

Pockemon Clustering Analysis



AI빅데이터융합경영학과
201965262 장예진
20195263 주고은

Index

01 Dataset

02 K-means Clustering

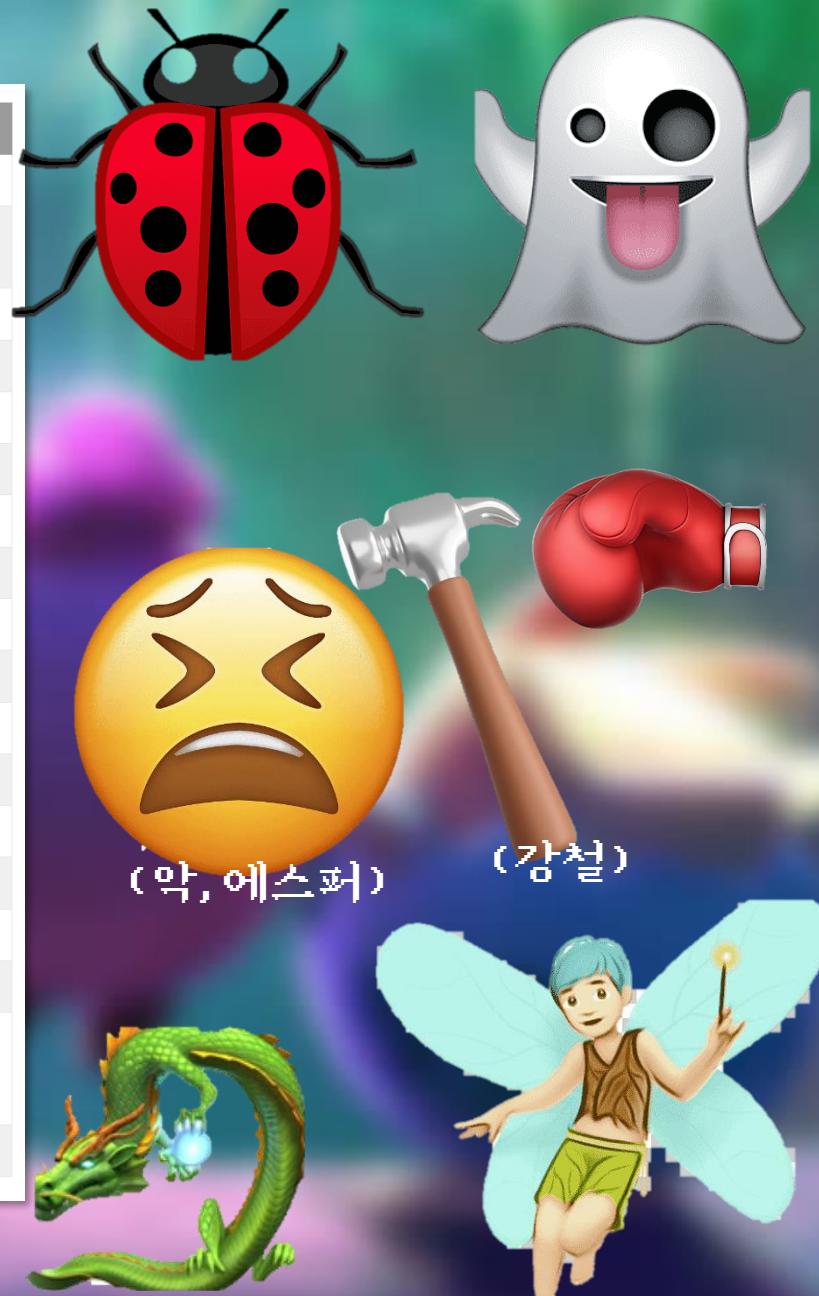
03 Analyzing Clustering

04 Conclusion

APPENDIX



상대포켓몬 타입	효과가 굉장하다	효과가 별로인 듯하다	효과가 없다
노말	격투		고스트
불꽃	물,땅,바위	불꽃,풀,얼음,벌레,강철,페어리	
물	전기,풀	불꽃,물,얼음,강철	
전기	땅	전기,비행,강철	
풀	불꽃,얼음,독,비행,벌레	물,전기,풀,땅	
얼음	불꽃,격투,바위,강철	얼음	
격투	비행,에스퍼,페어리	벌레,바위,악	
독	땅,에스퍼	풀,격투,독,벌레,페어리	
땅	물,풀,얼음	독,바위	전기
비행	전기,얼음,바위	풀,격투,벌레	땅
