



뉴스를 열심히 보면 주식을 잘할 수 있을까

주가와 증권뉴스 감성 상관관계

금준호, 김란, 노두호, 안지윤, 최규진



▶ 목차

주가와 증권뉴스 감성
상관관계

- 프로젝트 개요
- 데이터분석
- 상관분석 및 시각화
- 결론 및 의의

1

프로젝트 개요



주가와 증권뉴스 감성 상관관계

Google 학술검색

뉴스 감성 주가 상관관계

검색결과 약 2,740개 (0.08초)

학술자료

모든 날짜

2023 년부터

2022 년부터

2019 년부터

기간 설정...

관련도별 정렬

날짜별 정렬

모든 언어

한국어 웹

모든 유형

검토 자료

☐ 특허 포함

☐ 서지정보 포함

☒ 알림 만들기

[PDF] 뉴스와 주가: 빅데이터 감성분석을 통한 지능형 투자사결정모형
김유신, 김남규, 정승렬 - 지능정보연구, 한국지능정보시스템학회, 2012 - koreascience.or.kr
... 를 가지고 있었으며, 좀 더 세부적으로는 주식시장 개장 전 뉴스들과 주가지수의 등락과의 관계 또한 통계적으로 유의하여, 뉴스의 감성분석 결과를 이용해 주가지수의 변동성 예측이 가능할 ...
☆ 저장 99 인용 40회 인용 관련 학술자료 전체 4개의 버전

빅데이터에 나타난 감성 분석
이득환, 강형구, 김수현, 이창민 - 金融工學研究, 2013 - scholarworks.bwise.kr
... 뉴스의 감성 분석 결과와 주가지수 등락은 유의미한 관계가 있으므로, 뉴스의 감성분석 결과 를 이용하여 주가지수의 변동성 예측이 가능할 것으로 판단되었다. 한편 정청원, 김수경(2009)은 ...
☆ 저장 99 인용 8회 인용 관련 학술자료 전체 47개의 버전

[HTML] 아파트 매매가격과 부동산 온라인 뉴스의 교차상관관계와 인과관계 분석:
온라인 뉴스 기사의 비정형 빅데이터를 활용한 감성분석 기법의 적용
박재수, 이재수 - 국토계획, 2019 - kpaj.or.kr
... 감성분석 모형에 사용될 감성지수는 수집된 뉴스를 지역별, 일자별, 매체, 뉴스 유형별로 분류 한 후 뉴스 내용을 일련의 언어처리 과정인 형태소 분석 및 의존 구문 분석을 통해 어휘와 의미의 ...
☆ 저장 99 인용 5회 인용 관련 학술자료 전체 4개의 버전

BERT 를 활용한 뉴스 감성분석과 거시경제지표 조합을 이용한 주가지수 예측
장은아, 최희현, 이홍철 - 한국컴퓨터정보학회논문지, 2020 - dbpia.co.kr
... 뉴스 정보의 감성 분석은 최신 자연어처리 기법인 BERT와 NLTK VADER를 사용하고, 예측모델 은 주가예측모델로 적합하다고 알려진 딥러닝 예측모델 LSTM을 적용하여 가장 효과적인 지표 ...
☆ 저장 99 인용 3회 인용 관련 학술자료 전체 3개의 버전

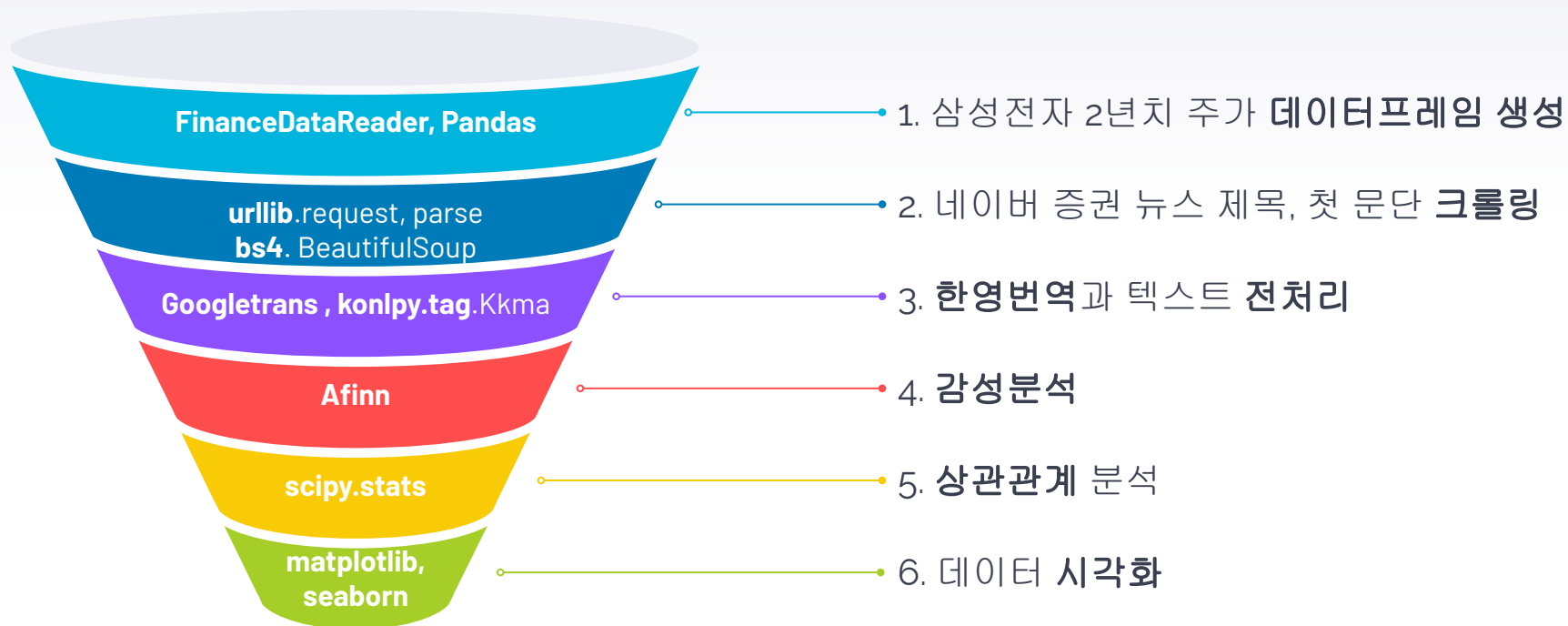
"그 결과 뉴스 콘텐츠의 감성분석 결과값과 주가지수 등락과는 유의한 관계를 가지고 있었으며, 좀 더 세부적으로는 주식시장 개장 전 뉴스들과 주가지수의 등락과의 관계 또한 통계적으로 유의하여, 뉴스의 감성분석 결과를 이용해 주가지수의 변동성 예측이 가능할 것으로 판단되었다"

뉴스와 주가: 빅데이터 감성분석을 통한 지능형 투자사결정모형

The count of the positive and negative sentiment of news articles for each day and variance of adjacent days close price along with historical data is used for prediction purpose and **an accuracy ranging from 65.30 to 91.2 % achieved with various machine learning techniques.**

Efficacy of News Sentiment for Stock Market Prediction (2019)

