



Trabalho Prático 1 – ETL

Integração de Sistemas de Informação

Trabalho realizado por:
Eduardo Gomes nº17006
LESI

Introdução

Foi nos proposto realizar um trabalho para a unidade curricular “Integração de Sistemas de Informação” com o foco no ETL - *Extract Transform Load* que consiste na extração de dados de fontes externas, transformação dos dados para atender às necessidades de negócio e carregamento dos dados.

O principal objetivo do trabalho proposto é explorar ferramentas que serão usados no ETL.

O trabalho foi realizado no programa *Pentaho Kettle*, o que oferece ao utilizador facilidade em extrair, transformar e carregar dados. No programa *Pentaho Kettle* foi escolhido extrair dados a partir de um ficheiro csv e tratá-los para que no final seja feito um output do ficheiro inicial em Json e XML para ser criado um html com uma stylesheet XSL.

Primeira Transformação

Segue-se a transformação feita no ficheiro csv no programa *Pentaho Kettle*.

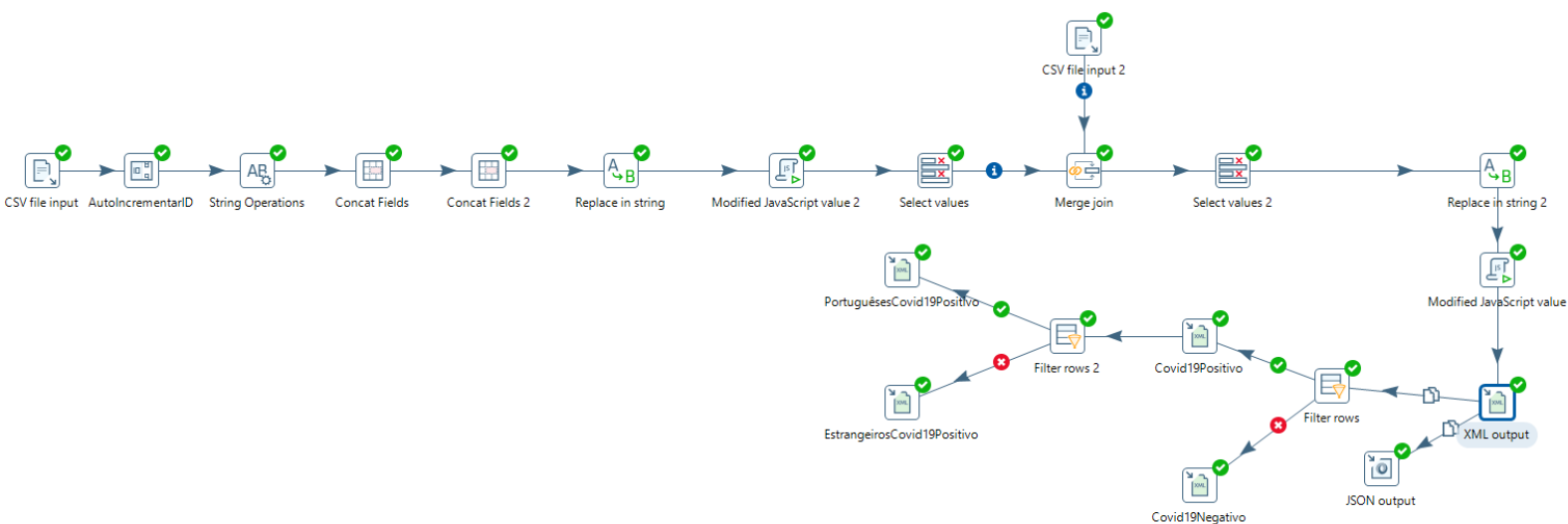


Figura 1 – O conjunto de transformações do ficheiro no Kettle (Aeroporto.ktr)

Extração de Dados

Primeiro foram extraídos dados de um ficheiro csv que representa uma lista de passageiros variados, que vão fazer viagens de avião. Cada passageiro tem os seus atributos, tais como o Nome, Sobrenome, Nacionalidade, Idade, Género, Número_Telemóvel, Companhia_Aérea, Id_Avião, Data, Hora, Destino e Preço.

Nome,Sobrenome,Nacionalidade,Idade,Género,Número_Telemóvel,Companhia_Aérea,Id_Avião,Data,Hora,Destino,Preço			
Miguel,FURTADO,português,23,m,914271152,TAP,1454,11/10/2020,15:30,Destino: Lisboa,35€			
david,Silva,Português,63,m,936399331,RYANAIR,1234,11/10/2020,10:20,Destino: Nièvre,150€			
FILIPA,moreira,PORTUGUÊS,73,f,9137312313,TAP,1421,11/10/2020,8:50,Destino: Nantes,143€			
rafael,martins,português,23,m,932748941,AIREUROPA,1212,11/10/2020,5:10,Destino: Berlim,123€			
Daniela,ferreira,Português,85,f,912923712,RYANAIR,1222,11/10/2020,11:10,Destino: Salzburg,153€			
Natália,Petejo,PORTUGUÊS,26,f,9134/88551,EASYJET AIRLINES,1242,11/10/2020,16:00,Destino: Madrid,96€			
VITOR,Gomes,Português,53,m,936511992,TAP,1111,11/10/2020,12:30,Destino: Texas,550€			
Leonardo,costa,português,23,m,9183+11654,AIREUROPA,1020,11/10/2020,11:00,Destino: Praga,340€			
laura,Cardoso,português,67,f,936523231,RYANAIR,1001,11/10/2020,17:15,Destino: Istambul,230€			
john,Morrison,Americano,34,m,194357425,AIREUROPA,1203,11/10/2020,3:10,Destino: Paris,154€			
SEAN,connery,inglês,65,m,917243823,TAP,1200,11/10/2020,9:50,Destino: Londres,135€			
sophie,Turner,Inglês,32,f,912322345,RYANAIR,1209,11/10/2020,4:40,Destino: Lisboa,53€			
ISABELLA,jones,inglês,19,f,916752832,EASYJET AIRLINES,1642,11/10/2020,21:20,Destino: Londres,145€			
George,Mcfly,AMERICANO,75,m,192325456,EASYJET AIRLINES,1931,11/11/2020,23:50,Destino: Ohio,650€			
José,mourinho,português,39,m,93zx3345721,RYANAIR,1099,11/11/2020,2:00,Destino: Lisboa,57€			
Raquel,gomes,PORTUGUÊS,78,f,919572736,EASYJET AIRLINES,1365,11/11/2020,5:10,Destino: Roma,234€			
Thiago,BATISTA,brazileiro,53,m,191241312,AZUL LINHAS AEREAS,1964,11/11/2020,15:50,Destino: Rio de Janeiro,451€			
Bernadette,Laurent,FRANÇÊS,64,f,913141124,TAP,1421,11/11/2020,13:20,Destino: Aveyron,132€			
Steve,AUSTIN,Americano,63,m,1931kk43245,AIREUROPA,1191,11/11/2020,15:05,Destino: Texas,860€			
Francisco,carvalho,português,23,m,938483848,RYANAIR,1865,11/11/2020,10:35,Destino: Lisboa,45€			
Charlotte,Flair,inglês,45,f,198483424,UNITED AIRLINES,1976,11/11/2020,5:00,Destino: Londres,130€			
CLAIRE,durand,FRANÇÊS,40,f,915255834,TAP,1921,11/11/2020,13:00,Destino: Paris,156€			
Eduardo,gomes,português,62,m,939483849,UNITED AIRLINES,1543,11/11/2020,16:10,Destino: Lisboa,54€			
Fábio,PINHEIRO,Português,36,m,91(8)388332,RYANAIR,1910,11/11/2020,19:40,Destino: Lisboa,49€			
JOE,miller,americano,67,m,193243521,AIREUROPA,1899,11/11/2020,11:30,Destino: Nova Iorque,550€			
charlie,jackson,inglês,50,m,9142x31314,TAP,1266,11/11/2020,2:25,Destino: Londres,137€			
Francine,DUBOIS,francês,23,f,953499421,TAP,1129,11/12/2020,19:50,Destino: Nantes,140€			
jacqueline,Simon,FRANÇÊS,27,f,953274231,RYANAIR,1875,11/12/2020,17:00,Destino: Paris,143€			
Bruno,pereira,PORTUGUÊS,34,m,938[24]3242,AIREUROPA,1900,11/12/2020,13:10,Destino: Barcelona,120€			
OLIVER,clark,Americano,61,m,194763997,UNITED AIRLINES,1676,11/12/2020,6:00,Destino: Geórgia,630€			
luna,DIAS,brazileiro,30,f,1921341115,TAP,1500,11/12/2020,1:50,Destino: Brasília,510€			
PEDRO,fernandes,português,54,m,934853688,RYANAIR,1061,11/12/2020,18:15,Destino: Lisboa,60€			
Josephine,MICHEL,FRANÇÊS,52,f,914321111,AIREUROPA,1953,11/12/2020,19:20,Destino: Paris,145€			
LUÍS,silva,português,19,m,9383pof82114,TAP,1972,11/12/2020,11:50,Destino: Bruxelas,160€			
Gabriela,LIMA,Brazileiro,24,f,192004313,AZUL LINHAS AEREAS,1992,11/12/2020,3:45,Destino: Rio de Janeiro,600€			
duarte,machado,PORTUGUÊS,33,m,930909380,AIREUROPA,1573,11/12/2020,9:20,Destino: Madrid,160€			
RANDY,brown,americano,66,m,(19062)2342,UNITED AIRLINES,1394,11/12/2020,16:40,Destino: Washington,645€			
João,santos,português,32,m,918304149,TAP,1652,11/12/2020,10:00,Destino: Barcelona,154€			
RITA,VIEIRA,Português,20,f,914422643,AIREUROPA,1249,11/12/2020,13:30,Destino: Madrid,147€			

Figura 2 - Dados originais no ficheiro csv

Como se pode observar, os dados precisam de ser tratados, tais como os nomes, sobrenomes, nacionalidades que apresentam só letras minúsculas, ou maiúsculas ou até mesmo a mistura dos dois e os números de telemóvel apresentam caracteres que não deviam pertencer ao número de telemóvel.

