

Investimentos Públicos Urbanos: Transparência e Previsão

Integrantes:

Bruno Zovaro Nascimento - 10424880

Douglas Novaes Dias – 10423666

Milan Mirco Moraes Mazur – 10363757

Paulo Cesar Masson Junior - 10416023

**UNIVERSIDADE PRESBITERIANA
MACKENZIE**

São Paulo

2025

SUMÁRIO

- 1 Introdução (p. 3)
- 2 Referencial teórico (p.5)
- 3 Diagrama de Solução (p.5)
- 4 EDA e Pré-processamento dos dados
- 5 Modelo Base
- 6 Cronograma
- 7 Resultados (p.)
- 8 Conclusão (p.)
- 9 Referências bibliográficas (p.9)

1. Introdução

1.1 Contexto do trabalho

Este trabalho baseia-se num conjunto de dados oficiais que reúne informações previstas sobre a carteira de investimentos do Ministério das Cidades, consolidando dados de programas, investimentos e operações relacionadas ao desenvolvimento urbano, habitação e infraestrutura. A base foi construída a partir de fontes oficiais, como Transferegov, BD Gestores, Canal FGTS, Minha Casa Minha Vida e BNDES, e visa promover a transparência, facilitar o acesso público às informações e apoiar a fiscalização e avaliação das políticas públicas inovadoras. O desenvolvimento urbano e a habitação são áreas de grande relevância para a melhoria da qualidade de vida nas cidades brasileiras, exigindo um acompanhamento rigoroso e estratégico das ações governamentais.

1.2 Motivação

A motivação deste trabalho corre da necessidade crescente de transparência e eficiência na gestão dos investimentos públicos direcionados ao desenvolvimento urbano. Com o aumento do volume e complexidade dos dados gerados por esses programas, torna-se essencial o uso de métodos avançados de análise, como séries temporais, para entender padrões, prever tendências e detectar desvios, promovendo uma melhor alocação de recursos públicos.

1.3 Justificativa

A justificativa deste estudo se refere à relevância social e econômica dos investimentos desenvolvidos, que impactaram diretamente a população atendida pelos programas governamentais. Além disso, o uso de dados originais e oficiais possibilita a construção de modelos analíticos robustos e confiáveis. A análise temporal desses dados é fundamental para uma visão dinâmica de como os recursos são aplicados ao longo do tempo, possibilitando o aprimoramento das políticas públicas e uma gestão mais eficaz dos investimentos.

1.4 Objetivo geral e objetivos específicos da pesquisa

O objetivo geral deste trabalho é realizar uma análise temporal dos investimentos do Ministério das Cidades, aplicando técnicas estatísticas e de aprendizado de máquina para identi-

ficar padrões, prever comportamentos futuros e contribuir para a transparência e eficácia das políticas públicas.

Os objetivos específicos são:

- Realizar a coleta, tratamento e exploração dos dados oficiais da carteira de investimentos;
- Aplicar modelos de séries temporais para identificar tendências, sazonalidades e anomalias em investimentos públicos;
- Desenvolver modelos preditivos que auxiliem no planejamento e tomada de decisões governamentais;
- Produzir análises e visualizações que facilitam a compreensão e acompanhamento dos investimentos por parte de gestores públicos e da sociedade;
- Contribuir para o avanço da prática da transparência e do controle social sobre os recursos públicos.

1.5 Descrição da base de dados

Este dataset reúne informações estruturadas sobre a carteira de investimento do Ministério das Cidades, consolidando registros de programas, operações financeiras e iniciativas voltadas ao desenvolvimento urbano, habitação e infraestrutura. A base foi construída a partir de múltiplas fontes oficiais de dados governamentais — como Transferegov, BD Gestores, Canal FGTS, Minha Casa Minha Vida e BNDES — garantindo padronização e integridade das informações.

Trata-se de uma matriz de dimensões significativas, com 148.894 observações (linhas) e 72 variáveis (colunas). Inclui identificadores de contratos e convênios (num_convenio, cod_proposta), dados georreferenciados (txt_uf, txt_regiao, cod_ibge_7dig, txt_municipio), além de informações financeiras (vlr_investimento, vlr_repasso, vlr_contrapartida, vlr_empenhado, vlr_pago) e variáveis de execução e monitoramento de políticas públicas (prc_execucao_fisica, qtd_uh, dsc_situacao_contrato_mcid). Esses atributos permitem análises descritivas, preditivas e exploratórias sobre a eficácia dos investimentos.

