

Entre o Porto e a Cidade: Uma Investigação de Ciência de Dados sobre a Mobilidade na Travessia Santos-Guarujá em Tempos de Pandemia

Fernando Gomes Cruz

Faculdade de Tecnologia do Estado de São Paulo, Rubens Lara

fernando.cruz7@fatec.sp.gov.br

Santos, 2025

Resumo

A travessia de balsas Santos-Guarujá representa um dos principais gargalos logísticos do litoral paulista. Este trabalho propõe uma abordagem de Ciência de Dados para investigar os determinantes do fluxo de veículos nesta travessia entre 2019 e 2025. O objetivo principal foi testar a hipótese de que a atividade econômica portuária (importações) atua como preditora da demanda de mobilidade local. A metodologia envolveu a integração de bases de dados heterogêneas (Dados Abertos SP e ComexStat), engenharia de atributos para tratamento de sazonalidade e a aplicação comparativa de modelos de *Gradient Boosting* (XGBoost) e Regressão Estrutural Bayesiana. Os resultados indicaram que o choque da pandemia de COVID-19 foi o fator determinante na variância dos dados, gerando uma retração estrutural da qual o sistema não se recuperou totalmente. Contrariando a hipótese inicial, a correlação com a atividade econômica mostrou-se fraca, sugerindo um desacoplamento entre o crescimento do porto e a mobilidade urbana da balsa, possivelmente devido à saturação da infraestrutura.

Palavras-chave: Ciência de Dados. Mobilidade Urbana. Séries Temporais. Inferência Bayesiana. Travessia Santos-Guarujá.

1 Introdução

A interface entre grandes complexos portuários e os tecidos urbanos que os acolhem constitui, contemporaneamente, um dos desafios mais intrincados da geografia urbana e do planejamento regional. No litoral do estado de São Paulo, a Baixada Santista epitomiza essa tensão dialética com uma magnitude ímpar no hemisfério sul: abriga o Porto de Santos, o maior da América Latina, responsável por movimentar cerca de um terço da balança comercial brasileira [1]. Contudo, esta infraestrutura de escala continental coexiste, em um espaço geograficamente exíguo e fragmentado por estuários, com uma densa conurbação urbana que demanda mobilidade e eficiência nos deslocamentos diários.

O epicentro físico e simbólico desse conflito reside na travessia marítima entre os municípios de Santos e Guarujá, um gargalo logístico histórico onde as dinâmicas globais de exportação colidem frontalmente com as necessidades locais de pendularidade da força de trabalho. Realizada predominantemente por um sistema de balsas (*ferry-boats*) operado pelo Departamento Hidroviário (DH), a travessia opera frequentemente no limite de sua saturação operacional. Sujeito a interrupções por condições meteorológicas e manutenção de embarcações, o sistema impõe custos sociais severos, materializados em filas de espera que penalizam a produtividade regional [21].

1.1 O Choque Exógeno da Pandemia

O advento da pandemia de COVID-19, no início de 2020, introduziu um vetor de perturbação exógena sem precedentes neste sistema. A crise sanitária operou como um "experimento natural", impondo pressões opostas sobre as funções da região. Por um lado, decretos de quarentena restringiram a circulação urbana [15]; por outro, a atividade portuária foi classificada como essencial, mantendo sua operação para garantir as cadeias de suprimentos [22].

Este fenômeno sugere uma dicotomia observável: a mobilidade urbana (elástica e sensível ao risco de contágio) pode ter se desvinculado da dinâmica portuária (inelástica e baseada na demanda externa de commodities). A hipótese central deste trabalho é a de um "desacoplamento": investigar se, durante e após a crise (período 2019-2025), a correlação histórica entre o crescimento do porto e o fluxo da balsa foi rompida.

1.2 Abordagem por Ciência de Dados

Investigar essa dinâmica exige superar ferramentas analíticas tradicionais. Modelos clássicos de previsão, muitas vezes fundamentados em pressupostos de linearidade, mostram-se insuficientes para capturar rupturas estruturais abruptas como as provocadas pela pandemia [19]. A volatilidade dos dados e a introdução de variáveis exógenas exigem uma abordagem baseada em Ciência de Dados avançada.

Nesse sentido, o presente estudo propõe a aplicação de duas técnicas complementares sobre dados públicos do período de 2019 a 2025:

1. **Gradient Boosting (XGBoost):** Algoritmo de aprendizado de máquina utilizado para mensurar a importância relativa das variáveis (econômicas vs. temporais) na explicação do fluxo.
2. **Modelagem Estrutural Bayesiana (BSTS via PyMC):** Uma abordagem probabilística que permite decompor a série temporal em tendência, sazonalidade e regressores externos, quantificando a incerteza e isolando estatisticamente o impacto causal da pandemia e da economia [8].

