스포츠 게임 회사의 데이터 기획과 분석

설득을 위한 데이터분석

강사 소개

최윤식 https://data-so-hard.tistory.com/

- 비투지게임즈 (2022년 1월 – 현재)
- 스포츠데이터에볼루션 (2020년 9월 – 2021년 11월)
- 더그아웃 매거진 (2018년 7월 – 2020년 8월)
- 세종대학교, 교육학 (2011년 3월 - 2019년 2월) - 학사

기타 활동

- 야구공작소 (2017년 2월 – 2018년 5월, 팟캐스트 야자수 PD)
- 네이버 라디오 MLB쇼 (2019년 1월 – 2019년 12월, PD)

목차

1. 판타지 게임
2. 데이터 기획 과정
3. 데이터 수집
4. 데이터 분석을 통한 데이터 가공
5. 데이터 직무에서 중요한 것들
6. 질문 Q&A

1. 판타지 게임

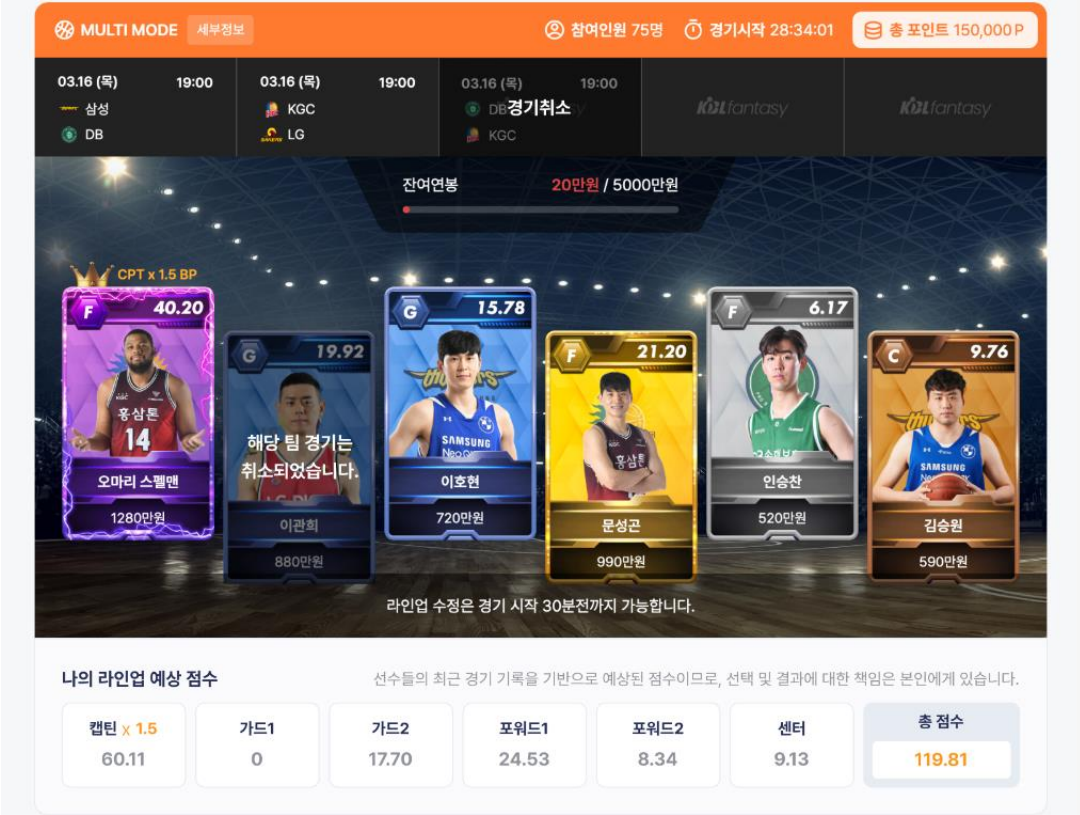
출처 - https://fantasy.kbl.or.kr/

판타지 게임이란?

실제 스포츠 경기를 기반으로 유저 간의 대결을 펼치는 게임

미국 판타지 게임 3대장

- DraftKings
- FanDuel
- Yahoo Fantasy



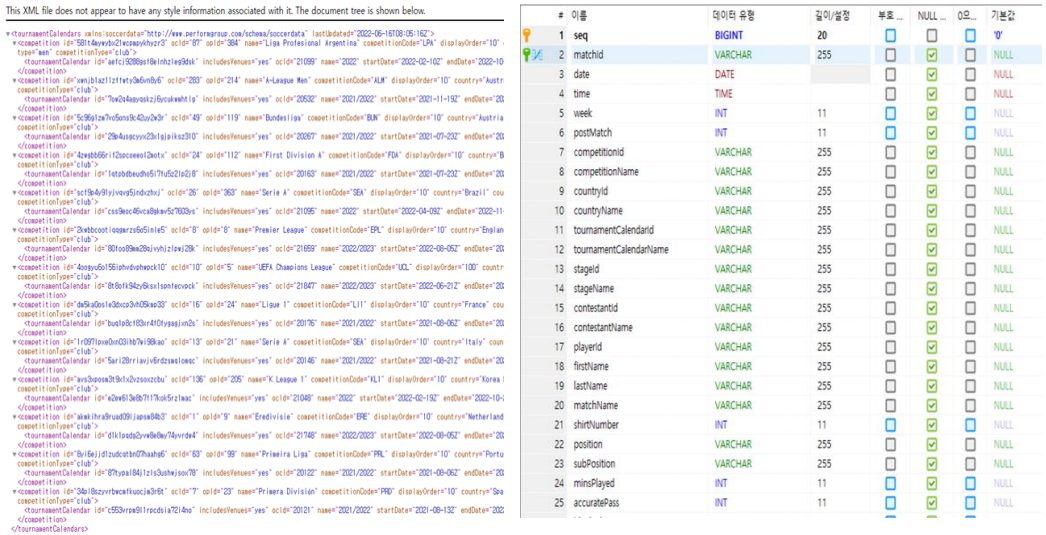
2. 데이터 기획 과정

데이터 분석가도 전체적인 흐름을 파악해야 한다!

3. 데이터 수집

출처 - creazilla

데이터 업체와 계약



- 계약 후 XML/JSON형태나 뷰테이블(View table)로 데이터를 제공해준다.

3. 데이터 수집

```python

for i in range(len(df)):
    try:
        df2 = df.loc[i]['id']
        url1 = "https://api.performfeeds.com/soccerdata/matchstats/7tzd7uu2opcq1jlnz0yw60qai/"
        url2 = "?_fmt=json&detailed=yes&_rt=b"
        url_r =  url1 + df2 + url2
        url_get = requests.get(url_r)
        text = url_get.text
        data = json.loads(text)
        df3 = pd.json_normalize(data['liveData'])
        #print(df3)

        lineUp = df3.loc[0]['lineUp']
        #print(lineUp)

        for dict_data in lineUp:
            players = dict_data['player']
            print("총 선수/코칭스탭 수:", len(players),"몇번째",i)

            for player in players:
                new_sr = pd.Series([])

                new_sr['playerId'] = player['playerId']
                new_sr['firstName'] = player['firstName']
                new_sr['lastName'] = player['lastName']
                new_sr['matchName'] = player['matchName']
                new_sr['position'] = player['position']

                try:
                    new_sr['subPosition'] = player['subPosition']
                except:
                    new_sr['subPosition'] = None

                for stat_dict in player['stat']:
                    new_sr[stat_dict["type"]] = stat_dict["value"]
```

- JSON, XML으로 제공 받는 데이터는

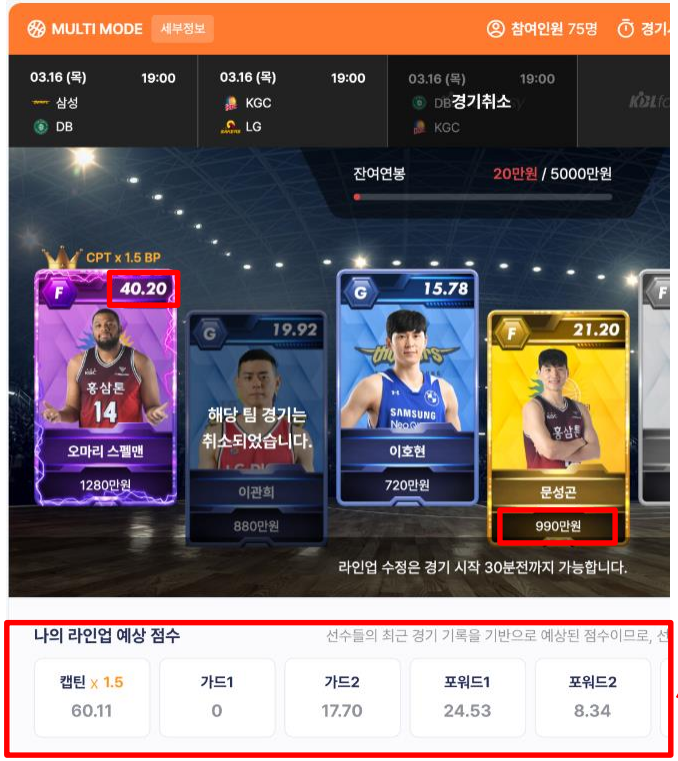
  크롤링 혹은 파싱 작업이 필요하다.

