

# Relatório ISI

Gonçalo Costa 26007 -LESI

Docente: Óscar Ribeiro

## Índice

Objetivos: .....	3
Estratégia: .....	4
Estrutura: .....	5
Base de Dados: .....	6
Dados Base de Dados: .....	7
Ficheiro.Json: .....	8
Definir ambiente de string.....	10
Filtragens: .....	11
Job- Enviar dado Processado por Email .....	12
Email: .....	13
Node-Red .....	14
Vídeo: .....	16
Conclusão: .....	16

## Objetivos:

O objetivo deste trabalho é aplicar técnicas de integração e transformação de dados utilizando a plataforma Pentaho Data Integration (PDI). O primeiro foco é a importação de dados a partir de múltiplas fontes, especificamente em formato JSON, garantindo que as informações sejam extraídas corretamente e de maneira eficiente. Em seguida, o trabalho visa validar e transformar esses dados utilizando expressões regulares (Regex), assegurando a consistência e integridade dos campos cruciais, como nome, marca e cor, conforme os requisitos estabelecidos. Além disso, o projeto tem como objetivo implementar a ordenação dos dados com base em critérios específicos, como a data de fabrico e o nome, facilitando a organização e a análise subsequente. O trabalho também busca garantir a exportação dos dados processados em diferentes formatos, como arquivos Excel e JSON, para assegurar a compatibilidade e a manipulação eficiente das informações em diversos ambientes. Por fim, o objetivo é automatizar todo o fluxo de processamento, permitindo a integração, transformação e exportação dos dados de forma reutilizável e com mínima intervenção manual, aumentando a eficiência e a precisão do tratamento das informações.

## *Estratégia:*

A estratégia adotada envolve a utilização do Pentaho como ferramenta principal para a execução do processo de extração, transformação e carga (ETL) dos dados. O Pentaho foi escolhido devido à sua interface intuitiva e às suas robustas capacidades de integração.

O primeiro passo foi a definição dos objetivos do projeto e a identificação das fontes de dados. A extração dos dados foi realizada a partir de arquivos JSON, que foram processados para facilitar a manipulação.

Em seguida, foram criados fluxos de trabalho dentro do Pentaho. Esses fluxos incluíram etapas para limpar e formatar os dados, calcular a idade dos veículos e filtrar os registros com base em critérios específicos, como data de fabricação e marca.

Após a transformação dos dados, foi feita a carga no banco de dados. Essa etapa garantiu que os dados fossem inseridos de forma organizada, utilizando a funcionalidade de auto-incremento para o campo identificador.

Por fim, a análise dos dados foi realizada para identificar tendências e padrões. Relatórios foram gerados em formatos XML e Excel para facilitar a visualização das informações.

Essa abordagem simplificada permite um controle eficiente sobre os dados e contribui para a geração de insights valiosos, assegurando que o projeto atinja os seus objetivos.

## Estrutura:

Input	23/10/2024 11:32	Pasta de ficheiros
Output	23/10/2024 16:17	Pasta de ficheiros
Relatorio	23/10/2024 18:43	Pasta de ficheiros

Figura 1-Estrutura Ficheiros

A estrutura apresentada é organizada em três pastas principais: Input, Output e Relatório, cada uma com um propósito específico dentro do fluxo de trabalho.

- **Input:** Esta pasta contém os arquivos de entrada que são utilizados para o processamento inicial. São os dados brutos que serão transformados e analisados.
- **Output:** Nesta pasta, são armazenados os resultados gerados após as transformações e análises dos dados. Aqui, o utilizador encontrará os arquivos processados e formatados prontos para uso.
- **Relatório:** Esta pasta é destinada ao relatório do trabalho.

