

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS
ESCOLA DE CIÊNCIAS EXATAS E DA COMPUTAÇÃO
CURSO DE ENGENHARIA DE COMPUTAÇÃO



**DESCOBERTA DE CONHECIMENTO NA ANÁLISE DE LICITAÇÕES NO ESTADO
DE GOIÁS**

MATHEUS ADÃO DE SOUZA E SILVA

GOIÂNIA
2020

MATHEUS ADÃO DE SOUZA E SILVA

DESCOBERTA DE CONHECIMENTO NA ANÁLISE DE LICITAÇÕES NO ESTADO
DE GOIÁS

Trabalho de Conclusão de Curso apresentado à Escola de Ciências Exatas e da Computação, da Pontifícia Universidade Católica de Goiás, como parte dos requisitos para a obtenção do título de Bacharel em Engenharia de Computação.

Orientador: Prof. Dr. Sibelius Lellis Vieira

Banca examinadora:

Prof. Me. Max Gontijo de Oliveira

Prof. Me. Geraldo Valeriano Ribeiro

GOIÂNIA

2020

MATHEUS ADÃO DE SOUZA E SILVA

DESCOBERTA DE CONHECIMENTO NA ANÁLISE DE LICITAÇÕES NO ESTADO
DE GOIÁS

Trabalho de Conclusão de Curso aprovado em sua forma final pela Escola de Ciências Exatas e da Computação, da Pontifícia Universidade Católica de Goiás, para obtenção do título de Bacharel em Engenharia de Computação, em ____/____/____.

Orientador: Prof. Dr. Sibelius Lellis Vieira

Prof. Me. Max Gontijo de Oliveira

Prof. Me. Geraldo Valeriano Ribeiro

Prof^a. Ma. Ludmilla Reis Pinheiro dos Santos
Coordenadora de Trabalho de Conclusão de
Curso

GOIÂNIA
2020

AGRADECIMENTOS

Agradeço em primeiro lugar a Deus por ter me dado forças para realizar este trabalho, à minha família, em especial a minha mãe, meu pai e meu irmão, que me apoiaram e incentivaram em meus estudos. Agradeço também aos colegas e professores do curso pelo convívio e solidariedade.

Agradeço ao professor Dr. Sibelius Lellis Vieira, pela oportunidade e encorajamento ao estudo científico, pelo apoio e incentivo que me foi dedicado, os meus mais sinceros e profundos votos de admiração e respeito.

Por fim, gostaria de agradecer ainda os meus amigos pelo carinho e pela compreensão nos momentos em que a dedicação aos estudos foi única.

A todos que colaboraram direta ou indiretamente para que este trabalho fosse realizado fica registrado os meus mais singelos agradecimentos.

RESUMO

Esta pesquisa tem o propósito de aplicar técnicas de mineração de dados no âmbito da Controladoria Geral do Estado de Goiás com o objetivo de relacionar possíveis indícios de irregularidades em processos licitatórios, auxiliando assim, em demonstrar 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 a redução dos recursos públicos.

Palavras-chave: licitações, cartelização, ciência de dados, simulação, descoberta de conhecimento.

ABSTRACT

This research aims to apply data mining techniques within the scope of the General Controllorship of the State of Goiás in order to relate possible indications of irregularities in bidding processes, thus helping to demonstrate its viability in inspection processes. For a better understanding of this study, a bibliographic review was carried out in order to know the facts and phenomena related to public tenders. Analytical intelligence concepts were explored, addressing business intelligence, data science and data mining techniques. Then the selection of data started, and it was possible to find suitable data available on the transparency portal of the State's Comptroller General. The dataframes related to the organ'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 Dice. The predictive analysis that followed provided the generation of rules associating two or more companies with joint interests and victories, and may present the indication of 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 the reduction of public resources.

Keywords: bids, cartelization, data science, simulation, knowledge discovery.

LISTA DE ILUSTRAÇÕES

FIGURAS

Figura 1. Processo <i>Knowledge Discovery in Databases</i> (KDD).....	28
Figura 2. Método para análise dos dados.....	41
Figura 3. Informações disponíveis na seção sobre as licitações em andamento.....	44
Figura 4. Informações disponíveis na seção sobre as licitações concluídas.....	45
Figura 5. Informações disponíveis na seção sobre os participantes de licitações.....	46
Figura 6. Informações disponíveis na seção sobre os vencedores das licitações.....	47
Figura 7. Organização das informações em 5 (cinco) <i>dataframes</i>	47
Figura 8. Palavras filtradas na descrição do objeto da licitação.....	53
Figura 9. Distribuição das licitações no período de 2016 a 2019.....	55
Figura 10. Número de licitações, valor estimado e valor total por modalidade.....	55
Figura 11. Número de propostas por modalidade.....	56
Figura 12. Porcentagem das propostas por modalidade.....	57
Figura 13. Número de propostas por tipo de disputa.....	57
Figura 14. Número de participantes por modalidade.....	58
Figura 15. Número de vencedores por modalidade.....	59
Figura 16. Número de propostas e número de licitações por ano.....	59
Figura 17. Número de itens de licitações e licitações por modalidade.....	61
Figura 18. Valor estimado e valor adjudicado total por modalidade.....	61
Figura 19. Participações por empresa.....	62
Figura 20. Vitórias por empresa.....	62
Figura 21. Empresas com maiores números de participações em órgãos diferentes.....	63

Figura 22. Empresas com maiores números de vitórias em órgãos diferentes.....	64
Figura 23. Número de itens de licitações e número de licitações por órgão.....	65
Figura 24. Quantidade de participantes e vencedores por órgão.....	65
Figura 25. As 4 (quatro) maiores empresas vencedoras por órgão.....	67
Figura 26. Área de trabalho da ferramenta RapidMiner.....	68
Figura 27. Configurações dos parâmetros do operador <i>Transform log to session</i> ...	69
Figura 28. Configurações do parâmetro do operador FP-Growth.....	70
Figura 29. Configurações do parâmetro do operador <i>Create Association Rules</i>	71
Figura 30. Resultado do processamento dos dados minerados.....	72
Figura 31. Processo para realizar a Clusterização no RapidMiner.....	74
Figura 32. Processo para realizar a Regra de Associação em cada <i>cluster</i> obtido..	76

QUADROS

Quadro 1. Técnicas e algoritmos que podem ser aplicados nas principais tarefas de mineração de dados.....	32
Quadro 2. Variáveis do “DATAFRAME 1 - propostas de licitações”.....	48
Quadro 3. Variáveis do “DATAFRAME 2 – licitações”.....	49
Quadro 4. Variáveis do “DATAFRAME 3 – participantes”.....	50
Quadro 5. Variáveis do “DATAFRAME 4 - participantes nos órgãos”.....	51
Quadro 6. Variáveis do “DATAFRAME 5 – órgãos”.....	51
Quadro 7. Relação de licitações e empresas.....	69
Quadro 8. Matriz esparsa das licitações e empresas.....	70
Quadro 9. Resultados execução do algoritmo FP-Growth para os dois suportes.....	72
Quadro 10. <i>Dataset 2</i> para realizar a Clusterização.....	73

Quadro 11. <i>Dataset 2</i> com a coluna de <i>clusters</i> gerados.....	74
Quadro 12. Quantidade de itens por <i>clusters</i> gerados.....	75
Quadro 13. Execução do algoritmo de associação FP-Growth para os <i>datasets</i> de <i>clusters</i>	76

TABELAS

Tabela 1. Primeira leitura da base de dados.....	35
Tabela 2. Quantidade de licitações totais obtidas por modalidade.....	52
Tabela 3. Quantidade de licitações obtidas para saúde e medicamentos por modalidade.....	53
Tabela 4. Média de participantes por órgão.....	66
Tabela 5. Valor estimado e valor total por órgão.....	66

LISTA DE SIGLAS

API	<i>Application Programming Interface</i>
BI	<i>Business intelligence</i>
BID	Banco Interamericano de Desenvolvimento
CART	<i>Classification and Regression Trees</i>
CGE/GO	Controladoria Geral do Estado de Goiás
CNPJ	Cadastro Nacional da Pessoa Jurídica
CPF	Cadastro de pessoas físicas
CSV	<i>Comma-Separated Values</i>
FIESP	Federação das Indústrias do Estado de São Paulo
FNS	Ferrovias Norte-Sul
FP-tree	<i>Frequent Pattern tree</i>
FTP	<i>File Transfer Protocol</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
IBM	<i>International Business Machines</i>
IPC	Índice de Percepção da Corrupção
KDD	<i>Knowledge Discovery in Databases</i>
KNN	<i>K Nearest Neighbor</i>
MPF	Ministério Público Federal
OECD	<i>Organisation for Economic Co-operation and Development</i>

PDF	<i>Portable Document Format</i>
RDC	Regime Diferenciado de Contratação
SAGRES	Sistema de Acompanhamento da Gestão dos Recursos da Sociedade
SAS	<i>Statistical Analysis System</i>
SEI	Sistema Eletrônico de Informação
SRP	Sistema de Registro de Preços
SVM	<i>Support Vector Machine</i>
TCE/PB	Tribunal de Contas do Estado da Paraíba
TCM/CE	Tribunal de Contas dos Municípios do Estado do Ceará
TCU	Tribunal de Contas da União
TI	Transparência Internacional
XML	<i>eXtensible Markup Language</i>

SUMÁRIO

1. INTRODUÇÃO	13
1.1 Contextualização	13
1.2 Justificativa	14
1.3 Objetivo	16
1.3.1 Objetivo geral	16
1.3.2 Objetivos específicos	16
1.4 Estrutura do trabalho	17
2. REFERENCIAL TEÓRICO	18
2.1 Licitações	18
2.1.1 Processo licitatório	18
2.1.2 Fatores para a prática de conluio/cartel em licitações	21
2.2 Inteligência analítica	25
2.2.1 Inteligência artificial, aprendizado de máquina, <i>business intelligence</i> e ciência de dados	25
2.2.2 Descoberta de conhecimento em bases de dados	27
2.2.3 Métodos de mineração de dados	30
2.2.4 Regra de associação e funcionamento do algoritmo FP-Growth	34
2.2.5 Clusterização e o algoritmo K-means	36
2.2.6 Ferramentas para mineração de dados	37
2.3 Estudos relacionados	38
3. PROCEDIMENTO METODOLÓGICO	40
4. RESULTADOS	43
4.1 Seleção dos dados e pré-processamento	43
4.2 Análise descritiva	54
4.3 Análise preditiva	67
5. CONSIDERAÇÕES FINAIS	79
5.1 Recomendações para trabalhos futuros	81
REFERÊNCIAS	82

1. INTRODUÇÃO

1.1 Contextualização

Através da Constituição Federal de 1988 os brasileiros têm conquistados diversos direitos enfatizando-se o incentivo à participação popular no processo de elaboração e discussão das leis orçamentárias e orçamentos, por meio de instrumentos de fortalecimento da participação do cidadão. Um desses instrumentos é a Lei de Responsabilidade Fiscal, que cita a responsabilidade na gestão fiscal, pressupondo a ação planejada e de forma transparente, prevenindo riscos e corrigindo desvios capazes de afetar o equilíbrio das contas públicas (BRASIL, 2000).

A Lei de Responsabilidade Fiscal reflete diretamente nas licitações e contratos administrativos, porque um dos princípios que ornamenta a administração pública e que a caracteriza, principalmente, é a obrigatoriedade do processo licitatório. Apesar de grande importância para a economia do setor público, as licitações são vulneráveis a ações fraudulentas e à corrupção, haja vista que suas características possibilitam diversas interações entre agentes públicos e privados (CARVALHO, 2010).

No Brasil, a administração pública tem se preocupado cada vez na oferta de serviços públicos com qualidade, disponibilizando ferramentas para controle social e prestação de contas de forma transparente. O uso da licitação pública tem o objetivo de reduzir custos de aquisição de bens públicos e minimizar, sempre que possível, desperdícios de recursos (DI PIETRO, 2019).

Nessa esteira, órgãos e entidades públicos, de todos os entes e poderes, como forma de atender ao princípio da transparência e às determinações legais, desenvolveram formas de divulgar de forma ampla, todos os dados da gestão dos recursos públicos, principalmente por meio dos portais da transparência e portal do cidadão. Sendo assim, a internet se tornou o meio essencial para a divulgação de informações e a oferta de serviços à população (PINHEIRO; MELO, 2016).

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.

A CGE/GO 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; e planejamento e orçamento de todos os órgãos do estado. Dessa forma, a CGE/GO disponibiliza mais um mecanismo de fiscalização a ser utilizado na sua atuação como órgão de controle interno do estado.

Diante do exposto, tem-se a necessidade de pesquisas científicas que forneçam bases teóricas e metodológicas com a finalidade de identificar anomalias para possíveis ações irregulares no mercado de licitações. Entretanto, o crescimento do volume de processos licitatórios específicos em intervalos de tempo mais curtos impõe vários desafios no âmbito dos órgãos de controle, não sendo possível realizar esse monitoramento de forma manual, necessitando então de ferramentas computacionais.

Encontram-se disponíveis técnicas computacionais que podem auxiliar na produção de conhecimento a partir de grandes volumes de dados, como são as bases de dados de órgãos públicos. A título de exemplo, técnicas baseadas em inteligência artificial, aprendizado de máquina e ciência de dados, são amplamente utilizadas para esse fim por diversas empresas, com o objetivo de identificar padrões ou informações relevantes para os negócios.

Nesse trabalho será abordado a área de ciência de dados, mais especificamente mineração de dados, para lidar com grandes volumes de informação. A utilização dessa técnica tem-se mostrado de grande valia na obtenção de informações potencialmente úteis. Este estudo pretende contribuir para o campo de investigação ao avançar em relação aos principais trabalhos de referência, a saber Morais (2016), Silva e Ralha (2011) e Grilo Júnior (2010).

1.2 Justificativa

A operação Lava Jato é a maior investigação sobre corrupção conduzida até hoje no Brasil. Ela teve início no Paraná, em 17 de março de 2014, unificando quatro ações que apuravam redes operadas por doleiros que praticavam crimes financeiros com recursos públicos. Desde então, a operação descobriu a existência de um vasto esquema de corrupção na Petrobras, envolvendo políticos de vários partidos e algumas das maiores empresas públicas e privadas do país, principalmente

empreiteiras. Os desdobramentos não ficaram restritos à estatal e às construtoras (OPERAÇÃO, 2020).

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” (SOUZA; MENDES; BRAGA, 2015).

Especificamente em relação ao Laudo de Perícia Criminal n.º 1342/2015-SETEC/SR/DPF/PR, concluiu-se que a probabilidade das três licitantes desclassificadas (UTC/Engevix, Odebrecht/OAS e MPE) terem apresentado propostas naqueles valores (superiores à proposta classificada) era de 0,0124%, ou seja, haveria necessidade de ocorrerem mais dez mil licitações para esse fato atípico novamente se repetir. Tal resultado é considerado consistente com a hipótese de que as citadas empresas tenham apresentado propostas com a finalidade de “dar cobertura” (SOUZA; MENDES; BRAGA, 2015).

O Tribunal de Contas da União (TCU) apresentou em relatório, estatisticamente, a existência do preço monopolístico praticado em contratos da operação Lava Jato, entre o ano de 2000 até 2014, estimando também a diferença entre o que a Estatal efetivamente pagou e o que teria dispendido se as contratações tivessem sido realizadas sem a existência do cartel. A atuação dos arranjos de preço reduziu em aproximadamente 17% o valor do desconto que seria ofertado no caso de um cenário competitivo. Em dimensão comparativa ao aludido percentual, foi verificado que o impacto da atuação do cartel é muito maior do que percentual pago de propina, de 3 a 5% (BRASIL, 2016).

Analizados apenas o conjunto de contratos com indícios e provas de condutas irregulares na Diretoria de Abastecimento, no período entre 2002 e 2015, o valor do dano encontrado foi da ordem de R\$ 5,7 bilhões e de R\$ 8,9 bilhões ao se aplicar um reajuste inflacionário. Ampliando para além da Diretoria de Abastecimento, foi

estimado que o prejuízo da atuação do Cartel na Petrobras poderia chegar a R\$ 29 bilhões (BRASIL, 2016).

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 consequente economia para os cofres públicos, esta pesquisa justifica-se por relacionar os possíveis indícios de fraudes em processos licitatórios com a implantação de técnicas de mineração de dados em órgãos da administração pública, focando no processo analítico das operações eletronicamente armazenadas, o acúmulo de informação e o seu potencial para análise e tomada de decisão. 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?

1.3 Objetivo

1.3.1 Objetivo geral

Aplicar técnicas de mineração de dados no âmbito dos dados disponibilizados pela Controladoria Geral do Estado de Goiás com vistas a relacionar possíveis indícios de irregularidades em processos licitatórios.

1.3.2 Objetivos específicos

Para se atingir o objetivo geral, propõem-se os seguintes objetivos específicos:

- Desenvolver uma investigação através de uma pesquisa exploratória para conhecer os fatos e fenômenos relacionados a licitações públicas;
- Explorar os conceitos de *data science* e técnicas de mineração de dados;
- 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;
- 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.

1.4 Estrutura do trabalho

Esse estudo apresenta-se estruturado em 5 (cinco) capítulos, sendo o capítulo 1 (um) referente a esta introdução. O capítulo 2 (dois) apresenta o referencial teórico, abordando uma revisão bibliográfica sobre conceitos e procedimentos das licitações, mineração de dados, técnicas, algoritmos e trabalhos correlatos. O capítulo 3 (três) descreve o procedimento metodológico utilizado nesse estudo. O capítulo 4 (quatro) descreve os resultados obtidos para a seleção dos dados e pré-processamento, análise descritiva e análise preditiva. O capítulo 5 (cinco) apresenta as considerações finais e sugestões para trabalhos futuros.

2. REFERENCIAL TEÓRICO

2.1 Licitações

Esta seção tem como objetivo apresentar conceitos sobre licitações, propósito e funcionamento do processo licitatório, abordando as modalidades, tipos, problemas com esse ato e características que são comuns na prática de conluíus.

2.1.1 Processo licitatório

Com o objetivo de reduzir custos de aquisição de bens públicos e minimizar desperdícios de recursos, a administração pública, inclusive no Brasil, tem feito uso do instrumento de licitação. A licitação é um procedimento administrativo pelo qual um ente público, no exercício da função administrativa, abre a todos os interessados, que se sujeitem às condições fixadas no instrumento convocatório, a possibilidade de formularem propostas dentre as quais selecionará e aceitará a mais conveniente para a celebração de contrato (DI PIETRO, 2019).

Pelo fato de licitação ser um ato público, não pode ser tratada de maneira sigilosa. A sociedade deve ter acesso aos procedimentos referentes a uma licitação. O processo licitatório deve afastar qualquer suspeita de favorecimento e garantir que o dinheiro público seja utilizado com cautela e eficiência. A licitação é a forma mais clara de se atender aos princípios das atividades da administração pública (SOUZA, 1997).

As licitações possuem algumas modalidades, que são as formas de realização do procedimento licitatório que pretendem ajustar-se às peculiaridades de cada tipo de negócio administrativo. Encontra-se disponível 6 (seis) modalidades de licitações, são elas: concorrência; tomada de preços; convite; concurso; leilão; pregão.

Segundo as Leis nº 8.666 e 10.520 (Brasil, 1993, 2002), a concorrência é a modalidade de licitação entre quaisquer interessados que, na fase inicial de habilitação preliminar, comprovem ter os requisitos mínimos de qualificação exigidos no edital para a execução de seu objeto. Tomada de preços refere-se aos interessados devidamente cadastrados ou que atenderem a todas as condições exigidas para o cadastramento até o terceiro dia anterior à data do recebimento das propostas, observada a necessária qualificação. O pregão é a modalidade de licitação destinada à aquisição de bens e serviços comuns.

Ainda conforme as Leis nº 8.666 e 10.520 (Brasil, 1993, 2002), o convite é a modalidade de licitação entre os interessados do ramo pertinente ao seu objeto, cadastrados ou não, escolhidos e convidados em número mínimo de 3 (três) pela unidade administrativa, a qual afixará, em local apropriado, cópia do instrumento convocatório e o estenderá aos demais cadastrados na correspondente especialidade que manifestarem seu interesse com antecedência de até 24 horas da apresentação das propostas. O concurso possibilita a quaisquer interessados para a escolha de trabalho técnico, científico ou artístico, mediante a instituição de prêmios ou remuneração aos vencedores, conforme critérios constantes de edital. O leilão é destinado para a venda de bens móveis inservíveis para a administração ou de produtos legalmente apreendidos ou penhorados, ou para a alienação de bens.

O processo licitatório possui ainda uma classificação em relação aos tipos, que se referem aos critérios de julgamento das propostas a ser observado no procedimento. São 4 (quatro) tipos existentes: menor preço; melhor técnica; técnica e preço; maior lance ou oferta.

O tipo de licitação menor preço alude-se ao critério de seleção da proposta mais vantajosa para a administração determinar que será vencedor o licitante que apresentar a proposta de acordo com as especificações do edital e ofertar o menor preço. Melhor Técnica menciona o fator preponderante para a escolha do vencedor não sendo o preço, mas critérios de qualidade e especificidade indispensáveis para o alcance do interesse público (BRASIL, 1993).

Técnica e preço relaciona os fatores “técnica” e “preço” e são considerados de acordo com proporções previamente consignadas em edital. Não há, pois, prevalência de qualquer dos fatores, mas um procedimento objetivo de aferição da média ponderada das valorizações das propostas técnicas e de preço. Maior Lance ou Oferta é considerada a melhor proposta aquela que corresponder ao maior valor nominal ofertado. O critério de julgamento deve ser utilizado nos casos de alienação de bens ou concessão de direito real de uso (BRASIL, 1993).

Conforme regulamenta o art. 37, inciso XXI, da Constituição Federal, que institui normas para licitações e contratos da Administração Pública, o art 3º da Lei nº 8.666 de 21 de junho de 1993 relata o seguinte:

A licitação destina-se a garantir a observância do princípio constitucional da isonomia, a seleção da proposta mais vantajosa para a administração e a promoção do desenvolvimento nacional sustentável e será processada e julgada em estrita conformidade com os princípios básicos da legalidade, da impessoalidade, da moralidade, da igualdade, da publicidade, da probidade administrativa, da vinculação ao instrumento convocatório, do julgamento objetivo e dos que lhes são correlatos (BRASIL, 1993).

