

미국 도소매 시장 동향 분석 및 중소기업 수출 기회 발굴 서비스

범주형자료분석팀

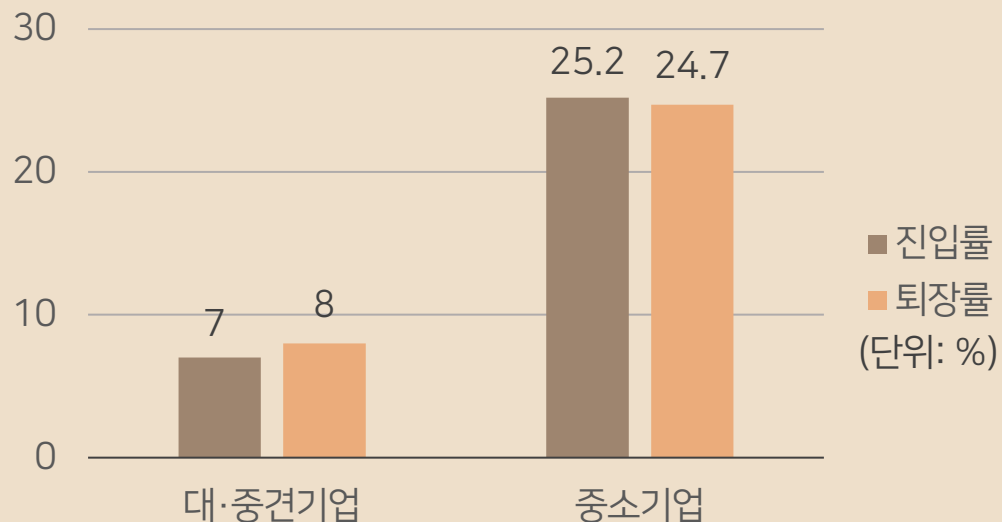
이정민
이경미
이현진
김준영
윤예빈

Part 1

주제 선정 배경

주제 선정 배경

◎ 대·중견기업 & 중소기업의 진입률 및 퇴장률



"소상공인은 인력 부족으로 현지 시장의 수요를 파악하기 어렵고, 그 수요에 접근할 수 있는 네트워크를 형성하기가 어렵다."*

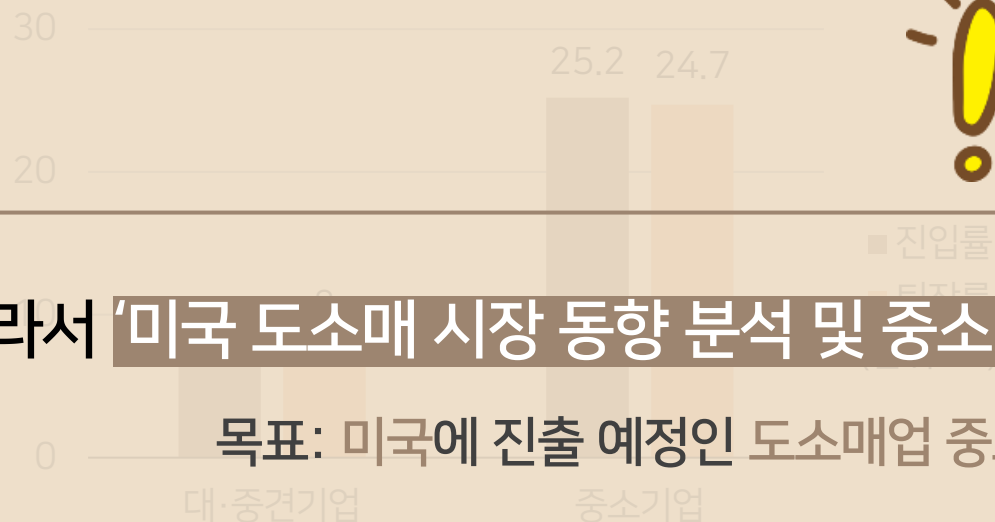


이진국 한국개발연구원 연구위원

한 해 중소기업의 25% 가량이 수출을 시도했다가 실패 후 퇴장함
이는 **시장 수요파악·네트워크 형성·수출 절차의 어려움** 때문으로 분석됨

주제 선정 배경

◎ 대·중견기업 & 중소기업의 진입률 및 퇴장률



"소상공인은 인력 부족으로 현지 시장의 수요를 파악하기 어렵고, 그 수요에 접근할 수 있는 네트워크를 형성하기가 어렵다"*

따라서 '미국 도소매 시장 동향 분석 및 중소기업 수출 기회 발굴 서비스'를 주제로 선정

목표: 미국에 진출 예정인 도소매업 중소기업에게 적절한 수출 타이밍 제시!



이진국 한국개발연구원 연구위원

한 해 중소기업의 25% 가량이 수출을 시도했다가 실패 후 퇴장함

이는 시장 수요파악·네트워크 형성·수출 절차의 어려움 때문으로 분석됨

서비스 흐름도

[Step 1]

BCG Matrix에서 사용되는 기준을 차용해 +/- 분류:

시장 경쟁력 (미국 기준 총 수입액 대비 對한국 수입액), 시장 성장률 (도소매 산업별 매출액 데이터)

[Step 2]

업종별 수요 예측

+

실시간 트렌드 분석

↳ Amazon 판매량/리뷰 감정 분석, 기후(계절 특성) 분석 등

시장 진입 위험도 지수 산출

-

- 기업 데이터 (기업/업장 수 등)
- 경제지표 (산업별 GDP 등)
- 규제 관련 데이터 (관세 등)

[Step 3]

현재 시점으로부터 1년 이내
미국 시장 진출 최적 타이밍 제시

위험 지수 산출 결과값
&
메인 risk 요인 출력

도소매 시장
內
업종 pool

시장 진입 여부
판단

서비스 흐름도

[Step 1]

BCG Matrix에서 사용되는 기준을 차용해 +/- 분류:

시장 경쟁력 (미국 기준 총 수입액 대비 對한국 수입액), 시장 성장률 (도소매 산업별 매출액 데이터)

[Step 2]



[Step 3]



종별 수요 예측

기존의 BCG Matrix에서는 '상대적 시장점유율', '시장성장률' 을 고려함

Amazon 판매량/리뷰 감정 분석, 기후(계절 특성) 분석 등

현재 시점으로부터 1년 이내
미국 시장 진출 최적 타이밍 제시

$$\text{상대적 시장 점유율} = \frac{\text{자사 시장 점유율}}{\text{주요 경쟁사 시장 점유율}}$$

장 진입 위험도 지수 산출

기업 데이터 (기업/업종 수 등)

경제지표 (산업별 GDP 등)

규제 관련 데이터 (관세 등)

$$\text{對미국 수출시장에서} \frac{\text{자사 시장 점유율}}{\text{주요 경쟁사 시장 점유율}} = \frac{\text{對한국 수입액}}{\text{총 수입액}} \quad (\text{미국 기준})$$

위험 지수 산출 결과값

메인 risk 요인 출력

과 유사한 의미를 가지는 것으로 파악함

시장 진입 여부 판단

◎ 시장 경쟁력 (미국 기준 총 수입액 대비 對한국 수입액)

미국의 對한국 수입액

HS code	2024-10	2024-09	2024-08	2024-07	...	2018-11
00-Miscellaneous Cargo	123184406.0	29691265.0	7.236021e+07	90320914.0	...	53959088.0
01-Live Animals	NaN	NaN	NaN	NaN	...	NaN
...
98-Special Classification Provisions	1662462.0	2238129.0	1.069399e+10	219312703.0	...	76326209.0

미국의 총 수입액

HS code	2024-10	2024-09	2024-08	2024-07	...	2018-11
00-Miscellaneous Cargo	482324502	1212221693	810778657	2320757470	...	1827618134
01-Live Animals	89232	247075	11337	200343	...	569905
...
98-Special Classification Provisions	13932041	630726847	23883264816	6469795064	...	466279326

시장 진입 여부 판단

◎ HS 코드란?

HS (Harmonized System) 코드

세계관세기구(WCO)가 개발한 국제 상품 분류 체계
수출입 상품의 종류와 세부 내역을 식별하는 데 사용

NAICS (North American Industry Classification System) 코드

미국, 캐나다, 멕시코 간 통일된 산업 분류 체계로, 산업의 유형에 따라 분류
2자리에서 6자리까지 구성된 코드로, 자리수가 많아질수록 더 세부적인 산업을 나타냄

시장 진입 여부 판단

◎ HS 코드란?

HS (Harmonized System) 코드

세계관세법주팀의 주제분석에서는 최종적으로 체계
모든 데이터들의 분류 체계를 **NAICS 코드로 통일**해야함!

NAICS (North American Industry Classification System) 코드

미국, 캐나다, 멕시코 간 통일된 산업 분류 체계로, 산업의 유형에 따라 분류
2자리에서 6자리까지 구성된 코드로, 자리수가 많아질수록 더 세부적인 산업을 나타냄

시장 진입 여부 판단

◎ 시장 경쟁력 데이터 결측치 처리



결측치가 존재하는 행들이 NAICS 코드에 매핑하기 어려운 항목들임



한국에서 해당 산업들이 차지하는 비율 등 산업적인 특징으로 인해 발생한 결측치임을 고려

⇒ 결측치가 존재하는 열을 제거하고 분석 진행

[제거된 결측치 존재 HS code 항목]

'01-Live Animals', '05-Other Products From Animals', '06-Trees', '14-Vegetable Plaiting Materials Etc', '18-Cocoa, Cocoa Prep',
'26-Ores, Slag And Ash', '31-Fertilizers', '41-Raw Hides, Leather', '43-Furs, Artificial Fur', '45-Cork', '50-Silk', '51-Wool',
'53-Vegetable Textile', '66-Umbrella, Walking-Sticks Etc', '78-Lead And Articles', '79-Zinc And Articles', '80-Tin And Articles',
'88-Aircraft, Spacecraft', '91-Clocks, Watches', '97-Arts, Collectible Pieces, Antiques'

시장 진입 여부 판단

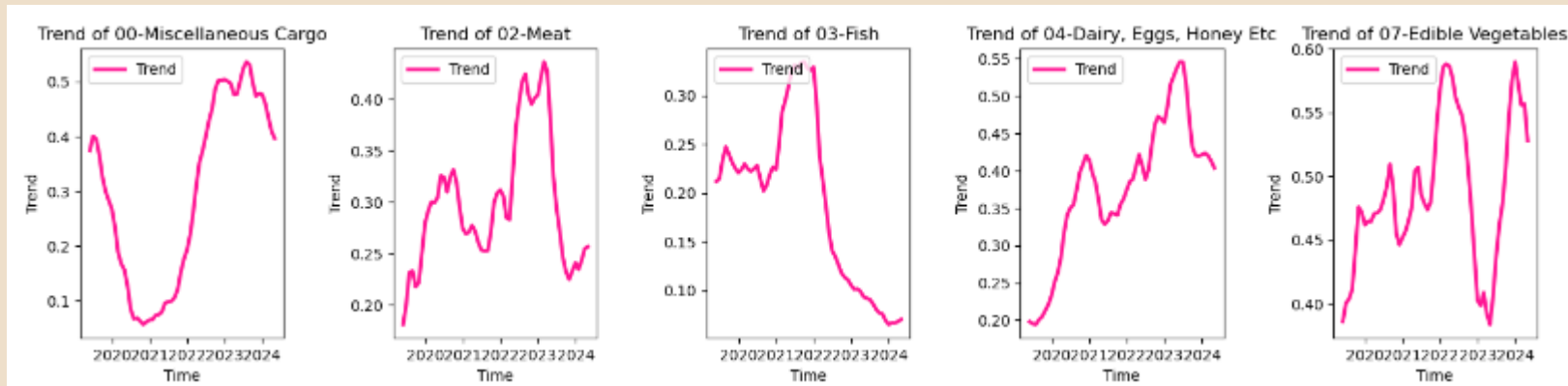
◎ 시장 경쟁력 데이터 EDA

시계열 덧셈 분해

시계열 데이터를 계절성, 추세, 불규칙성을 더한 합으로 표현하는 것

$$y_t = S_t + T_t + R_t$$

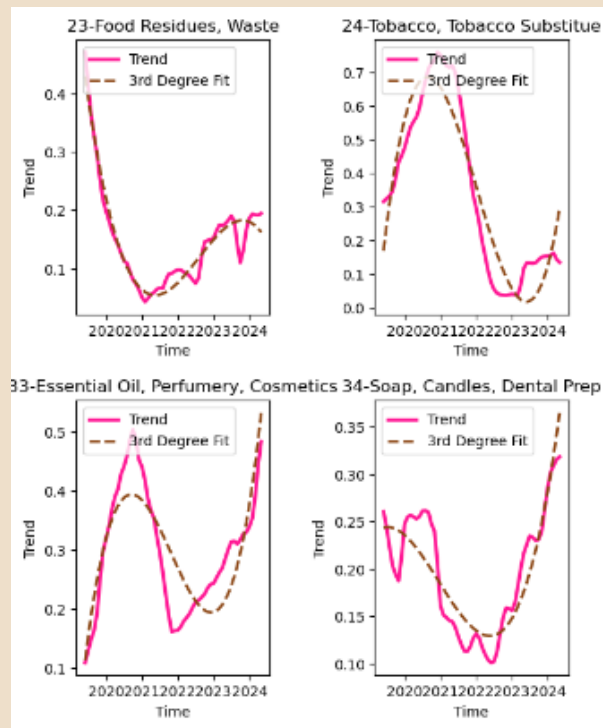
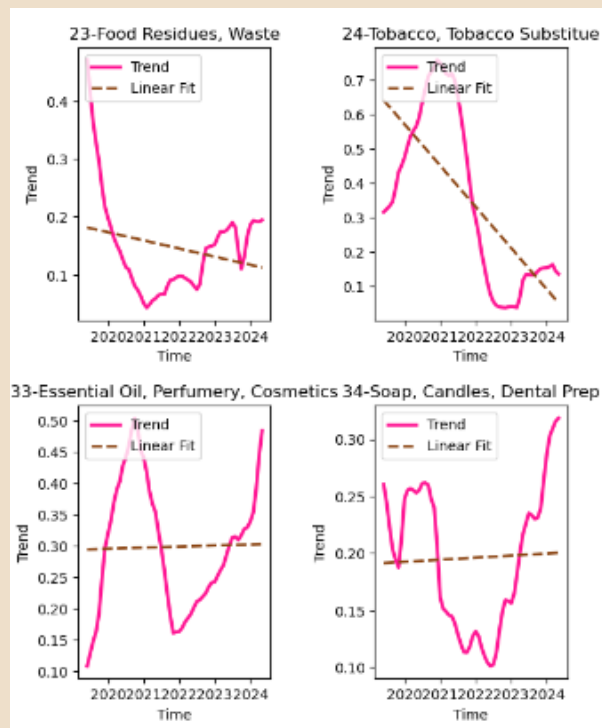
덧셈 분해를 이용해 추출한 **추세를 이용**하여 plus / minus 그룹을 분류하기로 결정



시장 진입 여부 판단

◎ 파생변수 생성

추세 그래프에 3차 회귀선 적합



(또 반함)



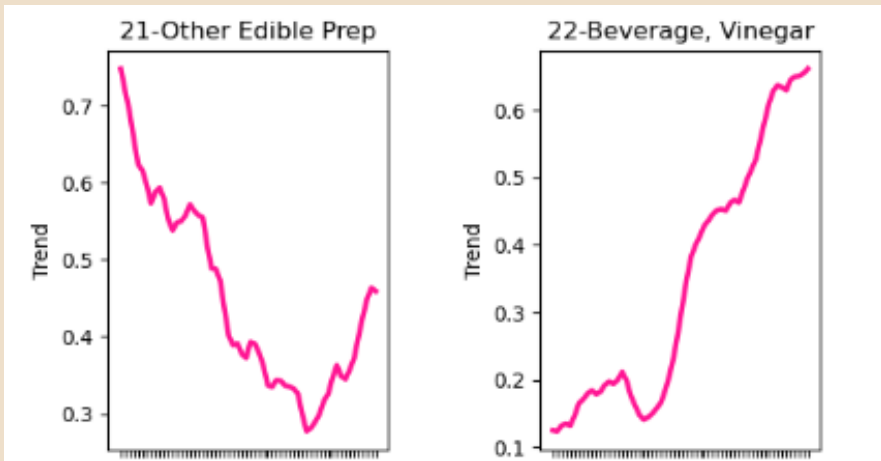
1차 회귀 직선에서
변동이 심한 직선의 추세를
잘 반영하지 못하던 문제 해결!



