

ANÁLISE DE DADOS: INTELIGÊNCIA APLICADA NA PREVENÇÃO E COMBATE À CRIMINALIDADE NO ESTADO DE GOIÁS.

DATA ANALYSIS: INTELLIGENCE APPLIED IN PREVENTING AND FIGHTING CRIMINALITY IN THE STATE OF GOIÁS.

FÉLIX, André Lucas de Souza¹

SILVA, Vânia Spindola Arantes da²

RESUMO

Com o uso da linguagem de programação Python para análise de dados é possível definir estratégias para coletar, tratar, analisar e visualizar dados que podem auxiliar na redução dos índices de criminalidade em determinadas áreas do Estado de Goiás. Neste contexto, a presente pesquisa busca avaliar dados mantidos pela Secretaria de Segurança Pública do Estado de Goiás e que estão relacionados com diversas ocorrências policiais. Para isso, foi utilizado dados coletados do Mportal dentro de um determinado intervalo de tempo. Esses dados foram analisados mediante o uso de ferramentas específicas para ciência de dados. A metodologia aplicada envolve o uso de conceitos difundidos em ciência de dados para a Análise Exploratória de Dados e para a Análise Explanatória de Dados (Visualização de Dados). O resultado da pesquisa busca apresentar a relevância dessas análises para o aperfeiçoamento da inteligência policial no combate a criminalidade, definindo novas estratégias no âmbito da Policial Militar do Estado de Goiás - PMGO.

Palavras-chaves: Dados, Informação, Conhecimento, Inteligência, Análise de Dados, Gestão da Informação, Gestão do Conhecimento.

ABSTRACT

Using the Python programming language for data analysis, it is possible to define strategies to collect, process, analyze and visualize data that can help reduce crime rates in certain areas of the State of Goiás. In this context, the present research seeks to evaluate data maintained by the Public Security Secretariat of the State of Goiás and which are related to various police incidents. For this, data collected from Mportal within a certain time interval was used. This data was analyzed using specific data science tools. The applied methodology involves the use of widespread concepts in data science for Exploratory Data Analysis and Explanatory Data Analysis (Data Visualization). The research result seeks to present the relevance of these

¹ Aluno do Curso de Formação e Aperfeiçoamento de Praças da Academia de Polícia Militar de Goiás – CAPM, graduado em Gestão Pública, <http://lattes.cnpq.br/2605608003968407>; E-mail: andrelucassfx@gmail.com, Goiânia-GO, janeiro de 2024.

² Professora orientadora: Especialista em Criminologia pela Universidade Federal de Goiás - UFG, vaniaspindola@gmail.com, Goiânia-GO, janeiro de 2024.

analyzes for improving police intelligence in the fight against crime, defining new strategies within the scope of the Military Police of the State of Goiás - PMGO.

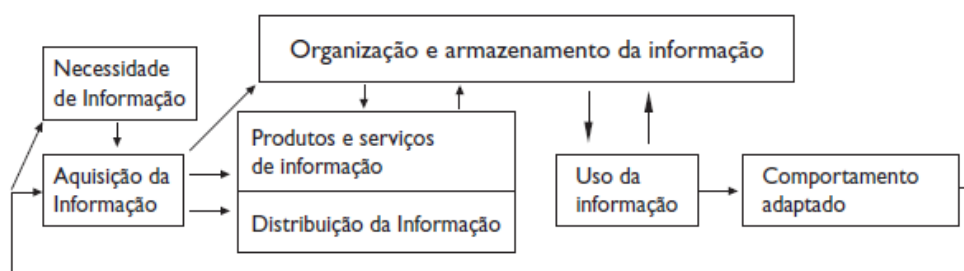
Keywords: Data, Information, Knowledge, Intelligence, Data Analysis, Information Management, Knowledge Management.

1. INTRODUÇÃO

Para Tarapanoff (2006), o conceito de sociedade do conhecimento está associado com as tarefas de administração organizacional. O principal ativo nesta abordagem administrativa é a informação, que gera inovação e conhecimento. Neste contexto, informação e conhecimento têm sido tratados extensamente na Teoria das Organizações e são conceitos que estão intimamente relacionados com a gestão da informação.

A gestão da informação pode ser analisada por diversos ângulos e está correlacionada com a gestão do conhecimento. Embora não haja um modelo padronizado para se trabalhar a gestão da informação, pode-se determinar métodos que auxiliam significativamente no processo de análise, tornando este tipo de gestão essencial na tomada de decisões estratégicas, táticas ou operacionais. De outra forma, pode-se dizer que a gestão da informação define como aplicar princípios administrativos na aquisição, organização, controle, disseminação e operacionalização efetiva do conhecimento produzido no âmbito corporativo; sendo uma ferramenta cíclica e que está cada vez mais em uso nas organizações modernas.

Figura 01: Ciclo da Gestão da Informação



Fonte: Tarapanoff (2006).

Além do cenário organizacional moderno, presencia-se atualmente o que se ocasionou chamar de quarta revolução digital (WIKIPÉDIA, 2023), em que a tecnologia tem papel crucial na vida das pessoas e da sociedade. Através da internet, uma quantidade gigantesca de dados sobre os mais variados assuntos está disponível a qualquer instante. A capacidade de adquirir, armazenar e manipular dados é facilitada por diversas fontes disponíveis nas plataformas de dados abertos. Dentre essas plataformas, encontram-se as disponibilizadas por órgãos e instituições governamentais, a exemplo das bases disponibilizadas pelo Governo Federal e acessível no portal de dados abertos (www.dados.gov.br).

