

## Descoberta de *insights* na análise de licitações no estado de Goiás

M.A.S. Silva <sup>a,\*</sup>, S.L. Vieira <sup>a,b</sup>

<sup>a</sup> Escola de Ciências Exatas e de Computação, Pontifícia Universidade Católica de Goiás, Goiânia (GO), Brasil

<sup>b</sup> Instituto de Criminalística Leonardo Rodrigues, Superintendência de Polícia Técnico-Científica, Goiânia (GO), Brasil

\*Endereço de e-mail para correspondência: [matheusadaosilva@gmail.com](mailto:matheusadaosilva@gmail.com). Tel.: +55-62-99601-7054.

Recebido em 30/06/2021; Revisado em 27/08/2023; Aceito em 31/08/2023

### Resumo

Este trabalho tem o objetivo de relacionar possíveis indícios de irregularidades em processos licitatórios, propiciando uma avaliação de sua viabilidade em processos de fiscalização. Para melhor entendimento desse estudo foi realizada uma revisão bibliográfica a fim de conhecer os fatos e fenômenos relacionado a licitações públicas. Foram explorados os conceitos de inteligência analítica, abordando *business Intelligence*, *data science* e técnicas de mineração de dados. Em seguida a seleção dos dados foi iniciada, sendo possível encontrar dados adequados disponibilizados no portal da transparência da Controladoria Geral do Estado. Os *dataframes* relacionados às licitações do órgão foram estruturados em dados associados aos períodos de janeiro de 2016 a dezembro de 2019. Através da análise descritiva, utilizando o Microsoft Power BI, foi possível empreender uma maior compreensão das variáveis presentes e identificação de indícios de padrões nos dados. A análise preditiva que se seguiu proporcionou gerar regras associando duas ou mais empresas com participações e vitórias conjuntas, podendo apresentar o indício de possíveis irregularidades. Os itens apresentados formam um potencial para o desenvolvimento de sistemas de apoio à decisão e definição de estratégias para fiscalizar o processo licitatório, pois geram informações que permitem maior assertividade em indicar possíveis irregularidades, permitindo assim uma melhor eficiência no uso dos recursos públicos.

*Palavras-Chave:* Licitações; Cartelização; Ciência de dados; Simulação; Descoberta de conhecimento.

### Abstract

This work aims to relate possible evidence of irregularities in bidding processes, providing an assessment of their viability in inspection processes. For a better understanding of this study, a bibliographical review was carried out in order to know the facts and phenomena related to public biddings. The concepts of analytical intelligence were explored, addressing business intelligence, data science and data mining techniques. Then the selection of the data began, and it was possible to find adequate data made available on the transparency portal of the General Controllershship of the State. The dataframes related to the agency's bids were structured on data associated with the periods from January 2016 to December 2019. Through descriptive analysis, using Microsoft Power BI, it was possible to undertake a greater understanding of the variables present and identification of evidence of patterns in the data. The predictive analysis that followed led to the generation of rules associating two or more companies with joint participations and victories, which may indicate possible irregularities. The items presented form a potential for the development of decision support systems and definition of strategies to supervise the bidding process, as they generate information that allows greater assertiveness in indicating possible irregularities, thus allowing a better efficiency in the use of public resources.

*Keywords:* Bids; Cartelization; Data science; Simulation; Knowledge discovery.

## 1. INTRODUÇÃO

A licitação é um procedimento administrativo em que um ente público abre a todos os interessados que se sujeitem às condições no instrumento convocatório a possibilidade de formularem propostas dentre as quais

selecionará a mais conveniente para a celebração de contrato [1].

A Controladoria Geral do Estado de Goiás (CGE/GO) é o órgão central dos sistemas de controle interno, correição, transparência e ouvidoria, integrante da Administração Direta do Poder Executivo do Estado de Goiás, na forma da Lei nº 20.491, de 25 de junho de 2019.

O órgão supracitado, conforme legislação disponibiliza o portal da transparência, uma plataforma *online* em que a população pode consultar informações como: despesas; receitas; convênios e repasses; gestão fiscal e governamental; pessoal; planejamento e orçamento; de todos os órgãos do estado. Dessa forma, o órgão disponibiliza mais um mecanismo de fiscalização a ser utilizado na atuação de controle interno do estado.

Entretanto, com o crescimento do volume de processos licitatórios em intervalos de tempo curtos, produzir informações que agreguem valor à tomada de decisão tornou-se um desafio, necessitando de pesquisas científicas que forneçam bases teóricas e metodológicas com a finalidade de identificar possíveis irregularidades nos processos de licitações [2].

Neste sentido, a ciência de dados é uma área interdisciplinar que relaciona dados com inteligência artificial, aprendizado de máquina, matemática e modelos estatísticos que tem, como sucedâneo, a descoberta de padrões nos dados. A mineração de dados, parte do processo desta descoberta, é um conjunto de tarefas, técnicas e algoritmos que efetiva automaticamente ou de forma semiautomática a descoberta de conhecimento nas bases de dados (DCDB). Os dados devem estar presentes em quantidades substanciais e os padrões descobertos devem ser significativos, pois levam a alguma vantagem [3].

Uma demonstração do potencial da análise técnica da informação foi evidenciada no Laudo de Perícia Criminal n.º 1342/2015-SETEC/SR/DPF/PR onde, por meio de estudo estatístico, concluiu-se que as empresas participantes das licitações das Refinarias RNEST e COMPERJ apresentaram propostas que não refletem a expectativa de distribuição estatística definida nas normas internacionais e pela própria Petrobras, indicando uma variação de preços muito superior à esperada e consistente com a hipótese de cartelização, tendo as demais empresas apresentado propostas com variação de preços consistente com a hipótese de “apresentação de proposta com a finalidade de dar mera cobertura” [4].

Desta forma, com vários casos de corrupção em licitações no Brasil e no estado de Goiás, em que os esforços de garantir a melhor opção de negócio para a administração pública, com a conseqüente economia para os cofres públicos, este artigo apresenta um estudo de caso com o objetivo de relacionar possíveis indícios de irregularidades em processos licitatórios, propiciando uma avaliação de sua viabilidade em processos de fiscalização. Sendo assim uma pergunta pode ser feita: a implantação de mineração de dados pode contribuir na identificação de possíveis irregularidades em processos licitatórios em dados disponibilizados pela CGE/GO?

Nesse trabalho, a utilização de estatística descritiva, em conjunto com tarefas de mineração de dados como

agrupamento e associação, através dos algoritmos relacionados, mostrou-se de grande valia na obtenção de informações potencialmente úteis. Em particular, foi possível identificar e analisar a aplicabilidade das técnicas de mineração de dados para o processo de fiscalização nos órgãos governamentais, mais especificamente voltado ao CGE/GO, bem como reconhecer entidades conforme recorrências em grupos suspeitos, levando em conta a tendência à cartelização, concentração em órgãos específicos e a presença de simulação via concorrentes fictícios.

Este estudo insere-se no campo de aplicações de técnicas inteligentes na área de análise de irregularidades em licitações, como [5-7]. Em [5] foi realizado um estudo cujo objetivo foi a detecção de cartéis de licitações públicas utilizando mineração de dados com regras de associação e clusterização. Em [6] foram propostos indicadores para investigação de licitações por meio de técnicas de reconhecimento de padrões estatísticos e mineração de dados, mostrando que, a partir da mineração de dados e utilização de técnicas de reconhecimento de padrões estatísticos, foi possível a obtenção de categorias de empresas que indique uma maior probabilidade de atuarem em licitações fraudulentas. Em [7] foi utilizado dados do Tribunal de Contas do Estado da Paraíba aplicando Regra de Associação nos dados obtidos, mediante o algoritmo FP-Growth, identificando a repetição dos mesmos participantes em vários processos licitatórios, indicando situações suspeitas de conluio entre as empresas.

## 2. PROCEDIMENTO METODOLÓGICO

De acordo com os meios utilizados para investigação, esse estudo inicia com uma pesquisa bibliográfica e documental, utilizando-se de livros, artigos acadêmicos, revistas, leis, dissertações, relacionado aos temas: Ciência de Dados, descoberta de conhecimento, mineração de dados e licitações públicas [8].

Foi realizado uma revisão bibliográfica nos estudos relacionados a *business Intelligence*, *data science*, técnicas de mineração de dados e licitações públicas no Brasil, de conceituação e exploração das técnicas, e de como o processo de licitações é realizado, mais especificamente no estado de Goiás, em artigos e livros dos últimos 10 anos. Após, foram levantados estudos que tiveram como objetivo identificar empresas cometendo fraudes em órgãos governamentais, pretendendo-se analisar e complementar as variáveis utilizadas para a obtenção de indicadores.