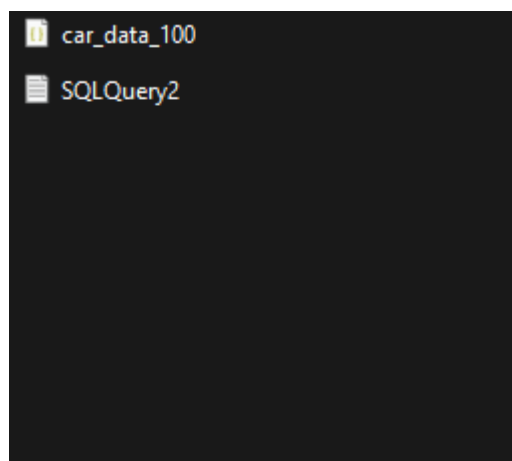


Figura 2-Ficheiro Input

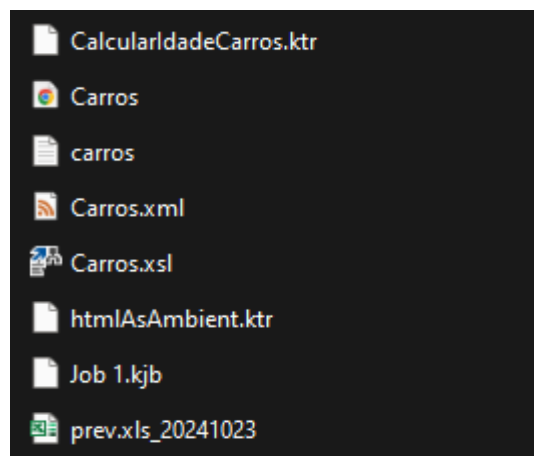
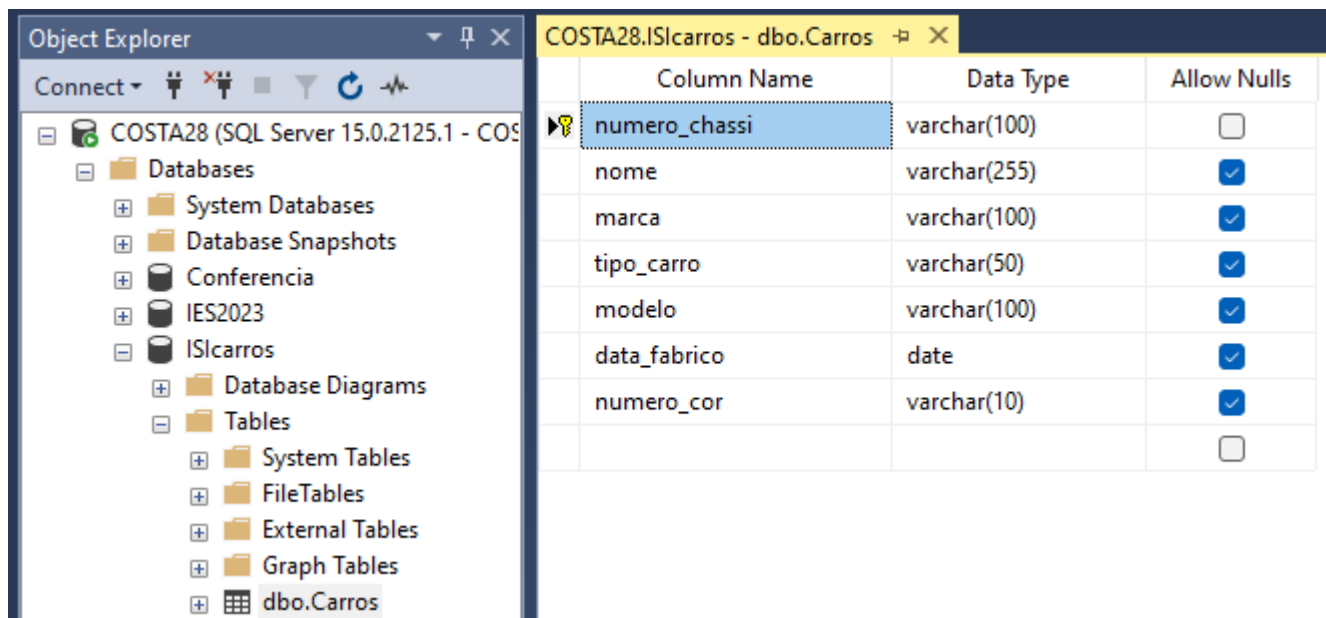