에스퍼	벌레,고스트,악	격투,에스퍼	
벌레	불꽃,비행,바위	풀,격투,땅	
바위	물,풀,격투,땅,강철	노말,불꽃,독,비행	
고스트	고스트,악	독,벌레	노말,격투
드래곤	얼음,드래곤,페어리	불꽃,물,전기,풀	
악	격투,벌레,페어리	고스트,악	에스퍼
강철	불꽃,격투,땅	노말,풀,얼음,비행,에스퍼,벌레,바위,드래곤,강철,페어리	독
페어리	독,강철	격투,벌레,악	드래곤



APPENDIX



이름: 피카츄

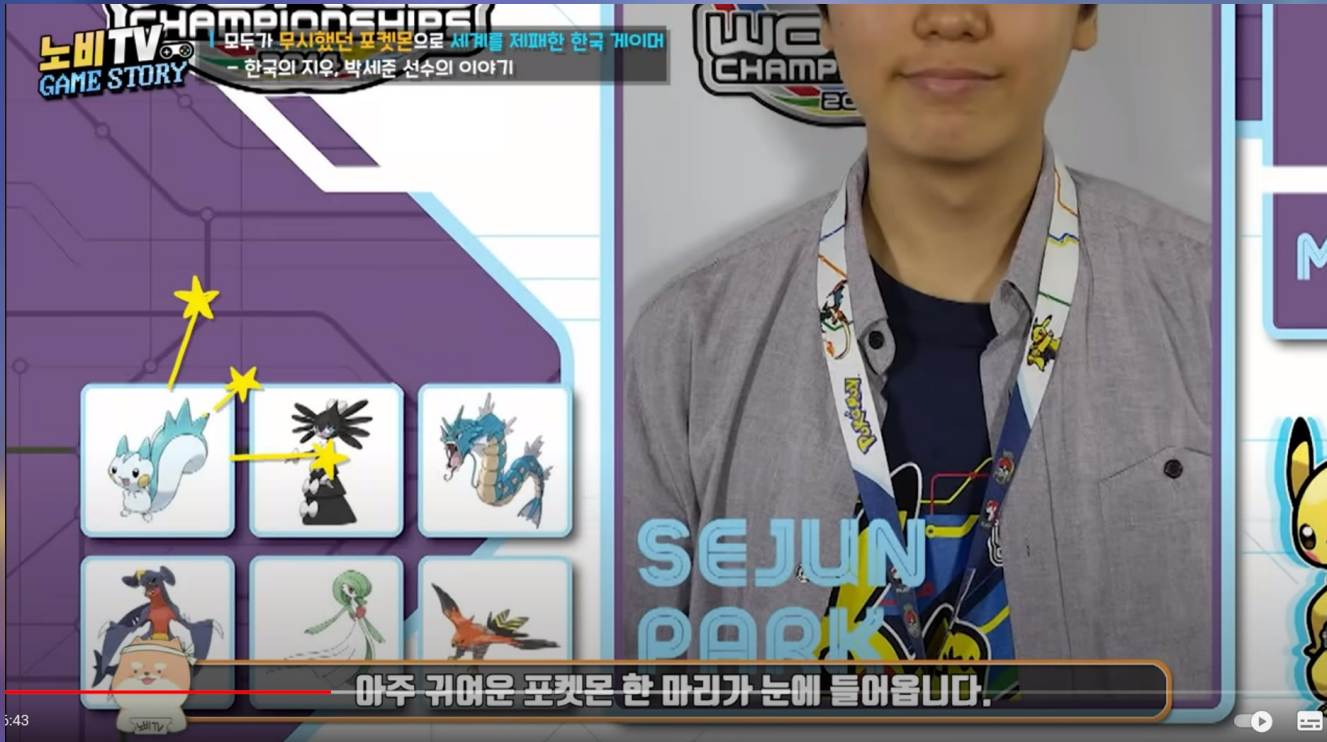
타입: 전기

약점(weakness): 땅 타입(ex.꼬마돌한테 약함)

강점(strength): 물 타입 (ex.꼬부기한테 강함)

***그렇지만 스탯을 꼭 확인해야 함!

Goal



흔히 대결에서 많이
쓰이는 드래곤 타입의
강한 포켓몬이 아닌
조합을 구성한 전략

>> 이러한 전략에 따른
최적의 포켓몬 조합을
군집분석을 통해 찾아
보자!



