

# Análise de Dados e Visualização: Desenvolvimento de um Dashboard Interativo para Análise de Gastos de Deputados Federais

Matheus Galdino dos Santos Lira

28 de abril de 2025

## Resumo

Este trabalho apresenta o desenvolvimento de um sistema de análise e visualização de dados dos gastos dos deputados federais brasileiros no ano de 2022. O projeto implementa um pipeline completo de ETL (Extract, Transform, Load) para coleta e processamento de dados da API da Câmara dos Deputados, seguido de análises estatísticas e visualizações interativas através de um dashboard desenvolvido com Streamlit. O sistema permite a exploração detalhada dos gastos parlamentares, oferecendo insights valiosos sobre o uso dos recursos públicos.

## Conteúdo

<b>1</b>	<b>Introdução</b>	<b>3</b>
1.1	Contextualização . . . . .	3
1.2	Objetivos . . . . .	3
<b>2</b>	<b>Estrutura do Projeto</b>	<b>3</b>
2.1	Organização dos Arquivos . . . . .	3
2.2	Exemplos de Código . . . . .	4
2.2.1	Configuração do Ambiente . . . . .	4
2.2.2	Extração de Dados . . . . .	4
2.2.3	Transformação de Dados . . . . .	4
2.2.4	Visualização no Dashboard . . . . .	5
<b>3</b>	<b>Revisão Bibliográfica</b>	<b>5</b>
3.1	ETL (Extract, Transform, Load) . . . . .	5
3.2	Visualização de Dados . . . . .	6
<b>4</b>	<b>Metodologia</b>	<b>6</b>
4.1	Tecnologias Utilizadas . . . . .	6
4.2	Arquitetura do Sistema . . . . .	6

<b>5</b>	<b>Desenvolvimento</b>	<b>6</b>
5.1	Processo ETL . . . . .	6
5.1.1	Extração . . . . .	6
5.1.2	Transformação . . . . .	7
5.1.3	Carga . . . . .	7
5.2	Análise de Dados . . . . .	7
5.3	Desenvolvimento do Dashboard . . . . .	7
<b>6</b>	<b>Análises e Visualizações</b>	<b>7</b>
6.1	Análises Implementadas . . . . .	7
6.1.1	Análise de Gastos por Categoria . . . . .	8
6.1.2	Análise Temporal . . . . .	8
6.1.3	Análise por Região . . . . .	8
6.2	Visualizações Implementadas . . . . .	8
6.2.1	Gráficos de Barras . . . . .	8
6.2.2	Gráficos de Pizza . . . . .	8
6.2.3	Tabelas Interativas . . . . .	8
6.3	Exemplos de Análises . . . . .	8
6.3.1	Distribuição de Gastos . . . . .	9
6.3.2	Evolução Temporal . . . . .	9
6.3.3	Ranking de gastos dos Deputados . . . . .	9
6.3.4	Gasto médio por Deputador por Partido . . . . .	10
<b>7</b>	<b>Desafios e Soluções</b>	<b>10</b>
7.1	Desafios Encontrados . . . . .	10
7.1.1	Desafios Técnicos . . . . .	10
7.1.2	Desafios de Negócio . . . . .	10
7.2	Soluções Implementadas . . . . .	11
7.2.1	Otimização de Performance . . . . .	11
7.2.2	Tratamento de Dados . . . . .	11
7.2.3	Melhorias na Interface . . . . .	11
<b>8</b>	<b>Resultados</b>	<b>11</b>
8.1	Visualizações Implementadas . . . . .	11
8.2	Insights Obtidos . . . . .	11
<b>9</b>	<b>Conclusão</b>	<b>12</b>
<b>10</b>	<b>Referências Bibliográficas</b>	<b>12</b>

## Lista de Figuras

1	Distribuição dos gastos por categoria . . . . .	9
2	Evolução dos gastos ao longo do tempo . . . . .	9
3	Top 20 deputados por valor total gasto . . . . .	10
4	Gasto médio por deputado por partido . . . . .	10

## Lista de Tabelas

# Agradecimentos

Agradeço a todos que contribuíram direta ou indiretamente para a realização deste trabalho, especialmente aos professores e colegas que ofereceram suporte e orientação durante todo o processo de desenvolvimento.

