



UNIVERSIDADE DO MINHO

**Mestrado Integrado em Engenharia Informática**

APLICAÇÕES INFORMÁTICAS NA BIOMEDICINA

FICHA 5

**Grupo:**

Guilherme Andrade - A80426

Guilherme Nunes - A80524

Mateus Silva - A81952

Rafaela Soares - A79034

Ricardo Caçador - A81064

Braga, Portugal  
19 de Novembro de 2019

## 0.1 Job 1

O principal objetivo da criação deste *job* recai sobre a necessidade de se elaborar uma "limpeza" do *dataset* fornecido, antes de este ser inserido numa tabela do *schema* criado para a resolução desta ficha. Para isso, foi necessário realizar uma análise sobre os campos que este *dataset* possui.

A partir desta análise, chegou-se à conclusão que o Género e a Idade foram campos de texto livre no questionário, pelo que o utilizador tinha liberdade total de indicar o termo com que se identifica.

Por esse mesmo motivo, este *dataset* possuía 42 géneros distintos, alguns destes perceptíveis como sendo *Male* ou *Female* ou outro qualquer género, mas escritos de forma distinta. Estes foram corrigidos para ficarem uniformizados, contudo outros simplesmente não correspondiam a nenhum género, pelo que foram removidos.

No que concerne à Idade, foi removida do *dataset* a informação de questionários com idades negativas.

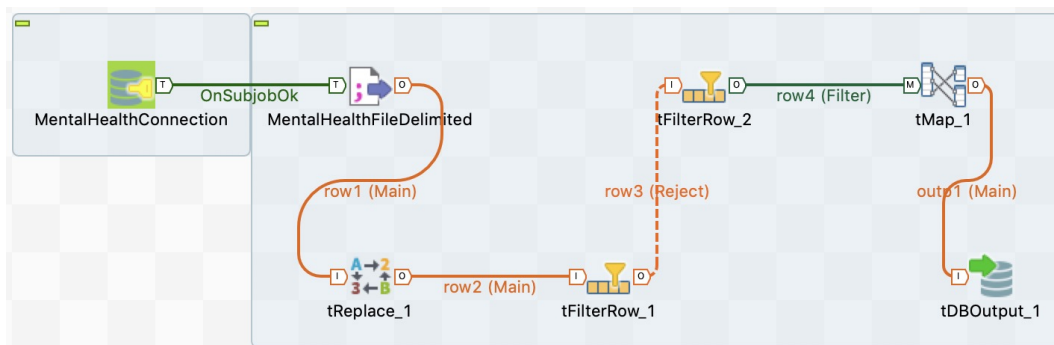


Figura 1: Job 1

No que diz respeito à concretização deste *job* no *Talend*, este começa com a conexão à base de dados usando o componente *tDBConnection*. Seguidamente, em caso de sucesso da operação anterior e usando o componente *tFileInputDelimited*, é aberto o ficheiro *csv* com o *dataset* e criado o *schema* dos dados conforme o tipo que este possuem no ficheiro.

Após isto, é feita a "limpeza" dos dados referida anteriormente, começando com a substituição dos géneros. Para tal, é usado o componente *tReplace*, seguindo-se de dois filtros para os géneros e idades inválidos, respetivamente. Para esta última operação foi usado o componente *tFilterRow*.

Para concluir, é usado um *tMap* que faz a correspondência entre o *schema* de dados criado anteriormente e o *schema* da tabela que irá ser guardada na base de dados.

Por fim, usando um *tDBOutput*, os dados são guardados na base de dados, à qual estamos conectados desde o início do processo, e guardados numa tabela especificada neste componente, tendo já sido aplicado o processo de "limpeza" de dados.

Desta forma, os próximos *jobs* podem usar já a tabela criada e que contém informação mais correta e intuitiva do que o *dataset* inicial.

## 0.2 Job 2

A partir do *job* referido anteriormente, considerou-se relevante analisar os dados relativos ao conhecimento dos funcionários acerca da disponibilização de benefícios de *mental health* que a empresa de Tecnologias de Informação(TIs), onde exercem funções, proporciona.

Esta análise é fulcral pois, cada vez mais, o tema esgotamento mental parece afirmar-se neste contexto laboral. Este pode provir de cariz pessoal, mas também do ambiente de *stress* que este tipo de empresa apresenta, uma vez que se exige uma carga cognitiva considerável e pelo facto de que as pessoas desejam que tudo seja feito o mais rápido possível.

