

Joshua N. Milligan

Poznaj Tableau 2022

Wizualizacja danych, interaktywna analiza danych
i umiejętność data storytellingu

Przedmowy:

Blair Hutchinson
starszy menedżer produktu,
Tableau, firma Salesforce

Mark Tossell
główny inżynier rozwiązań,
Tableau, firma Salesforce

Roberto Andreoli
starszy kierownik ds. inżynierii
rozwiązań, Tableau, firma
Salesforce (region South EMEA)

Wydanie V



<packt

Helion

Tytuł oryginału: Learning Tableau 2022: Create effective data visualizations, build interactive visual analytics, and improve your data storytelling capabilities, 5th Edition

Tłumaczenie: Anna Mizerska

ISBN: 978-83-8322-372-8

Copyright © Packt Publishing 2022. First published in the English language under the title 'Learning Tableau 2022 - Fifth Edition – (9781801072328)'

Polish edition copyright © 2023 by Helion S.A.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from the Publisher.

Wszelkie prawa zastrzeżone. Nieautoryzowane rozpowszechnianie całości lub fragmentu niniejszej publikacji w jakiegokolwiek postaci jest zabronione. Wykonywanie kopii metodą kserograficzną, fotograficzną, a także kopiowanie książki na nośniku filmowym, magnetycznym lub innym powoduje naruszenie praw autorskich niniejszej publikacji.

Wszystkie znaki występujące w tekście są zastrzeżonymi znakami firmowymi bądź towarowymi ich właścicieli.

Autor oraz wydawca dołożyli wszelkich starań, by zawarte w tej książce informacje były kompletne i rzetelne. Nie biorą jednak żadnej odpowiedzialności ani za ich wykorzystanie, ani za związane z tym ewentualne naruszenie praw patentowych lub autorskich. Autor oraz wydawca nie ponoszą również żadnej odpowiedzialności za ewentualne szkody wynikłe z wykorzystania informacji zawartych w książce.

Drogi Czytelniku!

Jeżeli chcesz ocenić tę książkę, zajrzyj pod adres

<https://helion.pl/user/opinie/pot225>

Możesz tam wpisać swoje uwagi, spostrzeżenia, recenzję.

Pliki z przykładami omawianymi w książce można znaleźć pod adresem:

<https://ftp.helion.pl/przyklady/pot225.zip>

Helion S.A.

ul. Kościuszki 1c, 44-100 Gliwice

tel. 32 230 98 63

e-mail: helion@helion.pl

WWW: <https://helion.pl> (księgarnia internetowa, katalog książek)

Printed in Poland.

- [Kup książkę](#)
- [Poleć książkę](#)
- [Oceń książkę](#)

- [Księgarnia internetowa](#)
- [Lubię to! » Nasza społeczność](#)

Spis treści |

Przedmowy	11
O autorze	15
O recenzentach	16
Wstęp	17
 ROZDZIAŁ 1.	
Rozpoczęcie pracy z Tableau	23
Łączenie się z danymi	24
Podstawy tworzenia wizualizacji	28
Miary i wymiary	28
Nieciągłe i ciągłe pola	30
Wizualizacja danych	33
Wykresy słupkowe	33
Wykresy liniowe	38
Tworzenie wizualizacji geograficznych	42
Stosowanie funkcji Show Me	49
Zebranie wszystkiego w pulpicie nawigacyjnym	52
Interfejs pulpitu nawigacyjnego	53
Budowa pulpitu nawigacyjnego	55
Podsumowanie	58
 ROZDZIAŁ 2.	
Łączenie się z danymi w Tableau	59
Paradygmat Tableau	59
Paradygmat Tableau w akcji	60
Łączenie się z danymi	65
Łączenie się z danymi w pliku	66
Łączenie się z danymi na serwerze	69
Łączenie się z danymi w chmurze	72
Skróty w celu łączenia się z danymi	75
Zarządzanie metadanymi źródła danych	75

Praca z wyciągami z danych zamiast połączeniami na żywo	78
Tworzenie wyciągów	79
Używanie wyciągów	82
Wydażność	84
Przenaszalność i bezpieczeństwo	85
Zastosowanie wyciągów	86
Używanie zestawów i grup	87
Filtrowanie danych	91
Filtrowanie nieciągłych (niebieskich) pól	92
Filtrowanie ciągłych (zielonych) pól	94
Filtrowanie dat	96
Inne opcje filtrowania	97
Kolejność działań w Tableau	98
Podsumowanie	100

ROZDZIAŁ 3.

Wizualizacje — kolejny krok101

Porównywanie wartości	102
Wykresy słupkowe	102
Odmiany wykresów słupkowych	105
Wizualizowanie dat i godzin	114
Części dat, wartości dat i dokładne daty	114
Odmiany wizualizacji dat i godzin	118
Wykresy Gantta	120
Dane jako część całości	123
Wykres skumulowany słupkowy	123
Mapy drzewa	126
Wykres obszarowy	129
Wykresy kołowe	131
Wizualizowanie rozkładu	133
Wykres ze znacznikami w postaci kółek	133
Skrzynki i wąsy	136
Histogramy	138
Wizualizowanie kilku osi w celu porównania różnych miar	142
Wykres punktowy	143
Wykresy z podwójnymi osiami i wykresy kombi	144
Podsumowanie	147

ROZDZIAŁ 4.**Początek przygody z obliczeniami i parametrami 148**

Wprowadzenie do obliczeń	149
Cztery główne typy obliczeń	149
Tworzenie i edytowanie obliczeń	150
Typy danych	153
Dodatkowe funkcje i operatory	154
Zrozumienie danych użytych w przykładach z tego rozdziału	155
Obliczenia na poziomie wiersza	155
Łączenie łańcuchów znaków	155
Operacje na ciągach znaków i logika warunkowa	156
Planowanie zmian w danych	159
Obliczenia agregujące	160
Znaczenie różnicy między obliczeniami na poziomie wiersza a agregacją	163
Parametry	166
Tworzenie parametrów	167
Praktyczne przykłady obliczeń i parametrów	170
Rozwiązywanie problemów z danymi	170
Poszerzanie danych	171
Ulepszanie interfejsu użytkownika, analizy i wizualizacji	173
Spełnianie wymagań zarządu firmy	175
Obliczenia ad hoc	177
Wydajność obliczeń	179
Podsumowanie	180

ROZDZIAŁ 5.**Obliczenia na różnych poziomach szczegółowości 182**

Ogólne informacje o poziomie szczegółowości	182
Obliczenia na poziomie szczegółowości	184
Składnia	185
Rodzaje poziomów szczegółowości	185
Znaczenie różnicy między poziomami szczegółowości	186
Przykłady obliczeń na poziomie szczegółowości typu fixed	188
Oznaczanie członków na podstawie historycznych danych	189
Najświeższe saldo każdego członka	192
Przykład obliczenia na poziomie szczegółowości typu include	195
Średnia liczba pożyczek na członka	195
Przykład obliczenia na poziomie szczegółowości typu exclude	199
Średnia ocena kredytowa na rodzaj pożyczki	199
Podsumowanie	201

ROZDZIAŁ 6.**Obliczenia tablicowe202**

Wprowadzenie do obliczeń tablicowych	203
Tworzenie i edytowanie obliczeń tablicowych	204
Szybkie obliczenia tablicowe	205
Podstawowe koncepcje obliczeń tablicowych	207
Względne a stałe obliczenie tablicowe	208
Zakres i kierunek	210
Adresowanie i partycjonowanie	214
Niestandardowe obliczenia tablicowe	219
Funkcje tablicowe zwracające metadane	219
Przeszukiwanie i poprzednia wartość	220
Funkcje kroczące	222
Funkcje okna	223
Funkcje rankingu	224
Funkcje skryptu	224
Funkcja Total	225
Praktyczne przykłady	226
Wzrost rok do roku	226
Dynamiczne tytuły z wartościami całkowitymi	229
Filtrowanie na pomocą obliczenia tablicowego (późne filtrowanie)	231
Podsumowanie	233

ROZDZIAŁ 7.**Tworzenie dobrze wyglądających i działających wizualizacji 234**

Co należy wziąć pod uwagę, przygotowując wizualizację?	235
Formatowanie w Tableau	236
Formatowanie na poziomie skoroszytu	238
Formatowanie na poziomie arkusza	239
Formatowanie na poziomie pola	243
Dodatkowe możliwości formatowania	252
Ulepszanie wizualizacji	254
Chmurki	257
Podsumowanie	261

ROZDZIAŁ 8.**Opowiadanie historii za pomocą pulpitu nawigacyjnego262**

Wprowadzenie do pulpitu nawigacyjnego	263
Cele pulpitu nawigacyjnego	263
Podejścia do tworzenia pulpitu nawigacyjnego	264

Projektowanie pulpitów nawigacyjnych w Tableau	266
Użycie obiektów	266
Pulpit nawigacyjny wyświetlający informacje o rentowności	270
Budowanie widoków	271
Tworzenie pulpitu nawigacyjnego	274
Dodanie akcji w celu prowadzenia historii	277
Filtrowanie kontekstowe	279
Projektowanie na różne wyświetlacze i urządzenia	282
Praca z akcjami	285
Akcje filtra	286
Akcje podkreślania	290
Akcje adresu URL	291
Akcje przechodzenia do arkusza	293
Akcje parametrów	293
Akcje zestawów	294
Pulpit nawigacyjny z wynikami poszczególnych regionów	298
Historie	303
Podsumowanie	310

ROZDZIAŁ 9.

Wizualna analityka danych — trendy, grupy, rozkłady 311

Trendy	312
Niestandardowe linie trendu	316
Modele trendu	320
Analizowanie modeli trendów	324
Funkcja Explain Data	330
Grupowanie	334
Rozkłady	338
Prognozowanie	342
Podsumowanie	346

Rozdział 10.

Zaawansowane wizualizacje 348

Zaawansowane wizualizacje — kiedy i dlaczego je stosować	349
Wykresy nachylenia i wykresy zmiany pozycji	349
Wykresy kaskadowe	352
Wykresy schodkowe i skokowe	353
Linie przebiegu w czasie	355
Wykresy hantli	356
Wykresy jednostkowe	357
Wykresy mozaikowe	360

Wizualizacje animowane	363
Wzbogacanie analizy przez animację	363
Wzbogacanie historii płynącej z danych przez animację	364
Podsumowanie	366

ROZDZIAŁ 11.

Dynamiczne pulpity nawigacyjne367

Przyciski Show/Hide	367
Zamiana arkuszy	373
Podstawowe reguły zamiany arkuszy	373
Zastosowanie zamiany arkuszy w celu zmiany widoków na pulpicie nawigacyjnym	373
Automatyczne pokazywanie i ukrywanie innych kontroltek	379
Podsumowanie	381

ROZDZIAŁ 12.

Odkrywanie pracy z mapami

i zaawansowanych funkcji geoprzestrzennych383

Przegląd map w Tableau	384
Renderowanie map w Tableau	384
Dostosowanie warstw map	385
Dostosowanie opcji map	388
Użycie danych geoprzestrzennych	389
Szerokość i długość geograficzna w Twoich danych	390
Importowanie definicji do geograficznej bazy danych Tableau	391
Wykorzystanie obiektów przestrzennych	391
Zastosowanie funkcji przestrzennych	394
Funkcje MAKELINE() oraz MAKEPOINT()	395
Funkcja DISTANCE()	397
Funkcja BUFFER()	398
Tworzenie niestandardowych terytoriów	399
Niestandardowe terytoria ad hoc	400
Niestandardowe terytoria określone przez pola	402
Warstwy map	403
Praca z mapami w Tableau — rady i wskazówki	408
Wykreślanie danych na obrazach w tle	410
Podsumowanie	414

ROZDZIAŁ 13.**Integracja zaawansowanych funkcji — rozszerzenia, skrypty i sztuczna inteligencja 415**

Rozszerzenie pulpitu nawigacyjnego w Tableau	416
Czym jest rozszerzenie pulpitu nawigacyjnego?	416
Używanie rozszerzenia pulpitu nawigacyjnego	418
Tworzenie rozszerzenia	424
Rozszerzenia analityczne w Tableau	432
Pisanie skryptów w Pythonie	433
Sztuczna inteligencja z CRM Analytics (Einstein Discovery)	440
Dane użyte w przykładach	441
Tworzenie historii Einstein Discovery	442
Wykorzystanie sztucznej inteligencji w Tableau	449
Podsumowanie	455

ROZDZIAŁ 14.**Modele danych, łączenie danych i blendowanie w Tableau 456**

Wyjaśnienie przykładowych danych użytych w tym rozdziale	457
Odkrywanie modelu danych w Tableau	459
Tworzenie modelu danych	459
Warstwy modelu danych	462
Używanie modelu danych	464
Łączenie danych	469
Rodzaje złączeń	469
Łączenie tabel	472
Złączenia — dodatkowe informacje	476
Zastosowanie blendowania	480
Zastosowanie blendowania w celu pokazania postępu w dążeniu do celu	482
Kiedy używać modelu danych, łączenia i blendowania?	486
Podsumowanie	488

ROZDZIAŁ 15.**Porządkowanie danych na potrzeby Tableau 489**

Strukturyzowanie danych dla Tableau	490
Dobrze ustrukturyzowane dane dla Tableau	491
Schemat gwiazdy	495
Cztery podstawowe przekształcenia danych	498
Przegląd przekształceń	498
Tabele przestawne (z kilkoma prostymi technikami czyszczenia danych)	499
Łączenie danych (operacje union)	504

Złączenia danych (operacje join)	508
Agregacja	513
Przegląd zaawansowanych rozwiązań problemów z danymi	518
Podsumowanie	519

ROZDZIAŁ 16.

Poskramianie danych za pomocą Tableau Prep521

Przygotowanie do odkrywania Tableau Prep Builder	522
Interfejs Tableau Prep Builder	523
Podstawowy paradygmat	526
Łączenie się z danymi	529
Porządkowanie danych	534
Obliczenia i agregacje w Tableau Prep	544
Filtrowanie danych w Tableau Prep	550
Przekształcanie danych pod kątem analizy	553
Używanie parametrów w Tableau Prep Builder	562
Automatyzacja przepływów	565
Podsumowanie	566

ROZDZIAŁ 17.

Dzielenie się swoją historią567

Prezentowanie, drukowanie i eksportowanie	567
Prezentacja	568
Drukowanie	569
Eksportowanie	572
Udostępnianie użytkownikom Tableau Desktop i Tableau Reader	574
Udostępnianie użytkownikom Tableau Desktop	574
Udostępnianie użytkownikom Tableau Reader	575
Udostępnianie użytkownikom Tableau Server, Tableau Cloud i Tableau Public	576
Publikowanie w Tableau Public	578
Publikowanie w Tableau Server i Tableau Cloud	579
Podsumowanie	589

Łączenie się z danymi w Tableau

Rozdział 2

Tableau oferuje możliwość połączenia się prawie z każdym źródłem danych. Wykorzystuje do tego moc i skuteczność istniejących silników baz danych (między innymi Snowflake i SQL Server) albo wyciąga dane lokalnie.

W rozdziale 14., „Modele danych, łączenie danych i blendowanie w Tableau”, zobaczymy, jak w różny sposób można powiązać tabele danych. W tym rozdziale skupimy się nad tym, by zrozumieć, jak Tableau łączy się i pracuje z danymi. Omówimy następujące tematy:

- paradygmat Tableau,
- łączenie się z danymi,
- zarządzanie metadanymi źródła danych,
- praca z wyciągami z danych zamiast połączeniami na żywo,
- używanie zestawów i grup,
- filtrowanie danych.

Zacniemy od zrozumienia paradygmatu pracy Tableau z danymi.

Paradygmat Tableau

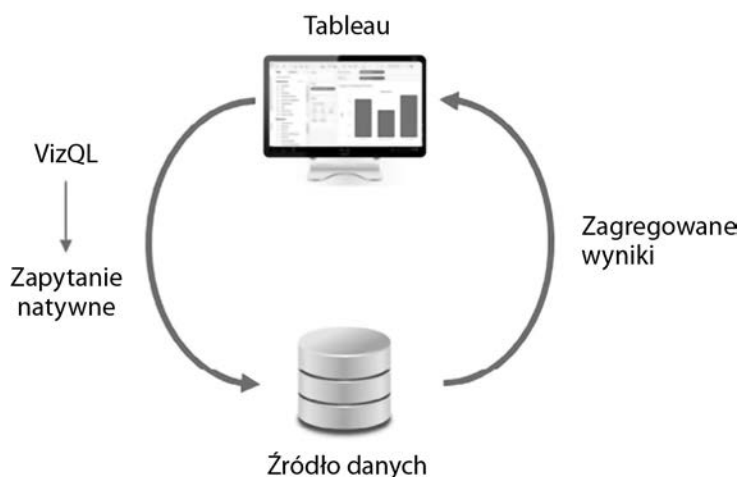
Tak unikatowy i fascynujący sposób pracy z danymi w Tableau jest zasługą języka **VizQL** (ang. *Visual Query Language*).

VizQL został stworzony w ramach projektu naukowego na Uniwersytecie Stanforda, który skupiał się na tym, jak ludzie postrzegają świat za pomocą wzroku i jak to można zastosować w wizualizacji danych. Naturalnie postrzegamy różnice rozmiaru, kształtu, położenia w przestrzeni i koloru. Gdy przeciągasz pola do środowiska wizualnego, VizQL umożliwia Tableau przekładanie Twoich działań na język zapytań, który definiuje, jak dane kształtują te elementy wizualne. Nigdy nie będziesz musiał czytać czy pisać kodu VizQL ani rozwiązywać problemów, które go dotyczą. Gdy przeciągasz pola na różne półki określające rozmiar, kolor, kształt czy położenie, Tableau w tle generuje VizQL. Dzięki temu Ty możesz skupić się na wizualizowaniu danych, a nie na pisaniu kodu!

Jedną z zalet języka VizQL jest spójny sposób opisywania, jak ułożenie różnych pól w widoku określa zapytanie związane z danymi.

Ten wspólny punkt odniesienia może zostać następnie przetłumaczony na różne odmiany języka SQL, MDX i **Tableau Query Language** (TQL, używany dla wyciągniętych danych). Tableau automatycznie wykona tłumaczenie z VizQL na język natywny, który może zostać uruchomiony przez silnik źródła danych.

W najprostszej postaci paradygmat Tableau dotyczący pracy z danymi wygląda tak jak na schemacie z rysunku 2.1.



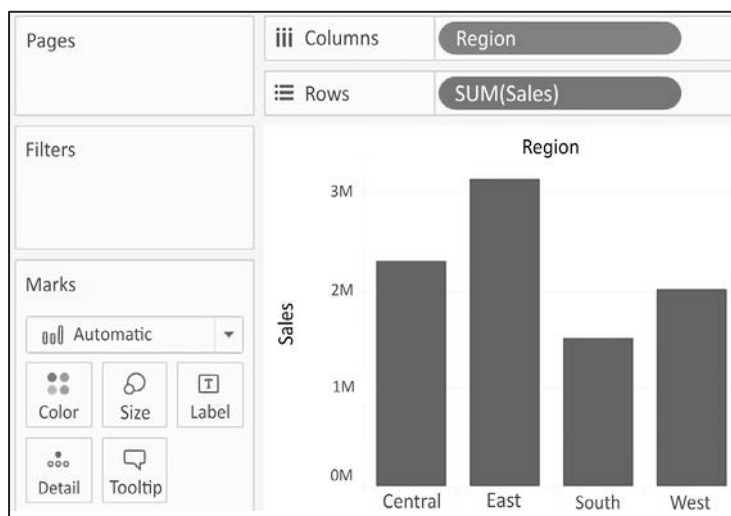
Rysunek 2.1. Podstawowy paradygmat Tableau związany z pracą z danymi

Zobaczmy na praktycznym przykładzie, jak działa ten paradygmat.

Paradygmat Tableau w akcji

Otwórz znajdujący się w folderze *r02* skoroszyt *Chapter 02 Starter.twbx* i przejdź do arkusza *Tableau Paradigm* (paradygmat Tableau). Ten widok, pokazany na rysunku 2.2, został stworzony przez przeciągnięcie wymiaru *Region* na półkę *Columns* i miary *Sales* na półkę *Rows*.

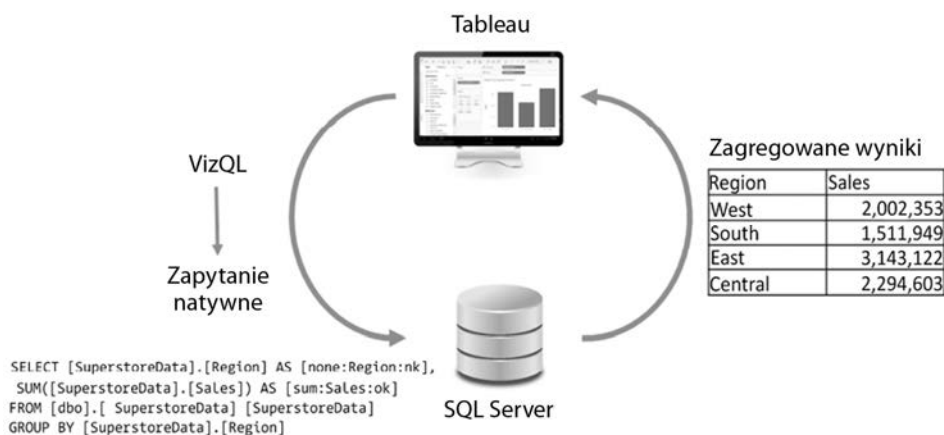
Ten widok definiują dwa pola. Pole *Region* jest tylko wymiarem, co oznacza, że określa poziom szczegółowości danych w widoku i dzieli miarę, w rezultacie czego będziemy mieli po jednym słupku na każdy region. Pole *Sales* jest używane jako miara zagregowana przez zsumowanie sprzedaży w każdym regionie. (Zauważ również, że pole *Region* jest nieciągłe, co przekłada się na nagłówki kolumn, podczas gdy pole *Sales* jest ciągłe i przekłada się na oś).



Rysunek 2.2. Ten wykres słupkowy jest wynikiem zapytania, które zwróciło cztery zagregowane wiersze danych

Na potrzeby tego przykładu (choć ta zasada dotyczy każdego źródła danych) założmy, że jesteś połączony na żywo z bazą danych SQL Server z danymi Superstore zapisanymi w tabeli. Gdy stworzyłeś widok z rysunku 2.2, Tableau wygenerował skrypt VizQL, który jest następnie tłumaczony na skrypt SQL i wysyłany do SQL Server. Silnik bazy danych SQL Server ocenia zapytanie i zwraca do Tableau zagregowane wyniki, które są renderowane.

Cały proces wygląda jak ten ze schematu na rysunku 2.3.



Rysunek 2.3. Platforma Tableau wygenerowała wykres słupkowy z rysunku 2.2 przy użyciu takiego paradygmatu

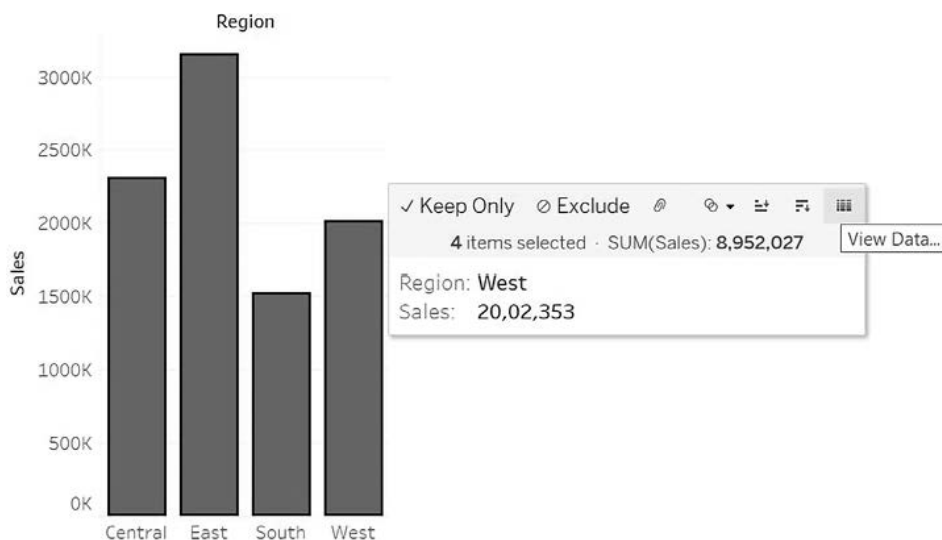
W bazie danych SQL Server mogą być setki, tysiące, a nawet miliony wierszy danych dotyczących sprzedaży. Jednak gdy SQL Server przetwarza zapytanie, zwraca zagregowane wyniki. W tym przypadku SQL Server zwraca do Tableau tylko cztery zagregowane wiersze danych — po jednym dla każdego regionu.

Wskazówka

Czasami administrator bazy danych może chcieć dowiedzieć się, jakie skrypty działają na określonej bazie danych, w celu rozwiązania problemu z wydajnością albo znalezienia wydajniejszego indeksowania lub struktury danych. Wiele baz danych zapewnia narzędzia do profilowania lub rejestry wykonywanych zapytań. Co więcej, wygenerowany przez Tableau skrypt SQL lub MDX możesz znaleźć w rejestrach zapisanych w folderze *Tableau Repository\Logs*.

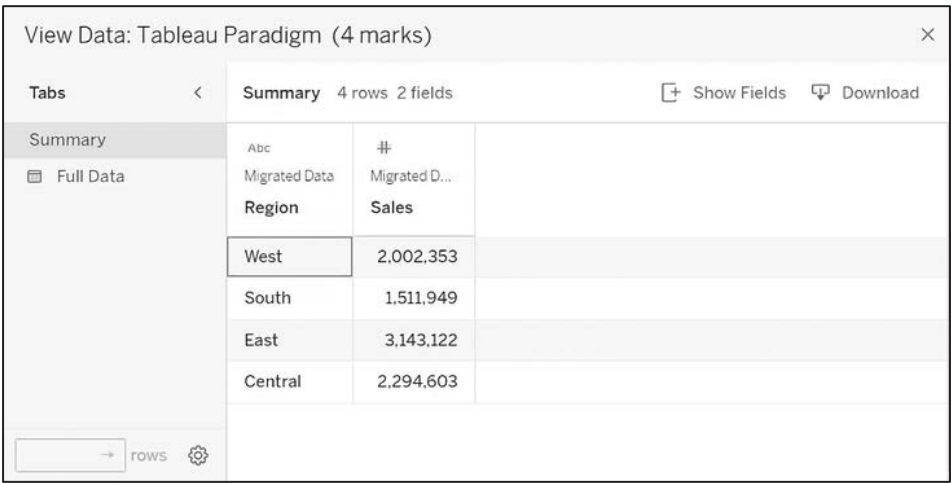
Możesz również użyć wbudowanej nagrywarki, by zlokalizować wykonane zapytania. Z górnego menu wybierz *Help/Settings and Performance/Start Performance Recording* (pomoc/ustawienia i wydajność/rozpocznij nagrywanie), następnie wykonaj kilka działań z widokiem i w końcu zakończ nagrywanie. Tableau otworzy pulpit nawigacyjny, w którym możesz zobaczyć wszystkie zadania, wydajność i zapytania wykonane podczas nagrania.

Aby zobaczyć zagregowane dane, których Tableau użyło do narysowania widoku, naciśnij *Ctrl+A*, by zaznaczyć wszystkie słupki, a następnie kliknij prawym przyciskiem myszy jeden z nich i wybierz *View Data...* (podgląd danych), tak jak pokazano na rysunku 2.4.



Rysunek 2.4. Użyj opcji View Data..., by zobaczyć podsumowanie danych, na podstawie których został utworzony znacznik

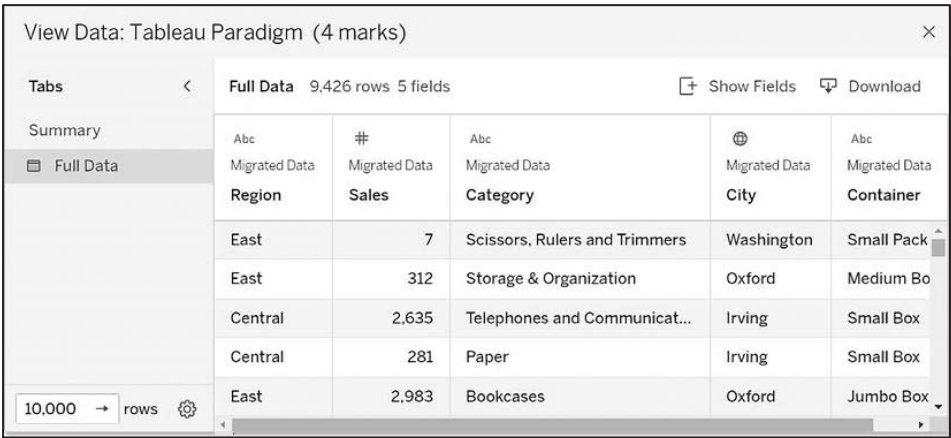
Otworzy się okno *View Data* (podgląd danych), pokazane na rysunku 2.5.



Rysunek 2.5. Zakładka Summary (podsumowanie) wyświetla zagregowane dane użyte przez Tableau w celu narysowania każdego znacznika w widoku

Okno *View Data* umożliwia obserwację danych w widoku. Zakładka *Summary* wyświetla zagregowane dane użyte do przygotowania widoku. Tutaj wartości *Sales* są sumą sprzedaży w każdym regionie. Po naciśnięciu zakładki *Full Data* (pełne dane) Tableau wyśle zapytanie do źródła danych, by uzyskać wszystkie rekordy składające się na zagregowane dane.

W tym przypadku mamy 9426 rekordów, co widać na rysunku 2.6, na pasku stanu.

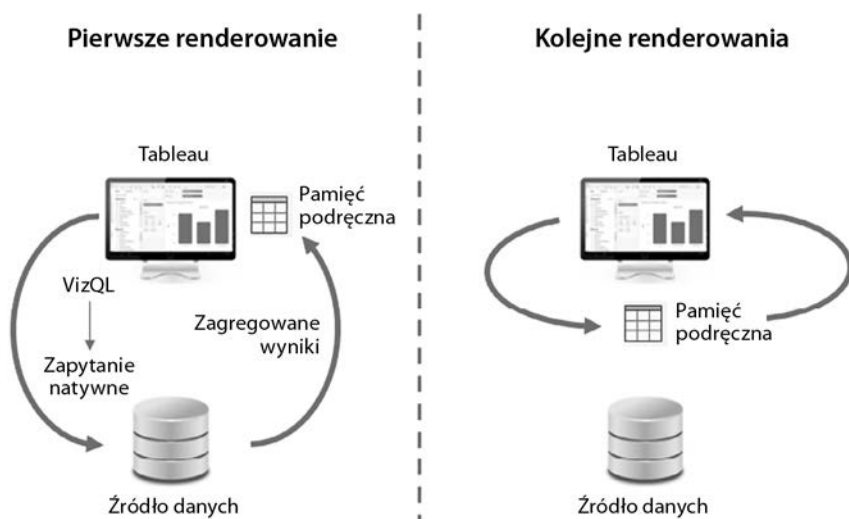


Rysunek 2.6. Zakładka Full Data odkrywa dane na poziomie wiersza

Platforma Tableau nie potrzebowała 9426 rekordów, by przygotować widok, i nie wnioskowała o nie do źródła danych, dopóki nie kliknięto zakładki *Full Data*.

Silniki baz danych są zoptymalizowane pod kątem agregacji danych. Zwykle te bazy danych znajdują się na serwerach o dużej mocy. Tableau wykorzystuje w pełni optymalizację i moc źródła danych. W ten sposób może zwizualizować ogromne zestawy danych przy stosunkowo małym, lokalnym przetwarzaniu danych.

Co więcej, Tableau wyśle zapytanie do źródła danych tylko wtedy, gdy dokonujesz zmian wymagających nowego zapytania czy odświeżenia widoku. W przeciwnym razie dla lepszej wydajności Tableau wykorzysta zagregowane wyniki zapisane w lokalnej pamięci podręcznej, tak jak pokazano na rysunku 2.7.



Rysunek 2.7. Pierwsze renderowanie przy danym zestawie zapytań wysyłanych bezpośrednio do bazy danych. Kolejne renderowania będą wysyłać zapytania do pamięci podręcznej, nawet jeśli te same pola zostały inaczej ułożone w widoku

W poprzednim przykładzie zapytanie z polem *Region* jako wymiarem i sumą pola *Sales* jako miarą zostanie wysłane do źródła danych tylko raz. Gdy cztery zagregowane pola zostaną zwrócone, zostaną zapisane w pamięci podręcznej. Jeśli później zechcesz przenieść pole *Region* na inną półkę, na przykład *Color*, lub pole *Sales* na inną półkę, na przykład *Size*, Tableau pobierze zagregowane wiersze z pamięci podręcznej i po prostu jeszcze raz przygotuje widok.

Jeśli wprowadziłbyś nowe pole do widoku, dla którego wyniki nie są zapisane w pamięci podręcznej, Tableau wysłałoby nowe zapytanie do źródła danych, pobrałoby zagregowane wyniki i dodało je do pamięci podręcznej.

Wskazówka

Możesz zmusić Tableau do pominięcia pamięci podręcznej i odświeżania danych bezpośrednio ze źródła danych — wystarczy, że naciśniesz klawisz *F5* lub wybierzesz źródło danych w menu *Data* i wybierzesz *Refresh* (odśwież). Zrób tak za każdym razem, gdy chcesz, by widok odzwierciedlał najświeższe zmiany w źródle danych, z którym jesteś połączony na żywo.

Gdy już rozumiesz ogólną zasadę działania Tableau, przejdźmy przez konkretne przykłady łączenia się z danymi.

Łączenie się z danymi

Praktycznie nie ma danych, których Tableau nie mogłoby zwizualizować! Prawie każda nowa wersja Tableau jest wzbogacana o nowe natywne konektory. Tableau cały czas dodaje nowe natywne łączniki dla danych chmurowych. Konektor z danymi sieciowymi pozwala na napisanie łącznika dla dowolnych danych online, które chciałbyś pobrać. **Tableau Hyper API** pozwala na odczytywanie i zapisywanie wyciągów danych w sposób zaprogramowany, dzięki czemu możesz mieć dostęp do danych z dowolnego źródła i zapisać je do natywnego formatu Tableau. Ponadto, jeśli dla jakiejś bazy danych nie ma wbudowanego łącznika, Tableau daje Ci możliwość użycia uniwersalnego połączenia **ODBC**.

W tym samym skoroszycie możesz mieć wiele źródeł danych. Każde źródło danych będzie wymienione w zakładce *Data* w bocznym pasku po lewej stronie.

Uwaga

Choć terminy *połączenie z danymi* i *źródło danych* są często używane naprzemiennie, warto je rozróżniać. **Połączenie** z technicznego punktu widzenia odnosi się do połączenia z danymi w jednym miejscu, na przykład z tabelami w jednej bazie danych lub z plikami tego samego typu w tej samej strukturze folderów. **Źródło danych** może zawierać więcej połączeń, które mogą być ze sobą złączone, na przykład tabela w SQL Server złączona z tabelami w Snowflake w tabelę w pliku Excel. Możesz na to patrzeć tak: skoroszyt Tableau może zawierać jedno lub więcej źródeł danych, a każde źródło danych może mieć jedno lub więcej połączeń. Tego rozróżnienia będziemy się trzymać w całej książce.

Ten podrozdział skupia się na kilku praktycznych przykładach łączenia się ze źródłami danych. Nie jest możliwe omówienie wszystkich dostępnych rodzajów połączeń, ale omówimy kilka najpopularniejszych. Możesz mieć dostęp do kilku źródeł danych wykorzystanych w następnych przykładach lub go nie mieć. Nie martw się, jeśli nie jesteś w stanie wykonać danego przykładu. Wtedy po prostu obserwuj, co się zmienia.

Łączenie się z danymi w pliku

Dane oparte na pliku obejmują wszystkie źródła danych, w których dane są zapisane w pliku. Źródła danych oparte na pliku to:

- **Wyciągi.** Plik *.hyper* lub *.tde* zawierający dane wyciągnięte z pierwotnego źródła.
- **Microsoft Access.** Plik bazy danych *.mdb* lub *.accdb*, stworzony w programie Access.
- **Microsoft Excel.** Plik skoroszytu *.xls*, *.xlsx* lub *.xlsm*, utworzony w Excelu. Kilka arkuszy Excela lub tabel może być ze sobą złączone lub związane w ramach jednego połączenia.
- **Plik tekstowy.** Plik tekstowy, w którym dane są oddzielone określonym znakiem, najczęściej są to pliki *.txt*, *.csv* lub *.tab*. Kilka plików tekstowych w jednym folderze może być złączone lub związane w ramach jednego połączenia.
- **Lokalny plik Cube.** Plik *.cub* zawierający wielowymiarowe dane. Te pliki są zwykle eksportowane z baz danych OLAP (ang. *Online Analytical Processing*, przetwarzanie informacji z wielowymiarowej bazy danych).
- **Adobe PDF.** Plik *.pdf*, który może zawierać tabele danych, sczytywane w odpowiedni sposób przez Tableau.
- **Plik przestrzenny.** Obsługiwany jest szeroki wachlarz formatów przestrzennych, między innymi *.kml*, *.shp*, *.tab*, *.mif*, przestrzenny JSON i pliki bazy danych ESRI. Te formaty zawierają obiekty przestrzenne, które Tableau może renderować.
- **Pliki statystyczne.** Plik *.sav*, *.sas7bdat*, *.rda* lub *.rdata* wygenerowany przez narzędzia do statystyki, takie jak SAS czy R.
- **Plik JSON.** Plik z rozszerzeniem *.json*, zawierający dane w formacie JSON.

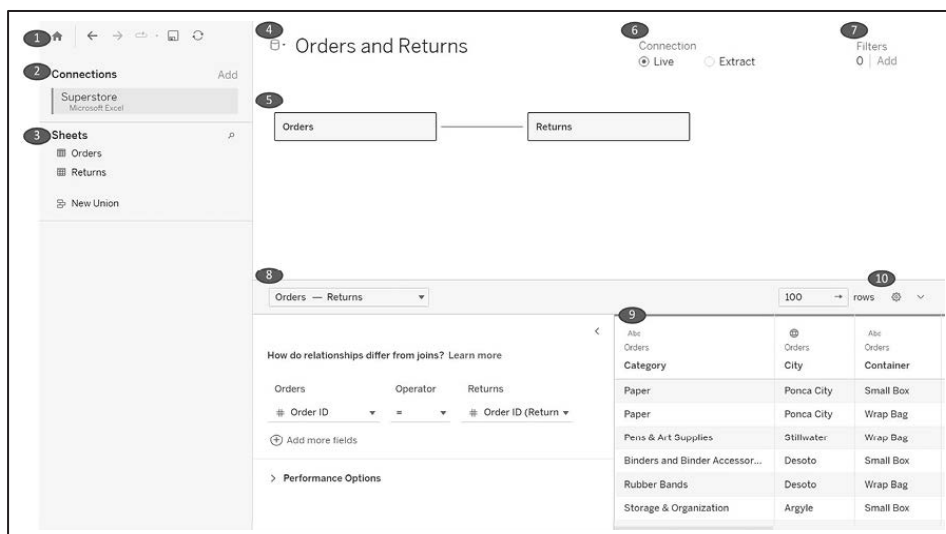
Poza plikami wymienionymi powyżej z Tableau możesz połączyć pliki w celu zaimportowania połączeń zapisanych w innym skoroszycie Tableau (*.twb* lub *.twbx*). Połączenie zostanie zaimportowane, a zmiany wpłyną tylko na bieżący skoroszyt.

Wykonaj następujące kroki, by ustawić połączenie z plikiem Excel:

1. Przejdź do arkusza *Connect to Excel* (połączenie z Excelem) w skoroszycie *Chapter 02 Starter.twbx*.
2. Z menu wybierz *Data/New Data Source* (dane/nowe źródło danych) i wybierz Microsoft Excel z listy dostępnych połączeń.
3. W oknie, które się pojawi, wybierz plik *Superstore.xlsx* z folderu *r02*. Tableau otworzy ekran *Data Source* (źródło danych). Po lewej stronie powinieneś zobaczyć dwa arkusze dokumentu Excel.

4. Kliknij dwukrotnie arkusz *Orders* (zamówienia), a następnie arkusz *Returns*. Tableau otworzy okno *Edit Relationship* (edycja relacji). Relacje omówimy szczegółowo w rozdziale 14., „Modele danych, łączenie danych i blendowanie w Tableau”. Na razie zamknij okno bez zmieniania domyślnych ustawień.

Twój ekran *Data Source* powinien wyglądać podobnie do tego z rysunku 2.8.



Rysunek 2.8. Ekran *Data Source* z dwoma obiektami (*Orders* i *Returns*)

Poświęć chwilę, by zapoznać się z interfejsem ekranu *Data Source*, który ma poniżej wymienione funkcje, ponumerowane na rysunku 2.8.

1. **Pasek narzędzi.** Pasek narzędzi ma kilka znanych Ci już przycisków, w tym *cofnij*, *wykonaj ponownie* i *zapisz*. Znajduje się tam również opcja odświeżania bieżącego źródła danych.
2. **Połączenia.** Są tu wymienione wszystkie połączenia w bieżącym źródle danych. Kliknij *Add* (dodaj), by dodać nowe połączenie do bieżącego źródła danych.
3. **Arkusze (lub tabele).** To lista wszystkich tabel danych dostępnych dla danego połączenia. Zaliczają się do tego arkusze, tabele i nazwane zakresy w Excelu; tabele, widoki i procedury składowane dla relacyjnych baz danych oraz inne opcje zależne od połączenia, na przykład *New Union* (nowe złączenie) lub *Custom SQL* (niestandardowy SQL).
4. **Nazwa źródła danych.** To jest nazwa aktualnie wybranego źródła danych. Możesz wybrać inne źródło danych za pomocą strzałki listy rozwijanej znajdującej się obok ikony bazy danych. Możesz kliknąć nazwę źródła danych, by ją zmienić.

5. **Płótno obiektu/modelu danych.** Przeciągnij arkusze i tabele z lewej strony do tego obszaru, by stanowiły część połączenia. Możesz dodać inne tabele przez ich przeciągnięcie lub podwójne kliknięcie. Każda tabela i każdy arkusz będą dodawane jako obiekty do obiektowego modelu. Możesz również dodać związane ze sobą tabele lub kliknąć dwukrotnie obiekt, by edytować tabele i zestawienia danych, na których ten obiekt się opiera. Więcej szczegółów znajdziesz w rozdziale 14., „Modele danych, łączenie danych i blendowanie w Tableau”. Na tym etapie zwróć tylko uwagę, że obiekty *Orders* i *Returns* są ze sobą powiązane za pomocą numeru zamówienia (*Order ID*).
6. **Opcje *Live* lub *Extract*.** W przypadku wielu źródeł danych możesz wybrać, czy chciałbyś być połączony z nimi na żywo (*Live*), czy uzyskać wyciąg z danych (*Extract*).

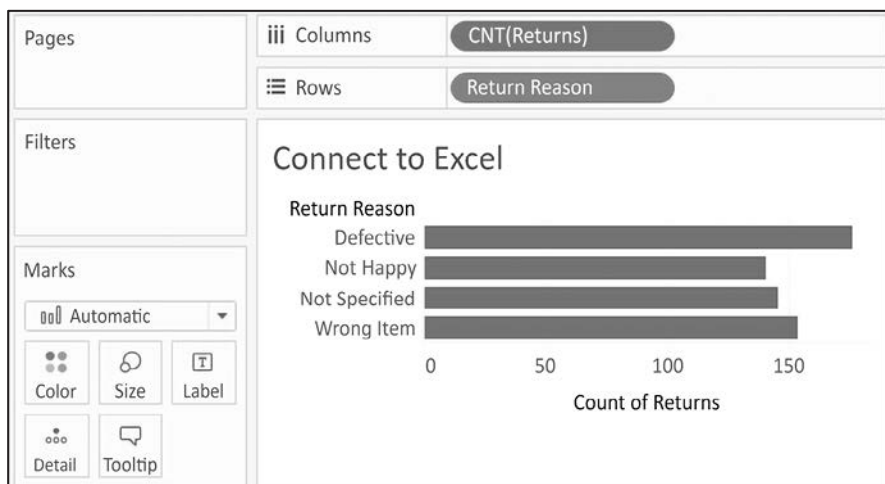
Przyjrzymy się temu bliżej później, w podrozdziale „Praca z wyciągami z danych zamiast połączeniami na żywo”.
7. **Filtry źródła danych.** Do źródła danych możesz dodać filtry, które będą zastosowane na poziomie źródła, a co za tym idzie, do wszystkich widoków danych w skoroszycie, które używają tego źródła danych.
8. **Lista rozwijana obiektów.** Tutaj możesz wybrać obiekt z listy wszystkich obiektów znajdujących się w Twoim połączeniu, w tym również pojedyncze tabele i relacje. W zależności od Twoich wyborów możesz zobaczyć, a nawet zmodyfikować pewne właściwości obiektu. Tutaj na przykład została wybrana relacja między *Orders* i *Returns*. To odkrywa kilka ukrytych opcji, które pozwalają Ci zdefiniować relację.
9. **Panel podglądu/widok metadanych.** W zależności od tego, jaką opcję wybrałeś, w tym obszarze zostaną wyświetlone albo podgląd danych, albo lista pól z dodatkowymi metadanymi. Zauważ, że ten widok daje Ci szeroki wachlarz możliwości, między innymi pozwala na zmianę typu danych, ukrywanie lub zmianę nazwy pól oraz stosowanie różnych funkcji przekształcania danych. Powiemy o kilku z tych możliwości później, w tym rozdziale i w następnych.
10. **Opcje panelu podglądu.** Te opcje pozwalają na określenie, czy chcesz mieć podgląd na dane, czy listę metadanych i w jaki sposób chciałbyś podglądać dane (na przykład możesz ustawić, czy chcesz widzieć aliasy i ukryte pola lub ile wierszy ma być wyświetlanych).

