



ALESANDRE EDSON GOMES DOS SANTOS

**PAINEL DE ACOMPANHAMENTO DAS AVALIAÇÕES TOXICOLÓGICAS
NA ANVISA**

Brasília, DF

2023

ALESANDRE EDSON GOMES DOS SANTOS

PAINEL DE ACOMPANHAMENTO DAS AVALIAÇÕES TOXICOLÓGICAS NA ANVISA

Projeto de Intervenção apresentado como exigência
parcial à obtenção de nota avaliativa no Curso de
Especialização em Ciência de Dados e Inteligência
Artificial da Faculdade de Educação e Ciências da Saúde
do Hospital Alemão Oswaldo Cruz

Orientador: Anderson Franca

Brasília, DF

2023

Aos anônimos trabalhadores do Sistema Único de Saúde
que perderam suas vidas na pandemia da COVID-19.

AGRADECIMENTOS

À Gerência-Geral de Gestão do Conhecimento, Inovação e Pesquisa da Anvisa por proporcionar aos servidores da Anvisa uma capacitação de excelência em Ciência de Dados e Inteligência Artificial.

Aos professores da Faculdade de Educação e Ciências da Saúde do Hospital Alemão Oswaldo Cruz, e em especial, Juliana Rodrigues e Anderson Franca, pela dedicação e pelo prazer em compartilhar o conhecimento.

Aos colegas da Gerência-Geral de Toxicologia da Anvisa que compartilharam suas experiências para o desenvolvimento deste trabalho.

Aos colegas do curso por vivenciar os mesmos sonhos, desafios, alegrias e angústias.

“Só eu posso pensar
Se deus existe, só eu
Só eu posso chorar quando estou triste
Só eu
Eu cá com meus botões de carne e osso
Eu falo e ouço”

Cérebro Eletrônico - Gilberto Gil

RESUMO

Toda empresa que deseja produzir, importar, exportar ou comercializar agrotóxicos no Brasil necessita de aprovação prévia Ministério da Saúde (MS) para minimizar os efeitos dos agrotóxicos na saúde da população; do Ministério do Meio Ambiente (MMA) para minimizar os efeitos dos danos ao meio ambiente e do Ministério da Agricultura, Pecuária e Abastecimento para avaliar a eficácia agronômica dos agrotóxicos. Atualmente na Anvisa, representante do MS, existe uma fila de 3.599 produtos que em alguns casos aguardam mais de oito anos para aprovação sendo que o prazo máximo, conforme a legislação atual, é de até três anos. Essa demora na análise dos processos coloca em risco a regulação do mercado de agrotóxicos e fortalece o argumento dos produtores que querem tirar competências da Anvisa e do Ibama (MMA) no registro de agrotóxicos e, por exemplo, conceder o registro automático para processos que aguardem mais de dois anos. Considerando que a Anvisa dificilmente conseguirá aumentar substancialmente o quadro de servidores, uma possibilidade de melhora dessa situação é a transformação digital da Gerência Geral de Toxicologia, unidade da Anvisa responsável pelo registro de agrotóxicos, com a implantação de uma cultura que utilize dados para as tomadas de decisões. Somente com informações confiáveis é possível que gestores, por exemplo, aloquem pessoal otimamente, evoluam sistemas, priorizem demandas ou antecipem problemas futuros com a identificação de gargalos operacionais. Assim, a proposta desse projeto de intervenção é criar uma ferramenta que apresente de maneira rápida e clara as principais informações da área de registros e pós-registro de agrotóxicos. Esse painel apresentará as estatísticas descritivas dos processos como número de entradas e de conclusões, tempo médio e mediano em dias das várias fases do processo por ano; estimará o tempo de conclusão de análise; e documentará toda a metodologia e tecnologia utilizada. A extração, tratamento e carga de dados do projeto foram feitas em python e as informações apresentadas em painéis criados no Power BI. O projeto de intervenção já apresenta alguns resultados que foram bem recebidos entre técnicos e gestores, mas ainda está em execução. Apesar dos bons resultados iniciais, o projeto encontrou algumas dificuldades tecnológicas como computadores lentos, impossibilidade de utilização de softwares livres e dificuldade de acesso às tabelas de sistema com os dados da área. Um achado importante do projeto foi a baixa qualidade das informações no sistema devido a possibilidade do não preenchimento pelo corpo técnico. Isso comprometeu a modelagem dos dados e somente foi possível estimar o tempo total de conclusão para dois produtos. Ainda assim, o projeto é bom e deve ser continuado desde que sejam oferecidas condições tecnológicas melhores e que haja um processo de conscientização da importância do preenchimento correto das informações nos sistemas.

Palavras-chave: Transformação digital, cultura *data-driven*, Painel de informações

LISTA DE SIGLAS

Agrofit – Sistema de Agrotóxicos Fitossanitários
Anvisa – Agência Nacional de Vigilância Sanitária
Aplan – Assessoria de Planejamento
COPSI – Coordenação de Processos Simplificados
Datavisa – Sistema de Produtos e Serviços sob Vigilância Sanitária
DIRE3 – Terceira Diretoria
DOU – Diário Oficial da União
FAO - Organização das Nações Unidas para Alimentação e Agricultura
FECS – Faculdade de Educação em Ciências da Saúde
GEAST – Gerência de Avaliação de Segurança Toxicológica
GEMAR – Gerência de Monitoramento e Avaliação do Risco
GGCIP - Gerência-Geral de Gestão do Conhecimento, Inovação e Pesquisa
GGTIN - Gerência-Geral de Tecnologia da Informação
GGTOX – Gerência-Geral de Toxicologia
GHS - Globally Harmonized System of Classification and Labelling of Chemicals
GPREQ – Gerência de Produtos Equivalentes
HAOC – Hospital Alemão Oswaldo Cruz
Ibama - Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis
Limites Máximos de Resíduos (LMR)
MAPA – Ministério da Agricultura, Pecuária e Abastecimento
Ministério da Agricultura, Pecuária e Abastecimento (MAPA)
Ministério do Meio Ambiente (MMA)
MS - Ministério da Saúde
OLAP - Processamento Analítico Online
OMS – Organização Mundial de Saúde
PARA – Programa de Análise de Resíduos de Agrotóxicos em Alimentos
RDC – Resolução da Diretoria Colegiada
RET – Registro Especial Temporário
SIPTOX - Sistema de Peticionamento da Toxicologia
SUS – Sistema Único de Saúde
TCU – Tribunal de Contas da União
VPN – Rede privada virtual

LISTA DE ILUSTRAÇÕES

GRÁFICOS

Gráfico 1: Quantidade de Expedientes de registro de agrotóxicos que entraram, saíram e tamanho da fila (2003-2023)	16
Gráfico 2: Estatísticas descritivas do tempo total de conclusão dos expedientes 5041 e 5065.....	32
Gráfico 3: Autocorrelação e Autocorrelação parcial dos expedientes 5041 e 5065.....	33
Gráfico 4: Previsões do tempo total dos expedientes 5041 e 5065	42

QUADROS

Quadro 1: Fontes de dados do Datavisa	22
Quadro 2: Fontes de dados do Tabela Assuntos.....	23
Quadro 3: Fontes de dados do Tabela Situação.....	24
Quadro 4: Fontes de dados do Critério Fila	25
Quadro 5: Cronograma do projeto de Intervenção	42

FIGURAS

Figura 1: Tamanho físico de um processo de Avaliação Toxicológica.....	15
Figura 2: Painel de processos em Andamento	37
Figura 3: Painel de Estatísticas Descritivas.....	39
Figura 4: Painel de Indicadores de Performance	40
Figura 5: Painel de Produtos Finalizados.....	41

LISTA DE TABELAS

Tabela 1: Quantidade de expedientes por ano de entrada e situação (2002-2023)	27
Tabela 2: Quantidade de expedientes por tipo de pleito e situação	28
Tabela 3: Quantidade de expedientes por fase do processo e tipo de pleito	29
Tabela 4: Leiaute das tabelas geradas de entrada e saída.....	29
Tabela 5: Tabela Indicadores Final (IND_FINAL)	30
Tabela 6: Quantidade de expedientes por assunto e ano de finalização (2005 - 2023).....	31
Tabela 7 : Resultados dos Critérios de Informação por modelo e Assunto	34
Tabela 8: Quantidade de expedientes em andamento por Tipo de produto e ano da última situação (2005 - 2023)	38

SUMÁRIO

1. INTRODUÇÃO.....	11
1.1 Registro de Agrotóxicos	12
1.2 Registro de Agrotóxicos na Anvisa	13
1.3 Situação Atual	14
1.4 A intervenção	16
1.5 Importância do Projeto	17
2 OBJETIVOS	19
2.1 Objetivo Geral	19
2.2 Objetivos Específicos.....	19
3 DESENVOLVIMENTO DO PROJETO	19
3.1 Coleta de Dados	21
3.2 Preparação dos Dados.....	25
3.3 Análise Exploratória dos Dados.....	27
3.4 Modelo de Previsão.....	30
3.5 Ferramentas Utilizadas.....	34
4 RESULTADOS DO PROJETO	36
4.1 Expedientes em Andamento	36
4.2 Estatísticas Descritivas	38
4.3 Indicadores de Performance	39
4.4 Expedientes Finalizados	40
4.5 Previsão de Tempo Total.....	41
5 CRONOGRAMA DE IMPLANTAÇÃO DO PROJETO	42
6 ORÇAMENTO	43
7 CONSIDERAÇÕES FINAIS	43
REFERÊNCIAS.....	47
APÊNDICES	48
Apêndice A – Script em python: Extração dos dados.....	49
Apêndice B – Script em python: Tratamento e análise descritiva	57
Apêndice C – Script em python: Modelo de predição	64
ANEXOS.....	70
ANEXO 1 – Solicitação de Instalação Anaconda.....	71
ANEXO 2 – Solicitação de desenvolvimento Cubo PARA	73

1. INTRODUÇÃO

Por ser um produto que traz grandes riscos à saúde da população e ao meio ambiente, os agrotóxicos exigem uma regulação do Estado para sua comercialização e produção. Esse controle é recente no mundo inteiro, trata dos anos de 1950, e no Brasil a lei atual, Lei 7.802/1989, coincide com redemocratização do país, sensível à questão ambiental, e assegura que, no Artigo 225 da Constituição Federal “Todos têm direito ao meio ambiente ecologicamente equilibrado”, e que “para assegurar a efetividade desse direito, incumbe ao Poder Público controlar a produção, a comercialização e o emprego de técnicas, métodos e substâncias que comportem risco para a vida, a qualidade de vida e o meio ambiente”.

A Lei 7.802/1.989 trouxe grandes avanços para a regulação do mercado de agrotóxicos no Brasil (LUCCHESE, 2013), mas setores organizados dos produtores de agrotóxicos e produtores agrícolas acusam o processo de registro de agrotóxicos de burocrático e lento o que impede que o agronegócio tenha disponibilidade de opções inovadoras e baratas e que acarreta perda de competitividade e tentam desde o ano 2.000 flexibilizar a lei.

Apesar dos esforços das instituições registrantes, o tempo total de registro para alguns agrotóxicos supera facilmente oito anos, sendo que o Decreto 10.833 de 2021 já aumentou o prazo anterior de 120 dias para até três anos. Atualmente existe uma fila de 3.599 produtos aguardando análise para de registro, o que reforça o argumento dos que defendem alteração na legislação.

Esse projeto de intervenção cria um painel de informações com estatísticas descritivas do processo de registro e pós-registro de agrotóxicos e um modelo de previsão do tempo necessário para a conclusão de uma análise com base nos conhecimentos da ciência de dados. Essa ferramenta auxiliará os gestores na tomada de decisão e na melhoria na eficiência dos processos para a diminuição da fila de produtos que aguardam análise para registros de agrotóxicos.

Nessa introdução são apresentados como é o registro de agrotóxicos, como se dá o registro de agrotóxicos na Agência Nacional de Vigilância Sanitária (Anvisa), qual a situação atual dos registros, a proposta de intervenção e a importância desse projeto.

1.1 Registro de Agrotóxicos

A Lei 7.802/1989 define agrotóxico como os produtos e os agentes de processos físicos, químicos ou biológicos, destinados ao uso nos setores de produção, no armazenamento e beneficiamento de produtos agrícolas, nas pastagens, na proteção de florestas, nativas ou implantadas, e de outros ecossistemas e também de ambientes urbanos, hídricos e industriais, cuja finalidade seja alterar a composição da flora ou da fauna, a fim de preservá-las da ação danosa de seres vivos considerados nocivos.

A legislação anterior, pouco rigorosa, que permitiu o avanço de centenas de substâncias tóxicas, muitas vezes proibidas em outros países (PELAES, TERRA e SILVA, 2013), deu lugar a uma nova lei atribuindo a três ministérios poder para analisar e permitir o registro de um novo agrotóxico no país, conforme o enfoque da sua área de competência: Ministério da Saúde (MS) para minimizar os efeitos dos agrotóxicos na saúde da população; o Ministério do Meio Ambiente (MMA) para minimizar os efeitos dos danos ao meio ambiente e o Ministério da Agricultura, Pecuária e Abastecimento (MAPA) para avaliar a eficácia agrônômica dos agrotóxicos.

Assim, toda empresa que deseja produzir, importar, exportar ou comercializar agrotóxicos no Brasil inicialmente necessita se cadastrar e se manter atualizada junto aos três órgãos registrantes. O registro de agrotóxico é um processo de avaliação e aprovação de documentos científicos, analisados em resposta às exigências de eficácia e segurança feitas ao produto proposto. Após o envio desses dossiês, os órgãos registrantes têm o ônus de provar que o produto em questão apresenta riscos – caso contrário, o produto é liberado.

Cabe a Anvisa, órgão representante do Ministério da Saúde, a análise dos dados referentes à toxicidade dos agrotóxicos, que são obtidos mediante provas em animais experimentais e outros procedimentos de laboratório. Esses dados permitirão a avaliação, a classificação toxicológica e a fixação de Limites Máximos de Resíduos (LMR) em alimentos, os quais deverão ser apresentados pela empresa registrante. Comprovado que o produto a ser registrado não apresenta características teratogênicas, carcinogênicas ou mutagênicas o produto é classificado conforme a rotulagem de produtos químicos em função da toxicidade aguda que passou ser adotada pela Resolução RDC/ANVISA Nº 294 de 29/07/2019 segundo o Sistema de Classificação Globalmente Unificado (*Globally Harmonized System of Classification and Labelling of Chemicals –GHS*).

Segundo informações do Sistema de Agrotóxicos Fitossanitários¹ (Agrofit) do MAPA, em 2023, são quase 500 ingredientes ativos registrados no país, incluindo os biológicos, com cerca de 3.215 produtos formulados. Os dados mais recentes de comercialização são de 2021, disponibilizados pelo IBAMA², que mostram que houve uma comercialização de 720.87 mil toneladas de agrotóxicos, o que coloca o Brasil como um dos maiores consumidores mundiais desses produtos entre 1990 e 2021, atrás apenas da China e se revezando com os Estados Unidos entre a 2^a e 3^a posição, segundo dados da Organização das Nações Unidas para Alimentação e Agricultura (FAO)³.

1.2 Registro de Agrotóxicos na Anvisa

A Gerência-Geral de Toxicologia (GGTOX) da Anvisa é a responsável pela análise toxicológica com fins de registro de agrotóxicos, representando o Ministério da Saúde. A GGTOX existe desde a criação da Anvisa pela Lei n. 9.782, de 1999, e sua estrutura é composta de três gerências e conta com o número de servidores que oscilou entre quinze e vinte direcionados para as análises toxicológicas ao longo desses anos.

Vários fatores fizeram com que o passivo de processo de registros de agrotóxicos se acumulasse ao longo dos anos. Em 2002, o Decreto 4.074 trouxe mais atribuições para a Anvisa com os processos para Registro Especial Temporário (RET) de produtos destinados à experimentação e o registro de Componentes, além de tornar a análise mais complexa levando-se a um maior tempo para a aprovação com a criação dos produtos técnicos equivalentes (PELAES, TERRA e SILVA, 2013).

Quanto aos processos de trabalho, a Auditoria Operacional do Tribunal de Contas da União (TCU), em 2012, verificou a ausência de sistema informatizado para o gerenciamento do registro de agrotóxicos, pois não foi criado o sistema que deveria interagir com os três órgãos, conforme consta no regulamento da Lei 7.802/1989 e também não se utilizava o Sistema de Produtos e Serviços sob Vigilância Sanitária (Datavisa), sistema de registros da Anvisa, controlando seus processos por meio de planilhas.

1 Disponível em: <<https://indicadores.agricultura.gov.br/agrofit/index.htm>> Acesso em 24 jul, 2023.

2 Disponível em: <<https://www.gov.br/ibama/pt-br/assuntos/quimicos-e-biologicos/agrotoxicos/relatorios-de-comercializacao-de-agrotoxicos>> Acesso em 24 jul, 2023.

