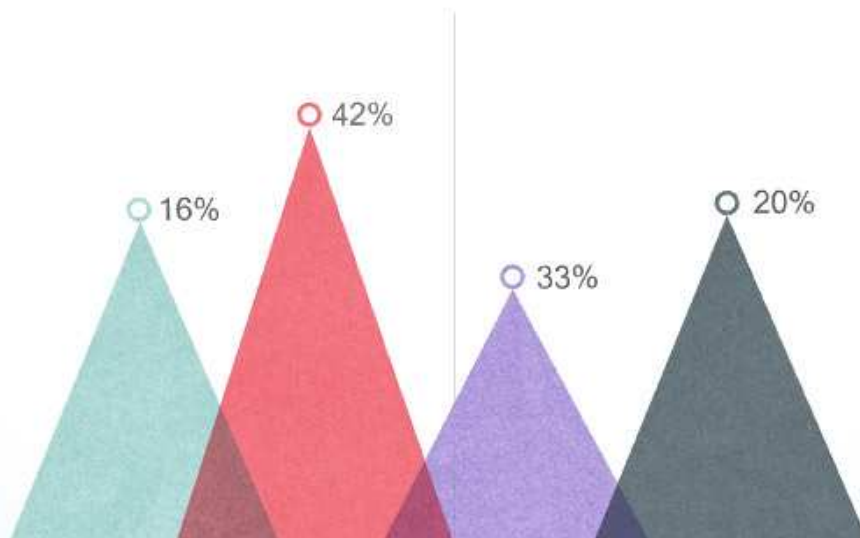


과제

[데이터 사이언스 과정] 프로젝트 과제



공지 및 프로젝트 소개

1. 프로젝트 과제는 **데이터사이언스 자격증 발급과 관련된 중요한 종합과제**입니다.
2. 모듈2 텍스트마이닝은 빅카인즈에서 제공되는 신문 기사를 이용한 '미래자동차 관련 뉴스의 텍스트마이닝' 과제입니다.
3. 정답이 정해져 있는 것이 아니기 때문에 세부 과제별로 최선의 방법을 모색하고, 학습자의 판단에 따라 분석결과를 정리해야 합니다. 평가 역시 **정답이 아닌 분석 로직(logic)과 분석 효율성을 중심으로** 하며, 학습 이외의 내용이 과제로 제출될 수 있습니다. 학습 이외의 과제는 다양한 정보(구글링 등)를 통합하여 해결하시기 바랍니다.
4. 본 과제 제출마감은 **2019년 7월 31일(수요일)**까지입니다. 과제물은 wise@wiseinc.co.kr로 제출 바랍니다.
 - 제출 과제물: 1) 결과물 정리 ppt 2) python 코드문서 3) 분석에 적용된 데이터파일
5. 본 과제의 풀이는 화상강의가 아닌 별도의 영상강의를 추후 제공할 예정입니다 (홈페이지 강좌 업로드)

1 과제

1. 데이터구조 및 과제1

빅카인즈에서 신문기사 다운로드 받기

- www.bigkinds.or.kr 접속 > 검색/분석 > 키워드 '미래자동차' 검색 (기간 1년)
- Step3 분석결과 및 시각화에서 '데이터다운로드' 하단의 다운로드를 선택하면 엑셀파일로 기사 파일 다운로드 가능

The screenshot shows the BigKinds website interface. At the top, there is a navigation bar with 'BIGkinds NEWS BIG DATA & ANALYSIS' and menu items like '검색·분석', '뉴스보기', '특별기획', '커뮤니티', and '빅카인즈 소개'. Below the navigation bar, there is a breadcrumb trail: '홈 / 검색·분석 / 뉴스 검색·분석'. The main content area is divided into three steps: 'Step 01. 뉴스 검색 - 미래자동차', 'Step 02. 검색 결과', and 'Step 03. 분석 결과 및 시각화'. Under 'Step 03', there is a sub-menu with '데이터 다운로드' highlighted. Below this, there is a table with columns for '뉴스 식별자', '일자', '언론사', '기고자', '제목', and '통합 분'. The table contains 20 rows of data. At the bottom right of the table, there is a '다운로드' button.

