

Evolução da Dívida, Investimentos e Renúncia Fiscal do Brasil entre 2014 e 2024

Bernardo Moro, Eduardo Stefanello, Gabriel Silveira

16 de junho de 2025

Resumo

O trabalho buscou encontrar uma correlação comparativa entre a evolução da dívida pública brasileira, dos investimentos públicos da União e da renúncia fiscal de tributos.

1 DIAGNÓSTICO E TEORIZAÇÃO

1.1 Identificação das partes interessadas e parceiros

As partes interessadas são os acadêmicos do curso de Big Data em Python, bem como a sociedade em geral, tendo em vista o interesse público nas análises de dados abertos governamentais que serão processados.

1.2 Problemática e/ou problemas identificados

Há anos a dívida pública do Brasil cresce. O objetivo desse trabalho é analisar o quanto a dívida pública poderia diminuir ou deixar de crescer se as renúncias fiscais do governo federal fossem controladas. Esse problema já foi analisado pelo Ministério da Economia em 2023 [3].

1.3 Justificativa

Uma vez que a disciplina é sobre Big Data, o que envolve a análise de grandes conjuntos de dados, analisar vastos dados econômicos governamentais ao longo de um determinado período parece pertinente ao objetivo deste trabalho.

1.4 Objetivos/resultados/efeitos a serem alcançados

1. Analisar grandes conjuntos de dados utilizando ferramentas computacionais.
2. Entender as ferramentas estatísticas que fornecerão inferências úteis.
3. Programar algoritmos em Python envolvendo ferramentas de Big Data.

1.5 Referencial teórico (subsídio teórico para propositura de ações da extensão)

2 PLANEJAMENTO E DESENVOLVIMENTO DO PROJETO

2.1 Plano de trabalho (usando ferramenta acordada com o docente)

A primeira etapa será definir os datasets utilizados para o processamento de dados. Isso deverá ser concluído até o dia 26 de maio em aula.

A segunda etapa será a normalização dos dados e adaptação para uso na ferramenta computacional utilizada. Isso deverá ser concluído até o dia 2 de junho, e cada integrante deverá programar a importação desses dados para a ferramenta.

A terceira etapa será a implementação de algoritmos de Machine Learning e a utilização da ferramenta Spark para processar a totalidade dos dados. Isso deverá ser concluído até o dia 16 de junho, em conjunto de forma online e síncrona.

A quarta e última será compilar os resultados para apresentar. Isso será feito até o dia 23 de junho, de forma online e assíncrona;

2.2 Descrição da forma de envolvimento do público participante na formulação do projeto, seu desenvolvimento e avaliação, bem como as estratégias pelo grupo para mobilizá-los.

Como os dados são públicos e abertos, o poder público atua de forma a disponibilizar os dados. Os estudantes do grupo atuam de forma a buscar nas bases públicas, processar e analisar os dados e apresentá-los à turma, que representarão a sociedade neste trabalho.

2.3 Grupo de trabalho (descrição da responsabilidade de cada membro)

Os integrantes do grupo são: Eduardo Stefanello, Gabriel Silveira e Bernardo Moro. As atribuições do Eduardo são: buscar e fazer o download dos dados públicos. Gabriel irá tratar os dados para processamento, para depois serem processados pelo grupo, com implementações em conjunto dos algoritmos de Machine Learning. Bernardo ficará responsável por compilar os dados e prepará-los para a apresentação.

2.4 Metas, critérios ou indicadores de avaliação do projeto

1. Será necessário fazer o download e o pré-processamento dos dados para eles poderem ser processados pelos algoritmos que serão implementados pelo grupo. Também será necessário instalar o Spark na ferramenta computacional e configurar o sistema. Spark é uma importante ferramenta em Big Data:

A utilização do Apache Spark tem se mostrado crucial para o processamento distribuído em ambientes de big data, pois permite a execução de tarefas de forma ágil e eficiente, otimizando a análise de grandes volumes de dados [9].

2. Iremos definir quais correlações de dados queremos fazer e que informações extrair do conjunto, pesquisar ferramentas estatísticas adequadas para essas análises.
3. Para este objetivo será necessário entender como funcionam os algoritmos de machine learning e como eles podem ser implementados em Python, quais bibliotecas utilizar e como entender o resultado recebido. Será utilizado o Python, por ser o objetivo da disciplina e também o pandas:

"O uso do pandas transforma o pré-processamento de dados em um processo ágil e eficiente, permitindo que o foco seja direcionado para a análise e interpretação dos resultados." [7]

2.5 Recursos previstos

Os recursos previstos serão os arquivos de dados, disponibilizados pelo poder público através de portais da internet.

Também haverá o recurso computacional, que serão o Databricks [4] e AWS Academy, definidos pelo professor.

