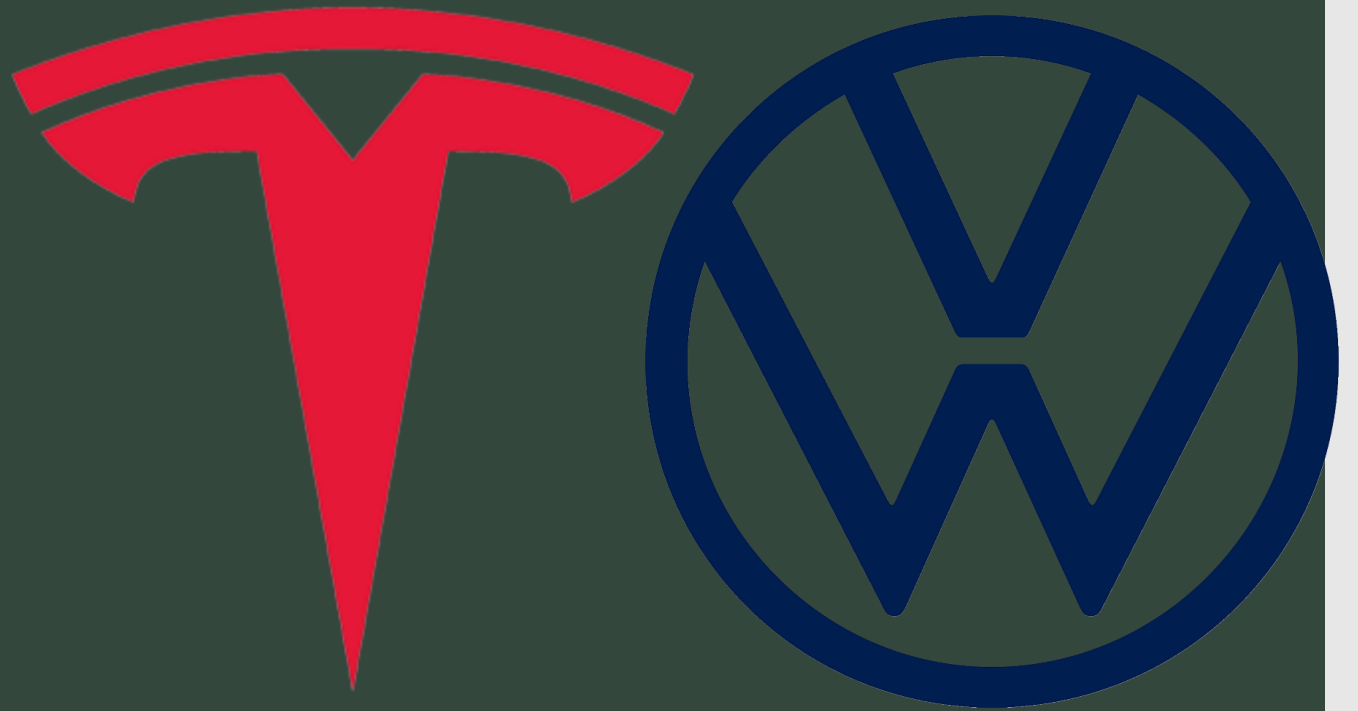


# 웹 스크레이핑을 이용한 글로벌 자동차업체 ESG 차이 요인 분석

## Tesla vs Volkswagen

---



# 목 차

1. 개 요

2. 데이터 수집 및 처리

3. 데이터 분석

4. 결론

# ESG 란?

• 사례



• 개념



- 지속가능경영과 사회적책임(CSR)이 진화하고 규범화 및 제도화된 형태
- 투자의사결정 및 장기적인 재무적 가치에 영향을 미칠 수 있는 중요한 비재무적 요인

## 이슈 : 자동차 업계의 화두 - ESG

- 탄소중립, ESG 경영 속도 내는 자동차 기업 - 테슬라
  - 전기차/태양광 발전/에너지 저장장치 : 기후위기 극복 위한 CO<sub>2</sub> 저감 노력으로 각인, 응원하는 투자자 급증

Tesla Model Y	21,812
Tesla Model 3	21,506
Toyota RAV4	15,990
Toyota Camry	12,257
Honda Civic	11,057

'22.1Q 캘리포니아주 차량 판매 순위

Luxury Compact SUV			Near Luxury		
Model	Regs.	Share	Model	Regs.	Share
Tesla Model Y	21812	55.5	Tesla Model 3	21506	65.2
Mercedes GLC-Class	4050	10.3	BMW 3-Series	2442	7.4
BMW X3	2833	7.2	Lexus ES	2411	7.3
Audi Q5	1842	4.7	BMW 4-Series	1711	5.2
Lexus NX	1808	4.6	Lexus IS	1250	3.8

'22.1Q 캘리포니아주 차량 판매 순위 (부문별)

- 전 세계적으로 이산화탄소 배출량을 줄이는 등 기후변화 대응 압력 지속 증가
  - 전통 완성차 업체들 : 기존 내연기관 차량에서 전기차 등 친환경차 전환에 '울인'



'35년부터 전기차만 만드는  
기업으로 탈바꿈 선언



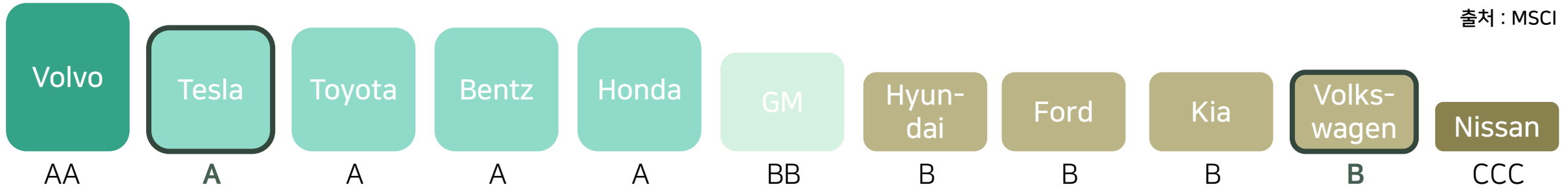
'30년까지 생산량의 40%,  
'탄소 0' 차량으로 생산 발표



'39년까지 탄소  
중립 실현 선언

## 과제 : 기업들간의 ESG 평가 차이 발생 원인은?

### • 자동차 기업들간 ESG 평가 차이



### • ESG 평가방식

- “최근 기업 평가, 투자 기준 등 때문에 ESG경영의 필요성이 급증하고 있지만 국내외 ESG 평가기관들은 평가 기준이나 평가 결과 도출 방식을 공개하지 않고 있다”
- “다우존스지속가능경영지수(DJSI), 모건스탠리캐피탈인터내셔널(MSCI), 세계경제포럼(WEF) 등 국내외 주요 13개 평가기관의 3000여 개 이상의 지표와 측정항목”

