

Title: MM802 - Visualization Mini Project

Date: March 17,2020

Names: Ashish Verma,Ankush Goyal

Project Title: Football 2020 Viz

1 Abstract

Our aim was to create a web application to visualize different characteristics of FIFA-2020 dataset. The project provides an easy to use interface which helps the users to display the number of players according to their country on the world map. Although the display alone is not enough and there are other questions that we answer for the users in this app. We have used various kinds of charts and graphs to answer these questions. The application is built from scratch and we haven't used any kind of template. All the charts are created from using Google chart Javascript.

2 Introduction

2.1 Application Domain

The project is specifically targeting the audience who are enthusiast to know even subtle information about football. Although visualizations are made taking into account the various aspects of only football game, this website can be extended to several other domains as well due to its ease of use from user perspective and ease of extension from the developer perspective. Use of this project depends on the data being used for displaying the visualization.

2.2 Questions of Interest with their answers

- To know the regions where football is more popular. We have developed an GeoChart visualization with map where user can directly visualise which region has more number of players according to the color density on map (dark green color means more dense area for number of players and light green means less dense area). We can directly co-relate the number of players coming from the area to the popularity of the football. User can also hover over specific region of interest to identify region name and the number of players in that specific region. We can observe that clusters of regions with high density of players are in South America and Europe where the sport is popular.
- User can get the idea for the wages of player according to his position in the club. We have visualised this wage component on a pie chart and visualised positions in different colors along with their descriptions. Various categories include SUB(substitute), RES(reserve) etc. and user has to simply hover over the specific region in pie chart to know various details such as category, wage and so on. There are in total 27 positions of players that are plotted on

pie chart. We can infer that clubs spend their majority of wages on SUBstitutes and REServe players.

- To display the relation between age of player and its overall rating , we used scatter plot for plotting the relation. User can easily conclude that which age category has the most skilled players (players whose rating is usually ≥ 85) and vice-versa. We can infer from the scatter plot the various clusters which lie as an outliers from the data. Players with high overall rating and younger in age are less which can be observed from data. Players with very high rating are oftenly in the prime of their careers so are between age 27 to 34.
- From the data we wanted to know the top clubs in the dataset. For the clubs part, we tried to visualise various details using the players belonging to particular club using bubble chart. As players are paid according to their capability, by hovering over bubbles, user can infer which club has how many players, their wage in euros and the estimated value of all the players of that particular club. After Visualization, We could see that clubs which having high wage structure and combined value of the players are playing much better football than their peers.
- We used the potential attribute in the dataset to answer how clubs are valuing the players with high potential in the transfer market. We tried to answer this question using line graph. Usually players in 30's are near to their retirement so we considered players whose age is near ≤ 25 (young players). It can be easily visualised that player whose potential rating is less or equal to 85 (which is a good rating) has steady increase in their worth and after that slope becomes large as value is increasing more drastically wrt to player's potential above 85. We could see that players with potential between 80 to 88 are greater in numbers with low estimated value. As the potential for the player increases so does the value for the club.

2.3 Dataset Used

- We used dataset from the source <https://www.kaggle.com/stefanoleone992/fifa-20-complete-player-dataset> which is publically available
- The size is approximately 48 MB and has 18K+ rows.
- We have preprocessed the data according to the requirements of the project. We have removed the columns for attributes which were not relevant for the visualization.

3 Project Status

As of now, we have accomplished all the targets which we decided early. Only thing we are working on presently is the designing part using CSS which we will be able to complete before our presentation.

Work	Team Member
Collecting the data	Ankush
Creating Page Layout	Ashish
Data Preprocessing	Ankush
Exporting to data to MySQL	Ankush
Connecting Nodejs with MySQL	Ashish
Making ajax calls for sending data from node to web server	Ankush
Interaction with database	Ashish and Ankush
Creating Charts	Ashish
Adding dynamic elements to the visualization and creating interactive visualizations	Ashish
Routing	Ankush

