

# **KNIME**

Création des graphiques de diagnostic sur l'égalité professionnelle femmes - hommes



### **Sommaire**

### **1** Contexte

- Diagnostic égalité professionnelle femmeshommes
- RGPD

### 3 Outil ETL

- Principales étapes de transformation des données
- Output

### 2 KNIME

- Présentation de l'outil
- Présentation du workflow

### 4 Outil de visualisation

Présentation et interprétation des graphiques générés



### **5 Conclusion**

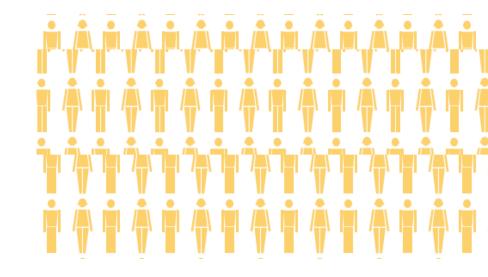
- Concernant l'outil
- Concernant l'index
- « égalité professionnelle femmes – hommes »

# CONTEXTE

### **Contexte**

### Diagnostic égalité professionnelle femmes -hommes :

- obligation : calcul et publication annuelle de l'index égalité femmes – hommes
  - l'index est calculé au moyen de **cinq indicateurs : rémunération**, **augmentation**, **promotion**, **parité** parmi les 10 **plus hauts salaires**, taux d'**augmentation** suite à **congés maternité**
- recommandation : produire un **diagnostic** sexué et une analyse de la situation respective des femmes et des hommes



### **Contexte**

### RGPD:

Les 5 grands principes des règles de protection des données personnelles :

- principe de finalité
- principe de **proportionnalité** et de **pertinence**
- principe d'une durée de conservation limitée
- principe de sécurité et de confidentialité
- respect des droits des personnes



# KNIME

# KNIME: présentation de l'outil

- Logiciel libre et open-source
- Version 4.4.1
- L'interface utilisateur graphique permet la construction de workflow par l'assemblage de nœuds réalisant une opération spécifique
- Opérations réalisées grâce aux nodes : le formatage des données (ETL: Extraction, Transformation, Chargement), l'analyse, la visualisation des résultats au sein de la même interface
- Il est également possible de concevoir ses propres nodes KNIME en Java



# KNIME: présentation du workflow

xlsx files are pasted in the "files input" folder, located in the "gender equality" folder

### Workflow gender equality:

- generate graphs on gender equality indicators write CSV file with anonymous data

#### File readers:

Excel reader nodes x3:

- info pro.xlsx
- remuneration.xslx
- salaries.xlsx

Each file has 256 rows Useless columns are discarded, data types are checked and modified if necessary



#### Data explorers:

- data info pro
- data remuneration
- data salaries
- Data explorer nodes data (ex: statistical indicators, missing

Data explorers metanode

- Data explorer nodes x3:
- display properties of input

values...)

#### Data transformation:

- Calculate"age"
- Binning "data values" (age, seniority)
- Calculate "salaire total"

**Data transformations** metanode

#### Data joiners:

lines

Joiner nodes x2: Inner join on "id salariés" Join result table has **256** 

Data ioiners metanode

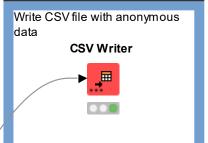
#### Columns transformers:

- Column filter: useless columns are discarded
- Column sorter: columns are organized before writing csv file
- Column combiner: columns "service" and "sex" are combined to be used in graphics views

Metanode columns transformations



CSV file is written in the gender\_equality\_workflow data folder. You can change this location by specifying a new one in the node "output location" setting



xlsx files are deleted from the "files input" folder, original datas are no longer available. To replay the worflow, paste the xlsx files into the "files input" folder

#### Delete Files/Folders

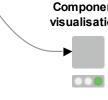


To open gender equality report: => right click on Component visualisation node

=> click "Interactive view"

Displays gender equality graphs usina Views nodes.

Pivoting and Groupby nodes are used to produce indicators per gender and also per department and gender / age and gender



Component visualisation

8

# ETL

# KNIME: présentation du workflow

xlsx files are pasted in the "files input" folder, located in the "gender equality" folder

### Workflow gender equality:

- generate graphs on gender equality indicators write CSV file with anonymous data

#### File readers:

Excel reader nodes x3:

- info pro.xlsx
- remuneration.xslx
- salaries.xlsx

Each file has 256 rows Useless columns are discarded, data types are checked and modified if necessary

File readers metanode



#### Data explorers:

- data salaries
- data (ex: statistical indicators, missing values...)

Data explorers metanode

- Data explorer nodes x3:
- data remuneration
- Data explorer nodes display properties of input

#### - data info pro - Binning "data

- values" (age, seniority)
  - Calculate "salaire total"

Calculate"age"

**Data transformations** metanode

#### Data transformation: Data joiners:

Joiner nodes x2:

### metanode

Inner join on "id salariés" Join result table has **256** lines

Data ioiners

#### Columns transformers:

- Column filter: useless columns are discarded
- Column sorter: columns are organized before writing csv file
- Column combiner: columns "service" and "sex" are combined to be used in graphics views

Metanode columns transformations



CSV file is written in the gender\_equality\_workflow data folder. You can change this location by specifying a new one in the node "output location" setting



**.**=

Delete Files/Folders

folder

xlsx files are deleted from the

datas are no longer available.

xlsx files into the "files input"

To replay the worflow, paste the

"files input" folder, original



To open gender equality report: => right click on Component visualisation node

=> click "Interactive view"

Displays gender equality graphs usina Views nodes.

