Université Mohammed V de RABAT
École Nationale Supérieure d'Informatique et d'Analyse des Systèmes
-Rabat- (ENSIAS)

Rapport de projet ML, Visualisation des donnees et MLOPS

Prédiction et Visualisation des données du Coupe d'Afrique des Nations CAN 2024

Filière : Génie de la data (GD)

Réalisé par :
AIT BOUAZZA Zaynab
ASLA Mohammed Amine
CHERKAOUI EL KHATTABI Anas
SOUSSOU Youness

Encadré par :
Pr S. EL FKIHI
Pr B. HDIOUAD
Pr M. HAMLAOUI

Devant le jury :
Pr S. EL FKIHI
Pr B. HDIOUAD
Pr M. HAMLAOUI

Année universitaire 2023-2024

Remerciement

Tout d'abord, nous exprimons notre gratitude envers Dieu le Tout-Puissant, qui nous a accordé la force et la patience nécessaires pour mener à bien ce modeste travail.

En second lieu, nous souhaitons témoigner de notre profonde reconnaissance envers Mme Sanaa EL FKIHI, Mme Boutaina HDIOUAD et M. Mahmoud HAMLIOUI pour leur soutien inestimable et leur accompagnement tout au long de ce projet. Leur expertise et leur engagement ont joué un rôle essentiel dans la concrétisation de cette initiative ambitieuse.

Nous tenons à souligner leur contribution significative en fournissant des conseils précieux sur les techniques de prédiction, la visualisation des données et les bonnes pratiques en matière de MLOPS. Grâce à leur passion et à leur dévouement, nous avons pu mener à bien ce projet de manière professionnelle et couronnée de succès.

Résumé

Le projet de prédiction et de visualisation de données sur la Coupe d'Afrique des Nations de football a pour objectif de créer une représentation visuelle permettant l'analyse des résultats passés et la prévision des événements futurs de cet événement sportif international, tout en adoptant une approche MLOps.

Ce projet implique la collecte exhaustive de données sur les matchs depuis la création de la Coupe d'Afrique des Nations en 1957, suivie d'une préparation minutieuse par le biais d'opérations de nettoyage et d'agrégation. L'utilisation de techniques de prédiction de données permettra de prévoir les résultats futurs en tenant compte des données sur les matchs passés et de l'équipe nationale idéale pour la CAN 2024.

Par la suite, le projet se consacrera à la création de visualisations de données variées, offrant une représentation claire et concrète des tendances et des comparaisons qui émergent des données. L'analyse des résultats obtenus permettra de tirer des conclusions sur les performances des différentes équipes et des joueurs au fil des années, ainsi que sur l'évolution générale de la Coupe d'Afrique des Nations de football.

Enfin, l'adoption de l'approche MLOps joue un rôle essentiel en améliorant non seulement l'efficacité du processus, mais également en favorisant une collaboration plus étroite entre les membres de l'équipe et les opérateurs. Cette approche intégrée renforce la qualité de l'analyse et de la prédiction tout en facilitant la gestion et la mise en œuvre des modèles prédictifs au sein du projet.

Mots clés : Machine Learning, Visualisation de données, Coupe d'Afrique des Nations (CAN), Football, Approche MLOps...

Abstract

The prediction and data visualization project on the Africa Cup of Nations football tournament aims to create a visual representation for analyzing past results and forecasting future events of this international sporting event, while adopting an MLOps approach.

This project involves comprehensive data collection on matches since the inception of the Africa Cup of Nations in 1957, followed by meticulous preparation through cleaning and aggregation operations. The use of data prediction techniques will enable forecasting future results, considering data from past matches and the ideal national team for CAN 2024.

Subsequently, the project will focus on creating various data visualizations, providing a clear and concrete representation of trends and comparisons emerging from the data. Analysis of the results obtained will draw conclusions on the performances of different teams and players over the years, as well as the overall evolution of the Africa Cup of Nations football tournament.

Finally, the adoption of the MLOps approach plays a crucial role in enhancing not only the efficiency of the process but also fostering closer collaboration between team members and operators. This integrated approach strengthens the quality of analysis and prediction while facilitating the management and implementation of predictive models within the project.

Table des matières

Table des figures

Introduction Générale

La Coupe d'Afrique des Nations de football incarne l'un des événements sportifs les plus emblématiques et populaires du continent africain. Depuis sa création en 1957, cette compétition internationale a réuni les meilleures équipes nationales africaines dans une lutte acharnée pour remporter la prestigieuse Coupe.

Le football, en tant que sport universel, voit de plus en plus l'importance des données, que ce soit pour analyser les performances des équipes et des joueurs ou pour anticiper les résultats des matchs, y compris la composition idéale des équipes. C'est dans cette perspective que notre projet s'inscrit, visant à mettre en œuvre une prédiction et une visualisation de données sur la Coupe d'Afrique des Nations de football, tout en adoptant une approche MLOps, nous permettant ainsi d'analyser les résultats passés et de prévoir avec précision les événements futurs.

Ce rapport synthétise les différentes étapes nécessaires à la réalisation de ce projet, débutant par la collecte minutieuse des données sur les matchs de la Coupe d'Afrique des Nations. Il explore ensuite les opérations de nettoyage et d'agrégation des données, tout en mettant en lumière l'approche MLOps adoptée pour la prédiction et la visualisation. L'objectif ultime est d'apporter une perspective enrichissante et prédictive à cet événement majeur du football africain.

Première partie
Machine Learning

Chapitre 1

Etude théorique

Dans ce chapitre, nous allons présenter le modèle que nous avons utilisé pour notre projet de prédiction du vainqueur de la Coupe d'Afrique des Nations 2024 et de la composition de notre équipe natinale. Nous allons fournir une vue d'ensemble détaillée de notre modèle, expliquant sa conception et sa mise en place de manière approfondie. De plus, nous allons décrire notre source de données en détail et présenter les différents prétraitements que nous avons appliqués pour garantir la qualité et la pertinence de nos résultats.

1.1 Source des données

1.1.1 La Collecte De Données :

1.1.1.1 Open Source :

Dans notre projet nous avons utilisé des données issues de la plateforme en ligne **Kaggle**. Kaggle est une source de données très riche qui propose une grande variété de données préparées pour l'analyse, ainsi que des outils et des ressources pour faciliter le travail des data scientists et des développeurs. Nous avons choisi d'utiliser Kaggle comme source de données pour notre projet car il offre un accès facile et rapide à des données de qualité qui peuvent être utilisées pour mettre en place des modèles de prédiction fiables et précis.

FIGURE 1.1 – Logo du Kaggle

1.1.1.2 Scraping Data :

Quelle que données qu'on a besoin n'existe pas dans des Open Source donc on a utilisée **Scraping Data** qui est le processus d'importation d'informations depuis un site web vers une feuille de calcul ou un fichier local enregistré sur votre ordinateur. C'est l'une des méthodes les plus efficaces pour obtenir des données sur le web, et dans certains cas, pour acheminer ces données vers un autre site web.
pour ce processus on a utilisé trois méthodes différentes :

— **BeautifulSoup :** Beautiful Soup est une bibliothèque Python permettant l'analyse de documents HTML et XML. Elle crée un arbre d'analyse pour les pages analysées, ce qui peut être utilisé pour extraire des données HTML.

on a utilisée **BeautifulSoup** pour extraire des données des sites statiques

FIGURE 1.2 – Logo du BeautifulSoup

— **Selenium :** est un projet open source regroupant une gamme d'outils et de bibliothèques destinés à prendre en charge l'automatisation des navigateurs. Il propose un outil de lecture pour la création de tests fonctionnels sur la plupart des navigateurs web modernes, sans avoir besoin d'apprendre un langage de script de test. on a utilisée **Selenium** pour extraire les données des sites dynamiques

FIGURE 1.3 – Logo du Selenium

— **Scrapy :** Scrapy est un framework de crawling web gratuit et open source écrit en Python. À l'origine conçu pour le web scraping, il peut également être utilisé pour extraire des données à l'aide d'API ou en tant que crawler web polyvalent. Il est actuellement maintenu par Zyte, une entreprise de développement et de services en web scraping.

on a utilisée **Scrapy** pour extraire les données du plusieurs sous-site de méme site

FIGURE 1.4 – Logo du Scrapy

1.1.2 Description des Données du Prédiction de joueurs convoqués et du joueurs titulaire :

Le jeu de données semble focalise sur les joueurs de football fournissant une variété de statistiques et d'informations notament leur performance dans les matches de saisons, leurs posistions, le classement des leagues dans ils jouent, et leur status dans les derniers matchs dans l'équipe national.

les colonnes sont les suivantes :

- **player_name_sofascore :** Le nom du joueur tel qu'il est répertorié sur le site SofaScore.
- **playername_transfer_market :** Le nom du joueur tel qu'il est répertorié sur le site TransferMarket
- **player_position :** La position du joueur (F pour un attaquant, M pour un milieu de terrain, D pour un défendeur et G pour un gardian de but)
- **player_market_value :** La valeur du joueur au marchée des joueurs.
- **total_game_played :** Le nombre total de matchs joués par le joueur dans la saison.
- **average_rating :** La note moyenne du joueur dans les matches jouées.
- **started_game_percentage :** Le pourcentage de matche dans lesquels le joueur a été titulaire.
- **minute_per_game :** Moyenne de minutes jouées par match.
- **goals_per_game :** Moyenne de buts marqués par match.
- **assist_per_game :** Moyenne de passes décisives par match.
- **dribbled_per_game :** Moyenne de dribbles par match.
- **pourcentage_passes :** Le pourcentage de passes réussies.
- **tackles_per_game :** Moyenne de tackles par match.
- **tackles_per_game :** Moyenne d'interceptions par match.
- **balls_recovered_per_game :** Moyenne de ballons récupérés par match.
- **possession_won :** Moyene de possession gagnées en match.
- **clean_sheets :**Le nombre de match sans encaisser de buts.
- **saves :** Moyenne d'arrêts pour le gardian par match
- **goals_conceded :** Moyenne de buts encaissés par match
- **number_match_called up :** Nombre de matchs pour lesquels le joueur a été convoqué dans les derniers 10 matchs du l'équipe national.
- **number_match_started :** Nombre de matchs que le joueur a été titulaire dans les derniers 10 matchs du équipe national.
- **player_status :** Le cible de notre dataset. il definie le statut actuel du joueur (called-up : convoqué, line-up : titulaire, not-called-up : non convoqué).
- **national_team :** L'équipe nationale à laquelle le joueur est associé.
- **league_ranking :** Le classement de la ligue dans laquelle le joueur joue.
- **player_team :** L'équipe à laquelle le joueur appartient.
- **player_league :** La ligue dans laquelle le joueur joue.

1.1.3 Description des Données du Prédiction de vainqueur du match :

Dans cette parties, nous avons sélectionné un jeu de données qui comprend les résultats de toutes les éditions de la Coupe d'Afrique depuis 1940. Ce jeu de données inclut des informations sur les équipes participantes à savoir :

- **FIFA World Ranking 1992-2022** qui décrit le classement de FIFA des équipes nationals du 1992 au 2022 et qui contient les caractéristiques suivantes :
  - **rank-date :** la date du calcule de rank
  - **countrie_full :** le nom complet de chaque équipe
  - **confederation :** FIFA confederations du l'équipe national
  - **rank :** rank du l'équipe national

— **International football results from 1872 to 2022** qui décrit tous les matches jouées par les équipes national du 1872 à 2022 et qui contient les caractéristiques suivantes :
  — **date :** la date du match
  — **home_team :** le nom de l'equipe a domicile
  — **away_team :** le nom de l'equipe à l'extérieur
  — **home_score :** score final de l'équipe à domicile après le temps réglementaire, y compris les prolongations, mais excluant les tirs au but
  — **away_score :** score final de l'équipe à l'extérieur après le temps réglementaire, y compris les prolongations, mais excluant les tirs au but
  — **tournament :** nom de tournament

1.2 Pré-traitement des données :

Le prétraitement des données consiste à nettoyer et préparer les données pour l'analyse. Cela inclut la suppression des valeurs manquantes, la normalisation des données et la transformation des variables. Nous avons effectué ces étapes pour rendre nos données utilisables pour notre projet de prédiction et visualisation de données sur la Coupe d'Afrique 2024. Le prétraitement des données est une étape cruciale pour garantir la qualité et la précision de nos résultats.

