

삼성 오픈소스 컨퍼런스

재미있게 경기할 수 있는 조기축구 상대팀 선정방법

데이터 수집부터 분석 및 적용까지

스타십벤딩머신(주) | 데이터팀 | 심상훈
20191016



Introduction

발표자, 팀, 프로젝트 배경을 소개합니다.



SOSCON 2019

SAMSUNG OPEN SOURCE CONFERENCE 2019

심상훈



- 소속 : 스타십벤딩머신(주), 데이터팀
- 축구 : 보는거 하는거 다 좋아함
- 경력 : 동네(조기, 생활) 축구 10년
- 포지션 : DMF, LWB, CB
- 역할 : 선수 겸 감독



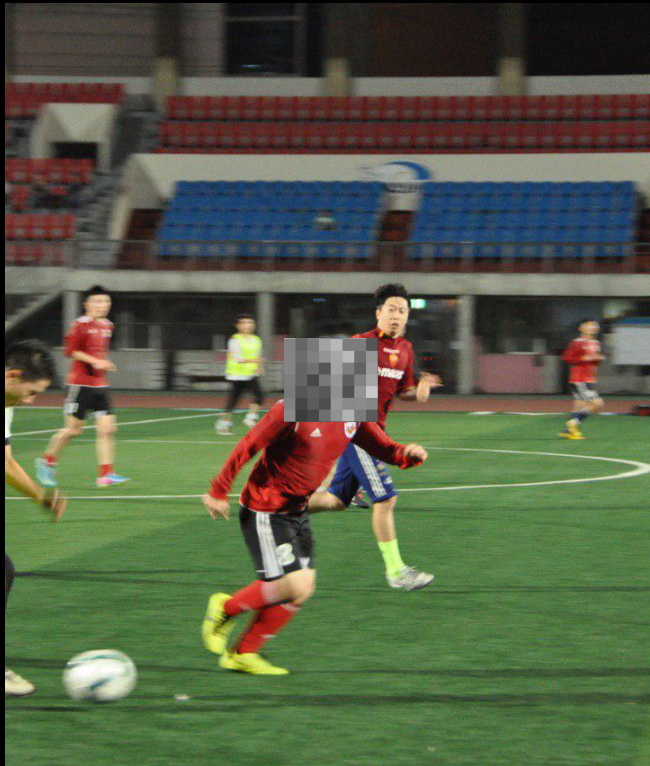
FC P.R.M.J



- 창단 : 2014년
- 인원 : 약 40명(이었는데 많이 줄어듬..)
- 평균 연령 : 31.1세
- 관계 : 친함
- 경기 : 매 주 토요일 아침

FC P.R.M.J





2017년, 주축 멤버 만 30세



다양한 변수

- 다양한 경조사 및 사건으로 매 주 인원변동이 심함
- 임원을 누가 맞느냐에 따라 경기 매칭 및 용병 섭외력에서 차이가 남
- 전반적인 조건이 안 좋을 때 잘하는 상대와 만나서 왕창 깨짐 : 재미도 없고 기분도 나쁨
- 전반적인 조건이 좋을 때 못하는 상대와 만나서 땀도 안남 : 재미도 없고 기분도 소소
- 해당 주에 조건에 맞는 상대를 만나 수준에 맞게 경기하면 땀도 나고 재밌음
 - 이기면(혹은 지더라도) 기분도 좋고, 차주 출석률도 좋고, 카톡방 분위기도 좋음



Moneyball, 2011



UEFA Champions League 2019



UEFA Champions League 2019 : 1472억



★ BEST



2018.09.02 23:41

댓글로 가기 23

배박겐두 배당률, 최근 경기력, 전적, 몸값, 현재 경기와 유사한 과거 경기의 스코어 등등 온갖 변수 다 때려박았습니다



2018.09.03 01:22

👍 🗨

원래는데이터 3천개로 xgboost랑 랜덤포레스트로 돌렸었어요 근데 성능이 고만고만하더라구요. 오버샘플링까지해서 레이블별 데이터 비율도 맞췄는데 f1스코어도 낮고 정확도도 낮고... 그래서 데이터 더 모으고 dnn, cnn으로 돌려보니 더 낮더라구요 132개 피처를 12*12 형태로 리세이프해서 cnn으로 돌렸습니다ㅋㅋ 근데 말씀하신 다른 알고리즘도 시도해봐야겠네요 하지만 성능이 크게 늘어날것같진 않아요 아직 제 데이터셋이 축구 경기 예측에 필요한 모든 변수들을 반영하고 있는것 같지가 않아서요



2018.09.03 07:48

👍 🗨

왜 딥러닝으로 한거예요? 피처가 많아요?



2018.09.03 07:54

👍 🗨

재밌는공들이놀이 피처가 132개입니다

원래는 트리계열 알고리즘으로 했었어요 트리계열 알고리즘 돌릴때 영향력 높은 피처 50개로 했었고요 그런데 딥러닝으로 예측한게 정확도가 더 높더라구요 그래서 딥러닝으로 했습니다



배팅을 위한 분석 예시

SOSCON 2019

포럼 디러닝 축구예측 현재까지 70.33% 적중률

2018.09.02 23:29

조회수 42273 추천수 222 댓글 191

<https://www.fmkorea.com/1248427433> [복사]

	date	match name	prob	odds	prediction	betting money	betting rate
0	180901	Leicester - Liverpool	0.69	4.10	무승부	5310.000000	0.590000
1	180901	West Ham - Wolverhampton Wanderers	0.66	2.55	홈패배	3965.806452	0.440645
2	180901	Brighton - Fulham	0.69	3.10	무승부	4881.428571	0.542381
3	180901	Chelsea - Bournemouth	0.60	1.25	홈승리	-9000.000000	-1.000000
4	180901	Crystal Palace - Southampton	0.69	3.10	무승부	4881.428571	0.542381
5	180901	Everton - Huddersfield	0.60	1.46	홈승리	-2426.086957	-0.269565
6	180901	Manchester City - Newcastle United	0.60	1.07	홈승리	-46028.571429	-5.114286
7	180902	Cardiff - Arsenal	0.66	1.54	홈패배	273.333333	0.030370
8	180902	Watford - Tottenham	0.66	1.54	홈패배	273.333333	0.030370
9	180902	Burnley - Manchester United	0.60	4.60	홈승리	4400.000000	0.488889
10	180831	AC Milan - Roma	0.60	2.39	홈승리	2810.071942	0.312230
11	180901	Bologna - Inter	0.66	1.65	홈패배	1232.307692	0.136923
12	180901	Parma - Juventus	0.66	1.21	홈패배	-8631.428571	-0.959048
13	180902	Fiorentina - Udinese	0.69	3.65	무승부	5157.169811	0.573019
14	180902	Lazio - Frosinone	0.60	1.19	홈승리	-13547.368421	-1.505263
15	180902	Chievo - Empoli	0.69	3.10	무승부	4881.428571	0.542381
16	180902	Sampdoria - Napoli	0.69	3.75	무승부	5195.454545	0.577273
17	180902	Sassuolo - Genoa	0.60	2.16	홈승리	2296.551724	0.255172
18	180902	Torino - SPAL 2013	0.60	1.70	홈승리	257.142857	0.028571
19	180902	Atalanta - Cagliari	0.66	4.45	홈패배	5053.043478	0.561449
20	180831	Getafe - Valladolid	0.60	1.75	홈승리	600.000000	0.066667

- 63 홈팀 선수 가치(Home_Value) - 0.22303100406996268
- 64 원정팀 선수 가치(Away_Value) - 0.18255675796300699
- 59 홈팀 10경기 MFE(Home_MFE_10) - 0.07016267202632027
- 10 원정팀 3경기 평균 전방 패스 수(Away_Pass_DIR_For_3) - 0.041273
- 41 원정팀 5경기 평균 전방 패스 수(Away_Pass_DIR_For_5) - 0.039712
- 56 원정팀 5경기 평균 터치 수(Away_Touch_5) - 0.03365044985133634
- 38 원정팀 5경기 평균 패스 수(Away_Pass_5) - 0.03348756893864298
- 47 원정팀 5경기 평균 롱패스 비율(Away_Pass_Long_ratio_5) - 0.0247
- 33 홈팀 5경기 평균 유효슈팅(Home_SoT_5) - 0.02361593450475436
- 40 홈팀 5경기 평균 전방 패스 수(Home_Pass_DIR_For_5) - 0.0184499
- 8 원정팀 3경기 평균 패스 수(Away_Pass_3) - 0.016352579590121374
- 42 원정팀 5경기 평균 미드필드로 패스 수(Away_Pass_TZ_MT_5) - 0.0
- 6 원정팀 3경기 평균 패스 성공률(Away_PS_3) - 0.01537711224335128
- 3 홈팀 3경기 평균 유효슈팅(Home_SoT_3) - 0.013319150746906188
- 60 원정팀 10경기 MFE(Away_MFE_10) - 0.012961789748764054
- 29 원정팀 3경기 평균 터치 수(Away_Touch_3) - 0.01236284386788393
- 57 홈팀 5경기 MFE(Home_MFE_5) - 0.012266115851441209
- 3 홈팀 5경기 평균 패스 수(Home_Pass_5) - 0.01202956753191691
- 38 원정팀 5경기 평균 패스 성공률(Away_PS_5) - 0.0107954518322018
- 13 홈팀 3경기 평균 전방으로 패스 수(Home_Pass_TZ_FT_3) - 0.00964
- 43 홈팀 5경기 평균 전방으로 패스 수(Home_Pass_TZ_FT_5) - 0.00892
- 49 홈팀 5경기 평균 슈팅 수(Home_Shots_5) - 0.008723761526874452
- 36 홈팀 5경기 평균 패스 성공률(Home_PS_5) - 0.008592242205170006
- 9 홈팀 3경기 평균 전방 패스 수(Home_Pass_DIR_For_3) - 0.00858717
- 46 홈팀 5경기 평균 롱패스 비율(Home_Pass_Long_ratio_5) - 0.00764
- 55 홈팀 5경기 평균 터치 수(Home_Touch_5) - 0.007106523893293825
- 62 6경기 상대 전적의 패배 수(Home_LOSE6) - 0.006932387240934464
- 28 홈팀 3경기 평균 터치 수(Home_Touch_3) - 0.0064946709906501575
- 61 6경기 상대 전적의 승리 수(Home_WINS) - 0.006187947260104416
- 53 홈팀 5경기 평균 공격(Home_RATE_5) - 0.006145300056456987



Goaline Technology



부심은 커녕 선수도 모자랄 때가 많음ㅋㅋㅋㅋㅋ...ㅠ,,ㅠ



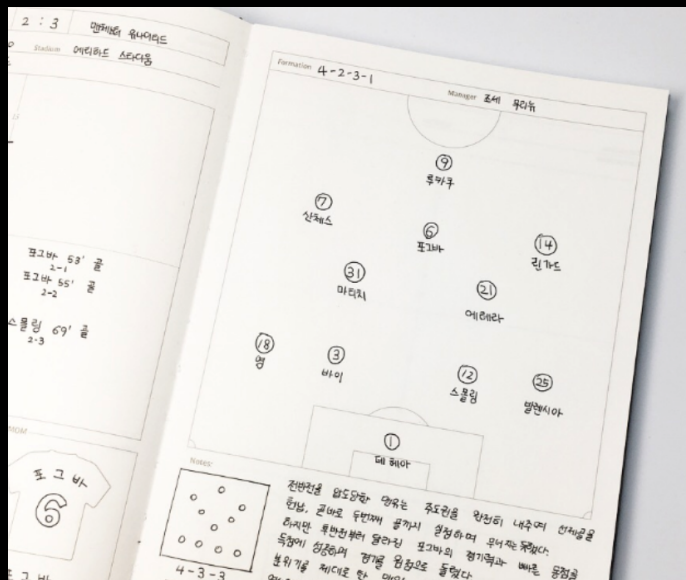
A: 모을 수 있는 것들만 하나씩 직접 기록할 수 밖에 ^^

= 일단 모을 수 있는 걸 하나씩 모아 보자!!!!

