

Pegar a base de dados da Alesp(assembleia legislativa de são paulo) e realizar os tratamentos necessários com as seguintes diretrizes:

- Utilizar obrigatoriamente pandas
- Extrair diretamente do site o arquivo XML dos gastos e cadastros de cada deputado
- Verificar e corrigir possíveis inconsistências nos dados que podem ser (Nomes diferentes para o mesmo item ex: TÁXI e taxi) ou valores ausentes ou duplicidade de dados
- verificar e contar os valores ausentes (NaN , NA , etc)
- realizar insights onde mostrem os valores totais dos ultimos 4 anos dos valores gastos por cada deputado
- mostrar os itens dos ultimos quatro anos onde se mais teve despesa
- mostrar os 3 deputados que mais e menos gastam
- realizar a conversão do dataframe para CSV e disponibiliza-lo em um bucket com acesso público
- Os insights não se limitam aos pedidos acima , onde gostaríamos de pelo menos mais 2
- Os resultados podem ou não trazer plotagem(trate como não obrigatório) mas é recomendado que traga
- Montar uma pequena apresentação no próprio notebook trazendo uma organização e fluxo de ETL

Enjoy

Ps :

Conectores

lib para conexão

```
import requests
```

Modelagem de dados

```
import pandas as pd
```

Ignorando avisos

```
import warnings
```

```
warnings.filterwarnings('ignore')
```

URL da API da Alesp

URL_Despesas =

'http://www.al.sp.gov.br/repositorioDados/deputados/despesas_gabinetes.xml'

URL_Cadastro =

'<https://www.al.sp.gov.br/repositorioDados/deputados/deputados.xml>'

Pegando os dados

```
Conexao_01 = requests.get( URL_Despesas )
```

```
Conexao_02 = requests.get( URL_Cadastro )
```

```
# Verificando se a API funcionou  
print( Conexao_01, Conexao_02 )
```