데이터 분석 계획



2 데이터 분석



데이터 분석 모델

a. 종가 높고 낮은 30일 직전 3일 뉴스 분석

2년 내 주식 최고가, 최저가
찍기 직전 3일 뉴스
감성분석

b. 수익률 큰 30일 직전 3일 뉴스 분석

주가 자체보다
많이 오르고 내린, 변동량이
큰 기간이 더 상관관계가 클
것이라 가설

c. 당일 뉴스 감성분석

365일 뉴스 감성 분석값
과 수익률과의 상관관계
분석

1. 뉴스가 주가에 반영되는 데에 3일 걸린다고 가정
2. 유의미한 표본양을 30주로 계산

1. 주가 데이터프레임 생성

```
def Finace_data(code, name):  
    # 데이터 얻기  
    data = fdr.DataReader(code,  
        start='2021-01-01',  
        end='2022-12-31')  
  
    # 결측치 제거  
    data.dropna(inplace=True)  
  
#함수 실행  
Finace_data('005930', 'Samsung')
```

```
Date,Open,High,Low,Close,Volume,Change  
2022-01-03,79400,79800,78200,78600,13502112,0.003831417624521105  
2022-01-04,78800,79200,78300,78700,12427416,0.0012722646310432406  
2022-01-05,78800,79000,76400,77400,25470640,-0.016518424396442133  
2022-01-06,76700,77600,76600,76900,12931954,-0.006459948320413411  
2022-01-07,78100,78400,77400,78300,15163757,0.01820546163849146  
2022-01-10,78100,78100,77100,78000,9947422,-0.003831417624521105  
2022-01-11,78400,79000,78000,78900,13221123,0.011538461538461497  
2022-01-12,79500,79600,78600,78900,11000502,0.0  
2022-01-13,79300,79300,77900,77900,13889401,-0.012674271229404344  
2022-01-14,77700,78100,77100,77300,10096725,-0.007702182284980763  
2022-01-17,77600,77800,76900,77500,8785122,0.002587322121604174  
2022-01-18,77600,77800,76600,77000,9592788,-0.006451612903225823  
2022-01-19,76500,76900,76100,76300,10598290,-0.009090909090909038  
2022-01-20,76200,76700,75900,76500,9708168,0.002621231979030192  
2022-01-21,75800,75800,74700,75600,15774888,-0.0117647058823529  
2022-01-24,75400,75800,74700,75100,13691134,-0.006613756613756627  
2022-01-25,74800,75000,73200,74000,17766704,-0.01464713715046606  
2022-01-26,73900,74400,73100,73300,12976730,-0.009459459459459474  
2022-01-27,73800,74000,71300,71300,22274777,-0.027285129604365577  
2022-01-28,71300,73700,71200,73300,21367447,0.028050490883590573  
2022-02-03,74900,74900,73300,73300,17744721,0.0  
2022-02-04,74300,74600,73400,74000,12730034,0.009549795361528002  
2022-02-07,73500,73600,72400,73000,14240838,-0.013513513513513487  
2022-02-08,73800,74200,73000,73500,11736666,0.006849315068493178
```


a. 종가 기준 모델

```
highest30 = ssdf.nlargest(n=30, columns='close', keep='all')
highest30.sort_values(by=['date'])
highest14_dates = [["2021-01-06", "2021-01-08"], ["2021-01-09", "2021-01-11"], ["2021-01-12", "2021-01-14"], ...]
lowest16_dates = [["2022-06-29", "2022-07-01"], ["2022-07-04", "2022-07-06"], ["2022-09-05", "2022-09-07"], ...]
dates = highest14_dates + lowest16_dates
```

b. 수익률 기준 모델

```
def top_N_changes(ssdf, N):
    ssdf['Abs_Change'] = ssdf['change'].abs()
    top_N = ssdf.sort_values(by='Abs_Change', ascending=False).head(N)
    top_N = top_N.drop('Abs_Change', axis=1)
    return top_N
```

```
top30 = top_N_changes(samsung, 30)
# 결과값 : ['2021-01-08', '2021-01-18', ...]
```

c. 당일 뉴스 기준 모델

```
Holidays = [] # 공휴일 지정
current_date = datetime.datetime.strptime(start_date, "%Y%m%d") #string to datetime
end_date = datetime.datetime.strptime(end_date, "%Y%m%d")
```

```
while current_date <= end_date:
    if current_date.weekday() >= 5 or current_date.strftime("%Y-%m-%d") in holidays:
```

2. 네이버 증권 크롤링

```
def crawling(date, lastpage):
    titles = [];    urls = [];    summaries = []

    for page in range(1, lastpage+1):
        url = URL
        response = requests.get(url)
        soup = BeautifulSoup(response.text, "html.parser")

        for article in soup.find_all("dd", class_="articleSubject"):
            title = article.a.text.strip()
            title = sub("[^ㄱ-ㅎㅌ-ㅣ가-힣 | A-Z |a-z ]","", title)
            url = article.a["href"]
            if title not in t_list:
                titles.append(title)
            if url not in urls:
                urls.append(url)

        for summary in soup.find_all("dd", class_="articleSummary"):
            summary_str = sub("[^ㄱ-ㅎㅌ-ㅣ가-힣 | A-Z |a-z ]","", summary_str)
            summary_str = summary_str.strip()
            summaries.append(summary_str)

    return titles, summaries
```

```
URL =
f"https://finance.naver.com/news/n
ews_search.naver?rcdate=&q=%BB%EF%
BC%BA%C0%FC%C0%DA&x=0&y=0&sm=all.b
asic&pd=4&stDateStart={date[0]}&st
DateEnd={date[1]}&page={page}"
```

```
2021-01-04, "삼성·LG, 부산엑스포 유치 응원 나섰다"
2021-01-04, "삼성전자, 세계 최초 '강아지용 스마트폰' 출시?"
2021-01-04, [유안타증권·주안추천주] 삼성전자·삼성엔지니어링·삼
2021-01-04, "대만 TSMC 팹, 어디까지 가봤니? <3> [강해형의
2021-01-04, ""중공장·영업기밀 어찌나""...美보조금·신형·시작
2021-01-04, "KLA, 용인에 반도체 트레이닝 센터 개소""...한국
2021-01-04, "코스피, 1% 오른 2476.86선·마감...코스닥 6거래일
2021-01-04, [마감·시황] 외국인·기관·쌍끌이·매수에 2470선·회
2021-01-04, "[데이리로 보는 증시] 삼성전자, 기관·외국인·코스
2021-01-04, [코스피·마감] 외국인·기관·동반·순매수...대형주·강세
2021-01-04, [오후·시황] 외국인·기관·쌍끌이·매수에 코스피 248
2021-01-04, "윤준호·삼성전자 부사장, 레인보우로보틱스 합류
2021-01-04, "삼성전자, 비스포크·냉장고·패널·할인행사 진행"
2021-01-04, "[다음주 증시 전망] 1분기 어닝시즌·이목 집중..."
2021-01-04, "비스포크·냉장고·사면·패널·'반값'...삼성전자, 5월
2021-01-04, "코스피, 외국인·기관·사자에 1%대 상승...2480선·회
2021-01-04, [오전·시황] 애플리인상·종료·기대감에 코스피·상
2021-01-04, 코스피·장 초반·강보합...힘빠진·에코프로株에 코스
2021-01-04, "코스피, 개인·외인·사자에 상승 출발...2450선·안
2021-01-04, ""삼성 3분기부터 실적 개선...목표가 8만원""
2021-01-04, "기재개·폐는 반도체株...코스피, 4월·안도렐리·탈
```

3.한영번역과 전처리, 감성분석

```
def TranslatorFunc (news):  
    eng_news = translator.translate(news, dest='en', src='ko')  
    return eng_news
```

```
def clean_text(texts) :  
    texts_re = [sub('[.,?!:;]', '', st) for st in texts]  
    texts_re2 = [sub('[@#$$%^&*()]',' ', st) for st in texts_re]  
    texts_re3 = [sub('[^a-z|A-Z]', '', st) for st in texts_re2]  
    texts_re4 = [sub('\s+', ' ', st) for st in texts_re3]  
    texts_re5 = [sub('[0-9]', '', st) for st in texts_re4]  
    return texts_re5
```

```
afinn = Afinn(language='en')  
sentiment = afinn.score(eng_news)
```

```
news_df = ['les gens pensent aux chiens','i hate flowers','he  
is kind and smart','we are kind to good people']
```

```
scores = [afn.score(article) for article in news_df]  
sentiment = ['positive' if score > 0  
             else 'negative' if score < 0  
             else 'neutral' for score in scores]
```

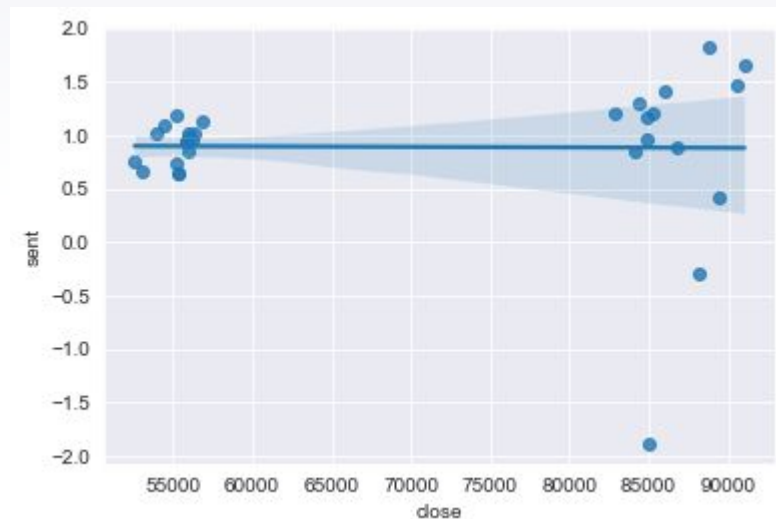
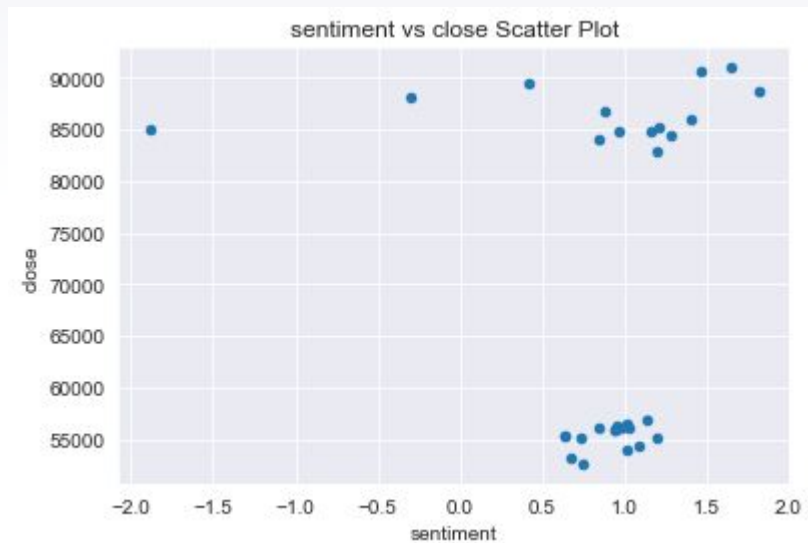
```
Date,Sentence  
2022-01-03,"Samsung Electronics' Municipal Ceremony L  
2022-01-04,"LG Electronics' Samsung All -Red Market R  
2022-01-05,"Samsung Electronics LG Electronics' New R  
2022-01-06,"Han Jong -hee Vice Chairman LG OLED Purch  
2022-01-07,"I came and came again a year ago.Accordin  
2022-01-10,"In reality, Samsung Electronics, which is  
2022-01-11,"Samsung Electronics 'CES, the world's lar  
2022-01-12,"LG Ensol Institution, and the market cap  
2022-01-13,"Weekly Biz Letter MRNA Century Aim of Rig  
2022-01-14,"Video Samsung Phone is a gender gangster  
2022-01-17,"LG Ensol's subscription war began, visite  
2022-01-18,"Sale in Osan Se -kyo, a brand of Korean -  
2022-01-19,"Investigating the iPhone in the iPhone r
```

	topic	scores	sentiments
0	les gens pensent aux chiens	0.0	neutral
1	i hate flowers	-3.0	negative
2	he is kind and smart	3.0	positive
3	we are kind to good people	5.0	positive

3 상관분석 및 시각화



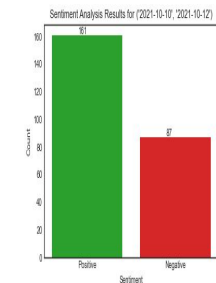
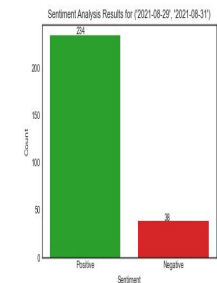
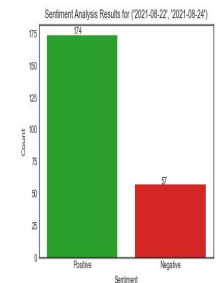
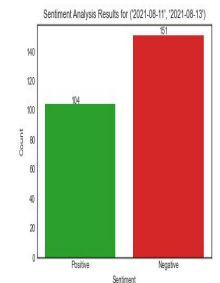
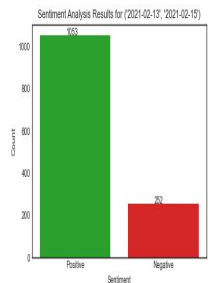
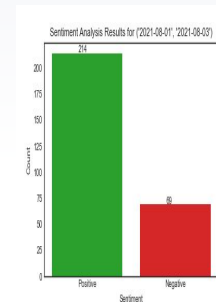
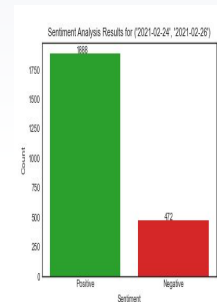
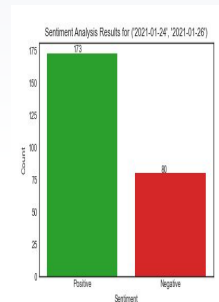
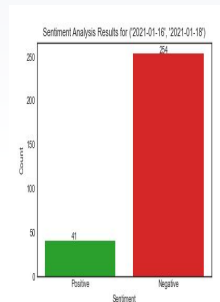
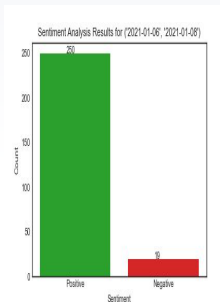
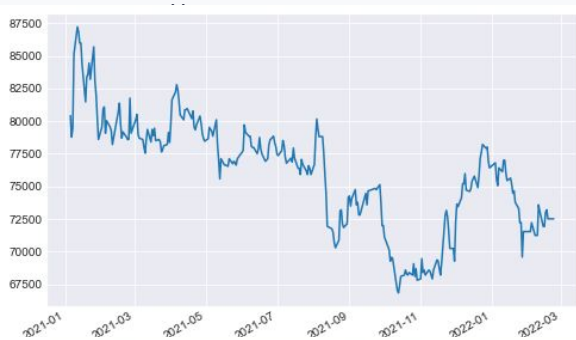
▶ a. 종가 기준 모델



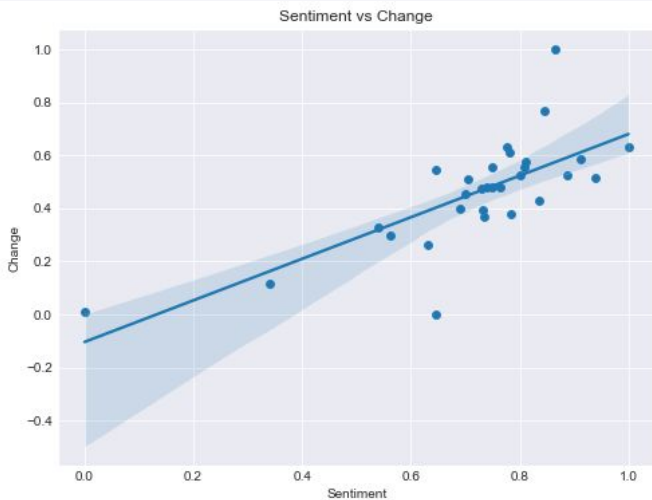
상관계수 : -0.0117, P-value : 0.9512

b. 수익률 기준 모델

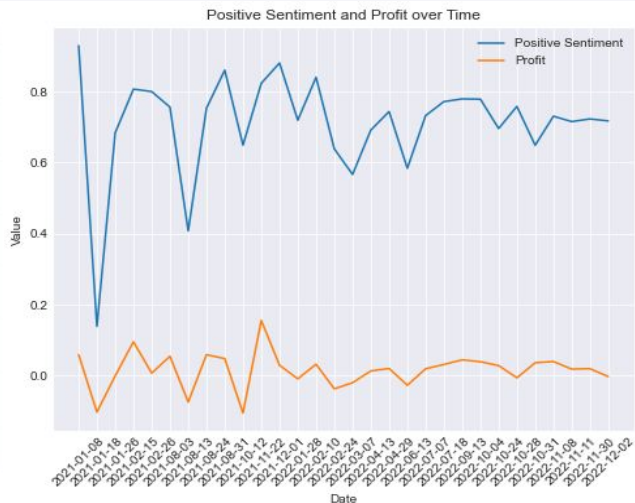
2021 - 01 ~ 2022 - 12 까지 수익률 변동폭이 큰 기간 **감성분석**



