
Cahier des charges

Objectif de la SAE

Les Situations d'Apprentissage et d'Évaluation (SAE), intégrées dans les programmes des Instituts Universitaires de Technologie (IUT), ont pour objectif de favoriser un apprentissage pratique et contextualisé tout en permettant une évaluation continue des compétences des étudiants. De manière générale, elles permettent :

- De favoriser l'apprentissage par la pratique
- De développer des compétences (soft skills)
- De renforcer l'autonomie et l'initiative
- De préparer les étudiants à l'insertion professionnelle

Dans le cadre de la programmation en Python, la SAE intitulée "**Écriture et lecture de fichiers de données**" a pour objectif de concevoir et de développer un programme en Python permettant de traiter un flux de données selon plusieurs étapes : la lecture, la transformation, et l'écriture des données, en respectant les spécifications fournies. Ces données seront traitées dans un but décisionnel, c'est-à-dire qu'elles devront permettre de générer des informations pertinentes pour l'analyse et la prise de décisions.

À cet effet, nous allons travailler sur les données ouvertes de la SNCF concernant les effectifs et les recrutements depuis 2012.

Rappel des besoins :

Pour y parvenir, nous devons être capables de concevoir un programme qui permette de manipuler ces données, il faudra donc utiliser le langage Python pour :

- Lire des fichiers de données (CSV) contenant des informations sur les métiers, les effectifs et les recrutements de la SNCF depuis 2012.
- Effectuer des transformations sur les données (nettoyage, filtrage, agrégation, etc.).
- Analyser les données afin de répondre à des questions décisionnelles liées aux effectifs, aux évolutions des métiers et des recrutements dans l'entreprise.
- Écrire les résultats dans un format de fichier adapté (CSV).
- Fournir des outils d'analyse permettant d'interroger les résultats et de les utiliser pour des prises de décisions.

Moyens nécessaires pour atteindre notre objectif :

Moyens logiciels

1. Langage et outils de développement :

- Python : principal langage pour développer le programme.
- Excel : pour la création du dictionnaire de données et la rédaction des rapports.
- Word : pour rédiger le cahier des charges et documenter les étapes.

2. Bibliothèques Python :

- random : pour des fonctions liées à la génération aléatoire (si besoin).
- csv : pour lire et écrire les fichiers CSV.
- os : pour la gestion des fichiers et des répertoires.

Moyens matériels

1. Ordinateurs performants :

- Ordinateurs avec une configuration suffisante pour exécuter Python et travailler sur des ensembles de données volumineux.

2. Accès aux données ouvertes :

- Les fichiers CSV des effectifs et recrutements de la SNCF, téléchargeables depuis des sources ouvertes ou fournis par le projet.

3. Connexion internet :

- Obligatoire pour accéder aux données, consulter la documentation en ligne et collaborer en temps réel.

4. Périphériques complémentaires :

- Écrans supplémentaires : pour travailler confortablement sur plusieurs fichiers ou outils.
- Claviers et souris ergonomiques : pour améliorer la productivité et réduire la fatigue.

Moyens humains

1. Pairs :

- Le travail en binôme ou en groupe permettra de répartir les tâches et d'avancer plus rapidement.

2. Encadrement pédagogique :

- Les enseignants ou tuteurs disponibles pour fournir des conseils, valider les étapes du projet, et corriger les erreurs et nous aider à déboguer nos programmes

Contraintes

▪ Techniques :

- ❖ Les fichiers doivent être manipulés avec des outils compatibles comme Python et Excel.
- ❖ Les scripts doivent être :
 - Documentés de manière claire et réutilisables.
 - Organisés selon une structure de dossiers définie.
- ❖ Une documentation claire doit être fournie à chaque étape.
- **Temps :**
 - ❖ Respect des délais imposés par le planning pédagogique.

Données :

1. Description des fichiers fournis :

- recrutement-metiers-sncf.csv : Contient les données sur les recrutements de la SNCF depuis 2012 (date, métier, type de contrat, genre).
- effectif-metiers-sncf.csv : Fournit les informations sur les effectifs en poste à la SNCF depuis 2012 (date, métier, type de contrat, genre, effectifs).

2. Dictionnaire de données :

- Un fichier Excel détaillant les variables des fichiers CSV (nom, type, description, modalités, etc.).

Organisation :

✓ Création de l'arborescence des dossiers :

SAE_SNCF/

└─ APPLICATION/

| └─ DATA/

| | └─ recrutement-metiers-sncf.csv

| | └─ effectif-metiers-sncf.csv

| | └─ dictionnaire_donnees.xlsx

| └─ EXPORT/

| | └─ EXTRACTIONS/

| | └─ STATISTIQUES/

| └─ sae.py

└─ CONCEPTION/

└─ rapport_activites.xlsx

└─ cahier_des_charges_et_conception.docx

└─ specifications.xlsx

- ✓ Écrire des fonctions pour charger les fichiers CSV. et afficher les données dans l'interface de Spyder .