A igualdade de tratamento entre os possíveis interessados é o ponto central da licitação, é condição indispensável da existência de competição real, efetiva e concreta. Entretanto, a aquisição de produtos e serviços por parte do setor público tem sido, alguma das maneiras de corrupção no Brasil, ocasionando uma perda significativa do gerenciamento dos recursos públicos, prejudicando assim toda sociedade (CASTRO, 2010).

A *Organisation for Economic Co-operation and Development* (OECD), conceitua corrupção como o abuso de agentes públicos e privados para obtenção de vantagens pessoais, aludindo não apenas ao recebimento de propina, mas incluindo a menção ao nepotismo, à fraude e à captura estatal. Segundo a organização, 57% do suborno estrangeiro envolve compras públicas (OECD, 2020).

A organização não-governamental Transparência Internacional (TI) entidade dedicada ao combate à corrupção, também menciona que, a corrupção está relacionada ao abuso de poder visando benefícios pessoais. O Índice de Percepção da Corrupção (IPC) é o principal indicador de corrupção no setor público do mundo, este foi produzido desde 1995 pela organização, o índice avalia 180 países e territórios e os avalia em uma escala na qual para o valor 0 (zero), o país é percebido como altamente corrupto e o valor 100 (cem) significa que o país é percebido como muito íntegro (ÍNDICE, 2019).

Segundo a Transparência Internacional, o Brasil no ano de 2019, manteve-se no pior patamar da série histórica do Índice de Percepção da Corrupção, com apenas 35 (trinta e cinco) pontos, ficando o segundo ano seguido com a pior nota da série histórica. Os 35 (trinta e cinco) pontos da nota brasileira equivalem ao valor mais baixo desde 2012, ano em que o índice passou por alteração metodológica e passou a permitir a leitura em série histórica (ÍNDICE, 2019)

O custo com a corrupção alcançou valores extraordinários. A Federação das Indústrias do Estado de São Paulo (FIESP) realizou um estudo no ano de 2010

afirmando que o custo médio da corrupção no Brasil é estimado entre 1,38% a 2,3% do PIB, isto é, aproximadamente, de R\$ 96,6 bilhões a R\$ 161 bilhões em 2019. Em Goiás, em relação ao último PIB divulgado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), em 2017, os números são aproximadamente, de R\$ 2,7 bilhões a R\$ 4,4 bilhões (CORRUPÇÃO, 2010).

Deve-se ter em mente que os países em que o custo corrupção existe tendem a receber menor quantidade de investimentos, reforçando o problema social. Segundo a Transparência Internacional, há outros efeitos negativos provenientes de práticas relacionadas à corrupção: prejudica-se a inovação tecnológica porque as empresas corruptas não têm interesse em fazê-la ou não precisam fazê-la para se manter no mercado, enquanto as demais não se sentem estimuladas a alocar recursos para essa finalidade; e destroem-se os empregos porque afeta-se o ambiente competitivo provocando o fim de empresas idôneas (FORTINI; MOTTA, 2016).

2.1.2 Fatores para a prática de conluio/cartel em licitações

Em licitação pública um dos grandes motivos relacionado ao desvio do dinheiro público é a formação de cartel, que consiste em um grupo de ofertadores que fizeram um acordo explícito para limitar competição entre eles em benefício próprio. O objetivo é atuar como se fosse um monopólio, mas enquanto no monopólio a decisão cabe apenas a um ente, no cartel consiste em uma associação voluntária de tomadores de decisão, cada qual consciente que o seu lucro depende do comportamento de todos os ofertadores da indústria. Portanto envolve muito mais que uma simples definição do nível de preço e quantidade máxima do lucro (CAMPOS, 2008).

Braga (2015) conceitua cartéis como acordos entre concorrentes para alcançar algum tipo de benefício coletivo em detrimento da concorrência. O senso comum é o de que cartéis são ajustes entre concorrentes para fixação de preços acima do preço de equilíbrio verificado no mercado. No entanto, esse mito deve ser imediatamente quebrado, já que essas empresas não visam ganhos unicamente via fixação de preços.

A OECD disponibilizou diretrizes para combater o conluio entre concorrentes em contratações públicas, segundo a organização, as formais mais comuns de conluio são classificadas em: propostas fictícias ou de cobertura; supressão de propostas; propostas rotativas ou rodízio; e divisão de mercado (DIRETRIZES, 2009).

Propostas fictícias ou de cobertura ocorrem quando indivíduos ou empresas combinam submeter propostas em que: um dos concorrentes aceita apresentar uma proposta mais elevada do que a proposta do candidato escolhido; um concorrente apresenta uma proposta que já sabe que é muito elevada para ser aceita; um concorrente apresenta uma proposta que contém condições específicas que sabe que serão inaceitáveis para o comprador. As propostas fictícias são concebidas para dar a aparência de uma concorrência genuína entre os licitantes (DIRETRIZES, 2009).

Supressão de propostas envolvem acordos entre os concorrentes nos quais uma ou mais empresas estipulam abster-se de concorrer ou retiram uma proposta previamente apresentada para que a proposta do concorrente escolhido seja aceita. Em propostas rotativas ou rodízio, as empresas conspiradoras continuam a concorrer, mas combinam apresentar alternadamente a proposta vencedora (DIRETRIZES, 2009).

Para divisão do mercado os concorrentes definem os contornos do mercado e acordam em não concorrer para determinados clientes ou em áreas geográficas específicas. As empresas concorrentes podem, por exemplo, atribuir clientes específicos ou tipos de clientes a várias empresas, para que os demais concorrentes não apresentem propostas para contratos ofertados por essas classes de potenciais clientes (DIRETRIZES, 2009).

A identificação de cartel não é uma tarefa simples de se realizar, e podem incluir muitas empresas da área e os clientes raramente estão em posição de detectar a existência de um cartel, principalmente em órgãos públicos, em que o número de colaboradores destinados para esse tipo de atividade geralmente não é suficiente. Como as licitações envolvem dinheiro público, ou seja, da sociedade, é necessário auxiliar e fiscalizar a formação de cartéis por vários meios e instrumentos que se encontram disponíveis.

Uma característica importante, que evidencia a existência de um possível cartel é a baixa taxa de utilização de capacidade. Isso quer dizer que as empresas produzem uma quantidade abaixo do que são capazes e vendem o produto a preços mais altos. Essa quantidade produzida e os preços dos produtos são combinados entre os participantes (TATSCH, 2012).

A formação do cartel se torna mais comum quando são poucas empresas que ofertam o mesmo produto ou serviço, em razão de ser mais fácil sua manutenção, trocas de informações, fiscalização, possível consciência de todos os empresários que

forneçam o produto e menor probabilidade de que algum dos participantes descumpra os acordos, garantindo, assim, maiores chances de sucesso do cartel (CARVALHO, 2018).

Ainda segundo Carvalho (2018) uma característica relevante trata-se das barreiras à entrada de novas empresas no setor. O cartel por natureza é instável, caso a entrada de novas empresas que ofereçam o mesmo produto seja constante ou crescente será extremamente difícil a manutenção desse cartel.

A homogeneidade dos produtos ou serviços facilita a constituição e manutenção do cartel, tendo em vista que possuindo produtos similares em qualidade e custos de produção numa fixação de preços será mais simples a definição de um valor base. Entretanto, caso os produtos ou serviços sejam heterogêneos não significa que não há cartel, pois, existem alternativas para esse tipo de situação (CARVALHO, 2018).

Quando há pouca ou nenhuma tecnologia no mercado que produz um determinado produto ou realiza um serviço, ou, suas tecnologias já estão consolidadas, sendo comum a aplicação da tecnologia existente, se torna mais fácil o surgimento e coordenação do cartel tendo em vista que, nesses casos, geralmente os produtos serão homogêneos (CARVALHO, 2018).

Existem algumas características comuns que ajudam a manter o conluio, como, pequeno número de empresas, quanto menor o número de fornecedores, mais fácil será para estes chegarem a um acordo quanto às propostas. O nível reduzido ou nulo de entradas no mercado é outro fator, quando empresas entraram recentemente no mercado ou quando for pouco provável que tal aconteça porque a entrada no mercado não é um processo simples, as empresas que já operam no mercado estão protegidas da pressão competitiva de potenciais novos concorrentes, fazendo com que os cartéis continuem (DIRETRIZES, 2009).

Condições do mercado é um fator considerável, pois mudanças significativas nas condições da procura ou oferta tendem a desestabilizar acordos de conluio em curso. Um fluxo de procura do setor público que seja constante e previsível tende a aumentar o risco de conluio. Licitações frequentes ajudam os membros de um acordo de cartel a distribuir entre si os contratos. Além disso, os membros de um cartel podem punir a empresa que não cumprir o acordo ficando com as propostas que originalmente lhe estavam atribuídas (DIRETRIZES, 2009).

É também um fator relacionado ao nível reduzido ou praticamente nulo de possibilidades, que faz com que os indivíduos e empresas que pretendem atuar em conluio fiquem mais seguros por saber que o responsável pelas aquisições tem pouca ou nenhuma alternativa e assim os esforços para aumentar os preços têm mais probabilidades de serem bem-sucedidos (DIRETRIZES, 2009).

Segundo a OECD no documento “Diretrizes para combater o conluio entre concorrentes nas contratações públicas”, existem alguns padrões para serem observados na fase de apresentação de propostas que podem estar em desacordo com as regras de um mercado competitivo, são elas: a proposta mais baixa é frequentemente do mesmo fornecedor; existe uma distribuição geográfica das propostas vencedoras; os fornecedores habituais não apresentam propostas para um processo de contratação no qual seria de esperar que o fizessem, mas continuam a concorrer em outros processos (DIRETRIZES,2009).

Ainda conforme o documento supracitado tem-se: alguns fornecedores retiram-se inesperadamente do concurso ou da licitação; determinadas empresas sempre apresentam propostas mas nunca são bem sucedidas; as empresas parecem ganhar os concursos de forma alternada; duas ou mais empresas apresentam uma proposta conjunta apesar de possuir capacidade para apresentar uma proposta independente; o concorrente vencedor subcontrata reiteradamente os demais concorrentes; o concorrente vencedor não aceita a adjudicação do contrato, vindo posteriormente a descobrir-se que foi subcontratado; os concorrentes encontram-se socialmente e têm reuniões pouco antes da apresentação de propostas (DIRETRIZES,2009).

Em relação aos preços, o documento mencionado informa ainda alguns tópicos interessantes, como por exemplo: os concorrentes aumentam os preços ou as faixas de preços de forma súbita e idêntica, situação que não pode ser explicada pelo aumento dos custos; descontos ou abatimentos previsíveis desaparecem inesperadamente; apresentação de preços idênticos; grande diferença entre o preço de uma proposta adjudicada e as restantes; a proposta de um determinado fornecedor para um contrato em particular é muito mais elevada do que a proposta do mesmo fornecedor para um contrato semelhante (DIRETRIZES,2009).

Os padrões e fatores observados, não são por si só fatos que configuram a existência de um cartel, porém podem ser o suficiente para começar uma investigação, que deve ser analisada de forma individual e de forma comparativa, em que pode se identificar a existência de um rodízio de propostas, por exemplo. É

necessária uma investigação de forma mais completa, pois as empresas podem apresentar justificativas plausíveis para a possível prática suspeita.

2.2 Inteligência Analítica

Esta seção tem como objetivo apresentar a relação de inteligência artificial, aprendizado de máquina, *business intelligence* e ciência de dados. Discorre-se sobre a descoberta do conhecimento em bases de dados, apresentar conceitos sobre *data mining* ou mineração de dados, técnicas e algoritmos para cada tipo de tarefa da mineração de dados.

2.2.1 Inteligência artificial, aprendizado de máquina, *business intelligence* e ciência de dados

Inteligência artificial, aprendizado de máquina, *business intelligence* e ciência de dados estão relacionados entre si. Essas técnicas são muitas vezes utilizadas alternadamente e confundidos um com o outro na mídia e comunicação corporativa. No entanto, são diferentes, dependendo do contexto. A inteligência artificial é sobre a condição de máquinas exercerem a capacidade de imitar o comportamento humano, funções particularmente cognitivas. A título de exemplo, pode-se citar: reconhecimento facial; condução automatizada; e triagem de correio com base no código postal (KOTU; DESHPANDE, 2019).

Ainda segundo Kotu e Deshpande (2019), aprendizado de máquina pode ser considerado um subcampo ou uma das ferramentas de inteligência artificial, fornecendo máquinas com a capacidade de aprendizagem, em que essas adquirem experiência a partir de dados. Ciência de dados é a aplicação de negócios de aprendizagem de máquina, inteligência artificial, e outros campos quantitativos como estatísticas, visualização, matemática e processo de mineração. É um campo interdisciplinar que extrai valor a partir de dados, com o objetivo de prever alguma informação, ou seja, analisar o que pode acontecer e o que pode ser feito.

Business Intelligence (BI) é definido como um conjunto de técnicas e ferramentas para a aquisição e transformação de dados brutos em informações significativas e úteis para fins de análise de negócios. As tecnologias de BI são capazes de lidar com grandes quantidades de dados estruturados e às vezes não estruturados para ajudar a identificar, desenvolver e criar oportunidades de negócios

estratégicos. O objetivo do BI é permitir a fácil interpretação de dados e realizar uma análise descritiva, para entender o que aconteceu e por que aconteceu (BENTLEY, 2017).

Para começar a gerar *insights*, ou seja, soluções, é preciso entender alguns conceitos, iniciando pela definição de ciclo da informação que é composto por 3 (três) principais itens: dados; informação; e conhecimento. Dados são fatos, imagens ou sons que podem ou não ser pertinentes ou úteis para uma tarefa particular, a título de exemplo, planilhas, banco de dados e internet. Informação são dados que foram organizados e processados de forma que tenham uma significância, ou seja, conjunto de dados coletados, transformando dados brutos em informação útil e aplicável a uma ou mais situações, como por exemplo gráficos e *dashboards*. Conhecimento é uma combinação de instintos, ideias, regras e procedimentos que guiam ações e decisões para auxiliar a tomada de decisão (ANTONELLI, 2009).

Um importante termo quando se trabalha com dados é o *Analytics* ou Analítica que consiste na descoberta, interpretação e comunicação de padrões significativos em dados. Especialmente valiosa em áreas ricas com informações registradas, a analítica depende da aplicação simultânea de estatísticas, programação de computadores e pesquisa operacional para quantificar o desempenho. A análise geralmente favorece a visualização de dados para comunicar resultados (BENTLEY, 2017).

Descriptive analytics ou análise descritiva é a interpretação de dados históricos para entender melhor as mudanças que ocorreram em um negócio. A análise descritiva descreve o uso de uma variedade de dados históricos para fazer comparações, fornecendo entradas para geração de padrões. Alguns exemplos para a análise descritiva são, mudanças de preços ano a ano, crescimento de vendas mês a mês, número de usuários ou receita total por assinante. Todas essas medidas descrevem o que ocorreu em uma organização durante um determinado período (FRANKENFIELD, 2019).

A análise descritiva usa os dados brutos e os analisa para tirar conclusões que sejam úteis e compreensíveis por gerentes, investidores e *stakeholders*. Esse tipo de análise utiliza uma gama completa de dados para fornecer uma imagem precisa do que aconteceu em uma empresa e como isso difere de outros períodos comparáveis. Essas métricas de desempenho podem ser usadas para sinalizar áreas de força e fraqueza para informar as estratégias de gerenciamento em um formato

fácil de entender. Essa análise se tornou fundamental. Entretanto, para aperfeiçoar a descoberta de conhecimento, mais esforço está sendo direcionado para campos mais novos de analítica, como a análise preditiva.

Predictive analytics ou análise preditiva abrange uma variedade de técnicas estatísticas de modelagem preditiva, aprendizado de máquina e mineração de dados que analisam fatos atuais e históricos para fazer previsões sobre eventos futuros ou desconhecidos. Nos negócios, os modelos preditivos exploram padrões encontrados em dados históricos e transacionais para identificar riscos e oportunidades. Os modelos capturam relacionamentos entre muitos fatores para permitir a avaliação do risco ou potencial associado a um determinado conjunto de condições, orientando a tomada de decisão para transações candidatas (BENTLEY, 2017).

A análise preditiva é definida com uma área de mineração de dados que lida com a extração de informações de dados e com o uso delas para prever tendências e padrões de comportamento. Frequentemente, o evento desconhecido de interesse está no futuro, mas a análise preditiva pode ser aplicada a qualquer tipo de incógnita, seja no passado, presente ou futuro. Por exemplo, identificar suspeitos depois que um crime foi cometido ou fraude de cartão de crédito conforme ela ocorre (BENTLEY, 2017).

Análise preditiva é baseado na captura de relacionamentos entre variáveis explicativas e as variáveis previstas de ocorrências passadas e em explorá-las para prever o resultado desconhecido. A precisão e usabilidade dos resultados contém uma dependência grande do nível de análise dos dados e da qualidade das suposições criadas.

2.2.2 Descoberta de conhecimento em bases de dados

Com o grande número de informações que são geradas e armazenadas atualmente, a análise dessas informações de forma manual torna-se inviável, necessitando do auxílio de mecanismos computacionais mais automáticos e inteligentes para que se possa realizar análise dos dados e auxiliar principalmente na tomada de decisão dentro do ambiente corporativo.

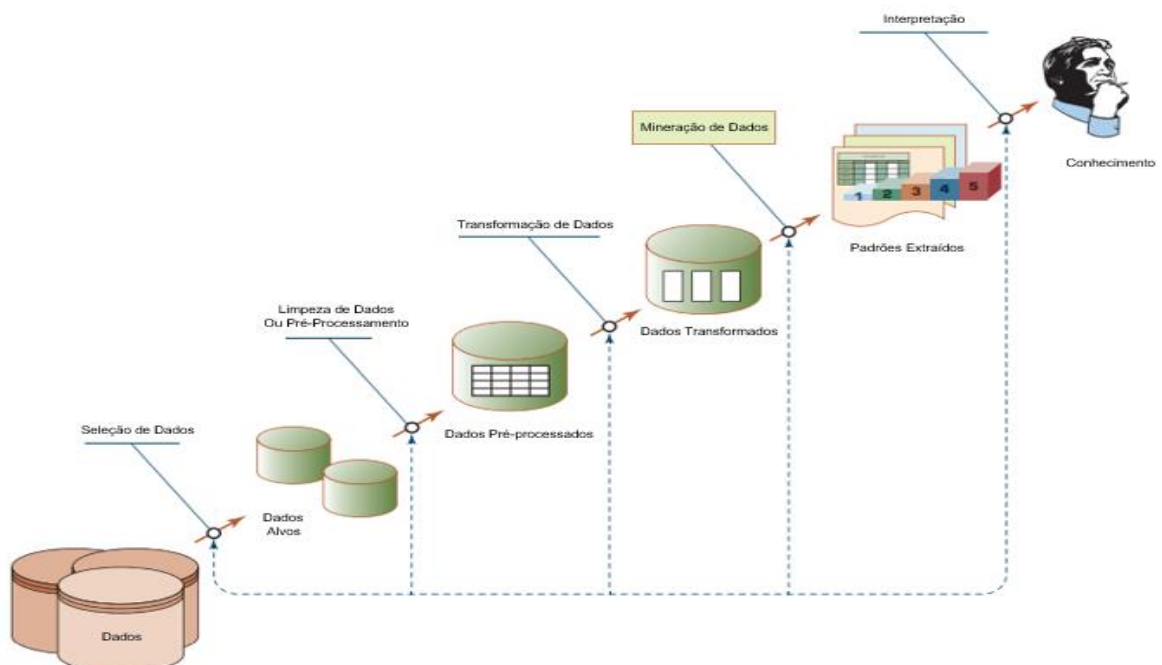
Dessa forma surgem algumas ferramentas computacionais, de maneira a compreender como processar certa quantidade de dados e, o mais importante, como analisar todo esse volume. *Knowledge Discovery in Databases* (KDD), ou descoberta

de conhecimento em bancos de dados, em português, é o processo não trivial de identificar em dados novos padrões que sejam válidos, que antes eram desconhecidos, e que são potencialmente úteis e compreensivos, possibilitando assim melhor entendimento de um problema ou procedimento de tomada de decisão (SILVA, 2002).

O processo do KDD exige três atores de distintas habilidades: usuário; especialista do domínio; analista de dados. O usuário é o demandante do trabalho, que irá potencialmente desfrutar os resultados obtidos. O especialista do domínio é quem conhece o tema que será estudado, normalmente um pesquisador ou profissional com larga experiência. O analista de dados, por sua vez, é quem deve executar o processo de verificação ou anulação das hipóteses criadas pelo especialista do domínio, que, por sua vez, irá entrar no ciclo de reformular as hipóteses para que sejam novamente testadas, seguindo uma “trajetória em espiral rumo à finalização do processo” (SOARES JUNIOR; QUINTELLA, 2005).

O KDD é uma sequência iterativa de etapas, que são ilustrados, na Figura 1.

Figura 1. Processo *Knowledge Discovery in Databases* (KDD)



Fonte: Adaptado de Sharda, Delen e Turban (2018).

Segundo Fayyad *et al.* (1996), o processo KDD refere-se ao processo de descoberta de conhecimentos úteis a partir de dados, composto por 5 (cinco) etapas,

conforme Figura 1, o processo de mineração de dados está dentro do KDD e refere-se à quarta etapa deste processo. Estas etapas são distribuídas em 3 (três) principais grupos: pré-processamento (seleção de dados, limpeza dos dados e tratamento de dados); processamento (mineração de dados); pós-processamento (interpretação).

A primeira etapa, seleção de dados, é a etapa dedicada a obtenção dos dados. Esta etapa pode-se tornar muito complexa, visto que os dados podem vir de diversas fontes, como planilhas e sistemas legados, além de possuírem os mais diversos formatos (FAYYAD *et al.*, 1996).

A segunda etapa, limpeza dos dados, tem como objetivo remover os dados que não são adequados, e baseada em algoritmos específicos, verifica a existência de dados que não estão padronizados e ausência de valores para determinadas variáveis, ou seja, registro com dados incompletos seja por falhas no processo de seleção ou de revisão, dessa forma, recomenda atribuir um valor padrão ou descartá-lo (FAYYAD *et al.*, 1996).

A etapa para transformação dos dados é dedicada para transformar os dados existentes em novos dados apropriadas para o processo de mineração de dados, de acordo com a técnica a ser utilizada. Na etapa mineração de dados pode utilizar ferramentas que, através de técnicas, e de algoritmos, permite-se buscar em uma determinada base de dados padrões não antes conhecidos (FAYYAD *et al.*, 1996).

A última etapa do processo KDD refere-se à interpretação dos dados, e nesta verifica-se o resultado obtido, analisando a qualidade do conhecimento descoberto, buscando identificar se a solução encontrada atende ao problema originalmente proposto para a realização do processo (FAYYAD *et al.*, 1996).

De acordo com Grilo Júnior (2010) os pesquisadores desta área do conhecimento procuram produzir técnicas, metodologias e instrumentos que apoiem a descoberta de conhecimento não conhecido anteriormente e acrescentem informações a um domínio de aplicação específica. A extração de conhecimento de bases de dados é uma área multidisciplinar que visa incorporar e relacionar técnicas adotadas em diversas áreas do conhecimento.

Através de vários campos de pesquisa, como ciência de dados, aprendizado de máquina, reconhecimento de padrões, banco de dados, estatística, e inteligência artificial, o processo de descoberta de conhecimento em bases de dados evoluiu, e continua evoluindo. Conseqüentemente, as técnicas utilizadas em KDD não devem ser vistas como substitutas de outros métodos e áreas de análise, mas de forma que

possam trabalhar juntas com o objetivo de melhorar os resultados das explorações feitas (FAYYAD *et al.*, 1996).

2.2.3 Métodos de mineração de dados

A mineração de dados está inserida no processo de descoberta de conhecimento em bases de dados. Esta é definida como o processo de descoberta de padrões nos dados, e esse processo deve ocorrer de forma automática ou semiautomático. Os padrões descobertos devem ser significativos, pois levam a alguma vantagem, como uma vantagem para a economia. Os dados estão invariavelmente presentes em quantidades substanciais (WITTEN *et al.*, 2017).

Segundo Cabena *et al.* (1998), mineração de dados é a ferramenta de extrair informação, previamente desconhecida e de máxima abrangência a partir de bases de dados, para ser utilizado na tomada de decisão. Han e Kamber (2000) definem mineração de dados como uma etapa na descoberta do conhecimento em bancos de dados que consiste no processo de analisar grandes volumes de dados sob diferentes perspectivas, a fim de descobrir informações úteis que normalmente não estão sendo visíveis.

A resolução de um problema utilizando mineração de dados depende de etapas de pré-processamento e pós-processamento. O pré-processamento corresponde a transformar dados brutos em um formato apropriado para realizar a mineração. As etapas de pós processamento estão relacionadas à avaliação do modelo de mineração de dados criado, de forma que seja assegurado que somente resultados válidos e úteis sejam incorporados aos sistemas de produção (MOURÃO, 2018).

Segundo Batista (2003) o processo de mineração de dados pode ser dividido em 4 (quatro) etapas, que são elas: escolha da tarefa de mineração de dados; escolha da técnica de mineração de dados; escolha do algoritmo a ser utilizado; aplicação do processo de mineração de dados.

Uma tarefa em mineração de dados consiste na especificação do que pretende buscar nos dados, que tipo de regularidades ou categoria de padrões deseja encontrar. Já a técnica de mineração consiste na especificação de métodos que garantam como descobrir os padrões que são relevantes. E o algoritmo é o procedimento passo a passo para determinar a solução do problema (SILVA; RALHA, 2011).

As tarefas de mineração de dados são divididas principalmente em descritivas ou preditivas. As descritivas buscam encontrar padrões para descrever dados, enquanto as preditivas utilizam algumas variáveis para prever valores não conhecidos. As tarefas mais utilizadas são: classificação; *clusterização*; regra de associação; estimativa; desvios.

A classificação é considerada a mais comum de todas as tarefas de mineração de dados. O objetivo da classificação é analisar os dados históricos armazenados em um banco de dados e gerar automaticamente um modelo que possa prever o comportamento futuro. Esse modelo induzido consiste em generalizações sobre os registros de um conjunto de dados de treinamento, que ajudam a distinguir classes predefinidas. Como exemplo dessa tarefa podemos citar a separação de clientes como bons ou mal pagadores, a partir do histórico de pagamento (SHARDA; DELEN; TURBAN, 2018).

A tarefa de clusterização ou segmentação é um processo de divisão de um grupo heterogêneo em vários subgrupos mais homogêneos. Neste processo não existem classes pré-definidas e os dados são agrupados de acordo com suas características próprias, sendo essa a principal diferença da tarefa de classificação. Um exemplo para essa tarefa é a aplicação de segmentação dos clientes de um supermercado conforme comportamentos de compra (QUEIROZ, 2016).

A regra de associação consiste em descobrir relações fortes entre determinados atributos. Essa tarefa tem como objetivo detectar padrões em forma de regras que associam valores de atributos em um determinado conjunto de dados. Em um determinado banco, 50% dos clientes que possuem cheque especial também adquirem cartão de crédito, é um exemplo para Regra de Associação (SILVA; RALHA, 2011).

A tarefa de estimativa tem como objetivo prever uma função desconhecida cuja saída tem um domínio de valores reais. Essa tarefa é teoricamente similar à tarefa de classificação, entretanto, a principal diferença é que o atributo a ser encontrado é contínuo ao contrário da classificação, em que o atributo é discreto. Como exemplo para essa tarefa, pode-se estimar o número de filhos em uma família, ou estimar a probabilidade de chuva em uma determinada região (QUEIROZ, 2016).

A tarefa desvios consiste em identificar registros de conjunto de dados cujas características não atendam aos padrões considerados normais no contexto, denominados como valores atípicos. Como exemplo para essa tarefa, considere um

banco de dados com informações sobre compras de clientes no cartão de crédito, essa tarefa pode ser utilizada para buscar compras cujas características não são comuns das compras realizadas pelo dono do cartão (GOLDSCHMIDT; PASSOS; BEZERRA, 2015).

Há diversas técnicas de mineração de dados, e a técnica a ser utilizada é determinada a partir do tipo de informação que se pretende determinar, e através da tarefa adequada. Conforme afirma Gonçalves (1999) nenhuma técnica resolve todos os problemas de mineração de dados, a familiaridade com diversas técnicas se torna necessário para encontrar o melhor caminho para a resolução do problema. No Quadro 1 são mostradas as principais tarefas de mineração de dados com as principais técnicas e algoritmos utilizados.

Quadro 1. Técnicas e algoritmos que podem ser aplicados nas principais tarefas de mineração de dados

Tarefa de mineração de dados	Técnicas	Algoritmos
Classificação	Indução por árvore de decisão; classificação bayseana; classificação por backpropagation; classificação baseada em conceitos da mineração de regras de associação; análise de vizinhança; algoritmos genéticos; abordagem por conjuntos fuzzy	<i>Support Vector Machine</i> (SVM); Algoritmo C4.5; CART; KNN; Classificadores Bayesianos; J48.
Clusterização	Métodos de particionamento; métodos hierárquicos; métodos baseados em densidade; métodos baseados em grid; métodos de <i>clustering</i> baseados em modelos (abordagem	K-means; K-modes; K-prototypes; Fuzzy K-means; Canopy; Cobweb;

	estatística e redes neurais); análise de outliers	
Regra de associação	Mineração de regras de associação (análise da cesta de venda – <i>market basket analysis</i>); minerando regras de associação booleanas unidimensionais a partir de banco de dados transacionais; minerando regras de associação em múltiplos níveis a partir de bancos de dados transacionais; minerando regras de associação multidimensionais a partir de bancos de dados transacionais e <i>data warehouse</i> ; da mineração de associação à análise de correlação; mineração de associação baseada em restrição	Apriori; FP-Growth; Eclat.
Estimativa	Regressão linear; regressão múltipla; regressão não linear; regressão logística; regressão de poisson	Back-Propagation; Lógica nebulosa.
Desvios	Ferramentas de consulta e técnicas de estatística; indução por árvores de decisão	Apriori; c4.5.

Fonte: Adaptação de Côrtes *et al.*, (2002).

2.2.4 Regra de associação e funcionamento do algoritmo FP-Growth

Conforme Silva e Ralha (2011), a Regra de Associação consiste em descobrir relações fortes entre determinados atributos. Essa tarefa tem como objetivo detectar padrões em forma de regras que associam valores de atributos em um determinado conjunto de dados.

Uma das técnicas para se realizar associação é a técnica *Market basket analysis*, que se trata de uma tecnologia criada com o objetivo de descobrir combinações de itens que ocorrem com frequência acima do esperado em uma base de dados. A partir das combinações de itens consideradas interessantes do ponto de vista estatístico, os algoritmos de *market basket analysis* geram um conjunto de regras de associação para os usuários. A título de exemplo, essa técnica traz resultados como, 70% dos clientes que compram o produto “strogonoff de frango (caixa)” também compram o produto “lasanha (caixa)”, 5,5% de todas as compras contêm os dois produtos (DEV MEDIA, 2020).

Algoritmos tradicionais que utilizam Regra de Associação adotam uma abordagem igual ou semelhante a do algoritmo Apriori, que se baseia na seguinte regra: se qualquer padrão de comprimento k não é frequente na base de dados, seu comprimento $(k + 1)$ não será frequente. A ideia é, por meio de um processo iterativo, gerar o conjunto de padrões de candidatos de comprimento $(k + 1)$ a partir do conjunto de padrões de frequência de comprimento k (para $k \geq 1$), e verificar suas frequências de ocorrência na base de dados. No entanto, a geração de conjunto candidato é ainda dispendiosa especialmente quando há muitos padrões e/ou estes são longos, ou seja, padrões formados por um número expressivo de itens. Também é dispendioso percorrer repetidas vezes a base de dados para verificar e testar todos os conjuntos de candidatos e seus padrões correspondentes (NANDI *et al.*, 2015).

Com a proposta de superar as limitações mencionadas acima e melhorar a eficiência da descoberta de regras de associação diversos métodos e algoritmos alternativos têm sido desenvolvidos. Como por exemplo, o algoritmo FP-Growth que 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. Isso possibilita uma melhor eficiência na geração das regras de associação, pois evita constantes acessos na base de dados (NANDI *et al.*, 2015).

Para o algoritmo FP-Growth funcionar corretamente é preciso informar pelo menos dois parâmetros de entrada, sendo eles, suporte e confiança. 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. Confiança é responsável pela força da regra, dessa forma, a confiança corresponde a frequência com que B ocorre, dentre os eventos que contêm A (TAN; STEINBACH; KUMAR, 2009).

Após definir os parâmetros de entrada e uma base de dados a ser analisada, o primeiro passo do algoritmo consiste em montar a estrutura FP-tree que irá armazenar os primeiros *itemsets* frequentes. Para isso, é necessário ler a base de dados duas vezes para que a FP-tree seja criada. Durante a primeira leitura é identificado o conjunto de *itemsets* frequentes F , de tamanho um, e seus respectivos suportes conforme é ilustrado na Tabela 1.

Tabela 1. Primeira leitura da base de dados

Transações	Itens	1 <i>Itemsets</i>
1	N, J, M	N:1, J:1, M:1
2	N, F, G	N:2, J:1, M:1, F:1, G:1
3	J, G, F, H	N:2, J:2, M:1, F:2, G:2, H:1
4	J, F, H	N:2, J:3, M:1, F:3, G:2, H:2
5	J, N, G	N:3, J:4, M:1, F:3, G:3, H:2
6	J, N, G, F	N:4, J:5, M:1, F:4, G:4, H:2
7	J, P	N:4, J:6, M:1, F:4, G:4, H:2, P:1
8	J, N, G	N:5, J:7, M:1, F:4, G:5, H:2, P:1
9	J, N, F	N:6, J:8, M:1, F:5, G:5, H:2, P:1
10	N, G, H, O	N:7, J:8, M:1, F:5, G:6, H:3, P:1, O:1

Fonte: Nandi *et al.* (2015).

Após a primeira leitura da base de dados os primeiros *itemsets* que não são frequentes, ou seja, que possuem valor de suporte menor que o especificado são removidos e os frequentes são ordenados em ordem decrescente e armazenados em uma lista L . Em seguida começa a segunda leitura da base de dados de modo que cada transação é lida novamente, os *itemsets* que não são frequentes de tal transação são desconsiderados e os frequentes são armazenados em uma lista T . A lista T é ordenada de acordo com o valor de suporte dos *itemsets* armazenados na lista L . Em

seguida a lista T é passada como parâmetro para o método que insere a transação na FP-tree (NANDI *et al.*, 2015).

Em seguida com todas as transações processadas e inseridas na estrutura FP-tree, esta estará corretamente criada. Ela irá representar a base de dados original e será passada como parâmetro para outro método do algoritmo FP-Growth responsável pela descoberta dos N *itemsets* frequentes. O método consiste no desenvolvimento de uma estratégia baseada na técnica de dividir para conquistar, em que o problema é organizado em subproblemas. O algoritmo faz isso considerando cada item da tabela de cabeçalhos. Após criadas, as FP-tree condicionais são utilizadas para encontrar os padrões frequentes que apresentam o *itemset* como sufixo (HAN; PEI; YIN, 2000).

Com o conjunto de N *itemsets* frequentes encontrados por meio da FP-tree, o algoritmo FP-Growth gera as regras de associação. A partir deste momento começa a fase de pós-processamento das regras geradas, objetivando a seleção das regras úteis e fortes. O algoritmo FP-Growth irá primeiro calcular a confiança.

Em seguida o algoritmo irá calcular o *Lift*, que é utilizado para avaliar dependências, indicando o quanto mais frequente torna-se o consequente da regra quando o antecedente ocorre. Depois o *Leverage* será calculado, medida que é utilizada para descobrir o valor da diferença entre o suporte real e o esperado de uma regra de associação. E por último a Convicção é calculada, medida que tem como objetivo avaliar uma regra de associação como uma implicação do antecedente no consequente. Somente as regras de associação que satisfizerem todas as medidas de qualidade são consideradas relevantes e úteis. (GUILLET; HAMILTON, 2007).

2.2.5 Clusterização e o algoritmo *K-means*

Conforme Queiroz (2016), a Clusterização consiste na divisão de um grupo heterogêneo em vários subgrupos mais homogêneos. Neste processo não existem classes pré-definidas e os dados são agrupados de acordo com suas características próprias.

A técnica de agrupamento em métodos de particionamento é uma das técnicas possíveis para a realização de Clusterização, que tem como objetivo encontrar a melhor partição dos n objetos em K grupos. Normalmente os K grupos encontrados possuem mais qualidade comparados com K grupos produzidos pelos métodos

hierárquicos. Um dos algoritmos mais famosos para essa técnica é o algoritmo denominado como *K-means* (RODRIGUES; 2009).

O *K-means* é um algoritmo de aprendizagem não supervisionado e de Clusterização, utilizado para particionar dados em *k* agrupamentos. Ele agrupa dados que compartilham características importantes e parecidas. De modo empírico uma boa solução para o processo de Clusterização é aquela em que os dados do grupo sejam mais semelhantes entre si, do que comparados com outro grupo (SOUSA; 2019).

O *K-means* é caracterizado como um algoritmo guloso, que em suas iterações escolhe o objeto que lhe parece mais promissor, que chama mais sua atenção, torna o objeto parte da solução do problema, mas não analisa as consequências de suas escolhas (SOUSA; 2019).

O algoritmo *K-means* pode ser descrito por 4 (quatro) etapas: atribui-se valores iniciais para os protótipos seguindo algum critério, por exemplo, sorteio aleatório desses valores dentro dos limites de domínio de cada atributo; atribui-se cada objeto ao grupo cujo protótipo possua maior similaridade com o objeto; recalcula-se o valor do centroide (protótipo) de cada grupo, como sendo a média dos objetos atuais do grupo; repete-se os passos 2 e 3 até que os grupos se estabilizem (FONSECA; BELTRAME, 2010).

2.2.6 Ferramentas para mineração de dados

Ferramentas de mineração de dados foram desenvolvidas com o objetivo de tornar a aplicação de mineração de dados uma tarefa mais usável e com menor dependência dos aspectos de programação, e com isto possibilitar que profissionais de qualquer área possam fazer uso dela. Há diversas ferramentas no mercado, cada uma com suas características e aplicada para determinadas tarefas de mineração de dados.

Weka é um software de aprendizado de máquina de código aberto testado e aprovado que pode ser acessado por meio de uma interface gráfica do usuário, aplicativos de terminal padrão ou uma *Application Programming Interface* (API) Java. É amplamente usado para aplicações de ensino, pesquisa e industriais, contém uma infinidade de ferramentas integradas para tarefas de aprendizado de máquina padrão e, adicionalmente, dá acesso transparente a caixas de ferramentas conhecidas, como

scikit-learn , R e Deeplearning4j. Pode-se utilizar para as tarefas de mineração, Classificação, Regra de associação, Clusterização e Estimativa (WEKA, 2020).

RapidMiner é um pacote de software que permite mineração de dados, mineração de texto e análise preditiva. O programa permite ao usuário inserir dados brutos, incluindo bancos de dados e texto, que são então analisados de forma automática e inteligente em larga escala. O software permite realizar análises com as tarefas, Classificação, Regra de associação, Clusterização e Estimativa (RAPIDMINER, 2020).

SAS Enterprise Miner é a ferramenta da *Statistical Analysis System* (SAS) que combina um extenso pacote de ferramentas integradas de mineração de dados com uma boa facilidade de utilização, capacitando os usuários a tirarem proveito de dados corporativos para obter vantagem estratégica, tudo em um único ambiente. Pode-se utilizar para as tarefas de mineração, Classificação, Regra de associação e Clusterização (SAS, 2020).

Intelligent Miner é a solução integrada da *International Business Machines* (IBM) para a análise sofisticada de dados. Ele permite coletar informações valiosas de bancos de dados e apresentar as informações aos usuários de negócios para análise e tomada de decisão. As tradicionais técnicas de mineração de dados são suportadas e são adicionados, ricos componentes de apresentação para possibilitar uma melhor análise visual dos resultados (IBM, 2020).

Há diversas outras ferramentas disponíveis no mercado, como, Clementine, Poly Analyst, WizRule, Tamanduá e oracle Data mining. Cada uma com recursos e características diferentes, necessitando analisar o contexto do problema a ser resolvido para selecionar a ferramenta mais adequada.

2.3 Estudos relacionados

Há vários trabalhos relevantes disponíveis relacionando mineração de dados com licitações públicas, sendo um desses o de Silva e Ralha (2011), no qual realizaram um estudo cujo objetivo é a detecção de cartéis de licitações públicas utilizando mineração de dados, e nesse caso foi utilizado as tarefas regras de associação e clusterização.

No estudo dos autores supracitados foi utilizado como estratégia organizar os *datasets* para procurar a associação entre empresas de forma que cada fornecedor

da base de dados, fosse um atributo booleano e cada instância um processo de licitação. Assim, para cada licitação, o atributo relativo a um determinado fornecedor é preenchido com o valor 'sim', caso aquele fornecedor tenha participado do certame, ou 'não', caso contrário. Para isso, foram utilizados dados do período de 2005 a 2008, extraídos do sistema “ComprasNet” do Governo Federal, que contém informações sobre as licitações realizadas na modalidade pregão eletrônico.

A partir de dados extraídos pelo Sistema de Acompanhamento da Gestão dos Recursos da Sociedade (SAGRES) da base de dados do Tribunal de Contas do Estado da Paraíba (TCE/PB), Grilo Júnior (2010) teve como objetivo relacionar possíveis indícios de fraudes em processos licitatórios nesse órgão. Nesse estudo foram utilizadas informações referentes aos participantes, propostas, fornecedores, contratos e aditivos. Os dados foram provenientes de licitações realizadas pelos municípios paraibanos no período de 2005 a 2009. O autor aplicou Regra de Associação nos dados obtidos, mediante o algoritmo FP-Growth, identificando a repetição dos mesmos participantes em vários processos licitatórios, indicando prováveis conluios entre as empresas.

Morais (2016), propôs em seu trabalho indicadores para investigação de licitações por meio de técnicas de reconhecimento de padrões estatísticos e mineração de dados. Nesse trabalho o autor realizou um estudo bibliográfico e identificou as principais variáveis necessárias para formação de indicadores de participação de cartéis em licitações, que são elas: objeto da licitação, órgão licitante, participantes e propostas de preços. Desenvolveu uma metodologia baseado nesses indicadores, utilizou a tarefa de mineração de dados Regra de Associação e realizou classificação não supervisionada, com o algoritmo *k-means*, e aplicou em dados disponibilizados no portal da transparência do Tribunal de Contas dos Municípios do Estado do Ceará (TCM/CE). Os resultados obtidos pelo autor mostraram 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.

3. PROCEDIMENTO METODOLÓGICO

Essa seção descreve o procedimento metodológico utilizado nesse estudo, assim como as suas etapas em relação as atividades realizadas, de forma a permitir uma compreensão dos processos que auxiliaram na obtenção dos resultados.

Segundo Gil (2017) pesquisa define-se como o procedimento racional e sistemático que busca proporcionar respostas aos problemas que são propostos. A pesquisa é requerida quando não se tem informações disponíveis para que se possa responder o problema, ou quando a informação disponível não se encontra organizada, fazendo com que não se possa obter uma relação com o problema.

As pesquisas científicas podem ser classificadas de acordo com a natureza, abordagem, objetivos e procedimentos técnicos. Quanto à natureza, esta pesquisa é denominada como aplicada, dedicada à geração de conhecimento para solução de problemas específicos, e dirigida à busca da verdade para determinada aplicação prática em situação particular (NASCIMENTO; SOUSA, 2015).

