

# Integração de Sistemas de Informação-LESI

TRABALHO PRÁTICO I

**Aluno (a):**

Eva Gomes a27484

## Lista de Abreviaturas e/ou Siglas

ETL – Extract, Transform, Load (Extração, Transformação e Carga)

ISI – Integração de Sistemas de Informação

KNIME – Konstanz Information Miner (Plataforma de análise e ETL)

API – Application Programming Interface (Interface de Programação de Aplicações)

CSV – Comma-Separated Values (Valores Separados por Vírgulas)

JSON – JavaScript Object Notation (Formato de intercâmbio de dados)

XML – eXtensible Markup Language (Linguagem de marcação extensível)

YAML – YAML Ain't Markup Language (Formato de serialização de dados)

KPI – Key Performance Indicator (Indicador-Chave de Desempenho)

ONG – Organização Não Governamental

BI – Business Intelligence (Inteligência de Negócio)

Power BI – Plataforma de análise e visualização de dados da Microsoft

Google Apps Script – Linguagem de script para automação no ecossistema Google

FTP – File Transfer Protocol (Protocolo de Transferência de Ficheiros)

ER – Expressões Regulares

Dashboard – Painel de controlo interativo para visualização de dados

## Índice

Lista de Abreviaturas e/ou Siglas .....	1
Índice de Ilustrações .....	4
Introdução .....	5
1. Tema do Trabalho .....	6
1.1 Promoção da Ciência em Escolas Primárias.....	6
2. Problema a Resolver .....	6
2.1 Fontes de Dados .....	6
2.2 Problemas Identificados .....	6
2.3 Necessidade do Sistema .....	7
3. Objetivos do Trabalho .....	7
3.1 Desenvolvimento do Processo ETL .....	7
3.2 Exportação de Dados.....	7
3.3 Visualização .....	7
3.4 Ferramentas Utilizadas .....	8
4. Estrutura dos Ficheiros de Dados.....	9
4.1 Ficheiro de Municípios.....	9
4.2 Ficheiro de Cientistas .....	9
5. Importação e Tratamento de Dados .....	10
5.1 Ficheiro de Municípios.....	10
5.2 Ficheiro de Cientistas.....	11
6. Conexão e Autenticação .....	12
7. Tratamento de Dados.....	13
7.1 Cientistas.....	13
7.2 Municípios .....	14
8. Integração e Análise de Dados.....	15
8.1 Integração dos Ficheiros.....	15
8.2 Análise de Municípios/Atividades e Cientistas Inscritos .....	16
9. Exportação dos Ficheiros .....	16
9.1 JSON e XML .....	17
9.2 Ficheiros XLS/XLSX (Registo de Linhas Rejeitadas).....	17
9.3 Ficheiros Excel para envio de Email .....	18

10. Envio de Emails .....	19
10.1 Email com Lista de Cientistas por Município .....	19
10.2 Email Automático de Novos Cientistas.....	19
11. Visualização de Dados .....	20
11.1 Gráficos no KNIME.....	20
11.2 Análise e Dashboards no Power BI .....	21
12. Trabalhos Futuros .....	25
Conclusão .....	27
Anexos .....	28
Anexo 1- Processos e Jobs no Knime .....	28
Anexo 2- QR Code para visualização do funcionamento dos processos .....	29
Referências Bibliográficas.....	30

## Índice de Ilustrações

Figura 1-Extrato do Ficheiro Municípios.....	9
Figura 2-Extrato do ficheiro Cientistas.....	10
Figura 3-Excerto Anexo 1- Importação de Dados .....	12
Figura 4--Excerto Anexo 1. Conexão e Autenticação.....	13
Figura 5-Excerto Anexo 1-Tratamento de Dados Cientistas .....	14
Figura 6--Excerto Anexo 1-Tratamentos de Dados Municípios .....	15
Figura 7-Excerto Anexo 1-Integração e Análise de Dados .....	16
Figura 8-Excerto Anexo 1 -Exportação JSON/XML.....	17
Figura 9-Excerto Anexo 1-Exportação XLS/XSLX .....	18
Figura 10--Excerto Anexo 1-Exportação Excel Email .....	19
Figura 11-Excerto Anexo 1-Envio de Emails.....	20
Figura 12-Excerto Anexo 1-Visualização de dados.....	20
Figura 13-Gráfico de Barras para número de pedidos por município no Knime ....	21
Figura 14-Gráfico de Barras-Comparação entre número de pedidos do por município e número de inscrições de cientistas por município .....	22
Figura 15-Gráfico de Barras-Comparação entre número de pedidos do por distrito e número de inscrições de cientistas por distrito .....	22
Figura 16-Tabela-Soma de pedidos por distrito .....	23
Figura 17-Gráfico de Barras-Número de cientistas inscritos por município .....	24
Figura 18-Pie Chart-Número de cientistas inscritos por distrito .....	24
Figura 19-Visualização em mapa-Número de cientistas inscritos por distrito .....	25
 Tabela 1-Código Apps Scripts.....	 11

## Introdução

Este relatório foi elaborado como trabalho individual no âmbito da unidade curricular Integração de Sistemas Informáticos, do curso de Licenciatura em Sistemas Informáticos (Ensino Pós-Laboral) no IPCA.

A promoção da ciência junto de crianças do ensino básico é um desafio que requer planeamento, organização e integração de dados de diversas fontes. Este trabalho foi desenvolvido no âmbito de uma iniciativa de uma ONG que visa levar cientistas a escolas primárias, nos municípios onde estudaram, criando uma ligação emocional e educativa entre os profissionais da ciência e as comunidades locais.

O objetivo principal deste projeto foi implementar um processo ETL (Extract, Transform, Load) capaz de consolidar, tratar, analisar e visualizar os dados de municípios e cientistas, permitindo:

- Organizar e padronizar os dados recolhidos por formulários e ficheiros Excel;
- Cruzar as informações entre municípios e cientistas;
- Criar estatísticas e dashboards para apoiar o planeamento das visitas;
- Automatizar a comunicação com os municípios e cientistas inscritos.

As ferramentas utilizadas incluíram Google Forms, Google Sheets, Google Apps Script, KNIME e Power BI, garantindo um fluxo automatizado e eficiente desde a recolha de dados até à análise e comunicação dos resultados.

# 1. Tema do Trabalho

## 1.1 Promoção da Ciência em Escolas Primárias

O presente trabalho propõe o desenvolvimento de um processo ETL para apoiar uma iniciativa de uma ONG dedicada à promoção da ciência junto de crianças do ensino básico.

A iniciativa consiste em levar cientistas a escolas primárias localizadas nos municípios onde estudaram, criando uma ligação emocional e educativa entre os profissionais da ciência e as comunidades locais.

## 2. Problema a Resolver

### 2.1 Fontes de Dados

A ONG recolhe informações de duas fontes principais:

- **Cientistas:** indicam a escola primária onde estudaram e o respetivo município.
- **Municípios:** manifestam interesse em receber visitas de cientistas e o número de atividades que pretendem desenvolver.

