

SoccerTact

Deepak Kumar
Fares Elattar
Grant Keebler
M. Atif Siddiqui
Rahul Purswani

Project Synopsis

A Data Analysis based Web Application that lets users analyse a soccer game.

Project Description

There is a popular saying in English football that “ League table never lies”, our project contradicts that by providing a data based approach to analyse soccer games. We go beyond that league table to provide the users with metrics that really defines who the best teams and players are.

Through our web-application users can select a game to analyse and see defining stats that produce the best results and performances. The application provides the user to analyse important events, team statistics and player statistics.

Our project aims at understanding and demystifying the game of soccer through data.

Project Milestones

First Semester:

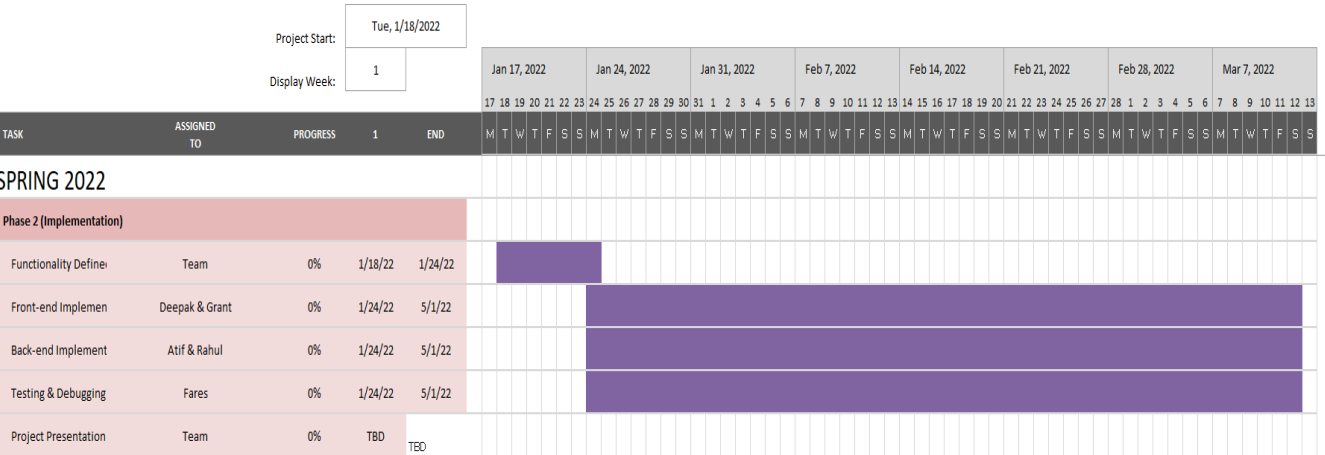
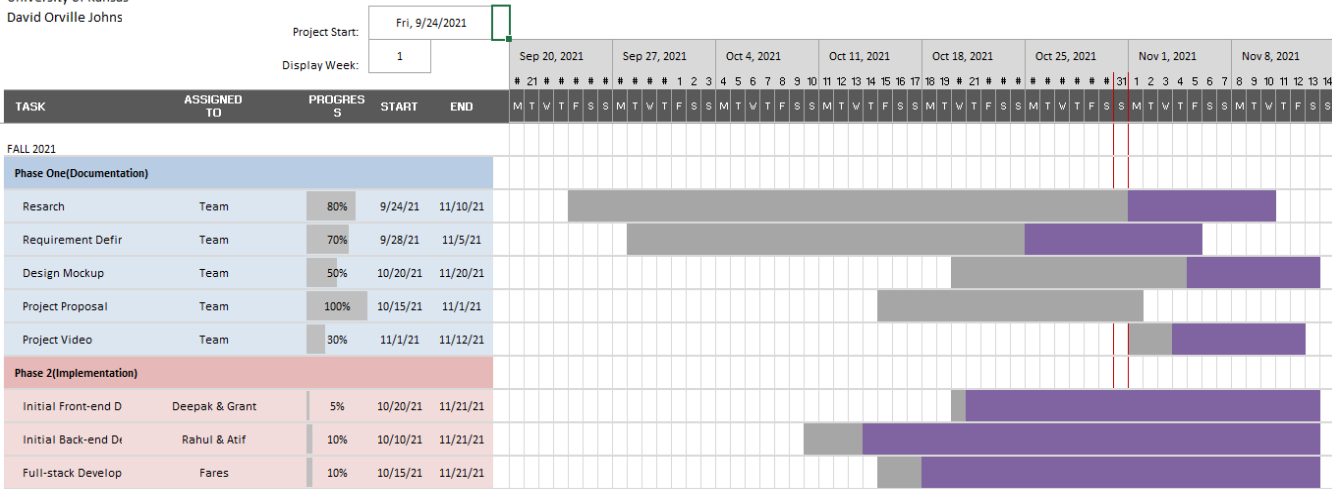
- Project Architecture and Design (Oct 8th)
- Backend and Frontend implementation started (Oct 30th)
- Data Collection (API's implementation and Documentation) (Nov 2nd)
- Project Proposal Video (Nov 1st)
- Basic Backend done (Nov 20th)
- Documentation (Nov 20th)
- Integrating Backend and Frontend (Nov 22nd)

