Building Your Tableau Toolkit

Josh Mills, Data Science Leader (Sr. Consultant)

Enterprise Analytics Office

2021

Nationwide®
is on your side

Agenda

- Extended hands-on examples using a data set with movie ratings, exploring features in Tableau such as:
  - Adding Interactivity with Filters
  - Working with Strings
  - Table Calculations
  - Working with Dates and Timestamps

- The data source “Movie Ratings” should have been sent in advance of this session (as the file **MOVIE RATINGS – LATEST (SMALL).hyper**). Create a new workbook using this data source before beginning the exercises.

Screenshots from Tableau Server 2018.1.5 are shown throughout this presentation. Nationwide IT regularly updates Tableau at least a couple times a year, so the screenshots shown may not reflect the most recent interface updates.

MovieLens Movie Ratings

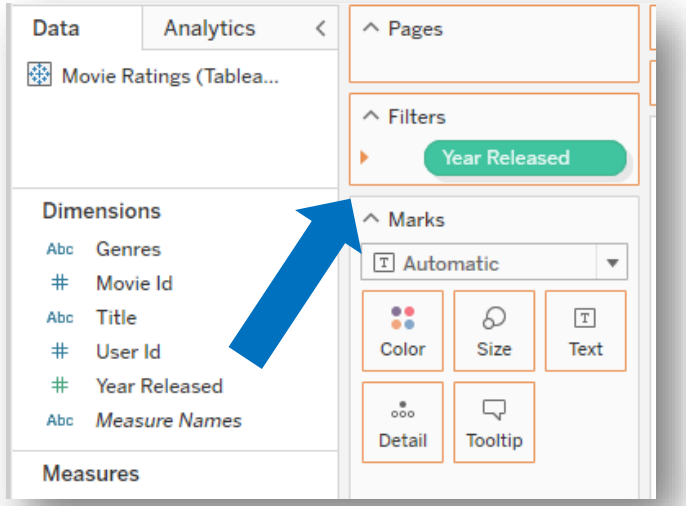
- Overview
  - 100,000 movie ratings from MovieLens, part of the GroupLens Research Project at the University of Minnesota. A subset of the full 27 million data set is used for this session.
  - **Source:** https://grouplens.org/datasets/movielens/latest/
  - **MovieLens web site:** https://movielens.org/

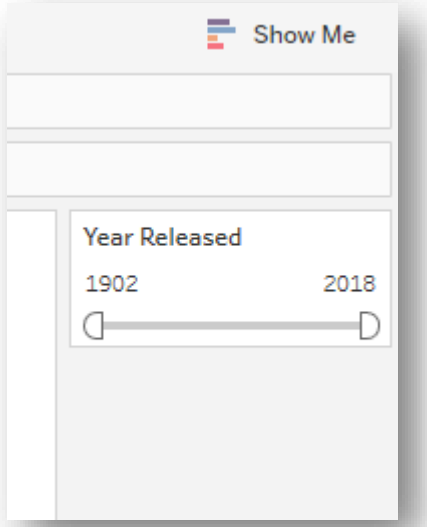
- Description of Key Columns
  - **Title:** Name of movie
  - **Rating:** 0-5 star ratings. Half-star ratings (e.g. 4.5) are possible.
  - **User ID:** Surrogate ID of user
  - **Movie ID:** Surrogate ID of movie
  - **Year Released:** Year movie was released
  - **Genre Indicators:** 0/1 indicator for whether a movie fell into a particular genre. A movie can be classified under more than one genre (**Genres** is a concatenated string upon which these indicators were built.)
  - **Timestamp:** Seconds since January 1, 1970 (UTC) representing when the user entered the rating for a given movie

Adding Interactivity with Filters

Interactive filters can be created by dragging a field to the "Filter" shelf. An interactive filter will automatically appear on the right

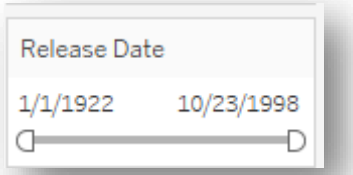
For this first example, drag "Year Released" into the **Filter** shelf. Right-click on the field and choose "Show Filter." An interactive filter will appear on the right-hand side



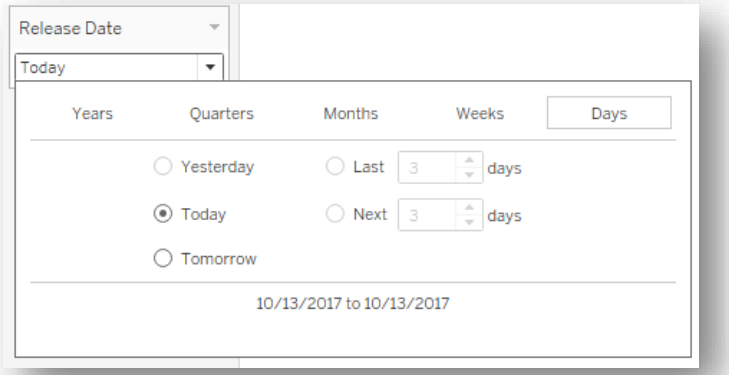


Many different types of filters are available, depending on the data type you are trying to filter

