Análise exploratória do SireneJud

ABJ

04/11/2022

Este relatório mostra uma análise exploratória de dados (*Exploratory Data Analysis*, EDA) da base de dados do SireneJud. A base foi baixada no dia 04/09/2022, a partir de um arquivo CSV, contendo 588.197 linhas e 26 colunas. A base foi baixada a partir do [link de dados abertos da ferramenta do SireneJud](https://sirenejud.cnj.jus.br/mapa/geral).

## Sobre a base de dados

A base de dados do SireneJud apresenta uma estrutura comum de bases de dados judiciais, contendo três unidades amostrais distintas: processos, partes e movimentações.

Os processos apresentam as seguintes colunas:

* FID: ID único, linha da base.
* geom: geometria do ponto de origem (latitude e longitude).
* ano: ano de distribuição do processo.
* uf: unidade federativa de distribuição do processo
* municipio: municipio de distribuição do processo
* grau: grau originário do processo (G1: primeiro grau, G2: segundo grau, JE: juizado especial, TR: turma recursal, TRU: turma recursal unificada).
* esfera: federal ou estadual.
* tribunal: tribunal de origem do processo
* orgao: órgão julgador (vara, câmara etc)
* porte: porte do tribunal, segundo classificação do CNJ disponível no relatório Justiça em Números
* classe: classe do processo segundo a SGT.
* assunto: códigos e nomes dos assuntos dos processos segundo a SGT
* datas: início do processo, julgamento e baixado. Caso o processo esteja ativo, a data de julgamento ou baixado é vazia.
* tempo: tempo de tramitação até o dia de extração dos dados.
* julgamento: indicador de processo concluído ou em andamento.
* complexo: indicador de processo de alta complexidade.

Os dados de partes apresentam as seguintes colunas:

* polo: ativo ou passivo
* nome: nome da parte
* sexo: sexo da parte (se pessoa física)
* tipo: jurídica ou física
* documento: CNPJ da parte, se juridica, e vazia caso contrário
* nacionalidade: nacionalidade da parte
* assistencia: assistência judiciária gratuita
* intimacao: intimação pendente

Os dados de movimentações apresentam as seguintes colunas:

* data: data da movimentação
* grau: grau da movimentação (primeiro grau, segundo grau, etc)
* nome\_situacao: nome da situação
* fase: fase processual
* tipo: tipo de procedimento
* natureza: natureza do procedimento
* nome: nome do movimento
* julgador: nome do julgador
* criminal: indicador de criminal

No momento, vamos analisar somente os dados de **processos** e **partes**. A análise das movimentações ficará para um estudo futuro.

## Filtros

O primeiro passo da análise é realizar filtros para obter uma base que faz parte do escopo da pesquisa. O escopo da pesquisa ainda não está totalmente definido no momento de escrita do presente relatório, mas vamos fazer algumas escolhas para facilitar a análise.

Os filtros aplicados no momento são:

* Apenas processos originários no primeiro grau ou segundo grau (remover juizado especial e turma recursal).
* Remover processos de alta complexidade
* Remover duplicatas de número de processo (ordenamos por grau antes de tirar as duplicatas)

## Informações básicas

Do total de 353385 processos, 293987 (83%) são estaduais. Dentre os processos estaduais, o tribunal mais comum é o TJSP, com 57129 (19.4%) dos processos estaduais.

As 10 classes mais comuns dos processos estão na [Tabela 1](#tbl-classe). Como as classes podem variar conforme a fase do processo (e.g. inquéito policial, cumprimento de sentença), é preciso tomar cuidado na hora de realizar filtros por essa variável.

Tabela 1: Dez classes mais comuns.

| sgt\_nm\_classe | n | prop |
| --- | --- | --- |
| Inquérito Policial | 73437 | 20.78% |
| Ação Civil Pública | 47544 | 13.45% |
| Procedimento Comum Cível | 41865 | 11.85% |
| Ação Penal - Procedimento Ordinário | 32487 | 9.19% |
| Crimes Ambientais | 23793 | 6.73% |
| Agravo de Instrumento | 20113 | 5.69% |
| Termo Circunstanciado | 16316 | 4.62% |
| Ação Penal - Procedimento Sumário | 15976 | 4.52% |
| Apelação Cível | 11654 | 3.30% |
| Cumprimento de sentença | 10879 | 3.08% |

Os 10 assuntos mais comuns estão na [Tabela 2](#tbl-assunto). É possível notar que assuntos genéricos como “Dano Ambiental” e “DIREITO AMBIENTAL” são bastante presentes nos dados.

