

Soccer Central Project Report

Nisha Poudel
CSCE
University of Nebraska Lincoln
Nebraska, USA
npoudel2@unl.edu

Nasimul Gani
CSCE
University of Nebraska Lincoln
Nebraska, USA
sgani2@unl.edu

Cuong Than
CSCE
University of Nebraska Lincoln
Nebraska, USA
cthan2@unl.edu

I. ABSTRACT

Soccer Central is a web application that uses the FIFA statistics of players to enable users to search and project the different attributes and abilities of players, country, club in the field of soccer. Careful structuring and visualisation of data retrieved from the extensive queries provide comprehensive understanding of a player or team. The application enables Fantasy League players to make an informed choice of players or team while playing in Career mode.

II. INTRODUCTION

With over 6 million players, Fantasy Premier League is the biggest Fantasy Football game in the world [4]. It is a platform for soccer enthusiasts around the world to come together in a friendly atmosphere to predict scores, create teams, and envision life as a soccer manager. However, this platform does not have many features that players would perhaps love to have. To illustrate, if somebody wants to build her dream team from the data available it can be a very tedious task. Thus, the purpose of the project is to create a web application that provides a User Interface for integration of all the attributes of a player, team or club based on FIFA statistics.

III. BACKGROUND

The web application is expected to be a place for looking up player information ranging from overall score to preferred foot and net worth. The data is structured and visualized using the attributes to provide a comprehensive understanding of a player or team based on numerical values of their attributes. For example, a feature is available for a user to find out the strongest team in terms of offense in the FIFA 2019 World Cup or the team that is stronger based on the combined overall of all its players.

A. Motivation for the Project

The motivation behind building the project is our love and interest for soccer and to provide a platform for other users who share our interest.

B. Objective of the project

- To build an web application in order to enable users with search facilities based on player name, nationality, Club and Position.

- To analyse the FIFA data set 19, specifically using the fields overall, potential, speed, strength, age, Nationality, Ball control, dribbling, position, players count and visualise the correlations.
- To analyse all the important attributes of a player by implementing k means clustering in order to categorise players with similar attributes.

C. Scope of the Project

The data set has been used to analyse the different dynamics of football player. Currently, the application only bases its features on the FIFA 2019 data set. The analysis was done from the data obtained after writing couple of aggregate queries. Following features have been included in the project

- Query the database to find the statistics and other attributes of players.
- Find all the players of a club/national team.
- Add filters on above to display players with certain attributes such as position/value/age etc.
- Correlate between age and overall rating, potential, speed and strength.
- Use k-mean clustering to segregate players with similar features.
- Correlate between Ball control and dribbling.

D. Database Schema and Fields

The following images describe the database schema of the application. It was generated using a NoSQL Database Schema Analyzer called Variety.js. The tables show each of the fields in the database along with the datatype of each field and the occurrences of a field in every document in the collection.

key	type	occurrences	percents
Acceleration	String	13725	100.0
Age	String	13725	100.0
Aggression	String	13725	100.0
Agility	String	13725	100.0
Balance	String	13725	100.0
Ball_Control	String	13725	100.0
Body_Type	String	13725	100.0
CAM	String (12423),null (1302)	13725	100.0
CB	String (12423),null (1302)	13725	100.0
CDM	String (12423),null (1302)	13725	100.0
CF	String (12423),null (1302)	13725	100.0
CM	String (12423),null (1302)	13725	100.0
Club	String (13533),null (192)	13725	100.0
Club_Logo	String	13725	100.0
Composure	String	13725	100.0
Contract_Valid_Until	String (12423),null (1302)	13725	100.0
Crossing	String	13725	100.0
Curve	String	13725	100.0
Dribbling	String	13725	100.0
FKAccuracy	String	13725	100.0
Finishing	String	13725	100.0
Flag	String	13725	100.0
GKDividing	String	13725	100.0
GKHandling	String	13725	100.0
GKkicking	String	13725	100.0
GKPositioning	String	13725	100.0
GKReflexes	String	13725	100.0
HeadingAccuracy	String	13725	100.0
Interceptions	String	13725	100.0
International_Reputation	String (12428),null (1297)	13725	100.0
Jersey_Number	String (12423),null (1302)	13725	100.0
Joined	null (2302),String (11423)	13725	100.0
Jumping	String	13725	100.0
LAM	String (12423),null (1302)	13725	100.0
LB	String (12423),null (1302)	13725	100.0
LCB	String (12423),null (1302)	13725	100.0
LCH	String (12423),null (1302)	13725	100.0
LDM	String (12423),null (1302)	13725	100.0
LF	String (12423),null (1302)	13725	100.0
LM	String (12423),null (1302)	13725	100.0
LS	String (12423),null (1302)	13725	100.0
LW	String (12423),null (1302)	13725	100.0

