

MISE EN PLACE D'UNE BASE DE DONNÉES

Diallo Coumba, Pamidjo Jocelyne, Ben Nadjelaa, Nguy Sioutyne

ENVIRONNEMENT

Scrum master : Sioutyne NGUY

Product owner : Nadjela BEN

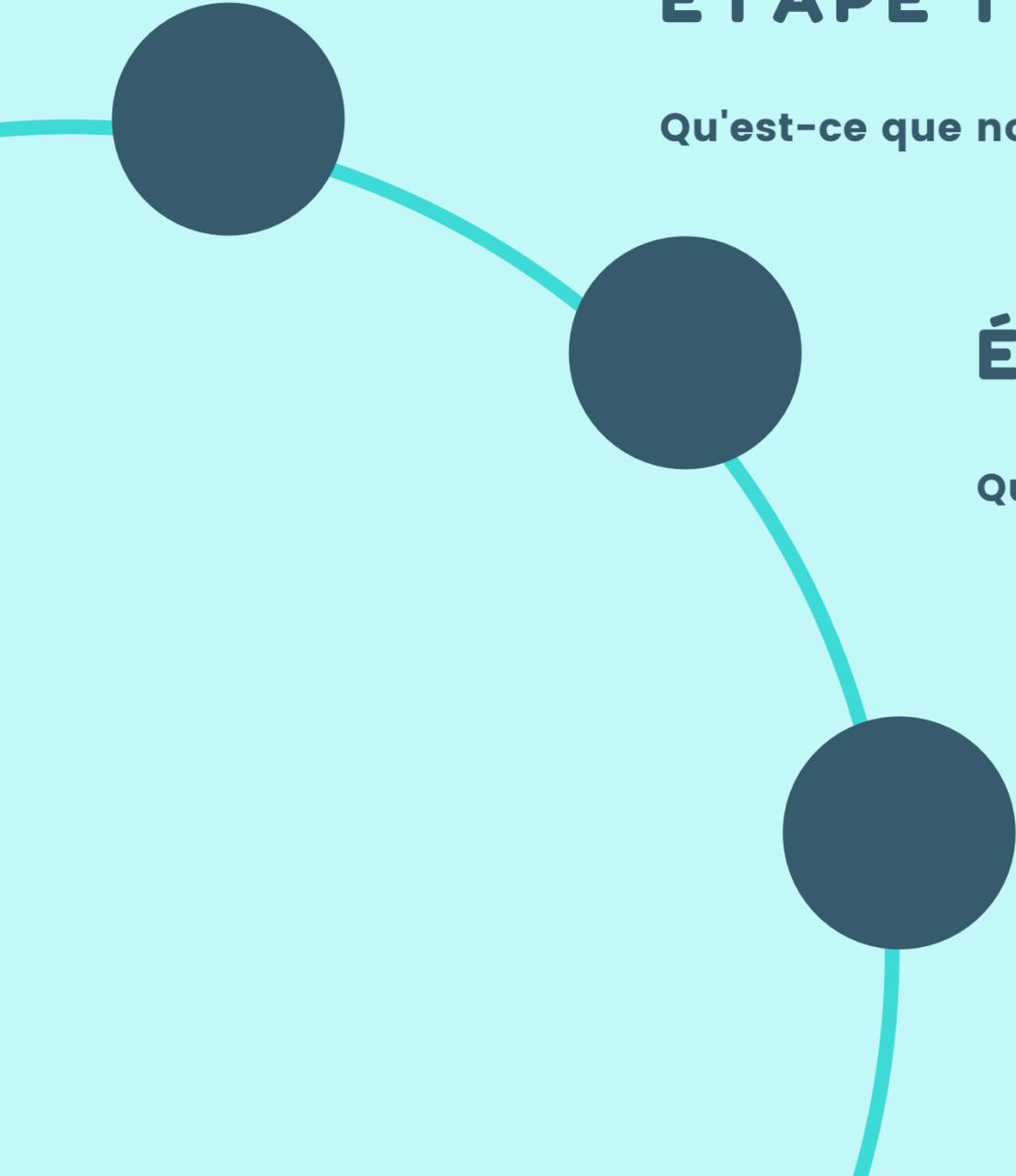
Développeurs :

- Coumba DIALLO
- Jocelyne TRAORE
- Nadjela BEN
- Sioutyne NGUY

CARNET DU PRODUIT

Description	Estimation Effort	Estimation Valeurs	Priorité
Catalogue	7	Moyen	1
Création des tables	6	Moyen	5
Mise en place de la base de données	8	Forte	3
Mise à disposition au niveau de Data paper	5	Moyen	7

SCRUM MEETING



ÉTAPE 1

Qu'est-ce que nous devons faire pour atteindre l'objectif de sprint?

ÉTAPE 2

Qu'est-ce que nous avons pour atteindre l'objectif de sprint?

