

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS
IEC – Instituto de Educação Continuada
Pós-graduação *Lato Sensu* em Ciência de Dados e Big Data

Luiza Castelar Rodrigues Pereira

Transformando dados em insights:
Análise e Tratamento de Dados de Empresas Brasileiras utilizando o POWER
BI

Belo Horizonte

2023

Luiza Castelar Rodrigues Pereira

Transformando dados em insights:

**Análise e Tratamento de Dados de Empresas Brasileiras utilizando o POWER
BI**

Belo Horizonte

2023

SUMÁRIO

1. Introdução	4
1.1. Contextualização	4
1.2. O problema proposto	4
1.3. Objetivos	5
2. Coleta de Dados	7
3. Processamento/Tratamento de Dados	13
4. Análise e Exploração dos Dados	19
5. Interpretação dos Resultados	41
6. Links	45
REFERÊNCIAS	46
APÊNDICE	47

1. Introdução

1.1. Contextualização

Atualmente no Brasil, para se empreender ou trabalhar por conta própria, é necessário a obtenção de CNPJ. O CNPJ é um registro que permite a emissão de notas fiscais, está relacionado ao pagamento de impostos e é fundamental para a regularização dessa condição de trabalho adversa do CLT.

Em 2022, aproximadamente 4 milhões de novos CNPJs foram abertos no país¹, mesmo em tempos de pandemia, totalizando aproximadamente 20 milhões de micro, pequenas, médias e empresas de grande porte².

Ao analisar o maciço conjunto de dados que caracterizam os CNPJs, é possível descobrir novos insights que permitirão avaliar o atual cenário das empresas brasileiras, permitindo a tomada de decisões pela sociedade, pelos próprios empreendedores e pelo governo.

1.2. O problema proposto

Ao analisar o cenário empreendedor brasileiro, podemos ajudar tanto o governo a tomar decisões de, por exemplo, incentivo às empresas, recuperações e redução/ajustamento de arrecadação tributária, como também ajudamos diretamente a população ao informar como estão organizadas as empresas brasileiras e quais suas características.

A partir da análise dos dados, um empreendedor pode tomar uma decisão de continuar seu negócio, se vai ou não abrir mais um negócio que pode competir com os demais existentes, verificar quanto de capital as empresas em média em um determinado segmento estão investindo para começar, etc.

Para alcançar o objetivo de forma mais minuciosa, analisarei o cenário de empresas que possuem estabelecimentos no estado de Minas Gerais (ou seja, que tenham sede ou filial nesse estado, sem levar em consideração se uma mesma empresa possa ter outra filial em outro estado ou até mesmo em outro país).

1.3. Objetivos

Com o intuito de auxiliar na avaliação do cenário empreendedor brasileiro, irei responder várias perguntas, dentre elas:

- 1) Qual a distribuição de porte entre empresas? Há mais empresas de pequeno, médio ou grande porte?
- 2) Qual a faixa de capital (reais) que as empresas estão injetando nos seus negócios inicialmente?
- 3) Qual a faixa etária dos empreendedores? Há alguma faixa que se destaca dentre as demais?
- 4) Onde se localizam essas empresas dentro de Minas?
- 5) Quantas empresas estão vinculadas a um ente federativo? E qual a relação entre elas e a principal atividade que exercem?
- 6) Quais as atividades primárias e secundárias as empresas mais estão associadas?
- 7) Qual a relação numérica entre estabelecimento e empresa? Há muitos estabelecimentos em poder de poucas empresas?
- 8) Há muitas empresas que utilizam o simples e o MEI simultaneamente? Como se comportam as empresas com relação à data de entrada e saída dessas duas opções? Há algum padrão nessas datas?

Para isso, farei a limpeza dos dados, removendo colunas e registros, farei a transformação dos dados, criarei novas tabelas além das já existentes, relacionarei as tabelas entre si, criarei gráficos univariados e multivariados com filtragens dos dados e analisarei todo esse conjunto de dados para tirar conclusões que ajudarão a entender melhor o cenário das empresas, utilizando como principal ferramenta o POWER BI.

Utilizarei como ferramental um Windows Server da AWS com 16 Gigas de RAM e 50 Gigas de HD do tipo t2.xlarge com POWER BI versão 2.116 (de Abril 2023) para poder trabalhar no dataset, devido ao tamanho do conjunto de dados.

2. Coleta de Dados

Os dados foram coletados a partir de bases públicas oferecidas pelo governo, através da Secretaria Especial da Receita Federal, atrelada ao Ministério da Economia. São **10** tabelas ao todo (algumas, pelo tamanho, já vem divididas em mais partes), juntamente a uma seção de Regime Tributário dividida em pequenas outras tabelas (Imune e Isentas, Lucro Arbitrado, Lucro Presumido e Lucro Real).

Devido ao tamanho já extenso das tabelas principais, não usarei as tabelas pertencentes à seção de Regime Tributário.

Os dados foram coletados no dia 06/04/2023, através do link dados.gov.br/dados/conjuntos-dados/cadastro-nacional-da-pessoa-juridica---cnpj. Essa base é atualizada mensalmente e a última atualização feita pelo governo até a confecção deste trabalho foi em 11/03/2023.

Neste conjunto de dados as tabelas que estão divididas em mais de um arquivo são as tabelas Empresas, Estabelecimento e Sócios. Usarei apenas um arquivo de cada, cuja numeração seja 0 (ou seja, Empresas0, Estabelecimentos0 e Socios0), novamente devido à grande quantidade de dados e a limitação de hardware para processá-los.

O detalhamento de cada uma das tabelas se encontra em: www.gov.br/receitafederal/dados/cnpj-metadados.pdf

As tabelas Natureza, Qualificações, CNAE, Países, Motivos e Municípios tem 2 campos, sendo um o **CODIGO**, em inteiro e outro a **DESCRICAÇÃO**, em string. O código é um identificador único e o campo DESCRICAÇÃO está especificado abaixo para cada tabela, sendo que cada um dos registros dessas tabelas não é repetido nem vazio.

- 1) Campo DESCRICAÇÃO da tabela Natureza = nome da natureza jurídica das empresas (ex: fundação privada, entidade sindical, fundo/clube de investimentos, etc)
Possui 90 registros.
- 2) Campo DESCRICAÇÃO da tabela Qualificações = nome da qualificação de sócios (ex: sócio, fundador, empresário, etc).
Possui 68 registros.
- 3) Campo DESCRICAÇÃO da tabela Países = o nome do país.
Possui 255 registros

4) Campo DESCRICAO da tabela CNAE = nome da atividade econômica exercida

Possui 1359 registros.

5) Campo DESCRICAO da tabela Motivos = nome do motivo da situação cadastral (ex: ativo, falência, fusão, etc)

Possui 61 registros

6) Campo DESCRICAO da tabela Municípios = o nome de cada município brasileiro

Possui 5571 registros

As demais tabelas estão descritas abaixo:

Empresas (possui 12.530.642 de registros e 7 colunas)

Nome da coluna/campo	Descrição	Tipo
CNPJ_BASICO	os 8 primeiros números do CNPJ	Texto
RAZAO_SOCIAL	nome sobre o qual a empresa se identifica	Texto
NATUREZA_JURIDICA	código da natureza jurídica	Número
QUALIFICAÇÃO_RESPONSÁVEL	código da qualificação do responsável pela empresa	Número
CAPITAL_SOCIAL_EMPRESA	valor inicial investido na empresa, disponível para ela	Número
PORTE_EMPRESA	código do porte/tamanho da empresa	Número
ENTE_FEDERATIVO_RESPONSÁVEL	relativo à uma entidade ligada ao governo	Texto

Estabelecimentos (possui 13.032.276 de registros e 31 colunas)

Nome da coluna/campo	Descrição	Tipo
CNPJ_BASICO	os 8 primeiros números do CNPJ	Texto
CNPJ_ORDEM	4 dígitos do meio que simbolizam o tipo de CNPJ (matriz ou filiais)	Texto
CNPJ_DV	2 últimos dígitos, de verificação, indicam se o número do CNPJ é válido	Texto
IDENTIFICADOR_MATRIZ/FILIAL	Número que identifica se o estabelecimento é matriz ou filial	Número
NOME_FANTASIA	Nome fantasia atrelado ao estabelecimento	Texto
SITUAÇÃO_CADASTRAL	Código da situação cadastral	Número
DATA_SITUAÇÃO_CADASTRAL	Data do evento da situação cadastral	Data
MOTIVO_SITUAÇÃO_CADASTRAL	Código do motivo da situação cadastral	Número
NOME_DA_CIDADE_NO_EXTERIOR	Nome da cidade no exterior em que o estabelecimento está	Texto
PAÍS	Código do país onde está o estabelecimento	Número
DATA_DE_INÍCIO_ATIVIDADE	Data de início da atividade do estabelecimento	Data
CNAE_FISCAL_PRINCIPAL	Código indicando a atividade principal declarada pelo CNPJ	Número

CNAE_FISCAL_SECUNDÁRIA	Código indicando as atividades secundárias declarada pelo CNPJ	Lista de números
TIPO_DE_LOGRADOURO	Descrição do logradouro (rua, estrada, avenida, etc)	Texto
LOGRADOURO	Nome da rua/avenida, etc	Texto
NÚMERO	Número onde se localiza o estabelecimento	Texto
COMPLEMENTO	Complemento da onde se localiza o estabelecimento	Texto
BAIRRO	Nome do bairro onde se localiza o estabelecimento	Texto
CEP	CEP do estabelecimento	Texto
UF	Sigla do estado onde o estabelecimento está	Texto
MUNICÍPIO	código do município onde o estabelecimento está	Número
DDD_1	DDD do telefone 1	Texto
TELEFONE_1	Número do telefone principal do estabelecimento	Texto
DDD_2	DDD do telefone 2	Texto
TELEFONE_2	Número do telefone secundário do estabelecimento	Texto
DDD_FAX	DDD do fax	Texto
FAX	Número do Fax do estabelecimento	Texto
CORREIO_ELETRÔNICO	Email do dono do estabelecimento	Texto

SITUAÇÃO_ESPECIAL	se a empresa está passando por uma situação especial, como ex. problema jurídico	Texto
DATA_DA_SITUAÇÃO_ESPECIAL	Data em que a situação especial foi registrada	Data

Dados do Simples (possui 34.695.558 de registros e 7 colunas)

Nome da coluna/campo	Descrição	Tipo
CNPJ_BÁSICO	os 8 primeiros números do CNPJ	Texto
OPÇÃO_PELoSIMPLES	se a empresa segue o regime tributário do simples (sim ou não)	Texto
DATA_DE_OPÇÃO_PELoSIMPLES	data com ano, mês e dia que a empresa começou a seguir o regime tributário do simples	Data
DATA_DE_EXCLUSÃO_DOSIMPLES	data com ano, mês e dia que a empresa deixou de seguir o regime do simples	Data
OPÇÃO_PELOMEI	se a empresa é mei ou não	Texto
DATA_DE_OPÇÃO_PELOMEI	data que a empresa passou a ser mei	Data
DATA_DE_EXCLUSÃO_DOMEI	data em que a empresa deixou de ser mei	Data

Sócios (possui 4.377.321 de registros e 11 colunas)

Nome da coluna/campo	Descrição	Tipo
CNPJ_BÁSICO	os 8 primeiros números do CNPJ	Texto

IDENTIFICADOR_DE_SÓCIO	se é pessoa física, jurídica ou estrangeiro	Número
NOME_SÓCIO_RAZÃO SOCIAL	nome do sócio se pessoa física e razão social se o sócio for juridico	Texto
CNPJ/CPF_SÓCIO	cpf para pessoa física. Se o sócio estiver como jurídico, esse campo é o cnpj	Texto
QUALIFICAÇÃO_SÓCIO	código da qualificação do sócio	Número
DATA_ENTRADA_SOCIEDADE	data que o sócio entrou para a sociedade	Data
PAÍS	se sócio é estrangeiro, o código do seu país	Número
REPRESENTANTE_LEGAL	número do cpf do representante legal	Texto
NOME_REPRESENTANTE	nome do representante legal, quando o sócio é pessoa jurídica	Texto
QUALIFICAÇÃO_REPRESENTANTE	código de qualificação do representante legal	Número
FAIXA_ETÁRIA	código correspondente à faixa etária do sócio	Número

3. Processamento/Tratamento de Dados

Foi feito um pré-processamento de filtrar a tabela de Estabelecimentos para que os estabelecimentos analisados fossem apenas os de Minas Gerais. Além desse filtro, apliquei também o filtro relativo à Empresas em cima de Estabelecimentos (ou seja, todo estabelecimento analisado tem que ter uma empresa correspondente). Estabelecimentos sem empresas, devido aos registros correspondentes estarem em outro arquivo não importados para esse modelo (como Empresa1, Empresa2, etc), não serão analisados.

A diminuição no número de registros, de 13.032.276 registros totais da tabela Estabelecimentos para 299.836 registros se faz necessária para que as análises de cada tabela e a conclusão façam sentido entre si.

As demais ligações, mesclagens e operações realizadas entre a tabela Estabelecimentos e as outras tabelas na hora de criar os gráficos também devem seguir a filtragem de Minas e a correspondência por Empresas.

Tabela Estabelecimentos

Transformei os campos CNPJ_BASICO, CNPJ_ORDEM e CNPJ_DV da tabela Estabelecimentos em texto (pois estavam em número) e juntei os valores para formar a coluna CNPJ_FINAL para testar se todos estabelecimentos estavam atrelados a apenas um CNPJ único.

Uma vez que verifiquei que todos os CNPJs eram únicos, removi as colunas CNPJ_ORDEM e CNPJ_DV, pois os valores presentes em CNPJ_ORDEM (que representa se aquele CNPJ é matriz ou filial) já estavam bem representados na coluna IDENTIFICADOR_MATRIZ/FILIAL e o significado de CNPJ_DV para o problema não era relevante, uma vez que é apenas um verificador que é calculado a partir dos números iniciais do CNPJ para dizer se ele é válido ou não.

Removi as colunas TIPO_LOGRADOURO, LOGRADOURO, NÚMERO, COMPLEMENTO e CEP pois eram dados muito específicos que também poderiam ser tratados pela coluna BAIRRO (uma vez que quero uma informação mais abrangente para plotar em um mapa de região). Das demais colunas que se referiam a endereço, juntei BAIRRO com MUNICÍPIO e

ESTADO e criei uma coluna específica de PAÍS, do tipo texto, diferente da que já vinha no dataset (que era um código para outra tabela), para formar uma só coluna ENDEREÇO, que pode ser inserida em gráfico apropriado. Como estamos filtrando os dados sendo apenas do estado de Minas Gerais, não era preciso definir o estado e país com as tabelas que já existiam nesse modelo, para esse caso da tabela Estabelecimentos. Um exemplo de registro de como ficou os valores da variável ENDEREÇO: SAVASSI,BELO HORIZONTE,MG,Brasil.

Ao plotar os valores de ENDEREÇO no gráfico, aproximadamente de 10 a 20 registros não foram localizados corretamente, sendo colocados em outros estados. Exclui o bairro desses registros específicos para corrigir esse problema.

Como a tabela auxiliar Municipio só servia para se ligar à uma coluna de todo o dataset (se ligar à coluna MUNICIPIO da tabela Estabelecimentos), pude mesclar as ocorrências iguais da tabela Município com a coluna MUNICIPIO do Estabelecimento e excluir a tabela auxiliar mencionada anteriormente.

Removi a coluna NOME_DA_CIDADE_EXTERIOR e PAÍS (coluna de número, original do dataset), uma vez que a análise somente contempla o estado de Minas Gerais (já que os dados estavam filtrados para Minas, o país seria sempre Brasil e não haveria o nome da cidade no exterior, uma vez que todas as cidades são brasileiras).

Transformei a coluna CORREIO_ELETRONICO em 0 ou 1, sendo 1 para presença de email e 0 para ausência do email.