3 Disponível em: <<https://www.fao.org/faostat/en/?#data/RP/visualize>> Acesso em 24 jul, 2023.

Esses fatores, juntamente com o aumento da demanda por registros de agrotóxicos e a oscilação da agenda regulatória da GGTOX sendo pautada por outros temas como a reavaliação, fiscalização de produtos e o Programa de Análise de Resíduos de Agrotóxicos em Alimentos (PARA), deixaram o processo de registro de agrotóxicos à margem das principais atividades da Gerência-Geral e fizeram com que a fila de registro de produtos se acumulassem ao longo dos anos. Em março de 2012 havia 1.040 pedidos de registro, sendo a capacidade estimada de 150 processos por ano (ROSA DOS SANTOS, 2012).

Em 2016, houve uma reestruturação na GGTOX que foi o ponto de partida de uma gestão motivada a simplificar os processos de trabalho e melhorar a gestão e a eficiência (AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA, 2016). Tais mudanças envolveram autoavaliação e análise de gestores, servidores, instituições governamentais, fabricantes de produtos agrotóxicos e representantes da sociedade. Esses esforços trouxeram bons resultados e, conforme os relatórios de gestão de 2016 e 2017, houve um aumento de processos decididos, passando de 232 em 2015 para 489 em 2017, o que correspondeu um aumento de 122%.

1.3 Situação Atual

Atualmente cerca de 3.600 processos de avaliação toxicológica aguardam análise da Anvisa⁴ para a obtenção do registro e consequente comercialização do produto. São processos diversos como registro de produto bioquímico, registro de produto microbiológico novo, avaliação toxicológica para alteração de intervalo de segurança ou avaliação toxicológica para alteração de frequência de aplicação que ao não cumprir o prazo legal para sua conclusão colocam em risco toda a regulação do mercado de agrotóxicos.

Em alguns casos, os processos de agrotóxicos chegam a esperar por mais de oito anos para a obtenção do registro, descumprindo em muito tempo o prazo legal estabelecido. Para não aguardar todo esse tempo por um registro, muitas empresas recorrem para terem seus pleitos analisados por demandas judiciais, em 2022, foram 60 processos analisados dessa forma. Recentemente o Decreto nº 10.833, de 2021 aumentou o prazo das análises, que era de 120 dias, para até três anos conforme a categoria do produto.

A utilização dos recursos da tecnologia de informação poderia diminuir esse problema do passivo de produtos para análise. Desde o ano 2.000, a Anvisa recebeu mais de 33.325

⁴ Fonte: <https://consultas.anvisa.gov.br/#/filas/>

processos que geraram cerca de 82.816 documentos relacionados. Alguns desses documentos chegam facilmente a 5.000 páginas (Figura 1) e ainda podem passar por até 70 situações distintas como por exemplo: “aguardando análise”, “em análise”, ou “em exigência” que aumenta o volume de dados armazenado. No passado era possível controlar esses documentos manualmente por planilhas, hoje a realidade mostra que é impossível trabalhar sem um sistema de informação eficiente. Uma possibilidade para melhorar esse cenário é uma transformação digital com a implantação de uma cultura *data-driven*, que utiliza os dados para as tomadas de decisões, aplicando ações estratégicas por meio de previsões e insights que os dados fornecem (SANTOS, 2021).

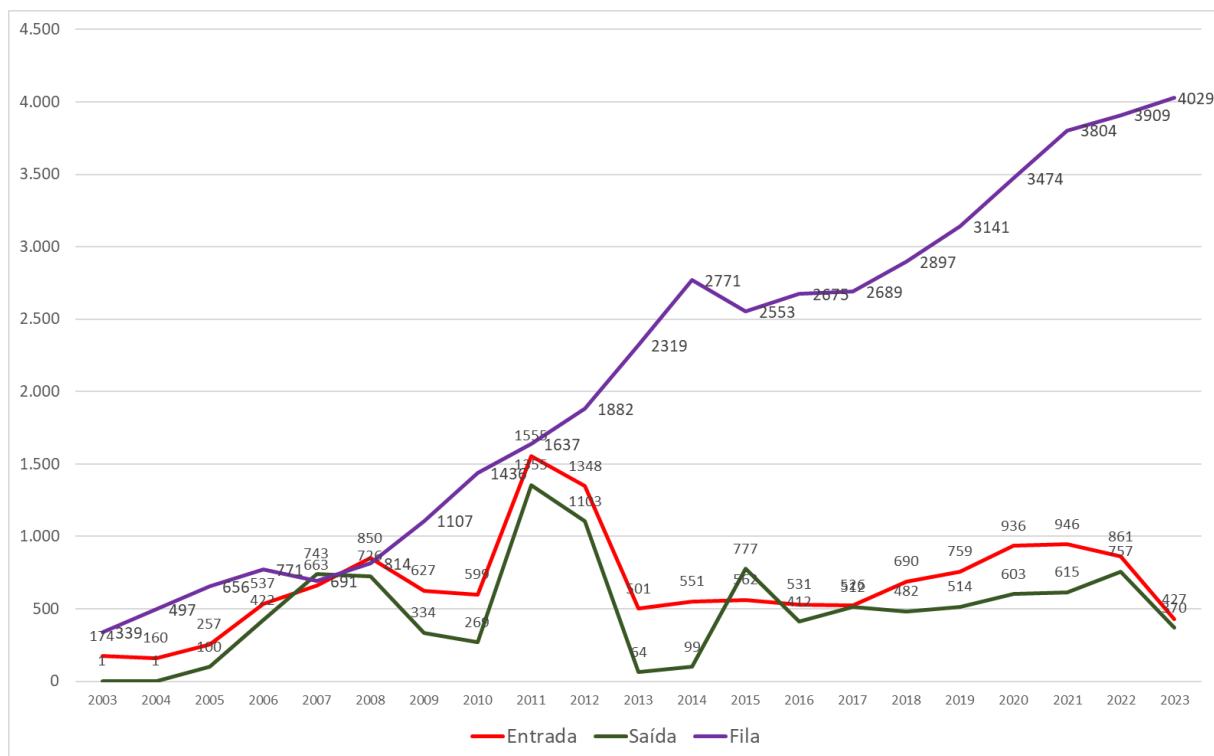
Figura 1: Tamanho físico de um processo de Avaliação Toxicológica



Fonte: Registro do autor

Apesar dos esforços recentes para o aumento de eficiência, o número de processos concluídos ainda é menor que os de processos recebidos, ou seja, a força de trabalho da Anvisa não é capaz de acompanhar a demanda de registros e como consequência a fila de registro só aumenta conforme pode ser observado no Gráfico 1.

Gráfico 1: Quantidade de Expedientes de registro de agrotóxicos que entraram, saíram e tamanho da fila (2003-2023)



Fonte: Datavisa – 30/06/2023

1.4 A intervenção

A proposta desse projeto de intervenção é criar uma ferramenta que apresente de maneira rápida e clara as principais informações da área de registros e pós-registro de agrotóxicos como a quantidade de processos que entraram, que estão em análise, quantas foram concluídas ou quantas estão em fila em um período de tempo específico como mês ou ano. Estas informações são vitais para o acompanhamento dos gestores com alocação de pessoal, evolução de sistemas, priorização de demandas ou ainda para antecipação de problemas futuros com a identificação de gargalos operacionais que podem comprometer os processos de trabalho.

Considerando que existem cerca de 65 tipos diferentes de registro ou pós-registro, 70 fases que os processos podem passar, uma série histórica de 23 anos, 276 meses, em quatro áreas e dois sistemas de origem distintas isso geraria uma quantidade grande de cruzamentos sendo impossível pré-formatar esses relatórios em planilhas ou documentos, por isso a melhor maneira para disponibilizar essa informação é através de painéis ou dashboards (CARVALHO E MELO, 2018), pois podem ser atualizados diariamente e podem ser confeccionados conforme a necessidade do usuário.

Além de apresentar estatísticas descritivas, esse painel vai também apresentar indicadores de desempenho, “key performance indicators” (KPI), que permitem avaliar de forma contínua a orientação e evolução de uma determinada atividade ou processo que é realizado numa instituição. Esses indicadores devem ir além do que o acompanhamento dos processos, pois pretendem avaliar, analisar, sugerir, decidir ou até antecipar o caminho do processo em análise (ANTUNES, 2020).

Um indicador de desempenho estabelecido dentro da área de registro de produtos da Anvisa é o tempo médio em dias que a área levou para concluir a análise de um processo ou ainda o tempo médio que um processo levou para ser registrado em todas as fases do processo. Conhecendo esse tempo médio, por exemplo, é possível descobrir se a produtividade da área está aumentando independentemente do número de servidores que estão realizando as análises.

Um outro resultado que esse dashboard deve incorporar é uma variável preditiva do tempo necessário para a conclusão dos processos com base nas informações passadas e considerando variáveis como o tipo de produto e o histórico das análises realizadas. Como esse tempo médio de conclusão vem caindo conforme a dinâmica de melhoria dos processos, esse indicador também ajudará o gestor na reorganização das filas de produtos a serem analisados. Aprimorado esse indicador, será possível transferir previsibilidade ao setor regulado estimando um tempo médio de quanto tempo seu produto poderá ser comercializado.

A GGTOX ainda está na fase inicial da sua transformação digital, pois passou a utilizar tardiamente o sistema Datavisa e a maioria das informações não estão sistematizadas, ou seja, são documentos anexados em formato (OCR – reconhecimento de caractere óptico). Os esforços da implantação da Ferramenta de Leitura Otimizada no Registro de Agrotóxicos (Flora) e do novo Sistema de Registro de Agrotóxicos (SIA) são justamente para avançar nessa transformação digital com a possibilidade de que muito do trabalho que hoje é realizado manualmente, como conferência da veracidade das informações, seja feito automaticamente.

1.5 Importância do Projeto

Esse projeto é importante para a GGTOX, pois auxilia os gestores com a tomada de decisões baseada em dados. Com um acompanhamento em tempo real é possível identificar processos de trabalhos que não estão funcionando plenamente e juntamente com os técnicos responsáveis propor soluções.

O quadro de pessoal da Anvisa, definido pela Lei de criação dos cargos e carreiras nas Agências Reguladoras⁵, é de 810 especialistas para todas as áreas, quantidade pequena quando comparada com a divisão de agrotóxicos da Agência de Proteção Ambiental dos Estados Unidos da América, que possui cerca de oitocentos funcionários em seus quadros (TRIBUNAL DE CONTAS DA UNIÃO, 2013). Apesar de recentemente o Ministério da Gestão e da Inovação em Serviços Públicos (MGI) ter autorizado concurso com 50 vagas para a Anvisa⁶, esse quantitativo não consegue nem cobrir o déficit atual do número de vagas estabelecidos na Lei de Criação de cargos e carreiras acima referida. Apesar da escassez de recursos humanos, as atribuições legais vão continuar as mesmas, obrigando os gestores a obterem soluções inovadoras e tendo que fazer mais com menos. Nesses casos, as soluções tecnológicas são as mais utilizadas, portanto essa ferramenta pode auxiliar os gestores na proposição de novas ações pautando suas demandas em dados e não em suposições.

Importante também destacar que ao oferecer essa ferramenta aos técnicos e gerentes, retornando informações úteis do dia a dia, instala-se uma cultura da informação, reforçando o papel de cada ator no processo de trabalho, aumentando por exemplo a responsabilidade do preenchimento das informações corretas no sistema. Esse pertencimento aumenta a sinergia da área que pode ser convertida em maior produtividade com mais avaliações toxicológicas sendo realizadas e um aumento da qualidade técnica dessas avaliações.

Por último, o projeto de intervenção é importante para ajudar a GGTOX a cumprir com a legislação quanto aos prazos estabelecidos para as avaliações toxicológicas e garantir o cumprimento das normas é o principal para que não haja pioras nesse modelo de regulação, como o Projeto de Lei 6.299/2002, recentemente aprovado na Câmara Federal e que aguarda votação no Senado Federal, que retira competências da Anvisa e do Ibama no processo de registro, concedendo registro automaticamente para qualquer produto que aguardar mais de dois anos pelo parecer final.

Importante também ter informações sólidas e confiáveis para dialogar com os setores organizados da sociedade, como a mídia, Organizações Não Governamentais (ONGs) ligadas ao meio-ambiente e academia, que alertam para o fato de o governo brasileiro registrar excessivamente agrotóxicos. Somente com essas informações é possível fazer um debate

⁵ LEI Nº 10.871, DE 20 DE MAIO DE 2004 disponível em http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2004/lei/l10.871.htm

⁶ <https://www.gov.br/anvisa/pt-br/assuntos/noticias-anvisa/autorizado-concurso-com-50-vagas-de-nivel-superior-para-a-anvisa>

transparente sobre os efeitos adversos dos agrotóxicos na saúde da população e no meio ambiente.

2 OBJETIVOS

2.1 Objetivo Geral

Criar um painel em formato de dashboard com as principais informações do processo de registro e pós-registro de agrotóxicos na Anvisa, que auxilie os gestores da Gerência de Toxicologia na tomada de decisões baseada em dados.

2.2 Objetivos Específicos

Descrever as principais estatísticas dos processos de registro e pós-registro como número de entrada e de conclusões, tempo médio e mediano em dias das várias fases do processo por ano.

Projetar o tempo estimado de conclusão de processo por código de assunto com base na série histórica utilizando conhecimento em ciência de dados.

Documentar toda a metodologia e a tecnologia utilizada para que outros técnicos possam reproduzir o painel, inclusive em outras áreas de registro da Anvisa.

3 DESENVOLVIMENTO DO PROJETO

Baseado na metodologia ativa de ensino baseada em problemas (PBL), o dashboard a ser criado pelo projeto de intervenção tem como objetivo responder perguntas simples que os gestores e técnicos fazem no dia a dia no acompanhamento dos processos de registro e pós-registro de agrotóxicos como por exemplo, quantos processos de registro de produto formulado com base em produto técnico equivalente foram concluídos nesse ano? Nesse mesmo período do ano passado já tinha concluído mais ou menos processos com essas mesmas características? A continuar nesse ritmo de conclusão de análises e considerando as novas entradas, a fila desses processos vai aumentar ou diminuir?

Essas são perguntas simples, mas no momento sem respostas na GGTOX. Apesar dos processos serem informatizados, o sistema antigo e de difícil atualização, não permite relatórios

acessíveis que respondam a esses questionamentos. O propósito dessa ferramenta é justamente atender a essa necessidade de informações, motivada pela melhoria dos processos de trabalho e ganho de eficiência culminando em análises mais rápidas e consequente diminuição da fila e cumprimento dos prazos.

Logo após a definição do problema, o próximo passo é identificar onde estão as informações que respondam às perguntas. Nessa fase são definidas as variáveis que entrarão no cubo de dados OLAP (Processamento Analítico Online). A construção dessa base normalizada e menor do que a base de dados original é de responsabilidade da equipe de inteligência de dados da Anvisa, área com expertise em retirar as informações do banco de dados operacional e disponibilizá-las para as áreas.

No projeto de intervenção, os interlocutores de dados da GGTOX, devem acessar esse cubo de dados, através da ferramenta de programação em python e combinar com tabelas auxiliares para gerar os outputs de tabelas que abastecerão o Power BI que é a ferramenta que construirá o painel.

No Power BI será construído o dashboard que disponibilizará as estatísticas descritivas com todos os cruzamentos possíveis por tipo de serviço ao longo do tempo. Também será disponibilizada a projeção do tempo total de um serviço ser concluído e uma listagem dos processos que estão em análise no momento. Tudo isso atualizado diariamente.

Assim, o processo com as etapas que compõe a estratégia desse projeto de intervenção são as seguintes:

Definição do Problema: criação de um dashboard com as principais informações do processo de registro e pós-registro de agrotóxico.

Coleta de Dados: identificação de como as informações são geradas, armazenadas e alteradas nos sistemas de informações da Anvisa.

Preparação dos Dados: Extração dos dados no sistema transacional e armazenamento em formato analítico no banco de dados SQL Server. Leitura e tratamento dessas informações pela GGTOX através de um script em python. Combinações com tabelas auxiliares e geração de output que alimentará o Power BI.

Análise exploratória dos dados: principais resultados como quantidade de processos, média e mediana do tempo levado em cada fase do processo. Tratamento das informações e seleção das variáveis com possibilidade de entrar no modelo de previsão.

Modelo de Previsão: desenvolvimento de um modelo de previsão do tempo para conclusão do processo por tipo de assunto.

Assim esse capítulo apresenta a estratégia do desenvolvimento do projeto e registra todos os passos metodológicos e conceitos. Esse registro é necessário para que outros técnicos possam chegar aos mesmos resultados e aperfeiçoamentos no futuro, ou ainda, realizar o mesmo painel em outras áreas da Anvisa. Esse capítulo detalha as principais etapas do desenvolvimento do projeto como a coleta e preparação de dados, a análise exploratória, e a apresentação do modelo de previsão além de trazer as ferramentas utilizadas pelo projeto.

3.1 Coleta de Dados

A primeira parte do desenvolvimento do painel de informações é especificar a origem dos dados, como elas são obtidas e o que significam. A principal fonte de informações está no chamado ciclo de vida do processo, que detalha as informações essenciais, por exemplo que empresa solicitou qual serviço, quando e qual o histórico do processo de requisição pelas áreas responsáveis pela análise. Todos esses dados estão armazenados no sistema Datavisa.

