

Analysez des indicateurs de l'égalité femme-homme avec Knime 5.1.0

Préparation des données avec le workflow KNIME

Conformité du fichier csv au RGPD

Index de l'égalité professionnelle

Indicateurs de diagnostic Egalité



Contexte

TransDigitConsulting

Le cabinet de consultant est spécialisé dans la transformation digitale des entreprises. Avec plus de 200 salariés, le cabinet est en plein développement. Le recrutement de consultants expérimentés devient donc un véritable enjeu stratégique.

La mission

En tant que Data Analyst, Laura me confie la mission de calculer l'index de l'égalité Femmes/Hommes du Cabinet

- Automatiser la création d'un rapport diagnostique sur l'égalité professionnelle femmes hommes.
- Générer un fichier csv respectant le RGPD.
- Création du workflow et des graphiques de diagnostic.

Analysez des indicateurs de l'égalité femme-homme avec Knime

Préparation des données avec le workflow KNIME

Conformité du fichier csv au RGPD



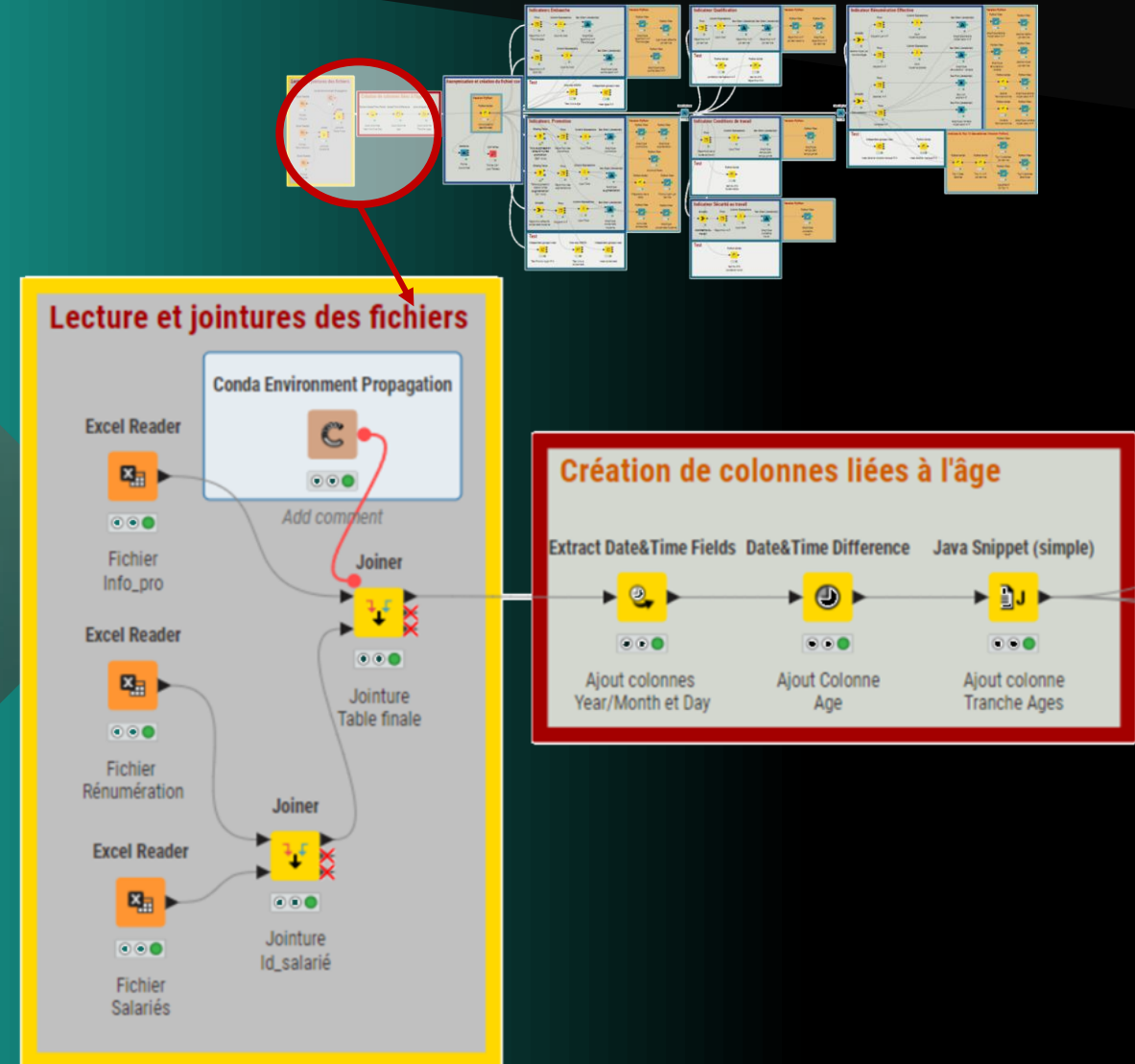
Workflow KNIME

Lecture et jointures des fichiers :

- Analyse exploratoire des différents fichiers effectuée au préalable sur Jupyter.
- Chargement des 3 fichiers sur le workflow
- Création d'une table reprenant toutes les données à l'aide de jointures réalisées sur l'unicité de la clé choisie (id_salarié).
- Ajout du node 'Conda' pour codage python ultérieur.

Création de colonnes liées à l'âge :

- Décomposition de la date de naissance (Year, month & day)
- Ajout d'une colonne avec l'âge actuel du salarié.
- Création d'une colonne 'Tranche âges' regroupant les âges par décennie.



Mise en conformité du fichier csv au RGPD

Identification des données à caractère personnel :

- Identifiant du salarié
- Nom et Prénom
- Numéro de téléphone
- Date de naissance

Base légale du traitement / Finalité du traitement :

Chaque année, nous avons l'obligation légale de calculer et publier sur notre site Internet l'index de l'égalité femmes-hommes.

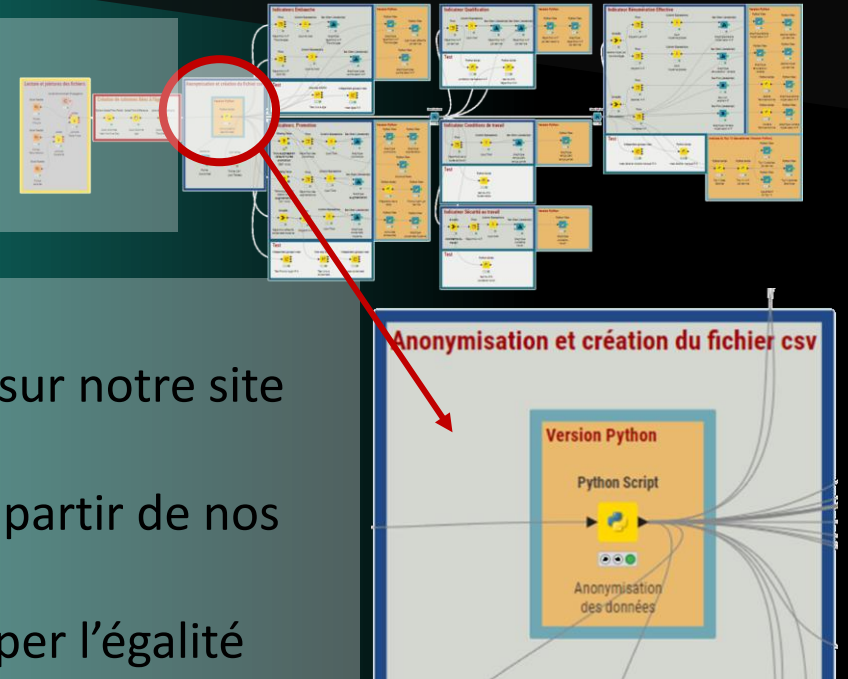
Dans ce cadre, nous devons traiter et analyser certaines données à partir de nos fichiers SIRH.

L'objectif étant d'appliquer une politique volontariste pour développer l'égalité femmes-hommes du cabinet. (argument pour le recrutement)

Minimisation des données :

Seules les données utiles à la publication de l'index sont conservées dans le traitement. Les données à caractère personnel sont anonymisées.

