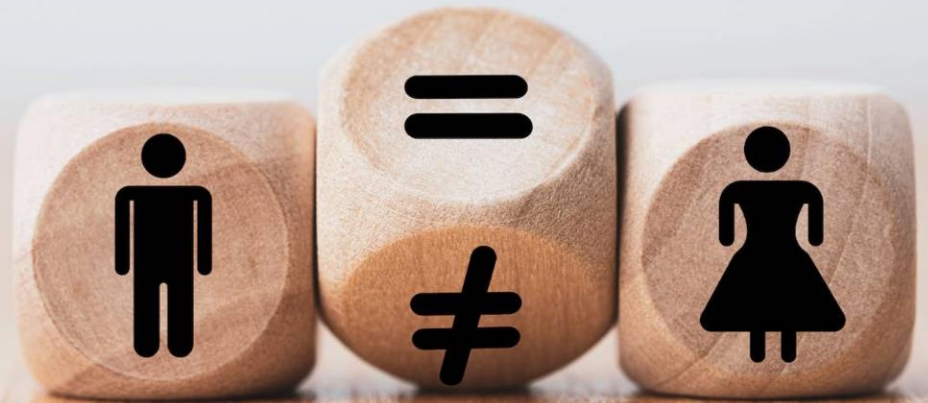


COMMUNIQUEZ VOS RÉSULTATS



OPENCLASSROOMS
Soutenance
01/06/2023 — 22:00

Parcours Data Analyst V.1
Projet 8
Vong-Raksmey OUM



INTRODUCTION

LES INÉGALITÉS DE GENRE DANS LA SPHÈRE PROFESSIONNELLE



- Cadre légal
 - lois maternité, discriminations et harcèlement
- Index d'égalité professionnelle
 - volonté sociétale de réduire ces écarts
- Étape significative
 - mais ne se concentre que sur la rémunération
- **Comment aider les entreprises à mesurer et garantir un niveau de parité sur différents aspects de la vie salariale et tout le long de la vie du contrat?**

MISE EN PLACE D'INDICATEURS CLÉ

- Inspiré de *L'outil diagnostic égalité sur Excel*






- Jeu de données *Description des emplois salariés en 2018*
- Analyses des données pour mise en place des indicateurs clé
- Application sur un cas d'étude dans un cabinet de consultants

DOMAINE DES INDICATEURS	INDICATEURS PRODUITS AUTOMATIQUEMENT PAR L'OUTIL SELON LE SEXE ET PAR CATEGORIE PROFESSIONNELLE	INDICATEURS COMPLEMENTAIRES DANS L'OUTIL PREVUS DANS LE CODE DU TRAVAIL POUR >300	DOMAINES D'ACTION EGALITE DES ACCORDS OU PLANS
EMBAUCHE	REPARTITION DES EFFECTIFS PAR CATEGORIE PROFESSIONNELLE		EMBAUCHE
	REPARTITION DES EFFECTIFS SELON L'AGE MOYEN		
	REPARTITION DES EFFECTIFS PAR TYPE DE CONTRAT	Répartition des congés de plus de 6 mois	
	REPARTITION DES EMBAUCHES PAR TYPE DE CONTRAT	Répartition selon les motifs de départ	
FORMATION	REPARTITION DU NOMBRE MOYEN D'HEURES DE FORMATION PAR SALAIRE ET PAR AN	Répartition du nombre moyen d'heures de formation selon le type de formation	FORMATION
PROMOTION	REPARTITION DES PROMOTIONS INTERNES REPARTITION DES EFFECTIFS SELON L'ANCIENNETÉ MOYENNE	Durée moyenne entre 2 promotions Répartition des effectifs selon l'ancienneté moyenne dans la CP	PROMOTION
QUALIFICATION	REPARTITION DES EFFECTIFS PAR SERVICE		QUALIFICATION
	REPARTITION DES EFFECTIFS PAR METIER		
	REPARTITION DES EFFECTIFS PAR EMPLOI		
CLASSIFICATION	REPARTITION DES EFFECTIFS SELON LES NIVEAUX D'EMPLOI DANS LA CLASSIFICATION		CLASSIFICATION
CONDITIONS DE TRAVAIL	REPARTITION DES EFFECTIFS SELON LA DUREE DU TRAVAIL	Répartition des effectifs selon l'exposition aux risques professionnels	CONDITIONS DE TRAVAIL
	REPARTITION DES EFFECTIFS SELON L'ORGANISATION DU TRAVAIL		
SECURITE ET SANTE AU TRAVAIL	REPARTITION DES ACCIDENTS DU TRAVAIL REPARTITION DES MALADIES PROFESSIONNELLES	Répartition des effectifs selon la pénibilité dont le caractère répétitif des tâches	SECURITE ET SANTE AU TRAVAIL
REMUNERATION EFFECTIVE obligatoire	EVENTAIL DES REMUNERATIONS ET REMUNERATION MOYENNE MENSUELLE	Rémunération médiane Répartition dans les 10 plus hautes rémunérations	REMUNERATION
ARTICULATION DES TEMPS	Pour les entreprises entre 50 et 299 salariés, l'entreprise doit produire au moins un indicateur d'articulation des temps parmi ceux proposés pour les entreprises de 300 salariés et plus	Existence de compléments de salaire (congé paternité, maternité, adoption)	ARTICULATION DES TEMPS
		Nombre de congés paternité pris par rapport aux jours théoriques	
		Formules d'organisation du travail facilitant l'articulation	
		Nombre d'accès au TP choisi et nombre de retours au temps plein	
		Participation aux modes d'accueil petite enfance	
		Dépenses éligibles crédit impôt famille	



1. DONNÉES PRINCIPALES

1.1 RECHERCHES

Sites Internet	Qualité / pertinence des <i>datasets</i>
	Trop générales, échelle U.E
	Incomplètes dans l'ensemble
	Trop générales, échelle mondiale
	Très intéressantes, mais insuffisantes
	Très intéressantes, mais déjà travaillées

✓ Description des emplois salariés en 2018 : jeu de données *FD_SALAAN_2018.csv*

1.2 DESCRIPTION

- Fichier .csv qui décrit les caractéristiques du salarié
- Chaque enregistrement = consolidation des périodes de travail d'un salarié sur 2018, en France métropolitaine
- Population de **2 331 440 individus** et de **31 variables**
- Intitulés + modalités précisés dans *Varmod_SALAAN_2018.csv*
- Qualités des données :
 - Valeurs manquantes = 1 987 105
 - Doublons = 44 079
 - Valeurs aberrantes : ['AGE'] max 127 ans / ['AGE'] min 0 ans
 - Conformes, cohérentes, pertinentes et accessibles

