

▼ MVP para Análise de Informações do INSS

Lucila da Conceição Castello Branco

▼ Definição do trabalho *

O dataset usado neste projeto será um arquivo do governo com informações da Comunicação de Acidentes de Trabalho (CAT), são somente as informações que governo disponibiliza no site do portal de transparência. Será realizada uma análise de acidentes de Janeiro a Maio de 2023 no Brasil para responder algumas questões:

Qual o ramo de negócio que gerou mais afastamentos no período? Responder essas perguntas tem por objetivo entender qual é o ramo de negócio no Brasil que menos se preocupa com a saúde dos colaboradores para que a fiscalização seja mais atuante nas empresas ou até mesmo gere encargos maiores.

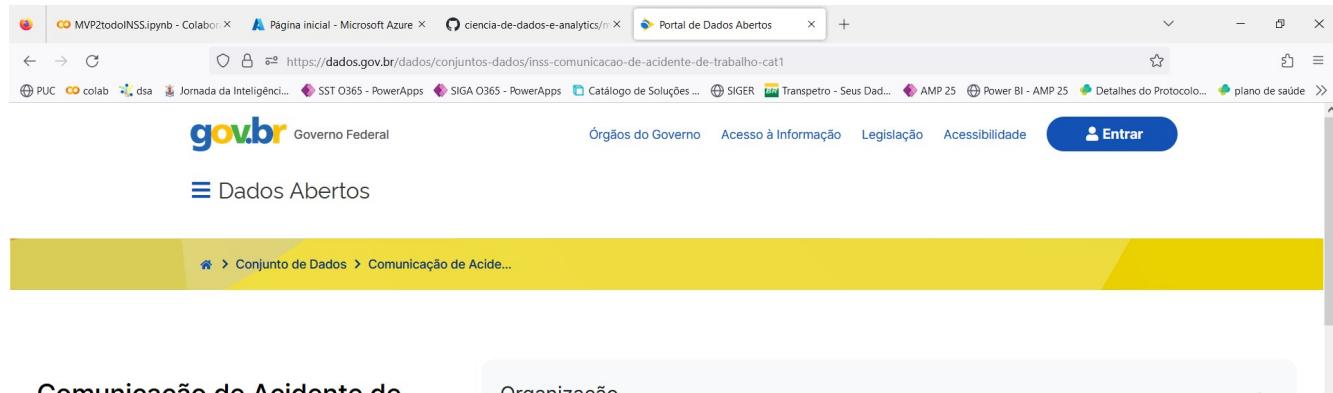
Quem sofre mais acidentes, homem ou mulher? O objetivo dessa pergunta é entender quem tem mais cuidado e atenção na prevenção de acidentes.

Qual é a parte do corpo que as pessoas mais se machucam? O objetivo dessa pergunta é preparar campanhas e/ou políticas para que as pessoas tenham mais atenção e cuidado.

Qual é o Estado que mais tem acidentes? O objetivo dessa pergunta é para entender e possivelmente exigir do governo uma disponibilização maior de recursos para equipar (caso não haja) e aprimorar hospitais e toda a equipe de Saúde para região de forma imediata e ter um olhar mais focado em políticas para mitigar os acidentes na região.

▼ 1. Origem dos dados

O site utilizado foi <https://dados.gov.br/dados/conjuntos-dados/inss-comunicacao-de-acidente-de-trabalho-cat1>



The screenshot shows the INSS website's 'Comunicação de Acidente de Trabalho - CAT - Plano de Dados Abertos 2016/2018' section. It features the INSS logo, a satisfaction scale (2/10), a search bar, and a navigation menu with links like '+ Seguir', 'Avaliar', 'Contato', and '35 conjuntos de dados'. The page also displays a message about the organization's role in providing previdenciários services to the Brazilian society.

3. Análise de Dados, tratamento e modelagem

Clique duas vezes (ou pressione "Enter") para editar

Para dar andamento ao projeto seria interessante conhecer a amplitude de informações do dataset. Para tal, como a nuvem utilizada é a Microsoft Azure, para empresas é disponibilizado o Microsoft Purview. É um recurso muito interessante para que haja governança e gestão dos dados, disponibiliza um controle para que os usuários tenham acesso somente às informações necessárias, permite catalogar todos os dados para que haja um entendimento das informações em bancos de dados. Infelizmente a conta disponibilizada para fazer o trabalho não é de estudante ou de empresa, portanto o recurso não pode ser utilizado, mas para manter os dados catalogados, o próprio site do INSS disponibiliza as informações de cada coluna como segue:

Agente Causador do acidente: Descrição e código do agente causador do acidente, é uma lista com 305 opções.

