

# Trabalho prático 01 - Integração de Sistemas de Informação

Tomás Ferreira Nºa20457

Outubro 2024

## Contents

<b>1</b>	<b>Introdução</b>	<b>3</b>
<b>2</b>	<b>Problema</b>	<b>4</b>
<b>3</b>	<b>Estratégia utilizada</b>	<b>5</b>
<b>4</b>	<b>Transformações</b>	<b>6</b>
4.1	Diagrama principal . . . . .	6
<b>5</b>	<b>Transformações - Problema 1</b>	<b>7</b>
<b>6</b>	<b>Transformações - Problema 2</b>	<b>8</b>
<b>7</b>	<b>Tranformações - Problema 3</b>	<b>10</b>
<b>8</b>	<b>Transformações - Ficheiros Excel e XML</b>	<b>11</b>
<b>9</b>	<b>Conclusão</b>	<b>12</b>
<b>10</b>	<b>Bibliografia</b>	<b>13</b>

# **1 Introdução**

Este projeto foi realizado no âmbito da disciplina Integração de Sistemas de Informação, o mesmo apresenta uma visão geral da análise de dados dos Pokémon realizada no KNIME. O projeto focou na classificação e filtragem dos Pokémon com base em características como geração e tipo. O objetivo é descrever as etapas do processo e os métodos aplicados, oferecendo uma análise clara e detalhada dos resultados obtidos.

## 2 Problema

O principal objetivo deste projeto é a demonstração de como podemos realizar uma ETL para transformar diversos tipos de dados. Para isso, utilizamos a plataforma **KNIME**. Foi desenvolvida uma ETL com o tema “Pokémon” com os seguintes objetivos principais:

- **Problema 1:** Identificar o número de Pokémon por cada geração.
- **Problema 2:** Contabilizar o número total de Pokémon por tipo (Fogo, Água, etc.).
- **Problema 3:** Calcular a média de cada estatística dos Pokémon, incluindo HP, Ataque, Defesa, Ataque Especial, Defesa Especial e Velocidade.

### 3 Estratégia utilizada

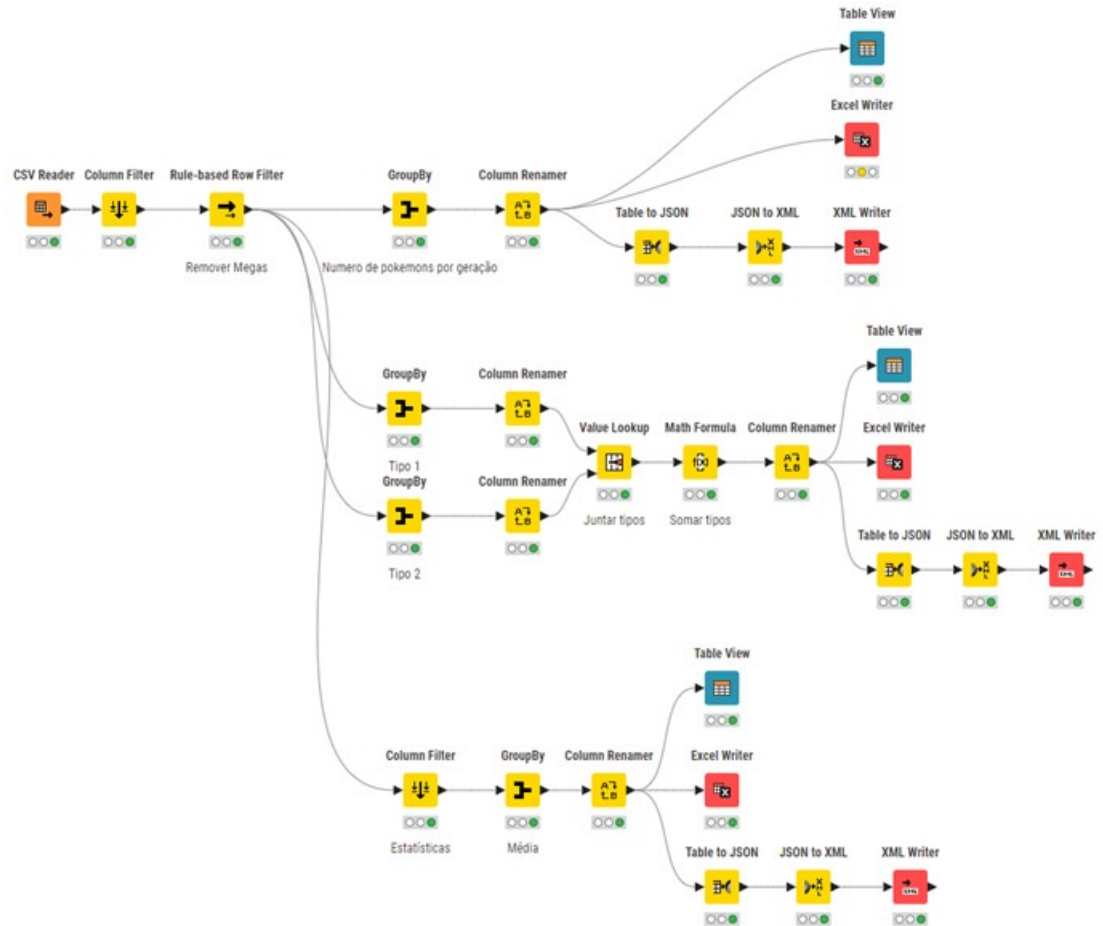
A obtenção de dados é obtida a partir de um ficheiro .csv. Utilizando o KNIME, iremos utilizar processos para manipular esses dados, tais como:

- Filtros de colunas.
- Filtros de linhas com regras.
- Operações matemáticas.
- Agrupamentos.