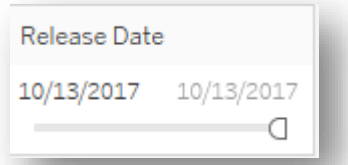
Date Variable Filters - Examples



Date Range

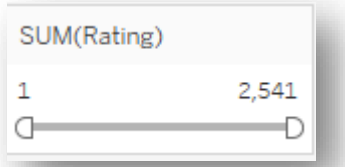


Relative Date

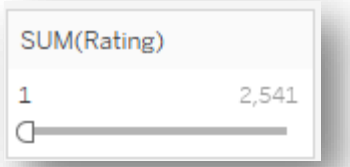


Start/End Date

Continuous Variable Filters - Examples



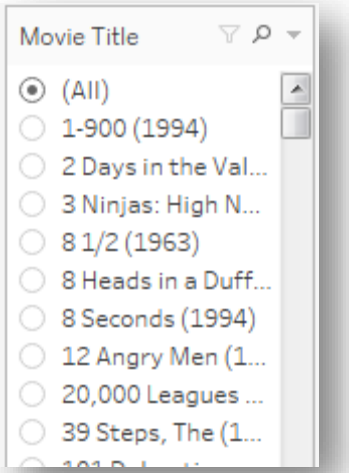
Range of Values



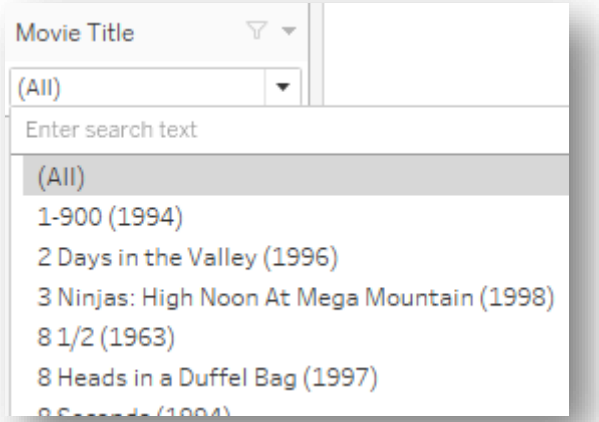
At Least/At Most

Many different types of filters are available, depending on the data type you are trying to filter

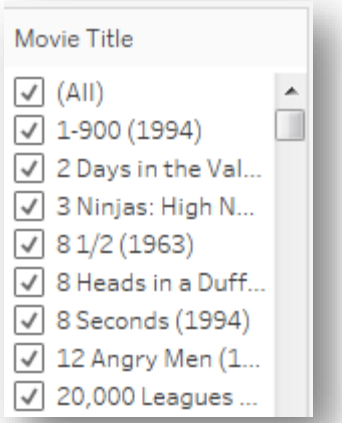
Discrete Variable Filters - Examples



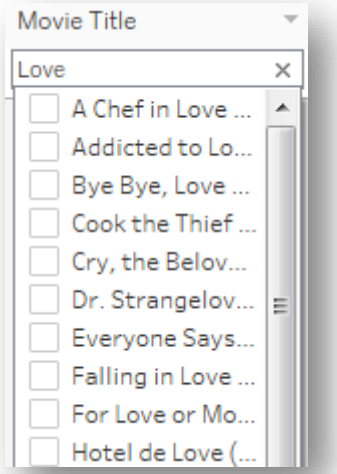
Single Value List



Single Value Dropdown

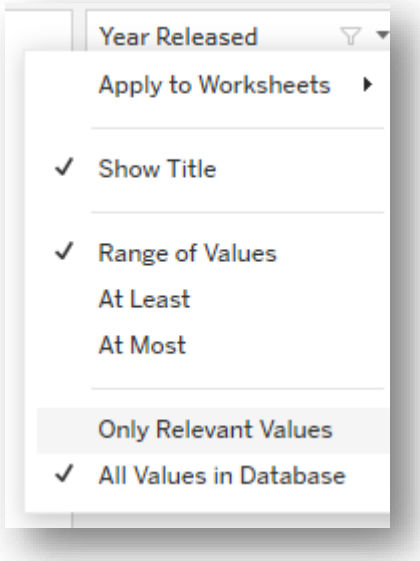


Multiple Value List/Dropdown



Custom List (Searchable)