Os recursos humanos serão os integrantes do grupo, além do professor que auxiliará no uso das diversas ferramentas [5].

2.6 Detalhamento técnico do projeto

• Ferramentas de desenvolvimento:

1. Google Collab (notebooks Jupyter);

2. Databricks (plataforma de análise de dados);
- **Ferramentas de programação em Python:**
 1. Pandas (biblioteca de manipulação de dados);
 2. Numpy (biblioteca para cálculos);
 3. Ploty (biblioteca para geração de gráficos);
 4. PySpark (biblioteca para paralelização de tarefas);
 5. Prophet (algoritmo de machine learning para predição de dados);
- **Ferramentas de inteligência artificial:**
 1. Chat GPT e Gemini (Chat com LLM);
 2. Notebook LM (Notebook com LLM);
- **Ferramentas organizacionais:**
 1. GitHub (Controle de código fonte);
 2. Trello (Organização de tarefas);
 3. Overleaf (Escrita de documentos acadêmicos)
 4. Gamma (Apresentações de slides)

3 ENCERRAMENTO DO PROJETO

3.1 Relato Coletivo

Durante o desenvolvimento de um projeto de análise de dados públicos, a equipe concentrou esforços na investigação de investimentos, dívida pública e renúncia fiscal no Brasil, abrangendo o período de 2014 a 2024.

Gabriel Silveira foi responsável pelo tratamento e modelagem preditiva dos dados utilizando PySpark no Databricks, aplicando Regressão Linear e Prophet, o que lhe permitiu identificar padrões como o aumento contínuo da dívida e a recuperação dos investimentos, mesmo enfrentando desafios com a integração e modelagem preditiva.

Eduardo Stefanello contribuiu significativamente na etapa de planejamento da extração e tratamento dos datasets, executando a limpeza, transformação e padronização dos dados, além de revisar e aprimorar o código-fonte em parceria com Gabriel, adotando uma abordagem colaborativa e estratégica que garantiu a qualidade analítica dos resultados.

Por sua vez, Bernardo Moro atuou na organização e comunicação dos progressos por meio do Trello e preparação dos slides, consolidando o valor da comunicação clara e da gestão efetiva das informações.

A experiência integradora proporcionou um aprendizado prático sobre o ciclo completo de um projeto de Big Data, reforçando a importância da aplicação de técnicas avançadas para embasar decisões estratégicas no setor público e consolidando competências técnicas e colaborativas entre os membros da equipe.

3.2 Avaliação de reação da parte interessada

Os estudantes e o professor da turma de Big Data em Python ficaram impressionados com a evolução da dívida pública e também dos valores de renúncia fiscal, principalmente à partir de 2020. Segundo o relatório do Fundo Monetário Internacional (FMI, 2021), os pacotes de estímulo e as políticas de apoio à saúde e à renda contribuíram para a escalada da dívida pública, gerando desafios importantes para a sustentabilidade fiscal nos próximos anos [1].

Todos acharam as conclusões muito interessantes e entenderam os desafios enfrentados pelo poder público e pelas empresas. Uma análise do Instituto de Pesquisa Econômica Aplicada (IPEA, 2021) aponta que os incentivos fiscais direcionados às empresas brasileiras registraram um aumento significativo a partir da pandemia. Segundo o estudo, essas medidas de alívio fiscal foram implementadas como parte dos esforços para mitigar os impactos da crise econômica, permitindo às empresas maior acesso a recursos e incentivando a retomada das atividades produtivas [2].

4 Relato de Experiência Individual (Pontuação específica para o relato individual)

4.1 Gabriel Silveira

4.1.1 CONTEXTUALIZAÇÃO

Participei do desenvolvimento de um projeto de análise de dados públicos com foco em investimentos, dívida pública e renúncia fiscal no Brasil (2014–2024). O trabalho foi realizado utilizando ferramentas como PySpark no Databricks, com foco em limpeza, análise e modelagem preditiva dos dados.

4.1.2 METODOLOGIA

A experiência foi remota e realizada no ambiente Databricks Community. Trabalhei com quatro bases de dados .csv, realizando as seguintes etapas: Tratamento e transformação dos dados com PySpark; Visualizações com Plotly e gráficos interativos; Aplicação de Regressão Linear e Prophet para previsão de investimentos.

4.1.3 RESULTADOS E DISCUSSÃO

O projeto superou minhas expectativas. Descobri padrões como o aumento constante da dívida, a alta repentina da renúncia fiscal e a recuperação dos investimentos.

- **Facilidades:** uso de Python e Pandas.
- **Dificuldades** uso do Prophet e integração de arquivos.
- **Aprendizado** Entendi melhor o ciclo completo de um projeto de Big Data e como aplicar análises preditivas com dados reais.

