

6일차 과제

6일차 과제를 시작하시기 전에 안내드린 동영상을 반드시 보시길 바랍니다 !

[Day6]

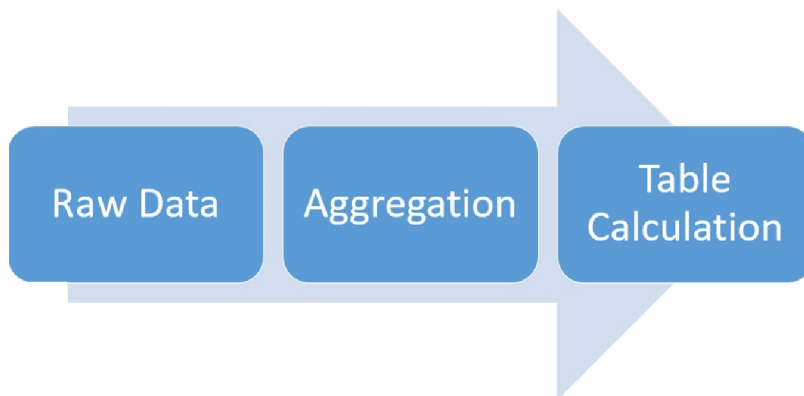
DAY4에 맛보았던 테이블 계산식을 좀 더 연습해 보면서 테이블 계산식과 세부 수준 계산식 (LOD)의 개념을 이해해보는 날입니다. 태블로를 통해 비즈니스 관련한 여러 가지 분석을 하기 위해 필요한 내용입니다. 오늘 내용을 끝까지 수행하신다면 태블로 고급 분석에 한 발 더 다가갈 수 있어요!!

과제 안내 작성을 완료하고 보니 너무 길어졌는데요. 시간이 좀 걸리시더라도 꼼꼼히 읽어보시고 개념을 확실히 잡고 시작하셨으면 좋겠습니다. **읽기를 포기하지 말아주세요!!!!!!!!!**

테이블 계산식의 원리는?

아래 그림처럼...

테이블 계산식은 “Raw Data”를 “집계한 값 (Aggregation)”을 가지고 “테이블 계산 (Table Calculation)”을 실행합니다.



“4일차 과제 - 국가별 가격 변동 비율 차이”시트를 열어서 함께 살펴볼까요?

우리는 국가별 가격 변동 비율을 구하기 위해서...

- (1) Dollar Price(Raw Data)의 집계(Aggregation) 를 구하고,
- (2) 그 집계 값을 가지고 “구성비율”이라는 테이블 계산식(Table Calculation)을 통해 변동 비율을 계산했습니다.
즉, 이미 집계된 Dollar Price 집계 값을 가지고 ‘재계산’ 했죠.

그런데 여기서 잠깐!!!! Dollar Price는 무엇을 기준으로 집계 값을 구했었나요?

- Dollar Price의 합계를 구한 기준은 연도(Date)와 국가명(Name)이었습니다.

기본적으로 태블로는 집계 값을 통해 시각화를 합니다. 이 때, 집계의 기준은 “시각화의 세부 수준”에 따라 결정되는데요.

시각화의 세부수준이 무엇인지 함께 살펴봅시다.

새로운 시트를 여시고 Dollar Price를 더블 클릭해보세요. 그럼 전체 데이터의 총 Dollar Price의 합계가 표현되는 것을 볼 수 있습니다. 그 상태에서 Name을 열 선반에 올려보세요. 이번엔 Dollar Price가 국가 수준으로 집계되는 것을 볼 수 있습니다.

이처럼 시각화에 추가되는 차원에 따라 집계의 기준이 변경 되는데요. 여기서 말하는 **측정값의 집계 기준을 시각화의 세부 수준 또는 뷰의 수준**이라고 합니다.

그리고 이러한 시각화 세부 수준에 영향을 미치는 곳은 아래 그림에 붉은색으로 표시된 곳입니다.

붉은색에 추가되는 “차원”이 시각화의 세부 수준을 결정 짓는다고 생각하시면 됩니다.



다시 테이블 계산 얘기로 돌아와 볼까요?

앞에서 보셨던 것처럼 측정값은 시각화의 세부 수준에 따라 집계되고, 이 집계 값을 갖고 다시 재계산 해야 하는 테이블 계산은 어떤 기준으로 계산을 해야할 지 지정해줘야 하는 것이죠. 이 때, 재계산이 되는 범위(테이블, 패널, 셀)와 방향, 기준에 따라 값이 달라 질 수 있기 때문에 범위와 방향을 신중하게 설정해줘야 합니다.

