

```
In [28]: import os
import googleapiclient.discovery
import pandas as pd
from konlpy.tag import Okt

import jpype
from collections import Counter
import numpy as np
from wordcloud import WordCloud
```

```
In [29]: import matplotlib.pyplot as plt

### 폰트 설정 라이브러리
from matplotlib import font_manager, rc

### 폰트 설정
plt.rc("font", family="Malgun Gothic")

### 마이너스기호 설정
plt.rcParams["axes.unicode_minus"] = False
```

```
In [30]: api_service_name = "youtube"
api_version = "v3"
DEVELOPER_KEY="AIzaSyCGHrcNmSoUKoRGosiIShFLW9o69acZwi4"

youtube = googleapiclient.discovery.build(
    api_service_name, api_version, developerKey=DEVELOPER_KEY)

request = youtube.commentThreads().list(
    part="snippet",
    videoId="5E9zOGX3oMo",
    maxResults=500
)

comments = []

# Execute the request.
response = request.execute()

# Get the comments from the response.
for item in response['items']:
    comment = item['snippet']['topLevelComment']['snippet']
    public = item['snippet']['isPublic']
    comments.append([
        comment['authorDisplayName'],
        comment['publishedAt'],
        comment['likeCount'],
        comment['textOriginal'],
        public
    ])

while (1 == 1):
    try:
        nextPageToken = response['nextPageToken']
    except KeyError:
        break
    nextPageToken = response['nextPageToken']
```

```

# Create a new request object with the next page token.
nextRequest = youtube.commentThreads().list(part="snippet", videoId="5E9zOGX3oMo", n
# Execute the next request.
response = nextRequest.execute()
# Get the comments from the next response.
for item in response['items']:
    comment = item['snippet']['topLevelComment']['snippet']
    public = item['snippet']['isPublic']
    comments.append([
        comment['authorDisplayName'],
        comment['publishedAt'],
        comment['likeCount'],
        comment['textOriginal'],
        public
    ])

df = pd.DataFrame(comments, columns=['author', 'updated_at', 'like_count', 'text', 'pub
df.info()

```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 167 entries, 0 to 166
Data columns (total 5 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   author      167 non-null    object
1   updated_at  167 non-null    object
2   like_count  167 non-null    int64
3   text        167 non-null    object
4   public      167 non-null    bool
dtypes: bool(1), int64(1), object(3)
memory usage: 5.5+ KB

```

In [31]: df

Out[31]:

	author	updated_at	like_count	text	public
0	@user-fr5ed3uh7e	2023-11-19T12:28:28Z	0	남는 전기로 수소에너지를 만들면 안되는 거? 정부 관계자들 대갈통 좀 잘 굴려봐라...	True
1	@SHKim-uv4wo	2023-09-14T11:37:56Z	0	공정부와 전력마피아들은 신재생에너지를 싫어합니다. 신재생에너지를 통한 수소 생성으로...	True
2	@user-zw8xy5zh8r	2023-09-07T06:53:45Z	0	문재인이 큰일했네 태양광 많이 해쳐먹었지? 거제대우조선소에도 몇번다녀가더니 망해버렸...	True
3	@user-hx8do8oq7y	2023-08-26T12:36:01Z	0	전라도에 인구도 없는데 전라도에 물뽕지원하니 전기가 남아돌지 송전시설도 없구만	True
4	@-iyagi.com-	2023-08-13T11:06:19Z	0	태양광발전소 출력제어 한다면서 \nLNG, 화력 발전소는 왜 계속 짓는건데	True
...	...	...	...	...	...
162	@user-lr3bq7mw9l	2023-04-20T08:52:52Z	17	남는 전기로 그린수소 만들어라!	True
163	@user-br6jr7zg4z	2023-04-20T08:47:40Z	30	산자부는 알고 있으면서 계속 미루다 미루다 터진거지나요? 알지만 일을 안한다. 누가...	True
164	@83nobleman	2023-04-20T08:46:36Z	29	re100 부족 하다던데 빨리 시설을 갖추길	True
165	@user-tt9nt2dt4t	2023-04-20T08:45:33Z	42	이 뉴스의 문제점은.. 무작정 전력망 확충만 이야기 한다면 점...\n중간에 그래프...	True
166	@jjk3264	2023-04-20T08:42:47Z	12	문재인판 바다이야기	True

167 rows × 5 columns

```
In [32]: max_like_count = df['like_count'].max()
max_like_text = df[df['like_count'] == max_like_count]['text'].values[0]
print(max_like_text)
```

RE100 인증 필요한 산업에 전력을 최대한 보내줘야지.....

```
In [33]: max_like_index = df[df['like_count'] == max_like_count].index[0]
print(max_like_index)
```

125

```
In [34]: df['like_count'].max()
```

Out[34]: 52

```
In [35]: save_path="./data/all_comments.csv"
df.to_csv(save_path, index=False)
```

```
In [36]: save_path="./data/all_comments.csv"
df1=pd.read_csv(save_path)
```

```
In [37]: okt = Okt()
okt
```

Out[37]: <konlpy.tag.\_Okt.Okt at 0x2bc447effd0>

```
In [38]: all_comment = []

for cmt in df["text"] :
    # print(oka.nouns(cmt))
    ### extend() : 리스트에 값만 추출하여 확장해서 추가하는 방식
    # - append() : 리스트에 형태(type) 자체를 추가하는 방식
    all_comment.extend(oka.nouns(cmt))

# print(all_comment)
```

```
In [39]: all_comment2 = [w for w in all_comment if len(w) > 1]
# all_comment2
```

```
In [40]: all_comment_count = Counter(all_comment2)
# print(all_comment_count)
```

```
In [41]: all_top_70 = {}
for k, v in all_comment_count.most_common(70):
    all_top_70[k] = v

all_top_70 = {k:v for k, v in all_comment_count.most_common(70)}
# all_top_70
```

```
In [42]: # Assuming all_top_50 is a dictionary with words as keys and counts as values

# Create a color map
cmap = plt.get_cmap('viridis')

# Extract values and keys
values = list(all_top_70.values())
keys = list(all_top_70.keys())

# Normalize values to use in colormap
norm = plt.Normalize(min(values), max(values))

# Create a figure and axis
plt.figure(figsize=(20, 10))

# Title
plt.title("all 리뷰의 단어 상위 (50개) 빈도 시각화", fontsize=17)

# Bar graph with gradient color
for key, value in zip(keys, values):
    # Additional condition to skip "태양광" or "전기"
    if key == "태양광" or key == "전기":
        continue

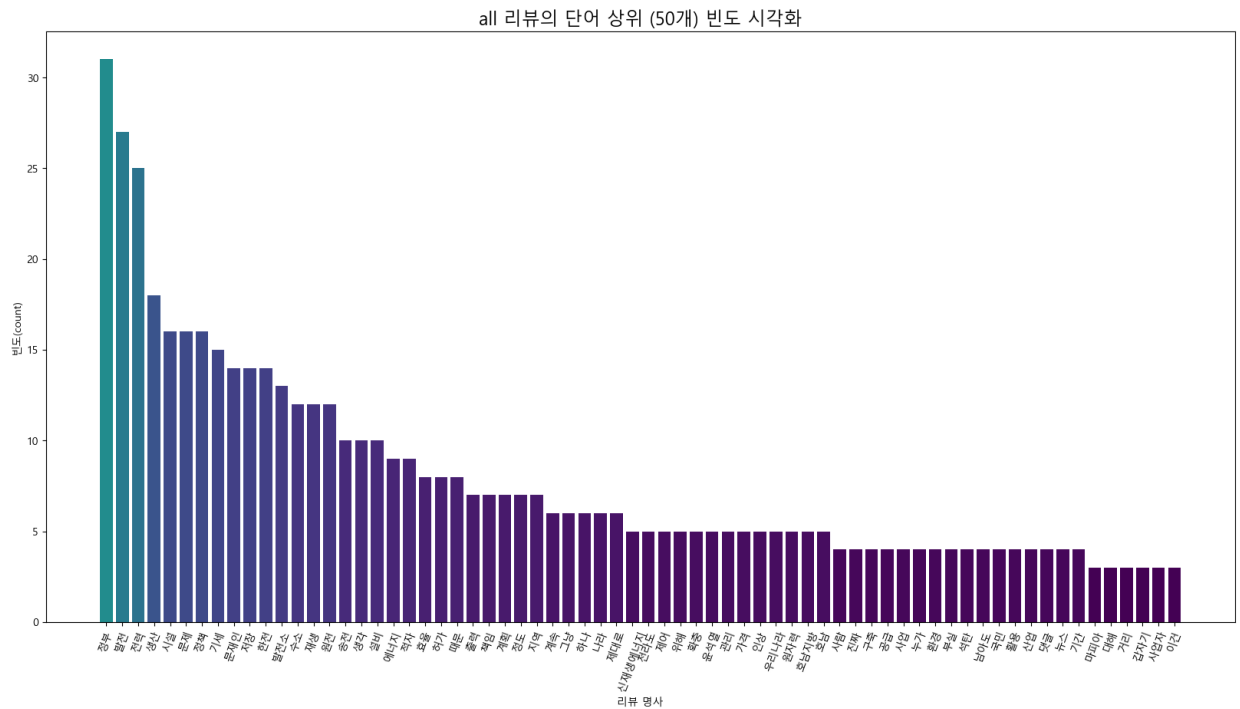
    color = cmap(norm(value))
    plt.bar(key, value, color=color)

# x-axis and y-axis Labels
plt.xlabel("리뷰 명사")
plt.ylabel("빈도(count)")

# Adjust x-axis Labels rotation
```

```
plt.xticks(rotation=70)
```

```
# Show the graph
plt.show()
```



```
In [43]: # all_top_70
```

```
In [44]: ### 긍정 리뷰 단어 워드클라우드 시각화
plt.figure(figsize=(8, 8))

### 그래프 제목
plt.title(" 리뷰 단어 워드클라우드 시각화")

### 사용할 폰트 파일 지정하기
font_path = "C:/Windows/Fonts/malguns1.ttf"

### 워드클라우드 그래프 속성 설정
wc = WordCloud(
    ### 폰트 지정
    font_path=font_path,
    ### 배경색 지정
    background_color="ivory",
    ### 그래프 너비
    width=800,
    ### 그래프 높이
    height=600
)

### 워드클라우드 그래프에 데이터 넣기
# - generate_from_frequencies() : 워드클라우드 이미지로 반환해줌
# cloud = wc.generate_from_frequencies(pos_top_20)

### 긍정 전체 단어 넣어보기
cloud = wc.generate_from_frequencies(all_top_70)

### 워드클라우드 이미지 보여주기
plt.imshow(cloud)
```



```
# Display WordCloud image
plt.imshow(cloud)
plt.axis("off")
plt.savefig("./img/리뷰_단어_워드클라우드_시각화.png")
plt.show()
```



```
In [47]: # filtered_top_70
```

```
In [48]: """
Analysis in English:
```

```
"1. Key words such as 'government', 'generation', 'power', 'production', 'facilities',
2. Government and Generation Policy: Words like 'government' and 'policy'
3. Words like 'generation', 'power', 'production', 'energy', etc., emerge
4. Energy Generation Facilities and Storage Technology: Words like 'facil
5. Regional Issues: Words like 'Jeolla-do', 'Honam region', 'Honam', etc.
6. Environment and Issues: The word 'environment' along with words like '
7. Other Issues: Words like 'price', 'increase', 'efficiency', 'permit',
8. Through the above analysis, it is evident that the provided word cloud
"""
```

```
Out[48]: '\nAnalysis in English:\n\n"1. Key words such as \'government\', \'generation\', \'power\', \'production\', \'facilities\', \'issues\', \'policy\', \'momentum\', \'Moon Jae-in\', \'storage\', etc., represent topics related to energy generation in Korea.\n\n"2. Government and Generation Policy: Words like \'government\' and \'policy\' have high frequencies, and the mention of \'Moon Jae-in\' suggests opinions related to the role of the Korean government and policy formulation for energy generation.\n\n"3. Words like \'generation\', \'power\', \'production\', \'energy\', etc., emerged as significant keywords related to energy generation in Korea, indicating the country\'s focus and importance on energy generation and power production.\n\n"4. Energy Generation Facilities and Storage Technology: Words like \'facilities\', \'power plant\', \'hydrogen\', \'regeneration\', \'nuclear power\', \'transmission\', \'facilities\', \'renewable energy\', etc., indicate keywords related to energy generation facilities. The mention of \'storage\' reflects an interest in energy storage technology.\n\n"5. Regional Issues: Words like \'Jeolla-do\', \'Honam region\', \'Honam\', etc., represent issues related to Korea\'s regional characteristics. There is a possibility that issues related to energy generation in specific regions are mentioned.\n\n"6. Environment and Issues: The word \'environment\' along with words like \'issues\', \'inadequate\', \'coal\', \'pollution\', etc., reflects concerns and issues related to environmental problems and inadequate policies in the context of energy generation.\n\n"7. Other Issues: Words like \'price\', \'increase\', \'efficiency\', \'permit\', \'plan\', \'deficit\', etc., indicate issues related to the economic aspects of energy generation. Terms like price, efficiency, and permits showcase the relevance between energy policy and economic activities.\n\n"8. Through the above analysis, it is evident that the provided word cloud reflects various topics and issues related to energy generation in Korea. This analysis can help understand opinions and issues related to energy policy formulation and serve as a basis for more effective policies and improvement strategies."\n'
```

```
In [49]:
```

```
"""
```

```
분석 결과:
```

1. 주요 단어들인 '정부', '발전', '전력', '생산', '시설', '문제', '정책', '기서
2. 정부와 발전 정책: '정부'와 '정책'이라는 단어가 빈도수가 높게 나타나며, '문
3. '발전', '전력', '생산', '에너지' 등의 단어들은 한국의 에너지 발전과 관련된
4. 에너지 발전 시설과 저장 기술: '시설', '발전소', '수소', '재생', '원전', '송
5. 지역과 관련된 이슈: '전라도', '호남지방', '호남' 등의 단어들은 한국의 지역적
6. 환경과 문제: '환경'이라는 단어와 함께 '문제', '부실', '석탄', '오염' 등의 단
7. 기타 이슈: '가격', '인상', '효율', '허가', '계획', '적자' 등의 단어들은 에
8. 위의 분석을 통해 주어진 워드 클라우드 결과에는 한국의 에너지 발전과 관련된 디

```
"""
```



Out[49]:

"\n분석 결과:\n

1. 주요 단어들인 '정부', '발전', '전력', '생산', '시설', '문제', '정책', '기세', '문재인', '저장' 등은 한국의 에너지 발전과 관련된 주제들을 나타내고 있습니다.\n
2. 정부와 발전 정책: '정부'와 '정책'이라는 단어들이 빈도수가 높게 나타나며, '문재인'이라는 인물도 언급되었습니다. 이는 한국 정부의 역할과 에너지 발전에 대한 정책 수립과 관련된 의견들을 나타냅니다.\n
3. '발전', '전력', '생산', '에너지' 등의 단어들은 한국의 에너지 발전과 관련된 주요 키워드로 나타났습니다. 이는 한국에서 에너지 발전과 전력 생산에 대한 관심과 중요성을 보여줍니다.\n
4. 에너지 발전 시설과 저장 기술: '시설', '발전소', '수소', '재생', '원전', '송전', '설비', '신재생에너지' 등의 단어들은 에너지 발전 시설과 관련된 키워드로 나타났습니다. '저장'이라는 단어는 에너지 저장 기술에 대한 관심을 보여줍니다.\n
5. 지역과 관련된 이슈: '전라도', '호남지방', '호남' 등의 단어들은 한국의 지역적인 특성과 관련된 이슈를 나타냅니다. 지역별 에너지 발전과 관련된 문제들이 언급되었을 가능성이 있습니다.\n
6. 환경과 문제: '환경'이라는 단어와 함께 '문제', '부실', '석탄', '오염' 등의 단어들이 나타났습니다. 이는 에너지 발전과 관련하여 환경 문제와 부실한 정책 등에 대한 우려와 이슈를 반영합니다.\n
7. 기타 이슈: '가격', '인상', '효율', '허가', '계획', '적자' 등의 단어들은 에너지 발전과 관련된 경제적인 측면과 이슈들을 나타냅니다. 가격, 효율, 허가 등의 단어들은 에너지 정책과 경제 활동 사이의 관련성을 보여줍니다.\n
8. 위의 분석을 통해 주어진 워드 클라우드 결과에는 한국의 에너지 발전과 관련된 다양한 주제와 이슈들이 반영되어 있음을 알 수 있습니다. 이러한 분석을 통해 에너지 정책 수립과 관련된 의견 및 이슈들을 파악할 수 있으며, 이를 바탕으로 보다 효과적인 정책과 개선 방안을 모색할 수 있습니다.\n"

In [4]:

```
import os
print(os.environ['PATH'])
```

```
C:\Users\sarak\anaconda3\envs\gj_env_01;C:\Users\sarak\anaconda3\envs\gj_env_01\Library\mingw-w64\bin;C:\Users\sarak\anaconda3\envs\gj_env_01\Library\usr\bin;C:\Users\sarak\anaconda3\envs\gj_env_01\Library\bin;C:\Users\sarak\anaconda3\envs\gj_env_01\Scripts;C:\Users\sarak\anaconda3\envs\gj_env_01\bin;C:\Users\sarak\anaconda3\condabin;C:\Windows\system32;C:\Windows;C:\Windows\System32\Wbem;C:\Windows\System32\WindowsPowerShell\v1.0;C:\Windows\System32\OpenSSH;C:\Program Files (x86)\NVIDIA Corporation\PhysX\Common;C:\Program Files\NVIDIA Corporation\NVIDIA NvDLISR;C:\ProgramData\chocolatey\bin;C:\Program Files\nodejs;C:\Program Files\Bandizip;C:\Program Files\ojdkbuild\java-11-openjdk-11.0.15-1\bin;C:\Program Files\Git\cmd;C:\Users\sarak\AppData\Local\Programs\Python\Python38-32\Scripts;C:\Users\sarak\AppData\Local\Programs\Python\Python38-32;C:\Program Files\MySQL\MySQL Shell 8.0\bin;C:\Users\sarak\anaconda3;C:\Users\sarak\anaconda3\Library\mingw-w64\bin;C:\Users\sarak\anaconda3\Library\usr\bin;C:\Users\sarak\anaconda3\Library\bin;C:\Users\sarak\anaconda3\Scripts;C:\Users\sarak\AppData\Local\Programs\Python\Launcher;C:\Users\sarak\AppData\Local\Microsoft\WindowsApps;C:\Users\sarak\AppData\Local\Programs\Microsoft VS Code\bin;C:\Users\sarak\AppData\Roaming\npm;C:\Program Files\JetBrains\IntelliJ IDEA 2023.2.2\bin;. ;C:\Program Files\JetBrains\IntelliJ IDEA Community Edition 2023.2.2\bin;. ;C:\Program Files\JetBrains\DataSpell 2023.2.4\bin;. ;C:\Users\sarak\AppData\Local\GitHubDesktop\bin;.
```

In [ ]: