

## ARTIGOS

### INTELIGÊNCIA ARTIFICIAL NA SEGURANÇA INSTITUCIONAL DO PODER JUDICIÁRIO: ANÁLISE DE DADOS CRIMINAIS

#### ARTIFICIAL INTELLIGENCE ON THE INSTITUTIONAL SECURITY OF THE JUSTICE SYSTEM: CRIME DATA ANALYSIS

*Pedro Ivo Vasconcellos da Costa Pinto*

Mestre em Linguística (2014) Especialista em Data Science e Analytics (2023). Bacharel em Letras: Português/Literaturas (2010). Agente da Polícia Judicial no TRT-1ª Região.

**Resumo:** : Apresenta a implementação de dois algoritmos de aprendizado de máquina não supervisionado que foram utilizados na análise de dados relacionados à criminalidade no estado do Rio de Janeiro. O objetivo é explorar quais as possíveis aplicações das técnicas de mineração de dados utilizadas atualmente no campo da inteligência artificial na atividade de segurança institucional do Poder Judiciário. A análise segue metodologia quantitativa e resulta em diversos outputs relevantes para apoiar as decisões do gestor da Polícia Judicial do Tribunal Regional do Trabalho da Primeira Região, um dos resultados é um índice de criminalidade e a formação de agrupamentos que reúnem municípios semelhantes. Espera-se que a pesquisa abra caminho para que a inteligência artificial, especialmente seu componente de ciência de dados, possam encontrar seu lugar nas atividades de segurança do Poder Judiciário.

**Palavras-Chave:** Inteligência artificial. Segurança institucional. Análise de dados de criminalidade. Aprendizado de máquina.

**Abstract:** This paper describes the implementation of two unsupervised machine learning algorithms to analyze data on crimes that occurred in the state of Rio de Janeiro. The goal is to explore potential applications of data mining techniques, currently utilized in the extensive field of Artificial Intelligence, in Institutional Security. Working with data is viewed as a valuable resource to support the decision-making process. The methodology is primarily quantitative, yielding several significant outputs that can be utilized to support the planning and decision-making processes conducted by those responsible for Institutional Security at the Labor Court within the state of Rio de Janeiro. The results include a crime index, a corresponding ranking based on this index, and the creation of clusters that aggregate the most similar cities in terms of crime occurrences. It is expected this research opens the doors for the Artificial Intelligence, especially its data mining component, to gradually integrate into the security activities of the Brazilian Justice System.

**Keywords:** Artificial Intelligence. Institutional Security of the Justice System. Crime data analysis. Machine learning.

## 1. INTRODUÇÃO

A prestação do serviço jurisdicional pelo Estado, atividade-fim de qualquer corte de justiça, deve ser vista como um macroprocesso sustentado por outros processos, que contribuem para que a organização judiciária atinja seu fim de forma mais eficiente e efetiva. Embora a segurança institucional não seja considerada um macroprocesso de apoio independente na cadeia de valor de diversos Tribunais do Brasil, é inegável a relevância do seu papel na estrutura do Poder Judiciário,

Principalmente em um país marcado por uma constante crise de segurança pública, a segurança institucional nos tribunais busca garantir que a violência não influencie ou mesmo interrompa a prestação jurisdicional. Por isso, o CNJ editou a Resolução CNJ n. 383/2021, que situa a segurança institucional como atividade essencial ao Poder Judiciário, pois não há como promover a justiça se a lei do mais forte sobrepujar a lei oficial e a violência eclipsar a racionalidade e o anseio comum por uma sociedade justa e democrática. É essencial que não tenha espaço para a violência nos ambientes em que se promove e aplica a justiça. Por isso, é necessário que os serviços administrativos dos Tribunais da esfera

federal aprimorem constantemente sua área de segurança, que se mostra cada vez mais estratégica para a consecução dos processos finalísticos da organização judiciária.

Em que pese sua importância na proteção dos Magistrados como instâncias de decisão independentes e imparciais (SEABRA; SILVA, 2020), além dos próprios servidores e jurisdicionados, os métodos de trabalho da segurança institucional precisam acompanhar a realidade do ambiente o qual estão inseridos. Nos últimos anos, vem se consolidando no Poder Judiciário o emprego de técnicas de mineração e análise de dados que se situam no campo cada vez mais pervasivo da inteligência artificial (CNJ, 2019). Este artigo trata da ciência de dados como um ramo da Inteligência Artificial e sua aplicação na atividade de segurança institucional do Tribunal Regional do Trabalho da Primeira Região (TRT-RJ).

Parte-se do princípio de que a Inteligência Artificial é uma ciência multidisciplinar e que sua forma mais acessível consiste em empregar métodos de análise de dados inspirados em habilidades cognitivas humanas, como classificar e categorizar percebendo padrões, semelhanças e distinções entre elementos do mundo. Apresenta uma aplicação prática desses

métodos de análise no contexto da atividade de segurança institucional, demonstrando que ciência de dados é capaz de extrair informações relevantes sobre a violência no Estado do Rio de Janeiro, região em que atua o TRT-RJ, que, à primeira vista, não estavam de todo modo disponíveis em uma análise apoiada unicamente na cognição humana. Espera-se que essas informações possam ser utilizadas pelo gestor da área na tomada de decisões mais ágeis e assertivas, visando sempre a consecução da missão da segurança institucional do Poder Judiciário.

O método empregado pelo presente estudo é predominantemente quantitativo. A partir de uma base de dados de grande dimensão e quantidade de dados numéricos, aplicam-se cálculos estatísticos e algébricos para reduzir a dimensionalidade dessa base de dados e poder extrair informações que possam ajudar o gestor da área de segurança institucional do TRT-RJ a planejar as ações da Polícia Judicial<sup>1</sup>. Preliminarmente, são feitas considerações sobre o contexto em que as técnicas são empregadas e a que tipo de problema podem ajudar a resolver.

O trabalho com os dados segue o fluxo básico em projetos de ciência de dados: coletar, limpar, analisar e apresentar visualmente os resultados (WICKHAM; GROLEMUND, 2016). Optou-se por essa forma simplificada do Processo Padrão de Indústria Cruzada para Exploração de Dados (CRISP-DM), uma vez que a pesquisa não foi conduzida a pedido da organização, em um contexto real de trabalho, mas de modo independente com o intuito de apontar um caminho para a segurança institucional. Essa condição impede que as etapas de avaliação e de implantação sejam efetivamente postas em prática. Como consequência, um processo que é de natureza repetitiva por excelência (PROVOST; FAWCETT, 2016), sequer chegaria ao fim do seu primeiro ciclo. Portanto, manteve-se a escolha por um fluxo de pesquisa mais adequado às condições em que o trabalho se desenvolveu.

## 2. DESENVOLVIMENTO

### 2.1 A INTELIGÊNCIA ARTIFICIAL NO PODER JUDICIÁRIO

A Inteligência Artificial (IA) é uma ciência interdisciplinar cujo propósito é compreender

<sup>1</sup> O recurso computacional utilizado é a linguagem de programação R por meio do *software RStudio*. Para conhecer os pacotes utilizados e avaliar o código que gerou os resultados aqui apresentados: [https://github.com/Pedrouivo/PCAEClusters\\_dados\\_de\\_criminalidade\\_RJ.git](https://github.com/Pedrouivo/PCAEClusters_dados_de_criminalidade_RJ.git)

as regras subjacentes ao intelecto humano e reproduzi-las em agentes racionais capazes não apenas de agir autonomamente, mas também de reagir ao ambiente, aprender, persistir, adaptar e mudar sempre com o intuito de atingir o melhor resultado (RUSSELL; NORVIG, 2010). Essa visão do “agente racional”, que evoca a ideia de automatismo e independência está mais relacionada ao que se convencionou chamar de “Inteligência artificial forte”; amiúde, é essa a IA que se vê em ficção científica e que acaba balizando o que se entende sobre o assunto.

Entretanto, IA pode ser algo mais acessível e, até mesmo, fazer parte do dia a dia das organizações cujo escopo de atuação não é a prestação de serviços de tecnologia da informação. Sob uma perspectiva despojada da ambição de criar agentes racionais autônomos e mais sintonizada com o que se pretende aqui, Dora Kaufman entende que:

A inteligência artificial hoje é fundamentalmente modelos estatísticos que, baseados em dados, calculam a probabilidade de eventos ocorrerem. Esse pequeno avanço tem sido responsável por transformações na economia, nas relações pessoais, na sociedade em geral, mas estamos a léguas de distância da chamada general AI (ou strong AI ou full AI), que, supostamente, seria uma inteligência artificial dotada de capacidades de nível humano. (KAUFMAN, 2022, p. 12)

O aumento do poder computacional das organizações combinado com a popularização

de técnicas de mineração de dados permitiu que a inteligência artificial começasse a ser utilizada nos mais diversos campos de atuação profissional: desde o comércio varejista, passando por aplicativos de cálculo de rotas para itinerários automotivos, até o ensino de idiomas.

No Poder Judiciário, particularmente, há algumas iniciativas que chamam a atenção. A principal delas é o sistema Sinapses, uma “Fábrica de Modelos de IA” (CNJ, 2019, p.21) cujo propósito é utilizar a inteligência artificial para agilizar e aprimorar o processo judicial, tornando-o mais eficiente e efetivo. Mantido pelo CNJ e regulamentado pela Resolução CNJ n. 332/2020, o Sinapses, análogo a uma plataforma de desenvolvimento de funcionalidades de arquitetura de computação em nuvem, possibilita que modelos de IA, como redes neurais e algoritmos de processamento de linguagem natural, sejam utilizados por usuários do Processo Judicial Eletrônico (PJe). Também digno de nota, o Projeto Solaris, conduzido no âmbito do Tribunal Regional do Trabalho da 9ª Região, implementa o uso de robôs para realizarem tarefas repetitivas as quais prescindem da intervenção humana. Nessa esteira, há ainda a Elis, desenvolvida pelo

---

Tribunal de Justiça do Estado de Pernambuco, igualmente voltada para automação e classificação e o Victor, do Supremo Tribunal Federal, com forte componente classificatório também (CNJ, 2019).

