

1 Lab5. Tableau를 이용한 Data Source 연동하기 II

※ Refer to <https://datatalking.tistory.com/7>

3

1. 다양한 데이터 연결하기

1) Tableau에서 [파일] > [새로 만들기]를 클릭하여 새 Tableau 문서를 엽니다.

2) 화면 왼쪽의 [데이터] > [데이터에 연결] 링크를 클릭한다.

3) [연결] 팝업창에서 왼쪽의 메뉴 중 [저장된 데이터 원본] > [슈퍼스토어 - 샘플]을 클릭한다.

-Orders, People, Returns 가 보인다.

9

4) 메뉴바의 제일 왼쪽의 Tableau 그림을 클릭하면 다시 [연결] 페이지가 나타난다.

5) 이번에는 동일한 파일이지만 로컬에 저장된 파일을 열기 위해 [연결]페이지의 [파일에 연결] > [Microsoft Excel]을 선택한다.

6) 다운로드 받은 Sample - Superstore.xls를 선택한다.

-Orders, People, Returns 가 보인다.

14

7) 다시 메뉴바의 Tableau 그림을 클릭하여 [연결] 페이지로 이동한다.

8) 이번에는 [서버에 연결] > [자세히...]를 클릭한다.

9) 목록에서 [Google Drive]를 선택한다.

10) 계정을 선택한다.

11) Google Drive에서 인증을 허용한다. 이렇게 해서 Google Drive에 인증이 만들어졌다.

12) Chrome창을 닫으면 Tableau의 [연결] 화면에 [Google Drive에 연결]이라는 팝업창이 나타난다.

13) 여기서 [내 드라이브] > [Tableau 교육] > Sample - Superstore.xls를 선택한다.

14) 약간의 시간이 흐른 뒤 다음과 같이 동일한 시트들을 확인할 수 있다.

-Orders, People, Returns 가 보인다.

24

15) 이번에는 [연결] > [서버에 연결] > [자세히...] > [Dropbox]를 선택한다.

10) Windows의 경우에는 [방화벽]의 경고창이 보일 수 있다. [엑세스 허용]을 클릭한다.

11) 아직 Dropbox에 로그인되어 있지 않으면 먼저, 로그인하라고 나온다.

12) Dropbox에 로그인한다.

13) Google Drive와 동일하게 [허용]을 클릭하여 인증이 만들어졌다.

14) Tableau [연결] 화면에 [Dropbox에 연결] 팝업창이 나타난다.

15) 여기서 [Tableau 교육] > Sample - Superstore.xls를 선택한다.

16) 약간의 시간이 흐른 뒤 다음과 같이 동일한 시트들을 확인할 수 있다.

-Orders, People, Returns 가 보인다.

34

17) 마지막으로 Microsoft OneDrive를 연결해 본다.

18) [연결] > [서버에 연결] > [자세히...] > [OneDrive 및 SharePoint Online]를 선택한다.

19) Microsoft에 로그인한다.

20) "이 앱이 회원님의 정보에 액세스 할 수 있도록 허용하시겠습니까?" 팝업창에서 [예]를 선택한다.

21) Chrome에서 인증이 됐다는 페이지가 나온다.

22) Tableau [연결] 화면에 [OneDrive 및 SharePoint Online에 연결] 팝업창이 나타난다.

23) 여기서 [Tableau 교육] > Sample - Superstore.xls를 선택한다.

24) 약간의 시간이 흐른 뒤 다음과 같이 동일한 시트들을 확인할 수 있다.

-Orders, People, Returns 가 보인다.

44

45

2. 다양한 데이터 준비하기

1) Sample - Superstore.xls를 Tableau에서 Open했으면, 화면 좌측의 [시트]의 Orders 테이블을 Drag & Drop 한다.

2) 전체 9994행 중 100개 행이 보여진다.

3) Order 테이블을 더블클릭하여 조인 캔버스(물리적 계층)을 엽니다.

4) 화면 좌측의 [연결] 옆에 [추가]를 클릭한다.

5) 이렇게 하면 굳이 메뉴바의 Tableau 아이콘을 통해 [연결] 페이지로 이동하지 않고, 바로 특정 연결을 수행할 수 있다.

6) [연결 추가] > [파일에 연결] > [Microsoft Excel]을 선택한다.

7) 로컬에 저장한 Sample - Superstore.xls를 선택하고 [열기]를 클릭한다.

8) 같은 파일이지만 하나는 서버로부터 연결된 파일이고, 하나는 로컬에 저장되어 있던 파일이다.

9) Tableau에서는 이 기중간의 테이블 연결이 가능하다.

10) Sample - Superstore는 OneDrive 및 SharePoint Online에서 로딩한 파일이고, Sample - Superstore(2)는 로컬에 저장되어 있던 파일이다.

11) Sample - Superstore에서 현재 Orders 테이블이 로딩되어 있으면, Sample - Superstore(2)에서는 Returns 테이블을 로딩해보자.

12) 약간의 시간이 걸리지만 두 테이블이 서로 조인되는 것을 확인할 수 있다.

13) 2개의 테이블 사이에 조인 아이콘이 있다.

14) 조인 아이콘을 클릭해보면 현재 두 개의 테이블이 각각 어떤 컬럼으로 연결되어 있는지를 보여준다.

15) 두 개의 테이블이 현재 Order ID를 기준으로 연결되어 있는 것을 볼 수 있다.

16) Orders 테이블은 하늘색 라벨, 그리고 Returns 테이블이 주황색 라벨이라면 데이터가 보여지는 화면 아래쪽의 표에서 확인해보면 Profit까지는 Orders

소속이고, 나머지 2개의 컬럼 즉, Returned와 Order ID(Returns)만 Returns 테이블 소속임을 색상으로도 확인할 수 있다.

17) 하지만 제일 마지막 컬럼인 Order ID(Returns)는 이미 Orders 테이블에도 Order ID가 있기 때문에 굳이 보여질 필요가 없다.

18) Order ID(Returns) 컬럼의 아래 화살표 > [숨기기]를 클릭하여 해당 컬럼을 숨긴다.

65

19) 화면의 오른쪽 상단에 보면 [연결] > [라이브]와 [연결] > [추출]이 있다.

20) [라이브]는 실시간 데이터 테이블과 연결되어 있기 때문에 신속한 업데이트가 되지만, 아무래도 테이블에 무리가 될 수 있다.

21) [추출]은 현재 데이터 테이블의 값을 Caching한 후, 그 Caching을 로컬 메모리에 저장하기 때문에 빠른 처리 속도, 그리고 모든 종류의 태블로 기능을 사용하기 가능하다는 장점이 있다.

22) 하지만, [추출] 연결은 상대적으로 느린 데이터 업데이트 주기가 될 것이다.

70

23) 화면 우측 상단에 보면 [필터]가 있다.

24) [필터]는 전체 데이터를 모두 로딩하기 보다 일부분만 필터링해서 테이블에 무리가 되지 않게 로딩하는 방법이다.

25) [필터] > [추가]를 클릭한다.

26) [데이터 원본 필터 편집]창이 보인다. 한번 더 [추가...]을 클릭한다.

27) [필터 추가]창의 [필드 선택]에서 Order Date를 선택하고 [확인]을 클릭한다.

28) [필터 필드] [필드 필드 [Order Date]] 창에서 "[필터 필드 [Order Date]]에 대해 어떤 방법으로 필터링하시겠습니까?"를 [년]을 선택함으로써 년도 기준으로 필터링하려고 한다.

29) [년]을 선택하고 [다음]을 클릭한다.

30) [필터 [Order Date]의 연도] 창이다. 여기서 2018년만 체크하고 [확인] 버튼을 클릭한다.

31) [데이터 원본 필터 편집] 창에서 [확인]을 클릭한다.

32) 결과적으로 2018년 한 해의 결과값만 로딩한 것을 확인할 수 있다.

33) 이것은 대용량 데이터 취급 시 적절하게 사용한다면 매우 유용할 것이다.