Pivoting and Groupby nodes are used to produce indicators per gender and also per department and gender / age and gender

> Component visualisation

### **KNIME: readers**

















Fichiers sources provenant du SIRH au format xlsx

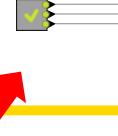
#### File readers:

Excel reader nodes x3:

- info\_pro.xlsx
- remuneration.xslx
- salaries.xlsx

Each file has **256** rows
Useless columns are
discarded, data types are
checked and modified if
necessary

File readers metanode



#### **Excel Reader**





read\_salaries

#### **Excel Reader**

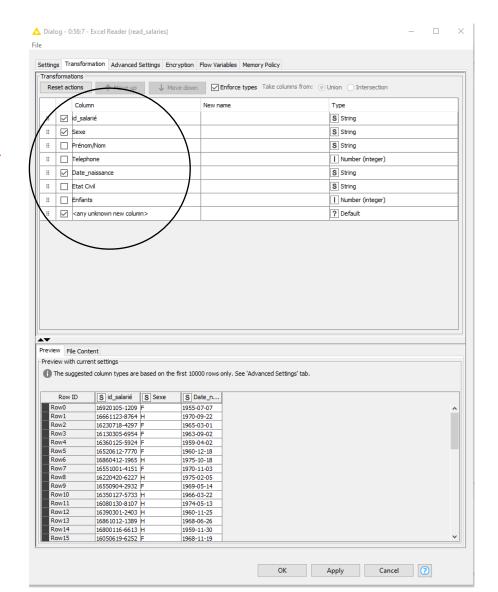


read\_remuneration

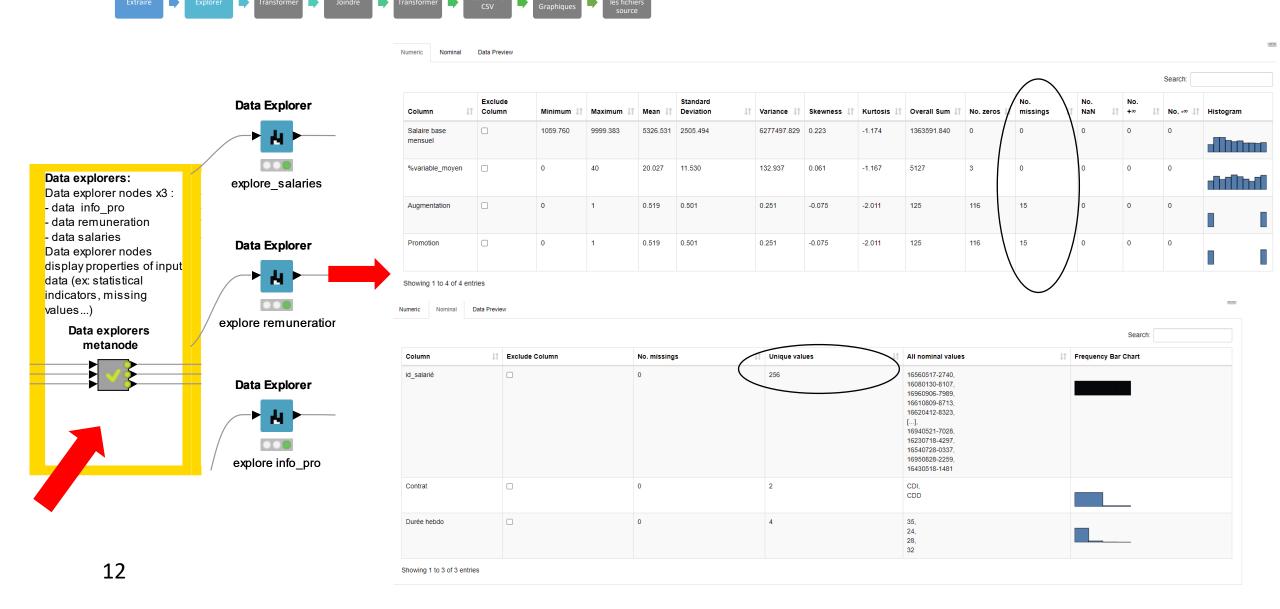
**Excel Reader** 



read info pro

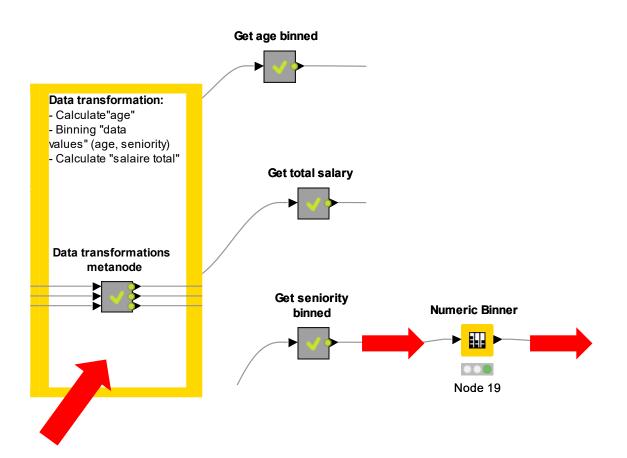


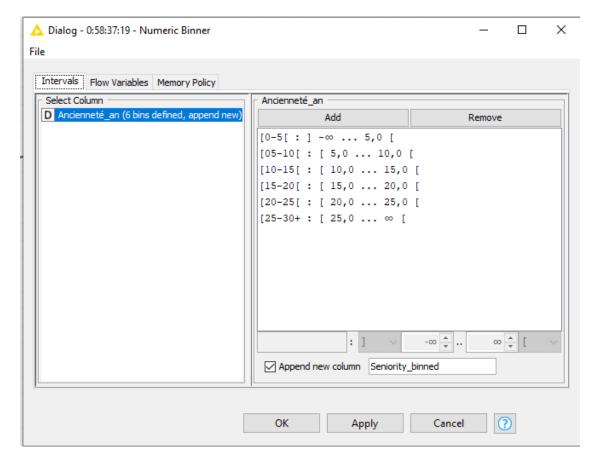
# **KNIME:** data explorers



### **KNIME: data transformation**

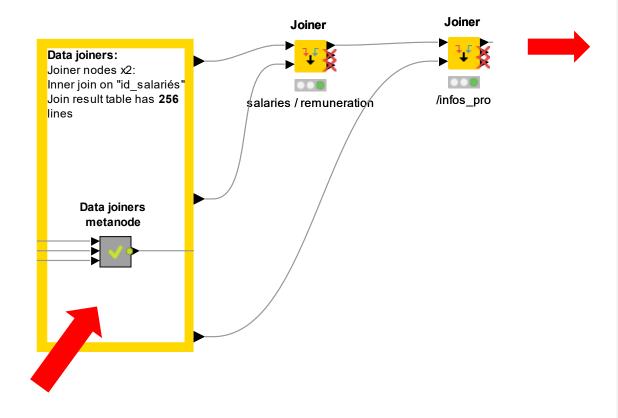


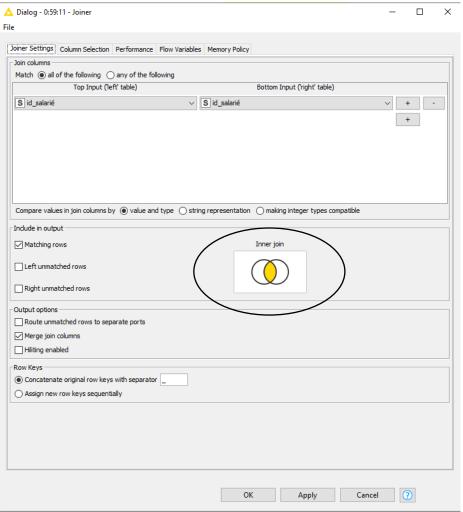




# **KNIME: joiners**

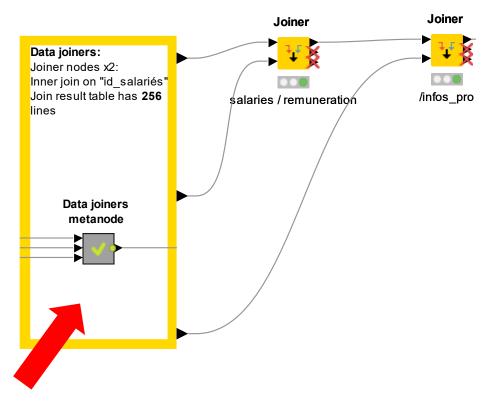


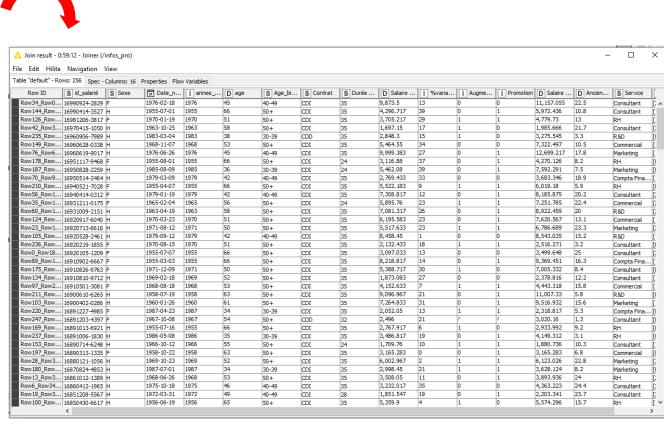




# **KNIME: joiners**

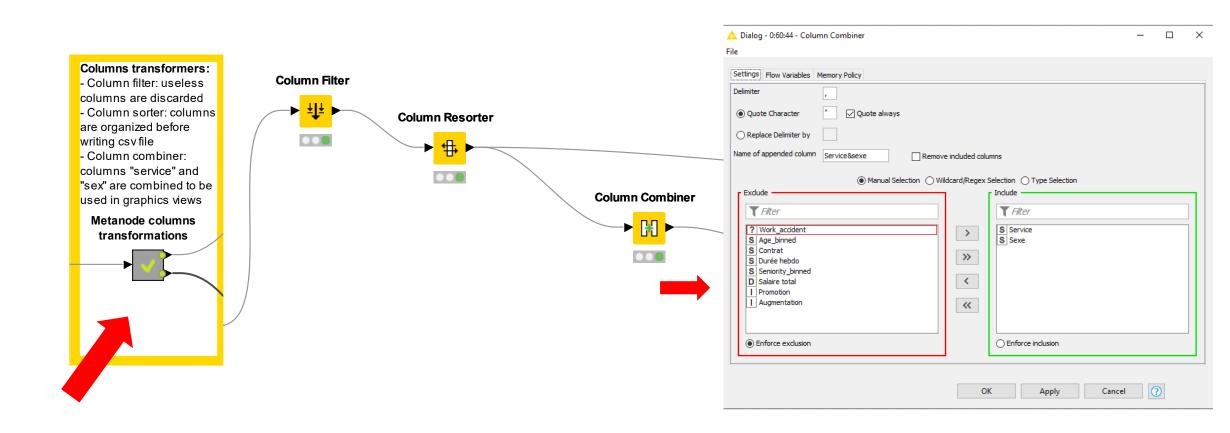






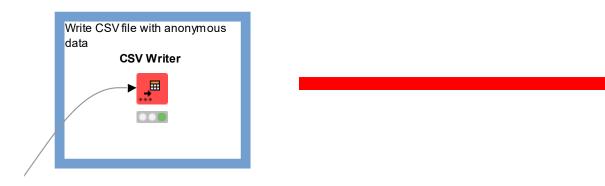
### **KNIME: columns transformers**

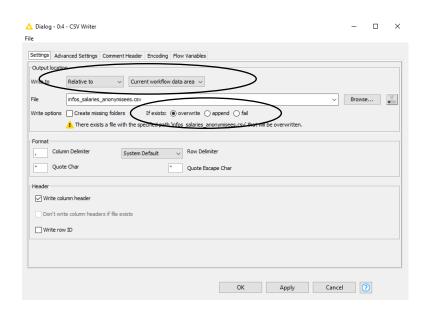


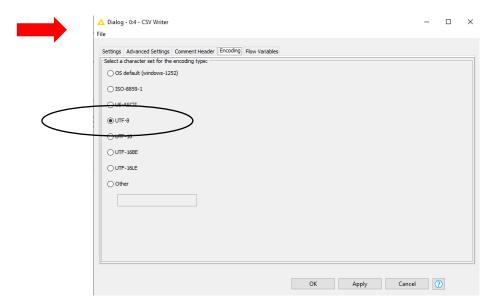