Wskazówka

Gdy już utworzyłeś i skonfigurowałeś swoje źródło danych, możesz kliknąć dowolny arkusz, by zacząć go używać.

Na zakończenie tego ćwiczenia wykonaj następujące kroki:

1. Naciśnij nazwę źródła danych w celu edycji tekstu i wpisz *Orders and Returns*.
2. Przejdź do arkusza *Connect to Excel* i przy użyciu źródła danych *Orders and Returns* stwórz szereg czasowy pokazujący liczbę zwrotów [CNT(Returns)] według powodów zwrotów (Return Reason). Twój widok powinien wyglądać jak ten z rysunku 2.9.



Rysunek 2.9. Liczba zwrotów według powodów zwrotów

Jeśli później będziesz musiał zmienić coś w połączeniu, z menu wybierz *Data*, znajdź swoje połączenie, a następnie wybierz *Edit Data Source* (edytuj źródło danych). Możesz również kliknąć prawym przyciskiem myszy dowolne źródło danych w zakładce *Data* w lewym pasku bocznym i wybrać *Edit Data Source*. Albo możesz także nacisnąć zakładkę *Data Source* (źródło danych), znajdującą się w lewym dolnym rogu Tableau Desktop.

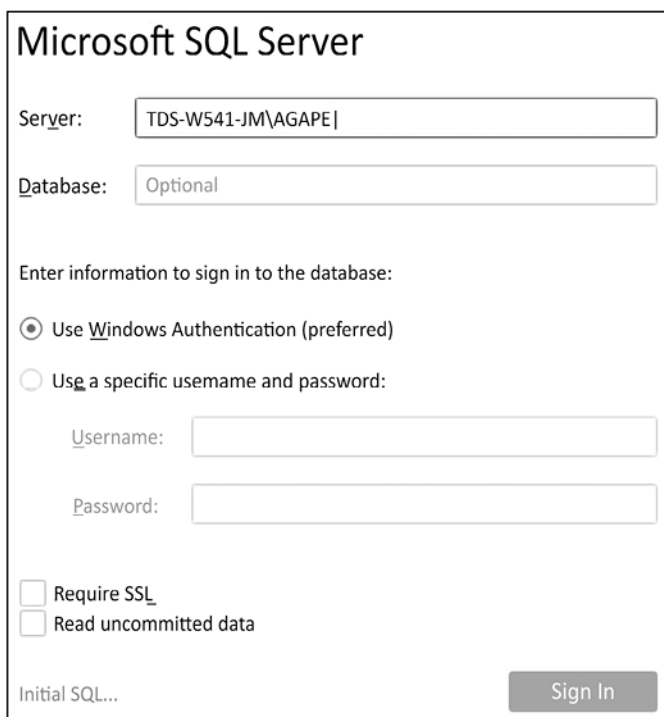
Łączenie się z danymi na serwerze

Serwery baz danych, takie jak SQL Server, Snowflake, Vertica czy Oracle, trzymają dane na jednej maszynie lub kilku i używają bardzo mocnych silników baz danych do zapisywania, agregowania, sortowania i dostarczania danych na podstawie zapytań z aplikacji klienckich. Tableau może w pełni wykorzystywać możliwości tych serwerów, by uzyskać dane na potrzeby wizualizacji i analizy. Dane mogą również zostać wyeksportowane z tych źródeł i zapisane lokalnie.

W ramach przykładu łączenia się z serwerem jako źródłem danych pokażemy, jak połączyć się z SQL Server. Jeśli masz dostęp do bazy danych na serwerze, możesz stworzyć

nowe źródło danych, by odkryć jego możliwości. Jednak tego przykładu nie znajdziesz wśród skoroszytów w materiałach do pobrania.

Po wybraniu połączenia Microsoft SQL Server pojawi się okno pokazane na rysunku 2.10. W tym oknie możesz skonfigurować połączenie.



Rysunek 2.10. Edycja połączenia dla Microsoft SQL Server

Połączenie z SQL Server wymaga podania nazwy serwera (*Server*), jak również informacji związanych z uwierzytelnianiem.

Administrator baz danych może skonfigurować SQL Server, by do baz danych można było się dostać za pomocą danych uwierzytelniania systemu Windows (*Use Windows Authentication*) lub z użyciem nadanej nazwy użytkownika i hasła. W przypadku SQL Server możesz również pozwolić na odczyt niezatwierdzonych danych.

To może potencjalnie korzystnie wpłynąć na jakość działania, ale z drugiej strony może również prowadzić do nieprzewidzianych wyników, jeśli dane są wstawiane, aktualizowane lub usuwane w tym samym czasie, w którym Tableau wysyła zapytanie. Co więcej, za pomocą linku *Initial SQL...*, znajdującego się na dole po lewej stronie, możesz zdecydować, by SQL był wykonywany w momencie połączenia.

Uwaga

W celu zachowania wysokich standardów bezpieczeństwa Tableau nie przechowuje hasła jako części połączenia ze źródłem danych. To znaczy, że jeśli udostępnisz skoroszyt komuś innemu za pomocą połączenia na żywo, to ta osoba będzie musiała mieć nazwę użytkownika i hasło, by uzyskać dostęp do danych. To również wiąże się z tym, że gdy po raz pierwszy otworzysz skoroszyt, będziesz musiał ponownie wpisać hasło dla wszystkich połączeń, które tego wymagają.

Gdy naciśniesz pomarańczowy przycisk *Sign In* (zaloguj się), zobaczysz ekran bardzo podobny do ekranu połączenia, który widziałeś dla Excela. Główna różnica polega na tym, że tutaj po lewej stronie znajdują się opcje wyboru bazy danych, co pokazano na rysunku 2.11.



Rysunek 2.11. Po połączeniu się z bazą danych Tableau wyświetli tabele, widoki i procedury składowane, które można dodać do modelu obiektowego

Po wybraniu bazy danych zobaczysz następujące opcje:

- **Table** (tabela). Tu są pokazane wszystkie tabele i widoki zawarte w wybranej bazie danych.
- **New Custom SQL** (nowy niestandardowy skrypt SQL). Możesz pisać własne, niestandardowe skrypty SQL i dodawać je jako tabele. Możesz je ze sobą łączyć, tak jak w przypadku każdej innej tabeli lub widoku.
- **New Union** (nowe złączenie). Możesz łączyć ze sobą tabele. Tableau dopasuje pole na podstawie nazwy i typu danych, a Ty możesz dodatkowo scalić pola, jeśli zajdzie taka potrzeba.
- **Stored Procedures** (procedury składowane). Możesz użyć procedury składowanej, która zwraca tabelę z danymi. Będziesz miał możliwość ustawienia wartości parametrów procedury składowanej albo użycia lub utworzenia parametru Tableau w celu przekazywania wartości.

Po skonfigurowaniu połączenia naciśnij zakładkę dowolnego arkusza, by rozpocząć wizualizowanie danych.

Łączenie się z danymi w chmurze

Czasami musimy połączyć się z danymi przechowywanymi w chmurze. Są to między innymi **Amazon RDS**, **Google BigQuery**, **Microsoft SQL Azure**, **Snowflake**, **Salesforce**, **Google Drive** i wiele innych. Omówienie każdego z tych połączeń szczegółowo jest poza zakresem tej książki, ale w ramach przykładu łączenia się z chmurowym źródłem danych omówimy połączenie z dokumentem Google Sheets zapisanym na dysku Google.

Aplikacja Google Sheets umożliwia tworzenie i zapisywanie online skoroszytów z danymi. Arkusze mogą być udostępniane i może na nich pracować kilku różnych użytkowników. Tutaj zobaczymy przykład łączenia się z arkuszem udostępnionym przez link.

Aby wykonać ten przykład, będziesz potrzebował darmowego konta Google. Używając swojej nazwy użytkownika i hasła, wykonaj poniższe kroki:

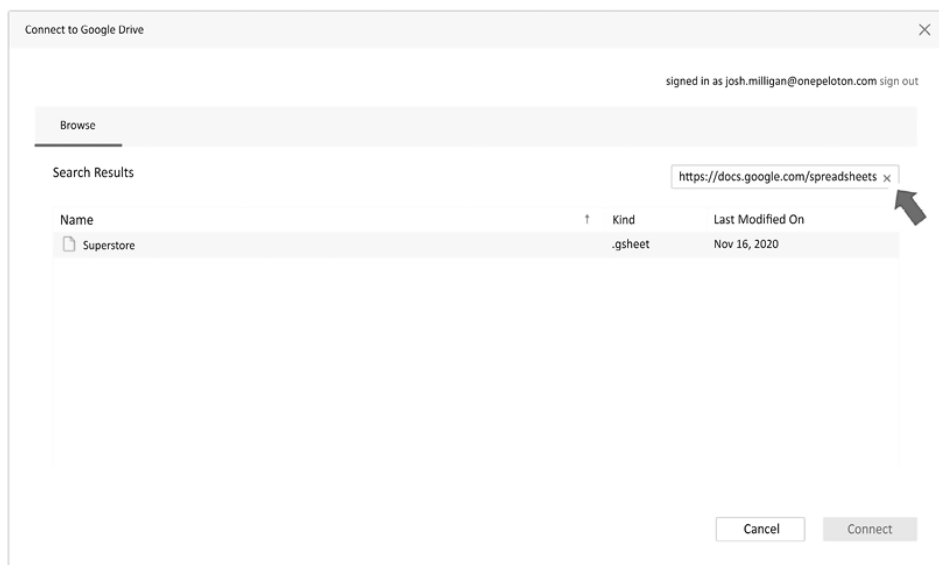
1. Naciśnij przycisk *Add new data source* (dodaj nowe źródło danych), znajdujący się w pasku narzędzi i zaznaczony na rysunku 2.12.



Rysunek 2.12. Przycisk dodawania nowych danych

2. Z listy dostępnych źródeł danych wybierz *Google Drive* (dysk Google). Możesz użyć pola wyszukiwania, by szybko zawęzić listę do *Google Sheets*.

3. Na następnym ekranie zaloguj się do konta Google i przyznaj programowi Tableau Desktop konieczne uprawnienia. Wyświetli się lista wszystkich Twoich dokumentów Google Sheets, wraz z podglądem i możliwościami wyszukiwania, tak jak pokazano na rysunku 2.13.



Rysunek 2.13. Łącząc się z dyskiem Google, możesz przeglądać jego zawartość. W tym przykładzie udostępniony link został wklejony w pole wyszukiwania

4. W polu wyszukiwania wpisz następujący adres URL: https://docs.google.com/spreadsheets/d/1YUZ1UnryuCX7cwbFNptGuzx0G4yth7i-m9Jrkce9_PE/edit?usp=sharing [dla Twojej wygody ten link został zawarty w skrószycie *Chapter 02 Starter* w zakładce *Connect to Google Drive* (łączenie się z dyskiem Google), skąd można go skopiować], a następnie naciśnij przycisk *Search* (szukaj).
5. Z listy wybierz arkusz *Superstore* i naciśnij przycisk *Connect* (połącz). Powinieneś widzieć teraz ekran *Data Source*.
6. Do pola połączenia przeciągnij arkusz *Orders*.
7. Kliknij nazwę *Data Source* i zmień ją na *Superstore* (Google Sheets), tak jak pokazano na rysunku 2.14.
8. Na potrzeby tego przykładu zmień typ połączenia z *Live* na *Extract*. Gdy łączysz się z własnymi danymi Google Sheets, możesz wybrać jedną z tych opcji (rysunek 2.15).
9. Naciśnij zakładkę arkusza *Connect to Google Sheets*. Zostaniesz zapytany, gdzie chcesz zapisać wyciąg. Nie zmieniaj domyślnej nazwy i zapisz plik

Uwaga

Możliwe, że będziesz musiał zmienić swoje ustawienia regionalne w Tableau, by stany poprawnie wyświetlały się na mapie. Z menu wybierz *File/Workbook Locale/More* (plik/ustawienia regionalne skoroszytu/więcej) i wybierz *English (United States)*.

Teraz, gdy już widziałeś kilka przykładów łączenia się z danymi, poznamy kilka skrótów i zobaczymy, jak zarządzać naszymi źródłami danych.

Skróty w celu łączenia się z danymi

Niektóre połączenia możesz wykonywać bardzo szybko. Te opcje umożliwią Ci szybsze rozpoczęcie analizy danych:

- Wklej dane ze schowka. Jeśli skopiowałeś dane do swojego schowka systemowego z dowolnego źródła (na przykład skoroszytu, tabeli na stronie internetowej czy pliku tekstowego), możesz je następnie wkleić bezpośrednio do Tableau. Możesz to zrobić za pomocą skrótu klawiszowego *Ctrl+V* lub z poziomu menu — wybierz *Data/Paste Data* (dane/wklej dane). Dane zostaną zapisane jako plik, a Ty zostaniesz przekierowany do jego lokalizacji, gdy zapiszesz skoroszyt.
- Z menu wybierz *File/Open* (plik/otwórz). To pozwoli Ci otworzyć dowolny plik z danymi obsługiwany przez Tableau, między innymi pliki tekstowe, pliki Excela, pliki Access (opcja niedostępna w systemach macOS), pliki przestrzenne, pliki statystyczne, pliki JSON, a nawet pliki *.cub* w trybie offline.
- Przeciągnij plik z Eksploratora Plików Windows lub Findera do obszaru roboczego Tableau. Każde ważne źródło danych oparte na pliku można przeciągnąć do obszaru roboczego Tableau, a nawet do skrótu programu Tableau na Twoim pulpicie lub pasku zadań.
- Powiel istniejące źródło danych. Możesz skopiować istniejące źródło danych przez kliknięcie go prawym przyciskiem myszy i wybranie *Duplicate* (powiel).

Te skróty zapewniają szybki sposób analizy danych. Teraz przejdźmy do zarządzania źródłami danych.

Zarządzanie metadanymi źródła danych

Źródła danych w Tableau przechowują informacje o połączeniu (połączeniach). Dodatkowo, poza danymi dotyczącymi samego połączenia (na przykład nazwą serwera bazy danych, bazą danych i/lub nazwami plików) źródło danych zawiera również

informacje o wszystkich dostępnych polach (takie jak nazwa pola, typ danych, domyślny format, komentarze, aliasy). Często te *dane o danych* są nazywane **metadanymi**.

