개인이 처음부터 끝까지 데이터를 Scraping을 수동으로 하는 것은 거의 불가능에 가까움. Scraping은 온체인 분석에서 가장 어렵고 전문적인 부분임.

1. 블록 탐색기 (Etherscan 등)를 이용한 수동 분석

* 무료로 이용 가능하나 매우 어렵고 느리며 엔티티 클러스터링 불가능
* 점수 계산은 불가능

1. 온체인 분석 플랫폼 활용: 가장 현실적인 방법

Axiom, Nansen, Arkham Intelligence, Dune Analytics 등…

데이터가 이미 식별, 클러스터링, 라벨링되어 있음.

* 우리가 원하는 것은 raw data…

그래서 Raw Data 확보 및 점수 계산을 위한 새로운 실행 계획: Easy One

1. 데이터 소스 선택 (by Queryable Platform): Table 형태의 Raw Data를 직접 조회(Query)하거나 추출(Export) 할 수 있는 플랫폼을 선택해야 함.

* Dune, Nansen, Arkham

1. 엔티티 기준 홀더 목록 추출: SQL 쿼리를 이용해 우리가 원하는 토큰의 전체 홀더 목록을 엔티티 기준으로 그룹화하여 추출한다.

* Input: 특정 토큰 컨트랙트 주소
* Process

1. 토큰의 모든 홀더 주소와 잔액을 가져온다.
2. 플랫폼이 미리 해놓은 지갑 라벨링 및 클러스터링 정보와 결합한다.
3. 동일 엔티티로 묶인 지갑들의 잔액을 합산한다.

* Output: [Entity A Holdings], [Entity B Holdings], …. 와 같은 데이터 목록

1. 직접 4가지 점수를 계산: 플랫폼에서 라벨링 돼있는 것을 기준으로 판단

하지만 플랫폼의 라벨마저도 100% 신뢰하지 않고 진짜 ‘원본’을 원한다면 Dune에서 SQL로 추출하거나 Etherscan의 Export 기능으로 CSV 파일로 확보해야 함.

* 하지만 이것도 수동 다운로드라서 우리의 프로젝트와는 맞지 않는다

자동화 파이프라인 설계 (Raw Data 수집->4대 리스크 계산-> 최종 위험점수)

1. 데이터 수집 계층: 실시간 블록체인 데이터 API를 사용
2. API 제공업체

* Alchemy / Infura:  가장 기본적인 블록체인 노드 접속을 제공합니다. 개발자는 이 노드를 통해 새로 생성되는 DEX 유동성 풀(PairCreated 이벤트)을 실시간으로 감시하고, 관련된 모든 트랜잭션 데이터를 직접 수집할 수 있습니다. 가장 자유도가 높지만, 다루기 가장 어렵습니다.
* Moralis / Covalent / The Graph: 한 단계 더 가공된 데이터를 제공합니다. 복잡한 원본 데이터를 사람이 읽기 쉬운 형태로 변환해주어 개발이 더 용이합니다. 예를 들어, 'The Graph'를 이용하면 특정 토큰의 모든 거래 내역을 효율적으로 쿼리할 수 있습니다.
* 유료 데이터 플랫폼 API: Nansen, Arkham, Dune 등은 API를 통해 자신들이 이미 분석하고 라벨링한 데이터를 제공. 가장 강력하지만 비용이 발생. 예를 들어, Arkham API를 호출하여 특정 지갑의 클러스터 정보를 받아오는 식.

1. 데이터 처리 및 분석 엔진
2. 개발 환경: 보통 파이선과 web3.py , pandas 같은 라이브러리를 사용
3. 구현 로직
4. 데이터 정제: API로 받아온 Raw Data(거래 목록)를 분석하기 쉬운 형태로 가공.
5. 핵심 로직 코드화: 우리가 지금까지 논의한 모든 분석 로직을 코드로 구현

* 엔티티 클러스터링 로직: '자금 출처 동일성' 등을 기준으로 지갑들을 자동으로 그룹화하는 알고리즘. (가장 복잡하고 어려운 부분)

라벨링 로직: 클러스터링된 엔티티가 'Dev', 'Sniper' 등의 기준에 부합하는지 판단하는 코드.

* 4대 점수 계산 로직: 각 평가 요소(Dev H., NHHI, Insiders, Snipers)의 점수를 계산하는 함수들.
* 최종 점수 및 페널티 로직: 가중치를 합산하고, '치명타 페널티' 규칙을 적용하여 최종 점수를 산출하는 코드.

1. 데이터베이스 및 알림 계층: 분석된 결과를 저장하고 사용자에게 알려주는 부분

* 데이터베이스: 분석이 끝난 모든 토큰과 그에 대한 위험 점수를 PostgreSQL, MongoDB 같은 데이터베이스에 저장하여 관리합니다.
* 알림 시스템:

1. 새로운 토큰이 분석될 때마다 결과를 텔레그램 봇이나 디스코드 웹훅으로 전송합니다.
2. "경고: 신규 토큰 'ABC' 최종 위험 점수 95점!" 또는 "기회: 신규 토큰 'XYZ' 최종 위험 점수 15점!" 과 같은 알림을 실시간으로 받아볼 수 있습니다.