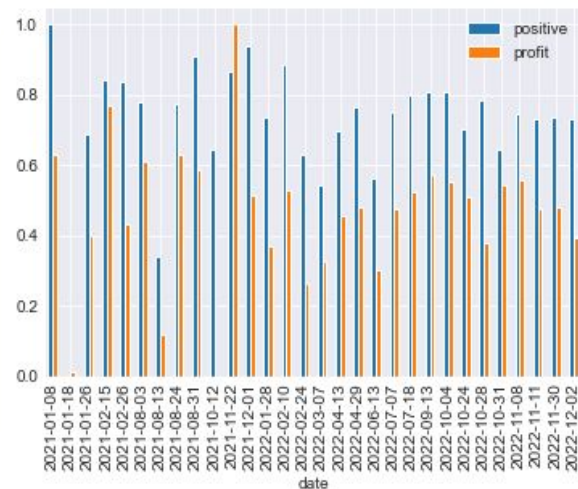
b. 수익률 기준 모델



산점도 그래프



라인 그래프



막대 그래프

상관계수 : 0.73001, P-value : 0.000005

C. 당일 뉴스 감성분석 모델

Affin 사용으로 결과 점수화

```
mean_sentiment = merged_df['Sentiment'].mean()
# 43.891
median_sentiment = merged_df['Sentiment'].median()
# 45.0

# 감성점수가 40점 이상이면 긍정, 미만이면 부정으로 나타냄
df_translated['Pos_Neg'] = df_translated['Afinn'].apply(lambda x:
'Positive' if x >= 40 else 'Negative')

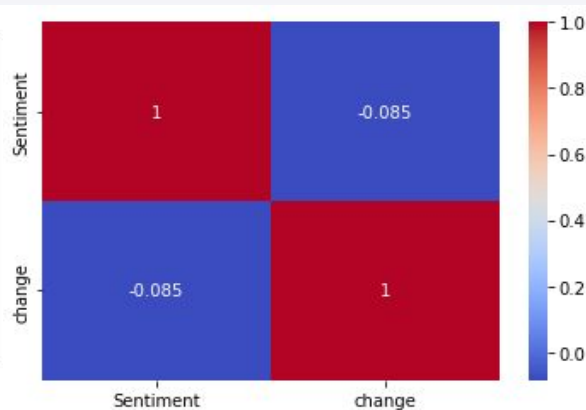
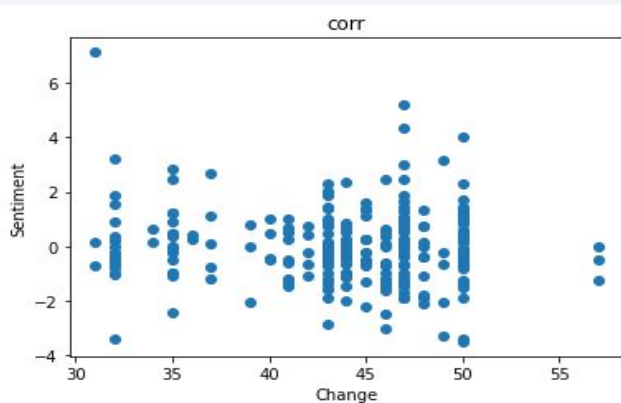
# 실제로 예측 성공한 날 True, 실패한 날 False로 출력
merged_df['score'] = np.where((merged_df['Sentiment'] >= 40) &
(merged_df['change'] >= 0), 'TRUE', 'FALSE')
```

Date,Afinn,Pos_Neg	Sentiment	score
2022-01-03,17.0,Negative	47.0	TRUE
2022-01-04,51.0,Positive	47.0	TRUE
2022-01-05,30.0,Negative	39.0	FALSE
2022-01-06,24.0,Negative	47.0	TRUE
2022-01-07,21.0,Negative	31.0	FALSE
2022-01-10,-6.0,Negative
2022-01-11,15.0,Negative	47.0	TRUE
2022-01-12,60.0,Positive	47.0	FALSE
2022-01-13,109.0,Positive	47.0	TRUE
2022-01-14,17.0,Negative	47.0	FALSE
2022-01-17,34.0,Negative	47.0	FALSE
2022-01-18,41.0,Positive	49.0	FALSE

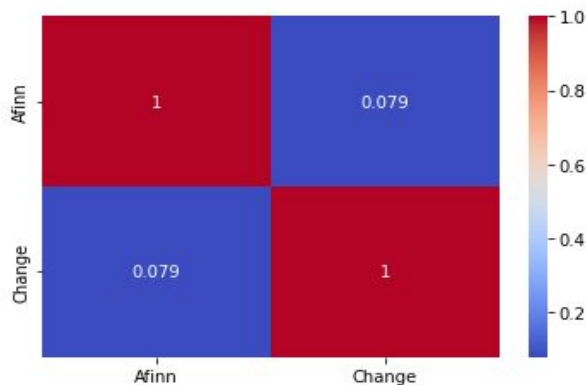
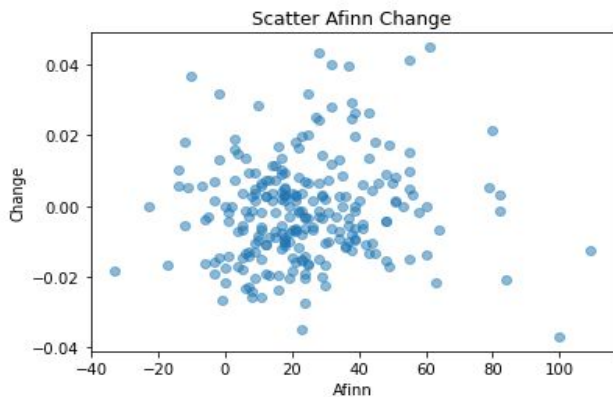
2021년 예측 성공한 날 : 95/248 → 정확도: 약 38%

2022년 예측 성공한 날 : 137/246 → 정확도: 약 56%

C. 당일 뉴스 감성분석 모델



2021년:
상관계수: -0.085,
p-value : 0.403



2022년:
상관계수: 0.079,
p-value : 0.215

결과 정리

a. 종가 높은 30일 직전 3일 뉴스 분석

상관계수 : -0.0117,
P-value : 0.9512

통계량이 낮고, P-값이 0.05 이상
→ 귀무가설 채택
→ 통계적으로 유의하지 않음

b. 수익률 큰 30일 직전 3일 뉴스 분석

상관계수 : 0.73001,
P-value : 0.000005

통계량이 높고, P-값이 0.05 이하
→ 귀무가설 기각
→ 통계적으로 유의

c. 당일 뉴스 감성분석

2021년: - 0.053, 0.403
2022년: 0.079, 0.215

통계량이 낮고, P-값이 0.05 이상
→ 귀무가설 채택
→ 통계적으로 유의하지 않음

공통 가설

H0 : 뉴스가 주가 변동율에 영향을 끼치지 않는다.

H1 : 뉴스가 주가 변동율에 영향을 끼친다.