### **KNIME: CSV writer**









### **KNIME: CSV writer**







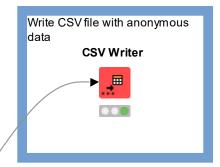












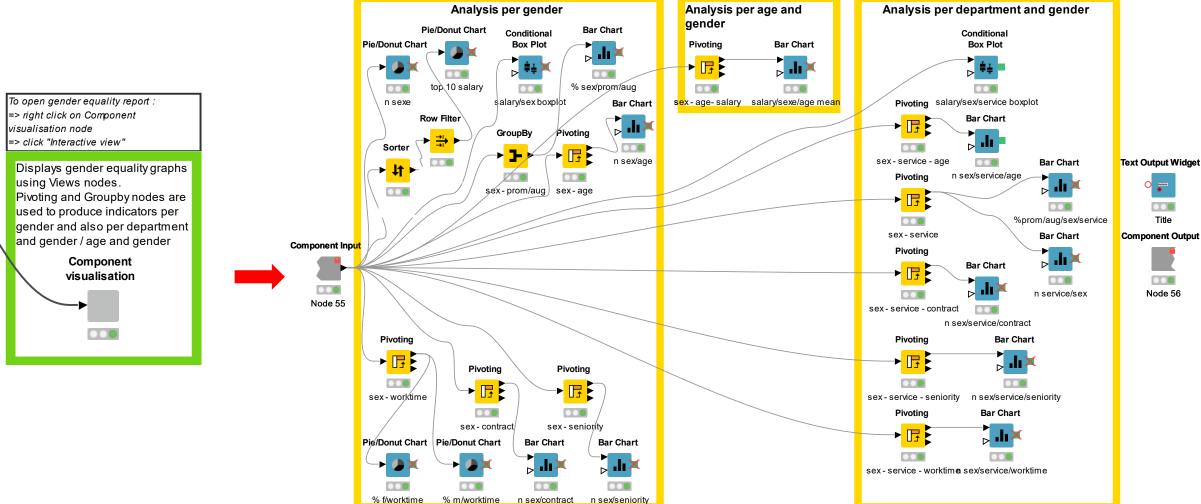
Rappel RGPD : principe de **finalité**, principe de **proportionnalité** et de **pertinence** , principe d'une durée de **conservation limitée** 

```
🚽 infos salaries anonymisees.csv 🗵
     "Service", "Sexe", (Age binned") "Contrat", "Durée hebdo", (Seniority binned") "Salaire total", "Promotion", "Augmentation"
                         "40<del>-49","CDI</del>","35","[20-25[",11157,0,0
     "Consultant", "H", "50+", "CDI", "35", "[10-15[", 5972, 1, 0
     "RH", "F", "50+", "CDI", "35", "[10-15[", 4780, 1, 1
     "Consultant", "H", "50+", "CDI", "35", "[20-25[", 1986, 0, 1
     "R&D", "H", "30-39", "CDD", "35", "[0-5[", 3276, 0, 1
     "Commercial", "H", "50+", "CDI", "35", "[10-15[", 7322, 0, 0
     "Marketing", "H", "40-49", "CDI", "35", "[15-20[", 12699, 1, 0
     "RH", "F", "50+", "CDI", "24", "[05-10[", 4270, 1, 0
    "Marketing", "H", "30-39", "CDI", "24", "[05-10[", 7592, 1, 0
    "Compta Finances", "H", "40-49", "CDI", "35", "[15-20[", 3683, 0, 0
    "RH", "F", "50+", "CDI", "35", "[05-10[", 6019, 1, 1
    "Consultant", "F", "40-49", "CDI", "35", "[20-25[",8186,1,0
     "Commercial", "F", "50+", "CDI", "24", "[20-25[", 7252, 1, 1
    "R&D","H","50+","CDI","35","[20-25[",8922,1,0
   "Commercial", "H", "50+", "CDI", "35", "[10-15[", 7621, 1, 0
