

# **PLAYING WITH SOCCER DATA FOR DATA ANALYSIS & PREDICTION**

A

**PROJECT SYNOPSIS**

FOR PROJECT

**BACHELOR OF TECHNOLOGY  
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

SUBMITTED BY

<b>ANUJ SHARMA</b>	<b>17ETCCS002</b>
<b>BHAVINI MITTAL</b>	<b>17ETCCS005</b>
<b>NIMISHA SHARMA</b>	<b>17ETCCS039</b>
<b>VIDIT JAIN</b>	<b>17ETCCS057</b>
<b>PARSHAVI BOLYA</b>	<b>17ETCCS041</b>
<b>KARTIK BOKADIA</b>	<b>17ETCCS019</b>

**UNDER THE GUIDANCE OF  
Mr. Gaurav Kumawat**



**Techno India NJR Institute of Technology  
2020-2021**

## **Playing with Soccer data for data analysis & prediction**

Prerequisite: R Studio basic configuration settings & student must know introduction to R Programming

### **Introduction:**

This project will drive you to build a dashboard for creating overall rating. The data given is having following attributes

**About Data:** The data is .sqlite format with following tables –

- a. Country ()
- b. League
- c. Match
- d. Player
- e. Player\_Attribute
- f. Team

As the data provided is not explicit/overt to put analytical insights over hence as III year student, this project will enhance skills that how to prepare the data before putting analytical insights by logical expressions. Visual Analytics based dashboard is expected for following queries -

Q – 1 Creating a Correlation between Overall rating & any other factors in order to identify that whether that factor is affecting the overall rating

Q – 2 Creating a dashboard to identify top n players on the basis of one more more (5 factors maximum) apart from overall rating to cross examine the overall rating?

Q – 3 Predict the number of goals of particular player in upcoming match.

Q – 4 Analyze the win % in home ground vs win % not in home ground vs draw for top 20 players determined by overall rating.

### **Architecture Flow:**

1. User imports the .CSV file in R Studio (Latest Version)

2. As per the required analytical insight, using logical expressions.
3. At this phase, any two visual insights is expected like Bar Graph, Line Graph, Pie Chart, Area Chart etc. over the dashboard against the selected query from drop down box.

**Included Components:**

Following libraries have been included –

- a. [dplyr](#)
- b. [RSQLite](#)
- c. [radarchart](#)
- d. [tidyr](#)

**Featured Technologies:**

- g. [R Programming](#)