One of the major downfalls of tracking data in football is the lack of a common language to describe actions that take place off the ball, particularly patterns of player movement. This approach provides a method for identifying and classifying off-ball in-possession runs into similar groups to allow for more generalisable analysis. The objective is to create a vocabulary of run types that can be used to better describe or analyse specific runs and be queried more easily than raw tracking data. These runs are identified by segmenting the raw tracking data using periods of high player speed and acceleration, then classifying them using a clustering method with functional cluster centres modelled as Bézier curves.

### Clustering - Bézier Curves
Cluster centres are represented by Bézier curves which provide an ideal functional form for measuring distances between trajectories of arbitrary lengths and cluster centres. These Bézier curves are defined as follows where \( \Theta \) is the matrix of control points (x,y coordinates) and \( P \) is the number of control points.

\[
B(t, \Theta) = \sum_{p=0}^{P} \binom{P}{p} t^p (1 - t)^{P-p} \Theta_p, \quad \text{for} \quad t \in [0,1]
\]

The cluster centres (control points) are determined by solving the following linear least squares problem and iterating until convergence when the average distance from each trajectory (\( r_i \)) of length \( d_i \) to its cluster centre is below a certain tolerance level.

\[
\min_{\Theta} \sum_{i=1}^{N} r_i = \sum_{i=1}^{N} \epsilon_i = \sum_{i=1}^{N} \left( \sum_{p=0}^{P} \binom{P}{p} t^p (1 - t)^{P-p} \Theta_p - X_i \right)^2
\]

### Coordinate Adjustment
The clustering is done on adjusted coordinates where all runs are in coordinates relative to the team centroid rather than the centre of the pitch (also all mirrored to originate from the left hand side of the pitch). This helps group runs of similar intent that occur in different parts of the pitch. Consider all of the runs grouped in one cluster mapped in both adjusted and actual coordinates.

### Overlapping Fullback Runs
A grouping of the most common run types made by fullbacks highlights different styles of “overlapping” runs.

### Player Archetyping - Most Common Runs
By looking at a player’s most common off-ball in-possession runs we can learn a lot about that player’s tendencies. Looking at these two anonymised players’ most common run types we can clearly identify that player 0 is an attacking player who gets wide to stretch the field and cut in while player 1 is a full back who makes lots of wide overlapping-style runs from positions roughly level with the team centroid.