

전자담배 현안 조사



전자담배의 금연 기여도, 중독성, 부작용을 중심으로

컴퓨터학과 20151024 윤지윤
정보통계학과 20161767 박보경
정보통계학과 20171037 주혜림
정보통계학과 20171013 박리나

START

Index

1. 문제 정의 및 전략 수립

2. 프로젝트 일정

3. 데이터 수집

4. 데이터 전처리

5. 데이터 분석 및 결과 적용

6. 한계점 & Discussion



Python, Jupyter Notebook

BeautifulSoup

Naver 검색 API

Selenium Webdriver

TF-IDF, LDA, Topic Modeling

Word Cloud

역할 분담



성명	담당 업무
윤지윤	최종 발표 준비 데이터 수집 (웹 크롤링 및 구글 폼 설문조사) 데이터 분석 및 결과 적용
박보경	최종 발표 준비 데이터 수집 (구글 폼 설문조사 문항 작성) 웹 크롤링 수집 데이터 전처리 데이터 분석 및 결과 적용
주혜림	중간 발표 및 최종 발표 준비 데이터 수집 (구글 폼 설문조사) 웹 크롤링 수집 데이터 전처리 데이터 분석 및 결과 적용
박리나	중간 발표 준비 데이터 수집 (구글 폼 설문조사) 웹 크롤링 수집 데이터 전처리

1. 문제 정의



상업적으로 판매되는 전자담배의 구성성분은 무엇인가

전자담배가 건강에 미치는 영향은 무엇인가

금연을 도와주는가, 아니면 단지 이중사용자를 양산할 뿐인가?

1. 문제 정의

전자담배의 위험과 이익에 대한 갑론을박

정보 제공 미흡에 따른 소비자들의 혼란 증가



{ 전자담배에 대한 정확한 실태 파악 필요 }

1. 전략 수립

전자담배의 금연 기여도

전자담배의 중독성

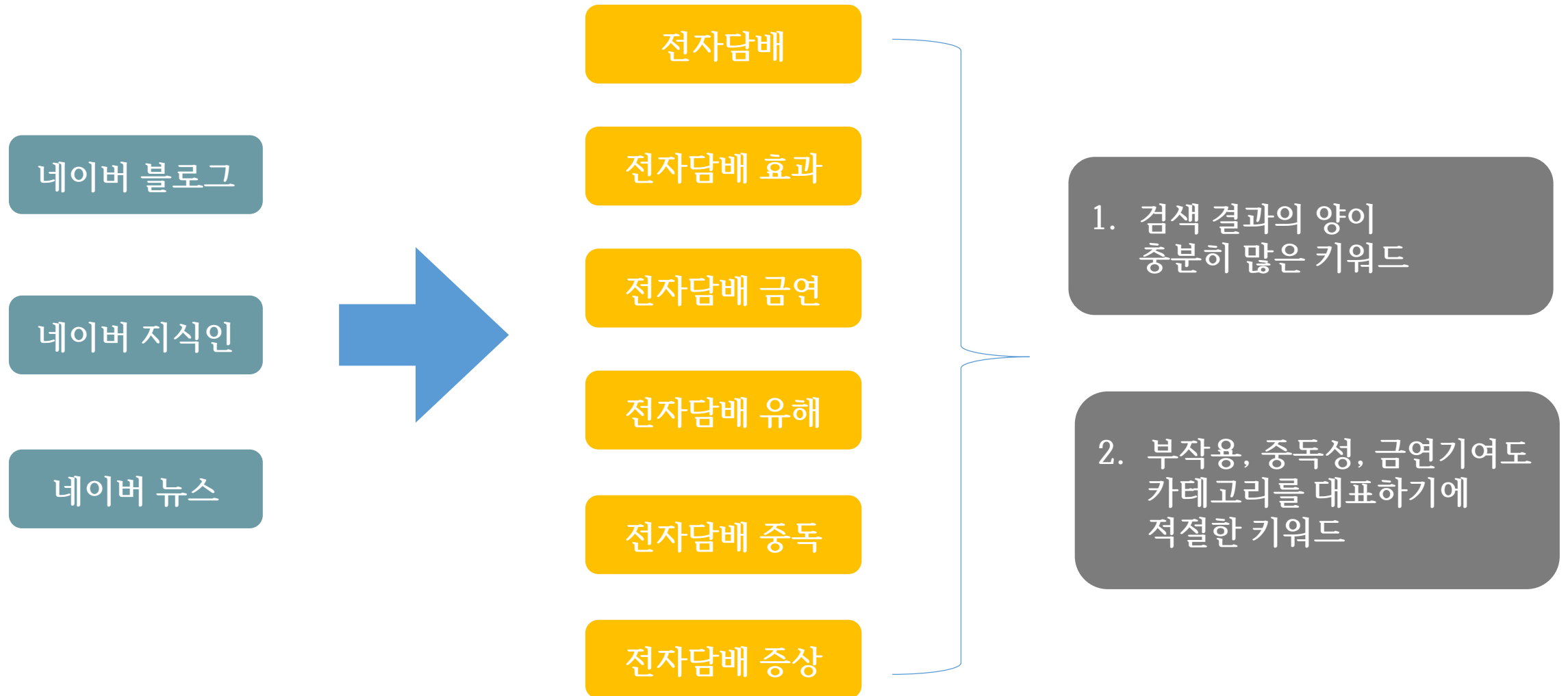
전자담배의 부작용

전자담배에 대한 정확한 정보 제공

2. 프로젝트 일정

구분	수행내용	3월			4월				5월				6월	
		3주	4주	5주	1주	2주	3주	4주	1주	2주	3주	4주	1주	2주
문제정의	전자담배 주제 선정	03.14 ~ 03.21												
전략수립	데이터 분석 방법 논의		03.22 ~ 03.28											
중간 발표	중간발표 준비			03.29 ~ 04.03										
	중간발표			04.04										
데이터 수집	텍스트 크롤링 코드 작성				04.05 ~ 05.07									
	네이버 텍스트 데이터 수집								05.07 ~ 05.16					
	설문조사 문항 작성								05.07 ~ 05.16					
	설문조사 시행									05.16 ~ 05.27				
데이터 분석	텍스트 데이터 전처리									05.16 ~ 05.27				
	분석 모델 구현												05.28 ~ 06.03	
결과적용	분석 결과 해석 및 적용												05.28 ~ 06.03	
최종 발표	최종발표 준비												06.04 ~ 06.06	
	최종발표												06.04 ~ 06.06	