1.2.1 outils et bibliothèques utilisés

1.2.1.1 Python

Python est un langage de programmation haut niveau, orienté objet, à sémantique dynamique et interprété. Ses structures de données intégrées de haut niveau, associées à un typage dynamique et une liaison dynamique, en font un choix idéal pour le développement rapide d'applications et comme langage de script ou de liaison pour connecter des composants existants. La syntaxe concise et facile à apprendre de Python favorise la lisibilité, ce qui réduit les coûts de maintenance logicielle.

Python prend en charge les modules et les packages, ce qui facilite la modularité des programmes et la réutilisation de code. L'interpréteur Python et sa bibliothèque standard complète sont gratuits à télécharger et à distribuer sous forme de code source ou binaire pour toutes les principales plates-formes.

FIGURE 1.5 – Python logo

1.2.1.2 Scikit-learn

Scikit-learn, souvent abrégé sklearn, est une bibliothèque open-source incontournable dédiée à l'apprentissage automatique (machine learning) en Python. Elle offre un ensemble complet d'outils pour la modélisation, la classification, la régression, le clustering, et bien plus encore. Scikit-learn s'intègre harmonieusement avec d'autres bibliothèques populaires telles que NumPy, SciPy et Matplotlib, facilitant ainsi le flux de travail des développeurs et des chercheurs.

La bibliothèque propose une vaste gamme d'algorithmes, des méthodes d'évaluation de modèles, et des utilitaires pour le prétraitement des données. Que ce soit pour la classification d'images, la prédiction de séries temporelles ou d'autres tâches, Scikit-learn demeure un choix de prédilection pour ceux qui cherchent à exploiter le potentiel de l'apprentissage automatique dans leurs projets Python.

FIGURE 1.6 – Sklearn logo

1.2.2 Pré-traitement des données du prédiction des joueurs convoquées et titulaire :

Dans une premièr temps, nous avons procédé à l'élimination des données jugées non pertinentes pour notre modèle de prédiction, telles que le nom individuel des joueurs.

Ensuite, nous avons élaboré une matrice de corrélation ainsi que des diagrammes en violons. Ces outils d'analyse nous ont permis d'identifier les données ayant une influence significative sur notre variable cible, ainsi que celles présentant une forte corrélation avec cette variable. De plus, nous avons examiné les relations importantes entre les différentes caractéristiques, mettant en évidence les corrélations essentielles pour comprendre le comportement de la variable cible. En parallèle, nous avons créé des matrices de corrélation spécifiques à chaque position, permettant ainsi une analyse plus approfondie et ciblée des données en fonction des rôles spécifiques des joueurs sur le terrain. Cette approche méticuleuse de la sélection et de l'analyse des données constitue un socle solide pour la construction d'un modèle de prédiction robuste.

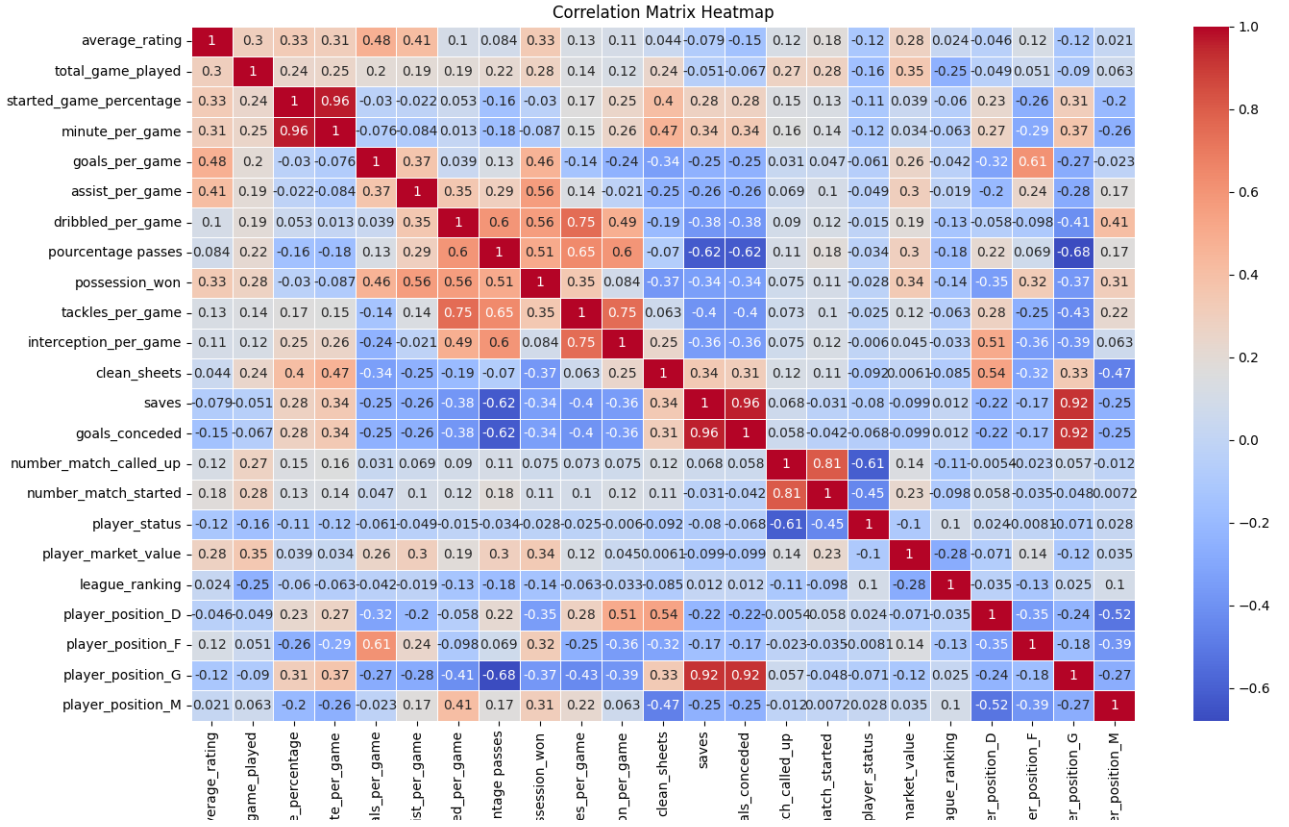

FIGURE 1.7 – Matrice de correlation total

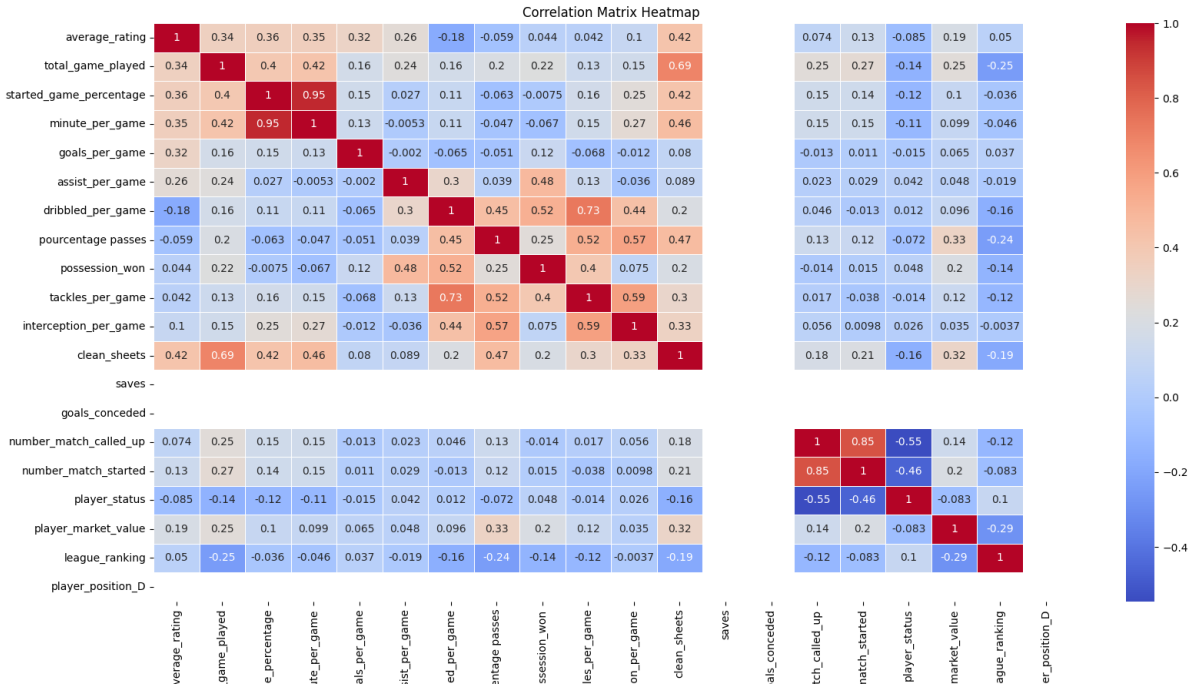

FIGURE 1.8 – Matrice de correlation du position défendeurs

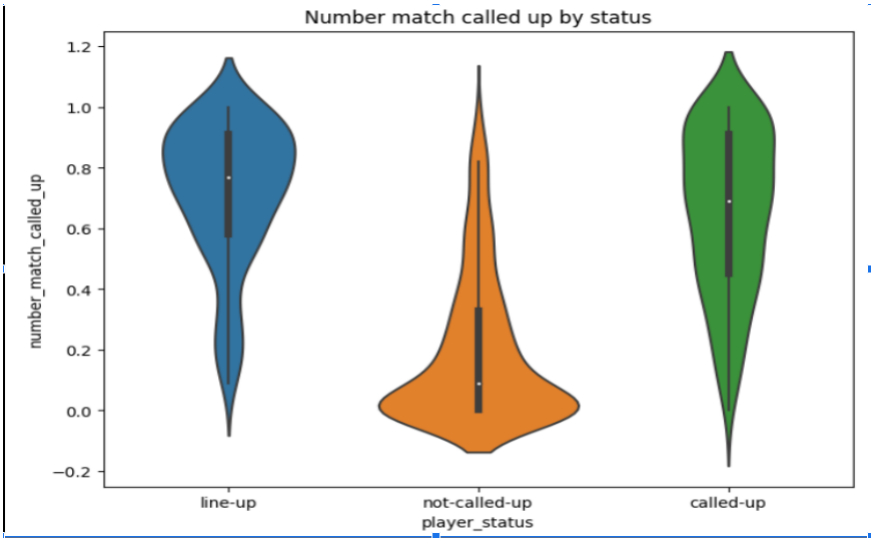

FIGURE 1.9 – Diagramme en violons du caractéristique number_match_called_up

puis on a classée ces valeurs selon leurs relation avec le cible et on a pris les décisions suivantes

- éliminer la colonne tackles_per_game car elle n'influence pas notre cible dans le cas du défandeurs.
- éliminer les colonnes pourcentage-passes et dribble-per-game qui n'ont pas une grande influance sur le cible pour toutes les positions.
- éliminer la colonne minute-per-game qui a une grande relation avec started-game-percentage et pas une influance sur le cible.
- combiner les deux colonnes saves-per-games et goals-conceded en une seule colonne pourcentage-saves (= saves_per_game / (saves_per_game + goals_conceded)) à cause de la grande relation entre eux.

1.2.3 Pré-traitement Des données du prédiction des vainqueurs dans coupe d'Afrique :

Ici on a concaténer les tables en un seul table puis on a éliminée les données qui non pas nécessaire puis on calculée plusieurs champs pour nous aidée à prédire et la table final a constituée de ces caractéristiques :

- **diff_classement :**différence entre classement de FIFA du deux équipes.
- **diff_but :**différence entre les moyennes des buts marquées dans les 5 derniers matches du coupe Afrique des deux équipes.
- **diff_but_subis :**différance entre les moyennes des buts subis dans les 5 derniers matchs du coupe Afrique
- **diff_win :**differance entre les matches vainquis dans les 5 derniers matches du coupe d'Afrique.
- **resultat :**la colonne cible qui a trois valeur home_win (l'équipe à domicile a vainquis le match) away_win (l'équipe à l'extérieur a vainquis le match) egalite (égalité entre les 2 équipes)

1.3 Le choix du modèle

1.3.1 Méthode du choix du modèle :

Pour choisir notre modèle on a utilisée les pipelines pour comparer les performances de tous les modéles et retourner la modéle avec la meilleurs performances.