Figura 3- Ficheiro Output

## Base de Dados:



Column Name	Data Type	Allow Nulls
numero_chassi	varchar(100)	<input type="checkbox"/>
nome	varchar(255)	<input checked="" type="checkbox"/>
marca	varchar(100)	<input checked="" type="checkbox"/>
tipo_carro	varchar(50)	<input checked="" type="checkbox"/>
modelo	varchar(100)	<input checked="" type="checkbox"/>
data_fabrico	date	<input checked="" type="checkbox"/>
numero_cor	varchar(10)	<input checked="" type="checkbox"/>
		<input type="checkbox"/>

Figura 4- Estrutura Base Dados

A estrutura da base de dados ISIcars é composta pela tabela Carros, que contém informações cruciais sobre os veículos. Os principais campos incluem:

- numero\_chassi: Identificação única do veículo (VIN).
- nome: Nome do proprietário.
- marca: Marca do veículo.
- tipo\_carro: Classificação do carro (ex: SUV, Sedan).
- modelo: Modelo do veículo.
- data\_fabrico: Data de fabricação do carro.
- numero\_cor: Código da cor do veículo.

## Dados Base de Dados:

	numero_chassi	nome	marca	tipo_carro	modelo	data_fabrico	numero_cor
1	0NXPY3UZWOAKN1GIJ	Juliana Costa	Mercedes-Benz	Sedan	GLA	2015-02-12	557
2	1UGLGS3RHM89NHXSP	Carlos Oliveira	BMW	Hatchback	i8	2019-06-13	454
3	275V3IFOF4EZIDDZH	Pedro Pereira	Mazda	Hatchback	CX-5	2014-07-23	4CC
4	37NAABZE4HI2555X6	Maria Oliveira	Toyota	Sedan	Prius	2017-09-11	B99
5	3ZGYX1UVK2JAFT6A3	Ana Silva	Mazda	Convertible	Mazda3	2023-08-01	3BA
6	4N0TMYJXMP8SPALOW	Carlos Costa	Honda	Sedan	Civic	2022-03-24	58C
7	4N6UKDF6NYCD1CMHY	Carlos Oliveira	Mercedes-Benz	Convertible	C-Class	2010-11-22	CA9
8	5OZUYAZD3Q62M7XM8	Juliana Costa	Mercedes-Benz	Truck	GLE	2017-02-11	EA5
9	6JSCY90UYOWHK1N7K	Pedro Santos	Volkswagen	Sedan	Passat	2000-09-07	238
10	74YF7FRP2RAOTMRFG	Juliana Oliveira	Honda	Truck	CR-V	2021-07-09	3BE
11	8D9P6WKNYXIUJ7GJ	Pedro Costa	Mazda	Truck	Mazda3	2023-11-19	1F8
12	9MP0ULU9B2QBTHNWI	Carlos Oliveira	Chevrolet	Sedan	Malibu	2005-04-09	1CD
13	AL62TY13RDGJ6V2C3	João Pereira	Volkswagen	Coupe	Jetta	2015-02-14	3FF
14	D7QT9X3TNSL2O22E6	João Santos	Honda	Hatchback	Accord	2009-05-07	310
15	DM5NQJBWXFGRGXILH	Juliana Costa	BMW	Coupe	M3	2014-03-16	CBD
16	DNUO29IEVN3LX939C	Juliana Costa	Honda	Coupe	CR-V	2020-08-06	110
17	DS6AY8MYE46XYTRSI	Carlos Pereira	Toyota	SUV	Corolla	2015-11-08	123
18	ECW1R77LLH042QHHK	Pedro Costa	Audi	Coupe	Q7	2012-04-29	91D
19	EKVE3CTPA7QWVFSMX	Juliana Santos	Mazda	SUV	Mazda3	2014-01-26	EAF
20	F0I5J3XJTQM00MGBE	Juliana Silva	Volkswagen	Hatchback	Jetta	2021-09-06	491
21	FABAIQNQ3PD75I24N	Carlos Santos	Volkswagen	SUV	Golf	2019-09-25	14B
22	G5KJLFJCHG7UPOSLU	Juliana Costa	Audi	Sedan	Q7	2014-01-12	12D
23	GTIVQKRDMBJA0KFJJ	João Pereira	Chevrolet	Truck	Spark	2020-11-22	644
24	HQ3666AOXZANENC08	Maria Costa	Nissan	Coupe	Sentra	2008-06-06	724

Figura 5- Dados Base de Dados

A figura 5 representa uma amostra dos dados extraídos da tabela Carros na base de dados ISIcars.

## Ficheiro.Json:

```
{ } car_data_100.json X
C: > Users > Utilizador > OneDrive > Ambiente de Trabalho > TP01-26007 > Input > { } car_data_100.json > ...
1
2 {
3     "nome": "Jo\u00e3o Silva",
4     "marca": "Chevrolet",
5     "tipo_carro": "SUV",
6     "modelo": "Malibu",
7     "data_fabrico": "2001-08-10",
8     "numero_chassi": "0R09N099B7NVDAJJ1",
9     "numero_cor": "14D"
10 },
11 {
12     "nome": "Carlos Santos",
13     "marca": "Volkswagen",
14     "tipo_carro": "SUV",
15     "modelo": "Golf",
16     "data_fabrico": "2019-09-25",
17     "numero_chassi": "FABAIQNQ3PD75I24N",
18     "numero_cor": "14B"
19 },
20 {
21     "nome": "Maria Pereira",
22     "marca": "Audi",
23     "tipo_carro": "Sedan",
24     "modelo": "Q7",
25     "data_fabrico": "2021-10-01",
26     "numero_chassi": "S3PMPANUFGU4PM872",
27     "numero_cor": "F00"
28 },
```

Figura 6- Ficheiro .Json

O arquivo car\_data\_100.json contém um conjunto estruturado de dados que representa informações sobre diferentes veículos. Este arquivo é crucial, pois serve como a fonte primária para inicializar a base de dados do projeto. Cada entrada no JSON inclui detalhes como o nome do proprietário, a marca do carro, o tipo de veículo, o modelo, a data de fabricação, o número do chassi e a cor do veículo.



## Cálculo da Idade dos Carros

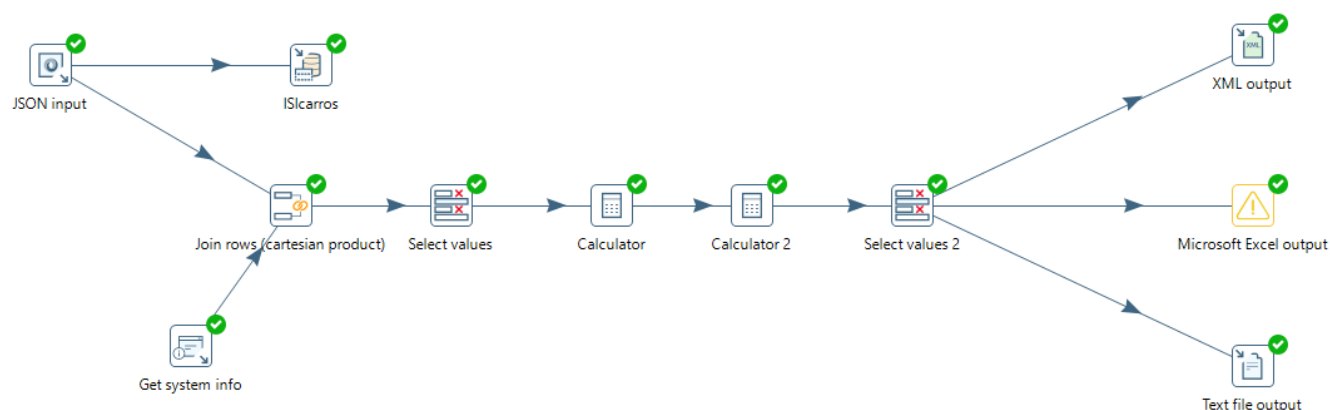


Figura 7- Transformação Cálculo Idade

O cálculo da idade dos carros foi realizado utilizando o Pentaho, com uma série de etapas cuidadosamente planejadas. Primeiro, extraímos os dados do arquivo JSON e do sistema por meio do passo "Get System Info", que forneceu a data atual necessária para os cálculos.

Na sequência, aplicamos um passo de "Select Values" para formatar a data atual no padrão desejado (yyyy/MM/dd), garantindo que estivesse em conformidade com o formato esperado para o cálculo. Esse formato é crucial para a comparação adequada entre as datas.

Posteriormente, utilizamos um passo "Calculator" para calcular a diferença entre a data atual e a data de fabricação dos veículos. Essa diferença foi inicialmente obtida em dias, e para convertê-la em anos, dividimos o resultado por 365, resultando na idade dos carros.

Em uma etapa final de "Select Values", removemos colunas desnecessárias, como "Constant365", "idade\_carro" e "data\_atual", mantendo apenas a coluna que representava a idade dos carros em anos. Essa abordagem não apenas simplificou os dados, mas também facilitou a visualização e análise dos resultados.

Além disso, como parte do processo de relatório, foi criado um arquivo Excel contendo esses dados, permitindo uma visualização organizada e intuitiva. Para complementar, também foram gerados um arquivo XML e um ficheiro de texto, proporcionando diferentes formatos de saída para atender a diversas necessidades de análise e apresentação dos dados.

## Definir ambiente de string



Figura 8- Transformação Ambiente String

Conforme exemplificado pelo professor, realizamos o carregamento de um arquivo gerado previamente em um Job, onde uma transformação de XML para HTML foi realizada. Nessa etapa, definimos que o conteúdo textual deve ser formatado como HTML, o que possibilita a manipulação do texto de maneira mais flexível e esteticamente agradável.

Ao formatar os dados dessa forma, garantimos que, no Job, será viável aplicar o passo de envio de e-mails de maneira eficiente. Isso significa que, ao final do processo, o texto gerado não só será legível e organizado, mas também poderá ser enviado diretamente por e-mail, facilitando a distribuição das informações entre os destinatários. Essa abordagem não apenas aprimora a apresentação dos dados, mas também otimiza a comunicação, assegurando que as partes interessadas recebam relatórios de forma clara e profissional.

