

통합과 조인의 차이점 비교

데이터 통합과 교차 데이터베이스 조인 비교에 대한 동영상에 오신 것을 환영합니다.

예제 통합 문서를 다운로드하여 사용자의 시스템에 설치된 Tableau 에서 직접 따라 해보실 수 있습니다.

간단히 말해 교차 데이터베이스 조인은 두 데이터 원본을 행 수준에서 통합하여 단일 데이터 원본을 만듭니다. 통합하면 각 데이터 원본에서 쿼리 결과를 분리하고 이를 뷰에 집계합니다. 교차 데이터베이스 조인과 데이터 통합에 대해 자세히 알아보려면 각각의 동영상을 시청하시기 바랍니다.

일대다 관계

시작을 위해 교차 데이터베이스 조인 동영상에서 사용한 제품 2016 과 매출 2016 의 데이터 원본 통합을 설정합니다. 이제 시장별 매출 범주를 간단히 분석해 보겠습니다.

- **제품** 데이터 원본에서 카테고리를 행으로 가져온 다음
- **매출** 데이터 원본으로 전환합니다.
 - 제품 ID 별 링크를 클릭하여 통합 관계를 설정하고
 - 매출을 열로 가져옵니다.
- 통합이 완료되어 이제 범주별로 전반적인 매출을 파악할 수 있습니다.
- 시장에 색상을 추가하면 어떻게 될까요? 대부분의 막대가 회색, 즉 별표입니다. 이유가 무엇일까요?

바로 시장과 일대다 관계에 있는 제품 ID 를 기준으로 통합했기 때문입니다. 열은 청록색 막대는 USCA 시장에만 판매된 제품 ID 를 표시하지만 모든 회색 막대는 여러 시장에 판매된 제품이기 때문에 별표가 있는 집계로 표시됩니다. 각 데이터 원본의 결과를 별도로 집계하고 뷰에 함께 표시하는 통합 수행 방식 때문에 일대다 관계에 있는 경우 데이터를 쉽게 시각화할 수 없습니다.

통합보다 교차 데이터베이스 조인 방법이 더 적합한 사례에 해당합니다. 이 데이터 원본은 제품 ID 를 기준으로 행 수준에서 조인됩니다. 첫 번째 뷰를 다시 만들어 보겠습니다.

- 범주를 행으로
- 매출을 열로
- 시장을 색상으로

이제 시장별 매출 범주를 분리하여 원하는 결과를 얻었습니다. 데이터를 조인할 때 매출 정보에 제품 정보를 단순히 추가했기 때문에 매출의 원래 레코드 수는 변경되지 않았습니다.

행 수를 늘리는 조인

그렇다면 교차 데이터베이스 조인이 항상 최고의 방법일까요? 다른 데이터 집합으로 전환해 보겠습니다.

이번에는 데이터 통합 동영상의 Office City 및 Coffee Chain 을 사용합니다. 이들은 동일한 회사에서 소유한 두 개의 소매 체인점입니다. Office City 는 미국 50 개 주 전체에서 영업하고 Coffee Chain 은 20 개 주에서 영업합니다. 이제 50 개 주 전체에서 두 조직의 결합된 매출을 본다고 가정해 보겠습니다.

여기에 공통 필드인 주를 기준으로 조인된 데이터 원본이 있습니다.

- 이 데이터 원본에 50 개 주가 모두 포함되어 있으므로 이 원본 내 Office City 연결에서 주를 가지고 옵니다.
- 마우스 오른쪽 버튼을 클릭하고 '추가된 매출'계산을 편집하면 Office City 매출 및 Coffee Chain 매출이 추가되고,
 - Zero Null 함수를 사용하여 Coffee Chain 에 매출이 없는 주의 경우 null 이 0 으로 처리됩니다.
- 그런 다음 추가된 매출을 열로 가져옵니다.
- 대단하죠? 몇 개 주에 수억 달러의 매출이 있다니 호황 사업이군요. 결국 이 값이 맞지 않다는 것을 알게 됩니다. 좀 더 자세히 살펴보겠습니다.

이 대시보드는 4 개 주의 Office City 매출과 동일한 4 개 주의 Coffee Chain 매출을 표시합니다. 결합된 매출은 중앙의 수식에 표시된 것, 즉 값을 더한 것이어야 합니다.

- 통합(노란색) 및 결합된 매출은 올바른 답을 제공하지만 조인된 데이터 원본(보라색) 값이 과도하게 부풀려졌습니다.

Office City, Coffee Chain 및 조인된 데이터 원본의 레코드 수를 살펴보면 이 값이 확실히 부풀려져 있음을 알 수 있습니다.

주를 기준으로 조인했기 때문에 Tableau 는 Coffee Chain 의 제품 및 날짜와 Office City 의 행 ID 라는 고유 정보를 사용하여 각 주를 결합한 행을 만들었습니다. 이러한 교차 제품은 매출 값을 여러 번 계산했음을 의미합니다.

이와 반대로 통합을 사용하면 주에 대해 하나의 데이터 원본의 매출 값과 해당 주에 대한 다른 데이터 원본의 매출 값을 반환했기 때문에 올바른 값을 제공했으며, 이 값을 더하기만 하면 되었습니다. 이 경우가 원하던 동작입니다.

여기에서 알 수 있듯이 어떤 필드를 조인 또는 통합할지 여부는 분석에 큰 영향을 미칩니다. 올바른 필드를 선택하고 이러한 통합이나 조인이 생성되는 방식을 이해해야만 무엇을 언제 사용할지 잘 알 수 있습니다.

각각의 이점

교차 데이터베이스 조인을 수행하여 통합된 데이터 원본을 만드는 경우 많은 이점이 있습니다.

- 조인은 행 수준으로 수행되므로 데이터 통합에서 직면하는 일대다 제한이 없으며, 따라서 별표가 발생하지 않습니다.
- 단일 데이터 원본을 추출, 저장, 게시할 수 있으며 데이터 통합과 달리 다른 사용자와 손쉽게 공유할 수 있습니다. 다른 데이터 원본과 마찬가지로 사용 가능합니다.

하지만 필드 조인의 세부 수준이 올바르지 않은 경우 조인을 수행하면 데이터 집합의 크기가 부자연스럽게 부풀려지고 잘못된 정보를 제공할 수 있습니다.

데이터 통합은 각 데이터 원본에 별도의 쿼리를 보내고 뷰에 원하는 세부 수준으로 집계하여 분석에 필요한 정보를 정확히 생성할 수 있습니다.

결론

통합 또는 교차 데이터베이스 조인을 사용한 데이터 원본 결합에 대한 동영상을 시청해 주셔서 감사합니다. 무료 교육 동영상을 통해 Tableau 사용에 대해 자세히 알아보시기 바랍니다.