A denominação dos tipos de conhecimento foi conceptualizada em conhecer ou não o facto de que a empresa disponibiliza benefícios de *mental health*.

Nesse sentido, utilizando a componente *tDBInput*, acedeu-se à base de dados, povoada devidamente pelo *job* 1. De seguida, inseriu-se novas alterações na base de dados, recorrendo à componente *tReplace*. Estas consistiram em trocar, na coluna *benefits*, os termos "*No*" e "*Don't know*" por "*Nao conhece*" e "*Yes*" por "*Conhece*".

De seguida, utilizou-se o componente *tAggregateRow*, de forma a agrupar no *output* os tipos de conhecimento acerca da disponibilização de benefícios e o número de pessoas associado aos mesmos.

Através do componente *tSortRow*, ordenou-se, de forma ascendente, a quantidade de pessoas associada a cada tipo de conhecimento.

No intuito de a informação ser extraída para um ficheiro *Excel*, recorreu-se à componente *tFileOutputExcel*.

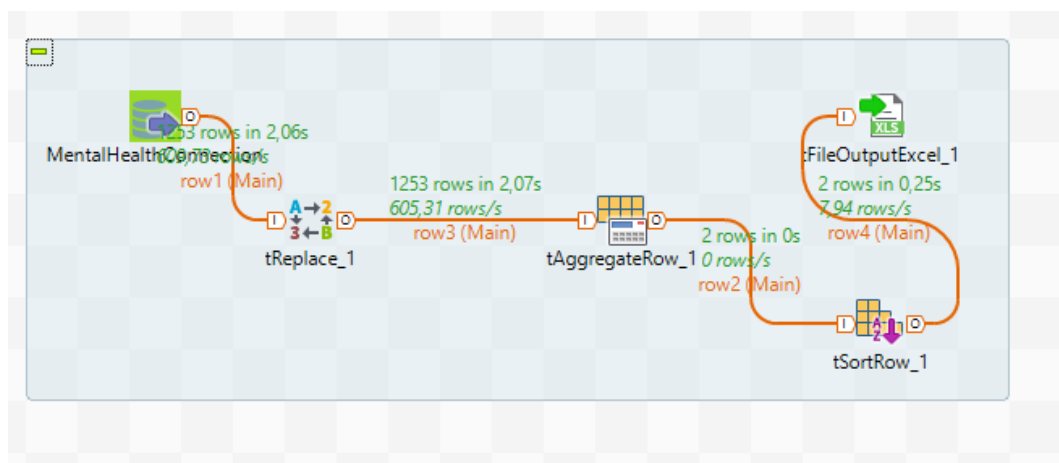


Figura 2: Job 2

	A	B
1	conhecimentoQueEmpresaFornecaBeneficios	numeroDePessoas
2	Conhece	473
3	Nao conhece	780
4		
5		
6		
7		

Figura 3: *Output* do Job 2

A partir do ficheiro produzido como *output*, evidencia-se que mais de metade dos funcionários (62,25%), alvo de estudo, desconhecem que a empresa dispõe de benefícios de *mental health*.

Tal facto é algo preocupante, pois os funcionários, como são submetidos a prazos de entrega e a *stress* profissional, deveriam de zelar pelo seu bem-estar.

Em suma, pode-se extrair, desta análise, que as empresas de Tecnologias de Informação(TIs) devem dar primazia à divulgação da disponibilização de benefícios de *mental health* e incentivar a que os seus colaboradores usufruam dos mesmos.

