

# 삼성전자 주식 등락률 보고서

요인 분류 모델 및 자연어 처리(NLP)





# CONTENTS

## 01. 분석 배경/목표

- 주식 시장에 대한 높아진 관심
- 주식에 영향을 미치는 매우 많은 요인

## 03. 알고리즘 개발

- 주식 예측 분석 자료들의 한계
- 분류 모델의 원리
- 자연어 처리 모델의 원리

## 02. EDA

- 데이터 이해
- 데이터 전처리

## 04. 프로세스 결과

- 활용 방안 / 기대 효과



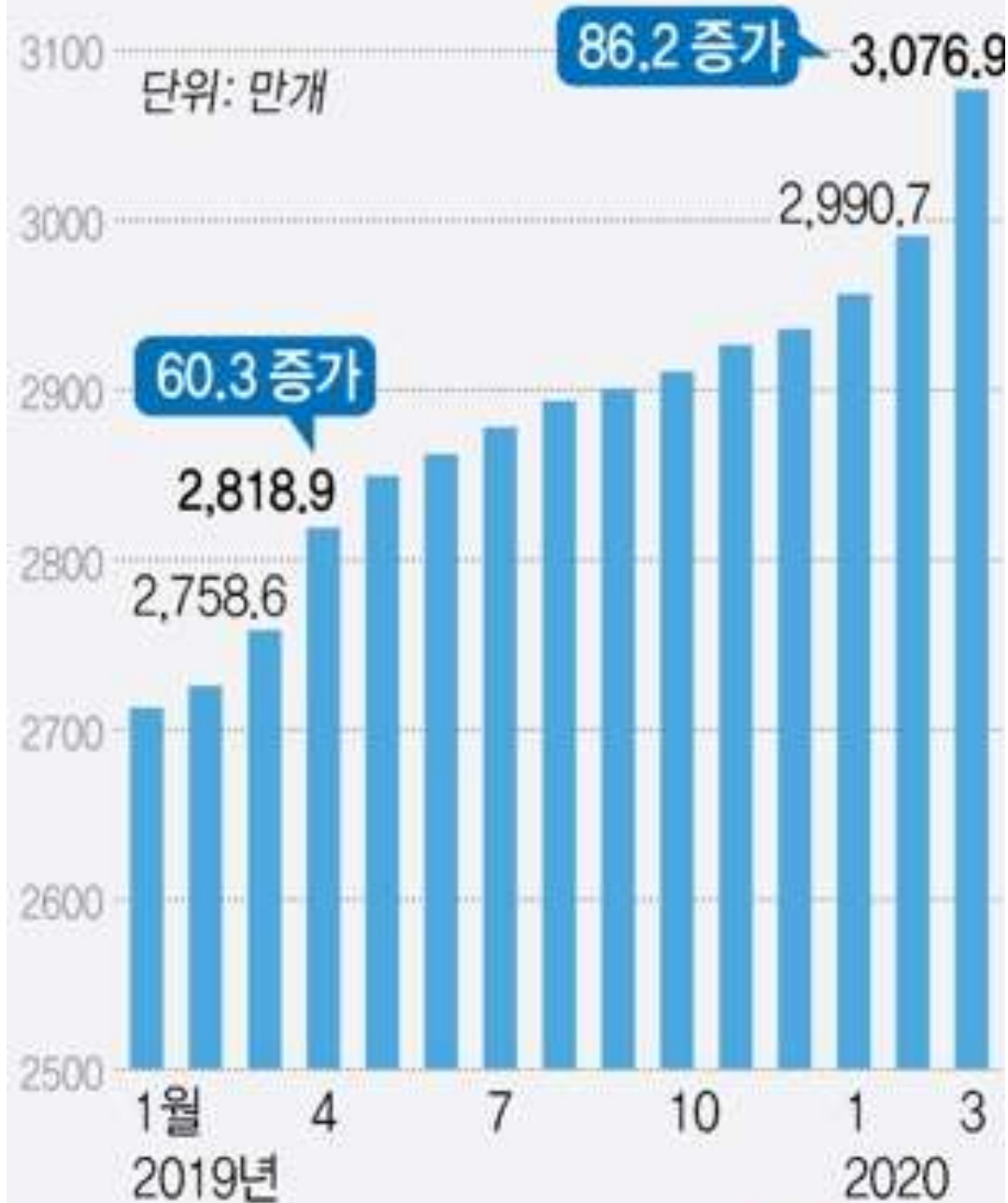
# 01. 분석 배경/목표



- 주식 시장에 대한 높아진 관심
- 주식 예측 분석 자료들의 한계
- 주식에 영향을 미치는 매우 많은 요인

# 주식 시장에 대한 높아진 관심

주식거래 활동계좌 수 추이



자료/ 금융투자협회 연암뉴스

개인 월별 순매수 규모



자료: 한국거래소·유가증권시장

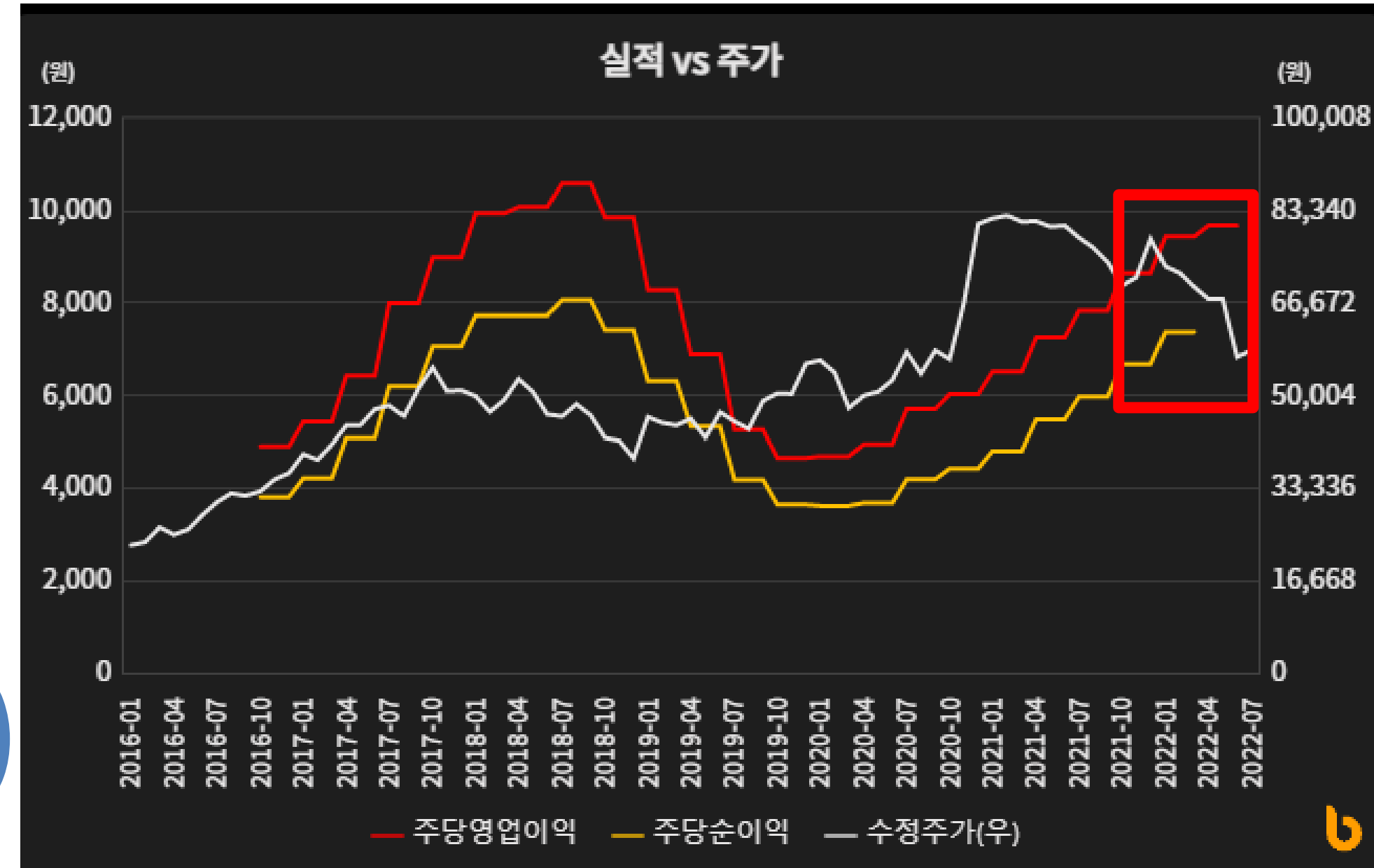
# 주식 예측 분석 자료들의 한계

주가 및 주식투자수익률의 변동요인

구분		변동요인
시장전체 변동요인	경제적 요인	경기, 물가, 금리, 재정수지, 무역수지, 환율, 기술혁신, 해외요인 (경기, 증시, 금리, 자금이동 등)
	경제 외적 요인	국내정세, 국제정세
	주식시장 내부요인	투자자 동향, 신용거래 규모, 차익거래 규모, 규제변화, 지분이동
기업의 개별요인		증자, 감자, 인수합병, 배당정책, 임원인사, 신제품 개발, 생산/수익 동향, 주주구성 변화 등

# WHY?

\*



\* 출처 : butler, [https://www.butler.works/reports/report?corp\\_code=00126380](https://www.butler.works/reports/report?corp_code=00126380)

