

[SLIDES 1 & 2]

Ces notes s'attachent à décrire les différentes étapes constituant les phases d'import et de préparation des données dans le workflow, avant d'aborder les principes d'égalité professionnelle femme-homme au travers de quelques statistiques descriptives et des indicateurs légaux obligatoires que les entreprises doivent produire chaque année ; on utilisera pour cela les différents graphiques et indicateurs produits par le workflow. Après avoir rappelé les principes essentiels du Règlement General de Protection des Données (le fameux RGPD) qui s'appliquent à Laflèche & Associés, on expliquera enfin comment les données ont été mises en conformité avant d'être exportées au format .csv. On conclura enfin avec quelques pistes pour l'amélioration de la qualité et de la sécurité des données et la mise en conformité avec les obligations légales du cabinet en la matière.

[SLIDE 3]

Tout d'abord, une vue d'ensemble du workflow avant d'en décrire les diverses étapes plus en détail. On a dans la partie en rouge à gauche l'import des fichiers avec les vérifications initiales, en orange foncé les jointures, en orange clair le type casting, la création de variables et la recherche d'outliers, en vert l'export du fichier .csv anonymisé et enfin en jaune sur la partie droite les différentes analyses statistiques descriptives et graphiques de l'égalité professionnelle femmes-hommes. Dans la partie tout à droite enfin, on peut voir un éclaté des différents composants du métanode de traitement des données.

[SLIDE 4]

Tout d'abord, donc, le téléchargement et la préparation des données. 3 fichiers ont été fournis par la DRH, tous au format MS Excel (extension .xlsx). Les traitements de départ ont été identiques pour les 3 fichiers, à savoir :

- Les fichiers ont été importés dans le workflow ;

- La dimension des tables a été vérifiée et elles ont été inspectées, ce qui a permis d'identifier, après élimination des doublons éventuels, la colonne `id_salarié` comme une clé primaire en vue de la jointure des tables ;
- Des valeurs manquantes ont été identifiées, aucune dans la table `info pro`, 82 dans la table `salariés` (uniquement dans la colonne `état civil`) et 30 dans la table `rémunération` (15 dans chacune des colonnes `augmentation` et `promotion` respectivement). Ces données ne pouvant être inférées d'autres éléments des tables sans un risque d'erreur considérable, on a donc choisi de ne faire aucun retraitement et de conserver les individus concernés dans l'analyse compte-tenu de la proportion relativement faible de données manquantes (15/256 soit moins de 6%).

[SLIDE 5]

Les tables ont ensuite été jointes, donc, en utilisant comme on l'a dit la colonne `id_salarié` comme clé. On a choisi de procéder à des jointures internes et d'utiliser les fonctionnalités du node pour analyser les échecs de jointure à droite et à gauche – aucun échec n'est à rapporter dans aucune des 2 jointures faites, ni à droite ni à gauche – chacun des employés existe donc dans les 3 tables et il n'y a pas d'employé dans les tables `info pro` et `rémunération` qui n'existerait pas dans le référentiel des salariés.

[SLIDE 6]

On a ensuite procédé au renommage des colonnes pour en supprimer les caractères spéciaux et les espaces, on a créé la variable `âge` à partir des dates de naissance des employés. On a également calculé le salaire équivalent temps-plein à partir des heures hebdomadaires et du salaire fixe et enfin on a assigné des valeurs booléennes à certaines données : `promotion`, `augmentation`, `enfants` (au lieu du nombre d'enfants), `temps plein/partiel` (au lieu du nombre d'heures hebdomadaires travaillées) et `accident du travail`. Tous ces traitements ont été regroupés dans un métanode. Nous avons également procédé à la création d'une clé artificielle après avoir mélangé les lignes de notre tableau final. Un certain nombre de ces modifications sont liées non seulement

aux bonnes pratiques en matière de data analysis mais à la mise en conformité avec le RGPD, on aura l'occasion d'y revenir plus en détail dans la 3^{ème} partie.

[SLIDE 7]

Voyons à présent ce qui caractérise l'égalité femme-homme, avec dans un premier temps quelques principes génériques contenus dans le Code du Travail et la transposition en droit français de diverses directives européennes. L'égalité professionnelle se constitue comme l'absence de discrimination entre les hommes et les femmes dans l'entreprise, un point important (mais non le seul) étant qu'à travail identique ou de valeur égale, la rémunération de l'employé(e) doit être identique. La garantie de l'égalité impose à l'employeur de formuler des objectifs proportionnels à la taille de l'entreprise, mais aussi de garantir la sécurité de ses salarié(e)s et de publier un indice d'égalité qui est obligatoire pour toutes les entreprises de plus de 50 salariés.