Le degré de précision concernant la distance au travail, le nombre d'enfants, niveau de satisfaction a été fortement abaissé.



Mise en conformité du fichier csv au RGPD

Transparence, information et droit des personnes concernées :

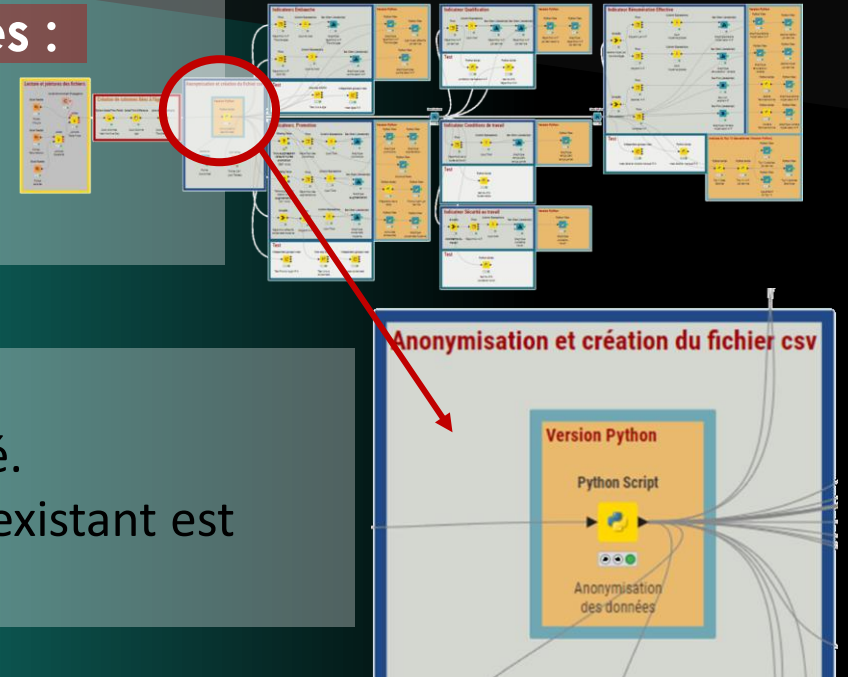
Nos collaborateurs sont informés du caractère légal et obligatoire du traitement des données.
Un droit de regard peut leurs être accordé sur demande formelle.

Sécurité des données et durée de conservation :

Le fichier stocké est accessible d'une manière restreinte et protégé.
Une mise à jour des données annuelle étant obligatoire, le fichier existant est supprimé et remplacé automatiquement d'une année sur l'autre.

Transfert des données et Analyse d'impact sur la protection des données (AIPD):

Aucun transfert n'est nécessaire.
Le traitement des données ne représente aucun risque pour les droits des personnes concernées. La réalisation d'un AIPD n'est pas utile dans notre cas.



Mise en conformité du fichier dans le Workflow KNIME

Anonymisation des données :

'id_salarié'

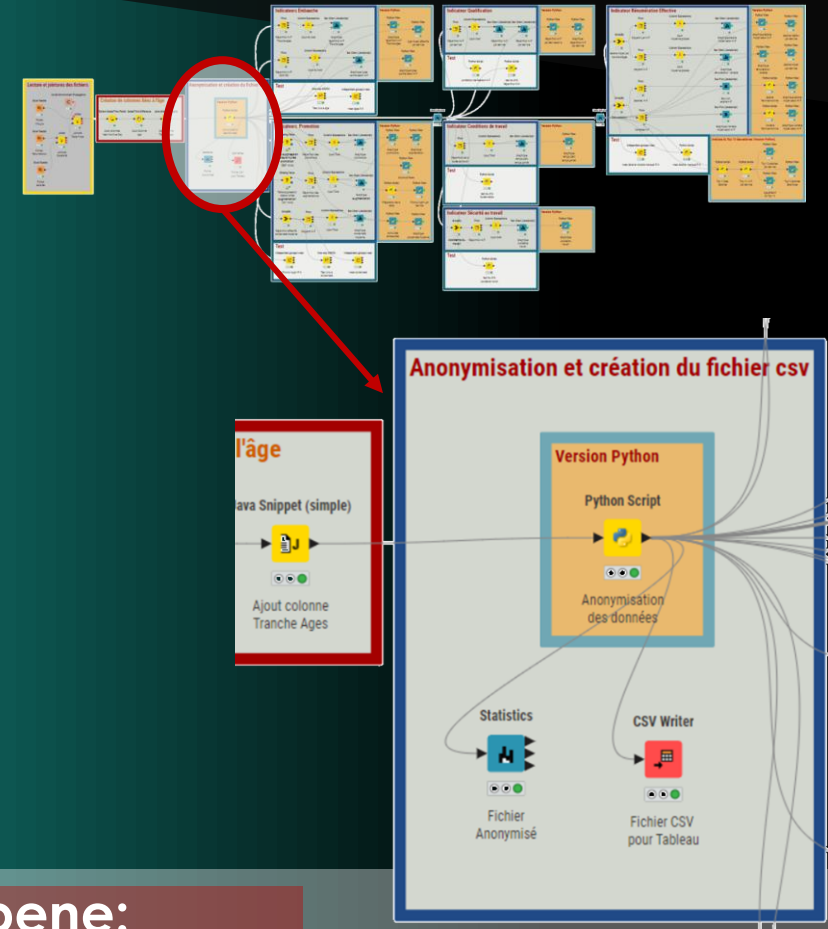
- Le node 'Anonymization' n'étant pas disponible sur la dernière version de Knime, j'ai utilisé le langage Python pour la réaliser.
- Création d'un id_salarié aléatoire avec vérification de l'unicité de chaque id. Les id_salarié d'origine sont écrasés par les id aléatoires.

Remplacements

- 'Niveau de satisfaction' par des mentions
- 'Distance Domicile/Travail' par tranches de 20km.
- 'La 'Durée hebdo' par TC ou TP.
- Le nombre 'Enfants' par Oui ou Non.

Autres

- Caviardage des colonnes 'Prénom/Nom', 'Téléphone' et 'Date_naissance'.
- Suppression les colonnes 'Year', 'Month' et 'Day of month'.



Nota bene:

'les données 'Ancienneté_an' sont conservées.
Une agrégation par tranches de 5 ans engendre une perte de précision dans l'analyse.

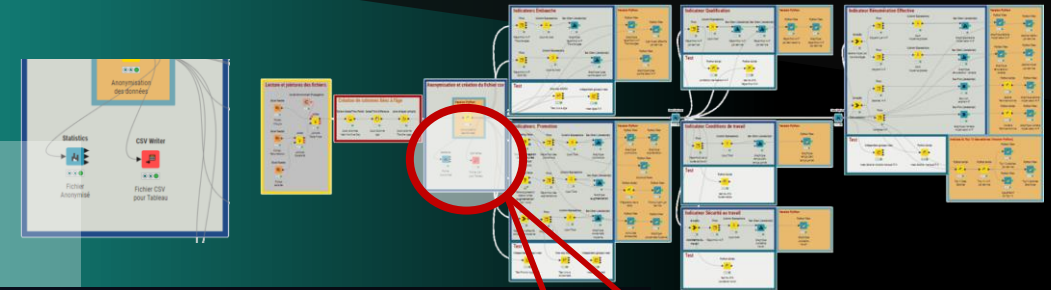
Workflow KNIME

Fichier csv avec données conformes au RGPD :

