금융팀

KUBIG CONFERENCE PROJECT 중간보고

박제재 조성윤 우다경 김수경

“주제: 해외주식 데이터를 이용한 국내/해외 종목 관계 분석”

제공된 데이터셋

- NASDAQ_FC_STK_IEM_IFO.CSV

  - 2023년 1월 1일 부터 2023년 8월 31일까지 미국 나스닥 거래소에서 시세를 제공하는 주문 가능한 종목 정보
  - ISIN_IEM_CD : ISIN 코드
  - TCK_IEM_CD : 종목 티커 코드
  - FC_SEC_KRL_NM : 해외주식 종목 한글명
  - FC_SEC_ENG_NM : 해외주식 종목 영문명

2. NASDAQ_DT_FC_STK_QUT.CSV

- 2023년 1월 1일부터 2023년 8월 31일까지 NASDAQ_FC_STK_IEM_IFO.csv에 있는 종목의 시세 정보
- TRD_DT : 거래일자
- TCK_IEM_CD : 티커종목코드
- IEM_ONG_PR : 종목시가
- IEM_HI_PR : 종목고가
- IEM_LOW_PI : 종목저가
- IEM_END_PR : 종목종가
- ACL_TRD_ATY : 누적거래수량
- SLL_CNS_SUM_QTY : 매도체결합계수량
- BYN_CNS_SUM_QTY : 매수체결합계수량

데이터 전처리

산업(SECTOR)를 어떻게 처리할 것인가?

- 국내 주식 데이터 크롤링: KRX 사이트

- 국내/해외 주식의 상이한 업종 문제

-> 분류를 통일! 어떤 기준?

-> 다양한 산업분류 기준 탐색

-> “GICS(글로벌산업분류기준)” 근거하여 11개로 통일

- 데이터 분류 기준을 얼마나 세분화시킬 것인가?
  - 대/중/소 분류 체계

-> 너무 세분화하면 데이터량의 한계에 따라 상관관계 확인이 어려울 것으로 판단

글로벌산업분류기준(GICS)

경제 섹터	산업군
에너지	에너지
소재	소재
산업재	자본재
	산업 전문 서비스
	운송
자유 소비재	자동차 및 부품
	내구소비재 및 의류
	소비자 서비스
	소매
필수 소비재	음식료 소매
	음식료, 담배
	가정 및 개인용품
건강 관리	건강관리 서비스 및 장비
	제약 및 생명과학
금융	은행
	다각화된 금융
	보험
정보기술	소프트웨어 및 IT 장비
	하드웨어 및 IT 장비
	반도체 및 반도체 장비
커뮤니케이션 서비스	통신 서비스
	미디어 및 엔터테인먼트
유틸리티	유틸리티
부동산	부동산

1. 이동평균

이동평균은 일정 기간의 주식 평균 가격을 의미하며 주식 가격의 전반적인

가까운 미래의 주식 가격은 최근 주식 가격들의 영향을 미치므로, 우리는 단
생변수로 지정.

5일 / 10일 / 20일 간격으로 주식 평균 가격을 계산하여 가격 동향을 파악할
성!

2.obv

obv 지표는 거래량을 분석하는 지표로서 매수세가 많을수록 주가가 상승할
하락할 확률이 높아지는 원리를 적용한 지표.

3. Stochastic

Stochastic(스토캐스틱)은 최근 N일간의 최고가와 최저가의 범위 내에서 한
도세보다 강할 때는 그 위치가 높게 형성되고, 매도세가 매수세보다 강할 때
한 것이다.

4. RSI

RSI 지표는 과매수, 과매도를 나타내는 지표.

과매수란 주식 가격이 기존보다 일정 수준 이상 상승할 경우, 많은 매수가 발
로 가격이 하락할 때, 많은 매도가 발생하는 것을 의미.

RSI의 공식을 구하기 위해서는 위와 같은 변수가 필요.

- RSI_U = 전날 주가 대비 오늘 주가 상승폭
- RSI_D = 전날 주가 대비 오늘 주가 하락폭
- RSI_AU = 일정기간(14일) 동안의 RSI_U의 평균값
- RSI_AD = 일정기간(14일) 동안의 RSI_D의 평균값
- RSI = RSI_AU / (RSI_AU + RSI_AD) * 100

외부데이터 결합

추가적으로 필요한 정보는 없을까?

- 거시경제 지표 (Economic Indicators)
  - 환율, 금리, 비트코인 달러 가격 등
  - 미국 지수: DOW JONES, S&P500
  - 한국 지수: KOSDAQ, KOSPI 200

- 기술지표 (Technical Indicatiors)

: 주식의 성과(수익률)을 기반으로 상관관계 고려하는 것이 중요하다 판단

-> 다양한 성과지표 코드로 구현하여 반영!