4

결론 및 의의





한계점



수익률이 낮은 기간
상관관계가 애매



여러 뉴스사에서
비슷한 기사 업로드



- 한영번역 한계
- 같은 단어라도 긍정/부정 가능
- 주식 특화 단어사전 부재



뉴스 감성 외에도
증시에 영향을 주는
여러 변수들 존재

발전방향



수익률과
변동률이 큰
주식을 선택



비슷한 뉴스 기사의
중복을 제거



영어권 주가정보 분석



다른 변수들 같이
고려한 상관관계
분석 가능

감사합니다

금준호, 김란, 노두호, 안지윤, 최규진



2. 네이버 증권 크롤링

```
stock_code = "005930" # 삼성전자 주식 코드
start_date = "시작날짜지정"
end_date = "종료날짜 지정"
Holidays = [] # 공휴일 지정
titles = [] # 제목저장
dates = [] # 날짜저장
previews = [] # 요약 저장
```

```
2021-01-04,"삼성·LG, 부산엑스포·유치·응원 나섰다"
2021-01-04,"삼성전자, 세계 최초 '강아지용 스마트폰' 출시?"
2021-01-04,[유안타겅·주간추천주]삼성전자·삼성엔지니어링·삼성SDI
2021-01-04,"대만·TSMC·팹, 어디까지 가겠니? <3> [강해영의·하이엔드·테크]"
2021-01-04,"""중공장·영업기밀·어쩌나""...美보조금·신청·시작, 머리싸맨·삼성·SK"
2021-01-04,"KLA, 용인에·반도체·트레이닝·센터·개소..."한국·거점·확보·의지""
2021-01-04,"코스피, 1% 오른 2476.86선·마감...코스닥 6거래일만에·하락[시황종합]"
2021-01-04,[마감·시황]·외인·기관·쌍끌이·매수에 2470선·회복
2021-01-04,"[데이터로 보는·중시]삼성전자, 기관·외국인·코스피·순매수 1위(3월·31)"
2021-01-04,[코스피·마감]외인·기관·동반·순매수·대형주·강세에 2470선·안착
2021-01-04,[오후·시황]·외인·기관·쌍끌이·매수에·코스피 2480선·돌파·시도
2021-01-04,"윤준호·삼성전자·부사장, 레인보우로보틱스·합류..."로봇·협력·본격화""
2021-01-04,"삼성전자, 비스포크·냉장고·패널·활인행사·진행"
2021-01-04,"[다음주·중시·전망] 1분기·어닝시즌·이목·집중..."삼성전자가·수급·가름
2021-01-04,"비스포크·냉장고·사면·패널·'반값'·삼성전자, 5월까지·활인·행사"
2021-01-04,"코스피, 외인·기관·사자에 1%대·상승 2480선·회복"
2021-01-04,[오전·시황]·美금리인상·종료·기대감에·코스피·상승
2021-01-04,코스피·장·초반·강보합...힘빠진·에코프로株에·코스닥은·악보합[개장시황]
2021-01-04,"코스피, 개인·외인·사자에·상승·출발 2450선·안착"
2021-01-04,"""삼전·3분기부터·실적·개선·목표가·8만원"""
2021-01-04,"기지개·폐는·반도체株...코스피, 4월·안도ल्ली·탈까"
```

```
current_date = datetime.datetime.strptime(start_date, "%Y%m%d") #string to datetime
end_date = datetime.datetime.strptime(end_date, "%Y%m%d")
```

```
while current_date <= end_date:
    if current_date.weekday() >= 5 or current_date.strftime("%Y-%m-%d") in
    holidays:
        current_date += datetime.timedelta(days=1)
        continue
    for page in tqdm(range(1, 4)):
        url = "네이버 증권 뉴스 주소"
        response = requests.get(url)
        soup = BeautifulSoup(response.text, "html.parser")
        articles = soup.find_all("dd", class_="articleSubject")
        if len(articles) == 0:
            break
        for article in articles:
            title = article.a.text.strip()
            preview_tag = article.find_next_sibling("dt")
            preview = preview_tag.text.strip() if preview_tag else ""
            titles.append(title)
            dates.append(current_date.strftime("%Y-%m-%d"))
            previews.append(preview)
            time.sleep(0.45)
        current_date += datetime.timedelta(days=1)
```


배운 슬라이드



$$E[r_j] = r_f + \beta(E[r_m] - r_f)$$

CAPM MODEL (자본자산 가격 결정 모델)

$$E[r_j] = r_f + \beta_1 F_1 + \dots + \beta_n F_n$$

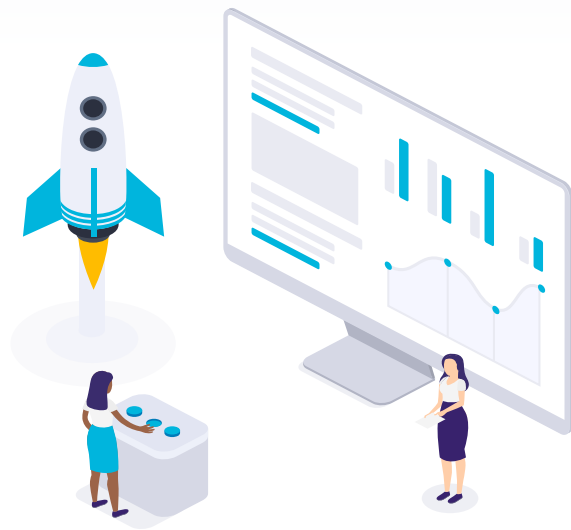
APT MODEL (차익 거래가 모델)

... 등등

FAMA FRENCH 3 MODEL (파마프렌치 3요인모델)



▶ 주가와 증권뉴스 감성분석 상관관계 있다 (?)



한계점

수집 데이터의 한계성

데이터의 중복성

데이터의 변동성

1. 한계점 : 긍정, 부정 기준을 잡기가 애매함
2. 아직 활성화된 한국어 감성사전이 없어서 번역해서 감성분석을 해야하는게 부정확함
3. 같은 단어가 긍정일수도 부정일수도 있음
4. 주식 시장의 단어사전이 아닌 일반 단어사전을 사용해서 정확도가 낮음
5. 매일매일을 분석했기 때문에 변화율의 변동폭이 적은날에는 점수가 애매함

2. 수익률 변동 기준 모델

수익률 높은
30일 추출

크롤링,
전처리

한영번역

감성분석

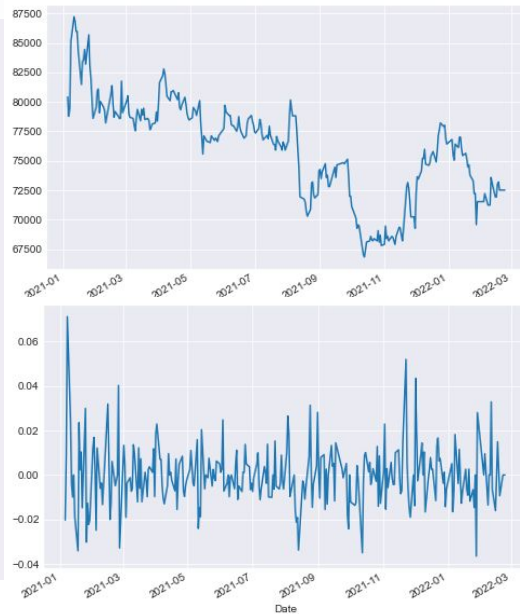
상관분석

데이터 시각화

```
def top_N_changes(finance_data, N):  
    finance_data['Abs_Change'] = finance_data['simple_return'].abs()  
    top_N = finance_data.sort_values(by='Abs_Change', ascending=False).head(N)  
    top_N = top_N.drop('Abs_Change', axis=1)  
    return top_N
```

```
top30 = top_N_changes(samsung, 30)
```

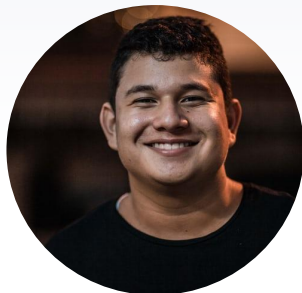
```
# 결과값 : ['2021-01-08', '2021-01-18', '2021-01-26', '2021-02-15', '2021-02-26',  
            '2021-08-03', '2021-08-13', '2021-08-24', '2021-08-31', '2021-10-12',  
            '2021-11-22', '2021-12-01', '2022-01-28', '2022-02-10', '2022-02-24',  
            '2022-03-07', '2022-04-13', '2022-04-29', '2022-06-13', '2022-07-07',  
            '2022-07-18', '2022-09-13', '2022-10-04', '2022-10-24', '2022-10-28',  
            '2022-10-31', '2022-11-08', '2022-11-11', '2022-11-30', '2022-12-02']
```



