|프로젝트 주제

● 데이터 시각화 공모전 참여를 통한 기업 트렌드 시각화

DACON 구직자를 위한 기업 트랜드 시각화 경진대회 (https://dacon.io/competitions/official/235866/overview/description)

| 프로젝트 목적 및 필요성

- 일반적으로 구직자들의 입장에서 좋은 기업이란 일정 수준 이상의 임금과 근로조건, 복지와 안전 수준이 보장되는 기업이라고 할 수 있다. 또한 해당 기업이 현재 활발하게 고용과 계약이 진행되고 있는 일명 '블루 오션(Blue Ocean)' 시장인지, 혹은 고용 포화 상태인 '레드 오션(Red Ocean)' 시장인지의 여부도 구직자의 입장에서는 매우 중요하다. 마지막으로, 기업의 성장 가능성 역시 구직자들의 회사 선택에 영향을 미치는 요인 중 하나이므로, 구직 시장의 트렌드를 파악할 수 있도록 만드는 것도 중요하다.
- 따라서 본 프로젝트에서는 크게 '구직 트렌드', '고용 현황' 두가지의 기준을 바탕으로 구직자들을 위한 기업 트렌드 대시보드를 제작하고자 한다.
- 위 두가지 요인을 바탕으로 산업 혹은 직종별로 비교하며, 구직자 본인이 취업하고자 하는 분야가 어느 정도의 안정성을 보이고 있는지 판단할 수 있도록 한다.
- 이미 구직 시장에 익숙하고 이직을 고려하고 있는 구직자들보다는, 구직 시장에 발을 들인지 얼마되지 않은 2030 사회 초년생 구직자들에게 도움을 줄 수 있는 대시보드를 제작할 필요가 있다. 또한, 발빠르게 변화하는 구직시장의 정보를 제공하기 위해 되도록 최근의 데이터를 바탕으로 대시보드를 구성하려 한다.

| 프로젝트 주요내용

● '기업 트렌드'라는 큰 주제 내에서 다섯가지의 상이한 하위 주제를 바탕으로 대시보드를 구성하는 것이므로, 각 주제에 알맞은 데이터들을 수집하여 필요한 정보만으로 구성되도록 정제하고, 효율적인 정보 전달을 가능하게 하는 시각화를 완성해낸다.

○ 구직 트렌드

- ✓ 취업 및 구인 사이트 데이터를 통해 가장 많이 언급되는 구직 키워드가 무엇인지 파악하고,
- ✓ 높은 비율로 요구되는 채용 조건이 무엇인지 파악하여
- ✓ 구직 시장 트렌드 및 구직자들의 관심사를 알아본다.

○ 고용 현황

- ✓ 빈 일자리, 채용 계획 인원 등을 통해 앞으로 고용이 활발할 것으로 예측되는 산업을 파악하고,
- ✓ 이와 더불어 최근 늘어나고 있는 2030 기간제 근로(비정규직) 비율을 고려하여, 기간제 근로 시장의 현황 역시 함께 살펴본다.

| 기대효과

- 다양한 데이터셋을 활용하며 기업 트렌드를 분석하고, 시각화해보며, 기업 트렌드 분석에 있어 다양한 방면으로 활용될 수 있는 기반을 마련한다.
- 데이터를 수집하고, 정제 및 가공, 시각화하는 전 과정을 상술하여, 초보자도 데이터 분석의 전반적인 과정을 실습해볼 수 있는 가이드 매뉴얼을 구성한다.
- 학생들이 수집된 데이터를 활용할 수 있는 Hands-On 실습 예제를 개발하여, 데이터 활용의 Best Practice 를 제공하고, 데이터를 실전적으로 분석 및 시각화할 수 있는 능력을 배양한다.

| 제작 기간

2021.12.20(월) - 2022.01.31(월)

| 대시보드 OVERVIEW





| 수집 데이터 정보 요약

- 구직 트렌드 데이터
 - O 잡코리아 채용 정보
 (https://drive.google.com/drive/folders/1BGSOtRSE0D8O4cwRROALdam8LzehRUvh?usp=s haring)
- 고용 시장 데이터
 - O 산업별 고용 현황 (현원, 구인 인원, 채용 계획 인원 등 포함)
 (https://docs.google.com/spreadsheets/d/1_kkWChKkSBIs40DoOKLaw-fbWHTXEMGA/edit?usp=sharing&ouid=111765825620961795095&rtpof=true&sd=true)
 - O 산업별 빈 일자리
 (https://docs.google.com/spreadsheets/d/1K_J3xNKS2lixvK7eDQGpJGDeZQ-X1Q1U/edit?usp=sharing&ouid=111765825620961795095&rtpof=true&sd=true)
 - O 산업별 기간제 근로자 수 (https://docs.google.com/spreadsheets/d/1uil-9LJA5TXTaQNK2yoKduAV2zd67RmU/edit?usp=sharing&ouid=111765825620961795095&r tpof=true&sd=true)
 - O 기간제 근로 계약 만료 후 조치 형태
 (https://docs.google.com/spreadsheets/d/11ekwK3ILQ9afptCP5C_jdYLGMM9Uofl/edit?usp=sharing&ouid=111765825620961795095&rtpof=true
 &sd=true)

|[구직 트렌드] 분석 및 시각화 시나리오 #1

- ✓ 현재 구직자들이 잡코리아 구직 사이트에서 가장 많은 관심을 보이는 채용 공고는 무엇인가요?
- ✓ 해당 채용 공고는 어떤 산업, 직종군의 채용공고이며, 경력 조건, 학력 제한 등, 채용 조건은 어떠한가요?
- ✓ 또한, 해당 채용 공고를 올린 기업은 어떤 기업인가요?
- ✓ 각 채용공고의 제목과 기업명, 그리고 구직자들의 관심도를 파악할 수 있는 지표는 한눈에 볼 수 있도록 텍스트로 구성해 시각화해주세요.



|[구직 트렌드] 분석 및 시각화 시나리오 #2

- ✓ 채용공고에 기재된 여러 조건 (경력 조건, 학력 제한 등)별로 등록된 채용공고가 전체 공고에서 어느 정도의 비율을 차지하고 있는지 알 수 있을까요?
- ✓ 경력 조건, 학력 제한 항목들이 전체 채용 조건에서 차지하는 비율을 확인할 수 있도록 파이차트 혹은 도넛차트로 시각화해주세요.