Fig. 1. Database Schema Page 1

LWB	String (12423),null (1302)	13725	100.0
Loaned_From	String (1069),null (12656)	13725	100.0
LongPassing	String	13725	100.0
LongShots	String	13725	100.0
Marking	String	13725	100.0
Name	String	13725	100.0
Nationality	String	13725	100.0
Overall	String	13725	100.0
Penalties	String	13725	100.0
Photo	String (13720),null (5)	13725	100.0
Position	String	13725	100.0
Positioning	String	13725	100.0
Potential	String	13725	100.0
Preferred_Foot	String	13725	100.0
RAM	String (12423),null (1302)	13725	100.0
RB	String (12423),null (1302)	13725	100.0
RCB	String (12423),null (1302)	13725	100.0
RCM	String (12423),null (1302)	13725	100.0
RF	String (12423),null (1302)	13725	100.0
RM	String (12423),null (1302)	13725	100.0
RS	String (12423),null (1302)	13725	100.0
RW	String (12423),null (1302)	13725	100.0
RWB	String (12423),null (1302)	13725	100.0
Reactions	String	13725	100.0
Real_Face	String	13725	100.0
Release_Clause	null (1268),String (12457)	13725	100.0
ST	String (12423),null (1302)	13725	100.0
ShortPassing	String	13725	100.0
ShotPower	String	13725	100.0
SkillMoves	String	13725	100.0
SlidingTackle	String	13725	100.0
Special	String	13725	100.0
SprintSpeed	String	13725	100.0
Stamina	String	13725	100.0
StandingTackle	String	13725	100.0
Strength	String	13725	100.0
Value	String	13725	100.0
Vision	String	13725	100.0
Volleys	String	13725	100.0
Wage	String	13725	100.0
Weak_Foot	String	13725	100.0
Work_Rate	String	13725	100.0
_id	ObjectId	13725	100.0

Fig. 2. Database Schema Page 2

IV. METHOD AND MATERIALS

A. Data Collection Method

Secondary data of all the attributes of players downloaded from Kaggle [6]

<https://www.kaggle.com/thec03u5/fifa-18-demo-player-dataset>

<https://www.kaggle.com/karangadiya/fifa19>

B. Scrubbing of Data

The data was initially in a CSV format. Importing it in MS Excel caused loss of some player name that contained diacritics. It was, thus, opened in Visual Studio with a CSV formatter. Upon validation, many fields had inconsistent naming styles and some fields were improperly formatted. This issue was dealt with using python scripts to replace such inconsistencies. Moreover, some of the fields like "real face" and "Specials" were removed as those were deemed unimportant for the purposes of this project.

C. Computer Methodologies

1) Front-end Development:

- **JavaScript:** JavaScript was used as a basic programming language as it does not provide low level access to memory and its capabilities depend upon the environment in which it is running. JavaScript is dynamic in terms of web page user-interaction, web server integration and web page manipulation. Node.js supports the functions that allow java script to read/write arbitrary files [1].

- **AngularJS:** User Interface was designed using AngularJS. AngularJS provides real-time DOM manipulation which allows us to create interactive pages that can render dynamic data. Consequently, AngularJS was chosen as the front-end structural framework for creating the dynamic web application. As this framework has both client-side model-view-controller (MVC) and model-view-view-model (MVVM) architectures, it simplifies the development and testing of the web application. It made use of the MongoDB database as a API endpoint which allowed secure procurement of data [9].

2) Back-end Development:

- **mongoDB:** We used the the mongoDB database to create a web app to query the database from the User Interface. MongoDB is a NoSQL database program that uses JSON-like documents with schema. It is a cross-platform document-oriented database program.It was used in our project for its simplicity and flexibility. NoSQL database is being popular in the company that has to deal with big data and real time web application [10]. It enables user to store and retrieve data in the form of collections, documents and fields unlike tabular relation used in relational database. MongoDB was used because of the relation-less nature of our data.

3) **Data Visualisation:** Data Visualisation was used to make our results more realistic and lively. D3.js was used for Data visualisation. D3.js is a JavaScript library for manipulating

documents and uses HTML, SVG, and CSS to render graphics on HTML elements [3]. D3 provides user with capabilities of modern browsers.