Tanto na iniciativa privada quanto na administração pública, a IA desponta como um elemento disruptivo na busca por soluções para problemas de negócio ou por aprimoramento dos seus serviços. No entanto, existe algo que perpassa todas essas atividades: o trabalho profissional com dados. Embora o nome Business Intelligence ou inteligência de negócios tenha ficado popular para designar essa área, atualmente o termo mais usado para o trabalho profissional com dados é ciência de dados ou, em inglês, Data Science. Refere-se a uma grande variedade de técnicas e algoritmos computacionais voltados para análises descritivas, preditivas ou prescritivas (PROVOST; FAWCETT, 2016).

Algumas técnicas ajudam a entender melhor os dados a partir da geração de relatórios ad hoc, interativos e dinâmicos; outras, apoiadas fortemente em matemática e ciência da computação, ajudam a classificar os dados, calcular probabilidades e realizar previsões. Ainda há aquelas que também realizam

previsões a partir de dados, mas se baseiam em técnicas de regressão oriundas da Estatística (SHARDA; DELEN; TURBAN, 2019). De qualquer modo, todas essas técnicas, até mesmo aquelas voltadas para processamento de linguagem natural, dependem de um intenso, específico e sistemático trabalho com dados.

## 2.2 O PROBLEMA: DADOS NA GESTÃO DA SEGURANÇA INSTITUCIONAL

O gestor da Coordenadoria de Polícia Judicial do TRT-RJ possui uma grande responsabilidade: supervisionar as atividades de segurança nas unidades espalhadas por todo o Estado do Rio de Janeiro. Isso envolve não apenas planejar a alocação de recursos, humanos e materiais, mas também compreender o contexto de cada um desses locais no que diz respeito à segurança. Nesse sentido, o coordenador precisa ter uma visão ampla de diferentes aspectos que podem impactar positiva ou negativamente a segurança do patrimônio do Tribunal e também das pessoas que frequentam suas unidades em busca da prestação jurisdicional. Um desses aspectos, como preconiza a própria Resolução CNJ n. 383/2021, deve ser o cenário criminal no qual os prédios da Justiça estão inseridos.

É comum que tenha de responder sobre decisões tomadas nas ocasiões de deslocamentos de autoridades e de eventos aos quais elas compareçam em toda área do TRT-RJ. O gestor sabe da importância de tomar decisões seguras, baseadas em fatos. Para isso, é desejável conhecer os dados da criminalidade das regiões onde se situam os prédios que abrigam varas do trabalho por todo o estado. Se ele recorrer ao portal do Instituto de Segurança Pública do Rio de Janeiro (ISP-RJ), instância responsável pela análise e publicação de dados relativos à Segurança Pública do Rio de Janeiro, encontrará uma interface de alto padrão, na qual, optando pela “visualização de dados”, será levado a um painel dinâmico e interativo. Esse painel (também chamado pelo nome em inglês *dashboard*) oferece, por meio de menus suspensos e filtros, diversas estatísticas relativas à Segurança Pública nos municípios do estado .

Trata-se de um recurso que possui grande valor não apenas pelas informações que fornece, mas também pela usabilidade intuitiva e agradável. Contudo, o relatório é geral e não permite que o gestor da Polícia Judicial do TRT-RJ veja apenas um grupo de crimes por vez. É necessário ao gestor algo mais adaptado à realidade de sua atividade específica, que está

relacionada, principalmente, ao planejamento e supervisão das ações de segurança com vistas à prevenção de crimes nas áreas de atuação do TRT-RJ.

A base de dados original, na qual se baseiam as análises que o usuário pode acessar na página ‘visualização de dados’, também não é o que o gestor da segurança institucional precisa. Essa base é disponibilizada no formato *comma separated values (csv)* com título Estatísticas de segurança: série histórica mensal por município desde 2014 com 10120 linhas e 60 colunas . A tabela, extremamente difícil de ler, traz dados sobre a criminalidade nos 92 municípios do Estado do RJ, mês a mês, desde 2014 até julho de 2023. Além dessas colunas contendo ano e mês, e das colunas com nomes e códigos dos municípios, há ainda dezenas de colunas (variáveis) descrevendo diferentes dimensões do fenômeno criminal.

O ISP-RJ, inclusive, disponibiliza quadro com metadados que ajudam a entender a base original . Percebe-se que a base de dados contém bastante informação, daí sua relevância, porém é igualmente notável sua alta complexidade, o que torna quase impossível sua utilização de forma eficaz sem o auxílio de ferramentas adequadas. Em face desse problema, apresenta-

se uma solução que envolve ciência de dados e IA. O objetivo da análise é justamente conseguir retirar informações úteis dessa extensa tabela.

### 2.3 MATERIAL E MÉTODOS

O projeto apresentado a seguir pode ser visto como uma aplicação de inteligência artificial, na medida em que, em consonância com a definição de Dora Kaufman (2022), coleta dados de uma área de interesse e os submete a procedimentos estatísticos para obter insights e conhecimentos que, na sua forma bruta, não estavam disponíveis.

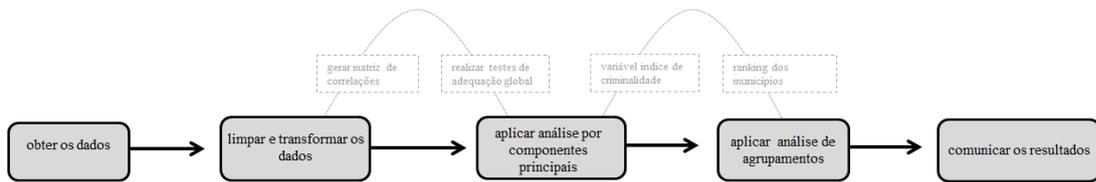
O objetivo último do estudo é oferecer ao gestor da área da segurança institucional uma análise originalmente feita no seio da instituição que pode auxiliar no processo decisório. Para isso, foram utilizados dois algoritmos de aprendizado de máquina não supervisionado: primeiramente, aplicou-se uma análise por componentes principais que gerou dois fatores que representam grande parte da variância das variáveis originais da base de dados primária. É um trabalho que, reduzindo a complexidade estrutural da base original, com perda de variância mínima, possibilitou, a partir dos dois fatores gerados, a criação de uma variável que os sintetiza, a que se chamou 'índice de criminalidade'.

Em seguida, utilizou-se um algoritmo de agrupamentos para formar grupos homogêneos internamente, porém heterogêneos entre si. Com esse resultado, o gestor da Polícia Judicial do TRT-RJ terá à sua disposição uma análise de dados personalizada para a atividade de segurança que desenvolve nesse Tribunal. O algoritmo utilizado classifica-se como aprendizado de máquina (ou modelagem) não supervisionado, justamente porque

[...] a modelagem supervisionada envolve a descoberta de padrões para prever o valor de uma variável alvo especificada, com base nos dados nos quais conhecemos os valores da variável alvo. A modelagem não supervisionada não se concentra em uma variável alvo. Em vez disso, busca outros tipos de regularidade em conjunto de dados. Não há modelo de treinamento e teste. (PROVOST; FAWCETT, 2016, p.164)

Além disso, o output ajuda a analisar as informações presentes na amostra de forma exploratória: não possuem variáveis explicativas e variável alvo, tampouco servem para inferências sobre outros grupos de dados não presentes na amostra, como é o caso de algoritmos de aprendizado de máquina supervisionado. A inserção de novos dados implicaria realizar novamente a análise. O projeto seguiu o fluxo baseado no que geralmente é recomendado para projetos em ciência de dados (WICKHAM; GROLEMUND, 2016):

**Figura 1 - Fluxo do projeto de dados.**

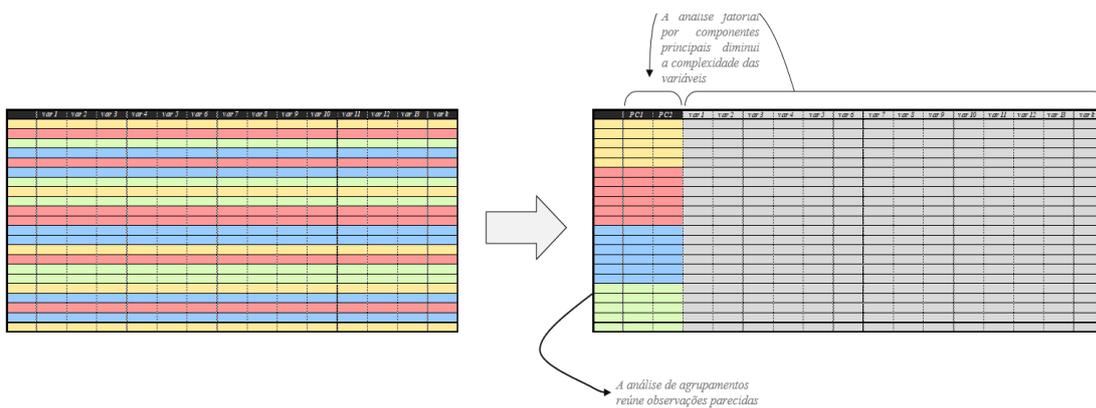


Fonte: elaboração do autor.

Portanto, o ponto inicial é a base de dados do ISP-RJ em formato csv, de alta complexidade e com informações dispersas e o resultado é uma base de dados própria que sintetiza a base original e que apresenta variáveis

resultantes de análises feitas sob medida para a atividade de segurança institucional do Poder Judiciário, além de gráficos que permitem obter informações com mais rapidez.

