

# Projet Data & IA - JO

De l'historique des Jeux à une app predictive interactive

DUNCAN

YANNIS

ABDELLATIF

ANYA

DUNEL



# Sommaire

- |    |                              |    |                           |
|----|------------------------------|----|---------------------------|
| 01 | Problématique & Objectifs    | 05 | Modèles utilisés          |
| 02 | Méthodologie & Plan d'Action | 06 | Résultats & Démonstration |
| 03 | Stack Tehcnique              | 07 | Synthèse du Projet        |
| 04 | Données                      | 08 | Conclusion                |



# Pourquoi analyser les Jeux Olympiques ?



## Contexte

- Plus de 120 ans d'histoire olympique
- Des milliers d'athlètes, d'épreuves et de médailles
- Un volume de données complexe à analyser sans outils adaptés



## Problématique

Comment exploiter efficacement les données olympiques pour :

- Comprendre les performances passées
- Analyser les tendances
- Et prédire les résultats futurs ?



# Objectifs du Projet



01

Structurer les données  
dans une base fiable

02

Explorer et visualiser les  
performances olympiques

03

Concevoir des modèles de  
prédition

04

Prédire les résultats des  
JO de Paris 2024

05

Proposer une App  
interactive

# Comment ?

1

## PROBLEME

Compréhension du problème et des questions à adresser.

2

## COMPRÉHENSION

Construction de la base de données olympique.

3

## ANALYSE

Exploration / dataviz pour comprendre les patterns.

4

## CONSTRUCTION

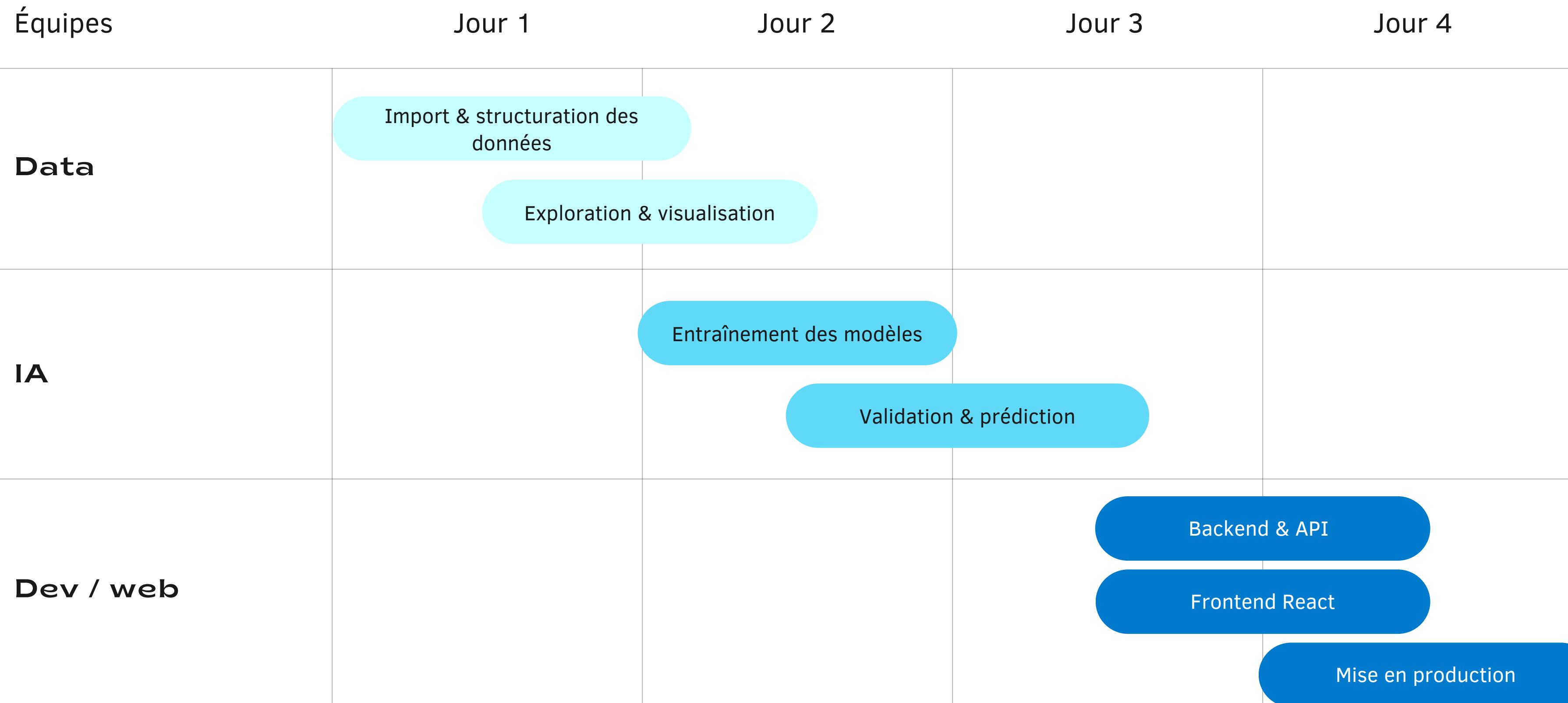
Construction et évaluation de 2 modèles de prédiction.