4 Development Environment

4.1 Basic Framework

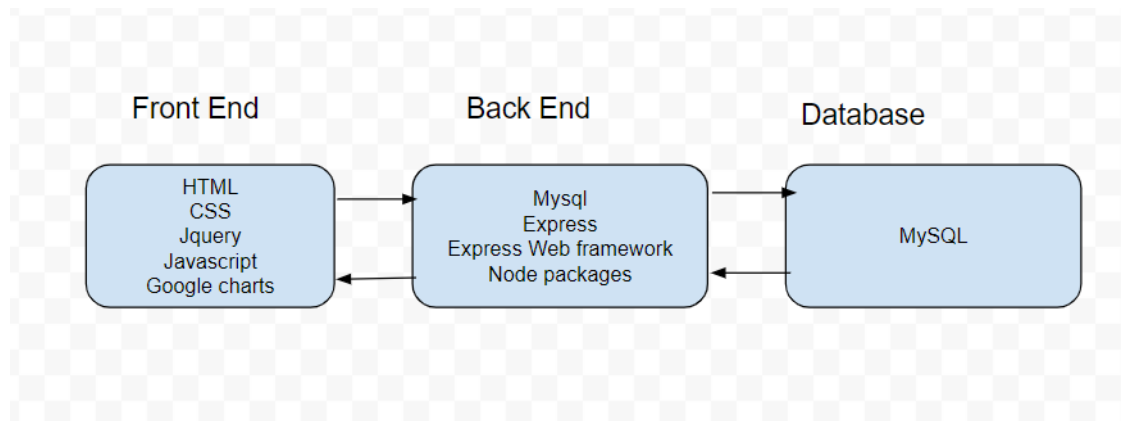


Figure 1: Framework

NPM(Node Package Manager) is used for this platform for handling dependencies. Node.js is used for server-side scripting. Node will act as a web server for this application and will be used for the routing application. Packages used in the node are

- Express
- MySQL

Package MySQL is used to connect node.js with the backend MySQL database. Data is load into MySQL workbench (database) installed locally from csv file. This is connect with node server using the MySQL db details such as credentials, port number and database name.

Google charts are used to get various charts and maps for data visualization. These are powerful, simple to use, and free tools.

EJS(Embedded JavaScript) is used for creating webpage as it has the luxury to add javascript content and HTML content inline.

Initially, a request is sent to server.js which act as server side of the project. This file interacts with the database and does necessary computations. The file also deals with setting up of listening PORT to run on the server and also routes various requests to corresponding functions. Default route of the application will redirect the user to index.html file.

This is a web page application, and the default route will display the results as per default input. It will display a world map where a user can hover over the regions to get no. of players and popularity of game in that particular region.

- Go the folder containing server.js file.
- Open command prompt from that folder.
- Run “node server.js” in the command prompt.
- The listening PORT address will be displayed in the command window and minimise this window.
- Go to browser and type url for the browser is “https://localhost:8080/”.