1.3 TRAITEMENT

Nettoyage	Code Python
1- Sélection des actifs (10 594)	<pre>1 df_salarie=df.loc[df['AGE'] > 15].copy() 2 df_salarie=df_salarie.loc[df_salarie['AGE'] < 70]</pre>
2- Sélection des variables pertinentes	<pre>1 df_salarie=df_salarie[['A6',"TRNNETO","AGE","CPFD","CS","SEXE","CONT_TRAV"]].copy() 2 df_salarie.rename(columns={'A6': 'Activité_6'}, inplace=True) 3 df_salarie.rename(columns={'TRNNETO': 'Salaire_net'}, inplace=True) 4 df_salarie.rename(columns={'AGE': 'Age'}, inplace=True) 5 df_salarie.rename(columns={'CPFD': 'Temps_travail'}, inplace=True) 6 df_salarie.rename(columns={'CS': 'Catégorie_SP'}, inplace=True) 7 df_salarie.rename(columns={'SEXE': 'Sexe'}, inplace=True) 8 df_salarie.rename(columns={'CONT_TRAV': 'Type_contrat'}, inplace=True)</pre>
3- Clarification des noms de variables	
4- Identification des valeurs manquantes restantes (145 202)	<pre>1 df_salarie.isna().sum() Activité_6 85368 Salaire_net 0 Age 0 Temps_travail 131206 Catégorie_SP 3525 Sexe 0 Type_contrat 97501 dtype: int64</pre>
5- Suppression des valeurs manquantes	<pre>1 df_salarie=df_salarie.dropna() 2 df_salarie = df_salarie.reset_index() 3 df_salarie=df_salarie.drop(columns=['index']).copy()</pre>
6- Mise à jour de l'index	
7- Modification des modalités:	<pre>1 np.unique(df_salarie['Sexe']) array([1, 2], dtype=int64) 1 df_salarie.loc[df_salarie["Sexe"] == 1, "Sexe"] = "Homme" 2 df_salarie.loc[df_salarie["Sexe"] == 2, "Sexe"] = "Femme" 1 np.unique(df_salarie['Temps_travail']) array(['C', 'N', 'P'], dtype=object) 1 df_salarie.loc[df_salarie["Temps_travail"] == "C", "Temps_travail"] = "Temps complet" 2 df_salarie.loc[df_salarie["Temps_travail"] == "N", "Temps_travail"] = "Non concerné" 3 df_salarie.loc[df_salarie["Temps_travail"] == "P", "Temps_travail"] = "Temps partiel"</pre>
<ul style="list-style-type: none"> ➤ à l'aide du fichier Varmod_SALAN_2018.csv ➤ pour toutes les variables codifiées ➤ sauf pour ['Salaire_net'] (faciliter les calculs futurs) 	

DATASET PRINCIPAL : df_salarie

```
1 df_salarie.sample(5)
```

	Activité_6	Salaire_net	Age	Temps_travail	Catégorie_SP	Sexe	Type_contrat
1856307	Administration publique, enseignement, santé h...	17	32	Temps complet	Professions intermédiaires de la santé et du t...	Femme	Autre type de contrat
1257331	Commerce de gros et de détail, transports, héb...	9	28	Temps partiel	Employés de commerce	Femme	CDI
284457	Administration publique, enseignement, santé h...	15	47	Temps partiel	Employés civils et agents de service de la fon...	Femme	Autre type de contrat
1661195	Commerce de gros et de détail, transports, héb...	22	48	Temps complet	Cadres de la fonction publique	Homme	CDI
1116431	Administration publique, enseignement, santé h...	13	58	Temps complet	Ouvriers qualifiés de type industriel	Femme	Autre type de contrat

➤ Population de **2 175 644 individus** et **7 variables**

➤ **2 type INTEGER** et **5 type OBJECT**

Secteur d'activité du salarié 0

Rémunération nette globale du salarié, en tranches 0

Âge du salarié 0

Durée du temps de travail hebdomadaire du salarié 0

Catégorie socioprofessionnelle du salarié 0

Genre du salarié 0

Contrat de travail du salarié 0

```
1 df_salarie.info()

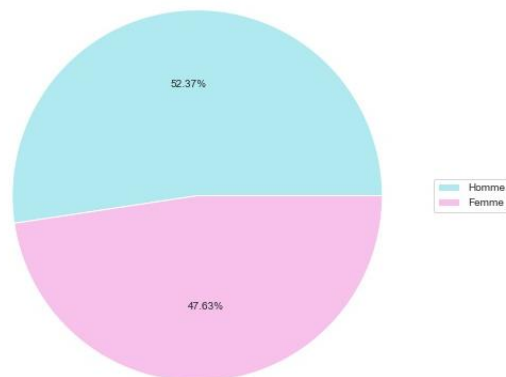
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2175644 entries, 0 to 2175643
Data columns (total 7 columns):
#   Column          Dtype
---  ---
0   Activité_6      object
1   Salaire_net     int64
2   Age             int64
3   Temps_travail  object
4   Catégorie_SP   object
5   Sexe            object
6   Type_contrat   object
dtypes: int64(2), object(5)
memory usage: 116.2+ MB
```



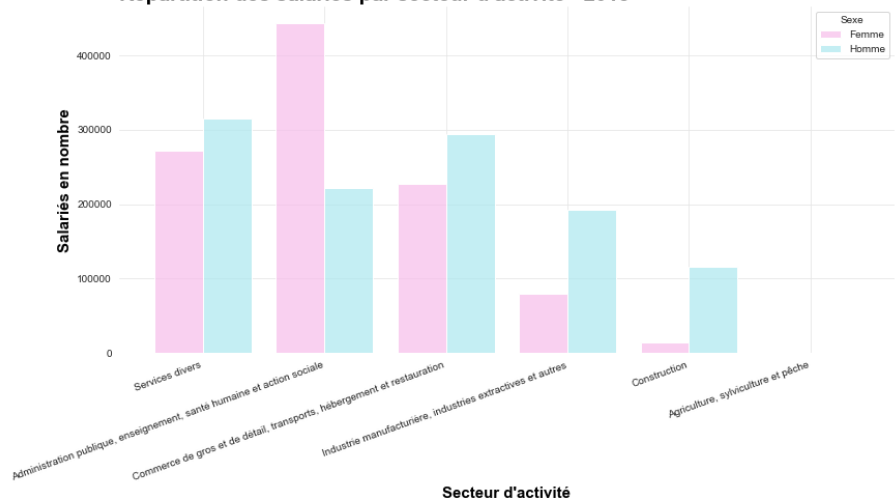
2. ANALYSES DES EMPLOIS SALARIÉS EN 2018

