**Documentação**

1. ALEPE – Verbas Indenizatórias
2. ALEPE – Projetos de leis
3. Câmara dos Deputados – CEAP
4. TSE – Boletim de urna
5. TSE – Filiação partidária
6. TSE – Receitas de campanhas
7. TSE – Despesas de campanhas
8. Receita Federal – Quadro de Sócios e Administradores (QSA)
9. **ALEPE – Verbas indenizatórias**

**O que é isto?**

Cada deputado estadual da ALEPE possui uma verba de R$15,450/mês para reembolso de despesas de sua atividade como parlamentar. Isto inclui gastos com serviços de aluguel de veículos, assessoria jurídica, consultorias técnicas, material de escritório, impressões em gráficas etc.

**Os dados capturados são:**

* Nome do parlamentar
* Período do pedido
* Código do pedido
* Valor do pedido
* Tipo de despesa
* CNPJ empresa
* Nome empresa
* Data da nota fiscal
* Valor da nota fiscal

**Como serão utilizados estes dados?**

* Criar ranking de 1) maiores gastadores de verbas indenizatórias, 2) empresas que mais recebem e 3) visão geral sobre os principais tipos de despesa
* Verificar se os sócios das empresas que recebem esta verba de reembolsos são 1) filiados a partidos políticos, 2) já foram/são servidores públicos, 3) já se candidataram a algum cargo eletivo, 4) possuem histórico na justiça, 5) São familiares de políticos
* Para realizar estas correlações será necessário cruzar a tabela de verbas indenizatórias com as tabelas do TSE,Receita Federal, e de Familiares (ainda não disponível).

**Fonte Original dos dados:**

<http://www.alepe.pe.gov.br/transparencia-vi/>

**Github:**

<https://github.com/CidadaoFiscal/alepe>

**Como funciona o robô da ALEPE (Xico\_alepe\_verba\_ind.py):**

* Código escrito em Python usando libraries: BeautifulSoup, Requests, Pandas, Intertools, csv, datetime, time.
* Os dados de verba indenizatória são publicados em um link que é a combinação de “código de deputado” + “ano” + “mes”. – Exemplo neste link: <http://www.alepe.pe.gov.br/transparencia-vi/?dep=1&ano=2015&mes=2> ... onde “1”, “2015” e “2” são os respectivos códigos.
* **Os passos do robô:**
  1. Obtem a lista atualizada de todos os códigos de deputados + cria uma lista com todas as combinações de “dep” + “ano” + “mês”.
  2. Cria a lista de links a serem acessados através das combinações de códigos acima
  3. Utilizando de BeautifulSoup, captura todos os dados relevantes da página. – É feito um FOR LOOP para repetir este passo para todas as páginas da lista criada no passo #2.
  4. Através do Pandas, os dados são tabulados em um dataframe e em seguida exportados para 2 arquivos CSV. Estes são:
     1. Reportverbas.csv 🡪 tabela com todos os dados dos reembolos
     2. Reportpedidos.csv 🡪 tabela com os dados dos pedidos.
  5. Infelizmente ... este script não consegue linkar os dados do gasto com o # do pedido. Esta junção de dados ocorre através de um template de excel chamado “Consolidador.xlsx”.

**Como os dados são tratados?**

* Após a obtenção dos dados brutos do site da ALEPE + consolidação dos reports de gastos+pedidos (através do excel “Consolidador”), é necessário analisar os dados e fazer devidas normalizações. Utilizamos o excel “Analise.xlsx” para este propósito.
* Na tab “Index” ... ajustamos os dados de “CNPJ/CPF”, “Nome do fornecedor”, tipo de gasto, etc.
* Após a normalização, salvamos a tabela de dados ajustada em um formato .csv para ser então importada em um banco de dados para visualização e investigação.

