

1일차 과제

부트캠프에 참여하신 모든 분들 격하게 환영합니다 :)

모든 과제는 학습 안내에 있는 자료를 모두 보신 후에 진행하시길 바랍니다.

제시한 과제와 시각화 결과물을 동일하게 만들기 위해서는 마크의 색상, 모양, 레이블 등 스스로 고민해야 하는 부분들이 포함되어 있습니다.
고민해 보시면서 태블로의 작동 원리를 이해하는 시간이 되셨으면 좋겠습니다.

기초를 다지고 시작해야 하다보니 1일차 내용이 좀 길어졌는데요~ 그래도 꼼꼼히 읽어보시고 부트캠프를 시작해 주세요 !!
부담감 보다는 즐거운 마음으로 참여해 보시길 바라며~ 그럼 1일차 과제 설명으로 고고!

[Day1]

1일차에는 "스타벅스 메뉴 데이터"와 "매장 정보 데이터"를 이용해 시각적 분석을 하고 대시보드를 만들어 보는 날입니다.

기초 실습 교육(1일차 동영상 안내 복사)은 잘 보셨나요?

태블로가 동작하는 원리를 떠올려 보면 ...

1. 필드는 정성적인 값과 정량적인 값에 따라 **차원**과 **측정값**으로 구분됩니다.

- **차원** : 정성적 데이터(제품명, 날짜, 지리명 등), 분석 기준이 되는 값, 불연속형 데이터로 측정값을 쪼개어 보는 하나의 관점
- **측정값** : 정량적 수치, 연속형 데이터로 집계되는 데이터

2. **측정값은 차원을 기준으로 집계되어 표현됩니다.**

여기서 말하는 **집계**란 무엇일까요?

집계는 **합계, 평균, 중앙값, 카운트, 카운트 (고유), 최소값, 최대값, 백분위수, 표준편차, 분산** 등 을 의미합니다.

(더 상세한 설명은 → 집계 (Aggregate): 📖 용어 해설)

즉, 하나의 측정값을 어떤 기준으로 집계해 보느냐에 따라서 다양한 분석을 할 수 있겠죠.

측정값을 화면에 가져다 놓으면 기본적으로 합계 값을 보여주는데요,

측정값을 합계가 아닌 평균 등 다른 집계 기준으로 보고 싶다면, 집계 형태를 변경해줘야 합니다.

집계의 형태를 변경해주는 방법

- 1) 행 또는 열, 마크 선반에 놓여진 알약에서 마우스 오른쪽 버튼 클릭 후 측정값의 집계 형태를 변경
- 2) 측정값을 가져올 때부터 집계 형태를 변경
 - Windows : 측정값에서 마우스 오른쪽 버튼을 누른 상태로 드래그
 - Mac : Option키 누른 상태로 드래그

3. 필드를 **행 선반에 놓으면 Y축**에 표현, **열 선반에 놓으면 X축**에 표현됩니다.

4. 마크 선반에 있는 마크의 형태, 색상, 크기, 레이블, 세부 정보, 도구 설명, (경로, 각도, 모양) 속성을 통해 원하는 형태로 다양한 시각화를 표현할 수 있습니다.

5. 필요한 값은 필터 선반을 통해 필터링 할 수 있습니다. (데스크탑 UI 구성 요소: 📖 용어 해설 문서를 참고하세요.)

6. 대시보드는 기존에 만들어 놓은 워크시트를 조합하여 만듭니다.

이러한 기본적인 내용을 가지고 가볍게 DAY1 과제를 시작해 볼까요?

아래 파일을 다운로드 받으세요 (최신 버전 다운로드를 권장합니다).

V2021.1 이상 : DAY1 스타벅스 메뉴 분석_Start_v2021.1.twbx

V2018.1 : DAY1 스타벅스 메뉴 분석_Start_v2018.1.twbx

1일차 과제 내용

1. 카테고리 별 평균 칼로리 & 평균 카페인
 2. 메뉴명 별 칼로리 & 카페인
 3. 카테고리별 메뉴명을 한 번에 살펴보기
 4. 당분 함유량과 칼로리 상관관계
 5. 시군구 별 매장 분포 현황
 6. 대시보드 만들기
- [추가 도전 과제]

1. 카테고리 별 평균 칼로리 & 평균 카페인

첫 번째 과제에서는 **막대 차트를 이용해 카테고리 별 평균 칼로리와 평균 카페인을 비교하는 시각화**를 그려봅니다.

