

*** 발표자: 박지윤 p.s. PPT 잘못된 내용 발견 시, 알려주세요 ☺

*** Image by, mscamilaalmeida on pixabay and All icons by, Surang on Flaticon

Superheroes 빅데이터를 활용한 ipywidgets

*** 발표자: 박지윤 p.s. PPT 잘못된 내용 발견 시, 알려주세요 ☺

*** Image by, mscamilaalmeida on pixabay and All icons by, Surang on Flaticon

PART1. 데이터셋 준비

[참고] 원본 데이터셋 링크(Kaggle)
<https://www.kaggle.com/datasets/claudiodavi/superhero-set>



구글 드라이브 호출

```
from google.colab import drive
drive.mount('/content/drive')
```

```
import pandas as pd
```

(외부)데이터셋 불러오기

```
info = pd.read_csv('/content/drive/MyDrive/heroes_information.csv')
abilities = pd.read_csv('/content/drive/MyDrive/heroes_abilities.csv')
```

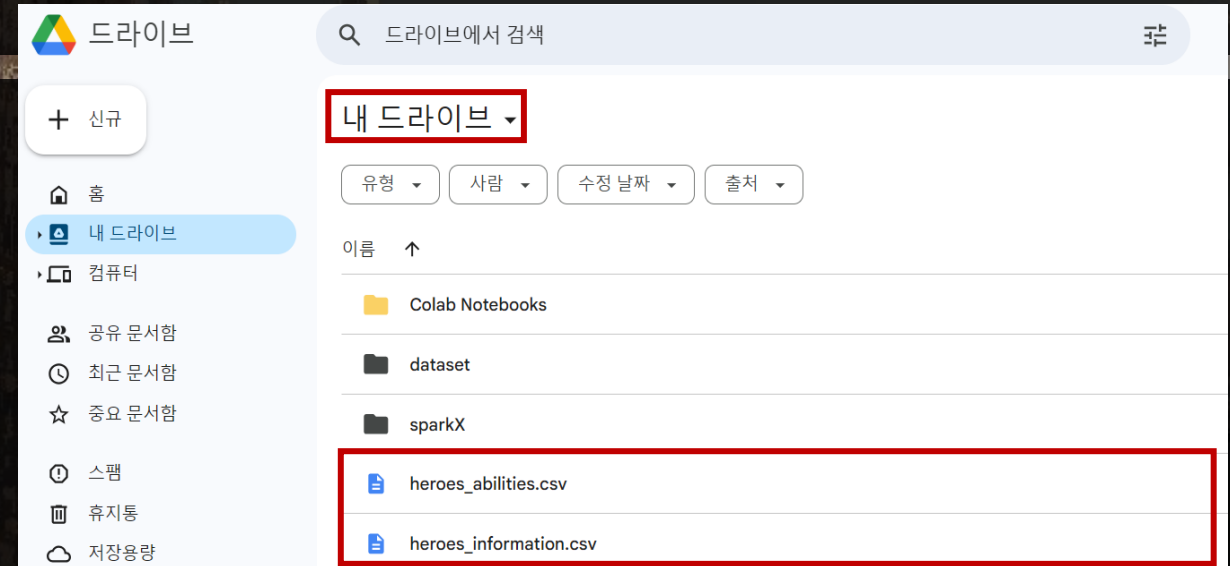
(외부)데이터셋 확인하기

```
print("----- heros_information.csv 확인하기 -----")
print(info.head(3))
```

```
print("----- heros_powers.csv 확인하기 -----")
print(abilities.head(3))
```

```
print(info.info()) # 캐릭터 기본 정보
```

```
print(abilities.info()) # 캐릭터 보유 초능력
```



*** 발표자: 박지윤 p.s. PPT 잘못된 내용 발견 시, 알려주세요 ☺

*** Image by, mscamilaalmeida on pixabay and All icons by, Surang on Flaticon

PART1. 데이터셋 준비

Q1. 첫 번째 데이터셋(info) 안에는 몇 개의 캐릭터 이름이 있을까요?

답:

?

Q2. 두 번째 데이터셋(abilities)는 몇 개의 초능력에 대해 TRUE/FALSE로 나타내고 있나요?

답:

?

*** info

	name	Gender	Eye color	Race	Hair	Height	Publisher	Skin color	Alignment	Weight
0	A-Bomb	Male	yellow	Human	No Hair	203.0	Marvel Comics	—	good	441.0
1	Abe Sapien	Male	blue	Ichthyo Sapien	No Hair	191.0	Dark Horse Comics	blue	Good	65.0
2	Abin Sur	Male	blue	Ungaran	No Hair	185.0	DC Comics	red	good	90.0

*** abilities

	name	Underwater breathing	Teleportation	Telekinesis	Energy Beams	...	Magnetism	Wind Control	Resurrection	Weather Control	Web Creation
0	3-D Man	False	False	False	False	...	False	False	False	False	False
1	A-Bomb	False	False	False	False	...	False	False	False	False	False
2	Abe Sapien	True	False	False	False	...	False	False	False	False	False

*** 발표자: 박지윤 p.s. PPT 잘못된 내용 발견 시, 알려주세요 ☺

*** Image by, mscamilaalmeida on pixabay and All icons by, Surang on Flaticon

PART2. 데이터셋 병합

Q3. 두 데이터프레임 info, abilities의 어떤 열(column)을 기준으로 병합이 될까요?

답:

?

Q4. 두 데이터프레임을 병합할 때 사용하는 pandas 모듈의 메서드는 무엇인가요?

답:

?

*** info

	name	Gender	Eye color	Race	Hair	Height	Publisher	Skin color	Alignment	Weight
0	A-Bomb	Male	yellow	Human	No Hair	203.0	Marvel Comics	—	good	441.0
1	Abe Sapien	Male	blue	Ichtyo Sapien	No Hair	191.0	Dark Horse Comics	blue	Good	65.0
2	Abin Sur	Male	blue	Ungaran	No Hair	185.0	DC Comics	red	good	90.0

*** abilities

	name	Underwater breathing	Teleportation	Telekinesis	Energy Beams	...	Magnetism	Wind Control	Resurrection	Weather Control	Web Creation
0	3-D Man	False	False	False	False	...	False	False	False	False	False
1	A-Bomb	False	False	False	False	...	False	False	False	False	False
2	Abe Sapien	True	False	False	False	...	False	False	False	False	False

*** 발표자: 박지윤 p.s. PPT 잘못된 내용 발견 시, 알려주세요 ☺

*** Image by, mscamilaalmeida on pixabay and All icons by, Surang on Flaticon

PART2. 데이터셋 병합

두 데이터프레임 병합

```
merged_data = pd.merge(info, abilities)
```

```
print(merged_data.head()) # 병합된 데이터프레임 확인하기
```

```
print(abilities.info()) # 캐릭터 기본 정보 & 보유 능력
```

Q5. 현재 데이터프레임에 숨은 “이상한 값”을 찾아보세요! (나중에 결측치로 처리)

답: ? * 힌트: 숫자 데이터

Q6. 일반적으로 데이터프레임에서 결측치는 뭐라고 표기되어 있나요?

답: ?

*** merged_data

	name	Gender	Eye color	Race	Hair	Height	...	Weight	Underwater breathing	Teleportation	...	Weather Control	Web Creation
0	A-Bomb	Male	yellow	Human	No Hair	203.0	...	441.0	False	False	...	False	False
1	Abe Sapien	Male	blue	Ichthy Sapien	No Hair	191.0	...	65.0	False	False	...	False	False
2	Abin Sur	Male	blue	Ungaran	No Hair	185.0	...	90.0	True	False	...	False	False
3	Abomination	Male	green	Human /Radiation	No Hair	203.0	...	441.0	False	False	...	False	False
4	Abraxas	Male	blue	Cosmic Entity	Black	-99.0	...	-99.0	False	True	...	False	False

*** 발표자: 박지윤 p.s. PPT 잘못된 내용 발견 시, 알려주세요 ☺

*** Image by, mscamilaalmeida on pixabay and All icons by, Surang on Flaticon

PART2. 데이터셋 병합

```
import numpy as np
```

```
# 결측치 처리하기
```

```
merged_data.replace(-99, np.nan, inplace=True)
```

```
# 결측치 확인하기(개수)
```

```
print(merged_data.isnull().sum())
```

Q7. 코드 출력 결과, NaN으로 바뀐 결측치가 Height 열과 Weight 열에서 각각 몇 개가 발견되었나요? (Publisher NaN 제외)

답: Height 열에서 ? Weight 열에서 ? 발견

```
# NaN이 있는 행(axis=0) 제거하기
```

```
final_data = merged_data.dropna(axis=0, how='any')
```

```
# NaN이 제거된 최종데이터 확인하기
```

```
print(final_data.head())
```

```
print(final_data.info()) # 최종데이터
```

Q8. 최종데이터에 근거하여, 위젯 만들기에 참여할 수 있는 캐릭터는 몇 개로 정리되었나요?

답: ?

*** 발표자: 박지윤 p.s. PPT 잘못된 내용 발견 시, 알려주세요 ☺

*** Image by, mscamilaalmeida on pixabay and All icons by, Surang on Flaticon

PART3. ipywidgets - (1) 드롭다운

```
import ipywidgets as widgets
from IPython.display import display, clear_output
```

빈 리스트 생성

```
abilities_list = []
```

```
for i in range(10, len(final_data.columns)):
    abilities_list.append(final_data.columns[i])
```

abilities_list 확인

```
print(abilities_list)
```

드롭다운 위젯 생성

```
ability_dropdown = widgets.Dropdown(options=abilities_list)
```

드롭다운 위젯 출력

```
display(ability_dropdown)
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
Index: 462 entries, 0 to 659
```

```
Data columns (total 26 columns):
```

#	Column	Non-Null Count	Dtype
0	name	462 non-null	object
1	Gender	462 non-null	object
2	Eye color	462 non-null	object
3	Race	462 non-null	object
4	Hair color	462 non-null	object
5	Height	462 non-null	float64
6	Publisher	462 non-null	object
7	Skin color	462 non-null	object
8	Alignment	462 non-null	object
9	Weight	462 non-null	float64
10	Underwater breathing	462 non-null	bool
11	Teleportation	462 non-null	bool
12	Telekinesis	462 non-null	bool
13	Energy Beams	462 non-null	bool
14	Fire Control	462 non-null	bool
15	Precognition	462 non-null	bool
16	Hypnokinesis	462 non-null	bool
17	Animal Control	462 non-null	bool
18	Levitation	462 non-null	bool
19	Water Control	462 non-null	bool
20	Omnilingualism	462 non-null	bool
21	Magnetism	462 non-null	bool
22	Wind Control	462 non-null	bool
23	Resurrection	462 non-null	bool
24	Weather Control	462 non-null	bool
25	Web Creation	462 non-null	bool

Q9. 현재 상태에서 드롭다운 위젯을 클릭해서 Underwater breathing를 다른 ability로 선택할 수 있다. (T/F)

답: ?

Q10. 현재 상태에서 다른 ability를 클릭했을 때, 해당 초능력을 보유한 캐릭터의 이름이 출력된다. (T/F)

답: ?



Underwater breathing



*** 발표자: 박지윤 p.s. PPT 잘못된 내용 발견 시, 알려주세요 ☺

*** Image by, mscamilaalmeida on pixabay and All icons by, Surang on Flaticon

PART3. ipywidgets - (1) 드롭다운

filter_heroes(ability) 함수를 통해 선택된 초능력을 가진 캐릭터 이름이 출력되고, 선택이 바뀌면 출력도 같이 바뀝니다.



Q11. 문제 10번이 T가 되려면, 해당 초능력을 보유한 캐릭터들의 이름을 출력하도록 ()를 정의해야 한다.

답: ?

드롭다운 위젯 생성

```
ability_dropdown = widgets.Dropdown(options=abilities_list)
```

드롭다운 위젯 출력

```
display(ability_dropdown)
```

함수 정의

```
def filter_heroes(ability):
    filtered_heros = final_data[final_data[ability] == True] # 데이터프레임 조건 필터링
    clear_output(wait=True) # 이전 출력이 있다면? clear하고 새로운 출력 준비
    display(ability_dropdown)
    display(filtered_heros['name']) # 선택된 초능력을 가진 캐릭터 이름 출력
```

함수 적용

```
ability_dropdown.observe(lambda change: filter_heroes(change.new), names='value')
```

< observe() 메서드 >

- handler function → lambda 함수
(그 안에 filter_heroes() 함수 호출)
- filter_heroes()의 매개변수는 새로 바뀐 change(new 속성)
- names → 추적할 속성(abilities_list의 선택된 value 변화)

Q12. 현재 상태에서 다른 ability를 클릭했을 때, 해당 초능력을 보유한 캐릭터의 이름이 출력된다. (T/F)

답: ?

Q13. 함수에서 clear_output(wait=True)를 없애고 실행하면, 드롭다운 선택을 바꿀 때마다 출력 결과가 계속 누적된다. (T/F)

답: ?