Tabela 2: Dez assuntos mais comuns. Um processo pode ter mais de um assunto.

| sgt\_nm\_assunto | n | prop |
| --- | --- | --- |
| Dano Ambiental | 69078 | 18.54% |
| Crimes contra a Flora | 64439 | 17.30% |
| Crimes contra o Meio Ambiente e o Patrimônio Genético | 50652 | 13.59% |
| Crimes contra a Fauna | 29993 | 8.05% |
| Da Poluição | 23550 | 6.32% |
| Revogação/Anulação de multa ambiental | 17227 | 4.62% |
| Flora | 15976 | 4.29% |
| DIREITO AMBIENTAL | 14841 | 3.98% |
| Indenização por Dano Ambiental | 14777 | 3.97% |
| Pesca | 10245 | 2.75% |

As combinações de classes e assuntos mais comuns estão na [Tabela 3](#tbl-classe-assunto). Aparentemente, as classes ajudam mais a pegar informações das fases (que estão disponíveis nas movimentações) do que do tipo de procedimento.

Tabela 3: Dez combinações de classes e assuntos mais comuns. Um processo pode ter mais de um assunto.

| sgt\_nm\_classe | sgt\_nm\_assunto | n | prop |
| --- | --- | --- | --- |
| Inquérito Policial | Crimes contra a Flora | 28933 | 7.77% |
| Procedimento Comum Cível | Dano Ambiental | 27007 | 7.25% |
| Ação Civil Pública | Dano Ambiental | 17965 | 4.82% |
| Inquérito Policial | Crimes contra o Meio Ambiente e o Patrimônio Genético | 16173 | 4.34% |
| Inquérito Policial | Crimes contra a Fauna | 11477 | 3.08% |
| Ação Penal - Procedimento Ordinário | Crimes contra o Meio Ambiente e o Patrimônio Genético | 10183 | 2.73% |
| Inquérito Policial | Da Poluição | 9479 | 2.54% |
| Ação Penal - Procedimento Ordinário | Crimes contra a Flora | 9370 | 2.51% |
| Crimes Ambientais | Crimes contra a Flora | 7697 | 2.07% |
| Ação Civil Pública | Flora | 7552 | 2.03% |

## Localização

A [Figura 1](#fig-mapa) mostra os pontos em que os processos foram distribuídos. A maior parte dos casos foi distribuída em Brasília, o que a princípio não faz muito sentido.

|  |
| --- |
| Figura 1: Pontos de distribuição dos processos |

A [Tabela 4](#tbl-municipios) mostra os 10 municípios com mais processos. A maioria são capitais, o que é esperado, já que são municípios mais populosos.

Tabela 4: Municípios com mais processos.

| loc\_uf | loc\_muni | n | prop |
| --- | --- | --- | --- |
| DF | BRASÍLIA | 17990 | 5.09% |
| SP | SAO PAULO | 13643 | 3.86% |
| MT | CUIABA | 6865 | 1.94% |
| PR | PARANAGUA | 6750 | 1.91% |
| PR | CURITIBA | 5859 | 1.66% |
| MG | BELO HORIZONTE | 5619 | 1.59% |
| RS | PORTO ALEGRE | 5558 | 1.57% |
| AM | MANAUS | 5355 | 1.52% |
| RO | PORTO VELHO | 5097 | 1.44% |
| SC | FLORIANOPOLIS | 4576 | 1.29% |

A [Tabela 5](#tbl-ilg) mostra os municípios com maior litigiosidade (processos por 100 mil habitantes). É possível identificar que vários municípios pequenos (com menos de 10 mil habitantes) figuram entre os mais litigiosos.