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	id salarié	Ancienneté_a	Distance domicile/T	Service	Work_accident	Niveau de satisfac	Contrat	Durée hebdo	Salaire base	variable_m	Augmentat	Promotio	Sex	Prénom/Nom	Telephone	Date_naissance	Etat Civil	Enfant	Ag	Tranche âges
2	16946023-5176	22.5	+ de 60 Km	Consultant	Non	Bien	CDI	Temps Complet	9873.5	13	0	0	F	#####	06-----	***.****	Marié	Oui	47	40 à 49 ans
3	16995805-1185	10.8	21-40 Km	Consultant	Non	Moyen	CDI	Temps Complet	4296.7167	39	0	1	H	#####	06-----	***.****	Célibataire	Non	68	60 ans et plus
4	16960469-8432	13	+ de 60 Km	RH	Oui	Bien	CDI	Temps Complet	3705.2167	29	1	1	F	#####	06-----	***.****		Oui	53	50 à 59 ans
5	16990950-5399	21.7	41-60 Km	Consultant	Non	Très bien	CDI	Temps Complet	1697.15	17	1	0	H	#####	06-----	***.****	Marié	Oui	59	50 à 59 ans
6	16941098-5387	3.3	0-20 Km	R&D	Oui	Très bien	CDD	Temps Complet	2848.3	15	1	0	H	#####	06-----	***.****		Non	40	40 à 49 ans
7	16913185-3593	10.5	0-20 Km	Commercial	Non	Moyen	CDI	Temps Complet	5464.55	34	0	0	H	#####	06-----	***.****		Non	54	50 à 59 ans
8	16973891-3722	17.8	41-60 Km	Marketing	Oui	Moyen	CDI	Temps Complet	9999.3833	27	0	1	H	#####	06-----	***.****	Célibataire	Non	47	40 à 49 ans
9	16987974-2204	8.2	41-60 Km	RH	Oui	Moyen	CDI	Temps Partiel	3116.88	37	0	1	F	#####	06-----	***.****		Oui	68	60 ans et plus
10	16988115-5688	3.6	41-60 Km	Marketing	Non	Moyen	CDI	Temps Partiel	5462.08	30	0	1	H	#####	06-----	***.****	Célibataire	Oui	47	40 à 49 ans

Information du node 'Statistics' :

Ancienneté_an	Distance domicile/Travail	Service	Work_accident	Niveau de satisfaction	Contrat	Durée hebdo	Salaire base mensuel	%variable_moyen	Augmentation	Promotion	Sexe	Etat Civil	Enfants	Tranche âges
No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 15	No. missings: 15	No. missings: 0	No. missings: 0
Top 20: 16.2 : 6 8.2 : 5 10.3 : 4 15.7 : 4 23.9 : 4 10.8 : 3 7.5 : 3 23.3 : 3 5.8 : 3 15.6 : 3 9.2 : 3 22.8 : 3 9.3 : 3 4.2 : 3 6.3 : 3 19.0 : 3 21.1 : 3 1.1 : 3 5.5 : 3 11.7 : 3	Top 20: 41-60 Km : 84 21-40 Km : 68 0-20 Km : 60 + de 60 Km : 44	Top 20: Consultant : 69 Commercial : 52 RH : 43 Compta Finances : 41 Marketing : 36 R&D : 15	Top 20: Oui : 129 Non : 127	Top 20: Moyen : 70 Bien : 69 Très bien : 59 Mauvais : 58	Top 20: CDI : 240 CDD : 16	Top 20: Temps Complet : 225 Temps Partiel : 31	Top 20: 2070.25 : 2 8056.5333 : 2 9573.2 : 2 8696.5667 : 2 7637.9333 : 2 9873.5 : 1 4296.7167 : 1 3705.2167 : 1 1697.15 : 1 2848.3 : 1 5464.55 : 1 9999.3833 : 1 3116.88 : 1 5462.08 : 1 2769.4333 : 1 5522.1833 : 1 7308.8167 : 1 5895.76 : 1 7081.3167 : 1 6195.5833 : 1	Top 20: 17 : 14 21 : 11 39 : 10 27 : 10 19 : 10 38 : 10 12 : 9 30 : 9 6 : 9 34 : 8 7 : 8 24 : 8 13 : 7 33 : 7 1 : 7 14 : 7 8 : 7 20 : 7 32 : 7 37 : 6	Top 20: 1.0 : 125 0.0 : 116 ?: 15	Top 20: 1.0 : 125 0.0 : 116 ?: 15	Top 20: H : 131 F : 125	Top 20: Marié : 92 Célibataire : 82 ?: 82	Top 20: Oui : 190 Non : 66	Top 20: 50 à 59 ans : 92 40 à 49 ans : 74 60 ans et plus : 62 30 à 39 ans : 28
Bottom 20: 14.0 : 1 6.4 : 1 24.8 : 1	Bottom 20:	Bottom 20:	Bottom 20:	Bottom 20:	Bottom 20:	Bottom 20:	Bottom 20: 3459.5167 : 1 9162.1833 : 1 1920.1 : 1	Bottom 20: 23 : 6 10 : 6 4 : 6	Bottom 20:	Bottom 20:	Bottom 20:	Bottom 20:	Bottom 20:	Bottom 20:



CSV Writer
Fichier CSV pour Tableau

Statistics
Fichier Anonymisé

Workflow KNIME

Exemple de l'analyse d'un indicateur :

Cadre 'Indicateur' :

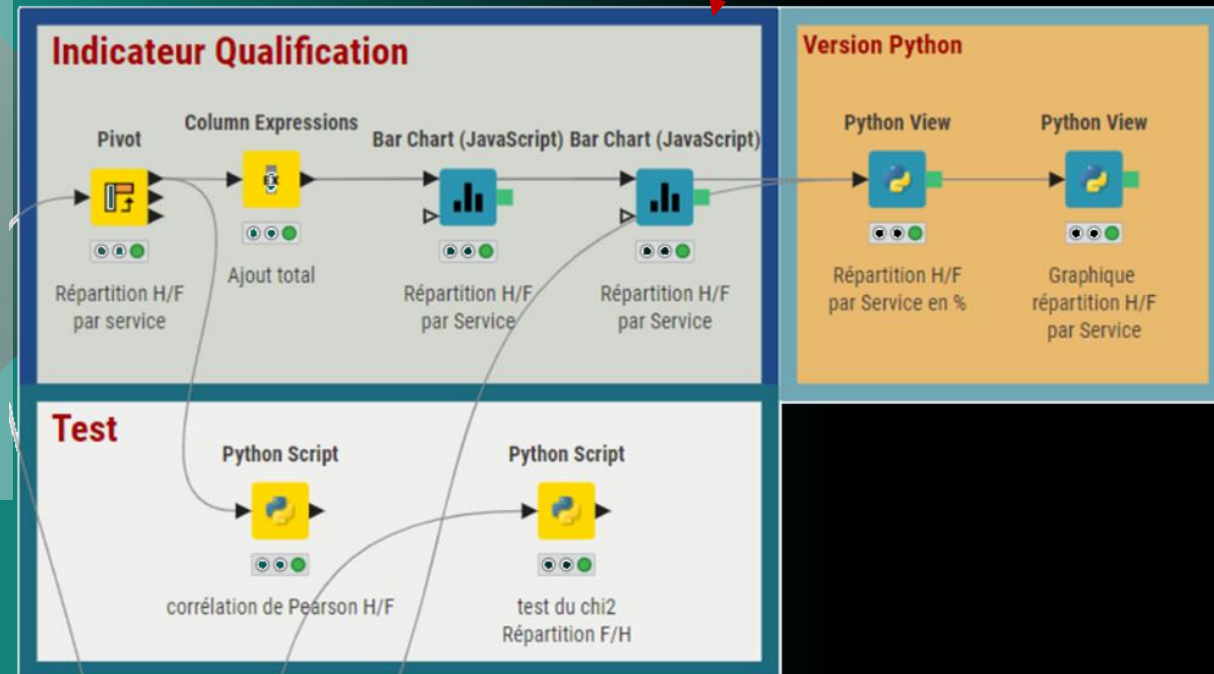
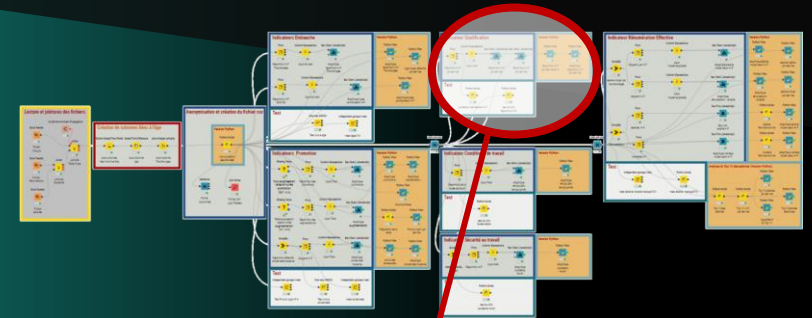
- Préparation des données utiles à l'indicateur
- Visualisation graphique des résultats

Cadre 'Version Python' :

- Adaptation et/ou complément d'analyse avec Python

Cadre 'Test' :

- Confirmation des résultats par test d'équité (Femmes/Hommes)



Analysez des indicateurs de l'égalité femme-homme avec Knime



Index de l'égalité professionnelle

Index de l'égalité professionnelle - Rappel

QU'EST-CE QUE L' « INDEX » ?