## Filtragens:

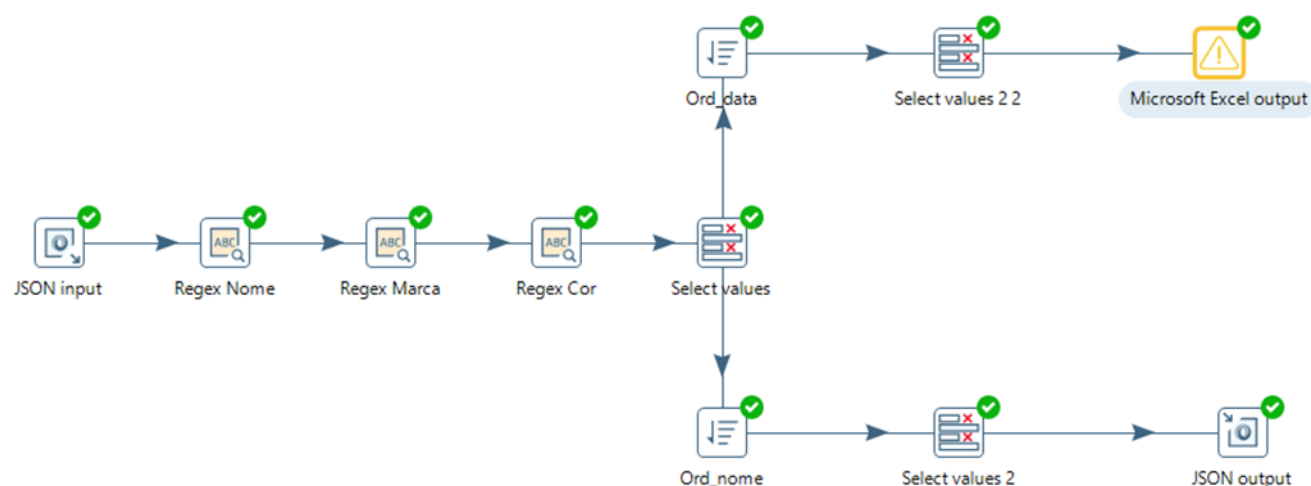


Ilustração 1- Filtragens

O fluxo de dados representado na imagem inicia com a leitura de um arquivo JSON, onde os dados são extraídos e posteriormente processados por meio de etapas de validação usando expressões regulares para os campos "nome", "marca" e "cor". Após a validação, os dados são filtrados e separados em dois caminhos. Um caminho ordena os dados pela data, enquanto o outro organiza os dados pelo nome. Em seguida, os dados de cada caminho são filtrados mais uma vez antes de serem exportados. Os dados ordenados por data são salvos em um arquivo Excel, e os dados organizados por nome são exportados em um arquivo JSON.

## *Job- Enviar dado Processado por Email*

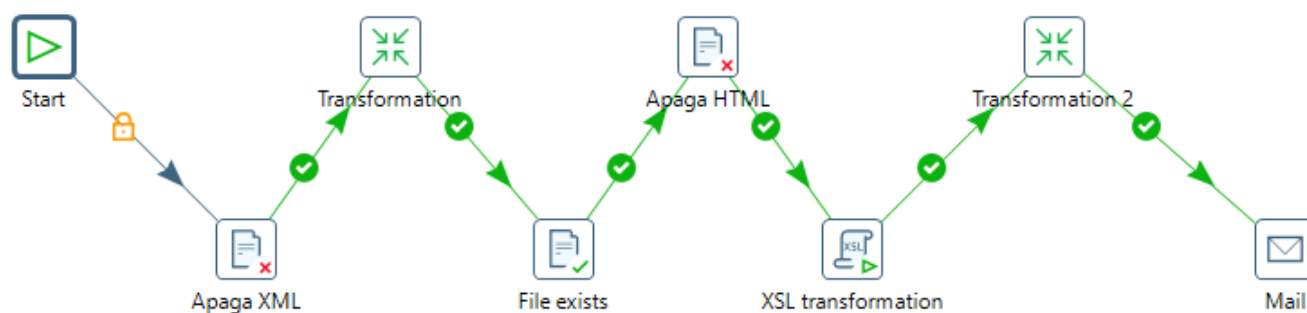


Figura 9- Transformação Email

Este Job realiza o processamento de dados em formato XML, preparando-os para envio por e-mail.

- Start: Inicia a execução do Job.
- Transformation: Aplica uma transformação inicial nos dados.
- Apaga XML: Remove arquivos XML existentes para evitar dados desatualizados.
- File Exists: Verifica a existência de um arquivo específico.
- Apaga HTML: Elimina arquivos HTML antigos para garantir dados atualizados.
- XSL Transformation: Converte dados XML em HTML usando uma folha de estilo XSL.
- Transformation 2: Realiza ajustes finais nos dados em HTML.
- Mail: Envia o conteúdo preparado por e-mail aos destinatários.

Email:

Carros

G

Gonçalo Oliveira da Costa

Para: Gonçalo Oliveira da Costa

Responder

Responder a todos

Reencaminhar

...