Removi as colunas de DDD correspondentes à TELEFONE_1, TELEFONE_2 e FAX e transformei TELEFONE_1, TELEFONE_2 e FAX em colunas de 0 ou 1, indicando presença ou falta do atributo (também renomeei as colunas para EXISTE_TELEFONE_1, EXISTE_TELEFONE_2 e EXISTE_FAX).

Removi as colunas SITUACAO_ESPECIAL e DATA_SITUACAO_ESPECIAL, pois todas as linhas se encontravam vazias de todos os registros, não representando informações relevantes.

Transformei as colunas DATA_SITUAÇÃO_CADASTRAL e DATA_INICIO_ATIVIDADE em formato DD/MM/YYYY, pois elas estavam em formato YYYYMMDD.

Criei uma nova tabela chamada de EstabelecimentosCNAE, que seria a ligação entre a tabela Estabelecimentos e os CNAES desses estabelecimentos, utilizando o CNPJ_FINAL criado anteriormente como meio de identificação de cada estabelecimento na tabela EstabelecimentosCNAE. Nessa tabela, adicionei as colunas de CNAE primário e secundário (excluindo ambas as colunas da tabela original de Estabelecimentos), e para distinguir os primários dos secundários criei outra tabela TipoCnae que se liga a EstabelecimentosCNAE.

A coluna NOME_FANTASIA é a única que não haverá qualquer modificação, pois seu valor é mais para conferir se a lógica dos resultados de outros gráficos está correta e não necessariamente fazer alguma agregação ou contagem no dado.

Após todas as manipulações anteriores, feitas no Power Query, criei uma coluna de nome EMPRESAS_CNPJ via DAX (essa coluna não fica visível no Power Query, somente nas análises). Essa coluna me indica qual empresa está associada à qual estabelecimento, para poder filtrar a tabela de Empresa com relação ao estado de Minas. Ela possui o valor de CNPJ_BASICO da tabela Empresa.

Tabela Empresas

Removi um registro da tabela de Empresas pois tinha o mesmo CNPJ_BASICO que outra empresa e estava vazio o restante dos registros, não representando informações relevantes:

1 Coluna = COUNTROWS(FILTER(Empresas, Empresas[CNPJ_BASICO] = EARLIER(Empresas[CNPJ_BASICO])))							Dados
CNPJ_BASICO	RAZAO_SOCIAL	NATUREZA_JURIDICA	QUALIFICACAO_RESPONSAVEL	CAPITAL_SOCIAL_EMPRESA	PORTE_EMI		Pesq
42938862		0	0	0			
42938862	IVAIR GONCALVES DINIZ E OUTRA	4120	59	0			CNA Emp Σ C C

Criei uma nova coluna a partir do ENTE_FEDERATIVO, pois ele continha o nome da cidade e o estado concatenados. Então separei em uma coluna própria de cidade e outra de estado, chamados respectivamente de ENTE_FEDERATIVO_CIDADE e

ENTE_FEDERATIVO_ESTADO.

Preenchi os valores nulos da coluna PORTE_EMPRESA com 0 (não informado) e arredondei os valores de CAPITAL_SOCIAL_EMPRESA, para não aceitar centavos. Também transformei CAPITAL_SOCIAL_EMPRESA em inteiro para facilitar as contas.

Limpei a coluna RAZAO_SOCIAL pois algumas pessoas colocaram seus nomes como razões sociais antecedidos ou sucedidos dos seus CPFs (removi os CPFs).

Adicionei uma coluna via DAX chamada EXISTE_ESTABELECIMENTO_FILTRADO, com valores de True ou False para indicar quais empresas possuíam estabelecimentos em MG. Essa coluna serve para filtrar a tabela Sócios através da tabela Empresas, pois Sócios se liga em Empresas, mas não em Estabelecimento.

Tabela Simples

Transformei os valores de DATA_EXCLUSAO_MEI, DATA_OPCAO_MEI, DATA_EXCLUSAO_PELoSIMPLES e DATA_OPCAO_PELoSIMPLES, que estavam no formato YYYYMMDD em data no formato DD/MM/YYYY.

Valores zerados na data foram substituídos por vazio (null).

Filtrei os valores de CNPJ_BASICO existentes nessa tabela para serem maiores que 40.000.000. Como qualquer valor abaixo desse não representa um CNPJ existente na tabela de Empresas quando filtrado pelo estado de Minas, o total de registros do Simples diminui (isso ajuda a diminuir o tamanho do arquivo. O Power BI só permite o upload online de arquivos com 1 Gb).

Tabela Sócios

Transformei DATA_ENTRADA_SOCIEDADE de YYYYMMDD para DD/MM/YYYY. Inseri o valor nulo em 2 linhas com datas inválidas (por ser um número insignificante, não me preocupei com o valor real dessas datas).

Transformei os códigos da coluna PAÍS de texto para número e mesclei com a tabela País. Como era somente essa tabela que usava da tabela País, pude excluir a tabela País do modelo.

Criei uma coluna no DAX chamada EMPRESA_FILTRADA dentro da tabela de Sócios, para trazer o valor da coluna EXISTE_ESTABELECIMENTO_FILTRADO da tabela Empresas para dentro da tabela de Sócios a fim de filtrar mais facilmente essa tabela.

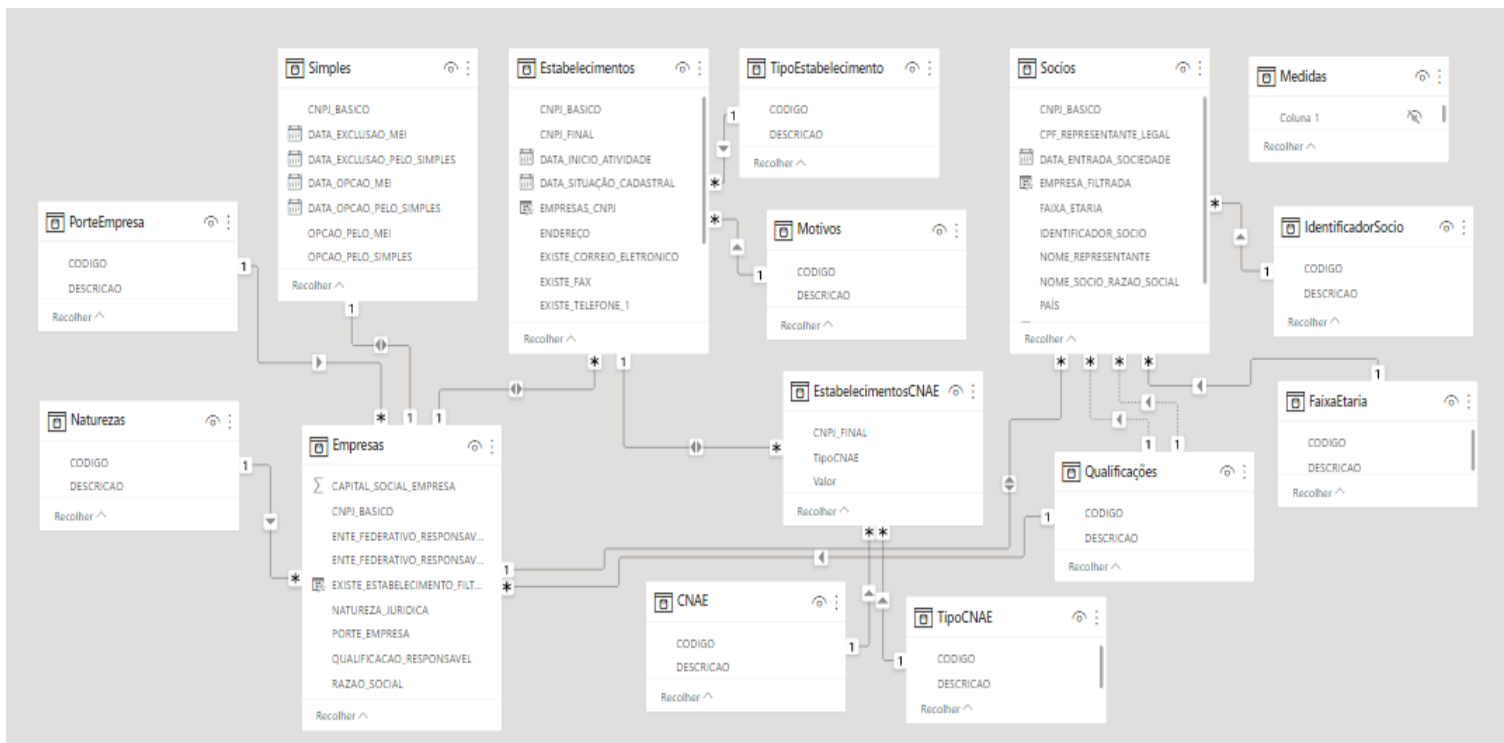
A coluna CNPJ_CPF_SOCIO tem a mesma informação que NOME_SOCIO_RAZAO_SOCIAL (ou seja, distinguir os sócios seja por nome ou cpf, como únicos). Como o CPF está com anonimização em alguns dígitos, podem haver números repetidos, o que faz com que o motivo de manter esse campo não seja mais válido (portanto removerei o CNPJ_CPF_SOCIO).

Nem a coluna NOME_REPRESENTANTE e REPRESENTANTE_LEGAL tem informações relevantes para análise, pois a presença ou não de representante pode ser inferida através da coluna QUALIFICAÇÃO_REPRESENTANTE, e como a informação não nula trazida por essas colunas, em comparação com o restante da amostragem é insignificante, resolvi não analisar com detalhes. Além disso, como o CPF na coluna REPRESENTANTE_LEGAL também está com anonimização, podem haver números repetidos, o que faz com que o motivo de manter esse campo seja ainda menos proveitoso.

Demais Tabelas

Criei os relacionamentos entre tabelas originais, e criei as tabelas extras “PorteEmpresa”, “TipoEstabelecimento”, “IdentificadorSócio”, “FaixaEtária” e “TipoCNAE” conforme imagem abaixo.

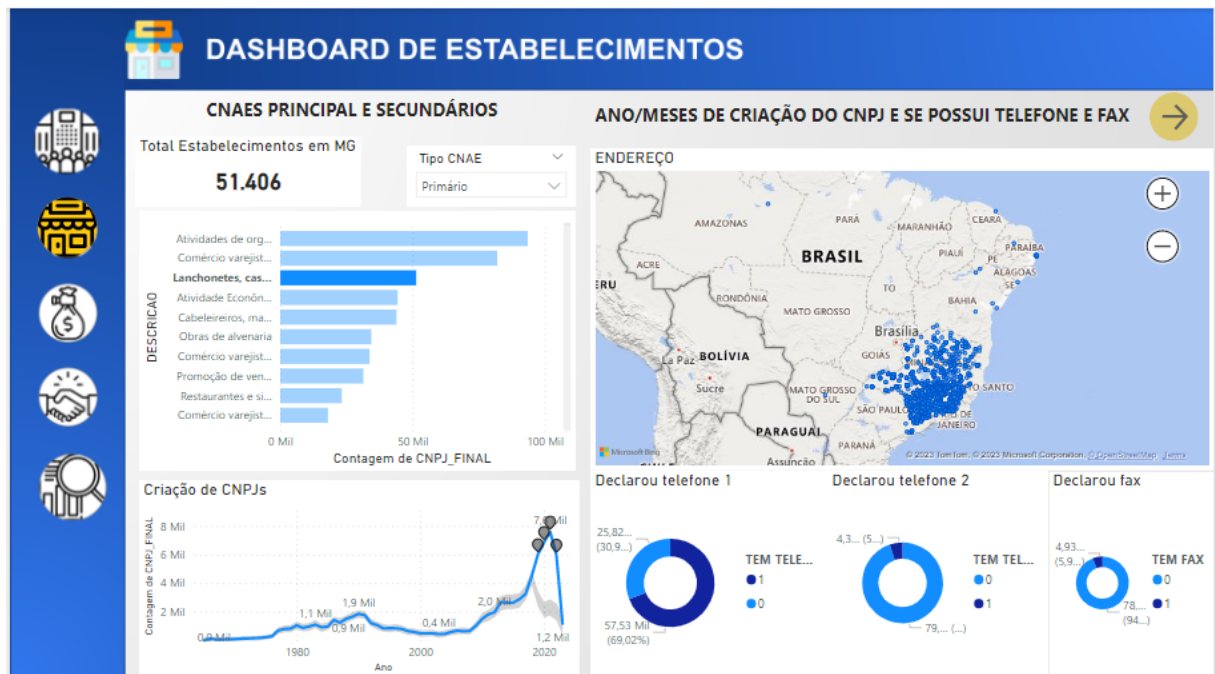
Também criei uma tabela de Medidas, que possui todas as medidas criadas via DAX para qualquer tabela do modelo.



Após todo o processamento, vemos que todos os relacionamentos são de 1:N ou 1:1 (não há relacionamentos N:N). O relacionamento que seria N:N, por exemplo, entre EstabelecimentoCNAE e CNAE, gerei uma tabela extra que é a EstabelecimentosCNAE, de forma a ter relacionamentos 1:N apenas e ficar mais fácil interpretar os resultados gráficos.

4. Análise e Exploração dos Dados

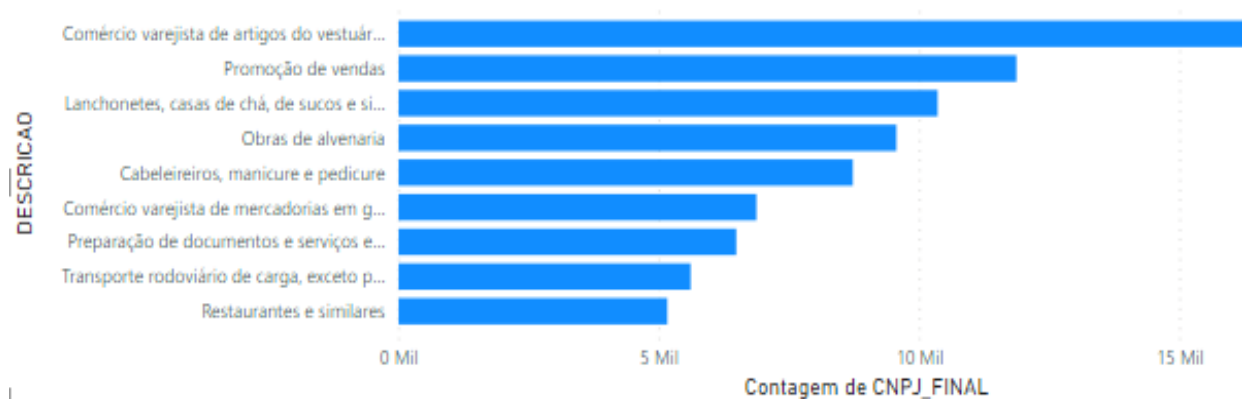
Separei cada tabela principal (Empresas, Estabelecimentos, Sócios e Simples) em páginas próprias para análise (vide foto abaixo de exemplo). A análise pode ser vista em: [powerBIWeb](https://powerbiweb.com).



Análise de Estabelecimentos

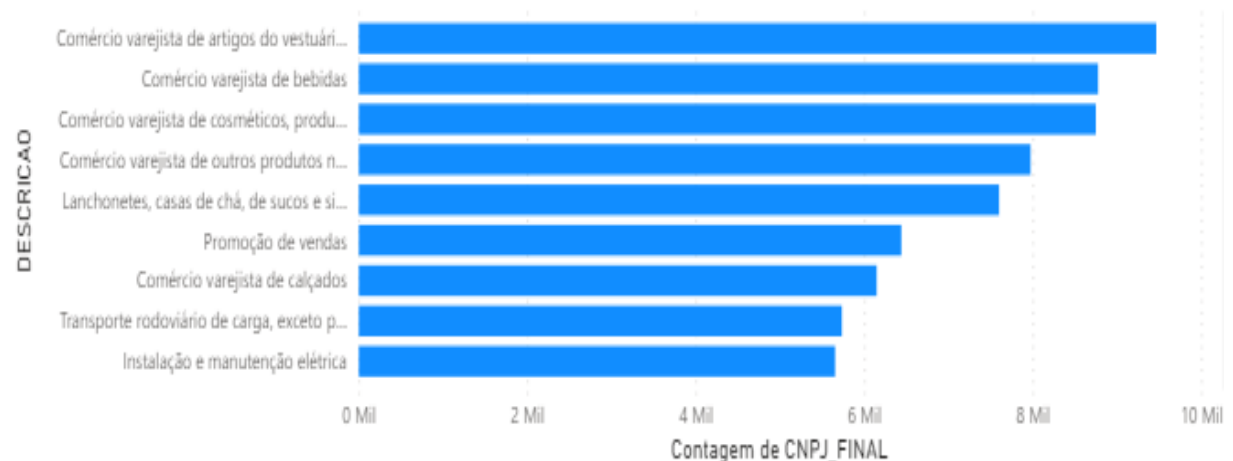
Ao analisar o que os empreendedores colocam como segmento principal de seus estabelecimentos, vemos que o top 3 segmento empresarial é de comércio varejista de vestuário (aproximadamente 16.278 estabelecimentos); promoção de vendas com 11.871 estabelecimentos e lanchonetes e semelhantes como terceiro (com 10.357).

