PROJECT PROPOSAL

Name of the project:

Ranking system of top 100 football teams in European top 5 leagues of the year 2017-18.

&

Ranking system of top 100 strikers in European Top 5 leagues of the year 2017-18.

Name: *SAUNAK MITRA*

Roll No: *Bsc 867*

Department of Statistics,6th Sem

Objective of the project:

This project is mainly conducted for the analysis of the ranking system of top 100 European football clubs on the basis of matches played ,won ,drawn, goals scored, goals conceded, goal difference and points won. And for that, a statistic(self made) has been developed, in order to rank the team according to the Z-scores .Again , Principal Component Analysis has been conducted to check the possible ranking of the teams, and to verify the accuracy of our statistic.

Simililarly, using another statistic we rank the top 100 European strikers according to the Z scores, and again we perform Principal Component Analysis to verify the accuracy of our statistic in ranking the strikers.

Data Collection:

Kaggle: *Clubs data has been collected.(Club names, Matches played, won, drawn, goals conceded, goals scored, goal difference, points won)*

Transfermarket: *Player's data has been collected.(Player's name, matches played, goals scored, assists provided)*

Solving and analysing the data:

- *First, we collect the data in an excel sheet.*
- *Next, we use mathematical formulas to express the individual Z scores of teams and players according to our statistic.*
- *Next, we open the R studio where we open the Excel file to do the ranking of players and teams according to the Z scores.*
- *Next, we open Minitab where we open the excel file and perform the Principal Component Analysis and generate the principal components, eigen values and scree plot.*
- *Next we again open excel to generate the PCA scores according to the principal components.*
- *Next we again open that excel file in R studio to find the rankings of the teams and players according to Principal Component Analysis.*
- *Finally, in Minitab we plot z score against PCA score.*

Inference:

In inference we detect how Z score and PCA score are correlated from the scatterplot.