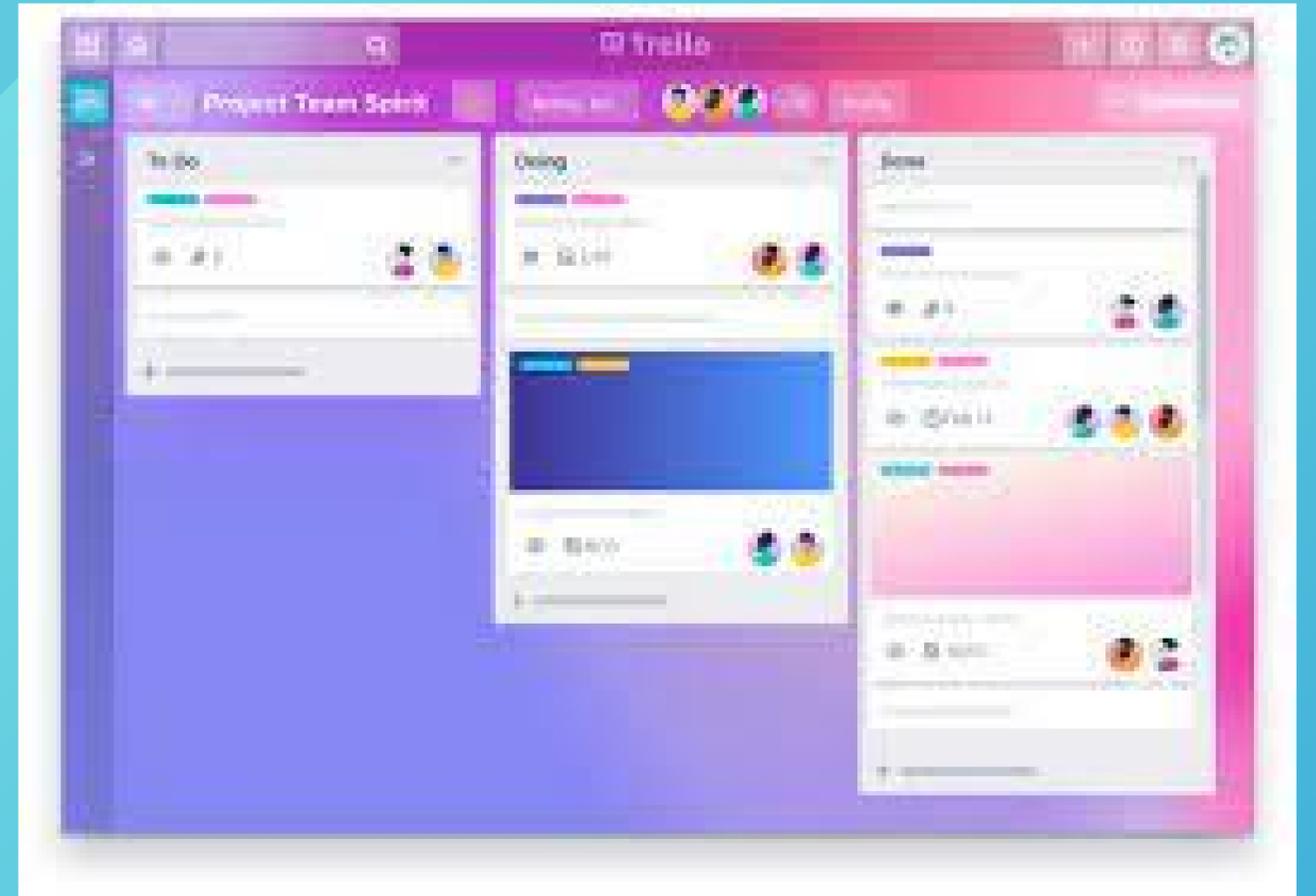
ÉTAPE 3

Comment doit-on procéder pour atteindre l'objectif de sprint ?

OUTILS DE COMMUNICATION ET GESTION DE PROJET



ild





- **Une messagerie directe permet le partage de fichiers**
- **Possibilité de lancer des discussions vidéo et faire de partage d'écran**