## 1 Introdução

### 1.1 Contextualização

A transparência no uso dos recursos públicos é um pilar fundamental da democracia. Neste contexto, a análise dos gastos dos deputados federais se torna uma ferramenta essencial para o controle social e a fiscalização do uso do dinheiro público. Este projeto visa desenvolver um sistema completo de análise e visualização desses dados, tornando-os acessíveis e compreensíveis para a sociedade.

### 1.2 Objetivos

- Desenvolver um pipeline ETL robusto para extração e processamento de dados da API da Câmara dos Deputados
- Implementar análises estatísticas dos gastos parlamentares
- Criar um dashboard interativo para visualização e exploração dos dados
- Documentar todo o processo de desenvolvimento e as decisões técnicas tomadas

## 2 Estrutura do Projeto

### 2.1 Organização dos Arquivos

O projeto está organizado da seguinte forma:

- `etl.py`: Script principal responsável pela extração, transformação e carga dos dados
- `dashboard/`: Diretório contendo os arquivos do dashboard interativo
  - `1__Homepage.py`: Página inicial do dashboard
  - `get_deputados.py`: Funções para obtenção de dados dos deputados
  - `get_despesas.py`: Funções para obtenção de dados de despesas
- `requirements.txt`: Lista de dependências do projeto
- `.env`: Arquivo de configuração com variáveis de ambiente

## 2.2 Exemplos de Código

### 2.2.1 Configuração do Ambiente

O projeto utiliza variáveis de ambiente para configuração. Exemplo de código para carregar configurações:

```
1 import os
2 from dotenv import load_dotenv
3
4 # Carregar variáveis de ambiente
5 load_dotenv()
6
7 # Configurar o do logging
8 path_logs = os.getenv("PATH_LOGS", "./logs")
9 os.makedirs(path_logs, exist_ok=True)
10
11 logging.basicConfig(
12     level=logging.INFO,
13     format="%(asctime)s - %(levelname)s - %(message)s",
14     handlers=[
15         logging.FileHandler(f"{path_logs}/logs.log"),
16         logging.StreamHandler()
17     ]
18 )
```

### 2.2.2 Extração de Dados

Exemplo de código para extração de dados da API:

```
1 def verificar_proxima_pagina(data):
2     for link in data['links']:
3         if link['rel'] == 'next' and link['href']:
4             return link['href']
5     return False
6
7 # Buscar lista de deputados
8 url_base = "https://dadosabertos.camara.leg.br/api/v2"
9 url_deputados = f"{url_base}/deputados"
10
11 response = requests.get(url_deputados, params={"idLegislatura":
12     56})
13 dados_deputados = response.json()
14 df_deputados = pd.DataFrame(dados_deputados["dados"])
```

### 2.2.3 Transformação de Dados

Exemplo de código para transformação dos dados:

```
1 # Normalizar o dos dados
2 df_deputados = df_deputados.rename(columns={
```

```

3     'id': 'id_deputado',
4     'nome': 'nome_deputado',
5     'siglaPartido': 'partido'
6 })
7
8 # Tratamento de valores ausentes
9 df_deputados = df_deputados.fillna({
10     'partido': 'SEM PARTIDO',
11     'uf': 'ND'
12 })

```

## 2.2.4 Visualização no Dashboard

Exemplo de código para criação de visualizações no Streamlit:

```

1 import streamlit as st
2 import plotly.express as px
3
4 # Configuração da página
5 st.set_page_config(
6     page_title="Análise de Gastos",
7     page_icon="📊",
8     layout="wide"
9 )
10
11 # Título
12 st.title("Análise de Gastos dos Deputados Federais")
13
14 # Gráfico de barras
15 fig = px.bar(
16     df_gastos,
17     x='deputado',
18     y='valor',
19     title='Gastos por Deputado'
20 )
21 st.plotly_chart(fig)