3. 데이터 수집



3. 데이터 수집

네이버 블로그 크롤링 (키워드 검색 결과 전체 텍스트) : TXT 파일	
전자담배	992개
전자담배 금연	975개
전자담배 유해	948개
전자담배 중독	954개
전자담배 증상	944개
전자담배 효과	953개
네이버 지식인 크롤링 (키워드 검색 결과 전체 텍스트) : TXT 파일	
전자담배	903개
전자담배 금연	946개
전자담배 유해	912개
전자담배 중독	929개
전자담배 증상	921개
전자담배 효과	933개
네이버 뉴스 크롤링 진행 (키워드 검색 결과 주요 텍스트) : JSON 파일	
전자담배	1000개
전자담배 금연	1000개
전자담배 유해	1000개
전자담배 중독	1000개
전자담배 증상	761개
전자담배 효과	1000개

3. 데이터 수집

데이터 수집 과정에서의 어려움

네이버 API

검색 결과 최대 1000개까지 지원

3 ~ 5 줄의 요약 텍스트만을 지원

뉴스 검색 결과는 클릭 시 각 언론사 홈페이지로 이동

다양한 웹 페이지 접속에 따른 많은 예외 (Exception) 발생

3. 데이터 수집


NAVER 전자담배

통합검색 이미지 뉴스 쇼핑 지도 카페 블로그 어학사전 더보기 검색옵션


정렬 기간 영역 유사문서 출처 옵션유지 편집 편집 상세검색

연관검색어 ? 액상 전자담배 전자담배 juul 줄 전자담배 액상 전자담배 추천 전자담배 간접흡연 줄 전자담배 아이코스 궤련형 전자담배 아이코스3 더보기


블로그 1-10 / 320,866건



전자담배액상 쥬스박스에서 공짜로 받는 팁! 2019.05.29.
특히 액상 전자담배는 담배 냄새가 전혀 나지 않는다고 해서 바꾸게 됐어요. 저는 평소엔 귀찮은 걸 정말 싫어하는 사람이라 그냥 일반 액상 전자담배 말고 csv...
FREAKS Ishpsy.blog.me/221549466506 블로그 내 검색



전자담배액상 호불호가 없더라고요! 3일 전
어떻게 담배를 끊었는지 여쭙보기 시작했습니다. 그런데 놀랍게도 전부 전자담배를 통해서 도움을 받았다고 말씀하시더라고요. 솔직히 처음 이 이야기를 들었을...
화이트레드작스 ghdwms960.blog.me/221553649233 블로그 내 검색



전자담배액상 정말 괜찮은 곳 여기! 어제
상향 전 자 담 배를 시작해서 글로벌한 21세기 문명인답게 전 자기기를 통해... 궤련형 전 자 담 배는 담 배 잎으로 만든 담 배 스틱을 홀더에 넣고, 홀더 부분의 금속...
그 사람이 내 안에 - blog.naver.com/crossstone/221555207659 블로그 내 검색

```
{  
  "description": "뭣썩었는데 전 자 담.배 액 상을 온라인에서 한번이라도 주문해보신 분들이라  
면 디케이에서 상 품을... 줄여서 담 배를 찾게되는 횟수도 현저히 줄어들더라구요! 저처럼 처음에 금  
단증상이 생겨서 <b>전 자 담 배</b>...",  
  "link": "https://blog.naver.com/1981kyh?Redirect=Log&logNo=221537356262",  
  "title": "<b>전자담배</b>액상 정말 맛있는거 찾는다면?"  
},  
{  
  "description": "여자들도 담배를 많이 피잖아요~? 전엔 연초가 많이 보였는데 술술 친구들 사  
이에서 <b>전자담배</b>가 유행을 하는지 전담을 가지고 다니는 애들이 어느순간부터 많이 보이더라고  
요 근데 담배갈지않고 약세사리갈... ",  
  "link": "https://blog.naver.com/issusi?Redirect=Log&logNo=221536538058",  
  "title": "청 주<b>전자담배</b> 가장 많은 이곳"  
},  
{  
  "description": "바로 연초에서 <b>전자담배</b>로 갈아탔기 때문입니다. 사실 처음에는 제가  
전담으로 바꿔보는게 어떨냐고... 멀리서 보면 <b>전자담배</b>라고는 전혀 생각도 못할 만큼 예쁘게  
만들어져 있어서 놀랐습니다 기존에 나와 있는...",  
  "link": "https://blog.naver.com/doutoury3?Redirect=Log&logNo=221534103789",  
  "title": "<b>전자담배</b> 잘 골랐습니다"  
},  
{  
  "description": "스트레스 심했었는데, <b>전자담배</b> SMP0를 만난 이후에는 정말 그런 고민  
하나 없이 연초 생각도 안 나고, 연기도 금세 사라지고 이아이 이용하면서 일반 담배를 손에 안 대고  
점점 줄여서 다음번 해가 오기 전까지 ",  
  "link": "https://blog.naver.com/apple5098?Redirect=Log&logNo=221527933882",  
  "title": "<b>전자담배</b> 향기로운 냄새로 갈아타자SMP0"  
},
```

