



UNIVERSIDADE DO MINHO

APLICAÇÕES INFORMÁTICAS NA BIOMEDICINA  
MESTRADO EM ENGENHARIA INFORMÁTICA  
2019/2020

**Grupo: 1**

---

## Resolução da ficha prática 5

---

***Trabalho realizado por:***

Filipa Parente  
João Almeida  
Leonardo Jesus  
Nuno Valente

***Número de Aluno:***

A82145  
A75209  
PG39261  
A81986

19 de Novembro de 2019

# Conteúdo

<b>Conteúdo</b>	<b>1</b>
<b>1 Resolução das questões propostas</b>	<b>2</b>
1.1 Questão 1 . . . . .	2
1.2 Questão 2 . . . . .	2
1.3 Questão 3 . . . . .	5

# 1 Resolução das questões propostas

## 1.1 Questão 1

1) Defina três jobs diferentes e justifique a utilidade de cada job com uma contextualização

Após a análise das informações do ficheiro csv, o grupo achou interessante a inclusão dos seguintes jobs:

- a) **peessoal 0.1:** Um job que lê o ficheiro csv original e filtra as informações gerais de cada paciente (*age, gender, country, state*). Este job foi definido para fins estatísticos.
- b) **bd\_mental\_health 0.1:** Um job que lê o ficheiro csv original e povoa uma tabela da base de dados (*ficha5*). Este job surgiu no contexto de criar uma alternativa à ferramenta já existente no MySQL workbench "Import Wizard".
- c) **filtros 0.1:** Um job que lê a base de dados criada no job anterior e filtra todos os pacientes dos Estados Unidos da América e subdivide-os por *gender* em 2 ficheiros csv diferentes, ordenados por ordem alfabética do *state*. Este job foi definido com o intuito de facilitar a pesquisa dos pacientes americanos existentes na base de dados.

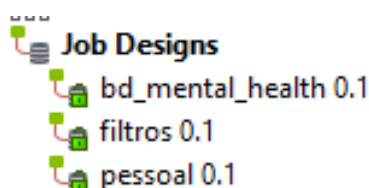


Figura 1: *listagem dos jobs criados no Talend*

## 1.2 Questão 2

2) Crie as jobs definidas no Talend, recorrendo às componentes adequadas. Deverá incluir, no mínimo, os seguintes elementos:

- a) **File Components:** tFileInputDelimited, tFileOutputExcel;
- b) **Processing components:** tMap, tFilterRow, tSortRow;
- c) **Database Components:** tDBConnection, tDBOutput;
- d) **Filtros de rejeição.**

Tendo em conta os jobs definidos já na questão 1 e os componentes mínimos exigidos, procedeu-se à sua criação no Talend.

### • Job "pessoal 0.1"

Para realizar este job, utilizaram-se 3 componentes diferentes:

- **tFileInputDelimited** para ler o ficheiro *mental\_health.csv*
- **tFilterCollumns** para filtrar os atributos pretendidos (*age, gender, country, state*)
- **tFileOutputExcel** para inserir os dados já filtrados no ficheiro *inf\_pessoal.xls*

Representado no seguinte esquema:

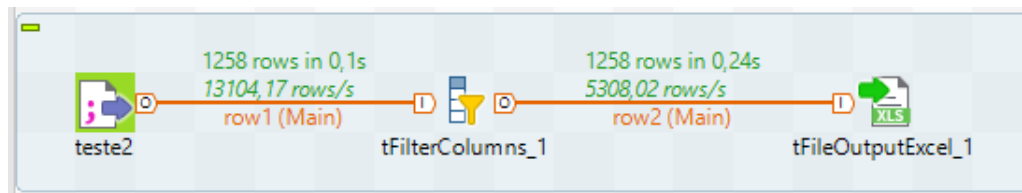


Figura 2: esquema do job "pessoal 0.1" definido na Questão 1

### • Job "bd\_mental\_health 0.1"

Para realizar este job, utilizaram-se 4 componentes diferentes:

- **tDBConnection** para estabelecer uma conexão com a base de dados onde se vão inserir os dados (*ficha5*)
- **tFileInputDelimited** para ler o ficheiro *mental\_health.csv*
- **tMap** para mapear os atributos presentes no ficheiro *mental\_health.csv* para os atributos respetivos da nova tabela a ser criada.
- **tDBOutput** para criar a tabela e inserir os dados na BD (*ficha5*)

Representado no seguinte esquema:

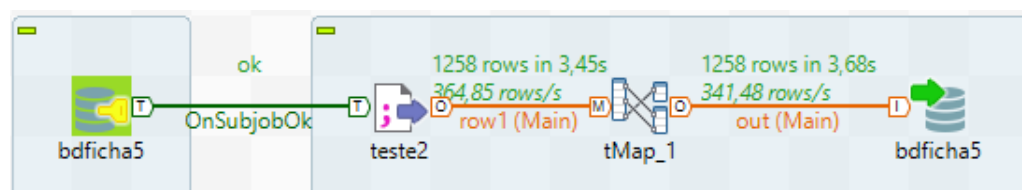


Figura 3: esquema do job "bd\_mental\_health 0.1" definido na Questão 1

### • Job "filtros 0.1"

Para realizar este job, utilizaram-se 10 componentes diferentes:

- **tDBConnection** para estabelecer uma conexão com a base de dados onde se vão ler os dados (*ficha5*)
- **tDBInput** para ler da base de dados já povoada (*ficha5*)
- **2 tFilterRow** para filtrar todos os pacientes norte-americanos (tFilterRow\_4) e o (tFilterRow\_1) para dividir os pacientes do sexo masculino (linha de cima) e os restantes (linha de baixo) através de um filtro de rejeição.
- **2 tFilterColumns** para filtrar os atributos pretendidos( Country e Gender).
- **2 tSortRow** para ordenar cada paciente (masculino ou feminino) por ordem alfabética do estado onde vive
- **2 tFileOutputExcel** para inserir os dados já filtrados nos ficheiros respetivos *tFileOutputExcel\_1* para o ficheiro *outmachosUSA.xls* e *tFileOutputExcel\_2* para o ficheiro *outFemeasUSA.xls*

Representado no seguinte esquema:

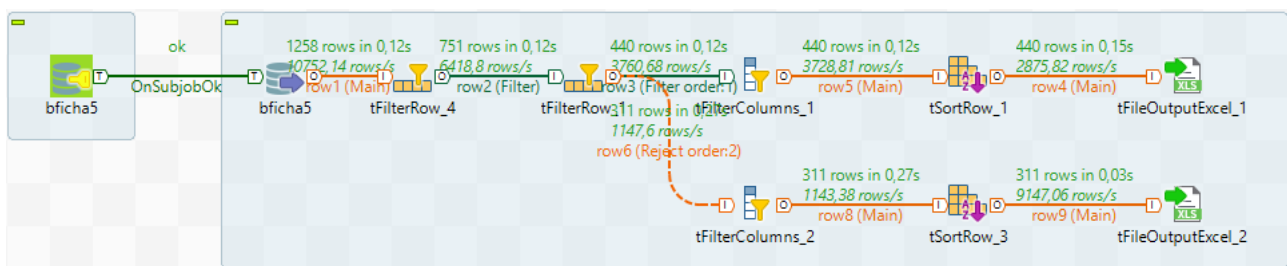


Figura 4: esquema do job "filtros 0.1" definido na Questão 1

Note-se que os três jobs em conjunto possuem os componentes mínimos pedidos no enunciado.

### 1.3 Questão 3

#### 3) Corra cada job definida e criada

Como se pode observar nas imagens 2, 3 e 4, todos os jobs definidos estão a funcionar normalmente.