    "Marketing", "H", "50+", "CDI", "35", "[20-25[", 6787, 1, 1
    "R&D","H","40-49","CDI","35","[15-20[",8543,0,0
    "Consultant", "F", "50+", "CDI", "35", "[0-5[", 2516, 1, 1
     "Consultant", "F", "50+", "CDI", "35", "[25-30+", 3500, 0, 0
    "Compta Finances", "F", "50+", "CDI", "35", "[15-20[", 9369, 1, 0
    "Consultant", "F", "50+", "CDI", "35", "[05-10[", 7005, 0, 1
    "Consultant", "H", "50+", "CDI", "35", "[10-15[", 2379, 0, 0
    "Commercial", "F", "50+", "CDI", "35", "[15-20[", 4443, 1, 1
    "R&D","H","50+","CDI","35","[05-10[",11007,1,0
26 "Marketing", "H", "50+", "CDI", "35", "[15-20[", 9517, 1, 0
27 "Compta Finances", "F", "30-39", "CDI", "35", "[05-10[", 2319, 1, 1
    "Consultant", "F", "50+", "CDD", "32", "[0-5[", 3020,,
    "RH", "H", "50+", "CDI", "35", "[05-10[", 2934, 0, 1
    "RH", "H", "30-39", "CDI", "35", "[0-5[", 4149, 1, 0
     "Consultant", "H", "50+", "CDI", "24", "[10-15[", 1881, 1, 1
    "Commercial", "F", "50+", "CDI", "35", "[05-10[", 3165, 1, 0
    "Marketing", "H", "50+", "CDI", "35", "[20-25[", 6123, 1, 1
    "Marketing", "H", "30-39", "CDI", "35", "[05-10[", 3628, 1, 1
    "RH", "H", "50+", "CDI", "35", "[20-25[", 3894, 1, 0
    "Consultant", "H", "40-49", "CDI", "35", "[20-25[", 4363, 0, 0
37 "Consultant", "H", "40-49", "CDT", "28", "[20-25[", 2203.1.0
```