#	이름	데이터 유형	길이/설정	부호 ...	NULL ...	0으...	기본값
1	seq	BIGINT	20	☐	☐	☐	'0'
2	matchId	VARCHAR	255	☐	☑	☐	NULL
3	date	DATE		☐	☑	☐	NULL
4	time	TIME		☐	☑	☐	NULL
5	week	INT	11	☐	☑	☐	NULL
6	postMatch	INT	11	☐	☑	☐	NULL
7	competitionId	VARCHAR	255	☐	☑	☐	NULL
8	competitionName	VARCHAR	255	☐	☑	☐	NULL
9	countryId	VARCHAR	255	☐	☑	☐	NULL
10	countryName	VARCHAR	255	☐	☑	☐	NULL
11	tournamentCalendarId	VARCHAR	255	☐	☑	☐	NULL
12	tournamentCalendarName	VARCHAR	255	☐	☑	☐	NULL
13	stageId	VARCHAR	255	☐	☑	☐	NULL
14	stageName	VARCHAR	255	☐	☑	☐	NULL
15	contestantId	VARCHAR	255	☐	☑	☐	NULL
16	contestantName	VARCHAR	255	☐	☑	☐	NULL
17	playerId	VARCHAR	255	☐	☑	☐	NULL
18	firstName	VARCHAR	255	☐	☑	☐	NULL
19	lastName	VARCHAR	255	☐	☑	☐	NULL
20	matchName	VARCHAR	255	☐	☑	☐	NULL
21	shirtNumber	INT	11	☐	☑	☐	NULL
22	position	VARCHAR	255	☐	☑	☐	NULL
23	subPosition	VARCHAR	255	☐	☑	☐	NULL
24	minsPlayed	INT	11	☐	☑	☐	NULL
25	accuratePass	INT	11	☐	☑	☐	NULL

- 파싱을 완료한 데이터를 토대로 테이블을 구축한다.

4. 데이터 분석을 통한 데이터 가공

- 수집한 데이터를 기반으로 기획에 맞춰 분석을 통해 게임에서 활용할 데이터로 가공한다.



프로젝션 산출 과정

```sql
/* 기본 프로젝션 공식 : 1분당 종목 프로젝션 * 평균 출전시간 */
SELECT CASE WHEN A.season_code = '39' THEN '2021-22'
 WHEN away_team = '37' THEN '2020-21'
 WHEN away_team = '35' THEN '2019-20' END AS SEASON, RANK() OVER (ORDER BY ProjPTS DESC) AS RANK, B.kname,
CASE WHEN B.team_code = '06' THEN 'KT'
 WHEN B.team_code = '10' THEN 'MOBIS'
 WHEN B.team_code = '15' THEN 'DB'
 WHEN B.team_code = '30' THEN 'ORION'
 WHEN B.team_code = '35' THEN 'SS'
 WHEN B.team_code = '50' THEN 'LG'
 WHEN  B.team_code = '55' THEN 'SK'
 WHEN  B.team_code = '60' THEN 'KCC'
 WHEN  B.team_code = '64' THEN 'KOGAS'
 WHEN  B.team_code = '70' THEN 'KGC' END AS TEAM, A.pos, B.back_num, (SUM(((A.score * 1)
+((A.d_r + A.o_r) * 1.3)
+(A.a_s * 1.6)
+(A.s_t * 2)
+(A.b_s * 2)
+(A.threep * 0.7)
+(A.gd * 1)
+(CAST(A.t_o AS INTEGER) * -1)
+((A.score >= 10 AND (A.o_r + A.d_r) >= 10 AND A.a_s >= 10) OR
```

player_daily_list (198r × 9c)

SEASON	RANK	kname	TEAM	pos	back_num	BPMin	AVGMin	ProjPTS
2021-22	1	저멀 워니	SK	C	34	1.45700	32	45.555555532
2021-22	2	오마리 스펠맨	KGC	FD	14	1.41895	31	43.888372080
2021-22	3	아셈 마레이	LG	C	50	1.38217	31	41.986792451
2021-22	4	라건아	KCC	C	20	1.32212	29	38.194444425
2021-22	5	머피 할로웨이	ORION	FD	14	1.33819	29	37.687755079
2021-22	6	캐디 라렌	KT	C	25	1.35578	26	35.173584900
2021-22	7	라숀 토마스	MOBIS	FD	25	1.45182	24	34.381818169
2021-22	8	토마스 로빈슨	SS	FD	0	1.47522	23	33.339999979
2021-22	9	앤드류 니콜슨	KOGAS	FD	44	1.46111	23	33.188095235
2021-22	10	최준용	SK	FD	2	1.14973	28	31.809259248
2021-22	11	아이제아 힉스	SS	FD	4	1.32870	24	31.058333328
2021-22	12	조니 오브라이언트	DB	FD	8	1.19732	26	30.976923069
2021-22	13	허웅	DB	GD	3	0.95342	30	28.390740740
2021-22	14	이재도	LG	GD	4	0.87390	32	27.964814784
2021-22	15	허훈	KT	GD	2	0.97015	29	27.867499979
2021-22	16	이대성	ORION	GD	43	0.89864	31	27.731999972
2021-22	17	안영준	SK	FD	8	0.94644	30	27.587037022
2021-22	18	이승현	ORION	FD	33	0.80556	34	27.187499981
2021-22	19	김낙현	KOGAS	GD	4	0.90006	30	27.001886790
2021-22	20	양홍석	KT	FD	11	0.92794	29	26.892307683

4. 데이터 분석을 통한 데이터 가공

데이터 가공은 어떻게?

5. 데이터 직무에서 중요한 것들

1. 커뮤니케이션
2. 풍부한 도메인 지식
3. 문서 작성 능력
4. 어려운 것보다 기본에 충실하라
5. 맺고 끊음을 할 수 있는 결단력

Q&A