**Figura 2: Da base de dados original até o resultado da análise**



Fonte: elaboração do autor.

## 2.4 LIMPEZA E TRANSFORMAÇÕES DA BASE DE DADOS

A segunda fase, normalmente a mais longa em projetos de ciência de dados, teve como objetivo chegar a uma base de dados apropriada para a análise pretendida: uma linha para cada um dos 92 municípios e variáveis correspondentes aos tipos penais particularizados. Para chegar a esse resultado foi necessário excluir relevante quantidade de colunas, que traziam dados referentes ao efeito do ato delituoso, como encontro de cadáver ou de ossada; da mesma forma, foram excluídas as variáveis com dados sobre a atividade policial, como cumprimentos de mandado e autos de prisão em flagrante. Foi necessário, ainda, que se retirassem da base original variáveis que

consistiam em cálculos, a partir das variáveis de interesse. É o caso das variáveis “total\_roubo”, “total\_furto” ou “cvli” (cf. Quadro 1).

Até essa etapa, verifica-se que já é possível atingir um resultado interessante: uma base de dados bem mais sintética e apropriada, que reúne dados quantitativos sobre diversos tipos penais ao longo dos doze meses dos anos de 2018 a 2022 (tabela com 5520 observações e 38 variáveis). Isso já seria suficiente para uma série de análises pertinentes, razão pela qual a base de dados resultante desse processo de limpeza e transformação de dados ficará salva nos arquivos da Coordenadoria de Polícia Judicial para posteriores estudos, consultas e análises:

**Quadro 1 - Variáveis na base de dados original do ISP-RJ**

Variável	Tipo
Município	qualitativa
Ano	qualitativa
Mês	qualitativa
Região	qualitativa
hom_doloso	quantitativa discreta
lesao_corp_morte	quantitativa discreta
Latrocínio	quantitativa discreta
hom_por_interv_policial	quantitativa discreta
tentat_hom	quantitativa discreta
lesao_corp_dolosa	quantitativa discreta
estupro	quantitativa discreta
roubo_transeunte	quantitativa discreta
roubo_celular	quantitativa discreta
roubo_em_coletivo	quantitativa discreta
roubo_ rua	quantitativa discreta
roubo_veiculo	quantitativa discreta
roubo_carga	quantitativa discreta
roubo_comercio	quantitativa discreta
roubo_residencia	quantitativa discreta
roubo_banco	quantitativa discreta
roubo_cx_eletronico	quantitativa discreta
roubo_conducao_saque	quantitativa discreta
roubo_apos_saque	quantitativa discreta
roubo_bicicleta	quantitativa discreta
outros_roubos	quantitativa discreta
furto_veiculos	quantitativa discreta
furto_transeunte	quantitativa discreta
furto_coletivo	quantitativa discreta
furto_celular	quantitativa discreta
furto_bicicleta	quantitativa discreta
outros_furtos	quantitativa discreta
Sequestro	quantitativa discreta
Extorsão	quantitativa discreta
sequestro_relampago	quantitativa discreta
Estelionato	quantitativa discreta
trafico_drogas	quantitativa discreta
Ameaça	quantitativa discreta
policiais_mortos	quantitativa discreta

Fonte: elaboração do autor.

## 2.5 ANÁLISE FATORIAL POR COMPONENTES PRINCIPAIS

Na seção anterior, apresentou-se a situação de uma base de dados criminais com sessenta variáveis e mais de dez mil linhas. Uma base assim, embora rica em informações, é extremamente complexa para uma análise feita unicamente pela cognição humana. Após alguns procedimentos de limpeza e transformação, conseguiu-se uma base mais resumida. Mas que ainda possui certa complexidade: mais de cinco mil linhas e trinta e oito observações.

Certamente, o gestor conseguirá realizar consultas pontuais para conferir a quantidade de ocorrência de crimes específicos cometidos em uma determinada localidade nos meses e anos que ali aparecem. O objetivo, entretanto, é que o trabalho realizado nesses dados forneça informação relevante, com uma visão abrangente das localidades de atuação do TRT-RJ.

Verifica-se, portanto, a necessidade de efetuar mais algumas transformações na base: agruparam-se as observações por municípios somando as linhas, assim um valor único na tabela corresponde ao total de crimes naquela localidade entre 2018 e 2022; finalmente, o município do Rio de Janeiro, capital, foi retirado da base de dados em razão do comportamento peculiar dos dados nessa localidade específica. Sendo uma realidade sensivelmente diferente, com valores muito mais altos que os outros, pelo menos no que diz respeito ao fenômeno criminal, convém estudar os dados da capital do estado separadamente em outra oportunidade. Na tabela 1, apresenta-se uma parte da base de dados que será analisada na próxima seção (a base foi reduzida para fins de apresentação):

**Tabela 1 – Base de dados para a análise: 91 linhas e 28 colunas**

municípios	hom_doloso	lesao_corp_morte	latrocinio	lesao_corp_dolosa
Angra dos Reis	367	0	6	3073
Aperibé	4	0	0	268
Araruama	229	1	3	2438
Areal	3	0	0	201
Armação dos Búzios	66	1	1	1140
Arraial do Cabo	57	0	0	755
Barra do Piraí	108	1	0	1888
Barra Mansa	227	2	0	2365
Belford Roxo	856	7	22	7243
Bom Jardim	19	0	2	409
Bom Jesus do Itabapoana	22	1	1	834
Cabo Frio	448	2	10	3592
Cachoeiras de Macacu	88	1	2	923
Cambuci	8	0	1	297
Campos dos Goytacazes	760	3	9	5162
Cantagalo	4	1	0	415
Carapebus	17	0	1	174
Cardoso Moreira	9	0	2	236
Carmo	1	0	0	547
Casimiro de Abreu	96	0	0	720

Fonte: elaboração do autor

Diante do cenário exposto, torna-se extremamente conveniente recorrer a um algoritmo de análise fatorial pelo método de componentes principais. Trata-se de uma técnica multivariada exploratória que, a partir da correlação entre as variáveis, extrai fatores que representam o comportamento conjunto das variáveis originais (FÁVERO; BELFIORE, 2021). Aplica-se esse algoritmo com dois propósitos específicos: 1) reduzir a dimensionalidade da base de dados originais e

2) criação de um ranking entre os municípios do Rio de Janeiro por meio da criação de um indicador de criminalidade, calculado a partir fatores extraídos pelo algoritmo. Ao fim, tem-se uma análise de caráter diagnóstico da situação criminal na área de interesse para o gestor da Polícia Judicial, que possuirá informações inéditas, produzidas internamente, nas quais poderá basear suas decisões.

O sucesso de uma análise fatorial por componentes principais depende muito de

haver correlações altas entre os fatores e, em um segundo momento, de testes que visam avaliar a adequação global da análise para a base de dados em questão. Inicialmente, houve um impasse quanto à necessidade de se proceder em novas manipulações da base de dados com o intuito de diminuí-la ainda mais no seu número de colunas. Com efeito, Matos e Rodrigues (2019) apresentam alegações em favor de um tamanho ideal da amostra para que a análise fatorial seja relevante.

Nessa perspectiva, deveria haver uma proporção de aproximadamente dez linhas para cada coluna da base de dados ou então um número mínimo de linhas, próximo a trezentas

observações. Fizemos alguns testes seguindo essas recomendações e os resultados não foram mais satisfatórios do que aqueles obtidos com a base de dados 91x28, tal qual apresentada na tabela 1. Fávero e Belfiore (2021), em obra referência de análise de dados, não fazem qualquer menção ao tamanho da amostra, uma vez que existem os testes de adequação global cujo resultado indica se o analista deve ou não prosseguir com a análise fatorial. Para não restarem dúvidas, cotejamos os resultados desses testes para as duas bases de dados, uma reduzida e uma em seu tamanho original:

**Tabela 2 – Comparação de testes de adequação global da análise entre bases de dados**

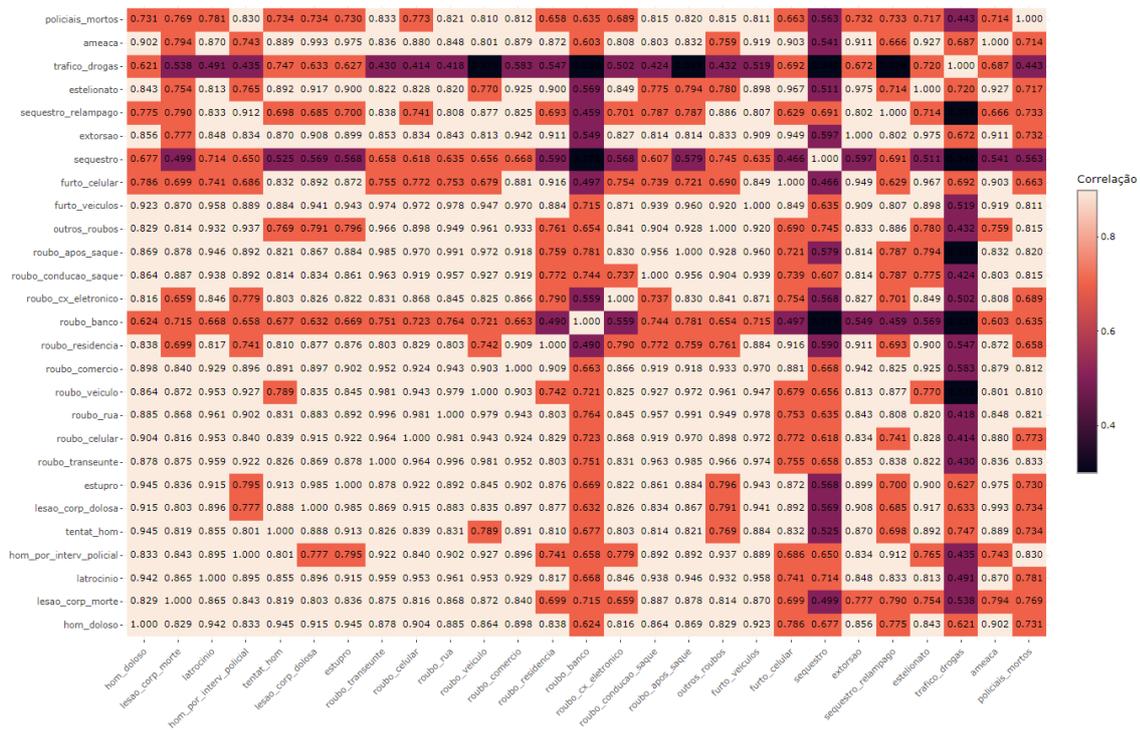
	Teste de esfericidade de Bartlett	Média da estatística Kaiser-Meyer-Olkin (KMO)
base original (91x28)	0	0.8918519
base com variáveis agrupadas (91x7)	0	0.8066667

