

마켓 데이터 수집과 저장



이승준 fb.com/plusjune

마켓 데이터

- 금융 데이터의 실시간은 얼마나 중요한가?
- 내가 원하는 데이터를 충분히 얻으려면 비용은 얼마나 드나?
- 어디서 어떤 데이터를 얼마나 얻을 수 있을까?
- 개인과 기관의 정보력의 차이는 얼마나 될까?
- 해외시장, 외국인이 국내 주식시장에 미치는 영향은?
- 금융투자에서 거시경제에 대한 이해가 얼마나 필요한가?

실시간 데이터, 정말 중요한가

트레이딩 스타일에 따라 다르다

Trading Style	매매 간격	보유기간	비고
Position Trading	장기간	수 개월~ 수 년	—
Swing Trading	짧은 기간	몇 일~ 몇 주	—
Day Trading	수 시간	하루는 넘기지 않음	오버나잇 포지션 없음
Scalp Trading	초단타	수초~ 수분 단위로 수십번 거래	오버나잇 포지션 없음

데이터 수집 자동화

- 크롤링 *Crawling* : 주기적으로 가져와 저장
- 데이터 구매: 해외 데이터, 시세 데이터, 과거 데이터
- 증권사 트레이딩 API: 실시간 가격
- 구글 알리미: 이메일
- 사설 정보: 이메일, 문자

저장 방법

- CSV (Plain text): 가능한 UTF-8 인코딩 권장
- RDB: Oracle, MySQL(MariaDB), SQLite
- NoSQL: MongoDB

주요 수집 대상 데이터

- 종목 코드 (심볼 목록)
- 종목 기본 정보, 상태
- 재무정보
- 주가 데이터 (틱,초,분,일,주,월) - 틱,분,일
- 뉴스, SNS 텍스트 등 비정형 데이터

마켓 데이터 소스 5 (로그인도 필요 없는)

1. KRX 한국거래소 <http://krx.co.kr>
증권 시장 전체 데이터, 증권 통계, 표준코드 시스템
2. 전자공시시스템(금융감독원) DART <http://dart.fss.or.kr>
기업 공시, 재무제표
3. 포털 증권 (네이버, Daum) <http://stock.naver.com>
종목 정보 상세, 뉴스, 투자 정보

마켓 데이터 소스 5 (로그인 도 필요 없는)

4. 야후 파이낸스 <http://finance.yahoo.com>
세계/미국 증시, 뉴스, 종목 정보

5. FRED 연방준비은행 경제 데이터

<http://research.stlouisfed.org/fred2>

거시 분석, 해외 시황 (포괄적인 경제 데이터)

한국거래소

유가증권, 선물등의 거래시장의 개설및 관리, 유가증권상장등을 담당하는 통합거래소

<http://www.krx.co.kr>

- 일자별 주가: 전체, 과거 10년치 이상
- 뉴스/공시: 공시 전체, 뉴스(과거 6개월)
- 투자자별 거래실적 (일일, 과거 10년치 이상)
- 기관외국인동향: 일자별 매수,매도

한국거래소 - 한국거래소 표준코드 시스템

- 표준코드 조회

<http://isin.krx.co.kr/srch/srch.do?method=srchList>

- 조회 시작 일을 19970101 로, 주권을 설정하여 조회하면 기존의 모든 표준코드를 조회할 수 있다.
- 전체 코드는 8천여건
- 상장되지 않은 종목은 큰 의미가 없음

전자 공시 시스템

- DART 금융감독원 - "발행정보 중심"

<http://dart.fss.or.kr>

- KIND 한국거래소 - "유통정보중심"

<http://kind.krx.or.kr>

DART 금융감독원

<http://dart.fss.or.kr/>

- 발행정보 중심 전자공시시스템
- (상장 회사가 아니더라도) 공시 의무가 있는 회사들의 공시 포함
- 사업보고서를 내야할 기준에 해당되어(비상장외감법인) 정기공시
- 공정위에서 지정한 대규모기업집단에 소속된 비상장회사
- 증권신고서 등의 효력발생 또는 정정요구 조치공시 (금감원이 하는 공시)

KIND - 한국거래소

<http://kind.krx.or.kr>

- 유통정보 중심 전자공시시스템
- 한국거래소의 시장감시역할 수행
- 거래소에서 발령하는 특정한 시장조치 포함 (예: 투자유의, 상장폐지)
- 상장주식수의 변동 공시 (예: BW/CB 등의 행사, 추가상장/변경상장)
- 안내공시 (예: 대표이사, 본점소재지 변경 등)

포탈 증권

네이버 증권 - 주식 기본, 세부 기업정보, 해외주식, 추천종목, 뉴스, 조건검색, 투자전략 등
<http://finance.naver.com>

- 주요지수 (코스피, 코스닥, 코스피200) 차트(실시간)
- 투자자별 매매동향(외국인, 기관, 프로그램), 외국인 순매수/순매도
- 뉴스, 리포트, 투자전략
- 조건검색 (필터링) 및 TOP 종목(상하한가, 업종, 테마)
- 주요 해외 지수, 시장지표(환율, 금리, 유가, 금 시세, 원자제)
- MY금융: 관심종목, 모의투자

공공기관 오픈 데이터

- <http://data.go.kr> 공공데이터포털
- <http://kosis.kr> 국가통계포털
- <http://kostat.go.kr> 통계청 홈페이지

세계 주요 시장

- 한국: KRX (KOSPI, KOSDAQ)
- 미국: NYSE, NASDAQ, AMEX
- 중국: 후강통(홍콩-상해), 선강통(홍콩-심천)
- 일본: TSE(도쿄)
- 유럽(영국 LSE, 프랑스 파리거래소, 독일 FWB), 동남아(인도네시아, 태국, 베트남)

FRED2

<http://research.stlouisfed.org/fred2>

- FRED II (Federal Reserve Economic Data II)
- 세인트루이스 연방은행의 통계 사이트
- 데이터의 차트를 볼 수 있으며 데이터를 XLS나 TXT형태로 다운로드
- 금융 뿐만 아니라 재정, 물가, 고용 등 일반적인 경제 관련 데이터도 수록
- 특히, Download All Data 메뉴 압축된 엑셀 파일 (약 14M) FRED 전체 다운로드 가능

FRED2 API 도구

- <http://api.stlouisfed.org/docs/fred>
- <http://goo.gl/B1YNk> FRED 엑셀 애드인
- <https://github.com/zachwill/fred> FRED API Python 구현