Une image contenant Emblème, symbole, cheval, logo

Description générée automatiquement

Application streamlit

stage été 2024

Nathan TALBOT

Clermont foot

01 janvier 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. Cette dernière a pour but d’étudier des données footballistiques dans le but d’obtenir des informations sur des équipes.

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

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

Lien de l’application : https://stage-cf63-rqkmdwxeknuypyitvywjjy.streamlit.app

Une image contenant texte, capture d’écran, logiciel, Police

Description générée automatiquement

***INTRODUCTION***

Cette application est composée de plusieurs pages, qui sont divisées en deux principaux thèmes :

1. L’étude des métriques

Les métriques en question sont des variables informatives sur des équipes ou des joueurs. Par exemple, une métrique sur un joueur peut être la distance que ce dernier a parcouru pendant un match, le nombre de passes en retrait… Une métrique sur une équipe peut-être le nombre total de buts inscrits par cette équipe lors d’une saison, la distance totale parcourue par les joueurs de cette équipe durant un match…

Cette partie est divisée en trois pages :

* Les métriques différenciant des groupes d’équipes d’un même championnat.
* L’évolution des métriques pour un groupe d’équipes ou une équipe en particulier au cours d’une saison.
* L’évolution des métriques pour un groupe d’équipes au cours de plusieurs saisons.

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

Les données de Stats Bomb utilisées proviennent de l’API « Team season stats », spécifique aux équipes et agrégées sur une saison, tandis que les données de Skill Corner sont divisées en quatre 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 de Skill Corner sont spécifiques aux joueurs et agrégées par match. Pour obtenir les données au niveau des équipes, 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 de cette équipe durant le match).

Note : Pour les métriques de type « ratio » (pourcentage de réussite d’une certaine action durant un match), j’ai calculé la moyenne des ratios de l’ensemble des 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’événements. Par exemple, pour les métriques liées aux passes, il est possible de faire une moyenne d’une métrique par 100 opportunités de passe, par 30 minutes de jeu en possession du ballon, ainsi que par match.

De plus, à l’exception des données physiques, les métriques sont divisées en plusieurs types. 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 (cartes de chaleur) apportant trois 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 centre (et leur zone de réception).

Pour créer ces Heatmaps, j’ai 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 significativement des groupes d’équipes d’un même championnat, sur une ou plusieurs saisons. Elle permet donc possiblement de connaitre les caractéristiques propres aux équipes performantes (ou non) d’un championnat.

Une image contenant texte, capture d’écran, Caractère coloré, nombre

Description générée automatiquementLes informations sont affichées dans un tableau dont les lignes correspondent aux 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 où un groupe ne contient qu’une équipe, l’écart type, le minimum et le maximum ne sont pas affichés (car ces informations correspondent à la moyenne).

Dans ce tableau, les métriques sont triées en fonction de la différence entre le Top et le Bottom, dans le cas où ce dernier est non vide ; sinon elles sont triées en fonction de la différence entre le Top et le Middle, si ce dernier est non vide.

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

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

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%

Une image contenant texte, nombre, capture d’écran

Description générée automatiquementEnfin, cette page offre la possibilité d’afficher les valeurs de certaines métriques pour chaque équipe d’un championnat sur une ou plusieurs saisons.

1. *Fonctionnalités de la page*

* Choisir le fournisseur de données.
* Choisir une compétition.
* 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 métriques gardées dans le tableau final. Par exemple, on peut décider de garder seulement les trois métriques les plus discriminantes entre deux 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 disponible avec les données de Stats Bomb car elles sont agrégées sur une saison et non sur un match.
* Choisir la catégorie de métrique.
* Choisir la moyenne d’agrégation des métriques (moyenne par 30 minutes de temps de jeu avec la possession, moyenne par match…).
* Dans le cas où on n’a pas choisi d’étudier les données physiques, il est également possible de filtrer les métriques, notamment en choisissant le type de la métrique (type de course ou de pression), de choisir les métriques « threat » (métriques liées à la dangerosité d’une action) …

Une image contenant texte, capture d’écran, Police, nombre

Description générée automatiquement

1. ***Évolution des métriques par journée***
2. *Présentation de la page*

Cette page a pour but d’étudier l’évolution d’une métrique au cours des journées d’un championnat sur une même saison, pour un ou plusieurs groupes d’équipes et/ou équipes.

