

Diagnostic Égalité Femmes-Hommes - Workflow KNIME Automatisé

Contexte du projet

Je travaille en tant que **Data Analyst** dans un cabinet de consultants spécialisé dans la transformation digitale des entreprises. Le cabinet compte plus de 150 salariés et est en pleine croissance.

Dans le cadre de ce projet académique (OpenClassrooms), j'ai été missionnée par la **Direction des Ressources Humaines (DRH)** pour automatiser la création du rapport annuel sur l'égalité professionnelle femmes-hommes, obligatoire avant le 1er mars pour toutes les entreprises de plus de 50 salariés.

Problématique métier

L'équipe RH souhaite développer une politique volontariste en matière d'égalité femmes-hommes pour :

- **Améliorer la marque employeur** et attirer les talents
- **Respecter l'obligation légale** de publication de l'index égalité F/H
- **Automatiser le processus** de diagnostic annuel via un workflow reproductible
- **Garantir la conformité RGPD** : les données SIRH contiennent des informations personnelles sensibles

Contrainte critique : Les données sources ne sont pas anonymisées et nécessitent un traitement conforme au RGPD avant toute analyse ou diffusion.

Objectifs de la mission

Créer un **workflow KNIME automatisé** permettant de :

1. **Collecter et agréger** les données issues du SIRH (Système d'Informations RH)
2. **Anonymiser** les données personnelles conformément au RGPD
3. **Calculer** minimum 5 indicateurs du diagnostic égalité F/H
4. **Générer automatiquement** les graphiques d'analyse
5. **Exporter** un fichier CSV prêt pour analyse dans Tableau Software
6. **Présenter** les résultats en plénière d'entreprise avec recommandations

Données et périmètre

Sources de données SIRH

Les données fournies proviennent de plusieurs tables du SIRH :

Table	Contenu	Champs clés
Salariés	Informations individuelles	ID, nom, prénom, genre, date naissance, ancienneté
Postes	Informations contractuelles	Type contrat, date embauche, statut, CSP
Rémunérations	Salaires et primes	Salaire brut, primes, augmentations
Temps de travail	Volume horaire	Temps plein/partiel, taux d'activité
Carrière	Évolutions professionnelles	Promotions, formations, mobilité

Indicateurs réglementaires (Index Égalité F/H)

L'index égalité professionnelle est calculé sur **100 points** répartis en 5 indicateurs :

Indicateur	Points max	Objectif métier
Écart de rémunération F/H par CSP et tranche d'âge	40	Identifier écarts salariaux injustifiés
Écart de taux d'augmentations individuelles	20	Vérifier équité évolutions salariales

Écart de taux de promotions	15	Analyser accès égal aux promotions
Nombre de salariées augmentées au retour congé maternité	15	Protéger parentalité
Nombre de personnes du sexe sous-représenté dans les 10 plus hautes rémunérations	10	Mesurer mixité top management

Seuil légal : Un score < 75/100 impose un plan d'action correctif.

Indicateurs complémentaires sélectionnés

Au-delà de l'index légal, j'ai choisi 5 indicateurs clés pour le diagnostic :

1. **Répartition F/H par catégorie socio-professionnelle (CSP)** : Cadres, Agents de maîtrise, Employés
2. **Écart de salaire moyen en équivalent temps plein (ETP)** : Neutralise l'effet temps partiel
3. **Taux de promotion par genre** : % d'évolution de poste par an
4. **Ancienneté moyenne par genre** : Déetecte potentiel turnover différencié
5. **Répartition F/H par type de contrat** : CDI/CDD/Stage/Alternance

Outils et compétences mobilisés

Stack technique

- **KNIME Analytics Platform** : ETL (Extract, Transform, Load) et automatisation workflow
- **Nœuds KNIME utilisés :**
 - **File Reader** : Import CSV/Excel
 - **Column Filter** : Sélection colonnes pertinentes
 - **Rule Engine** : Anonymisation et recodage
 - **GroupBy** : Agrégations (moyennes, comptages)
 - **Joiner** : Jointures entre tables
 - **Math Formula** : Calculs (écart %, ratios)
 - **Bar Chart, Pie Chart, Box Plot** : Visualisations
 - **CSV Writer** : Export données préparées

Compétences data

- **ETL / Data Wrangling** : Extraction, transformation, agrégation multi-sources
- **Anonymisation RGPD** : Suppression identifiants directs, pseudonymisation
- **Calculs statistiques** : Moyennes pondérées, écarts relatifs, taux de variation
- **Data Quality** : Vérification cohérence (salaires négatifs, dates incohérentes, doublons)
- **Data Visualization** : Graphiques adaptés aux indicateurs (barres, camemberts, boxplots)
- **Conformité réglementaire** : Respect méthodologie Ministère du Travail



Livrables

1. Workflow KNIME automatisé (.knwf)

Architecture du workflow :

[Import données SIRH]

