# SUMÁRIO

[SUMÁRIO 13](#_Toc140184042)

[1. INTRODUÇÃO 16](#_Toc140184043)

[1.1 Registro de Agrotóxicos 17](#_Toc140184044)

[1.2 Registro de Agrotóxicos na Anvisa 18](#_Toc140184045)

[1.3 Situação Atual 20](#_Toc140184046)

[1.4. A intervenção 21](#_Toc140184047)

[1.5 Importância do Projeto 23](#_Toc140184048)

[2. OBJETIVOS 24](#_Toc140184049)

[2.1 Objetivo Geral 24](#_Toc140184050)

[2.2 Objetivos Específicos 24](#_Toc140184051)

[3. DESENVOLVIMENTO DO PROJETO 24](#_Toc140184052)

[3.1 Coleta de Dados 26](#_Toc140184053)

[3.1 Datavisa 26](#_Toc140184054)

[3.1.2 Tabelas Auxiliares 28](#_Toc140184055)

[3.2 Preparação dos Dados 31](#_Toc140184056)

[3.3 Análise Exploratória dos Dados 33](#_Toc140184057)

[3.4 Modelo de Previsão 36](#_Toc140184058)

[3.5 Ferramentas Utilizadas 40](#_Toc140184059)

[4. RESULTADOS DO PROJETO 42](#_Toc140184060)

[4.1 Expedientes em Andamento 42](#_Toc140184061)

[4.2 Estatísticas Descritivas 44](#_Toc140184062)

[4.3 Indicadores de Performance 45](#_Toc140184063)

[4.4 Expedientes Finalizados 46](#_Toc140184064)

[5. CRONOGRAMA DE IMPLANTAÇÃO DO PROJETO 47](#_Toc140184065)

[6. ORÇAMENTO 47](#_Toc140184066)

[7. CONSIDERAÇÕES FINAIS 48](#_Toc140184067)

[REFERÊNCIAS 49](#_Toc140184068)

[APÊNDICES 50](#_Toc140184069)

[ANEXOS 51](#_Toc140184070)

# INTRODUÇÃO

Por ser um produto que traz grandes riscos à saude da população e ao meio ambiente, os agrotóxicos exigem uma regulação do governo 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, que coincide com redemocratização do país, sensível à questão ambiental, assegura que “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” (Artigo 225 da Constituição Federal).

A Lei 7.802/1.989 trouxe grandes avanços para a regulação do mercado de agrotóxicos no Brasil e a chamada Lei dos Agrotóxicos é considerada avançada em relação a outros países. 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 opções inovadoras e mais baratas que gera em perda de competitividade e tentam desde o ano 2.000 flexibilizar a lei.

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

Esse projeto de intervenção cria um dashboard 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 de 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 Anvisa, qual a situação atual dos registros na Anvisa, a proposta de intervenção e a importância desse projeto.

## 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, 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 agronômica 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 Informações Fitossanitários (AGROFIT) do Ministério da Agricultura, Pecuária e Abastecimento (MAPA), em 2023, são mais de 500 ingredientes ativos registrados no país com cerca de 3.215 produtos formulados, que em 2017 comercializaram 720.87 toneladas de agrotóxicos, tornando o Brasil um dos maiores consumidores mundiais desse produto entre 1990 e 2020 segundo dados da FAO.

Fonte: <https://indicadores.agricultura.gov.br/agrofit/index.htm>

<https://www.gov.br/ibama/pt-br/assuntos/quimicos-e-biologicos/agrotoxicos/relatorios-de-comercializacao-de-agrotoxicos>

https://www.fao.org/faostat/en/?#data/RP/visualize

## Registro de Agrotóxicos na Anvisa

A Gerência-Geral de Toxicologia (GGTOX) da Anvisa é a responsável pelo registro de agrotóxico representando o Ministério da Saúde. A GGTOX existe desde a criação da Anvisa e sua estrutura é composta de três gerências e conta com um quadro de servidores que oscilou entre quinze e vinte e quatro técnicos direcionados para as análises toxicológicas ao longo desses anos. Como comparação, observa-se que a divisão de agrotóxicos da Agência de Proteção Ambiental dos EUA possui cerca de oitocentos funcionários em seus quadros (FONTE).