CNAES de estabelecimentos



Analisando o top 3 de CNAES secundários que os empreendedores colocam para os estabelecimentos, vemos que o 1º é de comércio varejista de roupas, 2º comércio de bebidas e 3º comércio de cosméticos. Nem todos os estabelecimentos colocaram CNAES secundários (menos da metade, aproximadamente 142 mil, de um total de 299.836).

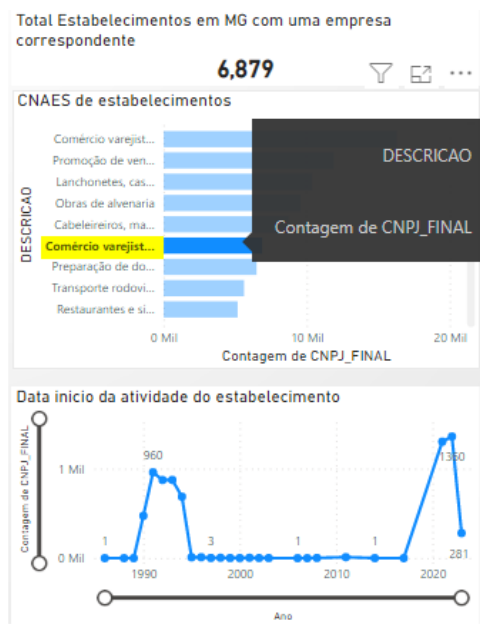
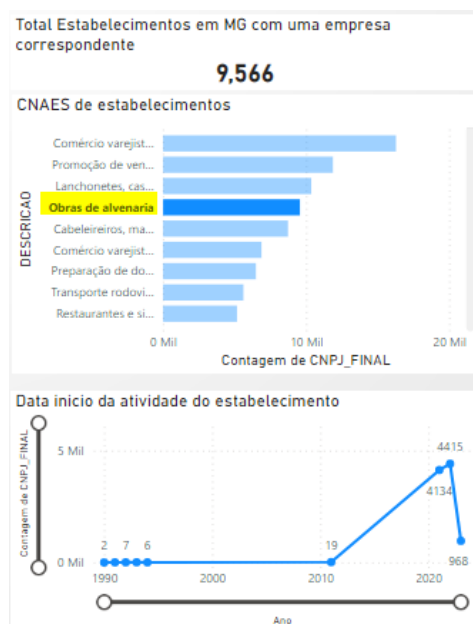
CNAES de estabelecimentos



Em relação à criação de estabelecimentos, temos comportamentos bem interessantes e diferentes relativos à cada CNAE. Os estabelecimentos varejistas de roupas, por exemplo, deram um salto enorme em 2020 para frente, o que não era tão expressivo no passado. Já os estabelecimentos de promoção de vendas também tiveram um salto semelhante ao de comércio de roupas, com seu pico em 2022.



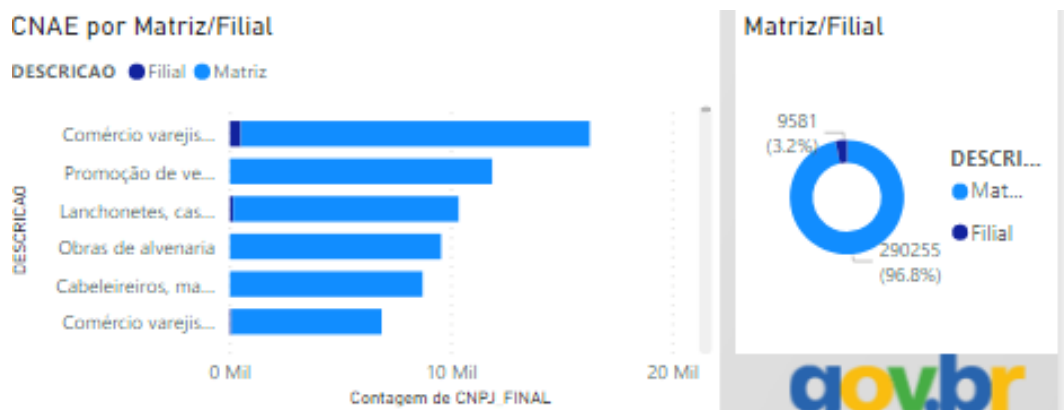
Já estabelecimentos como os ligados à obras de alvenarias, sempre tiveram crescimento constante, desde 2010, enquanto o de comércios ligados à comida tiveram um pico nos anos de 1991 a 1993, decaindo depois o número de empresas com esse tipo de CNAE, e aumentando novamente em 2020.



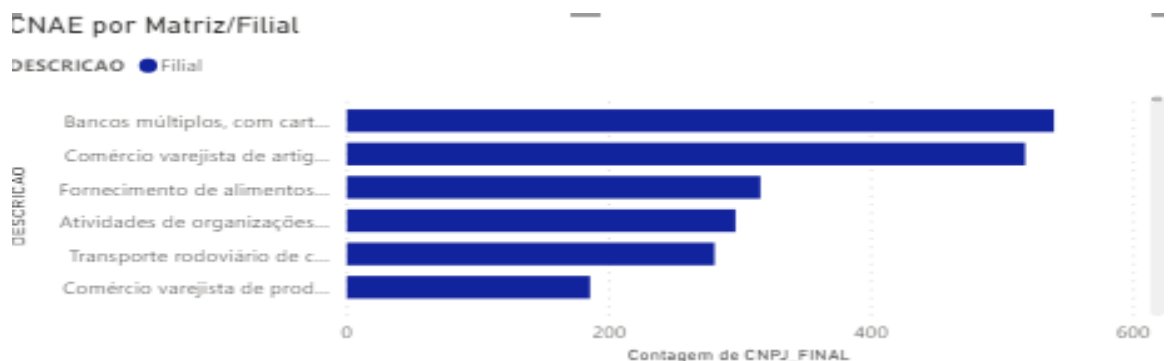
A análise por região em Minas mostra que há estabelecimentos espalhados por todo o território de Minas, porém vemos uma menor concentração no noroeste do Estado e a maior concentração na capital.



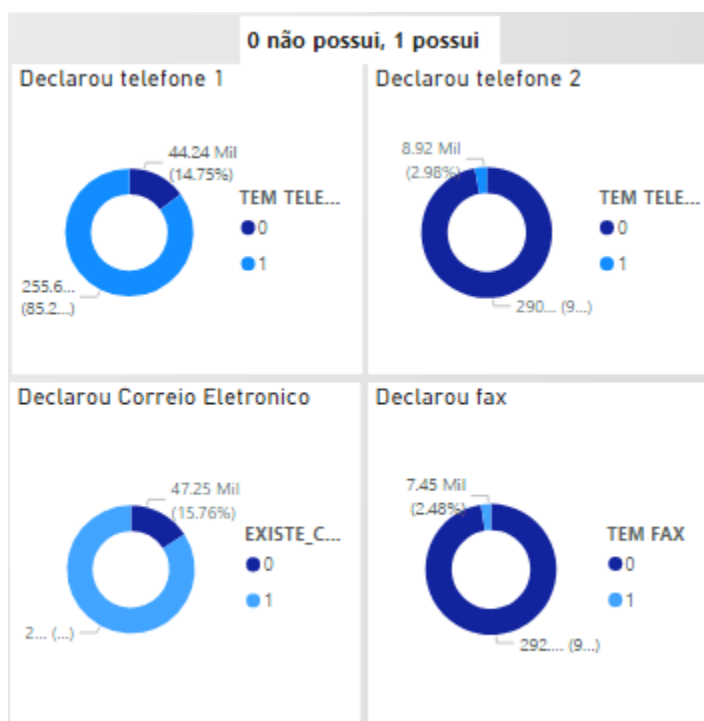
Vemos que a maioria dos estabelecimentos são matrizes. Um percentual muito pequeno (cerca de 3,2%) são filiais.



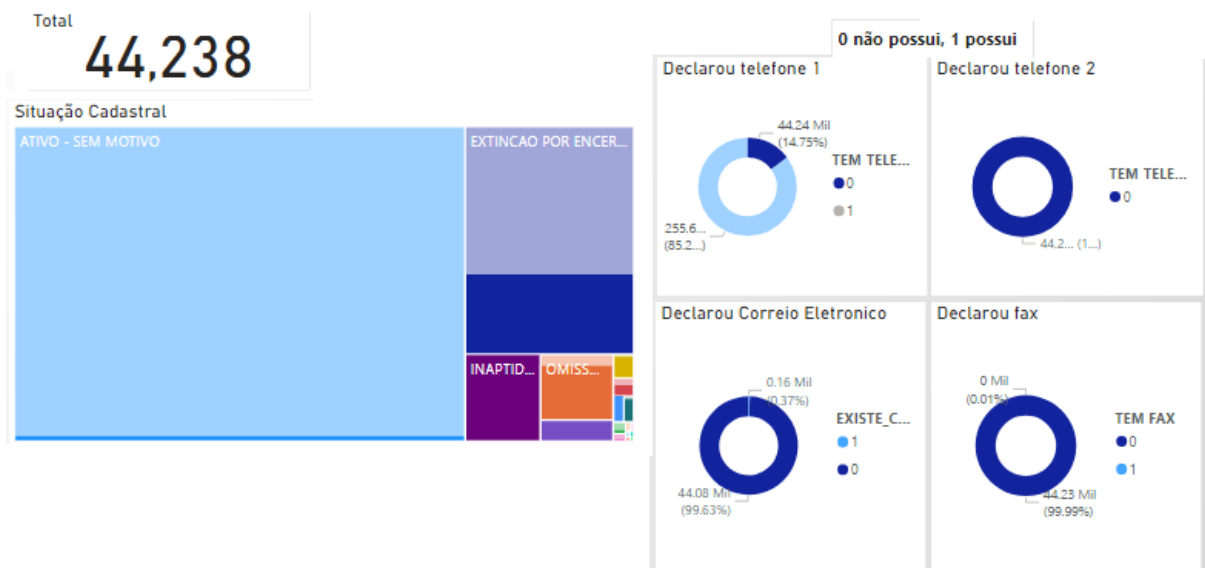
Verificamos que das poucas filiais que existem, a maioria está ligada a bancos. Em segundo lugar, os comércio de roupas, conforme ilustração abaixo:



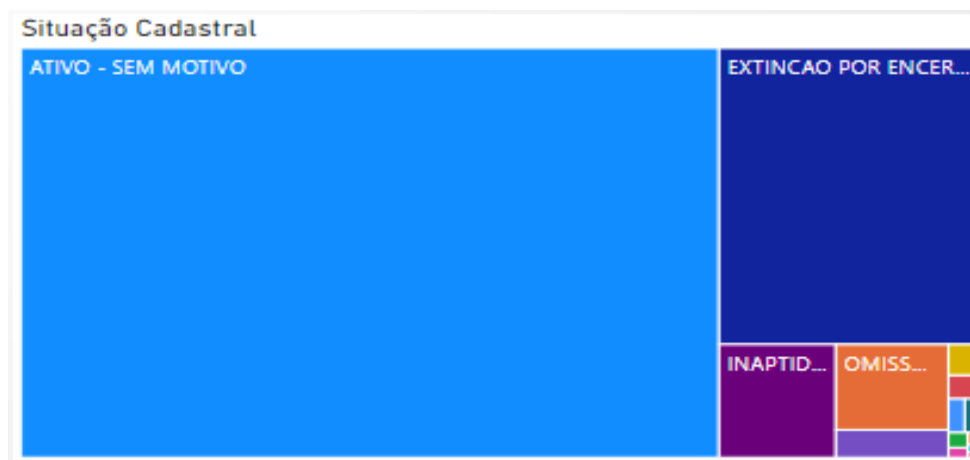
Comparando as colunas de telefone1, telefone2, correio eletrônico e fax, vemos que do total, aproximadamente 85% dos estabelecimentos têm pelo menos 1 telefone, mas não tem um segundo telefone. Poucos estabelecimentos têm fax (menos de 3%) e poucos têm um segundo telefone (também aproximadamente 3%).



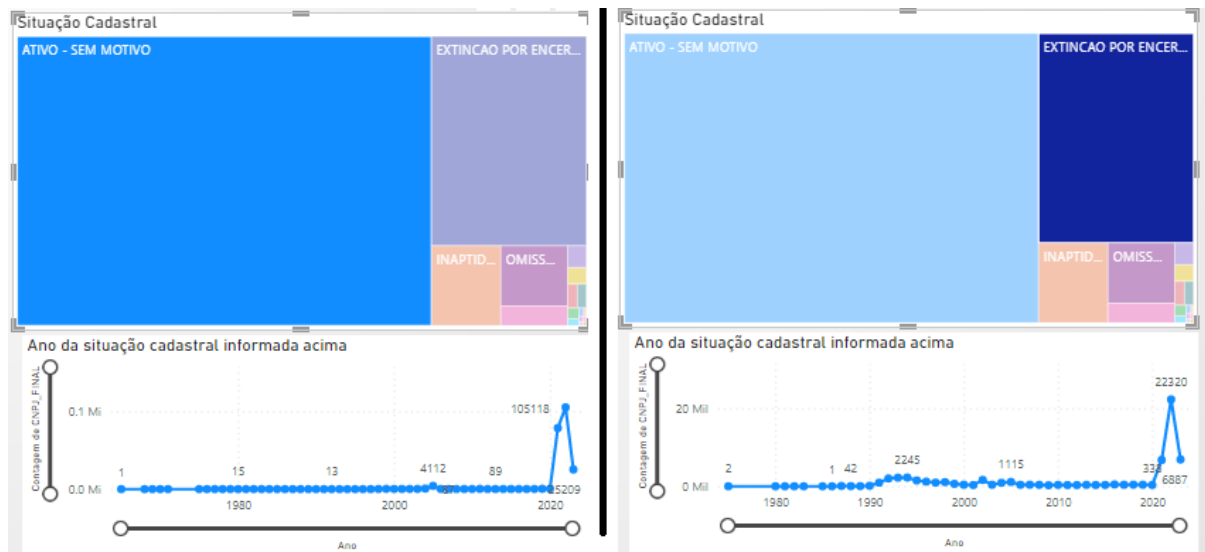
Dos que não tem o telefone 1 cadastrado na base, praticamente não tem telefone 2, correio eletrônico ou fax também. Interessante notar que embora o número dos que não tem telefone seja expressivo (44 mil empresas), vemos que sua situação cadastral é problemática (ou seja, está extinta, com omissão de informações, inapta, etc).



Dentre as situações em que se encontram os estabelecimentos, a grande maioria está regular.



Olhando para a data das situações cadastrais em que a grande maioria dos estabelecimentos se encontram vemos que houve um aumento do número de estabelecimentos ativos regulares e ao mesmo tempo houve um aumento de estabelecimentos que foram extintos, ambos ocorrendo nos anos da pandemia.



