



# **Relatório ISI**

## **Proposta de Sistema**

**Integração de Dados de Loja Social(ETL com KNIME e Exportação para Dashboard)**

No.20467 – Hugo Brandão

**Licenciatura em Engenharia Sistemas Informáticos**

**3ºano**

Barcelos | Outubro, 2024

## **Lista de Abreviaturas e Siglas**

**ETL**..... Extraction, Transform and Load,

**CSV**..... Comma-separated values,

**JSON**..... JavaScript Object Notation,

**API**..... Application Programming Interface,

## Índice de Figuras

Figura 1- Workflow KNIME.....	8
Figura 2- node CSV Reader .....	8
Figura 3- node String Manipulation(Nome) .....	9
Figura 4- node String Manipulation(Telemovel) .....	9
Figura 5- node Rule Engine .....	10
Figura 6- node Row Filter .....	10
Figura 7- node Table to JSON .....	11
Figura 8- node Group By .....	11
Figura 9- node JSON to XML.....	12
Figura 10. node POST Request JSON .....	12
Figura 11- node POST Request XML.....	13
Figura 12-Dashboard- Beneficiários.....	14
Figura 13- Número de visitas por ano em gráfico.....	14
Figura 14-QRCode do video .....	15

## Índice

1. Introdução .....	5
2. Enquadramento.....	6
3. Descrição do problema .....	6
4. Estratégia Utilizada.....	7
5. Transformações.....	8
6. Jobs.....	13
7. Vídeo com demonstração .....	15
8. Conclusão .....	16
9. Referências Bibliográficas .....	<b>Erro! Marcador não definido.</b>

## 1. Introdução

Através da integração de sistemas de informação, é possível combinar dados provenientes de fontes distintas, melhorar a qualidade da informação e facilitar a análise. Um dos processos fundamentais nesse contexto é o ETL, responsável pela extração, transformação e carregamento de dados para um sistema final ou base de dados.

O ETL é amplamente utilizado em cenários onde é necessário extrair dados de fontes externas (como ficheiros CSV, bases de dados ou APIs) para qualquer sistema necessário na ótica do utilizador. Este processo envolve a coleta de dados de várias fontes, a aplicação de transformações para garantir a consistência e qualidade, e, por fim, o carregamento desses dados em um destino adequado para análise ou armazenamento.

Este relatório descreve as etapas do desenvolvimento do processo ETL, detalha as ferramentas utilizadas, e discute os desafios enfrentados e soluções encontradas. A escolha do KNIME como ferramenta principal deve-se à sua flexibilidade e facilidade de uso na construção de processos de integração de dados. Oferece uma interface visual rica e uma vasta gama de nodes pré-configurados e personalizados.

## 2. Enquadramento

Este relatório é elaborado no contexto da Unidade Curricular de Integração de Sistemas de Informação do curso de Engenharia de Sistemas Informáticos.

O projeto visa a implementação de um processo de ETL com o auxílio da ferramenta KNIME para realizar uma integração de dados com um backend desenvolvido em Node.js e MongoDB (Mongoose) e uma dashboard para visualização dos dados processados.

O trabalho é desenvolvido por Hugo Soares Brandão, aluno com o número a20467.

## 3. Descrição do problema

Atualmente, muitas empresas lidam com uma variedade de dados provenientes de diferentes fontes e formatos. No contexto deste projeto, foi identificado um cenário onde uma Loja Social armazena parte dos seus dados em um ficheiro **CSV** e necessita de integrá-los numa base de dados para análise estatística e processamento. Para isso, surge a necessidade de enviar esses dados, devidamente transformados e estruturados, para uma **API** externa que os processa em tempo real.

Os ficheiros **CSV** são uma forma comum de armazenar dados de sistemas de gestão ou de outros ambientes onde a automação de dados é limitada. No entanto, este formato apresenta alguns desafios, tais como:

- **Qualidade dos dados:** Ficheiros CSV podem conter erros, como caracteres inválidos, dados mal formatados ou inconsistências nos campos.
- **Integração com outros sistemas:** Para que os dados possam ser utilizados por sistemas de gestão e análise de dados, os mesmos precisam ser convertidos para formatos apropriados (por exemplo JSON) e enviados para plataformas que suportam esse tipo de comunicação (ex: APIs REST).

Neste contexto, a possível solução é implementar um processo **ETL** que lide com os seguintes problemas:

1. **Extração dos Dados (Extraction):** É necessário extrair dados de ficheiros CSV de forma eficiente.
2. **Transformação dos Dados (Transform):** Durante a transformação, os dados precisam de ser “limpos”, exigindo a correção de erros de codificação (como caracteres especiais mal interpretados), e transformados nos formato JSON e XML, de modo a garantir que estejam prontos para envio à API.
3. **Carregamento dos Dados (Load):** Para finalizar o processo os dados transformados devem ser enviados para uma **API** externa pelo método **POST**. Além disso, a solução deve prever a possibilidade de armazenar os dados numa base de dados.

## 4. Estratégia Utilizada

Para a implementação deste projeto ETL, foi utilizado, da ferramenta KNIME, vários nodes para garantir a correta manipulação e validação dos dados. A estratégia inclui:

- **Leitura de Dados:** Leitura do ficheiro CSV com dados brutos.
- **Manipulação e Limpeza de Dados:** Uso de operadores de manipulação para garantir a padronização dos dados, incluindo a transformação de valores e a remoção de dados inválidos.
- **Validação:** Verificação de dados críticos como número de telemóvel e totalização de visitas.
- **Transformação:** Conversão de dados para JSON e XML antes de serem enviados para o backend.
- **Envio para o Backend:** Utilização de requisições POST para envio dos dados para um sistema backend desenvolvido em Node.js com MongoDB.
- **Dashboard:** Como objetivo final, foi desenvolvido uma dashboard para visualizarmos os dados processados, de modo a simular uma experiência real no contexto empresarial.

## 5. Transformações

### Diagramas

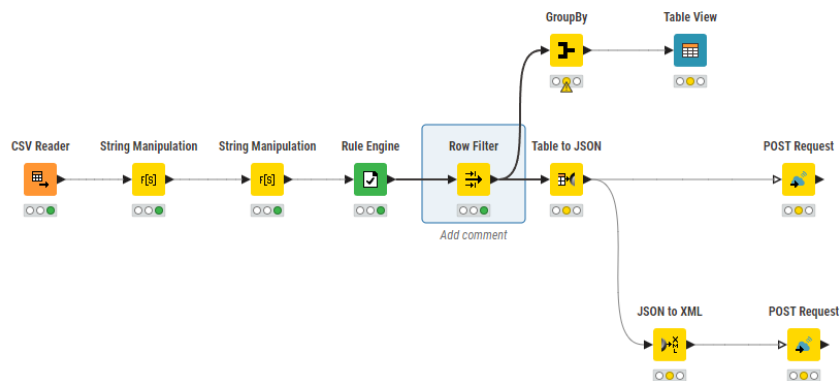


Figura 1- Workflow KNIME

### Explicação

1. **CSV Reader:** Node inicial que lê o ficheiro CSV com os dados a serem processados.

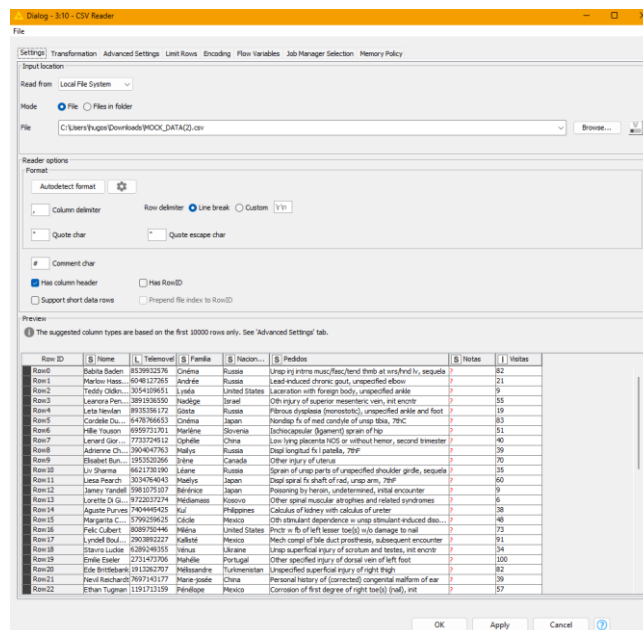


Figura 2- node CSV Reader



2. **String Manipulation (Nome):** Manipula o campo "Nome" para remover dados que não contenham apenas letras maiúsculas e minúsculas, garantindo consistência.