### 2.2 Problemas Identificados

Os dados recolhidos apresentam alguns desafios:

- Estão dispersos e sem ligação direta entre cientistas e municípios.
- Possuem erros e inconsistências, principalmente nos campos preenchidos manualmente.
- Os municípios não têm dropdowns para padronizar os nomes, resultando em preenchimentos incorretos.
- Há necessidade de organizar, tratar e cruzar os dados de forma eficiente.

## 2.3 Necessidade do Sistema

O sistema deve ser capaz de:

- Organizar e tratar os dados recolhidos.
- Fazer o cruzamento entre cientistas e municípios interessados.
- Criar informação útil para o planeamento das visitas.
- Permitir exportar os dados para formatos acessíveis e visualizáveis.

## 3. Objetivos do Trabalho

### 3.1 Desenvolvimento do Processo ETL

- **Extração:** obter dados de formulários ou ficheiros CSV/XLSX.
- **Transformação:** limpar e normalizar os dados (ex.: nomes de municípios e escolas).
- **Integração:** cruzar os dados entre cientistas e municípios interessados.
- **Cálculo de Estatísticas:** determinar, por exemplo, o número de cientistas por município.

### 3.2 Exportação de Dados

- Criar ficheiros para partilha com parceiros e integração com outras plataformas.

### 3.3 Visualização

- Criar dashboards no Power BI que apresentem:
  - Número de visitas planeadas por município.
  - Número de cientistas inscritos por município.
  - Cientistas que se inscrevem pela primeira vez para determinado município, com alertas automáticos (ex.: email).



### 3.4 Ferramentas Utilizadas

Durante o desenvolvimento deste trabalho foram utilizadas diversas ferramentas para extração, transformação, integração, análise e visualização de dados:

#### 1. **Google Forms**

- Para recolher automaticamente as inscrições dos cientistas.
- Permite padronizar alguns campos através de dropdowns (ex.: distrito, nível de educação).

#### 2. **Google Sheets**

- Armazena as respostas do Google Form em tempo real.
- Facilita a integração com KNIME através de conectores e APIs.

#### 3. **Google Apps Script**

- Criou uma API que transforma o Google Sheet em ficheiro .xlsx.
- Configuração de triggers para atualizar automaticamente os dados a cada nova submissão.

#### 4. **KNIME Analytics Platform**

- Ambiente principal para desenvolvimento do processo ETL: importação, tratamento, integração, agregação e exportação de dados.

#### 5. **Power BI**

- Ferramenta utilizada para visualização e análise interativa de dados.
- Importação dos ficheiros XML exportados do KNIME para criação de dashboards e estatísticas consolidadas.

## 4. Estrutura dos Ficheiros de Dados

### 4.1 Ficheiro de Municípios

- Contém colunas: Distrito, Município e Número de Atividades que desejam receber.
- Tanto os municípios quanto os distritos possuem uma designação própria criada pela instituição.
- Os ficheiros são preenchidos pelos colaboradores, mas não possuem dropdowns, resultando em muitos erros de preenchimento manual.

Concelho (nomenclatura Native ▾)	Distrito (nomenclatura Native ▾)	Número de Atividade ▾
Águeda	Aveiro	6
Albergaria-a-Velha	Aveiro	3
Anadia	Aveiro	
Arouca	Aveiro	
Aveiro	Aveiro	10
Castelo de Paiva	Aveiro	7
Espinho	Aveiro	
Estarreja	Aveiro	4
Santa Maria da Feira	Aveiro	6
Ílhavo	Aveiro	9
Mealhada	Aveiro	2
Murtosa	Aveiro	6

Figura 1-Extrato do Ficheiro Municípios

### 4.2 Ficheiro de Cientistas

- Inclui colunas como Distrito, Município, Escola, Email, Nome, Apelido e Nível de Educação.
- O campo Distrito possui dropdown, garantindo que segue as regras da instituição.
- O Nível de Educação também é selecionado através de dropdown (ex.: Ensino Secundário, Licenciatura).

- Apesar da padronização em alguns campos, ainda podem existir inconsistências nos dados preenchidos manualmente.

Timestamp	Nome	Apelido	Email	Nível de Educação concluído	Distrito	Município	Escola
11/10/2025 16:00:54	Alexandre	Vieira	alexandremmieira21@	Licenciatura	Porto	Vila do conde	Escola dos Peixinhos
11/10/2025 16:01:01	Eva	Gomes	evagomes00@gmail.co	Licenciatura	Braga	Vila Nova de Famalicão	EB Antas
14/10/2025 03:37:00	Tiago	Gomes	tiago.gomes@example.c	Ensino Secundário	Coimbra	Góis	Escola Básica de Góis
19/10/2025 23:44:00	Beatriz	Martins	beatriz.martins@examp	Ensino Básico	Viana do Castelo	Arcoos de Vaaldeeveez	Escola Básica de Arcos de Valde
14/10/2025 15:45:00	Sofia	Santos	sofia.santos@example.c	Mestrado	Setúbal	Seesimbria	Escola Básica de Sesimbria
20/10/2025 10:57:00	Tiago	Rodrigues	tiago.rodrigues@examp	Ensino Secundario	Viseu	Vilaa Noovaa dee Paaiva	Escola Básica de Vila Nova de P
14/10/2025 18:16:00	Rui	Pereira	rui.pereira@example.co	Ensino Basico	Castelo Branco	Vila Velha de Ródão	Escola Basika de Vila Velha de
20/10/2025 15:49:00	Miguel	Rodrigues	miguel.rodrigues@exam	Licenciatura	Braga	Fafe	Escola Basika de Fafe
12/10/2025 02:45:00	Maria	Rodrigues	maria.rodrigues@examp	Ensino Secundário	Braga	Fafe	Escola Basika de Fafe
20/10/2025 13:00:00	Ana	Costa	ana.costa@example.con	Ensino Primário	Faro	Olhão	Escola Básica de Olhão
12/10/2025 11:19:00	Inês	Ferreira	inês.ferreira@example.c	Ensino Secundário	Leiria	Castanheira de Pêra	Escola Básica de Castanheira de
14/10/2025 00:08:00	Tiago	Santos	tiago.santos@example.c	Mestrado	Setúbal	Sinees	Escola Basika de Sines

Figura 2-Extrato do ficheiro Cientistas

## 5. Importação e Tratamento de Dados

Para apoiar a iniciativa de promoção da ciência em escolas primárias, foi necessário consolidar e organizar dados de diferentes fontes. O processo de importação de dados foi dividido em duas principais fontes: municípios e cientistas.

### 5.1 Ficheiro de Municípios

Os dados sobre municípios foram fornecidos através de um ficheiro local em formato Excel. Para realizar a importação, utilizou-se a biblioteca Node Excel Reader, que permite ler e processar ficheiros Excel diretamente no ambiente Node.js.

