

분석 보고서

1. 서론

- a. 가설 : 유튜브 데이터와 주식 관련 데이터는 상관 관계를 가질 것이다
- b. 검증 : 유튜브 데이터와 주식 관련 데이터와의 시각화 및 분석

2. 분석 과정

a. 분석 시의 유튜브 데이터 및 주식 관련 데이터 설정

i. 팀원들과의 토의 결과 최근 국제정세에 가장 큰 영향을 미친 사건들과 관련 있는 주식을 선정

1. 그 결과 1. 코로나 발병 , 2. 러시아-우크라이나 전쟁을 국제 정세에 큰 영향을 미친 사건으로 규정
2. 코로나의 영향과 관련된 주식은 엔터주, 제약주, 항공주를 선정. 러시아-우크라이나 전쟁과 관련된 주식은 석유/가스주를 선정

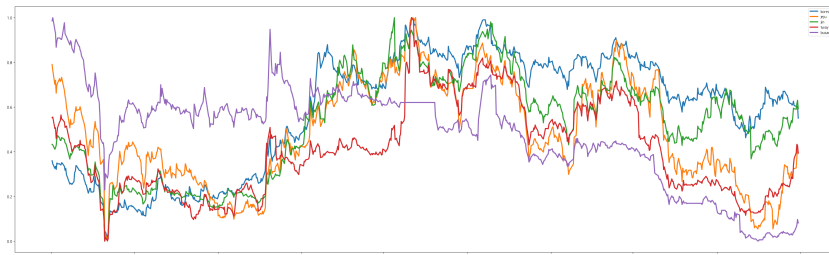
a. 코로나 관련 분석 기준 선정

i. 분석 및 시각화를 통해 각각 주들이 코로나의 영향을 받았는지 확인

▼ 엔터주

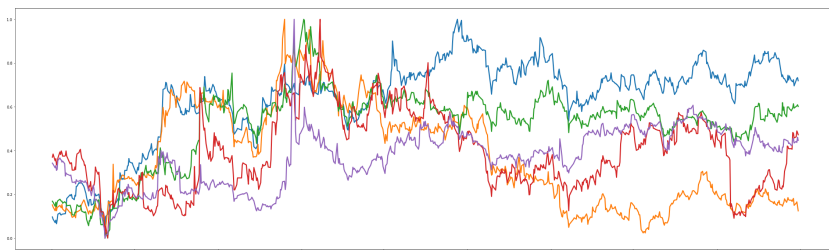
- 코스피에 상장된 기업들도 몇 없을 뿐더러, 코로나 발병 이후 상장한 주식들도 다수이므로 분석 및 시각화 진행 X

▼ 항공주



- 기간별 증가가 비슷한 흐름을 보이는 3개의 기업과 특이한 흐름을 보이는 1개의 기업 선정
- 대한항공, 제주항공, 진에어 (유사한 흐름), 에어부산 (특이한 흐름)

▼ 제약주



- 기간별 증가가 비슷한 흐름을 보이는 3개의 기업과 특이한 흐름을 보이는 1개의 기업 선정
- 셀트리온, 한미약품, 대웅제약 (유사한 흐름), 삼성 바이오 (특이한 흐름)

ii. 항공주, 제약주에서 분석할 카테고리 선정

▼ 카테고리 선정 과정

: **코로나 직후/백신/위드코로나를 기준으로 영향을 더 많이 받은 카테고리 선정**

코스피 지수 선 색상

- red: 코스피 지수(종가기준)
- blue: 각 기업별 코스피 지수(종가기준)

수직선 선 색상

- red: 코로나 시작 (2020, 3, 20)
- orange: 해외 백신 관련 기사 다량 배포 (2021, 2, 1)
- green: 국내 화이자 첫 접종 (2021, 2, 27)
- blue: 코스피 최대 (2021, 6, 25)
- black: 위드코로나 시행 (2021, 11, 1)

항공주 해석

- 가정: 코로나 발생 이후 항공주는 코로나 영향을 부정적으로 받은 주일 것이다.
- 결과:
 - 주가가 최저점을 찍었던 기간(2020.03.19)부터 위드 코로나가 시행되기 전(2021.11.1)까지는 코스피보다 대부분 낮은 주가 흐름을 보여줌.
 - 하지만 위드 코로나 시행(2021.11.1) 이후 잠시 동안의 하락을 보여준 뒤, 대체로 코스피보다 높은 주가 흐름을 보여줌.
 - 항공주의 대부분 비율을 차지하는 대한항공을 보았을 때, 위드 코로나 시행 전후(2021.11.1) 코스피와의 주가 비교 시, 주가 흐름의 명백한 차이가 보임.

=> 코로나로 인해 부정적인 영향을 받았음을 확인함과 동시에 위드 코로나와 함께 주가가 상대적으로 상승한 것을 확인할 수 있음.



