

Por ordem do



Ministério Federal
do Meio Ambiente, Proteção da Natureza,
Construção e Segurança Nuclear

da República Federal da Alemanha



MINISTÉRIO DO
MEIO AMBIENTE



Relatório Técnico

Projeto Política sobre Mudanças do Clima (PoMuC)

Identificação dos Gastos Orçamentários Públicos Brasileiros com a Política Nacional sobre Mudança do Clima

Produto Elaborado para: Ministério do Meio Ambiente &
Cooperação Alemã para o Desenvolvimento /
Deutsche Gesellschaft für Internationale Zusammenarbeit
(GIZ) GmbH

Elaborado por: International Policy Centre for Inclusive Growth (IPC-IG) &
Instituto de Pesquisa Econômica Aplicada (Ipea)

Autor: Bruno Miranda. Pesquisador Visitante do IPC-IG/PNUD na DIRUR-IPEA.

Pesquisa realizada pelo Instituto de Pesquisa Econômica Aplicada - IPEA (Diretoria de Estudos Regionais, Urbanos e Ambientais - DIRUR)¹ e pelo Centro Internacional de Políticas para o Crescimento Inclusivo (IPC-IG) do Programa das Nações Unidas para o Desenvolvimento (PNUD), coordenada pelo Ministério do Meio Ambiente (Departamento de Economia Ambiental e Acordos Internacionais, Secretaria de Relações Internacionais) no contexto da Iniciativa Internacional sobre Mudança do Clima (IKI) do Ministério Federal do Ambiente, Proteção da Natureza e Segurança Nuclear da Alemanha (BMU) e conta com o apoio técnico da Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, agência executora da cooperação técnica alemã.

Brasília, junho de 2020

¹ Acordo de Cooperação Técnica nº 29/2018-MMA entre o Ministério do Meio Ambiente (MMA) e o Instituto de Pesquisa Econômica Aplicada (IPEA).

Sumário

1. Análise exploratória dos dados	3
1.1. Os dados do SIOP.....	3
1.2. O processo da carga de dados.....	5
1.3. A necessidade de pré e pós processamentos.....	7
2. Seleção do escopo de análise.....	7
3. O projeto da Base de Dados	14
3.1. A estrutura de abordagem por fases	14
3.2. Fase 1 - o processo de carga de dados	15
3.3. Fase 2 – tratamento dos dados brutos	16
3.4. Fase 3 – Criação dos datasets (visões) e análise dos dados	17
4. Análise de informações processadas	17
4.1. Evolução dos gastos na função Gestão Ambiental (Identificador 18 no SIOP).....	18
4.2. Gastos em ações orçamentárias core, uma análise em busca textual	19
5. Tecnologias utilizadas	20
5.1. Definição das escolhas.....	20
SIGLAS	22
REFERÊNCIAS.....	22
ANEXOS	22

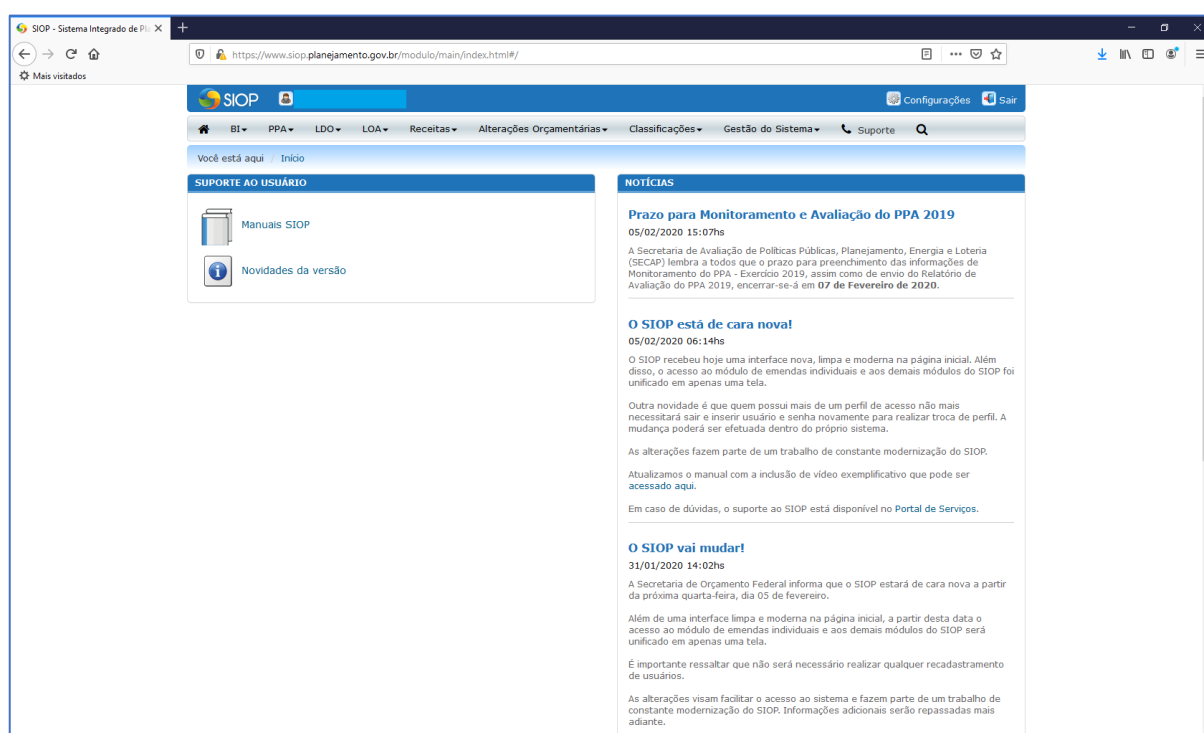
1. Análise exploratória dos dados

A forma como o projeto POMUC-Gastos obtém os dados relevantes para seu projeto empírico, é por meio do monitoramento dos gastos com as Ações Orçamentárias executadas via sistema SIOP. Trata-se de um dos principais sistemas estruturantes do Governo Federal, o qual tem por principal objetivo normalizar e fornecer o suporte aos processos de Planejamento e Orçamento.

1.1. Os dados do SIOP

Os dados do SIOP são disponibilizados usuário autenticado, onde exige-se cadastro prévio e avaliação por parte da equipe técnica mantenedora do sistema.

