



**INSTITUTO POLITÉCNICO
DO CÁVADO E DO AVE
ESCOLA SUPERIOR
DE TECNOLOGIA**

PROJETO I

Integração de Sistemas Informáticos

Realizado por:

Hugo Especial – 27963

Marco Cardoso – 27969

Conteúdo

Índice de imagens	2
Siglas	3
1. Introdução	4
1.2. Plataformas utilizadas	4
2. Desenvolvimento	5
2.1. Procedimento	5
2.1.1. Escolha do <i>dataset</i>	5
2.1.2. Conversão de ficheiros	6
2.1.3. Processo de preparação	7
2.1.4. Processo de ETL.....	9
2.1.5. Expressões regulares.....	11
2.1.6. Comparação do estado inicial com o final	12
2.2. Visualização de gráficos.....	13
2.3. Inclusão de funcionalidades extra	21
2.4. Vídeos de demonstração do fluxo	26
3. Conclusão	28
4. WebGrafia	29

Índice de imagens

Figura 1 - Conversão CSV para XML	6
Figura 2 - Fluxo de preparação dos Clientes	7
Figura 3 - Fluxo de preparação dos Livros	7
Figura 4 - Fluxo de preparação dos Vendas	7
Figura 5 - Preparação dos Dados	8
Figura 6 - Processo de ETL.....	10
Figura 7 - Estado inicial do processo ETL – KNIME	12
Figura 8 - Estado final do processo ETL – KNIME	12
Figura 9 - Estado inicial do processo ETL – PENTAHO	12
Figura 10 - Estado final do processo ETL – PENTAHO.....	13
Figura 11 - Fluxo para guardar o número de clientes por país	13
Figura 12 - Fluxo para guardar o número de livros lançados por ano	14
Figura 13 - Fluxo para guardar no número de vendas por ano	14
Figura 14 - Gráfico número de clientes por país	15
Figura 15 - Gráfico número de livros lançados por ano	15
Figura 16 - Gráfico número de vendas por ano	16
Figura 17 - Stream Lookup.....	17
Figura 18 - Fluxo de Vendas por Género.....	17
Figura 19 - Pedidos Totais por Género	18
Figura 20 - Stream Lookup de VendasPorPaíses	19
Figura 21 - Fluxo de Pedidos Totais por País	20
Figura 22 - Top 15 Países com mais pedidos	20
Figura 23 - Uso do nó "Send Email"	21
Figura 24 - Campos utilizados	21
Figura 25 - Fluxo de Enviar Email.....	22
Figura 26 - Dados de constantes Email	22
Figura 27 - Construção do corpo do email	22
Figura 28 - Ajuda Construção do email.....	23
Figura 29 - Envio Email I	23
Figura 30 - Envio Email II	24
Figura 31 - Envio Email III	24
Figura 32 - Job KNIME	25
Figura 33 - Job Pentaho	25

Siglas

- ETL – Extract, Transform, Load
- XML – eXtensible Markup Language
- JSON – JavaScript Object Notation
- CSV – Comma-Separated Values
- HTML – HyperText Markup Language
- CSS – Cascading Style Sheets
- JS – JavaScript

1. Introdução

Este projeto tem como objetivo aplicar boas práticas de ETL para realizar processos de extração, transformação e carregamento de dados. Pretende-se demonstrar a capacidade de integrar, limpar e normalizar informações provenientes de diferentes tipos de ficheiros, assegurando a sua consistência e qualidade.

Ao longo do projeto são realizadas operações como a validação de dados, deteção e remoção de valores ausentes ou inconsistentes, eliminação de duplicados e criação de *logs* automáticos de registos apagados. O trabalho reflete a importância da integração e tratamento de dados no contexto da unidade curricular de Integração de Sistemas de Informação, destacando o papel essencial dos processos de ETL na gestão eficiente de sistemas de informação modernos.

1.2. Plataformas utilizadas

Para realizar este tipo de método, é essencial recorrer a plataformas que permitam a criação, automação e gestão eficiente de processos de ETL. Estas ferramentas desempenham um papel fundamental na integração de dados provenientes de diferentes fontes, oferecendo um ambiente visual e interativo para o desenvolvimento de fluxos de trabalho.

Pentaho Kettle: A ferramenta oferece um ambiente gráfico (*Spoon*) que facilita a criação, teste e monitorização de fluxos de dados, permitindo integrar informações provenientes de diversas fontes — como bases de dados, ficheiros planos, *API's*, entre outros.

Knime: é uma plataforma open-source orientada para análise e integração de dados. Permite construir fluxos de trabalho de forma visual, através da ligação de nós que representam operações de extração, transformação e carregamento. A interface é intuitiva no qual facilita a manipulação, validação e limpeza de dados, tornando-o uma ferramenta eficaz tanto para tarefas simples de tratamento de informação como para processos de integração mais complexos.

2. Desenvolvimento

Nesta parte do relatório é descrito todo o procedimento seguido, desde a seleção do *dataset* inicial até à preparação final dos dados para exportação. O processo incluiu a conversão de ficheiros entre diferentes formatos, como XML e JSON, a extração e estruturação dos dados, bem como diversas etapas de verificação e limpeza, de modo a garantir a integridade e a consistência das informações. Foram também aplicadas técnicas de normalização e validação, assegurando que todos os registos cumprem o formato e as regras definidas antes de serem exportados para utilização posterior.

2.1. Procedimento

O procedimento iniciou-se com a procura e seleção do *dataset* mais adequado para os objetivos do projeto. Após a obtenção dos ficheiros, foi realizada a conversão para formatos compatíveis, permitindo que os dados fossem carregados e manipulados no ambiente de trabalho. Em seguida, foram efetuadas etapas de verificação, limpeza e normalização, garantindo que os dados estivessem consistentes e prontos para análise e posterior exportação.

2.1.1. Escolha do *dataset*

A escolha do *dataset* foi um passo fundamental, pois permitiu definir claramente quais dados seriam tratados e analisados ao longo do projeto. Optou-se por trabalhar com três ficheiros CSV, correspondendo a tabelas distintas: uma com informações sobre clientes, outra com detalhes dos livros e uma terceira com os registos de vendas, associando livros a clientes. Esta estrutura permitiu planejar desde a limpeza até à normalização dos dados, garantindo que cada tabela pudesse ser integrada e cruzada de forma eficiente durante as análises subsequentes.