- **Une traçabilité des actions menées**
- **Prise en main rapide et facile**

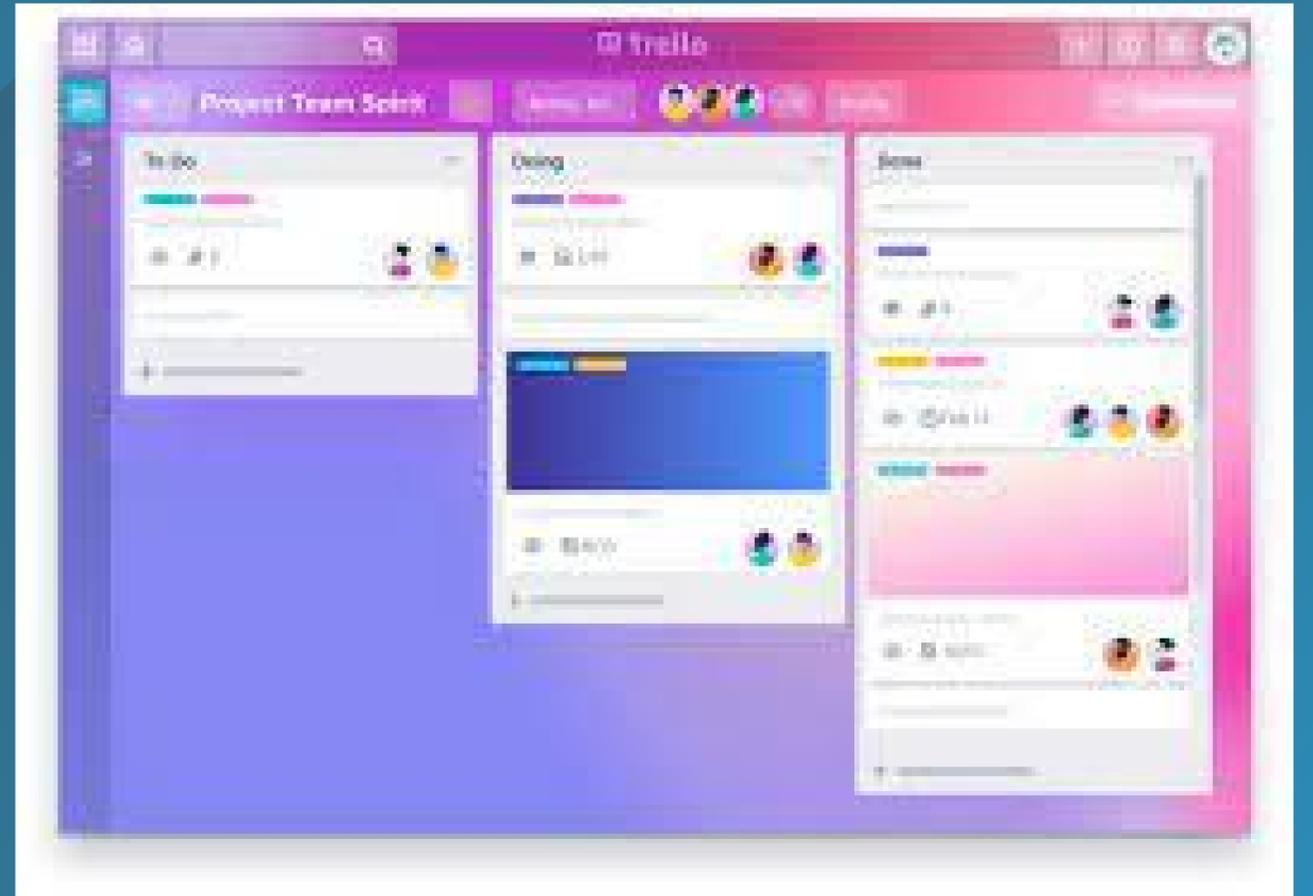


- **Moyen simple pour échanger des messages entre l'équipe**

OUTILS DE DEVELOPPEMENT



Python
PhpMyadmin
MySql
Google Data Studio





Interroger les tables et créer des jointures



Visualisation des KPI et fournir le Data paper

Projet SCRUM

MISE EN PLACE D'UNE BASE DE DONNÉES

Sprint 1

20/11/2021

Diallo Coumba, Pamidjo Jocelyne, Benmesmoudi Nadjelaa, Nguy Sioutyne

CHAPITRE 5



Identifier les sources de données



Tâches Réalisées

Définir les indicateurs RSE

Indicateurs sociétaux :

Nombre salarié Homme Femme

Nombre de turnover

Type de contrat

Taux d'absentéisme

Taux d'ancienneté

Indice de satisfaction de qualité de vie et de santé du travail

Taux d'activité selon le sexe

Ecart salariale entre homme et femme

Nombre de jours télétravail

Mobilité des salariés

Sources de données

DATA GOUV



Plateforme de diffusion de données
publiques de l'État français

INSEE



L'Institut national de la statistique et des
études économiques est chargé de la
production, de l'analyse et de la publication
des statistiques officielles en France



Jeux de données :

EGALITE FEMMES-HOMMES

https://www.insee.fr/fr/outil-interactif/5367857/tableau/40_SOC/44_EGF#

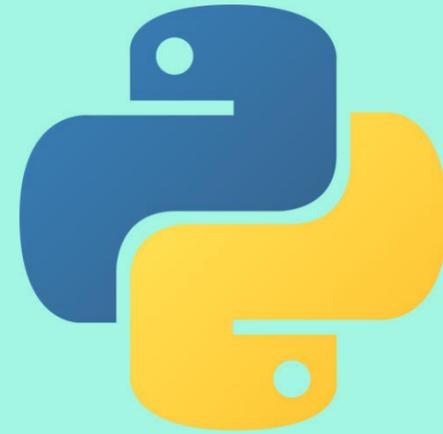
ÉCART SALARIALE ENTRE LES FEMMES ET LES HOMMES DANS LE SECTEUR PRIVE

<https://www.insee.fr/fr/statistiques/2407748#tableau-figure1>

SALAIRE HORAIRE MOYEN DU MOIS DE RÉFÉRENCE ET RÉMUNÉRATION BRUTE TOTALE DANS LES ENTREPRISES DE 10 SALARIÉS ET PLUS, Y COMPRIS FONCTION PUBLIQUE D'ÉTAT EN 2014

<https://www.insee.fr/fr/statistiques/3049114?sommaire=3049191&q=type+de+contrat+entreprise>

Outils utilisés



Collecter les données
avec Python

Parsehub, logiciel pour
faire du scrapping



WEB SCRAPING

Le web scraping, appelé aussi harvesting, est une technique d'extraction du contenu de sites Web, via un script ou un programme, dans le but de le transformer pour permettre son utilisation dans un autre contexte, par exemple le référencement.



Résultats

Fichier sous format csv, issu des jointures entre les deux base de donnée 'Variables Siren' et 'Egalité hommes-femmes'

Difficultés



Projet SCRUM

MISE EN PLACE D'UNE BASE DE DONNÉES

Sprint 2

27/11/2021

Diallo Coumba, Pamidjo Jocelyne, Benmesmoudi Nadjelaa, Nguy Sioutyne

Elaboration du modele de la base de données

Nom de la table

te_siren_admin

Séparation des libellés

Meilleure analyse du contenu de la table

=> L'objectif : avoir un seul modèle de base de donnée qui regroupent toutes les informations pertinentes des entreprises

Colonne	Type	Commun	Important	A supprimer
siren	number			
nic	number			
siret	number			
dateCreationEtablissement	YYYY-MM-DD HH:MIN:SEC			
tracheEffectifsEtablissement	number			
anneeEffectifsEtablissement	YYYY-MM-DD			
activitePrincipaleRegisteMetiersEtablissement	NULL			
dateDernierTraitementEtablissement	YYYY-MM-DD HH:MIN:SEC			
etablissementSiege	boolean			
nombrePeriodesEtablissement	number			
dateDebut	YYYY-MM-DD HH:MIN:SEC			
etatAdministratifEtablissement	STRING			
enseigne1Etablissement	NULL			
enseigne2Etablissement	NULL			
enseigne3Etablissement	NULL			
denominationUsuelleEtablissement	NULL			
activitePrincipaleEtablissement	STRING			
nomenclatureActivitePrincipaleEtablissement	STRING			
caractereEmployeurEtablissement	STRING			

Jeux de données II



ACTIVITÉ, EMPLOI ET CHÔMAGE EN 2020 ET EN SÉRIES LONGUES

Objectif : déterminer la population active et le taux d'activité par sexe

Piste : possibilité de faire une jointure entre le taux d'activité homme femme et la table des entreprises par le secteur d'activité

Problème : trouver des ressemblances entre les secteurs d'activité de cette table et celle de notre modèle de donnée

MASSE SALARIALE ET ASSIETTE CHÔMAGE PARTIEL MENSUELLES DU SECTEUR PRIVÉ, PAR RÉGION (2007 - 2020)

Objectif : est de comprendre où se retrouve le plus la masse salariale (dans le secteur public ? privée ?)

Aussi quelle région en France regroupe le plus de salaire

Problème : identification des entreprises privée / publique

Solution : Etablir une liste d'entreprise privée

NOMBRE D'ÉTABLISSEMENTS EMPLOYEURS ET EFFECTIFS SALARIÉS DU SECTEUR PRIVÉ, PAR DÉPARTEMENT X TRANCHE D'EFFECTIF (2006-2020)

Objectif : Identifier les entreprises du secteur privée - ce qui nous aider à construire notre table

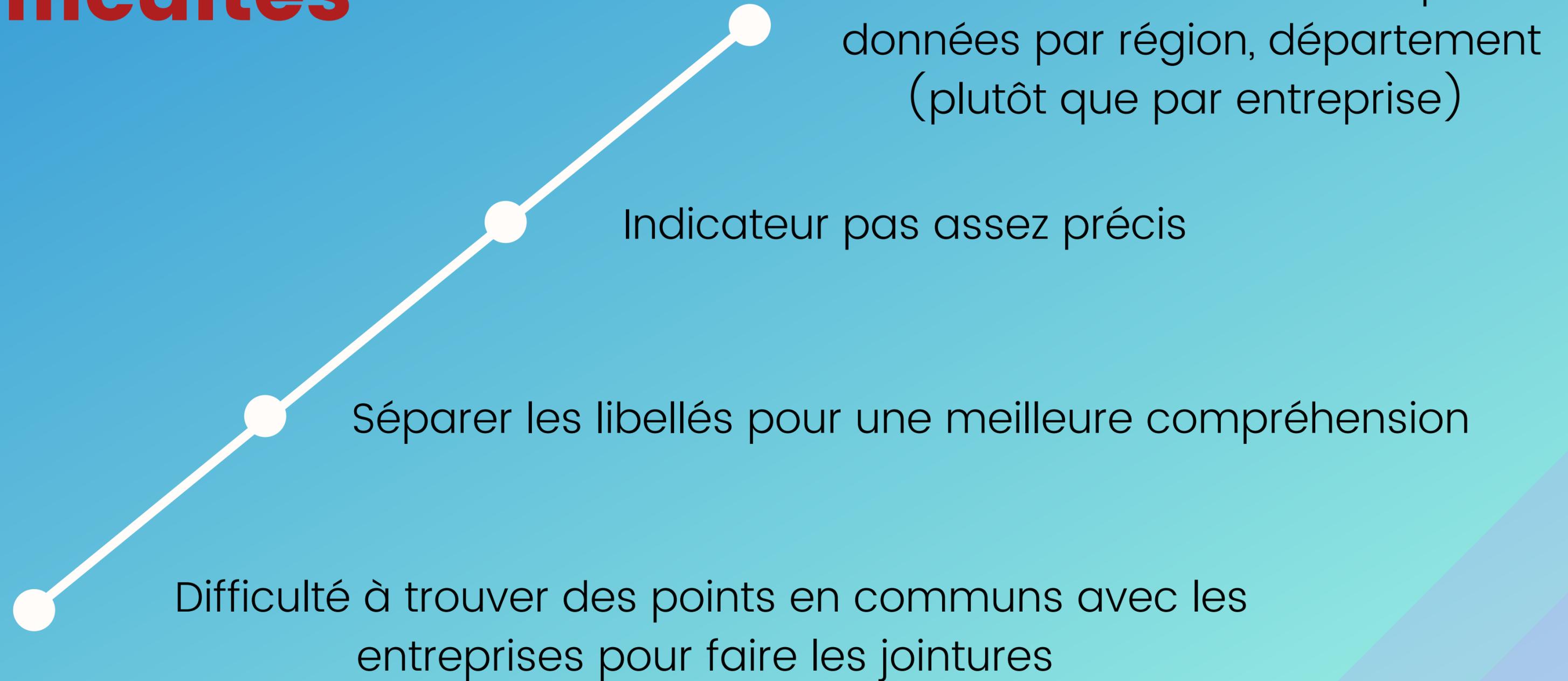
Solution : Si le code_APE est compris dans cette liste alors l'entreprise est privée sinon elle est publique