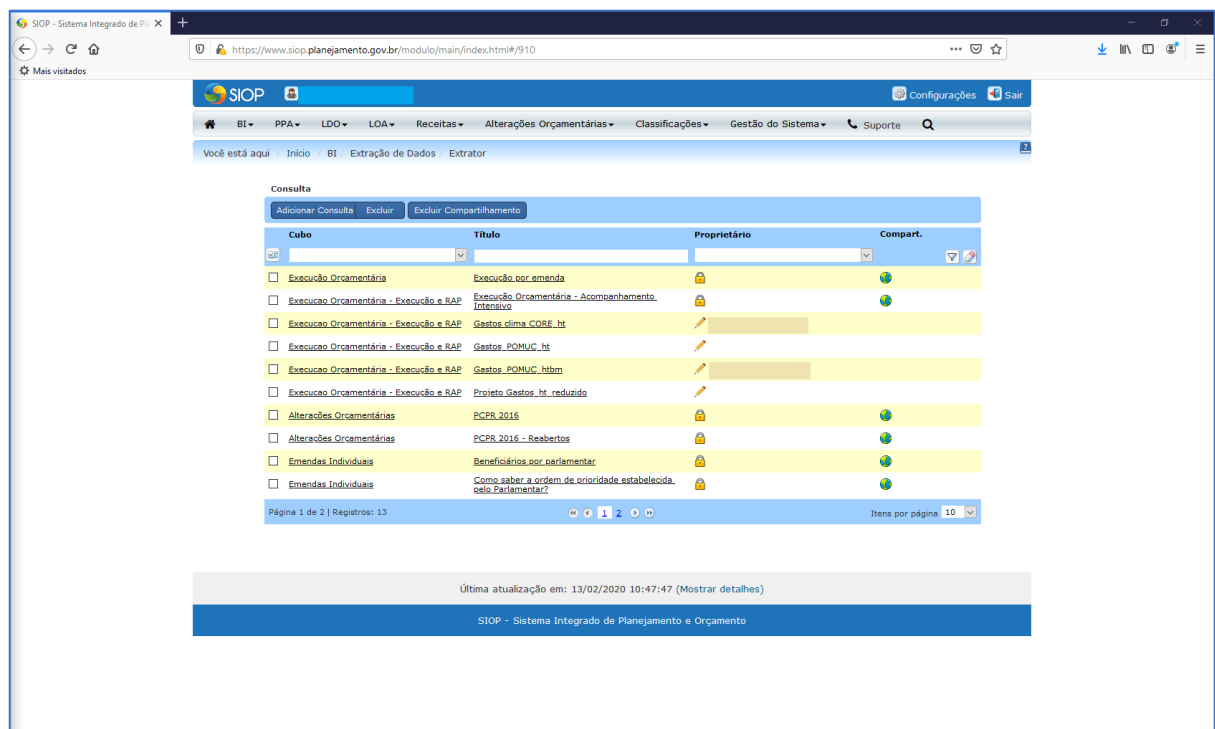
Figura 1 - Sistema SIOP



Fonte: Sistema SIOP

O sistema SIOP permite a criação de cubos de análise, onde se pode configurar consultas com as variáveis desejadas. Para o projeto PoMuC-Gastos, consultas específicas foram construídas.

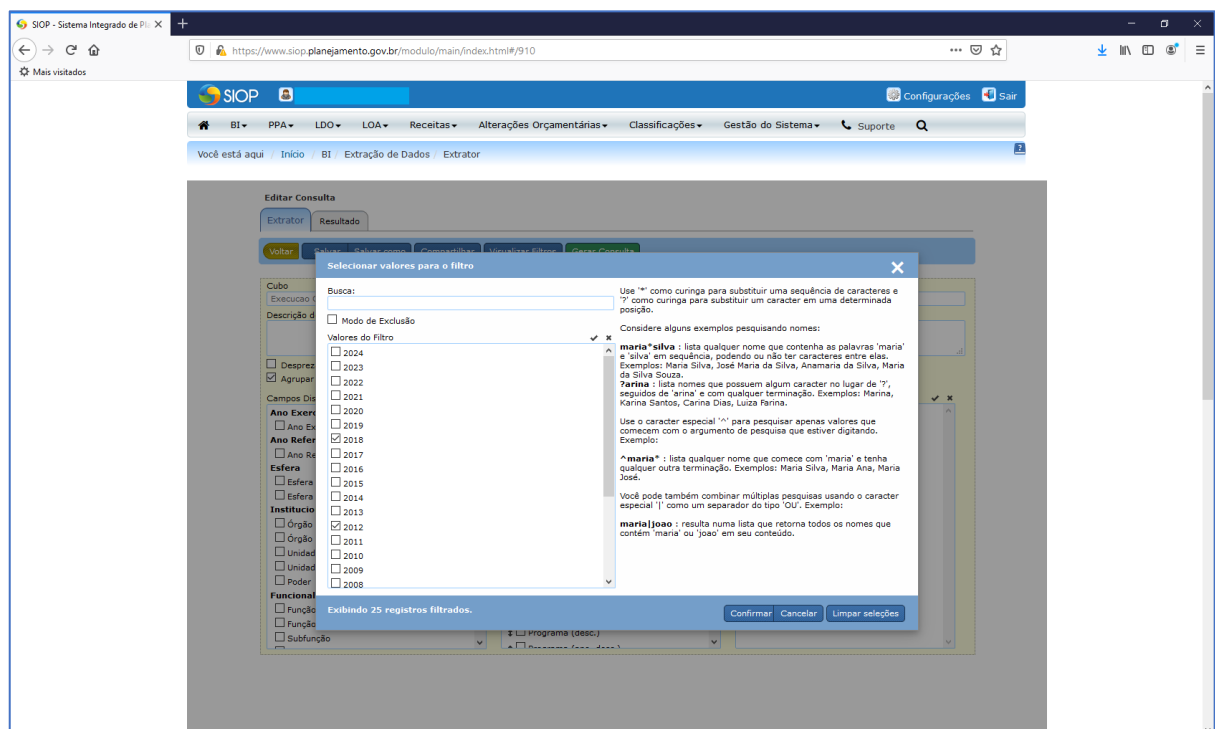
Figura 2 - Seleção dos cubos de dados



Fonte: Sistema SIOPI

A priori as consultas são referentes ao período de 2000-2019, sendo apenas o período de 2009-2019 alvo de análises do projeto.

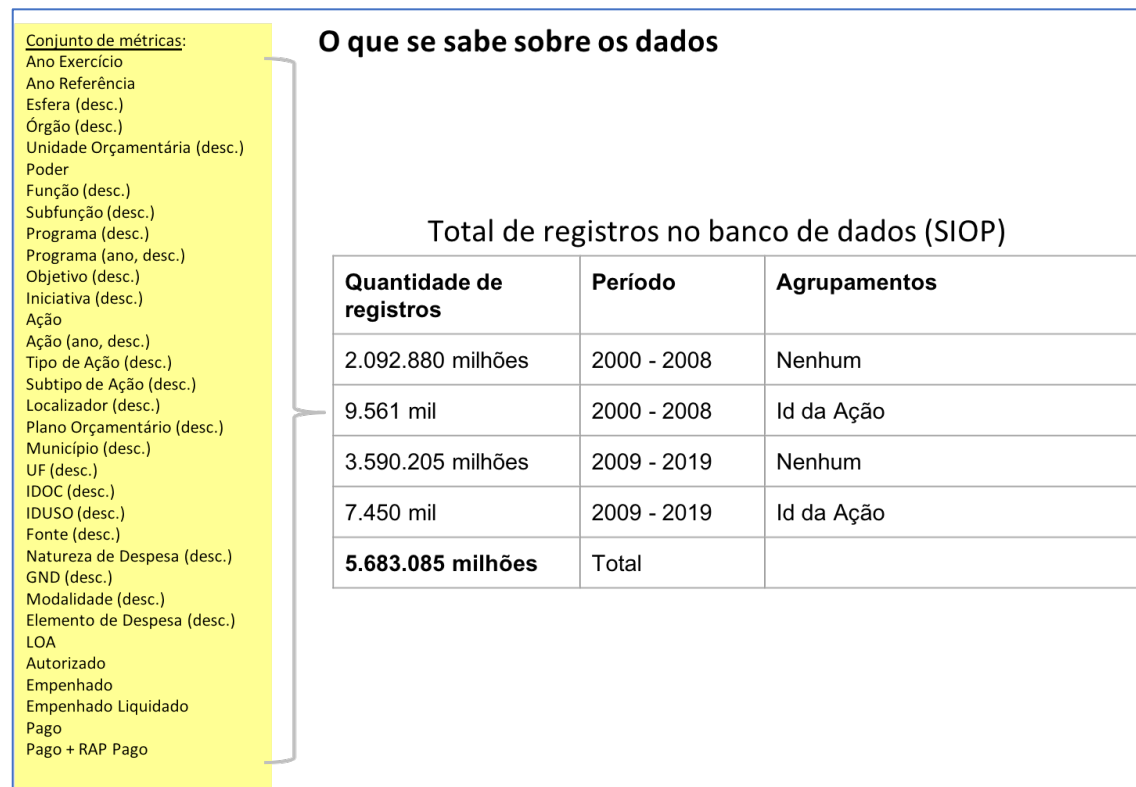
Figura 3 - Seleção do universo temporal da consulta



Fonte: Sistema SIOPI

Quarenta e três variáveis foram selecionadas para a construção das consultas, onde o tamanho do conjunto de dados ficou em aproximadamente cinco milhões e seiscentos mil registros.