Second Semester:

- Develop frontend to show more data insights (Feb)
- Develop backend to include more models for analysis (March - April)
- Finish User Interface (Late April)
- Finish Up (May)

Gantt Chart

University of Kansas
David Orville Johns



Work Plan:

- Backend And Data Science Team - M. Atif Siddiqui and Rahul Purswani
 - ❖ Creating a database of players records
 - ❖ Backend programming of the simulation
- Frontend Team- Deepak Kumar and Grant Keebler
 - ❖ Proposal Video
 - ❖ Design/UI
- Full Stack Developer
 - ❖ Floating between backend and the front end

Project Budget Preliminary Project Design

Item	Description	Date Needed	Cost
Data API	StatBombs Open API	10/4/2021	\$0
Website Domain	We are planning to host our project as a website, hence will need a Website Domain. We are looking for a domain from Google Domains and the average cost of a domain is about \$30.	02/15/2022	\$30
Total			\$30

Preliminary Project Design

Overview

SoccerTact is a web application that provides a complete analysis on soccer matches, teams, and individual players. It uses event data provided by StatsBomb (an OpenSource Data) to provide analysis on matches, individual teams and players. Event data records all the actions with the ball throughout the game.

This project is divided into two parts- frontend and backend. Backend does everything from cleaning and transforming the data to producing models for different factors such as passing, shots, etc. These models produce visualizations (graphs and figures) that give complete analysis for the selected matches, teams, or players. Frontend, on the other hand, handles the user interface. It helps the user filter the games, player, or the team user wants to analyze. UI then displays all the visualizations and analysis produced by the backend, specific to the selected game, team, or player. For game analysis, UI provides a complete timeline of all the major events that took place in the game. It also showcases the Match Sheet, which contains general information about the match like total shots, total passes, total possessions, etc for the selected teams or players.

→ Player Analysis Screen

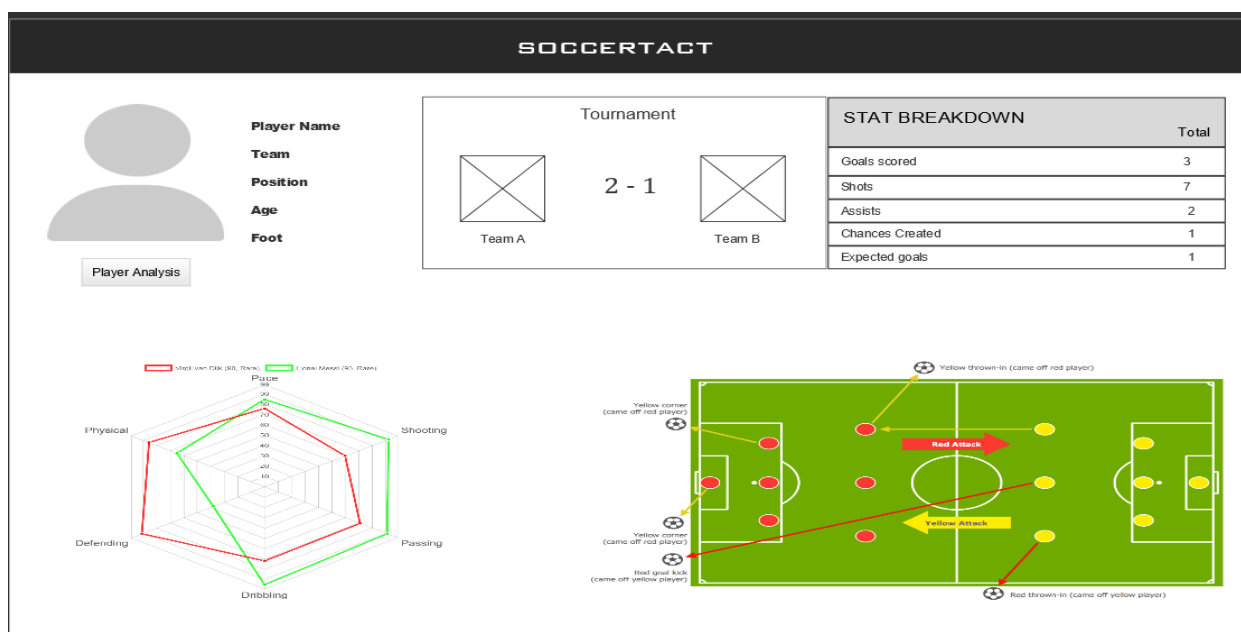
Users will also be able to see different kinds of player analysis just by clicking on them. SoccerTacts will show statistics like total shots, total shots on target, total successful passes, goals scored, goal assists, etc. for each player and how they have changed throughout the game or the entire history of the player. It will visualize this data with heatmaps to give a detailed overview of how the player dominates the field for a given game. It will show an analysis of the player for the current game as well as on the entire history of the player.