Outils utilisé



Python, Excel, Datastudio (problème de volumétrie)

Difficultés



Ce qui reste à faire



- **Consolidation d'une base de donnée commune des entreprises**
- **Faire un tri des jeux de données collectés entre sprint 1 et 2**
- **Travailler sur les jeux de données trouvés et les transformer afin de pouvoir les jointer avec notre BDD**

Projet SCRUM

MISE EN PLACE D'UNE BASE DE DONNÉES

Sprint 3

04/12/2021

Diallo Coumba, Pamidjo Jocelyne, Benmesmoudi Nadjelaa, Nguy Sioutyne

Réussi

- Ciblage des tables de données utilisés
- Définition des jointures
- Réalisation du schéma relationnel

Difficultés

- Impossible d'importer les tables : python avec pandas, php my admin, postgres (point bloquant)
- Meilleure démarche technique à suivre (travailler par bd ou par table)
- Temps limité

Ce qui reste à faire

- **Finaliser le schéma relationnel**
- **résoudre le problème d'importation (problème bloquant) pour pouvoir manipuler les tables**
- **Choix des statistiques descriptives**



Seance 1



- Définition des rôles
- Carnet du produit
- Scrum meeting
- Outil de communication
- Outil de developpement

+ Ajouter une carte



Sprint 1



- Définition des indicateurs RSE
- Trouver les bases de données pertinentes
- Difficultés rencontrées

+ Ajouter une carte



Sprint 2



- Elaboration du modele de la base de données
- Bien cibler les BDD
- Difficultés rencontrées

+ Ajouter une carte



Sprint 3



- Nettoyage de la BDD
- Définir les jointures
- Créer une architecture sous forme graphique
- Préparation du sprint review

+ Ajouter une carte



Projet SCRUM

Totally Sprint

SPRINT 4 FINAL

Présentation

10/12/2021

Diallo Coumba, Pamidjo Jocelyne, Benmesmoudi Nadjelaa, Nguy Sioutyne

**Étape 1 : Rappel du
Contexte et
objectifs**

Contexte

- **Partir de données existantes**
- **Croiser avec des données
recherchées**
- **Déduire des informations
pertinentes et utiles pour
l'évolution des entreprises**

Pourquoi Totally Sprint ?

Objectif

**Aider les entreprises à
s'améliorer en analysant
leur données**



ENVIRONNEMENT

Scrum master : Sioutyne NGUY

Product owner :
Nadjela BENMESMOUDI

Développeurs :

- Coumba DIALLO
- Jocelyne TRAORE
- Nadjela BEN
- Sioutyne NGUY

Responsabilité sociale des entreprises (RSE)

C'est la mise en place de l'ensemble des actions par les entreprises dans le but de respecter les principes du développement durable (social, environnemental et économique).



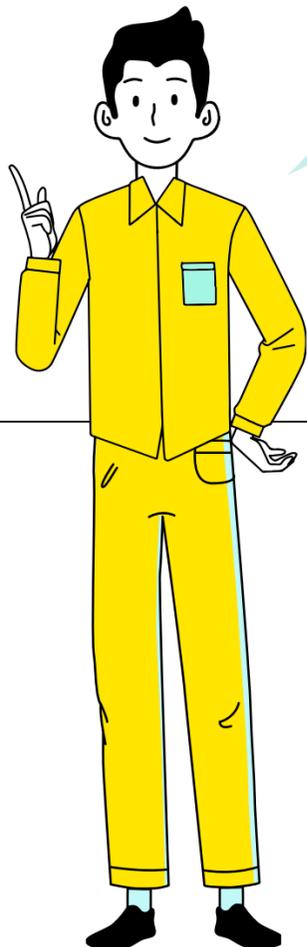
« Parce que la réussite collective passe par l'action de chacun »

Outils utilisés



Python, Excel, Trello

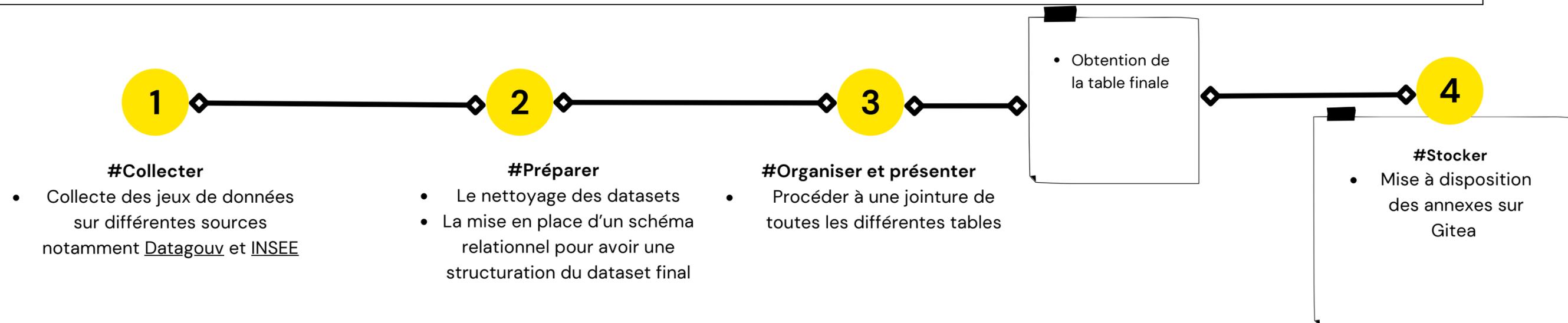
Étape 2 : Démarche et organisation de la base de données



La récolte de données RSE sociétales est souvent liée **à la recherche d'indicateurs paritaux** ou encore à ceux affiliés à des indicateurs comme l'égalité, à une géolocalisation des structures ou encore à des données liées à des critères beaucoup plus spécifiques.

Ainsi dans le cadre de la mise en place de nos jeux de données, certains jeux de données ont été jugés pertinents pour nos recherches pour aboutir à l'établissement d'un jeu de données final.

Ainsi plusieurs étapes successives sont notées dans ce processus



SOURCES

DATA GOUV



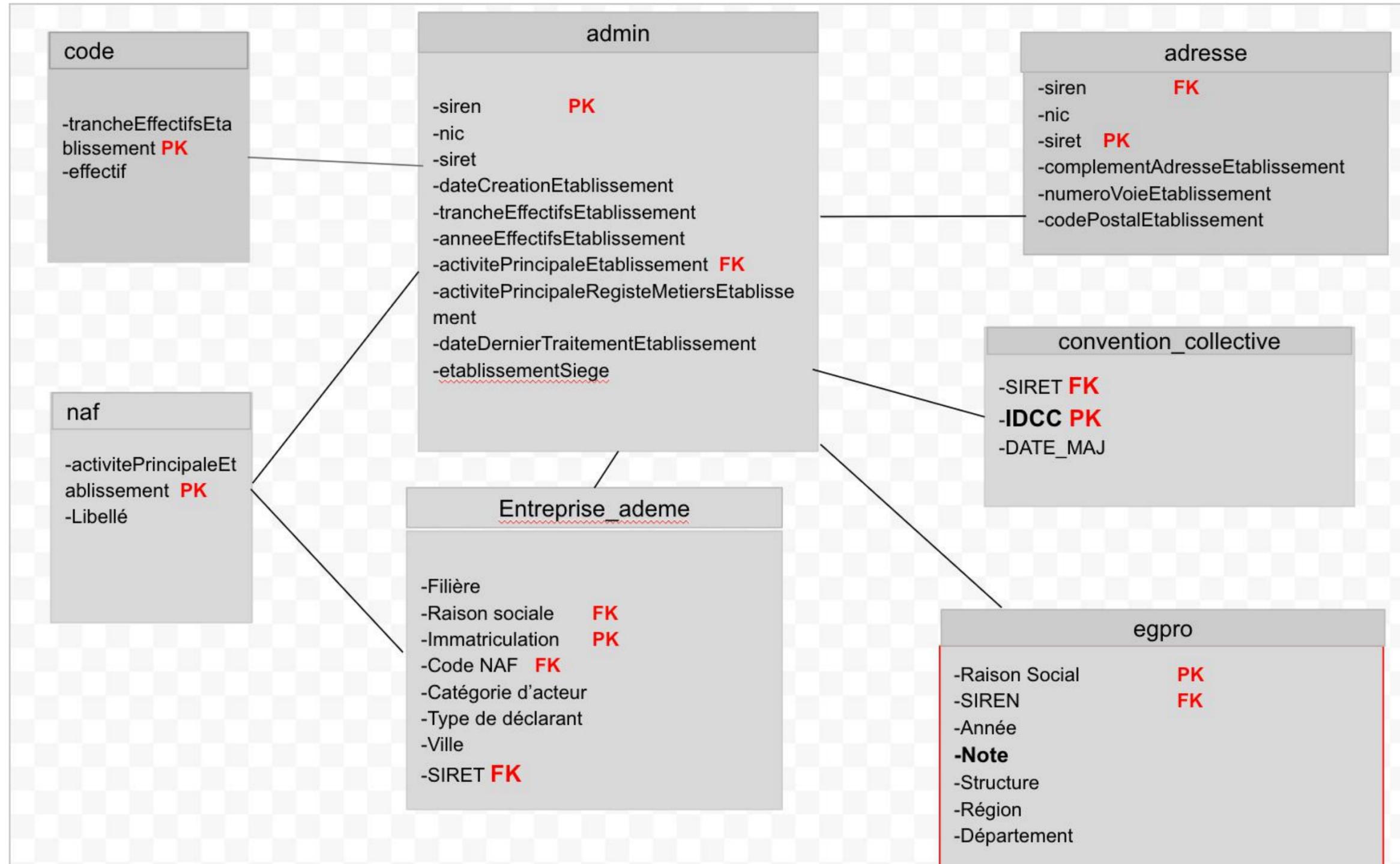
Plateforme de diffusion de données
publiques de l'État français

INSEE



L'Institut national de la statistique et des
études économiques est chargé de la
production, de l'analyse et de la publication
des statistiques officielles en France