Step name: CSV file input

Filename: \${Internal.Entry.Current.Directory}/Input/Aeroporto.csv

Delimiter: ,

Enclosure: "

NIO buffer size: 50000

Lazy conversion? ☒

Header row present? ☒

Add filename to result ☐

The row number field name (optional):

Running in parallel? ☐

New line possible in fields? ☐

Format: mixed

File encoding:

Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
Nome	String							none
Sobrenome	String							none
Nacionalidade	String							none
Idade	Integer							none
Género	String							none
Número_Telemóvel	String							none
Companhia_Aérea	String							none
Id_Avião	Integer							none
Data	String							none
Hora	String							none
Destino	String							none
Preço	String							none

Figura 3 - Extração de dados em csv usando o Kettle

Para que o programa kettle fosse capaz de extrair os dados do ficheiro csv foi necessário inserir todos os atributos e que tipo de atributo era, e usou-se como a “,” para delimitar as colunas

Rows of step: CSV file input (39 rows)

#	Nome	Sobrenome	Nacionalidade	Idade	Género	Número_Telemóvel	Companhia_Aérea	Id_Avião	Data	Hora	Destino	Preço
1	Miguel	FURTADO	português	23	m	914271152	TAP	1454	11/10/2020	15:30	Destino: Lisboa	35€
2	david	Silva	Português	63	m	936399331	RYANAIR	1234	11/10/2020	10:20	Destino: Nièvre	150€
3	FILIPA	moreira	PORTUGUÊS	73	f	9137312313	TAP	1421	11/10/2020	8:50	Destino: Nantes	143€
4	rafael	martins	português	23	m	932748941	AIREUROPA	1212	11/10/2020	5:10	Destino: Berlim	123€
5	Daniela	ferreira	Português	85	f	912923712	RYANAIR	1222	11/10/2020	11:10	Destino: Salzburg	153€
6	Natália	Petejo	PORTUGUÊS	26	f	9134/88551	EASYJET AIRLINES	1242	11/10/2020	16:00	Destino: Madrid	96€
7	VITOR	Gomes	Português	53	m	936511992	TAP	1111	11/10/2020	12:30	Destino: Texas	550€
8	Leonardo	costa	português	23	m	9183-11654	AIREUROPA	1020	11/10/2020	11:00	Destino: Praga	340€
9	laura	Cardoso	português	67	f	936523231	RYANAIR	1001	11/10/2020	17:15	Destino: Istambul	230€
10	john	Morrison	Americano	34	m	194357425	AIREUROPA	1203	11/10/2020	3:10	Destino: Paris	154€
11	SEAN	connery	inglês	65	m	917243823	TAP	1200	11/10/2020	9:50	Destino: Londres	135€
12	sophie	Turner	Inglês	32	f	912322345	RYANAIR	1209	11/10/2020	4:40	Destino: Lisboa	53€
13	ISABELLA	jones	inglês	19	f	916752832	EASYJET AIRLINES	1642	11/10/2020	21:20	Destino: Londres	145€
14	George	Mcfly	AMERICANO	75	m	192325456	EASYJET AIRLINES	1931	11/11/2020	23:50	Destino: Ohio	650€
15	José	mourinho	português	39	m	93zx3345721	RYANAIR	1099	11/11/2020	2:00	Destino: Lisboa	57€
16	Raquel	gomes	PORTUGUÊS	78	f	919572736	EASYJET AIRLINES	1365	11/11/2020	5:10	Destino: Roma	234€
17	Thiago	BATISTA	brazileiro	53	m	191241312	AZUL LINHAS AEREAS	1964	11/11/2020	15:50	Destino: Rio de Janeiro	451€
18	Bernadette	Laurent	FRANÇAIS	64	f	913141124	TAP	1421	11/11/2020	13:20	Destino: Aveyron	132€
19	Steve	AUSTIN	Americano	63	m	1931kk43245	AIREUROPA	1191	11/11/2020	15:05	Destino: Texas	860€
20	Francisco	carvalho	português	23	m	938483848	RYANAIR	1865	11/11/2020	10:35	Destino: Lisboa	45€
21	Charlotte	Flair	inglês	45	f	198483424	UNITED AIRLINES	1976	11/11/2020	5:00	Destino: Londres	130€
22	CLAIRE	durand	FRANÇAIS	40	f	915255834	TAP	1921	11/11/2020	13:00	Destino: Paris	156€
23	Eduardo	gomes	português	62	m	939483849	UNITED AIRLINES	1543	11/11/2020	16:10	Destino: Lisboa	54€
24	Fábio	PINHEIRO	Português	36	m	91(8)388332	RYANAIR	1910	11/11/2020	19:40	Destino: Lisboa	49€
25	JOE	millier	americano	67	m	193243521	AIREUROPA	1899	11/11/2020	11:30	Destino: Nova Iorque	550€
26	charlie	jackson	inglês	50	m	9142x31314	TAP	1266	11/11/2020	2:25	Destino: Londres	137€
27	Francine	DUBOIS	francês	23	f	953499421	TAP	1129	11/12/2020	19:50	Destino: Nantes	140€
28	jacqueline	Simon	FRANÇAIS	27	f	953274231	RYANAIR	1875	11/12/2020	17:00	Destino: Paris	143€
29	Bruno	pereira	PORTUGUÊS	34	m	938(24)3242	AIREUROPA	1900	11/12/2020	13:10	Destino: Barcelona	120€
30	OLIVER	clark	Americano	61	m	194763997	UNITED AIRLINES	1676	11/12/2020	6:00	Destino: Geórgia	630€
31	luna	DIAS	brazileiro	30	f	1921341115	TAP	1500	11/12/2020	1:50	Destino: Brasília	510€
32	PEDRO	fernandes	português	54	m	934853688	RYANAIR	1061	11/12/2020	18:15	Destino: Lisboa	60€
33	Josephine	MICHEL	FRANÇAIS	52	f	914321111	AIREUROPA	1953	11/12/2020	19:20	Destino: Paris	145€
34	LUIS	silva	português	19	m	9383prof82114	TAP	1972	11/12/2020	11:50	Destino: Bruxelas	160€
35	Gabriela	LIMA	Brazileiro	24	f	192004313	AZUL LINHAS AEREAS	1992	11/12/2020	3:45	Destino: Rio de Janeiro	600€
36	duarte	machado	PORTUGUÊS	33	m	930909380	AIREUROPA	1573	11/12/2020	9:20	Destino: Madrid	160€
37	RANDY	brown	americano	66	m	(19062)2342	UNITED AIRLINES	1394	11/12/2020	16:40	Destino: Washington	645€
38	João	santos	português	32	m	918304149	TAP	1652	11/12/2020	10:00	Destino: Barcelona	154€
39	RITA	VIEIRA	Português	20	f	914422643	AIREUROPA	1249	11/12/2020	13:30	Destino: Madrid	147€

Figura 4 - Resultados da execução de csv input no Kettle

Transformação de Dados

Auto Incrementar coluna ID_Passageiro

De seguida foi adicionada uma nova coluna ao ficheiro csv chamada “ID_Passageiro” com o objetivo de atribuir a cada passageiro o seu próprio ID.

Nome do step: AutoIncrementarID

Nome do valor: ID_Passageiro

Use a database to generate the sequence

Usa BD para obter a sequência? ☐

Connection: [dropdown] [Edit...] [New...] [Wizard...]

Schema name: [text] [Schemas...]

Nome da sequência: SEQ_ [Sequences...]

Use a transformation counter to generate the sequence

Usa o contador para calcular a sequência? ☒

Counter name (optional): [text]

Inicia no valor: 1

Incremento de: 1

Valor máximo: 999999999

[? Help] [OK] [Cancela]

Figura 5 – Inserção da coluna “ID_Passageiro”

Na primeira linha da coluna “ID_Passageiro” foi atribuído o valor “1”, em que vai incrementando um valor até chegar à última linha que corresponde ao último passageiro.

#	Nome	Sobrenome	Nacionalidade	Idade	Gênero	Número_Telemóvel	Companhia_Aérea	Id_Avião	Data	Hora	Destino	Preço	ID_Passageiro
1	Miguel	FURTADO	português	23	m	914271152	TAP	1454	11/10/2020	15:30	Destino: Lisboa	35€	1
2	david	Silva	Português	63	m	936399331	RYANAIR	1234	11/10/2020	10:20	Destino: Nièvre	150€	2
3	FILIPA	moreira	PORTUGUÊS	73	f	9133312313	TAP	1421	11/10/2020	8:50	Destino: Nantes	143€	3
4	rafael	martins	português	23	m	932748941	AIREUROPA	1212	11/10/2020	5:10	Destino: Berlim	123€	4
5	Daniela	ferreira	Português	85	f	912923712	RYANAIR	1222	11/10/2020	11:10	Destino: Salzburg	153€	5
6	Natália	Petejo	PORTUGUÊS	26	f	9134/88551	EASYJET AIRLINES	1242	11/10/2020	16:00	Destino: Madrid	96€	6
7	VITOR	Gomes	Português	53	m	936511992	TAP	1111	11/10/2020	12:30	Destino: Texas	550€	7
8	Leonardo	costa	português	23	m	9183+11654	AIREUROPA	1020	11/10/2020	11:00	Destino: Praga	340€	8
9	laura	Cardoso	português	67	f	936523231	RYANAIR	1001	11/10/2020	17:15	Destino: Istambul	230€	9
10	john	Morrison	Americano	34	m	194357425	AIREUROPA	1203	11/10/2020	3:10	Destino: Paris	154€	10
11	SEAN	connery	inglês	65	m	917243823	TAP	1200	11/10/2020	9:50	Destino: Londres	135€	11
12	sophie	Turner	Inglês	32	f	912322345	RYANAIR	1209	11/10/2020	4:40	Destino: Lisboa	53€	12
13	ISABELLA	jones	inglês	19	f	916752832	EASYJET AIRLINES	1642	11/10/2020	21:20	Destino: Londres	145€	13
14	George	Mcfly	AMERICANO	75	m	192325456	EASYJET AIRLINES	1931	11/11/2020	23:50	Destino: Ohio	650€	14
15	José	mourinho	português	39	m	93zx3345721	RYANAIR	1099	11/11/2020	2:00	Destino: Lisboa	57€	15
16	Raquel	gomes	PORTUGUÊS	78	f	919572736	EASYJET AIRLINES	1365	11/11/2020	5:10	Destino: Roma	234€	16
17	Thiago	BATISTA	brazileiro	53	m	191241312	AZUL LINHAS AEREAS	1964	11/11/2020	15:50	Destino: Rio de Janeiro	451€	17
18	Bernadette	Laurent	FRANÇAIS	64	f	913141124	TAP	1421	11/11/2020	13:20	Destino: Aveyron	132€	18
19	Steve	AUSTIN	Americano	63	m	1931kk43245	AIREUROPA	1191	11/11/2020	15:05	Destino: Texas	860€	19
20	Francisco	carvalho	português	23	m	938483848	RYANAIR	1865	11/11/2020	10:35	Destino: Lisboa	45€	20
21	Charlotte	Flair	inglês	45	f	198483424	UNITED AIRLINES	1976	11/11/2020	5:00	Destino: Londres	130€	21
22	CLAIRE	durand	FRANÇAIS	40	f	915255834	TAP	1921	11/11/2020	13:00	Destino: Paris	156€	22
23	Eduardo	gomes	português	62	m	939483849	UNITED AIRLINES	1543	11/11/2020	16:10	Destino: Lisboa	54€	23
24	Fábio	PINHEIRO	Português	36	m	91(8)388332	RYANAIR	1910	11/11/2020	19:40	Destino: Lisboa	49€	24
25	JOE	milller	americano	67	m	193243521	AIREUROPA	1899	11/11/2020	11:30	Destino: Nova Iorque	550€	25
26	charlie	jackson	inglês	50	m	9142x31314	TAP	1266	11/11/2020	2:25	Destino: Londres	137€	26
27	Francine	DUBOIS	français	23	f	953499421	TAP	1129	11/12/2020	19:50	Destino: Nantes	140€	27
28	jacqueline	Simon	FRANÇAIS	27	f	953274231	RYANAIR	1875	11/12/2020	17:00	Destino: Paris	143€	28
29	Bruno	pereira	PORTUGUÊS	34	m	938[24]3242	AIREUROPA	1900	11/12/2020	13:10	Destino: Barcelona	120€	29
30	OLIVER	clark	Americano	61	m	194763997	UNITED AIRLINES	1676	11/12/2020	6:00	Destino: Geórgia	630€	30
31	luna	DIAS	brazileiro	30	f	1921341115	TAP	1500	11/12/2020	1:50	Destino: Brasília	510€	31
32	PEDRO	fernandes	português	54	m	934853688	RYANAIR	1061	11/12/2020	18:15	Destino: Lisboa	60€	32
33	Josephine	MICHEL	FRANÇAIS	52	f	914321111	AIREUROPA	1953	11/12/2020	19:20	Destino: Paris	145€	33
34	LUI\$	silva	português	19	m	9383pof82114	TAP	1972	11/12/2020	11:50	Destino: Bruxelas	160€	34
35	Gabriela	LIMA	Brazileiro	24	f	192004313	AZUL LINHAS AEREAS	1992	11/12/2020	3:45	Destino: Rio de Janeiro	600€	35
36	duarte	machado	PORTUGUÊS	33	m	930909380	AIREUROPA	1573	11/12/2020	9:20	Destino: Madrid	160€	36
37	RANDY	brown	americano	66	m	(19062)2342	UNITED AIRLINES	1394	11/12/2020	16:40	Destino: Washington	645€	37
38	João	santos	português	32	m	918304149	TAP	1652	11/12/2020	10:00	Destino: Barcelona	154€	38
39	RITA	VIEIRA	Português	20	f	914422643	AIREUROPA	1249	11/12/2020	13:30	Destino: Madrid	147€	39

Figura 6 –Coluna “ID_Passageiro” inserida

Formatação de strings

A formatação das strings tem como objetivo que todos os dados estejam guardados de forma consistente.