5

## INTÉGRATION

Intégration des modèles dans une app.

# Plan d'Action

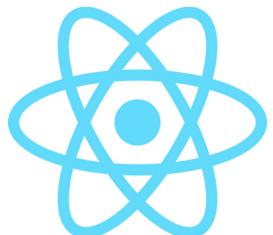


# Stack Technique

## Frontend

React

- Visualisations des analyses
- Exploration par pays / sports
- Affichage des prédictions générales et détaillées



## Backend

API REST (Node.js / Express)

- Requêtes vers la BDD



## Modèle IA

Scikit learn / pandas

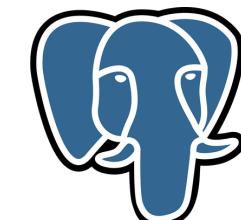
- Préparation des données
- Modèles de prédiction
- Évaluation des performances



## BDD

PostgreSQL

- Schéma relationnel structuré
- Contraintes & intégrité des données



# Exploration et préparation des données



**MÉDAILLES**  
~21 000



**RÉSULTATS**  
~162 000



**ATHLÈTES**  
~74 000



**PAYS HÔTES**  
52



## Préparation des données

- Suppression des doublons
- Gestion des valeurs manquantes
- Harmonisation des sports / pays
- Création de variables pertinentes

A	B	C	D	E	F	G	H	I	J	K
discipline_title	participant_country	3_le_games_participant	game_season	game_year	athlete_age	is_host	country_sport	athlete_prev_target		
1	discipline_title	participant_country	3_le_games_participant	game_season	game_year	athlete_age	is_host	country_sport	athlete_prev_target	
2	14	1	100	1	1	2022	23	0	2	1
3	14	1	100	2	1	2022	27	0	2	1
4	14	1	149	2	1	2022	36	0	4	2
5	14	1	149	2	1	2022	32	0	4	2
6	14	1	193	1	1	2022	25	0	2	1
7	14	1	193	4	1	2022	31	0	2	3
8	29	0	36	3	1	2022	30	0	36	2
9	29	0	193	2	1	2022	22	0	6	1
10	29	0	104	2	1	2022	25	0	6	1
11	29	0	151	2	1	2022	21	0	3	2
12	29	0	215	3	1	2022	32	0	35	2
13	29	0	215	2	1	2022	28	0	35	2
14	29	0	193	3	1	2022	31	0	6	1
15	29	0	149	1	1	2022	22	0	9	1
16	29	0	215	1	1	2022	25	0	35	1
17	29	0	168	2	1	2022	30	0	3	1
18	29	0	190	3	1	2022	36	0	16	1
19	29	0	190	1	1	2022	29	0	16	1
20	29	0	190	2	1	2022	22	0	16	3

# Choix des données

## Modèle 1

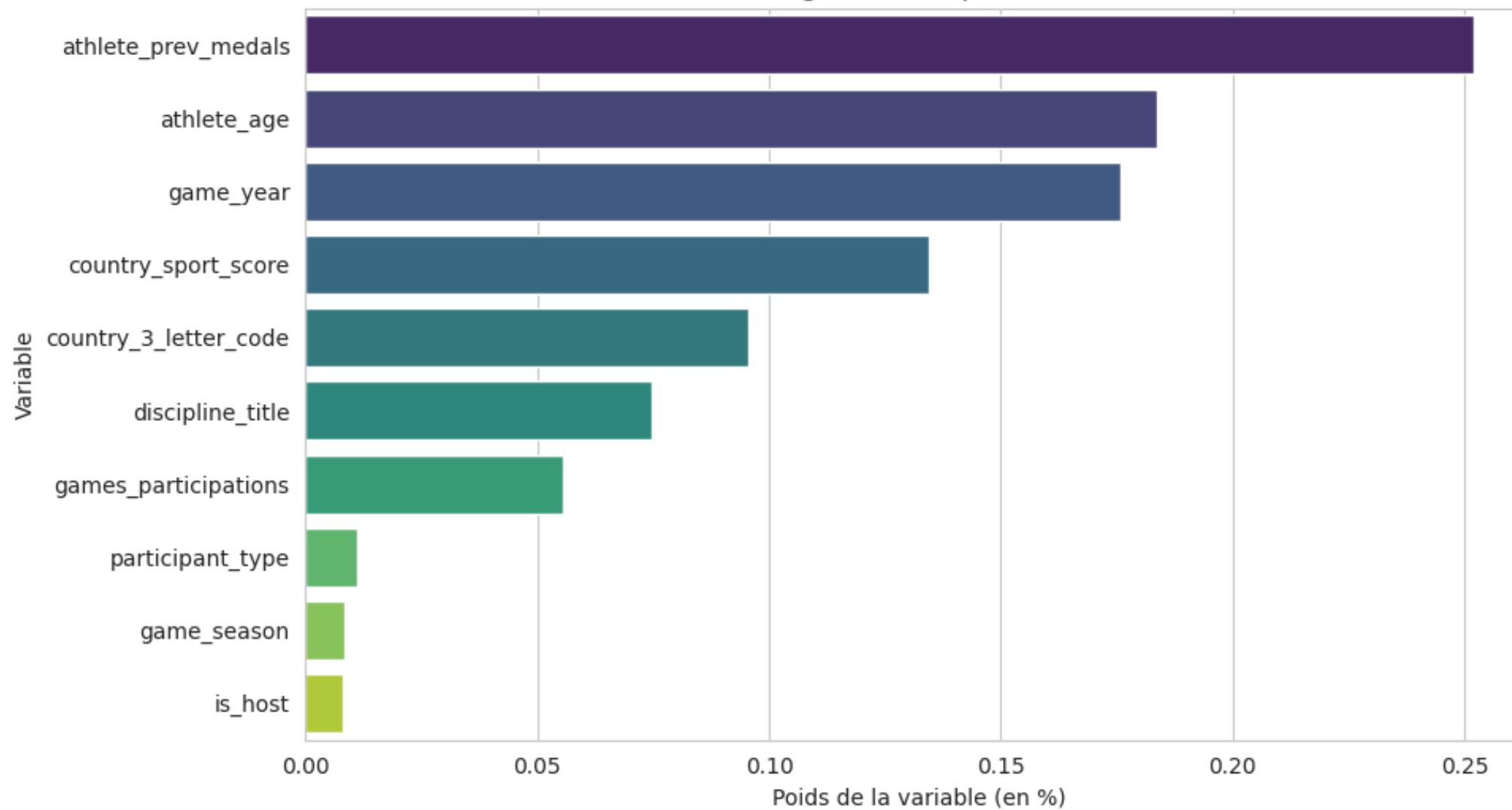
- Index
- Country name
- Discipline
- Game year
- Game season
- Host
- Medals won
- Delegation size
- Total medals available
- Country total medals year