Figura 4 - Etapa de entendimento das variáveis



Elaborado por: Pesquisadores do projeto PoMuC-Gastos

1.2. O processo da carga de dados

Para a referida etapa, foi desenvolvido um aplicativo codificado em linguagem de programação Python 3.7, onde, sua construção permitiu descobrir e mitigar diversas inconsistências nos dados oriundos exportados do sistema SIOP, dentre elas, a não unicidade dos registros no que diz respeito às Ações Orçamentárias, matéria-prima fundamental para análise e estudos dos gastos, objeto principal do PoMuC-Gastos.

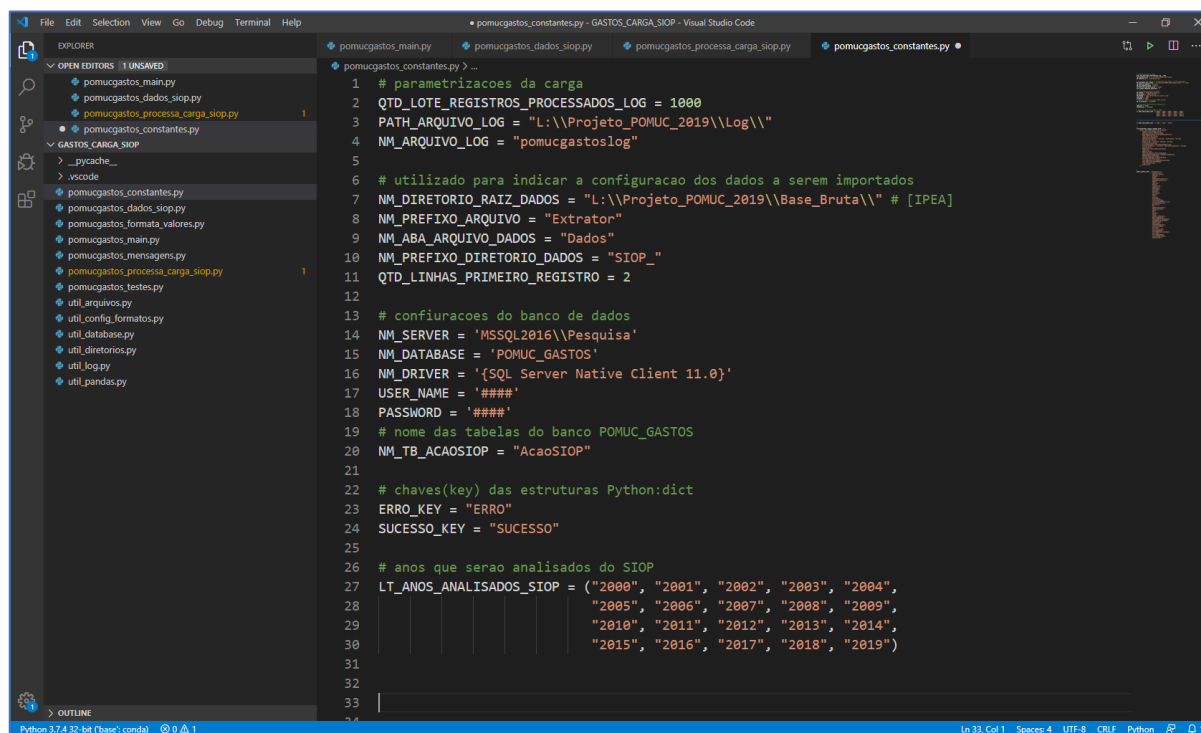
O aplicativo permite a parametrização de diversas variáveis para o processo de carga, sendo estes: log de operações em console e arquivo, local remoto dos arquivos de dados brutos, dados do banco de dados de destino e anos selecionados para a carga.

Essa estrutura permite que a carga de dados seja realizada de forma completa, parcial ou incremental, eliminando a necessidade de completa deleção e carga posterior, caso algum erro venha a ocorrer.

E ainda, o sistema de log permite ao operador da aplicação descobrir os erros em tempo real de carga, sabendo em qual diretório, arquivo e linha o aplicativo processou. Facilitando o monitoramento, rastreamento e correção de registros com defeito.

Na imagem abaixo é possível observar a tela de parametrização das variáveis indicadas:

Figura 5 - Parametrizações do aplicativo de carga de dados



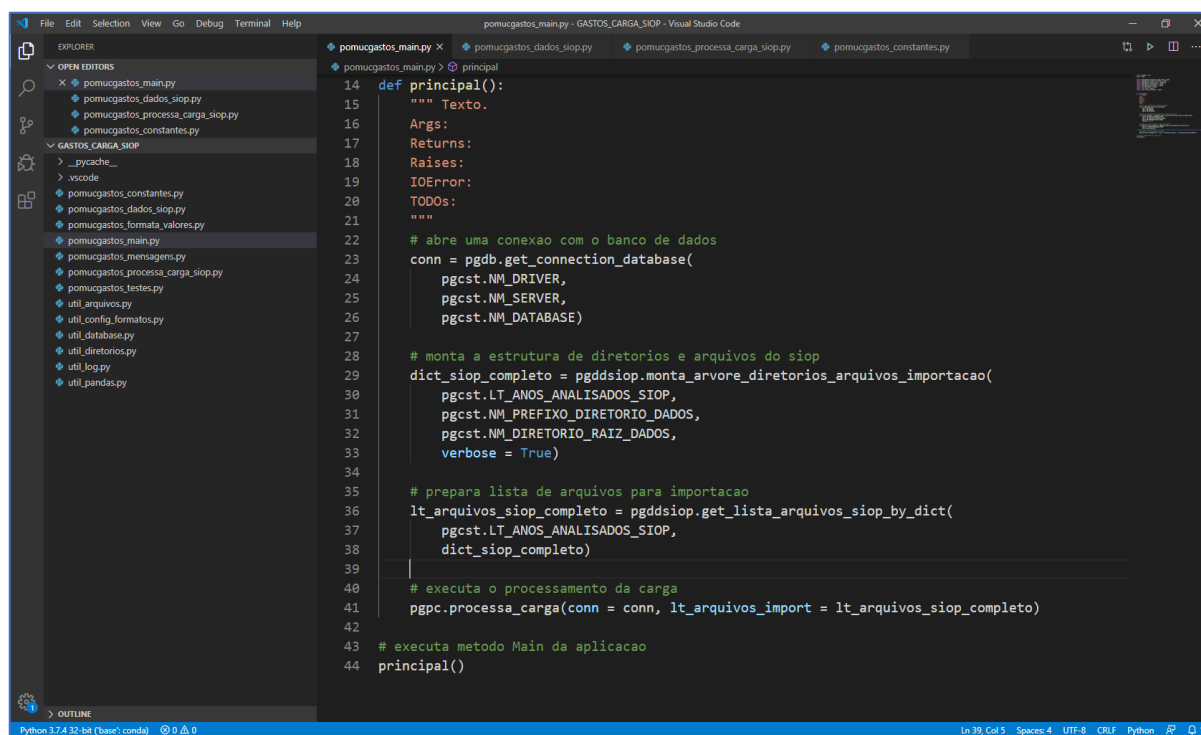
```
1 # parametrizacoes da carga
2 QTD_LOTE_REGISTROS_PROCESSADOS_LOG = 1000
3 PATH_ARQUIVO_LOG = "L:\\Projeto_POMUC_2019\\Log\\"
4 NM_ARQUIVO_LOG = "pomucgastoslog"
5
6 # utilizado para indicar a configuracao dos dados a serem importados
7 NM_DIRETORIO_RAIZ_DADOS = "L:\\Projeto_POMUC_2019\\Base_Bruta\\" # [IPEA]
8 NM_PREFIXO_ARQUIVO = "Extrator"
9 NM_ABA_ARQUIVO_DADOS = "Dados"
10 NM_PREFIXO_DIRETORIO_DADOS = "SIOP_"
11 QTD_LINHAS_PRIMEIRO_REGISTRO = 2
12
13 # configuracoes do banco de dados
14 NM_SERVER = 'MSSQL2016\\Pesquisa'
15 NM_DATABASE = 'POMUC_GASTOS'
16 NM_DRIVER = '{SQL Server Native Client 11.0}'
17 USER_NAME = '####'
18 PASSWORD = '####'
19 # nome das tabelas do banco POMUC_GASTOS
20 NM_TB_ACAOSIOP = "AcaoSIOP"
21
22 # chaves(key) das estruturas Python:dict
23 ERRO_KEY = "ERRO"
24 SUCESSO_KEY = "SUCESSO"
25
26 # anos que serao analisados do SIOP
27 LT_ANOS_ANALISADOS_SIOP = ("2000", "2001", "2002", "2003", "2004",
28                             "2005", "2006", "2007", "2008", "2009",
29                             "2010", "2011", "2012", "2013", "2014",
30                             "2015", "2016", "2017", "2018", "2019")
31
32
33
```