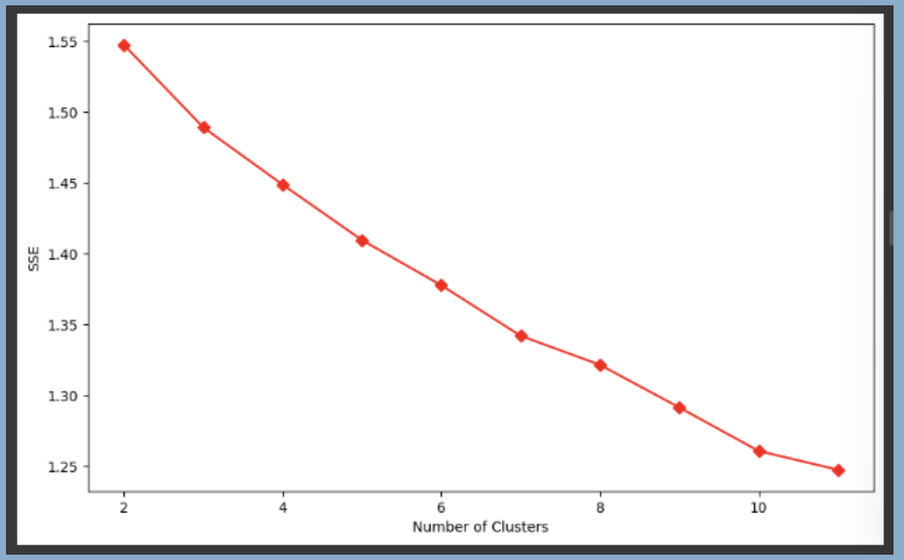
군집화

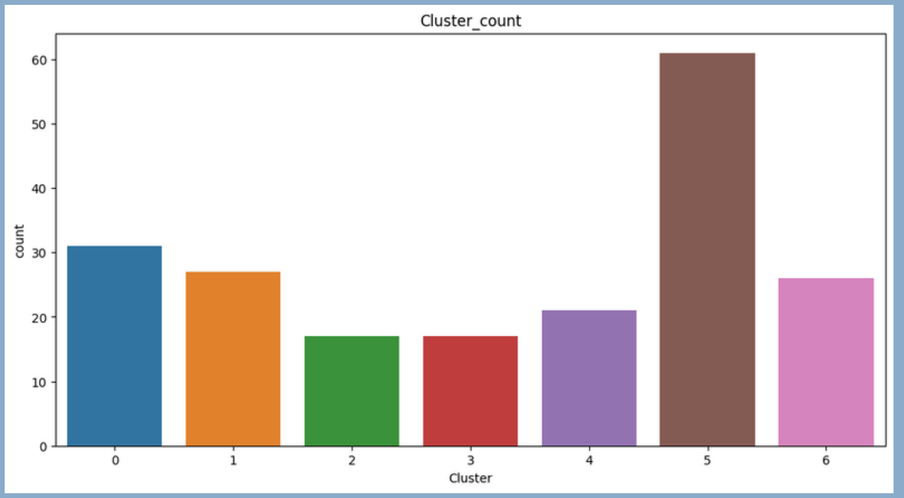
어떤 기준으로 그룹화를 진행할 것인가?

- “수익률” 높은 종목을 원할 것임
- 상관관계 분석 시 핵심 고려 사항: **포트폴리오 분산화**
  “계란을 한 바구니에 담지 마라.”
- 포트폴리오를 다각화하여 위험을 분산화하는 작업이 필요

- 어떤 기준?
  - Sector?
  - 변동성?
  - 가격 움직임?
  - 재무제표?
  - 수익률의 변화 패턴?

	도/소매업	화장품/백화점업	금융업	철강/건설/조선업	IT,반도체업	자동차업	화학/계약업
기준시점	100.00	100.00	100.00	100.00	100.00	100.00	100.00
비교시점	106.04	62.76	101.25	92.25	175.73	117.11	302.99
바이앤홀드수익률(%)	6.04	-37.24	1.25	-7.75	75.73	17.11	202.99





한계점

NASDAQ_RSS_IFO.CSV **활용 X**

- NASDAQ_RSS_IFO_202301.csv ~ NASDAQ_RSS_IFO_202308.csv
- RGS_DT : 발행일자
- TCK_IEM_CD : TCK_IEM_CD
- TIL_IFO : 제목정보
- CTGY_CFC_IFO : 카테고리분류정보
- MDI_IFO : 미디어정보
- NEWS_SMY_IFO : 뉴스요약정보
- RLD_OSE_IEM_TCK_CD : 관련해외종목티커코드
- URL_IFO : URL정보
- https://www.nasdaq.com/nasdaq-RSS-Feeds 에서 제공하는 Public한 나스닥 RSS 피드 컨텐츠
- 2023년 1월 1일 부터 2023년 8월 31일까지 발행된 정보(2023년 8월 31일 기준)

-> Fundamental에만 집중. 시장 Sentiment에 대한 정보 간과!

향후 계획

감성분석

- 뉴스 데이터, SNS 등을 분석하여 시장의 Sentiment 요소 함께 고려
- NLP 분야에 대한 공부
  - Transformer 기반 모델
  - GPT based 중요도 산정
  - 워드임베딩
- 그래프 기반 모델

포트폴리오 구현

국내/해외 종목 종합한 포트폴리오 구성

실질 투자에 도움이 될 수 있는 종목들 추천
(금융 서비스를 직접 만들어보기 위한 노력)

코드 공유

도메인 분야에 대한 지식 한계를 극복하기 위해 타 사례 코드 및 인사이트 참고

감사합니다!