Each filter type has different options that need to be carefully chosen



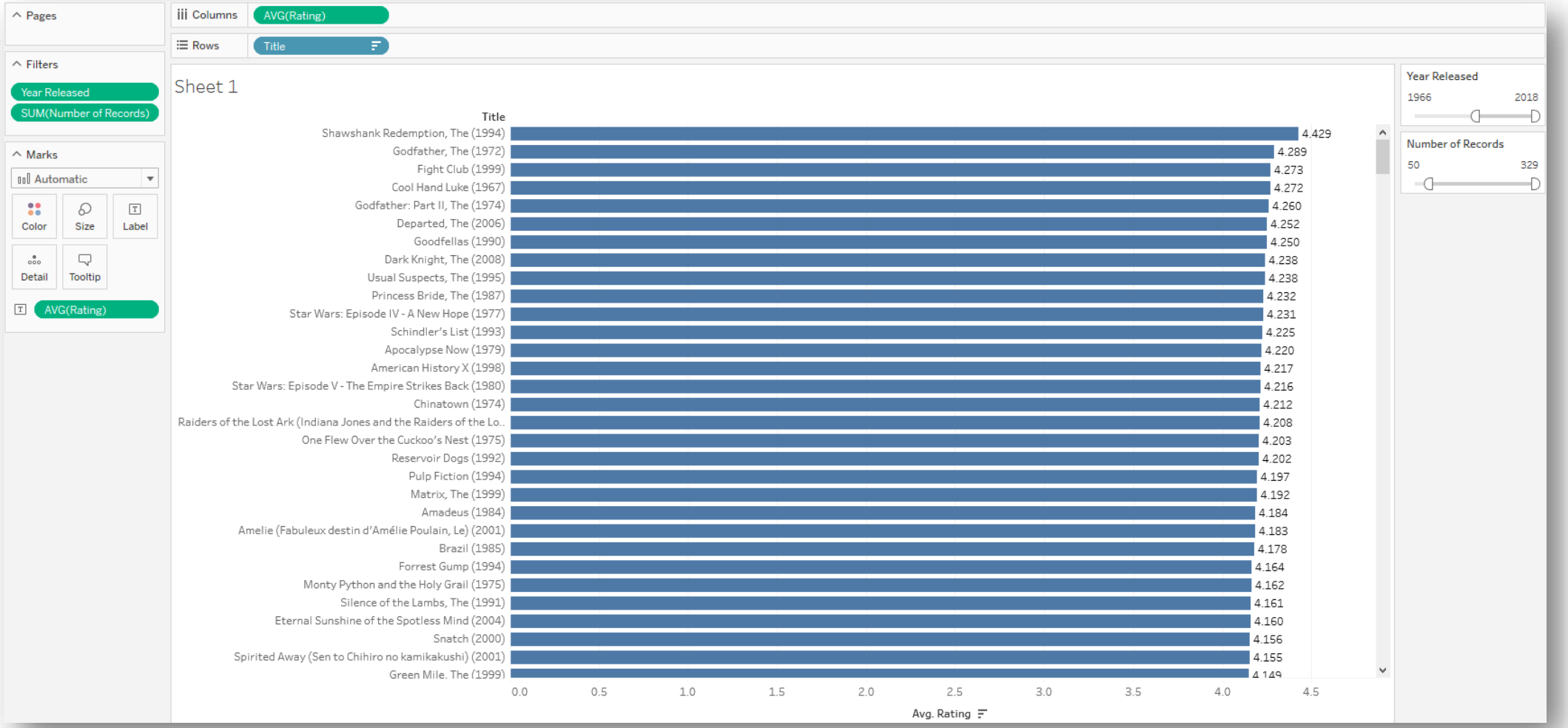
For example, you may want to show only relevant values in a filter rather than all possible values, particularly if the field you're using to filter has a high number of distinct values (> 500 typically)

HANDS-ON EXERCISE

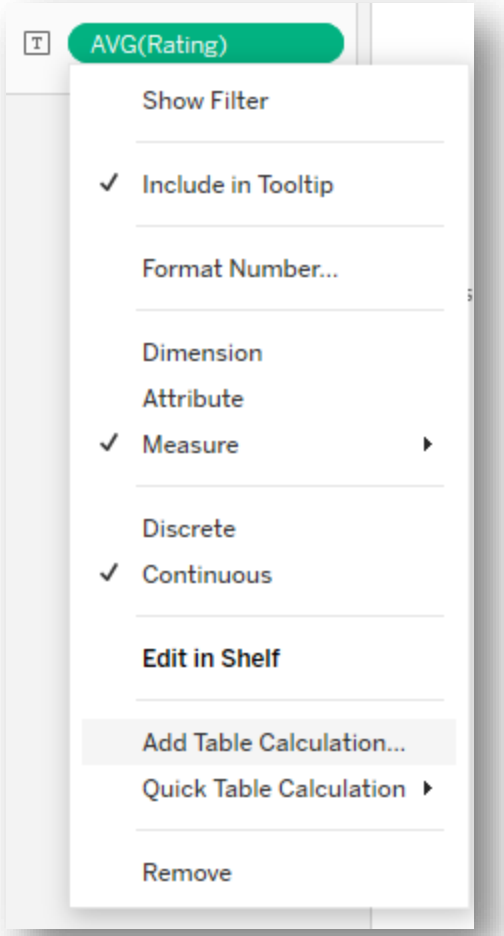
Create a new view (named **Movie Rankings**) that shows average movie rating by movie. Add filters for year released, and filter the view to show only movies released between 1966 and 2018. Then, add a filter to show only movies with at least 50 ratings. Sort the view by average movie rating, descending. Add labels to show average movie ratings.

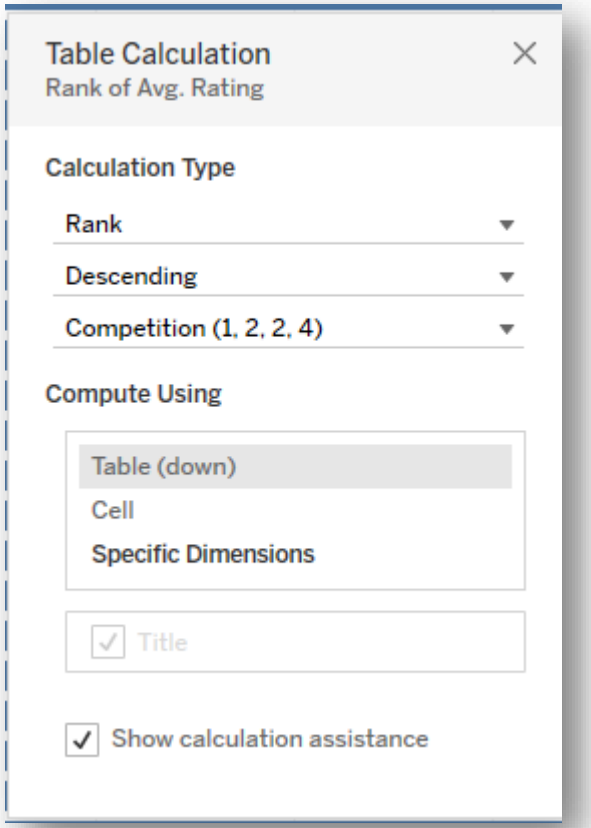
After this exercise, we'll add a way to show only the top 10 highest-rated movies.

