

Welcome

rse112's GitHub Profile

🔗 Skills



링크

블로그 : <https://rse112.github.io/>

github주소 : <https://github.com/rse112>

🌟 자기소개

안녕하세요 조찬규입니다

끊임없이 배움과 도전을 즐기며, 성장하는 데이터 분석가를 꿈꾸고 있습니다.

빠르게 발전하는 AI 기술에 발맞춰 새로운 기술을 익히고, 이를 통해 꾸준히 발전하는 과정에서 큰 보람을 느낍니다.

무엇이든 긍정적인 마인드로 임하며, 이러한 에너지가 주변에도 좋은 영향을 미칠 수 있도록 노력하고 있습니다. ☺

경력

- **학교:** 건국대학교
학과: 수학과 전공 / 응용통계학과 다전공

1. 건국대학교 감염병 통계/수학 모델링 연구실

2023년 5월 ~ 2023년 11월 (6개월)

- 감염병 모델링 및 통계 분석 연구 수행

2. (주)리얼데이터

2024년 2월 1일 ~ 2024년 6월 21일 (5개월)

- KB증권 M-able 어플 사용자 데이터 세그먼트 분류
- IRP/퇴직연금 관련 마케팅 데이터 분석

▶ 🧑‍💻 프로젝트 (FinTrendX)

Project Name	Description	GitHub Repository	Result Website
FinTrend	핫한 금융키워드를 추출하는 프로젝트	GitHub Link	View Result

💡 프로젝트 소개

FinTrendX는 사람들이 금융 관련 키워드를 네이버에 검색하면서 얻은 데이터를 통해, 현재 어떤 키워드가 핫한지, 어떤 키워드가 지속적으로 상승하고 있는지를 확인하기 위한 프로젝트입니다. 연관 검색어, 트렌드 데이터, 뉴스, 블로그 데이터를 수집하고 분석하여 급상승, 지속상승, 규칙성을 보이는 데이터를 분류하는 것을 목표로 합니다.

📁 프로젝트 구조

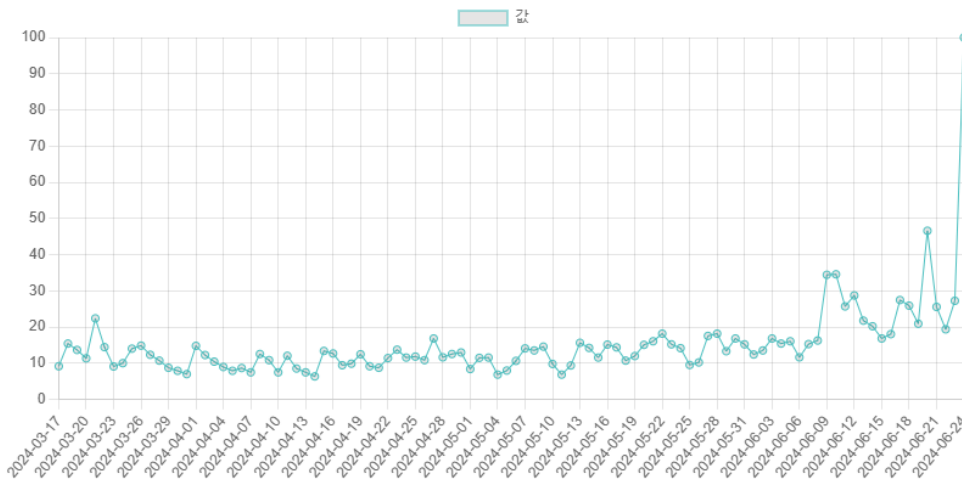
1. 메인 키워드별 연관검색어 집계
2. 각 키워드별 트렌드데이터 집계
3. 선별함수를 통한 급상승/지속상승/규칙성 데이터 분류
4. 네이버 뉴스데이터/구글검색어 데이터 집계 (비고: 구글검색어 데이터는 api 자체 오류 다분)
5. 블로그데이터를 통한 활동성 데이터 집계
6. 전체 데이터 Merge

📝 작업 흐름

1. 사용자가 지정한 대분류 키워드를 입력합니다.
2. 네이버 API를 통해 해당 키워드들의 연관검색어를 추출합니다.
3. 네이버 트렌드 API를 사용하여 추출한 연관검색어들의 검색어 추이 데이터를 수집합니다.
4. 선별 함수를 통해 일별/주별/월별 데이터를 분석하여 급상승, 지속상승, 규칙성을 보이는 키워드를 추출합니다.
5. 추출된 키워드들을 바탕으로 네이버 뉴스데이터와 구글 검색어 API를 사용하여 데이터를 가져옵니다. (비고: 구글 검색어 API는 오류가 발생할 수 있음)
6. 모든 데이터를 종합하여 하나의 데이터셋으로 Merge합니다.

