

Bilan personnel

Teodora MOLDOVAN

1. Contexte du projet

Le projet de traitement des données avec Python présente un double objectif :

- se familiariser et mettre en pratique les compétences acquises en manipulation de données avec des paquets Python comme numpy, pandas, matplotlib, scikit-learn ;
- extraire des statistiques à partir des données brutes.

Au début du projet, j'avais un niveau débutant en Python, n'ayant pas eu l'occasion d'en pratiquer avant l'ENSAI. J'étais motivée car je souhaite apprendre ce langage de programmation et j'estime qu'il sera utile dans ma future carrière à l'Insee. Mon objectif principal était d'atteindre les résultats attendus du projet et de me familiariser avec Python en approfondissant mes compétences au fil du travail. J'étais aussi un peu appréhensive face à la complexité potentielle des méthodes à utiliser.

2. Déroulement du projet :

Nous avons travaillé en équipe sur les données des Jeux Olympiques, présentant l'historique de résultats sur une période de 120 ans.

Sur le plan de l'organisation, j'ai d'abord travaillé pendant les créneaux dédiés, que j'ai particulièrement appréciés car ils m'ont permis d'avancer dans un cadre structuré. Par la suite, nous avons travaillé toujours ensemble lors de créneaux supplémentaires, avant de continuer chacun de notre côté selon les disponibilités de chacun. Nous avons travaillé plutôt en autonomie, chacun sur ses questions, et plus en collaboration sur la partie *apprentissage automatique*.

Côté programmation, une première étape dans l'analyse a été de bien comprendre le jeu de données et d'identifier certains points de vigilance, ce qui a facilité les étapes suivantes (par exemple dans le comptage des médailles bien considérer les cas ex-aequo et les épreuves collectives). J'ai notamment apprécié d'utiliser la bibliothèque pandas, que j'ai trouvée très efficace pour explorer et manipuler les données et plus facile à comprendre (le nommage est logique).

Ce projet m'a aussi permis de mieux comprendre la logique de Python : j'ai réalisé que, dans certains cas, il n'était pas nécessaire de définir une fonction complexe – des instructions simples comme des filtres, des boucles for ou des conditions if suffisent parfois à obtenir le résultat souhaité (exemple la 3^e question que j'ai traitée : « *Quels ont été les résultats spécifiques de Nadia Comăneci lors des JO de 1976* »). Le seul point d'attention ici a été d'effectuer une

recherche partielle et insensible à la casse, car les caractères spéciaux ont été supprimés dans la base de données.

3. Bilan final

J'estime avoir progressé au cours du projet, mais la progression est restée en deçà de mes attentes initiales. Je ne suis pas encore autonome pour programmer en Python : j'ai souvent dû me référer à mes cours, aux travaux dirigés, à des programmes réalisés par mes collègues, ou encore à des outils comme l'intelligence artificielle pour écrire certaines instructions, en particulier les fonctions complexes et les tests.

Les créneaux intégrés à l'emploi du temps ont été utiles et m'ont permis d'avancer dans un cadre guidé, mais ils se sont révélés insuffisants au vu de la charge réelle du projet. J'ai notamment sous-estimé le temps nécessaire pour la partie d'apprentissage automatique, qui m'a demandé un investissement conséquent, tout comme la rédaction du rapport final.

Je suis consciente que certains de mes programmes pourraient être simplifiés, mais je manque encore de recul, de compétences techniques en Python, et peut-être aussi d'esprit d'analyse et de synthèse pour les optimiser efficacement. L'automatisation de certaines étapes a également été difficile, je retiens la nécessité pour moi de progresser en programmation fonctionnelle.

Pour mes futurs projets, je pense qu'il sera essentiel d'améliorer mon organisation personnelle : mieux estimer le temps nécessaire pour les tâches complexes, prendre le temps de planifier les étapes, et surtout pratiquer davantage en dehors des cours pour renforcer ma maîtrise de Python.

Avec le recul, je pense que j'aurais dû noter plus systématiquement les difficultés rencontrées tout au long du projet. J'en ai probablement oublié certaines au moment de rédiger ce bilan, alors qu'elles auraient pu m'apporter des enseignements utiles. Je pense en particulier au traitement de la première question (« *Combien d'athlètes ont remporté des médailles aux Jeux olympiques d'hiver et d'été ?* »). J'ai commencé par écrire un programme assez simple mais qui n'affichait pas le résultat attendu. Je suis passée ensuite par plusieurs étapes de correction pour obtenir le même résultat que celui affiché par le *Comité International Olympique*. Cela fait partie des bonnes pratiques que je compte adopter à l'avenir, pour mieux capitaliser sur mes expériences.