3. 데이터 수집

```
def makeLinkList(keyword):
    items = json.load(open(keyword + "_naver_kin.json", "r", encoding="utf-8"))
    links = []

    i = 1
    for item in items:
        print(str(i) + " : " + item["link"])
        links.append(item["link"])

        if i == 5:
            break

        i += 1

    return links
```

```
def textCrawling(links, keyword):
    wfile = open(os.getcwd() + "/" + keyword + "_naver_kin.txt", mode='w', encoding='utf-8')

    idx = 1
    for link in links:
        try:
            driver = webdriver.Chrome("C:/Users/comdori/Downloads/chromedriver_win32/chromedriver.exe")
            driver.get(link)

            print("\naver_kin_" + str(idx) + "\n")

            html = driver.page_source

            bs = BeautifulSoup(html, 'html.parser')
            # print(bs.prettify())

            title = bs.select("#content > div.question-content > div > div.c-heading._questionContentsArea > div.c-heading_title > div > d
            question = bs.select("#content > div.question-content > div > div.c-heading._questionContentsArea.c-heading--default-old > div.c
            answer = bs.find_all('div', {'class': 'c-heading-answer__content-user'})

            wfile.write("\naver_kin_" + str(idx) + "\n\n")

            # title
            for t in title:
                contents = t.get_text().strip()
                print(contents)
                wfile.write(contents)

            # question
            for q in question:
                contents = q.get_text().strip()
                print(contents)
                wfile.write(contents)

            # answer
            for a in answer:
                contents = a.get_text().strip()
                print(contents)
                wfile.write(contents)

            wfile.write("\n")
```

3. 데이터 수집

데이터 수집 과정에서의 어려움

네이버 API

검색 결과 최대 1000개까지 지원

3 ~ 5 줄의 요약 텍스트만을 지원

뉴스 검색 결과는 클릭 시 각 언론사 홈페이지로 이동

다양한 웹 페이지 접속에 따른 많은 예외 (Exception) 발생

3. 데이터 수집

데이터 수집 과정에서의 어려움

```
except UnexpectedAlertPresentException as e:  
    print(type(e))  
    driver.switch_to.alert.accept()  
    print("Alert accepted")  
  
except Exception as e:  
    print(type(e))  
  
finally:  
    driver.close()  
    idx = idx + 1
```

3. 데이터 수집

전자담배의 금연 기여도

: 구글 폼을 이용한 설문조사

: 2019.05.16 ~ 2019.05.27
2주간 진행

Why ?

1. Tweepy API의 한계

: 최근 1주일 간의 tweet만 제공

2. 한정된 연령대의 SNS 사용

전자담배 금연기여도 설문조사

안녕하세요, 저희는 동덕여대 학생들로 전공과목 프로젝트를 진행함에 있어서 좀 더 많은 정보를 수집하고자 설문조사를 진행하게 되었습니다. 귀하의 답안은 오로지 프로젝트 진행에만 이용될 것을 약속합니다.

* 필수항목

1. 전자담배를 현재 사용하고 있거나, 사용해 본 경험이 있으십니까? *

☒ 예

☐ 아니오

2. 전자담배가 금연에 도움이 된다고 생각하십니까? *

☐ 예

☐ 아니오

3. 전자담배가 연초보다 몸에 덜 해롭다고 생각하십니까? *

전자담배 사용 미경험자의 경우, 3번 문항에서 설문종료

☐ 예

☐ 아니오

4. 전자담배를 이용하게 된 가장 큰 계기 및 이유는 무엇입니까?

☐ 금연을 하기 위해

☐ 흡연량을 조절하기 위해

4. 데이터 전처리

naver_blog_245

안녕하세요~ 하루에 몰아서 포스팅하는 틈덩이에요ㅠ
요즘 목군이 금연을 시작했어요ㅋㅋㅋ
맨날 구박받더니 이제서야!!

금연이라고는 못하지만
그래도 전자담배 하겠다고 결심한게 얼마나 기특한지ㅋㅋㅋ
3개월 뒤에는 전자담배도 끊기로 했어요
하루에 한갑피던 사람이.. 독하게 마음먹은거죠ㅋㅋㅋ

전자담배 사달라길래 단 1초의 고민도 없이 OK!!ㅋㅋㅋ
바로 가서 사왔어요
근데 은근히 비싸더라고요ㅠ

인터넷 검색해보다가 제일 가까운곳으로ㅋㅋㅋ
지나다니면서 엄청 봤는데 막상 가서 사려니까 안보이더라고요ㄷㄷ
네비찍고 가서 산 마사 전자담배

전자담배 기계를 한개에 8만원 2개에 12만원이더라고요
목군이 2개 사고싶어 하는 눈치길래
쿨하게 2개 샀어요
근데 지금 보니까 액상 돌린거를 2개사서 번갈아 가면서 피더라고요
은근 좋아하는 눈치ㅋㅋㅋㅋㅋㅋㅋㅋ

처음엔 헛구역질 난다더니 지금은 담배 생각도 안난다느니
이제 2주일인데ㅋㅋㅋ

```
data=data.replace('\n','')
for i in range(1,953):
    x="naver_blog_"+str(i)
    data=data.replace(x,"\\n\\n"+x+" "+"\\n")
f = open('전자담배_naver_blog_992.txt', 'w', encoding='utf-8')
f.write(data)
print(data)
```

naver_blog_245

대전 최초 24시 운영을 하는 유성전자담배를 소개하려고 해요 :-)
저는 흡연을 하지 않지만 요즘 전자담배를 사용하는 분들이 많더라구요!대전 유성전자담배는 신제품 기기 및 액상을 전국에서 최소 한달 먼저 빠르게 입고하고 있다고 해요 :-)유성 전자담배에서는 직접 손으로 만져보고 사용해보고 구매 할 수 있어서 좋아요! 사진에서 보는 것 같이 모든제품을 개봉해서 진열해놓았답니다! 그리고 리빌드를 매장에 방문해서 직접 편하게 할 수도 있다고 해요 또한 인터넷 구매시 사후 A/S 부분이 번거롭고 불가능 할 수 있지만 유성전자담배에서는 언제나 A/S 및 점검을 할 수 있다고 하니 대전 지역분들은 편리할 것 같아요!지금 입호흡 액상 2+1행사 / 모드액상도 2+1(일부제품)행사중이며기기 구매시 대학생/군인/폐북/인스타 등 10% 할인 이벤트를 하고 있어요! 유튜브 '별구베이프' 구독&좋아요 눌러주시면, 쿨링제10ml 선물 드리고 있다고 해요 :-)(정식 판매 되고 있는 최고급 원료 쿨링제)전자담배는 대전 유성전자담배를 찾아가 보시는게 어떠실까요!대전 유성구 봉명동 466-21번지 1층유성 리베라호텔 맞은편 롯데리아 전화번호 : 010-7502-5270 영업시간 : 24시간 / 연중무휴 / 단 일요일날 밤12시까지주차는'세움펠리피아 주상복합 아파트 지하 1층 무료주차'