Quanto a abordagem, essa pesquisa é do tipo quantitativa, porque utiliza uma abordagem ou método que emprega medidas padronizadas e sistemáticas, reunindo respostas pré-determinadas, facilitando a comparação e a análise de medidas estatísticas de dados (NASCIMENTO; SOUSA, 2015).

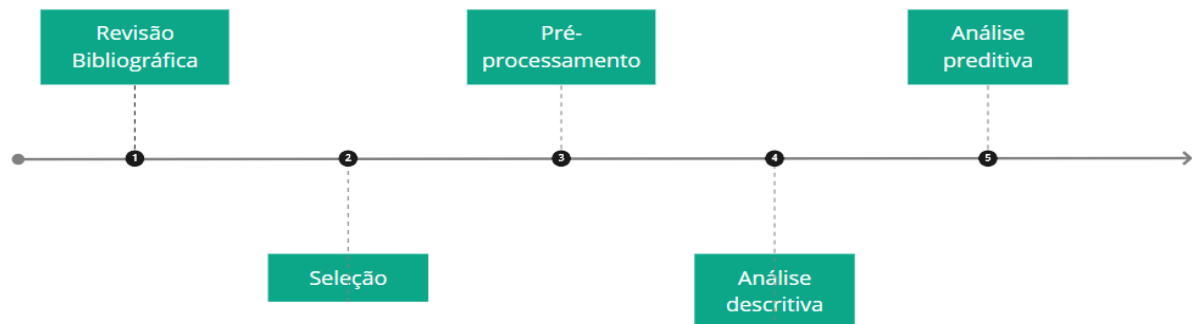
Quanto aos objetivos esse trabalho consiste em uma pesquisa exploratória e descritiva, pois foi desenvolvido com o objetivo principal de relacionar os principais aspectos e elementos necessários à compreensão da relação de ciência de dados com técnicas de mineração de dados e sua aplicação em licitações públicas, buscando anomalias que antes eram desconhecidas, descrevendo a eficiência da mineração de dados para essa situação problema (WAZLAWICK, 2014).

Quanto aos procedimentos técnicos esse estudo é uma pesquisa bibliográfica e experimental, pois segundo Gil (2017), a pesquisa bibliográfica propicia bases teóricas ao pesquisador para auxiliar no exercício reflexivo e crítico sobre o tema em estudo, sendo importante para despertar a curiosidade, inquietações e entendimento acerca do problema.

Conforme Nascimento e Sousa (2015), com a pesquisa experimental é estabelecido um objeto de estudo, selecionado as variáveis que podem influenciá-lo, definido mecanismos e formas de controle e de observação dos efeitos causados pelas variáveis selecionadas sobre o objeto pesquisado.

O método aplicado nesse estudo é apresentado na Figura 2.

Figura 2. Método para análise dos dados



Fonte: Elaborado pelo autor.

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 (WAZLAWICK, 2014).

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. 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 segunda etapa, 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 Moraes (2016) 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 quarta etapa, 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 quinta e última etapa, análise preditiva, foi dividido 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. Essa análise foi realizada baseando-se na metodologia desenvolvida por Silva e Ralha (2011), foi aplicado Regra de Associação com o apoio do algoritmo FP- Growth no *dataframe* de propostas de licitações, 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*. A última etapa foi realizada associação com os clusters obtidos, escolhidas as melhores regras e analisado as características das empresas nas participações e nas vitórias, com o apoio do software de aprendizado de máquina RapidMiner.

4. 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.

4.1 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 está disponível 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 contém 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.

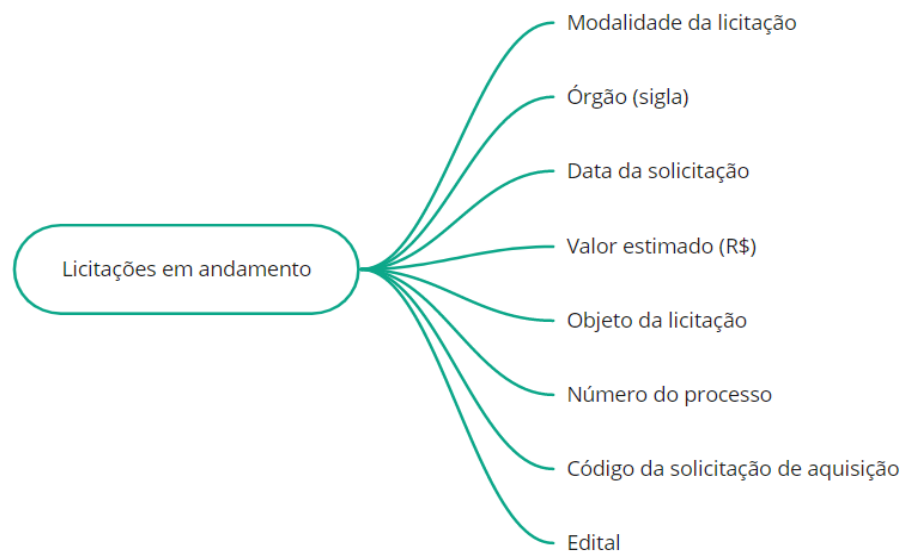
Na segunda opção de acesso aos dados de licitações, as informações são disponibilizadas em arquivos no formato *comma-separated values* (CSV), tipo de formato de arquivo de valores separados por vírgula amplamente utilizado para armazenar dados tabulares. Os dados são disponibilizados no servidor *File 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 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 3.

Figura 3. Informações disponíveis na seção sobre as licitações em andamento



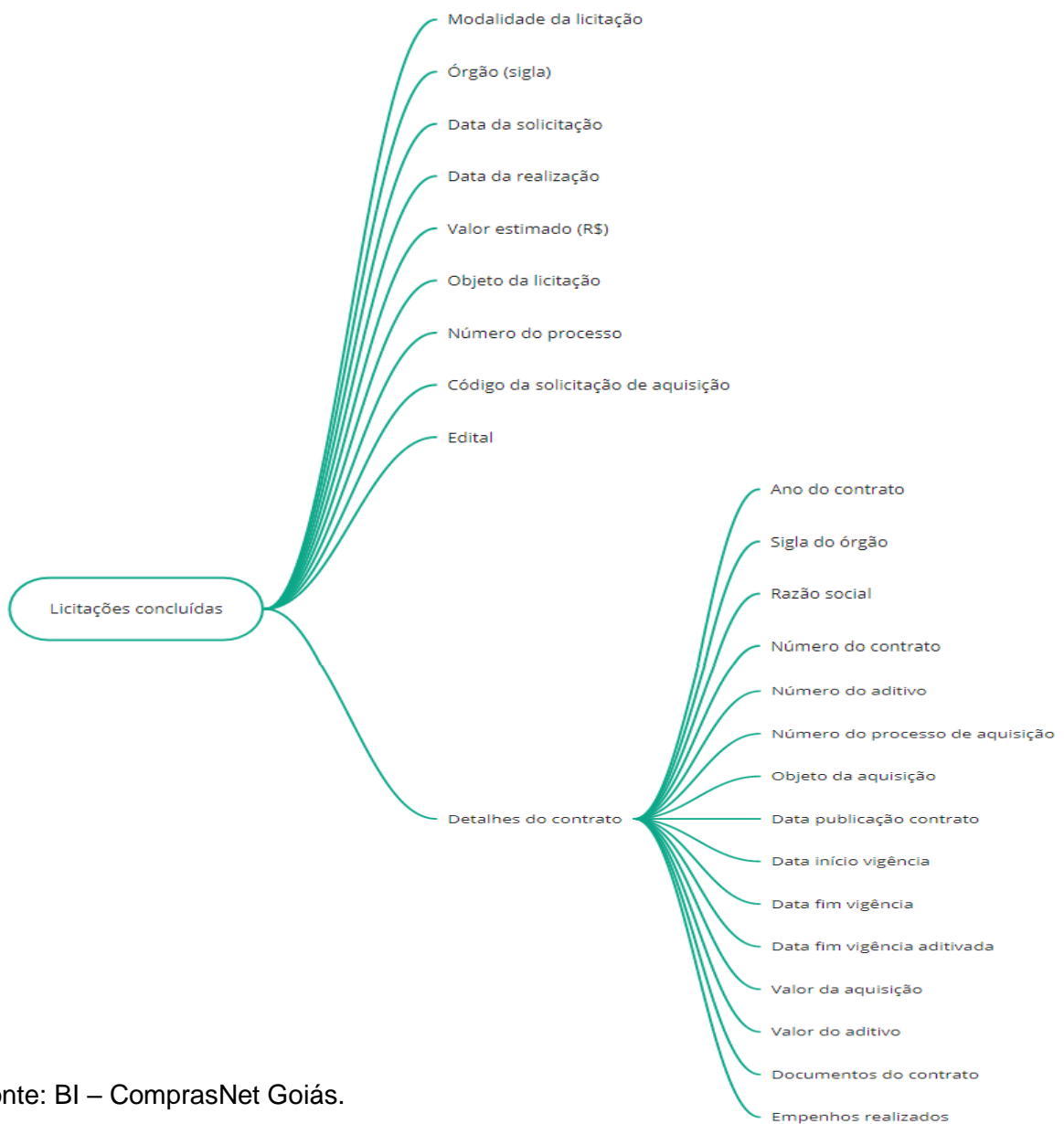
Fonte: BI – ComprasNet Goiás.

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 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 4.

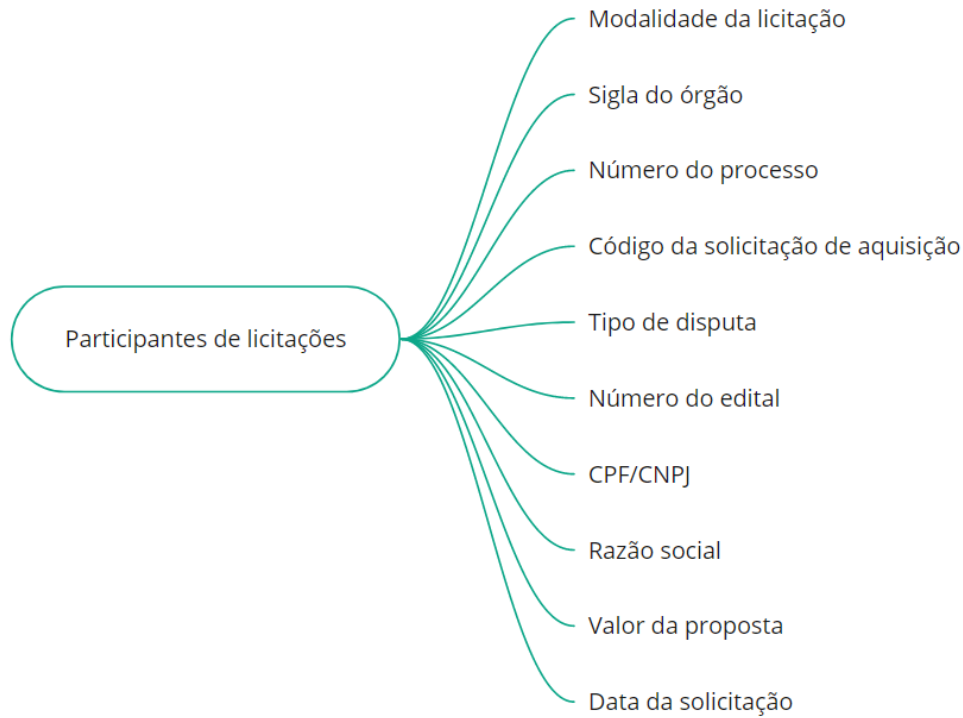
Figura 4. Informações disponíveis na seção sobre as licitações concluídas



Fonte: BI – ComprasNet Goiás.

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 5.

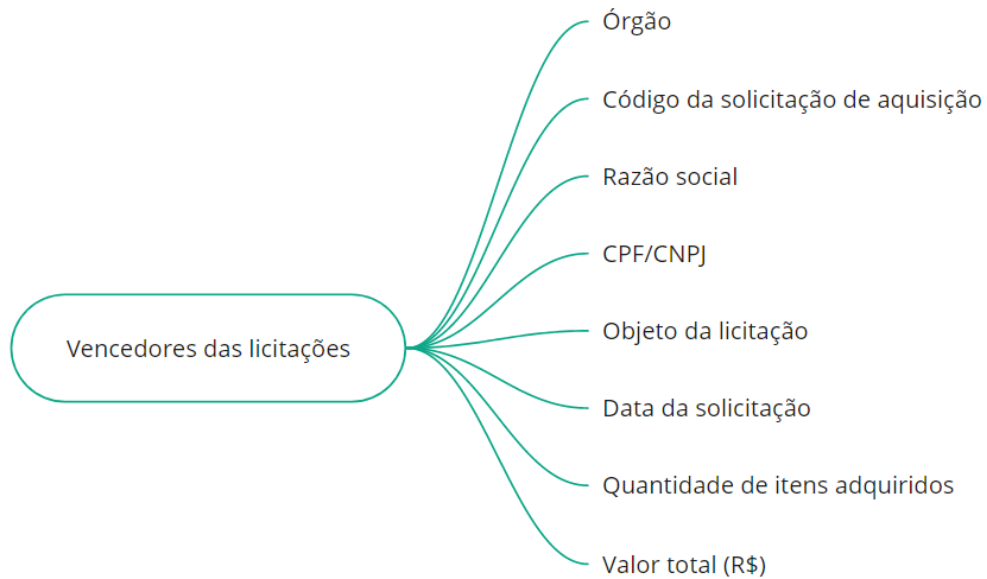
Figura 5. Informações disponíveis na seção sobre os participantes de licitações.



Fonte: BI – ComprasNet Goiás.

E na última seção de informações disponíveis é referente aos vencedores das licitações, podendo 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 6.

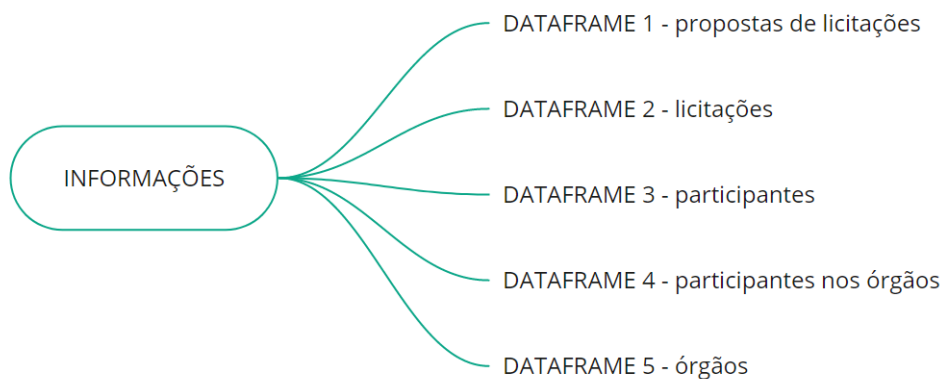
Figura 6. Informações disponíveis na seção sobre os vencedores das licitações



Fonte: BI – ComprasNet Goiás.

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 7.

Figura 7. Organização das informações em 5 (cinco) *dataframes*



Fonte: Adaptação de Moraes (2016).

O Quadro 2 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 (MORAIS, 2016).

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 2.

Quadro 2. Variáveis do “DATAFRAME 1 - propostas de licitações”

Metodologia Morais (2016)	Portal da transparência 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

Fonte: Elaborado pelo autor.

O Quadro 3 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 3.

Quadro 3. Variáveis do “DATAFRAME 2 – licitações”

Metodologia Morais (2016)	Portal da transparência 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

Fonte: Elaborado pelo autor.

O Quadro 4 é referente ao “DATAFRAME 3 – participantes”, apresentando as características gerais da empresa. 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 um participante único (empresa), e cada coluna representa as variáveis do Quadro 4.

Quadro 4. Variáveis do “DATAFRAME 3 – participantes”

Metodologia Morais (2016)	CGE/GO
CNPJ do participante	CPF/CNPJ
Nome do participante	Razão social
Porte do participante	
Número de participações	Vincular com <i>dataframe 1</i>
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 1</i>
Número total de vitórias em licitações	Vincular com <i>dataframe 2</i>
Número de órgãos em que participou licitações;	Vincular com <i>dataframe 1</i>
Número de órgãos em que venceu licitações;	Vincular com <i>dataframe 2</i>

Fonte: Elaborado pelo autor.

O Quadro 5 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 5. Nesse *dataframe* cada linha representa a combinação de um participante (empresa) + órgão único, e cada coluna representa as variáveis do Quadro 5.

Quadro 5. Variáveis do “DATAFRAME 4 - participantes nos órgãos”

Metodologia Morais (2016)	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 1</i>
Valor de todas as participações do participante no órgão	Vincular com <i>dataframe 1</i>
Valor total estimado de todas as licitações do participante no órgão	Vincular com <i>dataframe 1</i>
Número de vitórias do participante no órgão	Vincular com <i>dataframe 2</i>
Valor total de vitórias do participante no órgão	Vincular com <i>dataframe 2</i>
Valor total estimado das vitórias do participante no órgão	Vincular com <i>dataframe 2</i>

Fonte: Elaborado pelo autor.

E por fim, o Quadro 6 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 6. Nesse *dataframe* cada linha representa um órgão, e cada coluna representa as variáveis do Quadro 6.

Quadro 6. Variáveis do “DATAFRAME 5 – órgãos”

Metodologia Morais (2016)	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 2</i>
Número de itens de licitações do órgão	Vincular com <i>dataframe 2</i>

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 2</i>
Valor total das licitações no órgão	Vincular com <i>dataframe 2</i>
Valor total estimado das licitações no órgão	Vincular com <i>dataframe 2</i>
Número total de vencedores no órgão	Vincular com <i>dataframe 2</i>
Número total de participantes no órgão	Vincular com <i>dataframe 1</i>
CNPJ das quatro maiores vencedoras de licitação no órgão	Vincular com <i>dataframe 4</i>

Fonte: Elaborado pelo autor.

Os dados obtidos por modalidade de licitação são mostrados na Tabela 2.

Tabela 2. Quantidade de licitações totais obtidas por modalidade

Modalidade	Número de 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
Total	19658

Fonte: Elaborado pelo autor.

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 8 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.

Figura 8. Palavras filtradas na descrição do objeto da licitação



Fonte: Elaborado pelo autor.

Com o tratamento descrito, obteve-se os *dataframes* conforme a Tabela 3.

Tabela 3. Quantidade de licitações obtidas para saúde e medicamentos por modalidade

Modalidade	2016	2017	2018	2019	Total	%
Concorrência	1	4	4	2	11	0,49%
Concorrência SRP	1				1	0,04%
Convite		1			1	0,04%

Participante Registro de Preço	462	458	392	328	1640	72,76%
Pregão Eletrônico	46	40	34	19	139	6,17%
Pregão Eletrônico SRP	138	146	93	77	454	20,14%
Pregão Presencial	1	1	2		4	0,18%
Pregão Presencial RDC		2		1	3	0,13%
Tomada de Preço		1			1	0,04%
Total	649	653	525	427	2254	100%

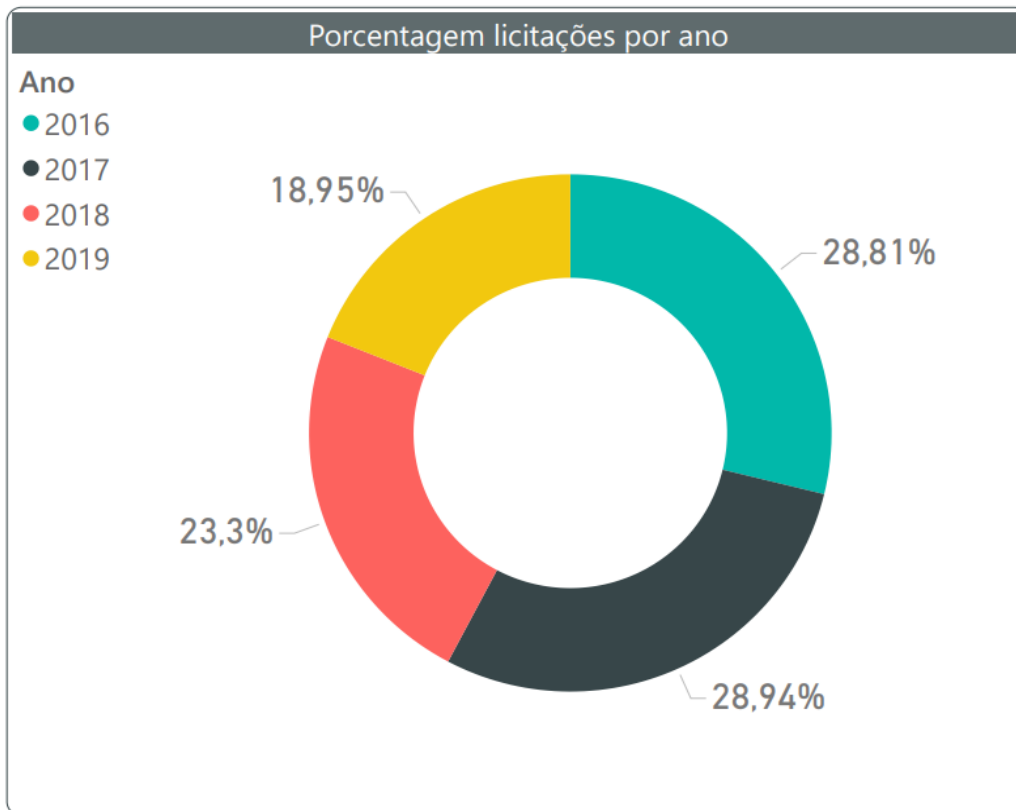
Fonte: Elaborado pelo autor.

4.2 Análise descritiva

Com os dados obtidos e organizados o processo de análise descritiva foi iniciado. 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 realizada 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 9 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, com 18,95%.