Les informations sont affichées dans un graphique, ou l’axe des abscisses correspond aux journées de la saison, et celui des ordonnées correspond aux valeurs de la métrique étudiée. Il est possible de superposer l’évolution d’une métrique pour plusieurs groupes d’équipes et/ou équipes.

Une image contenant texte, diagramme, Tracé, ligne

Description générée automatiquement

Les métriques étudiées sont obtenues grâce aux données de Skill Corner car Stats Bomb ne fournit pas de données agrégées par match (pour les équipes).

1. *Fonctionnalités de la page*

* Choisir une compétition.
* Choisir une saison à étudier.
* Choisir d’afficher l’évolution d’un ou plusieurs groupes d’équipes et modifier leur taille.
* Choisir d’afficher l’évolution d’une ou plusieurs équipes et possibilité de les sélectionner.
* Filtrer les métriques en choisissant la catégorie ainsi que la moyenne d’agrégation.
* Choisir les données provenant uniquement des joueurs ayant gagné leur match.
* Choisir la métrique à étudier une fois le filtrage effectué.

Une image contenant texte, capture d’écran, Police, nombre

Description générée automatiquement

1. ***Évolution des métriques par saison***
2. *Présentation de la page*

Cette page a pour but d’étudier l’évolution d’une ou plusieurs métriques au cours de plusieurs saisons pour un ou plusieurs groupes d’équipes d’un même championnat. Cela permet de connaitre les métriques qui sont en hausse (ou en baisse), et donc de pouvoir s’adapter à la tendance d’un championnat.

Les informations sont affichées dans un tableau, dont les lignes correspondent aux métriques sélectionnées, pour chaque groupe choisi. Les premières colonnes contiennent les valeurs de chaque métrique pour chaque groupe et pour chaque saison sélectionnée, et la dernière colonne correspond à l’évolution en pourcentage de ces métriques entre la première et la dernière saison étudiées. Pour chaque saison, hormis la première, la valeur de chaque métrique est affichée en vert si la métrique est en hausse par rapport à la saison précédente, et en rouge sinon. De plus, chaque ligne de la colonne « évolution des métriques » est remplie en vert si la métrique est en augmentation constante (la valeur de la métrique augmente chaque saison), en jaune si elle est en tendance haussière (augmentation et diminution entre les saisons et la métrique est en hausse sur l’ensemble des saisons), en orange si elle est en tendance baissière (augmentation et diminution entre les saisons et la métrique est en baisse sur l’ensemble des saisons), ou en rouge si elle est en diminution constante.

Une image contenant texte, capture d’écran, nombre, Police

Description générée automatiquementLes métriques sont triées en fonction de leur évolution entre les saisons pour le Top.

Enfin, la page offre une visualisation graphique de l’évolution des métriques choisies au cours des saisons, pour les groupes d’équipes sélectionnés.

Une image contenant texte, ligne, diagramme, Tracé

Description générée automatiquement

1. *Fonctionnalités de la page*

* Choisir le fournisseur de donnée.
* Choisir une compétition.
* Choisir le nombre de saison à comparer (si on choisit trois saisons, cela correspondra donc aux trois dernières saisons dont les informations sont disponibles).
* Choisir la taille des différents groupes.
* Choisir les métriques à afficher sur le graphique.

**Dans le cas ou Skill Corner est choisi :**

* Choisir les données provenant uniquement des joueurs sur les matchs qu’ils ont gagnés.
* Choisir la catégorie de métrique.
* Choisir la moyenne d’agrégation des métriques.
* Dans le cas où on n’a pas choisi d’étudier les données physiques, choisir le type de la métrique, les métriques « threat ») …

Une image contenant texte, capture d’écran, ligne, Police

Description générée automatiquement

1. ***Nombre de passes avant un but***
2. *Présentation de la page*

Le but de cette page est de fournir le nombre moyen de passes sur les actions menant à un but. Comme pour les précédentes pages, il est possible d’étudier plusieurs saisons, groupes d’équipes et équipes d’un même championnat. Il est donc possible de comparer les groupes d’équipes/équipes pour obtenir la tendance du nombre de passes avant un but.

Au niveau des types d’actions, on ne s’intéresse pas aux buts inscrits sur coup franc direct, penalty… Cependant, les buts contre leur camp sont comptabilisés.

Les données utilisées proviennent de l’API « Competition event ».

