INTRODUCTION TO BIG DATA

PROGRESS REPORT - EUROPEAN SOCCER ANALYSIS - GROUP 10

SAHAJ GANDHI | ANIKET GIRIYALKAR | ADITYA KALYAN JAYANTI

CSCI 620 | 04/04/2018

○ **PROJECT MOTIVATIONS**

Soccer is a global sport. It is played in almost every country in the world, and soccer competitions have global viewing numbers that blow all other sports (and even the Olympics) out of the water. The 2014 FIFA World Cup Final has been by far the most viewed sports event in recent times, with over 700 million viewers. It makes sense that its top professional leagues are different and superior in structure than those of any other sport in the world. European Soccer is more competitive than any other sports leagues because of the format of the leagues, the multiple competitions, and the intense rivalry between not only each team, but in each league.

To understand what makes European Soccer the most competitive league in the world, we have chosen to examine this database, which contains information about over 2500 matches, about 10000 players representing over 11 European Countries playing in the top tiers of their championships.

○ **LIST OF THE TASKS THAT THE TEAM HAS DONE SO FAR**

    i.   Found database in sqlite format
    ii.   Opened database in R/Rattle and pre-processed the data - handled empty/missing values and other similar issues
    iii.   Normalized the database tables to be 4NF compliant
    iv.   Opened the pre-processed database in MS SQL(SSMS)

○ **LIST OF TASKS THAT NEED TO BE DONE BEFORE THE DUE DATE**

    i.   Perform transformations on the database using R/Rattle
    ii.   Push database from SQLite format to MS SQL supported-format, and parse the data from the tables into the normalised form of database
    iii.   Using queries in SQL come up with interesting relationships and insights
    iv.   Perform meaningful analysis and come up with conclusions
    v.   Draw useful correlations, relations, and other useful conclusions from the database available in R/Rattle and SSMS(MS SQL)