<https://codingcoding.tistory.com/1066>



# 예측 불가능한 변수들



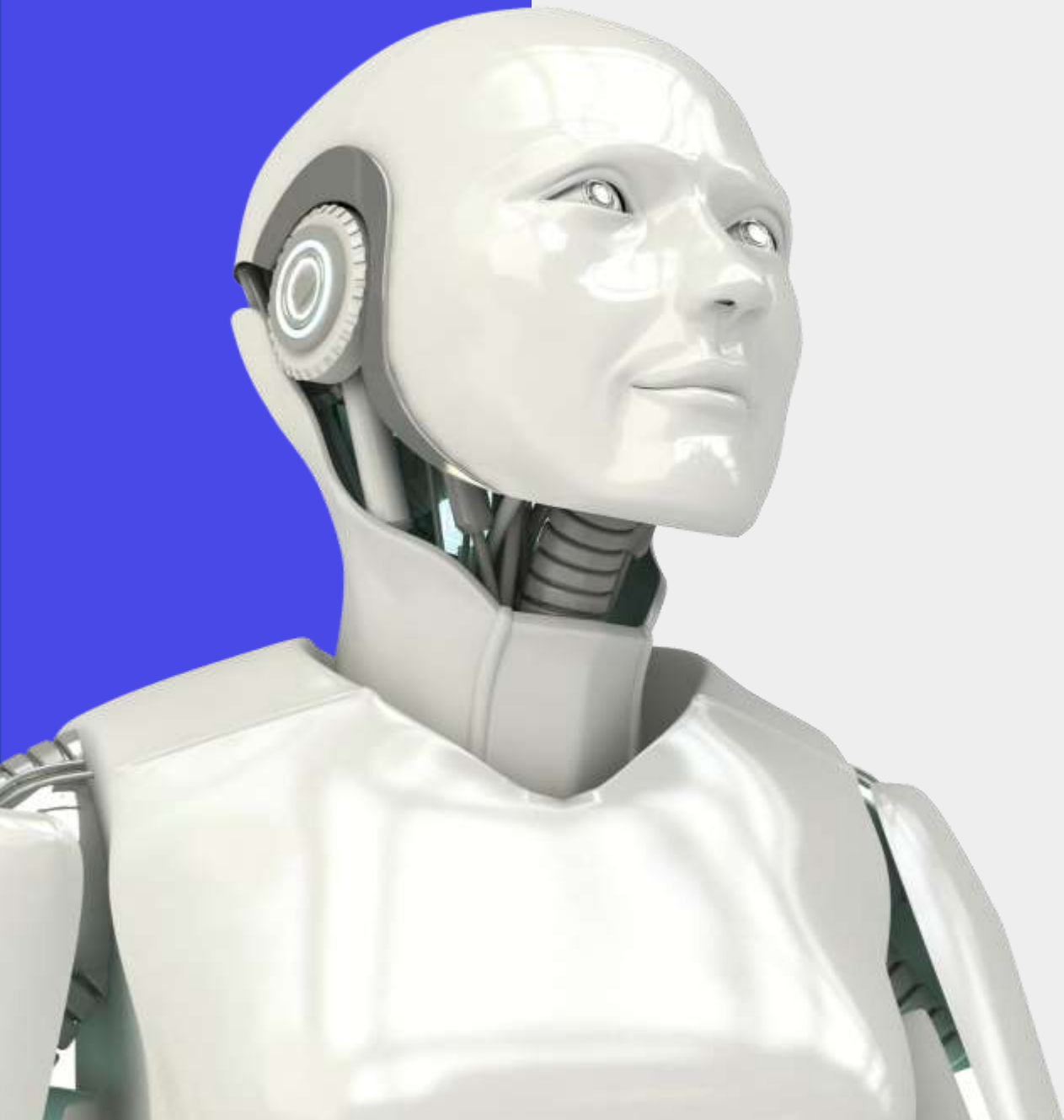


# 예측 불가능한 변수들





## 02. EDA

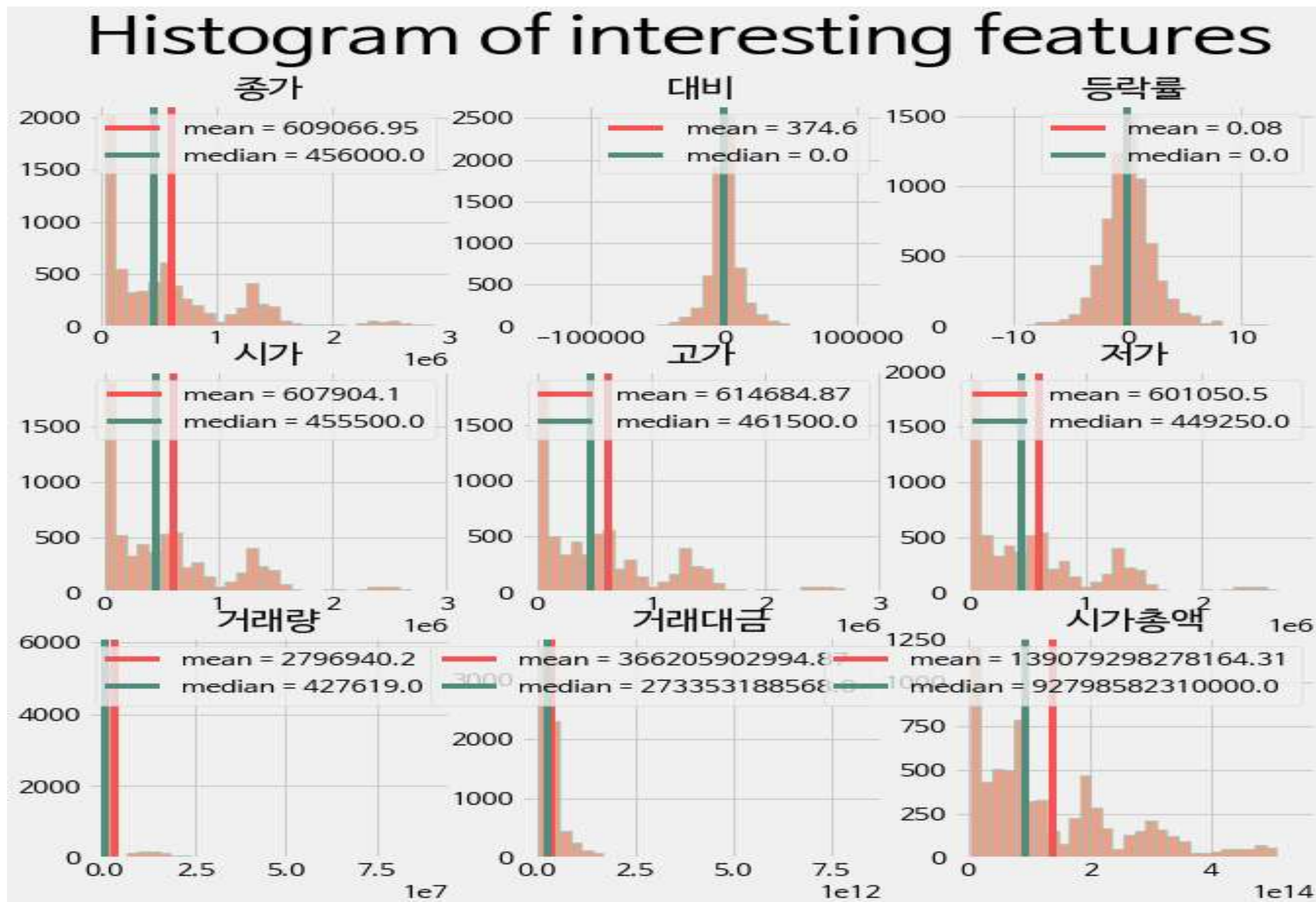


1. 데이터 이해
2. 데이터 전처리
3. 데이터 EDA

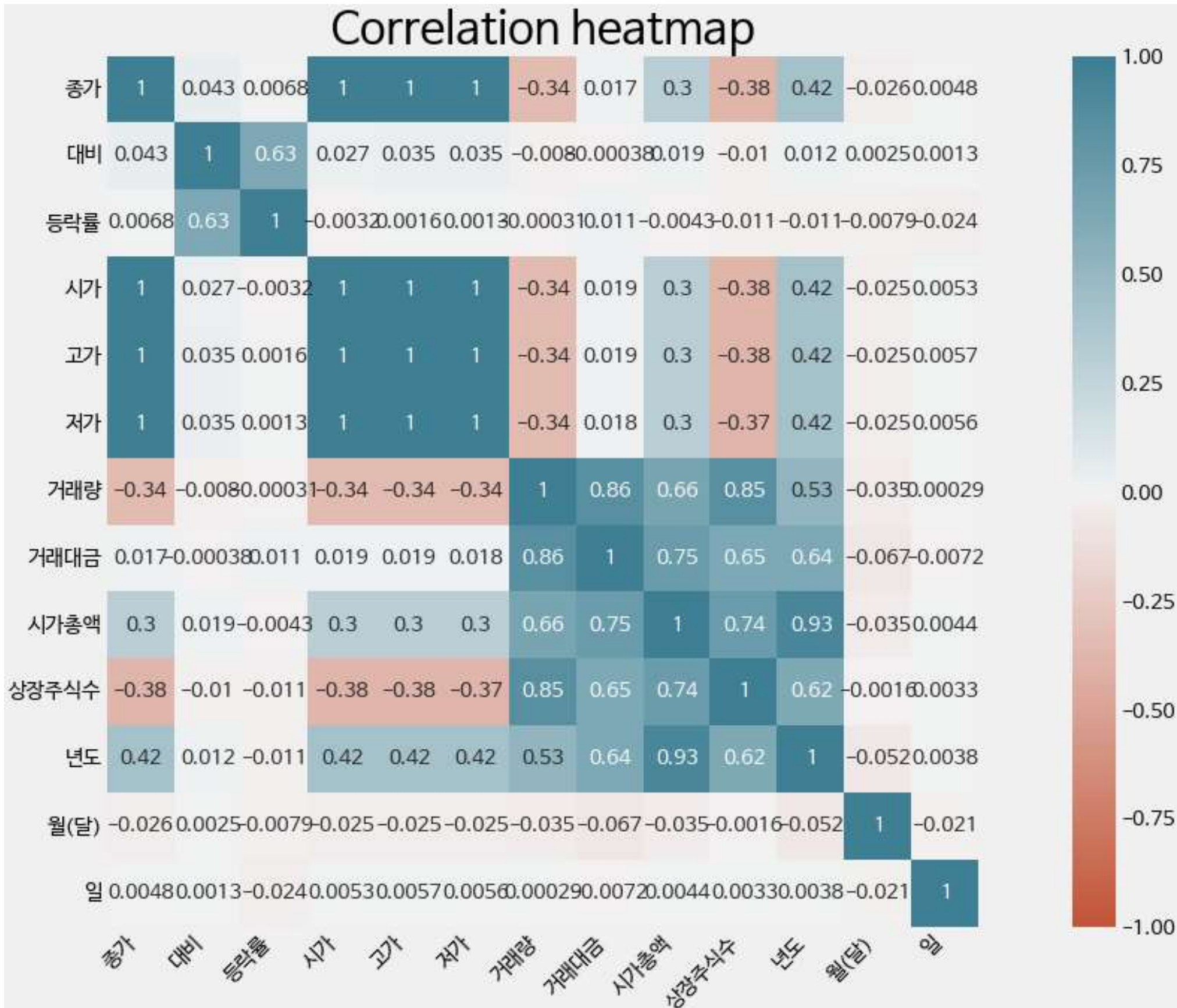


# 데이터 이해

- Features : 종가, 대비, 시가, 고가, 저가, 거래량, 거래대금, 시가총액
- Target : 등락률



# 데이터 EDA



1. 등락률과 대비는 0.63으로 어느정도  
연관성이 있는 상관 관계를 가지고 있습니다.

