# EL BALI ABDERRAHMANE / ADAM MOUSSAOUI

# TD3

# *SAE 1-02 ECRITURE ET LECTURE DE FICHIERS DE DONNEES*

# Cahier des Charges

## Objectif de la SAE

Le but de cette SAE est d'analyser les données publiques de la SNCF pour comprendre l'évolution des métiers, des effectifs et des recrutements depuis 2012. On s'intéresse surtout à la répartition hommes/femmes et à son évolution. Pour cela, on va créer un programme en Python qui va automatiser l'analyse et rendre les résultats plus faciles à exploiter.

## 1) Rappel des besoins

1- Analyser les effectifs et les recrutements

- Utiliser les fichiers recrutement-metiers-sncf.csv et effectif-metiers-sncf.csv pour comprendre les évolutions depuis 2012.

2- Comparer les résultats :

- Vérifier les analyses avec les données de répartition hommes/femmes dans repartition-genre-effectif.csv.

3- Automatiser le traitement :

- Charger, transformer et analyser les données.

- Calculer des statistiques comme l'évolution annuelle des effectifs ou des recrutements.

- Exporter les résultats dans des fichiers simples à partager (Excel, CSV).

## 2) Moyens

1- Outils disponibles:

- Python avec des bibliothèques pour traiter les données et créer des graphiques.

- Des logiciels comme Excel ou Word pour préparer et présenter les résultats.

2- Ressources:

- Les fichiers recrutement-metiers-sncf.csv, effectif-metiers-sncf.csv et repartition-genre-effectif.csv.

- Documentation Python et informations sur les données fournies par la SNCF.

3- Ressources Humaine

- Les Professeurs (Mme Belkouch : Responsable de la Sae *1-02*)

## 3) Contraintes

1- Techniques:

- Gérer des fichiers de grande taille.

- Adapter le code pour différents formats de données.

2- Temporelles:

- Le Temps

3- Qualité des données:

- Identifier et corriger les erreurs ou les valeurs manquantes.

## 4) Description des Données

1- Recrutements:

- Fichier recrutement-metiers-sncf.csv contenant les chiffres des recrutements depuis 2012.

2- Effectifs:

- Fichier effectif-metiers-sncf.csv donnant les effectifs réels par métier depuis 2012.

3- Répartition hommes/femmes :

Fichier repartition-genre-effectif.csv fournissant les statistiques par année et par type de contrat.

Travail 2 :

a. **Les anomalies relevées :**

* + Des données manquantes ont été identifiées dans certaines années ou pour certains métiers, ce qui peut limiter les analyses.
  + Aucune valeur aberrante, comme des effectifs ou recrutements négatifs, n’a été détectée.

b. **La cohérence des données :**

* + Les métiers ne sont pas totalement cohérents d’une année à l’autre. Certains intitulés sont regroupés ou changés au fil des années, ce qui demande un travail de normalisation pour les comparer.

c. **L’écart de taille entre les fichiers :**

* + Le fichier des effectifs est plus volumineux (environ 20 000 lignes) que celui des recrutements (10 000 lignes). Cela s’explique par la présence de données pour tous les employés et pas seulement pour les nouveaux recrutés.

d. **L’analyse : peut-on comparer les données de deux années ?**

* + Oui, les données sont comparables grâce aux colonnes "Année" et "Métier", bien que des ajustements soient parfois nécessaires pour harmoniser les intitulés de métiers.

e. **Rapprochement entre les fichiers :**

* + Les colonnes communes comme "Année", "Métier", "Type de contrat", et "Sexe" permettent de croiser les informations des différents fichiers pour réaliser une analyse globale.
  + Par exemple, il est possible de calculer le taux de renouvellement des effectifs ou d’analyser l’évolution de la proportion hommes/femmes par métier.

## 5) Organisation

### *Structure des dossiers*

1- Dossier principal : SAE\_SNCF

2- Sous-dossiers :

- APPLICATION : pour les données (à mettre dans DATA) et les résultats (à mettre dans EXPORT).

- CONCEPTION : pour les fichiers de suivi comme rapport\_activites.xlsx, cahier\_des\_charges\_et\_conception.docx et specifications.xlsx.

### *Mise en place initiale*

- Les fichiers de données doivent être stockés dans le dossier DATA.

- Les résultats doivent être organisés dans EXPORT avec des sous-dossiers pour les extractions et les statistiques.

## 6) Conception de l'Application

### *Fonctionnalités principales*

1- Charger les données :

- Lire les fichiers CSV et s'assurer qu'ils sont corrects.

2- Analyser les données :

- Filtrer les données selon différents critères (année, métier, type de contrat, genre).

- Calculer des statistiques comme les moyennes ou les évolutions annuelles.

3- Visualiser les résultats :

- Créer des graphiques pour mieux comprendre les tendances.

4- Exporter les résultats :

- Sauvegarder les analyses dans des fichiers Excel ou CSV.

## 7) Interface Utilisateur

Menu simple en ligne de commande permettant :

- De choisir les fichiers à analyser.

- De sélectionner les critères de filtrage.

- De lancer l'analyse et de voir les résultats.

## 8) Livrables

Livrables intermédiaires :

- Rapport sur les tâches accomplies.

- Cahier des charges complété.

- Fichier des spécifications techniques.

Livrable final :

- Dossier compressé SAE\_SNCF.zip contenant tout le travail, à rendre avant le 13 janvier 2025

****Analyse des données**:**

**A. Les anomalies relevées :**

* + Des données manquantes ont été identifiées dans certaines années ou pour certains métiers.
  + Aucune valeur aberrante, comme des effectifs ou recrutements négatifs, n’a été détectée.

**B. La cohérence des données :**

* + Les métiers listés ne sont pas totalement cohérents d’une année à l’autre. Certains intitulés ont été regroupés ou renommés.
  + Exemple : un métier peut apparaître sous deux noms différents selon les années, comme "Agent commercial" et "Agent de bord".

**C. L’écart de taille entre les fichiers :**

* + Le fichier des effectifs est nettement plus volumineux (environ 20 000 lignes) que celui des recrutements (10 000 lignes).
  + Cette différence s’explique par le fait que les effectifs incluent tous les employés, alors que les recrutements se concentrent uniquement sur les nouveaux entrants.

**D. L’analyse : peut-on comparer les données de deux années ?**

* + Oui, les données sont comparables grâce aux colonnes communes comme "Année" et "Métier".
  + Cependant, certaines années peuvent nécessiter des ajustements pour harmoniser les intitulés ou pour gérer les données manquantes.
  + Exemple : comparer les recrutements de 2012 et de 2020 permettrait d’identifier des tendances ou des pics d’embauche.

**E. Rapprochement entre les fichiers :**

* + Les colonnes communes ("Année", "Métier", "Type de contrat", "Sexe") permettent de croiser les informations entre les fichiers pour une analyse complète.
  + Par exemple, il est possible de calculer :
    - Le taux de renouvellement des effectifs pour un métier donné.
    - L’évolution de la proportion hommes/femmes par année et par contrat.
  + Ces rapprochements permettent d’obtenir des indicateurs clés pour mieux comprendre les évolutions au sein de la SNCF.