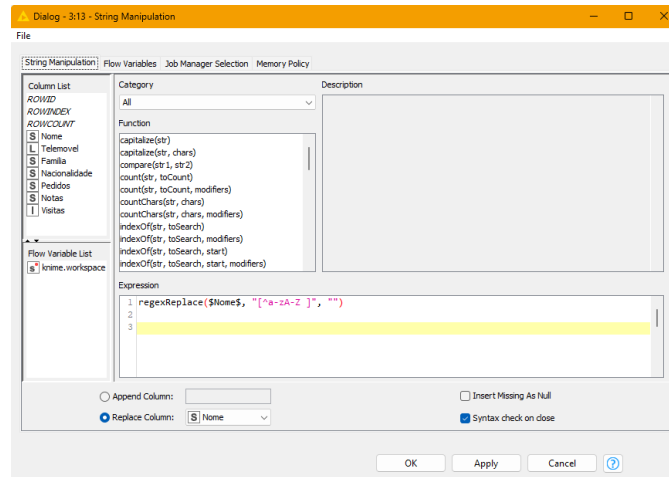


Figura 3- node String Manipulation(Nome)

3. **String Manipulation (Telemóvel):** Converte o campo "Telemóvel" para string, permitindo a configuração de expressões regulares de forma correta no node seguinte.

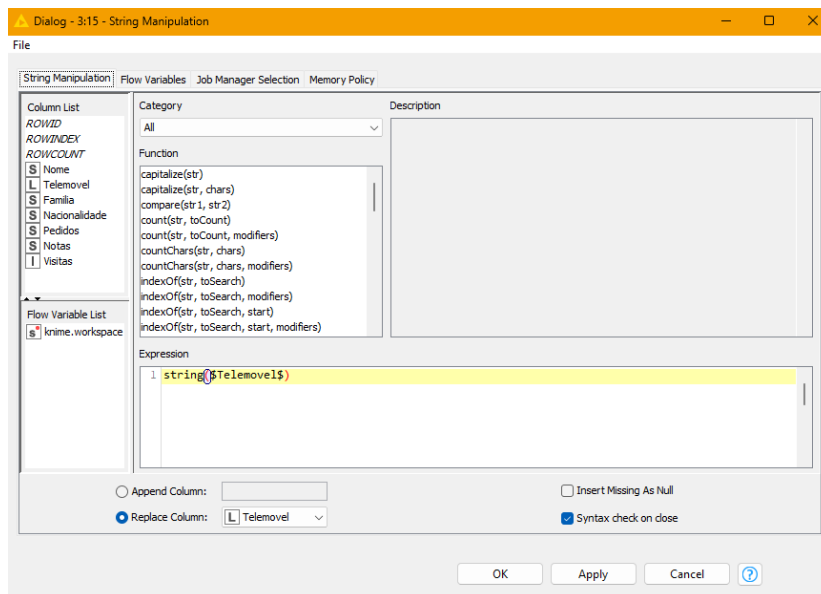


Figura 4- node String Manipulation(Telemovel)

4. **Rule Engine (Telemóvel):** Verifica se o número de telemóvel contém 9 dígitos. Se o número não for válido, o valor do campo é substituído por "Inválido".