81

82
83
84 3. 데이터 작업하기
85 1)데이터 원본을 태블로에 연결하면 컬럼에 따라 필드 역할과 유형이 결정된다.
86 2)역할(Role)
87 -필드는 차원 또는 측정값으로 역할이 부여된다.
88 3)유형(Type)
89 -필드의 유형은 자동으로 문자열, 정수, 날짜와 같은 타입이 설정된다.
90 -단, 해당 유형은 다른 유형으로 변경이 가능하다.
91 4)측정값
92 -기본적으로 숫자형식, 액션을 통해 집계 되면서 한 덩어리로 만들어진다.
93 5)차원
94 -한 덩어리의 측정값을 어떻게 나누어서 볼 것인지를 결정
95
96 6)질병관리청(<https://ncov.kdca.go.kr/>) > 발생현황 > (누적)확진 다운로드(코로나바이러스감염증-19_확진환자_발생현황_230525.xlsx)
97
98 7)Tableau에서 [데이터 연결] > [파일에 연결] > [Microsoft Excel] > 코로나바이러스감염증-19_확진환자_발생현황_230525.xlsx > [열기]
99 8)데이터 로딩 후 화면 좌측 [시트] > 시도별 발생(17개시도+검역) 테이블을 Drag & Drop
100 9)[필드]를 보면 테이블 데이터의 각 컬럼의 유형을 확인할 수 있는데, 일자는 날짜형, #은 숫자형, Abc는 글자형을 나타내고 있다.
101 10)DataGrid를 보면 현재 [일자]의 값이 null인 것을 확인할 수 있다.
102 11)실제로 Excel로 파일을 열어보면 일자(A5)가 한글로 누적(명)으로 되어 있어서 null로 표시된 것이다.
103 12)이 컬럼을 제외하기로 한다.
104 13)페이지 우측 상단의 [필터] > [추가]를 클릭한다.
105 14)[데이터 원본 필터 편집]창이 보인다. 한번 더 [추가...]을 클릭한다.
106 15)[필터 추가] 창이 [필드 선택]에서 "일자"를 선택하고 [확인] 버튼 클릭
107 16)[필터 필드 [필터 필드 [일자]] 창에서 "날짜 범위"를 선택하고 [다음]을 클릭
108 17)[필터 [일자]] 창에서 [특수]를 선택하고 [Null이 아닌 날짜]를 선택하고 [확인] 클릭
109 18)[데이터 원본 필터 편집] 창에서 한 번 더 [확인]을 클릭한다.
110 19)그러면 null로 표기되어 있던 행이 사라진다.
111 20)Tableau는 가로 방향으로 시도를 표기하고 있다.
112 21)이것을 세로 방향으로 피벗
113 22)계(명)을 선택하고 Shift키를 누르면서 제일 왼쪽의 컬럼인 검역을 선택하면 컬럼 영역 선택이 된다.
114 23)검역의 아래 화살표 > [피벗]을 선택한다.
115 24)이렇게 하면 세로 방향으로 각 시도의 값들이 정렬된다.
116 25)[피벗 필드명]에 "계(명)"을 제외하기로 한다.
117 26)페이지 우측 상단의 [필터] > [편집]을 클릭한다.
118 27)[데이터 원본 필터 편집]창이 보인다. 한번 더 [추가...]을 클릭한다.
119 28)[필터 추가] 창에서 [필드 선택] > [피벗 필드명]을 선택하고 [확인] 클릭
120 29)[필터 [피벗 필드명]] 창에서 [일반] 탭에서 "계(명)"을 체크하고, [제외(X)]를 체크하고 [확인] 클릭
121 30)[데이터 원본 필터 편집] 창에서 한 번 더 [확인]을 클릭한다.
122 31)변경된 "피벗 필드명"을 더블클릭하여 "시도"로 변경한다.
123 32)DataGrid 좌측의 [필드]를 보면 [원격 필드명]이 "피벗 필드명"인데, [필드명]이 "시도"로 변경됐음을 확인할 수 있다.
124 33)이번에는 DataGrid의 "피벗 필드 값"을 더블클릭하여 "확진자"로 변경한다.
125 34)"확진자"의 유형은 Abc 즉 글자형이기 때문에 유형 변경을 위해 Abc를 클릭하고 숫자(정수)를 선택한다.
126 35)DataGrid 좌측의 [이름]을 "시도별 확진자"로 변경한다.
