## **Relatório Técnico — Projeto de Ciência de Dados**

### **Pergunta Orientadora**

**Como é distribuído os gastos públicos entre áreas como saúde, educação e segurança ao longo do ano?**

### **Objetivo do Projeto**

O objetivo deste projeto é analisar a distribuição de gastos públicos em diferentes áreas temáticas (como saúde, educação e segurança) com base em dados públicos de despesas. Através dessa análise, buscamos identificar padrões, tendências e possíveis variações mensais ou sazonais ao longo do ano.

### **Fonte de Dados**

Os dados utilizados neste projeto são provenientes de arquivos .csv contendo informações sobre a execução de despesas públicas obtidos no portal gov, [Execução da Despesa](https://portaldatransparencia.gov.br/download-de-dados/despesas-execucao). Esses arquivos foram organizados em uma pasta no ambiente Colab (/content/content/) e lidos automaticamente com o auxílio da biblioteca glob.

### **Etapas Realizadas**

#### **1. Detecção de Codificação dos Arquivos**

Antes de carregar os dados, foi utilizada a biblioteca chardet para detectar a codificação correta dos arquivos .csv, garantindo a leitura adequada, especialmente considerando o uso de acentos e caracteres especiais em português.

#### **2. Leitura e Consolidação dos Arquivos**

Todos os arquivos CSV presentes no diretório foram lidos e armazenados em uma lista de DataFrames e consolidados em um só arquivo

#### **3. Leitura de Arquivo Consolidado**

Um arquivo chamado execucao\_despesa\_consolidado.csv foi baixado para análise no Power BI

#### **4. ETL no Power BI**

O arquivo consolidado foi usado como fonte de dados no Power BI, sendo feito o tratamento do dados e criação de novas colunas para facilitar na criação de gráficos e filtros

#### **5. Criação de medidas e gráficos**

Depois de ter sido feito o ETL, foram criadas medidas do tipo, Total Valor Pago, **Total Valor Liquidado** e **Diferença**, que seria a diferença entre o Total Valor Pago e Total Valor Liquidado.

Com as medidas prontas foi possível criar os seguintes gráficos, **Top 10 Funções com Maior Valor Pago** e **Top 10 Órgãos Superiores com Maior Valor Pago**. Os gráficos foram filtrados para mostrar apenas as 10 Funções e 10 Órgãos Superiores com maior valor pago, através do filtro N Superior na função ou órgão superior e passando a referência da soma valor pago, ou seja, o total do valor pago.

### **Ferramentas Utilizadas**

* **Python (Google Colab)**
* Bibliotecas: pandas, glob
* **Power BI**