📊 결과물

- 결과 페이지: <https://trendkey-7a41071967af.herokuapp.com/>



[그림 1] 트렌드 분석 결과 시각화

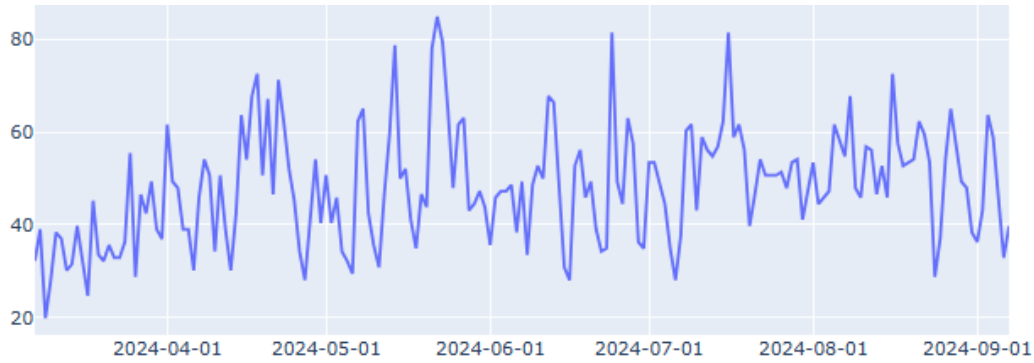
⚠ 데이터 선별함수에서 고려된 점

⚙ 일별 급상승 키워드 탐지 조건

일별 급상승 키워드를 탐지하기 위해 다음과 같은 조건들을 설정했습니다:

1. 검색량 2배 이상 증가 조건

- 현재 검색량이 이전 검색량보다 2배 이상 증가한 경우를 급상승 키워드로 간주합니다.



그러나, 위와 같은 그래프처럼 변동성이 크지만 일정 수준에 도달하지 못하는 키워드들이 포함되어, 유의미한 데이터를 도출하기 어려운 문제가 발생했습니다.

이를 해결하기 위해 추가적인 조건을 설정하였습니다:

- **현재 검색량이 2배 이상 증가했다**라도, **최소한 60 이상의 검색량**을 기록한 경우에만 급상승 키워드로 판단합니다. 이를 통해 검색량이 많이 상승했다라도 그 상승이 의미 있게 작용하지 않는 경우(작은 검색량 증가)는 제거합니다.

2. 검색량 95 이상일 때

- **현재 검색량이 95 이상**이고, **이전 검색량보다 많을 때** 급상승 키워드로 간주합니다. 이 조건을 통해 검색량이 일정 수준 이상 높은 키워드만을 선별하여 의미 있는 데이터를 얻을 수 있습니다.

3. 검색량 차이가 5 이상일 때

- **이전 검색량과 현재 검색량의 차이가 5 이상**일 경우에만 급상승으로 분류합니다. 이를 통해 작은 변동이 아닌 **유의미한 증가**를 포착할 수 있습니다. 예를 들어, 100에서 105로 증가한 경우 급상승으로 간주합니다.

⚙️ 지속 상승 키워드 탐지 조건

지속 상승 키워드를 탐지하기 위해 다음과 같은 조건을 설정하였습니다:

1. 추세의 상승 여부 확인

최근 검색량의 **상승 추세**를 확인하기 위해, OLS 회귀분석을 통해 계산된 기울기가 양수일 경우에만 키워드를 지속 상승 키워드로 간주합니다.

2. 최근 검색량의 비교

이전 검색량(**last_2**)보다 현재 검색량(**last**)이 커야 하며, **최근 15일간의 상승 추세**가 일정 범위 내에 있어야만 유의미한 상승 키워드로 판단합니다.

3. 모드별 추가 조건

- **주별/월별 모드**에서는 최근 검색량과 과거 검색량의 비교를 통해, **검색량이 일정 수준 이상**에서 지속적으로 증가하고 있는지 확인합니다.
- **일별 모드**에서는 최근 3일간의 검색량을 바탕으로 상승 패턴을 분석하여 지속 상승 여부를 판단합니다.

이러한 조건들을 통해 **일정한 상승 추세를** 보이며, **노이즈성 데이터가 아닌** 의미 있는 검색어만을 선별하게 됩니다.