

Analyse des Centres en Football :
Développement d'une Application Web pour
le Clermont Foot 63

Projet de 4ème année – Ingénierie

Mathématique et Data Science

Réalisé par : Nathan Talbot & Léo Bernardin

Année universitaire 2024-2025

Encadré par :

Rendu en :

Février 2025

Christophe de Vault

Sébastien Grillon

Résumé

L'objectif principal de ce projet était de développer une application web dédiée à l'analyse des centres effectués par les équipes de football. Pour ce faire, le club nous a fourni un ensemble de données footballistiques que nous avons dû traiter afin d'en extraire des informations exploitables.

Via ces données brutes à notre disposition, nous avons effectué divers traitements et analyses afin d'en extraire des informations utiles.

Les sujets des analyses ont été proposé par le Clermont Foot, afin de s'intéresser à des aspects spécifiques des centres réalisés par des équipes de Football.

Ces analyses ont ensuite été mises à disposition via une interface web offrant une visualisation intuitive et interactive, et permettant au staff du Clermont Foot d'avoir un usage pertinent des données footballistiques.

Mots clés : analyses de données, Football, SQLite, Streamlit, application web, Python, Pandas, Docker, GitHub.

Abstract

The main objective of this project was to develop a web application dedicated to analyzing crosses performed by football teams. To achieve this, the club provided us with a dataset containing football-related data, which we processed to extract actionable insights.

Using the raw data at our disposal, we performed various processing and analysis tasks to derive useful information. The topics of analysis were proposed by Clermont Foot, focusing on specific aspects of crosses made by football teams.

These analyses were then made accessible through a web interface offering an intuitive and interactive visualization, enabling the Clermont Foot staff to make effective use of football data.

Keywords: data analysis, football, SQLite, Streamlit, web application, Python, Pandas, Docker, GitHub.

Remerciements

À l'issue de ce projet, nous souhaitons adresser nos remerciements à toutes les personnes et structures qui ont contribué à sa réalisation.

Tout d'abord, nous remercions le Clermont Foot qui nous a fait confiance pour mener ce projet. Cela a été une expérience enrichissante et une grande chance de pouvoir disposer d'une base de données conséquente utilisant des mesures concrètes.

Nous remercions tout particulièrement Sébastien Grillon et Vincent Graillat, qui nous ont fourni les données nécessaires ainsi qu'une documentation précieuse pour avancer dans ce projet. Leur disponibilité, leurs retours constructifs et leur réactivité ont été essentiels pour nous permettre de progresser et d'apporter des améliorations concrètes à l'application.

Nous tenons également à remercier Christophe de Vault, notre professeur référent, pour son accompagnement tout au long de ce projet, grâce à des rendez-vous réguliers. Ses conseils avisés, son soutien constant et son suivi attentif nous ont permis de répondre aux attentes fixées.

Enfin, nous souhaitons exprimer notre gratitude envers Polytech, qui nous a offert cette opportunité de mener ce projet grâce à un partenariat avec le Clermont Foot, dans un cadre de travail enrichissant.

Table des matières

Introduction	5
I. Présentation du Clermont Foot.....	6
II. Présentation du projet.....	7
III. Outils utilisés.....	12
IV. Répartition du travail.....	15
V. Importation et stockage des données.....	16
VI. Données manquantes.....	20
VII. Transformation et exploitation des données.....	22
VIII. Application Streamlit	30
IX. Conclusion et perspectives	34
X. Bibliographie	35

Introduction

Ce projet a été réalisé dans le cadre de notre 4^e année à Polytech Clermont, en partenariat avec le club de football Clermont Foot 63. Il suit la tendance actuelle des clubs professionnels à utiliser les sciences des données pour analyser et améliorer les performances sportives.

Le football moderne repose de plus en plus sur l'analyse des données pour améliorer la performance des joueurs et déterminer les stratégies de jeu. Les clubs professionnels utilisent aujourd'hui des données variées, issues de statistiques événementielles (passes, tirs, interceptions...) et de tracking (positions des joueurs, vitesse, distance parcourue...). Ces informations permettent d'évaluer les qualités/défauts des joueurs et d'adapter les décisions tactiques en fonction des situations de jeu. Le Clermont Foot suit cette dynamique en cherchant à tirer parti des données pour améliorer son analyse des matchs. Notre projet vise à répondre à ce besoin en développant un outil d'analyse des centres, une action de jeu essentielle dans un match de football.

L'objectif principal était d'améliorer une application web existante, conçue par Nathan lors de son stage à l'été 2024, en y ajoutant de nouvelles fonctionnalités. La problématique centrale du projet est la suivante : Comment enrichir l'application existante pour qu'elle permette une meilleure exploitation des données footballistiques, en se concentrant sur les centres ? Pour y répondre, nous avons travaillé sur plusieurs axes : l'intégration des données de tracking (SkillCorner [1]) avec les données événementielles (StatsBomb [2]) et la création de nouveaux indicateurs. L'objectif final était de permettre aux analystes de mieux comprendre et exploiter les données afin d'améliorer la performance de l'équipe.

Notre approche a suivi une méthodologie structurée en plusieurs étapes : importation et nettoyage des données, structuration dans une base SQLite [3], transformation des données pour calculer des indicateurs spécifiques, et développement de l'application web via Streamlit [4]. L'application, déployée avec Docker [5], permet une visualisation des performances des joueurs et des équipes, facilitant ainsi la prise de décision lors des matchs.

Ce mémoire retrace le développement et l'amélioration de cette application web. Nous commencerons par présenter le contexte du projet et les données à disposition, ainsi que les objectifs initiaux. Nous détaillerons ensuite la gestion et l'exploitation des données, avant de présenter le développement de l'application Streamlit. Enfin, nous ferons le bilan des résultats obtenus et discuterons des perspectives d'amélioration. Nous mettrons également en lumière les défis rencontrés et les solutions apportées pour obtenir une application plus robuste et optimisée.

I. Présentation du Clermont Foot

Fondé en 1990 et basé à Clermont-Ferrand, le Clermont Foot 63 s'est progressivement imposé dans le football français. Actuellement en Ligue 2, le club s'est distingué par une approche réfléchie et structurée, où la recherche de performance se fait à travers une gestion stratégique et une vision à long terme.

Pour progresser, le club intègre des outils modernes, tels que l'analyse des données et les nouvelles technologies, afin d'optimiser les performances sur le terrain. Grâce à un recrutement astucieux et à un centre de formation de qualité, Clermont parvient à rivaliser avec des équipes aux moyens financiers plus conséquents, tout en consolidant son identité et son ambition.

L'histoire du club est marquée par une ascension progressive. Longtemps resté dans les divisions inférieures, le Clermont Foot a franchi un cap dans les années 2000 en s'installant durablement en Ligue 2. Après plusieurs saisons à lutter pour le maintien ou à viser le haut de tableau, le club a connu un tournant historique en accédant pour la première fois de son histoire à la Ligue 1 en 2021. Cette montée a marqué l'aboutissement de nombreuses années de travail et de structuration, confirmant le sérieux et l'ambition du projet clermontois.

II. Présentation du projet

1) Contexte de l'existant du projet (Stage de Nathan)

Durant son stage en 2024, Nathan a travaillé sur le traitement de données footballistiques afin d'en extraire des informations exploitables. Il a ensuite développé une application web pour les partager de manière interactive et accessible.

Cette application, conçue avec la bibliothèque Streamlit en Python, avait pour objectif de rendre les analyses compréhensibles même pour des personnes ne maîtrisant pas la programmation informatique. Elle était destinée aux membres du staff du Clermont Foot, notamment l'entraîneur et les analystes vidéo, afin de leur fournir des insights exploitables pour l'analyse des performances des équipes.

L'application initiale ne se limitait pas à l'étude des centres, mais proposait également des analyses sur d'autres métriques différenciant les équipes, comme la distance parcourue par match, le nombre de passes vers l'avant ou encore les zones de tir.

Toutefois, dans le cadre de notre projet, nous avons eu pour objectif de nous concentrer uniquement sur l'analyse des centres, en reprenant et améliorant cette partie du travail effectué par Nathan.

L'application développée lors du stage permettait d'analyser quatre aspects clés des centres :

- Les zones où ils sont effectués,
- Les zones où ils sont réceptionnés,
- La trajectoire des tirs réalisés dans un nombre défini d'événements suivant ces centres (ce nombre d'événements avait été défini à 5),
- Leur destination finale dans le but adverse.

Pour représenter ces données, plusieurs types de visualisations étaient intégrés à l'interface. Les zones d'exécution et de réception des centres étaient affichées sous forme de cartes de chaleur (Heatmap), tandis que la trajectoire des tirs était modélisée à l'aide de flèches sur un terrain de football stylisé.

La destination des tirs dans les cages adverses était, quant à elle, illustrée par des points positionnés sur une représentation graphique des buts.

Notre projet s'est donc appuyé sur cette base existante, avec pour objectif d'enrichir les analyses, d'optimiser les performances et d'améliorer l'expérience utilisateur afin de faciliter encore davantage l'exploitation des données par le Clermont Foot.

2) Présentation des données

Les données de StatsBomb sont événementielles, c'est-à-dire qu'elles contiennent des informations sur chaque événement clé du match, comme une passe, un tir, un corner, etc. Chaque événement est identifié de manière unique, et il contient des détails tels que le type de l'action, les coordonnées du joueur, ainsi que l'instant précis du match où cet événement a lieu. Les données de StatsBomb sont organisées en fichiers JSON, avec un fichier par match. En plus des événements, StatsBomb fournit des fichiers supplémentaires comme un fichier contenant des informations sur les alignements des équipes et des joueurs, ainsi que des identifiants d'équipes et de joueurs. Un autre fichier liste tous les matchs de la saison, précisant quelle équipe joue à domicile et quelle équipe joue à l'extérieur. Ces fichiers permettent d'obtenir une vue d'ensemble complète des événements d'un match ainsi que des équipes et des joueurs impliqués.

Les données de SkillCorner, quant à elles, sont de type "tracking" et concernent les positions des joueurs, leur vitesse et la distance parcourue à chaque instant du match. Ces données sont organisées en "frames", chaque frame représentant un instant précis du match. Le fichier principal de SkillCorner se concentre sur les centres, contenant des informations sur la position des joueurs avant, pendant et après chaque centre. Ces données sont particulièrement utiles pour analyser des situations spécifiques comme le nombre d'adversaires entre le ballon et le but, ou encore la position des joueurs qui peuvent potentiellement recevoir le centre.

3) Synchronisation des données

Le plus grand défi a été de synchroniser les données provenant de StatsBomb et de SkillCorner. En effet, bien que ces deux sources contiennent des informations complémentaires, elles utilisent des identifiants différents pour les matchs, les équipes et les joueurs, ce qui rend difficile le croisement des données.

Le problème réside dans le fait que chaque source attribue ses propres identifiants. Par exemple, StatsBomb utilise un identifiant pour chaque match, tandis que SkillCorner utilise un identifiant différent pour désigner les mêmes matchs. Le même problème se pose pour les équipes et les joueurs, qui ont des identifiants différents d'une source à l'autre. Par conséquent, il n'était pas

possible de simplement fusionner les données sans établir une correspondance entre ces identifiants.

Pour résoudre cette difficulté, nous avons utilisé des fichiers de "matching" fournis par le Clermont Foot. Un fichier permet de lier les identifiants des matchs de StatsBomb à ceux de SkillCorner. Un autre fichier fait le lien entre les identifiants des équipes dans les deux bases de données, et un troisième fichier associe les identifiants des joueurs de StatsBomb à ceux de SkillCorner. Grâce à ces fichiers, nous avons pu synchroniser les données événementielles de StatsBomb avec les données de tracking de SkillCorner, ce qui nous a permis de croiser les informations et d'effectuer des analyses détaillées sur les événements du match, tout en tenant compte des positions des joueurs.

Pour plus de détails, nous avons mis des liens renvoyant vers la documentation de SkillCorner et StatsBomb dans la bibliographie.

4) Organisation et gestion des données

Le projet s'appuie sur des données comportant 1 118 352 événements, répartis sur 306 matchs. Celles-ci présentaient des problèmes, notamment en termes de données manquantes ou incohérentes. Nous reviendrons sur cela plus tard dans le mémoire (cf. VI. Données manquantes). Une des exigences principales du projet était de rendre l'application facilement réutilisable et modifiable par d'autres développeurs. Pour cela, nous avons veillé à documenter chaque étape du processus, depuis l'importation des données jusqu'au déploiement de l'application. Nous avons également structuré le code, afin de permettre des ajustements futurs sans nécessiter une refonte complète.

5) Arborescence du projet

Pour réaliser ce projet, nous avons créé un repository GitHub [6] dédié. Afin d'optimiser son organisation et de respecter les différentes conventions de programmations, nous avons créé une arborescence épurée et intuitive.

a) Racine du projet

À la racine du projet, on retrouve les différents fichiers/dossiers propres à Git : « .gitignore », « README.me », Assets (contient les images utilisées dans le README.md), etc.

Il y a aussi le fichier « Dockerfile » permettant de créer l'image Docker de l'application Streamlit.

De plus, on y retrouve le dossier « Documentation » regroupant tous les rapports, diaporamas, etc.

Enfin, comme notre projet s'est divisé en deux principales tâches distinctes (traitement des données et développement de l'application web), nous avons décidé de créer deux sous-dossiers dédiés à ces parties.

Étant donné que nous avons codé en Python et que ces deux sections nécessitaient des bibliothèques Python différentes pour leur programmation, nous avons créé deux environnements virtuels distincts « env_traitement » et « env_app » pour respectivement le traitement des données et le développement de l'application Streamlit.

b) Dossier « Traitement donnees »

Ce sous-dossier correspond donc à la partie sur l'importation et la transformation des données, à qui on a d'ailleurs attribué deux sous-dossiers distincts « Importation_donnees » et « Transformation_donnees ».

Par ailleurs, le dossier dédié à l'importation des données contient le sous-dossier « Projet_centres_data » qui regroupe l'ensemble des données que le Clermont Foot avait mises à notre disposition.

De plus, nous avons créé le dossier « Donnees_manquantes » afin d'analyser les données manquantes dans l'ensemble des catégories de données à notre disposition (données de tracking, données événementielles « Event », etc.).

c) Dossier « Application streamlit »

Ce dossier contient les programmes de développement de l'application web. Le sous-dossier « .streamlit » permet de configurer cette dernière.

On y retrouve également les fichiers « variables.py » et « fonctions.py », qui contiennent respectivement les variables et fonctions génériques utilisées lors du développement de l'application.

Le fichier « Main.py » correspond au fichier principal de l'application : c'est lui qui est exécuté pour lancer cette dernière.

Le fichier « database.db » est la base de données résultantes de notre traitement des données. Elle contient les données utilisées par l'application.

Enfin, le dossier « pages » regroupe les différentes pages de l'application web. Pour la développer, nous avons créé un fichier Python par page.

III. Outils utilisés

1) Collaboration et gestion du projet avec GitHub

La collaboration entre nous deux a été grandement facilitée par l'utilisation de GitHub. Pour bien organiser notre travail, chaque membre de l'équipe a travaillé sur une branche dédiée, ce qui nous a permis de développer et tester nos fonctionnalités indépendamment. Une fois les développements terminés et vérifiés, nous avons fusionné les branches dans la branche principale (main) après avoir vérifié que tout fonctionnait correctement. Cela nous a offert de la flexibilité et une gestion efficace du code.

Afin d'éviter de surcharger le dépôt GitHub avec des fichiers volumineux, comme le dossier "Projet_centres_data" contenant les données brutes (fichiers JSON fournis par le Clermont Foot), nous avons utilisé un fichier « .gitignore ». Cela nous a permis d'exclure ces fichiers tout en maintenant une structure claire et légère du projet. L'arborescence du projet a été soigneusement organisée, avec des dossiers distincts pour les données, le code source, la documentation et les scripts, facilitant ainsi la navigation et la reprise du projet par d'autres développeurs.

Nous avons également créé un fichier README qui explique clairement les étapes à suivre pour prendre en main le projet. Il fournit des instructions détaillées pour lancer l'application via Docker, et guide sur la configuration des environnements virtuels, l'installation des dépendances, ainsi que l'organisation du projet pour faciliter son utilisation et sa modification.

2) Utilisation de Streamlit

L'application a été développée avec Streamlit, une bibliothèque Python qui permet de créer facilement des interfaces web interactives. Nous avons choisi Streamlit pour sa capacité à intégrer des visualisations dynamiques, ce qui correspondait parfaitement aux besoins des analystes du Clermont Foot. L'application offre des graphiques et des Heatmap interactifs pour visualiser les performances des équipes, rendant l'analyse des centres plus concrète.

3) Visual Studio Code (VS Code)

Pour ce projet, nous avons utilisé l'éditeur de code VS Code [7]. Son débogueur intégré a facilité le test et la correction du code, et sa personnalisation nous a permis d'ajouter des extensions, comme celle pour Git, simplifiant ainsi la gestion de version et la collaboration. De plus, il prend en charge Python et des formats comme JSON, ce qui a facilité le développement et la gestion des données.

4) Python et ses bibliothèques

Nous avons utilisé Python pour le développement, avec des bibliothèques comme Pandas [8] et Numpy [9] pour le traitement des données. Ces outils nous ont permis de manipuler facilement des dataframes et d'effectuer nos analyses de manière efficace.

5) Jupyter Notebook

Nous avons utilisé Jupyter Notebook [10] pour organiser notre code en cellules, ce qui a facilité la lecture et l'exécution. Les commentaires Markdown ont permis d'expliquer nos choix et démarches de manière claire.

6) Docker

Docker est un outil qui permet de créer et exécuter des applications dans des conteneurs, garantissant leur fonctionnement de manière identique sur n'importe quelle machine.

Nous avons utilisé Docker pour simplifier l'exécution de l'application sur différents systèmes.

7) Environnement virtuel Python

Pour gérer les dépendances de notre projet, nous avons mis en place un environnement virtuel Python. Cela a permis de créer un espace où nous pouvions installer les bibliothèques nécessaires sans risquer de rencontrer des problèmes de compatibilité entre les versions. Avec cet environnement, on s'est assurés que l'on utilisait les mêmes versions de bibliothèques. Pour rendre ça encore plus facile, nous avons créé un fichier requirements.txt. Ce fichier liste toutes les bibliothèques qu'il faut installer. Enfin, pour gérer les données du projet, nous avons choisi d'utiliser la bibliothèque Python SQLite. C'est une solution pratique pour notre besoin de centraliser les informations, comme les événements de match ou les actions des joueurs.

8) SQLite

SQLite nous a permis de stocker ces données dans une base de données relationnelle, ce qui a rendu leur manipulation et leur exploitation beaucoup plus faciles. De plus, étant directement intégrée à Python, elle ne nécessite pas d'installation ou de configuration complexe.

IV. Répartition du travail

Afin d'optimiser notre travail et de répondre le mieux possible aux attentes du Clermont Foot, nous nous sommes réparti les différentes tâches à effectuer durant ce projet.

De son côté, Nathan s'est occupé de :

- 1) La création/gestion du repository GitHub du projet : gestion des fichiers Git, des branches de travail, etc.
- 2) L'importation et le stockage SQLite des données (cf. V. Importation et stockage des données),
- 3) L'analyse des données manquantes (cf. VI. Données manquantes)
- 4) La création des environnements virtuels python
- 5) L'analyse du nombre d'adversaires entre le ballon et le but au cours des actions précédant des centres (cf. VII. Transformation et exploitation des données)
- 6) Le développement de l'application Streamlit (cf. VIII. Application Streamlit)
- 7) La conteneurisation de l'application via Docker.

Léo s'est quant à lui occupé de :

- 1) Elaboration des codes Python concernant la détection des différentes phases de jeu
- 2) Analyse des résultats concernant ces phases de jeu
- 3) L'analyse des données manquantes dans cette partie

V. Importation et stockage des données

1) Les données à notre disposition

a) Données initiales

Comme expliqué précédemment, nous avons accès à une variété de données fournissant des informations détaillées sur les joueurs, les équipes et le déroulement des matchs. Ces données nous ont été mises à disposition sous la forme d'un dossier intitulé "Projet_centres_data".

À l'intérieur de ce dossier, chaque type de donnée était organisé en sous-dossiers distincts. Par exemple, toutes les données associées aux événements des matchs étaient regroupées dans un dossier spécifique nommé "SB_events".

b) Format des données

L'ensemble des fichiers fournis était au format JSON. Leur organisation variait selon qu'ils étaient agrégés par match ou par saison. Par exemple, les fichiers contenant les événements de jeu "events" étaient structurés de manière à regrouper toutes les actions survenues lors d'un match donné, chaque match disposant ainsi de son propre fichier JSON. En revanche, d'autres fichiers, comme ceux relatifs aux rencontres de la saison "SB_matches", rassemblaient l'ensemble des données d'une saison en un seul document.

c) Spécification des données

Les données provenaient de deux fournisseurs distincts, chacun employant son propre système d'identification pour référencer les équipes, les joueurs et les matchs. Ainsi, un même club pouvait être désigné par des identifiants différents selon la source de données (SkillCorner ou StatsBomb). Afin d'assurer la correspondance entre ces références, le Clermont Foot nous a fourni des tables de correspondance (tables de matching) permettant d'établir un lien entre les identifiants des deux fournisseurs, notamment pour les équipes, les matchs et les joueurs (cf. image 1).

team_id_SB	team_id_SKC
168	85
144	66
164	69
136	70
165	72
130	74
143	78
156	93
141	97

Image 1 - Table de matching des équipes 1

d) Choix du format de stockage des données (SQLite)

Étant donné que les données initiales étaient stockées sous forme de fichiers JSON, leur exploitation nécessitait de charger et traiter un grand nombre de fichiers à chaque exécution du programme. Cette approche s'avérait inefficace et chronophage, c'est pourquoi nous avons opté pour la création d'une base de données SQLite. Cette base de données, que nous avons nommée "raw-database.db", a été conçue pour centraliser et structurer les données brutes dans des tables distinctes, simplifiant ainsi leur interrogation et leur manipulation.

Pour interagir avec cette base de données, nous avons utilisé la bibliothèque SQLite3 [11] de Python, qui nous permet d'exécuter des requêtes SQL et de manipuler facilement les tables de données.

De plus, afin de faciliter la visualisation du contenu de la base de données, nous avons utilisé l'API SQLite-web [12], offrant une interface graphique permettant d'accéder aux informations directement depuis un navigateur web.

2) Processus d'importation des données

Afin d'automatiser l'importation des données dans la base SQLite, nous avons développé un programme sous Jupyter Notebook. Ce programme se chargeait de parcourir les fichiers JSON fournis, d'extraire les données pertinentes et de les insérer dans la base de données. Chaque type de donnée a été stocké dans une table dédiée, ce qui permettait d'accéder facilement aux informations sans avoir à lire chaque fichier JSON séparément.

L'importation des données s'est déroulée en plusieurs étapes. Pour les fichiers contenant des données réparties par match, nous avons adopté une approche d'agrégation, fusionnant les différentes entrées en une seule table afin d'en faciliter l'exploitation.

En effet, au lieu d'avoir un fichier pour chaque match d'un type de données en particulier, nous n'avions plus qu'une seule table dans la base de données, contenant la fusion des données de l'ensemble des matches à dispositions.

Le programme d'importation, nommé "importation.ipynb" et localisé dans le dossier "Importation_donnees", a été conçu sous Jupyter Notebook afin de faciliter la compréhension du code grâce aux cellules Markdown permettant d'expliquer chaque étape du processus.

Pour manipuler les fichiers JSON et interagir avec la base de données, nous avons utilisé la bibliothèque Pandas, qui offre une gestion efficace des données tabulaires et permet notamment d'écrire directement dans une base SQLite grâce à la fonction "to_sql".

Lors de la lecture des fichiers JSON, nous avons utilisé la commande "pd.read_json(chemin_du_fichier)" afin de convertir les données en dataframes Pandas.

P.S. : Un dataframe est un objet conçu par Pandas pouvant être vu comme une table d'une base de données relationnelles.

Une fois les fichiers chargés, nous avons appliqué des traitements spécifiques avant de les insérer dans la base de données SQLite. Par exemple, certains fichiers nécessitaient un renommage des colonnes ou une transformation des données avant leur insertion.

L'un des défis majeurs rencontrés concernait le stockage de certaines variables complexes, notamment celles contenant des listes ou des dictionnaires. Par exemple, la colonne "location" des données "events" contenait des coordonnées sous forme de listes [x, y, z], correspondant à la position des événements recensés par Stats Bomb.

Étant donné que SQLite ne permet pas de stocker directement des valeurs de type liste ou dictionnaire, nous avons dû adapter la structure en créant trois colonnes distinctes "x_loc", "y_loc" et "z_loc" pour contenir ces valeurs.

De même, pour les données "Freeze frames", la colonne "data" contenait des listes de dictionnaires représentant les informations de tracking de chaque joueur sur le terrain à un instant (frame) donné. Afin d'exploiter ces données dans la base de données, nous avons créé un dataframe distinct contenant une ligne par joueur et une colonne pour chaque information de tracking, tout en assurant la liaison avec les identifiants des événements correspondants.

Enfin, les données "lineup" ont nécessité un traitement similaire. Chaque ligne de la colonne "lineup" contenait une liste de dictionnaires représentant les joueurs d'une équipe. Nous avons

extrait ces informations afin de structurer les données sous forme de table, associant chaque joueur à son équipe et au match concerné.

L'ensemble de ces traitements nous a permis d'optimiser l'importation et le stockage des données, garantissant une exploitation efficace dans l'application web développée ultérieurement.

VI. Données manquantes

1) Analyse des données manquantes

a) Freeze frames

La table `freeze_frames` contient des informations sur les événements précédant, succédant et durant chaque centre. Certaines valeurs sont absentes, mais elles ont une explication logique. Par exemple, les colonnes `timestamp` et `period` sont NaN pour les événements de début et de fin de période. De même, la colonne `group`, qui indique si une frame appartient à l'équipe à domicile ou à l'extérieur, est parfois vide. Cela arrive lorsque le ballon est en dehors du terrain, ce qu'on peut aussi voir avec les coordonnées `x` et `y`. Enfin, la colonne `trackable_object`, qui identifie le joueur en possession du ballon, est NaN quand personne ne le détient, ce qui correspond aux moments où le ballon est libre.

b) Matching matches

Cette table permet de comparer les correspondances entre SkillCorner et StatsBomb. Côté SkillCorner, aucune valeur n'est manquante, alors que quelques manquements sont présents pour StatsBomb. Il y a un match (id 1160044) identifié par SkillCorner qui n'a pas de correspondance chez StatsBomb. À l'inverse, tous les matchs StatsBomb sont bien présents chez SkillCorner. Tous les matchs de `freeze_frames` existent bien dans `matching_matches`, mais trois matchs sont présents dans `matching_matches` sans être dans `freeze_frames`. On trouve aussi un match présent dans `events` mais absent de `matching_matches`, et inversement. Malgré ces quelques incohérences, cela reste négligeable par rapport aux 306 matchs fournis par le Clermont Foot.

c) Matching players

Dans la table des joueurs, certaines informations sont absentes, notamment les dates de naissance. On remarque aussi que 24 joueurs ont un id dans SkillCorner mais pas dans StatsBomb, alors que l'inverse n'arrive jamais. En regardant de plus près, on voit que ces joueurs ont en moyenne 19,2 ans, et étaient donc encore plus jeunes au moment de la capture des données. Il est probable que StatsBomb n'ait pas encore enregistré ces joueurs, car ils venaient tout juste de devenir professionnels.

d) Matching teams

Ici, aucun problème à signaler. Toutes les équipes de StatsBomb ont une correspondance chez SkillCorner, et vice-versa. Il n'y a donc pas de valeur manquante ni d'incohérence dans cette table.

e) SB matches

Seule une valeur manque : l'affluence du stade (attendance) pour un match. Ce n'est pas un problème important car toutes les autres données sont complètes.

f) Tracking data

Cette table contient les données de vitesse et d'accélération des joueurs et du ballon. On voit que la vitesse et l'accélération du ballon (trackable_object = 55) sont toujours absentes. Pour les joueurs, certaines valeurs manquent aussi, mais il est difficile de savoir exactement quand et pourquoi. D'après la documentation de SkillCorner, ces données ne sont calculées que si au moins cinq frames successives sont disponibles pour un joueur, ce qui pourrait expliquer ces absences.

2) Conclusion

Même si certaines valeurs sont manquantes, elles sont souvent justifiées par la structure des données. Les absences liées aux jeunes joueurs ou aux calculs de vitesse ne remettent pas en cause l'analyse globale des matchs.

VII. Transformation et exploitation des données

1) Étude du nombre d'adversaires entre le ballon et le but adverse avant un centre

L'une des missions confiées par le Clermont Foot consistait à analyser le nombre d'adversaires situés entre le ballon et le but lors des actions précédant un centre.

Autrement dit, pour chaque centre réalisé par une équipe, l'objectif était de déterminer combien de joueurs adverses se trouvaient derrière le ballon au moment de l'événement précédent le centre (« event » au sens des données StatsBomb).

a) Données utilisées

Cette analyse nécessitait de croiser deux sources de données : celles de SkillCorner et celles de StatsBomb.

Les données "event" de StatsBomb ont été utilisées pour identifier les événements précédant chaque centre, car elles fournissent la liste détaillée des actions survenues durant un match (passes, tirs, touches, etc.).

Les données "freeze frame" de SkillCorner ont permis d'obtenir la position exacte des 22 joueurs et du ballon à chaque instant (frame) capturé.

Pour relier ces deux types de données, le Clermont Foot avait préalablement associé à chaque frame de SkillCorner un identifiant d'événement (event_id) provenant de Stats Bomb. Cette correspondance nous a permis d'exploiter les positions des joueurs au moment précis des événements précédant les centres.

b) Traitement des données

L'analyse a été réalisée dans un notebook Jupyter intitulé « nb_advers_ballon_but.ipynb », en suivant une méthodologie reposant sur plusieurs étapes successives.

Dans un premier temps, nous avons filtré les frames afin de ne conserver uniquement celles correspondant aux événements précédant un centre.

Une fois cette sélection effectuée, nous avons identifié l'équipe réalisant le centre en nous basant sur les données event de StatsBomb.

Pour chaque frame associée à un événement précédant un centre, nous avons ainsi attribué l'équipe ayant effectué le centre en question. Cela nous a permis, pour chaque événement précédant un centre, de connaître l'équipe responsable du centre qui suivait.

Ensuite, nous avons procédé à un nouveau filtrage en ne conservant que les joueurs appartenant à l'équipe adverse. Cette sélection nous a permis de nous concentrer uniquement sur les adversaires et d'analyser leur positionnement par rapport au ballon et leurs cages.

Enfin, nous avons déterminé le nombre de joueurs adverses situés entre le ballon et le but au moment de chaque frame analysée. Pour cela, nous avons exploité les coordonnées des joueurs et du ballon fournies par SkillCorner, ce qui nous a permis de comptabiliser précisément les joueurs adverses se trouvant derrière le ballon dans chaque situation étudiée.

c) Problématiques et difficultés rencontrées

L'un des premiers défis de cette analyse a été d'associer les données issues de StatsBomb et de SkillCorner, car ces deux sources de données utilisent des formats et des identifiants différents pour représenter les événements, les joueurs, les matches et les équipes.

Le Clermont Foot avait déjà réalisé un premier travail de synchronisation en attribuant à chaque frame SkillCorner un identifiant d'événement (`event_id`) correspondant dans les données StatsBomb. Cette correspondance nous a permis de relier les positions des joueurs aux actions du jeu.

Cependant, au-delà du simple appariement entre frames et événements, un problème plus complexe s'est posé : les identifiants des joueurs, des matches et des équipes n'étaient pas les mêmes pour les deux fournisseurs de données.

Dans les données "event" de StatsBomb, chaque joueur et chaque équipe est identifié par un ID propre à cette source de données.

Dans les données "freeze frame" de SkillCorner, les joueurs sont simplement référencés par des numéros et ne possèdent pas directement d'association avec leur équipe.

Nous avons donc dû utiliser une troisième source de données, les données « lineup » de StatsBomb, qui contiennent, entre autres, la liste des joueurs participant à chaque match ainsi que leur équipe respective. Grâce à ces données, nous avons pu établir les correspondances

nécessaires entre les joueurs présents dans les frames de SkillCorner et leur identification dans StatsBomb.

Ce travail a nécessité plusieurs étapes de traitement :

- i) Extraire, pour chaque match, la liste des joueurs et de leur équipe à partir des lineups de StatsBomb.
- ii) Associer ces informations aux frames de SkillCorner pour identifier l'appartenance des joueurs à leur équipe.
- iii) Vérifier la cohérence des données en nous assurant que chaque joueur était bien associé à la même équipe dans toutes les frames où il apparaissait.

Une fois cette correspondance établie, nous avons pu distinguer les joueurs adverses et ainsi compter ceux qui se trouvaient entre le ballon et le but lors de l'événement précédant le centre.

La seconde difficulté a été l'interprétation des coordonnées au sens de SkillCorner.

SkillCorner modélise le terrain sous la forme d'un rectangle de 105m x 68m, où l'origine du système de coordonnées (0,0) correspond au centre du terrain (cf. image 2).



Image 2 – Modélisation du terrain par SkillCorner

Cela signifie que :

- L'axe X représente la longueur du terrain, avec des valeurs positives et négatives de part et d'autre du milieu de terrain.
- L'axe Y représente la largeur du terrain, avec une valeur positive vers le haut et négative vers le bas du terrain.

Cependant, l'un des problèmes majeurs que nous avons rencontrés concernait l'orientation du terrain, car les équipes changent de camp entre la première et la seconde mi-temps.

Pour déterminer si un joueur adverse était placé entre le ballon et son propre but, il était essentiel de savoir dans quelle direction attaquait et défendait chaque équipe. Or, SkillCorner ne fournit pas directement cette information.

Nous avons donc utilisé les données « SKC_matches_data » du Clermont Foot, qui indiquaient pour chaque match :

- Quel camp l'équipe à domicile attaquait en première et en seconde mi-temps
- Par déduction, quel camp l'équipe adverse défendait à chaque période

À partir de ces données, nous avons pu standardiser l'orientation des matchs en considérant systématiquement que toutes les équipes attaquaient de gauche à droite (comme le fait Stats Bomb).

Concrètement, cela impliquait que :

- Si une équipe attaquait de droite à gauche, nous avons inversé toutes ses coordonnées pour qu'elles correspondent à une attaque de gauche à droite. C'est-à-dire nous avons multiplié par -1 les coordonnées de tous ses joueurs.
- Les positions des joueurs adverses ont également été ajustées pour que leur position relative au ballon et au but soit correcte dans ce référentiel standardisé.

Exemple d'application :

Si une équipe attaquant normalement de droite à gauche réalisait un centre en (-30,5), nous avons converti ces coordonnées en (30,-5).

Une fois toutes les équipes ramenées à un référentiel où elles attaquent de gauche à droite, nous avons pu compter facilement les adversaires situés à droite du ballon, ce qui correspondait aux défenseurs entre le ballon et le but.

Ce travail de normalisation a permis d'éviter les erreurs d'interprétation liées aux changements de côté entre les mi-temps et de garantir une analyse cohérente pour tous les matchs.

Le processus final a donc consisté à :

- Identifier, pour chaque frame précédant un centre, l'équipe qui a centré et sa direction d'attaque.
- Récupérer la position du ballon et celle de tous les joueurs adverses.
- Compter le nombre de joueurs adverses dont la coordonnée X était supérieure ou égale à celle du ballon.

Ce travail a permis d'obtenir une mesure fiable du nombre de défenseurs adverses placés entre le ballon et le but au moment précédant un centre, offrant ainsi au Clermont Foot une statistique pertinente pour évaluer les conditions dans lesquelles leurs joueurs effectuent leurs centres.

Cette analyse a nécessité un travail minutieux d'intégration et de transformation des données afin de croiser différentes sources (« event », « freeze frame », « SKC_matches_data », etc.) et de créer un référentiel cohérent.

En surmontant ces défis, nous avons pu extraire une information précieuse sur la densité défensive rencontrée avant les centres, qui pourra être utilisée par le staff du Clermont Foot pour affiner ses analyses tactiques.

Pour une compréhension plus détaillée des aspects techniques et du code implémenté, nous recommandons de consulter le fichier « nb_advers_ballon_but.ipynb ».

2) Étude des phases de jeu reliées aux centres

Dans cette partie du projet, nous avons cherché à identifier et à analyser deux types de phases de jeu spécifiques dans les matchs de football : l'Attaque Placée Numéro 1 et la Transition Numéro 1. L'objectif était de déterminer dans quelle phase de jeu chaque centre a été effectué, afin de comprendre les tendances et les comportements des équipes durant ces phases. Cette analyse permet de mieux cerner les contextes dans lesquels les centres sont utilisés et d'en tirer des conclusions tactiques.

a) Attaque placée Numéro 1

L'Attaque Placée Numéro 1 correspond à une séquence de jeu qui débute par une remise en jeu, comme un corner, un coup franc, une touche ou une remise en jeu standard, et se termine par une perte du ballon ou un arrêt du jeu. Cette phase est caractéristique des situations où une équipe tente de construire une attaque de manière organisée, en utilisant des schémas de jeu prédéfinis pour déstabiliser la défense adverse.

Pour identifier cette phase, nous avons d'abord ajouté une colonne à la table des événements dans la base de données. Cette colonne permet de marquer les événements appartenant à cette phase. Si la colonne existait déjà, nous l'avons réinitialisée à 0 pour éviter toute confusion. Ensuite, nous avons défini deux ensembles d'événements considérés comme des remises en jeu : ceux qui incluent des types de passes comme les corners, les coups francs et les touches, et les événements spécifiques comme le "Referee Ball-Drop".

Nous avons ensuite parcouru tous les événements de la table pour détecter les phases d'Attaque Placée Numéro 1. Une phase commence lorsqu'une remise en jeu est détectée et se termine lorsque la possession change ou que le match s'arrête. Les événements appartenant à cette phase sont marqués avec une valeur spécifique dans la colonne phase de jeu.

Au total, 576 940 événements ont été marqués comme appartenant à la phase d'Attaque Placée Numéro 1, ce qui représente 51.59 % de tous les événements analysés. Parmi les centres, 59.6 % ont été effectués durant cette phase, ce qui montre que les centres sont fortement associés aux attaques placées. Cela suggère que les équipes utilisent souvent les centres lors de ces phases organisées.

Les centres sont plus fréquents durant les phases d'Attaque Placée Numéro 1 que dans la moyenne des événements. Cela indique que les équipes privilégient les centres dans des situations où elles ont le temps de construire leur attaque de manière structurée, plutôt que dans des phases de jeu plus rapides ou désorganisées.

b) Transition Numéro 1

La Transition Numéro 1 correspond à une séquence de jeu qui débute par une récupération du ballon et se termine si l'équipe reperd le ballon dans les 4 événements suivants, hors événements de l'équipe adverse. Cette phase est typique des situations de contre-attaque ou de récupération rapide.

Pour identifier cette phase, nous avons d'abord défini les événements pertinents pour la phase de Transition Numéro 1. Nous avons sélectionné uniquement les événements directement liés au jeu, en excluant les actions défensives comme les pressions ou les changements tactiques. Les événements pertinents incluent les passes, les dribbles, les récupérations de ballon, etc. Ensuite, nous avons parcouru tous les événements de la table pour détecter les phases de Transition Numéro 1. Une phase commence lorsqu'un changement de possession est détecté, sans qu'il s'agisse d'une remise en jeu, et se termine si l'équipe reperd le ballon dans les 4 événements suivants.

Au total, 14 340 événements ont été marqués comme appartenant à la phase de Transition Numéro 1, ce qui représente 1.28 % de tous les événements analysés. Parmi les centres, seulement 0.33 % ont été effectués durant cette phase, ce qui est bien inférieur à la moyenne. Cela montre que les centres sont très peu utilisés dans les transitions rapides.

Les centres sont très peu représentés dans les phases de Transition Numéro 1. Cela suggère que les équipes privilégient d'autres types d'actions, comme les passes rapides ou les dribbles, lors des transitions, plutôt que les centres. Les centres nécessitent en effet plus de temps et de précision, ce qui les rend moins adaptés aux phases de jeu rapides.

c) Comparaison des Phases de Jeu

Lorsque l'on compare les deux phases de jeu, on observe des différences significatives dans la répartition des centres. L'Attaque Placée Numéro 1, qui représente 51.59 % des événements (576629 événements), est fortement associée aux centres, avec 59.6 % des centres effectués

durant cette phase. En revanche, la Transition Numéro 1, qui ne représente que 1.28 % des événements (14340 événements), est très peu associée aux centres, avec seulement 0.33 % des centres effectués durant cette phase.

Cette différence s'explique par la nature des deux phases. L'Attaque Placée Numéro 1 est une phase organisée, où les équipes ont le temps de construire leur attaque et de préparer des centres pour déstabiliser la défense adverse. En revanche, la Transition Numéro 1 est une phase rapide et désorganisée, où les équipes cherchent à profiter du déséquilibre de l'adversaire pour créer des occasions de but. Dans ce contexte, les centres sont donc moins fréquents.

Nous avons également vérifié qu'aucun événement n'appartenait à la fois à la phase d'Attaque Placée Numéro 1 et à la phase de Transition Numéro 1. Cette vérification confirme la cohérence de notre analyse et montre que les deux phases sont bien distinctes.

1	<code>select count(*)</code>
2	<code>from events</code>
3	<code>where phase_jeu = 1</code>

	count(*)
1	576629

1	<code>select count(*)</code>
2	<code>from events</code>
3	<code>where phase_jeu = 2</code>

	count(*)
1	14340

Image 3 & Image 4 – Nombre d'événements dans chaque phase

En effet, ces deux requêtes (cf. Image 3 & Image 4) ont été effectuées après avoir exécuté la recherche des deux phases l'une après l'autre, en gardant le marquage de la première phase pour chaque événement avant de faire le marquage de la deuxième phase. On retombe bien sur le nombre d'événements par phase cohérent avec le marquage des phases avec remise à zéro entre les deux marquages. Il n'y a donc aucun événement qui est considéré comme étant dans la phase 1 et phase 2. Le résultat est donc cohérent à ce niveau-là.

d) Implications tactiques

Les résultats de cette analyse ont des implications tactiques importantes pour les équipes. Tout d'abord, les centres sont importants pour créer des occasions de buts lors des phases d'Attaque Placée Numéro 1. Les équipes peuvent donc travailler sur des schémas de jeu spécifiques pour maximiser l'efficacité des centres dans ces phases. Par exemple, elles peuvent entraîner des

combinaisons de passes pour créer des espaces et permettre aux joueurs de centrer dans des positions avantageuses.

En revanche, les centres sont peu utilisés lors des phases de Transition Numéro 1. Cela suggère que les équipes privilégient des actions plus directes, comme les passes rapides ou les dribbles, pour profiter de l'effet de surprise et créer des occasions de but. Les entraîneurs peuvent donc se concentrer sur l'entraînement de ces actions pour améliorer l'efficacité de leur équipe lors des transitions rapides.

e) Limites et perspectives

Cette analyse présente certaines limites. Tout d'abord, la définition des phases de jeu repose sur des règles simples, comme le changement de possession ou le nombre d'événements. Des règles plus complexes, prenant en compte des facteurs supplémentaires comme la position des joueurs ou le contexte tactique, pourraient affiner l'analyse.

Cette analyse représente certaines limites. En effet, cette dernière aurait été plus complète si les deux autres phases suggérées par le Clermont Foot avaient été implémentées correctement. Malheureusement, je n'ai pas réussi à les intégrer dans cette étude, ce qui limite la portée des conclusions.

En résumé, cette analyse montre que les centres sont principalement associés aux phases d'Attaque Placée Numéro 1, tandis qu'ils sont très rares dans les transitions rapides.

VIII. Application Streamlit

1) Principe de l'application

Dans cette section, nous ne détaillerons pas l'implémentation technique ni le code de l'application web développée avec Streamlit, afin d'éviter une surcharge d'informations techniques. Cette bibliothèque Python, bien que puissante et intuitive, nécessite une approche détaillée pour son utilisation, ce qui pourrait nuire à la clarté du propos.

Nous nous concentrerons donc sur le fonctionnement global de l'application ainsi que sur ses fonctionnalités principales.

L'application repose sur une interface unique, regroupant l'ensemble des analyses menées sur les centres effectués au cours des matchs. Elle s'appuie sur une première version développée par Nathan lors de son stage, qui a servi de base à cette itération améliorée.

L'objectif principal de cette application est de fournir une analyse détaillée et interactive des centres, en explorant plusieurs axes d'étude :

- La localisation des centres sur le terrain.
- La zone de réception des centres.
- Les tirs effectués à la suite d'un centre dans un nombre défini d'événements.
- Le nombre d'adversaires positionnés entre le ballon et le but au moment des événements avant un centre.
- L'analyse des phases de jeu ayant précédé les centres.

2) Contenu de l'application

L'application présente ces analyses sous forme de visualisations interactives : des cartes de chaleur (Heatmap).

Ces Heatmap permettent une représentation intuitive des zones de réalisation et de réception des centres (cf. image 2).

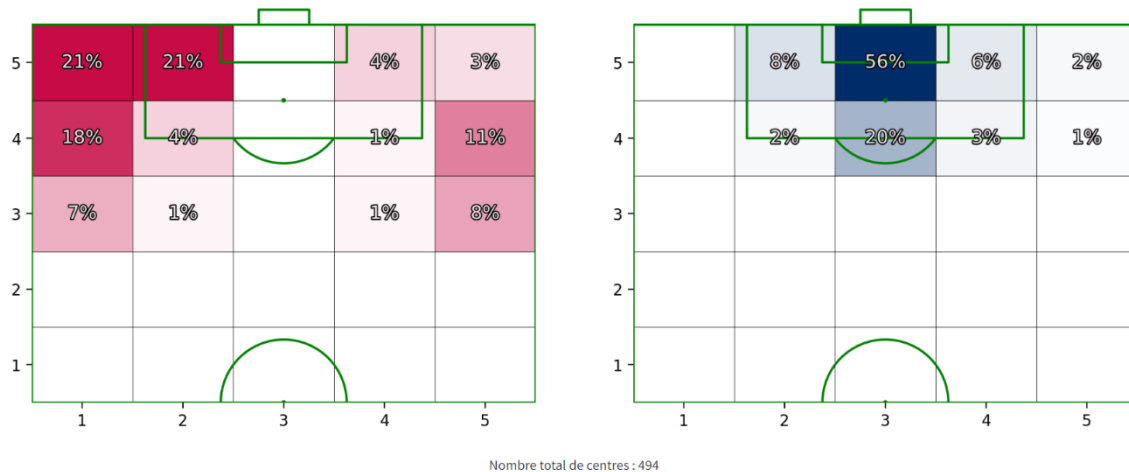


Image 5 – Heatmap des zones de centre (à gauche) et de leur zone de réception (à droite)

Étant donné que la majorité des centres sont réalisés et réceptionnés dans la moitié offensive du terrain, l'application ne représente que cette partie du terrain afin de maximiser la lisibilité.

L'affichage des tirs suivant un centre repose sur un système de flèches, conservé de la version précédente de l'application :

- Les tirs sont représentés par des flèches, indiquant leur direction (cf. image 3).
- Leur destination dans le but est matérialisée sur une vue en 2D des cages, où chaque tir est représenté par un point (cf. image 3).



Image 6 – Graphique de la trajectoire des tirs (à gauche) et de leur destination dans les cages (à droite)

Afin d'assurer une distinction visuelle claire entre les tirs convertis en but et les autres tentatives, les tirs ayant abouti à un but sont affichés en bleu.

A l'inverse, les tirs non marqués sont affichés en rouge.

En complément, l'application permet d'accéder à plusieurs informations contextuelles pour chaque centre analysé, telles que la journée du championnat, les équipes concernées, (domicile/extérieur), la minute de l'action, et d'autres indicateurs facilitant l'analyse détaillée (cf. image 4).

Tableau des tirs/buts pour la zone sélectionnée sur la Heatmap de gauche

Date	Journée	Domicile	Extérieur	Minute	Centreur	But	Tireur	Équipe
2024-02-11	21	139	137	31	7,212	Non	233463.0	137
2024-03-03	24	137	168	53	7,212	Non	233463.0	137
2024-03-30	27	137	156	26	27,336	Non	3599.0	137
2024-04-14	29	137	153	84	7,212	Non	3599.0	137
2023-08-20	2	129	138	29	7,424	Non	3012.0	138
2024-02-10	21	168	138	84	2,946	Non	51843.0	138
2024-03-31	27	138	134	70	7,424	Non	49225.0	138

Image 7 – Tableau informatif sur les centres sélectionnés par l'utilisateur

3) Fonctionnalités de l'application

L'application Streamlit développée dans le cadre de ce projet intègre diverses fonctionnalités permettant d'affiner l'analyse des centres et des actions qui en découlent. L'objectif principal était de concevoir un outil interactif et modulable, facilitant l'exploration des données sous différents angles et offrant aux analystes du Clermont Foot une meilleure compréhension des dynamiques de jeu.

Parmi les fonctionnalités majeures, l'application permet de filtrer les données en fonction de plusieurs critères.

L'utilisateur a la possibilité de sélectionner une compétition spécifique, bien que, dans notre cas, les données ne concernent qu'un seul championnat.

De même, il peut choisir une ou plusieurs saisons à analyser, mais cette option n'était pas exploitable dans notre projet, car nous disposons uniquement des données d'une saison. L'application offre également la possibilité d'étudier les performances d'une sélection d'équipes précises ou de comparer des groupes d'équipes définis en fonction de leur classement.

L'utilisateur peut ainsi sélectionner un groupe de haut de tableau « Top », de bas de tableau « Bottom » ou intermédiaire « Middle », tout en ayant la liberté de déterminer la taille de ces groupes. Cette fonctionnalité est particulièrement utile pour situer les performances du Clermont Foot par rapport à ses concurrents ou analyser le style de jeu de futurs adversaires.

L'un des aspects notables de l'application est la flexibilité offerte dans l'analyse des centres. Afin de faciliter l'interprétation des données, il est possible d'afficher l'ensemble des centres

effectués du même côté du terrain en appliquant une symétrie, permettant ainsi d'homogénéiser les visualisations.

De plus, l'utilisateur peut filtrer les centres en fonction de la partie du corps utilisée pour les réaliser, qu'il s'agisse du pied droit, du pied gauche ou d'une autre surface de contact.

Une autre fonctionnalité clé consiste à isoler les centres ayant conduit à un but dans une séquence définie d'événements. Cette option permet d'identifier les situations les plus décisives et de mieux comprendre les schémas de jeu efficaces.

De plus, l'utilisateur peut choisir d'afficher différentes métriques sur les Heatmap, comme le nombre total de centres réalisés par zone ou le pourcentage de centres rapporté au total.

Afin d'affiner encore davantage l'analyse, il est possible de modifier le quadrillage des Heatmap en ajustant le nombre de divisions du terrain, offrant ainsi une meilleure précision dans l'interprétation des tendances de jeu.

Un filtre important concerne le nombre d'adversaires situés entre le ballon et le but au moment de l'événement précédant un centre. Grâce à un curseur interactif, l'utilisateur peut sélectionner un nombre minimum et maximum d'adversaires afin de sélectionner les centres pour lesquels le nombre d'adversaires entre le ballon et le but sur les actions les précédents est compris entre le minimum et le maximum choisi.

Afin d'éviter une surcharge visuelle, l'affichage des tirs est conditionné par la sélection préalable d'une zone de départ des centres. En effet, si tous les tirs consécutifs à un centre étaient affichés simultanément, la lisibilité du graphique en serait fortement altérée. L'utilisateur peut également affiner son analyse en sélectionnant une zone de réception des centres. Dans ce cas, seuls les tirs effectués après des centres réceptionnés dans cette zone sont affichés, offrant ainsi une meilleure visibilité sur l'efficacité des centres selon leur destination.

L'ensemble de ces fonctionnalités permet d'adapter précisément les analyses aux besoins des analystes du Clermont Foot. Grâce aux divers filtres disponibles, il est possible de mettre en évidence des tendances de jeu spécifiques, de comparer les performances selon différentes configurations et d'optimiser la préparation des matchs en fonction des forces et faiblesses adverses.

IX. Conclusion et perspectives

Ce projet nous a permis d'analyser en profondeur la manière dont les équipes utilisent les centres en fonction du contexte tactique et des phases de jeu. Nous avons observé que l'Attaque Placée Numéro 1 est le cadre privilégié pour les centres, tandis que la Transition Numéro 1 s'appuie davantage sur des actions rapides et directes, avec un recours minimal aux centres.

L'un des aspects intéressants de l'étude a été l'analyse du nombre d'adversaires situés entre le ballon et le but avant un centre. Grâce aux données issues des freeze frames et aux positions des joueurs fournies par SkillCorner, nous avons pu quantifier précisément le degré d'opposition défensive rencontrée avant chaque centre. Cette approche permet de mieux comprendre dans quelles conditions les centres sont effectués et d'évaluer leur pertinence selon le nombre de défenseurs placés sur la trajectoire du ballon.

Cependant, plusieurs défis ont été rencontrés, notamment l'intégration des différentes sources de données (StatsBomb, SkillCorner) et la gestion des orientations du terrain. La standardisation des coordonnées pour que toutes les équipes attaquent de gauche à droite a été une étape clé pour garantir la cohérence des analyses.

Plusieurs pistes d'amélioration pourraient être explorées pour améliorer ces résultats. Tout d'abord, il serait intéressant d'étudier plus en détail l'efficacité des centres en fonction du nombre d'adversaires présents entre le ballon et le but. Par exemple, observer si le taux de conversion des centres en occasions de but varie en fonction de la densité défensive pourrait fournir informations intéressantes.

Enfin, l'ajout des phases d'Attaque Placée Numéro 2 et de Transition Numéro 2 permettrait d'avoir une vision encore plus complète du contexte des centres en fonction des différentes situations de jeu.

En conclusion, ce projet nous a permis de poser des bases solides pour l'analyse des centres en fonction des phases de jeu et du contexte défensif. L'exploitation des freeze frames et des événements a offert un aperçu détaillé des conditions dans lesquelles les centres sont effectués, ouvrant ainsi des perspectives intéressantes pour les analystes du Clermont Foot dans l'optimisation de leur stratégie offensive.

X. Bibliographie

- [1] SkillCorner : <https://skillcorner.crunch.help/en> (consulté entre septembre et février 2025).
- [2] StatsBomb : <https://statsbomb.com> (consulté entre septembre et février 2025).
- [3] SQLite : <https://www.sqlite.org> (consulté entre septembre et décembre 2025).
- [4] Streamlit : <https://docs.streamlit.io> (consulté entre septembre et février 2025).
- [5] Docker : <https://docs.docker.com/manuals/> (consulté en janvier 2025).
- [6] GitHub : <https://docs.github.com/en> (consulté entre septembre et février 2025).
- [7] VS Code : <https://code.visualstudio.com/docs> (consulté en septembre et octobre 2025).
- [8] Pandas : <https://pandas.pydata.org/docs/> (consulté entre septembre et février 2025).
- [9] Numpy : <https://numpy.org/doc/> (consulté entre septembre et février 2025).
- [10] Jupyter Notebook : <https://docs.jupyter.org/en/latest/> (consulté en janvier 2025).
- [11] SQLite3 : <https://docs.python.org/3/library/sqlite3.html> (consulté en octobre 2025).
- [12] SQLite-web : <https://github.com/coleifer/sqlite-web> (consulté entre octobre et janvier 2025).