Solution

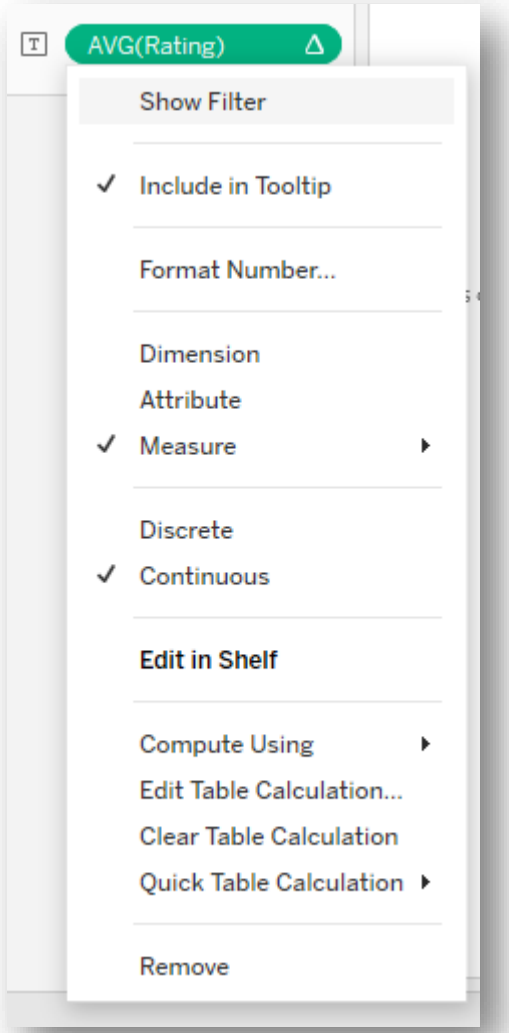


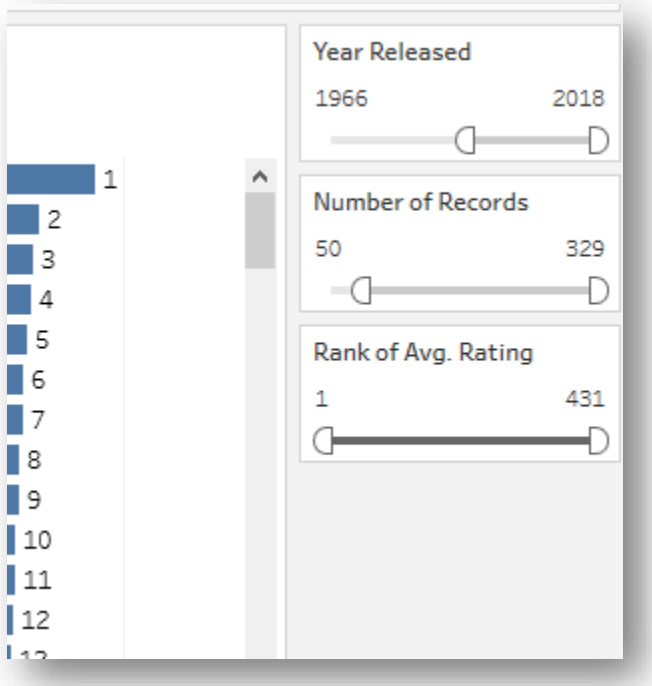
To show the top 10 highest rated movies, click on the drop-down menu for Average Rating and choose "Add Table Calculation", using the below settings





Click on the drop-down menu for Average Rating rank and choose “Show Filter”





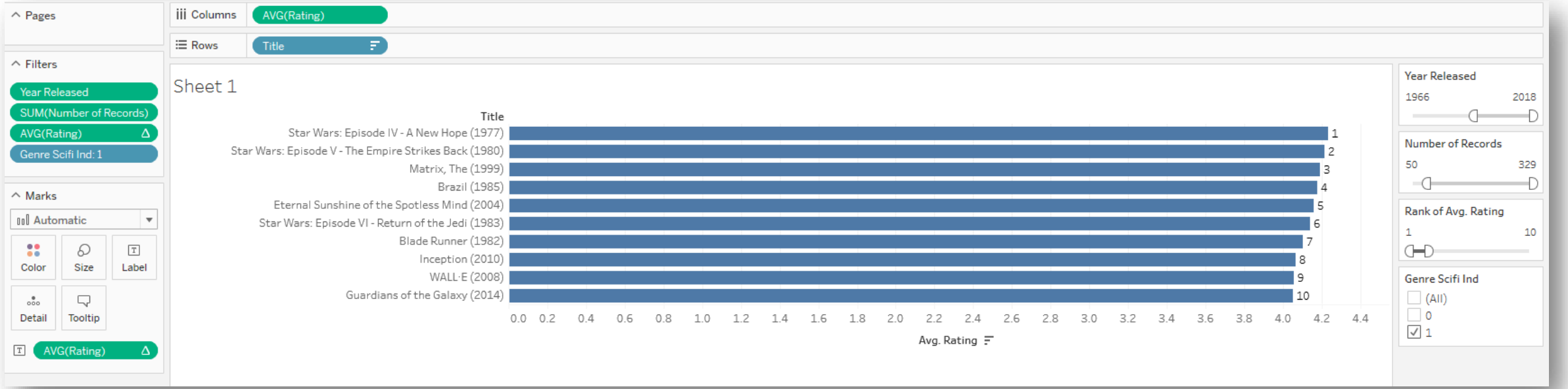
This filter can be modified to show only the top 10 highest rated movies

HANDS-ON EXERCISE

Modify the view to show only the top 10 highest-rated sci-fi movies.

HINT: The Genre indicators can be converted to dimensions.

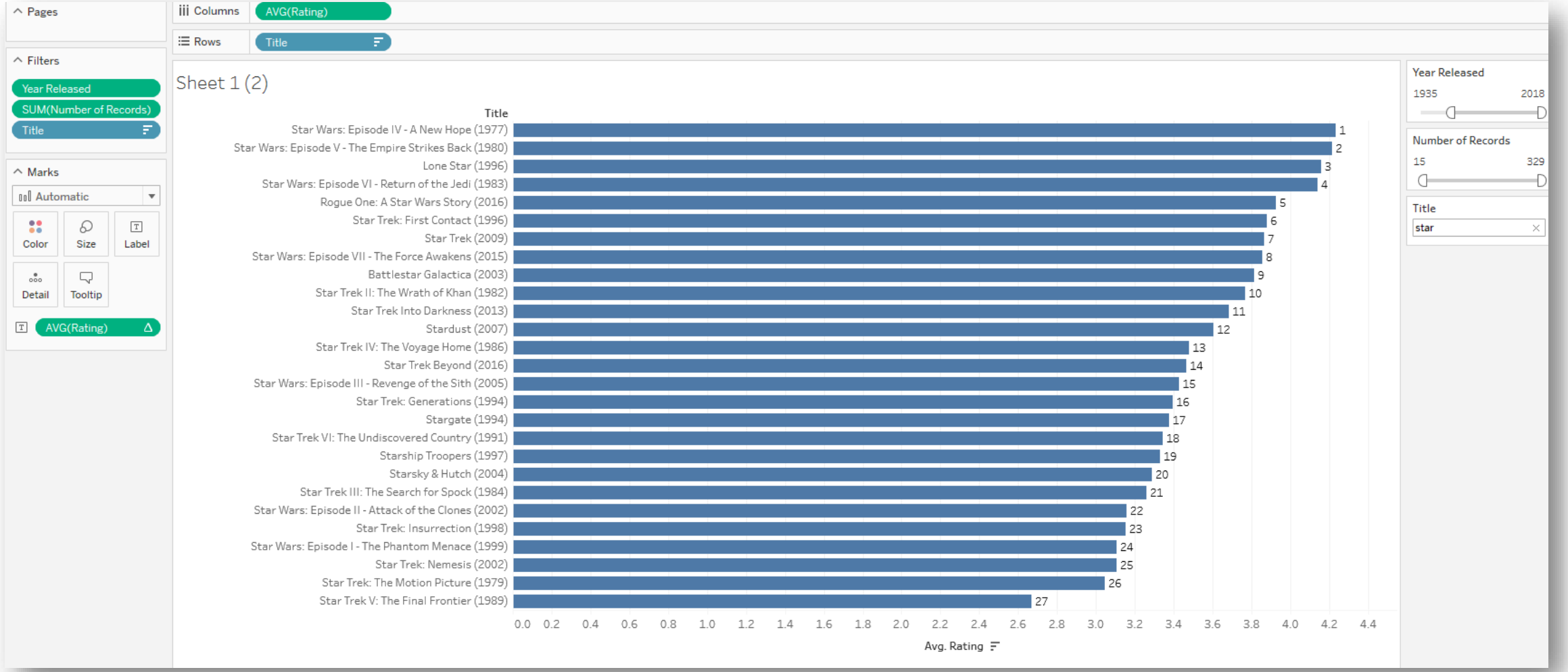
Solution



HANDS-ON EXERCISE

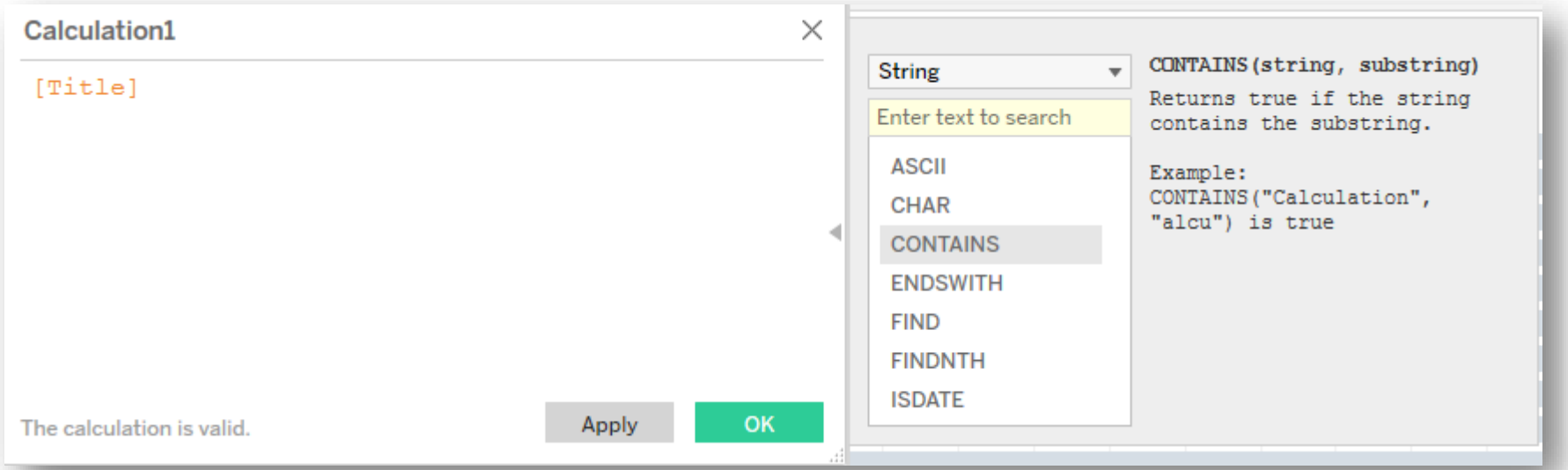
Create a copy of the view and add a filter for movie title. Remove the top X filter and the sci-fi filter. Reduce the minimum number of ratings to 15. Change the filter type to a wildcard match. Search for movies with the word “star” in the title.