고급 옵션

직접 기록

외부 장치에서 읽어오기
파일 기록
파일 제어
HMI 변수값 저장, 불러오기
시스템 설정 진입
사용자 액션 그룹 실행
사용자 액션 그룹 반복 중지
음성/부저 출력
스크립트
동적 키패드 실행
문자열 변수 제어



Data Collecting 1

데이터를 하나씩 꾸준히 모았습니다.



SOSCON 2019

SAMSUNG OPEN SOURCE CONFERENCE 2019

2018년, 심상훈 FC P.R.M.J 감독으로 취임



- 수집 기간

- 2018년 3월 ~ 2019년 1월
- 매년 2월, 8월은 휴식
- 52주 - 9주 - 3주: 40주

- 수집 방법

- 경기당 직접 기록 (선수 정보 별도 기록)

- 수집 데이터 종류

- 기본 데이터
- 우리팀 데이터
- 상대팀 데이터
- 기타 데이터



다음 카페 '서경축구클럽', '대한민국 FC'

2015, 2016, 2017, 2018
4년 연속 올해의 우수카페 선정
축구동호회 커뮤니티 행정 1위

서경축구클럽 총모임

국내 최대 축구클럽 커뮤니티 NO.1

최신글 보기 | 인기글 보기 | 공지사항 | 우리팀 소개 | 운영 신청/구립

2018 올해의 우수카페

통식 (공개)

카페지기 **홍복한하루^^**

회원수 70,591

방문수 14,683

카페업수 269

내 정보

카페 글쓰기

최신글 보기

[서경축구클럽 커뮤니티]

공지사항

가입인사(동업신청)

출석 체크

자유 게시판

이벤트

이벤트 후기

공동구매

축구용품백화점

축구교실/축구레슨

[영어 게시판]

글쓰기 | 답글

최신목록 | 목록 | 댓글 | 아랫글

10월19일(목) 안대대학교 인조잔디구장 상대 초청합니다~! _서울북부 주말경기

쉬우미안라면 조회 16 추천 0 2019.10.14. 12:44

http://cafe.daum.net/skfootball/1xv1/76297

1. 날짜/요일/시간 : 10월 19일 / 토요일 / 오전 10시 ~ 12시
2. 장 소 (구장정보) : 서울 동원구 일계동 안대대학교 인조잔디구장
3. 경기방식(인원) : 11 vs 11
4. 참가비(구장료) : 구장비는 받지 않습니다. 이운으로 5명으로 구장이 대체되었습니다~ 이운으로 5명입니다~! (물 x)
5. 우리팀 소개
- 팀 명 :
- 연 령 대 : 20대 중후반~30대 초반
- 실력(상,중,하) : 하 (선술 X) / 매너 : 상
- 유니폼(상,하) : 상의는 팔 부분: 흰색 / 몸통부분: 파랑색입니다.
- 대표자 연락처 :
6. 기타 전하는 말씀 : 구장이 정식규격에 비해 조금 작은 편인데 11대 11 가능한 경기장입니다~ 뒷 타일(12~2시)에 경기가 없는 경우가 대부분 이기때문에 상하 협의후 1~2편터 더 할 수 있습니다~

*** 선술 다수 있는 팀 정중히 사양합니다. ***
*** 파도방지를 받고 있습니다. (경기 후 돌려드립니다) 꼭 참고 하시고 연락 주세요 ***
*** 문자주실 때 간단한 팀소개(팀명, 나이대, 팀수준(선술유/무 등) 기재해주세요 ***
*** 매칭을 정독하고 연락 주시는 게 매너의 시작이라고 생각합니다 ***

NO.1 전국 축구 동호회 종합 커뮤니티

대한민국 FC (전국 축구 클럽 총모임)

글쓰기 | 답글

최신목록 | 목록 | 댓글 | 아랫글

[*친선팀 초청*] [11vs11] 10월 12일 잠실아울림 초청합니다~! _복서를 친선경기

슬럼프익 조회 16 추천 0 2019.10.07. 09:35

http://cafe.daum.net/soxxer/4q8m/108269

1. 날짜/요일/시간/장소(구장정보)
○ 2019년 10월 12일 16:00~18:00 잠실 어울림 구장

2. 경기방식(인원) : 11 vs 11

3. 경기장 및 참가비(구장료)
○ 잠실아울림구장 - 6.5만원

4. 우리팀 소개
- 팀 명 :
- 연 령 대 : 20대 중후반 ~ 30대 초중반
- 실력 (상,중,하) : 하 (선술 없음)
- 매너 (상,중,하) : 자신있게 최상이라고 말씀드립니다!
- 유니폼 (상/하) : 감정색/감정색
- 연락처 : < 문자 연락 부탁드립니다.

5. 전하는 말씀 :
안녕하십니까! 2012년 9월 친목위주로 창단 후 벌써 7년째 주말마다 즐겁게 운동하는 입니다!

동호인 커뮤니티
- 주말 즐거운 경기를 위하여 실력이 비슷한 팀이었으면 좋겠습니다^^ (선술 없는팀)
(부득이하게 용병을 모시고 오시더라도, 선술이나 현역을 데리고 오시지는 않았으면 합니다.)
- 참석 최소인원 13명 이상으로, 주/부심 이상일이 돌아가며 볼 수 있는 팀
- 시간 약속 잘지키고, 심판 판정에 잘 따르고, 파울이나 넘어졌을때 메너있게 일으켜주시는 팀

저희와 함께 주말 즐겁게 운동하시고, 이후 주기적으로 교류도 했으면 합니다!
감사합니다.

기본데이터

라운드, 날짜, 계절, 날씨, 경기시간, 경기쿼터수, 경기지역, 경기장소, 구장크기, 잔디상태

예) 1, 2018.3.3, 봄, 흐림, 08:00, 4, 서울지 광진구, 아차산배수지체육공원, 대, 중



SOSCON 2019

SAMSUNG OPEN SOURCE CONFERENCE 2019

우리팀 데이터

총참석인원, 우리팀참석인원, 용병참석인원, 총점수, 우리팀점수, 용병점수,
수비수수, 미드필더수, 공격수수, 평균나이, 평균키, 평균몸무게,
선수출신수, 에이스참석여부, 감독참석여부, 포메이션

예) 16명, 15명, 1명, 6.6점, 6.4점, 9.5점
6명, 3명, 7명, 30.7세, 174.8cm, 73.9kg,
2명, 참석, 불참석, 4-3-3

SOSCON 2019

SAMSUNG OPEN SOURCE CONFERENCE 2019



상대팀 데이터

상대팀명, 전적, 실력, 매너, 선수출신수, 나이대, 유니폼색

예) FC누구, 승, 중하, 상, 1명, 20~40대, 검정



SOSCON 2019

SAMSUNG OPEN SOURCE CONFERENCE 2019

기타 데이터

득점, 실점, 경기 결과

예) 8점, 3점, 승



SOSCON 2019

SAMSUNG OPEN SOURCE CONFERENCE 2019

Data Collecting 2

어라? 하는 것들을 발견하고 추가로 데이터를 수집하였습니다.



키퍼

예) 1명



추가 데이터 수집 _우리팀 : 전담프리킥커, 전담코너킥커

SOSCON 2019

전담프리킥커지정, 전담코너킥커지정

예) O, X



추가 데이터 수집 _우리팀 : 전날음주자, 에이스음주여부

SOSCON 2019

경기 전날 음주자수, 경기 전날 에이스 음주 여부

예) 5명, O



상대팀 홈 / 어웨이

예) 홈

이런 외적인 요인 외에도 근육과 생식기의 발육을 촉진하는 남성호르몬인 테스토스테론 수치가 증가해 선수들의 **자신감과 공격성을 높여준다고 합니다.** 영국 노섬브리아대의 샌디 울프슨 박사 등은 자국 프리미어리그 소속 19세 이하 선수들을 대상으로 연구한 결과 테스토스테론 수치가 홈경기에서 40% 증가했고, 라이벌 팀과의 홈경기에서는 67%나 증가한 것으로 조사됐다고 발표했습니다.

연구팀 관계자는 **"자기 영역을 수호하려는 다른 동물들처럼 축구선수들도 자신의 영역인 홈그라운드를 지키기 위해 더욱 정력적이고 능동적으로 변한다"**면서 "외부 팀으로부터 위협을 당할 때 더욱 활동적이고 자기 확신에 가득 차게 된다"고 설명했습니다.



부심 여부

예) O



추가 데이터 수집 _기타 : 키퍼장갑, 조끼, 휘슬

SOSCON 2019

키퍼 장갑 구비 여부, 조끼 구비 여부, 휘슬 구비 여부

예) O, O, X



데이터 수집 완료

SOSCON 2019

	이름	포지션	나이	키	몸무게	점수
1		DF	32	174	72	
2		MF	33	177	80	
3		FW	27	177	73	
4		FW	33	180	73	
5		MF	25	171	64	
6		DF	29	179	83	
7		FW	33	183	76	
8		MF	27	173	68	
9		FW	33	172	102	
10		DF	33	172	74	
11		MF	33	180	76	
12		DF	33	174	72	
13		FW	21	181	72	
14		DF	30	173	71	
15		DF	33	172	74	
16		DF	33	165	74	
17		MF	33	176	73	
18		GK	33	178	94	
19		MF	33	173	74	
20		FW	32	174	72	
21		DF	33	169	89	
22		MF	25	171	68	
23		FW	33	182	78	
24		DF	31	178	73	
25		FW	33	182	74	
26		DF	33	173	84	
27		FW	32	175	80	
28		DF	33	175	70	
29		FW	32	174	73	
30		GK	28	180	71	
31		GK	33	176	79	
32		DF	27	185	70	
33		MF	31	170	68	
34		MF	33	172	73	
35		FW	29	173	68	
36		MF	31	171	73	
37		DF	32	178	72	
38		DF	33	176	78	
39		DF	27	170	67	



데이터 수집 완료

SOSCON 2019

라운드	날짜	계절	날씨	시간	경기형태	지역	중점수	우리점수	용병점수	키비	수비수	미드필드	공격수	평균나이	평균키	평균몸무게	선수출신	에이스	나	음주자	에이스음주	상대전적	상대실력	상대매너	상대선술	상대평균나이	부심	키비상간	조끼	휘슬	독점	심정	결과	
1	2018.3.3	봄	비	11:00	4	서울시 광진구	6.9	6.6	7.5	1	4	4	3	30.5	176.1	74.4	1	0	0	3	0	무	하	하	0	30-40대	X	0	0	X	8	1	승	
2	2018.3.10	봄	흐림	8:00	4	고양시 덕양구	6.6	6.4	9.5	1	6	3	7	30.7	174.8	73.9	1	0	0	4	0	승	하하	상	0	20-30대	0	X	0	X	5	0	승	
3	2018.3.17	봄	맑음	6:00	4	서울시 강서구	6.5	6.5	6.7	0	5	3	6	31.2	175.9	76.2	2	0	X	2	X	무	하하	중	2	20-30대	0	X	X	0	3	6	패	
4	2018.3.24	봄	흐림	8:00	4	서울시 관악구	6.9	6.6	8.0	2	5	6	5	30.6	175.1	73.1	2	0	0	5	0	무	하	상	1	30대	0	X	0	0	14	10	승	
5	2018.3.31	봄	흐림	14:00	4	서울시 중랑구	6.3	6.4	6.1	1	2	2	6	29.6	174.7	73.6	0	X	X	4	X	패	중	상	3	20-30대	X	0	0	0	7	2	패	
6	2018.4.7	봄	눈	7:00	4	서울시 영등포구	6.6	6.4	7.7	2	6	4	3	32.5	174.0	79.4	2	0	0	4	0	승	하	상	0	20-40대	X	0	X	X	8	6	승	
7	2018.4.14	봄	비	7:00	4	서울시 광진구	6.5	6.7	6.0	1	5	3	6	30.6	175.2	74.4	1	0	X	5	X	무	중	중	2	20-40대	X	0	X	X	9	3	패	
8	2018.4.21	봄	흐림	8:00	4	고양시 덕양구	6.9	6.9		1	5	6	4	30.4	173.3	72.1	0	X	0	6	0	승	하하	상	0	20-30대	0	0	0	X	4	1	승	
9	2018.5.12	봄	비	8:00	4	고양시 덕양구	6.3	6.3	7.0	1	6	3	4	30.8	173.7	75.4	0	X	0	7	0	승	하하	상	0	20-30대	0	0	0	X	X	5	0	승
10	2018.5.19	봄	맑음	20:00	6	서울시 중랑구	6.5	6.7	5.3	0	9	3	5	31.7	174.0	75.5	1	0	0	6	X	무	최상	상	6	20-30대	0	X	0	0	12	1	패	
11	2018.5.26	봄	흐림	7:00	4	서울시 영등포구	6.5	6.5	7.0	1	5	6	5	31.9	176.6	77.1	2	0	0	4	0	승	하	중	0	30-40대	X	0	X	X	6	4	승	
12	2018.6.2	여름	흐림	11:00	4	서울시 강남구	6.2	6.7	5.2	0	5	6	4	30.0	175.6	70.5	1	0	0	2	X	무	상	중	2	30대	X	X	0	0	3	14	패	
13	2018.6.9	여름	비	7:00	6	서울시 성북구	6.8	6.8		0	7	5	5	30.6	174.7	72.9	2	0	0	7	0	무	하	상	1	20-50대	0	0	0	X	7	4	승	
14	2018.6.16	여름	맑음	20:00	6	서울시 중랑구	6.8	6.8		1	9	5	5	31.0	175.5	72.5	2	0	0	5	X	패	최상	상	6	20-30대	0	0	0	0	8	4	패	
15	2018.6.23	여름	흐림	10:00	6	고양시 덕양구	6.7	6.8	6.0	0	5	6	4	31.6	176.9	74.9	2	0	0	4	X	무	하	상	0	30-50대	X	0	0	X	2	5	패	
16	2018.6.30	여름	비	8:00	6	서울시 강서구	6.7	6.5	9.5	1	5	6	6	31.3	174.4	74.1	3	0	0	3	0	무	하	상	0	40대	0	X	X	X	7	2	승	
17	2018.7.7	여름	맑음	7:00	4	서울시 영등포구	6.8	6.8		1	6	6	2	30.9	174.9	74.8	2	0	0	1	X	승	하	중	0	30-40대	X	0	0	0	7	4	승	
18	2018.7.14	여름	비	10:00	4	고양시 덕양구	6.2	6.4	5.8	0	5	4	7	30.6	174.7	73.3	2	0	X	5	0	무	중상	상	3	20대	X	X	X	0	1	2	패	
19	2018.7.21	여름	흐림	20:00	4	광명시	6.7	6.8	6.5	0	7	3	2	31.4	173.5	75.3	1	0	0	9	0	무	중하	중	2	20-30대	X	0	X	0	3	3	무	
20	2018.7.28	여름	비	8:00	4	고양시 덕양구	6.8	6.9	4.0	1	5	6	4	31.8	174.7	76.1	2	0	0	4	0	무	하	상	0	30-50대	0	0	0	0	X	6	1	승
21	2018.9.1	가을	맑음	8:00	4	포천시	6.9	6.9		1	0	0	0	32.1	175.1	78.8	0	X	0	3	0	무	하	상	1	40대	0	0	0	X	0	5	3	승
22	2018.9.8	가을	맑음	10:00	6	고양시 덕양구	6.3	6.2	6.8	0	5	2	4	31.7	174.9	78.8	2	0	X	2	X	무	하	중	1	30-40대	0	0	0	0	6	7	패	
23	2018.9.15	가을	비	11:00	4	서울시 강남구	7.0	6.8	8.3	1	5	5	4	30.8	176.6	72.2	3	0	0	1	X	패	상	중	2	30대	0	0	0	0	X	5	5	무
24	2018.9.22	가을	맑음	12:00	4	서울시 구로구	6.1	6.2	4.0	1	5	7	4	31.4	173.9	75.0	0	X	0	3	0	무	하	중	0	30대	0	0	0	X	X	6	4	승
25	2018.9.29	가을	맑음	20:00	4	서울시 송파구	6.7	6.7		0	8	4	4	31.3	177.2	74.9	1	0	X	5	X	무	중하	중	1	20대	X	X	0	0	3	8	패	
26	2018.10.6	가을	비	10:00	6	고양시 덕양구	6.3	6.8	5.0	1	2	4	4	30.5	176.9	74.3	2	0	X	5	X	무	중	2	30-50대	X	0	X	X	2	8	패		
27	2018.10.13	가을	흐림	19:00	4	서울시 금천구	6.4	6.4		0	6	3	5	31.7	176.9	75.8	1	0	0	4	X	무	하	상	1	20-30대	X	0	0	0	2	7	패	
28	2018.10.20	가을	흐림	11:00	4	남양주시	7.0	6.9	7.2	1	4	5	2	31.0	174.9	71.4	1	0	0	2	0	무	하	상	0	20-40대	0	0	0	X	X	6	1	승
29	2018.10.27	가을	맑음	17:00	4	서울시 중랑구	5.9	6.6	5.0	0	5	2	5	30.3	174.3	75.3	0	X	0	1	X	무	중	2	20-30대	X	X	X	X	2	9	패		
30	2018.11.3	가을	흐림	8:00	4	서울시 도봉구	7.1	6.8	7.8	1	4	7	2	31.2	176.3	72.3	2	0	X	3	0	승	중	상	2	20-50대	0	0	0	X	X	6	5	승
31	2018.11.10	가을	비	20:00	4	서울시 송파구	6.8	6.8		1	3	3	4	30.4	175.8	73.5	1	0	0	6	X	무	중상	상	2	30-50대	X	0	X	0	1	5	패	
32	2018.11.17	가을	흐림	10:00	6	고양시 일산동구	6.3	6.5	5.9	0	4	3	7	31.4	176.1	75.8	1	0	0	9	X	패	중	상	2	30-50대	X	0	0	0	X	3	10	패
33	2018.11.24	가을	눈	10:00	4	서울시 은평구	6.9	6.7	7.7	1	5	5	2	31.2	174.1	75.1	1	0	0	4	0	무	하하	상	0	30대	0	0	0	0	6	2	승	
34	2018.12.8	겨울	맑음	8:00	4	서울시 서대문구	6.9	7.0	6.5	1	4	5	3	31.8	175.4	73.1	2	0	0	2	X	무	하	중	1	20-30대	0	0	0	X	0	4	1	승
35	2018.12.22	겨울	흐림	8:00	4	서울시 영등포구	6.8	7.0	5.0	0	4	3	5	30.6	175.6	71.6	1	0	X	3	X	무	하하	중	0	20-50대	X	X	0	X	2	5	패	
36	2018.12.29	겨울	맑음	20:00	4	서울시 도봉구	6.3	6.5	6.0	0	6	6	2	31.1	174.5	73.7	0	X	X	2	X	무	중하	중	1	20대	X	X	0	0	1	6	패	
37	2019.1.5	겨울	흐림	10:00	4	서울시 은평구	6.9	6.9		1	4	6	4	31.2	174.8	74.5	2	0	0	7	0	무	하	상	1	20-30대	0	0	0	X	X	7	3	승
38	2019.1.12	겨울	흐림	8:00	4	고양시 덕양구	7.0	6.9	7.3	1	5	8	4	31.0	175.1	72.1	3	0	0	4	X	무	하	상	1	20-30대	0	0	0	0	4	2	승	
39	2019.1.19	겨울	흐림	16:00	4	서울시 서대문구	6.8	6.8		1	6	4	2	31.8	174.0	75.7	1	0	0	2	X	무	중하	중	2	30-50대	X	0	0	X	0	5	5	무
40	2019.1.26	겨울	맑음	8:00	4	서울시 동작구	6.8	6.6	6.8	1	4	6	3	31.3	174.6	73.4	2	0	0	2	X	무	하하	상	0	30-40대	0	0	0	0	X	6	2	승

Data Analysis

수집한 데이터로 본격적인 분석을 시작합니다.



3.Data Exploration

```
print('구장크기 결측 비율 : ', (data.isnull().sum()/len(data))[7]*100, '%')
print('잔디상태 결측 비율 : ', (data.isnull().sum()/len(data))[8]*100, "%")
print('용병점수 결측 비율 : ', (data.isnull().sum()/len(data))[14]*100, "%")
print('상대매너 결측 비율 : ', (data.isnull().sum()/len(data))[33]*100, "%")
```

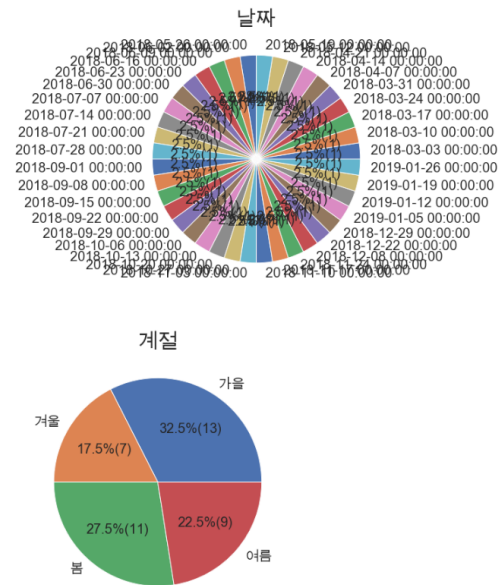
구장크기 결측 비율 : 15.0 %
 잔디상태 결측 비율 : 20.0 %
 용병점수 결측 비율 : 25.0 %
 상대매너 결측 비율 : 10.0 %

data.describe()

	경기쿼터수	총참석	우리참석	용병참석	총점수	우리점수	용병점수	키퍼	수비수	미드필더	공격수	평균나이	평균키	평균몸무게	선수출신	음주자	상대선출	득점	실점
count	40	40	40	40	40	40	30	40	40	40	40	40	40	40	40	40	40	40	40
mean	4	15	12	2	7	7	7	1	5	4	4	31	175	74	1	4	1	5	4
std	1	2	3	2	0	0	1	1	2	2	2	1	1	2	1	2	1	3	3
min	4	11	7	0	6	6	4	0	0	0	0	30	173	70	0	1	0	1	0
25%	4	13	10	1	6	6	6	0	4	3	3	31	174	73	1	2	0	3	2
50%	4	15	12	2	7	7	7	1	5	4	4	31	175	74	1	4	1	5	4
75%	4	16	15	3	7	7	7	1	6	6	5	31	176	75	2	5	2	7	6
max	6	20	20	5	7	7	10	2	9	8	7	33	177	79	3	9	6	14	14



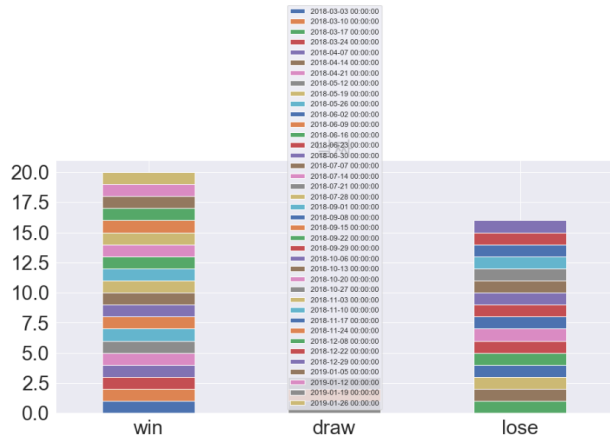
```
for i in range(len(data.columns)-1):
    plt.figure()
    pie(data.columns[i])
plt.show()
```



```
def bar(feature):  
    win = data[data["결과"]=="승"][feature].value_counts()  
    draw = data[data["결과"]=="무"][feature].value_counts()  
    lose = data[data["결과"]=="패"][feature].value_counts()  
    data_2 = pd.DataFrame([win, draw, lose])  
    data_2.index = ["win", "draw", "lose"]  
    data_2.plot(kind='bar', rot=0, figsize=(15,7), stacked=True, fontsize = 30)  
    plt.title(feature, fontsize=30)  
    plt.legend(loc = 'best')
```

```
for i in range(len(data.columns)-1):  
    plt.figure()  
    bar(data.columns[i])  
plt.show()
```

<Figure size 432x288 with 0 Axes>



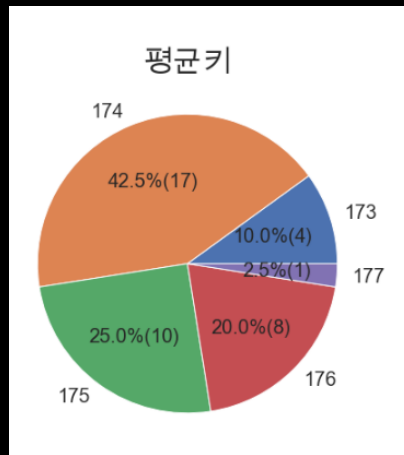
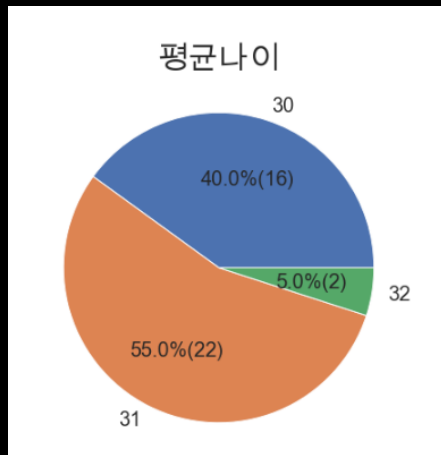
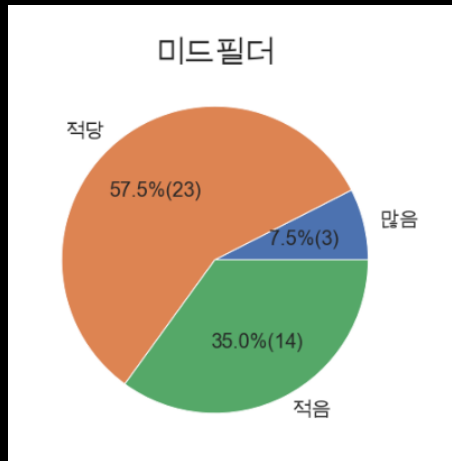
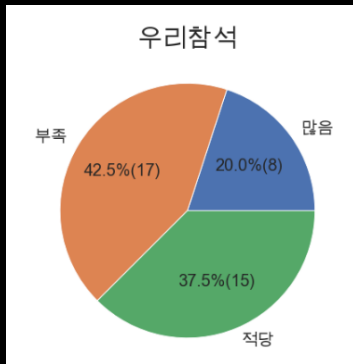
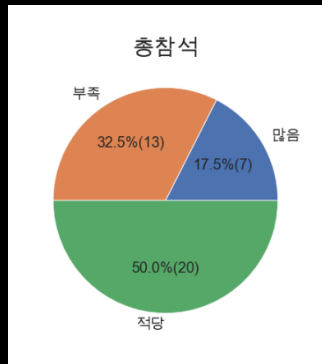
4.Data Preprocessing 1

Data Dictionay

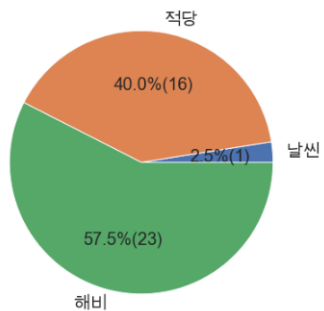
*좀 더 명확한 시각화 파악을 위한 데이터 전처리 : 주로 'Categorization'

- 시간 : 이른아침(~07:00), 아침(08:00~10:00), 점심(11:00~15:00), 저녁(16:00~)
- 지역 : 서울시, 경기도
- 총참석 : 부족(11~13), 적당(14~16), 많음(17~)
- 우리팀 참석 : 부족(7~11), 적당(12~15), 많음(16~)
- 용병 참석 : 완벽(0), 적당(1~2), 많음(3~)
- 총 점수 : 매우불안(~5.9), 불안(6.0~6.5), 해불만함(6.6~)
- 우리팀 점수 : 매우불안(~5.9), 불안(6.0~6.5), 해불만함(6.6~)
- 용병 점수 : 매우불안(~5.9), 불안(6.0~6.5), 해불만함(6.6~)
- 수비수 : 적음(0~3), 적당(4~6), 많음(7~)
- 미드필더 : 적음(0~3), 적당(4~6), 많음(7~)
- 공격수 : 적음(0~3), 적당(4~6), 많음(7~)
- 나이 : 30, 31, 32
- 나이 : 173, 174, 175, 176, 177
- 몸무게 : 날씬(~69), 적당(71~4), 해비(75~)
- 음주자 : 적음(0~2), 보통(3~5), 미친(6~)
- 포메이션 : 3백, 4백
- 상대팀 나이 : 25, 35, 45
- 상대팀 유니폼 : 검정, 파랑, 흰, 기타
- 득점 : 적음(0~3), 보통(4~6), 많음(7~)
- 실점 : 적음(0~3), 보통(4~6), 많음(7~)

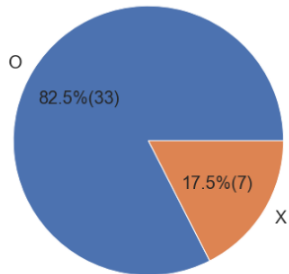




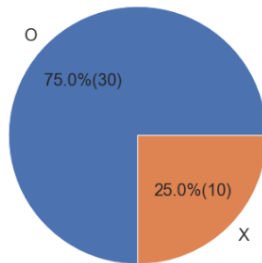
평균몸무게



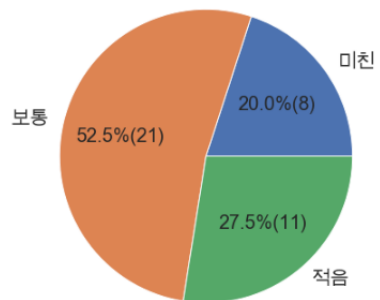
에이스



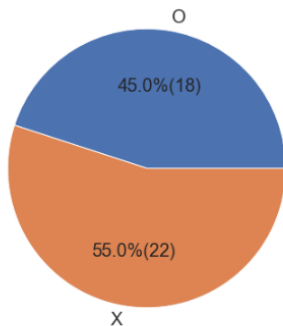
나



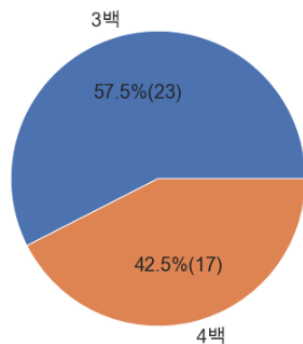
음주자

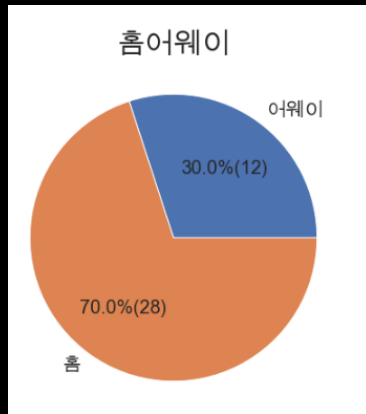
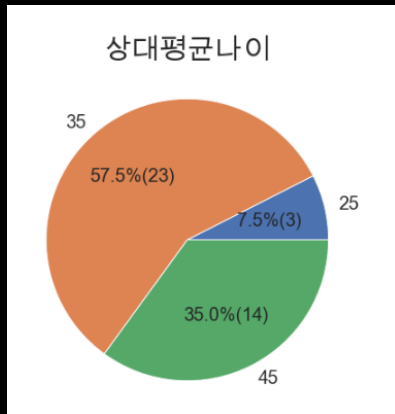
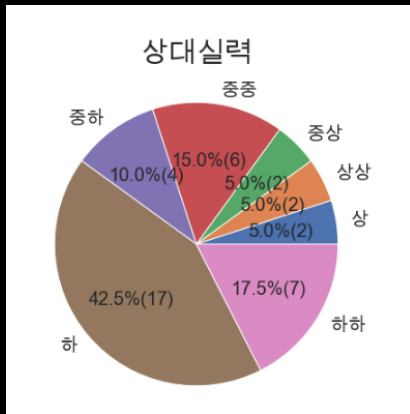


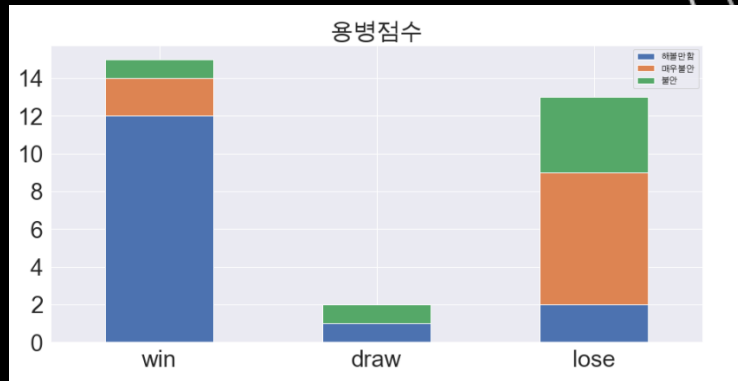
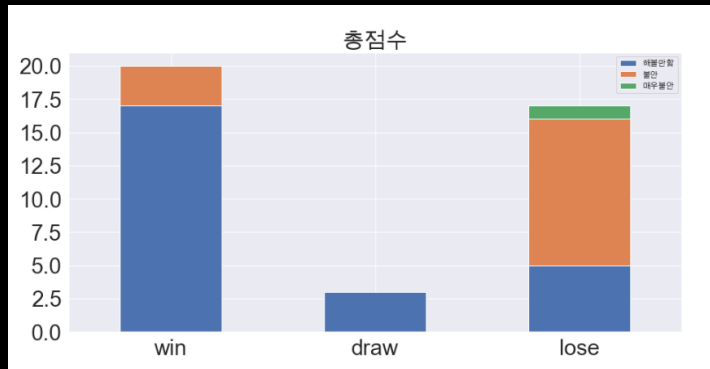
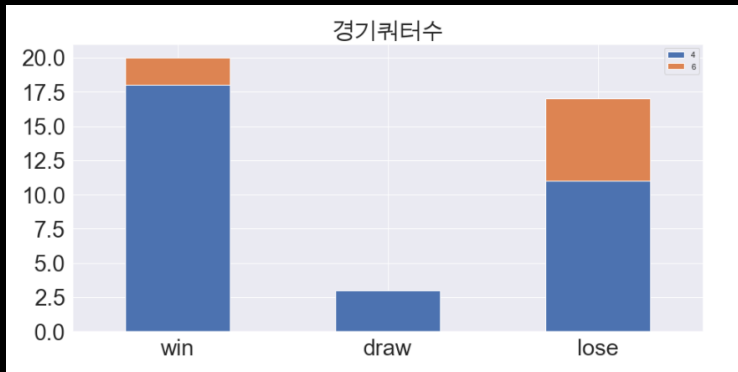
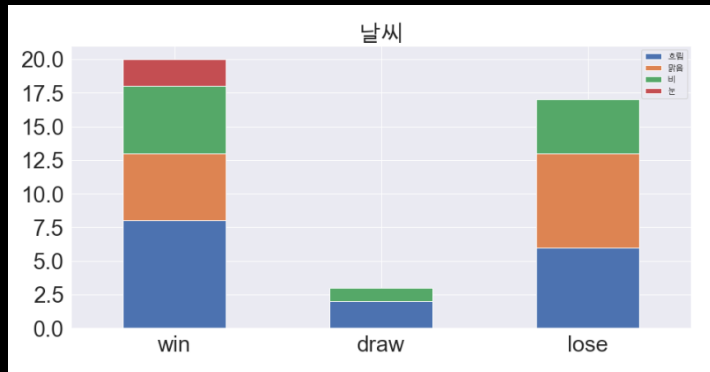
에이스음주

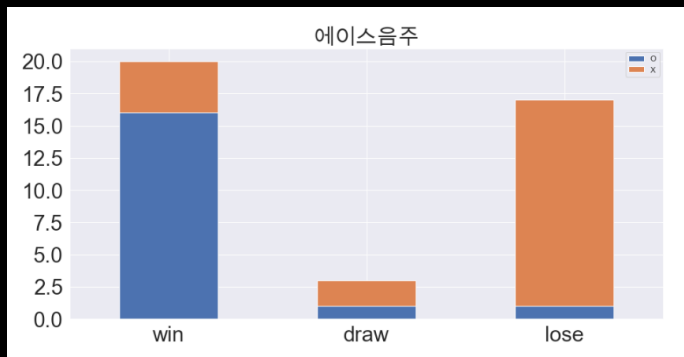
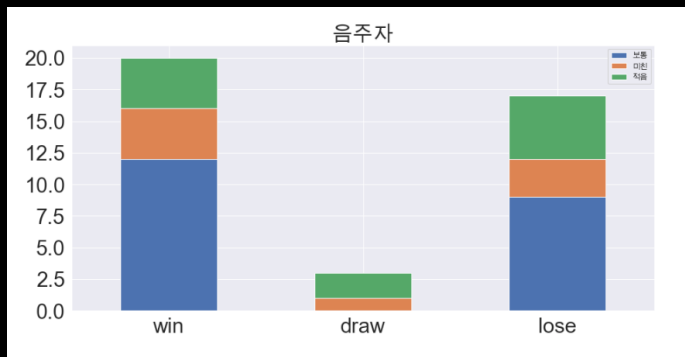
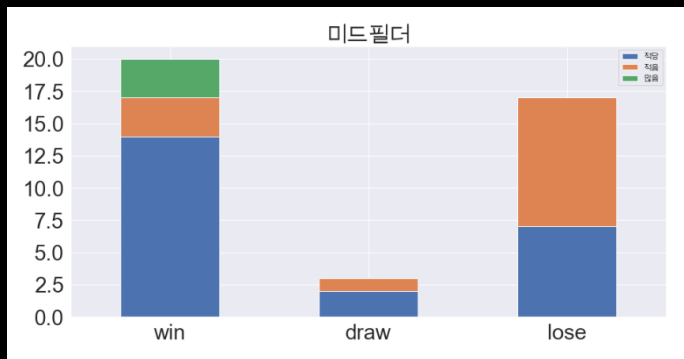
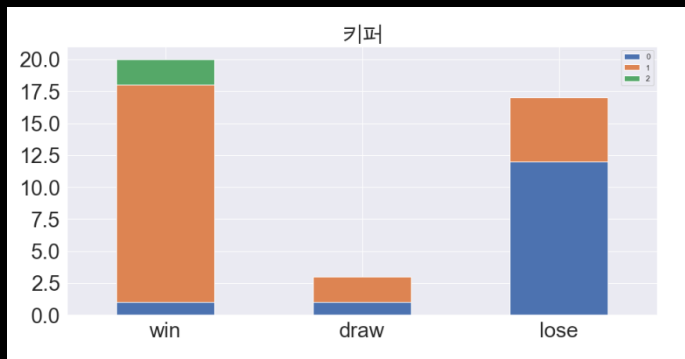


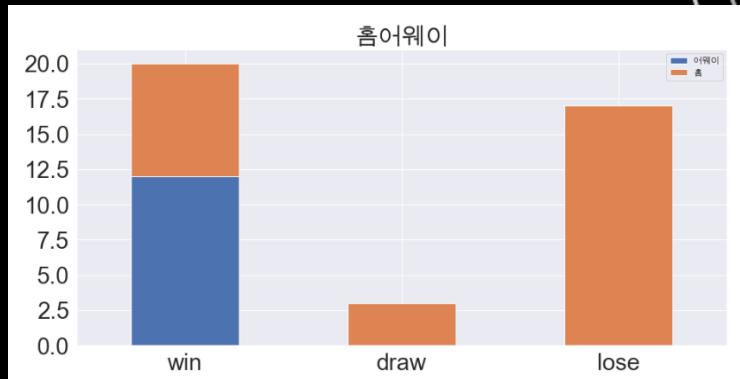
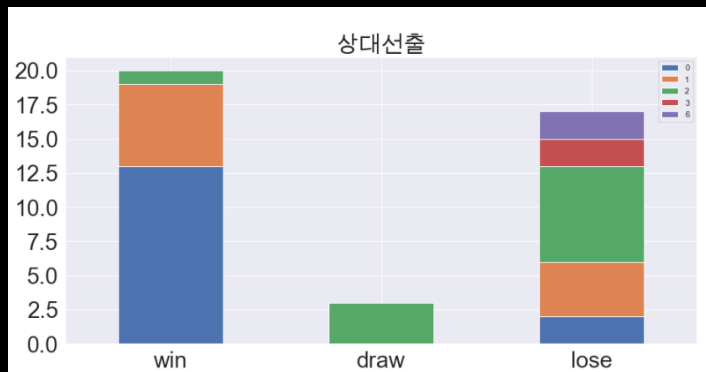
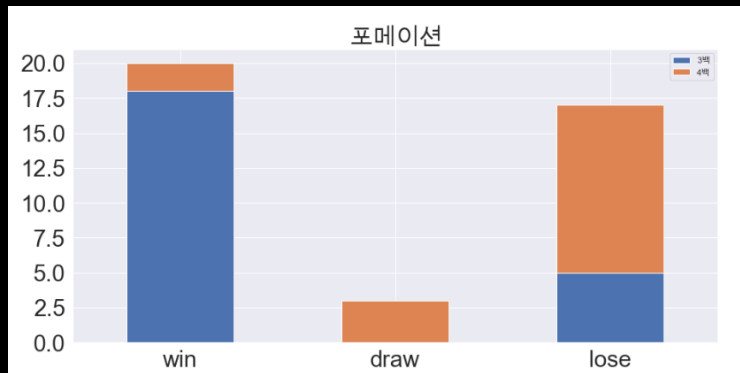
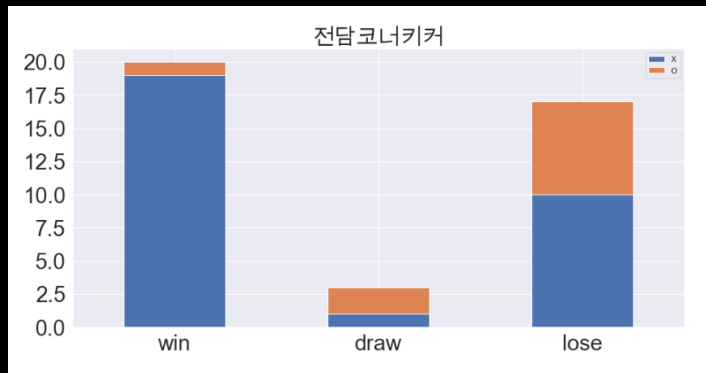
포메이션

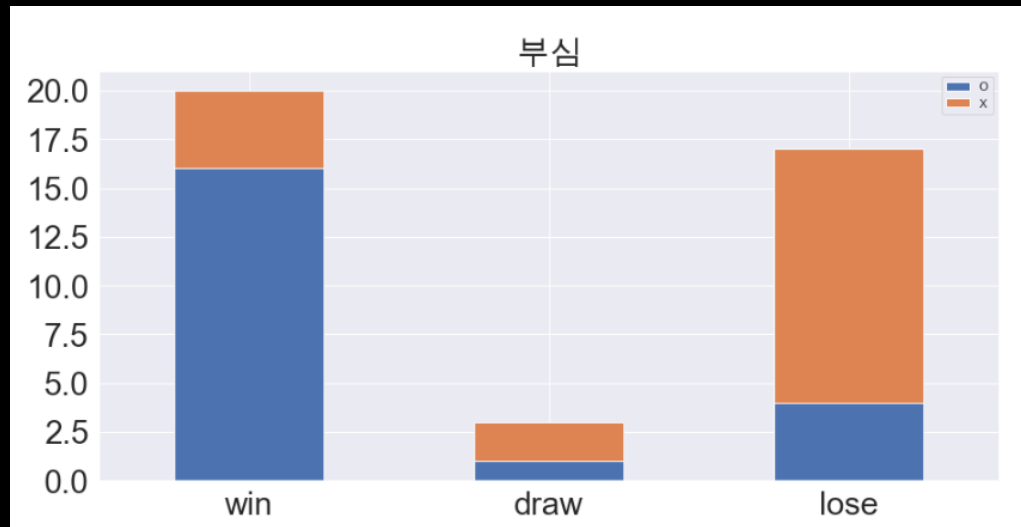












5.Feature Engineering

- feature(피쳐) : 각각의 컬럼
- feature engineering : 기계가 이해할 수 있도록 각각의 feature를 vector로 변환하는 작업 = 숫자의 배열로 만들어 주는 것
- 결측치 처리, 카테고리 데이터 처리 등

```
data.head()
```

	날짜	계절	날씨	시간	경기 쿼터수	지역	장소	구장크기	잔디상태	총참석	우리참석	용병참석	총점수	우리점수	용병점수	키퍼	수비수	미드필더	공격수	평균나이	평균키	평균몸무게	선수출신	에이스	나	음주자	에이스음주	전담프리카커	전담코너커	포메이션	상대팀명	상대실력	상대매너	상대선출	상대평균나이	상대유니폼색	홈어웨이	부심	키퍼장갑	조끼	휘슬	독점	실점	결과	
라운드	1	2018-03-03	봄	비	점심	4	서울시	아차산배수지체육공원	대	중	부족	부족	많음	해불만함	해불만함	해불만함	1	적당	적당	적음	30	176	해비	1	0	0	보통	0	X	X	3백	무	하	하	0	45	검정	어웨이	X	0	0	X	많음	적음	승
	2	2018-03-10	봄	흐림	아침	4	경기도	항공대학교	대	하	적당	적당	적당	해불만함	불만	해불만함	1	적당	적음	많음	30	174	적당	1	0	0	보통	0	X	X	4백	승	하	상	0	35	검정	홈	0	X	0	X	적당	적음	승
	3	2018-03-17	봄	맑음	이른아침	4	서울시	우장산체육구장	대	상	적당	부족	많음	불만	불만	해불만함	0	적당	적음	적당	31	175	해비	2	0	X	적음	X	X	X	4백	무	하	중	2	35	기타	홈	0	X	X	0	적음	적당	패
	4	2018-03-24	봄	흐림	아침	4	서울시	관악구체육공원	대	하	적당	적당	적당	해불만함	해불만함	해불만함	1	적당	적음	적음	30	174	적당	1	0	0	보통	0	X	X	3백	무	하	하	0	45	검정	어웨이	X	0	0	X	많음	적음	승
	5	2018-03-31	봄	맑음	아침	4	서울시	관악구체육공원	대	상	적당	부족	많음	불만	불만	해불만함	0	적당	적음	적당	31	175	해비	2	0	X	적음	X	X	X	4백	무	하	중	2	35	기타	홈	0	X	X	0	적음	적당	패

```
data.drop(["날짜", "장소", "평균나이", "평균키", "상대팀명", "상대유니폼색", "조끼", "휘슬"], 1, inplace = True)
```

```
data.shape
```

```
(40, 37)
```



```
print('구장크기 결측 비율 : ', (data["구장크기"].isnull().sum()/len(data))*100, '%')
print('잔디상태 결측 비율 : ', (data["잔디상태"].isnull().sum()/len(data))*100, "%")
print('용병점수 결측 비율 : ', (data["용병점수"].isnull().sum()/len(data))*100, "%")
print('상대매너 결측 비율 : ', (data["상대매너"].isnull().sum()/len(data))*100, "%")
```

구장크기 결측 비율 : 15.0 %
잔디상태 결측 비율 : 20.0 %
용병점수 결측 비율 : 25.0 %
상대매너 결측 비율 : 10.0 %

```
data["구장크기"].fillna("중", inplace=True)
data["잔디상태"].fillna("중", inplace=True)
data["용병점수"].fillna("불안", inplace=True)
data["상대매너"].fillna("중", inplace=True)
```

```
data.loc[data["결과"]=="패", "결과"] = 0
data.loc[data["결과"]=="무", "결과"] = 0
data.loc[data["결과"]=="승", "결과"] = 1
```

```
data.loc[data["경기쿼터수"]==4, "경기쿼터수"] = 0
data.loc[data["경기쿼터수"]==6, "경기쿼터수"] = 1
```



개별	날씨	시간	경기쿼터수	지역	구장경기	잔디상태	총합석	우리합석	양병합석	총점수	우리점수	양병점수	키퍼	수비수	미드필더	공격수	평균볼무게	선수출신	에이스	나	올주자	에이스올주	전담프리카커	전담코너키커	포메이션	상대전적	상대실력	상대매너	상대선술	상대공군나이	홀어웨이	부심	키퍼장갑	특점	실점	결과	
라운드																																					
1	봄	비	점심	4	서울시	대	중	부족	부족	많음	해불만함	해불만함	해불만함	1	적당	적당	적음	해비	1	0	0	보통	0	X	X	3백	무	하	하	0	45	어웨이	X	0	많음	적음	1
2	봄	흐림	아침	4	경기도	대	하	적당	적당	적당	해불만함	불만	해불만함	1	적당	적음	많음	적당	1	0	0	보통	0	X	X	4백	승	하하	상	0	35	홈	0	X	적당	적음	1
3	봄	맑음	이른아침	4	서울시	대	상	적당	부족	많음	불만	불만	해불만함	0	적당	적음	적당	해비	2	0	X	적음	X	X	X	4백	무	하하	중	2	35	홈	0	X	적음	적당	0
4	봄	흐림	아침	4	서울시	소	상	적당	적당	많음	해불만함	해불만함	해불만함	2	적당	적당	적당	적당	2	0	0	보통	0	0	X	3백	무	하	상	1	35	어웨이	0	X	많음	많음	0
5	봄	흐림	점심	4	서울시	대	하	부족	부족	많음	불만	불만	불만	1	적음	적음	적당	적당	0	X	X	보통	X	X	X	4백	패	중중	상	3	35	홈	X	0	많음	적음	0
6	봄	눈	이른아침	4	서울시	소	상	적당	부족	많음	해불만함	불만	해불만함	2	적당	적당	적음	해비	2	0	0	보통	0	0	0	3백	승	하	중	0	45	홈	X	0	많음	적당	1
7	봄	비	이른아침	4	서울시	대	중	적당	적당	많음	불만	해불만함	불만	1	적당	적음	적당	해비	1	0	X	보통	X	X	X	4백	무	중중	중	2	35	홈	X	0	많음	적음	0
8	봄	흐림	아침	4	경기도	대	하	적당	많음	완벽	해불만함	해불만함	불만	1	적당	적당	적당	적당	0	X	0	미친	0	X	X	3백	승	하하	상	0	35	어웨이	0	0	적당	적음	1
9	봄	비	아침	4	경기도	대	하	부족	적당	적당	불만	불만	해불만함	1	적당	적음	적당	해비	0	X	0	미친	0	X	X	3백	승	하하	상	0	35	어웨이	0	0	적당	적음	1
10	봄	맑음	저녁	6	서울시	대	중	많음	적당	적당	불만	해불만함	매우불만	0	많음	적음	적당	해비	1	0	0	미친	X	X	X	4백	무	상상	상	6	35	홈	0	X	많음	적음	0
11	봄	흐림	이른아침	4	서울시	소	상	많음	많음	적당	불만	불만	해불만함	1	적당	적당	적당	해비	2	0	0	보통	0	X	X	4백	승	하	중	0	45	홈	X	0	적당	적당	1
12	여름	흐림	점심	4	서울시	대	상	적당	부족	많음	불만	해불만함	매우불만	0	적당	적당	적당	날변	1	0	0	적음	X	0	X	4백	무	상	중	2	35	홈	X	X	적음	많음	0
13	여름	봄	이른아침	6	서울시	대	상	많음	많음	완벽	해불만함	해불만함	불만	0	많음	적당	적당	적당	2	0	0	미친	0	X	X	3백	무	하	상	1	35	어웨이	0	0	많음	적당	1
14	여름	맑음	저녁	6	서울시	대	중	많음	많음	완벽	해불만함	해불만함	불만	1	많음	적당	적당	적당	2	0	0	보통	X	X	X	3백	패	상상	상	6	35	홈	0	0	많음	적당	0
15	여름	흐림	아침	6	경기도	중	중	적당	적당	적당	해불만함	해불만함	불만	0	적당	적당	적당	해비	2	0	0	보통	X	0	0	4백	무	하	상	0	45	홈	X	0	적음	적당	0
16	여름	비	아침	6	서울시	대	상	많음	많음	적당	해불만함	불만	해불만함	1	적당	적당	적당	해비	3	0	0	보통	0	X	X	3백	무	하	상	0	45	어웨이	0	X	많음	적음	1
17	여름	맑음	이른아침	4	서울시	소	상	적당	적당	완벽	해불만함	해불만함	불만	1	적당	적당	적음	해비	2	0	0	적음	X	X	X	3백	승	하	중	0	45	홈	X	0	많음	적당	1
18	여름	봄	아침	4	경기도	대	상	적당	적당	많음	불만	불만	매우불만	0	적당	적당	많음	적당	2	0	X	보통	0	X	X	4백	무	중상	상	3	25	홈	X	X	적음	적당	0
19	여름	흐림	저녁	4	경기도	대	하	부족	부족	적당	해불만함	해불만함	불만	0	많음	적음	적음	해비	1	0	0	미친	0	0	0	4백	무	중하	중	2	35	홈	X	0	적음	적음	0
20	여름	비	아침	4	경기도	대	하	적당	적당	적당	해불만함	해불만함	매우불만	1	적당	적당	적당	해비	2	0	0	보통	0	X	X	3백	무	하	상	0	45	어웨이	0	0	적당	적음	1
21	가을	맑음	아침	4	경기도	대	상	적당	적당	완벽	해불만함	해불만함	불만	1	적음	적음	적음	해비	0	X	0	보통	0	X	X	3백	무	하	상	1	45	홈	0	0	적당	적음	1
22	가을	맑음	아침	6	경기도	대	상	부족	부족	적당	불만	불만	해불만함	0	적당	적음	적당	해비	2	0	X	적음	X	X	X	3백	무	하	중	1	45	홈	0	0	적당	많음	0
23	가을	비	점심	4	서울시	대	상	적당	적당	적당	해불만함	해불만함	해불만함	1	적당	적당	적당	적당	3	0	0	적음	X	0	0	4백	패	상	중	2	35	홈	0	0	적당	적당	0
24	가을	맑음	점심	4	서울시	소	상	많음	많음	적당	불만	불만	매우불만	1	적당	많음	적당	해비	0	X	0	보통	0	X	X	3백	무	하	중	0	35	어웨이	0	0	적당	적당	1
25	가을	맑음	저녁	4	서울시	대	하	적당	많음	완벽	해불만함	해불만함	불만	0	많음	적당	적당	해비	1	0	X	보통	X	X	0	4백	무	중하	중	1	25	홈	X	X	적음	많음	0
26	가을	비	아침	6	경기도	대	상	부족	부족	많음	불만	해불만함	매우불만	1	적음	적당	적당	해비	2	0	X	보통	X	0	0	4백	무	중중	중	2	45	홈	X	0	적음	적당	0
27	가을	흐림	저녁	4	서울시	대	상	적당	많음	완벽	불만	불만	불만	0	적당	적음	적당	해비	1	0	0	보통	X	0	0	3백	무	하	중	1	35	홈	X	0	적음	많음	0
28	가을	흐림	점심	4	경기도	대	중	부족	부족	많음	해불만함	해불만함	해불만함	1	적당	적당	적음	적당	1	0	0	적음	0	X	X	3백	무	하	상	0	35	홈	0	0	적당	적음	1
29	가을	맑음	저녁	4	서울시	대	중	부족	부족	많음	매우불만	해불만함	매우불만	0	적당	적음	적당	해비	0	X	0	적음	X	X	0	4백	무	중중	중	2	35	홈	X	X	적음	많음	0
30	가을	흐림	아침	4	서울시	대	하	적당	부족	많음	해불만함	해불만함	해불만함	1	적당	많음	적음	적당	2	0	X	보통	X	X	X	3백	승	중중	상	2	45	어웨이	0	0	적당	적당	1
31	가을	비	저녁	4	서울시	중	하	부족	부족	완벽	해불만함	해불만함	불만	1	적음	적음	적당	적당	1	0	0	미친	X	0	0	4백	무	중상	하	2	35	홈	X	0	적음	적당	0
32	가을	흐림	아침	6	경기도	중	중	적당	부족	많음	불만	불만	매우불만	0	적당	적음	많음	해비	1	0	0	미친	X	X	X	4백	패	중중	중	2	45	홈	X	0	적음	많음	0
33	가을	눈	아침	4	서울시	대	하	부족	부족	많음	해불만함	해불만함	해불만함	1	적당	적당	적음	해비	1	0	0	보통	0	X	X	3백	무	하하	상	0	35	어웨이	0	0	적당	적음	1
34	가을	맑음	아침	4	서울시	중	중	부족	부족	적당	해불만함	해불만함	불만	1	적당	적당	적음	적당	2	0	0	적음	X	X	X	3백	무	하	중	1	35	홈	0	0	적당	적음	1
35	가을	흐림	아침	4	서울시	소	중	부족	부족	적당	해불만함	해불만함	매우불만	0	적당	적음	적당	적당	1	0	X	보통	X	X	0	3백	무	하하	하	0	35	홈	X	X	적음	적당	0
36	가을	맑음	저녁	4	서울시	대	하	적당	부족	많음	불만	불만	불만	0	적당	적당	적음	적당	0	X	X	적음	X	X	X	3백	무	중하	중	1	25	홈	X	X	적음	적당	0
37	가을	흐림	아침	4	서울시	대	하	적당	적당	완벽	해불만함	해불만함	불만	1	적당	적당	적당	해비	2	0	0	미친	0	X	X	3백	무	하	상	1	35	어웨이	0	0	많음	적음	1
38	가을	흐림	아침	4	경기도	중	중	많음	적당	적당	해불만함	해불만함	해불만함	1	적당	많음	적당	적당	3	0	0	보통	X	0	X	3백	무	하	상	1	35	어웨이	0	0	적당	적음	1
39	가을	흐림	저녁	4	서울시	중	중	부족	적당	완벽	해불만함	해불만함	불만	1	적당	적당	적당	해비	1	0	0	적음	X	X	X	4백	무	중하	중	2	45	홈	X	0	적당	적당	0
40	가을	맑음	아침	4	경기도	중	중	적당	적당	적당	해불만함	해불만함	해불만함	1	적당	적당	적음	적당	2	0	0	적음	X	X	X	3백	무	하하	상	0	45	홈	0	0	적당	적당	0

#One-hot encoding

```
Image("https://hooon.s3.ap-northeast-2.amazonaws.com/SOSCON_2019/image/94.jpg")
```

color	color_red	color_blue	color_green
red	1	0	0
green	0	0	1
blue	0	1	0
red	1	0	0

```
from sklearn import preprocessing
minmax_scale = preprocessing.MinMaxScaler().fit(data_FE[['상대평균나이']])
scaled_data_FE_enemyAge = minmax_scale.transform(data_FE[['상대평균나이']])
data_FE["상대평균나이"] = scaled_data_FE_enemyAge
data_FE
```



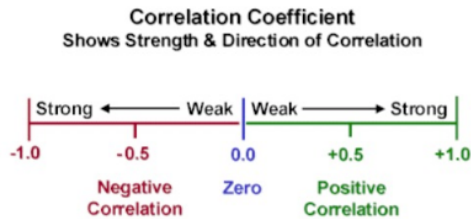
```
data_FE = pd.get_dummies(data)
```

行	列	列1	列2	列3	列4	列5	列6	列7	列8	列9	列10	列11	列12	列13	列14	列15	列16	列17	列18	列19	列20	列21	列22	列23	列24	列25	列26	列27	列28	列29	列30	列31	列32	列33	列34	列35	列36	列37	列38	列39	列40	列41	列42	列43	列44	列45	列46	列47	列48	列49	列50	列51	列52	列53	列54	列55	列56	列57	列58	列59	列60	列61	列62	列63	列64	列65	列66	列67	列68	列69	列70	列71	列72	列73	列74	列75	列76	列77	列78	列79	列80	列81	列82	列83	列84	列85	列86	列87	列88	列89	列90	列91	列92	列93	列94	列95	列96	列97	列98	列99	列100	列101	列102	列103	列104	列105	列106	列107	列108	列109	列110	列111	列112	列113	列114	列115	列116	列117	列118	列119	列120	列121	列122	列123	列124	列125	列126	列127	列128	列129	列130	列131	列132	列133	列134	列135	列136	列137	列138	列139	列140	列141	列142	列143	列144	列145	列146	列147	列148	列149	列150	列151	列152	列153	列154	列155	列156	列157	列158	列159	列160	列161	列162	列163	列164	列165	列166	列167	列168	列169	列170	列171	列172	列173	列174	列175	列176	列177	列178	列179	列180	列181	列182	列183	列184	列185	列186	列187	列188	列189	列190	列191	列192	列193	列194	列195	列196	列197	列198	列199	列200	列201	列202	列203	列204	列205	列206	列207	列208	列209	列210	列211	列212	列213	列214	列215	列216	列217	列218	列219	列220	列221	列222	列223	列224	列225	列226	列227	列228	列229	列230	列231	列232	列233	列234	列235	列236	列237	列238	列239	列240	列241	列242	列243	列244	列245	列246	列247	列248	列249	列250	列251	列252	列253	列254	列255	列256	列257	列258	列259	列260	列261	列262	列263	列264	列265	列266	列267	列268	列269	列270	列271	列272	列273	列274	列275	列276	列277	列278	列279	列280	列281	列282	列283	列284	列285	列286	列287	列288	列289	列290	列291	列292	列293	列294	列295	列296	列297	列298	列299	列300	列301	列302	列303	列304	列305	列306	列307	列308	列309	列310	列311	列312	列313	列314	列315	列316	列317	列318	列319	列320	列321	列322	列323	列324	列325	列326	列327	列328	列329	列330	列331	列332	列333	列334	列335	列336	列337	列338	列339	列340	列341	列342	列343	列344	列345	列346	列347	列348	列349	列350	列351	列352	列353	列354	列355	列356	列357	列358	列359	列360	列361	列362	列363	列364	列365	列366	列367	列368	列369	列370	列371	列372	列373	列374	列375	列376	列377	列378	列379	列380	列381	列382	列383	列384	列385	列386	列387	列388	列389	列390	列391	列392	列393	列394	列395	列396	列397	列398	列399	列400	列401	列402	列403	列404	列405	列406	列407	列408	列409	列410	列411	列412	列413	列414	列415	列416	列417	列418	列419	列420	列421	列422	列423	列424	列425	列426	列427	列428	列429	列430	列431	列432	列433	列434	列435	列436	列437	列438	列439	列440	列441	列442	列443	列444	列445	列446	列447	列448	列449	列450	列451	列452	列453	列454	列455	列456	列457	列458	列459	列460	列461	列462	列463	列464	列465	列466	列467	列468	列469	列470	列471	列472	列473	列474	列475	列476	列477	列478	列479	列480	列481	列482	列483	列484	列485	列486	列487	列488	列489	列490	列491	列492	列493	列494	列495	列496	列497	列498	列499	列500	列501	列502	列503	列504	列505	列506	列507	列508	列509	列510	列511	列512	列513	列514	列515	列516	列517	列518	列519	列520	列521	列522	列523	列524	列525	列526	列527	列528	列529	列530	列531	列532	列533	列534	列535	列536	列537	列538	列539	列540	列541	列542	列543	列544	列545	列546	列547	列548	列549	列550	列551	列552	列553	列554	列555	列556	列557	列558	列559	列560	列561	列562	列563	列564	列565	列566	列567	列568	列569	列570	列571	列572	列573	列574	列575	列576	列577	列578	列579	列580	列581	列582	列583	列584	列585	列586	列587	列588	列589	列590	列591	列592	列593	列594	列595	列596	列597	列598	列599	列600	列601	列602	列603	列604	列605	列606	列607	列608	列609	列610	列611	列612	列613	列614	列615	列616	列617	列618	列619	列620	列621	列622	列623	列624	列625	列626	列627	列628	列629	列630	列631	列632	列633	列634	列635	列636	列637	列638	列639	列640	列641	列642	列643	列644	列645	列646	列647	列648	列649	列650	列651	列652	列653	列654	列655	列656	列657	列658	列659	列660	列661	列662	列663	列664	列665	列666	列667	列668	列669	列670	列671	列672	列673	列674	列675	列676	列677	列678	列679	列680	列681	列682	列683	列684	列685	列686	列687	列688	列689	列690	列691	列692	列693	列694	列695	列696	列697	列698	列699	列700	列701	列702	列703	列704	列705	列706	列707	列708	列709	列710	列711	列712	列713	列714	列715	列716	列717	列718	列719	列720	列721	列722	列723	列724	列725	列726	列727	列728	列729	列730	列731	列732	列733	列734	列735	列736	列737	列738	列739	列740	列741	列742	列743	列744	列745	列746	列747	列748	列749	列750	列751	列752	列753	列754	列755	列756	列757	列758	列759	列760	列761	列762	列763	列764	列765	列766	列767	列768	列769	列770	列771	列772	列773	列774	列775	列776	列777	列778	列779	列780	列781	列782	列783	列784	列785	列786	列787	列788	列789	列790	列791	列792	列793	列794	列795	列796	列797	列798	列799	列800	列801	列802	列803	列804	列805	列806	列807	列808	列809	列810	列811	列812	列813	列814	列815	列816	列817	列818	列819	列820	列821	列822	列823	列824	列825	列826	列827	列828	列829	列830	列831	列832	列833	列834	列835	列836	列837	列838	列839	列840	列841	列842	列843	列844	列845	列846	列847	列848	列849	列850	列851	列852	列853	列854	列855	列856	列857	列858	列859	列860	列861	列862	列863	列864	列865	列866	列867	列868	列869	列870	列871	列872	列873	列874	列875	列876	列877	列878	列879	列880	列881	列882	列883	列884	列885	列886	列887	列888	列889	列890	列891	列892	列893	列894	列895	列896	列897	列898	列899	列900	列901	列902	列903	列904	列905	列906	列907	列908	列909	列910	列911	列912	列913	列914	列915	列916	列917	列918	列919	列920	列921	列922	列923	列924	列925	列926	列927	列928	列929	列930	列931	列932	列933	列934	列935	列936	列937	列938	列939	列940	列941	列942	列943	列944	列945	列946	列947	列948	列949	列950	列951	列952	列953	列954	列955	列956	列957	列958	列959	列960	列961	列962	列963	列964	列965	列966	列967	列968	列969	列970	列971	列972	列973	列974	列975	列976	列977	列978	列979	列980	列981	列982	列983	列984	列985	列986	列987	列988	列989	列990	列991	列992	列993	列994	列995	列996	列997	列998	列999	列1000
---	---	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-------

```
data_FE.shape
```

```
#Correlation value range
```

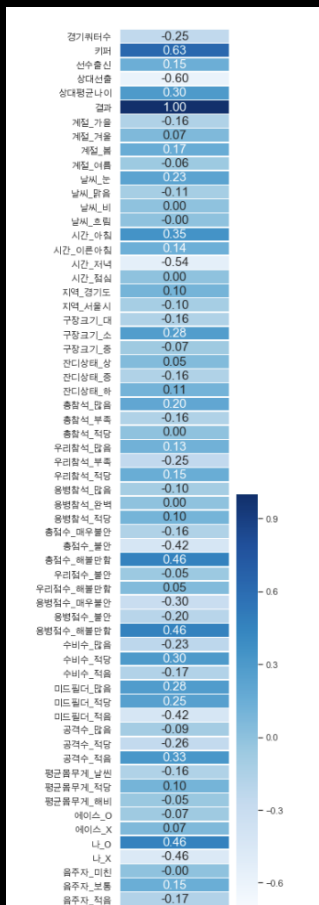
```
Image("https://hooon.s3.ap-northeast-2.amazonaws.com/SOSCON_2019/image/95.jpg")
```



```
data_FE_corr = data_FE.corr()
cols = data_FE_corr.columns.values
data_FE_corr_cols = list(cols)
for i in range(0, len(data_FE_corr_cols)):
    if data_FE.columns[i] != "결과":
        data_FE_corr.drop(data_FE_corr_cols[i], 1, inplace = True)

plt.figure(figsize = (3,30))
sns.heatmap(data_FE_corr, annot = True, fmt = '.2f', linewidths = .5, cmap = 'Blues', annot_kws={"size": 15})
```

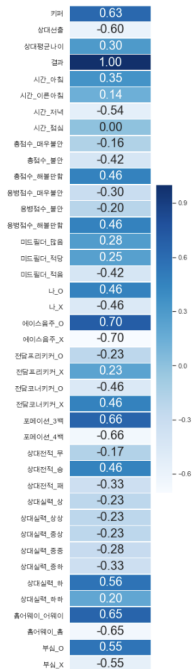




```
data.drop(["경기쿼터수", "선수출신", "계절", "날씨", "지역", "구장크기", "잔디상태", "총참석", "우리참석", "용병참석", "우리점수", "수비수", "공격수", "평균몸무게", "에이스", "음주자", "상대매너", "키퍼장갑", "득점", "실점"], 1, inplace = True)
```

data_FE.shape

(40, 41)

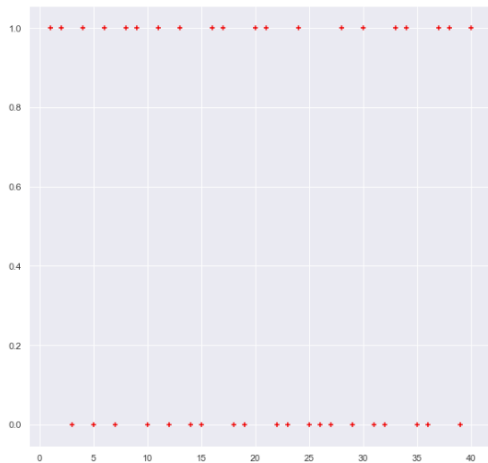


7. Modelling

- Logistic Regression
- kNN
- Decision tree
- Random Forest
- Naive Bayes
- SVM(support vector machine)

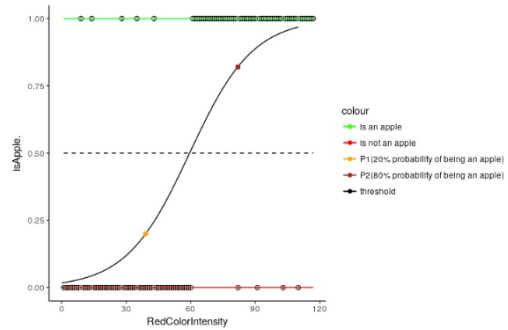
```
plt.figure(figsize = (10,10))  
plt.scatter(data_FE.index, data_FE["결과"], marker="+", color="red")
```

<matplotlib.collections.PathCollection at 0x124507390>



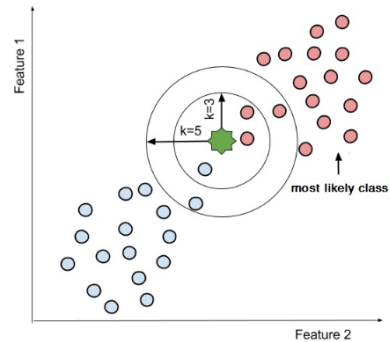
#Logistic Regression

Image(https://hooon.s3.ap-northeast-2.amazonaws.com/SOSCON_2019/image/100.jpg)



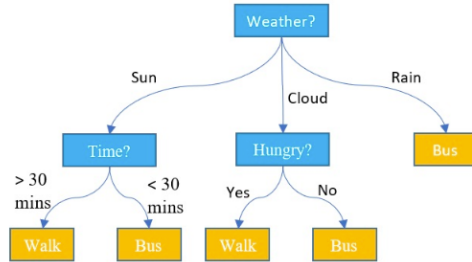
#kNN

Image(https://hooon.s3.ap-northeast-2.amazonaws.com/SOSCON_2019/image/101.jpg)



#Decision tree

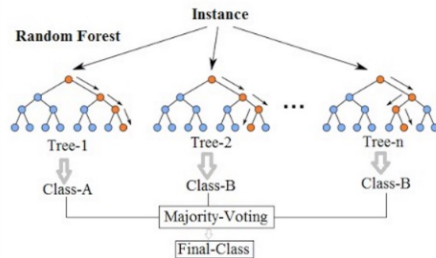
Image(https://hoon.s3.ap-northeast-2.amazonaws.com/SOSCON_2019/image/102.jpg)



#Random Forest

Image(https://hoon.s3.ap-northeast-2.amazonaws.com/SOSCON_2019/image/103.jpg)

Random Forest Simplified



#Naive bayse

Image("https://hoon.s3.ap-northeast-2.amazonaws.com/SOSCON_2019/image/104.jpg")

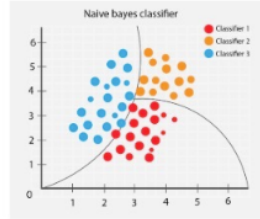
Naive Bayes

In machine learning, naive Bayes classifiers are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong (naive) independence assumptions between the features.

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

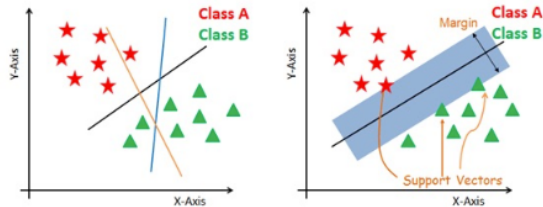
using Bayesian probability terminology, the above equation can be written as

$$\text{Posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$



#SVM(support vector machine)

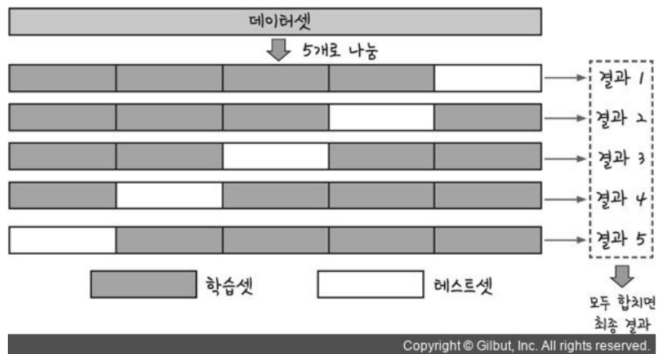
Image("https://hoon.s3.ap-northeast-2.amazonaws.com/SOSCON_2019/image/105.jpg")



```
from sklearn.linear_model import LogisticRegression
from sklearn.neighbors import KNeighborsClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn.svm import SVC
```

#k-fold cross validation

```
Image("https://hooon.s3.ap-northeast-2.amazonaws.com/SOSCON_2019/image/106.png")
```



```
x_train = data_FE.drop(["결과"], 1)
y_train = data_FE["결과"].values.reshape(data_FE["결과"].values.size, 1)
print(x_train.shape, y_train.shape)
y_train = data_FE["결과"].values
```

```
(40, 40) (40, 1)
```

```
from sklearn.model_selection import KFold
from sklearn.model_selection import cross_val_score
k_fold = KFold(n_splits=5, shuffle=True, random_state=0)
```



7.1.Logistic Regression

```
clf_logi = LogisticRegression()  
scoring = 'accuracy'  
score = cross_val_score(clf_logi, x_train, y_train, cv=k_fold, n_jobs=1, scoring=scoring)  
print("개별 점수 : ", score)  
print("평균 점수 : ", round(np.mean(score)*100, 2))
```

개별 점수 : [1. 1. 1. 1. 0.875]
평균 점수 : 97.5

7.2.kNN

```
clf_knn = KNeighborsClassifier(n_neighbors = 2)  
scoring = 'accuracy'  
score = cross_val_score(clf_knn, x_train, y_train, cv=k_fold, n_jobs=1, scoring=scoring)  
print("개별 점수 : ", score)  
print("평균 점수 : ", round(np.mean(score)*100, 2))
```

개별 점수 : [0.875 1. 1. 1. 0.875]
평균 점수 : 95.0

7.3.Decision tree

```
clf_dcs = DecisionTreeClassifier()  
scoring = 'accuracy'  
score = cross_val_score(clf_dcs, x_train, y_train, cv=k_fold, n_jobs=1, scoring=scoring)  
print("개별 점수 : ", score)  
print("평균 점수 : ", round(np.mean(score)*100, 2))
```

개별 점수 : [0.875 0.75 0.625 0.875 1.]
평균 점수 : 82.5

7.4.Random Forest

```
clf_rdf = RandomForestClassifier(n_estimators=10)  
scoring = 'accuracy'  
score = cross_val_score(clf_rdf, x_train, y_train, cv=k_fold, n_jobs=1, scoring=scoring)  
print("개별 점수 : ", score)  
print("평균 점수 : ", round(np.mean(score)*100, 2))
```

개별 점수 : [1. 1. 0.875 0.875 1.]
평균 점수 : 95.0

7.5.Naive Bayes

```
clf_nb = GaussianNB()  
scoring = 'accuracy'  
score = cross_val_score(clf_nb, x_train, y_train, cv=k_fold, n_jobs=1, scoring=scoring)  
print("개별 점수 : ", score)  
print("평균 점수 : ", round(np.mean(score)*100, 2))
```

개별 점수 : [0.875 1. 0.875 1. 0.875]
평균 점수 : 92.5

7.6.SVM(support vector machine)

```
clf_svm = SVC()  
scoring = 'accuracy'  
score = cross_val_score(clf_svm, x_train, y_train, cv=k_fold, n_jobs=1, scoring=scoring)  
print("개별 점수 : ", score)  
print("평균 점수 : ", round(np.mean(score)*100, 2))
```

개별 점수 : [1. 1. 0.875 0.875 0.75]
평균 점수 : 90.0

7.7.Coefficient - Logistic Regression

```
clf_logi.fit(x_train, y_train)
np.ravel(clf_logi.coef_)
dfCoef = pd.DataFrame({"coef" : data_FE.drop(["결과"], 1).columns, "weight" : np.ravel(clf_logi.coef_)})
dfCoef = dfCoef[(dfCoef['weight'] >= 0.3) | (dfCoef['weight'] <= -0.3)].set_index('coef')
dfCoef.style.background_gradient(cmap='summer')
```

coef	weight
커피	0.779536
상대선술	-0.845064
시간_저녁	-0.316962
총점수_불안	-0.355742
총점수_해불만함	0.336654
미드필드_척음	-0.381844
나_O	0.442686
나_X	-0.4717
에이스음주_O	0.586354
에이스음주_X	-0.615368
전담코너키커_O	-0.484787
전담코너키커_X	0.455774
포메이션_3백	0.445539
포메이션_4백	-0.474552
상대전적_승	0.413673
상대실력_하	0.365314
홀어웨이_어웨이	0.402252
홀어웨이_홀	-0.431266
부심_O	0.402868
부심_X	-0.431882



```
#test_1_1
```

[illegible]

상대선출	상대실력_중중	홈어웨이_홈
0	1	1

상대선출	상대실력_중상	홈어웨이_홈
3	1	1

상대선출	상대실력_중중	홈어웨이_어웨이
2	1	1

예측결과 : 승



9.HAPPY SOCCER

```
x_real_prediction = [1,1,0,1,1,0,1,1,1,1]
y_real =           [0,1,1,1,1,0,1,1,1,1]

count = 0

for i in range(0, len(y_real)):
    if y_real[i] == x_real_prediction[i]:
        count += 1

print("전체 경기 : ",len(y_real))
print("예측 성공 : ",count)
print("예측 성공률 : ",(count/len(y_real))*100,"%")
```

전체 경기 : 10
예측 성공 : 8
예측 성공률 : 80.0 %



Conclusion

분석의 결과로 행복한 추구를.



SOSCON 2019

SAMSUNG OPEN SOURCE CONFERENCE 2019

• 승패에 큰 영향을 주는 요소

- 키퍼 유무
- 상대팀 선수 출신 수
- 에이스 음주 여부
- 미디필더 수
- 포메이션
- 상대팀 홈구장 여부
- 부상 여부
- 감독의 참석 여부

coef	weight
키퍼	0.779536
상대선수	-0.845064
시간_저녁	-0.316962
총점수_불안	-0.355742
총점수_해불만함	0.336654
미드필더_적음	-0.381844
나_O	0.442686
나_X	-0.4717
에이스음주_O	0.586354
에이스음주_X	-0.615368
전담코너키퍼_O	-0.484787
전담코너키퍼_X	0.455774
포메이션_3백	0.445539
포메이션_4백	-0.474552
상대전적_승	0.413673
상대실력_하	0.365314
홈웨이_어웨이	0.402252
홈웨이_홈	-0.431266
부상_O	0.402868
부상_X	-0.431882



- 해당 주 상황에 맞는 재미있게 경기할 수 있는 상대를 선정하고, 이길 수 있는 확률을 높이는 방법
 - 일단 전담 키퍼 한 명은 무조건 나오게 한다. (GK 총 3명)
 - 상대팀에 선수 출신이 몇 명인지 정확히 파악한다. (거짓말하는지 여러번 체크)
 - 미드필더 수를 충분히 확보하고, 모자라면 용병은 가능한 미드필더로 구한다.
 - 경기 전날 에이스 술을 먹인다.
 - 포메이션은 무조건 3백으로 한다.
 - 경기장이 상대팀 홈 구장이면 피한다. 상대도 어웨이인 팀을 고른다.
 - 무조건 부심을 둔다.
 - 가능한 내가 참석 한다.



THANK YOU

SOSCON 2019

SAMSUNG OPEN SOURCE CONFERENCE 2019