Fonte: elaboração do autor

O teste de esfericidade de Bartlett teve bom resultado para ambas as bases; no entanto, a base de dados maior mostrou resultados melhores no segundo teste. Prosseguiu-se então, com a base original. O passo seguinte foi verificar como são as correlações entre as variáveis, utilizando o cálculo da correlação

de Pearson. O resultado pode ser avaliado no visual gerado pelo software:

Figura 3: Matriz de correlação entre as 27 variáveis da base de dados analisada



Fonte: elaboração do autor.

Há dois aspectos em particular da matriz na figura 2 que merecem destaque: primeiramente, é a alta correlação entre as variáveis. De acordo com a escala ao lado, quanto mais clara a cor da célula onde se encontra o coeficiente de correlação, mais próximo de um está esse coeficiente. Em uma rápida análise visual, conclui-se que entre as vinte e sete colunas contendo dados quantitativos relativos a diferentes tipos penais a correlação é, de modo geral, alta. Convém apontar, que pelo menos dois tipos penais destoam da tendência geral dos dados de apresentar correlação positiva

acima de 0,7: roubo a banco e tráfico de drogas. Esses coeficientes mais baixos poderiam prejudicar a qualidade da análise, o que não ocorreu como será detalhado mais adiante. Se fosse o caso, seria conveniente retirá-los da base de dados.

A tabela 2 mostrou que a base de dados é adequada para a análise proposta e a correlação predominantemente alta corrobora esse resultado. O próximo passo é iniciar o processamento da análise para que os fatores sejam calculados. Utilizou-se a função principal do pacote psych no software RStudio. Essa

função oferece diversos resultados que são relevantes para o raciocínio desenvolvido neste estudo. O objetivo desse algoritmo é extrair um fator numérico para cada variável da base de dados de entrada.

Cada um desses fatores é extraído a partir de um autovalor ou eigenvalue que, por essa razão, pode ser tomado como uma medida do quanto um determinado fator extraído é representativo

da variância global da amostra (MATOS; RODRIGUES, 2019). Portanto, quanto mais alto um eigenvalue associado a um fator em particular, mais representativo do conjunto de variáveis esse fator vem a ser. O resultado da análise por componentes principais da base de dados criminais do estado do Rio de Janeiro apresenta o seguinte resultado:

**Tabela 3 – Autovalores e proporção da variância**

	Autovalores	Prop. da Variância	Prop. da Variância Acumulada
PC1	21,872	0,810	0,810
PC2	1,636	0,061	0,871
PC3	0,942	0,035	0,906
PC4	0,606	0,022	0,928
PC5	0,459	0,017	0,945
PC6	0,328	0,012	0,957
PC7	0,296	0,011	0,968
PC8	0,242	0,009	0,977
PC9	0,161	0,006	0,983
PC10	0,132	0,005	0,988
PC11	0,087	0,003	0,991
PC12	0,065	0,002	0,993
PC13	0,053	0,002	0,995
PC14	0,036	0,001	0,996
PC15	0,019	0,001	0,997
PC16	0,016	0,001	0,998
PC17	0,013	0,000	0,998
PC18	0,010	0,000	0,998
PC19	0,007	0,000	0,998
PC20	0,006	0,000	0,998
PC21	0,005	0,000	0,998
PC22	0,003	0,000	0,998
PC23	0,001	0,000	0,998
PC24	0,001	0,000	0,998
PC25	0,001	0,000	0,998
PC26	0,001	0,000	0,998
PC27	0,000	0,000	0,998

Fonte: elaboração do autor.

---

Para cada variável, há um fator calculado a partir de um autovalor que mostra a importância desse fator. Para entender melhor a relevância de um autovalor, costuma-se calcular a proporção da variância global da amostra que ele captura. No caso da análise aqui apresentada, observa-se que o primeiro fator ou componente principal (PC1) se destaca dos demais concentrando mais de 80% da variância.

Considerando o propósito de encontrar fatores que representem as variáveis para reduzir o número de variáveis, não faria sentido manter os vinte e sete fatores gerados. É necessário escolher alguns dentre os vinte e sete fatores extraídos. Geralmente o critério utilizado para fazer essa escolha é o critério de Kaiser ou critério da raiz latente (FÁVERO; BELFIORE, 2021), segundo o qual devem ser

mantidos na análise apenas os fatores gerados a partir de autovalores maiores do que um. Portanto, para as próximas etapas, apenas os dois primeiros fatores, já destacados na tabela 3, estarão em foco, uma vez que são os únicos que atendem ao critério de Kaiser. Juntos, representam 87% da variância da amostra.

A tabela 4 apresenta o quanto cada variável contribui para a formação de desses fatores. Dessa forma, será possível identificar qual variável está bem representada e qual não está. Essa informação é obtida analisando outro output do algoritmo, o escore fatorial, que efetivamente relaciona as variáveis originais ao fator extraído, uma vez que funciona como um parâmetro que irá multiplicar cada valor original padronizado para formar os fatores:

**Tabela 4 – Escores fatoriais**

	PC1	PC2
hom_doloso	0,0431	0,0505
lesao_corp_morte	0,0405	-0,0610
latrocinio	0,0442	-0,0800
hom_por_interv_policial	0,0418	-0,1414
tenta_hom	0,0418	0,1430
lesao_corp_dolosa	0,0431	0,1314
estupro	0,0434	0,1102
roubo_transeunte	0,0444	-0,1305
roubo_celular	0,0437	-0,0719
roubo_rua	0,0443	-0,1239
roubo_veiculo	0,0432	-0,1773
roubo_comercio	0,0448	0,0173
roubo_residencia	0,0403	0,1525
roubo_banco	0,0324	-0,1346
roubo_cx_eletronico	0,0400	0,0329
roubo_conducao_saque	0,0428	-0,1256
roubo_apos_saque	0,0435	-0,1425
outros_roubos	0,0426	-0,1582
furto_veiculo	0,0453	-0,0190
furto_celular	0,0393	0,2591
sequestro	0,0304	-0,1107
extorsao	0,0426	0,1500
sequestro_relampago	0,0383	-0,1369
estelionato	0,0418	0,2088
trafico_drogas	0,0268	0,3873
ameaca	0,0423	0,1766
policiais_mortos	0,0381	-0,1000

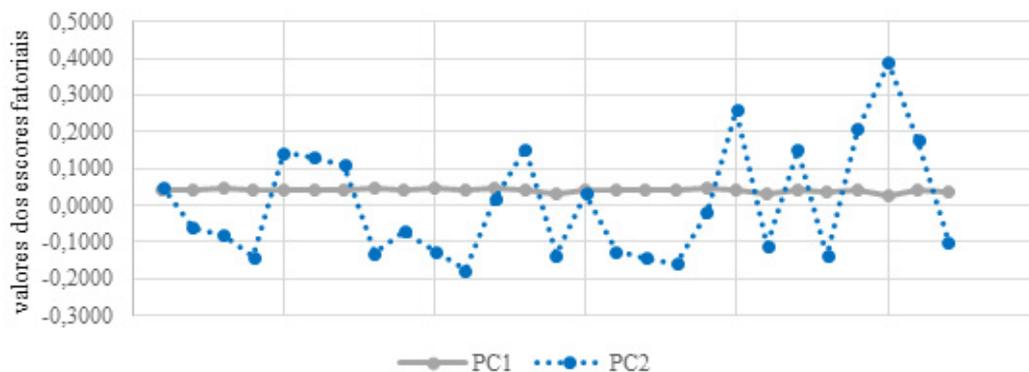
Fonte: elaboração do autor

Nesse ponto, é importante saber que o fator 1, que aparece nos outputs como PC1, em referência ao inglês principal component, é calculado usando os escores fatoriais da primeira coluna numérica da tabela 4 e o fator 2 (PC2), usando os escores fatoriais que aparecem associados à coluna PC2 da mesma tabela. Como já havia sido sinalizado anteriormente

na proporção da variância de cada eigenvalue, o fator 1 tem representatividade consideravelmente maior do que o fator 2. Embora em algumas variáveis o escore do segundo fator extraído (PC2) seja maior do que a do PC1, este apresenta maior quantidade de escores maiores e menor variação de valores, o que irá influenciar na representatividade do

fator gerado a partir desse escore. O gráfico apresentado na figura 3 deixa bem clara essa diferença:

**Figura 3:** Gráfico de dispersão dos escores fatoriais



Fonte: elaboração do autor.