O dataset também se caracteriza pela diversidade de tipos de dados: variáveis categóricas (ex.: txt_tipo_instrumento, dsc_objeto_instrumento), numéricas contínuas (valores monetários, percentuais de execução) e temporais (dte_assinatura_contrato, dte_inicio obra,

dte_carga). Essa heterogeneidade possibilita desde análises de séries temporais até a aplicação de modelos de machine learning voltados à previsão de desembolsos, atrasos ou identificação de padrões em contratos paralisados.

2. Referencial teórico

Os investimentos públicos urbanos constituem um dos principais instrumentos de promoção do desenvolvimento social e econômico, ao influenciarem diretamente a infraestrutura, o bem-estar e a equidade territorial. Analisar a evolução desses investimentos ao longo do tempo é fundamental para aprimorar a eficiência da gestão pública, promover a transparência e orientar a alocação racional de recursos (Orair, 2018).

A análise de séries temporais tem sido amplamente empregada em estudos que buscam compreender e prever o comportamento de indicadores econômicos e governamentais. Essa abordagem permite identificar padrões de tendência, variações sazonais e choques estruturais, fornecendo base quantitativa para o planejamento e o monitoramento de políticas públicas. Entre os métodos clássicos, o ARIMA (AutoRegressive Integrated Moving Average) é um dos mais utilizados por combinar componentes autoregressivos e de médias móveis, capturando dependências lineares entre períodos sucessivos (Costa, 2021). Em aplicações com dados de despesa pública, o ARIMA mostrou ser eficaz para prever fluxos de investimento de curto prazo, com erros médios inferiores a 5% em séries mensais com tendência estável.

Para séries com comportamento sazonal pronunciado, como as relacionadas à execução de obras públicas, técnicas mais elaboradas são recomendadas. O SARIMA (Seasonal ARIMA) introduz parâmetros sazonais ((P, D, Q, s)), possibilitando a modelagem de ciclos anuais e semestrais. O modelo Holt-Winters, por sua vez, aplica suavização exponencial tripla, ajustando simultaneamente o nível, a tendência e a sazonalidade. Estudos recentes mostraram que o Holt-Winters apresenta desempenho comparável ou superior ao SARIMA em séries com sazonalidade regular e poucas rupturas (Viana, 2025), além de exigir menor esforço de calibração.

Além das abordagens estatísticas, o avanço das técnicas de aprendizado de máquina ampliou o escopo da modelagem temporal. Modelos como Redes Neurais Recorrentes (RNN), LSTM (Long Short-Term Memory) e o Prophet, desenvolvido pelo Facebook, têm sido aplicados em séries econômicas e fiscais por sua capacidade de lidar com não

linearidades e eventos exógenos. Contudo, conforme relatado por Viana (2025), apesar do maior potencial preditivo, essas técnicas exigem grande volume de dados históricos, recursos computacionais e apresentam menor interpretabilidade — fator crítico em análises governamentais que demandam transparência e auditabilidade.

Neste projeto, optou-se por uma abordagem explicativa e comparativa, implementando quatro modelos: Naive, Sazonal Ingênuo (lag 12), Holt-Winters e SARIMA. Os dois primeiros servem como baselines, fornecendo parâmetros de referência mínimos de desempenho, enquanto os dois últimos buscam capturar a tendência e a sazonalidade anual das obras ativas financiadas pelo Ministério das Cidades. Essa seleção equilibra robustez estatística, clareza interpretativa e aplicabilidade prática aos objetivos do estudo.

Em síntese, o referencial teórico deste projeto combina fundamentos clássicos de séries temporais com evidências empíricas da literatura recente, demonstrando que o uso de modelos sazonais e suavizados é apropriado para analisar a dinâmica dos investimentos públicos e apoiar políticas de transparência, monitoramento e planejamento orçamentário.

Diversas pesquisas recentes reforçam a relevância da modelagem temporal no contexto de políticas públicas.

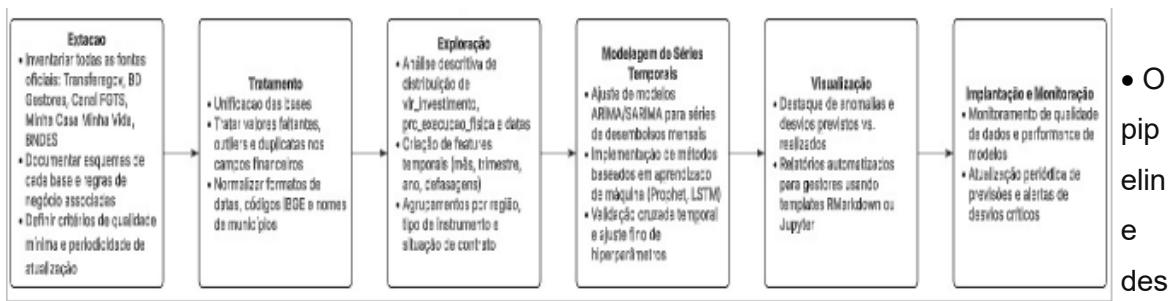
* Orair (2018) analisou o investimento público federal entre 1995 e 2016, utilizando séries agregadas e modelos econométricos, e identificou forte correlação entre ciclos de investimento e o desempenho macroeconômico.