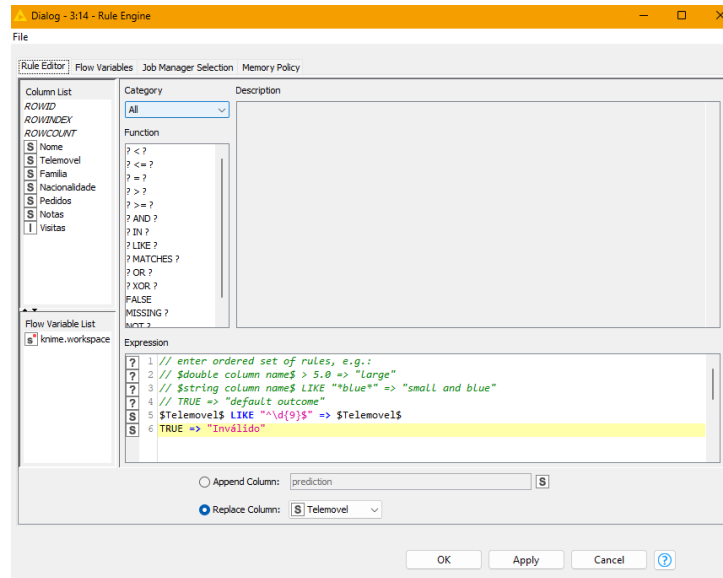


Figura 5- node Rule Engine

5. **Row Filter:** Filtra as linhas, mantendo apenas as que possuem um número de telemóvel válido.

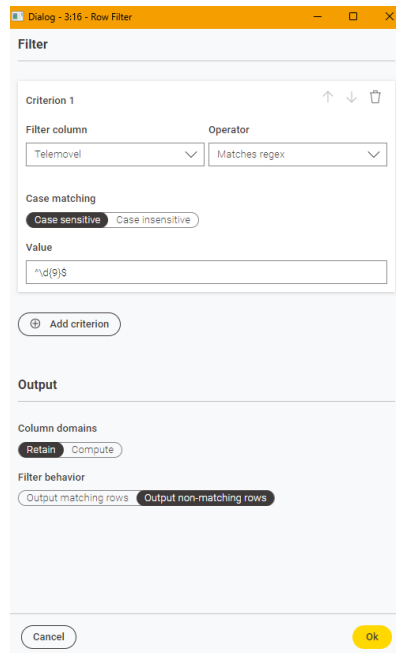


Figura 6- node Row Filter

6. **Table to JSON:** Converte os dados para o formato JSON, permitindo o envio no formato correto para o backend.

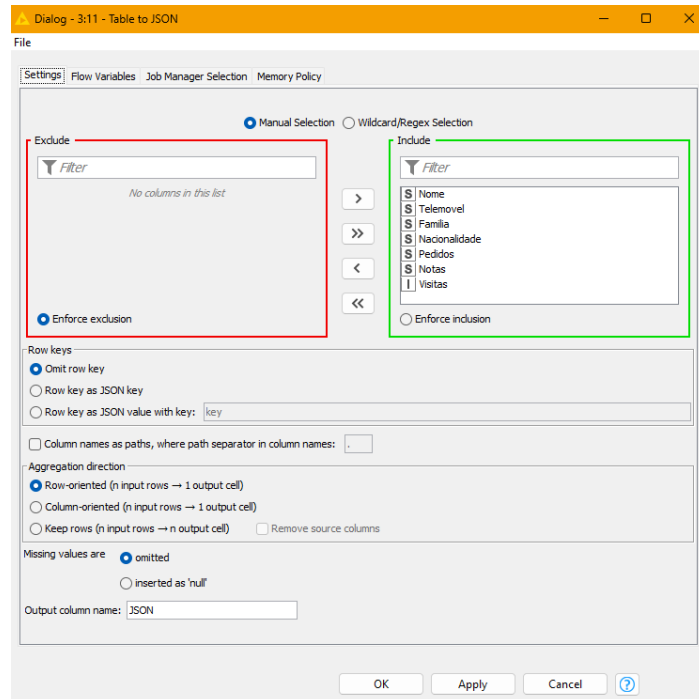


Figura 7- node Table to JSON

7. **Group By (Visitas):** Realiza a soma dos valores do campo "Visitas" para obter o total de visitas.

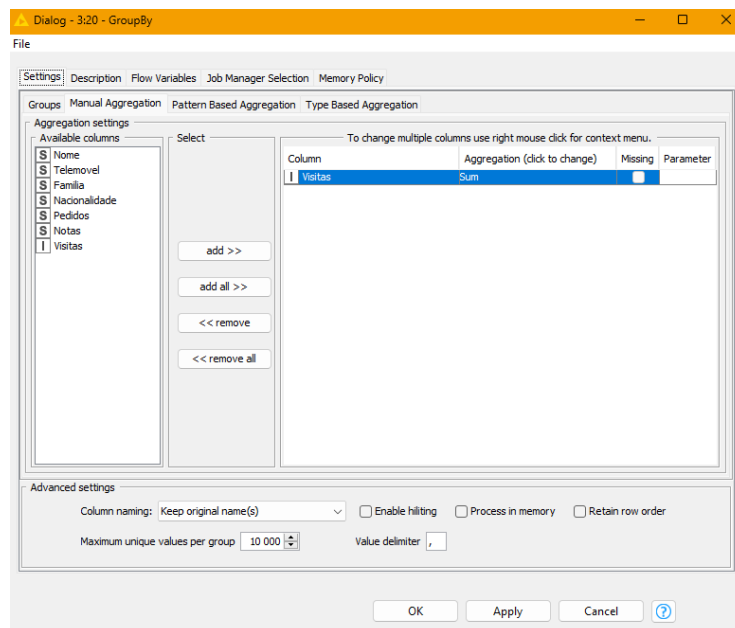


Figura 8- node Group By

8. **JSON to XML:** Converte os dados JSON em XML, possibilitando o envio de dados também neste formato para o backend.

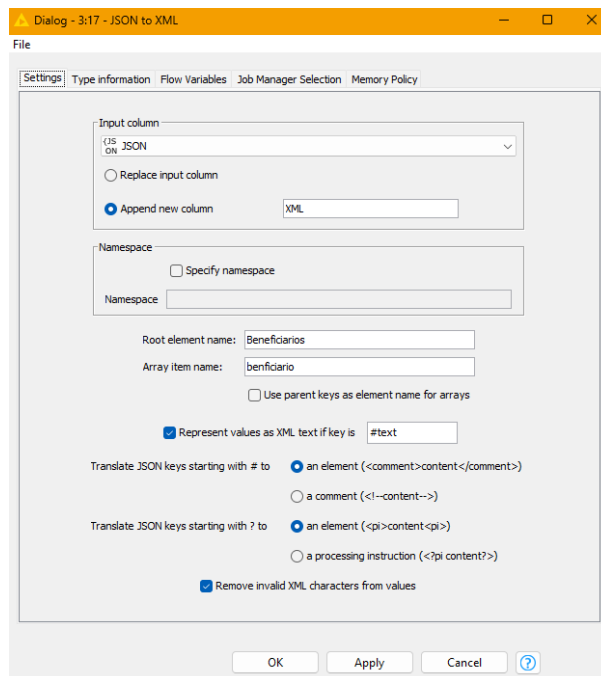


Figura 9- node JSON to XML

9. **Post Request:** Node para enviar os dados JSON e XML processados ao backend.