ANALYSE DE 1^{ER} NIVEAU

Répartition des salariés par genre - 2018

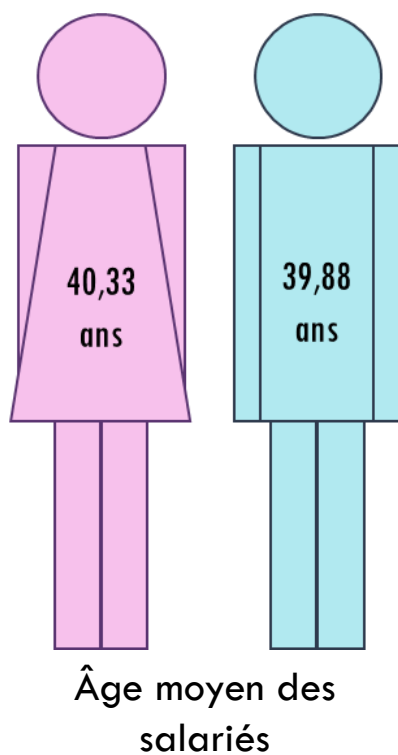


Répartition des salariés par secteur d'activité - 2018



- **47,63%** de femmes contre **52,37%** d'hommes : répartition plutôt égale
- **66,70%** de femmes dans les métiers du **care** et de l'**administration**
- **11,29%** de femmes dans les métiers de la **construction**

2.1 EMBAUCHE



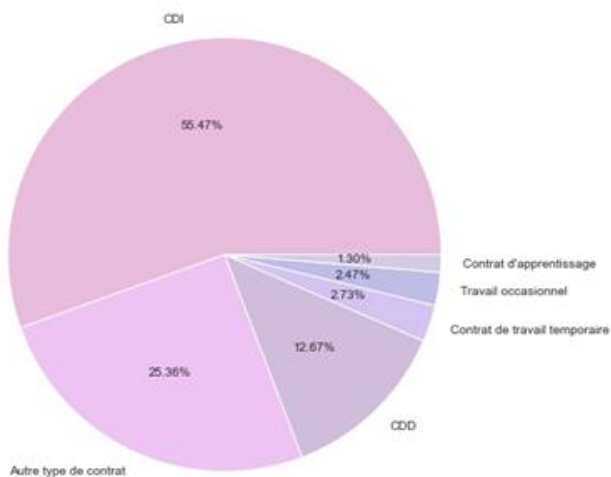
■ 5 CSP où les femmes sont le plus **sous-représentées**

	Sexe	Femme	Homme	Ensemble	% de femmes
Catégorie_SP					
Ouvriers qualifiés de type artisanal		12493	115766	128259	9.74
Chauffeurs		7761	62044	69805	11.12
Artisans		312	2262	2574	12.12
Contremaîtres, agents de maîtrise		4559	30421	34980	13.03
Ouvriers qualifiés de la manutention, du magasinage et du transport		5732	36871	42603	13.45

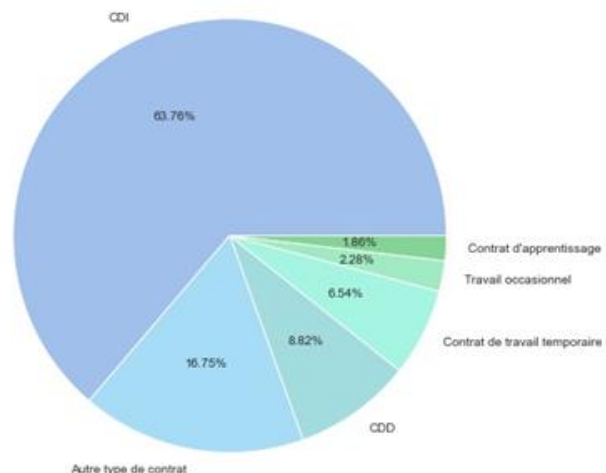
■ 5 CSP où les femmes sont le plus **sur-représentées**

	Sexe	Femme	Homme	Ensemble	% de femmes
Catégorie_SP					
Professions intermédiaires de la santé et du travail social		83183	19684	102867	80.86
Employés civils et agents de service de la fonction publique		143618	36175	179793	79.88
Professions libérales		3481	924	4405	79.02
Employés administratifs d'entreprise		140976	44579	185555	75.98
Professeurs des écoles, instituteurs et professions assimilées		54291	24153	78444	69.21

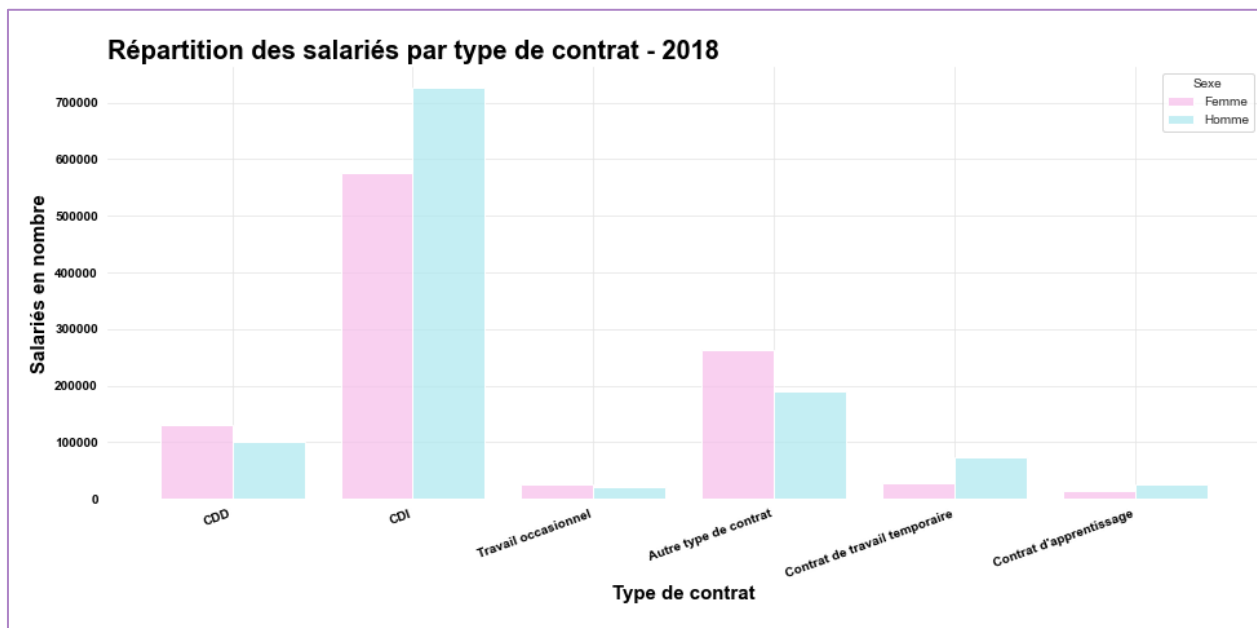
Répartition par type de contrat chez les Femmes - 2018