↓

[Vérification qualité] (valeurs manquantes, doublons, types)

↓

[Anonymisation RGPD] (suppression nom/prénom, génération ID anonyme)

↓

[Calcul salaire ETP] (salaire brut / taux activité)

↓

[Jointures tables] (Salariés ✕ Postes ✕ Rémunérations ✕ Temps travail)

↓

[Agrégations par genre] (moyennes, comptages, répartitions)

↓

[Calcul indicateurs] (écarts %, taux promotion, index égalité)

↓

[Génération graphiques] (5+ visualisations automatiques)

↓

[Export CSV final] → Tableau Software

Bonnes pratiques appliquées :

- **Commentaires exhaustifs** : chaque nœud documenté (objectif, transformation)
- **Nomenclature claire** : noms de colonnes explicites (`ecart_salaire_f_h_pct`, `tx_promotion_femmes`)
- **Modularité** : blocs fonctionnels réutilisables (anonymisation, calculs, exports)
- **Vérifications intermédiaires** : nœuds de contrôle qualité entre chaque étape

2. Fichier CSV anonymisé (.csv)

Contenu : Dataset agrégé prêt pour Tableau Software

- Lignes : Salariés anonymisés (ID pseudonyme uniquement)
- Colonnes : Genre, CSP, ancienneté, salaire ETP, type contrat, promotions
- **Conformité RGPD** : Aucune donnée identifiante (nom, prénom, adresse supprimés)

Structure type :

id_anonyme,genre,csp,anciennete_annees,salaire_etp,type_contrat,promotion_annee_n

EMP001,F,Cadre,5,45000,CDI,0

EMP002,H,Agent maitrise,3,32000,CDI,1

EMP003,F,Employe,2,25000,CDD,0

...

3. Présentation PowerPoint (15 slides max)

Plan de la présentation :

Slide	Contenu	Objectif
1	Contexte et enjeux	Obligation légale + amélioration marque employeur
2	Méthodologie collecte données	Sources SIRH, workflow KNIME
3	Anonymisation RGPD	Étapes de pseudonymisation

4	Vérification cohérence données	Contrôles qualité effectués
5-9	5 indicateurs clés (1 slide/indicateur)	Graphiques + analyse
10	Index égalité F/H	Score /100 + détail par indicateur
11	Synthèse forces/faiblesses	Diagnostic global
12-1 3	Recommandations actionnables	Plan d'action court/moyen terme
14	Méthodologie reproductible	Processus pour année N+1
15	Conclusion + Q&A	Ouverture discussion



Défis techniques et solutions

1. Anonymisation conforme RGPD

Problème : Les données SIRH contiennent nom, prénom, date de naissance (identifiants directs).

Solution appliquée dans KNIME :

[Rule Engine: Anonymisation]

Règle 1 : Supprimer colonnes "nom", "prenom", "adresse"

Règle 2 : Générer ID anonyme séquentiel (EMP001, EMP002...)

Règle 3 : Transformer date_naissance → tranche_age (18-25, 26-35, 36-45, 46-55, 56+)

Règle 4 : Si effectif < 5 dans une catégorie → Agréger avec catégorie supérieure

Pourquoi c'est efficient :

- **K-anonymat** : Chaque individu appartient à un groupe de minimum 5 personnes
- **Pseudonymisation** : ID anonyme irréversible sans table de correspondance externe
- **Minimisation** : Seules les données strictement nécessaires au diagnostic sont conservées

2. Calcul du salaire en équivalent temps plein (ETP)

Problème : Comparer salaires bruts biaise l'analyse (temps partiels surreprésentés chez les femmes).

Solution KNIME :

[Math Formula: Salaire ETP]

Formule : $\text{salaire_brut_mensuel} / (\text{taux_activite} / 100)$

Exemple :

- Salarié A : 2000€ brut à 80% → $2000 / 0.8 = 2500\text{€ ETP}$

- Salarié B : 3000€ brut à 100% → $3000 / 1.0 = 3000\text{€ ETP}$

Pourquoi c'est pertinent : Neutralise l'effet du temps partiel et permet comparaison équitable des rémunérations pour un même poste.

3. Détection des écarts de rémunération par CSP

Problème : L'écart global F/H peut masquer des disparités par catégorie socio-professionnelle.

Solution KNIME :

[GroupBy: Agrégation par Genre et CSP]

Groupes : genre, csp

Agrégations :

- AVG(salaire_etp) → salaire_moyen_etp
- COUNT(*) → effectif

[Pivot: Restructuration F/H en colonnes]

Lignes : csp

Colonnes : genre (F / H)

Valeurs : salaire_moyen_etp

[Math Formula: Calcul écart]

Formule : $((\text{salaire_H} - \text{salaire_F}) / \text{salaire_H}) * 100$

Insight obtenu : Écart de +8% chez les cadres, +3% chez agents de maîtrise, -2% chez employés → cibler actions sur population cadres.



