CS432/532: Final Project Report

Project Title: Data analysis of transfer investment of top tier soccer clubs in Europe (2010-2019)

Team Member(s): Bo Yan

I. Problem

With a combination of a dataset of transfer fee investments of top tier soccer clubs in Europe from 2010 to 2019, a data source of famous soccer players involved in these transfers, a data source of value and revenue of these clubs, and a data source of competition performance of these clubs and these players, a self-created database is implemented to perform analysis on transfer investment of these top tier soccer clubs, including normal data comparing and finding, investment return and overall performance of leagues, clubs and players, etc.

II. Software Design and Implementation

A. Software Design and NoSQL-Database and Tools Used

NoSQL-Database Used:

MongoDB

Programming Language Used:

Java

Software Design:

1. Database Implementation:

(1) Research and Data Gathering:

Different datasets or data sources from Internet are searched and gathered, from each of them, for project's concern, a small cut down set is kept for only top tier clubs and famous players, then all kept sets of data are translated into fields or sets of fields of a document.

(2) Database Creation:

All MongoDB style documents and collections are manually written and inserted into a self-created MongoDB database called project3 by me.

(3) Database Usage:

mongo < project3.js

2. User Interface Software Design:

(1) Connect MongoDB to Java

The latest version of the mongo-java-driver is downloaded and installed in the Java project classpath, and packages from com.mongo are imported and used in Java code to connect to, and CRUD the MongoDB database.

(2) User Interface Design

A terminal style interface is implemented using Java. When the program is started, user will see from the terminal a list of all supported queries and input of choice of query can be typed by user, then the result of selected query will be output in the terminal.

B. Supported Queries

Basic Queries:

Note although all basic queries can be directly answered by using the group and certain aggregation function supported by MongoDB itself, small data models are still implemented in Java project for user interface convenience.

1. Find the club with most transfer investment fee from 2010 to 2019

2. Find club which earned most number of competition titles from 2010 to 2019

3. Find player who transferred with the highest transfer fee from 2010 to 2019

Sophisticated Queries:

Note for all sophisticated queries, data models are implemented in Java project, after database is connected to Java and all needed data are read and stored into certain data model classes, deeper analysis are performed.

4. Analysis on a single transfer investment return

For this query, three parts of analysis will be performed on all single transfer investment:

(1) Analyze and find the single transfer investment that gives the related club hignest return at the competition level after the player joined the club. In the data model, different weights will be given for different competition results, return period of each transfer will be calculated, investment fee spent will be considered and finally the return will be computed and sorted to find the transfer with highest competition level return.

(2) Analyze and find the most successful player of the entire return period after transferred.

(3) Analyze and find the most successful player per year after transferred.

5. Analysis on transfer market trend and performance from 2010 to 2019 for clubs and leagues

For this query, two parts of analysis will be performed on the entire transfer market:

(1) Analyze and find the most successful club from 2010 to 2019, this will consider on two levels, club value and transfer performance in the entire 10-year-window.

Club value will be calculated based on the classic asset-revenue-ratio model and transfer performance model will consider all transfer performed, all player performance after the transfer, player's transfer fee, successful transfer ratio of all transfers, transfer investment fee, etc.

(2) Analyze and find the most successful league from 2010 to 2019, this will consider both competition results on UEFA level competitions and transfer market trend between leagues.

III. PROJECT OUTCOMES

GitHub:
https://github.com/byan13/cs532Project3.git
Demo Video:
https://youtu.be/si1tCpB1A7c

Presentation Slides:
https://drive.google.com/file/d/1PiuXq8bsG7ZC8OjuWFKc_OUcT1N7G1rv/view?usp=sharing

REFERENCES

[1] MongoDB 4.2 Manual
https://docs.mongodb.com/manual/

[2] mongo-java-driver download and installation
https://search.maven.org/classic/#search%7Cgav%7C1%7Cg%3A%22org.mongodb%22%20AND%20a%3A%22mongo-java-driver%22

[3] Data source for club values and revenues
https://en.wikipedia.org/wiki/Forbes%27_list_of_the_most_valuable_football_clubs

[4] Data source for players and expensive transfers
https://en.wikipedia.org/wiki/List_of_most_expensive_association_football_transfers

[5] Dataset for tansfer fee
https://www.statista.com/statistics/957904/europe-transfer-fee-spending-by-league

[6] Additional data source for clubs
https://en.wikipedia.org/wiki/Real_Madrid_CF
https://en.wikipedia.org/wiki/FC_Barcelona
https://en.wikipedia.org/wiki/Manchester_United_F.C.
https://en.wikipedia.org/wiki/FC_Bayern_Munich
https://en.wikipedia.org/wiki/Manchester_City_F.C.
https://en.wikipedia.org/wiki/Chelsea_F.C.
https://en.wikipedia.org/wiki/Arsenal_F.C.
https://en.wikipedia.org/wiki/Liverpool_F.C.
https://en.wikipedia.org/wiki/Tottenham_Hotspur_F.C.
https://en.wikipedia.org/wiki/Juventus_F.C.
https://en.wikipedia.org/wiki/Paris_Saint-Germain_F.C.
https://en.wikipedia.org/wiki/Atlético_Madrid
https://en.wikipedia.org/wiki/Borussia_Dortmund
https://en.wikipedia.org/wiki/FC_Schalke_04
https://en.wikipedia.org/wiki/Inter_Milan
https://en.wikipedia.org/wiki/A.S._Roma
https://en.wikipedia.org/wiki/West_Ham_United_F.C.
https://en.wikipedia.org/wiki/A.C._Milan
https://en.wikipedia.org/wiki/Everton_F.C.