4) *K means Clustering*: We used K-means clustering algorithm to partition players into group according to their ability[7]. The basic ideal of this method is treating each player as a multi-dimension point before partitioning them to each group. The object is to minimize the total distance between each point and its cluster's center. K-means algorithm's expected complexity is asymptotically polynomial, which makes it fast enough to implement for large dataset like FIFA. For new and young players, we can also predict their future position by finding their nearest center. In addition, we choose K-means because it works efficiently with the numeric and well-partitioned data like the set of player and their abilities.

5) *User Interface Features*: The homepage of the application is capable of displaying short description of each player in the data set using a find() query. Search feature is based on player attributes. Users can further search for players using Name, Club Name, Nationality or Position as filters. The application also has links to display and interact with the D3.js Charts.

V. RESULTS

Some of the prominent results of the Analysis of data set were as follows

A. Search Features

This Application enables user to search players based on different parameters like player name, club, Nationality. Some of the examples are provided below.

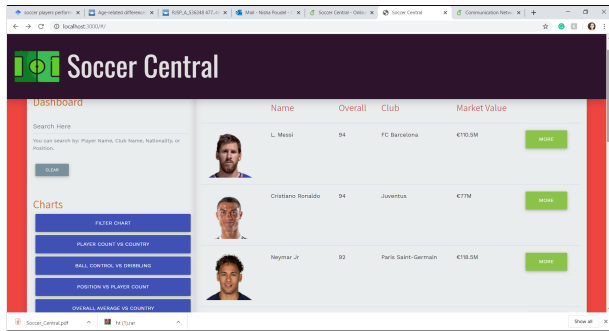


Fig. 3. Frontpage of the Webpage

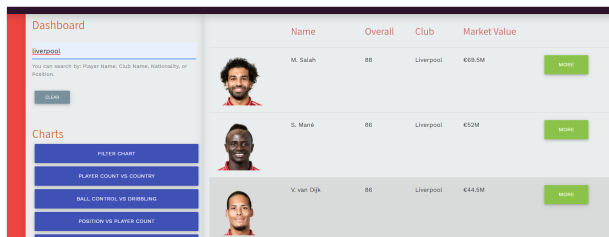


Fig. 4. Example when Search by Club



Fig. 5. Example when Search by Player Name

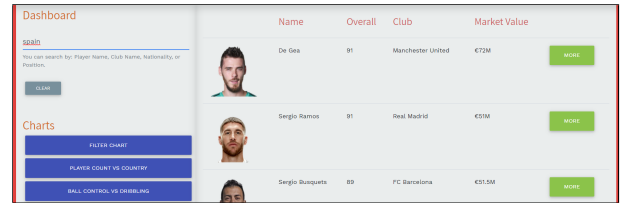


Fig. 6. Example when Search by Nationality Name

B. Number of players from a country

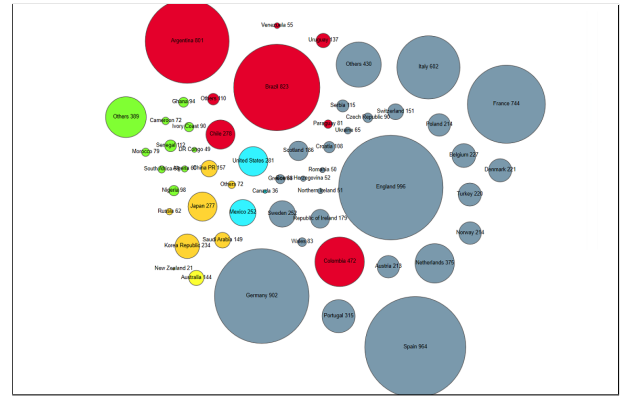


Fig. 7. Number of FIFA Players from Nations

As evident in Fig 7, Most of the FIFA 19 players come from England (996), Spain(964), and Germany(902) followed by South American country Brazil(823). Although, Asia is the biggest continent, the representation of players is proportionally very low with the number of players at 674.

C. Overall strength of Country as a Soccer Team

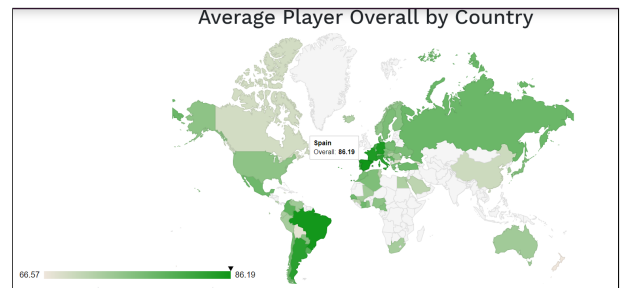


Fig. 8. Overall strength of Country as a Soccer Team

As evident in Fig 8, Spain is theoretically the strongest country as a team based on the 2019 data set. Spain is followed by Brazil, Germany and France. The darkness of the color green is synonymous to the strength of the Country's National Soccer Team.

