

Marcelo Abelha Peixoto Gomes

**Extração e internalização automatizada de dados
de licitações realizadas pelo portal Petronect**

Brasília

2020

Marcelo Abelha Peixoto Gomes

**Extração e internalização automatizada de dados
de licitações realizadas pelo portal Petronect**

Trabalho de conclusão do curso de pós-graduação lato sensu em Análise de Dados para o Controle realizado pela Escola Superior do Tribunal de Contas da União como requisito para a obtenção do título de especialista.

Orientador: Prof. Dr. Edans Flavius de Oliveira Sandes

Brasília

2020

REFERÊNCIA BIBLIOGRÁFICA

GOMES, Marcelo Abelha Peixoto. Extração e internalização automatizada de dados de licitações realizadas pelo portal Petronect. 2019. Trabalho de Conclusão de Curso (Especialização em 2020) – Escola Superior do Tribunal de Contas da União, Instituto Serzedello Corrêa, Brasília DF. 200 fl.

CESSÃO DE DIREITOS

NOME DO AUTOR: Marcelo Abelha Peixoto Gomes

TÍTULO: Extração e internalização automatizada de dados de licitações realizadas pelo portal Petronect

GRAU/ANO: Especialista/2020

É concedido ao Instituto Serzedello Corrêa (ISC) permissão para reproduzir cópias deste Trabalho de Conclusão de Curso e emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. Do mesmo modo, o ISC tem permissão para divulgar este documento em biblioteca virtual, em formato que permita o acesso via redes de comunicação e a reprodução de cópias, desde que protegida a integridade do conteúdo dessas cópias e proibido o acesso a partes isoladas desse conteúdo. O autor reserva outros direitos de publicação e nenhuma parte deste documento pode ser reproduzida sem a autorização por escrito do autor.

Marcelo Abelha Peixoto Gomes
marcelo.abelha@gmail.com

Ficha catalográfica

<p>Gomes, Marcelo Abelha Peixoto Extração e internalização automatizada de dados de licitações realizadas pelo portal Petronect / Marcelo Abelha Peixoto Gomes ; orientador, Edans Flavius de Oliveira Sandes , 2020. 80 p. Trabalho de Conclusão de Curso (pós - graduação) - Instituto Serzedello Corrêa, Análise de Dados para o Controle, Brasília, 2020. Inclui referências.</p>

1. Controle externo. 2. Mineração de dados. 3. Tribunal de Contas da União. 4. Petrobras. I. Edans Flavius de Oliveira Sandes. II. Instituto Serzedello Corrêa. Graduação em Administração. III. Título.

Marcelo Abelha Peixoto Gomes

Extração e internalização automatizada de dados de licitações realizadas pelo portal Petronect

Trabalho de conclusão do curso de pós-graduação lato sensu em Análise de Dados para o Controle realizado pela Escola Superior do Tribunal de Contas da União como requisito para a obtenção do título de especialista.

Brasília, 01 de março de 2020.

Banca Examinadora:

Prof. Edans Flavius de Oliveira Sandes, Dr.

Orientador

Prof. Saul Campos Berardo, Me.

RESUMO

O Tribunal de Contas da União, órgão auxiliar do Congresso Nacional para o exercício do controle externo, tem entre suas competências fiscalizar os editais de licitação, contratos e convênios praticados por responsáveis sujeito à sua jurisdição.

O objetivo do trabalho foi o de melhorar a atuação da Secretaria de Fiscalização de Infraestrutura de Petróleo e Gás Natural, secretaria especializada da Corte de Contas, por meio da disponibilização de informações acerca do processo licitatório de contratações e aquisições realizadas pelas empresas do Grupo Petrobras conduzidas pelo portal eletrônico Petronect.

Para atingir o objetivo estabelecido foi desenvolvido software para automatizar a extração e internalização dos dados do portal Petronect para banco de dados do Tribunal. Foi criado painel eletrônico para consulta das informações extraídas e disponibilizado ao sistema de Análise de Licitações e Editais (Alice) os documentos relacionados as licitações para sua análise.

Palavras-chave: Controle externo. Petronect. Mineração de dados. Tribunal de Contas da União. Petrobras.

ABSTRACT

The Tribunal de Contas da União, an auxiliary body of the National Congress for the exercise of accountability, has among its powers to supervise the bidding notices, contracts and agreements practiced by those responsible under its jurisdiction.

The objective of the work was to improve the performance of Secretaria de Fiscalização de Infraestrutura de Petróleo e Gás Natural by providing information about the bidding process for contracts and acquisitions carried out by the Petrobrás Group companies conducted in the Petronect electronic portal.

In order to achieve the established objective a software was developed to automate the extraction and internalization of data from the Petronect portal to the Court's database and an electronic panel was created to consult the extracted information. Furthermore, the documents related to the bids were made available to the Análise de Licitações e Editais system (Alice) for analysis.

Keywords: Accountability. Petronect. Data Mining. Tribunal de Contas da União. Petrobras.

LISTA DE ILUSTRAÇÕES

Figura 1 - Macrofluxo de compras e contratações - Petronect.....	17
Figura 2 - Fases do CRISP-DM	20
Figura 3 - Portal Petronect - Licitações públicas	25
Figura 4 - Disputas publicadas	26
Figura 5 - Detalhamento da licitação	26
Figura 6 - Itens da licitação	27
Figura 7 - Anexos da licitação.....	27
Figura 8 - Captcha - Anexos.....	28
Figura 9 - Código-fonte da página - DevTools	29
Figura 10 - Requisições - DevTools.....	30
Figura 11 - Requisição - Licitações publicadas - DevTools	30
Figura 12 - Detalhes da requisição de licitações publicadas	31
Figura 13 - Exploração de dados - Percentual de empresas contratantes das licitações	24
Figura 14 - Exploração de dados - Percentual de licitações por modos de disputa ...	25
Figura 15 - Exploração de dados - Modalidade	26
Figura 16 - Exploração de dados - Família dos itens	26
Figura 17 - Exploração de dados - Descrição dos itens das licitações	27
Figura 18 - Exploração de dados - Tipos dos itens	27
Figura 19 - Exploração de dados - Empresas que mais participaram	28
Figura 20 - Quantidade de cotações por porte da empresa.....	28
Figura 21 - Exploração de dados - Classificação das cotações	29
Figura 22 - Exploração de dados - Classificação das cotações - Licitações concluídas	30
Figura 23 - Exploração de dados - Maiores cotações.....	30
Figura 24 - Exploração de dados - Valores das cotações homologadas.....	31
Figura 25 - Exploração de dados - Quantidade de lances por pregão	32
Figura 26 - Exploração de dados - Frequência de licitações pela quantidade de lances.	33
Figura 27 - Exploração de dados - Número de empresas participantes em cada licitação	34
Figura 28 - Exploração de dados - Distribuição de valores das cotações	35

Figura 29 - E-mail enviado pelo sistema Alice	45
Figura 30 - Frequência dos valores estimados - Alice	47
Figura 31 - Status das cotações da empresa Diametral	48
Figura 32 - Painel - Tela inicial.....	56
Figura 33 - Painel - Tela de disputas.....	56
Figura 34 - Painel - Tela de objetos	57
Figura 35 - Painel - Tela de cotações	57
Figura 36 - Painel - Tela de pregões.....	58
Figura 37 - Painel - Tela de alertas.....	58
Figura 38 - Modelo Entidade Relacionamento.....	59

LISTA DE TABELAS

Tabela 1 - Requisições Petronect	15
Tabela 2 - Atributos retornados nas requisições R01, R02 e R03.....	15
Tabela 3 - Atributos da lista contida no atributo 'ANEXOS' da Tabela 2.....	16
Tabela 4 - Atributos da lista contida no atributo 'ITEMS' da Tabela 2.....	16
Tabela 5 - Atributos da lista contida no atributo 'ITEM_NOTES' da Tabela 4	17
Tabela 6 - Atributos da lista contida no atributo 'REGIONS' da Tabela 2.....	17
Tabela 7 - Atributos da requisição de histórico da licitação	17
Tabela 8 - Atributos da lista contida no atributo 'RELEVANT_DATES' da Tabela 7	18
Tabela 9 - Atributo da lista contida no atributo 'QUOTATION_LIST' da Tabela 7	18
Tabela 10 - Atributos da lista contida no atributo 'QUOTS' da Tabela 9	18
Tabela 11 - Atributos da lista contida no atributo 'AUCTION_LIST' da Tabela 7...	18
Tabela 12 - Atributos da lista contida no atributo 'MESSAGES_LIST' da Tabela 7	19
Tabela 13 - Atributos da lista contida no atributo 'ORDERED_CLASS_TOTAL' da Tabela 7	19
Tabela 14 - Atributos da lista contida no atributo 'ORDERED_CLASS_LOTE' da Tabela 7	19
Tabela 15 - Atributos da requisição de buscar mensagens da licitação	19
Tabela 16 - Atributos da lista contida no atributo 'SUBJECTS' da Tabela 15	19
Tabela 17 - Atributos da lista contida no atributo 'MESSAGES' da Tabela 15	20
Tabela 18 - Atributos da lista contida no atributo 'NOTES' da Tabela 15	20
Tabela 19 - Atributos da requisição de informações adicionais.....	20
Tabela 20 - Atributos da lista contida no atributo 'ATTACHMENTS' da Tabela 19	21
Tabela 21 - Atributos da requisição de Modos de Disputa	21
Tabela 22 - Atributos da requisição de Empresas	22
Tabela 23 - Atributos da requisição de Status das licitações concluídas	22
Tabela 24 - Atributos da requisição de Status das licitações em andamento	23
Tabela 25 - Dados selecionados - Licitações	36
Tabela 26 - Dados selecionados - Licitações - Anexos.....	37
Tabela 27 - Dados selecionados - Licitações - Itens	37
Tabela 28 - Dados selecionados - Licitações - Regiões	38
Tabela 29 - Dados selecionados - Histórico.....	38

Tabela 30 - Dados selecionados - 'RELEVANT_DATES'	39
Tabela 31 - Dados selecionados - 'QUOTATION_LIST'	39
Tabela 32 - Dados selecionados - 'QUOTS'	39
Tabela 33 - Dados selecionados - 'AUCTION_LIST'	40
Tabela 34 - Dados selecionados - MESSAGES_LIST.....	40
Tabela 35 - Dados selecionados - ORDERED_CLASS_TOTAL	40
Tabela 36 - Dados selecionados - ORDERED_CLASS_LOTE	41
Tabela 37 - Dados selecionados - 'MESSAGES'	41
Tabela 38 - Dados selecionados - 'ATTACH'	41
Tabela 39 - Dados selecionados - Informações adicionais.....	42
Tabela 40 - Dados selecionados - Modos de disputa	42
Tabela 41 - Dados selecionados - Empresas	42
Tabela 42 - Dados selecionados - Status	43

LISTA DE ABREVIATURAS E SIGLAS

Alice - Sistema de Análise de Licitações e Editais

ANP - Agência Nacional do Petróleo, Gás Natural e Biocombustíveis

ARPIA - Análise Robotizada de Projetos de Investimento e de Aquisições

Coinfra - Coordenação-Geral de Controle Externo de Infraestrutura

Seinfra - Secretaria de Fiscalização de Infraestrutura

SeinfraPetróleo - Secretaria de Fiscalização de Infraestrutura de Petróleo e Gás Natural

Segecex - Secretaria-Geral de Controle Externo

TCU - Tribunal de Contas da União

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Objetivos	18
2	METODOLOGIA	19
3	ENTENDIMENTO DO NEGÓCIO	20
3.1	Determinando os objetivos do negócio	20
3.2	Verificando a situação atual	21
3.3	Objetivos de mineração de dados	23
3.4	Plano do projeto.....	23
4	ENTENDIMENTO DOS DADOS	25
4.1	Coletando dados iniciais.....	25
4.2	Descrevendo os dados	15
4.2.1	Licitações Públicas - Publicadas/Em Andamento/Concluídas (R01 a R03)	15
4.2.2	Histórico (R04)	17
4.2.3	Buscar mensagens (R05)	19
4.2.4	Buscar informações adicionais (R06).....	20
4.2.5	Modos de disputa (R07)	21
4.2.6	Empresas (R08).....	22
4.2.7	Status das licitações concluídas (R09)	22
4.2.8	Status das licitações em andamento (R10).....	22
4.2.9	Baixar anexo (R11).....	23
4.3	Explorando os dados	23
4.4	Verificando a qualidade dos dados.....	35
5	PREPARAÇÃO DOS DADOS.....	36
5.1	Selecionando dados	36
5.1.1	Licitações Públicas - Publicadas/Em Andamento/Concluídas (requisições R01, R02 e R03)	36
5.1.2	Histórico (R04).....	38

5.1.3	Buscar mensagens (R05)	41
5.1.4	Buscar informações adicionais (R06).....	42
5.1.5	Modos de disputa (R07)	42
5.1.6	Empresas (R08)	42
5.1.7	Status das licitações concluídas e em andamento (R09 e R10).....	43
5.2	Limpendo dados	43
5.3	Construindo novos dados	43
5.4	Integrando dados	43
6	MODELAGEM	44
7	AVALIAÇÃO	46
8	CONCLUSÃO	50
8.1	Trabalhos Futuros.....	52
	REFERÊNCIAS	53
	GLOSSÁRIO	55
	ANEXOS	56
	Anexo I - Telas do Painel eletrônico desenvolvido	56
	Anexo II - Modelo Entidade-Relacionamento	59

1 INTRODUÇÃO

A Constituição Federal de 1988 em seu artigo 71 dispõe que o controle externo, a cargo do Congresso Nacional, será exercido com o auxílio do Tribunal de Contas da União (TCU) (Brasil, 1988).

Entre as competências do Tribunal listadas nos incisos de I a XI do mesmo artigo, destaca-se os incisos II e IV:

Art. 71. O controle externo, a cargo do Congresso Nacional, será exercido com o auxílio do Tribunal de Contas da União, ao qual compete:

...

II - julgar as contas dos administradores e demais responsáveis por dinheiros, bens e valores públicos da administração direta e indireta, incluídas as fundações e sociedades instituídas e mantidas pelo Poder Público federal, e as contas daqueles que derem causa a perda, extravio ou outra irregularidade de que resulte prejuízo ao erário público;

...

IV - realizar, por iniciativa própria, da Câmara dos Deputados, do Senado Federal, de Comissão técnica ou de inquérito, inspeções e auditorias de natureza contábil, financeira, orçamentária, operacional e patrimonial, nas unidades administrativas dos Poderes Legislativo, Executivo e Judiciário, e demais entidades referidas no inciso II;

...

Para atender as competências declaradas nos incisos II e IV do artigo 71 da Constituição, o artigo 41 da Lei Orgânica do TCU (Brasil, 1992) dispõe que o TCU irá acompanhar, pela publicação no Diário Oficial da União ou por outros meios, os editais de licitação, contratos e convênios praticados por responsáveis sujeito à sua jurisdição.

Art. 41. Para assegurar a eficácia do controle e para instruir o julgamento das contas, o Tribunal efetuará a fiscalização dos atos de que resulte receita ou despesa, praticados pelos responsáveis sujeitos à sua jurisdição, competindo-lhe, para tanto, em especial:

I - acompanhar, pela publicação no Diário Oficial da União, ou por outro meio estabelecido no Regimento Interno:

a) a lei relativa ao plano plurianual, a lei de diretrizes orçamentárias, a lei orçamentária anual e a abertura de créditos adicionais;

b) os editais de licitação, os contratos, inclusive administrativos, e os convênios, acordos, ajustes ou outros instrumentos congêneres, bem como os atos referidos no art. 38 desta Lei;

II - realizar, por iniciativa própria, na forma estabelecida no Regimento Interno, inspeções e auditorias de mesma natureza que as previstas no inciso I do art. 38 desta Lei;

III - fiscalizar, na forma estabelecida no Regimento Interno, as contas nacionais das empresas supranacionais de cujo capital social a União participe, de forma direta ou indireta, nos termos do tratado constitutivo;

...

Para auxiliar o exercício de suas competências, o Tribunal dispõe da Secretaria-Geral de Controle Externo (Segecex), que tem por finalidade gerenciar as atividades de controle externo, visando prestar apoio e assessoramento às deliberações do Tribunal (art. 2º - Portaria-Segecex nº 5/2017 (União, 2017)). Em sua estrutura, a Coordenação-Geral de Controle Externo de Infraestrutura (Coinfra) coordena as Secretarias de Fiscalização de Infraestrutura (Seinfra). As Seinfras são secretarias temáticas e, conseqüentemente, especializadas em setores específicos da área de infraestrutura nacional.

A Secretaria de Fiscalização de Infraestrutura de Petróleo e Gás Natural (SeinfraPetróleo) tem entre suas competências, conforme listado no art. 31 da Portaria-Segecex nº 5/2017 (União, 2017), a fiscalização dos empreendimentos de infraestrutura nos setores de petróleo e seus derivados, gás natural e biocombustíveis bem como os programas, sistemas e atos de gestão direcionados ao desenvolvimento do segmento.

Para realizar suas fiscalizações de maneira contínua e tempestiva, a SeinfraPetróleo precisa de informações sobre a atuação de seus jurisdicionados no setor que permitam avaliar se as normas legais estão sendo respeitadas. Uma informação relevante para o controle externo são os dados acerca das aquisições e contratações realizadas por meio de licitações.

Atualmente a Secretaria não possui tais dados estruturados e sua obtenção junto aos jurisdicionados muitas vezes se dá por meio de processos longos e demorados, dado que o volume de informações requisitados é muito grande.

Com o objetivo de melhorar a atuação da SeinfraPetróleo, foi criado o Projeto ARPIA (Análise Robotizada de Projetos de Investimento e de Aquisições). O projeto tem como objetivos extrair, analisar e apresentar em forma de painel e de forma automatizada todo o processo licitatório de projetos de investimento e aquisições realizadas pelas empresas que atuam no setor de óleo e gás. Com os dados obtidos será possível conhecer melhor as demandas das empresas atuantes no setor, os instrumentos mais utilizados para as contratações, a efetividade dos processos licitatórios e os fornecedores mais presentes nas disputas.

Um dos maiores jurisdicionados da SeinfraPetróleo é a Petróleo Brasileiro S.A. - Petrobras. Segundo o artigo 61 da lei 9.478 de 6 de agosto de 1997 (Brasil, 1997), a Petrobras é uma sociedade de economia mista vinculada ao Ministério de Minas e Energia, que tem como objeto a pesquisa, a lavra, a refinação, o processamento, o comércio e o transporte de petróleo proveniente de poço, de xisto ou de outras rochas, de seus derivados, de gás natural e de outros hidrocarbonetos fluidos, bem como quaisquer outras atividades correlatas ou afins, conforme definidas em lei.

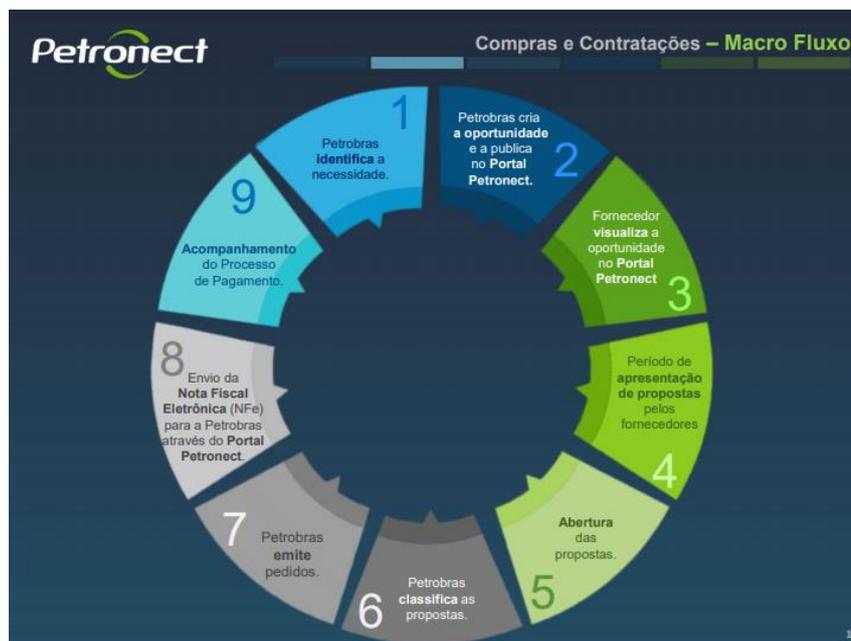
Por ser uma sociedade de economia mista, a Petrobras deve seguir os ditames da lei 13.303 de 30 de junho de 2016, conhecida como a Lei das Estatais (Brasil, 2016). Em seu artigo 28, a lei estabelece que as contratações a serem realizadas pela sociedade de economia mista serão precedidas de licitação.

A lei assegura, em seu artigo 31, que as licitações realizadas devem observar, entre outros princípios, o princípio da publicidade. Não obstante a existência de informações sobre as licitações realizadas pela Petrobras em seu sítio eletrônico (Petrobras), as informações divulgadas carecem de detalhamento.