C'est **un outil pour diagnostiquer les inégalités** de rémunération et de situation entre les femmes et les hommes.

Chaque année au plus tard le 1^{er} mars, les entreprises d'au moins 50 personnes salariées doivent calculer et publier sur leur site internet leur Index de l'égalité professionnelle femmes-hommes.

Elles doivent également **le communiquer**, avec le détail des différents indicateurs, à leur comité social et économique via la base de données sociales et économiques ainsi qu'à l'inspection du travail (Direccte).

COMMENT LE CALCULER ?

L'index, sur 100 points, est **calculé à partir de 4 à 5 indicateurs** selon que l'entreprise compte moins ou plus de 250 personnes salariées :

- ▶ L'écart de **rémunération** femmes-hommes, sur 40 points ;
- ▶ L'écart de taux d'**augmentations** individuelles, sur 20 points (35 points pour les entreprises de 50 à 250 personnes salariées) ;
- ▶ L'écart de taux de **promotions** (uniquement dans les entreprises de plus de 250 personnes salariées), sur 15 points ;
- ▶ Le pourcentage de salariées augmentées à leur retour de **congé de maternité**, sur 15 points ;
- ▶ Le nombre de personnes salariées du sexe sous-représenté parmi les **10 plus hautes rémunérations**, sur 10 points.

LES MESURES CORRECTIVES

En cas d'index inférieur à 75 points, l'entreprise doit mettre en place des mesures correctives adéquates et pertinentes pour atteindre le seuil de 75 points dans un délai maximal de trois ans.

Ces mesures annuelles ou pluriannuelles doivent être définies dans le cadre de la négociation obligatoire sur l'égalité professionnelle ou, à défaut d'accord, par décision unilatérale de l'employeur et après consultation du CSE.

En l'absence :

- ▶ de publication de son index,
- ▶ de définition, à la suite d'une note inférieure à 75 points, de mesures correctives adéquates et pertinentes,
- ▶ ou en cas d'inefficacité de ces mesures au bout de trois ans, l'index restant inférieur à 75, **l'entreprise s'expose à une sanction de pénalité financière jusqu'à 1 % de sa masse salariale annuelle.**

1 – Ecart de rémunération Femmes / Hommes

Sur la globalité des effectifs

L'écart de salaire peut se présenter de deux façons:

- Les hommes gagnent 2,20 % de plus que les femmes
- Les femmes gagnent 2,15 % de moins que les hommes

Sur les effectifs par service

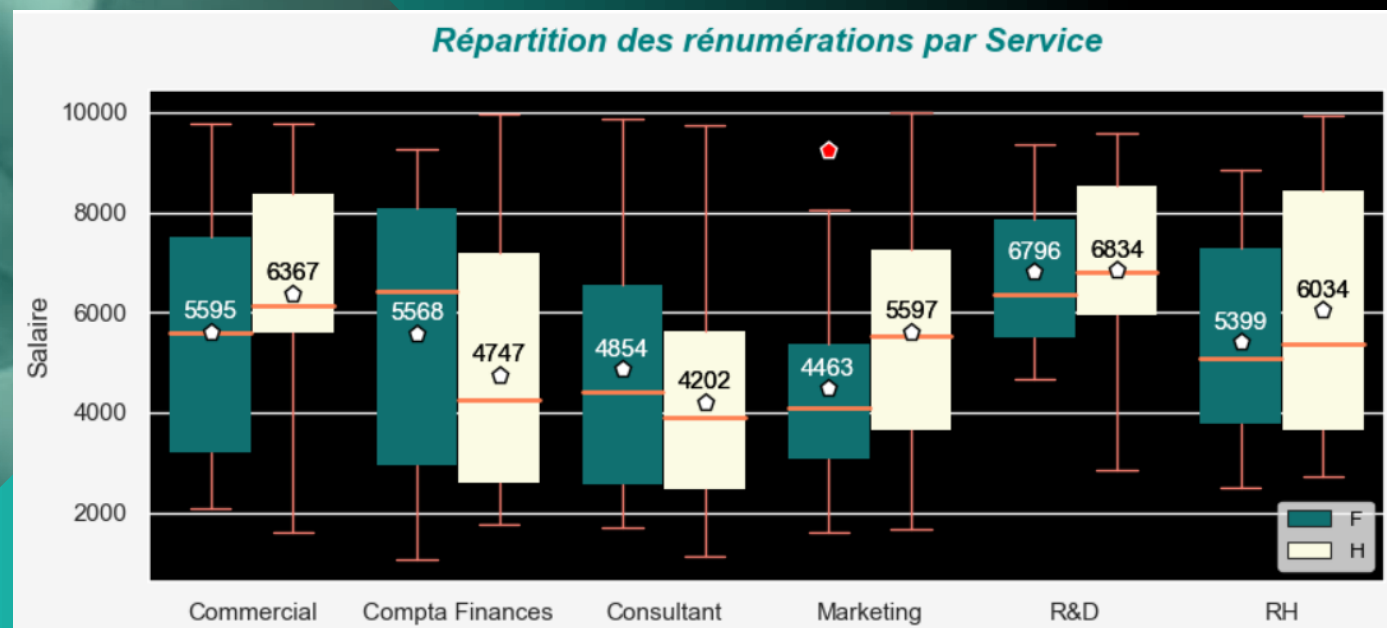
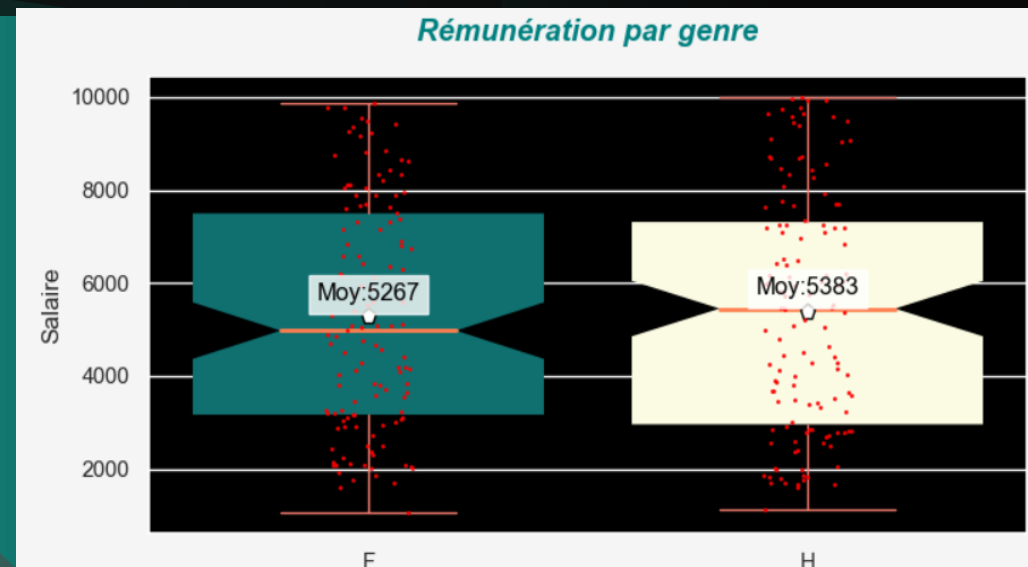
L'écart femme/homme de deux façons:

Comm. :	-12,12 %	ou	13,80 %
Compta :	17,30%	ou	-14,74%
Consult. :	15,52%	ou	-13,43%
Marktg. :	-20,74%	ou	25,41%
R&D :	-0,56%	ou	0,56%
RH :	-10,52%	ou	11,76%

Note :

Nous avons un écart de rémunération de 2,15%

$$\text{Points} = (1 - (\text{Écart} / 40)) \times 40 = 37,85$$



2 - Écart de répartition des augmentations/promotions individuelles

Augmentation :

Nous observons un écart de 2,4% favorable aux femmes.