Étape 3 : Description des variables et des tables



Analyse des données



27,4 %

Femmes



8,4 %

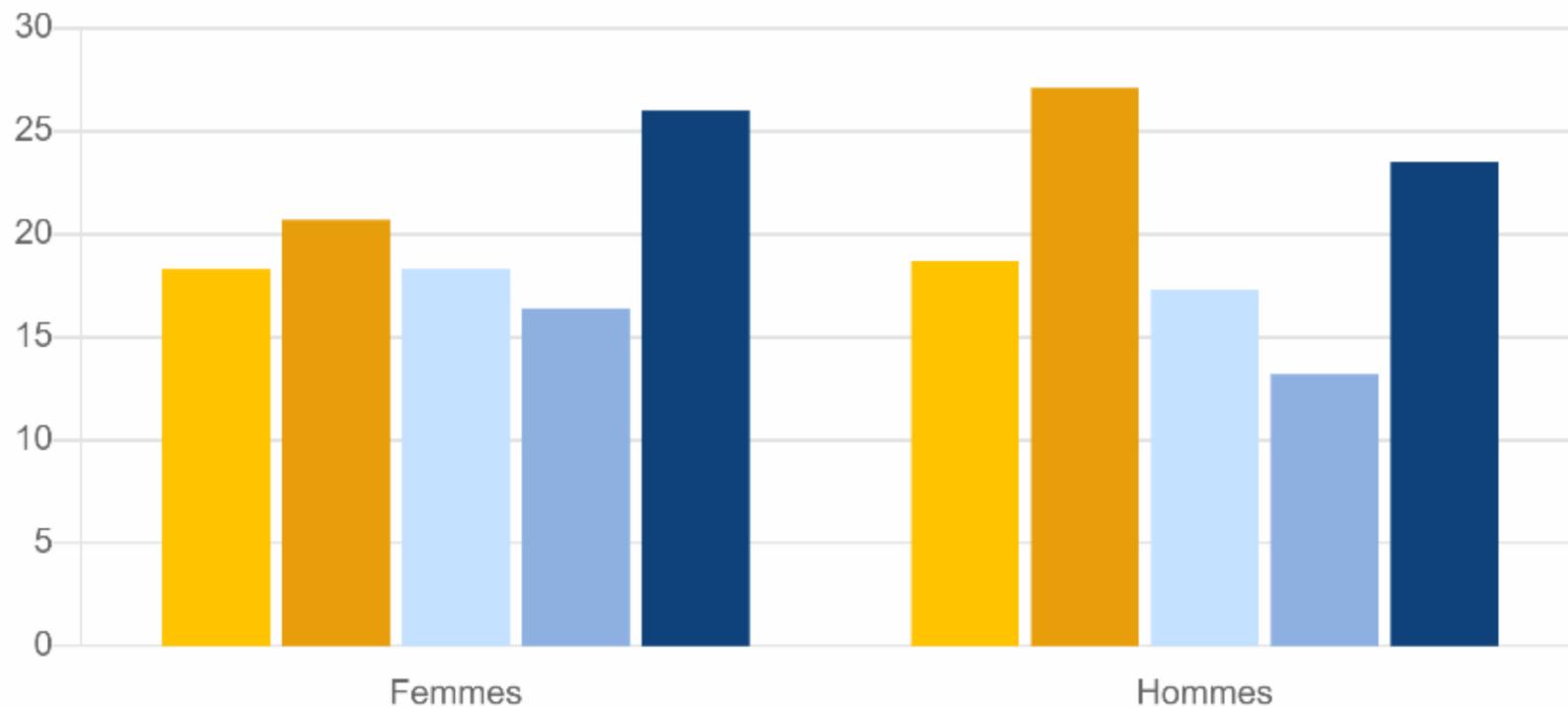
Hommes

Diplôme le plus élevé selon le sexe en 2020



Aucun diplôme, brevet des collèges CAP, BEP ou équivalent
Baccalauréat ou équivalent Bac + 2 Bac + 3 ou plus