O uso de dados abertos, particularmente os produzidos por gestores públicos, é a porta de entrada para fornecer canais de comunicação com a sociedade, servindo como fonte para elaboração eficiente de conhecimento e auxiliando na transparência pública. A análise de dados abertos também permite maior controle social sob a atuação de gestores públicos na condução de projetos e programas governamentais. Assim, o conjunto de dados abertos é a base inicial de

qualquer gestão da informação, entendê-los é o primeiro passo para iniciar uma análise exploratória e criar informação de qualidade (análise explanatória). No Brasil, a Lei de Acesso da Informação (Lei nº 12.527, de 18 de novembro de 2011) permite que qualquer cidadão possa solicitar e explorar dados mantidos e controlados por entidades governamentais; sendo fonte fidedigna de dados sobre a atuação em diversos setores sociais, tais como saúde, educação, infraestrutura urbana e segurança pública.

No que se refere a dados abertos produzido e controlados pelo Estado de Goiás, especificamente os dados que estão correlacionados com a segurança pública, é possível a qualquer cidadão acessá-los mediante solicitação direta e formal. O foco deste trabalho de pesquisa é justamente avaliar esses dados abertos mediante técnicas e ferramentas apropriadas para deduzir e produzir conhecimento relacionado com a atividade policial. A análise de dados utilizará métodos estatísticos por meio de linguagem de programação para obter conhecimento a partir de base de dados públicas. A principal fonte de dados está disponível no Mportal (www.ssp.go.gov.br). Os dados avaliados estão disponíveis por meio de acesso ao Mportal e, basicamente, são dados dos Registro de Atendimento Integrado – RAI, feito por policiais e outros agentes de segurança do Estado de Goiás.

Portanto, o objetivo principal desta pesquisa é avaliar os dados coletados e verificar a possibilidade de aplicar a gestão da informação e do conhecimento no processo de análise exploratória de dados. Com o uso da linguagem Python e de suas bibliotecas voltadas para atividades de ciência de dados; também, busca-se demonstrar que a análise exploratória de dados pode ser uma fonte de conhecimento de fácil utilização para visualizar e definir estratégias voltadas para o combate a criminalidade local no Estado de Goiás.

2. REVISÃO DE LITERATURA

Mckinney (2019) observa a importância de fazer a análise de dados estruturados no processo de gestão da informação, pois estes envolvem muitos formatos diferentes e genéricos de dados, como por exemplo dados tabulares ou do tipo planilha, em que cada coluna pode ser de um tipo diferente (string, numérico ou data). Essa avaliação é importante, pois inclui a maior parte dos dados comumente armazenados em bancos de dados relacionais ou em arquivos de texto delimitados com tabulação ou vírgula e que são amplamente utilizados em sistemas e aplicativos institucionais.

Quando estamos direcionados a analisar dados, é importante que se defina uma estrutura de análise voltada para o uso de ferramentas efetivas. Nos últimos anos, o suporte melhorado para a linguagem Python, incluindo suas bibliotecas (como o pandas, seaborn, matplotlib e o scikit-learn), transformou essa linguagem de programação em uma das mais efetivas para tarefas de análise exploratória e explanatória de dados.

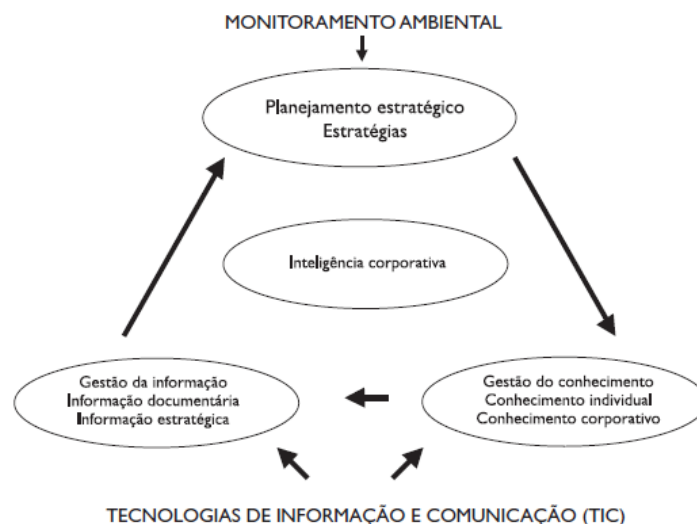
Mckinney (2019) define Python com a linguagem de programação com forte apelo para a ciência de dados. Desde o seu surgimento em 1991, Python se tornou uma das linguagens de programação interpretadas mais populares do mundo. A partir de 2015, Python passou a ser uma linguagem de computação científica e inovadora, sendo uma das linguagens mais importantes em ciência de dados, aprendizado de máquina (*machine learning*) e desenvolvimento de softwares em geral, tanto em ambiente acadêmico como no mercado de software.

No que se refere a métodos de análise de dados, Tarapanoff (2006) faz uma ponderação bastante importante, ao dizer que qualquer organização, seja pública ou privada, utiliza a informação em três fases estratégicas. A primeira delas envolve o levantamento, coleta, tratamento e a análise de dados, de modo a dar significado ao que está acontecendo diante de suas rotinas internas e externas, gerando informação; pode-se definir esse primeiro passo como a gestão da informação. Depois, a organização cria novos conhecimentos, combinando-os com a experiência do seu corpo técnico para fazer a gestão do conhecimento e apresentar os resultados esperados sobre a gestão da informação. Finalmente, a organização desenvolve o terceira fase, que envolve a análise da informação de modo a empreender ações apropriadas com o seu uso, aqui temos a inteligência estratégica ou corporativa.

Assim, o processo de gestão do conhecimento é uma atividade independente, mas que está fortemente ligado ao processo de gestão da informação e ao trabalho e análise da informação. A inteligência é estratégica e pode ser considerada a síntese de todo o processo de análise da informação e da gestão do conhecimento. Em uma organização, a informação, o conhecimento e a inteligência vão subsidiar a tomada de decisão e o planejamento estratégico. Por meio das estratégias avaliadas, as organizações aproveitam melhor as oportunidades e implementam ações que minimizam o impacto das ameaças sobre seus pontos fracos.