시청하신 테이블 계산식 동영상과 LOD 동영상, 그리고 정리된 문서를 보셨으면 본격적으로 과제를 시작해볼까요?

아래 파일을 다운로드 받으세요. (최신 버전 다운로드를 권장합니다.)

V2020.1 이상 : @DAY6 테이블 계산식 & LOD_Start.twbx

V2018.1 : @DAY6 테이블 계산식 & LOD_Start_v2018.1.twbx

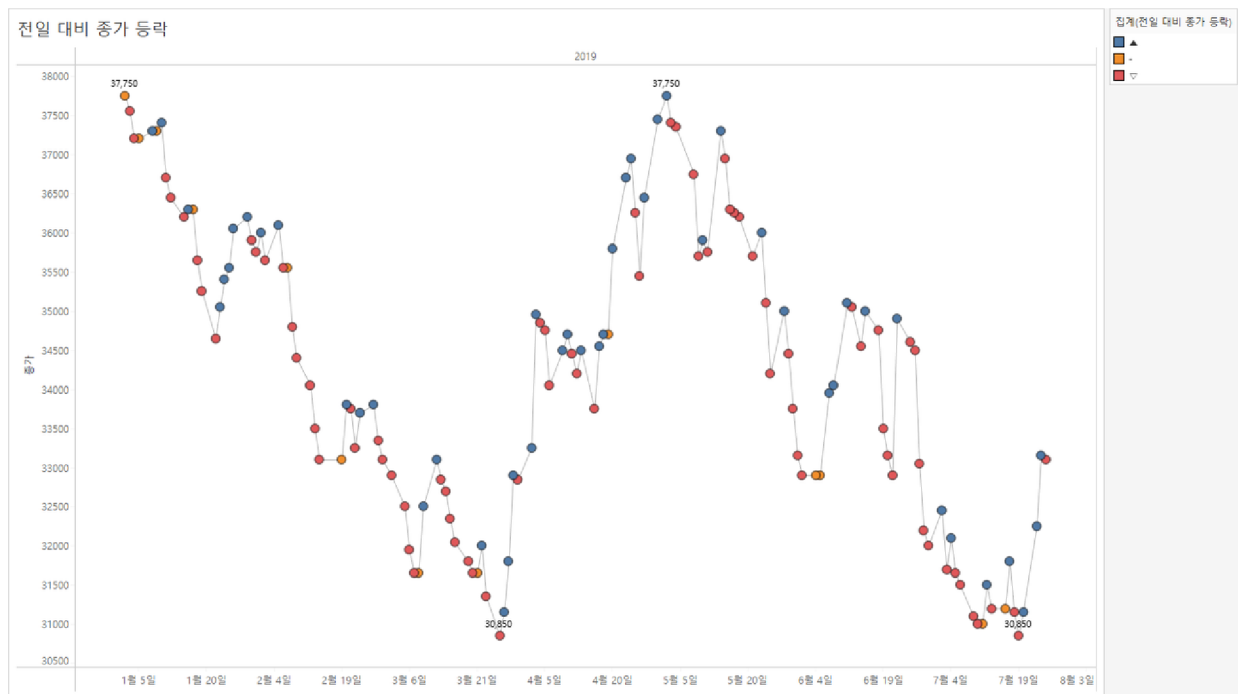
DAY6 과제 내용

- 1. 테이블 계산식 - 차이
- 2. 테이블 계산식 - 구성비율
- 3. 세부수준 계산식(LOD): 전국 기준 매출 구성비율 구하기
- 4. 세부 수준 계산식: 과거 고객 매출 기여도 및 신규 유입 고객 현황 분석
- 5. [추가 도전] 코호트 분석

1. 테이블 계산식 - 차이

아래와 같이 주식 데이터를 가지고 전일 대비 증가가 상승했는지 하락했는지 나타내는 차트를 아래와 같이 그려봅니다.

사용 Data : 주식 Data



이 과제에서는 몇 가지를 고려해 주셔야 해요.

[시각화를 위한 HINT]

1. 마크의 라인과 원을 이용해서 2019년 일 별 증가를 표현해 주세요.

- 증가를 하나는 라인, 하나는 원으로 나타내고 이중축을 이용하시면 되겠죠?
- “일”은 “연속형”으로 표현하는게 좋을 것 같네요.

2. 마크의 원은 “전일 대비 증가 등락 여부”로 색상을 표현해 봅니다. (상승은 파랑, 동일은 주황, 하락은 빨강)

1) “전일 대비 증가 등락”을 구하기 위해서는 테이블 계산식을 이용해야 할텐데요.

처음부터 계산식을 작성해서 구해내기 어려우시다면 다음과 같은 방법을 활용해 보세요.