Tabela 5: Municípios com mais litigiosidade. Considerada população do Censo IBGE de 2010.

| loc\_uf | loc\_muni | n | pop | ilg |
| --- | --- | --- | --- | --- |
| PR | BOCAIUVA DO SUL | 2389 | 10852 | 22014.375 |
| ES | ALFREDO CHAVES | 2570 | 13946 | 18428.223 |
| SE | SANTANA DO SAO FRANCISCO | 657 | 6946 | 9458.681 |
| MG | BONFIM | 347 | 6738 | 5149.896 |
| PR | PARANAGUA | 6750 | 139774 | 4829.224 |
| MG | AIURUOCA | 270 | 6047 | 4465.024 |
| MT | CLAUDIA | 400 | 10790 | 3707.136 |
| SP | IGUAPE | 1008 | 28565 | 3528.794 |
| SP | CANANEIA | 413 | 11900 | 3470.588 |
| SE | CEDRO DE SAO JOAO | 191 | 5598 | 3411.933 |

A [Tabela 6](#tbl-ilg-50k) mostra os municípios com maior litigiosidade, considerando somente municípios com mais de 50 mil habitantes. Novamente aparecem algumas capitais e figuram entre os mais litigiosos municípios do sul e sudeste.

Tabela 6: Municípios com mais litigiosidade.

| loc\_uf | loc\_muni | n | pop | ilg |
| --- | --- | --- | --- | --- |
| PR | PARANAGUA | 6750 | 139774 | 4829.224 |
| SP | UBATUBA | 2136 | 78243 | 2729.957 |
| PA | BARCARENA | 2128 | 97780 | 2176.314 |
| PA | ALTAMIRA | 2025 | 93243 | 2171.745 |
| SP | SAO SEBASTIAO | 1065 | 73523 | 1448.526 |
| SP | REGISTRO | 744 | 54058 | 1376.300 |
| MG | MARIANA | 680 | 53607 | 1268.491 |
| MT | CUIABA | 6865 | 546748 | 1255.606 |
| RO | PORTO VELHO | 5097 | 419354 | 1215.441 |
| PA | ITAITUBA | 1122 | 92834 | 1208.609 |

Considerando somente os estados que fazem parte da Amazônia Legal e populações acima de 50 mil pessoas, temos a [Tabela 7](#tbl-ilg-50k-amazon). Barcarena e Altamira são as regiões com maior litigiosidade, acompanhadas das capitais Cuiabá e Porto Velho.

Tabela 7: Municípios com mais litigiosidade.

| loc\_uf | loc\_muni | n | pop | ilg |
| --- | --- | --- | --- | --- |
| PA | BARCARENA | 2128 | 97780 | 2176.3142 |
| PA | ALTAMIRA | 2025 | 93243 | 2171.7448 |
| MT | CUIABA | 6865 | 546748 | 1255.6059 |
| RO | PORTO VELHO | 5097 | 419354 | 1215.4409 |
| PA | ITAITUBA | 1122 | 92834 | 1208.6089 |
| MT | SINOP | 1172 | 111929 | 1047.0924 |
| PA | SAO FELIX DO XINGU | 697 | 76357 | 912.8174 |
| PA | BREU BRANCO | 441 | 50769 | 868.6403 |
| RR | BOA VISTA | 1765 | 280694 | 628.7986 |
| PA | TUCURUI | 543 | 95670 | 567.5760 |

## Tempos

Como temos processos em andamento, a forma adequada de analisar os tempos é utilizando análise de sobrevivência (Colosimo e Giolo 2006). Faremos isso utilizando gráficos descritivos de Kaplan Meier. O resultado é a [Figura 2](#fig-km). Foram retirados tempos maiores de 10 mil dias e tempos menores que zero. A mediana é de aproximadamente 98 meses, ou seja, 8 anos.

|  |
| --- |
| Figura 2: Curva de sobrevivência dos tempos de tramitação dos processos |

A [Figura 3](#fig-km-classe) mostra a disribuição dos tempos por classe, considerando as cinco classes com maior volume processual. As curvas são parecidas em forma e apresentam medianas similares. Inquéritos policiais, ações civis públicas e crimes ambientais apresentam tempos medianos acima de 110 meses.

|  |
| --- |
| Figura 3: Curva de sobrevivência dos tempos de tramitação dos processos, por classe |

A [Figura 4](#fig-km-assunto) mostra a disribuição dos tempos por assunto, considerando os cinco assuntos com maior volume. As curvas são parecidas em forma e apresentam medianas similares.

|  |
| --- |
| Figura 4: Curva de sobrevivência dos tempos de tramitação dos processos, por assunto |

## Partes

Nessa seção, estudamos apenas o polo passivo das ações. A base de dados estava com o nome da coluna partes\_at\_desc com o valor “polo passivo”. No entanto, ao considerar que “at” significa “passivo”, os resultados não fizeram muito sentido (ministério público tinha a maioria dos casos). Portanto, consideramos os casos “pa” como “passivo”.

