

'빅데이터로 스포츠 산업 바꾼다'... 데이터분석 스타트업들 맹활약

김동진 기자 | 승인 2021.09.03 10:52 | 댓글 0

빅데이터를 활용한 스포츠 분석 도입 빠르게 확산중
핏투게더, 데이터축구로 시리즈B 117억원 투자유치 성공
비프로컴퍼니 픽셀스코프 큐엠아이티 등도 스포츠 시장공략



출처:픽사베이

빅데이터와 데이터 분석 분야가 대두됨에 따라 다양한 분야에서 데이터 분석을 접목하려는 시도가 이루어지고 있다. 이러한 흐름에 발맞춰 스포츠 분야 역시 각자의 종목에 데이터 분석을 접목해 좀 더 나은 결과를 이끌어 내고자 하고 있다. 현재 다양한 기업과 스타트업 기업들이 스포츠 분야에 대해 데이터 분석을 시도하고 있고, 이러한 회사들은 데이터의 분석 뿐만 아니라 데이터의 수집 및 전처리까지 전체적인 데이터와 관련된 모든 업무를 처리하고 있다.



스포츠 분야, 특히 축구 쪽에서 데이터 분석을 통해 괄목할만한 성과를 거둔 축구팀이 있다.

바로 영국 프리미어리그의 브랜트포드 라는 팀이다. 브랜트포드는 데이터 분석을 통해 팀의 어린 선수들이 들이는 비용에 비해 의미있는 성적을 내지 못한다는 것을 파악했다. 브랜트포드의 데이터 분석 결과 , 유망한 어린 선수들은 부유한 다른 구단들로 옮겨 나갔고 이외의 선수들은 1군 즉 성인 무대에서 활약하기에는 부족한 성장세를 보인다는 사실을 알아냈다. 이를 근거로 브랜트포드는 과감하게 어린 선수들에 대해 지원을 끊고 , 이에 따라 나오던 정부의 보조금 역시 받지 않기로 결정한다. 대신 어린 선수가 아닌 2군 선수들 , 즉 즉시 전력감의 선수들에게 전폭적인 지원을 하기로 한다. 따라서 이들은 잉글랜드 하부의 3부 4부 (브랜트포드는 2부에 머무르던 팀이었다. 이번시즌 처음으로 1부 승격에 성공) 혹은 스페인, 프랑스 같은 주류 축구 리그 이외에 노르웨이, 덴마크 같은 비교적 변방리그의 즉전감의 유망주들을 눈여겨 보았고 이를 통해 좋은 성과를 거두었다.

그러자 2부리그 중상위권에서 머무르던 브랜트포드는 2부리그 3위를 달성하며 최초로 1부리그에 입성하게 되었다. 이 배경에는 즉시전력감의 선수를 키우는데 현재 상황에서 이득이라는 브랜트포드의 데이터 분석 결과이기도 했다. 또한 저비용 고효율이라는 컨셉 아래에 영입한 선수들은 다른 팀으로 이적할때 팀에 큰 금액을 안겨주며 떠나 성적 , 자금 두마리 토끼를 모두 잡는 선순환 구조가 만들어졌다.

그 결과 브랜트포드는 현재 1/3 정도 경기를 치른 epl 리그에서 선수단 몸값은 17위에 속하지만 순위는 14위에 위치하고 있다. 시즌 초반 보여주었던 폭발력과 , 2부리그에서 올라온 팀들이 대부분 올라오자마자 한시즌만에 다시 2부리그로 강등된다는 사실을 보면 이 역시 대단한 결과라는 것을 알 수 있다.

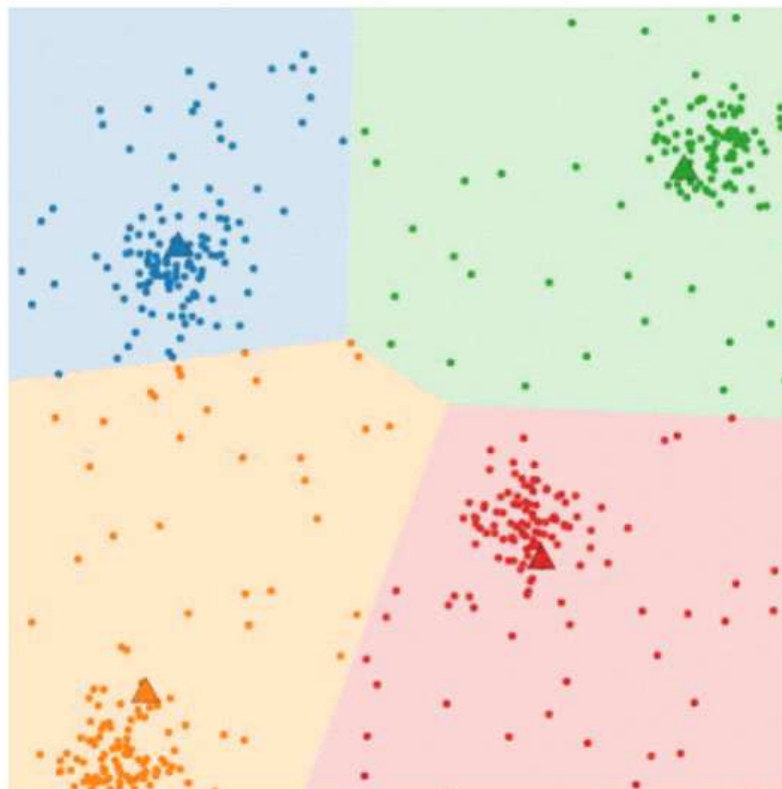
대부분의 스포츠 데이터 분석은 자신들의 팀을 분석하고 팀의 결함 , 팀의 장점등을 찾아내는 데에 치우친 반면 , 브랜트포드의 데이터 분석은 외부 선수를 분석하고 자신들의 팀에 얼마나 적합할지 혹은 성장 가능성이 얼마나 큰지 분석하는 스카우팅에 집중되어 있다. 브랜트포드의 경우를 통해 선수 스카우팅 분야 역시 , 데이터 분석이 유용하게 사용될 수 있다는 점을 알 수 있다. 따라서 이번 빅데이터 분석의 주제는 [K-means Clustering을 통한 축구 선수 군집화]로 결정하였다.

아래의 사진은 데이터 분석에 사용할 축구 선수에 관한 데이터들 이다. 데이터는 크게 3가지로 이루어져 있다. 공격(Offensive) , 수비(Defensive) , 패스(Passing) 으로 나누어 지고 이후 더 세분화 되어 나누어 진다. 이후 각각의 데이터에 대한 간략한 설명을 표로 정리하였고, 모델에 사용할 데이터 분석 알고리즘인 K-means Clustering에 대하여 소개한다.

-K-means Clustering

K-평균 알고리즘(K-means Clustering algorithm)은 주어진 데이터를 k개의 클러스터로 묶는 알고리즘으로, 각 클러스터와 거리 차이의 분산을 최소화 하는 방식으로 동작한다. 이 알고리즘은 자율 학습, 비지도 학습의 일종으로 레이블이 달려 있지 않은 입력 데이터에 대하여 레이블을 달아주는 역할을 수행한다. 이 알고리즘은 EM 알고리즘을 이용한 클러스터링과 비슷한 구조를 가지고 있다.

K-means Clustering 알고리즘은 클러스터링 방법 중 분할법에 속한다. 분할법은 주어진 데이터를 여러 파티션 (그룹) 으로 나누는 방법이다. 예를 들어 n개의 데이터 오브젝트를 받았다고 가정하면, 이때 분할법은 입력 데이터를 n보다 작거나 같은 k개의 그룹으로 나누는 것이다. 이때 각 그룹은 클러스터를 형성하고 데이터들을 한 개 이상의 데이터 오브젝트로 구성된 k개의 그룹으로 나누는 것이다. 이 때 그룹을 나누는 과정은 거리 기반의 그룹간 비유사도 (dissimilarity) 와 같은 비용 함수 (cost function)을 최소화 하는 방식으로 이루어지면, 이 과정에서 같은 그룹 내 데이터 오브젝트 끼리의 유사도는 증가하고, 다른 그룹에 있는 데이터 오브젝트와의 유사도는 감소하게 된다. K-means Clustering 알고리즘은 각 그룹의 중심 (Centroid)과 그룹 내의 데이터 오브젝트와의 거리의 제곱합을 비용 함수로 정하고, 이 함수 값을 최소화 하는 방향으로 각 데이터 오브젝트의 소속 그룹을 업데이트 해주어 클러스터링을 수행한다.



<K-mean Clustering의 예시 사진>

무작위 분할 알고리즘은 가장 많이 쓰이는 초기화 기법으로 각 데이터들을 임의의 클러스터에 배당한 후 , 각 클러스터에 배당된 점들의 평균값을 초기로 설정한다. 무작위 분할 기법의 경우 다른 기법들과 달리 데이터 순서에 대해 독립적이다.

무작위 분할의 경우 초기 클러스터가 각 데이터들에 대해 고르게 분포되기 때문에 각 초기 클러스터의 무게중심들이 데이터 집합의 중심에 가깝게 위치하는 경향을 띄게 된다. 이러한 특성 때문에 K-means Clustering 알고리즘에서는 무작위 분할이 선호된다.

또한 프로젝트를 진행함에 있어 Train Set과 Test Set , 모델이 이루어지는 과정과 결과는 팀원 모두가 동일 해야 한다. 같은 Input 데이터와 같은 과정으로 나온 Output을 통해 같은 결과를 도출해야 하기 때문이다.

따라서 Train Set과 Test set을 나누는 random_state와 , sklearn의 K-means Clustering 의 random_state를 동일하게 할 필요가 있다. random_state는 무작위 수를 생성하는 방법을 나타낸다. Train Set과 Test Set을 나누는 과정에서 사용되는 random_states는 numpy 에서 사용되는 np.random에서 사용되는 random_state 이다. K-means Clustering 알고리즘에서도 매개변수로 random_state를 지정할 수 있고 정수도 사용할 수 있다. 이 정수는 무작위 피드 , 즉 클래스 번호의 시작값을 나타낸다.

```
from sklearn.cluster import KMeans
kmeans = KMeans(n_clusters=n_cluster, random_state=42)
```

<random_state K-means 코드>

```
#모델에 대한 결과 확인 예시 코드 - train and test / 모델 = 42 / train-test = 13
from sklearn.cluster import KMeans
k = 300 #k-means 파라미터 k 개수 -> 50 ~ 300

kmeans = KMeans(n_clusters=k, random_state=42)
kmeans.fit(df_scaled_train)

df_test_result = kmeans.predict(df_scaled_test)

result_by_sklearn = df_scaled_train.copy()
result_by_sklearn["cluster"] = kmeans.labels_

test_result = df_scaled_test.copy()
test_result["cluster"] = df_test_result

result_by_sklearn.head()
result_by_sklearn.info()
```

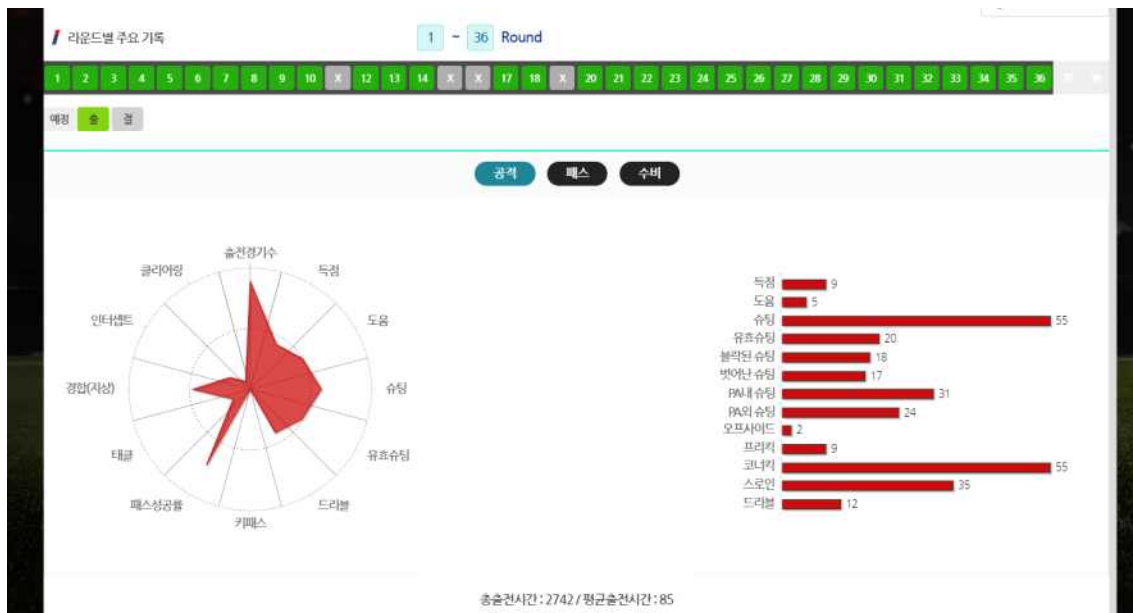
<최종 K-means Clustering 모델 코드>

Premier League Player Statistics

| Summary | | | | | | | | | | | |
|--|------|------|-------|---------|-----|-----|-----|------|------------|------|--------|
| Defensive | | | | | | | | | | | |
| Offensive | | | | | | | | | | | |
| Passing | | | | | | | | | | | |
| Detailed | | | | | | | | | | | |
| Overall | | | | | | | | | | | |
| Home | | | | | | | | | | | |
| Away | | | | | | | | | | | |
| Minimum apps | | | | | | | | | | | |
| All players | | | | | | | | | | | |
| Player | Apps | Mins | Goals | Assists | Yel | Red | SpG | PS% | AerialsWon | MotM | Rating |
| 1 Mohamed Salah Liverpool, 29, AM(CLR),FW | 11 | 990 | 10 | 7 | 1 | - | 4 | 81.3 | 0.4 | 4 | 8.18 |
| 2 Trent Alexander-Arnold Liverpool, 23, D(R),M(R) | 9 | 784 | 1 | 4 | 1 | - | 1.9 | 81.4 | 1 | 3 | 7.70 |
| 3 Mateo Kovacic Chelsea, 27, M(CL) | 8(1) | 705 | 1 | 5 | 1 | - | 1.3 | 88.6 | 0.2 | - | 7.56 |
| 4 Trevoh Chalobah Chelsea, 22, D(CR),DMC | 5(1) | 453 | 2 | - | - | - | 0.3 | 88.3 | 1.7 | 1 | 7.56 |
| 5 Conor Gallagher Crystal Palace, 21, AM(C) | 10 | 877 | 4 | 2 | 3 | - | 2.3 | 79.8 | 0.7 | 3 | 7.50 |
| 6 Thiago Silva Chelsea, 37, D(C) | 6(2) | 646 | 1 | - | 1 | - | 0.8 | 89.7 | 2.4 | 1 | 7.47 |
| 7 Marcos Alonso Chelsea, 30, D(CL),M(L) | 6 | 536 | 1 | 1 | 2 | - | 1.8 | 82.4 | 1.2 | 1 | 7.47 |
| 8 Declan Rice West Ham, 22, D(C),DMC | 11 | 990 | 1 | 3 | 3 | - | 0.6 | 91.7 | 1.4 | 1 | 7.46 |
| 9 Jarrod Bowen West Ham, 24, AM(CLR),FW | 11 | 966 | 2 | 4 | 1 | - | 2.5 | 77.9 | 0.6 | - | 7.42 |
| 10 Raphinha Leeds, 24, AM(LR) | 10 | 820 | 5 | - | 3 | - | 3.5 | 69.5 | 0.5 | 3 | 7.42 |

© WhoScored

Page 1/29 | Showing 1 - 10 of 288 first | prev | next | last



-데이터 Feature 설명

| 데이터 Feature | 설명 |
|-------------|-----------------------------------|
| Season | 선수가 경기를 뒀 해당 시즌 (ex 2020 - 2021) |
| League | 선수가 속한 리그 |
| Name | 선수의 이름, 이번 데이터에는 소속팀까지 포함 |
| Age | 선수의 나이 |
| Position1 | 선수의 주 포지션 |
| Position2 | 선수의 부 포지션 |
| Apps | 선수가 한시즌에 뒀 선발 경기수, 괄호 안은 교체 출전 |
| Mins | 선수가 한시즌에 뒀 경기 분수 (ex 900 => 900분) |

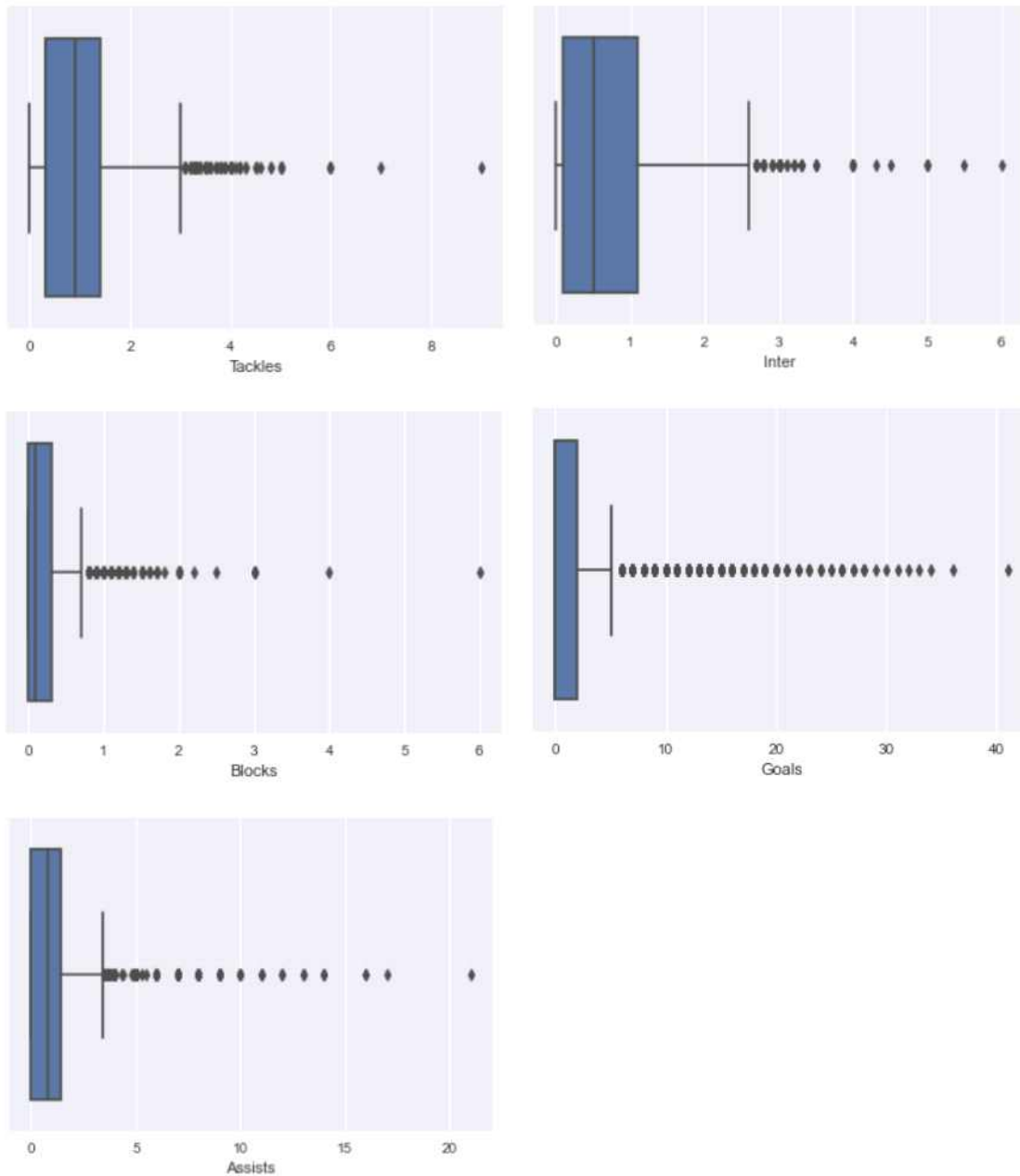
| | |
|----------|------------------------|
| Tackles | 경기당 평균 태클 수 |
| Inter | 경기당 평균 인터셉트(공을 빼앗은) 횟수 |
| Fouls | 경기당 경고 횟수 |
| Offsides | 경기당 오프사이드 승리 횟수 (수비) |
| Clear | 경기당 공을 걷어낸 횟수 |
| Drb(수비) | 경기당 드리블을 통과 시킨 횟수 (수비) |
| Blocks | 경기당 상대 선수를 막은 횟수 |
| OwnG | 기록한 자책골 수 |
| Goals | 전체 골 기록 수 |
| Assists | 전체 도움 기록 수 |
| SpG | 경기당 슈팅 횟수 |
| KeyP | 경기당 키 패스 (주요한 패스) 횟수 |
| Drb(공격) | 경기당 드리블 횟수 |
| Fouled | 경기당 파울당한 횟수 |
| Off | 경기당 오프사이드 횟수 |
| Disp | 경기당 공 소유권을 잃은 횟수 |
| UnsTch | 경기당 좋지 않은 볼 컨트롤 횟수 |
| AvgP | 경기당 패스 횟수 |
| PS% | 패스 성공률 (퍼센테이지 %) |
| Crosses | 경기당 크로스 횟수 |
| LongB | 경기당 롱볼 횟수 |
| ThrB | 경기당 던지기 횟수 |
| Rating | 선수의 평균 평점 |

| Player | CM | KG | Apps | Mins | Goals | Assists | SpG | KeyP | Drb | Fouled | Off | Disp | UnsTch | Rating |
|--|-----|----|------|------|-------|---------|-----|------|-----|--------|-----|------|--------|--------|
| 1  Son Heung-Min 29, M(CLR),FW | 183 | 78 | 10 | 875 | 4 | 1 | 2.4 | 1.8 | 1.5 | 0.7 | 0.8 | 2.1 | 2.3 | 6.88 |

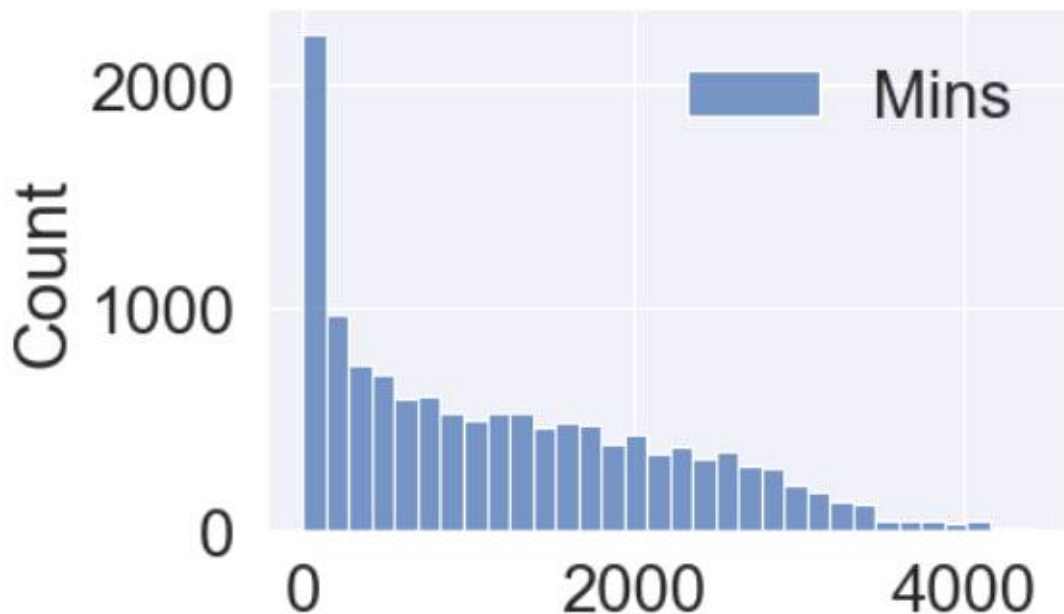
위의 사진은 손흥민 선수에 대한 Offensive 공격 데이터 이다. 이를 통해 손흥민 선수는 한경기당 평균 2.4개의 슈팅을 시도하고 현재까지 4골 , 1도움을 기록했다는 것을 알 수 있다. 또한 지금까지 보여준 키패스(주요한 패스)는 평균 1.8개 임을 알 수 있다.

축구 데이터 분야 쪽에서 어떤 선수를 살펴볼때 가장 중요하게 여겨지는 데이터 Feature는 'Tackles', 'Inter', 'Blocks', 'Goals', 'Assists'로 알려져 있다. 따라서 위 데이터들을 가지고 데이터 분석을 진행하기 전에 위의 주요한 데이터들이 어떠한 형태를 띠고 있는지 확인해볼 필요가 있다.

아래 그림들은 'Tackles', 'Inter', 'Blocks', 'Goals', 'Assists'에 대한 Box Plot들 이다. 해당 Feature들에 대하여 Box Plot들을 살펴본 결과 , Outlier가 생각 보다 많이 존재한다는 점을 알 수 있었다. 데이터 분석에 있어 Outlier를 제거하고 분석을 진행할지 , 그대로 진행할지에 대하여 팀원들과 회의를 진행하였고 , 결론은 제거하지 않고 진행하기로 하였다. 축구 데이터의 관점으로 Outlier를 살펴보았을 때 , 각각의 Outlier들은 일반적인 선수가 아닌 특별한 선수들 이라고 볼 수 있다. 이런 특별한 선수들을 제거하고 데이터 분석을 진행하는 것은 의미있는 결과를 도출하기 어렵다고 판단하였고 , 전체 데이터를 통해 데이터 분석을 진행하였다.



대신 Outlier를 바라보는 관점을 다시 보기로 하였다. 일반적인 경우 Outlier는 잘못 평가된 값으로, 결과적으로 잘못된 분석결과를 초래할 수 있는 값들을 말한다. 축구 데이터의 관점으로 잘못된 분석결과를 초래할 수 있는 값들을 살펴본 결과, 그러한 값들은 경기수가 부족한 선수들 이라고 판단하였다. 즉 한경기만 잘해 모든 피쳐값이 우수하게 나온 선수와 40경기를 잘해 모든 피쳐값이 우수하게 나온 선수를 같게 비교해서는 안된다는 것이다. 즉 선수 데이터의 경기수는 선수가 보여주는 퍼포먼스에 대한 신뢰도 이고 이런 신뢰도가 낮을 수록 노이즈한 데이터들이 자주 발생하게 된다. 이러한 관점을 바탕으로 Mins Feature (한 시즌에 뛴 경기 분수) 값이 특정값 이하인 데이터들을 Outlier 라고 판단하여 데이터 분석을 진행하였다.



위의 데이터를 통해 전체 데이터들을 살펴본 결과 Mins < 900 인 데이터가 8144개 존재함을 발견하였다. 이는 다시말해 한 시즌 즉 1년에 10경기도 채 못뛰는 선수 데이터들이 8144개 존재함을 의미하고 , 이는 전체 약 2만여개의 데이터들중 40%에 해당하는 수치다. 따라서 Mins Feature는 선수 데이터에 대한 신뢰도로 사용해 데이터를 분석하는 모델에 경계값을 넣어 사용할 데이터들을 선택할 수 있도록 두었다.

또한 일반적으로 주요하다고 알려진 데이터 Feature들인 'Tackles', 'Inter', 'Blocks', 'Goals', 'Assists'를 선택하여 데이터 분석을 진행하는 것처럼 , 원하는 Feature 만을 선택해 데이터 분석 모델을 사용할 수 있는 경우도 고려하였다. 만약 어떤 선수의 자책골이 5개라면 이러한 부분까지 닻을 필요는 없다고 생각할 수 있다. 이런 경우 자책골 Feature를 제외하고 데이터 분석을 진행한다.

따라서 수집한 축구 선수들의 데이터를 가지고 데이터 분석 모델을 세울때 , 선택할 수 있는 Parameter 들은 다음과 같다.

1. 제외할 데이터의 Mins 기준값
2. 시나리오에 따라 사용할 Feature
3. K-means Clustering 에서의 K

실제 스카우팅 과정을 살펴보고 모델이 적용될 수 있는 경우를 살펴본 결과 , 모델이 사용가능한 스카우팅 시나리오는 크게 2가지로 나뉜다.

1. 특출난 선수급의 활약을 보여주는 유명하지 않은 선수
2. 팀의 주전급 선수가 이탈하게 되어 대신 할 선수 찾기

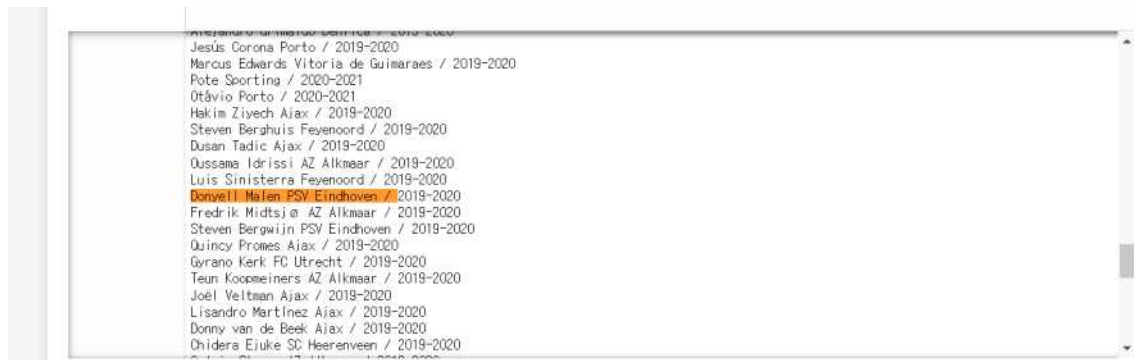
1번 시나리오의 경우 특출난 선수급의 활약을 보여주는 기준 선수를 찾아내야 한다. 이때 축구 전문가들이 heuristics 하게 선수의 퍼포먼스를 평가한 Rating Feature를 사용한다. Rating Feature를 Interquartile range로 확인하여 Outlier들을 찾아낸다. 앞서 축구 선수 데이터의 Outlier들은 특별하고 일반적이지 않은 선수들이었고, 따라서 선수의 퍼포먼스를 수치적으로 나타낸 Rating Feature에서 역시 동일 하게 적용된다.

이후 각 시나리오 별로 살펴볼 사항은 [한 클러스터 안에 어떠한 선수들이 포함되었는지], [한 클러스터에 속한 선수들이 얼마나 닮아있는지 Feature 값으로 확인하기] 다. 또한 K-means Clustering을 진행 할때, 시나리오에 대한 정확도를 검증하기 위하여 Train 데이터와 Test 데이터로 데이터를 나누어 데이터 분석을 진행한다.

예시로 살펴볼 1번과 2번 시나리오 에서는 Parameter를 다음과 같이 정하고 진행한다.

-시나리오 Parameter

1. Mins < 900
2. 전체 피쳐 모두 사용
3. K = 50 / 100 / 150 / 200 / 250 / 300



위 사진은 Rating Feature를 Interquartile range로 확인하여 찾아낸 Outlier 선수들의 목록이다. 위 사진에 포함된 선수들은 모두 특출나고 뛰어난 활약을 보여주고 있는 선수들이다. 그중 네덜란드 리그의 PSV Eindhoven 팀에 속한 Donyell Malen 이라는 선수와 유사한 선수를 찾아보도록 한다.

아래는 K의 경우에 따라 해당 선수가 속한 클러스터와 그 클러스터 안의 선수들 사진이다.

기준 선수

season 2019-2020
league Eredivisie
name Donyell Malen PSV Eindhoven
Apps 14
cluster 48
Name: 7042, dtype: object

k = 50

| season | league | name W |
|--------|-----------|--|
| 15 | 2020-2021 | SuperLig Cyle Larin Besiktas |
| 81 | 2019-2020 | RussiaPremierLeague Eldor Shomurodov FC Rostov |
| 148 | 2020-2021 | Ligue1 Gaëtan Laborde Montpellier |
| 245 | 2019-2020 | LaLiga Gerard Moreno Villarreal |
| 508 | 2019-2020 | Bundesliga2 Manuel Schäffler Wehen Wiesbaden |
| 626 | 2020-2021 | Premiership Odsonne Édouard Celtic |
| 666 | 2019-2020 | LigaNOS Moussa Marega Porto |
| 756 | 2020-2021 | LaLiga Kike Garcia Eibar |
| 873 | 2019-2020 | LeagueTwo James Vaughan Bradford |
| 888 | 2019-2020 | ItalySerieA Andrea Belotti Torino |
| 891 | 2019-2020 | LeagueOne Ronan Curtis Portsmouth |
| 1042 | 2020-2021 | Championship Kieffer Moore Cardiff |
| 1286 | 2019-2020 | Bundesliga Sebastian Andersson Union Berlin |
| 1329 | 2019 | Brazil Everaldo Chapecoense AF |
| 1417 | 2019-2020 | LeagueTwo Jerry Yates Swindon |
| 1481 | 2020-2021 | Championship André Ayew Swansea |
| 1482 | 2020-2021 | LeagueTwo James Vaughan Tranmere |
| 1570 | 2019-2020 | Championship James Collins Luton |
| 1715 | 2020-2021 | Eredivisie Sebastian Polter Fortuna Sittard |
| 1775 | 2020-2021 | LeagueOne Jerry Yates Blackpool |
| 1938 | 2020-2021 | JupilerProLeague Thomas Henry OH Leuven |
| 1984 | 2019 | USA Diego Rossi Los Angeles FC |
| 1988 | 2020-2021 | RussiaPremier Alexander Sobolev Spartak Moscow |
| 2080 | 2020-2021 | SuperLig Vincent Aboubakar Besiktas |
| 2108 | 2020-2021 | LeagueTwo Tom Nichols Crawley |
| 2214 | 2020-2021 | LeagueTwo Conor Wilkinson Leyton Orient |

기준 선수

season 2019-2020
league Eredivisie
name Donyell Malen PSV Eindhoven
Apps 14
cluster 46
Name: 7042, dtype: object

k = 100

| season | league | name W |
|--------|-----------|---|
| 245 | 2019-2020 | LaLiga Gerard Moreno Villarreal |
| 626 | 2020-2021 | Premiership Odsonne Édouard Celtic |
| 688 | 2019-2020 | RussiaPremierLeague Artem Dzyuba Zenit |
| 1806 | 2020-2021 | Ligue1 Andy Delort Montpellier |
| 1829 | 2020-2021 | Bundesliga2 Marvin Ducksch Hannover |
| 1984 | 2019 | USA Diego Rossi Los Angeles FC |
| 2064 | 2019-2020 | Bundesliga Serge Gnabry Bayern |
| 2080 | 2020-2021 | SuperLig Vincent Aboubakar Besiktas |
| 2128 | 2019-2020 | PremierLeague Marcus Rashford Man Utd |
| 2249 | 2019 | USA Wayne Rooney DC United |
| 2421 | 2020-2021 | Ligue1 Wissam Ben Yedder Monaco |
| 2435 | 2019-2020 | PremierLeague Raheem Sterling Man City |
| 2547 | 2020 | USA Kacper Przybylko Philadelphia |
| 3500 | 2020-2021 | Championship Arnaut Danjuma Bournemouth |
| 3704 | 2019-2020 | SuperLig Bangaly-Fodé Koita Kasimpasa |
| 3719 | 2020-2021 | RussiaPremier Jordan Larsson Spartak Moscow |
| 3824 | 2020-2021 | Eredivisie Rai Vloet Heracles |
| 3911 | 2020-2021 | PremierLeague Marcus Rashford Man Utd |
| 3936 | 2019-2020 | RussiaPremierLeague Sardar Azmoun Zenit |
| 3985 | 2020-2021 | JupilerProLeague Noa Lang Club Bruges |
| 4079 | 2019-2020 | Bundesliga Andrej Kramaric Hoffenheim |
| 4156 | 2019-2020 | Bundesliga2 Sonny Kittel Hamburg |
| 4276 | 2019-2020 | Championship Jarrod Bowen Hull |
| 4351 | 2020-2021 | JupilerProLeague Roman Yaremchuk Gent |
| 4402 | 2020-2021 | Eredivisie Danilo Twente |

기준 선수

season 2019-2020
league Eredivisie
name Donyell Malen PSV Eindhoven
Apps 14
cluster 74
Name: 7042, dtype: object

k = 150

| season | league | name W |
|--------|-----------|--|
| 173 | 2020-2021 | LigaProfesional Martin Ojeda Godoy Cruz |
| 183 | 2020-2021 | ItalySerieA Lorenzo Insigne Napoli |
| 245 | 2019-2020 | LaLiga Gerard Moreno Villarreal |
| 626 | 2020-2021 | Premiership Odsonne Édouard Celtic |
| 1984 | 2019 | USA Diego Rossi Los Angeles FC |
| 2128 | 2019-2020 | PremierLeague Marcus Rashford Man Utd |
| 2435 | 2019-2020 | PremierLeague Raheem Sterling Man City |
| 2554 | 2020-2021 | Ligue1 Kylian Mbappé PSG |
| 2972 | 2019-2020 | RussiaPremierLeague Aleksey Miranchuk Lokomotiv Moscow |
| 3719 | 2020-2021 | RussiaPremier Jordan Larsson Spartak Moscow |
| 3756 | 2020-2021 | ItalySerieA Domenico Berardi Sassuolo |
| 3824 | 2020-2021 | Eredivisie Rai Vloet Heracles |
| 3985 | 2020-2021 | JupilerProLeague Noa Lang Club Bruges |
| 4243 | 2019-2020 | ItalySerieA Paulo Dybala Juventus |
| 4276 | 2019-2020 | Championship Jarrod Bowen Hull |
| 5008 | 2020 | Brazil Claudinho Red Bull Bragantino |
| 5624 | 2020-2021 | LaLiga Gerard Moreno Villarreal |
| 6235 | 2020-2021 | LigaProfesional Julián Álvarez River Plate |
| 6376 | 2020-2021 | ItalySerieA Luis Muriel Atalanta |
| 6631 | 2019-2020 | Championship Said Benrahma Brentford |
| 7042 | 2019-2020 | Eredivisie Donyell Malen PSV Eindhoven |
| 7249 | 2019-2020 | PremierLeague Anthony Martial Man Utd |
| 7514 | 2020-2021 | LigaNOS Pote Sporting |
| 7521 | 2020-2021 | JupilerProLeague Théo Bongonda Genk |
| 7678 | 2020-2021 | RussiaPremier Nikola Vlasic CSKA Moscow |
| 7949 | 2019-2020 | RussiaPremierLeague Nikola Vlasic CSKA Moscow |

기준 선수

season 2019-2020
league Eredivisie
name Donyell Malen PSV Eindhoven
Apps 14
cluster 112
Name: 7042, dtype: object

k = 200

| season | league | name W |
|--------|-----------|--|
| 79 | 2019-2020 | LigaProfesional Nahuel Bustos Talleres |
| 140 | 2020 | USA Jordan Morris Seattle |
| 750 | 2020 | USA Romell Quioto Montreal |
| 853 | 2020-2021 | LigaNOS Paulinho Braga |
| 872 | 2020-2021 | LigaProfesional Sebastián Sosa Patronato de Parana |
| 1055 | 2019 | USA Brian Fernández Portland |
| 2366 | 2019-2020 | LeagueTwo Kevin van Veen Scunthorpe |
| 2547 | 2020 | USA Kacper Przybylko Philadelphia |
| 2716 | 2019 | USA Raúl Ruidiaz Seattle |
| 2746 | 2019-2020 | RussiaPremierLeague Djorde Despotovic FC Orenburg |
| 2865 | 2019-2020 | LeagueTwo Allie May Cheltenham |
| 2938 | 2020-2021 | JupilerProLeague Percy Tau Anderlecht |
| 3021 | 2019-2020 | LigaProfesional Gabriel Ávalos Patronato de Parana |
| 3704 | 2019-2020 | SuperLig Bangaly-Fodé Koita Kasimpasa |
| 3973 | 2020-2021 | Championship Josh Koroma Huddersfield |
| 4011 | 2020-2021 | Bundesliga Nicolás González Stuttgart |
| 4056 | 2019-2020 | RussiaPremierLeague Vladimir Obukhov FC Tambov |
| 4079 | 2019-2020 | Bundesliga Andrej Kramaric Hoffenheim |
| 4781 | 2020-2021 | Bundesliga Christopher Nkunku RBL |
| 4906 | 2019 | USA Gustavo Bou New England |
| 5648 | 2020-2021 | Eredivisie Eran Zahavi PSV Eindhoven |
| 5756 | 2019-2020 | Eredivisie Nicolai Jørgensen Feyenoord |
| 5969 | 2019-2020 | LigaProfesional Juan Manuel García Arsenal Sarandi |
| 6083 | 2020-2021 | JupilerProLeague Krépin Diatta Club Bruges |
| 7042 | 2019-2020 | Eredivisie Donyell Malen PSV Eindhoven |
| 7079 | 2020 | USA Raúl Ruidiaz Seattle |

기준 선수

```
season      2019-2020
league      Eredivisie
name        Donyell Malen PSV Eindhoven
Apps        14
cluster     20
Name: 7042, dtype: object
```

| | season | league | name | Apps | W |
|------|-----------|-----------------|----------------------------------|--------|---|
| 173 | 2020-2021 | LigaProfesional | Martin Ojeda Godoy Cruz | 21(1) | |
| 245 | 2019-2020 | LaLiga | Gerard Moreno Villarreal | 33(2) | |
| 626 | 2020-2021 | Premiership | Odsonne Édouard Celtic | 23(3) | |
| 1984 | 2019 | USA | Diego Rossi Los Angeles FC | 33(1) | |
| 2080 | 2020-2021 | SuperLig | Vincent Aboubakar Besiktas | 25(1) | |
| 2128 | 2019-2020 | PremierLeague | Marcus Rashford Man Utd | 31 | |
| 2435 | 2019-2020 | PremierLeague | Raheem Sterling Man City | 30(3) | |
| 2554 | 2020-2021 | Ligue1 | Kylian Mbappé PSG | 27(4) | |
| 3500 | 2020-2021 | Championship | Arnaut Danjuma Bournemouth | 29(4) | |
| 3719 | 2020-2021 | RussiaPremier | Jordan Larsson Spartak Moscow | 29 | |
| 4276 | 2019-2020 | Championship | Jarrod Bowen Hull | 29 | |
| 5591 | 2019-2020 | PremierLeague | Gabriel Jesus Man City | 21(13) | |
| 5624 | 2020-2021 | LaLiga | Gerard Moreno Villarreal | 30(3) | |
| 5644 | 2020-2021 | Ligue1 | Karl Toko Ekambi Lyon | 34(1) | |
| 6222 | 2019-2020 | SuperLig | Alexander Sørloth Trabzonspor | 34 | |
| 6235 | 2020-2021 | LigaProfesional | Julian Álvarez River Plate | 17(3) | |
| 7042 | 2019-2020 | Eredivisie | Donyell Malen PSV Eindhoven | 14 | |
| 7249 | 2019-2020 | PremierLeague | Anthony Martial Man Utd | 31(1) | |
| 7514 | 2020-2021 | LigaNOS | Pote Sporting | 32 | |
| 7533 | 2020-2021 | SuperLig | Aaron-Salem Boupemenda Hatayspor | 25(11) | |
| 7678 | 2020-2021 | RussiaPremier | Nikola Vlasic CSKA Moscow | 26 | |
| 7893 | 2019 | USA | Kacper Przybylko Philadelphia | 25(1) | |

기준 선수

```
season      2019-2020
league      Eredivisie
name        Donyell Malen PSV Eindhoven
Apps        14
cluster     52
Name: 7042, dtype: object
```

| | season | league | name | Apps | W |
|------|-----------|---------------|----------------------------------|-------|---|
| 626 | 2020-2021 | Premiership | Odsonne Édouard Celtic | 23(3) | |
| 750 | 2020 | USA | Romell Quioto Montreal | 15(1) | |
| 1988 | 2020-2021 | RussiaPremier | Alexander Sobolev Spartak Moscow | 16(6) | |
| 2080 | 2020-2021 | SuperLig | Vincent Aboubakar Besiktas | 25(1) | |
| 3500 | 2020-2021 | Championship | Arnaut Danjuma Bournemouth | 29(4) | |
| 3878 | 2020-2021 | PremierLeague | Michail Antonio West Ham | 24(2) | |
| 4402 | 2020-2021 | Eredivisie | Danilo Twente | 31(2) | |
| 4561 | 2019-2020 | Eredivisie | Myron Boadu AZ Alkmaar | 24 | |
| 4906 | 2019 | USA | Gustavo Bou New England | 13(1) | |
| 4931 | 2019-2020 | ItalySerieA | Duván Zapata Atalanta | 25(3) | |
| 5993 | 2020-2021 | PremierLeague | Mohamed Salah Liverpool | 34(3) | |
| 6056 | 2020-2021 | ItalySerieA | Duván Zapata Atalanta | 29(8) | |
| 7042 | 2019-2020 | Eredivisie | Donyell Malen PSV Eindhoven | 14 | |
| 7249 | 2019-2020 | PremierLeague | Anthony Martial Man Utd | 31(1) | |

K = 50 ~ 300 의 경우에 대하여 클러스터 안에 속한 전체 선수 숫자는 다음과 같다.

K = 50 : 100명 / 100 : 50명 / 150 : 26명 / 200 : 29명 / 250 : 22명 / 300 : 14명

K가 300 인 경우, 남아있는 14명의 선수는 대부분 Rating Feature에 대한 Outlier들 이었다. 또한 그렇지 않은 경우도 대부분의 사람들 혹은 축구 전문가들이 잘 알고있는 이미 유명한 선수 였다. 시나리오의 목표는 특출난 선수와 비슷한 퍼포먼스를 보여주는 유명하지 않은 선수를 찾는 것 이므로 , K가 300 이상인 모델의 결과는 의미 없다고 판단 하였다. 따라서 해당 데이터 사이즈와 1번 시나리오를 고려할 때 , K < 300 으로 진행하여야 의미있는 결과를 얻을 수 있다.

따라서 K=200인 경우에 대해 해당 클러스터에 속한 임의의 선수에 대한 Feature값을 이번 시나리오의 기준이 된 Donyell Malen 선수와 비교해보았다.

살펴보고 싶은 클러스터 번호 입력 : 112

살펴보는 클러스터 번호 : 112

| | season | league | name | Apps | cluster |
|------|-----------|------------------|----------------------------|-------|---------|
| 746 | 2020-2021 | Championship | Tyrese Campbell Stoke | 13(3) | 112 |
| 1001 | 2020-2021 | PremierLeague | Jesse Lingard West Ham | 16 | 112 |
| 1251 | 2019-2020 | LeagueOne | Liam Boyce Burton | 17(8) | 112 |
| 1484 | 2020-2021 | Eredivisie | Mimoun Mahi FC Utrecht | 14(8) | 112 |
| 1708 | 2020-2021 | JupilerProLeague | Tarik Tissoudali Beerschot | 16(3) | 112 |
| 1741 | 2019-2020 | LigaProfesional | Walter Bou Union | 20 | 112 |
| 2045 | 2020 | USA | Diego Rossi Los Angeles FC | 16 | 112 |
| 2160 | 2020-2021 | JupilerProLeague | Klauss Standard Liege | 10(3) | 112 |
| 2281 | 2020-2021 | PremierLeague | Gabriel Jesus Man City | 22(7) | 112 |

위의 결과를 확인한 결과 , K = 200 인 경우 Donyell Malen 선수는 112번 클러스터에 속해 있다. Test 데이터 셋에 속한 선수들을 해당 모델에 Fitting 한 결과 112번 클러스터에 속하는 선수들은 위와 같다. 해당 선수들 중 벨기에 리그의 Tarik Tissoudali 선수와 Donyell Malen 선수의 Feature 값을 비교해 유사도를 검증한다. 또한 시나리오는 특출난 선수와 비슷한 덜 유명한 선수이다. 해당 선수가 속한 벨기에 리그는 유럽 리그 랭킹 13위의 리그로 네덜란드가 7위 인것을 비교하면 리그 수준도 비교적 낮고 , 덜 유명하다는 것을 알 수 있다.

| season | league | name | age | position1 | position2 | Apps | Mins | Tackles | Inter | Fouls | Offsides | Clear | Drb | Blocks | OwnG | Goals | Assists |
|-----------|-------------|----------------|-----|-----------|-----------|-------|------|---------|-------|-------|----------|-------|-----|--------|------|-------|---------|
| 2019-2020 | Eredivisie | 7 Donyell M. | 22 | AM(L) | FW | 14 | 1191 | 0.3 | 0.1 | 0.6 | 0.6 | 0.5 | 0.4 | 0.1 | - | 11 | 2 |
| 2020-2021 | JupilerProL | 18 Tarik Tisso | 28 | AM(L) | FW | 16(3) | 1312 | 0.7 | 0.2 | 0.7 | 0.7 | 0.3 | 0.5 | - | - | 8 | 5 |

| SpG | KeyP | Drb | Fouled | Off | Disp | UnsTch | AvgP | PS% | Crosses | LongB | ThrB | Rating |
|-----|------|-----|--------|-----|------|--------|------|------|---------|-------|------|--------|
| 4.2 | 1.6 | 1.2 | 1.1 | 0.5 | 1.6 | 2.9 | 27.6 | 78 | 0.1 | 0.4 | 0.1 | 7.59 |
| 2.6 | 1.4 | 0.9 | 1.6 | 0.5 | 2.3 | 2.1 | 22.3 | 66.5 | 0.3 | 0.5 | 0.1 | 7.21 |

위 사진은 두 선수의 Feature를 비교한 사진이다. 골수가 11개와 8개, 도움 숫자가 2개와 5개 , 키퍼스 숫자가 1.6개와 1.4개 등 다양한 피쳐 부분에서 유사함을 보이고 있다.

Rating 이외에도 선수의 퍼포먼스를 heuristics 하게 알 수 있는 데이터가 또 한가지 존재한다. 바로 선수의 몸값 이다. 몸값은 선수가 리그에서 좋은 퍼포먼스를 보여주면 가파르게 상승하고 좋지 않은 모습을 보여주면 가파르게 하락한다.



위는 Tarik Tissoudali 선수의 몸값 변화 그래프다. 해당 데이터의 Season이 2020-2021 이므로 , 위쪽 그래프에서 빨간 박스쪽을 확인해야 한다. 선수가 축구를 시작한 2015년 부터 2020년도 사이의 몸값 상승분에 비해 특출난 선수와 비슷한 퍼포먼스를 보여준 2020-2021년

에는 몸값이 가파르게 상승한 것을 확인 할 수 있다.

2번 시나리오는 [팀의 주전급 선수가 이탈하게 되었다고 가정하고 , 선수를 대신할 다른 선수 찾기] 다. 이 경우 Test 데이터에 있는 선수가 이탈하였다고 가정하고 , 모델에 입력해 결과를 확인한다.

| | | | |
|------|-----------|---------------|-----------------------------------|
| 1881 | 2020-2021 | PremierLeague | Pierre-Emerick Aubameyang Arsenal |
| 1917 | 2020-2021 | LeagueOne | Colby Bishop Accrington |
| 1918 | 2019 | USA | Jozy Altidore Toronto |
| 1974 | 2019-2020 | PremierLeague | Sergio Agüero Man City |
| 1977 | 2019-2020 | PremierLeague | Neal Maupay Brighton |
| 1986 | 2019-2020 | SuperLig | Milan Skoda Rizespor |
| 2008 | 2020-2021 | Bundesliga | Max Kruse Union Berlin |
| 2051 | 2019 | USA | Adama Diomande Los Angeles FC |
| 2065 | 2019-2020 | LeagueOne | Colby Bishop Accrington |
| 2130 | 2020-2021 | ItalySerieA | Edin Dzeko Roma |

Test 데이터들중 임의로 선택한 선수는 영국 프리미어리그 에서 뛰고 있는 Neal Maupay 라는 선수다. 이 선수를 기준 삼아 $K = 50 \sim 300$ 인 경우에 대하여 이 선수와 함께 속해있는 Cluster를 살펴보도록 한다.

살펴보고 싶은 클러스터 번호 입력 : 34

k = 50

살펴보는 클러스터 번호 : 34

| season | league | name W |
|--------|-----------|--|
| 117 | 2020-2021 | RussiaPremier Vladimir Iljin FK Akhmat |
| 125 | 2020-2021 | Ligue1 Steve Mounié Brest |
| 208 | 2020-2021 | LeagueTwo Alfie May Cheltenham |
| 318 | 2020-2021 | Championship James Collins Luton |
| 397 | 2020-2021 | RussiaPremier Fedor Smolov Lokomotiv Moscow |
| 509 | 2019-2020 | PremierLeague Jamie Vardy Leicester |
| 608 | 2019-2020 | LeagueOne Daryl Murphy Bolton |
| 658 | 2020-2021 | JupilerProLeague Smail Previjak Eupen |
| 660 | 2020 | Brazil Gabi Flamengo |
| 768 | 2019-2020 | Bundesliga2 Pascal Testroet Erzgebirge Aue |
| 811 | 2020-2021 | Ligue1 Terem Moffi Lorient |
| 819 | 2019-2020 | Bundesliga Erling Haaland Borussia Dortmund |
| 905 | 2020-2021 | LaLiga Roger Marti Levante |
| 999 | 2020-2021 | PremierLeague Che Adams Southampton |
| 1016 | 2020-2021 | Championship Michael Smith Rotherham |
| 1055 | 2019 | USA Brian Fernández Portland |
| 1086 | 2019-2020 | LeagueOne Mohamed Eisa Peterborough |
| 1125 | 2019 | Brazil Wellington Paulista Fortaleza |
| 1194 | 2020-2021 | LeagueOne Jayden Stockley Charlton |
| 1273 | 2020 | Brazil Rafael Moura Goias |
| 1284 | 2019 | USA Gyasi Zardes Columbus |
| 1312 | 2020-2021 | PremierLeague Patrick Bamford Leeds |
| 1340 | 2020-2021 | LeagueTwo Ian Henderson Salford City |
| 1411 | 2019-2020 | Championship Lewis Grabban Nottingham Forest |
| 1459 | 2020-2021 | Championship Steven Fletcher Stoke |
| 1551 | 2020-2021 | Bundesliga2 Guido Burgstaller St. Pauli |
| 1670 | 2019 | Brazil Eduardo Sasha Santos FC |
| 1672 | 2020-2021 | Bundesliga2 Andreas Albers Jahn Regensburg |
| 1742 | 2020-2021 | Bundesliga2 Janni Serra Holstein Kiel |
| 1769 | 2020-2021 | LeagueTwo Conor McAleny Oldham |
| 1829 | 2020-2021 | Bundesliga2 Marvin Ducksch Hannover |
| 1841 | 2019-2020 | LigaProfesional Silvio Romero Independiente |
| 1857 | 2019-2020 | LeagueTwo Ryan Bowman Exeter |

살펴보고 싶은 클러스터 번호 입력 : 45

k = 100

살펴보는 클러스터 번호 : 45

| season | league | name W |
|--------|-----------|---|
| 79 | 2019-2020 | LigaProfesional Nahuel Bustos Talleres |
| 101 | 2019-2020 | Bundesliga2 Emmanuel Iyoha Holstein Kiel |
| 117 | 2020-2021 | RussiaPremier Vladimir Iljin FK Akhmat |
| 327 | 2019-2020 | LigaNOS Bruno Duarte Vitoria de Guimaraes |
| 658 | 2020-2021 | JupilerProLeague Smail Previjak Eupen |
| 660 | 2020 | Brazil Gabi Flamengo |
| 811 | 2020-2021 | Ligue1 Terem Moffi Lorient |
| 819 | 2019-2020 | Bundesliga Erling Haaland Borussia Dortmund |
| 872 | 2020-2021 | LigaProfesional Sebastián Sosa Patronato de Parana |
| 886 | 2019-2020 | Bundesliga2 Marcos Álvarez Osnabrueck |
| 1086 | 2019-2020 | LeagueOne Mohamed Eisa Peterborough |
| 1092 | 2020-2021 | Bundesliga2 Chris Führich Paderborn |
| 1240 | 2019-2020 | LeagueOne Marcus Forss AFC Wimbledon |
| 1257 | 2020-2021 | LigaProfesional Matías Tissera Club Atletico Platense |
| 1554 | 2020-2021 | Bundesliga2 Kenan Karaman Fortuna Duesseldorf |
| 1686 | 2020-2021 | PremierLeague Alexandre Lacazette Arsenal |
| 1742 | 2020-2021 | Bundesliga2 Janni Serra Holstein Kiel |
| 2189 | 2020-2021 | Ligue1 Boulaye Dia Reims |
| 2262 | 2020-2021 | LigaProfesional Miguel Merentiel Defensa y Justicia |
| 2571 | 2019-2020 | RussiaPremierLeague Fedor Chalov CSKA Moscow |
| 2646 | 2020-2021 | PremierLeague Dominic Calvert-Lewin Everton |
| 2688 | 2020-2021 | LeagueOne Chuks Aneke Charlton |
| 2716 | 2019 | USA Raúl Ruidiaz Seattle |
| 2729 | 2020-2021 | JupilerProLeague Fashion Sakala Oostende |
| 2768 | 2019-2020 | LigaNOS Toni Martínez Famalicão |
| 2934 | 2020-2021 | LeagueOne Ellis Simms Blackpool |
| 3000 | 2019-2020 | Championship Macauley Bonne Charlton |
| 3143 | 2020-2021 | Eredivisie Lois Openda Vitesse |
| 3200 | 2020-2021 | Ligue1 Habib Diallo Strasbourg |
| 3309 | 2020-2021 | Bundesliga Sasa Kalajdzic Stuttgart |
| 3503 | 2020-2021 | LeagueOne Kane Hemmings Burton |
| 3537 | 2020-2021 | LigaNOS Samuel Lino Gil Vicente |
| 3548 | 2020-2021 | LigaProfesional Gustavo Del Prete Estudiantes |

살펴보고 싶은 클러스터 번호 입력 : 126

k = 150

살펴보는 클러스터 번호 : 126

| season | league | name | Apps | W |
|--------|-----------|---------------------|-------------------------------|--------|
| 208 | 2020-2021 | LeagueTwo | Alfie May Cheltenham | 36(8) |
| 221 | 2019-2020 | LigaNOS | Fábio Abreu Moreirense | 27(7) |
| 306 | 2019-2020 | Championship | Harry Cornick Luton | 37(8) |
| 658 | 2020-2021 | JupilerProLeague | Small Prevljak Eupen | 23(4) |
| 811 | 2020-2021 | Ligue1 | Terem Moffi Lorient | 26(6) |
| 905 | 2020-2021 | LaLiga | Roger Martí Levante | 26(7) |
| 999 | 2020-2021 | PremierLeague | Che Adams Southampton | 30(6) |
| 1416 | 2020-2021 | LeagueOne | Lucas Akins Burton | 45 |
| 1417 | 2019-2020 | LeagueTwo | Jerry Yates Swindon | 30(1) |
| 1670 | 2019 | Brazil | Eduardo Sasha Santos FC | 34(3) |
| 1742 | 2020-2021 | Bundesliga2 | Janni Serra Holstein Kiel | 25(6) |
| 1769 | 2020-2021 | LeagueTwo | Conor McAleeny Oldham | 38(3) |
| 2189 | 2020-2021 | Ligue1 | Boulaye Dia Reims | 35(1) |
| 2194 | 2020-2021 | LeagueTwo | Jon Mellish Carlisle | 41(4) |
| 2571 | 2019-2020 | RussiaPremierLeague | Fedor Chalov CSKA Moscow | 29(1) |
| 2581 | 2020-2021 | LeagueOne | Mikael Mandron Crewe | 36(6) |
| 2604 | 2020-2021 | LeagueTwo | Lewis Alessandra Carlisle | 38(8) |
| 2646 | 2020-2021 | PremierLeague | Dominic Calvert-Lewin Everton | 32(1) |
| 2768 | 2019-2020 | LigaNOS | Toni Martínez Famalicão | 27(5) |
| 2903 | 2019-2020 | PremierLeague | Teemu Pukki Norwich | 33(3) |
| 3009 | 2019-2020 | Championship | Martyn Waghorn Derby | 36(7) |
| 3037 | 2019 | USA | C.J. Sapong Chicago | 29(3) |
| 3065 | 2019-2020 | LeagueTwo | Sam Hoskins Northampton | 35(1) |
| 3143 | 2020-2021 | Eredivisie | Lois Openda Vitesse | 30(3) |
| 3200 | 2020-2021 | Ligue1 | Habib Diallo Strasbourg | 27(5) |
| 3309 | 2020-2021 | Bundesliga | Sasa Kalajdzic Stuttgart | 23(10) |
| 3314 | 2020-2021 | LaLiga | Santi Mina Celta Vigo | 29(5) |
| 3503 | 2020-2021 | LeagueOne | Kane Hemmings Burton | 26(10) |
| 3711 | 2020-2021 | LigaNOS | Mario González Tondela | 25(2) |
| 3762 | 2019-2020 | Championship | Tom Barkhuizen Preston | 35(9) |
| 3823 | 2020-2021 | LaLiga | Alexander Isak Real Sociedad | 30(4) |
| 3854 | 2020-2021 | LeagueTwo | Jevani Brown Colchester | 28(12) |
| 3876 | 2020-2021 | LaLiga | José Luis Morales Levante | 30(8) |

살펴보고 싶은 클러스터 번호 입력 : 233

k = 250

살펴보는 클러스터 번호 : 233

| season | league | name | W |
|--------|-----------|---------------------|--|
| 658 | 2020-2021 | JupilerProLeague | Small Prevljak Eupen |
| 660 | 2020 | Brazil | Gabi Flamengo |
| 999 | 2020-2021 | PremierLeague | Che Adams Southampton |
| 1086 | 2019-2020 | LeagueOne | Mohamed Eisa Peterborough |
| 1257 | 2020-2021 | LigaProfesional | Matias Tissiera Club Atletico Platense |
| 1284 | 2019 | USA | Gyasi Zardes Columbus |
| 1742 | 2020-2021 | Bundesliga2 | Janni Serra Holstein Kiel |
| 2189 | 2020-2021 | Ligue1 | Boulaye Dia Reims |
| 2571 | 2019-2020 | RussiaPremierLeague | Fedor Chalov CSKA Moscow |
| 2646 | 2020-2021 | PremierLeague | Dominic Calvert-Lewin Everton |
| 2716 | 2019 | USA | Raúl Ruidiaz Seattle |
| 3711 | 2020-2021 | LigaNOS | Mario González Tondela |
| 3823 | 2020-2021 | LaLiga | Alexander Isak Real Sociedad |
| 3924 | 2020-2021 | JupilerProLeague | Iké Ugbo Cercle Bruges |
| 4059 | 2020-2021 | Premiership | Kevin Nisbet Hibernian |
| 4079 | 2019-2020 | Bundesliga | Andrej Kramaric Hoffenheim |
| 4756 | 2019-2020 | Championship | Conor Chaplin Barnsley |
| 5397 | 2020-2021 | Eredivisie | Lennart Thy Sparta Rotterdam |
| 6257 | 2019-2020 | Bundesliga2 | Marvin Ducksch Hannover |
| 6780 | 2019-2020 | LeagueOne | Freddie Ladapo Rotherham |
| 6905 | 2020-2021 | RussiaPremier | Djorđe Despotovic Rubin Kazan |
| 7094 | 2019-2020 | Bundesliga | André Silva Eintracht Frankfurt |
| 7376 | 2020-2021 | Bundesliga2 | Florian Krüger Erzgebirge Aue |
| 7632 | 2019-2020 | Bundesliga2 | Andreas Voglsammer Arminia Bielefeld |
| 7646 | 2019-2020 | LigaNOS | Carlos Vinicius Benfica |
| 8025 | 2020-2021 | PremierLeague | Danny Ings Southampton |

살펴보고 싶은 클러스터 번호 입력 : 129

k = 200

살펴보는 클러스터 번호 : 129

| season | league | name | W |
|--------|-----------|---------------------|------------------------------------|
| 267 | 2019-2020 | LeagueOne | Alex Gilbey Milton Keynes Dons |
| 306 | 2019-2020 | Championship | Harry Cornick Luton |
| 837 | 2019-2020 | LeagueOne | Niall Ennis Doncaster |
| 949 | 2019-2020 | LeagueTwo | Frank Nouble Colchester |
| 999 | 2020-2021 | PremierLeague | Che Adams Southampton |
| 1416 | 2020-2021 | LeagueOne | Lucas Akins Burton |
| 1430 | 2020-2021 | LeagueTwo | Luke James Barrow |
| 1511 | 2020-2021 | LeagueTwo | Jamie Reid Mansfield |
| 1601 | 2019-2020 | Championship | Bobby De Cordova-Reid Fulham |
| 1854 | 2020-2021 | Championship | Mason Bennett Millwall |
| 2194 | 2020-2021 | LeagueTwo | Jon Mellish Carlisle |
| 2571 | 2019-2020 | RussiaPremierLeague | Fedor Chalov CSKA Moscow |
| 2581 | 2020-2021 | LeagueOne | Mikael Mandron Crewe |
| 2604 | 2020-2021 | LeagueTwo | Lewis Alessandra Carlisle |
| 2737 | 2020-2021 | LeagueTwo | Abo Eisa Scunthorpe |
| 3065 | 2019-2020 | LeagueTwo | Sam Hoskins Northampton |
| 3143 | 2020-2021 | Eredivisie | Lois Openda Vitesse |
| 3200 | 2020-2021 | Ligue1 | Habib Diallo Strasbourg |
| 3242 | 2020-2021 | LeagueOne | Brennan Johnson Lincoln City |
| 3854 | 2020-2021 | LeagueTwo | Jevani Brown Colchester |
| 4117 | 2020-2021 | LeagueTwo | Elliott List Stevenage |
| 4183 | 2020-2021 | JupilerProLeague | Kamal Sowah OH Leuven |
| 4714 | 2019-2020 | Championship | Jamal Lowe Wigan |
| 4756 | 2019-2020 | Championship | Conor Chaplin Barnsley |
| 5362 | 2020-2021 | LeagueOne | Fejiri Okenabirhie Doncaster |
| 5474 | 2020-2021 | LeagueTwo | Brandon Thomas-Asante Salford City |
| 5539 | 2020-2021 | LeagueTwo | Kaiyne Woolery Tranmere |
| 5887 | 2019-2020 | LaLiga | Enes Ünal Real Valladolid |
| 6027 | 2020-2021 | Eredivisie | Jørgen Strand Larsen FC Groningen |
| 6673 | 2020-2021 | Championship | Lyndon Dykes QPR |
| 6708 | 2020-2021 | LeagueOne | Will Keane Wigan |
| 6946 | 2020-2021 | Championship | João Pedro Watford |
| 7248 | 2020-2021 | Championship | Pelly-Ruddock Mpanzu Luton |

살펴보고 싶은 클러스터 번호 입력 : 191

k = 300

살펴보는 클러스터 번호 : 191

| season | league | name | Apps | W |
|--------|-----------|---------------------|-------------------------------|--------|
| 208 | 2020-2021 | LeagueTwo | Alfie May Cheltenham | 36(8) |
| 306 | 2019-2020 | Championship | Harry Cornick Luton | 37(8) |
| 905 | 2020-2021 | LaLiga | Roger Martí Levante | 26(7) |
| 999 | 2020-2021 | PremierLeague | Che Adams Southampton | 30(6) |
| 1459 | 2020-2021 | Championship | Steven Fletcher Stoke | 30(7) |
| 1511 | 2020-2021 | LeagueTwo | Jamie Reid Mansfield | 31(8) |
| 1670 | 2019 | Brazil | Eduardo Sasha Santos FC | 34(3) |
| 2189 | 2020-2021 | Ligue1 | Boulaye Dia Reims | 35(1) |
| 2571 | 2019-2020 | RussiaPremierLeague | Fedor Chalov CSKA Moscow | 29(1) |
| 2604 | 2020-2021 | LeagueTwo | Lewis Alessandra Carlisle | 38(8) |
| 2903 | 2019-2020 | PremierLeague | Teemu Pukki Norwich | 33(3) |
| 3009 | 2019-2020 | Championship | Martyn Waghorn Derby | 36(7) |
| 3200 | 2020-2021 | Ligue1 | Habib Diallo Strasbourg | 27(5) |
| 3762 | 2019-2020 | Championship | Tom Barkhuizen Preston | 35(9) |
| 3854 | 2020-2021 | LeagueTwo | Jevani Brown Colchester | 28(12) |
| 3876 | 2020-2021 | LaLiga | José Luis Morales Levante | 30(8) |
| 4117 | 2020-2021 | LeagueTwo | Elliott List Stevenage | 36(8) |
| 4756 | 2019-2020 | Championship | Conor Chaplin Barnsley | 36(8) |
| 4895 | 2020-2021 | LeagueTwo | Jack Muldoon Harrogate Town | 34(8) |
| 5362 | 2020-2021 | LeagueOne | Fejiri Okenabirhie Doncaster | 30(9) |
| 5448 | 2020-2021 | LeagueTwo | Danny Johnson Leyton Orient | 36(6) |
| 5539 | 2020-2021 | LeagueTwo | Kaiyne Woolery Tranmere | 30(10) |
| 5863 | 2019-2020 | Championship | Hal Robson-Kanu West Brom | 24(15) |
| 6238 | 2020-2021 | Championship | Nahki Wells Bristol City | 36(10) |
| 6323 | 2019-2020 | ItalySerieA | Marco Mancosu Lecce | 28(5) |
| 6558 | 2020 | Brazil | Germán Cano Vasco da Gama | 33(1) |
| 6673 | 2020-2021 | Championship | Lyndon Dykes QPR | 36(6) |
| 6762 | 2020-2021 | Championship | Josh Windass Sheff Wed | 35(6) |
| 7376 | 2020-2021 | Bundesliga2 | Florian Krüger Erzgebirge Aue | 33(1) |
| 7698 | 2020-2021 | LeagueTwo | Ryan Loft Scunthorpe | 35(6) |

K = 50 ~ 300 에 따라 각 클러스터에 속한 선수들의 수는 다음과 같다.

K = 50 : 151명 / 100 : 76명 / 150 : 63명 / 200 : 34명 / 250 : 26명 / 300 : 30명

이후 선수의 Feature 값을 비교하는 경우, 비슷한 선수를 찾아내야 하므로 선택할 수 있는 K의 최대값인 300으로 결과를 비교한다. 만약 실제 스카우팅 과정에서는 유사도가 높은 선수만을 찾아보고 싶다면 높은 K 값으로, 조금 덜 유사해도 되니 많은 선수를 찾아보고 싶다면 낮은 K 값으로 데이터 분석을 진행하면 원하는 결과를 얻을 수 있다.

살펴보고 싶은 클러스터 번호 입력 : 191

살펴보는 클러스터 번호 : 191

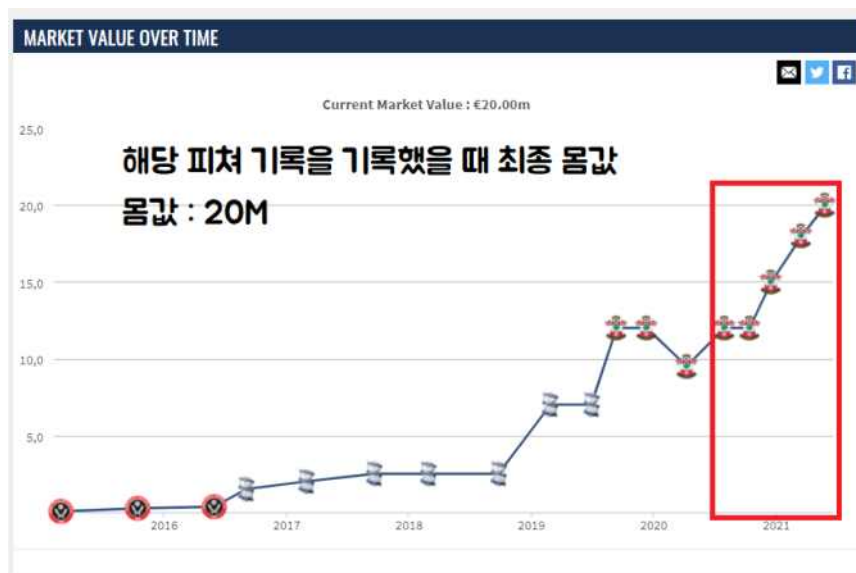
| | season | league | name | Apps | 冊 |
|------|-----------|---------------|-----------------------|-------|---|
| 208 | 2020-2021 | LeagueTwo | Alfie May Cheltenham | 36(8) | |
| 306 | 2019-2020 | Championship | Harry Cornick Luton | 37(8) | |
| 905 | 2020-2021 | LaLiga | Roger Martí Levante | 26(7) | |
| 999 | 2020-2021 | PremierLeague | Che Adams Southampton | 30(6) | |
| 1459 | 2020-2021 | Championship | Steven Fletcher Stoke | 30(7) | |
| 1511 | 2020-2021 | Championship | Lee Collins Bristol | 31(8) | |

K = 300 인 경우에 대해서 살펴본 결과, 같은 영국 프리미어리그에서 뛰고 있는 Che Adams 라는 선수와 비교한 결과이다.

| season | league | name | age | position1 | position2 | Apps | Mins | Tackles | Inter | Fouls | Offsides | Clear | Drb | Blocks | OwnG |
|-----------|------------|---------------|-----|-----------|-----------|-------|------|---------|-------|-------|----------|-------|-----|--------|------|
| 2020-2021 | PremierLea | 162 Che Adam | 25 | AM(CL) | FW | 30(6) | 2675 | 0.7 | 0.1 | 0.8 | 0.1 | 0.5 | 0.3 | 0.1 | - |
| 2019-2020 | PremierLea | 231 Neal Maup | 25 | AM(C) | FW | 30(7) | 2770 | 0.5 | 0.1 | 1 | 0.2 | 0.5 | 0.5 | - | - |

| Goals | Assists | SpG | KeyP | Drb | Fouled | Off | Disp | UnsTch | AvgP | PS% | Crosses | LongB | ThrB | Rating |
|-------|---------|-----|------|-----|--------|-----|------|--------|------|------|---------|-------|------|--------|
| 9 | 5 | 1.5 | 1 | 0.8 | 1.5 | 0.5 | 1.2 | 2.4 | 16.9 | 70.3 | - | 0.7 | 0.2 | 6.77 |
| 10 | 3 | 2.6 | 0.7 | 0.6 | 1.2 | 0.4 | 1.5 | 2.1 | 18.1 | 79.6 | 0.1 | 0.4 | - | 6.66 |

비교 해본 결과, 골 수는 9개와 10개, 도움수는 5개와 3개, 인터셉트 수는 0.1개 등, 다양한 부분에서 유사함을 보여주고 있음을 확인할 수 있다. 해당 퍼포먼스 Feature 들은 축구 전문가들이 해당 선수의 몸값을 결정하는 요소가 된다. 두 선수간 Feature값이 유사 하므로 몸값 역시 비슷할 것이라고 예측할 수 있다. 이를 검증하기 위해 실제로 두 선수의 몸값을 비교한다.



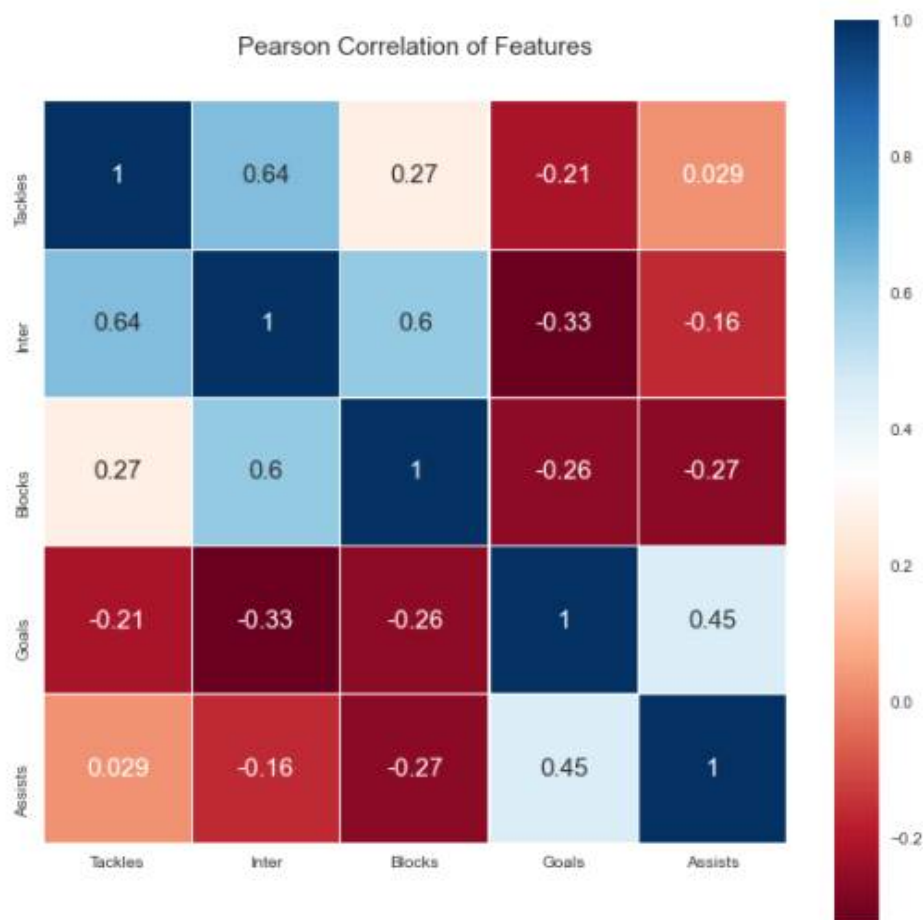


첫번째 그래프가 Che Adams 선수의 몸값 그래프, 두번째 그래프가 Neal Maupay 선수의 몸값 그래프다. 각 그래프 별로 해당 데이터의 년도수가 빨간색 박스로 표시되어 있다. 주목해야할 부분은 빨간색 박스의 마지막 부분이다. 위에서 비교한 두 데이터 모두 해당 시즌이 끝날때의 데이터 임으로 해당 시즌이 끝나는 시점의 몸값인 빨간색 박스가 끝나는 지점을 살펴보아야 정확한 몸값 비교가 가능하다. 그래프를 통해서 몸값을 비교해본 결과 두 선수 모두 20M 유로로 같은 것을 확인할 수 있다. 만약 두 선수가 뛰고 있는 리그가 다르다면 리그의 수준과 순위를 고려해야해 정확한 몸값 비교가 어려웠겠지만, 두 선수 모두 영국 프리미어리그에서 뛰고 있기 때문에 퍼포먼스 자체만으로 몸값을 비교할 수 있다.

실제 축구선수 스카우팅 시나리오에 모델을 적용해보았다. 이후 비 데이터 전문가들에게 이러한 사실을 전달할 필요가 있으므로, 시각화 작업 역시 진행한다. 시각화 작업의 경우 Feature의 개수가 많을 수록 어려워지므로, 축구 데이터에서 가장 대표적인 Feature로 알려져 있는 Feature 5개 (Goal, Assist, Tackle, Inter, Blcok)를 사용하여 먼저 진행한다.

-시각화 Feature 5개 (Goal, Assist, Tackle, Inter, Blcok) 사용

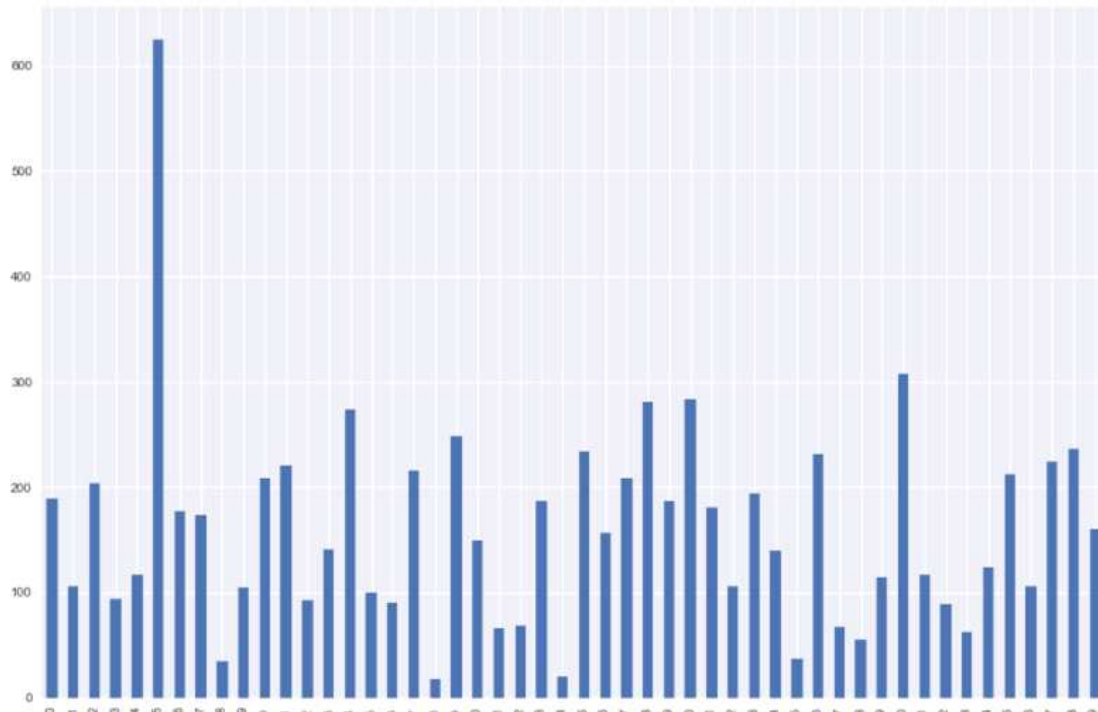
-Pearson Clorrelation



공격수 feature인 Goal과 Assist는 둘 사이에 0.45로 어느 정도 높은 상관관계를 보였고, 수비수 feature인 Tackle, Inter, Block 사이에서도 어느 정도 높은(0.64, 0.6) 상관관계를 확인할 수 있다.

이후 K=50 , 150 , 300 일때의 결과이다. K가 달라져도 결과는 비슷하나 극단적인 결과를 보기 위하여 이전 Final 발표내용에 더하여 K가 50, 300일 때를 추가하였다.

++K=50 일때 , Cluster 결과



모델과 시나리오 부분에서 outlier에 중점을 더 두었었기 때문에 클러스터에 속한 선수들이 가장 적은 클러스터 번호인 8번 클러스터(34명 선수), 18번 클러스터(17명 선수), 24번 클러스터(20명 선수)를 주의 깊게 확인하였다.

++이후 시각화 자료에서도 계속 나오지만 5번 클러스터를 보면 선수가 굉장히 많은 것을 볼 수 있다. 이 클러스터 안에 속한 선수들은 모든 feature 값이 0에 가까운 선수들이다.

살펴보는 클러스터 번호 : 18

| | Tackles | Inter | Blocks | Goals | Assists | Mins | Rating |
|------|---------|-------|--------|-------|---------|------|--------|
| 279 | 1.0 | 0.6 | 0.1 | 8 | 16 | 3377 | 7.41 |
| 688 | 0.4 | 0.1 | 0.1 | 17 | 12 | 2316 | 7.63 |
| 1130 | 0.8 | 0.4 | 0.0 | 16 | 16 | 2973 | 7.27 |
| 1671 | 0.8 | 0.5 | 0.0 | 18 | 12 | 2646 | 7.66 |
| 2653 | 1.4 | 0.4 | 0.1 | 8 | 21 | 2260 | 7.42 |
| 2777 | 0.9 | 0.4 | 0.0 | 14 | 17 | 2612 | 7.87 |
| 3053 | 1.5 | 0.6 | 0.1 | 8 | 17 | 2421 | 7.66 |
| 3763 | 1.5 | 0.7 | 0.0 | 18 | 12 | 3110 | 7.43 |
| 3970 | 0.8 | 0.5 | 0.0 | 7 | 16 | 2861 | 7.55 |
| 4026 | 0.6 | 0.2 | 0.0 | 25 | 21 | 2881 | 8.71 |
| 5906 | 0.4 | 0.4 | 0.0 | 17 | 16 | 2292 | 7.64 |
| 5971 | 1.5 | 0.6 | 0.0 | 18 | 14 | 2827 | 7.61 |
| 6011 | 0.6 | 0.2 | 0.1 | 9 | 16 | 2949 | 7.19 |
| 6646 | 1.3 | 0.5 | 0.1 | 13 | 20 | 2800 | 7.97 |
| 6937 | 0.7 | 0.3 | 0.1 | 11 | 14 | 2172 | 7.72 |
| 7550 | 1.9 | 0.4 | 0.0 | 13 | 14 | 1829 | 7.82 |
| 7571 | 0.8 | 0.4 | 0.1 | 15 | 13 | 2634 | 7.67 |

| | |
|---------|-------------|
| Tackles | 0.994118 |
| Inter | 0.423529 |
| Blocks | 0.047059 |
| Goals | 13.823529 |
| Assists | 15.705882 |
| Mins | 2644.705882 |
| Rating | 7.660000 |

살펴보는 클러스터 번호 : 24

| | Tackles | Inter | Blocks | Goals | Assists | Mins | Rating |
|------|---------|-------|--------|-------|---------|------|--------|
| 98 | 0.2 | 0.2 | 0.1 | 30 | 7 | 2610 | 8.00 |
| 165 | 0.5 | 0.1 | 0.0 | 41 | 7 | 2463 | 8.07 |
| 553 | 0.2 | 0.2 | 0.1 | 29 | 3 | 2803 | 7.61 |
| 789 | 0.6 | 0.2 | 0.0 | 34 | 4 | 2762 | 8.13 |
| 1115 | 0.5 | 0.2 | 0.0 | 30 | 9 | 3023 | 8.52 |
| 1716 | 0.4 | 0.2 | 0.0 | 28 | 5 | 3469 | 7.15 |
| 1935 | 0.7 | 0.3 | 0.0 | 34 | 10 | 2726 | 8.48 |
| 2554 | 0.1 | 0.1 | 0.0 | 27 | 7 | 2389 | 7.68 |
| 3154 | 0.6 | 0.2 | 0.1 | 26 | 4 | 3452 | 7.09 |
| 3353 | 0.4 | 0.2 | 0.1 | 27 | 3 | 2695 | 7.45 |
| 3464 | 0.4 | 0.2 | 0.1 | 31 | 5 | 2919 | 7.82 |
| 3786 | 0.3 | 0.1 | 0.1 | 27 | 6 | 2410 | 7.63 |
| 3934 | 0.5 | 0.2 | 0.2 | 31 | 3 | 3906 | 7.36 |
| 4494 | 0.6 | 0.2 | 0.1 | 31 | 10 | 3862 | 7.41 |
| 4833 | 0.3 | 0.1 | 0.0 | 28 | 5 | 2771 | 7.51 |
| 4897 | 0.3 | 0.1 | 0.1 | 27 | 3 | 2581 | 7.47 |
| 4966 | 0.7 | 0.1 | 0.1 | 29 | 3 | 2591 | 7.54 |
| 6222 | 0.6 | 0.0 | 0.1 | 24 | 8 | 2990 | 7.57 |
| 6903 | 0.2 | 0.3 | 0.1 | 36 | 9 | 3176 | 7.57 |
| 7235 | 0.6 | 0.1 | 0.0 | 25 | 8 | 2547 | 7.55 |

| | |
|---------|-----------|
| Tackles | 0.4350 |
| Inter | 0.1650 |
| Blocks | 0.0650 |
| Goals | 29.7500 |
| Assists | 5.9500 |
| Mins | 2907.2500 |
| Rating | 7.6805 |

| 살펴보는 클러스터 번호 : 8 | | | | | | | |
|------------------|-------------|-------|--------|-------|---------|------|--------|
| | Tackles | Inter | Blocks | Goals | Assists | Mins | Rating |
| 15 | 0.8 | 0.6 | 0.1 | 19 | 5 | 2887 | 7.18 |
| 508 | 0.9 | 0.3 | 0.2 | 19 | 3 | 2727 | 7.16 |
| 509 | 0.4 | 0.1 | 0.1 | 23 | 5 | 3034 | 7.22 |
| 1042 | 0.4 | 0.2 | 0.1 | 20 | 1 | 3421 | 7.56 |
| 1411 | 0.2 | 0.1 | 0.0 | 20 | 2 | 3919 | 6.76 |
| 1481 | 0.7 | 0.2 | 0.1 | 16 | 4 | 3655 | 7.14 |
| 1482 | 0.7 | 0.2 | 0.2 | 18 | 3 | 2379 | 7.46 |
| 1769 | 0.4 | 0.4 | 0.0 | 19 | 1 | 3219 | 6.82 |
| 1938 | 0.5 | 0.1 | 0.1 | 21 | 5 | 2779 | 7.20 |
| 1979 | 0.4 | 0.1 | 0.1 | 19 | 3 | 3726 | 6.87 |
| 2416 | 0.8 | 0.2 | 0.0 | 18 | 4 | 2173 | 7.35 |
| 2435 | 0.8 | 0.5 | 0.1 | 20 | 1 | 2661 | 7.35 |
| 3028 | 0.5 | 0.2 | 0.1 | 22 | 1 | 2739 | 7.17 |
| 3271 | 1.0 | 0.4 | 0.2 | 24 | 5 | 2842 | 7.91 |
| 4061 | 0.7 | 0.2 | 0.2 | 19 | 5 | 3196 | 7.14 |
| 4188 | 0.3 | 0.2 | 0.1 | 26 | 2 | 3590 | 7.24 |
| 4473 | 1.1 | 0.1 | 0.1 | 19 | 3 | 3698 | 6.87 |
| 4665 | 0.5 | 0.3 | 0.2 | 18 | 4 | 3187 | 7.05 |
| 4707 | 0.5 | 0.2 | 0.2 | 19 | 5 | 3407 | 6.90 |
| 4866 | 0.9 | 0.4 | 0.0 | 20 | 4 | 2932 | 7.06 |
| 4873 | 0.4 | 0.1 | 0.0 | 18 | 1 | 1246 | 6.95 |
| 5377 | 0.7 | 0.2 | 0.0 | 17 | 3 | 2816 | 7.37 |
| 5993 | 0.5 | 0.0 | 0.0 | 22 | 5 | 3062 | 7.07 |
| 6033 | 0.4 | 0.0 | 0.2 | 23 | 2 | 2984 | 7.30 |
| 6756 | 0.3 | 0.2 | 0.0 | 21 | 3 | 2521 | 7.16 |
| 6810 | 0.6 | 0.1 | 0.0 | 18 | 2 | 2354 | 7.09 |
| 6913 | 0.4 | 0.2 | 0.0 | 24 | 5 | 2727 | 7.36 |
| 7338 | 1.1 | 0.2 | 0.1 | 22 | 2 | 2815 | 7.12 |
| 7445 | 0.9 | 0.5 | 0.0 | 18 | 3 | 2507 | 7.19 |
| 7514 | 1.5 | 0.8 | 0.1 | 23 | 3 | 2735 | 7.49 |
| 7533 | 1.2 | 0.5 | 0.0 | 22 | 1 | 2501 | 7.25 |
| 7638 | 1.7 | 0.7 | 0.1 | 17 | 3 | 2900 | 7.29 |
| 7861 | 0.8 | 0.2 | 0.1 | 26 | 3 | 3681 | 7.35 |
| 8010 | 0.9 | 0.4 | 0.0 | 22 | 3 | 3139 | 7.03 |
| <hr/> | | | | | | | |
| Tackles | 0.702941 | | | | | | |
| Inter | 0.267647 | | | | | | |
| Blocks | 0.082353 | | | | | | |
| Goals | 20.352941 | | | | | | |
| Assists | 3.088235 | | | | | | |
| Mins | 2946.441176 | | | | | | |
| Rating | 7.189118 | | | | | | |

각각 18번 , 24번 , 8번에 대하여 살펴본 경우이다.

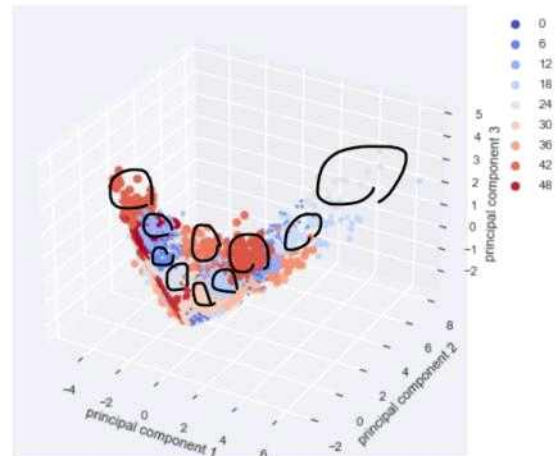
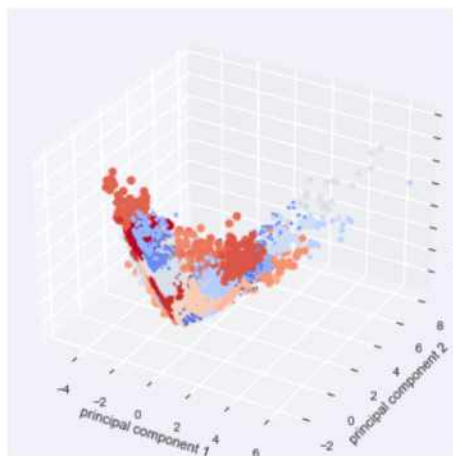
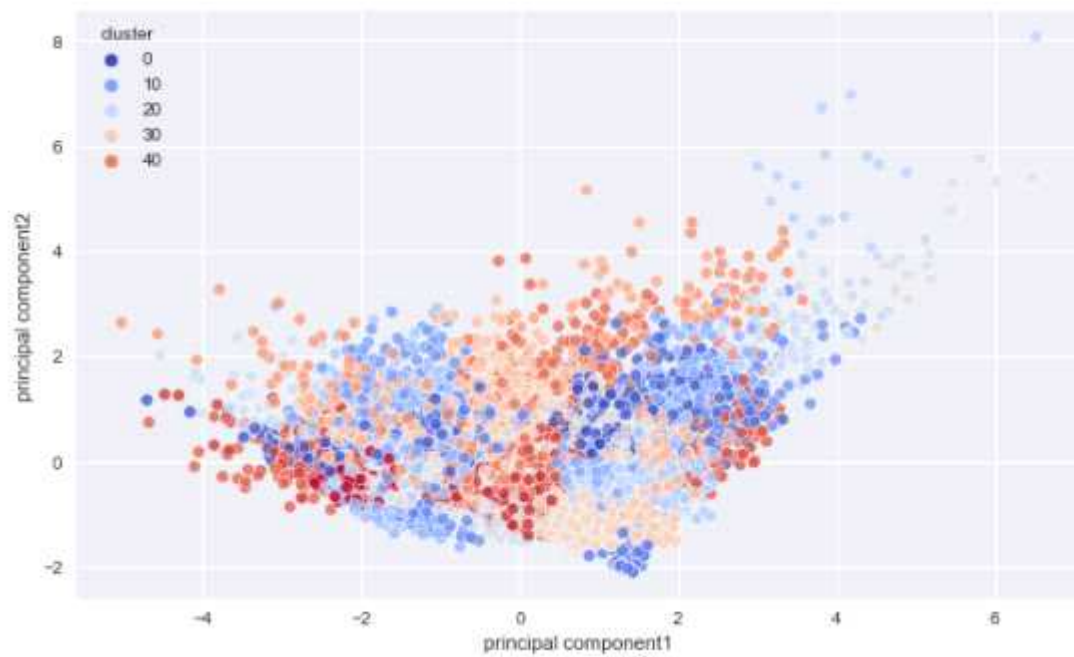
18번 클러스터를 확인해본 결과 , 18번 클러스터안에 속한 선수들은 어시스트 개수와 수비적인 능력, 특히 Tackle과 Intercept가 다른 클러스터보다 훨씬 뛰어난 것으로 확인되었다.

24번 클러스터안의 선수들은 골 개수 평균이 29.7로 다른 클러스터보다 공격적인 능력이 뛰어난 것으로 보였고 어시스트 개수 또한 적지 않은 것을 확인할 수 있다.

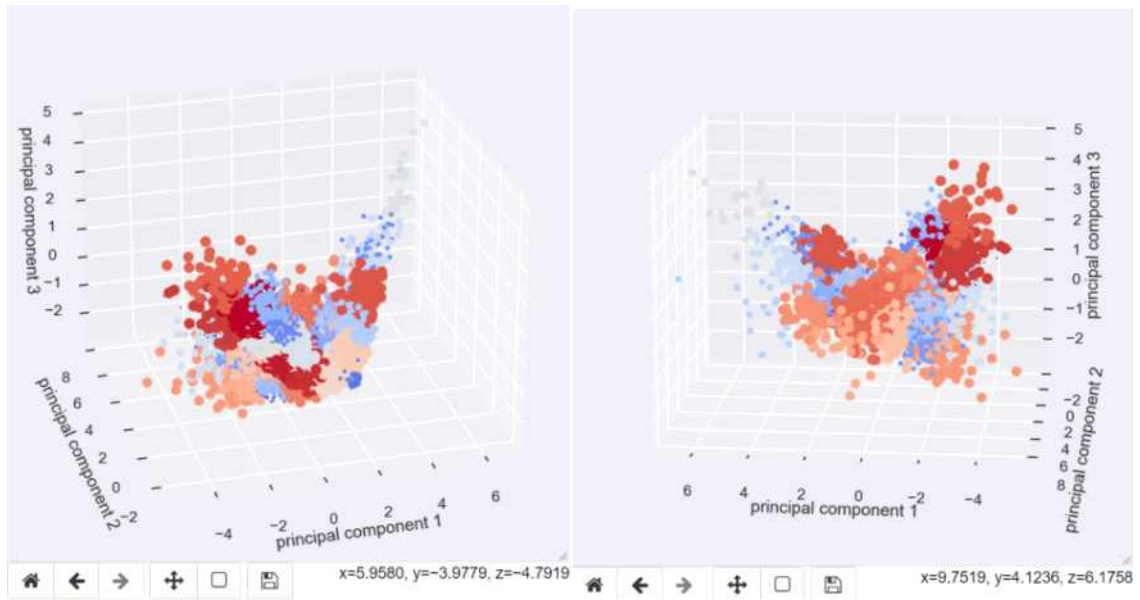
8번 클러스터의 경우, 골 개수 평균도 20.3으로 적지 않은 수치였고, 24번 클러스터보다 수비적인 능력에서 더 뛰어난 것으로 보아 8번 클러스터에 속한 선수들은 전체적인 실력이 우수한 선수들로 판단할 수 있다.

-PCA , T-sne를 이용한 시각화

1) PCA , K=50

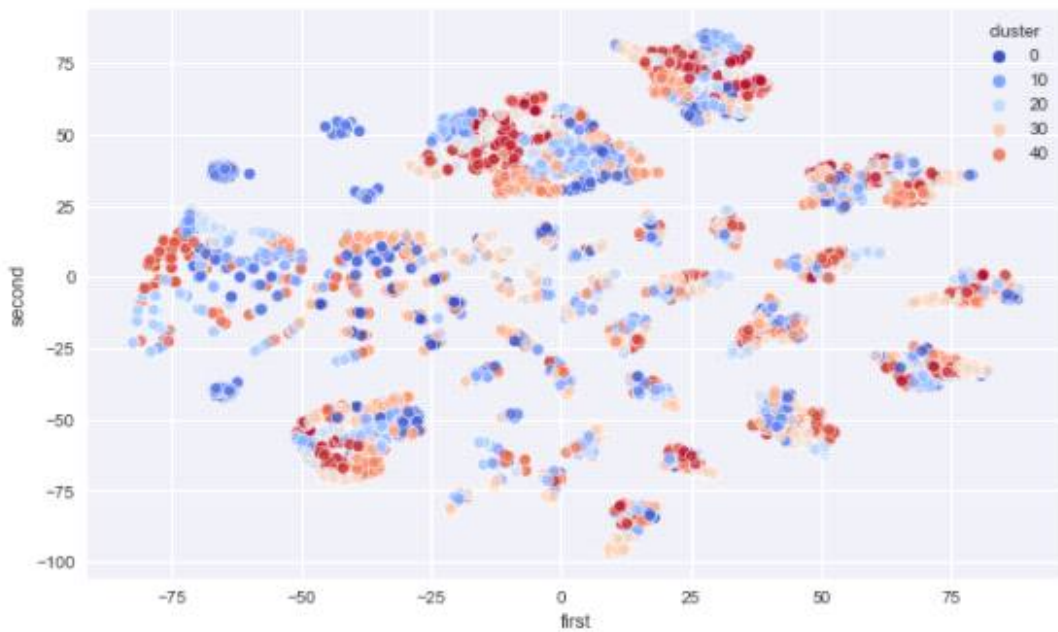


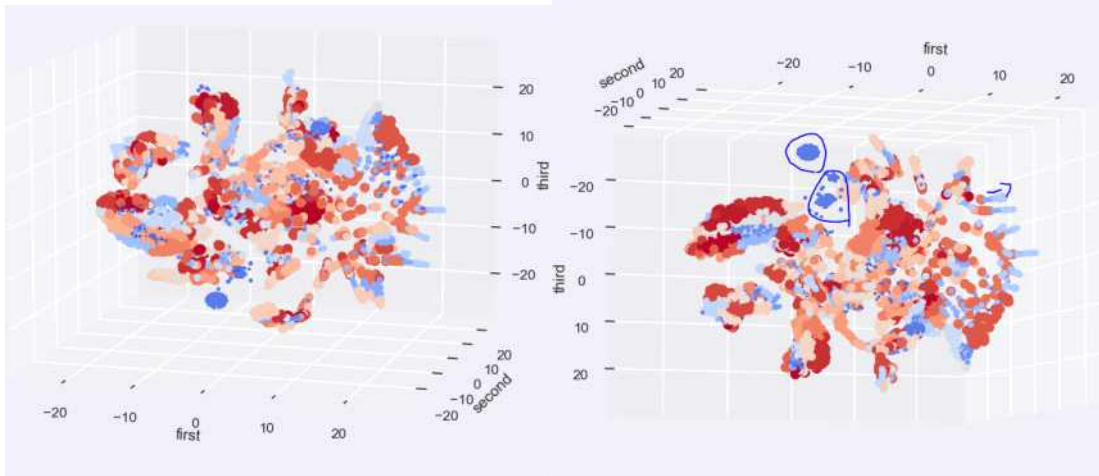
2차원으로 출력했을 때의 클러스터링 결과가 나쁘지 않아보인다. 더 자세히 보기위해 3차원으로도 출력해보았고, 출력된 해당 3차원 그림의 결과를 돌려가면서 안쪽도 확인하였다.



각각의 시각화 결과를 비교해본 결과 3차원으로 출력했을 때 , 결과를 보기 더 좋았다.

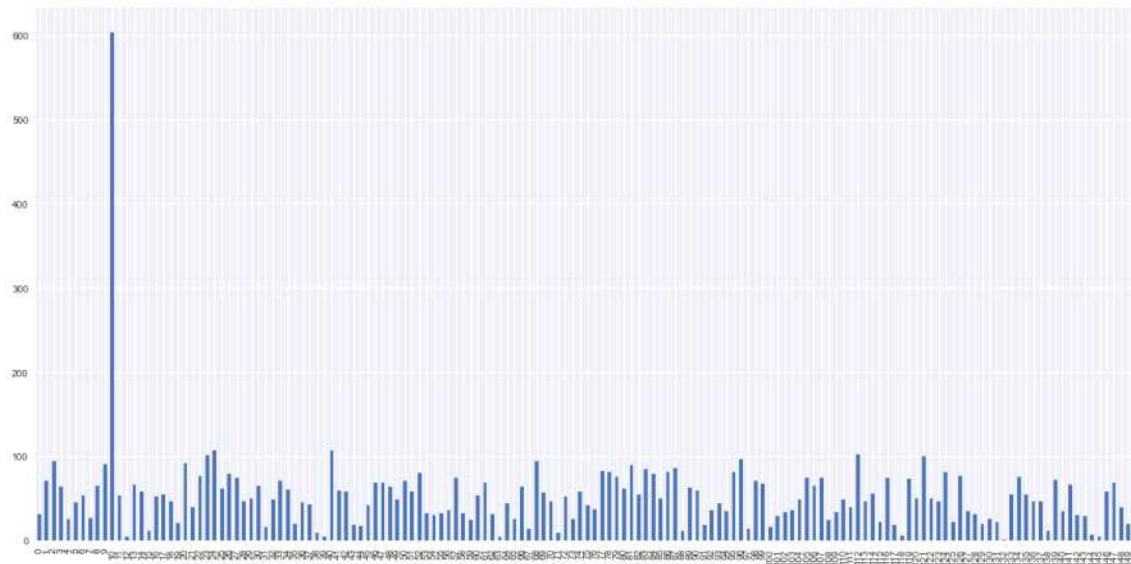
2) T-sne , K=50





이 경우 ,2차원에서는 별다른 결과를 얻을 수 없어서 3차원으로 출력했다. 3차원도 그림만으로는 해석하기 어려웠지만, 그나마 3차원 그림 중 오른쪽 사진을 보면 동그라미 친 2부분이 특히 모여있는 부분이 존재하는 것을 확인할 수 있고 , 같은 클러스터로 모여있었다는 점과 같은 사진에서 화살표를 하나 그려놨는데 화살표를 따라갈수록 색깔이 바뀌는 것으로 보아 중심에서 멀어질수록 같은 클러스터에 속한다고 판단할 수 있었다.

++K=150 Cluster 결과



이번 K=150인 경우에는 ,클러스터에 속한 선수들이 가장 적은 클러스터 번호인 132번 클러스터(1명 선수), 12번 클러스터(4명 선수), 39번 클러스터(4명 선수)를, 145번 클러스터(5명 선수), 63번 클러스터(5명 선수) 주의 깊게 살펴보았다.

| | | | | | | | | | | | | | | | |
|-------------------------|----------|--------|-------|---------|------|--------|------|--------------------------|-----------|--------|-------|---------|------|--------|------|
| 살펴보는 클러스터 번호 : 132 | | | | | | | | 살펴보고 싶은 클러스터 번호 입력 : 39 | | | | | | | |
| Tackles | Inter | Blocks | Goals | Assists | Mins | Rating | | 살펴보는 클러스터 번호 : 39 | | | | | | | |
| 4026 | 0.6 | 0.2 | 0.0 | 25 | 21 | 2881 | 8.71 | Tackles | Inter | Blocks | Goals | Assists | Mins | Rating | |
| Tackles | 0.60 | | | | | | | 1130 | 0.8 | 0.4 | 0.0 | 16 | 16 | 2973 | 7.27 |
| Inter | 0.20 | | | | | | | 2777 | 0.9 | 0.4 | 0.0 | 14 | 17 | 2612 | 7.87 |
| Blocks | 0.00 | | | | | | | 5906 | 0.4 | 0.4 | 0.0 | 17 | 16 | 2292 | 7.64 |
| Goals | 25.00 | | | | | | | 6646 | 1.3 | 0.5 | 0.1 | 13 | 20 | 2800 | 7.97 |
| Assists | 21.00 | | | | | | | Tackles | 0.8500 | | | | | | |
| Mins | 2881.00 | | | | | | | Inter | 0.4250 | | | | | | |
| Rating | 8.71 | | | | | | | Blocks | 0.0250 | | | | | | |
| dtype: float64 | | | | | | | | Goals | 15.0000 | | | | | | |
| | | | | | | | | Assists | 17.2500 | | | | | | |
| | | | | | | | | Mins | 2669.2500 | | | | | | |
| | | | | | | | | Rating | 7.6875 | | | | | | |
| | | | | | | | | dtype: float64 | | | | | | | |
| 살펴보고 싶은 클러스터 번호 입력 : 12 | | | | | | | | 살펴보고 싶은 클러스터 번호 입력 : 145 | | | | | | | |
| 살펴보는 클러스터 번호 : 12 | | | | | | | | 살펴보는 클러스터 번호 : 145 | | | | | | | |
| Tackles | Inter | Blocks | Goals | Assists | Mins | Rating | | Tackles | Inter | Blocks | Goals | Assists | Mins | Rating | |
| 98 | 0.2 | 0.2 | 0.1 | 30 | 7 | 2610 | 8.00 | 185 | 0.5 | 0.1 | 0.0 | 41 | 7 | 2463 | 8.07 |
| 2554 | 0.1 | 0.1 | 0.0 | 27 | 7 | 2389 | 7.68 | 1115 | 0.5 | 0.2 | 0.0 | 30 | 9 | 3023 | 8.52 |
| 3786 | 0.3 | 0.1 | 0.1 | 27 | 6 | 2410 | 7.63 | 1935 | 0.7 | 0.3 | 0.0 | 34 | 10 | 2726 | 8.48 |
| 7235 | 0.6 | 0.1 | 0.0 | 25 | 8 | 2547 | 7.55 | 4494 | 0.6 | 0.2 | 0.1 | 31 | 10 | 3862 | 7.41 |
| Tackles | 0.300 | | | | | | | 6903 | 0.2 | 0.3 | 0.1 | 36 | 9 | 3176 | 7.57 |
| Inter | 0.125 | | | | | | | Tackles | 0.50 | | | | | | |
| Blocks | 0.050 | | | | | | | Inter | 0.22 | | | | | | |
| Goals | 27.250 | | | | | | | Blocks | 0.04 | | | | | | |
| Assists | 7.000 | | | | | | | Goals | 34.40 | | | | | | |
| Mins | 2489.000 | | | | | | | Assists | 9.00 | | | | | | |
| Rating | 7.715 | | | | | | | Mins | 3050.00 | | | | | | |
| dtype: float64 | | | | | | | | Rating | 8.01 | | | | | | |
| | | | | | | | | dtype: float64 | | | | | | | |
| 살펴보는 클러스터 번호 : 63 | | | | | | | | | | | | | | | |
| Tackles | Inter | Blocks | Goals | Assists | Mins | Rating | | | | | | | | | |
| 553 | 0.2 | 0.2 | 0.1 | 29 | 3 | 2803 | 7.61 | | | | | | | | |
| 789 | 0.6 | 0.2 | 0.0 | 34 | 4 | 2762 | 8.13 | | | | | | | | |
| 3464 | 0.4 | 0.2 | 0.1 | 31 | 5 | 2919 | 7.82 | | | | | | | | |
| 3934 | 0.5 | 0.2 | 0.2 | 31 | 3 | 3906 | 7.36 | | | | | | | | |
| 4966 | 0.7 | 0.1 | 0.1 | 29 | 3 | 2591 | 7.54 | | | | | | | | |
| Tackles | 0.480 | | | | | | | | | | | | | | |
| Inter | 0.180 | | | | | | | | | | | | | | |
| Blocks | 0.100 | | | | | | | | | | | | | | |
| Goals | 30.800 | | | | | | | | | | | | | | |
| Assists | 3.600 | | | | | | | | | | | | | | |
| Mins | 2996.200 | | | | | | | | | | | | | | |
| Rating | 7.692 | | | | | | | | | | | | | | |
| dtype: float64 | | | | | | | | | | | | | | | |

132번 , 12번 , 39번 , 63번 클러스터에 대하여 살펴본 경우는 다음과 같다.

132번 클러스터는 선수가 한 명이긴 하지만 골 개수와 어시스트 개수가 다른 선수들에 비해 높은 것으로 보여진다.

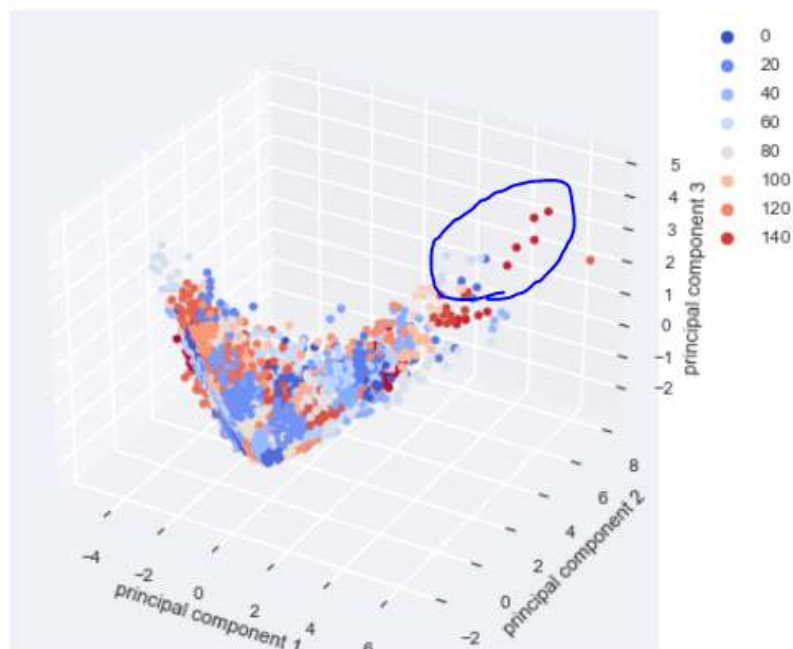
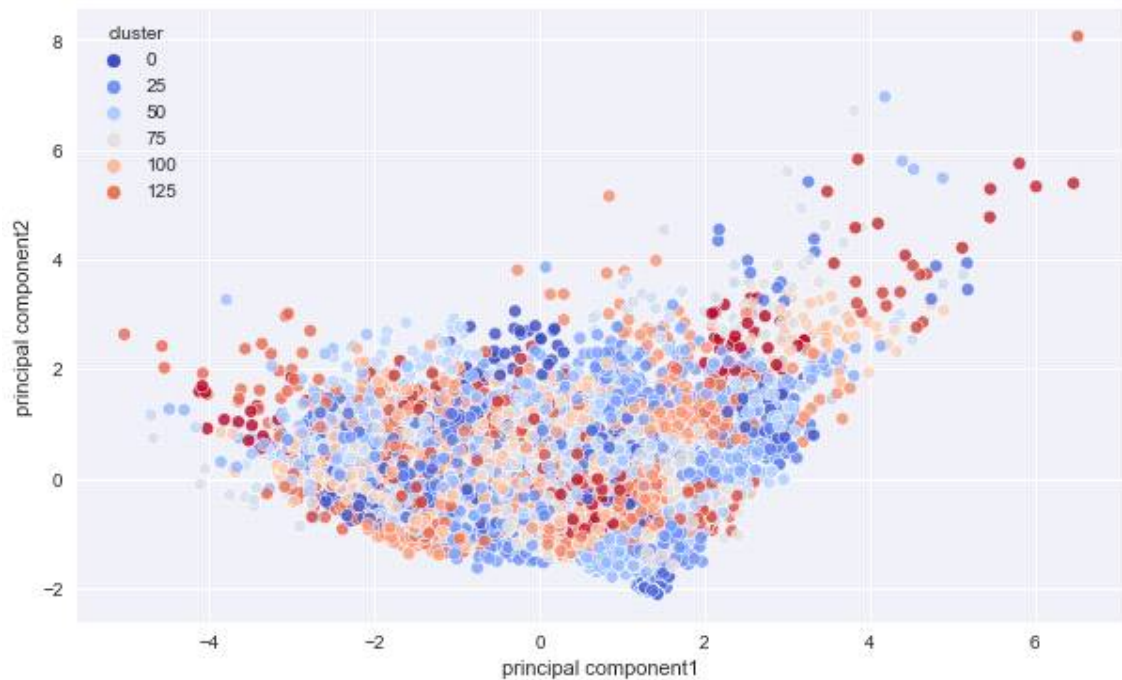
12번 클러스터에 속한 선수는 어시스트 개수에 비해 골 개수가 훨씬 많다. 이는 145번 클러스터도 마찬가지지만 145번 클러스터의 선수들은 수비 능력이 12번 클러스터에 속한 선수보다 약간 높은 것을 확인할 수 있다.

39번 클러스터에 속한 선수는 수비 능력이 다른 클러스터보다 뛰어났고 골과 어시스트 개수도 적지 않으므로 모든 부분에서 활약하는 선수로 보여진다.

63번 클러스터는 145번 클러스터와 비슷하지만 , 높은 골 개수에 비해 어시스트 개수가 적은 것을 확인할 수 있다.

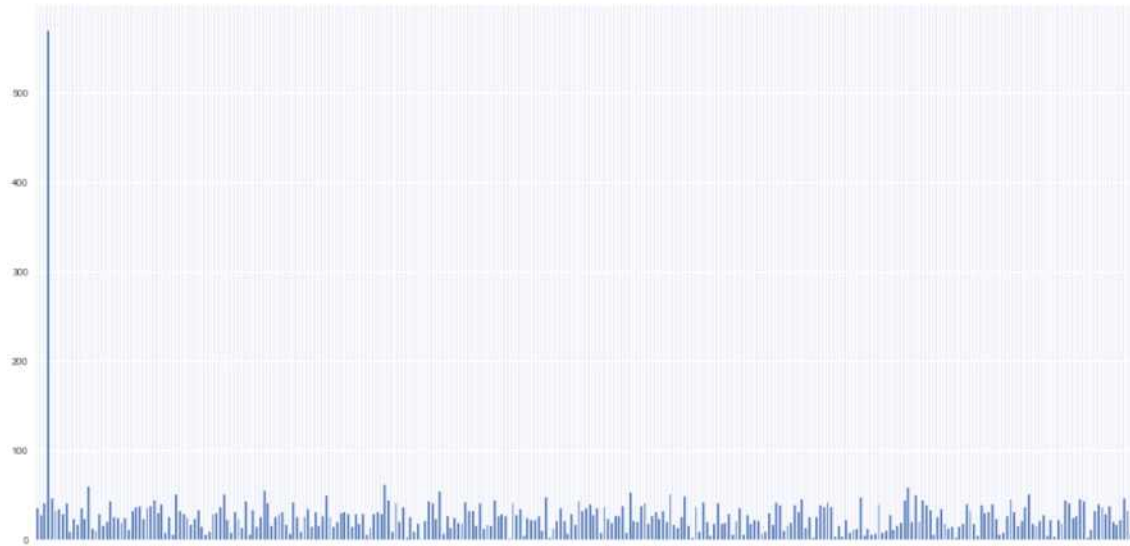
-PCA, T-sne를 이용한 시각화

1) PCA, K=150 (T-sne는 K=50 일때의 결과와 매우 유사한 결과를 도출하였다.)



동그라미 친 부분이 k가 50일 때는 모두 같은 색이었지만 k가 증가하니 다른 색이 추가된 것을 확인할 수 있다.

++K=300 Cluster 결과



이번에도 위에서 살펴보았던것과 같이 소수의 선수들이 속한 클러스터를 확인한다. 따라서 클러스터에 속한 선수들이 가장 적은 클러스터 번호인 105번 클러스터(1명 선수), 129번 클러스터(1명 선수), 251번 클러스터(2명 선수), 287번 클러스터(2명 선수), 212번 클러스터(2명 선수)를 주의 깊게 살펴본다.

```

살펴보는 클러스터 번호 : 105
      Tackles  Inter  Blocks  Goals  Assists  Mins  Rating
185         0.5    0.1    0.0     41         7  2463    8.07
-----
Tackles         0.50
Inter           0.10
Blocks          0.00
Goals           41.00
Assists         7.00
Mins            2463.00
Rating          8.07
dtype: float64

살펴보는 클러스터 번호 : 251
      Tackles  Inter  Blocks  Goals  Assists  Mins  Rating
3415         0.7    0.8    0.8     1         5  2530    6.82
5445         0.8    0.7    0.9     2         7  4140    6.90
-----
Tackles         0.75
Inter           0.75
Blocks          0.85
Goals           1.50
Assists         6.00
Mins            3335.00
Rating          6.86
dtype: float64
-----
살펴보고 싶은 클러스터 번호 입력 : 129

살펴보는 클러스터 번호 : 129
      Tackles  Inter  Blocks  Goals  Assists  Mins  Rating
4026         0.6    0.2    0.0     25         21  2881    8.71
-----
Tackles         0.60
Inter           0.20
Blocks          0.00
Goals           25.00
Assists        21.00
Mins            2881.00
Rating          8.71
dtype: float64

살펴보는 클러스터 번호 : 287

살펴보는 클러스터 번호 : 287
      Tackles  Inter  Blocks  Goals  Assists  Mins  Rating
1236         0.6    0.8    1.7     2         1  1454    7.08
5063         1.1    1.1    1.9     0         1  2663    6.96
-----
Tackles         0.85
Inter           0.95
Blocks          1.80
Goals           1.00
Assists         1.00
Mins            2058.50
Rating          7.02
dtype: float64

```

살펴보는 클러스터 번호 : 212

| | Tackles | Inter | Blocks | Goals | Assists | Mins | Rating |
|------|---------|-------|--------|-------|---------|------|--------|
| 1668 | 2.8 | 1.3 | 0.4 | 9 | 7 | 2523 | 7.67 |
| 5274 | 2.7 | 1.4 | 0.4 | 15 | 5 | 2705 | 7.39 |

```
Tackles      2.75
Inter        1.35
Blocks        0.40
Goals        12.00
Assists       6.00
Mins        2614.00
Rating        7.53
dtype: float64
```

따라서 위의 결과를 통해 105번 , 129번 , 251번 , 212번 클러스터를 살펴본 결과는 다음과 같다.

105번 클러스터는 선수가 한 명이긴 하지만 골 개수가 월등히 뛰어난 것으로 보인다.

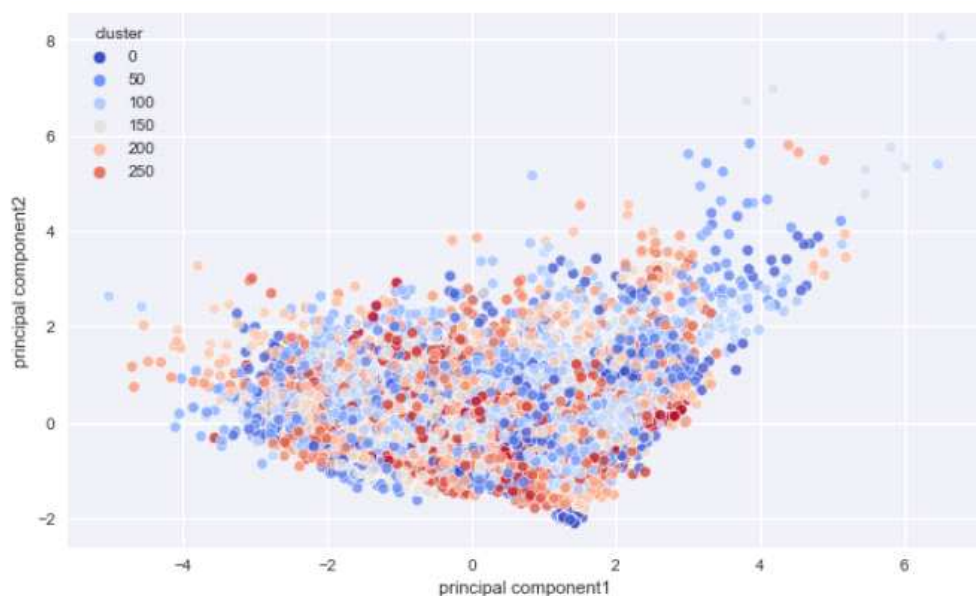
129번 클러스터에 속한 선수는 골 개수와 어시스트 개수 모두 어느정도 높은 값을 기록하였다.

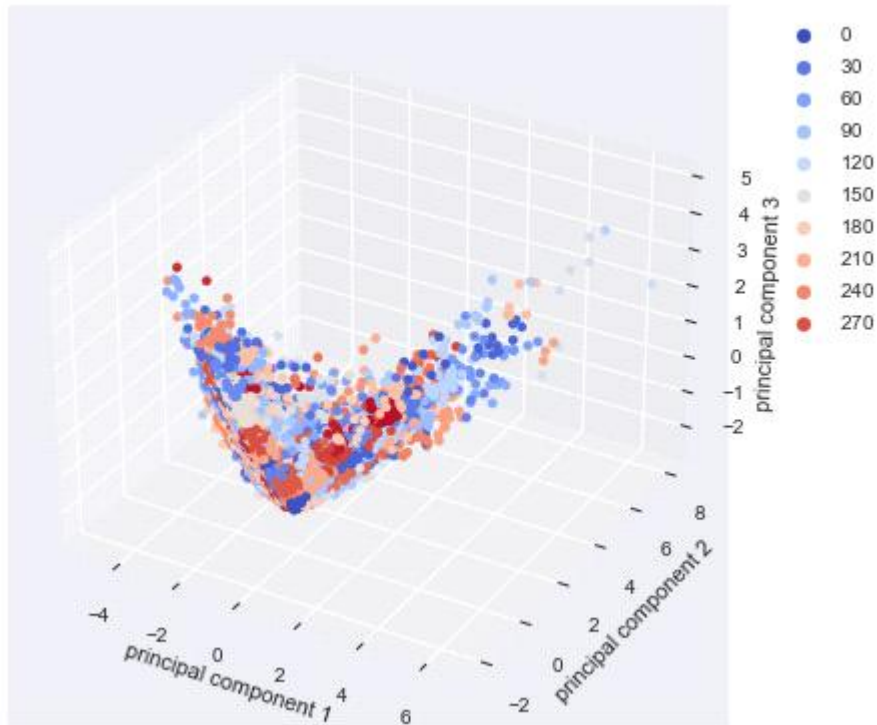
251번 클러스터에 속한 선수는 수비능력도 (287번만큼은 아니지만) 다른 클러스터보다 뛰어난 것으로 보이고 클러스터 들중 가장 유사한 287번 클러스터보다 어시스트 개수가 많다.

212번 클러스터는 골 개수가 적지 않으면서 Tackle 개수나 intercept개수가 지금까지 봤던 클러스터보다 월등히 높은 것을 볼 수 확인할 수 있다.

-PCA, T-sne를 이용한 시각화

1) PCA , K=300 (이번에도 T-sne는 K=50일때의 결과와 매우 유사한 결과를 도출하였다.)



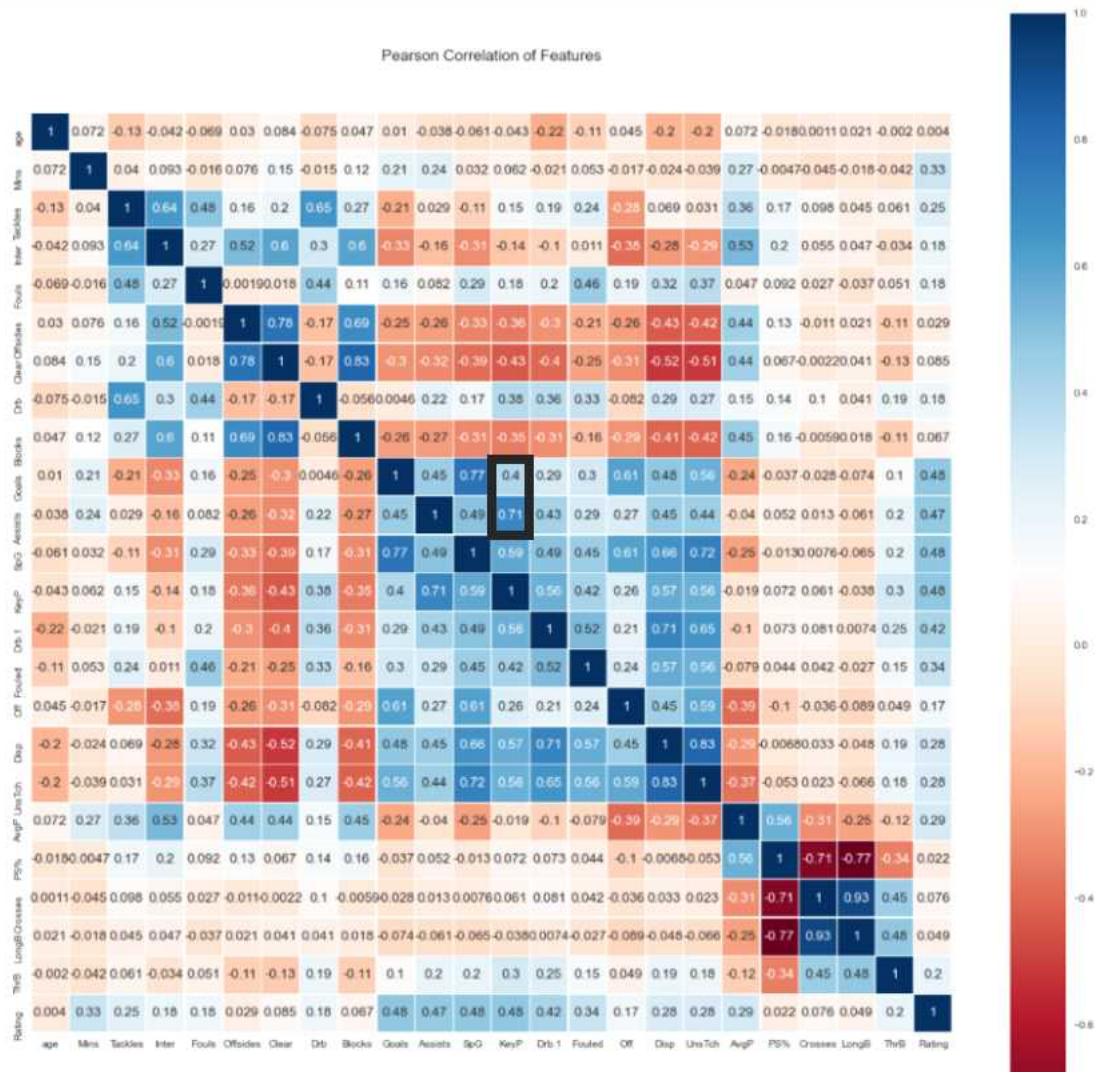


위의 결과를 앞선 결과들과 비교했을 때 , 큰 차이점이 보이지 않는 것을 확인할 수 있다.

앞선 시각화 결과들은 모두 Feature를 5개만 사용하여 확인한 결과들이다. Feature 5개를 통하여 어떤 Clsutering 결과가 등장하고 , 어떤 시각화 결과가 등장하는지 확인하였으니 Feature 24개를 모두 사용한 결과도 확인한다.

-시각화 Feature 24개 모두 사용

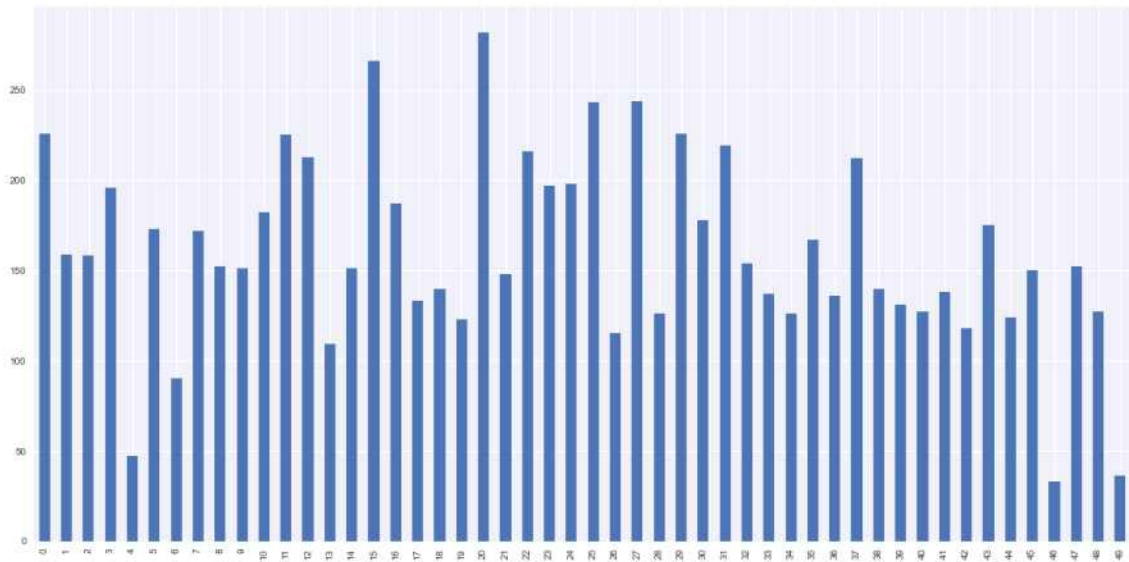
-Pearson Correlation



대체적으로 공격 능력들 사이의 상관관계가 높았고, 수비 능력 또한 사이에서도 상관관계가 높았다. 특히 KeyP라는 Feature는 해당 경기에서의 key pass(주요한 패스, 핵심적인 패스) 숫자를 말하는 것인데 골 개수와 어시스트 개수가 높을수록 key pass의 숫자가 높은 것을 확인할 수 있다. (굵은 색 네모) 이는 다시말해 골과 도움을 만들어 내는 선수들은 key pass 부분에서도 뛰어나다는 Insight를 얻을 수 있다.

++K=50 Cluster 결과

(결과가 대부분 비슷하기에 극단적인 K 값인 50, 300에 대하여 결과를 확인하였다.)



이번에도 마찬가지로 소수의 선수가 속한 클러스터를 중점적으로 확인한다. 따라서 클러스터에 속한 선수들이 가장 적은 클러스터 번호인 46번 클러스터(33명 선수), 49번 클러스터(36명 선수), 4번 클러스터(47명 선수)를 주의 깊게 확인한다. 또한 Feature를 모두 사용한 만큼 클러스터 안의 전체적인 Feature 평균값을 확인하여, 클러스터 안의 선수들에 대해 판단할 수 있게 한다.

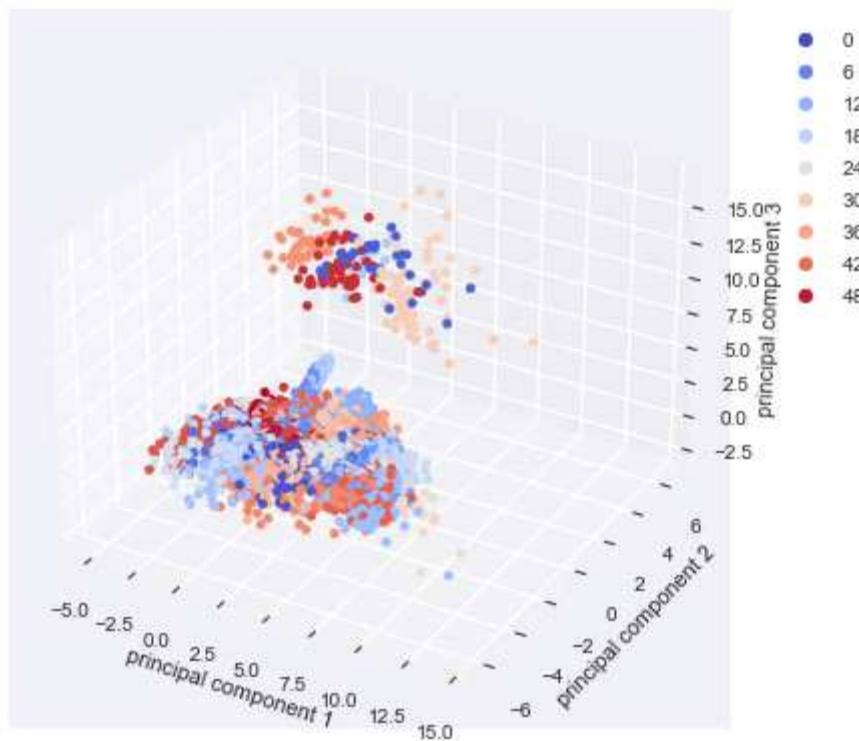
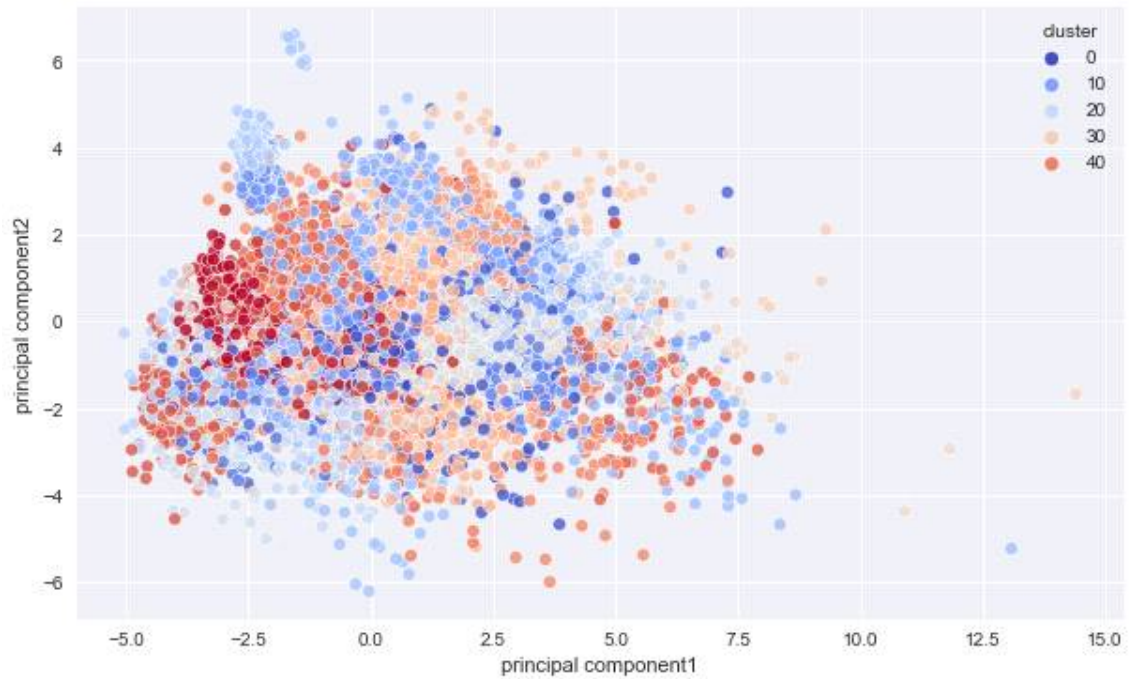
| 살펴보고 싶은 클러스터 번호 입력 : 46 | | 살펴보고 싶은 클러스터 번호 입력 : 49 | | 살펴보고 싶은 클러스터 번호 입력 : 4 | |
|-------------------------|-------------|-------------------------|-------------|------------------------|-------------|
| age | 26.363636 | age | 30.444444 | age | 27.234043 |
| Mins | 1499.878788 | Mins | 1701.250000 | Mins | 1997.808511 |
| Tackles | 1.645455 | Tackles | 1.488889 | Tackles | 3.038298 |
| Inter | 0.906061 | Inter | 1.325000 | Inter | 1.674468 |
| Fouls | 1.112121 | Fouls | 1.038889 | Fouls | 2.078723 |
| Offsides | 0.024242 | Offsides | 0.305556 | Offsides | 0.097872 |
| Clear | 0.703030 | Clear | 2.822222 | Clear | 1.472340 |
| Drb | 1.115152 | Drb | 0.708333 | Drb | 1.363830 |
| Blocks | 0.142424 | Blocks | 0.452778 | Blocks | 0.312766 |
| Goals | 1.454545 | Goals | 0.611111 | Goals | 0.851064 |
| Assists | 1.575758 | Assists | 0.583333 | Assists | 1.191489 |
| SpG | 0.869697 | SpG | 0.416667 | SpG | 0.687234 |
| KeyP | 0.845455 | KeyP | 0.302778 | KeyP | 0.627660 |
| Drb.1 | 1.118182 | Drb.1 | 0.383333 | Drb.1 | 0.787234 |
| Fouled | 1.027273 | Fouled | 0.688889 | Fouled | 1.831915 |
| Off | 0.045455 | Off | 0.036111 | Off | 0.034043 |
| Disp | 0.939394 | Disp | 0.266667 | Disp | 0.742553 |
| UnsTch | 1.375758 | UnsTch | 0.550000 | UnsTch | 1.125532 |
| AvgP | 1.575758 | AvgP | 0.583333 | AvgP | 42.431915 |
| PS% | 0.845455 | PS% | 0.302778 | PS% | 79.368085 |
| Crosses | 38.324242 | Crosses | 42.877778 | Crosses | 0.242553 |
| LongB | 84.000000 | LongB | 83.091667 | LongB | 2.870213 |
| ThrB | 0.315152 | ThrB | 0.100000 | ThrB | 0.004255 |
| Rating | 6.726364 | Rating | 6.649167 | Rating | 6.965319 |
| cluster | 46.000000 | cluster | 49.000000 | cluster | 4.000000 |

46번 클러스터와 49번 클러스터는 비슷한 feature 값을 띄어난 것으로 보였고 특히 LongB, cross 개수가 높은 것을 확인할 수 있다.

4번 클러스터는 Tackle 개수가 월등히 많은것을 확인할 수 있고, AvgP, PS% 같은 패스 능력이 뛰어난 것으로 확인되었다.

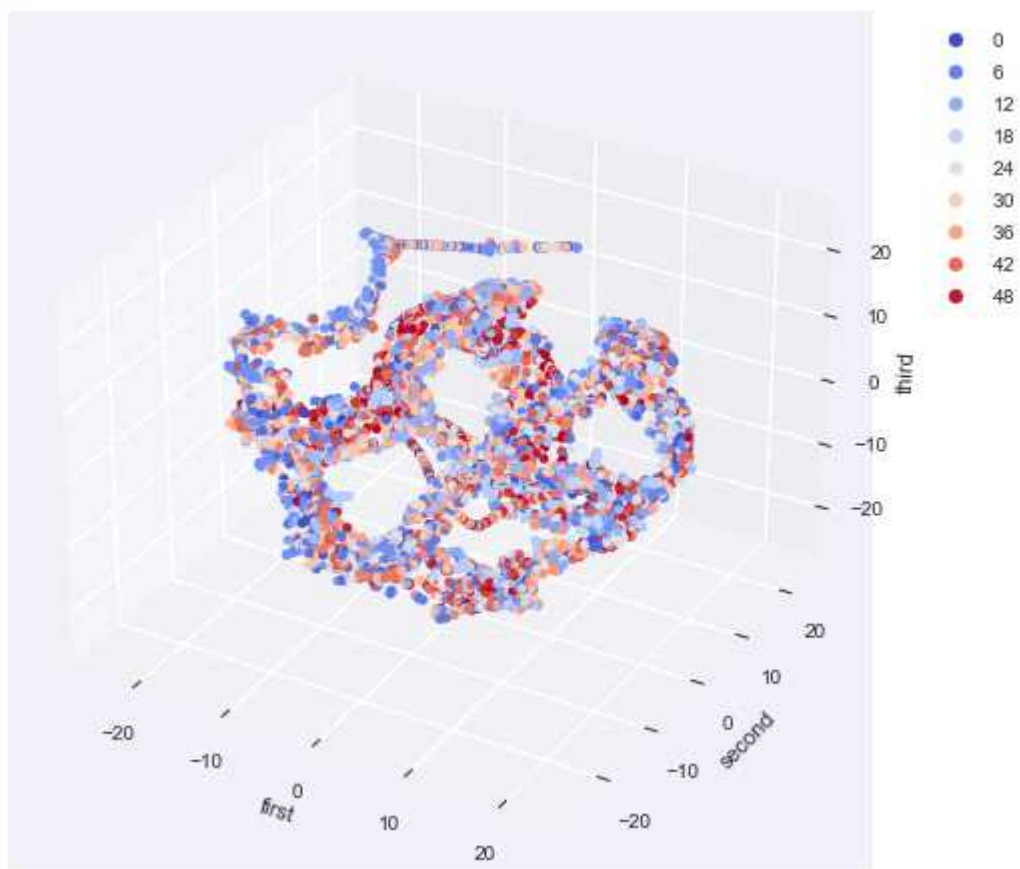
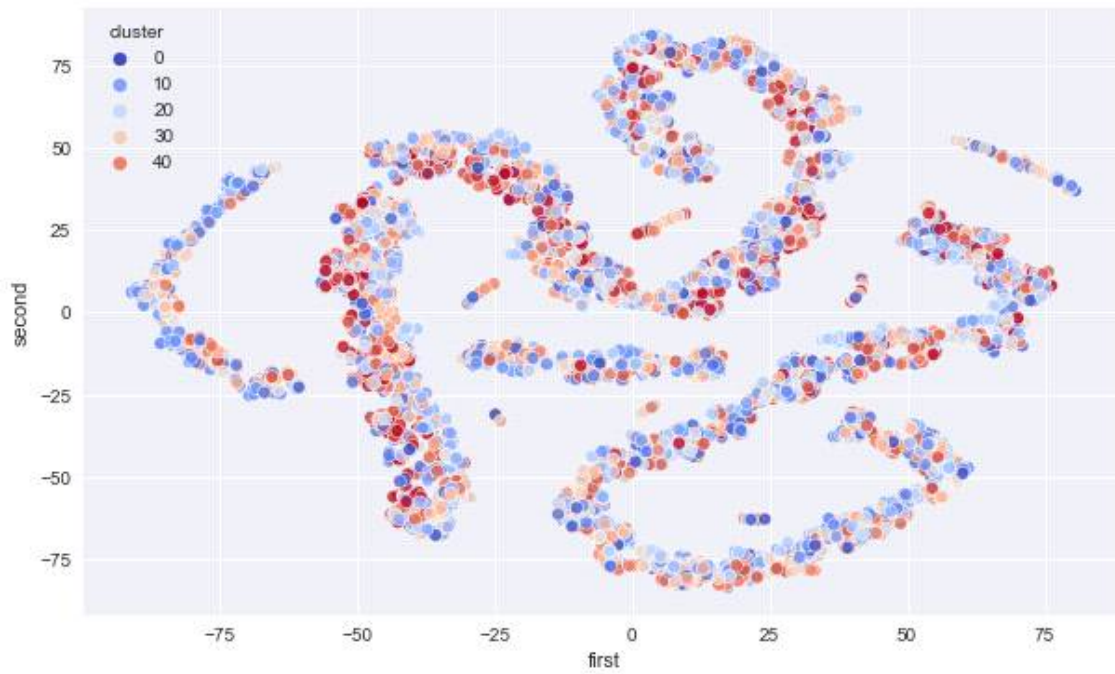
-PCA, T-sne를 이용한 시각화

1) PCA, K=50



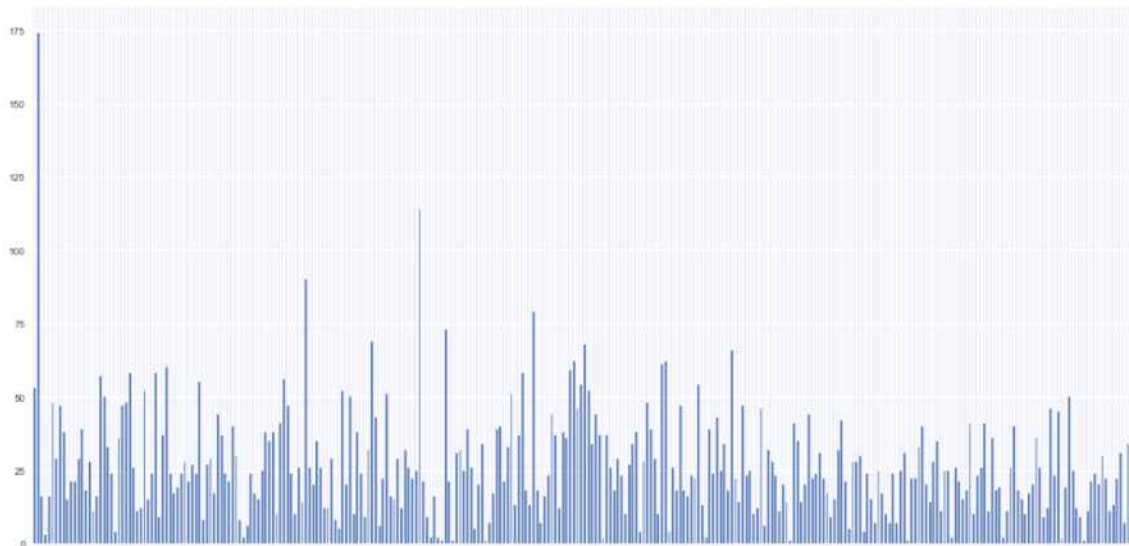
PCA 시각화에서는 어느 정도 클러스터가 모여 있는 것처럼 보이지만 별다른 insight를 얻지 못했다.

2) T-sne, K=50



이 경우에도 마찬가지로 의미있는 Insight를 얻지 못하였다.

++K=300 Cluster 결과



클러스터에 속한 선수들이 가장 적은 클러스터 번호인 206번 클러스터(1명 선수), 238번 클러스터(1명 선수), 286번 클러스터(1명 선수)를 주의 깊게 보았다.

| 살펴보고 싶은 클러스터 번호 입력 : 206 | | 살펴보고 싶은 클러스터 번호 입력 : 238 | | 살펴보고 싶은 클러스터 번호 입력 : 286 | |
|--------------------------|---------|--------------------------|---------|--------------------------|---------|
| age | 22.00 | age | 29.00 | age | 33.00 |
| Mins | 1074.00 | Mins | 1322.00 | Mins | 2005.00 |
| Tackles | 2.50 | Tackles | 0.80 | Tackles | 0.70 |
| Inter | 1.00 | Inter | 0.30 | Inter | 0.40 |
| Fouls | 1.00 | Fouls | 0.90 | Fouls | 0.50 |
| Offsides | 0.00 | Offsides | 0.00 | Offsides | 0.00 |
| Clear | 0.70 | Clear | 0.00 | Clear | 0.00 |
| Drb | 1.30 | Drb | 1.20 | Drb | 1.20 |
| Blocks | 0.10 | Blocks | 0.00 | Blocks | 0.10 |
| Goals | 0.00 | Goals | 13.00 | Goals | 8.00 |
| Assists | 0.00 | Assists | 6.00 | Assists | 14.00 |
| SpG | 0.50 | SpG | 4.70 | SpG | 2.80 |
| KeyP | 0.60 | KeyP | 2.60 | KeyP | 3.00 |
| Drb.1 | 2.20 | Drb.1 | 6.00 | Drb.1 | 2.60 |
| Fouled | 1.70 | Fouled | 4.10 | Fouled | 0.80 |
| Off | 0.00 | Off | 0.10 | Off | 0.20 |
| Disp | 1.00 | Disp | 3.10 | Disp | 1.10 |
| UnsTch | 1.10 | UnsTch | 4.90 | UnsTch | 2.00 |
| AvgP | 0.00 | AvgP | 6.00 | AvgP | 14.00 |
| PSx | 0.60 | PSx | 2.60 | PSx | 3.00 |
| Crosses | 31.90 | Crosses | 58.40 | Crosses | 41.30 |
| LongB | 81.40 | LongB | 80.60 | LongB | 78.50 |
| ThrB | 0.00 | ThrB | 0.80 | ThrB | 1.60 |
| Rating | 6.87 | Rating | 8.58 | Rating | 7.75 |
| cluster | 206.00 | cluster | 238.00 | cluster | 286.00 |

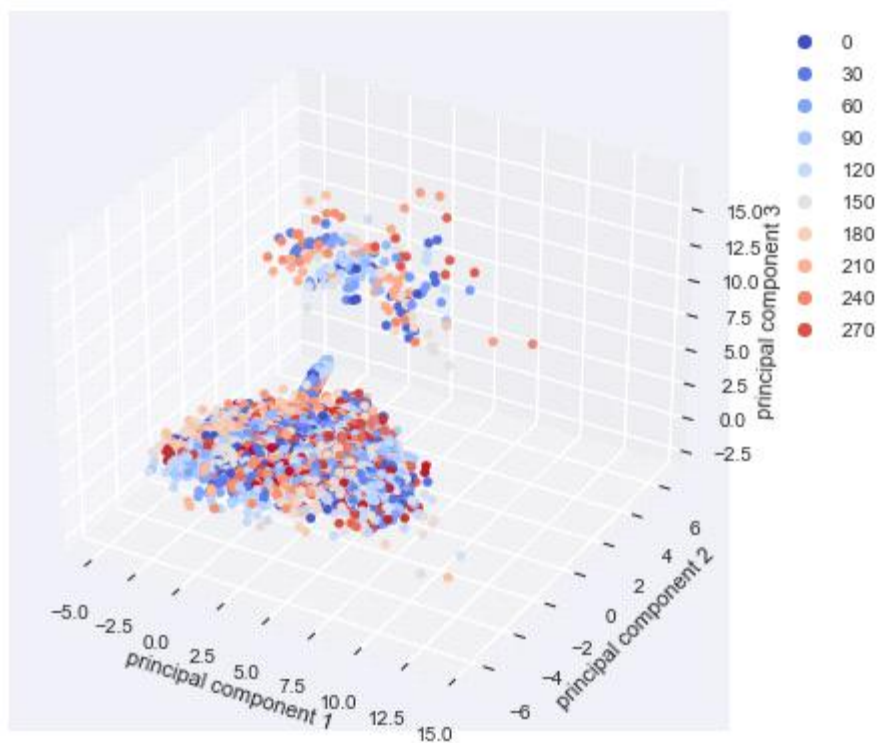
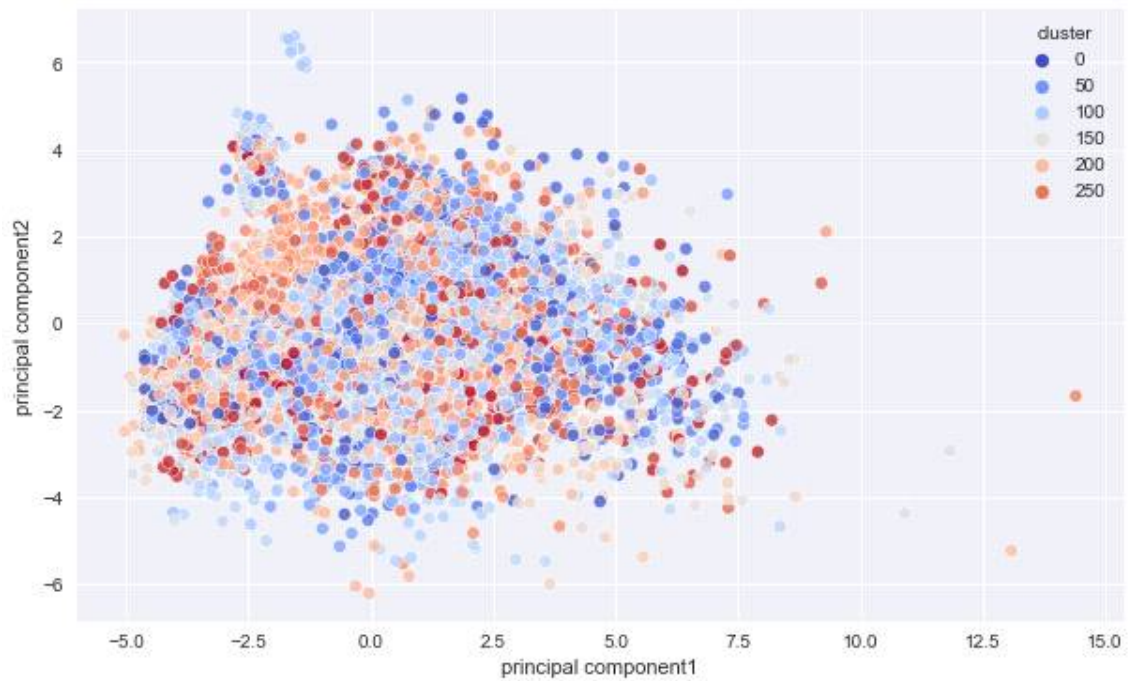
확인한 결과 , 206번 클러스터는 다른 클러스터에 비해 LongB, cross 개수가 높은 것을 확인할 수 있다.

238, 286번 클러스터에 속한 선수들은 골 능력과 수비 능력이 뛰어난 것을 확인할 수 있다.

-PCA, T-sne를 이용한 시각화

(T-sne의 경우 ,K=50의 결과와 동일한 결과를 도출하였다.)

1) PCA, K=300

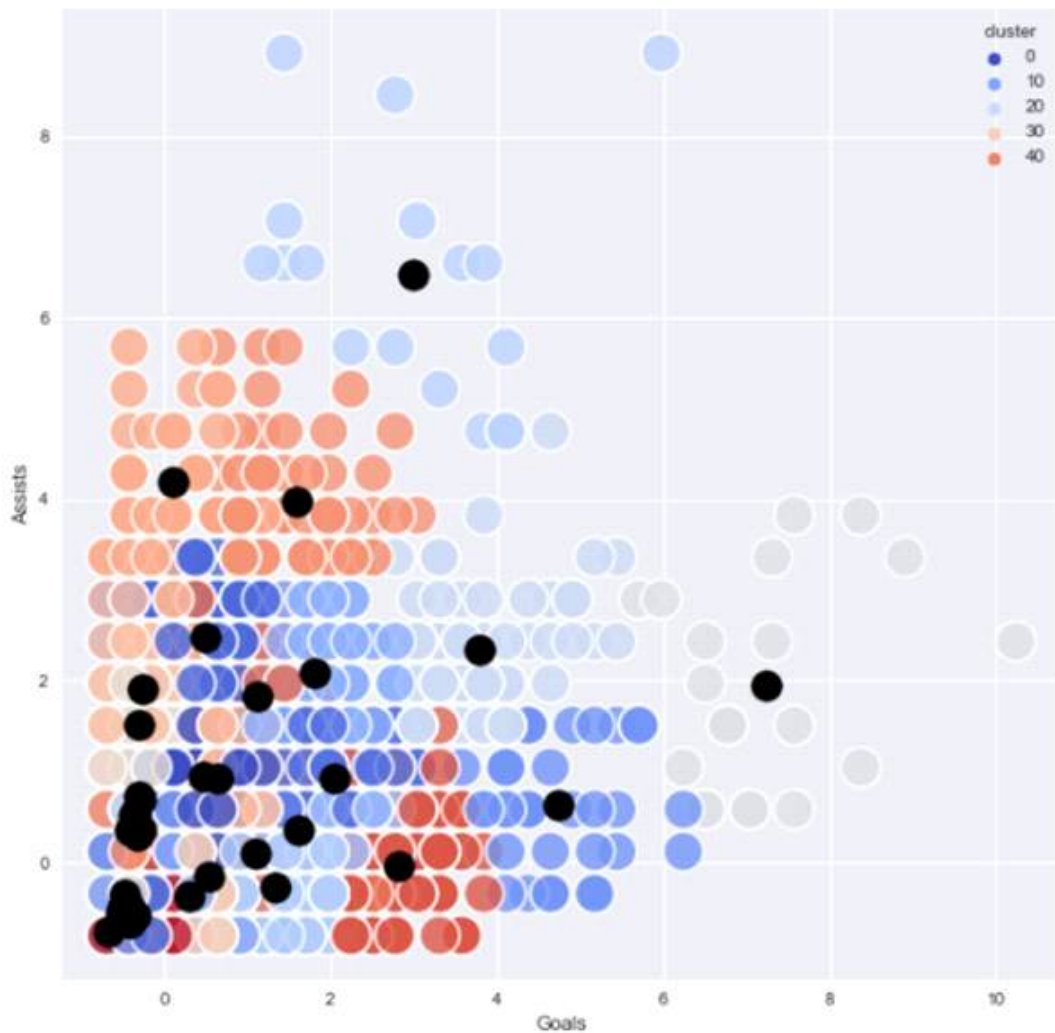


앞서 살펴보았던 결과들과 크게 다르지 않고 비슷하다.

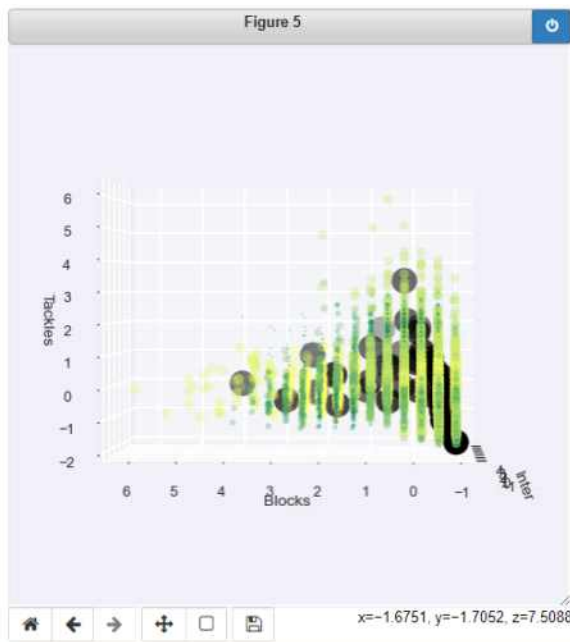
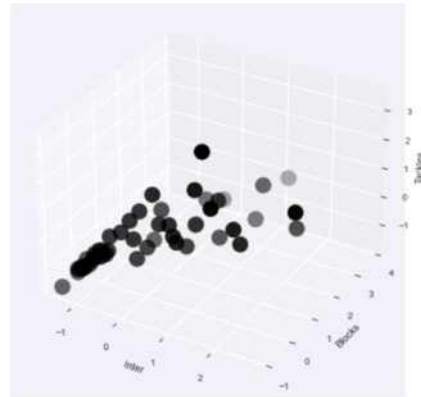
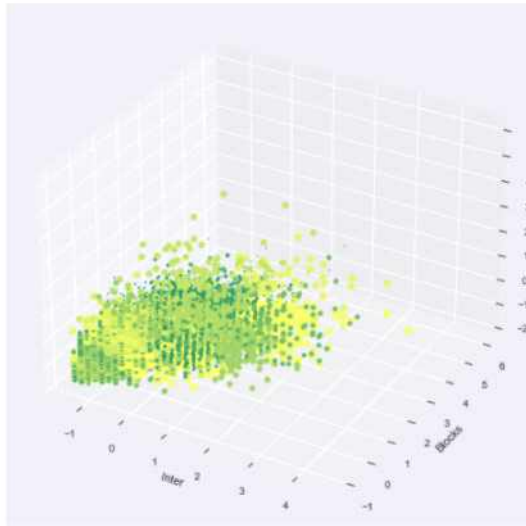
-Feature 값 시각화

일반적으로 축구 데이터들중 가장 중요하다고 알려져 있는 , 5개의 Feature 데이터들에 대해 더 자세히 시각화를 진행한다. 선형관계가 높은 공격 능력(Goal, Assist)과 수비 능력(Tackle, Intercept, Block)에 관하여 시각화를 진행하였다.

1) K=50 , Feature 5개 사용시 Feature 값

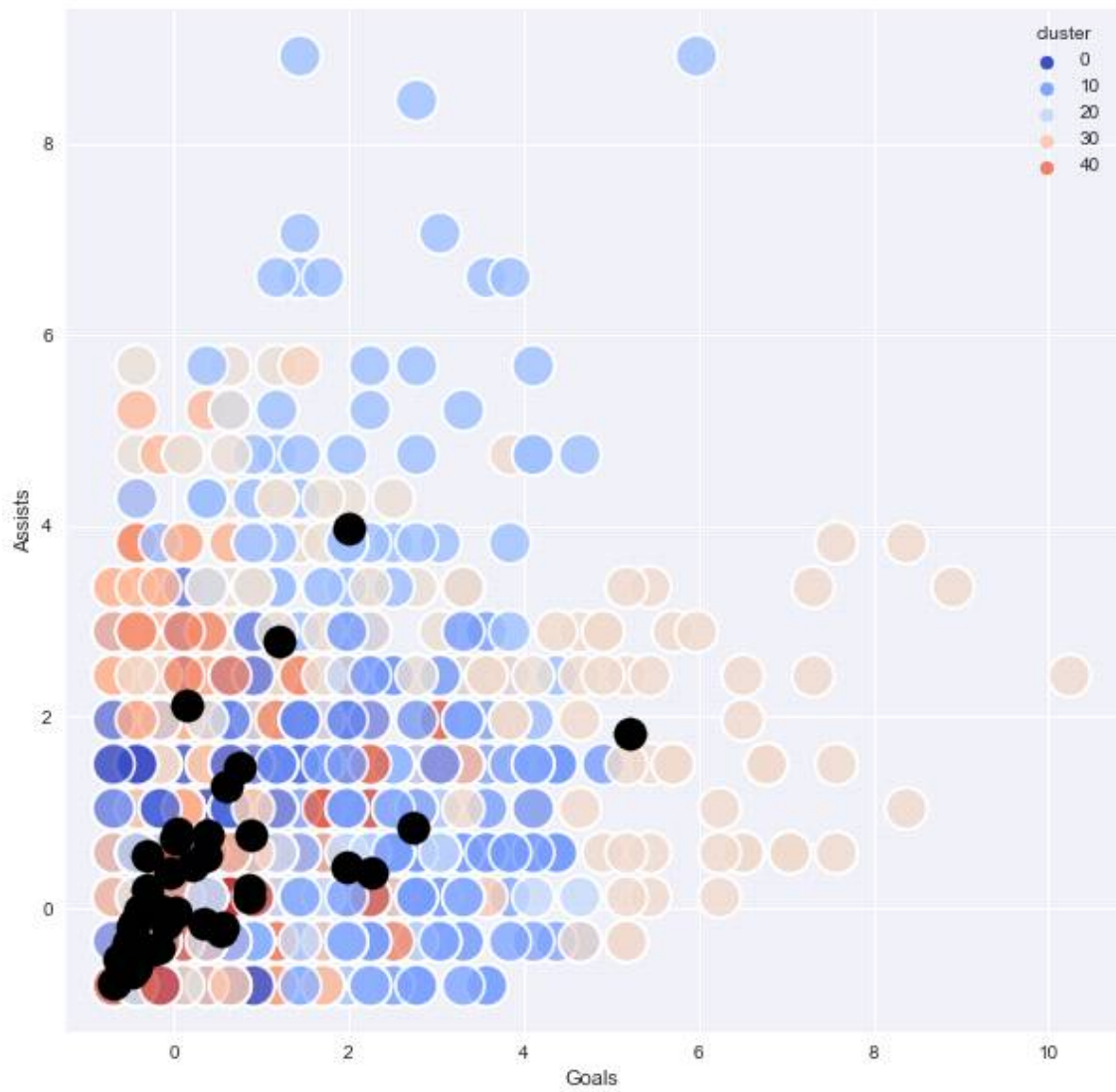


공격능력의 경우이다. 검정색 원은 클러스터 중심이고, outlier로 보이는 회색 원이나 연한 하늘색 원들이 클러스터링 잘 된 것으로 확인된다.

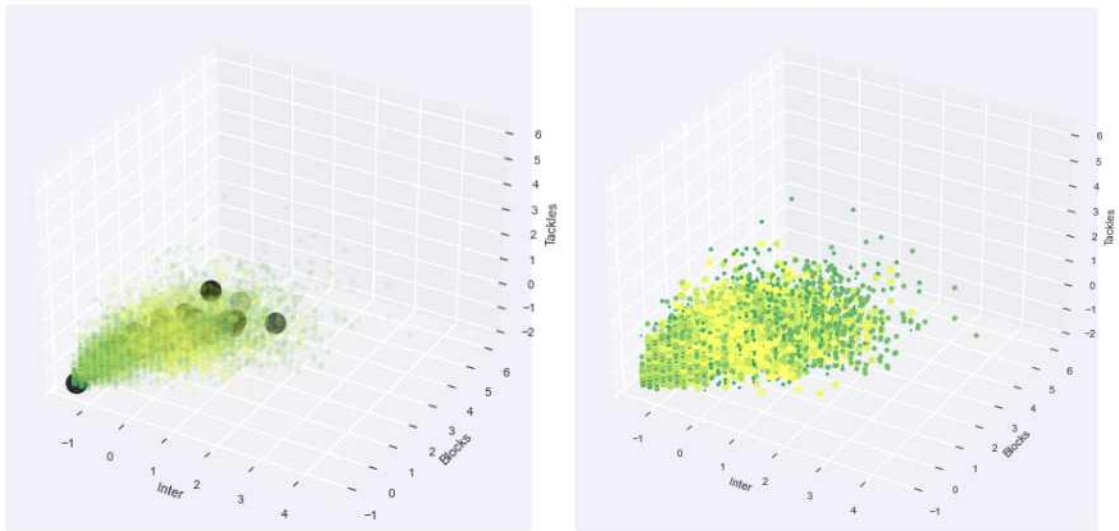


이번에는 수비 능력의 경우이다. 이 역시 마찬가지로 outlier로 보이는 데이터들끼리 잘 모인 것을 확인 할 수 있다.

2) K=50 , Feature 모두 사용시 Feature 값

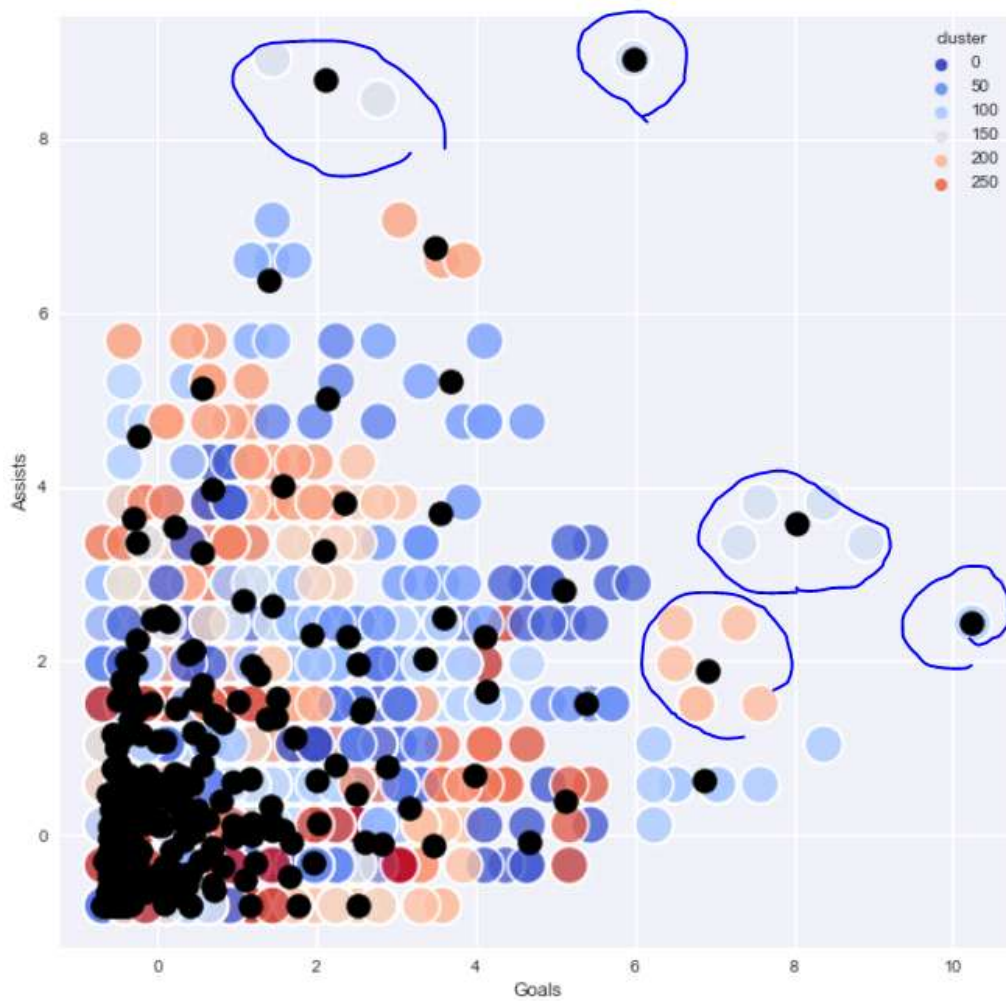


Feature를 모두 사용했을 때는 이전과 달리 더 많은 클러스터의 중심들이 0에 가까워진 것을 확인 할 수 있다.

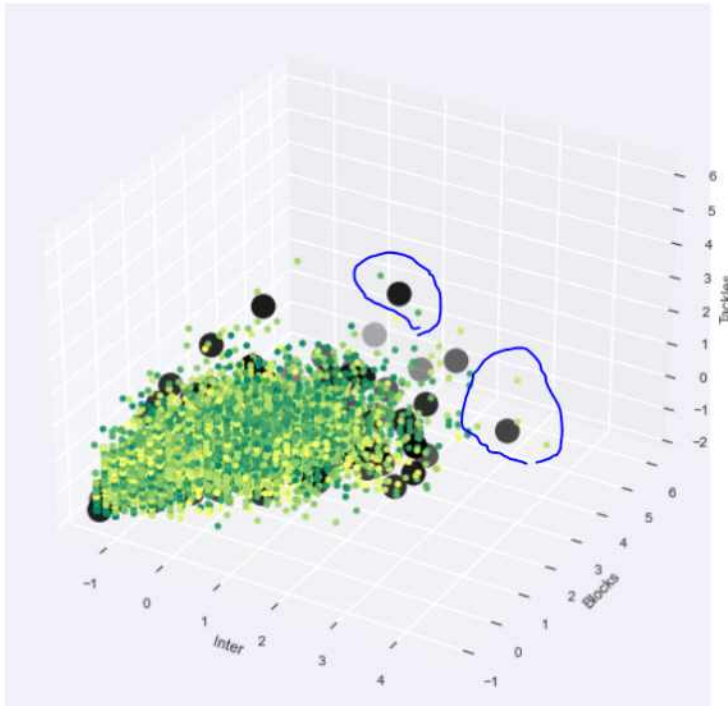


이 역시 앞선 결과들과 유사하다.

3) K=300 , Feature 5개 사용시 Feature의 값

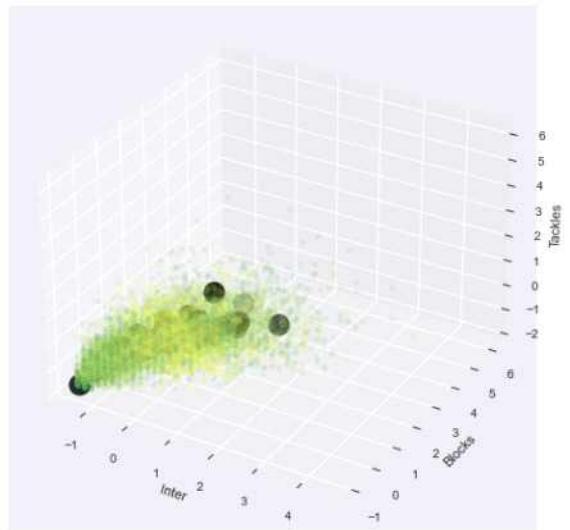
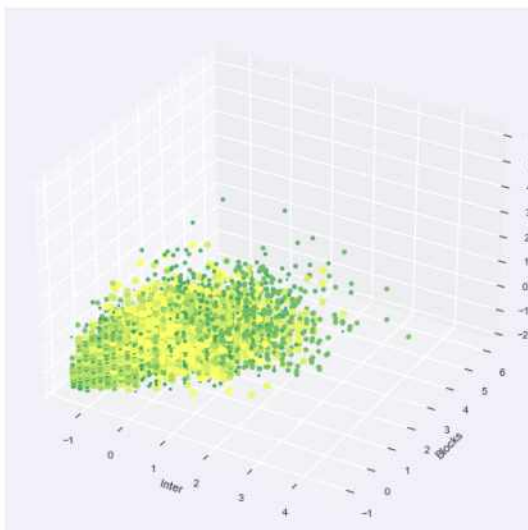
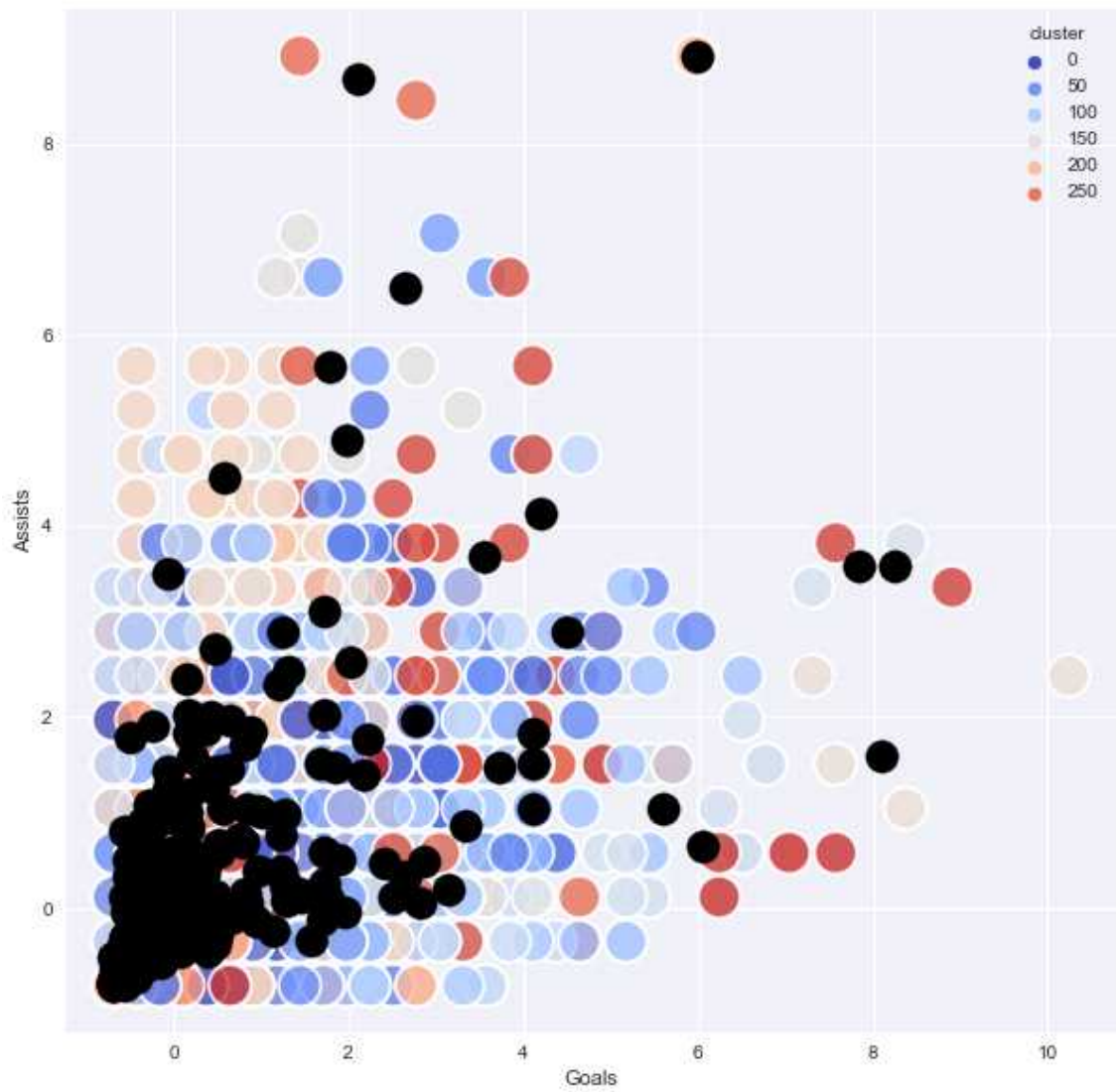


같은 클러스터 안에 포함된 선수들이 적은 선수들을 보면 outlier라고 판단할 수 있고, 이러한 클러스터들은 중심이 명확하게 보이는 것을 확인할 수 있다.



이 경우 역시 앞선 결과와 비슷하다.

4) $K=300$, Feature 모두 사용시 Feature 값

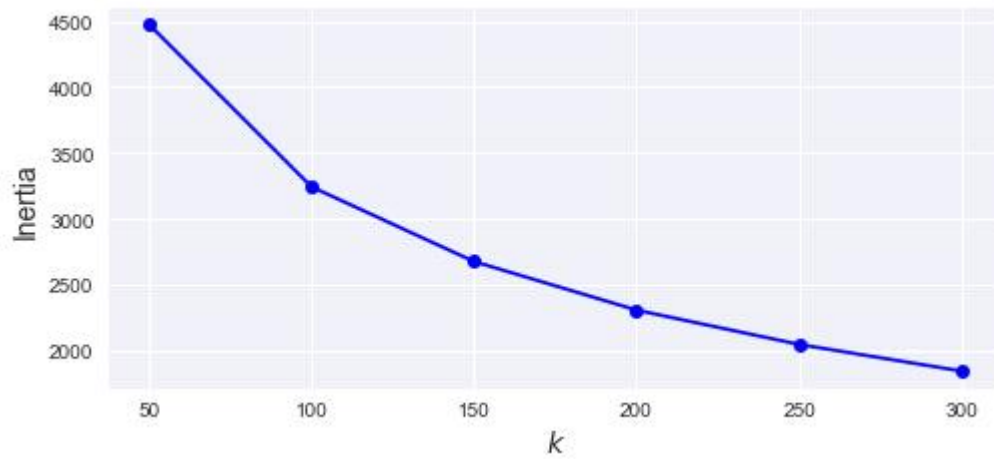


이번 결과는 K=50 일때의 결과가 비슷하다.

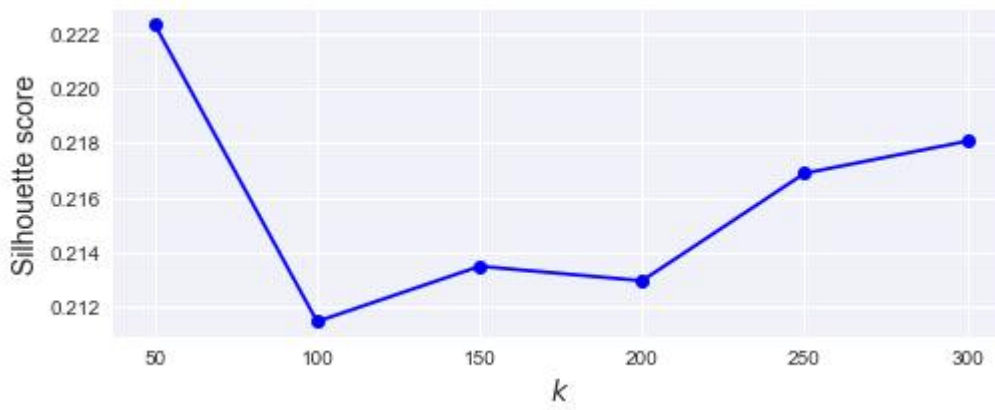
-Inertia, 실루엣 스코어 측정

1) Feature를 5개 사용한 경우

Inertia 측정

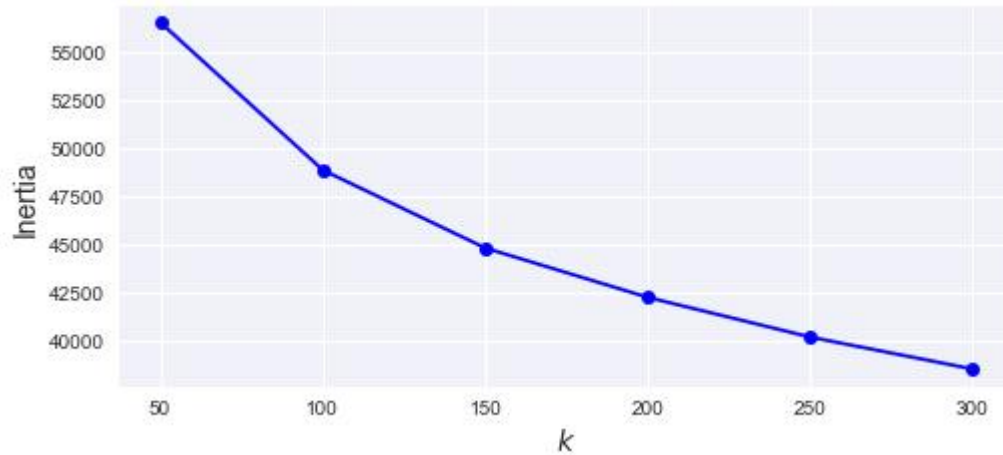


실루엣 스코어



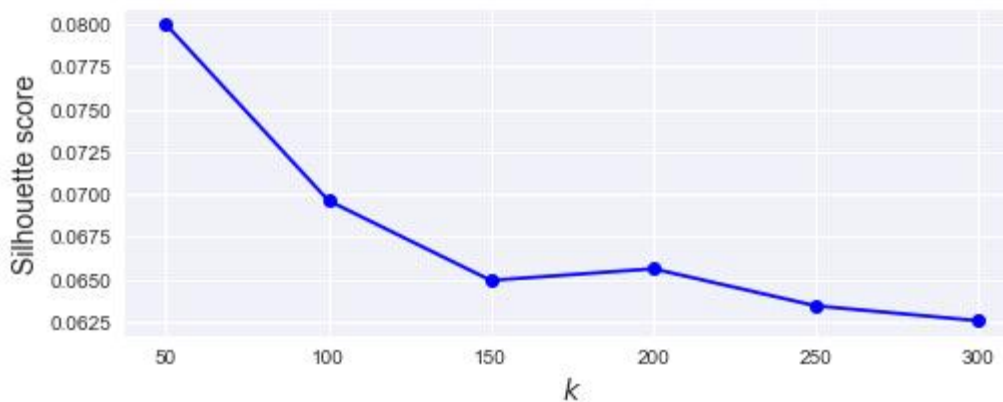
2) Feature 모두 사용한 경우

Inertia 측정



Feature를 5개 사용한 경우보다 Inertia 값이 대체적으로 증가한 것을 확인할 수 있다.

실루엣 스코어



-시각화에 대한 고찰

K가 늘어날 수록 더 세분화 되어가는 느낌이었고 , 더욱 비슷한 Feature 값을 가지는 선수들 끼리 클러스터링이 이루어지는 것을 눈으로 확인할 수 있었다.

-시각화에 대한 한계점

Feature가 24개만 되어도 눈으로 해석하기 매우 어려웠다. 수치적으로는 좋은 결과를 내지 못한 것 같다.

-시나리오 검증에 대한 고찰

실제 축구선수 스카우팅에 대한 프로세스를 찾아보고 최대한 유사한 시나리오를 선정하여 해당 모델을 검증하였다. 각 시나리오 별로 선수의 몸값을 통해 해당 모델의 결과를 검증하였고 이러한 부분이 실제 축구 선수 스카우터 들에게 결과의 설득력을 높이는 부분이 될 수 있다고 생각한다. 실제 다양한 스포츠 분야에서 이러한 데이터 분석을 활용한 선수 스카우팅이 이루

어지고 있다. 현 프로젝트에서는 분야만 축구 일뿐 , 당연히 다른 스포츠에도 적용 가능한 방법이라고 생각하고 이러한 부분에서 의미가 있다고 생각된다.

-시나리오 검증에 대한 한계점

2가지의 시나리오에 대하여 모델을 적용해보고 검증을 진행하였다. 비지도 학습인 만큼 결과의 정확성을 보여줄 수 없는 부분이 없는 점이 아쉽다. 또한 실제 축구 스카우터들 즉 실무자의 눈으로 바라본 모델은 또 다를 것이라 생각한다. 이러한 부분들이 보완되어야 더 나은 모델을 만들어 낼 수 있다고 생각한다.

-전체적인 Insight

축구 선수에 대한 군집화가 프로젝트의 주 주제였지만 , 프로젝트를 진행하면서 다양한 축구 Feature에 대한 Insight 들을 얻을 수 있었다. 예를 들면 골 개수와 도움 개수의 상관관계가 높으리라는 점은 축구를 보는 사람이라면 누구나 쉽게 생각할 수 있는 점이다. 하지만 이 사이에 KeyP(경기당 주요한 패스 횟수) 역시 높은 상관관계를 보인다는 것은 쉽게 생각하기 어렵다. 이러한 부분은 축구 Feature를 분석하는 부분에 특별한 Insight를 제공할 수 있다. 또한 시나리오 부분에서의 모델은 실제 축구팀의 스카우터들이 어떠한 선수를 스카우팅 해야될지 , 1차적으로 선수들을 걸러내는 Insight를 제공 할 수 있다고 생각된다. 국제 축구연맹 FIFA에 가입된 나라의 수가 UN의 회원국보다 많은 만큼 많은 양의 선수들을 1차적으로 걸러낼수 있다는 것은 분명 축구선수 스카우팅 과정에 있어서 의미가 있다고 생각한다.

-팀원 별 역할

강동현 : 데이터 수집 / 모델 생성 / 시나리오 수집 , 테스트 및 검증

박희재 : 데이터 수집 / 모델 생성 / 데이터 시각화

강영걸 : 데이터 수집 / 데이터 병합 / 모델 생성 / 사용 알고리즘 조사 및 설명