4.1.4 REFLEXÃO APROFUNDADA

A prática me ajudou a consolidar a teoria. A aplicação do Prophet com variáveis externas foi um grande avanço na compreensão de modelos de séries temporais com múltiplos fatores.

4.1.5 CONSIDERAÇÕES FINAIS

Foi uma experiência enriquecedora. Como aprimoramento, o uso de pipelines automatizados e integração com BI em tempo real seriam boas evoluções. O projeto mostrou o valor da análise de dados para decisões públicas.

4.2 Eduardo Stefanello

4.2.1 CONTEXTUALIZAÇÃO

Minha participação no desenvolvimento deste projeto se deu principalmente na parte de planejamento das estratégias de extração dos dados, pesquisa de Datasets e o tratamento/limpeza dos dados. Após a maior parte do desenvolvimento ser executada pelo Gabriel, revisei com ele em aula o código-fonte escrito e os resultados das análises.

4.2.2 METODOLOGIA

1. A metodologia de pesquisa consistiu na coleta de dados por meio da pesquisa e extração de datasets disponíveis em sites oficiais de dados públicos do governo federal. Foi realizada uma busca criteriosa nesses portais para identificar, selecionar e baixar datasets relevantes para o objeto de estudo, garantindo a utilização de fontes confiáveis e atualizadas para a análise subsequente.
2. Em relação ao tratamento e pré-processamento dos dados, foram realizadas as seguintes etapas:
 - (a) Importação e carregamento dos datasets em dataframes: os dados foram lidos a partir de fontes externas e convertidos em dataframes para facilitar o processamento.

- (b) Processamento dos dataframes: foram realizadas operações de manipulação dos dados, reorganização de registros e transformação dos dados brutos em formatos mais adequados para análise.
 - (c) Formatação das colunas: cada coluna foi convertida para o tipo de dado mais apropriado (por exemplo, datas, numéricos ou strings) e padronizada, garantindo consistência nos valores.
 - (d) Remoção de dados inválidos: registros com valores nulos, inconsistentes ou fora do padrão esperado foram identificados e removidos (ou tratados) para evitar distorções na análise.
 - (e) Seleção de colunas úteis: após a limpeza e formatação, foi feita uma seleção das colunas que realmente agregariam valor à análise, descartando aquelas que não eram relevantes.
3. Na revisão do código fonte das análises, foram identificadas algumas inconsistências nos gráficos gerados e no código fonte desenvolvido. Em parceria com o Gabriel, fizemos a correção em aula, modificando a forma como os gráficos ficaram ao final do projeto e melhorando o código, que estava apresentando alguns erros de execução até aquele momento.

4.2.3 RESULTADOS E DISCUSSÃO

Os resultados obtidos evidenciam a eficácia das estratégias de extração e tratamento dos dados adotadas no projeto.

A colaboração entre os membros da equipe, com discussões em aula e revisões do código-fonte executado com o Gabriel, permitiu não apenas a validação das técnicas aplicadas, mas também a identificação de pontos de melhoria para futuras análises. Auxiliar o Bernardo me ajudou a entender a totalidade do que desenvolvemos.

Em síntese, nossos resultados demonstram que a combinação de um planejamento estratégico bem definido e uma abordagem metodológica rigorosa na manipulação dos dados pode gerar insights valiosos, reforçando a importância de cada etapa do processo analítico no âmbito do projeto.

4.2.4 REFLEXÃO APROFUNDADA

Durante o desenvolvimento do projeto, assumi um papel de liderança que foi fundamental para garantir a eficiência do grupo. Minha atuação envolveu a delegação das tarefas de forma equilibrada, garantindo que cada membro pudesse contribuir com suas habilidades específicas para o alcance dos objetivos propostos. Como o professor definiu tarefas organizacionais interessantes para o andamento do trabalho, foi possível delegar determinadas tarefas ao Bernardo e outras ao Gabriel.

Esse papel de líder não foi apenas na distribuição de tarefas, mas também na promoção de um ambiente colaborativo, onde as ideias foram discutidas abertamente. A revisão conjunta dos códigos, em sessões dedicadas em aula, proporcionou momentos de reflexão e aprendizado mútuo, permitindo ajustes e melhorias que impactaram positivamente a qualidade do projeto.

4.2.5 CONSIDERAÇÕES FINAIS

Participar desse projeto foi uma experiência muito importante na minha formação acadêmica. Conforme Souza, Oliveira e Costa (2018), os projetos de extensão universitária exercem um papel indispensável na formação acadêmica, ao propiciar a aplicação prática do conhecimento teórico em contextos reais e promover o desenvolvimento de competências críticas e colaborativas [8].