왜냐면 저는 다이어트 중이라서 칼로리가 낮은 음료면서 상대적으로 카페인 함유가 적은 음료를 마시고 싶기 때문입니다!!!

그런데 잠깐!!!! 왜 하필 막대 차트를 이용하느니 한 번 알아보겠습니다.

우리는 값을 비교할 때 본능적으로 막대 차트를 사용했었습니다.

왜 막대 차트를 사용할까 그 이유를 한 번쯤 생각해 보신 적 있으신가요?

막대 차트는 값의 크기를 비교하는데 유용한 시각화입니다.

아래 예시를 한 번 살펴봅시다.

수치 데이터를 표현하는 방식은 다양한데요.

막대 차트처럼 값을 **길이**로 표현할 수도 있고, 버블 차트처럼 값을 원의 **크기**를 통해서 표현할 수도 있습니다.

그렇다면 오른쪽 버블 차트에서 독일, 멕시코, 인도의 매출을 한 번 비교해 보시겠어요? 어느 나라의 매출이 가장 큰가요?

막대 차트를 통해 값을 비교해 보면 독일이 가장 매출 값이 크다는 것을 알 수 있지만, 버블의 크기로는 값의 차이를 비교하기 쉽지가 않습니다.

우리의 뇌는 크기나 너비 보다는 길이를 비교하는데 좀 더 특화되어 있다고 해요.

이처럼 막대 차트는 **수치 데이터 값들 간의 작은 양적 차이를 비교하는데 유용**하기 때문에 가장 많이 사용 되는 시각화 중에 하나인거죠.

그리고 막대 차트는 특정 참조선(ex. 평균값, 중간값) 등을 표현해 해당 막대가 그 참조선 값에 도달했는지 도달하지 못했는지 비교할 수도 있고,

바인바(Bar in Bar) 차트 등을 통해 목표값에 도달했는지 아닌지 등도 살펴볼 수 있는 등 추가적으로 시각화를 발전시켜 나가기도 용이합니다.

그렇다면 버블 차트는 사용하지 말아야 할 시각화일까요?

버블 차트는 정확한 값의 비교가 아닌 전체적인 데이터의 트렌드를 본다가거나 대시보드에서 필터로 유용하게 사용할 수 있습니다.

다양한 시각화를 적재적소에 사용하면 좋겠죠?

그리고 막대 차트를 사용할 때 유의할 점 **

아래 예시의 오른쪽 막대 차트는 정렬되어 있지 않습니다.

그럼 화면에 보시는 것처럼 떨어져 있는 막대 간의 길이를 비교하는 것이 어려워집니다.

즉, 값의 비교가 쉬운 막대 차트의 장점이 줄어드는 거죠!

따라서 막대 차트를 사용하실 때는 **비슷한 값들의 비교를 명확하게 하기 위해서 데이터를 정렬하는 것을 권장합니다.**

그렇다면 정말!!!! 시각화를 시작 해볼까요?

아래와 같은 시각화를 완성해 보세요. (어렵게 느껴지신다면, 시각화를 위한 HINT를 참조하세요.)

각 시각화가 의미하는 것은 시각화 제목에 표현되어 있습니다!

사용 Data = 스타벅스 메뉴

[시각화를 위한 HINT]

1. **카테고리와 칼로리, 카페인**을 행 선반에 올려야 할 지, 열 선반에 올려야 할 지 고민해 보세요. (행 선반은 Y축, 열 선반은 X축)

2. **칼로리와 카페인의 집계 형태를 평균으로 변경**해주세요.

왜 집계를 평균으로 변경해야 할까요?

하나의 카테고리 안에는 여러 개의 메뉴가 있습니다. (예: 브루드 커피 카테고리에는 아이스 커피와 오늘의 커피 메뉴가 있죠.)