2. 총가와 시가, 고가, 저가는 같은 데이터로  
상관분석을 하여도 무방합니다.

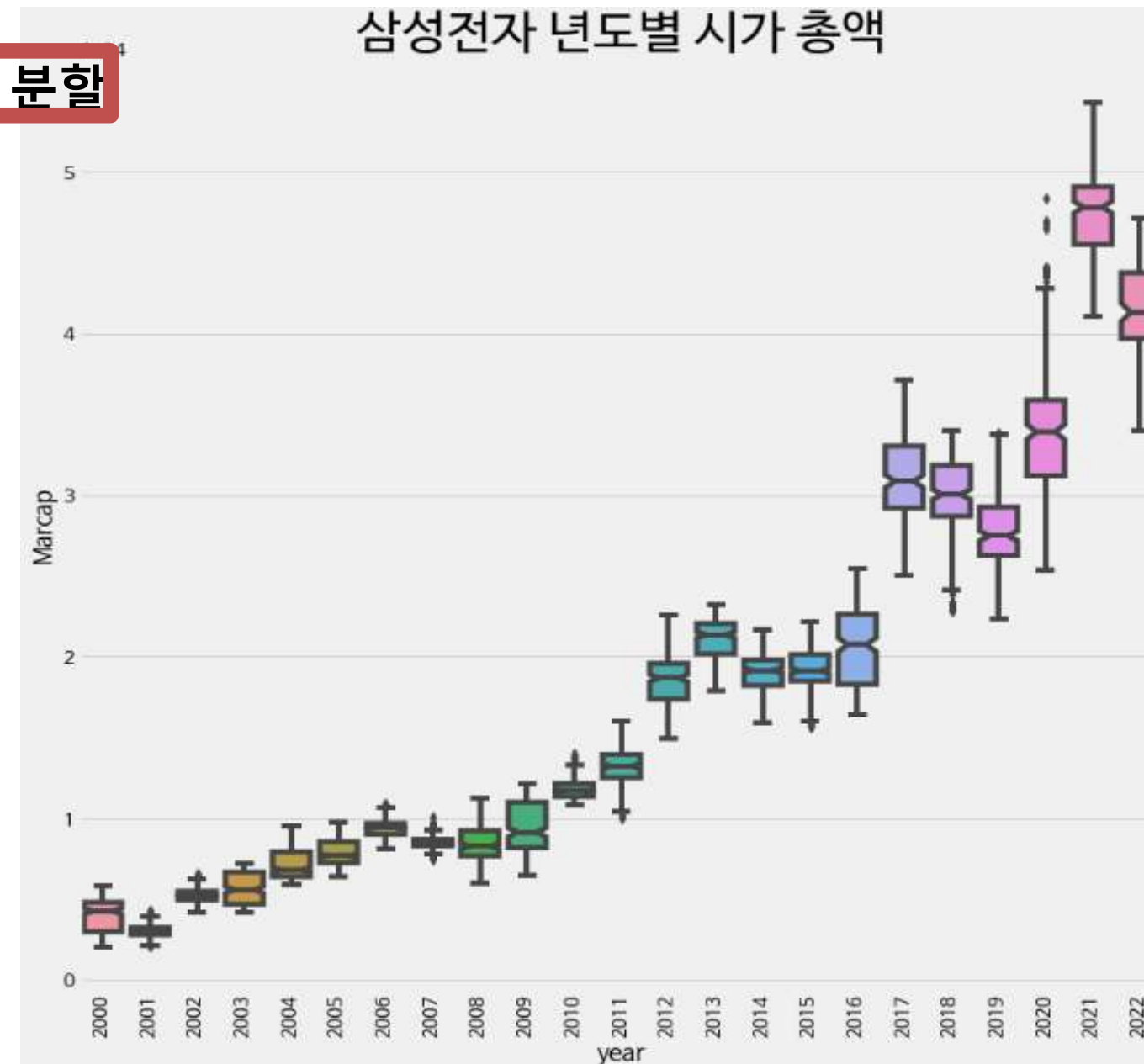
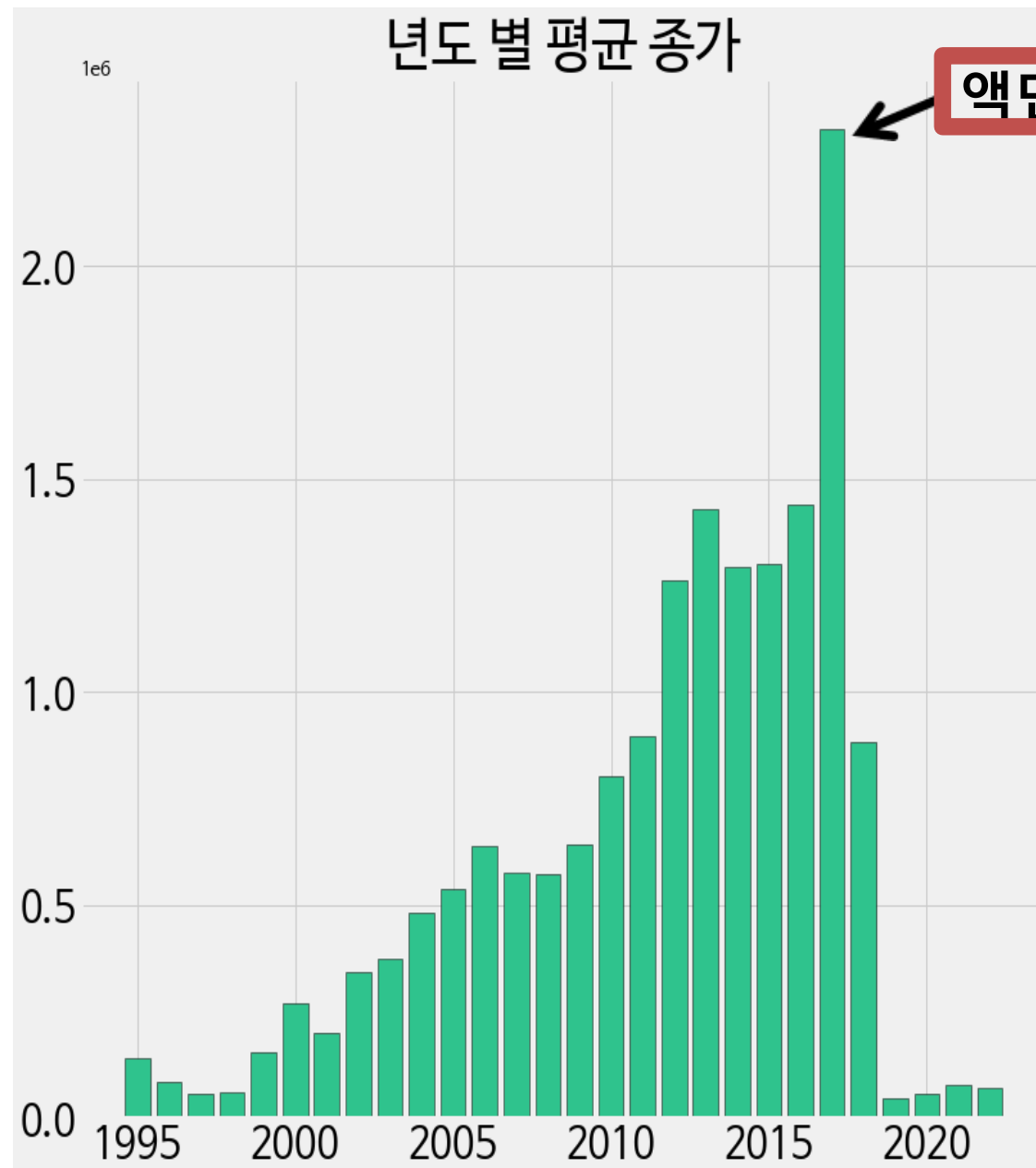
3. 거래량이 많아질수록 거래대금이 커지는  
것이 0.86으로 강한 비례 관계임을 알 수  
있습니다.

4. 연도가 흐를수록 시가총액은 커집니다.

5. 상장 주식 수가 많을수록 시가총액이  
커집니다.



# 데이터 EDA



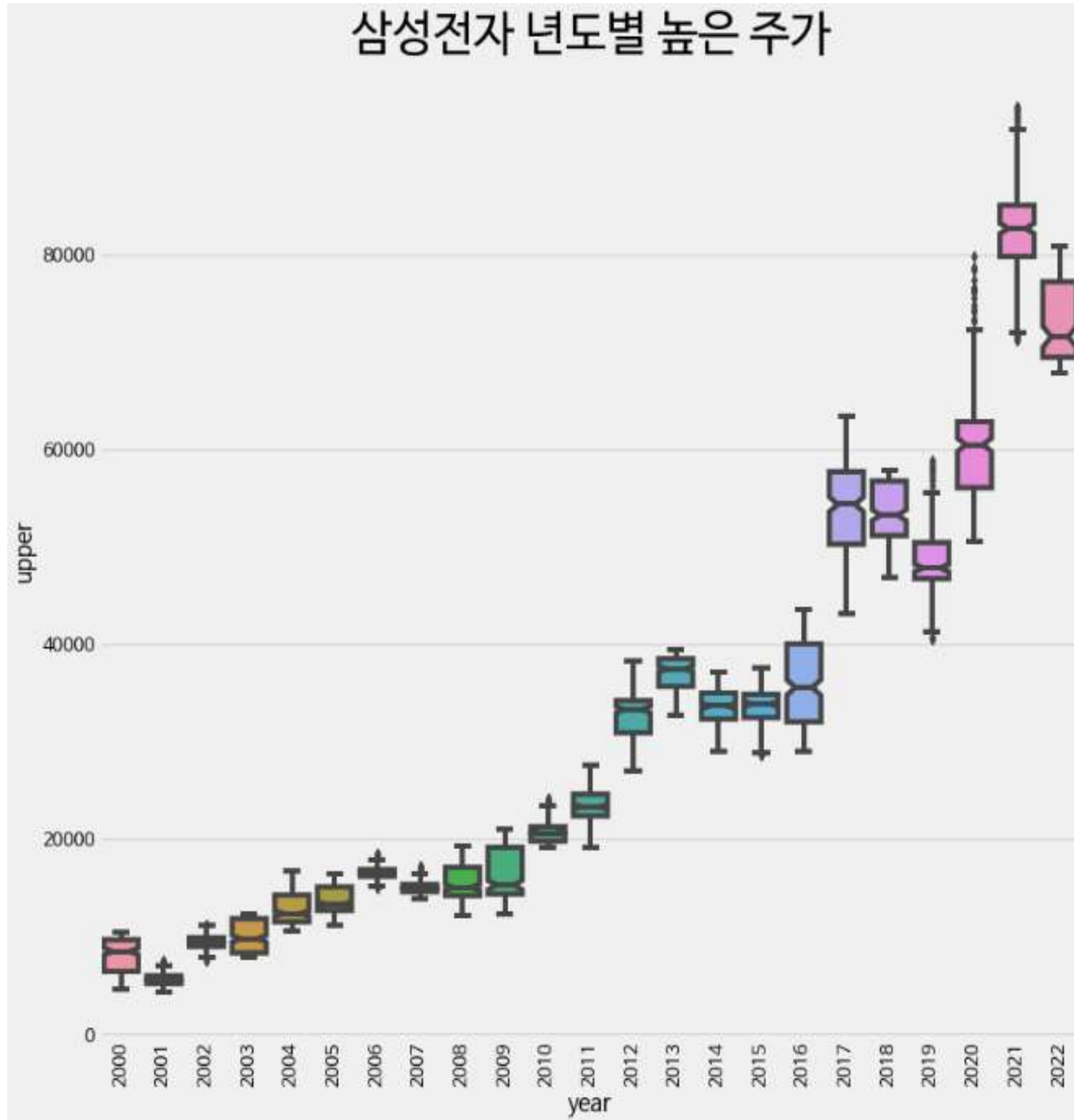
1. 삼성전자는  
2018.05.04일 액면 분할을  
실시 하였습니다.

2. 그래프에서는  
삼성전자의 시가 총액을  
연도가 지남에 따라  
비례하게 성장하는  
우상향을 보여 주고  
있습니다.

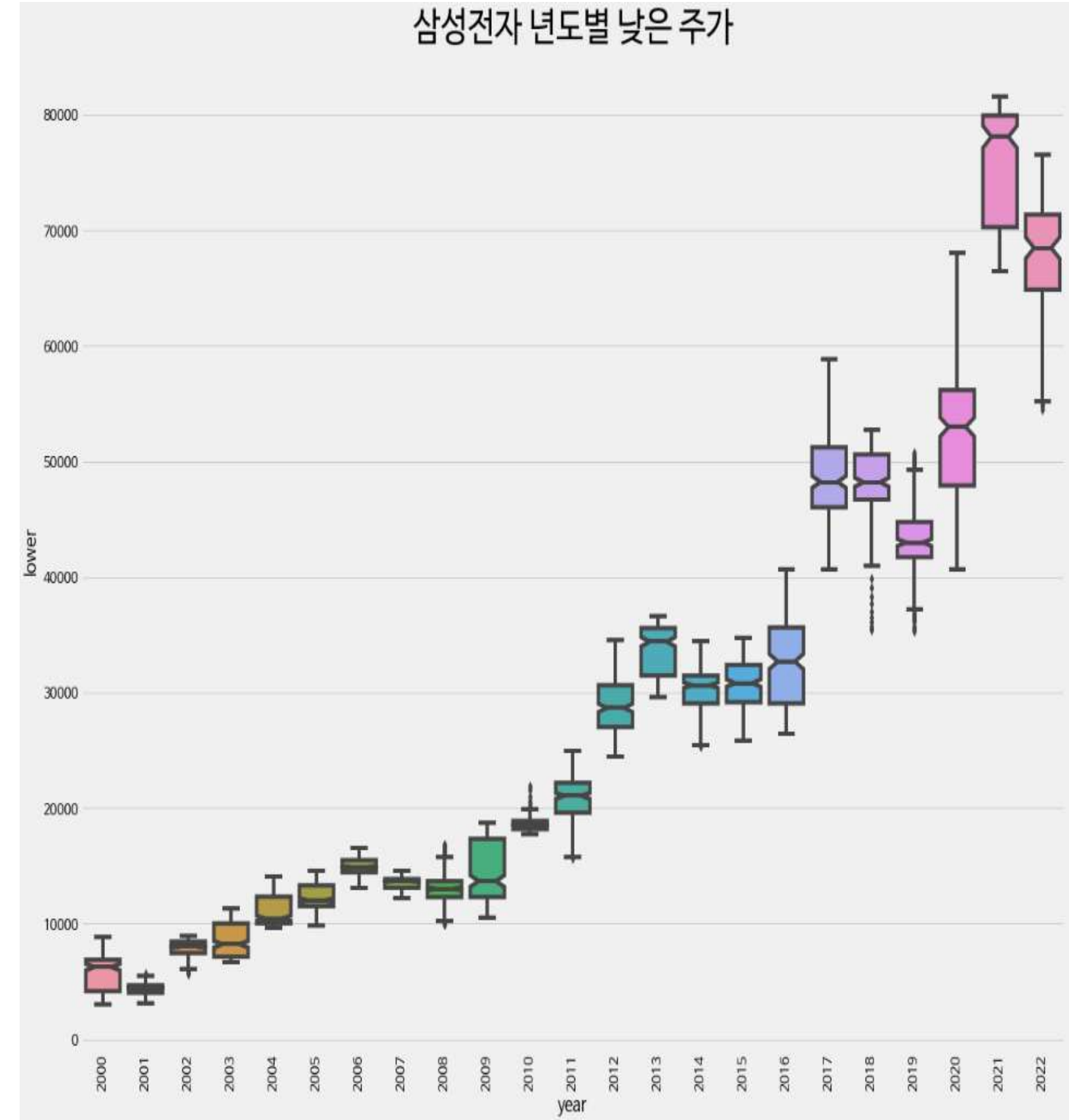
3. 2020년 삼성전자  
시가총액의 분산이 가장  
큰 것을 알 수 있습니다.

# 데이터 EDA

삼성전자 년도별 높은 주가



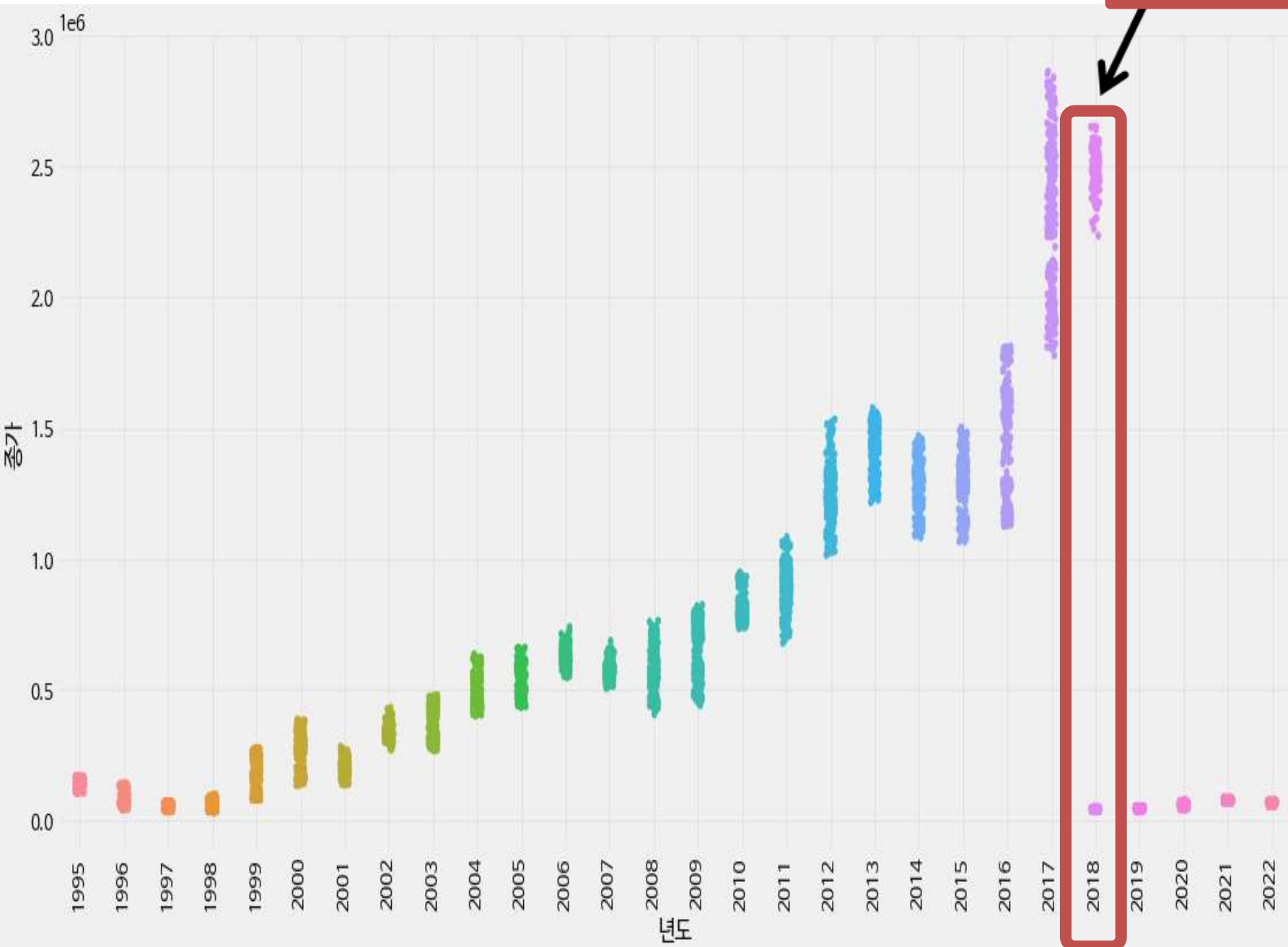
삼성전자 년도별 낮은 주가



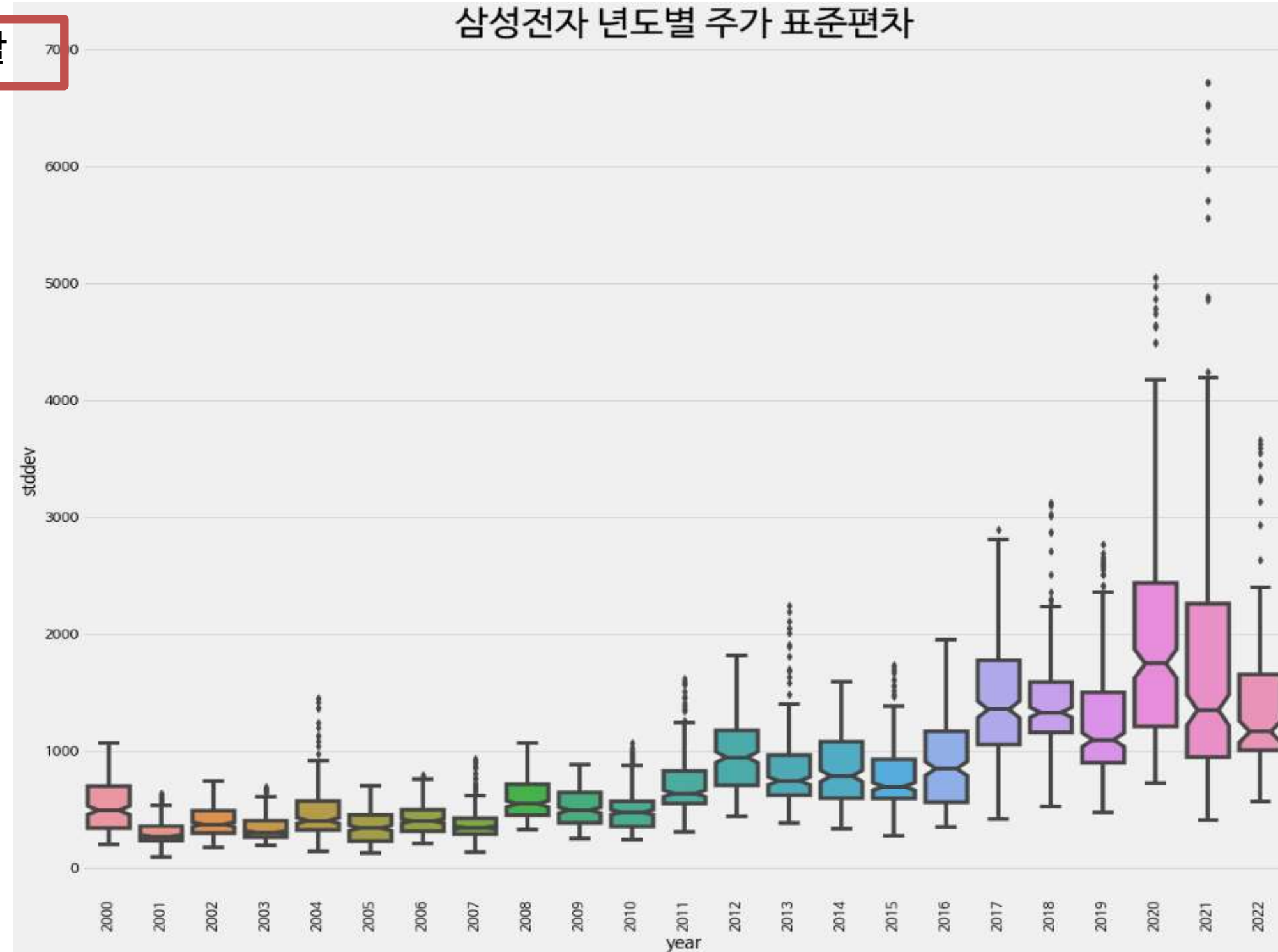


# 데이터 EDA

삼성전자 년도 별 증가 범위

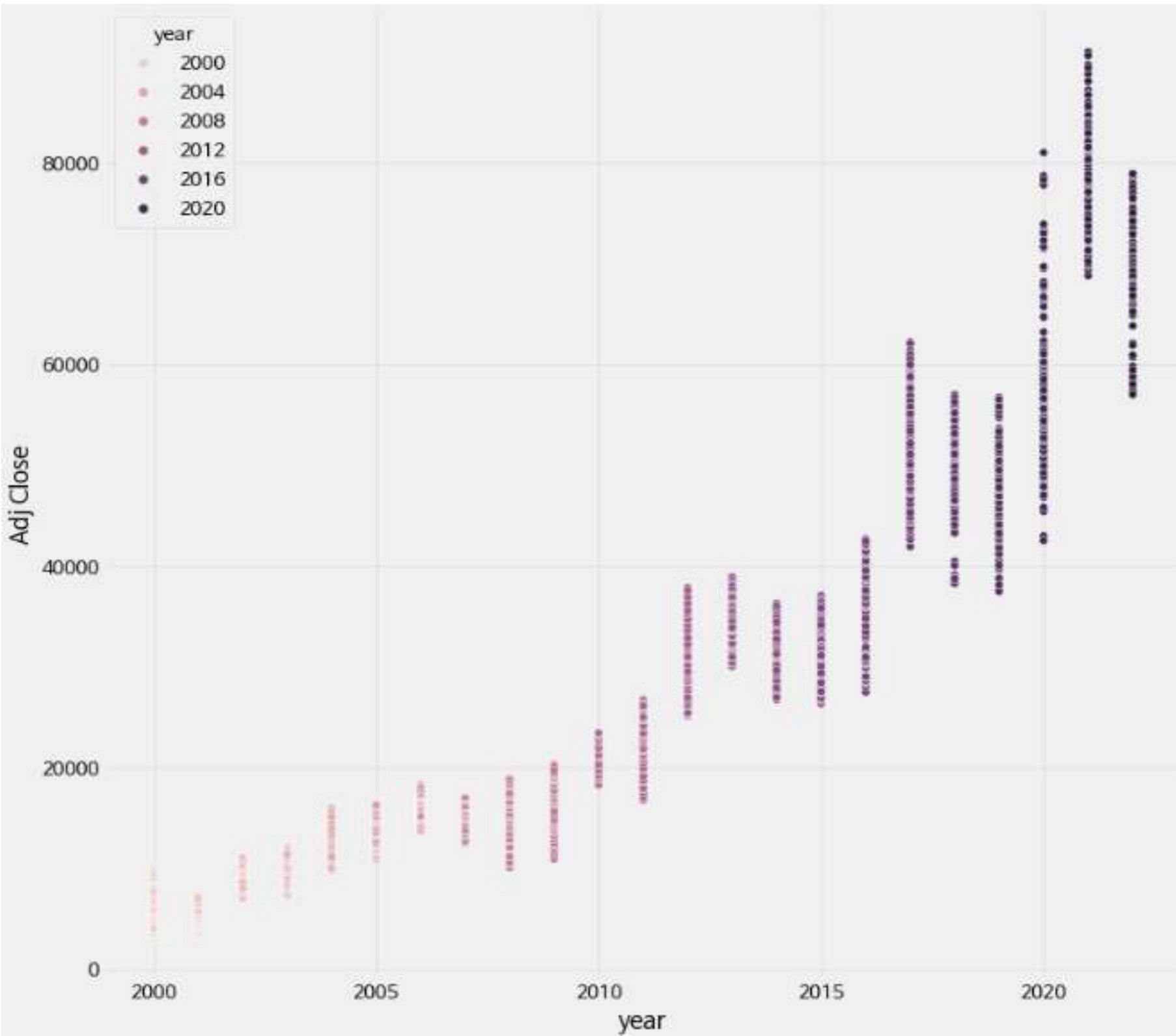


삼성전자 년도별 주가 표준편차

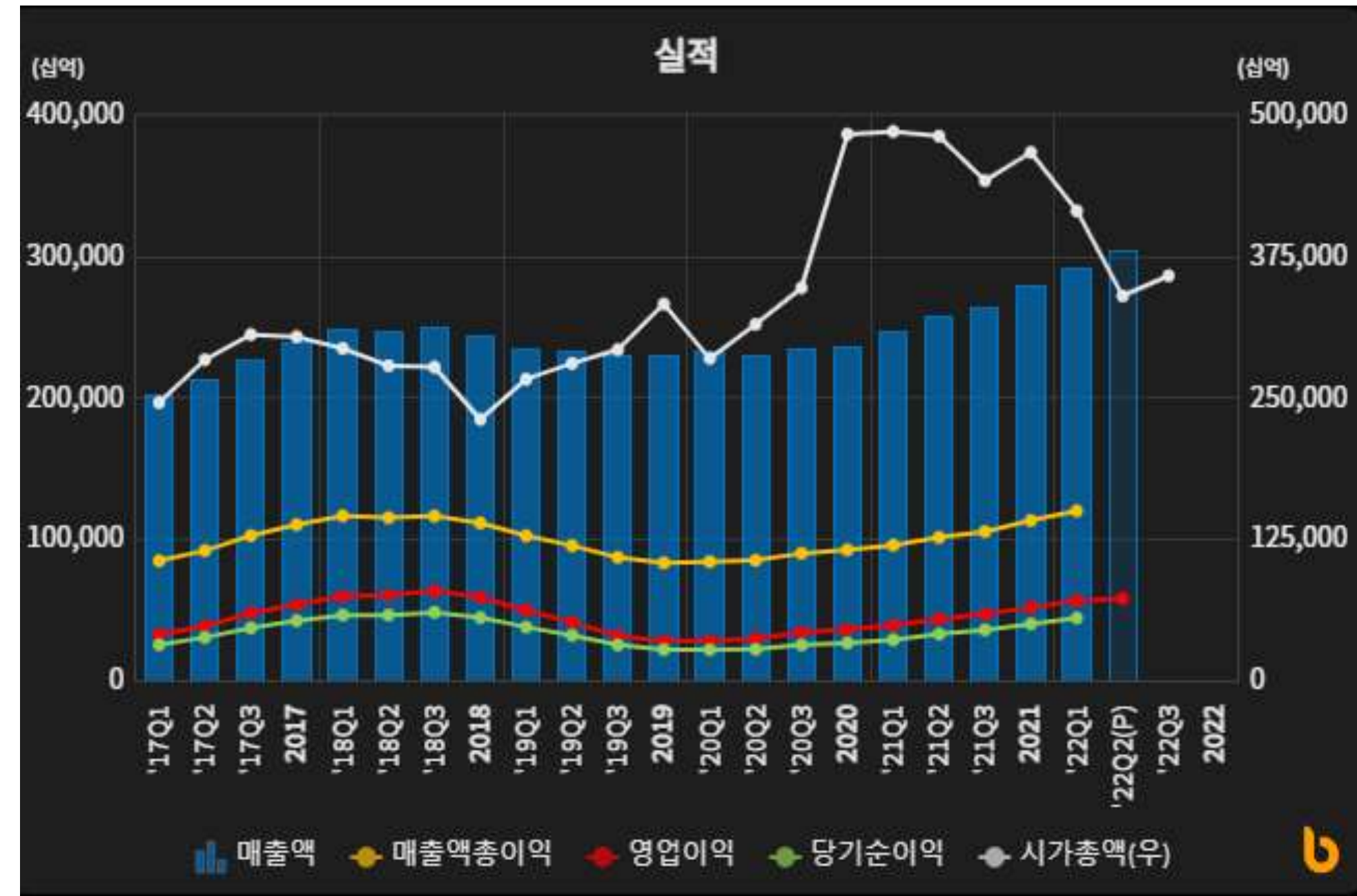


# 데이터 EDA

삼성전자 년도별 조정 종가



\*



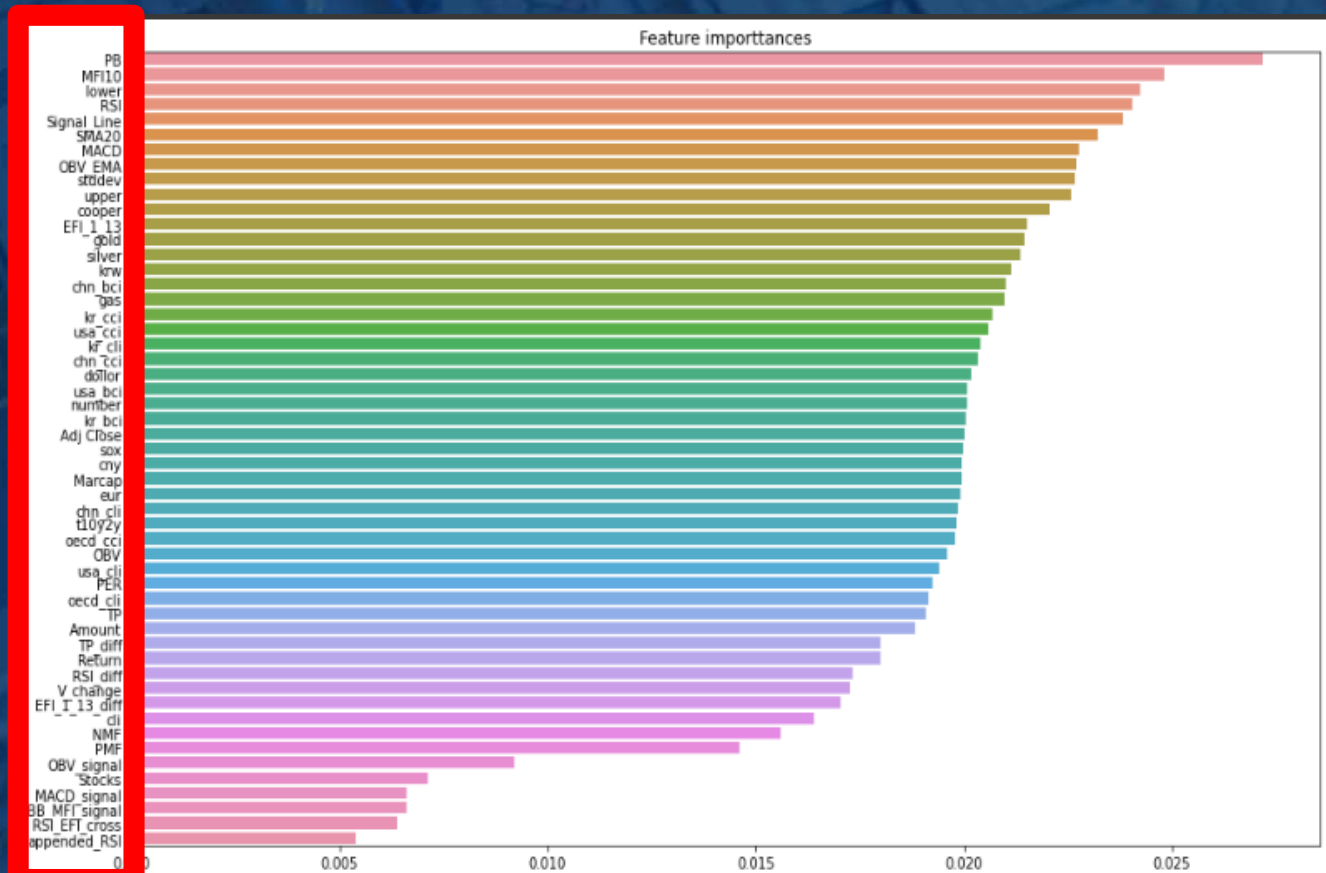


### 3. 알고리즘 개발

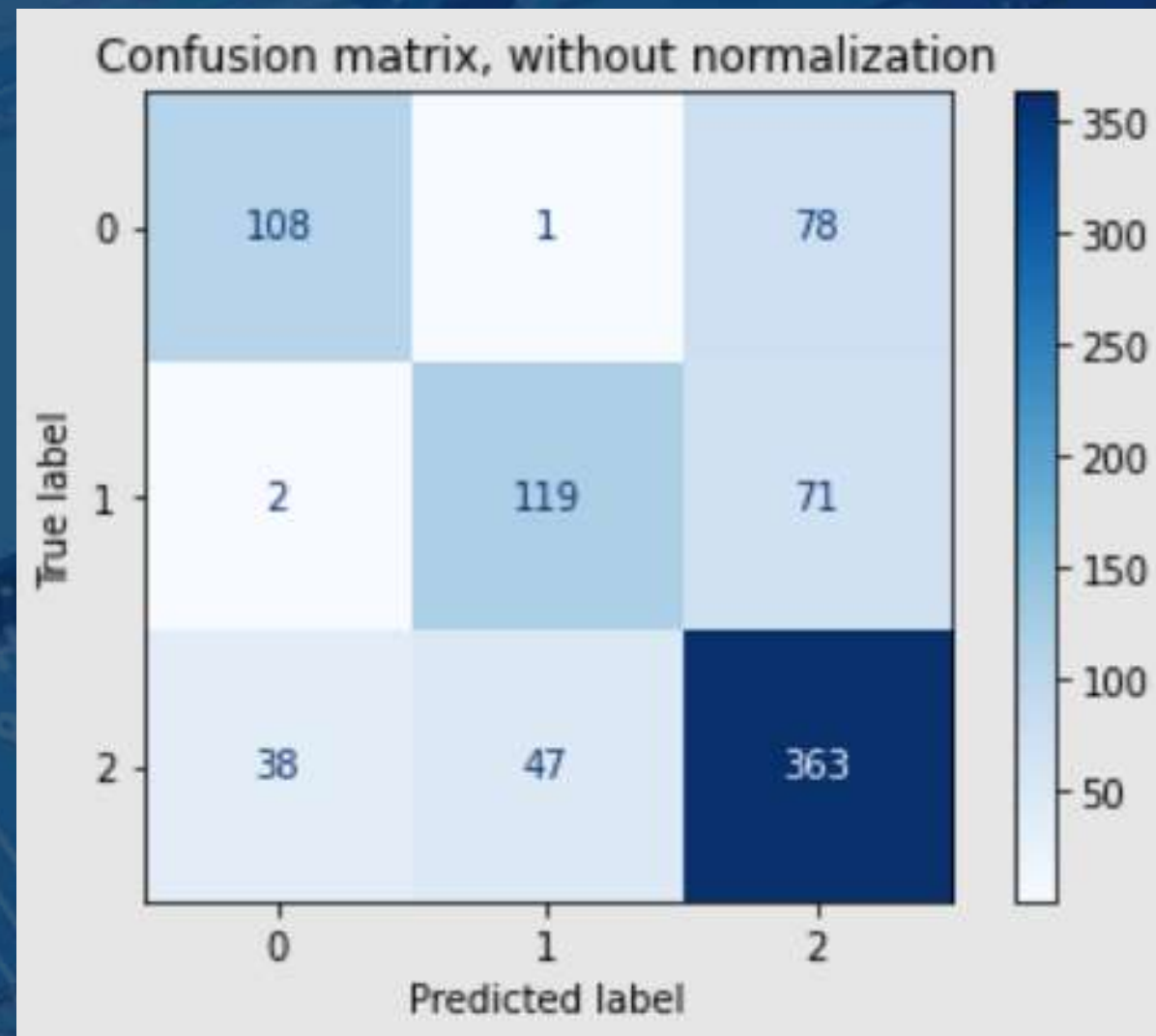


- 분류 모델의 원리  
- 자연어 처리 모델의 원리

# 분류 모델



중요한 피쳐



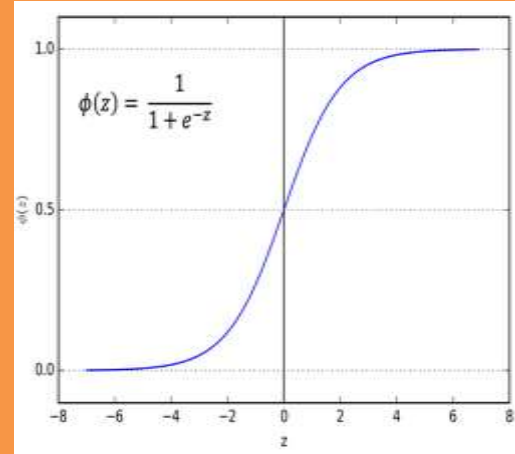
model\_acc = 0.713  
model\_0 acc (하락) = 0.729  
model\_1 acc (중립) = 0.712  
model\_2 acc (상승) = 0.708