Description des Données

a. Anomalies relevées

➤ Données manquantes :

Dans le fichier *Effectifs*, une valeur est manquante dans les colonnes Date et Métier (ligne 91 du fichier excel)

Le fichier recrutement *ne contient pas d'anomalie de logiciel ou de logique, mais après analyse de celui on se rend compte que pour chaque année le nombre de métier est différent :*

Par exemple en 2012 on a 5 métiers disponibles sauf le métier exécution tandis qu'en 2018 on a les 6 métiers disponibles (don exécution).

➤ Données aberrantes :

Nous n'avons constaté aucune donnée aberrante dans les deux fichiers excel, étant donné que les fichiers sont ouverts au public et distribué par la sncf directement, ça nous paraissait logique que des vérifications ont eu lieu avant la publication de ceux-ci.

b. Cohérence des données

- Une comparaison des colonnes Métiers (Recrutement) et Métier (Effectifs) montre des similitudes avec des métiers communs tels que :
 - *Agent commercial à bord des trains,*
 - *Conducteurs,*
 - *Exécution,*
 - *Ingénieurs et cadres.*
 - *Mainteneur, Aiguilleur, Agent commercial, gestionnaire.*

Cependant il y'a un métier qui diffère des deux:

- ☒ (Sélectionner tout)
- ☒ Agent commercial à bord des trains
- ☒ Conducteurs
- ☒ Exécution
- ☒ Ingénieurs et cadres
- ☒ Mainteneur, Aiguilleur, Agent comme
- ☒ Manager de proximité, maîtrise
- ☒ (Vides)

-Effectif métier : Manager de proximité, maîtrise.

☒ (Sélectionner tout)
☒ Agent commercial à bord des trains
☒ Conducteurs
☒ Exécution
☒ Ingénieurs et cadres
☒ Mainteneur, Aiguilleur, Agent comm
☒ Manager de proximité
☒ (Vides)

-Recrutement métier : Manager de proximité.

c. Écart de taille entre les fichiers

Le fichier *Recrutement* contient 212 lignes, tandis que le fichier *Effectifs* en contient 224, On peut donc dire que c'est un écart faible soit de 12 lignes.

d. Analyse : Peut-on comparer les données de deux années ?

Oui, mais certaines conditions doivent être respectées :

- Pour cela on doit faire la mise en commun des noms de colonnes et des métiers entre les fichiers.
- Vérification des dates pour s'assurer que les années correspondent exactement.
- Faire attention aux métiers exclusifs dans les fichiers, comme dit précédemment il y'a 1 métier qui n'apparaît pas dans le fichier effectif mais qui est présent dans le fichier recrutement. (voir le b)

e. rapprochement entre les différents fichiers

On peut faire le rapprochement avec les fichiers, grâce à ces colonnes :

1. **Date** : Les deux fichiers ont une colonne date.
2. **Métiers/Métier**
3. **Contrat de travail** : cette colonne est utile pour connaître le type du contrat des salariés.

Donc en soit, on peut notamment faire un tableau croisé dynamique avec les deux fichiers

Traitement des données ouvertes de la SNCF : Répartition par genre des effectifs et recrutements depuis 2012(Quelques comparaisons avec le fichier effectifs de la SNCF)

Observation

2013

On observe que, pour l'année 2013, le nombre de cadres permanents hommes dans notre tableau d'analyse correspond uniquement aux nombres d'agents au statut figurant dans la répartition "genre-effectifs" fournie par la SNCF. Cependant, pour les contractuels incluent les agents à temps complet et à temps partiel, car la somme de leurs effectifs (hommes et femmes) correspond à celle obtenue dans notre tableau d'analyse.

2015

En 2015, une anomalie est constatée. L'effectif des hommes ayant un contrat cadre permanent dans notre tableau d'analyse est légèrement inférieur à celui fourni par la SNCF, soit une différence de 270 agents (110 377 contre 110 107). Cependant, pour les contractuels, les données restent cohérentes et correspondent à la somme des effectifs des agents à temps complet et à temps partiel.

2017

En 2017, aucune correspondance n'est observée entre les données fournies par la SNCF et celles de notre tableau comparatif. Les effectifs d'hommes et de femmes, quel que soit le type de contrat, dans le tableau indicatif, ne correspondent pas aux données transmises par la SNCF.