Dataset

48kb
Size

32
Using variable

Kaggle
From



ROUNAK BANIK · UPDATED 6 YEARS AGO

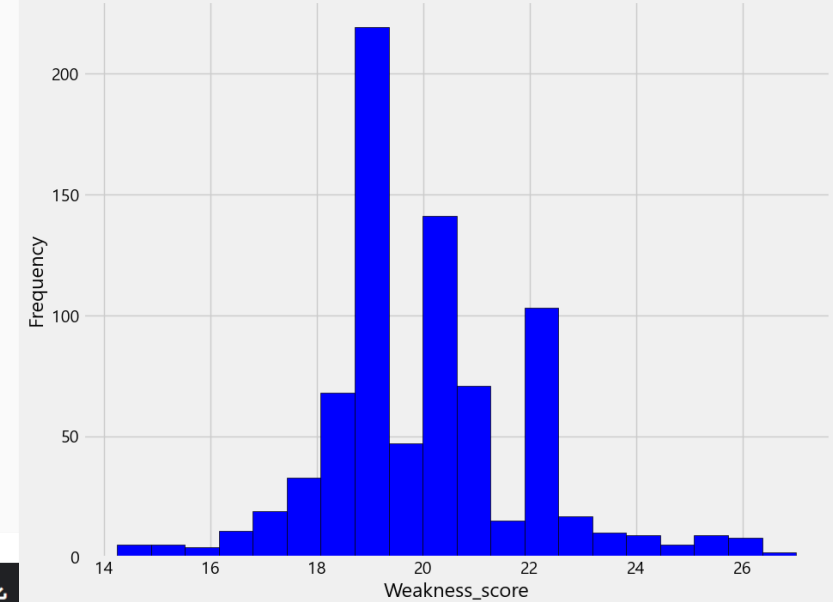
The Complete Pokemon Dataset

Data on more than 800 Pokemon from all 7 Generations.

Data Card Code (179) Discussion (11)

```
df1.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 801 entries, 0 to 800
Data columns (total 32 columns):
#   Column              Non-Null Count  Dtype
---  -
0   abilities            801 non-null   object
1   against_bug          801 non-null   float64
2   against_dark         801 non-null   float64
3   against_dragon       801 non-null   float64
4   against_electric     801 non-null   float64
5   against_fairy        801 non-null   float64
6   against_fight        801 non-null   float64
7   against_fire         801 non-null   float64
8   against_flying       801 non-null   float64
9   against_ghost        801 non-null   float64
10  against_grass        801 non-null   float64
11  against_ground       801 non-null   float64
12  against_ice          801 non-null   float64
13  against_normal       801 non-null   float64
14  against_poison       801 non-null   float64
15  against_psychic      801 non-null   float64
16  against_rock         801 non-null   float64
17  against_steel        801 non-null   float64
18  against_water        801 non-null   float64
19  attack               801 non-null   int64
20  base_total           801 non-null   int64
21  defense              801 non-null   int64
22  hp                   801 non-null   int64
23  name                 801 non-null   object
24  percentage_male      703 non-null   float64
25  sp_attack            801 non-null   int64
26  sp_defense           801 non-null   int64
27  speed               801 non-null   int64
28  type1                801 non-null   object
29  type2               417 non-null   object
30  generation           801 non-null   int64
31  is_legendary         801 non-null   int64
dtypes: float64(19), int64(9), object(4)
```

Distribution of Pokémon Scores





Dataset

1) 특성변수 abilities

: 각 포켓몬의 고유한 특성을 나타냄.

>본 연구에선 상성과 능력치만을 다루기에 제외시킴.