경기도 일자리 잡코리아채용정보

데이터 수집 및 획득

경기지역경제포털 (https://bigdata-region.kr/#/dataset?datecolumn=created&page=1)

1. 데이터 다운로드 시 회원가입 및 로그인 필요



지역경제 데이터를 한곳에서 확인해 보세요!

방대한 데이터를 쉽고 빠르게 찾아보세요!								Q	
인기검색어: 지역 화폐 정보 기업 일자리									
::	C				77=		0	₩	FREE
전체	지역화폐	기업	지역트랜드	취업	지역현황	기업신용보증	영상	유료 데이터	무료 데이터
			ID PASSWO		역경제포팀	4			

2. 해당 분석에서 이용된 데이터셋은 경기지역경제포털의 '취업' 카테고리에 해당하는 데이터



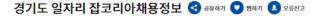
지역경제 데이터를 한곳에서 확인해 보세요!



3. 경기지역경제포털의 '경기도 일자리 잡코리아채용정보'

(https://bigdata-region.kr/#/dataset/d90b5384-6df8-49f4-999e-9172bb50a0b2)

데이터 '이용 신청' 후 '다운로드' 가능



경기도 일자리 관련하여 수집한 일자리 청년취업 취업 채용관련 정보





(활용목적은 어떤 항목을 선택해도 무관)



4. 제공된 데이터셋은 zip 파일 형태로 제공되므로, 압축 해제 후 사용 가능



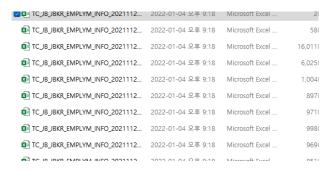
데이터셋 탐색 및 이해

데이터 원본 열람

zip 파일 열람 결과 18개의 엑셀 파일로 제공되어 있음

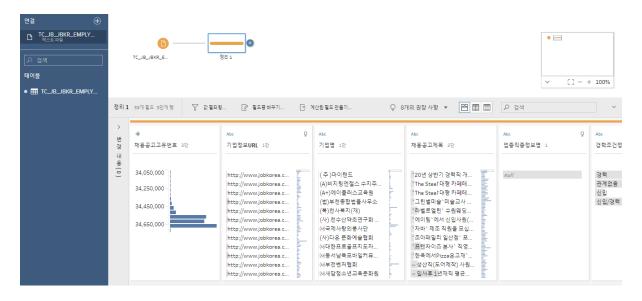
데이터 수집 기점에 따라 여러 개의 파일로 분할되어 제공하고 있음

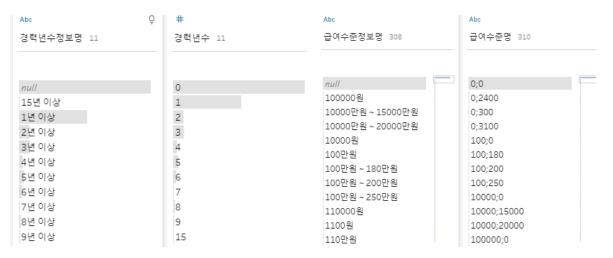
각각의 엑셀 파일이 지닌 칼럼(=column, 열, 필드)은 동일하며, 열 개수는 32 개, 행 개수는 각 파일별로 상이함



■ 필드 설명

해당 데이터 중 임의로 선정한 엑셀 파일 하나를 태블로 프렙에 올리고, 정리 단계를 추가하여 데이터가 지닌 필드 살펴보기

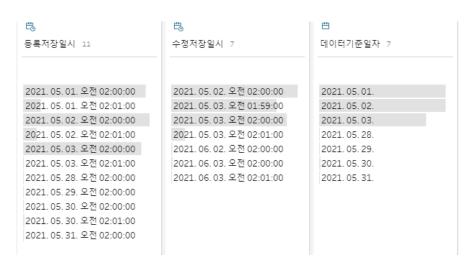




경력년수, 경력년수 정보명과 급여수준명, 급여수준정보명 필드는 같은 정보를 담고 있으며, 표기 방법만 다름



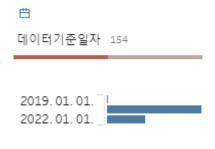
근무형태명, 해당채용공고키워드명, 채용직급직책명, 근무지역명과 같이 여러 개의 값이 들어갈 수 있는 필드는 ','를 구분자로 하여 하나의 행에 담김



