R기반의 데이터 분석 프로젝트 기획안

|  |  |
| --- | --- |
| **팀 명** | 9조\_발롱도르 |
| **팀 원** | 김선홍, 장소영 |
| **프로젝트 명** | 피파 2021 선수 능력치 예측과 선수 리뷰 |
| **프로젝트 주제 및 내용**  해결하고자 하는 문제 최종 산출물의 청사진 | FIFA는 실제 축구 게임을 반영하는 대표적인 게임이다. EA sports는25명의 데이터 수집가, 400명의 외부 데이터 기고가로 구성된 팀을 통해 선수의 데이터를 최신 상태를 반영할 수 있게 만든다. 또한 6000명 이상의 자원봉사자로 구성된 팀이 선수의 데이터 베이스를~. 이 게임은실제의 프로 축구 선수의 능력과 ,, 와 같은 다양한 정보를 반영하고 있는데,  저희 조는 게임의 데이터 셋을 통해 평소에 궁금했던 다음의 주제를 R을 통해 분석해 볼 예정입니다.  **1. 수비력이 좋은 팀, 공격력이 좋은 팀을 찾아보고, 2021 -2022 시즌의 실제 득점, 실점 과의 관계**  **2. 지난 해 발롱도르 순위와 선수 능력치의 상관관계**  **3. 챔피언스 리그 배당률과 게임에서의 급료와의 상관관계**  **4. 리그 점수와 선수 등급과의 상관관계**  **5. 왼발잡이가 더 잘하나?** |
| **주제 선정 배경**  주제 선정의 정당성  산출될 결과의 유용성 | 평소 축구에 관심이 많아서 주제 선정에 대해 대화를 해보다 좋아하는 분야에 대해 세미 프로젝트를 진행하는 것이 재밌을 것 같아 선정하게 되었습니다. |
| **프로젝트 수행 일정**  수행 일정 | 아이디어 도출 : 03.10~03.12  자료 수집 및 전처리 : 03.13~03.17  데이터 분석 : 03.18~03.22  결론 도출 및 보고서 작성 : 03.22 |

3월 18일 오전까지 제출

**1. 프로젝트 주제 선정 배경**

**2. 팀 구성 및 역할**

**3. 데이터 수집, 전처리 (어디서 수집했는지, 어떤 분석, 분석 결과를 완료 보고서에 넣어야)**

**, 지도, 상관분석 등등**

*k리그 데이터 포털 : 팀 스탯, 선수 스탯*

*주류 판매와 관중*

*xG 스탯과 선수 몸값*

*관중과 경기장의 입지의 상관관계*

*Market Basket Analysis : R에서는 MBA / 연관 규칙 마이닝에* ***\*arules\**** *패키지를 사용할 수 있습니다 . 또는 내 축구 이적과 같이 거래 순서가 중요한 경우* ***\*arulesSequences\**** *패키지를* ***\*사용해야***

*선수의 이적이 라이벌 팀간에는 거의 이루어지지 않는지?*

[*https://github.com/longhowlam/soccertransfers*](https://github.com/longhowlam/soccertransfers)

*피파게임에서 선수의 능력치와 선수의 스탯 간의 관계*

*피파22 에서의 능력치 예상*

*피파 / FM / 위닝일레븐의 선수 능력치 비교와 가장 잘 반영하는 게임은 무엇인지?*

[*http://sportstat.kr/%ed%8c%80-%ea%b8%b0%eb%8c%80-%eb%93%9d%ec%a0%90teg-2%ed%8e%b8/*](http://sportstat.kr/%ed%8c%80-%ea%b8%b0%eb%8c%80-%eb%93%9d%ec%a0%90teg-2%ed%8e%b8/)

*평일 경기와 주말 경기의 관중?*

*배당률과 승패의 상관관계*

*경륜 경정 경마 예측, 유로 2021 베팅*

*홈 팀의 승률이 더 높은가*

*날씨가 승패에 미치는 영향*

*골키퍼의 이적이 덜 활발한가?*

*이적료 총합과 구단의 순위*

<https://hjjeon27.tistory.com/17>

<http://www.inven.co.kr/webzine/news/?news=110917&site=fifaonline3>

<https://www.superookie.com/jobs/5eaf938c8b129f585f2d9f50>

<입학 성적과 순위와의 상관관계 >

<OTT 서비스> : 각 사이트가 어떤 영상 컨텐츠를 제공하는지, 평은 어떠한지 크롤링. …? 상관관계는 ?

<넷플릭스 > 캐글에 올라오는 데이터 셋을 보면서 궁금한거 하나씩 분석해보는 식 ..? 평점이 높은 건 무엇인지. 소셜에 언급이 많은 것이 평점이 높은지..?

2021 예선

4대리그 시즌마다 탑 상위 (공격, 수비) 수비가 중요한지 공격이 중요한지

플레이어 리뷰,, 워드 클라우드로 가져오기 ..~

데이터 셋 ~

최고 등급의 선수를 보유한 클럽, 국가 대표팀의 분석 ( 그냥 어느 팀인지 찾아보는 식으로)

**수비력이 좋은 팀, 공격력이 좋은 팀을 찾아보고, 2021 -2022 시즌의 실제 득점, 실점 과의 관계**

**지난 해 발롱도르 순위와 선수 능력치의 상관관계**

**챔피언스 리그 배당률과 게임에서의 급료와의 상관관계**

**리그 점수와 선수 등급과의 상관관계** (라리가에서 뛰는 상위 20명, EPL 에서 뛰는 상위 20명 --- 을 대상으로 )

선수 포지션을 구분해주는 주요한 변수가 무엇인지 PCA 분석

**왼발잡이가 더 잘하나 ..?**

왼발잡이 베스트 11 오른발잡이 베스트 11

공격력이 높으면 수비력은 낮은가의 가설검정

수비 스탯이 높았을 떄 정말 우승권에 가까워지나?

연령과 잠재력 사이의 상관 관계

강등권과 골키퍼 스탯의 상관관계

xg 값과 골의 상관관계