



트럼프 대통령 트윗 분석

요구사항 정리

| ≡ 기능 구분 | Aa 요구 사항 제목 | ≡ 요구사항 상세 내용 |
|------------|-----------------------|---|
| 라이브러리 설치 | <u>필요 라이브러리 가져오기</u> | 데이터 분석에 필요한 기본 라이브러리 : pandas, numpy 시각화에 필요한 라이브러리 : matplotlib, seaborn, wordcloud 결측치 파악에 필요한 라이브러리 : missingno 경고문 제외에 필요한 라이브러리 : warnings 문장 전치리에 필요한 라이브러리 : re, nltk, Counter |
| 문제 정의 | <u>주제 선정 및 가설 설정</u> | 트럼프 전 대통령의 트윗 데이터를 활용하여 분석할 주제 및 가설 설정 |
| 데이터 수집 | <u>데이터 가져오기</u> | 캐글 : https://www.kaggle.com/datasets/austinreese/trump-tweets?select=trumptweets.csv 에서 trumptweets.csv 데이터를 가져온 후 살펴보기 |
| 데이터 전처리 | <u>결측치 확인하기</u> | missingno를 활용하여 간단하게 결측치 관련 시각화를 한 후 결측치가 많은 변수는 우선 제거해준다. |
| 데이터 전처리 | <u>필요 변수 추출</u> | 이번 프로젝트에 필요하다고 생각하는 변수들을 정한 후 그에 맞는 데이터 만들기 |
| 데이터 전처리 | <u>date 변수 분리</u> | 트윗 게시물 수의 증감을 살펴보기 위해 date 변수를 연도, 월, 시간으로 분리(split 함수 이용) |
| 데이터 시각화 | <u>date 관련 변수 시각화</u> | 연도 별, 월 별, 시간 별 게시물 수를 시각화 하여 증감 추세 살펴보기 |
| 데이터 시각화 | <u>재임 전 데이터 시각화</u> | 2013 ~ 2016년 데이터를 추출한 후 좋아요 수 중앙값을 구한 뒤, 이상인 게시물과 미만인 게시물을 분리한 후 content 변수 전처리 후 많이 언급된 단어 top20과 wordcloud를 통해 언급된 단어들을 살펴보자 |
| 데이터 시각화 | <u>재임 기간 데이터 시각화</u> | 2017 이후 데이터를 추출한 후 좋아요 수 중앙값을 구한 뒤, 이상인 게시물과 미만인 게시물을 분리한 후 content 변수 전처리 후 많이 언급된 단어 top20과 wordcloud를 통해 언급된 단어들을 살펴보자 |
| 가설 검증 및 해석 | <u>시각화 자료를 통한 결과</u> | 재임 전 / 재임 기간 별 좋아요 수 중앙값에 따른 언급된 단어 빈도수에 대한 결과 해석 |
| 분석 의견 | <u>분석 결과에 대한 의견</u> | 분석 결과에 대한 개인적인 의견 제시 |
| 분석 의견 | <u>향후 계획</u> | 향후 데이터를 활용한 다른 방향성 제시 |

목차

- 1. 라이브러리 설치
 - 1.1 필요 라이브러리 가져오기
- 2. 문제 정의
 - 2.1 주제 선정 및 가설 설정
- 3. 데이터 수집
 - 3.1 데이터 가져오기
- 4. 데이터 전처리
 - 4.1 결측치 확인하기
 - 4.2 필요 변수 추출
 - 4.3 date 변수 분리
- 5. 데이터 시각화
 - 5.1 date 관련 변수 시각화
 - 5.2 재임 전 데이터 시각화
 - 5.3 재임 기간 데이터 시각화
- 6. 가설 검증 및 해석
 - 6.1 시각화 자료를 통한 결과
- 7. 분석 의견
 - 7.1 분석 결과에 대한 의견
 - 7.2 향후 계획

1. 라이브러리 설치

1.1 필요 라이브러리 가져오기

```
# 데이터 분석에 필요한 라이브러리
import pandas as pd
import numpy as np

# 시각화에 필요한 라이브러리
import matplotlib.pyplot as plt
import seaborn as sns
from wordcloud import WordCloud, STOPWORDS

# 결측치 파악에 필요한 라이브러리
import missingno as msno

# 경고문 제외에 필요한 라이브러리
import warnings
warnings.filterwarnings('ignore')

# 문장 전처리에 필요한 라이브러리
import re
from collections import Counter
import nltk
from nltk.corpus import stopwords
nltk.download('punkt')
nltk.download('averaged_perceptron_tagger')
nltk.download('wordnet')
nltk.download('omw-1.4')
nltk.download('stopwords')
```

2. 문제 정의

2.1 주제 선정 및 가설 설정

1. 트럼프가 재임 전 4년(2013 ~ 2016) 동안 올린 게시물의 좋아요 수 중앙값을 기준으로 이상인 게시물과 미만인 게시물에서 각각 많이 언급한 단어들을 통해 대통령이 되기 전 즉, 선거 기간에 언급해야 하는 것, 언급하지 말아야 할 것을 알아보자.
2. 재임 기간 동안 올린 게시물의 좋아요 수 중앙값을 기준으로 이상인 게시물과 미만인 게시물에서 각각 많이 언급한 단어들을 분석하여 트럼프 대통령에 대한 국민들의 인식을 파악한 후 국민들에게 좋은 이미지를 주기 위해 언급해야 할 것과 언급하지 말아야 할 것이 무엇인지 파악해보자.
3. 두 시기의 트럼프가 올린 게시물에서 언급된 단어들의 양상이 다를 것으로 예상된다. 재임 전 좋아요 수가 중앙값 미만인 게시물에서 언급한 단어들은 재임 기간 동안에는 대체로 언급하지 않았을 것 같다.

3. 데이터 수집

3.1 데이터 가져오기

```
df_trump = pd.read_csv('/content/drive/MyDrive/data/trumptweets.csv')
df_trump.head()
```

| | id | link | content | date | retweets | favorites | mentions | hashtags | geo |
|---|------------|---|---|---------------------|----------|-----------|----------|----------|-----|
| 0 | 1698308935 | https://twitter.com/realDonaldTrump/status/169... | Be sure to tune in and watch Donald Trump on L... | 2009-05-04 20:54:25 | 500 | 868 | NaN | NaN | NaN |
| 1 | 1701461182 | https://twitter.com/realDonaldTrump/status/170... | Donald Trump will be appearing on The View tom... | 2009-05-05 03:00:10 | 33 | 273 | NaN | NaN | NaN |
| 2 | 1737479987 | https://twitter.com/realDonaldTrump/status/173... | Donald Trump reads Top Ten Financial Tips on L... | 2009-05-08 15:38:08 | 12 | 18 | NaN | NaN | NaN |
| 3 | 1741160716 | https://twitter.com/realDonaldTrump/status/174... | New Blog Post: Celebrity Apprentice Finale and... | 2009-05-08 22:40:15 | 11 | 24 | NaN | NaN | NaN |
| 4 | 1773561338 | https://twitter.com/realDonaldTrump/status/177... | "My persona will never be that of a wallflower... | 2009-05-12 16:07:28 | 1399 | 1965 | NaN | NaN | NaN |

```
# 데이터 행/열 수 확인
df_trump.shape
```

(41122, 9)

```
# 데이터 변수 별 정보 확인
df_trump.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 41122 entries, 0 to 41121
Data columns (total 9 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   id           41122 non-null  int64
1   link         41122 non-null  object
2   content      41122 non-null  object
3   date         41122 non-null  object
4   retweets     41122 non-null  int64
5   favorites    41122 non-null  int64
6   mentions     22467 non-null  object
7   hashtags     5810 non-null   object
8   geo          0 non-null      float64
dtypes: float64(1), int64(3), object(5)
memory usage: 2.8+ MB

```

- 총 41122개의 데이터로 구성되어 있고 9개의 컬럼으로 이루어져 있다.
- mentions, hashtags, geo 변수에 결측치가 있다.
- 변수 설명
 - id : 임의의 id
 - link : 트위터 링크
 - content : 트위터 게시물
 - date : 게시 날짜
 - retweets : 리트윗 횟수
 - favorites : 게시물 좋아요 수
 - mentions : 게시물에 언급한 내용
 - hashtags : 게시물에 해시태크한 내용
 - geo : ??

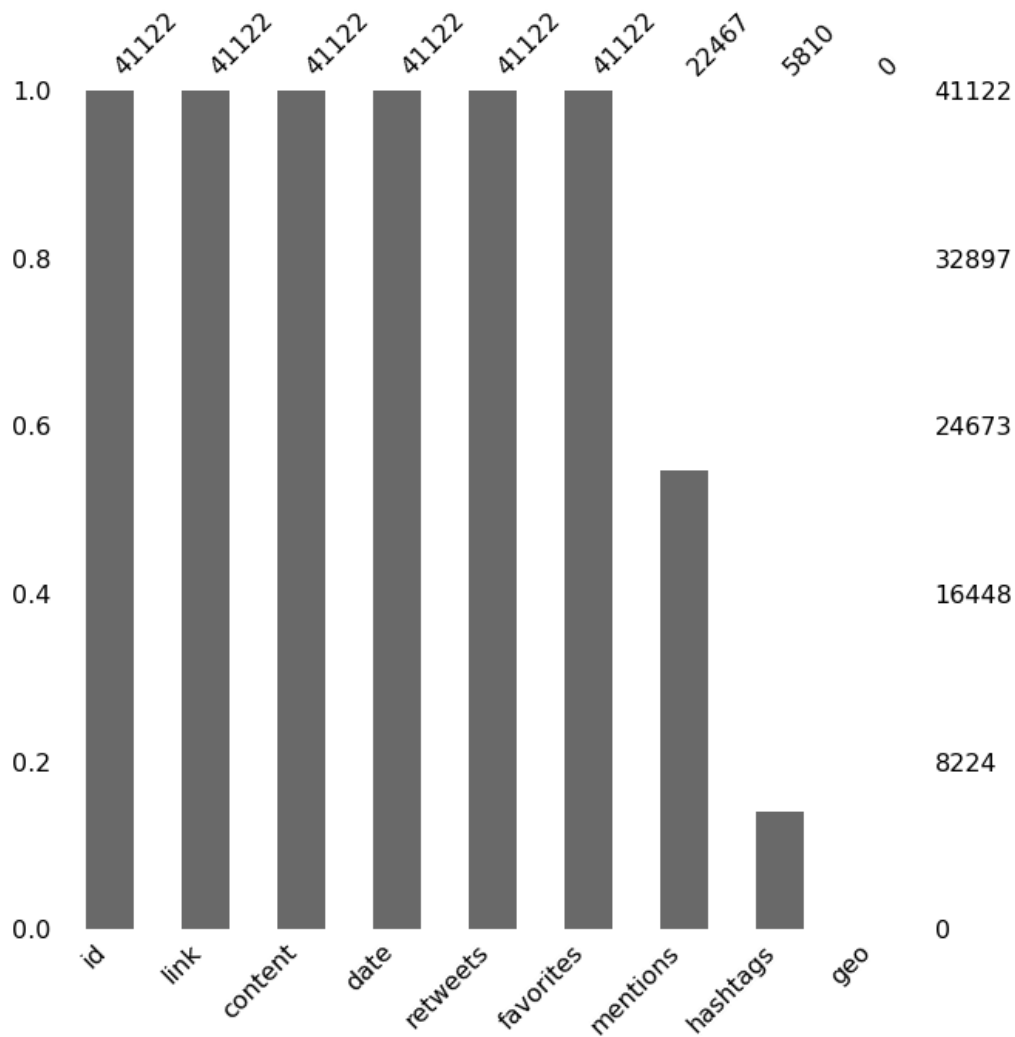
4. 데이터 전처리

4.1 결측치 확인하기

```

# 결측치 관련 시각화
msno.bar(df_trump, figsize=(10,10))
plt.show()

```



- mentions, hashtags, geo 변수는 결측치가 많기 때문에 제거해주는 것이 좋아보인다.

4.2 필요 변수 추출

- 필요 변수 - content, date, retweets, favorites
- 결측치가 많이 있는 mentions, hashtags, geo 변수는 제외한다.
- 임의로 정해진 id와 트위터 링크를 의미하는 link 변수도 제외한다.
 - 링크를 통해 얻을 수 있는 시각화나 인사이트가 없다고 판단

```
# 기존 데이터에서 필요한 변수만 포함한 데이터 만들기
df_trump_new = df_trump.loc[:, 'content': 'favorites']
df_trump_new
```

| | content | date | retweets | favorites |
|-------|---|---------------------|----------|-----------|
| 0 | Be sure to tune in and watch Donald Trump on L... | 2009-05-04 20:54:25 | 500 | 868 |
| 1 | Donald Trump will be appearing on The View tom... | 2009-05-05 03:00:10 | 33 | 273 |
| 2 | Donald Trump reads Top Ten Financial Tips on L... | 2009-05-08 15:38:08 | 12 | 18 |
| 3 | New Blog Post: Celebrity Apprentice Finale and... | 2009-05-08 22:40:15 | 11 | 24 |
| 4 | "My persona will never be that of a wallflower... | 2009-05-12 16:07:28 | 1399 | 1965 |
| ... | ... | ... | ... | ... |
| 41117 | I have never seen the Republican Party as Stro... | 2020-01-19 19:24:52 | 32620 | 213817 |
| 41118 | Now Mini Mike Bloomberg is critical of Jack Wi... | 2020-01-19 22:12:20 | 36239 | 149571 |
| 41119 | I was thrilled to be back in the Great State o... | 2020-01-20 01:27:07 | 16588 | 66944 |
| 41120 | "In the House, the President got less due proc... | 2020-01-20 02:16:00 | 20599 | 81921 |
| 41121 | A great show! Check it out tonight at 9pm. @ F... | 2020-01-20 02:57:49 | 7947 | 34902 |

41122 rows × 4 columns

4.3 date 변수 분리

```
# date 변수 분리

df_trump_new['year'] = df_trump_new['date'].str.split('-').str[0] # 연도
df_trump_new['month'] = df_trump_new['date'].str.split('-').str[1] # 월
df_trump_new['hour'] = df_trump_new['date'].str.split(' ').str[1].str.split(':')[0] # 시간

df_trump_new
```

| | content | date | retweets | favorites | year | month | hour |
|-------|---|---------------------|----------|-----------|------|-------|------|
| 0 | Be sure to tune in and watch Donald Trump on L... | 2009-05-04 20:54:25 | 500 | 868 | 2009 | 05 | 20 |
| 1 | Donald Trump will be appearing on The View tom... | 2009-05-05 03:00:10 | 33 | 273 | 2009 | 05 | 03 |
| 2 | Donald Trump reads Top Ten Financial Tips on L... | 2009-05-08 15:38:08 | 12 | 18 | 2009 | 05 | 15 |
| 3 | New Blog Post: Celebrity Apprentice Finale and... | 2009-05-08 22:40:15 | 11 | 24 | 2009 | 05 | 22 |
| 4 | "My persona will never be that of a wallflower... | 2009-05-12 16:07:28 | 1399 | 1965 | 2009 | 05 | 16 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 41117 | I have never seen the Republican Party as Stro... | 2020-01-19 19:24:52 | 32620 | 213817 | 2020 | 01 | 19 |
| 41118 | Now Mini Mike Bloomberg is critical of Jack Wi... | 2020-01-19 22:12:20 | 36239 | 149571 | 2020 | 01 | 22 |
| 41119 | I was thrilled to be back in the Great State o... | 2020-01-20 01:27:07 | 16588 | 66944 | 2020 | 01 | 01 |
| 41120 | "In the House, the President got less due proc... | 2020-01-20 02:16:00 | 20599 | 81921 | 2020 | 01 | 02 |
| 41121 | A great show! Check it out tonight at 9pm. @ F... | 2020-01-20 02:57:49 | 7947 | 34902 | 2020 | 01 | 02 |

41122 rows × 7 columns

5. 데이터 시각화

5.1 date 관련 변수 시각화

- 연도, 월, 시간 별 게시물 수 보기

```
# 연도
df_trump_new['year'].value_counts().sort_index()

# 월
df_trump_new['month'].value_counts().sort_index()

# 시간
df_trump_new['hour'].value_counts().sort_index()
```

```
2009      56
2010     143
2011     870
2012    4190
2013    8207
2014    5987
2015    7719
2016    3951
2017    2232
2018    2998
2019    4588
2020     181
Name: year, dtype: int64
```

```
01      3864
02      2978
03      2898
04      3438
05      3317
06      3149
07      3676
08      3148
09      3410
10      4508
11      3339
12      3397
Name: month, dtype: int64
```

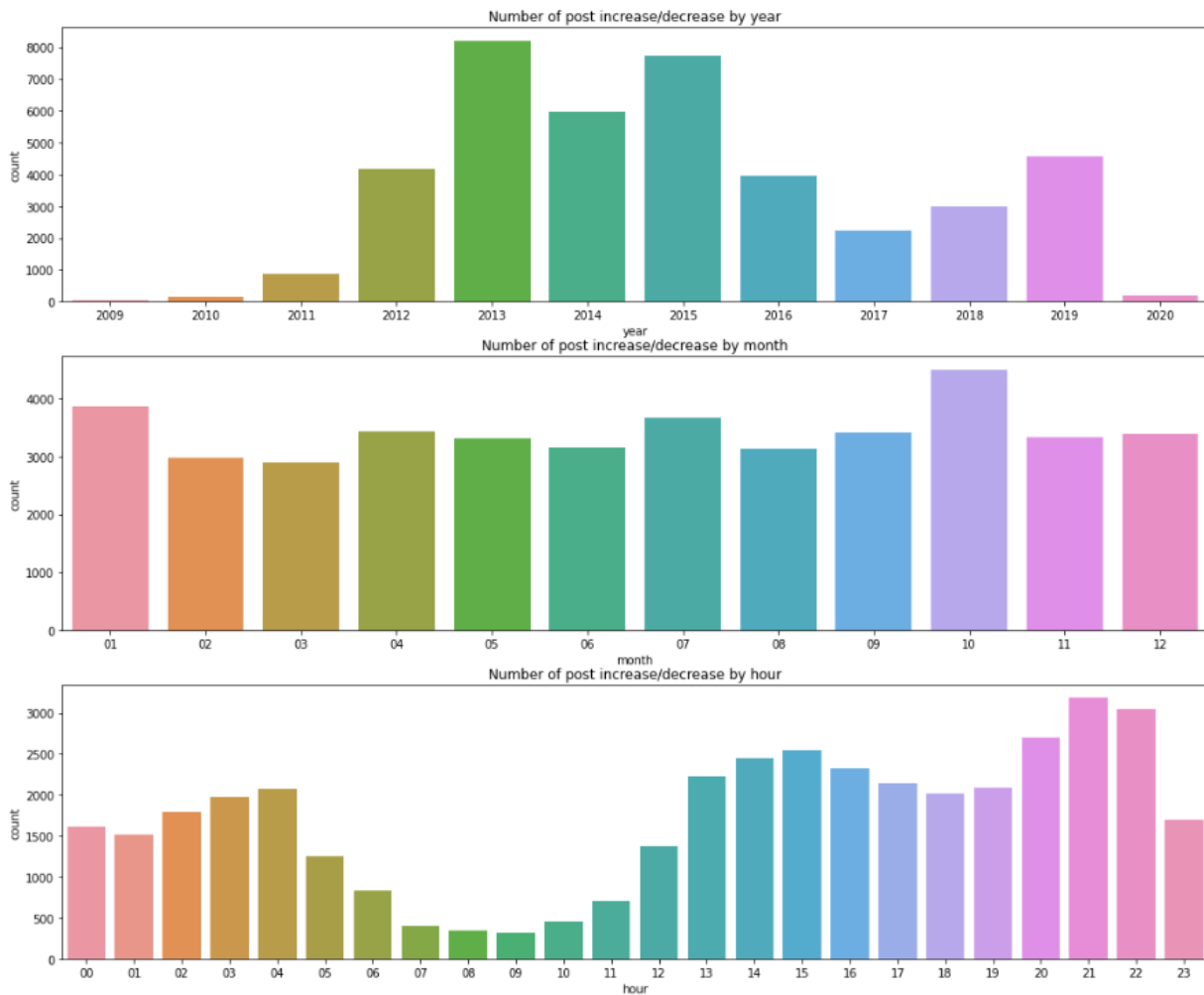
```
00      1620
01      1522
02      1795
03      1974
04      2076
05      1258
06       831
07       400
08       347
09       321
10       458
11       706
12      1377
13      2226
14      2448
15      2544
16      2319
17      2146
18      2025
19      2095
20      2696
21      3184
22      3050
23      1704
Name: hour, dtype: int64
```

- 시각화

```
f, ax = plt.subplots(3, 1, figsize = (18,15))

sns.countplot('year', data = df_trump_new, ax = ax[0])
ax[0].set_title('Number of post increase/decrease by year')
sns.countplot('month', data = df_trump_new, ax = ax[1], order = df_trump_new['month'].value_counts().sort_index().index)
ax[1].set_title('Number of post increase/decrease by month')
sns.countplot('hour', data = df_trump_new, ax = ax [2], order = df_trump_new['hour'].value_counts().sort_index().index)
ax[2].set_title('Number of post increase/decrease by hour')

plt.show()
```



1. 2013년 > 2015년 > 2014년 순으로 게시글 수가 많았다.
2. 10월 > 1월 > 7월 순으로 게시글 수가 많았다.
3. 21시 > 22시 > 20시 순으로 저녁 시간에 게시글 수가 많았다.

- 참고

- 2009년, 2020년 데이터에는 모든 월이 있지 않다.

- 2009년 → 1월 ~ 4월 x
- 2020년 → 2월 ~ 12월 x

```
# 2009년 데이터 확인
df_2009 = df_trump_new[df_trump_new['year'] == '2009']
df_2009['month'].value_counts()

# 2020년 데이터 확인
df_2020 = df_trump_new[df_trump_new['year'] == '2020']
df_2020['month'].value_counts()
```

```
05    21
06    11
08     7
07     5
10     4
09     3
11     3
12     2
Name: month, dtype: int64

, 01    181
   Name: month, dtype: int64
```

5.2 재임 전 데이터 시각화

- 필요 데이터 추출 및 좋아요 수 확인

```
# 2013 ~ 2016년 데이터 추출
df_1316 = df_trump_new[(df_trump_new['year'] >= '2013') & (df_trump_new['year'] <= '2016')]
df_1316
```

| | content | date | retweets | favorites | year | month | hour |
|-------|---|---------------------|----------|-----------|------|-------|------|
| 5259 | Republicans are making a mistake if they don't... | 2013-01-01 00:46:57 | 160 | 49 | 2013 | 01 | 00 |
| 5260 | Republicans have all the cards...but they just... | 2013-01-01 00:48:04 | 177 | 81 | 2013 | 01 | 00 |
| 5261 | Happy New Year everyone...it will be an intere... | 2013-01-01 00:49:20 | 522 | 402 | 2013 | 01 | 00 |
| 5262 | Obama and the Democrats are laughing at the de... | 2013-01-01 16:10:46 | 467 | 100 | 2013 | 01 | 16 |
| 5263 | I am a Republican...but the Republicans may be... | 2013-01-01 16:13:14 | 831 | 498 | 2013 | 01 | 16 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 31118 | Great move on delay (by V. Putin) - I always k... | 2016-12-30 20:41:33 | 29104 | 83039 | 2016 | 12 | 20 |
| 31119 | Join @ American32, founded by Hall of Fame leg... | 2016-12-30 20:46:55 | 6201 | 21635 | 2016 | 12 | 20 |
| 31120 | Russians are playing @ CNN and @ NBCNews for s... | 2016-12-30 23:18:18 | 19319 | 72360 | 2016 | 12 | 23 |
| 31121 | Happy New Year to all, including to my many en... | 2016-12-31 14:17:21 | 121041 | 314786 | 2016 | 12 | 14 |
| 31122 | Happy Birthday @ DonaldJTrumpJr! https://insta... | 2016-12-31 19:58:12 | 8082 | 48667 | 2016 | 12 | 19 |

25864 rows × 7 columns

```
# 좋아요 중앙값 구하기
df_1316['favorites'].median()
```

76.0

```
# 좋아요 수 76이상인 게시물 데이터
df_1316_good = df_1316[df_1316['favorites'] >= 76]

# 좋아요 수 76 미만인 게시물 데이터
df_1316_bad = df_1316[df_1316['favorites'] < 76]
```

• content 변수 전처리 및 빈도 분석

```
# content 변수 전처리해주는 함수 선언

def content_preprocessing(df):
    content = ''
    for text in df['content']:
        content = content + ' ' + text

    filtered_content = content.lower() # 대소문자 통일
    word_tokens = nltk.word_tokenize(filtered_content) # 단어 단위로 잘라주기
    token_pos = nltk.pos_tag(word_tokens) # 품사 찾아주기

    # 명사 단어만 추출하기
    NN_words = []

    for word, pos in token_pos:
        if 'NN' in pos: # NN도 여러개가 있기 때문
            NN_words.append(word)

    # 명사는 단수형으로 변환
    wlem = nltk.WordNetLemmatizer()
    lemmatized_words = []

    for word in NN_words:
        new_word = wlem.lemmatize(word)
        lemmatized_words.append(new_word)

    # a, an, the, this, that 같은 불용어 제거
    stopwords_list = stopwords.words('english') # nltk에서 제공하는 영어 불용어사전

    unique_NN_words = set(lemmatized_words)
    final_NN_words = lemmatized_words

    # 불용어 제거
    for word in unique_NN_words:
        if word in stopwords_list:
            while word in final_NN_words : final_NN_words.remove(word)

    # 추가로 제거하고 싶은 단어 제거
    customized_stopwords = ['@', 'http', '...', "'", '"', '%', '"', '....', '-']
    unique_NN_words1 = set(final_NN_words)
    for word in unique_NN_words1:
        if word in customized_stopwords:
            while word in final_NN_words: final_NN_words.remove(word)

    return final_NN_words
```

```
list_1316_good = content_preprocessing(df_1316_good) # 좋아요 수가 76개 이상인 게시물에서 나온 명사 리스트
list_1316_bad = content_preprocessing(df_1316_bad) # 좋아요 수가 76개 미만인 게시물에서 나온 명사 리스트
```

```
# 빈도 top 20 - 좋아요 76개 이상
```

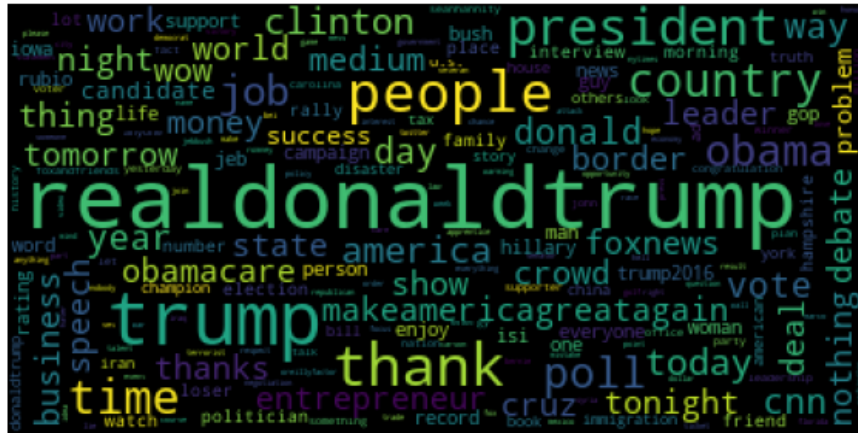
```
c_1316_good = Counter(list_1316_good)
c_1316_good.most_common(20)
```

```
[('realdonaldtrump', 2098),
 ('trump', 1560),
 ('thank', 890),
 ('people', 820),
 ('president', 710),
 ('time', 583),
 ('country', 515),
 ('obama', 473),
 ('poll', 471),
 ('job', 442),
 ('clinton', 330),
 ('today', 318),
 ('america', 316),
 ('year', 293),
 ('donald', 290),
 ('way', 276),
 ('vote', 272),
 ('cnn', 270),
 ('day', 264),
 ('makeamericagreatagain', 264)]
```

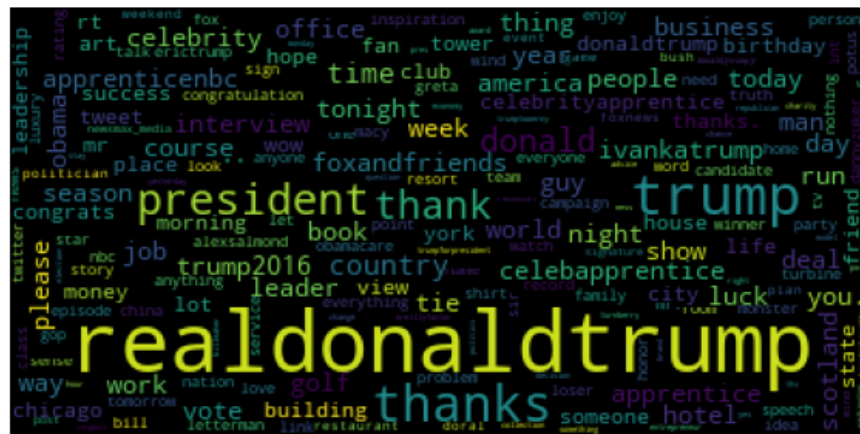
```
# 빈도 top 20 - 좋아요 76개 미만
c_1316_bad = Counter(list_1316_bad)
c_1316_bad.most_common(20)
```

```
[('realdonaldtrump', 5831),
 ('trump', 1869),
 ('thanks', 1438),
 ('president', 1016),
 ('thank', 790),
 ('donald', 532),
 ('country', 423),
 ('time', 396),
 ('course', 364),
 ('show', 351),
 ('people', 337),
 ('celebapprentice', 328),
 ('golf', 312),
 ('man', 300),
 ('apprenticenbc', 287),
 ('apprentice', 280),
 ('job', 267),
 ('please', 259),
 ('year', 251),
 ('hotel', 231)]
```

```
wordcloud = WordCloud(max_font_size=40, relative_scaling=.5).generate_from_frequencies(c_1316_good)
plt.figure(figsize = (10, 10))
plt.imshow(wordcloud)
plt.axis("off")
plt.show()
```



```
wordcloud = WordCloud(max_font_size=40, relative_scaling=.5).generate_from_frequencies(c_1316_bad)
plt.figure(figsize = (10, 10))
plt.imshow(wordcloud)
plt.axis("off")
plt.show()
```



- 필요 데이터 추출 및 좋아요 수 확인

```
# 2017년 이후 데이터 추출
df_17 = df_trump_new[df_trump_new['year'] >= '2017']
df_17
```

| | content | date | retweets | favorites | year | month | hour |
|-------|---|---------------------|----------|-----------|------|-------|------|
| 31123 | TO ALL AMERICANS- # HappyNewYear & many blessi... | 2017-01-01 06:00:10 | 28278 | 113734 | 2017 | 01 | 06 |
| 31124 | Well, the New Year begins. We will, together, ... | 2017-01-02 15:40:10 | 25093 | 110594 | 2017 | 01 | 15 |
| 31125 | Chicago murder rate is record setting - 4,331 ... | 2017-01-02 18:31:17 | 14785 | 55607 | 2017 | 01 | 18 |
| 31126 | @ CNN just released a book called "Unprecedent... | 2017-01-02 19:32:29 | 3359 | 12002 | 2017 | 01 | 19 |
| 31127 | Various media outlets and pundits say that I t... | 2017-01-02 19:37:10 | 7707 | 41498 | 2017 | 01 | 19 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 41117 | I have never seen the Republican Party as Stro... | 2020-01-19 19:24:52 | 32620 | 213817 | 2020 | 01 | 19 |
| 41118 | Now Mini Mike Bloomberg is critical of Jack Wi... | 2020-01-19 22:12:20 | 36239 | 149571 | 2020 | 01 | 22 |
| 41119 | I was thrilled to be back in the Great State o... | 2020-01-20 01:27:07 | 16588 | 66944 | 2020 | 01 | 01 |
| 41120 | "In the House, the President got less due proc... | 2020-01-20 02:16:00 | 20599 | 81921 | 2020 | 01 | 02 |
| 41121 | A great show! Check it out tonight at 9pm. @ F... | 2020-01-20 02:57:49 | 7947 | 34902 | 2020 | 01 | 02 |

9999 rows × 7 columns

```
# 좋아요 중앙값 구하기
df_17['favorites'].median()
```

74963.0

```
# 좋아요 수 74963이상인 게시물 데이터
df_17_good = df_17[df_17['favorites'] >= 74963]

# 좋아요 수 74963 미만인 게시물 데이터
df_17_bad = df_17[df_17['favorites'] < 74963]
```

• content 변수 전처리 및 빈도 분석

```
list_17_good = content_preprocessing(df_17_good) # 좋아요 수가 74963개 이상인 게시물에서 나온 명사 리스트
list_17_bad = content_preprocessing(df_17_bad) # 좋아요 수가 74963개 미만인 게시물에서 나온 명사 리스트
```

```
# 빈도 top 20 - 좋아요 74963개 이상

c_17_good = Counter(list_17_good)
c_17_good.most_common(20)
```

```
[('people', 770),
 ('country', 719),
 ('democrat', 659),
 ('president', 560),
 ('news', 556),
 ('year', 504),
 ('border', 486),
 ('time', 433),
 ('state', 429),
 ('job', 341),
 ('medium', 318),
 ('trump', 285),
 ('wall', 261),
 ('nothing', 259),
 ('thank', 253),
 ('deal', 230),
 ('election', 225),
 ('house', 222),
 ('crime', 216),
 ('day', 214)]
```

```
# 빈도 top 20 - 좋아요 74963개 미만
c_17_bad = Counter(list_17_bad)
c_17_bad.most_common(20)
```

```
[('president', 432),
 ('people', 390),
 ('job', 388),
 ('thank', 378),
 ('democrat', 353),
 ('country', 336),
 ('state', 326),
 ('trump', 325),
 ('today', 316),
 ('year', 268),
 ('time', 251),
 ('news', 243),
 ('border', 239),
 ('vote', 234),
 ('tax', 219),
 ('crime', 177),
 ('house', 174),
 ('honor', 167),
 ('deal', 166),
 ('day', 163)]
```

```
# wordcloud - 좋아요 74963개 이상

wordcloud = WordCloud(max_font_size=40, relative_scaling=.5).generate_from_frequencies(c_17_good)
plt.figure(figsize = (10, 10))
plt.imshow(wordcloud)
plt.axis("off")
plt.show()
```


- 재임 전에는 전체적으로 자신의 트위터 계정(realdonaldtrump)을 자주 언급하며 활발한 sns활동을 하고 있음을 보여주었지만 '좋아요'가 적은 게시물에 더 많이 언급되어 sns활동이 선거 활동에 좋은 영향을 준 것 같지는 않다.
 - '좋아요'가 많은 게시물 : 2098 언급
 - '좋아요'가 적은 게시물 : 5831 언급
- 재임 전 obama 대통령, 대통령 선거 후보였던 clinton을 언급한 글, 투표 관련 글(poll, vote), 자신의 슬로건(make america great again)을 언급한 글에 '좋아요' 수가 많았다.
- 재임 전 선거와 관련 없는 글(course, show, golf, hotel), 초보자, 초심과 같이 자신이 정치를 잘 알지 못한다는 의미의 글(celebapprentice, apprenticenbc, apprentice), 성 차별적인 단어(man)를 언급한 글에 '좋아요' 수가 적었다.
- 재임 기간 중에는 언론에 대한 글(medium), 미국-멕시코 국경 장벽 관련 글(wall), 대통령 선거에 출마할 것이라는 글(election)에 '좋아요' 수가 많았다.
- 재임 기간 중 투표 관련 글(vote), 세금과 관련된 글(tax), 명예와 관련된 글(honor)에 '좋아요' 수가 적었다.
- 재임 전에는 감사함에 대한 말(thank)을 언급을 자주 했는데 재임 기간에는 빈도 수가 조금 하락 하였다.

7. 분석 의견

7.1 분석 결과에 대한 의견

- 대통령이 되기 전에는 현 대통령과 이번 대통령 선거에 출마할 대통령 후보에 대한 언급(trump에 대한 언급)을 하는 것이 좋아 보인다.
- 또한, 선거와 관련된 말을 많이 하면서 대통령 선거에 관심이 있다는 것을 보여주고 민주당을 대표하고 사람들을 모을 수 있는 슬로건을 하나 만드는 것이 좋아 보인다.
- 정치에 대해서도 트럼프와는 다르게 경험이 많다는 것을 보여주는 것이 좋아 보인다.
- trump 대통령이 미국-멕시코 국경 장벽 공약을 지킴으로써 반응이 좋았는데 우리 또한 사람들에게 만족을 줄 수 있는 공약을 내세우고 지킬 수 있는 방안을 찾아야 할 것 같다.
 - 트럼프 대통령이 세금 관련해서 좋지 않은 평을 받고 있는데 세금 조정에 대한 방안을 모색하면 좋을 것 같다.

7.2 향후 계획

- retweets 변수도 활용
 - 데이터를 분할할 때 favorites 변수 외에도 retweets 변수로 분할 하거나 두 변수의 합을 구한 후 분할하여 이번 프로젝트와의 차이를 찾아보자.
- 형용사 분석
 - 단순히 명사에 관한 분석을 하였는데 트럼프 대통령이 그 명사에 대해 어떤 감정을 가지고 언급할 한 것인지 까지 분석해보자.
- 카테고리 분할
 - 월 별, 시간 별 게시물에 언급된 단어를 살펴보고 영향이 있는지 분석해보자.
 - 정치, 경제, 사회 등 언급한 단어 별로 카테고리를 만들어 관련 글끼리 묶어준 후 트윗 반응을 분석해보자.