Para o JSON

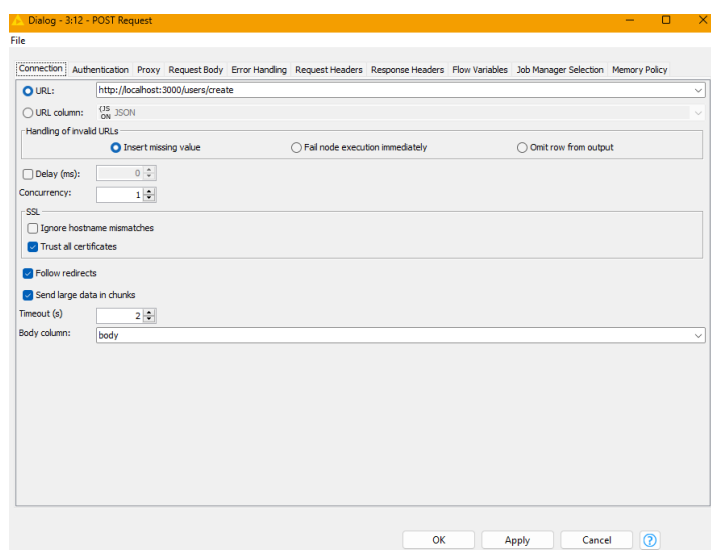


Figura 10. node POST Request JSON

### Para XML

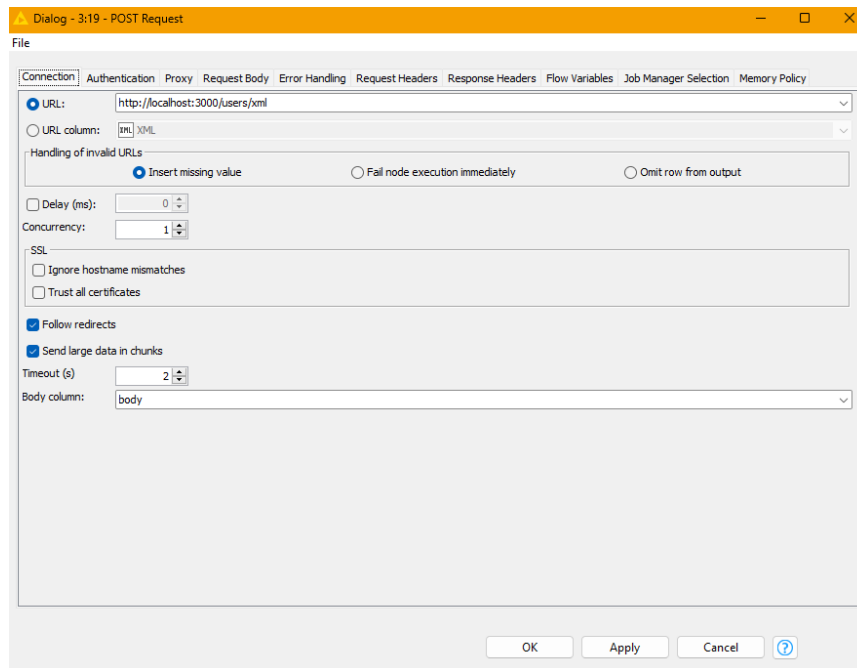


Figura 11- node POST Request XML

## 6. Jobs

Os principais jobs e processos configurados no KNIME foram:

1. **Processamento e Validação do Telemóvel:** Garantir que o campo "Telemóvel" seja válido antes de prosseguir, substituindo valores inválidos por "Inválido".
2. **Transformação de Dados para Backend:** Converte os dados para JSON e XML, permitindo que o backend possa armazená-los independentemente do formato.
3. **Agrupamento e Totalização de Visitas:** Processo para somar os valores do campo "Visitas", facilitando a análise do total de visitas no backend.
4. **Envio de Dados para Backend:** Através de requisições POST, envia os dados transformados para um backend em Node.js, que os armazena numa base de dados MongoDB.

5. **Visualização dos dados na dashboard:** Por fim, é possível visualizarmos os dados de cada utilizador, após o processo ETL, na dashboard:

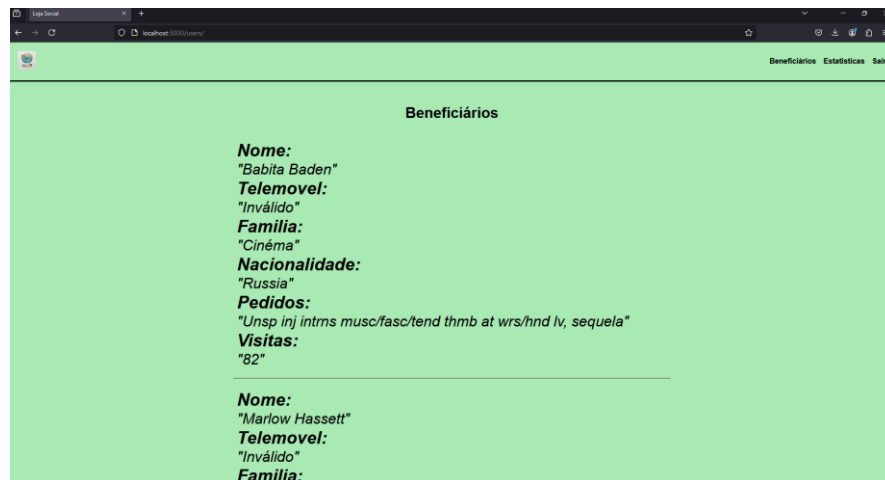


Figura 12-Dashboard- Beneficiários

6. **Estatísticas:** É possível visualizar as visitas do ano 2023 (Valor constante apenas para simulação do gráfico) e as visitas do ano 2024 contabilizados:

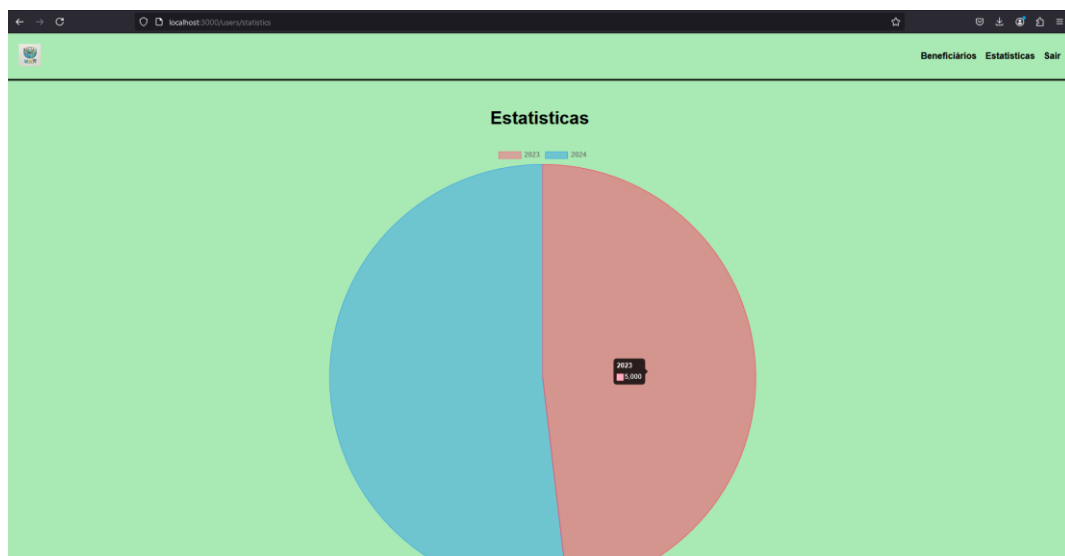


Figura 13- Número de visitas por ano em gráfico

## 7. Vídeo com demonstração

Foi criado, um vídeo demonstrativo, para apresentar de forma visual e prática o funcionamento do pipeline ETL desenvolvido com KNIME. O vídeo ilustra cada etapa do processo, desde a leitura dos dados do ficheiro CSV até à sua transformação, do envio para o backend em formatos JSON e XML até à sua visualização dos dados na dashboard.

Para aceder ao vídeo, é possível fazer um “scan” ao QRCode:



*Figura 14-QRCode do vídeo*

Caso não seja possível aceder ao vídeo pelo QRCode, temo como opção aceder ao link:

<https://www.youtube.com/watch?v=0bxE5LUYnq4>

## 8. Conclusão

Concluo com este projeto de ETL que exemplifica uma abordagem robusta para a integração de dados, desde a extração e transformação até ao carregamento em um sistema de armazenamento. Através de uma série de processos de manipulação e validação de dados, foi possível garantir que os dados tratados cumprissem os requisitos de qualidade e consistência. O uso de KNIME para implementar filtros, validações e conversões de formatos, assim como o backend construído em Node.js com MongoDB, proporcionou uma solução completa e escalável para o desafio de integração de sistemas.

A principal realização deste projeto foi o desenvolvimento de uma pipeline automatizada que não só valida informações, como também se adapta a diferentes formatos (JSON e XML) para uma integração mais flexível com o backend. O processo de transformação de dados permitiu melhorar a precisão dos dados, eliminando inconsistências e criando um conjunto de dados fiável e padronizado, pronto para análise ou para qualquer outra aplicação.



## 1. Bibliografia

*KNIME*. (2024). Obtido de KNIME: <https://www.knime.com/getting-started-guide>

*KNIME Hub*. (2024). Obtido de KNIME:

<https://hub.knime.com/knime/collections/Starter%20collection~Y0m-e65xNWBFcR20>

*MongoDB*. (2024). Obtido de <https://www.mongodb.com/docs/atlas/>

*NodeJS*. (2024). Obtido de <https://nodejs.org/docs/latest/api/>

*Stackoverflow*. (2024). Obtido de <https://stackoverflow.com/questions/tagged/knime>