## 4 Transformações

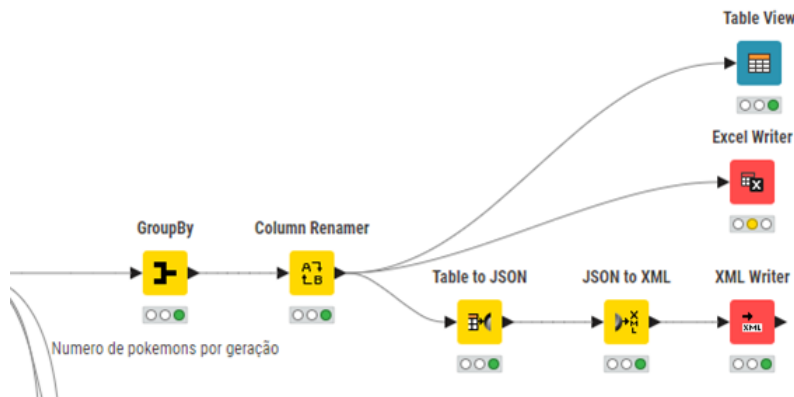
#### 4.1 Diagrama principal

Este diagrama contém o esquema que foi utilizado para resolver todos os problemas.

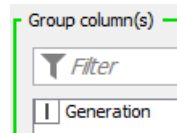


## 5 Transformações - Problema 1

A resolução para obter o número de Pokémon por cada geração foi a seguinte.



Usando um GroupBy, agrupando as gerações e fazendo uma contagem do número de Pokémon. O # representa o ID do Pokémon.



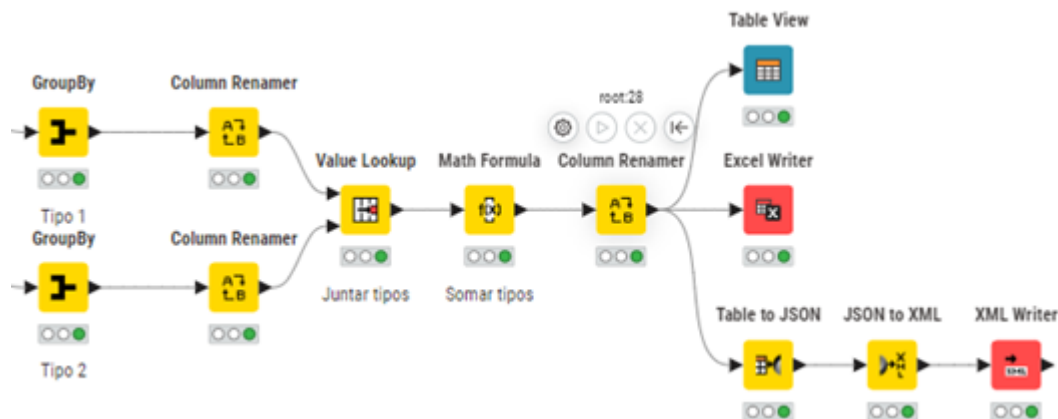
Column	Aggregation (click to change)	Missing	Parameter
#	Count	<input type="checkbox"/>	

Ficando este o resultado:

Generation <i>Number (integer)</i>	Count*(#) <i>Number (integer)</i>
1	151
2	99
3	140
4	116
5	164
6	81

## 6 Transformações - Problema 2

A resolução para obter o número total de Pokémon por tipo foi a seguinte.



Nesta resolução foi necessária ir por dois caminhos com o GroupBy, pois cada Pokémon pode ter dois tipos. Sendo assim, foi usado um GroupBy para contar cada tipo no Tipo 1 e outro GroupBy para cada tipo no Tipo 2. Com isto, usando o value Lookup foi criado esta tabela.

Type 1 <i>String</i>	Type 1 Count <i>Number (integer)</i>	Type 2 <i>String</i>	Type 2 Count <i>Number (integer)</i>
Bug	65	Bug	3
Dark	28	Dark	17
Dragon	26	Dragon	15
Electric	42	Electric	6
Fairy	17	Fairy	18
Fighting	25	Fighting	21
Fire	48	Fire	11
Flying	4	Flying	91
Ghost	30	Ghost	13
Grass	66	Grass	25
Ground	32	Ground	31



O que queremos agora é juntar isto tudo em uma coluna “total”. Usamos assim uma Math Formula.