CBO:Código Brasileiro de Ocupação, a lista tem 2424 opções

CBO Descrição: É a descrição de cada uma das opções do CBO.

CID: Código Identificador da Doença de acordo com o CID-10 - Código Código Internacional de Doenças, a lista tem 15086 opções.

CID Descrição: É uma breve descrição de cada código da CID.

CNAE: Classificação Nacional da Aividade Econômica no AEPS, são 87 opções numéricas.

CNAE Descrição: Breve descrição do código do CNAE
Tipo do Acidente: Pode ser típico, trajeto ou doença.
Emitente da CAT: É o emissor da CAT, pode ser o empregador, empregado, médico ou sindicato.
Espécie do Benefício: pode ser Auxilio Doença, Auxilio por morte ou PA.
Filiação do Segurado: Tipo de Filiação à Previdência Social do Segurado da CAT, pode ser Empregado, Trabalhador Avulso, Empregado Doméstico, Contribuinte individual, Segurado Especial e Segurados Facultativos.

Indicador de Óbito Acidente:Indicador de óbito do segurado, pode ser sim ou não.

Município Empregador: Atualmente a lista tem 5589 opções de Municípios.

Data Afastamento: Data em que ocorreu o afastamento do segurado, do seu trabalho, devido ao acidente de trabalho.

Data DDB: Data do Despacho do Benefício.

Data Acidente: Data do Acidente de Trabalho registrada na CAT

Data Nascimento: Data do Nascimento do Segurado, para efeito de limites, o ínicio deve ser depois de 1923.

Data Emissão da CAT: Dia que informaram a ocorrência no sistema.

Devido a janela de tempo, os campos data (exceto nascimento) tem valores que são de janeiro até maio de 2023, todos estilo AAAA/MM/DD.

Os dados foram baixados, organizados em uma planilha csv e inserido manualmente em um banco de dados do azzure

The screenshot shows the Microsoft Azure Storage Explorer interface. On the left, there's a sidebar with navigation links like 'Visão geral', 'Log de atividade', 'Marcações', 'Diagnosticar e resolver problemas', 'IAM (Controle de Acesso)', 'Migração de dados', and 'Eventos'. The main area shows a directory structure under 'acidente'. A file named 'D.SDA.PDA.005.CAT.202301_05.csv' is listed. Below the file, there are columns for 'Nome', 'Última modificação', 'Camada de acesso', 'Tipo de blob', and 'Tamanho'. The file is identified as a 'Blob de blocos' with a size of '125.3 MiB'.

Foi realizado um levantamento do formato das informações contidas no banco de dados.

The screenshot shows the Microsoft Data Factory interface. On the left, there's a sidebar with sections for 'Pipelines', 'Captura de Dados de Alterações...', 'Conjuntos de Dados', 'Fluxos de dados', and 'Power Query'. Under 'Fluxos de dados', 'dataflow1' is selected. The main area shows the 'Inspecionar' tab for 'dataflow1'. It displays the schema of the 'AzureSqlTable1' table, which contains 13 columns: Agente Causador Acidente, CNAE, Ramo_Empresa, Espécie do benefício, Filiação do Segurado, Munic Empr, Parte Corpo Atingida, Sexo, Tipo do Acidente, UF Munic. Acidente, UF Munic. Empregador, Data Acidente, and Data Nascimento.

Foram criados todos os recursos possíveis para atendimento da proposta.

The screenshot shows the Microsoft Azure Storage Explorer interface, similar to the one above. It shows a blob named 'acidente' containing a file 'D.SDA.PDA.005.CAT.202301_05.csv'. The blob details show it's a block blob with a size of 125.3 MiB.

Criar um recurso Todos os recursos Contas do Microsoft... Bancos de dados SQL Azure Databricks Serviços gratuitos Grade de Eventos Contas de armazenamento Registros de contêiner Mais serviços

Recursos

Recente FAVORITO

Nome	Tipo	Última visualização
acidentefactory	Data factory (V2)	4 horas atrás
sistemaabcd	Cofre de chaves	4 horas atrás
catalogo-acidente	Conta do Microsoft Purview	4 horas atrás
Acidentes	Grupo de recursos	6 horas atrás
acidentesqul	Banco de dados SQL	8 horas atrás
serveracidente	SQL Server	8 horas atrás
acidente	Conta de armazenamento	9 horas atrás
workers-vnet	Rede virtual	10 horas atrás
Acidentes23workapce	Serviço do Azure Databricks	13 horas atrás
Azure subscription 1	Assinatura	2 dias atrás

[Ver todos](#)

Vale lembrar que alguns foram desnecessários, já que exigem recursos que uma conta comum na nuvem para o desenvolvimento do projeto não tem acesso, somente conta corporativa ou de estudante.