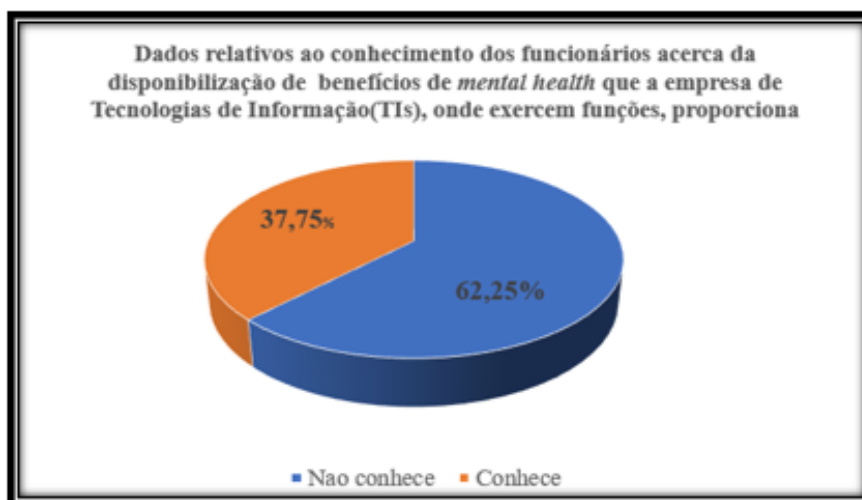


Figura 4: Dados relativos ao conhecimento dos funcionários acerca da disponibilização de benefícios de *mental health* que a empresa de Tecnologias de Informação(TIs), onde exercem funções, proporciona

### 0.3 Job 3

A realização deste *Job* tem como objetivo aprofundar conhecimentos em Talend, relativamente à análise e à manipulação de um determinado conjunto de dados.

Em concreto, o intuito deste *Job* é a apresentação de um *BarChart* relativo à amostra das idades da população que apresentam ter *MentalHealth* por país.

Tal apresenta extrema relevância, uma vez que se pode extrair quais as faixas etárias que apresentam maior incidência de problemas de saúde mental.

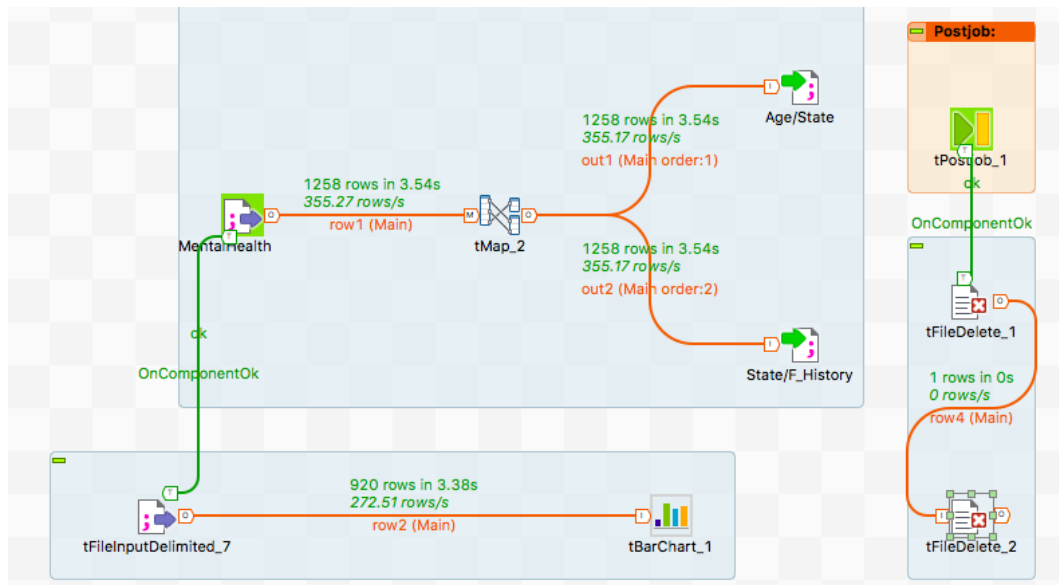


Figura 5: Job 3

Este *Job* utiliza o *dataSet Mentalhealth.csv*, pois os atributos utilizados não foram alterados pelo *Job1*, respetivamente a idade e o país.

Numa primeira fase, utiliza-se um *tMap* para se separar os atributos mencionados dos restantes e todos os dados são guardados num ficheiro temporário.

De referir que o *BarChart* não é feito sequencialmente com o *tMap*, pois o objetivo deste *Job* é que possam ser feitas várias análises simultâneas aos dados (*Age/State* e *State/FamilyHistory*, por exemplo).

Para além disso, o intuito também remete para a ocorrência de um *append* de todos estes ficheiros temporários em apenas um, que posteriormente seria usado para gerar a imagem. Contudo, devido ao elevado número de atributos nominais, apenas foi feita uma análise da idade média por país.

