

AI

NEWSLETTER

**K-water
AI Research Lab.**

18 JAN 2023

AI News

언어-이미지 모델 PaLI

PaLI를 통해 주어진 이미지를 묘사하는 자막을 생성하거나 질의응답을 할 수 있습니다. 총 100여개의 언어를 지원하며 이미지 분류, OCR, 텍스트 추론 등이 가능합니다.



Input: Answer in EN: what is the brand of this watch
Output: seiko



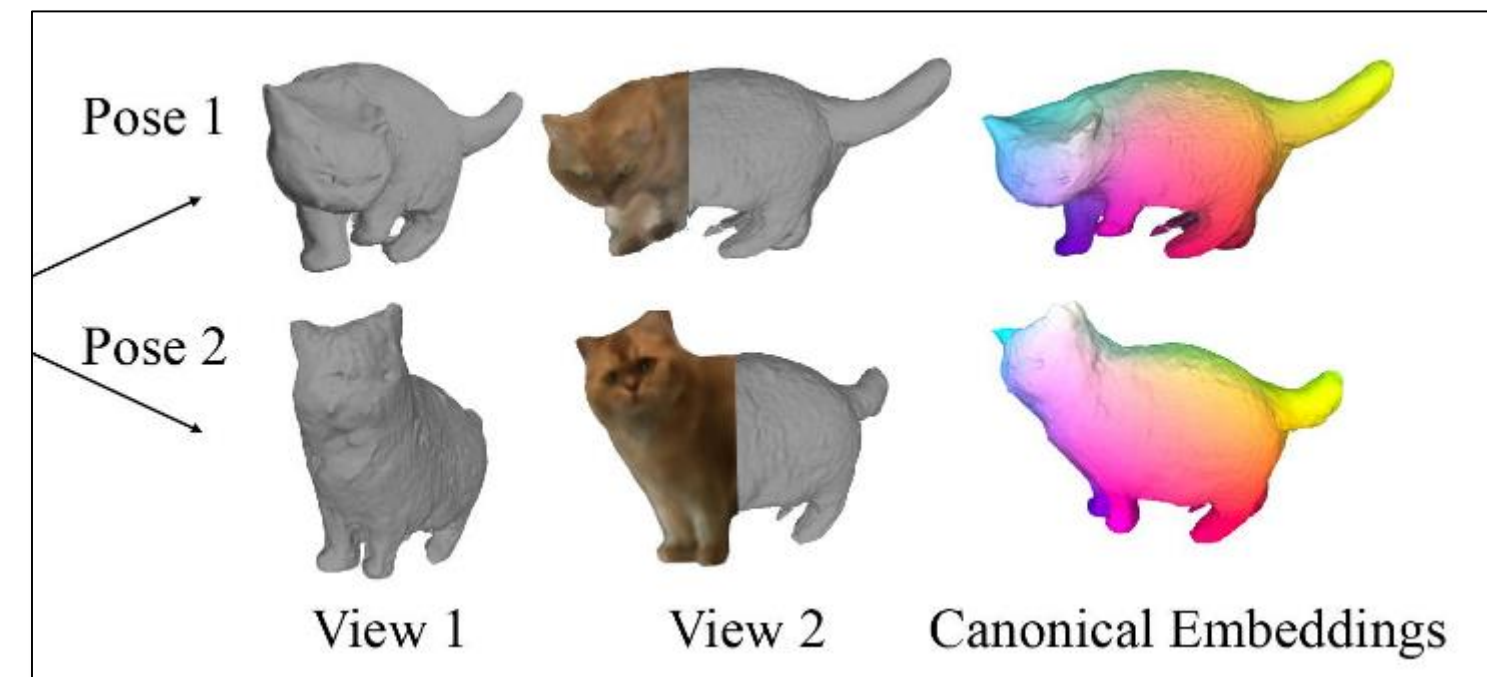
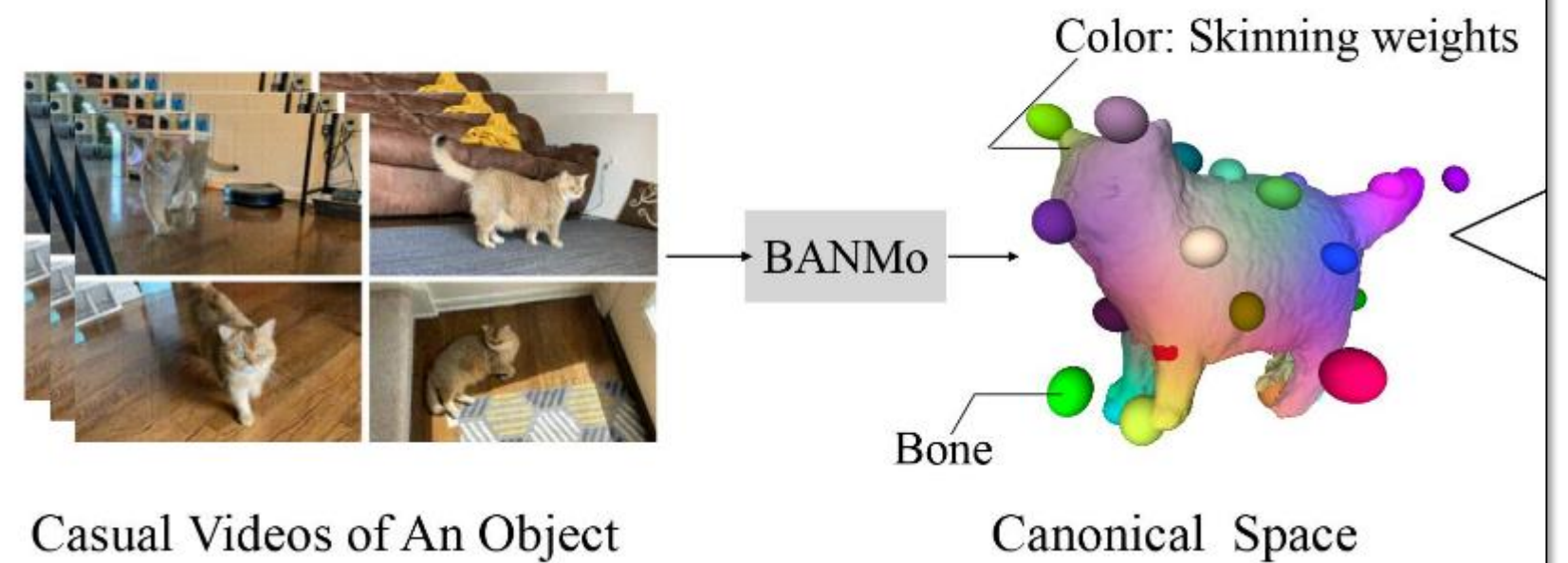
Input: Generate the alt_text in EN
Output: A cellar filled with barrels of wine



Input: Generate the alt_text in FR
Output: Un arbre debout dans un champ avec un ciel violet

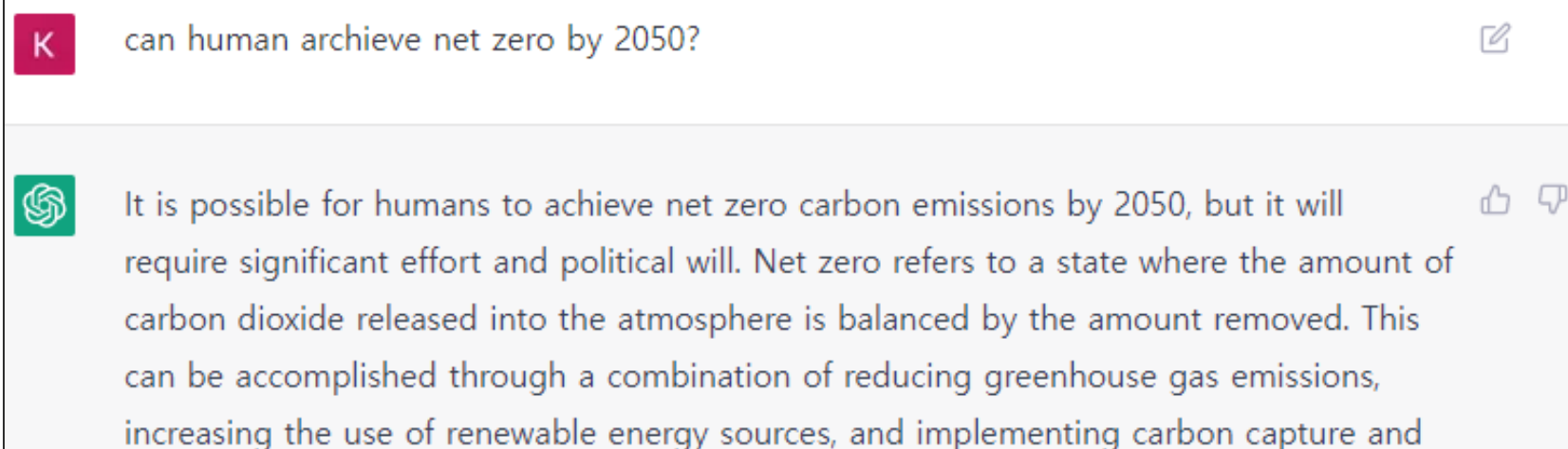
영상을 3D모델로! BANMo

CVPR 2022에서 비디오 자료를 활용하여 3D 애니메이션을 만들어주는 모델이 공개되었습니다.



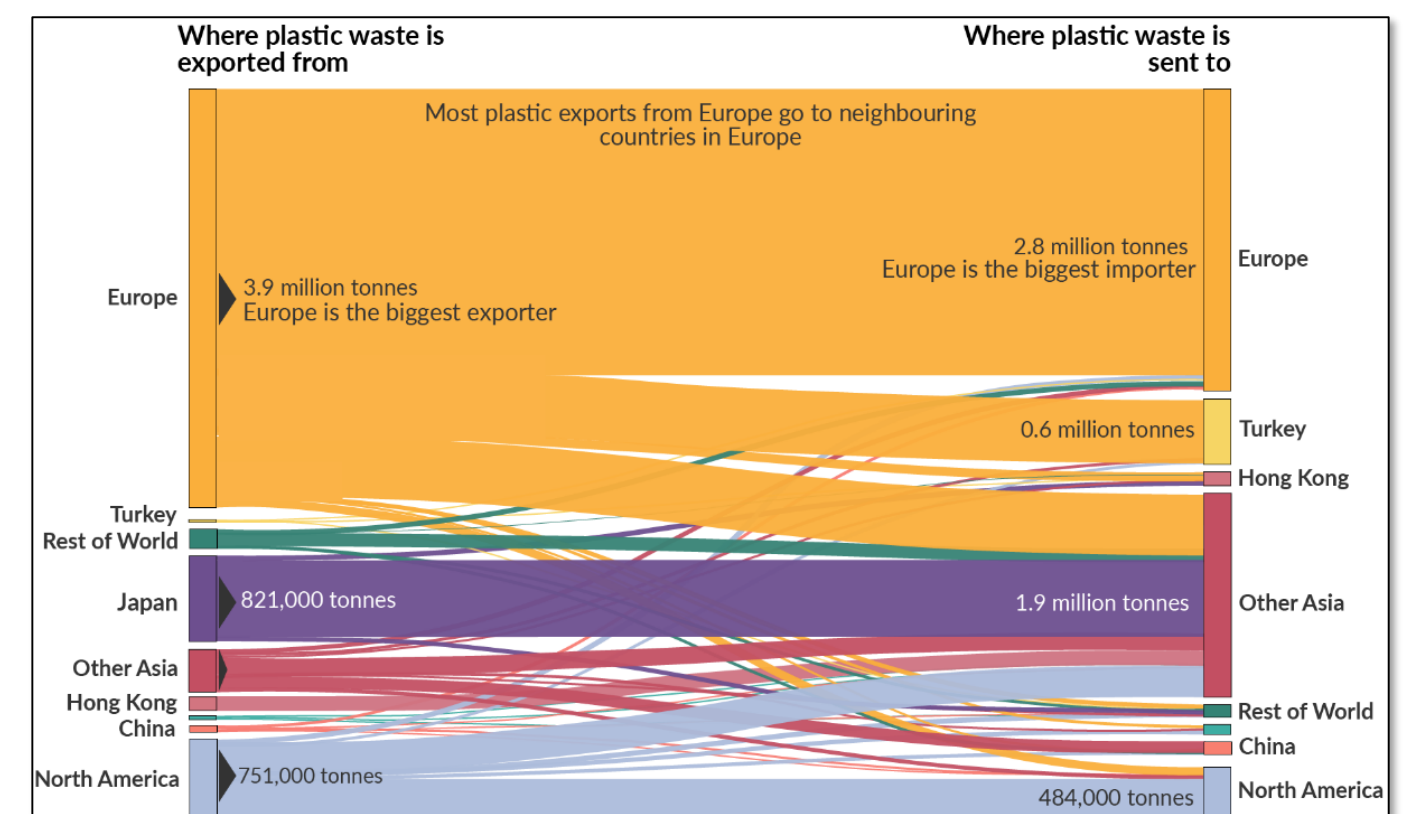
ChatGPT

요즘 **핫**한 ChatGPT입니다! 창간호에서 소개한 DALL-E2(우주비행사)를 기억하시나요? OpenAI에서 이번에는 대화형 인공지능을 발표했습니다. 영문으로 되어 있지만, 한글도 어느정도 가능합니다.
*Links: <https://chat.openai.com/>



시각화_생키 차트

전 세계적으로 플라스틱 폐기물의 이동은 어떻게 될까요? 아래 생키 차트를 통해 확인할 수 있습니다.





QR SCAN

Hands-On AI Project



QR SCAN

수도공급량 분석 (시계열)

1. 데이터를 준비 준비합니다.

- 구미 4공단 배수지의 유출유량 적산차를 활용합니다.

```
data_raw = pd.read_excel('water.xlsx', index_col='logTime')
```

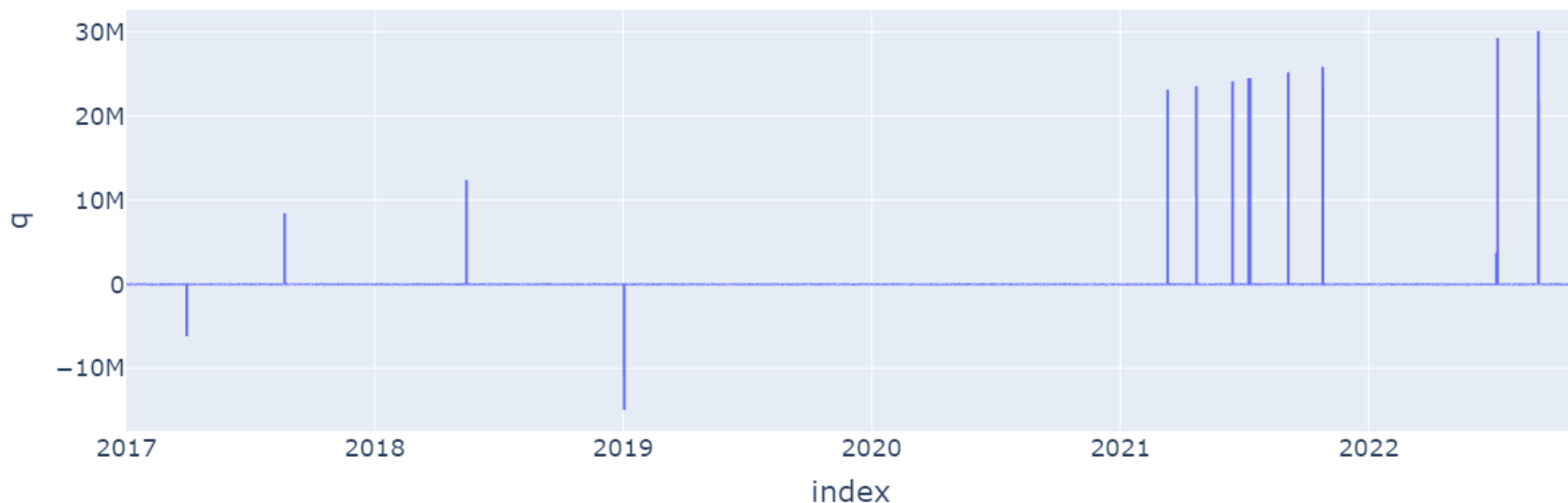
2. index타입을 datetime으로 변경합니다.

- 시간은 0~23의 숫자를 가져야 하는데, 수도데이터는 1~24의 값을 가지므로 이를 조절해줍니다.

```
index_new = []
for date in data_raw.index:
    hour = int(date.split(' ')[1].split(':')[0]) - 1
    idx_new = date.split(' ')[0] + " " + str(hour) + ":" +
    date.split(' ')[1].split(':')[1]
    index_new.append(idx_new)
data = data_raw
data.index = pd.to_datetime(index_new)
```

3. 전체적인 그림을 그려봅시다.

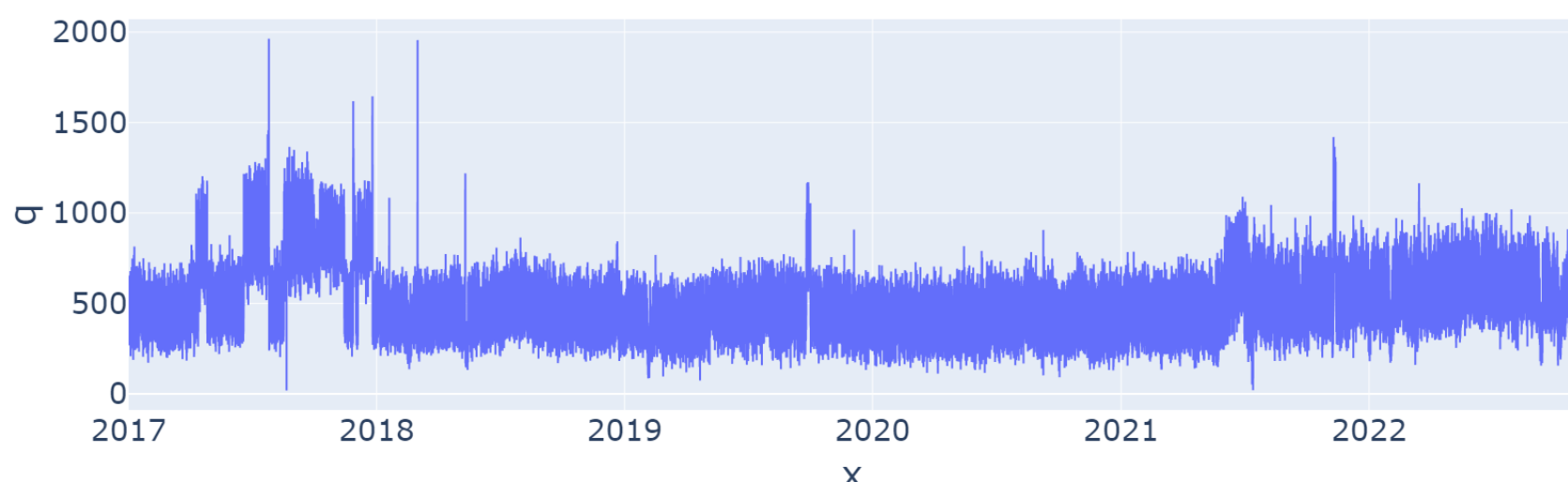
```
fig = px.line(data, x=data.index, y='q', width=1600, height=500)
```



4. 이상치를 처리해 줍니다.

- 0 이하, 2000을 초과하는 데이터는 제외합니다.

```
idx_minus = data['q'] <= 0
idx_big_number = data['q'] > 2000
```



5. 클리닝 된 것을 확인할 수 있습니다.

뉴스 텍스트 분석 (가짜뉴스 part2)

1. 데이터를 준비합니다 (GitHub)

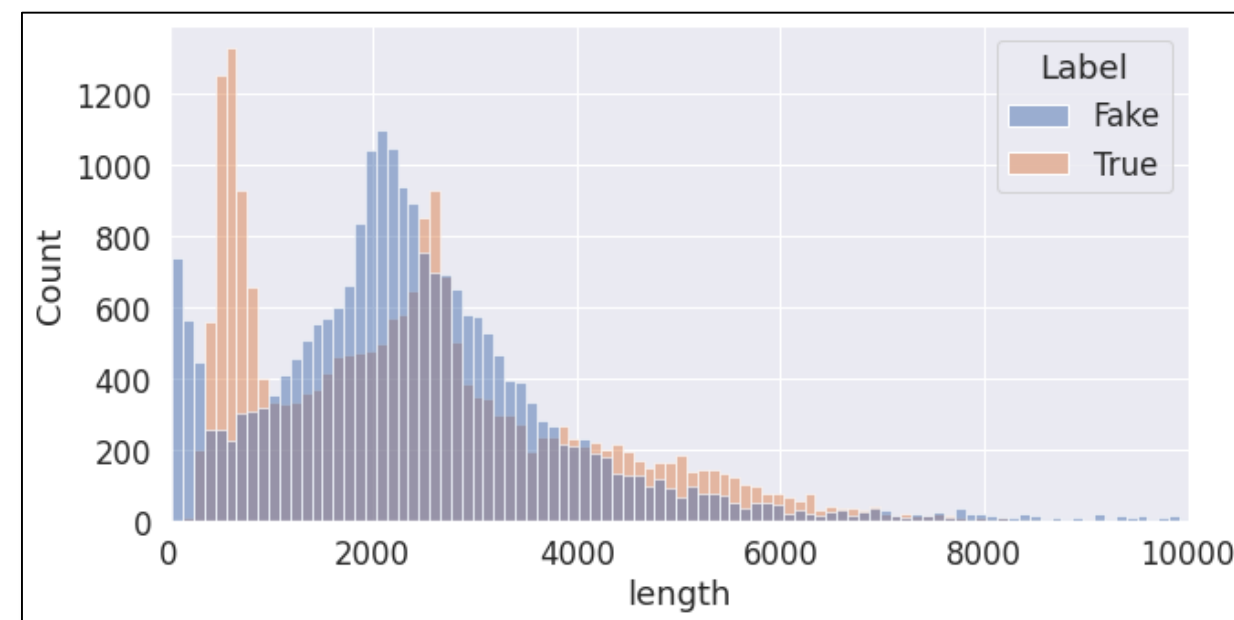
- 창간호의 '가짜뉴스 판단하기' 데이터와 동일합니다.

```
!wget https://github.com/muttinenisairohith/FakeNewsDe ....
```

2. 가짜/진짜 뉴스의 길이를 비교해 봅시다.

- 극단적으로 짧거나 1000~2000자는 가짜뉴스가, 400~1000자 사이는 진짜뉴스가 분포되어 있습니다.

```
data['length'] = data['text'].apply(lambda x: len(x))
sns.histplot(data=data, x='length', hue='Label')
```



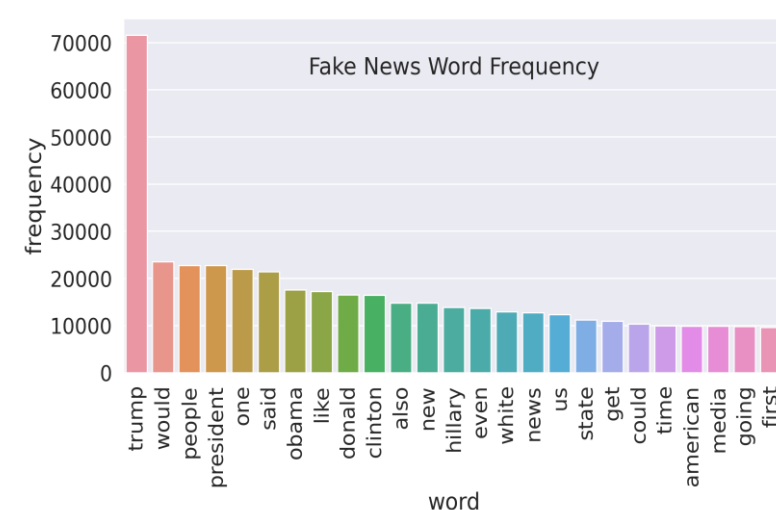
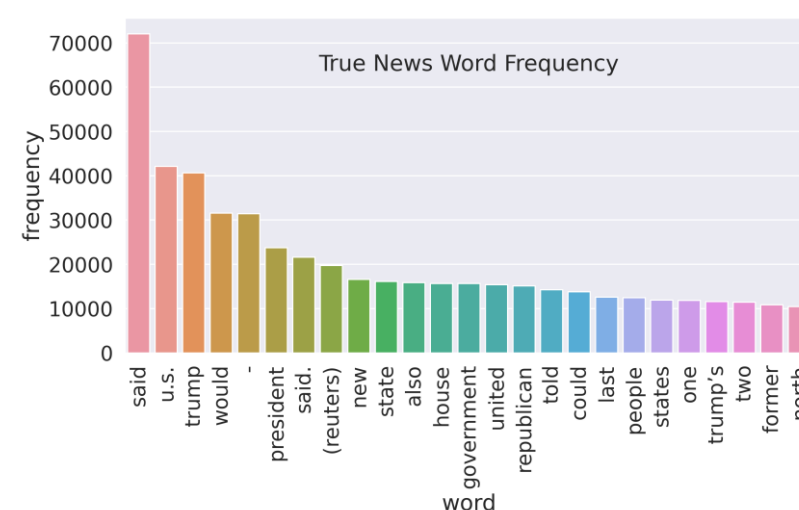
3. 어떤 단어들이 분포되어 있는지 확인해봅시다.

- 데이터 전처리(토큰화, 소문자처리, 불용어)를 합니다.

```
import nltk
from nltk.corpus import stopwords # NLTK 라이브러리의 불용어 활용
```

4. 단어들의 빈도를 계산해봅시다.

```
word_frequency_fake = dict()
for text in data['text_cleaned'][data['Label']=='Fake']:
    for word in text:
        word_frequency_fake[word] = word_frequency_fake.get(word, 0) + 1
sorted_fake = sorted(word_frequency_fake.items(), key=lambda x: x[1], reverse=True)
```



5. 다음 호에서는 전처리 및 형태소 분석 등을 적용해보겠습니다.

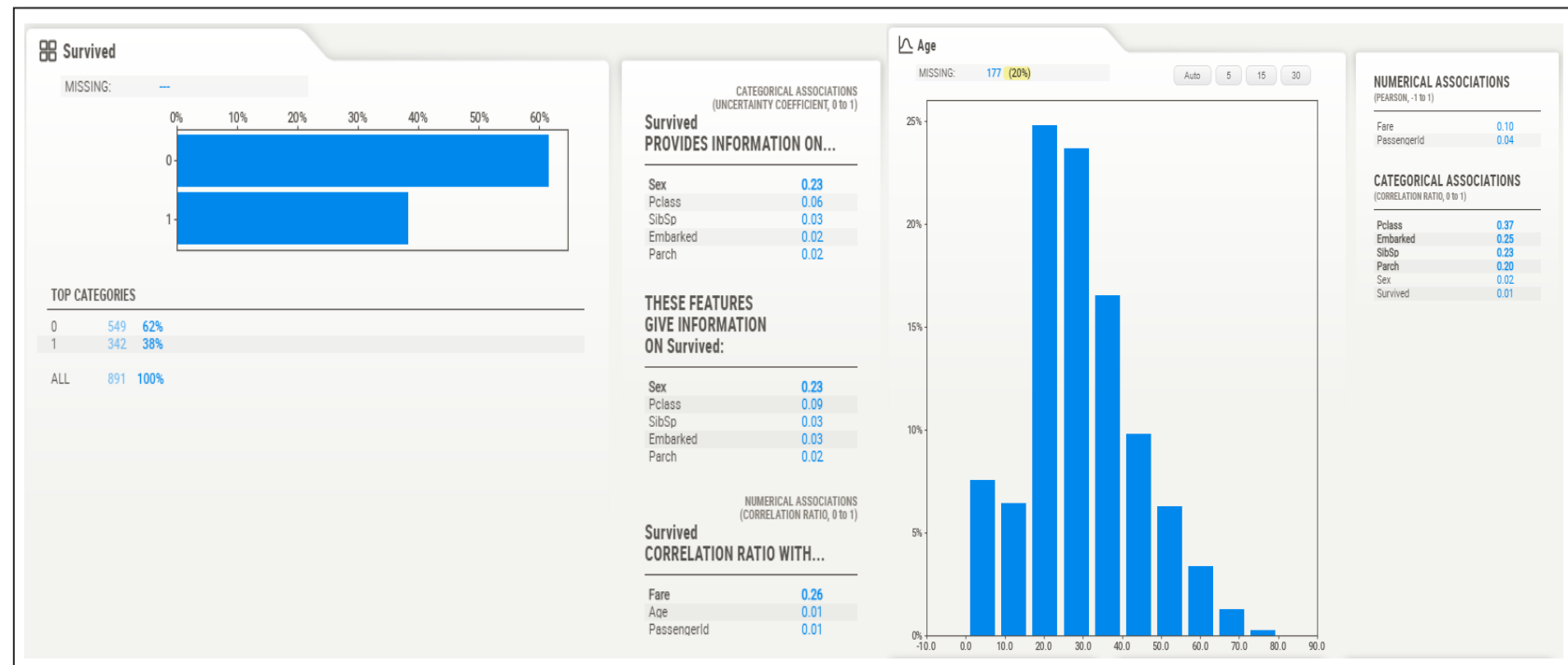
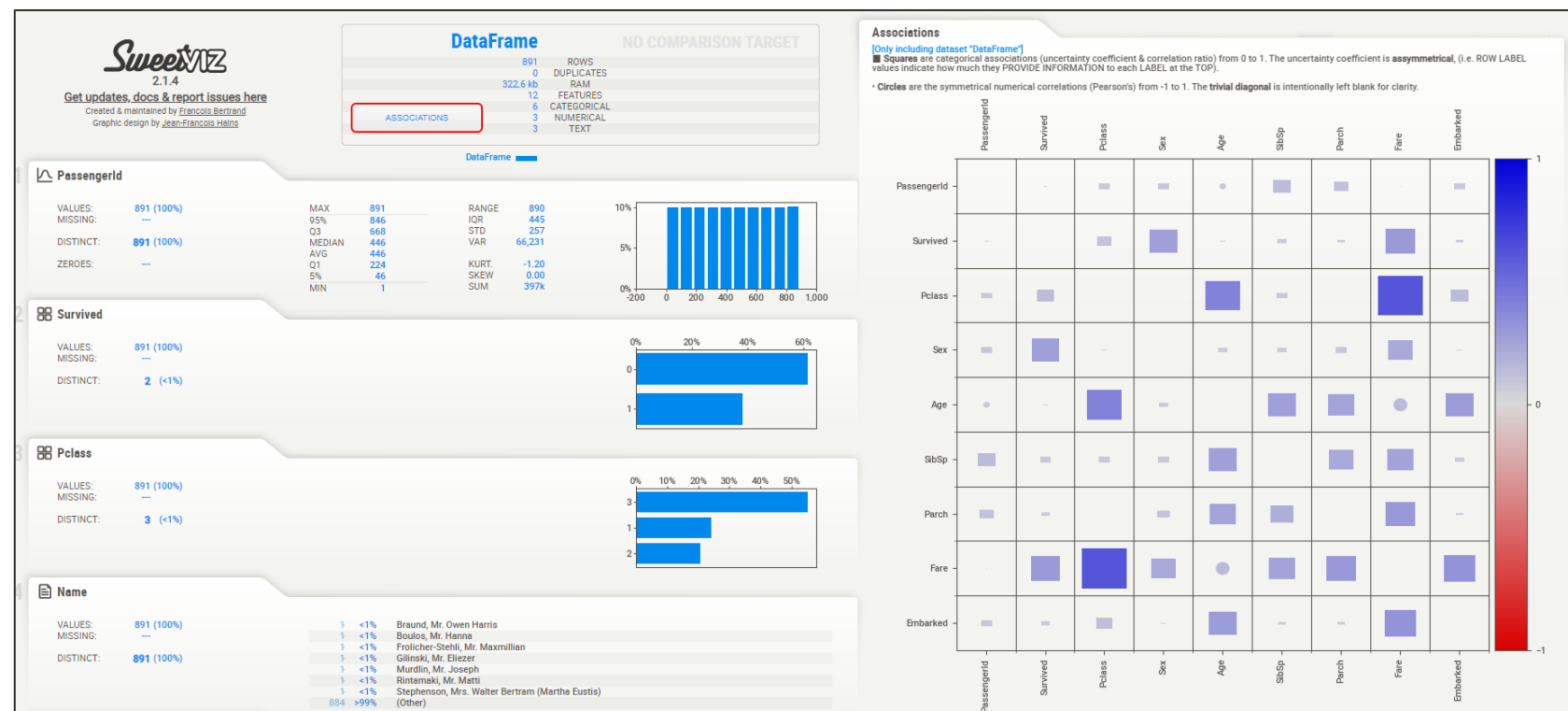
TIPs

Sweetviz

pandas.profile과 유사한 EDA 라이브러리 입니다.

```
!pip install sweetviz
data = pd.read_csv('titanic.csv')
data_report = sweetviz.analyze(data)
data_report.show_html('report_titanic.html', open_browser=True)
```

생성된 html을 통해 전체 변수에 대한 결측치와 기본 통계량 등을 확인할 수 있습니다. 각 자료를 클릭하면 보다 상세한 내용이 제공됩니다.



지도학습

□ 지도학습은 정답이 있는 데이터를 활용합니다.
- 입력값(X)과 그에 대한 정답(레이블, y)으로 학습시킵니다.

□ 지도학습 종류로는 회귀와 분류가 있습니다.
- 회귀는 정답 값이 연속적일 때 사용하고,
분류는 범주형 자료로 구분할 때 사용합니다.

□ 공부시간에 따른 성적 예측 유형 예시입니다.

- A. 공부시간에 따른 시험점수 예측 (회귀)
- B. 공부시간에 따른 시험합격 예측 (이중분류)
- C. 공부시간에 따른 학점 예측 (다중분류)

	입력	결과		
	공부시간	회귀	이진분류	다중분류
		시험점수	합격/불합격	학점
훈련	2시간	30점	불합격	D
	5시간	72점	불합격	C

	6시간	83점	합격	B
	8시간	95점	합격	A
테스트	7시간 공부하면?	점수 예측	합불 예측	학점 예측

zip(*iterable)

zip은 동일한 개수로 이루어진 데이터들을 묶어서 반환하는 함수입니다.
일반적으로 2개 이상의 리스트로 반복문을 만들 때 유용합니다.

```
# 이름과 성적 리스트
list_name = ['준우', '나리', '서연', '성민']
list_score = [89, 95, 76, 92]
for name, score in zip(list_name, list_score):
    print(f'{name}의 성적은 {score}점입니다.')
```

```
# OUTPUT
준우의 성적은 89점입니다.
나리의 성적은 95점입니다.
서연의 성적은 76점입니다.
성민의 성적은 92점입니다.
```


K-water AI Lab.

제2회 K-water AI 경진대회 성료

- AI경진대회를 성황리에 마쳤습니다👏 이번 대회는 농촌, 도시, 산업지역의 배수지 용수수요량을 예측하는 과제였습니다.
- 총 396 팀이 참가하였고 치열한 경쟁 끝에 4개 팀이 수상했습니다.



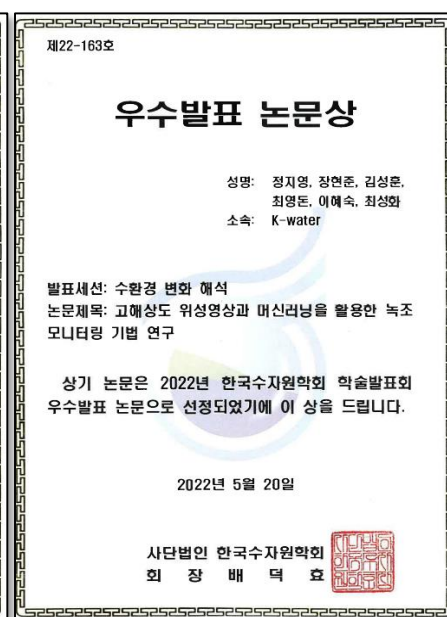
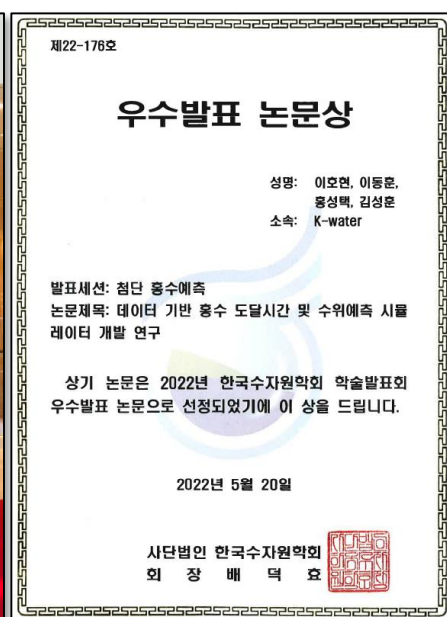
K-water연구원-육군분석평가단 MOU 체결

- 지난 12월 AI연구센터와 육군분석평가단이 상호 협력 및 지원을 위해 MOU를 체결했습니다.
- 이번 교류를 통해 AI, 빅데이터 등 디지털 기술을 공유하며 시너지 효과를 낼 것을 기대합니다.



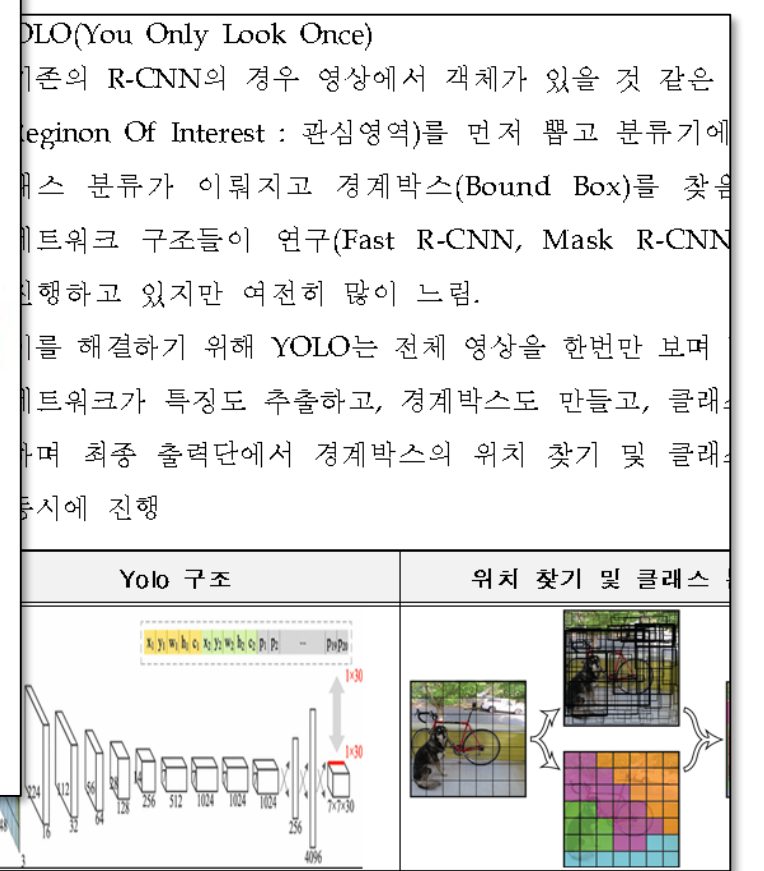
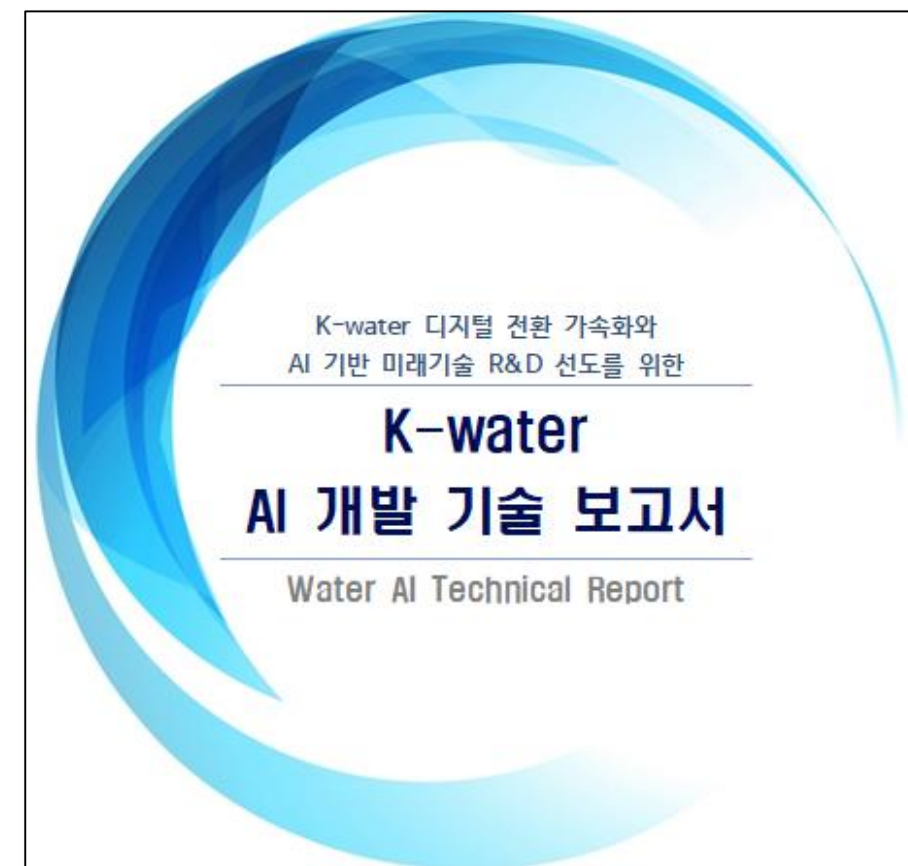
2022년 분야별 학술대회 참석

- 작년 말, 2022년 대한환경공학회에 참가해 그간의 연구성과를 발표했습니다.
- 또한 수자원학회 학술대회에서 논문 우수상 2건을 수상하였습니다.



AI 기술보고서 발간

- AI연구센터는 디지털 전환 가속화와 AI 기반 미래기술 R&D 선도를 위해 「K-water AI 기술 보고서」 발간을 준비하고 있습니다.
- AI기술의 일반사항, 물산업분야 AI기술 동향, AI모델 개발을 위한 가이드라인을 비롯하여 12개의 실무사례를 담았습니다.



문의

주경원 선임 (7840)
이소령 사원 (7341)

깃허브 & 홈페이지