Análise de Empresas

Pela coluna `EMPRESA_CNPJ` na tabela `Estabelecimentos`, vemos que há 299.836 registros de empresas ligadas à estabelecimentos (e não 1.418.212, que é o número de estabelecimentos em Minas). Isso ocorre, pois embora cada Estabelecimento tenha obrigatoriamente 1 empresa associada, como só estou usando o arquivo `Empresa0`, e não todo o conjunto de `Empresas`, devido ao enorme volume de dados, nem todos os estabelecimentos de Minas terão sua empresa correspondente. Consequentemente, estou trabalhando apenas com uma amostragem dos dados para o caso das empresas.

Ao mesclar `Empresa0` com `Empresa1`, para fins de teste, vemos que a quantidade de empresas correspondentes aos Estabelecimentos sobe para 397.268 registros.

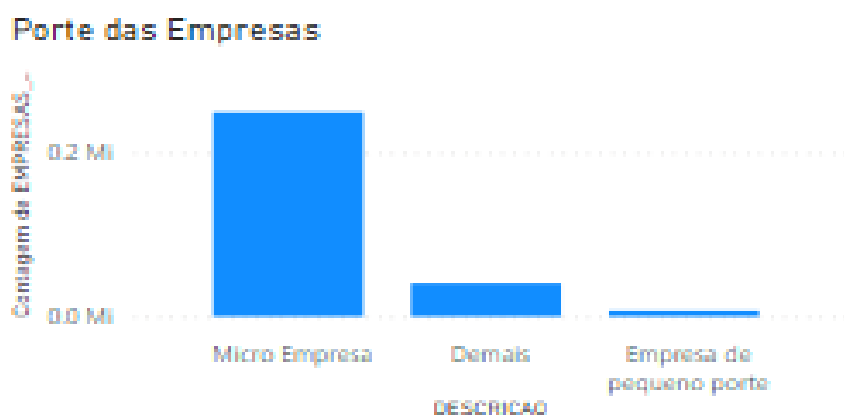
Como uma empresa pode ter mais de um estabelecimento, ao relacionar essas 2 tabelas, algumas empresas se repetem, aparecendo diversas vezes dentro dos 299.836 registros totais. Se considerarmos as empresas que só possuem um estabelecimento, sem nenhuma repetição, serão 293.975 empresas totais. Se considerarmos apenas empresas que possuem 2 ou mais estabelecimentos em seu CNPJ, teremos em torno de 5861 (contando com a matriz de cada uma e suas filiais).

A empresa que tem mais estabelecimentos em sua posse, tem 254 estabelecimentos e é o Bradesco. Vemos em 8º lugar, o MAC Donalds (Arcos Dourados).

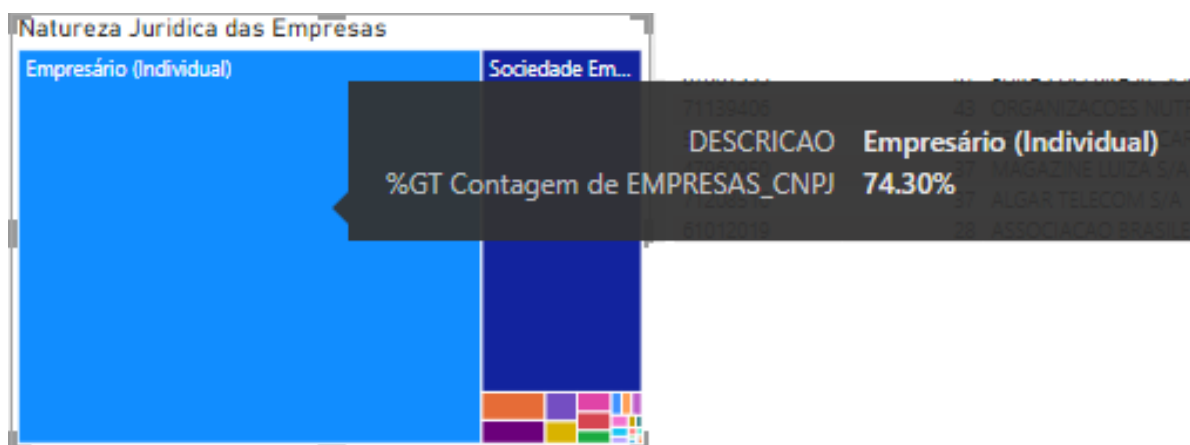
EmpresasCNPJ	Contagem de CNPJ_BASICO	RAZÃO SOCIAL
60746948	254	BANCO BRADESCO SA
62955505	183	IGREJA DO EVANGELHO QUADRANGULAR
60701190	157	ITAU UNIBANCO SA
49930514	88	SODEXO DO BRASIL COMERCIAL SA
67945071	87	SAPORE SA
61585865	66	RAIA DROGASIL S/A
90400888	62	BANCO SANTANDER (BRASIL) SA
42591651	48	ARCOS DOURADOS COMERCIO DE ALIMENTOS SA
87001335	47	PURAS DO BRASIL SOCIEDADE ANONIMA
71139406	43	ORGANIZACOES NUTRI DE REFEICOES COLETIVAS LTDA
51427102	38	TECNOLOGIA BANCARIA SA
47960950	37	MAGAZINE LUIZA S/A
71208516	37	ALGAR TELECOM S/A
61012019	28	ASSOCIACAO BRASILEIRA D'A IGREJA DE JESUS CRISTO
76543115	27	BANCO SISTEMA SA
60643228	26	FIBRIA CELULOSE S/A
61409892	26	COMPANHIA BRASILEIRA DE ALUMINIO
86573136	25	CONSTRUTORA MILIONARIO LTDA
60400075	25	NESTLE BRASIL LTDA
Total	5861	

Todas as demais análises da tabela Empresas serão feitas considerando a ocorrência distinta de empresas, que é de 295.668 (ou seja, as empresas que têm filiais ou que estão atreladas a mais de um estabelecimento diferente, são contadas apenas uma única vez).

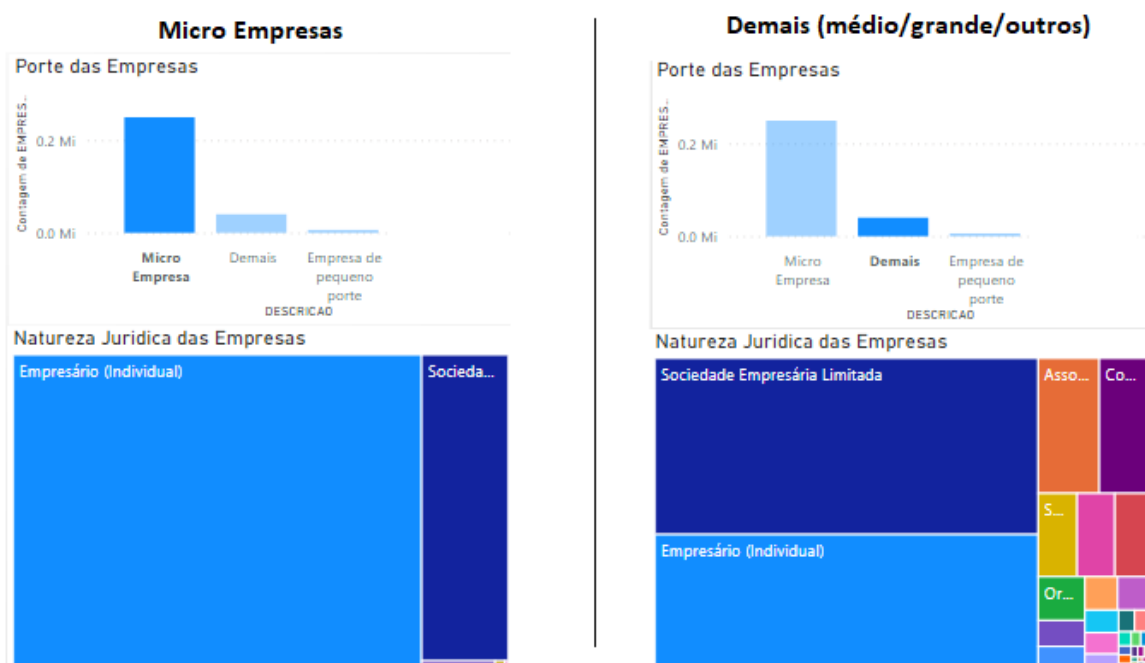
Ao analisarmos a coluna PORTE, vemos que a grande parte das empresas se consideram micro empresas. Em segundo lugar, estão as empresas de médio/grande porte/outros, que estão agrupadas pelo nome de “Demais” e por último, a minoria se considera empresa de pequeno porte.



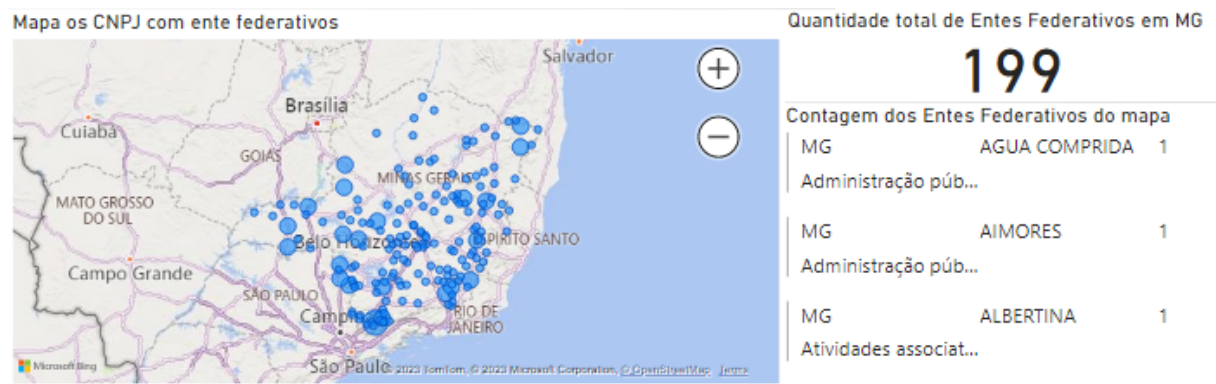
Ao analisar a Natureza Jurídica das Empresas, vemos que 74% são de empresários (individuais), enquanto 22% são de Sociedade Empresária Limitada (que é de 2 ou mais sócios na empresa).



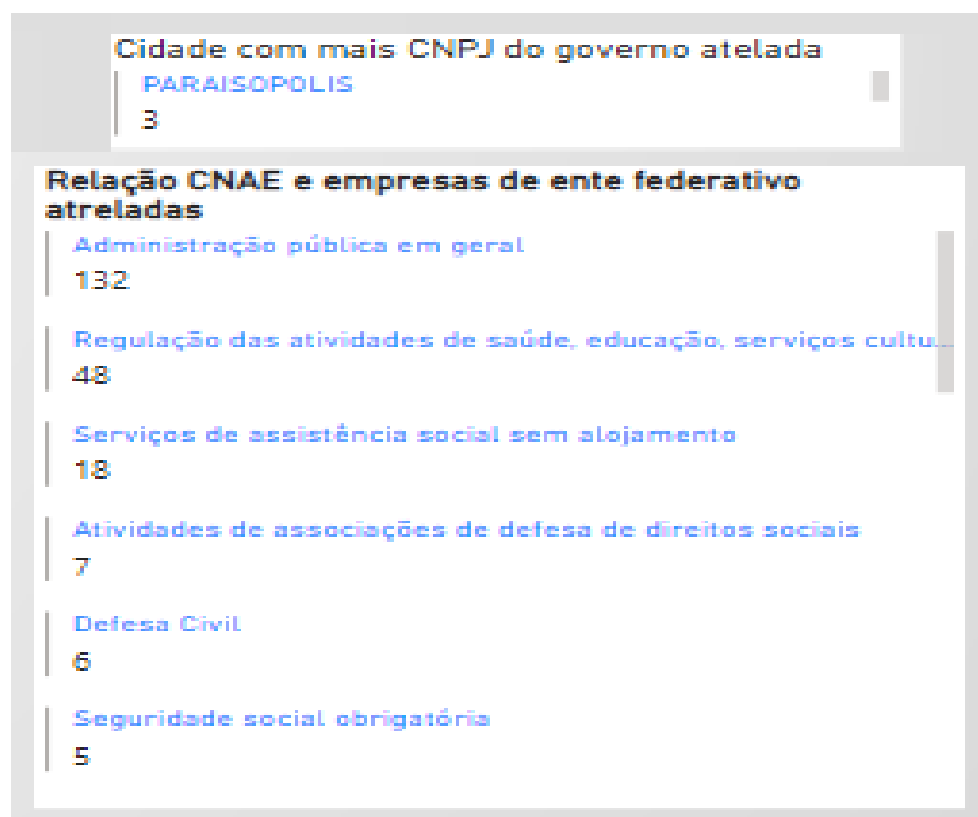
Ao comparar as micro empresas com as médio/grande/outro porte, vemos que para as micro empresas, só se apresentam 2 configurações: ou são Empresários Individuais ou são Sociedade Empresária Limitada. Já para as de médio/grande/outro porte, elas se enquadram em uma gama bem maior de natureza jurídicas, sendo em sua maioria Sociedade Empresária Limitada (aproximadamente 43%).



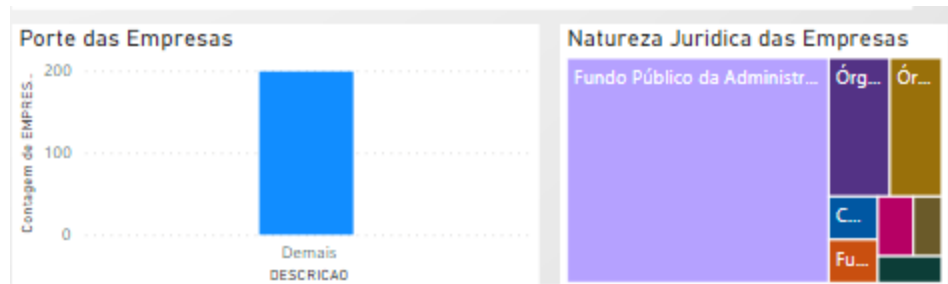
Ao verificar como estão distribuídas as empresas com ente federativo (ou seja, ligados ao governo), vemos que o número de CNPJ com ente é muito baixo (199 no total), também distribuídas por quase todo o território de Minas. Temos uma maior concentração na região sul/sudeste e centro.



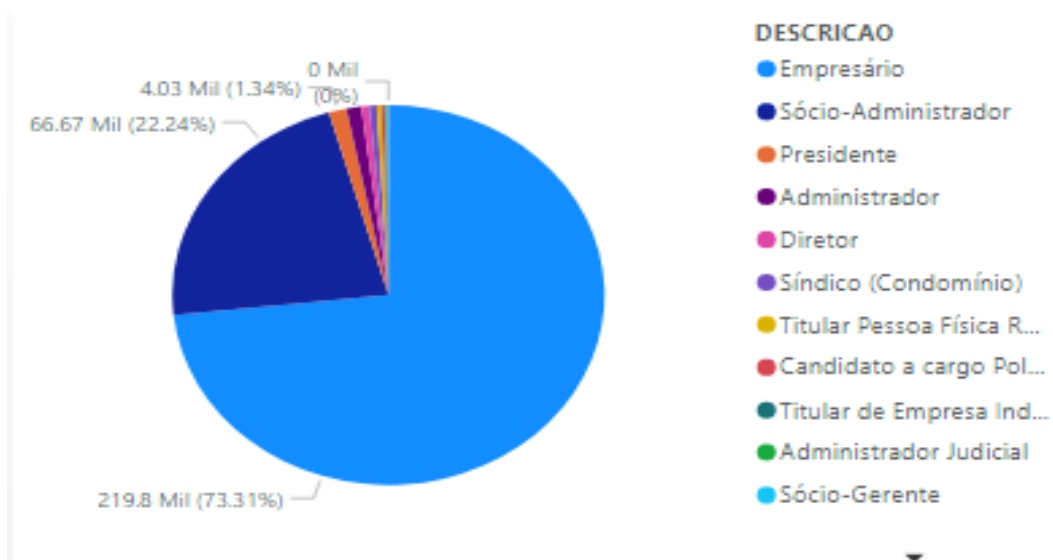
Vemos que a maioria das cidades tem de 1 a 2 CNPJ atrelados ao governo, e a cidade que tem o máximo de CNPJ de ente federativo atrelada é Paraisópolis com 3 CNPJs. Por CNAE, a categoria principal declarada é “Administração Pública em Geral”.



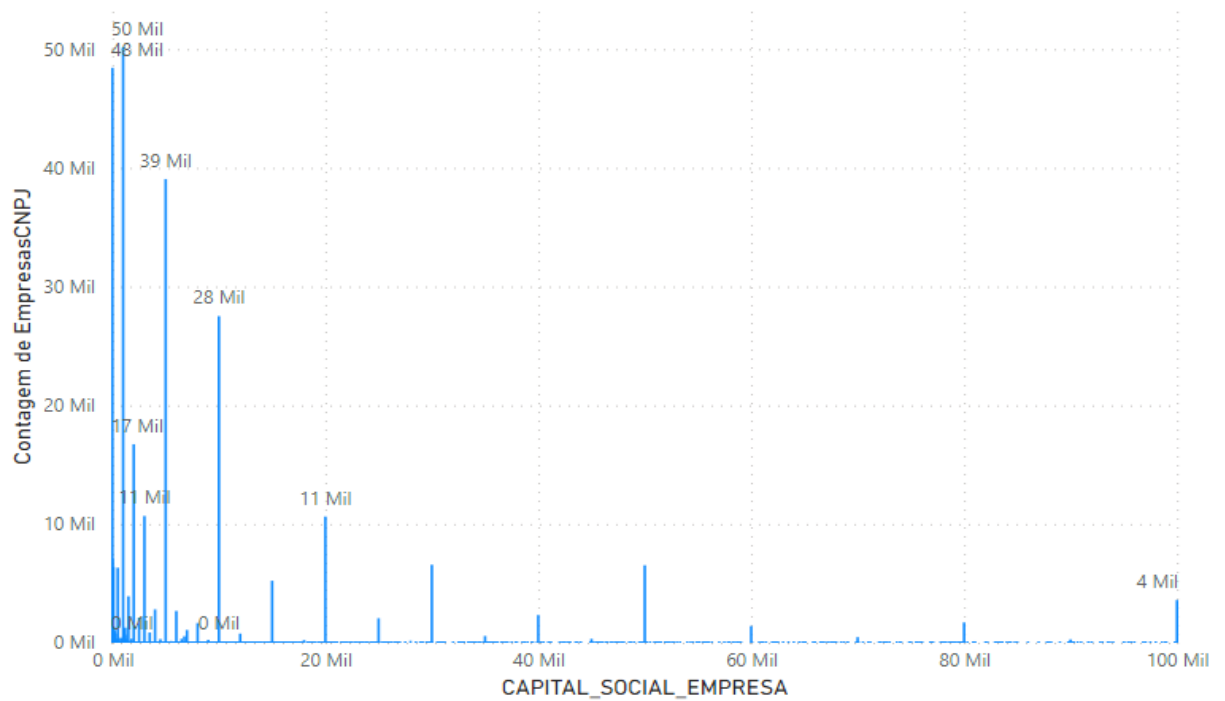
Vemos também que a grande maioria dos CNPJ são de NATUREZA “Fundos Públicos de Administração Municipal” e se categorizam no porte “demais”.



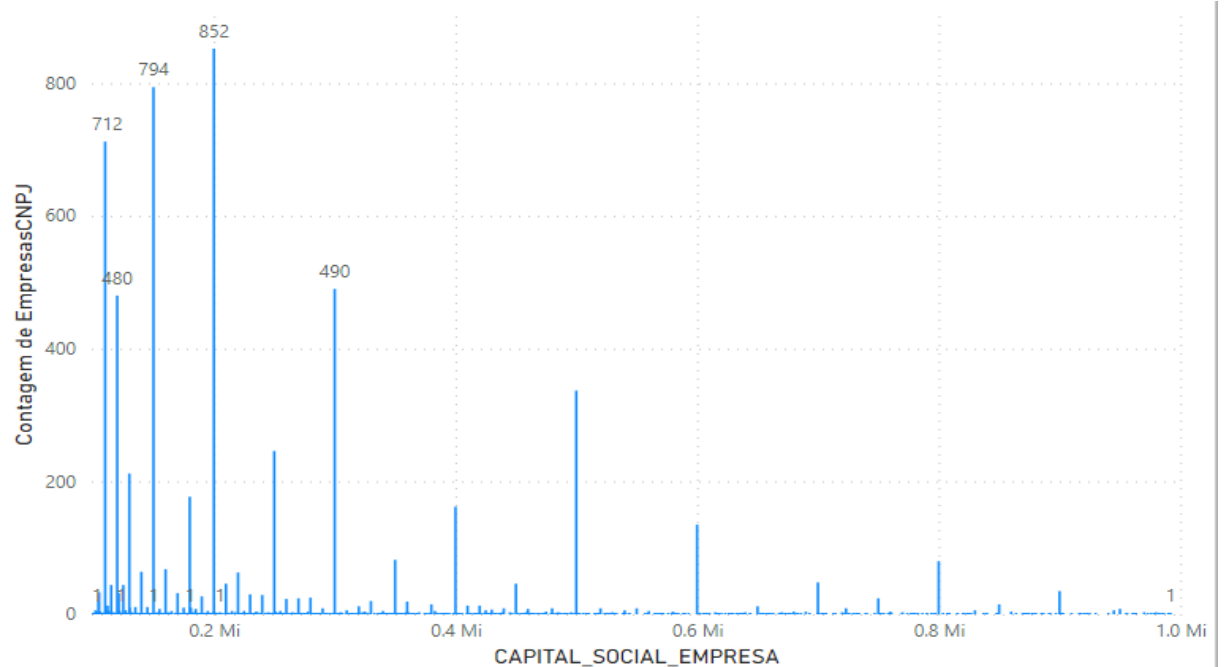
Saindo dos CNPJs com ente federativo e indo para o total de empresas novamente, ao analisarmos o campo Qualificações, vemos que o principal responsável pela empresa geralmente se considera empresário ou sócio-administrador da mesma.



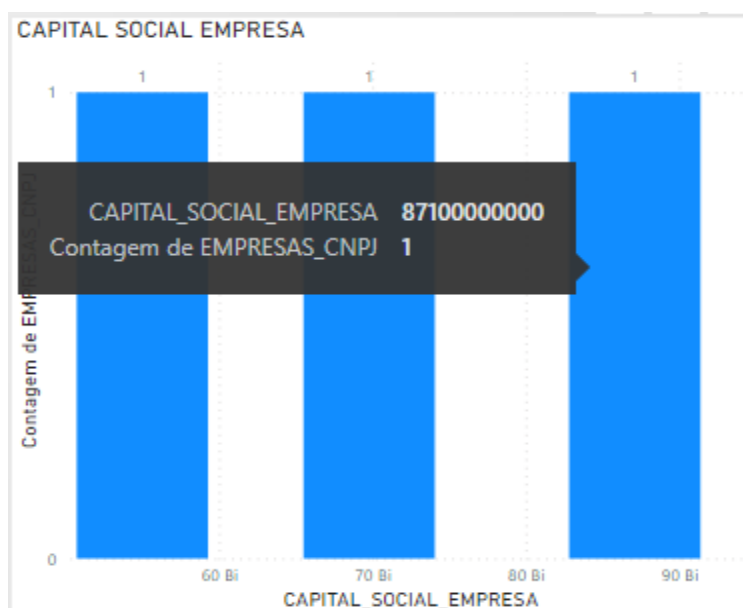
Para melhor analisar a coluna CAPITAL_SOCIAL_EMPRESA dividirei seu gráfico em 3, uma vez que há valores muito discrepantes entre si. Se olharmos o gráfico considerando aportes financeiros igual ou menor que 100 mil, vemos que aportes de valor mil reais ou 5 mil reais são a grande maioria entre as empresas. Um grande número de empresas está com o valor de aporte zerado, o que me leva a inferir que devem ser empresas do tipo MEI, uma vez que essa modalidade não tem obrigatoriedade de informar o aporte feito³.



Analisando empresas cujo aporte foi maior que 100 mil e menor/igual a um milhão, vemos que é baixo o número de empresas com esse aporte (por exemplo, 852 empresas com aporte de 200 mil em contraste com 50 mil empresas com aporte de mil reais). No gráfico abaixo, vemos que o maior número de empresas com aportes entre 100 mil a 1 milhão são as 852 empresas mencionadas anteriormente com aporte de 200 mil.



O total investido em todas as empresas é de R\$835.422.228.102 aproximadamente, e as empresas que mais investiram são o Santander, Bradesco e Itaú, sendo o Bradesco a empresa de maior aporte com um valor de R\$ 87.100.000.000.

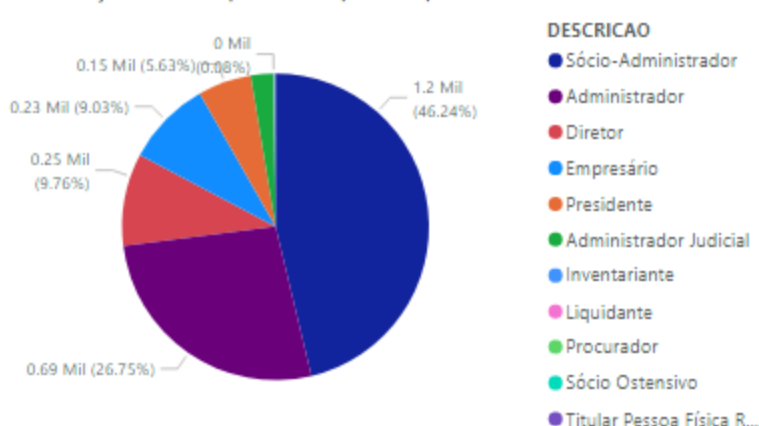


EMPRESAS_CNPJ	Soma de CAPITAL_SOCIAL_EMPRESA	RAZAO_SOCIAL
90400888	55000000000	BANCO SANTANDER (BRASIL) SA
60701190	69783971377	ITAU UNIBANCO SA
60746948	87100000000	BANCO BRADESCO SA
Total	211883971377	

Das empresas com aporte de um milhão para cima, vemos que são 2591 empresas e que deram juntas R\$ 831.280.253.097 de aporte, ou seja, esse valor representa aproximadamente 99% do total de aportes de todas as empresas.

Além disso, ao analisar como se classificam os responsáveis pelas empresas nesse segmento acima de um milhão, vemos que eles se dividem majoritariamente entre sócio-administrador e administrador.

Qualificações dos responsáveis pela empresa



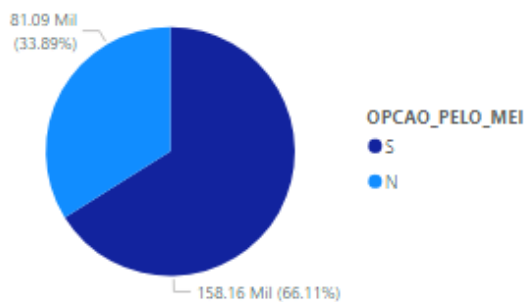
Análise do Simples Nacional

A tabela original possui **34.695.558** registros. Como filtrei ela para aparecer CNPJs cujos números eram maiores que 40.000.000 (já que abaixo desses valores, não existem registros na tabela Empresas, uma vez que o CNPJ_BASE é um número composto por pelo menos 8 dígitos), o total passou para **8.841.560** registros.

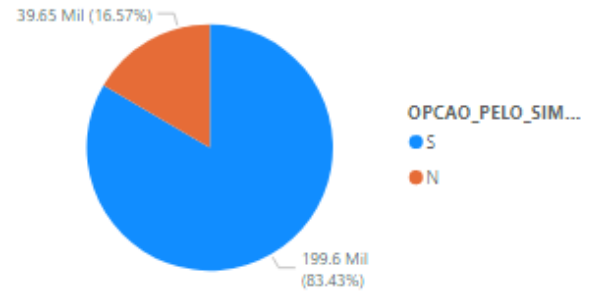
Desses, **239.250** registros são ligados a empresas, e não **295.668** registros, como é a ligação entre Estabelecimentos e Empresas. Isso ocorre, pois embora a tabela Simples seja constituída de apenas um arquivo e não vários como em Empresas e Estabelecimentos (o que pressuporia registros completos do Simples), ainda sim ele não possui todos os registros ligados a todas as Empresas, talvez por algum erro ou problema na fonte, ou mesmo coleta incompleta do governo. Sendo assim, todas as análises serão feitas levando em consideração o total de 239.250 registros da ligação do Simples com Empresas e filtradas por Estabelecimentos.

Vendo os gráficos abaixo, podemos concluir que 66% das empresas são MEI e 83% das empresas estão no Simples Nacional, atualmente. Lembrando que MEI é uma modalidade de empresa e o Simples um regime tributário, então um MEI pode fazer parte do Simples⁵.

Opção pelo MEI pelas empresas

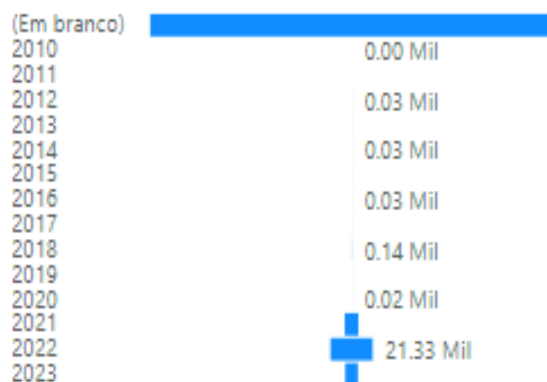


Opção do Simples pelas empresas

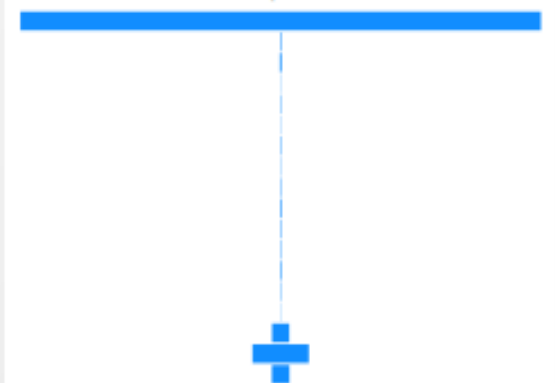


Independente ou não de ser MEI ou fazer parte do Simples, mais de 80% dos dados não possuem **data de exclusão** do MEI ou Simples.