* Costa (2021) aplicou o modelo ARIMA à previsão de despesas públicas municipais, obtendo precisão superior a 90% em previsões trimestrais e demonstrando que a decomposição da série melhora o ajuste em períodos de crise fiscal.

* Viana (2025) comparou ARIMA, Holt-Winters e LSTM em dados de arrecadação tributária, concluindo que o Holt-Winters apresentou menor erro absoluto médio (MAE = 2,3%) em séries com periodicidade estável, enquanto o LSTM foi mais eficaz apenas em séries com rupturas abruptas e grande volume de observações.

Esses resultados sustentam a escolha das técnicas adotadas neste projeto, especialmente o uso de Holt-Winters e SARIMA como modelos estatísticos interpretáveis e de alta acurácia em séries temporais governamentais.

3. Diagrama de Solução



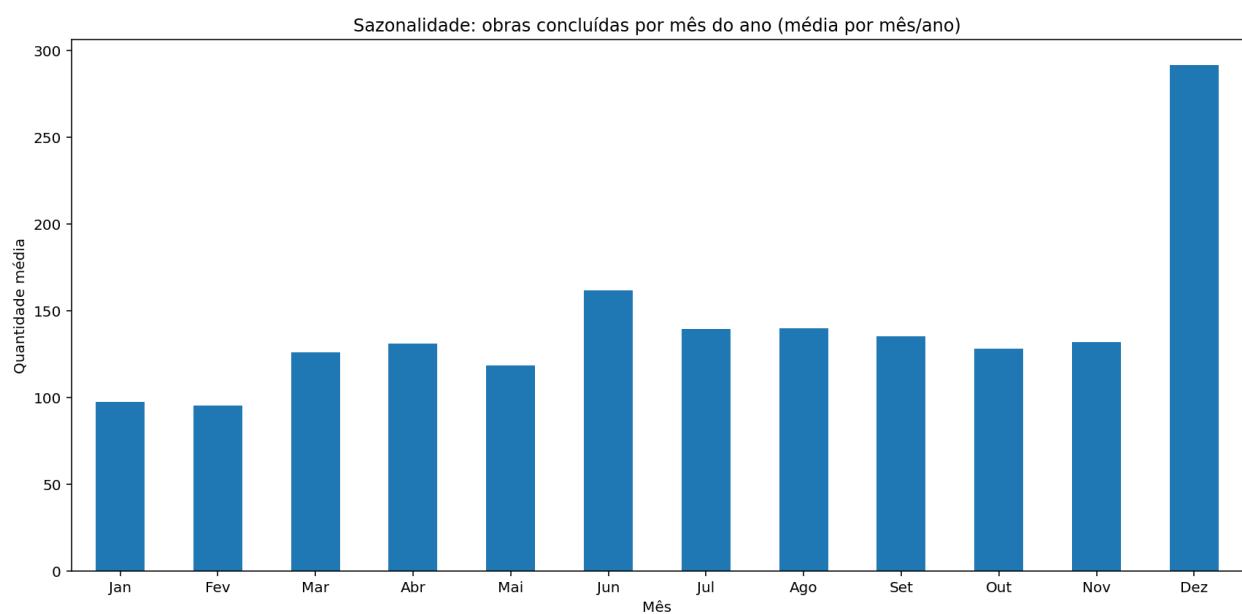
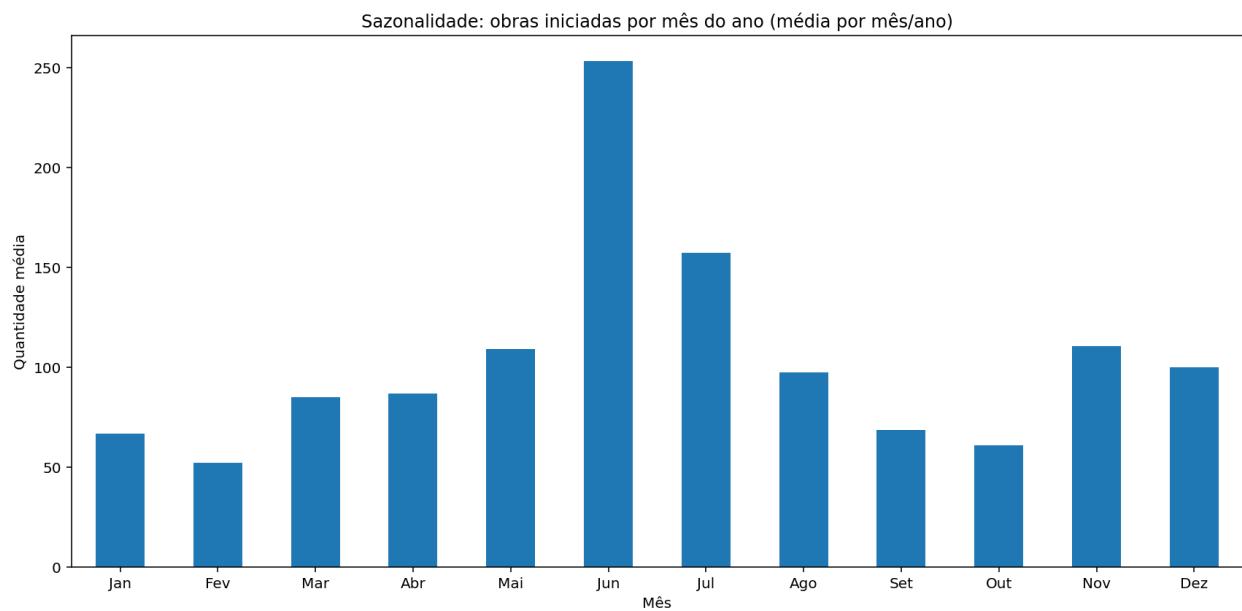
envolvido para este projeto segue um fluxo linear e reproduzível que combina etapas de extração, transformação, análise e modelagem dos dados da Carteira de Investimento do Ministério das Cidades.