O processo consistiu em:

1. Seleção do ficheiro local contendo os dados dos municípios.
2. Leitura das colunas relevantes (nome do município, distrito, número de atividades, etc.).
3. Conversão dos dados para um formato compatível com o ETL, permitindo a integração com os outros ficheiros do projeto.

## 5.2 Ficheiro de Cientistas

Os dados dos cientistas são recolhidos através de um Google Form, que regista automaticamente as respostas num Google Sheet. Para integrar este ficheiro ao processo ETL no KNIME:

1. Foi criada uma API em Google Apps Script que transforma o Google Sheet em ficheiro .xlsx. Foi configurado um trigger para que o ficheiro seja sempre atualizado automaticamente a cada nova submissão.

Tabela 1-Código Apps Scripts

```
function exportAsExcelXLSX() {
    var spreadsheet = SpreadsheetApp.getActiveSpreadsheet();
    var fileId = spreadsheet.getId();
    var fileName = spreadsheet.getName() + ".xlsx"; // Nome do
    ficheiro Excel

    // Exporte spreadsheet as XLSX via Drive API
    var url = "https://www.googleapis.com/drive/v3/files/" +
    fileId + "/export?mimeType=application/vnd.openxmlformats-
    officedocument.spreadsheetml.sheet";
    var options = {
        headers: {
            "Authorization": "Bearer " + ScriptApp.getOAuthToken()
        },
        muteHttpExceptions: true
    };

    var response = UrlFetchApp.fetch(url, options);
    var excelBlob = response.getBlob().setName(fileName);

    // Destiny Folder
    var folder =
    DriveApp.getFolderById("1CojjeJkpDxxi_D37Qdj_L3wfgCthVlvi");

    // Search file with the same name
    var files = folder.GetFilesByName(fileName);
    while (files.hasNext()) {
        var file = files.next();
        file.setTrashed(true); // move to trash the old file
    }

    // create new file
    folder.createFile(excelBlob);
}
```

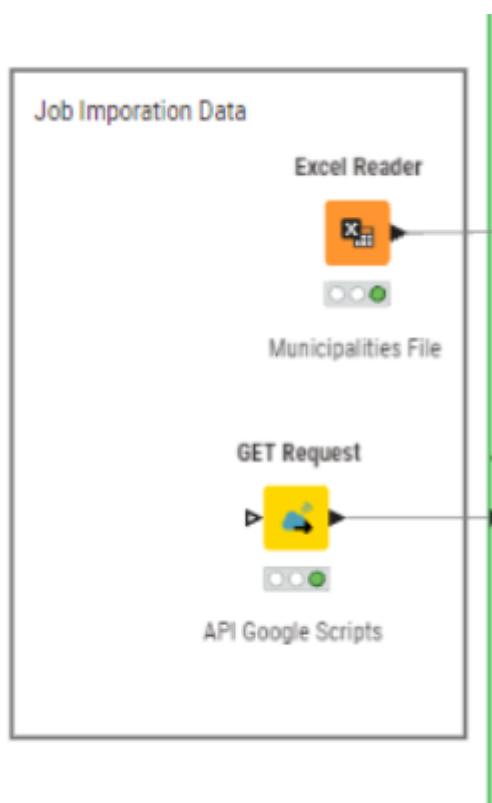
2. No KNIME, foi utilizado um nó GET Request para aceder diretamente à API e obter os dados em tempo real, eliminando a necessidade de fazer o download manual.

3. Os dados obtidos foram depois processados pelo mesmo fluxo de leitura usado para o ficheiro de municípios, garantindo consistência na estrutura dos dados.

**Vantagens deste fluxo:**

- Qualquer nova submissão no Google Form é automaticamente disponibilizada para o ETL, evitando processos manuais.
- A uniformização dos formatos dos ficheiros facilita a etapa de tratamento e integração de dados, abordada na secção seguinte.

Na figura seguinte temos um excerto do Anexo 1, que representa a importação dos dados no Knime.



*Figura 3-Excerto Anexo 1- Importação de Dados*

## 6. Conexão e Autenticação

Para possibilitar a exportação de ficheiros para o Google Drive, foi necessário estabelecer ligação com a conta da ONG. Para isso, foram utilizados os seguintes componentes no KNIME:

- **Google Authentication:** autenticação segura na conta do Google.
- **Google Sheets Connector:** conexão direta com ficheiros Google Sheets para leitura e escrita de dados.
- **Google Drive Connector:** permite exportar e armazenar os ficheiros processados diretamente no Google Drive.

Esta configuração garante que os ficheiros tratados no ETL possam ser automaticamente guardados e partilhados, evitando manipulação manual.

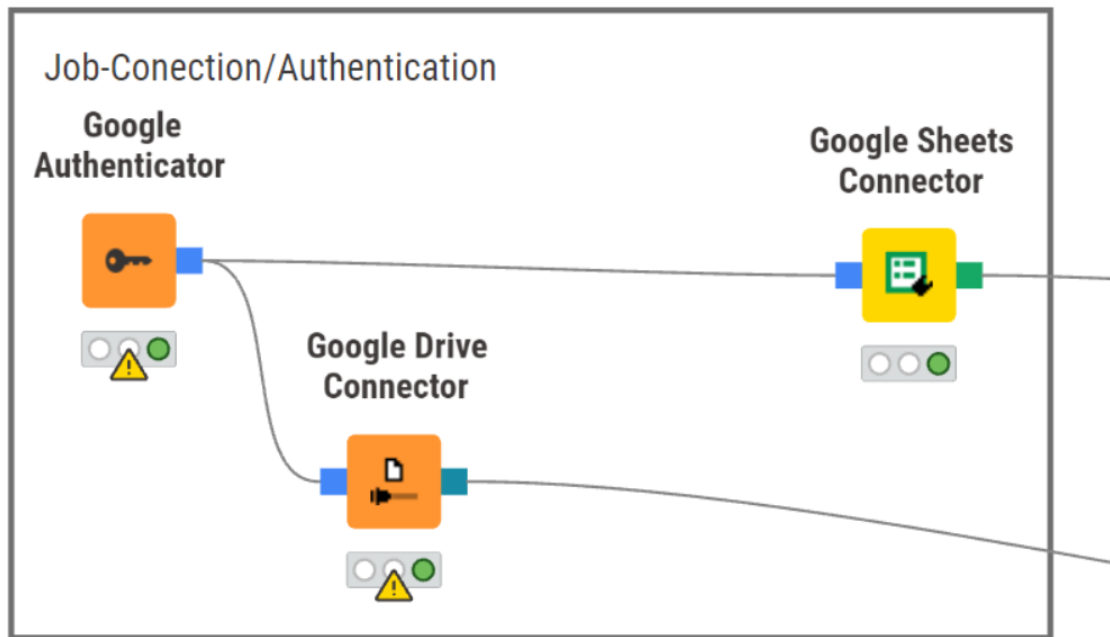


Figura 4--Excerto Anexo 1. Conexão e Autenticação

## 7. Tratamento de Dados

### 7.1 Cientistas

Os dados dos cientistas, obtidos via GET Request, são devolvidos em formato JSON, contendo todas as submissões do formulário. Para extrair os campos relevantes (nome do cientista, área de especialização, município de origem e escola primária), foi utilizado JSON Path, permitindo navegar na estrutura do JSON e selecionar apenas os dados necessários.

