

Trabalho Prático de:
Integração de Sistemas de
Informação

Docente:
Óscar Ribeiro

Realizado por:
Daniela Brito – n.º 25591

Ano Letivo 2024/2025

26 de outubro de 2024

Índice

1.	Enquadramento	3
2.	Problema	5
3.	Estratégia Utilizada	7
4.	Transformações.....	10
5.	Jobs	15
6.	Vídeo (QR Code)	16
7.	Conclusão	16
8.	Bibliografia	17

Índice de Figuras

FIGURA 1- DATA FLOW PRINCIPAL	9
FIGURA 2- DATA FLOW 2 (ENVIA EMAILS)	9
FIGURA 3- CSV INPUT	10
FIGURA 4- SORT ROWS.....	10
FIGURA 5- INPUT (EXEMPLO).....	11
FIGURA 6- SORT OUTPUT	11
FIGURA 7- UNIQUE ROWS	11
FIGURA 8- OUTPUT APÓS APLICAR UNIQUE ROWS	11
FIGURA 9- VALIDA EMAILS	12
FIGURA 10- REGISTOS, INCLUINDO ALGUNS EMAILS VÁLIDOS E INVÁLIDOS	12
FIGURA 11- EMAILS VÁLIDOS (PRONTOS PARA EXPORTAÇÃO)	12
FIGURA 12- EMAILS INVÁLIDOS (REGISTADOS NO FICHEIRO "FICHEIRO EMAILS INVÁLIDOS")	12
FIGURA 13- VALIDA PREÇOS INVÁLIDOS	13
FIGURA 14- EXEMPLO DE DADOS COM PREÇOS MISTOS.....	13
FIGURA 15- PREÇOS VÁLIDOS (PRONTOS PARA EXPORTAÇÃO)	13
FIGURA 16- PREÇOS INVÁLIDOS (FICHEIRO "FICHEIRO FORMATO PREÇOS INVÁLIDOS")	13
FIGURA 17- TRANSFORMAÇÃO: "MICROSOFT EXCEL WRITER"	14
FIGURA 18- TRANSFORMAÇÃO: "MICROSOFT EXCEL WRITER" (CONTENT)	14
FIGURA 19- O FICHEIRO EXCEL GERADO TERÁ O SEGUINTE ASPETO:	14

1. ENQUADRAMENTO

Este trabalho foi realizado no âmbito da disciplina de Integração de Sistemas de Informação (ISI) do curso de Engenharia de Sistemas Informáticos, com o objetivo de aplicar e experimentar processos de ETL (Extração, Transformação e Transferência) utilizando ferramentas adequadas, nomeadamente o Pentaho Kettle.

No cenário atual dos processos de negócio, a necessidade de integrar dados de diferentes fontes e formatos é um desafio constante. As empresas lidam com grandes volumes de dados provenientes de diversas origens e com diferentes estruturas. A capacidade de extrair, transformar e transferir esses dados de forma eficiente permite às organizações tomar decisões informadas e garantir a coerência dos seus sistemas de informação.

O objetivo deste trabalho é simular um cenário real de migração e integração de dados de clientes, com foco na conversão de dados a partir de um ficheiro CSV para formatos mais versáteis, como JSON e Excel. Este processo não só garante a limpeza e normalização dos dados, como também oferece saídas em formatos distintos, facilitando a utilização dos dados tanto em ambientes web (JSON) quanto em relatórios empresariais ou aplicações de escritório (Excel). Para completar o processo, é incluída a funcionalidade de envio automático do ficheiro Excel final por e-mail, permitindo o acesso ágil ao relatório por parte dos utilizadores finais.

O projeto proposto segue o desenvolvimento de um processo ETL com o seguinte fluxo:

- 1. Extrair os dados** a partir de um ficheiro CSV que contém informações de clientes.
- 2. Transformar os dados**, aplicando operações de limpeza e normalização (correção de endereços de e-mail, remoção de duplicatas).
- 3. Transferir os dados** transformados para dois formatos distintos:
 - JSON: para utilização em sistemas web ou aplicações que consomem este formato.
 - Excel: permitindo a visualização em formato tabular e a fácil manipulação dos dados em ferramentas de escritório, como o Microsoft Excel.
 - Enviar o ficheiro Excel final por e-mail: para fornecer o resultado da transformação de dados diretamente ao utilizador, garantindo o acesso imediato ao relatório de dados processados.