Coletor	Configurações	Erros	Mapeamento	Otimizar	Inspeccionar	Visualização de dados
Escada Perman...	52110-Vende...	M25.5 Dor Arti...	3702			Atividades Rel...
Queda de Pes...	414110-Armaz...	S90.9 Traum S...	5250			Atividades Rel...
Escada Perman...	141415-Ger. L...	S30.0 Contusa...	4789			Comercio Varei...
Escada Perman...	424105-Entreví...	S90.0 Contusa...	8411			Administracao ...
Chão - Superfí...	252105-Admin...	S82.2 Frat da D...	4329			Obras de Instal...
Queda de Pes...	251520-Psicól...	W01.0 Residen...	8660			Atividades de ...
Metal - Inclui L...	821445-Prepar...	S60.0 Contusa...	7490			Atividades Prof...
Rua e Estrada -...	333115-Prof.Es...	S92.3 Frat de ...	8599			Atividades de ...
Transportador ...	414105-Almox...	S62.2 Frat do P...	2854			Fabricacao de ...
Vidrania, Fibra ...	322205-Tec. de...	Y28.9 Local Ne	8411			Administracao ...
Impacto de Pe...	513115-Gover...	S62.6 Frat de ...	5510			Hoteis e Simila...

O recurso utilizado foi uma tabela estilo flat. No tratamento dos dados, algumas colunas foram retiradas porque eram desnecessárias para responder as questões levantadas. Outras que eram desnecessárias não foram retiradas porque poderiam surgir novas questões no decorrer da análise dos dados.

Todas as execuções de pipeline > pipeline1 - A atividade é executada > dataflow1

Tempo de inicialização do cluster: 1s 144ms Número de transformações: 3 Status do fluxo de dados: Success

Atualizar Atualização automática Ligado Editar o fluxo de dados

The screenshot shows the Microsoft Power BI Data Studio interface. On the left, there's a sidebar with sections for 'Runtimes e sessões', 'Runtimes de integração', 'Depuração de fluxo de dados', 'Notificações', and 'Alertas e métricas'. The main area is titled 'Coletores' and shows a table with one row for 'sink1'. The row details are: Coletor (sink1), Status (Bem-sucedido), Tempo de processamento (3s 58ms), Maior tempo de processamento (2s 546ms), Linhas gravadas (213k), Fases (empty), and Linhagem (empty).

A Base de dados foi enviada para o data studio para que sejam feitos alguns tratamentos e as questões levantadas pudessem ser respondidas.

The screenshot shows the Azure Data Studio interface. The left sidebar shows connections to 'serveracidente.database.windows.net' and the 'acidentesql' database. The main area displays the results of a query: 'select * from Acidente01_05'. The results grid has columns: 'Agente', 'Causador', 'Acidente', 'CNAE2.0 Empregador4', 'CNAE2.0 Empregador5', and 'Espécie do benefício'. The data includes various accident types like 'MOTOCICleta, motoneta', 'Queda de Pes. em Mesmo Nivel, Nic', and 'Impacto de Pes. Contra Objeto Parado', along with their corresponding codes and descriptions.