2.1.2. Conversão de ficheiros

Para cumprir os requisitos do enunciado, é necessário trabalhar com diferentes tipos de ficheiro. Como o *dataset* tem incluído 3 ficheiro CSV, então decidimos converter 2 ficheiros e XML, assim ficou as conversões:

- **Clientes.csv**: conversão para XML.
- **Livros.csv**: conversão para XML.
- **Vendas.csv**: mantido o CSV no KNIME e transformando em XML no PENTAHO.

Este foi o fluxo utilizado no Knime para permitir a conversão:

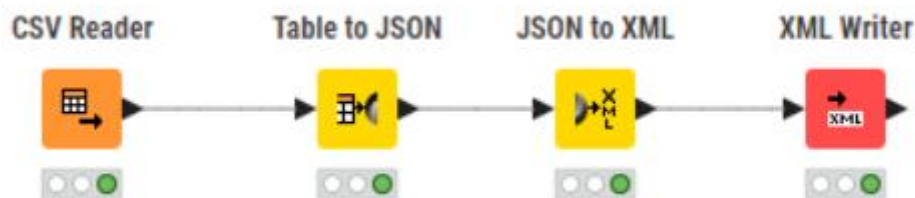


Figura 1 - Conversão CSV para XML

O nó “Table To Json” adiciona cada coluna do csv para um formato JSON. O nó “JSON to XML” transforma o conteúdo JSON para que seja legível escrever em XML.

Por último, o nó “XML Writer” escreve a transformação toda num ficheiro XML.

No **PENTAHO** a realização da conversão foi realizada ao mesmo momento que o processo de preparação.

2.1.3. Processo de preparação

Para utilizar métodos de verificação, normalização e remoção, é necessário trabalhar com dados que possam apresentar inconsistências ou erros. O *dataset* escolhido continha toda a informação bem estruturada e corretamente formatada, o que não permite testar completamente estas técnicas. Por isso, procedeu-se a uma “desformatação” controlada, introduzindo intencionalmente pequenas alterações nos dados, como valores incorretos, campos em formatos diferentes ou dados incompletos.

Fluxo de trabalho para preparar os dados no Knime:

- **Clientes.xml:** Desordenar linhas => Adicionar *strings* vazias => Atribuir *email's* errados.
- **Livros.xml:** Desordenar linhas => Adicionar *strings* vazias => Atribuir valores negativos para os preços.
- **Vendas.csv:** Desordenar linhas => Colocar valores negativos no preço total => Desorganizar a data de venda => Colocar quantidade vendida a 0 (zero) => Atribuir alguns *id's* como 0 (zero).

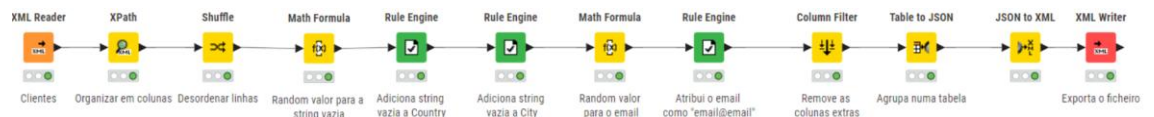


Figura 2 - Fluxo de preparação dos Clientes

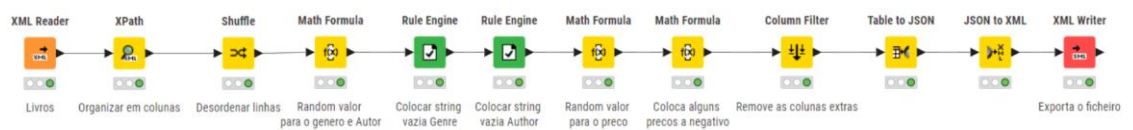


Figura 3 - Fluxo de preparação dos Livros

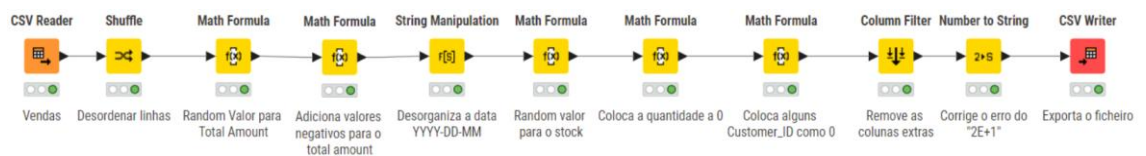


Figura 4 - Fluxo de preparação dos Vendas

Fluxo de trabalho para preparar os dados no PENTAHO:

- **Clientes.xml:** Desordenar linhas => Adicionar *strings* vazias => Atribuir *email's* errados.
- **Livros.xml:** Desordenar linhas => Adicionar *títulos* nulos
- **Vendas.csv:** Desordenar linhas => Colocar quantidade vendida a 0.

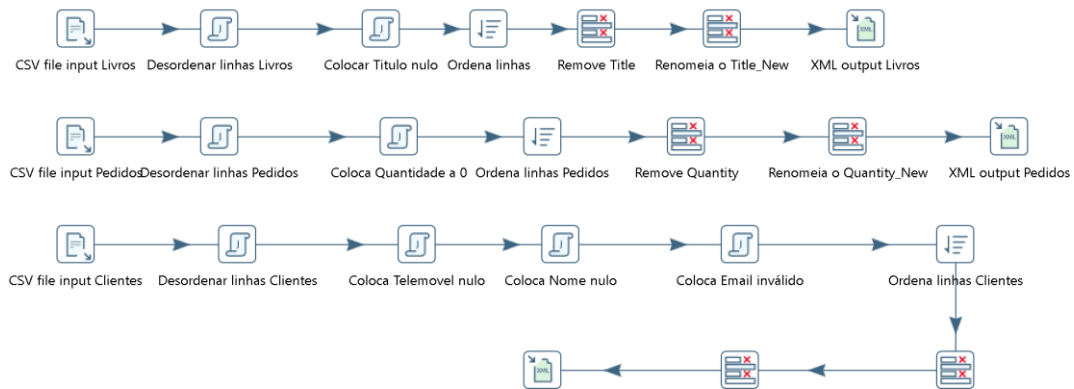


Figura 5 - Preparação dos Dados

2.1.4. Processo de ETL

O processo utilizado nas duas plataformas foi bastante semelhante, uma vez que o tema do trabalho é o mesmo e, por esse motivo, optámos por não introduzir diferenças significativas entre as abordagens.

KNIME: O processo foi centralizado num único *workflow*, de forma a facilitar a organização e a execução das tarefas. Este *workflow* foi utilizado essencialmente para realizar as transformações necessárias nas três tabelas associadas ao tema do projeto.

- **Tabela – Clientes.xml:** Eliminar *id's*/email's/telemóveis repetidos => Eliminar registos vazios ou colunas não atribuídas => Separar primeiro nome e apelido => Validar email => Validar telemóvel => Eliminar *email's* e telemóveis inválidos => Formatar o aspeto do telemóvel => Ordenar as linhas por *id* => Organizar e renomear as colunas => Guardar dados no formato JSON.
- **Tabela – Livros.xml:** Eliminar *id's* repetidos => Eliminar registos vazios ou colunas não atribuídas => Verificar se o preço é negativo, se for elimina => Arredondar o preço com uma casa decimal => Organizar as linhas por *id* => Guardar os dados no formato JSON.
- **Tabela – Vendas.csv:** Eliminar *id's* repetidos => Eliminar registos vazios ou colunas não atribuídas => Comparar o *id* da tabela Livros.xml com o *id* dos livros na tabela Vendas.csv => Comparar o *id* da tabela Clientes.xml com o *id* dos clientes na tabela Vendas.csv => Verificar se o preço final é negativo, se for elimina => Formatar as datas de venda => Organizar as linhas por *id* => Renomear as colunas => Guardar os dados no formato JSON.

PENTAHO: O processo foi centralizado numa única transformação, de forma a facilitar a organização e a execução das tarefas. Nesta transformação foi utilizado essencialmente para realizar as transformações necessárias nas três tabelas associadas ao tema do projeto.

- **Tabela – Clientes.xml:** Ordenar as linhas por *id* => Elimina espaços => Elimina *id*'s repetidos => Elimina *emails* repetidos => Validar *email*, *phone* e *name* => Guardar dados no formato XML.
- **Tabela – Pedidos.xml:** Ordenar as linhas por *id* => Elimina espaços => Elimina *id*'s repetidos => Verifica a quantidade => Guardar dados no formato XML.
- **Tabela – Livros.xml:** Ordenar as linhas por *id* => Elimina espaços => Elimina *id*'s repetidos => Verifica o título => Guardar dados no formato XML.

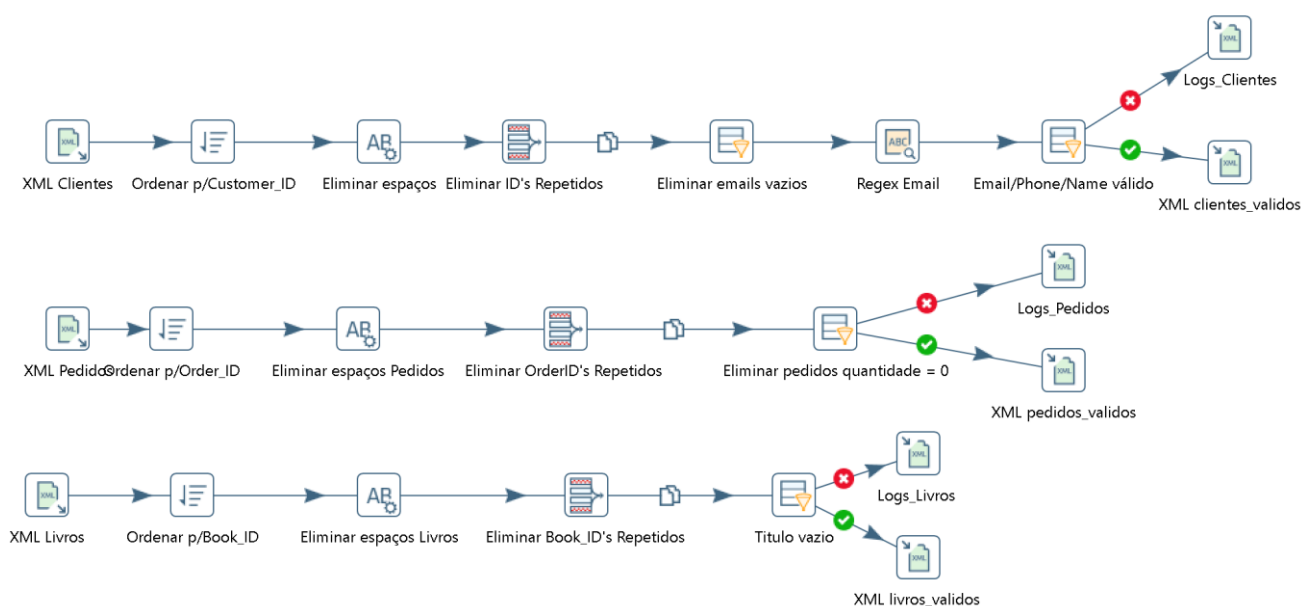


Figura 6 - Processo de ETL

2.1.5. Expressões regulares

Nesta parte do relatório vão ser explicadas as expressões regulares utilizadas no processo de ETL.

Email: `^[A-Za-z0-9._%+~]+@[A-Za-z0-9.-]+\.[A-Za-z]{2,}$`

- `^`: Indica o começo da string a ser analisada
- `[A-Za-z0-9._%+~]+` : Permite escrever texto antes do símbolo `@`.
- `@`: Verifica se o email tem o `@`.
- `[A-Za-z0-9.-]+` : Permite escrever texto, costuma ser o do domínio utilizado.
- `\.` : Permite escrever o ponto para o email.
- `[A-Za-z]{2,}` : Permite escrever a extensão do email, tem de ter no mínimo 2 caracteres.
- `$` : Indica o fim da string.

Telemóvel: `^[0-9]{6,15}$`

- `^`: Indica o começo da string a ser analisada.
- `[0-9]`: Indica qualquer dígito entre 0 a 9.
- `{6,15}` : Indica o número de vezes que o padrão anterior se pode repetir, neste caso entre 6 a 15 dígitos, para abranger qualquer número de telemóvel.
- `$`: Indica o fim da string.

2.1.6. Comparação do estado inicial com o final

Rows: 500 | Columns: 7

#	RowID	XML	Id	Nome	Email	Telemovel	Cidade	Pais
1	Row0_	<?xml version="1	174	Jonathon Strickland	ryan10@yahoo.com	1234568064	Bakerton	Dominica
2	Row0_	<?xml version="1	355	Manuel Guzman	dperez@hotmail.com	1234568245	Orrview	
3	Row0_	<?xml version="1	122	Colleen Navarro	darlenelawson@char	1234568012		
4	Row0_	<?xml version="1	280	Robert Perkins	drangel@gmail.com	1234568170	New Ronaldbury	
5	Row0_	<?xml version="1	83	Aaron Clark	jessica98@hotmail.c	1234567973	Thomashaven	
6	Row0_	<?xml version="1	218	Craig Berger	stacymccall@gmail.c	1234568108		
7	Row0_	<?xml version="1	404	Lisa Duran	christina28@espinoz	1234568294	Lake Tyler	Korea

Figura 7 - Estado inicial do processo ETL – KNIME

Rows: 129 | Columns: 10

#	RowID	Nome_Ar...	Nome_Ar...	XML	Id	Nome	Apelido	Email	Telemovel	Cidade	Pi
1	Row0_			<?xml versic	5	Marcus	Murphy	connerjohn@ya	1 234 567 895	Smithbury	Guinea
2	Row0_			<?xml versic	6	Stephen	Vasquez	ricemiguel@ya	1 234 567 896	Hamiltonstad	Rwand
3	Row0_			<?xml versic	11	Thomas	Garcia	rmiller@gmail.c	1 234 567 901	West Monicabu	Argenti
4	Row0_			<?xml versic	19	Marc	Nash	joshua59@gma	1 234 567 909	Davidton	French
5	Row0_			<?xml versic	22	Stacey	Adams	fjohnson@gmai	1 234 567 912	Clintonmouth	Icelandc
6	Row0_			<?xml versic	23	Hannah	Drake	sandersallen@t	1 234 567 913	Deannabury	Faroe I
7	Row0_			<?xml versic	23	Amanda	Martin	alana@sierra.c	1 234 567 921	Medtown	Asian

Figura 8 - Estado final do processo ETL – KNIME

Customer_ID	Name	Email	Phone	City	Country
99	Timothy Lane	janet99@davis.net	1234567989	South Heather	Turkey
98	Robert Blair	anicholson@gregory.com	1234567988	Josephfurt	United States Virgin Islands
97	Daniel Sanford	amanda13@yahoo.com	1234567987	Lake Stevenhaven	French Southern Territories
96	Levi Pierce	bayers@gmail.com	1234567986	Dayburgh	Bouvet Island (Bouvetoya)
95	Laura Fox	samanthaking@hogan.org	1234567985	New Theresa	Guadeloupe
94	Mr. David Cox	millermichael@gmail.com	1234567984	West Anthony	New Zealand
93	Jacob Cox	frank29@yahoo.com	1234567983	Robertburgh	San Marino
92	Heather Hamilton	kathy50@hotmail.com	1234567982	Davidshire	Croatia
91	Breanna Gonzalez	usimon@wright.com	1234567981	Andrewberg	Sierra Leone
90		moorebrittany@yahoo.com		West Christopher	Benin
9		naoeumemail		Rebeccafurt	Somalia
89	Erik Wang	gparks@yahoo.com	1234567979	New Desireebury	United Kingdom
88	Gregory Joseph	kimangela@yahoo.com	1234567978	Fletcherview	Algeria
87	Nathan Robbins	jenniferrodgers@garrison-baker.com	1234567977	East Tylertown	Bolivia
86	Molly Robinson	rebeccataylor@mcbride.info	1234567976	South Kathleen	Cuba
85	Sheila Coleman	cmyers@cooper-davies.com	1234567975	South Connie	Bangladesh
84	Gary Blair	leonardtaylor@bowers.net	1234567974	Lake Paul	Armenia
83	Aaron Clark	jessica98@hotmail.com	1234567973	Thomashaven	Fiji
82	Shaun Fernandez	greenmichael@hotmail.com	1234567972	Wrightburgh	Malaysia
81	Matthew Hall	patrick24@hotmail.com	1234567971	Lake Zachary	Suriname
80	Joy Fisher	kentchristian@gmail.com	1234567970	West Rogerview	Tuvalu
8	Matthew Johnson	austinkenneth@manning.net	1234567898	Kirstenborough	Israel
79	Daniel Brown	walkerrichard@young-grimes.com	1234567969	Villarrealand	Pakistan
78	Dr. Laurie Trevino	ryanbuchanan@hotmail.com	1234567968	Amberton	Saint Barthelemy
77	Jason Villegas	thomas00@stewart.com	1234567967	Fieldsland	Senegal
76	Kim Wang	jonesdylan@gmail.com	1234567966	Janicemouth	Turkey
75	Carla Webb	jennifer35@mann.biz	1234567965	Jasonmouth	Gabon
74	Bryan Moore	brendawong@juarez-camacho.biz	1234567964	Blackchester	Norfolk Island
73	Peter Krueger	vwelch@yahoo.com	1234567963	New Lawrenceton	Zimbabwe
72	Ms. Nicole Lee	mary21@henry-maynard.com	1234567962	West Ethan	Saint Barthelemy
71		douglasmccarthy@woodward-gallagher.com		West Sarah	United States Virgin Islands
70	Laura Duarte	jimenezchristine@dickson.com	1234567960	West Kimberlyhaven	Finland