일자 유형의 필드로는 등록저장일시, 수정저장일시, 데이터기준일자 필드가 있음

해당 데이터셋은 '채용공고고유번호' 필드를 id 필드로 하며, 데이터셋의 한 행은 채용공고 한 건의 정보를 담고 있음

데이터 기준 일자 탐색



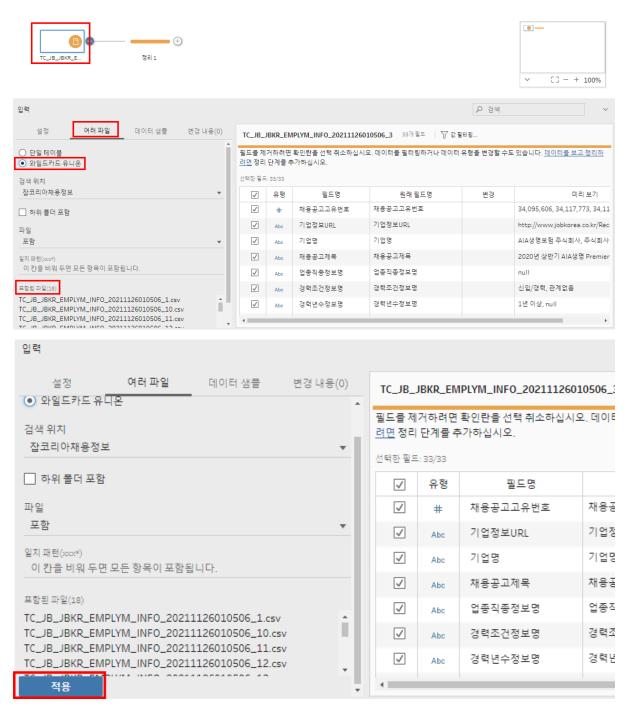
2019.01.01 ~ 2022.01.01

데이터셋 정제 기획

- 분석 시나리오와의 연계 이슈
 - O '업종직종정보명', '학력제한정보명', 경력조건정보명' 등, 잡코리아에 업로드된 채용 공고 세부 사항에 대한 정보를 담고 있으므로 해당 분석 시나리오에 부합함
 - O 구직자들의 관심도를 파악하기 위해서 필요한 채용공고의 '조회수'에 대한 정보를 담고 있음
 - O '경력년수', '경력년수정보명', 혹은 '급여수준명', '급여수준정보명' 등의 필드는 다른 이름을 가지고 있으나 유사한 정보를 가리키고 있으므로 둘 중 하나를 정제할 필요가 있음
 - O '등록저장일시', '수정저장일시', '사용여부', '서류접수마감일자', '채용공고등록일자', '잡코리아채용공고 QR 코드내용' 칼럼은 해당 시나리오에 필요한 정보가 아니므로 제거
 - O '연봉면접결정여부', '채용직급직책명', '모집인원명', '근무형태명' 등의 필드는 해당 분석 시나리오에 부합하는 필드일수도 있으나, 본 시각화는 구직자들의 입장에서 중요하다고 판단되는 정도에 따라 핵심 정보만을 선정해 효율적으로 전달하는 것으로 목표로 하고 있으므로 제외하였음
- 데이터 구조 및 결합 이슈
 - ㅇ 동일한 필드를 가진 18개의 테이블을 유니온하여, 하나의 테이블로 결합할 필요가 있음
- 데이터 클린징 이슈
 - O '공유수' 필드 값은 모두 0 이고, '상세내용' 필드 값은 모두 null 이므로 제거

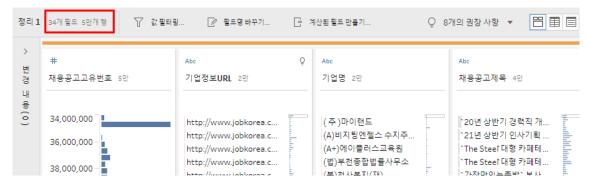
데이터셋 정제 실행

- 데이터 정제 방향
 - 1. 와일드카드 유니온을 통해 18개의 엑셀 파일을 유니온해 하나의 테이블로 만듦
 - 해당 분석 시나리오에 부합하는 필드들만 남기고 다른 필드들을 제거,
 서로 중복되는 정보를 담은 필드를 정제



태블로 프렙에 올린 테이블을 선택한 뒤, '여러 파일' → '와일드카드 유니온' → '적용'을 선택해당 테이블과 동일한 폴더에 담겨있는 18개의 엑셀파일이 유니온됨

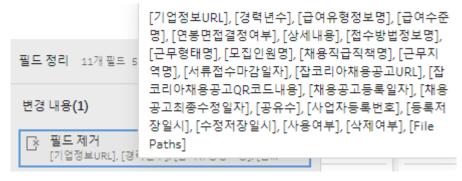




다시 정리 단계로 돌아와 열과 행 개수를 살펴보면, 34 개의 필드와 5 만개 행이 확인됨



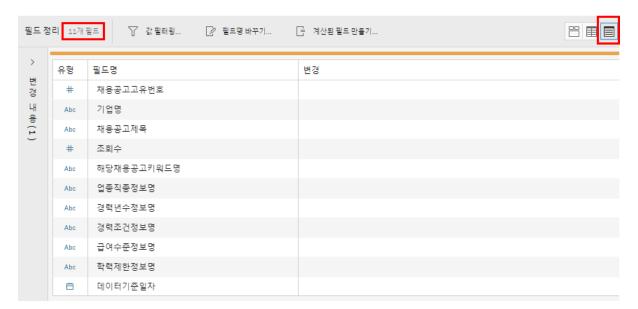
제거하고자 하는 필드들을 선택한 뒤, 제거



제거한 필드 목록은 다음과 같음

정제 결과 해당 데이터셋의 필드





정제 결과 해당 시나리오에 부합하는 11 개의 필드만 남겨짐



대시보드 시각화

1) 시나리오#1



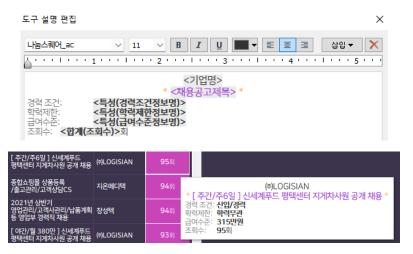
1. '채용공고제목', '기업명' 필드를 나란히 행 선반에 올림



2. '채용공고제목' 필드를 필터 선반에 올리고, '조회수' 필드 기중 상위 100 개로 필터링

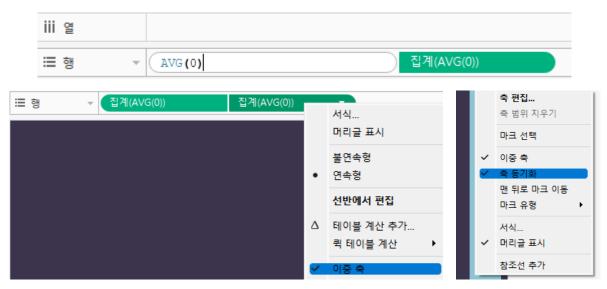


3. 마크 유형을 사각형으로, '조회수'를 색상과 레이블로 지정해줌



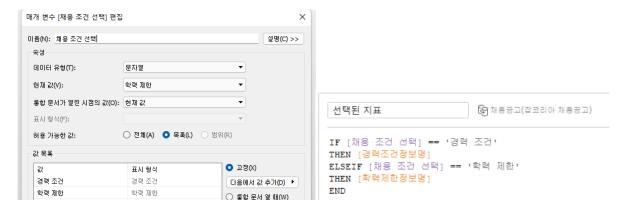
4. 도구설명에는 다음과 같은 필드를 포함

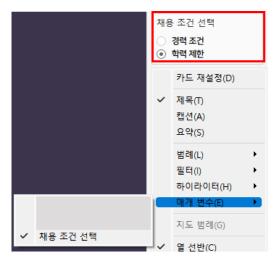
2) 시나리오#2



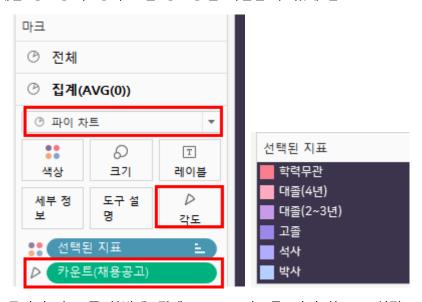
1. 행 패널에 'AVG(0)' 계산된 필드 두개를 올리고, 두번째 계산된 필드를 이중축 설정
→ 축 동기화