Quanto aos fatores, eles não dizem muita coisa sozinhos, é fundamental interpretá-los segundo os outros resultados que ora são comentados. Contudo, eles serão de grande utilidade para o cálculo do índice apresentado mais adiante, figurando na nova base de dados como “fator\_pc1” e “fator\_pc2”. No momento,

entretanto, talvez seja mais interessante para a análise verificar as cargas fatoriais, que nada mais são do que a correlação de Pearson entre os fatores gerados e os dados originais. Assim, pode-se avaliar com mais clareza o ajuste dos fatores extraídos com as variáveis originais:

**Tabela 5 – Cargas fatoriais**

Variáveis	PC1	PC2
hom_doloso	0,942	0,083
lesao_corp_morte	0,886	-0,100
latrocinio	0,966	-0,131
hom_por_interv_policial	0,914	-0,231
tentat_hom	0,914	0,234
lesao_corp_dolosa	0,942	0,215
estupro	0,950	0,180
roubo_transeunte	0,970	-0,213
roubo_celular	0,957	-0,118
roubo_ rua	0,969	-0,203
roubo_veiculo	0,945	-0,290
roubo_comercio	0,979	0,028
roubo_residencia	0,881	0,249
roubo_banco	0,709	-0,220
roubo_cx_eletronico	0,875	0,054
roubo_conducao_saque	0,937	-0,206
roubo_apos_saque	0,952	-0,233
outros_roubos	0,933	-0,259
furto_veiculos	0,990	-0,031
furto_celular	0,859	0,424
sequestro	0,664	-0,181
extorsao	0,933	0,245
sequestro_relampago	0,838	-0,224
estelionato	0,915	0,342
trafico_drogas	0,586	0,634
ameaca	0,924	0,289
policiais_mortos	0,833	-0,164

Fonte: elaboração do autor.

O resultado apresenta a correlação de Pearson entre as variáveis originais e os dois fatores extraídos. A alta proporção de variância capturada pelo fator 1, identificada quando se comentou os autovalores, é reiterada agora com os valores da segunda coluna da tabela 5: praticamente todos acima de 0,8. O gráfico de dispersão com dois eixos ajuda a perceber

melhor a força da correlação do fator 1 (PC1) com algumas variáveis da base de dados original:

Figura 4 - Gráfico de dispersão das cargas fatoriais



Fonte: elaboração do autor.

Como já havia sido apontado, três variáveis estão fracamente representadas pelos dois fatores. São elas “trafico\_drogas”, “sequestro” e “roubo\_banco”. Já no círculo mais ao alto, verifica-se um resultado interessante: um grupo de crimes que são bem representados pelos dois fatores extraídos. Entre eles estão alguns tipos de crime que têm maior impacto no planejamento da segurança institucional do que outros, pois apresentam maior probabilidade de ocorrer em um prédio da Justiça, como o furto de celular, ou a um Magistrado, como estelionato ou ameaça. Dessa forma, será possível visualizar que no índice criminal de cada município da análise, esses crimes têm uma representatividade considerável.

Já no círculo mais abaixo, estão crimes que são bem representados pelo PC1, mas cujo fator do PC2 foi irrelevante. Ainda assim, é relevante saber que o crime de roubo a veículo está bem representado na análise, tipo de delito que também preocupa a segurança institucional – que em suas atividades conduz veículos motorizados para deslocamentos com equipamentos sensíveis e autoridades.

Diante deste quadro, questiona-se o quanto das variáveis originais estão representadas nos fatores calculados ou o quanto de informação foi perdida com a redução imposta à base de dados original. Essa informação é dada por uma medida chamada comunalidade. Ela seria 100% para todas as variáveis, caso todos os fatores

gerados pelo algoritmo tivessem sido mantidos na análise (cf. tabela 3). Mas como o objetivo foi reduzir a dimensionalidade, aplicou-se o critério de Kaiser para selecionar apenas os fatores derivados de autovalores maiores do que um. Assim, evidentemente, perde-se um pouco de informação e isso reduz a comunalidade. No

entanto, como pode-se constatar com a tabela 6, ainda que a maior parte dos fatores extraídos tenham sido deixados de lado, os dois fatores mantidos capturam um alto nível de variância do grupo original de variáveis:

**Tabela 6 – Comunalidades**

variáveis	comunalidades
hom_doloso	0,895
lesao_corp_morte	0,795
latrocinio	0,951
hom_por_interv_policial	0,889
tentat_hom	0,891
lesao_corp_dolosa	0,934
estupro	0,935
roubo_transeunte	0,987
roubo_celular	0,929
roubo_ rua	0,980
roubo_veiculo	0,977
roubo_comercio	0,960
roubo_residencia	0,839
roubo_banco	0,551
roubo_cx_eletronico	0,768
roubo_conducao_saque	0,920
roubo_apos_saque	0,962
outros_roubos	0,937
furto_veiculos	0,981
furto_celular	0,917
sequestro	0,474
extorsao	0,930
sequestro_relampago	0,752
estelionato	0,953
trafico_drogas	0,745
ameaca	0,938
policiais_mortos	0,721

Fonte: elaboração do autor.

Os dois fatores gerados, portanto, são altamente representativos da variação da amostra. Viu-se que especialmente o primeiro fator captura grande parte da variância global das variáveis originais. As comunalidades confirmam que a análise gerou resultados satisfatórios, uma vez que se chegou a duas variáveis (PC1 e PC2) que representam a variância das outras vinte e sete variáveis quantitativas da base de dados original. Ainda é possível transformar essas duas variáveis em apenas uma, que servirá como um grande resumo de todas as outras, sendo, assim, uma espécie de índice criminal. Para isso, realizamos um cálculo simples, apresentado na fórmula 1, que soma os dois fatores ponderados pela proporção da variância que representam:

**Fórmula 1** – Índice criminal.

$$\text{Índice criminal} = (\text{PC1} \cdot \text{prop. da variância PC1}) + (\text{PC2} \cdot \text{prop. da variância PC2})$$

Fonte: elaboração do autor.

Onde PC1 e PC2 são respectivamente, o primeiro e o segundo fatores extraídos pelo algoritmo (cf. Anexo A), enquanto prop. da variância PC1 e prop. da variância PC2 correspondem, respectivamente, à proporção

da variância da amostra capturada pelos PC1 e PC2, valores apresentados na tabela 3.

Após inserir esse índice criminal na tabela, foram realizados procedimentos estatísticos simples para oferecer ao gestor um material apropriado ao seu uso profissional. Criou-se um ranking com base no índice criminal calculado e uma classificação de risco levando em consideração a posição desse índice na amostra. Se está entre os 25% mais altos, classificou-se como risco alto; se está entre os 25% mais baixos, o risco é baixo; entre esses dois limites, correspondentes ao primeiro e ao terceiro quartis, classificou-se como risco médio. Utilizou-se também um recurso de cores para evidenciar a transição das localidades com maior índice de criminalidade para aquelas com menor índice:

**Tabela 7 – municípios do Rio de Janeiro em rol com classificação de risco.**

ranking	municípios	risco	indice_criminal
1	São Gonçalo	alto	3,861
2	Duque de Caxias	alto	3,683
3	Nova Iguaçu	alto	3,111
4	Niterói	alto	2,201
5	São João de Meriti	alto	1,659
6	Belford Roxo	alto	1,414
7	Campos dos Goytacazes	alto	1,137
8	Itaboraí	alto	0,556
9	Cabo Frio	alto	0,548
10	Macaé	alto	0,528
11	Petrópolis	alto	0,509
12	Volta Redonda	alto	0,502
13	Magé	alto	0,419
14	Queimados	alto	0,295
15	Rio das Ostras	alto	0,257
16	Angra dos Reis	alto	0,256
17	Mesquita	alto	0,252
18	Maricá	alto	0,241
19	Nilópolis	alto	0,232
20	Itaguaí	alto	0,180
21	Nova Friburgo	alto	0,158
22	Teresópolis	alto	0,130
23	Araruama	alto	0,115
24	Barra Mansa	médio	0,075
25	Resende	médio	0,049
26	Três Rios	médio	-0,070
27	Seropédica	médio	-0,070
28	Japeri	médio	-0,085
29	Saquarema	médio	-0,105
30	Barra do Pirai	médio	-0,110
31	São Pedro da Aldeia	médio	-0,127
32	Itaperuna	médio	-0,163
33	Paraty	médio	-0,212
34	Valença	médio	-0,212
35	Guapimirim	médio	-0,229
36	Armação dos Búzios	médio	-0,230
37	Mangaratiba	médio	-0,251
38	Arraial do Cabo	médio	-0,252
39	Cachoeiras de Macacu	médio	-0,274
40	Rio Bonito	médio	-0,278
41	Casimiro de Abreu	médio	-0,283
42	São João da Barra	médio	-0,284
43	Paracambi	médio	-0,290
44	Santo Antônio de Pádua	médio	-0,302
45	São Francisco de Itabapoana	médio	-0,306
46	Itatiaia	médio	-0,315
47	Bom Jesus do Itabapoana	médio	-0,324
48	Paraíba do Sul	médio	-0,327
49	Miracema	médio	-0,331

50	Vassouras	médio	-0,333
51	Iguaba Grande	médio	-0,335
52	Tanguá	médio	-0,350

ranking	municípios	risco	indice_criminal	(conclusão)
53	Piraí	médio	-0,350	
54	São Fidélis	médio	-0,355	
55	Pinheiral	médio	-0,367	
56	Bom Jardim	médio	-0,367	
57	Quissamã	médio	-0,370	
58	Cordeiro	médio	-0,371	
59	Miguel Pereira	médio	-0,377	
60	Itaocara	médio	-0,379	
61	Porciúncula	médio	-0,385	
62	Paty do Alferes	médio	-0,386	
63	Quatis	médio	-0,388	
64	Silva Jardim	médio	-0,388	
65	Porto Real	médio	-0,388	
66	Conceição de Macabu	médio	-0,395	
67	Cardoso Moreira	médio	-0,397	
68	Cantagalo	médio	-0,399	
69	Mendes	baixo	-0,409	
70	Sapucaia	baixo	-0,411	
71	São José do Vale do Rio Preto	baixo	-0,414	
72	Carmo	baixo	-0,415	
73	Aperibé	baixo	-0,416	
74	Sumidouro	baixo	-0,417	
75	Italva	baixo	-0,417	
76	Natividade	baixo	-0,417	
77	Cambuci	baixo	-0,420	
78	Rio Claro	baixo	-0,424	
79	Duas Barras	baixo	-0,429	
80	Santa Maria Madalena Engenheiro Paulo de Frontin	baixo	-0,429	
81	Frontin	baixo	-0,430	
82	Carapebus	baixo	-0,432	
83	Laje do Muriaé	baixo	-0,438	
84	Rio das Flores	baixo	-0,439	
85	Trajano de Moraes	baixo	-0,440	
86	Areal	baixo	-0,440	
87	Comendador Levy Gasparian	baixo	-0,440	
88	Macuco	baixo	-0,443	
89	Varre-Sai	baixo	-0,443	
90	São Sebastião do Alto	baixo	-0,447	
91	São José de Ubá	baixo	-0,450	

Fonte: elaboração do autor.

