



# **DATA SCIENCE E POLÍTICA PÚBLICA: PROPOSTA METODOLÓGICA PARA A QUALIFICAÇÃO DE POLÍTICAS PARA O DESENVOLVIMENTO LOCAL/REGIONAL**

**DATA SCIENCE AND PUBLIC POLICY: A METHODOLOGICAL PROPOSAL  
FOR QUALIFYING POLICIES FOR LOCAL/REGIONAL DEVELOPMENT**

# DATA SCIENCE E POLÍTICA PÚBLICA: PROPOSTA METODOLÓGICA PARA A QUALIFICAÇÃO DE POLÍTICAS PARA O DESENVOLVIMENTO LOCAL/REGIONAL

## DATA SCIENCE AND PUBLIC POLICY: A METHODOLOGICAL PROPOSAL FOR QUALIFYING POLICIES FOR LOCAL/REGIONAL DEVELOPMENT

Victor da Silva Oliveira<sup>1</sup> | Tiago Costa Martins<sup>2</sup>

Arlindo Figueirôa Escobar Teixeira de Oliveira<sup>3</sup>

Recebimento: 17/01/2023

Aceite: 31/10/2023

<sup>1</sup> Doutor em Geografia (UFPE).  
Professor na Universidade Federal do Sul e Sudeste do Pará. Xinguara – PA, Brasil.  
E-mail: victorsoliveira@hotmail.com

<sup>3</sup> Doutorando em Ciência da Informação (UFPE).  
João Pessoa - PB, Brasil  
E-mail: arlindo.escobar@gmail.com

<sup>2</sup> Doutor em Desenvolvimento Regional (UNISC). Professor na Universidade Federal do Pampa. São Borja – RS, Brasil.  
E-mail: tiagomartins@unipampa.edu.br

## RESUMO

Postular uma intervenção na realidade local e regional com a pretensão de qualificar um ou mais atributos individuais, sociais e espaciais permeia uma gama de inquietações. Elas vão desde a concepção de desenvolvimento, passam pela definição das estratégias para a intervenção e chegam à definição de prioridades de investimentos — não apenas monetários — para atingir os objetivos elencados. Essas questões estão presentes nos bancos acadêmicos e nos formuladores de políticas públicas. Diante desse contexto e através da aplicação de *Data Science*, o presente artigo visa propor uma metodologia para qualificar políticas públicas que visem ao desenvolvimento local/regional a partir da ampliação da renda regional disponível. A proposta parte da releitura de clássicos do desenvolvimento econômico e regional com o pressuposto de que a diversificação produtiva pode ser um canal para ampliar a renda regional disponível. Utilizando metodologias de *Data Science* para mineração, sistematização e sintetização de informações espaciais atreladas às técnicas de análise regional, possibilita-se que gestores e acadêmicos identifiquem estratégias que para o desenvolvimento local e regional por meio de políticas públicas. A proposta traz alguns desafios, como a qualificação técnica e de recursos humanos para a sua efetiva aplicação, assim como a convergência para além da técnica que engloba a definição de uma política pública. Não obstante, a potencial aplicação dessa metodologia possibilita uma averiguação com considerável grau de acuidade de descrição e diagnóstico regional, atingindo a predição e a prescrição de alternativas viáveis para a transformação local/regional.

**Palavras-chave:** Análise regional. Desenvolvimento local e regional. Política local e regional.

## ABSTRACT

This article posits a methodological intervention to enhance various individual, social, and spatial attributes within local and regional settings. Concerns span from conceptualizing development to strategizing interventions and defining both monetary and non-monetary investment priorities to achieve stated goals. These topics are widely discussed in academia and among public policymakers. Utilizing Data Science, the paper proposes a method to qualify public policies that promote local/regional development by increasing regional disposable income. This approach revisits classical economic and regional development theories, suggesting that productive diversification may elevate regional disposable incomes. Data Science methods applied to spatial information mining, systematization, and synthesis—coupled with regional analysis techniques—equip managers and academics with tools to identify public policy strategies for local and regional development. The proposal introduces challenges, including the need for technical and human resource enhancement for effective implementation, and requires a convergence that extends beyond technique to the essence of public policy formulation. Nevertheless, applying this method can precisely describe and diagnose regional characteristics, thereby predicting and prescribing sustainable alternatives for local/regional transformation.

**Keywords:** Regional analysis. Local and regional development. Local and regional policy.

## INTRODUÇÃO

Diversas reflexões sobre o desenvolvimento local e regional foram e continuam sendo travadas no meio acadêmico em escala mundial. No caso brasileiro, Theis (2022) indagou sobre a noção aplicada nos Programas de Pós-Graduação que tratam da temática, ressaltando, em sua conclusão, a abrangência e a incapacidade de conciliar efetivamente ponderações que partam da base social, do capital e do Estado.

A questão se torna ainda mais complexa quando, para além dos bancos acadêmicos, depara-se com a prática social. As tentativas brasileiras de aplicar políticas de desenvolvimento local/regional, explícitas ou implícitas (Araújo, 2013), no passado e no presente, apresentam um leque diverso de concepções e modos de aplicação. Esse fato enaltece a riqueza dos conceitos, assim como faz surgir um desafio para qualquer postulante a intervir na realidade com o objetivo de modificar o *status quo* individual, social e espacial.

É na intersecção do desafio conceitual e prático que o presente artigo busca propor uma metodologia, através da aplicação de *Data Science*, para qualificar políticas públicas que objetivem o desenvolvimento local/regional a partir da ampliação da renda regional disponível. Buscou-se a



formulação de uma proposta baseada em dados reais que possibilitassem a formulação de conclusões preditivas e prescritivas que contribuíssem para a qualificação do desprendimento de esforço coletivo em vistas ao desenvolvimento local e regional por meio de políticas públicas.

O presente artigo está estruturado em quatro seções, além desta introdução e das considerações finais. Na próxima seção, é apresentada a construção teórica, alicerçada em clássicos da reflexão sobre o desenvolvimento econômico e regional, com um rearranjo que visa considerar a intenção de ampliação da renda regional disponível por meio da diversificação produtiva. Na seção seguinte, é tratado, teórica e metodologicamente, como a identificação pode ser realizada de modo estratégico. Posteriormente, de modo detalhado, é apresentada a metodologia proposta (baseada na construção teórica anterior). Nessa seção, são trazidas técnicas de análise regional e bancos de dados com informações que podem ser consultadas para, objetivamente, materializar os pressupostos teóricos apresentados. Não obstante, ainda na mesma seção, são tratados mecanismos de *Data Science* para a aplicação das técnicas e otimização dos dados citados. Por fim, na seção de resultados e discussões, são abordadas questões concretas que desnudam o potencial uso da metodologia apresentada, bem como são apresentadas algumas limitações e pontos de acuidade necessária.

## UMA ABORDAGEM DE DESENVOLVIMENTO LOCAL/REGIONAL

O título desta seção, propositalmente, anuncia “uma” com a intenção de sublinhar ao leitor o caráter restritivo que a proposta aqui apresentada assume. As teorias do desenvolvimento econômico e regional não são recentes nem parcas em quantidade. Para além desse debate, o rol de interpretações sobre os determinantes do desenvolvimento de atividades econômicas em um local/região é rico em diversidade.

Ao assumir a gama de teorias do desenvolvimento, galgando uma construção inédita para interpretar uma realidade específica, pode-se esbarrar em costumeiros resultados. Um desses — e talvez o mais emblemático — é o lançamento de “um novo arcabouço teórico-metodológico”. Em resultados desse tipo, pesquisadores com olhares mais atentos podem perceber extensas vinculações com teorias consagradas, mas são apresentadas embrulhadas em novas roupagens que transmitem uma atmosfera de novidade quando, na verdade, estão assentadas em concepções conhecidas.



Logo, para a construção do marco teórico sobre o desenvolvimento local/regional da metodologia aqui proposta, parte-se dos seguintes processos: (i) releitura de clássicos (operacional); (ii) atenção à diversificação produtiva (pressuposto); e (iii) ampliação da renda regional disponível (intenção). A partir dessas três considerações, é realizado o exercício de construção de um marco teórico que fundamenta uma proposta metodológica de análise local/regional com atributos novos. Mas isso ocorre não pela construção de uma nova teoria, e sim pelo rearranjo de concepções selecionadas de teorias consagradas do desenvolvimento econômico e regional. Isso tem como finalidade atender o objetivo do desenvolvimento por meio da diversificação socioeconômica e da ampliação da renda regional disponível<sup>1</sup>.

Como apontado, um dos pressupostos é a busca pela diversificação produtiva local/regional com sustentabilidade econômica ao longo do tempo, fomentando políticas públicas que ampliem a renda. Essa preocupação é um importante direcionamento sobre quais teorias do desenvolvimento econômico e regional podem ser revisitadas a fim de que subsídios para formulação do marco teórico sejam encontrados. A proposta de análise regional apresentada por Paiva (2013), com base nas indicações de North (1955; 1959), questiona o tipo de investimento induzido pelo bem de exportação e sua real capacidade de geração de economias externas.

Admite-se, como pressuposto, que a renda disponível ( $Y_d$ ) de um município/região é função de duas principais variáveis: o valor das exportações ( $X$ ) — fruto da comercialização de bens e serviços para fora<sup>2</sup> — e o valor da distribuição da renda ( $Y$ ). Como analisado por North (1955; 1959), setores de exploração de bens naturais tendem a apropriar grande parte da produção em forma de lucro, enquanto outra parcela menor é paga em impostos e outra, ainda menor, é apropriada como renda pelos trabalhadores.

Desse modo, a questão posta anteriormente pondera uma margem maleável percentualmente restrita ao escopo local/regional, haja vista que a apropriação de lucros e o massivo investimento no setor ocorrem com a importação de bens de outras regiões do país ou do exterior.

---

1 Boisier (2000) afirma que existe uma região em “términos reales” e outra em “términos potenciales”. Ou seja, há potencialidades regionais que podem ser fomentadas com vistas a um determinado desenvolvimento.

2 Nesse caso, exportação não é assumida necessariamente como uma ação internacional (para outro país), mas também como uma ação para fora do local/região.



O encadeamento direto e indireto induzido pelo setor exportador passa a ser preponderante. Há, entretanto, grandes diferenças entre os setores e sua capacidade de encadeamento próximo ao local da atividade principal. Vale pontuar as reflexões trazidas por Hirschman (1958) que, em suma, diferenciam a exportação de toras de madeira e a exportação de móveis, por exemplo.

Nesse ínterim, North (1955; 1959) postula que a chave para o desenvolvimento estaria na soma da renda das exportações e do potencial por ela ativado na geração de um mercado interno — ‘X + Yd’. Logo, a questão paira sobre quais setores têm essa dupla capacidade. Encontra-se parte da resposta em Hirschman (1958), que trata da existência de setores/atividades capazes de gerar maior encadeamento.

A partir de North (1955; 1959), pode-se aprofundar essa questão. O autor diferenciou renda interna disponível (Yd) de renda regional disponível (Yrd) por meio da propensão à importação (m) de cada segmento econômico. De modo genérico, em um município/região cuja base de exportação são bens naturais, há importação de bens de capital (I) e de bens de consumo de alto valor agregado (Ck) — pagos com a capitação de lucros e altos salários. Em síntese, nessas economias, a propensão a importar é a soma de I e Ck ( $m = I + Ck$ ). Em contrapartida, os bens de exportação (X) e os bens consumidos pelos trabalhadores (Cw) são produzidos localmente, com propensão baixa a importar. Nesse modelo, a renda regional disponível é representada pela seguinte função:

$$Yrd = (X + Cw + Gi) - m (Ck + I + Cw + T)$$

Onde:

X – Valor das Exportações

Cw – Bens Consumidos pelos trabalhadores

Gi – Gastos Governamentais

m – Propensão a Importar

Ck – Bens de Consumo capitalista

I – Bens de Capital

T – Impostos Pagos não retornados



## ABORDAGEM ESTRATÉGICA PARA AMPLIAÇÃO DA RENDA REGIONAL DISPONÍVEL

Diante dessa breve exposição, questionam-se quais estratégias podem ser adotadas para a ampliação da renda regional disponível e que, efetivamente, gerem maximização dos ganhos sociais através de um processo de diversificação socioeconômica. Como adiantado, a própria natureza de cada segmento demonstra caminhos e estratégias a seguir. Para garantir a permanência da renda gerada na região, o foco deve ser em ações com: (i) baixa propensão a importar, como no caso de bens de consumo dos trabalhadores que têm determinada expertise prévia; (ii) capacidade para se tornar produto de exportação — leia-se para fora do município/região; (iii) potencial de encadeamento próximo; e (iv) possibilidade de ser alvo factível de políticas para um projeto de planejamento e desenvolvimento local/regional.

Assim, os dois primeiros elementos da base teórica advêm de duas correntes teóricas — North e Hirschman — e da capacidade de determinadas atividades econômicas se tornarem vetores centrais de um processo de diversificação produtiva com sua comercialização externa e com o impulso para atividades adjacentes no interior do município/região<sup>3</sup>. Outrossim, as bases teórico-metodológicas que jogam luz sobre como e em quais segmentos intervir para garantir o desenvolvimento ao conjunto local perpassam.

Smith (1983), em questão trabalhada por Paiva (2013), aborda o *loop smithiano* e as possibilidades de superá-lo. Segundo o Paiva (2013), a riqueza de um determinado local é fruto direto da possibilidade de ampliação da produtividade do trabalho. Toda e qualquer atividade produtiva que almeja ganhar novos mercados consumidores necessita de competitividade, a qual, em suma, é representada pela oferta de bens e serviços de valor agregado pelo menor custo possível, ou seja, produzir mais, em menos tempo, com qualidade e baixo custo.

Esse esforço, por sua vez, é resultado da divisão do trabalho, que tem como limite o tamanho do mercado consumidor para os produtos, criados com índices de produtividade cada vez melhores. Assim, o tamanho do mercado é uma condição *sine qua non* da ampliação da divisão do trabalho. Para superar esse *loop*, a ampliação dos mercados, atingida pelo aumento da produtividade do trabalho, traz consigo a necessidade de adesão de novos consumidores. Em outras palavras, exportar para fora do local/região.

<sup>3</sup> As metodologias sobre como a identificação desses setores pode ser realizada e como pode ser considerada na qualificação das políticas públicas são apresentadas em tópico posterior.



Como indicado anteriormente, essa estratégia deve privilegiar segmentos que, em um primeiro momento, agreguem à renda regional disponível a partir da identificação e do fomento de cadeias produtivas, conforme tratado por Marshall (1996). Com o amadurecimento das produções advindas da ampliação da produtividade e do acúmulo de expertise, será gerada uma especialidade com vantagem em relação aos municípios/regiões do entorno, possibilitando que o produto/serviço seja um novo segmento a ser exportado<sup>4</sup>. Essa proposição, tal como reflete North (1955; 1959), reposiciona o município como ofertante de bens e serviços numa nova hierarquia urbana e regional. Como resultado, no entorno de um conjunto de atividades, criam-se diversas engrenagens que se retroalimentam em forma de cadeia produtiva.

O movimento de dinamização da renda a partir da manutenção da demanda elevada é, sobretudo, uma das principais contribuições de Keynes (1985) às reflexões de economia do desenvolvimento. O autor trata da atuação do Estado enquanto agente ativo, principalmente em momentos de crise. De fato, a atuação do poder público por meio de ações estrategicamente definidas pode exercer importante papel no incremento de renda em nichos do mercado que potencializam a renda regional disponível.

Essa estratégia estatal pode ser considerada a partir de duas formas de atuação: (i) direta ou indireta; (ii) pela demanda ou pela oferta. Elas não excluem-se entre si, tampouco projetam, em convergência, estratégias para o planejamento de políticas públicas para estimular a renda regional disponível de modo direto ao atacar a demanda de setores que, com base nas concepções anteriormente afirmadas, podem estimular a renda regional disponível por meio, por exemplo, de gastos públicos municipais<sup>5</sup>. Pela via da oferta, o poder público é agente central no processo de estruturação dos fluxos de mercadorias, informações e pessoas que impactam diretamente os custos de produção e são centrais em um processo de ganho de produtividade — como abordado por North (1959) e Smith (1983). Por fim, de modo indireto, o poder público do município/região pode aproximar agentes por

---

4 Entende-se como vendido para locais externos ao município/região. Vale destacar que existem dois tipos de produtos/serviços exportados: o que é produzido no local/região e comercializado para fora e que é produzido no local/região e o consumidor de fora se desloca até o local para consumir.

5 Deve-se ponderar a capacidade de atuação de cada município, sobretudo de modo direto. A envergadura de um investimento direto de um município com mais de um milhão de habitantes é muito distinta de um município com menos de 50 mil. Não obstante, aqueles que recebem *royalties* de qualquer origem pela exploração de algum bem natural também se distinguem, independentemente no número de habitantes.

vezes afastados do sistema produtivo e que podem auxiliar na qualificação da produção, ampliando a produtividade por meio da inovação.

Soma-se à discussão a contribuição de Schumpeter (1988) sobre o protagonismo da inovação no processo produtivo e na dinamização das economias. O autor trata da incorporação de novos mercados como um dos cinco pilares que o balizam, aproximando a segunda estratégia de atuação do Estado — aqui entendido a partir do poder municipal.

A compreensão da formação socioespacial como teoria e método (Santos, 1978) atribui protagonismo ao conceito de local/região quando se fala de uma proposta de intervenção que tem o objetivo de potencializar e/ou transformar atividades produtivas. Para tanto, deve-se partir do princípio de não isolamento regional e da necessidade de se apoiar no movimento histórico-cultural e econômico em que o(s) município(s) está(ão) inserido(s), com a possibilidade de transbordos regionais para o conjunto social em três dimensões: espacial, social e individual (Boisier, 2000).

Assim, parte-se do pressuposto de um fomento a atividades produtivas que configurem uma diversificada base e gerem renda regional somada às atividades já consolidadas de exportação. Diante disso, a proposta de análise regional apresentada em Paiva (2013) indica a distinção entre dois grupos de atividades econômicas a partir da sua função dinâmica.

O primeiro grupo corresponde às atividades propulsivas, cujas características são voltadas à obtenção de renda básica para a economia local. São atividades que produzem algo vendido para fora — externo ao município/região — em troca de um valor monetário que passa a circular e mobiliza a economia local. Há também tipos específicos de atividades e serviços voltados ao consumo no local por pessoas de outros municípios, assim como as atividades governamentais mantidas por impostos oriundos de fora do município/região<sup>6</sup>.