	뉴스 식별자	일자	언론사	기고자	제목	통합 분
15	02100851.20190605091141001	20190605	아주경제	김해원	현대차, 자유학기제 '미래차 학교' 2학기 모집...해외서도 신청 가능	사회>교육
16	02100801.20190605090218001	20190605	아시아경제	김지희	현대차, 자유학기제 특화 프로그램 2학기 참여학교 모집	경제>자
17	02100101.20190605085741001	20190605	매일경제	박소현	현대차, 2학기 자유학기제 참여 중학교 모집...14일까지	경제>자
18	01500801.20190605081917001	20190605	영남일보	/박종문기자 kpjm@yeongnam.com	영남대에 13억원 상당 소프트웨어 기증	IT_과학>
19	01100401.20190605030212001	20190605	동아일보		바이오헬스 강국, 국민이 만든다[기고/박능후]	경제>산업
20	02100501.20190605173315002	20190605	파이낸셜뉴스	구자윤	[인터뷰] 박영태 캠퍼스 대표 "3대 신산업 선정 '전기차'... 규제 완화 있어야 미래 있다"	경제>산업

1 과제

1. 데이터구조 및 과제1

빅카인즈에서 신문기사 다운로드 받기

- 다운로드 된 엑셀파일은 약 3,300여건의 '미래자동차' 관련 뉴스임 (만약 컴퓨터 사양이 낮다면, 기사수를 300건 내외로 줄여서 분석)
- '본문'컬럼의 기사 내용이 주요 분석 대상이 됨. 또한 일자별 추세를 분석을 위해서 '일자' 컬럼도 필요