▶ 팀원 구성



김란



김준호



노두호

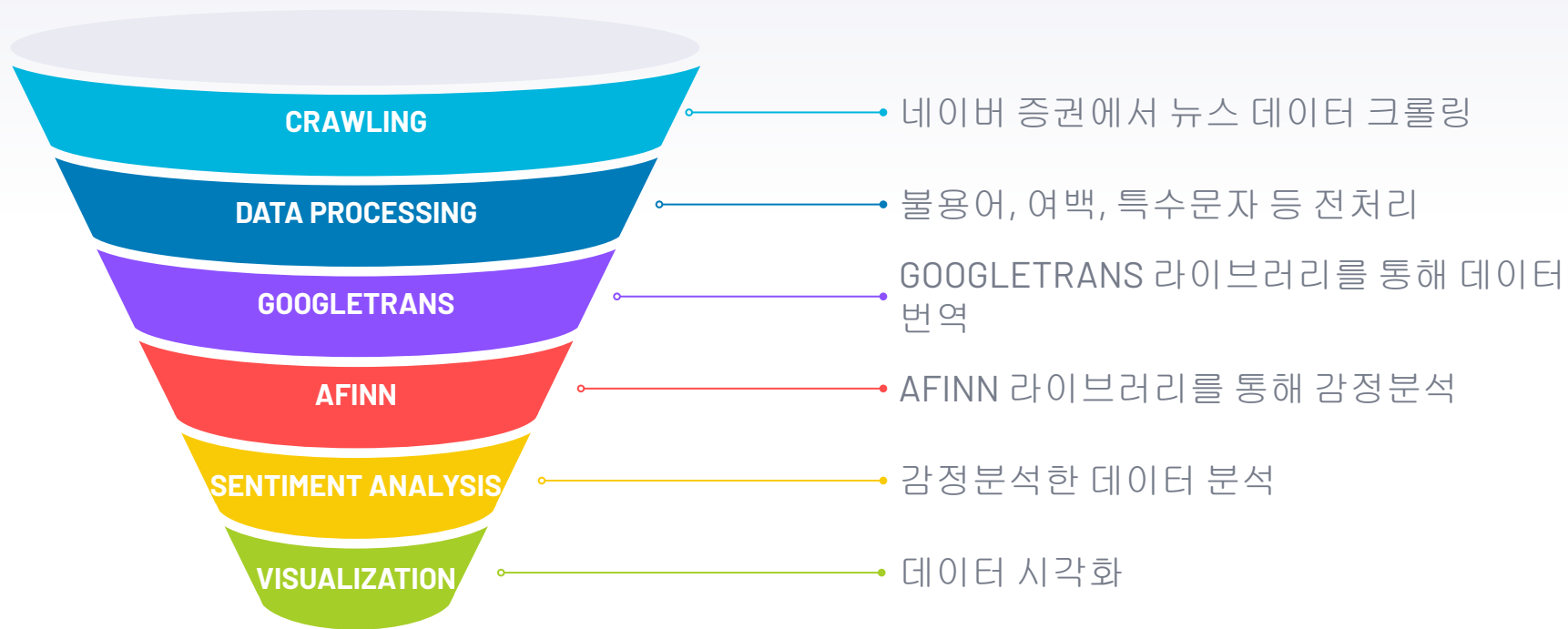


안지윤



최규진

DATA 전처리 과정



▶ a. 종가 기준 모델

종가 높은 30일 추출

```
highest30 = ssdf.nlargest(n=30, columns='close', keep='all')
highest30.sort_values(by=['date'])

highest14_dates = [ ["2021-01-06", "2021-01-08"], ["2021-01-09", "2021-01-11"], ["2021-01-12", "2021-01-14"], ...]

lowest30 = ssdf.nsmallest(n=30, columns='close', keep='all')
lowest30.sort_values(by=['date'])

lowest16_dates = [ ["2022-06-29", "2022-07-01"], ["2022-07-04", "2022-07-06"], ["2022-09-05", "2022-09-07"], ...]

dates = highest14_dates + lowest16_dates
```

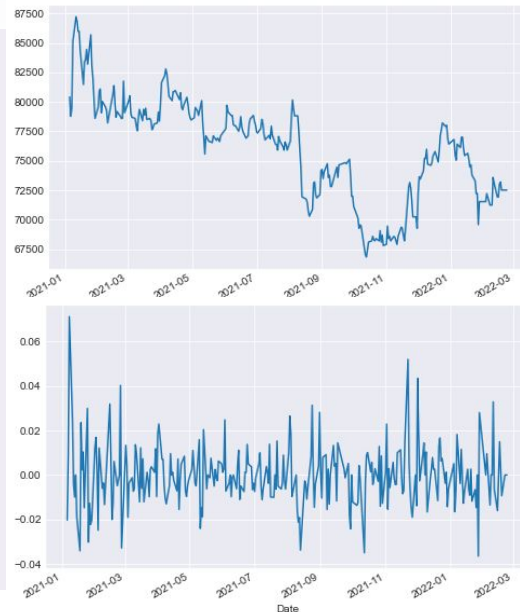

b. 수익률 변동 기준 모델

수익률 변동이 큰 30일 추출

```
def top_N_changes(finance_data, N):  
    finance_data['Abs_Change'] = finance_data['simple_return'].abs()  
    top_N = finance_data.sort_values(by='Abs_Change', ascending=False).head(N)  
    top_N = top_N.drop('Abs_Change', axis=1)  
    return top_N
```

```
top30 = top_N_changes(samsung, 30)
```

```
# 결과값 : ['2021-01-08', '2021-01-18', '2021-01-26', '2021-02-15', '2021-02-26',  
            '2021-08-03', '2021-08-13', '2021-08-24', '2021-08-31', '2021-10-12',  
            '2021-11-22', '2021-12-01', '2022-01-28', '2022-02-10', '2022-02-24',  
            '2022-03-07', '2022-04-13', '2022-04-29', '2022-06-13', '2022-07-07',  
            '2022-07-18', '2022-09-13', '2022-10-04', '2022-10-24', '2022-10-28',  
            '2022-10-31', '2022-11-08', '2022-11-11', '2022-11-30', '2022-12-02']
```



수익률 변동 기준 모델

수익률(RETURN) 변동이 큰 기간
뉴스데이터 분석



뉴스의 긍정 부정단어가 주가에 미치는 정확도 계산

금준호, 최규진

당일 뉴스 감성 분석 모델

금준호, 최규진



전처리 과정

주식 데이터
라이브러리로 삼성전자의
데이터를 불러옴

1

생성된 날짜와 url을
기반으로 크롤러 함수를
사용하여 네이버
증권에서 데이터 수집

3

번역된 영문 텍스트를
Afinn 라이브러리로 감정
분석

5

불러온 데이터의
기간으로 날짜와 url 생성

2

수집된 데이터를 구글
번역 라이브러리로
영문으로 번역

4

감정 분석시 긍정일때
주가가 오르고, 부정일때
주가가 내려간 수를
종합하여 정확도를 계산

6

샘플



► You can also split your content

White

Is the color of milk and fresh snow, the color produced by the combination of all the colors of the visible spectrum.

Black

Is the color of ebony and of outer space. It has been the symbolic color of elegance, solemnity and authority.



Use diagrams to explain your ideas

Vestibulum congue tempus

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor.



Vestibulum congue tempus

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor.



Vestibulum congue tempus

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor. Ipsum dolor sit amet elit, sed do eiusmod tempor.

► And tables to compare data

	A	B	C
Yellow	10	20	7
Blue	30	15	10
Orange	5	24	16



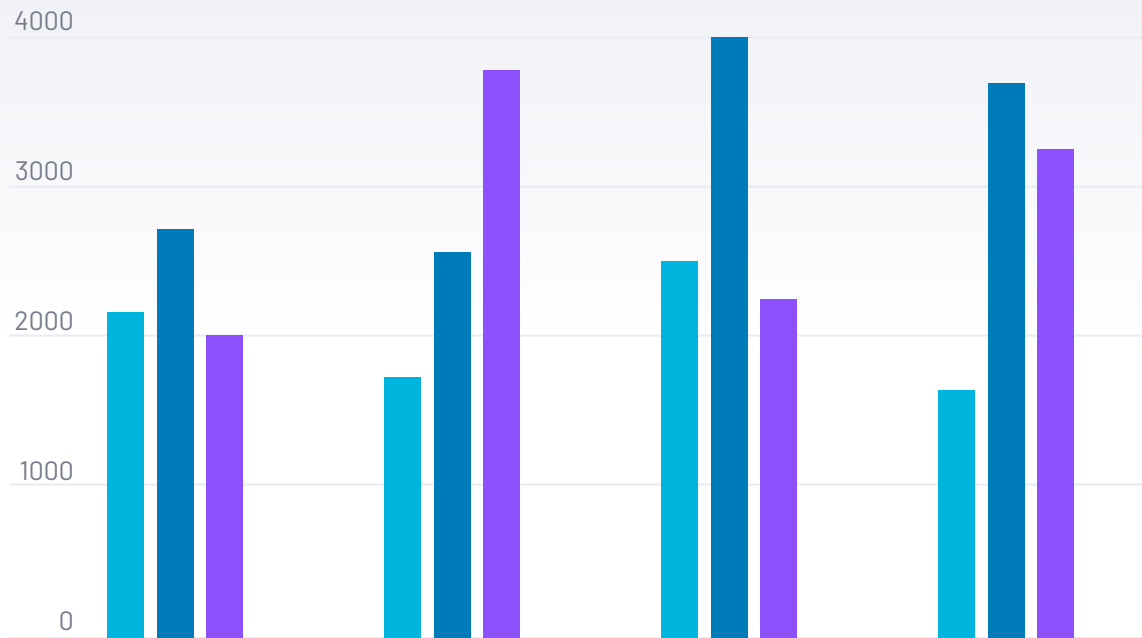
► This is a slide title

- ▶ Here you have a list of items
- ▶ And some text
- ▶ But remember not to overload your slides with content

Your audience will listen to you or read the content, but won't do both.