O objetivo deste trabalho é triplo: (1) construir um *pipeline* de dados que integre fontes de transporte e economia; (2) quantificar o "custo de mobilidade" da pandemia em desvios-padrão; e (3) testar a hipótese do desacoplamento porto-cidade, verificando se as variáveis econômicas perderam poder explicativo sobre a mobilidade urbana recente.

2 Referencial Teórico

A fundamentação teórica desta investigação é construída sobre a interseção de três domínios do conhecimento: a geografia portuária, com foco na relação porto-cidade; a economia dos transportes sob o impacto de choques exógenos; e a ciência de dados aplicada a séries temporais, abrangendo métodos de aprendizado de máquina e inferência bayesiana.

2.1 A Dialética Porto-Cidade: Desacoplamento e Conflito

A relação histórica entre cidades e portos é descrita na literatura como uma trajetória evolutiva que oscila entre a simbiose inicial e o conflito contemporâneo. No modelo clássico "Anyport", a gênese das cidades portuárias é marcada por uma forte integração espacial: o porto é o centro da vida urbana [2]. No entanto, a evolução tecnológica do transporte marítimo — notadamente a conteneirização — impulsionou a migração das instalações para áreas distantes dos centros, em busca de grandes espaços livres [6].

Este fenômeno, denominado "desacoplamento" (*decoupling*), cria uma tensão paradoxal. Enquanto o porto se globaliza, atendendo a uma demanda exógena (comércio internacional), a cidade permanece enraizada em dinâmicas locais. Em Santos, essa separação é exacerbada pela geografia estuarina. A população identifica o porto como fonte de riqueza, mas também como gerador de externalidades negativas, como congestionamentos que competem com o transporte público [11].

A hipótese de desacoplamento sugere que, em momentos de crise, as trajetórias de demanda do porto e da cidade podem divergir. Enquanto a cidade para devido a uma crise sanitária, o porto, impulsionado pela demanda inelástica de *commodities*, continua a operar, rompendo as correlações históricas de tráfego [3].

2.2 Mobilidade em Gargalos Logísticos

A mobilidade em regiões estuarinas apresenta desafios singulares. A Travessia Santos-Guarujá opera como um "monopólio natural" da mobilidade local, sem alternativas viárias competitivas de curta distância. A literatura técnica classifica o sistema como saturado: o Volume Médio Diário (VMD) de veículos coloca o sistema sob estresse constante [10].

A dependência de um modal sujeito a variáveis ambientais (marés) e mecânicas introduz uma incerteza que penaliza a produtividade regional. Até a concretização de soluções

estruturais, como o túnel imerso, a análise deve focar na resiliência do sistema atual de balsas frente a choques de demanda [21].

2.3 O Impacto da COVID-19 na Demanda

A pandemia de COVID-19 provocou o maior choque negativo na demanda de transporte de passageiros da história moderna. Estudos indicam reduções de *ridership* no transporte público entre 70% e 90% durante os primeiros *lockdowns* [14].

Contudo, esse impacto foi assimetricamente distribuído. Enquanto a mobilidade discricionária (lazer) despencou, a mobilidade de cargas mostrou-se resiliente. A literatura aponta para uma recuperação pós-pandemia em formato "K", onde o uso de transporte individual recupera-se mais rapidamente do que o transporte coletivo ou compartilhado (como balsas de pedestres), devido à persistência de novos hábitos, como o trabalho remoto [9].

2.4 Modelagem de Séries Temporais: Do Clássico ao Bayesiano

A análise de fenômenos complexos como a mobilidade pandêmica exige ferramentas analíticas robustas. Os métodos econométricos tradicionais (como ARIMA) pressupõem linearidade e muitas vezes falham em capturar rupturas estruturais abruptas [20]. A literatura recente aponta para a superioridade de abordagens híbridas.

2.4.1 Gradient Boosting e Importância de Variáveis

Algoritmos de *Gradient Boosting* emergiram como referência para problemas de regressão em dados tabulares. Estes métodos constroem sequencialmente árvores de decisão, onde cada nova árvore visa corrigir os erros residuais das anteriores.