Tarapanoff (2006) define outro ponto importante: não tem como fazer análise de inteligência pelo acesso passivo da informação, ou seja, somente organizando os dados brutos. A inteligência deve ser criada e, neste momento, o processo da gestão da informação e do conhecimento vão definir um sistema útil para a organização. Este sistema deverá estar integrado em sua cultura e em seus cenários voltados para o futuro e para a tomada de decisão; estes elementos se retroalimentam formando a inteligência corporativa.

Figura 02: Integração da gestão da informação e gestão do conhecimento ao planejamento estratégico.



Fonte: Tarapanoff (2006).

Knaflíc (2017) faz uma consolidação dos principais pontos a serem observados na Análise Explanatória de Dados, ou seja, basicamente aborda em sua obra, conceitos importantes

para a visualização dos resultados da exploração de dados, abordando conceitos importantes para criar gráficos, tabelas de forma eficiente e com foco específico para a gestão da informação e do conhecimento.

Fávero (2017) apresenta as principais técnicas estatísticas e de modelagem multivariada e que podem ser adaptadas para utilização em ferramenta de análise de dados, como as que utilizam linguagem e bibliotecas de programação Python voltadas para ciência de dados, dando a devida ênfase na importância da análise de dados em ambientes acadêmicos e organizacionais, considerando a importância da modelagem voltada à tomada de decisão.

De acordo com Gibbs (2008), constitui uma boa prática fazer a análise dos dados simultaneamente com a coleta, pois isso permite levantar questionamentos sobre a pesquisa realizada. Entende-se que a coleta dos dados servirá de base para a análise exploratória, uma vez que é a primeira tarefa e deve ser baseada em questionamentos sobre a importância dos dados para a gestão da informação.

Oliveira (2018) estabelece um modelo de workflow (fluxo de trabalho) para iniciar o processo de gestão da informação. Sendo extremamente relevante a abordagem que é apresentada para se fazer a Análise exploratória dos dados. Definindo um fluxo de trabalho simples e ao mesmo tempo extremamente eficiente para buscar, limpar, transformar, modelar e analisar dados no processo de gestão da informação e do conhecimento.

Portanto, a literatura analisada busca definir conceitos que se aplicam na gestão da informação e do conhecimento. Os processos de análise exploratória de dados estão embasados em conceitos observados em livros técnicos sobre a análise de dados e sobre a utilização de ferramentas voltadas para essa finalidade.

3. METODOLOGIA

Fávero (2017) estabelece um modelo de workflow (fluxo de trabalho) para iniciar o processo de gestão da informação. Entretanto, é válido ressaltar que não existe apenas uma forma de estruturar e aplicar os conhecimentos de ciência de dados na gestão da informação. A forma de aplicação varia bastante, de acordo com a necessidade do projeto ou do objetivo que se busca alcançar.

Nesta pesquisa, optou-se por usar um modelo de workflow dividido em duas etapas, a primeira delas envolve a Análise Exploratória dos Dados, já a segunda etapa envolve especificamente a visualização e apresentação das análises. Esta segunda Etapa é definida como Análise Explanatória de Dados. Apesar da maioria das bibliografias defenderem que a parte de visualização de dados seria apenas uma fase da Análise Exploratória, há diferenças entre esses conceitos.

Conforme defendido por Knaflitz (2017), há uma distinção importante a fazer entre análise exploratória e explanatória. A análise exploratória seria aquilo que fazemos para compreender os dados e descobrir o que pode ser digno de apresentar para os tomadores de decisão ou outras partes interessadas. Já na análise explanatória, estamos comunicando nossa análise para um determinado público. Na explanatória, significa que temos algo específico para explicar, uma ‘história para contar’ e que provavelmente apresentaremos por meio de visualização da dados (gráficos, tabelas, relatórios etc.).

3.1. ETAPA EXPLORATÓRIA DE DADOS

A Etapa Exploratória de Dados pode ser dividida em quatro fases: Importação, Limpeza, Exploração e Comunicação.

3.1.1. Importação (Import)

Basicamente envolve a coleta dos dados de alguma fonte e consolida essa base para importação. Nesta fase, define-se como os dados serão analisados e quais ferramentas serão utilizadas para a análise. O formato de dados também é definido nesta fase, aqui o arquivo de dados pode ser importado em diversos formatos (json, csv, xls etc.).

3.1.2. Limpeza (Tidy)

A limpeza é a fase em que há um tratamento mínimo na base de dados de forma a torná-la mais fácil de manipular. Nesta fase são excluídos dados nulos, feitas correções em dados, retirados os dados duplicados, e atributos (características) pouco relevantes.

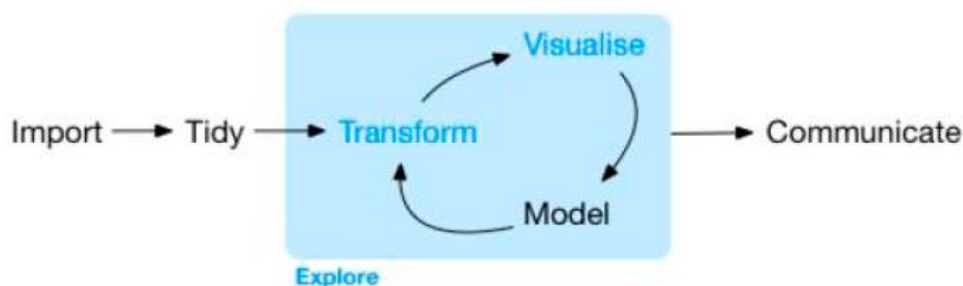
3.1.3. Exploração (Explore)

A exploração é a fase mais complexa da análise de dados, pois pode ser necessário transformar alguns dados em tipos específicos, fazer agrupamentos e definir modelos de dados para gerar algum tipo de visualização inicial, de forma a facilitar a gestão da informação e gerar conhecimento sobre os dados que estão sendo analisados. Em torno de 70% da análise de exploratória dos dados é feita nesta fase.

3.1.4. Comunicação (Communicate)

Esta fase é a conclusão da Análise Exploratória, aqui tem que haver algum tipo de relatório que consolide as informações geradas. Deve haver a explicação de como os dados foram analisados e quais foram os resultados alcançados. É uma fase prévia a Etapa seguinte que envolve a visualização dos dados para apresentação final.

Figura 03: Modelo de Workflow para gestão da informação.



Fonte: Oliveira (2018).

3.2. ETAPA EXPLANATÓRIA DE DADOS

Também conhecida como visualização de dados, é a Etapa em que são gerados tabelas e gráficos para diferentes contextos de análise. Nesta Etapa são exploradas as bibliotecas da Linguagem Python para visualização eficiente de informações. Os gráficos que compõem essa Etapa podem ser de variados tipos tais como: tabelas, histogramas, gráficos em colunas e de barras (verticais e horizontais), gráficos de dispersão etc.

Aqui diversos conceitos aplicados a estatística descritiva são utilizados pelas bibliotecas Python, de forma a facilitar a gestão da informação e produzir conteúdo condizente com a Etapa anterior (Análise Exploratória dos Dados). De acordo com Knaflitz (2017), as saídas geradas nesta Etapa podem ser tabelas ou gráficos, que interagem com nosso sistema visual. Quanto mais bem elaborado a saída de informação, mais eficiente será sua visualização. Isso significa que um gráfico bem projetado normalmente comunicará a informação mais rápido que uma tabela bem projetada. Os tipos de gráficos que são frequentemente usados são: pontos, linhas, barras e área.

4. RESULTADOS E DISCUSSÕES

4.1. DA ANÁLISE EXPLORATÓRIA DE DADOS

4.1.1. Fases: Importação, Limpeza e Exploração

Inicialmente, houve uma busca por dados mantidos no Mportal, definindo previamente um escopo de pesquisa. Este escopo foi por ocorrências sobre o uso de armas de fogo, os dados coletados compõem o Registro de Atendimento Integrado – RAI. Basicamente, são dados relacionados a ocorrências policiais registradas em um determinado período. Por meio da utilização de ‘Consulta Avançada de Ocorrências’, buscou-se dados relacionados as ocorrências no Município de Goiânia entre o período de **01/05/2023 a 01/11/2023**; a escolha do município foi aleatória.

Ao todo, 774 (setecentos e setenta e quatro) registros foram encontrados neste período. A base de dados foi nomeada como “mportal_ocorrencia_export_geral.csv”. Esta primeira tarefa envolveu a coleta e a importação dos dados em um formato de dados abertos (formato ‘csv’).

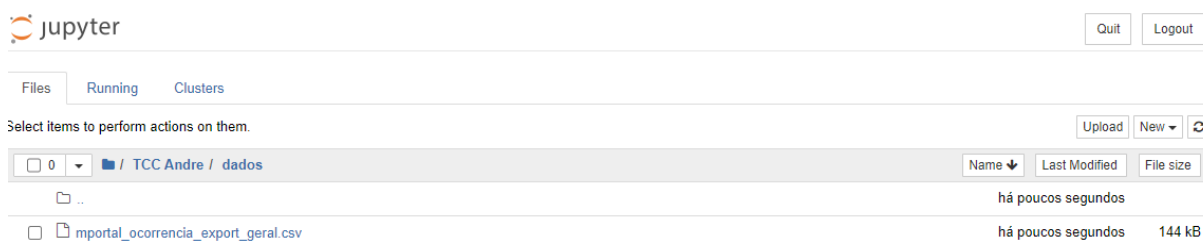
Figura 03: Coleta no Mportal – Filtro através de Consulta Avançada de Ocorrências.

Fonte: Mportal (2023).

Como ferramenta para escrever o código que analisará os dados, optou-se pelo Jupyter Notebook ³, ferramenta que permite usar diversas bibliotecas da linguagem Python ⁴ integradas com outras ferramentas tais como NumPy, Pandas, Seaborn para visualização de dados, Matplotlib, entre outras. Dentre as ferramentas citadas, o Pandas ⁵ é a principal delas, trata-se de uma biblioteca para análise de dados em Python, que possui diversos recursos para se trabalhar com dados; sendo possível tratar, manipular, analisar os dados de forma bastante interativa e rápida.

Primeiramente, houve a importação da base de dados do Mportal para o Jupyter Colaby a importação do arquivo “mportal_ocorrencia_export_geral.csv”; para isso, basta fazer o upload do arquivo coletado no Mportal para o Jupyter Notebook, conforme apresentado na Figura 04.

Figura 04: Carregamento de arquivo no Jupyter Notebook.



³ Jupyter Notebook, acesso em: <https://jupyter.org/>.

⁴ Linguagem Python: linguagem de programação versátil e poderosa que oferece uma ampla gama de recursos e ferramentas para análise de dados. Seu uso é cada vez mais comum em várias áreas, desde finanças, vendas, saúde, marketing e principalmente na pesquisa científica.

⁵ Documentação sobre os principais comandos na biblioteca do Pandas, acesso disponível em: <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.plot.html>.

Fonte: Autor (2023).

Após a importação do arquivo ‘csv’ para o Jupyter, foi criado um caderno de anotação (notebook). Isto permitirá registrar as anotações sobre as análises de dados ao mesmo tempo em que é possível aplicar os comandos básicos no arquivo importado. Essas informações posteriormente poderão ser consolidadas em relatório de análise, sendo a saída para a visualização dos dados (fase communicate). A Figura 05 mostra como criar um notebook, clique em Python 3 e será criado um caderno de anotação para utilizar comandos e escrever seus relatórios sobre a análise de dados. Depois, basta renomear o notebook criado; para esta pesquisa, o nome do notebook é ‘Análise de Dados – Exploracao’.

Figura 05: Criar um Notebook no Jupyter.



Fonte: Autor (2023).

No Python, existem diversas bibliotecas para a leitura de arquivos no formato csv, tratamento de dados, visualização de dados etc. Portanto, devemos iniciar a análise de dados importando as bibliotecas necessárias para este fim. No caso, fizemos a importação de três bibliotecas: Pandas, Seaborn⁶ e Matplotlib⁷. As bibliotecas Seaborn e Matplotlib serão bastante usadas na Etapa de análise explanatória dos dados, mas já foram instaladas para facilitar as análises na fase de comunicação de dados (communicate) e permitir a exploração de dados por meio de visualizações iniciais.

⁶ Seaborn é uma biblioteca de visualização de dados Python baseada em matplotlib. Ele fornece uma interface de alto nível para desenhar gráficos estatísticos atraentes e informativos. Acesso a documentação: seaborn.pydata.org.

⁷ Matplotlib é uma biblioteca abrangente para criar visualizações estáticas, animadas e interativas em Python. Matplotlib torna as coisas fáceis e as difíceis possíveis. Acesso a documentação: matplotlib.org.

Tabela 01: Comandos para importação de Bibliotecas Python.

```
import pandas as pd
```

```
import seaborn as sns
```

```
import matplotlib.pyplot as plt
```

Fonte: Autor (2023).

Feito isso, pode-se atribuir outro nome para a base de dados, no caso esta base recebeu um nome de ‘base_original’ e abrir o arquivo csv. Os comandos aplicados permitem visualizar a base de dados como um dataframe (tabela), no caso são mostrados apenas cinco registros do arquivo “mportal_ocorrencia_export_geral.csv”, conforme apresentado na Tabela 02.

Tabela 02: Dataframe do Arquivo csv.

```
base_original = pd.read_csv('dados/mportal_ocorrencia_export_geral.csv', sep=';')
```

```
base_original.head(5)
```

	NÚMERO	DATA FATO	HORA	DIA	CATEGORIA	NATUREZAS	ÍNDICE	REGIAO	BAIRRO	LOGRADOURO
0	32660415	01/11/2023	11:33	quarta-feira	posse ou porte de arma uso restrito	art. 16 caput do estatuto do desarmamento poss...	Ocorrências Criminais Graves	CENTRAL	SETOR SOL NASCENTE	AVENIDA FUSUAMA
1	32645939	31/10/2023	12:59	terça-feira	roubo uso arma de fogo	art. 157, § 2º a, inc. i do cpb roubo com empr...	Ocorrências Criminais Graves	VALE MEIA PONTE	SETOR MARECHAL RONDON (FAMA)	RUA 3
2	32610829	28/10/2023	23:01	sábado	homicídio simples	homicídio -> art. 121 caput do cpb homicídio s...	Ocorrências Criminais Graves	MENDANHA	SETOR CÂNDIDA DE MORAIS	RUA CM10
3	32610829	28/10/2023	23:01	sábado	ações policiais	averiguação em chamada para atendimento	Ocorrências Criminais Graves	MENDANHA	SETOR CÂNDIDA DE MORAIS	RUA CM10
4	32598023	28/10/2023	04:53	sábado	homicídio simples	homicídio -> art. 121 caput do cpb homicídio s...	Ocorrências Criminais Graves	MENDANHA	SETOR CÂNDIDA DE MORAIS	RUA CM9

Fonte: Autor (2023).

Nesta primeira análise, os dados coletados do Mportal já apresentam algumas características importantes. Ao todo foram coletados 773 registros (linhas), sendo que a maioria dos dados são categóricos e textuais (object). O dataframe possui 10 (dez) variáveis e a base de dados não possui registros nulos (non-null).

Tabela 03: Tipos de dados.

```
base_original.shape[0]
```

```
773
```

```
base_original.columns
```

```
Index(['NÚMERO', 'DATA FATO', 'HORA', 'DIA', 'CATEGORIA', 'NATUREZAS',  
      'ÍNDICE', 'REGIAO', 'BAIRRO', 'LOGRADOURO'],  
      dtype='object')
```

```
base_original.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 773 entries, 0 to 772  
Data columns (total 10 columns):  
#   Column      Non-Null Count  Dtype  
---  ---  
0   NÚMERO      773 non-null    int64  
1   DATA FATO  773 non-null    object  
2   HORA        773 non-null    object  
3   DIA         773 non-null    object  
4   CATEGORIA   773 non-null    object  
5   NATUREZAS   773 non-null    object  
6   ÍNDICE      773 non-null    object  
7   REGIAO      773 non-null    object  
8   BAIRRO      773 non-null    object  
9   LOGRADOURO  771 non-null    object  
dtypes: int64(1), object(9)  
memory usage: 60.5+ KB
```

Fonte: Autor (2023).

A Tabela 03 confirma que apenas a primeira coluna é do tipo numérica (int64), sendo as demais colunas classificadas como objeto (strings textuais). Ressalte-se que seguindo a fase

de exploração (transform - Figura 03), algumas variáveis foram tratadas e houve a transformação delas antes da importação da base para a ferramenta Jupyter. No caso, as variáveis ‘CATEGORIA’, ‘REGIAO’ e ‘ÍNDICE’ foram incluídas na base originalmente coletada.

Isso é relevante, Gibbs (2008), inclusive, relata que a análise em dados qualitativos (categóricos) envolve um processo de transformação, ou seja, antes de fazer uma análise clara, compreensível, criteriosa, confiável e até original para gerar ideias analíticas (Gibbs, 2008, p. 16 apud Miles e Huberman, 1994; Maykut e Morehouse, 2001; Ritchie e Lewis, 2003) deve haver a transformação e o tratamento da base de dados. Assim, a seleção e a busca em todos os dados deve ser feita com base em boas evidências de sustentação. Este é o grande desafio e requer organização e uma abordagem estruturada dos dados. Nesta concepção, pode-se entender que não só se deve excluir variáveis (colunas) inservíveis, mas incluir outras variáveis condizentes com as análises que se deseja executar na fonte original.