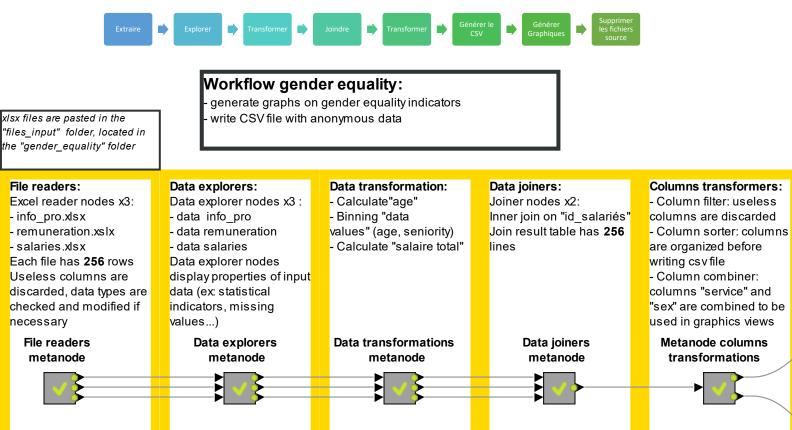
18 Normal text file length: 13 821 lines: 257 Ln: 1 Col: 66 Pos: 67 Windows (CR LF) UTF-8

### **KNIME: views nodes**





### KNIME: delete files node



CSV file is written in the gender\_equality\_workflow data folder.
You can change this location by specifying a new one in the node "output location" setting

Write CSV file with anonymous data

**CSV Writer** 



xlsx files are deleted from the "files\_input" folder, original datas are no longer available. To replay the worflow, paste the xlsx files into the "files\_input" folder

#### Delete Files/Folders



To open gender equality report :
=> right click on Component
visualisation node
=> click "Interactive view"

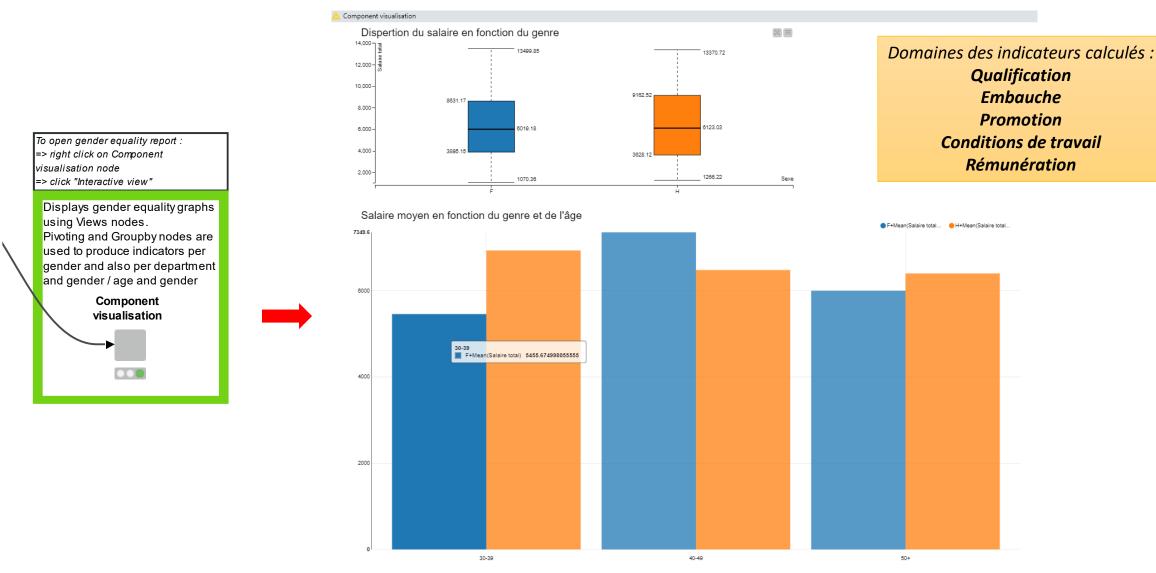
Displays gender equality graphs using Views nodes.
Pivoting and Groupby nodes are used to produce indicators per gender and also per department and gender / age and gender