Em outubro de 2002, por iniciativa da Petrobras, foi criada a Petronect, uma sociedade com participação da Petrobras, através de sua subsidiária e-Petro, SAP e Accenture, para promover serviços de comércio eletrônico relacionados à aquisição de bens e serviços (Petronect, 2002).

A Petronect mantém no ambiente de Internet o Portal de Compras e Contratações com funcionalidades para aquisição de bens e serviços (Petronect, 2019). Através do sistema, as empresas do Grupo Petrobras selecionam empresas, disponibilizam os editais de licitação, esclarecem dúvidas do mercado, e comunicam os vencedores de cada processo (Fornecedor, 2018). O Macrofluxo de compras e contratações da Petronect está ilustrado na Figura 1.

Figura 1 - Macrofluxo de compras e contratações - Petronect



Fonte: (Petronect, 2019)

O portal da Petronect possui dados abundantes e de grande relevância para as atividades de fiscalização da SeinfraPetróleo. Entretanto, as informações disponibilizadas somente são acessíveis por meio de consultas pontuais, muitas vezes protegidas por códigos de validação (*Captchas*), que dificultam o acesso amplo e automatizado. Estas características limitam a atuação abrangente das atividades de controle.

1.1 Objetivos

O presente trabalho tem como objetivo geral subsidiar a SeinfraPetróleo com as informações das licitações públicas realizadas e disponibilizadas no portal Petronect de maneira estruturada em painel eletrônico.

Como objetivos específicos, tem-se o desenvolvimento de software para extrair e internalizar as informações contidas no portal e a construção de painel eletrônico para consolidação das informações adquiridas. Além disso, os arquivos dos editais obtidos serão entregues ao sistema de Análise de Licitações e Editais (Alice) para verificar a possibilidade de utilizar o sistema para apontamentos de possíveis restrições de competitividade presentes nos certames.

Considerando que o projeto ARPIA (criado pela SeinfraPetróleo) pretende, em primeira etapa, internalizar informações das contratações e aquisições realizadas pelas empresas do Grupo Petrobras (Petrobras e suas subsidiárias) conduzidas por meio do sistema eletrônico Petronect, o resultado deste trabalho de conclusão de curso atenderá a necessidade inicial do projeto ARPIA.

2 METODOLOGIA

Para desenvolver o presente trabalho foi utilizado a metodologia CRISP-DM (*Cross Industry Standard Process for Data Mining*) (IBM, 2020).

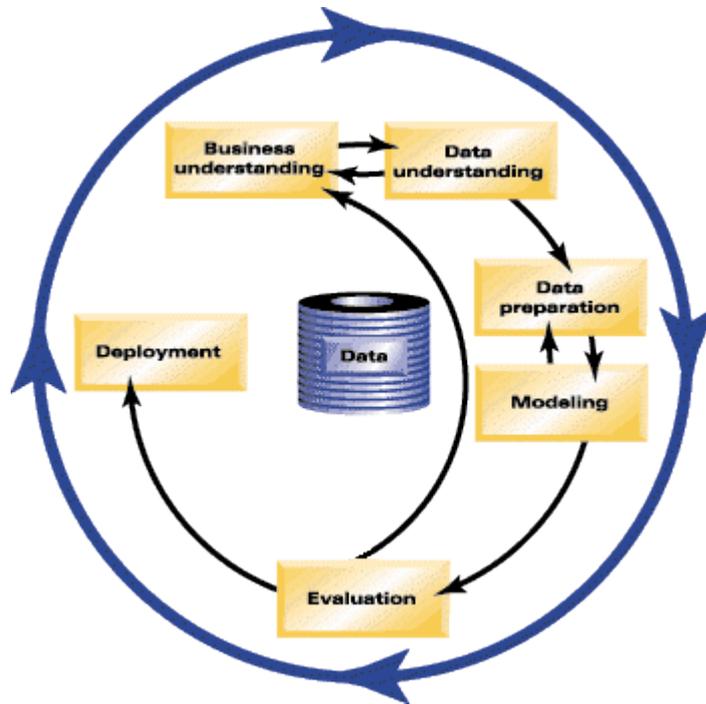
O CRISP-DM é uma metodologia independente de indústria e tecnologia, desenvolvida como um modelo de processos genéricos para facilitar e estruturar o desenvolvimento de projetos de mineração de dados (Wirth, et al.). A metodologia engloba etapas de planejamento, desenvolvimento e avaliação do projeto no qual foi aplicada.

O CRISP-DM consiste em seis fases e cada fase possui tarefas genéricas (SPSS, 2000):

- **Entendimento do Negócio:** Consiste em identificar os objetivos de negócio a serem atendidos pelo projeto, analisando-se o contexto do projeto e identificando-se os objetivos da mineração de dados;
- **Entendimento dos Dados:** Consiste na coleta inicial dos dados e sua exploração para familiarização e avaliação de sua qualidade;
- **Preparação dos Dados:** Consistem em atividades para a transformação dos dados para adequá-los aos algoritmos de mineração;
- **Modelagem:** Nesta fase, são selecionadas as técnicas e modelos que serão utilizadas na mineração de dados;
- **Avaliação dos Resultados:** Fase utilizada para avaliar os resultados das fases anteriores para determinar se o modelo aplicado alcança os objetivos propostos;
- **Implantação:** Aplicação dos artefatos gerados nos processos de trabalho que motivaram o projeto.

A Figura 2 apresenta as fases estabelecidas pela metodologia CRISP-DM e suas principais dependências. A figura também deixa claro a natureza cíclica da metodologia, incentivando a execução de suas fases de forma a permitir que fases já executadas possam ser revisadas ao longo do projeto.

Figura 2 - Fases do CRISP-DM



Fonte: (IBM, 2020)

O restante deste trabalho foi estruturado seguindo as fases descritas pelo CRISP-DM com algumas adaptações.

3 ENTENDIMENTO DO NEGÓCIO

Nesta seção serão delineados os objetivos do projeto buscando entender o problema a ser resolvido, a situação atual e as ferramentas disponíveis, os objetivos e os critérios de sucesso da mineração de dados e o plano do projeto.

3.1 Determinando os objetivos do negócio

A SeinfraPetróleo possui a necessidade de obter os dados das licitações públicas realizadas pela Petrobras para melhorar sua atuação de fiscalização e acompanhamento do setor de petróleo e gás.

O Grupo Petrobras disponibiliza no portal Petronect grande conjunto de dados sobre suas licitações. Contudo, a obtenção manual das informações é penosa dado o grande volume de dados e a necessidade de preenchimentos de diversos *Captchas* ao longo da navegação pelas páginas do portal.

Assim, estabeleceu-se como objetivo do negócio melhorar a atuação da SeinfraPetróleo por meio da disponibilização de forma estruturada e visual das informações a serem obtidas do portal Petronect possibilitando a análise tanto de forma individual como agregada dos dados.

Para atingir este objetivo foi desenvolvido software para extrair e internalizar as informações contidas no portal e foi construído painel eletrônico para consolidação das informações adquiridas. Além disso, os arquivos dos editais obtidos foram entregues ao sistema Alice para avaliar se o sistema consegue identificar possíveis restrições de competitividade presentes nos certames. Os resultados da análise do Alice foram agregados ao painel eletrônico desenvolvido.

Espera-se que ao final do trabalho seja possível acompanhar tempestivamente todas as contratações realizadas por meio do portal da Petronect; diminuir o tempo, esforço e dispêndio de recursos gasto na preparação e obtenção dos dados de contratações; tomar decisões de maneira mais confiante e inteligente sobre qual objeto de controle selecionar (com base em indicadores de risco).

3.2 Verificando a situação atual

O portal eletrônico Petronect disponibiliza publicamente as informações sobre as licitações públicas das empresas do Grupo Petrobras que se encontram em um de três estágios: publicadas, em andamento e concluídas. Adicionalmente, o portal disponibiliza informações mais detalhadas em área de acesso restrito, mediante autenticação de usuário e senha. Contudo, é possível extrair grande parte das informações sobre as licitações diretamente da área pública. Os dados que alimentam o portal podem ser recuperados utilizando interfaces de requisições https e podem ser armazenados em um banco de dados relacional.

O Alice, Sistema de Análise de Licitações e Editais, é uma ferramenta utilizada para varrer grandes quantidades de documentos que compõe os editais e atas de realização de pregões eletrônicos. Este sistema busca por padrões textuais pré-configurados que possam sinalizar indícios de irregularidades ou riscos no processo licitatório.

Neste projeto o sistema Alice será utilizado para a execução das buscas textuais já existentes nos novos documentos adquiridos do portal Petronect.

Como exigências para o atingimento do objetivo proposto têm-se que:

- O trabalho deveria ser concluído até o início de março de 2020.

- Os arquivos recuperados deveriam ser armazenados em sistema de arquivos que façam parte da infraestrutura de avaliação do sistema Alice. O formato de gravação dos dados deveria obedecer ao padrão pré-estabelecido utilizado pelo sistema Alice.
- Os metadados obtidos do portal deveriam ser gravados em servidor de banco de dados que permita seu acesso por ferramentas de visualização de dados.

As seguintes premissas foram adotadas durante o desenvolvimento do projeto:

- As informações obtidas do portal Petronect são públicas.
- Informações acerca do valor, itens e empresas contratadas estarão presentes nos dados extraídos.
- Os dados obtidos serão disponibilizados para os usuários por meio de ferramentas de visualização de dados (Dashboards).
- O sistema Alice será capaz de aplicar as tipologias preexistentes nos arquivos novos disponibilizados por esse projeto.

Contudo, deve-se observar que o trabalho desenvolvido possui as seguintes limitações:

- A extração dos dados ocorre em datas e horários predefinidos, o que pode gerar diferenças entre os dados presentes no portal e os dados disponibilizados na última extração.
- As informações disponibilizadas na área de acesso restrito do portal não são recuperadas, devido ao maior grau de esforço de obtenção necessário e ao curto prazo para conclusão do presente projeto.

Entre os riscos do projeto, encontrou-se a possibilidade de haver alterações no portal Petronect que impactassem ou inviabilizassem a obtenção dos dados. Além disso, os dados que foram obtidos poderiam não possuir informações esperadas pela área de negócio ou conter possíveis inconsistências existentes na origem dos dados. A complexidade da forma como os dados são disponibilizados pelo portal também poderia trazer dificuldades na realização do projeto, principalmente considerando-se o prazo para sua conclusão.

Para mitigar os riscos de mudanças no portal Petronect afetarem ou inviabilizarem a obtenção dos dados, a construção de um serviço web ou um mecanismo de envio dos dados das disputas diretamente pelo Petronect ao Tribunal seria melhor do que extrair as informações por meio de raspagem de dados. Esse mecanismo, além de combater possíveis alterações nas estruturas de disponibilização de dados do portal, poderia ser utilizado para aumentar a assertividade do significado dos dados obtidos.

Os possíveis benefícios do projeto incluem o ganho de conhecimento acerca do volume de compras, tipos de itens, empresas contratadas e valores envolvidos nas compras realizadas pelas empresas do grupo Petrobras.

Além disso, poderão surgir informações que levem a novas formas de atuação nas fiscalizações por parte da SeinfraPetróleo e ações de controle mais eficientes.

Com a utilização em conjunto do sistema Alice para a análise dos editais, novas tipologias de indícios de irregularidades poderão ser agregadas ao sistema.

Possivelmente o método empregado poderia inclusive ser replicado para extração de informações relacionadas a outras empresas públicas.

3.3 Objetivos de mineração de dados

Este projeto teve como objetivo de mineração de dados o desenvolvimento de um sistema que obtenha os documentos disponibilizados pelo portal Petronect de forma automática. Conforme definido na metodologia CRISP-DM, o tipo de problema de mineração abordado é o de descrição e sumarização de dados.

Para atingir o objetivo de mineração de dados, teve-se como objetivos secundários: a modelagem dos dados obtidos, de forma que fosse possível construir uma base de dados; a disponibilização de painel com as informações (metadados) obtidas durante o processo de coleta de documentos, para facilitar a compreensão pela área interessada e do público em geral; integração dos resultados com o sistema de Análise de Licitações e Editais (Alice), de maneira a permitir a execução dos algoritmos de mineração de dados existentes nesta outra ferramenta.

3.4 Plano do projeto

O plano do projeto foi dividido em 4 atividades: 1) Estudo inicial; 2) Desenvolvimento do módulo de obtenção de dados; 3) Integração com o sistema Alice; 4) Desenvolvimento do Painel Eletrônico.

Estudo Inicial: O portal Petronect foi estudado com o objetivo de entender quais são os dados disponíveis e como os dados são apresentados. Neste ponto, foi necessário solicitar aos gestores do Petronect informações de usuário e senha para acesso a área restrita do portal. Em conjunto com integrantes da SeinfraPetróleo, foi estabelecido quais dados, além dos documentos das licitações, são relevantes para o negócio.

Após essa etapa, foi necessário realizar estudo mais aprofundado das requisições realizadas durante a interação com o portal para verificar quais seriam as técnicas mais apropriadas a serem empregadas para fazer a raspagem dos dados. As verificações do código fonte das páginas assim como a análise do tráfego de dados na rede durante as interações com o portal foram realizadas durante o estudo. Foram analisados tanto as páginas da área pública quanto da área restrita do portal. Para definição das técnicas foram levados em consideração o volume de informações que seriam extraídas, a complexidade da realização da raspagem de dados e o prazo final para a entrega do projeto.

Desenvolvimento do módulo de obtenção de dados: Concluído o estudo inicial, foi iniciado o desenvolvimento do módulo que obtém os dados e documentos do portal. Para o desenvolvimento do módulo de extração dos dados foi utilizado a linguagem Python no editor de código-fonte Visual Studio Code desenvolvido pela Microsoft. Foram utilizadas, entre outras, as bibliotecas ‘requests’, ‘json’ e ‘sqlalchemy’ desenvolvidas em Python para realizar as interações com o portal e gravar as informações em banco de dados.

Com base nos dados que foram extraídos, foi realizada etapa de modelagem dos dados para internalizar as informações em banco de dados relacional. Para fornecer a infraestrutura de banco de dados, foi utilizado o LabContas - Laboratório de Informações de Controle - plataforma do TCU composta de bases de dados internalizadas, catálogo de informações, ferramentas e soluções para tratamento e análise de dados destinada ao exercício das atividades de controle externo.

O LabContas possui sistema de gerenciamento de banco de dados, que utiliza a linguagem SQL (Linguagem de Consulta Estruturada, do inglês Structured Query Language) como interface.

Com a obtenção dos primeiros conjuntos de dados, foi realizada nova etapa de estudo para melhorar a compreensão das informações. O conjunto inicial de dados foi analisado e disponibilizado para a equipe da Secretaria de Fiscalização para nova rodada de verificação se os objetivos do projeto estavam sendo alcançados. Como a resposta foi afirmativa, o projeto avançou para a finalização do desenvolvimento do módulo.

Integração com o sistema Alice: Para a conclusão do desenvolvimento do módulo também foi necessário envolver a equipe responsável pela manutenção do sistema Alice no Tribunal. Com esta atividade, alinhou-se os produtos gerados neste trabalho com os insumos esperados pelo sistema Alice para que fosse possível realizar a integração de ambos.

Desenvolvimento do Painel Eletrônico: Concluído o desenvolvimento do módulo de obtenção de dados, foi iniciado o desenvolvimento do painel eletrônico que permitiu aos interessados consultar as informações obtidas. O desenvolvimento do painel foi realizado coletando feedback com a área de negócio periodicamente para otimizar a forma de visualização das informações. Para construção do painel de informações, optou-se pelo QlikView desenvolvido pela empresa QlikTech.

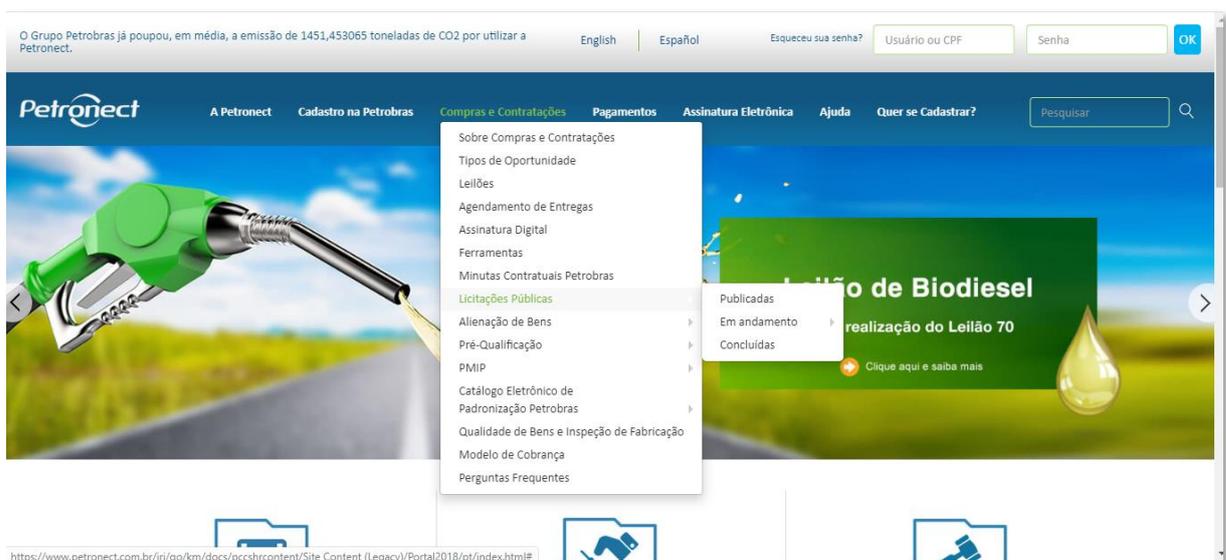
4 ENTENDIMENTO DOS DADOS

Antes de se iniciar o desenvolvimento do software de extração e internalização dos dados, se fez necessário coletar alguns dados iniciais e explorá-los para identificar quais informações seriam relevantes para o atingimento dos objetivos propostos.

4.1 Coletando dados iniciais

O portal Petronect organiza os dados das licitações públicas em 5 áreas distintas separando-as pelo estágio em que se encontra a licitação. As áreas são acessadas por meio da opção ‘Compras e Contratações’ da barra de menu horizontal do portal. Nessa opção está disponibilizada a sub-opção ‘Licitações Públicas’. Este sub-menu, por sua vez, dá acesso às páginas que contém as informações sobre as licitações que estão publicadas, em andamento e concluídas, conforme exibido na Figura 3.

Figura 3 - Portal Petronect - Licitações públicas



Fonte: (Petronect, 2020)

Acessando-se a opção ‘Publicadas’, o portal exibe listagem das disputas que já foram publicadas no Petronect.

Figura 4 - Disputas publicadas

Número	Objeto	Sub-status	Empresa	Data início	Data fim	Opções
7002156592	Contrato Global de Selos Mecânicos	Recebimento de Propostas	Petróleo Brasileiro S. A.	27/08/2018	28/02/2020	<input type="button" value="Tenho Interesse"/>
7002771580	Serv-recuper-valv-conj-engasat-anal-comp	Recebimento de Propostas	Petróleo Brasileiro S. A.	30/10/2019	07/02/2020	<input type="button" value="Tenho Interesse"/>
7002789255	Serviços de Adequação em Lotos de Cabos	Recebimento de Propostas	Petróleo Brasileiro S. A.	24/01/2020	14/02/2020	<input type="button" value="Tenho Interesse"/>
7002875408	Água potável	Recebimento de Propostas	Petróleo Brasileiro S. A.	21/11/2019	08/01/2021	<input type="button" value="Tenho Interesse"/>
7002875410	Suporte à manutenção de equipamentos	Recebimento de Propostas	Petróleo Brasileiro S. A.	12/12/2019	06/02/2020	<input type="button" value="Tenho Interesse"/>
7002911712	Compressor para fuso	Recebimento de Propostas	Petróleo Brasileiro S. A.	09/01/2020	14/02/2020	<input type="button" value="Tenho Interesse"/>
7002914461	FORN ADT ANTIOXIDANTE UBC UBMC	Recebimento de Propostas	PETROBRAS BIOCOMBUSTIVEL	18/12/2019	21/02/2020	<input type="button" value="Tenho Interesse"/>
7002959272	Global Metil Isobutil Cetona MIBK granel	Recebimento de Propostas	Petróleo Brasileiro S. A.	24/12/2019	12/02/2020	<input type="button" value="Tenho Interesse"/>
7002945472	BASE E MÓDULO DO PLC ACS00	Recebimento de Propostas	BREITENER JARAQUI S.A.	23/01/2020	06/02/2020	<input type="button" value="Tenho Interesse"/>
7002945573	BASE E MÓDULO PLC ACS00	Recebimento de Propostas	BREITENER TAMBAQUI S.A.	23/01/2020	06/02/2020	<input type="button" value="Tenho Interesse"/>

Fonte: (Petronect, 2020)

A listagem de disputas apresenta opções que, ao serem acessadas, exibem detalhes da licitação (Figura 5), arquivos anexos (Figura 6) e listagem dos itens da licitação (Figura 7).