Gibbs (2008) ainda pondera que a análise qualitativa geralmente busca melhorar os dados e aumentar seu volume, sua densidade e sua complexidade. Na base original coletada do Mportal, a princípio, havia aproximadamente 350 (trezentos e cinquenta) registros, mas ao se observar os registros de ocorrência, verificou-se que estes estavam relacionados a mais de um delito ou ocorrência criminosa; por isso foi necessário transformar e tratá-los. Este tratamento foi complementado com a categorização dos dados qualitativos, que permitiu atribuir novas variáveis para organizar os dados em subgrupos.

Tabela 04: Amostra da Base de Dados tratada.

NÚMERO	DATA FATO	HORA	DIA	CATEGORIA	NATUREZAS	ÍNDICE	REGIAO	BAIRRO	LOGRADOURO
32660415	01/11/2023	11:33	quarta-feira	posse ou porte de arma uso restrito	art. 16 caput do estatuto do desarmamento poss...	Ocorrências Criminais Graves	CENTRAL	SETOR SOL NASCENTE	AVENIDA FUSUAMA
32645939	31/10/2023	12:59	terça-feira	roubo uso arma de fogo	art. 157, § 2º a, inc. i do cpb roubo com empr...	Ocorrências Criminais Graves	VALE MEIA PONTE	SETOR MARECHAL RONDON (FAMA)	RUA 3
32610829	28/10/2023	23:01	sábado	homicídio simples	homicídio -> art. 121 caput do cpb homicídio s...	Ocorrências Criminais Graves	MENDANHA	SETOR CÂNDIDA DE MORAIS	RUA CM10
32610829	28/10/2023	23:01	sábado	ações policiais	averiguação em chamada para atendimento	Ocorrências Criminais Graves	MENDANHA	SETOR CÂNDIDA DE MORAIS	RUA CM10
32598023	28/10/2023	04:53	sábado	homicídio simples	homicídio -> art. 121 caput do cpb homicídio s...	Ocorrências Criminais Graves	MENDANHA	SETOR CÂNDIDA DE MORAIS	RUA CM9

Fonte: Autor (2023).

Observa-se que não houve a necessidade de deletar variáveis (colunas) da base de dados original. Pelo contrário, houve a necessidade de transformação dos dados nesta primeira etapa, pois a base original carecia de categorização e classificação dos registros apresentados. O crucial nesta análise de dados será extrair informação útil dos dados textuais (categóricos) que estão dispostos principalmente nas variáveis ‘NATUREZAS’ e ‘BAIRRO’, tarefa para a exploração de dados. Na exporação de dados é importante categorizar e organizar melhor os dados para gerar informação de qualidade. Dessa forma, avaliou-se primeiramente a variável ‘BAIRRO’.

Tabela 05: Amostra da Base de Dados tratada.

NÚMERO	DATA FATO	HORA	DIA	CATEGORIA	NATUREZAS	ÍNDICE	REGIÃO	BAIRRO	LOGRADOURO
32660415	01/11/2023	11:33	quarta-feira	posse ou porte de arma uso restrito	art. 16 caput do estatuto do desarmamento poss...	Ocorrências Criminais Graves	CENTRAL	SETOR SOL NASCENTE	AVENIDA FUSUAMA
32645939	31/10/2023	12:59	terça-feira	roubo uso arma de fogo	art. 157, § 2º a, inc. I do cpb roubo com empr...	Ocorrências Criminais Graves	VALE MEIA PONTE	SETOR MARECHAL RONDON (FAMA)	RUA 3
32610829	28/10/2023	23:01	sábado	homicídio simples	homicídio -> art. 121 caput do cpb homicídio s...	Ocorrências Criminais Graves	MENDANHA	SETOR CÂNDIDA DE MORAIS	RUA CM10
32610829	28/10/2023	23:01	sábado	ações policiais	averiguação em chamada para atendimento	Ocorrências Criminais Graves	MENDANHA	SETOR CÂNDIDA DE MORAIS	RUA CM10
32598023	28/10/2023	04:53	sábado	homicídio simples	homicídio -> art. 121 caput do cpb homicídio s...	Ocorrências Criminais Graves	MENDANHA	SETOR CÂNDIDA DE MORAIS	RUA CM9

Fonte: Autor (2023).

Nesta análise, verificou-se que 207 (duzentos e sete) bairros do município de Goiânia tiveram algum tipo de ocorrência. Nota-se também que pelo menos 25 (vinte e cinco) desses bairros tiveram mais de 7 (sete) ocorrência no período de análise. A informação está disposta na Tabela 06. Nesta fase exploratória, também iniciou-se o modelo de dados (model) e a transformação dos dados (transform), o agrupamento dos dados também foi feito nesta fase.

Tabela 06: Bairros com maiores ocorrências registradas.

```
ocorrencias_bairros = base_original['BAIRRO'].value_counts()
ocorrencias_bairros.shape
```

(207,)

```
ocorrencias_bairros.head(25)
```

PARQUE INDUSTRIAL JOÃO BRÁZ	15
BOA VISTA	13
JARDIM GOIÁS	13
CONJUNTO VERA CRUZ	12
SETOR SANTOS DUMONT	12
JARDIM NOVA ESPERANÇA	11
SETOR MARISTA	11
SETOR NORTE FERROVIÁRIO	11
SETOR COIMBRA	11
SETOR BUENO	10
SETOR SÃO JOSÉ	10
SETOR CAMPINAS	10
RESIDENCIAL GOIÂNIA VIVA	9
SETOR GENTIL MEIRELLES	9
PARQUE OESTE INDUSTRIAL	9
RESIDENCIAL ORLANDO MORAIS	9
SETOR SUL	9
CIDADE JARDIM	9
SETOR URIAS MAGALHÃES	9
JARDIM AMÉRICA	8
VILA PEDROSO	8
CAPUAVA	8
RESIDENCIAL BUENA VISTA II	8
GARAVELO B	7
SÃO FRANCISCO	7

Name: BAIRRO, dtype: int64

Fonte: Autor (2023).

