

## Documents Données 2 - IA1 - Projet:

Abraham - Fabien - Vohizy - Mirado

STD21017 - STD21101 - STD21005 - STD21081

### Contexte :

Nous voulons exploiter un API de football pour mieux peaufiner nos recherches sur les potentielles futurs talents du monde du football afin de simplifier le processus de recrutement de joueurs.

Pour cela, notre étude se tournera sur l'API suivante :

[https://apifootball.com/documentation/?gclid=CjwKCAjw8symBhAqEiwAaTA\\_FHOpU1tvtkPyd9Kq2CkYNy4z1voKLIXDjwRvuOmvwmhqVebHnzE-RoC6ngQAvD\\_BwE](https://apifootball.com/documentation/?gclid=CjwKCAjw8symBhAqEiwAaTA_FHOpU1tvtkPyd9Kq2CkYNy4z1voKLIXDjwRvuOmvwmhqVebHnzE-RoC6ngQAvD_BwE)

### Technologies :

Quant aux choix technologiques, après une longue discussion nous nous sommes mis d'accord sur les outils et les langages que nous allons utiliser tout au long du projet. A cet effet, le langage de script sera principalement Python. L'implémentation du tableau de bord utilisé Tableau pour la visualisation des données et utilise enfin le service S3 d'Amazon Web Service pour le stockage.

### Tâches:

La répartition des tâches dans notre projet s'est principalement faite par le biais de github : plus précisément avec l'outil Project de notre organization composé de : **RAMAROSANDY Abraham Prince** (STD 21017), **RAHARINAIIVOSOA Radintsoa Mirado** (STD 21081), **ANDRIATSIMIADY Ramiandrisoa Vohizy** (STD 21005) et **MALALA-ZO Raharison Fabien** (STD 21101) ; Afin de simplifier votre compréhension de notre répartition de tâches, voici un lien amenant à notre projet :

<https://github.com/orgs/G6-Data-Project/projects/1/views/1>

On a tous autant participer au projet mais certaines responsabilités ce sont quand même départager vis-à-vis des compétences de chacun : **Vohizy** (OPS), **Fabien** (PO), **Abraham** (Full-Stack sur chatgpt) et **Mirado** (Back)

### Processus:

#### **1. Extraction des données de l'API :**

A partir de l'API football, il faut créer un code qui puisse extraire en .csv les informations dont on a besoin.

Elle consiste concrètement à faire un appel de la liste de toutes les ligues qui incluent la liste de toutes les teams et des players de chaque ligues.

## **2. Transformation des données obtenus :**

Tout d'abord, nous devons soumettre les données à un filtre global pour identifier les caractéristiques générales recherchées chez un joueur professionnel générique. Ensuite, nous appliquons un deuxième filtre pour extraire la liste des meilleurs joueurs pour chaque poste spécifique. Après cela, nous calculons un score en fonction des statistiques précédemment filtrées, afin de créer un classement des joueurs. Ce classement nous permettra finalement d'obtenir la liste des meilleurs joueurs pour chaque poste, en tenant compte de leurs évaluations spécifiques et de leurs performances. Après les transformations précédentes, on obtient un nouveau fichier.csv

## **3. Orchestration via Apache Airflow**

La manœuvre consiste à créer une DAG qui automatise les étapes précédentes (extraction des données de l'API et transformations des données obtenus) tous les 3 mois. La logique dans la DAG sera la suivante : faire les imports requis, la configuration de la DAG (nom,intervalle de temps d'exécution de la DAG), déterminer en tâches l'extraction et les transformations précédentes et la connexion à la S3,donner l'ordre d'exécution des tâches données.

### **Documentations:**

- <https://awstip.com/how-to-use-apache-airflow-to-upload-data-to-an-s3-bucket-on-aws-529643695867>
- <https://medium.com/international-school-of-ai-data-science/setting-up-a-pache-airflow-in-windows-using-wsl-dbb910c04ae0>

### **Conclusion:**

En conclusion, le projet IA1, porté par l'équipe composée d'Abraham, Fabien, Vohizy et Mirado, vise à exploiter les opportunités offertes par l'API de football pour affiner la recherche de futurs talents dans le monde du football. L'objectif ultime est de simplifier le processus de recrutement de joueurs en identifiant de manière efficace les talents émergents.

Grâce à l'utilisation de la technologie Python pour la création de scripts, de Tableau pour la visualisation des données, et du service Amazon S3 d'Amazon Web Services pour le stockage, notre approche est résolument moderne et orientée vers la performance.

Le projet a été organisé avec soin à l'aide de GitHub, et chaque membre de l'équipe a apporté ses compétences pour contribuer au succès collectif. Les responsabilités ont été attribuées en fonction des forces individuelles de chacun : Vohizy pour les opérations, Fabien pour la gestion du produit, Abraham en tant que Full-Stack, et Mirado en charge du développement backend.

Le processus du projet s'articule autour de deux étapes majeures. Tout d'abord, l'extraction des données de l'API football pour obtenir une source de données stable et exploitable, qui est ensuite convertie en fichier .csv. Ensuite, ces données sont soumises à une série de filtres successifs, permettant d'identifier les caractéristiques générales recherchées chez un joueur professionnel, d'extraire les meilleurs joueurs pour chaque poste spécifique, et de calculer un classement basé sur des scores issus des statistiques filtrées.

L'automatisation du processus est assurée par une DAG via Apache Airflow, qui orchestre l'extraction et la transformation des données à intervalles réguliers de trois mois. Cette approche permet de garantir que les informations restent à jour et pertinentes au fil du temps.

En somme, le projet IA1 représente un engagement dynamique et collaboratif pour exploiter pleinement les technologies modernes et l'expertise de chaque membre de l'équipe, avec pour objectif de révolutionner le processus de repérage de futurs talents dans le monde du football.