- 측정값에 있는 증가를 행 선반에 가져다 놓고, 쉼표 테이블 계산을 이용해서 “차이”를 구해 보세요.
- 테이블 계산 범위는 전체 일자가 되어야 하니 테이블이 되어야겠죠?
- 계산 기준을 “전일”로 해야 이전 값을 가져와서 차이를 구할 수 있을 것 같아요.

2) 구해 놓은 테이블 계산 값을 차트에서 확인해 보세요.

- 전날 보다 증가가 상승했으면 양수, 전날 보다 하락했으면 음수, 아니라면 0 값이 나타나겠죠. 그런데 첫 번째 날 (2019-01-02)은 비교할 이전 값이 없으니, 그 날짜에는 테이블 계산 값이 안보일거예요. 화면 하단에 1 Null이 보이 시나요. 이처럼 계산이 되지 않은 값이 있음을 확인할 수 있습니다.

3) 제대로 계산이 되었다면, 테이블 계산식을 이용해 구해 놓은 증가를 “데이터 창”으로 드래그 해보세요!

새로운 측정값이 생성되나요? 이름을 “전일 대비 증가 등락”으로 해주세요.

- 어떤 식으로 계산되었는지 식을 확인해 봅시다!
 - 측정값에 있는 “전일 대비 증가 등락”에서 마우스 오른쪽 버튼을 클릭하고 “편집”을 클릭해보세요.
 - $ZN(SUM([증가])) - LOOKUP(ZN(SUM([증가])), -1)$
 - 위의 함수는 기준 일자의 증가 합계 값과, LOOKUP 함수를 이용해서 현재 기준 일자의 -1일 자의 집계된 증가를 가져와서 계산을 하고 있네요. (DAY6 동영상 안내에 테이블 계산 함수를 보시면 다양한 함수를 볼 수 있어요.)
 - 즉, 기준일자 증가 합계에서 이전 일자의 합계를 뺀거죠. 결과가 왜 양수, 음수, 0 값이 나왔는지 아시겠죠?

4) “전일 대비 증가 등락”을 이용해서 “증가 여부 색상”의 계산식을 만들어 보세요.

- 전일 대비 증가 등락 > 0 이라면 등락
- 전일 대비 증가 등락 < 0 이라면 하락
- 그도 아니라면 동일

[추가] 워크시트 작업(동작) 활용하기

차원에 있는 코드명을 이용해서 해당 주식에 대한 정보 사이트로 이동해 봅시다.

※ ‘작업’이라는 용어가 최신 버전에서 ‘동작’으로 변경되어서 병행 표기하였습니다.

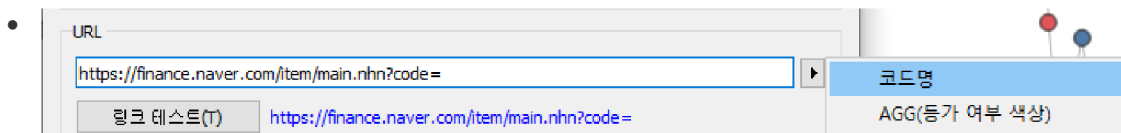
1. “코드명”을 “원 마크”의 “세부 정보”에 추가해 주세요.

2. 메뉴 > 워크시트 > 작업(동작)을 클릭해 주세요.

3. 작업(동작) 추가에서 URL로 이동을 클릭해 주세요.

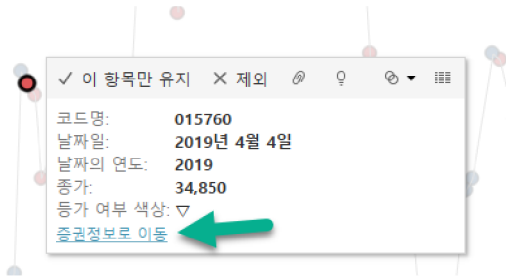
4. URL 추가 작업(동작)에서 아래와 같이 해보세요.

- 이름 : 증권정보로 이동
- 작업(동작) 실행 조건 : 메뉴
- URL에 아래 주소 입력
 - <https://finance.naver.com/item/main.nhn?code=>
- 위 주소에서 code= 뒤에 코드명을 넣어 주세요.



- 확인 클릭

5. 원 마크를 클릭하고 도구 설명에 있는 “증권정보로 이동”을 클릭해보세요.

