Une image contenant Emblème, symbole, cheval, logo

Description générée automatiquement

streamlit APPLICATION

Summer Internship 2024

Nathan TALBOT

Clermont foot 63

***Document introduction***

This document aims to present the Streamlit application that I developed during my internship at Clermont Foot in the summer of 2024. The application is designed to analyze football data to obtain insights about teams.

This data is provided by two suppliers: Stats Bomb and Skill Corner.

Streamlit is a Python library that allows the development of applications, offering tools for a pleasant and practical user interaction.

Application link : https://stage-cf63-rqkmdwxeknuypyitvywjjy.streamlit.app

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

Description générée automatiquement

***INTRODUCTION***

This application is composed of several pages, divided into two main themes:

1. The Study of Metrics

The metrics in question are informative variables about teams or players. For example, a player metric could be the distance they covered during a match, or the number of backward passes made. A team metric might be the total number of goals scored by the team in a season or the total distance covered by the team’s players during a match.

This section is divided into three pages:

* Metrics that differentiate groups of teams within the same league.
* The evolution of metrics for a group of teams or a specific team throughout a season.
* The evolution of metrics for a group of teams over multiple seasons.

A group of teams refers to either the Top (the top five teams in the league), the Bottom (the last three teams in the league), or the Middle (teams ranked between the Top 5 and the Bottom 3).

The Stats Bomb data used comes from the "Team season stats" API, which is specific to teams and aggregated over a season. Meanwhile, Skill Corner data is divided into four categories:

* Physical: <https://skillcorner.crunch.help/en/glossaries/physical-data-glossary>
* Off-ball Runs: <https://skillcorner.crunch.help/en/glossaries>
* Passing: <https://skillcorner.crunch.help/en/glossaries/passing-metrics-1>
* Pressing: <https://skillcorner.crunch.help/en/glossaries/overcoming-pressure>

Skill Corner data is player-specific and aggregated per match. To obtain team-level data, I aggregated the players' data from the same team for a given match for each metric. For example, for a metric based on the distance a player covered during a match, the team’s metric would be the sum of all players’ distances in that match.

Note: For "ratio" metrics (e.g., percentage of successful actions during a match), I calculated the average ratio of all players on the same team for a match rather than summing the percentages for each player.

For each metric category, it is possible to analyze metrics by averaging them over a period, ball possession, or event count. For example, for passing-related metrics, you can average a metric per 100 passing opportunities, per 30 minutes of ball possession, or per match.

Additionally, except for physical data, the metrics are divided into several types. Pressing data is categorized by pressure intensity (high, medium, and low), while running and passing data are categorized by run type (runs behind the defense, support runs, etc.).

Run types: <https://skillcorner.crunch.help/en/glossaries/run-detection-and-classification>

1. Heatmaps

These pages visualize heatmaps providing three types of information about teams: the starting zones of actions leading to a shot (or a goal), shooting zones, and crossing zones (and their receiving zones).

To create these heatmaps, I used the "Competition event" API, which provides a list of all events identified by Stats Bomb during a season for a given competition.

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

1. ***Discriminating Metrics***
2. *Page Overview*

This first page aims to analyze metrics that significantly differentiate groups of teams within the same league over one or more seasons. This allows identifying potential characteristics of successful (or unsuccessful) teams in a league.

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

Description générée automatiquementThe information is displayed in a table where rows represent metrics, and columns contain information about these metrics for each group: average, difference between groups, standard deviation, minimum, and maximum. If a group contains only one team, the standard deviation, minimum, and maximum are not displayed (since these values match the average).

Metrics are sorted based on the difference between the Top and Bottom groups (if the Bottom group is not empty), or between the Top and Middle if the Bottom is empty. To study the difference between two groups for a given metric, I calculated the percentage difference using the following formula:

Difference of a with b = 100\*(a – b)/ |b| (en %)

For example, if we analyze the metric "total distance covered during a match" for the Top 3 and Bottom 5 of Ligue 2 in the 2023/2024 season, the Top 3 average per match is 101,715 meters, and the Bottom 5 average is 102,444 meters. The difference of the Top 3 with the Bottom 5 is:

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

Une image contenant texte, nombre, capture d’écran

Description générée automatiquementFinally, this page offers the ability to display certain metric values for each team in a league over one or more seasons.

1. *Page Features*

* Choose the data provider.
* Choose a competition.
* Choose one or more seasons to analyze.
* Choose the size of the groups. Note that the Top contains at least 1 team and can contain all the teams in the league, the Bottom can be empty or include all teams outside the Top, and the Middle is the remaining teams not in the Top or Bottom.
* Choose the number of metrics to keep in the final table. For example, you may decide to keep only the three most discriminating metrics between two groups.
* Choose the information displayed for each metric (average, standard deviation, etc.). If Skill Corner is chosen:
* Choose data from matches where the players won. This option is unavailable with Stats Bomb data, as it is aggregated by season and not by match.
* Choose the metric category.
* Choose the aggregation average for metrics (e.g., average per 30 minutes of possession, average per match).
* If physical data is not selected, you can filter metrics by type (e.g., run or pressing type) or choose "threat" metrics (related to the danger level of an action).