“복잡한 ESG 등급, 이슈와 요인을 쉽게 분석할 수 있는 방법은 없을까?” (과제1)

“기업들간의 ESG 평가 차이 발생 원인은 무엇인가?” (과제2)

# 데이터 수집 과정

## 데이터 정의

- 데이터: Tesla와 Volkswagen의 ESG 뉴스기사 데이터
- 비교 기업: Tesla, Volkswagen

## 데이터 수집

- 웹사이트: Financial Times 국제 경제 뉴스 사이트
- 수집 데이터: ESG 뉴스기사의 제목, 본문 일부, 날짜
- 딕셔너리: ESG 딕셔너리 활용

E, S, G 각각의 클래스에 대하여 논문 기반으로 구축



	Environmental	Social	Governance
1	Biodiversity	Community	Audit
2	Climate	Customer	Committee
3	Ecosystem	Data	Board
4	Change	Privacy	Composition
5	Environment	Discrimination	Business
6	Policy	Employee	Ethics
7	Greenhouse	Engagement	Corruption
8	Gas	Human	Instability
9	Emission	Rights	Employee
10	Pollution	Labor	Diversity
11	Renewable	Relations	Management
12	Energy	Standards	Shareowner
13	Resource	Politic	Rights
14	Depletion	Contribution	Tax
15	Chemical	Safety	Transparency
16	Waste	Quality	Voting
17	Toxic	Harassment	Governance

# 데이터 수집 과정

## 논문 및 리포트 기반 딕셔너리 구축

- From the stockholder to the stakeholder: How sustainability can drive financial outperformance  
Clark, Gordon L., Andreas Feiner, and Michael Viehs. (2015). *Arabesque Asset Management*. Cited by 365
- Environmental, social and governance reporting in annual reports: A textual analysis  
Baier, P., Berninger, M., & Kiesel, F. (2020). *Financial Markets, Institutions & Instruments*. Cited by 23

TABLE 1: SELECTION OF MATERIAL ESG FACTORS<sup>27</sup>

ENVIRONMENTAL ("E")	SOCIAL ("S")	GOVERNANCE ("G")
Biodiversity/land use	Community relations	Accountability
Carbon emissions	Controversial business	Anti-takeover measures
Climate change risks	Customer relations/product	Board structure/size
Energy usage	Diversity issues	Bribery and corruption
Raw material sourcing	Employee relations	CEO duality
Regulatory/legal risks	Health and safety	Executive compensation schemes
Supply chain management	Human capital management	Ownership structure
Waste and recycling	Human rights	Shareholder rights
Water management	Responsible marketing and R&D	Transparency
Weather events	Union relationships	Voting procedures

TABLE 1 Topics, categories and subcategories of the ESG word list

Governance		
Corporate governance	Business ethics	Sustainability management and reporting
Audit and control	Bribery and corruption	Disclosure and reporting
Board structure	Political influence	Stakeholder engagement
Remuneration	Responsible marketing	UNGC compliance
Shareholder rights	Whistle-blowing system	Governance of sustainability issues
Transparency <sup>a</sup>		
Talent <sup>a</sup>		
Environmental		
Ecosystem service	Climate change	Environmental management
Access to land	Biofuels	Environmental standards
Biodiversity management	Climate change strategy	Pollution control
Water	Emissions management and reporting	Product opportunities
		Waste and recycling
		Supply chain environmental standards
Social		
Public health	Human rights	Labor standards
Access to medicine	Community relations	Diversity
HIV and AIDS	Privacy and free expression	Health and safety
Nutrition	Security	ILO core conventions
Product safety	Weak governance zones	Supply chain labor standards
Society <sup>a</sup>		
Charity <sup>a</sup>		
Education <sup>a</sup>		
Employment <sup>a</sup>		

<sup>a</sup>Additional (sub-)category added to the categorization by Dimson et al. (2015).  
<sup>b</sup>Renamed (original: Transparency and performance).

# 데이터 수집 과정



## 데이터 스크래핑 과정

1. URL에 기업명과 ESG 딕셔너리 서브클래스 입력 (①)
2. 전체 페이지 수와 현재 페이지 수로 반복문 횟수 정하기 (⑤, ⑥)
3. 기사 제목, 본문 일부, 날짜 스크래핑 (②, ③, ④)

① URL

https://www.ft.com/search?q=tesla%2BClimate&page=2&sort=date

(1) (2)

(1) 기업명과 ESG 딕셔너리 서브클래스

(2) 페이지 수

② 기사 제목

③ 기사 본문 일부

④ 기사 날짜

⑤ 현재 페이지 수

⑥ 전체 페이지 수



## 데이터 수집 결과

### 데이터 수집 결과

- 데이터 개수: Tesla 20,247개, Volkswagen 21,498개, 총 41,745개

	E	S	G	Total
Tesla	7,551	6,689	6,007	20,247
Volkswagen	7,248	7,312	6,938	21,498
Total	14,799	14,001	12,945	41,745

### 중복 데이터 처리

- 중복 기사 데이터 제거 후 개수: Tesla 8,603개, Volkswagen 10,367개, 총 18,970개

	E	S	G	Total
Tesla	2,893	3,126	2,584	8,603
Volkswagen	3,209	3,847	3,311	10,367
Total	6,102	6,973	5,895	18,970

## 데이터 변수 소개

### 데이터 변수

1. X: 기사 아이디
2. Company: 기업명
  - Tesla: 8,603개
  - Volkswagen: 10,367개
3. Esg: E, S, G 클래스 (총 3개)
4. Subclass: ESG 디렉터리에서 정의된 서브클래스 (16개씩 총 48개)
5. Title: 기사 제목
6. Text: 기사 본문 일부
7. Date: 기사 작성 날짜
  - Tesla: 2007년 ~ 2022년 (16년치 데이터)
  - Volkswagen: 2004년 ~ 2022년 (19년치 데이터)

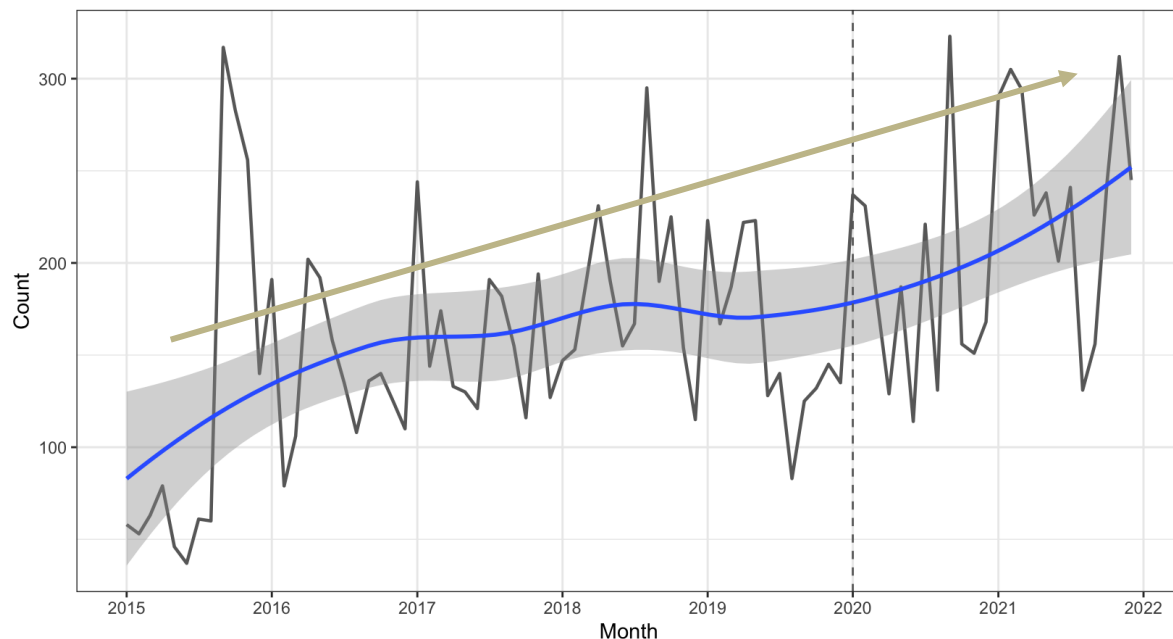
```
> tesla.esg
# A tibble: 8,603 x 7
```

	X	Company	Esg	Subclass	Title	Text	Date
	<int>	<chr>	<chr>	<chr>	<chr>	<chr>	<date>
1	1	tesla	Environment	Climate	The war on 'woke capit~	"...~	2022-05-27
2	2	tesla	Environment	Climate	Aviva's Amanda Blanc: ~	"...~	2022-05-27
3	3	tesla	Environment	Climate	Forget ESG. Bring on th~	"...~	2022-05-27
4	4	tesla	Environment	Climate	German companies strugg~	"...~	2022-05-25
5	5	tesla	Environment	Climate	'Do-good' measures do~	"...~	2022-05-23
6	6	tesla	Environment	Climate	The last metro? Crossra~	"...~	2022-05-22
7	7	tesla	Environment	Climate	Here's why S&P dropped~	"...~	2022-05-20
8	8	tesla	Environment	Climate	US faces summer fuel pr~	"...~	2022-05-19
9	9	tesla	Environment	Climate	Big Oil fends off share~	"...~	2022-05-17
10	10	tesla	Environment	Climate	FirstFT: New York mayor~	"...~	2022-05-17

→ 데이터 분석 시 Title+Text 분석 진행

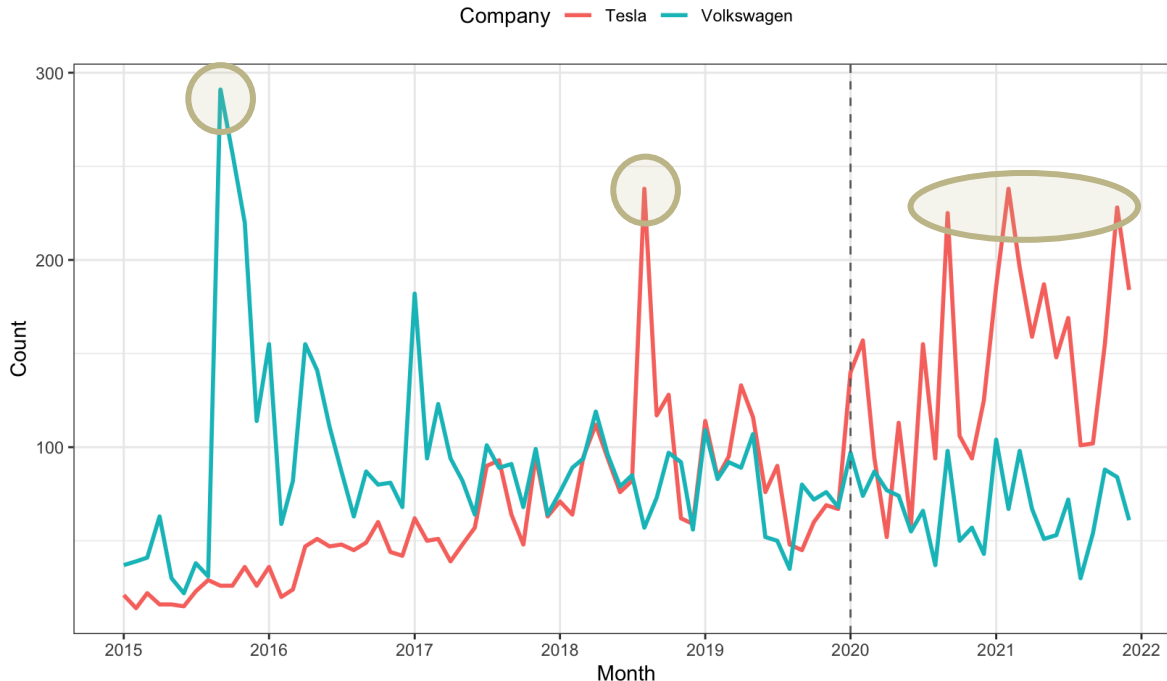
# ESG 뉴스기사 빈도 추이 분석

Total News Counts over Time (Monthly)



- ESG 기사 빈도는 증가 추세
- ESG 기사 빈도 평균: 코로나 전 < 코로나 후

News Counts by Tesla vs Volkswagen over Time (Monthly)

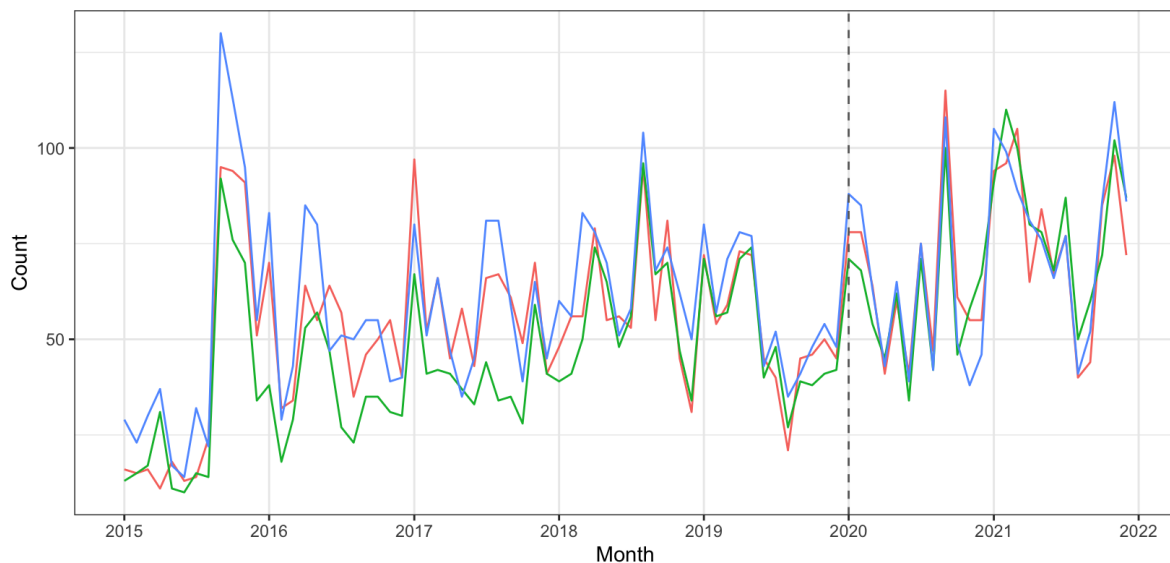


- 코로나 전, ESG 기사 빈도 평균: Tesla < Volkswagen
- 코로나 후, ESG 기사 빈도 평균: Tesla > Volkswagen
- 2015년 Volkswagen에 특이점, 2018년 Tesla에 특이점, 코로나 후 Tesla에 다수 특이점 보임

# ESG 뉴스기사 빈도 추이 분석

News Counts by ESG over Time (Monthly)

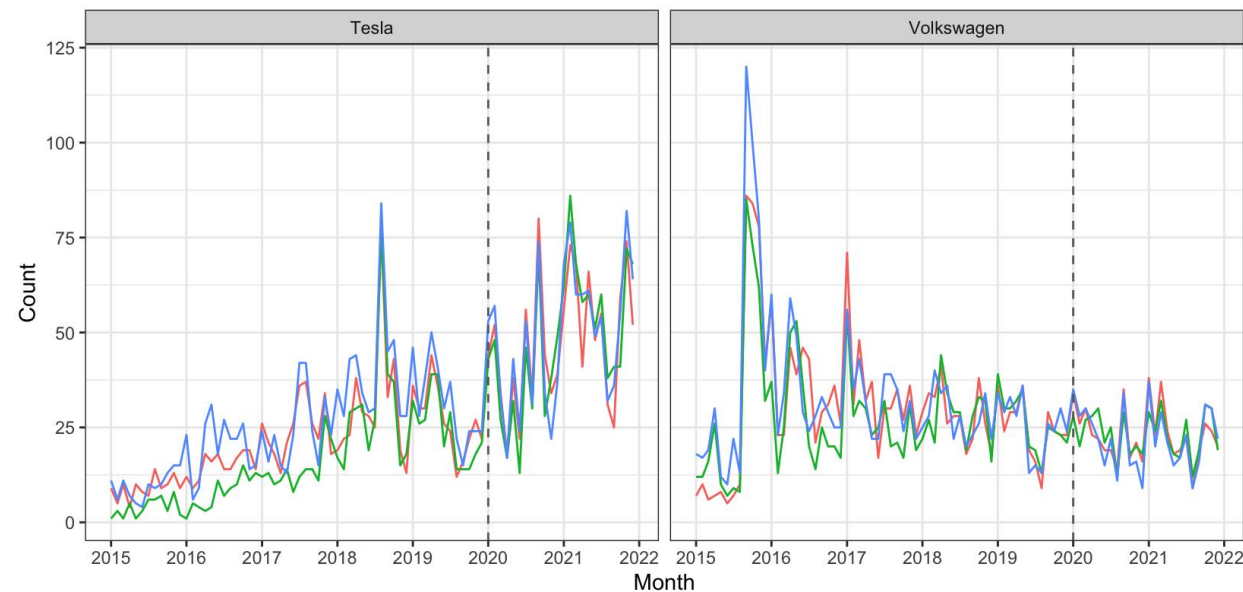
ESG — Environment — Governance — Social



- ESG 별 기사 빈도는 비슷한 비중을 보임
- ESG 별 기사 빈도는 비슷한 추이를 보임

News Counts by ESG over Time (Monthly)

ESG — Environment — Governance — Social



- Tesla 경우, COVID 전, ESG 별 기사 빈도:  $S > E > G$
- Volkswagen 경우, 2015년 특이점에서:  $S \gg E > G$
- 하지만 각 기업의 ESG 별 기사 빈도는 전반적으로 비슷한 비중 및 추이를 보임

## ESG 뉴스기사 빈도 추이 분석

- “ESG에 대한 관심도는 꾸준히 증가하고 있다.”
  - ESG 기사 빈도는 증가 추세에 있고 빈도 평균 또한 코로나 후가 코로나 전에 비해 높다.
  - ESG 기사 빈도 증가에는 특히 Tesla의 기여도가 높음을 회사별 빈도수 분석을 통해 확인할 수 있었다.
- “Tesla는 2018년, Volkswagen은 2015년 이슈가 있었을 것으로 예측할 수 있다.”
  - Tesla 경우 2018년, Volkswagen 경우 2015년 특이점(outlier)에 해당하는 피크를 보였다.
  - Textmining 후 그 결과와 실제 기사의 내용을 비교하여 textmining 분석 결과의 유의성을 검증할 수 있다.
- “E, S, G 클래스 중 두드러진 중요도를 보이는 클래스는 발견되지 않았다.”
  - ESG 개별 클래스에 대한 빈도수의 편향성이 확인되지 않았다.
  - 이는 (1) ESG 클래스 모두가 ESG 기사 빈도 증가에 기여, 즉 ESG의 세 클래스 모두 중요하게 다뤄지고 있다고 해석할 수 있다. 혹은 ESG의 화두는 많으나 (2) 언론이 전문적으로 ESG 클래스 별 분석 및 보도를 하지 못하고 있음을 시사한다고 볼 수도 있다.

# ESG 뉴스기사에 대한 감성 분석

## ① 텍스트 토큰화

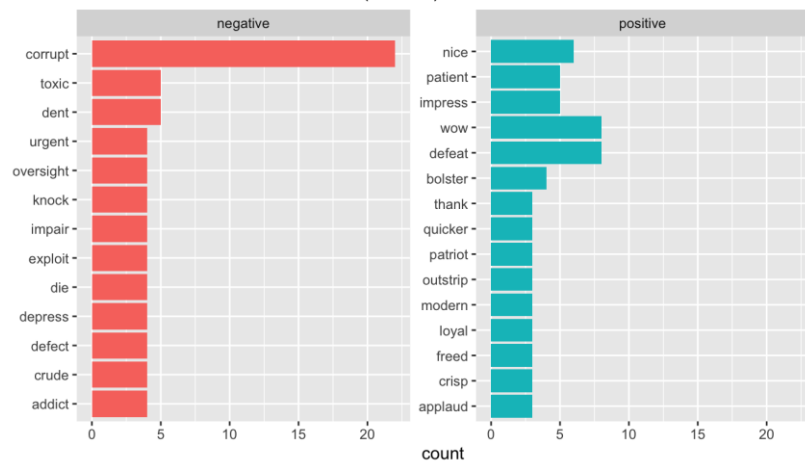
## 데이터 전처리 및 토큰화

```
esg.words <- esg.all %>%
  mutate(Month=floor_date(esg.all$Date, unit="month"))
  filter(Month >= "2015-01-01" & Month < "2022-01-01")
  mutate(text=str_replace_all(text, pattern="[^\w\d@-]"), pattern="[^\w\d@-]")
  unnest_tokens(word, text) %>%
  anti_join(stop_words, by="word") %>% #불용어 제거
  mutate(word=wordStem(word)) %>% #어간 추출
  mutate(word=gsub(pattern="can|could|will|would|is", replacement="", word)) %>% #구문
  filter(nchar(word) > 2) %>% #단어 길이
  count(Company, ESG=Esg, word, sort=TRUE, name="count")
  ungroup()
```

```
> esg.words
# A tibble: 36,342 x 4
  Company ESG word count
  <chr> <chr> <chr> <int>
1 Tesla Social tesla 2230
2 Tesla Environment tesla 1998
3 Tesla Governance tesla 1743
4 Volkswagen Environment volkswagen 1699
5 Volkswagen Social volkswagen 1699
6 Volkswagen Governance volkswagen 1482
7 Volkswagen Environment emiss 757
8 Volkswagen Environment car 753
9 Volkswagen Social car 727
10 Tesla Environment electr 672
# ... with 36,332 more rows
```

## ② TF-IDF 시행

Tesla's sentiment word counts (tf-idf>0)

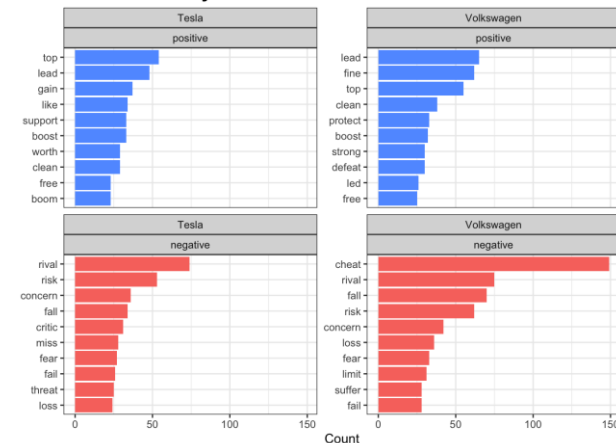


- TF-IDF>0인 단어들의 비교적 낮은 빈도 수
- TF-IDF 아닌 절대 빈도 사용

## ③ 감성 분석 : 'bing' 사전

```
> e.sentiment
# A tibble: 40 x 5
  Company ESG word count sentiment
  <chr> <chr> <chr> <int> <chr>
1 Tesla Environment rival 74 negative
2 Tesla Environment risk 53 negative
3 Tesla Environment concern 36 negative
4 Tesla Environment fall 34 negative
5 Tesla Environment critic 31 negative
6 Tesla Environment miss 28 negative
7 Tesla Environment fear 27 negative
8 Tesla Environment fail 26 negative
9 Tesla Environment threat 25 negative
10 Tesla Environment loss 24 negative
# ... with 30 more rows
```