Underwater breathing

*** 발표자: 박지윤 p.s. PPT 잘못된 내용 발견 시, 알려주세요 ☺

*** Image by, mscamilaalmeida on pixabay and All icons by, Surang on Flaticon

PART3. ipywidgets - (2) 검색 위젯(텍스트+버튼)

```
import ipywidgets as widgets
from IPython.display import display, clear_output
```

텍스트 위젯 생성

```
search_box = widgets.Text(placeholder='Enter character name', description='Search:')
```

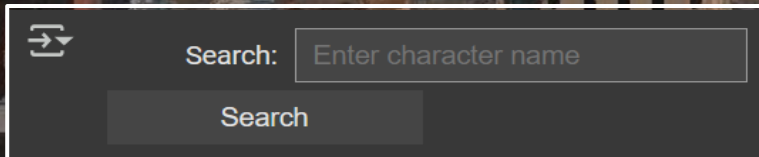
버튼 위젯 생성

```
search_button = widgets.Button(description='search')
```

두 위젯 출력

```
display(search_box, search_button)
```

Q14. 현재 출력된 텍스트 위젯에 검색할 캐릭터 이름을 입력하고 버튼 위젯을 클릭하면 기본 정보가 출력이 될까요? 답: ?



The screenshot shows a dark-themed IPython widget interface. On the left, there is a small icon of a document with a magnifying glass. To its right, the text "Search:" is displayed. Next to "Search:" is a text input field with the placeholder text "Enter character name". Below the input field is a button with the text "Search".

*** 발표자: 박지윤 p.s. PPT 잘못된 내용 발견 시, 알려주세요 ☺

*** Image by, mscamilaalmeida on pixabay and All icons by, Surang on Flaticon

PART3. ipywidgets - (2) 검색 위젯(텍스트+버튼)

```
# 함수 정의
def on_button_click(event):    # event는 함수 내 매개변수로 들어간 게 아니라 버튼 클릭 "이벤트 객체"
    search_term = search_box.value
    if not len(search_term) == 0:
        filter_by_name(search_term)    # 검색어가 있는 경우, filter_by_name() 함수 호출
    else:
        print("No name entered. Please try again.")    # 검색어가 없는 경우, 오류 메시지 출력

def filter_by_name(search_name):
    filtered_data = final_data[final_data['name'].str.contains(search_term, case=False)]
    clear_output(wait=True)
    display(search_box, search_button)
    display(filtered_data.iloc[:, :10])

# 함수 적용
# 버튼 클릭 → on_button_click() 함수 호출
search_button.on_click(on_button_click)
```



Q15. 나의 눈동자 색은?

- Height: 178.0(cm)
- Weight: 74.0(kg)



Q16. Hello, I'm Thor.

- Height: ? (cm)
- Weight: ? (kg)



Q17.

"I am Groot"
"I am Groot"
"I am Groot"
"I am Groot"

Q15. Spider-Man의 눈동자 색(Eye color)은 어떤 색깔일까요?

답: ?

Q16. 토르의 키(cm)와 몸무게(kg)를 각각 검색 결과에서 찾아주세요.

답: ? / ?

Q17. 그루트의 성별과 종족을 각각 검색 결과에서 찾아주세요.

답: ? / ?

*** 발표자: 박지윤 p.s. PPT 잘못된 내용 발견 시, 알려주세요 ☺

*** Image by, mscamilaalmeida on pixabay and All icons by, Surang on Flaticon

[추가] 데이터프레임 시각화 - matplotlib

퀴즈 정답은 "한국어"로 공개하겠다!!