Com a criação da Anvisa fez-se necessário a criação de um sistema que substituísse os vários sistemas de registros de produtos da saúde que estavam nas unidades do Ministério da Saúde. Para isso foi criado o Sistema de Produtos e Serviços sob a Vigilância Sanitária – Datavisa, que inicialmente tinha a função de acompanhar os andamentos de processos, mas conforme o tempo foi sendo adaptado para gerenciar informações de tipos de medicamentos, alimentos, tabaco, cosméticos e os demais produtos objetos de regulação. O Datavisa também gerencia processos de trabalho como autorização de funcionamento de empresas, importações de produtos, processos administrativos sanitários. Ou seja, o Datavisa é um sistema imenso, presente em todas as áreas da Anvisa e com tecnologia obsoleta há mais de uma década. Esse sistema se propõe a gerenciar várias áreas, mas não o faz da melhor forma e atualmente está em andamento uma modernização dessa forma de peticionamento. Especificamente para a GGTOX, desde 2021 está em desenvolvimento o Sistema de Informações de Agrotóxicos, que criará uma porta de entrada única para o registros de agrotóxicos e organizará todo o trabalho dos três órgão em uma única plataforma. A previsão de implantação desse novo sistema é a partir do segundo semestre de 2024.

O processo de peticionamento é quando uma empresa previamente cadastrada pleiteia um serviço para a Anvisa. Essa solicitação é feita através do Solicita⁷, que é uma aplicação amigável do Datavisa que coleta dados específicos e documentos em pdf e realiza o tramite para o pagamento de taxa, conforme o tipo de serviço ou assunto. Nesse processo é que são geradas as principais informações do solicitante como nome da empresa, tipo de serviço e data da solicitação.

Importante informar que quando a empresa cria uma nova petição, por exemplo, o registro de um novo produto, é gerado um número de processo único. Quando são feitas novas solicitações para aquele mesmo produtos são criados novos expedientes para esse mesmo processo. Assim as principais informações que serão utilizadas relativas ao processo são as listadas no Quadro 1. Destacando que as últimas três linhas são referentes ao mesmo expediente, ou seja, um mesmo expediente pode passar por várias situações até a sua conclusão.

Quadro 1: Fontes de dados do Datavisa

Variável	Tipo	Definição
CO_ASSUNTO	Inteiro	Código do Assunto – Tipo de serviço que a empresa está solicitando à Anvisa
DS_ASSUNTO	Texto	Descrição do Assunto – Descrição do Tipo de serviço que a empresa está solicitando à Anvisa
DT_RECEBIMENTO_ANVISA_TOX	Data	Data de entrada do pleito na Anvisa
NU_PROCESSO	Número formatado	Número do processo de Pleito
NU_EXPEDIENTE	Número formatado	Número do Expediente do Processo
NU_CNPJ_EMPRESA	Número formatado	CNPJ da empresa solicitante do Serviço
NO_RAZAO_SOCIAL_EMPRESA	Texto	Nome da empresa solicitante do Serviço
NO_PRODUTO	Texto	Nome do Produto que está pleiteando o serviço
DS_SITUACAO_ASSUNTO_DOC	Texto	Situação do expediente
DT_INICIO_SITUACAO	Data	Data de início da situação do expediente
DT_FIM_SITUACAO	Data	Data de fim da situação do expediente

Fonte: Elaboração do autor

Outras informações secundárias, são as tabelas de informações específicas da GGTOX que traduzem os dados do Datavisa em informações relevantes. São as chamadas tabelas ‘de para’ de Assuntos, de Situação das Fases do Processo e de Critérios das Filas.

⁷<https://www.gov.br/anvisa/pt-br/sistemas/peticionamento>

A tabela auxiliar de assunto trata da principal informação da petição, nela é especificada qual serviço a empresa solicita à Anvisa. Pelo código de assunto pode por exemplo especificar se a empresa deseja registrar um produto agrotóxico novo ou apenas atualizar uma informação de pós-registro. Para sistematizar essas informações de assuntos na GGTOX foi criada essa tabela “de para” que designa qual código de assunto vai ser analisado pelas gerências específicas, por exemplo o assunto 5065 - Produto Formulado com base em Produto Técnico Equivalente, que é um registro de produto foi classificado como Tipo de serviço: 1. Registro e será tratado pela Área: GPREQ – Gerência de Produtos Equivalentes.

Um mesmo código de Assunto, e são 118 entre ativos e inativos, pode ser agrupado em até quatro tipos conforme a informação que o usuário deseja detalhar. Dessa forma uma tabela pode ter níveis diferenciados em hierarquias que facilitam a visualização do serviço. Por exemplo um mesmo código de assunto pode ser visto em um nível mais simples como de registro, registro de produto formulado ou registro de produto formulado de agente biológico de controle, conforme Quadro 02:

Quadro 2: Fontes de dados do Tabela Assuntos

Variável	Tipo	Definição
CO_ASSUNTO	Texto	Código do Assunto – Tipo de serviço que a empresa está solicitando à Anvisa
CO_ASSUNTO_N	Inteiro	Código do Assunto em formato numérico
DS_ASSUNTO_O	Texto	Descrição do Assunto – Descrição do Tipo de serviço que a empresa está solicitando à Anvisa
COD_DESC	Texto	Código do Assunto + Descrição do Assunto
COD_DESC_RESU	Texto	Código do Assunto + Descrição do Assunto Resumida
AREA_GGTOX	Texto	Área da GGTOX responsável pela análise do Serviço: GEAST, GPREQ, GEMAR ou COPSI.
TIPO_PUBLICACAO	Texto	Tipo de serviço: 1. Registro, 2. Pós-registro 3. Outros
TIPO_REGISTRO_0	Texto	Tipo de serviço nível 0: Exemplo 12. Produto Formulado
TIPO_REGISTRO_1	Texto	Tipo de serviço nível 0: Exemplo 122. Produto Formulado de Uso agrícola de Baixo Impacto
TIPO_REGISTRO_2	Texto	Tipo de serviço nível 0: Exemplo 1221. Produtos Biológicos e Orgânicos
TIPO_REGISTRO_3	Texto	Tipo de serviço nível 0: Exemplo Produto Agente Biológico de Controle
SISTEMA	Texto	Sistema em que o serviço é tratado: Datavisa ou Sistema de Peticionamento da Toxicologia (SIPTOX)
PRAZO_336	Texto	Prazo para análise do serviço
SITUAÇÃO	Texto	Situação do código do Assunto: 0-Ativo 1-Inativo
OBS	Texto	Observações sobre a classificação do assunto

Fonte: Elaboração do autor

Outra tabela auxiliar utilizada é a de Situação. Conforme o ciclo de vida do processo, um expediente pode passar por várias situações desde sua entrada até sua conclusão. Por essa tabela por exemplo pode-se identificar se um processo está em andamento ou finalizado, ou ainda especificar se o processo foi indeferido ou deferido para os processos finalizados.

Cada situação foi categorizada em fases para fins de contagem de tempo do processo. Assim, por exemplo se um processo está nas situações “Aguardando análise”, “Aguardando triagem e distribuição”, “Distribuído para a área responsável” ou “Em conferência documental” ele será classificado na fase “Fila”. São cinco as fases resumidas que um processo pode passar: 1. Sobrestado, que é aguardando alguma restrição para poder entrar na fila, 2. Fila, 3.1 Análise, 3.2 Empresa e 4. Recurso. As demais informações da tabela auxiliar de situação podem ser visualizadas no Quadro 3.

Quadro 3: Fontes de dados do Tabela Situação

Variável	Tipo	Definição
DS_SITUACAO_ASSUNTO_DOC	Texto	Situação do Assunto – Tipo de serviço que a empresa está solicitando à Anvisa
STATUS_2_GGTOX	Texto	Situação do Expediente - Em processamento ou Finalizado
STATUS_3_GGTOX	Texto	Resultado do Expediente - Deferido ou Indeferido
CICLO_GGTOX	Texto	Fase do Processo
ORDEM_CICLO	Texto	Ordem da Fase do Processo
STATUS	Texto	Status da Situação: 0-Ativo 1-Inativo

Fonte: Elaboração do autor

Por último a tabela Auxiliar de Critério de Fila que agrupa os tipos de serviços em grupos para análise. Na GGTOX um processo pode aguardar em fila ou lista para ser analisado. A lista é uma situação diferenciada que não segue a ordem cronológica e aguarda a análise de um outro processo. Por exemplo, para os produtos técnicos equivalentes, a Anvisa aguarda a avaliação do MAPA para depois disso, conforme o resultado, analisar a petição colocá-la na fila.

Assim as filas e listas dos serviços da Anvisa obedecem às regras estabelecidas conforme o assunto e a situação da petição. Esses critérios estão dispostos em uma tabela

específica e pode ser consultado nos Dados Abertos⁸. O Quadro 4 detalha as informações dos critérios das filas. Quanto ao tipo de produto será utilizado somente os da Toxicologia.

Quadro 4: Fontes de dados do Critério Fila

Variável	Tipo	Definição
#TIPO_PRODUTO	Texto	Área de interesse, ou tipo de produto: alimentos, autorização de funcionamento, cosméticos etc. Usaremos somente TOXICOLOGIA
TIPO_LISTA_FILA	Texto	Identifica se trata-se da configuração de uma fila ou de uma lista de análise.
FILA	Texto	Nome da fila ou lista de análise configurada, por exemplo: REGISTRO, PÓS-REGISTRO, Alterações, Revalidações...
SUBFILAS	Texto	Nome da Subfila ou sublista de análise configurada, representando subdivisões específicas da fila ou lista, normalmente relativas a diferentes categorias de produto ou assuntos.
COD_ASSUNTO	Texto	Código do assunto configurado para determinada fila ou lista de análise
ASSUNTO	Texto	Descrição do Assunto configurado para determinada fila ou lista de análise
SITUACAO	Texto	Descrição da situação documental configurada para determinada fila ou lista de análise. Exemplo: Aguardando análise, Aguardando inspeção, Anuído etc.

Fonte: Elaboração do autor

3.2 Preparação dos Dados

A Gerência-Geral de Gestão do Conhecimento, Inovação e Pesquisa (GGCIP) é a responsável inteligência de dados da Anvisa, que tem a competência de preparar as informações que estão disponíveis nos sistemas em produção e disponibilizá-las para as áreas fins. Desde o início de 2022, a GGTOX já tem a sua disposição a tabela do Quadro 01 com informações do Datavisa, que está armazenada no SQL Server e pode ser acessada através da ferramenta de ODBC atualizada diariamente. Importante destacar que essa tabela pode ser incrementada com mais informações de interesse, basta solicitar através da Central de Atendimento *servicedesk*.

O pré-processamento consiste em ler essas informações do Datavisa combinando com outras tabelas auxiliares que estarão alocadas no *sharepoint* da Anvisa de maneira que se possa obter as estatísticas descritivas do projeto. Considerando que as inconsistências são normais em

⁸ em <https://dados.gov.br/dados/conjuntos-dados/configuracoes-das-filas-de-analise-da-anvisa>

um banco de dados grande, serão feitos tratamentos que diminuam esses erros para a obtenção de informações confiáveis e consequente geração de outputs que serão consumidos pelo painel.

Os itens abaixo detalham o algoritmo responsável por essa Extração, transformação e carregamento (ETL) das informações do sistema Datavisa e o código completo desse script em python se encontra no Apêndice A.

1. Leitura das tabelas de [PROCESSO], [ASSUNTO], [SITUAÇÃO] e [CRITÉRIO DE FILA].
2. Exclusão de dois 'EXPEDIENTES' com "DATA DE ENTRADA" inválidas como: '1201-11-07 00:00:00.0000000'.
3. Manutenção de "TIPO DE PRODUTO" com valor *missing*, pois será considerado como principal informação o "CODIGO DE ASSUNTO".
4. Manutenção de "CNPJ", "RAZÃO SOCIAL" e "NOME PRODUTO" com valor *missing*, pois não são informações relevantes para as estatísticas e para o modelo.
5. Os campos "DATA FIM SITUAÇÃO" e "DATA PUBLICAÇÃO" em branco serão tratados conforme os itens 10 e 11, abaixo.
6. Criação da informação se o processo entrou em "RECURSO" pelo menos uma vez. Essa informação é importante, pois processos em recurso tem desfechos diferentes dos outros processos e podem contaminar a informação. É definido pela existência de um expediente com código de assunto 5062 em qualquer momento.
7. Criação da informação se o processo foi "CANCELADO" ou paralisado em qualquer momento. Informação também relevante para diferenciar dos outros processos. Considerando "CANCELADO" com processos em pelo menos uma das "SITUAÇÕES": 'Cancelado a pedido da empresa', 'Desistência a pedido', 'Petição encerrada', 'Arquivado', 'Arquivado a pedido'
8. Seleção dos expedientes com "TIPO DE PRODUTO" da tabela de Assunto com valores igual a '1. Registro' ou '2. Pós-Registro'. Nesse filtro foram descartados os expedientes '3. Outros' como 'Auto de Infração Sanitária – GGTOX' ou 'Reconstituição de documentação, pois são documentos que não estão relacionados aos tempos do processo.
9. Identificação da "SITUAÇÃO MAIS RECENTE" dentre todas as datas das "SITUAÇÕES" do expediente. Nos casos em que a última data/hora tem duas situações iguais, causada por erro de banco, será considerada apenas a situação que não tem data final de situação.
10. Identificação da situação do processo com a situação mais recente do expediente.
11. Para os processos finalizados que não tem data de publicação será considerada a data inicial da última situação como data de finalização. Para os que tem data de publicação, essa será considerada como data de finalização.
12. Identificação dos processos finalizados como "Status do processo" = 'Finalizados'
13. Criação da variável "DIAS", que é o número de dias que o expediente ficou na situação. Considerando que em alguns casos não existe a informação da data final da situação e que os tempos de situações diferentes se sobrepõe, será considerada a data de início da próxima situação como a data final da situação atual (lag).
14. Criação de uma situação com o 'Tempo Total' de dias do processo. Diferença entre a data de entrada e a data finalização do expediente. Em alguns casos esse tempo total não bate com a soma dos tempos dos status.
15. OUTPUT = [Andamento]. Expediente com situação em andamento.
16. OUTPUT = [Entrada_DV]. Quantidade de expedientes que entraram por código de assunto e data.

17. OUTPUT = [Saida_DV]. Quantidade de expedientes finalizados por código de assunto e data.
18. OUTPUT = [IndFinal]. Tabela com os processos detalhados por situação e quantidade de dias em que ficou na situação e na fase. Selecionados somente os registros com data de situação menor que a data de finalização.

Esse script em python será programado para ser executado diariamente através da ferramenta do Windows *Task Manager*. Os resultados serão salvos no *sharepoint* da Anvisa e serão atualizados automaticamente pelo Power BI. Para fins de documentação, os resultados apresentados aqui foram congelados com a base de dados atualizada até o dia 30/06/2023, ou seja, final do primeiro semestre.

3.3 Análise Exploratória dos Dados

O cubo de dados gerado com as informações do Datavisa hospedado no servidor da GGCIP tem 244.472 linhas, ou seja, 36.457 processos de registro ou pós-registro com vários expedientes e situações. Desses processos, 290 estiveram em recursos e 2.278 em situação que foi paralisada em qualquer momento como: 'Cancelado a pedido da empresa', 'Desistência a pedido', 'Petição encerrada', 'Arquivado' ou 'Arquivado a pedido'

Considerando que um processo pode ter mais de um expediente, foram 54.962 expedientes que entraram desde o ano de 2.000, sendo que 15.952 estão em andamento e 39.010 estão finalizados. Mais da metade dos processos (55,6%) entraram na Anvisa antes de 2016, conforme Tabela 1:

Tabela 1: Quantidade de expedientes por ano de entrada e situação (2002-2023)

Ano de Entrada	Em Processamento	Finalizados	Total	% Finalizados	% Total Global
Até 2016	3.889	26.662	30.551	87,3%	55,6%
2017	1.026	1.120	2.146	47,8%	3,9%
2018	1.316	1.402	2.718	48,4%	4,9%
2019	1.647	2.037	3.684	44,7%	6,7%
2020	2.289	3.968	6.257	36,6%	11,4%
2021	2.531	2.293	4.824	52,5%	8,8%
2022	2.189	1.205	3.394	64,5%	6,2%
2023	1.065	323	1.388	76,7%	2,5%
Total Global	15.952	39.010	54.962	29,0%	100,0%

Fonte: Datavisa – 30/06/2023

Conforme Tabela 2, do total de expedientes da série histórica, 36.183 (66%) são de registros, sendo que quase metade desses expedientes (48%) são de Registro Especial Temporário, 17.245. Na classe dos registros, mais três tipos se destacam e acabam se sobressaindo sobre os demais: Produto Técnico Equivalente, Produto Formulado com base em Produto Técnico e Registro de Componentes.