127 36)화면 상단의 "시도별 발생(17개 시도...)"를 클릭하여 "코로나 19 확진자 현황"으로 수정한다.
128
129 37)화면 하단의 [시트 1]로 이동한다.
130 38)화면 좌측의 [테이블] > [확진자]의 right-click하여 [측정값으로 변환]을 선택하여 측정값으로 변환한다.
131 39)확진자를 [행]으로 Drag & Drop 한다.
132 40)즉, 누적 확진자의 값인데, 마우스를 올리면 그 값이 표시된다.
133 41)[테이블] > [시도]를 더블클릭한다.
134 42)그러면 누적 확진자를 시도별 막대 그래프로 표시한다.
135 43)세로 막대 그래프를 가로 막대 그래프로 변경하려면 톨바에서 [행과 열 바꾸기]를 클릭하면 된다.
136 44)그 버튼 바로 오른쪽의 소팅 버튼을 클릭하여 오름차순으로 혹은 내림차순으로 가로 막대 그래프를 변경이 가능하다.
137 45)톨바의 [표준]을 [전체 보기]로 변경해 보자.
138
139 46)화면 좌측의 [테이블] > [시도]를 [마크] > [색상]으로 Drag & Drop 하면, 각 시도별로 색상이 다르게 보여진다.
140 47)[시트 1]의 이름을 "시도별 확진자"로 변경한다.
141 48)방금 변경한 "시도별 확진자" 시트 right-click 하여 [복제]를 선택한다.
142 49)여기서 [마크]에서 시도를 Drag out하여 제거하고, 화면 좌측의 [테이블] > [확진자]를 [마크] > [색상]으로 Drag & Drop 해보자.
143 50)이렇게 하면 최소 확진자부터 최대 확진자까지 그라데이션으로 색상을 표현하고 있다. 화면 우측의 합계(확진자) 범례 참조.
144 51)[마크] > [색상]을 클릭하고 [색상 편집...]에서 기본 파란색을 붉은 색으로 변경하면, 그라데이션의 막대 그래프 색상이 붉은 색 계열로 변경된다.
145
146 52)[새 워크시트] 클릭하여 새로운 시트를 생성한다.
147 53)화면 좌측의 [테이블] > [확진자]를 더블클릭하여 시트에 막대그래프로 표시한다.
148 54)화면 좌측의 [테이블] > [확진자]를 더블클릭하여 시트에 막대그래프를 년도별 라인그래프로 변경된다.
149 55)년과 월기준으로 변경하기 위해 "년"앞의 +기호 클릭하고, "분기"앞의 +기호 클릭하고, "분기"는 Drag out하여 년도와 월기준으로 막대 그래프를 생성한다.
150 56)화면 좌측의 [테이블] > [확진자]를 [마크] > [레이블]로 Drag & Drop하면 라인 그래프에 라벨링이 된다.
151 57)[행]에 있는 "합계(확진자)"를 평균으로 변경해본다.
152 58)[행]에 있는 "합계(확진자)" 아래 화살표 > [측정값 (합계)] > [합계]를 [평균]으로 변경한다.
153 59)동일한 방법으로 [마크]에 있는 "합계(확진자)"도 "평균(확진자)"로 변경한다.
154
155 60)현재 그래프가 시도 확진자 전체이기 때문에 "서울"만 보려고 한다.
156 61)화면 좌측의 [테이블] > [시도]를 [마크]위의 [필터]로 Drag & Drop해 보자.
157 62)[필터 [시도]]창에서 [목록에서 선택(S)] 선택, "서울"만 체크 후, [확인] 클릭한다.
158
159 63)화면 좌측의 [데이터]가 아니라 그 옆의 [분석]을 선택한다.
160 64)[요약] > [평균 라인]을 [시트]에 Drag 하여 [참조선] > [테이블]에 Drop하면 그래프에 [평균] 선이 나타난다.
161 65)한편 동일하게 [평균 라인]을 [참조선] > [패널]에 Drag & Drop하면, 각 년도별로 평균라인이 표시된다.
162
163 66)[새 워크시트] 클릭하여 새로운 시트를 생성한다.
164 67)화면 좌측의 [테이블] > [확진자]를 더블클릭하여 시트에 막대그래프로 표시한다.
165 68)화면 좌측의 [테이블] > [확진자]를 더블클릭하면 선 그래프로 표시된다.