A [Tabela 8](#tbl-tipo-pessoa) mostra os tipos de pessoa. A maior parte é de pessoas jurídicas, sendo que quase 1% da base possui autoridades, orgão de representação ou valores vazios.

Tabela 8: Tipos de pessoa

| tipo\_pessoa | n | prop |
| --- | --- | --- |
| FISICA | 577286 | 60.63% |
| JURIDICA | 366865 | 38.53% |
| AUTORIDADE | 4130 | 0.43% |
| ORGAOREPRESENTACAO | 3680 | 0.39% |
|  | 234 | 0.02% |
| Total | 952195 | 100.00% |

Vamos analisar apenas as pessoas jurídicas. Dentre as partes, aproximadamente 10% tem CNPJ vazio e 90% preenchido. No total, são cerca de 48000 CNPJs distintos e com 14 dígitos, que foram pesquisados na base de dados abertos da Receita Federal do Brasil (RFB). Quase a totalidade (99%) foram encontrados. Alguns resultados disso foram colocados a seguir.

A [Tabela 9](#tbl-litigantes) mostra os quinze maiores litigantes. O nome que mais aparece é o vazio, que são casos em que o campo do CNPJ está vazio. Em seguida, vem a Samarco e a Vale.

Tabela 9: 10 Maiores litigantes nos processos

| nome\_fantasia | n\_processos | prop |
| --- | --- | --- |
| NA | 81842 | 13.91% |
| SAMARCO | 51502 | 8.76% |
| VALE - AREA OPERACIONAL GOVERNADOR VALADARES | 8035 | 1.37% |
| IBAMA SUPERINT REGIONAL NO GOIAS | 4575 | 0.78% |
| SANEPAR | 4122 | 0.70% |
| PLUMBUM | 3995 | 0.68% |
| IBAMA/CGFIN - COORDENACAO GERAL DE FINANCAS | 3702 | 0.63% |
| PETROBRAS | 3214 | 0.55% |
| COORDENACAO GERAL DE ORCAM FINAN E ANALISE CONTABIL | 2880 | 0.49% |
| SUPERINTENDENCIA DE ADMINISTRACAO DO MF/ RJ | 2791 | 0.47% |
| NORTE ENERGIA | 2523 | 0.43% |
| PETROBRAS - EDISE | 2389 | 0.41% |
| IBAMA - CPB | 2338 | 0.40% |
| MPF GABINETE DO MINISTRO | 1956 | 0.33% |
| PROCURADORIA GERAL DE JUSTICA | 1908 | 0.32% |

A [Tabela 10](#tbl-cnae) mostra as atividades principais das empresas no polo passivo, de acordo com a informação da “Classe” da tabela CNAE. Os casos de “administração pública em geral” provavelmente se referem a diferentes entidades públicas.

Tabela 10: Quantidade de partes-processos por tipo de atividade.

| nm\_classe | n\_processos | prop |
| --- | --- | --- |
| Administração pública em geral | 55703 | 9.47% |
| Extração de minério de ferro | 54556 | 9.28% |
| Outras sociedades de participação exceto holdings | 24850 | 4.22% |
| Transporte ferroviário de carga | 9200 | 1.56% |
| Holdings de instituições não financeiras | 8217 | 1.40% |
| Geração de energia elétrica | 8211 | 1.40% |
| Captação tratamento e distribuição de água | 7396 | 1.26% |
| Atividades imobiliárias de imóveis próprios | 5166 | 0.88% |
| Bancos múltiplos com carteira comercial | 4696 | 0.80% |
| Desdobramento de madeira | 4652 | 0.79% |

A tabela da RFB permite fazer muitos outros cruzamentos. Ela contém informações como localização da sede, situação da empresa, natureza da empresa, quantidade de sócios, entre outras.

Colosimo, Enrico Antonio, e Suely Ruiz Giolo. 2006. *Análise de Sobrevivência Aplicada*. Editora Blucher.