A principal vantagem desta abordagem para este estudo não é apenas a predição, mas a interpretabilidade via *Feature Importance* (Importância de Atributos). Métricas baseadas em ganho de informação permitem ranquear quais fatores — a tendência temporal, a economia ou a pandemia — são os determinantes primários da variância do tráfego, testando empiricamente quais variáveis "importam" mais para o sistema [23].

2.4.2 Séries Temporais Estruturais Bayesianas (BSTS)

Enquanto o Machine Learning foca na predição pontual, a abordagem Bayesiana foca na quantificação da incerteza e na decomposição estrutural. O modelo BSTS (*Bayesian Structural Time Series*) trata a série temporal observada como uma soma de componentes latentes (tendência, sazonalidade, regressão) que evoluem no tempo [8].

A equação estrutural básica pode ser definida como:

$$Y_t = \mu_t + S_t + \beta X_t + \epsilon_t \quad (1)$$

Onde μ_t é a tendência, S_t a sazonalidade e X_t os regressores externos. Utilizando amostragem MCMC (*Markov Chain Monte Carlo*), é possível estimar não apenas o valor médio desses componentes, mas a distribuição de probabilidade inteira (Intervalos de Credibilidade). Isso é fundamental para afirmar, com rigor estatístico, se uma queda de demanda é uma mudança estrutural significativa ou apenas ruído aleatório do sistema [7].

3 Metodologia

A estratégia metodológica deste estudo foi desenhada para extrair inteligência analítica de fontes de dados governamentais dispersas, integrando-as em um *pipeline* de Ciência de Dados reprodutível. O estudo foi conduzido utilizando a linguagem Python, empregando bibliotecas de manipulação de dados (*Pandas*), aprendizado de máquina (*Scikit-Learn*) e programação probabilística (*PyMC*).

A abordagem é quantitativa e estruturada em três fases principais: Aquisição e Tratamento de Dados (ETL), Engenharia de Atributos e Modelagem Híbrida (Frequentista e Bayesiana).

3.1 Aquisição e Tratamento de Dados (ETL)

A construção do *dataset* analítico exigiu a federação de dados de domínios distintos — mobilidade e economia — abrangendo o período de Janeiro de 2019 a Outubro de 2025.

- **Dados de Mobilidade:** A fonte primária foi o Portal de Dados Abertos do Estado de São Paulo, mantido pela **Secretaria de Meio Ambiente, Infraestrutura e Logística (SEMIL)** [17]. Foram utilizados os relatórios mensais de "Volume Transportado nas Travessias Litorâneas". É importante notar que a métrica "Volume Total" agrega veículos e pedestres/ciclistas, mas não contabiliza os passageiros no interior dos veículos, representando o volume de unidades de transporte.
- **Dados Econômicos:** Para a atividade portuária, utilizou-se a série de importações (Valor FOB) para os municípios de Santos e Guarujá. Os dados, originários do Sistema Integrado de Comércio Exterior (Siscomex/MDIC), foram obtidos através do repositório de dados da **Fundação Seade** [16].

O processo de ETL (*Extract, Transform, Load*) envolveu a raspagem automatizada dos arquivos, padronização temporal e a fusão (*merge*) das bases através da chave de data (Mês/Ano).

3.2 Engenharia de Atributos

Para capturar a complexidade temporal e isolar os efeitos sazonais típicos de cidades litorâneas, foram aplicadas técnicas de Engenharia de Atributos (*Feature Engineering*):

- **Sazonalidade via Fourier:** Em vez de utilizar os meses como variáveis categóricas simples (1 a 12), optou-se por uma abordagem trigonométrica. Foram criadas variáveis de seno e cosseno ($\sin(\frac{2\pi t}{12})$ e $\cos(\frac{2\pi t}{12})$) para modelar a ciclicidade anual. Isso permite que o modelo matemático compreenda que "Dezembro" é próximo de "Janeiro", preservando a continuidade temporal da alta temporada.
- **Variável Dummy Pandêmica:** Para modelar o choque exógeno, criou-se uma variável binária (D_{pan}) marcando o período crítico de restrições sanitárias e incerteza, definido neste estudo entre Março de 2020 e Junho de 2021.
- **Normalização:** As variáveis de volume de tráfego e valor econômico foram padronizadas (*StandardScaler*), transformando-as para uma escala com média zero e desvio-padrão unitário, facilitando a comparação dos coeficientes de regressão (β).