Kliknij prawym przyciskiem myszy pole w panelu danych, a pokażą się opcje dla metadanych. Kilka z tych opcji zostanie przedstawionych w kolejnych ćwiczeniach, inne wyjaśnimy w dalszej części tej książki. Oto niektóre z dostępnych opcji po kliknięciu prawym przyciskiem myszy:

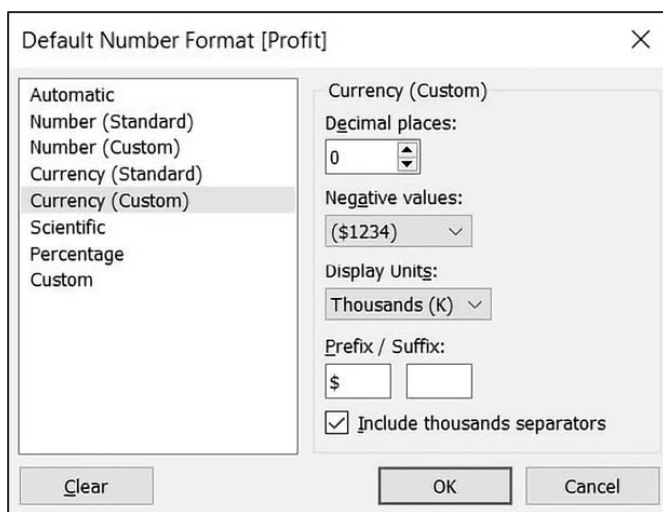
- zmiana nazwy pola,
- ukrywanie pola,
- zmiana aliasów na wartości wymiaru niebędącego datą,
- tworzenie pól obliczeniowych, grup, zestawów, kontenerów lub parametrów,
- rozdzielanie pola,
- zmiana domyślnego użycia daty lub pola liczbowego na nieciągłe albo na ciągłe,
- zmiana definicji pola na wymiar lub na miarę,
- zmiana typu danych pola,
- przypisywanie roli geograficznej do pola,
- zmiana domyślnego sposobu wyświetlania pól w wizualizacji, między innymi domyślnych kolorów i kształtów, formatu liczb lub daty, kolejności sortowania (dla wymiarów) lub typu agregacji (dla miar),
- dodanie domyślnego komentarza do pola [który będzie wyświetlany w chmurce po najechaniu kursorem myszki na pole w panelu danych lub jako część opisu po wybraniu *Describe...* (opisz) z menu],
- dodanie pola do hierarchii albo usunięcie go.

Uwaga

Opcje metadanych, które odnoszą się do wizualnego przedstawienia pola, takie jak domyślna kolejność lub domyślny format liczbowy, definiują ogólne domyślne właściwości pola. Jednak możesz zmienić te domyślne ustawienia dla każdego pola osobno w widoku przez kliknięcie prawym przyciskiem myszy aktywnego pola na półce i wybranie interesującej Cię opcji.

Aby zobaczyć, jak to działa, użyj widoku z kartogramem dla zysku według stanów, który stworzyłeś wcześniej w arkuszu *Connect to Google Sheets*. Jeśli nie utworzyłeś tego widoku, możesz użyć źródła danych *Orders and Returns*, ale uzyskany przez Ciebie widok będzie wyglądał nieco inaczej. Gdy masz otwarty przed sobą kartogram, wykonaj następujące kroki:

1. Prawym przyciskiem myszy kliknij pole *Profit* w panelu danych i wybierz *Default Properties/Number Format...* (domyślne właściwości/format liczby). Otwarte w ten sposób okno daje Ci wiele możliwości dotyczących formatów liczbowych.
2. Ustaw format liczbowy na *Currency (Custom)* [waluta (niestandardowy)], a *Decimal places* (liczba miejsc po przecinku) na 0 i *Display Units* (jednostki wyświetlanych wartości) na *Thousands (K)* (tysiące), na końcu naciśnij *OK*, tak jak pokazano na rysunku 2.17.



Rysunek 2.17. Zmiana domyślnego formatu liczby pola

3. Po kliknięciu *OK* powinieneś zauważyć, że etykiety na mapie zostały zaktualizowane i mają teraz symbol waluty.
4. Jeszcze raz kliknij prawym przyciskiem myszy pole *Profit* i wybierz *Default properties/Color...* (domyślne wartości/kolor). Otworzy się nowe okno, w którym będziesz mógł wybrać i zmienić domyślną kolorystykę pola *Profit*. Poeksperymentuj z różnymi paletami kolorów i ustawieniami. Zauważ, że za każdym razem, gdy naciśniesz przycisk *Apply* (zastosuj), wizualizacja się zmienia, tak jak na rysunku 2.18.

Wskazówka

Różne palety kolorów (palety, w których jeden kolor przechodzi w drugi) działają szczególnie dobrze dla pól takich jak *Profit*, które mogą mieć dodatnie i ujemne wartości. Dzięki domyślnej wartości środkowej, którą jest 0, na podstawie koloru można łatwo określić wartości dodatnie i ujemne.



Rysunek 2.19. Dane z oryginalnego źródła danych są wyciągane w formie zrzutu

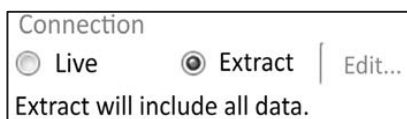
Gdy używasz połączenia na żywo, Tableau wysyła zapytania bezpośrednio do źródła danych (lub używa danych w pamięci podręcznej, jeśli to możliwe). Gdy wyciągasz dane, Tableau eksportuje część lub wszystkie dane z oryginalnego źródła danych i zapisuje w pliku. Wersje Tableau starsze od 10.5 używały pliku *.tde* (Tableau Data Extract). Począwszy od wersji 10.5, Tableau używa plików *.hyper* i przekształca pliki *.tde* na *.hyper*, gdy zaktualizujesz starsze skoroszyty.

Podstawowy paradygmat pracy Tableau z danymi nie zmienił się, ale możesz zauważyć, że teraz Tableau wysyła zapytania i odbiera wyniki z wyciągu. Dane mogą zostać pobrane ze źródła ponownie, by zaktualizować wyciąg. Dlatego każdy wyciąg jest zrzutem (ang. *snapshot*) danych ze źródła w momencie ostatniego odświeżenia. Wyciągi mają tę przewagę, że są bardzo wydajne i można je łatwo przenosić.

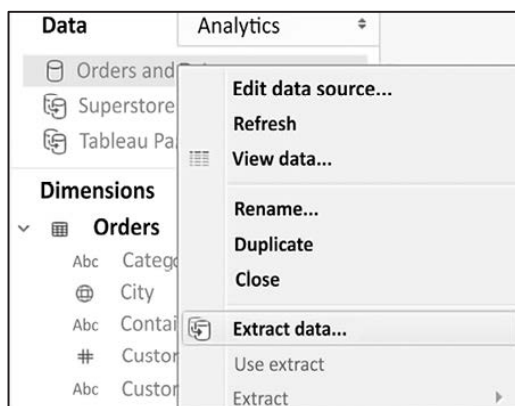
Tworzenie wyciągów

Wyciąg możesz stworzyć na wiele sposobów:

- Na ekranie *Data Source* (widocznym na rysunku 2.20) wybierz *Extract*. Link *Edit...* (edycja) pozwala na konfigurację wyciągu.
- Wybierz źródło danych z menu *Data* lub kliknij prawym przyciskiem myszy źródło danych w panelu danych i wybierz *Extract Data...*. Na rysunku 2.21 pokazano ustawienia dla wyciągu, które będziesz mógł dostosować.



Rysunek 2.20. Wybór trybu Live lub Extract dla połączenia oraz konfiguracja ustawień wyciągu za pomocą linku Edit



Rysunek 2.21. Opcja Extract data...

- Programiści mogą tworzyć wyciągi za pomocą Tableau Hyper API. API pozwala na użycie języków programowania takich jak Python, Java, C++ lub C#/.NET, by za pomocą kodu czytać i zapisywać wyciągi Hyper. Szczegóły tego podejścia są poza zakresem tej książki, ale na stronie Tableau znajdziesz niezbędną dokumentację (w języku angielskim): https://help.tableau.com/current/api/hyper_api/en-us/index.html.
- Niektóre narzędzia, takie jak Alteryx czy Tableau Prep, mogą zwracać wyciągi Tableau. W rozdziale 16., „Poskramianie danych za pomocą Tableau Prep”, omówimy możliwości zwracania danych przez Tableau Prep.

Będziesz mógł skonfigurować wyciąg na kilka sposobów. Aby to zrobić, wybierz *Extract*, a następnie *Edit...* na ekranie *Data Source* lub *Extract Data...* z menu kontekstowego połączenia w panelu danych. Podczas konfiguracji wyciągu zostaniesz zapytany o wybór pewnych opcji, tak jak pokazano na rysunku 2.22.

Możesz skonfigurować wiele ustawień związanych z wyciągiem. Oto różne opcje i ich wpływ na wydajność i elastyczność:

- W zależności od źródła danych i stworzonego przez Ciebie modelu obiektowego możesz wybierać między tabelami logicznymi (*Logical Tables*) a tabelami fizycznymi (*Physical Tables*). Więcej szczegółów znajdziesz w rozdziale 14., „Modele danych, łączenie danych i blendowanie w Tableau”.

Extract Data [X]

Specify how to store data in the extract:

Data Storage

☒ Logical Tables ☐ Physical Tables

Store data using one table for each logical table. [Learn more](#)
Use this option if you need to use extract filters, aggregation, or other extract settings.

Specify how much data to extract:

Filters (optional)

Filter	Details
Region	keeps Central and South
Category	keeps Office Machines

[Add...] [Edit...] [Remove]

Aggregation

☐ Aggregate data for visible dimensions

☐ Roll up dates to [Year]

Number of Rows

☒ All rows ☐ Incremental refresh

☐ Top: [Orders] by [] rows

☐ Sample: [Orders] by [] [rows]

[History...] [Hide All Unused Fields] [Extract] [Cancel]

Rysunek 2.22. Okno Extract Data daje sporo możliwości konfiguracji wyciągu

- Możesz też dodać filtry (*Filters*), które ograniczą liczbę wyciągniętych danych do podzbioru oryginalnego źródła. W tym przykładzie w wyciągu zostaną uwzględnione tylko rekordy z centralnego (Central) lub południowego (South) regionu (Region), a kategoria (Category) to maszyny biurowe (Office Machines).
- Możesz zagregować wyciąg przez zaznaczenie pola wyboru w części *Aggregation*. Dane będą zagregowane do poziomu szczegółowości widocznych rozmiarów lub do określonego poziomu daty, na przykład roku lub miesiąca.

Uwaga

Widoczne pola to te, które są pokazane w panelu danych. Możesz ukryć pole z poziomu ekranu *Data Source* lub w panelu danych przez kliknięcie pola prawym przyciskiem i wybranie *Hide* (ukryj). Ta opcja nie będzie aktywna, jeśli dane pole jest używane w którymś z widoków w skoroszytcie. Ukryte pola nie mogą być użyte w widoku i nie są ujęte w wyciągu, jeśli nie były widoczne podczas jego tworzenia lub optymalizacji.

W poprzednim przykładzie, gdyby wymiary *Region* i *Category* były widoczne, przygotowany wyciąg zawierałby tylko dwa wiersze danych (jeden wiersz dla regionu *Central* i drugi dla *South*). Ponadto każda miara zostałaby zagregowana na poziomie *Region/Category* zgodnie z ustawionymi filtrami wyciągu. Na przykład, pole *Sales* zawierałoby sumę sprzedaży w regionie *Central* dla kategorii *Office Machines* i w regionie *South* dla tej samej kategorii. Wszystkie miary są agregowane zgodnie ze swoją domyślną agregacją.

Możesz dostosować liczbę wierszy w wyciągu przez uwzględnienie wszystkich wierszy lub tylko *n* pierwszych wierszy w zestawie danych. Jeśli wybierzesz wszystkie wiersze, możesz wybrać odświeżanie przyrostowe. Jeśli Twoje źródło danych dodaje rekordy przyrostowo i masz pole z identyfikatorem lub pole daty, których można z powodzeniem użyć do określenia, czy dany rekord jest nowy, wtedy wyciąg przyrostowy pozwoli Ci dodać do wyciągu tylko te rekordy, bez potrzeby ponownego tworzenia całego wyciągu. W poprzednim przykładzie w następnym odświeżeniu przyrostowym zostaną uwzględnione wszystkie nowe wiersze, których identyfikator wiersza (*Row ID*) jest większy niż najwyższa wartość poprzedniego wyciągu.

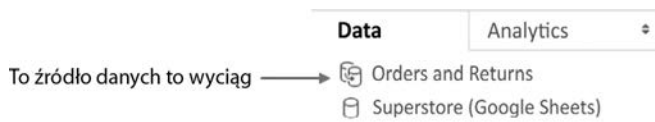
Wskazówka

Odświeżenia przyrostowe mogą być doskonałą metodą radzenia sobie z bardzo dużymi ilościami danych, które z czasem się powiększają. Jednak odświeżania przyrostowego używaj ostrożnie, ponieważ zostaną dodane tylko nowe wiersze na podstawie określonego przez Ciebie pola. Nie pobierzesz zmian w istniejących wierszach i nie usuniesz wierszy, których już nie ma w źródle danych. Co więcej, pominiesz nowe wiersze, jeśli wartość sprawdzanego pola będzie mniejsza niż maksymalna wartość istniejącego wyciągu. Jeśli któryś z tych scenariuszy jest prawdopodobny w Twoim źródle danych, lepiej zrezygnować z odświeżania przyrostowego.

Gdy już wiesz, jak stworzyć lub skonfigurować wyciągi, dowiedzmy się, jak ich używać.

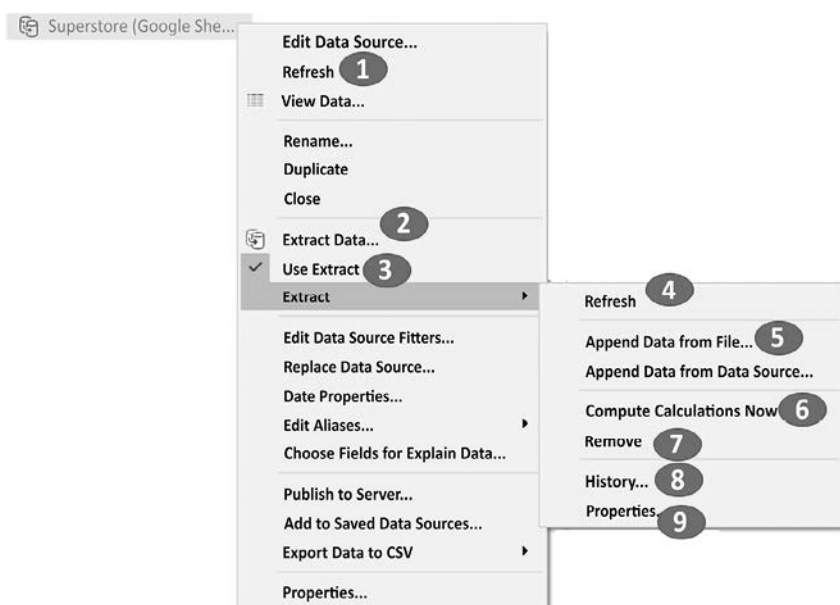
Używanie wyciągów

Każde źródło danych z trybem *Extract* będzie miało osobą ikonę (pokazaną na rysunku 2.23), wskazującą na to, że dane zostały pobrane z oryginalnego źródła do wyciągu.