Além disso, os dados passaram por um processo de limpeza e normalização, incluindo:

- Eliminação de espaços em excesso e letras duplicadas utilizando expressões regulares.
- Uniformização de letras maiúsculas e minúsculas para coincidir com as designações usadas pelos municípios.
- Correção de inconsistências nos nomes de municípios e escolas, garantindo compatibilidade com o ficheiro de municípios.
- Filtragem de dados: inscrições de pessoas que não possuam, no mínimo, licenciatura foram automaticamente eliminadas. Os dados removidos foram guardados em ficheiro separado para posterior análise.

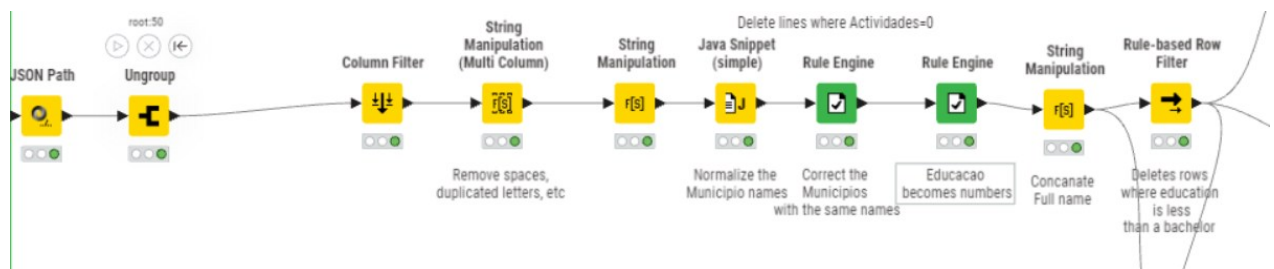


Figura 5-Excerto Anexo 1-Tratamento de Dados Cientistas

## 7.2 Municípios

O ficheiro de municípios também passou por tratamento de dados para garantir consistência:

- Municípios cujas atividades estavam vazias ou como null tiveram o valor substituído por 0.
- Municípios que não solicitaram atividades foram eliminados do ficheiro principal, sendo armazenados separadamente, apenas para manter registo histórico.

Estas operações garantem que apenas os municípios com interesse real em receber visitas de cientistas sejam considerados no planeamento e cruzamento de dados.

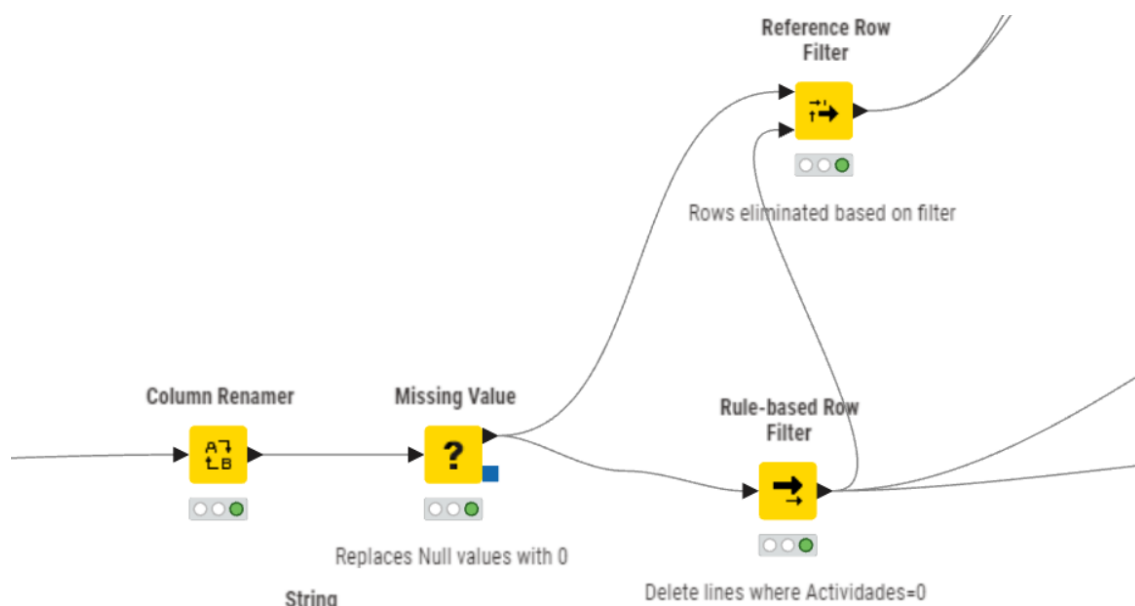


Figura 6--Excerto Anexo 1-Tratamentos de Dados Municípios

Foram também utilizados em ambos os ficheiros outros nós com os objetivos de renomear colunas, de enumerar por ordem descendente, entre outros.

## 8. Integração e Análise de Dados

Para consolidar e analisar os dados de cientistas e municípios, foi realizado um processo de integração e agregação utilizando os nós Joiner e Group By no KNIME.

### 8.1 Integração dos Ficheiros

- Os ficheiros de cientistas e municípios foram unidos através da coluna Município usando o nó Joiner.
- Este processo permite associar cada cientista ao respetivo município, garantindo que os dados estejam alinhados para análise posterior.
- Após o Joiner, outro nó Group By foi aplicado para criar uma tabela detalhada contendo:
  - Lista de nomes dos cientistas inscritos por município.
  - Emails dos cientistas correspondentes.
  - Quantidade total de cientistas inscritos por município.



## 8.2 Análise de Municípios/Atividades e Cientistas Inscritos

- Foi utilizado um Group By adicional para identificar os municípios que solicitaram atividades e a quantidade de atividades pedidas. Esta tabela permite visualizar quais municípios estão efetivamente interessados em receber visitas de cientistas e quantas atividades cada município deseja realizar.
- O nó Group By foi utilizado para criar uma tabela com o número de cientistas inscritos por município.

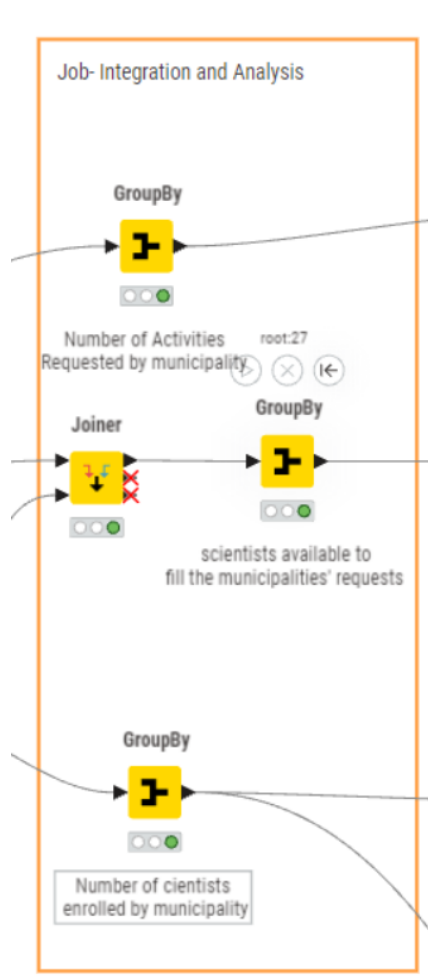


Figura 7-Excerto Anexo 1-Integração e Análise de Dados

## 9. Exportação dos Ficheiros

Após o tratamento e análise dos dados, os ficheiros finais foram exportados em três formatos principais, cada um com uma finalidade específica.

## 9.1 JSON e XML

- Foram criados ficheiros JSON a partir das tabelas resultantes dos três Group By.
- O processo incluiu os seguintes passos:
  - **Table to JSON:** conversão da tabela para JSON.
  - **JSON to XML:** transformação dos dados JSON para XML.
  - **XML Writer:** escrita final em ficheiro XML.
- Estes ficheiros permitem partilhar os dados de forma estruturada e também integrar com outras plataformas que utilizem XML.



Figura 8-Excerto Anexo 1 -Exportação JSON/XML

## 9.2 Ficheiros XLS/XLSX (Registo de Linhas Rejeitadas)

- Estes ficheiros contêm os dados filtrados pelo Rule-based Row Filter, ou seja, registos que não cumpriram os critérios, como cientistas sem licenciatura.
- A criação dos ficheiros utilizou:
  - Google Connector para integração com Google Drive.
  - Java Edit Variable, que permite criar nomes de ficheiros que incluem a data, mantendo um histórico das execuções.
- **Limitações:**
  - Com o Google Sheet Writer, não é possível selecionar a pasta de destino, ficando os ficheiros na raiz do Drive.
  - Para contornar, foi usado o Google Drive Connector + Excel Writer, permitindo escolher a pasta de destino.

- Porém, não foi possível adicionar data e hora automaticamente em cada execução usando este método.

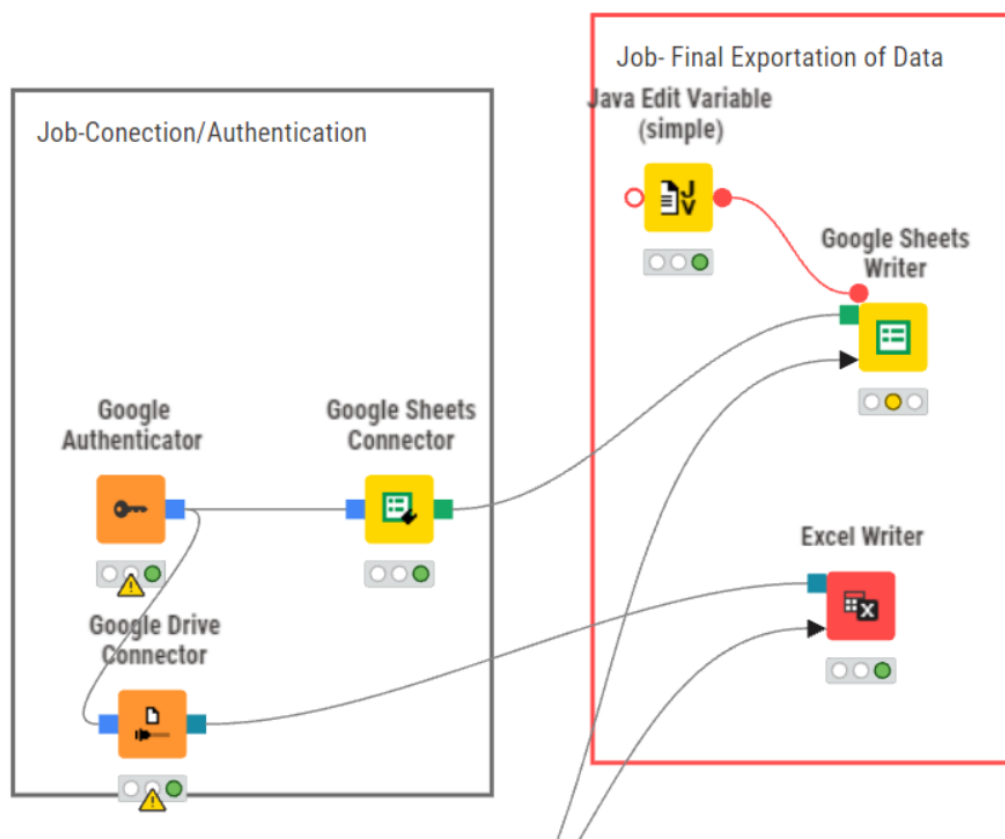


Figura 9-Excerto Anexo 1-Exportação XLS/XSLX

### 9.3 Ficheiros Excel para envio de Email

- Foi criado também um ficheiro Excel contendo a lista de nomes e emails dos cientistas que se inscreveram e correspondem aos pedidos dos municípios.
- Este ficheiro será usado para envio de emails, informando os cientistas sobre as visitas que irão realizar, caso sejam selecionados.

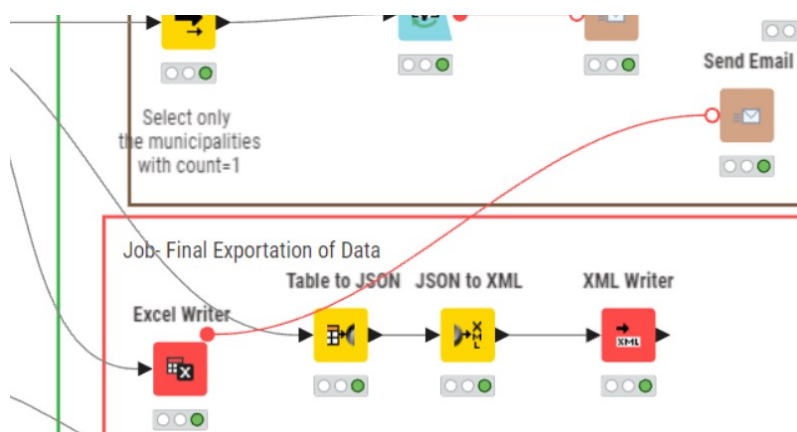


Figura 10--Excerto Anexo 1-Exportação Excel Email

## 10. Envio de Emails

Após a integração e exportação dos dados, foram configurados dois tipos de envio de emails automáticos para os municípios e para monitoramento das inscrições de cientistas.

### 10.1 Email com Lista de Cientistas por Município

- É enviado um email contendo um ficheiro Excel com a lista de nomes e emails dos cientistas que se inscreveram e correspondem aos pedidos de atividades dos municípios.
- Este email permite que os municípios tenham acesso direto às informações dos cientistas que irão participar das visitas.