Dans ce contexte, on a constaté que faire la prédiction des joueurs titulaire et les joueurs convoqués a donnée un mieux performance qu'on divise ces deux prédictions. Donc on a choisis de prédire les joueurs convoqués puis faire une autre prédiction sur ces joueurs pour prédire les joueurs titulaires

1.3.2 Les modèles choisis :

1.3.2.1 Logistic Regression :

Pour la prediction des joueurs titulaire et des joueurs convoqués on a choisis la regression logistique qui est un modèle statistique permettant d'étudier les relations entre un ensemble de variables qualitatives Xi et une variable qualitative Y. Il s'agit d'un modèle linéaire généralisé utilisant une fonction logistique comme fonction de lien.

Un modèle de régression logistique permet aussi de prédire la probabilité qu'un événement arrive (valeur de 1) ou non (valeur de 0) à partir de l'optimisation des coefficients de régression. Ce résultat varie toujours entre 0 et 1. Lorsque la valeur prédite est supérieure à un seuil, l'événement est susceptible de se produire, alors que lorsque cette valeur est inférieure au même seuil, il ne l'est pas.

- **l'avantage du régression logistique :**
  - **Interprétabilité :**Les résultats de la régression logistique sont faciles à interpréter. Les coefficients de régression représentent la force et la direction de l'association entre chaque variable indépendante et la variable dépendante.
  - **Efficient pour des problèmes de classification binaire :** La régression logistique est bien adaptée aux problèmes de classification binaire, où l'objectif est de prédire une des deux classes.
  - **Moins enclin au surajustement (overfitting) :** Comparé à des modèles plus complexes, la régression logistique a moins de risques de surajustement, ce qui la rend efficace avec des ensembles de données de taille modérée.
  - **Utilisable avec des données peu nombreuses :** Elle peut être utilisée avec des ensembles de données relativement petits tout en fournissant des résultats robustes.
  - **Calcul rapide et peu gourmand en ressources :** La régression logistique est computationnellement efficace et peut être implémentée rapidement, même sur des ensembles de données volumineux.
  - **Peu de paramètres à ajuster :** En général, il y a moins de paramètres à ajuster par rapport à d'autres algorithmes, ce qui simplifie le processus de réglage du modèle.
- **l'inconvénient du régression logistique :**
  - **Linéarité des données :** La régression logistique suppose une relation linéaire entre les variables indépendantes et la log-odds de la variable dépendante. Si cette relation n'est pas respectée, le modèle peut ne pas fonctionner de manière optimale.
  - **Limité aux problèmes de classification binaire :** La régression logistique est principalement conçue pour des problèmes de classification binaire. Elle doit être étendue pour résoudre des problèmes de classification multiclasse.
  - **Sensible aux outliers :** La régression logistique peut être sensible aux valeurs aberrantes (outliers), ce qui peut affecter les coefficients de manière significative.
  - **Pas appropriée pour des relations complexes :** Si la relation entre les variables indépendantes et dépendantes est complexe, la régression logistique peut ne pas être en mesure de la modéliser efficacement.
  - **Assumption d'indépendance des observations :** La régression logistique suppose que les observations sont indépendantes. Si les données présentent une structure de dépendance, cela peut affecter les performances du modèle.

— **Non adaptée aux données non structurées :** La régression logistique fonctionne mieux avec des données bien structurées. Pour des tâches impliquant des données non structurées, d'autres méthodes peuvent être plus appropriées.

1.3.2.2 Gradient Boosting :

On a choisis le gradient boosting pour prédire les vainqueurs des matchs. Le gradient boosting est une technique d'apprentissage automatique qui construit un modèle prédictif robuste en combinant séquentiellement plusieurs modèles plus faibles, généralement des arbres de décision, pour corriger les erreurs des modèles précédents. Il commence par construire un modèle de base, calcule les résidus entre les prédictions et les valeurs réelles, puis ajuste un nouveau modèle pour prédire ces résidus. Ce processus est répété plusieurs fois, chaque ajout de modèle améliorant la précision globale du modèle en corrigeant les erreurs résiduelles accumulées. Le résultat final est obtenu en combinant les prédictions de tous les modèles de manière pondérée. Le gradient boosting est particulièrement puissant et largement utilisé pour des tâches de régression et de classification, avec des implémentations populaires telles que XGBoost et LightGBM. Cependant, il peut être sensible au surajustement, et des techniques de régularisation sont souvent utilisées pour atténuer ce risque.

— **l'avantage du GradientBoosting :**
  — **Puissance Prédictive :** Le gradient boosting est connu pour sa capacité à produire des modèles très performants, souvent en tête des compétitions de machine learning.
  — **Adaptabilité :** Il peut être utilisé pour résoudre une variété de problèmes, y compris la régression et la classification, et peut être adapté à différentes fonctions de perte.
  — **Gestion des Données Hétérogènes :** Il peut gérer naturellement des ensembles de données hétérogènes, comprenant des variables de différents types (numériques, catégorielles) sans nécessiter une préparation de données extensive.
  — **Capturation de Relations Non Linéaires :** Grâce à la construction séquentielle de modèles, le gradient boosting peut capturer des relations non linéaires complexes entre les variables.
— **l'inconvénient du GrandientBoosting :**
  — **Sensibilité aux Paramètres :** Il peut être sensible au choix des hyperparamètres, et un mauvais réglage peut conduire au surajustement ou à une convergence lente.
  — **Temps de Formation :** La construction séquentielle de modèles peut rendre le processus d'entraînement plus lent par rapport à d'autres méthodes.
  — **Risque de Surajustement :** Sans une bonne régularisation, le gradient boosting peut être sujet au surajustement, en particulier si le nombre d'itérations est trop élevé.
  — **Interprétabilité Limitée :** Comme avec d'autres modèles d'ensemble, l'interprétation des résultats peut être complexe en raison de la combinaison de multiples modèles.

Chapitre 2

Réalisation du projet

2.1 Introduction

Dans cette partie, nous présenterons la partie entrainement du modèle, puis la validation du modèle après l'entrainement et enfin les résultats de la prédiction des joueurs convoqués et la prédiction des joueurs titulaire et la prédiction des résultat des matches dans coupe d'Afrique

2.2 Accuracy et Validation des modèles

2.2.1 Définition des paramètres de validation :

2.2.1.1 Matrice de confusion :

La matrice de confusion est une table qui est souvent utilisée pour décrire les performances d'un modèle de classification sur un ensemble de données. Elle compare les prédictions du modèle avec les vraies classes des échantillons. La matrice de confusion a quatre entrées principales : vrais positifs (True Positives - TP), faux positifs (False Positives - FP), vrais négatifs (True Negatives - TN) et faux négatifs (False Negatives - FN).

Voici comment ces termes sont définis dans le contexte d'une matrice de confusion :

- **Vrais Positifs (True Positives - TP) :**Le modèle a correctement prédit que l'échantillon appartient à la classe positive.
- **Faux Positifs (False Positives - FP) :** Le modèle a incorrectement prédit que l'échantillon appartient à la classe positive alors qu'en réalité, il appartient à la classe négative.
- **Vrais Négatifs (True Negatives - TN) :** Le modèle a correctement prédit que l'échantillon appartient à la classe négative.
- **Faux Négatifs (False Negatives - FN) :** Le modèle a incorrectement prédit que l'échantillon appartient à la classe négative alors qu'en réalité, il appartient à la classe positive.

La matrice de confusion est généralement représentée comme suit :

$$\text{Matrice de Confusion} = \begin{bmatrix} \text{TP} & \text{FP} \\ \text{FN} & \text{TN} \end{bmatrix} \tag{2.1}$$

Cette matrice est utile pour évaluer les performances d'un modèle de classification en fournissant une vue détaillée des types d'erreurs qu'il commet. À partir de cette matrice, plusieurs mesures d'évaluation, telles que la précision, le rappel, le F1-score, et l'exactitude, peuvent être calculées.

2.2.1.2 Exactitude (Accuracy) :

L'exactitude mesure la correction globale du modèle. Elle représente le ratio d'instances correctement prédites par rapport au nombre total d'instances.

L'exactitude est définie comme :

$$\text{Exactitude} = \frac{\text{Vrais Positifs} + \text{Vrais Négatifs}}{\text{Total d'Instances}} \tag{2.2}$$

Interprétation : Une exactitude élevée indique une bonne performance globale, mais elle peut être trompeuse en présence de classes déséquilibrées.

2.2.1.3 Précision :

La précision est la capacité du modèle à identifier correctement les instances positives parmi toutes les instances prédites comme positives.

La précision est définie comme :

$$\text{Précision} = \frac{\text{Vrais Positifs}}{\text{Vrais Positifs} + \text{Faux Positifs}} \tag{2.3}$$

Interprétation : Une précision élevée signifie que lorsque le modèle prédit positif, il a de fortes chances d'être correct. C'est important lorsque la minimisation des faux positifs est cruciale.

2.2.1.4 Rappel (Sensibilité ou Taux de Vrais Positifs) :

Le rappel est la capacité du modèle à identifier correctement toutes les instances positives parmi toutes les instances réellement positives.

Le rappel, Sensibilité ou Taux de Vrais Positifs) est défini comme :

$$\text{Rappel} = \frac{\text{Vrais Positifs}}{\text{Vrais Positifs} + \text{Faux Négatifs}} \tag{2.4}$$

Interprétation : Un rappel élevé indique que le modèle est bon pour capturer les instances positives. C'est important lorsque la minimisation des faux négatifs est cruciale.

2.2.1.5 F1-score :

Le F1-score est la moyenne harmonique de la précision et du rappel. Il offre un équilibre entre précision et rappel.

Le F1-score est défini comme :

$$\text{F1-score} = 2 \times \frac{\text{Précision} \times \text{Rappel}}{\text{Précision} + \text{Rappel}} \tag{2.5}$$

Interprétation : Le F1-score est particulièrement utile en présence d'une distribution inégale des classes, car il équilibre le compromis entre précision et rappel.

2.2.2 Paramètre de validation pour la prédiction des joueurs convoqués :

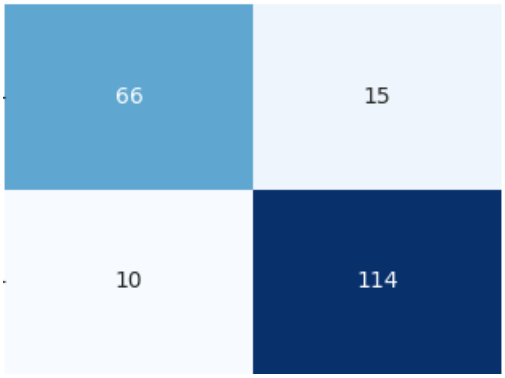

FIGURE 2.1 – Matrice de confusion pour la prédiction des joueurs convoqués

— **Accuracy :** 87.80
— **Precision :** 88.37
— **Recall :** 91.93
— **F1-score :** 90.11

2.2.3 Paramètre de validation pour la prédiction des joueurs titulaires :

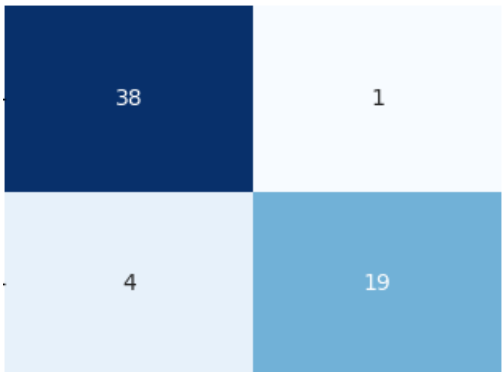

FIGURE 2.2 – Matrice de confusion pour la prédiction des joueurs titulaires

— **Accuracy :** 91.93
— **Precision :** 95
— **Recall :** 82.60
— **F1-score :** 88.37

2.2.4 Paramètre de validation pour la prédiction des vainqueurs des matches :

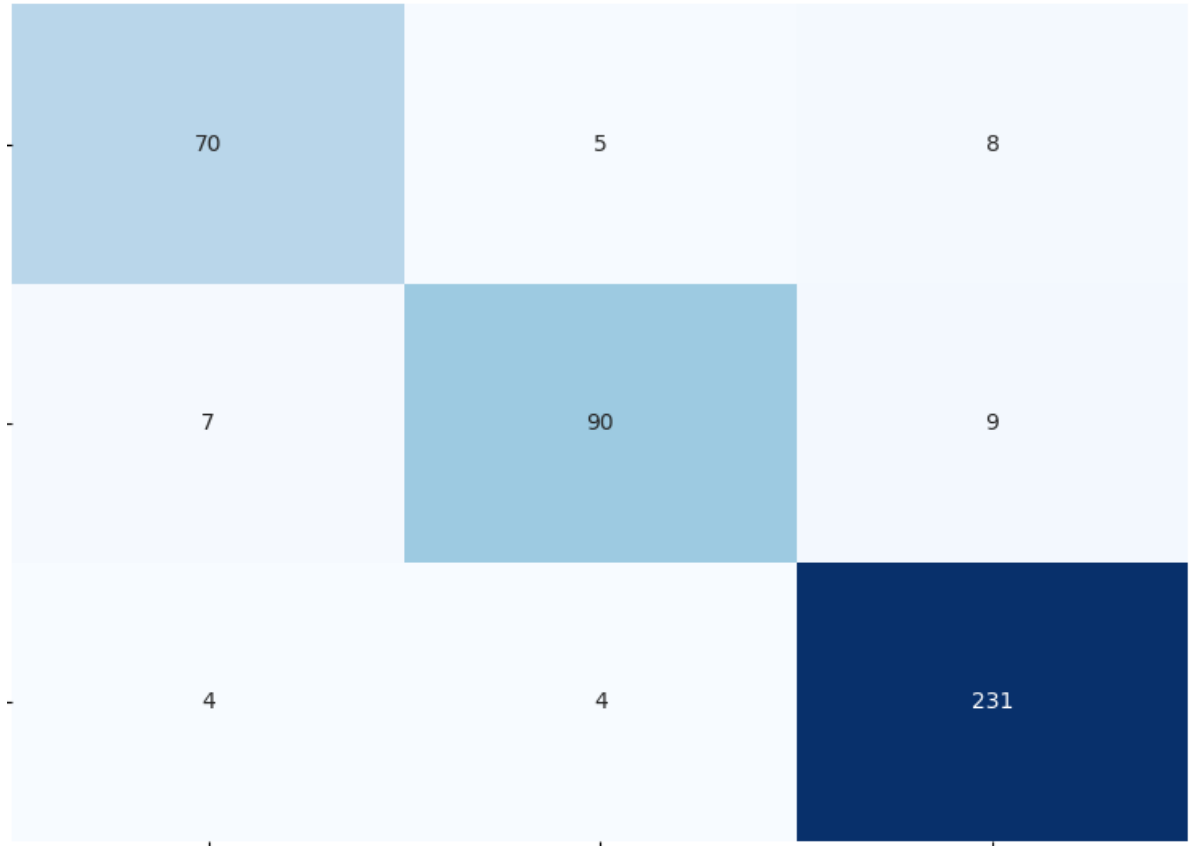

FIGURE 2.3 – Matrice de confusion pour la prédiction des vainqueur des matches

— **Accuracy :** 91.35
— **Precision :** 90.15
— **Recall :** 88.63
— **F1-score :** 89.34

2.2.5 interpretation :

Les valeurs de précision, de rappel, de score F1 et d'exactitude de tous les prédictions indiquent des performances équilibrées et généralement bonnes des modèles sur plusieurs métriques.

2.3 Simulation des modéles de machines learning :

Maintenant on vas faire la prédiction du joueurs convoquées et titulaire au équipe marocain et on vas prédire le gagnant du coupe d'Afrique 2023

2.3.1 Prediction des joueurs convoqués par l'équipe marocain :

En premier lieu, ici c'est la liste convoqués par notre coach Walid Regragui :

FIGURE 2.4 – liste d'équipe national marocain

et la et la liste prédit par notre modéle :

FIGURE 2.5 – liste d'équipe national marocain prédit

On voit les joueurs que notre modéle a mal prédite par la couleur rouge, cas on peut l'explique quelque example par :

- Quelques joueurs ont toujours participé avec l'équipe national et l'autre n'as jamais participé (par example Chadi Riad et Jawad El Yamiq)
- La décision de le coach de ne pas convoquer des joueurs car il n'as pas satisfait par leurs jeux méme si leurs statistique est mieux que tout les joueurs convoqués (par example ABDRAZAK HAMDALLAH)

2.3.2 Prédiction des joueurs titulaire par l'équipe national :

Ici c'est la prédiction des joueurs titulaire par l'équipe national marocain :

FIGURE 2.6 – line-up d'équipe national marocain prédite

2.3.3 Prédiction des resultat du CAN 2023 :

La figure ci-dessous montre la prédiction du resultat de la coupe d'Afrique 2023 d'aprés notre modéle :

FIGURE 2.7 – la phase du groupe dans la CAN 2023 prédite

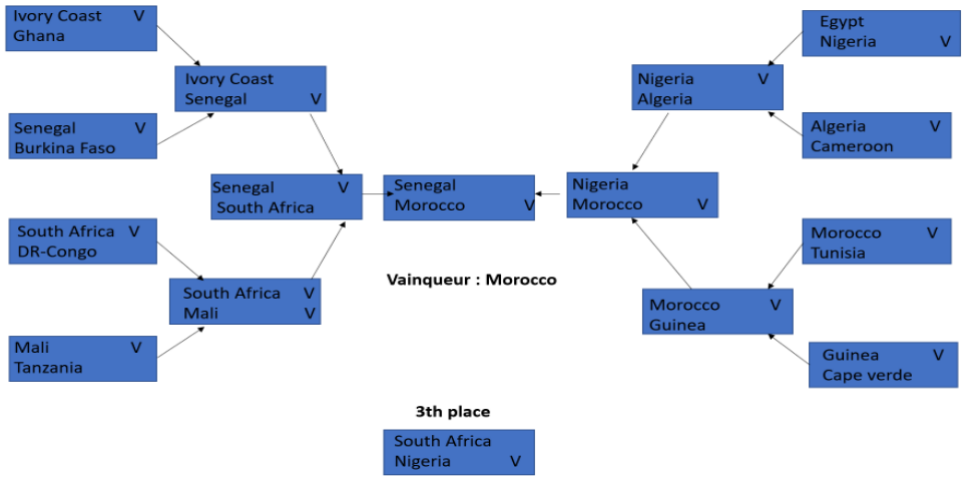

FIGURE 2.8 – la phase du Knock out dans la CAN 2023 prédite

comme on voit notre modéle a prédit que le maroc est le vainqueur du coupe d'Afrique 2023 (on n'a pas la méme opinion)

2.4 Conclusion

Méme si on peut utiliser les statistiques pour prédire les évenement de foot mais c'est impossible de être 100% sure de ce que vas passer car il y a beaucoup des événements qu'on peut pas prédire par example la montalité des joueurs le jour de match, un carton rouge dans les premiers minute, un but dans les premiers minutes...

Deuxième partie
Visualisation de données

Chapitre 3

Étude théorique

Au cours de ce chapitre, nous explorerons en détail les concepts fondamentaux et les approches théoriques qui constitueront la base de notre projet de visualisation de données sur la Coupe d'Afrique des Nations. Nous introduirons les différentes méthodes et outils que nous mettrons en œuvre dans le futur pour collecter, préparer et visualiser les données, tout en identifiant les sources de données qui guideront notre analyse.

3.1 Source des données

Dans la phase préliminaire de notre projet, la première étape consiste à identifier avec précision les sources de données qui seront exploitées. Initialement, nous avons fait usage de données officielles concernant la Coupe d'Afrique des Nations de football. Ces données, émanant de la Fédération Internationale de Football Association (FIFA), comportent des détails exhaustifs sur les matchs de la Coupe d'Afrique des Nations, ainsi que des informations sur les équipes et les joueurs ayant pris part à ces rencontres.

En complément de ces données officielles, nous avons également intégré des bases de données en ligne dédiées au football, telles que celles fournies par des sites spécialisés comme Sofascore. De plus, nous avons exploré des bases de données disponibles sur des plateformes telles que Kaggle. Ces ressources supplémentaires nous ont permis d'enrichir les informations relatives aux matchs de la Coupe d'Afrique des Nations. Elles offrent notamment des statistiques détaillées sur les performances des nations participantes ainsi que des analyses tactiques approfondies des stratégies employées par les équipes.

3.2 Préparation des données

Une fois les données nécessaires à la visualisation de la Coupe d'Afrique des Nations de football réunies, la phase cruciale suivante consiste à les préparer pour une intégration efficace dans notre analyse. Cette préparation garantit la propreté et la fiabilité des données, les rendant prêtes à être exploitées dans les différentes étapes de notre travail. Notre approche comprend plusieurs étapes clés.

Tout d'abord, nous réalisons une opération de nettoyage pour éliminer toute information manquante ou incorrecte, assurant ainsi l'intégrité des données. Ensuite, nous effectuons un traitement des données pour les structurer de manière adaptée à notre objectif, impliquant souvent leur transformation pour les rendre compatibles avec nos outils de visualisation.

Une étape essentielle de la préparation des données est l'agrégation, regroupant les données de manière à les rendre plus faciles à manipuler et à visualiser. Cela peut inclure le calcul de statistiques sur les performances des équipes et des joueurs, ou encore le regroupement par année, par équipe, ou par phase du tournoi.

Enfin, une vérification minutieuse de la qualité des données préparées est effectuée, garantissant qu'elles répondent aux critères nécessaires pour être utilisées dans les phases ultérieures du projet. Cette approche méticuleuse de la préparation des données constitue la base robuste de notre analyse, assurant la fiabilité des résultats obtenus.

3.3 Outils de travail

3.3.1 Tableu

Tableau est un outil de visualisation de données se distinguant par sa popularité et sa capacité à générer des graphiques et des tableaux de bord à partir de diverses sources de données. Sa pertinence s'affirme particulièrement lorsqu'il s'agit de manipuler des volumes importants de données, offrant ainsi une flexibilité remarquable dans la création de visualisations personnalisées.
L'atout majeur de Tableau réside dans sa capacité à rendre les tendances et les comparaisons issues des données de la Coupe d'Afrique des Nations de football claires et concrètes, que ce soit pour représenter les performances des équipes, les statistiques des joueurs, ou encore les évolutions au fil des éditions du tournoi.

FIGURE 3.1 – Tableau Logo

3.3.2 Seaborn

Seaborn se présente comme une bibliothèque de visualisation de données remarquable pour Python, offrant une plateforme riche pour créer des graphiques de qualité à partir de données statistiques. Ce qui distingue Seaborn, c'est sa capacité à générer des visualisations esthétiques et informatives tout en simplifiant le processus pour les utilisateurs. Elle propose une variété de graphiques prédéfinis, facilitant ainsi la représentation visuelle des résultats statistiques. De plus, son adaptabilité se manifeste à travers des options avancées de personnalisation.

Dans le cadre de notre projet, nous avons choisi Seaborn pour donner vie aux résultats de nos prédictions de données de manière à la fois précise et professionnelle. Cette bibliothèque sera notre alliée dans la création de graphiques expressifs, facilitant ainsi la communication efficace des résultats de notre travail d'analyse et de prédiction.

FIGURE 3.2 – Seaborn Logo

3.3.3 Plotly

Plotly se distingue en tant que bibliothèque de visualisation de données interactive et dynamique, offrant des fonctionnalités avancées pour la création de graphiques et de tableaux de bord.

En utilisant Plotly dans notre projet, nous cherchons à transcender la simple représentation statique des données. Cette bibliothèque nous permettra de créer des visualisations engageantes qui peuvent être explorées en détail, renforçant ainsi la compréhension des tendances et des relations au sein de nos données.

FIGURE 3.3 – Plotly Logo

3.4 Conclusion

Au fil de ce chapitre, nous avons examiné en détail les concepts fondamentaux et les approches théoriques qui forment la base de notre projet de visualisation de données sur la Coupe d'Afrique des Nations. Nous avons introduit les divers outils et méthodes que nous mettrons en œuvre pour la collecte, la préparation et la visualisation des données, tout en identifiant les sources de données que nous prévoyons d'utiliser.

Chapitre 4

Réalisation du projet et résultats

Dans cette section consacrée à l'étude pratique et aux résultats de notre projet, l'objectif principal est de dévoiler les résultats obtenus à travers des tableaux, des graphiques et des visualisations interactives. Ces représentations mettront en avant les tendances marquantes et les comparaisons significatives, tirées des données concernant la Coupe d'Afrique des Nations (CAN) de football.

4.1 Source des données

4.1.1 A partir des sites web à l'aide de Selenium Et BeautifulSoup

Selenium et Beautiful Soup jouent un rôle crucial dans l'extraction et la manipulation de données pour notre projet. En exploitant les capacités de collecte de données de Selenium en tandem avec l'analyse structurée de Beautiful Soup, nous avons la capacité d'extraire divers types de données des pages web, couvrant à la fois les données tabulaires et celles qui ne sont pas présentées dans des tableaux.
Les sites web ciblés comprennent :

- le Site officiel de la FIFA
- Sofascore
- 365scores
- Wikipedia
- Transfer market

Cette combinaison habile de Selenium et Beautiful Soup offre une approche robuste et complète pour acquérir les données essentielles à notre projet de visualisation, garantissant ainsi une base solide pour nos analyses ultérieures.

4.1.2 A partir de Kaggle

Les données extraites dans la section consacrée au Machine Learning depuis Kaggle sont également mises à profit dans cette section pour élaborer des visualisations basées sur les historiques des joueurs et des équipes.

4.2 Nettoyage des données

Pour assurer des analyses précises, un nettoyage des données s'est avéré nécessaire en raison de certaines valeurs erronées. Les actions de nettoyage entreprises comprennent :

— supprimer des lignes vides et des doublons
— renommer des colonnes
— modifier des types de données
— supprimer des relations
— identifier et corriger des erreurs

À l'issue de ces étapes, la fiabilité de nos données a été garantie, les rendant ainsi prêtes à être utilisées dans notre analyse.

4.3 Visualisation de données

4.3.1 Visualisation des equipes ayant gagnés la CAN

— Carte à bulles décrivant la distribution des pays gagnants

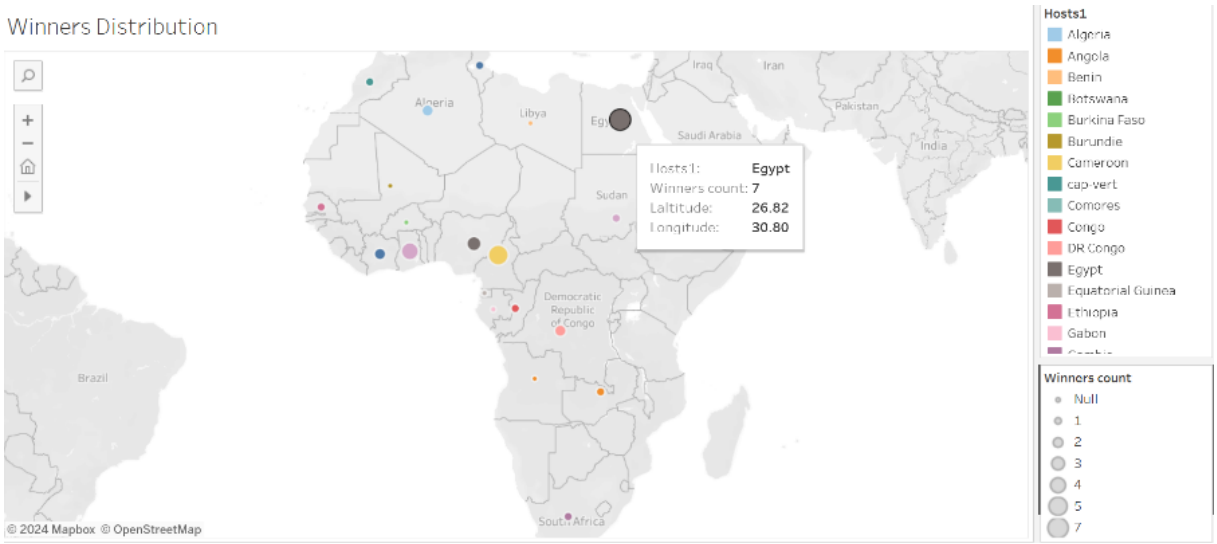

FIGURE 4.1 – la distribution des pays gagnants

Notre carte à bulles dévoile les parcours triomphants des nations qui ont marqué l'histoire de la Coupe d'Afrique des Nations (CAN) depuis ses débuts.

La taille imposante des bulles évoque la grandeur des exploits de chaque nation. L'Égypte, le Cameroun, le Nigeria, et d'autres encore, s'affichent en majesté, témoins de leurs victoires glorieuses au fil des éditions de la CAN.

— Diagramme Sankey représentant les pays gagnants pour chaque éditions de CAN

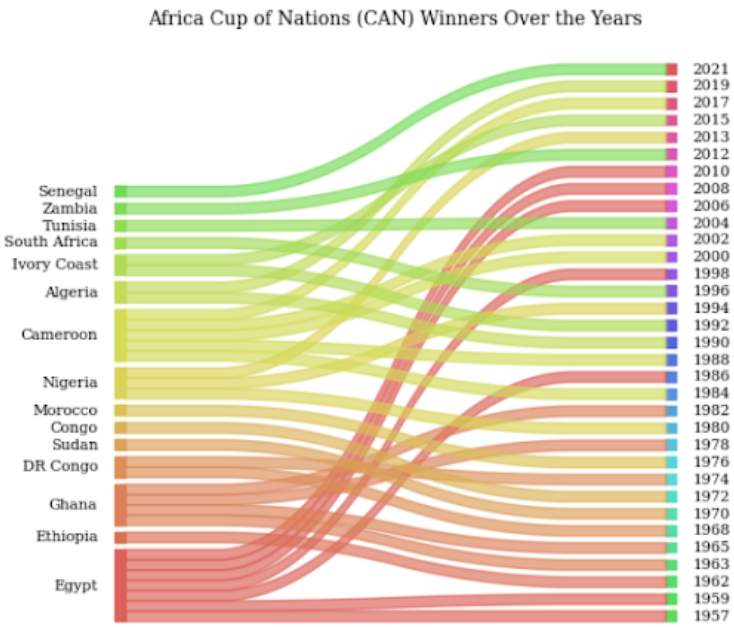

FIGURE 4.2 – Les gagnants de la CAN tout au long des années

Chaque flux sur le graphique représente un pays, et la largeur du flux est proportionnelle au nombre de fois où ce pays a atteint le sommet du football africain. Chaque lien entre les blocs indique une victoire dans une année spécifique. Ainsi, ce diagramme crée une narration visuelle captivante de l'évolution des champions au fil des éditions.

4.3.2 Visualisation des buts marqués

— Graphe à barres empilées décrivant les buts marqués par chaque nation pour les différentes édition du CAN

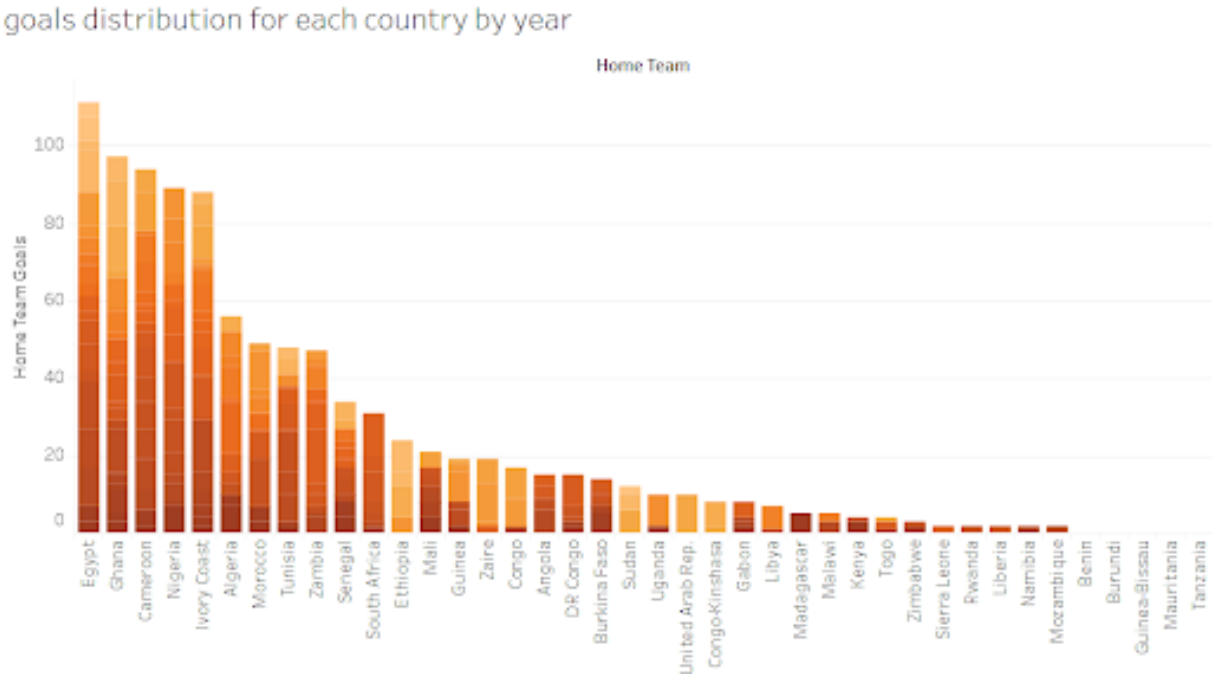

FIGURE 4.3 – Buts marques par chaque équipe dans la CAN

Ce graphe decrit les buts marqués par chaque équipe au cours des différentes éditions de la Coupe d'Afrique des Nations (CAN). Chaque barre incarne une nation, tandis que les segments empilés offrent une vue chronologique de ses performances au fil des années.

La hauteur de la barre représente le total des buts marqués par cette nation tout au long des différentes éditions, offrant une perspective globale des performances, et mettant l'Egypt en premiere place.

— Graphe en ligne représentant le total de buts marqués dans chaque édition de la CAN

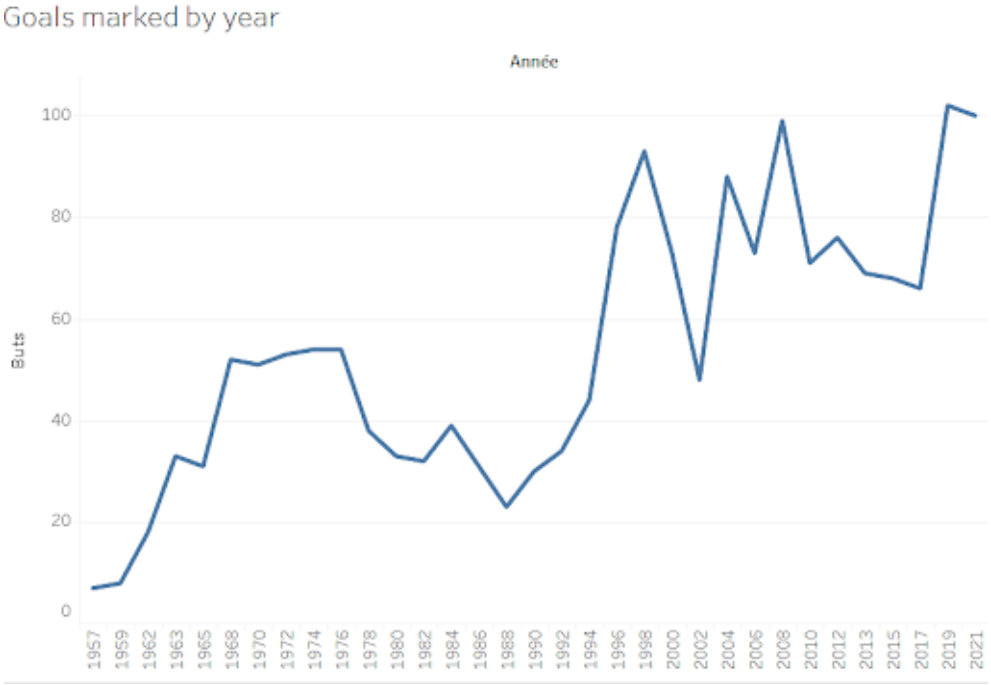

FIGURE 4.4 – Le total de buts marqués chaque année

Ce graphe donne une perspective temporelle sur l'évolution du nombre de buts marqués chaque edition de la Coupe d'Afrique des Nations (CAN). Cette ligne chronologique offre un voyage à travers les éditions du tournoi, mettant en lumière les moments où les filets ont tremblé et les records ont été établis.

4.3.3 Visualisation des performances de chaque nation

— Diagramme en arbre décrivant les points cumulés pour chaque équipe.

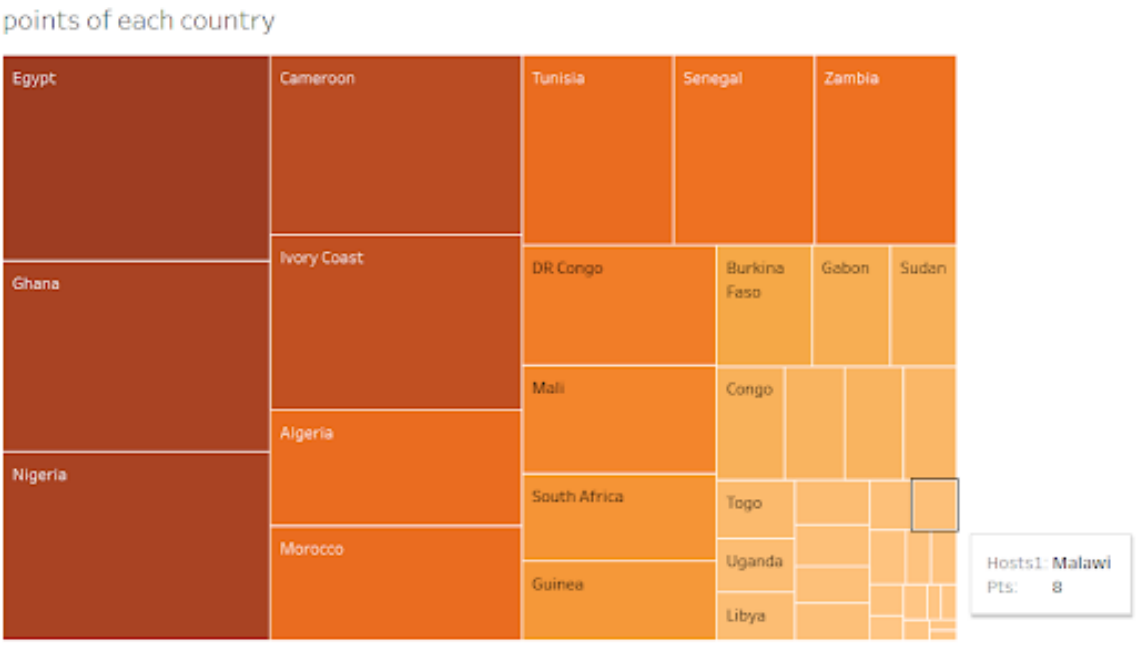

FIGURE 4.5 – Total de points de chaque nation

Ce diagramme (Treemap) offre une perspective unique sur les points cumulés marqués par chaque pays depuis les premières éditions de la Coupe d'Afrique des Nations (CAN).

Les blocs plus grands indiquent une présence durable et des performances solides(Egypt,Ghana, Nigeria..), tandis que les blocs plus petits signalent une contribution ponctuelle.

— Graphe à barres horizontales présentant le nombre de qualification au final pour chaque nation

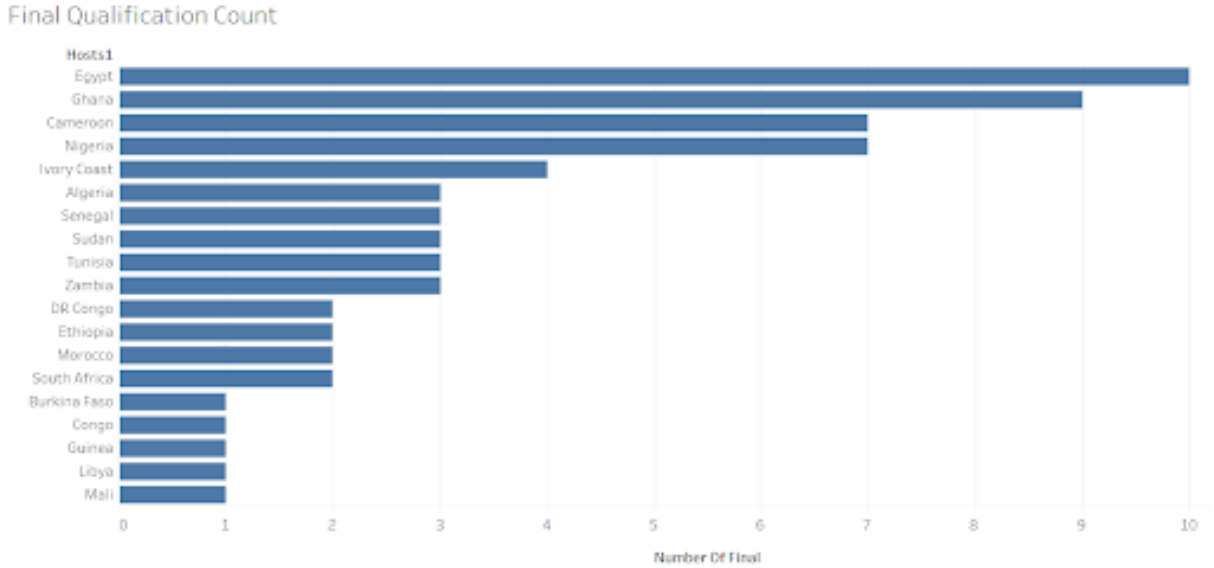

FIGURE 4.6 – Nombre de qualification au final pour chaque nation

Chaque barre représente une nation, offrant un aperçu visuel clair du nombre de fois qu'elle a brillamment atteint la phase finale du prestigieux tournoi.

Chaque barre horizontale est un témoignage des performances exceptionnelles d'une nation sur la scène de la CAN.

4.3.4 Visualisation d'une analyse comparative

— Histogramme décrivant le récapitulatif des résultats : Victoires, Défaites et Matchs Nuls par Équipe

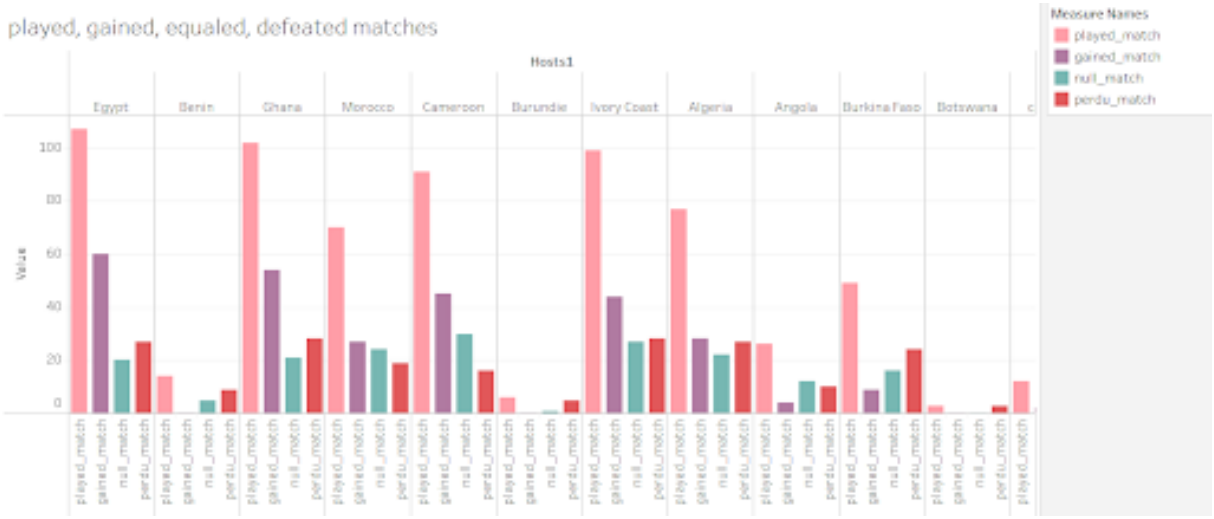

FIGURE 4.7 – Nombre de fois qu'une équipe a gagné,égalisé ou perdu un match

Le graphe résume les victoires, défaites et matchs nuls de chaque équipe. Il offre un aperçu précis des performances spécifiques de chacune, distinguant les équipes leaders en victoires, celles ayant subi le plus de défaites et celles obtenant fréquemment des résultats d'égalité.

Cette visualisation permet une comparaison directe des performances en termes de résultats de matchs, révélant les tendances et les performances globales des équipes tout au long de la compétition.

— Carte géographique montrant le pourcentage de victoire de toutes les nations de la CAN

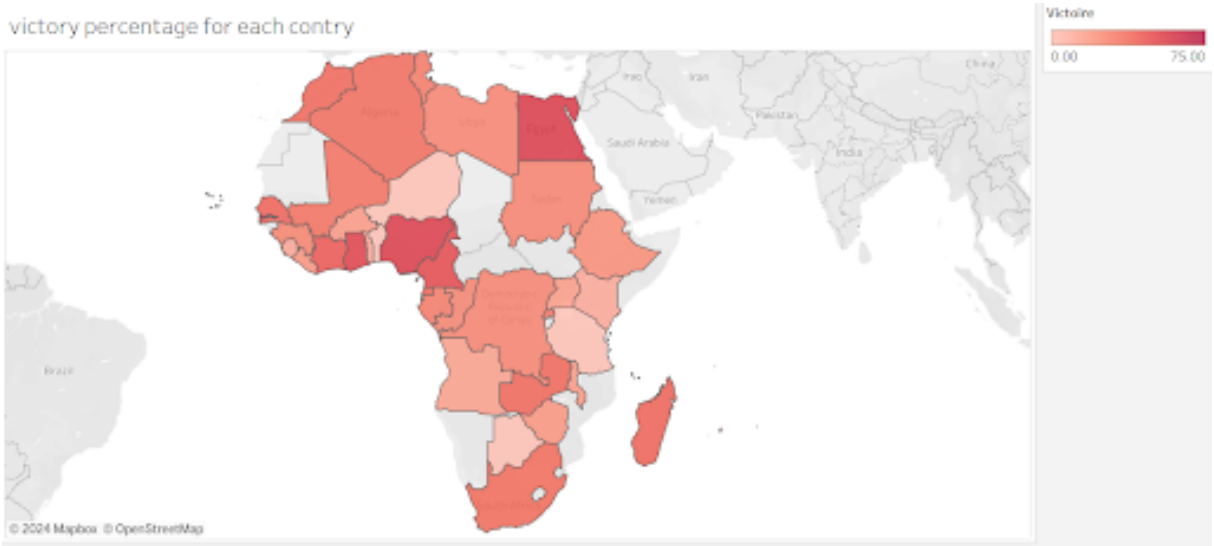

FIGURE 4.8 – Pourcentage de victoire de chaque equipe

Le graphe illustre le pourcentage de victoires de chaque équipe. Il met en avant les performances relatives des équipes, permettant d'identifier celles ayant le plus haut taux de succès. Cette visualisation offre une comparaison directe des performances globales de chaque nation durant la compétition.

4.3.5 Visualisation des performances de notre equipe nationale

— Barplot qui montre la distribution des positions de chaque joueur

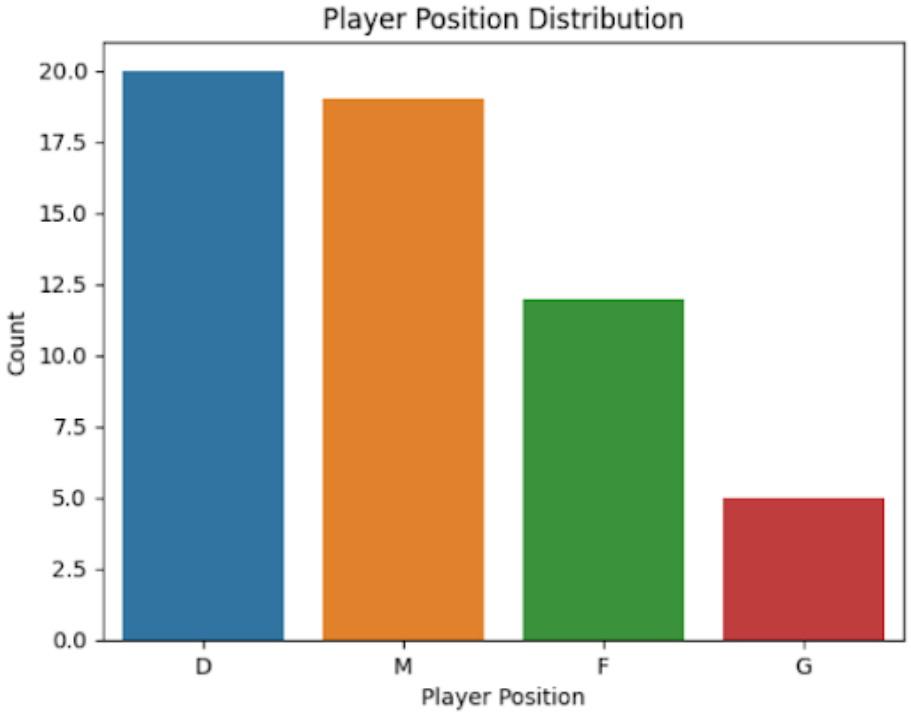

FIGURE 4.9 – Distribution des positions des joueurs

Le graphe montre la répartition des positions des différents joueurs marocains pouvant jouer avec l'équipe nationale, offrant une vue claire des postes occupés. Il met en lumière la diversité des rôles, illustrant la répartition des défenseurs, milieux de terrain, attaquants et autres postes clés.

— Graphique en double barres qui montre les meilleurs buteurs/ joueurs ayant des assistes

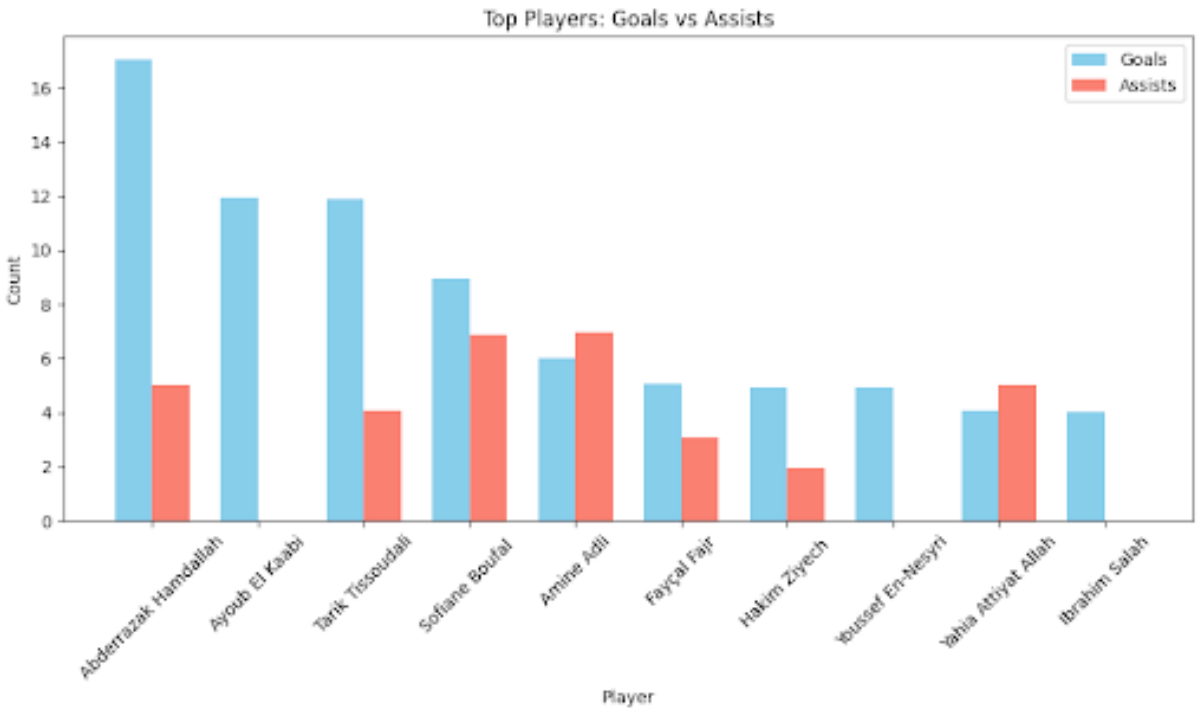

FIGURE 4.10 – Les meilleurs buteurs/ayant des assistes dans notre équipe nationale

Le graphique révèle les meilleurs buteurs et les joueurs les plus décisifs en termes d'assistances qui peuvent jouer dans l'équipe nationale. Il met en avant les performances individuelles des joueurs pour marquer des buts et créer des occasions.

Cette visualisation permet d'identifier les joueurs clés dans la création d'opportunités et de buts, offrant ainsi un aperçu des contributions individuelles à la performance globale de l'équipe.

— Spiderchart montrant les performances des gardiens de buts marocains

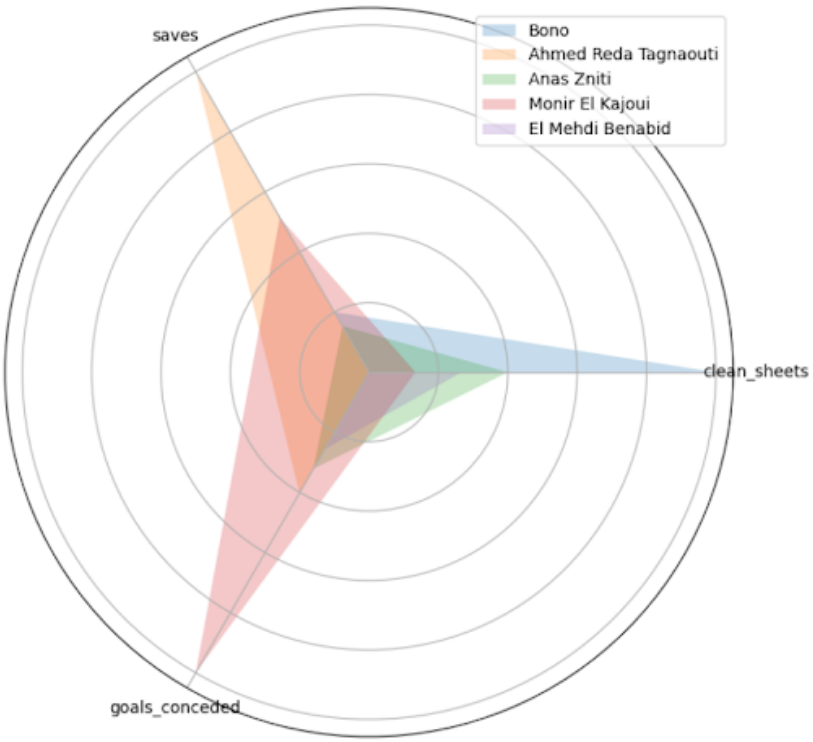

FIGURE 4.11 – Les performances de nos gardiens de but

le grapge présente les performances des gardiens de but marocains. Chaque joueur est représenté par une série de lignes reliant les différentes catégories de performance, telles que le nombre de buts encaissés, le nombre de matches sans encaisser de buts et le nombre d'arrêts.

Ce type de visualisation permet une analyse rapide et visuelle des forces et des faiblesses de chaque joueur, mettant en évidence leurs points forts dans certaines compétences spécifiques tout en identifiant les domaines où ils pourraient améliorer leurs performances.

Cette représentation graphique facilite également la comparaison entre les joueurs, permettant ainsi de prendre des décisions informées en matière de composition d'équipe ou d'évaluation des performances individuelles.

— Graphique de densité qui montre la distribution des performances des joueurs

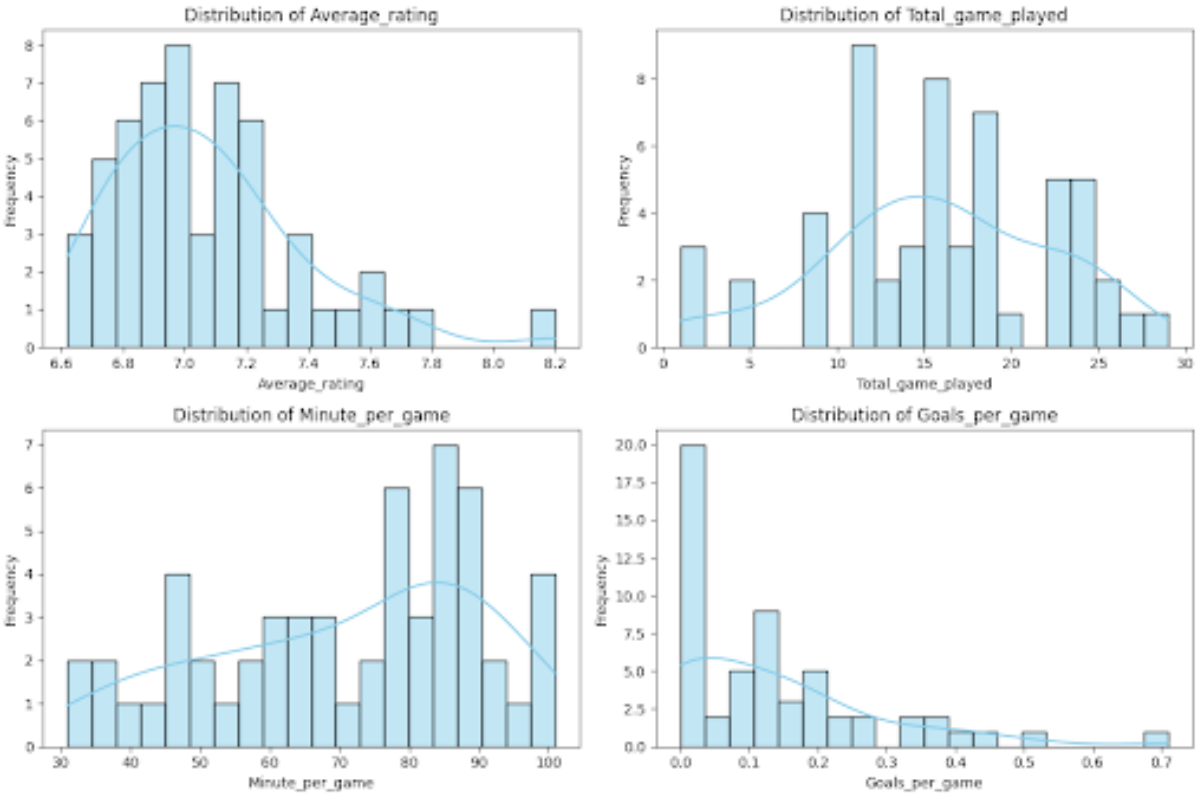

FIGURE 4.12 – Distribution des performances

Le graphique détaille les performances individuelles des joueurs à travers quatre critères clés : la note moyenne, le total de matchs joués, les minutes par match et les buts marqués. Cette visualisation met en lumière la moyenne de notation, la régularité dans les matchs disputés, l'implication en jeu par minute et la fréquence de marquage de buts.

Elle permet une comparaison directe des performances individuelles, offrant ainsi une vue complète des contributions des joueurs à l'équipe nationale.

4.4 Conclusion

Au terme de cette exploration dédiée à la Coupe d'Afrique des Nations de football, les résultats présentés à travers diverses visualisations et analyses dépeignent un panorama riche et informatif. Ces représentations visuelles offrent un aperçu des tendances émergentes, des performances des équipes et des dynamiques de victoires et de défaites.
Cette plongée approfondie dans les données de la CAN a non seulement éclairé sur les performances passées, mais peut également fournir des indications précieuses pour évaluer et anticiper les scénarios futurs de cette compétition prestigieuse.

Troixième partie
MLops

Chapitre 5

Approche Devops

Dans le domaine en évolution rapide du machine learning, l'efficacité et la reproductibilité des modèles sont primordiales. Notre projet a relevé ce défi en intégrant des pratiques DevOps innovantes, en particulier en utilisant Docker et GitHub.

5.1 Préparation à la Dockerisation

L'étape initiale de notre projet de dockerisation a été axée sur la création d'un environnement stable et reproductible pour chaque modèle de machine learning. Cette étape a nécessité une attention particulière aux détails pour garantir que les conteneurs Docker répondent précisément aux besoins de nos modèles.

5.1.1 Identification des Dépendances

Nous avons commencé par un examen minutieux des exigences spécifiques de chaque modèle. Cela incluait les versions de Python nécessaires, étant donné que différents modèles peuvent exiger des versions spécifiques pour fonctionner correctement. Nous avons également catalogué toutes les bibliothèques externes utilisées, telles que Sickit-learn, NumPy,et Pandas, en prêtant une attention particulière à leurs versions pour éviter les incompatibilités.

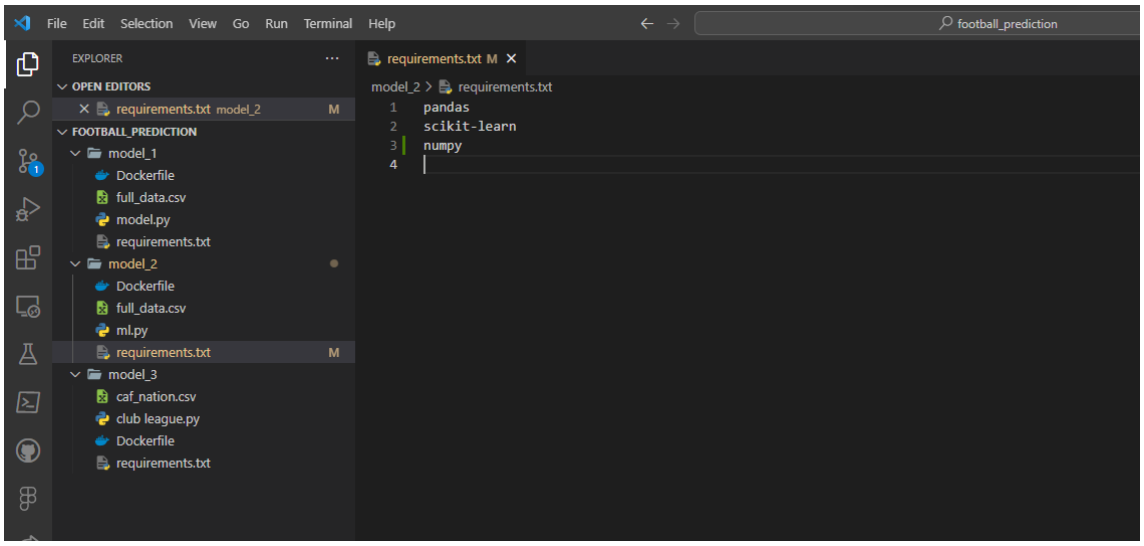

FIGURE 5.1 – Dépendances

5.1.2 Processus de Dockerisation

Pour assurer une gestion efficace et isolée de chaque modèle de machine learning, nous avons adopté une approche systématique pour la création des Dockerfiles. Chaque Dockerfile a été conçu pour répondre aux exigences spécifiques d'un modèle donné, garantissant ainsi que l'environnement de chaque conteneur Docker soit parfaitement adapté à ses besoins.

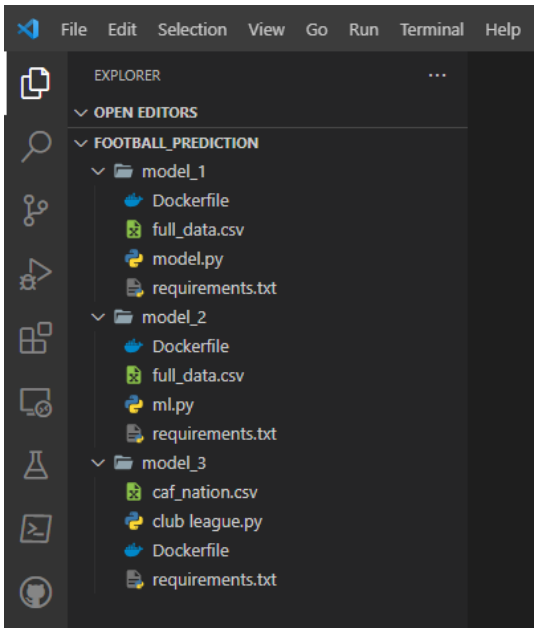

FIGURE 5.2 – Organisation des fichiers

5.1.3 Création des Dockerfiles

Nous avons débuté par la définition d'une image de base dans chaque Dockerfile, souvent une version spécifique de Python, correspondant à celle utilisée dans notre environnement de développement. Cela a été suivi par l'installation des bibliothèques nécessaires, telles que Sickit-Learn et Pandas, en s'appuyant sur le fichier requirements.txt que nous avions préparé. Pour garantir la précision, chaque commande d'installation a été scrupuleusement vérifiée pour éviter les problèmes de dépendance.

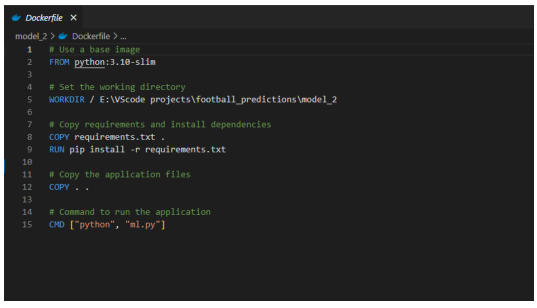

FIGURE 5.3 – Création Dockerfiles

5.1.4 Construction et Exécution des Images

Une fois les Dockerfiles finalisés, nous avons procédé à la construction des images Docker correspondantes. Cette étape a été réalisée en utilisant la commande docker build, suivie du

tag de l'image pour faciliter son identification et sa gestion. Après la construction, chaque image a été exécutée dans un conteneur Docker isolé en utilisant docker run. Cette exécution nous a permis de vérifier que chaque modèle fonctionnait comme prévu dans cet environnement nouvellement créé.

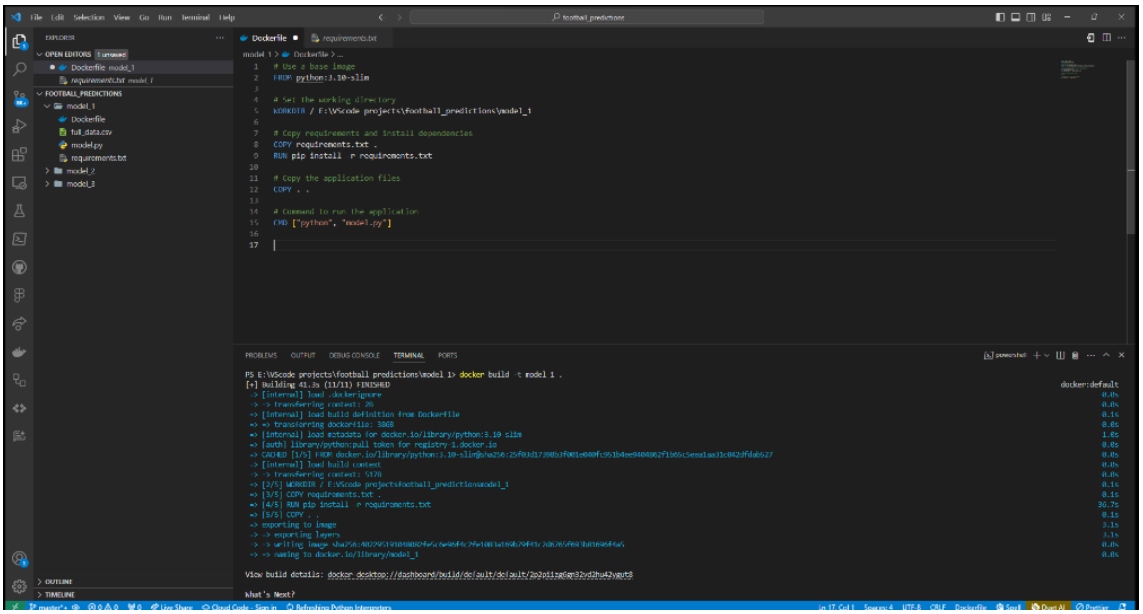

FIGURE 5.4 – Construction des images

```
PS E:\VScode projects\football_predictions> docker images
REPOSITORY                TAG         IMAGE ID       CREATED          SIZE
model_2                   latest      35b1292e473b   7 minutes ago    518MB
model_3                   latest      35b1292e473b   7 minutes ago    518MB
model_1                   latest      402295191048   3 hours ago      518MB
model                     latest      306555bc86cb   3 hours ago      507MB
<none>                    <none>      65537a747702   3 hours ago      507MB
<none>                    <none>      d5ae67b97ffd   4 hours ago      507MB
```

FIGURE 5.5 – Les Docker images des trois modèles

5.2 Utilisation de GitHub pour la Collaboration d'Équipe

En plus de notre travail avec Docker, un aspect crucial de notre projet a été l'utilisation de GitHub comme plateforme centrale pour la collaboration et la gestion du code. Nous avons créé un répertoire GitHub spécifique pour notre projet, qui a servi de noyau pour la coordination d'équipe et la collaboration efficace.

5.2.1 Création et Gestion du Répertoire

Nous avons commencé par établir un nouveau répertoire sur GitHub, conçu pour stocker non seulement nos Dockerfiles mais aussi les scripts de modèles de machine learning, les fichiers de données, et la documentation pertinente. Ce répertoire a été structuré de manière à faciliter la navigation et la compréhension, avec des dossiers et des fichiers clairement nommés et organisés.

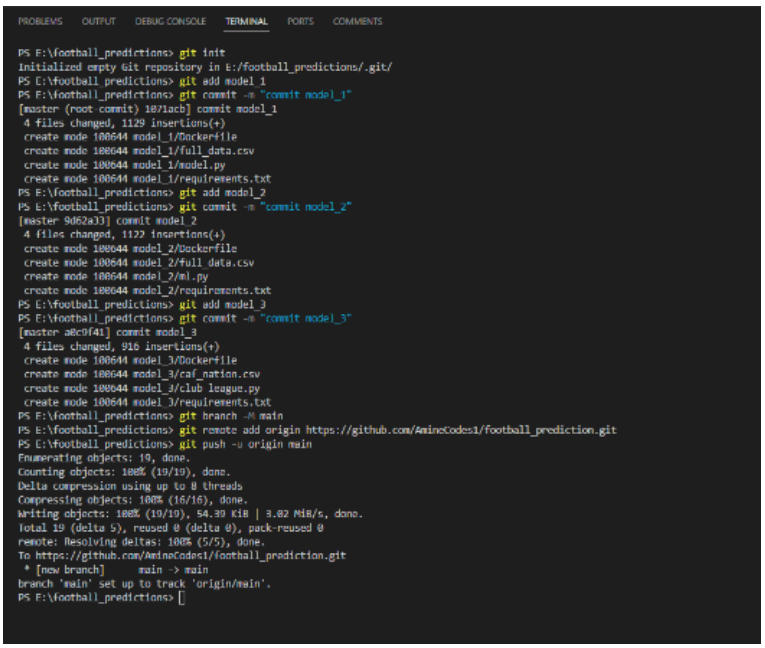

FIGURE 5.6 – Initialisation du répertoire en Github

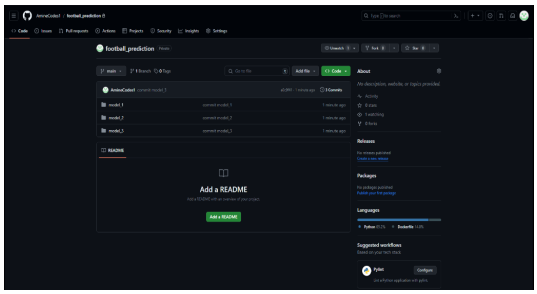

FIGURE 5.7 – Répertoire en Github

5.3 Conclusion

En conclusion, l'adoption de Docker pour encapsuler nos modèles de machine learning a représenté un pas important vers une gestion plus efficace et une meilleure standardisation des processus de développement et de déploiement. La combinaison de Docker et GitHub a non seulement simplifié la collaboration et le versionnage, mais a également augmenté la fiabilité et l'efficacité de notre projet de machine learning.

Conclusion Générale

En conclusion, notre projet intégrant la prédiction, la visualisation de données et l'approche MLOps pour la Coupe d'Afrique des Nations a été une opportunité enrichissante pour mettre en pratique les connaissances acquises tout au long de notre parcours. La qualité des données, provenant du machine learning, a été un élément clé dans l'élaboration de modèles de prédiction précis et fiables.

L'utilisation d'outils performants tels que Seaborn et Tableau pour la visualisation des données a permis la création de graphiques et de tableaux clairs, offrant une analyse approfondie des résultats. L'adoption de l'approche MLOps a joué un rôle essentiel dans la gestion efficace du cycle de vie des modèles, favorisant ainsi la robustesse et la reproductibilité des résultats.

Les résultats obtenus ont non seulement mis en lumière les tendances et les performances clés de chaque équipe pendant la Coupe d'Afrique des Nations 2024, mais ont également permis d'explorer diverses approches de prédiction. La comparaison des résultats a contribué à déterminer la méthodologie la plus adaptée à notre projet.

Nous espérons que ces résultats seront d'une utilité manifeste et d'un intérêt particulier pour tous les passionnés de la Coupe d'Afrique des Nations et pour ceux qui s'intéressent de près aux données de performance associées à cet événement majeur du football africain. Cette étude vise à enrichir la compréhension de l'évolution des performances sportives à travers une approche novatrice et méthodique.

Webographie

-**[Site officiel FIFA]** https ://www.fifa.com et https ://www.fifaratings.com/teams

-**[Nettoyage des données]** : https ://www.talend.com/fr/resources/what-is-data-cleansing/

-**[Sofascore]** https ://www.sofascore.com/

-**[Les étapes de réaliser un projet de visualisation]** https ://bluebearsit.com/power-bi

-**outils de MLOps**, https ://www.data-transitionnumerique.com/docker-tuto-complet/