Tabela 2: Quantidade de expedientes por tipo de pleito e situação

Tipo de Pleito	Em Processamento	Finalizados	Total	% Total Geral
Registro	7.971	28.212	36.183	66%
111. Produto Técnico	24	1.121	1.145	3%
112. Produto Técnico Equivalente	1.572	2.186	3.758	10%
1211. Produto Formulado com base em Produto Técnico	82	4.551	4.633	13%
1212. Produto Formulado com base em Produto Técnico Equivalente	2.343	1.881	4.224	12%
1213. Produto Formulado Emergencial	-	5	5	0%
1214. Produto Formulado Novo	96	211	307	1%
1221. Produtos Biológicos e Orgânicos	58	467	525	1%
1222. Produto Fitossanitário para Agricultura Orgânica	11	278	289	1%
123. Produto de uso não agrícola	6	46	52	0%
124. Preservativo de Madeira	-	67	67	0%
13. Registro Especial Temporário (RET)	177	17.068	17.245	48%
14. Pré-Mistura	4	19	23	0%
15. Registro para Exportação (REX)	-	83	83	0%
16. Registro de Componentes	3.598	229	3.827	11%
Pós-Registro	7.981	10.798	18.779	34%
2.5 Alteração de pós-registro de componente	1.580	13	1.593	8%
21. Alteração de formulação e reclassificação toxicológica	48	1.492	1.540	8%
22. Inclusão de culturas e alterações relacionadas a dose maior para aplicação, intervalo de segurança, frequência de aplicação e modalidade de emprego	92	2.335	2.427	13%
23. Alterações de Implementação Imediata	4.355	1.530	5.885	31%
24. Outras alterações pós-registro	1.906	5.428	7.334	39%
Total Geral	15.952	39.010	54.962	29,0%

Fonte: Datavisa – 30/06/2023

Quando analisados mais criteriosamente, considerando os processos finalizados, as fases e as datas das situações, alguns expedientes finalizados não passam por críticas simples como data de início da situação menor que a data fim situação. Assim, excluindo esses processos com erros impeditivos, apenas 36.578 podem ser analisados pela ótica da fase do processo. Desse total, por exemplo, apenas 13.585 tem tempo computado pela fase “Análise”, o que configura uma fragilidade da informação no preenchimento do sistema. Por falha do Datavisa, a alteração da situação é uma opção do técnico na análise do processo e é de conhecimento que muitas vezes esse campo somente é preenchido no final da análise, por mais agilidade do técnico no cumprimento da meta de processos a se concluir.

Tabela 3: Quantidade de expedientes por fase do processo e tipo de pleito

Fase do Processo	1. Registro	2. Pós-Registro	Total	% Total Geral
1.Sobrestado	2.709	10	2.719	7,4%
2.Fila	21.346	10.479	31.825	87,0%
3.1.Analise	8.922	4.663	13.585	37,1%
3.2.Empresa	3.135	403	3.538	9,7%
4.Recurso	163	1	164	0,4%
5.Nao_classificado	203	486	689	1,9%
6.Tempo_total	25.957	10.621	36.578	100,0%

Fonte: Datavisa – 30/06/2023

O primeiro *output* gerado que será consumido pelo painel é a lista de processos em andamento. Considerando que os processos de Registro de Componentes 5015 são automaticamente recebidos, a tabela de expedientes em andamento é um pouco menor do que apresentado anteriormente, 12.353 expedientes. Os campos dessa tabela deverão ser filtrados no painel conforme necessidade usuário.

As outras duas tabelas de interesse geradas pelo ETL em python são a quantidade de expediente que entraram e que saíram por código de assunto e por data: ENTRADA_DV e SAIDA_DV. Para poder se relacionar com outras informações no Power BI esses outputs foram agrupados por data, código de assunto e a quantidade, conforme os leiautes da Tabela 4 abaixo.

Tabela 4: Leiaute das tabelas geradas de entrada e saída

ENTRADA_DV (21.648 Linhas)			SAÍDA_DV (6.596 Linhas)		
co_assunto	data_entrada	entrada_dv	co_assunto	data_finalizacao	saida_dv
5000	24-abr-02	1	5000	08-jul-05	3
5000	28-mai-02	1	5000	11-jul-05	1
5000	10-jun-02	1	5000	12-jul-05	1
5000	20-jun-02	1	5000	14-jul-05	3
...
5138	02-jun-23	1	5138	10-mai-23	1
5138	06-jun-23	1	5138	01-jun-23	1

Fonte: Datavisa – 30/06/2023

Por último a tabela com os expedientes, fases do processo e dias em cada fase está no *output* IND_FINAL, como denominação para indicadores final e um resumo das suas informações pode ser consultada na TABELA 05.

Tabela 5: Tabela Indicadores Final (IND_FINAL)

DS_SITUACAO_ASSUNTO_DOC	dt_inicio_situacao	dt_fim_situacao	data_entrada	data_finalizacao	CICLO_GGTOX	dias
Distribuído para a área responsável	17-set-15	23-set-15	08-set-15	23-fev-22	2.Fila	6
Guarda temporária - aguardando análise	23-set-15	05-nov-19	08-set-15	23-fev-22	2.Fila	1504
Aguardando análise	05-nov-19	06-nov-19	08-set-15	23-fev-22	2.Fila	1
Guarda temporária - aguardando análise	06-nov-19	22-fev-21	08-set-15	23-fev-22	2.Fila	474
Aguardando análise	22-fev-21	06-out-21	08-set-15	23-fev-22	2.Fila	226
Em análise	06-out-21	14-dez-21	08-set-15	23-fev-22	3.1.Analise	69
Aguardando análise do cumprimento de exigência	14-dez-21	21-dez-21	08-set-15	23-fev-22	3.1.Analise	7
Em análise do cumprimento de exigência	21-dez-21	04-jan-22	08-set-15	23-fev-22	3.1.Analise	14
Em exigência	04-jan-22	06-jan-22	08-set-15	23-fev-22	3.2.Empresa	2
Aguardando análise do cumprimento de exigência	06-jan-22	13-jan-22	08-set-15	23-fev-22	3.1.Analise	7
Encaminhado para avaliação de outro setor	13-jan-22	23-fev-22	08-set-15	23-fev-22	3.1.Analise	41
	08-set-15	23-fev-22	08-set-15	23-fev-22	6.Tempo_total	2360

Fonte: Datavisa – 30/06/2023

3.4 Modelo de Previsão

A proposta inicial do projeto era criar um modelo que conseguisse estimar os dias em que o expediente passou nas fases “tempo total” de conclusão e “tempo de análise” dado que o serviço fosse solicitado hoje. Considerando o “tempo total” como a diferença em dias entre a data de entrada e a data de conclusão do expediente e nesse tempo estão inclusos todos os tempos das outras fases como o tempo de fila, cumprimento da exigência pela empresa e análise. O “tempo de análise” é o tempo em que o técnico efetivamente trabalhou no processo, ou seja, o tempo entre o recebimento do processo para análise e conclusão do parecer técnico.

Considerando que para concluir um modelo de previsão significativo é necessário ter uma série histórica consistente para a confirmação da existência de tendências, sazonalidade ou ciclos, apesar do banco armazenar conclusões de expedientes desde o ano 2.005, nem todos os assuntos possuem registros suficientes para isso. Como exemplo, quase 40 assuntos têm menos de 100 observações nesse período de 18 anos. Muito se deve a não solicitação do serviço pelas empresas, ou a variação da existência dos assuntos no período, pois muitos serviços são criados e outros extintos. Por exemplo o registro de produtos biológicos e suas classificações são

recentes entre os serviços oferecidos pela Anvisa. A tabela abaixo mostra o número de expedientes solicitado por ano desde 2.005 e nela se constata que somente oito assuntos têm mais de 1.000 observações nesse período. Assim, para a previsão dos tempos foi necessária uma restrição dos serviços que entrarão no modelo.

Tabela 6: Quantidade de expedientes por assunto e ano de finalização (2005 - 2023)

Assunto	Até 2016	2017	2018	2019	2020	2021	2022	2023	TOTAL
5004-Registro Especial Temporário (RET)	13.933	319	334	468	459	843	499	213	17.068
5002-Produto Formulado	4.035	49	76	33	118	87	61	44	4.503
5051-Avaliação Toxicológica para Alteração de Rótulo e Bula	28	3	3	33	2.119	119	66	17	2.388
5041-Produto Técnico Equivalente	677	192	204	282	216	214	280	120	2.185
5048-Avaliação Toxicológica para Alteração de Produto Técnico em Produto Formulado	494	225	191	327	320	491	2	-	2.050
5000-Avaliação Toxicológica para Inclusão de Culturas	675	86	50	159	156	165	283	89	1.663
5065-Produto Formulado com base em Produto Técnico Equivalente	424	206	154	121	183	217	176	111	1.592
5045-Avaliação Toxicológica para Alteração de Formulador/ Manipulador	864	9	1	179	8	1	-	-	1.062
...									
5010-Inclusão de Pragas	1	-	-	-	-	-	-	-	1
5092-Registro simplificado nível I - Produto Técnico Equivalente	-	-	-	-	-	-	1	-	1

Fonte: Datavisa – 30/06/2023

Considerando que os Produtos Técnicos Equivalentes e os Produto Formulado com base em Produto Técnico Equivalente respondem por 88% da fila de espera na data de 30/06/2023 e que os outros serviços, na maioria, são atendidos dentro do prazo estabelecido, será feito um recorte no modelo de previsão descartando esses assuntos que não tem grande demanda e consequentemente sem fila.

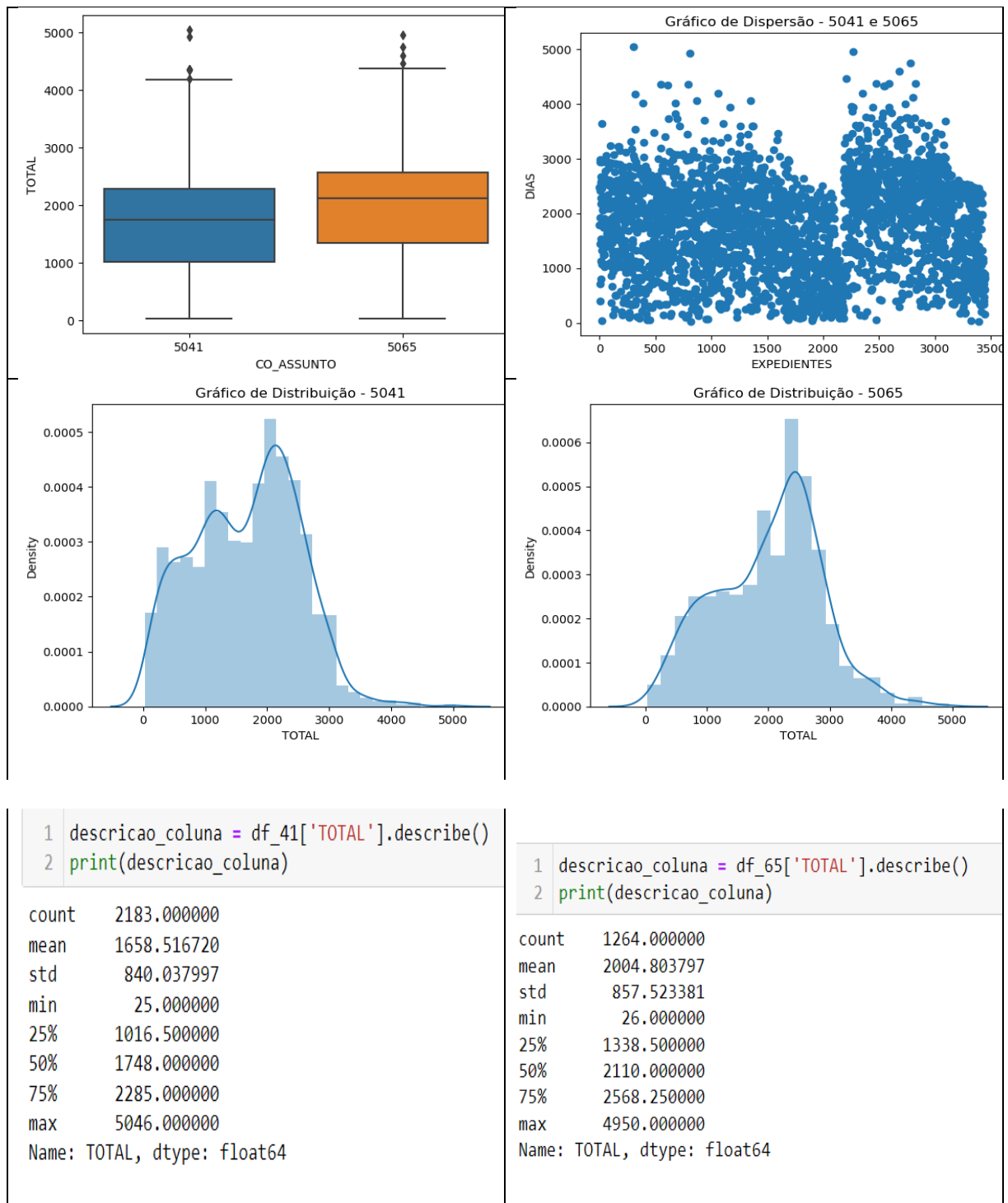
Considerando também que o sistema não guarda todas as fases dos processos e que menos de 38% dos expedientes conseguem ter medidos “tempo de análise”, conforme Tabela 3, não será possível fazer a estimativa de tempo de análise. Portanto, por limitações na base de dados, foram testados os modelos para a previsão somente do “tempo total” dos seguintes assuntos:

- 5041-Produto Técnico Equivalente
- 5065-Produto Formulado com base em Produto Técnico Equivalente

Foram 3.447 expedientes finalizados dos assuntos 5041 e 5065 entre 2010 e 2023. Sendo que o “tempo total” para a conclusão desses expedientes varia entre 25 e 5.046 dias. Um tempo médio de “total de análise” de 1.658 dias para expedientes 5041 e 2.004 dias para

expedientes 5065 com grande variação e outliers conforme pode ser observado no Gráfico 2. Muito das inconsistências dessas informações se deve a falhas do Datavisa e no preenchimento das informações. Esses estudos foram obtidos através do script em python que está no Apêndice B.

Gráfico 2: Estatísticas descritivas do tempo total de conclusão dos expedientes 5041 e 5065

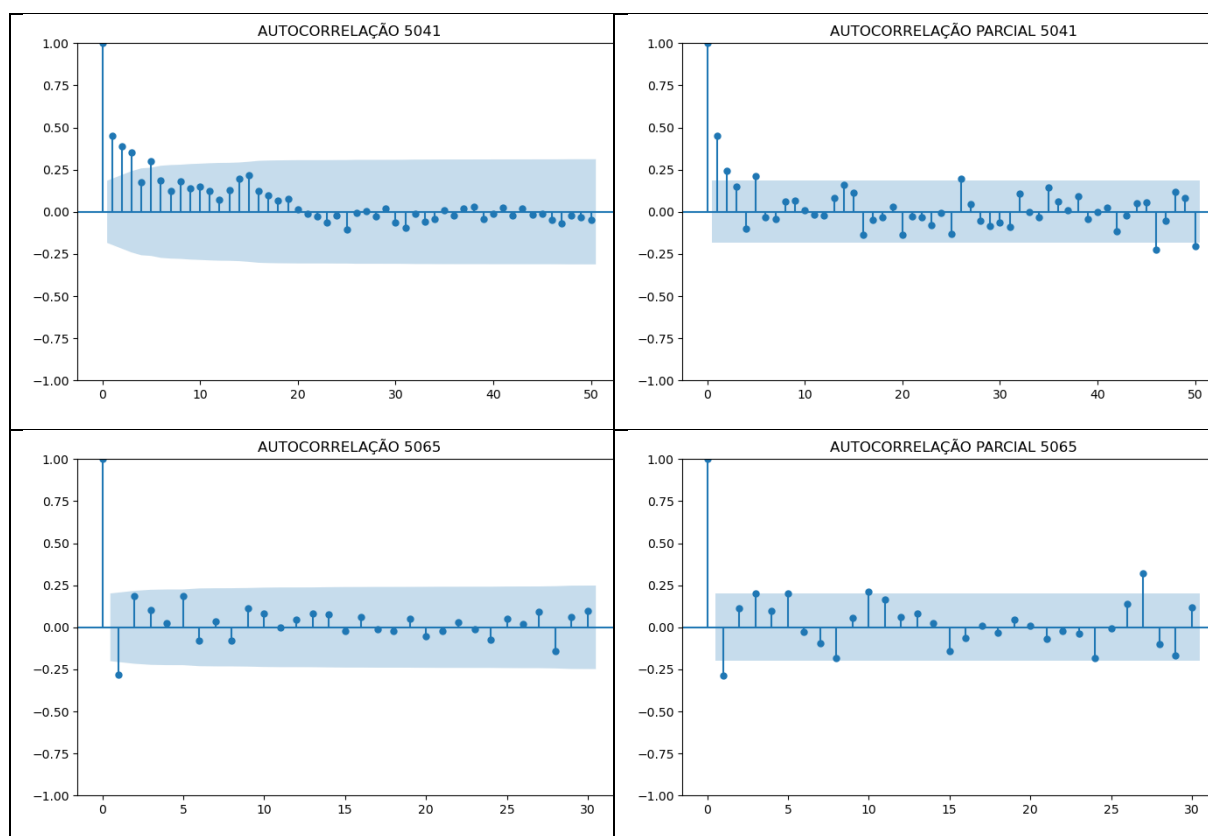


Fonte: Datavisa – 30/06/2023

Para a detecção dos outliers desse “tempo total” foi utilizado a ferramenta de *IsolationForest* da biblioteca Sklearn⁹ no python, que é um modelo de detecção de anomalias baseada em árvore de decisões. Como resultado, mais de 25% das informações foram consideradas como anomalias. Apesar da sugestão de exclusão, esses valores foram mantidos e foram trabalhados manualmente no próprio modelo.

