

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS
NÚCLEO DE EDUCAÇÃO A DISTÂNCIA
Pós-graduação *Lato Sensu* em Ciência de Dados e Big Data

Arthur Pereira dos Santos

**Análise comparativa de ocorrências policiais e os gastos em segurança
pública no estado do Rio de Janeiro**

Belo Horizonte

2022

Arthur Pereira dos Santos

Análise comparativa de ocorrências policiais e os gastos em segurança pública no estado do Rio de Janeiro

Trabalho de Conclusão de Curso apresentado ao Curso de Especialização em Ciência de Dados e Big Data como requisito parcial à obtenção do título de especialista.

Belo Horizonte

2022

SUMÁRIO

1. Introdução	4
1.1. Contextualização	4
1.2. O problema proposto	4
1.3. Objetivos	5
3. Processamento/Tratamento de Dados.....	12
4. Análise e Exploração dos Dados	16
5. Criação de Modelos de Machine Learning.....	21
6. Interpretação dos Resultados.....	28
7. Apresentação dos Resultados	31
8. Links	32
REFERÊNCIAS	33
APÊNDICE	Erro! Indicador não definido.

1. Introdução

1.1. Contextualização

Conhecido por seus pontos turísticos e pluralidade cultura, o Rio de Janeiro sempre foi referenciado como um cenário de destaque tanto nacional quanto no mundo. De eventos políticos como a RIO-92, festivais nacionais e internacionais como a Bienal do Livro e o Rock in Rio e até os Jogos Olímpicos de Verão de 2016, sua notoriedade turística e televisiva sempre foi ponto chave na economia local.

Uma preocupação, porém, sempre foi a violência e alto número de casos de policiais que ocorrem por todo o estado. Entre noticiários e retratações, é possível ver um cenário preocupante tanto com a violência vivida por sua população quanto com o descaso das autoridades do governo.

Com a degradação do programa do UPP (Unidades de Polícia Pacificadoras), os altos índices de desemprego, o histórico de violência policial e a constante crise financeira vivida pelo Estado, o cenário de violência e sensação de abandono do poder público é instaurado, conforme descrito por Carneiro, Julia (2019).

Porém, com a possibilidade de utilizar os portais de transparência, é possível realizar estudos e entender melhor qual o cenário da violência no estado, bem como a destinação de seus recursos.

1.2. O problema proposto

Com base nos dados de transparência disponibilizados pelo governo do estado e pela secretaria de segurança pública, é possível criar uma relação de gastos com a segurança pública e o quanto isso tem influenciado na violência registrada no estado. Utilizando a técnica dos 5 porquês [1], chegamos ao seguinte cenário:

(Why?) Por que esse problema é importante?

Esse problema é importante porque nos permite entender melhor o cenário dos gastos públicos sobre a área da segurança no estado e as ocorrências policiais registradas ao longo dos anos.

(Who?) De quem são os dados analisados? De um governo? Um ministério ou secretaria? Dados de clientes?

Os dados foram disponibilizados pelos órgãos do governo do estado do Rio de Janeiro. Os registros de gastos distribuídos por ano foram extraídos do portal de transparência do Estado [2] e a relação de casos pelas regiões foi retirado do site do Instituto de Segurança Pública [3].

(What?): Quais os objetivos com essa análise? O que iremos analisar?

O objetivo é entender se há uma relação entre os gastos do governo do estado e a quantidade de registros de ocorrência no Estado do Rio de Janeiro. Com base nesses levantamentos, é possível entender o melhor o cenário atual do setor de segurança da região, e se é possível imaginar ou prever o cenário futuro.

(Where?): Trata dos aspectos geográficos e logísticos de sua análise.

O estudo trata da região do estado do Rio de Janeiro em sua totalidade. Na análise exploratória de dados também é possível identificar através de filtragens pelas categorizações dos dados setores menores (municípios) de todo o estado.

(When?): Qual o período está sendo analisado? A última semana? Os últimos 6 meses? O ano passado?

O período a ser analisado, com base nos dados obtidos, é de janeiro de 2016 até dezembro de 2021.

1.3. Objetivos

O objetivo do estudo é entender a relação entre os gastos com segurança pública do governo do estado e a quantidade de ocorrências policiais registradas pelas delegacias do Rio de Janeiro. Com base nessas informações, é possível entender melhor o cenário de violência vivido pela população da região e preparar previsões estratégias para que a quantidade de registros diminua futuramente.

2. Coleta de Dados

Para esse estudo, foram utilizadas duas bases de dados principais e uma base auxiliar para apoio na criação de indicadores.

A primeira base, nomeada como “ocurrences”, foi extraída do site do Instituto de Segurança Pública (ISP) [3], e contém 9900 registros. Os seguintes dados estão registrados na tabela:

Nome da coluna	Descrição	Tipo	Observações
CISP	Código Identificador das Circunscrições Integradas de Segurança Pública - CISP onde a ocorrência foi registrada	Numérico inteiro	
Mês	Mês do ano, no formato numérico, das ocorrências	Numérico inteiro	
Ano	Ano das ocorrências	Numérico inteiro	
mes_ano	União dos campos mês e ano, com separação pelo caractere "m", indicando o mês e ano	Texto	
AISP	Código Identificador das Áreas Integradas de Segurança Pública - AISP em que a ocorrência foi registrada	Numérico inteiro	
RISP	Código Identificador das Regiões Integradas de Segurança Pública - RISP em que a ocorrência foi registrada	Numérico inteiro	
Munic	Nome do Município onde a ocorrência foi registrada	Texto	
Mcirc	Código IBGE de 7 dígitos do município da circunscrição	Numérico inteiro	
Regiao	Região do estado em que foi registrada a ocorrência	Texto	
hom_doloso	Quantitativo de vítimas de casos de homicídios dolosos	Numérico inteiro	
lesao_corp_morte	Quantitativo de vítimas de Lesão Corporal e Morte	Numérico inteiro	
Latrocínio	Quantitativo de vítimas de casos de homicídios dolosos	Numérico inteiro	
Cvli	Quantitativo de vítima de crimes Violentos Letais Intencionais	Numérico inteiro	Homicídio doloso + Lesão corporal seguida de morte + Latrocínio

hom_por_inter v_policial	Quantitativo de vítimas de morte por intervenção de agente do Estado	Numérico inteiro	
letalidade_viole nta	Quantitativo de ocorrências de letalidade violenta	Numérico inteiro	Homicídio doloso + Lesão corporal seguida de morte + Latrocínio + Morte por intervenção de agente do Estado
tentat_hom	Quantitativo de vítimas de tentativa de homicídio	Numérico inteiro	
lesao_corp_dol osa	Quantitativo de vítimas de lesão corporal dolosa	Numérico inteiro	
Estupro	Quantitativo de vítimas de estupro	Numérico inteiro	Em 2009, o Artigo 214 do Código Penal passou a caracterizar Atentado Violento ao Pudor (AVP) como Estupro. Por isso, antes de 2009, a variável "estupro" é na verdade a soma de "estupro" e "AVP".
hom_culposo	Quantitativo de vítimas de Lesão Corporal e Morte	Numérico inteiro	Atropelamento + colisão + outros
lesao_corp_cul posa	Quantitativo de vítimas de lesão corporal culposa (trânsito)	Numérico inteiro	Atropelamento + colisão + outros
roubo_transeu nte	Quantitativo de ocorrências de roubo a transeunte	Numérico inteiro	
roubo_celular	Quantitativo de ocorrências de roubo de telefone celular	Numérico inteiro	
roubo_em_cole tivo	Quantitativo de ocorrências de roubo em coletivo	Numérico inteiro	
roubo_rua	Quantitativo de ocorrências de roubo de rua	Numérico inteiro	Roubo a transeunte + Roubo de celular + Roubo em coletivo
roubo_veiculo	Quantitativo de ocorrências de roubo de veículo	Numérico inteiro	
roubo_carga	Quantitativo de ocorrências de roubo de carga	Numérico inteiro	
roubo_comerci o	Quantitativo de ocorrências de roubo a estabelecimento comercial	Numérico inteiro	
roubo_residenc ia	Quantitativo de ocorrências de roubo a residência	Numérico inteiro	
roubo_banco	Quantitativo de ocorrências de roubo a banco	Numérico inteiro	
roubo_cx_eletr onico	Quantitativo de ocorrências de roubo de caixa eletrônico	Numérico inteiro	
roubo_conduca o_saque	Quantitativo de ocorrências de roubo com condução da vítima para saque em instituição financeira	Numérico inteiro	

roubo_apos_sa que	Quantitativo de ocorrências de roubo após saque em instituição financeira	Numérico inteiro	
roubo_bicicleta	Quantitativo de ocorrências de roubo de bicicleta	Numérico inteiro	
outros_roubos	Quantitativo de ocorrências de roubos que não se enquadram nas demais categorias	Numérico inteiro	Inclui contagem de ocorrências de latrocínio.
total_roubos	Quantitativo de roubos (total da soma das categorias de roubos)	Numérico inteiro	
furto_veiculos	Quantitativo de ocorrências de furto de veículo	Numérico inteiro	
furto_transeun te	Quantitativo de ocorrências de casos de furto a transeunte	Numérico inteiro	
furto_coletivo	Quantitativo de ocorrências de furto em coletivo	Numérico inteiro	
furto_celular	Quantitativo de ocorrências de furto de telefone celular	Numérico inteiro	
furto_bicideta	Quantitativo de ocorrências de furto de bicicleta	Numérico inteiro	
outros_furtos	Quantitativo de ocorrências de furtos que não se enquadram nas demais categorias	Numérico inteiro	
total_furtos	Quantitativo de furtos (total da soma das categorias de furtos)	Numérico inteiro	
Sequestro	Quantitativo de extorsão mediante sequestro (sequestro clássico)	Numérico inteiro	
Extorsão	Quantitativo de ocorrências de extorsão	Numérico inteiro	
sequestro_rela mpago	Quantitativo de vítimas de extorsão com momentânea privação da liberdade (sequestro relâmpago)	Numérico inteiro	
estelionato	Quantitativo de ocorrências de estelionato	Numérico inteiro	
apreensao_dro gas	Quantitativo de ocorrências de apreensão de drogas	Numérico inteiro	Refere-se à aglutinação dos títulos de Uso/Porte, Tráfico e Apreensão de substância entorpecente.
posse_drogas	Quantitativo de registros que possuem algum título referente a posse de drogas	Numérico inteiro	A soma dessas variáveis pode ser maior que o número de registros de apreensão de drogas, pois um mesmo registro pode conter mais de um tipo de apreensão de drogas.
trafico_drogas	Quantitativo de registros que possuem algum título referente a tráfico de drogas	Numérico inteiro	
apreensao_dro gas_sem_autor	Quantitativo de registros que possuem algum título referente a apreensão de drogas sem autor	Numérico inteiro	

recuperacao_v eiculos	Quantitativo de recuperação de veículos	Numérico inteiro	Refere-se à recuperação de veículos, não necessariamente roubados/furtados durante o mês, e/ou roubados/furtados na mesma área.
Apf	Quantitativo de autos de prisão em flagrante	Numérico inteiro	
Aaapai	Quantitativo de autos de apreensão de Adolescentes por prática de ato infracional	Numérico inteiro	
Cmp	Quantitativo de cumprimento de mandados de prisão	Numérico inteiro	
Cmba	Quantitativo de cumprimento de mandados de busca e apreensão	Numérico inteiro	
Ameaça	Quantitativo de pessoas ameaçadas	Numérico inteiro	
peessoas_desap arecidas	Quantitativo de pessoas desaparecidas	Numérico inteiro	
encontro_cada ver	Quantitativo de cadáveres encontrados	Numérico inteiro	
encontro_ossa da	Quantitativo de ossadas encontradas	Numérico inteiro	
pol_militares_ mortos_serv	Quantitativo de policiais militares mortos em serviço	Numérico inteiro	
pol_civis_mort os_serv	Quantitativo de policiais civis mortos em serviço	Numérico inteiro	
registro_ocorre ncias	Quantitativo de Total de Casos ocorridos no mês, na CISP	Numérico inteiro	
Fase	Código de Identificação de situação dos registros	Numérico inteiro	2 - Consolidado sem errata 3 - Consolidado com errata

A segunda base, nomeada como “Gastos Convertidos”, foi extraída do portal de transparência do Estado do Rio de Janeiro [2], e corresponde aos gastos do governo. O arquivo é a junção de 6 bases (divididas anualmente), totalizando 2778 registros. Os seguintes dados estão registrados na tabela:

Campo	Descrição	Tipo	Obs
Ano do Gasto	Ano do Registro	Numérico Inteiro	
Elemento	Detalhamento dos Gastos	Texto	
Fonte de Recursos1	Detalhamento da fonte dos recursos gastos	Texto	
Programa1	Detalhamento de programa ao qual os gastos foram direcionados	Texto	

Projeto / Atividade	Detalhamento do projeto ou atividade ao qual os gastos foram direcionados	Texto	
Subfunção1	Categorização do gasto	Texto	
Table Names	Campo gerado para indicar de qual tabela o dado foi extraído	Texto	
Órgão1	Órgão responsável por efetivar o gasto	Texto	
Abr	Valor gasto no mês de Abril	Valor Monetário	
Ago	Valor gasto no mês de Agosto	Valor Monetário	
Dez	Valor gasto no mês de Dezembro	Valor Monetário	
Fev	Valor gasto no mês de Fevereiro	Valor Monetário	
Jan	Valor gasto no mês de Janeiro	Valor Monetário	
Jul	Valor gasto no mês de Julho	Valor Monetário	
Jun	Valor gasto no mês de Junho	Valor Monetário	
Mai	Valor gasto no mês de Maio	Valor Monetário	
Mar	Valor gasto no mês de Março	Valor Monetário	
Nov	Valor gasto no mês de Novembro	Valor Monetário	
Out	Valor gasto no mês de Outubro	Valor Monetário	
Set	Valor gasto no mês de Setembro	Valor Monetário	
Valor Empenhado	Valor total gasto no registro	Valor Monetário	Quantitativo não é igual à soma dos meses

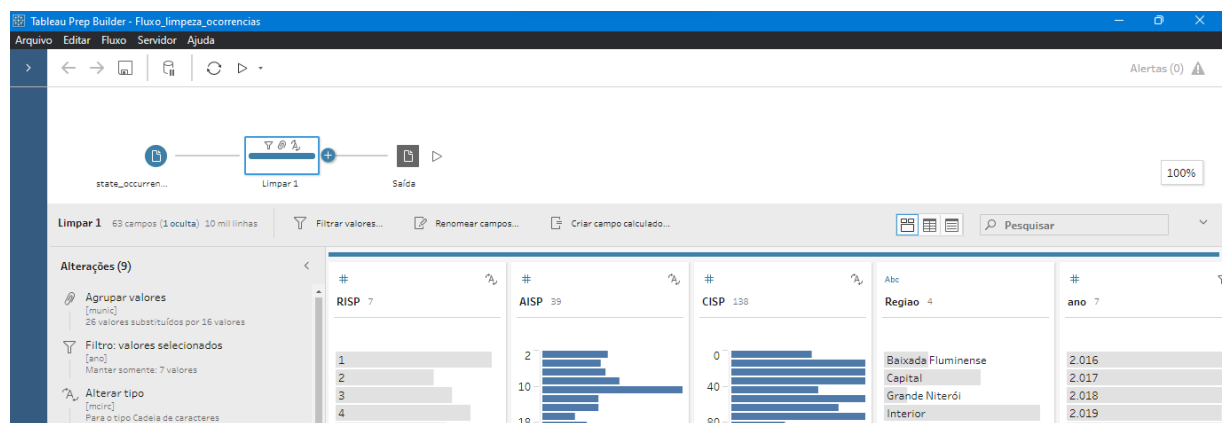
A base auxiliar também foi extraída do Instituto de Segurança Pública (ISP) [3], e contém às relações entre códigos de delegacias e suas regiões de atuação.

Campo	Descrição	Tipo	Obs
CISP	Código Identificador das Circunscrições Integradas de Segurança Pública - CISP onde a ocorrência foi registrada	Numérico Inteiro	
Município	Nome do município de atuação da CISP	Texto	
Região de Governo (Relacao)	Nome da Região em que o CISP atua	Texto	
Unidade Territorial	Nome do Bairro / Região Administrativa que o CISP atua	Texto	
AISP (Relacao)	Código Identificador das Áreas Integradas de Segurança Pública - AISP em que a ocorrência foi registrada	Numérico Inteiro	
RISP (Relacao)	Código Identificador das Regiões Integradas de Segurança Pública - RISP em que a ocorrência foi registrada	Numérico Inteiro	

3. Processamento/Tratamento de Dados

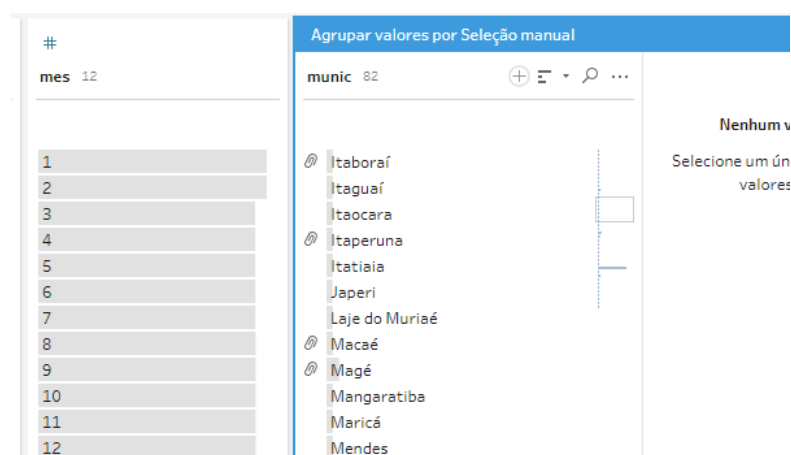
Para o processamento e tratamentos dos dados, foram utilizadas duas ferramentas. O Tableau Prep foi utilizado para identificar e tratar campos nulos, bem como realizar filtragens e criação de campos e grupos para criação de indicadores. O Microsoft Excel foi utilizado para transformação de tipos de campos.

O dataset “ocurrences” recebeu o tratamento no tableau prep conforme figura abaixo:



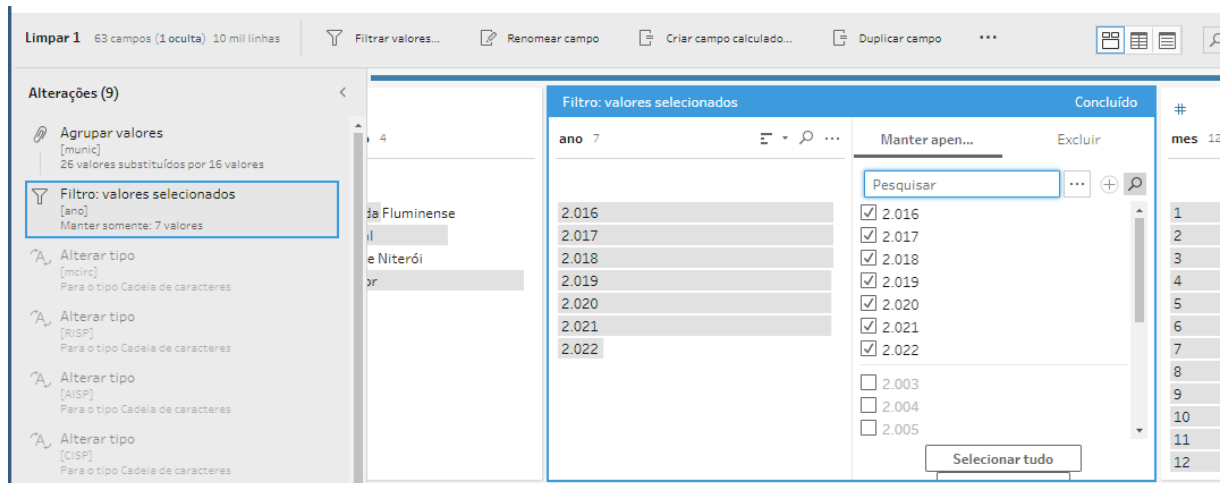
1. Tableau Prep – Preparação de dataset “ocurrences”

A partir da seleção manual, foi possível realizar o agrupamento dos municípios que continham o mesmo nome com alguma variação de caractere:



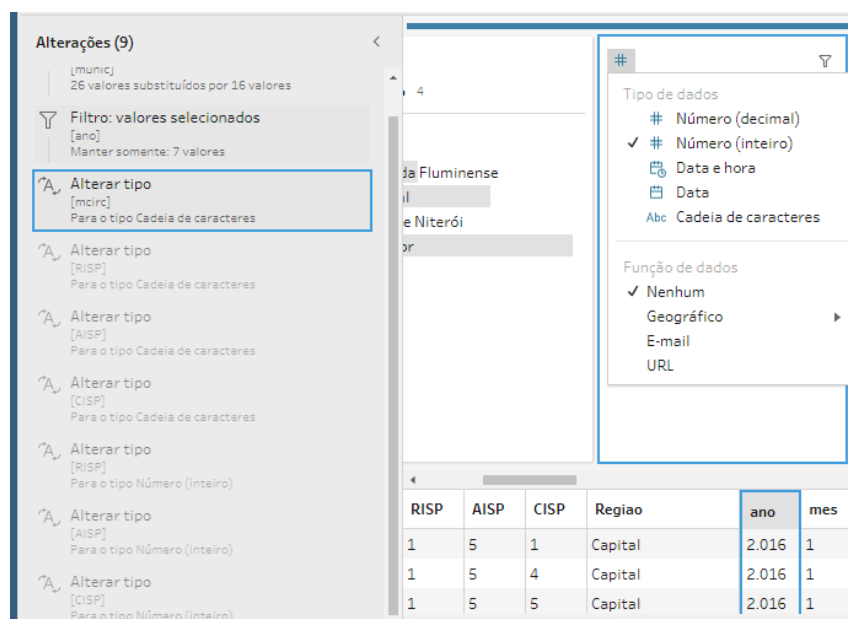
2. Tableau Prep – Agrupando Municípios

Foi utilizado o método de filtragem para a remoção de registro provenientes de anos que não interessassem ao estudo. Foram mantidos apenas os anos de 2016 até 2021:



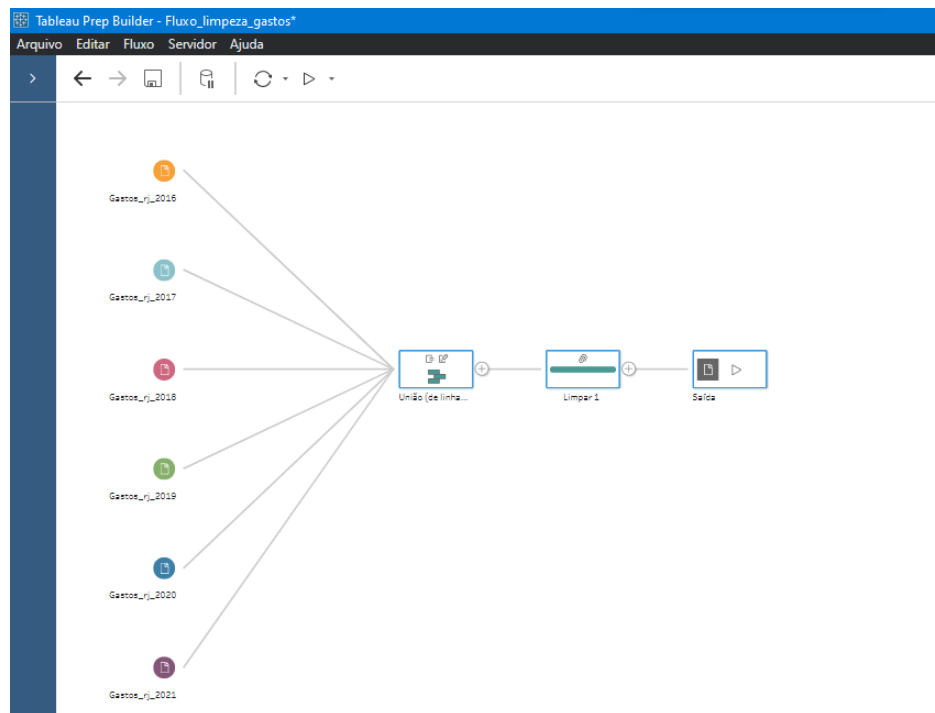
3. Tableau Prep – Filtragem de Anos

Os campos numéricos que não foram reconhecidos pelo Tableau foram manualmente convertidos:



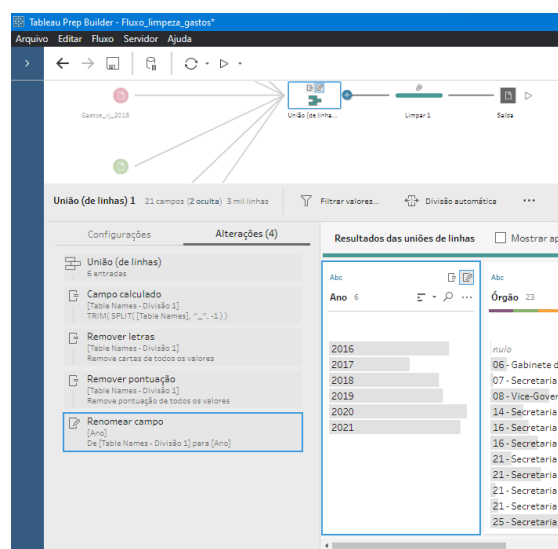
4. Tableau Prep – Conversão de campos

O dataset “Gastos Convertidos” foi tratado no Tableau Prep para união de tabelas distribuídas por ano, bem como tratamento de campos nulos:



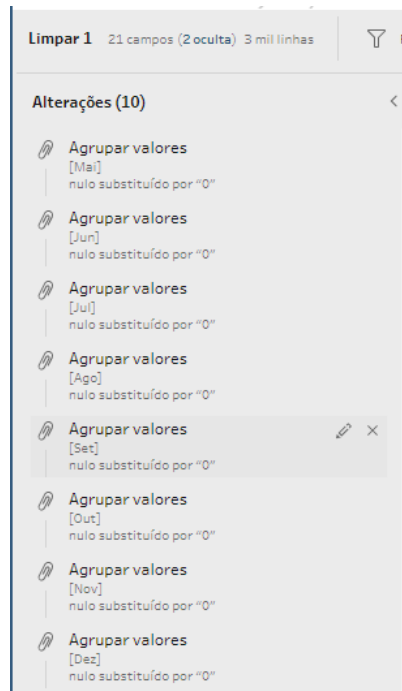
5. Tableau Prep – Ajustes da tabela de Gastos

Foi realizada a união de linhas das tabelas de 2016 até 2021. Também foi realizada a transformação do campo Table Names, gerado automaticamente pelo Tableau para criação de campo “Ano”:



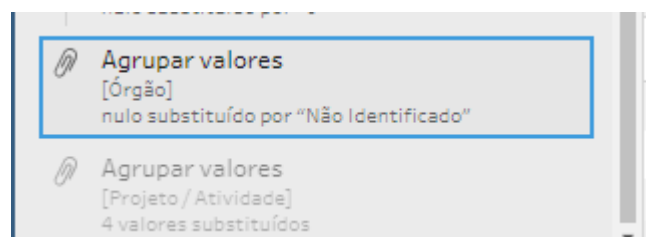
6. Tableau Prep – Atualizando campo Table Names

Foi realizada a atualização dos campos de gastos dos meses, convertendo os valores nulos em “0”. Essa estratégia foi definida por se tratar de um campo de valor monetário.



7. Tableau Prep – Conversão de valor nulo para 0

Para valores nulos identificados no campo “Órgão”, foi adicionado o valor “Não identificado”, por se tratar de um campo textual, utilizado para categorização de registros. O agrupamento de valores no campo “Projeto / Atividade” também foi realizado manualmente para valores com diferenças pequenas de caracteres:

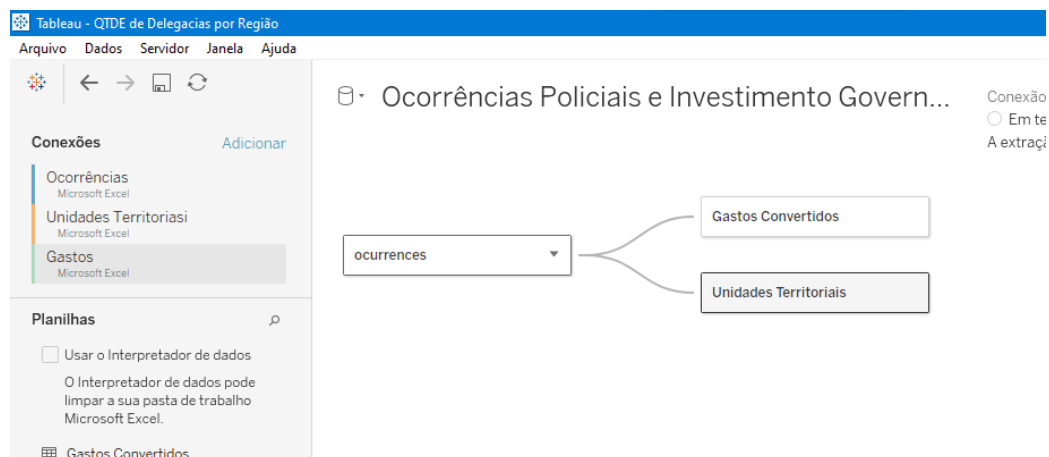


8. Tableau Prep – Alteração de campos nulos em órgão

No Excel, os campos de valores mensais e o campo de valor empenhado foram atualizados manualmente para conversão em números no formato monetário.

4. Análise e Exploração dos Dados

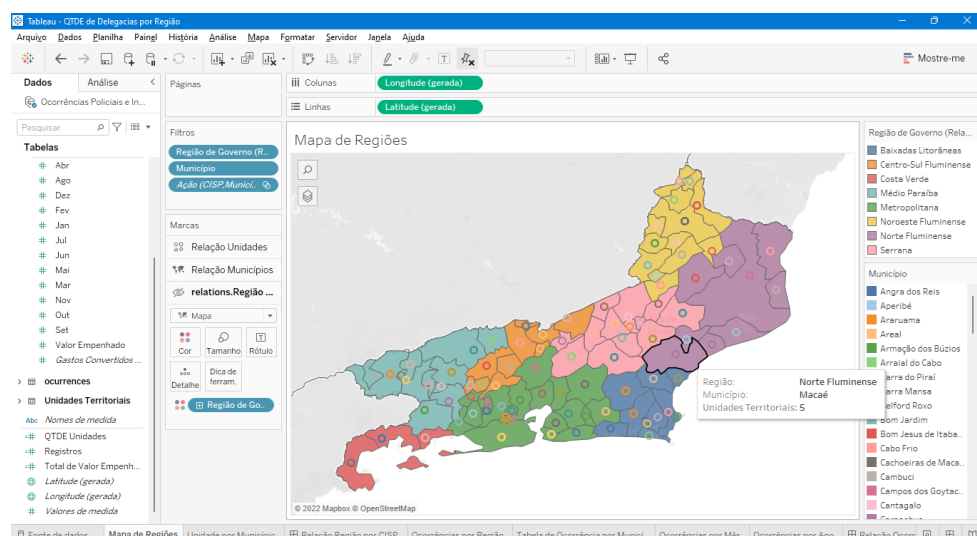
Durante a fase de análise de dados, foi utilizado o Tableau para identificação de valores e criação de indicadores. Com base nos dados obtidos, foi possível estabelecer uma relação entre os registros de ocorrências e os registros de gastos pelo ano. A tabela auxiliar foi relacionada pelo código CISP, correspondente à delegacia que fez o registro de ocorrência.