## 2.6 ANÁLISE DE AGRUPAMENTOS

Até aqui, a análise concentrou-se sobretudo em diminuir a dimensionalidade da base de dados original disponível na web por meio da extração de fatores representativos dos dados registrados nas variáveis. Como ilustra a figura 1, este artigo também se propõe a processar as observações da base de dados de forma a complementar a análise exploratória com dados da criminalidade na área de abrangência do TRT-RJ.

A análise de agrupamento ou análise de aglomerados (em inglês, clustering) é geralmente vista como um modelo de aprendizado não supervisionado (RUSSEL; NORVIG, 2010). É eficaz como método de análise exploratória de dados e, de fato, simula computacionalmente uma habilidade típica da inteligência humana: a percepção de semelhança e de diferença entre itens de um grupo. Considerando que as noventa e uma linhas da base de dados com que se trabalha aqui forme um grupo em si e que cada uma separadamente forme noventa e um grupos com um integrante, o objetivo da análise de agrupamento é, por meio de cálculos algébricos sequenciais, alocar cada observação em uma quantidade pequena de “agrupamentos

homogêneos internamente e heterogêneos entre si” (FÁVERO; BELFIORE, 2021, p. 309, grifo do autor). Portanto, espera-se, ao fim da análise, avaliar os municípios localizados na área de atuação do TRT-RJ em grupos parecidos internamente e diferentes entre si.

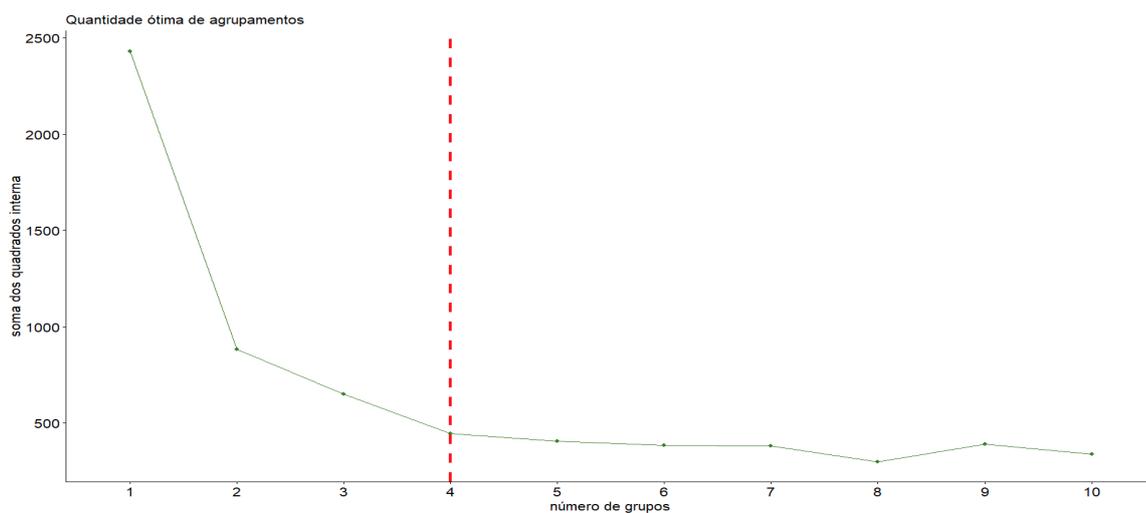
Há dois métodos de agrupamento que podem ser utilizados no presente estudo: o hierárquico e o não hierárquico. O primeiro baseia-se na distância entre as observações. Existem alguns tipos de cálculos para essa distância; entretanto, utiliza-se aqui o mais comum, que é o cálculo da distância euclidiana. O estágio inicial do algoritmo considera que cada observação da amostra é um agrupamento. A partir desse ponto, elas vão sendo combinadas de acordo com o método de encadeamento escolhido. O resultado pode ser avaliado em um gráfico chamado dendrograma em que a ideia de hierarquia fica nítida. O analista pode, então, a partir da observação do gráfico, escolher um ponto de corte a partir do qual a quantidade de aglomerados será definida.

Já o método não hierárquico, geralmente identificado com o algoritmo baseado em centroides k-means, o número de agrupamentos é definido previamente. Procede-se, na

seqüência, ao cálculo reiterado da distância entre os valores das observações e esses centroides, que nada mais são do que “médias aritméticas dos valores ao longo de cada dimensão para as instâncias do agrupamento” (PROVOST; FAWCETT, 2016, p.170). Na ausência de qualquer outra pista para o analista escolher o número de agrupamentos com que começar, costuma-se utilizar um método visual:

o método de elbow. Trata-se de um gráfico em que se pode avaliar a diminuição da soma dos quadrados interna à medida em que aumenta o número de agrupamentos. O ponto de inflexão da linha é onde a variação interna dos grupos foi minimizada sem aumentar excessivamente o número de grupos:

**Figura 5 - Método de elbow para a escolha da quantidade ótima de agrupamentos**



Fonte: elaboração do autor.

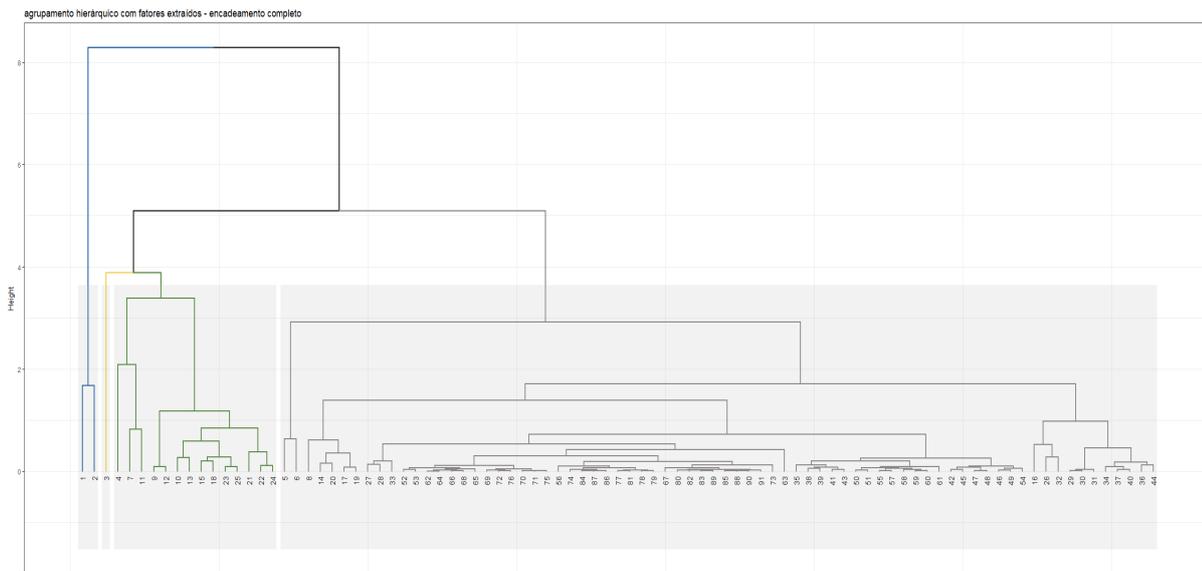
A figura 4 mostra que a partir de quatro agrupamentos não há ganho significativo de informação que justifique o aumento da quantidade de grupos. Adotou-se, portanto, quatro grupos para os dois algoritmos, porque

além do que mostra a figura 4 com o gráfico produzido no software, esse também é o número de unidades que a Coordenadoria de Polícia Judicial do TRT-RJ possui no interior do Rio de Janeiro.