Expression

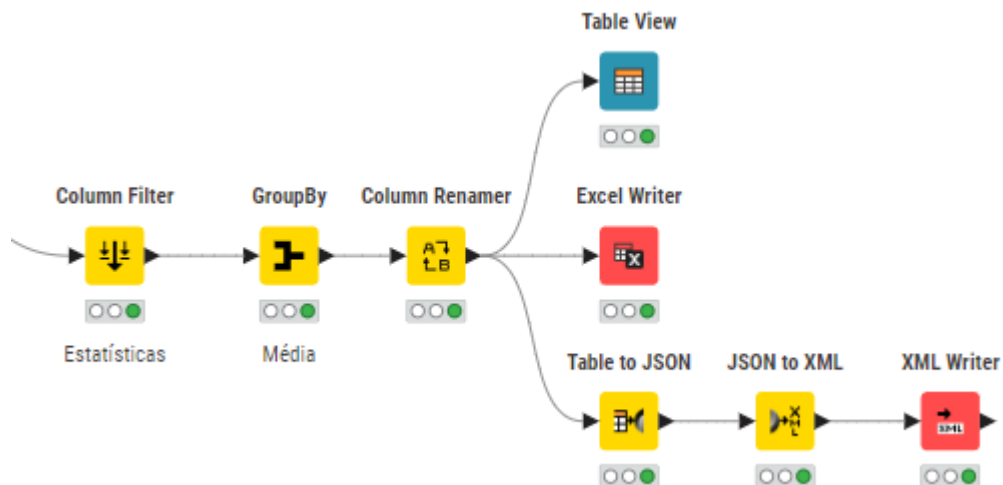
1 \$Type 1 Count\$ + \$Type 2 Count\$

Com isto a tabela final ficou assim, adicionando uma nova coluna com o total de cada tipo.

Type 1 <small>String</small>	Type 1 Count <small>Number (Integer)</small>	Type 2 <small>String</small>	Type 2 Count <small>Number (Integer)</small>	TotalTipos <small>Number (double)</small>
Bug	65	Bug	3	68
Dark	28	Dark	17	45
Dragon	26	Dragon	15	41
Electric	42	Electric	6	48
Fairy	17	Fairy	18	35
Fighting	25	Fighting	21	46
Fire	48	Fire	11	59
Flying	4	Flying	91	95
Ghost	30	Ghost	13	43
Grass	66	Grass	25	91
Ground	32	Ground	31	63
...	...	...	...	...

## 7 Transformações - Problema 3

A resolução para a média de cada estatística dos Pokémon foi a seguinte.



Começamos com um Column Filter para mostrar apenas as estatísticas. Depois é usado um GroupBy, este GroupBy não agrupa nada em específico, mas faz antes a agregação de cada coluna, fazendo assim a média.

Column	Aggregation (click to change)
HP	Mean
Attack	Mean
Defense	Mean
Sp. Atk	Mean
Sp. Def	Mean
Speed	Mean

Ficando este o resultado.

Média HP Number (double)	Média Ataque Number (double)	Média Defesa Number (double)	Média Sp. Ataque Number (double)	Média Sp. Defesa Number (double)	Média Velocidade Number (double)
68.67	75.979	71.744	70.213	70.109	66.607

## 8 Transformações - Ficheiros Excel e XML

Usando Excel Writers e XML Writers no KNIME, foram criados ficheiros Excel e XML para cada problema.

```
▼<root>
  ▼<item>
    <Generation>1</Generation>
    <Total>151</Total>
  </item>
  ▼<item>
    <Generation>2</Generation>
    <Total>99</Total>
  </item>
  ▼<item>
    <Generation>3</Generation>
    <Total>140</Total>
  </item>
  ▼<item>
    <Generation>4</Generation>
    <Total>116</Total>
  </item>
  ▼<item>
    <Generation>5</Generation>
    <Total>164</Total>
  </item>
  ▼<item>
    <Generation>6</Generation>
    <Total>81</Total>
  </item>
</root>
```

	A	B
1	Generation	Total
2	1	151
3	2	99
4	3	140
5	4	116
6	5	164
7	6	81
8		

## 9 Conclusão

Este projeto proporcionou uma oportunidade para aprofundar conhecimentos sobre a integração e manipulação de dados utilizando a plataforma KNIME. Ao longo do trabalho, foi possível realizar uma análise detalhada dos dados dos Pokémon, aplicando técnicas de ETL para transformar informações brutas.

Além disso, a criação de arquivos em formato Excel e XML evidenciou a versatilidade do KNIME na exportação de dados, permitindo a integração com outras aplicações e sistemas. Este trabalho não só fortaleceu as competências técnicas na utilização de ferramentas de análise, mas também destacou a importância da organização e interpretação dos dados para a tomada de decisões informadas.

Resumindo, este projeto foi um passo significativo nas competências de manipulação de dados e criação de arquivos.

## 10 Bibliografia

Ficheiro de dados utilizado: <https://gist.github.com/armgilles/194bcff35001e7eb53a2a8b441e8b2c6>