

PDF에 연결

PDF 커넥터에 대한 동영상에 오신 것을 환영합니다. PDF를 다운로드하여 설치된 Tableau에서 직접 따라 해보실 수 있습니다.

PDF에는 Tableau 분석이 필요한 중요 데이터가 포함될 수 있습니다. Tableau의 PDF 커넥터는 테이블에서 데이터를 추출할 수 있도록 설계되었습니다. PDF에는 데이터에 대한 메타데이터가 부족하기 때문에, 분석하기에 앞서 몇 가지 게시물 연결 작업을 해야 할 수 있습니다.

PDF 데이터 테이블에 연결

Tableau는 각 행에 한 줄의 데이터가 있는 행과 열로 구성된 PDF에서 크로스탭 데이터 테이블을 읽을 수 있습니다. 이 문서 2페이지의 주식 데이터부터 살펴보겠습니다. Tableau에 연결해 보겠습니다. 연결 패널에서 PDF 파일을 선택하면 페이지 선택 도구가 표시됩니다. 전체 문서, 특정 페이지 또는 페이지 범위를 확인할 수 있습니다. 2-8페이지를 살펴보겠습니다. 그리고 첫 번째 페이지를 캔버스에 가져오겠습니다. PDF 테이블 머리글의 라인 중 하나가 Tableau 작업을 혼란스럽게 하지만 데이터 해석기를 사용하면 원하는 머리글을 얻을 수 있습니다.

Union

각 페이지에 테이블이 별도로 표시되지만 동일한 열 제목을 사용하여 같은 방식으로 구성되기 때문에 Union을 사용하면 손쉽게 통합할 수 있습니다. 이미 캔버스에 있는 테이블의 경우, 원하는 다른 페이지의 테이블을 첫 번째 테이블 아래의 Union 드롭 영역으로 끌어서 이동합니다. 테이블 이름에 해당하는 새 열이 표시되기 때문에 아래로 스크롤하면 3페이지의 데이터가 통합되어 2페이지에 완벽하게 정렬되어 있음을 확인할 수 있습니다.

불완전한 테이블 정리

이 PDF는 데이터를 가져오기에 효율적으로 구성되었지만 모든 PDF가 효율적으로 구성되어 있는 것만은 아닙니다. 일반적으로 Tableau는 라인당 하나의 데이터 행, 계층이 없거나 중첩되지 않은 머리글, 하위 테이블이 없는 등의 테이블 구조를 갖춘 PDF에 가장 잘 연결합니다. 그러나 작업 시 PDF가 꼭 완벽할 필요는 없습니다. 다른 파일에 연결해 보겠습니다. 첫 번째 파일만큼 쉽지 않습니다. 14페이지의 테이블을 참조하십시오. (참고 사항: Tableau는 절대적인 페이지 번호를 보여줍니다. 문서에 매겨진 페이지 번호와 일치하지 않을 수도 있습니다.)

또 다른 데이터 소스를 추가해 보겠습니다. 14페이지를 참조하십시오. 데이터 연결 드롭다운 아래의 'PDF 다시 스캔' 옵션을 사용하면 살펴볼 페이지를 다시 선택할 수 있습니다. 14페이지에는 테이블이 1개만 있지만 왼쪽에는 3가지 옵션이 표시됩니다. Tableau에서 이 테이블을 가져올 수 있는 3가지 방법을 인식했기 때문입니다. 한 번에 테이블 하나씩 가져오는 경우 그 내용을 확인할 수 있습니다. (참고 사항: 이전 작업처럼 여러 페이지의 테이블을 통합하는 경우, 같은 페이지의 여러 버전이 아닌 각 페이지에서 정확한 테이블 버전을 통합해야 합니다.)

테이블 1을 가져와 데이터 해석기를 사용해 보겠습니다. 테이블 1에 모든 정보가 있는 것처럼 보이지만 어떤 이유에서인지 1995-1997년도 데이터가 단일 열로 읽혀지고 있습니다. 이 버전의 데이터를 사용하려는 경우, 필드에서 사용자 지정 분할을 수행하고 공백에서 모든 열을 분할하여 데이터를 정리한 다음, 분할된 필드의 이름을 적절한 연도로 바꿉니다.

테이블 2와 3의 내용을 살펴봅시다. 테이블 2는 원래 테이블의 하단처럼 보이고, 테이블 3은 원래 테이블의 상단처럼 보입니다. 테이블 2와 3에 사용된 열 구분을 좀 더 선호하므로 이 테이블로 작업해 보겠습니다. 먼저, 테이블 2를 밖으로 끌어서 테이블 3 아래로 이동하여 이 두 테이블을 통합해 보겠습니다. '유입'과 F1 열 사이에 불일치가 존재합니다. 두 항목을 모두 선택하고 '불일치 필드 병합'을 선택하면 이제 해당 열이 올바르게 표시됩니다. 이 열의 이름을 '식수원'으로 변경합니다.