D. Relation of age with different parameters

Age has always been the matter of interest in the soccer world. Most of the people believe that potential, speed, overall capability decrease with age. Our attempt to compare age with different parameters has shown that it is indeed true even in the most competitive game in the world.

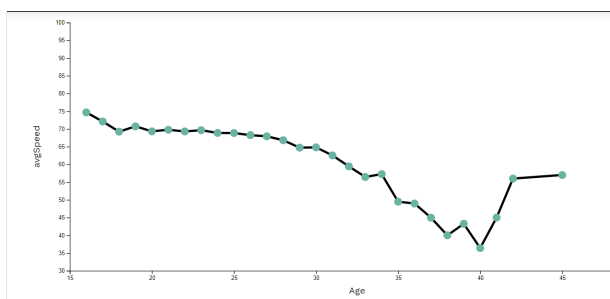


Fig. 9. Age Verses Speed

1) *Age versus speed*: As evident in Fig 9, the speed of the players is at highest in their teens. It decreases slightly and is constant in their 20s. It then continues to decrease in their 30s with slight increase in between(34) which is an anomaly due to the lack of data for players above that age. A probable reason could be that most players move out of the international limelight after 34 or they retire.

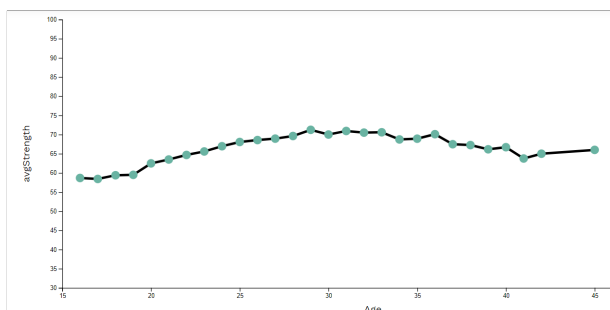


Fig. 10. Age Versus Strength

2) *Age versus strength*: As evident in Fig 10, The average strength of player seem to be at highest in age 29 followed by 36. The strength of a player seem to slight steady increase in their 20s and remain somewhat stagnant in their 30s.

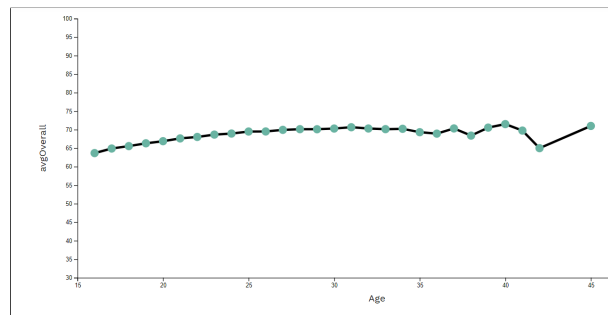


Fig. 11. Age Versus Overall

3) *Age versus overall*: As evident in Fig 11, The overall is the average of all the attributes of a player. The graph shows that the overall capability of a players is static from 25 to 34, slightly decrease in 35 and re-increase in 36 years and decrease abruptly.

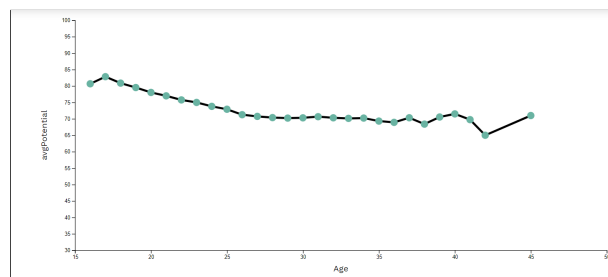


Fig. 12. Age Versus Potential

4) *Age versus potential*: As evident in Fig 12, The potential is the projection of average of all the utmost potential attributes of a specific player. The result does not have much difference except for during early 20s when players start off with high potential unlike their overall.

E. Correlation of Ball Control and Dribbling

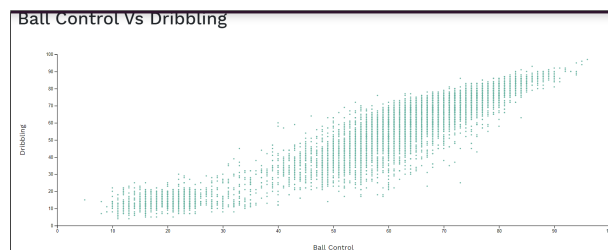


Fig. 13. Ball Control Versus Dribbling

As evident in Fig 13, The graph is somewhat scattered till the Ball control count is 40. However, after 40 count, the graph is linear depicting the relationship as somewhat directly proportional.