Em seguida, iniciou-se a etapa de seleção dos dados, através do BI da ComprasNet, que é a organização que disponibiliza os dados do CGE/GO no portal da transparência do estado, localizando o maior número de indicadores possível. A coleta dos dados foi realizada entre agosto de 2020 e setembro de 2020 e restringindo as

licitações no período de janeiro de 2016 a dezembro de 2019.

Com a seleção dos dados concluída, a etapa de pré-processamento dos dados foi iniciada, realizando todo o processo de organização e transformação, com o objetivo de eliminar inconsistências e redundâncias dos dados, utilizando como base a metodologia desenvolvida por [6] que utiliza técnicas de reconhecimento de padrões estatísticos para realizar a extração dos dados, separação das variáveis e obtenção dos indicadores, fazendo uma relação com os itens propostos por ele e adaptando aos itens disponíveis no CGE/GO. As informações foram divididas inicialmente em 5 (cinco) *dataframes*, sendo eles: propostas de licitações; licitações; participantes; participantes nos órgãos; órgãos

A etapa seguinte, análise descritiva dos dados, foi realizada utilizando a ferramenta de visualização de dados Microsoft Power BI Desktop, usando painéis interativos que relacionam os atributos característicos às propostas das licitações, às licitações, às empresas participantes, aos participantes no órgão e aos órgãos. Essa análise é importante para a avaliação das características das licitações, pois permite um entendimento do problema e das variáveis relevantes.

Na última etapa, análise preditiva, foi dividida em 3 (três) sub etapas, sendo elas: Etapa 1 – Aplicação de Regra de Associação; Etapa 2 – Aplicação de Clusterização; Etapa 3 – Aplicação de Regra de Associação. Estas etapas foram realizadas com o apoio do software de aprendizado de máquina RapidMiner.

A análise foi realizada baseando-se na metodologia desenvolvida por [5] em que foi aplicada a Regra de Associação com o apoio do algoritmo FP- Growth no *dataframe* de propostas de licitações. O Algoritmo supracitado codifica o conjunto de dados em uma estrutura de dados compacta em forma de árvore chamada *Frequent Pattern tree* (FP-tree) e extrai os conjuntos de itens frequentes diretamente desta estrutura [9].

Para o algoritmo FP-Growth funcionar corretamente é preciso informar pelo menos dois parâmetros de entrada, o suporte e a confiança. O suporte é a métrica utilizada para encontrar todos os  $N$  *itemsets*. O suporte de uma regra de associação  $X, A \Rightarrow B$  corresponde a frequência com que A e B ocorrem em uma base de dados. A confiança é responsável pela força da regra, ou seja, a confiança corresponde a frequência com que B ocorre, dentre os eventos que contêm A [10].

Em seguida, utilizando o mesmo *dataframe* procurou-se encontrar grupos de empresas que reunissem comportamentos e características semelhantes, utilizando a tarefa de mineração de dados *clusterização*, com técnica de agrupamento em métodos de particionamento e o algoritmo *K-means* [11]. O *K-means* é um algoritmo que implementa a clusterização utilizando a técnica de

centróides, agrupa dados que compartilham características importantes e similares, escolhendo o objeto que lhe parece mais promissor [12].

A última etapa foi realizada a regra de associação novamente com os clusters obtidos, feito a escolha das melhores regras e analisado as características das empresas nas participações e nas vitórias.

### 3. RESULTADOS

Nessa seção descreve-se os resultados que foram obtidos nesse estudo, divididos em 3 (três) partes, sendo eles, resultados para as etapas de seleção dos dados e pré-processamento, resultados para a etapa da análise descritiva e resultados para a etapa da análise preditiva.

#### 3.1. Seleção dos dados e pré-processamento

A Controladoria Geral do Estado de Goiás divulga no endereço eletrônico <http://www.transparencia.go.gov.br/portaldatransparencia/> o portal da transparência do órgão. Nesse portal estão disponíveis informações divididas em 6 (seis) categorias, sendo elas: despesas; receitas; convênios e repasses; pessoal; planejamento e orçamento; gestão fiscal e governamental. Dentro da categoria despesas encontra-se informações de advogados dativos, contratos, empenhos e pagamentos, execução orçamentária, fornecedores, gastos governamentais, gastos com publicidade e propaganda, ordem cronológica de pagamentos e informações sobre licitações, sendo o último o foco desse estudo.

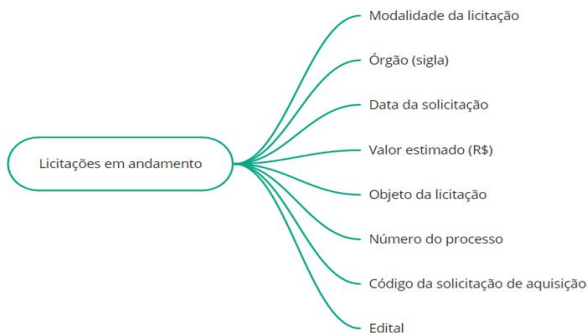
O portal apresenta duas opções para obtenção de informações sobre licitações. A primeira opção é acessando através da categoria despesas e depois em licitações, tendo o acesso redirecionado para o painel de BI da ComprasNet do estado de Goiás. A segunda opção é através da seção dados abertos, seção que permite que pessoas interessadas obtenham todas as informações que são apresentadas nos painéis analíticos do Portal. Estas informações são disponibilizadas de forma tabular simples para tratamento em planilhas, bancos de dados, ferramentas analíticas ou programas próprios, permitindo assim a realização de análises complementares às análises apresentadas nas outras seções do Portal.

Na segunda opção de acesso aos dados de licitações, as informações são disponibilizadas em arquivos no formato *comma-separated values* (CSV), tipo de formato de arquivo de valores separados por vírgula amplamente utilizado para armazenar dados tabulares. Os dados são disponibilizados no servidor *File Transfer Protocol* (FTP) do estado de Goiás, organizados por diretório, bastando apenas acessar o que deseja e realizar o *download*. Essa seria a opção mais fácil de obter os dados. Entretanto, foi observado que os dados disponíveis nesse local são disponibilizados sempre no dia 25 (vinte e cinco) de cada

mês e depois não são atualizados novamente, acarretando informações incompletas por não contemplar os demais dias do mês. Dessa forma, optou-se pela seleção dos dados através da primeira opção, painel da ComprasNet, em que as informações são atualizadas diariamente.

Esse painel contém uma seção denominada Visão geral, contendo informações gerais das licitações, tais como, quantidade de licitações e valores adjudicados por modalidade e histórico de valores adjudicados por mês, podendo filtrar as informações por ano de solicitação, mês da solicitação e nome do órgão. Contém ainda mais 4 (quatro) seções, sendo elas: licitações em andamento; licitações concluídas; participantes de licitações; vencedores das licitações.

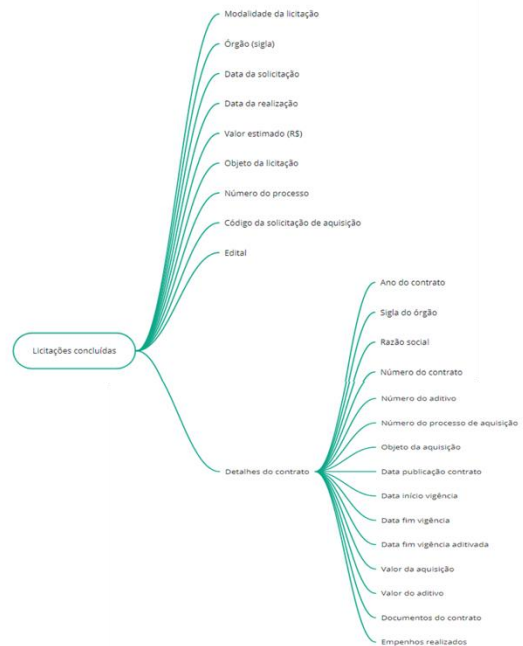
Na seção sobre as licitações em andamento é possível filtrar por ano da solicitação, mês da solicitação, órgão, modalidade da licitação, número do processo, valor estimado e data da solicitação, disponíveis conforme a **Figura 1**.



**Figura 1.** Informações disponíveis na seção sobre as licitações em andamento, elaborado pelo autor.

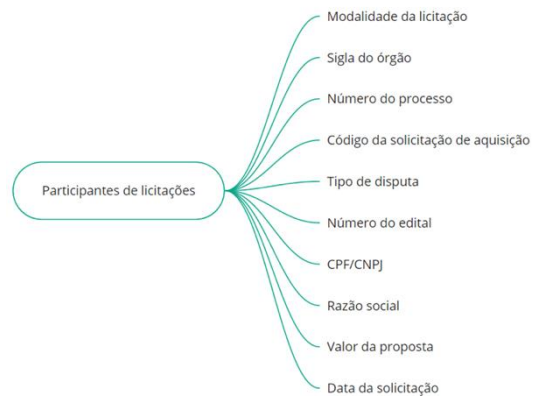
As informações de número do processo, código da solicitação de aquisição e edital possuem *hyperlink*. No número do processo é possível ser redirecionado para a página do Sistema Eletrônico de Informação (SEI), podendo obter informações da data de início do processo administrativo. Ao clicar no código da solicitação de aquisição o acesso é direcionado para uma página do ComprasNet, podendo-se obter diversos dados acerca da licitação, tais como os itens que compõe a licitação, participantes, vencedor, quantidade de itens, local de entrega dos itens, natureza, formas de pagamento e informações de valores. Ao clicar em edital é exibido o edital referente a licitação no formato *Portable Document Format* (PDF).

Na seção sobre as licitações concluídas também é possível filtrar por ano da solicitação, mês da solicitação, órgão, modalidade da licitação, número do processo, valor estimado e data da solicitação, estando disponíveis as informações conforme **Figura 2**.



**Figura 2.** Informações disponíveis na seção sobre as licitações concluídas, elaborado pelo autor.

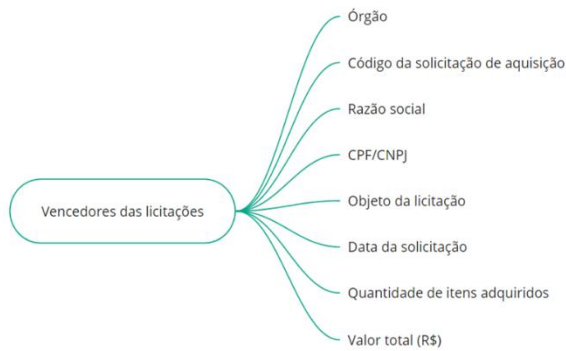
Na seção sobre os participantes de licitações é possível filtrar por ano da solicitação, mês da solicitação, modalidade da licitação, órgão, tipo de disputa, razão social, CPF/CNPJ, número do edital e número do processo, estando disponíveis as informações conforme **Figura 3**.



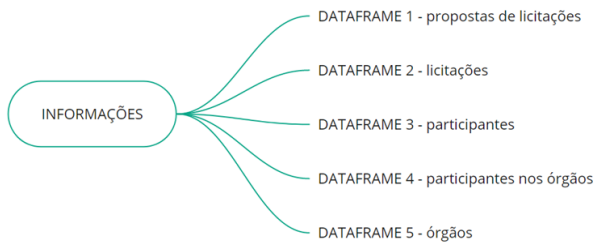
**Figura 3.** Informações disponíveis na seção sobre os participantes de licitações, elaborado pelo autor.

A última seção de informações disponíveis é referente aos vencedores das licitações, podendo-se filtrar por ano da solicitação, mês da solicitação, nome do órgão, razão social, CPF/CNPJ, valor total e data da solicitação, estando disponíveis as informações conforme **Figura 4**.

Para esse estudo restringiu-se todas as informações disponíveis no período de janeiro de 2016 a dezembro de 2019. As informações foram divididas inicialmente em 5 (cinco) *dataframes*, conforme é exibido na **Figura 5**.



**Figura 4.** Informações disponíveis na seção sobre os vencedores das licitações, elaborado pelo autor.



**Figura 5.** Organização das informações em 5 (cinco) *dataframes*, elaborado pelo autor.

O **Quadro 1** aborda as variáveis do “DATAFRAME 1 - propostas de licitações”, reunindo as variáveis que identificam tanto o participante como as licitações das quais participaram. Tais variáveis são necessárias não apenas para a obtenção direta de indicadores, mas para a realização de vínculos futuros com outros *dataframes* que possibilitem outros tipos de agrupamentos [6].

Para a construção desse *dataframe* foi utilizado a seção participantes de licitações do BI da ComprasNet, e para obter a informação de valor estimado foi associado o código da solicitação de aquisição com o mesmo código presente nas seções de licitações em andamento e concluída. Nesse *dataframe* cada linha representa uma proposta de um participante, e cada coluna representa as variáveis do **Quadro 1**. A primeira coluna dos Quadros 1 a 5 informa os atributos das licitações conforme a metodologia aplicada em [6].

O **Quadro 2** é relacionado ao “DATAFRAME 2 – licitações”. Para a construção desse *dataframe* foi utilizado a seção vencedores das licitações do BI da ComprasNet, e para obter a informação de valor estimado foi associado o código da solicitação de aquisição com o mesmo código presente nas seções de licitações em andamento e concluída. Nesse *dataframe* cada linha representa um item de cada licitação vencedora, e cada coluna representa as variáveis do **Quadro 2**.

**Quadro 1.** Variáveis do “DATAFRAME 1 - propostas de licitações”.

Metodologia	CGE/GO
Código do participante da licitação, por exemplo CNPJ	CPF/CNPJ
Código de identificação da licitação a qual participou	Número Processo / Código da solicitação de aquisição
Código de identificação do órgão licitante	Código do órgão / Sigla do órgão/ Nome do órgão
Valor da proposta apresentada	Valor da proposta
Identificação de inabilitação, caso tenha sido desclassificada no processo de habilitação	
Identificação da participação em consórcio, caso faça parte de consórcio de empresas na licitação;	
Data da licitação, para delimitações futuras e comparações com diferentes épocas de atuação de cartéis	Data da solicitação
	Valor estimado da licitação
	Modalidade da licitação
	Tipo de disputa
	Número do edital
	Razão social

O **Quadro 3** é referente ao “DATAFRAME 3 – participantes”, apresentando as características gerais do participante. Para a construção desse *dataframe* foi utilizado o “DATAFRAME 1 - propostas de licitações” para selecionar os participantes e o “DATAFRAME 2 – licitações” para compor associações e encontrar novos indicadores conforme o **Quadro 3**. Nesse *dataframe* cada linha representa um participante único (empresa), e cada coluna representa as variáveis do **Quadro 3**.



**Quadro 2.** Variáveis do “DATAFRAME 2 – licitações”.

Metodologia	CGE/GO
Código do órgão licitante	Código do órgão / Sigla do órgão
Código de identificação da licitação	Número Processo / Código da solicitação de aquisição
Modalidade da licitação	Modalidade licitação
Tipo de licitação	Tipo de licitação
Data da licitação	Data solicitação
Natureza do objeto	
Descrição do objeto	Objeto licitação
Valor estimado	Valor estimado
Valor vencedor	Valor Vencedor
Valores das propostas	
CNPJ do vencedor	CNPJ do vencedor
	Razão social do vencedor
	Valor adjudicado unitário
	Quantidade de itens

**Quadro 3.** Variáveis do “DATAFRAME 3 – participantes”.

Metodologia	CGE/GO
CNPJ do participante	CPF/CNPJ
Nome do participante	Razão social
Porte do participante	
Número de participações	Vincular com <i>dataframe</i> 1
Número de licitações realizadas pelos órgãos em que participou – Número total de licitações realizadas pelos órgãos nos quais a empresa participou de licitação tanto as que participou quanto as que não participou	Vincular com <i>dataframe</i> 1
Número total de vitórias em licitações	Vincular com <i>dataframe</i> 2
Número de órgãos em que participou licitações;	Vincular com <i>dataframe</i> 1
Número de órgãos em que venceu licitações;	Vincular com <i>dataframe</i> 2

O **Quadro 4** é referente ao “DATAFRAME 4 - participantes nos órgãos”. Para a construção desse *dataframe* foi utilizado o “DATAFRAME 1 - propostas de licitações” para selecionar os participantes e o “DATAFRAME 2 – licitações” para compor associações e encontrar novos indicadores conforme o **Quadro 4**. Nesse *dataframe* cada linha representa a combinação de um participante (empresa) + órgão único, e cada coluna representa as variáveis do **Quadro 4**.