Null 값 처리

'저장 공간 변경'과 같은 하위 머리글을 데이터 행으로 읽거나 '수력 발전소에서 취수'와 같은 머리글을 2개의 서로 다른 행으로 읽기 때문에 Null을 가진 여러 행이 생겨날 수 있습니다. 이러한 Null을 제거하기 위해 오른쪽 상단에서 추가를 클릭하여 데이터 원본 필터를 추가하겠습니다. 필터를 추가하고 이러한 Null을 가진 모든 열을 선택합니다. 여기서는 F10을 사용해 보겠습니다. 확인을 클릭하고 Null과 제외를 차례로 선택합니다. 그런 다음 확인을 두 차례 클릭합니다. 이제 해당 행이 사라지고 데이터만 남게 됩니다.

하지만 실제 합계에 포함되어야 할 여러 유형의 식수원도 존재하기 때문에 실행을 취소하겠습니다. 해당 값과 Null을 필터링으로 한 번에 제거하려고 합니다. 필터 > 추가로 이동한 다음 식수원을 선택합니다 원본 PDF의 머리글 또는 합계나 값에 대한 Null 행에 해당하는 항목을 확인합니다. 수력 발전과 같은 머리글의 경우 한 번은 데이터로 또 한 번은 Null로 표시되므로 데이터가 삭제되지 않도록 지금은 그대로 두겠습니다. 제외와 확인을 차례로 클릭합니다.

머리글 고정 및 피벗

이제 수력 전기에 해당하는 데이터 행과 Null 행을 확보했지만, 잠시 이러한 행을 정리해보도록 하겠습니다. 첫 번째 열을 제외하고는 어떠한 머리글도 가지고 있지 않지만 원본 PDF의 값을 상호 참조하여 어떤 머리글이 이에 해당하는지 알 수 있습니다. 1995년부터 2010년에 해당하는 열이 표시되어야 합니다. 좀 더 빨리 살펴보도록 하겠습니다.

데이터에 피벗 기능을 사용하여 연도에 해당하는 열 및 백만 세제곱 미터에 해당하는 열을 확인할 수 있습니다. 테이블 이름 열은 숨기고 연도별 날짜 데이터 유형을 변경합니다. 백만 세제곱 미터의 데이터 유형을 숫자(정수)로 변경하고, 이 데이터 유형은 마지막 열 유형에 해당되므로 이 열에서 모든 Null을 필터링할 수 있습니다. 필터 수정 > 추가로 이동하여 백만 세제곱 미터를 선택한 다음 Null 제외를 선택합니다.

필드 멤버 재별칭

식수원 이름과 관련하여 몇 가지 문제가 있습니다. 해당 필드의 메뉴를 불러온 다음 별칭을 클릭합니다. 여기에서 주어진 필드 멤버를 다시 재별칭할 수 있습니다. 별칭을 더블 클릭하고 바꾸려는 이름을 입력합니다.

- 수력 발전소에서 방출
- 지하수
- 수력 발전 취수

그룹 및 계층 다시 만들기

식수원 카테고리과 같이 원본 PDF 테이블에 구조가 있었던 기억이 납니다. 시트 1을 클릭하여 데이터 패널에서 이러한 구조를 만들 수 있습니다. 먼저 백만 세제곱 미터는 실제 측정값이므로 이 아래로 끌어오겠습니다.

여기서도 그룹으로 다시 만들 수 있습니다. 식수원을 열로 가져오고 첫 번째 그룹으로 만듭니다. Control 키를 누른 상태에서 같은 그룹으로 묶을 멤버를 클릭합니다.

- 수력 발전 취수
- 수력 발전소에서 방출 및
- 증발
- 바다로 방출 및 순수 취수

도구 설명의 종이 클립 아이콘을 사용하여 그룹화해보겠습니다. 반복하고 다음 항목을 다시 그룹으로 만듭니다. 지하수, 얼음, 호수 및 저수지, 눈, 토양 수분

데이터 패널에서 이 새로운 필드를 마우스 오른쪽 버튼으로 클릭한 다음 그룹을 편집합니다. 먼저, 강수량을 클릭하여 그룹으로 만듭니다. 그룹으로 묶을 항목이 강수량 하나일지라도 자체 그룹으로 만들 수 있습니다. 이제 유입, 저장 공간 변경, 유출로 이름을 변경해 보겠습니다. 이제 그룹으로 묶은 필드 카테고리 이름을 지정해 보겠습니다. 식수원을 이 카테고리 필드의 상단으로 끌어와서 계층을 만들 수 있으므로 뷰에서 드릴다운할 수 있습니다.

PDF 작업할 때 유용한 도움말

주어진 PDF 게시물 연결을 정리하는 정밀한 단계는 다를 수 있지만, 이 동영상에서는 분석 준비를 위한 데이터 정리에 유용한 몇 가지 도구를 소개했습니다. 참고로, Tableau에서 하위 테이블, 머리글의 계층 구조, 단일 행으로 해석되어야 하는 여러 콘텐츠 행 등이 포함된 PDF를 처리할 때 문제가 발생할 수 있습니다. 마지막으로, PDF가 셀 및 데이터 테이블로 구문 분석되어야 하므로 색상과 음영은 데이터 해석 방식에 따라 변경될 수 있습니다.

결론

PDF 커넥터 교육 동영상을 시청해 주셔서 감사합니다. 계속해서 무료 교육 동영상을 통해 Tableau 사용에 관해 자세히 알아보시기 바랍니다.