категори를 기준으로 합계로 집계해주면 카테고리 안에 있는 메뉴들의 칼로리와 카페인 값이 모두 더해져서,

카테고리 별로 메뉴들의 총 합계 칼로리, 총 합계 카페인 값이 보여지겠죠.

우리는 카테고리 별로 평균 칼로리와 평균 카페인 값을 보는 것이 좋을 것 같습니다.

3. 평균 칼로리와 평균 카페인을 표현하기 위해 각각 2개의 축이 생겼을텐데요.

평균 카페인을 표현하는 오른쪽 막대에만 “평균 카페인”을 이용해 색상을 표현해 주세요.

(색상을 표현할 때는 카페인 값이 높을수록 붉은색을 띠도록 해주세요.)

4. 평균 칼로리를 기준으로 정렬해 주세요. (정렬을 해야하는 이유 이제는 다 아시죠?)

우리가 처음 던졌던 질문(칼로리가 낮으면서 카페인 값이 적은 카테고리는 무엇일까)에 대한 답을 찾으실 수 있으셨나요?

그럼 좀 더 시각화를 해볼까요?

2. 메뉴명 별 칼로리 & 카페인

1번에서는 카테고리 수준에서 칼로리와 카페인을 나타냈다면, 더 아래 수준인 메뉴명 수준까지 칼로리와 카페인을 함께 표현해 보아요!

이번 시각화에서 사용할 시각화는 트리맵입니다.

트리맵은 계층 구조의 데이터를 표시하는데 적합한 시각화로 전체 대비 부분의 비율이 얼마나 되는지 비교하는데 많이 사용됩니다.

사각형의 크기와 색상에 따라 데이터의 패턴을 확인할 수 있을 뿐만 아니라 많은 데이터를 한 번에 볼 수 있다는 장점이 있는데요.

이번 시각화에서는 메뉴명 수준으로 전체 데이터를 한 번에 보일 수 있도록 트리맵을 사용했습니다!

아래와 같이 **칼로리는 사각형의 크기로, 카페인**은 **색상**으로 표현해 보세요.

[시각화를 위한 HINT]

1. 마크의 유형을 사각형으로 변경해주세요.
2. 칼로리를 마크의 크기에 놓아주세요.
3. 카페인을 마크의 색상에 놓아주세요.
4. 메뉴명을 마크의 레이블에 놓아주세요.
5. 마크의 색상을 클릭해서 색상 편집을 클릭해 주세요.
6. 색상을 아래와 같이 변경해 주세요.

그런데, 왜 이번 시각화에서는 칼로리와 카페인의 집계를 평균으로 변경하지 않았을까요?

태블로는 화면에 포함된 차원에 따라서 측정값을 집계한다고 설명을 드렸어요.

- 1번 과제에서는 카테고리를 기준으로 칼로리와 카페인의 평균 값을 계산 했다면,
2번 과제에서는 메뉴명을 기준으로 칼로리와 카페인 값을 계산하는 것이죠.

그런데 현재 데이터의 가장 낮은 행 수준이 메뉴명이예요. 즉, 메뉴명은 유일하게 구분되고 중복되지 않는 값이죠.
따라서 하나의 메뉴명에는 하나의 칼로리, 하나의 카페인 값이 있어요. 아래 데이터를 보면 아시겠죠?

즉, 메뉴명을 기준으로 칼로리, 카페인 값을 합계로 계산하나 평균으로 계산하나 결과는 동일합니다.
따라서 별도로 집계를 변경해주지 않은 것입니다.

3. 카테고리과 메뉴명을 한 번에 살펴보기

카테고리 기준으로 만든 시트와 메뉴명 기준으로 만든 시트를 한 번에 살펴 봅시다.
아래와 같이 카테고리에 마우스 오버하면, 해당 카테고리에 해당되는 메뉴를 살펴보도록 합니다.

태블로 기능 중 도구 설명을 이용해 볼게요.

1. 카테고리 별 평균 칼로리&카페인 시트로 이동해 주세요.
2. 평균(칼로리(Kcal)) 마크 카트를 클릭하고, 도구 설명 클릭해 주세요.