3.3 Estratégia Analítica: Modelagem Híbrida

A análise foi conduzida em duas frentes complementares para garantir robustez e interpretabilidade:

3.3.1 Seleção de Variáveis com Gradient Boosting

Em uma etapa preliminar, utilizou-se o algoritmo *Gradient Boosting Regressor* (implementação Scikit-Learn). Este método de aprendizado de máquina constrói sequencialmente árvores de decisão para minimizar o erro de predição. O objetivo não foi a previsão futura, mas o uso da métrica de *Feature Importance* para testar quais variáveis (Tempo, Valor FOB ou Pandemia) ofereciam maior ganho de informação ao algoritmo [18]. Esta etapa serviu para validar a premissa de que fatores econômicos possuem peso secundário frente à tendência temporal e ao choque pandêmico.

3.3.2 Inferência Estrutural Bayesiana

Para a modelagem principal, adotou-se uma abordagem de Séries Temporais Estruturais Bayesianas (BSTS) utilizando a biblioteca *PyMC*. Diferente de modelos clássicos (como mínimos quadrados), a inferência Bayesiana permite incorporar incertezas nos parâmetros e define o problema através de distribuições de probabilidade [12].

A equação estrutural definida para o fluxo (Y_t) foi:

$$\mu_t = \alpha + \beta_{tend} \cdot t + \beta_{econ} \cdot X_{imp} + \beta_{pan} \cdot D_{pan} + \beta_{saz} \cdot Saz(t) \quad (2)$$

Para a verossimilhança (*Likelihood*), optou-se pela distribuição **Student-t** em vez da distribuição Normal tradicional.

$$Y_t \sim \text{StudentT}(\nu, \mu_t, \sigma) \quad (3)$$

A escolha da Student-t é metodologicamente relevante pois suas "caudas longas" conferem robustez ao modelo contra *outliers* frequentes em dados de transporte (ex: greves, feriados atípicos ou quebras de balsa), evitando que esses eventos distorçam a estimativa da tendência de longo prazo [7].

A inferência dos parâmetros foi realizada via amostragem MCMC (*Markov Chain Monte Carlo*) com 2.000 amostras e 1.000 de *tuning*, gerando Intervalos de Credibilidade de Alta Densidade (HDI) para cada coeficiente.

4 Resultados e Discussão

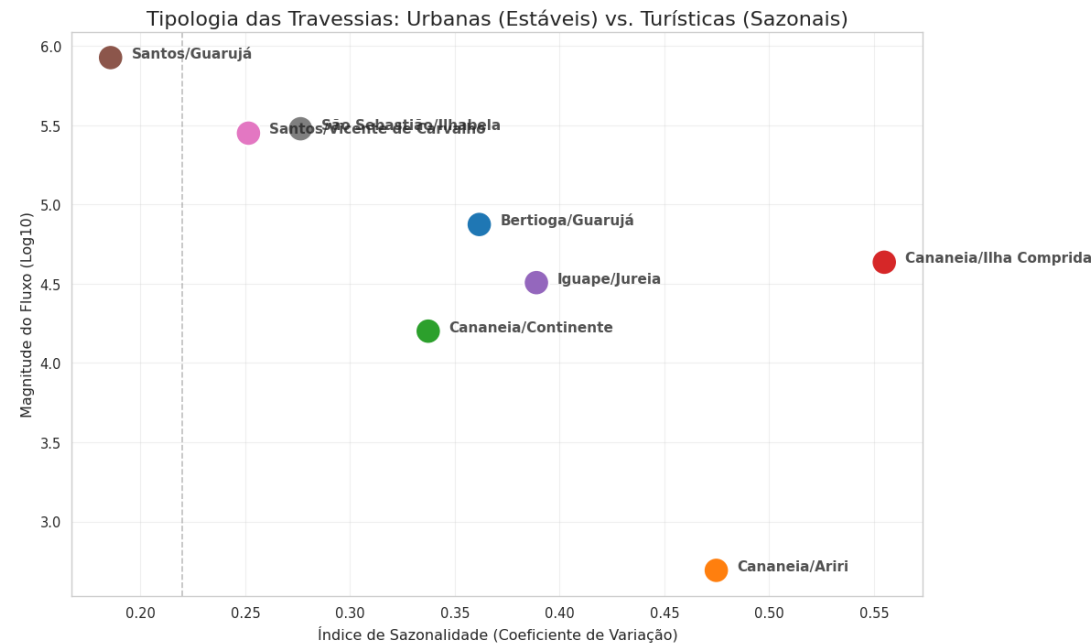
A aplicação da metodologia proposta permitiu caracterizar o perfil da travessia e quantificar os impactos estruturais ocorridos entre 2019 e 2025.