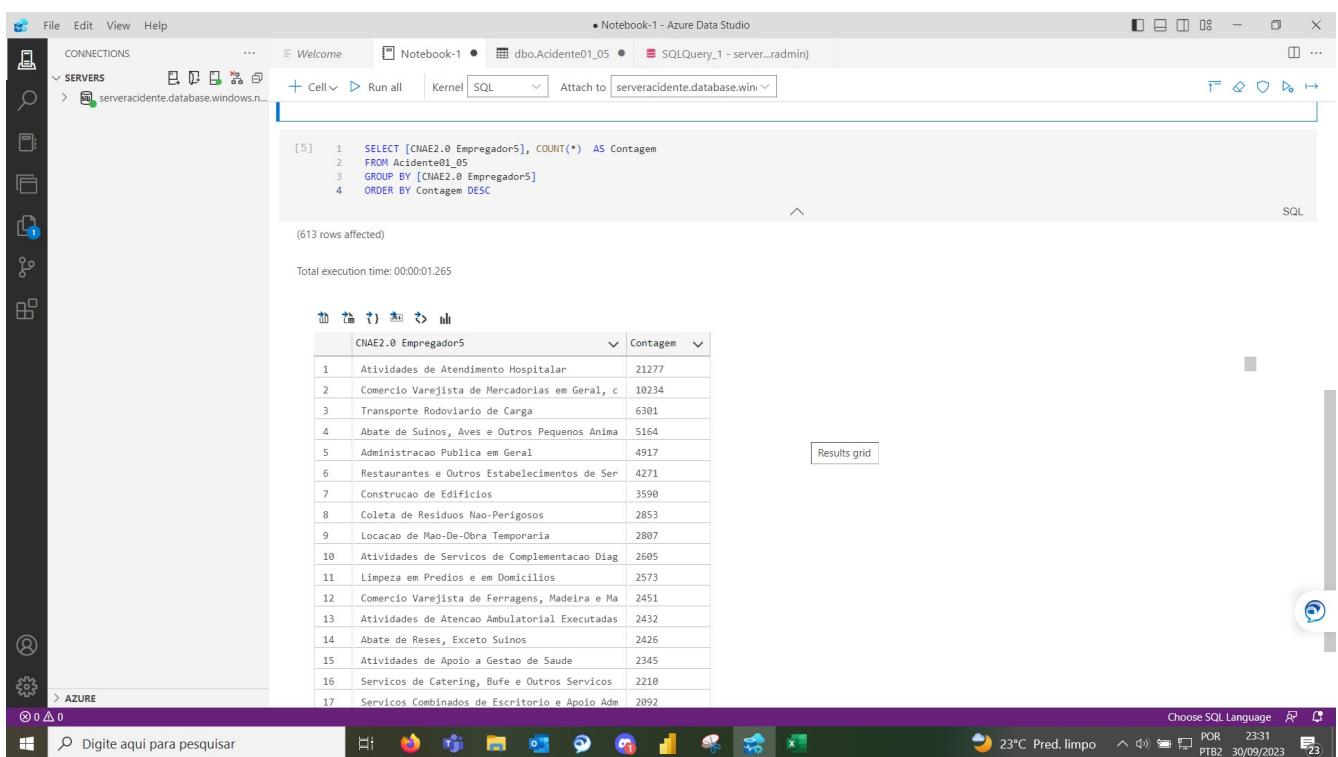
Entre as questões levantadas, a primeira questão a ser respondida seria saber o Estado que apresenta o maior índice de acidentes de trabalho.

The screenshot shows the Azure Data Studio interface. The left sidebar shows connections to 'serveracidente.database.windows.net' and the 'acidentesql' database. The main area displays the results of a query: 'SELECT [UF] Munic., [Acidente], COUNT(*) AS Contagem FROM Acidente01_05 GROUP BY [UF] Munic., [Acidente] ORDER BY Contagem DESC'. The results grid has columns: 'UF', 'Munic.', 'Acidente', and 'Contagem'. The data shows the top states by accident count, with Maranhão having the highest count at 78351.



14	Mato Grosso do Sul	3416
15	Acre	2935
16	Alagoas	1671
17	Rio Grande do Norte	1406
18	Pernambuco	1146
19	Piauí	781
20	Amapá	612
21	Sergipe	604
22	Rio Grande Norte	306
23	Paraíba	208

Com base nos resultados apresentados, fica claro que o estado do Maranhão tem uma necessidade muito alta de recursos para atendimento aos acidentados do INSS, talvez seja interessante conhecer as indústrias da região.



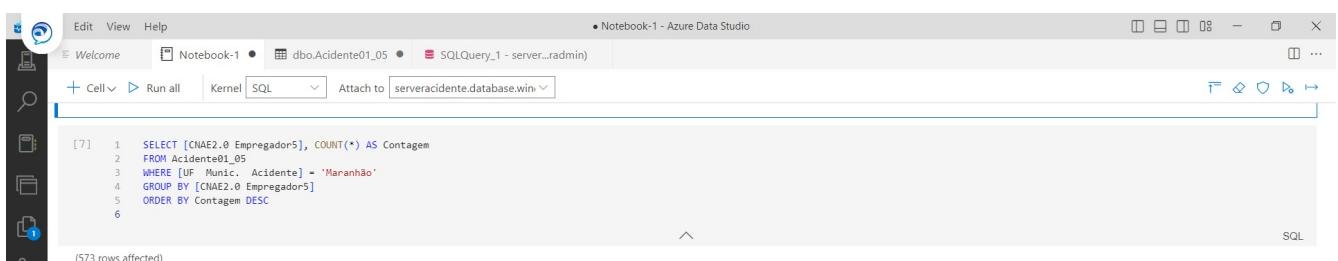
```
[5] 1 SELECT [CNAE2.0 Empregador$], COUNT(*) AS Contagem
2 FROM Acidente01_05
3 GROUP BY [CNAE2.0 Empregador$]
4 ORDER BY Contagem DESC
```