Elaborado por: Pesquisadores do projeto PoMuC-Gastos

O aplicativo de carga de dados foi desenvolvido para ser executado de maneira *standalone* por meio de tecnologia de processamento paralelo, tal qual Apache Spark ou similares. Porém, também pode ser executado direto pelo console de uma IDE de desenvolvimento Python (3.7 ou superior) ou em servidores que possuam a linguagem disponível previamente instalada e configurada. O tempo de processamento depende necessariamente de três fatores: a) quantidade de anos selecionados para a carga; b) tipo de execução; e c) infraestrutura de banco de dados de destino.

Para se dirimir dificuldades quanto à configuração de tecnologias não disponíveis no momento da execução do projeto, para o PoMuC-Gastos o aplicativo foi executado direto na console da ferramenta de desenvolvimento Microsoft Visual Code com destino em um servidor de banco de dados Microsoft SQL Server 2016, para os anos de 2000-2019 anos de dados oriundos do SIOP. O tempo de execução ficou em média oito horas ininterruptas de processamento.

Figura 6 - Classe main do aplicativo de carga de dados desenvolvido em Python 3.7



Elaborado por: Pesquisadores do projeto PoMuC-Gastos

1.3. A necessidade de pré e pós processamentos

Do total de variáveis selecionadas (43) inicialmente, foram estendidas para cinquenta e uma, pois devido ao acoplamento da base de dados disponibilizada pelo sistema SIOP, alguns atributos de Id e descrição (por exemplo) possuem uma estrutura agregada.

Esse desdobramento dos campos, bem como outras operações de formatação, possibilita a utilização de filtros (consultas/subconsultas) mais precisos no futuro, na etapa de análise dos dados.

2. Seleção do escopo de análise

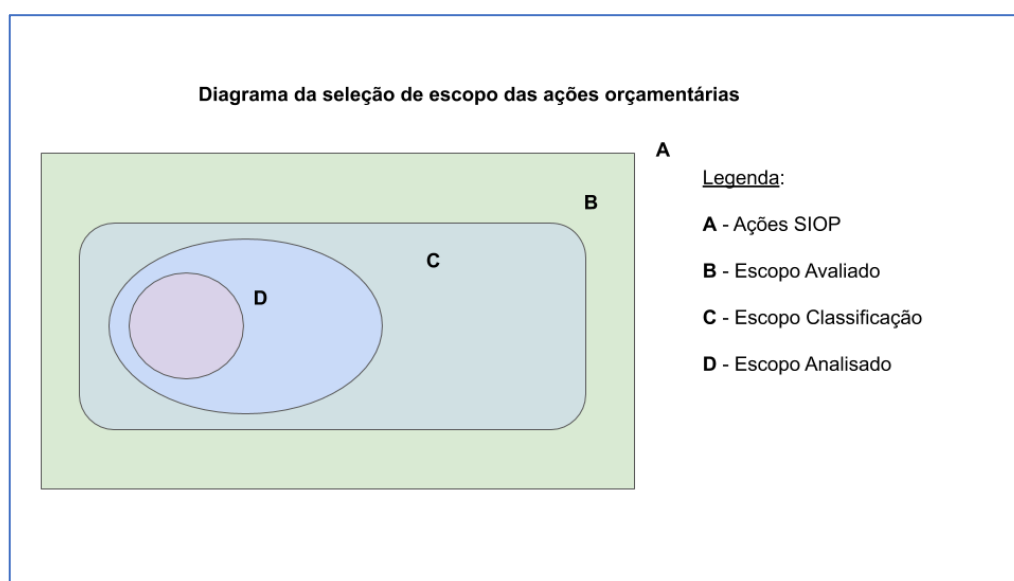
De acordo com consultas realizadas ao banco de dados do projeto PoMuC-Gastos, o SIOP possui o total de treze mil e cento e vinte e oito ações orçamentárias (13.128), agrupadas no pelas variáveis 'Id' e 'Nome da Ação', onde tal quantitativo corresponde ao período de 2000-2019, anos disponíveis para download em momento de definição do escopo do projeto. Cabendo a ressalva, que apenas o período jan/2009 a dez/2019 será o foco primordial do estudo.

Para tanto, impossível seria classificar e analisar todo o escopo das ações orçamentárias disponíveis com os recursos de tempo, investimento e técnicas disponíveis para o projeto. Não

obstante, sabe-se que quantitativo expressivo dos registros oriundos do SIOP, nada se alinham às questões de políticas de mitigação e adaptação à mudança climática efetuadas por meio do Governo Federal.

Dessa forma, para a definição de quais ações orçamentárias irão compor o escopo de análise, a seguinte classificação e diagrama foram desenhados:

Figura 7 - Etapa de seleção de escopo das ações alvo do estudo



Elaborado por: Pesquisadores do projeto PoMuC-Gastos

De acordo com o banco de dados (importado do SIOP e enriquecido em etapas de análise) do projeto PoMuC-Gastos, os seguintes quantitativos de registros se apresentam:

Tabela 1 - Tipo das ações orçamentárias identificadas na base de dados do projeto Gastos

Tipo	Quantitativo	Descrição
Ações SIOP	Agrupadas pelo Id: Desagrupadas: 5.683.085 milhões	São os registros das ações orçamentárias importadas no SIOP e sem qualquer alteração, ou seja, os dados em seu estágio bruto.
Escopo Avaliado	Agrupadas pelo Id:	Correspondem aos registros que de algum modo foram avaliados por meios das técnicas de seleção de escopo utilizadas, sendo estas: inclusão (ações que serão foco de análise no estudo) ou exclusão (ações que não será foco de análise do estudo).

Escopo Classificação	Agrupadas pelo Id:	Correspondem aos registros de ações que serão submetidas a uma regra de classificação preliminar, ou seja, se são ou não ALINHADAS às políticas de mitigação/adaptação a agenda mudança climática.
Escopo Analisado	Agrupadas pelo Id:	São os registros das ações orçamentárias que receberam o esforço de enriquecimento/classificação quanto a sua agenda de mitigação/adaptação à política de mudança climática, ou seja, são ações que possuem status de CORE ou ALINHADAS às diversas políticas implementadas.

