import pandas as pd  
import re  
from konlpy.tag import Okt  
from collections import Counter  
import openpyxl  
  
#데이터가 들어있는 엑셀파일 읽어오기  
df = pd.read\_excel("C:/Users/Dong-In Kang/PycharmProjects/koala/Lookie/"+keyword+".xlsx",sheet\_name= 'Rawdata')  
  
#데이터파일에서 contents열에 있는 정보를 message에 저장  
message =df['contents']  
  
total\_message = ""  
  
#message에 저장된 데이터들을 total\_message에 모두 저장  
for m in message:  
 total\_message += str(m).strip()  
# total\_message에서 숫자제거  
total\_message = re.sub('[0-9]','',total\_message)  
print('number erased')  
# total\_message에서 영문자제거  
total\_message = re.sub('z[A-Za-]+','',total\_message)  
print('English erased')  
# total\_message에서 특수문자제거  
total\_message = re.sub('[-=+,#/\?:^$.@\*\"※~&%ㆍ·!』\\‘’|\(\)\[\]\<\>`\'…》]','',total\_message)  
print('signs erased')  
  
#결과를 저장한 엑셀파일 열기  
wb = openpyxl.Workbook()  
sheet =wb.active  
okt=Okt()  
#total\_message에 있는 데이터들을 형태소분석 (pos : 품사태깅// 예를들어 ("코딩","Noun") 형식으로 반환)  
morphs = okt.pos(total\_message)  
  
Final\_word = []  
  
#불용어설정  
stopwords=[]  
  
#명사,동사,형용사,해시태그에 해당하는 단어 중에서 불용어가 아닌 단어만 Final\_word에 저장  
for word, pos in morphs:  
 if pos in ['Noun', 'Verb', 'Adjective', 'Hashtag']:  
 if word not in stopwords:  
 Final\_word.append(word)  
#Final\_word에 저장되어있는 각 단어들의 개수 계산 및 저장  
count2 = Counter(Final\_word)  
for v in count2:  
 if count2[v]!=1:  
 sheet.append([v,count2[v]])  
wb.save(keyword+"\_selected\_frequency.xlsx")