### 10.2 Email Automático de Novos Cientistas

- Um segundo tipo de email é enviado em loop, utilizando o Rule-based Row Filter para identificar inscrições de cientistas que se inscrevem pela primeira vez para determinado município.
- Cada município recebe automaticamente uma notificação informando que há um novo cientista inscrito que poderá participar nas atividades.
- Este processo automatizado garante comunicação rápida e precisa, permitindo que os municípios estejam sempre atualizados sobre novos participantes.

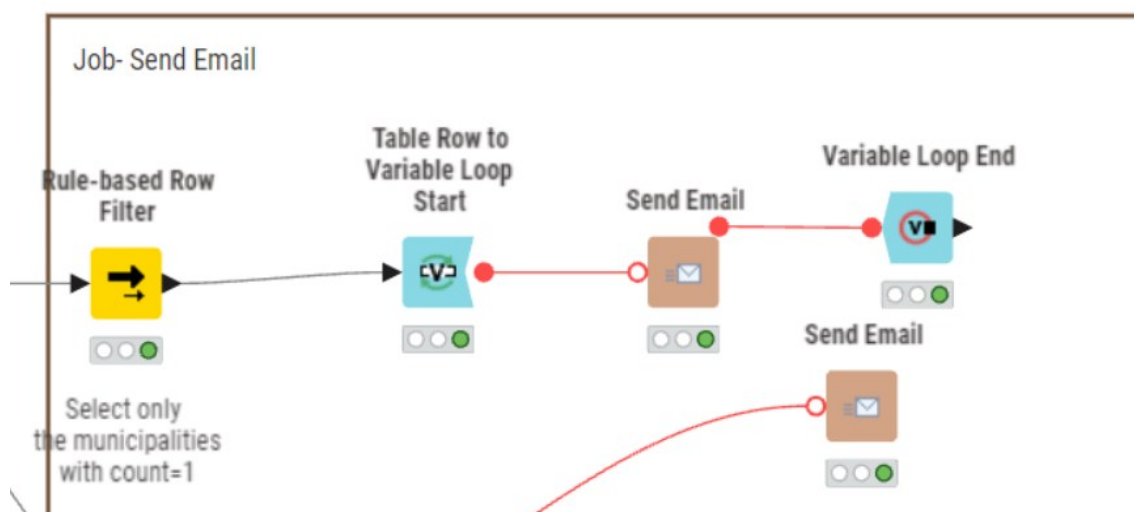


Figura 11-Excerto Anexo 1-Envio de Emails

## 11. Visualização de Dados

Após a integração, análise e exportação dos dados, foi realizada a visualização de resultados para facilitar a interpretação e apoio ao planeamento das visitas.

### 11.1 Gráficos no KNIME

- Foi criado um exemplo de gráfico de barras diretamente no KNIME.
- A tabela de dados foi conectada aos nós de visualização e o texto foi utilizado como título do gráfico, configurado através dos nós de formatação.
- Esta visualização serve como exemplo rápido de análise e verificação dos dados dentro do KNIME, mas não é o principal método de análise.

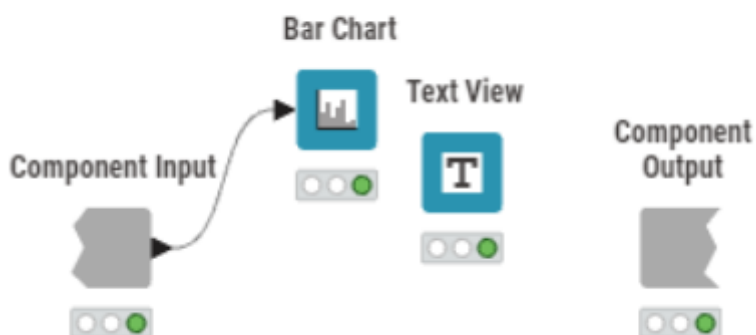


Figura 12-Excerto Anexo 1-Visualização de dados

## Número de Cientistas por Município

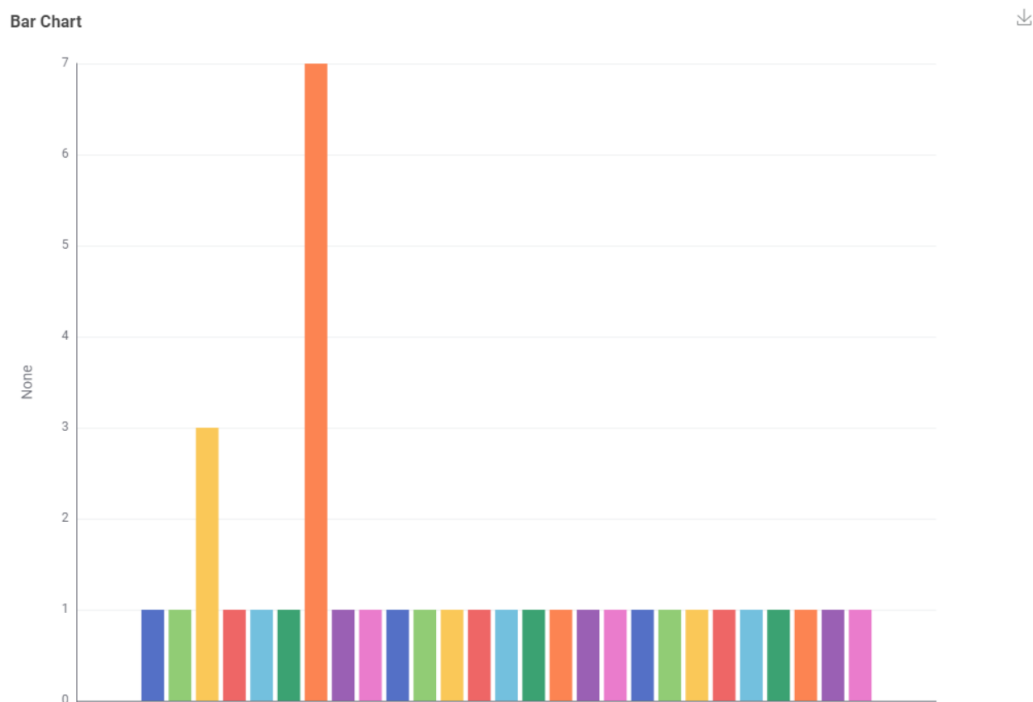


Figura 13-Gráfico de Barras para número de pedidos por município no Knime

## 11.2 Análise e Dashboards no Power BI

- Para análises mais avançadas e relatórios interativos, os ficheiros XML exportados do KNIME foram importados para o Power BI.
- No Power BI, foram criados dashboards que permitem:
  - Comparação entre número de pedidos do por município/distrito e número de inscrições de cientistas por