A Figura 10 exibe as 9 (nove) modalidades com a quantidade de licitações, o valor estimado e o valor total. Para o escopo de saúde e medicamentos, observa-se que a modalidade com maior quantidade de licitações é a de Participante e Registro de Preço, obtendo 1640 licitações de 2254 no total. Em seguida a modalidade Pregão Eletrônico SRP com 454 licitações, Pregão Eletrônico com 139, Concorrência com 11, Pregão presencial com 4, Pregão presencial RDC com 3, Concorrência SRP com 1, Convite com 1 e Tomada de Preço com 1. Há uma diferença muito grande também do valor estimado e do valor total das licitações, podendo ser um erro no banco de dados.

Figura 9. Distribuição das licitações no período de 2016 a 2019

Fonte: Elaborado pelo autor.

Figura 10. Número de licitações, valor estimado e valor total por modalidade

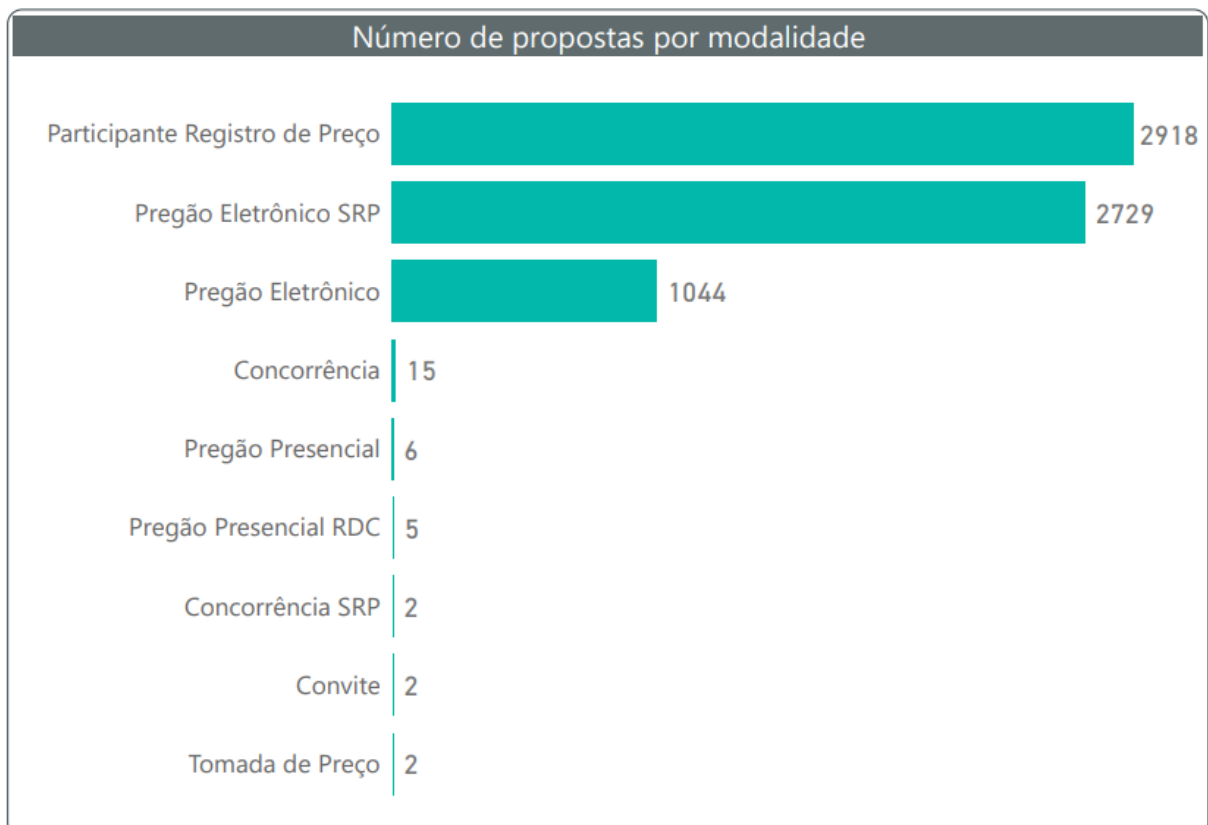
Valores das licitações analisadas			
Modalidade	Nº licitações	Valor estimado	Valor total
Participante Registro de Preço	1640	4.636.699.100,48	644.830.856,01
Pregão Eletrônico SRP	454	19.413.720.525,40	2.025.371.517,24
Pregão Eletrônico	139	656.000.327,87	49.150.125,60
Concorrência	11	124.352.577,86	104.352.577,86
Pregão Presencial	4	60.850.541,79	10.183.008,12
Pregão Presencial RDC	3	17.077.012,01	17.077.012,01
Concorrência SRP	1	63.105.700,00	3.712.100,00
Convite	1	435.000,00	87.000,00
Tomada de Preço	1	73.440.872,00	4.590.054,50
Total	2254	25.045.681.657,41	2.859.354.251,34

Fonte: Elaborado pelo autor.

A Figura 11 exibe o número de propostas por modalidade de todo o período (2016-2019). A modalidade Participante Registro de Preço obteve 2918 propostas, Pregão Eletrônico SRP com 2729, Pregão Eletrônico com 1044, Concorrência com 15, Pregão presencial com 6, Pregão Presencial RDC com 5, Concorrência SRP com 2, Convite com 2 e Tomada de Preço com 2 propostas. A Figura 12 exibe a mesma informação, porém de forma percentual. As modalidades Participante Registro de Preço e Pregão Eletrônico SRP correspondem juntas a mais de 80% de todas as propostas.

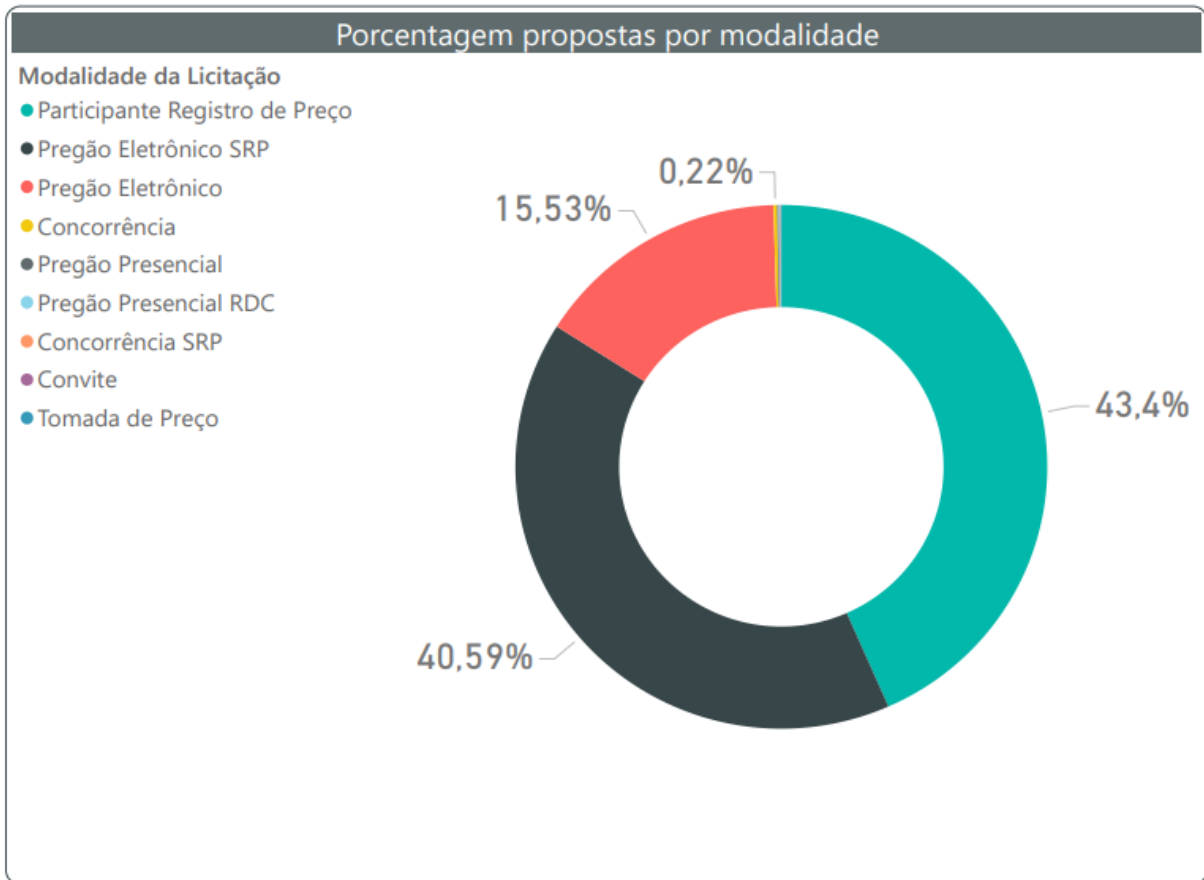
A Figura 13 aborda o número de propostas por tipo de disputa. Para o escopo desse estudo somente 2 (dois) tipos de disputa estiveram presentes, sendo eles, Menor Preço e Maior Desconto. Menor preço correspondendo a 6721 propostas e Maior Desconto com apenas 2 propostas.

Figura 11. Número de propostas por modalidade



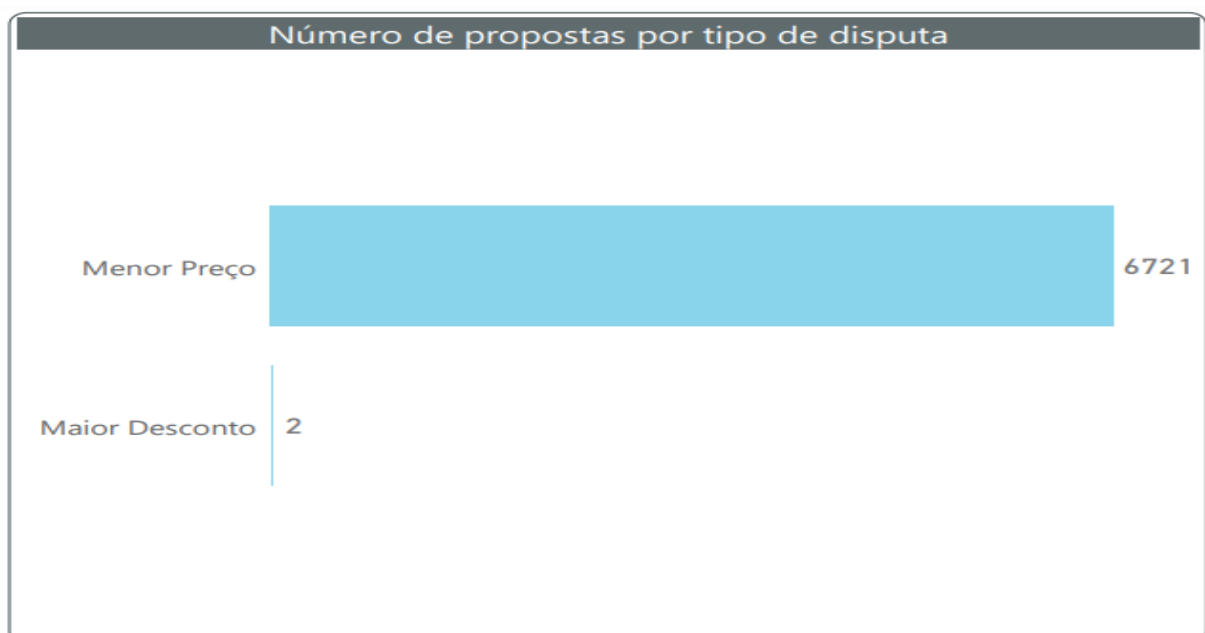
Fonte: Elaborado pelo autor.

Figura 12. Porcentagem das propostas por modalidade



Fonte: Elaborado pelo autor.

Figura 13. Número de propostas por tipo de disputa



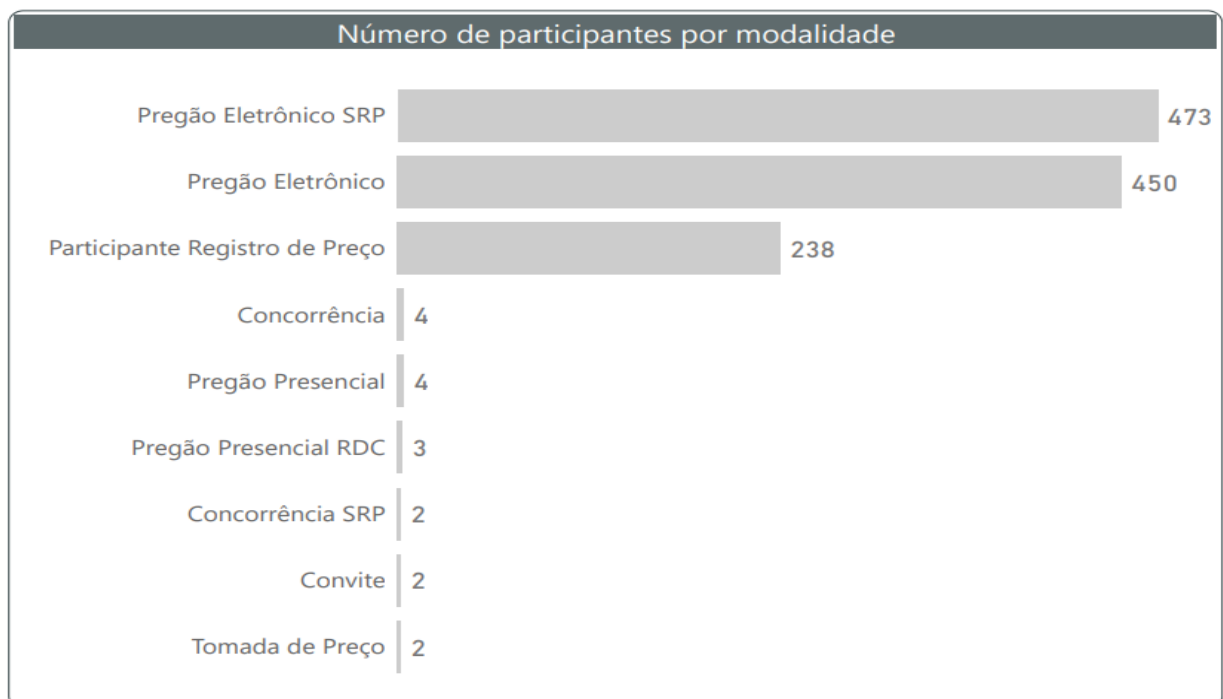
Fonte: Elaborado pelo autor.

A Figura 14 exibe o número de participantes por modalidade. Há no total 846 participantes, alguns participando de mais de uma modalidade. A modalidade Pregão Eletrônico SRP contém 473 participantes, Pregão Eletrônico com 450, Participante Registro de Preço com 238, Concorrência com 4, Pregão presencial com 4, Pregão Presencial RDC com 3, Concorrência SRP com 2. Convite com 2 e Tomada de Preço com 2.

A Figura 15 ilustra o número de vencedores por modalidade. A modalidade Pregão Eletrônico SRP contém 247 vencedores, Participante Registro de Preço com 235, Pregão Eletrônico com 179, Concorrência com 3, Pregão presencial com 3, Concorrência SRP com 1, Convite com 1, Pregão Presencial RDC com 1 e Tomada de Preço com 1.

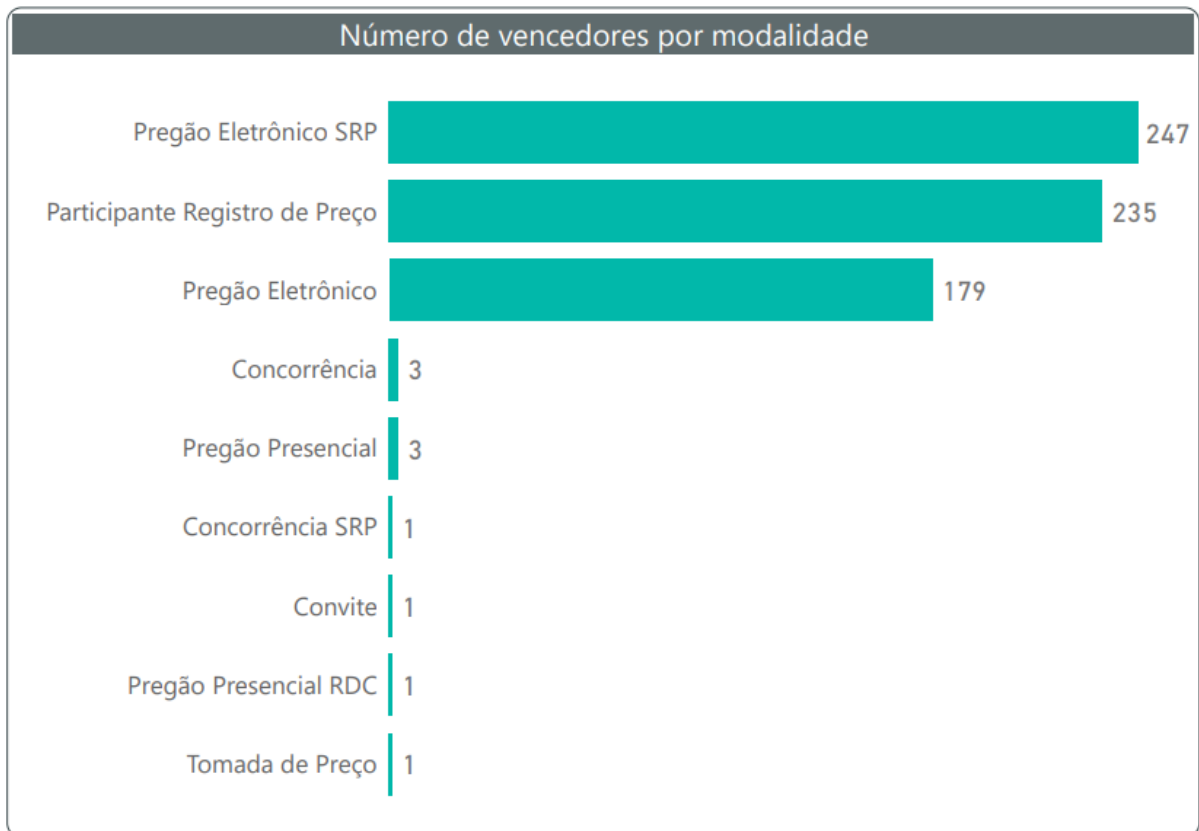
O número de propostas e número de licitações por ano é exibido na Figura 16. No ano de 2016 temos 2090 propostas para 649 licitações, em 2017 contém 1901 propostas para 652 licitações, 2018 com 1597 propostas para 525 licitações e para o ano de 2019 contém 1135 propostas para 427 licitações. Os anos 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 14. Número de participantes por modalidade.



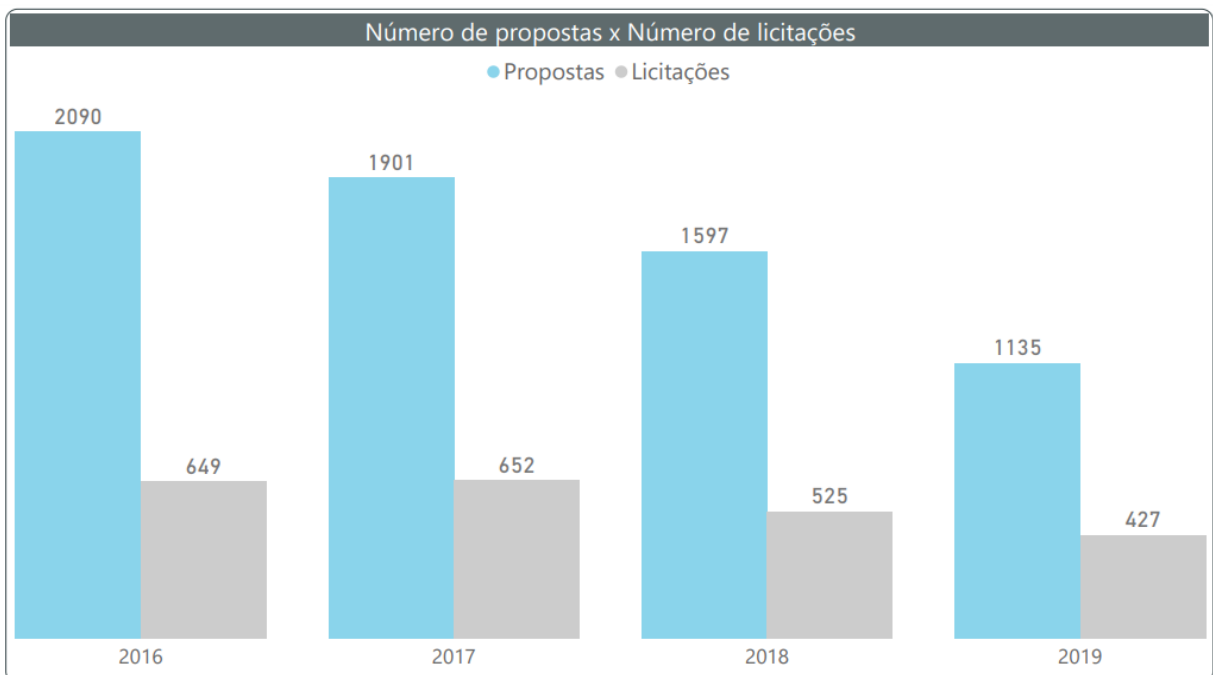
Fonte: Elaborado pelo autor.

Figura 15. Número de vencedores por modalidade.



Fonte: Elaborado pelo autor.

Figura 16. Número de propostas e número de licitações por ano



Fonte: Elaborado pelo autor.