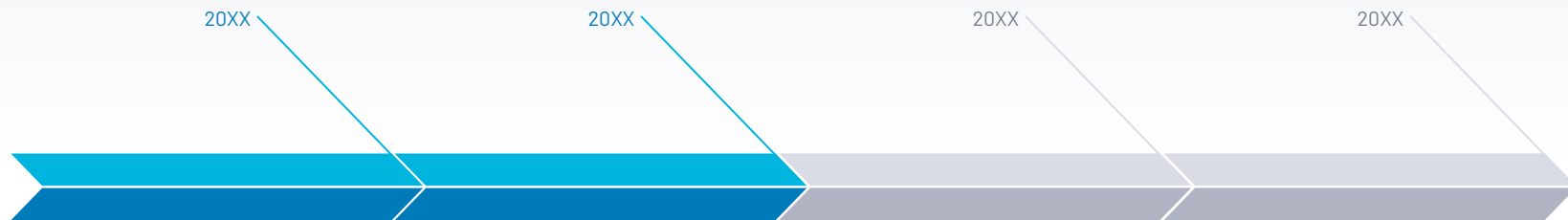






You can insert graphs from Excel or Google Sheets

Our process is easy



Lorem Ipsum

Lorem ipsum dolor sit amet, consectetur adipiscing.

Lorem Ipsum

Lorem ipsum dolor sit amet, consectetur adipiscing.

Lorem Ipsum

Lorem ipsum dolor sit amet, consectetur adipiscing.

Lorem Ipsum

Lorem ipsum dolor sit amet, consectetur adipiscing.

▶ Our process is easy

기사크롤링

전처리

데이터 번역

감정분석

상관분석

데이터 시각화

▶ Let's review some concepts

Yellow

Is the color of gold, butter and ripe lemons. In the spectrum of visible light, yellow is found between green and orange.

Yellow

Is the color of gold, butter and ripe lemons. In the spectrum of visible light, yellow is found between green and orange.

Blue

Is the colour of the clear sky and the deep sea. It is located between violet and green on the optical spectrum.

Blue

Is the colour of the clear sky and the deep sea. It is located between violet and green on the optical spectrum.

Red

Is the color of blood, and because of this it has historically been associated with sacrifice, danger and courage.

Red

Is the color of blood, and because of this it has historically been associated with sacrifice, danger and courage.

Credits

Special thanks to all the people who made and released these awesome resources for free:

- ▶ Presentation template by [SlidesCarnival](#)
- ▶ Illustrations by [Sergei Tikhonov](#)
- ▶ Photographs by [Unsplash](#)

Presentation design

This presentation uses the following typographies:

- ▶ Titles: Raleway
- ▶ Body copy: Barlow

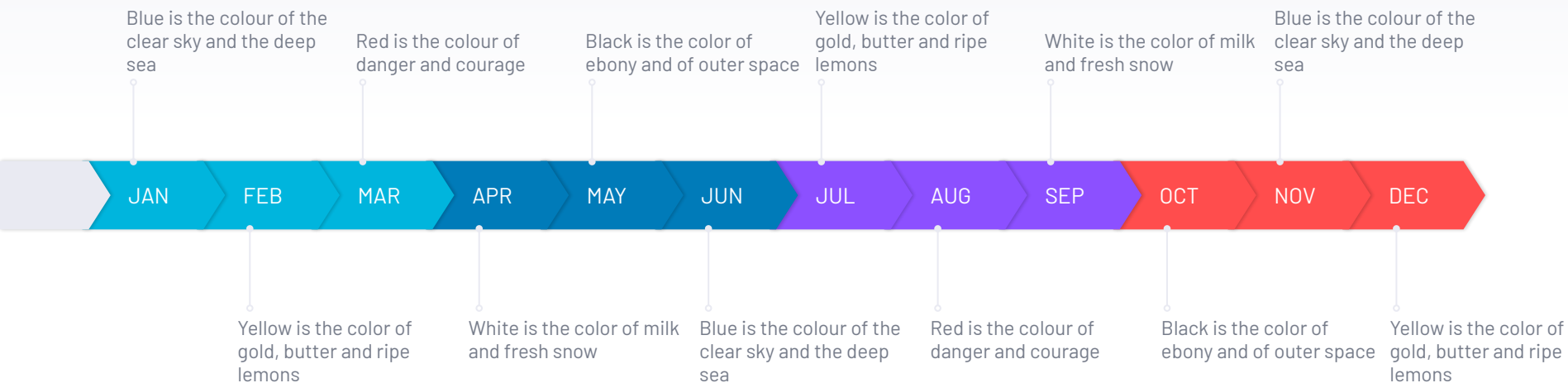
Download for free at:

<https://www.fontsquirrel.com/fonts/raleway>

<https://www.fontsquirrel.com/fonts/barlow>

You don't need to keep this slide in your presentation. It's only here to serve you as a design guide if you need to create new slides or download the fonts to edit the presentation in PowerPoint®

Timeline



주식 데이터 라이브러리로 삼성전자 주식 데이터를 불러오기

Special thanks to all the people who made and released these awesome resources for free:

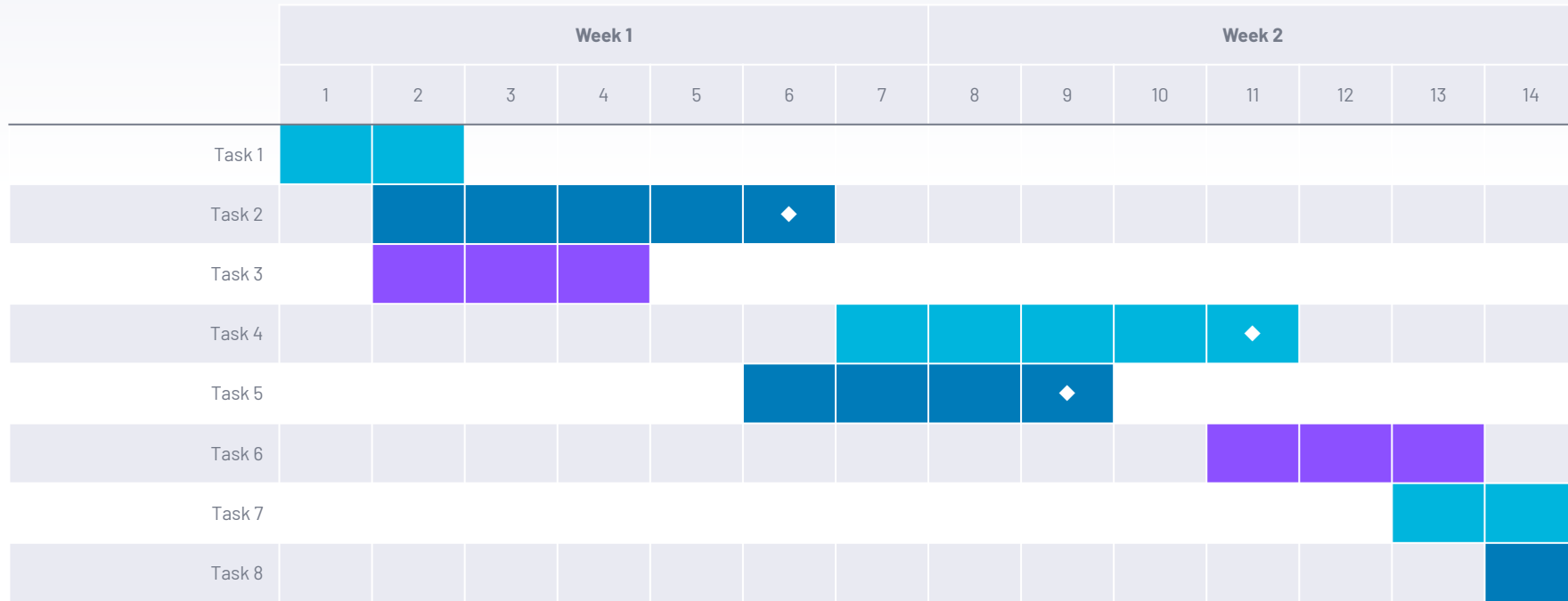
- ▶ Presentation template by [SlidesCarnival](#)
- ▶ Illustrations by [Sergei Tikhonov](#)
- ▶ Photographs by [Unsplash](#)

Credits

Special thanks to all the people who made and released these awesome resources for free:

- ▶ Presentation template by [SlidesCarnival](#)
- ▶ Illustrations by [Sergei Tikhonov](#)
- ▶ Photographs by [Unsplash](#)

Gantt chart



SWOT Analysis

STRENGTHS

Blue is the colour of the clear sky and the deep sea

S

WEAKNESSES

Yellow is the color of gold, butter and ripe lemons

W

Black is the color of ebony and of outer space

OPPORTUNITIES

O

White is the color of milk and fresh snow

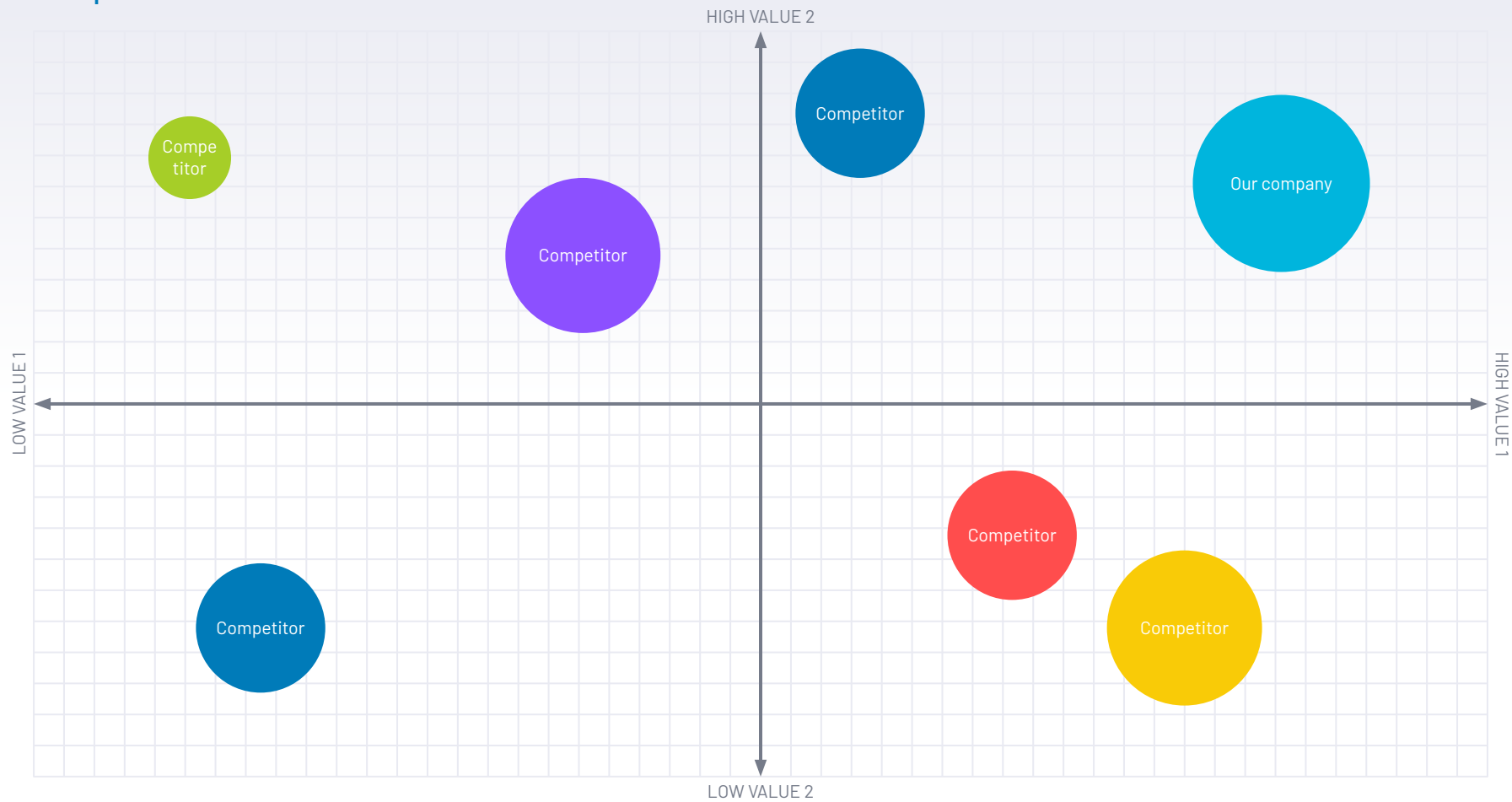
THREATS

T

Funnel



Competitor Matrix



Weekly Planner

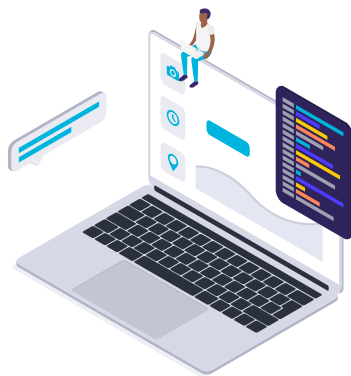
	SUNDAY	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY
09:00 - 09:45	Task	Task	Task	Task	Task	Task	Task
10:00 - 10:45	Task	Task	Task	Task	Task	Task	Task
11:00 - 11:45	Task	Task	Task	Task	Task	Task	Task
12:00 - 13:15	✓ Free time	✓ Free time	✓ Free time	✓ Free time	✓ Free time	✓ Free time	✓ Free time
13:30 - 14:15	Task	Task	Task	Task	Task	Task	Task
14:30 - 15:15	Task	Task	Task	Task	Task	Task	Task
15:30 - 16:15	Task	Task	Task	Task	Task	Task	Task

Extra resources

Illustrations created by [Sergei Tikhonov](#).

Free illustrations published under the MIT License. You can use them for personal and commercial projects, without the need to include attribution.

[See license](#).



Extra resources

Illustrations created by [Sergei Tikhonov](#).

Free illustrations published under the MIT License. You can use them for personal and commercial projects, without the need to include attribution.

[See license](#).





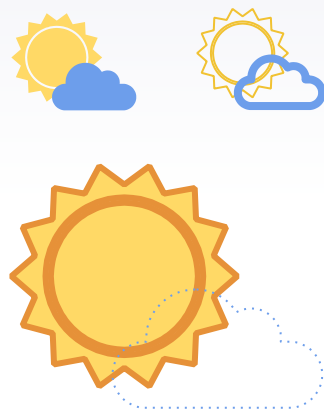
SlidesCarnival icons are editable shapes.

This means that you can:

- ▶ Resize them without losing quality.
- ▶ Change fill color and opacity.
- ▶ Change line color, width and style.

Isn't that nice? :)

Examples:



Find more icons at
slidescarnival.com/extra-free-resources-icons-and-maps

Diagrams and infographics



You can also use any emoji as an icon!
And of course it resizes without losing quality.

How? Follow Google instructions <https://twitter.com/googledocs/status/730087240156643328>



many more...



Free templates for all your presentation needs



For PowerPoint and
Google Slides



100% free for personal
or commercial use



Ready to use,
professional and
customizable



Blow your audience
away with attractive
visuals