Figura 9 - Estado inicial do processo ETL – PENTAHO

Customer_ID	Name	Email	Phone	City	Country	Email_valido
1	Deborah Griffith	baljoseph@wright-keith.net	1234567891	South Craigfort	denmark	Y
10	Ronald Osborn	staciekelley@heath.com	1234567900	Lake Benjamin	cameroon	Y
100	Karla Guerrero	wsimpson@yahoo.com	1234567990	Dustinview	argentina	Y
101	Brooke Kline	shirleythomas@allen-skinner.net	1234567991	New James	georgia	Y
102	Christopher Williams	ufigueroa@hotmail.com	1234567992	Cherylport	macedonia	Y
103	Ashley Castro	watsonmolly@yahoo.com	1234567993	Leepport	swaziland	Y
104	Anthony Ortiz	amartinez@gmail.com	1234567994	Michaelfort	nicaragua	Y
105	Billy Ochoa	mstrong@yahoo.com	1234567995	Samanthaburgh	andorra	Y

Figura 10 - Estado final do processo ETL – PENTAHO

2.2. Visualização de gráficos

Durante o processo de ETL também foram criados outros tipos de *output*. Na plataforma KNIME foram retiradas estas informações das tabelas:

- Número de clientes por país
- Número de livros lançados por ano
- Número de vendas por ano

Para agrupar o número de clientes por país este foi o fluxo:

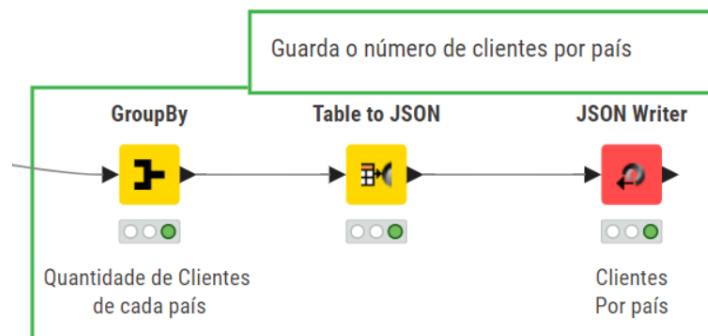


Figura 11 - Fluxo para guardar o número de clientes por país

Neste fluxo, o nó “Group By” vai agrupar cada cliente por país, de seguida o nó “Table to Json” vai servir para gerar o conteúdo em JSON através das colunas e por fim escreve num ficheiro.

Para agrupar o número de livros lançados por ano:

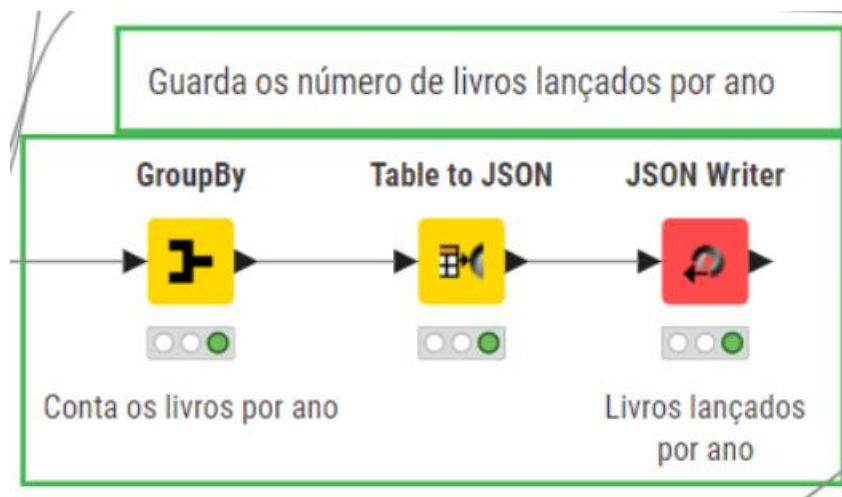


Figura 12 - Fluxo para guardar o número de livros lançados por ano

Neste fluxo, o nó "Group By" vai agrupar cada livro lançado por ano, de seguida o nó "Table to Json" vai servir para gerar o conteúdo em JSON através das colunas e por fim escreve num ficheiro.

Para agrupar o número de vendas por ano, o fluxo foi parecido:

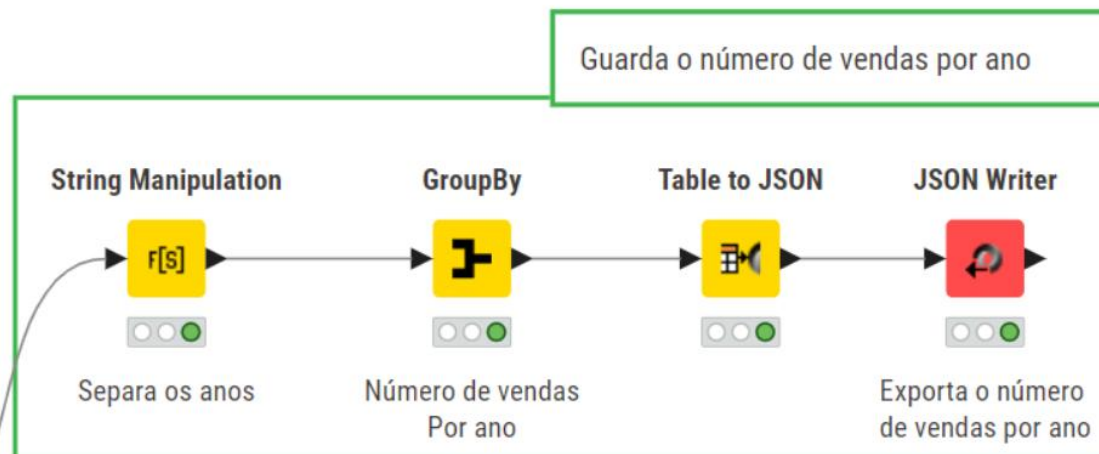


Figura 13 - Fluxo para guardar no número de vendas por ano

Neste fluxo, primeiro o nó "String Manipulation" serve para separar o ano da data da venda, o nó "Group By" vai agrupar cada venda por ano, de seguida o nó "Table to Json" vai servir para gerar o conteúdo em JSON através das colunas e por fim escreve num ficheiro.