Para a modelagem, os dados foram agrupados por ano e mês e conforme o Teste Dickey-Fuller as séries foram consideradas estacionárias, ou seja, não foram detectadas evidências de tendência ou sazonalidade, com p-value de 0.011 para os expedientes 5041 e p-value 0.0002 para os expedientes 5065. Foram também testadas graficamente as correlações entre um valor em um determinado lag (atraso) e os valores anteriores da série temporal. Esses modelos foram desenvolvidos através do script em python que está no Apêndice C.

Gráfico 3: Autocorrelação e Autocorrelação parcial dos expedientes 5041 e 5065



Fonte: Datavisa – 30/06/2023

⁹ <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.IsolationForest.html>

Considerando que os gráficos de autocorrelação e autocorrelação parcial (Gráfico 3) não são conclusivos foram calculados os critérios de informação AIC (Akaike Information Criterion) e BIC (Bayesian Information Criterion) para os modelos Autorregressivos (AR), de Média móvel (MA), e Modelo de Média Móvel Autorregressiva (ARMA). Conforme a Tabela 7 os modelos Autoregressivos foram que obtiveram os menores valores indicando o melhor ajuste.

Tabela 7: Resultados dos Critérios de Informação por modelo e Assunto

MODELO	5041		5065	
	AIC	BIC	AIC	BIC
Autoregressivos (AR)	1.664	1.675	1.423	1.433
Media móvel (MA)	1.716	1.724	1.467	1.475
Modelo de Média Móvel Autorregressiva (ARMA)	1.677	1.682	1.458	1.463

Fonte: Datavisa – 30/06/2023

3.5 Ferramentas Utilizadas

Para demonstrar como o painel pode ser desenvolvido e para auxiliar outros técnicos que reproduzirão as informações do painel, as ferramentas utilizadas no desenvolvimento do projeto que foram GitHub, Sharepoint, ODBC, Python e o Power BI e o ChatGPT são apresentadas nesse tópico.

O GitHub¹⁰ é uma plataforma aberta na internet que pode ser acessada por qualquer pessoa e serve como hospedagem e compartilhamento de arquivos. Tem a vantagem de poder criar repositórios públicos ou privados e fazer com que grupo de colaboradores possam trabalhar em conjunto obtendo melhores códigos fontes dos arquivos de programação e mostrando exatamente o que mudou a cada versão. Foi criado um repositório público chamado IndicadoresGTOX com acesso em <https://github.com/Alesandre-santos/IndicadoresGGTOX> onde serão disponibilizados as tabelas auxiliares e os scripts em python.

¹⁰ <https://github.com/>

A GGCIP também disponibilizou um ambiente seguro para armazenar, organizar, compartilhar e acessar informações para o desenvolvimento de ciência de dados na Anvisa através do Sharepoint, que é uma ferramenta colaborativa da Microsoft. A pasta privada¹¹ somente pode ser acessada pelos interlocutores de dados da área previamente cadastrados na GGCIP.

Para acessar a tabela principal com as informações do Datavisa no SQL Server será utilizada uma ferramenta de ODBC (acrônimo para Open Database Connectivity) que é um protocolo de conexão a uma fonte de dados externa por vários aplicativos, incluídos o Excel, o Access e o Power BI. A configuração é feita da seguinte forma: Executar ODBC na tela de pesquisar do Windows e adicionar uma nova conexão com nome: BI_GGTOX, descrição: BI_GGTOX e Servidor: anvssdf522. O acesso tem que ser com autenticação do SQL Server usando ID de logon e senha. A senha de acesso ao servidor é disponibilizada pela GGCIP.

Os scripts em python desse projeto foram desenvolvidos no Jupiter do Anaconda. O Python é uma linguagem de programação de alto nível, ou seja, uma linguagem mais próxima da linguagem humana, que pode ser usada em aplicações da Web, desenvolvimento de software, ciência de dados e machine learning (ML). Com o python será dada a carga da tabela final no software que mostrará os dados em tabelas e gráficos. Assim, o python definirá a data de entrada e conclusão dos processos e contará o número de dias que o processo passou em cada fase. Através dos scripts do python também foi possível também criar o modelo de previsão do “tempo total” que será utilizado.

O Power BI é a ferramenta da Microsoft que organiza e interpreta grandes volumes de dados. Com ela é possível extrair os dados com o Power Query e relacionar tabelas, criar novas variáveis e apresentar os dados. Ainda assim foi utilizado o python por ter melhor possibilidade de utilização dos códigos escritos e por apresentar melhor performance. O Power BI foi utilizado no projeto apenas para apresentação dos dados, pois apresenta muitas possibilidades de combinações e nessa ferramenta é possível especificar qual variável filtrar em tabelas e gráficos.

O ChatGPT é um programa de computador que tenta simular um ser humano na conversação com as pessoas e utiliza como base a inteligência artificial através de redes neurais

11

<https://dados.anvisa.gov.br/repdados/DocumentosRepdados/Forms/AllItems.aspx?RootFolder=%2frepdados%2fDocumentosRepdados%2fPOWERBIRS&FolderCTID=0x012000FC8ACDF518531C479E32DB0103134FAB>

e machine learning. No projeto, o chatGPT foi utilizado como auxílio na simplificação dos códigos em python, oferecendo na maioria das vezes os códigos corretos e enxutos.

4 RESULTADOS DO PROJETO

O projeto ainda está sendo desenvolvido, mas já apresenta alguns resultados que acrescentam inovações aos processos de trabalho: expedientes em andamento, estatísticas descritivas, indicadores de performance, expedientes finalizados e previsão do tempo total.

4.1 Expedientes em Andamento

O primeiro painel do dashboard é uma lista com processos que estão em andamento, ou seja, processos que saíram da fila, começaram a ser analisados, mas ainda não foram finalizados. Como foi dito, o Datavisa não gera essa útil informação, pois os técnicos somente conseguem listar os processos que estão na fila, a partir do portal de consultas da Anvisa¹². Depois que esses processos saem da fila não existe um acompanhamento e muitos deles ficam indefinidamente pendentes, ou seja, sem finalização, caso as empresas não avisem à Anvisa.

No painel esses processados podem ser filtrados pelo código de assunto, situação, nome do produto ou empresa, conforme pode ser visto na Figura 2. Podem também ser ordenados pela data de entrada ou pela data da última situação, como no exemplo listado que tem três expedientes de 5041 – Produto Técnico Equivalente que aguardam análise do cumprimento de exigência desde 2020.

¹² <https://consultas.anvisa.gov.br/#/>

Figura 2: Painel de processos em Andamento

ANVISA
GERÊNCIA GERAL DE TOXICOLOGIA

Última Atualização: 29/06/2023

Lista de Processos em Andamento (Não finalizados) - DATAVISA

Tipo: Tipo 1 Tipo 2 Tipo 3 Assunto: 5041-Produto Técnico Equivalente

Expediente: Fase Produto Situação: Aguardando análise do cumprimento de exigên...

data_entrada	co_assunto	nu_processo	nu_expediente	no_produto	ciclo_ggtox	ds_situacao_assunto_doc	dt_inicio_ul_situacao
30/10/17	5041	2535161			3.1.Analise	Aguardando análise do cumprimento de exigência	15/07/2020
20/12/18	5041	2535185			3.1.Analise	Aguardando análise do cumprimento de exigência	18/09/2020
05/11/19	5041	2535163			3.1.Analise	Aguardando análise do cumprimento de exigência	10/02/2021
28/04/16	5041	2535103			3.1.Analise	Aguardando análise do cumprimento de exigência	18/02/2021
30/10/17	5041	2535161			R 3.1.Analise	Aguardando análise do cumprimento de exigência	29/03/2021
03/12/19	5041	2535169			3.1.Analise	Aguardando análise do cumprimento de exigência	18/05/2021
05/08/19	5041	2535145			3.1.Analise	Aguardando análise do cumprimento de exigência	18/05/2021
06/02/17	5041	2535107			3.1.Analise	Aguardando análise do cumprimento de exigência	25/05/2021
31/10/18	5041	2535175			3.1.Analise	Aguardando análise do cumprimento de exigência	25/08/2021
06/05/19	5041	2535126			3.1.Analise	Aguardando análise do cumprimento de exigência	11/04/2022
21/07/14	5041	2535143			3.1.Analise	Aguardando análise do cumprimento de exigência	12/04/2022
08/05/20	5041	2535139			3.1.Analise	Aguardando análise do cumprimento de exigência	26/04/2022
22/12/14	5041	2535100			M 3.1.Analise	Aguardando análise do cumprimento de exigência	01/06/2022
25/05/16	5041	2535109			3.1.Analise	Aguardando análise do cumprimento de exigência	08/07/2022
27/08/20	5041	2535187			3.1.Analise	Aguardando análise do cumprimento de exigência	17/08/2022
29/06/18	5041	2535137			3.1.Analise	Aguardando análise do cumprimento de exigência	31/10/2022
31/07/14	5041	2535145			3.1.Analise	Aguardando análise do cumprimento de exigência	17/03/2023
11/06/19	5041	2535134			3.1.Analise	Aguardando análise do cumprimento de exigência	28/04/2023
27/08/21	5041	2535117			3.1.Analise	Aguardando análise do cumprimento de exigência	02/05/2023
10/10/18	5041	2535171			3.1.Analise	Aguardando análise do cumprimento de exigência	03/05/2023
30/11/18	5041	2535180			3.1.Analise	Aguardando análise do cumprimento de exigência	15/05/2023

Fonte: Datavisa – 30/06/2023

Dos 12.353, quase 8.000 (64,5%) são de expedientes de processos de pós-registro. Alguns desses expedientes são de implementação imediata e muitas vezes os *status* não são finalizados, o que configuraria apenas um erro na atualização do preenchimento do sistema. Mas dos expedientes de registros, Tabela 8, mais de 55% (2.421) estão parados desde 2021. Esse resultado exige uma iniciativa dos gestores para verificar se os erros são apenas de atualização das informações no sistema ou se realmente os processos estão parados há tanto tempo.

Tabela 8: Quantidade de expedientes em andamento por Tipo de produto e ano da última situação (2005 - 2023)

Tipo de Produto	Antes de 2021	2022	2023	Total
Total Registros	2.421	923	1.029	4.373
112. Produto Técnico Equivalente	7	6	11	24
1211. Produto Formulado com base em Produto Técnico	939	355	278	1.572
1212. Produto Formulado com base em Produto Técnico Equivalente	27	25	30	82
1214. Produto Formulado Novo	1.319	497	527	2.343
1221. Produtos Biológicos e Orgânicos	49	28	19	96
1222. Produto Fitossanitário para Agricultura Orgânica	1	8	49	58
123. Produto de uso não agrícola	-	-	11	11
13. Registro Especial Temporário (RET)	3	3	-	6
14. Pré-Mistura	73	-	104	177

Fonte: Datavisa – 30/06/2023

4.2 Estatísticas Descritivas

Nesse painel da Figura 3 é apresentado um gráfico de linha com o número de expedientes que entraram e foram finalizados por assuntos ao longo da série histórica. Os assuntos podem ser filtrados pelo código, unidade mínima, ou também por até quatro tipos de agrupamentos. A linha do tempo pode ser visualizada por data completa, mês trimestre e ano e o período pode ser filtrado por ano, como exemplo 2010 até 2023.

Figura 3: Painel de Estatísticas Descritivas



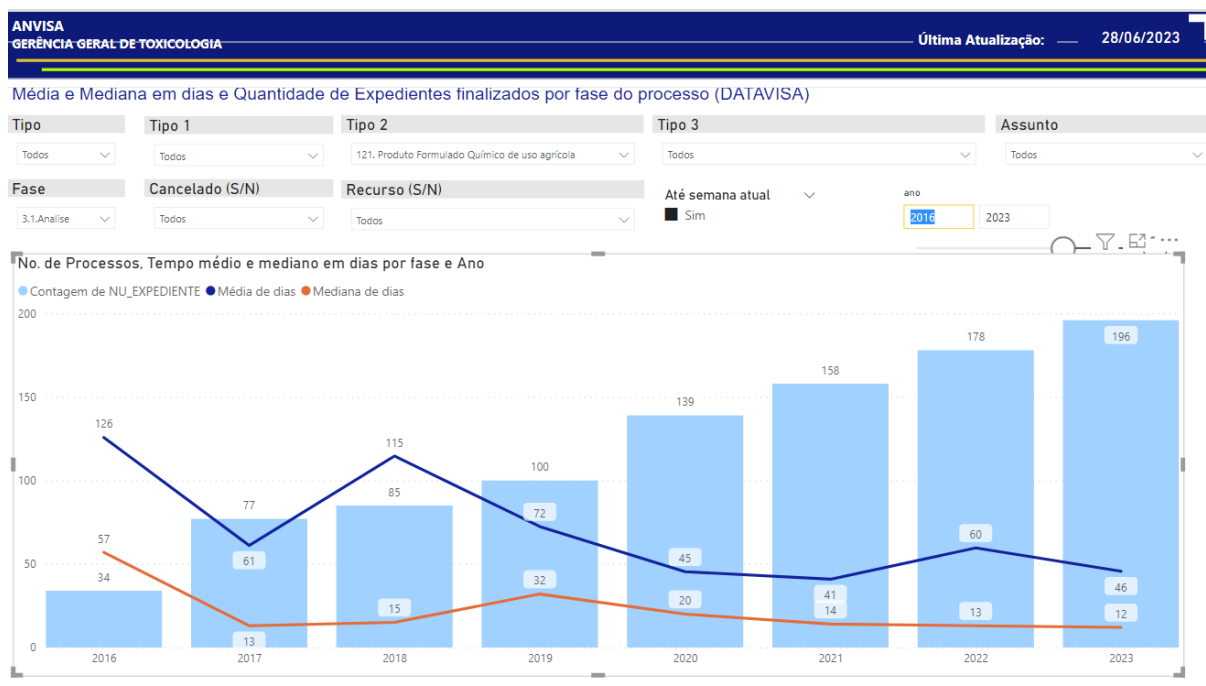
Fonte: Datavisa – 30/06/2023

Com essa ferramenta é possível que o gestor acompanhe semanalmente a produtividade da sua área (“Área GGTOX”), pois através do filtro “Até a semana atual” é possível que se compare períodos iguais, pois não se considerará todas as quantidades dos anos anteriores. No exemplo da figura acima são listados expedientes que foram iniciados e finalizados no primeiro semestre, por ano, desde 2010.

4.3 Indicadores de Performance

Nesse painel é apresentado um gráfico que contém o tempo médio e mediano em dias da fase do processo e também a quantidade de expedientes concluídos no período. Naturalmente podem ser filtrados o tipo de processo e o período da série histórica com a opção de selecionar até a mesma semana dos períodos anteriores.

Figura 4: Painel de Indicadores de Performance



Fonte: Datavisa – 30/06/2023

No Exemplo da Figura 4 é apresentado o tempo médio e mediano em dias da análise dos processos concluídos até o primeiro semestre dos anos entre 2016 e 2023. Esse painel acompanha o Indicador 3.1 do Plano de Gestão Anual da Anvisa, que é aumentar em 30% as decisões em avaliações toxicológicas para fins de registro de agrotóxicos químicos para uso agrícola em relação à média dos últimos 3 anos. Com o acompanhamento até a semana atual é possível ver que a área está cumprindo sua meta e no momento tem concluída mais avaliações que nos últimos três anos.

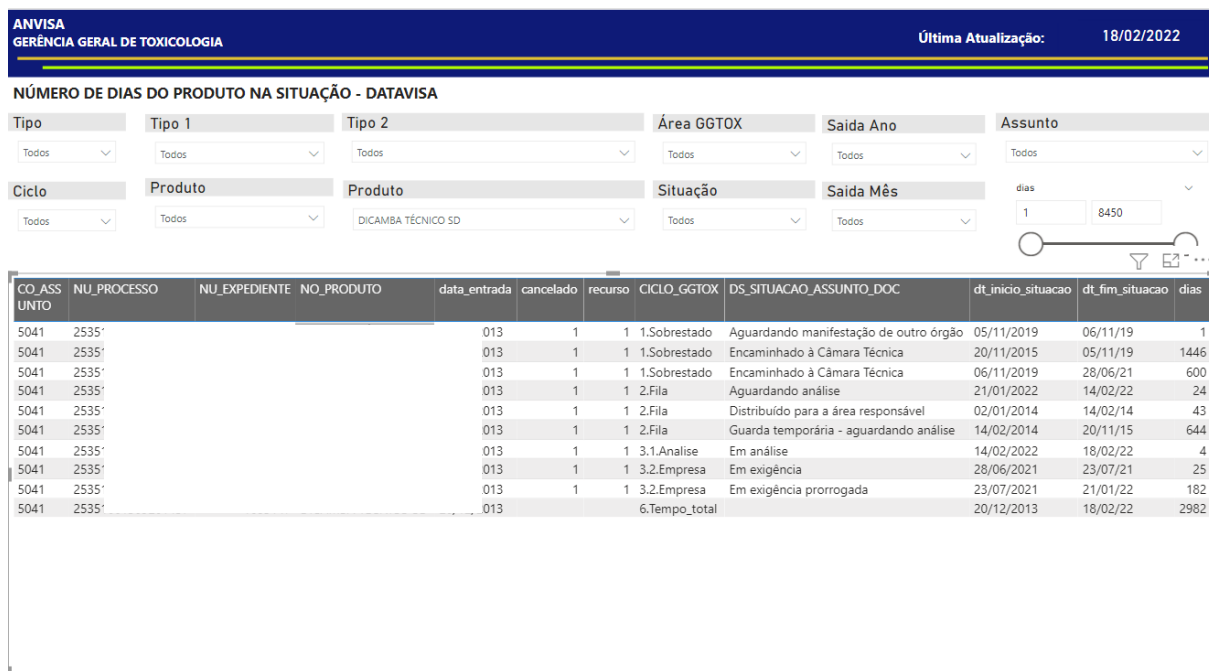
4.4 Expedientes Finalizados

Nesse painel é possível listar os produtos finalizados conforme os filtros: assuntos, área da GGTOX, Ano e mês da conclusão, nome do produto, número do expediente e fase ou situação com a especificação da quantidade de dias. Assim é possível listar os seguintes relatórios, como exemplo:

- Processos de Registro Especial Temporário (5004) finalizados em maio de 2022 que tiveram tempo total de conclusão maior do que 200 dias.
- Processos da Geast de produtos biológicos que ficaram em fila (fase do processo) entre 50 e 100 dias no período de 2020 a 2022;

- Processo que ficaram mais de 100 dias em análise e quais foram as situações desse processo.

Figura 5: Painel de Produtos Finalizados

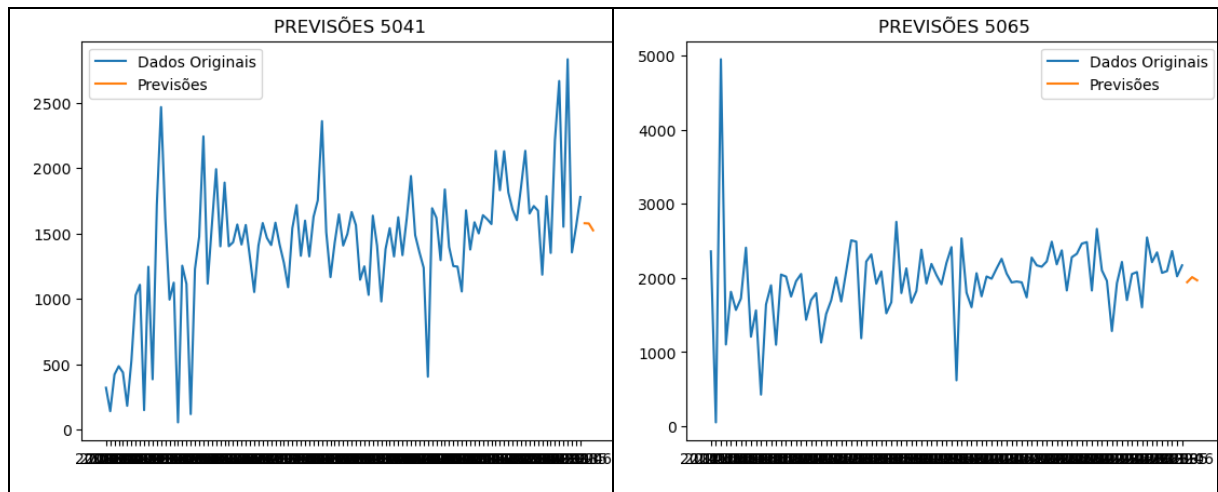


Fonte: Datavisa – 30/06/2023

Esse painel detalha o histórico de quantos dias o processo ficou em cada situação e fase, explicitando aos usuários a metodologia que levou ao cálculo dos indicadores. Serve para os gestores e técnicos criticarem ou justificarem a demora nas fases. Esse retorno da informação aos usuários do sistema é importante para o aperfeiçoamento do método de cálculo em função de como os técnicos utilizam o sistema e também para o desenvolvimento de novos painéis que são de necessidade dos usuários.

4.5 Previsão de Tempo Total

Ainda não foi possível incorporar os dados no modelo de previsão ao painel final mas os resultados já são descritos a partir da modelagem no python. Utilizando o modelo Autoregressivo as previsões para os três meses seguintes, ou seja, julho, agosto e setembro são os seguintes valores: 5041: 1578, 1577 e 1524 e 5065: 1941, 2010 e 1967, conforme Gráfico 4

Gráfico 4: Previsões do tempo total dos expedientes 5041 e 5065

Fonte: Datavisa – 30/06/2023

5 CRONOGRAMA DE IMPLANTAÇÃO DO PROJETO

Considerando que algumas etapas do projeto de intervenção já foram concluídas e outras estão em andamento, o Quadro 5 define o cronograma do projeto com uma entrega parcial no final de 2023 e reservando o ano de 2024 para receber as críticas e sugestões dos técnicos para a melhoria das versões.

Quadro 5: Cronograma do projeto de Intervenção

Etapa do Projeto	2022				2023				2024			
	1T	2T	3T	4T	1T	2T	3T	4T	1T	2T	3T	4T
Definição do Problema												
Desenvolvimento conceitual do projeto												
Compreensão dos dados												
Preparação dos Dados e Tratamento												
Análise Exploratória												
Modelo de Previsão												
Desenvolvimento do Painel												
Automatização dos processos												
Atualização e Melhorias												

Fonte: Elaboração do autor

6 ORÇAMENTO

Os principais custos do projeto que são os recursos com softwares e pessoal já estão incorporados à instituição. Na Gerência-Geral de Conhecimento, Inovação e Pesquisa existe uma contratação de desenvolvimento, manutenção, configuração, implantação, operação e sustentação de soluções relacionadas a Business Intelligence (Bi), Analytics, Big Data, Ciência de Dados e gestão de dados no Valor de R\$ 10.778.076,80. Está em processo de contratação, ainda sem valor definido, o software estatístico SAS, que pode auxiliar futuramente o tratamento de dados desse projeto. O software Anaconda, em que os scripts do python são rodados, pode ser utilizado em uma versão sem custos para a instituição. A Anvisa já dispõe de 200 licenças do Power BI Pro e 280 do Premium que foram negociadas em um pacote mais amplo com a Microsoft.

Os custos com pessoal na GGTOX serão com a alocação de dois servidores, responsáveis para a manutenção e atualização dos painéis. Esses servidores são os interlocutores de dados da área, que fazem parte da política de gestão do conhecimento, mas também continuam com suas atribuições da área, ou seja, o servidor interlocutor de dados é o mesmo quem faz a publicação dos produtos aprovados no Diário Oficial da União (DOU). Portanto não existe como distinguir um custo específico do projeto, pois muitos desses recursos já estão alocados e concorrem com outras atividades.

7 CONSIDERAÇÕES FINAIS

Desde 2016, com as mudanças na gestão visando a melhoria dos processos de trabalho, já está em andamento a implementação da cultura de dados na GGTOX. Uma parte desse acompanhamento da produtividade era feita a partir dos expedientes publicados no DOU em planilhas Excel. Esse processo manual implicava em retrabalho e informações poucos confiáveis e limitadas, pois somente trazia conclusões de processos publicáveis. Assim, o conceito desse projeto já existe há algum tempo como uma necessidade da área e das instâncias supervisoras: Assessoria de Planejamento (APLAN) e a Terceira Diretoria (DIRE3). O avanço desse projeto de intervenção consiste na criação de uma metodologia transparente de classificação dos eventos e na automatização na criação dessas informações com a mínima intervenção para evitar retrabalho e erros.

Os resultados desse projeto já são consumidos conforme os painéis listados e a cada dia cresce o interesse dos técnicos e gestores em acessar mais informações, proporcionando uma espiral desenvolvedora em que os usuários recebem a informação, questionam os dados, aperfeiçoam o modelo e solicitam mais informações. Esse é o princípio da cultura da informação em que cada colaborador é agente dessa transformação. Assim a implementação desses painéis de informação, ao longo desses dois anos, é exitosa e merece incentivo da gestão. Ainda assim, existem pontos que merecem atenção para o pleno sucesso do projeto como as barreiras tecnológicas e a cultura organizacional resistente ao fluxo livre das informações.

Uma das barreiras tecnológicas que existem na Anvisa é defasagem dos computadores pessoais que são extremamente lentos e não permitem o manuseio de grande quantidade de dados. Muitas vezes é mais produtivo para o técnico trabalhar de casa com a VPN do que trabalhar no espaço físico da Anvisa. Comum também ver servidores trabalharem com notebooks pessoais na Anvisa e afirmarem ser impossível trabalhar com os computadores locais. Apesar dos computadores não serem tão antigos, a lentidão na maioria das vezes é causada pela ação do antivírus e justificada na GGTIN pela segurança das informações.

Outra dificuldade tecnológica é a impossibilidade de utilizar e testar novos softwares diferentes dos estabelecidos pela Anvisa. Até mesmo softwares livres como R não são incentivados pela GGTIN. Como exemplo, em fevereiro de 2023 foi solicitada a instalação do Anaconda e há mais de cinco meses essa solicitação está pendente sem nenhuma resposta (Anexo 1). Em alguns casos até os computadores pessoais utilizados no teletrabalho da Anvisa pelos servidores tem os softwares livres bloqueados pelo antivírus quando instalada a VPN.

Quanto à cultura organizacional, a proibição do acesso das áreas fins, que são as responsáveis pelos sistemas, às tabelas com os seus dados é o maior entrave à formação de uma cultura de gestão baseada em dados. Atualmente quem faz a interlocução com a área de infraestrutura de banco de dados e de sistemas é a GGCIP, mas nem mesmo essa área tem apoio técnico e nem sempre consegue consolidar as informações. Um exemplo é a própria GGCIP ser incapaz de conseguir replicar internamente uma informação que é pública e ter que recorrer a um processo automatizado para retirar dados do portal da Anvisa. Em outra ocasião, a GGTOX solicitou acesso ao modelo de dados do Datavisa e foi informada pela área de Administração de Dados que essa informação é restrita e não pode ser divulgada. Esses comportamentos destoam da política de dados abertos implementada pelo governo federal e estão em descompasso com a Lei de Acesso à informação. Somente as áreas fins conhecem os dados e

regras de seus negócios e somente elas podem traduzir e compilar a veracidade das informações. Além da segurança, outra justificativa para a impossibilidade de acesso aos dados é a complexidade dos sistemas, por exemplo o Datavisa chegar a ter mais de mil tabelas, mas essa justificativa não é razoável, pois existem outros sistemas de exclusividade da GGTOX que são mais novos e mais simples como Sistema de Peticionamento da Toxicologia, Cadastro de Estudos de Resíduos e o Sistema de Gerenciamento de Amostras do PARA que também tem seu acesso negado.

Os Avanços da Anvisa na área de Governança de Dados são notórios é merecem destaque a própria criação da GGCIP que promove a cultura da informação na Anvisa diariamente, e teve como ações a criação dos interlocutores de dados e essa capacitação em Ciência de Dados. Mas é importante destacar que esse modelo de intermediação entre as áreas fins e os dados dos sistemas não é funcional. De um lado são centenas de áreas ávidas por informações e do lado da GGCIP são poucos técnicos para viabilizarem esses pedidos. O resultado é que nem todas as demandas são atendidas em tempo e comprometem a qualidade dos serviços da Anvisa. Como exemplo, foi solicitado com urgência um cubo de dados do SISGAP, que é o Sistema que Gerencia as Amostras do PARA (Anexo 2), para o desenvolvimento de relatórios gerenciais em setembro de 2022 e até hoje o pedido não foi concluído.

A gerência de toxicologia tem um corpo técnico altamente capacitado com mestres e doutores com conhecimentos dificilmente encontrados na sociedade em geral. A análise toxicológica de um produto técnico novo, por exemplo, leva quase um ano de estudos com pesquisas intensas equiparadas a uma tese de doutorado. Essa especificidade das análises faz com que muitos técnicos não reconheçam a importância do correto preenchimento das informações no sistema que consideram um trabalho menor e menos importante do que o parecer físico. Para alguns, o preenchimento das informações corretas no sistema serve como controle burocrático e retrabalho que pouco agrega à área. Por outro lado, a grande demanda de avaliações de produtos obriga os técnicos a analisarem muitos processos mensalmente para cumprir a meta institucional definida pelo teletrabalho sem perder a qualidade. Como consequência desse foco dos servidores nas análises e pelo obsoleto Datavisa não obrigar o preenchimento dos campos temos baixa qualidade dos dados e informações ruins. Isso pode ser verificado nos resultados do projeto com poucas informações aproveitáveis dos tempos de fase de cada processo.

Em resumo, o projeto é bom e pode ser melhorado desde que sejam oferecidas condições tecnológicas como computadores novos e configurações de rede que não comprometam a agilidade das máquinas, liberdade para instalação de softwares livres e acesso pleno das áreas aos seus dados em tabelas e servidores de consulta com uma governança de dados estabelecida e transparentes para todos os atores da instituição.

Quanto à resistência dos servidores, é importante implementar uma política de conscientização da importância do preenchimento das informações no sistema e o retorno pleno dessas informações para os técnicos. Além disso o novo SIA que está em fase final de desenvolvimento também ajudará na transformação digital da GGTOX. Uma informação importante que futuramente pode ser trabalhada no painel é o técnico responsável pela análise do processo. Essa é uma informação que envolve características pessoais e que pode trazer uma animosidade na equipe. Mas, se bem trabalhada pelos gestores e técnicos pode ser utilizada no futuro como ferramenta de avaliação de desempenho ou até mesmo como relatório de cumprimento de metas da Gestão Orientada para Resultados (PGOR), que é o Teletrabalho do Anvisa.

REFERÊNCIAS

ANTUNES, D. **Balanced Scorecard e Dashboard Estratégico: Caso da FersiReis,Lda**. Dissertação de Mestrado – Instituto Universitário de Lisboa. Lisboa, p. 95. 2020

AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA. **Relatório de Atividades GGTOX**. 2016. Brasília, DF, Disponível em: < <https://www.gov.br/anvisa/pt-br/centraisdeconteudo/publicacoes/agrotoxicos/publicacoes/relatorio-de-atividades-ggtox-2016.pdf/view> > Acesso em 25 jul. 2023.

BRASIL. **Decreto nº 4.074, de 4 de janeiro de 2002**. Diário Oficial da República Federativa do Brasil, Poder Executivo, Brasília, DF, 4 jan 2002. Disponível em: < http://www.planalto.gov.br/ccivil_03/decreto/2002/d4074.htm > Acesso em 24 jul. 2023.

_____. **Lei nº 7.802, de 11 de julho de 1989**. Diário Oficial da República Federativa do Brasil, Poder Executivo, Brasília, DF, 12 jul 1989. Disponível em: <http://www.planalto.gov.br/ccivil_03/LEIS/L7802.htm> Acesso em 24 jul. 2023.

CARVALHO, Rafael de C.; MELO, Claudia de O. **Tomada de decisão baseada em dados: avaliando a visualização de informação em dash boards**. In: WORKSHOP DE INICIAÇÃO CIENTÍFICA EM SISTEMAS DE INFORMAÇÃO - SIMPÓSIO BRASILEIRO DE SISTEMAS DE INFORMAÇÃO (SBSI), Porto Alegre: Sociedade Brasileira de Computação, 2018.

LUCCHESI, G. **Agrotóxicos – construção da legislação**. 2005, Biblioteca digital da Câmara dos Deputados. Disponível em:< <https://bd.camara.leg.br/bd/handle/bdcamara/2227> >Acesso em: 24 jul. 2023.

PELAEZ, V. & TERRA, F. h. b & SILVA, L. R. **A regulamentação dos agrotóxicos no Brasil: entre o de mercado e a defesa da saúde e do meio ambiente**. Revista de Economia, v. 36, n. 1 (ano 34), p. 27-48, jan./abr. 2010. Editora UFPR.

ROSA DOS SANTOS, G. **Características, Sistema de Registros de Produtos e Concorrência no Mercado de Agrotóxicos no Brasil**. Radar: Tecnologia, Produção e Comércio Exterior: n. 20, jun. 2012. Brasil. Instituto de Pesquisa Econômica Aplicada (Ipea)

SANTOS, A. N. **Desenvolvimento de cultura organizacional data-driven: uma visão dos profissionais de agências de publicidade**. Dissertação de Mestrado – Universidade Católica de Brasília. Brasília, p. 97. 2021.

TRIBUNAL DE CONTAS DA UNIÃO, **Relatório de auditoria operacional. Fragilidades estruturais, no controle interno e no fluxo de trabalho, relativo à emissão de informe de avaliação toxicológica**. 2013. Disponível em: <<https://portal.tcu.gov.br/data/files/20/24/80/9F/63D0E410504D80E42A2818A8/011.726%202013-0%20%20Anvias%20informe%20agrot%3%b3xicos.pdf>> Acesso em 25 jul. 2023.