qua, 23/10/2024 16:17

Informações dos Carros

Número do Chassi	Nome	Marca	Tipo do Carro	Modelo	Data de Fabrico	Número de Cor
S3PMPANUFGU4PM872	Maria Pereira	Audi	Sedan	Q7	2021-10-01	F00
MSZHQQWBESE77INMN	Maria Pereira	Audi	Coupe	Q5	2004-07-27	EC6
6YBLCIX1GAOPS3PP0	Carlos Costa	Nissan	Sedan	Versa	2015-02-27	650
HKZGFLSJA1UUE74AQ	Pedro Santos	Mazda	Truck	Mazda6	2012-04-27	836
NGG86KZF0ZZU3T9P0	Ana Pereira	Ford	Convertible	Mustang	2003-05-30	485
NGV8RKEPUKZE05JP5	Pedro Oliveira	Mercedes-Benz	Hatchback	C-Class	2012-05-22	094
YQAVKFKX6PJ2XYT2Q	Juliana Santos	Mazda	Convertible	CX-5	2012-10-08	887
OVF66Z8Q3F6ZB7KOH	Juliana Oliveira	Volkswagen	Sedan	Golf	2021-09-22	DDD
QA0TXEYZEYZGZTXDC	Maria Pereira	Mercedes-Benz	Hatchback	GLA	2018-11-01	A01
5Z3NID8MF3VKFEF65	João Silva	Mercedes-Benz	Convertible	GLE	2005-06-07	DFB
AQACK2299OK30RD19	Ana Pereira	BMW	SUV	320i	2003-05-01	95B
HO8OACCTAXEU6CAZE	Maria Costa	Honda	Truck	Civic	2012-04-01	F0C
I9WJGKTEAUS4QTQTT	Ana Oliveira	Audi	Hatchback	A4	2008-05-15	6DE
C07CWP439D7HI2S20	Ana Costa	Ford	Truck	Mustang	2022-10-01	01F
7Y1B59AZL7MCYSQCL	Maria Santos	Toyota	Hatchback	Corolla	2000-04-08	6DE
Y5364HBYI6465RMNT	Carlos Costa	Volkswagen	Sedan	Golf	2000-02-25	D77
5YUZ9OWDL5P999GLG	Juliana Costa	Audi	Coupe	A4	2006-08-28	A14

Message date: 2024/10/23 16:17:06.594

Responder

Reencaminhar

Figura 10- Email

Na Figura 10 esta representado a informação que foi enviada para o respetivo email de envio.

## Node-Red

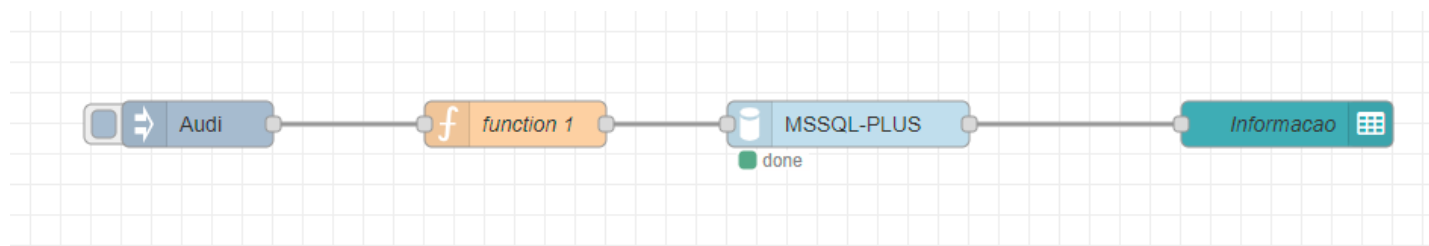


Figura 11- Note-Red Marca

Foi criado um fluxo no Node-RED com o objetivo de filtrar as marcas de carros, especificamente para identificar todas as pessoas que possuem um veículo da marca Audi. Utilizando a robustez do Node-RED, a implementação permite a manipulação eficiente de dados em tempo real.

numero_c...	nome	marca	tipo_carro	modelo	data_fabri...	numero_cor
ECW1R77LLH0...	Pedro Costa	Audi	Coupe	Q7	2012-04-29T00...	91D
G5KJLFJCHG7U...	Juliana Costa	Audi	Sedan	Q7	2014-01-12T00...	12D
IH4WD4OD53J...	Pedro Santos	Audi	Truck	A3	2003-09-24T00...	A44
KY5G7IWGXHY...	Ana Silva	Audi	Convertible	A3	2002-09-20T00...	323
LZE18K5MF7F...	Maria Santos	Audi	Hatchback	A3	2017-02-28T00...	F08
ZGXQJA3K6D9...	Juliana Silva	Audi	Sedan	Q5	2000-05-31T00...	7A5

Figura 12- Dados do Note-Red Marca

Na Figura 12 temos a tabela gerada pelo Node-Red que representa todos os utilizadores que possuem um carro da marca Audi.

