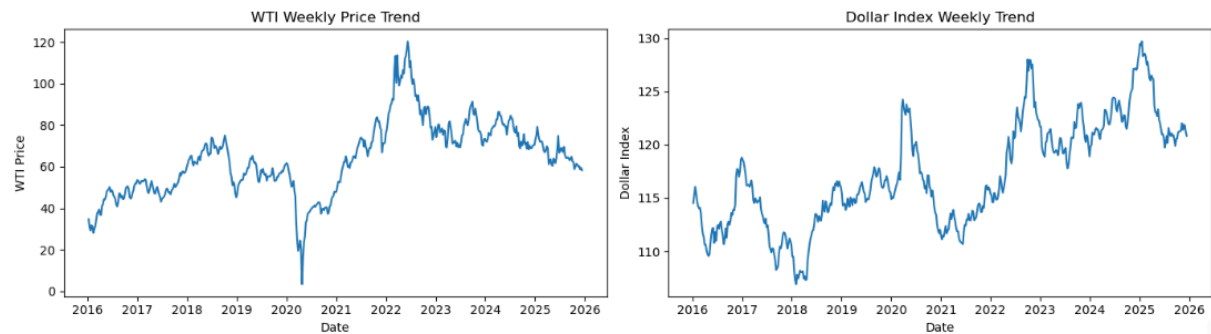


# WTI 원유 가격 리스크 모니터링 및 예측 시스템

본 프로젝트는 WTI 주간 유가와 환율, 금리, 재고, 달러지수 등 외생 변수를 결합해 가격 자체의 예측을 넘어 '변동성 확대 구간'을 사전에 경고하는 모델 구축을 목표로합니다.

## 1. EDA 분석 정리



### WTI 주간 가격 추세 (WTI Weekly Price Trend)

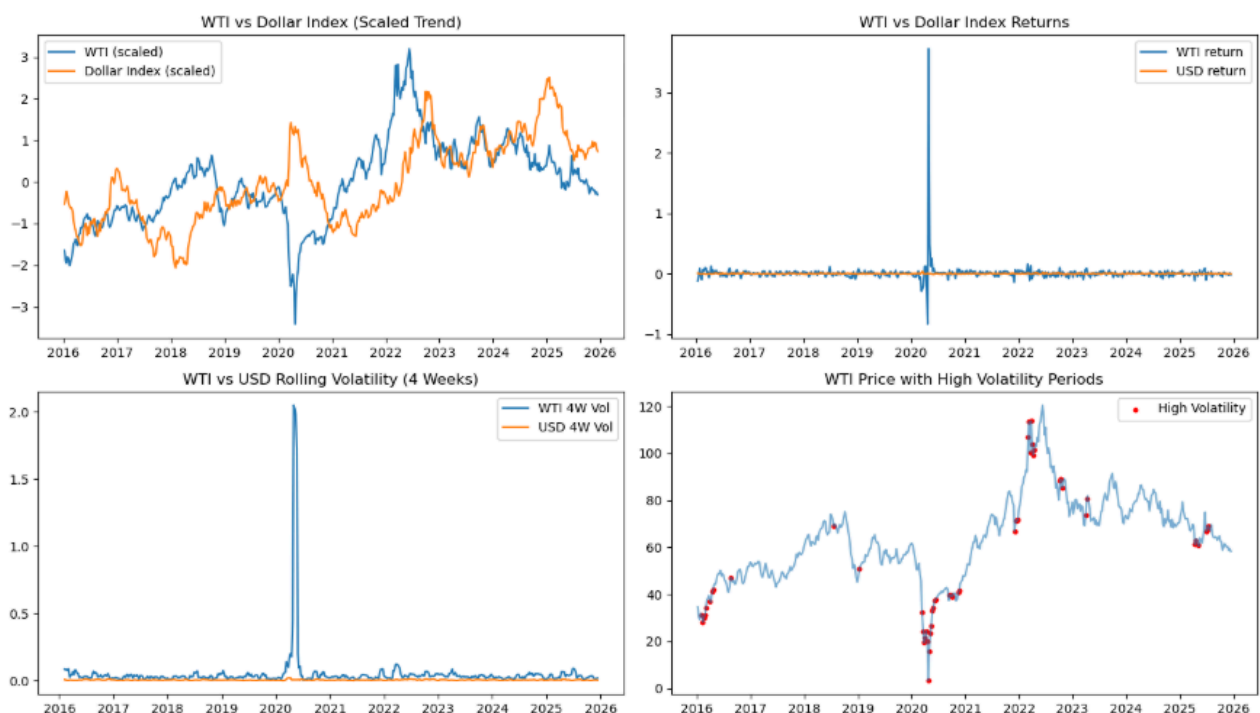
WTI 주간 가격은 장기적인 상승과 하락 국면이 반복되지만 일정한 주기성이나 계절적 패턴은 뚜렷하게 관찰되지 않는다.

특정 시점에서는 외생 이벤트로 인한 급격한 가격 변동이 발생하며 이러한 비정기적 충격이 전체 가격 흐름에 큰 영향을 미치는 구조를 보인다.

### 달러지수 주간 추세 (Dollar Index Weekly Trend)

달러지수는 급격한 상승 이후 비교적 완만한 조정 국면이 이어지는 패턴이 반복적으로 나타난다.

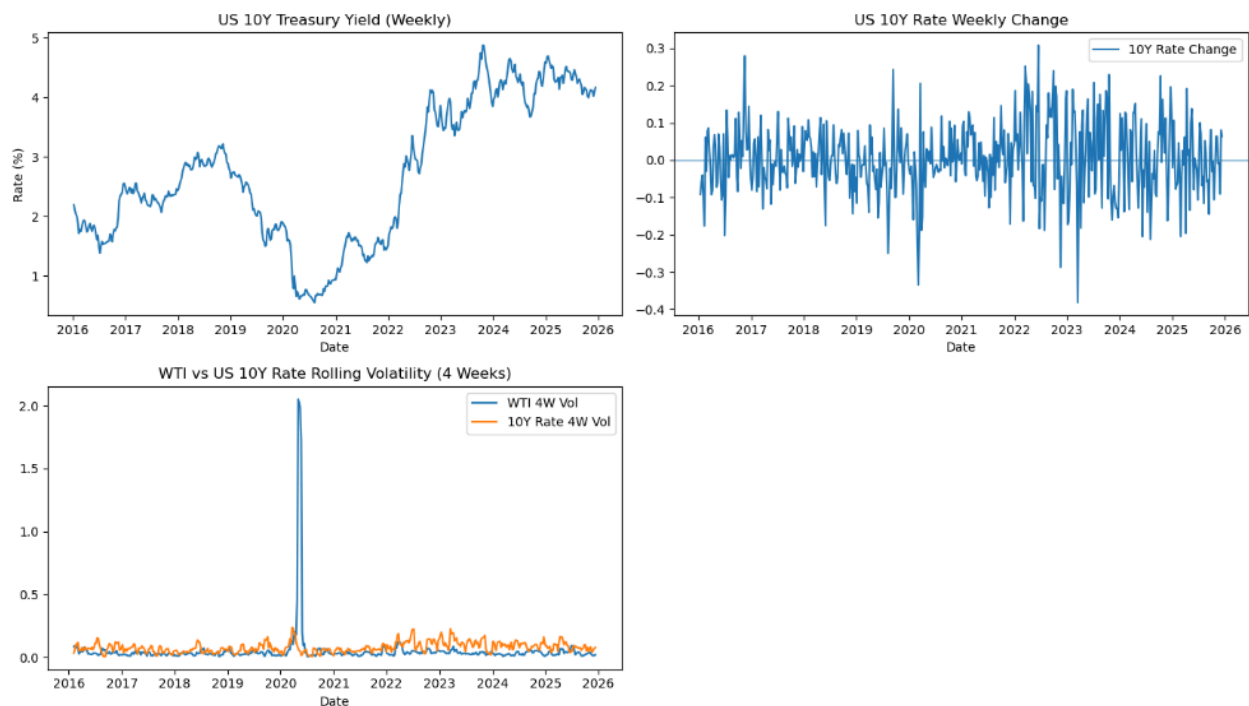
이는 통화 정책과 같은 거시 금융 요인의 영향을 반영한 결과로 원자재 가격과의 관계를 분석하기 위한 대표적인 외생 변수로 활용할 수 있다.