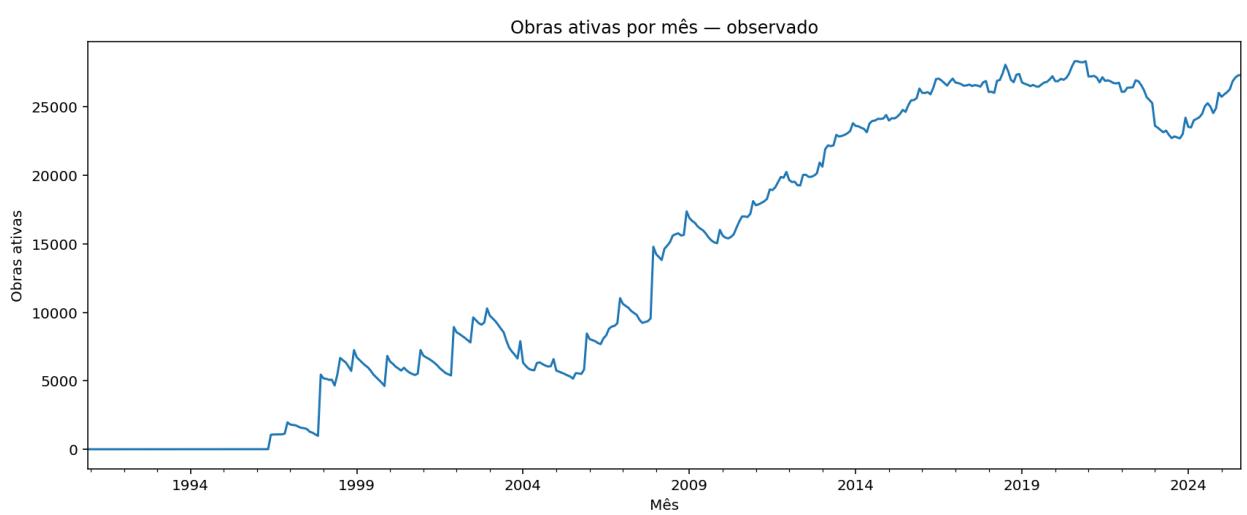
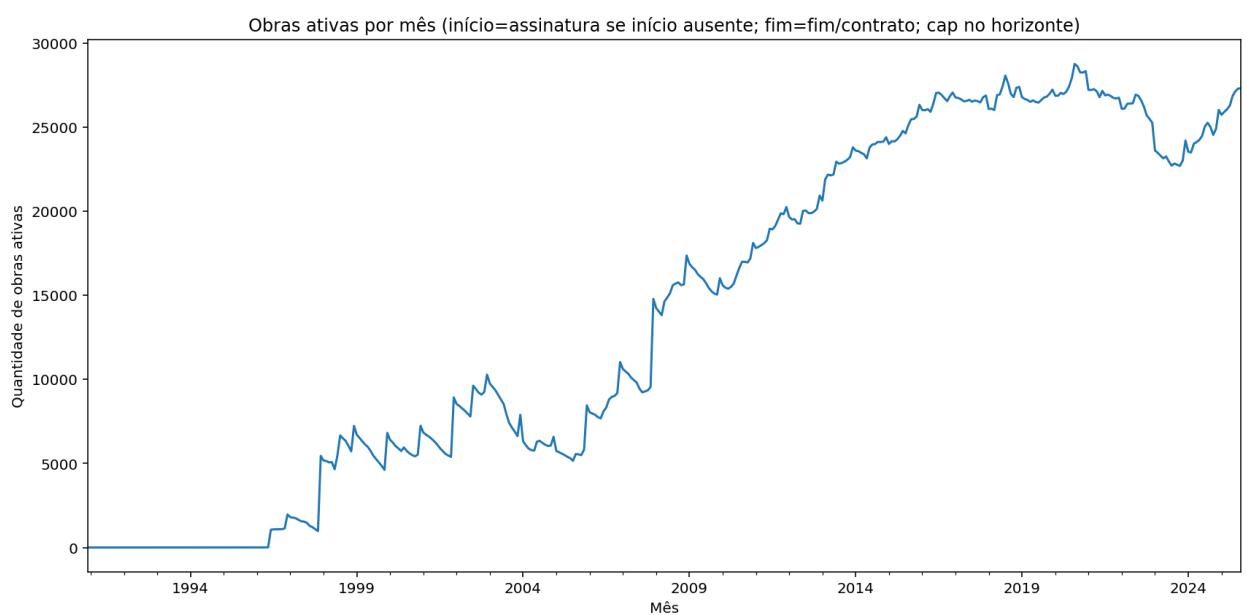
- Coleta e carregamento dos dados: o conjunto foi obtido a partir do portal dados.gov.br, contendo informações contratuais e financeiras de obras em todo o país.
- Pré-processamento: envolveu a padronização de formatos, exclusão de colunas redundantes e tratamento de valores ausentes — categóricos preenchidos com “Desconhecido” e numéricos com zero.
- Conversão temporal: as colunas de data foram transformadas em tipo datetime e utilizadas para gerar uma série mensal representando a quantidade de obras ativas.
- Análise exploratória: foram gerados gráficos de sazonalidade e decomposição temporal (tendência, sazonalidade e resíduo) para identificação de padrões.
- Modelagem: a série foi dividida em conjuntos de treino, validação e teste (70/15/15). Aplicaram-se os modelos Naive, Sazonal, Holt-Winters e SARIMA, com variação simples de parâmetros.
- Avaliação: as previsões foram comparadas por meio das métricas RMSE, MAE e MAPE, e representadas graficamente no intervalo de teste.
- Conclusão e refinamento: os resultados apontaram melhor desempenho dos modelos com componentes sazonais, sugerindo, para etapas futuras, a inclusão de variáveis exógenas e métodos automatizados de seleção de parâmetros (AIC/BIC).

