Authors: Albert Garrell & Gabriel Gausachs
Contact Info: albert.garrell@gmail.com & g4a4b4r4i@gmail.com

We are not submitting on behalf of a commercial organisation
Parma-led proposal: Categorising Player Styles To Support Recruitment Profiling

Offensive Player Clustering using a combination of Opta Vision events and spatial tracking data

Research Question: How can we group similar attacking players together using data analysis, and how might this help teams improve their scouting methods and the analysis of possible replacements or undervalued players?

Objective of this Research: The objective of this research is to develop a model for categorising offensive football players based on their playing styles. Going beyond traditional statistics like goals and assists, we aim to generate metrics that provide insights into a player's typical position on the field, their ability to create goal-scoring opportunities, the frequency of good decisions taken, and their performance in defensive tasks, such as pressure. The ultimate goal is to enhance our understanding of offensive player dynamics and to uncover correlations between some metrics that define those players. This involves creating a clustering model for effective player categorization. By doing so, we aim not only to identify groups of similar players but also to compare those similar players on the metrics that best represent their way of playing.

Rationale For Topic: The subjective nature of the scouting and recruitment area makes it the perfect sector to deep down on the data, aiming to find some patterns or information previously unknown to the human eye. The balance between analysing players through tapes or statistics, with the more personal analysis of the person behind the football player, is very thin and uncertain, and with the emergence of data tracking a new approach for analysing player performance arises. Nowadays, it is essential to make use of this data in order to be able to evaluate players and detect similar characteristics, interesting players to sign and evaluate their performance. In this project we try to merge data tracking with event-data to find the underlying nature of the players performance to find similar offensive players, with the corresponding analysis of each type of player based on key metrics for their roles.

Our approach: In our approach to categorise offensive football players, firstly we extract a combination of metrics using Opta Vision event data and spatial tracking. We start by collecting detailed traditional information from key events such as shots – for example average distance, position and conversion -, passes, like number of line breaking passes per 90 minutes, crosses and other traditional statistics. Also, we combine event-data with tracking data to take into account the position of the players given each event, the pressure applied and received, or the possible targets to understand better every action taken by the players. In addition, we believe it's important to identify relevant features so we extract metrics beyond traditional statistics. Below we explain three of the metrics we collect that we think are crucial for the clustering. After that, we apply machine learning clustering algorithms such as K-Means, Hierarchical and DBSCAN to group similar players based on these characteristics. To optimise the clustering model, we use different methods like the elbow method to determine the optimal number of clusters in K-Means or dimensionality reduction techniques like PCA to highlight significant features.

Finally, based on the results of our model, each player type undergoes a detailed analysis, leveraging key metrics tailored to their specific roles on the field. In doing so, this research aspires to not only bridge the gap between subjective and data-driven evaluations, but also to empower football clubs with actionable insights for refined player recruitment strategies and enhanced performance analysis. These are three of the metrics we will be using to compute the model and classify the players:

(I) Position Analysis: As expected, position analysis is a key metric to compare and classify players. Given all the raw data on the position of every player, we extract a heat map evaluating the % of time that each player spends in each zone (See Exhibit A). To reduce our analysis to only offensive players, we decide to just take into account those players that spend a minimum of 50% of the time in the offensive half of the field. To analyse the position on the offensive half, we divide the attacking half in four zones: left-wing, right-wing, center-up and center-down. We also aim to evaluate the rate conversion of each player given different types of actions (i.e. shots, passes, crosses), and given the case, taking into account the pressure received.

(II) PAx (Passes above expected) per 100 passes: PAx is a metric created by us using the data provided, based on various articles found on the internet. We find this metric highly valuable as it characterises a **player's skill in completing passes**, allowing us to assess how proficient or lacking a player is in executing passes.

In our dataset, each pass is associated with different pieces of information, including the *pass target* and the *outcome*. Within the pass target field, we find the *expected pass completion (xP)*, which refers to the likelihood of the pass being successfully completed. For each player, we aggregate the completed passes, attempted passes and the sum of the probabilities associated with their passes. We then apply the formula: [(Completed passes – xP) / Attempted passes] to get the *passes above expected per pass*. Finally, we extrapolate the values to per 100 passes to get a more relative perspective and easy interpretation of the metric. (Exhibit B). What's more, with this metric and the data provided, we could go further and detect the quality of the player in making passes on goal scoring opportunities using metrics such as expected threat or expected receiver.

(III) Pressure value per 90 minutes: Offensive players are known for their passing, dribbling and shooting, but coaches and football analysts are also interested in the defensive work they do. That's why we've created this metric called pressure value. This metric tells us **in a quantitative way how much pressure each player is exerting**. In the data, some events like pass, ball touch, take ons, have a field called *pressure.player* that tells us which players are pressing in that event and how (high, medium, low). For each of these events, we take the players who are pressing and assign them a pressure value (high=0.99, medium=0.66 and low=0.33). Then we do a sum per player and we get the total pressure value and we extrapolate it to get the pressure value per 90 minutes. (Exhibit C)

Application within the real world: Our research in categorising offensive players will help teams improve their scouting methods and the analysis of possible replacement or undervalued players, offering a fresh perspective in player evaluation, hence maximising the chances to improve the quality of their team and results.

Authors: Albert Garrell & Gabriel Gausachs
Contact Info: albert.garrell@gmail.com & g4a4b4r4i@gmail.com

Exhibit A

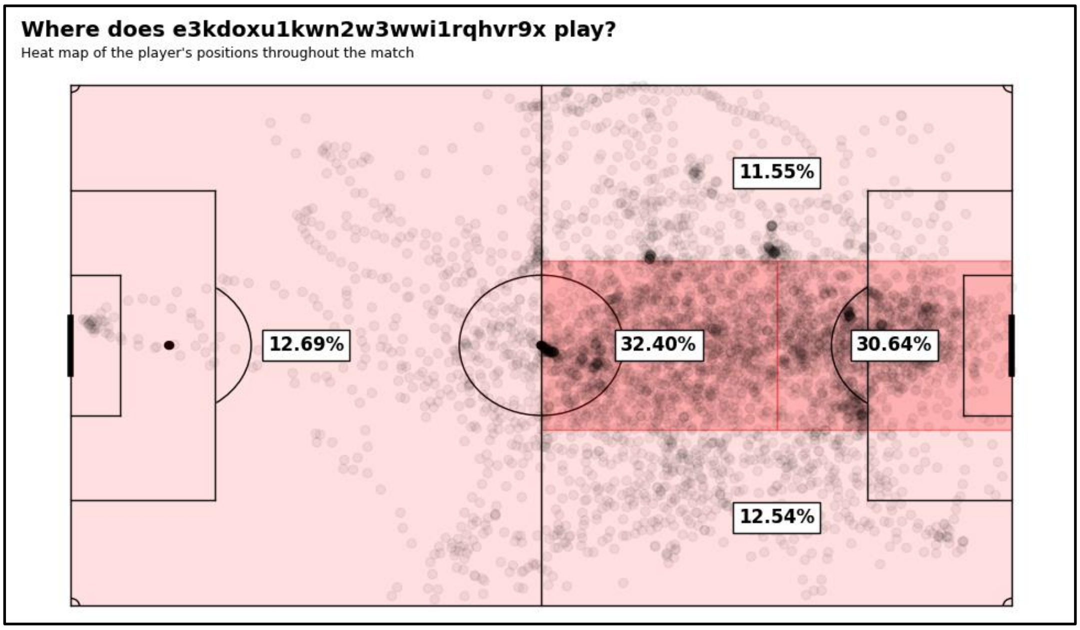
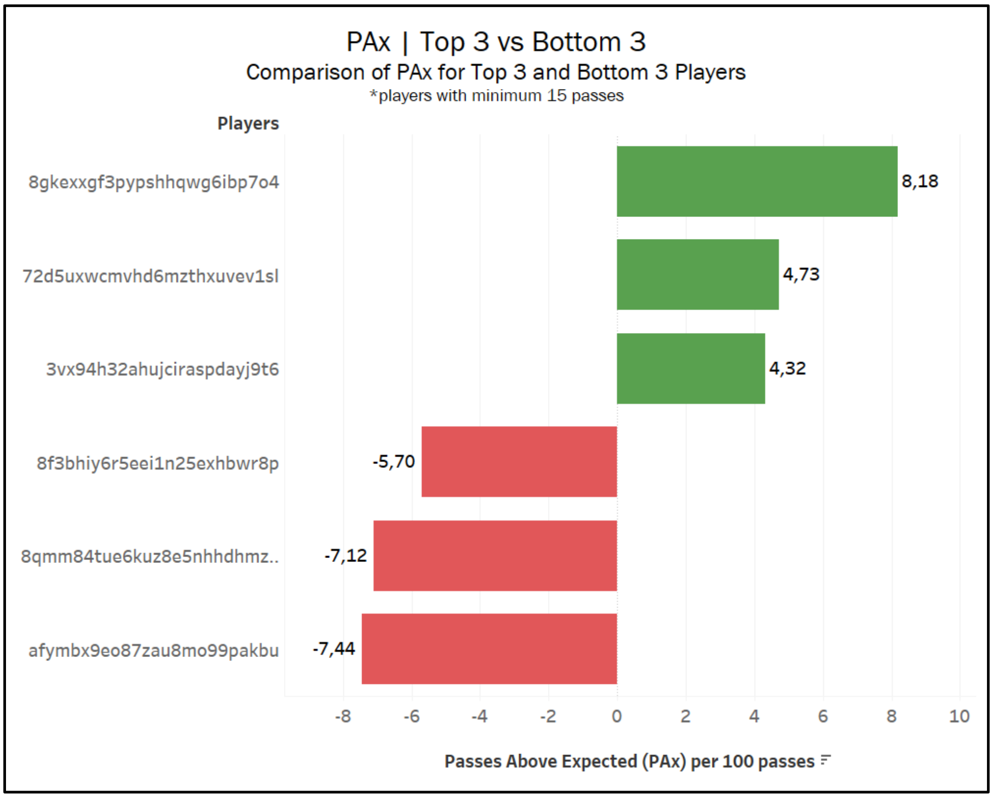
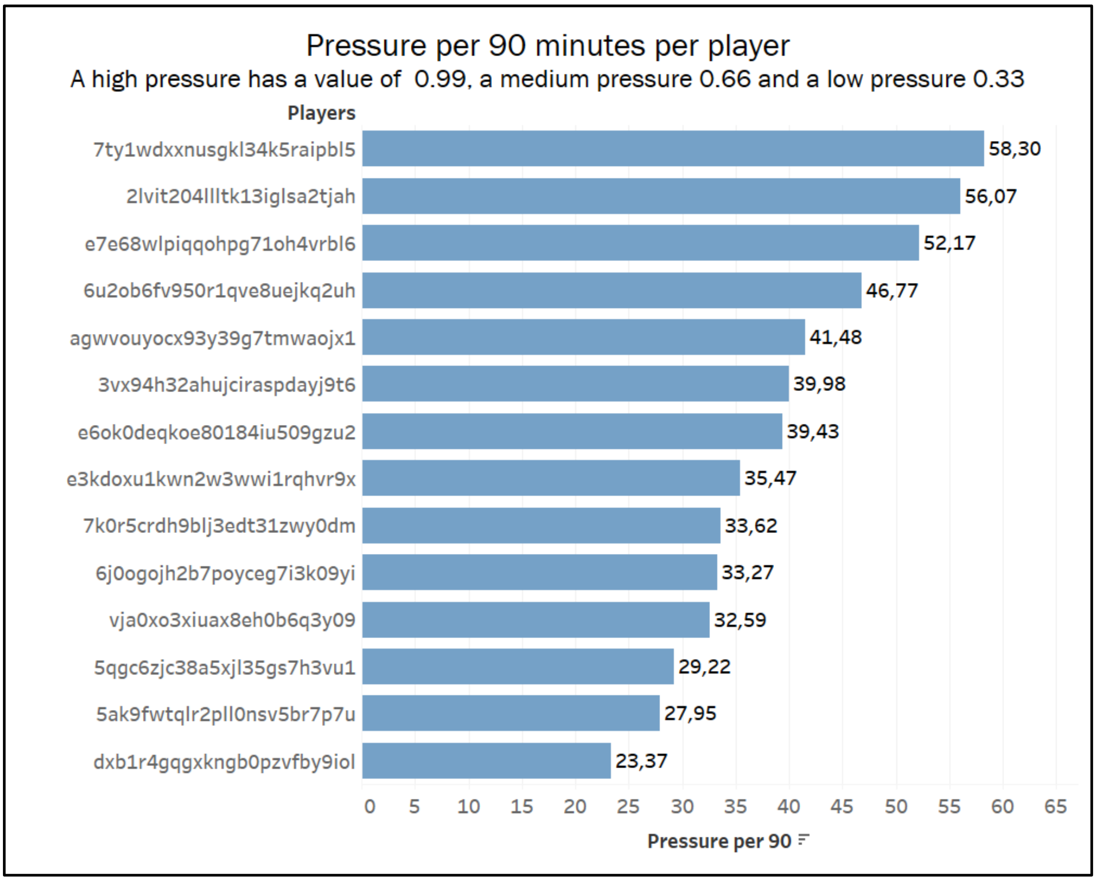


Exhibit B



Authors: Albert Garrell & Gabriel Gausachs
Contact Info: albert.garrell@gmail.com & g4a4b4r4i@gmail.com

Exhibit C



Inspirational work:

https://careyanalytics.wordpress.com/2018/02/22/quantifying-player-profiles-the-evolution-of-the-full-back/

https://worldfootballanalytics.com/2021/09/05/introduction-to-football-analytics-top-6-analytics-metrics-you-should-know/

https://theanalyst.com/eu/2021/09/expected-pass-completion-explained/