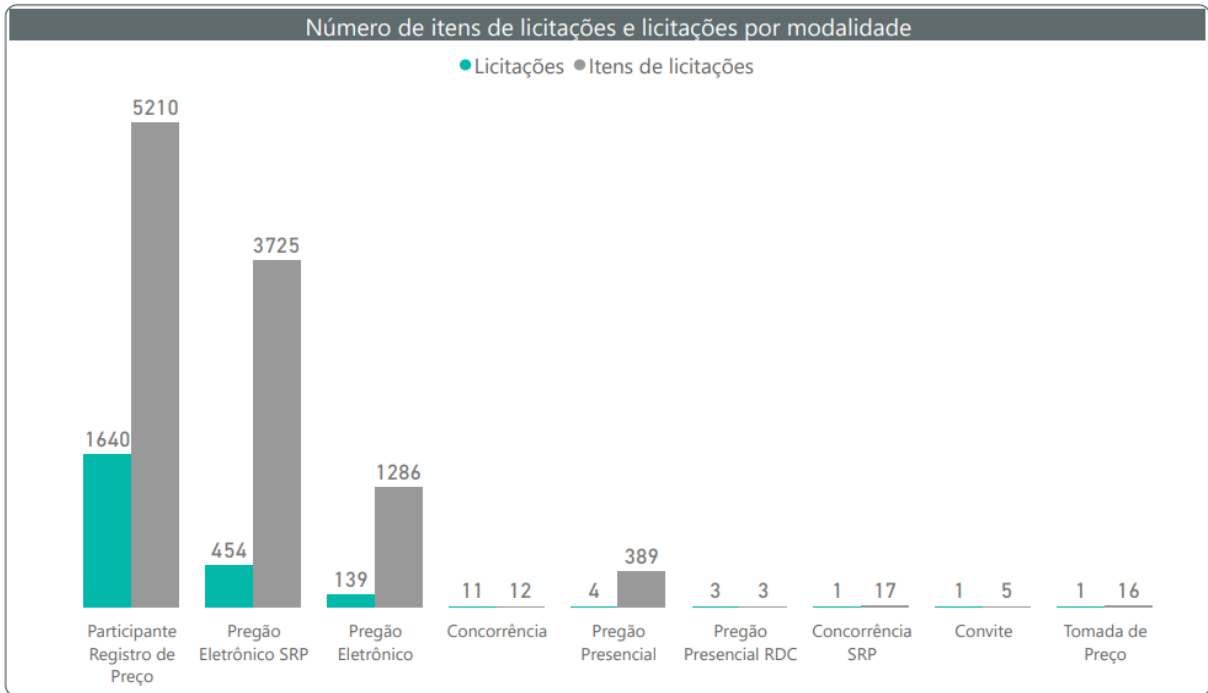
A Figura 17 ilustra o número de itens de licitações e o número de licitações por modalidade. A modalidade Participante Registro de Preço possui 5210 itens de licitações distribuídas em 1640 licitações, Pregão Eletrônico SRP possui 3725 itens em 454 licitações, Pregão Eletrônico possui 1286 itens em 139 licitações, Concorrência possui 12 itens em 11 licitações, Pregão presencial possui 389 itens em 4 licitações, Pregão Presencial RDC possui 3 itens em 3 licitações, Concorrência SRP possui 17 itens em 1 licitação, Convite possui 5 itens em 1 licitação e Tomada de Preço possui 16 itens em 1 licitação.

A Figura 18 fornece o valor estimado e o valor adjudicado total por modalidade. Para a modalidade Pregão Eletrônico SRP consta R\$ 19,41 Bilhões estimado e R\$ 2,03 Bilhões adjudicado total. Participante Registro de Preço possui R\$ 4,64 Bilhões estimado e R\$ 0,64 Bilhões adjudicado total. Pregão Eletrônico possui R\$ 0,66 Bilhões estimado e R\$ 0,05 Bilhões adjudicado total. Concorrência possui R\$ 0,12 Bilhões estimado e R\$ 0,1 Bilhões adjudicado total. Tomada de Preço possui R\$ 0,07 Bilhões estimado e R\$ 4,59 milhões adjudicado total. Concorrência SRP possui R\$ 63,11 milhões estimado e R\$ 3,71 milhões adjudicado total. Pregão Presencial possui R\$ 60,85 milhões estimado e R\$ 10,18 milhões adjudicado total. Pregão Presencial RDC possui R\$ 17,08 milhões estimado e R\$ 17,08 adjudicado total. Convite possui R\$ 0,44 milhões estimado e R\$ 0,09 milhões adjudicado total.

A Figura 19 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.

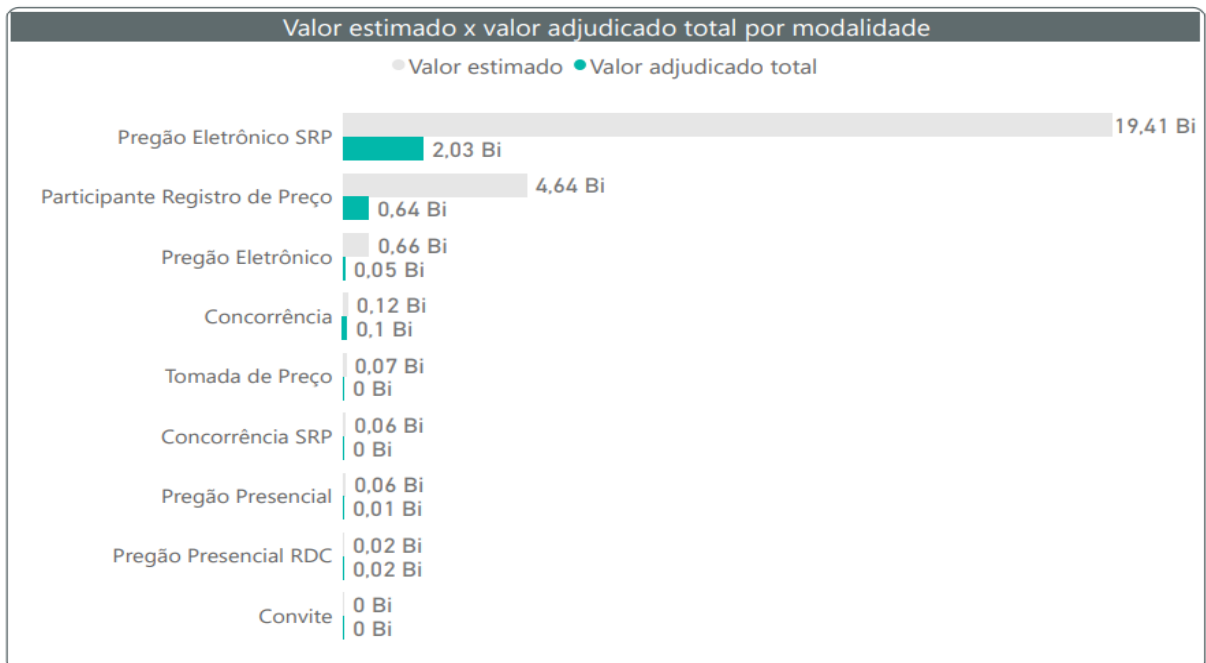
A Figura 20 ilustra uma estimativa das vitórias por empresa. Neste está listado as empresas que mais vencem licitações para as empresas com menores vitórias. A EMPRESA_36 está entre as que mais vencem, em seguida as EMPRESA_20 e EMPRESA_32.

Figura 17. Número de itens de licitações e licitações por modalidade

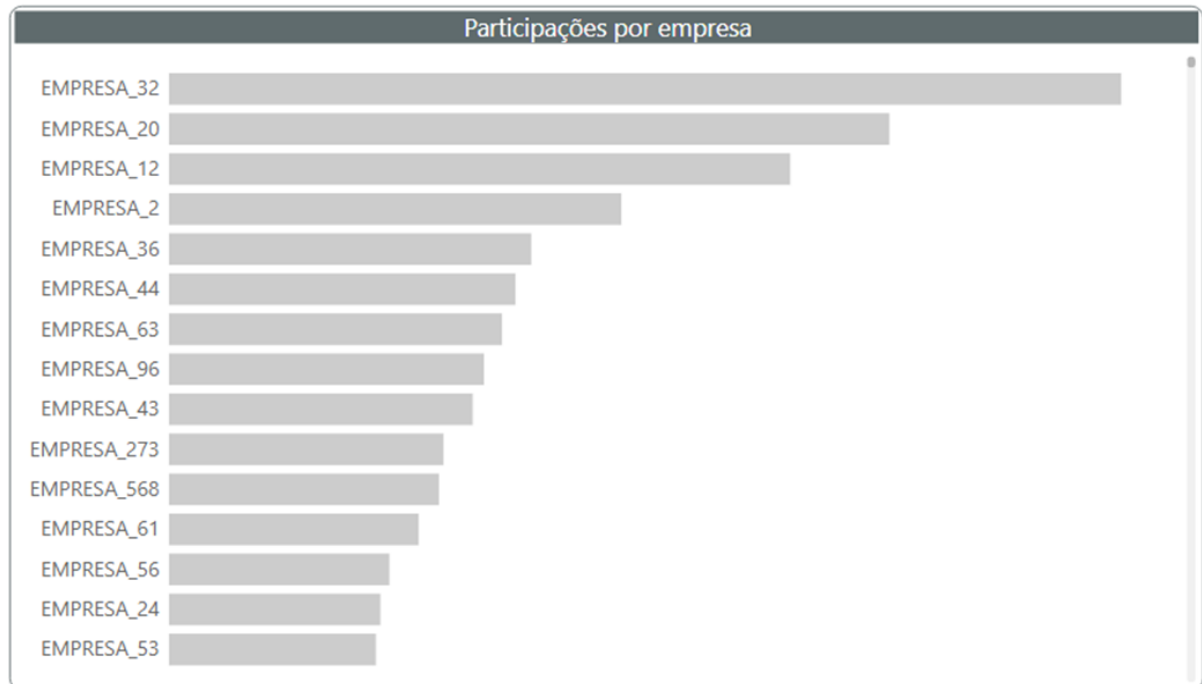


Fonte: Elaborado pelo autor.

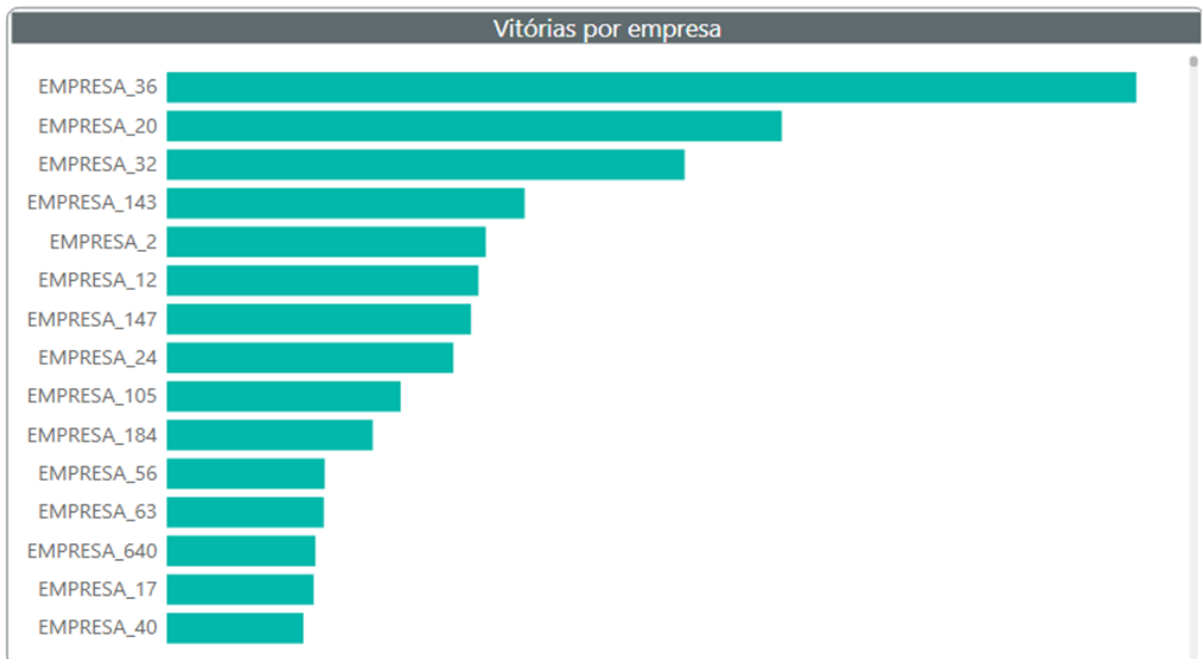
Figura 18. Valor estimado e valor adjudicado total por modalidade



Fonte: Elaborado pelo autor.

Figura 19. Participações por empresa

Fonte: Elaborado pelo autor.

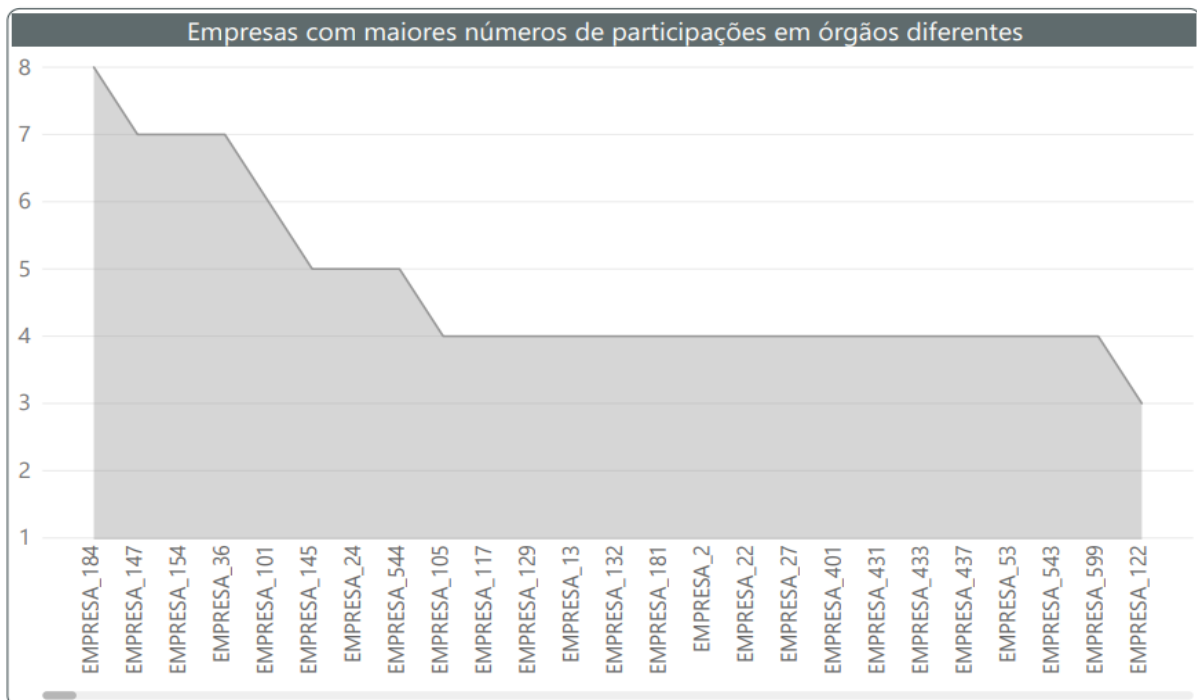
Figura 20. Vitórias por empresa.

Fonte: Elaborado pelo autor.

A Figura 21 ilustra as empresas com maiores números de participações em órgãos diferentes. A EMPRESA_184 está no topo, participando em 8 órgãos distintos. As EMPRESA_147, EMPRESA_154, EMPRESA_154 e EMPRESA_36 estão participando em 7 órgãos diferentes. A EMPRESA_101 está participando de 6 órgãos diferentes. As EMPRESA_145, EMPRESA_24 e EMPRESA_544 estão participando de 5 órgãos diferentes. As EMPRESA_105, EMPRESA_117, EMPRESA_129, EMPRESA_13, EMPRESA_132, EMPRESA_181, EMPRESA_2, EMPRESA_22, EMPRESA_27, EMPRESA_401, EMPRESA_431, EMPRESA_433, EMPRESA_437, EMPRESA_53, EMPRESA_543, EMPRESA_599 estão participando de 4 órgãos diferentes. A EMPRESA_122 está participando de 3 órgãos distintos.

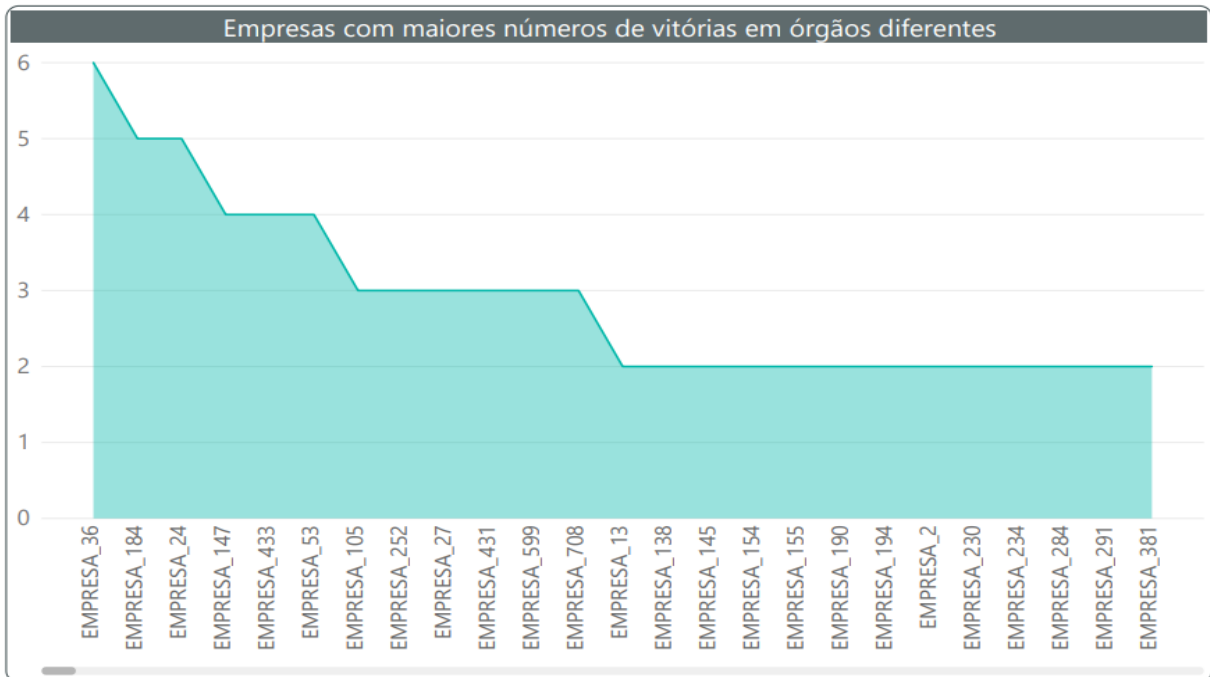
A Figura 22 traz a visão das maiores vitórias em órgãos diferentes por empresa. A EMPRESA_36 venceu em 6 órgãos diferentes. As EMPRESA_184 e EMPRESA_24 venceram em 5 órgãos. As EMPRESA_147, EMPRESA_433 e EMPRESA_53 venceram em 4 órgãos. As EMPRESA_105, EMPRESA_252, EMPRESA_27, EMPRESA_431, EMPRESA_599 e EMPRESA_708 venceram em 3 órgãos. As demais empresas desse gráfico venceram em 2 órgãos diferentes.

Figura 21. Empresas com maiores números de participações em órgãos diferentes



Fonte: Elaborado pelo autor.

Figura 22. Empresas com maiores números de vitórias em órgãos diferentes

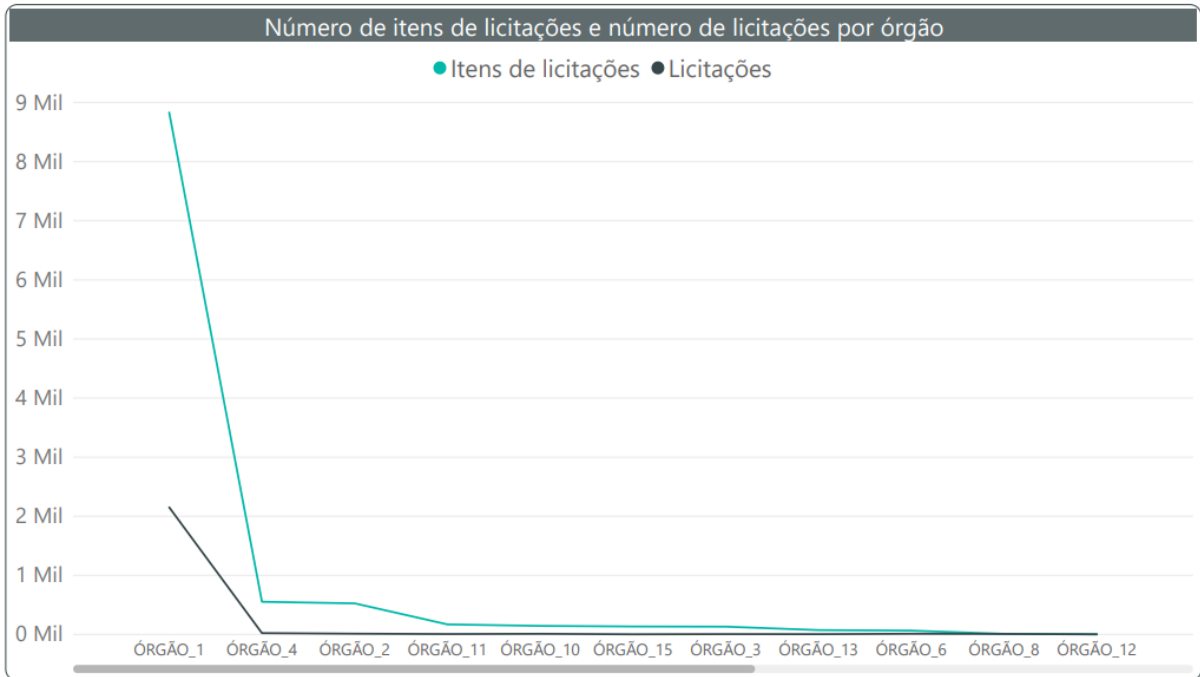


Fonte: Elaborado pelo autor.