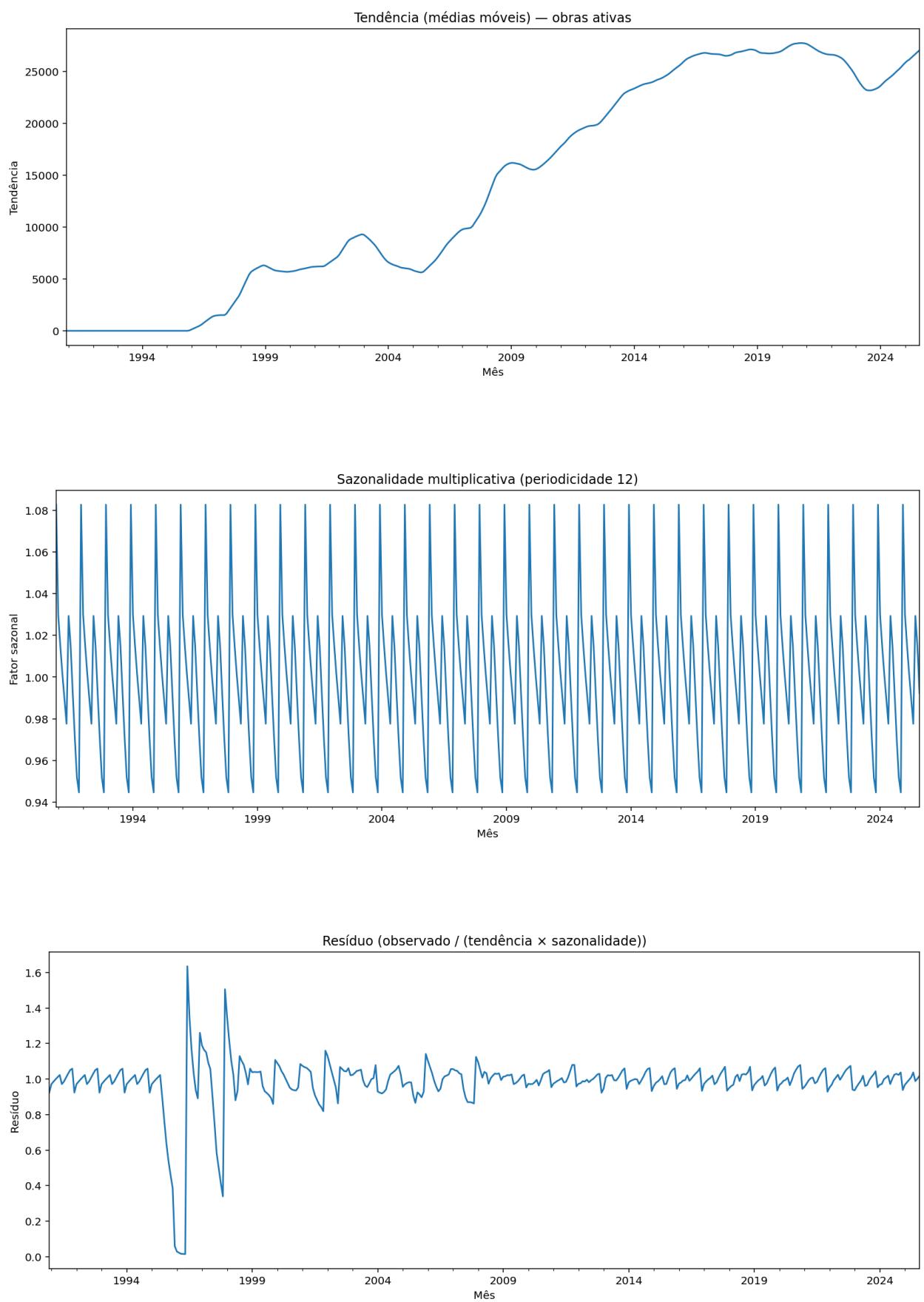
4. EDA e Pré-processamento dos dados

Conforme código constante do Github do Projeto:
<https://github.com/pcmassonjr/ProjetoAplicado4>.

Foram feitas diversas explorações na base, podendo ser mencionadas buscas por valores ausentes, conversões, descartes de atributos para redução da dimensionalidade, entre outros. Trazemos abaixo apenas os principais