Este trabalho cumpre os seguintes objetivos:

- Consolidar os conceitos associados à Integração de Sistemas de Informação.
- Aplicar ferramentas de suporte a processos ETL, nomeadamente o Pentaho Kettle.
- Realizar operações de transformação de dados, como limpeza e normalização.
- Demonstrar a importação e exportação de dados entre diferentes formatos (CSV para JSON e Excel).

- Automatizar a distribuição de resultados com o envio do ficheiro Excel final por e-mail.
- Contribuir para o desenvolvimento prático e experiência no uso de ferramentas de ETL.

2. PROBLEMA

Neste cenário prático, o problema em análise centra-se na migração e integração de dados de um ficheiro de produtos em formato CSV para formatos JSON e Excel com recurso a um processo ETL. As empresas, ao lidar com dados vindos de diferentes fontes e estruturas, frequentemente encontram dificuldades em assegurar a consistência, qualidade e adequação dos dados aos sistemas que os irão consumir. Este trabalho foca-se em resolver desafios típicos dessa migração, incluindo a limpeza, normalização e validação dos dados.

O ficheiro inicial, em formato CSV, contém dados de produtos como o *nome*, *descrição*, *preço*, *quantidade*, *e-mail do fornecedor* e *nome do fornecedor*. Porém, os dados apresentam alguns problemas que dificultam a integração noutros sistemas. Em particular, os *preços*, *e-mails* e *quantidades* encontram-se por vezes mal formatados, há entradas duplicadas e valores em falta em alguns registos. Este conjunto de dados mal estruturado compromete a precisão da informação e a sua fiabilidade para uso empresarial.

Desafios a Resolver

1. Limpeza e Normalização de Dados

- **Preços:** Os preços dos produtos encontram-se em formatos variados e incorretos para o sistema, como o uso de ponto em vez de vírgula para separar as casas decimais, ou até mesmo valores não numéricos. O processo ETL inclui a transformação para um formato coerente de preços (ex.: NN,NN€), garantindo que todos os valores estão prontos para futuras operações financeiras e análises.
- **E-mails:** Os endereços de e-mail dos fornecedores apresentam erros como o uso de caracteres incorretos e falta de estrutura válida (ex.: falta do "@" ou domínio incorreto). A validação e normalização dos e-mails permitirá a integração correta com outras aplicações e a criação de canais de contacto consistentes com os fornecedores.

2. Remoção de Registos Duplicados

Alguns produtos, como "Cadeira" e "Mesa", aparecem duplicados, contendo os mesmos valores em todas as suas colunas. O sistema ETL é configurado para identificar e remover estas entradas repetidas, assegurando que cada produto tem apenas uma linha no relatório final, evitando redundâncias.

3. Tratamento de Dados Incompletos

Alguns registos estão incompletos, como é o caso de produtos com a quantidade em branco. Estes dados serão tratados de forma a serem identificados para correção ou verificação adicional, ou marcados para revisão.

Objetivo do Processo ETL

Para além das etapas de transformação, este trabalho integra a funcionalidade de envio automático do ficheiro Excel final por e-mail. Esta solução oferece ao utilizador final um acesso direto e imediato ao relatório de produtos em formato Excel, uma prática comum no contexto empresarial onde a partilha de relatórios é fundamental. Assim, o trabalho pretende demonstrar a execução de cada uma das fases do processo ETL, com especial atenção à correção e consistência dos dados e à sua apresentação final em diferentes formatos, possibilitando a integração dos dados em sistemas web e de análise empresarial.

Para atingir estes objetivos, é utilizada a plataforma Pentaho Kettle, uma ferramenta de ETL, que facilita a automação de processos, assegura a qualidade dos dados e permite a interoperabilidade entre formatos distintos, essencial em qualquer projeto de Integração de Sistemas de Informação.

3. ESTRATÉGIA UTILIZADA

Para a resolução do problema de **integração e transformação de dados de produtos**, adotou-se uma abordagem estruturada de ETL (Extração, Transformação e Transferência) através da ferramenta *Pentaho Kettle*. A estratégia foca-se na extração de dados de um ficheiro CSV, transformação com correções e validações, e transferência dos dados para formatos JSON e Excel, garantindo a sua adequação a diferentes contextos de uso empresarial. Este processo também inclui o envio do relatório final em Excel para um e-mail específico, facilitando a distribuição e consulta dos dados.