A Figura 23 mostra os maiores número de itens de licitações e número de licitações por órgão. O ÓRGÃO_1 possui 8834 itens e 2152 licitações. O ÓRGÃO_4 possui 553 itens e 23 licitações. O ÓRGÃO_2 possui 526 itens e 13 licitações. O ÓRGÃO_11 possui 170 itens e 7 licitações. O ÓRGÃO_10 possui 145 itens e 11 licitações. O ÓRGÃO_15 possui 135 itens e 3 licitações. O ÓRGÃO_3 possui 132 itens e 6 licitações. O ÓRGÃO_13 possui 74 itens e 5 licitações. O ÓRGÃO_6 possui 65 itens e 12 licitações. O ÓRGÃO_8 possui 10 itens e 9 licitações. O ÓRGÃO_12 possui 4 itens e 2 licitações. Observa-se grandes quantidades nos 3 primeiros órgãos, sendo que nos demais a quantidade se torna menor e mantém uma linearidade.

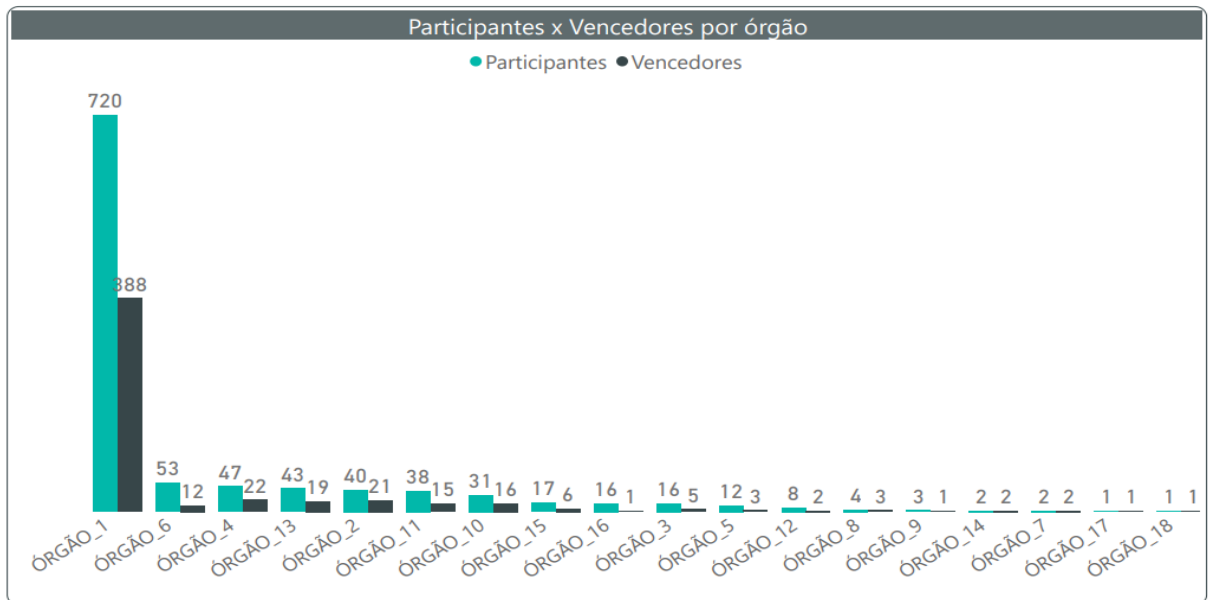
A Figura 24 aborda o número de participantes e o número de vencedores por órgão. O ÓRGÃO_1 lidera o ranking com 720 participantes e 388 vencedores. O ÓRGÃO_6 em seguida com 53 participantes e 12 vencedores. O ÓRGÃO_4 com 47 participantes e 22 vencedores. Os demais órgãos possuem números de participações e vitórias menores, tendendo a único participante e um único vencedor.

Figura 23. Número de itens de licitações e número de licitações por órgão



Fonte: Elaborado pelo autor.

Figura 24. Quantidade de participantes e vencedores por órgão



Fonte: Elaborado pelo autor.

A Tabela 4 ilustra a média de participantes por órgão. A Tabela 5 traz uma visão do valor estimado e valor total por órgão. A Figura 25 ilustra os 4 maiores vencedores por órgão. A partir da Figura 25 conseguimos mapear as empresas que mais tem vencido, se vence sempre no mesmo órgão ou em órgãos diferentes, um alerta para comparações na próxima etapa, análise preditiva.

Tabela 4. Média de participantes por órgão

Média de participantes por órgão	
Órgão	Média
ÓRGÃO_16	16,00
ÓRGÃO_13	8,60
ÓRGÃO_5	6,00
ÓRGÃO_15	5,67
ÓRGÃO_11	5,43
ÓRGÃO_6	4,42
ÓRGÃO_12	4,00
ÓRGÃO_2	3,08
ÓRGÃO_9	3,00
ÓRGÃO_10	2,82
ÓRGÃO_3	2,67
ÓRGÃO_4	2,04
ÓRGÃO_14	1,00
ÓRGÃO_17	1,00
ÓRGÃO_7	1,00
ÓRGÃO_18	0,50
ÓRGÃO_8	0,44
ÓRGÃO_1	0,33
Total	68,00

Fonte: Elaborado pelo autor.

Tabela 5. Valor estimado e valor total por órgão

Valor estimado X Valor total por órgão		
Órgão	Valor estimado	Valor total
ÓRGÃO_1	23.700.336.834,95	2.709.763.605,62
ÓRGÃO_4	684.771.537,24	6.093.099,72
ÓRGÃO_10	280.120.732,27	6.777.116,97
ÓRGÃO_8	122.352.577,86	102.352.577,86
ÓRGÃO_13	97.318.818,44	6.334.082,01
ÓRGÃO_2	56.889.021,04	608.866,14
ÓRGÃO_11	34.076.152,78	2.345.307,28
ÓRGÃO_6	24.304.042,54	17.854.442,38
ÓRGÃO_15	24.223.310,99	566.804,73
ÓRGÃO_3	8.110.402,00	266.634,00
ÓRGÃO_12	7.813.500,00	2.665.500,00
ÓRGÃO_16	3.019.796,64	1.509.898,32
ÓRGÃO_18	2.000.000,00	2.000.000,00
ÓRGÃO_17	112.428,00	37.476,00
ÓRGÃO_5	95.324,70	89.662,35
ÓRGÃO_7	84.000,00	36.000,00
ÓRGÃO_9	35.997,96	35.997,96
ÓRGÃO_14	17.180,00	17.180,00
Total	25.045.681.657,41	2.859.354.251,34

Fonte: Elaborado pelo autor.

Figura 25. 4 maiores empresas vencedoras por órgão

4 maiores empresas vencedoras por órgão				
Órgão	1ª 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

Fonte: Elaborado pelo autor.

4.3 Análise preditiva

A análise descritiva uma vez finalizada, observa-se 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, foram utilizados 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:

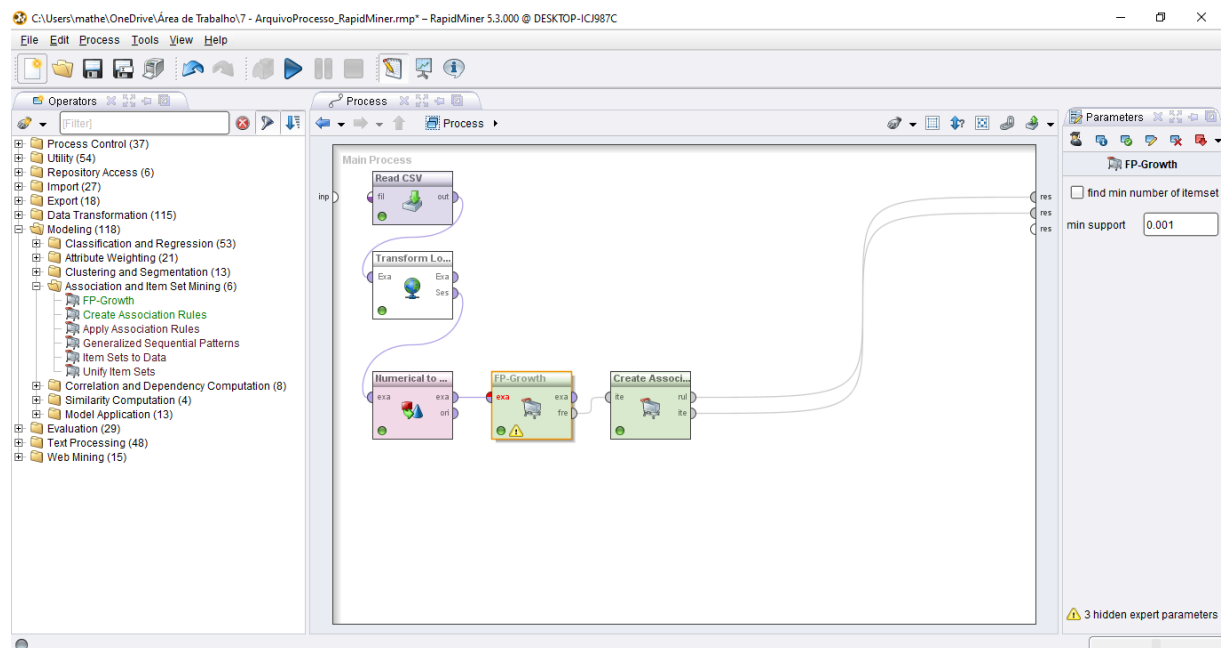
- Etapa 1 – Aplicação de Regra de Associação
- Etapa 2 – Aplicação de Clusterização
- Etapa 3 – Aplicação de Regra de Associação

Conforme Grilo Júnior (2010), 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 informe-os. 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.

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 26. 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 26. Área de trabalho da ferramenta RapidMiner



Fonte: 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 7. Esse *dataset* foi composto por 5555 linhas fora o cabeçalho, distribuídas nas 2254 licitações e nos 846 participantes, ou seja, cada linha representando uma licitação, 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 7. Dataset 1 contendo as licitações e empresas

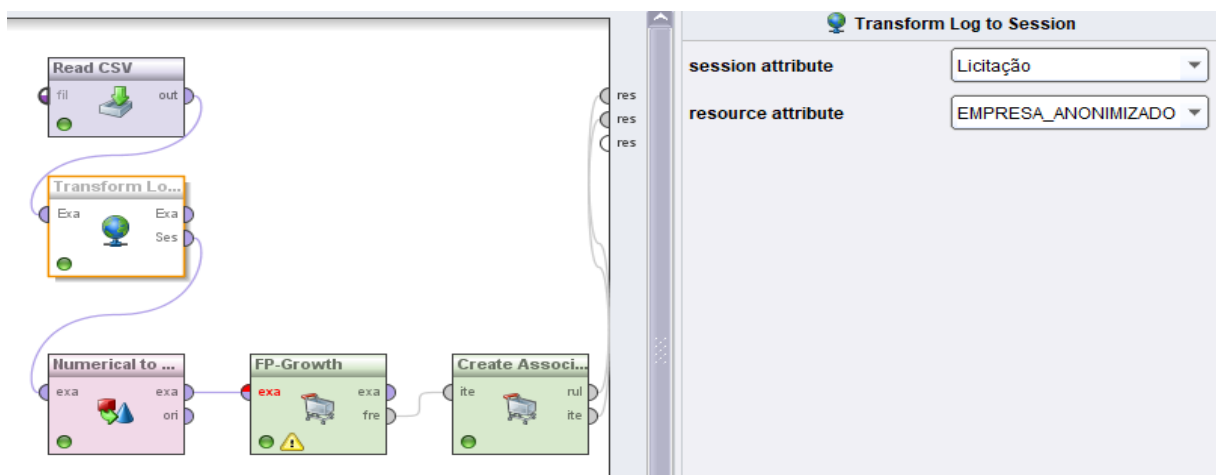
Licitação	Empresa
LICITAÇÃO_1	EMPRESA_1
LICITAÇÃO_1	EMPRESA_2
LICITAÇÃO_2	EMPRESA_3
...	...
LICITAÇÃO_n	EMPRESA_n

Fonte: Elaborado pelo autor.

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 26. 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 27.

Figura 27. Configurações dos parâmetros do operador *Transform log to session*

Fonte: 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 Binominal*, gerando os dados similar aos do Quadro 8.

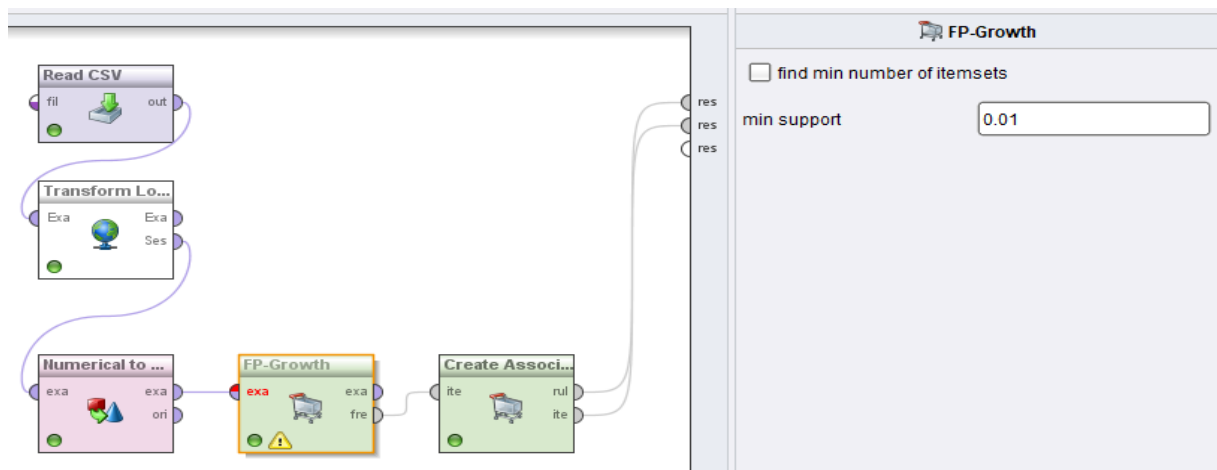
Quadro 8. Matriz esparsa das licitações e empresas

Licitações	EMPRESA_1	EMPRESA_2	EMPRESA_3	EMPRESA_n
LICITAÇÃO_1	<i>true</i>	<i>true</i>	<i>true</i>	...
LICITAÇÃO_2	<i>false</i>	<i>true</i>	<i>false</i>	...
LICITAÇÃO_3	<i>false</i>	<i>true</i>	<i>true</i>	...
LICITAÇÃO_n

Fonte: Elaborado pelo autor.

Depois de utilizar o operador *Numerical to Binominal* foi possível fazer uso do operador FP-Growth para a identificação dos itens que são frequentes. A Figura 28 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 28. Configurações do parâmetro do operador FP-Growth

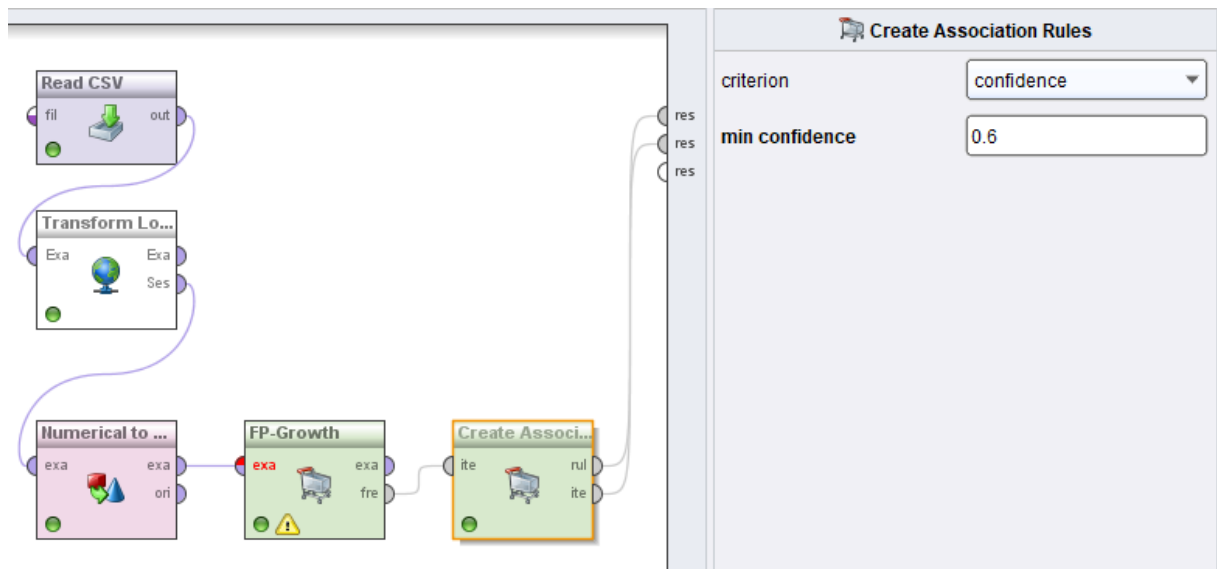


Fonte: Elaborado pelo autor.

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 29 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 29. Configurações do parâmetro do operador *Create Association Rules*



Fonte: 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 30.

Figura 30. Resultado do processamento dos dados minerados

No.	Premises	Conclusion	Support	Confid...	LaPla...	Gain	p-s	Lift	Convi...
1	EMPRESA_58	EMPRESA_32	0.012	0.605	0.993	-0.027	0.008	3.221	2.055
2	EMPRESA_24	EMPRESA_36	0.025	0.606	0.984	-0.058	0.022	8.539	2.360
3	EMPRESA_56	EMPRESA_20	0.028	0.643	0.985	-0.059	0.022	4.526	2.402
4	EMPRESA_67	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	15.32	2.870
9	EMPRESA_32, EMPRESA_56	EMPRESA_20	0.016	0.686	0.993	-0.030	0.012	4.832	2.735
10	EMPRESA_214	EMPRESA_20	0.016	0.692	0.993	-0.030	0.013	4.874	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	4.976
16	EMPRESA_56, EMPRESA_214	EMPRESA_20	0.011	0.857	0.998	-0.014	0.009	6.035	6.006

Fonte: 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 antecedente 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 Quadro 9. 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.

Quadro 9. Resultados execução do algoritmo FP-Growth para os dois suportes

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

Fonte: Elaborado pelo autor.

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 mercados 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 mercado 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 10. Esse *dataset* foi composto por todos os itens da base mencionada acima, contendo 6724 registros, ou seja, cada linha representando licitações e os itens, caso a licitação possua mais de um item.

Quadro 10. *Dataset 2* para realizar a Clusterização

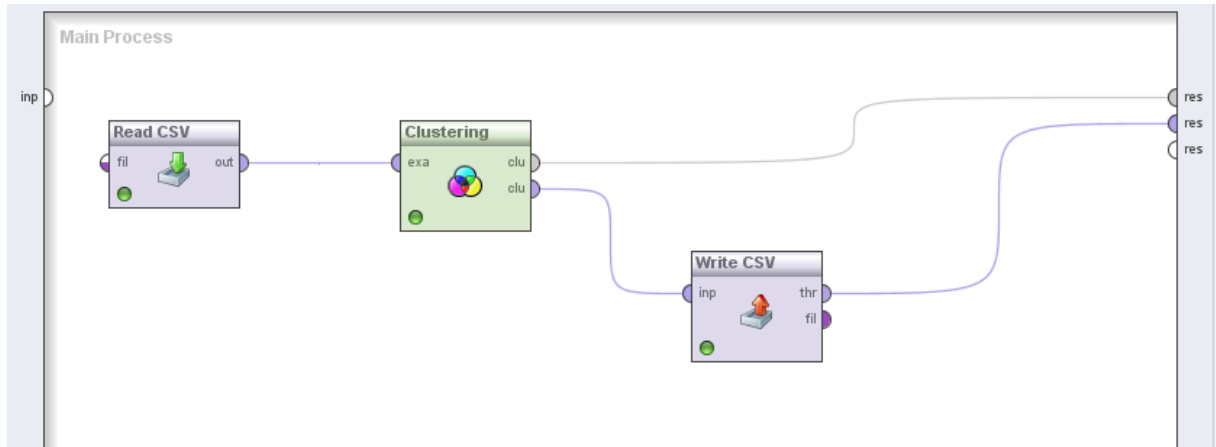
Licitação	Empresa	Órgão
LICITAÇÃO_1	EMPRESA_1	ÓRGÃO_1
LICITAÇÃO_1	EMPRESA_2	ÓRGÃO_2
LICITAÇÃO_2	EMPRESA_3	ÓRGÃO_3
...
LICITAÇÃO_n	EMPRESA_n	ÓRGÃO_n

Fonte: Elaborado pelo autor.

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 31. 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 11. Para finalizar o processo, a segunda saída é conectada no

operador WriteCSV, operador que permite exportar os dados gerados em um arquivo CSV.

Figura 31. Processo para realizar a Clusterização no RapidMiner



Fonte: Elaborado pelo autor.

Quadro 11. 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

Fonte: Elaborado pelo autor.

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 no Quadro 12.

Quadro 12. Quantidade de itens por *clusters* gerados

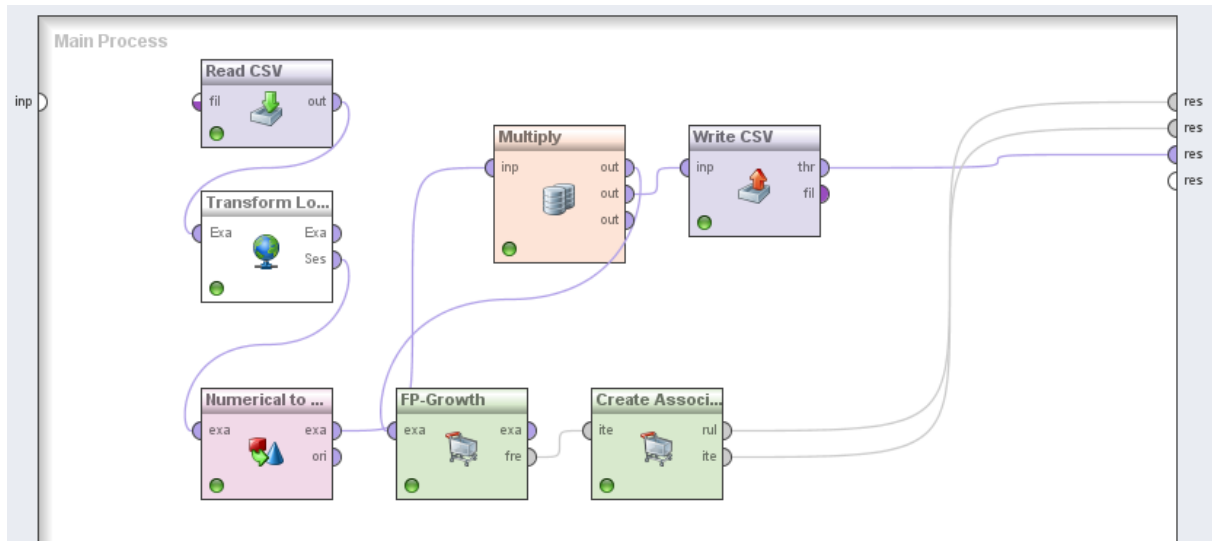
Cluster	Quantidade de itens
CLUSTER_0	963
CLUSTER_1	927
CLUSTER_2	1280
CLUSTER_3	870
CLUSTER_4	966
CLUSTER_5	850
CLUSTER_6	867
Total	6723

Fonte: Elaborado pelo autor.