Elaborado por: Pesquisadores do projeto PoMuC-Gastos

E desse modo, para o conjunto universo das Ações SIOP, aplicou-se as seguintes regras de seleção de escopo:

Tipo	Seleção do escopo
Inclusão	1- ESCOPO AÇÕES CORE
Inclusão	2- ESCOPO AÇÕES PPCDAm
Inclusão	3- ESCOPO AÇÕES Plano ABC
Inclusão	4- ESCOPO AÇÕES Plano Saúde
Inclusão	5- ESCOPO ÓRGÃOS 71000 e 74000
Inclusão	6 - ESCOPO ÓRGÃO MINISTERIO DA SAUDE INCLUIR OS PROGRAMAS: 0002 - Prevenção e Controle de Doenças Transmitidas por Vetores 0013 - Vigilância Epidemiológica e Ambiental em Saúde 0119 - Saneamento Básico 0122 - Saneamento Ambiental Urbano 0150 - Proteção e Promoção dos Povos Indígenas 1036 - Integração de Bacias Hidrográficas 1138 - Drenagem Urbana Sustentável 1187 - Vigilância, Prevenção e Controle das Doenças Transmitidas por Vetores e Zoonoses 1203 - Vigilância Epidemiológica e Ambiental em Saúde 1289 - Vigilância Sanitária de Produtos, Serviços e Ambientes 1308 - Vigilância, Prevenção e Controle da Malária e da Dengue

	1371 - Vigilância Ambiental em Saúde 2067 - Resíduos Sólidos 2068 - Saneamento Básico 2083 - Qualidade Ambiental 8007 - Resíduos Sólidos Urbanos
Inclusão	7- ESCOPO ÓRGÃOS 22000 - Ministério da Agricultura, Pecuária e Abastecimento e 49000 - Ministério do Desenvolvimento Agrário INCLUIR OS PROGRAMAS: 0135 - Assentamentos Sustentáveis para Trabalhadores Rurais 0137 - Desenvolvimento Sustentável de Projetos de Assentamento 0273 - Luz para Todos 0351 - Agricultura Familiar - PRONAF 0365 - Minimização de Riscos no Agronegócio 0368 - Manejo e Conservação de Solos na Agricultura 1062 - Desenvolvimento da Educação Profissional e Tecnológica 1156 - Pesquisa e Desenvolvimento para a Competitividade e Sustentabilidade do Agronegócio 1161 - Pesquisa e Desenvolvimento Agropecuário e Agroindustrial para a Inserção Social 1169 - Desenvolvimento do Cooperativismo e do Associativismo Rural 1225 - Desenvolvimento da Agricultura Orgânica - Pró-Orgânico 1270 - Proambiente 1334 - Desenvolvimento Sustentável de Territórios Rurais 1336 - Brasil Quilombola 1350 - Educação do Campo (PRONERA) 1409 - Desenvolvimento da Agroenergia 1426 - Conservação, Manejo e Uso Sustentável da Agrobiodiversidade 1427 - Assistência Técnica e Extensão Rural na Agricultura Familiar 1442 - Desenvolvimento Sustentável do Agronegócio 2012 - Agricultura Familiar 2014 - Agropecuária Sustentável, Abastecimento e Comercialização 2018 - Biodiversidade 2022 - Combustíveis 2029 - Desenvolvimento Regional, Territorial Sustentável e Economia Solidária 2042 - Pesquisa e Inovações para a Agropecuária 2052 - Pesca e Aquicultura 2065 - Proteção e Promoção dos Direitos dos Povos Indígenas 2066 - Reforma Agrária e Ordenamento da Estrutura Fundiária 2069 - Segurança Alimentar e Nutricional 2077 - Agropecuária Sustentável

	2078 - Conservação e Uso Sustentável da Biodiversidade
Inclusão	<p>8- ESCOPO ÓRGÃO MINISTÉRIO DO MEIO AMBIENTE 44000 - Ministério do Meio Ambiente</p> <p>INCLUIR OS PROGRAMAS:</p> <p>0052 - Educação Ambiental para Sociedades Sustentáveis</p> <p>0104 - Recursos Pesqueiros Sustentáveis</p> <p>0122 - Serviços Urbanos de Água e Esgoto</p> <p>0150 - Proteção e Promoção dos Povos Indígenas</p> <p>0151 - Proteção de Terras Indígenas, Gestão Territorial e Etnodesenvolvimento</p> <p>0497 - Gestão da Política Nacional de Recursos Hídricos</p> <p>0499 - Áreas Protegidas do Brasil</p> <p>0503 - Prevenção e Combate ao Desmatamento, Queimadas e Incêndios Florestais - Florescer</p> <p>0506 - Nacional de Florestas</p> <p>0508 - Conservação e Uso Sustentável da Biodiversidade e dos Recursos Genéticos</p> <p>0508 - Conservação, Uso Sustentável e Recuperação da Biodiversidade</p> <p>0512 - Zoneamento Ecológico-Econômico</p> <p>0515 - Infra-Estrutura Hídrica</p> <p>1080 - Combate à Desertificação</p> <p>1102 - Agenda 21</p> <p>1107 - Probacias - Conservação de Bacias Hidrográficas</p> <p>1122 - Ciência, Tecnologia e Inovação para Natureza e Clima</p> <p>1145 - Comunidades Tradicionais</p> <p>1287 - Saneamento Rural</p> <p>1305 - Revitalização de Bacias Hidrográficas em Situação de Vulnerabilidade e Degradação Ambiental</p> <p>1332 - Conservação e Recuperação dos Biomas Brasileiros</p> <p>1346 - Qualidade Ambiental</p> <p>1375 - Desenvolvimento do Ensino da Pós-Graduação e da Pesquisa Científica</p> <p>1426 - Conservação, Manejo e Uso Sustentável da Agrobiodiversidade</p> <p>2012 - Agricultura Familiar</p> <p>2018 - Biodiversidade</p> <p>2021 - Ciência, Tecnologia e Inovação</p> <p>2022 - Combustíveis</p> <p>2026 - Conservação e Gestão de Recursos Hídricos</p> <p>2029 - Desenvolvimento Regional e Territorial</p> <p>2029 - Desenvolvimento Regional, Territorial Sustentável e Economia Solidária</p> <p>2032 - Educação Superior - Graduação, Pós-Graduação, Ensino, Pesquisa e Extensão</p> <p>2036 - Florestas, Prevenção e Controle do Desmatamento e dos Incêndios</p> <p>2040 - Gestão de Riscos e Resposta a Desastres</p>

	2045 - Licenciamento e Qualidade Ambiental 2046 - Mar, Zona Costeira e Antártida 2046 - Oceanos, Zona Costeira e Antártica 2050 - Mudança do Clima 2050 - Mudanças Climáticas 2052 - Pesca e Aquicultura 2054 - Planejamento Urbano 2065 - Proteção e Promoção dos Direitos dos Povos Indígenas 2067 - Resíduos Sólidos 2068 - Saneamento Básico 2069 - Segurança Alimentar e Nutricional 2078 - Conservação e Uso Sustentável da Biodiversidade 2080 - Educação de qualidade para todos 2083 - Qualidade Ambiental 2084 - Recursos Hídricos 8007 - Resíduos Sólidos Urbanos
Inclusão	9- ESCOPO ÓRGÃO 20000; 20102; 34000; 60000; 93000
Inclusão	10- ESCOPO ÓRGÃO 30000 - Ministério da Justiça e Segurança Pública
Inclusão	11- ESCOPO ÓRGÃO 52000 - Ministério da Defesa
Inclusão	12- ESCOPO ÓRGÃO 35000 - Ministério das Relações Exteriores
Inclusão	13- ESCOPO ÓRGÃO 38000 - Ministério do Trabalho e Emprego
Inclusão	14- ESCOPO ÓRGÃO 42000 - Ministério da Cultura
Inclusão	15- ESCOPO ÓRGÃO 81000 e ÓRGÃO 57000 - Ministério da Mulher, da Família e dos Direitos Humanos
Inclusão	16- ESCOPO ÓRGÃO 58000 - Ministério da Pesca e Aquicultura
Inclusão	17- ESCOPO ÓRGÃO 25000 - Ministério da Economia
Inclusão	18- ESCOPO ÓRGÃO 26000 - Ministério da Educação
Exclusão	19- ESCOPO ÓRGÃO 24000 - Ministério da Ciência, Tecnologia, Inovações e Comunicações (OK) EXCLUIR APENAS OS PROGRAMAS (já analisados, não há ações orçamentárias aderentes): 0901 - Operações Especiais: Cumprimento de Sentenças Judiciais 2106 - Programa de Gestão e Manutenção do Ministério da Ciência, Tecnologia e Inovação 2106 - Programa de Gestão e Manutenção do Ministério da Ciência, Tecnologia, Inovações e Comunicações 0909 - Operações Especiais: Outros Encargos Especiais 2025 - Comunicações para o Desenvolvimento, a Inclusão e a Democracia 0750 - Apoio Administrativo 0089 - Previdência de Inativos e Pensionistas da União