Já o segundo grupo diz respeito a atividades com função dinâmica distinta, mas complementares à economia local: são as atividades reflexas. Elas são voltadas ao abastecimento do mercado local, ou seja, ao comércio varejista e aos serviços básicos de atendimento ao munícipe<sup>7</sup>. A depender da entrada de recursos por meio das atividades propulsivas, essas atividades tenderão

6 Todas essas atividades compõem parte da primeira função exposta anteriormente, que define a renda interna da região -  $(X + Gi)$ .

7 Representado na função anterior pelo  $Cw$ .



a ganhar vigor, ou seja, reflexo. Ainda, há as atividades mistas, cuja natureza se encontra entre as propulsivas e as reflexas.

Analisar cada tipo de atividade de função dinâmica tem objetivos estratégicos diferentes. As atividades propulsivas podem ajudar a compreender o nível atual de especialização produtiva em atividades com real capacidade de incrementar renda circulante no local. O diagnóstico de apenas uma atividade sobressaliente, por exemplo, pode indicar uma potencial satelitização da economia local. Assim, seu contraponto seria uma economia local diversificada, autônoma e crescente, com diversos setores propulsivos em destaque — indicam-se os caminhos a serem fomentados por ações com essa pretensão.

Já a revisão das atividades reflexas busca compreender o nível de evasão da renda básica. Um local/região com poucas atividades é, sobretudo, um indicativo de que os consumidores potenciais que aqueceriam o mercado local estão consumindo em outros locais. Ainda, pode indicar que o poder de compra das famílias é baixo e não há potencialização do comércio, havendo, portanto, geração de renda e riqueza que não alimenta a inclusão.

## **PROPOSTA METODOLÓGICA: ANÁLISE REGIONAL E DATA SCIENCE PARA POLÍTICA PÚBLICA LOCAL/REGIONAL**

A partir de tais pressupostos, para a apresentação do “mapa metodológico” sugerido neste estudo, é preciso se ater à base empírica. A proposta do estudo inicia com a utilização do banco de microdados da Relação Anual de Informações Sociais (Rais) do Ministério da Economia, utilizando a Classificação Nacional de Atividades Econômicas 2.0 (CNAE), a qual classifica as atividades econômicas em 1.357 subclasses. O foco desta metodologia pode ser local (municípios) ou regional.

As revisões propostas devem fazer análises comparativas típicas de uma análise regional, uma vez que só há clareza da dimensão dos indicadores encontrados se comparados com realidades circunvizinhas, regiões inseridas e o próprio Estado. A proposta do estudo é utilizar o recorte de regiões de comparação dado pela última regionalização do Instituto Brasileiro de Geografia Estatística (IBGE), o qual indica regiões intermediárias, os municípios que as compõem e suas posições na hierarquia urbana.



Os municípios de comparação, integrantes das regiões intermediárias ou adjacentes, podem ter como indicativos para analogia os municípios com posição hierárquica urbana e regional superior e, no caso de análise de uma região, a comparação com outras regiões e os indicadores estaduais e nacionais. No âmbito dos microdados da Rais, são tratadas três informações-base que servem para outros momentos desta metodologia.

A primeira diz respeito aos vínculos formais das atividades econômicas por segmento. Esse dado é comparado proporcionalmente ao estadual, com a finalidade de perceber se há especialização produtiva em cada um dos setores por meio do Quociente locacional (QI)<sup>8</sup>. Segundo North (1955; 1959), quando resultado for superior a 1, o QI indica que, proporcionalmente, determinada atividade econômica está gerando empregabilidade no local de análise acima da média e do padrão vigente no Estado — no caso, o território de referência. Logo, é um sinal de que essa atividade não se destina apenas ao mercado interno, sendo, portanto, uma atividade propulsiva. O QI é dado pela seguinte equação:

$$QI_{ij} = \frac{\left[ \frac{E_{ij}}{\sum_j E_{ij}} \right]}{\left[ \frac{\sum_i E_{ij}}{\sum_j \sum_i E_{ij}} \right]}$$

Em que:

$QI_{ij}$  = Quociente locacional da classe  $i$  no município  $j$ ;

$E_{ij}$  = Emprego na classe  $i$  no município  $j$ ;

$\sum_j E_{ij}$  = Emprego em todas as classes no município  $j$ ;

$\sum_i E_{ij}$  = Emprego na classe  $i$  na região de referência;

$\sum_j \sum_i E_{ij}$  = Emprego de todas as classes na região de referência.

A segunda informação baseada nos microdados da Rais são os salários pagos aos trabalhadores formais de cada uma das atividades de acordo com sua função dinâmica — propulsivas ou reflexas. Ter ciência do incremento de renda de cada segmento vai além do valor monetário circulante que cada setor produtivo insere na economia local em forma de salário aos trabalhadores. Segmentos

8 Para mais informações sobre o uso dos QIs para a análise regional aplicada ver: Bitencourt e Guimaraes (2012); Mattei e Mattei (2017); e Crocco *et al.* (2006).

propulsores, tendo em vista sua especialização e capacidade produtiva, tendem a pagar salários superiores ao conjunto das atividades reflexas. Somado aos indicadores de análise regional, esse dado potencializa a identificação de cada atividade econômica segundo sua função dinâmica; além disso, fundamenta a seleção de atividades a serem promovidas com vistas à ampliação da renda regional disponível.

Por fim, no âmbito da Rais, extraem-se informações sobre os estabelecimentos e suas principais características a partir da classificação CNAE 2.0. Elementos como a natureza, o tamanho do estabelecimento e a quantidade de trabalhadores são relevantes na exposição de como cada segmento é estruturado no município. Em paralelo, essa identificação permite entender a dimensão das empresas atuantes em cada função dinâmica, visto que estratégias distintas são necessárias para micro, pequenas, médias e grandes empresas.

Com finalidade semelhante às descritas, busca-se compilar informações sobre os microempreendedores individuais (MEIs). Esse esforço busca sanar parcialmente o limite dos microdados da Rais, que apontam indicadores apenas do trabalho formal. Dessa forma, agregar os MEIs potencializa a análise em termos de generalidade do estudo, uma vez que, para além do recurso técnico/metodológico, é crescente o número de cadastros de MEIs no Brasil.