Operadores e Processos Envolvidos

1. Extração dos Dados (Extração)

A primeira etapa do processo ETL consiste na leitura do ficheiro CSV, que contém campos como *Produto*, *Descrição*, *Preço*, *Quantidade*, *Email_Fornecedor* e *Fornecedor*. A ferramenta Kettle utiliza um operador de entrada que permite definir o ponto-e-vírgula como delimitador dos campos. Esta configuração permite interpretar corretamente os valores monetários em euros, onde a vírgula separa as casas decimais (ex.: `159,99€`).

2. Transformação dos Dados (Transformação)

a) **Limpeza de Dados:** Esta etapa inclui operações de verificação e correção para garantir a consistência dos dados:

- **Remoção de registos duplicados:** Identificação e eliminação de registos repetidos (ex.: duplicações de produtos como "Cadeira" e "Mesa") através do step *Unique Rows*, para garantir que cada produto está representado de forma única no ficheiro final.
- **Preenchimento de campos vazios:** Identificação de valores ausentes (ex.: campos de quantidade em branco) para preenchimento ou marcação como inválidos, de forma a manter a integridade dos dados.

b) **Normalização e Correção de Preços:**

Para os preços, são aplicadas transformações que substituem pontos por vírgulas e garantem o formato correto, incluindo as duas casas decimais (ex.: transformar `29.99` em `29,99€`). Esta etapa assegura que os dados estejam consistentes para processamento e análise.

Exemplo de expressão regular:

`^\\d{1,3}(\\.\\d{2})?€`

- `^`: Início da linha.
- `\\d{1,3}`: Entre 1 a 3 dígitos antes da vírgula.

- `(,\d{2})?`: Uma vírgula seguida de dois dígitos (decimal opcional).
- `€`: Símbolo do euro.
- `$`: Fim da linha.

c) Validação de E-mails:

A correção dos e-mails de fornecedores utiliza expressões regulares (ER) que garantem que todos os endereços estão no formato padrão (ex.: `nome@dominio.com`). Esta validação abrange:

- **Início e fim da string:** Utilização de símbolos `^` e `\$` para assegurar que o e-mail corresponde exatamente ao formato.
- **Verificação de caracteres permitidos antes do @:** Apenas letras, números e os caracteres `.` , `_` , `%` , `+` e `-`.
- **Formato de domínio:** Garantir que o e-mail inclui um domínio válido após o `@` , com uma extensão mínima de dois caracteres, como `.com` ou `.pt`.

Exemplos como `fornecedor8@exemplo,com` (com uma vírgula em vez de ponto) e `fornecedor10@exemplo` (domínio incompleto) são corrigidos ou identificados para revisão, garantindo a adequação dos contactos de fornecedores.

Expressão Regular:

`^[a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}$`

- `^`: Início da linha.
- `[a-zA-Z0-9._%+-]+`: Permite letras, números e certos caracteres especiais antes do @.
- `@`: Símbolo obrigatório.
- `[a-zA-Z0-9.-]+`: Letras, números, pontos e hífens no domínio.
- `\.`: Ponto literal.
- `[a-zA-Z]{2,}`: Extensão do domínio com pelo menos 2 letras.
- `$`: Fim da linha.

3. Transferência dos Dados (Transferência)

a) A última etapa consiste na **exportação dos dados transformados** para dois formatos distintos:

- **JSON:** Um formato leve e amplamente usado em aplicações web e APIs. A exportação para JSON permite que os dados sejam facilmente integrados em sistemas web e reutilizados em ambientes de desenvolvimento digital.
- **Excel:** A exportação para Excel permite que os dados sejam visualizados em formato tabular e manipulados por aplicações de escritório, como o Microsoft Excel, facilitando a análise e preparação de relatórios.

b) Envio do Relatório Final por E-mail:

Para automatizar a distribuição dos dados, o ficheiro Excel final é enviado por e-mail a um destinatário específico, utilizando um Job que inclui a configuração do servidor de e-mail, destinatário e anexos. Esta etapa é essencial para garantir a rápida disponibilização dos dados tratados aos utilizadores finais, diretamente no seu ambiente de trabalho.