F. Position and Players count

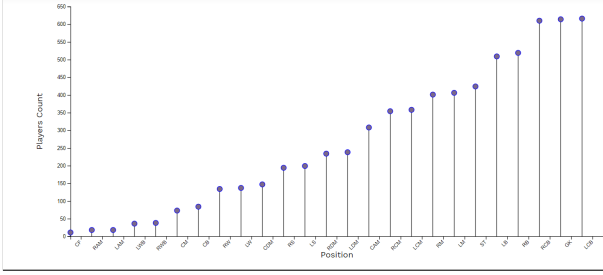


Fig. 14. Number of players playing in a specific position

As evident in Fig 14, The number of players playing in LCB, GK and RCB are the highest while CF, RAM AND LAM is lowest. This could indicate the need for players to be well-rounded when they are scouted to maximize their chances to play professionally. Most players change their preferred position when they sign for a club and thus the graph tends to show the range of options in players for each position.

G. Number of clusters

For each number from 1 to 20, we used k-means algorithm and calculated the average distortion of our model. The result can be seen in Fig 15

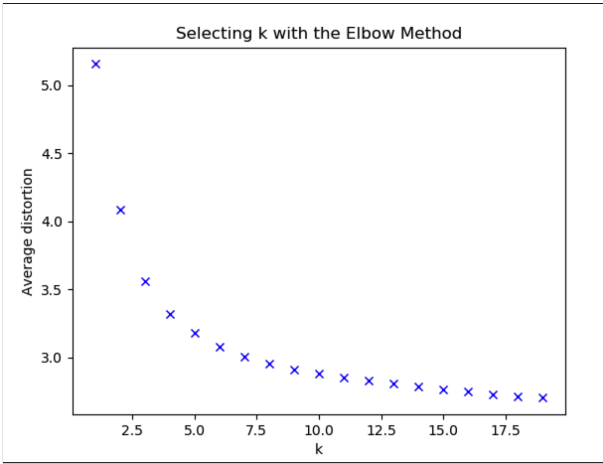


Fig. 15. Learning Curve

Using the elbow method, we can find out the number of cluster is 8. This mean we can divide the set of players into 8 groups, each group corresponds to one (or some) position(s) in the pitch. For example, in our result, group 2 contains all goals keepers while group 1 contains defenders with excellent physical abilities.

VI. DISCUSSIONS

The application is expected to be helpful for all soccer enthusiasts to obtain general information on players and could help many people who look for player information before Fantasy leagues and playing Career Mode on FIFA video

games. This website could provide a quick reference to such users.

A. Future Plans

Using multiple datasets from previous years would provide more insight and help track change in player attributes.

B. Challenges

Firstly, Scrubbing and Analysis of data took long due to the many inconsistencies in the data set described earlier. Secondly, a lack of experience in Web Development added to the learning curve among the team members. Finally, integrating all the different frameworks used in the application amassed to the challenges faced in the project.

REFERENCES

- [1] *An Introduction to JavaScript*. URL: <https://javascript.info/intro>. (accessed: 13.12.2019).
- [2] *Bubble Chart*. URL: <https://observablehq.com/explore>. (accessed: 12.03.2019).
- [3] *Data Driven Documents*. URL: <https://d3js.org/>. (accessed: 13.12.2019).
- [4] *Fantasy - Documentation*. URL: <https://fantasy.premierleague.com/>. (accessed: 09.25.2019).
- [5] *GeoChart*. URL: <https://developers.google.com/chart/interactive/docs/gallery/geochart>. (accessed: 12.09.2019).
- [6] *Kaggle*. URL: <https://www.kaggle.com/karangadiya/fifa19>. (accessed: 10.09.2019).
- [7] J. MacQueen. "Some methods for classification and analysis of multivariate observations". In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*. Berkeley, Calif.: University of California Press, 1967, pp. 281–297. URL: <https://projecteuclid.org/euclid.bsmsp/1200512992>.
- [8] *Stackflow*. URL: <https://stackoverflow.com/questions/14812314/google-geochart-all-countries>. (accessed: 12.8.2019).
- [9] *What Is AngularJS?* URL: <https://docs.angularjs.org/guide/introduction>. (accessed: 13.12.2019).
- [10] *What Is mongoDB?* URL: <https://www.mongodb.com/what-is-mongodb>. (accessed: 13.12.2019).