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*.

Foram gerados 7 (sete) *datasets*, um para cada *cluster*, por meio do processo ilustrado na Figura 32, 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 7. O Quadro 13 mostra as principais informações obtidas ao aplicar associação em cada *cluster*.

Figura 32. Processo para realizar a Regra de Associação em cada *cluster* obtido



Fonte: Elaborado pelo autor.

Quadro 13. Execução do algoritmo de associação FP-Growth para os *datasets* de *clusters*

Cluster	Licitações	Atributos	Suporte mínimo	Confiança mínima	Itens frequentes	Regras
CLUSTER_0	340	251	1%	60%	120	160
CLUSTER_1	296	274	1%	60%	217	600
CLUSTER_2	398	307	1%	60%	428	1771
CLUSTER_3	321	203	1%	60%	94	46
CLUSTER_4	314	245	1%	60%	107	31
CLUSTER_5	269	286	1%	60%	140	139
CLUSTER_6	316	182	1%	60%	158	274

Fonte: Elaborado pelo autor.

Posteriormente, a partir do “DATAFRAME 2 – licitações” foi gerado um *dataset* dos vencedores, contendo os atributos de licitação e empresa vencedora, em seguida, gerado uma matriz esparsa, similar ao apresentado no Quadro 8. 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.

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 licitações e celebraram contrato juntas em 8 licitações. Das 11 licitações que participaram juntas, a EMPRESA_2 venceu sozinha em apenas 1 licitação, e a EMPRESA_36 venceu sozinha em 2 licitações. Essas duas empresas possuíam um número relativamente baixo de participações no *cluster*, a EMPRESA_2 contém 17 participações e a EMPRESA_36 contém 25 participações nesse *cluster*, na base toda a EMPRESA_2 possui 100 participações e a EMPRESA_36 possui 160 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 licitações e celebrou contrato em 2 licitações. Das 5 licitações que o grupo participou, a EMPRESA_20 venceu as 5 licitações, a EMPRESA_63 venceu 4 licitações e a EMPRESA_12 venceu 3 licitações. A EMPRESA_20 possui 48 participações, a EMPRESA_63 possui 26 participações e a EMPRESA_12 contém 43 participações nesse *cluster*. Na base toda, as EMPRESA_20, EMPRESA_63 e EMPRESA_12 possuem respectivamente 320, 148 e 276 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 e 88 respectivamente, sendo que, entre essas, elas participaram juntas de 18 licitações e ganharam juntas 9 licitações. Vale ressaltar que, das 20 licitações que a EMPRESA_43 participou, a EMPRESA_32 não participou somente de 2 licitações.

Esse estudo 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.

5. CONSIDERAÇÕES FINAIS

A investigação envolvendo a ciência de dados, mais especificamente mineração de dados é muito relevante, porque o volume de dados aumenta exponencialmente, e é necessário saber extrair informações desses dados. Sendo assim, a existência de uma ciência dedicada a isso se torna indispensável.

Nesse estudo foi abordado um dos instrumentos da Administração Pública mais importantes, porque é com esse instrumento que órgãos e entidades públicas realizam seus negócios de forma transparente e eficiente. Como envolve também dinheiro público, o abuso de agentes públicos e privados para obtenção de vantagens pessoais ocorre, e não somente em território nacional, mas em todo o mundo.

Devido ao número de colaboradores dedicados para essa atividade de fiscalização não ser suficiente, se torna necessário a utilização de ferramentas e técnicas computacionais para auxiliar nesse processo de forma mais automática. Uma dessas técnicas é a mineração de dados, que é capaz de produzir conhecimento a partir de grandes volumes de dados, como são as bases de dados relacionadas aos processos de licitação.

A mineração de dados está inserida dentro de um processo denominado *Knowledge Discovery in Databases* (KDD), ou Descoberta de conhecimento em bancos de dados, sendo esse o processo que identifica novos padrões que sejam válidos, e que antes eram desconhecidos. O KDD possui 5 (cinco) etapas: seleção de dados; limpeza dos dados; tratamento de dados; mineração de dados; interpretação.

Nesse estudo foi utilizado o processo de mineração de dados dividindo esse processo em 4 (quatro) etapas, que são: escolha da tarefa de mineração de dados; escolha da técnica de mineração de dados; escolha do algoritmo a ser utilizado; aplicação do processo de mineração de dados.

Dessa forma, essa pesquisa teve como objetivo geral investigar a contribuição da aplicação de mineração de dados nos processos licitatórios utilizando dados no portal da CGE/GO com vistas a identificar e relacionar possíveis irregularidades. Para atendê-lo foram propostos os seguintes objetivos específicos: desenvolver uma investigação através de uma pesquisa exploratória para conhecer os fatos e fenômenos relacionado a licitações públicas; explorar os conceitos de *data science* e técnicas de mineração de dados; 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; e reconhecer entidades conforme recorrências em grupos suspeitos, levando em conta a tendenciosa conquista, concentração em órgãos específicos e a presença de concorrentes fictícios.

Procurou-se atingir os objetivos, geral e específicos, por meio de pesquisa bibliográfica e experimental, voltados a responder a seguinte pergunta de pesquisa: a implantação de mineração de dados pode contribuir na identificação de possíveis irregularidades em processos licitatórios analisados pela CGE/GO?

As considerações finais decorrentes desse estudo, e a resposta para a questão mencionada acima, são apresentadas a seguir, devidamente norteadas pelos objetivos geral e específicos propostos.

Considerando o objetivo de desenvolver uma investigação através de uma pesquisa exploratória para conhecer os fatos e fenômenos relacionado a licitações públicas, esse estudo buscou na literatura e em trabalhos relacionados os principais conceitos relacionados as licitações públicas no Brasil, dando uma visão geral sobre o propósito e funcionamento do processo licitatório, abordando as modalidades, tipos, problemas com esse ato e características que são comuns na prática de conluio.

Em relação ao item de explorar os conceitos de *data science* e técnicas de mineração de dados, buscou-se trabalhar com autores renomados na área, utilizando livros e trabalhos científicos recentes. Foi apresentado a relação de inteligência artificial, aprendizado de máquina, *business intelligence*, ciência de dados, análise descritiva e análise preditiva. Exposto sobre a descoberta do conhecimento em bases de dados, conceitos sobre *data mining*, técnicas e algoritmos para cada tipo de tarefa da mineração de dados.

Quanto ao item para 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, foram examinados trabalhos importantes relacionados, em que os autores obtiveram bons resultados, comparando os métodos abordados e aplicando uma análise descritiva nos dados disponíveis selecionados no portal da transparência da CGE/GO, para a identificação das variáveis e da melhor técnica a ser utilizada nesse problema.

Em relação ao item para reconhecer entidades conforme recorrências em grupos suspeitos, levando em conta as informações dos vencedores, concentração de vencedores em órgãos específicos e a presença de concorrentes fictícios, observa-se que através da análise descritiva e análise preditiva muitas informações e

correlações que antes eram desconhecidas e/ou dificilmente encontradas aparecem, o que proporciona um ganho considerável de tempo nas análises iniciais de licitações. A utilização de regras de associação e de clusterização foram utilizadas neste processo.

Sendo assim, dadas as evidências obtidas nesse estudo, conclui-se que com a utilização de métodos adequados de mineração de dados, pode contribuir na análise e identificação de possíveis irregularidades em processos licitatórios analisados 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.

5.1 Recomendações para trabalhos futuros

Essa pesquisa apresentou aspectos preliminares para contribuições da mineração de dados nos processos licitatórios, podendo expandir e realizar outros testes, objetivando o controle eficiente dos recursos públicos, reduzindo a prática de eventos irregulares nas licitações públicas.

Para trabalhos futuros sugere-se:

- Testar a metodologia em outros órgãos que realizam controle de fiscalização, para verificar se a utilização de mineração de dados contribui positivamente;
- Criação de um *data warehouse* para que se possa armazenar os dados de forma histórica, para a facilitação nas comparações de participantes e vencedores e obtenção de informações específicas;
- Desenvolvimento de um método para melhor escolha das regras geradas;
- Aplicação de novos métodos utilizando novas variáveis.

REFERÊNCIAS

ANTONELLI, Ricardo Adriano. Conhecendo o Business Intelligence (BI): uma ferramenta de auxílio à tomada de decisão. **Revista Tecap**, Rio de Janeiro, v. 3, n. 3, p. 79-85, 2009.

BATISTA, Gustavo Enrique de Almeida Prado Alves. **Pré-processamento de Dados em Aprendizado de Máquina Supervisionado**. 2003. 232 f. Tese (Doutorado) - Curso de Ciências - Ciências de Computação e Matemática Computacional, Universidade de São Paulo, São Carlos, 2003.

BENTLEY, Drew. **Business Intelligence and Analytics**. New York: Larsen And Keller Education, 2017.

BRAGA, Tereza Cristine Almeida. CADE, Cartéis e Licitações: um Novo Nicho da Política Antitruste Brasileira. **Revista de Defesa da Concorrência**, Brasília, v. 3, n. 1, p. 108-132, maio 2015.

BRASIL. Constituição (1988). Lei Complementar nº101, de 4 de maio de 2000. Estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências. **Diário Oficial da União**. Brasília, DF, 4 de maio de 2000.

BRASIL. Constituição (1988). Lei Nº 8.666, de 21 De Junho de 1993. Regulamenta o art. 37, inciso XXI, da Constituição Federal, institui normas para licitações e contratos da Administração Pública e dá outras providências. **Diário Oficial da União**. Brasília, DF, 21 de junho de 1993.

BRASIL. Constituição (1988). Lei Nº 10.520, De 17 De Julho De 2002. Institui, no âmbito da União, Estados, Distrito Federal e Municípios, nos termos do art. 37, inciso XXI, da Constituição Federal, modalidade de licitação denominada pregão, para aquisição de bens e serviços comuns, e dá outras providências. **Diário Oficial da União**. Brasília, DF, 17 de julho de 2002.

BRASIL. **RELATÓRIO TC 016.119/2016-9**. Brasília: Tcu, 2016.

CABENA P. *et al.* **Discovering data mining: from concept to implementation**. Englewood Cliffs: Prentice Hall, 1998.

CAMPOS, Francisco. As práticas de conluio nas licitações públicas à luz da teoria dos jogos. **Revista Análise Econômica**, Porto Alegre, v. 1, n. 50, p. 185-206, set. 2008.

CARVALHO, Daniel Bulha de. As influências da Lei de Responsabilidade Fiscal nas Licitações e Contratos Administrativos. **Revista do Tcu**, Brasília, v. 1, n. 118, p. 25-34, 2010.

CARVALHO, Lorena Moreira Seal. **O Cartel na Licitação Pública**. 2018. 115 f. Dissertação (Mestrado) - Curso de Ciências Jurídico-políticas Com Menção em Direito Administrativo, Universidade de Coimbra, Coimbra, 2018.

CASTRO, Fabiano de. **A CORRUPÇÃO NO ORÇAMENTO: FRAUDES EM LICITAÇÕES E CONTRATOS COM O EMPREGO DE EMPRESAS INIDÔNEAS**. 2010. 42 f. Monografia (Especialização) - Curso de Orçamento Público, Instituto Serzedello Corrêa Tcu, Brasília, 2010.

CÔRTEZ, Sérgio da Costa; PORCARO, Rosa Maria; LIFSCHITZ, Sérgio. **Mineração de Dados - Funcionalidades, Técnicas e Abordagens**. 2002. Disponível em: ftp://obaluae.inf.puc-rio.br/pub/docs/techreports/02_10_cortes.pdf. Acesso em: 26 maio 2020.

CORRUPÇÃO: custos econômicos e propostas de combate. FIESP, 2010. Disponível em: <https://www.fiesp.com.br/arquivo-download/?id=2021>. Acesso em 15 de maio de 2020.

DEVMEDIA. **Mineração de dados com Market Basket Analysis**. 2020. Disponível em: <https://www.devmedia.com.br/mineracao-de-dados-com-market-basket-analysis-revista-sql-magazine-111/27853>. Acesso em: 31 out. 2020.

DI PIETRO, Maria Sylvia Zanella. **Direito Administrativo**. 32. ed. Rio de Janeiro: Forense, 2019.

DIRETRIZES para combater o conluio entre concorrentes em contratações públicas. **OECD**, 2009. Disponível em: <https://www.oecd.org/daf/competition/cartels/44162082.pdf>. Acesso em 15 de maio de 2020.

FAYYAD, U.; SHAPIRO, G. P.; SMYTH P. **From data mining to knowledge discovery in databases**. *AI Magazine*, 17(3): 37-54, Fall 1996.

FONSECA, Felipe Cesar Stanzani; BELTRAME, Walber Antonio Ramos. **Aplicações Práticas dos Algoritmos de Clusterização Kmeans e Bisecting K-means**. 2010. Disponível em: <https://www.researchgate.net/publication/327121358>. Acesso em: 18 nov. 2020.

FORTINI, Cristiana; MOTTA, Fabrício. Corrupção nas licitações e contratações públicas: sinais de alerta segundo a Transparência Internacional. **Revista de Direito Administrativo & Constitucional**, Belo Horizonte, v. 1, n. 64, p. 93-113, jun. 2016.

FRANKENFIELD, Ake. **Descriptive Analytics**. 2019. Disponível em: <https://www.investopedia.com/terms/d/descriptive-analytics.asp#:~:text=Descriptive%20analytics%20is%20the%20interpretation,historic%20data%20to%20draw%20comparisons.&text=These%20measures%20all%20describe%20what,business%20during%20a%20set%20period..> Acesso em: 31 out. 2020.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 6. ed. São Paulo: Atlas, 2017.

GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel; BEZERRA, Eduardo. **Data Mining: conceitos, técnicas, algoritmos, orientações e aplicações**. 2. ed. Rio de Janeiro: Elsevier, 2015.

GONÇALVES, Lóren Pinto Ferreira. **Mineração de dados em supermercados: o caso do supermercado "tal"**. 1999. 36 f. Dissertação (Mestrado) - Curso de Administração, Universidade Federal do Rio Grande do Sul, Porto Alegre, 1999.

GUILLET, F; HAMILTON, H. J. **Quality Measures in Data Mining**. Studies in Computational Intelligence 43, Springer, 2007.

GRILO JÚNIOR, Tarcísio Ferreira. **Aplicação de Técnicas de Data Mining para Auxiliar no Processo de Fiscalização no Âmbito do Tribunal de Contas do Estado da Paraíba**. 2010. 103 f. Dissertação (Mestrado) - Curso de Engenharia de Produção, Centro de Tecnologia, Universidade Federal da Paraíba, João Pessoa, 2010.

HAN, J.; KAMBER, M. **Data mining: concepts and techniques**. New York: Morgan Kaufmann, 2000.

HAN, J.; PEI, J.; YIN, Y. **Mining Frequent Patterns Without Candidate Generation**, In: Inf. Conf. on Management of Data (SIGMOD'00), Dallas, Texas, p. 1-12, 2000.

IBM. **Intelligent Miner Visualizer**. Disponível em: https://www.ibm.com/support/knowledgecenter/en/SSEPGG_10.5.0/com.ibm.im.overview.doc/c_ibm_db2_intelligent_miner_visualization.html. Acesso em: 31 out. 2020.

ÍNDICE De Percepção Da Corrupção 2019. **TRANSPARÊNCIA INTERNACIONAL**, 2019. Disponível em: <https://transparenciainternacional.org.br/ipc/>. Acesso em: 22 maio 2020.

KOTU, Vijay; DESHPANDE, Bala. **Data Science: concepts and practice**. 2. ed. Cambridge: Elsevier, 2019.

MORAIS, Cirilo Max Macedo de. **Proposição De Indicadores Para Investigação De Licitações Por Meio De Técnicas De Reconhecimento De Padrões Estatísticos E Mineração De Dados**. 2016. 146 f. Dissertação (Mestrado) - Curso de Informática Forense e Segurança da Informação, Universidade de Brasília, Brasília, 2016.

MOURÃO, Roberto Nunes. **Mineração de Dados para Previsão de Renda de Clientes com Contas-Correntes Digitais**. 2018. 70 f. Dissertação (Mestrado) - Curso de Computação Aplicada, Universidade de Brasília, Brasília, 2018.

NANDI, Júlio César Borba; PEREIRA, Ruano Marques; FELIPPE, Gabriel. O Algoritmo de Associação Frequent Pattern-Growth na Shell Orion Data Mining Engine. **Anais SULCOMP**, v. 7, 2015.

NASCIMENTO, Francisco Paulo do; SOUSA, Flávio Luís Leite. **Metodologia da Pesquisa Científica: Teoria e Prática**. Brasília: Thesaurus, 2015.

OECD Work on Anti-corruption and Integrity. **OECD**. Disponível em: <https://www.oecd.org/corruption-integrity/about/>. Acesso em: 15 de maio de 2020.

OPERAÇÃO Lava Jato. **Folha De São Paulo**. Disponível em: <http://arte.folha.uol.com.br/poder/operacao-lava-jato/>. Acesso em: 15 maio 2020.

PINHEIRO, André Alves; MELO, Telma Maria Escóssio. **O Portal De Licitações Dos Municípios Do TCM/CE Sob A Luz Do Princípio Da Transparência**. Brasília, 2016.

QUEIROZ, Altamira de Souza. **Algoritmos de inteligência computacional utilizados na detecção de fraudes nas redes de distribuição de energia elétrica**. 2016. 86 f. Dissertação (Mestrado) - Curso de Engenharia Elétrica e Computação, Universidade Estadual do Oeste do Paraná, Foz do Iguaçu, 2016.

RAPIDMINER. **RapidMiner Best Data Science & Machine Learning PLataform**. Disponível em: <https://rapidminer.com/>. Acesso em: 31 out. 2020.

RODRIGUES, Fabiene Silva. **Métodos de agrupamento na análise de dados de expressão gênica**. 2009. 95 f. Dissertação (Mestrado) - Curso de Estatística, Universidade Federal de São Carlos, São Carlos, 2009.

SAS. **Software de Business Analytics e Business Intelligence**. Disponível em: https://www.sas.com/pt_br/home.html. Acesso em: 31 out. 2020.

SHARDA, Ramesh; DELEN, Dursun; TURBAN, Efraim. **BUSINESS INTELLIGENCE, ANALYTICS, AND DATA SCIENCE: a managerial perspective**. 4. ed. Boston: Pearson, 2018.

SILVA, Carlos Vinícius Sarmiento; RALHA, Célia Ghedini. **Detecção de cartéis em licitações públicas com agentes de mineração de dados**. Revista Eletrônica de Sistemas de Informação, Curitiba, v. 10, n. 1, 2011. Disponível em: https://repositorio.unb.br/bitstream/10482/12315/1/ARTIGO_Deteccao%20Carteis%20Licitacoes.pdf Acesso: 15 mai. 2020. doi:10.5329/RESI.2011.1001008

SILVA, Marcelino Pereira dos Santos. **SKDQL: uma linguagem declarativa de especificação de consultas e processos para descoberta de conhecimento em bancos de dados e sua implementação**. 2002. 113 f. Dissertação (Mestrado) - Curso de Ciência da Computação, Centro de Informática, Universidade Federal de Pernambuco, Recife, 2002.

SOARES JUNIOR, Jair Sampaio; QUINTELLA, Rogério Hermida. Descoberta de conhecimento em bases de dados públicas: uma proposta de estruturação metodológica. **Revista de Administração Pública (RAP)**, Rio de Janeiro, v. 1, n. 1, p. 1077-1107, out. 2005.

SOUSA, Maria Cristina Cordeiro. **UMA ANÁLISE DO ALGORITMO K-MEANS COMO INTRODUÇÃO AO APRENDIZADO DE MÁQUINAS**. 2019. 75 f. TCC (Graduação) - Curso de Matemática, Universidade Federal do Tocantins, Araguaína, 2019.

SOUZA, Audrey Jones de; MENDES, Raphael Borges; BRAGA, Jefferson Ribeiro Bastos. **Laudo de Perícia Criminal (Contábil-Financeira) Nº 2311/2015-SETEC/SR/DPF/PR**, 2015.

SOUZA, Fatima Regina de. **Manual básico de licitação**. São Paulo: Nobel, 1997.

TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. **Introdução ao Data Mining**. Rio de Janeiro: Editora Ciência Moderna, 2009.

TATSCH, Dante Rodrigo Delibero. **CARACTERIZAÇÃO DOS CARTÉIS NO BRASIL**. 2012. 36 f. Monografia (Especialização) - Curso de Economia, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2012.

WAZLAWICK, Raul Sidnei. **Metodologia de Pesquisa para Ciência da Computação**. 2. ed. Rio de Janeiro: Elsevier, 2014.

WEKA. **WEKA The workbench for machine learning**. Disponível em: <https://www.cs.waikato.ac.nz/ml/weka/>. Acesso em: 31 out. 2020.

WITTEN, Ian H. *et al.* **Data Mining: practical machine learning tools and techniques**. 4. ed. Cambridge: Elsevier, 2017.