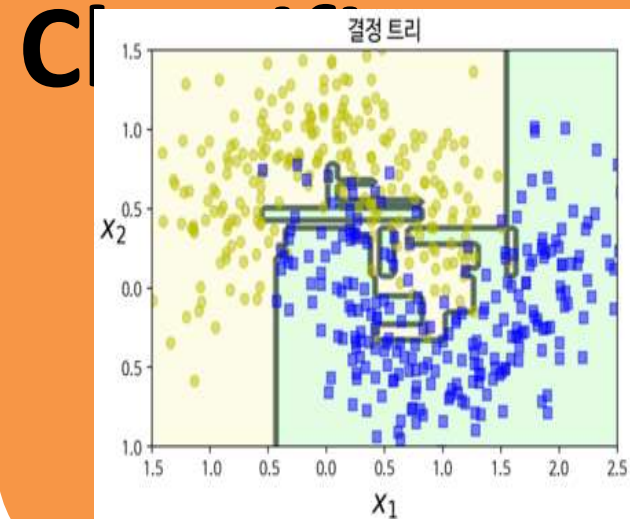
# 분류 모델

등락률 (%)
-1.72
-2.36
1.02
0.68
1.74
-0.35
-1.54
-0.34
-1.84
-1.81
0.33
-1.94
-0.32

Sigmoid 함수



ExtraTrees



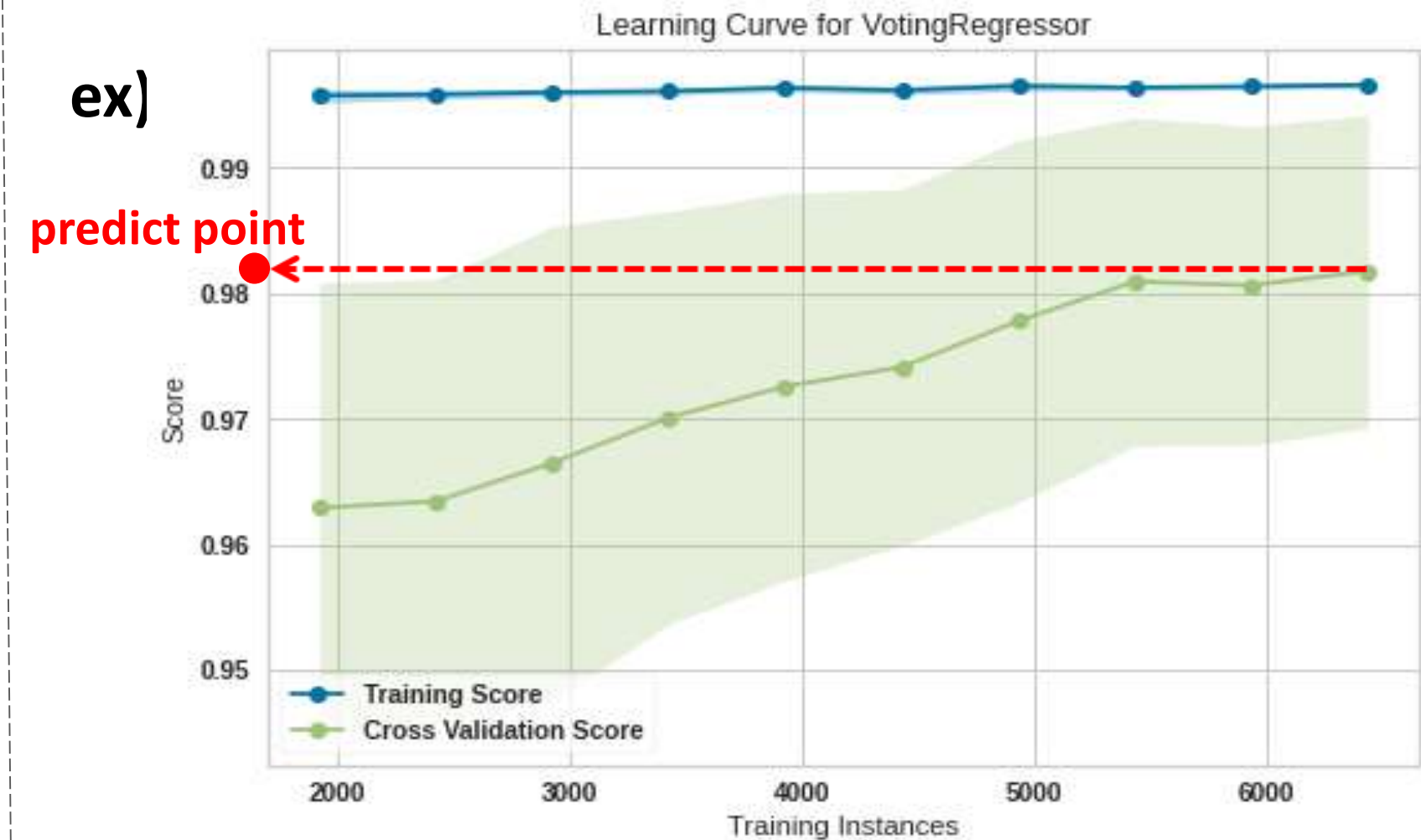
↑ : 1

↓ : 0

Expected value

\* 상승 / 하락 예측 정확도

ex)



# 자연어 처리 감정 분석

Unnamed: 0	date	tokens	event Ret	sent_event
0	0 2000-01-04	올해 상위 주요 그룹 일제 사무 새해 출발 징표 올해 경상 이익 상위 사표 인터넷 ...	0.000000	0.002268
1	1 2000-01-05	주가 급등 장외 종목 관심 폭발 새천년 증시 출발 세계 최고 새천년 증시 출발 오늘...	0.000000	0.001925
2	2 2000-01-06	분기 조정 국면 회복 가주 군단 이모저모 디지털 장선 가전 경쟁 가열 새해 게임 제...	-0.093809	-0.000272
3	3 2000-01-07	주가 이틀 폭락 안팎 쌍방울 야구단 인수 리노 공업 채운 사장 쉐러리맨 엑소더스 스...	0.034450	0.001517
4	4 2000-01-10	표준 협회 밀레니엄 프론티어 대상 제품 선정 국내 연구 개발 투자 기업 삼성 쉐러리...	0.027281	0.002349
...	...	...	...	...
5545	5545 2022-06-24	반도체 격차 인텔 세계 삼성 돌격 종목 매수 관망 바닥 고심 개미 레이더 정부 경제...	0.020799	0.001963
5546	5546 2022-06-27	정의선 장녀 결혼식 재계 총수 출동 삼성 스마트 시티 비대 봉사 활동 통해 상당 물...	0.034475	0.002563
5547	5547 2022-06-28	눈물 개미 종간 배당 막차 타자 배당 수익 수도 이재용 베르사체 원피스 가격 코스피...	-0.006516	0.001349
5548	5548 2022-06-29	임금 인상 자제 물가 상승 노동자 책임 반발 추경호 발언 상무 한미 경제 협력 삼성...	-0.030606	0.000681
5549	5549 2022-06-30	매직 삼성 가전 결합 스페셜 렌탈 서비스 인기 품목 확대 반도체 쇼크 코스피 코스닥...	0.000000	0.001430