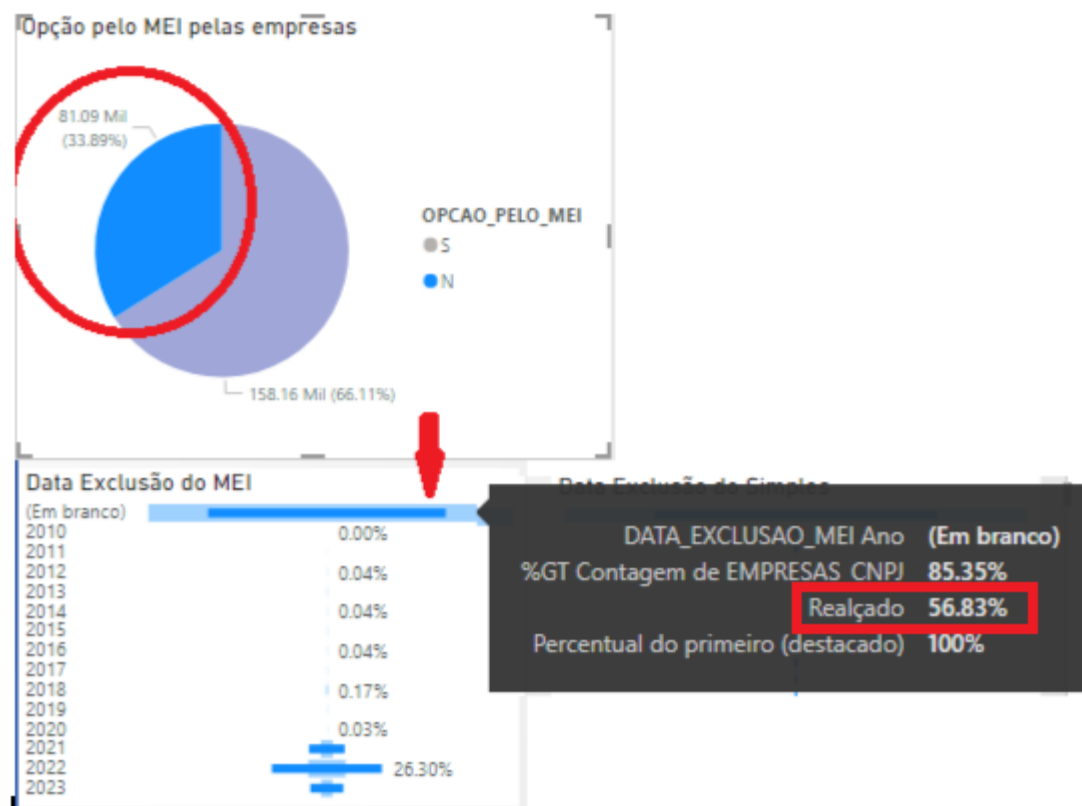
Data Exclusão do MEI



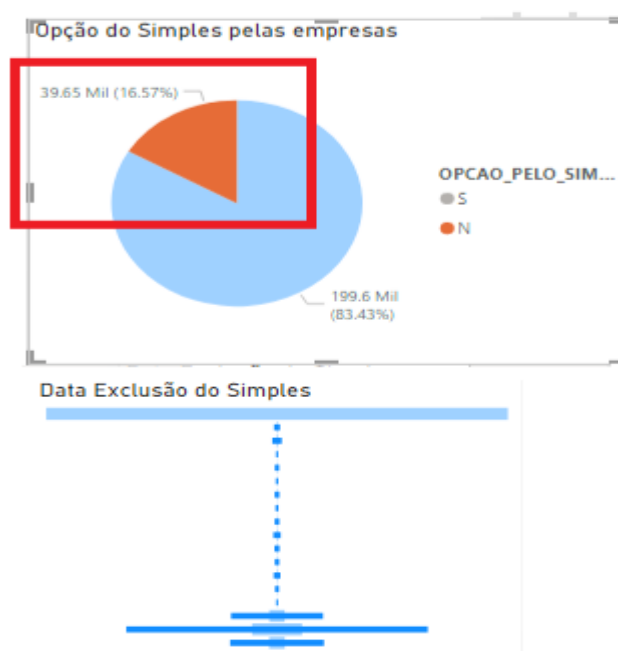
Data Exclusão do Simples



Dos que optaram não ser MEIs, mais da metade não declarou data de exclusão. Isso me faz questionar se os que não optaram ser MEI podem ter sido MEI em algum instante e não o são mais (por isso teriam a data de exclusão) ou nunca nem foram MEI (e sim se encaixaram em outra categoria, portanto nem teriam essa data de exclusão).

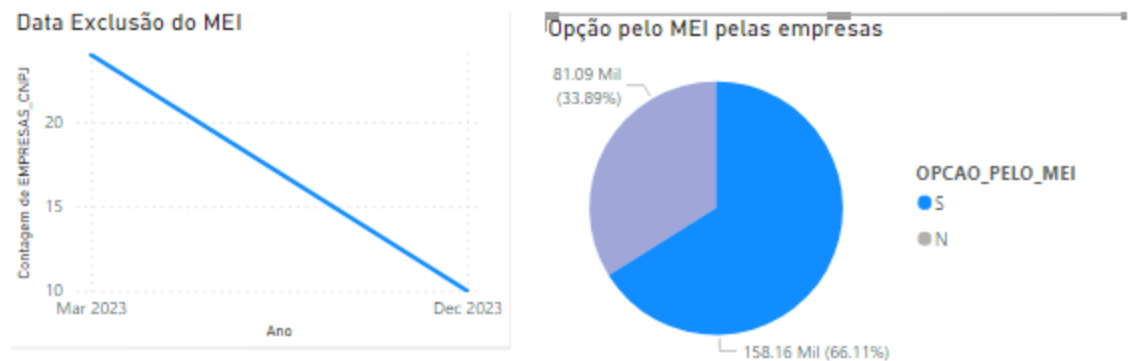


Os que não optaram ser do Simples, todos declararam alguma data de exclusão (abaixo, o traçado azul claro é o valor total de registros nulos, independente da análise. Os pontos em traços em azul mais forte/escuro é que são os valores representativos do Simples para não optantes. Todos os valores azuis mais forte/escuro representam uma data).

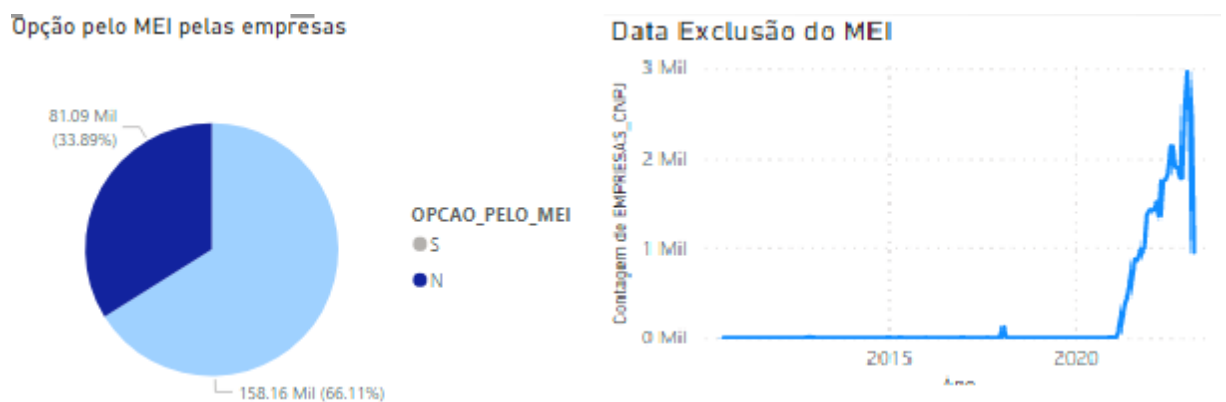


Das **datas declaradas para exclusão** do MEI e do Simples, vemos que:

- 1) Para o MEI, aqueles que se declararam MEI e têm data de exclusão, todas as datas são de Março de 2023 até Dezembro 2023 (lembrando que o empresário pode pedir a baixa do MEI, mas a mesma só ser efetivada posteriormente no governo). Totalizam menos de 50 registros⁶.



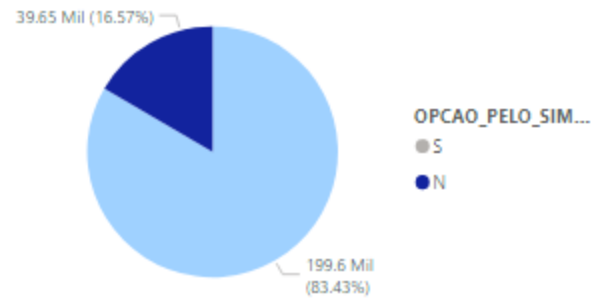
- 2) Já para aqueles que optaram não ser mais MEI, houve um aumento muito grande entre 2021 a 2023, especialmente em janeiro de 2023 (pode ser devido à extinção da empresa, problemas/irregulares com o MEI ou mesmo mudança do MEI para outra categoria).



- 3) Dos que optaram por não ser Simples, há um comportamento semelhante às datas dos que não são MEI, ou seja, aumentou muito em 2020 e teve pico em 2022.

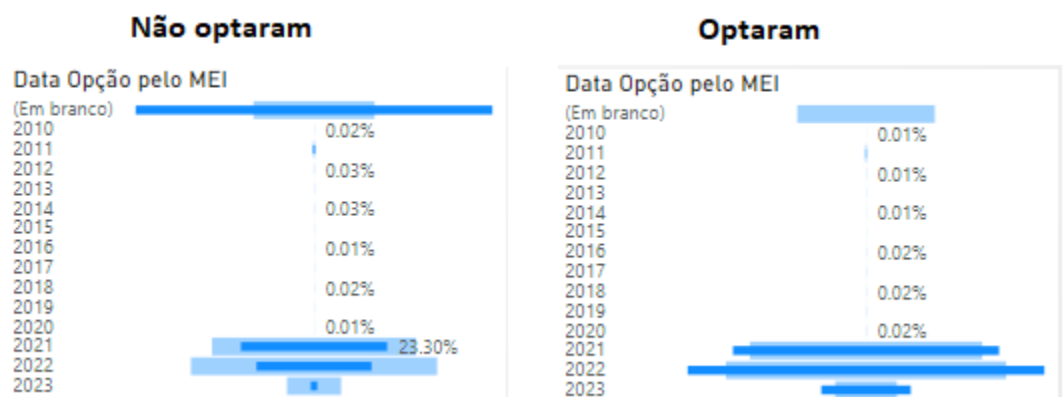


Opção do Simples pelas empresas



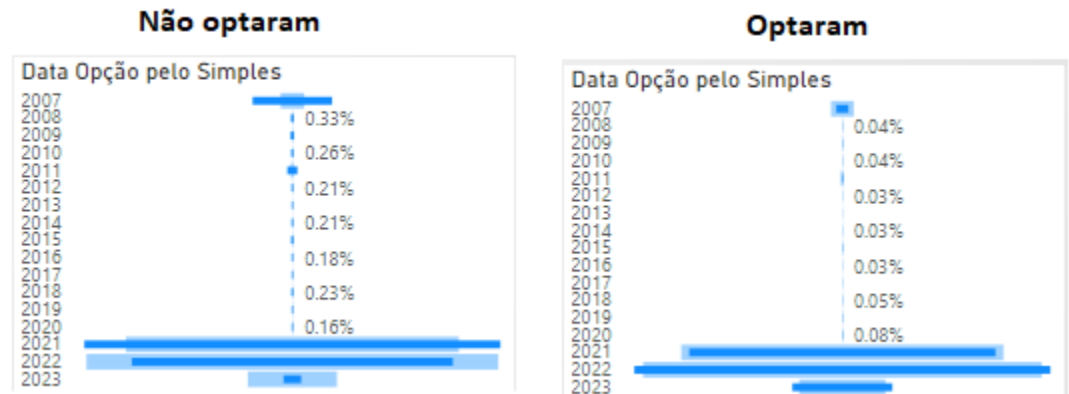
Com relação à data de adesão ao MEI ou ao Simples, vemos que:

- 1) Para o MEI, todos que são declarados MEI têm data da opção, e os que não são há tantos valores nulos/brancos como datas preenchidas. Os que não são, podem ter sido em algum momento MEI e deixado de ser, por isso tem data de opção inicial. Abaixo, o azul escuro representa o valor analisado e o azul claro representa o total sem filtrar (sem colocar se optou ou não).

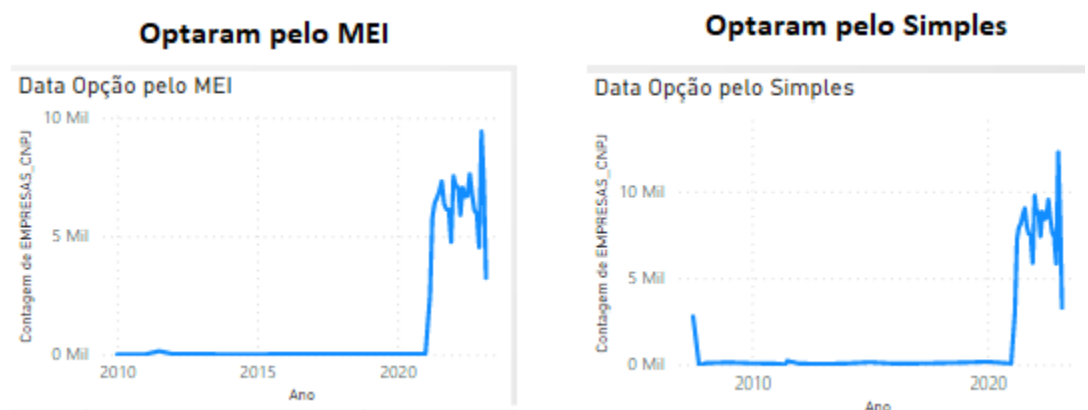


- 2) Já para o Simples, tanto aqueles que optaram como não optaram, todos tem a data de entrada do Simples (poderia ser um erro na fonte? Um valor padrão colocado para todos?). Vemos abaixo que não há a categoria

“nulo/branco/vazio”. A data começa por 2007 e vai até 2023.



Das datas de opção pelo MEI e Simples, vemos que o desenho da onda é parecido em ambos, o que me faz deduzir que quem opta ser MEI muito provavelmente também opta por ser Simples mais ou menos nas mesmas datas:

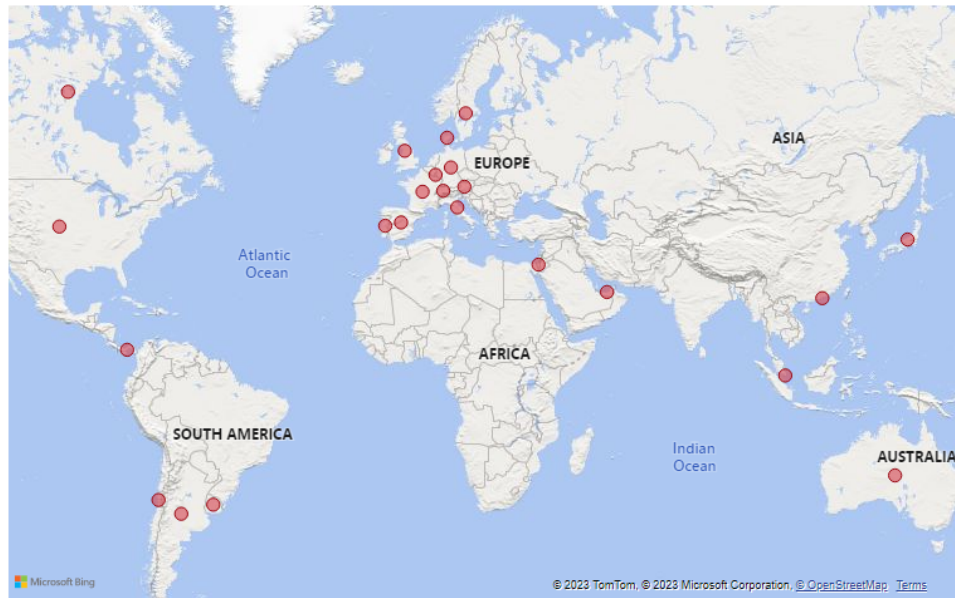


Análise dos Sócios

A tabela Sócios possui 4.377.321 registros, mas filtrados e ligados à Empresa são apenas 65.409 registros.

Desses 65.409 registros, temos sócios com nacionalidade estrangeira, com origem especialmente da Europa, totalizando 28 países totais da onde eles vieram, sendo 123 estrangeiros.