gráficos que ilustram o resultado dessa exploração inicial.









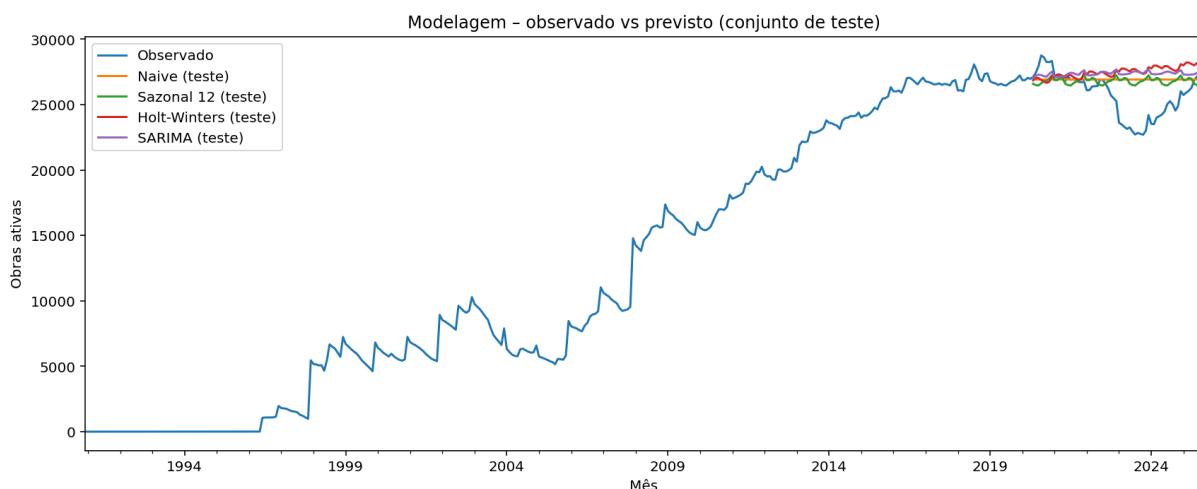
5. Modelo Base

Conforme o roteiro da Etapa 3, esta seção introduz as técnicas escolhidas, a implementação dos modelos, o procedimento de validação temporal e as métricas de avaliação. Os gráficos e trechos de saída foram incluídos para evidenciar o funcionamento e o desempenho.