Note :

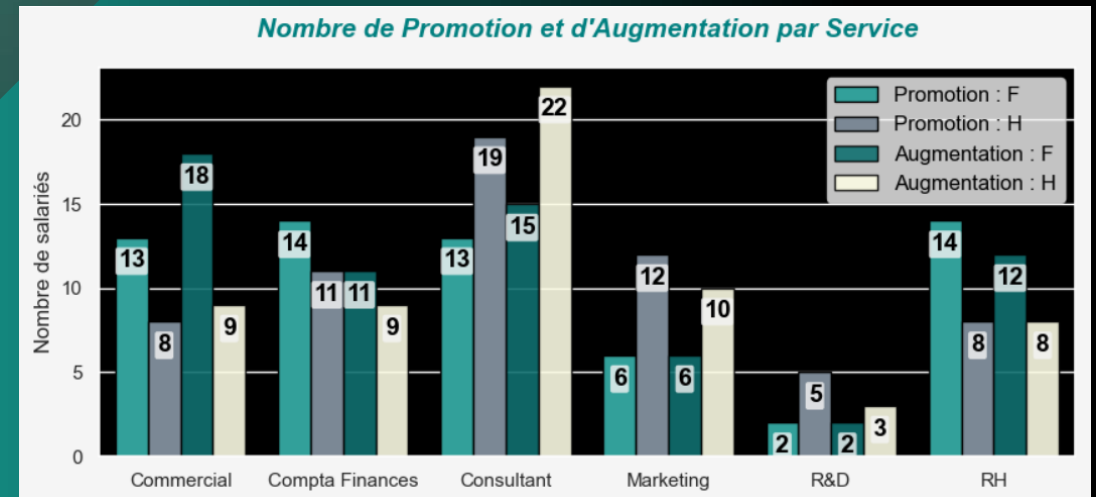
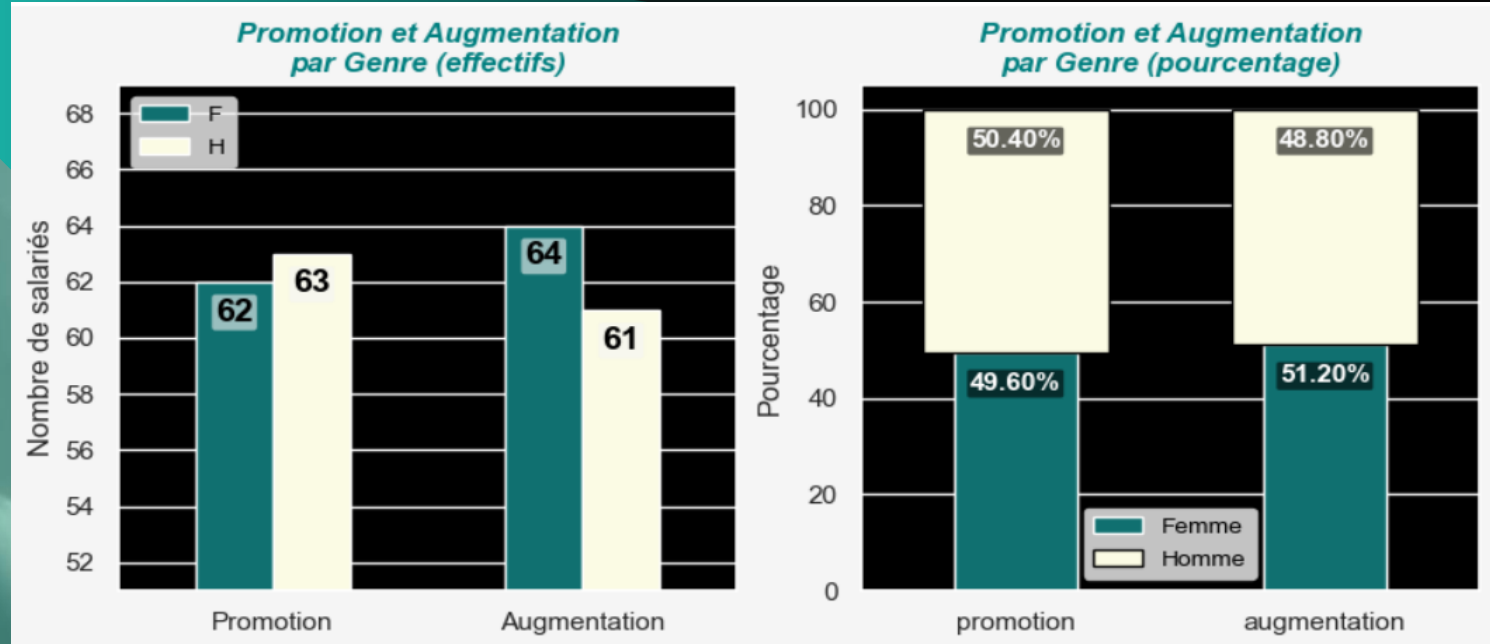
Selon les règles de calcul en vigueur, nous obtenons une note de 20 points.

Promotion : (plus de 250 salariés)

Nous observons un écart de d'1 personne favorable aux hommes.

Note :

L'égalité étant à moins de 2 personnes près, nous obtenons une note de 15 points.



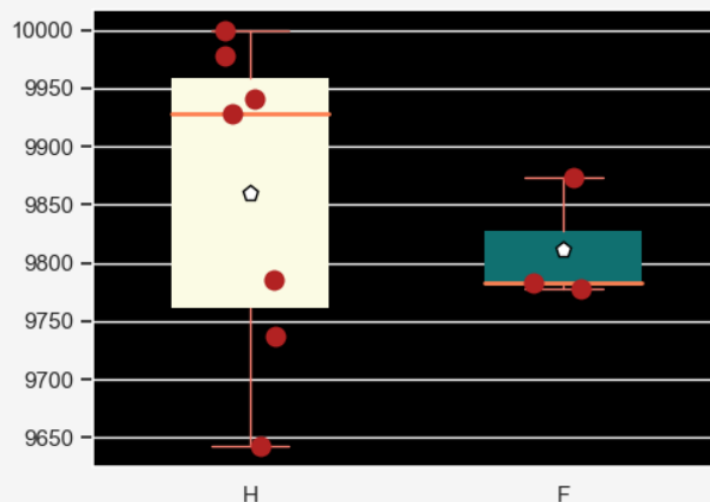
3 - Parité parmi les 10 plus hautes rémunérations

Parité Top 10:

Pour obtenir les 10 points, la présence de 4 femmes dans le top est nécessaire.