Rysunek 2.23. Ikona znajdująca się obok źródła danych wskazuje, czy to wyciąg, czy nie

Pierwsze połączenie z danymi w panelu danych z rysunku 2.23 to wyciąg, a drugie nie. Po utworzeniu wyciągu możesz zdecydować, czy go użyć, czy nie. Gdy klikniesz prawym przyciskiem myszy źródło danych (lub *Data* z menu, a potem źródło danych), zobaczysz opcje pokazane i oznaczone na rysunku 2.24.



Rysunek 2.24. Menu kontekstowe połączenia z danymi w panelu danych z ponumerowanymi opcjami dla wyciągu

Przyjrzyjmy się tym opcjom bliżej.

1. **Refresh.** Opcja odświeżania po prostu informuje Tableau, by odświeżać lokalną pamięć podręczną z danymi. W przypadku połączenia na żywo ta opcja będzie wymagać ponownego wysłania zapytania do źródła. Mając źródło w trybie wyciągu, pamięć podręczna jest czyszczona i niezbędny jest wyciąg, ale opcja *Refresh* nie odświeża wyciągu z oryginalnego źródła danych. W tym celu użyj opcji *Refresh* znajdującej się w podmenu *Extract* (patrz punkt 4. na tej liście).
2. **Extract Data....** Ta funkcja tworzy nowy wyciąg ze źródła danych (zastępujący istniejący wyciąg).

3. **Use Extract.** Ta opcja jest aktywna, jeśli istnieje wyciąg dla danego źródła danych. Jej odznaczenie spowoduje, że Tableau będzie używało połączenia zamiast wyciągu. Wyciąg nie zostanie usunięty i można go użyć ponownie — wystarczy, że w dowolnym momencie zaznaczysz tę opcję. Jeśli w skoroszybie pierwotne źródło danych nie jest dostępne, Tableau spyta Cię, gdzie je znaleźć.
4. **Refresh.** Ta opcja odświeża wyciąg z danymi z pierwotnego źródła danych. Nie optymalizuje wyciągu pod kątem dokonanych przez Ciebie zmian (takich jak ukrywanie pól czy tworzenie nowych obliczeń).
5. **Append Data from File...** lub **Append Data from Data Source....** Te opcje pozwalają na dołączenie dodatkowych plików lub źródeł danych do istniejącego wyciągu, pod warunkiem że mają taką samą strukturę danych jak pierwotne źródło. Do Twojego istniejącego wyciągu zostaną dodane wiersze, ale nie nowe kolumny.
6. **Compute Calculations Now.** By wyciąg był jak najwydajniejszy, ta opcja zmienia jego strukturę na podstawie zmian, które wprowadziłeś od pierwszego utworzenia wyciągu. Niektóre pola obliczeniowe mogą być obliczane tylko raz z zapisaniem wyniku, a nowo ukryte kolumny lub usunięte obliczenia zostaną usunięte z wyciągu.
7. **Remove.** Ta opcja usuwa definicję wyciągu. Może również usunąć plik wyciągu i wznowić połączenie na żywo z pierwotnym źródłem danych.
8. **History....** Za pomocą tej opcji możesz wyświetlić historię wyciągu i odświeżania.
9. **Properties....** Ta opcja umożliwia wyświetlenie właściwości wyciągu, takich jak lokalizacja, źródło danych, filtry czy ograniczenia związane z wierszami.

W dalszej części tego podrozdziału dowiemy się, jak tworzyć wyciągi z uwzględnieniem wpływu na wydajność podczas korzystania z nich.

Wydajność

W Tableau mamy dwa rodzaje ekstraktów:

- Tableau Data Extracts (pliki *.tde*). Przed wersją Tableau 10.5 to był jedyny dostępny rodzaj wyciągów.
- Wyciągi Hyper (pliki *.hyper*), dostępne w wersji 10.5 i późniejszych.

W zależności od skali i wielkości zarówno wyciągi *.hyper*, jak i *.tde* mogą działać szybciej niż większość tradycyjnych połączeń na żywo z bazą danych. W większości przypadków Tableau domyślnie będzie tworzyć wyciąg Hyper. Jeśli nie używasz starszych wersji Tableau, jest mało powodów, by używać starszego *.tde*. Niesamowita wydajność wyciągów Tableau opiera się na kilku czynnikach:

- Wyciągi Hyper wykorzystują połączenie przetwarzania OLTP (ang. *Online Transactional Processing*) i modeli OLAP, a silnik określa optymalne zapytanie. Wyciągi *.tde* są kolumnowe i bardzo sprawnie działają z zapytaniami.
- Wyciągi są ustrukturyzowane, więc można je bardzo szybko wgrywać do pamięci bez konieczności dodatkowego przetwarzania oraz przenosić między pamięcią a dyskiem. Rozmiar nie jest zatem ograniczony do dostępnego miejsca w pamięci RAM, a RAM jest skutecznie wykorzystywany do poprawy wydajności.
- W wyciągu wiele pól obliczeniowych jest zapisywanych jako obliczone wartości. Zapisana w wyciągu wcześniej obliczona wartość często może zostać odczytana dużo szybciej niż w przypadku wykonywania obliczenia przy każdym zapytaniu. Wyciągi Hyper robią tak również dla wielu agregacji.

Możesz wybierać między wyciągami, by zwiększyć wydajność, a tradycyjnymi bazami danych. Aby zwiększyć wydajność, weź pod uwagę następujące działania:

- Przed utworzeniem wyciągu ukryj nieużywane pola. Jeśli stworzyłeś wszystkie zaplanowane wizualizacje, w oknie *Extract* możesz nacisnąć przycisk *Hide Unused Fields* (ukryj nieużywane pola), by ukryć wszystkie pola, które nie są częścią wizualizacji ani nie biorą udziału w obliczeniach.
- Jeśli to możliwe, użyj podzbioru danych z pierwotnego źródła. Na przykład, jeśli masz historyczne dane za 10 ostatnich lat, ale musisz przeprowadzić analizę tylko dla dwóch ostatnich, zastosuj w wyciągu filtr na polu *Date* (data).
- Zoptymalizuj wyciąg po utworzeniu lub zmianie pól obliczeniowych albo po usunięciu lub ukryciu pól.
- Zapisz wyciąg na dysku twardym.

Choć głównym powodem używania wyciągów jest wydajność, należy wziąć pod uwagę jeszcze inne czynniki, co zrobimy w następnej kolejności.

Przenaszalność i bezpieczeństwo

Powiedzmy, że Twoje dane są przechowywane w bazie danych na serwerze, do którego masz dostęp, tylko będąc podłączonym do sieci biurowej. By pracować z danymi, musisz być na miejscu lub połączyć się za pomocą sieci VPN. Nawet chmurowe źródła danych wymagają połączenia z internetem. W przypadku wyciągu możesz wziąć swoje dane ze sobą i pracować nad nimi offline.

Plik z wyciągiem zawiera dane wyeksportowane ze źródła. Podczas zapisu skoroszytu możesz wybrać plik skoroszytu Tableau (*.twb*) lub plik spakowanego skoroszytu Tableau (*.twbx*). Zobaczmy, jakie są między nimi różnice:

- Skoroszyt Tableau (.twb) zawiera definicje dla wszystkich połączeń, pól, wizualizacji i pulpitów nawigacyjnych, ale nie zawiera żadnych danych czy zewnętrznych plików, na przykład obrazów. Skoroszyt Tableau można edytować w programie Tableau Desktop i opublikować w aplikacji Tableau Server.
- Spakowany skoroszyt Tableau (.twbx) zawiera wszystko to, co plik .twb, ale jego częścią są również wyciągi i zewnętrzne pliki, spakowane do jednego pliku ze skoroszytem. Spakowany skoroszyt z wyciągami można otworzyć w programie Tableau Desktop albo Tableau Reader i opublikować w aplikacji Tableau Public lub Tableau Online.

Wskazówka

Spakowany skoroszyt Tableau (.twbx) jest tak naprawdę po prostu skompresowanym plikiem .zip. Jeśli zmienisz rozszerzenie pliku z .twbx na .zip, możesz zobaczyć jego zawartość, tak jak w przypadku standardowych plików .zip.

Jest kilka kwestii dotyczących bezpieczeństwa, które należy wziąć pod uwagę podczas używania wyciągu. Przede wszystkim, po utworzeniu wyciągu nie będzie działała żadna warstwa bezpieczeństwa ograniczająca dostęp do danych. Wyciąg nie wymaga podania nazwy użytkownika i hasła. Każdy może przeczytać wszystkie dane w wyciągu. Po drugie, można łatwo zobaczyć wszystkie dane widocznych (nieukrytych) pól zawartych w pliku wyciągu .hyper lub .tde albo w wyciągu będącym częścią spakowanego skoroszytu (.twbx), nawet jeśli te dane nie są pokazane w wizualizacji. Bądź ostrożny w pracy z wyciągami, by niepowołane osoby nie miały dostępu do wrażliwych lub osobistych danych.

Zastosowanie wyciągów

Decydując się na użycie wyciągu, powinieneś wziąć pod uwagę kilka różnych czynników. Czasami nie masz wyboru (na przykład OLAP wymaga połączenia na żywo, a niektóre chmurowe źródła danych wymagają wyciągów). W pozostałych przypadkach należy ocenić dostępne możliwości.

Z zasady wyciągów używaj w następujących sytuacjach:

- Potrzebujesz lepszej wydajności niż ta oferowana przez połączenie na żywo.
- Musisz mieć możliwość przenoszenia danych.
- Musisz użyć funkcji, które nie są obsługiwane przez silnik bazy danych (na przykład funkcja MEDIAN nie jest obsługiwana w przypadku połączenia na żywo z SQL Server).

- Chcesz udostępnić spakowany skoroszyt. To jest szczególnie ważne, jeśli chcesz udostępnić spakowany skoroszyt komuś, kto używa darmowego programu Tableau Reader, za pomocą którego można tylko odczytywać spakowane skoroszyty z wyciągniętymi danymi.

Natomiast unikaj wyciągów w następujących przypadkach:

- Masz wrażliwe dane, do których mogą mieć dostęp tylko uprawnieni użytkownicy, albo nie wiesz, kto będzie miał dostęp do wyciągu. Przed utworzeniem wyciągu możesz ukryć wrażliwe pola, przez co nie będą już jego częścią. Jednak wyciągi w aplikacji Tableau Server można zablokować na poziomie wiersza.
- Musisz zarządzać bezpieczeństwem przy użyciu nazwy użytkownika i hasła. (Jednak jeżeli masz Tableau Server, wciąż możesz używać wyciągów przechowywanych na Tableau Server, które wymagają zalogowania. Udostępnianie pracy w Tableau Server zostanie omówione w rozdziale 17., „Dzielenie się swoją historią”).
- Musisz widzieć zmiany w danych źródłowych w czasie rzeczywistym.
- Ilość danych sprawia, że czas potrzebny do ich wyciągnięcia jest zbyt długi. Liczba rekordów, które mogą zostać wyciągnięte w rozsądnym czasie, zależy od takich czynników, jak typy danych pól, liczba pól, prędkość źródła danych i przepustowość sieci. Silnik Hyper zwykle buduje wyciągi *.hyper* dużo szybciej, niż to było w przypadku plików *.tde*.

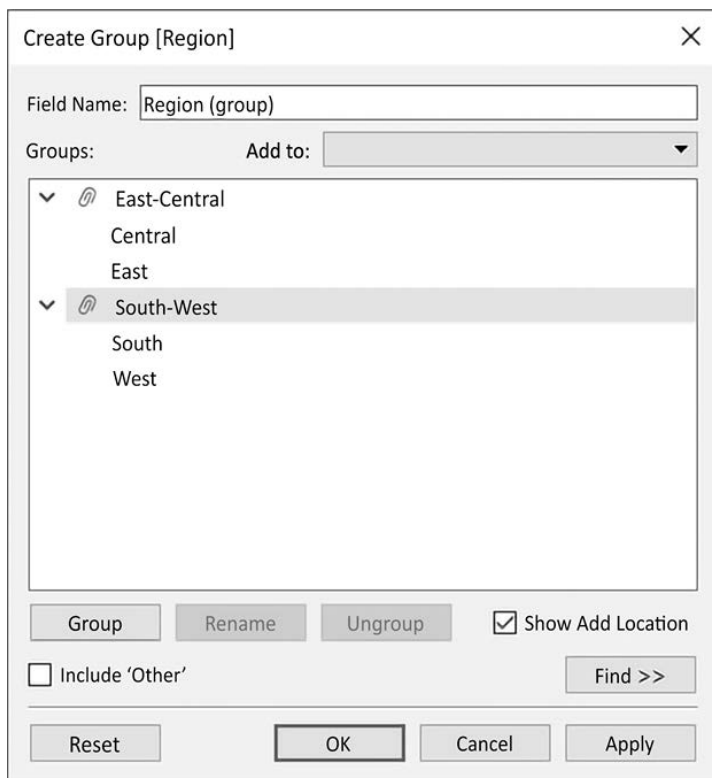
Gdy już rozumiesz, jak tworzyć wyciągi, zarządzać nimi i ich używać (i kiedy ich nie stosować), przejdziemy do zestawów i grup danych, które stanowią alternatywną metodę grupowania powiązanych ze sobą danych, a następnie odkryjemy różne sposoby filtrowania danych w Tableau.

Używanie zestawów i grup

W Tableau zestawy i grupy są powiązаныmi koncepcjami umożliwiającymi grupowanie powiązanych ze sobą danych. W obu przypadkach grupowanie jest definiowane w Tableau, a nie w źródle danych.

Grupa jest połączeniem wartości z jednego wymiaru lub ich większej liczby. Na przykład w przypadku czterech regionów: *Central*, *East*, *South* i *West*, możesz stworzyć dwie grupy: *East-Central* i *South-West*. To pozwoli Ci porównać dane z nieco innej perspektywy, co może przełożyć się na nowe wnioski na wyższym poziomie. Grupę możesz utworzyć przez wybranie *Create/Group* (utwórz/grupuj) z menu kontekstowego pola.

Otworzy się okno, pokazane na rysunku 2.25, w którym możesz zdefiniować grupy.



Rysunek 2.25. Okno Create Group umożliwia grupowanie wartości danego wymiaru

Wynikiem jest pole `Region (group)`, którego możesz użyć tak jak każdego innego wymiaru w Twoich danych.

Wskazówka

Możesz również tworzyć grupy *ad-hoc* przez zaznaczenie znaczników w swojej wizualizacji. Przykład tworzenia grup *ad-hoc* został pokazany w punkcie „Odmiany wykresów słupkowych” w rozdziale 3., „Wizualizacje — kolejny krok”.

Zestaw w Tableau definiuje zbiór rekordów z bazy danych. Na poziomie wiersza każdy rekord jest albo w zestawie, albo poza zestawem. Są dwa rodzaje zestawów:

- dynamiczne (czasami nazywane zestawami obliczeniowymi),
- stałe.

Zestaw dynamiczny jest obliczany dla pojedynczego wymiaru na podstawie zdefiniowanego przez Ciebie warunku. Wraz ze zmianą danych wyniki warunku mogą się zmieniać i rekordy mogą wchodzić do zestawu lub poza niego wypadać.

Na przykład gdybyś w menu rozwijanym w polu Customer Name (nazwa klienta) w panelu danych wybrał *Create/Set*, mógłbyś zdefiniować warunek określający, które rekordy należą do zestawu (patrz rysunek 2.26).