Partindo agora para a análise dos resultados finais presentes nas seguintes imagens:

	A	B	C	D	E
1	37	Female	United Sta	IL	
2	44	M	United Sta	IN	
3	32	Male	Canada	NA	
4	31	Male	United Kin	NA	
5	31	Male	United Sta	TX	
6	33	Male	United Sta	TN	
7	35	Female	United Sta	MI	
8	39	M	Canada	NA	
9	42	Female	United Sta	IL	
10	23	Male	Canada	NA	
11	31	Male	United Sta	OH	
12	29	male	Bulgaria	NA	
13	42	female	United Sta	CA	
14	36	Male	United Sta	CT	
15	27	Male	Canada	NA	
16	29	female	United Sta	IL	
17	23	Male	United Kin	NA	
18	32	Male	United Sta	TN	
19	46	male	United Sta	MD	
20	36	Male	France	NA	
21	29	Male	United Sta	NY	
22	31	male	United Sta	NC	
23	46	Male	United Sta	MA	
24	41	Male	United Sta	IA	
25	33	male	United Sta	CA	
26	35	male	United Sta	TN	
27	33	male	United Sta	TN	
28	35	Female	United Sta	CA	
29	34	male	United Sta	OH	
30	37	Male	United Kin	NA	

Figura 5: conteúdo do ficheiro *info\_pessoas.xls*

ID	Timestamp	Age	Gender	Country	state	self_employed	family_history	treatment	work_interfere	no_employees	remote_work	tech_company	benefits	care_options	wellness_program	seek_help	anonymity
1	2014-08-27 11:29:31	37	Female	United States	IL	NA	No	Yes	Often	6-25	No	Yes	Yes	Not sure	No	Yes	Yes
2	2014-08-27 11:29:37	44	M	United States	IN	NA	No	No	Rarely	More than 1000	No	No	Don't know	No	Don't know	Don't know	Don't know
3	2014-08-27 11:29:44	32	Male	Canada	NA	NA	No	No	Rarely	6-25	No	Yes	No	No	No	No	Don't know
4	2014-08-27 11:29:46	31	Male	United Kingdom	NA	NA	Yes	Yes	Often	26-100	No	Yes	No	Yes	No	No	No
5	2014-08-27 11:30:22	31	Male	United States	TX	NA	No	No	Never	100-500	Yes	Yes	Yes	No	Don't know	Don't know	Don't know
6	2014-08-27 11:31:22	33	Male	United States	TN	NA	Yes	No	Sometimes	6-25	No	Yes	Yes	Not sure	No	Don't know	Don't know
7	2014-08-27 11:31:50	35	Female	United States	MI	NA	Yes	Yes	Sometimes	1-5	Yes	Yes	No	No	No	No	No
8	2014-08-27 11:32:05	39	M	Canada	NA	NA	No	No	Never	1-5	Yes	Yes	No	Yes	No	No	Yes
9	2014-08-27 11:32:39	42	Female	United States	IL	NA	Yes	Yes	Sometimes	100-500	No	Yes	Yes	Yes	No	No	No