Une image contenant texte, Police, capture d’écran, ligne

Description générée automatiquementLes informations sont affichées dans un tableau, dont les lignes correspondent aux saisons étudiées et les colonnes aux groupes et équipes.

De plus, la page permet d’obtenir, pour chaque saison, le nombre moyen de passes de chaque équipe du championnat sélectionné.

Enfin, dans le cas ou plusieurs saisons sont sélectionnées, la page offre une visualisation graphique de l’évolution du nombre de passes au cours des saisons pour les groupes et équipes choisis.

Une image contenant texte, ligne, diagramme, Tracé

Description générée automatiquement

1. *Fonctionnalités de la page*

* Choisir une compétition.
* Choisir les saisons à étudier.
* Choisir la taille des différents groupes.
* Choisir les types de début d’action menant aux buts (corners, touches, open plays, six-mètres…).
* Sélectionner des groupes d’équipes.
* Sélectionner des équipes.
* Sélectionner les saisons pour lesquelles on souhaite afficher le nombre moyen de passes de chaque équipe du championnat.

Une image contenant texte, capture d’écran, Police, nombre

Description générée automatiquement

1. ***Heatmap des zones de début d’action***

***menant à un tir***

1. *Présentation de la page*

Cette page permet de visualiser sur deux heatmaps les zones du terrain ou démarrent les actions menant à un tir ou à un but, pour des équipes ou un groupe d’équipes d’un championnat au cours d’une ou plusieurs saisons.

Une image contenant texte, diagramme, capture d’écran, carré

Description générée automatiquementLes deux heatmaps sont affichées sur deux terrains, côte à côte. La particularité de celle de gauche est que le terrain est divisé en zones rectangulaires, chacune ayant une couleur fixe et une valeur au choix (pourcentage des débuts d’actions dans chaque zone, nombre de débuts d’actions…), tandis que la heatmap de droite ne contient aucune zone, et la couleur de cette dernière est affichée de manière dégradée en fonction de la densité des débuts d’actions à chaque endroit du terrain. Ces deux heatmaps offrent deux affichages complémentaires des informations sur les zones de début d’action, permettant ainsi une meilleure compréhension des informations souhaitées.

Enfin, le nombre total de tirs (ou de buts) est affiché en bas de la page.

1. *Fonctionnalités de la page*

* Choisir une compétition.
* Choisir les saisons à étudier.
* Choisir d’afficher soit un groupe d’équipes, soit une ou plusieurs équipes.
* Dans le cas où on choisit d’afficher un groupe d’équipe, choisir la taille des groupes et le groupe à afficher.
* Sinon, choisir une ou plusieurs équipes.
* Choisir les types de début d’action menant aux tirs.

**Pour la heatmap de gauche :**

* Choisir le nombre de colonnes et de lignes (donc choisir le nombre de zones).
* Sélectionner uniquement les débuts d’actions menant à un but.
* Choisir le type de comptage de chaque zone parmi : pourcentage, pourcentage sans le signe « % », valeur (nombre totale de début d’action), ou aucune valeur.
* Dans le cas où on choisit des équipes, il est possible de sélectionner une zone de la Heatmap et d’obtenir les informations sur les débuts d’action dans cette zone. Cette option n’est pas disponible quand on choisit un groupe, car dans la plupart des cas, la quantité d’informations serait bien trop importante.

Une image contenant texte, capture d’écran, Police, nombre

Description générée automatiquement

1. ***Heatmap des zones de tir***
2. *Présentation de la page*

Cette page permet quant à elle de visualiser les positions de tir. Elle garde les mêmes principes que la page précédente : informations basées sur des équipes ou un groupe, sur une ou plusieurs saisons d’un même championnat, affichage de deux heatmaps (une avec des zones et une autre sans), affichage du nombre total de tirs…

Cependant, les heatmaps sont affichées dans le dernier tiers du terrain. En effet, la plupart des tirs proviennent de cette zone, il est donc inutile de conserver l’autre partie de ce dernier.

Une image contenant diagramme, texte, capture d’écran, Plan

Description générée automatiquementEnfin, il est important de noter que pour la cohérence des données, les penaltys ne sont pas pris en compte.

1. *Fonctionnalités de la page*

* Choisir une compétition.
* Choisir les saisons à étudier.
* Choisir d’afficher soit un groupe d’équipes, soit une ou plusieurs équipes.
* Dans le cas où on choisit d’afficher un groupe d’équipes, modifier la taille des groupes et sélectionner le groupe à afficher.
* Sinon, sélectionner une ou plusieurs équipes.

**Pour la heatmap de gauche :**

* Choisir le nombre de colonnes et de lignes.
* Choisir le type de comptage de chaque zone parmi : pourcentage, pourcentage sans le signe « % », nombre de tirs, ou aucune valeur.
* Une image contenant texte, capture d’écran, nombre, Police

  Description générée automatiquementUne image contenant texte, Police, ligne, nombre

  Description générée automatiquementDans le cas où on choisit des équipes, sélectionner une zone de la heatmap, et obtenir les informations sur les tirs présents dans cette zone.

1. ***Heatmap des zones de centre***
2. *Présentation de la page*

Cette dernière page offre des informations sur les zones de centre et leur destination. Comme pour les autres pages proposant des heatmaps, les informations sont obtenues via les données de Stats Bomb, dont les informations sur les centres sont disponibles en bas de la documentation de l’API « event ».

A noter que pour une meilleure cohérence des données, j’ai choisi de ne pas inclure les centres provenant de corners.

Une image contenant texte, diagramme, Plan, ligne

Description générée automatiquementIl y’a toujours deux heatmaps affichées, cependant, ces dernières sont chacune contenues dans des demi-terrains, et ces derniers sont divisés en zones (comme les heatmaps de gauche des pages précédentes). La heatmap de gauche contient les informations sur les zones de départ des centres et celle de droite sur les zones de réception de ces centres. Le nombre total de centres est affiché en dessous de ces deux heatmaps.

Il est possible de choisir une zone du terrain sur les deux heatmaps.

Choisir une zone sur la heatmap de gauche permet de :

* Afficher uniquement sur la heatmap de droite les zones de réception des centres provenant de la zone sélectionnée sur la heatmap de gauche.
* Afficher deux figures côte à côte en dessous des heatmaps ; celle de gauche est un terrain (dont on a gardé seulement la dernière partie) qui permet de visualiser sous forme de flèche les positions de départ et d’arrivé des tirs provenant de la zone de centre sélectionnée. Les tirs pris en compte ont eu lieu dans les cinq évènements (au sens de Stats Bomb) survenus après le centre. Les flèches correspondant aux tirs ayant terminés au fond des filets sont affichées en bleu, tandis que les autres sont affichées en rouges. La figure de droite représente des cages, dans lesquelles sont affichées les localisations des tirs sélectionnés précédemment. De même que sur la figure de gauche, les buts sont en bleu et les autres tirs sont en rouge.

Une image contenant ligne, diagramme, Tracé, Rectangle

Description générée automatiquementDans le cas où une zone est choisie sur la heatmap de gauche, il est possible de sélectionner une zone de réception sur la heatmap de droite pour davantage filtrer les tirs affichés sur les figures en dessous.

1. *Fonctionnalités de la page*

* Choisir une compétition.
* Choisir les saisons à étudier.
* Choisir d’afficher soit un groupe d’équipes, soit une ou plusieurs équipes.
* Dans le cas où on choisit d’afficher un groupe d’équipes, choisir la taille des groupes et le groupe à afficher.
* Sinon, choisir une ou plusieurs équipes.
* Filtrer les centres ayant amenés à un but dans les cinq évènements suivants.
* Afficher tous les centres du côté gauche de la heatmap (faire une symétrie des centres se trouvant du côté droit ainsi que de leurs zones de réception).
* Choisir le nombre de colonnes et de lignes des deux heatmaps.
* Choisir le type de comptage de chaque zone parmi : pourcentage, pourcentage sans le signe « % », nombre de centre, aucune valeur, pourcentage des centres ayant amenés à un but dans chaque zone et pourcenatge de but sans le signe « % ».
* Choisir une zone sur la heatmap de gauche.
* Une image contenant texte, capture d’écran, Police, nombre

  Description générée automatiquementDans le cas où une zone est choisie sur cette dernière, choisir une zone sur la heatmap de droite.
* Dans le cas où on étudie des équipes et qu’une zone a été choisie sur la heatmap de gauche, afficher les informations sur les tirs/buts provenant de cette zone de centre. Si une zone de réception a été choisie, les informations affichées ne concernent que les tirs/buts contenus dans cette zone de réception.

Une image contenant texte, capture d’écran, ligne, Police

Description générée automatiquement