Síntese

Desta forma consigo aplicar o ETL para resolver problemas de integração de dados em diferentes formatos e assegurar a consistência e qualidade das informações. Este fluxo de trabalho simula um cenário empresarial real, onde dados provenientes de um ficheiro CSV são convertidos, validados e distribuídos automaticamente, oferecendo uma solução prática e eficiente para a gestão e integração de sistemas de informação.

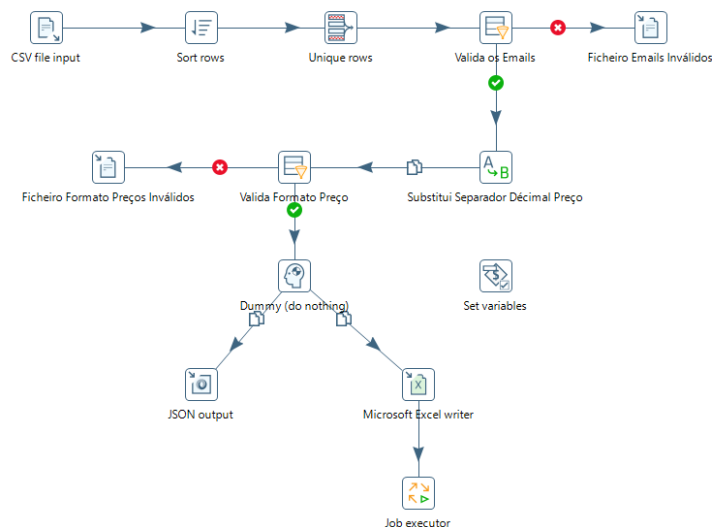


Figura 1- Data Flow Principal

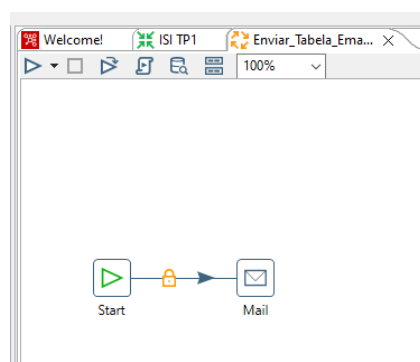


Figura 2- Data Flow 2 (Envia Emails)

4. TRANSFORMAÇÕES

1. CSV File Input

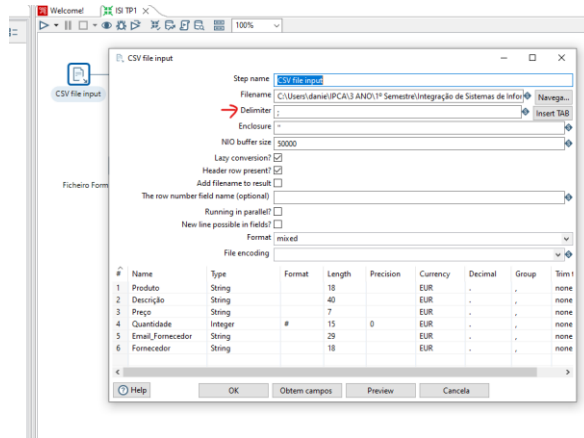


Figura 3- CSV Input

- **Objetivo:** Extrair dados de um ficheiro CSV que contém informações sobre produtos.
- **Configurações:** No input do CSV, é definido o caminho do ficheiro, o delimitador de campos (neste caso, um ponto e vírgula), e é especificado que a primeira linha contém os nomes dos campos.
- **Resultado Esperado:** Os dados são carregados para o sistema, prontos para as transformações seguintes.

2. Sort Rows

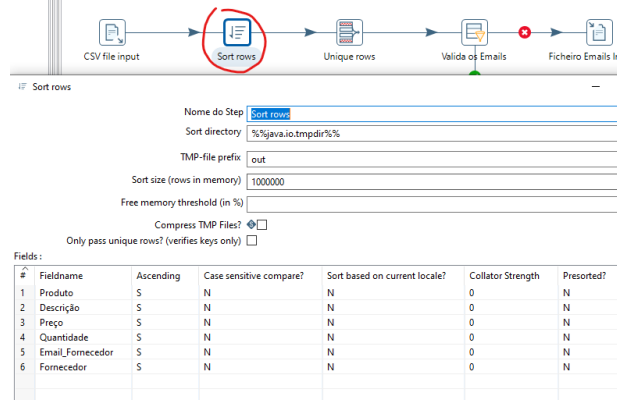


Figura 4- Sort Rows

- **Objetivo:** A ordenação dos dados é essencial para assegurar que as entradas duplicadas sejam consecutivas, permitindo ao passo seguinte "Unique rows" identificar e eliminar qualquer redundância
- **Configuração:** A ordenação é feita de forma ascendente em cada um dos campos seleccionados.
- **Resultado:**

Produto	Descrição	Preço	Quantidade	Email_Fornecedor	Fornecedor
Cadeira	Madeira	30.00€	10	fornecedor1@exemplo.com	Móveis ABC
Mesa	Vidro	50.00€	5	fornecedor2@exemplo.com	Móveis ABC
Cadeira	Madeira	30.00€	10	fornecedor1@exemplo.com	Móveis ABC
Armário	Aço	120.00€	3	fornecedor3@exemplo.com	Móveis XYZ
Cadeira	Madeira	35.00€	10	fornecedor1@exemplo.com	Móveis ABC

Figura 5- Input (Exemplo)

Produto	Descrição	Preço	Quantidade	Email_Fornecedor	Fornecedor
Armário	Aço	120.00€	3	fornecedor3@exemplo.com	Móveis XYZ
Cadeira	Madeira	30.00€	10	fornecedor1@exemplo.com	Móveis ABC
Cadeira	Madeira	30.00€	10	fornecedor1@exemplo.com	Móveis ABC
Cadeira	Madeira	35.00€	10	fornecedor1@exemplo.com	Móveis ABC
Mesa	Vidro	50.00€	5	fornecedor2@exemplo.com	Móveis ABC

Figura 6- Sort Output

3. Unique Rows

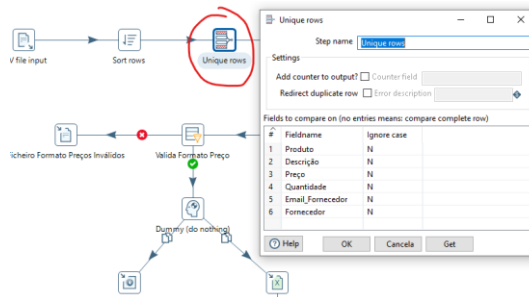


Figura 7- Unique Rows

- **Objetivo:** Remove registos duplicados da entrada de dados, garantindo que cada produto seja representado apenas uma vez. Uma "Cadeira de Madeira" com as mesmas especificações (preço, quantidade, fornecedor) deve aparecer uma única vez. Apenas os registos em que todos estes campos coincidirem com a linha anterior serão considerados duplicados e, portanto, removidos.
- **Configurações:** São definidos os campos para determinar a unicidade, como "Produto", "Descrição", etc.
- **Resultado Esperado:** Lista de dados sem duplicação.

Produto	Descrição	Preço	Quantidade	Email_Fornecedor	Fornecedor
Armário	Aço	120.00€	3	fornecedor3@exemplo.com	Móveis XYZ
Cadeira	Madeira	30.00€	10	fornecedor1@exemplo.com	Móveis ABC
Cadeira	Madeira	35.00€	10	fornecedor1@exemplo.com	Móveis ABC
Mesa	Vidro	50.00€	5	fornecedor2@exemplo.com	Móveis ABC