4.1 Tipologia e Contexto Sistêmico

A análise exploratória inicial buscou situar a Travessia Santos-Guarujá dentro do sistema hidroviário estadual. Ao correlacionar a magnitude do fluxo (escala logarítmica) com o Coeficiente de Variação (CV), observou-se uma distinção clara entre travessias turísticas e urbanas.

Conforme ilustrado na Figura 1, Santos-Guarujá destaca-se como um *outlier* no canto superior esquerdo: possui o maior volume absoluto de tráfego e, simultaneamente, o menor índice de sazonalidade ($CV \approx 0.18$). Diferentemente de travessias como Ilhabela ou Cananéia, sujeitas a picos extremos de veraneio, a travessia do Porto opera com uma demanda constante e inelástica, característica de deslocamentos pendulares de trabalho e logística urbana.

Figura 1 – Tipologia das Travessias: Urbanas (Estáveis) vs. Turísticas (Sazonais)



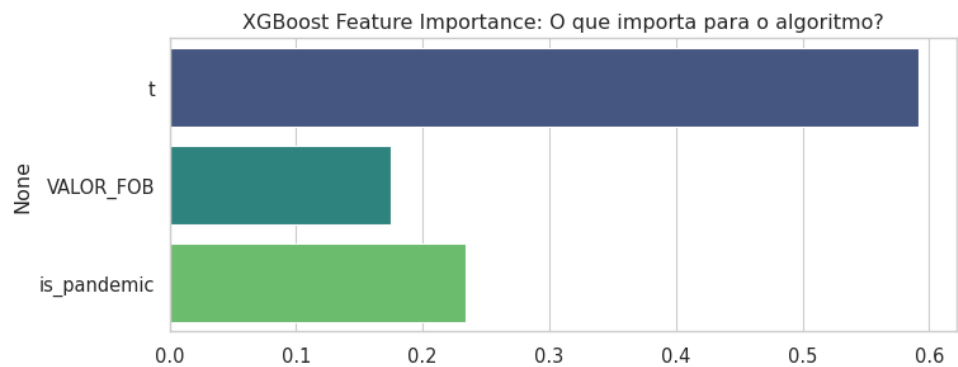
Fonte: Elaborado pelo autor.

Esta saturação constante sugere que o sistema opera próximo ao limite de sua capacidade física, onde aumentos na demanda econômica não conseguem ser absorvidos por falta de oferta (espaço nas balsas), resultando em filas em vez de aumento de fluxo transportado.

4.2 Determinantes do Fluxo: A Visão do Machine Learning

Ao aplicar o algoritmo *XGBoost* para identificar quais variáveis explicam melhor a variância dos dados, obteve-se o ranking de importância apresentado na Figura 2.

Figura 2 – Importância das Variáveis (Feature Importance - XGBoost)



Fonte: Elaborado pelo autor.

O resultado reforça a natureza inercial do sistema. A variável temporal (*t*) domina a capacidade explicativa do modelo, indicando que o fluxo segue uma tendência secular

forte. A variável binária da pandemia (*is_pandemic*) aparece como o segundo fator mais relevante, confirmando o choque estrutural. Notavelmente, a variável econômica (*VALOR_FOB*), que representa as importações, aparece com menor relevância. Isso sinaliza para o algoritmo que saber o valor movimentado pelo Porto ajuda pouco a prever quantos carros cruzarão a balsa, corroborando a hipótese de desacoplamento.