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 (FONTE).

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.

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 artigo publicado pelo Revista Radar nº 20 de Junho de 2012 - Características, Sistema de Registros de Produtos e Concorrência no Mercado de Agrotóxicos no Brasil de Gesmar Rosa dos Santos já demonstrava o tamanho do problema com os registros de agrotóxicos: Em março de 2012, havia 1.040 pedidos de registro, sendo a capacidade estimada de 150 processos por ano (considerando-se o caso da Anvisa, agência onde o processo tem maior complexidade).

Em 2016, houve uma reestruturação na GGTOX que foi o ponto de partida uma de uma gestão motivada a simplificar os processos de trabalho e melhorar a gestão e a eficiência. Tais mudanças envolveram autoavaliação e análise de gestores, servidores, instituições governamentais, fabricantes de produtos agrotóxicos e representantes da sociedade. Esses esforços trouxeram bons resultados, 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%.

~~A GGTOX ainda agilizou suas análises com a instituição em 2017, do procedimento simplificado de avaliação toxicológica para o registro e alterações de pós-registro de produtos técnicos (SIPTOX) e o desenvolvendo a Ferramenta de Leitura Otimizada no Registro de Agrotóxicos (FLORA).~~ Mas, apesar desses esforços 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 como consequência a fila de registro só aumenta conforme pode ser observado no **Gráfico 1**.

Gráfico 1: Evolução de Número de processos que entram que saem e a fila (2003-2022)

Gráfico, Gráfico de linhas

Descrição gerada automaticamente

Fonte: Datavisa

## Situação Atual

Atualmente cerca de 3.600 processos de avaliação toxicológica aguardam análise da Anvisa 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 todo a regulação do mercado de agrotóxicos.

Em alguns casos, os processos de agrotóxicos chegam a esperar por mais de dez 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 ao judiciário 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 processos que geraram cerca de 82.816 documentos relacionados. Alguns desses documentos chegam facilmente a 5.000 folhas (**Figura 01**) 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. Mais ainda, é impossível trabalhar sem transformar esses dados em informações relevantes que tragam conhecimento para a área. Hoje qualquer instituição que almeja sucesso no seu objetivo tem uma necessidade de transformar dados em conhecimento.

Figura 1: Tamanho físico de um processo de Avaliação Toxicológica

Uma imagem contendo no interior, cozinha, mesa, balcão

Descrição gerada automaticamente

Fonte: Registro do autor

## 1.4. A intervenção

O 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 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, sendo os dados mais confiáveis a partir de 2016, ou ainda 276 meses, em quatro áreas e dois sistemas de origem distintos 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 **dashboard**. 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 continua 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 (Parmenter, 2007).

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 um 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 podendo, por exemplo, reorganizar as 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 do 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.

Mas o grande passo em direção à transformação digital no registro de agrotóxicos se dará com a implementação de uma cultura *data-driven* entre gestores e servidores da área com as decisões sendo tomada com base em dados. Berntsson Svensson e Taghavianfar (2020) sustentam que cultura data-driven é aquela em que as decisões da organização são baseadas em informações, dados e estatísticas, significando dizer que a cada dia a intuição vem dando lugar às ações apoiadas por números e, assim, apontando para um caminho mais promissor.

## 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.

Dificilmente a Anvisa vai conseguir aumentar seu quadro de servidores, mas as atribuições legais vão continuar as mesmas obrigando que os gestores tenham soluções inovadoras fazendo mais com menos. As soluções tecnológicas são as mais utilizadas, portanto essa ferramenta pode auxiliar os gestores na avaliação de novas soluções tecnologias. Dialogar com o setor e com a alta gerência. Quantos servidores são necessários? / Mais eficiência fazer mais com menos

Importante também destacar que ao oferecer essa ferramenta aos técnicos e gerentes, retornando informações uteis do dia-a-dia, instala-se uma cultura da informação, reforçando o papel de cada ator no processo de trabalho, aumentado 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 que a GGTOX cumpra as normas quanto ao prazo estabelecido para as avaliações toxicológicas. Garantir o cumprimento dessas normas é o início para garantir uma regulação mais eficaz nesse mercado tão importante para a saúde da população.