Step name
String Operations

The fields to process:

#	In stream field	Out stream field	Trim type	Lower/Upper	Padding	Pad char	Pad Length	InitCap	Escape	Digits	Remove Special character
1	Nome		both	lower	none			S	None	remove	none
2	Sobrenome		both	lower	none			S	None	remove	none
3	Número_Telemóvel		both	lower	none			N	None	only	none
4	Nacionalidade		both	lower	none			S	None	remove	none

Help
OK
Get fields
Cancela

Figura 7 –Operações com Strings

Começamos por fazer um trim both, isto é remover espaços ao início e fim dos dados. Todos os campos com letras são transformados em lowercase e de modo a que os nomes e cidades comecem com maiúscula foi ativada a opção InitCap. Também foi selecionada apenas permitir dígitos nos números de telefone e a opção de remover qualquer dígito que possa existir em nomes ou cidades.

Junção de colunas

De seguida foi feita a junção de colunas entre o “Nome”, “Sobrenome” e “Data”, “Hora” numa só coluna, “Nome Completo” e “Data_Hora” respetivamente. Na junção foi usada um separador que foi o “espaço” para ser identificável o nome, o sobrenome, a data e a hora.

Nome do Step: **Concat Fields**

Target Field Name: **Nome_Completo**

Length of Target Field: **0**

Separator: Insert TAB

Enclosure:

Fields | Advanced

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim Type	Nu
1	Nome	String							both	
2	Sobrenome	String							both	

Figura 8 –Junção das colunas “Nome” e “Sobrenome” numa só chamada “Nome_Completo”

Nome do Step: **Concat Fields 2**

Target Field Name: **Data_Hora**

Length of Target Field: **0**

Separator: Insert TAB

Enclosure:

Fields | Advanced

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim Type	Nu
1	Data	String							both	
2	Hora	String							both	

Figura 9 –Junção das colunas “Data” e “Hora” numa só chamada “Data_Hora”

#	Nome	Sobrenome	Nacionalidade	Idade	Género	Número_Telemóvel	Companhia_Aérea	Id_Avião	Data	Hora	Destino	Preço	ID_Passageiro	Nome_Completo	Data_Hora
1	Miguel	Furtado	Português	23	m	914271152	TAP	1454	11/10/2020	15:30	Destino: Lisboa	35€	1	Miguel Furtado	11/10/2020 15:30
2	David	Silva	Português	63	m	936399331	RYANAIR	1234	11/10/2020	10:20	Destino: Nièvre	150€	2	David Silva	11/10/2020 10:20
3	Filipa	Moreira	Português	73	f	913312313	TAP	1421	11/10/2020	8:50	Destino: Nantes	143€	3	Filipa Moreira	11/10/2020 8:50
4	Rafael	Martins	Português	23	m	932748941	AIREUROPA	1212	11/10/2020	5:10	Destino: Berlim	123€	4	Rafael Martins	11/10/2020 5:10
5	Daniela	Ferreira	Português	85	f	912923712	RYANAIR	1222	11/10/2020	11:10	Destino: Salzburq	153€	5	Daniela Ferreira	11/10/2020 11:10
6	Natália	Petejo	Português	26	f	913488551	EASVJET AIRLINES	1242	11/10/2020	16:00	Destino: Madrid	96€	6	Natália Petejo	11/10/2020 16:00
7	Vitor	Gomes	Português	53	m	936511992	TAP	1111	11/10/2020	12:30	Destino: Texas	550€	7	Vitor Gomes	11/10/2020 12:30
8	Leonardo	Costa	Português	23	m	918311654	AIREUROPA	1020	11/10/2020	11:00	Destino: Praga	340€	8	Leonardo Costa	11/10/2020 11:00
9	Laura	Cardoso	Português	67	f	936523231	RYANAIR	1001	11/10/2020	17:15	Destino: Istambul	230€	9	Laura Cardoso	11/10/2020 17:15
10	John	Morrison	Americano	34	m	194357425	AIREUROPA	1203	11/10/2020	3:10	Destino: Paris	154€	10	John Morrison	11/10/2020 3:10
11	Sean	Connery	Inglês	65	m	917243823	TAP	1200	11/10/2020	9:50	Destino: Londres	135€	11	Sean Connery	11/10/2020 9:50
12	Sophie	Turner	Inglês	32	f	912322345	RYANAIR	1209	11/10/2020	4:40	Destino: Lisboa	53€	12	Sophie Turner	11/10/2020 4:40
13	Isabella	Jones	Inglês	19	f	916752832	EASVJET AIRLINES	1642	11/10/2020	21:20	Destino: Londres	145€	13	Isabella Jones	11/10/2020 21:20
14	George	Mcfly	Americano	75	m	192325456	EASVJET AIRLINES	1931	11/11/2020	23:50	Destino: Ohio	650€	14	George Mcfly	11/11/2020 23:50
15	José	Mourinho	Português	39	m	933345721	RYANAIR	1099	11/11/2020	2:00	Destino: Lisboa	57€	15	José Mourinho	11/11/2020 2:00
16	Raquel	Gomes	Português	78	f	919572736	EASVJET AIRLINES	1365	11/11/2020	5:10	Destino: Roma	234€	16	Raquel Gomes	11/11/2020 5:10
17	Thiago	Batista	Brazileiro	53	m	191241312	AZUL LINHAS AEREAS	1964	11/11/2020	15:50	Destino: Rio de Janeiro	451€	17	Thiago Batista	11/11/2020 15:50
18	Bernadette	Laurent	Françês	64	f	913141124	TAP	1421	11/11/2020	13:20	Destino: Aveyron	132€	18	Bernadette Laurent	11/11/2020 13:20
19	Steve	Austin	Americano	63	m	193143245	AIREUROPA	1191	11/11/2020	15:05	Destino: Texas	860€	19	Steve Austin	11/11/2020 15:05
20	Francisco	Carvalho	Português	23	m	938483848	RYANAIR	1865	11/11/2020	10:35	Destino: Lisboa	45€	20	Francisco Carvalho	11/11/2020 10:35
21	Charlotte	Flair	Inglês	45	f	198483424	UNITED AIRLINES	1976	11/11/2020	5:00	Destino: Londres	130€	21	Charlotte Flair	11/11/2020 5:00
22	Claire	Durand	Françês	40	f	915255834	TAP	1921	11/11/2020	13:00	Destino: Paris	156€	22	Claire Durand	11/11/2020 13:00
23	Eduardo	Gomes	Português	62	m	939483849	UNITED AIRLINES	1543	11/11/2020	16:10	Destino: Lisboa	54€	23	Eduardo Gomes	11/11/2020 16:10
24	Fábio	Pinheiro	Português	36	m	918388332	RYANAIR	1910	11/11/2020	19:40	Destino: Lisboa	49€	24	Fábio Pinheiro	11/11/2020 19:40
25	Joe	Miller	Americano	67	m	193243521	AIREUROPA	1899	11/11/2020	11:30	Destino: Nova Iorque	550€	25	Joe Miller	11/11/2020 11:30
26	Charlie	Jackson	Inglês	50	m	914231314	TAP	1266	11/11/2020	2:25	Destino: Londres	137€	26	Charlie Jackson	11/11/2020 2:25
27	Francine	Dubois	Françês	23	f	953499421	TAP	1129	11/12/2020	19:50	Destino: Nantes	140€	27	Francine Dubois	11/12/2020 19:50
28	Jacqueline	Simon	Françês	27	f	953274231	RYANAIR	1875	11/12/2020	17:00	Destino: Paris	143€	28	Jacqueline Simon	11/12/2020 17:00
29	Bruno	Pereira	Português	34	m	938243242	AIREUROPA	1900	11/12/2020	13:10	Destino: Barcelona	120€	29	Bruno Pereira	11/12/2020 13:10
30	Oliver	Clark	Americano	61	m	194763997	UNITED AIRLINES	1676	11/12/2020	6:00	Destino: Geórgia	630€	30	Oliver Clark	11/12/2020 6:00
31	Luna	Dias	Brazileiro	30	f	192341115	TAP	1500	11/12/2020	1:50	Destino: Brasília	510€	31	Luna Dias	11/12/2020 1:50
32	Pedro	Fernandes	Português	54	m	934853688	RYANAIR	1061	11/12/2020	18:15	Destino: Lisboa	60€	32	Pedro Fernandes	11/12/2020 18:15
33	Josephine	Michel	Françês	52	f	914321111	AIREUROPA	1953	11/12/2020	19:20	Destino: Paris	145€	33	Josephine Michel	11/12/2020 19:20
34	Luís	Silva	Português	19	m	938382114	TAP	1972	11/12/2020	11:50	Destino: Bruxelas	160€	34	Luís Silva	11/12/2020 11:50
35	Gabriela	Lima	Brazileiro	24	f	192004313	AZUL LINHAS AEREAS	1992	11/12/2020	3:45	Destino: Rio de Janeiro	600€	35	Gabriela Lima	11/12/2020 3:45
36	Duarte	Machado	Português	33	m	930909380	AIREUROPA	1573	11/12/2020	9:20	Destino: Madrid	160€	36	Duarte Machado	11/12/2020 9:20
37	Randy	Brown	Americano	66	m	190622342	UNITED AIRLINES	1394	11/12/2020	16:40	Destino: Washington	645€	37	Randy Brown	11/12/2020 16:40
38	João	Santos	Português	32	m	918304149	TAP	1652	11/12/2020	10:00	Destino: Barcelona	154€	38	João Santos	11/12/2020 10:00
39	Rita	Vieira	Português	20	f	914422643	AIREUROPA	1249	11/12/2020	13:30	Destino: Madrid	147€	39	Rita Vieira	11/12/2020 13:30

Figura 10 –Criadas as novas colunas e as strings já se apresentam normais

Substituir Strings

De seguida, foi decidio substituir as strings que estão ligadas ao atributo “Género” em que a letra “m” irá ser substituído pela palavra “masculino” e “f” substituído por “feminino”.