가운데 축을 중심으로 하는 도넛 차트 생성





2. 매개변수 '채용 조건 선택'과 계산된 필드 '선택된 지표'를 생성해 매개변수 선택에 따라 '학력 제한 정보명'과 '경력 조건 정보명'을 확인할 수 있게 함



3. 생성된 두가지 마크 중 첫번째 '집계(AVG(0))' 마크를 '파이 차트'로 설정,

각 구직 조건에 따른 채용공고의 수를 알고 싶은 것이므로, '채용공고(카운트)' 필드를 각도로 지정

색상으로는 '선택된 지표' 필드를 지정



4. 레이블로는 '선택된 지표' 필드와 '카운트(채용공고)'필드를 구성비율로 퀵 테이블 계산하여 지정



5. 두번째 '집계(AVG(0))' 마크 유형은 원으로 설정하고, 선택한 채용 조건을 표시하기 위해 '채용 조건 선택' 필드를 레이블로 지정, 가운데 맞춤으로 표시

|[구직 트렌드] 분석 및 시각화 시나리오 #3

✓ 채용공고에 기재된 채용 업종·직종 정보 빈도수를 확인하여 가장 활발하게 고용이 진행되고 있는 업종·직종이 무엇인지 파악할 수 있을까요?



|[구직 트렌드] 분석 및 시각화 시나리오 #4

- ✓ 채용공고별로 가지고 있는 키워드를 추출하여 가장 많이 언급된 키워드가 무엇인지 확인할 수 있을까요?
- ✓ 구직 시장에서 가장 많이 언급되는 키워드는 무엇인가요? 한눈에 확인할 수 있도록 워드클라우드로 시각화해주세요.



데이터셋 정제 기획

● 분석 시나리오와의 연계 이슈



- O 현재 확보한 잡코리아 채용 정보 데이터셋에 시각화하고자 하는 '업종직종정보명', '해당채용공고키워드명' 필드가 존재함
- 데이터 클린징 이슈
 - O 그러나 두 필드 모두 여러 내용이 ';'를 기준으로 한 행으로 묶여 있기 때문에 어떤 업종·직종명 혹은 키워드가 많이 언급되고 있는지를 보다 상세하게 파악하기 위해서는 별도의 정제 과정이 필요할 것으로 보임

데이터셋 정제 실행

파이썬을 활용하여 '채용공고정보명' 칼럼과 '해당채용공고키워드명' 칼럼 문자열 분할하여 별도의 xlsx 파일로 저장

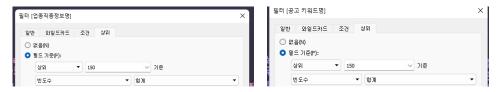
잡코리아 업종직종정보명 빈도수 정제 결과 데이터셋

(https://docs.google.com/spreadsheets/d/1q2DeKKtOE-M9mY0PwK1CzrpYly5hJnWJ/edit?usp=sharing&ouid=111765825620961795095&rtpof=true&sd=true)

잡코리아 공고 키워드명 빈도수 정제 결과 데이터셋

(https://docs.google.com/spreadsheets/d/1ODZTPGy_voFLszTDtLoQ-xhs4ZV9vsl-/edit?usp=sharing&ouid=111765825620961795095&rtpof=true&sd=true)

대시보드 시각화



1. '업종 직종 정보명' 또는 '공고 키워드명' 필드를 필터 선반에 올린 뒤, '빈도수' 필드를 기준으로 상위 150개만 필터링해 시각화

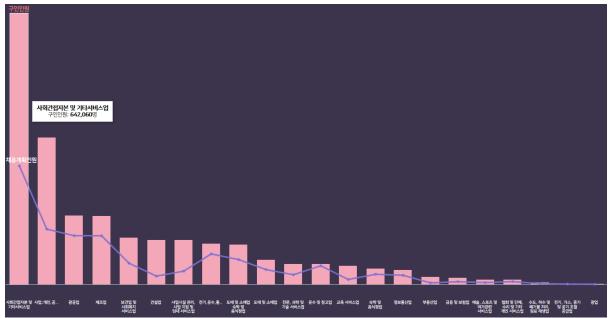


2. 마크 유형을 텍스트로 지정하고, '업종직종정보명'과 '공고 키워드명' 필드를 텍스트로 표시, '빈도수' 필드를 색상과 크기로 지정해 워드 클라우드 생성

|[고용시장 현황] 분석 및 시각화 시나리오 #1

- ✓ 산업별로 고용 현황을 파악하고 서로 비교할 수 있나요?
- ✓ 현원, 구인인원, 채용인원, 부족인원이 얼마나 되는지 파악하여 각 산업의 고용 시장 파이를 확인할 수 있을까요?
- ✓ 앞서 파악한 현원, 구인인원, 채용인원, 부족인원과 채용 계획 인원을 비교하여 앞으로의 고용 전망을 예측할 수 있을까요?
- ✓ 추가적으로 미충원인원에 대한 정보를 통해 해당 산업 고용 시장의 고용률이 어느정도인지 파악할 수 있을까요?





산업별·규모별 현원, 구인·채용·미충원·부족·채용계획

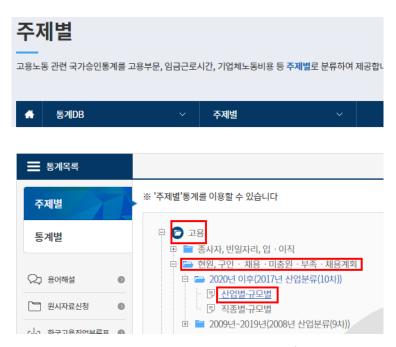
데이터 수집 및 획득

고용노동통계 (http://laborstat.moel.go.kr/hmp/index.do)

고용노동통계는 로그인하지 않아도 데이터 사용 가능



본 데이터 분석에 활용된 고용노동통계의 데이터셋은 '통계 DB' → '주제별' 카테고리에서 제공되는 데이터셋을 활용하였음



'고용' 주제의 '산업별·규모별 현원, 구인·채용·미충원·부족·채용계획'