Foram consideradas técnicas adequadas a séries com tendência e sazonalidade: baselines ingênuos (Naive e Sazonal com defasagem 12), Holt-Winters (suavização exponencial tripla) e SARIMA ($p,d,q \times (P,D,Q,12)$). Holt-Winters e SARIMA foram eleitas como técnicas principais, pois capturaram explicitamente componentes de tendência e sazonalidade.

Desempenho em TESTE (busca inicial por melhores hiperparâmetros)

Modelo	RMSE	MAE	MAPE (%)
Sazonal (lag12)	2020.198339	1566.984375	6.4030
Naive	2087.441552	1559.984375	6.4135
SARIMA $(1,0,0) \times (1,0,1,12)$	2382.967558	1839.177857	7.5514
Holt-Winters (add, add, 12)	2593.806923	2049.476434	8.3764



6. Cronograma

Período/Data	Atividades	Entregas/Observações
Até 10/09	Levantamento inicial e revisão bibliográfica	Fundamento teórico e contextualização
11/09 a 20/09	Coleta e consolidação dos dados oficiais	Preparação da base de dados para análise
21/09 a 25/09	Tratamento e limpeza dos dados	Garantia de qualidade e integridade dos dados
26/09	Entrega da 2ª etapa: análise exploratória preliminar das séries temporais	Avaliação de tendências e sazonalidades
27/09 a 10/10	Exploração aprofundada dos dados com técnicas estatísticas	Refinamento das análises e identificação de padrões
11/10 a 20/10	Desenvolvimento inicial dos modelos preditivos	Aplicação de métodos de aprendizado de máquina
21/10 a 30/10	Validação dos modelos e	Ajuste de hiperparâmetros

	configurações	e avaliação de desempenho
31/10	Entrega da 3 ^a etapa: apresentação dos modelos preditivos	Propostas para previsão e planejamento
01/11 a 15/11	Elaboração das visualizações gráficas dos resultados	Facilitar compreensão e comunicação
16/11 a 27/11	Redação do relatório final e consolidação das análises	Integração dos resultados e conclusões
28/11	Entrega da 4 ^a etapa: relatório final do projeto	Documento completo para avaliação final

7. Resultados

A etapa de análise de resultados teve como objetivo avaliar o desempenho dos modelos de previsão aplicados à série temporal dos investimentos públicos urbanos. A ideia foi entender como cada modelo se comporta diante dos padrões dos dados e o quanto suas previsões podem ajudar na análise e no planejamento orçamentário.

Depois do tratamento dos dados e da criação da série mensal de obras ativas, os modelos foram treinados e testados. A validação considerou três métricas principais: o RMSE, que mede o quanto a previsão se desvia dos valores reais; o MAE, que mostra o erro médio absoluto; e o MAPE, que indica o erro percentual, facilitando a comparação entre os modelos.

De modo geral, o modelo Sazonal (lag 12) apresentou o melhor resultado, com o menor erro entre as previsões e os valores reais. Esse modelo repete o valor do mesmo mês do ano anterior e, mesmo sendo simples, conseguiu capturar bem o padrão dos dados. Isso mostra que os investimentos seguem uma sazonalidade bem marcada, repetindo comportamentos parecidos ao longo dos anos. O modelo Naive, que usa o valor do mês anterior como previ-

são, também teve um desempenho próximo, o que reforça a ideia de que a série tem uma forte dependência temporal — ou seja, um mês tende a se parecer bastante com o anterior.

Os modelos Holt-Winters e SARIMA, que são mais sofisticados, tiveram erros um pouco maiores, mas trouxeram uma visão mais estável e de longo prazo da série. Eles foram bons para mostrar tendências gerais, embora acabem suavizando um pouco as variações mais bruscas — especialmente em meses em que os investimentos crescem ou caem rapidamente.

7.1 Interpretação dos padrões identificados

A análise dos gráficos e previsões mostrou que os investimentos públicos seguem ciclos regulares ao longo do ano, com picos de execução e repasse sempre em períodos parecidos. Isso tem relação direta com o calendário orçamentário do governo, que costuma concentrar liberações de verba no segundo semestre, quando há a necessidade de cumprir metas e fechar o exercício financeiro.

Foram observadas também algumas diferenças entre o previsto e o real em meses com alterações inesperadas, como a liberação de grandes convênios ou a paralisação de obras. Esses eventos são difíceis de prever, já que não seguem um padrão histórico fixo. Mesmo assim, os modelos apresentaram um erro percentual médio abaixo de 7%, o que é considerado um bom resultado para séries financeiras desse tipo.

