Hi buddies, this post is the first part of a project that I am doing to increase my knowledge about getting and manipulating data. The idea came from a job interview that a did on my path to becoming a data scientist

No processo seletivo eu tinha um case de um problema e tinha que resolve-lo e apresentar os resultados. Passei o fim de semana resolvendo esse case (o processo começou na sexta-feira e devia ser entregue na segunda), talvez devido minha pouca experiencia foi bem difícil fazer tudo nesse tempo (foi meu primeiro processo seletivo para essa area, além disso, a empresa era de grande porte, uma emissora de televisão conhecida internacionalmente então me empenhei o máximo que pude pois queria muito passer. Spoiler: não passei hahah), pois bem, a parte em que fui mal no case foi na apresentação da resolução do problema, o “deploy”, onde o recrutador estava esperando algo mais “profissional” e com o pouco tempo que tive, apresentei um power point que até foi elogiado, porém, não era o que estava esperando. Aprendi muito fazendo esse case, e a partir dele resolve que um dia faria algo parecido. Lembro de ouvir muito dele que utilizavam a linguagem R e faziam o deploy com shiny com gráficos interativos e etc (era isso o que ele esperava do deploy). Sendo assim resolve criar um projeto como uma meta pessoal em que ia fazer todo o processo desde a extração dos dados até a apresentação, mas agora não mais em power point, mas em Power BI, Google data Studio além de um dashboard publicado na web desenvolvido com ajuda de bibliotecas do python (não é shiny ainda mas acredito que sejam semelhantes hehe).

As etapas desse projeto serão:

1º WebScrape - Extração dos dados a partir de uma API publica.

Aproveitando que estamos em ano eleitoral, resolvi pegar os dados publicos dos atuais deputados e os gastos realizados com dinheiro público desde o início da pandemia até atualmente (2020-2022).

Palavras-chave: Python, webscrape, API

2º Cloud and database integration with python.

Nesta etapa iremos criar uma conta na OracleCloudInfrastructure (OCI), onde poderemos criar um banco de dados e armazenar os dados extraídos.

Palavras-chave: Python, Oracle Cloud, Database, MySQL, Alwaysfree, VPN

3º Power BI

4º Google Data Studio

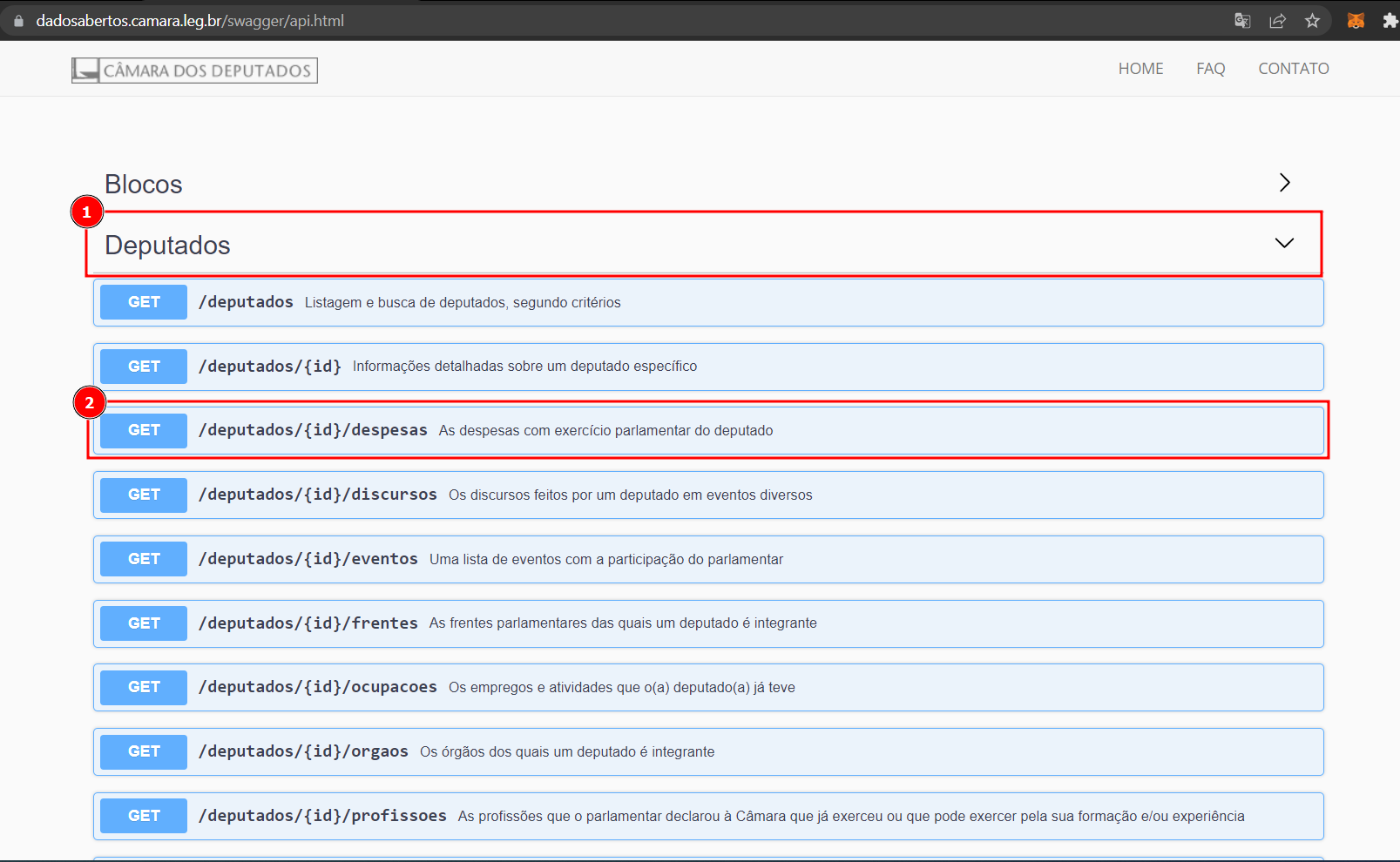
5º Python dashboard

In this section I want to show step by step how to get some data from a public API using python.

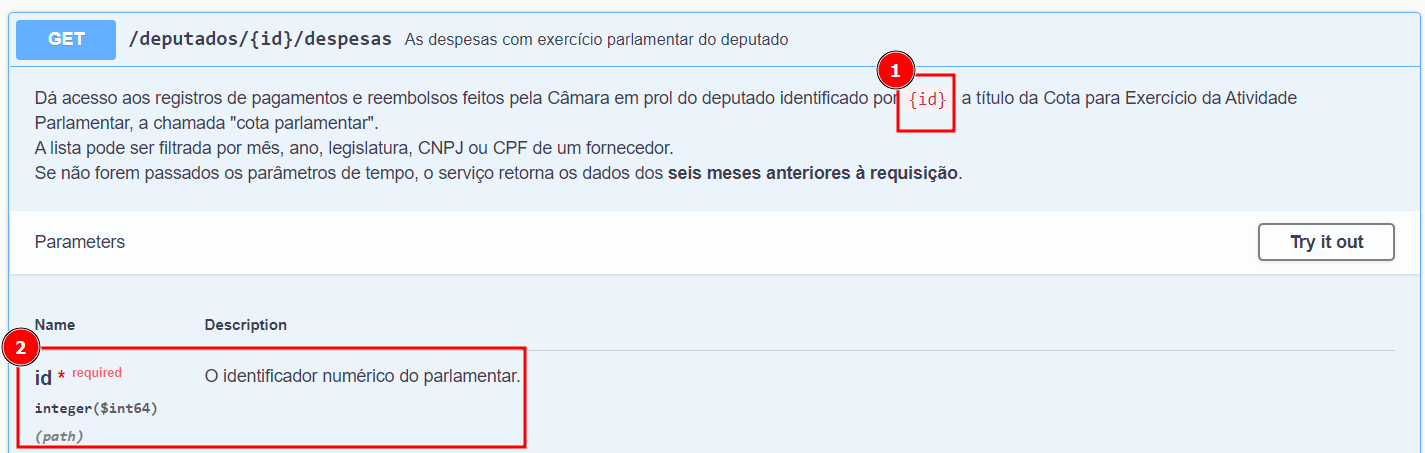
Alright, let’s get started.