[SLIDE 8]

Avant d'aborder l'index à proprement parler et ses différents constituants, quelques statistiques descriptives. Le cabinet compte 131 hommes sur 256 employés, soit environ 51%, et 49% de femmes, donc. La majorité des postes est occupée par des femmes dans les services commerciaux, la comptabilité-finance et les RH, les hommes formant la majorité des effectifs dans les services de consulting, marketing et recherche et développement.

[SLIDE 9]

80% des salariéES ont des enfants contre 69% de leurs confrères, et 56% des salariéES ont bénéficié d'une augmentation de salaire contre 48% des hommes seulement (ces chiffres étant données sur la base on le rappelle de 256 moins 15 salariés pour lesquels l'information est manquante), les autres indicateurs étant relativement proches pour les 2 sexes, avec notamment une prévalence des accidents du travail identique à 50%.

L'indice de satisfaction médian s'établit à 51 chez nos collègues féminines et 49 chez leurs confrères.

[SLIDE 10]

Voyons à présent 2 indicateurs de précarité. Parmi 16 employés en CDD, 11 sont des femmes, soit plus des 2 tiers. On sait par ailleurs qu'au niveau national, 80% des emplois à temps partiel sont occupés par des femmes, or dans le cabinet, sur 31 employés à temps partiel, seuls 14 sont des femmes, soit moins de 50%, statistique sur laquelle le cabinet fait beaucoup mieux que la moyenne nationale.

[SLIDE 11]

Quelques indicateurs de distribution enfin. Les salaires variables et l'âge médians sont très proches chez nos collègues masculins et féminins...

[SLIDE 12]

... alors que l'ancienneté médiane des hommes est de 22% plus élevée environ que celle des femmes, qui quant à elles ont un trajet travail médian de 13% supérieur à celui des hommes.

[SLIDE 13]

Passons à présent au calcul des différentes composantes de l'index d'égalité professionnelle femmes-hommes. Pour rappel, son calcul est obligatoire pour toutes les entreprises de plus de 50 salariés et il doit être publié avant le 1^{er} mars de chaque année. Outre conditionner l'accès aux marchés publics, cet index doit également être supérieur à 75 points sur 100, faute de quoi l'entreprise a 3 ans pour mettre en place des mesures correctives avec (et on insiste sur ce point) OBLIGATION DE RESULTATS sous peine de sanctions financières pouvant s'élever jusqu'à plusieurs % du CA.

Pour les entreprises de plus de 250 salariés comme la nôtre, cet index se décompose en 5 indicateurs :

- L'écart de rémunération femmes-hommes (sur 40 points) ;
- L'écart de taux d'augmentations individuelles (sur 20 points) ;
- Le pourcentage de salariéES augmentées à leur retour de congé de maternité (sur 15 points) ;
- Le nombre de salariés du sexe sous-représenté parmi les 10 plus hautes rémunérations (10 points) ; et
- L'écart de taux de promotions (sur 15 points).

Ce dernier indicateur est facultatif pour les entreprises de moins de 250 salariés et est alors remplacé par une surpondération de 15 points du 2^{ème} indicateur, afin de maintenir la base totale de calcul de l'indice à 100 points.

[SLIDE 14]

Concernant le premier indicateur, l'écart de rémunération femmes-hommes, il s'établit à +2.80% en moyenne en faveur des hommes et +12.6% en médiane, toujours en faveur des hommes – pour rappel le calcul a été effectué ici sur le salaire équivalent temps plein. Nous faisons donc plutôt mieux que la moyenne nationale qui est de +17% en moyenne en faveur des hommes. Les taux de salaire variables étant similaires dans notre cabinet (39% pour les femmes et 40% pour les hommes), leur inclusion ne contribuerait pas à changer significativement ce résultat si l'on considérait le salaire total au lieu du salaire fixe.

[SLIDE 15]

Nous ne sommes malheureusement pas en mesure de calculer le second indicateur, qui est celui de l'écart de taux d'augmentation individuelle, cette information n'étant pas incluse dans les données de la DRH. On a donc essayé de l'approximer.