## município/distrito

Soma de Número de Cientistas e Soma de Pedidos do Município por Município

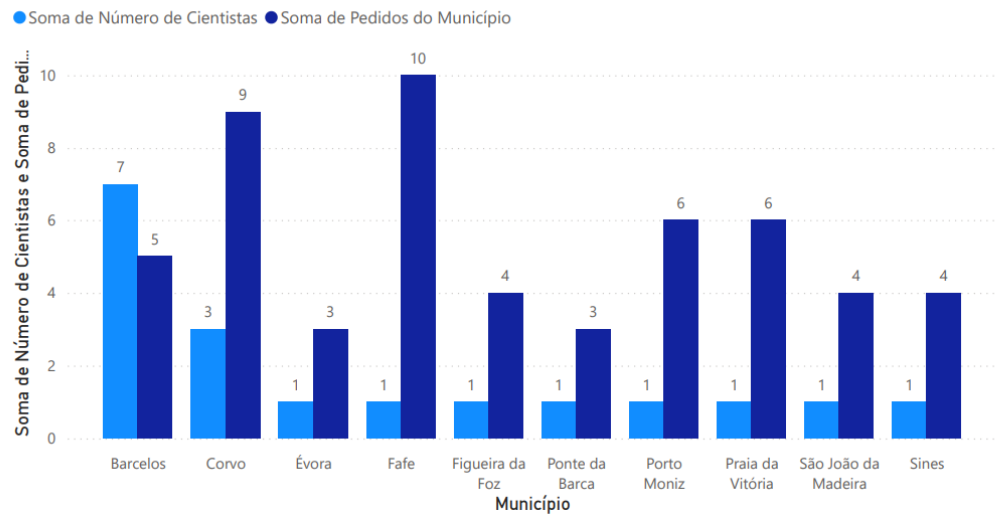


Figura 14-Gráfico de Barras-Comparação entre número de pedidos do por município e número de inscrições de cientistas por município

Soma de Pedidos do Município e Soma de Número de Cientistas por Distrito

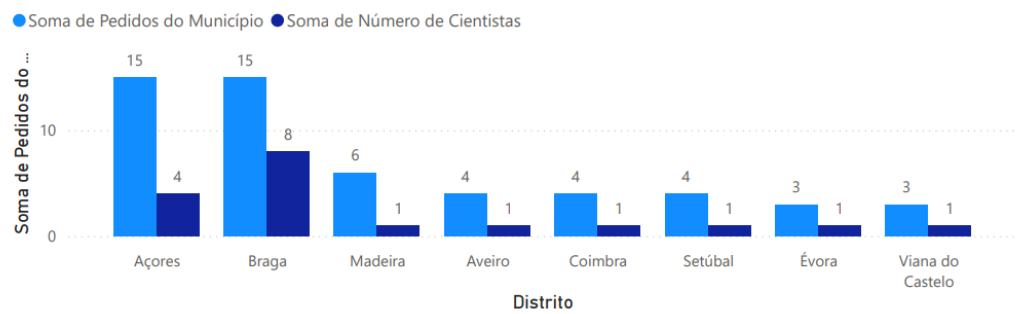


Figura 15-Gráfico de Barras-Comparação entre número de pedidos do por distrito e número de inscrições de cientistas por distrito

- Consultar soma de pedidos por distrito

Soma de Pedidos	Distrito
19	Açores
19	Aveiro
14	Beja
14	Braga
12	Bragança
11	Castelo Branco
17	Coimbra
14	Évora
16	Faro
14	Guarda
16	Leiria
16	Lisboa
11	Madeira
15	Portalegre
18	Porto
21	Santarém
13	Setúbal
10	Viana do Castelo
14	Vila Real
24	Viseu
<b>308</b>	

Figura 16-Tabela-Soma de pedidos por distrito



○ Visualizar o número de cientistas inscritos por município/distrito

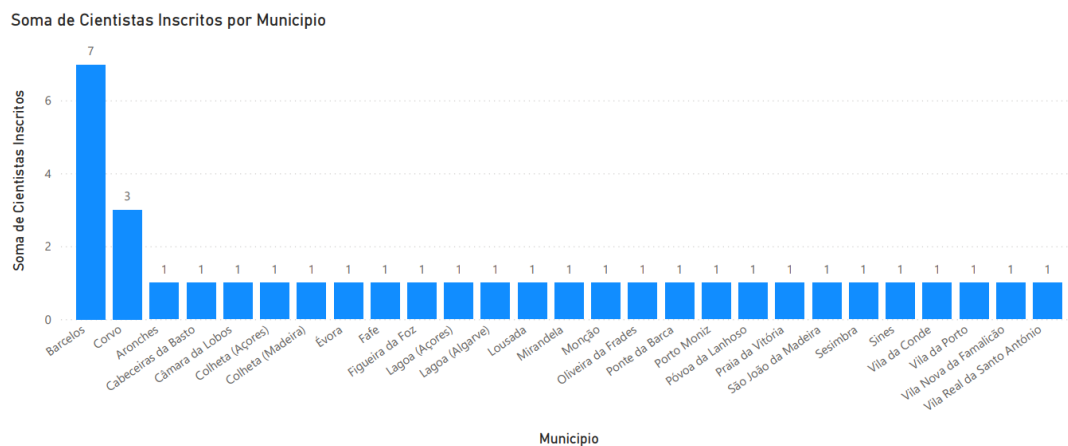


Figura 17-Gráfico de Barras-Número de cientistas inscritos por município

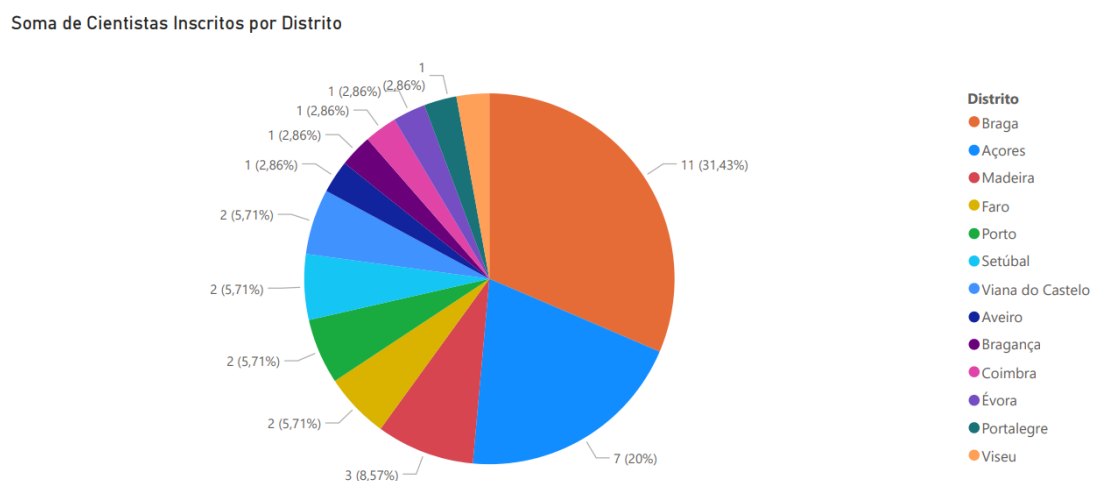


Figura 18-Pie Chart-Número de cientistas inscritos por distrito

Soma de Cientistas Inscritos por Distrito



Figura 19-Visualização em mapa-Número de cientistas inscritos por distrito

## 12. Trabalhos Futuros

Com vista à evolução e melhoria contínua do projeto, identificam-se várias oportunidades de desenvolvimento que poderão aumentar a sua relevância e impacto:

- **Dashboard Online e Acessível**

Pretende-se criar um dashboard interativo acessível via web, permitindo que todos os colaboradores consultem os dados em tempo real. Essa dashboard poderá ser implementado em plataformas como Power BI Service ou Google Data Studio, garantindo que os ficheiros sejam atualizados automaticamente após cada execução do processo ETL.