2)상성변수 against_

: 각 타입에 대한 포켓몬의 상성을 나타냄. 0,0.5,1.0,2.0으로 이루어져있으며 숫자가 클수록 해당 타입에 약함을 나타냄

3)능력치 변수 base_total, attack, defense, sp_attack, sp_defense, speed

:포켓몬의 능력치를 나타내며 총 능력치 합, 공격, 방어, 특수공격, 특수방어, 스피드를 나타냄

4)이름변수 name

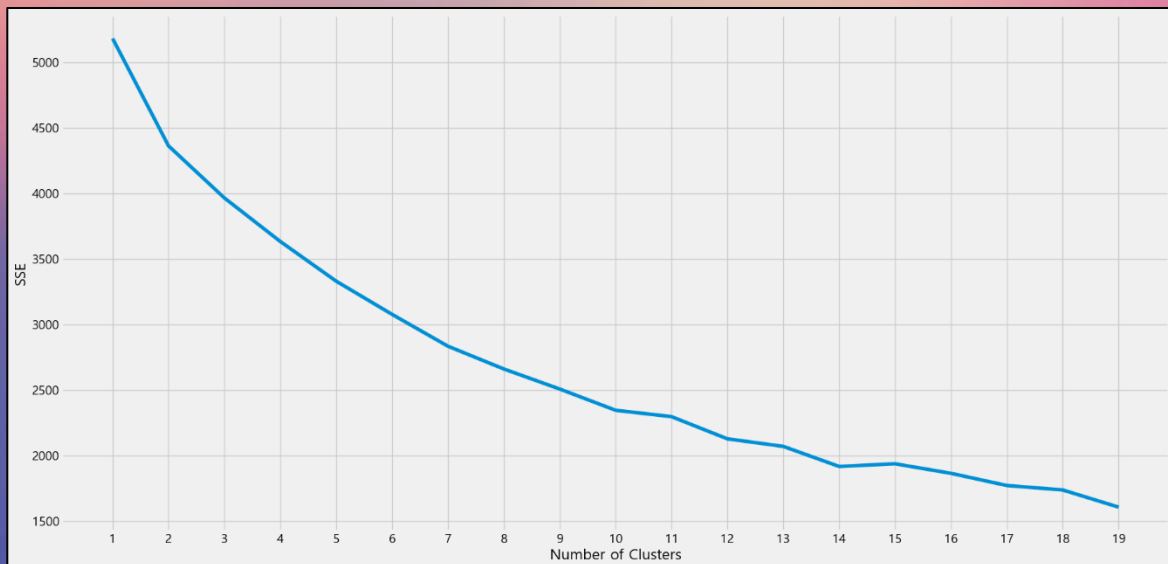
5)타입변수 type1, type2

:벌레, 악, 드래곤, 전기, 요정, 격투, 불, 비행, 고스트, 풀, 땅, 얼음, 노멀, 독, 에스퍼, 바위, 강철, 물 18가지의 타입으로 이루어져있음. 각 포켓몬은 type1과 type2를 배정받으며 type1만 있는 단일타입에 경우 type2는 nan값.