Solution

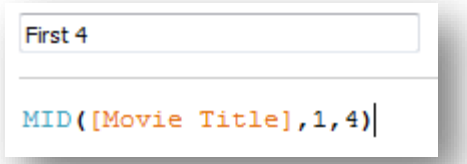


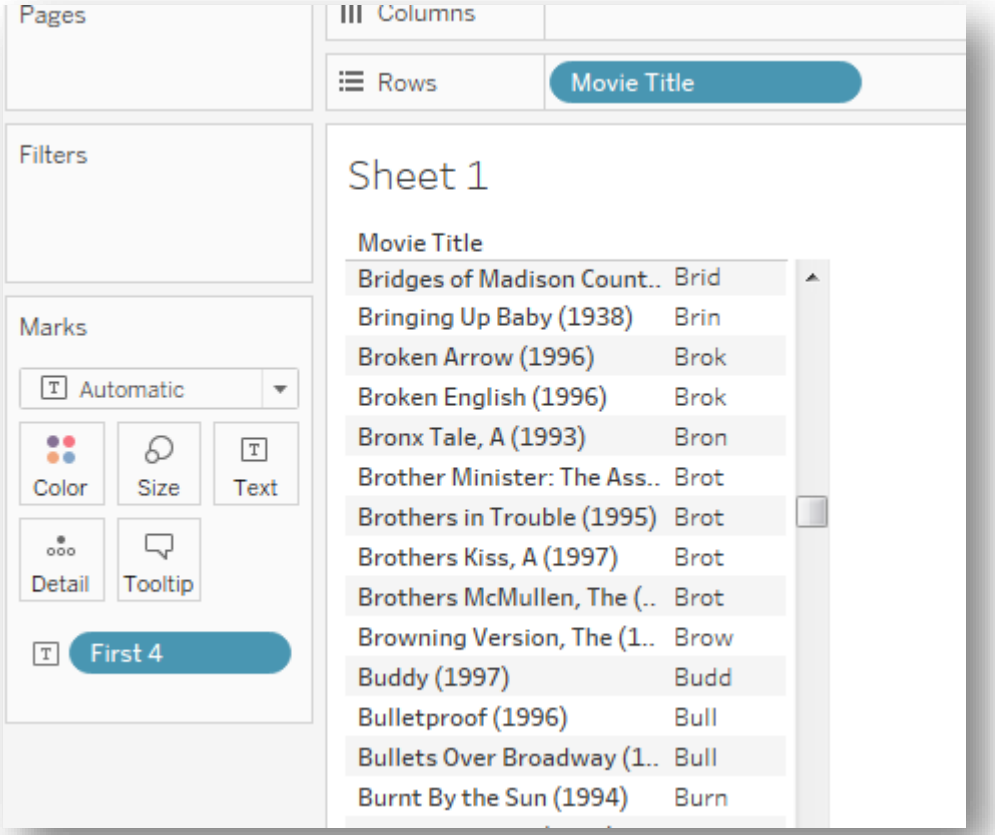
Working with Strings

Tableau contains a wide variety of string functions



The MID function is the equivalent of substring functions in SQL and other programming languages