- **Automatização da Execução Diária**

Será fundamental configurar mecanismos de agendamento automático para que o fluxo ETL seja executado diariamente, sem intervenção manual. Garantindo que os dados tratados sejam exportados para os formatos necessários e sincronizados com o dashboard online.

- **Tratamento das Escolas Indicadas pelos Cientistas**

O processo ETL deverá ser ampliado para incluir a normalização e validação das escolas primárias indicadas pelos cientistas nos formulários. Cada escola será associada à designação oficial da organização, evitando duplicidades e erros de escrita.

- **Integração com Indicadores Socioeconómicos**

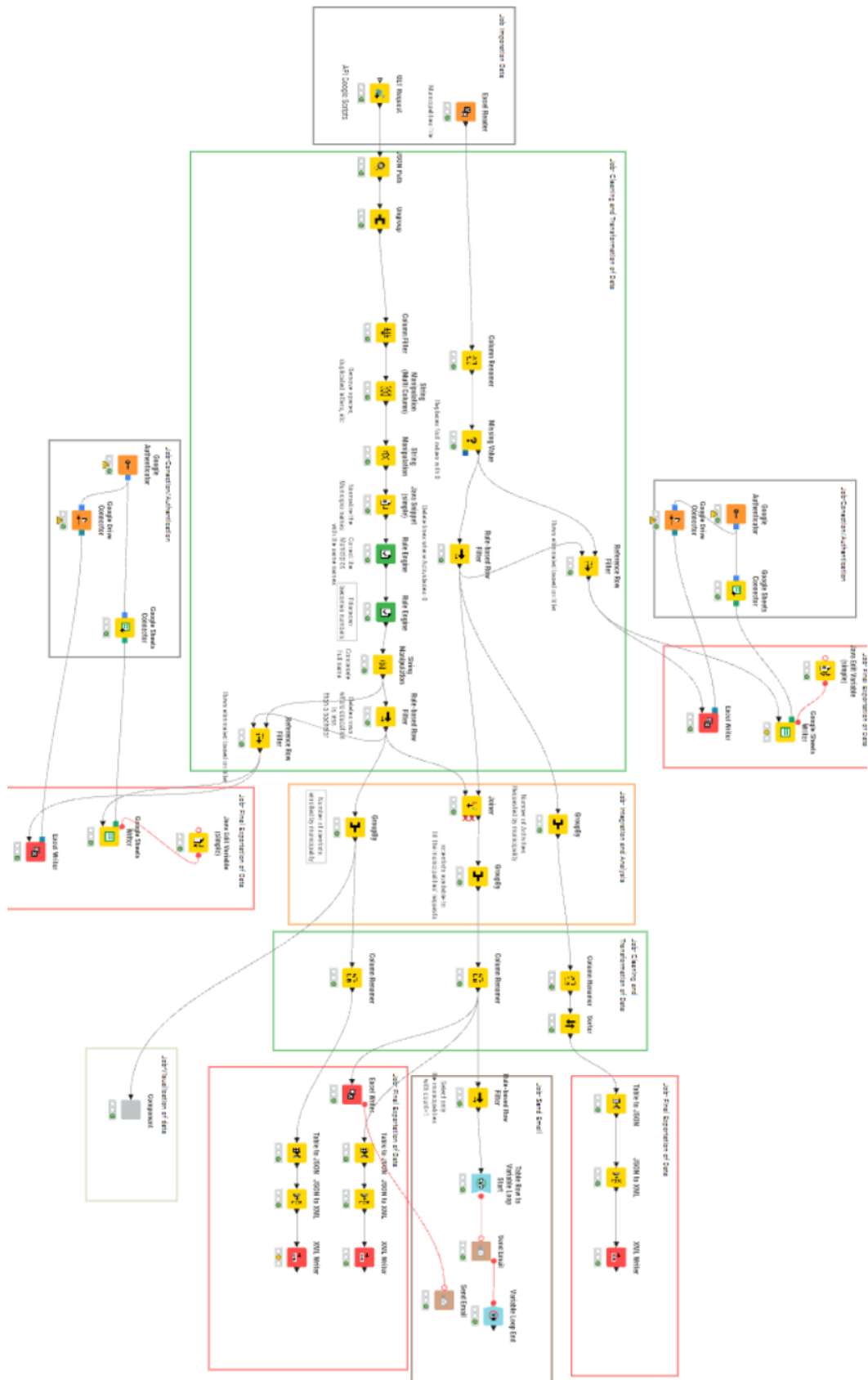
Realizar um cruzamento dos dados tratados com indicadores estatísticos que identifiquem os municípios com maior índice de pobreza e isolamento social. Essa integração permitirá priorizar ações em regiões mais vulneráveis, aumentando o impacto social da iniciativa e contribuindo para uma distribuição mais equitativa das atividades.

## Conclusão

O desenvolvimento deste projeto permitiu criar um processo ETL completo e automatizado, que consolida dados de diferentes fontes, trata inconsistências e cria informação útil para o planeamento das visitas de cientistas às escolas primárias.

Este trabalho demonstra como a aplicação de processos ETL e ferramentas de análise de dados pode melhorar significativamente a organização, comunicação e eficácia de uma iniciativa educativa, garantindo que os recursos disponíveis sejam utilizados de forma eficiente e que os colaboradores recebam informações claras e oportunas.

## Anexo 1- Processos e Jobs no Knime



## Anexo 2- QR Code para visualização do funcionamento dos processos



(link:

[https://drive.google.com/file/d/1XVQUByC5sn8Doj40DioSEzJNu9a8TXo8/view?usp=drive\\_link](https://drive.google.com/file/d/1XVQUByC5sn8Doj40DioSEzJNu9a8TXo8/view?usp=drive_link) )

## Referências Bibliográficas

Microsoft. Power BI Documentation. Microsoft Docs. Disponível em: <https://docs.microsoft.com/power-bi>

Google. Google Apps Script Documentation. Disponível em: <https://developers.google.com/apps-script>

Knime. Knime Documentation. Disponível em: <https://docs.knime.com/>