	0999 - Reserva de Contingência 1008 - Inclusão Digital 2038 - Democracia e Aperfeiçoamento da Gestão Pública
Inclusão	20- ESCOPO ÓRGÃO 47000 - Ministério do Planejamento, Desenvolvimento e Gestão
Inclusão	21- ESCOPO ÓRGÃO 53000 - Ministério do Desenvolvimento Regional INCLUIR OS PROGRAMAS: 2084 - Recursos Hídricos 0379 - Irrigação e Drenagem 0498 - Desenvolvimento Sustentável do Pantanal 0508 - Conservação e Uso Sustentável da Biodiversidade e dos Recursos Genéticos 0512 - Zoneamento Ecológico-Econômico 0515 - Proágua Infra-estrutura 0757 - Gestão da Política de Desenvolvimento Regional e Ordenamento Territorial 1027 - Prevenção e Preparação para Emergências e Desastres 1029 - Resposta aos Desastres e Reconstrução 1036 - Integração de Bacias Hidrográficas 1047 - Desenvolvimento Integrado e Sustentável do Semi-Árido - CONVIVER 1138 - Drenagem Urbana Sustentável 1305 - Revitalização de Bacias Hidrográficas em Situação de Vulnerabilidade e Degradação Ambiental 1343 - Desenvolvimento Sustentável da Aquicultura 1388 - Ciência, Tecnologia e Inovação para a Política Industrial, Tecnológica e de Comércio Exterior (PITCE) 1430 - Desenvolvimento Macrorregional Sustentável 2013 - Agricultura Irrigada 2021 - Ciência, Tecnologia e Inovação 2026 - Conservação e Gestão de Recursos Hídricos 2029 - Desenvolvimento Regional, Territorial Sustentável e Economia Solidária 2040 - Gestão de Riscos e Resposta a Desastres 2048 - Mobilidade Urbana e Trânsito 2051 - Oferta de Água 2052 - Pesca e Aquicultura 2054 - Planejamento Urbano 2068 - Saneamento Básico 2069 - Segurança Alimentar e Nutricional 2076 - Turismo 2077 - Agropecuária Sustentável 2084 - Recursos Hídricos

Inclusão	22- ESCOPO ÓRGÃO 56000 - Ministério das Cidades INCLUIR OS PROGRAMAS: 0122 - Serviços Urbanos de Água e Esgoto 1036 - Integração de Bacias Hidrográficas 1138 - Drenagem Urbana e Controle de Erosão Marítima e Fluvial 1138 - Drenagem Urbana Sustentável 1295 - Descentralização dos Sistemas de Transporte Ferroviário Urbano de Passageiros 2040 - Gestão de Riscos e de Desastres 2040 - Gestão de Riscos e Resposta a Desastres 2048 - Mobilidade Urbana e Trânsito 2054 - Planejamento Urbano 2068 - Saneamento Básico 6001 - Apoio ao Desenvolvimento Urbano de Municípios 6001 - Apoio ao Desenvolvimento Urbano de Municípios de Pequeno Porte 6002 - Apoio ao Desenvolvimento Urbano de Municípios de Médio e Grande Porte 8007 - Resíduos Sólidos Urbanos 9989 - Mobilidade Urbana
Inclusão	23- ESCOPO ÓRGÃO Ministério da Cidadania
Inclusão	24- ESCOPO ÓRGÃO 28000 - Ministério da Indústria, Comércio Exterior e Serviços
Inclusão	25- ESCOPO TRANSPORTE: L:\Projeto_POMUC_2019\POMUC_Gastos_Database\Consultas Escopo\POMUC_Gastos - ações agrupadas e palavras-chave TRANSPORTE
Inclusão	26- ESCOPO ÓRGÃO: 32000 - Ministério de Minas e Energia
Inclusão	27- ESCOPO: Ações Alinhadas Agregadas

Elaborado por: Pesquisadores do projeto PoMuC-Gastos

3. O projeto da Base de Dados

3.1. A estrutura de abordagem por fases

Com a estratégia de abordagem por fases, pretendeu-se obter a maior liberdade de trabalho entre os pesquisadores, bem como sistematizar um processo de desenho da informação o mais alinhado possível a etapa principal do trabalho do projeto PoMuC-Gastos, ou seja, a classificação das ações orçamentárias que são core para o entendimento do fenômeno de gastos para mitigação/adaptação à mudança climática no Brasil.

Grosso modo, se obteve com a estratégia de segmentação por fases:

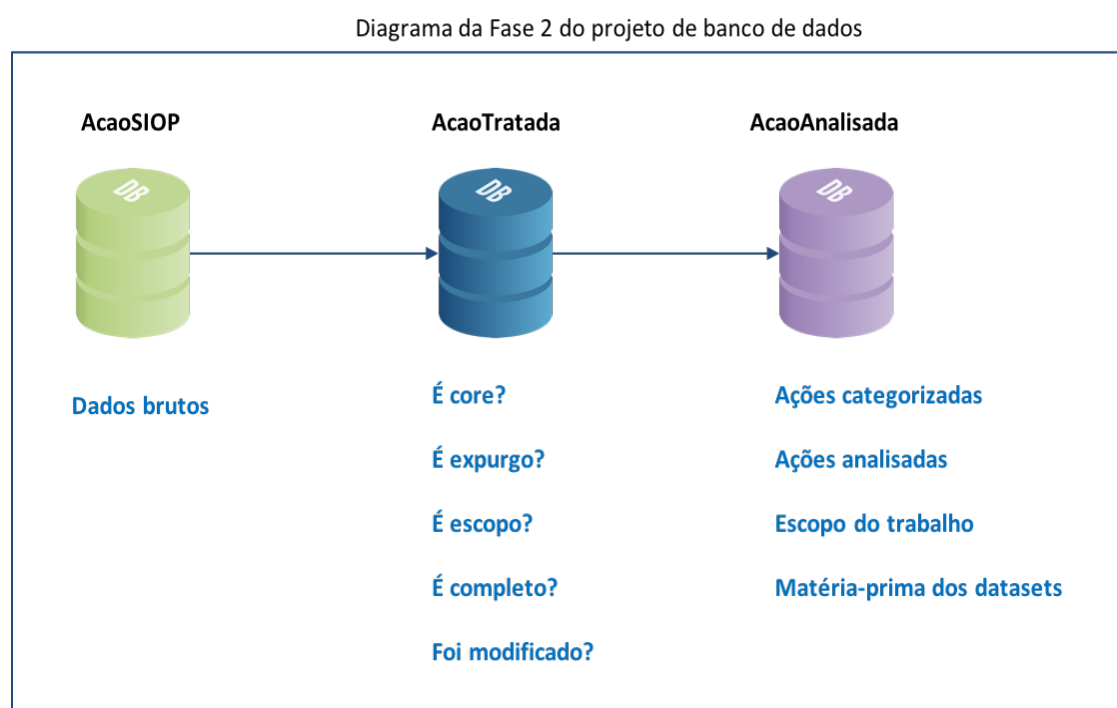
- Possibilidade de junções de dados;
- Possibilidade de validação e replicação da informação;

- Disponibilização do banco de dados na infraestrutura de Rede/Servidores do IPEA;
- Exploração inicial dos dados.

