

# 9주차 회의록

회의일시	2020년 10월 30일 (금요일) 14:00 – 16:30	장소	한티역
팀원	박채린, 이혜승		
회의안건	<ul style="list-style-type: none"> <li>3개의 포털 트렌드 정보를 하나의 DB 테이블로 merge시키기</li> <li>포털 트렌드 검색 비율과 경기동행지수의 비교( 과연 정말 경기가 안좋을 때, 포털에의 '경제' 검색량이 많을지 확인 및 증명 by 시각화, 상관계수 )</li> </ul>		

논의 및 결정 사항

내용

비고

1. 3개의 포털 트렌드 정보를 하나의 DB 테이블로 merge시키기

i. Google trends

ii. Kakao trends

iii. Naver datalab

➔ DB에 portal\_trends\_ratio 테이블로 저장[코드]

: 포털이 제공하지 않는 기간에 대해서는 0으로 저장했음

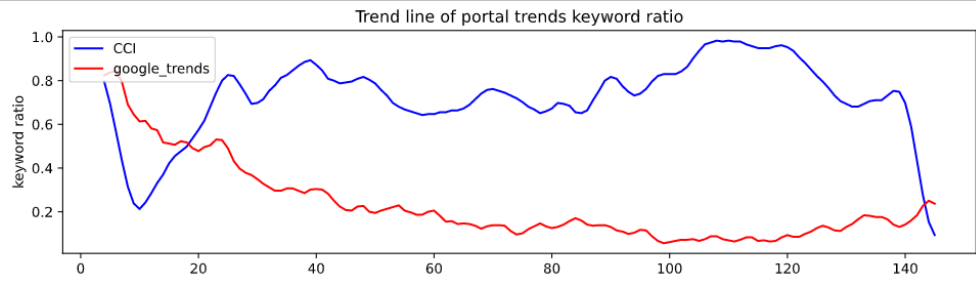
	id	year	month	day	google	kakao	naver
0	1	2015	9	7	69.0	0.0	0.00000
1	2	2015	9	14	74.0	0.0	0.00000
2	3	2015	9	21	69.0	0.0	0.00000
3	4	2015	9	28	62.0	0.0	0.00000
4	5	2015	10	5	61.0	0.0	0.00000
...	...	...	...	...	...	...	...
264	265	2020	9	28	63.0	19.0	25.29102
265	266	2020	10	5	81.0	21.0	28.63240
266	267	2020	10	12	82.0	24.0	26.85177
267	268	2020	10	19	93.0	21.0	27.25273
268	269	2020	10	26	90.0	0.0	19.11270

2. 포털 트렌드 검색 비율과 경기동행지수의 비교 [코드]

(구글의 월간 검색 비율(2008년7월~)데이터 활용)

i. 시각화

경기동행지수와 구글에 '경제' 키워드 검색 비율이 실제로 음의 관계를 보이고 있음을 시각적으로 확인할 수 있다. (이때 경기동행지수가 100미만이면 불황, 100이상이면 호황이다.) 즉, '경제'에 관심을 갖고 검색을 많이 해보는 때는 실제 경제 상황이 좋지 않음을 확인할 수 있다.



Moving average(windows = 5)로 smooth하게 표현해본 결과이다.

ii. 회귀분석과 피어슨 상관계수 확인

- 피어슨 상관계수: -0.401755

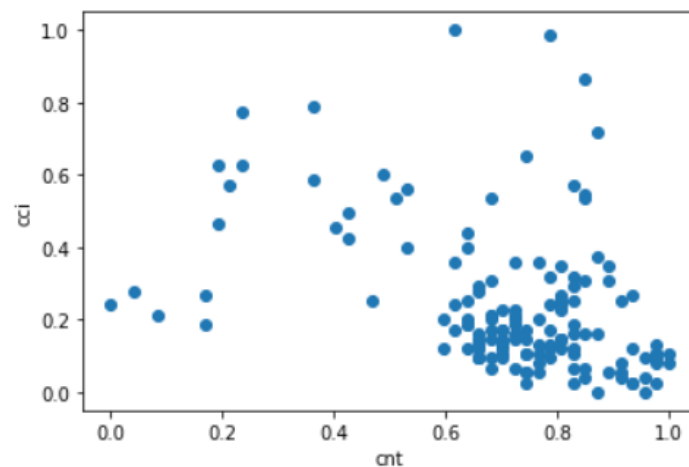
(-0.7 ~ -0.3은 강한 음의 상관 관계를 가진다. 즉, cci와 트렌드 검색어비율은 음의 상관 관계를 가진다.)

- 단순 회귀분석 결과

: cnt의 계수는 -0.41로 두 변수는 음의 관계를 보이며, 유의수준 0.05에서 **p-value는 0.05보다 작으므로 통계적으로 유의미하다고** 할 수 있다.

( 결정계수가 0.16으로 선형의 관계가 강하지는 않다고 해석할 수 있다. 따라서 cnt와 cci는 선형관계가 아닌 비선형관계일 수도 있다. )

- 산점도



**Issue 및 질문** 이러한 방법으로 포털 트렌드 변수의 유의성을 확인해도 될지? ( '과연 정말 경기가 안좋을 때, 포털에의 '경제' 검색량이 많을지'에 대하여)

**Issue)** 네이버 뉴스에서 검색 기간 내 '뉴스 기사 수' 스크래핑 불가... (네이버 사이트 개편)

➔ 빅카인즈에서 csv파일 다운로드 받아야할 지 고민중.

공  
지  
사  
항

- 카카오/ 네이버도 월간으로 확인해보기(변수 유의성)

	계획
10 주차(11/2 ~ 8)	11/4 교수님 미팅 전 ● 뉴스 기사 수 키워드(침체 불경기 소득감소 고용위기) DB table 저장.
11 주차(11/9 ~ 15)	● 경제 단기 지표 생성
12 주차(11/16 ~ 22)	11/18 교수님 미팅 ● 경제 단기 지표 생성
13 주차(11/23 ~ 29)	프로젝트 완성 목표> ● 유의성 확인(소비자심리지수와 비교)
14 주차(11/30 ~ 12/6)	12/2 교수님 미팅 ● 발표 준비 및 영상 제작 12/11 프로젝트 발표 영상 제출