Top 10 des salaires

Sexe	Salaire base mensuel	%variable_moyen	Age
H	9999.38	27	47
H	9978.15	34	53
H	9940.23	24	47
H	9928.1	27	38
F	9873.5	13	47
H	9785.53	28	42
F	9782.5	38	50
F	9777.95	4	47
H	9737.0	12	58
H	9642.97	17	60

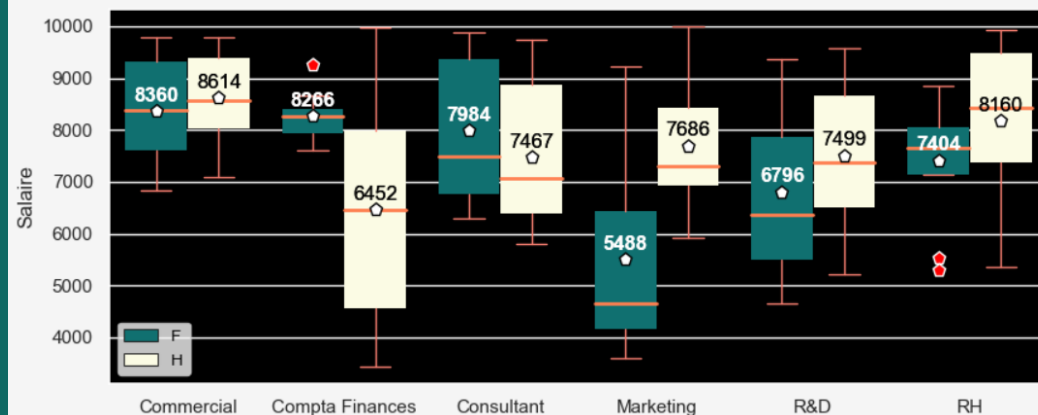


Note :

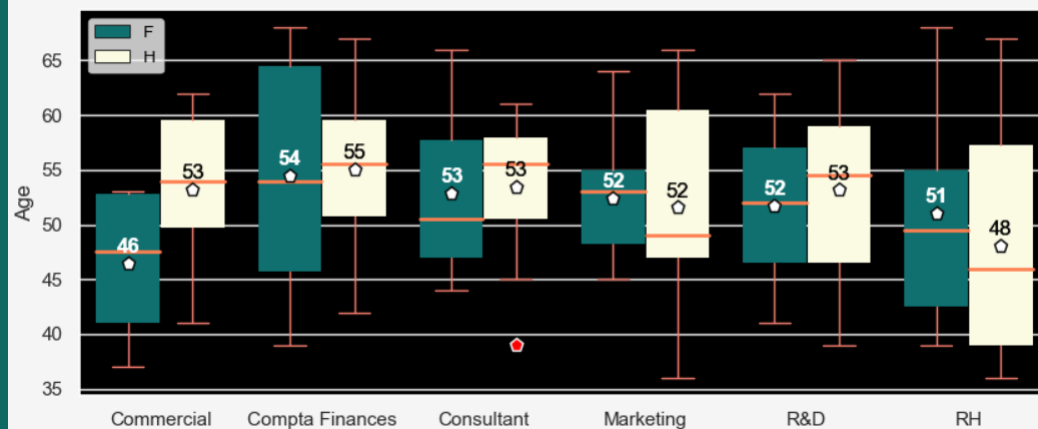
Pour 3 femmes présentes, le calcul est:
 $(4/3) * 10 = 7,5$ points

Répartition par service :

Top 10 des salaires par service



Age des effectifs du top 10



4 -Index de l'égalité professionnelle - Résultat

Sur la globalité des effectifs

Nous avons une note de 37,85 points.

Augmentation :

Nous avons une note de 20 points.

Promotion :

Nous avons une note de 15 points.

Augmentée suite congé de maternité :

N'ayant pas la possibilité de le vérifier. Nous considérons que cela est une pratique du Cabinet pour attirer de nouveaux talents.
Soit une note de 15 points.

Parité Top 10:

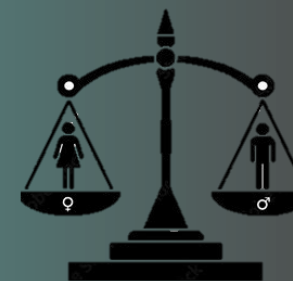
Nous avons une note de 7,5 points.

Avec une Note globale de

95,35 / 100

cette année

Nous avons un argument fort pour recruter de nouveaux talents



Analysez des indicateurs de l'égalité femme-homme avec Knime



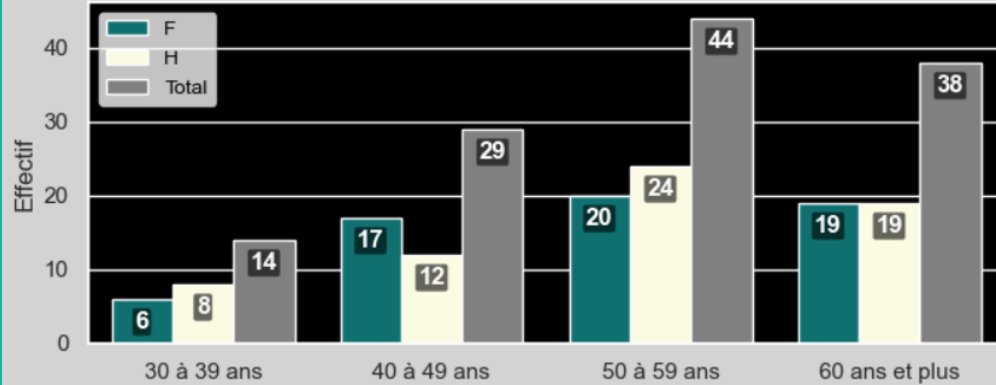
Indicateurs de diagnostic Égalité

Indicateurs : Rémunération effective et Promotion

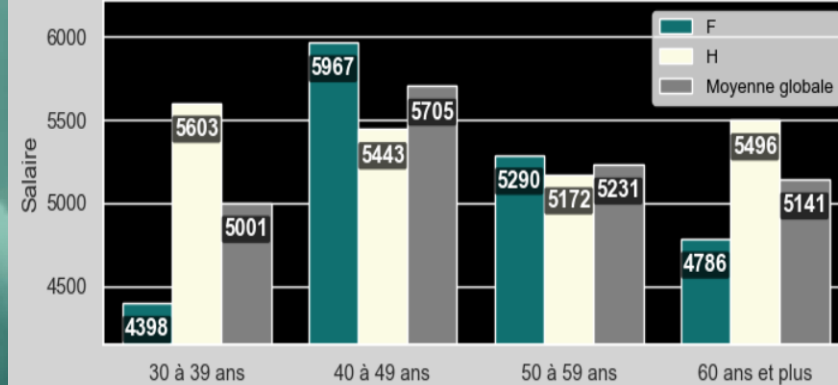
Analyse :

- Indicateurs déjà analysés dans l'index d'égalité professionnelle.
- Visualisations complémentaires suivant les tranches d'âges.

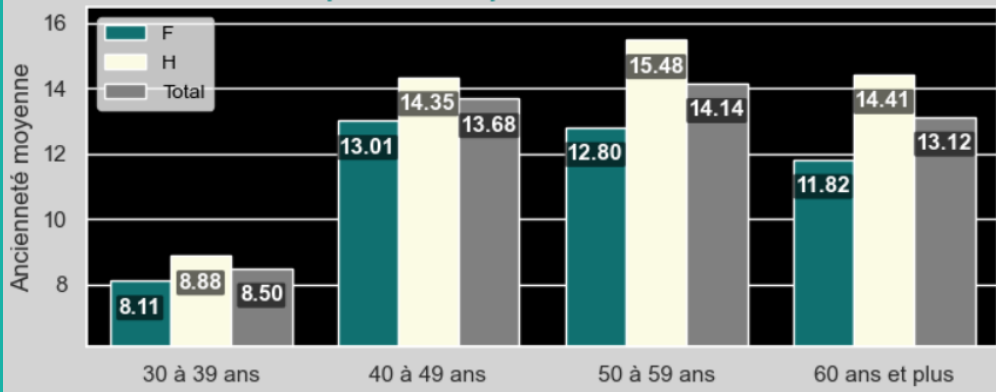
Répartition des promotions par genre



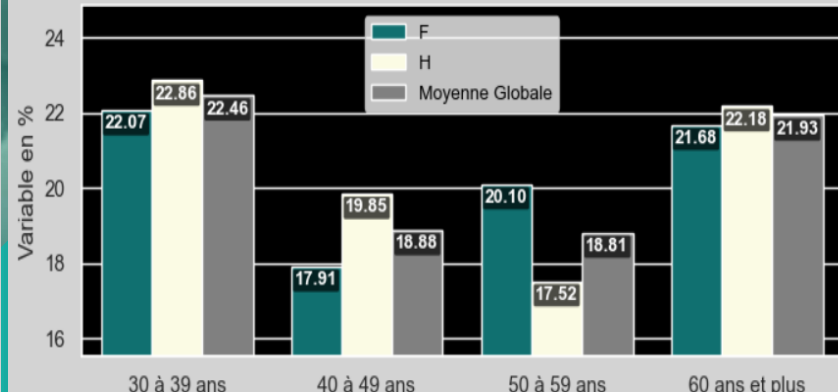
Salaire moyen par tranche d'âges et par genre



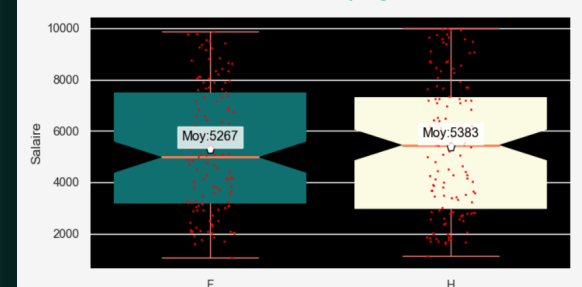
Répartition F/H par suivant l'ancienneté



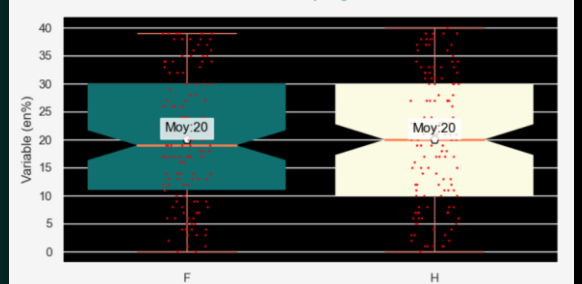
Rémunération moyenne variable par genre



Rémunération par genre



%Variable par genre



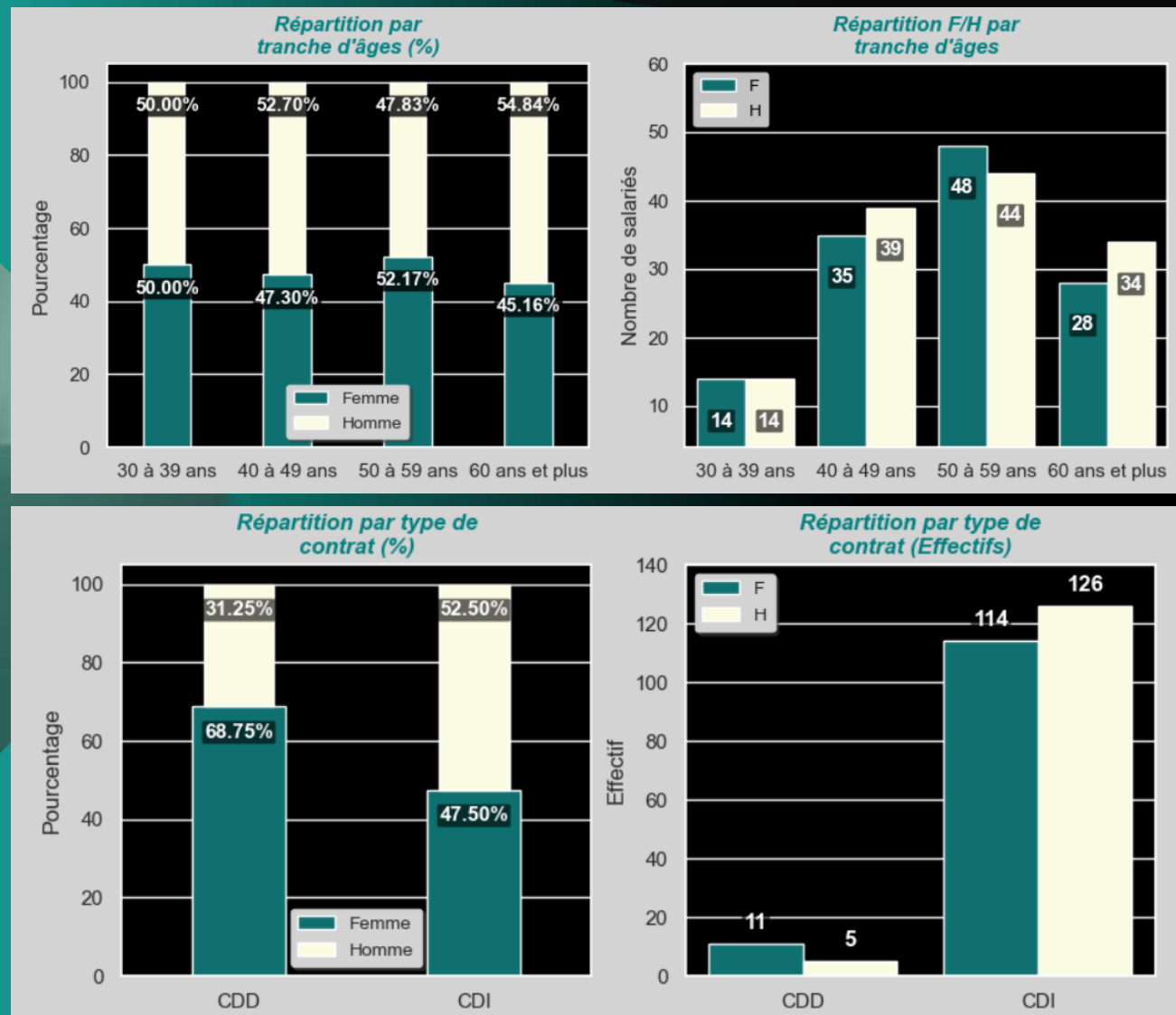
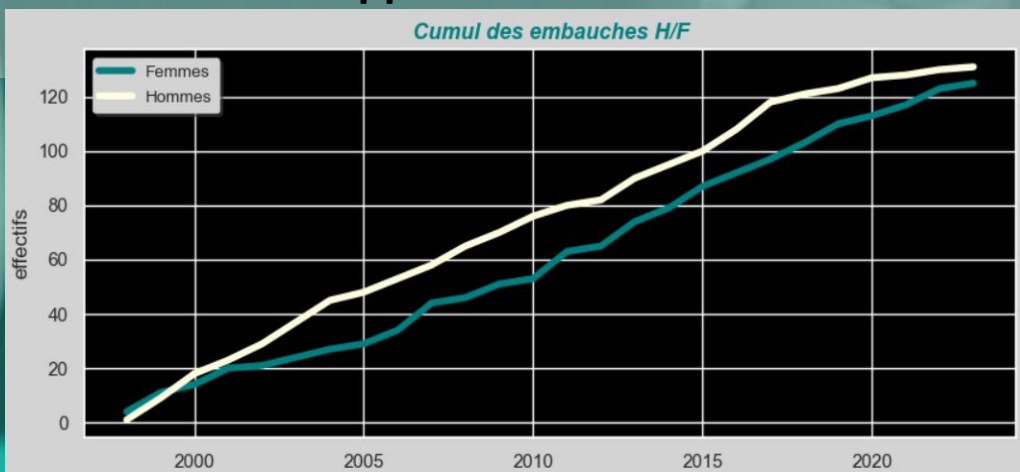
Indicateur : Embauche

Analyse :

- Les effectifs F/H sont relativement bien répartis par tranches d'âges.
- CDI : Nous avons 9,5% moins de femmes
- CDD : Nous avons deux fois plus de femmes

Effectifs actuels du cabinet :

