

Aplicações Informáticas na Biomedicina

Ficha 5

João Pimentel (a80874) Pedro Gonçalves (a82313)
Nuno Rei (a81918) Fábio Senra (a82108)

Novembro 2019

Universidade do Minho
Mestrado Integrado em Engenharia Informática

1 Pergunta 1

Neste primeiro *job* (Figura 2) foi efetuado o carregamento da totalidade do ficheiro *csv* para uma tabela num *schema* associado à Ficha em questão.

De seguida, foi executada uma projeção das colunas *country*, *age*, *family_history*, *treatment* e *seek_help*.

Tendo o resultado desta projeção, é efetuada uma filtração sobre os dados, em que se rejeitam idades inferiores a 18 anos e superiores a 50 anos, bem como *seek_help* não ambíguas, como se vê na Figura 1.

Conditions	Input column	Function	Operator	Value
	seek_help	Upper case	!=	"DON'T KNOW"
	Age	Empty	>=	18
	Age	Empty	<=	50

Figura 1 - Filtro Condicional sobre Idade e Ajuda

Aos valores que cumprem os requisitos, é aplicada uma ordenação ascendente pela idade, sendo que os registos são direcionados para um ficheiro *excel*. Já os que não cumprem, são direcionados para um filtro que ignora valores de idade inferiores a 50 anos e maiores que 100, bem como valores ambíguos de ajuda, sendo depois ordenados pela idade e direcionados para um ficheiro. Quer isto dizer que, os registos que passam no primeiro filtro correspondem a pessoas com idades entre os 18 e os 50 anos, e os que são rejeitados correspondem a pessoas com mais de 50 anos, permitindo relacionar a idade com a saúde mental.

Assim, é expectável que, com uma certa análise sobre os dados obtidos seja possível determinar uma relação entre as colunas, por exemplo, a relação entre a existência de casos de uma saúde mental precária com a relutância das pessoas na procura de tratamento, bem como com a sua idade.

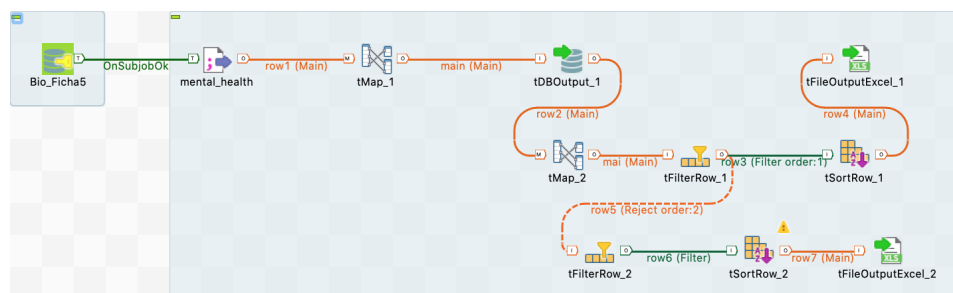


Figura 2 - Job Talend 1.

2 Pergunta 2

No segundo *job* a ideia passou por verificar de que forma são afetados os trabalhadores tendo em conta o seu local e condições de trabalho.

Tendo em conta que o *dataset* já estava carregado numa tabela, bastou utilizar a tabela em questão como *input*, como se pode ver na Figura 3.

Em seguida, foi efetuada uma projeção sobre as colunas *self_employed*, *work_interfere*, *remote_work*, *leave*, *benefits*, *seek_help*, *treatment* e *mental_vs_physical*. Note-se que era possível fazer esta projeção logo na importação da BD, no entanto foi decidido que seriam utilizadas maioritariamente as funcionalidades do *Talend* (*tMap*).

Tendo apenas as colunas necessárias, são efetuadas duas filtrações. A primeira retira todos os registos com valores abstratos, ou seja, valores de incerteza como *Don't Know* e *NA*, obtendo-se um *dataset* menos ambíguo. O segundo filtro diferencia os trabalhadores por conta própria, dos trabalhadores por conta de outrem.

Sendo assim, é possível relacionar aspetos inerentes aos locais de trabalho, juntamente com a facilidade e vontade em encontrar ajuda para os problemas de saúde mental. Além disso, devido à filtração efetuada, pode-se relacionar o tipo de trabalhador (por conta própria ou outrem) com o seu estado mental.

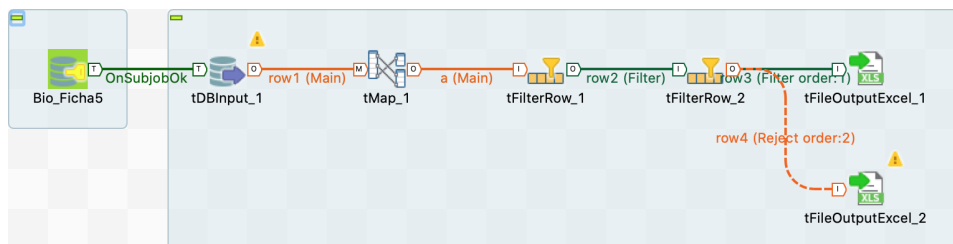


Figura 3 - *Job Talend 2*.

3 Pergunta 3

O terceiro *job* teve como objetivo fazer uma análise sobre o quão confortáveis os trabalhadores se sentem em discutir assuntos relacionados com saúde mental no seu local de trabalho, não só com colegas, mas também com os seus supervisores. Foi, também, considerado que a divisão entre organizações tecnológicas e não tecnológicas seria um ponto de interesse para a análise a efetuar.

Dito isto, foram projectadas as colunas *mental_health_consequence*, *coworkers*, *phys_health_consequence* e *supervisor*. Esta projecção teve o intuito de permitir uma filtragem para obter registos com propensão a discutir problemas de saúde mental com indivíduos do seu local de trabalho e, também, sujeitos da opinião de que a discussão deste tipo de problemas não traria consequências negativas na carreira profissional.

Em seguida, foi aplicado um filtro aos resultados obtidos anteriormente, de forma a separar as organizações tecnológicas das restantes. Assim, os resultados foram separados em dois ficheiros *excel*, com várias outras colunas para além das mencionadas.

Os resultados deste *job* demonstram que 138 registos estão associados a empresas da área tecnológica, restando apenas 14 casos contrários. Se o padrão se mantiver para um maior número de registos, poderá ser visível a evolução na prevenção e ajuda perante problemas de saúde mental por parte deste tipo de empresas, fornecendo um ambiente que permite um certo à vontade para discussão destes assuntos. Outra análise interessante seria uma filtragem pelas colunas *benefits* e *care_option*, permitindo, idealmente, a identificação das empresas que dedicam mais tempo à prevenção e melhoria de problemas associados à saúde mental dos seus funcionários.

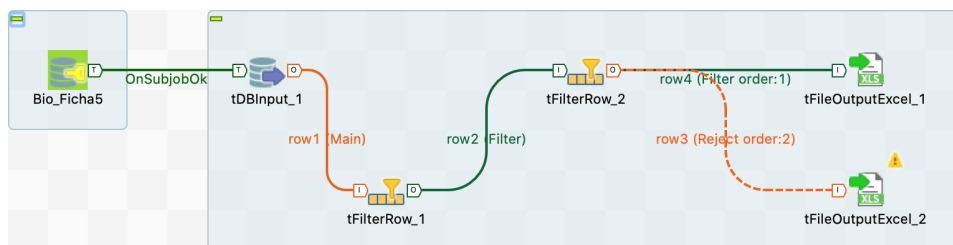


Figura 4 - Job Talend 3.