3.3. Fase 2 – tratamento dos dados brutos

Durante a fase 2 o foco foi o processamento e tratamento nos dados brutos oriundos do SIOP. O objetivo principal foi prover um espelhamento entre a base do sistema original e uma base off-line para o projeto PoMuC-Gastos e dessa forma, poder realizar enriquecimentos nos dados, conforme explicitado pelo diagrama abaixo:

Figura 9 - Etapa do desenho do banco de dados do projeto



Elaborado por: Pesquisadores do projeto PoMuC-Gastos

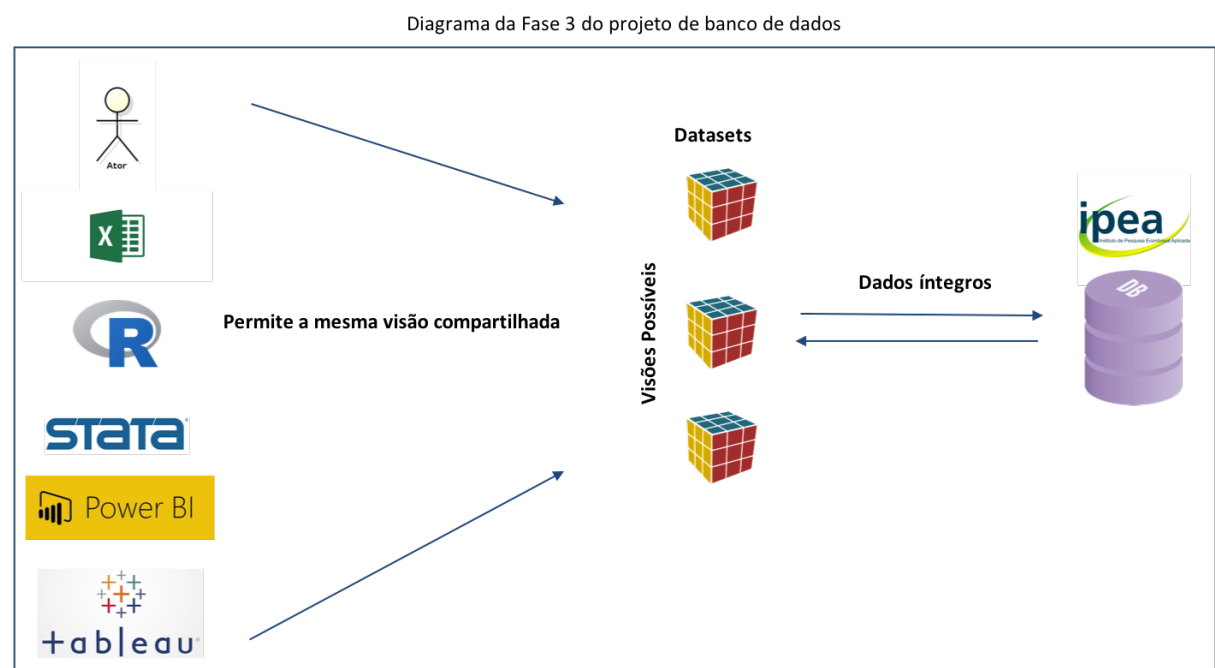
As atividades do processo da fase 2 foram:

- Processamento e tratamento dos dados brutos;
- Melhoramento do banco de dados para fins de desagregação e enriquecimento;
- Acréscimo de variáveis (atributos e categóricas);
- Elaboração de estatísticas descritivas da base de dados;
- Validação e testes dos registros armazenados;
- Definição de regras para seleção do escopo das análises.

3.4. Fase 3 – Criação dos *datasets* (visões) e análise dos dados

A arquitetura final visa tirar a maior parte da complexidade da manipulação / tratamento dos dados dos pacotes estatísticos e colocar as fases iniciais de extração, limpeza e transformação, em um software de banco de dados. Em seguida, foram criados repositórios menores de dados (*datasets*) para que se respondam perguntas específicas de negócio e orientados a determinado formato (série temporal, corte transversal, etc.) de análise.

Figura 10 - Etapa de definição das consultas negociais ao banco de dados enriquecido



Elaborado por: Pesquisadores do projeto PoMuC-Gastos

As atividades do processo da fase 3 foram:

- Criação de visão consolidada suportando informações específicas que se desejam obter;
- Evita trabalho “ad hoc” para se obter as respostas desejadas;
- Possibilita extração de informação por meio de diversas plataformas e softwares;
- Mais focado às perguntas de negócio;
- Normalização dos formatos de dados.

4. Análise de informações processadas

As análises a seguir são fruto de um trabalho que começou ainda na fase de entendimento do projeto, uma vez que até mesmo a decisão sobre quais fontes de dados usar, carrega seu impacto

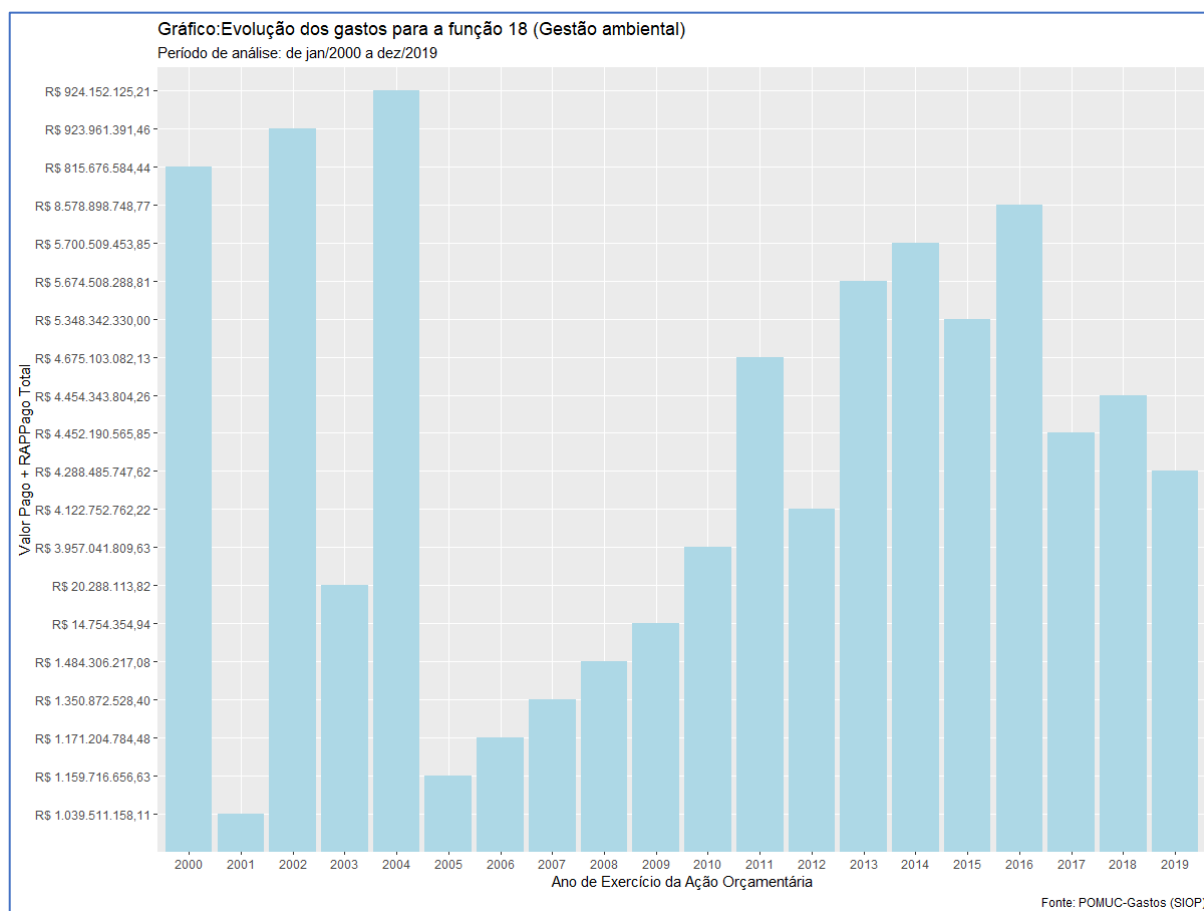
para essa etapa do projeto. E nesse aspecto, os dados oriundos do SIOP foram de suma importância por possuírem boa estrutura, ampla documentação técnica, API de acesso, bem como dicionário de dados.

Cabe a ressalva que tais resultados, não possui a intenção de informar resultados do projeto, tendo o objetivo apenas de explorar resultados técnicos de modo a exemplificar as técnicas desenvolvidas pelos pesquisadores do projeto, para se chegar aos futuros resultados finalísticos do projeto PoMuC-Gastos.

4.1. Evolução dos gastos na função Gestão Ambiental (Identificador 18 no SIOP)

O gráfico (barras) a seguir é uma análise das variáveis ‘Valor Pago’ + ‘RAPPago Total’ somadas e acumuladas por cada Ação Orçamentária que possuam a variável ‘Função’ com o identificador 18 do SIOP, ou seja, são Ações do tipo “Gestão Ambiental”. O escopo da análise se configura pelo período janeiro/2020 a dezembro/2020.