# 자연어 처리 감정 분석

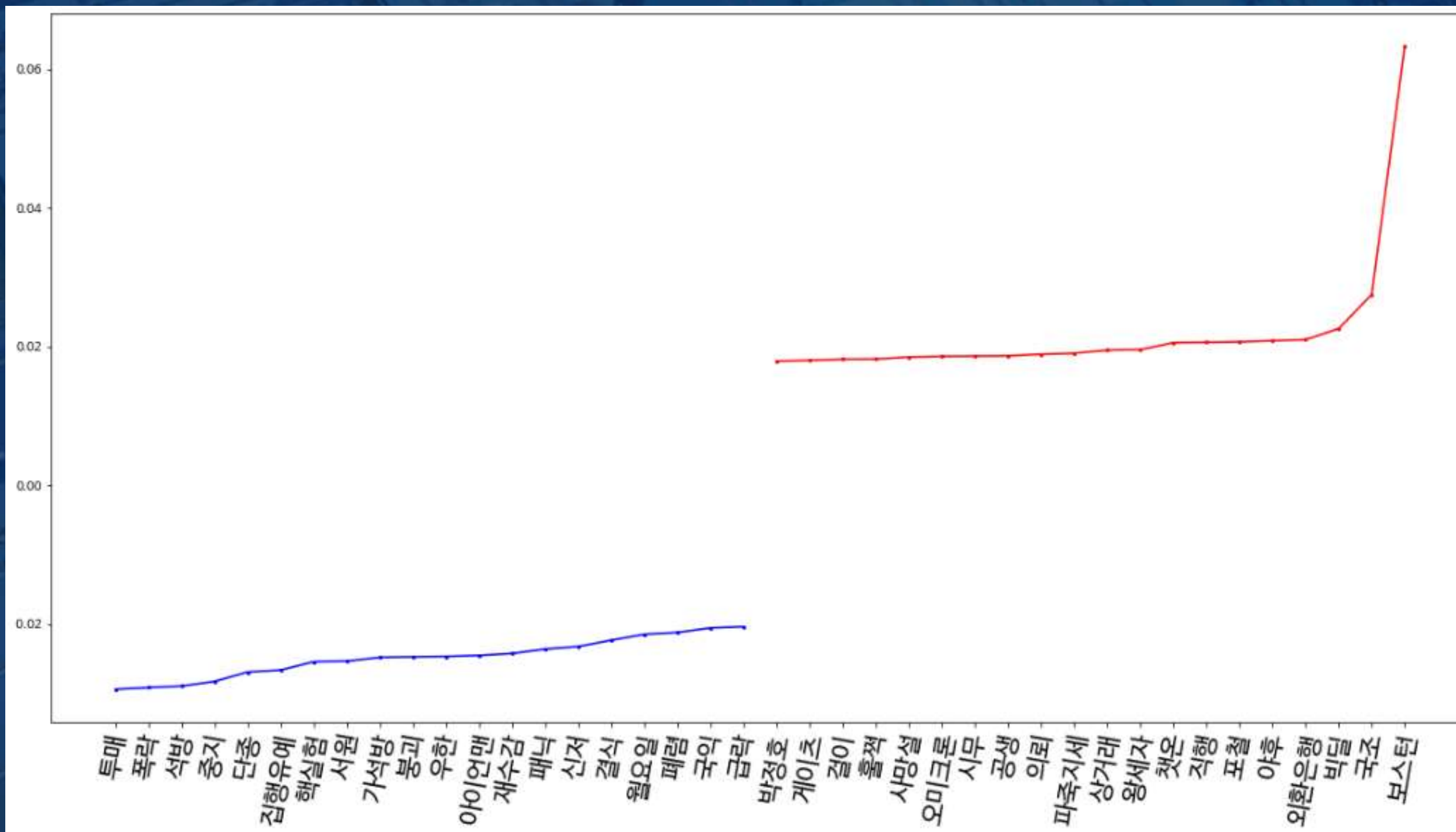


# 자연어 처리를 통하여 회사와 연관성이

있는 키워드들을 얻어 낼 수 있습니다.



# 자연어 처리 감정 분석

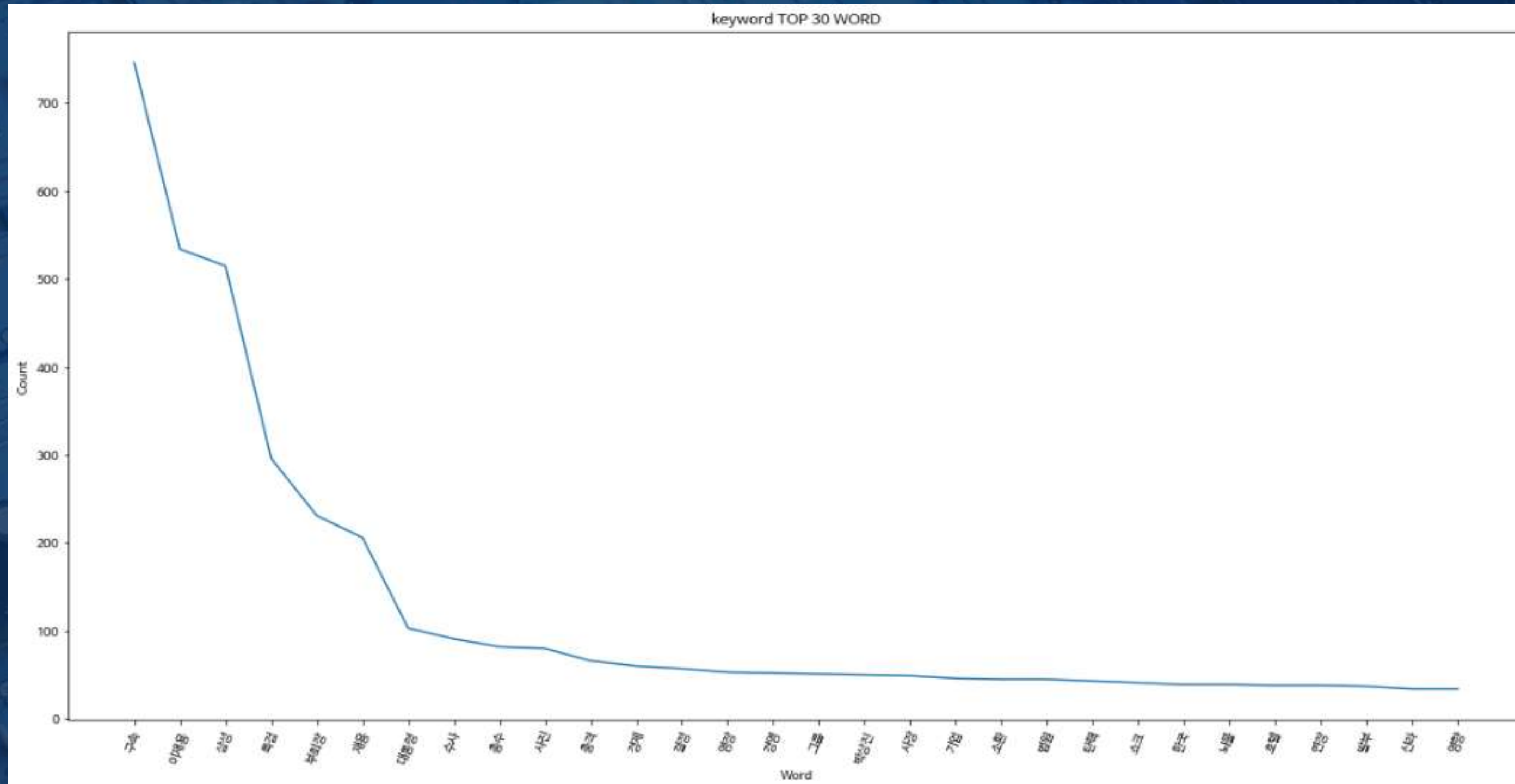


긍정 요인(빨간색), 부정 요인(파란색) Top 20 < 감정점수 기준 >

업종 데이터를 통해 기업의 폭넓은 뉴스데이터를 적용할 수 있습니다.

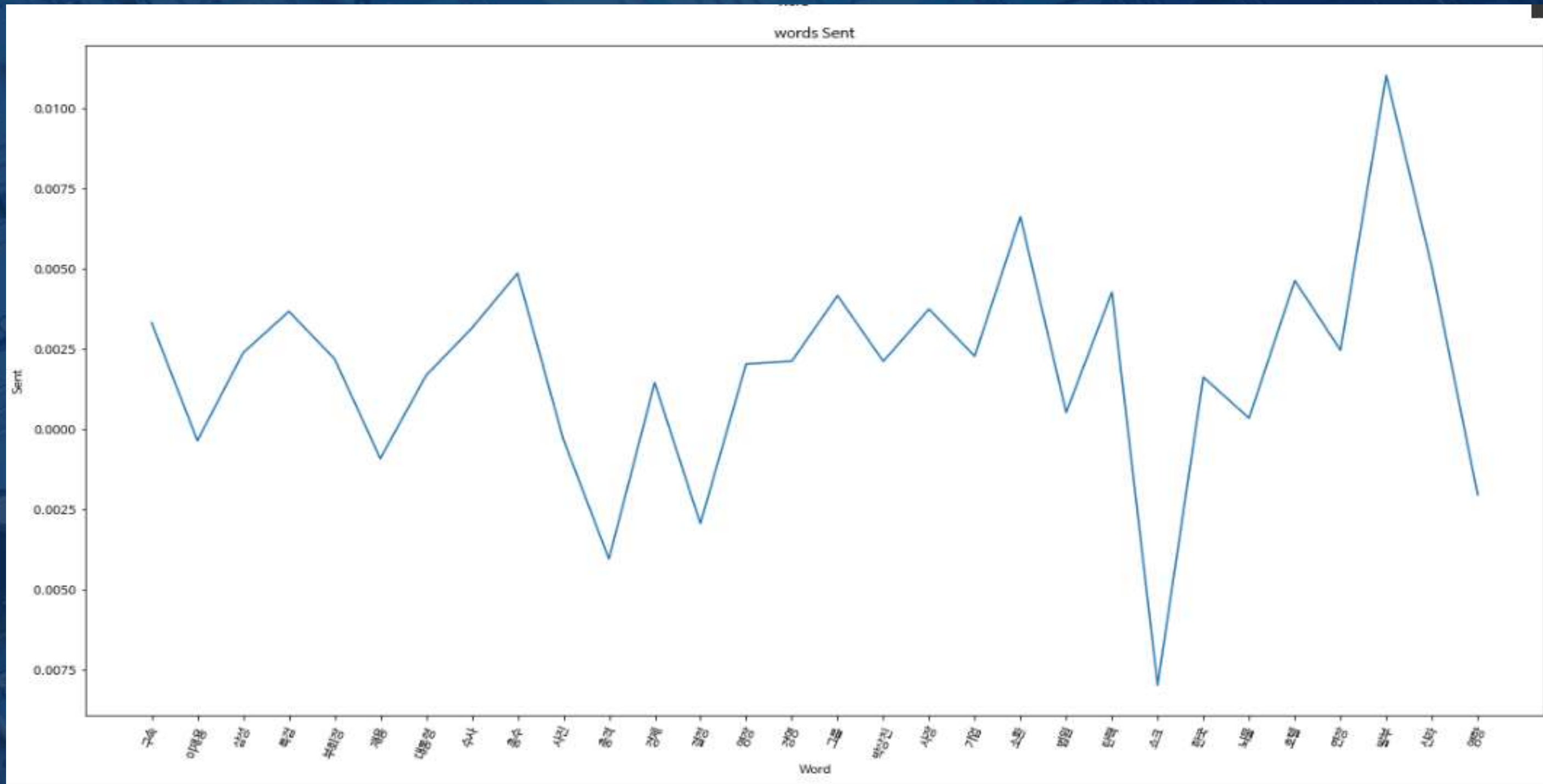


# 자연어 처리 감정 분석



일별 단어 빈도 수 ( 2017.02.17)

# 자연어 처리 감정 분석



## 단어의 감성 점수 그래프 (2017.02.17)



## 04. 프로세스 결과



- 활용 방안 / 기대효과

# 활용 방안 / 기대 효과



\* 출처 : <https://m.blog.naver.com/tradepeople/221760337868>



# 활용 방안 / 기대 효과