1. **ALEPE – Projetos de Lei**

**O que é isto?**

Tabela com os dados de todas os projetos de leis apresentados pelos deputados estaduais da ALEPE:

**Os dados capturados são:**

* Parlamentar
* titulo do projeto de lei
* resumo
* textocompleto
* justificativa
* historico
* situacao
* Data de aprovações
* Anexos

**Como serão utilizados estes dados?**

* Criar rankings de 1) quais são os tipos de projetos de leis mais frequentes 2) quais são os deputados que mais apresentam projetos por categoria, 3) qual o % de projetos rejeitados por inconstitucionalidade, etc
* Permitir que o cidadão comum saiba que tipos de projetos de leis foram apresentados pelo seu representante eleito

**Fonte Original dos dados:**

<http://www.alepe.pe.gov.br/proposicoes/>

**Github:**

<https://github.com/CidadaoFiscal/alepe>

**Como funciona o robô da ALEPE para projetos de lei:**

* Código escrito em Python usando libraries: Selenium, BeautifulSoup, Requests, Pandas, csv, datetime, time.
* São duas tabelas importantes a serem capturadas:
  1. Lista de todos os projetos de leis apresentados: Incluindo Autor, Título do projeto, Data da publicação e Link do Projeto
  2. Dados sobre o projeto (ao clicar no link do projeto)
* **Os passos do robô:**
  1. ***Xico\_alepe\_propostas\_todas.py:***
     1. Este script de Python utiliza selenium para capturar a lista de todos os projetos de leis
     2. Ao terminar a captura, esta lista é exportada em um arquivo chamado “reportpropostas.csv”.
  2. ***Xico\_dados\_propostas.py:***
     1. Este script de Python utiliza BeautifulSoup para acessar todos os links salvos no arquivo “reportpropostas.csv” e capturar todos os dados relevantes sobre os projetos de lei.
     2. Ao terminar a caputra, os dados são salvos em um arquivo chamado “report\_dados\_propostas.csv”

**Como os dados são tratados?**

* Após o fim do processo do robô ... utilizamos um template de excel chamado de “Analise\_projetos\_de\_lei\_Alepe.xlsx” para categorizar todos os projetos de leis e realizar análises.
* Infelizmente ... o arquivo “report\_dados\_propostas.csv” não inclui o nome do Autor do projeto. Portanto ... precisamos fazer um VLOOKUP no excel para buscar o autor do projeto no arquivo “reportpropostas.csv”. – Este vlookup é feito utilizando o Link do projeto como chave.

1. **Câmara dos Deputados - CEAP**

**O que é isto?**

CEAP ou Cota Para o Exercício da Atividade Parlamentar (també conhecido como “Cotão”) é a verba para reembolso de despesas dos deputados federais no Congresso. Esta verba varia de 30 mil até 44 mil reais por mês, dependendo do estado. A verba para deputados de Pernambuco é de R$41.676 por mês.

Esta verba pode ser utilizada para pagamento de alimentação, passagens aéreas, matéria de divulgação da atividade parlamentar (Gráficas, Outdoors, etc).

**Os dados capturados são:**

* Parlamentar (Nome, partido)
* Data Nota Fiscal
* Cópia digitalizada da Nota Fiscal
* CNPJ do Fornecedor
* Nome do Fornecedor
* Valor da Nota Fiscal
* Valor restituído (as vezes a câmara aprova a restituição parcial, caso parte do gasto não possa ser reembolsado)
* Tipo de Serviço realizado
* Número do pedido/reembolso
* Entre outros

**Como serão utilizados estes dados?**

* Criar ranking de 1) maiores gastadores da verba CEAP, 2) empresas que mais recebem e 3) visão geral sobre os principais tipos de despesa
* Verificar se os sócios das empresas que recebem esta verba são 1) filiados a partidos políticos, 2) já foram/são servidores públicos, 3) já se candidataram a algum cargo eletivo, 4) possuem histórico na justiça, 5) São familiares de políticos
* Para realizar estas correlações será necessário cruzar a tabela de verbas indenizatórias com as tabelas do TSE,Receita Federal, e de Familiares (ainda não disponível).