Figura 8- Output após aplicar Unique Rows

4. Valida os Emails

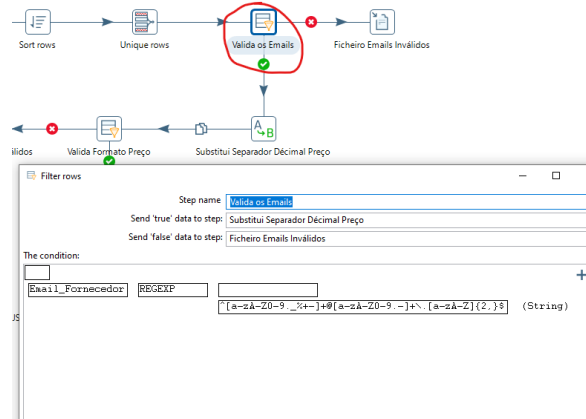


Figura 9- Valida Emails

- **Objetivo:** o passo "Valida Emails" é utilizado para verificar a conformidade dos endereços de email dos fornecedores, identificando e separando os que não seguem o formato padrão, utilizando uma expressão regular.
- **Configurações:** 1. aplicada exclusivamente ao campo "Email_Fornecedor". Expressão regular **`^[a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}$`**
Verifica se o email começa com letras, números ou certos caracteres permitidos (._%+-);
Verifica se inclui um @ seguido por um domínio válido (letras, números, . ou -);
Verifica se termina com uma extensão mínima de dois caracteres, como .com ou .pt.
- **Resultado:**

Produto	Descrição	Preço	Quantidade	Email_Fornecedor	Fornecedor
Cadeira	Madeira	30.00€	10	fornecedor1@exemplo.com	Móveis ABC
Mesa	Vidro	50.00€	5	fornecedor2@exemplo.com	Móveis ABC
Armário	Aço	120.00€	3	fornecedor3@exemplo.com	Móveis XYZ
Cadeira	Madeira	35.00€	10	fornecedor8@exemplo.com	Móveis ABC
Mesa	Plástico	20.00€	8	fornecedor10@exemplo	Móveis DEF

Figura 10- registos, incluindo alguns emails válidos e inválidos

Produto	Descrição	Preço	Quantidade	Email_Fornecedor	Fornecedor
Cadeira	Madeira	30.00€	10	fornecedor1@exemplo.com	Móveis ABC
Mesa	Vidro	50.00€	5	fornecedor2@exemplo.com	Móveis ABC
Cadeira	Madeira	35.00€	10	fornecedor8@exemplo.com	Móveis ABC

Figura 11- Emails Válidos (prontos para exportação)

Produto	Descrição	Preço	Quantidade	Email_Fornecedor	Fornecedor
Armário	Aço	120.00€	3	fornecedor3@exemplo.com	Móveis XYZ
Mesa	Plástico	20.00€	8	fornecedor10@exemplo	Móveis DEF

Figura 12- Emails Inválidos (registados no ficheiro "Ficheiro Emails Inválidos")

5. Valida Formato Preço

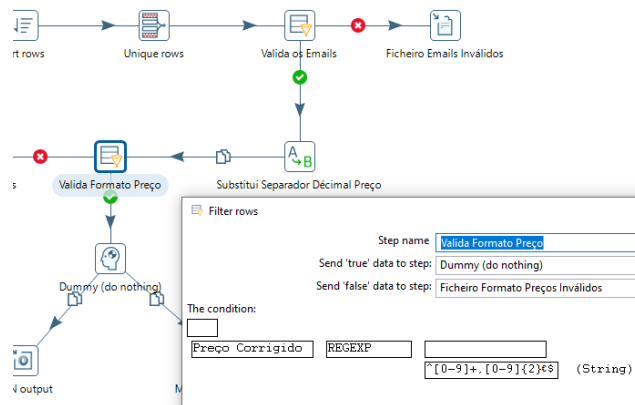


Figura 13- Valida Preços inválidos

- **Objetivo:** Verificar se os preços estão no formato correto, necessário para garantir a consistência dos dados financeiros. Esta verificação identifica os valores de preço que não seguem o formato esperado e separa-os para uma análise e possível correção (ex.: falta do símbolo "€", uso de ponto em vez de vírgula).
- **Configurações:** Aplica-se ao campo "Preço Corrigido". Utiliza uma expressão regular para assegurar que os preços estão no formato correto, com duas casas decimais e o símbolo "€" no final.