Além disso, a análise de Big Data tem sido considerada essencial na atualidade, pois possibilita a compreensão profunda de grandes volumes de dados, contribuindo para decisões estratégicas mais embasadas e inovadoras em diversos setores [6]. Portanto, experiências como esta na universidade tem potencial de alavancar a carreira profissional.

4.3 Bernardo Moro

4.3.1 CONTEXTUALIZAÇÃO

Colaborei em um projeto voltado para a análise de dados públicos brasileiros, com foco em investimentos, dívida pública e renúncia fiscal, abrangendo o período de 2014 a 2024.

Enquanto a equipe técnica se concentrava no uso de ferramentas como PySpark no Databricks para as etapas de limpeza, análise e modelagem preditiva, minha principal contribuição foi na criação e gestão do Trello.

Além disso, fui responsável por desenvolver os slides de apresentação, que serviram para comunicar o progresso e os resultados do projeto em suas diversas fases.

4.3.2 METODOLOGIA

O projeto foi desenvolvido de forma remota, com a equipe técnica utilizando o ambiente Databricks Community para analisar quatro bases de dados no formato .csv.

Ao observar o trabalho e os resultados dos meus colegas diretamente na plataforma Databricks, consegui acompanhar de perto e aprender sobre as etapas metodológicas que estavam sendo aplicadas.

Em paralelo, o Trello funcionou como ferramenta central para registrar e verificar o andamento dessas atividades técnicas. Os slides que preparei foram o meio pelo qual todo esse progresso técnico foi comunicado de maneira visual e organizada.

4.3.3 RESULTADOS E DISCUSSÃO

Através do acompanhamento das atividades no Trello e da preparação dos slides, pude perceber que o projeto alcançou resultados bastante positivos, superando as expectativas da equipe. As análises revelaram tendências importantes, como o aumento contínuo da dívida, um crescimento súbito na renúncia fiscal e uma posterior recuperação dos investimentos.

Sobre o processo da equipe, notei que havia uma familiaridade maior com ferramentas como Python e Pandas. Meu papel foi garantir que tanto os resultados quanto os aprendizados da equipe fossem transmitidos de forma clara e organizada por meio das apresentações. Ter contato com o ciclo completo de um projeto de Big Data, mesmo atuando na parte organizacional, e visualizar a aplicação de análises preditivas foi uma experiência muito instrutiva para mim.

4.3.4 REFLEXÃO APROFUNDADA

Minha prática na organização das tarefas e na criação do material de apresentação foi essencial para consolidar a importância da comunicação clara em projetos complexos. A necessidade de traduzir informações técnicas em slides compreensíveis ampliou minha percepção sobre como facilitar o entendimento de modelos de séries temporais com múltiplos fatores para diferentes stakeholders.

4.3.5 CONSIDERAÇÕES FINAIS

Esta foi uma experiência bastante enriquecedora para mim, especialmente nos aspectos de organização e comunicação. Acredito que o uso efetivo do Trello e a qualidade das apresentações foram importantes para o bom andamento e para a clareza geral do projeto. Observando o trabalho da equipe, ficou evidente o valor da análise de dados para a tomada de decisões no setor público.

Referências

- [1] Brazil: 2020 article iv consultation-press release; staff report; and statement by the executive director for brazil. IMF Report, 2021.
- [2] Incentivos fiscais e seus impactos: Uma análise dos incentivos à recuperação econômica no brasil pós-pandemia. Technical report, Instituto de Pesquisa Econômica Aplicada (IPEA), 2021. Relatório acessado em 16 de junho de 2025.
- [3] Relatório de monitoramento da renúncia fiscal, 2023.
- [4] Databricks Inc. Databricks unified data analytics platform. <https://databricks.com/product/unified-data-analytics-platform>, 2021. Acessado em 16 de junho de 2025.
- [5] John H. Marler and John W. Boudreau. An evidence-based review of hr analytics for talent management. *Human Resource Management Review*, 27(3):521–536, 2017.

- [6] Viktor Mayer-Schönberger and Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Houghton Mifflin Harcourt, New York, 2013.
- [7] Wes McKinney. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. O'Reilly Media, 2 edition, 2017.
- [8] Fernando Souza, Carlos Oliveira, and Ana Costa. A participação de estudantes em projetos de extensão: contribuições para sua formação acadêmica. In *Anais do Encontro Nacional de Pesquisa em Educação*, pages 123–130, Salvador, BA, 2018. Universidade Federal da Bahia.
- [9] Matei Zaharia, Reynold Xin, Patrick Wendell, Tathagata Das, Michael J. Franklin, Scott Shenker, and Ion Stoica. Apache spark: A unified engine for big data processing. *Communications of the ACM*, 59(11):56–65, 2016.