**Quadro 4.** Variáveis do “DATAFRAME 4 - participantes nos órgãos”.

Metodologia	CGE/GO
CNPJ do participante	CPF/CNPJ
Código do órgão	Código do órgão / Sigla do órgão
Natureza do objeto	
Número de participações do participante no órgão	Vincular com <i>dataframe</i> 1
Valor de todas as participações do participante no órgão	Vincular com <i>dataframe</i> 1
Valor total estimado de todas as licitações do participante no órgão	Vincular com <i>dataframe</i> 1
Número de vitórias do participante no órgão	Vincular com <i>dataframe</i> 2
Valor total de vitórias do participante no órgão	Vincular com <i>dataframe</i> 2
Valor total estimado das vitórias do participante no órgão	Vincular com <i>dataframe</i> 2

**Quadro 5.** Variáveis do “DATAFRAME 5 – órgãos”.

Metodologia	CGE/GO
Código do órgão	Código do órgão / Sigla do órgão
Nome do órgão	Nome do órgão
Número de licitações do órgão	Vincular com <i>dataframe</i> 2
Número de itens de licitações do órgão	Vincular com <i>dataframe</i> 2
Média de participantes por licitação do órgão – total de participantes em todas as licitações no órgão dividido pelo total de licitações no órgão	Vincular com <i>dataframe</i> 2
Valor total das licitações no órgão	Vincular com <i>dataframe</i> 2
Valor total estimado das licitações no órgão	Vincular com <i>dataframe</i> 2
Número total de vencedores no órgão	Vincular com <i>dataframe</i> 2
Número total de participantes no órgão	Vincular com <i>dataframe</i> 1
CNPJ das quatro maiores vencedoras de licitação no órgão	Vincular com <i>dataframe</i> 4

E por fim, o **Quadro 5** apresenta o “DATAFRAME 5 – órgãos”. Para a construção desse *dataframe* foi utilizado o “DATAFRAME 1 - propostas de licitações” para selecionar os órgãos, “DATAFRAME 2 – licitações” e

“DATAFRAME 4 - participantes nos órgãos” para compor associações e encontrar novos indicadores conforme o **Quadro 5**. Nesse *dataframe* cada linha representa um órgão, e cada coluna representa as variáveis do **Quadro 5**.

Os dados obtidos por modalidade de licitação são mostrados na **Tabela 1**.

**Tabela 1.** Quantidade de licitações totais obtidas por modalidade.

Modalidade	Número licitações
Aquisições BID	16
Compra Direta - Dispensa de Licitação	2182
Compra Direta - Obras e Serviços de Enge	39
Concorrência	731
Concorrência SRP	1
Concurso	76
Convite	43
Inexigibilidade	2229
Informação não disponível	7126
Outras Dispensas	1550
Participante Registro de Preço	2607
Pregão Eletrônico	1985
Pregão Eletrônico SRP	647
Pregão Presencial	154
Pregão Presencial RDC	3
Pregão Presencial SRP	9
Tomada de Preço	260
<b>Total</b>	<b>19658</b>

Buscou-se trabalhar apenas com licitações cuja participação possibilitasse a identificação de padrões para tentar reconhecer duas situações: participantes que combinam previamente a atuação conjunta em licitações, caso de cartel; participantes que nunca concorreram, por terem sócios comuns, fantasmas ou promover esquema laranja, passando a atuar em licitações como se fossem concorrentes.

Dessa forma, para ter mais precisão e assertividade nos resultados restringiu-se a metodologia para o escopo de saúde e medicamentos. A partir da descrição do objeto presente no “DATAFRAME 2 – licitações” foi filtrado as licitações cuja descrição havia as palavras da **Figura 6** e suas variações sem acentos, cedilhas e plural. Em seguida foram eliminados dados relacionados às modalidades Compra direta – dispensa licitação, Compra direta – obras e engenharia, Inexigibilidade, Outras dispensas e Informação não disponível.

Com o tratamento descrito, obteve-se os *dataframes* conforme a **Tabela 2**.

Medicamento	Remédio	Fármaco	Medicação
Xarope	Curativo	Hospital	Laboratório
Médico	Hospitalar	Farmacêutico	Saúde

**Figura 6.** Palavras filtradas na descrição do objeto da licitação, elaborado pelo autor.

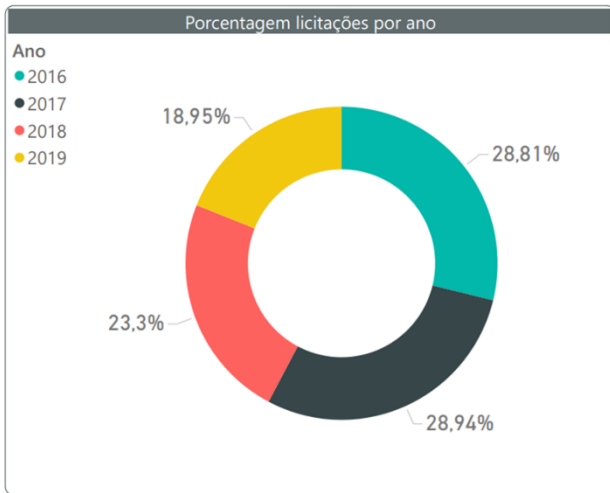
**Tabela 2.** Quantidade de licitações obtidas para saúde e medicamentos por modalidade.

Modalidade	2016	2017	2018	2019
Concorrência	1	4	4	2
Concorrência SRP	1			
Convite		1		
Participante Registro de Preço	462	458	392	328
Pregão Eletrônico	46	40	34	19
Pregão Eletrônico SRP	138	146	93	77
Pregão Presencial	1	1	2	
Pregão Presencial RDC		2		1
Tomada de Preço		1		
<b>Total</b>	<b>649</b>	<b>653</b>	<b>525</b>	<b>427</b>

### 3.2. Análise descritiva

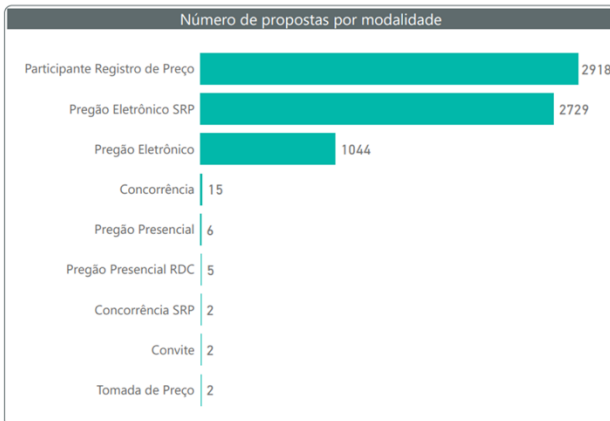
Com os dados obtidos e organizados, o processo de análise descritiva pode ser realizado. Essa etapa é muito importante porque visa resumir e explorar o comportamento dos dados obtidos, ou seja, traduzir diversas linhas e colunas em informações de forma visual. A análise descritiva foi feita usando o Power BI Desktop, amplamente utilizado para aplicativos de inteligência de dados, permitindo interatividade entre gráficos e identificação de padrões. As visualizações obtidas com a ferramenta ilustram o comportamento dos dados dos 5 (cinco) *dataframes* obtidos na etapa anterior.

A **Figura 7** ilustra um gráfico de pizza com a porcentagem de licitações por ano. Observa-se um equilíbrio na quantidade de licitações de 2016 a 2019. O ano de 2016 corresponde a 28,81%, 2017 corresponde 28,94%, 2018 com 23,3% e o ano que contém menos licitações, 2019, contempla 18,95% do total.



**Figura 7.** Distribuição das licitações no período de 2016 a 2019, elaborado pelo autor.

A **Figura 8** exibe o número de propostas por modalidade de todo o período (2016-2019). A modalidade Participante Registro de Preço registrou 2918 propostas, Pregão Eletrônico SRP registrou 2729, Pregão Eletrônico registrou 1044, Concorrência registrou 15, Pregão presencial registrou 6, Pregão Presencial RDC com registro de 5, Concorrência SRP com 2, Convite registrou 2 e Tomada de Preço registrou apenas 2 propostas.

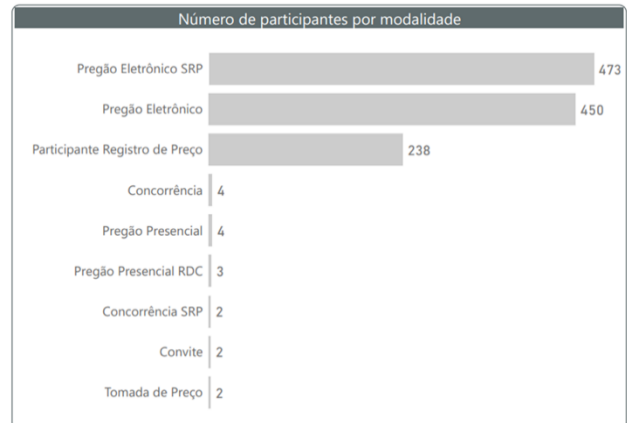


**Figura 8.** Número de propostas por modalidade, elaborado pelo autor.

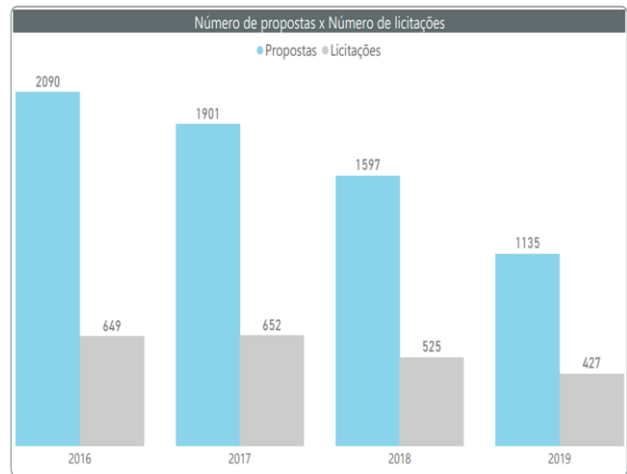
A **Figura 9** exibe o número de participantes por modalidade. Há no total 846 participantes, alguns participando em mais de uma modalidade. A modalidade Pregão Eletrônico SRP contém 473 participantes, Pregão Eletrônico contém 450, Participante Registro de Preço contém 238, Concorrência contém 4, Pregão presencial com 4, Pregão Presencial RDC com 3, Concorrência SRP, Convite e Tomada de Preço com apenas 2.

O número de propostas e número de licitações por ano é exibido na **Figura 10**. No ano de 2016 ocorreram 2090 propostas para 649 licitações, em 2017 foram 1901 propostas para 652 licitações, 2018 com 1597 propostas para 525 licitações e para o ano de 2019 ocorreram 1135 propostas para 427 licitações. Os anos 2016, 2017, 2018 e

2019, representaram respectivamente, 31%, 28%, 24% e 17% da quantidade de propostas. Observa-se uma queda média de 4,6% de propostas ao longo dos anos.

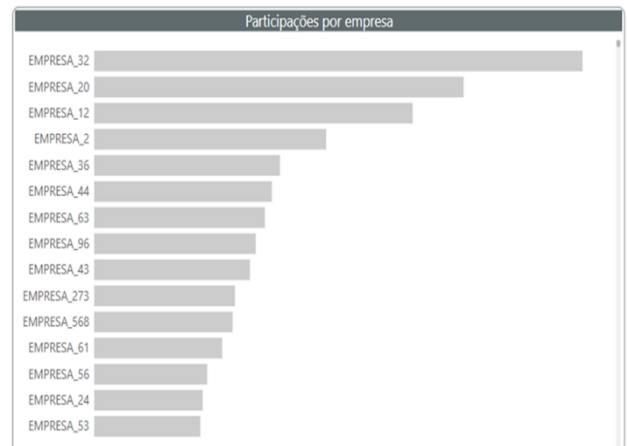


**Figura 9.** Número de participantes por modalidade, elaborado pelo autor.



**Figura 10.** Número de propostas e número de licitações por ano, elaborado pelo autor.

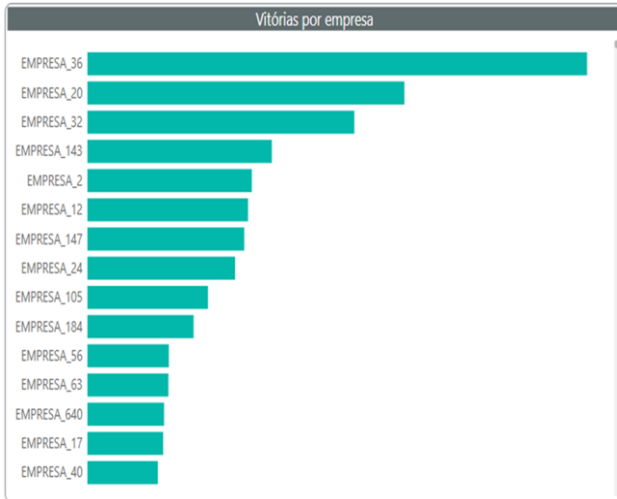
A **Figura 11** ilustra uma estimativa das participações por empresa. Nesse gráfico o campo de indicador da empresa foi substituído por um código sequencial para preservar a identidade do proponente. Neste está listado as empresas com maiores participações para as empresas com menores participações.



**Figura 11.** Participações por empresa, elaborado pelo autor.

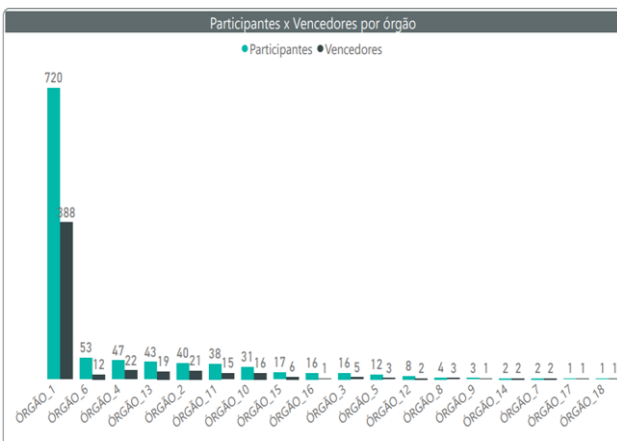


A **Figura 12** ilustra uma estimativa das vitórias por empresa. Neste está listado as empresas que mais vencem licitações para as empresas com menores vitórias. Os elementos envolvidos, órgãos, licitações e participantes são apresentados respectivamente como ORGAO\_n, LIC\_n e EMPRESA\_n, onde n é o número do respectivo elemento entre os elementos do conjunto. A EMPRESA\_36 está entre as que mais vencem, em seguida as EMPRESA\_20 e EMPRESA\_32.



**Figura 12.** Vitórias por empresa, elaborado pelo autor.

A **Figura 13** apresenta o número de participantes e o número de vencedores por órgão. O ÓRGÃO\_1 lidera o ranking com 720 participantes e 388 vencedores. O ÓRGÃO\_6 em seguida com 53 participantes e 12 vencedores. O ÓRGÃO\_4 com 47 participantes e 22 vencedores. Os demais órgãos possuem números de participações e vitórias menores, tendendo a único participante e um único vencedor.



**Figura 13.** Quantidade de participantes e vencedores por órgão, elaborado pelo autor.

A **Figura 14** ilustra os 4 (quatro) maiores vencedores por órgão. A partir da Fig. 25 foi possível mapear as empresas que mais tem vencido, se vence sempre no

mesmo órgão ou em órgãos diferentes, servindo de um alerta para comparações na próxima etapa, análise preditiva.

Órgão	1º vencedor	2º vencedor	3º vencedor	4º vencedor
ÓRGÃO_1	EMPRESA_36	EMPRESA_20	EMPRESA_32	EMPRESA_12
ÓRGÃO_10	EMPRESA_143	EMPRESA_36	EMPRESA_147	EMPRESA_105
ÓRGÃO_11	EMPRESA_291	EMPRESA_24	EMPRESA_410	EMPRESA_147
ÓRGÃO_12	EMPRESA_147	EMPRESA_22	EMPRESA_592	EMPRESA_739
ÓRGÃO_13	EMPRESA_255	EMPRESA_180	EMPRESA_154	Não há vencedor
ÓRGÃO_14	EMPRESA_184	EMPRESA_36	EMPRESA_185	EMPRESA_550
ÓRGÃO_15	EMPRESA_641	EMPRESA_252	Não há vencedor	Não há vencedor
ÓRGÃO_16	EMPRESA_284	EMPRESA_483	EMPRESA_704	Não há vencedor
ÓRGÃO_17	EMPRESA_298	Não há vencedor	Não há vencedor	Não há vencedor
ÓRGÃO_18	EMPRESA_184	EMPRESA_36	EMPRESA_324	EMPRESA_53
ÓRGÃO_2	EMPRESA_184	EMPRESA_431	EMPRESA_612	EMPRESA_510
ÓRGÃO_3	EMPRESA_448	EMPRESA_524	Não há vencedor	Não há vencedor
ÓRGÃO_4	EMPRESA_569	EMPRESA_840	EMPRESA_27	EMPRESA_33
ÓRGÃO_5	EMPRESA_727	EMPRESA_728	Não há vencedor	Não há vencedor
ÓRGÃO_6	EMPRESA_184	EMPRESA_708	EMPRESA_599	EMPRESA_577
ÓRGÃO_7	EMPRESA_750	Não há vencedor	Não há vencedor	Não há vencedor
ÓRGÃO_8	EMPRESA_252	Não há vencedor	Não há vencedor	Não há vencedor
ÓRGÃO_9	EMPRESA_284	Não há vencedor	Não há vencedor	Não há vencedor

**Figura 14.** As 4 (quatro) maiores empresas vencedoras por órgão, elaborado pelo autor.

### 3.3. Análise preditiva

A análise descritiva, uma vez finalizada, permite uma observação que é de grande relevância pois permite um maior entendimento do problema e uma visão de todas as variáveis disponíveis, propiciando o início da análise preditiva. Para essa análise não foram utilizados todos os indicadores dos 5 (cinco) dataframes descritos na seção anterior, e sim os indicadores que representasse os participantes, vencedores e licitações, utilizando a ferramenta de mineração de dados RapidMiner e dividindo essa análise em 3 (três) etapas, sendo elas:

- Etapa 1 – Aplicação de Regra de Associação
- Etapa 2 – Aplicação de Clusterização
- Etapa 3 – Aplicação de Regra de Associação

As operações no RapidMiner funcionam por meio de processos, ou seja, para realizar qualquer tarefa de mineração é preciso incluir operadores para cada processo que antecede a tarefa final desejada, sendo que alguns operadores possuem parâmetros a serem ajustados, necessitando que o usuário os informe. Os operadores, sequência de execução e parâmetros são armazenados em um arquivo do tipo eXtensible Markup Language (XML), permitindo editar para ajustes ou para salvar as configurações do experimento realizado para futura reutilização [7].

A ferramenta possui uma interface gráfica em que o processo de tratamento dos dados pode ser definido inserindo os operadores dentro de um fluxo, localizado na parte central da **Figura 15**. Os operadores que podem ser inseridos no processo estão disponíveis ao lado esquerdo e os detalhes de configurações estão ao lado direito da interface.

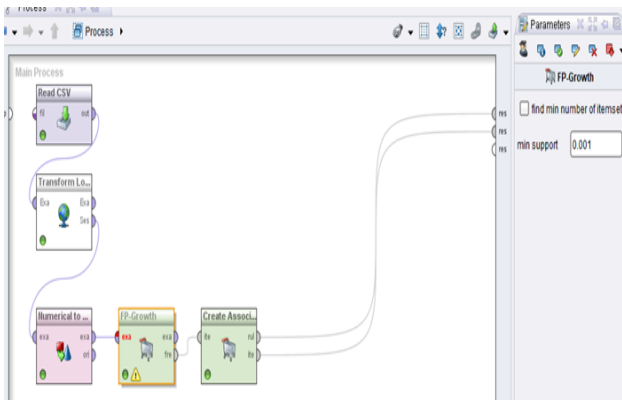


Figura 15. Área de trabalho da ferramenta RapidMiner, elaborado pelo autor.

Para a realização da primeira etapa, a partir da base de dados do “DATAFRAME 1 - propostas de licitações” foi gerado um *dataset* 1 contendo a relação de licitações e as empresas participantes, conforme ilustrado no Quadro 6. Esse *dataset* foi composto por 5555 linhas excetuando o cabeçalho, distribuídas nas 2254 licitações e nos 846 participantes, ou seja, cada linha representando uma licitação, e em licitações que contém mais de uma empresa participante, o número da licitação se repete, porém com o indicador da empresa diferente, representando uma nova empresa para a licitação.

Quadro 6. Dataset 1 contendo as licitações e empresas.

Licitação	Empresa
LICITAÇÃO_1	EMPRESA_1
LICITAÇÃO_1	EMPRESA_2
LICITAÇÃO_2	EMPRESA_3
...	...
LICITAÇÃO_n	EMPRESA_n

Feito essa relação das licitações e empresas participantes, o arquivo foi salvo no formato CSV. Para ler esse arquivo no RapidMiner foi utilizado o operador ReadCSV, conforme a Figura 15. Este operador lê um arquivo CSV e gera um conjunto de treinamento, uma tabela com duas colunas e 5556 linhas que representam as licitações e as propostas.

A tabela criada a partir do operador ReadCSV pode ser acessada por qualquer outro operador através da linha *out*. Neste estudo, o operador que foi utilizado em seguida foi o que transforma a tabela gerada pelo operador ReadCSV em uma matriz esparsa. O operador que realiza esse procedimento é o *Transform log to session*.

No operador *Transform log to session*, é preciso configurar dois parâmetros, sendo eles: *session attribute* e *resource attribute*. Nesse caso, o *session attribute* recebe o nome da coluna que contém as licitações e *resource attribute* o nome da coluna que contém os participantes, conforme Figura 16.

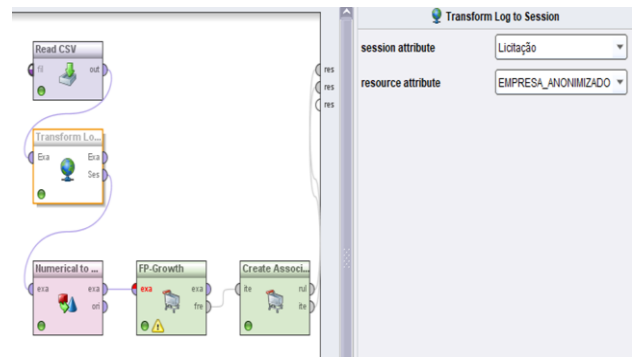


Figura 16. Configurações dos parâmetros do operador Transform log to session, elaborado pelo autor.

O operador *Transform log to session* possui duas portas de saída, uma com o resultado do processamento e outra com os dados originais, tendo sido utilizado somente a saída que fornece o resultado do processamento. Após obter a matriz esparsa, não é possível utilizar o *Transform log to session* diretamente na entrada do operador FP-Growth, que é o operador responsável por identificar os conjuntos de itens que são frequentes, porque esse operador somente manipula dados do tipo binário. Desta forma, é necessário transformar todos os dados do tipo numérico em dados do tipo binário, e o operador responsável por esse procedimento é o *Numerical to Binomial*, gerando os dados similar aos do Quadro 7.

Quadro 7. Matriz esparsa das licitações e empresas.

Licitações	EMPRESA_1	EMPRESA_n
LICITAÇÃO_1	true	...
LICITAÇÃO_2	false	...
LICITAÇÃO_3	false	...
LICITAÇÃO_n	...	...

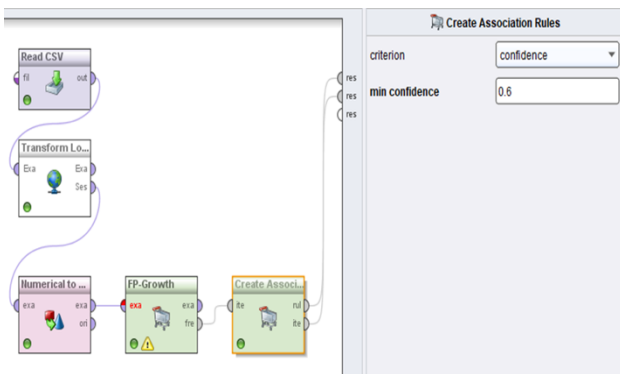
Depois de utilizar o operador *Numerical to Binomial* foi possível fazer uso do operador FP-Growth para a identificação dos itens que são frequentes. A Figura 17 ilustra o único parâmetro ajustável desse operador, o *min support*, que é o valor do suporte mínimo a ser considerado de acordo com as características do problema.



Figura 17. Configurações do parâmetro do operador FP-Growth, elaborado pelo autor.

Os itens frequentes do operador FP- Growth são a entrada para o operador *Create Association Rules*. Com base nos itens frequentes e no valor do *min support*, o operador *Create Association Rules* gera as regras de associação. A **Figura 18** ilustra o único parâmetro ajustável desse operador, *min confidence*, ou seja, o valor mínimo para a confiança.

No RapidMiner, o processo de mineração de itens frequentes é dividido nesses dois operadores, FP- Growth realizando a geração de itens frequentes, e o *Create Association Rules* realizando as regras de associações a partir do conjunto do operador anterior.



**Figura 18.** Configurações do parâmetro do operador *Create Association Rules*, elaborado pelo autor.

Nesse primeiro experimento foi utilizado o suporte mínimo de 0.01 (1%) e confiança mínima de 0.60 (60%). Com estes parâmetros foram encontradas 16 regras de associação com 84 conjuntos de itens frequentes conforme é ilustrado na **Figura 19**.