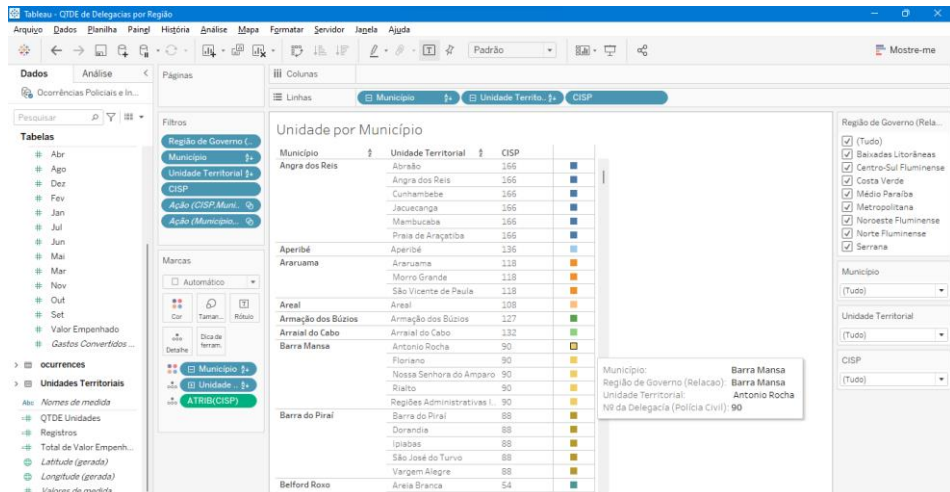
9. Tableau – Relações entre bases

A partir desse processo, foi possível realizar a montagem de diversos indicadores, e explorar tanto os dados de registro, quanto os dados de gastos do governo do estado.

A primeira exploração foi referente às áreas de atuação das delegacias, e aos quantitativos de registros nas regiões:

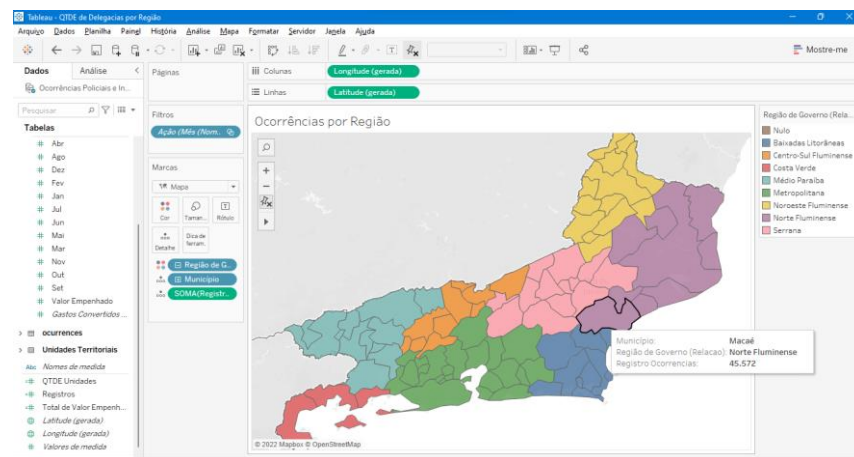


10. Tableau – Dados de Territórios por Município e Região



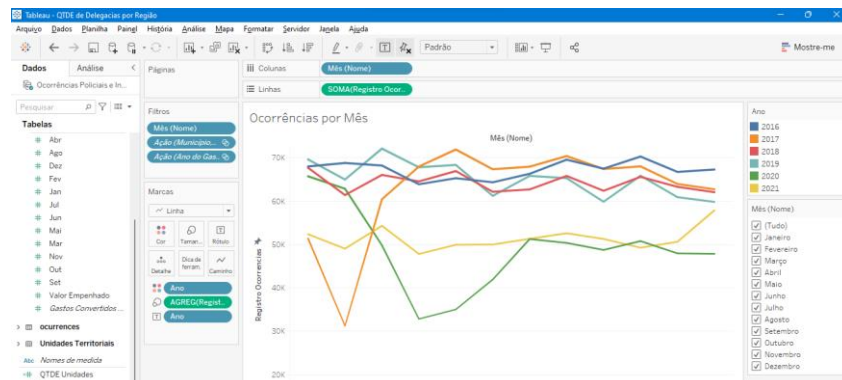
11. Tableau – Relação de Unidades por CISP

Após essa análise, foi possível criar um relacionamento entre os municípios e a quantidade de ocorrências:



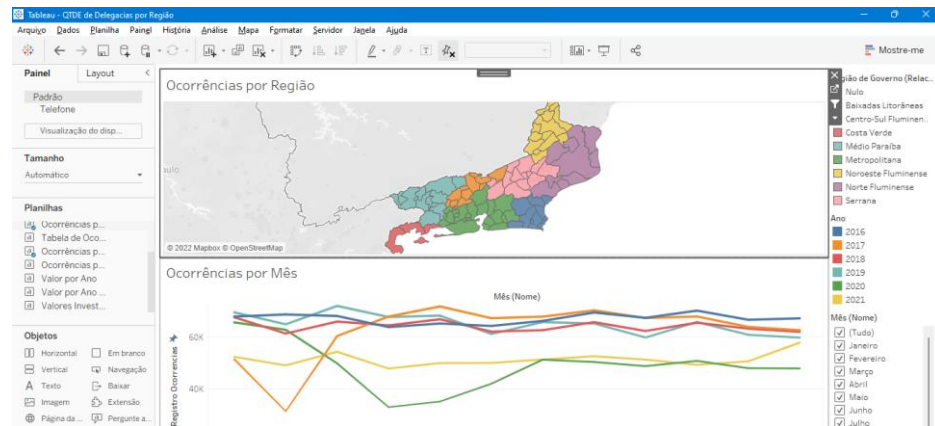
12. Tableau – Relação de Municípios e Ocorrências

Foi possível montar uma relação de quantidade de ocorrências distribuídas por mês e ano:



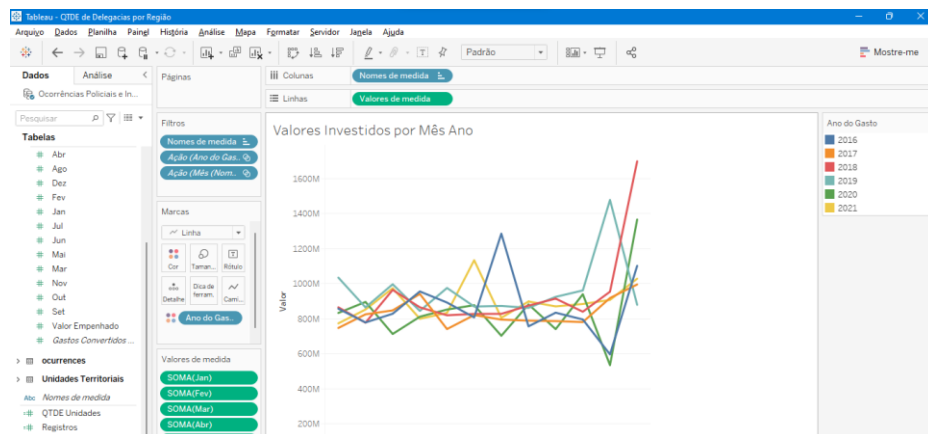
13. Tableau – Quantidade de ocorrências por Mês e Ano

Finalmente, ao combinar os dados explorados, foi possível estabelecer indicadores por tempo e região:



14. Tableau – Relação de ocorrências por Região e Tempo

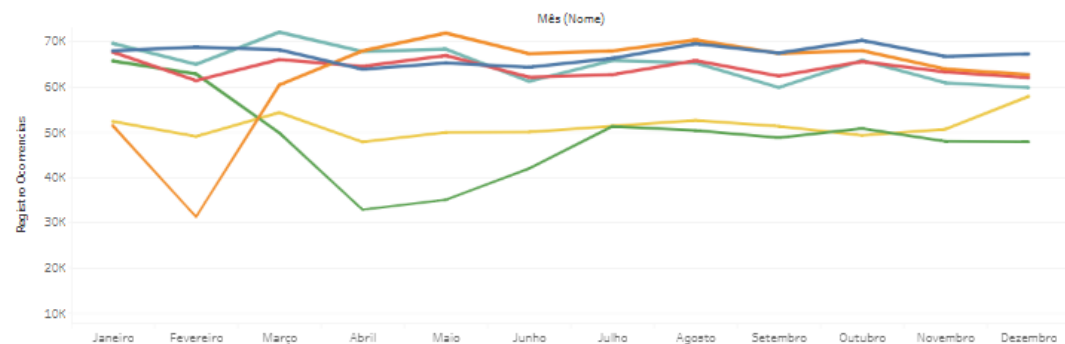
Uma vez explorada a parte de ocorrências, foi feita uma análise na parte de gastos, e sua distribuição pelos anos:



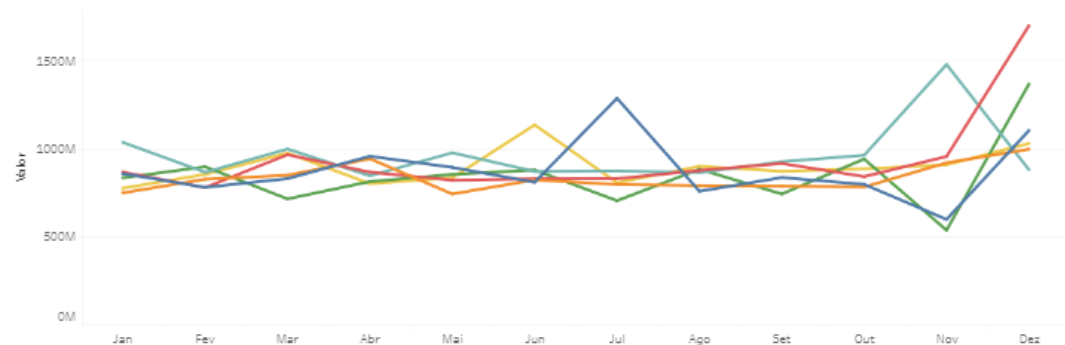
15. Tableau – Distribuição de Gastos por meses e Anos

Assim, foi possível equipara as quantidades de investimento e ocorrências por meses do ano:

Ocorrências por Mês



Valores Investidos por Mês Ano



16. Tableau – Relação de Ocorrências e Gastos por Meses e Anos

Um Ranking criado foi a de Gastos pelas classificações de subfunção e elemento:

Valores por Função e Elemento

Subfunção	Elemento	
122 - Administração Geral	31901200 - Vencimentos e Vantagens Fixas - Pessoal Milit	25.220.528.353
	31901100 - Vencimentos e Vantagens Fixas - Pessoal Civil	13.212.980.750
	31901700 - Outras Despesas Variáveis - Pessoal Militar	7.217.846.029
	31911300 - Obrigações Patronais	7.475.131.220
	33903900 - Outros Serviços de Terceiros - Pessoa Jurídica	2.321.649.822
	31901600 - Outras Despesas Variáveis - Pessoal Civil	1.779.366.522
	33903000 - Material de Consumo	543.852.109
	33909200 - Despesas de Exercícios Anteriores	433.976.408
	44905200 - Equipamentos e Material Permanente	63.781.507
	33904900 - Auxílio Transporte	508.604.627
	33913900 - Outros Serviços de Terceiros - Pessoa Jurídica	5.266.195
	33904600 - Auxílio Alimentação	440.588.239
	31909200 - Despesas de Exercícios Anteriores	276.138.116
	33913000 - Material de Consumo	244.592.704
	33900800 - Outros Benefícios Assistenciais do Servidor e	175.764.067
	44905100 - Obras e Instalações	128.184.217
	33904000 - Serviços de Tecnologia da Informação e Comuni	14.217.914
	31900400 - Contratação Por Tempo Determinado	129.961.287
	33901700 - Outras Despesas Variáveis - Pessoal Militar	126.494.674
	33911700 - Outras Despesas Variáveis - Pessoal Militar	121.318.122
	33903400 - Outras Despesas de Pessoal Decorrentes de Con	12.588.568
	31900300 - Despesas de Pessoal de Apoio	100.000.000
	31900200 - Despesas de Pessoal de Apoio	100.000.000
	31900100 - Despesas de Pessoal de Apoio	100.000.000
	31900000 - Despesas de Pessoal de Apoio	100.000.000
	31900000 - Despesas de Pessoal de Apoio	100.000.000
	31900000 - Despesas de Pessoal de Apoio	100.000.000

17. Tableau – Ranking de gastos por Subfunção e Elemento.

Por fim, com base na análise expiratória, foi identificado que a relação entre os gastos em segurança pública e a quantidade de ocorrências é de que não existe influência no quanto é gasto pelo governo e a quantidade de casos.

Isso pode ser percebido tanto pelos picos de investimento nos meses finais do ano e a continuidade dos níveis de casos nos mesmos períodos e meses subsequentes, quanto pela rápida retomada ao patamar médio, mesmo que os investimentos fossem mantidos. Com exceção de dois momentos (fevereiro de 2017, durante a greve de Policiais Militares [4] e abril de 2020, mês subsequente ao decreto de estado de pandemia causada pelo Covid-19 [5]), as quantidades de registros se mantiveram constantes ao longo dos anos.

Foi identificado também que os maiores gastos do Estado no setor de segurança são na área Administração Geral, em vencimentos para pessoal militar, seguido por vencimentos para pessoal civil. Nenhuma outra categorização de gastos chega na Dezena de bilhões de Reais no recorte do início 2016 até o fim de 2021 como essas categorizações.

5. Criação de Modelos de Machine Learning

Ao utilizar os dados históricos dos portais de transparência e com o objetivo de entendimento do cenário de violência com base nos registros de ocorrência, para assim identificar estratégias para melhorar a situação de segurança do estado, uma estratégia que pode ser adotada é a utilização de modelos de Machine Learning para predição de resultados futuros.

Com base no modelo de Vasandani [10] de fluxo de trabalho para utilização de modelos de Machine Learning, foi possível realizar a preparação e a identificação dos resultados.

Título: Análise comparativa de ocorrências policiais e os gastos em segurança pública no estado do Rio de Janeiro		
<p>1 Estabelecendo o Problema:</p> <p>O Problema proposto é o entendimento da relação entre os gastos do governo na área de segurança pública e a quantidade de casos de ocorrências registradas pelo Estado.</p> <p>O estudo busca identificar motivos para o cenário de caos e insegurança vivido pelo estado, mesmo com orçamentos de bilhões ao longo dos anos.</p>	<p>2 Resultados / Previsões</p> <p>Com base na série histórica de casos por mês, separados por diversas categorias durante os anos de 2016 até o final de 2021, é possível realizar métodos de machine learning para previsão do cenário de violência dos primeiros meses do ano de 2022.</p> <p>Os resultados mostram que não há uma relação direta entre os gastos do governo e os casos de violência, visto que a maior parte dos gastos é referente ao setor administrativo e não em equipamentos e setores de inteligência.</p>	<p>3 Aquisição de Dados</p> <p>Todas as bases de dados utilizadas no estudo foram retiradas de sites oficiais do governo do estado. Os dados orçamentários foram retirados do portal de transparência do governo, e os dados de casos policiais foram extraídos do portal do instituto de segurança pública do estado.</p> <p>A ideia inicial era utilizar a base de dados de um recorte histórico maior (trazendo dados de 2003 até 2021), porém os portais o portal orçamentário exibiam erros durante as extrações, por isso o recorte foi limitado, porém suficiente para a análise proposta.</p>
<p>4 Modelagem</p> <p>Os modelos de Machine Learning adequados para o problema proposto são os modelos referentes à análise de séries temporais, com base nas variáveis que podem ser utilizadas para auxiliar nas previsões de resultados mais fiéis possíveis com a realidade.</p> <p>Com base na utilização de identificação de sazonalidade, entendimento de linhas de tendência e ajustes por diferenciação de por meio de decomposição de sinais, foram identificados os modelos de SARIMA e regressão linear simples e múltipla previsões dos meses iniciais do ano de 2022.</p>	<p>5 Avaliação do Modelo</p> <p>Com base na utilização dos modelos e suas previsões foi possível identificar que os modelos de regressão linear e SARIMA trouxeram dados mais fiéis ao resultado real, enquanto a regressão linear múltipla não trouxe dados adequados. Os ajustes realizados nas variáveis de controle como a sazonalidade do SARIMA fizeram os resultados tornarem-se ainda mais fiéis. Possivelmente com um recorte maior, os dados da predição poderiam ser ainda mais precisos.</p>	<p>6 Preparação de Dados</p> <p>Para a utilização dos dados nos modelos de machine learning, foi necessário realizar a filtragem pelos dados que seriam importantes para a predição (quantidade de ocorrências), a transformação de dados por meio de agrupamento por período mensal e o tratamento dos dados nulos, transformando-os em "0" para a quantidade de registros de ocorrências.</p>

Como a base traz uma série histórica, foram testados modelos referentes à análise temporal para previsão quantidade de ocorrências nos meses futuros (dos meses de janeiro à junho de 2022). Com base nos dados disponibilizados até a data de encerramento deste estudo (abril de 2022) é possível ainda utilizar uma comparação com os dados dos quatro primeiros meses com os resultados dos modelos.

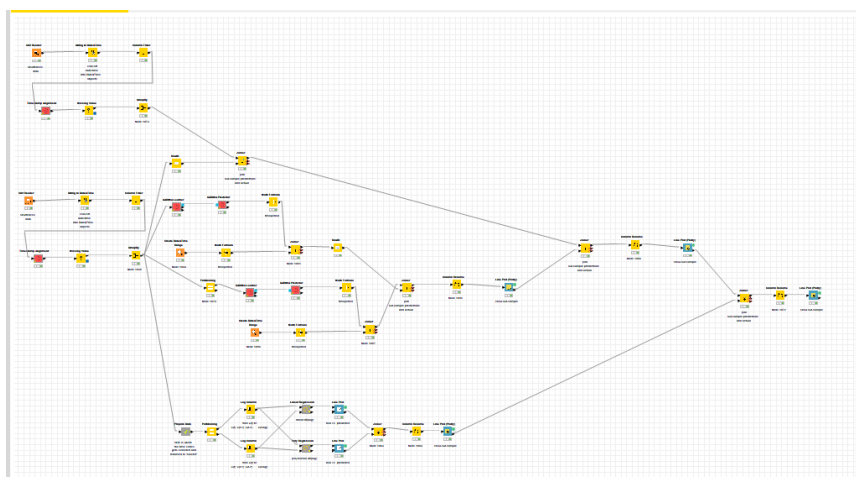
Os modelos escolhidos para a análise temporal foram:

- SARIMA (Seasonal Autoregressive Integrated Moving Average), uma variação do modelo de Machine Learning por média móvel integrada autorregressiva (em inglês, ARIMA), com o componente de análise de sazonalidade. Sua escolha se deu pelo fato da análise do aspecto da sazonalidade;

- Regressão Linear simples, modelo que utiliza o conceito de correlação linear, traçando uma previsão com base no ponto médio dos dados históricos (analisado pelo coeficiente de correlação);
- Regressão Linear Multipla, modelo que a conceituação de regressão linear simples com mais de uma variável quantitativa (no caso, utilizando a previsão com base nos outro modelos para a análise).

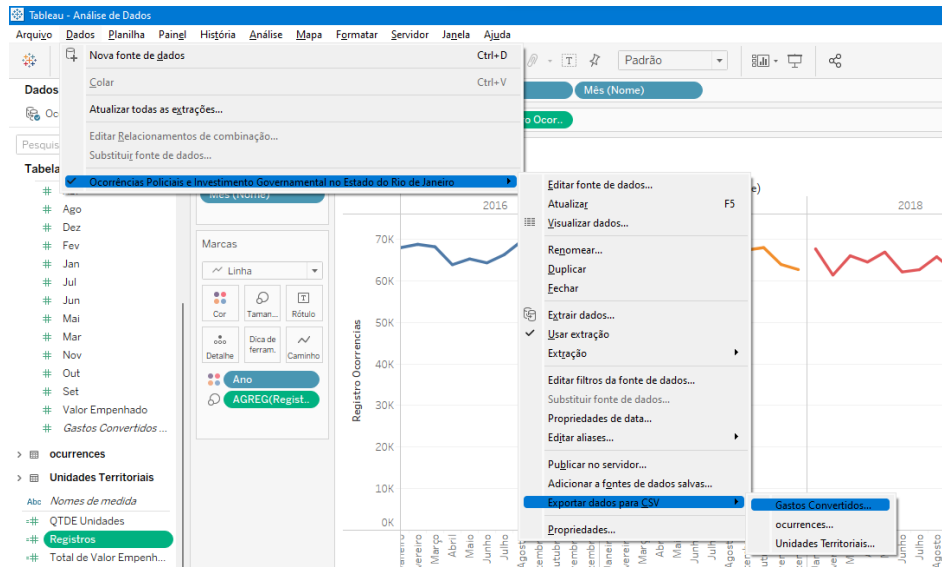
Para a análise final, foi utilizado o modelo SARIMA para a previsão pela maior proximidade dos resultados encontrados com a trajetória real da série (durante a validação inicial, foi utilizada a previsão dos seis últimos meses de 2021, o que seria o final da série histórica).

Para a utilização dos modelos, foi utilizada a ferramenta “Knime Analytics Platform”, um programa de Recuperação e manipulação de dados e execução de modelos de machine learning em formato de fluxo de trabalho.



18 – Fluxo de análise de série histórica de ocorrências com Knime Analytics Platform.

Com o Knime, cada elemento do fluxo de trabalho realiza alguma operação. É possível distribuir cada setor do fluxo em série de operações para realizar determinadas transformações. Com a ferramenta, carregamos os mesmos dados extraídos dos portais de transparência. Caso deseje utilizar os dados já tratado, também é possível utilizar dados exportados do tableau para realização das análises por Machine Learning.

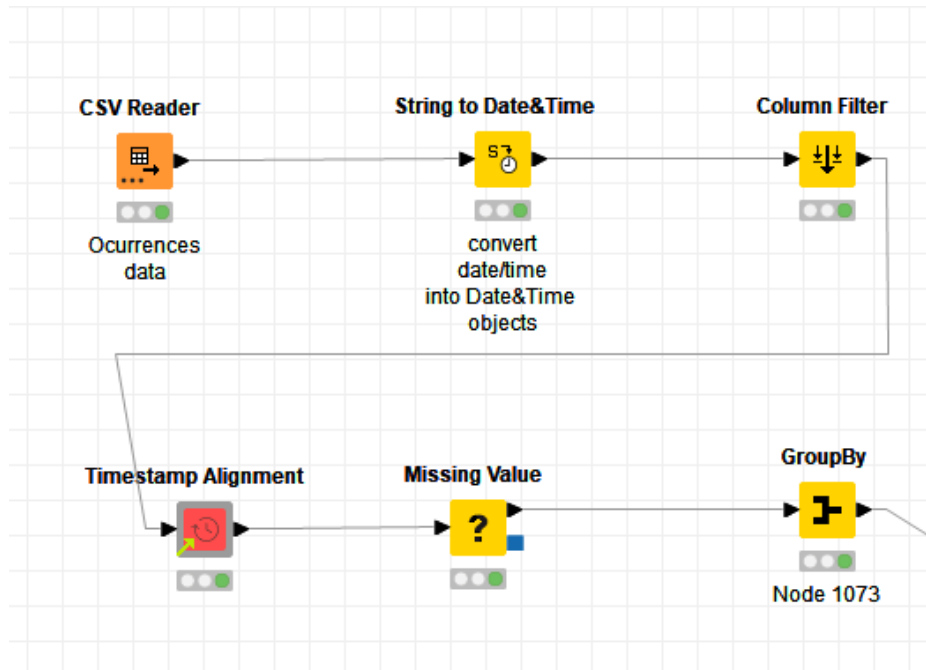


19 - Exportação de Dados para utilização no Knime.

Com os dados exportados, podemos utilizar o componente de leitura de arquivo csv para carregar os dados no fluxo de trabalho. Com os dados carregados, podemos dar início às transformações e filtrações necessárias para adequação ao modelo.

No primeiro bloco utilizaremos o elemento de leitura de dados:

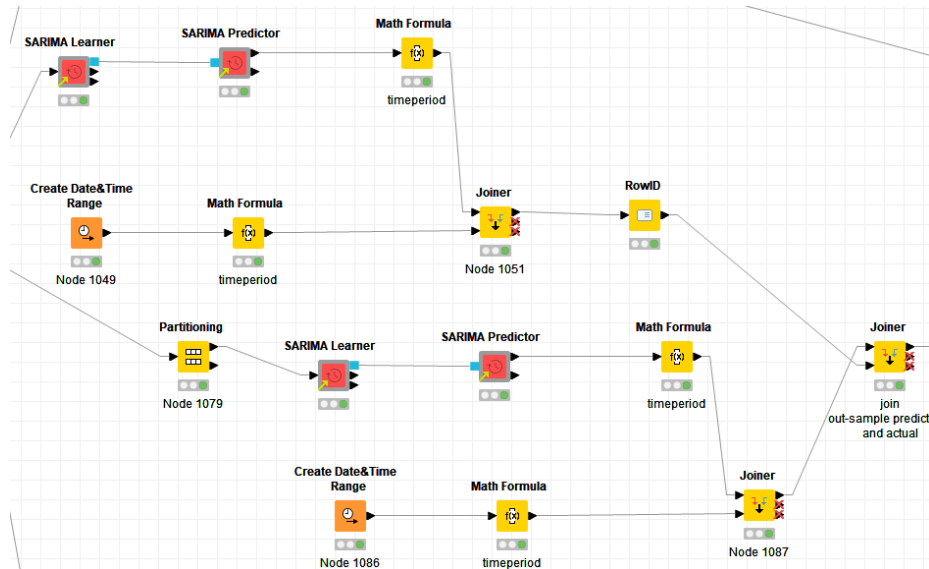
- Transformação de dado de texto para data;
- Filtragem de colunas para manter apenas os elementos necessários para análise (no caso, o total de registro de ocorrências da linha e a coluna de data transformada);
- Elemento de adequação de campo de data para informação interpretada no Knime (transformação de data em granularidade que será utilizada, no caso, o mês do ano);
- Tratamento de dados nulos (não foi identificado dados nesse momento, mas será mantido nesse bloco para futuras análises) – substituição do valor nulo para “0” no campo de total de ocorrências;
- Agrupamento de valores por valor mensal – identificado através de elemento de transformação por granularidade).



20 – Bloco de Leitura e preparação de dados para série histórica.

O segundo bloco contém os elementos referentes ao fluxo de execução do modelo SARIMA:

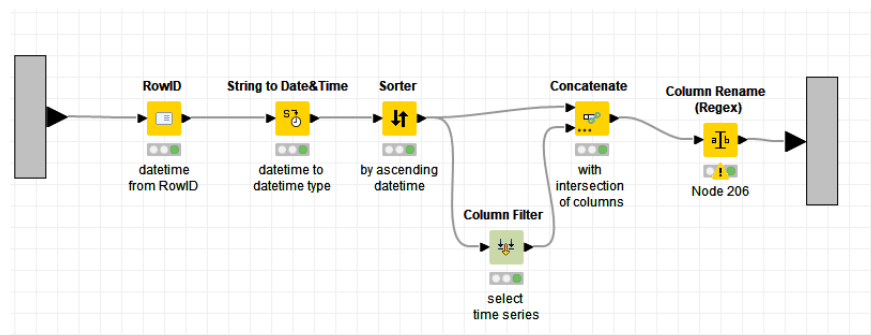
- SARIMA Learner – Elemento de aprendizado do modelo, que executa as leituras dos dados e inicia o processo de identificação das variáveis de controle para identificação dos dados. Para a aplicação desse modelo, foi utilizada a sazonalidade de 12 meses por tratar de série anual completa;
- SARIMA Predictor – Elemento de identificação e exposição da predição realizada pelo SARIMA Learner;
- Elemento de fórmula matemática para cálculo e identificação de campos na tabela de predição (operacional);
- Elemento de criação de faixa de datas, para identificação e controle de qual data seria prevista;
- Elemento de junção de tabelas de predição com tabela de dados reais;
- Elemento de partição de dados para treinamento e validação;



21 – Bloco de Modelo SARIMA e validações

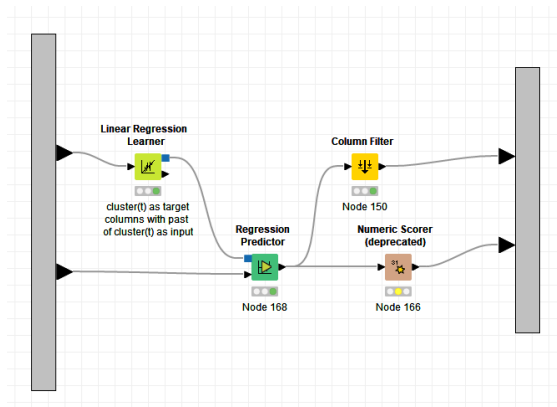
O terceiro bloco realiza as operações referentes aos modelos de Regressão Linear Simples e Regressão Linear Multipla:

- Elemento de Preparação de Dados – Nó de elementos internos do Knime para validação, transformação de dados em Data, ordenação de dados e filtragem de colunas;

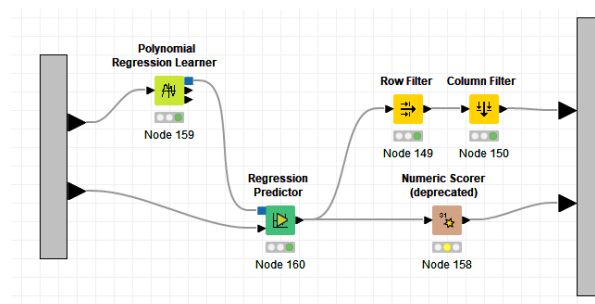


22 – Sub Fluxo de trabalho de elemento de preparação de dados

- Particionamento de Dados – Para particionamento entre dados de validação e treinamento;
- Elemento de calculo matemático para identificação de variável de controle dos Modelos;
- Elemento de Regressão Linear (Simples e Multipla) – Com aplicação de elemento de regressão linear, predição por elemento de regreção e dados de validação, elemento de validação de acertos do modelo e filtragem de dados;

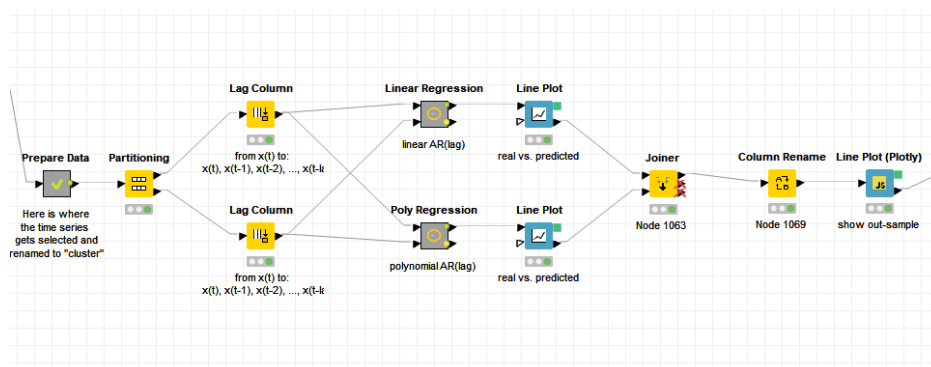


23 – Sub Fluxo de trabalho de regressão linear simples



24 – Sub Fluxo de trabalho de regressão linear múltipla

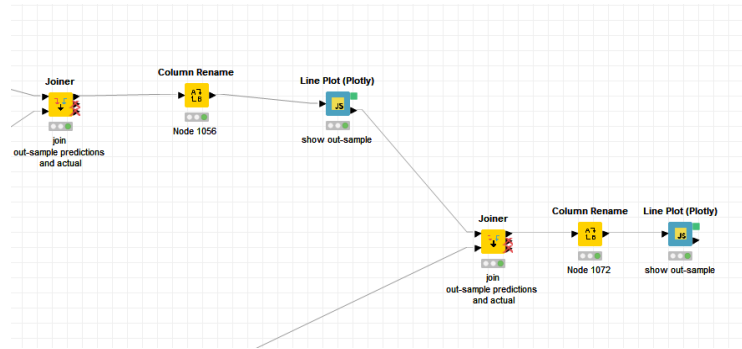
- Elemento de plotagem para verificação visual do resultado da predição;
- Elemento de junção dos resultados das predições;
- Elemento de renomeação de coluna para controle interno de dados.



25 – Bloco de modelos de regressão linear simples e múltipla

O quarto bloco reúne as predições e a série histórica real, realizando criação de imagens para apresentação dos dados resultantes dos modelos de machine learning:

- Elemento de junção de dados de modelos;
- Elemento de renomeação de colunas resultantes de modelos;
- Elemento de criação de imagem das linhas temporais e predições;



26 – Bloco de visualização final de resultados

Com base nos resultados dos modelos, pode-se ver que o modelo SARIMA traz a predição mais próxima dos dados reais com relação aos últimos 6 meses do ano de 2021, e por isso foi utilizado para a predição dos primeiros seis meses do ano de 2022.

Para ajudar na validação do resultado final, foram extraídos novos dados do portal de transparência do governo com relação aos números de segurança do Estado. Foram extraídos (com base no mesmo processo descrito nos capítulos anteriores) os dados de janeiro à abril de 2022 (mês de encerramento e entrega desse estudo).

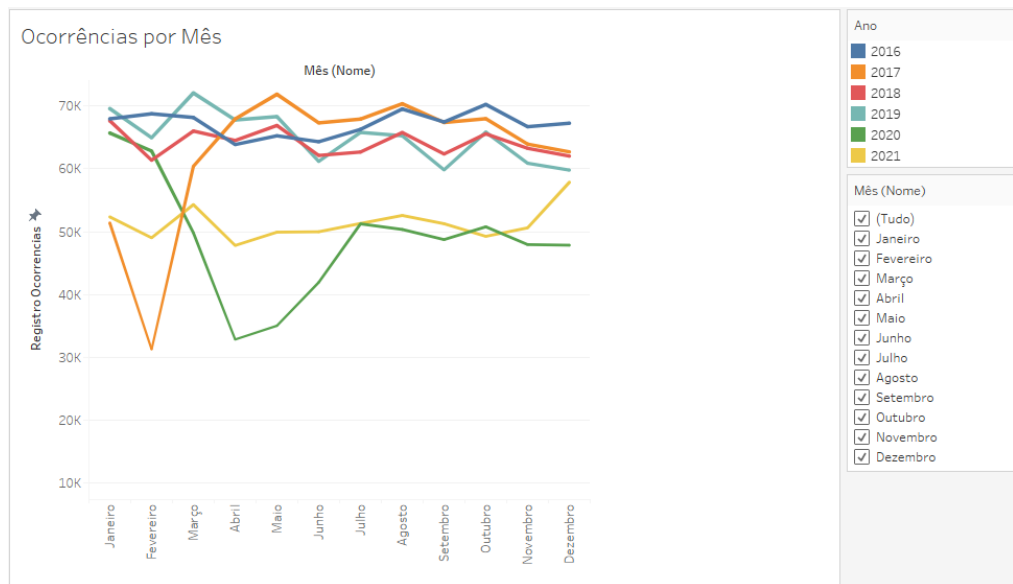
Os dados extraídos receberam o mesmo tratamento identificado no primeiro bloco diretamente no Knime, trazendo uma série histórica dos primeiros meses de 2022. Esses dados foram incorporados aos resultados finais do modelo com o elemento de junção.



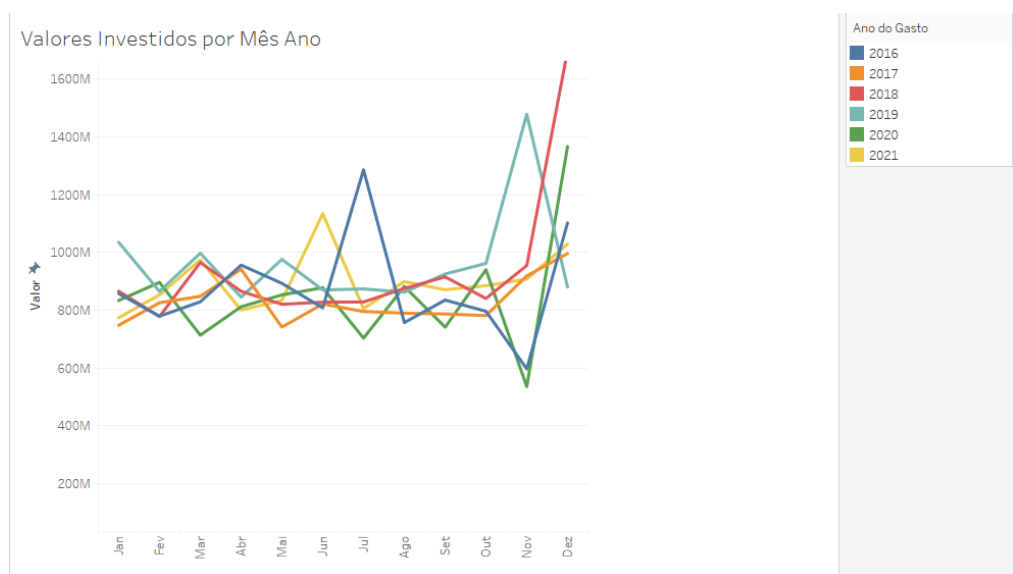
27 – Exibição final de resultados de previsões por cada modelo e de dados reais

6. Interpretação dos Resultados

Com base na análise exploratória de dados e nas análises identificadas pelos modelos de machine learning, pudemos observar que o nºs de casos ao longo da série histórica não costuma tender para um declínio nem uma alavancagem. Os pontos mais baixos da série não são reflexos em si de nenhuma alta no investimento ou métodos de prática de segurança, mas sim de momentos particulares da história (Fevereiro de 2017 – Mês da chegada da Greve dos policiais e protestos de suas famílias por melhores condições de trabalho; e Março de 2020 – Decreto de estado de Pandemia Global pela COVID - 19):



28 – Tela de montagem de grafico de ocorrências por mês



29 – Tela de montagem de valores investidos por mês

Também é possível identificar pelos gráficos, que mesmo em momentos de picos de gastos em segurança, a quantidade de casos manteve-se dentro de sua constância, mesmo em momentos de conhecida movimentação turística, como os meses de carnaval e o período dos jogos olímpicos de verão de 2016.

Analisando os dados em si e não sua visualização histórica, identificamos alguns fatores que podem auxiliar a entender o cenário:

Valores por Função e Elemento

Subfunção	
122 - Administração Geral	60.967.892.823
125 - Normatização e Fiscalização	2.601.676.873
181 - Policiamento	2.465.971.647
306 - Alimentação e Nutrição	1.498.089.647
182 - Defesa Civil	594.597.693
422 - Direitos Individuais, Coletivos ..	571.245.369
302 - Assistência Hospitalar e Ambu..	440.079.863
332 - Relações de Trabalho	97.333.536
421 - Custódia e Reintegração Social	91.931.497
782 - Transporte Rodoviário	80.732.982
126 - Tecnologia da Informação	70.125.266
781 - Transporte Aéreo	48.525.116
128 - Formação de Recursos Human..	24.712.635
183 - Informação e Inteligência	23.764.224
243 - Assistência à Criança e ao Adol..	14.211.576
334 - Fomento ao Trabalho	191.341

30 – Ranking de Gastos do Estado

Com base no ranking da figura 30, é possível ver que o setor responsável pelo maior gasto durante o período entre janeiro de 2016 até dezembro de 2021 foi o setor Administrativo (quase R\$ 61 Bilhões). Abrindo a visualização para uma categorização de granularidade menor, podemos entender melhor como esse gasto está sendo distribuído dentro do setor:

Subfunção	Elemento	
122 - Administração Geral	31901200 - Vencimentos e Vantagens Fixas - Pessoal Milit	25.220.528,35
	31901100 - Vencimentos e Vantagens Fixas - Pessoal Civil	13.212.980,75
	31901700 - Outras Despesas Variáveis - Pessoal Militar	7.217.846,02
	31911300 - Obrigações Patronais	7.475.131,22
	33903900 - Outros Serviços de Terceiros - Pessoa Jurídica	2.321.649,82
	31901600 - Outras Despesas Variáveis - Pessoal Civil	1.779.366,52
	33903000 - Material de Consumo	543.852,10
	33909200 - Despesas de Exercícios Anteriores	433.976,40
	44905200 - Equipamentos e Material Permanente	63.781,50
	33904900 - Auxílio Transporte	508.604,62
	33913900 - Outros Serviços de Terceiros - Pessoa Jurídica	5.266,19
	33904600 - Auxílio Alimentação	440.588,23
	31909200 - Despesas de Exercícios Anteriores	276.138,11
	33913000 - Material de Consumo	244.592,70
	33900800 - Outros Benefícios Assistenciais do Servidor e	175.764,06
	44905100 - Obras e Instalações	128.184,21
	33904000 - Serviços de Tecnologia da Informação e Comuni	14.217,91
	31900400 - Contratação Por Tempo Determinado	129.961,28
	33901700 - Outras Despesas Variáveis - Pessoal Militar	126.494,67
	33911700 - Outras Despesas Variáveis - Pessoal Militar	121.318,12

31 – Ranking de Gastos por categorização de Elemento

É possível identificar que os gastos com pagamentos (vencimentos para pessoas físicas ou jurídicas, militares e civis) são responsáveis pelos maiores gastos do orçamento, enquanto gastos em categorias como equipamentos, materiais de consumo e segurança, serviços de apoio a inteligência têm dimensão bem menor (na maior parte os gastos não chegam a R\$ 200 Mil).