Exemples de visualisations clés

1. Répartition F/H par CSP (Graphique en barres empilées)

Cadres : [redacted] 40% F | [redacted] 60% H

Agents maîtrise: [redacted] 55% F | [redacted] 45% H

Employés : [redacted] 70% F | [redacted] 30% H

Interprétation : Sous-représentation des femmes dans l'encadrement (plafond de verre).

2. Écart de salaire moyen ETP par CSP (Graphique en barres)

Écart F/H (%)

Cadres : [redacted] +8% (défavorable aux F)

Agents : [redacted] +3%

Employés : [redacted] -2% (favorable aux F)

Global : [redacted] +5%

Recommandation : Audit des grilles salariales cadres + politique de ratrapage ciblée.

3. Taux de promotion par genre (Camembert ou barres)

Femmes : 12% promus sur l'année

Hommes : 18% promus sur l'année

Écart : -6 points (défavorable aux F)

Action : Révision des critères de promotion + sensibilisation managers aux biais inconscients.

Résultats et recommandations

Score Index Égalité F/H

Score total entreprise : 78/100  (seuil légal = 75/100)

Indicateur	Score obtenu	Score max	Analyse
Écart rémunération	32/40	40	Écart +5% global, +8% cadres
Taux augmentations	18/20	20	Quasi-égalité (écart < 2%)
Taux promotions	10/15	15	Écart -6 points défavorable F
Retour maternité	15/15	15	100% des retours = augmentation
Top 10 rémunérations	3/10	10	3 femmes / 7 hommes (30% F)

Forces identifiées

1.  **Politique parentalité exemplaire** : 100% retours maternité = augmentation
2.  **Égalité augmentations collectives** : Écart négligeable F/H
3.  **Mixité employés** : 70% de femmes, bonne représentation

Faiblesses à corriger

1.  **Écart salarial cadres** : +8% en défaveur des femmes
2.  **Taux de promotion** : -6 points pour les femmes

3.  **Plafond de verre** : Seulement 40% de femmes cadres, 30% dans TOP 10 rémunérations

Recommandations actionnables

Court terme (0-6 mois) :

1. **Audit salarial cadres** : Analyser poste par poste les écarts injustifiés
2. **Plan de rattrapage** : Budget dédié pour résorber écarts identifiés
3. **Sensibilisation managers** : Formation aux biais inconscients dans les promotions

Moyen terme (6-12 mois) : 4. **Objectifs chiffrés** : 50% de femmes dans les promotions cadres 5. **Mentorat féminin** : Programme d'accompagnement talents féminins 6.

Transparence grilles salariales : Publication fourchettes par poste

Long terme (1-3 ans) : 7. **Pipeline talents** : Recrutement paritaire sur postes cadres 8. **Flexibilité horaires** : Faciliter articulation vie pro/perso (tous genres) 9. **KPIs mensuels** : Dashboard RH temps réel pour suivi continu

Évolution historique (si données disponibles)

Score Index Égalité F/H :

Année N-2 : 72/100 (sous seuil légal)

Année N-1 : 75/100 (seuil atteint)

Année N : 78/100 (amélioration +3 pts)

Constat : Efforts continus depuis 2 ans, tendance positive à poursuivre.

Structure des fichiers

```
📁 Projet-KNIME-Egalite-FH/
  ├── 📄 README.md (ce fichier)
  └── 📁 data/
    |   ├── raw/
    |   |   └── donnees_sirh.zip (sources SIRH)
    |   └── clean/
    |       └── BERNARD_Olivia_2_fichier_csv_092025.csv
  └── 📁 workflow/
```

```
|   └── BERNARD_Olivia_1_workflow_KNIME_092025.knwf
|   ├── presentation/
|   └── BERNARD_Olivia_3_support_presentation_092025.pptx
└── docs/
    ├── methodologie_anonymisation_RGPD.md
    └── guide_reproductibilite_workflow.md
```

Valeur ajoutée professionnelle

Ce projet démontre ma capacité à :

- **Maîtriser KNIME** : Conception workflow ETL complexe, automatisation complète
- **Garantir conformité RGPD** : Anonymisation, pseudonymisation, minimisation données
- **Calculer indicateurs réglementaires** : Index égalité F/H selon méthodologie officielle
- **Automatiser reporting RH** : Workflow reproductible année après année sans intervention manuelle
- **Communiquer insights** : Présentation claire adaptée à audience non-technique (plénière)
- **Proposer actions concrètes** : Recommandations priorisées court/moyen/long terme

Impact métier : Gain de 5 jours/an pour l'équipe RH + amélioration continue score égalité F/H (78 → objectif 85/100 en N+1).

Ressources utiles

- [Guide Index Égalité F/H - Ministère du Travail](#)
 - [Documentation KNIME Analytics Platform](#)
 - [RGPD : Anonymisation des données - CNIL](#)
-