Apesar de nos últimos anos o número de registros aprovados ter crescido substancialmente, essa fila é uma das justificativas para a flexibilização do registro de agrotóxicos. O Projeto de Lei 6.299/2002, recentemente aprovado na Câmara Federal e que aguarda votação no Senado, tira competências da Anvisa e do Ibama na aprovação do registro, concedendo registro automaticamente para qualquer substância que aguardar mais de 2 anos pelo laudo final.

Por outro lado, os efeitos adversos dos agrotóxicos na saúde da população humana e no meio ambiente despertam a atenção do tema para os mais diversos setores da sociedade. Além da mídia, Organizações Não Governamentais (ONG’s) ligadas ao meio-ambiente e pesquisadores acadêmicos alertam para o fato do governo brasileiro registrar excessivamente agrotóxicos, muitas vezes não permitidos em outros países, desprezando critérios científicos.

# OBJETIVOS

## 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.

## 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 de processos 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 outras pessoas possam reproduzir o painel futuramente, inclusive em outras áreas de registro da Anvisa

# DESENVOLVIMENTO DO PROJETO

Baseado na metodologia 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 de processo 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. É justamente para atender a essa necessidade de informações que essa ferramenta se encaixa, 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 respondem 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 PowerBI que é a ferramenta que construirá o painel.

No PowerBI 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 ***PIPELINE*** 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 PowerBI.

**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.

(https://blog.dsacademy.com.br/mas-afinal-o-que-e-pipeline-de-dados/)

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 PIPELINE 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 da ferramenta 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.

Outras informações secundárias, são 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. Essas fontes de dados serão detalhadas nos próximos tópicos.

### 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 gerência processos de trabalho como autorização de funcionamento de empresas, importações de produtos, processos administrativos sanitários. Ou seja, o Datavisa é um sistema grande, 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 de forma perfeita e atualmente está em andamento uma modernização desse sistema de peticionamento.

O processo de peticionamento é quando uma empresa previamente cadastrada no sistema Datavisa pleiteia um serviço para a Anvisa. Essa solicitação é feita através do Solicita, que é uma aplicação amigável do sistema Datavisa que coleta dados específicos e documentos em pdf e realiza o pagamento da 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 3 linhas são referentes ao mesmo expediente, ou seja, um mesmo expediente pode passar por várias situações até a sua conclusão.

**Quadro 2: Fontes de dados do Sistema 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

Uma informação importante que futuramente pode ser trabalhada no painel é o servidor 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.

### 3.1.2 Tabelas Auxiliares

O assunto é a principal informação da petição, nele é especificado qual serviço a empresa solicita da 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 códigos de assuntos vão ser analisados 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 grupos conforme o tipo de 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 (Total de Assuntos=118)**

|  |  |  |
| --- | --- | --- |
| **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. (13 Tipos) |
| TIPO\_REGISTRO\_1 | Texto | Tipo de serviço nível 0: Exemplo 122. Produto Formulado de Uso agrícola de Baixo Impacto (16 Tipos) |
| TIPO\_REGISTRO\_2 | Texto | Tipo de serviço nível 0: Exemplo 1221. Produtos Biológicos e Orgânicos. (21 Tipos) |
| TIPO\_REGISTRO\_3 | Texto | Tipo de serviço nível 0: Exemplo Produto Agente Biológico de Controle (32 Tipos) |
| SISTEMA | Texto | Sistema em que o serviço é tratado: Datavisa ou 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 diagrama do 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. Sobreestrado, que é aguardando alguma restrição para poder entrar na fila, 2.Fila, 3.1.Analise, 3.2.Empresa e 4.Recurso.

As demais informações da tabela auxiliar de situação podem ser visualizadas no quadro 03.

**Quadro 3: Fontes de dados do Tabela Situação (Total de Assuntos=118)**

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

O quadro 5 detalha as informações dos critérios das filas. Quanto ao tipo de produto será utilizado somente os da Toxicologia.