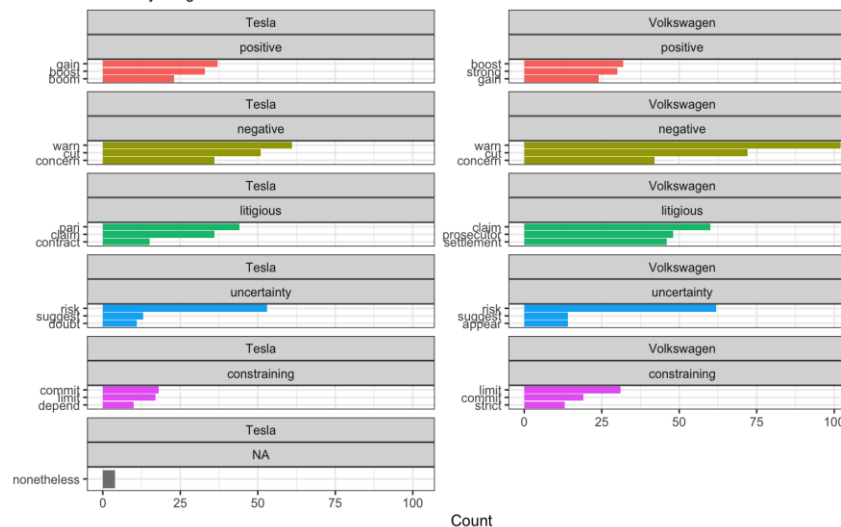
Sentiment Analysis in Environment



## ④ 감성 분석 : 'loughran' 사전

Sentiment Analysis in Environment

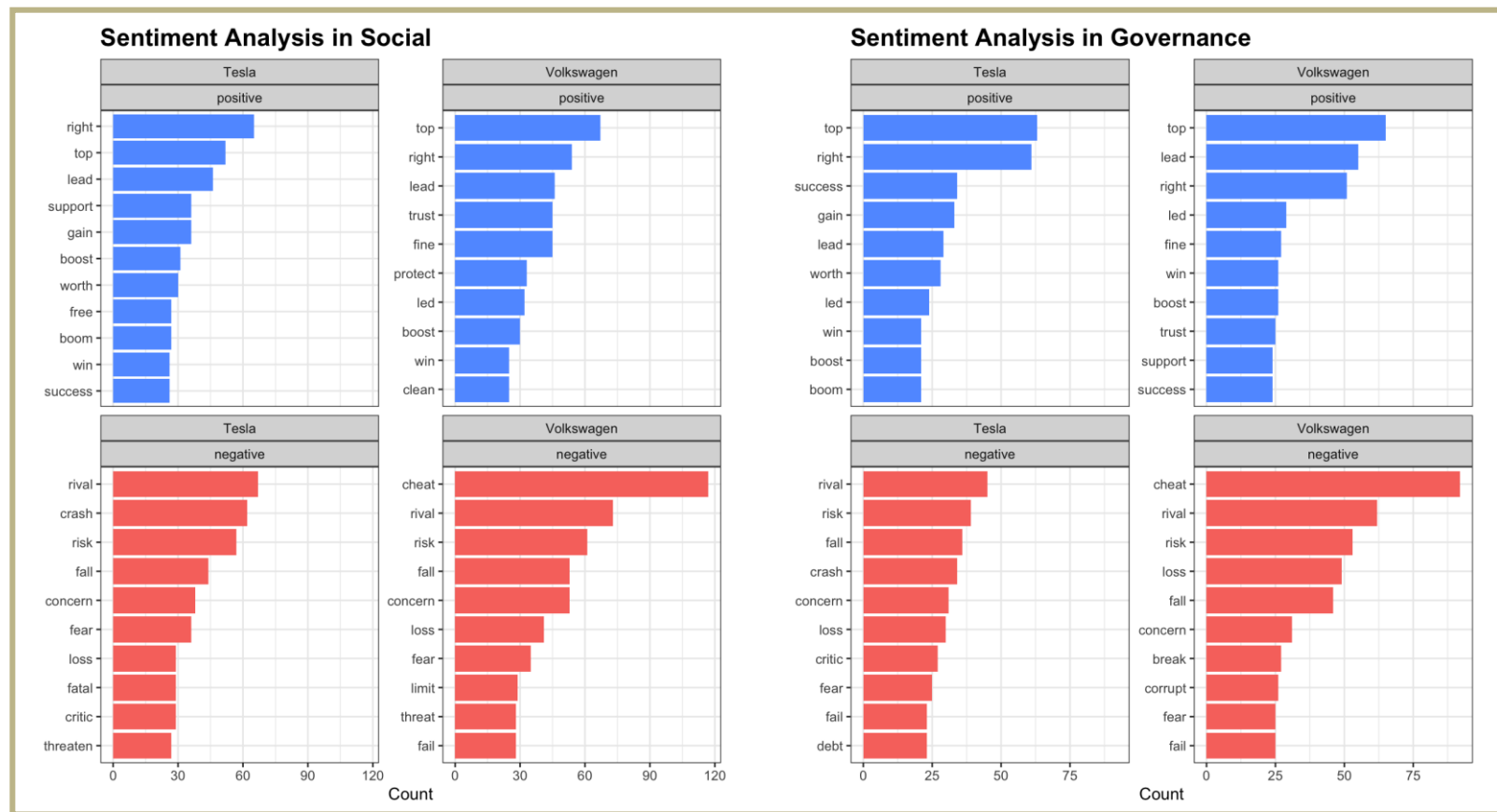
Dictionary: loughran



- 5단계로 구분할 만큼의 감성 단어 간 차별점을 보이지 않음
- Loughran 아닌 bing 사용

# ESG 뉴스기사에 대한 감성 분석

## ⑤ Environmental / Social+Government로 클래스 재구성

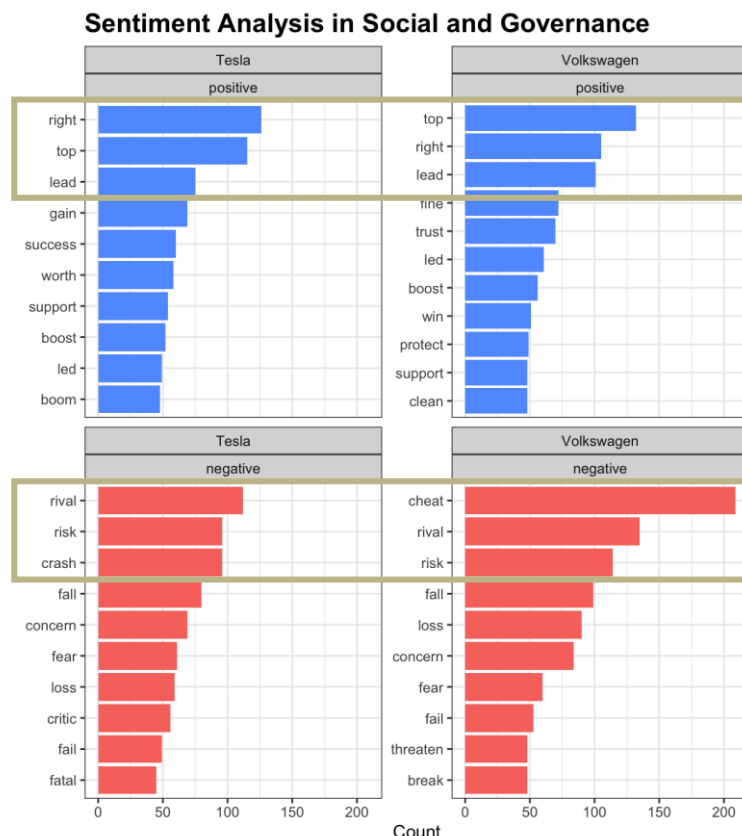
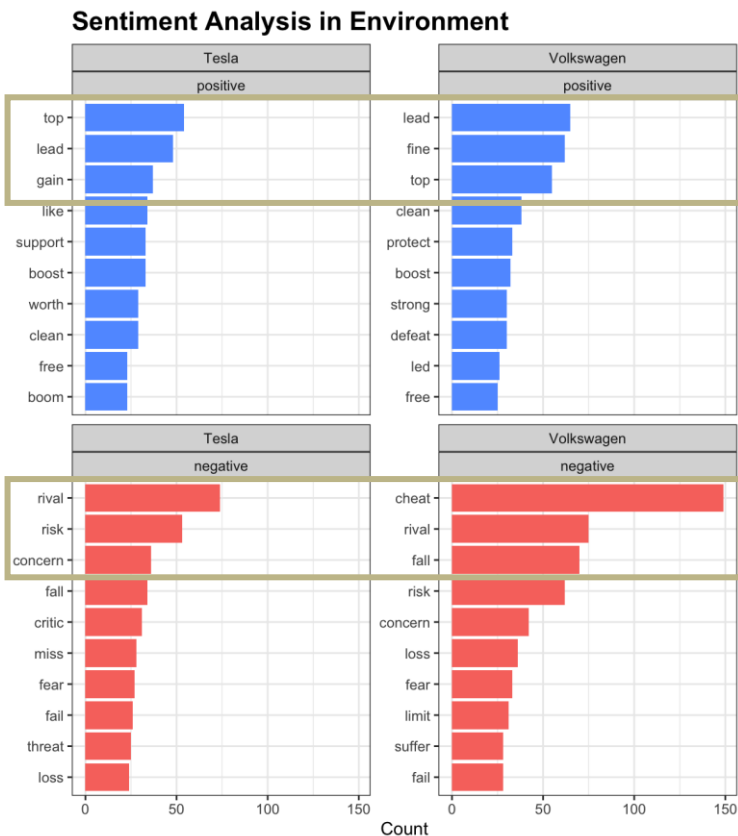


- Social vs. Governance: 단어 경향이 매우 비슷한 결과를 보임 (실제 Social과 Governance는 그 기준이 모호한 경우가 많다)

# ESG 뉴스기사에 대한 감성 분석

## Environmental: Tesla vs Volkswagen

P	
T	<u>gain</u> , support, worth, boom
V	<u>fine</u> , defeat, protect
N	
T	critic, miss, threat
V	<u>cheat</u> , limit, suffer



## Social+Governance: Tesla vs Volkswagen

P	
T	gain, success, worth, boom
V	fine, trust, win, protect, clean
N	
T	<u>crash</u> , critic, fatal
V	<u>cheat</u> , loss, threaten, break

E	S+G
<u>top</u> , <u>lead</u> , boost, <u>clean</u> , free, rival, risk, concern, fall, fear, fail, loss	<u>right</u> , <u>top</u> , <u>lead</u> , support, boost, led, rival, risk, fall, concern, fear, loss, fail



# ESG 뉴스기사에 대한 감성 분석

Tesla

Volkswagen

Tesla

Volkswagen

E

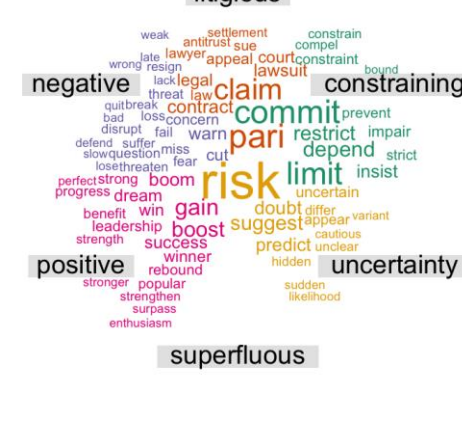


negative

S + G



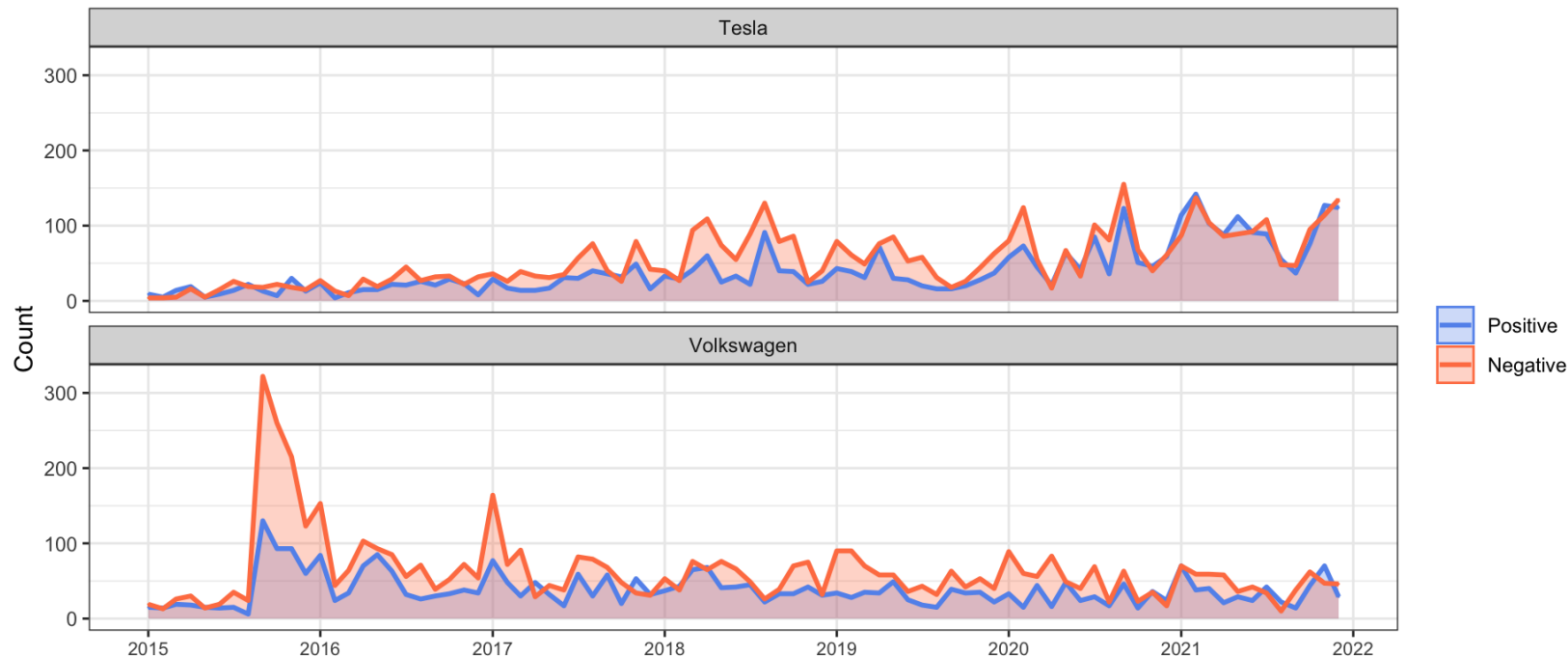
negative



- 공통적으로 자주 출현한 감성어휘는, top, lead, rival, risk
- 대표 감성어휘로 E에서는 clean, S+G에서는 right, trust
- E의 대표 감성어휘는, Tesla의 경우 gain, Volkswagen의 경우 fine, cheat
- S+G의 대표 감성어휘는, Tesla의 경우 crash, fatal, Volkswagen의 경우 cheat, threat, break
- 추가적으로 Volkswagen에서 litigious 용어들이 주요하게 출현