Figura 5 - Detalhamento da licitação

O Grupo Petrobras já poupou, em média, a emissão de 1451,452966 toneladas de CO2 por utilizar a Petronect. English | Español Esqueceu sua senha?

Petronect A Petronect Cadastro na Petrobras Compras e Contratações Pagamentos Assinatura Eletrônica Ajuda Quer se Cadastrar?

LICITAÇÕES PÚBLICAS > PUBLICADAS

Nome:

Descrição do Objeto de Contratação:

Status:

Data de início:

Data Fim:

Fuso Horário:

Moeda:

Condição de Pagamento:

Enquadramento de Processo:

Fonte: (Petronect, 2020)

Figura 6 - Itens da licitação

Oportunidade 7002156592 - Informações

Licitação publicada no Diário Oficial da União em 27/08/2018

Petróleo Brasileiro S. A. Número da Oportunidade: 7002156592 Objeto de Contratação: Contrato Global de Selos Mecânicos Data início: 27/08/2018 às 10:00:00 Horas Data fim: 28/02/2020 às 18:00:00 Horas

Itens			
8001	LOTE A	Quantidade: 1	Unidade de Fornecimento:
0001	LOTE A - FOLE SIMPLES E DUPLO	Quantidade: 1	Unidade de Fornecimento: UN
8002	LOTE B	Quantidade: 1	Unidade de Fornecimento:
0002	LOTE B - MOLA MÚLTIPLA ALTA PRESSÃO	Quantidade: 1	Unidade de Fornecimento: UN

Recebimento de Propostas BREITENER TAMBAQUI S.A. 23/01/2020 06/02/2020

Fonte: (Petronect, 2020)

Figura 7 - Anexos da licitação

Oportunidade 7002156592 - Anexos

Descrição	Ações
Adendo_A Minuta Inst Contratual	Download
Anexo 10 REG-DP-061-0509_FL0001	Download
Anexo 9 REG-DP-061-0508_FL0002	Download
Anexo 8 REG-DP-061-0508_FL0001	Download
Anexo 7 REG-DP-061-0507_FL0002	Download

Recebimento de Propostas BREITENER JARAQUI S.A. 23/01/2020 06/02/2020

Recebimento de Propostas BREITENER TAMBAQUI S.A. 23/01/2020 06/02/2020

Fonte: (Petronect, 2020)

Para acessar as informações da listagem dos anexos e as informações dos itens das licitações, é necessário o preenchimento de um *captcha* conforme a Figura 8.

Figura 8 - Captcha - Anexos



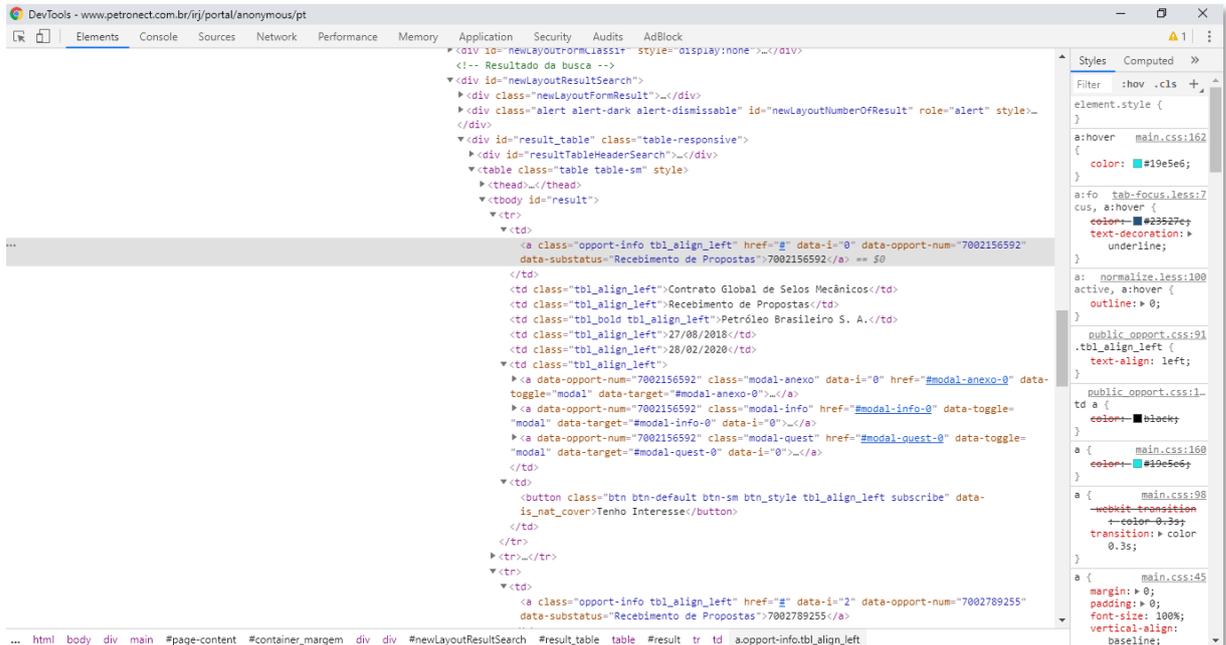
The image shows a web browser window titled "Oportunidade 7002156592 - Anexos". In the center, there is a captcha box containing the alphanumeric code "fpk1R 390". Below the code is an input field with the placeholder text "Insira aqui o código acima" and a refresh button. To the left of the input field is a button labeled "Enviar". At the bottom of the browser window, the address bar shows "Recebimento de Propostas", "Petróleo Brasileiro S. A.", "27/08/2018", and "28/02/2020".

Fonte: (Petronect, 2020)

Captchas são testes utilizados em sítios eletrônicos para identificar se a interação está ocorrendo com humanos ou com outras máquinas. A ferramenta é comumente utilizada para evitar acessos rápidos automatizados que possam degradar a qualidade do serviço ou que possam sobrecarregar o fornecedor do serviço ao ponto de inviabilizar seu acesso.

Conhecida a árvore de páginas do portal, buscou-se, inicialmente, identificar no código-fonte de cada página as estruturas que continham as informações desejadas. Para visualizar o código-fonte das páginas, foi utilizada as 'Ferramentas de desenvolvedor' (Devtools) do navegador web Google Chrome.

Figura 9 - Código-fonte da página - DevTools

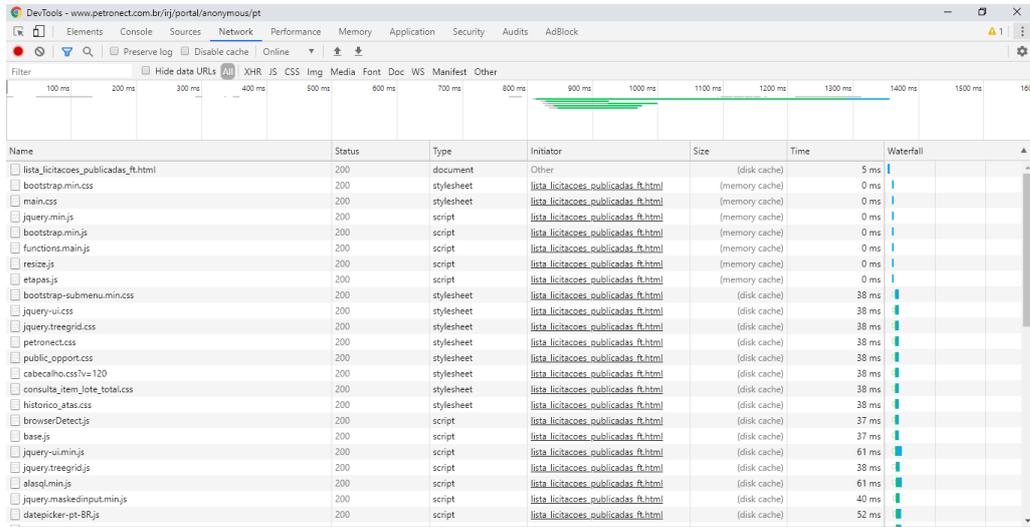


Fonte: própria

Durante o estudo dos códigos-fontes de cada página necessária, verificou-se que algumas das páginas eram exibidas utilizando-se iFrames. Um iFrame é um código em HTML que permite que uma página seja exibida dentro de outra página, entretanto seu código-fonte não fica disponível na própria página que executa o iFrame. Assim, para acessar seu código-fonte seria necessário fazer uma nova requisição para a origem do iFrame.

Aprofundando-se mais o estudo da origem dos dados almejados, resolveu-se analisar as requisições que eram enviadas para o portal. Novamente, utilizando-se a 'Ferramentas de desenvolvedor' do navegador de internet, verificou-se as diversas requisições realizadas entre o navegador e o portal, conforme mostrado na Figura 10.

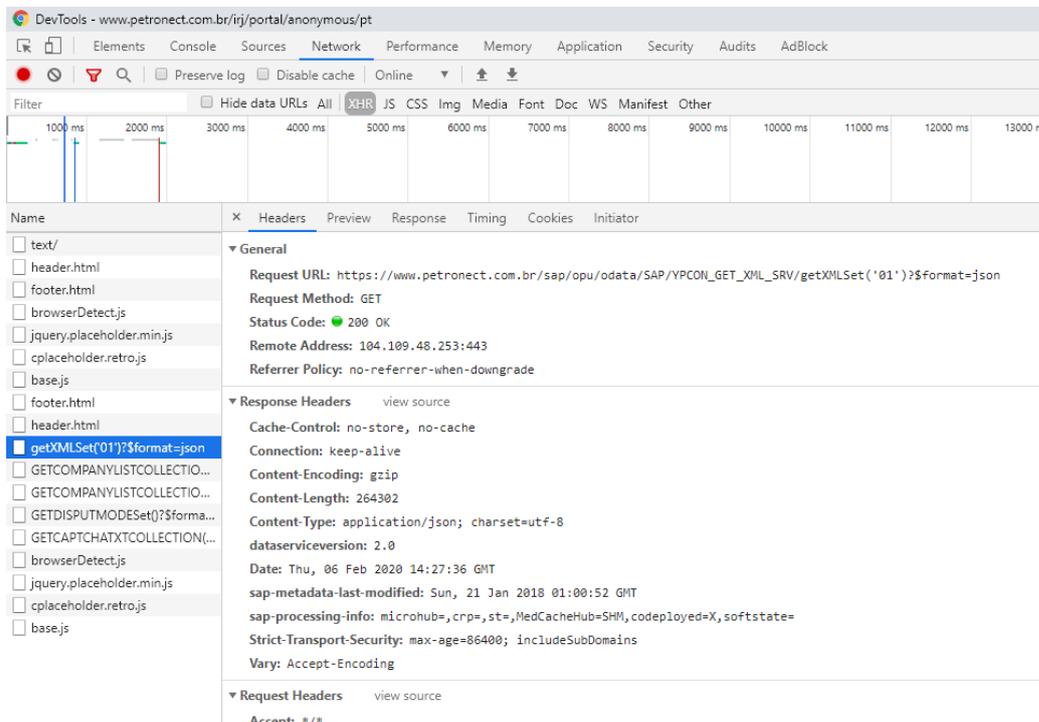
Figura 10 - Requisições - DevTools



Fonte: própria

Muitas das requisições correspondiam à transferência de scripts javascript e de folhas de estilo CSS. Todavia, algumas requisições retornavam conteúdos no formato JSON, exemplificadas na Figura 11.

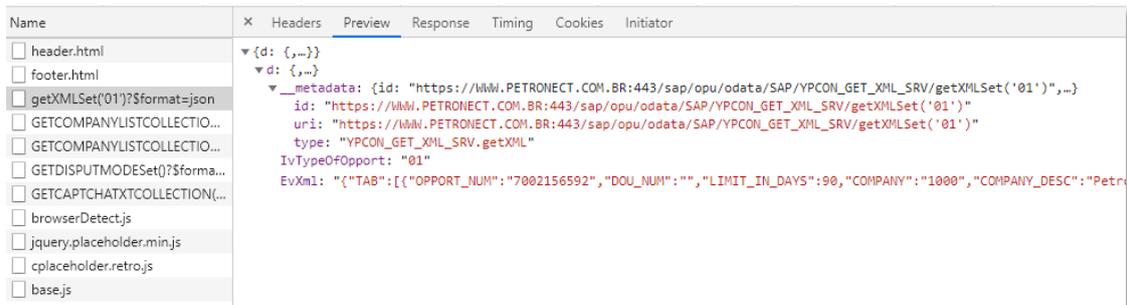
Figura 11 - Requisição - Licitações publicadas - DevTools



Fonte: própria

Uma das requisições identificadas retornou os dados utilizados pelo portal para a construção da lista de licitações publicadas, conforme Figura 12.

Figura 12 - Detalhes da requisição de licitações publicadas



Fonte: própria

Após estudo das requisições realizadas durante a interação com o Portal Petronect nessas 5 áreas distintas, verificou-se que os dados disponibilizados na área pública do portal poderiam ser adquiridos por meio de requisições específicas que retornam os dados em formato JSON.

Foram identificadas um total de 11 requisições que retornam informações acerca das licitações públicas realizadas por meio do Petronect. As requisições são realizadas utilizando-se comando HTTP GET passando os valores da URL específica, conforme descrito na Tabela 1. Cada requisição retorna um conjunto de dados diferentes e complementares ao trabalho. As requisições foram numeradas no formato “Rxx”, onde xx indica o número da requisição.

Tabela 1 - Requisições Petronect

Tipo de Requisição	URL	Parâmetros
R01: Licitações Públicas Publicadas	https://www.petronect.com.br/sap/opu/odata/SAP/YPCON_GET_XML_SRV/getXMLSet('01')?\$format=json	O valor passado para se obter a lista de licitações públicas publicadas é o código 01.
R02: Licitações Públicas Em Andamento	https://www.petronect.com.br/sap/opu/odata/SAP/YPCON_GET_PUB_LIC_IN_PROGRESS_SRV/GETLISTLICITINPROGRESSSet(DisputeMode='%s',ObjectId='',Description='',CompanyCode='%s',SubStatus='07',Regions='',PublishDate='%s',PublishDateFinal='%s',EndDate='',EndDateFinal='',Lang='P')?\$format=json	<i>DisputeMode</i> - Código do Modo de Disputa; <i>CompanyCode</i> - Código da Empresa contratante, e <i>PublishDate</i> e <i>PublishDateFinal</i> - Intervalo de datas de publicação que se deseja buscar as licitações
R03: Licitações Públicas Concluídas	https://www.petronect.com.br/sap/opu/odata/SAP/YPCON_GET_PUBL_LICIT_CONC_SRV/GETLISTLICITCONCSet(DisputeMode='04',ObjectId='',Description='',CompanyCode='%s',SubStatus='01',Regions='',PublishDate='',PublishDateFinal='',EndDate='%s',EndDateFinal='%s',Lang='P')?\$format=json	<i>CompanyCode</i> - Código da Empresa contratante; <i>EndDate</i> e <i>EndDateFinal</i> - Intervalo de datas de conclusão de licitações que se deseja buscar.
R04: Histórico	https://www.petronect.com.br/sap/opu/odata/SAP/YPCON_PUB_HISTORIC_SRV/historyCollection(DisputeMode='%s',ObjectId='%s',ItemType='02')?\$format=json	<i>DisputeMode</i> - Código do Modo de Disputa; <i>ObjectId</i> - Código identificador da licitação
R05: Buscar mensagens	https://www.petronect.com.br/sap/opu/odata/SAP/YPCON_GET_MESSAGES_INFO_SRV/messagesInfoSet('%s')?\$format=json	Código identificador da licitação
R06: Buscar informações adicionais	https://www.petronect.com.br/sap/opu/odata/SAP/YPCON_GET_HEADER_INFO_SRV/headerInfoSet('%s')?\$format=json	Código identificador da licitação
R07: Modos de disputa	https://www.petronect.com.br/sap/opu/odata/SAP/YPCON_GET_DISPUT_MODE_SRV/GETDISPUTMODESet()?\$format=json&\$filter=IvLangu eq 'P'	Não há.
R08: Empresas	https://www.petronect.com.br/sap/opu/odata/SAP/YPCON_GET_COMPANY_LIST_SRV/GETCOMPANYLISTCOLLECTION?\$format=json	Não há.
R09: Status das licitações concluídas	https://www.petronect.com.br/sap/opu/odata/SAP/YPCON_GET_STATUS_OPP_CONC_SRV/GETSTATPUBOPPCONCSet()?\$format=json&\$filter=IvLangu eq 'P'	Não há.
R10: Status das licitações em andamento	https://www.petronect.com.br/sap/opu/odata/SAP/YPCON_GET_STAT_OPP_IN_PROGRESS_SRV/GETSTATPUBOPPInProgressSet()?\$format=json&\$filter=IvDisputeMode eq '%s' and IvLangu eq 'P'	<i>DisputeMode</i> - Código do Modo de Disputa;
R11: Baixar anexo	https://www.petronect.com.br/sap/opu/odata/SAP/YPCON_PUB_ATTACHMENT_DOWNLOAD_SRV/attachmentSet('%s')/\$value	Código identificador do anexo

4.2 Descrevendo os dados

Nesta seção serão apresentados o detalhamento dos objetos retornados em formato JSON pelas requisições R01 a R11.

4.2.1 Licitações Públicas - Publicadas/Em Andamento/Concluídas (R01 a R03)

Os dados obtidos nas requisições Licitações Públicas - Publicadas (R01), Em Andamento (R02) e Concluídas (R03) possuem estruturas iguais em formato JSON. Foram identificados 31 atributos simples e 3 atributos compostos por listas de objetos que possuem outros atributos internamente. A estrutura dos dados é apresentada na Tabela 2.

Tabela 2 - Atributos retornados nas requisições R01, R02 e R03

Atributo	Informação	Formato
AUC_START_DATE	Data de início do pregão	Data (yyyy-mm-dd)
AUC_START_TIME	Horário de início do pregão	Hora (hh:mi:ss)
COMPANY	Código da Empresa que está realizando a licitação	Numérico
COMPANY_DESC	Nome da Empresa que está realizando a licitação	Alfanumérico
DESC_DETAIL	Informações adicionais	Alfanumérico
DISPUTE_MODE	Código do modo de disputa da licitação	Numérico
DOU_NUM	Número do diário oficial	Numérico
DOU_PUBL_DATE	Data de publicação no diário oficial	Data (yyyy-mm-dd)
START_DATE	Data de início da licitação	Data (yyyy-mm-dd)
START_HOUR	Horário de início da licitação	Hora (hh:mi:ss)
END_DATE	Data de fim da licitação	Data (yyyy-mm-dd)
END_HOUR	Horário de fim da licitação	Hora (hh:mi:ss)
HAS_PREQUALIFIED	Se houve pré-qualificação	Alfanumérico
IS_EQUALIZED	(não identificado)	-
IS_PREQUALI	(não identificado)	-
LIMIT_IN_DAYS	Limite em dias	Numérico
NAT_COVERAGE	Código de cobertura da Licitação	Alfanumérico
OPEN_DATE	(não identificado)	Data (yyyy-mm-dd)
OPEN_HOUR	(não identificado)	Hora (hh:mi:ss)
OPPORT_DESCR	Nome atribuído a licitação	Alfanumérico
OPPORT_NUM	Número identificador da licitação	Numérico
OPPORT_TYPE	Código do tipo da licitação	Alfanumérico
POSTING_DATE	Data de publicação da licitação no portal	Data (yyyy-mm-dd)
PQ_VENDOR_LIST_DATE	(não identificado)	Data (yyyy-mm-dd)

PQ_VENDOR_LIST_HOUR	(não identificado)	Hora (hh:mi:ss)
STATUS	Código do estado da licitação	Numérico
STATUS_DESC	Descrição do estado da licitação	Alfanumérico
SUB_STATUS	Código do sub-estado da licitação	Numérico
ANEXOS	Lista dos arquivos associados a licitação	Lista
ITEMS	Lista dos itens que serão adquiridos pela licitação	Lista
REGIONS	Lista de regiões atendidas pela licitação	Lista

Cada anexo na lista do atributo 'ANEXOS' apresenta dois atributos de informação. Os anexos são os documentos publicados na licitação. Esses atributos serão relevantes para o atendimento do objetivo do projeto em obter os documentos referentes à licitação. A estrutura do anexo é apresentada na Tabela 3.

Tabela 3 - Atributos da lista contida no atributo 'ANEXOS' da Tabela 2

Atributo	Informação	Formato
DESCRIPTION	Nome do anexo	Alfanumérico
PHIO_OBJID	Código identificador do anexo	Alfanumérico

A lista de itens que serão adquiridos pela licitação é a informação contida no atributo 'ITEMS' da tabela. Cada item da licitação apresenta um conjunto de 12 atributos simples e 1 atributo composto por uma lista. A estrutura dos dados dos itens da licitação é apresentada na Tabela 4.