4. 데이터 전처리



1. 부사어 및 형용사	‘꼭좀’, ‘무작정’, ‘차라리’, ‘지난번’, ‘몰래’, ‘게다가’, ‘괜찮은’
2. 일반적인 명사	‘아무나’, ‘이하’, ‘상태’, ‘개월’, ‘개념’, ‘사람’
3. 의성어, 의태어, 감탄사	‘빠끔빠끔’, ‘질끈’, ‘쑹쑹쑹’, ‘오오오’
4. 지명 및 인물명	‘남동구’, ‘부산’, ‘충청남도’, ‘방선희’, ‘김태원’
5. 오타 혹은 형태소 분석기의 성능 문제	‘님이’, ‘권찮’, ‘잘하셨는’, ‘비릿하’, ‘없드라구’, ‘슬림해’
6. 사전적인 의미를 알 수 없는 단어	‘르그’, ‘있즌’, ‘읍니’, ‘오시’, ‘레소’, ‘노보’
7. 접속사	‘그리고’, ‘그런데’, ‘또한’, ‘하지만’
8. 고유명사	‘조선일보’, ‘이데일리’, ‘뉴스투데이’, ‘한경닷컴’

5. 데이터 분석 및 결과 적용

블로그



전자담배



전자담배 효과



전자담배 금연



전자담배 유해



전자담배 중독



전자담배 증상



지식인

5. 데이터 분석 및 결과 적용

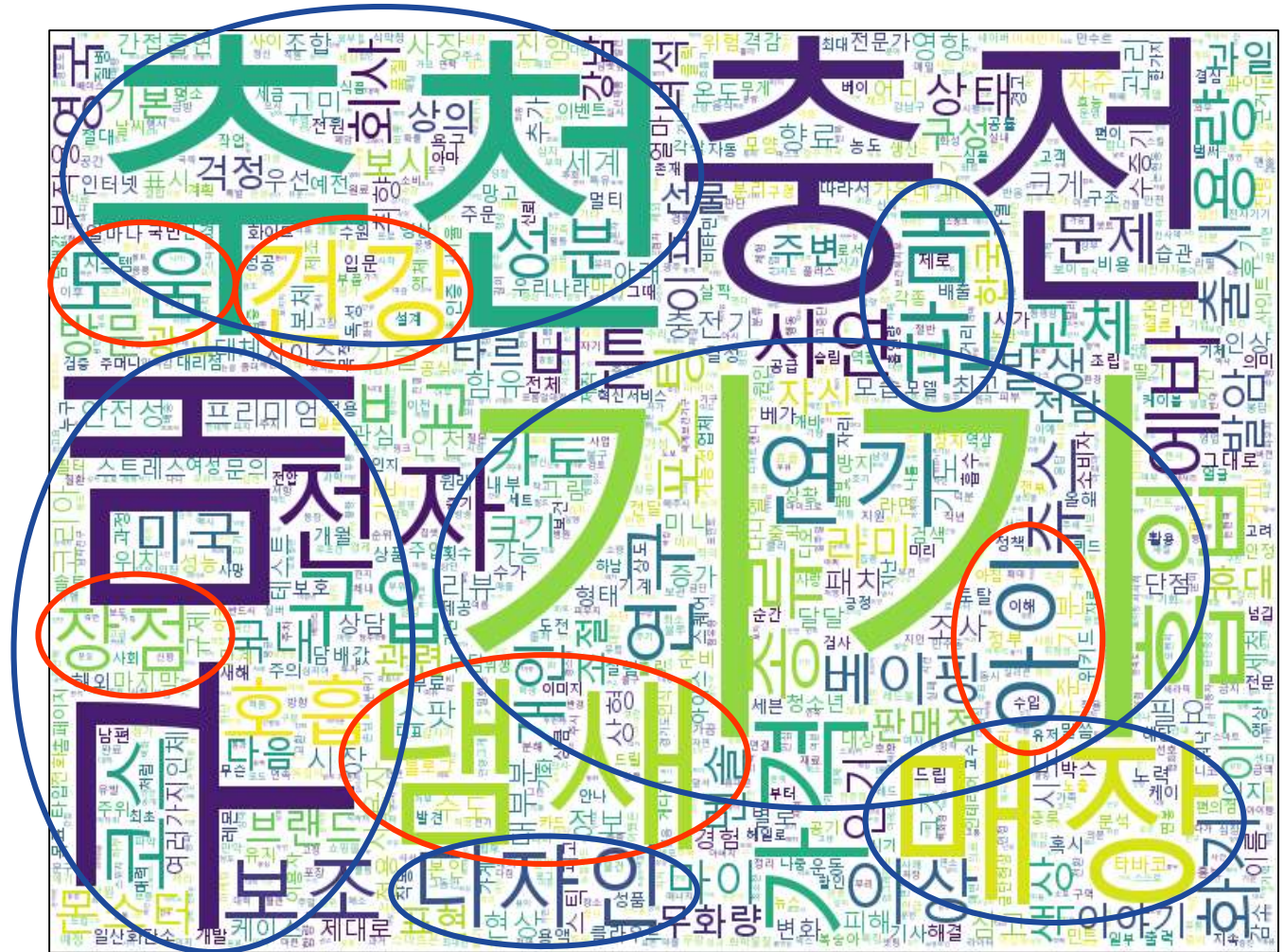
	뉴스 Topic Modeling								
	전자담배	전자담배 효과	전자담배 유해			전자담배 금연		전자담배 중독	
1	국내	지난해	니코틴	발표	식약처	흡연	담뱃갑	담뱃갑	흡연
2	출시	전체	소비자	실험	규제	과태료	경고	금연	금연
3	시장	인상	정보	포름알데히드	타르	일반	정책	암세포	소비자
4	브랜드	담뱃값	품질	분석	분석	감소	세계	수위	강력
5	미국	올해	공식	비교	세금	예방	강화	문구	노화
6	하반기	세금	법적	정보공개	경고	미국	보건복지부	효과	마리오네트
7	중국	시장	시장	소송	연구	성분	문구	그림	정책
8	매출	필립모리스	함량	인식	화학	성인	그림	표현	경고
9	제품	매출	분석	실제	근거	단속	청소년	니코틴	마약
10	퀄런	경쟁	판매	건강	정책	증가	단속	판매	판매
11	판매	사망	감소	세부	청소년	흡연식	흡연	방송	하와이
12	업그레이드	퀄런	간접흡연	결과	캠페인	인체	연기	포함	뉴질랜드
13	전체	전망	신제품	거부처분	교육	폐암	발생	도박	인체
14	상륙	수출	시판	데이터	세계보건기구	판매량	도입	전면	한국
15	신제품	국내	출시	제기	논란	치료	예방	규제	미국