**Quadro 4: Fontes de dados do Tabela Critério Fila (Total de Assuntos=118)**

|  |  |  |
| --- | --- | --- |
| **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... |
| SUBFILA | 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. |

1. Fonte: Elaboração do autor

## 3.2 Preparação dos Dados

A Gerência-Geral de Gestão do Conhecimento, Inovação e Pesquisa (GGCIP) é a responsável pela área de 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 a 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 estão alocadas no *sharepoint* da Anvisa de maneira que se possa obter as estatísticas descritivas do projeto. Considerando que as inconsistências são normais em um banco de dados grande, serão feitos tratamentos que diminuam esses erros desses 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 de leitura, tratamento e transformações se encontra no Anexo 01.

1. Leitura das tabelas de [PROCESSO], [ASSUNTO], [SITUAÇÃO] e [CRITÉRIO DE FILA].
2. Exclusão de dois ‘EXPEDIENTES’ com “DATA DE ENTRADA” inválida: ‘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 foi descartado 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 o tempo 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 executado todo dia através da ferramenta do Windows que executa uma tarefa previamente programada, *Task Manager*. Os resultados serão salvos no *sharepoint* da Anvisa e serão atualizados automaticamente pelo PowerBI. Para fins de documentação, os resultados apresentados aqui for considerada 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 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:

Tabela 01: Quantidade de Expedientes por ano de entrada e Situação



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 01: Quantidade de Expedientes por tipo de pleito e Situação



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 1.749 passaram pela fase “Análise” o que configura que uma fragilidade da informação no preenchimento do sistema. A alteração da situação é uma opção do técnico na análise do processo e é de conhecimento que esse campo não é preenchido corretamente ao longo da análise e o técnico na maioria das vezes só preenche no momento da conclusão.

Tabela 01: Quantidade de Expedientes distintos que passaram pelas fases no processo e Tipo de pleito



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: CO\_ASSUNTO, NU\_CNPJ\_EMPRESA, NO\_RAZAO\_SOCIAL\_EMPRESA, NU\_PROCESSO, NU\_EXPEDIENTE, NO\_PRODUTO, DATA\_ENTRADA, DS\_SITUACAO\_ASSUNTO\_DOC, DT\_INICIO\_UL\_SITUACAO, AREA\_GGTOX, CICLO\_GGTOX

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 PowerBi esses outputs foram agrupados por dia, código de assunto e a quantidade, conforme os layouts da tabela abaixo.



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 03.



707253118 EXPEDIENTE DOBRADO!!!

## 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” de um expediente considerando que o serviço fosse solicitado hoje. Lembrando que o “tempo total” é 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 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 guardar registros 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 constata-se que somente 8 assuntos têm mais de 1.000 observações nesse período. Assim, na previsão dos tempos foi necessária uma restrição dos serviços que entrarão no desenvolvimento do modelo.



Considerando também que a fila de espera na data de 30/06/2023 é de 3.788 e que os Produtos Técnicos Equivalentes e os Produto Formulado com base em Produto Técnico Equivalente respondem por 88% dessa fila 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 preenchimento do sistema não é corretamente utilizado e que menos de 5% dos expedientes tem um “tempo de análise” estabelecido pelo sistema, conforme TABELA 0X não será possível fazer essa estimativa. Portanto, por limitações na base de dados foram testados os modelos para a previsão 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 5047 entre 2010 e 2023. Sendo que o “tempo total” para a conclusão desses expedientes varia entre 25 e 5.046 dias. Uma média de 2000 para expedientes 5041 e 2000 para 5065 com variação muito grande e muitos outliers. Como consequência de falhas no preenchimento dessas informações. A distribuição desse tempo pode ser verificada conforme os gráficos a seguir:

|  |  |
| --- | --- |
| Gráfico, Gráfico de caixa estreita  Descrição gerada automaticamente | Gráfico, Histograma  Descrição gerada automaticamente |
| Gráfico, Gráfico de dispersão  Descrição gerada automaticamente | Interface gráfica do usuário, Texto, Aplicativo  Descrição gerada automaticamente |

Para a detecção dos outliers desse “tempo total” foi utilizado a ferramenta de *IsolationForest* da biblioteca Slearn no python. Como resultado, mais de 25% das informações foram consideradas como anomalias. Apesar da sugestão de exclusão, esses valores foram mantidos para serem trabalhados 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.

|  |  |
| --- | --- |
|  |  |
|  |  |

Considerando que os gráficos de autocorrelação e autocorrelação parcial não são conclusivos foram calculados outros critérios de informação como 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 XXX os modelos Autoregressivo foram que obtiveram os menores valores indicando o melhor ajuste do modelo.