3차 계수 값을
추세선 증감 지표로 활용 !

시장 진입 여부 판단

◎ 파생변수 생성



감소했다가 증가하는 21번 항목보다,
지속적으로 증가하는 22번의 항목의 분산이 더 크게 계산됨



즉, 단순한 분산 값으로는 변동성을 정확하게 반영할 수 없음

추세의 1차 미분값의 분산을 변동성 지표로 사용

시장 진입 여부 판단

◎ 파생변수 생성

변동성이 클수록 시장 진입에 불리하므로 역수를 취함

(3차 계수) / (추세 1차 미분값의 분산)

회귀선의 3차 계수를 통해 **추세선의 증감**을 고려하고 분산을 통해 **변동성**을 고려함



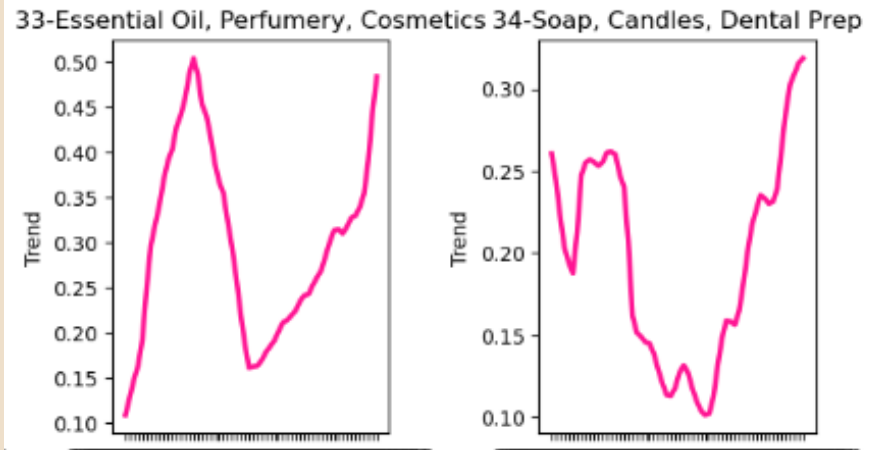
추세의 스케일이 작은 경우 그래프 상에서
변동이 심해도 분산이 너무 작게 계산됨



스케일링을 통해서 해당 분산 값이 커질 수 있도록
조정이 필요하다고 판단해
모든 과정은 minmax scaling을 적용하고 진행함

시장 진입 여부 판단

◎ 파생변수 생성

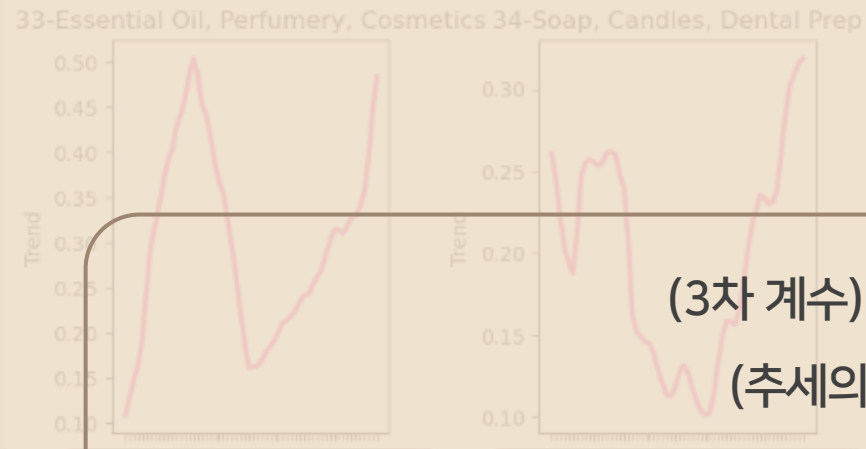


두 항목 모두 큰 폭으로 감소했다가 증가하지만
33번 항목은 과거의 max값과 비슷한 정도로 증가하는 상태이고
34번 항목은 과거보다 더 많이 증가하고 있음

추세의 **가장 최신값/최대값**을 추가적으로 이용
해당 업종의 '잠재력' 고려

시장 진입 여부 판단

◎ 파생변수 생성



두 항목 모두 큰 폭으로 감소했다가 증가하지만
33번 항목은 과거의 max값과 비슷한 정도로 증가하는 상태이고
34번 항목은 과거의 max값보다 더 많이 증가하고 있음

두 변수 모두 임계값을 넘을 경우 Plus Group, 하나라도 넘지 못할 경우 Minus Group으로 분류

추세의 가장 최신향/최댓값을 추가적으로 이용
해당 업종의 '잠재력' 고려

위험도 지수 산출

○ 위험도 지수 카테고리

미국의 기업, 사업체, 직원 수 데이터 활용



기업

AHS, MFN 관세율 데이터 활용



관세

GDP, PPI, 주가지수 데이터 활용



경제지표

◎ 파생변수 생성

Establishments / Firms

사업체 수를 회사 수로 나눈 값으로, 회사의 평균 사업체 수를 의미
경쟁 회사의 규모를 알 수 있는 지표로 이 값이 커지면 수출 위험도는 높아짐

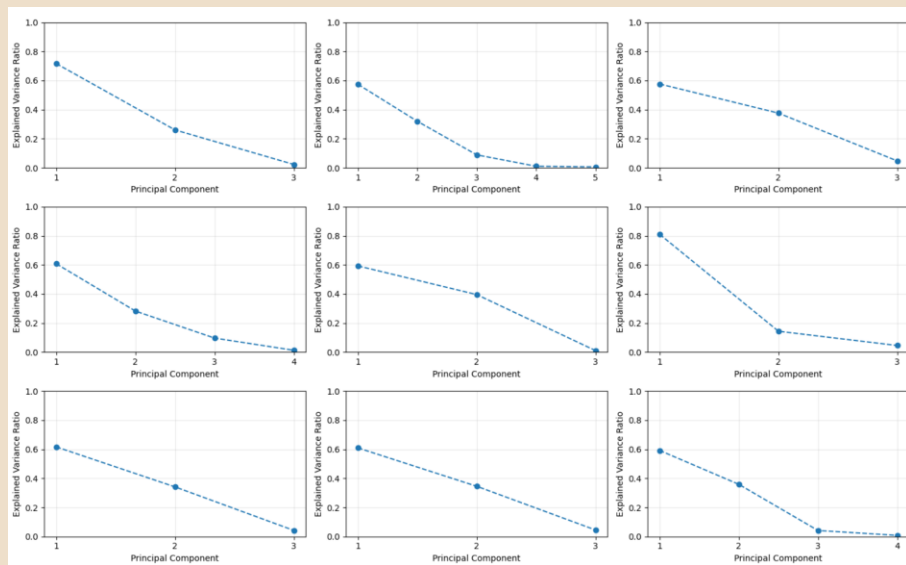
Employment / Establishments

직원 수를 사업체 수로 나눈 값으로, 사업체의 평균 직원 수를 의미
경쟁 회사의 인력을 알 수 있는 지표로 이 값이 커지면 수출 위험도는 높아짐

◎ 주성분 분석

주성분 분석 (Principal component analysis)

여러 개의 독립변수들을 잘 설명해줄 수 있는 주된 성분을 추출하는 기법으로
원 데이터의 분포를 최대한 보존하면서 고차원 공간의 데이터들을 저차원 공간으로 변환



파생변수까지 고려하여 PVE 값을 계산한 결과
변수 Firms, Establishments, Employment만
고려했을 때 PC1의 PVE 값이 가장 높았음

◎ 주성분 분석

주성분 분석 (Principal component analysis)

여러 개의 독립변수들을 잘 설명해줄 수 있는 주된 성분을 추출하는 기법으로
원 데이터의 분포를 최대한 보존하면서 고차원 공간의 데이터들을 저차원 공간으로 변환

그럼 시작합니다



따라서 **Firms, Establishments, Employment** 세 가지 변수로 PCA 진행!

파생변수까지 고려하여 PVE 값을 계산한 결과
변수 Firms, Establishments, Employment만

고려했을 때 PC1의 PVE 값이 가장 높았음

◎ 주성분 분석

주성분 분석에서의 loading 값의 의미

주성분을 구성하기 위해 원래 변수(feature)에 할당된 가중치로,
실제 변수가 주성분 결정에 얼마나 많은 영향을 주었는지를 의미하는 값

	PC1	PC2	PC3
Firms	0.805933	-0.593471	0.136648
Establishments	0.990239	-0.022039	-0.198365
Employment	0.750024	0.666807	0.115063

Ex. Firms의 가중치 (W_{Firms})

$$= \frac{0.805933}{0.805933 + 0.990239 + 0.750024}$$

세 가중치를 합하면 1이 되도록 전체 PC1의 loading 값의 합에 대한
각각의 변수의 loading 값의 비율을 가중치로 사용 !

◎ 기업 지수 산출



최종 기업 지수 산출 공식

X는 모두 표준화된 값!

$$W_{Firms}X_{Firms} + W_{Establishments}X_{Establishments} + W_{Employment}X_{Employment}$$

NAICS	Firms	Establishments	Employment	Total_sum
42	0.056784	-0.109045	-0.154292	-0.073272
423	-0.286719	-0.604420	-0.470022	-0.474591
...
454390	1.870651	1.431153	0.282364	1.238070

1주차 REVIEW

기업 | 관세 | 경제지표

◎ 관세 지수 산출 in 1주차



수출 시장에 갓 진입하는 중소기업은 해당 업종의 시장 규모보다는
관세가 실제로 얼마나 부과 되는지의 영향을 더 많이 받는다고 판단



중소기업 입장에서 더 중요한 관세 자체 수치, 그 중에서도 실질적으로 적용되는
관세율을 의미하는 AHS Weighted Average를 사용하기로 결정!

1주차 REVIEW

기업 | 관세 | 경제지표

◎ 관세 지수 산출 in 1주차



수출 시장에 갓 진입하는 중소기업은 해당 시장 업종의 규모보다는
관세가 실제로 얼마나 부과 되는지의 영향을 더 많이 받는다고 판단



중소기업 입장에서 더 중요한 관세 자체 수치, 그 중에서도 실질적으로 적용되는
관세율을 의미하는 AHS Weighted Average를 사용하기로 결정 !

위험도 지수 산출

기업 | 관세 | 경제지표

○ 경제지표 카테고리

GDP

주가지수

PPI

매출액

위험도 지수 산출

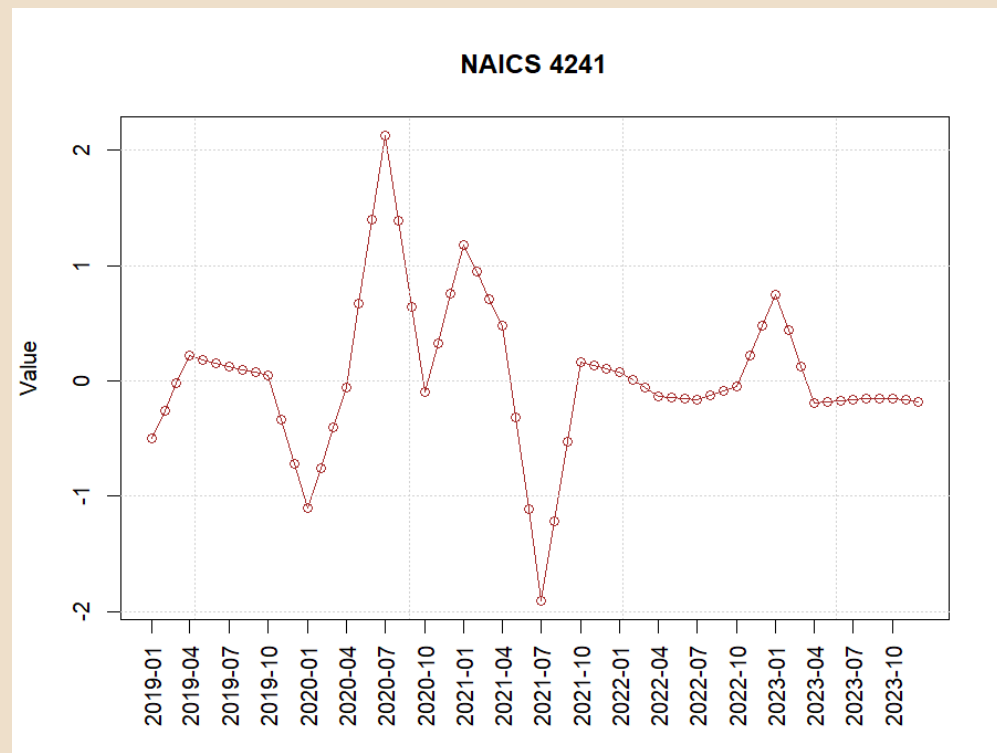
○ GDP 데이터 설명

선형 보간 (Linear Interpolation)

두 점 사이의 값을 선형적으로 채워 넣는 방법

EX) 분기별 데이터가 있을 때
분기 사이의 빈 월을 두 분기 값을 잇는
직선 상의 위치에 각 월의 값을 배치

선형보간 적용 결과



GDP가 일정 기간 내에 생산되는 최종 생산물의 “합”인 점을 고려하여 선형 보간을 사용 !

○ 데이터의 문제점

특정 시점 이전의 데이터가 없는 경우

Prophet은 주로 미래 예측에 사용되지만
데이터의 계절성과 추세가 안정적이고 반복적이라면 **과거 예측**에도 적절하게 활용



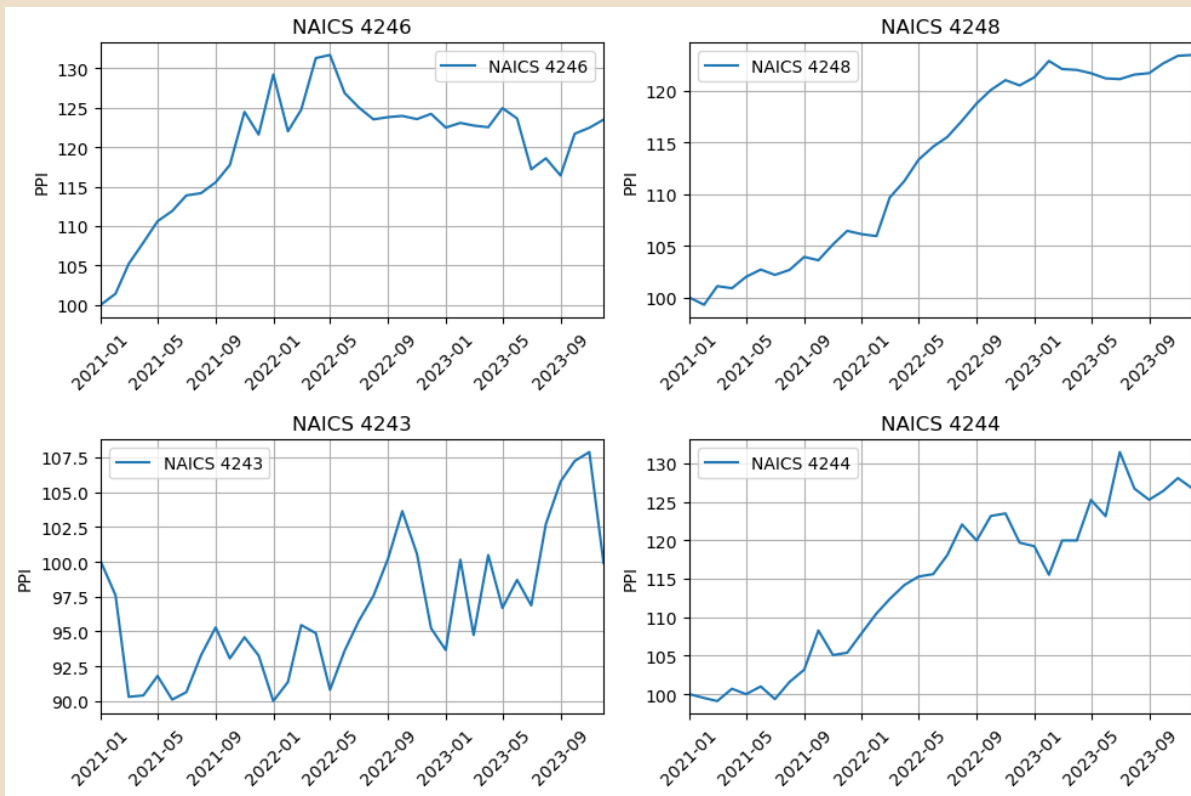
2021년 이전의 데이터가 없는
4243, 4244, 4246, 4248, 4249의 기존 데이터의 추세를 먼저 확인

위험도 지수 산출

기업 | 관세 | 경제지표

○ 데이터의 문제점

기존 데이터의 추세



비교적 안정적으로 증가하는 추세를 보임



Prophet을 이용하여

2019년 1월 ~ 2020년 12월의

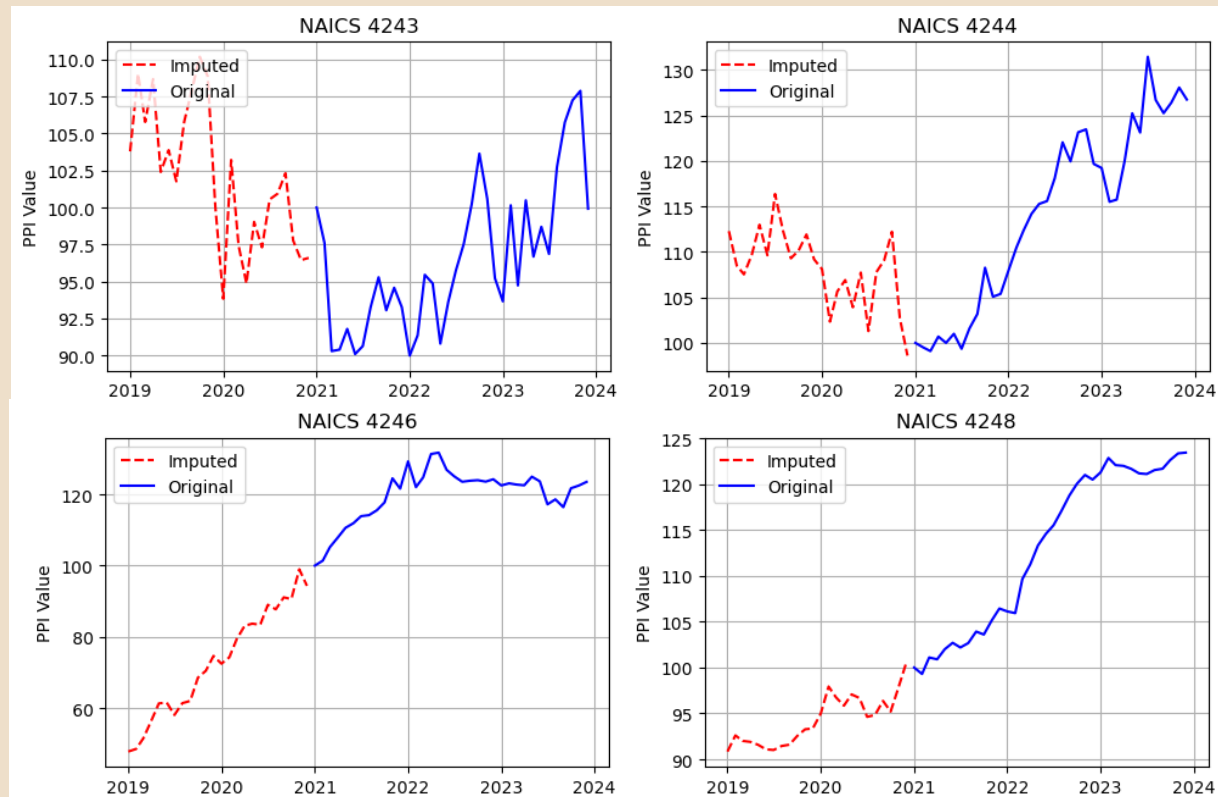
결측치를 보간 !

위험도 지수 산출

기업 | 관세 | 경제지표

○ 데이터의 문제점

보간된 데이터



빨간 점선이 **보간된 결측치**
파란 실선이 **원본 데이터**로
모델이 추세와 계절성, 변동을 반영하여
결측치를 보간한 것을 확인 가능



○ 데이터의 문제점

특정 시점 이후의 데이터가 없는 경우

442110, 442299, 443142, 445530, 446110, 448110, 448120,
448210, 448310, 451120, 451211, 452111, 453210, 453220 는
2022년 8월 까지만 데이터가 있고 2022년 9월 ~ 2023년 12월의 데이터가 없음

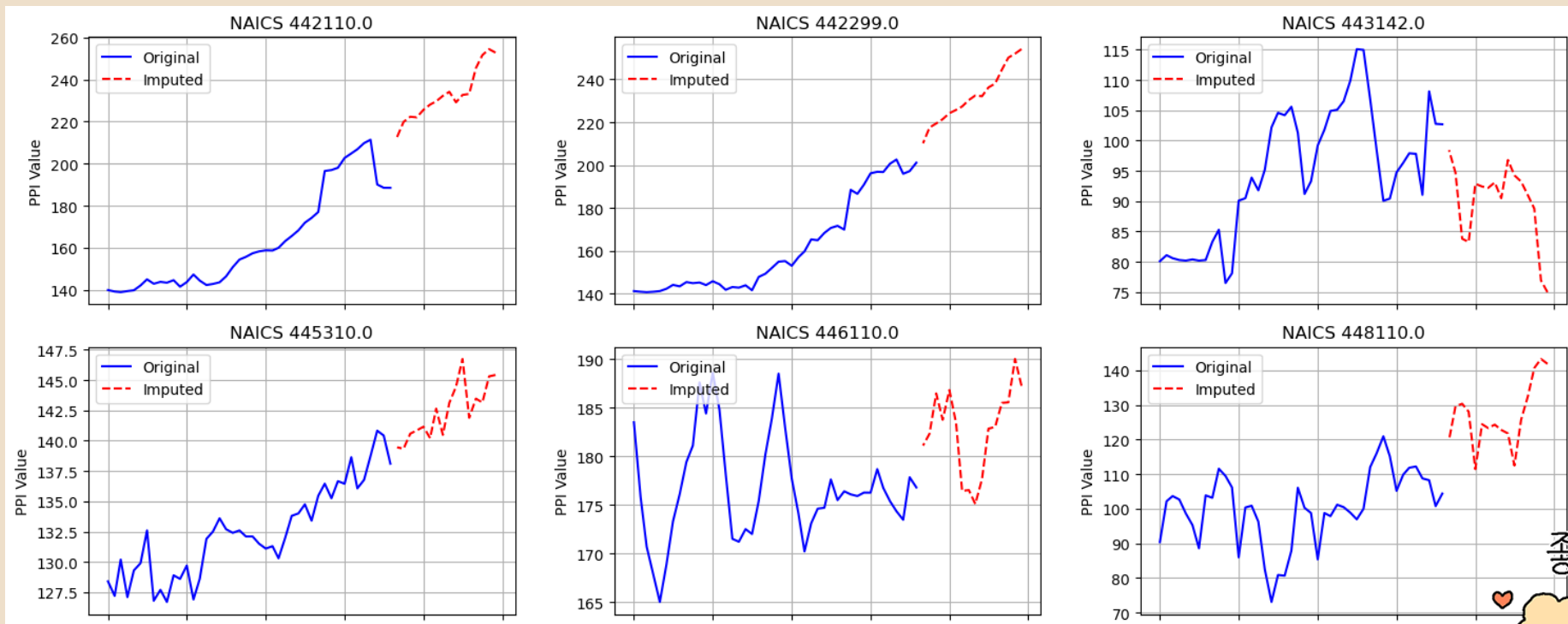


마찬가지로 Prophet을 이용하여 결측치를 보간!

위험도 지수 산출

기업 | 관세 | 경제지표

◎ 최종 데이터 plot



NAICS 코드별 월별 PPI 데이터의 문제 해결 완료 !

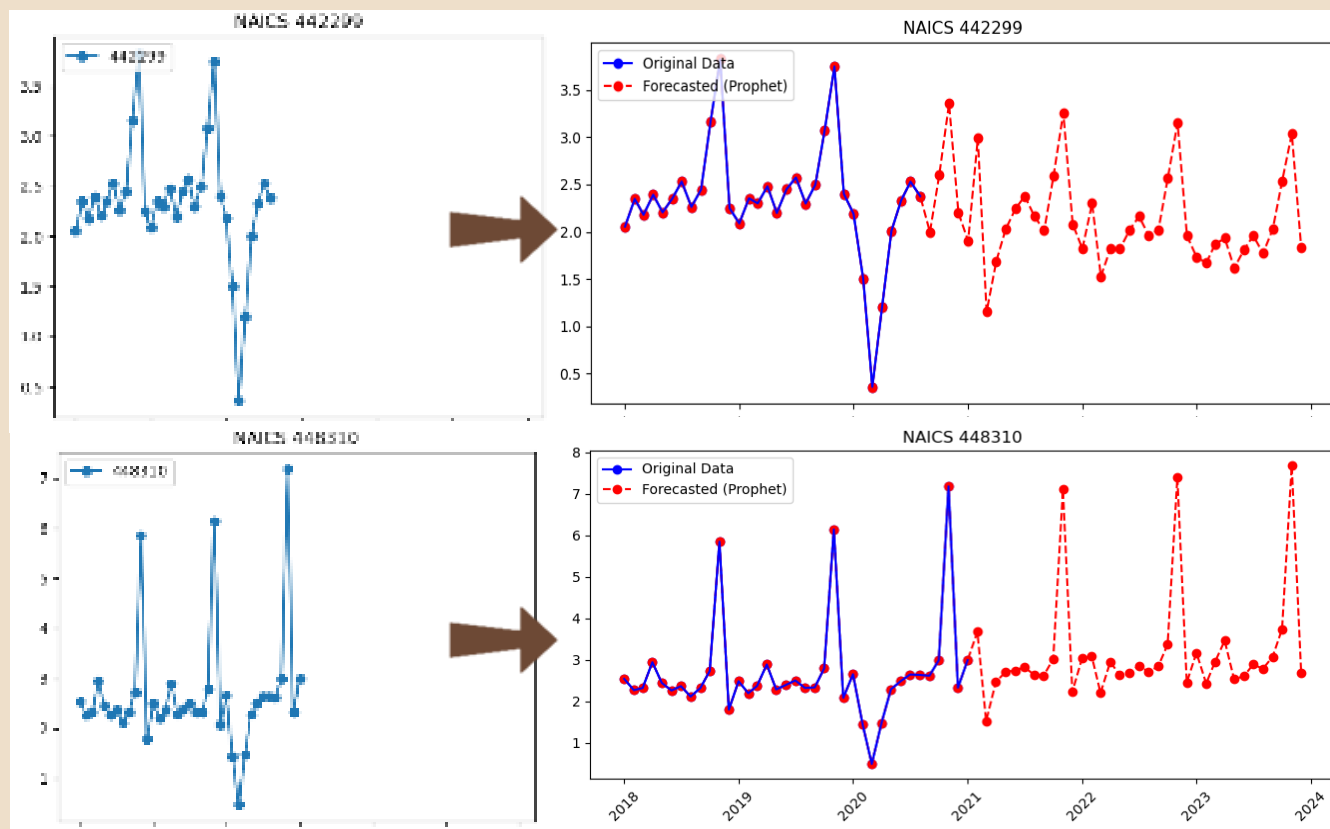


위험도 지수 산출

기업 | 관세 | 경제지표

매출액 결측치 보간

빨간 점선이 보간된 결측치, 파란 실선이 원래 데이터!