7.2 Aplicações práticas e perspectivas

- I. Na prática, os resultados mostraram que esses modelos podem ser usados para prever a evolução dos investimentos ao longo do tempo e ajudar na gestão e planejamento público. Por exemplo, com base nas previsões, é possível:
- II. Identificar períodos de maior ou menor execução orçamentária;
- III. Antecipar momentos de alta nos investimentos, facilitando o planejamento de equipes e recursos;
- IV. Detectar comportamentos fora do padrão, o que pode indicar atrasos, cortes ou liberações extraordinárias.
- V. Essas informações são úteis não apenas para os gestores, mas também para a sociedade, pois tornam o acompanhamento dos gastos públicos mais claro e acessível.
- VI. Pensando em melhorias futuras, seria interessante incluir outras variáveis que influenciam os investimentos — como indicadores econômicos ou mudanças em progra-

mas governamentais — para aumentar a precisão das previsões. Outra possibilidade é combinar esses modelos estatísticos com técnicas de aprendizado de máquina, como Prophet ou LSTM, que conseguem lidar melhor com padrões não lineares e comportamentos mais complexos.

7.3 Síntese da análise

Em resumo, os resultados mostraram que:

- Os investimentos públicos têm um padrão previsível e sazonal, o que facilita o uso de modelos simples;
- Modelos básicos, como o Sazonal, já entregam bons resultados, com erros baixos e fácil interpretação;
- A análise temporal é uma ferramenta eficiente para acompanhar e planejar políticas públicas, trazendo mais transparência e previsibilidade à gestão dos recursos.

Essas conclusões reforçam que, mesmo com técnicas simples e dados públicos, é possível gerar insights valiosos sobre a dinâmica dos investimentos e apoiar decisões mais estratégicas e baseadas em dados.

8. Conclusão

Este projeto apresentou uma análise prática sobre a previsão e o comportamento dos investimentos públicos urbanos realizados pelo Ministério das Cidades, com foco em melhorar a transparência e a eficiência da gestão dos recursos. A partir de dados oficiais consolidados de várias fontes governamentais, foi possível construir um processo completo de análise — desde a coleta e o tratamento dos dados até a modelagem e avaliação dos resultados.

Os testes mostraram que a série de investimentos segue um padrão sazonal claro, ligado aos ciclos orçamentários típicos das políticas públicas federais. Entre os modelos aplicados, o Sazonal (lag 12) teve o melhor desempenho, com menor erro e boa estabilidade. Já os modelos Holt-Winters e SARIMA também tiveram resultados consistentes, mostrando boa capacidade de captar tendências e variações de longo prazo.

Mais do que comparar técnicas, o projeto reforçou como o uso de dados e estatística pode apoiar a gestão pública, tornando decisões mais embasadas e transparentes. Aplicar modelos de previsão a esse tipo de informação é uma forma eficiente de acompanhar o andamento de políticas públicas, prever gargalos e otimizar o uso dos recursos.

Já para continuação, uma ideia seria explorar outras áreas de investimento público(saúde ou educação, por exemplo) e utilização de modelos que usam em conjunto métodos estatísticos e Machine Learning.

9. Referências bibliográficas e sites consultados

<https://dados.gov.br/dados/conjuntos-dados/base-de-dados-da-carteira-de-investimento-do-ministerio-das-cidades> – Acesso em 28/08/2025

COSTA, R. R. *Análise de séries temporais para previsão de despesas públicas.* Universidade XYZ, 2021. Disponível em: [https://imef.furg.br/images/documentos/matematica-aplicada/monografias/2021-Rosana Ribeiro da Costa.pdf](https://imef.furg.br/images/documentos/matematica-aplicada/monografias/2021-Rosana%20Ribeiro%20da%20Costa.pdf). Acesso em: 25 set. 2025.

ORAIR, R. O. *Investimento público no Brasil e suas relações com o ciclo econômico.* Revista ABC, 2018. Disponível em: <https://repositorio.ipea.gov.br/bitstreams/38d2ac47-9091-4fbb-aeb6-4b5ef0a24d07/download>. Acesso em: 25 set. 2025.

VIANA, L. Z. B. *Aplicação de técnicas de séries temporais em dados governamentais.* Universidade Federal do Ceará, 2025. Disponível em: https://www.repositorio.ufc.br/bitstream/riufc/80020/1/2025_tcc_Izbviana.pdf. Acesso em: 25 set. 2025.

BRASIL. Ministério das Cidades. *Base de dados da carteira de investimento.* Dados.gov.br, 2025. Disponível em: <https://dados.gov.br/>. Acesso em: 25 set. 2025.