Step name

Fields string

#	In stream field	Out stream field	use RegEx	Search	Replace with	Set empty string?	Replace with field	Whole Word	Case sensitive	Is Unicode
1	Género		S	f	Feminio	N		S	N	N
2	Género		S	m	Maculino	N		S	N	N

Figura 11 –Substituição de Strings

Criar uma nova coluna com “Modified JavaScript value”

De seguida foi criada uma nova coluna ao ficheiro csv usando o “Modified JavaScript value”, uma coluna chamada “Email”. Sabendo que cada “Email” é constituído pelo primeiro e último nome de cada passageiro, foi usado o “Modified JavaScript value” para juntar o primeiro e último nome mas sem espaço entre os dois que foi usado na junção de colunas anteriormente.

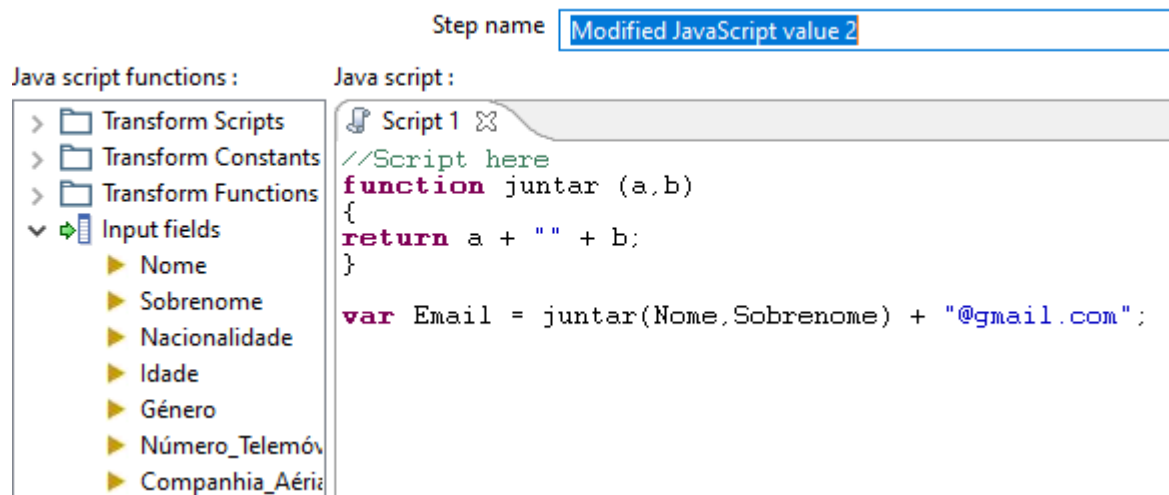


Figura 12 –Criação da coluna “Email” pelo “Modified JavaScript value”

Como se pode verificar na figura 12, foi criada uma função “juntar” que tem como objetivo juntar as variáveis selecionadas. Foi criada uma variável chamada “Email” que será criada graças à função “juntar”, que irá juntar o “Nome” e o “Sobrenome” de cada passageiro e depois será adicionado “@gmail.com”.

Reorganização dos dados

Recorrendo à opção “Select values” vai ser reorganizado a ordem das colunas e mudar o tipo da variável “Data_Hora” de string para date.

Step name: Select values

Select & Alter Remove Meta-data

Fields:

#	Fieldname	Rename to	Length	Precision
1	ID_Passageiro			
2	Nome_Completo			
3	Nacionalidade			
4	Idade			
5	Género			
6	Número_Telemóvel			
7	Email			
8	Companhia_Aérea			
9	Id_Avião			
10	Data_Hora			
11	Destino			
12	Preço			

Figura 13 –Reorganização da ordem das colunas

Step name

Select values

Select & Alter

Remove

Meta-data

Fields to alter the meta-data for :

#	Fieldname	Rename to	Type	L
1	Data_Hora		Date	

Figura 14 –Mudar o tipo da variável “Data_Hora” de string para date

#	ID_Passageiro	Nome_Completo	Nacionalidade	Idade	Gênero	Número_Telemóvel	Email	Companhia_Aérea	Id_Avião	Data_Hora	Destino	Preço
1	1	Miguel Furtado	Português	23	Maculino	914271152	MiguelFurtado@gmail.com	TAP	1454	11/10/2020 15:30	Destino: Lisboa	35€
2	2	David Silva	Português	63	Maculino	936399331	DavidSilva@gmail.com	RYANAIR	1234	11/10/2020 10:20	Destino: Nièvre	150€
3	3	Filipa Moreira	Português	73	Feminio	913312313	FilipaMoreira@gmail.com	TAP	1421	11/10/2020 08:50	Destino: Nantes	143€
4	4	Rafael Martins	Português	23	Maculino	932748941	RafaelMartins@gmail.com	AIREUROPA	1212	11/10/2020 05:10	Destino: Berlim	123€
5	5	Daniela Ferreira	Português	85	Feminio	912923712	DanielaFerreira@gmail.com	RYANAIR	1222	11/10/2020 11:10	Destino: Salzburg	153€
6	6	Natália Petejo	Português	26	Feminio	913488551	NatáliaPetejo@gmail.com	EASYJET AIRLINES	1242	11/10/2020 16:00	Destino: Madrid	96€
7	7	Vitor Gomes	Português	53	Maculino	936511992	VitorGomes@gmail.com	TAP	1111	11/10/2020 12:30	Destino: Texas	550€
8	8	Leonardo Costa	Português	23	Maculino	918311654	LeonardoCosta@gmail.com	AIREUROPA	1020	11/10/2020 11:00	Destino: Praga	340€
9	9	Laura Cardoso	Português	67	Feminio	936523231	LauraCardoso@gmail.com	RYANAIR	1001	11/10/2020 17:15	Destino: Istambul	230€
10	10	John Morrison	Americano	34	Maculino	194357425	JohnMorrison@gmail.com	AIREUROPA	1203	11/10/2020 03:10	Destino: Paris	154€
11	11	Sean Connery	Inglês	65	Maculino	917243823	SeanConnery@gmail.com	TAP	1200	11/10/2020 09:50	Destino: Londres	135€
12	12	Sophie Turner	Inglês	32	Feminio	912322345	SophieTurner@gmail.com	RYANAIR	1209	11/10/2020 04:40	Destino: Lisboa	53€
13	13	Isabella Jones	Inglês	19	Feminio	916752832	IsabellaJones@gmail.com	EASYJET AIRLINES	1642	11/10/2020 21:20	Destino: Londres	145€
14	14	George Mcfly	Americano	75	Maculino	192325456	GeorgeMcfly@gmail.com	EASYJET AIRLINES	1931	11/11/2020 23:50	Destino: Ohio	650€
15	15	José Mourinho	Português	39	Maculino	933345721	JoséMourinho@gmail.com	RYANAIR	1099	11/11/2020 02:00	Destino: Lisboa	57€
16	16	Raquel Gomes	Português	78	Feminio	919572736	RaquelGomes@gmail.com	EASYJET AIRLINES	1365	11/11/2020 05:10	Destino: Roma	234€
17	17	Thiago Batista	Brazileiro	53	Maculino	191241312	ThiagoBatista@gmail.com	AZUL LINHAS AEREAS	1964	11/11/2020 15:50	Destino: Rio de Janeiro	451€
18	18	Bernadette Laurent	Françês	64	Feminio	913141124	BernadetteLaurent@gmail.com	TAP	1421	11/11/2020 13:20	Destino: Aveyron	132€
19	19	Steve Austin	Americano	63	Maculino	193143245	SteveAustin@gmail.com	AIREUROPA	1191	11/11/2020 15:05	Destino: Texas	860€
20	20	Francisco Carvalho	Português	23	Maculino	938483848	FranciscoCarvalho@gmail.com	RYANAIR	1865	11/11/2020 10:35	Destino: Lisboa	45€
21	21	Charlotte Flair	Inglês	45	Feminio	198483424	CharlotteFlair@gmail.com	UNITED AIRLINES	1976	11/11/2020 05:00	Destino: Londres	130€
22	22	Claire Durand	Françês	40	Feminio	915255834	ClaireDurand@gmail.com	TAP	1921	11/11/2020 13:00	Destino: Paris	156€
23	23	Eduardo Gomes	Português	62	Maculino	939483849	EduardoGomes@gmail.com	UNITED AIRLINES	1543	11/11/2020 16:10	Destino: Lisboa	54€
24	24	Fábio Pinheiro	Português	36	Maculino	918388332	FábioPinheiro@gmail.com	RYANAIR	1910	11/11/2020 19:40	Destino: Lisboa	49€
25	25	Joe Miller	Americano	67	Maculino	193243521	JoeMiller@gmail.com	AIREUROPA	1899	11/11/2020 11:30	Destino: Nova Iorque	550€
26	26	Charlie Jackson	Inglês	50	Maculino	914231314	CharlieJackson@gmail.com	TAP	1266	11/11/2020 02:25	Destino: Londres	137€
27	27	Francine Dubois	Françês	23	Feminio	953499421	FrancineDubois@gmail.com	TAP	1129	11/12/2020 19:50	Destino: Nantes	140€
28	28	Jacqueline Simon	Françês	27	Feminio	953274231	JacquelineSimon@gmail.com	RYANAIR	1875	11/12/2020 17:00	Destino: Paris	143€
29	29	Bruno Pereira	Português	34	Maculino	938243242	BrunoPereira@gmail.com	AIREUROPA	1900	11/12/2020 13:10	Destino: Barcelona	120€
30	30	Oliver Clark	Americano	61	Maculino	194763997	OliverClark@gmail.com	UNITED AIRLINES	1676	11/12/2020 06:00	Destino: Geórgia	630€
31	31	Luna Dias	Brazileiro	30	Feminio	192341115	LunaDias@gmail.com	TAP	1500	11/12/2020 01:50	Destino: Brasília	510€
32	32	Pedro Fernandes	Português	54	Maculino	934853688	PedroFernandes@gmail.com	RYANAIR	1061	11/12/2020 18:15	Destino: Lisboa	60€
33	33	Josephine Michel	Françês	52	Feminio	914321111	JosephineMichel@gmail.com	AIREUROPA	1953	11/12/2020 19:20	Destino: Paris	145€
34	34	Luís Silva	Português	19	Maculino	938382114	LuísSilva@gmail.com	TAP	1972	11/12/2020 11:50	Destino: Bruxelas	160€
35	35	Gabriela Lima	Brazileiro	24	Feminio	192004313	GabrielaLima@gmail.com	AZUL LINHAS AEREAS	1992	11/12/2020 03:45	Destino: Rio de Janeiro	600€
36	36	Duarte Machado	Português	33	Maculino	930909380	DuarteMachado@gmail.com	AIREUROPA	1573	11/12/2020 09:20	Destino: Madrid	160€
37	37	Randy Brown	Americano	66	Maculino	190622342	RandyBrown@gmail.com	UNITED AIRLINES	1394	11/12/2020 16:40	Destino: Washington	645€
38	38	João Santos	Português	32	Maculino	918304149	JoãoSantos@gmail.com	TAP	1652	11/12/2020 10:00	Destino: Barcelona	154€
39	39	Rita Vieira	Português	20	Feminio	914422643	RitaVieira@gmail.com	AIREUROPA	1249	11/12/2020 13:30	Destino: Madrid	147€

Figura 15 –Coluna organizada

Junção de dois ficheiros csv

De seguida, foi decidido juntar um ficheiro csv ao ficheiro que está a ser trabalhado. O ficheiro que vai ser adicionado contém duas colunas, “ID_Passageiro” e “Covid19”.