**When Skill Corner is Selected:**

* Choose data from only those players who played in matches they won. This option is not available with Stats Bomb data because it is aggregated over a season rather than a match.
* Select the metric category.
* Choose the aggregation average for the metrics (average per 30 minutes of play with possession, average per match, etc.).
* If physical data is not selected for study, it is also possible to filter metrics by type (e.g., type of run or pressure), select "threat" metrics (metrics related to the danger level of an action), etc.

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

Description générée automatiquement

1. ***Metric Evolution by Matchday***
2. *Page Overview*

The purpose of this page is to study the evolution of a metric across matchdays of a season for one or more groups of teams and/or individual teams.

Une image contenant texte, diagramme, Tracé, ligne

Description générée automatiquementInformation is displayed in a graph where the x-axis represents the matchdays of the season, and the y-axis represents the values of the studied metric. It is possible to overlay the evolution of a metric for multiple groups of teams and/or individual teams.

The metrics studied are obtained using Skill Corner data as Stats Bomb does not provide match-level aggregated data (for teams).

1. *Page features*

* Choose a competition.
* Select a season to study.
* Choose to display the evolution of one or more groups of teams and adjust their size.
* Choose to display the evolution of one or more individual teams and select them.
* Filter metrics by category and aggregation average.
* Select data from only players who won their matches.
* Choose the metric to study after filtering.

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

Description générée automatiquement

1. ***Metric Evolution by Season***
2. *Page overview*

This page aims to study the evolution of one or more metrics across multiple seasons for one or more groups of teams within the same competition. This helps to identify metrics that are increasing (or decreasing) and thus adapt to the trends of a competition.

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

Description générée automatiquementInformation is displayed in a table where rows correspond to the selected metrics for each chosen group. The first columns contain the values of each metric for each group and each selected season, and the last column shows the percentage change of these metrics between the first and last seasons studied. For each season, except the first, the value of each metric is displayed in green if the metric has increased compared to the previous season, and in red otherwise. Additionally, each row in the "metric evolution" column is filled with green if the metric shows a constant increase (metric value increases every season), yellow if it shows an upward trend (fluctuating but overall increasing), orange if it shows a downward trend (fluctuating but overall decreasing), or red if it shows a constant decrease. Metrics are sorted based on their evolution between seasons for the Top.

Finally, the page offers a graphical visualization of the evolution of the selected metrics over the seasons for the selected groups of teams.

Une image contenant texte, ligne, diagramme, Tracé

Description générée automatiquement

1. *Page features*

* Choose the data provider.
* Select a competition.
* Choose the number of seasons to compare (e.g., selecting three seasons will correspond to the three most recent seasons for which information is available).
* Choose the size of different groups.
* Select the metrics to display on the graph.

**If Skill Corner is chosen:**

* Select data from only players who played in matches they won.
* Choose the metric category.
* Choose the aggregation average for the metrics.
* If physical data is not studied, choose the type of metric, "threat" metrics, etc.

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

Description générée automatiquement

1. ***Number of Passes Leading to a Goal***
2. *Page Overview*

The purpose of this page is to provide the average number of passes leading to a goal. As with previous pages, it is possible to study multiple seasons, groups of teams, and individual teams within the same competition. This allows for comparison of groups/teams to obtain trends in the number of passes leading to a goal.

Regarding types of actions, goals scored from direct free kicks, penalties, etc., are not considered. However, own goals are included. Data is obtained from the "Competition event" API.

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

Description générée automatiquementInformation is displayed in a table where rows correspond to the studied seasons and columns to groups and individual teams.

Additionally, the page provides the average number of passes for each team in the selected competition for each season. If multiple seasons are selected, the page offers a graphical visualization of the evolution of the number of passes over the seasons for the chosen groups and teams.

Une image contenant texte, ligne, diagramme, Tracé

Description générée automatiquement

1. *Page Features*

* Choose a competition.
* Select the seasons to study.
* Choose the size of different groups.
* Select the types of actions leading to goals (corners, throw-ins, open plays, goal kicks, etc.).
* Select groups of teams.
* Select individual teams.
* Une image contenant texte, capture d’écran, Police, nombre

  Description générée automatiquementSelect the seasons for which to display the average number of passes for each team in the competition.

1. ***Heatmap of Starting Zones Leading to a Shot***
2. *Page Overview*

This page allows you to visualize on two heatmaps the zones of the pitch where actions leading to a shot or goal start, for teams or a group of teams in a competition over one or more seasons.