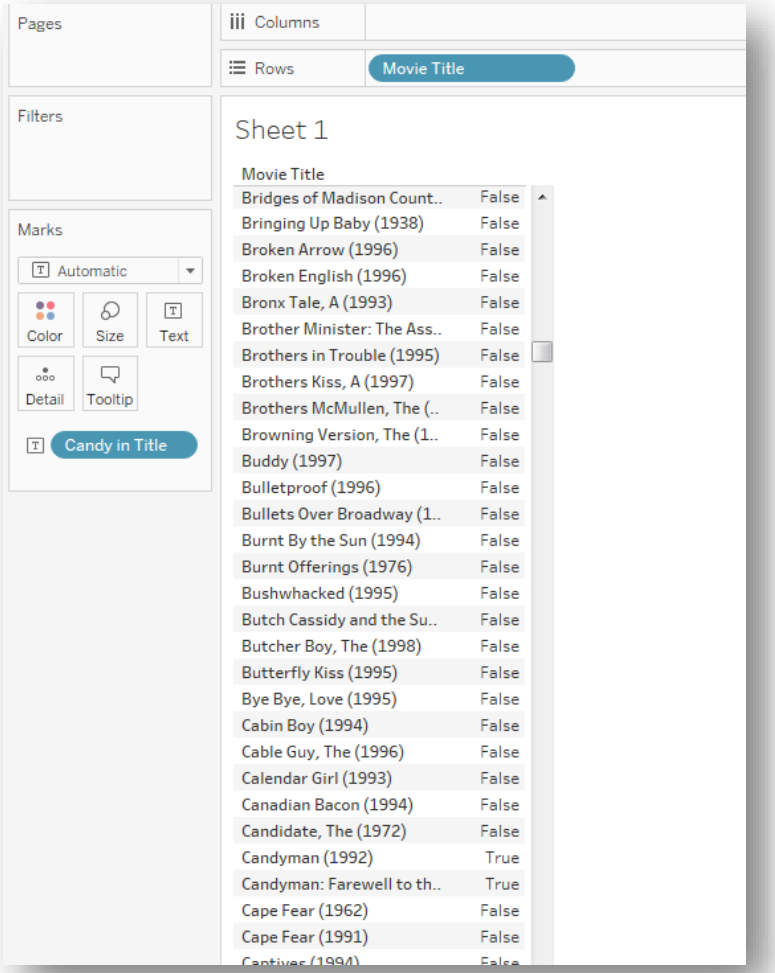


CONTAINS can be used to search for specific substrings within a string

Candy in Title

```
CONTAINS([Movie Title],"Candy")
```

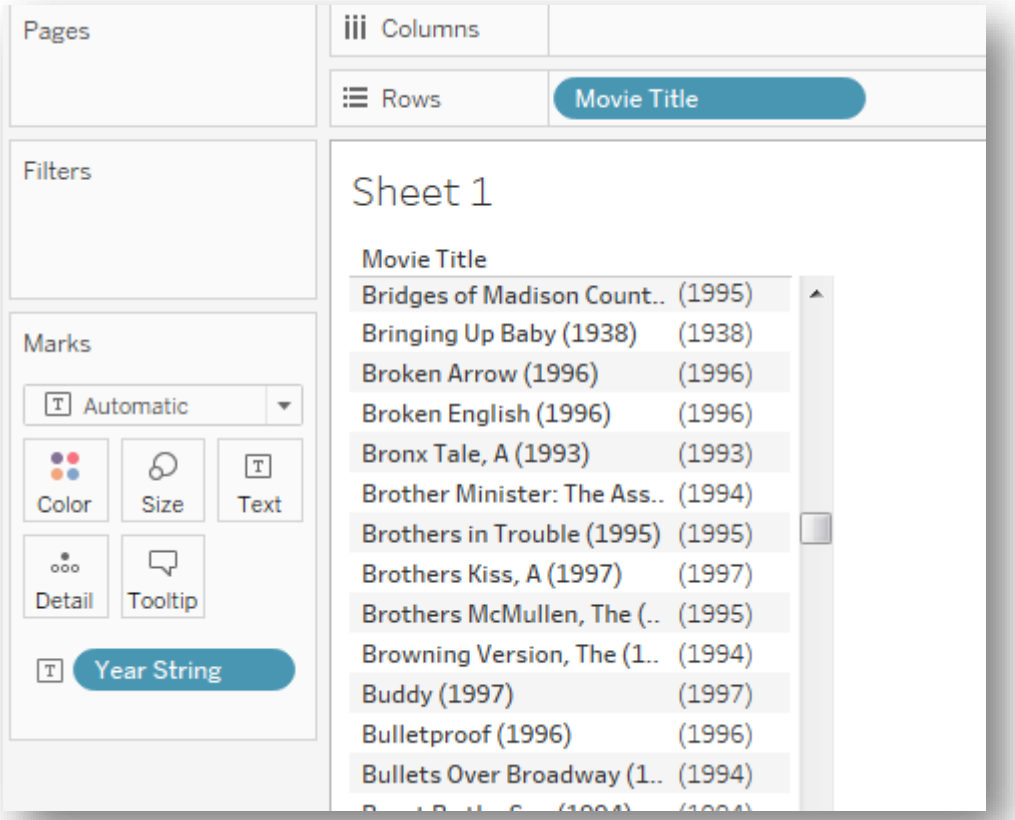
NOTE: Strings are case-sensitive in Tableau!



The RIGHT function can be used to extract the last series of characters from a string

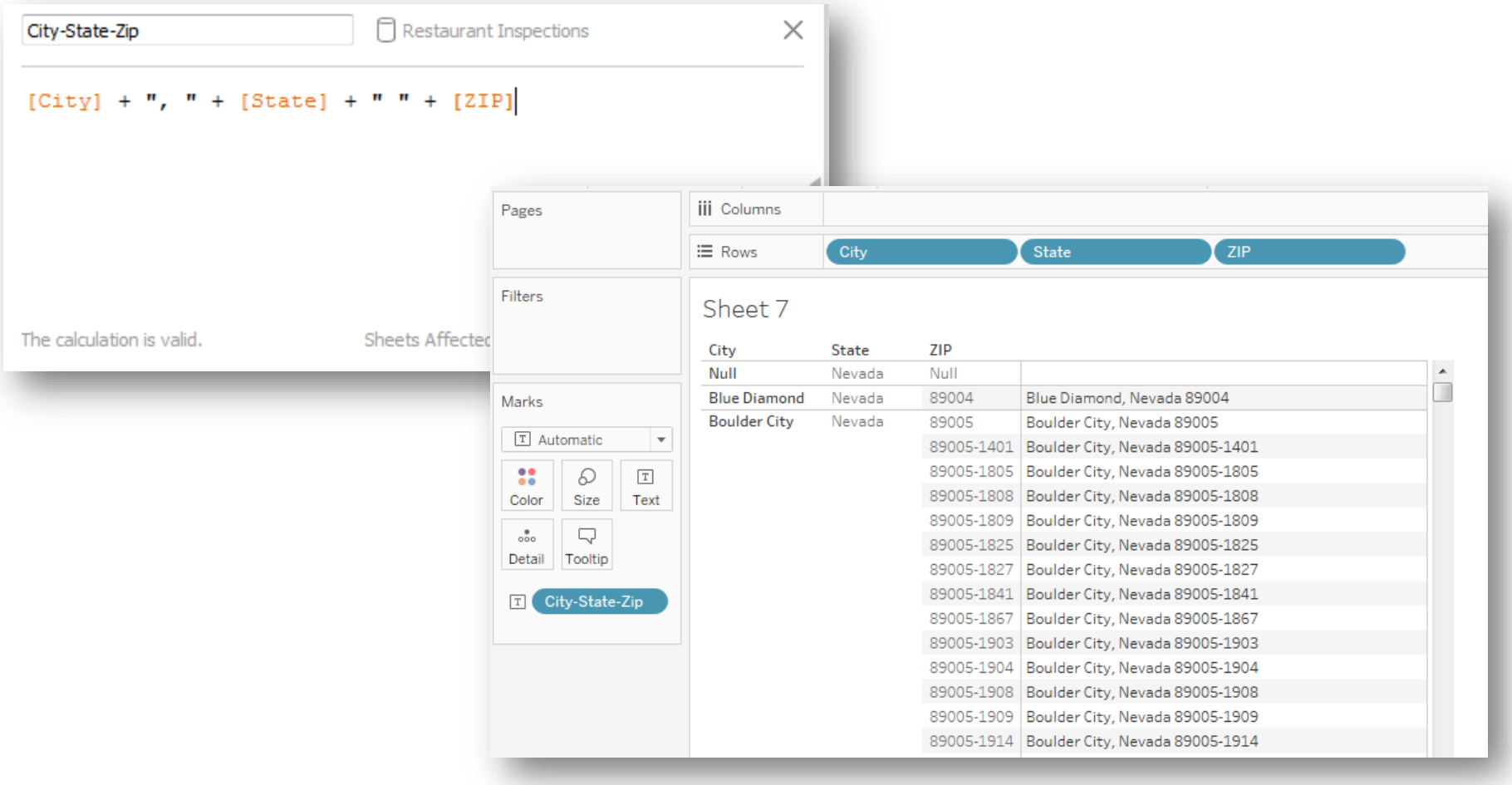
Year String

```
RIGHT([Movie Title],6)
```