5. 데이터 분석 및 결과 적용

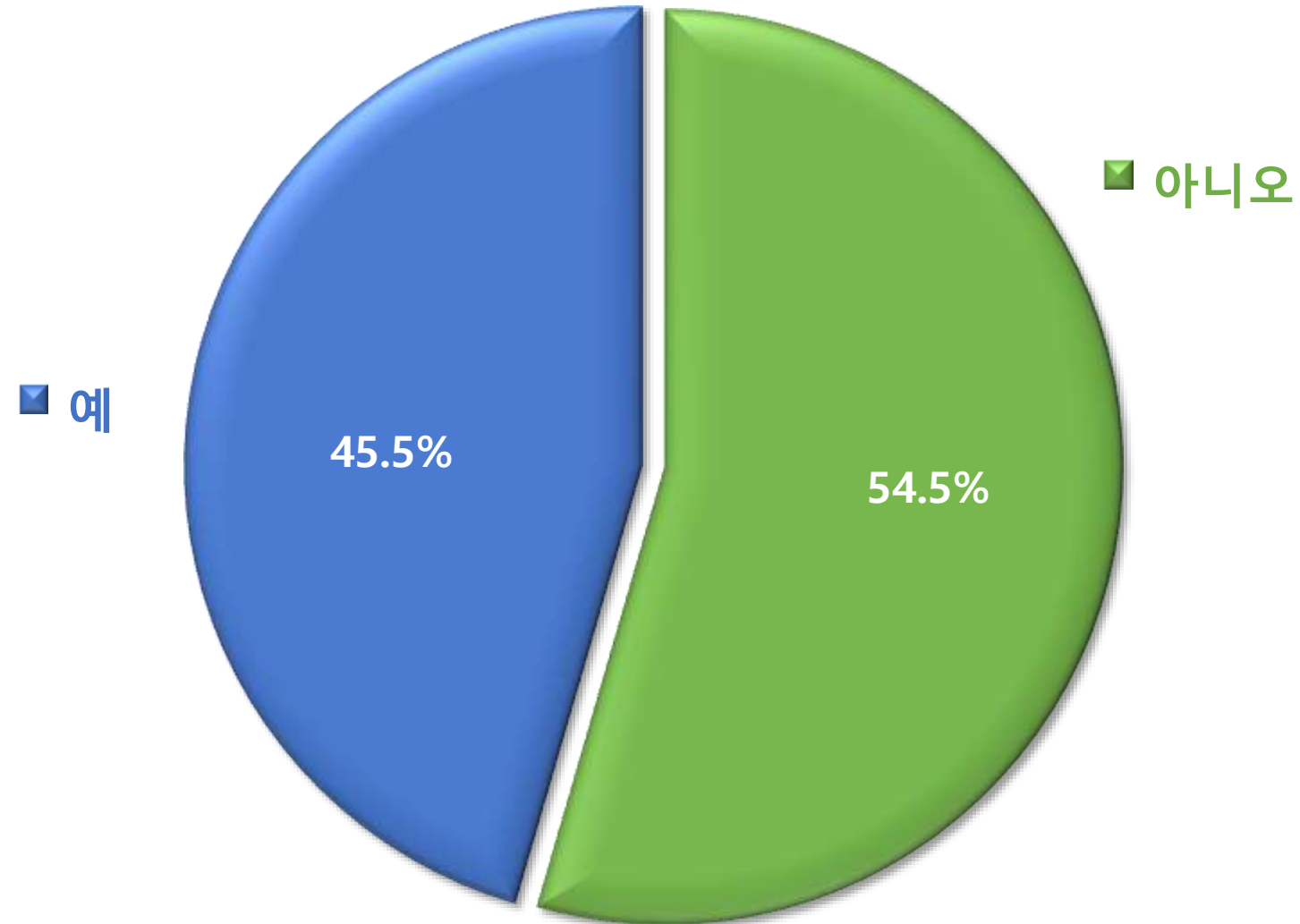
블로그



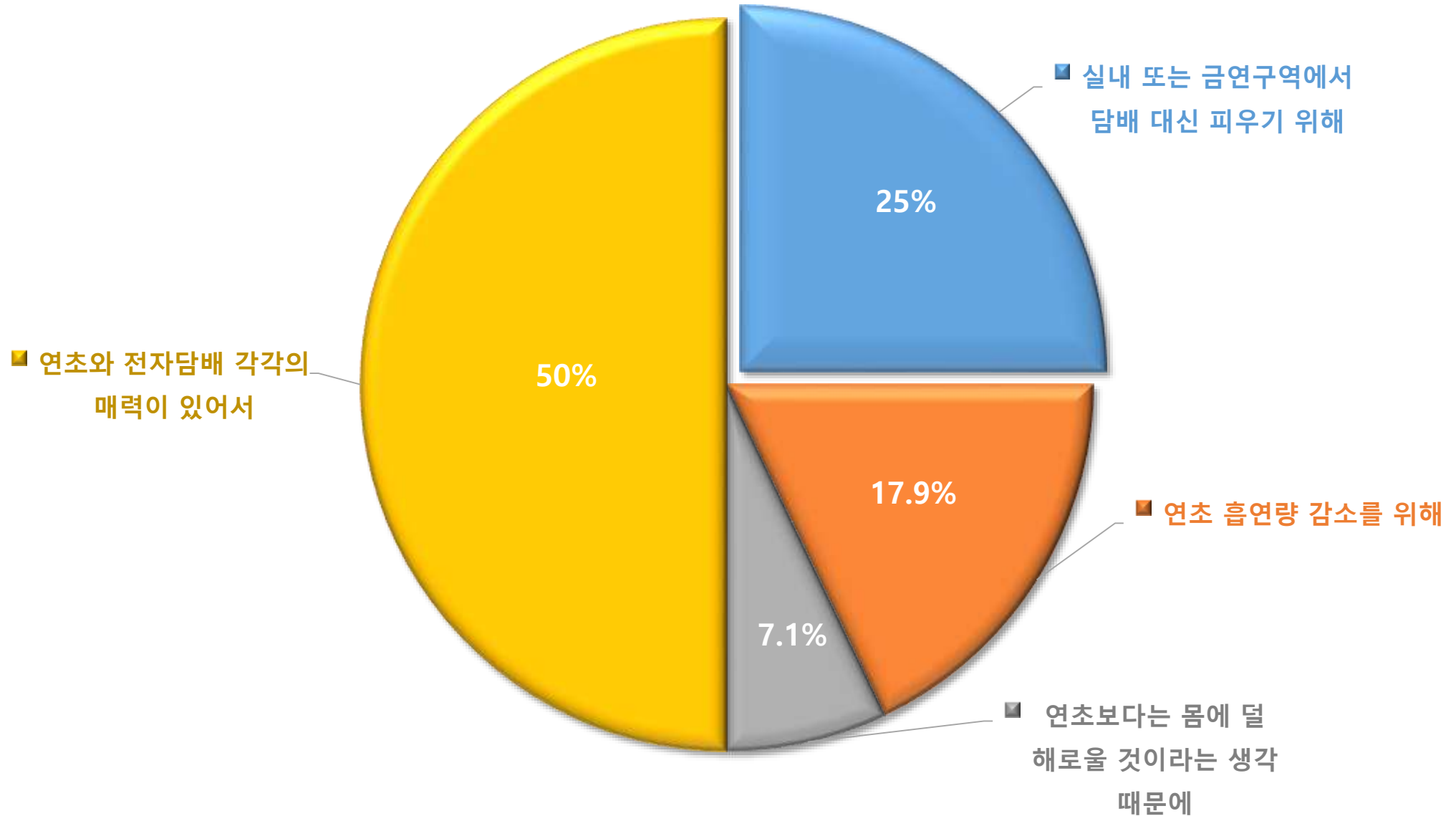
전자담배 효과



전자담배가 연초보다 몸에 덜 해롭다고 생각하십니까?



전자담배와 연초를 함께 피우는 이유는 무엇입니까?

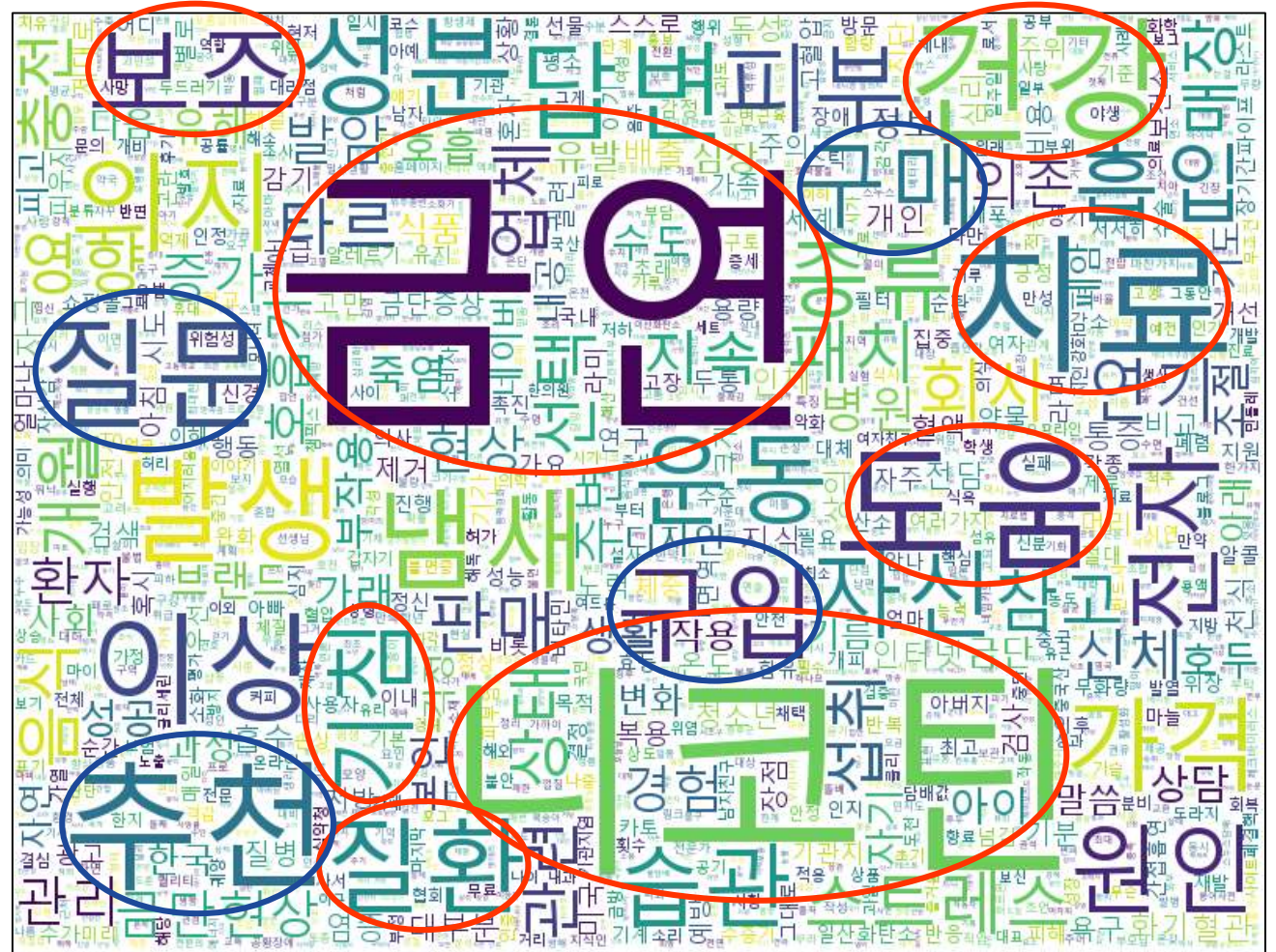


5. 데이터 분석 및 결과 적용

지식인



전자담배 효과



5. 데이터 분석 및 결과 적용



	전자담배 효과		
	지식인		
	Topic 1	Topic 2	Topic 3
1	전자담배	액상	금연
2	니코틴	전자담배	니코틴
3	금연	배터리	전자담배
4	이상	매장	흡연
5	액상	사용	이상
6	일반	금연	금단
7	흡연	추천	제품
8	사용	코일	호흡기
9	효과	브랜드	부작용
10	금단	니코틴	중국산
11	제품	구입	질환
12	배터리	제품	신장
13	추천	디자인	실패
14	코일	구매	성공
15	성공	성공	정보

Topic 1

금연하려고 하는데 전자담배 사용이 효과가 있나요?
액상 전자담배 제품추천 부탁드립니다.

Topic 2

금연을 결심하였으나 바로 끊는 것에 무리가 있어 니코틴이
함유된 액상 전자담배 제품을 구매하여 사용하고자 합니다.
금연에 성공할 수 있을까요?

5. 데이터 분석 및 결과 적용

NAVER 지식백과

니코틴 금단

Q

통합검색

☰ 카테고리 보기 ▾

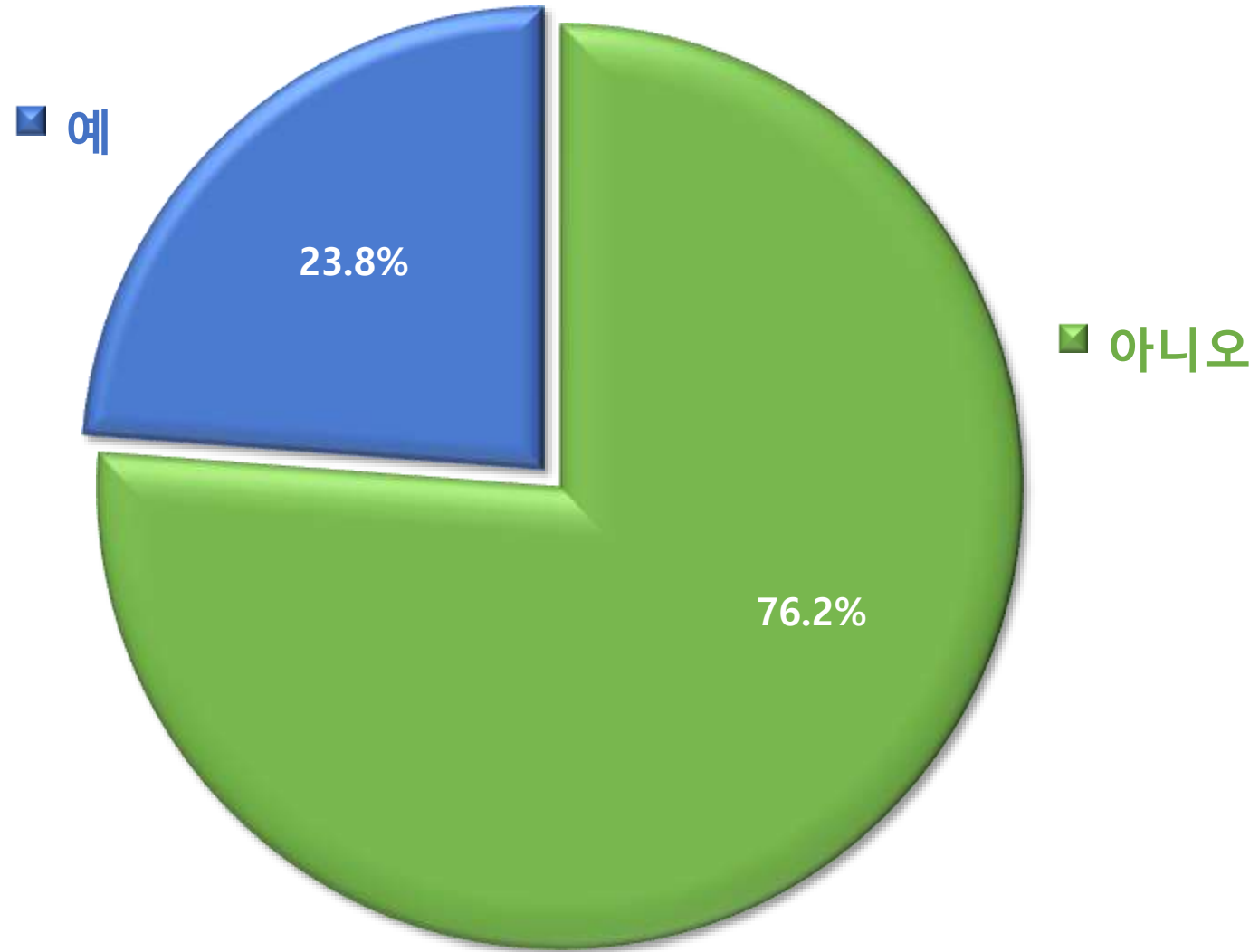
4. 니코틴 금단

적어도 몇 주 이상 니코틴을 매일 사용하다가 니코틴 사용을 급격하게 중단하거나 사용량을 줄였을 때 24시간 이내에 여러 가지 부적응적인 징후가 나타나는 경우를 말한다. 니코틴 금단 증상은 (1) 불쾌한 기분 또는 우울한 기분, (2) 불면, (3) 자극 과민성, 좌절감, 분노, (4) 불안, (5) 집중력 장애, (6) 안절부절못함, (7) 심장 박동 수 감소, (8) 식욕 증가 또는 체중 증가이며 이 중 4개 이상의 징후가 나타나 사회적, 직업적 또는 다른 중요한 기능에서 현저한 곤란이 초래될 때 니코틴 금단으로 진단한다. 보통 니코틴 사용을 중단하면 금단 증상이 90-120분 후부터 나타나기 시작해 24-48시간에 최고조에 이르고 몇 주 또는 몇 개월간 지속된다.

장기간의 흡연과 관련된 의학적 문제에는 폐암, 후두암, 기종 및 심장혈관 질환 등이 있다. 하루에 담배 두 갑을 30년 정도 피운 사람은 비흡연자보다 사망률이 2배 더 높다 (Bennett, 1980). 흡연자는 호흡기 증상과 폐결핵 유병률이 비흡연자보다 유의하게 높다. 즉, 흡연자 가운데 기침, 가래, 호흡 곤란 등의 호흡기 증상을 호소하는 비율은 31.5%로 비흡연자의 11.4%에 비해 현저하게 높다. 20개비 이상 피우는 사람 가운데 호흡기 증상을 보이는 사람은 무려 40%가 넘는다. 폐결핵 유병률도 비흡연자의 2.2배나 된다(대한결핵협회, 1994).

	전자담배 증상				
	지식인				
	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
1	전자담배	천식	치료	니코틴	피부
2	금연	가래	전자담배	담배	치료
3	담배	이상	호흡	공황장애	증상
4	니코틴	담배	질환	중독	수면
5	액상	감기	자각	이상	금연
6	금단증상	전자담배	담배	질환	전자담배
7	스트레스	호흡기	알레르기	체온조절	혈관
8	간접흡연	기침	심장	만성	얼굴
9	성공	기관지	호흡기	긴장	고생
10	사용	항생제	재발	두드러기	기분
11	어지러움	이비인후과	니코틴	금단현상	부작용
12	질문	비타민	전문의	혈압	장기간
13	원인	폐암	수술	가족	흡입
14	개선	역류성	원인	금연	발병
15	심장	후두	음식	변비	구내염

실제 전자담배 사용이 금연에 도움을 미쳤습니까?



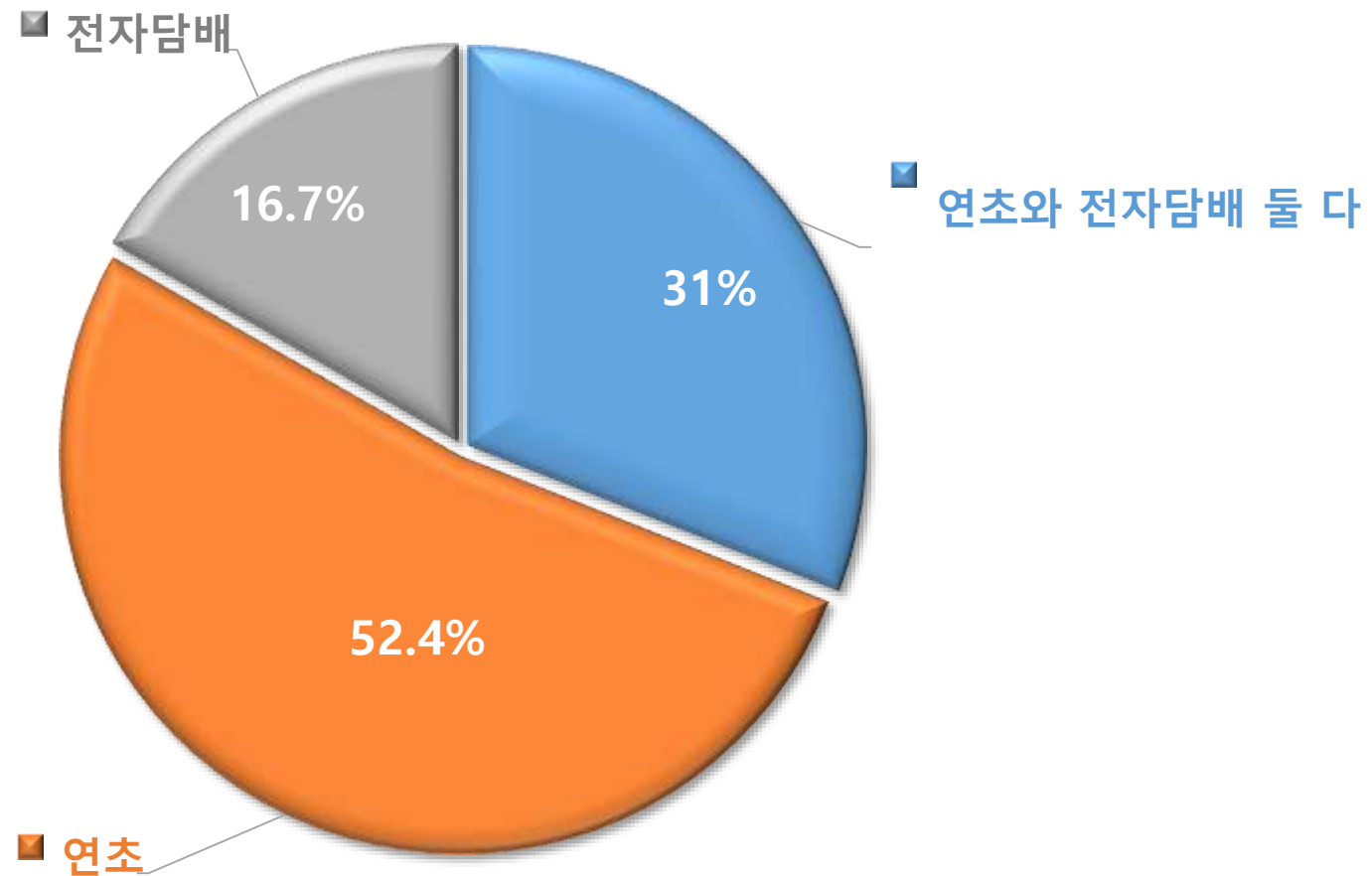
5. 데이터 분석 및 결과 적용

“ 전자담배 업체는 명확한 정보 제공의 자세 필요 ”

“ 정부는 무분별한 광고에 대해 적절한 규제와 처방 필요 ”

6. 한계점 & Discussion

현재 사용하는 담배의 종류는 무엇입니까?



감사합니다.



END