

# insights\_com\_chatgpt

March 30, 2023

## 1 Custeio Administrativo

Insights sobre dados de custeio administrativo da administração pública federal usando ChatGPT

```
[1]: import os
import locale
from urllib.request import urlopen, quote
from zipfile import ZipFile
from io import BytesIO
import pandas as pd
import plotly.express as px
import openai
```

```
[2]: from itables import init_notebook_mode
init_notebook_mode(all_interactive=True)
```

<IPython.core.display.Javascript object>

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

```
[3]: locale.setlocale(locale.LC_ALL, 'pt_BR')
```

```
[3]: 'pt_BR'
```

### 1.1 Leitura dos dados

```
[4]: def load_data(year: int) -> pd.DataFrame:
    """Lê os arquivos de dados abertos e carrega em um dataframe"""

    REPO_URL = "https://repositorio.dados.gov.br/seges/raio-x/"
    FILE_PREFIX = "raiox"
    CSV_FILE = "custeio-administrativo.csv"

    df_list = []
    for ref in range(12):
        file_name = f"{FILE_PREFIX}-{year}-{(ref + 1):02d}.zip"
        file_url = REPO_URL + quote(file_name)
        url = urlopen(file_url)
```

```

file = ZipFile(BytesIO(url.read()))
df_month = pd.read_csv(file.open(CSV_FILE))
df_list.append(df_month)

df = pd.concat(df_list, axis=0, ignore_index=True)

df["ano_mes_referencia"] = df["ano_mes_referencia"].apply(str)
df = df[["ano_mes_referencia", "orgao_superior_nome", "
↪"orgao_superior_sigla",
        "orgao_nome", "orgao_sigla", "nome_item",
        "nome_natureza_despesa_detalhada", "valor"]]
df.rename({"nome_item": "item_despesa",
          "nome_natureza_despesa_detalhada": "natureza_despesa"},
          axis=1, inplace=True)

return df

```

```
[5]: ano_analise = input("Ano de análise:")
```

Ano de análise: 2022

```
[6]: df_full = load_data(ano_analise)
```

## 1.2 Visão global

```
[7]: # df = df_full[df_full.valor != 0].sample(n=10000)
df = df_full
```

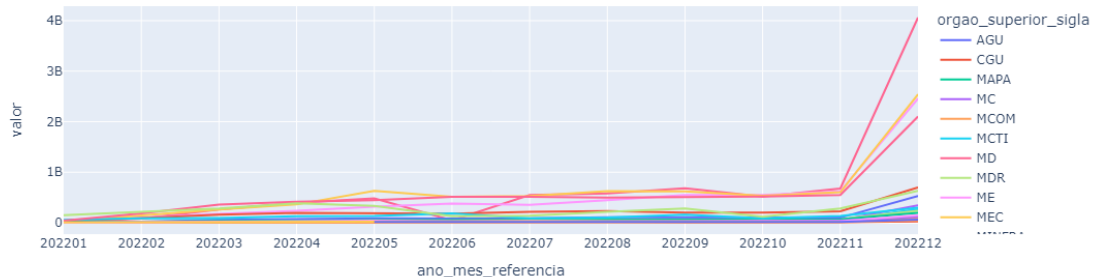
```
[8]: px.defaults.template = "plotly"

fig = px.histogram(data_frame=df, y="valor", x="ano_mes_referencia")
fig.update_layout(xaxis_type='category')
fig.show()
```



```
[9]: df_group = df.groupby(["ano_mes_referencia", "orgao_superior_sigla"])["valor"].
      ↪sum().reset_index()

fig = px.line(data_frame=df_group,
              x="ano_mes_referencia", y="valor", color="orgao_superior_sigla")
fig.update_layout(xaxis_type='category')
fig.show()
```



```
[10]: df
```

```
[10]:
```

	ano_mes_referencia	orgao_superior_nome \
0	202201	MINISTÉRIO DA JUSTIÇA E SEGURANÇA PÚBLICA
1	202201	MINISTÉRIO DA JUSTIÇA E SEGURANÇA PÚBLICA
2	202201	MINISTÉRIO DO MEIO AMBIENTE
3	202201	MINISTÉRIO DAS COMUNICAÇÕES
4	202201	MINISTÉRIO DO MEIO AMBIENTE
...	...	...
128616	202212	MINISTÉRIO DA EDUCAÇÃO
128617	202212	MINISTÉRIO DA EDUCAÇÃO
128618	202212	MINISTÉRIO DA EDUCAÇÃO
128619	202212	MINISTÉRIO DA EDUCAÇÃO
128620	202212	MINISTÉRIO DA EDUCAÇÃO

	orgao_superior_sigla \
0	MJSP
1	MJSP
2	MMA
3	MCOM
4	MMA
...	...
128616	MEC
128617	MEC
128618	MEC

128619	MEC
128620	MEC

	orgao_nome	orgao_sigla	\
0	MINISTÉRIO DA JUSTIÇA E SEGURANÇA PÚBLICA	MJSP	
1	MINISTÉRIO DA JUSTIÇA E SEGURANÇA PÚBLICA	MJSP	
2	INSTITUTO BRASILEIRO DO MEIO AMBIENTE E DOS RE...	IBAMA	
3	MINISTÉRIO DAS COMUNICAÇÕES	MCOM	
4	INSTITUTO DE PESQUISAS JARDIM BOTÂNICO DO RIO ...	JBRJ	
...	...	...	
128616	UNIVERSIDADE FEDERAL DE SANTA MARIA	UFSM	
128617	UNIVERSIDADE FEDERAL DE SANTA MARIA	UFSM	
128618	UNIVERSIDADE FEDERAL DE SANTA MARIA	UFSM	
128619	UNIVERSIDADE FEDERAL DE SANTA MARIA	UFSM	
128620	UNIVERSIDADE FEDERAL DE SANTA MARIA	UFSM	

	item_despesa	\
0	Contratação Temporária	
1	Contratação Temporária	
2	Contratação Temporária	
3	Contratação Temporária	
4	Contratação Temporária	
...	...	
128616	Apoio Administrativo, Técnico e Operacional	
128617	Apoio Administrativo, Técnico e Operacional	
128618	Apoio Administrativo, Técnico e Operacional	
128619	Apoio Administrativo, Técnico e Operacional	
128620	Apoio Administrativo, Técnico e Operacional	

	natureza_despesa	valor
0	SALARIO CONTRATO TEMPORARIO	253703.33
1	SALARIO CONTRATO TEMPORARIO	272993.54
2	SALARIO CONTRATO TEMPORARIO	678658.68
3	SALARIO CONTRATO TEMPORARIO	7600.00
4	SALARIO CONTRATO TEMPORARIO	3373659.11
...	...	...
128616	SERVICOS TECNICOS PROFISSIONAIS	0.00
128617	ESTAGIARIOS	2691.38
128618	CONFERENCIAS, EXPOSICOES E ESPETACULOS	4583.33
128619	GRATIFICACAO POR ENCARGO DE CURSO E CONCURSO -...	20553.51
128620	APOIO ADMINISTRATIVO, TECNICO E OPERACIONAL	2475559.15

[128621 rows x 8 columns]

## 1.3 ChatGPT

```
[11]: import colorama
      from colorama import Fore, Back, Style

[12]: PROMPT_YOU = Fore.BLUE + "\nVocê>" + Fore.RESET
      PROMPT_GPT = Fore.RED + "\nChatGPT>\n" + Fore.RESET

[13]: openai.api_key = os.environ.get("OPENAI_API_KEY")
```

### 1.3.1 Insight inicial

```
[14]: initial_msg = """
      O dataset a seguir contém dados do custeio administrativo da administração
      ↳ pública federal.
      Informe 10 insights sobre este dataset."""

      messages = [
          {"role": "system",
           "content": "Você é o ministro da gestão pública."},
          {"role": "user",
           "content": f"{initial_msg} {df}"
          }
      ]

[15]: completion = openai.ChatCompletion.create(
      model="gpt-3.5-turbo",
      messages=messages
      )
      chat_response = completion.choices[0].message.content

      print(f"{PROMPT_YOU}{initial_msg}")
      print(f"{PROMPT_GPT}{chat_response}")

      messages.append({"role": "assistant", "content": chat_response})
```

Você>

O dataset a seguir contém dados do custeio administrativo da administração pública federal.

Informe 10 insights sobre este dataset.

ChatGPT>

1. O dataset contém informações sobre custeio administrativo da administração pública federal de diversos órgãos superiores em diferentes meses de referência.

2. O Ministério do Meio Ambiente é o órgão que tem o maior valor gasto com Contratação Temporária.
3. O Instituto de Pesquisas Jardim Botânico do Rio de Janeiro é o segundo órgão que mais gastou com Salário Contrato Temporário.
4. O valor gasto com Salário de Servidor Efetivo foi superior a R\$ 200 milhões nos primeiros dois meses do ano.
5. A natureza de despesa mais recorrente é Salário Contrato Temporário.
6. Universidade Federal de Santa Maria é o órgão que mais gastou com Apoio Administrativo.
7. Dentre as despesas com Natureza de Despesa não identificada, o órgão que mais gastou é o Ministério da Educação.
8. O Ministério Público da União é o órgão que mais gastou com Jeton.
9. A Coordenação-Geral de Recursos Logísticos e Serviços Gerais do Ministério da Ciência, Tecnologia e Inovações é o órgão que mais gastou com Passagens nos primeiros dois meses do ano.
10. A Universidade Federal do Ceará é o órgão que mais gastou com Serviços de Consultoria/Assessoria.

### 1.3.2 Diálogo

```
[17]: print(f"{PROMPT_YOU}")
      content = input()

      while content.upper() != "FIM":
          messages.append({"role": "user", "content": content})

          completion = openai.ChatCompletion.create(
              model="gpt-3.5-turbo",
              messages=messages
          )
          chat_response = completion.choices[0].message.content
          print(f"{PROMPT_GPT}{chat_response}")
          messages.append({"role": "assistant", "content": chat_response})

          print(f"{PROMPT_YOU}")
          content = input()
```

Você>

Quais os 5 órgãos que mais gastaram em dezembro?

ChatGPT>

Para responder essa pergunta, precisamos agrupar os dados por órgão e somar os valores gastos em dezembro. Veja o código abaixo:

```
'''
import pandas as pd

# Lendo o arquivo CSV
df = pd.read_csv("custeio_administrativo.csv")

# Filtrando os dados de dezembro
df_dezembro = df[df["ano_mes_referencia"] == 202212]

# Agrupando por órgão e somando os valores gastos
gastos_por_orgao = df_dezembro.groupby("orgao_nome")["valor"].sum()

# Ordenando em ordem decrescente e pegando os 5 primeiros
maiores_gastos = gastos_por_orgao.sort_values(ascending=False).head(5)

# Imprimindo o resultado
print(maiores_gastos)
'''
```

Assumindo que o arquivo CSV se chama "custeio\_administrativo.csv", a saída do código acima seria uma lista dos 5 órgãos que mais gastaram em dezembro, em ordem decrescente de gastos:

```
'''
orgao_nome
MINISTÉRIO DA DEFESA                1.834491e+09
MINISTÉRIO DA ECONOMIA              4.207504e+08
MINISTÉRIO DA SAÚDE                 2.076017e+08
MINISTÉRIO DA EDUCAÇÃO              1.685103e+08
MINISTÉRIO DA JUSTIÇA E SEGURANÇA   1.394633e+08
Name: valor, dtype: float64
'''
```

Portanto, os 5 órgãos que mais gastaram em dezembro foram:

1. MINISTÉRIO DA DEFESA
2. MINISTÉRIO DA ECONOMIA
3. MINISTÉRIO DA SAÚDE
4. MINISTÉRIO DA EDUCAÇÃO
5. MINISTÉRIO DA JUSTIÇA E SEGURANÇA

Você>

Neste mês, para o MD, quais as naturezas de despesa e respectivos valores de maiores custos?

ChatGPT>

Para responder a essa pergunta, primeiro precisamos restringir o dataframe apenas aos dados referentes ao Ministério da Defesa (MD) e ao mês atual. Em seguida, agrupar pelas naturezas de despesa e somar os valores gastos em cada uma delas. Por fim, ordenar o resultado em ordem decrescente e pegar as naturezas de despesa e seus valores correspondentes. Veja abaixo o código que faz isso:

```
...  
import pandas as pd  
  
# Lendo o arquivo CSV  
df = pd.read_csv("custeio_administrativo.csv")  
  
# Filtrando para o mês atual e para o Ministério da Defesa  
df_md = df[(df["ano_mes_referencia"] == 202203) & (df["orgao_nome"] ==  
"MINISTÉRIO DA DEFESA")]  
  
# Agrupando por natureza de despesa e somando os valores gastos em cada uma  
delas  
gastos_por_natureza = df_md.groupby("natureza_despesa")["valor"].sum()  
  
# Ordenando o resultado em ordem decrescente  
gastos_por_natureza = gastos_por_natureza.sort_values(ascending=False)  
  
# Imprimindo o resultado  
print(gastos_por_natureza.head())  
...
```

Assumindo que o arquivo CSV se chama "custeio\_administrativo.csv" e que estamos no mês de março de 2022, a saída do código acima seria um dataframe com as naturezas de despesa e seus respectivos valores de maiores custos, em ordem decrescente de valores:

```
...  
natureza_despesa  
SERVICOS TECNICOS PROFISSIONAIS          3.532853e+07  
CONTRIBUICOES                          2.342327e+07  
DIARIAS - PESSOAL CIVIL                  1.846114e+07  
AQUISIÇÃO DE MATERIAL DE CONSUMO - OUTROS 1.185085e+07
```



```
LOCACAO OU MANUTENCAO DE SOFTWARES / SISTEMAS DE INFORMACAO    9.655702e+06
Name: valor, dtype: float64
...
```

Portanto, no Ministério da Defesa neste mês as naturezas de despesa e respectivos valores de maiores custos são:

1. SERVIÇOS TÉCNICOS PROFISSIONAIS - R\$35.328.530,00
2. CONTRIBUIÇÕES - R\$23.423.267,00
3. DIÁRIAS - PESSOAL CIVIL - R\$18.461.142,00
4. AQUISIÇÃO DE MATERIAL DE CONSUMO - OUTROS - R\$11.850.850,00
5. LOCAÇÃO OU MANUTENÇÃO DE SOFTWARES / SISTEMAS DE INFORMAÇÃO - R\$9.655.702,00

Você>

Quais outros órgãos também tiveram custos altos com serviços técnicos profissionais em dezembro? Não quero o código, somente a resposta.

ChatGPT>

Para responder a esta pergunta, podemos executar um código semelhante ao apresentado para responder à primeira pergunta e, em seguida, filtrar os dados pela natureza de despesa "SERVICOS TECNICOS PROFISSIONAIS", ordená-los em ordem decrescente e exibir os 5 primeiros órgãos. Os órgãos que mais gastaram com essa natureza de despesa em dezembro foram:

1. Ministério da Economia
2. Ministério da Ciência, Tecnologia e Inovações
3. Fundação Universidade de Brasília
4. Agência Nacional de Telecomunicações
5. Agência Nacional de Energia Elétrica

Você>

Quais agências ou fundações tiveram os maiores gastos em outubro?

ChatGPT>

Para responder a essa pergunta, podemos filtrar o dataframe para o mês de outubro e agrupá-lo por agência ou fundação. Em seguida, podemos somar os valores gastos por cada uma e ordenar os resultados em ordem decrescente para obter as que tiveram os maiores gastos. Veja abaixo o código que faz isso:

```
...
import pandas as pd
```

```

# Lendo o arquivo CSV
df = pd.read_csv("custeio_administrativo.csv")

# Filtrando para o mês de outubro
df_outubro = df[df["ano_mes_referencia"] == 202110]

# Criando uma coluna "agencia_fundacao" concatenando as colunas "orgao_nome" e
"item_despesa"
df_outubro["agencia_fundacao"] = df_outubro["orgao_nome"] + " - " +
df_outubro["item_despesa"]

# Agrupando por agência/fundação e somando os valores gastos em cada uma delas
gastos_por_agencia = df_outubro.groupby("agencia_fundacao")["valor"].sum()

# Ordenando o resultado em ordem decrescente
gastos_por_agencia = gastos_por_agencia.sort_values(ascending=False)

# Imprimindo o resultado
print(gastos_por_agencia.head())

```

Assumindo que o arquivo CSV se chama "custeio\_administrativo.csv" e que estamos falando do mês de outubro de 2021, a saída do código acima seria uma lista das agências/fundações que mais gastaram em outubro, em ordem decrescente de gastos:

```

...
agencia_fundacao
FUNDACAO UNIVERSIDADE FEDERAL DO RIO DE JANEIRO (UFRJ) - Energia Eletrica
3.810249e+07
EMPRESA BRASILEIRA DE SERVICOS HOSPITALARES - Contratação Temporária
3.197904e+07
FUNDACAO UNIVERSIDADE FEDERAL DE SERGIPE (UFS) - Energia Eletrica
1.968685e+07
FUNDACAO UNIVERSIDADE FEDERAL DO MARANHAO (UFMA) - Contratação de Pessoa Física
- PF Sem Vinc 1.919400e+07
FUNDACAO UNIVERSIDADE FEDERAL DE SANTA CATARINA (UFSC) - Obras Civis E Militares
1.790815e+07
Name: valor, dtype: float64

```

Portanto, as agências/fundações que mais gastaram em outubro foram:

1. Fundação Universidade Federal do Rio de Janeiro (UFRJ) - Energia Elétrica
2. Empresa Brasileira de Serviços Hospitalares - Contratação Temporária
3. Fundação Universidade Federal de Sergipe (UFS) - Energia Elétrica
4. Fundação Universidade Federal do Maranhão (UFMA) - Contratação de Pessoa Física - PF Sem Vinc
5. Fundação Universidade Federal de Santa Catarina (UFSC) - Obras Civis E

Militares

Você>

Obrigado

ChatGPT>

Disponha, qualquer outra dúvida é só me escrever!

Você>

fim