동일하게 Prophet 모델을 이용해
과거의 값들로 결측치 예측

Growth를 'linear'에서 'logistic'으로
바꾸고 상한과 하한을 설정하여 예측값이
너무 극단적이지 않도록 파라미터 조정

◎ 경제지수 산출

DFM (Dynamic Factor Model)

여러 시계열 변수 간에 공통적인 요인(factor)을 추출하여
시간 변화에 따른 상관 구조를 분석하는 모델로 다변량 시계열 데이터에 많이 사용됨

$$f_t = \Phi f_{t-1} + \epsilon_t$$

f_t : 추정된 공통 요인, Φ : 요인 변동의 상관성을 나타내는 행렬, ϵ_t : 오차항

$$x_t = \Lambda f_t + \eta_t$$

x_t : 각 시점에서의 관측된 데이터, Λ : 로딩값, η_t : 변수별 고유 오차

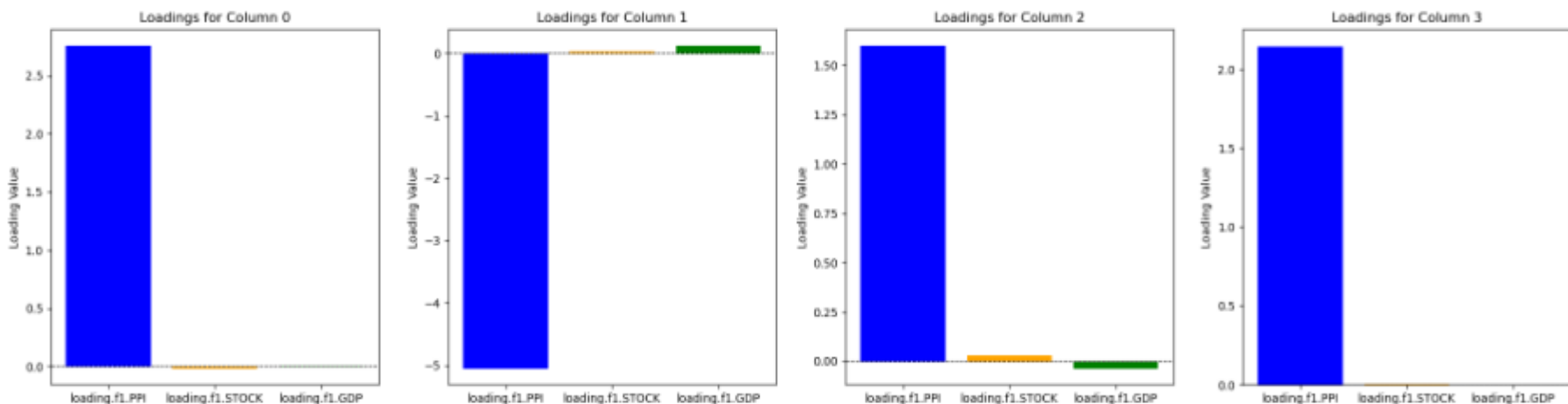
위험도 지수 산출

기업 | 관세 | 경제지표

○ 경제지수 산출



요인 분석이 적절하지 않은 데이터 형태



Loading 값을 출력했을때 PPI 변수가 압도적으로 큰 로딩값을 가짐

해당 요인은 거의 PPI로 설명되고 있다는 의미이므로

세 개의 시계열 변수가 하나의 요인으로 압축되었다고 보기에는 어려움

◎ 경제지수 산출

공적분

정상성을 만족하기 위해 필요한 차분 횟수인 적분 차수(Order of integration)가 모두 d 일때,
시계열의 선형 결합의 적분 차수가 d 보다 작다면 시계열 사이에 공적분 관계가 존재

비정상 시계열을 차분하는 과정에서 장기적인 관계에 대한 정보는 잃어버림
소득과 소비 등 경제 분야 시계열 자료들에서는 각각은 비정상적이지만 장기적으로
균형 관계를 갖는 경우가 많기 때문에 해당 이론 적용

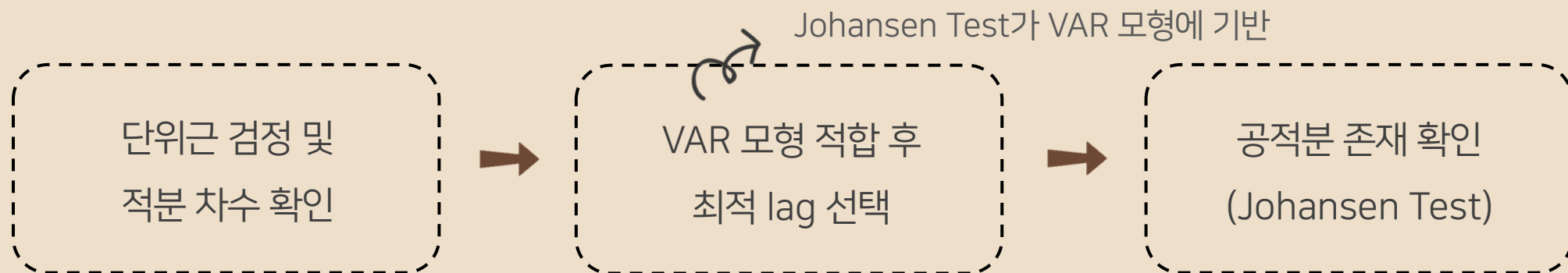
1주차 REVIEW

기업 | 관세 | 경제지표

○ 경제지수 산출

공적분

정상성을 만족하기 위해 필요한 차분 횟수인 적분 차수(Order of integration)가 모두 d 일때,
시계열의 선형 결합의 적분 차수가 d 보다 작다면 시계열 사이에 공적분 관계가 존재



지수 산출을 위해 위의 과정에서 모든 변수 업종 별로 **min-max scaling** 실시

1주차 REVIEW

기업 | 관세 | 경제지표

○ 경제지수 산출

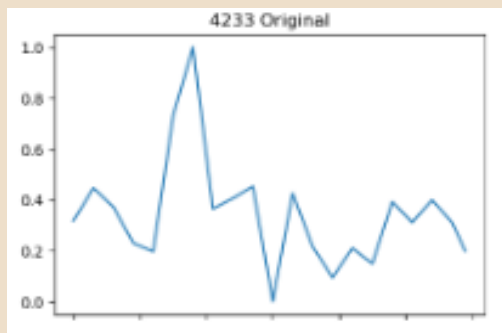


단위근 검정 및 적분 차수 확인

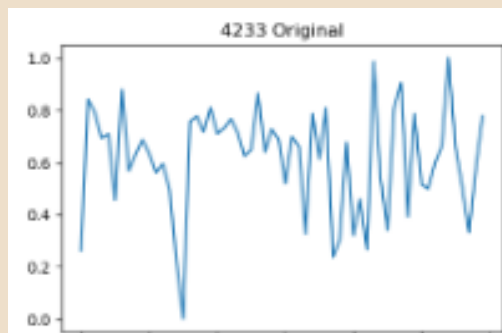
각 변수의 timeplot, ACF plot 확인 후 **ADF 검정** 실시

대표적인 단위근 검정 방법

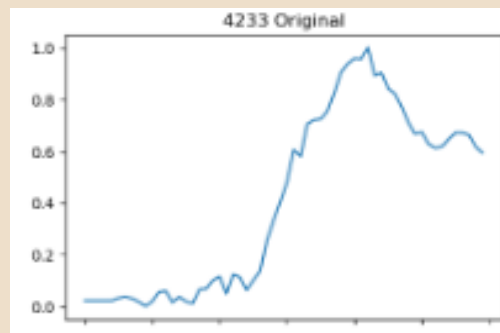
Ex) NAICS code 4233 업종의 timeplot



GDP



주가지수



PPI



매출액

1주차 REVIEW

기업 | 관세 | 경제지표

○ 경제지수 산출



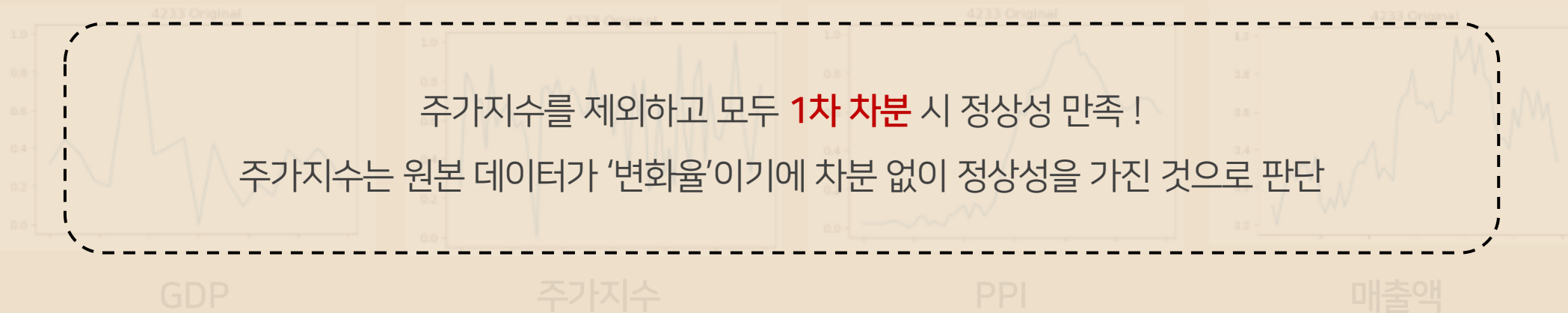
단위근 검정 및 적분 차수 확인

각 변수의 timeplot, ACF plot 확인 후 ADF 검정 실시

대표적인 단위근 검정 방법



Ex) NAICS code 4233 업종의 timeplot



1주차 REVIEW

기업 | 관세 | 경제지표

○ 경제지수 산출



VAR 모형 적합 후 최적 lag 선택

단위근 검정이 종속변수(y_t)와 설명변수(y_{t-1}) 간의 상관관계 존재 유무를 나타내는 계수의 유의성을 파악하는 것이라면,
VAR 모형을 이용한 공적분 검정에서는 VAR(p) 모형을 따르는 벡터 시계열 Y_t 의 구성 변수 간 공적분 관계가 존재하는지를 검증

검증 이전에 최적의 lag를 선택하기 위해 VAR 모형 적합



BIC 기준 $p=2$ 가 선택됨 !

1주차 REVIEW

기업 | 관세 | 경제지표

○ 경제지수 산출



공적분 존재 확인 (Johansen Test)

차분 차수 d 가 1로 동일했던 GDP, PPI, 매출액에 대해

Johansen Test를 실시해 공적분 존재를 확인



대부분의 NAICS 코드에서 공적분이 2개 존재한다는 결과 도출 !

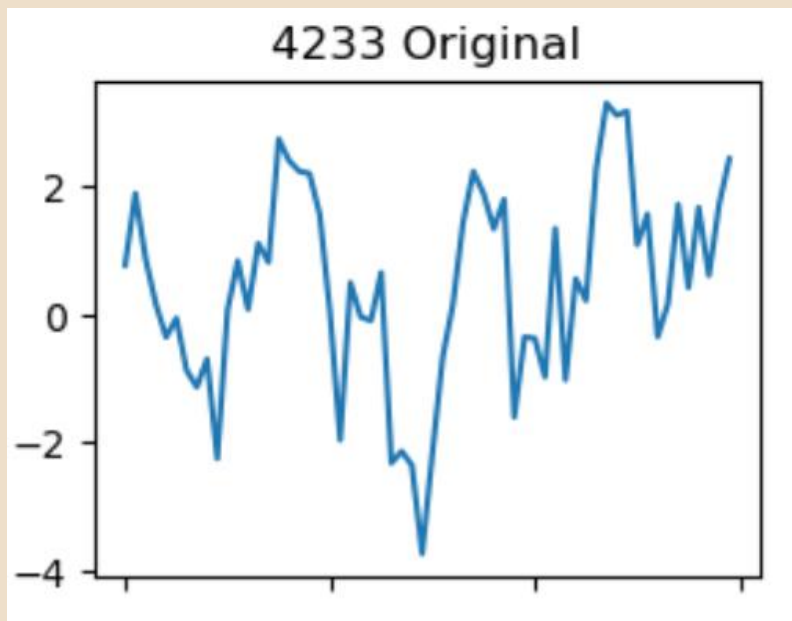


해당 변수 3개 모두 장기 균형 관계에 있다는 의미로 해석 가능

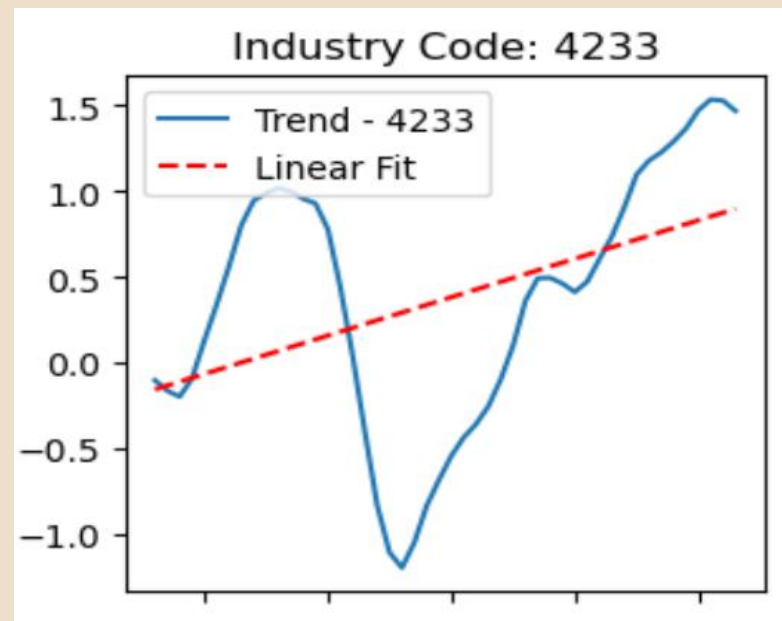
1주차 REVIEW

기업 | 관세 | 경제지표

○ 경제지수 산출



공적분 벡터와 원본 데이터 내적을 활용해서
세 변수의 공통 요인을 새로운 시계열로 도출

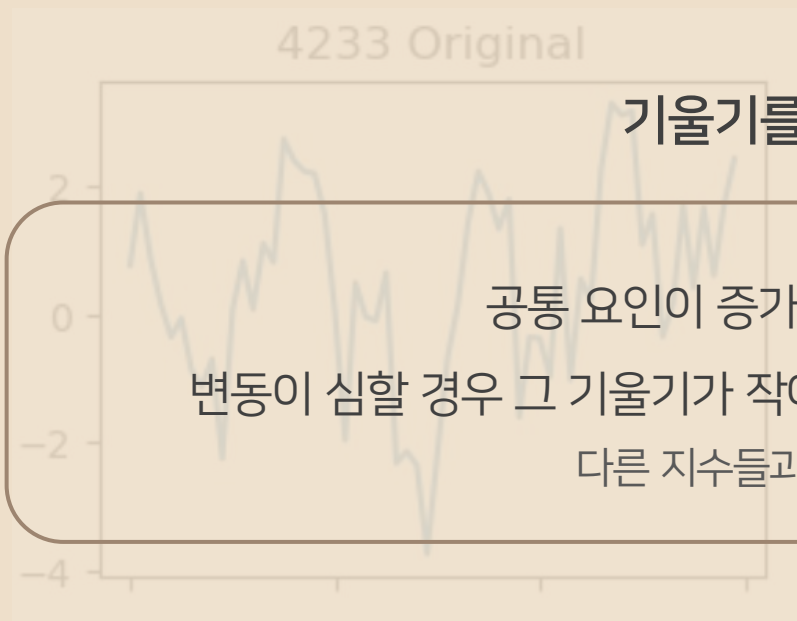


시계열 덧셈 분해 후 1차 회귀선 적합
이 때의 **기울기**를 경제 지수로 사용!

1주차 REVIEW

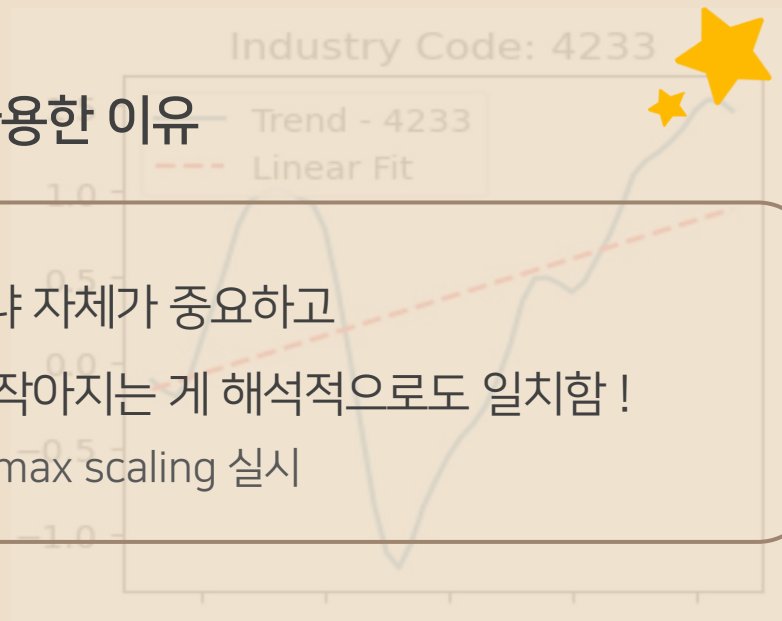
기업 | 관세 | 경제지표

◎ 경제지수 산출

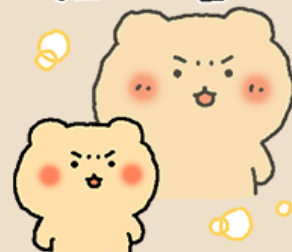


기울기를 지수 값으로 사용한 이유

공통 요인이 증가하느냐 감소하느냐 자체가 중요하고
변동이 심할 경우 그 기울기가 작아져서 지수 값도 작아지는 게 해석적으로도 일치함!
다른 지수들과 마찬가지로 min-max scaling 실시



(또 반함)



공적분 벡터와 원본 데이터 내적을 활용해서
세 변수의 공통 요인을 새로운 시계열로 도출

시계열 덧셈 분해 후 1차 회귀선 적합
이 때의 기울기를 경제 지수로 사용!

1주차 REVIEW

◎ 최종 위험도 지수 산출

위험도지수 공식

$$(\text{위험도 지수}) = (\text{기업 지수}) + (\text{관세 지수}) - (\text{경제 지수})$$

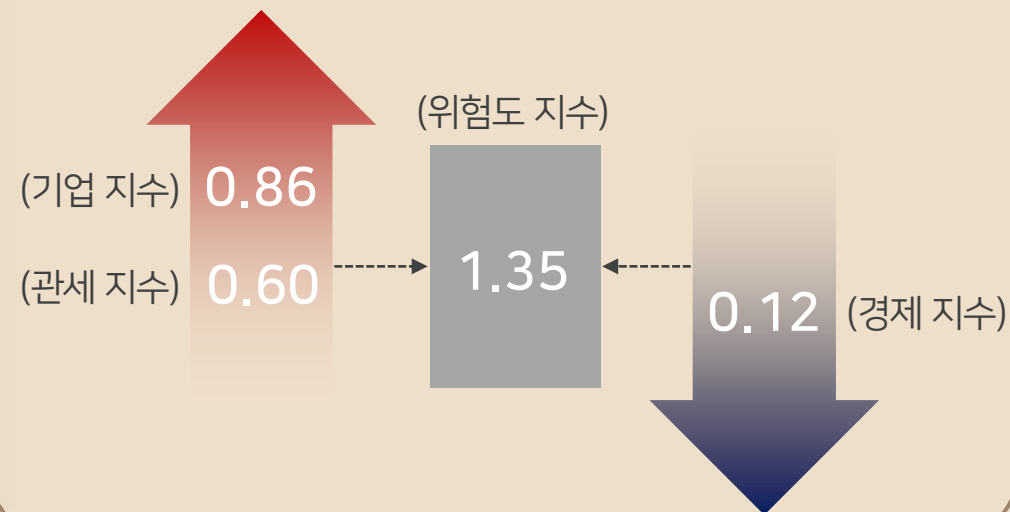
NAICS code	기업 지수	관세 지수	경제 지수	위험도 지수
4233 Lumber and Other Construction Materials Merchant Wholesalers	0.485072838	0	0.302644341	0.182428497
4236 Electrical and Electronic Goods Merchant Wholesalers	0.678342858	0	0.481577787	0.196765071
...
453220 Gift, Novelty, and Souvenir Stores	0.584795403	1	0.310099758	1.274695645

1주차 REVIEW

⊙ 위험도 지수 타당성 입증

사례 1. 식료품 도매업 (NAICS 코드: 4244)

NAICS code	기업 지수	관세 지수	경제 지수	위험도 지수
4244 Grocery and Related Product Merchant Wholesalers	0.86	0.60	0.12	1.35



$$\text{위험도 지수}(1.35) = \text{기업 지수}(0.86) + \text{관세 지수}(0.60) - \text{경제 지수}(0.12)$$

식료품 도매업(NAICS 4244)의 높은 위험도 지수는 1) 높은 기업 지수와 2) 낮은 경제 지수에 기인함

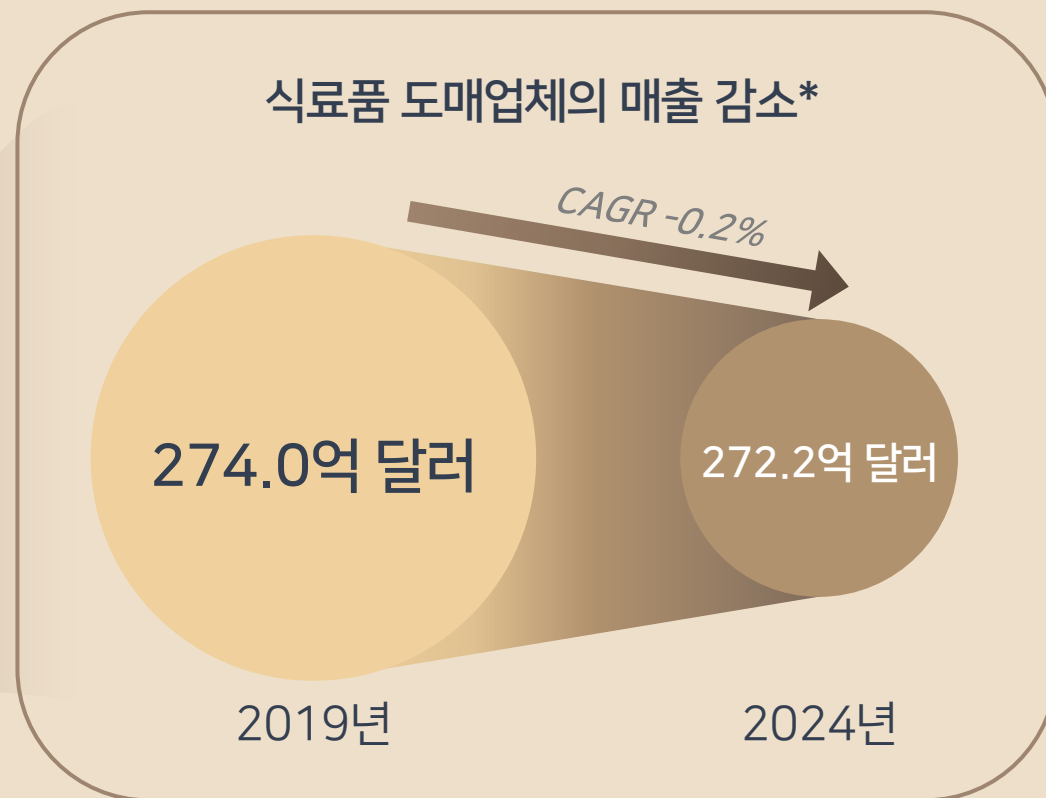
1주차 REVIEW

⊙ 위험도 지수 타당성 입증

사례 1. 식료품 도매업 (NAICS 코드: 4244)

NAICS code	기업 지수	관세 지수	경제 지수	위험도 지수
4244 Grocery and Related Product Merchant Wholesalers	0.86	0.60	0.12	1.35

평균 = 0.3432, 표준편차 = 0.2063



식료품 도매업(NAICS 4244)의 경제지수는 비교적 낮으며,
실제로 식료품 도매업 매출은 CAGR -0.2%로 수축해 옴

1주차 REVIEW

○ 위험도 지수 타당성 입증

사례 1. 식품품 도매업 (NAICS 코드: 4244)

NAICS code	기업 지수	관세 지수	경제 지수	위험도 지수
4244 Grocery and Related Product Merchant Wholesalers	0.86	0.60	0.12	1.35

평균 = 0.49794, 표준편차 = 0.2731

현지 물류기반 없어, 수출에 실패한 사례*

"한 국내 중소기업이 미국 대형 유통업체와 수출
계약 직전까지 갔으나, 현지 물류 창고 요구로
원가 부담이 커져 계약이 성사되지 않았다.
미국 바이어들은 물류 창고 보유를 선호하며,
이는 중소기업에 큰 부담을 주고 있다."



KOTRA
대한무역투자진흥공사

미국 식품품 도매업 시장에서 신규 진입 한국 중소기업은
현지 물류 인프라 부족으로 대형 유통업체와의 경쟁에서 불리한 위치임

식품품 도매업(NAICS 4244)의 기업지수는 비교적 높으며,
실제로 현지 경쟁력 부족으로 미국 진출 포기한 사례 존재함

1주차 REVIEW

◎ 위험도 지수 타당성 입증

사례 2. 의류·원단 도매업 (NAICS 코드: 4243)

NAICS code	기업 지수	관세 지수	경제 지수	위험도 지수
4243 Apparel, Piece Goods, and Notions Merchant Wholesalers	0.56	1	0.28	1.34

평균 = 0.41071, 표준편차 = 0.4585

섬유업계, 한미 FTA 혜택 미비*

“한미 FTA의 원사기준으로 국내 섬유업체들이
실 생산 기반 부족 등으로 관세 혜택을 받지
못하고, 미국 시장 진출에 어려움을 겪고 있다.
이에 원사기준 완화에 대한 정부 차원의 대책이
필요하다는 목소리가 커지고 있다.”

kotra

KOTRA
대한무역투자진흥공사

미국 의류·원단 도매업 시장에서 한국 중소기업은 관세 규제로
중국 대비 가격 경쟁력 확보하지 못해, 미국 진출 포기하는 경우가 많음

의류·원단 도매업(NAICS 4243)의 관세지수는 비교적 높으며,
실제로 관세 혜택을 받지 못해 가격 경쟁력을 확보하지 못함

Trend 자연어처리

[Step 2]

+ 업종별 수요 예측
실시간 트렌드 분석
↳ Amazon 판매량/리뷰 감정 분석, 기후(계절 특성) 분석 등



[Step 3]

현재 시점으로부터 1년 이내
미국 시장 진출 최적 타이밍 제시

실시간 판매 정보 등을 활용해 세부적인 상품에 대한 소비자들의 수요와 관심도를 분석하고자 했으나
구체적이고 단기간인 관심도 정보는 기반이 마련되지 않은 중소기업의 수출로 이어지기에는 어렵다고 판단



NAICS 코드 기준 "업종" 에 대한 **장기적인 관심**을 분석대상으로 설정 !



Trend 자연어처리


[Step 2]

+ 업종별 수요 예측
실시간 트렌드 분석
↳ Amazon 판매량/리뷰 감정 분석, 기후(계절 특성) 분석 등



[Step 3]

현재 시점으로부터 1년 이내
미국 시장 진출 최적 타이밍 제시

(+) 그룹 최종 흐름 정리 

트렌드 지수

↓
Part 2에서 진행 예정



매출 및 수출 관련 지표

↓
Part 3에서 후술 예정

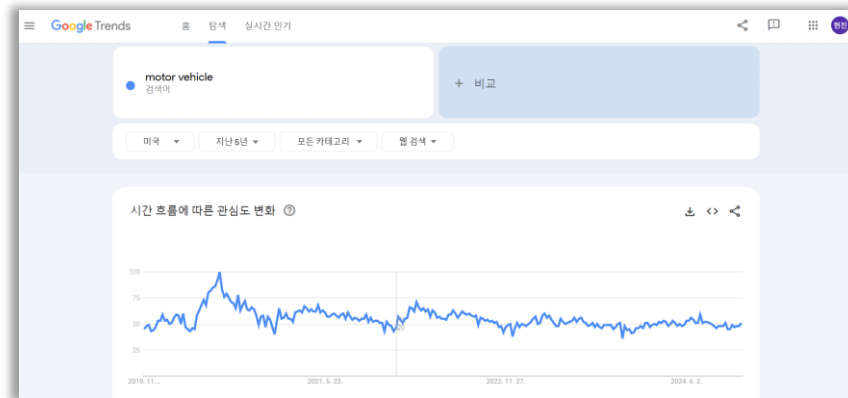


현재 시점으로부터 1년 이내
미국 시장 진출 최적 타이밍 제시 !

Trend 자연어처리

◎ 구글 검색량

Google Trends



하지만 특정 단어를 검색하면 **검색량 시계열 데이터 사용 가능**

Trend 자연어처리

◎ Trend 지수 생성 흐름



NAICS 코드를 검색하여
업종별로 관련도가 높은
뉴스 기사 크롤링

업종명으로 검색해서 관련도 순 상위 50개 기사 사용



$\sum_{\text{all 키워드}} (\text{유사도의 총합} \times \text{대표 키워드 검색량})$
trend 시계열 생성

구글 트렌드 활용 !



각각의 기사 내에서 유사도가 높은
trend 키워드 상위 10개 추출



업종별 50개의 기사에서의
상위 10개 키워드 유사도를
모두 sum !



Ex) 기사 1의 단어1 유사도 + 기사 2의 단어2 유사도 + ... = 유사도 총합

Trend 자연어처리

◎ 키워드 추출

KeyBERT

원본 문서를 가장 잘 나타내는 **중요한 용어 또는 구문을 찾아내는** 작업을 위한 모델로,
텍스트 임베딩을 형성하는 단계에서 BERT를 사용하는 BERT-based solution

----- KeyBERT 작동 과정 -----

1. 문서의 각 문장을 단어로 쪼갠 뒤 BERT를 통해 각 단어에 대한 임베딩 벡터를 추출
2. Phrase에 대해서도 문서 전체에 적용한 것과 같은 방식으로 처리
3. 코사인 유사도를 사용하여 가장 많이 사용되는 단어/문구 찾기
4. 전체 문서를 가장 잘 설명하는 단어/문구 추출

Trend 자연어처리

○ 키워드 추출

Ex) NAICS 코드 4231_Motor vehicle

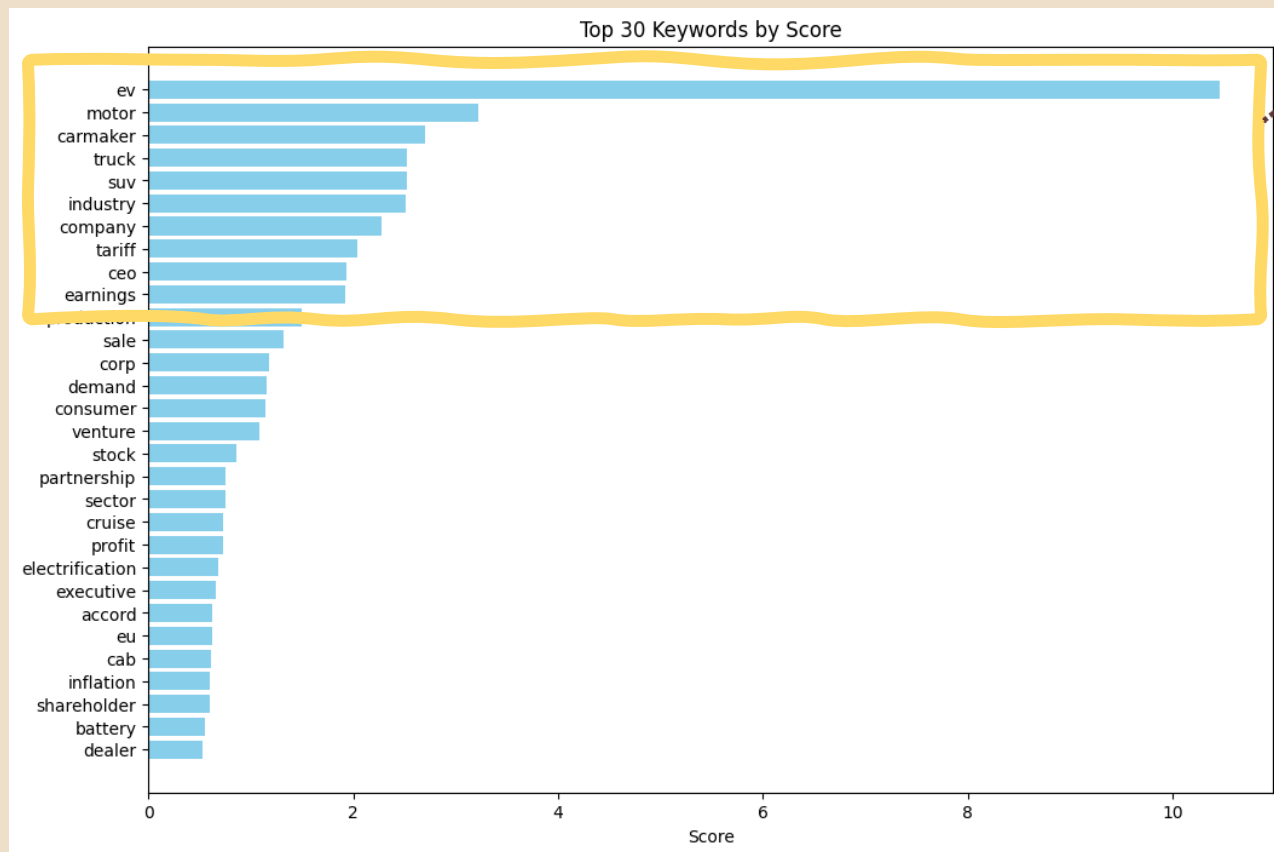
Title	Content	Extracted_Keywords_with_Scores
Bentley Motors further delays...	In this article Bentley ...	bentley:0.4894, vehicle:0.3540, ev:0.3414...
Ford's October vehicle sales...	In this article DETROIT —Ford..	labor:0.3665, ford:0.3581, strike:0.3365...
GM's newest EV is a Cadillac...	In this article DETROIT —General...	cadillac:0.5776, ev:0.4248, suv:0.4180...
Volkswagen's Scout Motors...	In this article NASHVILLE, Tenn...	volkswagen:0.4340, ford:0.3680, vehicle:0.3629...

기사 별로 유사도 점수 상위 10개 단어를 {keyword : score} 형태로 출력하여
'Extracted_Keywords_with_Scores' 열 생성

Trend 자연어처리

○ 대표 키워드 추출

Ex) NAICS 코드 4231_Motor vehicle



모든 열의 점수를 더한 뒤 상위 10개를
해당 업종의 대표 키워드로 선정!

Keyword	Score
ev	6.8907
tesla	6.0927
Ford	5.3904
gm	3.7509
toyota	2.7246
...	...



Trend 자연어처리

◎ Trend 시계열 생성



$\sum_{all \text{ 키워드}}$ (유사도의 총합 X 대표 키워드의 검색량)을 계산하여 trend 시계열 생성

Keyword	Score
ev	6.8907
tesla	6.0927
Ford	5.3904
gm	3.7509
toyota	2.7246
...	...

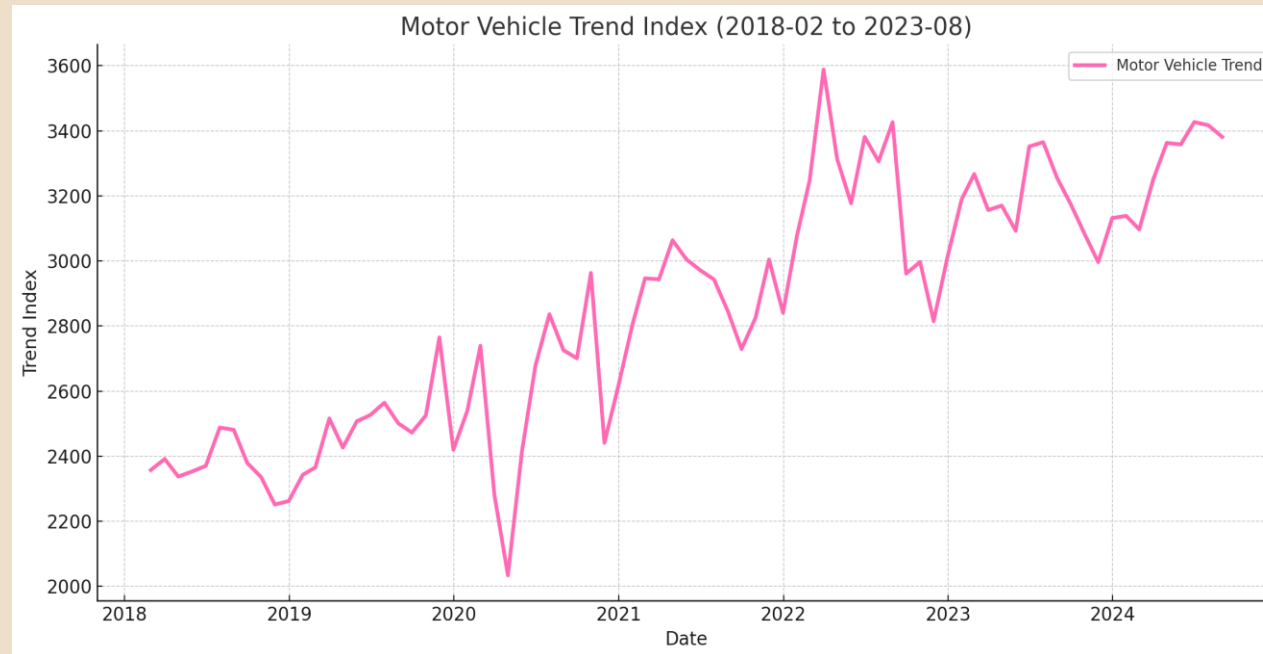


18-Feb	18-Mar	18-Apr	18-May	18-Jun	18-Jul	18-Aug
18	19	19	20	19	20	20
43	32	33	36	35	39	43
74	79	79	78	79	83	81
69	69	66	64	65	64	66
72	78	73	77	79	84	82
...

Trend 자연어처리

◎ 최종 Trend 지수

Ex) NAICS 코드 4231_Motor vehicle



업종별로 소비자들의 "관심도"를 반영한 trend 지수 생성 완료 !

매출액/수출액 예측

○ 예측과정

Date	Sales_4231	Sales_4232	Sales_423430	수출액_4231	수출액_4232	수출액_423430	기타 x 변수들
2018-02	35287	6862	17317	63503927	36731496	23351004	...
2018-03	40599	7592	22125	54541280	36483062	23416763	...
2018-04	38550	7556	20173	67699081	34131740	24044088	...
...
2024-08	54103	9438	25881	72755845	44111377	68979880	...
2024-09	예측이 필요한 구간						?
2024-10							

매출액과 수출액 모두 예측이 필요한 미래 시점에 대한 x변수들의 값이 존재하지 않음

매출액/수출액 예측

예측과정

Date	Sales_4231	Sales_4232	Sales_423430	수출액_4231	수출액_4232	수출액_423430	기타 x 변수들
2018-02	35287	6862	17317	63503927	36731496	23351004	...
2018-03	40599	7592	22125	54541280	36483062	23416763	...
2018-04	38550					24044088	...
...							...
2024-08	54103	9438	25881	72755845	44111377	68979880	...
2024-09							?
2024-10							?

미래 시점을 예측하기 위해 사용 가능한 방법



매출액과 수출액 데이터 만을 활용해 **단변량 예측 모델링**

..SARIMA, Prohpet 등 시계열 모델 혹은 N-BEATS 등 딥러닝 모델을 활용



X변수들에 대한 예측을 먼저 진행한 뒤 이를 이용한 모델링

X변수 데이터셋이 완성되면 트리모델 등 머신러닝 기법을 활용할 수 있음

매출액 / 수출액 모두 예측이 필요한 미래 시점에 대한 x변수들의 값이 존재하지 않음

매출액/수출액 예측

◎ 단변량 모델링



매출액과 수출액 데이터만을 활용해 단변량 예측 모델링
SARIMA, Prophet 등 시계열 모델 혹은 N-BEATS 등 딥러닝 모델을 활용

N-BEATS

N-BEATS는 딥러닝 기반의 시계열 예측 모델로, 블록 구조를 기반으로 설계
각 블록은 포워드 예측과 백워드 잔차를 생성하여
시계열 데이터의 패턴을 점진적으로 학습하고 잔여 오류를 줄이는 데 기여

매출액/수출액 예측

◎ 단변량 모델링



매출액과 수출액 데이터만을 활용해 단변량 예측 모델링
SARIMA, Prophet 등 시계열 모델 혹은 N-BEATS 등 딥러닝 모델을 활용



비교용으로 딥러닝 모델도 시도한 것 ...

하지만 매출액과 수출액의 데이터셋이 크지 않아
딥러닝보다는 최대한 정통 시계열 모델로 예측해 과적합을 방지하고자 했음
또한 모델의 성능 확인을 위해 train 값의 일부를 test set으로 활용

매출액/수출액 예측

◎ 예측 성능 평가지표

MAPE (Mean Absolute Percentage Error)

실제 값에 대한 예측 오차의 절댓값을 평균하여 백분율로 측정한 값



$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100$$

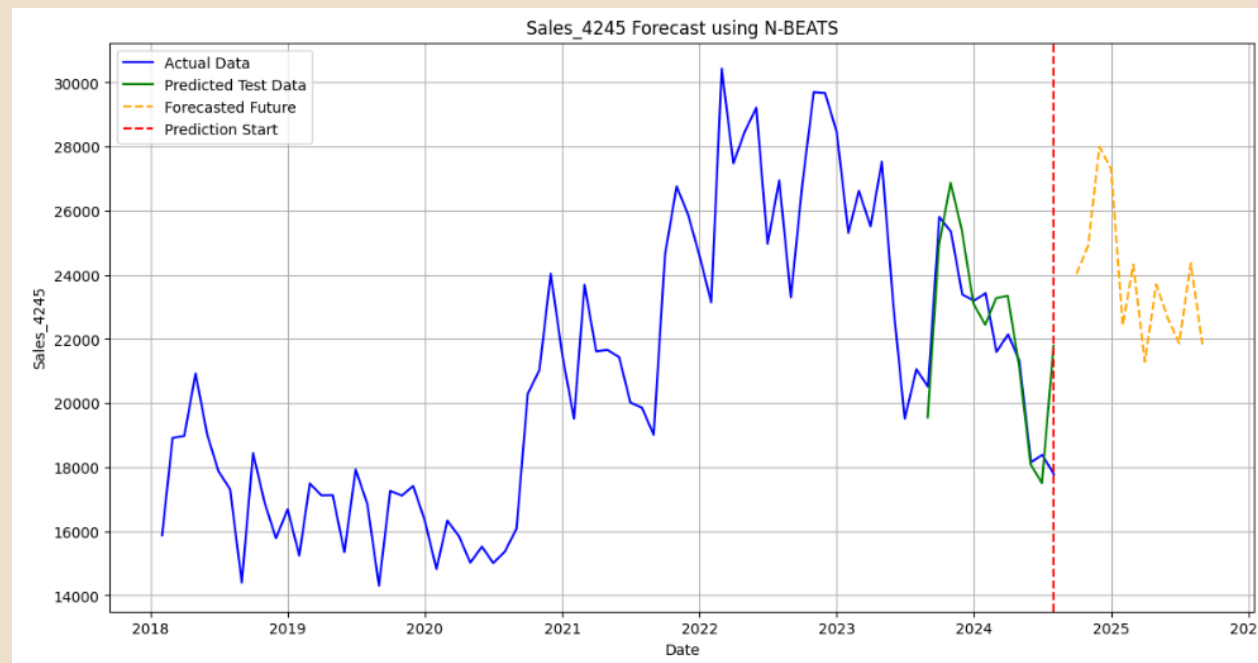
실제값과 예측값이 몇 퍼센트 차이나는지 직관적으로 해석하기 용이해 평가지표로 사용

매출액/수출액 예측

매출액 | 수출액

○ 매출액 단변량 예측

업종	Prophet	SARIMA	N-BEATS
4231	2.99%	2.49%	2.07%
4232	2.93%	3.61%	12.37%
423430	4.41%	5.55%	12.23%
4235	8.67%	5.32%	9.78%
...



Prophet, SARIMA, N-BEATS 중 **MAPE가 평균적으로 가장 낮은 모델**을 최종 모델로 선정

◎ 최종 데이터셋 구성



트렌드, 즉 사람들의 관심도를 반영하는 변수로 매출액 설명변수 후보에 추가함 !

◎ 그레인저 인과검정

그레인저 인과관계 (Granger Causality)

1969년 쿨라이브 그레인저가 제안한 개념으로 X의 과거 정보가 Y의 미래 값을 예측하는 데 통계적으로 유의미한 도움을 준다면 X는 Y를 그레인저 인과한다고 함



그레인저 인과 검정을 위해서는 데이터가 정상성을 가져야 함

◎ 그레인저 인과검정

대부분의 데이터가 정상성을 가지지 않음

↓ 로그 차분

로그 차분한 업종별 Y변수와 로그 차분한 X 변수들로 그레인저 인과 검정 진행

↓ 각 변수별 업종별로 lag를 구함

X 변수들의 시차를 조정

Ex. 환율의 4231 업종에 대한 lag가 3이면, 이에 대한 환율 데이터(환율_4231)를 3개월 뒤로 조정



◎ 그레인저 인과검정

각 변수별 업종별로 lag 만큼 데이터를 조정한 최종 형태
X 변수들의 데이터는 18년 이전 데이터가 존재해 lag 만큼 이전 데이터를 추가함

	환율_4231	...	GDP기여도_4231	...	인구_4235	...	인구_454310
2018-02	1087.13		-0.23		327528		328085
2018-03	1066.37		-0.1333		327699		328219
2018-04	1069.08		-0.3667		327848		328364
...							...
2024-08	1381.68	...	0.23	...	336550	...	337362

◎ 설명 변수 선택

Lasso 회귀

일반 회귀식의 MSE에 라쏘회귀는 L1-norm이라는 패널티를 부과하여 회귀계수를 추정하는 방법으로, 덜 중요한 특성의 가중치를 0으로 보내는 데 활용

업종마다 서로 다른 변수로 모델링 하는 것은 무리라고 판단 !

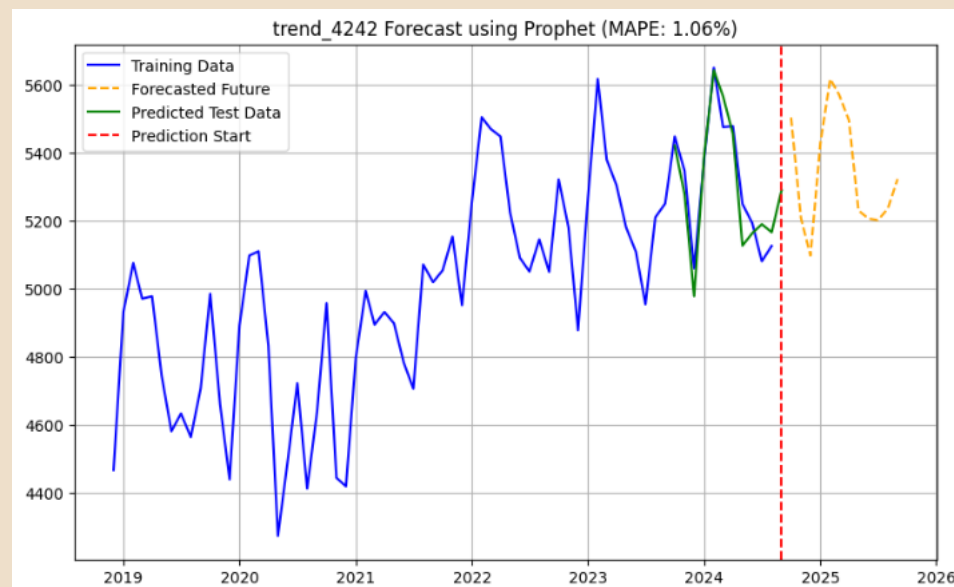


업종 별로 선택된 변수를 **hard voting** 하여 최종 설명 변수를 선택하기로 함

설명 변수 단변량 예측

Ex) Trend 지수 예측 성능 비교

업종	Prophet	SARIMA	N-BEATS
4231	2.06%	5.95%	2.19%
4413	1.54%	3.30%	6.69%
423430	2.01%	3.83%	4.67%
4242	1.06%	3.60%	4.10%
...



종속변수와 동일하게 SARIMA, Prohpet, N-BEATS 활용해 단변량 예측 진행

Trend의 경우, 가장 성능이 좋은 Prophet으로 예측 진행

설명 변수 단변량 예측

date	Sales_452990	Sales_453310	Sales_454110	...	PCE	temperature_LA	unemployment_rate	flag_holiday_us	trend_4231	...
2018-02-01	5.662	1.372	40.785	...	13738.7	14.7	4.1	1	2357.7089	...
2018-03-01	6.509	1.553	46.395	...	13793.2	14.8	4.0	0	2391.6099	...
2018-04-01	6.033	1.446	45.191	...	13838.4	16.7	4.0	0	2337.5566	...
...
2025-01-01	예측구간			...	20206.7852	14.534476	4.0718	1	3420.01812	...
2025-02-01				...	20398.0027	13.053089	4.005674	2	3600.23968	...
2025-03-01				...	20476.8445	12.822212	4.215965	1	3412.43636	...
2025-04-01				...	20530.8719	14.317209	4.108794	0	3356.712	...
2025-05-01				...	20622.4122	15.585768	4.111073	0	3413.84272	...
2025-06-01				...	20686.2299	16.014367	4.310308	1	3533.5737	...
2025-07-01				...	20744.7394	17.955365	4.310017	1	3547.19066	...
2025-08-01				...	20856.3867	20.038024	4.354249	1	3481.8897	...
2025-01-01				...	20914.8889	21.170318	4.460864	1	3325.85401	...

단변량 예측이 완료된 설명 변수 데이터를 활용해 종속변수 다변량 예측 진행

◎ 다변량 예측 (예측된 설명 변수 활용)

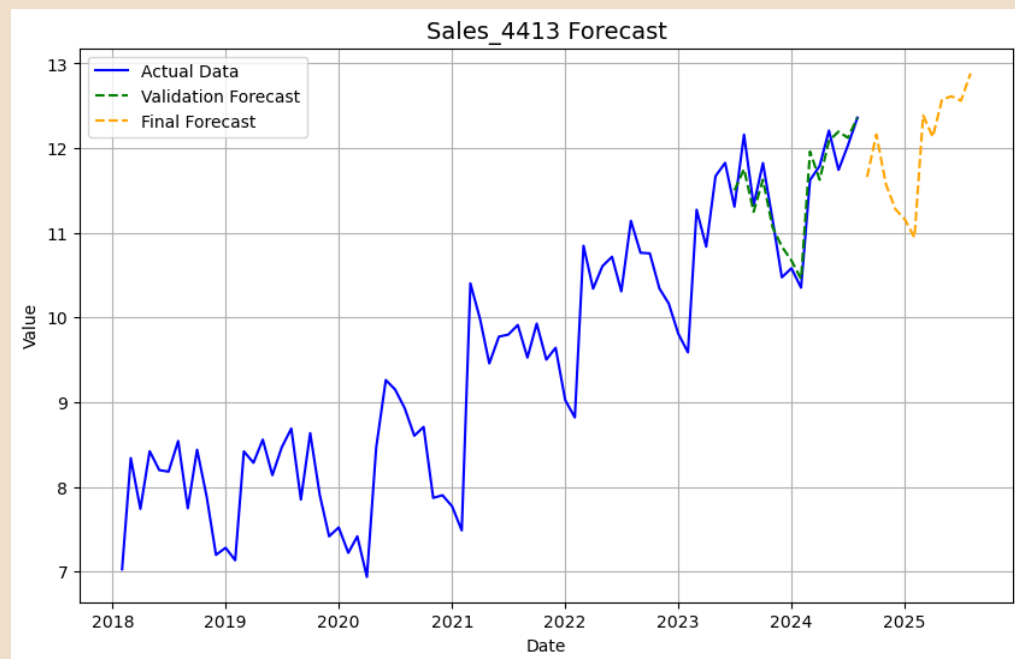
Model	Random Forest	선형회귀	다항회귀	SVR
Sales_448140	9.39%	13.12%	17.12%	8.80%
Sales_448190	9.51%	9.05%	29.92%	9.54%
Sales_451110	15.89%	23.05%	16.00%	17.60%
Sales_452112	7.32%	8.09%	17.96%	5.74%
Sales_452990	5.44%	5.51%	9.05%	6.00%
...

다양한 모델에 적합해본 결과 성능에 큰 차이가 관찰되지 않음

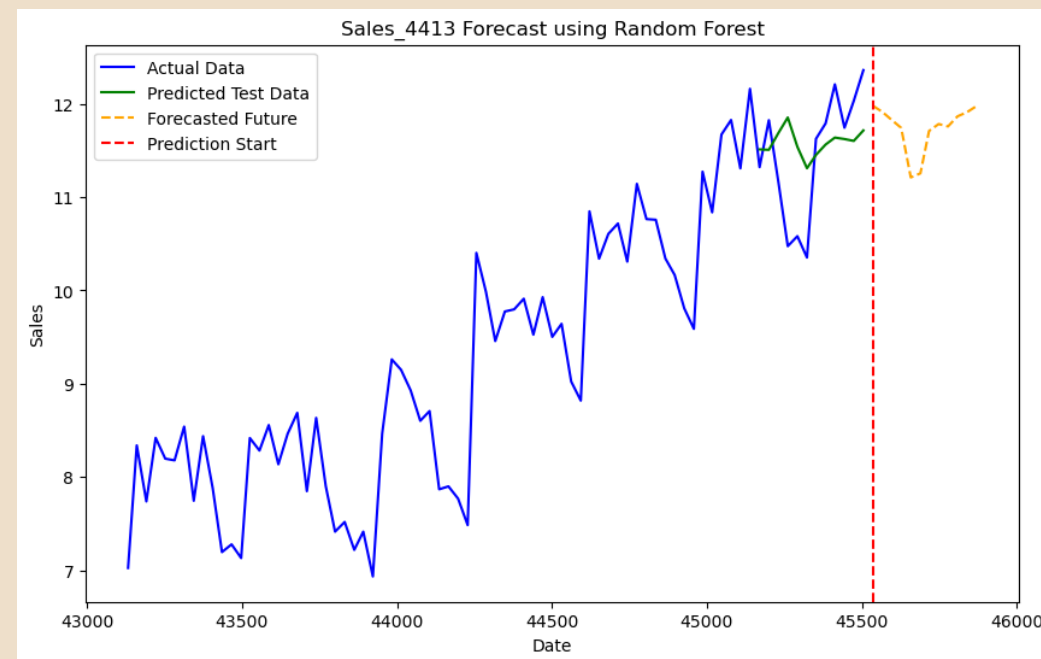
여러 업종에 대해 안정적으로 좋은 결과를 내는 **Random Forest**를 최적의 모델로 선정

○ 단변량 / 다변량 예측 결과 비교

Ex) NAICS 코드 4413



단변량 예측 결과 (SARIMA)_MAPE: 1.67%



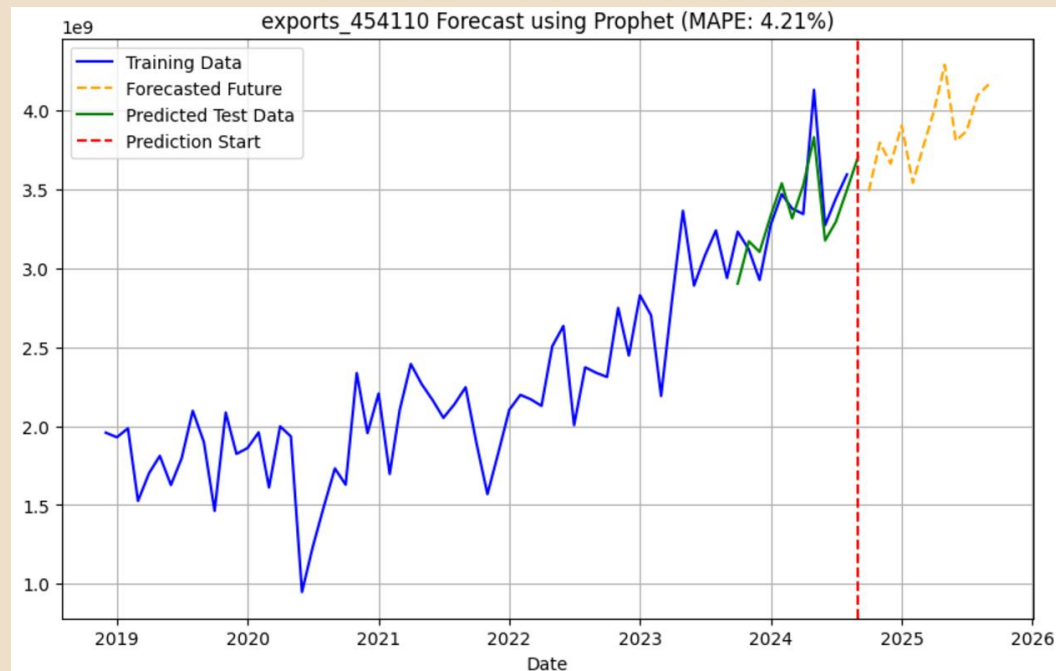
다변량 예측 결과 (랜덤포레스트)_MAPE: 4.88%

매출액/수출액 예측

매출액 | 수출액

○ 수출액 단변량 예측

업종	Prophet	SARIMA	N-BEATS
4235	6.04%	10.79%	10.53%
4241	11.10%	21.38%	9.52%
4245	21.83%	13.31%	11.12%
454110	4.21%	11.99%	8.53%
...



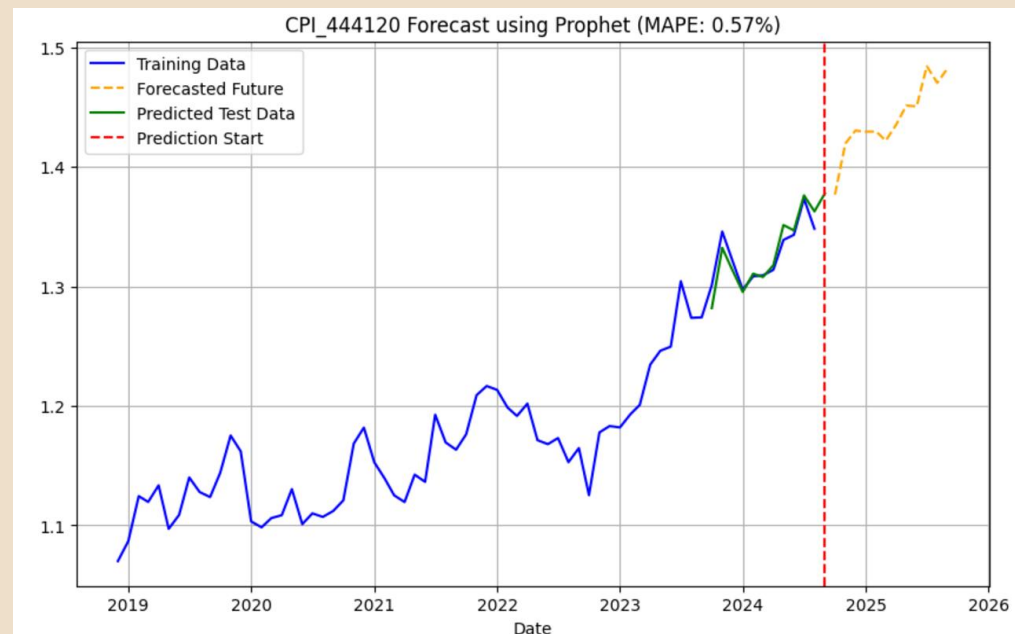
매출액 예측과 동일하게 Prophet, SARIMA, N-BEATS 중

MAPE가 평균적으로 가장 낮은 Prophet을 최종 모델로 선정

○ 설명 변수 단변량 예측

Ex) CPI 지수 예측 성능 비교

업종	Prophet	SARIMA	N-BEATS
4231	0.48%	1.19%	%
4413	0.54%	1.43%	%
423430	0.92%	0.87%	%
4242	0.53%	1.04%	%
...



종속변수와 동일하게 SARIMA, Prohpet, N-BEATS 활용해 단변량 예측 진행
모든 변수에 대해 가장 성능이 좋았던 Prohpet으로 예측 진행

예측 결과에 대한 MAPE는 모두 10% 미만!

매출액/수출액 예측

매출액 | 수출액

설명 변수 단변량 예측

date	exports_4231	exports_4232	exports_423430	...	exchange	pop	CPI_4231	...	GDP_4231	...
2018-02-01	63503926.72	36731496.47	23351003.59	...	1084.21	328085	1	...	-0.23	...
2018-03-01	54541280.06	36483061.82	23416763.46	...	1060.25	328219	1.0089	...	-0.1333	...
2018-04-01	67699080.98	34131740	24044088.01	...	1069.33	328364	1.0142	...	-0.0367	...
...
2025-01-01	예측구간			...	1349.90	337929.74	0.9210	...	0.4879	...
2025-02-01				...	1349.73	338035.03	0.9197	...	0.1783	...
2025-03-01				...	1387.43	338169.98	0.9272	...	-0.0879	...
2025-04-01				...	1405.57	338311.70	0.9295	...	-0.0097	...
2025-05-01				...	1404.17	338438.78	0.9302	...	0.0314	...
2025-06-01				...	1404.31	338591.20	0.9308	...	0.0639	...
2025-07-01				...	1407.51	338760.82	0.9364	...	0.0883	...
2025-08-01				...	1413.41	338938.71	0.9389	...	0.0363	...
2025-09-01				...	1447.41	339116.41	0.9417	...	-0.1241	...

매출액과 동일하게 단변량 예측이 완료된 설명 변수 데이터를 활용해 종속변수 다변량 예측 진행

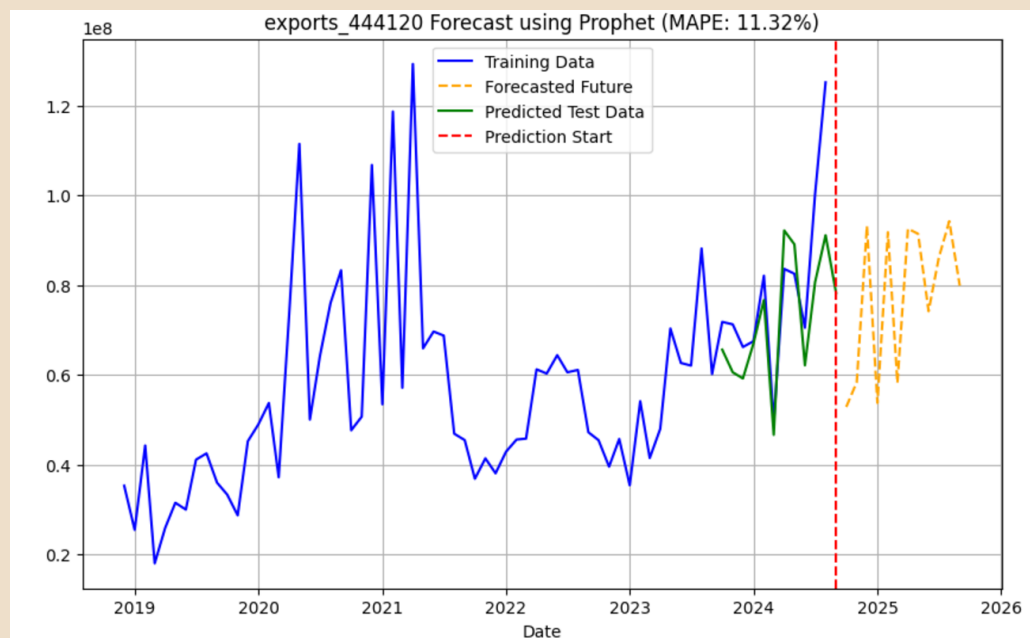
◎ 다변량 예측 (예측된 설명 변수 활용)

Model	Random Forest	선형회귀	다항회귀	XGBoost
Exports_423430	17.89%	56.06%	67.20%	28.97%
Exports_4235	12.04%	15.12%	48.14%	11.23%
Exports_4241	16.33%	10.71%	32.01%	16.94%
Exports_4242	18.35%	46.64%	106.42%	20.78%
Exports_452990	10.53%	99.02%	75.90%	16.33%
...

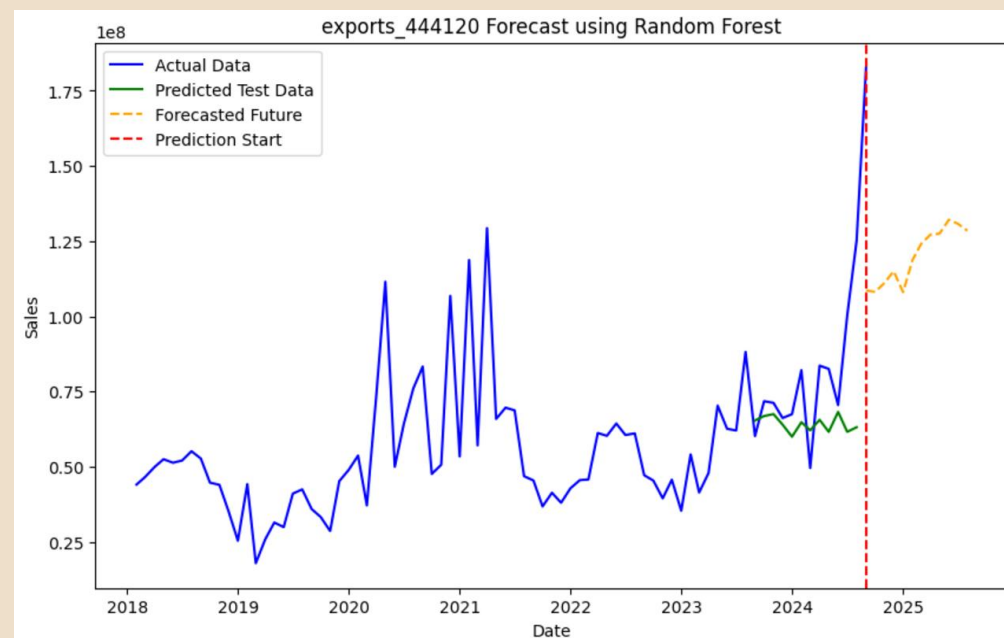
다양한 모델에 적합해본 결과 Random Forest와 XGBoost의 성능이 비교적 좋았음
그 중에서도 성능이 조금 더 좋았던 **Random Forest**를 최적의 모델로 선정

단변량 / 다변량 예측 결과 비교

Ex) NAICS 코드 444120



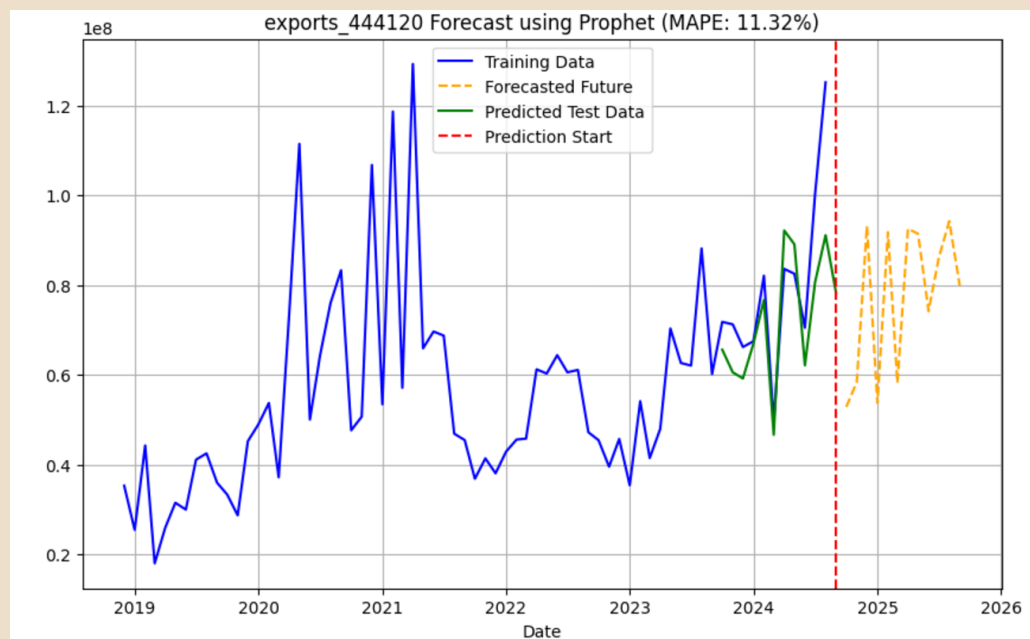
단변량 예측 결과 (Prophet)_MAPE: 11.32%



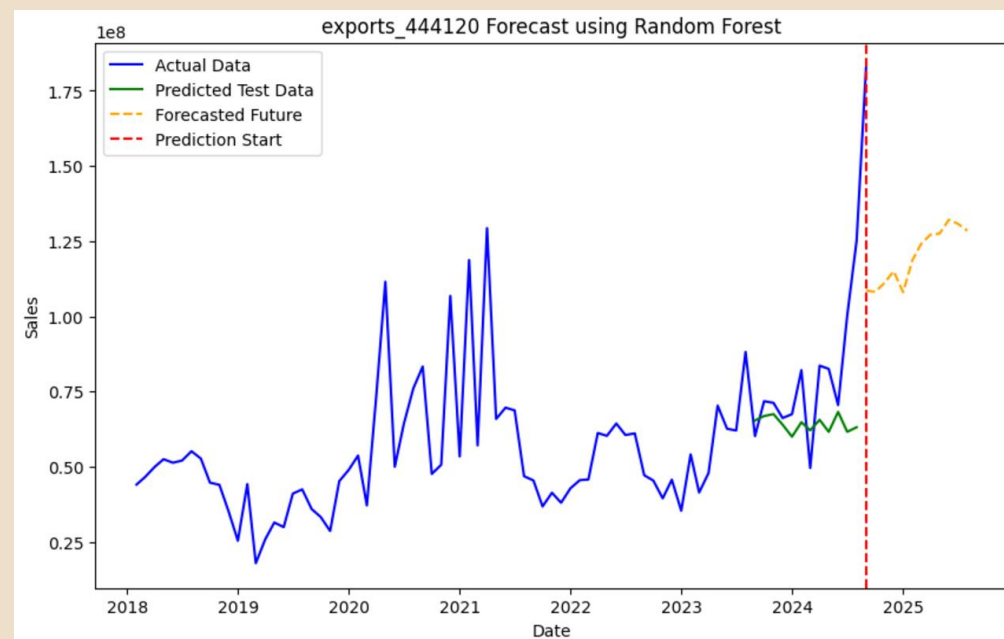
다변량 예측 결과 (랜덤포레스트)_MAPE: 23.90%

단변량 / 다변량 예측 결과 비교

Ex) NAICS 코드 444120



단변량 예측 결과 (Prophet)_MAPE: **11.32%**



다변량 예측 결과 (랜덤포레스트)_MAPE: 23.90%

미국 시장 진출 타이밍 제시

◎ 진출시기 예측 지표

예측에 활용되는 지표 구성

트렌드 지수 + 매출액 + 수출액 + (수출액/매출액)



트렌드 지수는 해당 업종에 대한 미국 소비자들의 관심도를 보여줌
트렌드 지수와 매출액의 크기가 크게 차이날 때는 시장의 성장 / 축소 가능성을 내포함

미국 시장 진출 타이밍 제시

◎ 진출시기 예측 지표

예측에 활용되는 지표 구성

트렌드 지수 + 매출액 + 수출액 + (수출액/매출액)



미국의 업종별 매출액은 미국 내 특정 업종의 시장 규모와 소비를 보여줌
즉, 미국 내에서의 해당 업종의 시장 성장성을 평가할 수 있음

미국 시장 진출 타이밍 제시

◎ 진출시기 예측 지표

예측에 활용되는 지표 구성

트렌드 지수 + 매출액 + 수출액 + (수출액/매출액)



업종별 대미 수출액은 해당 업종에서 한국 제품에 대한 미국 내 수요를 알 수 있음

미국 시장 진출 타이밍 제시

◎ 진출시기 예측 지표

예측에 활용되는 지표 구성

트렌드 지수 + 매출액 + 수출액 + (수출액/매출액)

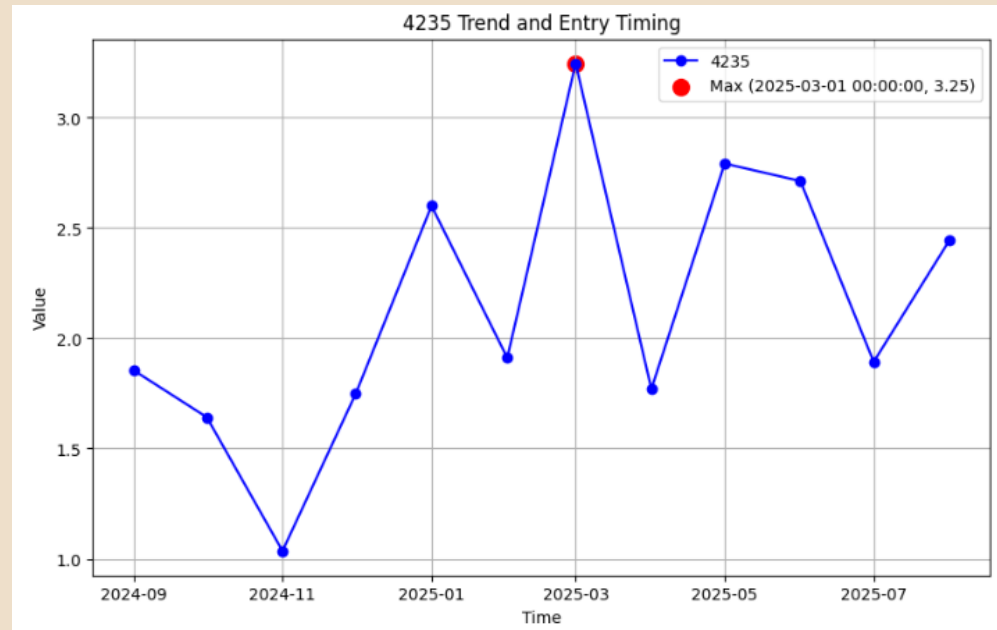


업종별 수출액/매출액은 미국의 해당 업종에서 한국 기업이 차지하는 비중을 알 수 있음

미국 시장 진출 타이밍 제시

○ 진출시기 제시

Ex) NAICS 코드 4235

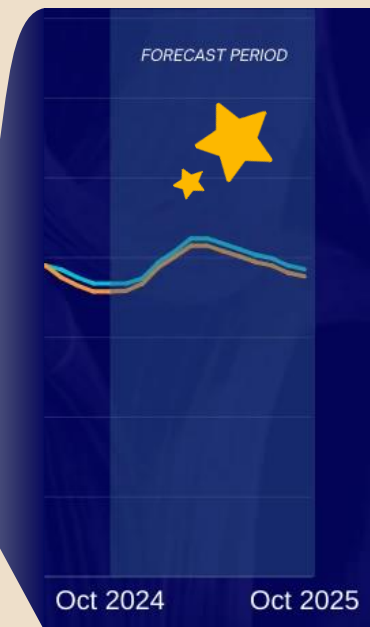


4235 업종의 경우 2025년 3월이 향후 1년 중 가장 진입하기 적절한 시기로 판단됨

미국 시장 진출 타이밍 제시

○ 진출시기 타당성 입증

4235(철강 제품) 시장 전망



인프라 투자 및 금리 인하, 철강 가격 상승의 전조*

"2024년 하반기 가격 정체와 수요 부진을 겪고 있으나, 2025년 4월 가격이 정점에 이를 것이다. 미국 정부의 대규모 인프라 투자 프로그램과 금리 인하가 맞물리며 철강 제품에 대한 수요 증가를 이끌 것으로 보인다."



MEPS
철강시장 전문
분석기업

2025년 3월 철강 제품 가격은 상승세로 전환될 가능성이 높으며, 미국 내 건설 및 인프라 프로젝트가 증가함에 따라 철강 도매업체들은 안정적인 수요 증가와 가격 상승 혜택을 누릴 전망이다

실제로 해당 품목의 가격 및 수요 전망을 고려할 때,
최종 진출 시기로 예측된 2025년 3월은 매우 적절한 시기인 것을 확인 가능