

시각화 과제 &gt;

## [2주차] 20기 이호림



20기 이호림 준회원

2023.01.26. 17:43 조회 0

댓글 0 URL 복사

첨부파일 모아보기 1

## 사용한 데이터:

- 1) 데이콘 병원 개/폐업 분류 예측 train.csv 데이터(학기 중 기말 프로젝트 때 사용했던 파일로, 데이터 전처리까지 완료된 채로 이용)
- 2) 태블로 내장 데이터- 슈퍼 스토어
- 3) 청소년 정신 건강 데이터 (2주차 실습 데이터)

20기 시각화 2주차 과제\_이호림.twbx

## 1. 계산된필드&amp;목록형 매개변수-병원운영기간과 직원 수에 따른 병원 개/폐업 여부

career

hospital

DATEDIFF('year', [Open Year], #2018#)|

전체

검색

ABS  
ACOS  
AND  
AREA  
ASCII  
ASIN  
ATAN  
ATAN2  
ATTR  
AVG  
BUFFER

ABS (숫자)

주어진 숫자의 절대값을 반환합니다.

예: ABS (-7) = 7

계산이 유효합니다.

2 종속성

적용

확인

우선, 계산된 필드 기능을 이용하여 2018(년)에서 open year(병원 설립 연도)를 빼주어 병원의 운영 연도인 career 변수를 새로 만들어 주었습니다. 현재 연도인 2023년이 아닌, 2018에서 빼준 이유는 해당 데이터 파일이 업로드된 연도가 2018년이기 때문입니다. **DATEDIFF** 함수를 이용하여 날짜 계산에 이용하였습니다 (구글링 짱...bb 표현하고자 하는 바가 있을 때 구글링 하면 다 나와서 아주 편리해요!)

### 3. 의료기관의 조직요인(organizational factor)과 폐업

의료기관의 조직요인과 폐업 간의 관계를 살펴본다. 조직요인으로는 전체 중 전문의 비율과 간호사 비율을 각각 설정하였다. 다만, 치과의 경우는 치과 전문의와 간호사+치과위생사 비율로 하였다. 분석 결과 의원의 경우 전문의와 간호사의 비율이 낮을수록 폐업할 가능성이 높았으며, 통계적으로 유의하였다(OR, 0.989; 95% CI, 0.978 to 1.000;  $p=0.04$ ; OR, 0.989; 95% CI, 0.979 to 0.998;  $p=0.02$ ). 이러한 경향은 치과의원에서도 관찰되었다(OR, 0.995; 95% CI, 0.993 to 0.997;  $p<0.001$ ; OR, 0.997; 95% CI, 0.994 to 0.999;  $p=0.03$ ). 그러나 병원급 의료기관에서는 이들 변수가 관련성이 없었다(표 3).

\*출처: <https://repository.hira.or.kr/bitstream/2019.oak/2986/2/2022%20HIRA%20Research%202%ea%b6%8c%201%ed%98%b8%2010.pdf>

[병원, 의원, 치과의원의 폐업 관련 요인] 논문을 활용하여, 병원의 개/폐업에 영향을 주는 변수를 해당 병원의 직원 수로 설정을 하였습니다. 고용한 총 직원 수인 변수 employee1,2(2017년,2016년)을 이용하여 직원 수가 적을 수록 병원의 폐업(close)에 큰 영향을 주는지 확인하기 위하여 막대그래프를 그려보았습니다.



병원운영기간을 1년으로 설정(매개변수 이용)



병원운영기간을 6년으로 설정(매개변수 이용)

위에서 생성한 career 변수를 목록형 매개변수 '병원운영기간'으로 설정하여 운영 기간과 고용한 총 직원 수에 따라 병원의 개/폐업 여부가 어떻게 변화하는지 시각화 하였습니다. 직접 파일을 열어 확인해 보면, 병원 운영기간이 1,3,4,5,6,11,13,19,35(년) 일 때 병원이 문을 닫았고 35(년)을 제외한다면 운영 기간이 20년 이하일 때 문을 닫는다고 예측해 볼 수 있습니다.

추가로 평균적으로 병원이 개업했을 때 폐업했을 때보다 5배 정도 직원 수가 많은 것을 확인할 수 있습니다. 이는 운영기간을 6(년)으로 했을 때 더 자세히 살펴볼 수 있는데, 폐업의 직원 수는 15명으로 극히 적었고 반대로 개업의 직원 수는 약 1800명 정도로 높게 나타났습니다.

결론적으로 위에서 살펴본 논문의 **직원 수가 적을 수록 폐업할 가능성이 높고 통계적으로 유의하였다**는 문장처럼 직원의 수가 병원의 폐업에 영향을 주는 것이 유의미하다고 볼 수 있습니다.

2. LOD 계산 지역별 국가/지역의 평균 매출(INCLUDE, EXCLUDE, FIXED)

1) 행에 지역만 설정

LOD 계산(행에 지역만-평균 매출)

지역	평균 INCLUDE국가/지역	EXCLUDE국가/지역	FIXED지역	평균 매출
동남아시아	293.0	282.7	282.7	₩283
북아시아	310.9	363.1	363.1	₩363
오세아니아	244.4	315.5	315.5	₩316
중앙아시아	300.1	370.1	370.1	₩370

지역별로 구분된 평균 INCLUDE국가/지역,EXCLUDE국가/지역,FIXED지역및평균 매출입니다.

EXCLUDE 국가/지역의 값과 FIXED 지역의 값이 같은 이유는 전자는 국가/지역을 제외한 상위 범주 지역끼리의 평균을 구했고, 후자는 상위 범주인 지역을 고정하여(결국 그냥 지역별로 평균을 낸 것과 같음) 평균을 구했기 때문에 값이 같다.

지역의 하위 범주인 국가/지역 변수를 INCLUDE 하면 각 지역 별 국가들끼리 평균 낸 다음 다시 그 값들을 평균을 내기 때문에 나머지 3개의 값들과 다르게 나온다. (평균 매출은 그냥 지역별로 바로 평균을 내기 때문에 EXCLUDE 국가/지역, FIXED 지역과 값이 마찬가지로 동일)

2) 행에 지역과 국가/지역을 설정

LOD 계산(행에 지역&국가/지역-평균 매출)

지역	국가/지역	평균 INCLUDE국가/지역	EXCLUDE국가/지역	평균 FIXED지역	평균 매출
동남아시아	Cambodia	388.4	282.7	282.7	₩388
	Indonesia	291.5	282.7	282.7	₩291
	Malaysia	348.6	282.7	282.7	₩349
	Myanmar (Burma)	251.0	282.7	282.7	₩251
	Philippines	269.3	282.7	282.7	₩269
	Singapore	285.7	282.7	282.7	₩286
	Thailand	261.2	282.7	282.7	₩261
	Vietnam	248.3	282.7	282.7	₩248
북아시아	China	372.1	363.1	363.1	₩372
	Japan	403.2	363.1	363.1	₩403
	Mongolia	40.2	363.1	363.1	₩40
	South Korea	192.6	363.1	363.1	₩193
	Taiwan	546.3	363.1	363.1	₩546
오세아니아	Australia	326.1	315.5	315.5	₩326
	New Zealand	273.9	315.5	315.5	₩274
	Papua New Guinea	133.1	315.5	315.5	₩133
중앙아시아	Afghanistan	394.1	370.1	370.1	₩394
	Bangladesh	480.1	370.1	370.1	₩480
	India	383.1	370.1	370.1	₩383
	Nepal	185.4	370.1	370.1	₩185
	Pakistan	236.4	370.1	370.1	₩236
	Sri Lanka	121.7	370.1	370.1	₩122

지역및국가/지역별로 구분된 평균 INCLUDE국가/지역,EXCLUDE국가/지역,평균 FIXED지역및평균 매출입니다.

행에 지역 변수와 국가/지역 변수를 동시에 설정하면, 첫 번째 사진과는 다르게 INCLUDE 국가/지역과 평균 매출 값이 동일하게 나온다.

예를 들어 동남아시아의 하위 범주인 8개의 국가 별로 각각의 매출의 평균 낸 값이 평균 매출 값인 것이고 이 8가지 평균 매출 값들의 평균 값이 바로 첫 번째 사진의 293이라는 값이 나오는 것이다.