Após diversos processamentos do código, chegou-se ao melhor resultado. A figura 5 a seguir mostra os agrupamentos da análise com o método hierárquico. Cada número na base

do gráfico da figura 5 corresponde à posição do município no ranking elaborado na seção anterior:

**Figura 6 - Dendrograma de agrupamentos pelo método hierárquico**



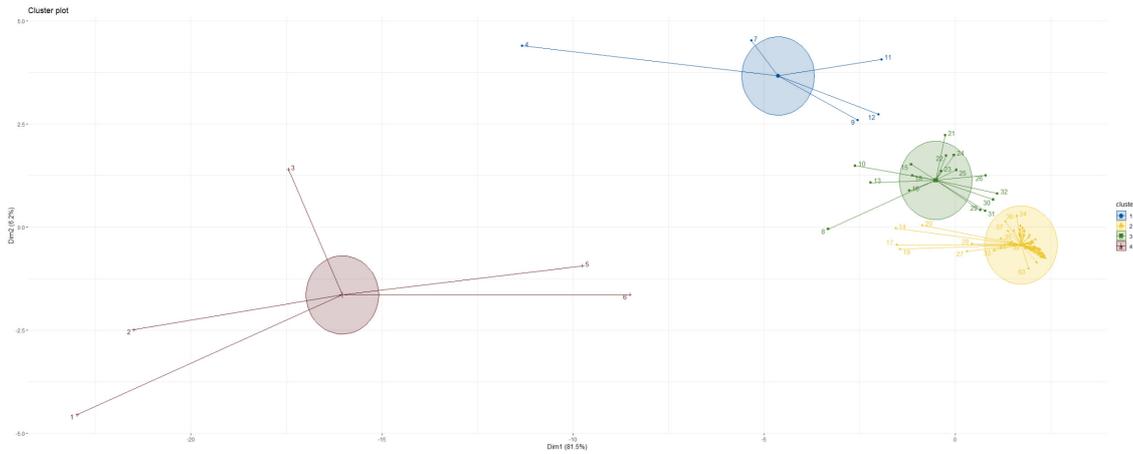
Fonte: elaboração do autor.

Há um grupo consideravelmente grande e outros três menores. Inclusive há um grupo formado por um único município, o terceiro no ranking. Esse não é um resultado ideal muito proveitoso. Afinal, será que esse município é tão diferente assim dos outros? Para responder essa pergunta, convém conferir os resultados gerados pelo método não hierárquico k-means.

a organização dos grupos em torno dos centroides:

Os resultados do método não hierárquico mostram um melhor equilíbrio entre as observações. Na figura 6, fica evidente

**Figura 7 - Agrupamentos pelo método k-means**

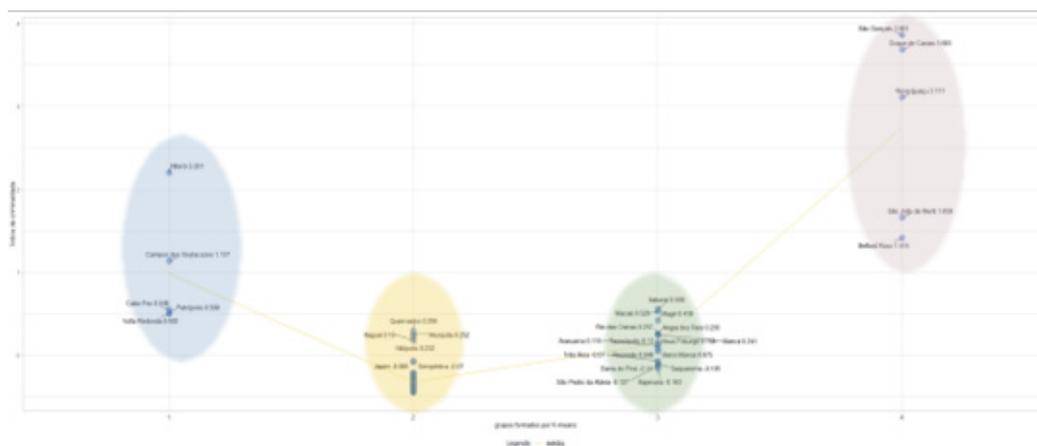


Fonte: elaboração do autor

De forma semelhante aos agrupamentos formados pelo método hierárquico vistos na figura 5, há dois grupos concentrando grande parte das observações e outros dois grupos com menor número de observações. No entanto, esses dois grupos com menor número de observações reúnem os municípios com maior média de índice de criminalidade, como mostra a linha amarela na figura 7. Esse resultado se mostra mais satisfatório que o anterior pois,

apesar de grande concentração de observações nos grupos dois e três, a distribuição entre os grupos ficou mais equilibrada. A tabela 7 ajuda a entender a diferença entre os grupos. Esse já é um resultado interessante porque permite que medidas de segurança mais vigorosas sejam implantadas para os municípios que estão nesses grupos, especialmente para o grupo 4.

**Figura 8 - Dispersão dos agrupamentos com média do índice de criminalidade**



Fonte: elaboração do autor.

Contudo, o resultado mais interessante está na tabela 8. Trata-se da estatística-F, calculada através da análise de variância. Ela é o resultado da divisão do quadrado médios entre os grupos pelo quadrado médio dentro dos grupos. Se o que se deseja com uma análise de agrupamentos é chegar a poucos grupos homogêneos internamente e heterogêneos entre eles, o resultado que indica essa condição é uma estatística-F alta. É justamente o que se verifica na tabela 8, especialmente quando se considera a variável criada na análise por componentes principais na seção anterior (fator\_pc1), que,

como apresentado, é a mais representativa das variáveis originais. Além disso, a estatística-F do método não hierárquico é notavelmente maior que no método hierárquico. Essa constatação vem a ser uma confirmação estatística do que foi percebido visualmente com o auxílio das figuras 5 e 6: os agrupamentos alcançados por meio do método não hierárquico k-means são aqueles que devem ser tomados como referência da divisão dos municípios do estado do Rio de Janeiro no que diz respeito ao índice de criminalidade:

**Tabela 8 – Estatística-F dos agrupamentos**

	Hierárquico		K-means	
	fator_pc1	fator_pc2	fator_pc1	fator_pc2
Estatística-F	101	96,73	141	96,73

Fonte: elaboração do autor.

### 3. CONSIDERAÇÕES FINAIS

Identificam-se muitas iniciativas dedicadas a conferir celeridade e efetividade ao processo judicial, assim como diversas ações envolvendo inteligência artificial para melhoria dos fluxos relacionados à atividade-fim dos Tribunais de Justiça. Em que pese sua inegável importância, se não houver um planejamento de segurança ao redor dos atores e do próprio patrimônio do Poder Judiciário, a prestação jurisdicional pode ser seriamente comprometida. Por isso, a segurança institucional tem sido vista como uma atividade primordial para a independência da Justiça (CNJ, 2020b, 2021a, 2021b). Nesse contexto, a prevenção é a melhor estratégia para lidar com os riscos, sejam eles internos ou externos, como é o caso da violência e criminalidade.

Este trabalho apresentou um caminho para que técnicas de aprendizado de máquinas próprias da inteligência artificial possam fazer parte da rotina dos gestores

das unidades de Polícia Judicial, onde estão lotados os servidores que operacionalizam o planejamento de segurança institucional dos tribunais. Primeiramente, obteve-se na web uma base de dados sobre a criminalidade com sessenta colunas e milhares de linhas. Esse tipo de estrutura não facilita o trabalho do gestor da Polícia Judicial do TRT-RJ. Após submetê-la a uma análise fatorial por componentes principais foi possível reduzir essa complexidade a dois fatores altamente representativos da variância original da amostra. Na sequência, esses dois fatores foram utilizados em um cálculo ponderado para a criação de um índice de criminalidade que permitiu ordenar de forma decrescente os municípios do estado do Rio de Janeiro.

Ao passo em que essa primeira análise teve como foco as muitas colunas da base de dados original, a segunda análise deteve-se nas noventa e uma linhas às quais a tabela original foi reduzida. A aplicação de um algoritmo de

---

aprendizado de máquina não supervisionado resultou em quatro grupos que reúnem os municípios da base original, com exceção da capital, de acordo com suas semelhanças e dissemelhanças no que diz respeito aos dados da criminalidade. Seja no método hierárquico, seja no método K-means, cujos resultados foram os mais satisfatórios, esse tipo de análise certamente será útil para, por exemplo, decidir sobre a alocação de recursos de segurança pelas unidades subordinadas à Coordenadoria de Polícia Judicial que se distribuem por todo o estado do Rio de Janeiro.

Com efeito, as análises apresentadas não são o que há de mais sofisticado em inteligência artificial e Ciência de Dados. Contudo, sua relevância reside no fato de ser um passo inicial para que esse campo do conhecimento encontre seu lugar na segurança institucional do Poder Judiciário. Os resultados apresentados em forma de tabelas encerram informações que podem ser de grande utilidade na atividade de inteligência policial: sendo necessário avaliar o risco de uma missão de apoio a cumprimento de mandado por Oficial de Justiça, informações analíticas sobre os níveis de criminalidade de uma localidade são de extrema importância. De forma análoga, essas mesmas informações

servem para a decisão sobre a instalação de agências bancárias dentro de prédios do Tribunal.

Muitos benefícios podem ser colhidos pela segurança institucional caso decida integrar às suas rotinas técnicas de ciência de dados que já são amplamente empregadas no campo da inteligência artificial. Concordando com a afirmação de que “as organizações devem estar concentradas em buscar casos práticos e de impacto imediato” (CNJ, 2019), acredita-se que o trabalho apresentado neste artigo ganhe relevância ao desenvolver uma aplicação prática de algoritmos de aprendizado de máquina e às atividades de segurança do Tribunal Regional do Trabalho da Primeira Região.

Ressalta-se que este estudo foi conduzido com dados externos ao Tribunal de Justiça. Para que, eventualmente, essa importante ciência esteja plenamente integrada aos processos da área da segurança dos órgãos do Poder Judiciário, é fundamental fortalecer a visão de que os dados são ativos organizacionais. O caminho para isso seria a implementação de um programa de governança de dados, definindo políticas, erigindo princípios e estabelecendo responsabilidades, com o objetivo de criar e fortalecer processos que garantam a qualidade

dos dados produzidos pelas unidades vinculadas à segurança institucional.

## REFERÊNCIAS

CONSELHO NACIONAL DE JUSTIÇA. **Inteligência artificial no Poder Judiciário brasileiro**. Brasília: CNJ, 2019. Disponível em: <https://bibliotecadigital.cnj.jus.br/jspui/bitstream/123456789/98/1/Intelig%20aancia%20Artificial%20no%20Poder%20Judiciario%20Brasileiro.pdf>. Acesso em: 10 set. 2023.