Rows of step: CSV file input 2 (39 rows)

#	ID_Passageiro	Covid19
1	1	N
2	2	N
3	3	N
4	4	N
5	5	P
6	6	N
7	7	N
8	8	P
9	9	N
10	10	N
11	11	N
12	12	N
13	13	P
14	14	N
15	15	N
16	16	P
17	17	N
18	18	N
19	19	N
20	20	N
21	21	P
22	22	P
23	23	P
24	24	N
25	25	N
26	26	N
27	27	N
28	28	N
29	29	P
30	30	N
31	31	P
32	32	P
33	33	P
34	34	N
35	35	N
36	36	N
37	37	N
38	38	P
39	39	N

Figura 16 –Novo ficheiro csv que confirma se um “ID_Passageiro” está infetado por Covid-19

Cada “ID_Passageiro” tem o seu atributo “Covid19” correspondente, em que confirma se um dado passageiro está infetado pelo vírus Covid-19 em que a letra “P” corresponde positivo e “N” corresponde negativo.

Para que a junção de dois ficheiros csv seja possível foi utilizado o “Merge Join” em que a coluna “Covid19” corresponda ao passageiro que possui o mesmo “ID_Passageiro”.

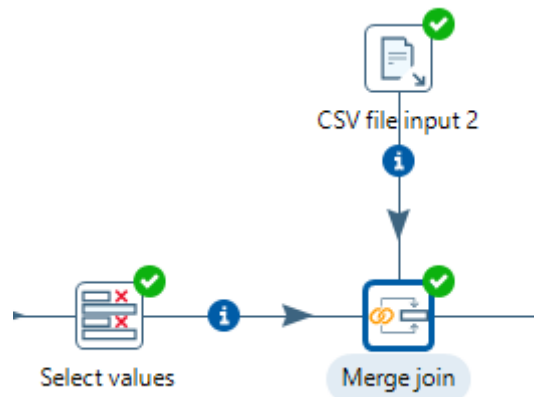


Figura 17 –Merge join entre “Select values”, em que foi feito a reorganização da ordem das colunas, e o novo ficheiro csv que confirma os casos de Covid-19(CSV file input 2)

Step name	Merge join		
First Step:	Select values		
Second Step:	CSV file input 2		
Join Type:	FULL OUTER		
Keys for 1st step:		Keys for 2nd step:	
#	Key field	#	Key field
1	ID_Passag...	1	ID_Passag...

Figura 18 –Merge join já referido na “Figura 16” em que o “ID_Passageiro” do ficheiro csv que está ser trabalhado irá corresponder ao “ID_Passageiro” do novo ficheiro csv

Depois de ser feita a junção, foi criada uma coluna “ID_Passageiro” a mais e com o “Select Values” essa mesma coluna foi eliminada.

#	ID_Passageiro	Nome_Completo	Nacionalidade	Idade	Gênero	Número_Telemóvel	Email	Companhia_Aérea	Id_Avião	Data_Hora	Destino	Preço	Covid19
1	1	Miguel Furtado	Português	23	Maculino	914271152	MiguelFurtado@gmail.com	TAP	1454	11/10/2020 15:30	Destino: Lisboa	35€	N
2	2	David Silva	Português	63	Maculino	936399331	DavidSilva@gmail.com	RYANAIR	1234	11/10/2020 10:20	Destino: Nièvre	150€	N
3	3	Filipa Moreira	Português	73	Feminio	913312313	FilipaMoreira@gmail.com	TAP	1421	11/10/2020 08:50	Destino: Nantes	143€	N
4	4	Rafael Martins	Português	23	Maculino	932748941	RafaelMartins@gmail.com	AIREUROPA	1212	11/10/2020 05:10	Destino: Berlim	123€	N
5	5	Daniela Ferreira	Português	85	Feminio	912923712	DanielaFerreira@gmail.com	RYANAIR	1222	11/10/2020 11:10	Destino: Salzburg	153€	P
6	6	Natália Petejo	Português	26	Feminio	913488551	NatáliaPetejo@gmail.com	EASYJET AIRLINES	1242	11/10/2020 16:00	Destino: Madrid	96€	N
7	7	Vitor Gomes	Português	53	Maculino	936511992	VitorGomes@gmail.com	TAP	1111	11/10/2020 12:30	Destino: Texas	550€	N
8	8	Leonardo Costa	Português	23	Maculino	918311654	LeonardoCosta@gmail.com	AIREUROPA	1020	11/10/2020 11:00	Destino: Praga	340€	P
9	9	Laura Cardoso	Português	67	Feminio	936523231	LauraCardoso@gmail.com	RYANAIR	1001	11/10/2020 17:15	Destino: Istambul	230€	N
10	10	John Morrison	Americano	34	Maculino	194357425	JohnMorrison@gmail.com	AIREUROPA	1203	11/10/2020 03:10	Destino: Paris	154€	N
11	11	Sean Connery	Inglês	65	Maculino	917243823	SeanConnery@gmail.com	TAP	1200	11/10/2020 09:50	Destino: Londres	135€	N
12	12	Sophie Turner	Inglês	32	Feminio	912322345	SophieTurner@gmail.com	RYANAIR	1209	11/10/2020 04:40	Destino: Lisboa	53€	N
13	13	Isabella Jones	Inglês	19	Feminio	916752832	IsabellaJones@gmail.com	EASYJET AIRLINES	1642	11/10/2020 21:20	Destino: Londres	145€	P
14	14	George Mcfly	Americano	75	Maculino	192325456	GeorgeMcfly@gmail.com	EASYJET AIRLINES	1931	11/11/2020 23:50	Destino: Ohio	650€	N
15	15	José Mourinho	Português	39	Maculino	933345721	JoséMourinho@gmail.com	RYANAIR	1099	11/11/2020 02:00	Destino: Lisboa	57€	N
16	16	Raquel Gomes	Português	78	Feminio	919572736	RaquelGomes@gmail.com	EASYJET AIRLINES	1365	11/11/2020 05:10	Destino: Roma	234€	P
17	17	Thiago Batista	Brazileiro	53	Maculino	191241312	ThiagoBatista@gmail.com	AZUL LINHAS AEREAS	1964	11/11/2020 15:50	Destino: Rio de Janeiro	451€	N
18	18	Bernadette Laurent	Françês	64	Feminio	913141124	BernadetteLaurent@gmail.com	TAP	1421	11/11/2020 13:20	Destino: Aveyron	132€	N
19	19	Steve Austin	Americano	63	Maculino	193143245	SteveAustin@gmail.com	AIREUROPA	1191	11/11/2020 15:05	Destino: Texas	860€	N
20	20	Francisco Carvalho	Português	23	Maculino	938483848	FranciscoCarvalho@gmail.com	RYANAIR	1865	11/11/2020 10:35	Destino: Lisboa	45€	N
21	21	Charlotte Flair	Inglês	45	Feminio	198483424	CharlotteFlair@gmail.com	UNITED AIRLINES	1976	11/11/2020 05:00	Destino: Londres	130€	P
22	22	Claire Durand	Françês	40	Feminio	915255834	ClaireDurand@gmail.com	TAP	1921	11/11/2020 13:00	Destino: Paris	156€	P
23	23	Eduardo Gomes	Português	62	Maculino	939483849	EduardoGomes@gmail.com	UNITED AIRLINES	1543	11/11/2020 16:10	Destino: Lisboa	54€	P
24	24	Fábio Pinheiro	Português	36	Maculino	918388332	FábioPinheiro@gmail.com	RYANAIR	1910	11/11/2020 19:40	Destino: Lisboa	49€	N
25	25	Joe Miller	Americano	67	Maculino	193243521	JoeMiller@gmail.com	AIREUROPA	1899	11/11/2020 11:30	Destino: Nova Iorque	550€	N
26	26	Charlie Jackson	Inglês	50	Maculino	914231314	CharlieJackson@gmail.com	TAP	1266	11/11/2020 02:25	Destino: Londres	137€	N
27	27	Francine Dubois	Françês	23	Feminio	953499421	FrancineDubois@gmail.com	TAP	1129	11/12/2020 19:50	Destino: Nantes	140€	N
28	28	Jacqueline Simon	Françês	27	Feminio	953274231	JacquelineSimon@gmail.com	RYANAIR	1875	11/12/2020 17:00	Destino: Paris	143€	N
29	29	Bruno Pereira	Português	34	Maculino	938243242	BrunoPereira@gmail.com	AIREUROPA	1900	11/12/2020 13:10	Destino: Barcelona	120€	P
30	30	Oliver Clark	Americano	61	Maculino	194763997	OliverClark@gmail.com	UNITED AIRLINES	1676	11/12/2020 06:00	Destino: Geórgia	630€	N
31	31	Luna Dias	Brazileiro	30	Feminio	192341115	LunaDias@gmail.com	TAP	1500	11/12/2020 01:50	Destino: Brasília	510€	P
32	32	Pedro Fernandes	Português	54	Maculino	934853688	PedroFernandes@gmail.com	RYANAIR	1061	11/12/2020 18:15	Destino: Lisboa	60€	P
33	33	Josephine Michel	Françês	52	Feminio	914321111	JosephineMichel@gmail.com	AIREUROPA	1953	11/12/2020 19:20	Destino: Paris	145€	P
34	34	Luís Silva	Português	19	Maculino	938382114	LuísSilva@gmail.com	TAP	1972	11/12/2020 11:50	Destino: Bruxelas	160€	N
35	35	Gabriela Lima	Brazileiro	24	Feminio	192004313	GabrielaLima@gmail.com	AZUL LINHAS AEREAS	1992	11/12/2020 03:45	Destino: Rio de Janeiro	600€	N
36	36	Duarte Machado	Português	33	Maculino	930909380	DuarteMachado@gmail.com	AIREUROPA	1573	11/12/2020 09:20	Destino: Madrid	160€	N
37	37	Randy Brown	Americano	66	Maculino	190622342	RandyBrown@gmail.com	UNITED AIRLINES	1394	11/12/2020 16:40	Destino: Washington	645€	N
38	38	João Santos	Português	32	Maculino	918304149	JoãoSantos@gmail.com	TAP	1652	11/12/2020 10:00	Destino: Barcelona	154€	P
39	39	Rita Vieira	Português	20	Feminio	914422643	RitaVieira@gmail.com	AIREUROPA	1249	11/12/2020 13:30	Destino: Madrid	147€	N

Figura 19 –Coluna “Covid19” adicionada com sucesso

Substituir Strings (2)

Foi usado outra vez a substituição de strings para as colunas “Covid19” e “Destino”. Para a coluna “Covid19” deu-se a substituição da letra “N” para a palavra “Negativo” e a letra “P” para a palavra “Positivo”. Para a coluna “Destino” decidiu-se eliminar a expressão “Destino: “ com uma expressão regular, que aparece em todas as linhas.

Step name

#	In stream field	Out stream field	use RegEx	Search	Replace with	Set empty string?	Replace with field	Whole Word	Case sensitive	Is Unicode
1	Destino		S	[a-zA-Z]{7}:		N		S	N	N
2	Covid19		S	N	Negativo	N		S	N	N
3	Covid19		S	P	Positivo	N		S	N	N

Figura 20 –Substituição de strings nas colunas “Destino” e “Covid19”

Na coluna “Destino” foi usado a expressão regular “[a-zA-Z]{7}:" para encontrar em cada linha a expressão “Destino: “ com o objetivo de a eliminar, substituindo essa expressão por nada.