Para representar os gráficos o método escolhido foi utilizar a biblioteca *ChartJs* para desenhar os gráficos de formas mais prática e eficiente. Foi desenvolvido código html,css e javascript para a demonstração dos gráficos.

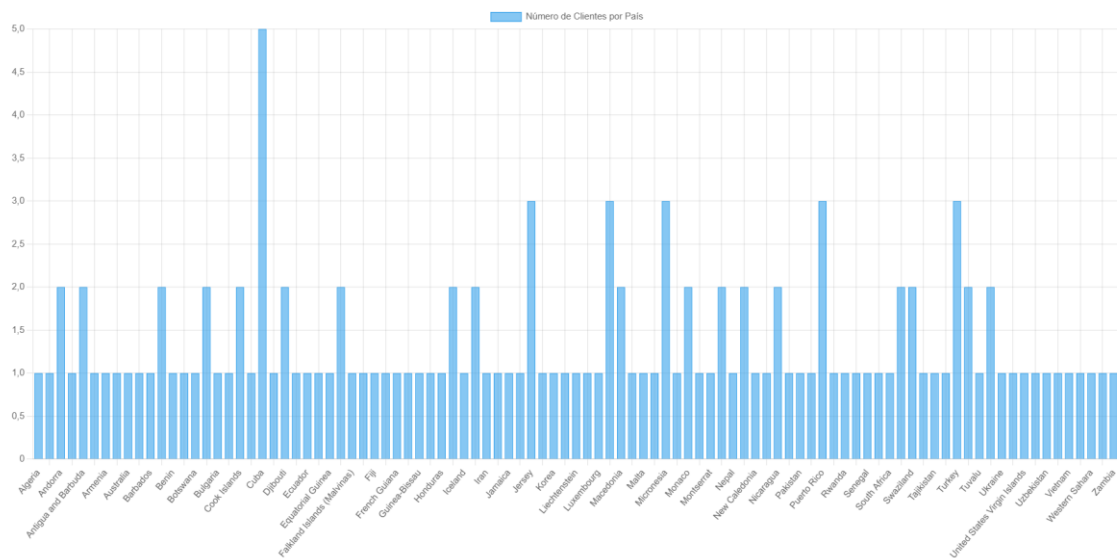


Figura 14 - Gráfico número de clientes por país

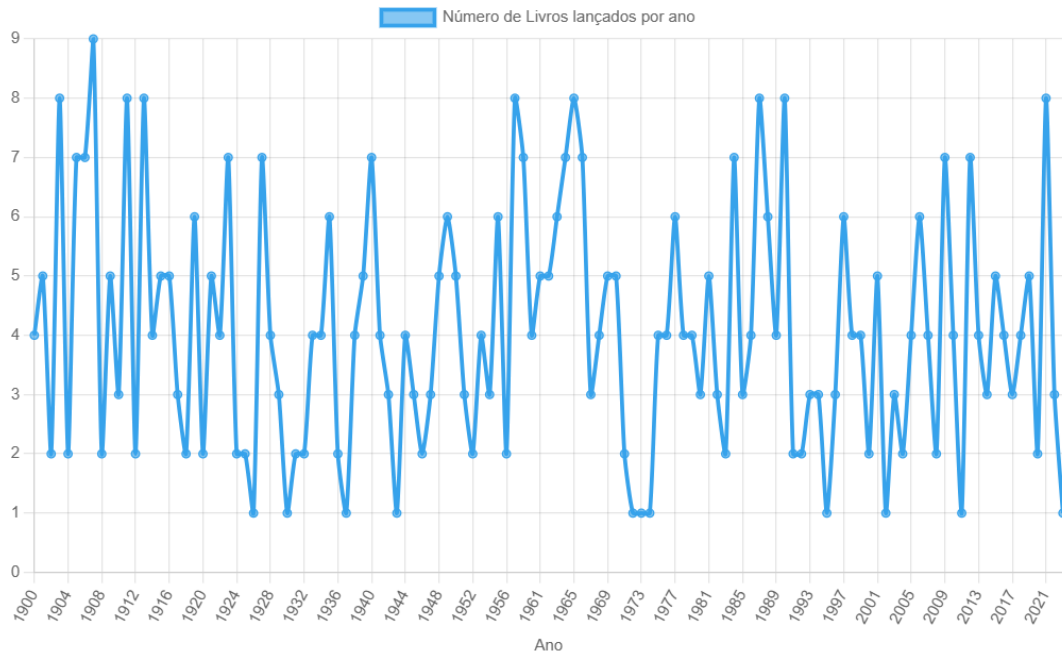


Figura 15 - Gráfico número de livros lançados por ano

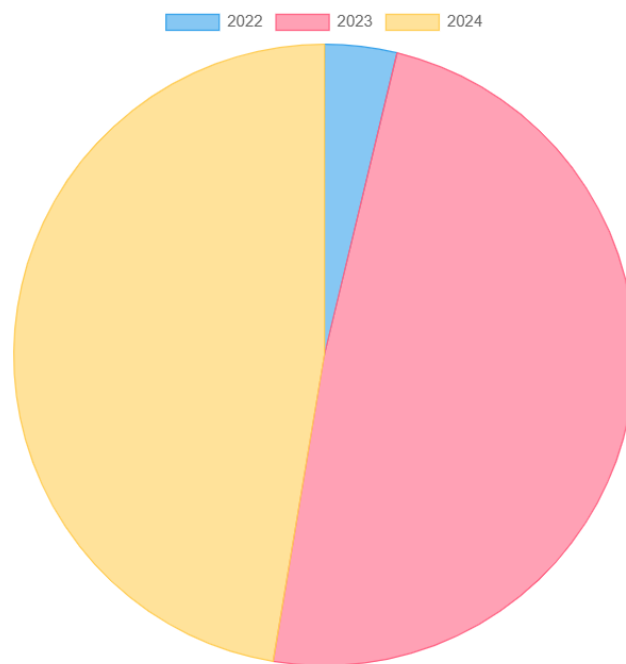


Figura 16 - Gráfico número de vendas por ano

Durante o processo de ETL também foram criados outros tipos de *output*. Na plataforma PENTAHO foram retiradas estas informações das tabelas:

- Número de livros vendidos por Género
- Número de vendas por País

Para representar os gráficos o método escolhido foi utilizar a biblioteca *ChartJs* para desenhar os gráficos de formas mais prática e eficiente. Foi desenvolvido código *html*, *css* e *javascript* para a demonstração dos gráficos.

A transformação “LivrosVendidosPorGenero.ktr” foi desenvolvida no PENTAHO com o objetivo de integrar dados provenientes de ficheiros XML e gerar um resumo agregado das vendas de livros por género. O processo segue uma sequência de sete passos principais:

1. Leitura de ficheiros XML de livros e pedidos através dos componentes *Get data from XML*, onde são extraídos registos de <Livro> e <Pedido> respetivamente.

2. Junção dos fluxos de dados via o passo *Stream Lookup*, que associa cada pedido ao livro correspondente, permitindo relacionar vendas com os livros (funciona como se fosse um *JOIN* em SQL).

#	Field	LookupField
1	Book_ID	Book_ID

#	Field	New name	Default	Type
1	Genre			None

Figura 17 - Stream Lookup

3. Agrupamento e agregação através do passo *Group by Genero*, calculando o total de vendas por género literário.
4. Ordenação dos resultados com *Sort rows Genero*, garantindo a apresentação estruturada e sequencial dos dados.
5. Filtragem condicional com *Filter Rows*, para filtrar caso tenha algum género nulo.
6. Exportação do resultado com *JSON Output*, que grava o conjunto de dados num ficheiro JSON.

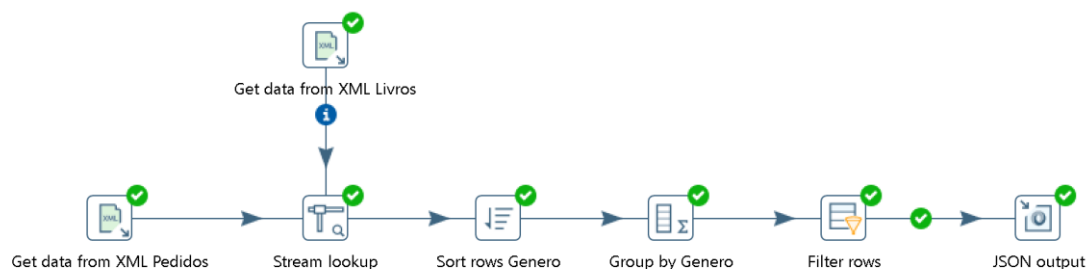


Figura 18 - Fluxo de Vendas por Género

Através de html css e javascript analisamos os dados que estão no ficheiro JSON referente a Pedidos por género, usando uma biblioteca de javascript para auxiliar na criação dos gráficos (Chart.js).



Figura 19 - Pedidos Totais por Género

A transformação *VendasPorPaíses.ktr* foi desenvolvida no PENTAHO com o objetivo de integrar dados provenientes de ficheiros XML e gerar um resumo agregado das vendas de livros por país. O processo segue uma sequência de sete passos principais:

1. Leitura de ficheiros XML de livros e pedidos através dos componentes *Get data from XML*, onde são extraídos registos de <Cliente> e <Pedido> respetivamente.
2. Junção dos fluxos de dados via o passo *Stream Lookup*, que associa cada pedido ao cliente correspondente, permitindo relacionar clientes com os pedidos (funciona como se fosse um *JOIN* em SQL).

Figura 20 - Stream Lookup de *VendasPorPaíses*

3. Ordenação dos resultados com *Sort rows Países*, garantindo a apresentação estruturada e sequencial dos dados.
4. Agrupamento e agregação através do passo *Group by País*, calculando o total de vendas por país.
5. Filtragem condicional com *Filter Rows*, para filtrar caso tenha algum país nulo.
6. Ordenação dos resultados com *Sort rows Total de pedidos*, garantindo a apresentação estruturada e sequencial dos dados.
7. Exportação do resultado com *JSON Output*, que grava o conjunto de dados num ficheiro JSON.

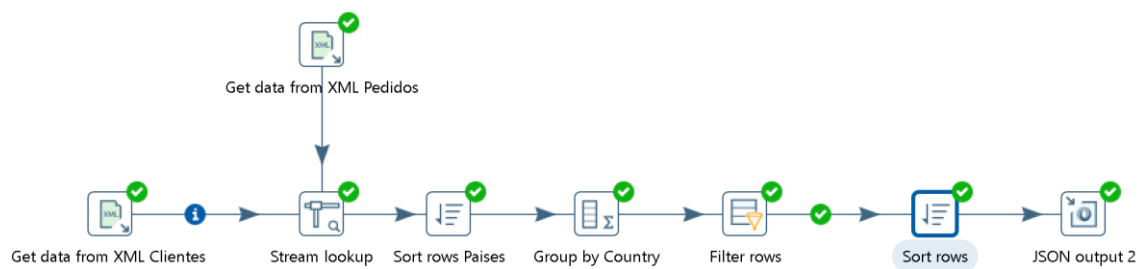


Figura 21 - Fluxo de Pedidos Totais por País

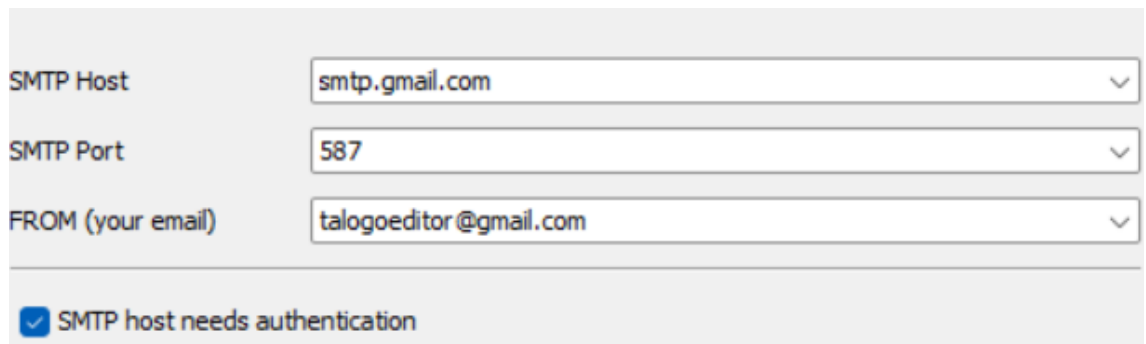
Através de html css e javascript analisamos os dados que estão no ficheiro JSON referente a Pedidos por género, usando uma biblioteca de javascript para auxiliar na criação dos gráficos (Chart.js).



Figura 22 - Top 15 Países com mais pedidos

2.3. Inclusão de funcionalidades extra

Para utilizar mais funcionalidades permitidas pelas plataformas utilizadas, no KNIME foi usado o serviço de envio de emails. Para verificar o uso do email, foram modificados dois registos, um no Livros.csv e outro no Clientes.xml para que fosse possível visualizar o funcionamento desta funcionalidade.



SMTP Host	smtp.gmail.com
SMTP Port	587
FROM (your email)	talogoeditor@gmail.com

☒ SMTP host needs authentication

Figura 23 - Uso do nó "Send Email"

```
Caro(a) ${SNome} ${SApelido},  
  
Este livro chamado "${STitulo}" vai ser lançado em ${IAno_Publicado}!
```

Figura 24 - Campos utilizados

Para utilizar mais funcionalidades permitidas pelas plataformas utilizadas, no PENTAHO foi usado o serviço de envio de emails.



Figura 25 - Fluxo de Enviar Email

O envio de email e dividido em 5 fases em que a primeira se trata de importar o ficheiro json PedidosTotaisPorGenero_0.json. Em seguida adicionamos constantes para os seguintes campos:

Adiciona valores constantes

Nome do Step Add constants

Campos :

#	Nome	Tipo	Formato	Tamanho	Precis	Moeda	Decimal	Grupo	Valor
1	email	String							a27963@alunos.ipca.pt
2	password	String							Hugo_1206612066
3	sender	String							Pentaho
4	smtp_port	String							587
5	smtp	String							smtp.office365.com
6	Title	String							Pentaho - JSON
7	email_message	String							

OK Cancela

Help

Figura 26 - Dados de constantes Email

Em terceiro constrói o conteúdo que vai no corpo da mensagem do email.

User defined Java expression

Nome do Step User defined Java expression

Fields:

#	New field	Java expression	Value type
1	mensagem	"Género: " + Genre + " - Total Pedidos: " + Total_Pedidos + "\n"	String

OK Cancela

Help

Figura 27 - Construção do corpo do email

Em quarto lugar fazemos um *memory group by* para nos auxiliar na construção do envio do email.

Step name Memory group by

Always give back a result row ☐

The fields that make up the group:

#	Group field	
1		

Get Fields

Aggregates :

#	Name	Subject	Type	
1	final_text	mensagem	Concatenate strings separated by	
2	email_dest	email	First value	
3	email_title	Title	First value	
4	email_sender	sender	First value	
5	smtp_server	smtp	First value	
6	smtp_port_num	smtp_port	First value	
7	pass	password	First value	

Get lookup fields

Help OK Cancela

Figura 28 - Ajuda Construção do email

E por fim o envio do email usando as contantes criadas nos passos anteriores.

Step name Mail

Addresses Server EMail Message Attached Files Embedded images

Destination

Destination address email_dest

Cc

BCc

E-mail sender

Sender name email_sender

Sender address email_dest

Reply to

Contact

Contact phone

Help OK Cancela

Figura 29 - Envio Email I

Addresses Server **Email Message** Attached Files Embedded images

SMTP Server

SMTP Server smtp_server ▼

Port smtp_port_num ▼

Authentication

Use authentication ☒

Authentication user email_dest ▼

Authentication password pass ▼

Use secure authentication ☒

Secure connection type TLS ▼

Figura 30 - Envio Email II

Addresses Server **Email Message** Attached Files Embedded images

Message Settings

Include date in message? ☐

Only send comment in mail body? ☐

Use HTML format in mail body? ☐

Encoding UTF-8 ▼

Manage priority ☐

Priority Normal ▼

Importance Normal ▼

Sensitivity Normal ▼

Message

Subject email_title ▼

Comment final_text ▼

Figura 31 - Envio Email III

Para tornar o trabalho ainda mais simples, foram criados *job's* para facilitar o fluxo utilizado. No primeiro *metanode* está o fluxo responsável pela preparação dos dados e no segundo *metanode* está o fluxo principal do processo ETL.



Figura 32 - Job KNIME

No PENTAHO foi utilizado um *job* em que chama todas as transformações para facilitar todos os fluxos utilizados.



Figura 33 - Job Pentaho

2.4. Vídeos de demonstração do fluxo

Demonstração fluxo do KNIME:



Demonstração fluxo do PENTAHO:



3. Conclusão

A realização deste projeto permitiu compreender de forma prática a importância dos processos de tratamento e gestão de dados, desde a recolha e preparação até à análise final. Foi possível perceber que, para além da execução técnica, o sucesso de um projeto de ETL que depende também da capacidade de interpretar os dados, identificar incoerências e aplicar métodos de verificação e normalização.

O uso do KNIME tornou o desenvolvimento do fluxo de trabalho mais visual, intuitivo e automatizado. Durante o projeto, foram desenvolvidas competências técnicas e de planeamento, reforçando a importância do tratamento adequado da informação para garantir a qualidade dos dados e a fiabilidade das decisões. Apesar das dificuldades iniciais, estas foram ultrapassadas, permitindo concluir o projeto com sucesso.

O uso do PENTAHO permitiu que o desenvolvimento dos processos de integração de dados fosse mais estruturado, automatizado e eficiente. Ao longo do projeto, foram adquiridas competências técnicas e de planeamento, destacando-se a importância de um tratamento adequado e consistente da informação para assegurar a qualidade dos dados e a confiabilidade das decisões baseadas neles. Apesar dos desafios iniciais na configuração e no desenho dos fluxos de transformação, estes foram superados com sucesso, possibilitando a conclusão eficaz do projeto e o fortalecimento das boas práticas de integração de dados.

4. WebGrafia

Repositório utilizado: [link](#)