Répartition par type de contrat chez les Hommes - 2018



- Les femmes ont plus tendance à être en CDD ou sous d'autres types de contrats par rapport aux hommes.
- Les autres types de contrat sont :
 - Les postes d'adultes-relais
 - Le titre emploi-service entreprise (TESE)
 - Le chèque-emploi associatif (CEA)
 - Le chèque emploi-service universel (CESU)
 - Le CESU « préfinancé »
- Niveau de sécurité dans l'emploi inférieur aux hommes



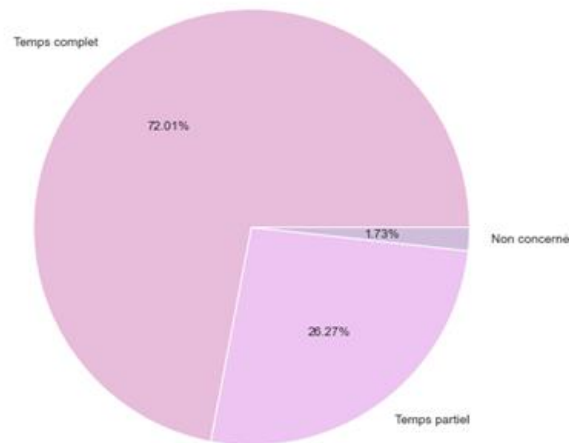
Type_contrat	Sexe	Femme	Homme	Ensemble	% de femmes	Ratio femmes / hommes	Ratio hommes / femmes
Autre type de contrat		262805	190803	453608	57.94	0.73	1.38
CDD		131253	100449	231702	56.65	0.77	1.31
Travail occasionnel		25571	21163	46734	54.72	0.83	1.21
CDI		574792	726491	1301283	44.17	1.26	0.79
Contrat d'apprentissage		13514	25953	39467	34.24	1.92	0.52
Contrat de travail temporaire		28296	74554	102850	27.51	2.63	0.38

Indicateurs clé à l'embauche créés :

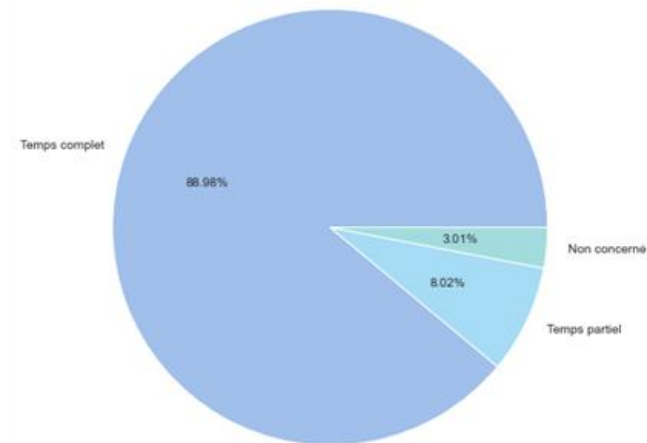
- ❑ **Ratio CDI = 1,26** : les hommes sont 1,26 fois plus en CDI que les femmes
- ❑ **Ratio CDD = 1,31** : les femmes sont 1,31 fois plus en CDD que les hommes

2.2 CONDITIONS DE TRAVAIL

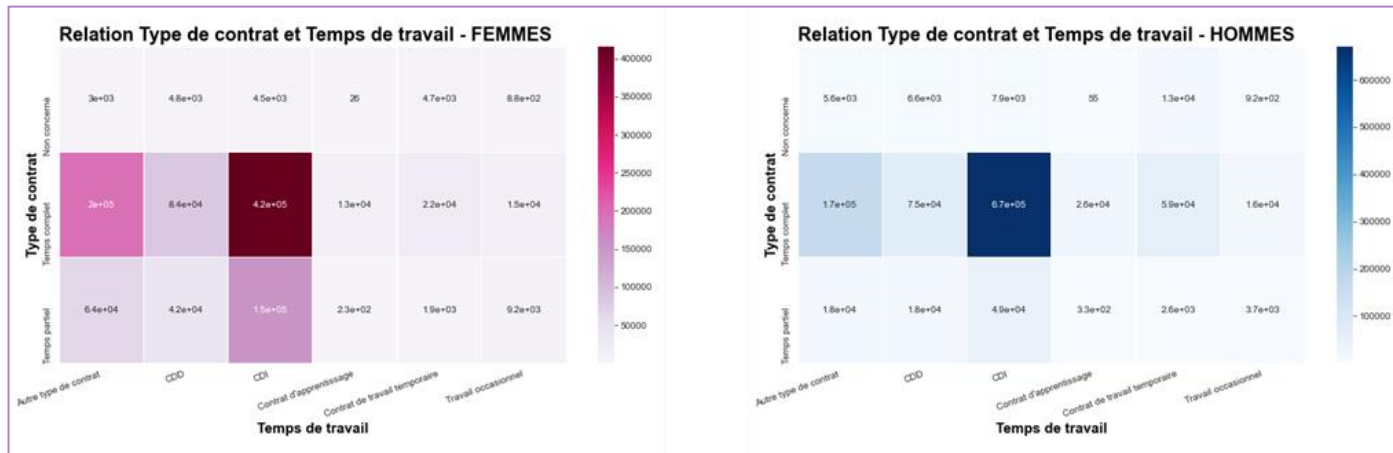
Répartition du temps de travail chez les Femmes - 2018



Répartition du temps de travail chez les Hommes - 2018



- Les femmes ont plus tendance à avoir recours au travail à temps partiel
- Explications principales du travail à temps partiel :
 - Par contrainte = durée de temps de travail rattaché au poste
 - Par choix = à la demande du salarié
- Relation de la durée de temps de travail avec le type de contrat à vérifier