3. 데이터 원본 필터&차원필터&목록형 매개변수

**필터 [지역]** X

일반 와일드카드 조건 상위

☒ 목록에서 선택(S) ☐ 사용자 지정 값 목록(C) ☐ 모두 사용(U)

검색 텍스트 입력

☐ 대전  
☒ 부산  
☒ 서울  
☐ 세종  
☐ 울산  
☐ 인천  
☐ 전남  
☐ 전북  
☒ 제주  
☐ 충남  
☐ 충북

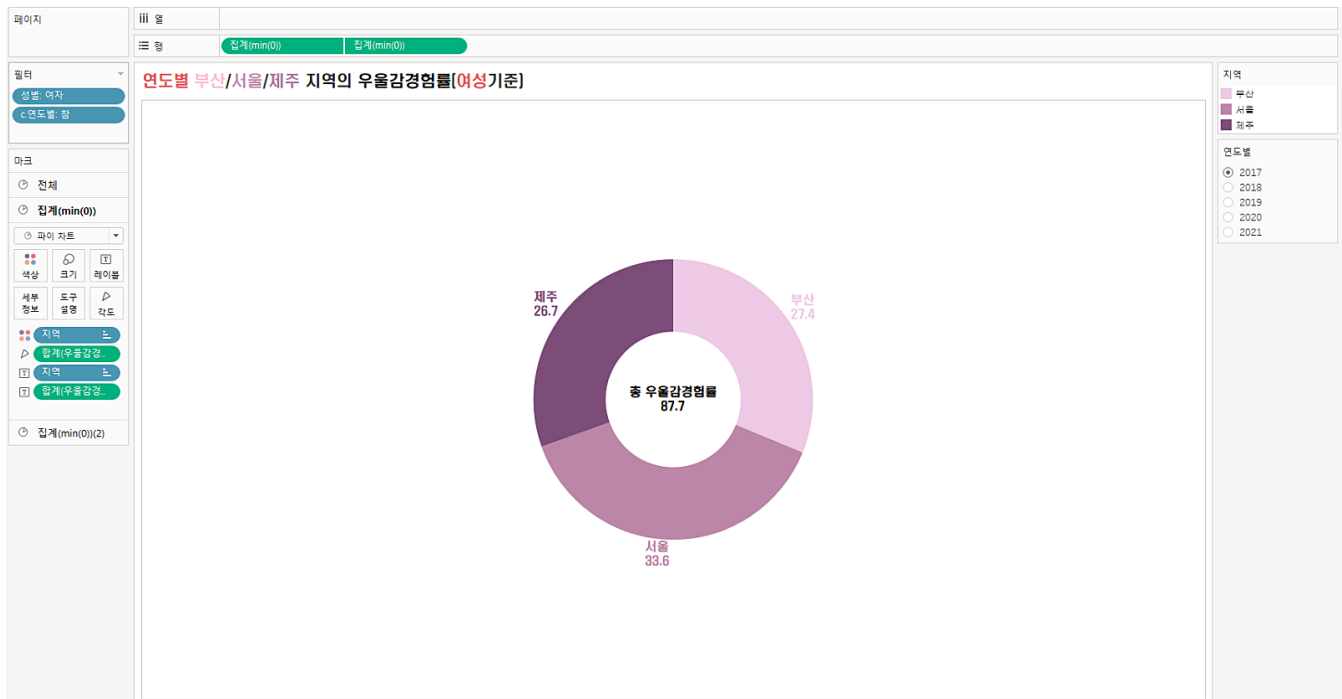
전체(L) 없음(O) ☐ 제외(X)

요약

필드: [지역]  
 선택 내용: 3개(총 17개)의 값을 선택함  
 와일드카드(W): 전체  
 조건(D): 없음  
 제한(M): 없음

재설정 확인 취소

데이터 원본 필터를 이용하여 지역을 부산, 서울, 제주만 추출하였습니다(서울은 제가 사는 곳이라 ㅎㅎ 부산이랑 제주는 여행 가고 싶은 곳이라서 골랐어요!)



이후 차원 필터를 이용하여 성별을 여자로만 추출하였고 각 지역 별 우울감 경험률을 보여주는 파이차트를 시각화하였습니다. 연도별 목록형 매개변수를 생성하여 연도에 따라서 비율이 어떻게 변하는지 한눈에 볼 수 있도록 설정하여 가시성을 높이려 하였습니다.

레이블 서식을 이용하여 글씨체, 크기, 색상(마크와 색상 일치)도 적용해 보았습니다! 실습하면서 느낀 건데 레이블 서식 같은 기능을 잘 활용하면 더 깔끔한 시각화가 되는 것 같아요 ”.

이번 주에 굉장히 많은 기능들과 함수들을 배웠는데 되게 재밌었고 신기했어요! 앞으로 더 열공해서 적재적소에 자유자재로 잘 활용하도록 하겠습니다 :) 양질의 녹화강의 만들어 준 다운 멘토님 감사당! 짱 bb 🥰

## 댓글

20기 이호림

댓글을 남겨보세요



등록