### Component visualisation

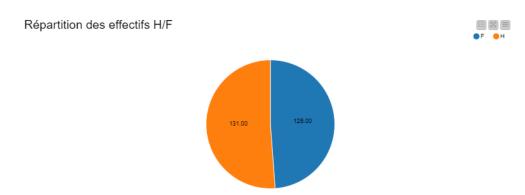
**—** 

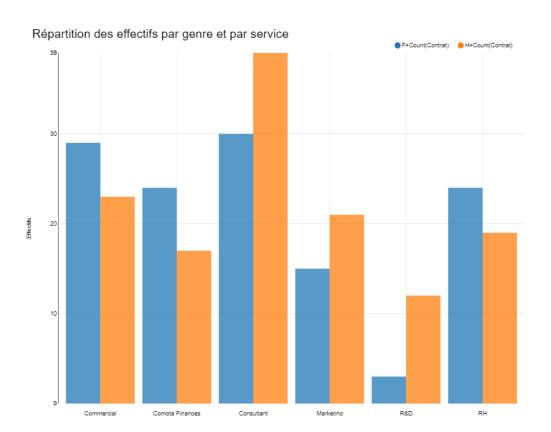
### Outil de visualisation

### KNIME: outil de visualisation



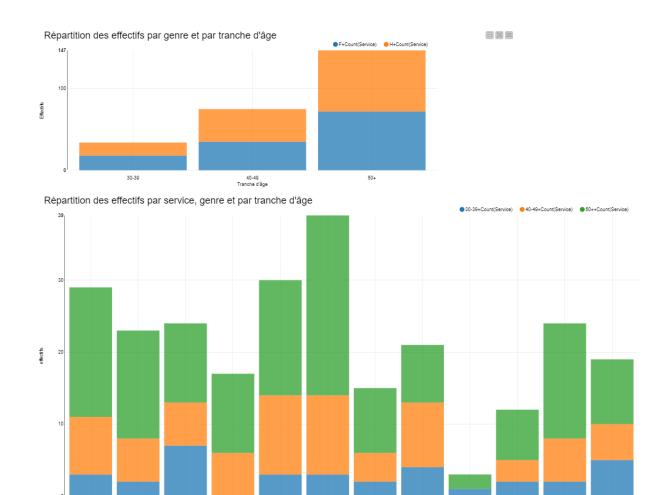
## QUALIFICATION : répartition des effectifs





- 256 employés : 125 femmes, 131 hommes
- Services majoritairement féminins : commercial, compta finance, RH
- Services majoritairement masculins : consultant, marketing, R&D

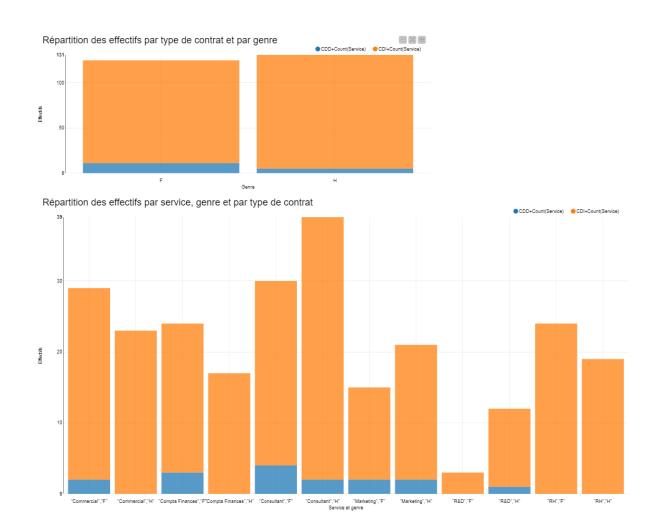
# EMBAUCHE : répartition des effectifs



- La tranche d'âge 50 + compte le plus grand nombre de salariés
- Pas de différence marquée dans la répartition des effectifs par tranche d'âge entre les femmes et les hommes

"Commercial", "H" "Compta Finances", "F"Compta Finances", "H" "Consultant", "F"

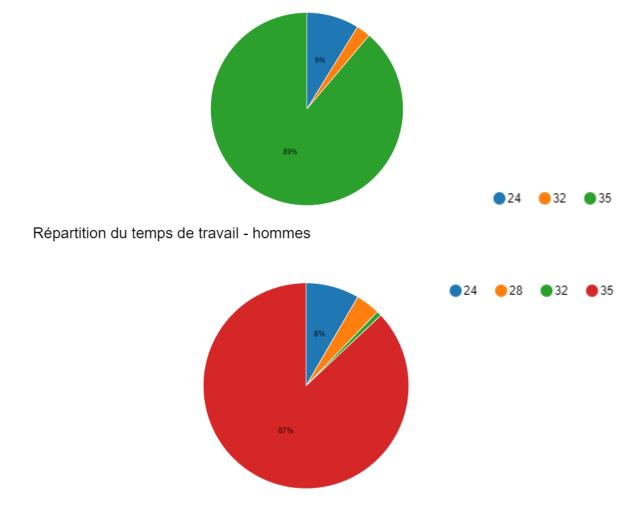
### **EMBAUCHE: types de contrats**



- La part de salariés en CDD est plus importante chez les femmes que chez les hommes
- Le recours au CDD a lieu dans tous les services, excepté en RH
- Au sein des services commercial et compta finances les CDD concernent uniquement les femmes