O QI também deve ser buscado para atividades eminentemente rurais, como pecuária e agricultura. Porém, o índice de informalidade típico dessas atividades fragilizaria a obtenção de resultados claros por meio dos microdados da Rais. Como alternativa, indica-se utilizar dados da produção rural do Censo Agropecuário IBGE, como o valor da produção animal, o efetivo rebanho e as produções de lavoura temporária e permanente.

Além disso, os dados de exportação internacional de produtos do município/região corroboram a verificação da especialização produtiva. Parte-se do pressuposto de que a entrada de determinado produto no amplo e concorrido mercado internacional é sinal de um elevado índice de produtividade. Do mesmo modo, partes significativas de insumos e maquinários para atividades com grande produtividade são adquiridas do estrangeiro; assim, dados de importação potencializam sua identificação.

Os microdados do Ministério da Economia referentes ao comércio exterior proporcionam acesso a informações como volumes exportados e importados, seus respectivos valores em dólar, países parceiros e a transformação da inserção municipal/regional desde 1989.



Para além da especialização produtiva em si — em que o QI tem protagonismo — outros quatro indicadores complementares são utilizados com razões específicas. O Índice de Herfindahl-Hirschman (IHH) é usado para determinar a concentração de um respectivo setor no município, ou seja, o QI indica se há especialidade. Já o IHH informa se esse setor é marcado por um poder de atração devido à especialização ou, inversamente, um baixo poder de atração. Esse indicador se refere ao nível de concorrência entre os setores, configurando-os de modo gradual, que varia desde a concorrência perfeita até o monopólio. Para essa metodologia, como discutido anteriormente, busca-se verificar setores com potencial de geração de produtos e serviços para exportação que, paralelamente, não tenham tendência à concentração, haja vista sua menor capacidade de impacto na renda regional disponível. A equação do IHH é dada da seguinte forma:

$$IHH = \frac{P_{ij}}{P_{it}} - \frac{P_{tj}}{P_{tt}}$$

Em que:

$P_{ij}$  = Setor i do município j;

$P_{it}$  = Total do setor i;

$P_{tj}$  = Total do município j;

$P_{tt}$  = Total da região de referência.

Outro indicador associado a questões de especialização é o coeficiente de localização. Ele varia entre zero e um, possibilitando identificar a dispersão entre os vários setores da atividade econômica e classificando aqueles com maior ou menor tendência de concentração espacial. A classificação é realizada em três grupos: os mais bem dispersos, próximos ao resultado de zero; os intermediários; e os com setores de muita concentração, com resultado próximo a um. A equação do coeficiente de localização é a seguinte:

$$CL = \frac{(j_{ei} - \sum_i j_{ei})}{2}$$

Em que:

$j_{ei}$  = Emprego no i no município e;

$\sum_i j_{ei}$  = Emprego de todas as classes na região de referência.



Ainda, para verificar a diversificação e a especialização das atividades produtivas, é utilizado o indicador de coeficiente de especialização. Ele também varia entre zero e um e verifica se o município/região tem estrutura produtiva semelhante à região de referência. Se o resultado for próximo a um, ou seja, com estrutura produtiva diferente da região de referência, o setor produtivo será considerado especializado, pois apresenta uma elevada presença em paralelo ao presente na região de referência. A equação utilizada é a seguinte:

$$C E_i = \sum_i \frac{(i_{ej} - \sum_j i_{ej})}{2}$$

Em que:

$\sum_i$  = Emprego de todas as classes na região de referência;

$i_{ej}$  = Participação do setor e no município j;

$\sum_j i_{ej}$  = Participação percentual do setor i na região de referência.

Para verificar se a associação entre a distribuição geográfica entre diferentes setores é semelhante, utiliza-se o coeficiente de associação geográfica. Esse indicador, que varia entre zero e um, mostra se os padrões locais têm associação geográfica e, se sim, em qual grau. Ao realizar a análise conjunta com o IHH, espera-se compreender a complementaridade entre os diferentes setores da estrutura produtiva, especialmente aqueles que exigem a existência do outro para realização com maior eficiência, aproximando-se, assim, da compreensão de atividades propulsivas e reflexas. A equação utilizada é a seguinte:

$$CAG_{ik} = \sum_j \frac{(j_{ei} - j_{ek})}{2}$$

Em que:

$\sum_i$  = Emprego de todas as classes na região de referência;

$j_{ei}$  = Participação do setor i no município e;

$j_{ek}$  = Participação do setor k no município e.

Com a obtenção dos resultados dos cinco indicadores descritos, é possível verificar, com elevado grau de precisão, a estrutura produtiva do local/região em estudo, assim como vislumbrar os segmentos propícios às intervenções estratégicas de políticas públicas para a diversificação

produtiva. A partir da aplicação dessas técnicas de análise regional e com base nos pressupostos teóricos anteriormente apresentados, estarão sendo expostos a descrição e o diagnóstico da estrutura produtiva, além dos primeiros indicativos situacionais da retenção e distribuição da renda regional disponível. O resultado traz indícios dos setores de atividades com potencial para indução do desenvolvimento.

Em síntese, os cinco indicadores possibilitarão perceber os seguintes aspectos (Quadro 1):

**Quadro 1** | Síntese interpretativa dos indicadores selecionados.

Indicador	Síntese dos resultados
Quociente locacional	Localização fraca / média / significativa
Índice de Herfindahl-Hirschman	Concorrência perfeita / monopolização
Coefficiente de localização	Dispersão / concentração
Coefficiente de especialização	Diversificação / especialização
Coefficiente de associação geográfica	Associação fraca / média / significativa

Fonte: Produzido pelos autores.

Com as leituras obtidas a partir dos indicadores, resgatam-se os pressupostos de Keynes (1985) ao indicar o estado — no caso em tela representado pelos municípios — como vetor de indução para o desenvolvimento. Operacionalmente, podem ser utilizadas duas estratégias: de atuação direta/indireta e de demanda/oferta. Elas fazem uso de dois bancos de dados.

O primeiro é o Portal da Transparência do(s) município(s) em análise. Ele disponibiliza, para acesso público, todos os empenhos realizados por todos os órgãos de competência municipal. Desse banco de dados, devem ser segmentados os empenhos por tipo de atividade/serviço, de modo paralelo aos segmentos estratégicos diagnosticados na etapa anterior. Além disso, os valores gastos de acordo com o município de origem do produto/serviço contratado devem ser mapeados.

Para essa atividade, é utilizado o segundo banco de dados a partir dos CNPJ dos contratados — disponibilizados no Portal de Transparência — e sua posterior busca junto às informações da Receita Federal sobre a localização do empreendimento e a sua atividade econômica principal e secundária, classificadas pelo CNAE. A intenção dessa etapa é gerar uma matriz de possibilidades de empenhos realizados fora do município/região que poderiam, juntamente às ações indiretas, convergir e induzir setores estratégicos para a geração de renda interna regional e fomentar a diversificação e a especialização socioeconômica municipal/regional.



Ainda, no corte passível de atuação dos governos municipais na potencialização do seu desenvolvimento, quando as atividades com potencial para ampliação da renda regional disponível são identificadas, o Estado pode atuar como interlocutor entre os segmentos produtivos e os centros de competências científicas e tecnológicas presentes nas ICTs. Essa ação visa facilitar a combinação entre as demandas dos setores público e privado com aquilo que é produzido na academia.

A prospecção pode ser realizada por meio do Diretório de Grupos de Pesquisa do CNPq (DGP-CNPq) e sistematizada com informações como: líder do grupo; área do conhecimento; ICT do grupo; endereço; contatos (e-mail e telefone); resumo descritivo das atividades desenvolvidas; linhas de pesquisa; redes de cooperação; recursos humanos; instituições parceiras; e equipamentos disponíveis. A partir dessas informações, o gestor público poderá ter acesso rápido e fácil a qualquer competência científica e tecnológica para fomento em seu município/região, ampliando a expertise e apostando em transformações e inovações que ampliem a produtividade do trabalho (Smith, 1983; Schumpeter, 1988).

Por fim, entre as fontes de dados que podem ser consultadas para a aplicação da presente metodologia, estão a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) e a Plataforma de Periódicos da CAPES. Há certo consenso de que essas são as duas principais bases de dados em que se encontra a maioria substancial da produção científica de intelectuais brasileiros e estrangeiros. No caso em tela, a revisão sobre a formação socioespacial do município/região de intervenção tem como alternativa a busca por quantidade e qualidade considerável de trabalhos já realizados sobre as distintas regiões brasileiras.

Todas as técnicas e bases de dados apresentadas até aqui são habitualmente utilizadas em análises regionais com foco na estrutura produtiva, como apontado no início desta seção. Invariavelmente, recorre-se a esses meios para a obtenção de interpretações e sínteses com elevado grau de representatividade da realidade analisada. No entanto, apesar de tratarem de aspectos relativamente próximos e bancos de dados complementares, as técnicas de análise desses indicadores são aplicadas majoritariamente de modo singular e com interpretações conjuntas realizadas pela “soma” das sínteses individuais. Apesar de essa prática ter importante relevância científica, há possibilidades de potencialização. É nessa lacuna que a *Data Science* adentra na



metodologia em tela.

*Data Science* é assumida como eixo estruturante da proposta metodológica pelo fato de realizar a interpretação desses indicadores de modo conjunto, e não de modo singular. Recentemente, esse ramo das ciências ditas “exatas” vem adentrando as ciências sociais e humanas. Consiste no estudo científico de dados e informações em grande volume, variedade, veracidade e velocidade de geração, que possibilitam analisar um determinado problema por diversos ângulos e, por fim, oferecer um produto com valor e resultado para sua solução (Marquesone, 2016).

A Figura 1 sintetiza o trajeto analítico com base na *Data Science*. Tradicionalmente, as duas primeiras etapas, descritivas e diagnósticas, são amplamente difundidas não apenas no âmbito do fazer científico, mas também na gestão pública e privada. Em anos recentes, com a intensificação da produção de dados e das capacidades operacionais de armazená-los e processá-los, há cada vez mais buscas por análises que visam prever o que acontecerá. Para isso, tomam-se por base condições historicamente postas e, posteriormente, o que é possível realizar, tendo por objetivo alterar resultados indesejados, como a partir de políticas públicas.

**Figura 1** | Sintetização da análise através da *Data Science*.



Fonte: Adaptado de Verma *et al.* (2016).

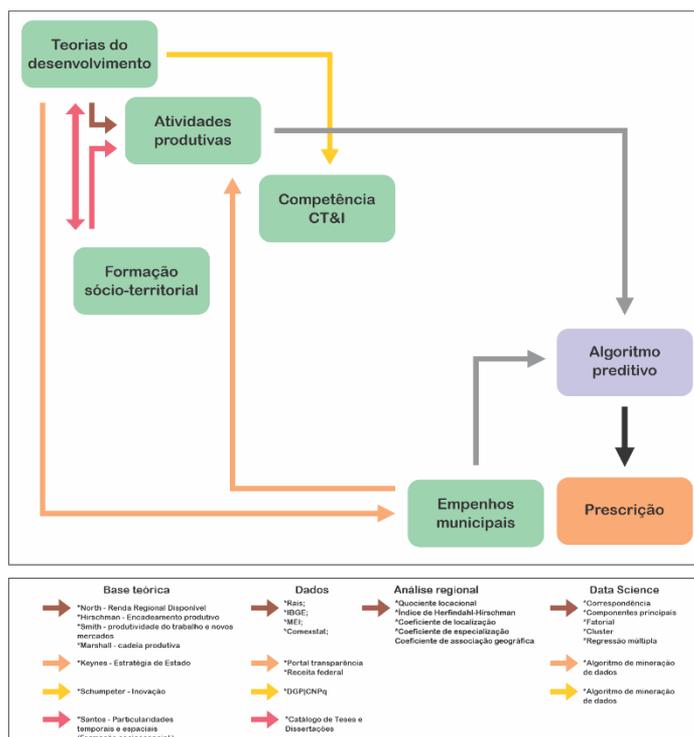
Inicialmente, como base teórica que fundamenta a proposta metodológica em torno da diversificação socioeconômica para o incremento de renda regional disponível por meio de políticas públicas diretas e indiretas, assumem-se alguns pressupostos provenientes de teorias do desenvolvimento e da análise regional.

As análises de *Data Science*, primeiramente em caráter exploratório, devem permear técnicas de correspondência, componentes principais, multifatoriais, cluster e regressão múltipla sobre os indicadores de análise regional para a busca das melhores possibilidades explicativas e preditivas, objetivando a diversificação e a ampliação da renda regional disponível.

O potencial do uso, primeiramente exploratório e posteriormente para obter indicações e conclusões das técnicas apontadas anteriormente, justifica-se pela ampla gama de indicadores brutos e agregados apontados como centrais para um diagnóstico preciso das funções dinâmicas do município/região. Para além da construção de cada indicador, a título de exemplo, a utilização das 1.357 subclasses (CNAE) para análise regional de modo desagregado ou indicadores selecionados é impensável sem artifícios técnicos como os possibilitados pela *Data Science* e apontados anteriormente.

A Figura 2 apresenta as concepções postas no tópico anterior, com o acréscimo das informações apresentadas nele.

**Figura 2** | Sintetização da proposta teórica e metodológica.



Fonte –Produzido pelos autores.

## RESULTADOS E DISCUSSÃO

A proposta metodológica apresentada explora, de modo particular, possibilidades desempenhadas pela *Data Science* na qualificação de políticas públicas destinadas a impulsionar o desenvolvimento local e regional (Boisier, 2000). Justificativas concretas sinalizam o potencial do uso dos mecanismos apresentados, assim como limitações e pontos de atenção em eventuais aplicações. Nesta seção, destacam-se tais questões: i) mineração e sistematização dos dados; ii) interposição de técnicas de análise regional; e iii) prescrição estratégica.

Quanto à mineração e sistematização, os dados indicados para consulta na construção da matriz analítica podem ser exemplificados pelos obtidos junto aos microdados da Rais e MEI. Neste caso, cita-se a Rais vínculos, havendo também o uso dos dados de estabelecimentos. Os microdados são apresentados por mesorregião do Brasil; assim, independente do(s) município(s) no qual se busca aplicar a metodologia aqui proposta, o pesquisador precisará trabalhar inicialmente com um número de vínculos elevado — na região Sudeste, por exemplo, havia quase 24 milhões de vínculos ativos em 2021 — e apresentados com aproximadamente 60 variáveis com mais de 10 mil possibilidades de atributos.

Nesse sentido, a *Data Science* se mostra uma ferramenta para exploração e aproveitamento dos bancos de dados, revelando potencial na área de mineração e sistematização de informações, sobretudo na identificação das funções dinâmicas de cada atividade econômica na região em análise, como exposto anteriormente. Com softwares específicos, como RStudio, o pesquisador pode realizar os filtros necessários para otimizar a pesquisa e a capacidade de uso de um hardware convencional para a atividade de pesquisa.

No âmbito da mineração de dados, sua obtenção nos portais da transparência municipais gera grande desafio, uma vez que eles não têm um padrão entre os municípios do país. Além disso, costumeiramente, não há mecanismos ágeis para obter uma grande quantidade de informações. Logo, a geração de algoritmos computacionais específicos para captar os dados de empenhos do(s) município(s), assim como para filtrar informações pertinentes para a pesquisa — como valores, destino, CNPJ das empresas contratadas, entre outros — mostra-se como elemento viável para a aplicação da metodologia.

Não obstante, a vinculação dos dados de CNPJ com o banco de dados da Receita Federal para verificação da atividade principal da empresa contratada e sua localização só é possível, via plataforma da receita, com a pesquisa de um CNPJ é por vez. O número grande de empenhos de uma prefeitura —



mesmo de pequeno porte — impossibilita a realização dessa atividade sem apoio de um mecanismo de mineração de dados automatizado como a *Data Science* proporciona. Nesse aspecto, esse dado auxilia no diagnóstico e evidencia possibilidade de redirecionamentos dos gastos públicos que podem induzir a manutenção de uma demanda efetiva (Keynes, 1985) local/regional, corroborando com os pressupostos da proposta teórica de diversificação produtiva.

Quanto à interposição de técnicas de análise regional, a *Data Science* se apresenta como um mecanismo para identificar tendências, padrões e correlações ocultas tanto em caráter exploratório, ao longo da pesquisa, quanto para a análise dos segmentos produtivos locais regionais — sobretudo, preditivos. Como salientado, a quantidade de dados e variáveis que a metodologia propõe utilizar é vasta, assim como a operacionalização de indicadores que contemplem a releitura de clássicos apontados nas primeiras seções deste texto.

A aplicação de apenas uma ferramenta estatística — como identificação de tendências — que englobe os indicadores de análise regional apresentados, com base em vários bancos de dados, um exercício complexo e de difícil realização por *softwares* e *hardwares* habilmente utilizados no dia a dia da pesquisa científica na área em questão. Logo, a interposição proposta é possível a partir da arquitetura de dados e do processamento em nuvem oferecidos pela *Data Science*.

De modo pragmático, outra potencialidade que auxilia a qualificação de políticas públicas a partir da metodologia proposta é a possibilidade de predição estatística de retorno à renda regional disponível — a intenção explicitada teórica e metodologicamente — ao executar determinadas ações estratégicas no âmbito das compras governamentais, por exemplo, para a identificação e fomento de atividades econômicas com potencial encadeamento (Hirschman, 1958) e que possam gerar cadeias produtivas (Marshall, 1996).

A prescrição estratégica encerra o potencial uso da metodologia apresentada. A tarefa de qualificação de políticas públicas para o desenvolvimento local/regional impõe, mesmo com os riscos postos ao pesquisador, a indicação pragmática de possíveis alternativas para obtenção dos resultados esperados. Munida da inflexão teórica, dos mecanismos de análise regional interpostos via ferramentas estatísticas e predições, a confecção de políticas públicas amplia a acuidade da alocação escassa de recursos públicos e minimiza esforços realizados por inspiração ou sem base teórica/técnica.



## CONSIDERAÇÕES FINAIS

A metodologia apresentada neste artigo é rigorosa e focada em questões técnicas, com base em um sólido arcabouço teórico e análises regionais fundamentadas em dados confiáveis e fontes qualificadas. Essa abordagem visa oferecer, ao poder público — seja em nível municipal ou regional — a capacidade de abordar questões específicas relacionadas ao desenvolvimento, com ênfase na ampliação da renda disponível e na melhoria dos aspectos individuais, sociais e espaciais.