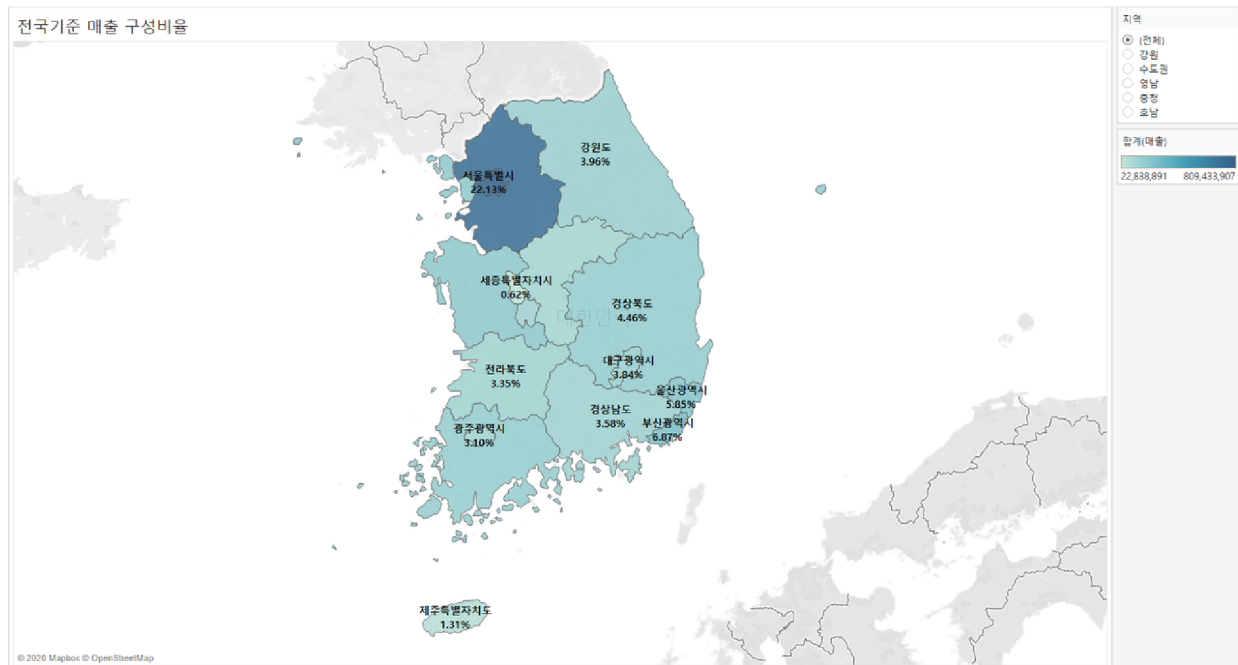


원본 데이터에는 코드명이 존재하지 않아서 임의로 하나의 데이터 값을 만들어 놓았습니다. 따라서, 하나의 사이트만 연결이 되었지만, 다양한 코드를 가진 필드가 있다면, 해당 코드에 대한 사이트로 연결될 수 있도록 활용할 수 있겠죠?

2. 테이블 계산식 - 구성비율

테이블 계산식을 활용해서 시도별 매출 구성비율을 살펴봅니다.

사용 Data : 주문+반품 추출



[시각화를 위한 HINT]

1. 국가, 시도를 상세 정보에 넣고, 마크의 색상은 매출 합계로 나타내 주세요.
2. 매출의 “구성 비율”을 레이블로 표시해 주세요.
 - 1) 매출을 레이블로 드래그 해 주세요
 - 2) 킷 테이블 계산의 “구성 비율”을 이용해보세요. 범위는 테이블로 해야 전국을 기준으로 비율이 계산 되겠죠?

- 구성비율 계산식을 더블클릭해서 살펴보면 아래와 같은 계산식을 볼 수 있습니다.
 - $SUM([매출]) / TOTAL(SUM([매출]))$
 - 매출 값을 전체 매출의 합계(TOTAL) 값으로 나눠주고 있죠. (**TOTAL 함수는 테이블 계산함수 입니다.**)

3. 지역을 필터로 추가해주세요! (단일 값 선택만 가능하게 옵션을 조정해주세요.)

“수도권” 지역 필터를 선택해 보세요. 서울특별시의 전국 대비 매출 구성비율 22.13%의 값이 그대로 보이시나요?

왜 비율이 46.82%로 바뀌었을까요?

“구성 비율”을 구할 때 계산의 범위를 “테이블”로 지정해 했습니다.

지역 필터를 적용하기 전에는 화면에 보이는 전체 시도에 대해 범위가 지정이 됐다면,
필터를 적용한 후에는 해당 지역에 포함된 '시도'에 대한 범위로 계산 범위가 바뀌게 됩니다.