166 69)[열]의 "년(일자)" 아래 화살표를 클릭해보면 컨택스트 메뉴에서 년, 분기, 월, 일이 가로선 기준으로 두개가 있는 것을 볼 수 있다.
167 70)위의 년, 분기, 월, 일은 불연속형이고 아래의 년, 분기, 월, 일은 연속형이다.
168 71)아래의 년을 클릭한다. 그러면 라인 그래프가 연속형으로 변경된다.
169 72)불연속형과 달리 [열]의 "년(일자)"앞의 +를 클릭하면 새로 다른 그래프가 나오는 것이 아니라 분기, 월, 주, 일로 연속형 그래프를 볼 수 있다.
170
171 73)화면 좌측의 [테이블] > [시도]를 [필터]로 Drag & Drop한다.
172 74)[필터 [시도]]창에서, "서울"을 체크하고 [확인]을 클릭하여 서울 데이터만 확인할 수 있다.
173 75)좌측의 [테이블]을 [마크] > [색상]으로 Drag & Drop한다.
174 76)[필터] > "시도:서울" 아래 화살표 > [필터 편집...]을 클릭하여, 수도권 즉 서울, 경기, 인천을 체크하여 수도권만 반영한다.
175
176 77)추세선을 넣어보자.
177 78)화면 좌측의 [데이터] 옆의 [분석] > [모델] > [추세선]을 [시트]에 Drag하여 [추세선] > [다항식]에 Drop한다.
178
179 79)[새 워크시트] 클릭하여 새로운 시트를 생성한다.
180 80)화면 좌측의 [시도]를 더블클릭한다. 시도가 현재 문자형이기 때문에 표 형태로 보여진다.
181 81)[행]에서 시도를 Drag out하여 지운다.
182 82)화면 좌측의 [테이블] > [시도]에 아래 화살표 > [지리적 역할] > [주/시/도] 클릭한다.
183 83)이렇게 하면, [테이블] > [시도]가 지구모양 아이콘으로 변경되고, 경도(생성됨), 위도(생성됨)이 자동 생성된다.
184 84)이제 [테이블] > [시도]를 더블클릭하면 시트에 지도가 표시된다.
185 85)지도 우측 하단에 보면 [1개의 알 수 없는 항목]이 보이고, 이것을 클릭하면 [[시도]에 대한 특수 값] 창이 나오게 되는데, 여기서 [위치 편집...]을 클릭한다.
186 86)그 이유는 바로 "검역"이라는 데이터 때문이라는 것을 알 수 있다.
187 87)[위치 편집]창을 닫고, "검역"은 특수한 데이터이기 때문에 "인천공항" 기준으로 변경하기로 한다. 왜냐하면 검역은 외국에서 들어오는 입국자 정보이기 때문이다.
188 88)지도에 마우스 right-click [주석 추가] > [지점...] 선택한다.
189 89)[주석 편집]이 나타난다.
190 90)툴바에서 [삽입] > [위도(생성됨)] 선택하여 편집기 창에 나오게 하고, 엔터키를 눌러서 그 다음 라인에는 [삽입] > [경도(생성됨)]을 넣는다. 그리고 [확인] 버튼을 클릭한다.
191 91)그러면 지도에 위도와 경도 값이 표시된다.
192 92)이 값을 Drag 하여 인천 영종도 쪽에 맞추면 대략 인천공항의 위도와 경도가 표시된다.
193 93)메뉴바의 [맵(M)] > [위치 편집...]을 클릭하여 [위치 편집] 창을 띄운다.
194 94)[위치 일치] 항목 중에 현재 "검역"은 "인식할 수 없음"이니까 선택하여 텍스트를 지우고 아래 화살표 > "위도 및 경도 입력.." 메뉴를 선택한다.
195 95)[위도 및 경도 입력]창에서, [위도] : 37.447, [경도] : 126.428을 입력하고 [확인] 버튼 클릭한다. 또 한번 [확인]을 클릭한다.
196 96)그러면 해당 지역은 "검역"으로 인식하고, 다른 시도처럼 파란색 원으로 표시된다. 그리고 값이 표시되어 있는 네모상자는 지운다.