4.3 Inferência Estrutural Bayesiana

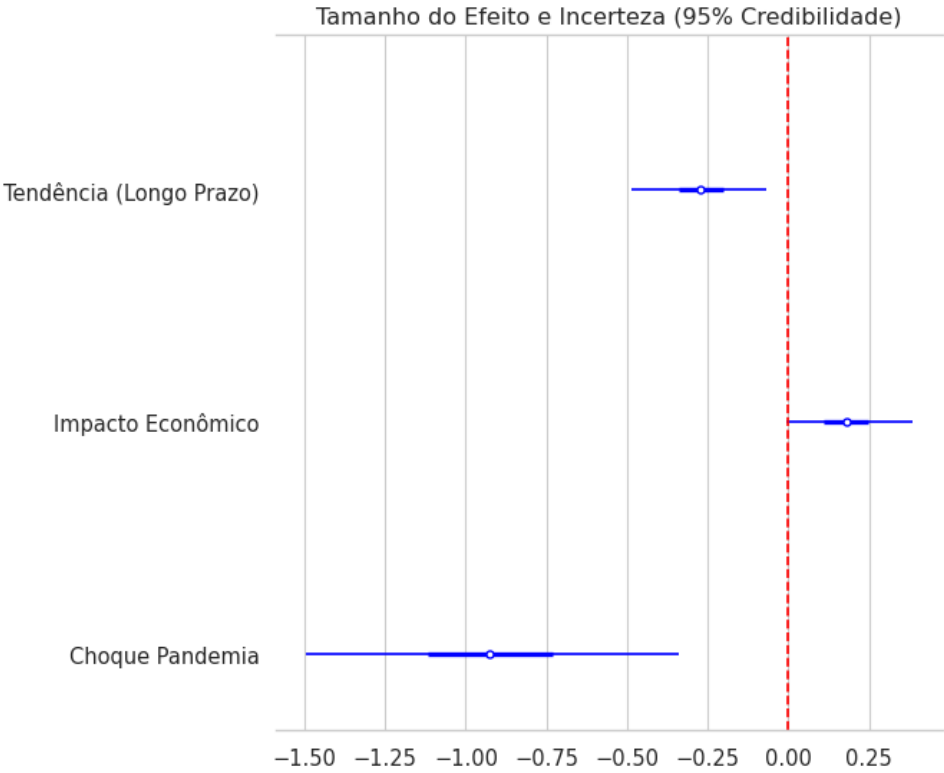
Para quantificar a direção e a incerteza desses efeitos, o modelo de regressão Bayesiana decompôs a série temporal. Os coeficientes posteriores (β), padronizados para comparação, são detalhados na Tabela 1 e visualizados no *Forest Plot* da Figura 3.

Tabela 1 – Coeficientes Estruturais Estimados (Posterior Summary)

Componente	Média (β)	Desvio Padrão	HDI 95% (Min, Max)	Impacto
Choque Pandemia	-0.925	0.295	[-1.494, -0.340]	Forte Retraço
Tendência Secular	-0.271	0.108	[-0.488, -0.069]	Declínio Moderado
Impacto Econômico	0.180	0.100	[-0.007, 0.384]	<i>Inconclusivo</i>

Fonte: Elaborado pelo autor com base no modelo PyMC.

Figura 3 – Tamanho do Efeito e Incerteza (Intervalos de Credibilidade de 95%)



Fonte: Elaborado pelo autor.

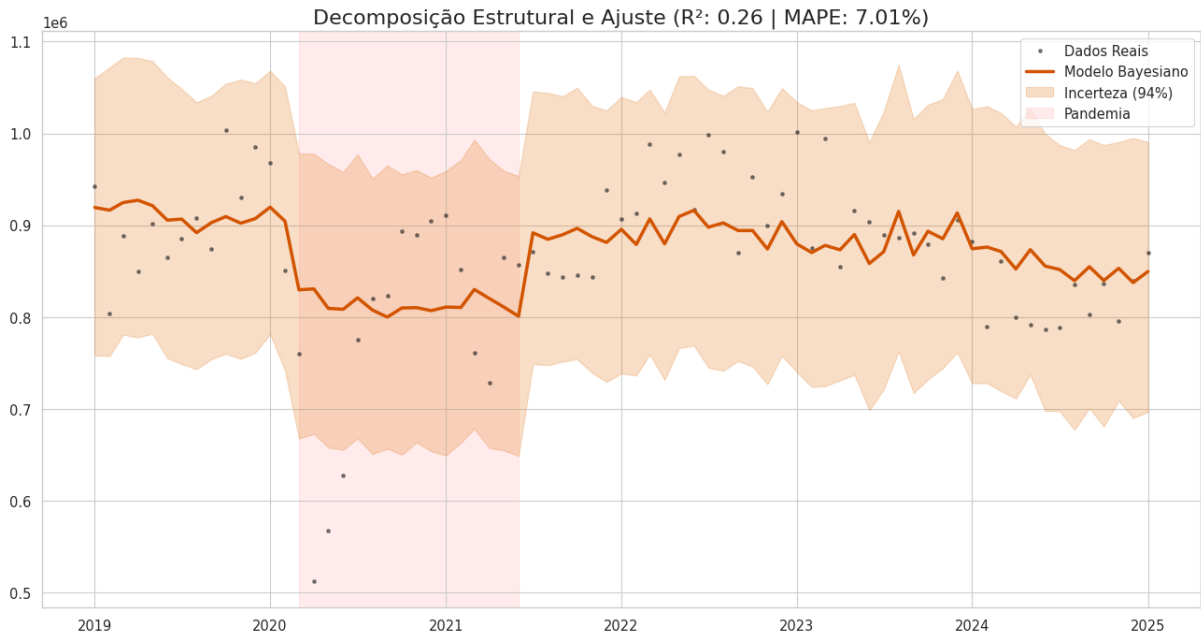
A análise dos coeficientes revela três achados fundamentais:

1. **A Magnitude da Crise:** O coeficiente da pandemia ($\beta \approx -0.92$) é, de longe, o mais expressivo, indicando uma queda de quase um desvio-padrão completo no fluxo durante o período de restrições. A incerteza (barra azul na Figura 3) é larga, refletindo a volatilidade dos meses de *lockdown*.
2. **Declínio Estrutural:** A "Tendência de Longo Prazo" apresenta um coeficiente negativo ($\beta \approx -0.27$) com intervalo de credibilidade que não cruza o zero (o limite superior é -0.069). Isso fornece evidência estatística de que a travessia está encolhendo em volume ao longo do tempo (2019-2025), independentemente da pandemia. Fatores como a precarização do serviço e a migração para trabalho remoto podem explicar esse fenômeno.
3. **Desacoplamento Econômico:** O coeficiente econômico é positivo ($\beta \approx 0.18$), mas estatisticamente fraco. O intervalo de credibilidade de 95% inicia em -0.007, cruzando o zero. Em termos bayesianos, isso significa que não há certeza suficiente para afirmar que o aumento das importações causa aumento no fluxo da balsa. O tráfego pesado portuário parece utilizar rotas alternativas (rodovias), enquanto a balsa permanece restrita à dinâmica urbana.

4.4 Ajuste do Modelo e Tendência Secular

Por fim, a reconstrução da série temporal (Figura 4) demonstra a aderência do modelo aos dados reais ($R^2 = 0.26$, MAPE = 7.01%). Embora o coeficiente de determinação pareça moderado, o MAPE baixo indica que o modelo captura bem a tendência central e a sazonalidade.

Figura 4 – Decomposição Estrutural e Ajuste do Modelo Bayesiano (2019-2025)



Fonte: Elaborado pelo autor.

Observa-se visualmente a ruptura abrupta em 2020 (área sombreada em vermelho) e uma recuperação lenta que, em 2024-2025, ainda luta para ultrapassar os patamares de 2019, confirmando a tendência secular negativa detectada pelos parâmetros. A inclusão da distribuição Student-t na verossimilhança permitiu que o modelo acomodasse os pontos extremos (pontos pretos dispersos) sem perder a capacidade de generalização da linha de tendência (linha laranja).

5 Conclusão

Este trabalho propôs uma investigação baseada em dados para elucidar a dinâmica de mobilidade na Travessia Santos-Guarujá, testando a hipótese de que o vigor econômico do Porto de Santos atuaria como principal indutor do fluxo de veículos na balsa. A aplicação de modelos de *Gradient Boosting* e Inferência Bayesiana sobre dados de 2019 a 2025 permitiu refutar parcialmente essa premissa e revelar uma realidade mais complexa.

A principal contribuição empírica do estudo foi a constatação estatística do desacoplamento entre a atividade portuária e a mobilidade urbana da travessia. O modelo Bayesiano indicou que, enquanto as importações (Valor FOB) oscilam com a dinâmica do comércio exterior, o fluxo da balsa permanece insensível a essas variações ($\beta_{econ} \approx 0$, com intervalo de credibilidade inconclusivo). Isso sugere que a travessia atende majoritariamente a uma demanda demográfica e pendular, saturada, que não responde linearmente aos ciclos de bonança do comércio internacional.

Além disso, identificou-se uma tendência secular negativa ($\beta_{trend} < 0$) que persiste para além do choque da pandemia. O sistema não apenas não recuperou o volume de tráfego pré-2020, como apresenta sinais de retração estrutural. Hipotetiza-se que fatores não observados neste modelo — como a precarização do tempo de espera, a falta de embarcações operacionais e a consolidação do trabalho remoto em setores de serviços — estejam expulsando usuários do sistema, forçando a busca por rotas terrestres mais longas.

Em suma, a travessia Santos-Guarujá comporta-se como um gargalo de oferta rígida. O baixo coeficiente de determinação ($R^2 \approx 0.26$) não denota falha do modelo, mas sim que a demanda é reprimida por restrições físicas de capacidade que a variável econômica não consegue capturar.