```

# 3 Revisão Bibliográfica

## 3.1 ETL (Extract, Transform, Load)

O processo ETL é fundamental para a preparação de dados para análise. No contexto deste projeto, o ETL foi implementado para:

- Extração: Obtenção dos dados da API da Câmara dos Deputados
- Transformação: Limpeza, normalização e preparação dos dados
- Carga: Armazenamento em banco de dados relacional (MySQL/SQlite)

## 3.2 Visualização de Dados

A visualização de dados é uma ferramenta poderosa para compreensão e comunicação de informações. Neste projeto, utilizamos o Streamlit para criar um dashboard interativo que permite:

- Visualização geral dos gastos dos deputados
- Análise detalhada por deputado
- Comparativos entre diferentes períodos
- Filtros por tipo de despesa

## 4 Metodologia

### 4.1 Tecnologias Utilizadas

- Python: Linguagem principal de programação
- Pandas: Biblioteca para manipulação e análise de dados
- Streamlit: Framework para desenvolvimento de dashboards
- SQLAlchemy: ORM para interação com banco de dados
- MySQL/SQLite: Sistemas de gerenciamento de banco de dados

### 4.2 Arquitetura do Sistema

O sistema foi desenvolvido seguindo uma arquitetura modular, com componentes separados para:

- Processamento de dados (ETL)
- Análise estatística
- Visualização e interface do usuário

## 5 Desenvolvimento

### 5.1 Processo ETL

O processo de ETL foi implementado utilizando Python e bibliotecas especializadas. A seguir, detalhamos cada etapa:

#### 5.1.1 Extração

A extração dos dados foi realizada através da API da Câmara dos Deputados ([dadosabertos.camara.leg.br](https://dadosabertos.camara.leg.br)), utilizando a biblioteca requests. Foram coletados:

- Dados básicos dos deputados
- Informações detalhadas de cada parlamentar
- Despesas realizadas no ano de 2022

### **5.1.2 Transformação**

O processo de transformação incluiu:

- Normalização dos dados
- Tratamento de valores ausentes
- Criação de chaves de relacionamento
- Estruturação em tabelas relacionais

### **5.1.3 Carga**

Os dados foram armazenados em um banco de dados relacional, com suporte a:

- MySQL: Para ambientes de produção
- SQLite: Para desenvolvimento e testes

## **5.2 Análise de Dados**

As análises implementadas incluem:

- Distribuição dos gastos por categoria
- Comparativo entre deputados
- Análise temporal dos gastos
- Identificação de padrões e outliers

## **5.3 Desenvolvimento do Dashboard**

O dashboard foi desenvolvido utilizando Streamlit, oferecendo:

- Interface intuitiva e responsiva
- Visualizações interativas
- Filtros dinâmicos
- Exportação de dados

# **6 Análises e Visualizações**

## **6.1 Análises Implementadas**

O sistema implementa diversas análises sobre os dados dos deputados e seus gastos:

### **6.1.1 Análise de Gastos por Categoria**

- Distribuição dos gastos por tipo de despesa
- Comparativo entre diferentes categorias
- Identificação de padrões de gastos

### **6.1.2 Análise Temporal**

- Evolução dos gastos ao longo do tempo
- Sazonalidade nos gastos
- Identificação de períodos de maior gasto

### **6.1.3 Análise por Região**

- Comparativo de gastos entre estados
- Média de gastos por região
- Identificação de padrões regionais

## **6.2 Visualizações Implementadas**

O dashboard oferece diversas visualizações interativas:

### **6.2.1 Gráficos de Barras**

- Comparação de gastos entre deputados
- Distribuição de gastos por categoria
- Evolução temporal dos gastos

### **6.2.2 Gráficos de Pizza**

- Proporção de gastos por categoria
- Distribuição de deputados por partido
- Distribuição por região

### **6.2.3 Tabelas Interativas**

- Lista detalhada de despesas
- Ranking de deputados por gastos
- Detalhamento por categoria

## **6.3 Exemplos de Análises**

A seguir, apresentamos alguns exemplos de análises realizadas:



6.3.1 Distribuição de Gastos

A Figura 1 mostra a distribuição dos gastos por categoria:

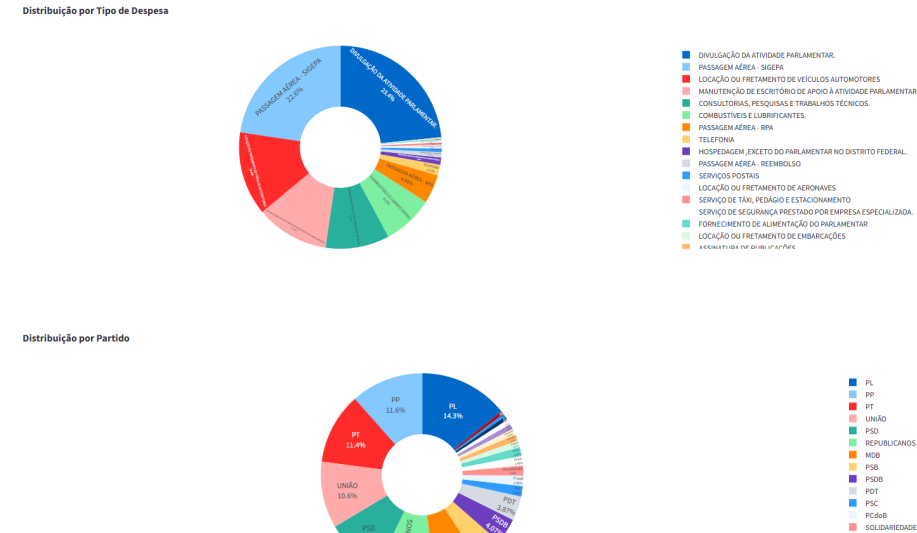


Figura 1: Distribuição dos gastos por categoria

6.3.2 Evolução Temporal

A Figura 2 apresenta a evolução dos gastos ao longo do tempo:

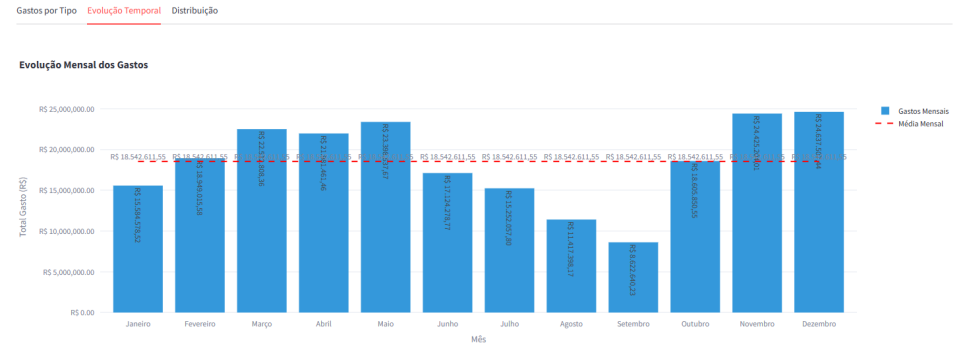


Figura 2: Evolução dos gastos ao longo do tempo

6.3.3 Ranking de gastos dos Deputados

A Figura 3 apresenta o ranking dos 20 deputados que mais gastaram:

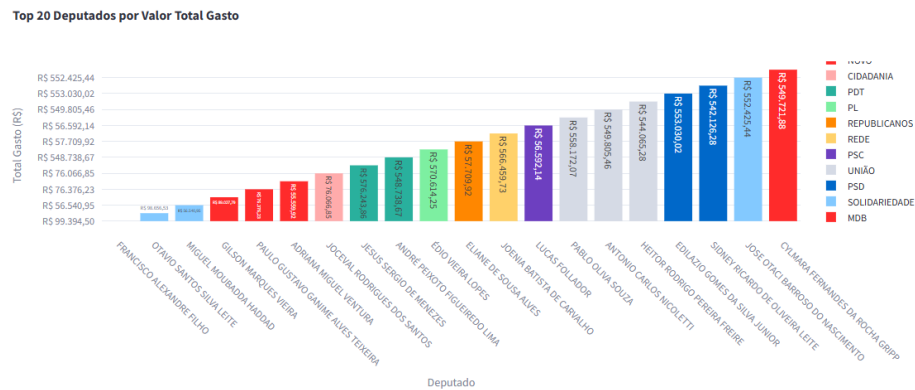


Figura 3: Top 20 deputados por valor total gasto

### 6.3.4 Gasto médio por Deputador por Partido

A Figura 4 apresenta o gasto médio por deputado por partido:

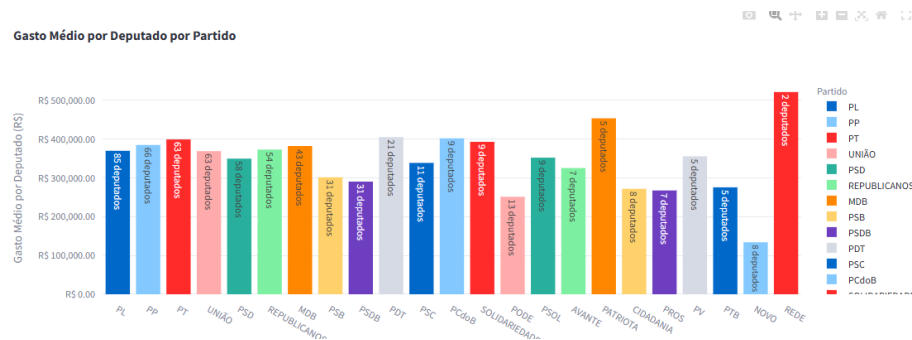


Figura 4: Gasto médio por deputado por partido

## 7 Desafios e Soluções

### 7.1 Desafios Encontrados

Durante o desenvolvimento do projeto, foram enfrentados diversos desafios:

#### 7.1.1 Desafios Técnicos

- Volume de dados: O grande volume de dados exigiu otimização no processamento
- Limitações da API: Necessidade de implementar paginação e tratamento de erros
- Performance do dashboard: Otimização das visualizações para melhor performance

#### 7.1.2 Desafios de Negócio

- Qualidade dos dados: Necessidade de tratamento e validação
- Interpretação dos dados: Desenvolvimento de visualizações claras e intuitivas
- Atualização dos dados: Implementação de processo de atualização periódica

## **7.2 Soluções Implementadas**

Para superar os desafios, foram implementadas as seguintes soluções:

### **7.2.1 Otimização de Performance**

- Implementação de cache para dados frequentemente acessados
- Uso de índices no banco de dados
- Otimização das consultas SQL

### **7.2.2 Tratamento de Dados**

- Implementação de validação de dados
- Tratamento de valores ausentes
- Normalização de dados

### **7.2.3 Melhorias na Interface**

- Design responsivo
- Feedback visual para o usuário
- Documentação clara das funcionalidades

## **8 Resultados**

### **8.1 Visualizações Implementadas**

O dashboard oferece diversas visualizações, incluindo:

- Gráficos de barras para comparação de gastos
- Gráficos de pizza para distribuição de despesas
- Tabelas interativas com dados detalhados
- Mapas de calor para análise temporal

### **8.2 Insights Obtidos**

Através da análise dos dados, foi possível identificar:

- Padrões de gastos por região
- Diferenças significativas entre deputados
- Categorias de despesas mais comuns
- Comportamentos atípicos que merecem atenção

## 9 Conclusão

Este projeto demonstrou a viabilidade e importância da análise de dados no contexto da transparência pública. Através do desenvolvimento de um sistema completo de ETL e visualização, foi possível criar uma ferramenta valiosa para o acompanhamento dos gastos parlamentares. O dashboard desenvolvido facilita o acesso e compreensão dos dados, contribuindo para o fortalecimento do controle social.

## 10 Referências Bibliográficas

### Referências

- [1] Streamlit Documentation. <https://docs.streamlit.io/>
- [2] Pandas Documentation. <https://pandas.pydata.org/docs/>
- [3] Câmara dos Deputados. Dados Abertos. <https://dadosabertos.camara.leg.br/>
- [4] SQLAlchemy Documentation. <https://www.sqlalchemy.org/>