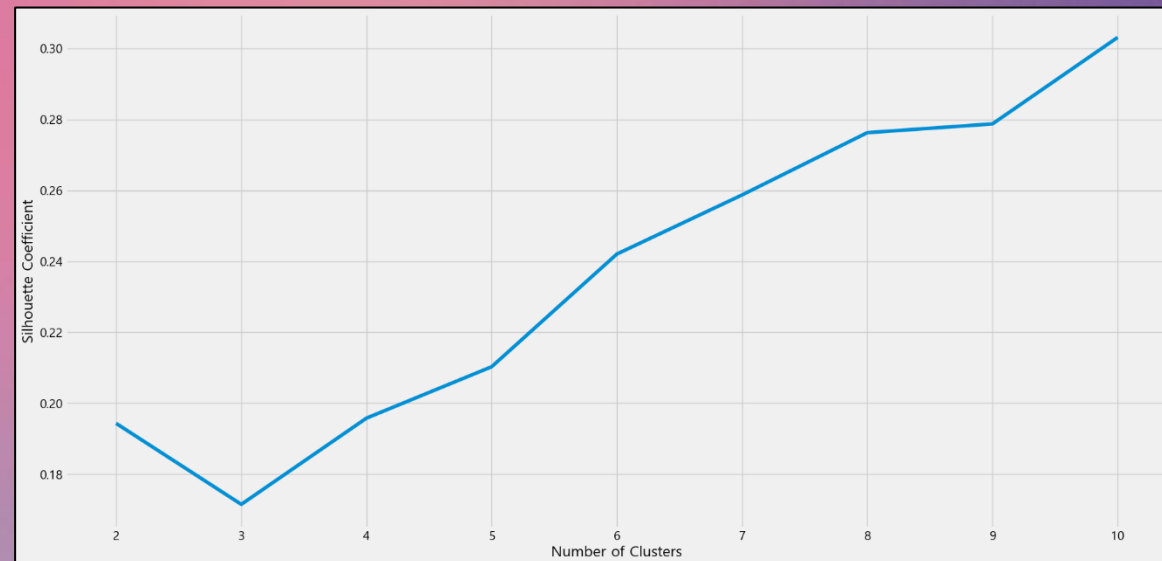
결측치 제거를 위해 '없음'으로 대체함



K-means Clustering



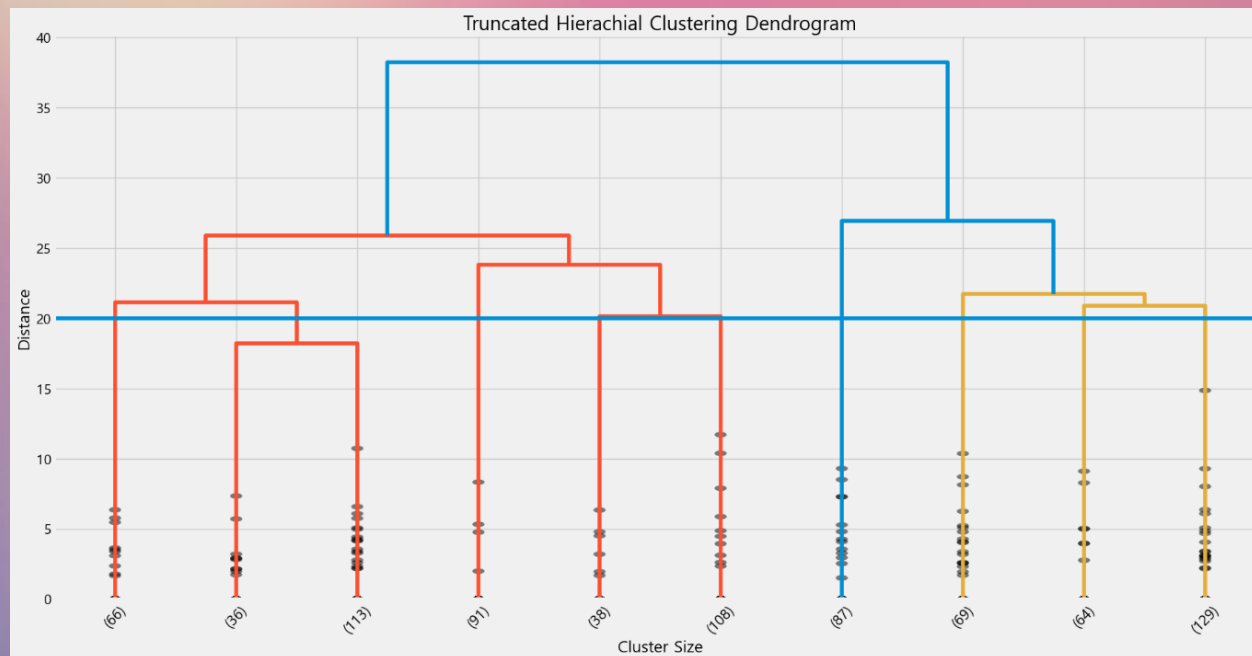
군집의 수에 따른 SSE를 나타내는 Elbow 그래프:
K = 5 정도에서 눈에 띄게 줄어듦.



실루엣 계수(Silhouette Coefficient) 그래프:
군집의 수가 5 이상일 때 군집화가 잘 이루어짐.



Hierarchical Clustering

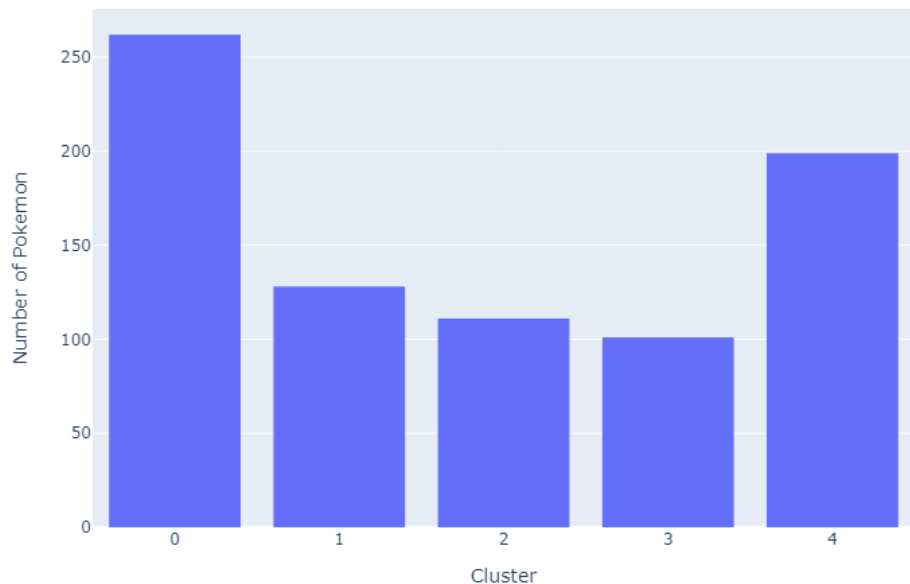


계층적 군집분석 덴드로그램 : 데이터 포인트들을 서로 유사한 그룹으로 분류
덴드로그램 아래쪽에 있는 괄호안의 숫자는 각 군집의 해당 분기점 아래에 있는 데이터포인트의 수를 나타냄.
군집들의 색상은 파랑, 빨강, 노랑, 주황 4가지로 구분

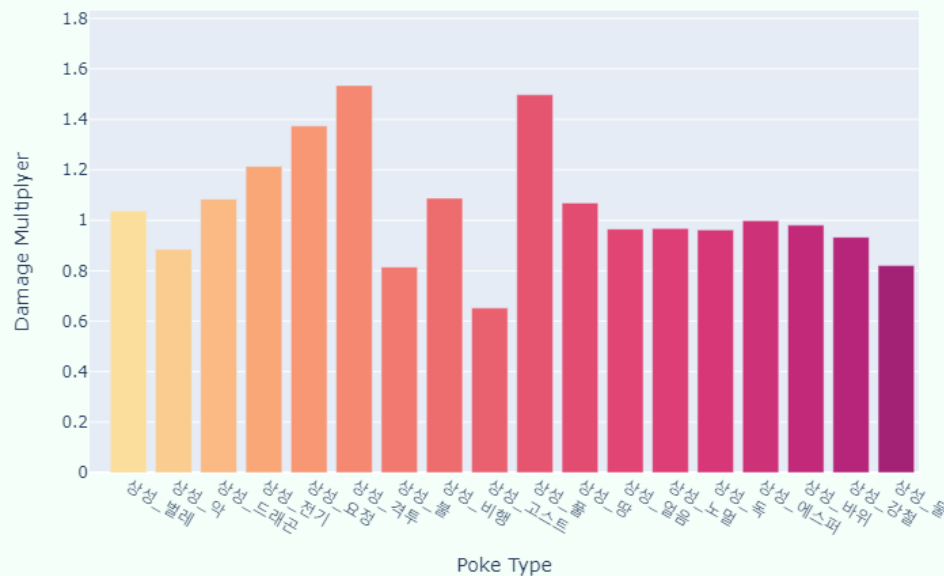


Analyzing Clustering

Number of Pokemon in Each Cluster



Cluster 1 Weakness Chart



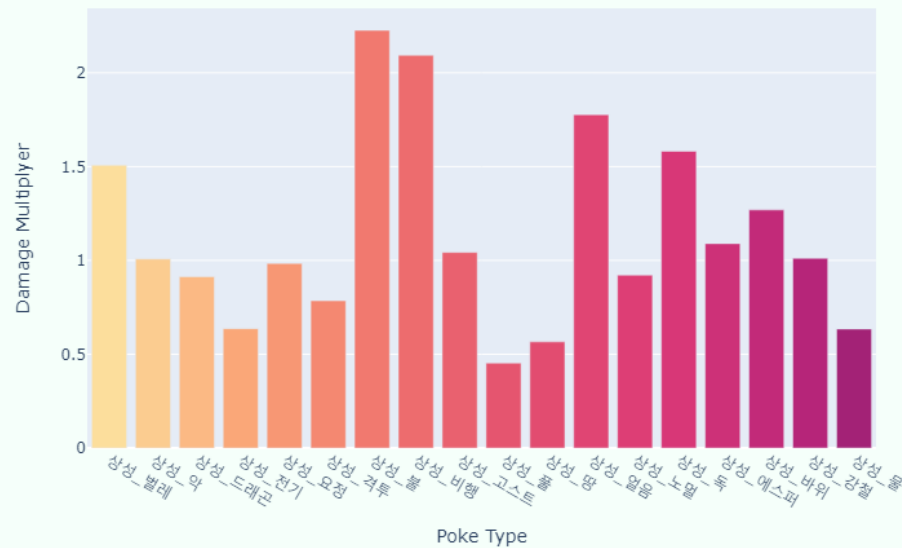
각 클러스터에 속하는 포켓몬의 수
0~4번까지 5개의 클러스터가 존재
첫번째 클러스터에는 가장 많은 포켓몬 존재.