¹⁾ 산업별·규	모별				
자료갱신일: 2	021-12-29 / 수록기간:	반기 2020 1/2 ~ 2021 2/	/2 / 자료문의처 : 044-202	2-7252	
일괄설정 +	항목[7/7]	시도별(17개)[1/…	규모별[1/8]	산업분류별[1/93]	시점[1/4] 🔺 🔽
현재까지 선택된	데이터는 항목(7) × 분위	류(1+1+1) × 시점(1)=7 셀	[20,000 (조회) 200,0	00 (다운로드) 다운로드	불가]
▶ 시도별(177	ㅐ)[□전체레벨 1]				
□ 1 레벨 전체	세선택				
☑ 🛅 전국					
☐ <u></u>					
□ ■ 부산					
대구					
¹⁾ 산업별·규	口用				
GBSII					
자료갱신일: 2	021-12-29 / 수록기간:	반기 2020 1/2 ~ 2021 2/3	2 / 자료문의처 : 044-202-	7252	
일괄설정 +	항목[7/7]	시도별(17개)[1/18]	규모별[1/8]	산업분류별[1/93]	시점[1/4] 🔺 🔽
현재까지 선택된	데이터는 항목(7) × 분류	류(1*1*1) × 시점(1)=7 셀『	20,000 (조회) 200,000) (다운로드) 다운로드 불	가 1
▶ 규모별 [_	전체레벨 1]				
□ 1 레벨 전체	세선택				
☑ ■ 전규모(5인이상)					
□ <u>1</u> 10인이상					
□ <u>↑</u> 중소규모(300인 미만) □ <u>↑</u> 1규모(5~9인)					
	±(5~351) ⊒(10~360])				

해당 분석 시나리오는 지역별, 규모별로 비교하는 것을 목표로 하고 있지 않으므로, 시도별 항목은 '전국' 항목만 선택하고, 규모별은 그대로 '전규모' 항목만 선택



산업분류에 따른 비교를 하는 것이 목적이므로, 산업분류 항목은 1 레벨 중 집계된 산업분류를 제외하고, 개별의 산업분류 17 개만을 선택

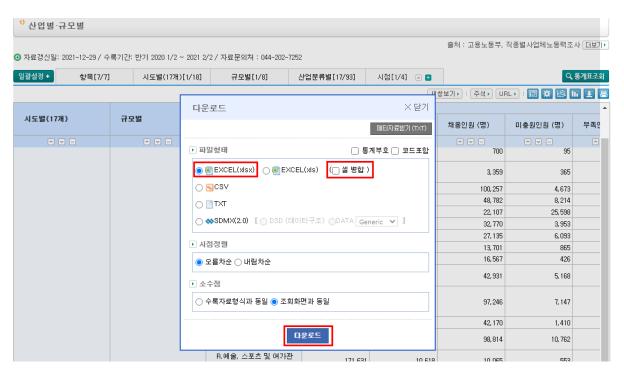
각 산업의 하위 산업에 따른 비교까지 하는 것을 목표로 하지 않으므로 1 레벨만 선택하도록 함



다운로드 받을 칼럼을 모두 선택했으면 '통계표조회'를 선택하여 결과 테이블 확인



다운로드 버튼 클릭



엑셀 파일 형태를 선택하고,

현재 '시도별', '규모별' 칼럼이 병합되어 있으므로 '셀 병합' 체크박스를 해제한 후 '다운로드'

데이터셋 탐색 및 이해

데이터 원본 열람

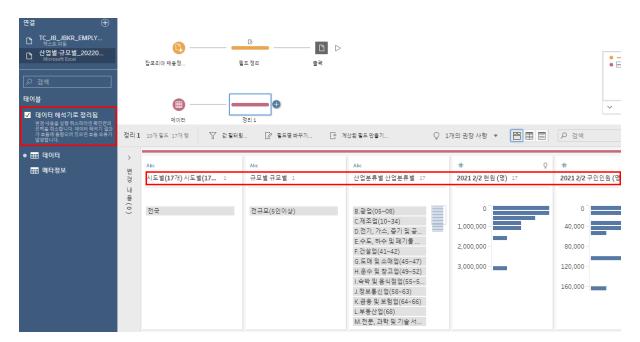


태블로 프렙에 다운로드받은 산업별·규모별 고용현황 테이블을 올린 후, '데이터' 테이블을 올려살펴봄





원본 테이블을 그대로 올릴 경우 필드명이 제대로 표시되지 않으므로, '데이터 해석기 사용'을 체크하여 데이터 해석기 기능을 사용해 다시 불러오기



정리 단계를 추가해 데이터 해석기로 정리한 결과를 살펴보면 필드명이 불러와진 것을 확인할 수 있음

테이블 열 개수: 10 개

테이블 행 개수: 17 개

■ 필드 설명

- 차원 필드로는 산업분류별, 규모별, 시도별 필드가 있음 산업분류별: 한국표준산업분류 10차 산업분류

규모: 산업 종사자 수 규모

- 측정값 필드는 부족률(%) 필드를 제외하고 모두 명수로 기재되어 있음

부족률(%): 부족 인원 / (현원 + 부족인원) x 100

- 해당 데이터셋은 '산업분류별' 필드를 id 필드로 하여, 데이터셋의 한 행은 산업분류별고용 정보를 담고 있음

데이터 기준 일자 탐색



2021 년 하반기 (2021.07~2021.12)

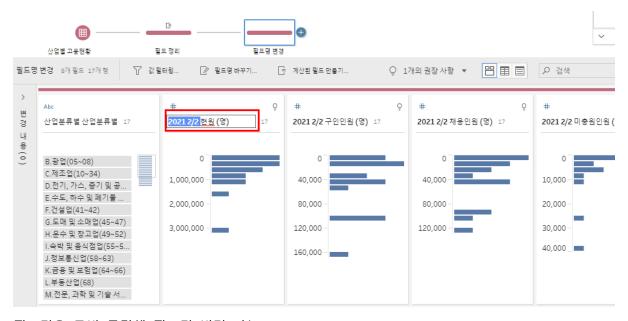
데이터셋 정제 기획

- 분석 시나리오와의 연계 이슈
 - O 해당 시나리오에서 제공하고자 하는 산업별 고용 현황에 대한 여러 요소를 칼럼으로 가지고 있으므로 적절한 연계성을 지님
 - 지역과 규모별로 비교하지 않으므로 '시도별(17개)', '규모별' 필드 제거
- 데이터 구조 및 결합 이슈
 - O 다른 테이블과의 결합 없이 바로 활용 가능
- 데이터 클린징 이슈
 - O 데이터 해석기로 불러온 결과 테이블의 필드명을 정리해야 함
 - O 산업분류 필드 각 행에 존재하는 알파벳과 괄호 뒤 숫자를 제거해야 함

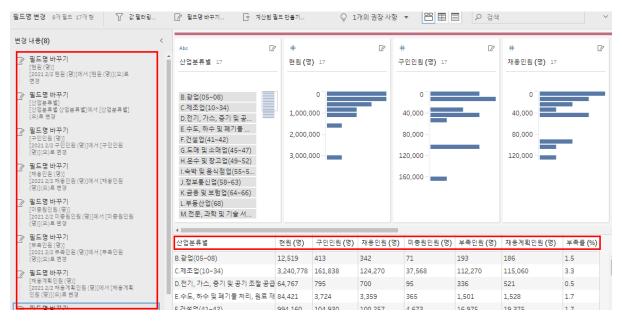
- 데이터 정제 방향
 - 1. 데이터 해석기를 통해 불러와진 필드명을 원하는 이름으로 수정
 - 2. '시도별', '규모별' 필드 제거
 - 3. '산업분류별' 필드 내용 분할



'시도별' 필드와 '규모별' 필드 제거



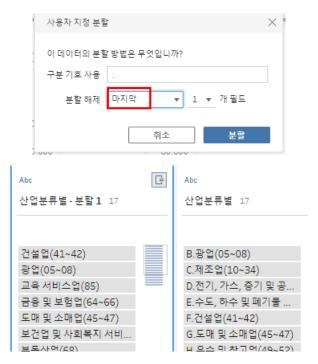
필드명을 두번 클릭해 필드명 변경 가능



필드명 변경 완료



'산업분류별' 필드 -> '값 분할' → '사용자 지정 분할' 선택



': 뒷부분의 내용을 분할하여 알파벳 제거



'(' 앞부분의 내용을 분할하여 산업분류별 필드 분할 완료



분할 과정 중 생성된 두 필드를 제거하고, 최종 결과 필드명을 '산업분류별'로 변경



정제 결과 해당 데이터셋의 필드

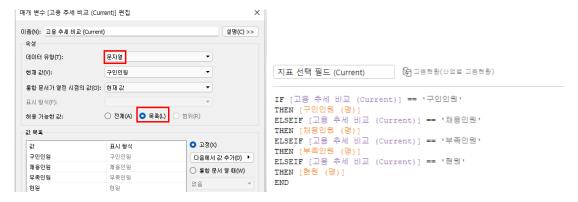
(https://docs.google.com/spreadsheets/d/16_hXdKVixUv-SoLalQo5jRi9-HM74li-/edit?usp=sharing&ouid=111765825620961795095&rtpof=true&sd=true)

유형	필드명
Abc	산업분류별
#	현원(명)
#	구인인원(명)
#	채용인원(명)
#	미층원인원(명)
#	부족인원(명)
#	채용계획인원(명)
#	부족률 (%)

대시보드 시각화

현원, 구인인원, 채용인원, 부족인원 필드를 통해 현재의 고용 시장 상황을 파악하고,

위 필드들과 채용계획인원, 미충원인원 필드와의 비교를 통해 앞으로의 고용시장을 전망해볼 예정



1. '구인인원', '채용인원', '부족인원', '현원' 필드를 목록 형태로 담은 '고용 추세 비교 (Current)' 매개변수와 '지표 선택 필드 (Current)' 선택된 필드를 생성



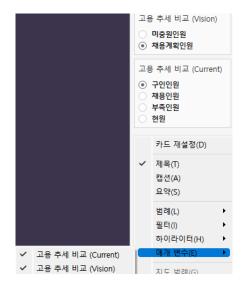
2. '미충원인원', '채용계획인원' 필드를 담은 '고용 추세 비교 (Vision)' 매개변수와 '지표 선택 필드 (Vision)' 계산된 필드를 생성



3. '산업분류' 필드를 열 패널에, 앞서 만든 계산된 필드 2개를 행 선반에 올리고, '지표 선택 필드 (Vision)' 필드를 이중축으로 설정 → 축 동기화



4. 두 필드의 마크 유형을 서로 다르게 해 한 화면에서 비교할 수 있게 만듦 위 화면에서는 '지표 선택 필드 (Current)'를 막대 유형으로, '지표 선택 필드 (Vision)'를 라인 유형으로 설정



5. 매개변수를 추가해 각각의 매개변수 별로 항목을 선택할 수 있도록 설정

| [고용시장 현황] 분석 및 시각화 시나리오 #2

- ✓ 산업별로 빈 일자리 현황을 파악하고 싶습니다.
- ✓ 빈일자리는 객관적 지표 (인원 수 등)와 상대적 지표 (백분위수, 비율 등)를 모두 파악할 수 있나요?
- ✓ 텍스트 패널로 시각화해 필터 결과에 따라 여러 정보를 한번에 파악할 수 있도록 해주세요.



산업/규모별 고용

고용 노동통계의 주제별 통계에서 데이터셋 다운로드



'고용' 주제의 '산업별·규모별 종사자, 빈일자리, 입·이직'



빈일자리 현황을 파악하고 싶으므로 전체 항목 중 빈일자리 관련 항목만 선택해 수집 상용/임시일용으로 구분하여 비교할 것이 아니기 때문에 '전체'에 해당하는 항목만 수집



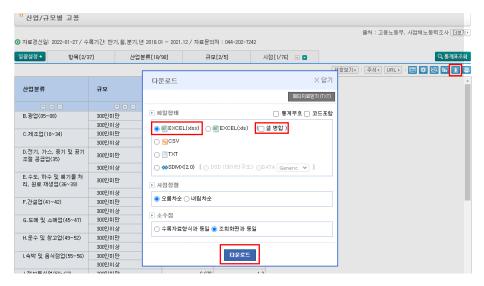
개별 산업분류별로 비교하기 위해 '전체' 항목 체크를 해제



개별 산업 규모별로 비교하기 위해 '전체' 항목 체크 해제



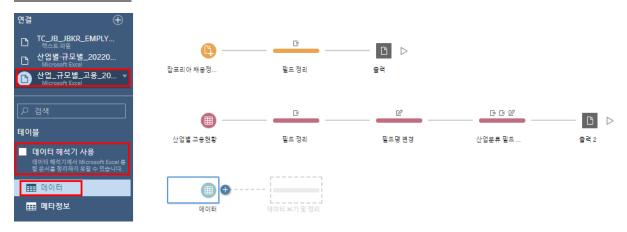
시점별로 비교하는 것이 아니므로 가장 최근의 시점만 선택한 뒤, '통계표조회' 선택



셀 병합을 해제한 뒤 '다운로드'

데이터셋 탐색 및 이해

데이터 원본 열람



태블로 프렙에 다운로드 받은 데이터 원본을 올린 뒤, '데이터 해석기 사용' 선택



정리 단계를 추가해 데이터셋 살펴보기

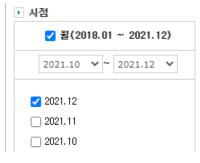
테이블 열 개수: 4개

테이블 행 개수: 36 개

■ 필드 설명

- 산업분류: 한국 표준 산업분류 10 차 산업 대분류
- 규모: 산업 종사자 수 규모
- 빈일자리_전체 (명): 구인 활동을 하고 있으며, 한 달 이내에 일이 시작될 수 있는 일자리 수
- 빈일자리율_전체(%): 결원율이라고도 하며, 산출식은 다음과 같음, 빈 일자리율 = 빈 일자리 수 / (빈 일자리 수 + 근로자 수) x 100
- 해당 데이터셋은 '산업분류'와 '규모' 필드를 함께 id 필드로 하여, 산업분류, 규모별 빈일자리 현황 정보를 담고 있음

데이터 기준 일자 탐색



2021 년 12 월

데이터셋 정제 기획

- 분석 시나리오와의 연계 이슈
 - O 산업분류 필드가 존재하므로 시나리오대로 산업과 그 규모에 따른 빈 일자리 현황 파악 가능
 - O '규모별' 필드를 추가로 활용하여 보다 세부적인 빈 일자리 현황을 파악할 수 있을 것으로 보임
 - O '빈일자리 _전체' 필드의 단위는 '명'이므로 인원수를 파악할 수 있고, '빈일자리율_전체' 필드의 단위는 %이므로 백분위수 파악 가능
- 데이터 구조 및 결합 이슈
 - O 다른 테이블과의 결합없이 사용 가능
- 데이터 클린징 이슈
 - O 산업분류 필드 분할 (알파벳, 괄호 뒤 숫자 제거) 필요

데이터셋 정제 실행

- 데이터 정제 방향
 - 1. 데이터 해석기를 통해 불러와진 필드명을 원하는 이름으로 수정
 - 2. '산업분류' 필드 내용 분할



정제 결과 해당 데이터셋의 필드

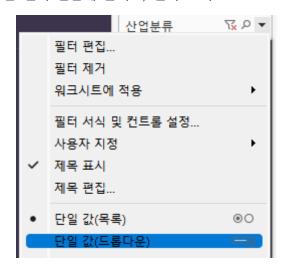
(https://docs.google.com/spreadsheets/d/1Vp7TulNf7At7e73SRsel8Lzvd71K-d1t/edit?usp=sharing&ouid=111765825620961795095&rtpof=true&sd=true)

유형	필드명
Abc	산업분류
Abc	규모
#	빈일자리_전체 (명)
#	빈일자리율_전체 (%)

대시보드 시각화



1. '산업분류', '규모' 필드를 필터 선반에 올리고, 필터 표시



2. 필터 서식을 단일 값(드롭다운)으로 설정



3. 마크유형을 텍스트로 설정한 뒤, '빈 일자리 전체 (명)' 필드와 '빈일자리율 전체 (%)' 필드를 텍스트 선반에 올려 표시

|[고용시장 현황] 분석 및 시각화 시나리오 #3

- ✓ 산업별로 기간제 근로자 비율을 확인할 수 있나요?
- ✓ 다른 산업과의 기간제 근로자 비율을 비교할 수 있나요?

기간제 근로자 비율		
전체	12.5%	
기타서비스업	19.4%	
서비스업 등	14.6%	
농림어업	13.6%	
도소매음식숙박	10.7%	
전기운수통신금융	6.9%	
제조업	5.1%	
광업·제조업	5.1%	
건설업	4.5%	

산업별 기간제근로자 수

데이터 수집 및 획득

고용 노동통계의 주제별 통계에서 데이터셋 다운로드



'고용' 주제의 '산업별 기간제근로자 수'



개별 산업분류를 비교하기 위해 '전체' 항목 체크박스 해제



가장 최신 시점인 2021 1/2 항목만을 선택



통계표조회 > 우측 상단 다운로드 버튼 > 엑셀 파일 선택, 셀 병합 해제, 다운로드

데이터셋 탐색 및 이해

데이터 원본 열람



태블로 프렙에 다운로드 받은 데이터 원본을 올린 뒤, '데이터 해석기 사용' 선택



정리 단계를 추가해 데이터셋 살펴보기

테이블 열 개수: 4개

테이블 행 개수: 16 개

■ 필드 설명

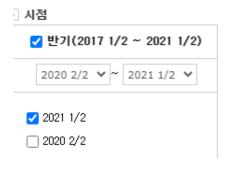
- 산업분류별: 한국표준산업분류 10차 산업대분류

- 인원수: 근로자 구분에 따른 인원 수

- 비율: 전체 근로자 대비 기간제 근로자의 비율

- 해당 데이터셋은 '산업분류별', '근로자 구분' 필드를 함께 id 필드로 하여, 데이터셋의 한 행은 산업분류, 근로자 구분별 근로자 수와 비율에 대한 정보를 담고 있음

데이터 기준 일자 탐색



2021 년 상반기 (2021.01~2021.06)

데이터셋 정제 기획

- 분석 시나리오와의 연계 이슈
 - O 산업별로 기간제 근로자 수를 파악할 수 있으며, 전체 근로자 수 대비 기간제 근로자 수의 비율을 파악할 수 있으므로 적절함
- 데이터 구조 및 결합 이슈
 - O 별도의 결합없이 바로 사용 가능
- 데이터 클린징 이슈
 - O '비율(%)' 필드명의 의미가 명확하지 않으므로 보다 정확한 필드명으로 변경

데이터셋 정제 실행

- 데이터 정제 방향
 - 1. 데이터 해석기를 통해 불러와진 필드명을 원하는 이름으로 수정
 - 2. '비율(%)' 필드명을 '근로자 비율(%)'로 수정