## Modèle 2

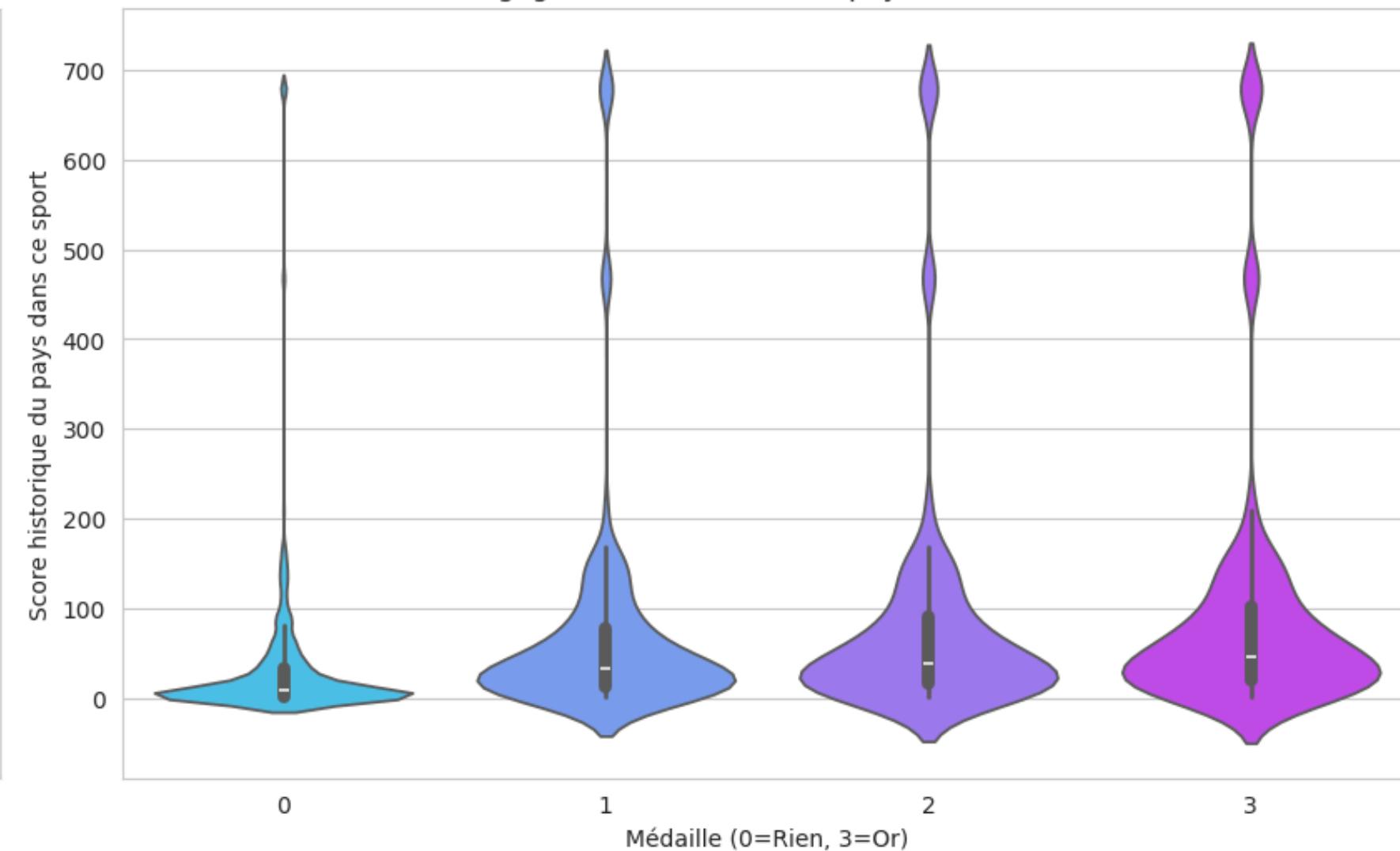
- Index
- Discipline
- Participant type
- Country code
- Games participations
- Game season
- Game year
- Athlete age
- Host
- Country sport score
- Athletes prev medals

# Analyse des données

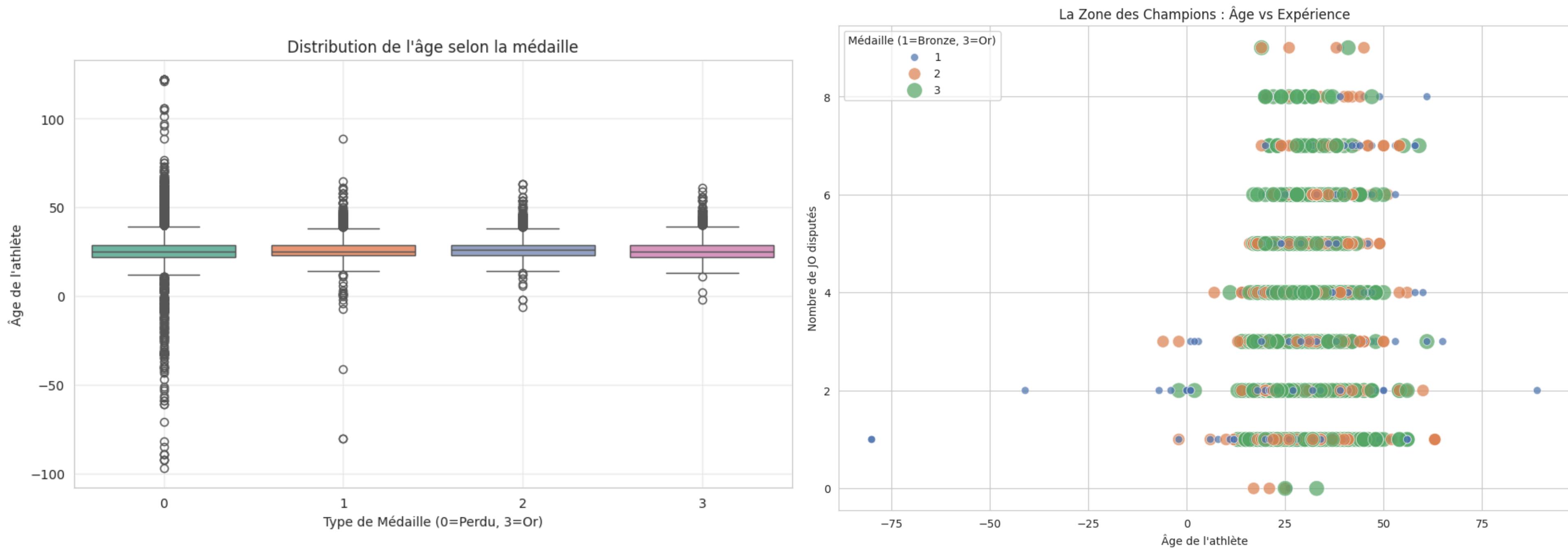
Quels facteurs ton IA regarde-t-elle pour attribuer une médaille ?