풀, 격투, 페어리(요정), 전기 타입에 약함
고스트, 불, 물 타입에 강함
이 클러스터1에 속하는 포켓몬은
“거북왕, 대잠이, 게을킹, 약어리, 그레닌자”

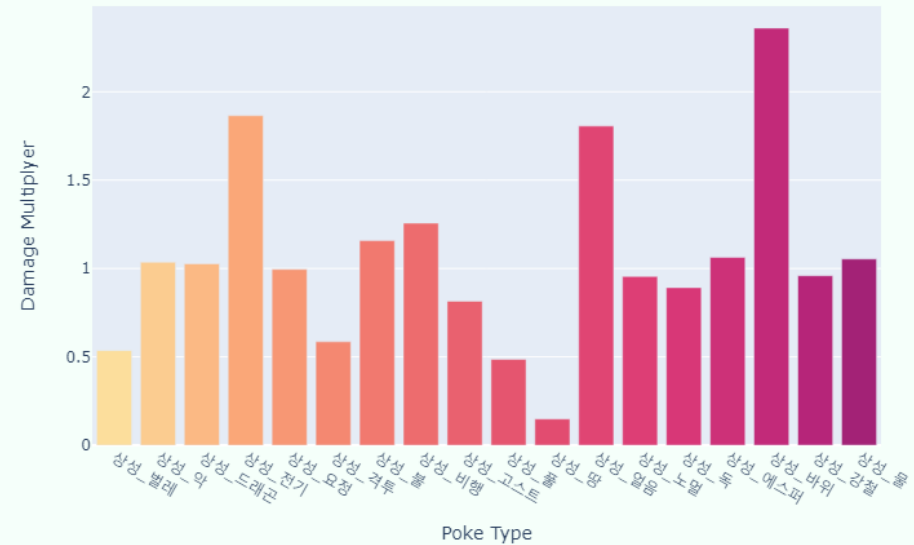


Analyzing Clustering

Cluster 2 Weakness Chart



Cluster 3 Weakness Chart



불, 비행, 얼음 타입에 약함.

풀, 땅, 전기에 강함

해당 클러스터2에 속하는 포켓몬으로

“햇삼, 이상해꽃, 브사이저, 나무킹, 헤라크로스”

바위, 얼음, 전기 타입에 약함.

땅, 격투, 벌레에 강하고 영향을 덜 받음.

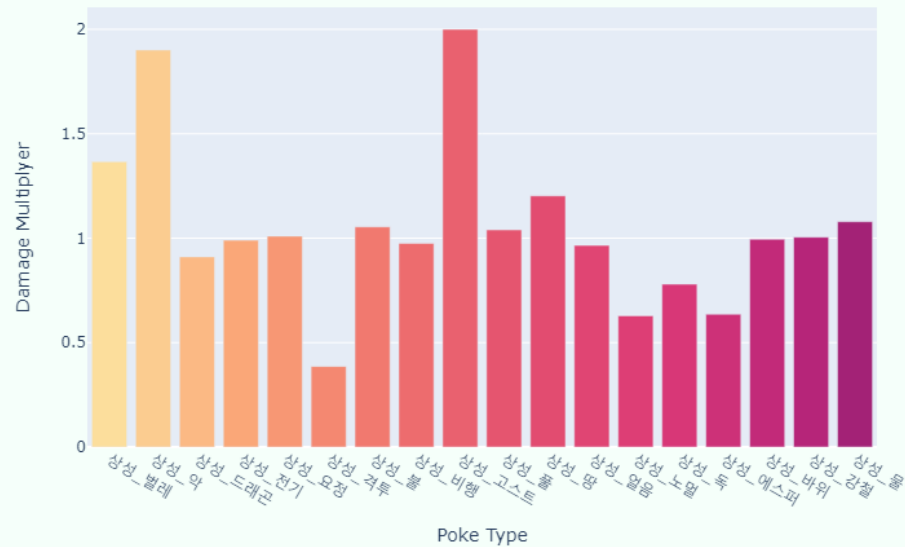
해당 클러스터3에 속하는 포켓몬으로

“리자몽, 가라도스, 프레타, 망나뇽, 보만다”