정제 결과 해당 데이터셋의 필드

 $\label{lem:mass} $$ \frac{\text{(https://docs.google.com/spreadsheets/d/11jyhreM6oL0M0nw3MqfRBhR1BKAs0FuW/edit?usp=sharing&ouid=111765825620961795095&rtpof=true&sd=true) $$ $$$

유형	필드명
Abc	산업분류별
Abc	근로자 구분
#	인원 수 (명)
#	근로자 비율 (%)

대시보드 시각화



- 1. '산업분류별' 필드를 행 선반에 올림
- 2. 마크 유형은 사각형으로 설정
- 3. '근로자 비율 (%)' 필드를 색상과 레이블로 설정

| [고용시장 현황] 분석 및 시각화 시나리오 #4

- ✓ 산업별로 기간제 근로 계약 만료 후 어떤 조치가 많은 비율을 차지하고 있는지 파악하고, 비교할 수 있나요?
- ✓ 산업별로 계약 만료 후 조치 항목의 비율이 각각 어떠한지 확인하고, 다른 산업들과의 비교를 통해 어떤 형태의 계약 만료 조치 비율이 높은지를 한번에 파악하기 위해서 간트차트로 시각화해주세요.



산업별 계약기간 만료자 조치현황

데이터 수집 및 획득

고용 노동통계의 주제별 통계에서 데이터셋 다운로드



'고용' 주제의 '산업별 계약기간 만료자 조치현황'



개별 근속기간 별로 비교할 것이므로 '전체' 항목 체크 해제



개별 산업분류 별로 비교할 것이므로 '전체' 항목 체크 해제



'계약만료자' 항목은 기간제 근로 계약 종료 후 '계약종료', '정규직전환', '계속고용', '기타' 조치사항에 해당하는 인원 수를 모두 합한 인원 수이므로, '계약만료자' 항목을 제외하고 선택



가장 최신 시점인 2021 1/2 항목을 선택



통계표조회 → 우측 상단 다운로드 버튼 → 엑셀 파일 선택, 셀 병합 해제, 다운로드

데이터셋 탐색 및 이해

데이터 원본 열람



태블로 프렙에 다운로드 받은 데이터 원본을 올린 뒤, '데이터 해석기 사용' 선택 정리 단계를 추가해 데이터셋 살펴보기

테이블 열 개수: 5 개

테이블 행 개수: 64 개

■ 필드 설명

- 인원수: 계약 만료 후 취해진 조치별 인원 수, 계약만료 필드의 '계약만료자'와 대응되는 인원수 필드의 값은 계약종료, 계속고용, 정규직전환, 기타에 해당하는 인원 수가 합쳐진 전체 값임
- 비율(%): 전체 계약만료자 중 특정 조치에 해당하는 인원 수
- 해당 데이터셋은 '산업분류별', '근속기간', '계약만료 후 조치' 필드를 함께 id 필드로 하여, 데이터셋의 한 행은 산업분류와 근속기간별로 계약만료 후 조치별로 근로자 수와, 각각의 계약만료 후 조치 사항이 차지하는 비율에 대한 정보를 담고 있음

데이터 기준 일자 탐색



2021 년 상반기 (2021.01~2021.06)

데이터셋 정제 기획

- 분석 시나리오와의 연계 이슈
 - O 계약 만료 후 조치 필드가 '정규직 전환', '계속 고용', '계약 종료', '기타'의 항목으로 나뉘어져 있으므로 적절함
 - O '근속기간' 필드가 있으므로 추가적으로 근속기간에 따른 계약만료 후 조치 현황을 비교할 수 있을 것으로 보임
- 데이터 구조 및 결합 이슈
 - O 별도의 결합없이 바로 사용 가능

- 데이터 클린징 이슈
 - O '계약만료' 필드명의 의미가 명확하지 않으므로 필드명 변경이 필요
 - O '비율(%)' 필드명의 의미가 명확하지 않으므로 필드명 변경

데이터셋 정제 실행

- 데이터 정제 방향
 - 1. 데이터 해석기를 통해 불러와진 필드명을 원하는 이름으로 수정
 - 2. '계약만료' 필드명을 '계약만료 후 조치'로 수정
 - 3. '비율(%)' 필드명을 '조치 비율(%)'로 수정



정제 결과 해당 데이터셋의 필드

(https://docs.google.com/spreadsheets/d/13hpL36Ozgg3VSN3MisT-E10UrziHmRq4/edit?usp=sharing&ouid=111765825620961795095&rtpof=true&sd=true)

유형	필드명
Abc	근속기간
Abc	산업분류별
Abc	계약만료 후 조치
#	인원 수 (명)
#	조치 비율 (%)

대시보드 시각화



1. '계약 만료 후 조치' 필드에 따른 '조치 비율 (%)' 필드를 간트 차트로 표시하기 위해 계산된 필드 3개 생성



2. 생성한 계산된 필드를 이용해 'Gantt Start'와 'Gantt Percent' 계산된 필드를 생성해 간트 차트 표시



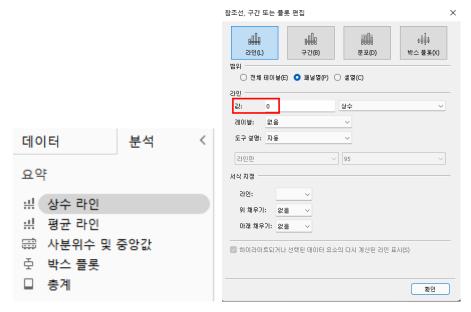
3. 근속기간별로 비교할 것이므로, '근속기간' 필드와 생성한 계산된 필드 'Gantt Percent'를 열 선반에 놓고, '산업분류별' 필드를 행 선반에 둠



4. 마크 유형을 간트 차트로, '조치 비율 (%)' 필드를 크기와 레이블로 설정 '계약만료 후 조치' 필드를 색상과 레이블로 설정



5. 표현하고 싶은 '조치 비율' 필드를 '계약만료 후 조치' 필드를 기준으로, 간트 차트에 알맞은 값으로 변환하기 위해 Gant Percent 필드를 '계약 만료 후 조치' 차원을 기준으로 퀵 테이블 계산



6. 분석 패널 → 상수라인을 0 으로 지정하여 0 에 해당하는 라인을 표시해 각 항목을 효과적으로 비교할 수 있게 함