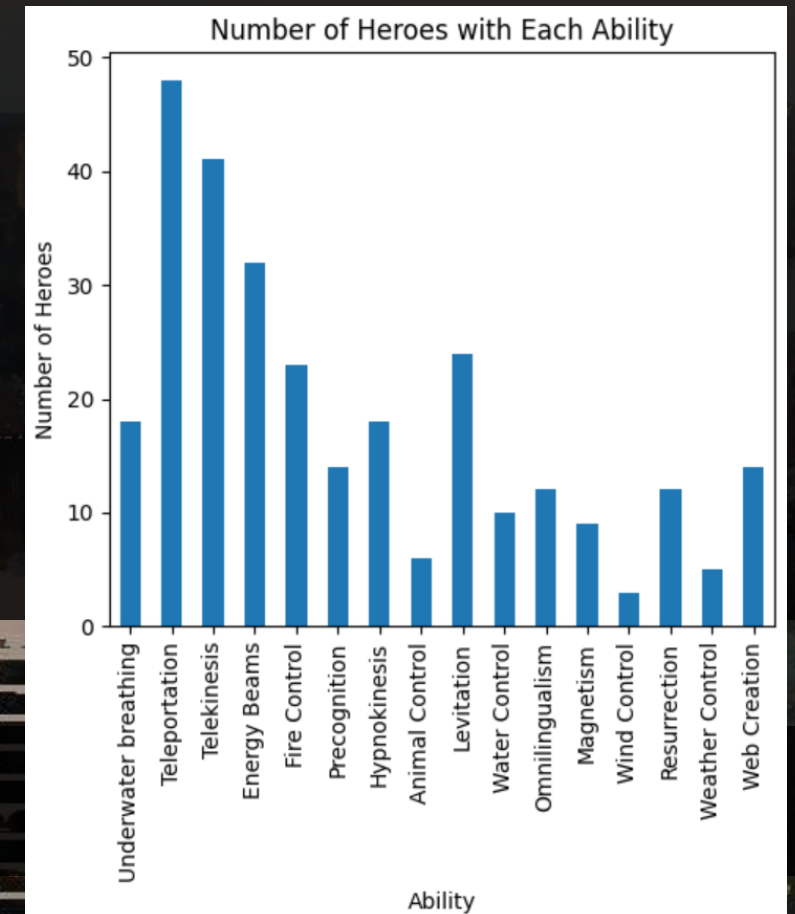
```
import matplotlib.pyplot as plt
```

```
ability_counts = final_data.iloc[:, 10:].sum() # 숫자 기반 슬라이싱 .iloc
# final_data.loc[:, final_data.columns[10:].sum()과 같음 (레이블 기반 슬라이싱 .loc)
print(ability_counts)
```

```
plt.figure(figsize=(5, 6))
```

```
ability_counts.plot(kind='bar') # 막대그래프
```

```
plt.title("Number of Heroes with Each Ability") # (제목) 각 초능력을 보유한 캐릭터 수
plt.xlabel("Ability") # (x레이블) 캐릭터 보유 초능력
plt.ylabel("Number of Heroes") # (y레이블) 해당하는 캐릭터 수
plt.tight_layout()
plt.show()
```



Q. 어떤 초능력이 가장 많은 캐릭터들이 보유하고 있나요? 답: ?

Q. 어떤 초능력이 가장 적은 캐릭터들이 보유하고 있나요? 답: ?

*** 발표자: 박지윤 p.s. PPT 잘못된 내용 발견 시, 알려주세요 ☺

*** Image by, mscamilaalmeida on pixabay and All icons by, Surang on Flaticon

[추가] 데이터프레임 시각화 - matplotlib

퀴즈 정답은 "한국어"로 공개하겠다!!



```
import matplotlib.pyplot as plt
```

```

ability_counts = final_data.iloc[:, 10:].sum()      # 숫자 기반 슬라이싱 .iloc
# final_data.loc[:, final_data.columns[10:].sum()과 같음 (레이블 기반 슬라이싱 .loc)
print(ability_counts)

plt.figure(figsize=(5, 6))

ability_counts.plot(kind='bar')      # 막대그래프

plt.title("Number of Heroes with Each Ability") # (제목) 각 초능력을 보유한 캐릭터 수
plt.xlabel("Ability")                  # (x레이블) 캐릭터 보유 초능력
plt.ylabel("Number of Heroes")        # (y레이블) 해당하는 캐릭터 수
plt.tight_layout()
plt.show()

```

Q. 어떤 초능력이 가장 많은 캐릭터들이 보유하고 있나요? 답: ?

Q. 어떤 초능력이 가장 적은 캐릭터들이 보유하고 있나요? 답: ?

1. 수중 호흡(잠수)	Underwater breathing	18
2. 순간 이동	Teleportation	48
3. 염력	Telekinesis	41
4. 에너지 빔	Energy Beams	32
5. 화염 제어	Fire Control	23
6. 예지력	Precognition	14
7. 최면	Hypnokinesis	18
8. 동물 조종	Animal Control	6
9. (공중)부양	Levitation	24
10. 물 다루기	Water Control	10
11. 전언어 능력	Omnilingualism	12
12. 자기력	Magnetism	9
13. 풍력 제어	Wind Control	3
14. 부활(생명 재생)	Resurrection	12
15. 날씨 제어	Weather Control	5
16. 웹(거미줄) 생성	Web Creation	14