The two heatmaps are displayed on two side-by-side pitches. The left heatmap divides the pitch into rectangular zones, each with a fixed color and a chosen value (percentage of actions starting in each zone, number of actions, etc.), while the right heatmap does not have any zones and the color is displayed in a gradient based on the density of actions starting at each point on the pitch. These two heatmaps provide complementary views of the starting action zones, allowing for a better understanding of the desired information.

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

Description générée automatiquement

The total number of shots (or goals) is displayed at the bottom of the page.

1. *Page features*

* Choose a competition.
* Select the seasons to study.
* Choose to display either a group of teams or one or more individual teams.
* If choosing to display a group of teams, select the group size and the specific group to display.
* Otherwise, select one or more individual teams.
* Choose the types of starting actions leading to shots.

**For the left heatmap:**

* Select the number of columns and rows (i.e., the number of zones).
* Choose to display only the starting actions leading to goals.
* Select the type of count for each zone from: percentage, percentage without the "%" sign, value (total number of starting actions), or no value.
* Une image contenant texte, capture d’écran, Police, nombre

  Description générée automatiquementIf teams are selected, you can select a zone on the heatmap and obtain information on the starting actions in that zone. This option is not available when selecting a group, as the amount of information would generally be too extensive.

1. ***Heatmap of Shot Zones***
2. *Page Overview*

This page allows you to visualize shot positions. It follows the same principles as the previous page: information based on teams or a group, for one or more seasons of the same competition, displaying two heatmaps (one with zones and one without), and showing the total number of shots.

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

Description générée automatiquementHowever, the heatmaps are displayed in the last third of the pitch. Since most shots originate from this area, it is unnecessary to keep the rest of the pitch. Additionally, to maintain data consistency, penalties are not included.

1. *Page features*

* Choose a competition.
* Select the seasons to study.
* Choose to display either a group of teams or one or more individual teams.
* If choosing to display a group of teams, adjust the group size and select the group to display.
* Otherwise, select one or more individual teams.

**For the left heatmap:**

* Select the number of columns and rows.
* Une image contenant texte, Police, ligne, nombre

  Description générée automatiquementChoose the type of count for each zone from: percentage, percentage without the "%" sign, number of shots, or no value.
* If individual teams are selected, select a zone on the heatmap and get information on the shots in that zone. Une image contenant texte, capture d’écran, nombre, Police

  Description générée automatiquement

1. ***Heatmap of Crossing Zones***
2. *Page Overview*

This final page provides information on crossing zones and their destinations. As with the previous heatmap pages, information is obtained from Stats Bomb data, with crossing information available at the end of the "event" API documentation.

To ensure data consistency, crosses from corners are excluded.

Une image contenant texte, diagramme, Plan, ligne

Description générée automatiquementTwo heatmaps are displayed, each within a half-pitch, divided into zones (similar to the left heatmaps on previous pages). The left heatmap contains information on crossing zones, and the right heatmap shows the receiving zones of these crosses. The total number of crosses is displayed below these heatmaps.

It is possible to select a zone on both heatmaps.

Selecting a zone on the left heatmap allows you to:

* Display only the receiving zones on the right heatmap that come from the selected zone on the left heatmap.
* Display two figures side by side below the heatmaps: The figure on the left is a pitch (showing only the last part) that visualizes starting and ending positions of shots from the selected crossing zone with arrows. Shots ending in the goal are shown in blue, while others are in red. The figure on the right shows goalscoring areas with the locations of the previously selected shots. As with the left figure, goals are in blue and other shots in red.

Une image contenant ligne, diagramme, Tracé, Rectangle

Description générée automatiquementIf a zone is selected on the left heatmap, you can further filter the shots by selecting a receiving zone on the right heatmap.

1. *Page features*

* Choose a competition.
* Select the seasons to study.
* Choose to display either a group of teams or one or more individual teams.
* If choosing to display a group of teams, select the group size and the specific group to display.
* Otherwise, select one or more individual teams.
* Filter crosses that led to a goal in the following five events.
* Display all crosses on the left side of the heatmap (mirroring the crosses on the right side and their receiving zones).
* Select the number of columns and rows for both heatmaps.
* Choose the type of count for each zone from: percentage, percentage without the "%" sign, number of crosses, no value, percentage of crosses leading to a goal in each zone, and percentage of goals without the "%" sign.
* Select a zone on the left heatmap.
* If a zone is selected on the left heatmap, choose a receiving zone on the right heatmap.

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

Description générée automatiquement

* If studying individual teams and a zone has been selected on the left heatmap, display information on shots/goals from that crossing zone. If a receiving zone has been selected, the information displayed pertains only to the shots/goals in that receiving zone.

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

Description générée automatiquement