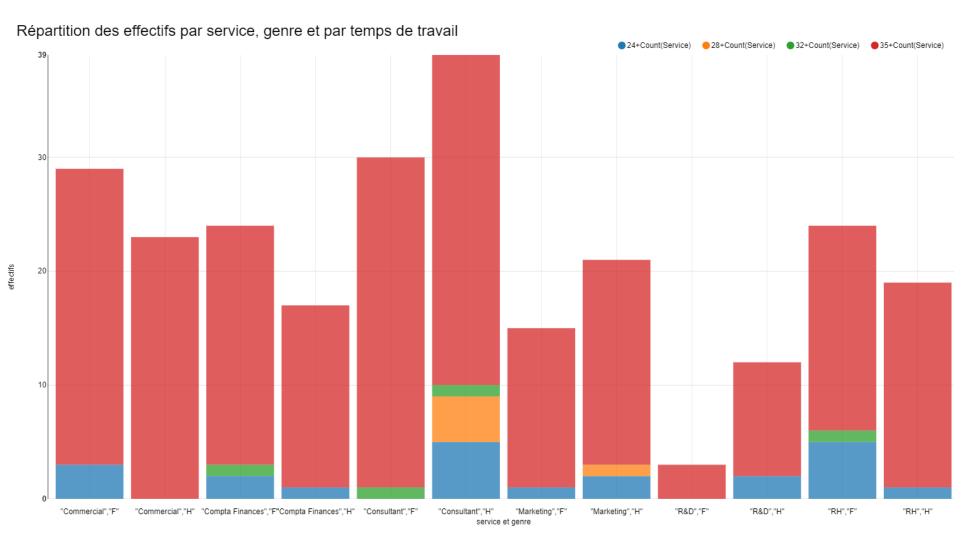
### **CONDITIONS DE TRAVAIL : durée du travail**

Répartition du temps de travail - femmes



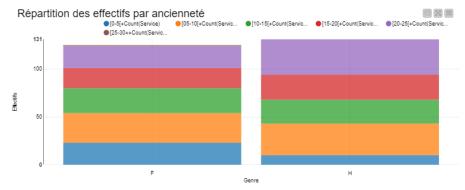
13 % des hommes exercent un temps partiel contre 10 % des femmes

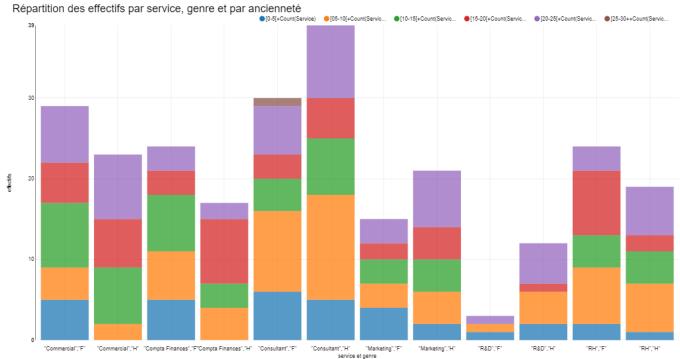
### **CONDITIONS DE TRAVAIL : durée du travail**



- Le temps partiel est plus utilisé chez les hommes que chez les femmes dans les services majoritairement masculins : consultant, marketing, R&D
- Le temps partiel est plus utilisé chez les femmes que chez les hommes dans les services majoritairement féminins : commercial, compta finance, RH

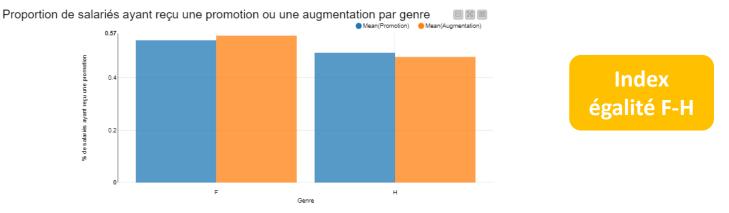
### PROMOTION: ancienneté





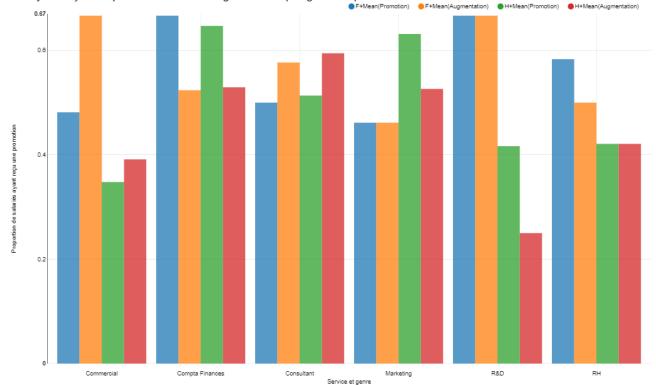
Globalement, les hommes ont plus d'ancienneté que les femmes au sein de l'entreprise, et ce au sein de tous les services