**Fonte Original dos dados:**

<https://dadosabertos.camara.leg.br/swagger/api.html#staticfile>

**Github:**

<https://github.com/CidadaoFiscal/CEAP>

**Como os dados são tratados?**

* Os dados da CEAP são publicados já em formato .CSV no site da câmara. Não sendo necessário o uso de robôs para captura dos dados.
* Não temos nenhum template específico para o tratamento dos dados
* Apesar de haver alguns erros de classificação das despesas (Por exemplo: uma nota fiscal de gasolina foi erroneamente classificada como gasto com refeição), no geral, os dados da CEAP não requerem muito esforço de normalização. A qualidade dos dados é boa.

1. **TSE – Boletim de Urna**

**O que é isto?**

Boletim de urna é a tabela de dados mais completa com o resultado das eleições. Nesta tabela é possível saber quantos votos o candidato recebeu por seção eleitoral (esta é a informação mais granular registrada pela urna, visto que o CPF do eleitor não é registrado por questão de sigilo eleitoral).

**Os dados capturados são:**

* Candidato Nome
* Candidato CPF
* Candidato Partido
* Candidato Cargo
* Votos obtidos por seção eleitoral
* Resultado (eleito vs. Não eleito)
* Vários outros detalhes registrados pela Urna

**Como serão utilizados estes dados?**

* Servirá para mostrar se uma determinada pessoa já foi candidata a cargo eletivo e qual foi o seu resultado no pleito. Este dado será a fonte do “Histórico na política” de cada pessoa
* Será possível realizar análises para indicar em quais localidades o candidato obteve mais voto... o que ajuda a entender qual é a “base eleitoral” do candidato.

**Fonte Original dos dados:**

<http://www.tse.jus.br/eleicoes/estatisticas/repositorio-de-dados-eleitorais-1/repositorio-de-dados-eleitorais>

**Github:**

https://github.com/CidadaoFiscal/TSE

**Como os dados são tratados?**

* O Boletim de Urna já é publicado no formato .CSV no site do TSE, não sendo necessário o uso de robôs para captura dos dados.
* Não é necessário realizar qualquer tipo de normalização. Os dados são precisos e confiáveis.
* Esta tabela não inclui uma coluna para “Bairro” e “Local de votação”. – Entretanto... existe uma outra tabela do TSE que indica esta informação. Portanto é necessário fazer o cruzamento.

1. **TSE – Filiação partidária**

**O que é isto?**

Esta tabela contém a lista de todas as pessoas filiadas a partidos políticos no Brasil. É possível também checar o histórico de filiação.

**Os dados capturados são:**

o Nome do filiado;

o Sigla do partido político;  
o Nome do partido político;  
o Unidade da federação;  
o Código do município;  
o Nome do município;  
o Zona eleitoral;  
o Seção eleitoral;  
o Data da filiação;  
o Situação do registro de filiação;  
o Tipo do registro de filiação;  
o Data do processamento do registro de filiação; o Data da desfiliação;  
o Data do cancelamento do registro de filiação; o Data da regularização do registro de filiação;  
o Motivo do cancelamento do registro de filiação

**Como serão utilizados estes dados?**

* Servirá para mostrar se uma determinada pessoa já foi filiada a algum partido político.
* Servirá para saber se sócios de empresas contratadas ou comissionados são filiados ao mesmo partido do político contratante. Esta informação pode indicar suspeitas de “aparelhamento da máquina pública”. (Apenas como exemplo, é muito comum que dirigentes de estatais sejam filiados de partidos da base do governo).

**Fonte Original dos dados:**

<http://www.tse.jus.br/partidos/filiacao-partidaria/relacao-de-filiados>

**Github:**

<https://github.com/CidadaoFiscal/TSE>

**Como os dados são tratados?**

* Esta tabela é disponível em um arquivo .CSV no site do TSE.
* Infelizmente, só é possível baixar os arquivos por partido e por estado. Sendo assim, é necessário o download de diversos arquivos para obtenção da base completa de filiados no Brasil.

