

프로젝트 주제 : 관세청 수출 데이터로 주식가격 예측하기

과제 개요

2년 만에 대기업 시총도 제쳤다...한 달 사이 주가 22% 급등한 기업

13일 한국거래소에 따르면 에이피알은 지난 11일까지 한 달간 주가가 13만6000원에서 16만6300원으로 22.28% 상승했다. 같은 기간 달바글로벌은 55.93% 급등했다. 지난 5월 말 상장 당시 공모가(6만6300원) 대비 3.47배 수준인 23만원까지 치솟았다. ... 에이피알의 해외 매출 비중은 2023년 46%에서 2024년 64%, 올 1분기 71%로 늘었다. 달바글로벌 역시 해외 비중이 2023년 22%에서 2024년 46%, 올 1분기 56%로 급증했다.

<https://www.hankyung.com/article/202507113442i?>

여러분은 헤지펀드의 퀀트 트레이더가 되어, 수출 데이터를 활용한 투자 전략을 설계하는 실전 미션을 수행하게 됩니다.

실제 금융 시장에서 수출입 데이터는 단순 통계를 넘어 산업의 활력, 글로벌 수요, 기업 실적의 선행지표로 활용됩니다. 관세청이 매달 초 발표하는 수출 데이터는 기관투자자들이 면밀히 추적하는 대표적인 대체 데이터이며, 이를 먼저 분석하고 해석하는 능력은 시장에서 경쟁력이 됩니다.

제공 데이터

None

```
project/  
  data/  
    price_daily/  
      open.csv  
      close.csv  
      volume.csv  
    price_minutely/  
      open.csv  
      close.csv  
      volume.csv  
    export_value.csv  
    gics.csv
```

```
kospi.csv
src/
...
```

파일 스키마

Price Data

	symbol1	symbol2	...	symbol n
20200101				
...				
20241230				

export_value

date	symbol	export_value
2020-01-31	ZPWVVS	2848715.0
...

gics (sector, industry_group, industry, sub_industry)

symbol	value
AEGQRD	35101010

kospi

open	high	low	close
..

분석 환경

- Python 3.11 이상 권장
- 패키지 (pandas, numpy, matplotlib, scikit-learn, ...) 설치시 requirements.txt에 버전명시
- Long/Short 포지션 제약, 슬리피지, 스프레드, 거래수수료는 없다고 가정합니다

과제

여러분들은 실제 퀀트 트레이더가 새로운 데이터를 발굴했을 때 거치는 프로세스를 따라가게 됩니다.

1) 무역 수출 서프라이즈 정의하기

"이 데이터에서 어떤 시그널을 추출할 것인가?"

과거 데이터를 기반으로 스스로 예상치를 만들고, 그 결과를 활용해 수출 발표가 어느 정도 예상을 벗어났는지를 정의하세요. 또한 이후의 예측 모델링에 사용할 수 있도록 서프라이즈 지표를 명확히 설정하세요.

- 이 지표는 2, 3, 4번 과제의 공통 입력값으로 활용됩니다.
- 사용한 방법의 근거를 간단히 설명하고, 산출된 지표의 분포·결측치 처리 방식·기본적인 상관관계 등을 함께 제시합니다.
- 수출 실적이 없는 달은 어떻게 처리해야 하나요?
- 수출이 어쩌다 한 번 발생한 회사는 어떻게 처리해야 하나요?
- 수출 데이터는 명확한 계절 패턴(연말/연초, 명절 등)이 존재합니다.

2) 벤더 퀄리티 검증하기

****이 데이터는 정말 가치가 있는가?***

수출 데이터를 제공하는 벤더는 관세청에서 데이터를 마감한 후,, 다음 달 **1일 오전 10시 20분경** 데이터를 전달합니다. 1일이 주말인 경우에도 데이터를 업데이트하여 제공합니다. 해당 벤더는 데이터를 클라이언트에게 동시에 제공하며, 자사가 데이터를 독점적으로 보유하고 있어 가장 빠르게 제공한다고 주장합니다. 이 주장이 사실이라면, 그 시점 전후로 주가의 움직임에 차이가 나타나야 합니다.

아무리 좋은 시그널이라도, 시장이 이미 알고 있다면 무용지물입니다. 실제 금융사는 고가의 데이터 피드를 구독하기 전에 반드시 "이 데이터가 정말 먼저 나오는지", "가격에 반영되기 전에 받을 수 있는지" 검증합니다. 만약 벤더의 주장이 거짓이라면, 이 데이터의 투자 가치는 크게 떨어집니다.

공개 직전과 직후 구간을 구분하여 가격 반응을 비교하고, 벤더의 주장이 타당한지 검증하세요. 만약 그렇지 않다면, 반박할 수 있는 근거를 제시하세요.

- 적절한 통계적 검정 방법을 선택하고 그 결과를 제시하세요
- 전체 시장이 동시에 오르거나 내릴 때는 어떻게 처리해야 하나요?

3) 산업별 민감도 분석

****어떤 데이터에 집중할 것인가?***

모든 산업이 수출 변화에 동일하게 반응하는 것은 아닙니다. 제공된 **GICS** 코드를 활용하여 산업별로 수출 서프라이즈에 대한 주가 반응을 비교하고, 어떤 산업이 상대적으로 더 민감하게 움직이는지 분석하세요. 한정된 자본으로 최대 수익을 내려면, 가장 민감하게 반응하는 산업에 집중해야 합니다. 이 분석을 통해 여러분의 전략이 어떤 유니버스에서 작동할지 정의합니다.

- GICS: <https://www.msci.com/indexes/index-resources/gics>
- GICS 그룹 구성원이 너무 적은 경우는 어떻게 처리해야 하나요?

4) 규칙 기반 전략 제시

앞선 3단계의 분석 결과를 통합하여 재현 가능한 투자 전략을 설계하세요. 전략은 파이썬 코드 형태일 수도 있고, 학습된 모델을 기반으로 할 수도 있습니다. 어떤 방식이든 재현 가능한 형태로 제출해야 하며, 결과가 일관되게 도출되어야 합니다.

금융에서는 아무리 훌륭한 분석도 실제 수익률로 증명되지 않으면 의미가 없습니다. 여러분의 전략은 실제 **2025년** 데이터에서 테스트되며, 그 성과가 곧 여러분의 과제 점수입니다.

- **Forward Looking Bias** 방지 - 수출 데이터는 매월 1일 10시 20분에 공개됩니다. 당일 수익률을 예측 타겟으로 사용하면 미래 데이터를 사용하는 것입니다. 전략 설계 시 "데이터 공개 시점 이후"의 수익률만 사용하세요

평가

- 단순 시각화보다는 논리적 설계 근거와 재현성을 중점 평가합니다.
- 보고서 분량은 **A4 기준 5페이지 이내**, 발표는 **10분** 내외로 제한됩니다.

본 자료와 데이터, 코드, 산출물의 저작권은 **Presto Labs Pte. Ltd.**에 있으며, 외부 유출·재배포를 금합니다.