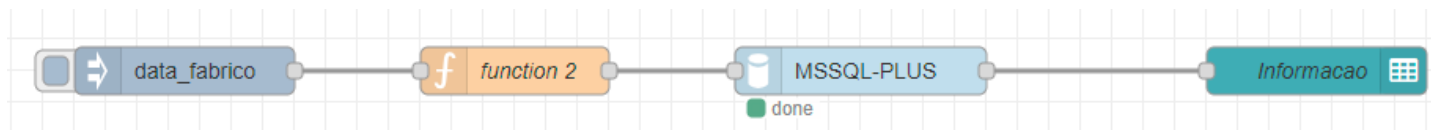


Figura 13- Node-Red Data

Foi criado um fluxo no Node-RED com o objetivo de filtrar os carros com data de fabricação superior a 2015. Este fluxo utiliza a robustez do Node-RED para manipulação eficiente de dados em tempo real, permitindo uma análise eficaz e dinâmica.

numero_c...	nome	marca	tipo_carro	modelo	data_fabri...	numero_cor
1UGLGS3RHM...	Carlos Oliveira	BMW	Hatchback	i8	2019-06-13T00...	454
37NAABZE4HI...	Maria Oliveira	Toyota	Sedan	Prius	2017-09-11T00...	B99
3ZGYX1UVK2J...	Ana Silva	Mazda	Convertible	Mazda3	2023-08-01T00...	3BA
4N0TMYJXMP8...	Carlos Costa	Honda	Sedan	Civic	2022-03-24T00...	58C
5OZUYAZD3Q...	Juliana Costa	Mercedes-Benz	Truck	GLE	2017-02-11T00...	EA5
74YF7FRP2RA...	Juliana Oliveira	Honda	Truck	CR-V	2021-07-09T00...	3BE
8D9P6WKNYXI...	Pedro Costa	Mazda	Truck	Mazda3	2023-11-19T00...	1F8
DNUO29IEVN3...	Juliana Costa	Honda	Coupe	CR-V	2020-08-06T00...	110
F0I5J3XTQM0...	Juliana Silva	Volkswagen	Hatchback	Jetta	2021-09-06T00...	491
FABAIQNQ3PD...	Carlos Santos	Volkswagen	SUV	Golf	2019-09-25T00...	14B
GTIVQKRDMBJ...	João Pereira	Chevrolet	Truck	Spark	2020-11-22T00...	644
LZE18K5MF7F...	Maria Santos	Audi	Hatchback	A3	2017-02-28T00...	F08
PSUC2D18YNH...	Ana Costa	Toyota	Truck	Camry	2022-12-22T00...	D8A
Q6GEMVT2238...	Carlos Pereira	BMW	Truck	320i	2019-02-05T00...	12F
TEZZB3ZFWAQ...	Juliana Costa	Ford	Convertible	Fiesta	2021-10-23T00...	AED
UAYVMBVEFO...	João Oliveira	Honda	Sedan	CR-V	2018-02-27T00...	96E
UYKVKB45UT0...	Carlos Oliveira	Honda	SUV	HR-V	2022-06-20T00...	F6A
YBTAH4TGJLCX...	Maria Costa	Volkswagen	Hatchback	Tiguan	2021-06-22T00...	7EE

Figura 14- Node-Red Dados Data

Na Figura 14 temos a tabela gerada pelo Node-Red que representa todos os utilizadores que possuem um carro de data de fabrico superior a 2015.

## Vídeo:



## Conclusão:

O presente trabalho focou na análise e transformação de dados no setor automotivo, utilizando ferramentas como Pentaho Spoon e Node-RED. Através do Pentaho, foi possível realizar uma série de transformações que incluíram o cálculo da idade dos veículos, a formatação e a filtragem dos dados. O Spoon facilitou o processo de ETL (Extração, Transformação e Carga), permitindo a manipulação eficiente de grandes volumes de dados e a criação de relatórios detalhados.

Além disso, a integração com o Node-RED possibilitou a automação do fluxo de trabalho, contribuindo para uma gestão mais ágil e eficaz das informações. A experiência adquirida ao longo deste projeto demonstra a importância das ferramentas de integração e transformação de dados para a tomada de decisões informadas e para o aprimoramento das operações no setor automotivo.