



## 이종 포켓몬의 존재

---

# 지원자 : 문경배

# Pokemon데이터 EDA분석

## <분석 목차>

---

1. 데이터 해석
  2. 분석 Process & 해석
  3. 결론
  4. 최종포켓몬 & 아쉬운점
-

## Pokemon 도메인

### 교배

- FACT** - 포켓몬들은 “교배” 시스템을 통해 기존의 알을 낳는다.
- 교배 시스템에서는 ‘수컷’과 ‘암컷’이 알을 낳으면 자식은 ‘암컷’의 종족을 따른다.
  - ex) 수컷 ‘아보’와 암컷 ‘신농’의 교배시 태어나는 포켓몬은 ‘신농’의 하위 포켓몬인 미농
  - 즉, 포켓몬은 ‘이종교배’를 하더라도 한 가지의 종(암컷의 종) 밖에 따를 수 없다.

### 유전

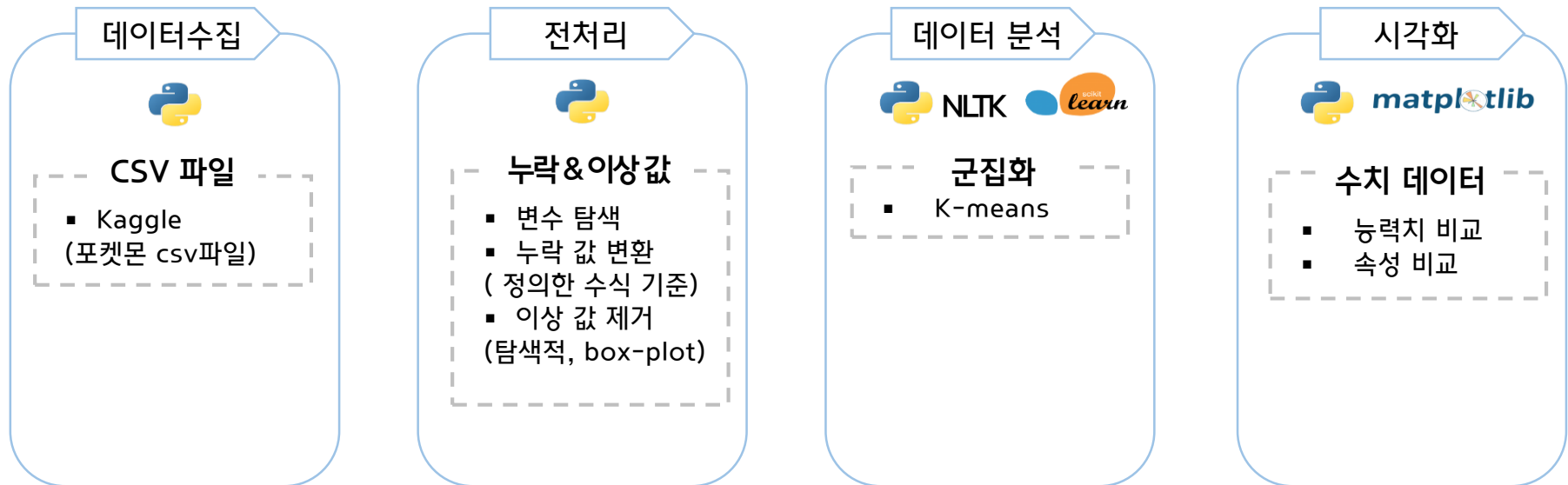
- FACT** - 교배로 태어난 ‘자식’ 포켓몬은 부모의 유전정보 (개체 값)을 따른다.

- 가설** - 이때, ‘동종교배’(수컷 ‘피카츄’ + 암컷 ‘피카츄’ = 자식 ‘피카츄’)를 하면 부모의 평균치 처럼 같은 유전성질을 가질 확률이 높다.
- 반면 ‘이종교배’(수컷 ‘파이리’ + 암컷 ‘피카츄’ = 자식 ‘피카츄’)를 하면 부모의 ‘종족’이 다르므로 자식은 엄마와 똑같은 피카츄지만 유전성질은 수컷과 암컷의 조합으로 ‘매우 다양할 수 있다.’

 ‘순종’의 특성을 많이 가진 종족과 ‘이종’의 특성을 많이 가진 종족(포켓몬)을 구별하는 데이터 분석을 해보자!

- Pokemon Go에서 수집된 포켓몬들의 데이터
- 총 7세대에 걸친 801종의 포켓몬이 hp, attack, defense 등의 변수를 가지고 있다.
- 이중 ‘상성 값’(물, 불, 바람 등)을 종족이 나타내는 고유한 특성이라 생각하여 이 변수를 중심으로 ‘이종’, ‘순종’ 포켓몬을 찾겠다.

### Data Analyzing Process



➡ 자세한 분석 과정은 Jupyter notebook 스크립트에 기록

가설 : 사전적으로 정의된 포켓몬과 다른 "이중" 포켓몬이 존재할 것이다.

전제1 : 포켓몬은 결혼을 통해 자식을 낳는 설정이 있다.

전제2 : 따라서 다른 종간의 번식을 통해 생긴 "이중"포켓몬은 사전적으로 정의된 속성과 다른 능력치를 가지고 있을것이다.(종족의 섞였기 때문에)

귀무가설 : 이중, 순종 포켓몬의 능력치는 같을 것이다.



데이터 분석을 통해 p-value 0.063으로 90%신뢰구간에서 귀무가설을 기각



‘이중‘, ‘순종’ 포켓몬간의 능력치(변수들)의 비교를 통해 “성장성, 가성비, 능력치 ” 를 비교 추천  
\* 순종포켓몬은 일반적인 상성능력치를 가지고 있어 배틀시 불리하다. 따라서 상성 값이 다양한 이중포켓몬을 사용하자!

가성비 좋은 포켓몬 : electric\_ground OR fighting 속성의 포켓몬

성장성이 기대되는 포켓몬 : fighting 속성의 포켓몬

⇒ 가장 좋다 생각되는 포켓몬 fighting(가성비 순위 8위로 조금 낮지만 성장성, 능력치가 너무 좋음)  
Fighting 속성 포켓몬 중

```
In [114]: final_pokemon[['name', 'base_total', 'experience_growth', 'gaseong', 'capture_rate']]
```

```
Out[114]:
```

	name	base_total	experience_growth	gaseong	capture_rate
21	Poliwrath	510	1059860	0.099609	45
73	Heracross	600	1250000	0.093750	45
153	Gallade	618	1250000	0.120703	45
207	Keldeo	580	1250000	0.028320	3
241	Stufful	340	1000000	0.088542	140
242	Bewear	500	1000000	0.130208	70
252	Buzzwole	570	1250000	0.018555	25
253	Pheromosa	570	1250000	0.018555	255



“Gallade” 강력추천!!!



“엘레이드”는 실제 포켓몬 세계관에서도  
매우 강력한 친구

## 아쉬운 점

- 데이터 양의 부족으로 효과적인 군집을 이루지 못했다.
- 데이터의 양이 너무 부족하여 통계적 검증부분에서 근거가 부족하였다.
- 또한, 추가적인 머신러닝(회귀)을 활용하여 인사이트(예측)를 도출하는데 한계가 있었다.
- 만약, 801종의 포켓몬별로 각각 1개의 샘플이 아닌 여러 데이터가 있었다면 더 좋은 분석이 가능했을거 같아 아쉬움이 남는다...
- 더 많은 데이터를 확보하여 “군집”의 성능을 강화할 필요가 있다.