en %

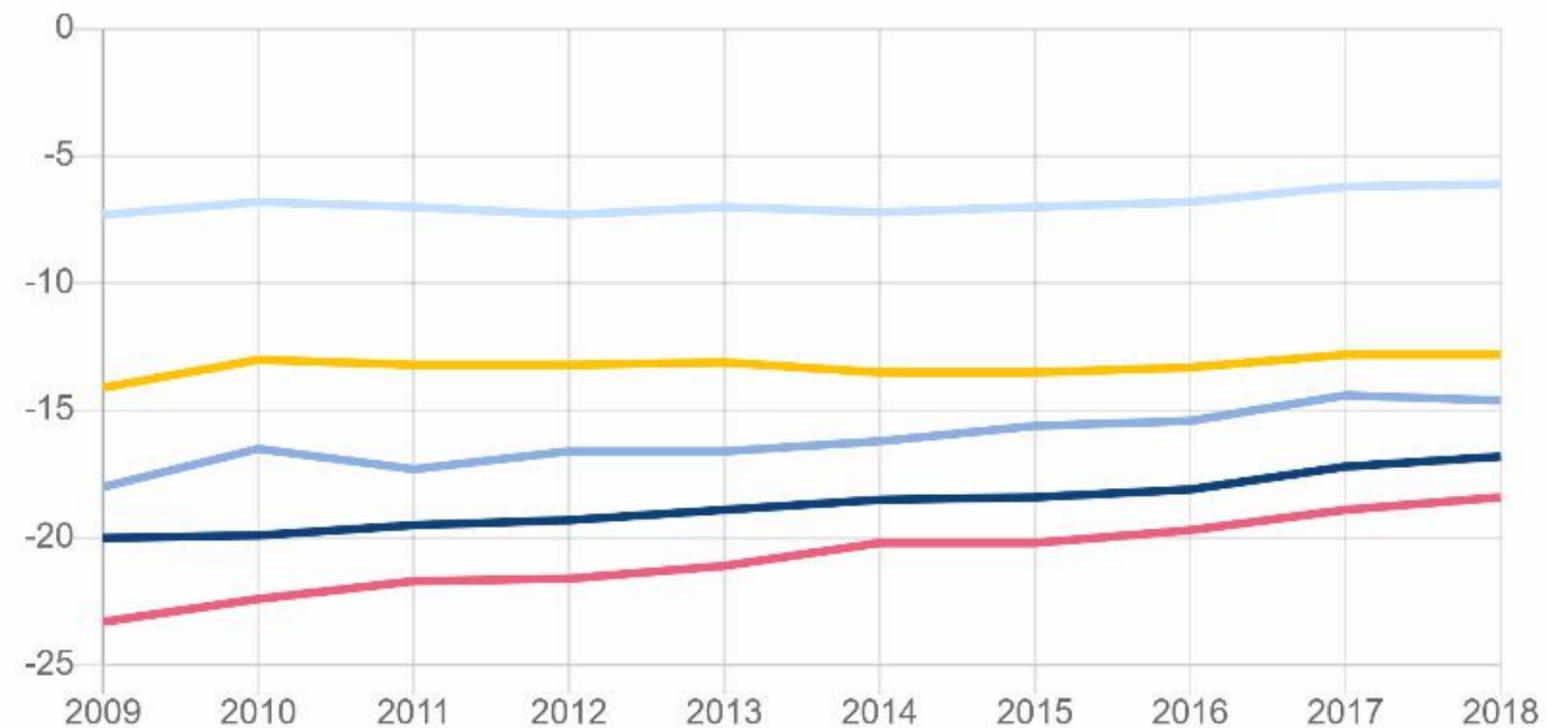


Écart de salaires en équivalent temps plein entre femmes et hommes



Cadres¹ Professions intermédiaires Employés Ouvriers Ensemble

en %



Nos Indicateurs RSE Sociétaux

Note (egpro)

L'Index de l'égalité professionnelle a été conçu pour faire progresser au sein des entreprises l'égalité salariale entre les femmes et les hommes.

Il permet aux entreprises de mesurer, les écarts de rémunération entre les sexes et de mettre en évidence leurs points de progression. Lorsque des disparités salariales sont constatées, des mesures de correction doivent être prises.

L'index est calculé à partir de 5 indicateurs portant sur l'écart des rémunérations, promotions et augmentations, mais également sur le congé maternité et les plus hauts salaires de l'entreprise.

IDCC

Le colonne IDCC représente le numéro de convention collective

S'affilier aux conventions collectives offre aux salarié et les parties prenantes de l'entrepris des meilleurs conditions sociales de travail

Une entreprise qui a un numero IDCC, est une entreprise plus responsable

REP

Ademe est l'agence de la transition écologique,

Elle enregistre des producteurs qui sont affilié au dispositif REP(responsabilité élargie du producteur)

Ce dispositif permette d'évaluer l'entreprise sur son organisation de la prévention et la gestion des déchets.

L'objectif est de traiter les déchets produits, mais aussi de les prévenir en agissant sur l'ensemble du cycle de vie des produits.



ETAPE 4 : Exploitation et usages

TE_SIREN_ADMIN

Cette table nous permet l'identification des entreprises et des établissements

ENTREPRISE_ADEME (REP-EMB)

Ensuite nous évaluons la capacité à gérer correctement les résidus de fabrication de produits (pour les entreprises productrices) et lister les entreprises responsables

01

02

03

04

EGPRO

Une fois l'entreprise identifiée nous mesurons son Index de l'égalité professionnelle

L'UTILITE

- Alerter l'entreprise sur son fonctionnement
- Mettre en évidence les points de progression sur lesquels agir
- La finalité est qu'à partir de nos extractions qu'on puisse avertir une entreprise sur sa façon de fonctionner et qu'elle puisse prendre des actions afin d'éviter de mauvaises notations

Réussi

- Bonne organisation de l'équipe (bonne répartition des tâches)
- Finalisation du schéma relationnel
- Manipulation de données sur Python

Difficultés

- Volume de données
- Trouver des jeux de données facile à intégrer
- Temps limité
- Importer les données sur système de gestion de BDD (mysql, postgres)

Roadmap du projet

The screenshot displays a Trello project board for "SCRUM Project" in a Kanban view. The board is organized into five columns representing different stages of the project:

- Seance 1:** Contains six task cards: "Trouver les bases de données pertinentes", "Définition des rôles", "Carnet du produit", "Scrum meeting", "Outil de communication", and "Outil de developpement".
- Sprint 1:** Contains four task cards: "Définition des indicateurs RSE", "Difficultés rencontrées", "Collection des données", and "Recherche de jeux de données".
- Sprint 2:** Contains three task cards: "Elaboration du modele de la base de données", "Bien cibler les BDD", and "Difficultés rencontrées".
- Sprint 3:** Contains four task cards: "Nettoyage de la BDD", "Définir les jointures", "Créer une architecture sous forme graphique", and "Préparation du sprint review".
- JOUR J : Présentation:** Contains three task cards: "Préparer la présentation", "Avancement sur le data paper", and "Répartition des parties".

Each task card includes a progress bar at the top, indicating the completion status of the task. The board interface includes a top navigation bar with "Trello" logo, workspace name "SCRUM Project", and various utility buttons like "Inviter", "Automatisation", and "Filtre".

AXES D'AMELIORATION

REDUIRE LA TAILLE DU
FICHIER

Avoir un fichier
moins volumineux
afin de permettre
une meilleure
exploitation

Avoir un code qui
permettrait une
rapide excécution
du script

OPTIMISER LE SCRIPT
PYTHON

Afin de permettre une
meilleure utilisation

OPTIMISER LA
COLLECTE ET LA MISE
A JOUR

Merci !

N'hésitez pas à nous contacter
si vous avez des questions.