1. **TSE – Receitas de Campanha**

**O que é isto?**

Esta tabela contém a lista de todas as receitas obtidas para a campanha eleitoral. Exemplos de receitas são: Fundo Partidário, Fundo Eleitoral e doação de empresas e pessoas físicas.

**Os dados capturados são:**

* Candidato Nome
* Candidato Partido
* Origem Receita
* Nome doador
* Valor
* Entre outros dados

**Como serão utilizados estes dados?**

* Servirá para mostrar quanto cada político recebeu de dinheiro público ou de doadores primados para a sua campanha eleitoral
* Servirá para identificar se algum doador é servidor comissionado ou sócio de empresa que presta serviço via licitação ou através de outras verbas

**Fonte Original dos dados:**

<http://www.tse.jus.br/eleicoes/estatisticas/repositorio-de-dados-eleitorais-1/repositorio-de-dados-eleitorais>

**Github:**

<https://github.com/CidadaoFiscal/TSE>

**Como os dados são tratados?**

* Esta tabela já é publicada no formato .CSV no site do TSE, não sendo necessário o uso de robôs para captura dos dados.
* Não é necessário realizar qualquer tipo de normalização. Os dados são precisos e confiáveis.

1. **TSE – Gastos de Campanha**

**O que é isto?**

Esta tabela contém a lista de todas as despesas de campanha dos candidatos a cargos eletivos. Como muitas campanhas são financiadas quase que exclusivamente com dinheiro público... esta tabela pode indicar o destino final do dinheiro do contribuinte.

**Os dados capturados são:**

* Candidato Nome
* Candidato Partido
* Origem do recurso
* Nome do fornecedor
* Tipo de despesa
* Valor
* Entre outros dados

**Como serão utilizados estes dados?**

* Servirá para mostrar quanto cada político gastou em suas campanhas eleitorais.
* Servirá para identificar se algum fornecedor é servidor comissionado ou sócio de empresa que presta serviço via licitação ou através de outras verbas.

**Fonte Original dos dados:**

<http://www.tse.jus.br/eleicoes/estatisticas/repositorio-de-dados-eleitorais-1/repositorio-de-dados-eleitorais>

**Github:**

<https://github.com/CidadaoFiscal/TSE>

**Como os dados são tratados?**

* Esta tabela já é publicada no formato .CSV no site do TSE, não sendo necessário o uso de robôs para captura dos dados.
* Não é necessário realizar qualquer tipo de normalização. Os dados são precisos e confiáveis.

1. **Receita Federal – Quadro de Sócios e Administradores**

**O que é isto?**

Esta tabela contém a lista de todas as empresas registradas no Brasil e seus respectivos sócios

**Os dados capturados são:**

* Tipo da Empresa
* CNPJ
* Nome da Empresa
* Nome do sócio
* Qualificação do sócio
* outros

**Como serão utilizados estes dados?**

* Será uma das principais tabelas utilizadas pela Cidadão Fiscal. Iremos verificar quem são os sócios de todas as empresas que forem contratadas pelos parlamentares e checar se os mesmos possuem alguma relação direta. Checaremos se os sócios filiados ao mesmo partido do contratante, se fizeram doação para campanhas ou se já foram candidatos em outras eleições.

**Fonte Original dos dados:**

<http://idg.receita.fazenda.gov.br/orientacao/tributaria/cadastros/cadastro-nacional-de-pessoas-juridicas-cnpj/dados-abertos-do-cnpj>

**Github:**

https://github.com/CidadaoFiscal/ReceitaFederal

**Como os dados são tratados?**

* Esta tabela já é publicada no formato .CSV no site do TSE, não sendo necessário o uso de robôs para captura dos dados.
* Não é necessário realizar qualquer tipo de normalização. Os dados são precisos e confiáveis.