Individual Players Analysis Features

- Shows Player Match report (Example Below)
- Heat Map for each individual player (Uses Non-Neg Matrix Factorization) (OS)
- Ranking player among other players.
- Comparing different techniques used for shooting and passing.

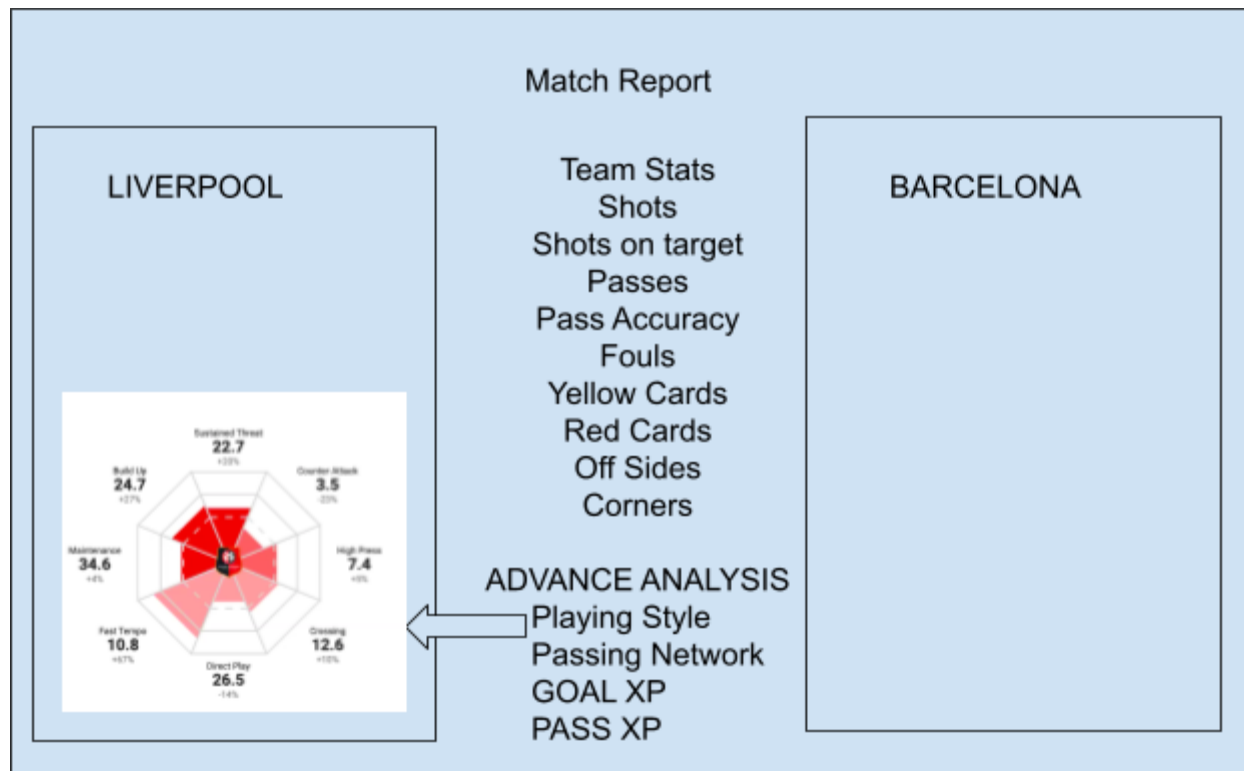


SOC CERTACT



→ Match Analysis Screen

The UI provides the user with Stats reports about the game. These include the total shots, passes, assists and fouls by each team. This screen also provides advanced analysis for each team. It shows a visualization of different playing styles used by the two teams. The user is also provided with features to compare two events from the game and see how these events are weighted against each other.