Tabela 4 - Atributos da lista contida no atributo 'ITEMS' da Tabela 2

Atributo	Informação	Formato
ITEM_NUM	Número identificador do item	Numérico
EXLIN	Código identificador do item	Alfanumérico
ITEM_DESC	Descrição do item	Alfanumérico
ITEM_NOTES	Lista com notas explicativas relativas ao item	Lista
UNIT	Unidade	Alfanumérico
QUANTITY	Quantidade	Numérico
DELIV_DATE	Data de entrega	Data (yyyy-mm-dd)
GROUPING_LEVEL	Nível de agrupamento	Alfanumérico
ITEM_PROCESS_TYP	Código do tipo do item	Alfanumérico
FAMILY	Código identificador da família do item	Numérico
FAMILY_DESCR	Descrição da família do material	Alfanumérico
NUM_MATERIAL	Número do material	Numérico
FAMILY_CLASSIF	Código da classificação da família do item	Numérico

A lista de notas explicativas do item possui 2 atributos apresentados na Tabela 5.

Tabela 5 - Atributos da lista contida no atributo 'ITEM_NOTES' da Tabela 4

Atributo	Informação	Formato
TDFORMAT	(não identificado)	Alfanumérico
TDLINE	Texto da nota	Alfanumérico

Os dados contidos na lista 'REGIONS', que aparenta possuir informações das localidades atendidas pela licitação, possuem 3 atributos simples e são apresentados na Tabela 6.

Tabela 6 - Atributos da lista contida no atributo 'REGIONS' da Tabela 2

Atributo	Informação	Formato
COUNTRY	Código do país	Alfanumérico
REGION	Código da Unidade Federativa	Alfanumérico
REGION_DESCRIPTION	Nome da Unidade Federativa	Alfanumérico

4.2.2 Histórico (R04)

A requisição de histórico da licitação (R04) traz 6 atributos do tipo lista e 1 atributo simples. As informações remontam as ações que foram realizadas durante o tempo de vida da licitação. A estrutura do objeto retornado no formato JSON está descrita nas Tabelas 7 a 14.

Tabela 7 - Atributos da requisição de histórico da licitação

Atributo	Informação	Formato
RELEVANT_DATES	Lista contendo informações de datas relacionadas à licitação	Lista
QUOTATION_LIST	Lista contendo informações acerca de cotações recebidas na licitação	Lista
AUCTION_LIST	Lista contendo informações dos lances nas licitações do tipo pregão	Lista
MESSAGES_LIST	Lista de mensagens trocadas durante o pregão licitação	Lista
ORDERED_CLASS_TOTAL	Lista com informações da classificação das empresas participantes por valor total	Lista
ORDERED_CLASS_LOTE	Lista com informações da classificação de empresas participantes por lote	Lista
HAS_DIVULG_REL	Indicativo de divulgação do histórico	Alfanumérico

Tabela 8 - Atributos da lista contida no atributo 'RELEVANT_DATES' da Tabela 7

Atributo	Informação	Formato
START_DATE	Data de início da licitação	Data (yyyy-mm-dd)
START_TIME	Horário de início da licitação	Hora (hh:mi:ss)
RECURSAL_PERIOD	Data do período de recursos	Data (yyyy-mm-dd)
END_TIME	Data de fim da licitação	Data (yyyy-mm-dd)
END_DATE	Horário de fim da licitação	Hora (hh:mi:ss)
COMMER- CIAL_PROP_OPEN_TIME	Horário de abertura para recebimento de cotações	Hora (hh:mi:ss)
COMMER- CIAL_PROP_OPEN_DATE	Data de abertura para recebimento de cotações	Data (yyyy-mm-dd)
PUBL_DATE	Data de publicação no diário oficial	Data (yyyy-mm-dd)
DIVULG_DATE	Data de publicação da licitação	Data (yyyy-mm-dd)

Tabela 9 - Atributo da lista contida no atributo 'QUOTATION_LIST' da Tabela 7

Atributo	Informação	Formato
NUMBER_INT	Número identificador do item da licitação de que se trata a cotação	Alfanumérico
DESCRIPTION	Descrição do item da licitação de que se trata a cotação	Alfanumérico
QUOTS	Lista com informações das cotações	Lista

Tabela 10 - Atributos da lista contida no atributo 'QUOTS' da Tabela 9

Atributo	Informação	Formato
TAXNUM	CNPJ da empresa	Numérico
SUPPLIER_NAME	Nome da empresa	Alfanumérico
MEPP	Indicativo de microempresa	Alfanumérico
QUANTITY	Quantidade do item da licitação cotado	Numérico
UNIT_VALUE	Valor unitário do item cotado	Numérico
TOTAL_VALUE	Valor total da cotação	Numérico
CREATED_AT_DATE	Data da cotação	Data (yyyy-mm-dd)
CREATED_AT_TIME	Horário da cotação	Hora (hh:mi:ss)
CHARACTERISTIC	Status da cotação	Alfanumérico

Tabela 11 - Atributos da lista contida no atributo 'AUCTION_LIST' da Tabela 7

Atributo	Informação	Formato
TAXNUM	CNPJ da empresa	Numérico
AUC_VALUE	Valor do lance	Numérico
CREATED_AT_DATE	Data da cotação	Data (yyyy-mm-dd)
CREATED_AT_TIME	Horário da cotação	Hora (hh:mi:ss)

Tabela 12 - Atributos da lista contida no atributo 'MESSAGES_LIST' da Tabela 7

Atributo	Informação	Formato
AUTOR	Autor da mensagem	Alfanumérico
CREATED_AT_DATE	Data da cotação	Data (yyyy-mm-dd)
CREATED_AT_TIME	Horário da cotação	Hora (hh:mi:ss)
MESSAGE	Conteúdo da mensagem	Alfanumérico

Tabela 13 - Atributos da lista contida no atributo 'ORDERED_CLASS_TOTAL' da Tabela 7

Atributo	Informação	Formato
TAXNUM	CNPJ da empresa	Numérico
SUPPLIER_NAME	Nome da empresa	Alfanumérico
MEPP	Indicativo de microempresa	Alfanumérico
TOTAL_VALUE	Valor total da cotação	Numérico

Tabela 14 - Atributos da lista contida no atributo 'ORDERED_CLASS_LOTE' da Tabela 7

Atributo	Informação	Formato
TAXNUM	CNPJ da empresa	Numérico
SUPPLIER_NAME	Nome da empresa	Alfanumérico
MEPP	Indicativo de microempresa	Alfanumérico
TOTAL_VALUE	Valor total da cotação	Numérico

4.2.3 Buscar mensagens (R05)

A requisição de buscar mensagens (R05) traz as mensagens que foram publicadas no portal Petronect para cada licitação. A estrutura do objeto retornado no formato JSON está descrita nas Tabelas 15 a 18.

Tabela 15 - Atributos da requisição de buscar mensagens da licitação

Atributos	Informação	Formato
SUBJECTS	Assuntos	Lista
MESSAGES	Mensagens	Lista
NOTES	Notas	Lista

Tabela 16 - Atributos da lista contida no atributo 'SUBJECTS' da Tabela 15

Atributos	Informação	Formato
SUBJECT	Assunto da mensagem	Lista
MESSAGES		Lista
SUBJECT_ID	Código identificador do assunto	Alfanumérico

	SUBJECT_TITLE	Título da mensagem	Alfanumérico
	CREATED_AT_DATE	Data de criação da mensagem	Data (yyyy-mm-dd)
	CREATED_AT_TIME	Horário de criação da mensagem	Hora (hh:mi:ss)
	YPCOMT_LINES		Lista
	TDFORMAT	Campo de controle	Alfanumérico
	TDLINE	Conteúdo da mensagem	Alfanumérico
	ATTACH	Anexos a mensagem	Lista

Tabela 17 - Atributos da lista contida no atributo 'MESSAGES' da Tabela 15

Atributo	Informação	Formato	
SUBJECT_ID	Código identificador do assunto	Alfanumérico	
SUBJECT_TITLE	Título da mensagem	Alfanumérico	
CREATED_AT_DATE	Data de criação da mensagem	Data (yyyy-mm-dd)	
CREATED_AT_TIME	Horário de criação da mensagem	Hora (hh:mi:ss)	
YPCOMT_LINES		Lista	
	TDFORMAT	Campo de controle	Alfanumérico
	TDLINE	Conteúdo da mensagem	Alfanumérico
ATTACH	Anexos a mensagem	Lista	

Tabela 18 - Atributos da lista contida no atributo 'NOTES' da Tabela 15

Atributo	Informação	Formato
TDFORMAT	Campo de controle	Alfanumérico
TDLINE	Conteúdo da mensagem	Alfanumérico

4.2.4 Buscar informações adicionais (R06)

A requisição de informações adicionais (R06) complementa as informações trazidas pelas requisições que buscam a lista de licitações publicadas, em andamento e concluídas. A estrutura do objeto JSON retornado está descrita na Tabela 19.

Tabela 19 - Atributos da requisição de informações adicionais

Atributo	Informação	Formato
PREQUALIF_OPPORT	(não identificado)	-
P_DESCRIPTION	(não identificado)	-
HISTORY	Código indicativo que a licitação possui histórico	Alfanumérico
HISTORY_DYN_NAME	(não identificado)	-
MESSAGES	Código indicativo que a licitação possui mensagens	Alfanumérico
CO_CODE	Código da empresa	Numérico
EMP_DESCRIPTION	Nome da empresa	Alfanumérico
DESCRIPTION	Nome atribuído a licitação	Alfanumérico
STAT	Código do status da licitação	Alfanumérico
TXT30	(não identificado)	-

START_DATE	Data de início da licitação	Data (yyyy-mm-dd)
START_TIME	Horário de início da licitação	Hora (hh:mi:ss)
QUOT_DEAD	(não identificado)	-
QUOT_DEAD_TIME	(não identificado)	-
CREATED_AT_DATE	Data de criação da licitação	Data (yyyy-mm-dd)
CREATED_AT_TIME	Horário de criação da licitação	Hora (hh:mi:ss)
TZONE	Fuso horário	Alfanumérico
CURRENCY	Moeda	Alfanumérico
YPPAYCON	(não identificado)	-
YPCON_INCO1	(não identificado)	-
YPCON_MODALITY	Código da modalidade da licitação	Numérico
YPCON_MODALITY_NAME	Nome da modalidade da licitação	Alfanumérico
YPCON_QUOTE_REQUEST	(não identificado)	Numérico
ATTACHMENTS	Documentos da licitação	Lista
DISPLAY_PROP_LANCE	(não identificado)	Alfanumérico
PUBL_DOU_DATE	Data de publicação no DOU	Data (yyyy-mm-dd)
DISPUTE_MODE	Código do modo de disputa	Numérico
MEPP	Indicativo de Microempresa	Alfanumérico
AUC_START_DATE	Data início do pregão	Data (yyyy-mm-dd)
AUC_START_TIME	Data fim do pregão	Hora (hh:mi:ss)

O atributo 'ATTACHMENTS' apresenta lista com as informações dos documentos anexados a licitação. Sua estrutura é apresentada na Tabela 20.

Tabela 20 - Atributos da lista contida no atributo 'ATTACHMENTS' da Tabela 19

Atributo	Informação	Formato
DESCRIPTION	Nome do anexo	Alfanumérico
PHIO_OBJID	Código identificador do anexo	Alfanumérico

4.2.5 Modos de disputa (R07)

A requisição de modos de disputa (R07) é uma requisição auxiliar disparada para compor os campos de filtro das buscas do portal Petronect e retorna lista dos possíveis modos de disputas utilizados nas licitações. Sua estrutura é apresentada na Tabela 21

Tabela 21 - Atributos da requisição de Modos de Disputa

Atributos	Informação	Formato
id	Identificador da requisição do modo de disputa	Alfanumérico
uri	URI da requisição do modo de disputa	Alfanumérico
type	Tipo da requisição	Alfanumérico
lvLangu	Informação de Língua do modo de disputa	Alfanumérico

EvVal	Código do modo de disputa	Numérico
EvDescription	Descrição do modo de disputa	Alfanumérico

4.2.6 Empresas (R08)

A requisição de empresas (R08), assim como a requisição dos modos de disputa (R07), é uma requisição auxiliar utilizada para compor os campos de filtro das buscas que podem ser realizadas pelo portal Petronect. A requisição retorna, em formato JSON, as empresas do grupo Petrobras que utilizam o Petronect para compras e contratações. Sua estrutura é apresentada na Tabela 22.

Tabela 22 - Atributos da requisição de Empresas

Atributo	Informação	Formato
id	Identificador da requisição da empresa	Alfanumérico
uri	URI da requisição da empresa	Alfanumérico
type	Tipo da requisição	Alfanumérico
Code	Código identificador da empresa	Numérico
Description	Nome da empresa	Alfanumérico
Logsys	(não identificado)	-
StructSearch	(não identificado)	-

4.2.7 Status das licitações concluídas (R09)

A requisição de status das licitações concluídas (R09) retorna, em formato JSON, o conjunto de status que uma licitação concluída pode assumir. É uma requisição auxiliar utilizada para compor os campos de filtro das buscas que podem ser realizadas pelo portal Petronect. Os dados retornados na requisição estão descritos na Tabela 23.

Tabela 23 - Atributos da requisição de Status das licitações concluídas

Atributo	Informação	Formato
id	Identificador da requisição do status	Alfanumérico
uri	URI da requisição do status	Alfanumérico
type	Tipo da requisição	Alfanumérico
lvLangu	Informação de língua do status	Alfanumérico
EvVal	Código do status	Numérico
EvDescription	Descrição do status	Alfanumérico

4.2.8 Status das licitações em andamento (R10)

A requisição de status das licitações em andamento (R10), assim como para as licitações concluídas (R09), retorna o conjunto de status que uma licitação em andamento pode assumir. É uma requisição auxiliar utilizada para compor os campos de filtro das

buscas que podem ser realizadas pelo portal Petronect. Sua estrutura é apresentada na Tabela 24.

Tabela 24 - Atributos da requisição de Status das licitações em andamento

Campo	Informação	Formato
id	Identificador da requisição do status	Alfanumérico
uri	URI da requisição do status	Alfanumérico
type	Tipo da requisição	Alfanumérico
lvLangu	Informação de língua do status	Alfanumérico
EvVal	Código do status	Numérico
EvDescription	Descrição do status	Alfanumérico

4.2.9 Baixar anexo (R11)

A requisição de baixar anexo (R11) retorna o arquivo requisitado.

Verificou-se que uma licitação pode possuir arquivos nos formatos PDF, DOC, DOCX e ZIP. Todos os arquivos listados como anexo da licitação são baixados pelo software desenvolvido e entregues ao sistema Alice, independentemente do seu formato. O sistema Alice, por sua vez, realiza a extração de arquivos compactados, converte documentos em texto e realiza análises textuais.

4.3 Explorando os dados

Para as requisições de licitações em andamento e concluídas, é necessário buscar por intervalos de datas. Inicialmente foi tentado buscar os dados relativos a um período de 90 dias anteriores a data da consulta. Contudo, a interface da requisição de licitações concluídas exige que o período pesquisado tenha um intervalo máximo de 30 dias. Assim, foram realizadas requisições limitadas a períodos de 30 dias entre os dias 01/11/2019 e 17/01/2020. Ao final do período, foram extraídos dados de 1472 licitações, somando-se os resultados das requisições de licitações publicadas, em andamento e concluídas.

Analisando-se os dados coletados, verificou-se que a maioria dos valores dos campos são alfanuméricos. Há, também, muitos campos com o formato de data e hora que poderão ser condensados em apenas um parâmetro para representar a data em sua completude.

Algumas requisições possuem informações em comum. Por exemplo, as informações principais da requisição que lista as licitações publicadas (R01), em andamento (R02) e concluídas (R03) também constam nas requisições de informações adicionais (R06).

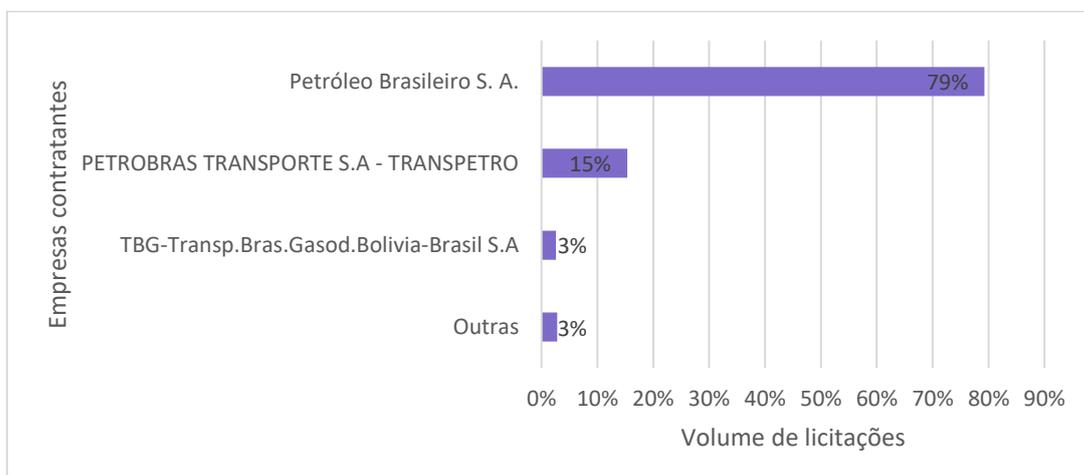
Durante a exploração, não foi possível identificar o significado de alguns campos retornados pelas requisições porque os campos não retornaram valores. Entretanto, foi possível identificar diversos campos que poderão ser úteis para o atingimento do objetivo proposto de melhoria da atuação da Secretaria de Fiscalização. Entre eles estão campos com informação do número identificador da licitação, datas de início e fim da licitação, modo de disputa, modalidade, empresa responsável pela licitação, os itens que são objetos da licitação, as cotações recebidas para cada item e os lances realizados durante uma licitação na modalidade Pregão.

O campo 'OPPORT_NUM', que contém o código identificador da licitação, é passado como parâmetro principal das requisições de informações de histórico (R04), informações adicionais (R06) e mensagens (R05), portanto, foi utilizado como chave no momento de unir as diversas informações.

Os dados extraídos foram explorados com o auxílio das seguintes bibliotecas python: 'Pandas Profiling', 'matplotlib', 'seaborn'. Adicionalmente, algumas visualizações foram construídas em painel eletrônico no QlikView e no Excel.

Utilizando-se tais ferramentas, verificou-se que todas as licitações apresentaram informações da empresa contratante e que 79% das licitações retornadas foram realizadas pela empresa Petróleo Brasileiro S. A., 15% foram realizadas pela empresa Petrobras Transporte S. A. - Transpetro, 3% foram realizadas pela empresa TBG-Transp.Bras.Gasod.Bolivia-Brasil S.A. e 3% pelas demais empresas, conforme apresentado na Figura 13.

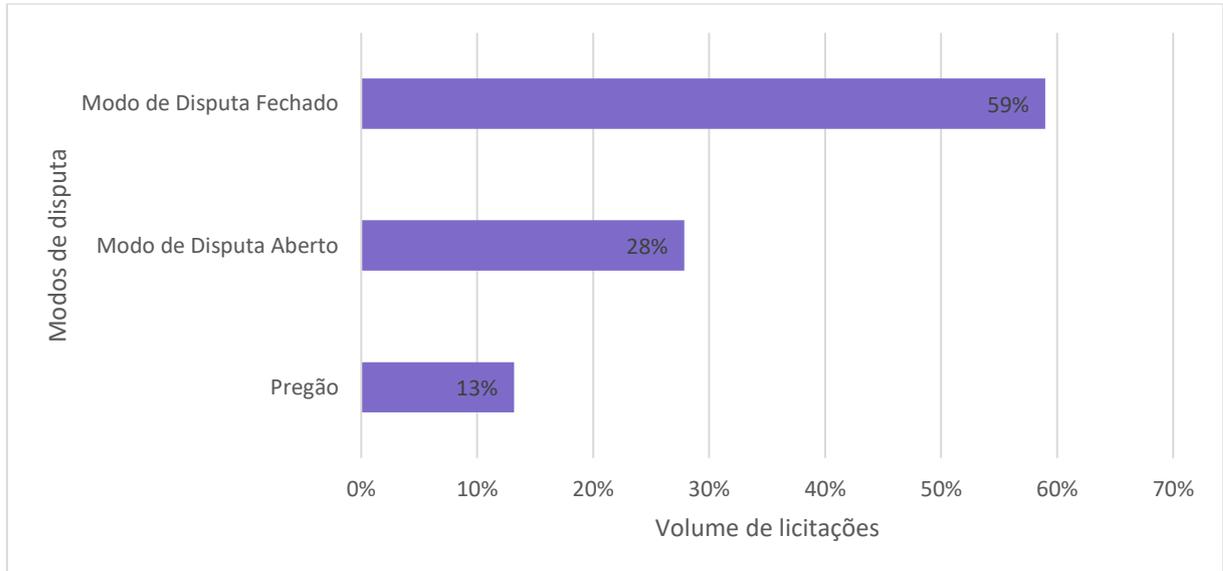
Figura 13 - Exploração de dados - Percentual de empresas contratantes das licitações



Fonte: própria

Além disso, todas as licitações apresentaram o modo de disputa em que foram realizadas. O modo de disputa mais utilizado nos dados coletados foi o Modo de Disputa Fechado, representando 59% das licitações (Figura 14).

Figura 14 - Exploração de dados - Percentual de licitações por modos de disputa



Fonte: própria

De acordo com o Regulamento de Licitações e Contratos da Petrobras (RLCP) (Petrobras, 2018) as licitações da Petrobras podem ser realizadas nos seguintes modos de disputa:

- rito do pregão;
- modo de disputa aberto;
- modo de disputa fechado;
- modo de disputa combinado

No modo de disputa aberto, os participantes apresentam em sessão pública lances públicos e sucessivos.

No modo de disputa fechado, as propostas apresentadas pelos participantes são sigilosas até a data e hora designadas para sua divulgação.

Com relação a modalidade utilizada nas licitações, 85% das licitações utilizaram a modalidade Licitação, Lei 13.303, Art. 28, CAPUT e 12,7% das licitações utilizaram a

modalidade Pregão, Lei 13.303, Art. 32, inciso IV (Figura 15). Ressalta-se que 2,3% das licitações obtidas não tiveram o campo indicativo de modalidade preenchido.

Figura 15 - Exploração de dados - Modalidade

Value	Count	Frequency (%)
Licitação, Lei 13.303, Art. 28, CAPUT	1251	85.0%
Pregão, Lei 13.303, Art. 32, inciso IV	187	12.7%
	34	2.3%

Fonte: própria

Cada licitação pode possuir um ou mais itens que a empresa contratante deseja adquirir e cada um dos itens pode ser adquiridos na quantidade definida no edital da disputa. Com relação aos itens das licitações, foi possível identificar 3424 descrições de itens distintos nos dados coletados em um total de 5637 itens. Foi possível recuperar as descrições, tipos, quantidades e unidades dos itens. Os dados também apresentam informações acerca da família dos itens, contudo mais de 95% dos itens não apresentaram esse campo preenchido, conforme mostrado na Figura 16.

Figura 16 - Exploração de dados - Família dos itens

Value	Count	Frequency (%)
	5371	95.3%
Bens Simplificado Nova Lei	74	1.3%
Bens corporativo Nova Lei	55	1.0%
Serviços corporativo Nova Lei	26	0.5%
Família Padrão para Serviços - Registro Corporativo	15	0.3%
Máquinas, ferramentas hidráulicas, seus componentes e acessórios	12	0.2%
Unidade hidráulica para sistema de controle hidráulico direto	7	0.1%
Anéis elastoméricos perfilados	6	0.1%
Parafuso estojo para ligações críticas	6	0.1%
Família Padrão para Bens RQT - Registro Corporativo	4	0.1%
Other values (34)	61	1.1%

Fonte: própria

Do total dos itens retornados, notou-se que os mais frequentes possuem descrição genérica como ‘MANUTENÇÃO EM MÁQUINAS/EQUIPAM. - S/MAT’ ou ‘VALOR TOTAL DA PPU’ (Figura 17).

Figura 17 - Exploração de dados - Descrição dos itens das licitações

Value	Count	Frequency (%)
MANUTENÇÃO EM MÁQUINAS/EQUIPAM. - S/MAT.	88	1.6%
LIMPEZA FINAL DA OBRA/DESMOBILIZAÇÃO	52	0.9%
ADEQUAÇÃO DRENOS E SUSPIROS C/ MATERIAL	51	0.9%
VALOR TOTAL DA PROPOSTA	51	0.9%
VALOR TOTAL DA PPU	50	0.9%
LEVANTAMENTO DE CAMPO E PROJ. EXECUTIVO	50	0.9%
MOBILIZAÇÃO / CANTEIRO DE OBRA	50	0.9%
AS BUILT E DATA BOOK	50	0.9%
ANIDRO REDUC	44	0.8%
FEIXE TUBULAR P/PERMUTADOR	41	0.7%
Other values (3414)	5110	90.7%

Fonte: própria

Foram identificados 3 valores referentes aos tipos dos itens (Figura 18), sendo que 58% são relacionados a Materiais ('MATL'), 23,4% relacionados a Serviços ('SERVICE'). O valor 'OUTL', que representou 15,9% dos tipos de itens, parece representar um agrupador de itens diferentes dentro da mesma licitação. Algumas disputas apresentaram apenas itens do tipo 'OUTL'. Itens sem classificação por tipo representaram 2,7% dos valores.

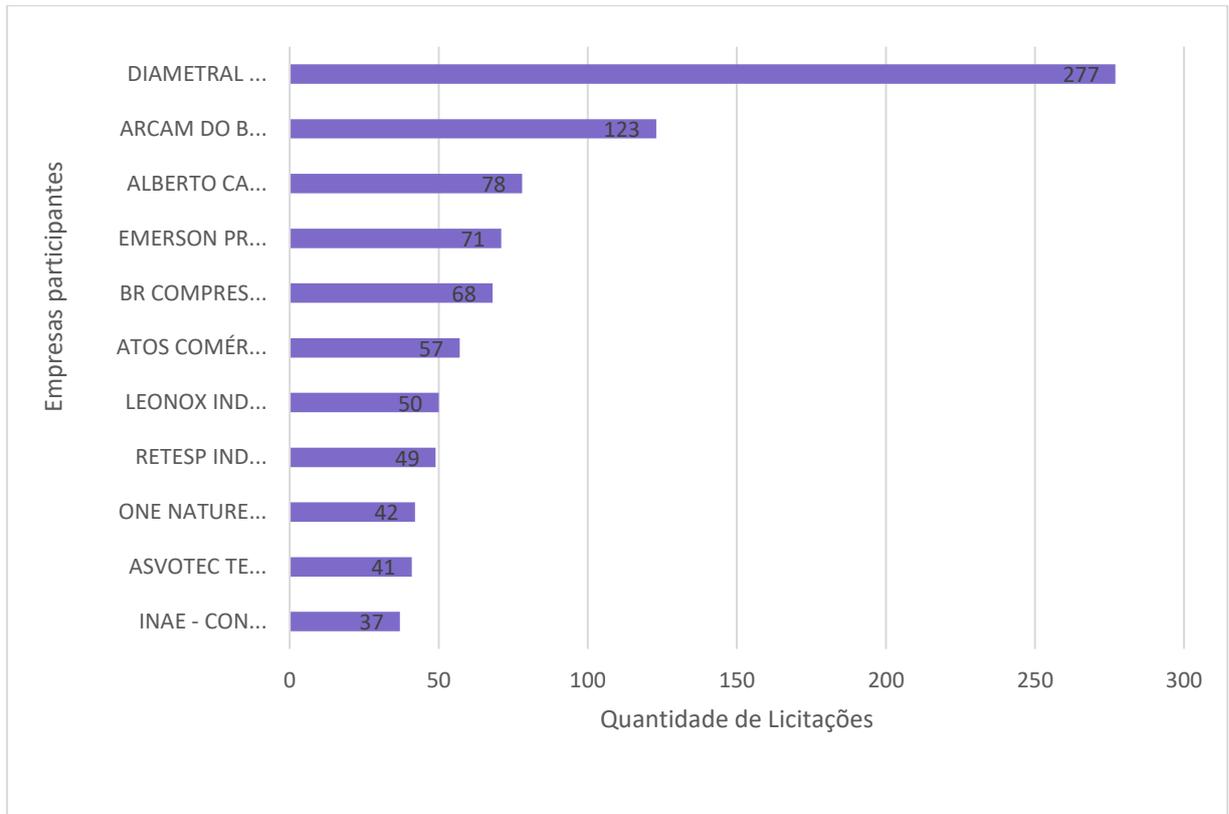
Figura 18 - Exploração de dados - Tipos dos itens

Value	Count	Frequency (%)
MATL	3268	58.0%
SERVICE	1321	23.4%
OUTL	898	15.9%
	150	2.7%

Fonte: própria

Com relação aos dados de cotações dos itens, foi possível identificar dados sobre as empresas que forneceram cotações, como CNPJ e nome, os valores das cotações e a data da cotação. A Figura 19 apresenta as 10 empresas que mais participaram das licitações extraídas.

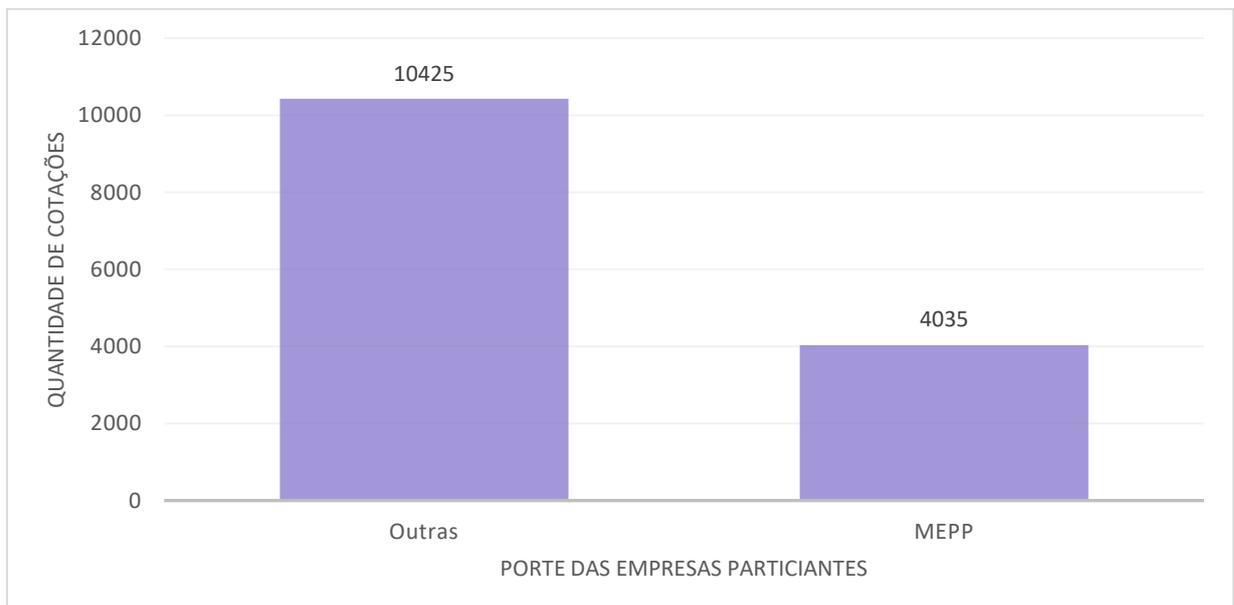
Figura 19 - Exploração de dados - Empresas que mais participaram



Fonte: própria

Foi possível identificar também, conforme a Figura 20, a distribuição da participação de micro e pequenas empresas (MEPP) no fornecimento das cotações.

Figura 20 - Quantidade de cotações por porte da empresa



Fonte: própria

Um dos campos recuperados apresenta a classificação das cotações realizadas pelas empresas concorrentes. Avaliando-se a frequência de cada classificação atribuída às cotações, nota-se uma alta frequência do valor ‘Proposta não Avaliada’, conforme mostrado na Figura 21.

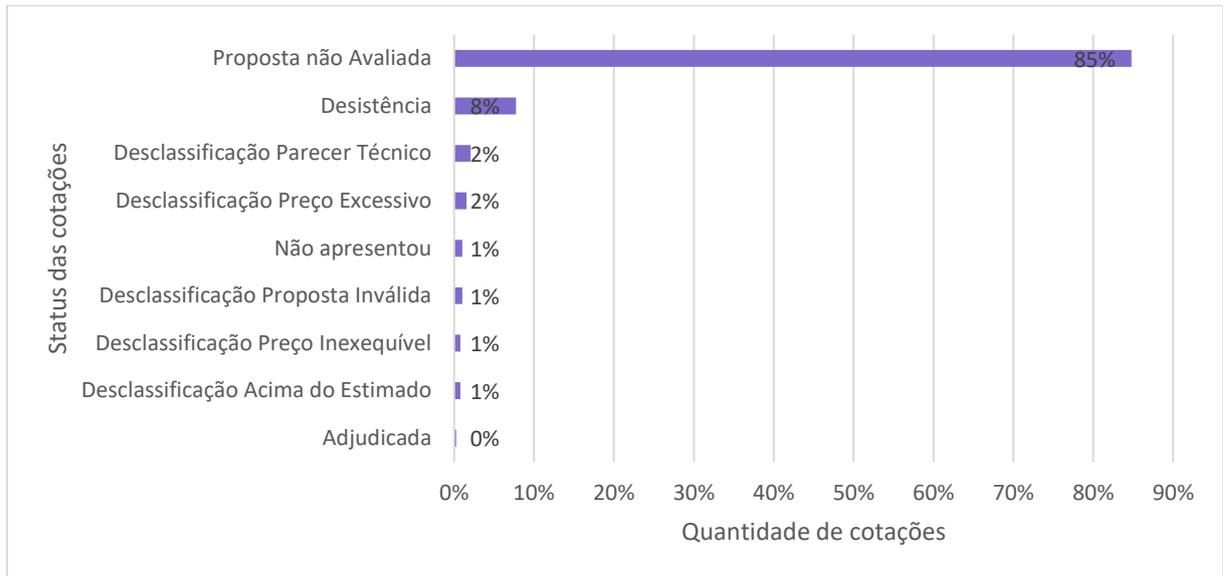
Figura 21 - Exploração de dados - Classificação das cotações

Value	Count	Frequency (%)	
Proposta não Avaliada	13357	92.4%	
Desclassificação Preço Excessivo	398	2.8%	
Desistência	178	1.2%	
Desclassificação Proposta Inválida	148	1.0%	
Desclassificação Parecer Técnico	123	0.9%	
Habilitada	96	0.7%	
Classificada/Aceita	59	0.4%	
Desclassificação Acima do Estimado	40	0.3%	
Não apresentou	24	0.2%	
Desclassificação Preço Inexequível	13	0.1%	
Other values (3)	24	0.2%	

Fonte: própria

Filtrando-se apenas os dados de licitações concluídas, o percentual de cotações no estado ‘Proposta não Avaliada’ também apresentou alta frequência, cerca de 85% (Figura 22).

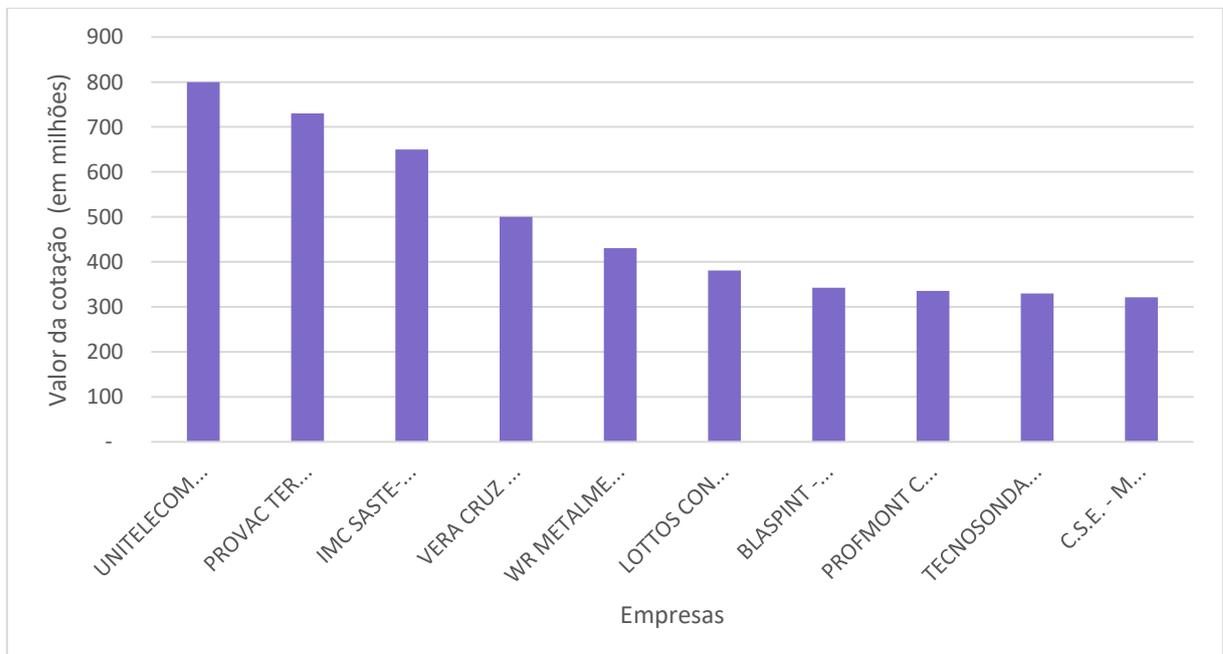
Figura 22 - Exploração de dados - Classificação das cotações - Licitações concluídas



Fonte: própria

Com relação aos valores das cotações foi possível identificar os 10 maiores valores ofertados e as empresas que realizaram a oferta. O maior valor ofertado em uma cotação foi de R\$ 800 milhões, conforme mostrado na Figura 23. Entretanto, essa cotação foi classificada como 'Proposta não Avaliada'.

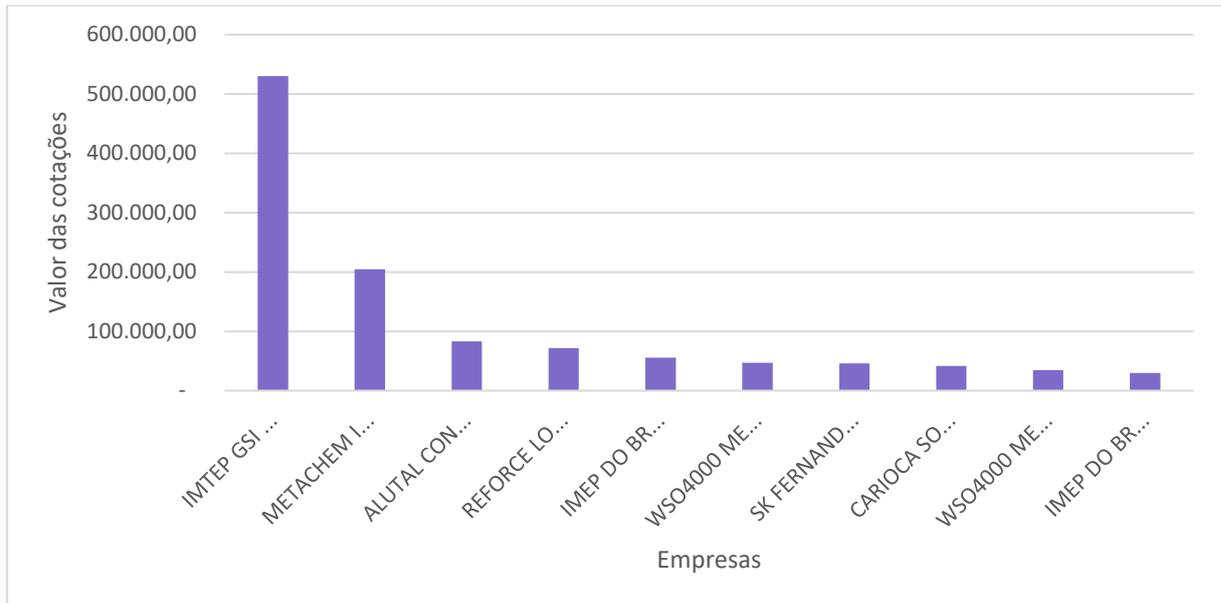
Figura 23 - Exploração de dados - Maiores cotações



Fonte: própria

Filtrando-se as cotações com classificação 'Homologada' (Figura 24), os valores apresentados são bem inferiores aos trazidos na Figura 23.

Figura 24 - Exploração de dados - Valores das cotações homologadas



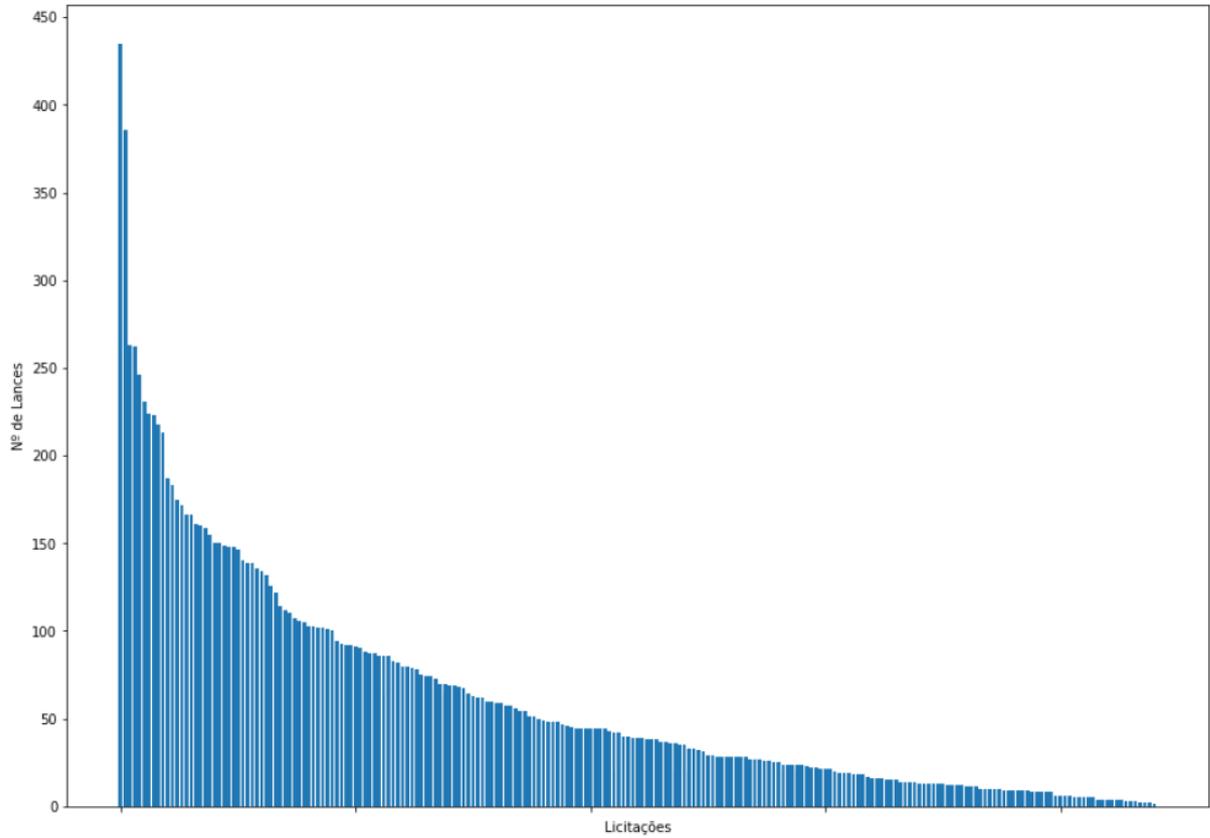
Fonte: própria

Em licitações na modalidade 'Pregão', a disputa entre as empresas se dá pela apresentação dos valores ofertados por meio de lances. Os dados relacionados aos lances efetuados em licitações dessa modalidade trazem informações sobre a empresa que efetuou o lance, a data e hora do lance e o valor do lance.

A Figura 25 mostra que foi possível identificar qual o pregão que obteve o maior número de lances. Foram apresentados 435 lances e participaram dessa licitação, 62 empresas.

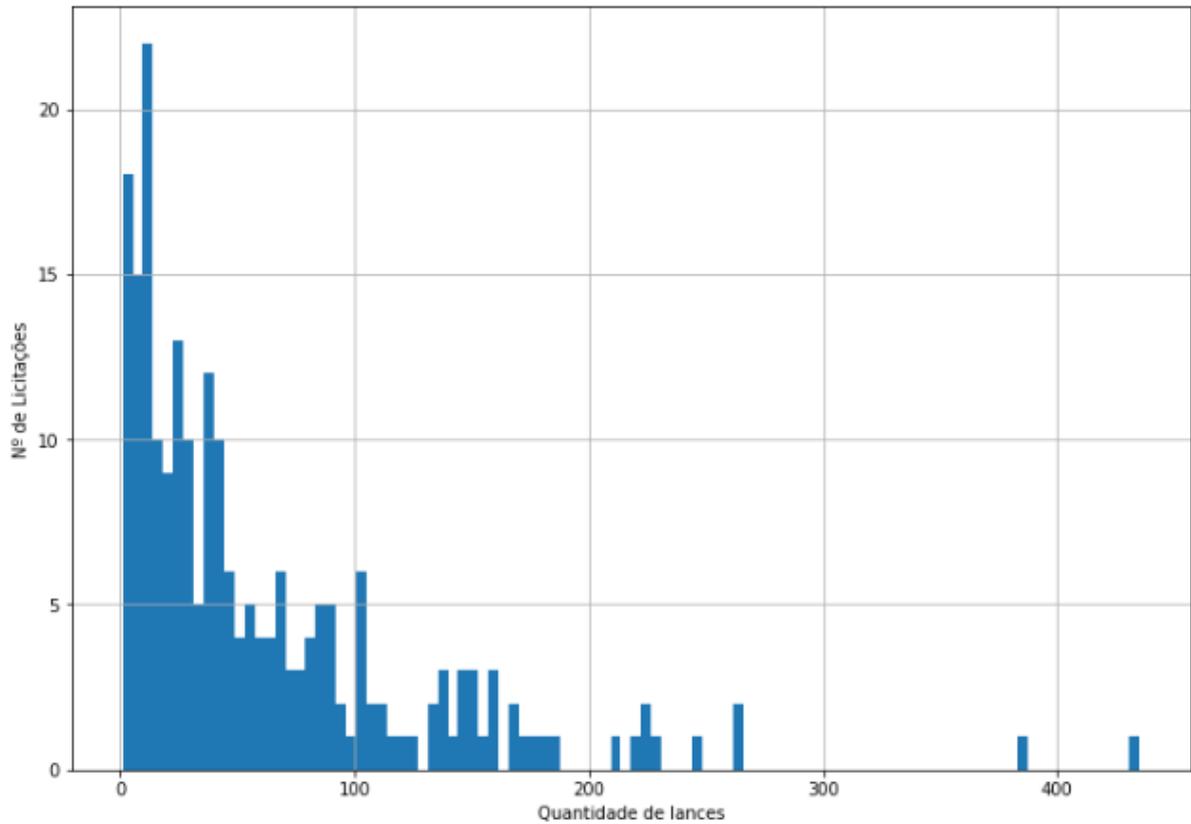
Analisando-se a distribuição de frequência de licitações relacionadas ao número de lances, conforme mostrado na Figura 26, percebe-se uma maior concentração de licitações que tiveram menos de 100 lances.

Figura 25 - Exploração de dados - Quantidade de lances por pregão



Fonte: própria

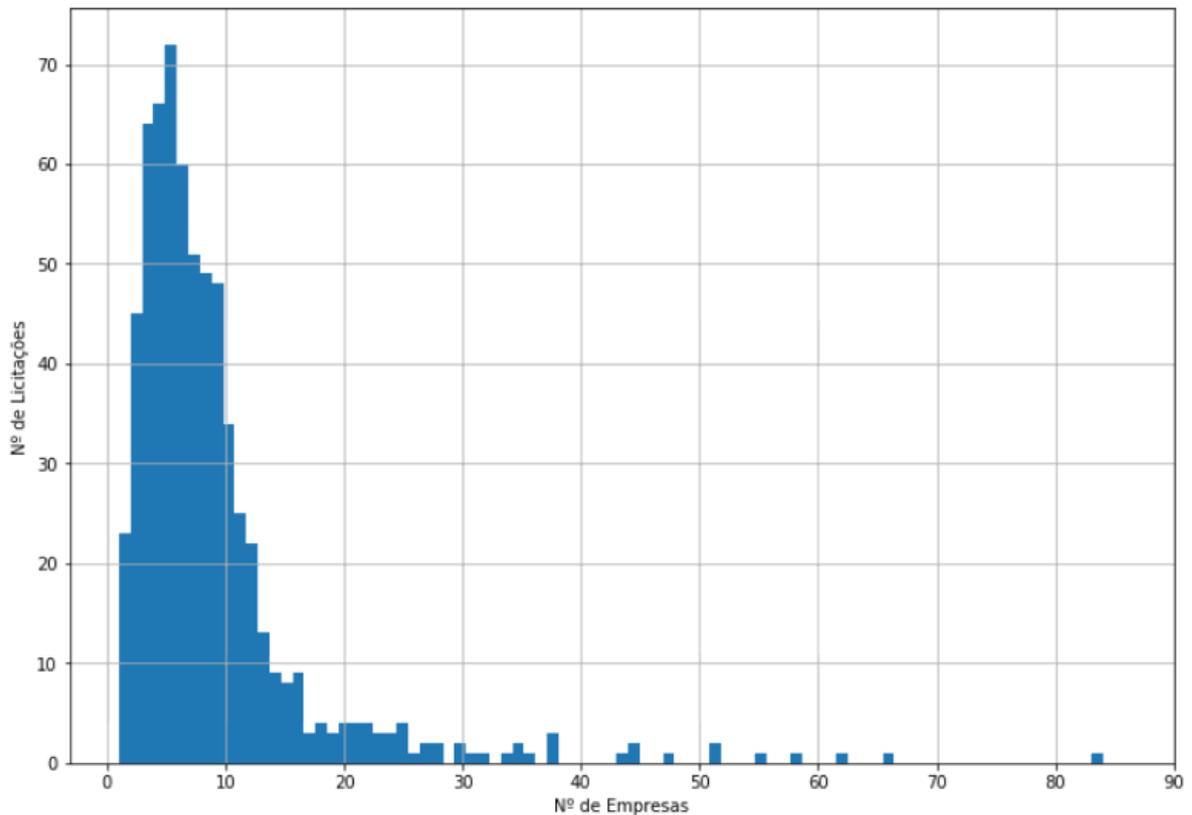
Figura 26 - Exploração de dados - Frequência de licitações pela quantidade de lances.



Fonte: própria

Para avaliar a competitividade das licitações, plotou-se o gráfico da Figura 27. Nela é possível verificar o número de empresas que participaram nas licitações. Observa-se que a maioria das licitações tiveram até 10 empresas concorrentes.

Figura 27 - Exploração de dados - Número de empresas participantes em cada licitação



Fonte: própria

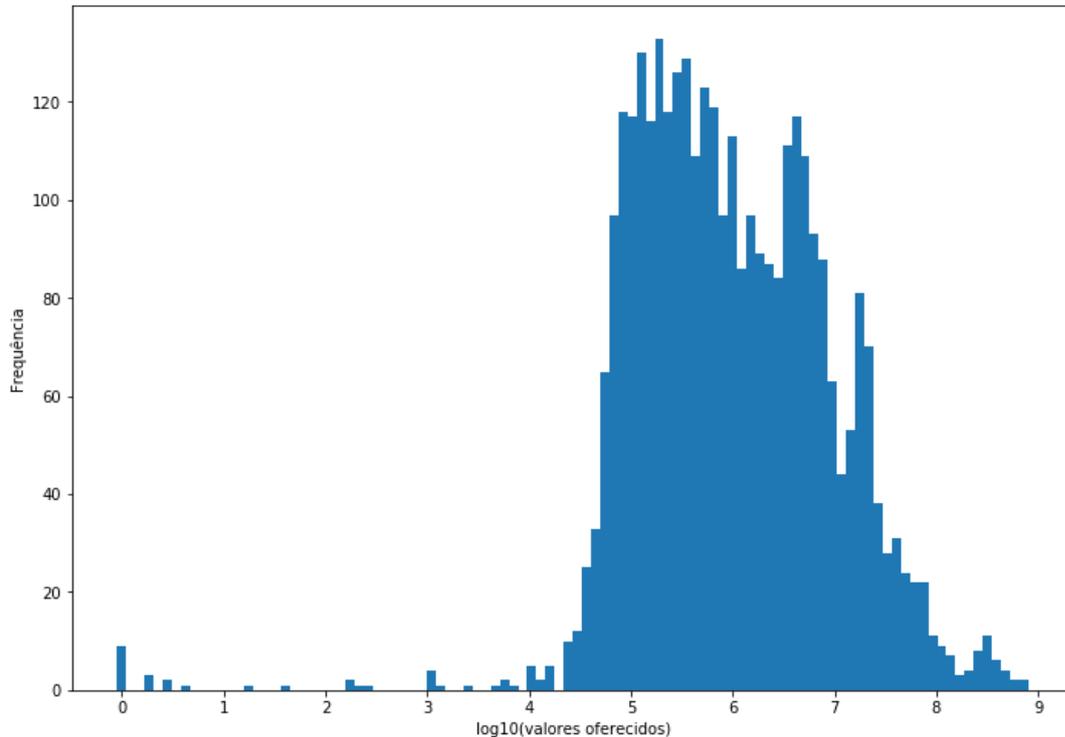
Foi possível recuperar também as mensagens trocadas entre o pregoeiro e os participantes ao longo do pregão. Entretanto, o teor das mensagens se refere apenas a controles de início e fim do pregão e dúvidas relacionadas à operacionalização do pregão.

Identificou-se que, terminado a fase de lances no pregão, o último lance de cada empresa participante é registrado como uma cotação da licitação.

O resultado dos pregões ou cotações parecem ser disponibilizados na lista 'ORDER_CLASS_TOTAL'. Essa lista apresenta informações do CNPJ e nome das empresas, indicador de micro ou pequena empresa, valores totais oferecidos. A lista aparece ordenada pelo valor total de forma crescente.

A Figura 28 mostra a frequência dos valores ofertados em escala logarítmica. Observa-se que os valores mais frequentes se encontram no intervalo de R\$100 mil a R\$10 milhões.

Figura 28 - Exploração de dados - Distribuição de valores das cotações



Fonte: própria

Verificou-se também que dentre as informações coletadas existem alguns campos de algumas requisições que parecem ser de controle do sistema gerador da informação (Logsys, StructSearch) e que não possuem valor para este projeto que busca informações sobre as licitações.

4.4 Verificando a qualidade dos dados

De maneira geral os dados parecem estar consistentes e com poucos valores faltantes.

Relacionados a requisição de informações sobre as licitações, os campos que apresentaram maior quantidade de valores faltantes (60,5%) foram a data de início do pregão ('AUC_START_DATE' e 'AUC_START_TIME') e a data de finalização do pregão ('AUC_END_DATE' e 'AUC_END_TIME'), o que pode estar ligado ao fato de a maioria das licitações não serem da modalidade pregão.

As informações de substatus da licitação também apresentou alto índice de valores faltantes (44%). Isso se deve ao fato das licitações que estão no status de ‘publicadas’ serem maioria da amostra e não possuem informações de substatus.

Nos dados relacionados aos itens das licitações, os campos relacionados à descrição da família do item apresentaram alto percentual de valores faltantes, cerca de 95% dos itens não possuem a descrição da família a que pertencem. Contudo, o código da família apresentou um percentual menor de valores faltantes, 24%.

Alguns campos retornados pelas requisições não trouxeram quaisquer valores preenchidos. Assim, não foi possível determinar seu real significado e utilidade para o projeto.

5 PREPARAÇÃO DOS DADOS

Após coletar e explorar os dados iniciais, se faz necessário selecionar os dados que serão armazenados e estruturá-los de forma a tornar possível sua internalização e disponibilização. Nesta sessão, serão apresentados os atributos selecionados e os campos em que a informação foi estruturada e armazenada no banco de dados.

5.1 Selecionando dados

Com o objetivo de fornecer informações que possam melhorar a atuação da Secretaria de Fiscalização de Petróleo e Gás Natural, buscou-se selecionar campos que apresentaram informações que caracterizem as licitações e seus itens, assim como identificar os valores e empresas envolvidas nas contratações. As informações acerca dos documentos integrantes da licitação também foram mantidas para possibilitar a integração com o sistema Alice.

5.1.1 Licitações Públicas - Publicadas/Em Andamento/Concluídas (requisições R01, R02 e R03)

Da requisição principal de informações de licitações publicadas, em andamento e concluída, optou-se em internalizar os atributos nos campos do banco de dados descritos na Tabela 25.

Tabela 25 - Dados selecionados - Licitações

Atributos	Campos (tabela.campo)
AUC_START_DATE	licitacao.data_inicio_pregao
AUC_START_TIME	
COMPANY	empresa.cod_empresa
COMPANY_DESC	empresa.nome_empresa

DISPUTE_MODE	licitacao.cod_modulo_disputa
DOU_PUBL_DATE	licitacao.data_publicacao_dou
START_DATE	licitacao.data_inicio
START_HOUR	
END_DATE	licitacao.data_fim
END_HOUR	
OPPORT_DESCR	licitacao.descr_licitacao
OPPORT_NUM	licitacao.cod_licitacao
OPPORT_TYPE	licitacao.tipo_licitacao
STATUS	status.cod_status_petro
STATUS_DESC	status.descr_status
SUB_STATUS	substatus.cod_substratus_petro
ANEXOS	anexo
ITEMS	item_licitacao
REGIONS	regiao

Os dados dos anexos serão todos mantidos para possibilitar a recuperação dos documentos. Eles foram gravados na tabela ‘anexo’ do banco de dados conforme a Tabela 26.

Tabela 26 - Dados selecionados - Licitações - Anexos

Atributo	Campo (tabela.campo)
DESCRIPTION	anexo.descr_anexo
PHIO_OBJID	anexo.phio_objid

Os itens das licitações podem ser caracterizados pelas informações retornadas pelas requisições R01, R02 e R03. Portanto todos os atributos relacionados a eles foram mantidos conforme a Tabela 27.

Tabela 27 - Dados selecionados - Licitações - Itens

Atributo	Campo (tabela.campo)
ITEM_NUM	item_licitacao.numero_item
EXLIN	item_licitacao.exlin
ITEM_DESC	item_licitacao.descr_item_licitacao
ITEM_NOTES	nota_item
UNIT	item_licitacao.unidade_item
QUANTITY	item_licitacao.quantidade
DELIV_DATE	item_licitacao.data_entrega
GROUPING_LEVEL	item_licitacao.nivel_agrupamento

ITEM_PROCESS_TYP	item_licitacao.tipo_item
FAMILY	item_licitacao.cod_familia
FAMILY_DESCR	item_licitacao.descr_familia
NUM_MATERIAL	item_licitacao.num_material
FAMILY_CLASSIF	item_licitacao.classif_familia

Com relação à lista de notas apresentada para cada item, decidiu-se manter o atributo 'TDLINE' que apresenta o conteúdo da nota no campo 'descr_nota' da tabela 'nota_item'.

Os dados contidos na lista 'REGIONS' também foram mantidos para possibilitar análise regionalizada das licitações. Assim, os dados foram armazenados de acordo com a Tabela 28.

Tabela 28 - Dados selecionados - Licitações - Regiões

Atributo	Campo (tabela.campo)
COUNTRY	região.pais
REGION	região.uf
REGION_DESCRIPTION	região.desc_uf

Alguns atributos foram retirados do projeto por apresentarem informações que não agregariam ao alcance do objetivo final do projeto ou até mesmo atributos que não apresentaram dados preenchidos.

5.1.2 Histórico (R04)

A requisição de histórico traz dados adicionais que poderão ser úteis nas análises referentes às concorrências geradas durante a licitação. Dados dos valores contratados, dos lances do pregão e da classificação das empresas participantes são de grande valia para fins fiscalizatórios.

O único atributo que se julga desnecessário é o 'HAS_DIVULG_REL' que aparenta ser utilizado pelo portal Petronect para fins de controle de exibição da informação e por esse motivo será descartado.

Os demais atributos presentes nas listas dessa requisição serão mantidos integralmente, conforme a Tabela 29.

Tabela 29 - Dados selecionados - Histórico

Atributo	Campo (tabela.campo)
-----------------	-----------------------------

RELEVANT_DATES	data_relevante
QUOTATION_LIST	cotacao
AUCTION_LIST	lance
MESSAGES_LIST	mensagem_pregao
ORDERED_CLASS_TOTAL	classificacao_total
ORDERED_CLASS_LOTE	classificacao_lote
HAS_DIVULG_REL	licitacao.houve_divulgacao

Os dados do atributo ‘RELEVANT_DATES’ foram armazenados conforme Tabela 30.

Tabela 30 - Dados selecionados - ‘RELEVANT_DATES’

Atributo	Campo (tabela.campo)
START_DATE	data_relevante.data_inicio
START_TIME	
END_DATE	data_relevante.data_fim
END_TIME	
COMMERCIAL_PROP_OPEN_DATE	data_relevante.inicio_prop_comercial
COMMERCIAL_PROP_OPEN_TIME	
PUBL_DATE	data_relevante.data_publicacao
RECURSAL_PERIOD	data_relevante.data_publicacao

Os dados selecionados do atributo ‘QUOTATION_LIST’ foram armazenados conforme a Tabela 31 e Tabela 32

Tabela 31 - Dados selecionados - ‘QUOTATION_LIST’

Atributo	Campo (tabela.campo)
NUMBER_INT	cotacao.cod_item_licitacao
QUOTS	cotacao

Por sua vez, a Tabela 32 apresenta a lista de informações das cotações que foram selecionadas.

Tabela 32 - Dados selecionados - ‘QUOTS’

Atributo	Campo (tabela.campo)
TAXNUM	cotacao.taxnum
SUPPLIER_NAME	cotacao.supplier_name

MEPP	cotacao.mepp
QUANTITY	cotacao.quantity
UNIT_VALUE	cotacao.unity_value
TOTAL_VALUE	cotacao.total_value
CREATED_AT_DATE	cotacao.created_at
CREATED_AT_TIME	
CHARACTERISTIC	cotacao.characteristic

A Tabela 33 apresenta os dados selecionados do atributo 'AUCTION_LIST':

Tabela 33 - Dados selecionados - 'AUCTION_LIST'

Atributo	Campo (tabela.campo)
TAXNUM	lance.taxnum
AUC_VALUE	lance.auc_value
CREATED_AT_DATE	lance.created_at
CREATED_AT_TIME	

Os dados do atributo 'MESSAGES_LIST' foram armazenados conforme Tabela 34.

Tabela 34 - Dados selecionados - MESSAGES_LIST

Atributo	Campo (tabela.campo)
AUTOR	mensagem_leilao.autor
MESSAGE	mensagem_leilao.mensagem
CREATED_AT_DATE	mensagem_leilao.created_at
CREATED_AT_TIME	mensagem_leilao.created_at

Os dados do atributo 'ORDERED_CLASS_TOTAL' foram armazenados conforme Tabela 35.

Tabela 35 - Dados selecionados - ORDERED_CLASS_TOTAL

Atributo	Campo (tabela.campo)
TAXNUM	classificacao_total.taxnum
SUPPLIER_NAME	classificacao_total.supplier_name
MEPP	classificacao_total.mepp
TOTAL_VALUE	classificacao_total.total_value

Os dados do atributo ‘ORDERED_CLASS_LOTE’ foram armazenados conforme Tabela 36.

Tabela 36 - Dados selecionados - ORDERED_CLASS_LOTE

Atributo	Campo (tabela.campo)
TAXNUM	classificacao_lote.taxnum
SUPPLIER_NAME	classificacao_lote.supplier_name
MEPP	classificacao_lote.mepp
TOTAL_VALUE	classificacao_lote.total_value

5.1.3 Buscar mensagens (R05)

A requisição de buscar mensagens traz as mensagens que foram publicadas no portal Petronect para cada licitação. Optou-se por internalizar os dados das mensagens para ser possível futura análise dos conteúdos das mensagens. Entretanto, as três listas apresentam o mesmo conteúdo. Assim, serão descartados os dados contidos nas listas ‘SUBJECTS’ e ‘NOTES’. Serão mantidos apenas os dados da lista ‘MESSAGES’.

O atributo de controle ‘TDFORMAT’ não apresenta informações relevantes e será descartado.

Os atributos foram armazenados na tabela ‘mensagem_licitacao’ conforme Tabela 37.

Tabela 37 - Dados selecionados - ‘MESSAGES’

Atributo	Campo (tabela.campo)
SUBJECT_TITLE	mensagem_licitacao.titulo_mensagem
CREATED_AT_DATE	mensagem_licitacao.data_mensagem
CREATED_AT_TIME	
YPCOMT_LINES	
TDLINE	mensagem_licitacao.descr_mensagem
ATTACH	anexo_mensagem

As informações dos anexos das mensagens foram armazenadas na tabela ‘anexo_mensagem’ conforme Tabela 38.

Tabela 38 - Dados selecionados - 'ATTACH'

Atributo	Campo (tabela.campo)
DESCRIPTION	anexo_mensagem.descr_anexo

ATTACH_OBJID	anexo_mensagem.phio_objid
--------------	---------------------------

5.1.4 Buscar informações adicionais (R06)

Os dados obtidos na requisição de dados adicionais serão utilizados para enriquecer as informações sobre as licitações. Essa requisição possui muitos campos que já são obtidos nas requisições de licitações publicadas, em andamento e concluídas. Assim, foram agregados apenas os dados complementares de moeda utilizada na licitação, código e descrição da modalidade utilizada. Os dados selecionados foram armazenados conforme a Tabela 39.

Tabela 39 - Dados selecionados - Informações adicionais

Atributo	Campo (tabela.campo)
CURRENCY	licitacao.moeda
YPCON_MODALITY	licitacao.cod_modalidade
YPCON_MODALITY_NAME	licitacao.tipo_modalidade

5.1.5 Modos de disputa (R07)

Os atributos selecionados da requisição dos modos de disputa relacionam-se aos campos de código e descrição do modo de disputa e serão mantidos pois possuem relacionamento direto com as informações contidas nos dados das licitações.

Os atributos foram armazenados na tabela 'modo_disputa' conforme Tabela 40

Tabela 40 - Dados selecionados - Modos de disputa

Atributo	Campo (tabela.campo)
EvVal	modo_disputa.cod_mododo_disputa
EvDescription	modo_disputa.descr_mododo_disputa

5.1.6 Empresas (R08)

Os dados selecionados sobre as empresas são o código e a o nome da empresa contratante. Esses dados se relacionam diretamente com os dados das licitações e foram armazenados na tabela 'empresa' conforme a Tabela 41.

Tabela 41 - Dados selecionados - Empresas

Atributo	Campo (tabela.campo)
Code	empresa.cod_empresa
Description	empresa.nome_empresa

5.1.7 Status das licitações concluídas e em andamento (R09 e R10)

As requisições de status das licitações concluídas e andamento retornam dados similares e complementares e por isso serão mantidas apenas os dados referentes ao código e descrição dos status. Os dados foram armazenados na tabela ‘status’

Tabela 42 - Dados selecionados - Status

Atributo	Campo (tabela.campo)
EvVal	status.cod_status_petro
EvDescription	status.descr_status

5.2 Limpando dados

Os campos que não apresentaram dados ou não continham informações relevantes para o projeto foram descartados.

5.3 Construindo novos dados

Os atributos retornados pelas requisições com informações de data e hora foram fundidos em apenas um campo de data no banco de dados.

Os atributos que foram utilizados da requisição de informações adicionais foram incorporados como informações das licitações.

Para as tabelas ‘anexo’ e ‘anexo_mensagem’ foi criado campo para guardar a data em que o arquivo foi baixado do portal Petronect.

Para guardar os resultados entregues pelo sistema Alice, foi necessário criar uma tabela, com o nome de ‘alerta’, contendo os campos retornados da avaliação.

5.4 Integrando dados

Cada requisição retorna um conjunto de dados específico. Com o objetivo de internalizar esses dados, foi necessário agregar as fontes de dados distintas em um banco de dados relacional. Assim, separou-se os dados de acordo com objeto a que se referem e foram criadas tabelas específicas para cada objeto. Quando necessário, foram criados campos do tipo chave primária e chave estrangeira nas tabelas para guardar as relações entre os objetos. O modelo entidade-relacionamento (MER) está descrito no Anexo II - Modelo Entidade-Relacionamento.

O banco de dados utilizado foi o Microsoft SQL Server que já está disponível no LabContas, ambiente do TCU em que se pretende manter os dados coletados ao final do projeto.

6 MODELAGEM

Para analisar os editais extraídos pela ferramenta será utilizado o Sistema de Análise de Licitações e Editais - Alice.

O Alice é uma ferramenta que foi desenvolvida pelo Ministério da Transparência, Fiscalização Controle/Controladoria-Geral da União em 2015 e foi cedido para o TCU. O Tribunal, desde então, trabalhou em adequações e aprimoramentos da ferramenta para adaptá-la aos objetivos pretendidos pela Corte de Contas.

O sistema busca padrões textuais nos documentos disponibilizados que possam sinalizar indícios de irregularidades ou riscos no processo licitatório. Esses padrões são chamados de ‘Tipologias’ ou ‘Trilhas de auditoria’. As tipologias buscam principalmente por informações que possam indicar restrições de competitividade nas licitações. Nesse modelo, é realizada a sumarização por extração onde o trecho identificado é extraído do documento e apresentado para o usuário. A descrição e sumarização de dados ajuda a identificar quais documentos deverão ser verificados em sua íntegra com base em trechos extraídos identificados pelas tipologias.

Além disso, os trechos extraídos podem identificar similaridades entre conteúdos de diversos documentos distintos, revelando a uma relação inesperada entre os documentos e entre as licitações. Outro tipo de análise realizada pelo sistema é o de cruzamento de dados. Nessas análises o sistema busca nas atas de realização dos pregões informações referentes aos licitantes em outras bases de dados.

Por fim, modelos específicos baseados em aprendizado de máquina são capazes de extrair o valor estimado da licitação a partir dos editais e anexos publicados. O valor estimado é utilizado como forma de priorização baseada em materialidade.

Atualmente, o sistema busca no Comprasnet, o Portal de Compras do Governo Federal, os editais e atas de pregão publicados e os analisa, gerando e-mails de alertas para as unidades técnicas do TCU.

O atual projeto busca agregar os editais recuperados do sistema Petronect à base de editais que o sistema Alice analisa.

A partir da análise dos editais do Petronect com as tipologias pré-existentes no sistema Alice, poderá ser verificado se as tipologias atuais já seriam suficientes para indicar possível

irregularidades nos editais de compras do Grupo Petrobras. Caso as tipologias atuais não sejam suficientes para atingir os objetivos pretendidos, novas tipologias poderão ser desenvolvidas especificamente para atender as particularidades que eventualmente os editais possam ter.

Para disponibilizar os editais extraídos pelo software desenvolvido no presente trabalho será utilizada a mesma metodologia utilizada pelos Tribunais de Contas Estaduais no projeto Alice Nacional.

A metodologia consiste em disponibilizar os arquivos adquiridos em um diretório do disco virtual do TCU em formato pré-definido para que o sistema possa efetuar as devidas análises.

Após a análise, o sistema dispara e-mail com o extrato dos resultados. Os documentos analisados em que foram identificados trechos indicativos de possíveis restrições são apontados no corpo do e-mail (Figura 29). Os detalhes de cada resultado podem ser acessados posteriormente por meio de sistema eletrônico.

Figura 29 - E-mail enviado pelo sistema Alice

15 de jan de 2020

Informe de Licitações - PETRONECT



Mostrando 312 licitações de seu interesse, de 312 carregadas em 10 de jan de 2020.

Unidade Responsável	Edital	Estimativa da Unidade*	Alertas
Petróleo Brasileiro S. A. Código: 1000	7002817082/2019 - 20/11/2019 Objeto: Serv mov cargas descarga empilhadeira	R\$ 1.200,00	⚠
Petróleo Brasileiro S. A. Código: 1000	7002949702/2020 - 13/01/2020 Objeto: Lic SW, man, sup, cab estr Data Center	R\$ 1.200,00	⚠
Petróleo Brasileiro S. A. Código: 1000	7002943384/2020 - 03/01/2020 Objeto: 03 válvulas esfera metálicas, 14 pol.	Não Identificado	⚠
TBG- Transp.Bras.Gasod.Bolivia- Brasil S.A Código: 7000	7002949794/2020 - 09/01/2020 Objeto: AQUISIÇÃO TURBO ALIMENTADOR WAUKESHA	Não Identificado	⚠
TBG- Transp.Bras.Gasod.Bolivia- Brasil S.A Código: 7000	7002949522/2020 - 09/01/2020 Objeto: SOBRESSALENTES PARA MOTORES WAUKESHA	Não Identificado	⚠
Petróleo Brasileiro S. A. Código: 1000	7002896572/2019 - 02/12/2019 Objeto: Serv op logística porto do RJ para BANIT	R\$ 69.767.285,47	-
Petróleo Brasileiro S. A. Código: 1000	7002951906/2020 - 13/01/2020 Objeto: Floculante	R\$ 33.817.581,78	-
Petróleo Brasileiro S. A. Código: 1000	7002928108/2020 - 10/01/2020 Objeto: Ampliação Capac Geracao Energia de Emerg	R\$ 23.019.120,60	-

Fonte: própria

7 AVALIAÇÃO

O software desenvolvido conseguiu realizar a extração dos dados disponibilizados nas requisições ao portal Petronect. Além disso, foram internalizadas, no LabContas, informações de 1472 disputas ao longo do desenvolvimento do projeto.

Foram baixados e entregues automaticamente para o Sistema Alice um total de 6834 arquivos.

Os dados do portal Petronect são incrementados e atualizados diariamente. O extrator é responsável por baixar os novos dados e manter os dados de licitações baixadas anteriormente atualizadas. Logo, os dados avaliados nesse projeto representam uma fotografia dos dados do Petronect tirada entre os dias 01/11/2019 e 17/01/2020.

Com relação às análises realizadas pelo sistema Alice, das 1472 licitações analisadas 26 (1,8%) apresentaram possíveis irregularidades apontadas pelas tipologias já existentes no sistema. Este percentual está bem abaixo do observado nas licitações do Comprasnet, onde verifica-se uma taxa acima de 5% de licitações em risco. Em relação à extração do valor estimado, 573 licitações (39%) tiveram seu valor estimado pelo sistema. Este percentual também está bem abaixo do observado nas licitações do Comprasnet, onde 75% das licitações possuem seu valor estimado extraído.

O valor baixo de detecção pode ser reflexo das diferenças existentes entre editais cadastrados no Comprasnet e dos editais cadastrados no Petronect. A natureza e itens de contratação diferenciadas entre os dois sistemas também podem contribuir para esse baixo valor, haja vista que as tipologias atuais do sistema Alice foram desenvolvidas com base em amostras de editais extraídos do Comprasnet.

Os valores das licitações estimadas pelo Alice somaram aproximadamente R\$ 467 milhões. Contudo, da análise da frequência dos valores estimados utilizando-se o histograma dos valores em base logarítmica (Figura 30), identificou-se que o valor R\$ 1.200 foi estimado como o valor da licitação em 273 licitações diferentes.

Dada a elevada frequência desse valor, foi realizada uma análise manual em pequena amostra buscando verificar a veracidade desse valor.

Como resultado, identificou-se que o valor R\$ 1.200,00 está relacionado, em todas as licitações da amostra, como o valor de multas contratuais enumeradas em arquivos integrantes ao edital, conforme excerto destacado:

8.1.7 - Pela não informação à PETROBRAS da ocorrência de acidente de trabalho nas dependências da PETROBRAS, com seus empregados ou

terceiros a seu serviço, vinculados à execução deste Contrato: R\$ 1.200,00 (mil e duzentos reais), por cada ocorrência.

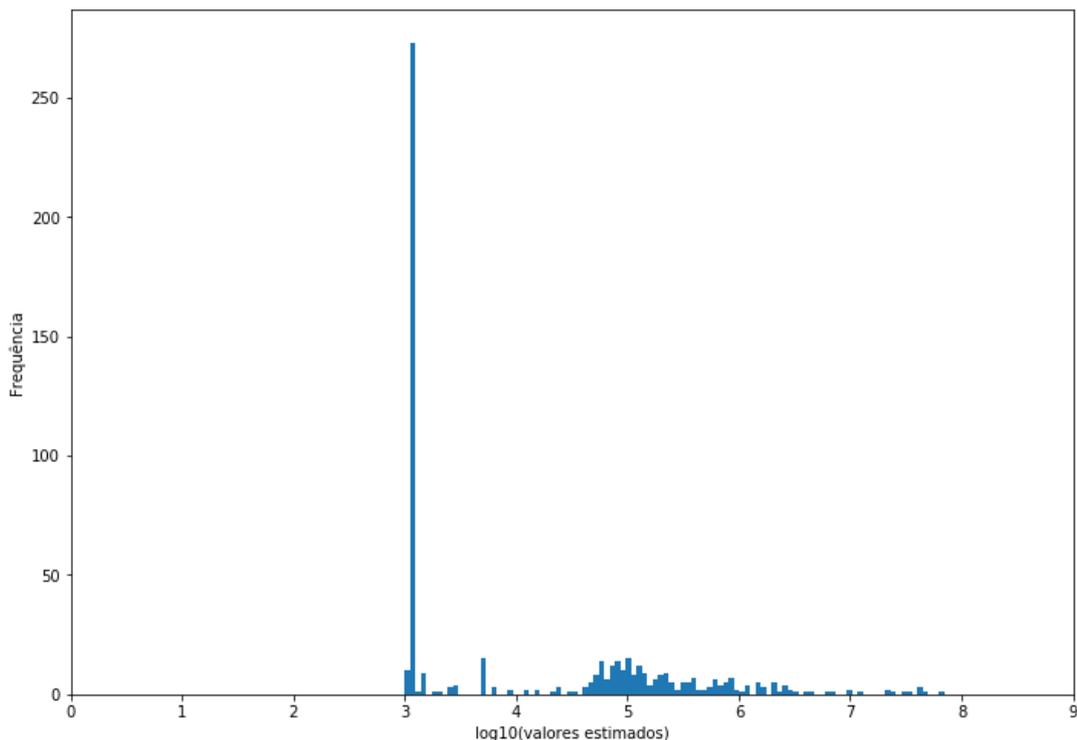
8.1.8 - Pela não emissão da Comunicação de Acidente de Trabalho - CAT, nas condições e nos prazos legais, em caso de acidente de trabalho ocorrido nas dependências da PETROBRAS, com seus empregados vinculados à execução deste Contrato: R\$ 1.200,00 (mil e duzentos reais), por cada ocorrência.

Dentro da mesma gama amostral, não foi possível identificar os valores estimados das licitações, o que pode indicar que não há a divulgação de tais valores nos editais em questão.

Foi realizado também outras análises em editais cujo valor levantado pelo Alice foi superior a R\$ 1.200,00. Dessa análise, não foi possível confirmar se o valor estimado pelo sistema é realmente o valor do edital.

Assim, o levantamento dos valores realizados pelo sistema Alice se mostrou prejudicado.

Figura 30 - Frequência dos valores estimados - Alice



Fonte: própria

Da análise das informações coletadas, utilizando-se o painel desenvolvido, foi possível trazer a conhecimento da SeinfraPetróleo que a maioria das licitações cadastradas no portal são da empresa Petróleos Brasileiro S. A.

Os dados obtidos mostraram os itens das licitações mais frequentes são os classificados como ‘materiais’ (58%).

A modalidade ‘Pregão’ não foi a modalidade mais utilizada apesar do artigo 32 da Lei 13.303/2016 (Brasil, 2016) estabelecer em seu inciso IV que a modalidade denominada pregão deverá ser adotada preferencialmente para a aquisição de bens e serviços comuns.

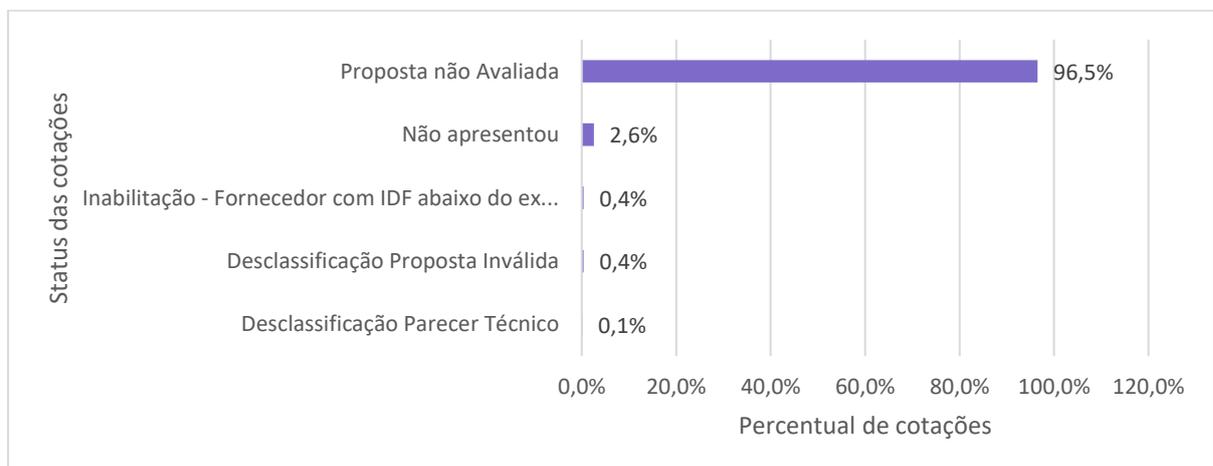
O modo de disputa mais utilizado foi o modo de disputa fechado, onde as propostas apresentadas pelos participantes são sigilosas até a data e hora designadas para sua divulgação.

Verificou-se que as propostas cadastradas no portal Petronect estão, em sua maioria, com o estado ‘Proposta não Avaliada’.

O elevado número de cotações em tal estado pode ser um reflexo do número de licitações que ainda estão em andamento. Contudo, mesmo após filtrar os dados para as licitações que já foram concluídas, o percentual de propostas neste estado manteve-se alta (85%).

Interessante apontar que, entre os dados coletados, a empresa Diametral Representações, Comercio e Serviços LTDA que participou da maior quantidade de cotações, 277 participações, não teve, durante o intervalo de tempo analisado, nenhuma cotação habilitada ou até mesmo classificada conforme mostrado na Figura 31.

Figura 31 - Status das cotações da empresa Diametral



Fonte: própria

Observou-se que menos da metade das cotações foram apresentadas por microempresas ou empresas de pequeno porte.

Outro ponto de interesse levantado para a SeinfraPetróleo foi a utilização recorrente de nomenclatura genérica para itens de licitação, diminuindo a transparência da disputa.

Itens nomeados como ‘MANUTENÇÃO EM MÁQUINAS/EQUIPAM. - S/MAT’, ‘VALOR TOTAL DA PROPOSTA’ e ‘VALOR TOTAL DA PPU’ foram identificados em

diversas disputas e não são suficientemente claros para se caracterizar o item desejado a ser comprado.

A lei 13.303/2016 (Brasil, 2016) determina em seu artigo 33 que o item da licitação e do contrato decorrente da licitação será definido de forma sucinta e clara no instrumento convocatório. Assim, entende-se que a publicidade dos itens da licitação no sistema com nomenclatura genérica pode ferir a disposição legal.

8 CONCLUSÃO

O objetivo do negócio que se buscou alcançar no presente trabalho foi o de melhorar a atuação da SeinfraPetróleo por meio da disponibilização de informações acerca do processo licitatório de contratações e aquisições realizadas pelas empresas do Grupo Petrobras conduzidas pelo portal eletrônico Petronect.

Como critério de sucesso do negócio foi estabelecido o acompanhamento das contratações realizadas pelo Petronect de forma tempestiva, diminuindo o tempo e esforço para a obtenção e preparação dos dados das disputas.

Para satisfazer os critérios de sucesso, foi estabelecido como objetivo de mineração de dados o desenvolvimento de software para automatizar a extração e internalização dos dados do portal Petronect.

Ademais, as informações obtidas foram disponibilizadas em painel eletrônico para consulta e os documentos relacionados às licitações foram entregues ao sistema Alice para análise.

Inicialmente, estudou-se o portal Petronect avaliando quais as informações estavam disponibilizadas em área pública e como obtê-las de maneira automatizada. Verificou-se por meio de ferramenta de análise de tráfego que as requisições realizadas entre o navegador de internet e o portal eram respondidas em formato JSON e exibidas e formatadas por meio de scripts desenvolvidos em javascript.

Como as requisições que continham as informações que seriam utilizadas nesse projeto retornavam JSONs contendo os dados, optou-se em realizar essas requisições e tratar os dados a partir do JSON de resposta ao invés de extrair os dados diretamente do código fonte de cada página.

Isso possibilitou que não fosse necessário passar por validações de CAPTCHA uma vez que a validação CAPTCHA só era solicitada durante a navegação das páginas do portal e não como pré-requisito para a realização de requisição dos dados.

Assim, foi desenvolvido, em Python, utilizando-se o editor de código-fonte Visual Studio Code, software que realiza requisições ao portal Petronect e, após preparação dos dados, os grava no banco de dados do LabContas. O software automatiza a obtenção dos dados diminuindo tanto o tempo quanto o esforço da tarefa que exigiria diversas interações com o portal por parte do usuário.

Ademais, o software realiza o download dos arquivos que compõe cada edital de cada licitação e os disponibiliza em pasta de rede do Tribunal para o sistema Alice. Com os arquivos disponibilizados, o Alice pode então realizar sua varredura por tipologias previamente cadastradas.

O software realiza, também, varredura nos arquivos resultantes da validação do sistema Alice e grava os achados no banco de dados do LabContas.

Para visualização e análise dos dados gerados tanto pela extração das informações do portal Petronect como dos resultados obtidos pelo sistema Alice, foi desenvolvido um painel eletrônico utilizando-se a ferramenta Qlikview (Anexo I - Telas do Painel eletrônico desenvolvido).

Em breve análise das informações coletadas e exibidas no painel construído, a SeinfraPetróleo conseguiu conhecer melhor os volumes de compras realizadas, as modalidades utilizadas, os itens mais frequentes nas disputas, as principais empresas participantes e diversas outras informações que ajudarão a secretaria a melhorar sua atuação. Ademais, foi possível identificar possíveis irregularidades e pontos de questionamentos a serem verificados junto a Petrobras.

As análises realizadas pelo sistema Alice, aplicando-se as tipologias preexistentes, indicou possíveis problemas em 26 editais. Além disso, o sistema apontou o valor estimado de 573 licitações.

Os motivos de baixa aderência das tipologias aplicadas ainda deverão ser estudados, entretanto a natureza e itens de contratação diferenciadas entre os sistemas Petronect e Comprasnet também podem estar relacionados ao baixo valor de indícios levantados, pois as tipologias aplicadas pelo sistema Alice foram desenvolvidas utilizando-se editais extraídos do Comprasnet.

Com relação aos valores estimados pelo sistema, identificou-se que os números apontados não estavam diretamente relacionados com o valor real das licitações. Muitos dos valores foram retirados de minutas contratuais presentes no rol de documentos das licitações e se referiam a muitas contratuais. Assim, o valor estimado da licitação pelo sistema não se mostrou acurado.

8.1 Trabalhos Futuros

O portal Petronect possui uma área de acesso restrito a usuários logados que pode apresentar mais informações que não estão disponibilizadas na área pública do portal sobre as disputas.

Para buscar melhorar os resultados do projeto, vislumbra-se que, em etapa posterior, seja realizado um estudo acerca dos dados disponíveis no ambiente seguro para possibilitar a incorporação das informações complementares à base de dados geradas no presente projeto.

Ademais, será necessário criar uma forma de agendamento para execução do software desenvolvido de forma que o processo fique completamente automatizado e o painel desenvolvido deverá ser disponibilizado na intranet do Tribunal para que seu acesso fique disponível para os servidores da casa.

Além disso, os resultados entregues pelo sistema Alice deverão ser estudados para identificar sua aderência as tipologias preexistentes no sistema e buscar melhorias.

Novas tipologias deverão ser desenvolvidas para o sistema Alice para buscar particularidades pertinentes aos editais de contratação do Grupo Petrobras.

REFERÊNCIAS

Brasil. 1988. Constituição Federal de 1988. *Constituição da República Federativa do Brasil*. 1988.

—. **2016.** Lei nº 13.303 de 30 de junho de 2016. *Lei das Estatais*. 30 de jun de 2016.

—. **1992.** Lei nº 8.443 de 16 de julho de 1992. *Lei Orgânica do Tribunal de Contas da União*. 16 de jul de 1992.

—. **1997.** Lei nº 9.478 de 6 de agosto de 1997. *Política energética nacional*. 6 de ago de 1997.

Fornecedor, Canal. 2018. <https://canalfornecedor.com.br/as-utilidades-do-petronect-petrobras-2/AS-4-UTILIDADES-DO-PETRONECT,O-SISTEMA-DE-COMPRAS-PETROBRAS>. [Online] Canal Fornecedor, 04 de setembro de 2018. [Citado em: 02 de 03 de 2020.]

IBM. 2020. CRISP-DM Help Overview. *IBM Knowledge Center*. [Online] IBM, 05 de fevereiro de 2020. https://www.ibm.com/support/knowledgecenter/SS3RA7_15.0.0/com.ibm.spss.crispdm.help/crisp_overview.htm.

Petrobras. Portal da Transparência. *Licitações e contratos*. [Online] [Citado em: 11 de 02 de 2020.] <http://transparencia.petrobras.com.br/licitacoes-contratos/licitacoes>.

—. **2018.** Regras de contratação. [Online] 26 de 06 de 2018. [Citado em: 13 de 02 de 2020.] https://canalfornecedor.petrobras.com.br/media/filer_public/fc/62/fc62e81b-7a2a-44da-9dba-bd5463d2d7db/cartilha-rlcp_rev01.pdf.

Petronect. 2019. Apresentação Institucional. *Petronect*. [Online] 2019. [Citado em: 1º de Setembro de 2019.] <https://www.petronect.com.br/irj/go/km/docs/pccshrcontent/TreinamentosForn/apresentacao-institucional-pt-07-2019.pdf>.

—. **2002.** Institucional. *Site da Petronect*. [Online] Petronect, 2002. [Citado em: 15 de 09 de 2019.] [https://www.petronect.com.br/irj/go/km/docs/pccshrcontent/Site%20Content%20\(Legacy\)/Portal2018/pt/APetronect_Institucional.html](https://www.petronect.com.br/irj/go/km/docs/pccshrcontent/Site%20Content%20(Legacy)/Portal2018/pt/APetronect_Institucional.html).

—. **2020.** Pagina Inicial. *Petronect*. [Online] 2020. [Citado em: 05 de fevereiro de 2020.] <https://www.petronect.com.br/irj/portal/anonymous/pt>.

SPSS. 2000. CRISP-DM 1.0. *The Modeling Agency*. [Online] 2000. [Citado em: 05 de 03 de 2020.] <https://www.the-modeling-agency.com/crisp-dm.pdf>.

União, Tribunal de Contas da. 2017. Portaria-Segecex nº 5/2017. *Dispõe sobre as competências, a estrutura e a alocação de funções de confiança da Secretaria-Geral de Controle Externo*. 7 de mar de 2017.

Wirth, Rüdiger e Hipp, Jochen. CRISP-DM: Towards a Standard Process Model for Data. [Online] [Citado em: 05 de 03 de 2020.] <http://www.cs.unibo.it/~danilo.montesi/CBD/Beatriz/10.1.1.198.5133.pdf>.

GLOSSÁRIO

CSS - Cascading Style Sheets - Folha de Estilo em Cascatas

Dashboards - Painéis eletrônicos de visualização de dados

HTML - *Hypertext Markup Language* - Linguagem de Marcação de HiperTexto

HTTP - *Hyper Text Transfer Protocol Secure* - protocolo de transferência de hipertexto

HTTPS - *Hyper Text Transfer Protocol Secure* - protocolo de transferência de hipertexto seguro

Javascript - Linguagem de programação que permite criar conteúdos dinâmicos em páginas HTML

JSON - *JavaScript Object Notation* - Notação de Objetos JavaScript

Python - Linguagem de programação interpretada livre, de alto nível e amplamente utilizada no contexto de ciências de dados

Sistema Alice - Sistema de Análise de Licitações e Editais que possibilita a avaliação tempestiva e automatizada de editais de licitação e atas de pregão, com a identificação de indícios de irregularidades, fraudes, desvios e desperdícios de recursos públicos, possibilitando ações de controle mais eficientes e efetivas.

Raspagem de dados - ato de baixar automaticamente os dados de uma página web e extrair informações específicas dela.

ANEXOS

Anexo I - Telas do Painel eletrônico desenvolvido

Figura 32 - Painel - Tela inicial



Figura 33 - Painel - Tela de disputas

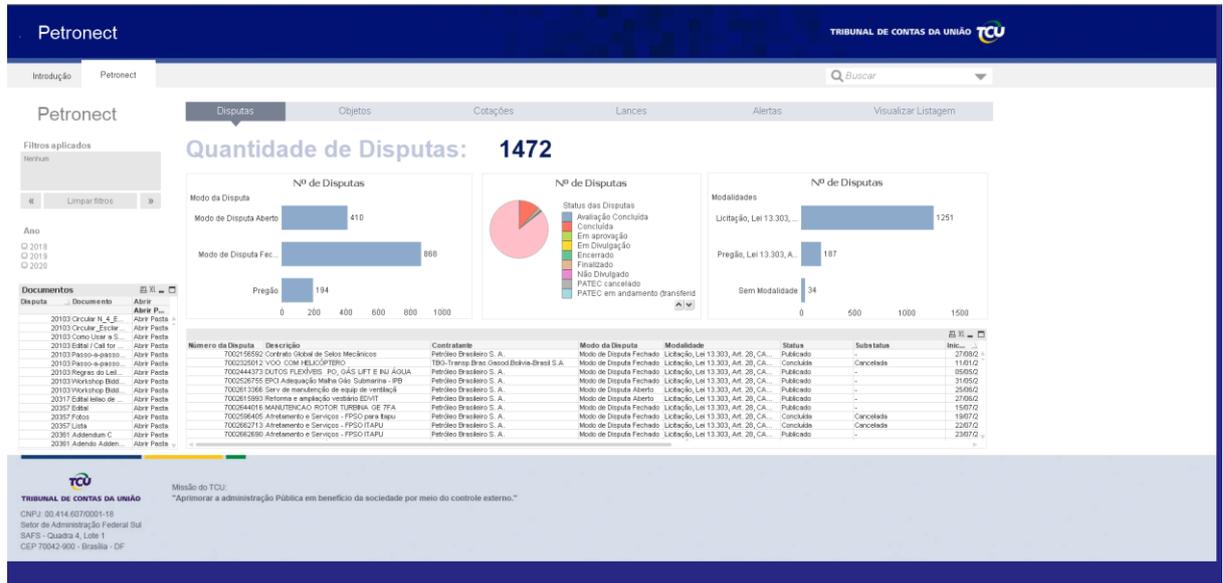


Figura 34 - Painel - Tela de objetos

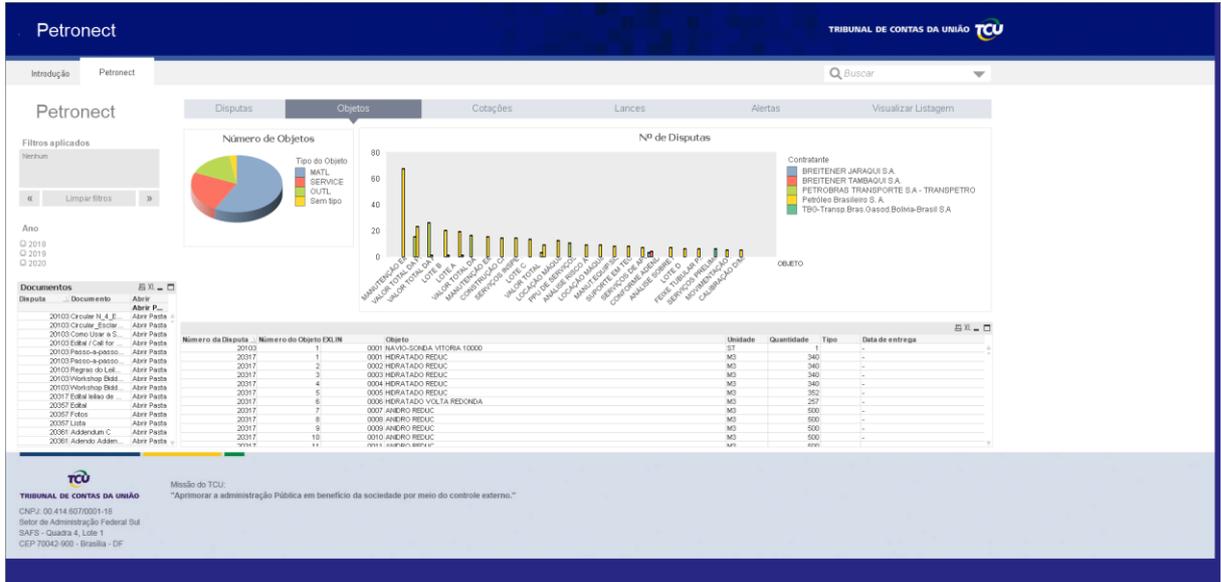


Figura 35 - Painel - Tela de cotações

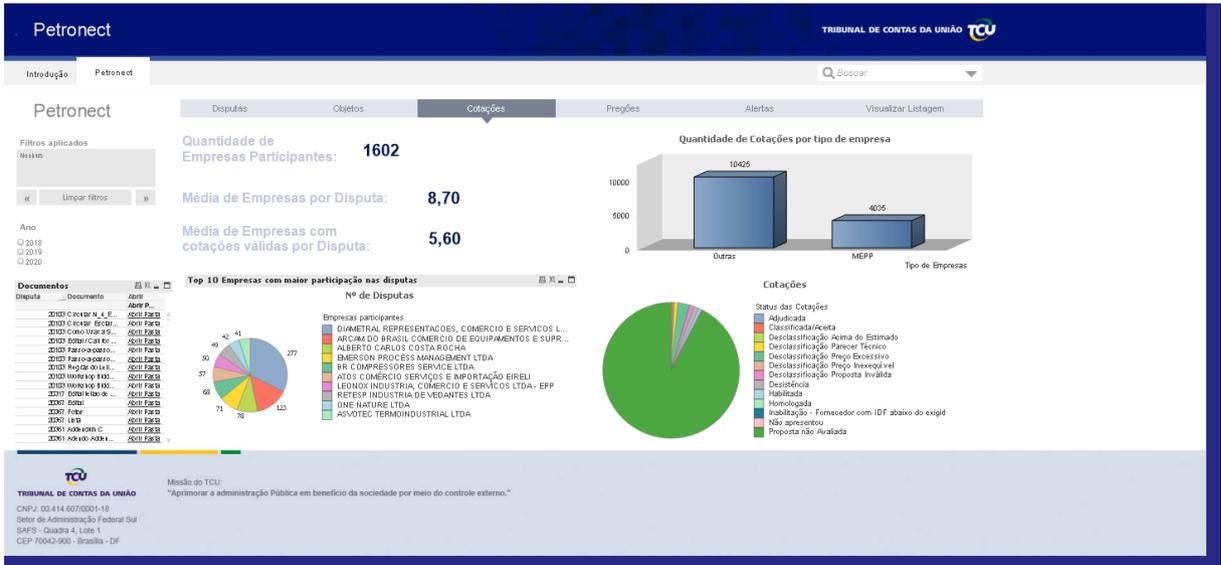


Figura 36 - Painel - Tela de pregões

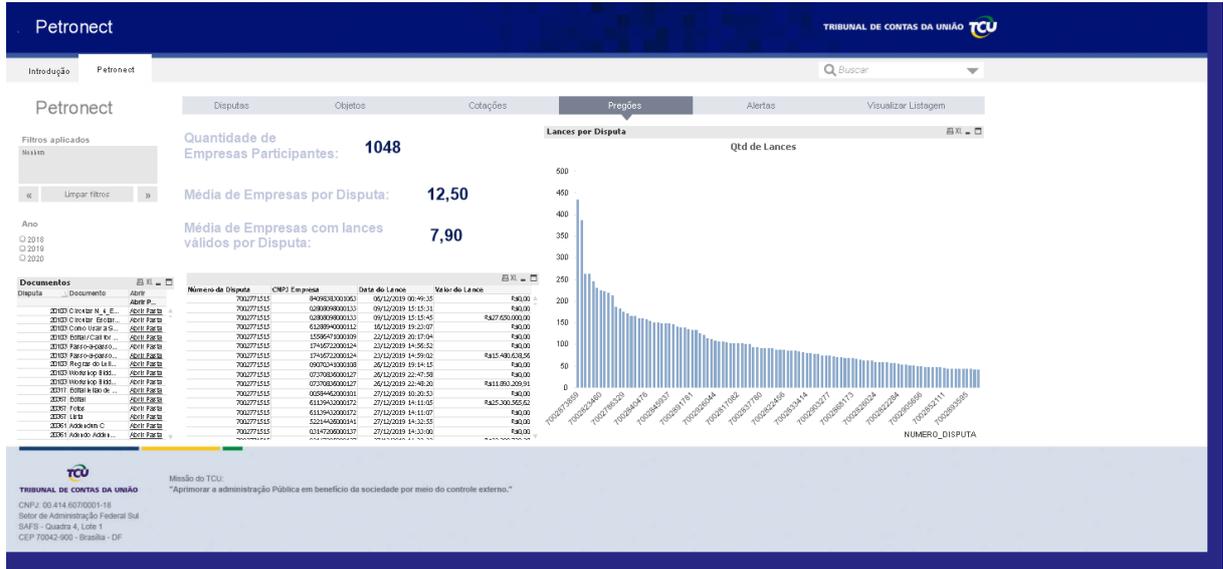
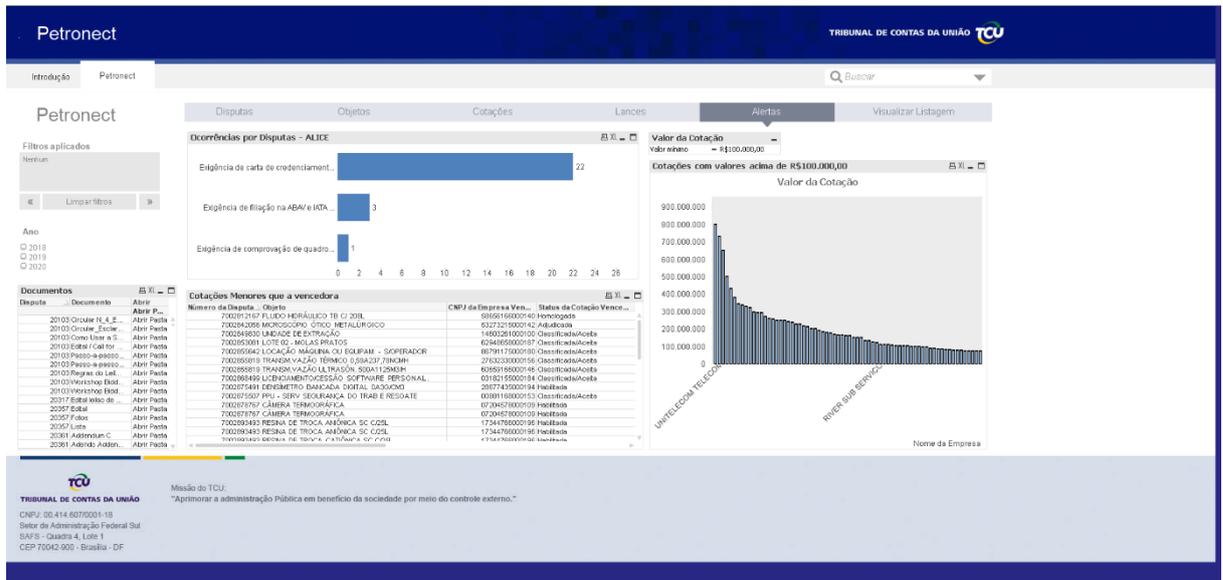


Figura 37 - Painel - Tela de alertas



Anexo II - Modelo Entidade-Relacionamento

Figura 38 - Modelo Entidade Relacionamento