Apesar de ser uma informação relevante, ela pode ser melhor explorada ao se estabelecer uma outra variável na base de dados original. No caso, buscamos atrelar as ocorrências registradas por região do município de Goiânia, de forma a agrupar melhor as informações necessárias e consolidar o resultado pela variável 'REGIÃO', o que permite observar melhor a frequência de ocorrências no período previamente determinado. A Tabela 07 consolida esta análise.

Tabela 07: Região com maiores ocorrências registradas.

```

ocorrencias_regiao = base_original['REGIAO'].value_counts().to_frame().reset_index()
ocorrencias_regiao.columns = ['Setor', 'Total']
ocorrencias_regiao.head(11)

```

	Setor	Total
0	OESTE	146
1	SUDOESTE	101
2	CENTRAL	94
3	MENDANHA	78
4	SUL	72
5	NOROESTE	67
6	LESTE	67
7	NORTE	62
8	VALE MEIA PONTE	48
9	CENTRO-OESTE	37
10	APARECIDA DE GOIÂNIA	1

Fonte: Autor (2023).

Pode-se concluir que a região Oeste do Município concentra o maior número de ocorrência e a região Centro-Oeste, o menor número de ocorrências. Também é possível deduzir que houve uma ocorrência registrada fora do Município de Goiânia, no caso em Aparecida de Goiânia.

Outro ponto importante é que pode-se criar subgrupos e categorizar as ocorrências. Para isso definimos uma outra variável chamada 'CATEGORIA', através da análise feita a partir dessa variável, podemos definir o agrupamento das principais ocorrências. Na Tabela 08 é possível verificar todas as ocorrências agrupadas em categorias, observa-se que as principais ocorrências estão associadas a porte e posse irregular de armas de fogo e também as diversas ações policiais, tais como abordagem a pessoas, veículos, atendimento a chamados, averiguações gerais, cumprimento de diligências entre outras verificações.

Tabela 08: Região com maiores ocorrências registradas.

```
serie_temporais_numeros = base_original['CATEGORIA'].value_counts().to_frame().reset_index()
serie_temporais_numeros.columns = ['Ocorrências', 'Total']
print(serie_temporais_numeros)
```

	Ocorrências	Total
0	ações policiais	121
1	porte arma de fogo	107
2	posse irregular de arma	77
3	posse ou porte de arma uso restrito	58
4	tráfico de drogas	43
5	abordagem	42
6	homicídio simples	41
7	resgate bombeiros	37
8	exclusão de ilicitude	33
9	furto	32
10	ameaça	27
11	roubo	18
12	disparo de arma de fogo	17
13	recuperação de veículos	11
14	perda ou extravio de arma de fogo	10
15	receptação	8
16	outros fatos atípicos	8
17	lesão corporal contra mulher	7
18	foragido da justiça	7
19	lesão corporal	7
20	roubo uso arma de fogo	7
21	crime de trânsito	7
22	injúria	6
23	furto qualificado	4
24	feminicídio	4
25	dano	4
26	associação criminosa	3
27	homicídio qualificado	3
28	desobediência	3
29	desacato	3
30	falsificação documento público	2
31	consumo de drogas	2
32	comércio de arma de fogo	2
33	estelionato	2
34	estupro vulnerável	2
35	descumprir medida protetiva	2
36	perturbação sossego	1
37	extorsão	1
38	contrabando	1
39	corrupção ativa	1
40	sequestro e cárcere privado	1
41	lesão corporal	1

Fonte: Autor (2023).

Uma outra forma de verificar efetivamente os dados coletados é categorizá-los em uma variável que faça o agrupamento dos crimes e delitos em uma variável mais compacta, classificando as ocorrências por grau de gravidade. Para isso, foi definido uma outra variável chamada de ‘ÍNDICE’. A Tabela 09 mostra as ocorrências agrupadas, sendo que houve o registro de 396 (trezentos e noventa e seis) crimes considerados graves, 137 (centos e trinta e sete) crimes de menor potencial ofensivo, dentre outros registros.

Tabela 09: Agrupamento de crimes.

```
serie_ocorrencia_indice = base_original['ÍNDICE'].value_counts().to_frame().reset_index()
serie_ocorrencia_indice.columns = ['Subgrupo', 'Total']
print(serie_ocorrencia_indice)
```

	Subgrupo	Total
0	Ocorrências Criminais Graves	396
1	Atendimento de Rotina PMGO	181
2	Ocorrências de Menor Potencial Ofensivo	137
3	Resgate e Socorro	37
4	Ocorrências Relevantes	22

Fonte: Autor (2023).

4.2. DA ANÁLISE EXPLANATÓRIA DE DADOS

A análise explanatória pode ser definida como a visualização eficiente dos dados por meio de gráficos e tabelas. Nesta Etapa, tem-se toda a consolidação da análise exploratória (Etapa anterior) de forma efetiva e como gestão da informação. Essa visualização de dados pode ser apresentada e utilizada como conhecimento para subsidiar a tomada de decisão e definir estratégias táticas e operacionais. A conclusão de todo um processo de análise explanatória faz parte da Inteligência Corporativa (Figura 02), conforme definido por Tarapanoff (2006) como “sistema integrado para a tomada de decisão”.