(613 rows affected)

Total execution time: 00:00:01.265

CNAE2.0 Empregador\$	Contagem
Atividades de Atendimento Hospitalar	21277
Comercio Varejista de Mercadorias em Geral, c	18234
Transporte Rodoviário de Carga	6301
Abate de Suínos, Aves e Outros Pequenos Animais	5164
Administracao Publica em Geral	4917
Restaurantes e Outros Estabelecimentos de Serviços	4271
Construcao de Edificios	3590
Coleta de Residuos Nao-Perigosos	2853
Locacao de Mao-De-Obra Temporaria	2807
Atividades de Servicos de Complementacao Diagnóstica	2605
Limpeza em Predios e em Domicilios	2573
Comercio Varejista de Ferragens, Madeira e Materiais de Construcao	2451
Atividades de Atencao Ambulatorial Executadas	2432
Abate de Reves, Exceto Suínos	2426
Atividades de Apoio a Gestao de Saude	2345
Servicos de Catering, Buffe e Outros Servicos	2210
Servicos Combinados de Escritorio e Apoio Administrativo	2092

Com base nos dados, é curioso imaginar que no Brasil a maior quantidade de acidentes no trabalho ocorram justamente em empresas de Atendimento hospitalar, já que o objetivo é justamente cuidar das pessoas, mas também é compreensível, já que o ambiente de trabalho é justamente com doenças, são vírus e bactérias o tempo todo no ambiente, qualquer descuido pode acarretar sérias consequências. Como o Maranhão apresentou maior quantidade de acidentes, cabe acrescentar uma "lupa" para analisar somente as ocorrências neste estado.



```
[7] 1 SELECT [CNAE2.0 Empregador$], COUNT(*) AS Contagem
2 FROM Acidente01_05
3 WHERE [UF Munic. Acidente] = 'Maranhão'
4 GROUP BY [CNAE2.0 Empregador$]
5 ORDER BY Contagem DESC
6
```

(573 rows affected)

Total execution time: 00:00:00.713

CNAE2.0 EmpregadorS		Contagem
1	Atividades de Atendimento Hospitalar	7588
2	Comercio Varejista de Mercadorias em Geral, c	3648
3	Transporte Rodoviario de Carga	2239
4	Administracao Publica em Geral	1998
5	Locacao de Mao-De-Obra Temporaria	1985
6	Restaurantes e Outros Estabelecimentos de Ser	1811
7	Atividades de Atencao Ambulatorial Executadas	1397
8	Limpeza em Predios e em Domicilios	1260
9	Construcao de Edificios	1175
10	Comercio Varejista de Ferragens, Madeira e Ma	1029
11	Servicos de Catering, Bufe e Outros Servicos	970
12	Atividades de Servicos de Complementacao Diag	942
13	Coleta de Residuos Nao-Peligrosos	859
14	Atividades de Apoio a Gestao de Saude	846
15	Fabricacao de Pecas e Acessorios para Veiculo	825
16	Fabricacao de Artefatos de Material Plastic	783

Choose SQL Language 22°C Pred. limpo POR 23:48 PTBZ 30/09/2023

Digite aqui para pesquisar

É interessante analisar que no Maranhão o atendimento hospitalar também é o setor que gera mais acidentes em relação aos outros setores.

Ao analisar o Sexo das pessoas, foi identificado que o banco de dados precisa de um pequeno ajuste, ao tentar ajustar não encontrei o caminho mais adequado, como o valor é insignificante, mantive por que o mais importante é a informação apresentada:

Welcome serveracidente.database.windows.net:accidentesql Notebook-1 dbo.Acidente01_05 SQLQuery_1 - server...radmin

Cell ▾ Run all Kernel SQL Attach to serveracidente.database.win

17	Fabricacao de Acucar em bruto	723
18	Atividades de Servicos Prestados Principalmen	681
19	Comercio de Pecas e Acessorios para Veiculos	681
20	Atividades de Atencao a Saude Humana na Fene	659

```
[8] 1 SELECT [Sexo], COUNT(*) AS Contagem
 2 FROM Acidente01_05
 3 GROUP BY [Sexo]
 4 ORDER BY Contagem DESC
```

(4 rows affected)

Total execution time: 00:00:02.984

Sexo		Contagem
1	Masculino	136754
2	Feminino	74493
3	Não Informado	2384
4	Indeterminado	3

A diferença entre homens e mulheres que sofrem acidente de trabalho é significativa, há a possibilidade das mulheres estarem mais atentas com sua segurança em relação aos homens, mas para uma análise mais profunda, é preciso conhecer o universo de homens e mulheres que tem emprego, procurei esse banco de dados para anexar ao trabalho, mas infelizmente um banco de dados com essa divisão para o período, não foi encontrado.