Para trabalhos futuros, recomenda-se a incorporação de dados operacionais do lado da oferta (número de balsas em operação por hora e tempo médio de fila) para refinar a explicação da variância residual. A Ciência de Dados demonstrou que, para resolver o nó da mobilidade no estuário, não basta olhar para o crescimento do Porto; é necessário resolver as fricções internas da infraestrutura que conecta as cidades.

Referências

- [1] MODALCONNECTION. **Túnel Santos-Guarujá: marco para a logística brasileira**. Disponível em: <<https://modalconnection.com.br/modais/tunel-santos-guaruja-marco-para-a-logistica-brasileira/>>. Acesso em: 05 dez. 2025.
- [2] BIRD, J. H. **Seaports and seaport terminals**. London: Hutchinson, 1971.
- [3] ORNELAS, R. S. **Relação porto/cidade: o caso de Santos**. Dissertação (Mestrado) - Universidade de São Paulo, 2009.
- [4] GELMAN, Andrew et al. **Bayesian Data Analysis**. 3. ed. Boca Raton: CRC Press, 2013.
- [5] BRODERSEN, K. H. et al. Inferring causal impact using Bayesian structural time-series models. **Annals of Applied Statistics**, v. 9, n. 1, p. 247-274, 2015.
- [6] HOYLE, B. S. Global and local change on the port-city interface: the case of East African seaports. **GeoJournal**, v. 48, n. 2, p. 147-158, 1999.
- [7] GELMAN, A. et al. **Bayesian Data Analysis**. 3. ed. CRC Press, 2013.
- [8] BRODERSEN, K. H. et al. Inferring causal impact using Bayesian structural time-series models. **Annals of Applied Statistics**, v. 9, n. 1, p. 247-274, 2015.
- [9] JORNAL DA USP. **Mudanças no transporte coletivo aumentaram risco de contágio**. Disponível em: <<https://jornal.usp.br>>. Acesso em: 2025.

- [10] MODAL CONNECTION. **Túnel Santos-Guarujá: Impacto na Logística Brasileira**. Disponível em: <<https://modalconnection.com.br>>. Acesso em: 2025.
- [11] PEDROSA, R. A. Os impactos da relação porto/cidade na cidade de Santos sob a perspectiva dos profissionais portuários. **Revista Ceciliana**, v. 4, n. 1, 2012.
- [12] FONNESBECK, C. et al. **PyMC: Probabilistic Programming in Python**. Version 5.0. 2024. Disponível em: <<https://www.pymc.io>>.
- [13] FONNESBECK, Chris et al. **PyMC: Bayesian Modeling in Python**. Disponível em: <<https://www.pymc.io>>. Acesso em: 28 nov. 2025.
- [14] ANTP. **Impactos da Covid-19 no Transporte Público**. Associação Nacional de Transportes Públicos, 2020.
- [15] SCIELO. Impactos da Covid-19 na Mobilidade e na Acessibilidade. Acesso em 2025.
- [16] FUNDAÇÃO SEADE. **Comércio Exterior dos Municípios Paulistas**. Repositório de Dados. Fonte: MDIC/Siscomex. Disponível em: <<https://repositorio.seade.gov.br/dataset/comercio-exterior>>. Acesso em: 05 dez. 2025.
- [17] SÃO PAULO (Estado). Secretaria de Meio Ambiente, Infraestrutura e Logística. **Volume Transportado nas Travessias Litorâneas**. Portal de Dados Abertos. Disponível em: <<https://dadosabertos.sp.gov.br/dataset/volume-trav-lit>>. Acesso em: 05 dez. 2025.
- [18] PEDREGOSA, F. et al. Scikit-learn: Machine Learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825-2830, 2011.
- [19] STITCH FIX. Sorry ARIMA, but I'm Going Bayesian. 2016.
- [20] TAYLOR, S. J.; LETHAM, B. Forecasting at Scale. **The American Statistician**, v. 72, n. 1, p. 37-45, 2018.
- [21] A TRIBUNA. **Filas sem fim e horas de espera: travessia de balsas entre Santos e Guarujá vive caos**. Publicado em: 04 dez. 2024.
- [22] PEDROSA, R. A. Os impactos da relação porto/cidade na cidade de Santos. Periodicos Unisanta.
- [23] CHEN, T.; GUESTRIN, C. XGBoost: A Scalable Tree Boosting System. In: **KDD '16**. ACM, 2016. p. 785-794.