Backend

Data Extraction and Transformation

- Convert data from JSON to CSV format:

The Data we are using is StatsBomb Event data. This comes in as a cluster of JSON files. We use data extraction and transformation techniques to convert this JSON data to Pandas Dataframe. It is then converted to CSV format to store the visualizations and data produced that will be used by the frontend.

- Cleaning the data:

The data collected from Statsbomb is massive with thousands of event points. The data cleaning model makes lists of important events by adding features or events that are useful and discarding the events that are not important for our design.

- Data Visualization:

The new parsed data is used to create visualization that helps in understanding the game better. There are 2 different kinds of visualization SoccerTact provides, Team and Player based. These visualizations include graphs and charts that tell the visual story of a match.

For Match and Team analysis

- Event Comparison model:

Using feature Vectors, the model compares 2 events for similar kinds. This will help in determining how events are weighted against each other. The general idea of the model is to assign a pre-action and post-action value to the events and see how they change the possibility of either scoring or conceding a goal. This analysis feature uses Soccer Action open-source package.

- Passing Models:

This project consists of two Passing models- Passing Networks and Passing Difficulty models. Both models are implemented using Neural Networks to provide the user the ability to analyse how different passes are weighted against each other. The Application also provides the user to compare different types of playing style used by a team throughout a game.

- Playing Style and Tactics comparison:

Using Event data we extract the different playing styles used by a team throughout a game. This data is visualized to show how the team used different tactics and styles of playing throughout the game.

For Player Analysis

- Match report for each player:

Using event data we extract the data related to individual players to develop a player report for each match. This consists of total shots and passes attempted by the player. Shows the percentage of successful passes and shots completed. The report also contains data on players Name, country, age. It gives a stats breakdown for the actions performed by the player.

- Heat maps for each player:

The match report shows a heat map for individual players to figure out the action space of the player. This can be used to develop action maps for both teams.

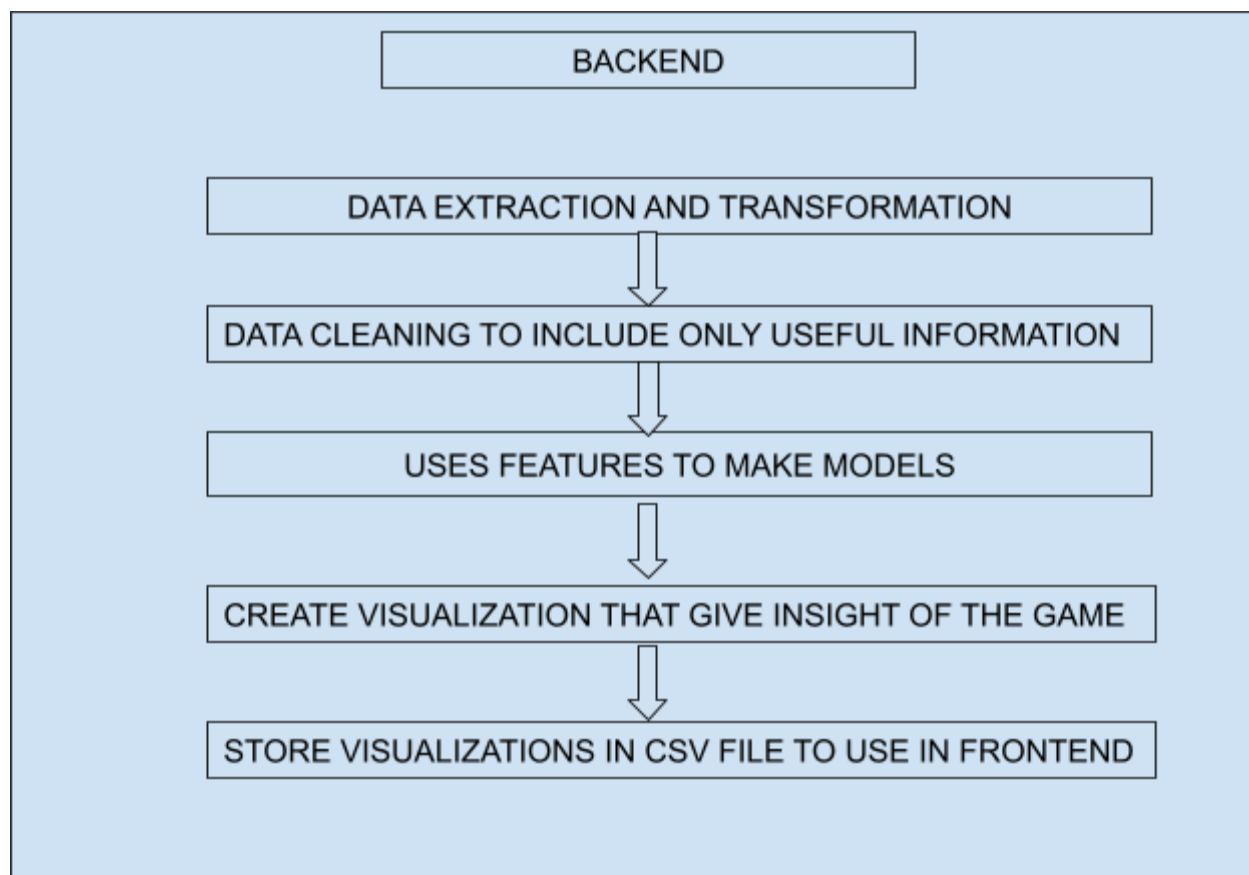
- Compare different techniques for shooting and passing used by each player:

Using event data we see how different techniques are used by a player. It helps in figuring out the percentage of each technique that results in successful pass and shot completions.

- Shot prediction using logistic regression

Using Logistic Regression we provide the prediction for the shot completion. This model takes into account the location the shot was taken from and the technique used to shoot the ball.

The most important feature of SoccerTact is that it is scalable. We have planned to increase the analysis feature provided to the user.



Design Constraints

One of the major constraints we are facing is the limited availability of event data for different games. This reduces the accuracy and scope of our models in the backend (for team and player specific). Soccer Clubs maintain confidentiality for their team-specific and player-specific data. This makes our analysis limited to the teams and players we have in our data. We also couldn't find any tracking data for any matches or teams. Tracking data records the positions and the actions of all the players throughout the match. Unavailability of tracking data limits our analysis to actions with the ball. We cannot analyze many player statistics like speed of the player, speed of the shot, player movements without the ball, etc. with just event data.

Ethical Issues

The web application will be open for anyone to use with no discrimination. However, we are limited by the set of data of matches and players. Our idea was motivated by our passion towards soccer and the curiosity to understand and analyze the game. After some research we came across similar ideas that could inspire some features that we will add to our project. We will make sure to clearly give credits to those resources. The data we will be using is open-sourced therefore, we are not exposing any private club data. There is some other data that the clubs keep private for their own use, but we do not use that data.


Intellectual Property Issues

Third Party Stats:

So in our project the stats and data are not fully ours as we are using StatBomb and we are using game data from google to display we have to make sure that we credit those that provided this data to us even if it was free as of course we did not go and collect this data and we have to ensure they are not only properly credited but of course even if we tried to make money on this project I don't it would be a huge issues to get the rights to use these stats monetarily sense its free to use.

Sharing Our Program to the World

One major issue I think any project would have is when we let people use our service, what is stopping them from copying our work and using it for themselves or



making something and claiming it as theirs or selling it to others. There is also the problem of people editing our code and making changes and providing that to users without our consent. We could sell people use of our program to solve this or maybe get a copyright on our code. This is likely not a perfect solution; however, having our program be completely available will very likely cause a handful of intellectual property issues.

Copyrights and Patents

Obviously the difficulty with getting any kind of patents or copyrights is that we are using third party stats and resources so those things we cannot claim with the exception of the code we write specifically so in this regard it would be somewhat difficult to patent our entire project if it is even possible so that will have to be a hurdle that needs to be crossed for intellectual property. I don't think it would be possible to copyright or patent stats provided by StatsBomb or Google however we can patent and copyright everything else in the project it will just be difficult to get our the statistics we provide and display.

Change Log

Through our research we found that as the granularity of data increases its availability decreases. Soccer data is available in 3 formats. Matchsheet data, event data and tracking data. The simulator that we planned to build was heavily dependent on the availability of tracking data, which is hard to get. Tracking data is private to clubs and is not open. There are some companies that do sell tracking data but it is really expensive. Therefore we decided to change the project to only include event data.