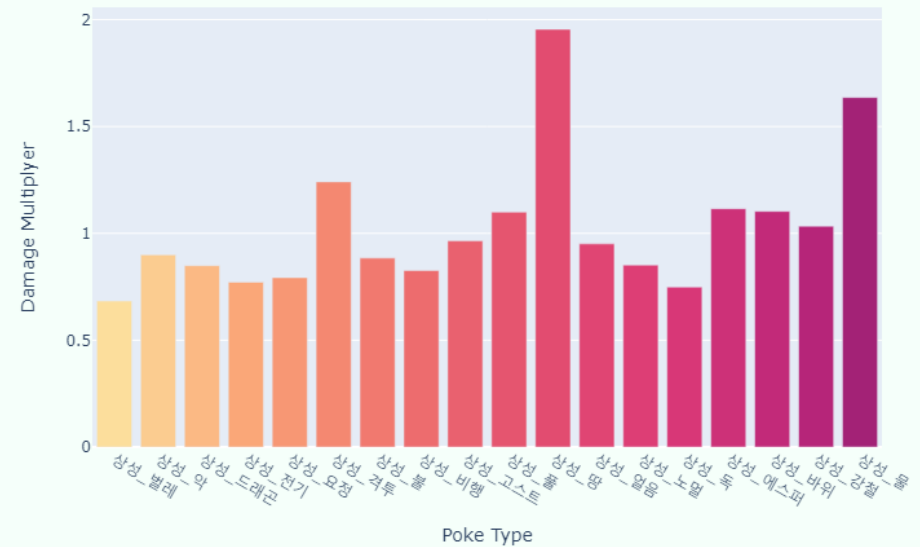
Analyzing Clustering

Cluster 4 Weakness Chart



고스트, 악, 벌레 타입에 약함
격투, 노말, 에스퍼, 독에 강함
해당 클러스터4에 속하는 포켓몬으로
“메타그로스, 팬텀, 가디안, 후딘, 엘레이드”

Cluster 5 Weakness Chart



땅, 물, 격투에 약함.
벌레, 독에 강함
해당 클러스터5에 속하는 포켓몬으로는
“루카리오, 강철톤, 전룡, 헬가, 번치코”

Best Clustering Combination



Best Stats Combination

스탯 총합(공격, 방어, 특수방어, 특수공격, 스피드, hp)을 고려한 각 클러스터의 상위 점수 포켓몬 들 추출

공격, 스피드에 능력치에 따른 변수= physical sweeper

특수공격과 스피드 = special sweeper

물리공격, 특수공격, 스피드를 가진 올라운더 = Mixed sweeper

> 각 클러스터에서의 스탯에 따른 점수에 따른 조합으로 팀 구성

Best Stats Combination



(마기라스)



(프테라)



(리자몽)



(나무킹)



(루카리오)



(후딘)



결과 및 보완점

결과:

(1) 군집분석 (K-means Clustering)을 활용하여 각 군집의 상위 점수에 따른 조합 선정

(2) 군집분석(K-means Clustering)을 활용하여 각 군집의 특성에 대한 필터링을 통해 종합 스탯 상위 점수에 따른 최적의 조합을 선정

> 능력치(hp)가 높은 포켓몬만 선정하는 것이 아닌 각 상성과 타입을 고려한 최적의 시너지 조합 도출 가능!

한계점:

(1) 일반적인 포켓몬 대회 기준에 맞추고자 최대한 노력했지만, 개최되는 각 대회의 기준이 조금씩 상이하여 최적의 기준을 가지는 팀 조합이 약간의 오류 있을 가능성 고려

(2) 포켓몬은 각기 다른 성격과 특성, 기술, 소지품에 따라 배틀 방식이 달라지기 때문에 이러한 변수까지 고려하여 세심한 분석이 필요함

(3) 실질적인 예측을 평가하기 위해 머신러닝으로 모델을 개발하여 추후 보강해볼 예정



Thank you!