Criar uma nova coluna com “Modified JavaScript value” (2)

De seguida foi criada mais uma nova coluna ao ficheiro csv usando o “Modified JavaScript value”, uma coluna chamada “Lugar” que corresponde ao lugar do avião em que passageiro vai ficar. O “Lugar” de cada passageiro irá ser determinado graças à coluna “Covid19” em que, se o passageiro estiver infetado com Covid-19, este ficará num lugar reservado do avião para que não haja contágio, e se o passageiro não estiver infetado, este ficará num lugar normal do avião.

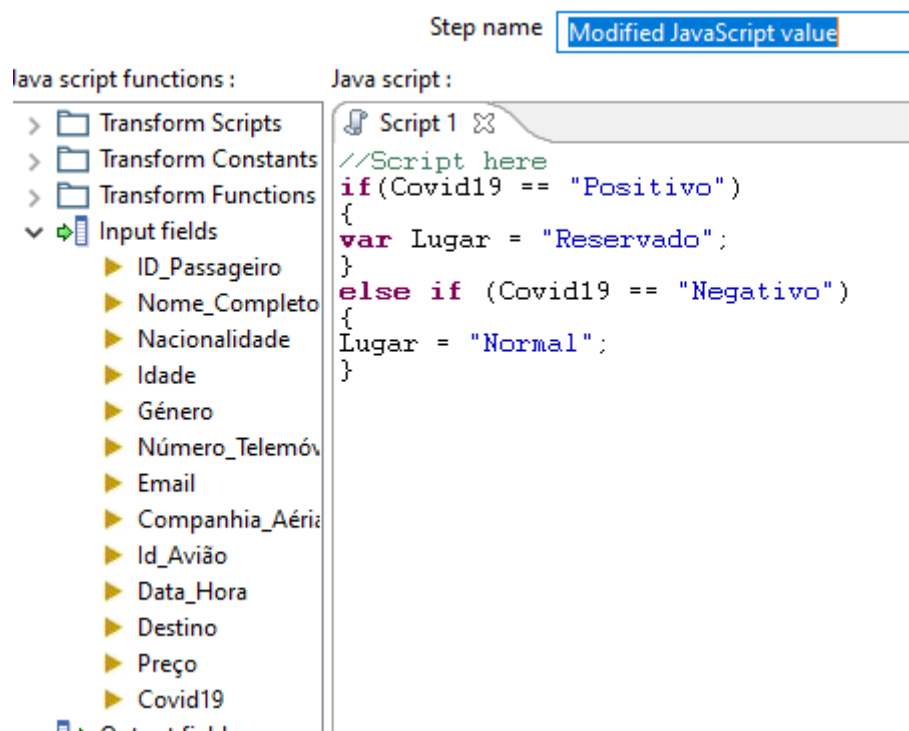


Figura 21 –Criação da coluna “Lugar” seguindo as condições do valor do “Covid19”

#	ID_Passageiro	Nome_Completo	Nacionalidade	Idade	Género	Número_Telemóvel	Email	Companhia_Aérea	Id_Avião	Data_Hora	Destino	Preço	Covid19	Lugar
1	1	Miguel Furtado	Português	23	Maculino	914271152	MiguelFurtado@gmail.com	TAP	1454	11/10/2020 15:30	Lisboa	35€	Negativo	Normal
2	2	David Silva	Português	63	Maculino	936399331	DavidSilva@gmail.com	RYANAIR	1234	11/10/2020 10:20	Nièvre	150€	Negativo	Normal
3	3	Filipa Moreira	Português	73	Feminio	913312313	FilipaMoreira@gmail.com	TAP	1421	11/10/2020 08:50	Nantes	143€	Negativo	Normal
4	4	Rafael Martins	Português	23	Maculino	932748941	RafaelMartins@gmail.com	AIREUROPA	1212	11/10/2020 05:10	Berlim	123€	Negativo	Normal
5	5	Daniela Ferreira	Português	85	Feminio	912923712	DanielaFerreira@gmail.com	RYANAIR	1222	11/10/2020 11:10	Salzburg	153€	Positivo	Reservado
6	6	Nátalia Petejo	Português	26	Feminio	913488551	NátaliaPetejo@gmail.com	EASYJET AIRLINES	1242	11/10/2020 16:00	Madrid	96€	Negativo	Normal
7	7	Vitor Gomes	Português	53	Maculino	936511992	VitorGomes@gmail.com	TAP	1111	11/10/2020 12:30	Texas	550€	Negativo	Normal
8	8	Leonardo Costa	Português	23	Maculino	918311654	LeonardoCosta@gmail.com	AIREUROPA	1020	11/10/2020 11:00	Praga	340€	Positivo	Reservado
9	9	Laura Cardoso	Português	67	Feminio	936523231	LauraCardoso@gmail.com	RYANAIR	1001	11/10/2020 17:15	Istambul	230€	Negativo	Normal
10	10	John Morrison	Americano	34	Maculino	194357425	JohnMorrison@gmail.com	AIREUROPA	1203	11/10/2020 03:10	Paris	154€	Negativo	Normal
11	11	Sean Connery	Inglês	65	Maculino	917243823	SeanConnery@gmail.com	TAP	1200	11/10/2020 09:50	Londres	135€	Negativo	Normal
12	12	Sophie Turner	Inglês	32	Feminio	912322345	SophieTurner@gmail.com	RYANAIR	1209	11/10/2020 04:40	Lisboa	53€	Negativo	Normal
13	13	Isabella Jones	Inglês	19	Feminio	916752832	IsabellaJones@gmail.com	EASYJET AIRLINES	1642	11/10/2020 21:20	Londres	145€	Positivo	Reservado
14	14	George Mclfy	Americano	75	Maculino	192325456	GeorgeMclfy@gmail.com	EASYJET AIRLINES	1931	11/11/2020 23:50	Ohio	650€	Negativo	Normal
15	15	José Mourinho	Português	39	Maculino	933345721	JoséMourinho@gmail.com	RYANAIR	1099	11/11/2020 02:00	Lisboa	57€	Negativo	Normal
16	16	Raquel Gomes	Português	78	Feminio	919572736	RaquelGomes@gmail.com	EASYJET AIRLINES	1365	11/11/2020 05:10	Roma	234€	Positivo	Reservado
17	17	Thiago Batista	Brazileiro	53	Maculino	191241312	ThiagoBatista@gmail.com	AZUL LINHAS AEREAS	1964	11/11/2020 15:50	Rio de Janeiro	451€	Negativo	Normal
18	18	Bernadette Laurent	Frânces	64	Feminio	913141124	BernadetteLaurent@gmail.com	TAP	1421	11/11/2020 13:20	Aveyron	132€	Negativo	Normal
19	19	Steve Austin	Americano	63	Maculino	193143245	SteveAustin@gmail.com	AIREUROPA	1191	11/11/2020 15:05	Texas	860€	Negativo	Normal
20	20	Francisco Carvalho	Português	23	Maculino	938483848	FranciscoCarvalho@gmail.com	RYANAIR	1865	11/11/2020 10:35	Lisboa	45€	Negativo	Normal
21	21	Charlotte Flair	Inglês	45	Feminio	198483424	CharlotteFlair@gmail.com	UNITED AIRLINES	1976	11/11/2020 05:00	Londres	130€	Positivo	Reservado
22	22	Claire Durand	Frânces	40	Feminio	915255834	ClaireDurand@gmail.com	TAP	1921	11/11/2020 13:00	Paris	156€	Positivo	Reservado
23	23	Eduardo Gomes	Português	62	Maculino	939483849	EduardoGomes@gmail.com	UNITED AIRLINES	1543	11/11/2020 16:10	Lisboa	54€	Positivo	Reservado
24	24	Fábio Pinheiro	Português	36	Maculino	918388332	FábioPinheiro@gmail.com	RYANAIR	1910	11/11/2020 19:40	Lisboa	49€	Negativo	Normal
25	25	Joe Miller	Americano	67	Maculino	193243521	JoeMiller@gmail.com	AIREUROPA	1899	11/11/2020 11:30	Nova Iorque	550€	Negativo	Normal
26	26	Charlie Jackson	Inglês	50	Maculino	914231314	CharlieJackson@gmail.com	TAP	1266	11/11/2020 02:25	Londres	137€	Negativo	Normal
27	27	Francine Dubois	Frânces	23	Feminio	953499421	FrancineDubois@gmail.com	TAP	1129	11/12/2020 19:50	Nantes	140€	Negativo	Normal
28	28	Jacqueline Simon	Frânces	27	Feminio	953274231	JacquelineSimon@gmail.com	RYANAIR	1875	11/12/2020 17:00	Paris	143€	Negativo	Normal
29	29	Bruno Pereira	Português	34	Maculino	938243242	BrunoPereira@gmail.com	AIREUROPA	1900	11/12/2020 13:10	Barcelona	120€	Positivo	Reservado
30	30	Oliver Clark	Americano	61	Maculino	194763997	OliverClark@gmail.com	UNITED AIRLINES	1676	11/12/2020 06:00	Geórgia	630€	Negativo	Normal
31	31	Luna Dias	Brazileiro	30	Feminio	192341115	LunaDias@gmail.com	TAP	1500	11/12/2020 01:50	Brasília	510€	Positivo	Reservado
32	32	Pedro Fernandes	Português	54	Maculino	934853688	PedroFernandes@gmail.com	RYANAIR	1061	11/12/2020 18:15	Lisboa	60€	Positivo	Reservado
33	33	Josephine Michel	Frânces	52	Feminio	914321111	JosephineMichel@gmail.com	AIREUROPA	1953	11/12/2020 19:20	Paris	145€	Positivo	Reservado
34	34	Luís Silva	Português	19	Maculino	938382114	LuísSilva@gmail.com	TAP	1972	11/12/2020 11:50	Bruxelas	160€	Negativo	Normal
35	35	Gabriela Lima	Brazileiro	24	Feminio	192004313	GabrielaLima@gmail.com	AZUL LINHAS AEREAS	1992	11/12/2020 03:45	Rio de Janeiro	600€	Negativo	Normal
36	36	Duarte Machado	Português	33	Maculino	930909380	DuarteMachado@gmail.com	AIREUROPA	1573	11/12/2020 09:20	Madrid	160€	Negativo	Normal
37	37	Randy Brown	Americano	66	Maculino	190622342	RandyBrown@gmail.com	UNITED AIRLINES	1394	11/12/2020 16:40	Washington	645€	Negativo	Normal
38	38	João Santos	Português	32	Maculino	918304149	JoãoSantos@gmail.com	TAP	1652	11/12/2020 10:00	Barcelona	154€	Positivo	Reservado
39	39	Rita Vieira	Português	20	Feminio	914422643	RitaVieira@gmail.com	AIREUROPA	1249	11/12/2020 13:30	Madrid	147€	Negativo	Normal

Figura 22 –Ficheiro csv final

Como se pode verificar na coluna “Destino”, já não se encontra nas suas linhas a expressão “Destino: “ e a coluna “Lugar” cumpre sempre a condição feita anteriormente.

Criação dos ficheiros XML e json

Depois do ficheiro csv ficar como pretendido, chegou o momento em converter este para um ficheiro xml e de seguida um ficheiro json.