Utilizando o modelo Autoregressivo as previsões para os pque os gráficos de autocorrelação e autocorrelação parcial não são conclusivos foram calculados outros critérios de informação como 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 XXX os modelos Autoregressivo foram que obtiveram os menores valores indicando o melhor ajuste do modelo. Com as seguintes previsões para os próximos três meses: 5041: 1578, 1577 e 1524 e 5065: 1941, 2010 e 1967, conforme gráfico XXX

|  |  |
| --- | --- |
|  |  |

## 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 PowerBI.e o ChatGPT.

O GitHub (https://github.com/) 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 versões dos 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.

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 está disponível em [https://idados.anvisa.gov.br/repdados/DocumentosRepdados/Forms/AllItems.aspx?RootFolder=%2frepdados%2fDocumentosRepdados%2fPOWERBIRS&FolderCTID=0x012000FC8ACDF518531C479E32DB0103134FAB](https://idados.anvisa.gov.br/repdados/DocumentosRepdados/Forms/AllItems.aspx?RootFolder=%2Frepdados%2FDocumentosRepdados%2FPOWERBIRS&FolderCTID=0x012000FC8ACDF518531C479E32DB0103134FAB) e somente pode ser acessada pelos interlocutores de dados 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 conecção a uma fonte de dados externa por vários aplicativos, incluídos o Excel, o Access e o PowerBI. A configuração é feita da seguinte forma: Executar ODBC Data Source 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 foi possível também criar o modelo de previsão do tempo total que será utilizado.

O PowerBi é a ferramenta da Microsoft que organiza e interpreta grandes volumes de dados. Com ela é possível extrair os dados com o PowerQuery e relacionar tabelas, criar novas variáveis e apresentar os dados. Ainda assim será utilizado o python por ter melhor possibilidade de utilização dos códigos escritos e por apresenta melhor performance. O PowerBI 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 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 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.

# RESULTADOS DO PROJETO

Desde 2016, com as mudanças na gestão visando a melhoria dos processos de trabalho, já existe um processo de medição da produtividade da área, mas feitas a partir dos expedientes publicados no Diário Oficial da União (DOU). Esses indicadores eram feitos a partir de planilhas feitas pelas gerências e preenchidas pelos técnicos. Esse processo artesanal que implicava em retrabalho e informações erradas começou a ser substituído a partir de 2020, substituindo a fonte de dados pelas informações do Datavisa, pois evitava o retrabalho de preencher planilhas de acompanhamento, as informações eram mais confiáveis, acompanharia processos que não tinham publicação no DOU e a entrada de novos expedientes, pois somente são publicados processos finalizados.

Assim, o conceito desse projeto já está em andamento ao longo dos anos e é uma necessidade das gerências da GGTOX e também das áreas supervisoras da Anvisa, como a Assessoria de Planejamento (APLAN) e a própria Terceira Diretoria, que é a responsável da área. Assim, proposta desse projeto de intervenção foi criar uma metodologia de classificação dos eventos transparente e automatizar a criação dessas informações de maneira que ocorra com a mínima intervenção humana para evitar retrabalho e erros.

Atualmente o projeto já apresenta alguns resultados, que ainda que não são definitivos, mas que acrescentam inovações aos processos de trabalho: Expedientes em Andamento, Estatísticas Descritivas, Indicadores de Performance e Expedientes Finalizados.

## 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 não foram finalizados ainda. Como já foi dito, o Datavisa não gera essa informação de extrema importância, pois os técnicos somente conseguem listar os processos que estão na fila, a partir do portal de consultas da Anvisa. 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 solicitem à Anvisa o resultado da análise.

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

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

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 no preenchimento do sistema. Mas dos expedientes de registros, 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.



## 4.2 Estatísticas Descritivas

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

Interface gráfica do usuário, Gráfico

Descrição gerada automaticamente

Com essa ferramenta é possível que o gestor acompanhe semanalmente sua produtividade, 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 filtrada 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.