Dialog box titled "Edit Set [Customers who purchased more than \$100]".

Name: Customers who purchased more than \$100

General Condition Top

☐ None

☒ By field:

Sales Sum

> 100

Range of Values

Min: Load

Max:

☐ By formula:

Reset OK Cancel Apply

Rysunek 2.26. Zestaw dynamiczny oparty na warunku

W tym przykładzie stworzyliśmy zestaw dynamiczny o nazwie *Customers who purchased more than \$100* (klienci, którzy dokonali zakupu powyżej 100 dolarów) z warunkiem mówiącym, że suma sprzedaży musi być większa niż 100. Zauważysz pewnie też opcje obliczania według formuły (*By formula*) lub pierwszych *n* wartości (*Top N*). Wszystkie te warunki będą na poziomie zagregowanych wartości (w całym zestawie danych lub w całym kontekście, jeśli są zastosowane filtry), a następnie każdy rekord jest oceniany, czy należy do zestawu, czy jest poza nim.

W tym przypadku całkowita kwota sprzedaży dla każdego klienta będzie obliczana w całym zestawie danych, a następnie każdy rekord zostanie sprawdzony, by dowiedzieć się, czy sprzedaż dla klienta z tego rekordu jest większa niż 100 dolarów.

Stały zestaw to lista wartości dla jednego wymiaru lub ich większej liczby. Jeśli wartości dla pojedynczego rekordu pasują do listy zdefiniowanej przez zestaw, wtedy rekord jest częścią zestawu, a w przeciwnym razie jest poza nim. Na przykład możesz utworzyć inny zestaw na podstawie pola Customer Name, tym razem z zakładką *General* (patrz rysunek 2.27).

Create Set

Name: Customers with the first name of Aaron

General Condition Top

☒ Select from list ☐ Custom value list ☐ Use all

Enter search text

- ☒ Aaron Davies Bruce
- ☒ Aaron Day
- ☒ Aaron Dillon
- ☒ Aaron Fuller Davidson
- ☒ Aaron Riggs
- ☐ Aaron Shaffer
- ☐ Adam Barton
- ☐ Adam G Sawyer
- ☐ Adam McKinney
- ☐ Adam Saunders Gray
- ☐ Alan Atkins

All None ☐ Exclude

Summary

Field: [Customer Name]

Selection: Selected 5 of 2708 values

Wildcard: All

Condition: None

Limit: None

Reset OK Cancel

Rysunek 2.27. Zestaw statyczny opierający się na wyborze członków

Tutaj możesz wybrać pojedyncze wartości, które określają, co jest w zestawie, a co poza nim. Zwróć uwagę na opcję wykluczania z zestawu zamiast określania, co ma być częścią zestawu. W tym przypadku stworzyliśmy zestaw o nazwie *Customers with first name of Aaron* (klienci o imieniu Aaron). Każdy rekord, który w polu Customer Name ma wartość pasującą do jednej z 6 zaznaczonych, będzie częścią zestawu. Wszystkie pozostałe będą wykluczone. Ponieważ jest to stały zestaw, wartości nie są nigdy obliczane ani ponownie obliczane. Jeśli w przyszłym tygodniu pojawi się rekord dla klienta o imieniu i nazwisku Aaron Burr, to będzie cały czas znajdował się poza zestawem.

W rozdziale 8., „Opowiadanie historii za pomocą pulpitu nawigacyjnego”, zobaczymy dokładniej, jak to działa, gdy będziemy omawiać, jak używać zestawu działań, by wzbogacić możliwości interakcji i analizy w pulpitych nawigacyjnych. Na razie podsumujmy to, czego dowiedzieliśmy się o używaniu danych w Tableau, przez spojrzenie na różne sposoby filtrowania danych.

Filtrowanie danych

Często będziesz miał potrzebę wyfiltrowania danych w Tableau w celu przeprowadzenia analizy na podzbiorze danych, zawężenia danych, na których chcesz się skupić, lub zagłębienia się w dane. Tableau oferuje wiele sposobów filtrowania danych.

Choć możesz zastosować dowolną liczbę filtrów w dowolnym widoku, jeśli chcesz ograniczyć zakres swojej analizy do podzbioru danych, możesz wyfiltrować dane u źródła za pomocą jednej z tych technik:

- **Filtry źródła danych** są stosowane przed wszystkimi innymi filtrami i są przydatne, gdy chcesz ograniczyć swoją analizę do podzbioru danych.
- **Filtry wyciągu** ograniczają dane przechowywane w wyciągu (.tde lub .hyper). Filtry źródła danych są często przekształcane w filtry wyciągu, gdy są obecne podczas eksportowania danych.
- **Niestandardowe filtry SQL** mogą zostać zastosowane przy połączeniu na żywo przy użyciu niestandardowego zapytania SQL, które ma parametr Tableau w wyrażeniu WHERE. Więcej szczegółów na temat parametrów znajdziesz w rozdziale 4., „Początek przygody z obliczeniami i parametrami”.

Jeśli Twoje filtry dotyczą tylko określonych widoków, możesz zastosować filtry do tych widoków za pomocą jednej z następujących technik:

- Przeciągnij pola z panelu danych na półkę *Filters*.
- Zaznacz jeden lub więcej znaczników lub nagłówki w widoku, a następnie wybierz *Keep Only* (zachowaj tylko) lub *Exclude* (wyklucz), tak jak pokazano na rysunku 2.28.

✓ Keep Only ✕ Exclude

Rysunek 2.28. W zależności od wyboru znaczników możesz zaznaczyć, by zachować te wartości lub je wykluczyć

- Kliknij prawym przyciskiem myszy dowolne pole w panelu danych lub w widoku i wybierz *Show Filter* (pokaż filtr). Filtr zostanie pokazany jako kontrolka (na przykład lista rozwijana albo pole wyboru), by użytkownik końcowy widoku lub pulpitu nawigacyjnego mógł zmieniać wartości filtra.
- Użyj filtra akcji. W rozdziale 8., „Opowiadanie historii za pomocą pulpitu nawigacyjnego”, przyjrzymy się bliżej filtrom i filtrom akcji, gdyż są ściśle związane z pulpitem nawigacyjnym.

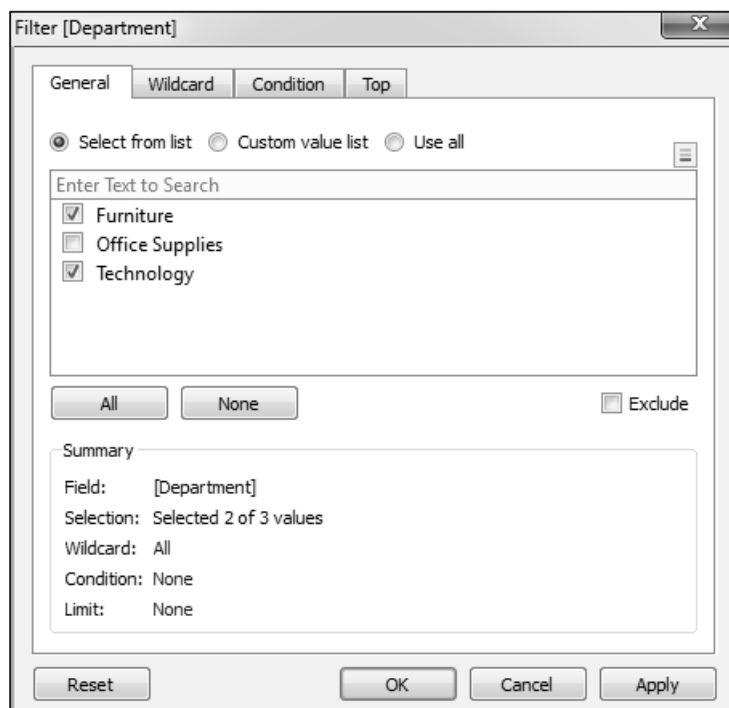
Każda z tych opcji dodaje jedno lub więcej pól na półkę *Filters* widoku. Gdy wrzucisz pole na półkę *Filters*, zostaniesz zapytany o zdefiniowanie filtra. Opcje filtra będą się znacząco różnić w zależności od tego, czy pole jest ciągłe, czy nieciągłe. To, czy pole jest filtrowane jako wymiar, czy jako miara, będzie miało duży wpływ na sposób zastosowania filtra i na wyniki.

Filtrowanie nieciągłych (niebieskich) pól

Gdy filtrujesz za pomocą pola nieciągłego będziesz miał możliwość wybrania pojedynczych wartości, by je zachować lub wykluczyć. Na przykład, gdy przeciągniesz nieciągły wymiar *Department* na półkę *Filters*, Tableau da Ci opcje pokazane na rysunku 2.29.

Opcje filtra mają zakładki *General* (ogólne), *Wildcard* (wieloznacznik), *Condition* (warunek) oraz *Top* (pierwsze/ostatnie wartości). Twój filtr może obejmować wybrane opcje z każdej zakładki. Część *Summary* (podsumowanie) w zakładce *General* pokaże wszystkie wybrane opcje.

- Zakładka *General* pozwala wybrać elementy z listy (możesz użyć niestandardowej listy, by dodawać elementy ręcznie, jeśli wymiar zawiera bardzo dużą liczbę wartości, które długo się wgrywają). Możesz użyć opcji *Exclude*, by wykluczyć wybrane elementy.
- Zakładka *Wildcard* umożliwia dopasowanie łańcuchów znaków, które zawierają podaną wartość, zaczynają się od niej lub się na nią kończą albo dokładnie pasują do wybranej wartości. Na przykład możesz wpisać *John*, by ograniczyć listę klientów tylko do tych, którzy mają na nazwisko *Johnson* lub *Johnston*.
- Zakładka *Condition* pozwala na określenie warunków opartych na agregacjach innych pól spełniających warunki (na przykład warunek, by zachować każdy dział, dla którego suma sprzedaży była większa niż 1 000 000 dolarów). Co więcej, możesz napisać własne obliczenie, by tworzyć złożone warunki.



Rysunek 2.29. Filtr dla nieciągłego pola pokaże opcje w celu włączania lub wykluczania pojedynczych wartości

O obliczeniach dowiesz się więcej w rozdziale 4., „Początek przygody z obliczeniami i parametrami”, oraz rozdziale 6., „Obliczenia tablicowe”.

- Zakładka *Top* umożliwia ograniczenie filtra tylko do górnych lub dolnych wierszy. Na przykład możesz postanowić zachować tylko pięć pierwszych elementów listy sum sprzedaży.

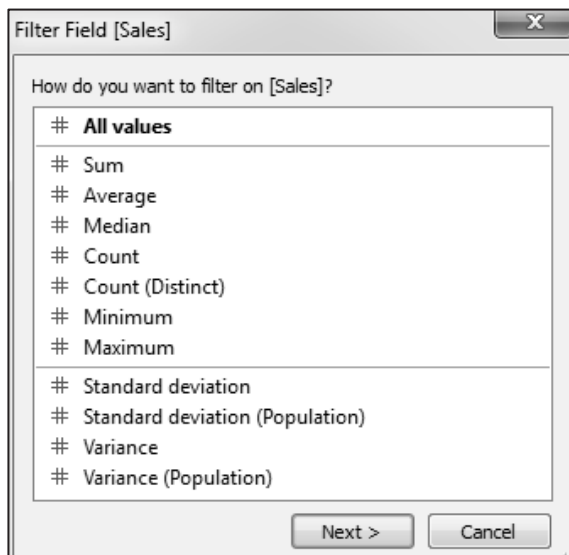
Wskazówka

Nieciągłe miary (poza polami obliczeniowymi używającymi obliczeń tablicowych) nie mogą być dodawane na półkę *Filters*. Jeśli to pole przechowuje datę lub wartość liczbową, możesz przed zastosowaniem filtra przekształcić je na pole ciągłe. Inne typy danych będą się wiązać z koniecznością utworzenia pola obliczeniowego w celu przekształcenia tych wartości, które chcesz przefiltrować na ciągłe wartości liczbowe.

Zobaczmy, jak filtrowane są pola ciągłe.

Filtrowanie ciągłych (zielonych) pól

Jeśli przeciągniesz ciągły wymiar na półkę *Filters*, będziesz miał inny zestaw ustawień do wyboru. Często najpierw zostaniesz zapytany, jak chcesz przefiltrować pole, co pokazano na rysunku 2.30.



Rysunek 2.30. Dla wartości liczbowych często będziesz widział opcję dla agregacji jako części filtra

Te opcje są tutaj podzielone na dwie główne kategorie:

- **Wszystkie wartości.** Filtr będzie się opierał na każdej pojedynczej wartości pola, wiersz po wierszu. Na przykład filtr *All values* (wszystkie wartości), zachowujący tylko wartości powyżej 100 dolarów, obliczy każdy rekord w danych i zachowa tylko pojedynczą sprzedaż powyżej 100 dolarów.
- **Agregacja.** Filtr będzie się opierał na określonej agregacji [na przykład suma (*Sum*), średnia (*Average*), minimum (*Minimum*), maksimum (*Maximum*), odchylenie standardowe (*Standard deviation*) i wariancja (*Variance*)], a agregacja zostanie przeprowadzona na poziomie szczegółowości widoku. Na przykład filtr zachowujący tylko sumę sprzedaży powyżej 100 000 dolarów w widoku na poziomie kategorii zostawi tylko te kategorie, które mają sprzedaż przynajmniej na poziomie 100 000 dolarów.

Po dokonaniu wyboru (lub jeśli wybór nie dotyczył danego pola) otworzy się kolejne okno (patrz rysunek 2.31), w którym będziesz mógł ustawić docelowy filtr.



Rysunek 2.31. Opcje filtra dla pola Sales [jako suma (SUM)]

Tutaj zobaczysz opcje dla filtrowania wartości ciągłych na podstawie zakresu z wartością początkową lub wartością końcową albo oboma. W zakładce *Special* (specjalne) możesz ustawić, by pokazywane były wszystkie wartości, puste wartości lub niepuste wartości.

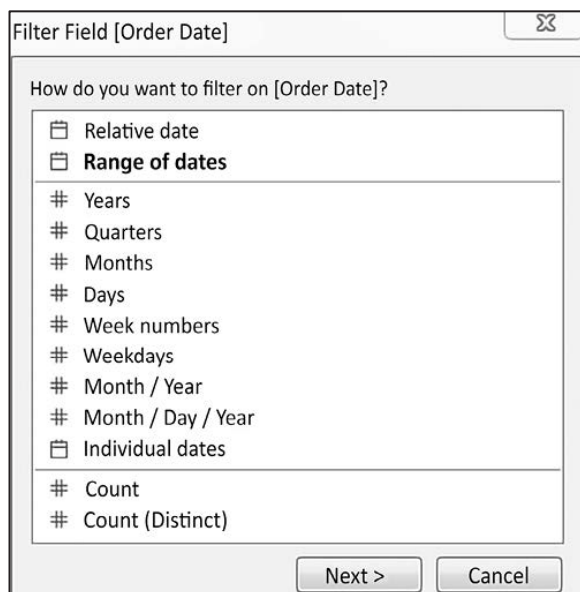
Z perspektywy interfejsu użytkownika największa różnica w opcjach filtrowania wynika z tego, czy pole jest ciągłe, czy nieciągłe. Jednak powinieneś zawsze wiedzieć, czy używasz pola jako **filtra wymiaru**, czy **filtra miary**, by rozumieć, jakiego rodzaju wyniki otrzymasz, gdyż na to wpływa kolejność działań, które zostaną omówione w punkcie „Kolejność działań w Tableau”.