Les gagnants viennent-ils de pays dominants ?



# Analyse des données



# Modèles

## Modele 1

Prédire le podium des sports qui vont apporter le plus de médailles pour chaque pays

Prédiction Globale

## Modele 2

Prédire différentes stats sur les JO  
exemple => à domicile, jeune contre vétéran, etc

Prédiction Détalée

# Modèle 1

Nouvelles data :

- historique des médailles
- Potentiel Sport
- nombre d'athlètes par sport
- is host

Modèle utilisé :  
**HistGradientBoosting**

Entrainement avec 9 features avancées...  
Score R<sup>2</sup> (Test) : 0.7765

index	country_name	discipline_title	game_year	game_season	is_host	medals_won	delegation_size	total_medals_available_in_sport	country_total_medals_year	prev_medals_sport	prev_country_power	season_encoded	country_encoded	sport_encoded
0	Afghanistan	Boxing	2012	Summer	0	0	1	91	1	0.0	0.0	0	0.0	12.0
1	Afghanistan	Judo	2012	Summer	0	0	1	98	1	0.0	0.0	0	0.0	38.0
2	Afghanistan	Judo	2016	Summer	0	0	1	98	0	0.0	1.0	0	0.0	38.0
3	Afghanistan	Shooting	2020	Summer	0	0	1	89	0	0.0	0.0	0	0.0	49.0
4	Afghanistan	Taekwondo	2012	Summer	0	1	2	56	1	0.0	0.0	0	0.0	62.0

# Modèle 2

## Random Forest Classifier

- Relation complexe
- Robustesse
- Probabilités

## Fonctionnement

- Un arbre = une partie des données
- Hasard des données, critères
- Mise en commun