Acessando a URL: <https://dadosabertos.camara.leg.br/swagger/api.html>

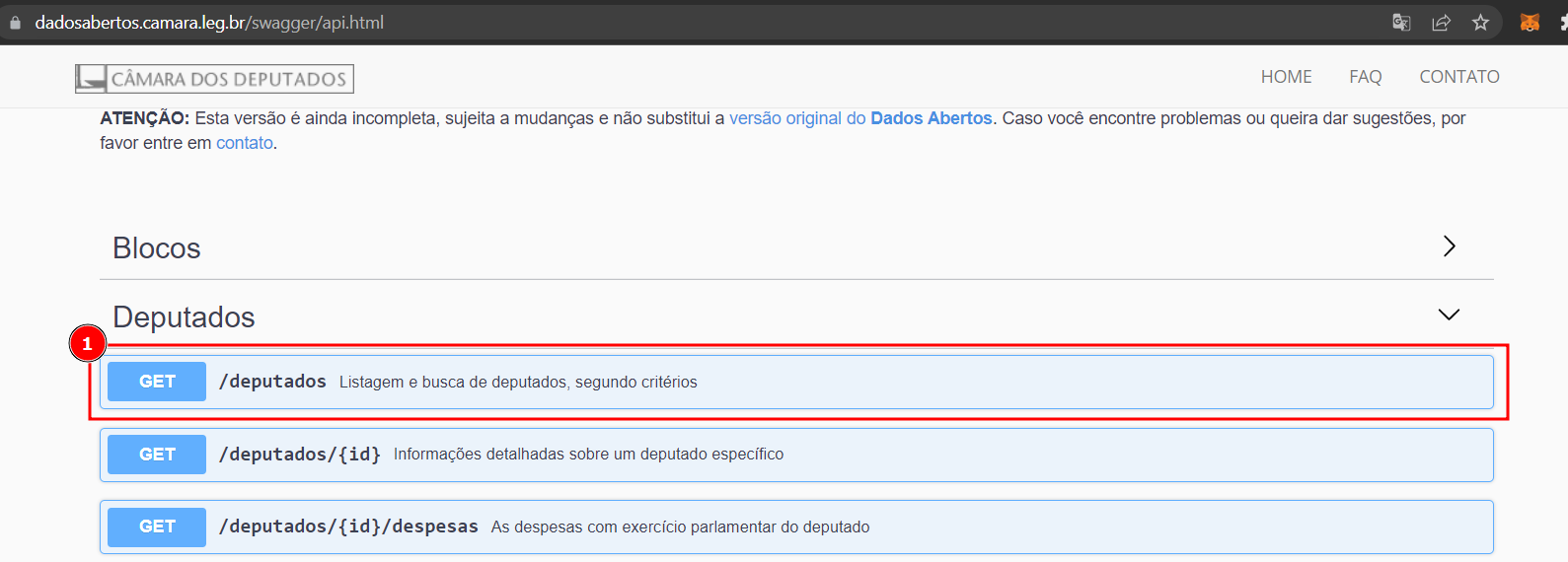
Temos a toda a documentação da API dos dados abertos. Nessa página buscamos pelas despezas do deputados:



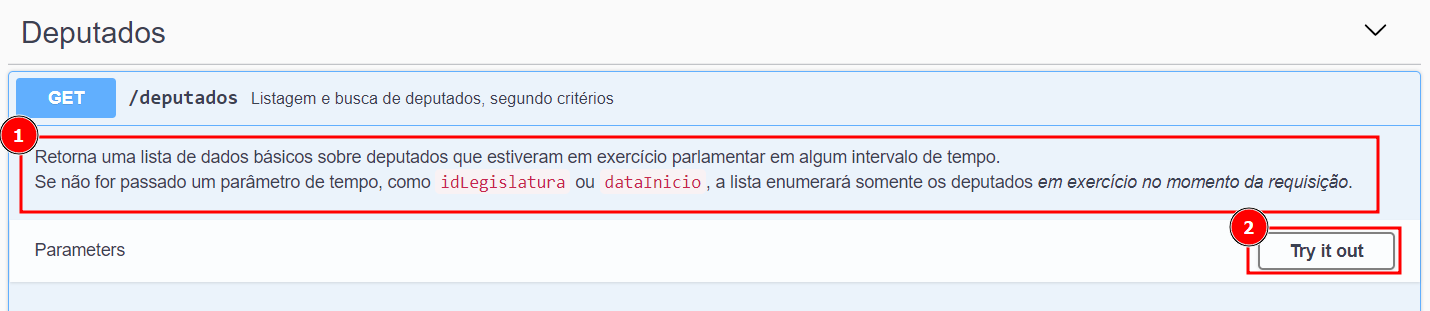
Clicando na aba de despezas, Podemos observar que para conseguirmos os dados, temos que obrigatóriamente ter o “ID” do deputado como é demonstrado na documentação.



No caso, queremos os dados das despezas de todos os deputados, então clicamos na aba de deputados pois lá Podemos encontrar a URL para assim, fazer requisições que trarão informações dos deputados e entre essas informações, os “ID’s” de cada um deles.



Clicando na aba deputados temos o seguinte:



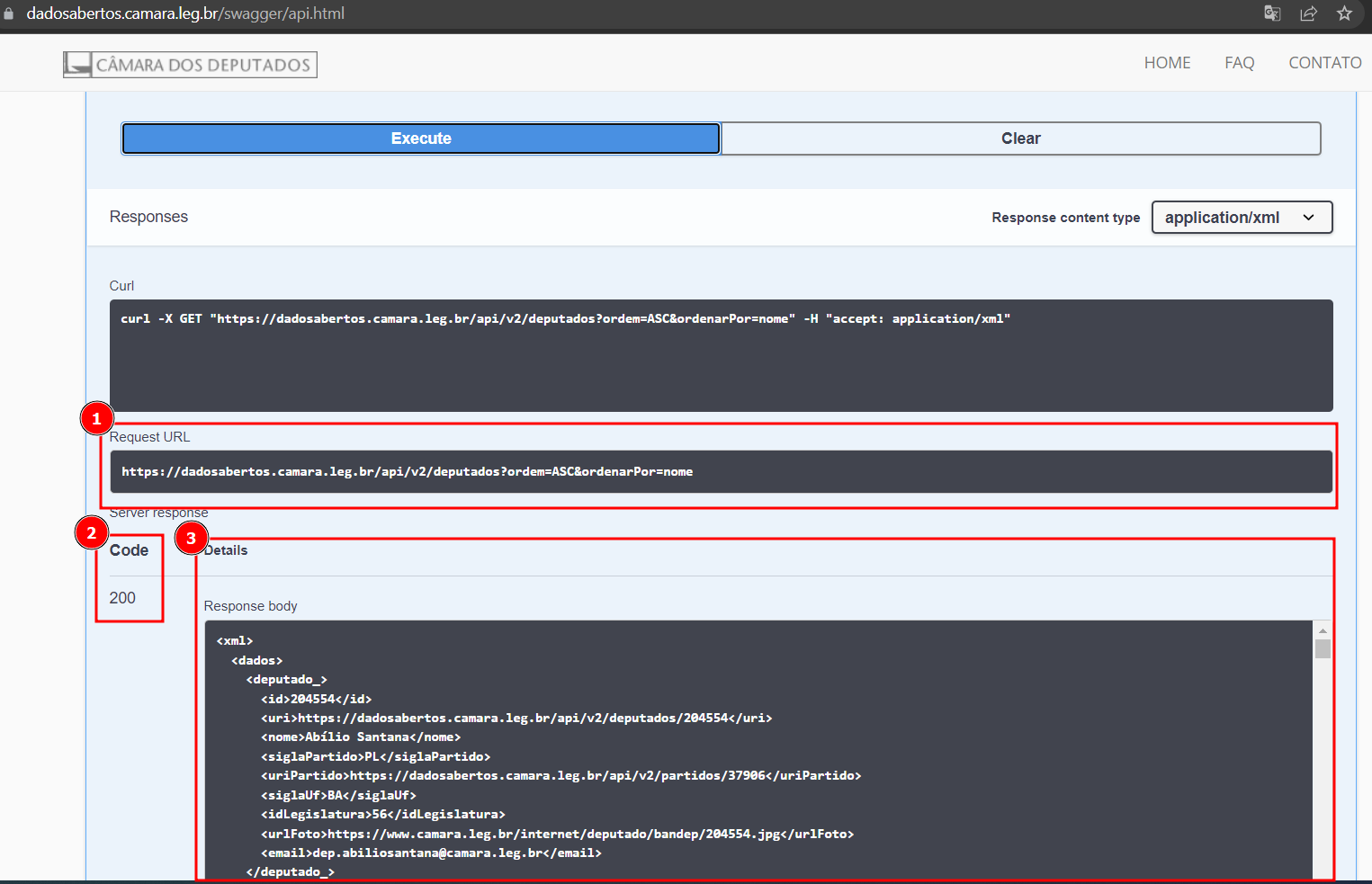
1- Uma breve descrição do que a requisição irá nos retornar

2- Botão para ‘testarmos’ a API e ver o que ela vai nos retornar de informações.

Clicando em Try it out, Rolando a página é possível notar um botão para a execução:



Clicando em execute, o site nos retorna o seguinte:



1 – URL utilizada para fazermos a requisição dos dados

2 – Código que a requisição nos retorna, 200 no caso, significa que a requisição foi bem sucedida.

3 – Os dados que precisamos.

Ótimo, agora que temos a URL para a requisição podemos iniciar nosso código em Python. Nesta primeira etapa do projeto, iremos utilizer apenas duas bibliotecas: Pandas e Requests. Caso não possua elas instaladas, digite em seu terminal:

Pip install pandas

Pip install requests

Caso tenha alguma dificuldade, aqui estão as páginas oficiais

https://pandas.pydata.org/

<https://docs.python-requests.org/en/latest/>

Com as bibliotecas instaladas, Podemos começar: