|  |  |  |
| --- | --- | --- |
| **Metric** | **Definition** | **Macro-Averaged Correlation with EOS Points** |
| **EDA + Fundamental Analysis** | | |
| Passing Volume | Number of passes (successful/unsuccessful) executed by a team in a given match. | 0.8 |
| Mean Passes Per Player | The expected value for the number of passes a single player in a team is involved in. Calculated independently from each game using node degree from player network graphs | 0.21 |
| Standard Deviation of Passes across Players | The standard deviation for the number of passes across players. A high value indicates the coexistence of players with high passing volume and low passing volume in the game ie: there are specialized play-maker roles | 0.58 |
| Mean Passes Per Zone | The expected value for the number of passes executed from within each zone to a different zone. Calculated independently from each game using node degree from 9-zone network graphs. |  |
| Standard Deviation of Passes across Zones | Analogous to player network graph standard deviation with a 9-zone mapping. A high z indicates coexistence of “hot” and “cold” zones |  |
| Passing Lane Intensity | Proportional to the frequency of passes among lanes between zones in the 9-zone discretization. Normalized to sum to one.  Yield a total of 36 features per team, per game. | Note: Did clustering analysis and found 2 distinct clusters representing different kinds of performances in terms of passing lane intensity. Statistically significant performance difference was also found across the two clusters. |
| Advance Ratio | Here, Y is the lateral direction (left/right) and X is the vertical direction (towards/away from the opponent’s goal). A high advance ratio indicates more lateral activity ie: crosses. | 0.68 |
| **Graph Theory Analysis (Player Networks)** | | |
| Clustering Coefficient | Calculates the (weighted) number of triangles centered around the node as with respect to the number of possible triangles there could have been. Measures local robustness in a player network. | 0.52 |
| Algebraic Connectivity | Second smallest eigenvalue of the Laplacian(D-A) matrix of the graph. Quantifies integration/segregation between players.  Fault tolerance interpretation: A high algebraic connectivity means a network is tolerant to faults in certain nodes. This means if a player is marked (by an opponent’s defender) or off the field(red card or injury), the network as a whole is tolerant highly to it in terms of ball movement. | 0.56 |
| Average All Pairs Shortest Path | Shortest path between all pairs of nodes(players) in a network graph where edge weights are defined as 1/frequency of passes between the nodes. A low value indicates movement of the ball between 2 random nodes, on average, was performed frequently. | -0.73 |
| **Defensive Metrics – Count Based, Derived from Event Tags** | | |
| Yellow Cards/Game | Normalized by possession | -0.43 |
| Dangerous Ball Lost | Normalized by possession | -0.33 |
| Total Interceptions | Normalized by possession | -0.67 |
| Interceptions inside box | Normalized by possession | -0.67 |
| Total slide tackles | Normalized by possession | -0.56 |
| Slide tackles inside box | Normalized by possession | -0.57 |
| Fraction of slide tackles in box |  | -0.25 |
| Fraction of interceptions in box |  | -0.48 |
| Centroid <x, y> per match for slide-tackle and interception |  | 0.64 in x, -0.06 in y |
| Fouls |  | -0.34 |
| Yellow card per foul (in a given match) |  | -0.01.  Note:  Committing fouls gives us a negative correlation but this is not because of bookings. See that the number of yellow cards per foul committed is an incredibly weak signal. Thus, the performance degradation that comes with fouls is likely to have come from set pieces rather than bookings |