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	뉴스 식별	일자	언론사	기고자	제목	통합 분류	통합 분류	통합 분류	사건/사고	사건/사고	사건/사고	인물	위치	기관	키워드	특성추출	본문	JURL	분석제외 여부	
2	02100101	20190605	매일경제	연론사	문 대통령 정치>외교 정치>청와대							정치해,문경남	창원컨벤션	미세먼지,미세먼지,미세먼지,미세먼지	창원컨벤션	창원시,창원시,창원시,창원시	창원시,창원시,창원시,창원시	http://news.mk.co.kr/newsRead.p		
3	02100801	20190605	아시아경제	손선희	문 대통령 '미분류'							이이 문,허 문,창원,경	창원컨벤션	창원시,창원시,창원시,창원시	창원시,창원시,창원시,창원시	창원시,창원시,창원시,창원시	창원시,창원시,창원시,창원시	http://www.asiae.co.kr/article/2019		
4	01100801	20190605	조선일보	장윤서 기	제약바이오경제>산업 IT_과학>경제>자원							오제세,원,아시아,캐	구글,보건,제약산업,제약산업,제약산업	제약산업,제약산업,제약산업,제약산업	제약산업,제약산업,제약산업,제약산업	제약산업,제약산업,제약산업,제약산업	제약산업,제약산업,제약산업,제약산업	http://biz.chosun.com/site/data/		
5	07101201	20190605	디지털타임	김양혁	현대차, '미사회'교육 시험														http://www.dt.co.kr/contents.htm	
6	02100801	20190605	아시아경제	이은결	[단독]눈물 경제>산업 경제>무인 경제>취업_창업								울산,군산,기업진출	눈물,완화,중소기업,미세먼지	눈물,완화,중소기업,미세먼지	눈물,완화,중소기업,미세먼지	눈물,완화,중소기업,미세먼지	http://www.asiae.co.kr/article/2019		
7	02100801	20190605	아시아경제	박형수	[특징주]두 IT_과학>경제>자동차														http://www.asiae.co.kr/article/2019	
8	02100701	20190605	헤럴드경제	정찬수 기	중학생 진:사회>교육_시험														http://biz.heraldcorp.com/view.p	
9	02100311	20190605	서울경제	이재용 기	현대차, 자 경제>자동차 사회>교육 경제>산업_기업									현대모터스	한국잡월드	현대차, 자 자동차, 현대차 자동차	현대차, 자 자동차, 현대차 자동차	현대차, 자 자동차, 현대차 자동차	http://www.sedaily.com/NewsVie	
10	01101001	20190605	한겨레	김진철	현대차, 자 경제>자동차 사회>교육 경제>산업_기업									현대모터스	한국잡월드	현대차, 자 자동차, 현대차 자동차	현대차, 자 자동차, 현대차 자동차	현대차, 자 자동차, 현대차 자동차	http://www.hani.co.kr/arti/econo	
11	02100801	20190605	아시아경제	유현석	두울, 생체 IT_과학>고 경제>산업 경제>자동차														http://www.asiae.co.kr/article/2019	
12	07100501	20190605	전자신문	박태준	현대차, 중 사회>교육_시험														http://www.etnews.com/2019060	
13	02100101	20190605	매일경제	배준희	[카드뉴스] 경제>반도체 경제>산업 IT_과학>과학									문,수혜주	평화산업,수혜주,수혜주,수혜주	평화산업,수혜주,수혜주,수혜주	평화산업,수혜주,수혜주,수혜주	http://new.예외		
14	01100401	20190605	동아일보	정진수	현대차, 중 사회>교육_시험									현대모터스	한국잡월드	현대차, 자 자동차, 현대차 자동차	현대차, 자 자동차, 현대차 자동차	현대차, 자 자동차, 현대차 자동차	http://news.donga.com/3/all/201	
15	02100701	20190605	헤럴드경제	정찬수 기	현대차, 중 사회>교육_시험														http://biz.heraldcorp.com/view.p	
16	02100851	20190605	아주경제	김해원	현대차, 자 사회>교육_시험									현대차	한국잡월드	현대차, 자 자동차, 현대차 자동차	현대차, 자 자동차, 현대차 자동차	현대차, 자 자동차, 현대차 자동차	http://www.ajunews.com/view/20	
17	02100801	20190605	아시아경제	김지희	현대차, 자 경제>자동차 사회>교육 경제>산업_기업														http://www.asiae.co.kr/article/2019	
18	02100101	20190605	매일경제	박소현	현대차, 2차 경제>자동차 사회>교육 경제>산업_기업									현대모터스	한국잡월드	현대차, 자 자동차, 현대차 자동차	현대차, 자 자동차, 현대차 자동차	현대차, 자 자동차, 현대차 자동차	http://news.mk.co.kr/newsRead.p	
19	01500801	20190605	영남일보	박종문 기	영남대에 'IT_과학>고 IT_과학>보안									이,서길수,	한국MS,한 영남대,13'영남대,자 한국MSC	한국MS,한 영남대,13'영남대,자 한국MSC	한국MS,한 영남대,13'영남대,자 한국MSC	한국MS,한 영남대,13'영남대,자 한국MSC		
20	01100401	20190605	동아일보	구자윤	바이오헬스 경제>산업 경제>취업 경제>자원									박능후	보건복지부 바이오헬스 바이오헬스 병에 걸리	보건복지부 바이오헬스 바이오헬스 병에 걸리	보건복지부 바이오헬스 바이오헬스 병에 걸리	http://news.donga.com/3/all/201		
21	02100501	20190605	파이낸셜뉴스	구자윤	[인터뷰] 북경제>산업 경제>자동차 IT_과학>IT_과학일반									박	중국,송도,쌍용자동차 전기차,박 전기차,자 카메라 모	중국,송도,쌍용자동차 전기차,박 전기차,자 카메라 모	중국,송도,쌍용자동차 전기차,박 전기차,자 카메라 모	http://www.fnnews.com/news/20		
22	01500301	20190604	경상일보	이우사 기	"스마트공: 경제>산업 지역>울산									임재홍,김	경제부시중 담회,(주)스마트공 자동차,울산 산지역	경제부시중 담회,(주)스마트공 자동차,울산 산지역	경제부시중 담회,(주)스마트공 자동차,울산 산지역	www.ksilbo.co.kr/news/articleVie		
23	01400551	20190604	충북일보	최대만	오제세, '47 경제>산업 경제>자원									오제세,오	청주	더불어민주 오제세,4차4차,산업혁 [충북일보	더불어민주 오제세,4차4차,산업혁 [충북일보	더불어민주 오제세,4차4차,산업혁 [충북일보	http://www.inews365.com/news/	

1 과제

2. 과제

과제1(배점 20점)

월별 '미래자동차'의 기사건수를 제시하시오

단어의 정제과정을 거쳐 wordcloud를 수행하시오

아웃풋 제시

1 과제

2. 과제

과제2(배점 20점)

'미래자동차'의 상위 20개 단어만 선별하여 이 단어와 같이 언급된 단어간 네트워크를 작성하시오

아웃풋 제시

1 과제

2. 과제

과제3(배점 20점)

Clustering 기법을 통해 하위 기사를 군집화 하고, 묶은 단어/내용을 파악하여 군집별 주제명을 정하십시오

아웃풋 제시

1 과제

2. 과제

과제4(배점 20점)

LDA 기법을 적용하여 하위 Topic을 찾아 네이밍을 하시오

아웃풋 제시

1 과제

2. 과제

과제5(배점 20점)

Topic별로 월별 추세를 파악해보고 어떤 topic이 최근 주목받는지 파악해보시오

아웃풋 제시



수고하셨습니다!

W와이즈인컴퍼니