APÊNDICES

Apêndice A – Script em python: Extração dos dados

Extração e Tratamento dos Dados

In [2]:

```
1 #pip install pyodbc
2 # importar bibliotecas
3 import pandas as pd
4 import numpy as np
5 import io
6 import pyodbc
7
```

In [2]:

```
1 # Conexão direta com o banco SQLServer Importação da tabela de expedientes Situações
2 conn = pyodbc.connect('Driver={SQL Server};Server=anvssdf522;UID=alesandre.santos;PW
3 query = 'SELECT * FROM ta_historico_peticao'
4 df = pd.read_sql_query(query, conn)
5 #df = df0.head(1000)
```

...

In [13]:

```
1 # Leitura do arquivo congelado em 01/07/2023
2 df = pd.read_excel(r'C:\Users\alesa\OneDrive - ANVISA - Agencia Nacional de Vigilanc
3                      header=0, sheet_name="Planilha1")
4
```

In [3]:

```
1 #df.head()
2 #df.info()
3 #base_a = base_a.dropna()
4 #base_a.info()
5 #base_a.isnull().sum()
6 #base_a.head()
```

In [9]:

```

1 # Importação das tabelas auxiliares e seleção de colunas
2 assuntos = pd.read_excel(r'C:\Users\alesa\OneDrive - ANVISA - Agencia Nacional de Vi
3                             header=0, sheet_name="ASSUNTOS")
4 status = pd.read_excel(r'C:\Users\alesa\OneDrive - ANVISA - Agencia Nacional de Vi
5                             header=0, sheet_name="STATUS_NIVEL_2")
6
7 colunas_para_excluir0 = ['ContarDeCO_ASSUNTO2', 'MáxDaData_SITUACAO', 'STATUS_3_GGTO
8                             'ciclo' ]
9 status = status.drop(colunas_para_excluir0, axis=1)
10
11 crit_fila = pd.read_excel(r'C:\Users\alesa\OneDrive - ANVISA - Agencia Nacional de
12                             header=0, sheet_name="FI
13 crit_fila = crit_fila[['CO_ASSUNTO', 'DS_SITUACAO_ASSUNTO_DOC', 'TIPO_LISTA_FILA', 'FI
14 #crit_fila.info()

```

In [16]:

```

1 # Transformando datas do formato texto para o formato data e excluindo expediente co
2 df = pd.DataFrame(df[(df['NU_EXPEDIENTE'] != 977988114) &
3                     (df['NU_EXPEDIENTE'] != 36605116)])
4 df['DT_ENTRADA'] = pd.to_datetime(df['DT_RECEBIMENTO_ANVISA_TOX'], format="%Y-%m-%d %
5 df['DT_INICIO_SITUACAO'] = pd.to_datetime(df['DT_INICIO_SITUACAO'], format="%Y-%m-%d
6 df['DT_FIM_SITUACAO'] = pd.to_datetime(df['DT_FIM_SITUACAO'], format="%Y-%m-%d %H:%M:
7 df['DT_PUBLICACAO'] = pd.to_datetime(df['DT_PUBLICACAO'], format="%Y-%m-%d %H:%M:%S.%
8 df['DT_CARGA'] = pd.to_datetime(df['DT_CARGA'], format="%Y-%m-%d %H:%M:%S.%f")
9 df['CO_ASSUNTO'] = df['CO_ASSUNTO'].astype('int64')
10 #df.info()

```

In [17]:

```

1 # Define a função para formatar o CNPJ e Expediente
2 def formatar_cnpj(NU_CNPJ_EMPRESA):
3     return "{}.{}.{} / {}-{}".format(NU_CNPJ_EMPRESA[:2], NU_CNPJ_EMPRESA[2:5], NU_CNPJ_EMPRESA[5:8], NU_CNPJ_EMPRESA[8:11], NU_CNPJ_EMPRESA[11:14])
4
5 def formatar_exped(NU_EXPEDIENTE):
6     return "{} / {}-{}".format(NU_EXPEDIENTE[:7], NU_EXPEDIENTE[7:8], NU_EXPEDIENTE[8:11])
7
8 # Cria a nova coluna com o CNPJ formatado
9 df['NU_CNPJ_EMPRESA'] = df['NU_CNPJ_EMPRESA'].astype('str')
10 df['NU_CNPJ_EMPRESA'] = df['NU_CNPJ_EMPRESA'].apply(formatar_cnpj)
11
12 df['NU_EXPEDIENTE'] = df['NU_EXPEDIENTE'].astype('str')
13 df['NU_EXPEDIENTE_f'] = df['NU_EXPEDIENTE'].apply(formatar_exped)
14
15 df['NU_EXPEDIENTE'] = df['NU_EXPEDIENTE'].astype('int64')
16
17 #df.head()

```

In [18]:

```
1 # Criando df de Recursos
2 df_r = pd.DataFrame((df[(df['CO_ASSUNTO'] == 5062)]))
3
4 # Criando df de Cancelados
5 df_c = pd.DataFrame(df[(df['DS_SITUACAO_ASSUNTO_DOC'] == "Cancelado a pedido da empr
6                        (df['DS_SITUACAO_ASSUNTO_DOC'] == "Desistência a pedido") |
7                        (df['DS_SITUACAO_ASSUNTO_DOC'] == "Petição encerrada") |
8                        (df['DS_SITUACAO_ASSUNTO_DOC'] == "Arquivado") |
9                        (df['DS_SITUACAO_ASSUNTO_DOC'] == "Arquivado a pedido"))])
10
11 #df_c.head()
12 #Agrupando df de Recursos e Cancelados por Número do Processo
13 df_ra = df_r.groupby('NU_PROCESSO').size().reset_index(name='RECURSO')
14 df_ca = df_c.groupby('NU_PROCESSO').size().reset_index(name='CANCELADO')
```

In [19]:

```
1 # Relacionando Tabela principal com tabela de Recursos e Cancelados
2 df_001 = df.merge(df_ra, on="NU_PROCESSO", how="left").merge(df_ca, on="NU_PROCESSO")
3
4 # Relacionando Tabela principal com tabela de Assuntos
5 df_002 = df_001.merge(assuntos, on="CO_ASSUNTO", how="left")
6
7 # Exclui colunas que não são de interesse
8 colunas_para_excluir = ['Fato_Gerador', 'obs', 'Tipo_agrofit', 'DT_RECEBIMENTO_ANVI
9                        'prazo_336']
10 df_002 = df_002.drop(colunas_para_excluir, axis=1)
11
12 # Seleciona somente expedientes de Registro e Publicação
13 df_002 = pd.DataFrame(df_002[(df_002['TIPO_PUBLICACAO'] == "1. Registro") |
14                             (df_002['TIPO_PUBLICACAO'] == "2. Pós-Registro")])
15
16 #df_002.info()
```

In [22]:

```

1  # Criando a informação de ultima situação no processo do expediente mais recente
2  ult_status = df_002.groupby(['NU_PROCESSO', 'NU_EXPEDIENTE'])['DT_INICIO_SITUACAO'].
3
4  base_dv_0 = df_002.merge(ult_status, on=['NU_PROCESSO', 'NU_EXPEDIENTE', 'DT_INICIO_
5
6  ult_status_r = base_dv_0.groupby(['NU_PROCESSO', 'NU_EXPEDIENTE', 'DT_INICIO_SITUACAO
7
8  ult_status_r = pd.DataFrame(ult_status_r[(ult_status_r['QTD_ERRO'] >= 2)])
9
10 base_dv_1 = base_dv_0.merge(ult_status_r, on=['NU_PROCESSO', 'NU_EXPEDIENTE', 'DT_IN
11 base_dv_1['MARCADOR'] = np.where((base_dv_1['QTD_ERRO'] >= 2) & (base_dv_1['DT_FIM_S
12 base_dv_1 = pd.DataFrame(base_dv_1[(base_dv_1['MARCADOR'] == 0)])
13 base_dv_1['ULT_SIT'] = base_dv_1['DS_SITUACAO_ASSUNTO_DOC']
14 base_dv_1['DT_INICIO_UL_SITUACAO'] = base_dv_1['DT_INICIO_SITUACAO']
15
16 base_dv_1 = base_dv_1[['NU_PROCESSO', 'NU_EXPEDIENTE', 'ULT_SIT', 'DT_INICIO_UL_SITUACAO
17 #base_dv_1.info()
18 df_003 = df_002.merge(base_dv_1, on=['NU_PROCESSO', 'NU_EXPEDIENTE'], how="left")
19 df_004 = df_003.merge(status, left_on="ULT_SIT", right_on="DS_SITUACAO_ASSUNTO_DOC",
20
21 df_004 = df_004.drop(['DS_SITUACAO_ASSUNTO_DOC_y'], axis=1)
22 df_004['DS_SITUACAO_ASSUNTO_DOC'] = df_004['DS_SITUACAO_ASSUNTO_DOC_x']
23 df_004['STATUS_2_GGTOX_P'] = df_004['STATUS_2_GGTOX']
24
25 df_004 = df_004.drop(['DS_SITUACAO_ASSUNTO_DOC_x', 'STATUS_2_GGTOX', 'CICLO_GGTOX', 'c
26
27 df_005 = df_004.merge(status, left_on="DS_SITUACAO_ASSUNTO_DOC", right_on="DS_SITUACAO
28 df_005['data_p_nula'] = df_005['DT_PUBLICACAO'].isnull().astype(int)
29
30 # Se Finalizado e data de publicação = nulo então data finalização = data situação e
31 df_005['DT_FINALIZACAO'] = np.where((df_005['data_p_nula'] == 1) & (df_005['STATUS_2
32                                     df_005['DT_INICIO_UL_SITUACAO'], df_005['DT_PUBL
33 df_005['DATA_FINALIZACAO'] = df_005['DT_FINALIZACAO'].dt.date
34 df_005['DATA_ENTRADA'] = df_005['DT_ENTRADA'].dt.date
35 df_005 = df_005.drop(['data_p_nula'], axis=1)
36
37 df_006 = df_005.merge(crit_fila, on=['CO_ASSUNTO', 'DS_SITUACAO_ASSUNTO_DOC'], how="
38 df_006 = pd.DataFrame(df_006[(df_006['ULT_SIT'] == df_006['DS_SITUACAO_ASSUNTO_DOC']
39                             (df_006['DT_INICIO_UL_SITUACAO'] == df_006['DT_INICIO_S

```

In [23]:

```
1 # Criando outputs: Andamento, Saída e Entrada
2
3 ANDAMENTO = pd.DataFrame(df_006[(df_006['STATUS_2_GGTOX_P'] == "Em Processamento") &
4                                (df_006['CO_ASSUNTO'] != 5015) &
5                                (df_006['ULT_SIT'] == df_006['DS_SITUACAO_ASSUNTO'] &
6                                (df_006['DT_INICIO_UL_SITUACAO'] == df_006['DT_IN
7
8 FINALIZADOS = pd.DataFrame(df_006[(df_006['STATUS_2_GGTOX_P'] == "Finalizados") &
9                                (df_006['ULT_SIT'] == df_006['DS_SITUACAO_ASSUNTO'] &
10                                (df_006['DT_INICIO_UL_SITUACAO'] == df_006['DT_IN
11
12 SAIDA_DV = FINALIZADOS.groupby(['CO_ASSUNTO', 'DATA_FINALIZACAO']).size().reset_index
13
14 ENTRADA_DV = df_006.groupby(['CO_ASSUNTO', 'DATA_ENTRADA']).size().reset_index(name=''
```

In [24]:

```

1  # Criando outputs: Indicadores finalizados (Situações, fases, dias)
2
3  IND_FINALIZADOS = pd.DataFrame(df_005[(df_005['STATUS_2_GGTOX_P'] == "Finalizados")])
4
5  IND_FINALIZADOS = IND_FINALIZADOS.sort_values(by=['NU_PROCESSO', 'NU_EXPEDIENTE', 'CO_ASSUNTO'],
6          ascending=[True, True, True, False, True])
7
8  IND_FINALIZADOS['DT_FIM_SITUACAO_LAG'] = IND_FINALIZADOS['DT_INICIO_SITUACAO'].shift(1)
9
10 IND_FINALIZADOS = pd.DataFrame(IND_FINALIZADOS[(IND_FINALIZADOS['DT_INICIO_SITUACAO'] < IND_FINALIZADOS['DT_FIM_SITUACAO_LAG'])])
11
12 IND_FINALIZADOS['DIAS'] = np.where((IND_FINALIZADOS['NU_PROCESSO'] == IND_FINALIZADOS['NU_EXPEDIENTE'] &
13     (IND_FINALIZADOS['NU_EXPEDIENTE'] == IND_FINALIZADOS['CO_ASSUNTO']) &
14     IND_FINALIZADOS.apply(lambda row: (row['DT_FIM_SITUACAO'] - row['DT_INICIO_SITUACAO']).days > 0)),
15     1, 0)
16 IND_FINALIZADOS = pd.DataFrame(IND_FINALIZADOS[(IND_FINALIZADOS['DIAS'] > 0)])
17
18 IND_FINALIZADOS['CICLO_GGTOX'] = np.where((IND_FINALIZADOS['CICLO_GGTOX'].isnull()),
19     IND_FINALIZADOS['SAIDA_A'] - IND_FINALIZADOS['DT_FINALIZACAO'].dt.year,
20     IND_FINALIZADOS['SAIDA_M'] - IND_FINALIZADOS['DT_FINALIZACAO'].dt.month,
21     IND_FINALIZADOS['ENTRADA_A'] - IND_FINALIZADOS['DT_ENTRADA'].dt.year,
22     IND_FINALIZADOS['ENTRADA_M'] - IND_FINALIZADOS['DT_ENTRADA'].dt.month,
23     0)
24 # Criando Fase '6. Tempo Total'
25 IND_FINALIZADOS_T = IND_FINALIZADOS.groupby(['SAIDA_A', 'SAIDA_M', 'ENTRADA_A', 'ENTRADA_M', 'CO_ASSUNTO', 'NU_PROCESSO', 'NU_EXPEDIENTE']).agg({'DIAS': 'sum'})
26
27
28
29
30
31 IND_FINALIZADOS_T['DT_INICIO_SITUACAO'] = IND_FINALIZADOS_T['DT_ENTRADA']
32 IND_FINALIZADOS_T['DT_FIM_SITUACAO'] = IND_FINALIZADOS_T['DT_FINALIZACAO']
33 IND_FINALIZADOS_T['CICLO_GGTOX'] = '6.Tempo_total'
34
35 IND_FINALIZADOS_T['DIAS'] = IND_FINALIZADOS_T.apply(lambda row: (row['DT_FINALIZACAO'] - row['DT_INICIO_SITUACAO']).days, axis=1)
36
37 IND_FINAL = pd.concat([IND_FINALIZADOS_T, IND_FINALIZADOS])
38 IND_FINAL = pd.DataFrame(IND_FINAL[(IND_FINAL['DIAS'] > 0)])
39 IND_FINAL_1 = IND_FINAL[['SAIDA_A', 'SAIDA_M', 'ENTRADA_A', 'ENTRADA_M', 'CO_ASSUNTO', 'NU_PROCESSO', 'NU_EXPEDIENTE', 'CICLO_GGTOX', 'DIAS', 'DT_FINALIZACAO', 'DT_ENTRADA', 'DT_INICIO_SITUACAO', 'NO_RAZAO_SOCIAL_EMPRESA']]
40
41 #IND_FINAL.info()
42

```

In [30]:

```

1  # Exportando saidas para Excel
2
3  ANDAMENTO.to_excel('ANDAMENTO.xlsx', sheet_name='Planilha1', index=False)
4  SAIDA_DV.to_excel('SAIDA_DV.xlsx', sheet_name='Planilha1', index=False)
5  ENTRADA_DV.to_excel('ENTRADA_DV.xlsx', sheet_name='Planilha1', index=False)
6  IND_FINAL_1.to_excel('IND_FINAL.xlsx', sheet_name='Planilha1', index=False)

```


Apêndice B – Script em python: Tratamento e análise descritiva

1 # Estatísticas Descritivas e Tratamento dos Dados

In [1]:

```
1 #pip install pyodbc
2 # importar bibliotecas
3 import pandas as pd
4 import numpy as np
5 import io
```

2.1 ESTATISTICAS DESCRITIVAS ¶

In [2]:

```
1 #Importa arquivo final do tratamento de dados
2 IND_FINAL = pd.read_excel(r'C:\Users\alesa\OneDrive - ANVISA - Agencia Nacional de
3                          header=0, sheet_name="Planilha1")
4 assuntos = pd.read_excel(r'C:\Users\alesa\OneDrive - ANVISA - Agencia Nacional de Vi
5                          header=0, sheet_name="ASSUNTOS")
```

In [4]:

```
1  # Selecionando variáveis de interesse
2  IND_FINAL_0 = IND_FINAL[['SAIDA_A',
3                          #'SAIDA_M',
4                          #'ENTRADA_A',
5                          #'ENTRADA_M',
6                          'CO_ASSUNTO',
7                          #'NU_PROCESSO',
8                          'NU_EXPEDIENTE',
9                          'RECURSO',
10                         'CANCELADO',
11                         'DT_FINALIZACAO',
12                         'DT_ENTRADA',
13                         #'DT_INICIO_SITUACAO',
14                         #'DT_FIM_SITUACAO',
15                         'CICLO_GGTOX',
16                         'DIAS',
17                         #'NU_CNPJ_EMPRESA',
18                         #'NO_RAZAO_SOCIAL_EMPRESA',
19                         #'NO_PRODUTO',
20                         #'TIPO_PUBLICACAO',
21                         #'DT_FIM_SITUACAO_LAG'
22                        ]]
23
24  IND_FINAL_0 = pd.DataFrame(IND_FINAL_0[(IND_FINAL_0['CICLO_GGTOX'] == "3.1.Analise")
25                                         (IND_FINAL_0['CICLO_GGTOX'] == '6.Tempo_total')])
26
27  IND_FINAL_0 = pd.DataFrame(IND_FINAL_0[(IND_FINAL_0['CICLO_GGTOX'] == "6.Tempo_total")
28                                         (IND_FINAL_0['CICLO_GGTOX'] == '6.Tempo_total')])
29
30  IND_FINAL_0 = pd.DataFrame(IND_FINAL_0[(IND_FINAL_0['CO_ASSUNTO'] == 5065) |
31                                         (IND_FINAL_0['CO_ASSUNTO'] == 5041)])
32
33  # Transformando Recurso e Cancelado maior que 1 em 1
34  IND_FINAL_0['RECURSO'] = np.where((IND_FINAL_0['RECURSO'] >= 1), 1, 0)
35  IND_FINAL_0['CANCELADO'] = np.where((IND_FINAL_0['CANCELADO'] >= 1), 1, 0)
36
37  IND_FINAL_0.info()
```

...

