

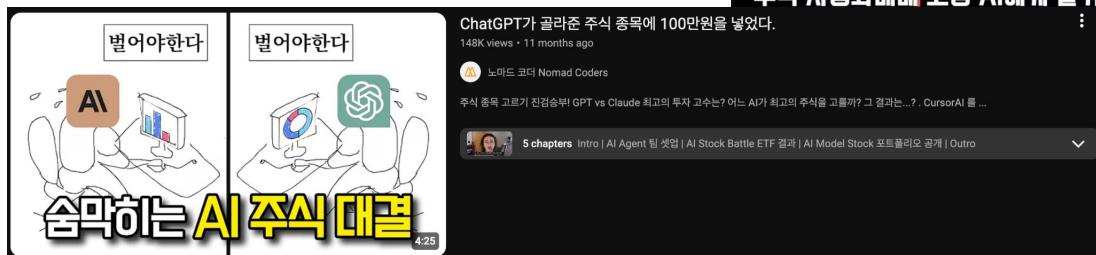
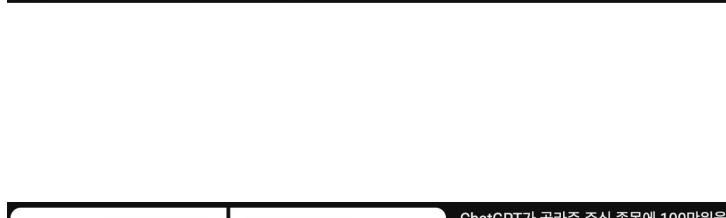
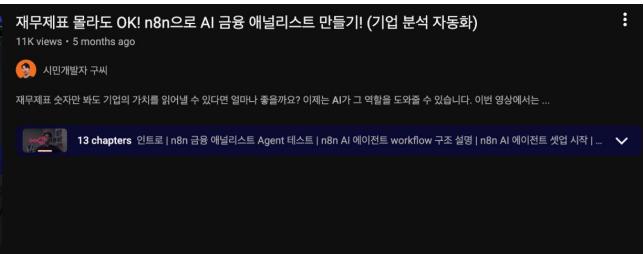
# Project 1

관세청 수출 데이터로 주식가격

예측하기

이수환, Vendor Data Team Lead, Senior Data Engineer

# AI? 주식?



# 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/
  ...

```

	AEGQRD	AIJFBS	AJAAJF	ANKPII	AOWNEI	
6256.451102	34434.599908	31132.638990	116101.486956	NaN	3	
6150.081764	34046.057173	30149.973197	113560.545323	NaN	2	
6469.189780	33317.539544	30819.972602	108087.747958	NaN	2	
5918.003207	33900.353647	31221.972244	111410.517786	NaN	3	
4941.339279	33171.836019	29256.640658	106719.548616	NaN	2	
...	...	...	...	...	...	
8260.000000	24858.045977	36289.159892	339239.080460	6832.0	1	
8100.000000	24957.279693	36045.934959	342420.689655	6808.0		
8040.000000	25900.000000	35900.000000	346000.000000	6720.0	1	
7800.000000	24750.000000	35650.000000	346800.000000	6596.0	1	
8110.000000	24750.000000	36350.000000	355600.000000	6796.0		

	date	symbol	export_value
0	2020-01-31	ZPWVVS	2865954
1	2020-01-31	YRTSYY	10836746
2	2020-01-31	OOVKZQ	861314197
3	2020-01-31	DKEETE	16163843
4	2020-01-31	AUBZAL	2657843
...	...	...	...
22654	2025-10-31	RNJWLR	5880114
22655	2025-10-31	CFFEFN	22842425
22656	2025-10-31	TVTYVW	13155702
22657	2025-10-31	YGUFCU	867967
22658	2025-10-31	IQRCIA	12899355

22659 rows × 3 columns

# 수출데이터

요즘 경제 뉴스를 보면...

 한국 경제 뉴스: 환율, 환율, 환율...

 미국 경제 뉴스: 고용률, 실업수당 청구...

→ 한국은 수출 중심 경제

수출이 GDP의 40% 이상 (2023년 기준)

미국은 약 10%

환율 → 수출 경쟁력 → 기업 실적

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?>

# 과제

## 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.

## 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.

# 과제

## 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.

## 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.

- Long/Short 포지션 제약, 슬리피지, 스프레드, 거래수수료는 없다고 가정합니다

Q & A