No	Premises	Conclusion	Support	Confid.	LaPla	Gain	p-s	Lift	ComL
1	EMPRESA_58	EMPRESA_32	0.012	0.695	0.993	-0.027	0.009	3.221	2.955
2	EMPRESA_24	EMPRESA_36	0.025	0.609	0.984	-0.058	0.022	8.539	2.960
3	EMPRESA_56	EMPRESA_20	0.028	0.643	0.985	-0.059	0.022	4.526	2.402
4	EMPRESA_57	EMPRESA_20	0.011	0.658	0.994	-0.023	0.009	4.632	2.508
5	EMPRESA_32, EMPRESA_43	EMPRESA_20	0.018	0.667	0.991	-0.036	0.014	4.694	2.574
6	EMPRESA_20, EMPRESA_43	EMPRESA_32	0.018	0.667	0.991	-0.036	0.013	3.551	2.437
7	EMPRESA_20, EMPRESA_214	EMPRESA_32	0.011	0.667	0.995	-0.021	0.008	3.551	2.437
8	EMPRESA_20, EMPRESA_214	EMPRESA_56	0.011	0.667	0.995	-0.021	0.010	15.322	2.670
9	EMPRESA_32, EMPRESA_56	EMPRESA_20	0.016	0.686	0.993	-0.030	0.012	4.632	2.735
10	EMPRESA_214	EMPRESA_20	0.016	0.692	0.993	-0.030	0.013	4.674	2.788
11	EMPRESA_12, EMPRESA_43	EMPRESA_32	0.010	0.697	0.996	-0.019	0.007	3.712	2.680
12	EMPRESA_12, EMPRESA_56	EMPRESA_32	0.010	0.719	0.996	-0.018	0.008	3.828	2.888
13	EMPRESA_12, EMPRESA_43	EMPRESA_20	0.011	0.758	0.997	-0.018	0.009	5.334	3.539
14	EMPRESA_12, EMPRESA_56	EMPRESA_20	0.011	0.781	0.997	-0.017	0.009	5.500	3.922
15	EMPRESA_32, EMPRESA_214	EMPRESA_20	0.011	0.828	0.998	-0.015	0.009	5.827	4.979
16	EMPRESA_56, EMPRESA_214	EMPRESA_20	0.011	0.857	0.998	-0.014	0.009	6.035	6.006

**Figura 19.** Resultado do processamento dos dados minerados, elaborado pelo autor.

Foi possível perceber que foram geradas poucas regras de associação. Além disso as regras obtidas tiveram o suporte e a confiança não tão altos, a confiança máxima obtida foi de 0.85 (85%) e o suporte máximo de 0.028 (2,8%). Para a situação de cartel, mais especificamente rodízio de licitações, o valor alto de confiança garante que a frequência de ocorrência dos participantes do lado

antecessor da regra seja aproximada à frequência de ocorrência dos participantes do lado consequente da regra.

Para se obter mais regras com maior confiança o valor do suporte é diminuído para 0.001 (0.1%), implicando um aumento no número de regras para 12.163 com 4.390 itens frequentes conforme **Tabela 3**. Porém, mesmo com o número elevado da quantidade de regras, algumas poderiam ficar de fora, devido a ocorrência delas serem menor, e podendo ser uma regra de qualidade, ou seja, dessa forma não se tem uma precisão, que é tão importante para essa situação.

**Tabela 3.** Resultados execução do algoritmo FP-Growth para os dois suportes.

	Suporte mínimo	Confiança mínima	Número de regras	Itens frequentes
<i>Dataset</i> 1	0.01 (1%)	60%	16	84
<i>Dataset</i> 1	0.001 (0.1%)	60%	12.163	4.390

Como a base de dados da CGE/GO abordada nesse estudo contém licitações de diversos órgãos do período de 2016 a 2019, uma solução foi agrupar todos os itens por similaridade, utilizando a tarefa de Clusterização. Um novo *dataset* foi preparado para que a Clusterização encontrasse os prováveis grupos que estão simbolizando os processos de licitações. Desta forma, poder-se-ia executar a Regra de Associação novamente em cada *cluster* gerado, já que ter-se-ia em cada grupo um provável conjunto de licitações, dividindo desta forma o espaço de soluções e obtendo mais precisão nos *insights*.

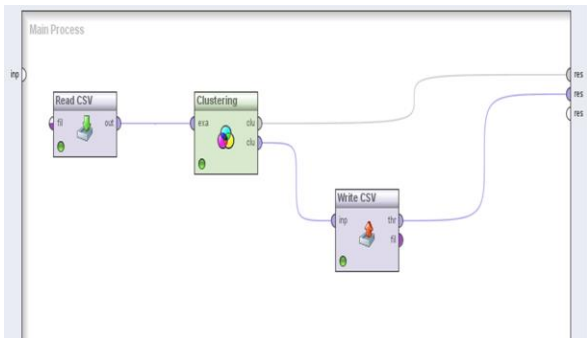
Nesta segunda etapa, a partir da base de dados do “DATAFRAME 1 - propostas de licitações” foi preparado um *dataset* 2 contendo os atributos de licitação, empresa participante e órgão, conforme ilustrado no **Quadro 8**. Esse *dataset* foi composto por todos os itens da base supracitada, contendo 6724 registros, ou seja, cada linha representando licitações e os itens, caso a licitação possua mais de um item.

**Quadro 8.** Dataset 2 para realizar a Clusterização.

Licitação	Empresa	Órgão
LICITAÇÃO_1	EMPRESA_1	ÓRGÃO_1
LICITAÇÃO_1	EMPRESA_2	ÓRGÃO_2
LICITAÇÃO_2	EMPRESA_3	ÓRGÃO_3
...	...	...
LICITAÇÃO_n	EMPRESA_n	ÓRGÃO_n

Em seguida o *dataset* 2 foi salvo no formato CSV e inserido no RapidMiner, efetuando a leitura do arquivo através do operador ReadCSV conforme é ilustrado na

**Figura 20.** Após a leitura do arquivo, a saída do operador ReadCSV é conectada na entrada do operador Clustering, operador responsável por realizar o modelo de clusterização. Em seguida, a primeira saída do operador Clustering, que é responsável por gerar o modelo é conectada na entrada do resultado do processamento. A segunda saída do operador Clustering possibilita visualizar o cluster gerado correspondente a cada registro da base de dados similar ao que é exibido no Quadro 9. Para finalizar o processo, a segunda saída é conectada no operador WriteCSV, operador que permite exportar os dados gerados em um arquivo CSV.



**Figura 20.** Processo para realizar a Clusterização no RapidMiner, elaborado pelo autor.

**Quadro 9.** Dataset 2 com a coluna de clusters gerados.

Licitação	Empresa	Órgão	Cluster
LICITAÇÃO_1	EMPRESA_1	ÓRGÃO_1	CLUSTER_1
LICITAÇÃO_1	EMPRESA_2	ÓRGÃO_2	CLUSTER_1
LICITAÇÃO_2	EMPRESA_3	ÓRGÃO_3	CLUSTER_2
...	...	...	...
LICITAÇÃO_n	EMPRESA_n	ÓRGÃO_n	CLUSTER_n

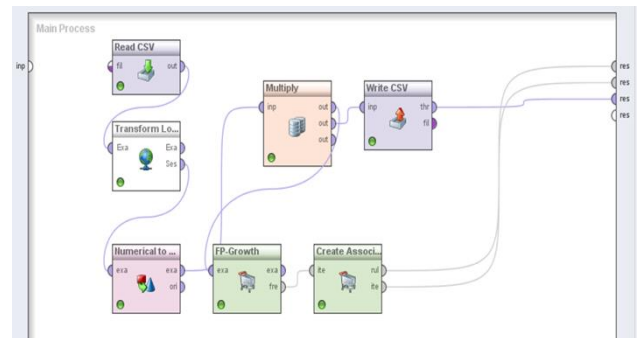
Esse experimento foi realizado utilizando técnica de agrupamento em métodos de particionamento, com o apoio do algoritmo *K-means*. Com o modelo gerado foi possível obter 7 (sete) clusters com a quantidade de itens em cada cluster conforme é exibido na Tabela 4.

Com os grupos encontrados na aplicação do algoritmo de clusterização *K-means*, foi possível iniciar a etapa 3 da análise preditiva, aplicando Regra de Associação novamente, porém agora em cada cluster obtido, na tentativa de identificar grupos de empresas associadas atuando especificamente nos novos grupos definidos. Além disso, com menos registros, foi possível obter mais regras em cada grupo, com a confiança e suporte maiores, proporcionando mais qualidade na escolha das regras e mais precisão na geração de insights.

**Tabela 4.** Quantidade de itens por clusters gerados.

CLUSTER	QUANTIDADE DE ITENS
CLUSTER_0	963
CLUSTER_1	927
CLUSTER_2	1280
CLUSTER_3	870
CLUSTER_4	966
CLUSTER_5	850
CLUSTER_6	867
<b>TOTAL</b>	<b>6723</b>

Foram gerados 7 (sete) datasets, um para cada cluster, por meio do processo ilustrado na Figura 21, similar ao realizado na etapa 1, com o acréscimo do operador WriteCSV para conseguir exportar os dados obtidos e realizar análises das melhores regras. Os atributos utilizados nessa etapa também foram a licitação e a empresa participante, similar ao apresentado no Quadro 6. A Tabela 5 mostra as principais informações obtidas ao aplicar associação em cada cluster com suporte mínimo de 1% e confiança mínima de 60%.



**Figura 21.** Processo para realizar a Regra de Associação em cada cluster obtido, elaborado pelo autor.

**Tabela 5.** Execução do algoritmo de associação FP-Growth para os datasets de clusters.

Cluster	Licitações	Atributos	Itens frequentes	Regras
CLUSTER_0	340	251	120	160
CLUSTER_1	296	274	217	600
CLUSTER_2	398	307	428	1771
CLUSTER_3	321	203	94	46
CLUSTER_4	314	245	107	31
CLUSTER_5	269	286	140	139
CLUSTER_6	316	182	158	274

Posteriormente, a partir do “DATAFRAME 2 – licitações” foi gerado um dataset dos vencedores, contendo os atributos de licitação e empresa vencedora, e em seguida, gerado uma matriz esparsa, similar ao apresentado no Quadro 7. Com as regras de associação dos clusters gerados foi possível escolher as melhores regras por meio

dos parâmetros fornecidos como, Suporte, Confiança, *Lift* e Convicção, para analisar com a matriz esparsa dos vencedores, verificar o comportamento das empresas nas participações e nas vitórias, para averiguar características irregulares, como cartel e rodízio.

#### 4. DISCUSSÃO

Algumas das regras de associação obtidas nesse trabalho são apresentadas a seguir. A detecção de empresas com participações em conjunto e as vitórias em licitações podem levar a indícios de conluio, e 3 (três) regras foram selecionadas com o intuito de ilustrar os resultados alcançados.

Uma regra presente no “CLUSTER\_0” que possui o valor de suporte mais alto desse *cluster*, sendo de 0.032, informa que sempre que a EMPRESA\_2 participa a EMPRESA\_36 também participa em 64,7% dos casos. Essas duas empresas participaram juntas de 11 (onze) licitações e celebraram contrato juntas em 8 (oito) licitações. Das 11 (onze) licitações que participaram juntas, a EMPRESA\_2 venceu sozinha em apenas 1 (uma) licitação, e a EMPRESA\_36 venceu sozinha em 2 (duas) licitações. Essas duas empresas possuíam um número relativamente baixo de participações no *cluster*: a EMPRESA\_2 contém 17 (dezesete) participações e a EMPRESA\_36 contém 25 (vinte e cinco) participações nesse *cluster*. Na base toda a EMPRESA\_2 possui 100 (cem) participações e a EMPRESA\_36 possui 160 (cento e sessenta) participações em licitações.

Uma regra presente no “CLUSTER\_0” que possui o valor de confiança em 100%, com 0.015 de suporte, informa que sempre que as EMPRESA\_20 e a EMPRESA\_63 participam, a EMPRESA\_12 também participa em 100% dos casos. Esse grupo participou de 5 (cinco) licitações e celebrou contrato em 2 (duas) licitações. Das 5 (cinco) licitações que o grupo participou, a EMPRESA\_20 venceu as 5 licitações, a EMPRESA\_63 venceu 4 (quatro) licitações e a EMPRESA\_12 venceu 3 licitações. A EMPRESA\_20 possui 48 (quarenta e oito) participações, a EMPRESA\_63 possui 26 (vinte e seis) participações e a EMPRESA\_12 contém 43 (quarenta e três) participações nesse *cluster*. Na base toda, a EMPRESA\_20, a EMPRESA\_63 e a EMPRESA\_12 possuem respectivamente 320 (trezentos e vinte), 148 (cento e quarenta e oito) e 276 (duzentos e setenta e seis) participações.

Uma regra no “CLUSTER\_2” com valor de suporte 0.045 e confiança de 90% estabeleceu a relação de duas empresas, sendo elas as EMPRESA\_43 e EMPRESA\_32, que contém participações em apenas um órgão, possuem participações individuais em licitações de 20 (vinte) e 88 (oitenta e oito) respectivamente, sendo que, entre essas, elas participaram juntas de 18 (dezoito) licitações e

ganham juntas 9 (nove) licitações. Vale ressaltar que, das 20 (vinte) licitações que a EMPRESA\_43 participou, a EMPRESA\_32 não participou somente de 2 (duas) licitações.

#### 5. CONSIDERAÇÕES FINAIS

O estudo aborda um dos pilares cruciais da Administração Pública: o processo licitatório. Esse instrumento é vital para que órgãos e entidades governamentais conduzam suas operações de maneira transparente e eficiente. No entanto, dada a natureza que envolve recursos públicos, o abuso de agentes tanto públicos quanto privados para obter vantagens pessoais é um fenômeno global.

Esse trabalho corresponde a uma parcela das possibilidades que ciência de dados e mineração de dados podem oferecer, utilizando-se de técnicas adequadas, *insights* de grandes valores podem surgir. As regras descritas acima não significam que de fato existem irregularidades e/ou formação de cartel nas participações e vitórias conjuntas, mas proporciona uma nova visão aos profissionais que fazem as análises dos processos licitatórios com dados de portais abertos como os da CGE/GO, podendo ser o item inicial para uma investigação mais detalhada para constatação da veracidade dos fatos relatados.

Espera-se que com os resultados de técnicas de mineração de dados aplicados no âmbito de dados como os disponibilizados pela CGE/GO um ganho de produtividade e mais rapidez nos processos decorrentes das análises da área responsável por fiscalização possa ser atingido. Espera-se também a possibilidade de realização de novos cruzamentos com os dados armazenados no banco de dados, utilizando-se novos algoritmos e técnicas que possam trazer novos *insights* e visão para todos os órgãos de controle externo que queiram fazer uso desta tecnologia.

Dessa forma, dadas as evidências obtidas nesse estudo, é possível verificar que a utilização de métodos adequados de mineração de dados pode contribuir na identificação de possíveis irregularidades em processos licitatórios em dados disponibilizados pela CGE/GO, apresentando o potencial para o desenvolvimento de sistemas de apoio à decisão e definição de estratégias, gerando informações que permitem maior assertividade, permitindo a eficiência do uso dos recursos públicos.

Para trabalhos futuros, recomenda-se testar a metodologia em outros órgãos de fiscalização, criar um *data warehouse* para armazenamento histórico de dados, desenvolver métodos de melhor escolha das regras geradas e explorar novas abordagens e variáveis.

#### REFERÊNCIAS BIBLIOGRÁFICAS



- [1] M.S.Z. Di Pietro. *Direito Administrativo, Editora Forense*, Brasil, 2019, 755-756.
- [2] A.A. Fraga. Detecção de casos suspeitos de fraudes em licitações realizadas nos municípios da Paraíba: uma aplicação de técnicas de mineração de dados. *Dissertação de Mestrado*, Centro de Ciências Sociais Aplicadas, Universidade Federal da Paraíba, 2017.
- [3] I.H. Witten, E. Frank, M.A. Hall, C.J. Pal. *Data Mining: practical machine learning tools and techniques*, Elsevier, United States of America, 2017, 3-9.
- [4] A.J. Souza, R.B. Mendes, J.R.B. Braga. Laudo de Perícia Criminal (Contábil-Financeira) Nº 2311/2015-SETEC/SR/DPF/PR, 2015.
- [5] C.V.S. Silva, C.G. Ralha. Detecção de cartéis em licitações públicas com agentes de mineração de dados, *Revista Eletrônica de Sistemas de Informação* **10**, 1-19, 2011.
- [6] C.M.M. Moraes. Proposição De Indicadores Para Investigação De Licitações Por Meio De Técnicas De Reconhecimento De Padrões Estatísticos E Mineração De Dados. *Dissertação de Mestrado*, Departamento de Engenharia Elétrica, Universidade de Brasília, 2016.
- [7] T.F. Grilo Júnior. Aplicação de Técnicas de Data Mining para Auxiliar no Processo de Fiscalização no Âmbito do Tribunal de Contas do Estado da Paraíba. *Dissertação de Mestrado*, Centro de Tecnologia, Universidade Federal da Paraíba, 2010.
- [8] R.S. Wazlawick. *Metodologia de Pesquisa para Ciência da Computação*, Elsevier, Brasil, 2014, 21-26.
- [9] J.C.B. Nandi, R.M. Pereira, G. Felipe, K. Madeira, A.C.G. Barbosa, M.C.M. Garcia. O Algoritmo de Associação Frequent Pattern-Growth na Shell Orion Data Mining Engine, *Anais Sulcomp* **7**, 1-10, 2015.
- [10] P. Tan, M. Steinbach, V. Kumar. *Introdução ao Data Mining Mineração de Dados*, Editora Ciência Moderna, 2009, 433-437.
- [11] V. Kotu, B. Deshpande. *Data Science: concepts and practice. 2. ed.* Cambridge: Elsevier, United States of America, 2019, 235-247.
- [12] C.S.C. Arantes, J. L. Beneti, M.M. Pereira, R.G. Loiola, S.L. Vieira. Aplicação do *k-means clustering* no problema de formação de arranjo físico celular. *XXI Simpósio de Engenharia de Produção*, Bauru, 2014.