

La place de l'égalité homme femme au sein des entreprises d'Ile de France

BELDJOUDI Sarah, MAWETE Anisca
Étudiantes en Master 2 Data Management à l'École Paris School of Business

Tout d'abord, nous tenons à remercier notre professeur Mr MAMAVI Olivier et toute la classe de Master 2 Data Management de l'École Paris School of Business.

Résumé :

L'objectif principal de la plupart des entreprises a toujours été de créer du profit, mais il existe plusieurs façons pour elles de prendre en compte leurs impacts sociaux et environnementaux ; c'est la Responsabilité Sociétale des Entreprises. Depuis plusieurs années maintenant, l'enjeu environnemental et sociétal est devenu un sujet central. En France, l'État régit les pratiques de la RSE à l'aide de la norme ISO26000 où il est énoncé ce qu'il est attendu en matière environnementale, sociale et sociétale.

Compte tenu de l'importance du sujet sociale au sein des entreprises, nous avons créé une base de données comportant 657 observations reposant sur l'égalité homme femme et la santé économique des entreprises de plus de 50 salariés d'Ile de France entre 2018 et 2020. Les données ont été classées par SIREN, par domaine d'activités et par effectif. La base de donnée est à retrouver en cliquant [ici](#).

Mots clés : RSE, entreprise, égalité homme femme, Ile de France, France, chiffre d'affaires, domaine d'activités, index ega pro

Titre du projet : La place de l'égalité homme femme au sein des entreprises d'Ile de France

Auteurs : BELDJOUDI Sarah (producteur owner, collection identifier, data collector, data manager, data publisher), MAWETE Anuisca (scrum master, collection identifier, data collector, data manager, data publisher)

Contexte :

Cette recherche a été faite dans le but d'observer l'implication RSE des entreprises localisées en Ile de France. Ici, avec l'index egapro, nous pouvons mesurer l'écart des salaires entre les sexes avec une note sur 100. Cette note est attribuée selon plusieurs critères (écart de rémunération, écart de répartition, nombre de salariés augmentés, présence d'au moins 4 femmes parmi les plus hautes rémunérations) et doit être de 75/100 minimum pour les entreprises sous peine de recevoir des sanctions financières si l'écart n'est pas réduit dans les trois années qui suivent. La particularité de cette donnée est de pouvoir mesurer une évolution sur plusieurs années afin de déterminer si les organisations se sont impliquées davantage en matière de RSE.

À la suite de ces analyses, nous avons par ailleurs remarqué que le secteur du transport était celui où il y avait une plus grande égalité homme femme. Entre 2018 et 2020, nous remarquons une augmentation de l'index egapro, qui peut être effectivement liée aux sanctions encourues.

Description des variables :

SIREN (string) :	Numéro d'identité de l'unité légale attribué par l'INSEE à toutes les personnes physiques ou morales inscrites au répertoire ainsi qu'à leurs établissements.
annee (string) :	Période étudiée, de 2018 à 2020
CA (string) :	Chiffre d'affaire de chaque établissement
note (string) :	Note de l'index egapro attribué à chaque établissement
section_et (integer) :	Activité principale de l'établissement
effectif (integer) :	Nombre de personne dans l'entreprise par tranche
NAF_unite_legale (integer) :	Code attribué par l'INSEE à chacun des secteurs d'activités économiques

Détails du projet :

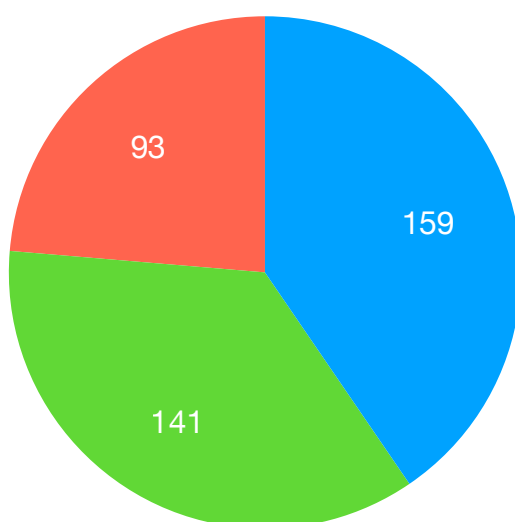
En premier lieu, nous avons été confronté à un problème de volume des données notamment sur les bases comprenant les informations générales des entreprises françaises. La solution a donc été de filtrer directement par région Ile de France et par effectif en se procurant chaque base de donnée une par une. Cela nous a permis de rassembler toutes les bases entre elles en en gardant une seule afin de nettoyer les valeurs aberrantes et inutiles à notre projet.

Nous avons répété la même opération pour la base de donnée concernant le chiffre d'affaire où nous avons seulement gardé les chiffres d'affaire de 2018 à 2020 de manière à pouvoir les joindre à notre index egapro uniquement disponible sur cette période. Le fait de se baser sur une période de deux ans au moins rend les résultats plus intéressants à interpréter.

Concernant l'utilisation de nos outils, nous avons extrait nos tables sous Excel avant de pouvoir les soumettre sous le logiciel R Studio. Ce logiciel nous a permis de visualiser, de nettoyer et de joindre nos tables de manière efficace.

Domaines d'activités les plus représentés- en nombre d'observations

- Commerce et répartition automobiles et de motocycles
- Activités spécialisés scientifiques et techniques
- Industrie manufacturières



Évolution de la moyenne de la note egapro de 2018 à 2020