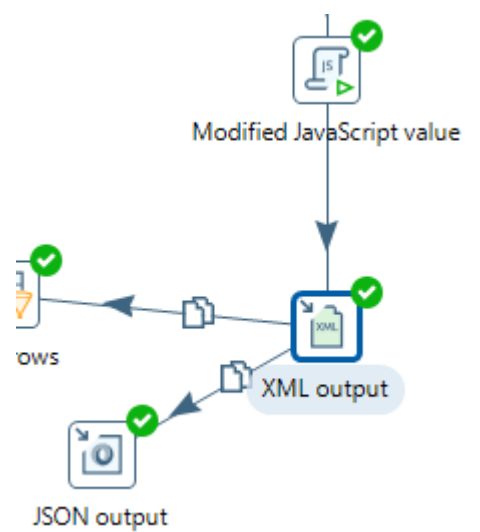


Figura 23 –Conversão do ficheiro csv para XML e de seguida de XML para Json (ver ficheiro Aeroporto.xml e ver pasta Aeroporto.Json)

Filtração das linhas

Para criar dois novos ficheiros XML em que um apenas apresenta os passageiros infetados com Covid-19 e outro em que apresenta os não infetados, foi necessário um “Filter Rows” em foi criada um condição em que se a linha da coluna “Covid19” estiver apresentado como “Positivo”, crirará um ficheiro XML só com os infetados(Covid19Positivo) e se apresentar “Negativo”, crirará outro ficheiro XML que apresenta quem não está infetado (Covid19Negativo).

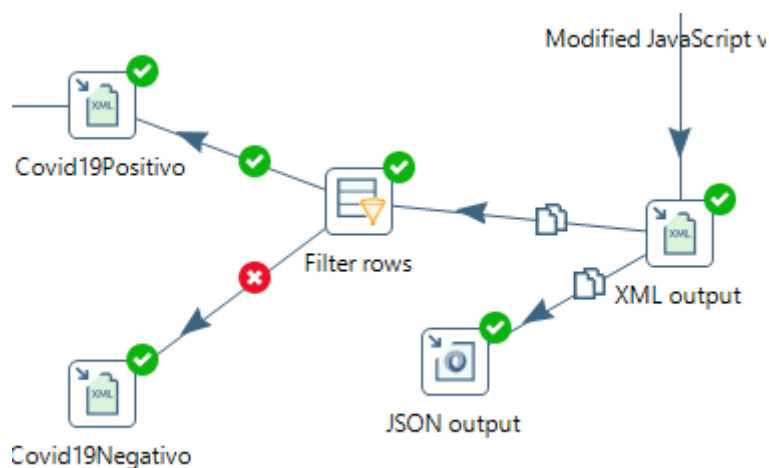


Figura 24 – Filtração apartir do XML output

Step name	Filter rows
Send 'true' data to step:	Covid19Positivo
Send 'false' data to step:	Covid19Negativo

The condition:

Covid19 = [Positivo]

Figura 25 – Condição que irá fazer dois ficheiros XML se a respeitar

Com o ficheiro que apresenta os passageiros com Covid-19 (Covid19Positivo.xml), foi feito outro “Filter Rows” para filtrar esse mesmo ficheiro para criar outros dois ficheiros em que um apresenta quem são os portugueses infectados (PortuguêsesCovid19Positivo.xml) e outro quem são os estrangeiros infectados (EstrangeirosCovid19Positivo.xml).

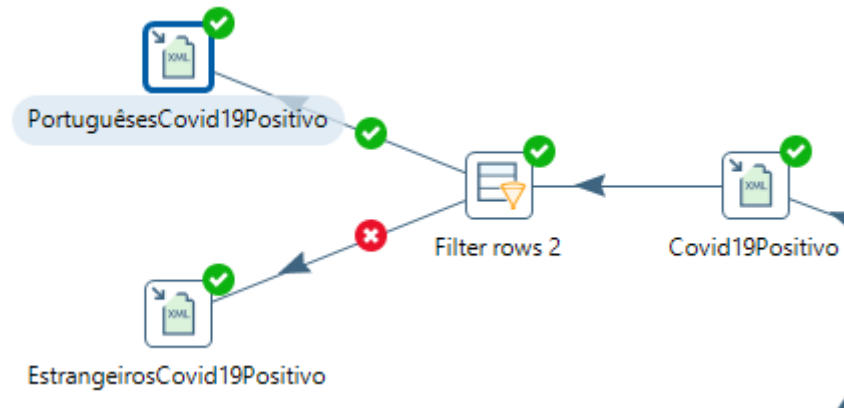


Figura 26 – Filtração a partir do ficheiro Covid19Positivo.xml

Step name	Filter rows 2		
Send 'true' data to step:	PortuguêsesCovid19Positivo		
Send 'false' data to step:	EstrangeirosCovid19Positivo		

The condition:

☐

Nacionalidade

=

Português

(String)

Figura 27 – Condição que irá fazer dois ficheiros XML se a respeitar

Construção do Job

Um job consiste num conjunto de transformações, em que no Job criado para este trabalho, irá executar toda a transformação que foi explicada anteriormente e criará um ficheiro html ,que vai ser permitida graças a um stylesheet XSL que foi criado, que terá o conteúdo do ficheiro XML criado na transformação anterior (Aeroporto.xml) e, de seguida, o html já criado, será enviado para o meu email.



Figura 28 – Job que irá executar a transformação já explicada anteriormente (Transformation), em que vai criar um ficheiro html que vai ser usado na Transformation(2) (que depois será explicada) para enviar para o email

Job entry name: XSL transformation

General **Advanced**

Files

Get filenames from previous result ☐

XML File name:

XSL File name:

Output Filename:

Settings

Transformer factory: ▼

If file exists: ▼

Add output filename to result ☐

Figura 29 – XSL transformation em que será necessário o ficheiro Aeroporto.xml que foi criado na transformação anterior, e um ficheiro XSL para transformar num ficheiro html

```

<?xml version="1.0"?>
- <xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform" version="1.0">
  <xsl:output encoding="ISO-8859-1" method="html"/>
  - <xsl:template match="/">
    - <html>
      - <head>
        <title>Passageiros</title>
      </head>
      - <body>
        <h1>Passageiros</h1>
        - <table border="3">
          - <tr>
            <th>ID_Passageiro</th>
            <th>Nome_Completo</th>
            <th>Nacionalidade</th>
            <th>Idade</th>
            <th>Género</th>
            <th>Número_Telemóvel</th>
            <th>Email</th>
            <th>Companhia_Aérea</th>
            <th>Id_Avião</th>
            <th>Data_Hora</th>
            <th>Destino</th>
            <th>Preço</th>
            <th>Covid19</th>
            <th>Lugar</th>
          </tr>
          - <xsl:for-each select="Passageiros/Passageiro">
            - <tr>
              - <td>
                <xsl:apply-templates select="ID_Passageiro"/>
              </td>
              - <td>
                <xsl:apply-templates select="Nome_Completo"/>
              </td>
              - <td>
                <xsl:apply-templates select="Nacionalidade"/>
              </td>
              - <td>
                <xsl:apply-templates select="Idade"/>
              </td>
              - <td>
                <xsl:apply-templates select="Género"/>
              </td>
              - <td>
                <xsl:apply-templates select="Número_Telemóvel"/>
              </td>
              - <td>
                <xsl:apply-templates select="Email"/>
              </td>
              - <td>
                <xsl:apply-templates select="Companhia_Aérea"/>
              </td>
              - <td>
                <xsl:apply-templates select="Id_Avião"/>
              </td>
              - <td>
                <xsl:apply-templates select="Data_Hora"/>
              </td>
              - <td>
                <xsl:apply-templates select="Destino"/>
              </td>
              - <td>
                <xsl:apply-templates select="Preço"/>
              </td>
              - <td>
                <xsl:apply-templates select="Covid19"/>
              </td>
              - <td>
                <xsl:apply-templates select="Lugar"/>
              </td>
            </tr>
          </xsl:for-each>
        </table>
      </body>
    </html>
  </xsl:template>
</xsl:stylesheet>

```

Figura 30 – Ficheiro XSL que vai ser usado para a criação do ficheiro html

Segunda Transformação (“html_to_field”)



Figura 31 – Segunda Transformação

Esta transformação tem como objetivo de agrupar os dados do html e definir as suas variáveis para que seja enviado para o email o conteúdo do html.

No “Text file input” foi necessário para fazer input do ficheiro html criado anteriormente (Aeroporto.html).

Nome do Step: **Text file input**

ds

File or directory:

Regular Expression:

Exclude Regular Expression:

Selected files:

#	File/Directory	Wildcard (RegExp)	Exclude wildcard	Required	Include s
1	\${Internal.Entry.Current.Directory}/Output/Aeroporto.html			N	N

Figura 32 – Input do ficheiro html no “Text file input”

Nome do Step: **Text file input**

#	Name	Type	Format	Position	Length	Precision	Currency	Decimal	Group	Null if	Default	Trim type	Repeat
1	html	String		0	80							none	N

Figura 33 – Definir o tamanho do ficheiro html e também dar um nome e definir o seu tipo

Step name:

Include all rows? ☐

Temporary files directory:

TMP-file prefix:

Add line number, restart in each group ☐

Line number field name:

Always give back a result row ☐

The fields that make up the group:

#	Group field
1	

Aggregates :

#	Name	Subject	Type	Value
1	html	html	Concatenate strings separated by	

Figura 34 – Definir o subjetivo do agregado e que terá que concatenar strings

Este “Group by” tem como função agrupar os campos que pretendo, que será o html que já está no “Text file input”, para isso foi necessário especificar o grupo que desejaria agregar, foi dado o nome “html” (Name) ao campo resultante, escolher o assunto (Subject) que é o html que se encontra no “Text file input” e qual o método de agregar que vou usar, que foi “Concatenate strings that are separated by”.

Step name:

Apply formatting ☒

Field values:

#	Field name	Variable name	Variable scope type
1	html	html_body	Valid in the Java Virtual Machine

Figura 35 – Definir o nome da variável do campo já escolhido e definir o tipo de escopo variável

Enviar para o Email

Voltando outra vez para o Job(AeroportoJob.kjb) foi necessário fazer uma ligação da Transformation(2) (transformação anteriormente explicada) com o “Mail”, com o objetivo de enviar o conteúdo do html para o meu email.



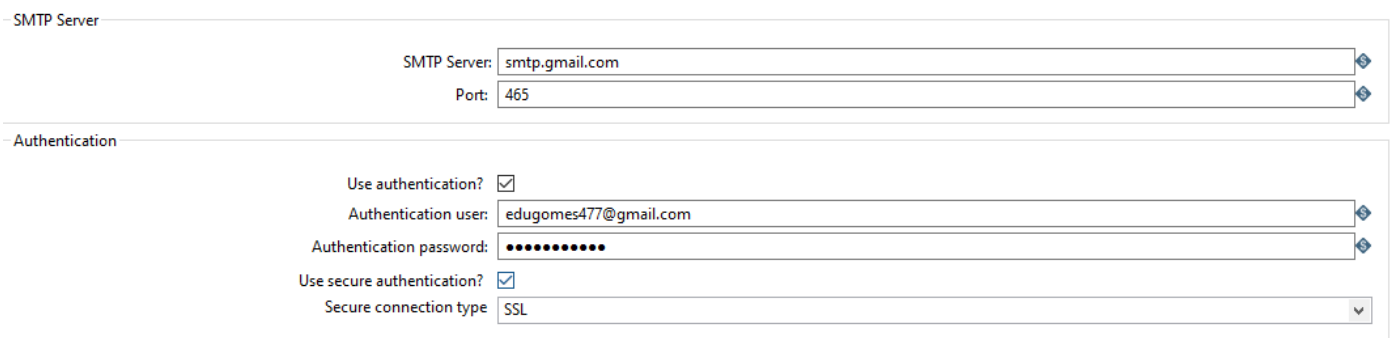
The screenshot shows the 'Mail' configuration window with two main sections: 'Destination' and 'Sender'. In the 'Destination' section, the 'Destination address' is set to 'edugomes477@gmail.com', while 'Cc' and 'BCc' are empty. In the 'Sender' section, the 'Sender name' is 'Eduardo' and the 'Sender address' is 'edugomes477@gmail.com'. Below these, the 'Reply to' field is also 'edugomes477@gmail.com', and 'Contact person' and 'Contact phone' are empty.

Destination	
Destination address:	edugomes477@gmail.com
Cc:	
BCc:	

Sender	
Sender name:	Eduardo
Sender address:	edugomes477@gmail.com
Reply to:	edugomes477@gmail.com
Contact person:	
Contact phone:	

Figura 36 – Endereço de destino e o Endereço que o vai enviar

Para que o envio para o Email seja concedido, é necessário inserir qual o endereço que vai enviar o conteúdo do html (Sender address), também foi necessário inserir o nome que possui esse mesmo endereço (Sender name), e inserir o endereço que vai receber o conteúdo enviado (Destination address). Como se pode verificar, vou usar o meu email para enviar o html para o meu próprio email.



The screenshot shows the 'SMTP Server' and 'Authentication' configuration sections. In the 'SMTP Server' section, the 'SMTP Server' is 'smtp.gmail.com' and the 'Port' is '465'. In the 'Authentication' section, 'Use authentication?' is checked, 'Authentication user' is 'edugomes477@gmail.com', 'Authentication password' is masked with dots, 'Use secure authentication?' is checked, and 'Secure connection type' is set to 'SSL'.

SMTP Server	
SMTP Server:	smtp.gmail.com
Port:	465

Authentication	
Use authentication?	<input checked="" type="checkbox"/>
Authentication user:	edugomes477@gmail.com
Authentication password:	••••••••
Use secure authentication?	<input checked="" type="checkbox"/>
Secure connection type	SSL

Figura 37 – SMTP Server e Autentificação do meu email

Para enviar o email foi necessário inserir o “SMTP Server” e número do “Port” e usar autenticação, que é o ato de estabelecer ou confirmar algo como autêntico, para que o kettle tenha acesso ao meu email, foi necessário inserir o meu email e a minha palavra-passe, e no final usar autenticação segura (“Use secure authentication”) com o tipo de conexão “SSL”.

Addresses

Server

Email Message

Attached Files

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: Passageiros

Comment: \${html_body}

Figura 38 – Enviar a mensagem

De seguida, foi necessário meter no “Email Message” o que queria enviar, foi ativado a opção de incluir na mensagem a data “Include date in message” e usar o formato de html na mensagem “Use HTML format in mail body”. No assunto (Subject) da mensagem foi escrito “Passageiros” e na própria mensagem (Comment) foi necessário escrever “\${html_body}” que irá enviar todo o conteúdo do html obtido na Transformation(2).

Passageiros

ID_Passageiro	Nome_Completo	Nacionalidade	Idade	Género	Número_Telemóvel	Email	Companhia_Aérea	Id_Avião	Data_Hora	Destino	Preço	Covid19	Lugar
1	Miguel Furtado	Português	23	Maculino	914271152	MiguelFurtado@gmail.com	TAP	1454	11/10/2020 15:30	Lisboa	35€	Negativo	Normal
2	David Silva	Português	63	Maculino	936399331	DavidSilva@gmail.com	RYANAIR	1234	11/10/2020 10:20	Nièvre	150€	Negativo	Normal
3	Filipa Moreira	Português	73	Feminino	913312313	FilipaMoreira@gmail.com	TAP	1421	11/10/2020 08:50	Nantes	143€	Negativo	Normal
4	Rafael Martins	Português	23	Maculino	932748941	RafaelMartins@gmail.com	AIREUROPA	1212	11/10/2020 05:10	Berlim	123€	Negativo	Normal
5	Daniela Ferreira	Português	85	Feminino	912923712	DanielaFerreira@gmail.com	RYANAIR	1222	11/10/2020 11:10	Salzburg	153€	Positivo	Reservado
6	Natália Petejo	Português	26	Feminino	913488551	NatáliaPetejo@gmail.com	EASYJET AIRLINES	1242	11/10/2020 16:00	Madrid	96€	Negativo	Normal
7	Vitor Gomes	Português	53	Maculino	936511992	VitorGomes@gmail.com	TAP	1111	11/10/2020 12:30	Texas	550€	Negativo	Normal
8	Leonardo Costa	Português	23	Maculino	918311654	LeonardoCosta@gmail.com	AIREUROPA	1020	11/10/2020 11:00	Praga	340€	Positivo	Reservado
9	Laura Cardoso	Português	67	Feminino	936523231	LauraCardoso@gmail.com	RYANAIR	1001	11/10/2020 17:15	Istambul	230€	Negativo	Normal
10	John Morrison	Americano	34	Maculino	194357425	JohnMorrison@gmail.com	AIREUROPA	1203	11/10/2020 03:10	Paris	154€	Negativo	Normal
11	Sean Connery	Inglês	65	Maculino	917243823	SeanConnery@gmail.com	TAP	1200	11/10/2020 09:50	Londres	135€	Negativo	Normal
12	Sophie Turner	Inglês	32	Feminino	912322345	SophieTurner@gmail.com	RYANAIR	1209	11/10/2020 04:40	Lisboa	53€	Negativo	Normal
13	Isabella Jones	Inglês	19	Feminino	916752832	IsabellaJones@gmail.com	EASYJET AIRLINES	1642	11/10/2020 21:20	Londres	145€	Positivo	Reservado
14	George Mcfly	Americano	75	Maculino	192325456	GeorgeMcfly@gmail.com	EASYJET AIRLINES	1931	11/11/2020 23:50	Ohio	650€	Negativo	Normal
15	José Mourinho	Português	39	Maculino	933345721	JoséMourinho@gmail.com	RYANAIR	1099	11/11/2020 02:00	Lisboa	57€	Negativo	Normal
16	Raquel Gomes	Português	78	Feminino	919572736	RaquelGomes@gmail.com	EASYJET AIRLINES	1365	11/11/2020 05:10	Roma	234€	Positivo	Reservado
17	Thiago Batista	Brazileiro	53	Maculino	191241312	ThiagoBatista@gmail.com	AZUL LINHAS AEREAS	1964	11/11/2020 15:50	Rio de Janeiro	451€	Negativo	Normal
18	Bernadette Laurent	Francês	64	Feminino	913141124	BernadetteLaurent@gmail.com	TAP	1421	11/11/2020 13:20	Aveyron	132€	Negativo	Normal
19	Steve Austin	Americano	63	Maculino	193143245	SteveAustin@gmail.com	AIREUROPA	1191	11/11/2020 15:05	Texas	860€	Negativo	Normal
20	Francisco Carvalho	Português	23	Maculino	938483848	FranciscoCarvalho@gmail.com	RYANAIR	1865	11/11/2020 10:35	Lisboa	45€	Negativo	Normal
21	Charlotte Flair	Inglês	45	Feminino	198483424	CharlotteFlair@gmail.com	UNITED AIRLINES	1976	11/11/2020 05:00	Londres	130€	Positivo	Reservado
22	Claire Durand	Francês	40	Feminino	915255834	ClaireDurand@gmail.com	TAP	1921	11/11/2020 13:00	Paris	156€	Positivo	Reservado
23	Eduardo Gomes	Português	62	Maculino	939483849	EduardoGomes@gmail.com	UNITED AIRLINES	1543	11/11/2020 16:10	Lisboa	54€	Positivo	Reservado
24	Fábio Pinheiro	Português	36	Maculino	918388332	FábioPinheiro@gmail.com	RYANAIR	1910	11/11/2020 19:40	Lisboa	49€	Negativo	Normal
25	Joe Miller	Americano	67	Maculino	193243521	JoeMiller@gmail.com	AIREUROPA	1899	11/11/2020 11:30	Nova Iorque	550€	Negativo	Normal
26	Charlie Jackson	Inglês	50	Maculino	914231314	CharlieJackson@gmail.com	TAP	1266	11/11/2020 02:25	Londres	137€	Negativo	Normal
27	Francine Dubois	Francês	23	Feminino	953499421	FrancineDubois@gmail.com	TAP	1129	11/12/2020 19:50	Nantes	140€	Negativo	Normal
28	Jacqueline Simon	Francês	27	Feminino	953274231	JacquelineSimon@gmail.com	RYANAIR	1875	11/12/2020 17:00	Paris	143€	Negativo	Normal
29	Bruno Pereira	Português	34	Maculino	938243242	BrunoPereira@gmail.com	AIREUROPA	1900	11/12/2020 13:10	Barcelona	120€	Positivo	Reservado
30	Oliver Clark	Americano	61	Maculino	194763997	OliverClark@gmail.com	UNITED AIRLINES	1676	11/12/2020 06:00	Geórgia	630€	Negativo	Normal
31	Luna Dias	Brazileiro	30	Feminino	192341115	LunaDias@gmail.com	TAP	1500	11/12/2020 01:50	Brasília	510€	Positivo	Reservado
32	Pedro Fernandes	Português	54	Maculino	934853688	PedroFernandes@gmail.com	RYANAIR	1061	11/12/2020 18:15	Lisboa	60€	Positivo	Reservado
33	Josephine Michel	Francês	52	Feminino	914321111	JosephineMichel@gmail.com	AIREUROPA	1953	11/12/2020 19:20	Paris	145€	Positivo	Reservado
34	Luís Silva	Português	19	Maculino	938382114	LuísSilva@gmail.com	TAP	1972	11/12/2020 11:50	Bruxelas	160€	Negativo	Normal
35	Gabriela Lima	Brazileiro	24	Feminino	192004313	GabrielaLima@gmail.com	AZUL LINHAS AEREAS	1992	11/12/2020 03:45	Rio de Janeiro	600€	Negativo	Normal
36	Duarte Machado	Português	33	Maculino	930909380	DuarteMachado@gmail.com	AIREUROPA	1573	11/12/2020 09:20	Madrid	160€	Negativo	Normal
37	Randy Brown	Americano	66	Maculino	190622342	RandyBrown@gmail.com	UNITED AIRLINES	1394	11/12/2020 16:40	Washington	645€	Negativo	Normal
38	João Santos	Português	32	Maculino	918304149	JoãoSantos@gmail.com	TAP	1652	11/12/2020 10:00	Barcelona	154€	Positivo	Reservado
39	Rita Vieira	Português	20	Feminino	914422643	RitaVieira@gmail.com	AIREUROPA	1249	11/12/2020 13:30	Madrid	147€	Negativo	Normal

Conclusão

Os processos de Extract, Transformation and Load (ETL) foram executados através da utilização da ferramenta Pentaho Kettle. Com a utilização das funcionalidades de transformação fornecidas pelo Pentaho foi possível concluir o objetivo definido que passava por extrair os dados de um ficheiro csv, transformar esses dados e criar ficheiros XML, Json e html e pegando no html com o objetivo de enviar por email. ETL mostrou-se um processo muito útil e pratico na transformação de dados e neste caso como a ferramenta Pentaho Kettle dispunha de diversas formas de transformações tinha á disposição muitas opções de conversão de dados.

Bibliografia:

<https://wiki.pentaho.com/>

<https://forums.pentaho.com/>

<https://stackoverflow.com/>

<http://www.regexlib.com/>

Material fornecido pelo professor que consta no moodle