즉, **필터 적용 전**에는 서울특별시 매출 합계 / 전국 매출 합계 (TOTAL의 범위가 전국)로 구성비율이 구해졌다면,
필터 적용 후에는 서울특별시 매출 합계 / 인천,서울,경기도 매출 합계 (TOTAL의 범위가 수도권)로 계산 되다보니
구성비율의 수치가 달라지게 된거죠.

그렇다면, 필터를 적용해도 전국 대비 매출 구성비율의 수치가 바뀌지 않게 하려면 어떻게 해야 할까요?
다음 과제에서 알아보시다.

3. 세부수준 계산식(LOD) : 전국 기준 매출 구성비율 구하기

세부수준 계산식을 이용해 지역 기준 / 전국 기준 매출 구성비를 함께 살펴봅시다.

과제를 시작하기 전에 세부 수준식 (LOD) 개념을 다시 한 번 짚어 볼까요?

동영상을 시청하셨다면 세부 수준식을 어느 때 사용할 수 있는가를 아셨을텐데요.
세부 수준식을 3번 과제에 왜 적용할 수 있는지 한 번 살펴봅시다!

다운로드 받은 워크북에서 “지역별/전국구 구성비 함께 보기”시트를 열어주세요.

현재 시트의 시각화 세부 수준을 한 번 생각해 보세요.

이 시트의 시각화 세부 수준은 “국가”와 “시도”입니다. 이 뷰에서 매출의 합계 값을 구한다면 가장 낮은 시각화 세부 수준인 “시도” 기준으로 매출 합계 값이 구해지게 되겠죠.

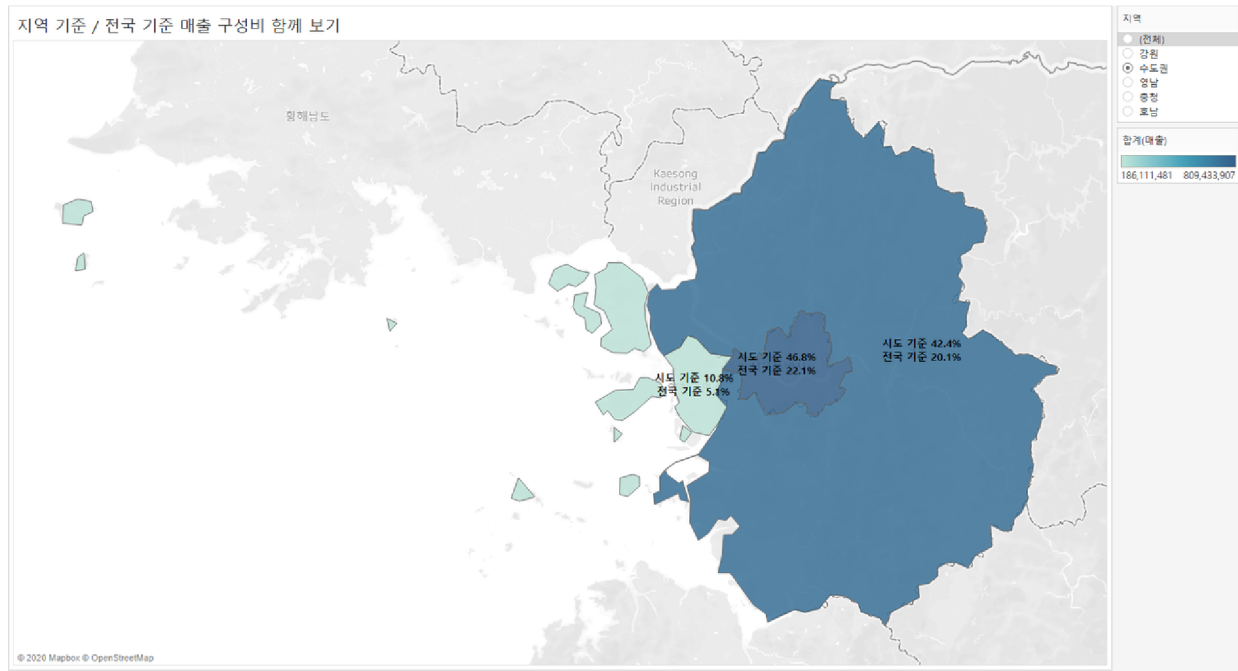
그런데 우리의 최종 목적인 전국 기준 매출 구성 비율을 구하려면 “시도 매출의 합계 값”을 “전국 매출의 합계 값”으로 나눠야 했습니다.

즉, 우리에게 필요한 “전국 기준”의 매출 합계 값은 이 뷰의 수준에서는 절대 구할 수가 없는거죠.

이처럼 “현재 뷰의 시각화 세부 수준”과 “실제 내가 필요한 집계 수준의 기준”이 상이할 때 사용할 수 있는 것이 “세부 수준 계산식”입니다.