Figura 11 - Exemplo 1 de análise utilizando os datasets



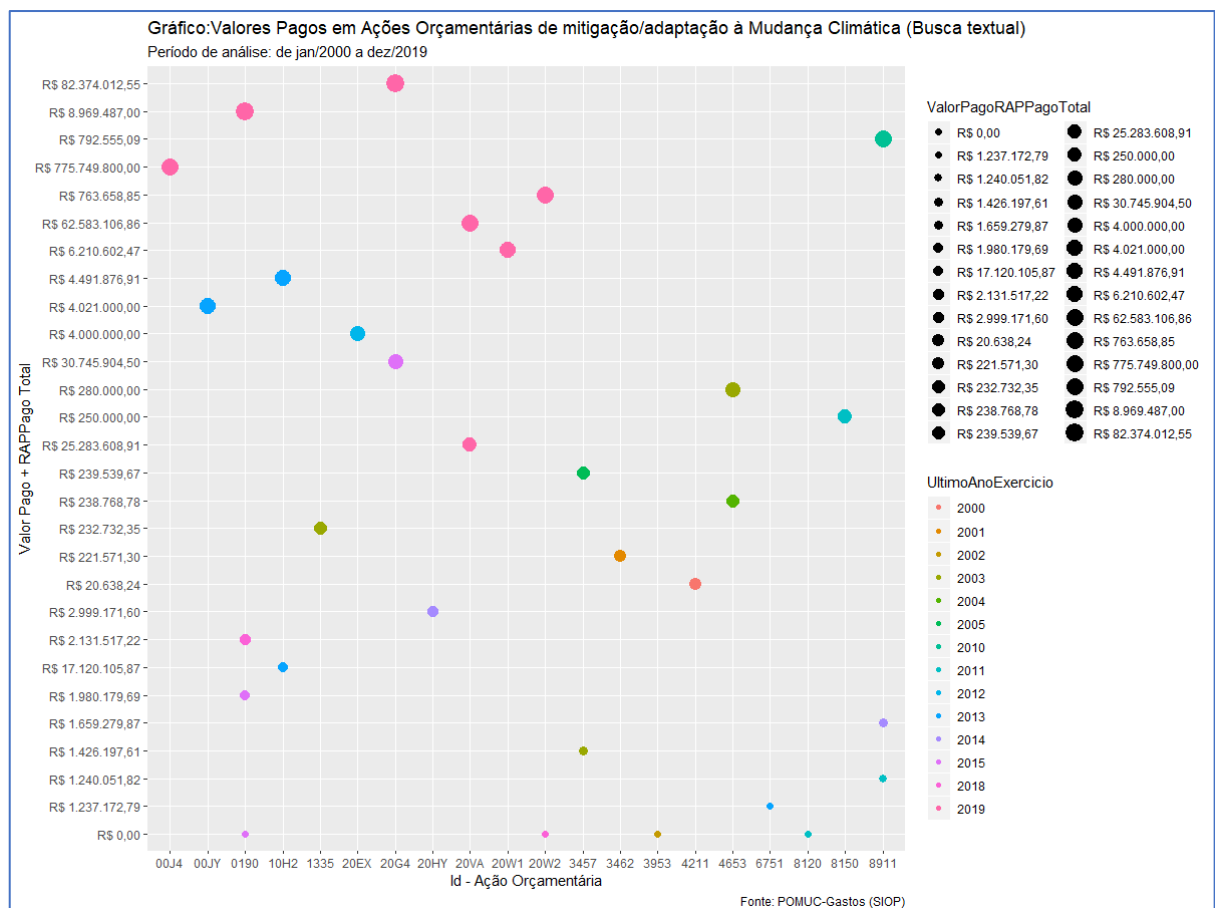
Elaborado por: Pesquisadores do projeto PoMuC-Gastos

4.2. Gastos em ações orçamentárias core, uma análise em busca textual

O gráfico (dispersão) a seguir é uma análise das variáveis ‘Valor Pago’ + ‘RAPPago Total’ somadas e acumuladas por cada Ação Orçamentária que possuam em sua descrição os termos “mudança” e “climática”. O escopo da análise se configura pelo período janeiro/2020 a dezembro/2020 e a consulta retornou um total de vinte ações.

Cabe a ressalva que tais ações foram fruto de uma busca textual (data mining) em um processo simples de mineração de dados, não devendo ser analisadas como resultado finalístico e ou parcial para a dimensão de gastos governamentais em agenda para o clima.

Figura 12 - Exemplo 2 de análise utilizando os datasets



Elaborado por: Pesquisadores do projeto PoMuC-Gastos

5. Tecnologias utilizadas

Ao se iniciar um projeto de cunho tecnológico, algumas perguntas e critérios precisam ser observados durante o processo de escolha de ferramentas e tecnologias a serem utilizadas, sob risco do sucesso ou não da iniciativa, e desse modo para o a solução técnica dada ao projeto PoMuC-Gastos, vislumbrou dentre outros, os seguintes aspectos: a) privilegiou-se ferramentas amplamente utilizadas no meio científico e que se apresentaram robustas para a tarefa; b) uso de soluções open source; c) documentação atual e vigente; d) amplo apoio da comunidade técnica e científica.

5.1. Definição das escolhas

Desse modo, elencam-se as ferramentas utilizadas abaixo:

Nome	Descrição	Justificativa
Microsoft SQL Server 2016	Software de banco de dados proprietário.	O Ipea, bem como sua equipe técnica da Divisão de Sistemas possui larga experiência na plataforma de dados do fabricante Microsoft. Fornecendo ao projeto total apoio em sua criação e acesso aos pesquisadores do banco de dados necessário. Não menos importante, o fator licenciamento já se encontrava dimensionado, bem como toda a estrutura de administração e backup por parte do time técnico do Órgão executor do projeto. Ademais, a plataforma é líder de mercado quanto à gestão de dados, amplamente difundido, confiável e robusto para a missão.
Python 3 (+ pacotes)	Linguagem de programação open source de múltiplo propósito.	É uma linguagem de programação amplamente utilizada no meio técnico e científico, a qual possui diversos pacotes (bibliotecas) e facilidades para se trabalhar com manipulação de arquivos de dados, bem como análise de dados. Para o projeto PoMuC-Gastos foi utilizada na codificação do aplicativo de carga de dados.
R 3.6	Linguagem de programação estatística open source.	É uma linguagem de programação amplamente utilizada no meio técnico e científico, a qual possui diversos pacotes (bibliotecas) e facilidades para se trabalhar com manipulação de arquivos de dados, bem

		como análise de dados. Para o projeto PoMuC-Gastos foi utilizada na codificação das análises de dados, acessando diretamente os <i>datasets</i> construídos.
RStudio (+ pacotes)	IDE de desenvolvimento para a linguagem R open source.	Trata-se de uma ferramenta robusta tecnicamente para desenvolvimento em R e não ocasiona custos financeiros ao projeto.
Microsoft Visual Code	IDE de desenvolvimento de propósito geral open source, fabricada e mantida pela Microsoft.	Trata-se de uma ferramenta robusta tecnicamente para desenvolvimento em Python e não ocasiona custos financeiros ao projeto.
Microsoft Visual Studio 2019	IDE de desenvolvimento de propósito geral open source, fabricada e mantida pela Microsoft.	A ferramenta em questão é o repositório padrão das soluções desenvolvidas baseadas nas tecnologias Microsoft, desse modo, o projeto banco de dados, bem como o formulário de enriquecimento de dados se utilizaram das facilidades e incremento da produtividade fornecidos pelo ecossistema fornecido pelo fabricante.

Elaborado por: Pesquisadores do projeto PoMuC-Gastos

SIGLAS

- API – Application Programming Interface
- IDE – Integrated Development Interface
- SIOP – Sistema Integrado de Planejamento e Orçamento

REFERÊNCIAS

ANEXOS