PAISES ONDE OS SÓCIOS ESTRANGEIROS NASCERAM



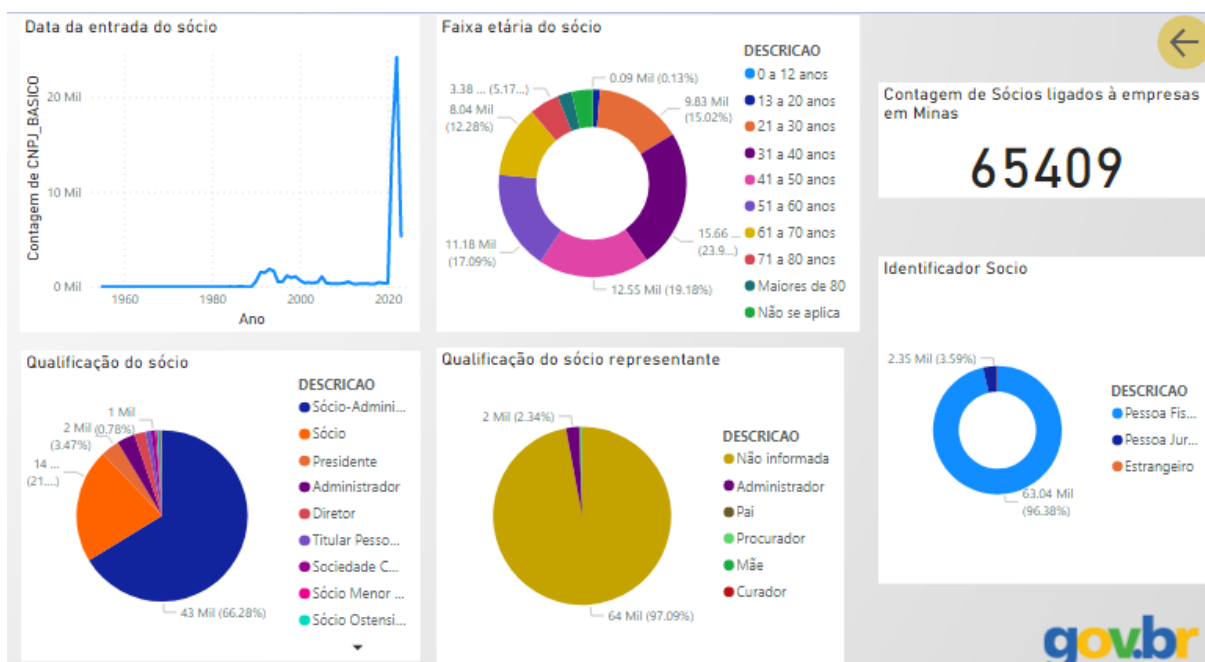
O número de sócios máximo que uma empresa chegou a ter foi de 105 (Sinai Empreendimentos), sendo seguido pelo Bradesco (com 103). A média de sócios por empresa é de 1,67 (2 pessoas por empresa).

RAZAO_SOCIAL	Contagem de NOME_SOCIO_RAZAO_SOCIAL
SINAI EMPREENDIMENTOS IMOBILIARIOS LTDA	105
BANCO BRADESCO SA	103
AZEVEDO SETTE ADVOGADOS	91
ASPEN SEGUROS E PREVIDENCIA LTDA	80
R	45
CONSORCIO FRUTAL GERACAO COMPARTILHADA	36
M	36
G	35
GTEC CAFES LTDA	30
JLT BRASIL CORRETAGEM DE SEGUROS LTDA	30
ADMINISTRA SOLUCOES EM SAUDE LTDA	28
RPM SERVICOS ADMINISTRATIVOS LTDA	28
DIVMED SERVICOS EM SAUDE LTDA	27
GP	27
MULTIMED SERVICOS MEDICOS E HOSPITALARES LTDA	27
XPTO SERVICOS MEDICOS LTDA	26

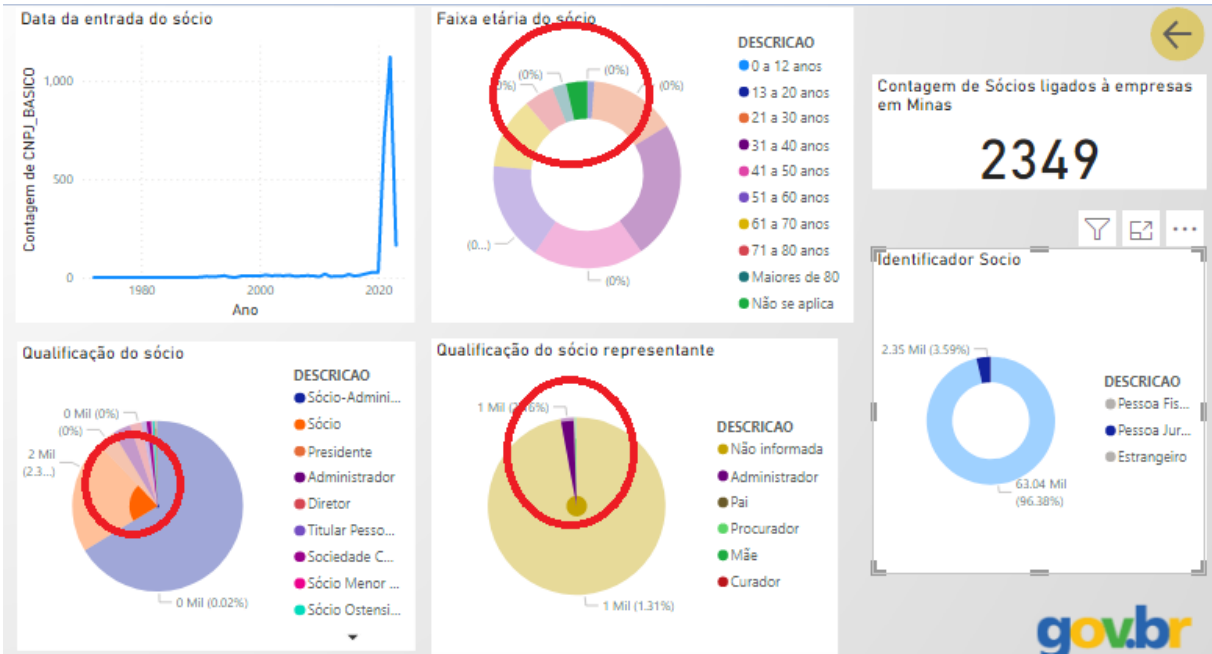
Vemos também que o número de empresas que um mesmo sócio participa é em média 1 empresa/sócio, sendo que o máximo que um sócio participou de empresas foi em 18 empresas, como mostra a figura abaixo:

NOME_SOCIO_RAZAO_SOCIAL	Contagem de NOME_SOCIO_RAZAO_SOCIAL
LEANDRO RODRIGUES MARINO	18
EBES SISTEMAS DE ENERGIA SA	16
FINCO ASSESSORIA FINANCEIRA LTDA	16
LUIZ OTAVIO FONTES JUNQUEIRA	16
OSVALDO ANTUNES CRUZ JUNIOR	16
RODOLFO MOLINARI FILHO	16
ROGERIO MARCHINI SANTOS	16
SURYA GUEDES MENDONCA	16
LCM CONSTRUCAO E COMERCIO S.A	14
CONATA ENGENHARIA LTDA	13
INFRACON ENGENHARIA E COMERCIO LTDA	13
Total	65409

Analisando os Sócios, vemos que grande parte deles entrou em sociedade nos anos 2020, tendo o pico em 2022. A faixa etária desses sócios é de 21 a 70 anos (especialmente entre 31 a 40), eles se qualificam na sua maioria como sócios-administrador ou sócios, são pessoas físicas e não possuem outro sócio representando eles.



Ao analisar os sócios pessoa jurídica vemos que são muito poucos (2349), não se aplica a categoria de idade para esses e que os seus representantes se consideram administradores.



5. Interpretação dos Resultados

Estabelecimentos

Ao analisar o CNAE secundário, vemos que o CNAE de comércio varejista de vestuário (comércio que vende para o consumidor final) é o mais cadastrado entre os estabelecimentos. Infiro que essa realidade pode se referir a comércios de multiprodutos, por exemplo, uma loja que vende roupas também vende perfume, ou uma loja que vende calçados também tem artigos de vestuário. A diversificação de itens daquele que era proposto originalmente pelo estabelecimento é uma realidade crescente, e o consumidor final acaba consumindo mais naquele estabelecimento, uma vez que ele tem de tudo.

Com relação à localidade, vemos que a maioria dos estabelecimentos se encontram na capital e não há muitos estabelecimentos especialmente na região noroeste de Minas, o que pode ser uma oportunidade para um empreendedor tentar montar um estabelecimento diferenciado ou com propósito diversificado nessa região, lucrando com a população das cidades do interior.

Há mais estabelecimentos do tipo matriz do que filiais espalhados em Minas. Pode parecer estranho, uma vez que temos por exemplo, diversos estabelecimentos “iguais” (a exemplo, de um MAC Donalds, em vários bairros da cidade). Não devemos confundir filiais (de um mesmo dono/CNPJ) com franquias (donos diferentes), que neste trabalho não serão analisadas a fundo. O CNAE com maior número de filiais é de bancos, análise essa reforçada quando olhamos quantos estabelecimentos estão ligados à empresas (na aba de análises de Empresas), em que vemos que o Bradesco possui 254 estabelecimentos ligados a ele, sendo a maior empresa a possuir estabelecimentos, considerando a análise com filtro em Minas Gerais.

Quase todo estabelecimento prioriza ter pelo menos um telefone, e se ele não tem telefone, ele também não terá correio eletrônico, fax nem outros meios de comunicação dos analisados. Desses que não tem nem um telefone, infiro pela análise gráfica demonstrada anteriormente que esses estabelecimentos não estão em funcionamento regular em sua maioria (podendo até nem estar funcionando mais, porém ainda tendo o registro de CNPJ).

Verifiquei que havia um grande número de estabelecimentos abrindo e ao mesmo tempo fechando na pandemia, e o principal motivo de fechamento dos mesmos era “extinção por encerramento/liquidação voluntária”. Essa abertura e fechamento foi acentuada em tempos de pandemia, começando em 2020. A abertura de novas empresas/estabelecimentos em tempos de pandemia, pode ser explicada pela mudança na fonte de renda do trabalhador, devido ao desemprego em massa. Com essa crescente demissão das empresas, essas pessoas precisaram encontrar novos meios de sustento, portanto, abrem micro/pequenas empresas para trabalhar de forma independente e regular como empreendedores¹.

Por fim, infiro que setores como de varejo de roupas e promoção de vendas tiveram um crescimento expressivo em 2020 e em diante especialmente pelo uso da internet, por exemplo, quando o varejo de roupas agora focava em vendas em seus sites, e-commerces e outros, sem que o cliente precisasse sair de casa.

Empresas

Com relação ao cenário das empresas em si, vemos que a grande maioria se consideram microempresas e poucas empresas detêm múltiplos estabelecimentos (5861 empresas detêm 2 ou mais estabelecimentos de um total de 295.668 empresas distintas).

A grande maioria é de natureza jurídica de Empresários Individuais (74%) e para empresas de médio a grande porte (ou aquelas com características especiais⁴), a natureza jurídica é variada, mas a maioria são de Sociedade Empresária Limitada.

Já os CNPJs com ente federativo (ligadas ao governo), são bem poucas, espalhadas por todo o estado, numa proporção em média de 1 CNPJ por estado, com CNAE de "administração pública" em grande maioria e maioria de natureza jurídica de “Fundos Públicos de Administração Municipal”.

Em geral, a qualificação do responsável pela empresa é de empresário (73%) ou sócio-administrador (22%). Em empresas com aportes acima de 1 milhão de reais, os responsáveis são qualificados como sócio-administradores ou administradores.

A maioria dos aportes iniciais das empresas é de mil e 5 mil reais. As empresas com maior volume de aporte inicial são os bancos e o valor dado pelas 2591 empresas com maior aporte são equivalentes à 99% do total de aportes de todas as empresas do estado de Minas.

Simples Nacional

Sobre o Simples e o MEI, vemos que a maioria é MEI (66%) e a grande maioria usa o Simples (83%).

Uma empresa pode começar sendo MEI, pode deixar de ser MEI a medida do tempo por diversos fatores ou nunca nem ter sido MEI (se encaixando desde seu início em outra categoria). Assim sendo, uma mesma empresa pode ter declarado ser MEI e ter data de exclusão do MEI (uma vez que não necessariamente é imediato, ou seja, posso pedir baixa hoje e ser oficializado pelo governo daqui um mês⁶), ou mesmo não estar declarada como MEI e ter data de adesão ao MEI (pelo mesmo motivo acima citado).

Por algum motivo, é inserido a data de opção pelo Simples nas empresas (o que pode me levar a crer que há muito incentivo para as empresas aderirem ao Simples, ou “o Simples já começaria como requisito básico da abertura da mesma”, então mesmo que uma empresa não esteja usando do Simples, ela consequentemente poderá ter a data de exclusão do Simples.

A data de opção tanto para MEI como para Simples teve um crescimento enorme nos anos de 2020 para frente, muito provavelmente pela quantidade de novas empresas/estabelecimentos abertas (conforme já elucidado na conclusão de Estabelecimentos, acima). O gráfico de ambos tem formato semelhante, o que me faz deduzir que muitas empresas abriram MEI e aderiram ao Simples, pela facilidade e vantagens que esse regime tributário proporciona.

Sócios

Vemos que a média de sócios por empresa é de 1 a 2 sócios, e a média do mesmo sócio em diferentes empresas também é 1 o que condiz com as empresas da modalidade

MEI (que não podem ter mais de um sócio), que vimos em evidência no tópico do Simples Nacional.

Vemos também que a idade dos sócios é variada, desde 21 a 70 anos, porém com ênfase entre 31 a 40 anos.

Vemos que o cenário brasileiro é bem marcado por MEIs ou pequenas empresas, e por ser assim, geralmente os sócios não são pessoas jurídicas e sim físicas.

6. Links

Link de acesso às bases de dados e dicionário temas deste trabalho:

dados.gov.br/dados/conjuntos-dados/cadastro-nacional-da-pessoa-juridica---cnpj

Link de acesso aos gráficos e análises do POWER BI:

<https://app.powerbi.com/view?r=eyJrIjoibWE3NDczNTMtNmJhZi00N2FkLWJiZDAzM2E4ZTJhYmZlNmNliiwidCI6IjE0Y2JkNWE3LWVjOTQtNDZiYS1iMzE0LWNjMGZjOTcyYTE2MSIsImMiOjh9&pageName=ReportSection060793dbb10507000c23>

Link de acesso ao arquivo bruto de tratamento dos dados do POWER BI:

<https://drive.google.com/file/d/1G-HlqIH3sckAOtqACu385jRfHcyLNjpr/view?usp=sharing>

REFERÊNCIAS

¹ BUSSOLA. **Brasil teve aproximadamente 4 milhões de CNPJS abertos em 2022.** Publicado em 2023.

exame.com/bussola/brasil-teve-aproximadamente-4-milhoes-de-cnpj-abertos-em-2022

² MINISTÉRIO DA ECONOMIA. **Mais de 1,3 milhões de empresas são criadas no país em quatro meses.** Publicado em 2022.

gov.br/economia/pt-br/assuntos/noticias/2022/julho/mais-de-1-3-milhao-de-empresas-sao-criadas-no-pais-em-quatro-meses

³ GULARTE, Charles do CONTABILIZEI. **O que é o Capital Social de uma Empresa e Como Definir o Valor.** Publicado em 01/2022.

www.contabilizei.com.br/contabilidade-online/o-que-e-capital-social-e-como-definir-o-valor

⁴ GULARTE, Charles do CONTABILIZEI. **Porte de Empresa: Quais são as Classificações? Veja como é definido nas diferentes instituições.** Publicado em 01/2022.

www.contabilizei.com.br/contabilidade-online/porte-de-empresa

⁵ FACILITE TECNOLOGIA. **Simples Nacional e MEI: Descubra quais são as suas diferenças de uma vez por todas.** www.facilite.co/simples-nacional-mei

⁶ GULARTE, Charles do CONTABILIZEI. **Desenquadramento MEI: Quais as opções? Veja como fazer.**

Publicado em 05/2023. www.contabilizei.com.br/contabilidade-online/desenquadramento-mei/

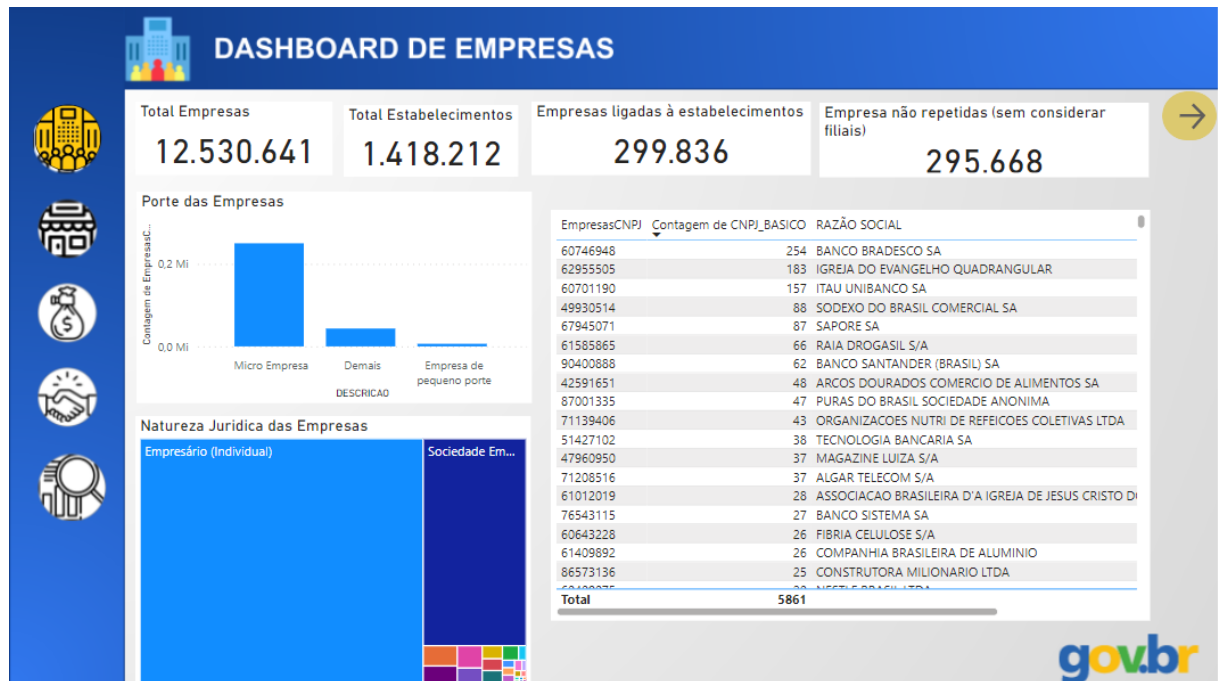
Canal do HashTag Treinamentos sobre POWER BI no Youtube.