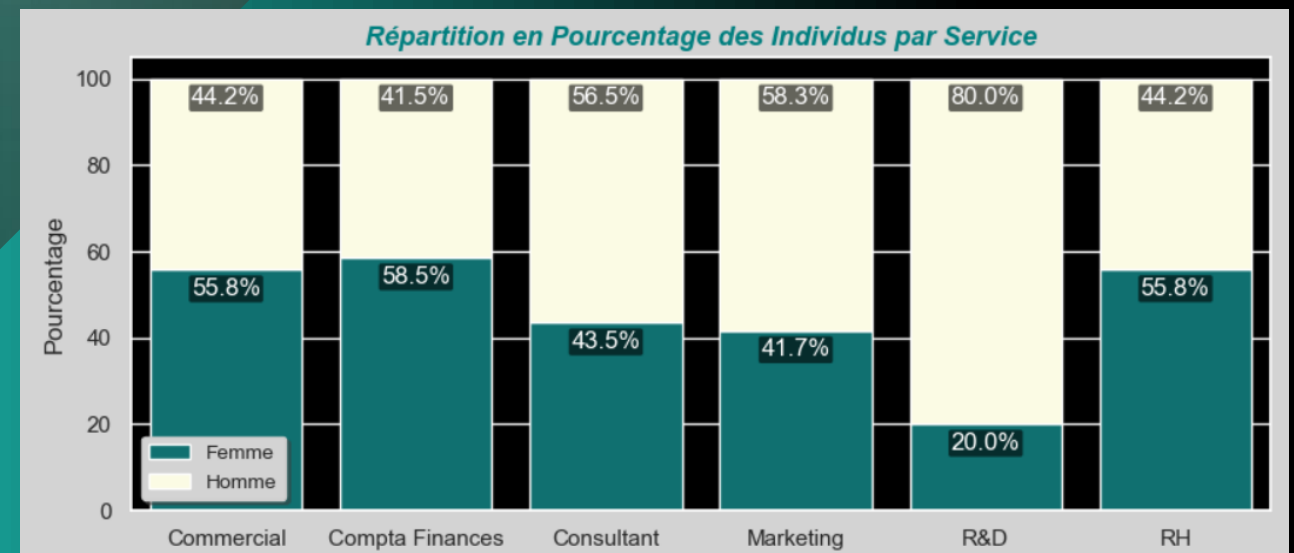
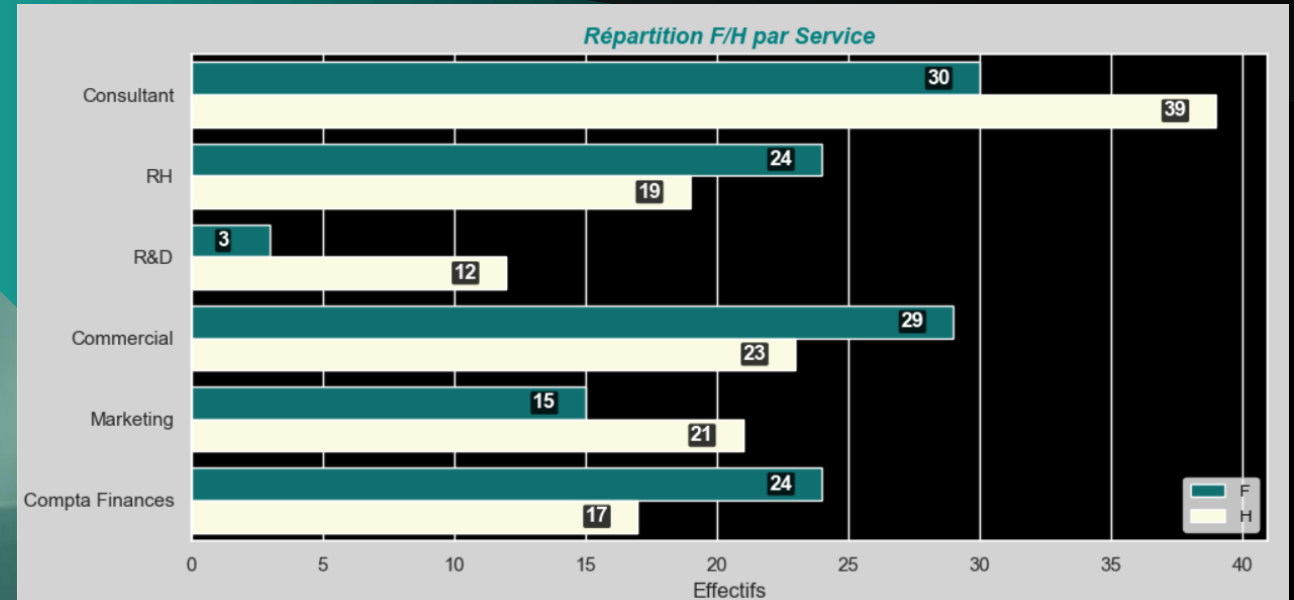
- Les nombres de salariés Femmes / Hommes tendent à se rapprocher



Indicateur : Qualification

Analyse :

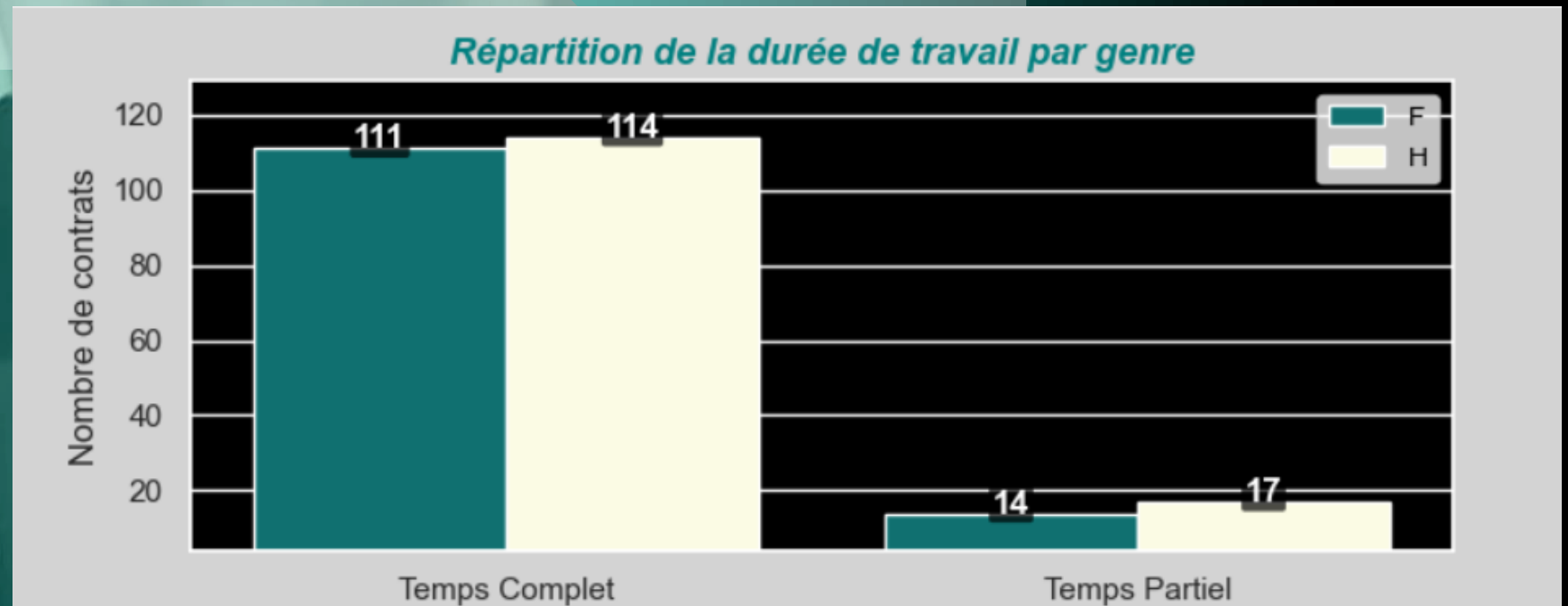
- Les effectifs F/H sont relativement bien équilibrés dans les services.
- Seule la R&D est très déséquilibrée avec seulement 3 femmes pour 12 hommes.



Indicateur : Conditions de travail

Analyse :

- Le temps complet concerne 225 salariés
- 31 personnes sont en temps partiels.
- Le cabinet offre la possibilité aux salariés d'adapter leur temps de travail.



Indicateur : Sécurité et Santé au Travail

Analyse :

- Avec 63 accidents concernant les femmes et 66 pour les hommes, la répartition est plutôt homogène.
- Il serait intéressant de comprendre pourquoi les accidents représentent la moitié des effectifs du cabinet. En effet, les métiers du cabinet n'exposent pas ses salariés à de risques importants en termes de sécurité physique.