이러한 개념을 가지고, 아래와 같이 시도 기준 매출 비율과 전국 기준 매출 비율을 함께 나타내는 시각화를 나타내 봅시다!

사용 Data : 주문+반품 추출



[시각화를 위한 HINT]

1. FIXED 함수를 이용해 “전국 기준 매출” 계산식을 만들어 보세요.

- FIXED는 지정된 특정 차원을 기준으로 집계하는 방법입니다.
 - 시도의 전체 합계 매출을 구하려면 그보다 높은 수준의 차원인 “국가”를 기준으로 매출의 합계 값을 구하면 되겠죠?

2. “전국 기준 구성비율” 계산식을 만들어 보세요.

- 아래와 같이 **매출 합계**를 이전에 구해 놓은 **전국 기준 매출 합계**로 나누면 되겠죠.
 - $\text{SUM(매출)} / \text{SUM}([전국 기준 매출])$

*애초에 뷰의 세부 수준이 “시도”이니 **매출 합계**는 **시도의 매출 합계**를 나타내는 것! 이제 확실히 이해가 가시나요?
(더 이상의 설명은 생략한다.)

3. 새로 만든 “전국 기준 구성비율”을 ‘레이블’에 추가해주세요. 레이블을 클릭해서 위의 결과와 같이 나타나게 해주세요.

자 이제 필터를 바꿔보세요. 전국 기준 매출 구성 비율 값이 안 바뀌는 것이 보이시나요?

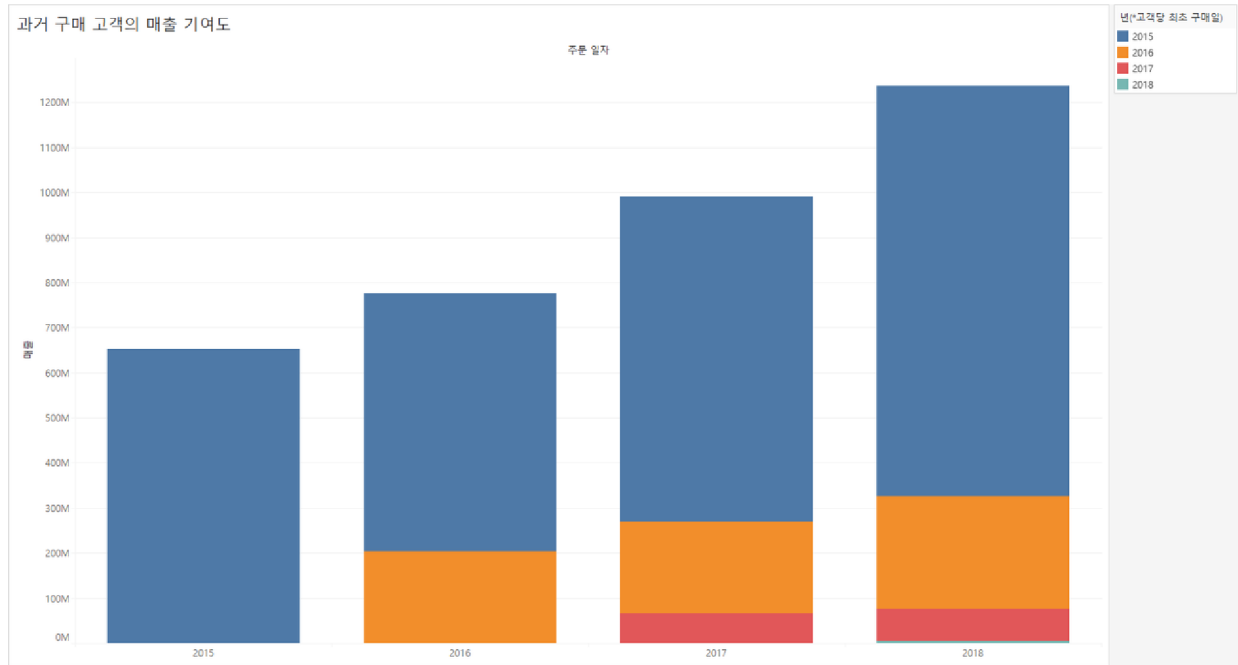
4. 세부 수준 계산식 : 과거 고객 매출 기여도 및 신규 유입 고객 현황 분석

이번 과제에서는 과거 구매 고객 매출 기여도와 신규 유입 고객 현황을 살펴봅시다.

