Une image contenant Emblème, symbole, cheval, logo

Description générée automatiquement

Application streamlit

stage été 2024

Nathan TALBOT

Clermont foot

2024

***Présentation document***

Ce document a pour but de présenter l’application Streamlit que j’ai développé lors de mon stage au Clermont Foot durant l’été 2024. Elle a pour but d’étudier des données et d’obtenir diverses informations sur des clubs de différents championnats, sur différentes saisons.

Ces données sont obtenues car à deux fournisseurs : Stats Bomb et Skill Corner.

Streamlit est une librairie Python qui permet de développer des applications/sites web, en proposant des outils permettant une interaction agréable et pratique pour l’utilisateur.

***INTRODUCTION***

Les pages sont divisées en 2 principaux thèmes :

1. L’étude des métriques

Cette partie est divisée en 3 pages : les métriques différenciants des groupes d’équipe d’un même championnat, l’évolution des métriques pour un groupe d’équipe ou une équipe en particulier au cours d’une saison, ainsi que l’évolution des métriques au cours de plusieurs saisons, pour un groupe d’équipe.

Un groupe d’équipe désigne soit un Top (exemple : les 5 premières équipes du championnat), un Bottom (les 3 dernières équipes du championnat ou un Middle (les équipes classées entre le Top 5 et le Bottom 3).

Les données de Stats Bomb que j’ai utilisées proviennent de l’API « Team season stats », spécifiques aux équipes et agrégées sur une saison, tandis que les données de Skill Corner sont divisées en 4 catégories :

* Physiques : <https://skillcorner.crunch.help/en/glossaries/physical-data-glossary>
* Courses sans ballon : <https://skillcorner.crunch.help/en/glossaries>
* Passes : <https://skillcorner.crunch.help/en/glossaries/passing-metrics-1>
* Pressing : <https://skillcorner.crunch.help/en/glossaries/overcoming-pressure>

Ces données Skill Corner sont spécifiques aux joueurs et étudiées sur un match. Pour ramener ces données sur des équipe, j’ai donc, pour chaque métrique, regroupé et additionné les données des joueurs d’une même équipe sur un même match (e.g. : métrique basée sur la distance parcourue par un joueur durant un match : la métrique pour l’équipe est donc la somme de toutes les distances parcourues par les joueurs d’une équipe durant un match).

Nb : Pour les métriques de type « ratio » (pourcentage de réussite d’une action durant un match), j’ai calculé la moyenne pour tous les joueurs d’une même équipe sur un match, et non additionné les pourcentages de chaque joueur.

Pour chaque catégorie de métrique, il est possible d’étudier les métriques en faisant une moyenne sur une durée, une période avec ou sans possession du ballon, ou encore un nombre d’évènement. Par exemple, pour les métriques de passes, il est possible de faire une moyenne d’une métrique par 100 opportunités de passe, par 30 minutes de jeu avec la possession du ballon, ainsi que par match.

De plus, les métriques sont divisées en plusieurs types, à l’exception des données physiques. Les données de pressing sont classées par intensité de pression (forte, moyenne et basse) tandis que les données de course et de passe sont classées par type de course (course dans le dos de la défense, course de soutien…).

Type de course : <https://skillcorner.crunch.help/en/glossaries/run-detection-and-classification>

1. Les Heatmaps

Ces pages ont pour but de visualiser des Heatmaps (carte de chaleur) apportant 3 types d’informations sur les équipes : les zones de début d’action menant à un tir ou à un but, les zones de tir ainsi que les zones de centres (et de leur zone de réception).

Pour créer ces Heatmaps, j’ai également utilisé l’API « Competition event », qui permet d’obtenir la liste de tous les évènements identifiés par Stats Bomb sur une saison pour une compétition donnée.

API event : <https://github.com/statsbomb/open-data/blob/master/doc/Open%20Data%20Events%20v4.0.0.pdf>

1. ***Métriques discriminantes***
2. *Présentation de la page*

Cette première page a pour but d’étudier les métriques différenciant le plus des groupes d’équipes d’un même championnat, sur une ou plusieurs saisons.

Les informations sont affichées dans un tableau dont les lignes correspondent aux différentes métriques et les colonnes correspondent aux informations sur ces métriques pour chaque groupe : moyenne, différence entre chaque groupe, écart type, minimum, maximum. Dans le cas ou un groupe ne contient qu’une équipe, l’écart type, le minimum et le maximum ne sont pas affichés (car correspondent à la moyenne).

Cette page permet également de sélectionner des métriques dans ce tableau, et d’afficher les valeurs de ces métriques pour chaque équipe du championnat pour chaque saison sélectionnée.

Pour étudier la différence entre 2 groupes pour une même métrique, j’ai décidé de calculer la différence en % en utilisant la formule :

Différence de a avec b = 100\*(a – b)/(valeur absolue de b) %

Par exemple, si on étudie la métrique « distance totale parcourue pendant un match » pour le Top 3 et le Bottom 5 de Ligue 2 lors de la saison 2023/2024, la moyenne du Top 3 sur cette saison par match est de 101715 mètres et celle du Bottom 5 est de 102444 mètres. La différence du Top 3 avec le Bottom 5 est donc de :

100\*(101715 – 102444)/|102444| = -0.71%

1. *Fonctionnalités de la page*

* Choisir le fournisseur de donnée.
* Choisir une ou plusieurs saisons à étudier.
* Choisir la taille des différents groupes. A noter que le Top contient au moins 1 équipe et peut contenir toutes les équipes du championnat, le Bottom peut être vide ou contenir toutes les équipes en dehors des équipes du Top, et le Middle correspond au reste des équipes qui ne sont ni dans le Top ni dans le Bottom.
* Choisir le nombre de métrique gardée dans le tableau final. Par exemple, on peut décider de garder seulement les 3 métriques les plus discriminantes entre 2 groupes.
* Choisir les informations affichées pour chaque métrique (moyenne, écart type…).

Dans le cas ou Skill Corner est choisi :

* Choisir les données provenant uniquement des joueurs sur les matchs qu’ils ont gagnés. Cette option n’est pas possible avec les données Stats Bomb car elles sont agrégées sur une saison et non sur un match.
* Choisir la catégorie de métrique.

Dans le cas ou la catégorie de métrique choisie n’est pas la catégorie de donnée physique :

* C