Como o Maranhão tem sido foco de interesse devido ao alto número de Comunicações de Acidentes de Trabalho, foi realizado esse acompanhamento analisando em paralelo ao país, ele também apresenta as mesmas características.

```
[19] 1 SELECT [Sexo], COUNT(*) AS Contagem
2   FROM Acidente01_05
3  WHERE [UF Munic. Acidente] = 'Maranhão'
4 GROUP BY [Sexo]
5 ORDER BY Contagem DESC
```

(3 rows affected)

Total execution time: 00:00:00.490

	Sexo	Contagem
1	Masculino	48804
2	Feminino	28724
3	Não Informado	823

Para ajustar o erro apresentado e apresentar imagens mais interativas, será usado o powerbi, ferramenta liberada pelo professor, para obter os dados da SQL e assim enriquecer as informações apresentando imagens. Para efeitos do trabalho ficarão congeladas, mas para apresentações do relatório, oferece muito mais interatividade e riqueza de informações. Para transferir para o powerBi, baixei o arquivo do suporte para ter interação direta com o banco de dados, ou através do powerBi, ele oferece recursos para importar dados em nuvem que ao publicar mantém interação em caso de alguma mudança para que os dados estejam sempre atualizados.

The screenshot shows the Microsoft Azure Data Factory interface. On the left, there's a sidebar with 'Recursos do Data Factory' (Resources) expanded, showing 'Pipelines' with one item named 'pipeline1'. The main area is titled 'Atividades' (Activities) and shows a list of available activities: 'Mover e transformar', 'Synapse', 'Azure Data Explorer', 'Azure Functions', 'Serviço em Lote', 'Databricks', 'Data Lake Analytics', 'Geral', 'HDInsight', 'Iteração e condicionais', 'Machine Learning', and 'Power Query'. A 'Fluxo de dados' (dataflow) card is selected, showing 'dataflow1'. To the right, there's a context menu with options like 'Monitor', 'Clonar', 'Renomear', etc. The top navigation bar includes 'Microsoft Azure', 'Data Factory', 'Validar tudo', 'Pesquisar alocador e documentação', and user information.

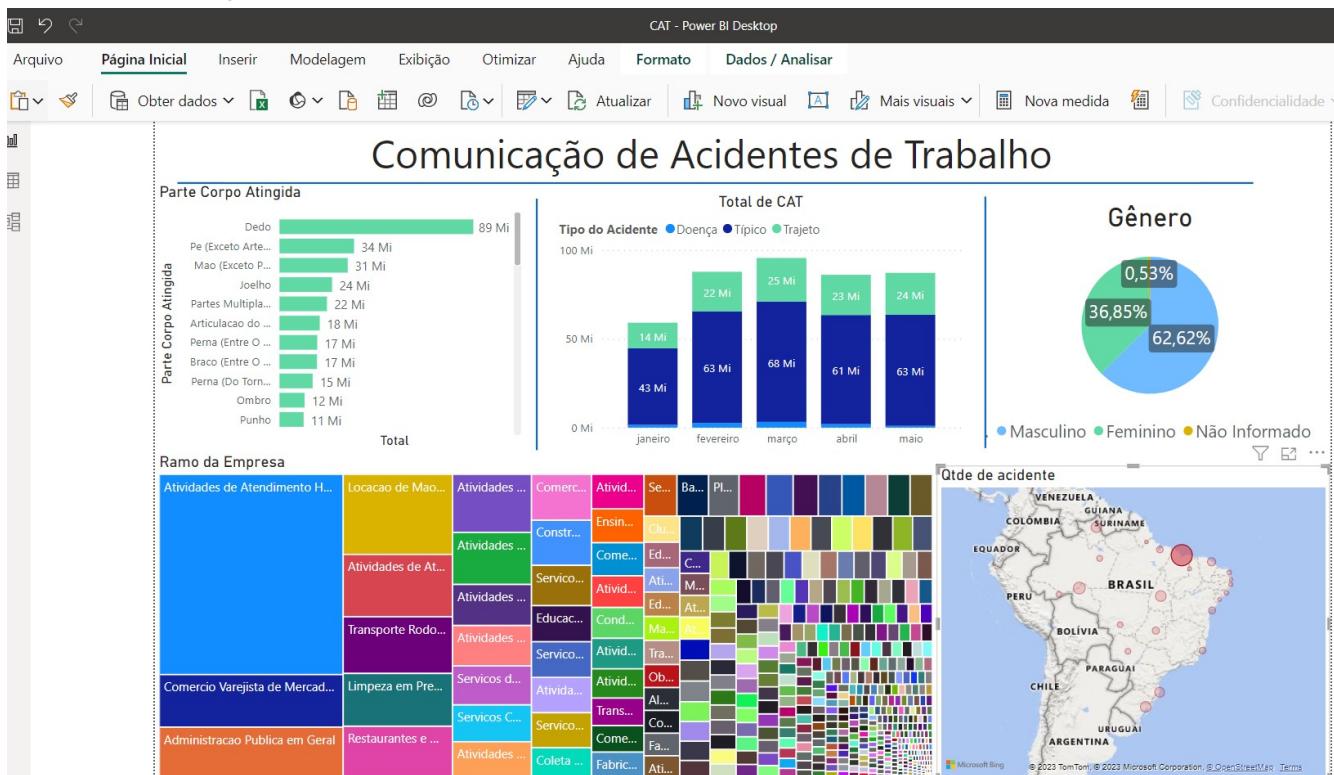
Com os dados transferidos para o power BI, foi criado um dashboard:





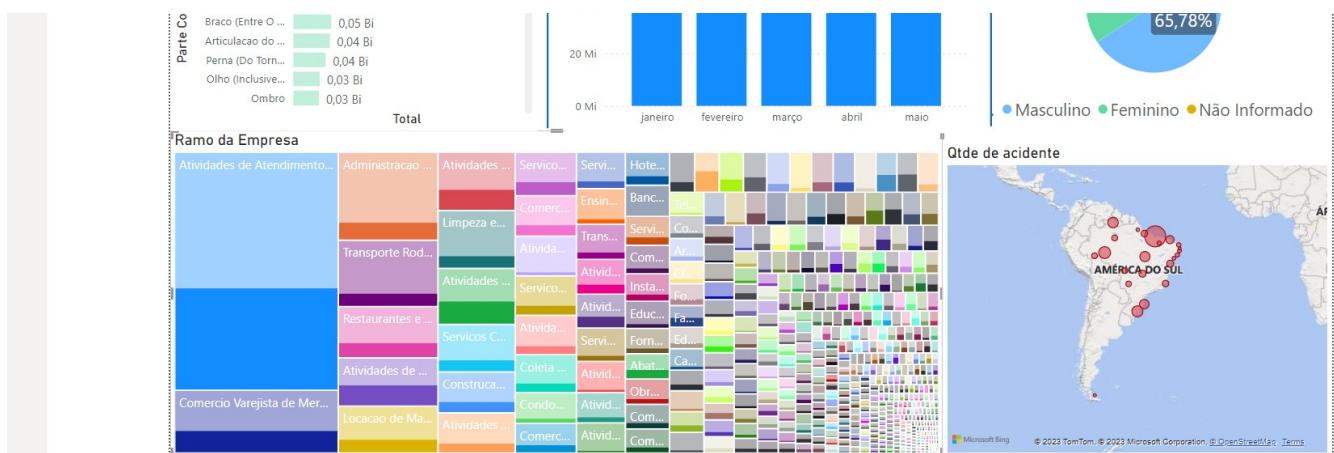
Foi acrescentado um gráfico sobre os tipos de acidente para identificar se o acidente é realmente trabalhando ou em deslocamento. Acidentes em deslocamento a responsabilidade é do governo, pois o colaborador está fora da empresa.

Ao selecionar apenas o estado Maranhão, observa-se a mesma tendência:

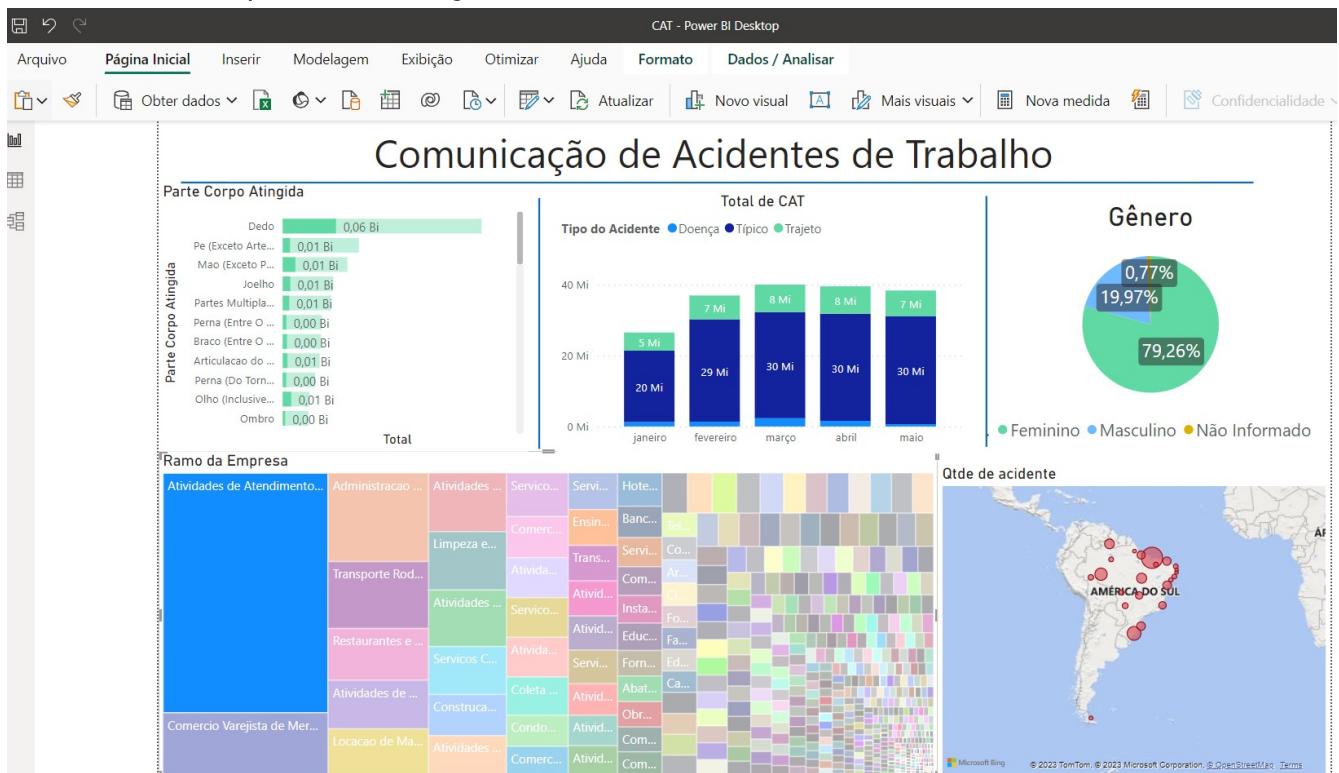


Através das imagens, percebe-se que a quantidade de machucados com mãos e dedos é significativa, mas não existe um ramo de mercado específico, tende a ser proporcional em todos os ramos de mercado.





O curioso dos acidentes é que a maioria dos colaboradores que se machucam no atendimento hospitalar são do gênero feminino:



Conclusão

Respondendo as questões levantadas no início do projeto, o ramo de Atendimento Hospitalar apresenta o maior índice de acidentes de trabalho, podem ser levantadas várias questões, como o ambiente de trabalho, já que os colaboradores estão expostos a diversos tipos de doenças, como bactérias, vírus e etc. por tratar de pessoas que já estão doentes. Por outro lado, os dedos são os mais afetados, cabe um cuidado maior com a segurança dos equipamentos utilizados devido ao risco associado.

O Maranhão é o estado que apresenta a maior quantidade de acidentes, mas no geral,

acompanha a proporção dos outros indicadores do Brasil.

Outra questão levantada foi a parte do corpo que mais sofre em caso de acidentes. No geral, mãos e principalmente dedos são os mais afetados, a quantidade de pessoas que machucam o dedo é mais que o dobro do 2º lugar, que é o pé, seguido da mão, portanto mãos e dedos requerem mais atenção para a segurança.

No universo de trabalhadores masculinos e femininos, os homens são a grande maioria, mas essa questão não pode ser validada pois faz-se necessário uma análise na proporção de homens e mulheres inseridos no ambiente de trabalho, mas esse dado não foi disponibilizado no INSS e não foi encontradoo outro banco de dados com tal informação. Por outro lado, as mulheres que mais se machucam estam trabalhando no atendimento hospitalar, mas não significa que neste ramo as mulheres são mais descuidadas com a segurança que os homens, ter um levantamento da proporção entre homens e mulheres no ambiente de trabalho é essencial para fazer qualquer afirmação categórica. O gênero LGBTQIA+ não foi analisado por falta de informações no INSS.