2015년 처음 비즈니스를 시작한 슈퍼스토어입니다. 2015년에 첫 구매가 발생한 고객이 2018년도 까지 얼마나 매출에 기여하고 있을까요? 아래 시각화를 살펴보니 2015년에 첫 구매를 발생시킨 고객이 여전히 2018년도까지 많은 매출에 기여하고 있음을 알 수가 있는데요. 반면에 매년 새로 유입되는 고객은 줄어드는 것을 볼 수가 있는데요.

아래와 같은 시각화를 통해 고객의 충성도와 신규 유입 고객 현황을 살펴 봅시다.

사용 Data : 주문+반품 추출



[시각화를 위한 HINT]

1. 주문일자를 이용해 연도를 옆에 놓고, 매출은 행에 놓아주세요.
2. 위의 내용을 시각화 하기 위해 고객별 최초 구매 연도를 색상으로 표현해 봅시다.

예를 들어, 아래와 같이 데이터가 있을 때, 각 “고객별 최초 구매일”을 구하고, 고객 별 최초 구매일의 연도를 색상으로 표현하면 매출 막대의 색상은 고객의 유입 년도에 따라 구분되겠죠?

	A	B	C
1	고객ID	고객명	주문일자
2	1	신호임	1/3/2015
3	1	신호임	3/13/2015
4	1	신호임	1/13/2016
5	1	신호임	2/13/2017
6	1	신호임	3/13/2018
7	2	김민지	1/3/2015
8	2	김민지	2/3/2016
9	3	최수현	1/4/2016
10	3	최수현	1/3/2018
11	4	이수림	7/5/2017
12	4	이수림	3/1/2018

- “고객별 최초 구매일”이라는 이름으로 계산식을 만들어 주세요.

- 현재 시각화 세부 수준은 “연도”이기 때문에, “고객”을 기준으로 최초 주문 일자를 구하려면 우리는 세부 수준식을 써야하겠네요.
- FIXED 함수를 이용해서 “고객”으로 차원을 고정하고, MIN 함수를 이용해서 최초 주문일자를 구해보세요.

{ FIXED [고객번호]: MIN([주문 일자]) }

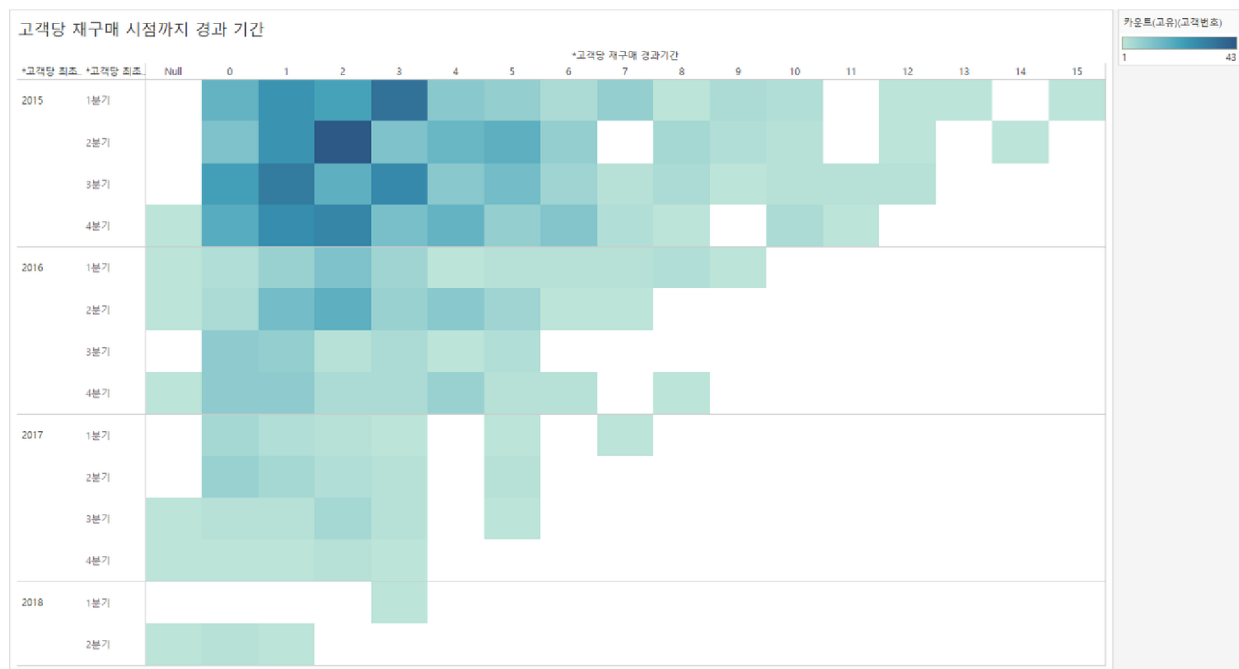
3. 새로 만든 “고객별 최초 구매일”을 색상에 놓아주세요.