In [5]:

```
1 # Transpor os dados e gerando as duas bases de 5041 e 5065
2
3 df_transposed = IND_FINAL_0.pivot_table(index=['CO_ASSUNTO', 'NU_EXPEDIENTE', 'RECURSO',
4         'DT_ENTRADA', 'DT_FINALIZACAO'],
5         columns='CICLO_GGTOX',
6         values='DIAS').reset_index()
7
8 # Renomear as colunas resultantes
9 df_transposed.columns.name = None
10
11 # Renomear as colunas adicionando um prefixo
12 df_transposed = df_transposed.rename(columns=lambda x: '' + str(x))
13
14 df_transposed.rename(columns={'3.1.Analise': 'ANALISE'}, inplace=True)
15 df_transposed.rename(columns={'6.Tempo_total': 'TOTAL'}, inplace=True)
16
17 assunto_tipo = assuntos[['CO_ASSUNTO', 'TIPO_PUBLICACAO']]
18
19 df_transposed = df_transposed.merge(assunto_tipo, on="CO_ASSUNTO", how="left")
20 df_transposed = pd.DataFrame(df_transposed[(df_transposed['TIPO_PUBLICACAO'] == "1.
21 df_transposed = df_transposed.drop(['TIPO_PUBLICACAO'], axis=1)
22 #df_transposed.head()
23 #df_transposed.info()
24
25 df_41 = pd.DataFrame(df_transposed[(df_transposed['CO_ASSUNTO'] == 5041)])
26 df_65 = pd.DataFrame(df_transposed[(df_transposed['CO_ASSUNTO'] == 5065)])
27
28 df_41.head()
```

...

In [6]:

```
1 #Verificando nulos
2 df_transposed.isnull().sum()
```

...

In [7]:

```
1 import matplotlib.pyplot as plt
2 import seaborn as sns
```

In [8]:

```
1 #Boxplot por categoria
2 sns.boxplot(x="CO_ASSUNTO",
3             y="TOTAL",
4             data=df_transposed)
5 plt.show()
```

...

In [9]:

```
1 sns.distplot(df_41['TOTAL'])
2 plt.title('Gráfico de Distribuição - 5041')
3 plt.show()
```

...

In [10]:

```
1 sns.distplot(df_65['TOTAL'])
2 plt.title('Gráfico de Distribuição - 5065')
3 plt.show()
```

...

In [12]:

```
1 # Gráfico de Dispersão
2
3 # Crie o gráfico de dispersão
4 plt.scatter(df_transposed.index, df_transposed['TOTAL'])
5
6 # Adicione rótulos aos eixos x e y
7 plt.xlabel('EXPEDIENTES')
8 plt.ylabel('DIAS')
9
10 # Adicione um título ao gráfico
11 plt.title('Gráfico de Dispersão - 5041 e 5065')
12
13 # Exiba o gráfico
14 plt.show()
```

...

In [19]:

```
1 # Gráfico de Dispersão
2
3 # Crie o gráfico de dispersão
4 plt.scatter(df_65.index, df_65['TOTAL'])
5
6 # Adicione rótulos aos eixos x e y
7 plt.xlabel('EXPEDIENTES')
8 plt.ylabel('DIAS')
9
10 # Adicione um título ao gráfico
11 plt.title('Gráfico de Dispersão - 5065 ')
12
13 # Exiba o gráfico
14 plt.show()
15
```

...

In [20]:

```
1 descricao_coluna = df_41['TOTAL'].describe()
2 print(descricao_coluna)
```

...

In [21]:

```
1 descricao_coluna = df_65['TOTAL'].describe()
2 print(descricao_coluna)
```

...

Detecção de anomalias - Outliers

In [22]:

```
1 #outliers semana 12 - detecção de anomalias
2 #importar bibliotecas
3 from sklearn.ensemble import IsolationForest
```

In [23]:

```
1 #df_transposed = pd.DataFrame(df_transposed[(df_transposed['CO_ASSUNTO'] == 5041)])
2
3 #Definir modelo de Isolation Forest - SOMENTE TOTAL
4 modelo=IsolationForest(n_estimators=100,max_samples='auto',random_state=0)
5 #Visualizar parâmetros do modelo
6 print(modelo.get_params())
7
8 variaveis_anomalia = ['TOTAL']
9
10 #Ajustar modelo multivariado
11 modelo.fit(df_transposed[variaveis_anomalia])
12
13 #Criar coluna chamada score
14 df_transposed['scores'] = modelo.decision_function(df_transposed[variaveis_anomalia])
15
16 df_transposed.info()
```

...

In [24]:

```
1 #Definir modelo de Isolation Forest - CO_ASSUNTO e TOTAL
2 modelo=IsolationForest(n_estimators=100,max_samples='auto',random_state=0)
3 #Visualizar parâmetros do modelo
4 print(modelo.get_params())
5
6 colunas_analise = ['TOTAL']
7
8 #Ajustar modelo multivariado
9 modelo.fit(df_transposed[colunas_analise])
10
11 #Criar coluna chamada score
12 df_transposed['scores'] = modelo.decision_function(df_transposed[colunas_analise])
13
14 df_transposed.head()
15
16
```

...

In [28]:

```
1 df_transposed.to_excel('anomalias.xlsx', sheet_name='Planilha1', index=False)
```

In []:

```
1 # Exportando arquivo final para a modelagem
2 #TESTE.to_excel('TESTE.xlsx', sheet_name='Planilha1', index=False)
3 df.to_excel('df.xlsx', sheet_name='Planilha1', index=False)
```

Apêndice C – Script em python: Modelo de predição

Modelo Previsão Tempo Total

In [1]:

```
1 #pip install pyodbc
2 # importar bibliotecas
3 import pandas as pd
4 import numpy as np
5 import io
```

In [2]:

```
1 # Tratamento da base: agrupando a serie por ano e mês
2 df_0 = pd.read_excel(r'C:\Users\alesa\OneDrive - ANVISA - Agencia Nacional de Vig
3                      header=0, sheet_name="Planilha1")
4
5 df_0['DATA_FINALIZACAO'] = df_0['DT_FINALIZACAO'].dt.date
6 df_0['DATA_FINALIZACAO_AM'] = df_0['DT_FINALIZACAO'].dt.strftime('%Y-%m')
7
8 df_41 = pd.DataFrame(df_0[(df_0['CO_ASSUNTO'] == 5041)])
9 df_65 = pd.DataFrame(df_0[(df_0['CO_ASSUNTO'] == 5065)])
10
11 #df = df_0.groupby(['CO_ASSUNTO', 'CO_ASSUNTO_T', 'DATA_FINALIZACAO'])['TOTAL'].mean()
12 df = df_0.groupby(['CO_ASSUNTO', 'CO_ASSUNTO_T', 'DATA_FINALIZACAO_AM'])['TOTAL'].n
13 df41 = df_41.groupby(['CO_ASSUNTO', 'CO_ASSUNTO_T', 'DATA_FINALIZACAO_AM'])['TOTAL'].n
14 df65 = df_65.groupby(['CO_ASSUNTO', 'CO_ASSUNTO_T', 'DATA_FINALIZACAO_AM'])['TOTAL'].n
15
16 #df65.head()
17
```

Séries Temporais

In [3]:

```
1 # Verificando a estacionaridade da série
2
3 from statsmodels.tsa.stattools import adfuller
4 import matplotlib.pyplot as plt
5
```

In [4]:

```
1 #result = adfuller(df['TOTAL'])
2 result41 = adfuller(df41['TOTAL'])
3 result65 = adfuller(df65['TOTAL'])
4
5 # Extraia os resultados do teste
6 #p_value = result[1]
7 p_value41 = result41[1]
8 p_value65 = result65[1]
9
10
11 # Imprima os resultados
12 #print(f"Valor p: {p_value}")
13 print(f"Valor p41: {p_value41}")
14 print(f"Valor p65: {p_value65}")
15
```

Valor p41: 0.011022706431745681

Valor p65: 0.0002491862947946675

In [5]:

```
1 # Verificar a estacionariedade com valor-p
2 #if p_value < 0.05:
3 #     print("A série é estacionária.")
4 #else:
5 #     print("A série não é estacionária.")
6
7 if p_value41 < 0.05:
8     print("A série41 é estacionária.")
9 else:
10     print("A série41 não é estacionária.")
11
12 if p_value65 < 0.05:
13     print("A série65 é estacionária.")
14 else:
15     print("A série65 não é estacionária.")
```

A série41 é estacionária.

A série65 é estacionária.

In [6]:

```
1 ##### Gráficos ACF e PACF
2 from statsmodels.graphics.tsaplots import plot_acf
3 from statsmodels.graphics.tsaplots import plot_pacf
```

In [7]:

```
1 # Definir a coluna ano como índice
2 df41.set_index('DATA_FINALIZACAO_AM',inplace=True)
3 # Alterar o nome do índice de Any para ano
4 # Alterar o nome do índice de Any para ano
5 df41.index.name = 'AnoMes'
6 df41 = df41[['TOTAL']]
```

In [8]:

```
1 # Definir a coluna ano como índice
2 df65.set_index('DATA_FINALIZACAO_AM', inplace=True)
3 # Alterar o nome do índice de Any para ano
4 # Alterar o nome do índice de Any para ano
5 df65.index.name = 'AnoMes'
6 df65 = df65[['TOTAL']]
7 #df65.head()
```

In [11]:

```
1 # Verificando Autocorrelação
2 fig, ax = plt.subplots(figsize=(8,5))
3 plot_acf(df65['TOTAL'], lags=30, ax=ax)
4 plt.title('AUTOCORRELAÇÃO 5065')
5 plt.show()
6
```

...

In [12]:

```
1 # Verificando Autocorrelação Parcial
2 fig, ax = plt.subplots(figsize=(8,5))
3 plot_pacf(df65['TOTAL'], lags=30, ax=ax)
4 plt.title('AUTOCORRELAÇÃO PARCIAL 5065')
5 plt.show()
```

...

Modelo Autoregressivo (AR)

In [21]:

```
1 # AUTOREGRESSIVO
2 import pandas as pd
3 from statsmodels.tsa.ar_model import AutoReg
```

In [24]:

```
1 ar_modelo = AutoReg(df65[['TOTAL']], lags=2).fit()
2 ar_modelo.summary()
```

...

In [25]:

```
1 # Previsão do Modelo
2 pred = ar_modelo.predict(start=len(df65), end=len(df65)+2, dynamic=False)
3 pred
```

...

In [26]:

```

1  # Gráfico de previsão
2
3  # Ajustar o modelo ARIMA aos dados
4  #model = ARIMA(df41['TOTAL'], order=order)
5  #model_fit = model.fit()
6
7  ar_modelo = AutoReg(df41[['TOTAL']], lags=2).fit()
8  #modelo_ma = sm.tsa.arima.ARIMA(df65[['TOTAL']], order=(0,0,1)) # Ordem do modelo MA
9
10 # Fazer previsões
11 n_steps = 3 # Número de passos futuros a serem previstos
12 forecast = ar_modelo.forecast(steps=n_steps)
13 #forecast1 = modelo_ma.forecast(steps=n_steps)
14
15 # Plotar os dados originais e as previsões
16 plt.plot(df41['TOTAL'], label='Dados Originais')
17 #plt.plot(np.arange(len(df65['TOTAL']), len(df65['TOTAL']) + n_steps), forecast, for
18 plt.plot(np.arange(len(df41['TOTAL']), len(df41['TOTAL']) + n_steps), forecast, label='Previsões')
19 plt.title('PREVISÕES 5041')
20 plt.legend()
21 plt.show()

```

...

Modelo de Médias Moveis (MA)

In [18]:

```
1 import statsmodels.api as sm
```

In [19]:

```

1 #MODELO MA
2 modelo_ma = sm.tsa.arima.ARIMA(df65[['TOTAL']], order=(0,0,1)) # Ordem do modelo MA
3 resultado_ma = modelo_ma.fit()
4 print(resultado_ma.summary())

```

...

Modelo de Média Movel Autoregressivo (ARMA)

In [90]:

```

1 #MODELO ARMA
2 modelo_ma = sm.tsa.arima.ARIMA(df65[['TOTAL']], order=(0,1,1)) # Ordem do modelo MA
3
4 resultado_ma = modelo_ma.fit()
5 print(resultado_ma.summary())

```

...

ANEXOS

ANEXO 1 – Solicitação de Instalação Anaconda

432765 - Detalhes da solicitação

[Adicionar comentário](#)

[Anexar documento](#)

Data/hora da abertura	Status	Prioridade	Área da solicitação	Área do solicitante
10/02/2023 15:15:08	Aberto	3-BAIXA	Tecnologia da Informação.Equipamentos de TI.Estação de Trabalho.Solicitar instalação de aplicativo	GGTOX
Previsão do início do atendimento			Previsão do fim do atendimento	
Descrição da solicitação				
Prezados, Solicito a instalação do software livre Anaconda para trabalho com ciência de dados. Instalar no computador da Anvisa. Segue autorização do gestor,				
Propriedades				
Nome		Valor		Exemplo
Histórico				
Contato	Data	Tipo		Resumo
Arcanjo Xavier Marti	10/02/2023 15:27:07	Transferir		Transferir Grupo de 'Triagem Suporte' para 'Servidores da TI
Santos, Alesandre	10/02/2023 15:22:24	Vincular documento		Vincular documento
Santos, Alesandre	10/02/2023 15:15:08	Inicial		Contact: Santos, Alesandre Telefone: (61) 3462-6787 Email: Alesandre.Santos@anvisa.gov.br
Anexos				
Documento	Descrição		Anexado em	Status
anaconda.pdf			10/02/2023 15:21:44	Instalado

ANEXO 2 – Solicitação de desenvolvimento Cubo PARA

415540 - Detalhes da solicitação

Editar

Criar requisição de mudança

Criar incidente

Criar solicitação

Solicitante	Usuário final afetado	Categoria da solicitação	Status	Prioridade
	Reis, Arthur	Gestão do Conhecimento, Inovação e Pesquisa.Inteligência de Dados.Desenvolvimento	Em andamento	3-BAIXA

Detalhes

Relatado por	Sistema	Grupo	Responsável		
Santos, Alesandre		N3 Infra Projetos - BI	Anvisa, BI BI		
Data/hora do início do atendimento	Data/hora do fim do atendimento	Ticket de sistema externo	Área do solicitante	RDM	Multiplicador
			GGTOX		

Informações resumidas

Resumo	Tempo total de atividade
P: SISGAP	00:10:26
Descrição	
Prezados,	
Solicito a criação de uma view no BI_TOXICOLOGIA com as informações do SISGAP que gerencia as amostras do Programa de Análise de Resíduos de Agrotóxicos em Alimentos (PARA), Esse cubo é necessário pois trataremos as informações diariamente com o início das respostas do programa e é impossível solicitar os relatórios à GGTIN todos os dias. Com esse procedimento de trabalho ganhamos tempo e evitamos o retrabalho.	
São dois relatórios que já foram feitos pela AD/GGTIN e as queries podem ser requisitadas através dos CAs: - Planilha de Gerenciamento: CA 374347 e 372286 - Planilha de detecções: CA 366505	
Estamos à disposição para qualquer informação adicional	
Data/hora da abertura	Última modificação em
12/09/2022 09:37:44	25/07/2023 07:55:29
Data/hora da resolução	Data/hora de fechamento

1 . Informações adicionais

2 . Logs

3 . Gerenciamento de conhecimento

4 . Relacionamentos

1. Propriedades

2. Anexos

3. Tipo de serviço

4. Ciclo de Vida

Propriedades

Nome	Valor	Exemplos
	Não há propriedades definidas para este(a) Área de incidente/problema/solicitação	