Gráfico

Descrição gerada automaticamente

No Exemplo acima 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 acompanhar 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 agrotóxicosquímicos para uso agrícola em relação à média dos últimos 3 anos. Esse indicador foi definido para o ano de 2023 e considerou a média de avaliações concluídas entre 2020 e 2022, que foi 398. 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 selecionados e identificar os tempos em dias em cada fase. Exemplos de consultas:

* Quais processos de Registro Especial Temporário (5004) foram finalizados em maio de 2022 que tiveram tempo total de conclusão maior do que 200 dias.
* Quais processos de pós registro ficaram em fila (fase do processo) entre 50 e 100 100 dias entre 2020 e 2022;
* Identificado que um processo ficou mais de 600 dias em análise quais foram as fases e situações desse processo.

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Esse painel demonstra o histórico de quantos dias o processo ficou em cada situação e como a metodologia de cálculo dos indicadores está mostrando isso. Server para os gestores identificar os outliers e entender porque os processos ficaram parados tanto tempo em cada fase.

TODOS COM RECURSO E CANCELADO

# CRONOGRAMA DE IMPLANTAÇÃO DO PROJETO

Considerando que algumas etapas do projeto de intervenção já foram concluídas e outras estão em andamento o quadro 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.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **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 |  |  |  |  |  |  |  |  |  |  |  |  |

# 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 (GGCIP) 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 PowerBI 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 devem ser os interlocutores de dados da área, que fazem parte de uma política de gestão do conhecimento e tratamento da informação, que também continuam com outras atribuições. Portanto não existe como discriminar um custo específico do projeto, pois muitos desses recursos já estão alocados e concorrem com outros projetos.

# CONSIDERAÇÕES FINAIS

Neste item deve inserir a percepção em relação a este Projeto de Intervenção, lições aprendidas, o foi desenvolvido, as dificuldades durante a elaboração. Os objetivos alcançados, o resultado final do trabalho.

Avanços da Anvisa na Gestão dos Dados – Saída do bi, próprio curso de capacitação estamos no caminho certo.

Barreira Tecnológica – Instalação de software livre / antivírus. Dificuldade de testar novas soluções gratuitas. Barreira para instalação do Anaconda

Barreira GGCIP – Uma área para atender todas as demandas. Limitada. Não oferece acesso às tabelas. As áreas fins são as que conhecem os dados.

Barreira AD da GGTIN – impossibilidade de acesso ao modelo de dados do sistema

Grande resistência dos servidores à cultura de dados. Servidores muito capacitados em análises toxicológicas mas ...

Dados muito ruins nos períodos antigos os erros continuam

Sistema muito ruim – Datavisa permite erros.

Avanços - Discriminar o técnico responsável pela análise

# REFERÊNCIAS

AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA. **Relatório Anual da Ouvidoria da Anvisa**. 2001, 2008, 2012, 2014, 2015, 2016, 2017 e 2018. Brasília, DF, Disponível em: <<http://portal.anvisa.gov.br/ouvidoria/publicacoes>> Acesso em 7 out. 2019.

\_\_\_\_\_\_. **Relatório de Atividades da Anvisa**. 2005 a 2017. Brasília, DF, Disponível em: < <http://portal.anvisa.gov.br/relatorios-de-gestao1>> Acesso em 7 out. 2019.

\_\_\_\_\_\_. **Relatório de Gestão da Anvisa**. 2000 a 2019. Brasília, DF, Disponível em: < <http://portal.anvisa.gov.br/relatorios-de-gestao1>> Acesso em 7 out. 2019.

BRASIL. **Lei nº 12.527, de 18 de novembro de 2011**. Diário Oficial da República Federativa do Brasil, Poder Executivo, Brasília, DF, 18 nov 2011. Disponível em: <<http://www.planalto.gov.br/ccivil_03/LEIS/L7802.htm>> Acesso em 25 nov. 2019.

\_\_\_\_\_\_. **Constituição (1988)**. **Constituição da República Federativa do Brasil.** Brasília, DF, 5 out 1988. Senado Federal: Centro Gráfico, 1988.

\_\_\_\_\_\_. **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 03 out. 2019.

\_\_\_\_\_\_. **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 03 out. 2019.

# APÊNDICES

**Apêndice A – Scrip em Python Tratamento dos Dados**

# ANEXOS

**ANEXO 1 – Caso 1 - Transparência Ativa da fila para análise de registros de Agrotóxicos**