- Filtry wymiaru zostawiają tylko *wiersze danych z określonymi wartościami*. Na przykład wyfiltrowanie regionu centralnego spowoduje wykluczenie wszystkich wierszy dla tego regionu. Nie zobaczysz żadnego stanu z tego regionu, a Twoje zagregowane wyniki, takie jak SUM(Sales), nie będą zawierać żadnych wartości z tego regionu.
- Filtry miary zostawiają tylko *zagregowane wiersze danych* na poziomie określonym przez wymiary w Twoim widoku. Na przykład jeśli Twój filtr zostawia tylko te wiersze, gdzie SUM(Sales) jest większe niż 100 000 dolarów, a Twój widok zawiera pola Region i Month, wtedy otrzymany w rezultacie widok będzie obejmował tylko wartości, dla których dany region miał sprzedaż wyższą niż 100 000 dolarów w danym miesiącu.

Poza filtrowaniem pól ciągłych i nieciągłych zauważysz jeszcze kilka ustawień filtrowania dat, które teraz omówimy.

Filtrowanie dat

W rozdziale 3., „Wizualizacje — kolejny krok”, przyjrzymy się specjalnej metodzie obsługi dat przez Tableau. Na razie omówimy opcje dostępne po przeciągnięciu pola *Order Date* na półkę *Filters*, pokazane na rysunku 2.32.



Rysunek 2.32. Początkowe ustawienia filtra dla pola daty

Mamy do wyboru następujące opcje:

- **Relative date** (data względna). Ta opcja pozwala na wyfiltrowanie daty na podstawie konkretnej daty (na przykład zachowanie tylko trzech ostatnich tygodni, licząc od dzisiaj, lub sześć miesięcy od 1 stycznia).
- **Range of dates** (zakres dat). Ta opcja pozwala Ci filtrować datę na podstawie zakresu z datą początkową, datą końcową lub oboma.
- **Część daty**. Ta opcja umożliwia filtrowanie na podstawie części daty, na przykład roku (*Years*), miesiąca (*Months*), dni (*Days*), lub zestawienia tych części, na przykład miesiąc/rok (*Month/Year*). W zależności od Twojego wyboru będziesz miał różne możliwości filtrowania oraz będziesz mógł ustawić datę domyślną na najświeższą datę w Twoich danych.

- **Individual dates** (pojedyncze daty). To ustawienie pozwala filtrować pod kątem każdej pojedynczej wartości pola daty.
- **Count** (zlicz) lub **Count (Distinct)** [zlicz (odrębne)]. Ta opcja umożliwia filtrowanie pod kątem liczby wystąpień lub liczby odrębnych wystąpień wartości daty.

Inne opcje filtrowania

Warto również znać poniżej podane opcje filtrowania:

- Możesz wyświetlać kontrolkę filtrowania dla niemal każdego pola. Wystarczy, że prawym przyciskiem myszy klikniesz dane pole i wybierzesz *Show Filter* (pokaż filtr). Rodzaj kontrolki zależy od rodzaju pola i od tego, czy jest ciągłe, czy nieciągłe. Można go dostosować za pomocą małej strzałki listy rozwijanej, znajdującej się w prawym górnym rogu kontrolki filtra.
- Filtry można dodawać do **kontekstu**. Kontekst został zilustrowany w punkcie „Kolejność działań w Tableau”, a w całej książce będziesz miał okazję zobaczyć wiele różnych przykładów kontekstu. Zwróć uwagę, że możliwość dodania pola użytego jako filtr jest dostępna w menu rozwijanym kontrolki *Filter* lub pola na półce *Filters*.
- Filtry mogą być ustawione, by pokazywać wszystkie wartości w bazie danych, wszystkie wartości w kontekście, wszystkie wartości w hierarchii lub tylko te wartości, które są właściwe ze względu na pozostałe filtry. Te opcje są dostępne w menu rozwijanym kontrolki *Filter* lub pola na półce *Filters*.
- Gdy używasz Tableau Server, możesz zdefiniować filtry użytkownika, które pozwalają zabezpieczyć dane na poziomie wiersza. Filtrowanie będzie się odbywać na podstawie nazwy użytkownika i hasła.
- Domyślnie każde pola umieszczone na półce *Filters* określa filtr szczególny dla bieżącego widoku. Jednak możesz określić zakres filtra za pomocą menu pola na półce *Filters*. Wybierz *Apply to* (zastosuj do), a potem jedną z następujących opcji:
 - **Wszystkie powiązane źródła danych.** Wszystkie źródła danych będą filtrowane według określonych wartości. Relacje między polami są tworzone tak samo jak łączenie [to znaczy domyślnie przez dopasowanie nazwy i typu danych lub niestandardowo za pomocą opcji w menu *Data/Edit Relationships* (dane/edycja relacji)]. Ten filtr będzie stosowany we wszystkich widokach używających dowolnego z powiązanych źródeł danych. Ta opcja jest czasami określana jako **filtrowanie krzyżowe źródła**.

- **Bieżące źródło danych.** Źródło danych dla tego pola będzie filtrowane. Ten filtr będzie stosowany we wszystkich widokach używających dowolnego z powiązanych źródeł danych.
- **Wybrane arkusze.** Filtr będzie stosowany w każdym wybranym arkuszu używającym danego źródła danych tego pola.
- **Bieżący arkusz.** Filtr będzie stosowany tylko w bieżącym widoku.

W książce zobaczymy bardzo dużo praktycznych przykładów filtrowania danych, a wiele z nich będzie korzystało z tych opcji. Na koniec tego rozdziału przyjrzymy się krótko kolejności działań w Tableau, gdzie filtry odgrywają główną rolę.

Kolejność działań w Tableau

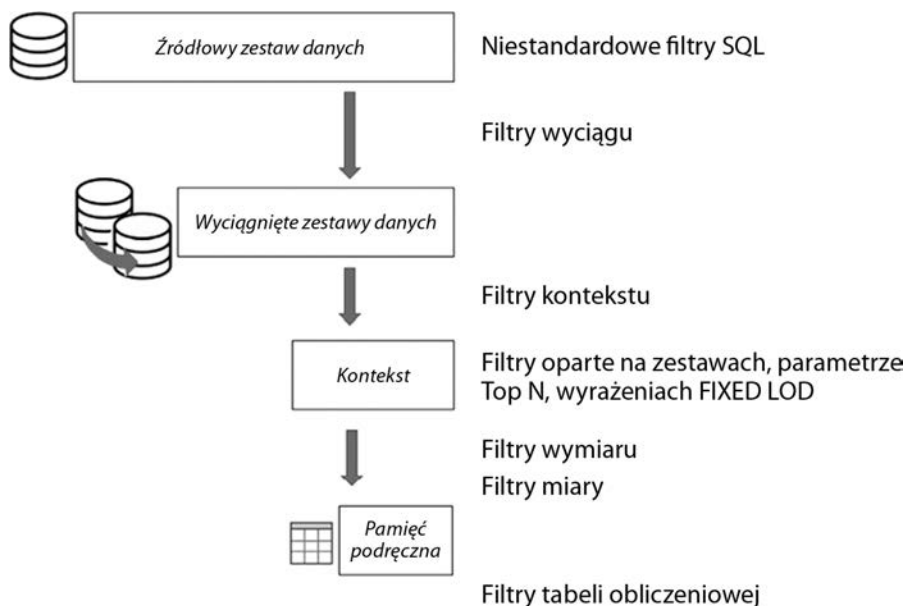
Jak zapewne pamiętasz, na początku tego rozdziału omawialiśmy podstawowy paradygmat działania Tableau. Podczas cyklu wykonywania zapytań i wizualizacji wyników Tableau będzie stosować różne filtry i obliczenia na podstawie tego, jak stworzyłeś daną wizualizację. Jeszcze nie mówiliśmy o obliczeniach, ale uwzględnimy je tutaj przy okazji omawiania kolejności działań, ponieważ ważne jest, by zobaczyć pełny obraz. Możliwe, że będziesz chciał wrócić do tego punktu po przeczytaniu rozdziałów o różnych typach obliczeń, by lepiej zrozumieć ich działanie.

Model myślowy przedstawiający kolejność stosowania filtrów i obliczeń przez Tableau można zbudować na wiele sposobów. Można to przedstawić w formie stopni lub schematów przepływu, które znajdziesz w internecie. Tutaj przyjrzymy się stopniowo malejącym zestawom danym, na których może pracować Tableau (patrz rysunek 2.33).

Pierwotny **zestaw źródłowy** to Twój plik Excel, tabele Snowflake lub każde inne pierwotne źródło danych, z którym się połączyłeś. Zawiera wszystkie dane, które mogą zostać użyte przez Tableau. Możesz pisać niestandardowe wyrażenia SQL będące częścią Twojego połączenia Tableau i one zostaną wykonane przed wszystkimi innymi działaniami.

Tworzenie wyciągu dla większości zestawów danych nie jest obowiązkowe. Jeśli jednak utworzysz **wyciąg**, wtedy możesz użyć filtrów wyciągu, by ograniczyć dane ze źródła zapisane w wyciągu.

Jak już zostało wspomniane wcześniej, możesz dodać filtry do **kontekstu**. To jeszcze bardziej ogranicza zestaw danych, na którym będzie pracować Tableau. W następnych rozdziałach omówimy zestawy oraz obliczenia LOD (*Level Of Detail*, poziom szczegółowości). Na tym etapie zwróć uwagę na to, że te i pozostałe filtry opierające się na nich, wraz z filtrami *Top N*, działają w obrębie kontekstu.

**Rysunek 2.33. Kolejność działań**

Widzieliśmy, jak wymiary i miary mogą być używane jako filtry. Zauważ, że wymiary są używane jako pierwsze (ponieważ mogą wykluczać wiersze, które nie powinny być uwzględniane w agregacji filtra miary), a później stosowane są filtry miary.

Wynik końcowy jest zapisywany w **tabeli w pamięci podręcznej** — zagregowanej tabeli wyników, których Tableau potrzebuje do wizualizacji. Więcej o obliczeniach tablicowych dowiesz się z rozdziału 6., „Obliczenia tablicowe”. Na razie po prostu zapamiętaj, że obliczenia tablicowe są wykonywane w pamięci podręcznej, więc każdy filtr oparty na obliczeniach tablicowych jest stosowany przez Tableau jako jedno z ostatnich działań.

Uwaga

Jest sporo działań, które Tableau wykonuje z wykorzystaniem pamięci podręcznej, w tym prognozy, obliczenia tablicowe, sumy częściowe, sumy ogólne, trendy i linie odniesienia. Z tego filtry obliczeń tablicowych wpływają tylko na linie trendu i odniesienia.

Teraz nie musisz rozumieć każdego aspektu kolejności wykonywania działań przez Tableau. Tak naprawdę, najczęściej nawet nie będziesz o nich myślał. Jednak możliwe, że będziesz chciał wrócić do tego punktu po przeczytaniu kolejnych rozdziałów tej książki i odkryciu kilku funkcjonalności Tableau. Znajomość kolejności działań pomoże Ci uniknąć kilku błędów, a nawet w opracowaniu innowacyjnych rozwiązań trudnych wyzwań!

Podsumowanie

Ten rozdział omówił kluczowe koncepcje pracy Tableau z danymi. Choć zwykle nie będziesz przejmować się tym, jakie zapytania generuje Tableau i jakie zapytania wysyła do silników danych, zrozumienie paradygmatu Tableau bardzo pomoże Ci podczas analizy danych.

Zobaczyliśmy wiele przykładów różnych połączeń z różnymi źródłami danych, poznaliśmy wady i zalety używania wyciągów, dowiedzieliśmy się, jak zarządzać metadanymi, i omówiliśmy opcje filtrowania danych.

Praca z danymi jest podstawą wszystkich czynności wykonywanych w Tableau. Zrozumienie, jak łączyć się z różnymi źródłami danych, kiedy pracować z wyciągami i jak dostosowywać metadane, będzie niezbędne, gdy zaczniesz przeprowadzać głębszą analizę i przygotowywać bardziej złożone wizualizacje, takie jak te z rozdziału 3., „Wizualizacje — kolejny krok”.

PROGRAM PARTNERSKI

— GRUPY HELION —



1. ZAREJESTRUJ SIĘ
2. PREZENTUJ KSIĄŻKI
3. ZBIERAJ PROWIZJĘ

Zmień swoją stronę WWW w działający bankomat!

Dowiedz się więcej i dołącz już dzisiaj!

<http://program-partnerski.helion.pl>

GRUPA
Helion 

Przekonaj się, co Tableau wyczaruje z Twoich danych!

Dane stały się paliwem rozwoju cywilizacji. Wykorzystanie ich potencjału jest jednak trudne: problemem okazuje się wyciąganie z nich informacji i wniosków, aby można było na ich podstawie podejmować trafne decyzje. Jednym z narzędzi ułatwiających tę pracę jest Tableau — program do analityki danych, który umożliwia ich zrozumienie, interpretację i prowadzenie na nich efektywnych działań.

Dzięki tej książce płynnie rozpoczniesz pracę z Tableau 2022 i nauczysz się wizualizacji danych. Dowiesz się, jak je analizować i przedstawiać w formie graficznej, a także jak opowiadać oparte na nich historie. W tym wydaniu ujęto najnowsze funkcje programu, między innymi rozszerzenia pulpitu nawigacyjnego, Explain Data i integrację z CRM Analytics (Einstein Analytics), służącym do modelowania predyktywnego w Tableau. Pokazano również, jak używać tabel i obliczeń na różnych poziomach szczegółowości i stosować wizualną analitykę statystyczną. Następnie zademonstrowano techniki łączenia różnych źródeł danych z funkcjonalnościami modeli danych Tableau wraz z mapami i wizualizacjami geoprzestrzennymi. Z tego przewodnika dowiesz się też, jak korzystać z funkcji Tableau Prep Builder, by skutecznie oczyścić i zorganizować dane.

Dzięki książce nauczysz się:

- tworzyć zachwycające wizualizacje złożonych danych
- budować interaktywne pulpity nawigacyjne
- korzystać z obliczeń w celu rozwiązywania problemów i uzupełniania analiz
- wzbogacać możliwości Tableau poprzez rozszerzenia, skrypty i CRM Analytics
- oczyszczać i organizować dane w Tableau
- tworzyć angażujące historie oparte na danych

Joshua N. Milligan — wielokrotny laureat tytułu Tableau Visionary (dawniej Zen Master), światowy finalista zawodów Iron Viz. Od 2004 roku pracował jako konsultant w firmie Teknion Data Solutions, a po przejęciu jej w 2022 roku przez Resultant działał z klientami z różnych branż. Mieszka w Tulsie z żoną i czworgiem dzieci.

Helion 	KOD KORZYŚCI Sięgnij po więcej! 	
 helion.pl	ISBN 978-83-8322-372-8	
 HELION SA ul. Kościuszki 1c 44-100 Gliwice tel.: 32 230 98 63 helion@helion.pl	 9 788383 223728	
Cena: 139,00 zł		

<packt