10	2014-08-27 11:32:43	23	Male	Canada	NA	NA	No	No	Never	26-100	No	Yes	Don't know	No	Don't know	Don't know	Don't know
11	2014-08-27 11:32:44	31	Male	United States	OH	NA	No	Yes	Sometimes	6-25	Yes	Yes	Don't know	No	No	No	Don't know
12	2014-08-27 11:32:49	29	male	Bulgaria	NA	NA	No	No	Never	100-500	Yes	Yes	Don't know	Not sure	No	No	Don't know
13	2014-08-27 11:33:23	42	female	United States	CA	NA	Yes	Yes	Sometimes	26-100	No	No	Yes	Yes	No	No	Don't know
14	2014-08-27 11:33:26	36	Male	United States	CT	NA	Yes	No	Never	500-1000	No	Yes	Don't know	Not sure	No	Don't know	Don't know
15	2014-08-27 11:33:57	27	Male	Canada	NA	NA	No	No	Never	6-25	No	Yes	Don't know	Not sure	Don't know	Don't know	Don't know
16	2014-08-27 11:34:00	29	female	United States	IL	NA	Yes	Yes	Rarely	26-100	No	Yes	Yes	Not sure	No	No	Don't know
17	2014-08-27 11:34:20	23	Male	United Kingdom	NA	NA	No	Yes	Sometimes	26-100	Yes	Yes	Don't know	No	Don't know	Don't know	Don't know
18	2014-08-27 11:34:37	32	Male	United States	TN	NA	No	Yes	Sometimes	6-25	No	Yes	Yes	Yes	No	Don't know	Don't know
19	2014-08-27 11:34:53	46	male	United States	MD	Yes	Yes	No	Sometimes	1-5	Yes	Yes	Yes	Not sure	Yes	Don't know	Yes
20	2014-08-27 11:35:08	36	Male	France	NA	Yes	Yes	No	NA	6-25	Yes	Yes	No	No	Yes	No	Yes
21	2014-08-27 11:35:12	29	Male	United States	NY	No	Yes	Yes	Sometimes	100-500	No	Yes	Yes	Yes	No	No	No
22	2014-08-27 11:35:24	31	male	United States	NC	Yes	No	No	Never	1-5	Yes	Yes	No	No	No	No	Yes
23	2014-08-27 11:35:48	46	Male	United States	MA	No	No	Yes	Often	26-100	Yes	Yes	Yes	Yes	No	No	Don't know
24	2014-08-27 11:36:24	41	Male	United States	IA	No	No	Yes	Never	More than 1000	No	No	Don't know	No	No	Don't know	Don't know
25	2014-08-27 11:36:48	33	male	United States	CA	No	Yes	Yes	Rarely	26-100	No	Yes	Yes	Not sure	Don't know	Yes	Yes
26	2014-08-27 11:37:08	35	male	United States	TN	No	Yes	Yes	Sometimes	More than 1000	No	No	Yes	Yes	No	Don't know	No
27	2014-08-27 11:37:23	33	male	United States	TN	No	No	No	NA	1-5	No	Yes	Don't know	Not sure	No	Don't know	Don't know