제약주 해석

- 가정: 코로나 발생 이후 제약주는 코로나 영향을 받은 최대 수혜주일 것이다.
- 결과:
 - 해외 백신 관련기사 다량배포일(2021. 2. 1) 전까지는 대부분의 제약주들은 코로나로 인한 수혜받음 (셀트리온이 가장 수혜많이 받음)
 - 해외 백신 관련 기사 다량 배포+ 국내에 해외백신 접종 이후, 코스피 지수보다 크게 제약주들의 주가가 폭락함
 - 위드 코로나 발표 이후, 코스피 지수와 격차가 크게 줄어들지 않음

=> 코로나 수혜주일 것이라는 가정과 아예 다른 결과를 나타냄



결론 - 항공주가 가정과 유사한 방향성을 보이므로, 항공주로 채택

- iii. 유튜브 데이터 검색 시 키워드 및 필터 설정
 - a. 항공 + 코로나 - 스포츠 (코로나와 관련된 항공업계 데이터를 수집하기 위하여, 스포츠 관련 영상들은 제외)
 - b. 항공사 (항공사 관련 주가 데이터를 활용하기 때문에 항공사 관련 영상 검색)
 - c. 필터 : 조회수, 관련성 (타 필터는 사람들의 관심도를 의미한다고 보기 어려움)
- b. 러시아-우크라이나 관련 주식 선정
 - i. 석유/가스 관련 분석 기준 선정
 - ii. 코스피에 상장된 유가 관련된 회사들 중, 시가총액이 상위 5개인 기업 선정
 - i. SK이노베이션 - 152,013 (096770)
 - ii. S-Oil - 99,411 (010950)
 - iii. HD현대 - 49,924 (267250)
 - iv. GS - 41,998 (078930)
 - v. LX인터내셔널 - 13,585 (001120)

ii. 유튜브 데이터 검색 시 키워드 및 필터 설정

- 국제 유가 (전쟁으로 인한 유가 변동의 영향을 제일 잘 보여주는 키워드)
- 유가 전쟁 (러시아-사우디아라비아 사이의 유가 전쟁이 있기도 하였으며, 우크라이나와의 전쟁 및 석유 관련해서도 나타내는 키워드)
- 필터 : 조회수, 관련성 (타 필터는 사람들의 관심도를 의미한다고 보기 어려움)

b. 정해진 카테고리의 주식 관련 데이터들과 유튜브 데이터를 분석하기

▼ 항공주

1. 사용할 주식 관련 데이터를 증가와 변화율, 거래량으로 설정

- 각 회사들의 변화율, 증가, 거래량을 한 달 단위로 resampling

	kor_change_4w	kor_close_4w	kor_volume_4w	jeju_change_4w	jeju_close_4w	jeju_vol
Date						
2020-01-06	-0.023956	17647.333333	5.219850e+05	-0.018120	20126.000000	220631.6
2020-02-03	-0.004854	16873.222222	5.601328e+05	-0.007252	18545.666667	148783.0
2020-03-02	-0.003631	15952.800000	1.049362e+06	-0.004818	16700.700000	214289.0

2. Youtube Data API V3를 통해 불러올 수 있는 데이터들 중

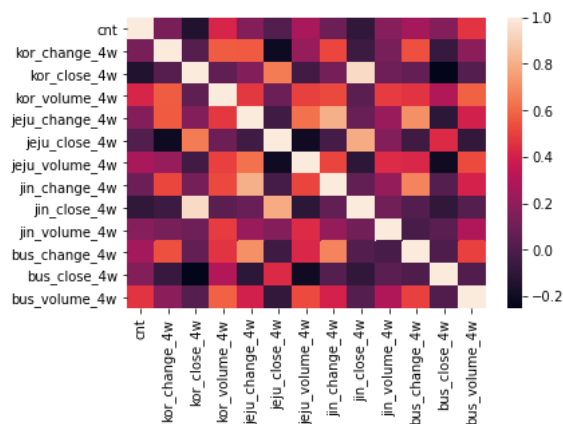
필터는 조회수로 하고,
업로드 날짜 별 영상 개수를 가져오기

Date	
2020-02-03	7
2020-03-02	30
2020-03-30	27
2020-04-27	24

3. Data별로 주식 관련 데이터와 영상 개수를 모두 합치고 MinMaxScaler를 통해 정규화

	cnt	kor_change_4w	kor_close_4w	kor_volume_4w	jeju_change_4w	jeju_close_4w
Date						
2020-01-06	0.423077	0.000000	0.255382	0.000000	0.000000	0.806034
2020-02-03	0.115385	0.469574	0.217910	0.003509	0.282688	0.668437
2020-03-02	1.000000	0.499626	0.173356	0.048507	0.346002	0.507798