3. 도구 설명 편집 창에 추가할 시각화에 대한 설명을 입력해 주세요.

4. 삽입 클릭 > 시트 > 메뉴 별 칼로리&카페인 선택

5. 넓이(maxwidth)와 높이(height) 조정

6. 확인 클릭

4. 당분 함유량과 칼로리 상관관계

당분 함유량이 높을 수록 칼로리가 높을까요? 당분과 칼로리의 상관관계를 살펴봅시다!

이번 시각화에서는 스캐터 플롯 (산점도) 시각화를 이용해 볼 예정이에요.

스캐터 플롯은 2개의 연속형 데이터에 대한 상관관계를 분석하는데 가장 많이 사용되는 시각화 입니다.

두 개의 축으로 데이터가 얼마나 퍼져 있는지 분포를 살펴 볼 수 있고, 상수 라인 / 평균 라인 / 사분위수 및 중앙값 / 추세선 등과 같은 참조 라인을 추가하여 값의 분포를 비교하기에도 유용합니다.

아래와 같은 시각화를 완성해 보세요. (어렵게 느껴지신다면, 시각화를 위한 HINT를 참조하세요.)

사용 Data = 스타벅스 메뉴

[시각화를 위한 HINT]

1. **당류**와 **칼로리**를 각각 알맞게 열 선반과 행 선반에 위치 시켜보세요.

- 당류는 X축, 칼로리는 Y축에 놓아야겠죠?

2. **메뉴명 수준(기준)**으로 당류와 칼로리를 살펴봅시다.

1번 스텝을 진행하셨다면 현재 시각화에는 전체 메뉴의 합계 당류와 합계 칼로리가 나타나는 것을 볼 수가 있습니다.

메뉴명을 마크 선반의 세부 정보에 가져다 놓으세요. 시각화의 집계 기준이 전체에서 메뉴명 수준으로 변경됩니다.

3. **마크의 형태**를 “**원**”으로 변경해 주세요.

4. **카페인**으로 **마크(원)의 색상**과 **크기**를 표현하세요.

겹쳐지는 제품들을 함께 볼 수 있도록 **색상의 불투명도를 조정**하고, **테두리를 추가**해 명확하게 볼 수 있도록 해보세요.

당류가 높을 수록 칼로리가 높은 것을 볼 수 있고, 비슷한 당류가 들어가도 칼로리가 메뉴에 따라 달라진다는 것을 알 수가 있네요.

5. 시군구 별 매장 분포 현황

어디 시군구에 스타벅스가 가장 많을까요? 시군구 별 매장의 분포를 맵을 이용해 살펴봅시다.

사용 Data = 스타벅스 매장 정보

[시각화를 위한 HINT]

1. 데이터 창에서 사용 데이터가 “스타벅스 매장 정보”로 선택되었는지 확인해 주세요.

2. 시도, 시군구 필드를 이용해서 매장이 존재하는 시군구를 표현해봅시다.

3. 마크의 크기와 색상을 스타벅스 매장 수로 나타내보세요.

스타벅스 매장수는 “매장코드”를 카운트 해서 크기와 색상에 표현하면 되겠죠?

6. 대시보드 만들기

1~3번 과제에서 만든 ‘카테고리 별 평균 칼로리와 평균 카페인’, ‘당분 함유량과 칼로리 상관관계’ 시트를 이용해 대시보드를 만들어 봅시다!

[추가 도전 과제]

1. 칼로리가 낮으면서 카페인이 높지 않은 메뉴를 마시고 싶을 때, 어떤 시각화로 답을 찾아볼 수 있을까요?

위에서 만든 차트와 유사하게 메뉴별 칼로리와 카페인의 상관관계를 보여주는 차트를 만들면 쉽게 볼 수 있습니다.

이 문제는 2일차에서 함께 진행해 볼 예정이니 자유롭게 시각화 해보세요.

2. “스타벅스 매장 정보” 데이터에 있는 경도/위도 데이터를 이용해서 아래와 같이 서울시의 실제 매장 위치를 표현해 보세요.