- Ces graphiques montrent qu'il semble exister une corrélation entre la durée de temps de travail et le type de contrat
- Les femmes sont plus souvent à temps partiel que les hommes, peu importe le type de contrat
- Le test d'indépendance du **CHI2** vient nous confirmer cette corrélation :

```

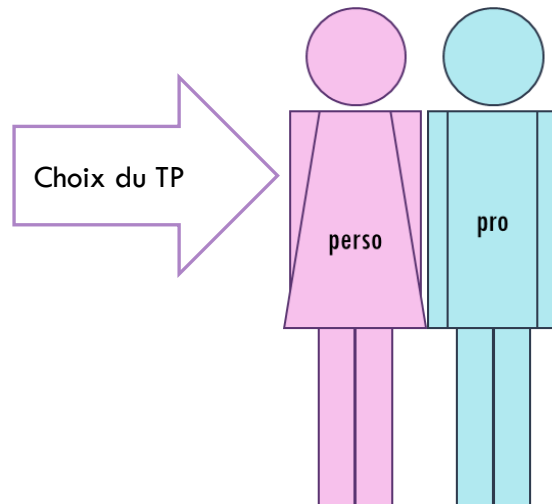
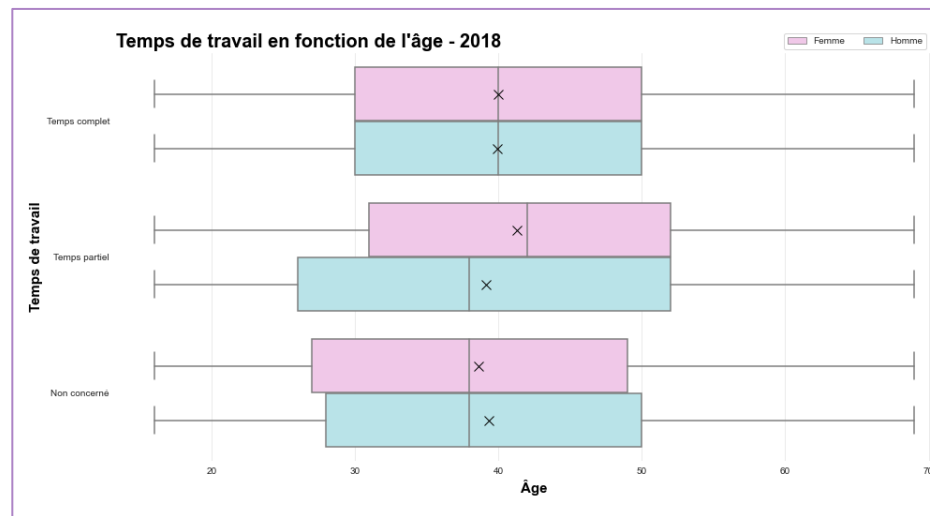
1 x = "Temps_travail"
2 y = "Type_contrat"
3
4 heat_wom = temp_wom[[x,y]].pivot_table(index=x,columns=y,aggfunc=len).fillna(0).copy().astype(int)
5
6 st_chi2, st_p, st_dof, st_exp = st.chi2_contingency(heat_wom)
7 st_chi2

```

55770.559773695866

1	st_p
0.0	

✓ P_valeur < 0,5 : H0 est rejetée au niveau de test 5% - les variables sont dépendantes



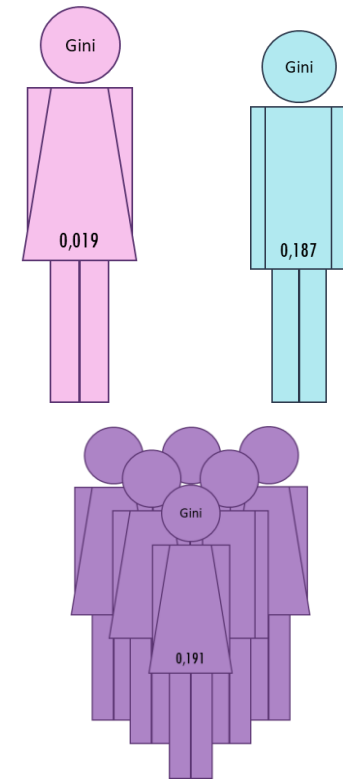
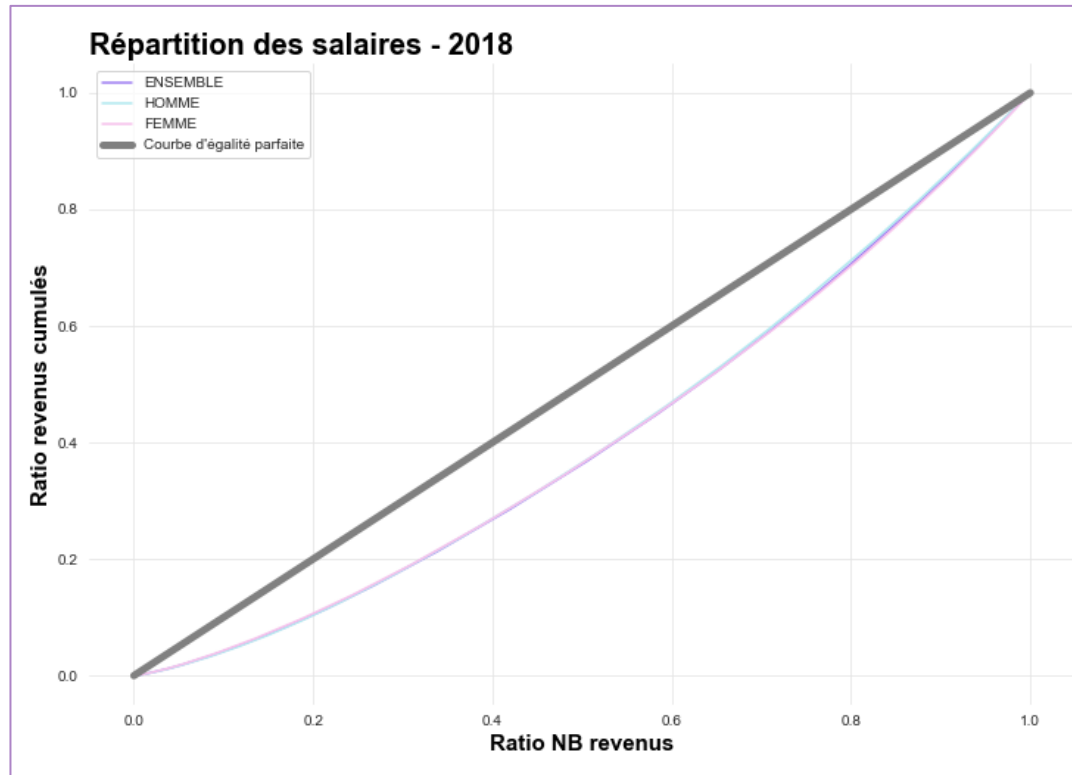
- Âge moyen à TP = femme **41,40 ans** > homme **39,17 ans**
- Âge moyen d'une primo maternité en 2018 = **30,6 ans**

	Sexe	Femme	Homme	Ensemble	% de femmes	Ratio femmes / hommes	Ratio hommes / femmes
Temps_travail							
Temps partiel		272181	91338	363519	74.87	0.34	2.98
Temps complet		746164	1013806	1759970	42.40	1.36	0.74
Non concerné		17886	34269	52155	34.29	1.92	0.52

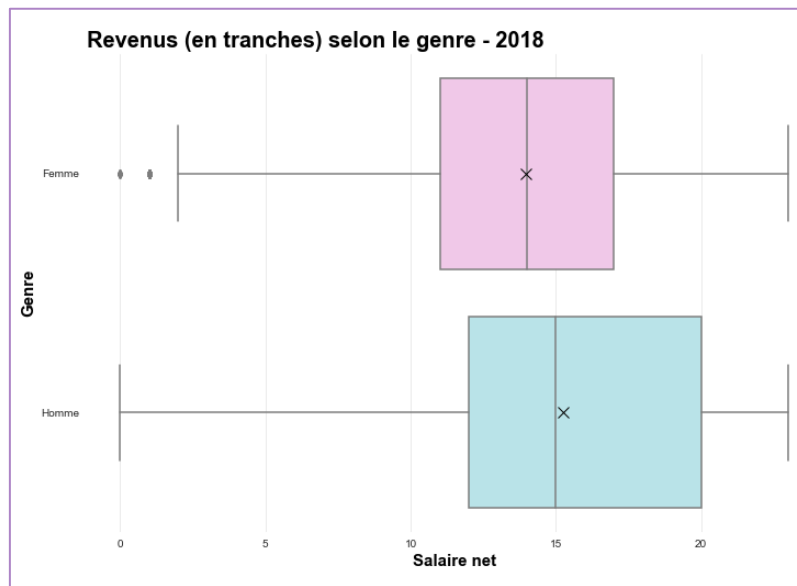
Indicateurs clé aux conditions de travail créés :

- ❑ **Ratio TC = 1,36** : les hommes sont 1,36 fois plus à TC que les femmes
- ❑ **Ratio TP = 2,98** : les femmes sont 2,98 fois plus à TP que les hommes

2.3 RÉMUNÉRATION

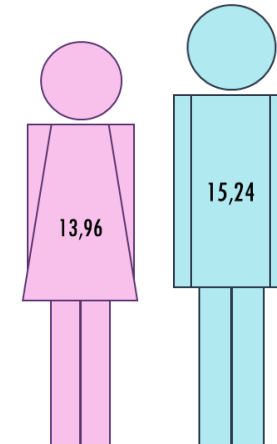


➤ Répartition des salaires très légèrement plus inégalitaire chez les femmes



Modalités de la variable ["Salaire_net"]	
0	Moins de 200 euros
1	200 à 499 euros
2	500 à 999 euros
3	1 000 à 1 499 euro
4	1 500 à 1 999 euro
5	2 000 à 2 999 euros
6	3 000 à 3 999 euros
7	4 000 à 5 999 euros
8	6 000 à 7 999 euros
9	8 000 à 9 999 euros
10	10 000 à 11 999 euros
11	12 000 à 13 999 euros
12	14 000 à 15 999 euros
13	16 000 à 17 999 euros
14	18 000 à 19 999 euros
15	20 000 à 21 999 euros
16	22 000 à 23 999 euros
17	24 000 à 25 999 euros
18	26 000 à 27 999 euros
19	28 000 à 29 999 euros
20	30 000 à 34 999 euros
21	35 000 à 39 999 euros
22	40 000 à 49 999 euros
23	50 000 euros et plus

Tranche de salaire moyenne



- **50% des femmes** sont dans une tranche de salaire **< à 14**
- **3^{ème} quartile** : un homme peut gagner 20 000€-30 000€, alors qu'une **femme** ne gagnera qu'entre **18 000€-26 000€**
- Test **ANOVA** : confirme que la différence entre les moyennes des populations femmes/hommes sont **statistiquement significatives**:

```
1 st.f_oneway(df_salarie['Salaire_net'][df_salarie['Sexe'] == 'Homme'],
2           df_salarie['Salaire_net'][df_salarie['Sexe'] == 'Femme'])
```

F_onewayResult(statistic=37659.19706678251, **pvalue=0.0**)

- ✓ P_valeur < 0,5 : H0 est rejetée au niveau de test 5% - les moyennes de chaque population ne sont pas égales

Indicateurs clé à la rémunération créés :

- ❑ **Rémunération médiane** : femme 18 000-19 999€ / homme 20 000-21 999€
- ❑ **Répartition dans les 10 plus hauts salaires***

* si les données le permettent



3. CAS D'ÉTUDE DANS UN CABINET DE CONSULTANTS

3.1 PRÉSENTATION DES DONNÉES

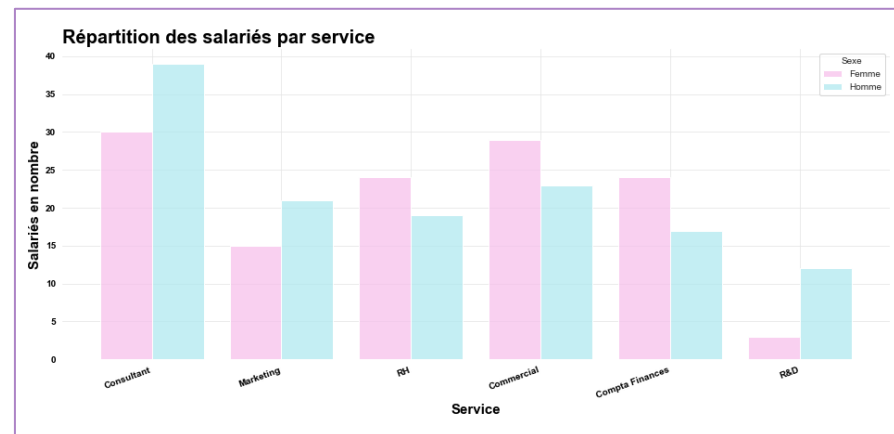
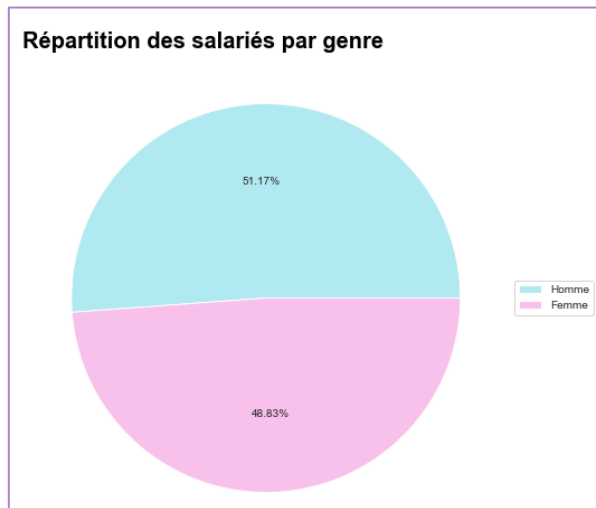
Cabinet de consultants spécialisé dans la transformation digitale des entreprises en plein développement, souhaitant appliquer une politique de recrutement de qualité afin de renforcer ses effectifs.

- Fichier .xlsx fourni par Openclassrooms
- 1 individu = 1 salarié avec 1 identifiant numérique unique
- Population de **256 individus** et de **17 variables**
- Traitement des données effectués :
 1. Remplacement des valeurs manquantes par 0.0 (salariés en CDD sans augmentation, ni promotion)
 2. Calcul de l'âge du salarié à partir de sa date de naissance
 3. Retrait des variables jugées inutiles pour le diagnostic
 4. Retrait des variables ne respectant pas le RGPD
 5. Clarification des noms de variables

1	df_oc							
	id_salarié	Contrat	Temps_trav	Salaire_brut_mens	Ancienneté_an	Service	Sexe	Age
0	16990924-2829	CDI	Temps complet	9873.5000	22.5	Consultant	Femme	47
1	16990414-3527	CDI	Temps complet	4296.7167	10.8	Consultant	Homme	67
2	16981206-3817	CDI	Temps complet	3705.2167	13.0	RH	Femme	53
3	16970415-1050	CDI	Temps complet	1697.1500	21.7	Consultant	Homme	59
4	16960906-7989	CDD	Temps complet	2848.3000	3.3	R&D	Homme	40
...
251	16020725-8401	CDI	Temps complet	7304.2667	7.9	Consultant	Femme	44
252	16020701-3376	CDI	Temps complet	8328.0167	10.2	Compta Finances	Femme	41
253	16020327-1465	CDI	Temps partiel	8860.8000	11.9	RH	Femme	49
254	16011205-0331	CDI	Temps partiel	4974.3200	11.8	Commercial	Femme	66
255	16000206-8278	CDI	Temps complet	6583.8500	12.6	Marketing	Femme	64

256 rows x 8 columns

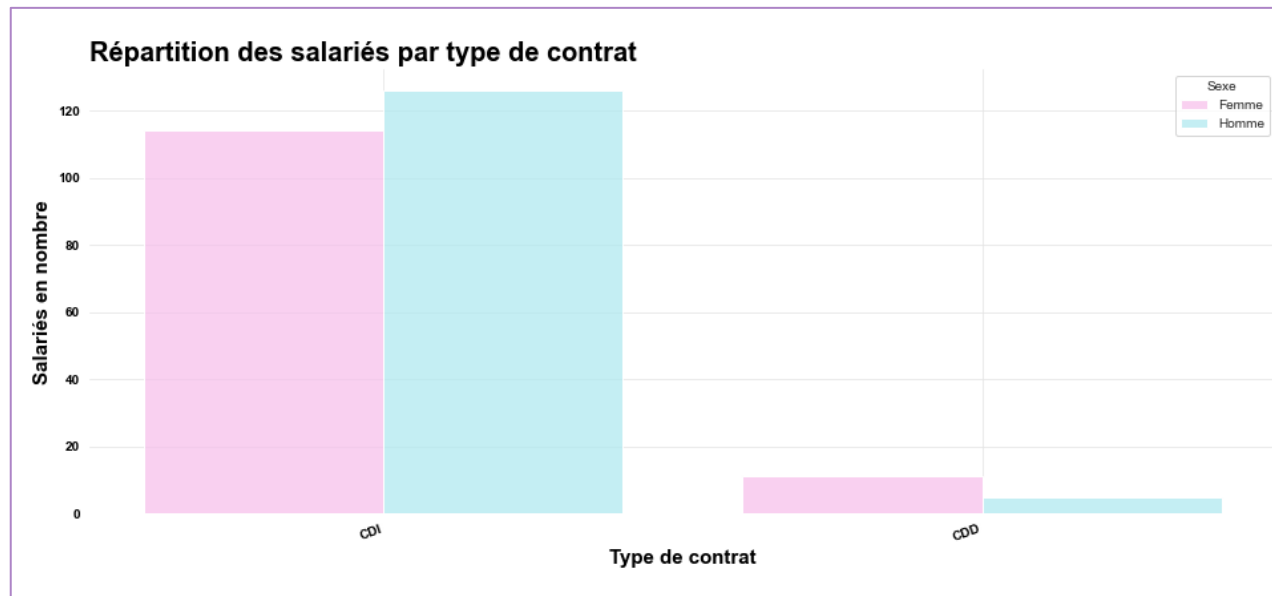
125 femmes pour 131 hommes



- Femmes > hommes en nb dans le service Commercial
- Sur-représentées : 58,54% dans le service Compta Finances
- < en nb et sous-représentées : 20% dans le service R&D

3.2 DIAGNOSTIC 1:

Répartition des effectifs par type de contrat



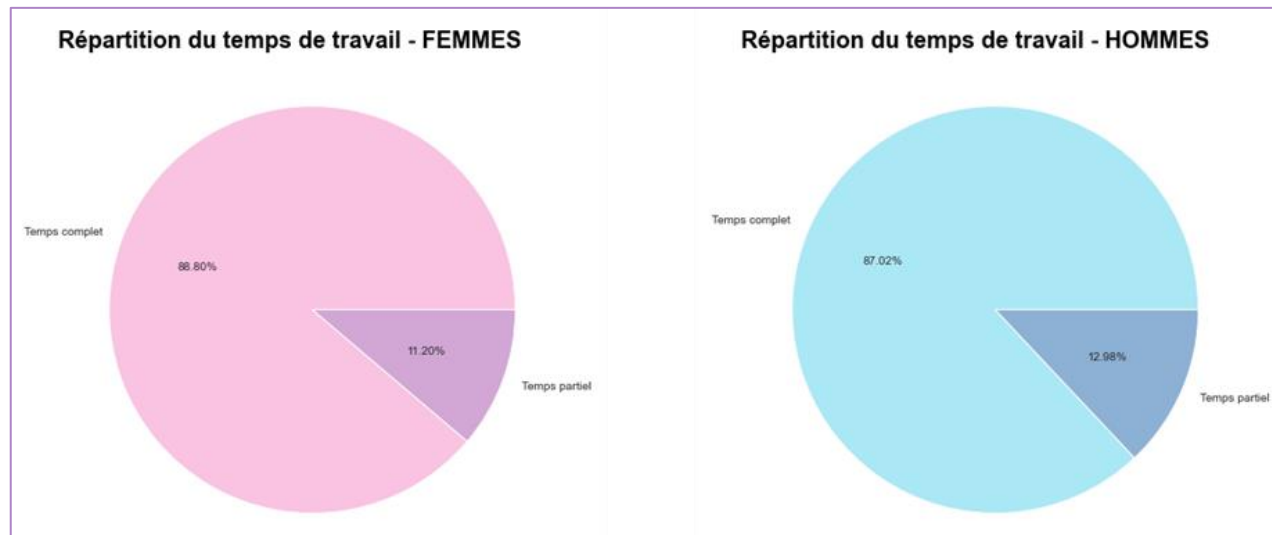
Indicateurs clé EMBAUCHE :

✓ **Ratio CDI = 1,11** : représentation proportionnée des femmes au vu de l'ensemble

✓ **Ratio CDD = 2,20** : nouvelles embauches

3.3 DIAGNOSTIC 2:

Répartition des effectifs selon la durée de travail



Indicateurs clé CONDITIONS DE TRAVAIL :

✓ **Ratio TC = 1,03:** égalité quasi parfaite

✓ **Ratio TP = 0,82 :** les hommes sont 1,21 plus nombreux en TP

3.4 DIAGNOSTIC 3: Rémunération médiane



Indicateur clé REMUNERATION :

✗ Salaire médian = 5 085€ : celui de la femme est inférieur de 8,72% à celui de l'homme

3.4 DIAGNOSTIC 3bis:

Représentation dans les 10 plus hauts salaires

```
1 top10=df_oc.sort_values('Salaire_brut_mens', ascending=False).head(10)
2 top10
```

	id_salarié	Contrat	Temps_trav	Salaire_brut_mens	Ancienneté_an	Service	Sexe	Age
6	16960619-9017	CDI	Temps complet	9999.3833	17.8	Marketing	Homme	46
135	16470418-3807	CDI	Temps complet	9978.1500	13.0	Compta Finances	Homme	52
120	16560122-3612	CDI	Temps complet	9940.2333	22.7	Marketing	Homme	47
124	16550903-3725	CDI	Temps complet	9928.1000	10.0	RH	Homme	38
0	16990924-2829	CDI	Temps complet	9873.5000	22.5	Consultant	Femme	47
107	16600919-9578	CDI	Temps complet	9785.5333	17.2	Commercial	Homme	41
248	16030220-7121	CDI	Temps complet	9782.5000	6.3	Consultant	Femme	50
118	16560416-4227	CDI	Temps complet	9777.9500	10.5	Commercial	Femme	47
62	16750821-7945	CDI	Temps complet	9737.0000	7.4	Consultant	Homme	58
201	16240230-3065	CDI	Temps complet	9642.9667	19.0	RH	Homme	60

Indicateur clé REMUNERATION :

 Répartition dans les 10 salaires les + hauts = 30% de femmes

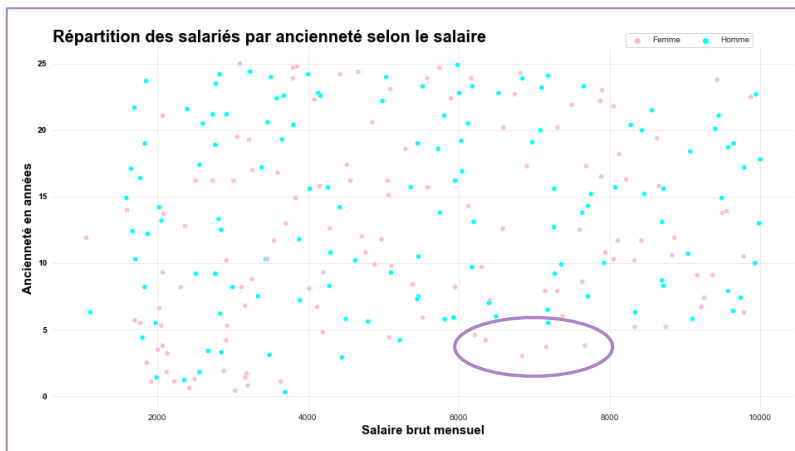
À noter:

Âge moyen : **48,6 ans** / Salaire moyen : **9 844,53€** / Ancienneté moyenne : **14,64 ans**
Tous en CDI à temps complet

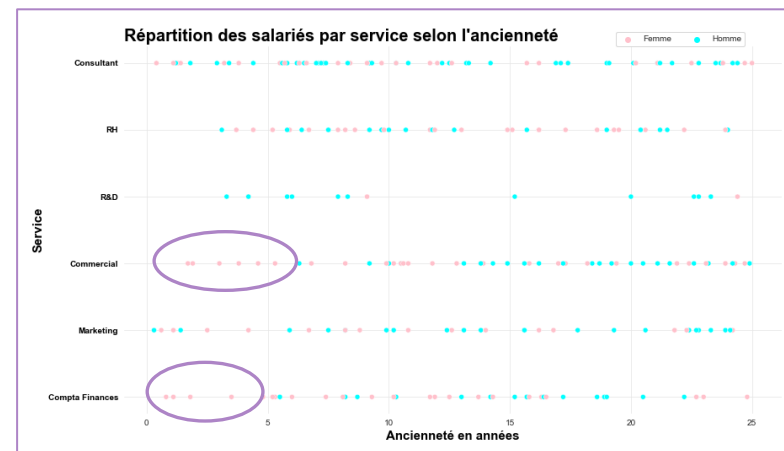
→ **Volonté de l'employeur de fidéliser ses salariés.**

3.5 PRÉCONISATIONS

Les résultats des indicateurs clé mis en place montrent que le cabinet tend vers une diminution des inégalités des sexes. Les graphiques ci-dessous en témoignent clairement:



- Salaires les plus hauts attribués majoritairement aux femmes, parmi les salariés de -5 ans d'ancienneté



- Recrutement des 5 dernières années concernant majoritairement des femmes dans 2 services

- Résultat : politique de recrutement de qualité en termes d'égalité femmes-hommes
- **Préconisation: axer le recrutement sur le service R&D**



CONCLUSION

VERS UNE DIMINUTION DES INÉGALITÉS?

- Indicateurs clé → accompagnement des entreprises → mesure des inégalités → identification des leviers → actions en faveur de la lutte contre l'inégalité des sexes dans le milieu professionnel
- Changements sociétaux:
 - Allongement du congé paternité en juillet 2021
 - Engagements RSE de la part de l'entreprise (69% sondés favorables – étude Hellowork de 2023)
- Disparités persistantes dans certains secteurs :
 - Vision de la place sociétale de la femme dans la sphère privée
 - Promotion des filières de formation
- Création d'un index de l'inégalité dans la formation



😊 **MERCI DE VOTRE ATTENTION**

OPENCLASSROOMS

Parcours Data Analyst V.1

Projet 8

« Communiquez vos résultats »

Vong-Raksmey OUM