4. 정규화 된 데이터로 히트맵을 작성



▼ 석유주

1. 사용할 주식 관련 데이터를 증가와 변화율, 거래량으로 설정

- 각 회사들의 변화율, 증가, 거래량을 한 달 단위로 resampling

	sk_change_4w	sk_close_4w	sk_volume_4w	soil_change_4w	soil_close_4w	soil_vol
Date						
2020-01-06	-0.004296	148000.000000	2.698130e+05	-0.006445	92933.333333	3.142103
2020-02-03	-0.009726	133388.888889	5.297488e+05	-0.013078	82722.222222	3.821017
2020-03-02	-0.003710	130300.000000	6.644577e+05	-0.004663	74475.000000	3.693485

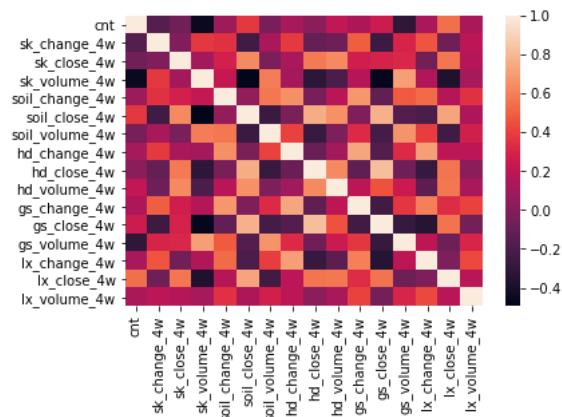
2. Youtube Data API V3를 통해 불러올 수 있는 데이터들 중
필터는 관련성으로 하고,
업로드 날짜 별 영상 개수를 가져오기

Date	
2020-01-06	3
2020-02-03	4
2020-03-02	4
2020-03-30	10
2020-04-27	44

3. Data별로 주식 관련 데이터와 영상 개수를 모두 합치고 MinMaxScaler를 통해 정규화

	cnt	sk_change_4w	sk_close_4w	sk_volume_4w	soil_change_4w	soil_close_4w
Date						
2020-01-06	0.037975	0.235913	0.296005	0.018556	0.238906	0.642606
2020-02-03	0.050633	0.062106	0.223971	0.149536	0.000000	0.475331
2020-03-02	0.050633	0.254677	0.208742	0.217415	0.303108	0.340227

4. 정규화 된 데이터로 히트맵을 작성



3. 결론

a. 항공주

- i. 업로드 날짜에 따른 유튜브 영상 개수와 주식 데이터와의 상관분석 시, 각 회사의 거래량과 영상 개수 간의 상관 계수가 다른 데이터보다 양의 상관 관계가 큰 것을 알 수 있다.
- ii. 그와 반대로, 유튜브 영상 개수와 종가,변화율은 상관 관계가 없음을 알 수 있었다.
- iii. 이는 초반에 설정했던 '유튜브 데이터와 주식 관련 데이터는 상관 관계를 가질 것이다' 라는 가정에 완전히 부합하진 않지만 특정 데이터의 경우에는 가정에 부합함을 알 수 있으며, 유튜브 내, 사람들의 관심이 높아지면 거래량에는 영향을 미치지만 주식 가격에는 변동을 주지 못한다는 점을 시사할 수 있다.

b. 유가주

- i. 업로드 날짜에 따른 유튜브 영상 개수와 주식 데이터와의 상관분석 시, SK이노베이션과 GS는 거래량이 음의 상관 관계를, S-OIL과 LX인터내셔널은 종가와 양의 상관 관계를 가졌음을 알 수 있었다.
- ii. 그와 반대로, HD현대는 영상 개수와 상관 관계를 가지는 주가 데이터가 없었으며, 변화율과는 어떤 회사와도 상관 관계가 없었다.

iii. 이는 같은 석유 관련 회사임에도, 회사들마다 유튜브 데이터에 따른 상관관계가 다 다를 수 있었고, 특정 회사는 유가 관련 영상이 많아질수록 종가 또한 상승함을 알 수 있었다.

따라서, 특정 회사를 매수 혹은 매도할 시에 유튜브 영상 데이터를 활용할 수 있음을 알게 되었다.

4. 한계점

- a. 관련성 순위나 조회수 순으로 유튜브 영상을 불러올 때, 상위 동영상의 경우 가중치를 주어야한다고 생각했지만, 그것이 잘 되지 않음.
- b. 만약 유튜브 데이터가 주가 데이터에 반영되는 데에 시간이 소요된다면, 해당 분석에서 진행한 상관관계 분석은 의미가 없음을 알 수 있음.
- c. 주가에는 단순 사람들의 관심도 뿐만이 아닌, 해당 회사의 실적이나 미래에 대한 기댓값도 포함되어 있으므로, 단순 유튜브 데이터만을 통한 주가 분석은 한계가 있었음.