197 97)화면 좌측의 [테이블] > [확진자]를 [마크] > [크기]로 Drag & Drop 한다.
198 98)각 원의 사이즈가 너무 작으면, [마크] > [크기]를 선택하여 좀 더 크게 마우스로 오른쪽 방향으로 Drag한다.
199 99)또, [시도]를 [마크]의 [레이블]에 Drag & Drop하여 지도의 각 원에 시도가 표시되게 한다.
200 100)[마크] > [레이블]을 클릭한다.
201 101)[레이블 모양] > [맞춤]이 현재는 "자동"이다.
202 102)아래 화살표를 클릭하여 [가로] 가운데, [세로] 가운데 버튼을 클릭한다.
203 103)그러면 원안에 시도의 이름이 들어가지만 서울과 경기를 제외하고는 원의 크기가 작아서 글씨가 잘 보이지 않는다.
204 104)이럴 경우, 화면 우측의 범례 즉, "합계(확진자)"의 아래 화살표 > [크기 편집...]을 클릭한다.
205 105)[크기 편집 [확진자]]창에서 [크기 범위(S)]를 "범위별"로 맞추고, [마크 크기 범위(W)]의 왼쪽 시작부분을 조금 오른쪽으로 Drag하여 시작값을 조금 더 크게 시작하기로 하고, [확인]버튼을 클릭한다.
206 106)이렇게 하면 오히려 값이 큰 서울과 경기의 글씨가 가려질 수 있는데, 이럴 경우에는 [마크] > [레이블]을 클릭하여 [오프선] > "레이블이 다른 마크와 겹치도록 허용"을 체크하면 된다.
207
208 107)색상을 현재 파란색 계열에서 붉은 색 계열로 변경해 본다.
209 108)화면 좌측의 [테이블] > [확진자]를 [마크] > [색상]으로 Drag & Drop하여 파란색 색상을 그라데이션으로 변경한다.
210 109)[마크] > [색상]을 클릭하고, [색상 추가...]를 클릭하여 [색 선택] 창을 띄운뒤, 붉은 색을 선택하면, 붉은 색 그라데이션을 표시한다.
211 110)또, [색상]에서 [효과] > [테두리]를 "자동" 아니라 "검은색"으로 지정하면 원 테두리가 검은색으로 변경된다.
212
213 111)현재 지도의 각 시도는 확진자의 총합을 표시하고 있어서 최근 날짜 값으로 변경하려고 한다.
214 112)화면 좌측의 [테이블] > [일자]를 [필터]로 Drag & Drop 한다.
215 113)[필터 필터 [일자]]창이 오픈되면, [개별 날짜]를 선택하고 [다음] 버튼 클릭
216 114)가장 최근 날짜를 보여주도록 하기 위해, [필터 [일자]]창에서 제일 아래쪽의 [통합 문서를 연 최근 날짜 값으로 필터링] 체크하면 창의 위쪽의 목록이 다시 정렬되는데, 마우스로 제일 아래로 스크롤하여 가장 최근 날짜가 체크되어 있음을 확인한다. 그리고 [확인] 버튼을 클릭한다.
217 115)각 원의 값이 최근 날짜 기준으로 변경된 것임을 확인할 수 있다.
218 116)화면 아래로 와서 지금 작업한 시트를 "시도별 일간 확진자"라고 변경한다.
219
220 117)이제 지금까지 작업한 시트를 Dashboard에 넣기로 한다. 화면 아래에서 [새 대시보드]를 클릭한다.
221 118)화면 좌측의 [시트]에서 방금 작업했던 [시도별 일간 확진자]를 Drag & Drop하고, 왼쪽의 범례는 모두 제거한다.
222 119)지도위의 "시도별 일간 확진자" 글자를 더블클릭하여 [제목 편집]창을 띄운다.
223 120)[제목 편집] 창에서 [삽입] > [일자] 후 [확인]을 클릭하여, 제목이 <시트이름>(<일자>기준) 이렇게 설정하고 이 글자들을 전체 선택하여 가운데 정렬하고 [확인]버튼 클릭한다.
224 121)지금까지 작업한 파일을 저장한다.
225