A qualificação das políticas públicas a partir da aplicação da metodologia proposta pode ser encarada por espectros distintos que se complementam. Citam-se três. O primeiro é a ação estratégica para alocação de esforços e recursos públicos. Como pontuado ao longo do artigo, não são únicos os caminhos para a intervenção na realidade local/regional. A proposta trazida traça um aspecto; logo, havendo uma abordagem inicial translúcida, as trajetórias infindas da atuação do poder público encontram alternativas concretas que, ao serem fomentadas, podem gerar benefícios locais/regionais no médio e longo prazo.

Em segundo lugar, tem-se a possibilidade do acompanhamento concreto para a avaliação dos resultados alcançados. A partir dos indicadores preditivos de aplicação/retorno, criam-se mecanismos sociais de controle da atuação pública regional, uma vez que a metodologia oferece, de modo objetivo, os índices que se pretendem alcançar.

Por fim, considera-se a validade da aplicação da metodologia baseada em *Data Science* para a qualificação das políticas públicas destinadas ao desenvolvimento local/regional, dadas as suas capacidades de análise robusta, predição precisa e otimização de recursos. A *Data Science* permite mineração e sistematização eficazes de grandes volumes de dados, facilitando a identificação de tendências, padrões ocultos e correlações cruciais para uma análise regional abrangente.

Além disso, a capacidade de previsão estatística oferecida pela *Data Science* possibilita avaliar resultados esperados de ações estratégicas antes mesmo de sua implementação, tornando o processo de formulação de políticas mais informado e eficaz.

Entretanto, é importante reconhecer que essa proposta tem limitações inerentes. Ela se concentra, principalmente, nas fases técnicas da construção de políticas públicas, deixando de fora o processo inicial de decisão, os objetivos da intervenção e a implementação prática, que são



aspectos intrinsecamente políticos.

Sobre a aplicação prática dessa abordagem, enfrentam-se desafios substanciais, incluindo a necessidade de recursos humanos qualificados e meios técnicos adequados para lidar com a vasta quantidade de dados coletados, tabulados e processados. Não obstante, a disponibilização sem padrão de dados impossibilita o simples replicar de algoritmos de mineração e sistematização dos dados, havendo a necessidade de readequação.

Para estudos futuros, sugere-se uma maior integração entre abordagens técnicas e políticas, bem como a análise de casos práticos que demonstrem a eficácia dessa metodologia na promoção do desenvolvimento local/regional ao longo do tempo. Isso porque ações como as que se busca fomentar a partir da aplicação da presente metodologia ocorrem no médio e longo prazo.

## **AGRADECIMENTOS**

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela subvenção ao projeto de pesquisa (processo: 307567/2022-2).



## REFERÊNCIAS

- ARAUJO, Tania Bacelar de. **Brasil: territorialidade e políticas públicas: curso de ambientação para analista técnico de políticas sociais**. Brasília: ENAP, 2013.
- BITENCOURT, Rosimeire Sedrez; GUIMARAES, Lia Buarque de Macedo. Aplicação do coeficiente de Gini locacional e do Quociente locacional como apoio à delimitação geográfica de Sistemas locais de produção: o setor coureiro calçadista doRs. In: **XXXII Encontro nacional de engenharia de produção**. Bento Gonçalves/RS, 2012.
- BOISIER, Sergio. Desarrollo (Local): ¿ de qué estamos hablando? In: BECKER, Dinizar Fermiano.; BANDEIRA, Pedro Silveira. (Orgs.) **Desenvolvimento Local-Regional: Determinantes e desafios contemporâneos**, v. 1. Santa Cruz: Edunisc, 2000.
- CROCCO, Marco Aurélio. **Metodologia de identificação de aglomerações produtivas locais**. Economia e sociedade brasileiras. Revista nova econ. vol.16 n.2 Belo Horizonte,2006.
- HIRSCHMAN, Albert Otto. **The strategy of economic development**. New Haven: Yale University Press, 1958.
- KEYNES, John Maynard. **A teoria geral do emprego, do juro e da moeda**. Os Economistas. São Paulo: Nova Cultura, 1985.
- MARQUESONE, Rosângela. **Big Data: Técnicas e Tecnologias para Extração de Valores dos Dados**, 1. ed. São Paulo – SP: Editora Casa do Código, 2016.
- MARSHALL, Alfred. **Princípios de economia**. São Paulo: Nova Cultural, 1996.
- MATTEI, Taise Fátima; MATTEI, Tatiane Salete. **Métodos de Análise Regional: um estudo de localização**. Revista paranaense de desenvolvimento, Curitiba, v.38, n.133, 2017.
- NORTH, Douglass. **A agricultura no crescimento econômico**. Journal of Farm Economics, nº 41(5), 1959.
- \_\_\_\_\_. **Teoria da localização e crescimento econômico regional**. Journal of Political Economy, nº63(3), 1955.
- PAIVA, Carlos ÁguedoNagel. **Fundamentos da análise do planejamento de economias regionais**. Foz do Iguaçu: Editora Parqueltaipú, 2013.
- SANTOS, Milton. **Por uma geografia nova**. São Paulo: HUCITEC-EDUSP, 1978.
- SCHUMPETER, Joseph Alois. **Teoria do desenvolvimento econômico: uma investigação sobre lucros, capital, crédito, juro e o ciclo econômico**. Traduzido por Maria Sílvia Possas. 3. ed. São Paulo: Nova Cultural, 1988
- SMITH, Adam. **A riqueza das Nações: investigação sobre sua natureza e suas causas**. São Paulo: Abril cultural, 1983.
- THEIS, Ivo Marcos. **Hic et nunc: qual concepção de Desenvolvimento quando se trata de Desenvolvimento regional?**. Revista Brasileira de Estudos Urbanos e Regionais, v.24, 2022.
- VERMA, JaiPrakash *et al.* **Big data analytics: challengesandapplications for text, audio, video, andsocial media data**. InternationalJournalon Soft Computing, Artificial IntelligenceandApplications (IJSCAI), v.5, n.1, 2016.