## WTI 가격 VS 달러지수

WTI 가격과 달러지수의 관계를 다양한 관점에서 분석한 결과, 두 지표는 장기적으로 일관된 동조 관계를 보이기는 하지만 시기별로 상이한 반응을 나타내는 구조를 가진 것으로 확인되었다. 달러지수의 절대적인 상승이나 하락이 WTI 가격 변동을 직접적으로 설명하지는 않았으며 변화율과 변동성 관점에서도 두 지표 간 반응은 비대칭적으로 나타났다. 특히 WTI 수익률과 변동성은 평상시에는 안정적인 수준을 유지하다가 특정 시점에서 급격하게 확대되는 스파이크 형태를 보였으며 이는 원유 시장이 점진적인 변동보다 외생 이벤트에 의해 단기간에 충격을 받는 특성을 가진다는 점을 보여준다.

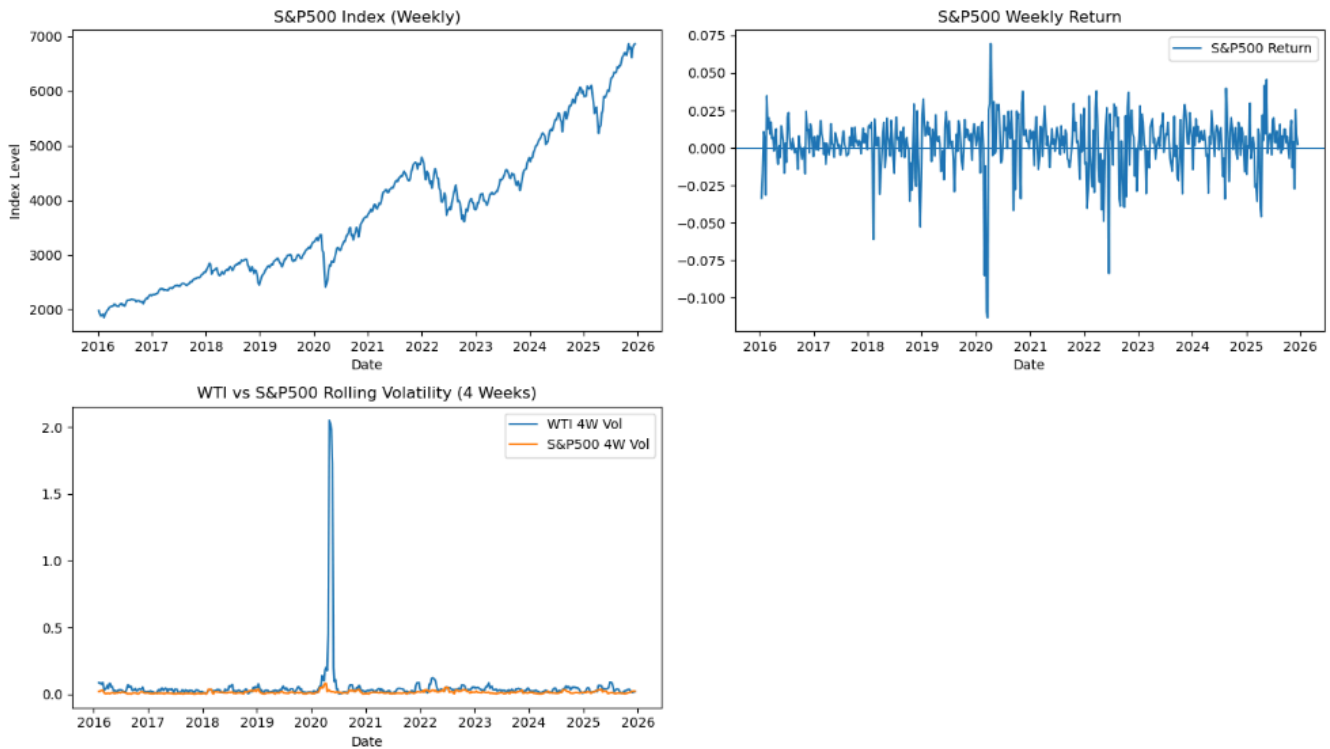
고변동성 구간을 중심으로 추가 분석한 결과, 해당 구간에서는 달러지수의 평균 수준 변화보다 변동성 자체가 소폭 확대되는 경향이 관찰되었으며 이는 원유 시장의 위기 국면이 글로벌 금융 시장 전반의 불확실성 확대와 일부 동조하는 환경에서 발생함을 시사한다. 다만 시차 상관 분석 결과 달러지수는 WTI 변동성을 선행적으로 설명하는 핵심 변수라기보다는, 원유 시장 리스크가 확대되는 거시적 환경을 보조적으로 설명하는 외생 변수로 해석하는 것이 타당하다. 이러한 분석을 통해 WTI 가격 리스크는 평균적인 가격 예측보다 변동성 급증 구간을 사전에 식별하는 접근이 더 적합하며 본 프로젝트에서는 이를 기반으로 고변동성 구간을 분류하는 경고형 모델링 전략을 채택하였다.



## WTI 가격 VS 미국 10년 국채 금리

금리의 절대적인 수준 자체는 유가 변동성 확대를 직접적으로 설명하지는 못하는 반면 금리 변화가 불안정해지는 국면에서는 유가 변동성 역시 확대되는 경향이 관찰되었다. 특히 2020년 이후 저금리 국면에서 고금리 국면으로 전환되는 과정에서 금리 변화량과 변동성이 커진 시점에 WTI 변동성도 일부 동조하는 모습을 보였으나 이러한 관계는 전 기간에 걸쳐 일관되게 유지되지는 않았다.

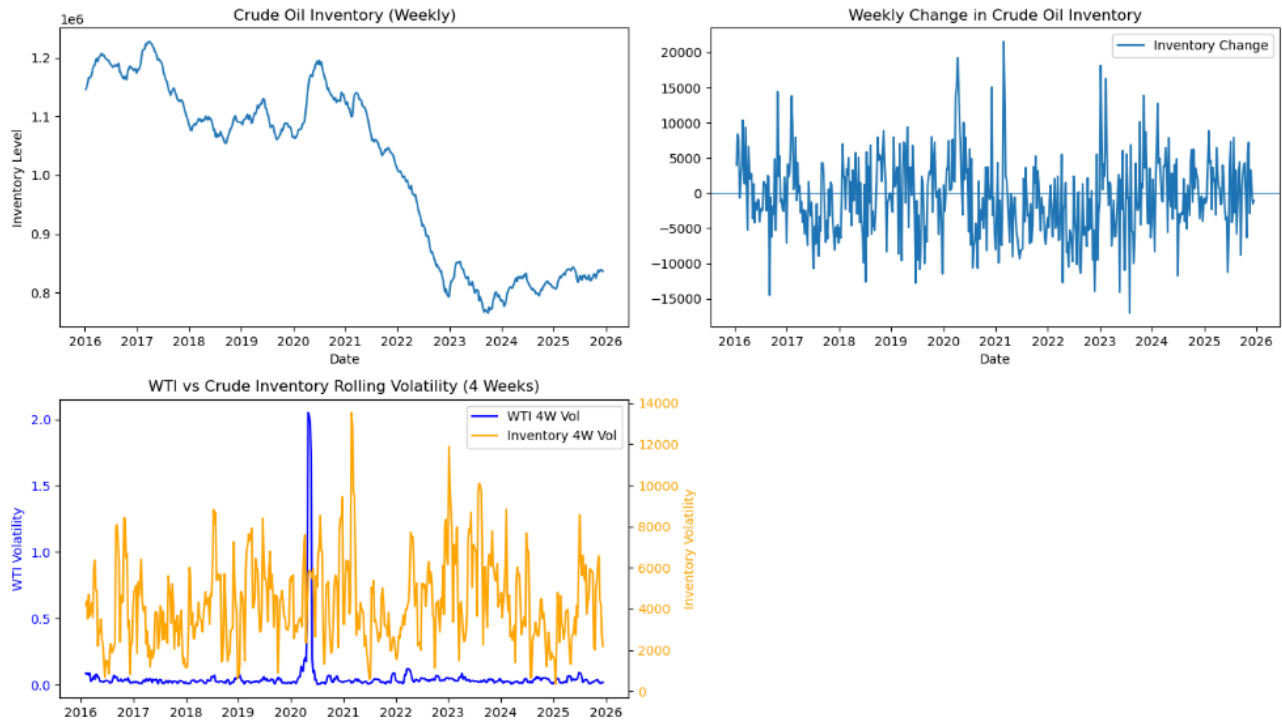
이는 유가 리스크가 특정 금리 수준에 의해 결정되기보다는 금융 환경 전반의 불확실성과 유동성 변화가 커지는 국면에서 함께 확대되는 성격을 가진다는 점을 시사한다. 따라서 미국 10년물 금리는 유가 변동성의 직접적인 원인 변수라기보다는 원유 시장 리스크가 확대되는 거시적 환경을 보조적으로 설명하는 외생 변수로 해석하는 것이 적절하다.



## WTI 가격 VS S&P 500

S&P500은 분석 기간 전반에 걸쳐 장기적인 우상향 추세를 유지하며 위험자산 시장의 구조적 성장 흐름을 반영하는 반면, WTI 가격 변동성은 이러한 장기 추세와 직접적으로 연결되기보다는 개별 이벤트와 수급 충격에 의해 좌우되는 성격이 강한 것으로 나타났다. S&P500 주간 수익률과 변동성은 2020년과 같은 극단적 위기 국면에서는 급격히 확대되었으나 대부분의 기간에서는 WTI 변동성과 일관된 동조 관계를 보이지 않았다.

WTI 고변동성 구간을 기준으로 추가 분석한 결과 해당 구간에서 S&P500 지수의 절대적 레벨은 상대적으로 낮게 관측되었으나 이는 지수의 장기 추세가 반영된 결과로 해석에 주의가 필요하다. 반면 S&P500의 단기 변동성은 WTI 고변동성 구간에서 평균적으로 더 높게 나타났으며 특히 WTI 고변동성 구간 중 약 44%에서 주식시장 역시 고변동성 상태에 해당하였다. 이는 유가 변동성 확대가 항상 주식시장 불안정성과 동반되지는 않지만 상당한 비중의 구간에서는 금융시장 전반의 위험 확대 국면과 함께 나타나는 경향이 있음을 시사한다. 종합하면 S&P500은 WTI 변동성의 직접적인 설명 변수라기보다는 원유 시장 리스크가 금융시장 전반의 불확실성 확대 국면과 중첩되는지를 판단하는 보조적인 시장 심리 지표로 해석하는 것이 적절하다.

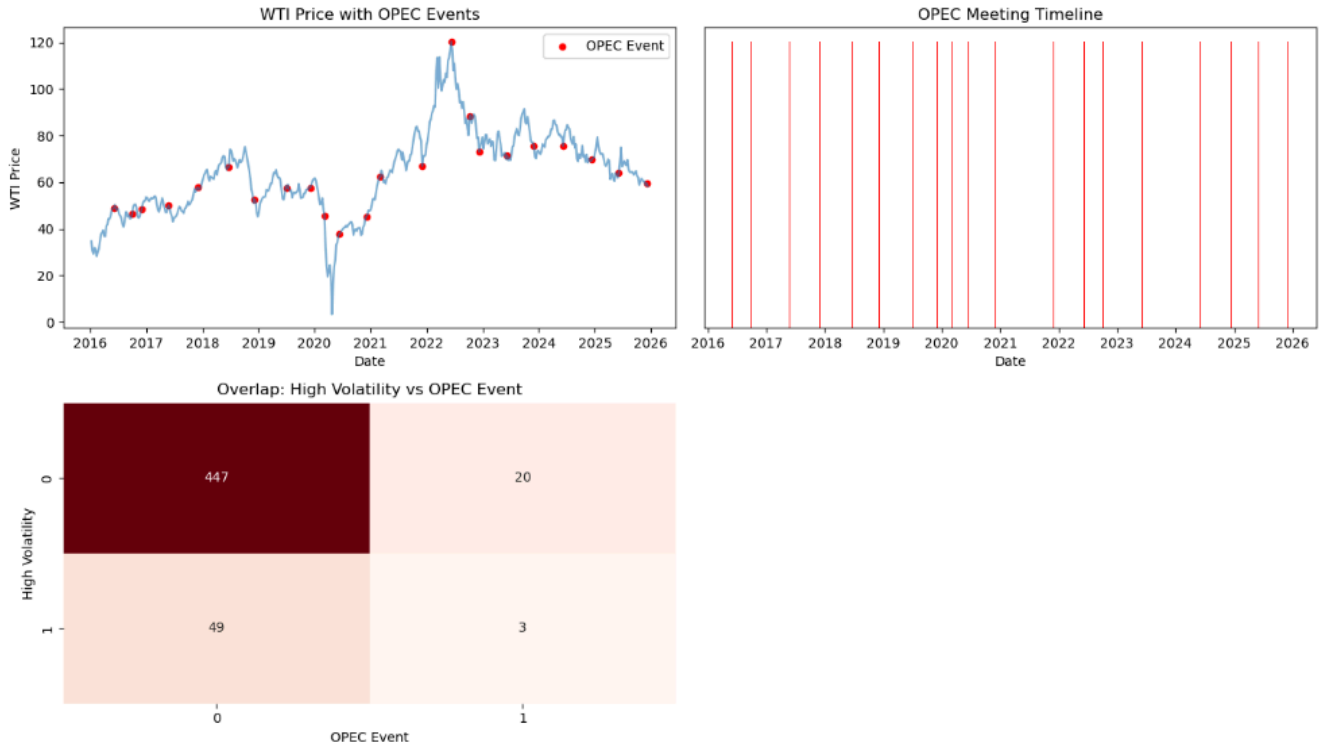


### WTI 가격 VS 원유 재고 (CrudeInventory)

원유 재고 추이를 분석한 결과, 2016년부터 2019년까지는 비교적 안정적인 범위에서 증감이 반복되었으나 2020년을 전후로 재고 구조에 뚜렷한 전환이 발생하며 이후에는 장기적인 감소 국면이 형성된 것으로 나타났다. 주간 재고 변화량은 대부분 제한적인 범위 내에서 움직였으나, 일부 시점에서는 단기간에 큰 폭의 증감이 집중적으로 발생하며 수급 충격이 국지적으로 나타나는 특성을 보였다.

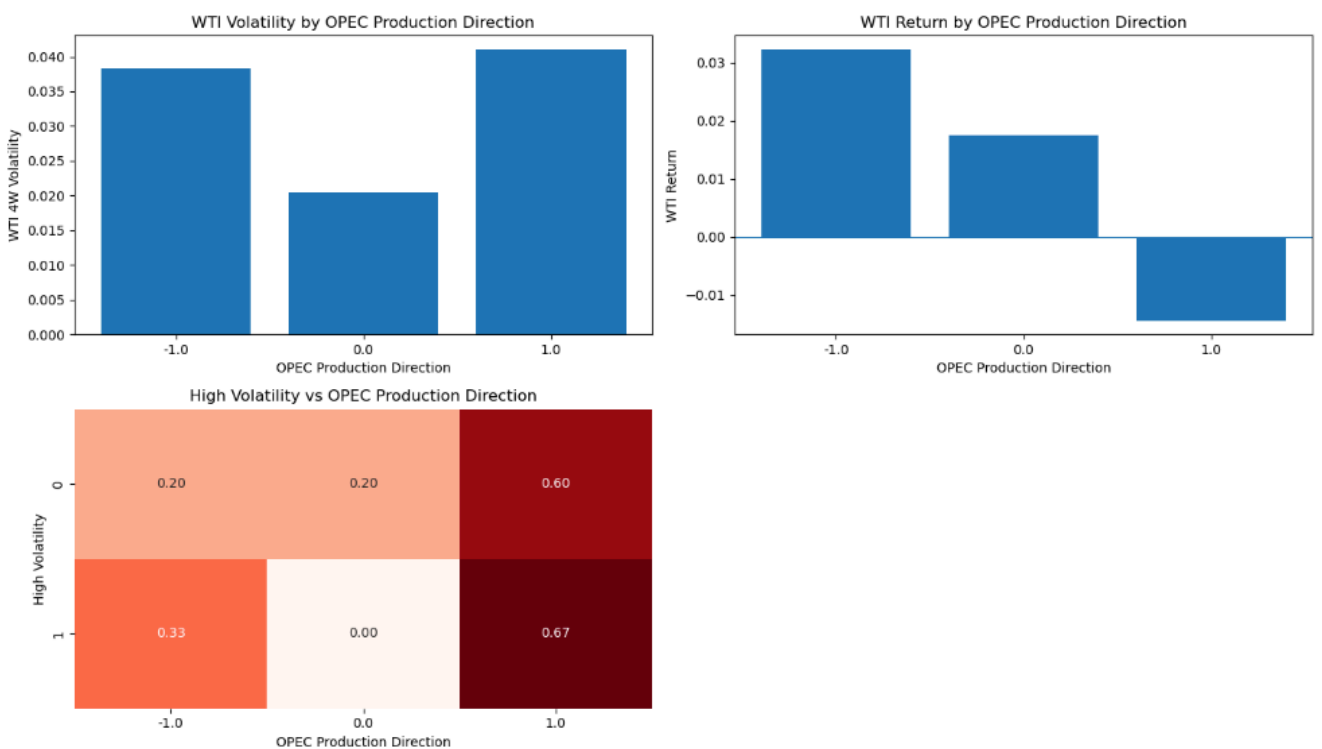
WTI 변동성과 원유 재고 변동성을 함께 비교한 결과는 재고 변동성은 전 기간에 걸쳐 비교적 큰 진폭으로 움직이는 반면 WTI 변동성은 특정 시점에서만 급격히 확대되는 위기형 구조를 보였다. 두 변수의 극단값 발생 시점은 완전히 일치하지 않았으며 시차 상관 분석에서도 재고 변동성이 WTI 변동성을 일관되게 선행하거나 후행하는 관계는 관찰되지 않았다. 다만 WTI 고변동성 구간에서는 재고 수준과 재고 변동성이 비고변동성 구간 대비 평균적으로 더 높고 재고 변화 방향 역시 증가 쪽으로 기울어지는 경향이 확인되었다.

이는 원유 가격 리스크가 단순한 재고 수준보다는 재고가 누적되는 불안정한 수급 환경과 결합될 때 확대될 가능성을 시사하며 원유 재고는 유가 변동성의 직접적 예측 변수라기보다 리스크 발생 가능성을 높이는 조건 변수로 해석하는 것이 타당하다.



## WTI 가격 VS OPEC 회의

OPEC 회의 시점을 WTI 가격 흐름과 함께 시각화한 결과 회의는 무작위적으로 분포하기보다는 가격 급변이나 추세 전환이 발생한 국면 인근에서 상대적으로 자주 관찰되었다. 또한 고변동성 구간과 OPEC 이벤트의 중첩을 확인한 결과 전체 기간 대비 고변동성 구간에서 OPEC 이벤트가 발생할 비율이 더 높게 나타났다. 이는 OPEC 회의가 유가 변동성을 즉각적으로 증폭시키는 직접적 트리거라기보다는 이미 불안정해진 시장 환경에서 향후 가격 방향성에 대한 신호를 제공하는 조건적 이벤트로 작용했을 가능성을 시사한다.



## WTI 가격 VS OPEC 증산/감산 방향 (OPEC\_ProdDirection)

OPEC 생산 방향 변수를 회의 시점에 발표된 결정 이벤트로 정의하여 분석한 결과 생산 정책이 동결된 경우보다 증산 또는 감산과 같이 방향성 있는 결정이 발표된 주에 WTI 변동성이 더 높게 나타났다. 정책 방향별 평균 수익률을 비교하면 감산 발표 시에는 양의 반응, 증산 발표 시에는 음의 반응이 관찰되었으나 표본 수가 제한적이므로 방향성 확인 수준의 결과로 해석하는 것이 적절하다.

WTI 고변동성 구간을 기준으로 OPEC 생산 방향 분포를 살펴본 결과 고변동성 구간에서는 동결보다 증산이나 감산과 같은 정책 변화가 발생한 비율이 더 높게 나타났다. 이는 특정 정책 방향 자체보다는 정책 변화가 발생했다는 사실 자체가 시장 불확실성을 자극하며 변동성 확대와 더 밀접하게 연결되어 있음을 시사한다. 종합적으로 OPEC 생산 방향 변수는 장기적인 정책 효과를 설명하기보다는, 고변동성 국면에서 시장 반응을 증폭시키는 이벤트성 조건 변수로 해석하는 것이 타당하다.

## Part 2. 고변동성 구간(high\_vol) 분류 모델링

Logistic Regression을 이용해 고변동성 구간(high\_vol) 진입 여부를 분류한 결과 검증 구간에서 모델은 high\_vol 클래스를 한 번도 예측하지 못했으며 확률 점수 역시 구분력이 거의 없는 수준으로 나타났다. 이는 검증 구간 내 high\_vol 표본 비중이 매우 낮은 불균형 구조에서 기본 설정의 로지스틱 회귀가 보수적인 예측을 수행한 결과로 해석할 수 있다. 실제로 모델은 대부분의 관측치를 정상 구간으로 분류하며 정확도는 높게 나타났으나, 고변동성 구간을 탐지하는 본 분석의 목적에는 부합하지 않는 결과를 보였다.

ROC AUC 역시 0.52 수준에 머물러 확률 기반 분류에서도 유의미한 구분력을 확보하지 못했으며, 이는 단순한 선형 분류기와 기본 임계값 설정만으로는 고변동성 이벤트를 효과적으로 포착하기 어렵다는 점을 시사한다. 따라서 이후 단계에서는 입력 변수 스케일링, 클래스 불균형 대응 기법, 그리고 임계값 조정을 통해 정확도 중심 평가가 아닌 고변동성 탐지에 적합한 Recall 중심의 평가 체계로 모델을 재설계한다. 이 베이스라인 결과는 이후 개선된 분류 모델의 성능 향상을 비교·해석하기 위한 기준으로 활용된다.

==Logistic Regression == ROC AUC : 0.5214

	precision	recall	f1-score	support
0	0.95	1.00	0.97	146
1	0.00	0.00	0.00	8
accuracy			0.95	154
macro avg	0.47	0.50	0.49	154
weighted avg	0.90	0.95	0.92	154

**StandardScaler**와 **class\_weight='balanced'**를 적용한 **Logistic Regression** 모델을 통해 클래스 불균형 환경에서도 **WTI** 고변동성 구간을 일정 수준 이상 탐지할 수 있음을 확인하였다. 기본 임계값 **0.5** 기준에서 **high\_vol**에 대한 재현율이 **0.75** 수준으로 개선되었으며 이는 이전 베이스라인 모델이 고변동성 구간을 전혀 탐지하지 못했던 결과와 대비되는 유의미한 변화이다. 정확도와 정밀도는 낮아졌으나, 본 분석의 목적이 고변동성 국면을 사전에 놓치지 않는 조기 경고에 있다는 점에서 이러한 **trade-off**는 허용 가능한 결과로 해석한다.

표준화된 회귀 계수 해석 결과, 주식시장 변동성(**sp500\_vol\_4w**)이 고변동성 발생 가능성을 가장 강하게 설명하는 변수로 나타났으며 **WTI** 단기 수익률, 원유 재고 변화, **OPEC** 생산 방향과 같은 수급·정책 요인 또한 고변동성 구간과 구조적으로 연결되어 있음을 확인할 수 있었다.

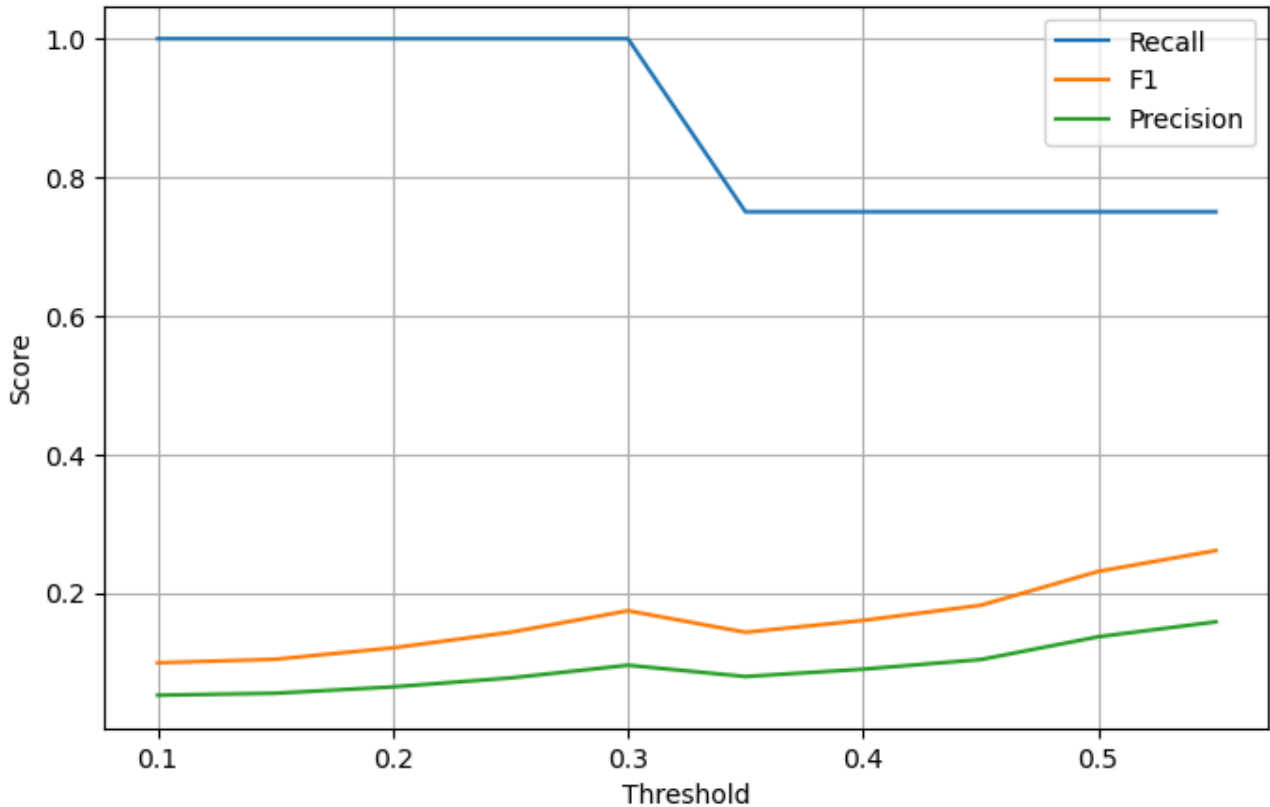
반면 달러 수익률은 음의 계수를 보여 달러 강세 국면에서는 상대적으로 고변동성 발생 확률이 낮아지는 경향이 관찰되었다. 이러한 결과는 **lag** 변수를 사용하지 않은 단순한 구조에서도 거시·금융·이벤트 신호가 고변동성 탐지에 의미 있는 정보를 제공함을 보여주며 이후 임계값 조정과 시차 변수 추가를 통해 추가적인 성능 개선 가능성이 있음을 시사한다.

== Logistic Regression | threshold 0.50 ==

	precision	recall	f1-score	support
<b>0</b>	0.982	0.740	0.844	146
<b>1</b>	0.136	0.750	0.231	8
<b>accuracy</b>			0.740	154
<b>macro avg</b>	0.559	0.745	0.537	154
<b>weighted avg</b>	0.938	0.740	0.812	154

== Logistic Regression 회귀 계수 확인 ==

	feature	coef
<b>6</b>	sp500_vol_4w	1.330227
<b>0</b>	wti_return	0.299621
<b>7</b>	inventory_change	0.292213
<b>10</b>	OPEC_ProdDirection	0.234391
<b>8</b>	inventory_vol_4w	0.192955
<b>4</b>	rate_vol_4w	0.183855
<b>5</b>	sp500_return	0.134475
<b>2</b>	usd_vol_4w	0.086128
<b>9</b>	OpecEventDummy	0.044064
<b>3</b>	rate_change	0.024160
<b>1</b>	usd_return	-0.243266



**Threshold**를 조정하며 고변동성(**high\_vol**) 탐지 성능을 분석한 결과 낮은 **threshold** 구간에서는 모든 고변동성 사례를 탐지하여 재현율이 1.00에 도달했으나 오탐이 크게 증가하며 정밀도와 **F1-score**가 낮게 나타났다.

반대로 **threshold**를 높일수록 정밀도와 **F1-score**는 개선되었지만 재현율이 급격히 감소하여 실제 고변동성 사례를 놓치는 문제가 발생하였다. 특히 **F1-score** 최대 기준으로 선택된 **threshold**는 고변동성 탐지율이 매우 낮아 조기 경고 목적에는 부적합한 결과를 보였다.

이에 본 프로젝트에서는 정확도나 **F1** 최대화가 아닌 고변동성 구간을 놓치지 않는 탐지 성능을 최우선 목표로 설정하였다. 재현율 1.00을 만족하는 **threshold** 후보군 중에서 정밀도가 가장 높은 값을 선택한 결과, **threshold** 0.30이 사전 경고 목적에 가장 적합한 운영 기준으로 판단되었다. 해당 기준에서는 검증 구간의 모든 고변동성 사례를 탐지하는 대신 일정 수준의 오탐을 감수하는 **trade-off**가 존재하지만 이는 리스크 관리 관점에서 놓침 비용을 최소화하기 위한 합리적인 선택으로 해석된다.

**Lag** 변수를 추가할 경우 결측치 증가로 인해 학습 및 검증 표본 수가 급격히 감소하였으며 이로 인해 시차 효과를 안정적으로 검증하기 어렵다고 판단하였다. 이에 본 단계에서는 **lag** 변수를 제외하고 현재 시점의 거시·금융·수급·이벤트 변수 조합만을 사용하여 고변동성 구간 탐지 모델링에 집중하였다.

**Logistic Regression**을 통해 구조적 신호의 존재를 확인한 이후 변수 간 비선형 관계와 상호작용을 학습하기 위해 트리 기반 부스팅 모델인 **XGBoost**를 적용하였다. 기본 **threshold(0.5)** 기준에서 **XGBoost**는 전체 정확도는 높게 나타났으나 고변동성(**high\_vol**) 탐지 성능은 매우 제한적이었으며 대부분의 변동성 확대 구간을 놓치는 보수적인



예측 성향을 보였다. 이는 변동성 확대 탐지 문제에서 모델 구조보다도 **threshold** 설정이 성능에 결정적인 영향을 미친다는 점을 시사한다.

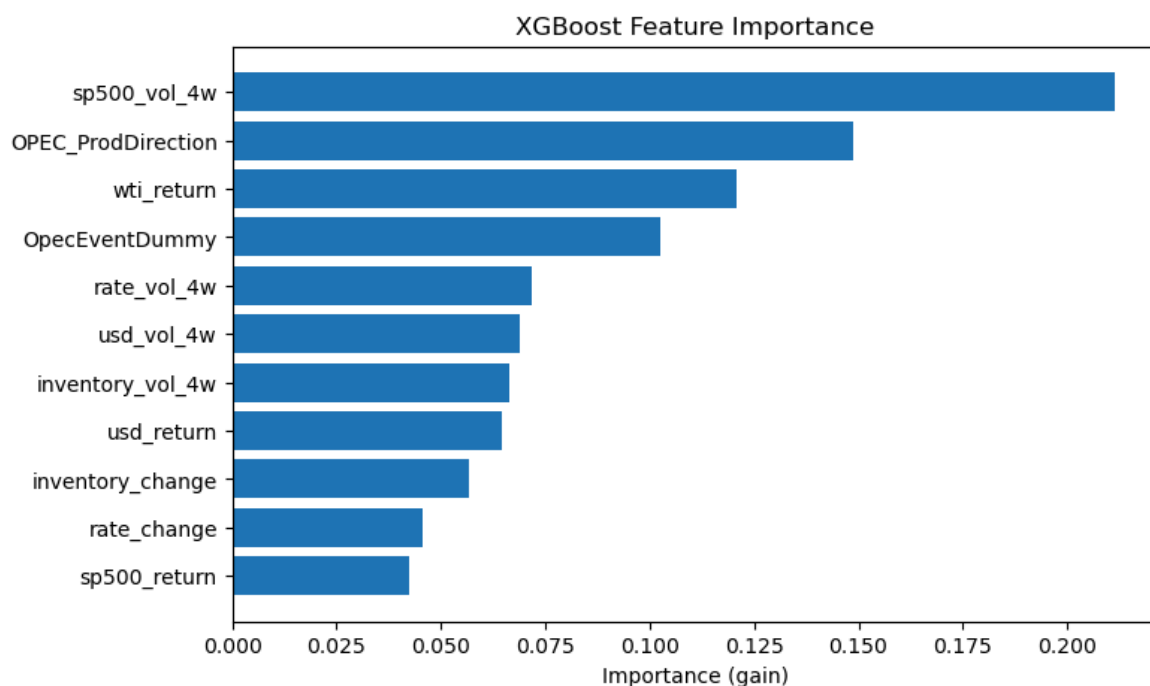
== XGBoost | threshold 0.50 ==

	precision	recall	f1-score	support
<b>0</b>	0.952	0.959	0.956	146
<b>1</b>	0.143	0.125	0.133	8
<b>accuracy</b>			0.916	154
<b>macro avg</b>	0.548	0.542	0.544	154
<b>weighted avg</b>	0.910	0.916	0.913	154

**Threshold**를 조정하며 성능 변화를 분석한 결과 낮은 **threshold** 구간에서는 **high\_vol**에 대한 재현율이 크게 개선되는 반면 오탐이 증가하는 **trade-off**가 명확히 관찰되었다.

본 데이터의 규모와 클래스 분포 특성상 재현율을 1.0으로 완전히 유지하는 **threshold**는 존재하지 않았으며 이에 따라 일정 수준 이상의 재현율을 확보하는 현실적인 탐지 기준을 선택하였다. **Recall  $\geq$  0.75** 조건 하에서 **threshold**를 비교한 결과, 약 0.13~0.14 구간에서 **F1-score**가 가장 높게 나타났으며 해당 기준은 전체 고변동성 사례의 약 75%를 탐지하면서도 오탐을 과도하게 증가시키지 않는 균형적인 지점으로 해석할 수 있다.

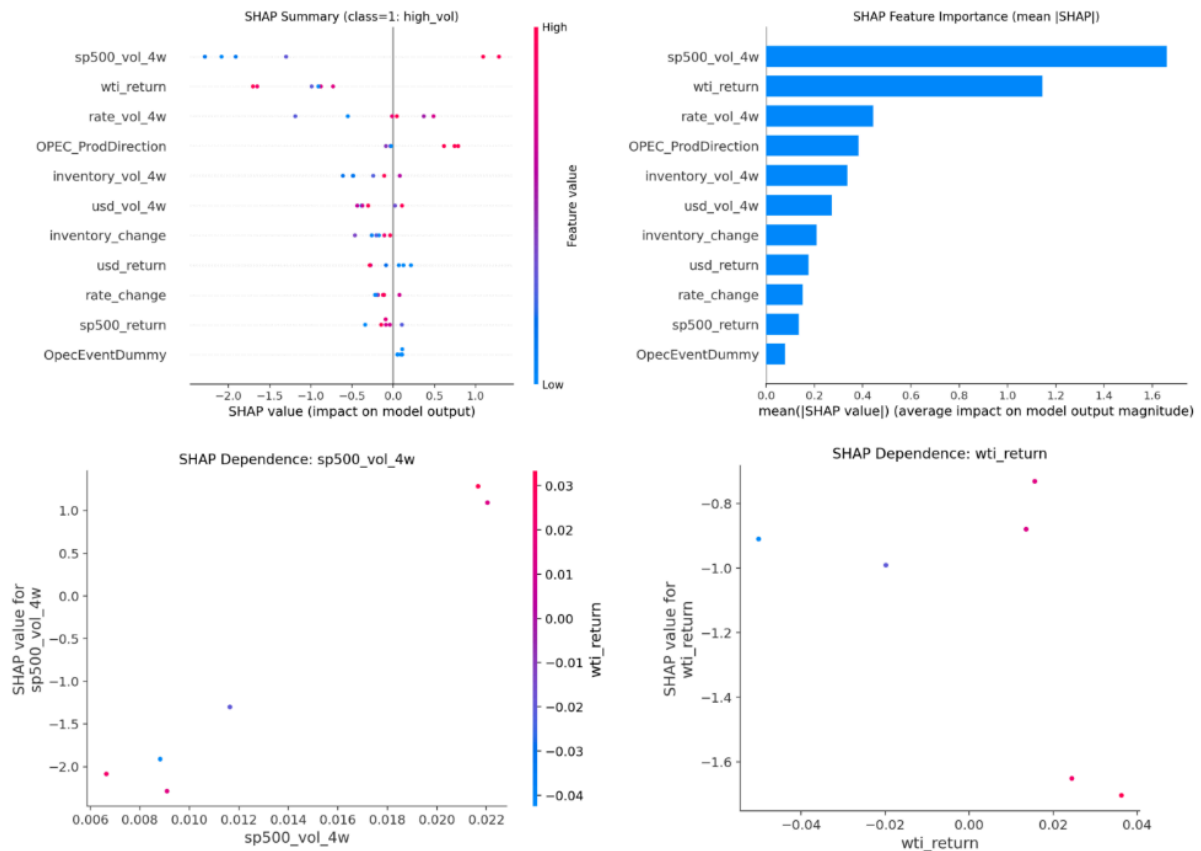
종합적으로 **XGBoost** 모델은 기본 분류 기준에서는 고변동성 탐지에 적합하지 않았으나, **threshold** 조정을 통해 조기 경고 목적에 부합하는 탐지 성능을 확보할 수 있었다. 이 결과는 변동성 확대 탐지가 단순한 이진 분류 문제가 아니라, 모델의 예측 확률을 어떻게 해석하고 운영 기준으로 활용할 것인지가 핵심인 확률 기반 의사결정 문제임을 명확히 보여준다.



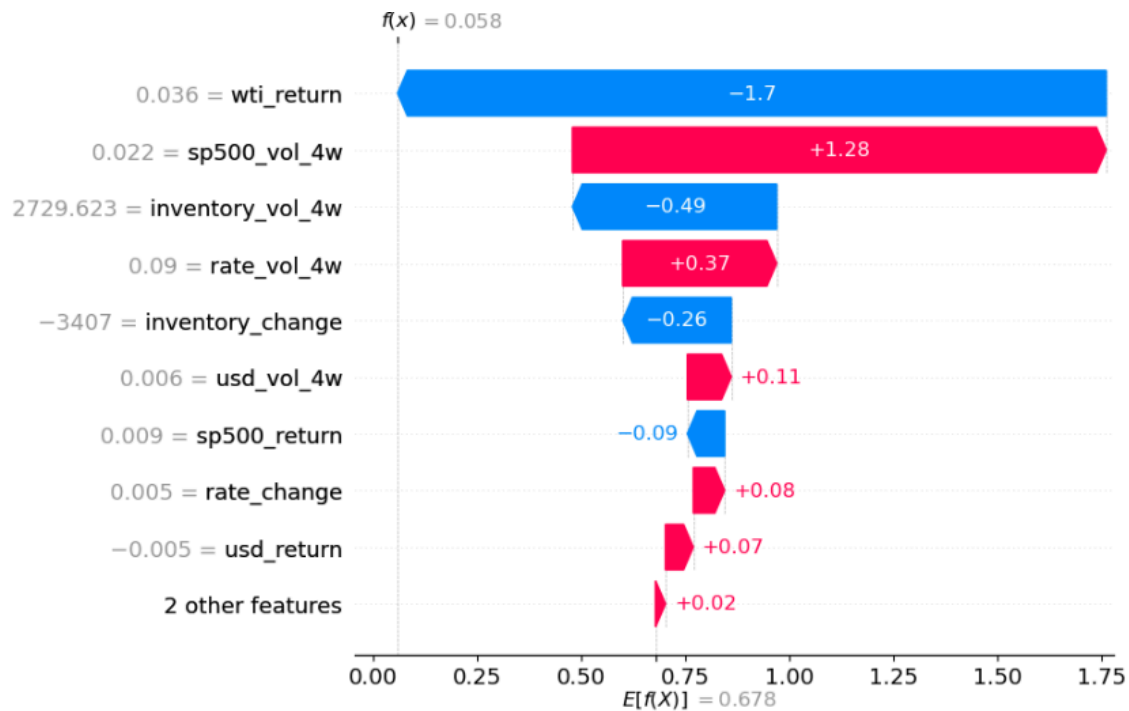
**XGBoost** 모델의 **feature importance**를 분석한 결과 고변동성(**high\_vol**) 탐지에는 개별 가격 수준 변수보다 변동성 관련 지표와 금융 시장 신호가 더 핵심적인 역할을 수행하는 것으로 나타났다. 특히 주식시장 변동성(**sp500\_vol\_4w**)이 가장 높은 중요도를 보였으며 이는 유가 변동성 확대가 원유 시장 내부 요인에 국한되지 않고 금융시장 전반의 위험 신호와 밀접하게 연결되어 있음을 시사한다.

또한 **OPEC** 생산 방향과 **OPEC** 이벤트 더미가 상위 중요 변수로 나타난 점은, 정책 결정 자체의 장기 효과보다는 정책 변화가 발생한 시점의 불확실성 신호가 변동성 확대 탐지에 의미 있는 정보를 제공함을 보여준다.

달러, 금리, 재고 관련 변수들 역시 중간 수준의 중요도로 고르게 분포되어 있어 고변동성 국면이 단일 변수에 의해 설명되기보다는 거시·금융·수급·이벤트 요인이 복합적으로 작용한 결과임을 확인할 수 있었다. 이러한 **feature importance** 결과는 본 프로젝트에서 설정한 고변동성 탐지 문제가 특정 가격 예측 문제가 아니라 여러 외생 신호를 종합적으로 해석해 리스크 국면을 식별하는 문제라는 점을 모델 차원에서도 뒷받침한다.



**SHAP** 분석 결과 고변동성(**high\_vol**) 구간은 단일 변수의 영향보다는 금융시장 변동성(**sp500\_vol\_4w**)과 원유 단기 가격 충격(**wti\_return**)이 결합될 때 발생하는 구조를 보였다. 금융시장 변동성이 일정 수준 이상으로 상승하면 고변동성 경고 확률이 비선형적으로 증가하며, 동일한 금융 환경에서도 **WTI** 가격의 급격한 변동이 함께 발생할 경우 경고 신호가 더욱 강화되는 상호작용 패턴이 관찰되었다. 이는 본 모델이 개별 이벤트 여부가 아닌, 시장 불안정성의 누적과 가격 충격이 동시에 발생하는 시점을 고변동성 국면으로 식별하고 있음을 시사한다.



검증 구간에서 경고 기준(threshold=0.13)을 적용한 결과 총 3주에서 high\_vol 경고가 발생하였으며 그중 예측 확률이 가장 높은 1건(약 51.5%)을 대표 사례로 선택해 SHAP Waterfall로 예측 근거를 분해하였다.

“경고를 낸 주” 기여 상위 3개 변수 해석

#### WTI 단기 수익률 (wti\_return)

해당 주의 WTI 단기 수익률은 양(+)의 값을 보였으며, 이는 고변동성 확률을 가장 크게 낮추는 요인으로 작용하였다. 즉, 가격 방향 자체는 안정적이었으며, 본 모델은 가격 방향보다 급격한 가격 충격 여부를 변동성 신호로 더 중요하게 반영하고 있음을 보여준다.

#### S&P500 변동성 (sp500\_vol\_4w)

S&P500 변동성은 고변동성 확률을 가장 강하게 끌어올린 핵심 요인으로 나타났다.

이는 글로벌 금융시장 불안이 원유 시장 변동성으로 전이되는 구조를 시사하며, WTI 변동성이 금융 환경과 밀접하게 연결되어 있음을 보여준다.

#### 원유 재고 변동성 (inventory\_vol\_4w)

재고 변동성은 고변동성 확률을 낮추는 방향으로 작용하였다.

해당 시점에서는 수급 측면의 충격이 제한적이었으며, 재고 요인은 금융시장 변동성에 따른 경고 강도를 일부 완충하는 역할을 수행하였다.

### Part 3. WTI 변동성 수준(wti\_vol\_4w) 회귀 모델링 - 결과 요약

본 파트에서는 Part 2에서 사용한 외생 변수들이 WTI 변동성 확대 여부(high\_vol)뿐 아니라 변동성 수준(wti\_vol\_4w)의 크기 자체도 설명할 수 있는지를 검증하였다. 이를 위해 동일한 입력 변수 구성을 유지한 채 Linear Regression과 XGBoost Regressor를 적용하여 회귀 성능을 비교하였다.

#### 회귀 모델 결과 요약

== Model Comparison ==

Linear | MAE: 0.0165 | RMSE: 0.0195 | R2: -6.4475

XGBoost | MAE: 0.0213 | RMSE: 0.0241 | R2: -10.3692

#### Linear Regression

MAE와 RMSE 기준 절대 오차는 비교적 작은 수준이었으나 결정계수( $R^2$ )가 음수로 나타나 변동성 수준에 대한 설명력이 매우 제한적이었다. 이는 단순 평균 예측보다도 변동성의 변화를 잘 설명하지 못했음을 의미한다.

#### XGBoost Regressor

비선형 관계와 변수 간 상호작용을 학습할 수 있음에도 불구하고 제한된 데이터 규모와 관측 구간으로 인해 설명력( $R^2$ ) 개선은 나타나지 않았다. 오히려 과적합 위험이 커지며 회귀 성능은 더욱 불안정해졌다.

#### 해석 및 시사점

본 결과는 WTI 변동성 수준이 개별 외생 변수의 선형 또는 비선형 조합만으로 안정적으로 예측되기보다는 구조적 변화(regime shift), 이벤트성 충격, 국면 전환의 영향을 크게 받는 복합적인 시계열 특성을 지님을 시사한다.

반면 Part 2의 분류 모델에서는 변동성 수준을 정확히 맞히지 않더라도 고변동성 국면(high\_vol)에 진입할 가능성 자체를 사전에 탐지하는 데에는 의미 있는 성능을 확보할 수 있었다.

#### Part 3 결론

WTI 시장 분석에서 변동성 수준의 정밀한 회귀 예측보다는 확률 기반 분류를 통해 위험 구간을 조기에 식별하는 접근이 데이터 구조와 실제 활용 목적에 더 부합하는 전략임을 확인하였다. 이에 따라 본 프로젝트의 핵심 모델링 방향은 회귀 예측이 아닌 고변동성 국면 탐지를 위한 조기 경고 시스템으로 정리한다.

## 최종 결론

### WTI 원유 가격 리스크 조기 경고 모델

본 프로젝트는 WTI 주간 유가를 단순 예측하는 대신 달러·금리·주식시장 변동성·원유 재고·OPEC 정책 이벤트를 결합하여 유가 변동성 확대(**high volatility**) 국면을 사전에 탐지하는 **Early Warning** 모델을 구축하는 것을 목표로 했다. 시계열 분석 결과, WTI 변동성은 반복적 주기보다는 금융·수급·정책 충격이 동시에 누적되는 국면에서 급격히 확대되는 특성을 보였으며 이에 따라 회귀 예측보다 분류 기반 리스크 탐지 문제로 재정의하였다.

**Logistic Regression**과 **XGBoost** 분류 모델을 통해 **S&P500** 변동성, 원유 단기 가격 충격, 금리·달러 변동성, **OPEC** 정책 이벤트가 고변동성 발생 확률을 구조적으로 설명함을 확인하였다. 특히 **threshold** 조정을 통해 “놓치지 않는 경고(**Recall** 중심)”라는 실제 운영 목적에 맞는 의사결정 기준을 설계하였다.

반면 변동성 수준을 직접 예측하는 회귀 모델은 데이터 규모와 비정상성 한계로 인해 설명력이 낮았으며 이를 통해 WTI 시장에서는 ‘정밀 예측’보다 ‘위험 구간 조기 탐지’가 더 현실적인 접근임을 도출하였다.

본 프로젝트는 원유 시장 리스크를 확률 기반 경고 시스템으로 해석하고 운영 관점에서 모델을 설계한 사례이다.