Si l'on s'intéresse au nombre d'employés augmentés, on constate que :

- 64 femmes ont été augmentées, 50 n'ont pas été augmentées et nous n'avons pas d'information pour 11 d'entre-elles ;
- Chez les hommes, 61 ont été augmentés, 66 ne l'ont pas été et l'information est indisponible pour 4 d'entre eux.

Parmi les employés pour lesquels l'information est disponible, plus de 51% des femmes ont été augmentées, contre 48% des hommes.

On constate par ailleurs que les employés sans enfant forment une plus grande proportion (33/116 soit 28%) des employés non augmentés que des employés augmentés, dont 25% (31/125) seulement sont sans enfants.

[SLIDE 16]

Par ailleurs, si l'on compare les taux d'augmentation par sexe des employés parents aux non-parents, on constate que :

- 56% des femmes sans enfants ont été augmentées (14/25)
- Un peu plus de 41% des hommes sans enfants ont été augmentés (17/41)
- 50% des femmes ayant des enfants ont été augmentées (50/100) et
- Pres de 49% des hommes ayant des enfants ont été augmentés (44/90)

Le fait d'avoir des enfants semble donc avoir un impact positif pour les hommes en termes d'augmentation dans notre cabinet, et négatif pour les femmes.

[SLIDE 17]

Passons à présent au 3^{ème} indicateur, l'écart de taux de promotion. Si l'on s'intéresse au nombre d'employés promus, on constate que :

- 62 femmes ont été promues, 52 n'ont pas été promues et nous n'avons pas d'information pour 11 d'entre-elles ;
- Chez les hommes, 63 ont été promus, 64 ne l'ont pas été et l'information est indisponible pour 4 d'entre eux.

Parmi les employés pour lesquels l'information est disponible, plus de 54% des femmes ont été promues, contre 49% des hommes.

On constate par ailleurs que les employés sans enfant forment une plus grande proportion (28% soit 33/116) des employés non promus que des employés promus, dont 25% (31/125) seulement sont sans enfants.

[SLIDE 18]

Par ailleurs, si l'on compare les taux de promotion par sexe des employés parents aux non-parents, on constate que :

- 52% des femmes sans enfants ont été promues (13/25)
- Pres de 44% des hommes sans enfants ont été promus (18/41)
- 49% des femmes ayant des enfants ont été promues (49/100) et
- 50% des hommes ayant des enfants ont été promus (45/90)

Le fait d'avoir des enfants semble donc ici encore avoir un impact positif pour les hommes en termes de promotion dans notre cabinet, et négatif pour les femmes.

[SLIDE 19]

Passons au 4^{ème} indicateur, celui du pourcentage de salariées augmentées à leur retour de congés maternité. Encore une fois, faute de données pertinentes, nous ne pouvons pas calculer cet indicateur, les données reçues de la DRH ne précisant pas si nos employéES ont bénéficié ou non d'un congé maternité depuis leur embauche.

Nous avons donc approximé ce chiffre en utilisant la proportion de femmes ayant des enfants et ayant été augmentées, soit 50% comme on l'a dit précédemment, ce qui revient à supposer que toutes les employéES ayant des enfants les ont eus depuis leur arrivée dans l'entreprise et qu'elles ont toutes bénéficié d'un congé maternité.

[SLIDE 20]

Et enfin le dernier indicateur ; le sexe sous-représenté parmi les 10 plus hauts salaires – ce sont les femmes, au nombre de 3 sur 10.

Toutes les données évoquées ici étant de nature personnelle, il nous faut à présent évoquer nos obligations en la matière vis-à-vis du RGPD.

[SLIDE 21]

Le RGPD est issu de la transposition en droit français du paquet européen de protection des données (2016). Il repose sur une obligation de contrôle *a posteriori* avec obligation de documenter la conformité des traitements. Il s'applique à toutes les personnes morales ou physiques qui collectent des données personnelles destinées à être contenues dans un fichier, que ce fichier soit électronique ou non et traité automatiquement ou pas. Le RGPD garantit le droit à l'effacement, à la rectification et à la portabilité des données, et il interdit par principe le traitement de certains types de données comme l'origine ethnique, la religion ou la santé sauf dans de très rares exceptions très encadrées par la loi. Enfin, le responsable du traitement des données (à savoir l'entreprise, même dans le cas de l'utilisation d'un sous-traitant) est dans l'obligation de déclarer les violations à la CNIL.

Les données personnelles collectées doivent être traitées de manière licite, loyale et transparente et collectées seulement pour des finalités déterminées, explicites et légitimes. Les données doivent également être adéquates, pertinentes et limitées à ce qui est nécessaire pour la finalité du traitement ; en outre, l'individu dont les données sont collectées doit en être informé, doit donner son consentement et être informé de la façon dont ses données vont être utilisées et pour quel but. L'individu conserve un droit d'accès et de rectification qui se traduit par une obligation pour l'organisme collecteur de maintenir les données exactes et à jour. Les données qui sont conservées sous une forme permettant l'identification des personnes concernées ne doivent pas l'être pendant une durée excédant celle strictement nécessaire pour la finalité du traitement, et elle doivent l'être de manière à garantir leur sécurité, tous ces processus étant placés sous la responsabilité du responsable du traitement, sur lequel

repose une obligation de documentation de la conformité et de déclaration à la CNIL des violations, avec dans certains cas graves une obligation de notification des personnes dont les données ont été compromises.

[SLIDE 22]

Pour nous placer en conformité avec ces exigences légales vis-à-vis du RGPD, nous avons donc opéré un certain nombre de retraitements sur les données de nos 3 fichiers RH avant leur export au format .csv :

- ❖ Les données ont été anonymisées : les noms & numéros de téléphone ont été supprimés, ce qui est une meilleure option que la pseudonymisation (plus sûre et offre plus de flexibilité en matière de durée maximale de conservation des données) ;
- ❖ Les données inutiles ont été supprimées, comme l'état civil ;
 - Les accidents du travail ont été conservés – même si le RGPD interdit par principe le traitement des données de santé, nous avons considéré qu'il ne s'agissait pas d'une donnée de santé à proprement parler et qu'un simple oui/non ne permettait pas de déduire une information sur l'état de santé de la personne ou les traitements reçus ;
- ❖ Certaines données ont été retraitées pour rendre l'identification impossible "avec des moyens raisonnables", pour reprendre les termes mêmes de la loi : l'employe_id a été remplacé par une clé artificielle après le mélange des lignes dans la table, la date de naissance a été remplacée par l'âge à l'année entière la plus proche, le nombre d'enfants a été remplacé par une variable booléenne ;
- ❖ Concernant la durée de conservation, elle n'est pas limitée si les données sont anonymisées mais en pratique 3 ans serait souhaitable, ce qui correspond à la limite légale pour les mesures correctives prises pour remédier aux indices inférieurs à 75 points ;

- ❖ Concernant la mise à jour régulière, elle est assurée par les RH dans le cadre de leur mission ;
- ❖ Enfin, concernant la licéité du traitement, elle est ici assurée par le fait que les données sont utilisées pour répondre à une obligation légale du cabinet, à savoir la publication de l'indice annuel d'égalité femmes-hommes.

[SLIDE 23]

Pour conclure, on soulignera les points suivants :

- Nous avons un problème d'intégrité des données avec des données manquantes qui affectent le calcul de plusieurs indicateurs obligatoires (augmentations et promotions par exemple, dans lesquelles nous avons des NULLs) ;
- Les données qui nous ont été fournies par les RH ne permettent pas de remplir notre obligation légale de calcul de l'indice d'égalité car les informations pour 2 des 5 indicateurs sont manquantes ;
- Les données pourraient être enrichies pour fournir des indicateurs plus pertinents (nombre de femmes cadres ou dans des postes dirigeants, congés maternité & paternité, formation, harcèlement & violences au travail, télétravail etc...) ;
- Les fichiers reçus des RH ne sont pas conformes au RGPD car ils ne sont pas sécurisés – *a minima*, il faudrait un mot de passe ;
- Les fichiers reçus des RH ne respectent pas la 1NF (nom/prénom dans même colonne etc...) ;
- Et enfin, il faudrait mettre en place une documentation de conformité des traitements avec le RGPD.

Et enfin, le chiffre d'accidents du travail identiques pour les 2 sexes ne doit pas masquer son importance – 1 employé sur 2 c'est énorme, surtout dans un cabinet de consulting où la pénibilité est moindre *a priori* que dans des emplois ouvriers par exemple et il faudrait envisager une analyse plus fine de cette donnée et un plan d'action le cas échéant.