Figura 6: conteúdo da tabela criada (*ficha5.mental\_health*)

# 1 RESOLUÇÃO DAS QUESTÕES PROPOSTAS

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	2014-08-27	AL	No	Yes	Yes	Sometime: 26-100	Yes	No	Yes	Don't know	No	No	Don't know	Yes	Somewhat	Maybe	No	Some of th	Some of th	No	Yes
2	2014-08-27	AL	No	No	Yes	Sometime: 26-100	Yes	Yes	Yes	Yes	Yes	Don't know	No	Don't know	Very easy	No	No	Yes	Yes	Yes	Yes
3	2014-08-27	AL	No	Yes	Yes	Sometime: 100-500	No	No	Don't know	No	No	No	No	Don't know	Don't know	No	No	Yes	Yes	Maybe	Yes
4	2014-11-16	AL	No	Yes	Yes	Sometime: 26-100	Yes	Yes	Yes	No	No	No	Don't know	Don't know	Don't know	Maybe	No	Some of th	Some of th	No	No
5	2014-12-16	AL	No	Yes	Yes	Sometime: 6-25	No	Yes	Yes	Yes	Yes	No	No	Yes	Don't know	Maybe	No	Some of th	No	No	No
6	2014-08-27	AZ	No	No	Yes	Rarely	26-100	No	Yes	Yes	Yes	Yes	No	Don't know	Don't know	No	No	Some of th	Yes	No	No
7	2014-08-27	AZ	No	No	No	NA	More than No	No	No	No	No	No	Don't know	Don't know	Don't know	Maybe	Maybe	No	No	No	No
8	2014-08-25	AZ	No	Yes	Yes	Often	26-100	No	Yes	Yes	Yes	No	Yes	Don't know	Yes	No	No	Some of th	No	No	No
9	2014-08-27	CA	No	Yes	No	Sometime: More than Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Don't know	Don't know	Yes	Maybe	Some of th	No	No	No
10	2014-08-27	CA	No	No	No	Never	More than No	No	Yes	Don't know	Not sure	Yes	Don't know	Don't know	Don't know	Maybe	No	Some of th	Some of th	No	Maybe
11	2014-08-27	CA	No	Yes	No	Never	More than No	No	Yes	No	No	No	No	Don't know	Don't know	Yes	No	No	No	No	Maybe
12	2014-08-27	CA	No	No	Yes	Sometime: 26-100	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Somewhat	No	No	Some of th	Some of th	No	Maybe
13	2014-08-27	CA	No	Yes	Yes	Sometime: More than No	No	Yes	Don't know	No	No	Don't know	Don't know	Don't know	Don't know	No	No	No	Some of th	No	Maybe
14	2014-08-27	CA	No	No	No	Never	More than No	No	Yes	Yes	No	No	Yes	Don't know	Don't know	No	No	Some of th	Some of th	No	No
15	2014-08-27	CA	No	No	Yes	Sometime: More than No	No	Yes	Don't know	Not sure	Don't know	Don't know	Don't know	Don't know	Don't know	Yes	Yes	Some of th	Some of th	No	No
16	2014-08-27	CA	No	No	Yes	Rarely	6-25	No	Yes	No	No	No	No	Yes	Don't know	No	No	Some of th	Yes	No	No
17	2014-08-27	CA	No	No	No	Rarely	More than No	Yes	Yes	No	No	No	No	Yes	Somewhat	No	No	Some of th	Yes	No	Maybe
18	2014-08-27	CA	No	Yes	Yes	Sometime: 26-100	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Somewhat	No	No	Some of th	Yes	No	No
19	2014-08-27	CA	No	No	No	Never	More than No	No	Yes	Yes	No	No	Don't know	Don't know	Don't know	Maybe	No	Some of th	Some of th	No	No
20	2014-08-27	CA	No	Yes	Yes	Sometime: More than No	No	Yes	Yes	No	Yes	Yes	Yes	Don't know	Don't know	Yes	No	No	No	No	Maybe
21	2014-08-27	CA	No	Yes	Yes	Sometime: More than No	No	Yes	Yes	Not sure	No	Don't know	Don't know	Don't know	Yes	No	No	No	No	No	No
22	2014-08-27	CA	No	Yes	No	NA	More than No	Yes	Don't know	No	Don't know	Don't know	Don't know	Don't know	Very easy	No	No	Yes	Yes	Maybe	Maybe
23	2014-08-27	CA	No	No	Yes	Sometime: More than No	No	Yes	Yes	No	No	Don't know	Don't know	Don't know	Don't know	Yes	No	Some of th	Some of th	No	Maybe
24	2014-08-27	CA	No	No	Yes	Rarely	6-25	No	Yes	Yes	No	Don't know	No	Don't know	Don't know	Maybe	No	No	No	No	No
25	2014-08-27	CA	No	Yes	Yes	Sometime: More than No	No	Yes	Don't know	Not sure	Don't know	Don't know	Yes	Yes	Don't know	Don't know	No	Yes	Yes	No	No
26	2014-08-27	CA	No	No	No	Sometime: More than No	No	Yes	Don't know	Not sure	Don't know	Don't know	Don't know	Don't know	Very easy	Yes	No	No	No	No	Maybe
27	2014-08-27	CA	No	Yes	Yes	Never	More than No	No	Yes	Don't know	No	Yes	Yes	Yes	Don't know	Don't know	No	No	Yes	No	No

