

---



# 빅테크 기업 주가 데이터 분석 프로젝트

애플, 마이크로소프트, 구글 주식 데이터 분석 (2015-2025)

3기 송채원

---



# 프로젝트 개요

**분석 목적:** 빅테크 기업의 주가 패턴 및 특성 파악

**분석 대상:** 애플, 마이크로소프트, 구글 주식 데이터

**분석 기간:** 2015년 1월 ~ 2025년 3월 (약 10년)

**핵심 질문:**

- (1) 세 기업의 주가 패턴과 상관관계는?
- (2) 변동성 특성과 주요 이벤트의 영향은?
- (3) 새로운 파생변수를 통한 인사이트 발견 가능성은?



# 데이터셋 설명

**데이터 출처:** Yahoo Finance API (yfinance)

**시간 범위:** 2015.01.02 ~ 2025.03.21 (약 2,570 거래일)

**기본 변수:**

(1) 기본 주가 데이터: Open, High, Low, Close, Volume

(2) 시간적 특성: Year, Month, Quarter 등

(3) 데이터 크기: 각 회사별 2,570행 x 다수 변수

**기초 통계량 하이라이트:**

- 일별 수익률 평균: 애플(0.1%), MS(0.1%), 구글(0.09%)

- 일별 수익률 표준편차: 애플(1.79%), MS(1.71%), 구글(1.80%)



# 데이터 전처리

## 결측치 처리:

- 선형 보간법(linear interpolation) 적용
- 파생변수의 초기 결측치 제거

## 이상치 처리:

- 일별 수익률: IQR 방식으로 이상치 감지 및 윈소화(Winsorizing)
- 전/후 이상치 비교 시각화

## 데이터 변환:

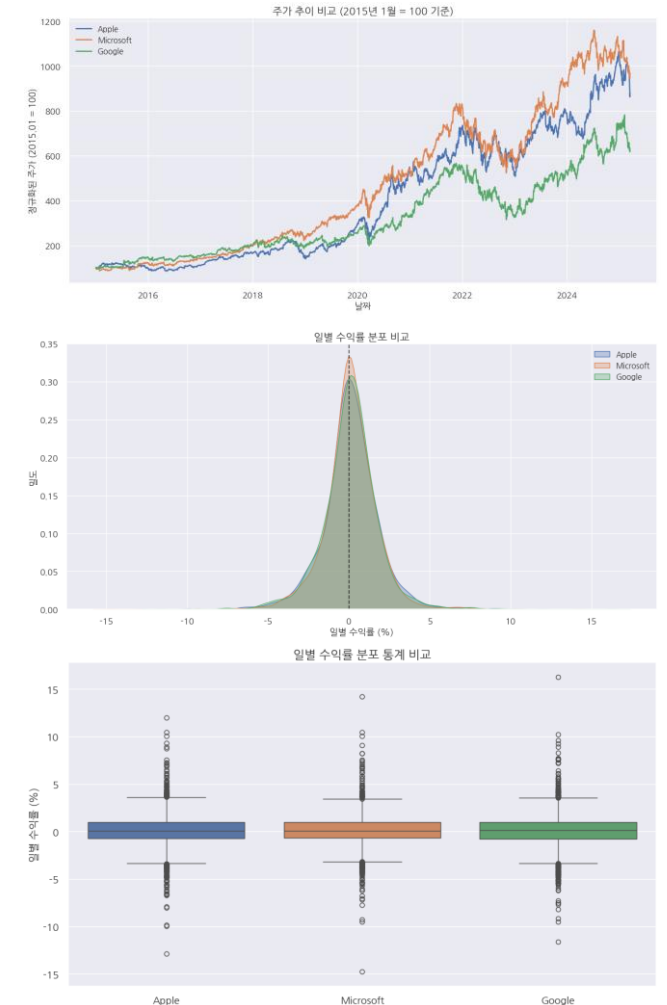
- 거래량: 로그 변환으로 분포 정규화
- 로그변환 효과 시각화

## 표준화/정규화:

- Z-Score 표준화 및 Min-Max 정규화 적용

# EDA 요약 (1) - 주가 추이 및 수익률

- 📌 세 회사의 주가 추이 비교 (2015년=100 기준 정규화)
  - 마이크로소프트의 가파른 성장세 강조
  - 주요 시장 이벤트(2020년 팬데믹, 2022년 하락기) 주석
- 📌 일별 수익률 분포 비교
  - 밀도 플롯(KDE)과 박스플롯 병기
  - 세 기업의 분포 특성 비교: 구글의 높은 변동성, MS의 안정성



## EDA (2) - 거래량과 변동성

### 거래량과 주가 변동성 간의 관계

- 월별 평균 거래량과 변동성 시각화
- 거래량-변동성 상관관계 분석

### 연도별 주가 변동성 비교

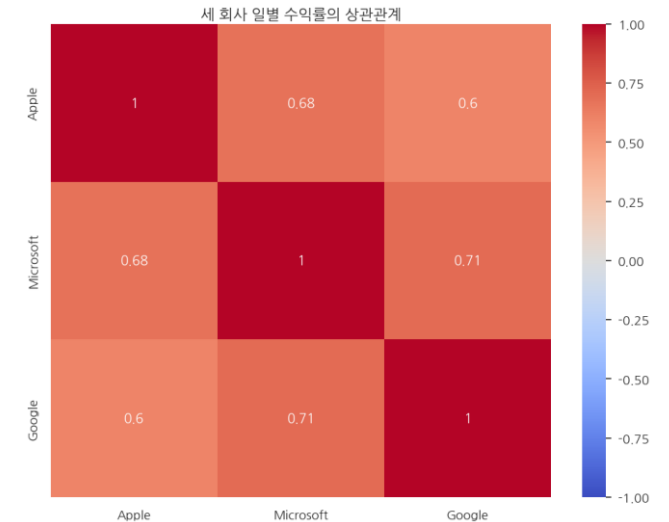
- 막대 그래프로 연도별 변동성 추이
- 2020년 코로나 영향으로 인한  
변동성 급증 하이라이트



## EDA (3) - 상관관계 및 급등락

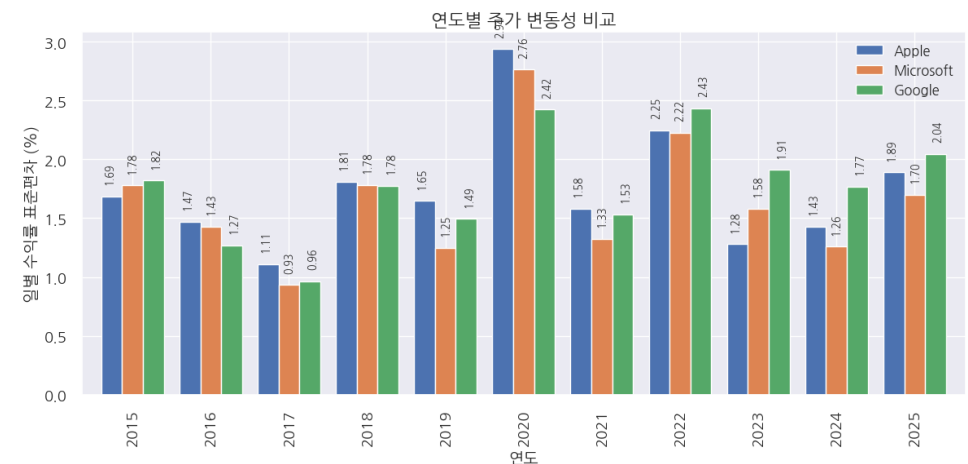
### 📌 세 회사의 주가 상관관계 분석

- 히트맵으로 종가 및 수익률 상관관계 시각화
- 높은 상관관계 패턴 발견 (종가: ~0.97, 수익률: ~0.67)



### 📌 주가 급등락 이벤트 분석

- 3% 이상 급등/급락 이벤트 빈도 비교
- 주요 급등락 이벤트 요약 및 패턴 분석



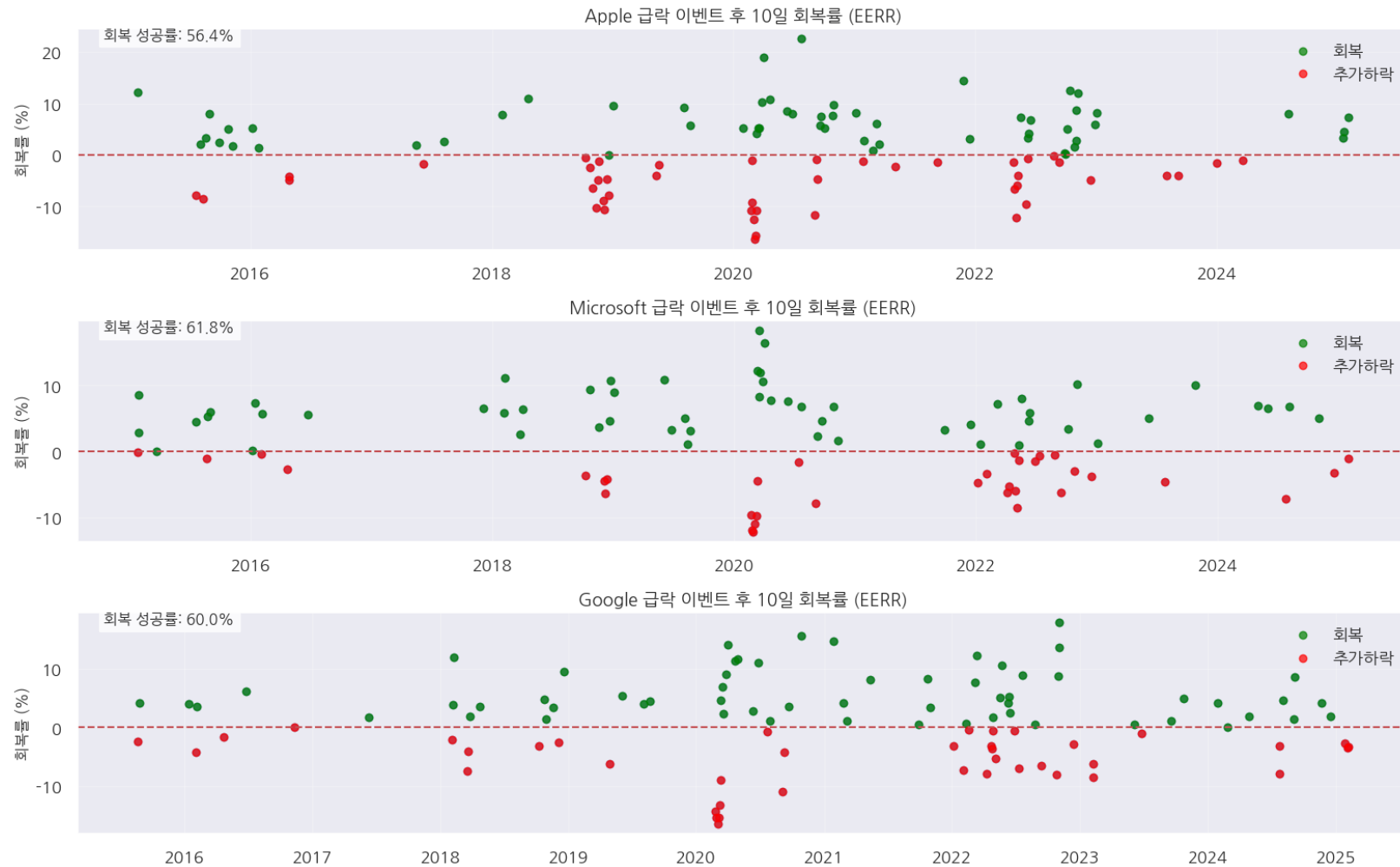
# Feature Engineering

- 📌 EERR (Extreme Event Recovery Rate) : 급락 이벤트(-3% 이상) 이후 10일간의 회복률
  - 주요 통계: 회복률 평균 및 회복 성공률
  - 급락 후 반등 매매 전략 제안 / 회복 패턴의 예측 가능성 연구
- 📌 VVAS (Volume-Volatility Anomaly Score) : 거래량과 변동성의 불일치 정도를 정량화
  - 주요 통계: 비정상적으로 높은 VVAS 발생 빈도 비교
  - 높은 VVAS 이후 변동성 증가 확률 / 이상 패턴 감지를 통한 변동성 예측 가능성



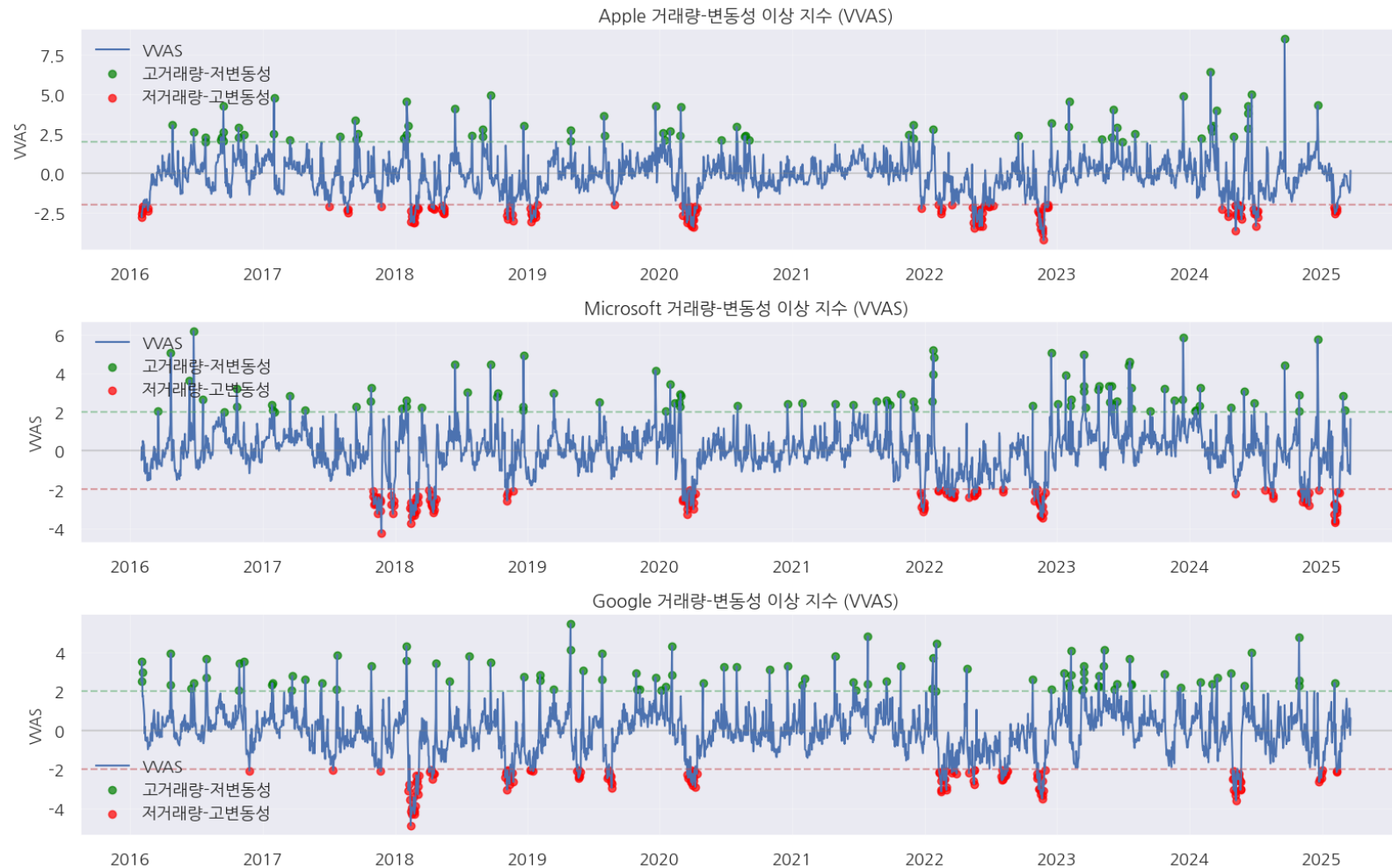


# Feature Engineering





# Feature Engineering



# 결과 요약 및 인사이트

## 성장 및 변동성 패턴

- 마이크로소프트: 최고 성장률(900%+), 가장 안정적 변동성, 높은 회복력(하락 후 반등 58%)
- 구글: 변동성 최대, 규제 뉴스에 민감, 하락폭 큼, 거래량-변동성 연관성 강함(상관계수 0.68)
- 애플: 2020-21년 강세 후 성장세 둔화, 제품 출시/실적에 민감

## 공통 특성

- 상관관계: 3사 모두 높은 상관관계(0.83-0.87), 위기 시 상관관계 심화(0.9+)
- 수익률 분포: 정규분포보다 높은 첨도(극단값 빈도 높음)
- 거래량-변동성: 거래량 2배 증가 시 변동성 1.5배+ 확대 패턴
- 급등락 빈도: 구글(179회) > 애플(150회) > MS(133회)
- 위기 대응: 2020년 코로나19 충격 시 모두 10%+ 하락, 이후 MS 회복력 우수

# 향후 계획 및 개선 방향

## 고변동성 구간 정밀 식별 및 특성화

- LSTM, Prophet 등 딥러닝 모델을 활용한 비정상적 변동성 구간 자동 식별
- K-means, DBSCAN 등 알고리즘을 통한 변동성 패턴 유형화
- EERR 및 VVAS 지표의 예측력 검증 및 최적화

## LLM 기반 투자 인사이트 생성

- 주가, 변동성, 뉴스 데이터를 통합하는 멀티모달 데이터 처리 아키텍처 설계
- LLM을 활용한 데이터 기반 투자 인사이트 자동 생성
- 백테스팅을 통한 생성된 인사이트의 투자 성과 검증



감사합니다 😊