아래 내용은 추가적으로 도전해 보시고 싶으신 분만 수행해 주세요!!!

[추가 도전] 코호트 분석

최초 구매가 일어난 후, 두번째 구매가 발생하기까지의 기간이 어느 정도 되는지를 시각화 해봅니다.
즉, 2015년 2분기에 처음 구매를 한 고객은 2분기가 지난 후 가장 많이 재구매를 했음을 나타냅니다.

사용 Data : 주문+반품 추출



[시각화를 위한 HINT]

1. 최초 구매일은 이미 4번 과제에서 “고객별 최초 구매일”로 만들어 놓았습니다.
2. 최초 구매일을 활용해서 2번째 구매일 (“고객별 최초 재구매일”)을 구해볼까요?

우리는 아래와 같은 시각화를 그리게 될텐데요.

4번 과제와 마찬가지로 우리가 구해야하는 “고객 별 최초 재구매일” 에 사용되어야 하는 집계 기준과 “시각화 세부수준 (뷰의 세부수준)”이 다르기 때문에 FIXED 세부 수준식을 사용해야 합니다.

아래 식을 이용해서 “고객별 최초 재구매일” 계산식을 만들어 주세요.

```
{ FIXED [고객번호] :
    MIN(
        IIF( [고객별 최초 구매일] < [주문 일자], [주문 일자], NULL )
    )
}
```

식이 의미하는 바를 가장 안쪽 식부터 차근차근 살펴볼까요?

(1) IIF 함수는 제일 첫 번째 조건이 만족하면, 2번째 값을, 아니라면 3번째 값을 적용한다는 의미입니다.
즉, “고객별 최초 구매일”이 “주문일자” 보다 작으면 주문일자를 가져오고, 아니라면 NULL 값이라는 얘기인데요.

아래 표와 같은 결과가 나오겠죠.

	A	B	C	D	E
1	고객ID	고객명	주문일자	고객별 최초 구매일	IIF 함수를 이용했을 때까지의 주문일자
2	1	신호임	1/3/2015	1/3/2015	NULL
3	1	신호임	3/13/2015	1/3/2015	3/13/2015
4	1	신호임	1/13/2016	1/3/2015	1/13/2016
5	1	신호임	2/13/2017	1/3/2015	2/13/2017
6	1	신호임	3/13/2018	1/3/2015	3/13/2018
7	2	김민지	1/3/2015	1/3/2015	NULL
8	2	김민지	2/3/2016	1/3/2015	2/3/2016
9	3	최수현	1/4/2016	1/4/2016	NULL
10	3	최수현	1/3/2018	1/4/2016	1/3/2018
11	4	이수림	7/5/2017	7/5/2017	NULL
12	4	이수림	3/1/2018	7/5/2017	1/3/2018

(2) IIF를 통해 구해진 주문일자 중 가장 최소값을 갖고 오게 되니, 고객별 두번째 구매날짜를 구할 수 있겠죠??
(위의 표에서 파란색 값을 가져오게 될거예요.)

3. “고객별 최초 구매일”과 “고객별 최초 재구매일”을 가지고 DATEDIFF 함수를 이용해서 ‘quarter’기준으로 재구매가 일어나기 까지의 시간 차이를 구해보세요. 계산식 이름은 “고객별 재구매 경과기간”으로 합니다.

DATEDIFF('quarter', [고객별 최초 구매일], [고객별 최초 재구매일])

4. “고객별 재구매 경과기간”의 결과 값은 숫자이기 때문에, 새로운 계산식이 측정값에 위치하는 것을 볼 수 있습니다.
하지만 우리는 “고객별 재구매 경과기간”을 집계 형태로 사용하는 것이 아니라 하나의 “차원”으로 사용해야 겠죠.
“고객별 재구매 경과기간” 필드를 드래그 하여 측정값에서 차원으로 이동시켜 주세요.

5. 구해 놓은 “고객별 최초 구매일”을 행 선반에 불연속형 “년”, “분기” 수준으로 놓아주세요.

6. “고객당 재구매 경과기간”을 열 선반에 놓아 주세요.

7. “고객번호”에서 마우스 오른쪽 버튼을 누른채 (Mac의 경우 Option을 누른채) 로 색상 에 드래그 하고,
고유 카운트로 집계해주세요. 해당 기간에 해당하는 고객을 카운트 합니다.

| 여기까지 끝내셨다면 정말 큰 박수 보내드립니다 $\pi\pi$!!!