Após esta operação ter terminado, utiliza-se um *fileInputDelimiter* para ler o ficheiro temporário e gera-se *BarChart* com essa informação, como se pode observar na seguinte imagem 6.

Para terminar, é efetuado um *PostJob* que efetua uma limpeza dos ficheiros temporários gerados anteriormente.

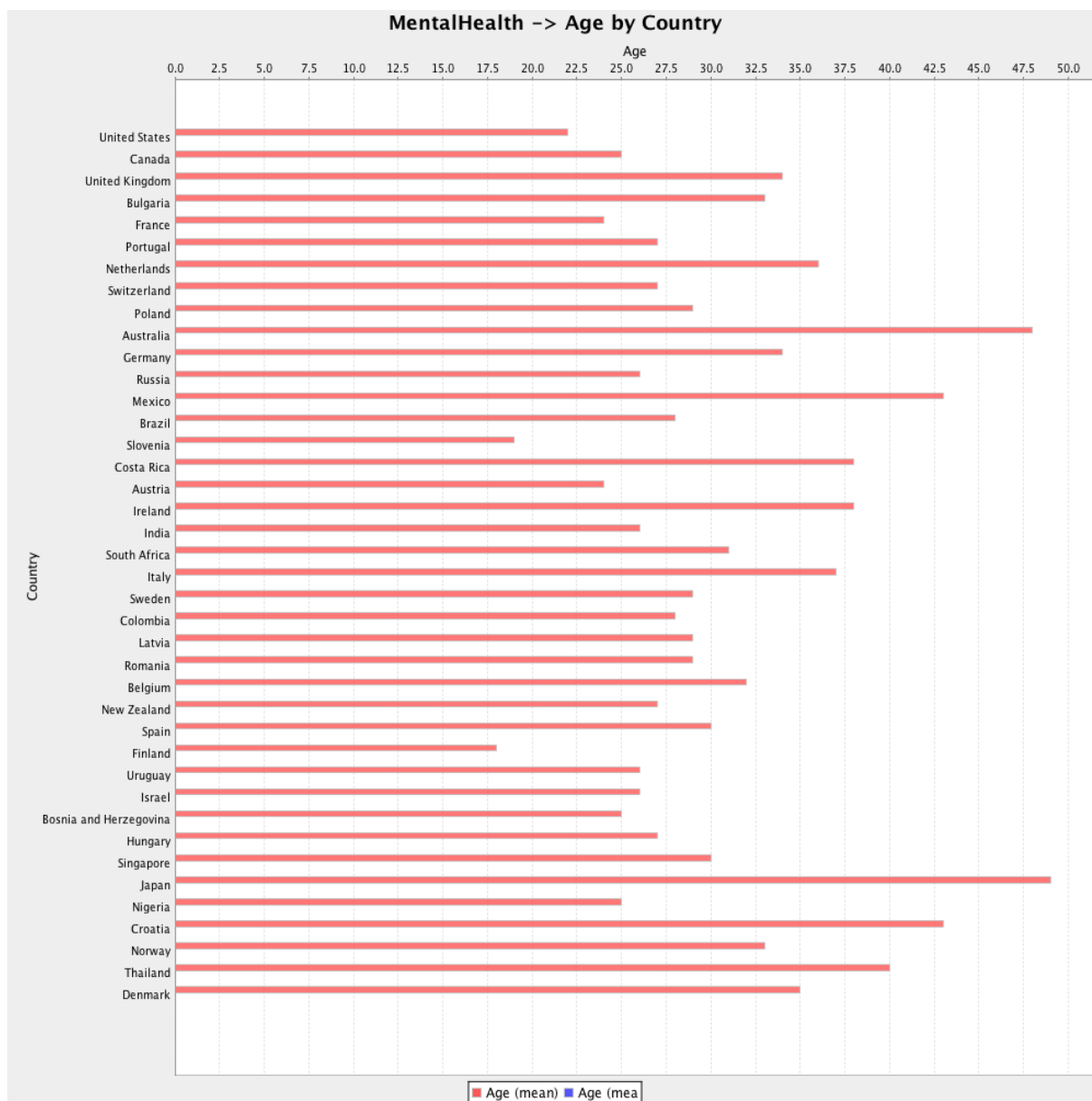


Figura 6: Idade da população por país