## Source de données

- Nettoyage
- Fusion
- Encodage

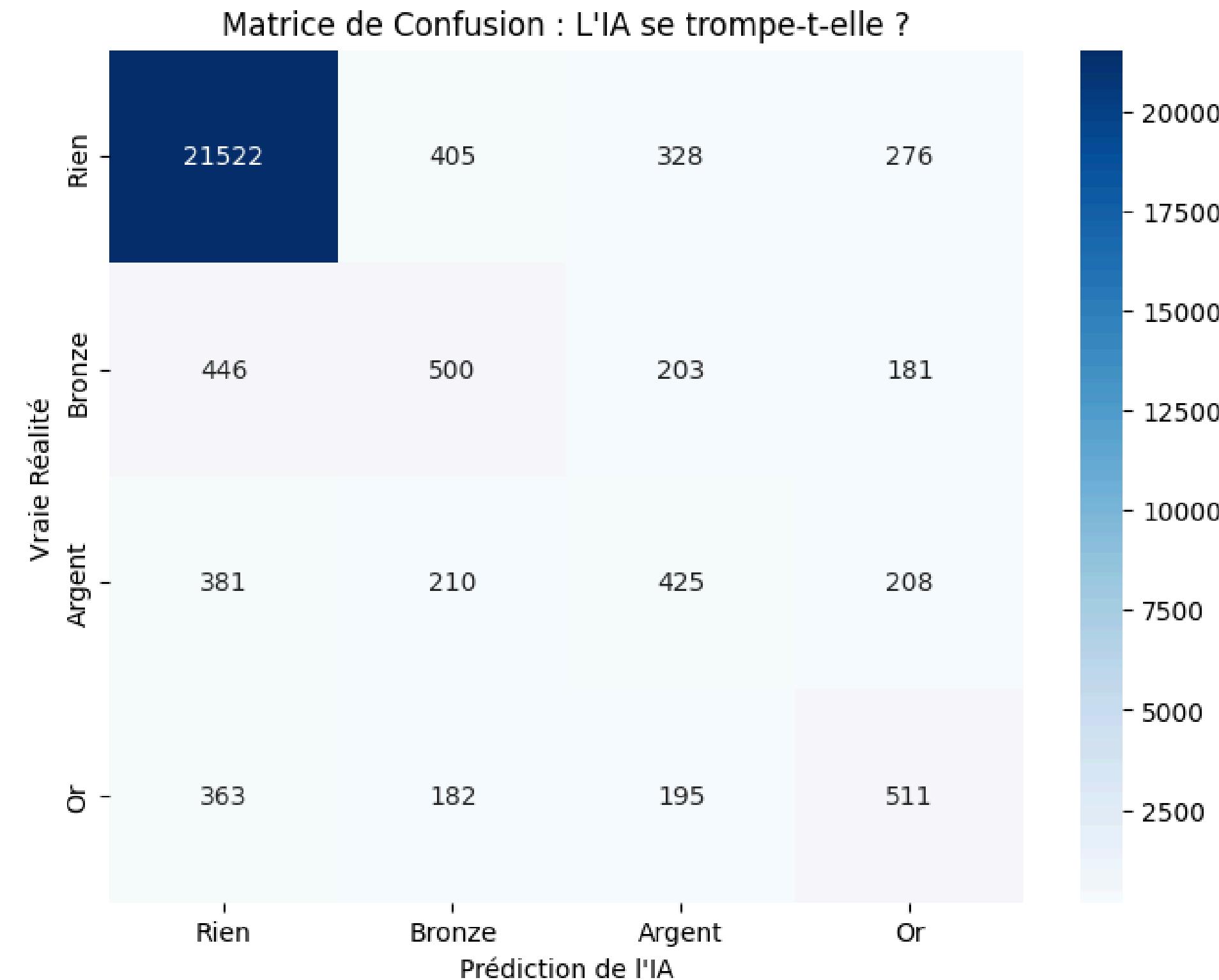
## Valeurs ajoutées

- country\_sport\_score
- athlete\_prev\_medals
- is\_host

Entraînement 80% / Test 20%

Précision de 87,17%

# Modèle 2



# Fonctionnalités de l'application



ANALYTICS

## Analyses

Visualisations avancées et analyses statistiques des données olympiques



## Visualisations

Graphiques interactifs avec Plotly pour explorer les données



## Prédictions IA

Algorithmes avancés pour prédire les performances des pays

## Analyse Historique

Exploration des données olympiques : médailles, pays, sports et athlètes...

## Visualisation

Comparaison des performances des pays / athlètes / médailles cumulées grâce à des graphiques, matrices ...

## Prédictions

Prédiction des performances futures à l'aide de nos modèles de machine learning

mpics Paris 2024

Home Analysis Predictions i

## Fonctionnalités

**Prédictions IA**  
Algorithmes avancés pour prédire les performances des pays

**Visualisations**  
Graphiques interactifs avec Plotly pour explorer les données

**Précision**  
Analyse basée sur les données historiques et tendances

## Notre Équipe

Les talents derrière ce projet.

# Synthèse du projet

## Difficultés rencontrées

- Données : Nettoyage des pays, gestion des éditions manquantes, cohérence des clés.
- Modèles : Choix des features

## Apprentissages

- Vision bout-en-bout : de la donnée brute à une application utilisable.
- Importance de la qualité des données pour obtenir des modèles crédibles.
- Nécessité de penser « utilisateur final » et pas seulement « algorithme ».



# Conclusion

## Bilan du projet

- ➡ Historique des JO → analyse → modèles prédictifs → app
- ➡ Une vision data-driven des performances olympiques.

## Perspectives

Les prochaines étapes qu'on pourrait améliorer :

- Ajouter des données socio-éco, améliorer les modèles, suivi en quasi temps réel pendant de futurs Jeux...