CONSELHO NACIONAL DE JUSTIÇA. **Resolução n. 332, de 21 de agosto de 2020**. Dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência artificial no Poder Judiciário e dá outras providências. Brasília: CNJ, 2020a. Disponível em: <https://atos.cnj.jus.br/files/original191707202008255f4563b35f8e8.pdf>. Acesso em: 30 ago. 2023.

CONSELHO NACIONAL DE JUSTIÇA. **Resolução n. 344, de 9 de setembro de 2020**. Regulamenta o exercício do poder de polícia administrativa no âmbito dos tribunais, dispondo sobre as atribuições funcionais dos agentes e inspetores da polícia judicial. Brasília: CNJ, 2020b. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/3460>. Acesso em: 20 set. 2023.

CONSELHO NACIONAL DE JUSTIÇA. **Resolução n. 383 de 25 de março de 2021**. Cria o Sistema de Inteligência de Segurança Institucional do Poder Judiciário e dá outras providências. Brasília: CNJ, 2021a. Disponível em: <https://atos.cnj.jus.br/files/original22430820210329606257fc252cc.pdf>. Acesso em: 30 ago. 2023.

CONSELHO NACIONAL DE JUSTIÇA. **Resolução n. 435, de 28 de outubro de 2021**. Dispõe sobre a política e o sistema nacional de segurança do Poder Judiciário e dá outras providências. Brasília: CNJ, 2021b. Disponível em: <https://atos.cnj.jus.br/files/original152110202111036182a8e64e88e.pdf>. Acesso em: 30 ago. 2023.

FÁVERO, Luiz Paulo; BELFIORE, Patrícia. **Manual de análise de dados**. Rio de Janeiro: LTC, 2021.

KAUFMAN, Dora. **Desmistificando a inteligência artificial**. Belo Horizonte: Grupo Autêntica, 2022. Disponível em: <https://integrada.minhabiblioteca.com.br/#/books/9786559281596/>. Acesso em: 10 set. 2023.

MATOS, Daniel Abud Seabra; RODRIGUES, Erica Castilho. **Análise fatorial**. Brasília: Enap, 2019. 74 p. E-book. Disponível em: <https://repositorio.enap.gov.br/bitstream/1/4790/1/Livro%20An%C3%A1lise%20Fatorial.pdf>. Acesso em: 9 ago. 2023.

PROVOST, Foster.; FAWCETT, Tom. **Data Science para negócios**. Rio de Janeiro: Alta Books, 2016. 408 p.

RUSSELL, Stuart J.; Norvig, Peter. **Artificial intelligence: a modern approach**. 3 ed. Pratices Hall, 2010.

SHARDA, Ramesh.; DELEN, Dursun.; TURBAN, Efraim. **Business Intelligence e análise de dados para gestão do negócio**. 4 ed. Porto Alegre: Bookman, 2019. 584 p.

SEABRA, Marcelo.; SILVA, Antônio da. A atividade de inteligência no Poder Judiciário: a defesa dos Princípios de Bangalore e sua aplicação nos núcleos de pesquisa patrimonial. **Revista CNJ**, Brasília, v 4, n. 1, p. 145-157, jan/jun, 2020. Disponível em: <https://bibliotecadigital.cnj.jus.br/jspui/bitstream/123456789/163/1/A%20atividade%20de%20intelig%20aancia%20no%20Poder%20Judici%20a%20rio.pdf>. Acesso em: 30 ago. 2023.

WICKHAM, Hadley; Mine; GROLEMUND, Garrett. **R for Data Science: Import, Tidy, Transform, Visualize, and Model Data**. [S. l.]: O'Reilly Media, Inc., 2017. 518 p. Disponível em: <https://r4ds.had.co.nz/index.html>. Acesso em: 30 ago. 2023.

## ANEXOS A- Resultados das análises

Anexo A

							(continua)
ranking	municípios	risco	indice_criminal	fator_pc1	fator_pc2	cluster_H	cluster_K
1	São Gonçalo	alto	3,861			1	4
2	Duque de Caxias	alto	3,683	4,69475438		1	4
3	Nova Iguaçu	alto	3,111		1,270750478	2	4
4	Niterói	alto	2,201		3,416408918	3	1
5	São João de Meriti	alto	1,659			4	4
6	Belford Roxo	alto	1,414		-1,2198466	4	4
7	Campos dos Goytacazes	alto	1,137		3,570573697	3	1
8	Itaboraí	alto	0,556		0,072256075	4	3
9	Cabo Frio	alto	0,548		2,102891526	3	1
10	Macaé	alto	0,528		1,1651961	3	3
11	Petrópolis	alto	0,509		3,227328483	3	1
12	Volta Redonda	alto	0,502	0,46771478	2,03706363	3	1
13	Magé	alto	0,419		0,930638835	3	3
14	Queimados	alto	0,295			4	2
15	Rio das Ostras	alto	0,257		1,248447283	3	3
16	Angra dos Reis	alto	0,256		0,661885366	4	3
17	Mesquita	alto	0,252			4	2
18	Maricá	alto	0,241		1,051118576	3	3
19	Nilópolis	alto	0,232			4	2
20	Itaguaí	alto	0,18	0,22718559	-0,07278224	4	2
21	Nova Friburgo	alto	0,158		1,681853613	3	3
22	Teresópolis	alto	0,13		1,30606641	3	3
23	Araruama	alto	0,115		1,106981367	3	3
24	Barra Mansa	médio	0,075		1,387175	3	3
25	Resende	médio	0,049		1,11541162	3	3
26	Três Rios	médio	-0,07		0,911946986	4	3
27	Seropédica	médio	-0,07		-0,48909693	4	2
28	Japeri	médio	-0,085			4	2
29	Saquarema	médio	-0,105		0,358176933	4	3
30	Barra do Piraí	médio	-0,11		0,369875239	4	3
31	São Pedro da Aldeia	médio	-0,127		0,338234936	4	3
32	Itaperuna	médio	-0,163		0,654609771	4	3
33	Paraty	médio	-0,212			4	2
34	Valença	médio	-0,212			4	2
35	Guapimirim	médio	-0,229			4	2
36	Armação dos Búzios	médio	-0,23		0,13017894	4	2
37	Mangaratiba	médio	-0,251			4	2
38	Arraial do Cabo	médio	-0,252			4	2
39	Cachoeiras de Macacu	médio	-0,274			4	2
40	Rio Bonito	médio	-0,278			4	2
41	Casimiro de Abreu	médio	-0,283			4	2
42	São João da Barra	médio	-0,284			4	2
43	Paracambi	médio	-0,29		-0,22990972	4	2
44	Santo Antônio de Pádua	médio	-0,302		0,034815666	4	2
45	São Francisco de Itabapoana	médio	-0,306			4	2

Anexo A

ranking	municípios	risco	indice_criminal	fator_pc1	fator_pc2	cluster_H	(conclusão) cluster_K
46	Itatiaia	médio	-0,315			4	2
47	Bom Jesus do Itabapoana	médio	-0,324			4	2
48	Paraíba do Sul	médio	-0,327		-0,06672624	4	2
49	Miracema	médio	-0,331			4	2
50	Vassouras	médio	-0,333			4	2
51	Iguaba Grande	médio	-0,335			4	2
52	Tanguá	médio	-0,35			4	2
53	Piraí	médio	-0,35			4	2
54	São Fidélis	médio	-0,355			4	2
55	Pinheiral	médio	-0,367			4	2
56	Bom Jardim	médio	-0,367			4	2
57	Quissamã	médio	-0,37			4	2
58	Cordeiro	médio	-0,371		-0,25581026	4	2
59	Miguel Pereira	médio	-0,377			4	2
60	Itaocara	médio	-0,379			4	2
61	Porciúncula	médio	-0,385			4	2
62	Paty do Alferes	médio	-0,386		-0,39314607	4	2
63	Quatis	médio	-0,388			4	2
64	Silva Jardim	médio	-0,388			4	2
65	Porto Real	médio	-0,388			4	2
66	Conceição de Macabu	médio	-0,395			4	2
67	Cardoso Moreira	médio	-0,397			4	2
68	Cantagalo	médio	-0,399			4	2
69	Mendes	baixo	-0,409			4	2
70	Sapucaia	baixo	-0,411			4	2
71	São José do Vale do Rio Preto	baixo	-0,414			4	2
72	Carmo	baixo	-0,415			4	2
73	Aperibé	baixo	-0,416			4	2
74	Sumidouro	baixo	-0,417			4	2
75	Italva	baixo	-0,417			4	2
76	Natividade	baixo	-0,417			4	2
77	Cambuci	baixo	-0,42			4	2
78	Rio Claro	baixo	-0,424			4	2
79	Duas Barras	baixo	-0,429			4	2
80	Santa Maria Madalena	baixo	-0,429			4	2
81	Engenheiro Paulo de Frontin	baixo	-0,43			4	2
82	Carapebus	baixo	-0,432			4	2
83	Laje do Muriaé	baixo	-0,438			4	2
84	Rio das Flores	baixo	-0,439			4	2
85	Trajano de Moraes	baixo	-0,44			4	2
86	Areal	baixo	-0,44			4	2
87	Comendador Levy Gasparian	baixo	-0,44			4	2

---

88	Macuco	baixo	-0,443		-0,56460213	4	2
89	Varre-Sai	baixo	-0,443		-0,55117658	4	2
90	São Sebastião do Alto	baixo	-0,447		-0,5612774	4	2
91	São José de Ubá	baixo	-0,45			4	2

Fonte: elaboração do autor

**Pedro Ivo Vasconcellos da Costa Pinto**

Bacharel em Letras: Português/Literaturas (2010). Mestre em Linguística (2014) Especialista em Data Science e Analytics (2023). Agente da Polícia Judicial no TRT-1ª Região.

