadação concentrados em períodos estratégicos, como:

- **Janeiro:** Decorrente do encerramento de prazos de recolhimento de tributos anuais e declarações fiscais.
- **Meses trimestrais (março, junho, setembro e dezembro):** Devido ao fechamento de contas e recolhimento de tributos trimestrais.
- **Dezembro:** Destacado pela regularização de pendências fiscais e recolhimentos extras relacionados à massa salarial do fim do ano (décimo terceiro salário).

Esses padrões ajudam a fluxos de caixa planejados e estratégias para a administração pública, permitindo maior previsibilidade nos ciclos fiscais.

5. Importância das Contribuições Previdenciárias

As contribuições previdenciárias representaram uma parcela significativa da arrecadação, refletindo diretamente a formalização do mercado de trabalho e da base salarial. Estados com maior concentração de empregos formais, como São Paulo e Minas Gerais, tiveram destaque.

Esse dado reforça a necessidade de:

- **Estimular a formalização do trabalho em estados com arrecadação previdenciária baixa**, como os do Norte e Nordeste.
 - **Monitorar o impacto de alterações legislativas ou políticas tributárias** que possam afetar a base de arrecadação previdenciária.
-

6. Impacto de Políticas Econômicas e Tributárias

Alguns resultados apontam para a influência direta de políticas econômicas sobre a arrecadação, como:

- Reduções temporárias no IPI para estimular o consumo em momentos de crise econômica.
- Incentivos fiscais que, embora atrativos para setores específicos, podem gerar perda de arrecadação em estados menos desenvolvidos.

É essencial avaliar o custo-benefício dessas políticas, considerando tanto os efeitos econômicos de curto prazo quanto o impacto fiscal de longo prazo.

7. Necessidade de Modernização e Integração Tecnológica

A análise revelou a importância de processos automatizados para coleta, tratamento e análise de dados fiscais. A transposição manual de conjuntos de dados e a padronização das variáveis, embora eficaz para este estudo, expõe a necessidade de:

- **Plataformas integradas** para gerenciamento de dados tributários.
 - **Ferramentas de aprendizado de máquina** para identificar padrões em grandes volumes de dados.
 - **Sistemas mais transparentes e acessíveis** para estados e gestores públicos, permitindo tomadas de decisão baseadas em dados concretos.
-

8. Insights Estratégicos para Políticas Públicas

As tendências observadas sugerem que as políticas públicas podem ser otimizadas para:

- **Reduzir desigualdades regionais:** Estimulando o desenvolvimento econômico em estados com baixa arrecadação.
 - **Melhorar a comunicação e a transparência tributária:** Facilitando o cumprimento das obrigações fiscais por empresas e cidadãos.
 - **Fomentar a automação e modernização tecnológica:** Investindo em soluções digitais para maximizar a eficiência da administração tributária.
-

Síntese das Conclusões

O estudo destaca a robustez do sistema de arrecadação tributária no Brasil, mas também expõe desafios significativos, como desigualdades regionais, sazonalidades marcantes e a necessidade de modernização tecnológica. A análise reafirma a importância de alinhar políticas fiscais a objetivos econômicos e sociais mais amplos, promovendo maior equilíbrio e eficiência na arrecadação tributária nacional e estadual.