### **PROMOTION: augmentations et promotions**



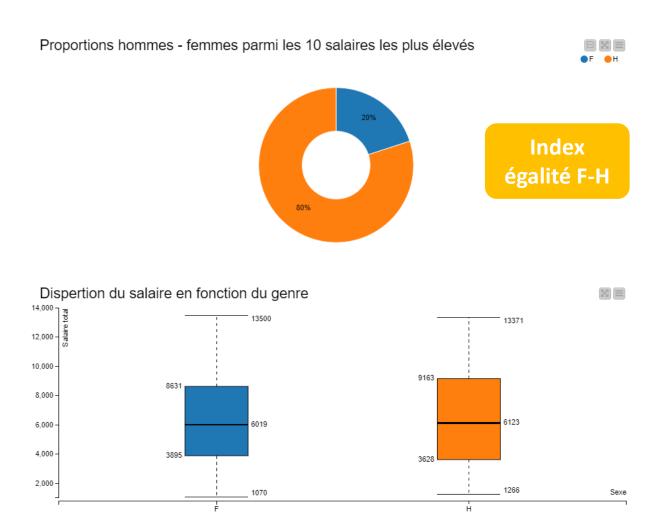
Proportion de salariés ayant reçu une promotion ou une augmentation par genre et par service

29



- Les parts de salariés ayant reçu une promotion / une augmentation sont plus importantes chez les femmes que chez les hommes
- Services où les parts de femmes ayant reçu une / augmentation sont plus importantes : commercial, R&D, RH
- Services où les parts d'hommes ayant reçu une promotion / augmentation sont plus importantes : marketing

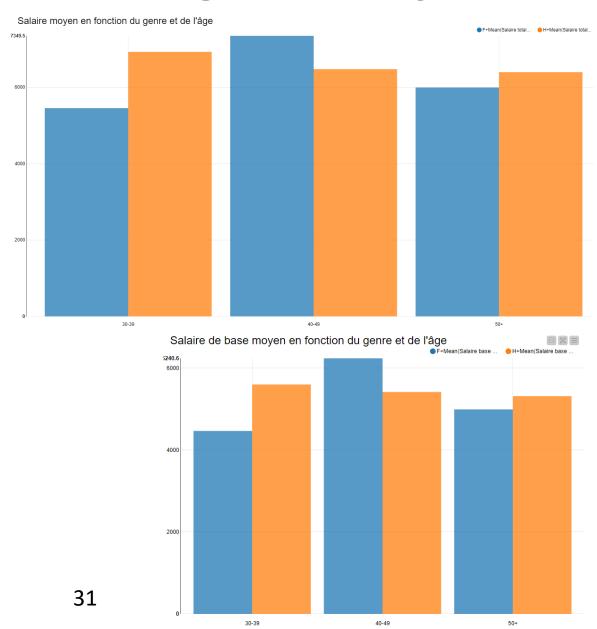
### REMUNERATION EFFECTIVE



- Parmi les 10 plus hauts salaires 8 salaires sur 10 sont des hommes
- La dispersion du salaire est globalement similaire chez les femmes et chez les hommes, bien qu'on remarque une étendue plus importante chez les femmes ainsi qu'un salaire médiant et moyen légèrement plus bas

Salaire moyen F	Salaire moyen H
6297 €	6486 €

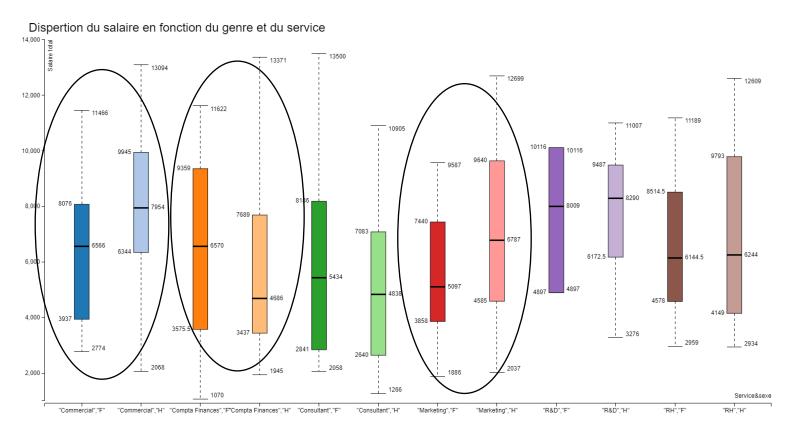
### REMUNERATION EFFECTIVE



Index égalité F-H

C'est parmi les plus jeunes que la différence de salaire moyen est la plus marquée, en faveur des hommes

### REMUNERATION EFFECTIVE



- Services au sein desquels le salaire médiant est plus élevé chez les femmes que chez les hommes : compta finances (malgré des temps partiels plus répandus chez les femmes), consultant (temps partiels plus répandus chez les hommes)
- Services au sein desquels le salaire médiant est plus élevé chez les hommes que chez les femmes : commercial (temps partiels plus répandus chez les femmes), marketing (malgré des temps partiels plus répandus chez les hommes), RH

### Index égalité professionnelle

### **Calcul index égalité femmes - hommes**

Ecart de rémunération : 39/40

Ecart augmentation : 20/20

Ecart promotion: 15/15

Augmentation suite à maternité : non calculable

Parité 10 plus hautes rémunérations : 5/10

Index estimé = 79/85 → 93/100



# CONCLUSION

# Principaux éléments d'analyse

### **Egalité professionnelle femmes - hommes**



- Index estimé à 93/100, supérieur au seuil réglementaire fixé à 75
  - → pour un calcul exhaustif de l'index : ajouter les données concernant la CSP des salariés ainsi que les informations concernant les augmentations dans l'année suivant un congés maternité
  - → pas d'obligation de mise en œuvre d'un plan d'action
- Toutefois il pourrait être intéressant de mener des analyses complémentaires et actions le cas échéant concernant :
  - la parité parmi les plus hautes rémunérations
  - la répartition des effectifs par genre au sein du service R&D
  - le recours au CDD

### **KNIME**



- Outil no code facile à prendre en main et répondant au besoin
- Niveau de contrôle et personnalisation des graphiques limité



# Merci de votre attention