Diversas são as formas de visualizar as informações geradas na análise exploratória, nesta Etapa não cabe mais fazer análise de dados. Aqui, deve-se estar completamente focado na gestão da informação e do conhecimento. Cada informação criada deverá gerar conhecimento, independente do seu tamanho; ou seja, não importa se sua análise exploratória gerou apenas um ou dois números como informação, um simples texto associado a ela pode ser uma ótima maneira de comunicá-la. Conforme se observa na Figura 05:

Figura 05: Região com maior Índice de Ocorrências.

19%
dos **crimes** registrado entre maio e
novembro de 2023, em **Goiânia**,
foram registrados na região **Oeste**.

Fonte: Autor (2023).

Essa mesma informação consta na Tabela 07, mas de forma bem mais discreta. Observe como esta visualização da informação, com o uso de uma frase explicativa, apresenta uma informação muito mais relevante para o contexto comunicativo.

Da mesma forma, a informação apresentada na Tabela 09, poderia ser visualmente mais relevante se fosse transmitida por meio de uma Tabela apresentada como um ‘Mapa de Calor’ (heatmap). Ao analisá-la, de imediato, já seria possível inferir que entre as ocorrências registradas nos finais de semana, incluindo as sextas-feiras, estes dias apresentam as ocorrências criminais mais graves (54%). Visualmente esta forma de apresentação traz a informação maior mais rápida para quem irá fazer a análise.

Tabela 10: Agrupamento de crimes.

Mapa de Calor (Heatmap)

Baixo - Alto

OCORRÊNCIAS	Sexta	Sábado	Domigo
Atendimento de Rotina PMGO	23%	17%	24%
Ocorrências Criminais Graves	54%	53%	46%
Ocorrências de Menor Potencial Ofensivo	11%	25%	21%
Ocorrências Relevantes	4%	2%	3%
Resgate e Socorro	8%	4%	6%

Fonte: Autor (2023).

Segundo Knafllic (2017), as tabelas interagem com nosso sistema verbal, já os gráficos interagem com nosso sistema visual, assim trazem a informação mais rapidamente, o que significa que um gráfico bem projetado normalmente comunicará a informação mais rápido que uma tabela bem projetada. Pode-se fazer uma verificação das principais ocorrências registradas em determinados meses utilizando gráfico. Como exemplo, vamos visualizar as principais ocorrências dentro do intervalo (de maio a outubro). Observa-se que neste período, houve 166 ocorrências em maio, sendo este o mês com maior numero de ocorrências registradas no Município.

Gráfico 01: Visualização de Ocorrências por Mês.



Fonte: Autor (2023).

A análise explanatória de dados envolve diverso tipos de representações visuais que podem ser exploradas de diversas maneiras. Há casos de uso para outros tipos de apresentações e pode-se utilizar as bibliotecas Searbon e Matplotlib para abordá-las. Em muitos casos, não há uma única representação visual correta; pois frequentemente existem diferentes tipos de apresentações que podem atender a necessidade específica. O importante é ter essa necessidade

claramente enunciada: uma representação visual que torne a informação mais eficiente no processo de gestão da informação.

5. CONSIDERAÇÕES FINAIS

Com o avanço da tecnologia, a área de análise de dados emergiu auxiliando instituições públicas no processo de tomada de decisão. Neste cenário, foi avaliado dois tipos de análise de dados atualmente em voga: a análise de dados exploratória e a análise de dados explanatória, sendo esta mais voltada para a visualização de dados no processo de gestão da informação. Em paralelo, apresentou-se modelos para definir um fluxo de trabalho para a análise exploratória de dados, observando os principais conceitos e definições sobre gestão da informação e do conhecimento.

Por meio da coleta e da importação de dados estruturados, mantidos no Portal da Secretaria de Segurança Pública do Estado de Goiás, no caso o Mportal, demonstrou-se de forma inicial como é feito o processo de análise de dados com a Linguagem de programação Python e suas bibliotecas mais utilizadas. A conclusão que se chega é a de que pode-se utilizar dados já existentes em sites da Secretaria de Segurança Pública do Estado de Goiás para coletar e explorar base de dados estruturados e apresentá-los de forma organizada e visualmente simples, servindo de base para a tomada de decisão institucional.

Ao finalizar esta pesquisa, observa-se que a tendência das instituições policiais é consolidar cada vez mais o uso de tecnologia para balizar a tomada de decisão e definir as estratégias institucionais nas ações policiais; sendo que os dados existentes podem contribuir sobremaneira nesta empreitada, melhorando a relação e a prestação das atividades policiais para a sociedade civil.

REFERÊNCIAS

DADOS ABERTOS. Portal de Dados Abertos. Disponível em: <https://dados.gov.br/dados/>. Acesso em: 25 ago. 2023.

FÁVERO, Luiz Paulo; BELFIORE, Patrícia. **Manual de Análise de Dados**, Estatística e Modelagem Multivariada. 2.ed. Rio de Janeiro, RJ-Brasil: 2017. 1219p.

GIBBS, Graham. **Análise de Dados Qualitativos**. 1.ed. . São Paulo, SP-Brasil: 2008. 192p.

KNAFLIC, Cole Nussbaumer. **Storytelling com Dados: Um guia sobre visualização de dados para profissionais de negócios**. 1.ed. . São Paulo, SP-Brasil: 2017. 257p.

TARAPANOFF, Kira. **Inteligência, Informação e Conhecimento**. IBICT, UNESCO 1.ed. Brasília, DF-Brasil: 2006. 439p.

MCKINNEY, Wes. **Python para Análise de Dados**, Tratamento de Dados Com Pandas Numpy e Ipython. 2.ed. São Paulo, SP-Brasil: 2019. 892p.

OLIVEIRA, Paulo Felipe. **Ciência de Dados com R: Introdução**. 1.ed. Brasília, DF-Brasil: 2018. 198p.

WIKIPÉDIA. **Indústria 4.0 ou Quarta Revolução Industrial**. Disponível em: https://pt.wikipedia.org/wiki/Indústria_4.0, Acesso em: 23 set. 2023.