4.2 Functionality Demonstration

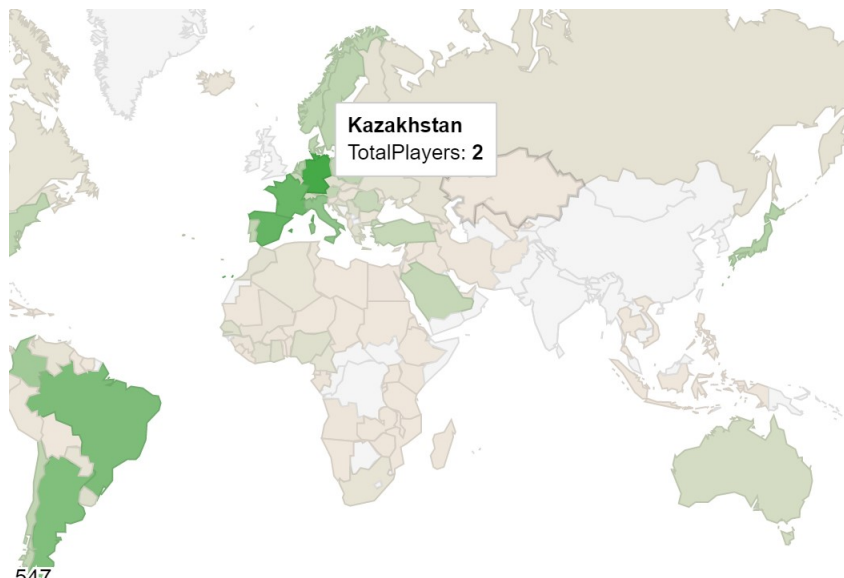


Figure 2: Chart for player population

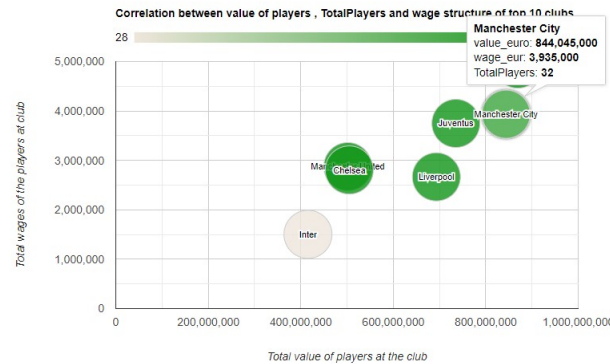


Figure 3: Bubble chart to compare player value VS wage

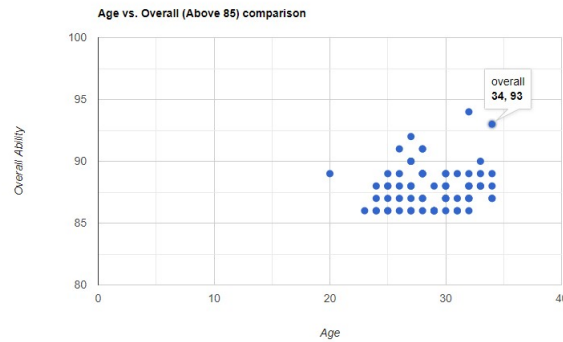


Figure 4: Scatter Plot to compare between age VS Overall ability

5 Concluding Remarks

Our team has been able to complete all the objectives we initially set up for ourselves except designing. We did the designing ourselves without relying on any external prebuilt template. We tried to include all the visualizations that we thought would help the user in efficiently navigating through the data and the problem set we chose. Graphic Visualizations are much more intuitive and gives better representation of the data. Graphics can give insight with much less effort as compared to raw data.

Future Scope: There are few more options that could help the user make the best decision. Application can be made more dynamic and scope can be extended to other sports as well as non-sports categories as well. Also , data from various years can be used to compare different trends between them using various charts

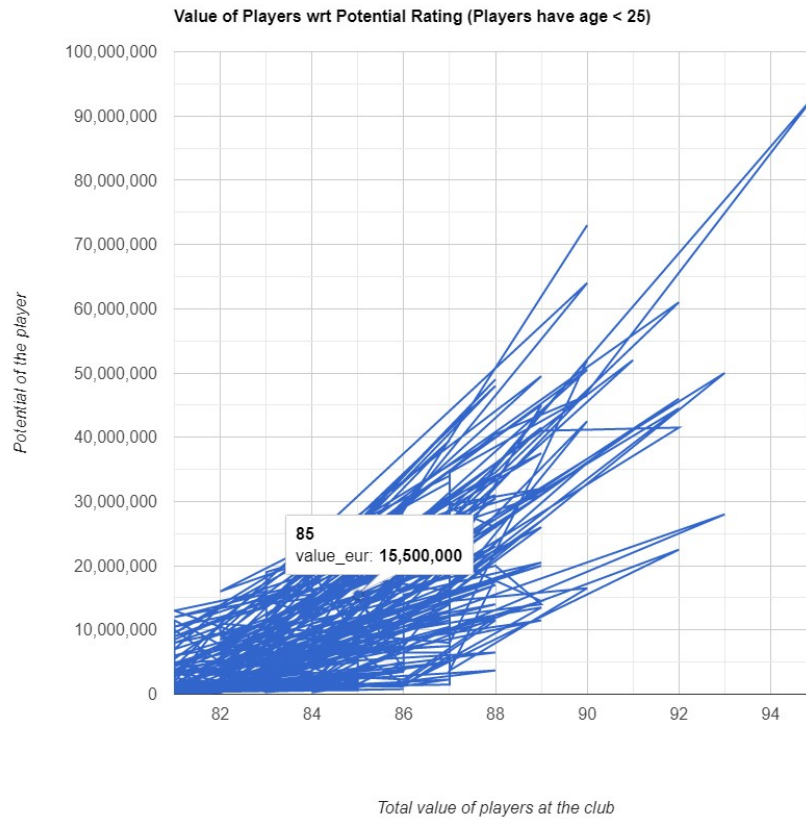


Figure 5: Line Chart for comparison between total player value VS player's potential

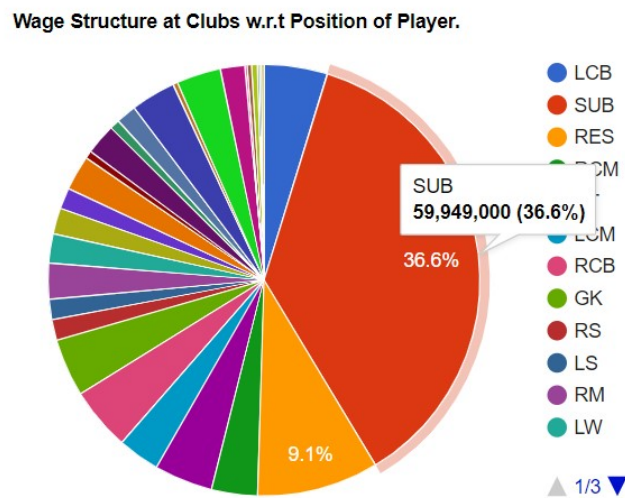


Figure 6: Wage Structure at Clubs