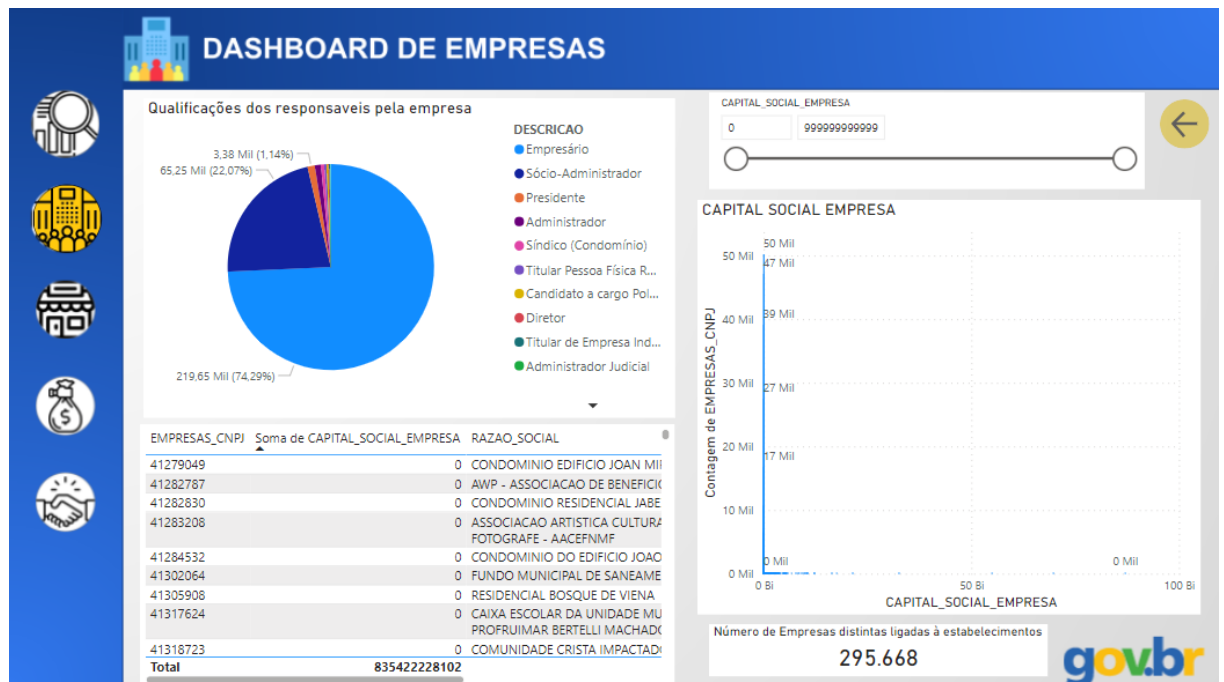
www.youtube.com/@HashtagTreinamentos

APÊNDICE

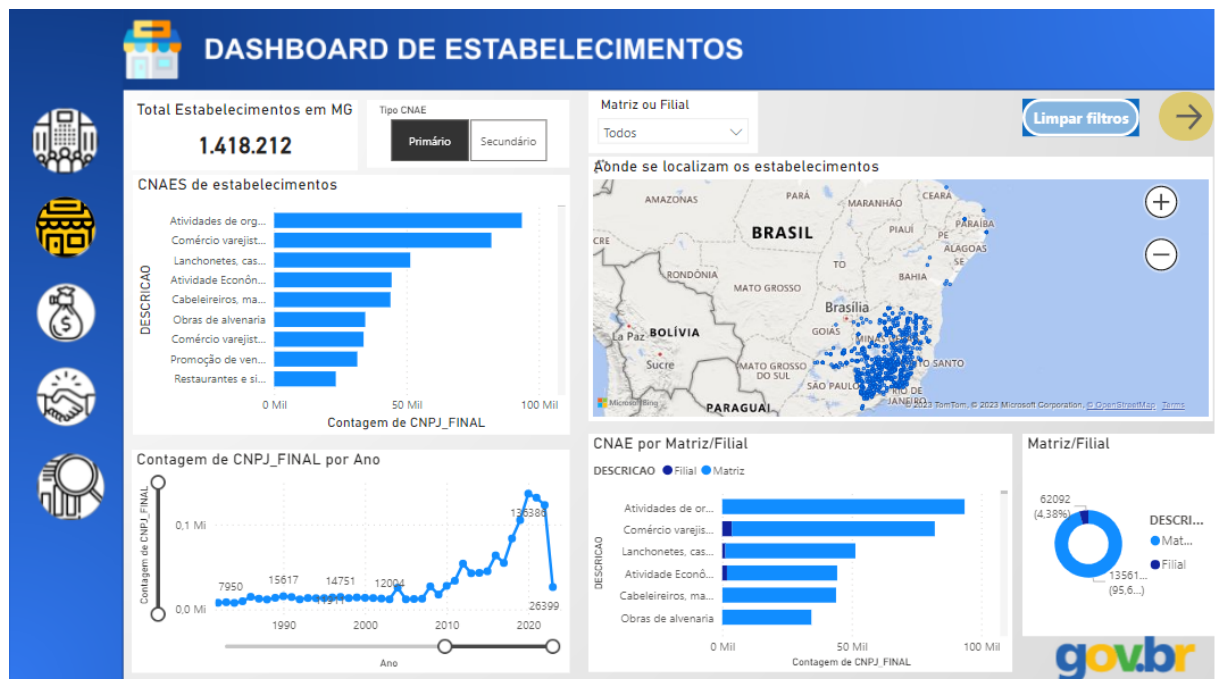
Caso o link do trabalho no POWER BI não esteja mais acessível externamente, seguem abaixo as telas criadas com as análises gráficas:

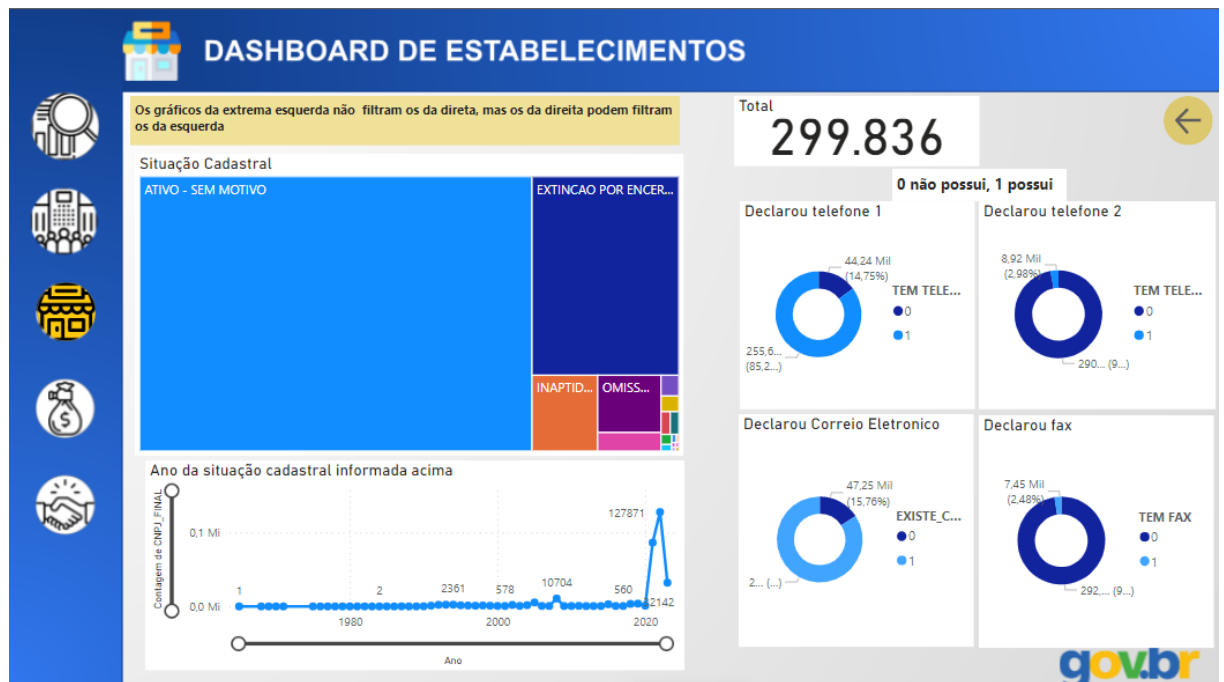
1. Telas relativas à Empresas



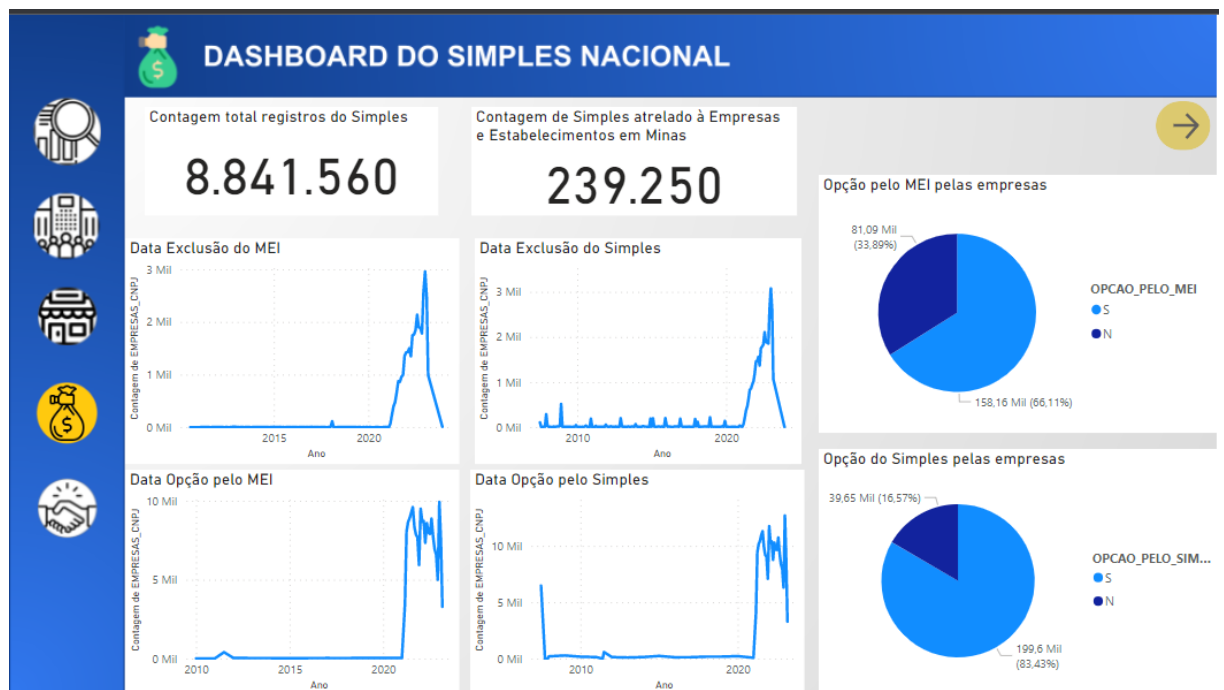


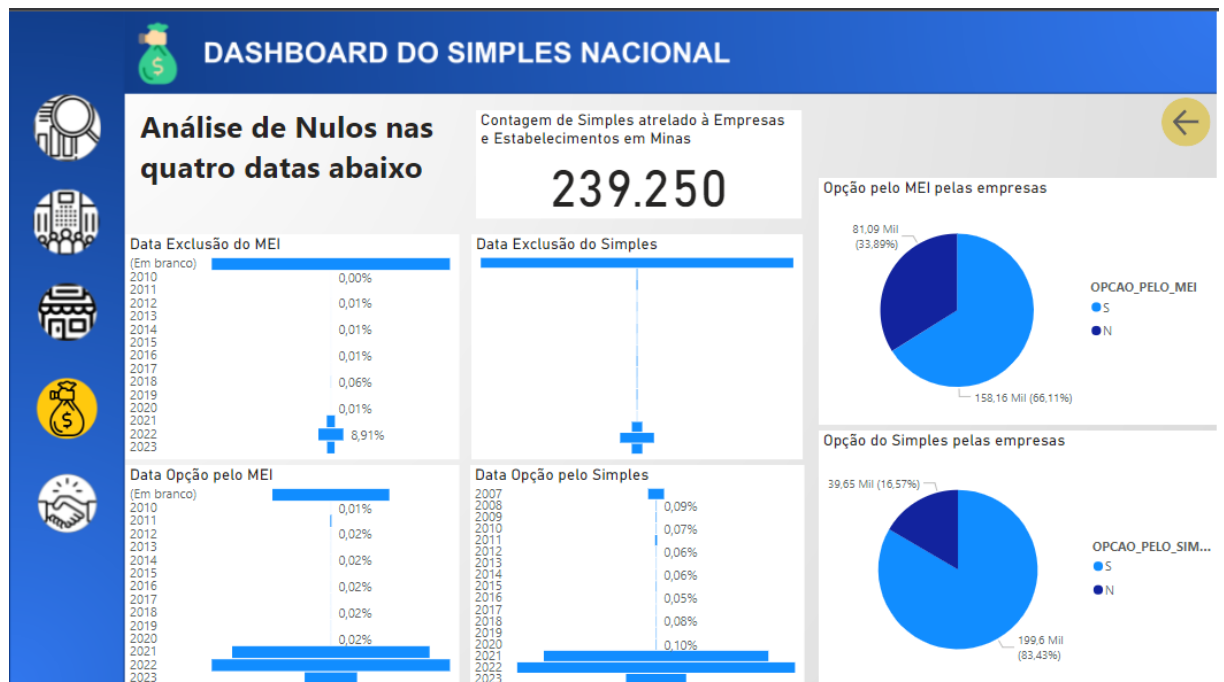
2. Telas relativas à Estabelecimentos





3. Telas relativas ao Simples Nacional





4. Tela relativa aos Sócios

