

Domain situation

WorldCup 2022, one of the most famous and exciting sport tournaments, is just happening recently. Different teams' match strategies and performances have become an interesting topic in people's conversation in the last two weeks. In this project, we fetched the most up-to-date worldCup 2022 dataset, which includes various statistics for both sides of each match, including total attempts, defense pressure applied, ball possession, etc. Here we ask several interesting questions. For example, does higher ball possession rate bring more goals? Do different time slots of the match make differences in the player's general performance? Does the player know their opponents, or are they familiar with the players in the opponent's team? All the questions are answered by the four graphs on the html page *combined.html*.

Data/task abstraction

We have two data sets. The first one is the huge dataset with numerous columns that represent the statistics of two teams in a specific match.

One major data transformation we have to conduct to answer some questions is to convert the data from "matchwise" to "teamwise". For example, if we want to get the total number of attempts of Argentina in all the past teams. We have to iterate through each match that involves Argentina, collect the data, and add them up to get the final answer. In another example, we want to get the match data separately for each different time slot. Very similarly, we have to go through each row that row, fetch the match time slot, and group it together with other matches that share the same time slot. After we finish wrangling the data, we will be able to feed the data into the visualization process.

The other dataset is particularly used for the third graph. It is a json file that collects each player in the final game as a node and their relations as links. It is used for exploring whether the players in the final match are at the same level and are familiar with each other.

Visual Encoding and Interaction Idiom

Graph1: This is a scatterplot that explores whether the strategy of controlling the ball can help to win the match. Specifically, it tries to check whether there's a positive correlation between the ball possession / number of passes and the number of goals the team is able to make in the match. Each point represents a team in a match. X-axis represents the possession rate, and the y-axis represents the number of goals. The left slider controls the number of passes in the game. The match point will turn red if the number of passes exceeds the threshold set by the slider.

Graph2: This bar chart attempts to find if the difference in the match will affect the player's general performance, which is indicated by the number of goal attempts in the game. Normally when the players are playing well, more chances to goal will be created, which leads to more goal attempts and on-target attempts. The selection button on the top can change the view

mode between total goal attempts and on-target attempts in the five different match hours in the worldCup.

Graph3: This node-link chart tries to find out whether the players of two teams in the final match - Argentina and France - know each other. This graph has nodes to be players in Argentina and France and edges to indicate if the two players are playing in the same league. Each blue node represents an Argentina player and each purple node represents a French player. The player name will be shown when hovering the mouse on the node. Here we can see a majority of players in the two teams are playing premier league and La Liga. They have been playing against each other many times in the club, so they do know each other and they are familiar with each other to some degree.

Graph4: The last graph is a pie chart that shows the statistics of the final four teams - Argentina, Croatia, France, and Morocco. It provides three views to the data - total attempts, defensive pressure applied, and switches of play completed. Switching between different views we can find that Argentina and France tend to be more aggressive with higher total attempts. However, Croatia and Morocco get into the semi-final thanks to their great defense and their counter-attack strategy which can be shown on higher defensive pressure and switches of play completed.