Com base nesses dados podemos identificar que o investimento em infraestrutura, inteligência, equipamentos e serviços de apoio a qualidade de trabalho dos profissionais de segurança claramente ficam despriorizados, enquanto as folhas salariais e contratos de serviço ocupa a maior porcentagem dos gastos. Esse entendimento encontra-se alinhado tanto com os fatos históricos já referenciados no estudo (como a greve policial de 2017) quanto em reportagens que podem ser encontradas em veículos de imprensa.

7. Apresentação dos Resultados

O cenário de violência policial é um problema que assola a população do estado do Rio de Janeiro a muito tempo, e não há previsão de mudança quanto a isso. Os esforços governamentais, ações policiais, apoio militar e movimentos de apoio à cultura, lazer e educação não são o suficiente para que as ocorrências diminuam, seja por conta de crime organizado, guerra de facções ou pontos isolados.

Dito isso, fica claro pelos constantes relatos midiáticos que a população se sente abandonada pelos seus governantes enquanto que não conseguem realizar ações contundentes. Ao analisar os dados é possível ver que mesmo com o passar dos anos, os avanços tecnológicos e grandes eventos na região, não há recuo na violência sofrida.

A maior parte do orçamento do setor de segurança é gasto com pagamentos e à servidores civis e militares, despesas de gestões anteriores e contratos/licitações, enquanto há pouquíssimo investimento em setores de inteligência, melhorias de equipamentos, instalações e itens de consumo de agentes de segurança. A icônica cena dos policiais recorrendo à mecânicos do subúrbio carioca e tendo que pagar quase que do próprio bolso o montante para o reparo de suas viaturas no filme “Tropa de Elite” retrata bem o momento vivido pela corporação.

Ao mesmo tempo, Diante dos gráficos montados, é possível ver a distribuição dos registros pelas regiões do estado, sua enorme concentração da região da capital, e sua série histórica. Os momentos de queda nos registros, que poderiam indicar uma melhora no cenário, são facilmente identificados por eventos que podem indicar subnotificação nos registros. Há uma clara falta de coordenação nos gastos.

Apesar do estudo ser apenas de um pequeno recorte, uma vez que os sistemas de transparência do governo não permitiram extrações maiores durante o período de montagem, consegue demonstrar que não há cenário de esperança por mudanças, visto que não é possível identificar relação entre os dois expoentes do estudo.

Contudo, foi possível identificar vários fatores que contribuem para o cenário, com indicadores e mapas de distribuição dos dados, análises de gráficos com evolução temporal e descritivos de totais.

O fluxo de trabalho com os modelos escolhidos para as tentativas e previsão final, bem como o recorte dos dados utilizados para análise e predição e os painéis criados para visualização de dados estão disponíveis com link no fim do arquivo.

8. Links

Aqui você deve disponibilizar os links para o vídeo com sua apresentação de 5 minutos e para o repositório contendo os dados utilizados no projeto, scripts criados, etc.

Link para o vídeo: <https://youtu.be/8YdNR94AnOY>

Link para o repositório:
https://github.com/ArthurPSantos/tcc_puc_serie_historica_ocorrencias_policiais

Links para Dashboards montados com base na exploração de dados:

Relação de Delegacias por Região:
<https://public.tableau.com/app/profile/arthur.pereira.dos.santos/viz/QTDEdeDelegaciasporRegio/RelaoRegioporCISP?publish=yes>

Relação de ocorrências por Região e tempo:
<https://public.tableau.com/app/profile/arthur.pereira.dos.santos/viz/QTDEdeDelegaciasporRegio/RelaoOcorrnciasporRegioeMs?publish=yes>

Relação de distribuição de orçamento por Ano e Mês:
<https://public.tableau.com/app/profile/arthur.pereira.dos.santos/viz/QTDEdeDelegaciasporRegio/ValoresInvestidosporAno?publish=yes>

Relação de Ocorrências e valores investidos em segurança por ano e mês:
<https://public.tableau.com/app/profile/arthur.pereira.dos.santos/viz/QTDEdeDelegaciasporRegio/DistribuiodeInvestimentoeOcorrnciasporMs?publish=yes>

Ranking de Gastos por Função e Elemento:
<https://public.tableau.com/app/profile/arthur.pereira.dos.santos/viz/QTDEdeDelegaciasporRegio/DistribuiodeValoresporElemento?publish=yes>

REFERÊNCIAS

1. IT BEST PRACTICES, UNIVERSITY of NEBRASKA–LINCOLN. **Remember the 5 W's**. Disponível em: <<https://its.unl.edu/bestpractices/remember-5-ws>>. Acesso em: 10 de abril de 2022.
2. GOVERNO DO ESTADO DO RIO DE Janeiro. **Portal de Transparência Fiscal**. Disponível em: <<http://www.transparencia.rj.gov.br/transparencia/>>. Acesso em: 10 de abril de 2022.
3. INSTITUTO DE SEGURANÇA PÚBLICA. **Conjunto de Dados**. Disponível em: <<http://www.ispdados.rj.gov.br/EstSeguranca.html>>. Acesso em: 10 de abril de 2022.
4. REDAÇÃO ÉPOCA. **A greve dos policiais militares no Espírito Santo e no Rio de Janeiro**. Disponível em: <<https://epoca.oglobo.globo.com/brasil/noticia/2017/02/greve-dos-policiais-militares-no-espírito-santo-e-no-rio-de-janeiro.html>>. Acesso em: 10 de abril de 2022.
5. GOV.BR. **DECRETO LEGISLATIVO Nº 6 DE 20 DE MARÇO DE 2020**. Disponível em: <<https://legislacao.presidencia.gov.br/atos/?tipo=DLG&numero=6&ano=2020&ato=b1fAzZU5EMZpWT794>>. Acesso em: 10 de Abril de 2022.
6. Carneiro, Julia. **Cinco motivos que levaram o Rio à pior crise de segurança em mais de uma década**. Disponível em: <<https://www.bbc.com/portuguese/brasil-39816208>>. Acesso em: 10 de Abril de 2022.
7. Brito, Ana Carolina; Gomes, Alexandre; Pio, João. **CRIMINALIDADE NA CIDADE DO RIO DE JANEIRO (RJ) As influências das políticas públicas e as relações a curto e longo prazos**. Disponível em: <<https://www.scielo.br/j/rbcsoc/a/P8NPDMrGjMyJDKMRz8WHRvK/?lang=pt&format=pdf>>. Acesso em: 10 de Abril de 2022.
8. Coordenadoria de Estatística do Instituto de Segurança Pública. **SÉRIES HISTÓRICAS ANUAIS DE TAXA DE LETALIDADE VIOLENTA NO ESTADO DO RIO DE JANEIRO**. Disponível em: <<http://www.ispdados.rj.gov.br/Arquivos/SeriesHistoricasLetalidadeViolenta.pdf>>. Acesso em: 10 de Abril de 2022.
9. Almeida, Livia; Coelho, Diogo. **Rio de Janeiro: a Segurança Publica em números. Evolução dos principais indicadores de criminalidade e atividade policial no Estado do Rio de Janeiro – 2003 a 2015**.
10. Vasandani, Jasmine. **A Data Science Workflow Canvas to Kickstart Your Projects**. Disponível em: <<https://towardsdatascience.com/a-data-science-workflow-canvas-to-kickstart-your-projects-db62556be4d0>>. Acesso em: 10 de Abril de 2022.
11. Weisinger, Corey. **Time Series Analysis with Components**. Disponível em: <<https://www.knime.com/blog/time-series-analysis-with-components>>. Acesso em: 10 de Abril de 2022.

12. Ganzaroli, Dennis. **Time Series Analysis with KNIME — an introduction**. Disponível em: < <https://medium.com/mlearning-ai/time-series-analysis-with-knime-an-introduction-7ce01a7ce055>> Acesso em: 10 de Abril de 2022.
13. Brownlee, Jason. **A Gentle Introduction to SARIMA for Time Series Forecasting in Python**. Disponível em: < <https://machinelearningmastery.com/sarima-for-time-series-forecasting-in-python/>>. Acesso em: 10 de Abril de 2022.
14. DATA SCIENCE BLOG. **UMA VISÃO GERAL SOBRE MACHINE LEARNING – REGRESSÃO E CLUSTERING**. Disponível em: < <https://operdata.com.br/blog/uma-visao-geral-sobre-machine-learning-regressao-clustering/>>. Acesso em: 10 de Abril de 2022.
15. Souza, Carlos. **Séries temporais com Machine Learning — Parte 4**. Disponível em: < <https://medium.com/data-hackers/s%C3%A9ries-temporais-com-machine-learning-parte-4-6beb3efb2e99>>. Acesso em: 10 de Abril de 2022.
16. Junior, Airton. **Séries temporais e componentes: aplicando ARIMA para forecast em dados do COVID-19**. Disponível em: < <https://imasters.com.br/data/series-temporais-e-componentes-aplicando-arima-para-forecast-em-dados-do-covid-19>>. Acesso em: 10 de Abril de 2022.

SOBRENOME DO AUTOR, Nome do autor. **Título do livro ou artigo**. Cidade: Editora, ano.