`^[0-9]+,[0-9]{2}€$`

Esta expressão verifica:

- Que o valor começa com um ou mais dígitos (inteiro);
 - Tem uma vírgula como separador decimal;
 - Seguido de exatamente dois dígitos após a vírgula;
 - Termina com o símbolo "€".
- **Resultado:**

Produto	Descrição	Preço Corrigido	Quantidade	Email_Fornecedor	Fornecedor
Cadeira	Madeira	30,00€	10	fornecedor1@exemplo.com	Móveis ABC
Mesa	Vidro	50,00€	5	fornecedor2@exemplo.com	Móveis ABC
Armário	Aço	120,00€	3	fornecedor3@exemplo.com	Móveis XYZ
Cadeira	Madeira	35,00€	10	fornecedor8@exemplo.com	Móveis ABC
Mesa	Plástico	20,00	8	fornecedor10@exemplo	Móveis DEF

Figura 14- Exemplo de Dados com Preços Mistos

Produto	Descrição	Preço Corrigido	Quantidade	Email_Fornecedor	Fornecedor
Cadeira	Madeira	30,00€	10	fornecedor1@exemplo.com	Móveis ABC
Mesa	Vidro	50,00€	5	fornecedor2@exemplo.com	Móveis ABC
Cadeira	Madeira	35,00€	10	fornecedor8@exemplo.com	Móveis ABC

Figura 15- Preços Válidos (prontos para exportação)

Produto	Descrição	Preço Corrigido	Quantidade	Email_Fornecedor	Fornecedor
Armário	Aço	120,00€	3	fornecedor3@exemplo.com	Móveis XYZ
Mesa	Plástico	20,00	8	fornecedor10@exemplo	Móveis DEF

Figura 16- Preços Inválidos (ficheiro "Ficheiro Formato Preços Inválidos")

6. Exportação para JSON e Excel

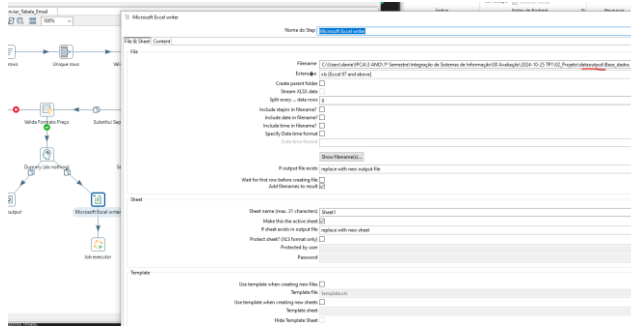


Figura 17- Transformação: "Microsoft Excel writer"

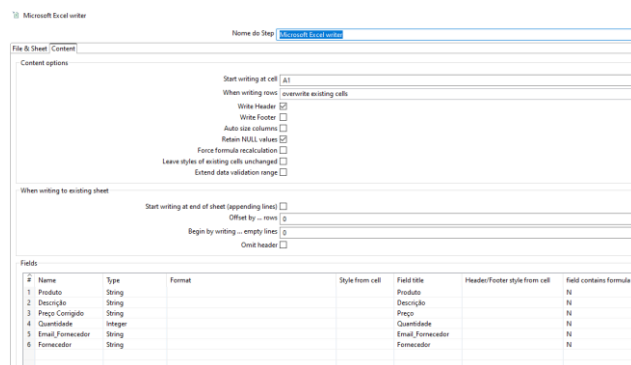


Figura 18- Transformação: "Microsoft Excel writer"(Content)

- **Objetivo:** Exportar os dados para um ficheiro Excel, facilitando a análise tabular e a visualização em ferramentas de office.
- **Configurações:** O ficheiro é configurado para ter uma folha ativa inicial com os dados na célula A1. O ficheiro gerado é do tipo .xls.
- **Resultado:**

Produto	Descrição	Preço Corrigido	Quantidade	Email_Fornecedor	Fornecedor
Cadeira	Madeira	30,00€	10	fornecedor1@exemplo.com	Móveis ABC
Mesa	Vidro	50,00€	5	fornecedor2@exemplo.com	Móveis ABC

Figura 19- O ficheiro Excel gerado terá o seguinte aspeto:

5. JOBS

1. Envio Automático do Ficheiro Excel

The image displays three screenshots of a 'Mail' configuration window, likely from a data integration tool. The window has a sidebar with a 'Mail' icon and a main content area with tabs: 'Addresses', 'Server', 'Email Message', and 'Attached Files'.

- Destination Tab:** Shows fields for 'Name of mail job entry' (Mail), 'Destination address' (a25591@alunos.ipca.pt), 'Cc', 'Bcc', 'Sender name' (Aluno), 'Sender address' (a25591@alunos.ipca.pt), 'Reply to', 'Contact person', and 'Contact phone'.
- SMTP Server Tab:** Shows fields for 'SMTP Server' (smtp.office365.com), 'Port' (587), 'Authentication' (checked), 'Authentication user' (a25591@alunos.ipca.pt), 'Authentication password' (masked with dots), 'Use secure authentication?' (checked), and 'Secure connection type' (TLS).
- Message Settings Tab:** Shows checkboxes for 'Include date in message?' (checked), 'Only send comment in mail body?' (checked), and 'Use HTML format in mail body?' (checked). It also has 'Encoding' (UTF-8), 'Manage priority' (unchecked), 'Priority' (Normal), 'Importance' (Normal), and 'Sensitivity' (Normal). The 'Message' section shows 'Subject' (Relatório de vendas) and 'Comment' (Segue em anexo o relatório final em Excel).

- **Objetivo:** O Job "Enviar Tabela por Email" automatiza o envio do ficheiro Excel gerado no processo ETL para os utilizadores finais por email. Este Job permite que o ficheiro, com os dados já limpos e transformados, seja distribuído de forma eficiente e automática, eliminando a necessidade de envio manual e assegurando que os dados estejam acessíveis de forma rápida
- **Configurações:**
 - a) Servidor e Autenticação de Email:
 - Servidor SMTP: smtp.office365.com, que permite o envio seguro do email.
 - Porta: 587, configurada para a utilização de TLS.
 - Autenticação: Email e password encriptada, com autenticação segura habilitada para proteger o envio de dados.

b) Destinatário e Mensagem:

- Destinatário: a25591@alunos.ipca.pt, email associado ao utilizador final.
- Assunto do Email: "Relatório de vendas".
- Mensagem: Inclui uma breve nota, "Segue em anexo o relatório final em Excel.", alertando para o ficheiro anexo.
- Formato HTML: Habilitado para uma melhor apresentação do email.

c) Anexos: O Job indica que o ficheiro Excel gerado na transformação anterior está anexado ao email, permitindo ao destinatário aceder diretamente ao relatório completo de forma prática e rápida.

- **Resultado:** Email enviado

6. VÍDEO (QR CODE)

Demonstração da execução do trabalho em Pentaho Kettle:



<http://gofile.me/7s8bt/kQQdaBt8l>

7. CONCLUSÃO

Este trabalho prático realizado no âmbito da Unidade Curricular de Integração de Sistemas de Informação do 3º ano da Licenciatura em Engenharia de Sistemas Informáticos (LESI) permitiu consolidar conhecimentos teóricos e práticos sobre processos ETL (Extração, Transformação e Transferência). Com a utilização da ferramenta Pentaho Kettle, foi possível aplicar técnicas de limpeza, normalização e exportação de dados em diferentes formatos, incluindo JSON e Excel, assegurando a qualidade e a consistência dos dados finais.

Ao longo do trabalho, foram implementadas transformações específicas para corrigir e validar dados críticos, como o formato de emails e preços, e foi desenvolvido um Job para o envio automático do relatório em Excel por email, automatizando a entrega da informação aos utilizadores finais. Esta experiência prática contribuiu para o desenvolvimento de competências essenciais em integração de sistemas e automação de processos de dados, que são fundamentais na área da Engenharia de Sistemas Informáticos.

8. BIBLIOGRAFIA

- **Ribeiro, Ó. ETL - Introdução à Integração de Dados.** Instituto Politécnico do Cávado e do Ave, Escola Superior de Tecnologia, 2024-2025. Disponível no Moodle IPCA.
- **Ribeiro, Ó. Expressões Regulares com C# - Aplicações em ETL.** Instituto Politécnico do Cávado e do Ave, Escola Superior de Tecnologia, 2024-2025. Inclui exemplos de validação de dados em processos ETL. Disponível no Moodle IPCA.
- **Ribeiro, Ó. Pentaho Kettle - Tutorial e Aplicações em ETL.** Instituto Politécnico do Cávado e do Ave, Escola Superior de Tecnologia, 2024-2025. Documento de referência para utilização do Pentaho Data Integration (Kettle) no desenvolvimento de transformações e automação de processos.
- **Ribeiro, Ó. Integração ao Nível dos Dados - XML e JSON.** Instituto Politécnico do Cávado e do Ave, Escola Superior de Tecnologia, 2023-2024. Aborda o uso de XML e JSON para representação e integração de dados. Disponível no Moodle IPCA.