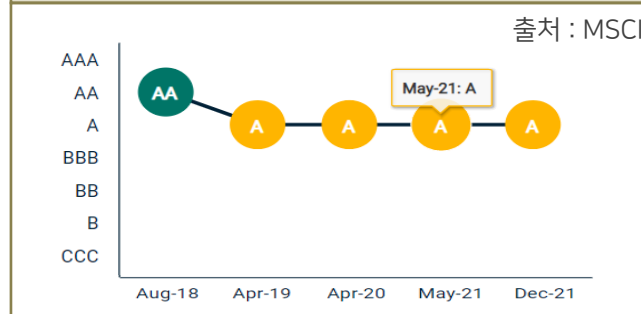
# ESG 뉴스기사에 대한 감성 분석

Positive/Negative Word Counts over Time (Monthly)

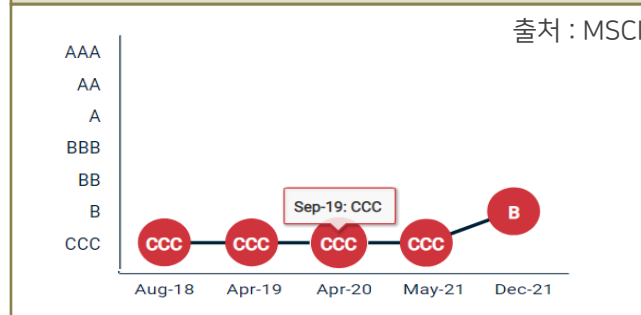


- | Negative - Positive | 차이: Tesla < Volkswagen
- | Negative - Positive | 는 ESG 등급을 반영하는 경향을 보임
- 2018년 Tesla의 Negative 어휘가 다소 높아지는 경향, 2015년 Volkswagen의 Negative 어휘가 매우 높아지는 현상을 보임

Tesla's ESG Rating History



Volkswagen's ESG Rating History



## ESG 뉴스기사에 대한 감성 분석

- ESG를 대표하는 단어를 발견할 수 있었다. 하지만 E, S, G 클래스를 구분하는 감성어휘는 다소 부족한 경향을 보였다.
- “실제 사건과 일치하는 대표 감성어휘를 통하여 ESG 등급에 영향을 끼치는 요인을 분석할 수 있었다.”
  - 2018년 발생한 Tesla의 자율주행으로 인한 치명적인 충돌 사고가 해당 기업의 ESG 주요 이슈
  - 2015년 발생한 Volkswagen 배기가스 치팅 사건이 (소비자들과의 법정 다툼으로도 이어지면서) ESG 평가에 영향을 끼침
  - 특히, Volkswagen 사건은 환경에 직접적 영향을 끼치고 소비자를 기만하는 경우로 ESG 등급 평가에 큰 영향을 주었을 것으로 예측.
  - 따라서, ESG 홍보 차원에서 보도를 내보내는 것도 중요하지만 제품 기본 기능과 소비자 신뢰를 지키는 것이 중요.



## 결론 및 의의

- ESG에 대한 관심은 지속적으로 증가하는 추세에 있다.
  - ESG 뉴스기사 데이터 분석 결과, ESG 기사 빈도는 증가 추세에 있음을 확인
  - 코로나 전후 비교시에도 ESG 기사 빈도 평균이 늘어나고 있는 상황
  - 하지만 ESG에 대한 관심 증가 대비, 아직 ESG 각 항목에 대한 언론의 전문적인 분석, 보도는 미흡하다는 부분이 추가적 시사점이 되겠다.
- (과제1) ESG 등급 요인은 뉴스기사 대표 감성어휘 추출과 긍부정 단어 빈도수 비교를 통해 분석할 수 있다.
  - 아직 E, S, G별로 세분화하여 전문적으로 보도하고 있진 않기 때문에 단순히 클래스별 빈도수를 분석하는 것만으로는 의미 없음.
  - 대표 감성어휘를 통해 구체적인 Tesla와 Volkswagen의 이슈를 발견할 수 있었다.
  - 긍부정 단어 빈도수는 ESG 등급을 반영하는 주요 변수라 할 수 있다.
  - 따라서 ESG 평가 기준이 불투명한 상황에서, 일반인이 ESG 요인을 분석하거나 ESG 투자를 하고자할 때 ESG 뉴스기사 감성분석은 유의한 분석 수단이 될 수 있다.

## 결론 및 의의

- (과제2) 자동차 산업 내에서 ESG 등급 평가 차이 요인은 안전 그리고 환경적 리스크와 밀접한 관련이 있다.
  - ESG 등급에 공통적으로 영향을 끼치는 단어(risk)를 확인하였으며, 긍부정 단어 측면에서 평가 등급에 끼치는 영향이 긍정적 단어보다 부정적 단어가 크다는 것을 발견하였다.
  - 특히 Tesla는 충돌 사건, Volkswagen에서는 배기가스 치팅 사건이 주요 이슈로 발견
  - 그외에도 clean, protect, strong 등의 단어 발견은 자동차 산업 내에서 ESG 등급 평가에 안전 및 환경이 큰 요인으로 작용하고 있음을 시사
- ESG에서는 기본적인 소비자 신뢰, 제품의 기능 등등을 지키는 것이 중요
  - 워드클라우드를 통해 Volkswagen에서 litigious 용어가 주요하게 출현함을 확인, 그리고 Tesla에서도 다수 출현을 확인
  - 기업별로 상당한 ESG 홍보가 이뤄지고 있지만 trust, right 등의 공통 감성어휘 출현 확인을 통해 기본적인 소비자와의 신뢰가 중요함을 확인
  - ESG 감성분석을 다른 산업군 및 여러 기업들에 활용해 볼 수 있는 기반 마련

## 한계점 및 제언

- 데이터 가용성 측면에서의 한계 경험
  - 아직까지 각 기업들이 ESG 정보를 구체적으로 공개하고 있지 않음
  - 또한 Financial times 검색 엔진의 한계 상존. 다른 데이터 소스도 찾아보면 좋을 것으로 판단됨
- ESG 각 분야에서의 분석 미흡
  - ESG 디렉터리를 따로 구축했음에도 불구하고 ESG 개별 항목에서의 유의미한 차이는 도출하지 못함
  - 아직까지 뉴스에서 전문적으로 ESG 각각을 다루고 있지는 못한 것도 ESG 각 분야 분석의 장애물
- 향후 연구 방향 : 다양한 산업군 및 회사로의 확장
  - 자동차 산업 및 테슬라 vs. 폭스바겐 外 다른 산업 및 업체들로 분석 확대 가능
  - 확대 적용 과정을 통한 보완 과정 등을 통해 더욱 정교한 분석으로 거듭날 수 있을 것으로 기대

감사합니다 😊