Figura 7: conteúdo do ficheiro outmachosUSA.xls

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	27/08/2014 15:25	AL	No	No	No	Never	100-500	No	Yes	Don't know	Not sure	No	No	Don't know	Don't know	Maybe	No	No	Some of th	No
2	2014-08-29 18:39	AL	Yes	No	Yes	Sometime: 1-5	No	No	No	Yes	No	No	No	Very diffic	No	No	No	No	No	No
3	2014-08-30 20:55	AL	Yes	Yes	Yes	Often	1-5	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Very easy	Yes	Yes	Yes	Yes	Yes
4	2014-08-27 12:18	AZ	No	No	Yes	Sometime: 6-25	No	Yes	No	Yes	No	No	No	Don't know	Somewhat	Maybe	No	Some of th	No	No
5	2014-08-27 16:29	AZ	No	No	Yes	NA	26-100	No	Yes	Yes	Yes	No	No	Yes	Don't know	Maybe	Maybe	No	No	No
6	2014-08-28 11:50	AZ	Yes	Yes	Yes	Sometime: 1-5	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Somewhat	Maybe	Yes	Some of th	No	No
7	2014-08-28 23:13	AZ	No	Yes	Yes	Sometime: 500-1000	No	Yes	Yes	Yes	No	Don't know	Yes	Somewhat	Maybe	No	Some of th	Yes	No	No
8	2014-08-27 11:33	CA	NA	Yes	Yes	Sometime: 26-100	No	No	Yes	Yes	No	No	Don't know	Somewhat	Yes	Yes	Yes	Yes	Maybe	Maybe
9	2014-08-27 11:36	CA	No	Yes	Yes	Rarely	26-100	No	Yes	Yes	Not sure	Don't know	Yes	Yes	Don't know	No	No	Yes	Yes	No
10	2014-08-27 11:37	CA	No	Yes	Yes	Rarely	6-25	Yes	Yes	Yes	Yes	Don't know	Don't know	Don't know	Don't know	No	No	Yes	Yes	Maybe
11	2014-08-27 12:35	CA	No	No	No	Rarely	1-5	Yes	Yes	No	Yes	No	No	Don't know	Very diffic	Maybe	No	Yes	Some of th	No
12	2014-08-27 12:40	CA	No	Yes	No	Never	26-100	No	Yes	Don't know	Not sure	No	No	Don't know	Very easy	No	No	Some of th	Some of th	No
13	2014-08-27 12:48	CA	No	No	No	NA	6-25	No	Yes	No	Yes	No	No	Don't know	Very diffic	Maybe	No	Yes	Yes	No
14	2014-08-27 12:51	CA	No	No	Yes	Sometime: 1-5	No	Yes	Yes	No	No	No	No	Yes	Somewhat	No	No	Yes	Yes	No
15	2014-08-27 12:53	CA	No	Yes	Yes	Never	More than No	No	Yes	Yes	Not sure	Don't know	No	Yes	Somewhat	Maybe	No	No	No	No
16	2014-08-27 13:42	CA	No	Yes	Yes	Rarely	100-500	Yes	Yes	Yes	Yes	Yes	Don't know	Yes	Very easy	No	No	Some of th	Yes	No
17	2014-08-27 13:55	CA	No	Yes	Yes	Often	26-100	No	Yes	Yes	Yes	No	No	Yes	Don't know	Maybe	No	Some of th	Some of th	No
18	2014-08-27 14:14	CA	No	No	No	Never	More than No	No	Yes	Don't know	No	No	Don't know	Don't know	Don't know	Yes	Maybe	Some of th	Some of th	No
19	2014-08-27 14:16	CA	No	No	No	NA	500-1000	No	Yes	Don't know	Not sure	No	Don't know	Don't know	Don't know	Maybe	Maybe	Some of th	No	No
20	2014-08-27 14:18	CA	No	No	No	IA	More than No	No	Yes	Yes	No	Don't know	Yes	Don't know	Don't know	No	No	Some of th	Yes	Maybe
21	2014-08-27 14:20	CA	No	No	Yes	Sometime: More than No	No	Yes	Yes	Not sure	Don't know	Don't know	Yes	Somewhat	No	No	Yes	Yes	No	No
22	2014-08-27 14:21	CA	No	No	Yes	Rarely	More than No	No	Yes	Yes	Yes	Don't know	Yes	Don't know	Somewhat	No	No	Some of th	Yes	No
23	2014-08-27 14:28	CA	No	Yes	Yes	Sometime: More than No	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Don't know	Maybe	No	Some of th	No	No
24	2014-08-27 14:30	CA	No	No	No	IA	More than No	No	Yes	Don't know	Not sure	Don't know	Don't know	Don't know	Don't know	Maybe	No	No	No	No
25	2014-08-27 14:38	CA	Yes	No	No	Sometime: More than No	No	Yes	Yes	Not sure	Don't know	No	Don't know	Very easy	Maybe	Maybe	Some of th	Some of th	No	No
26	2014-08-27 14:47	CA	No	Yes	Yes	Sometime: More than No	No	Yes	Yes	Yes	Yes	Don't know	Don't know	Don't know	Don't know	Maybe	No	Some of th	Some of th	No
27	2014-08-27 14:52	CA	No	Yes	Yes	Often	More than No	No	Yes	Yes	Not sure	Don't know	Yes	Don't know	Don't know	Maybe	No	Some of th	No	No

Figura 8: conteúdo do ficheiro outFemeasUSA.xls

Após a análise dos resultados finais, é possível concluir que os três jobs criados, estão a devolver os resultados corretos.