Concatenating Strings

- Strings can be combined simply by using the + operator



HANDS-ON EXERCISE

Suppose the release year was not available in the data set. Use a combination of the previously described string functions to extract the release year from the movie title, creating a new variable named **Year (Derived).**

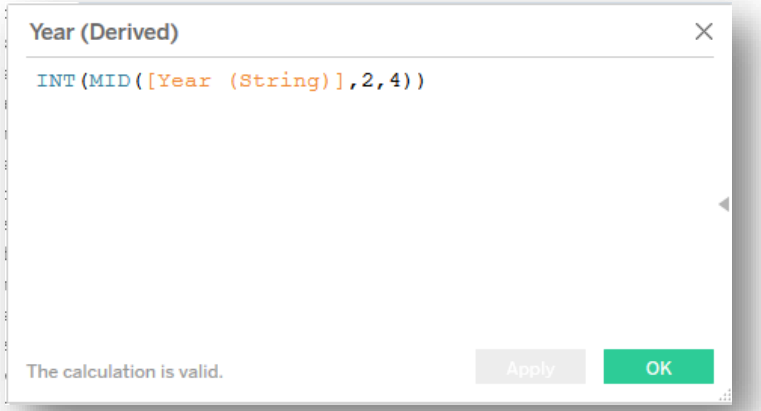
HINTS: Use the INT function to convert strings to integers. Create a new view to check your work.

BONUS: Create a new calculated field that only contains the movie title without the year, named **Title Only**. Then, recreate the original movie title field using Tableau's concatenation functions, named **Recreated Title**. You will need to use a combination of the LEFT function (which performs a similar function as RIGHT, except it uses the start of the string) and the LEN function (which returns the length of a string)

Solution – Year (Derived)

Create a new calculated field named **Year (String)**, using the RIGHT and RTRIM functions

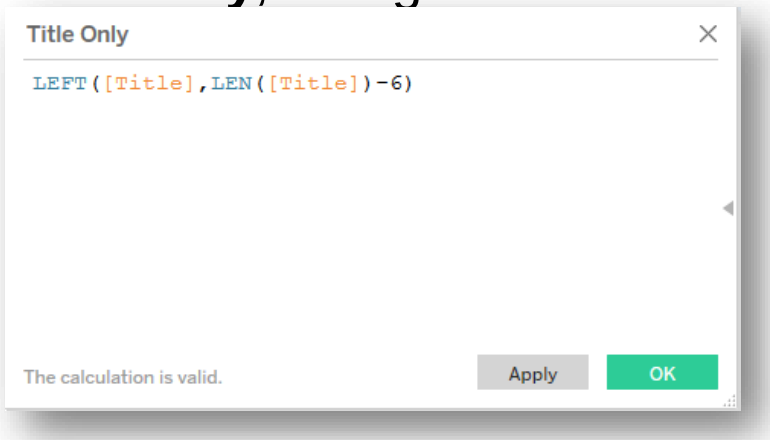
Then create a new calculated field named **Year (Derived)** using the MID and INT functions



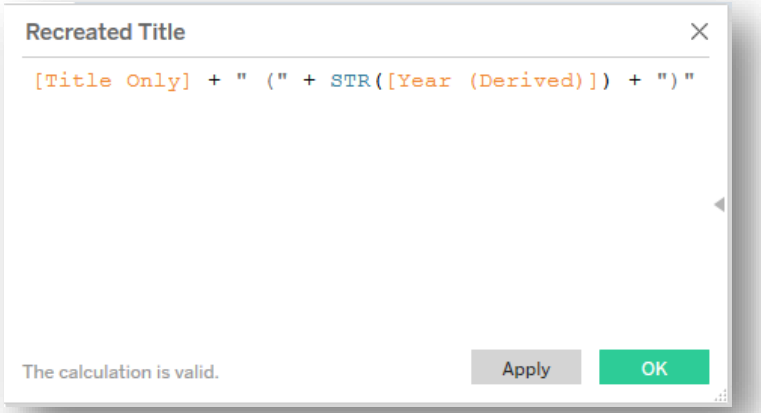
**Thanks to Kunyao (Richard) Xu and Tatum McPhillips who identified issues with the original solution.*

Solution – Recreated Title

Create a new calculated field named **Title Only**, using the LEFT and LEN functions



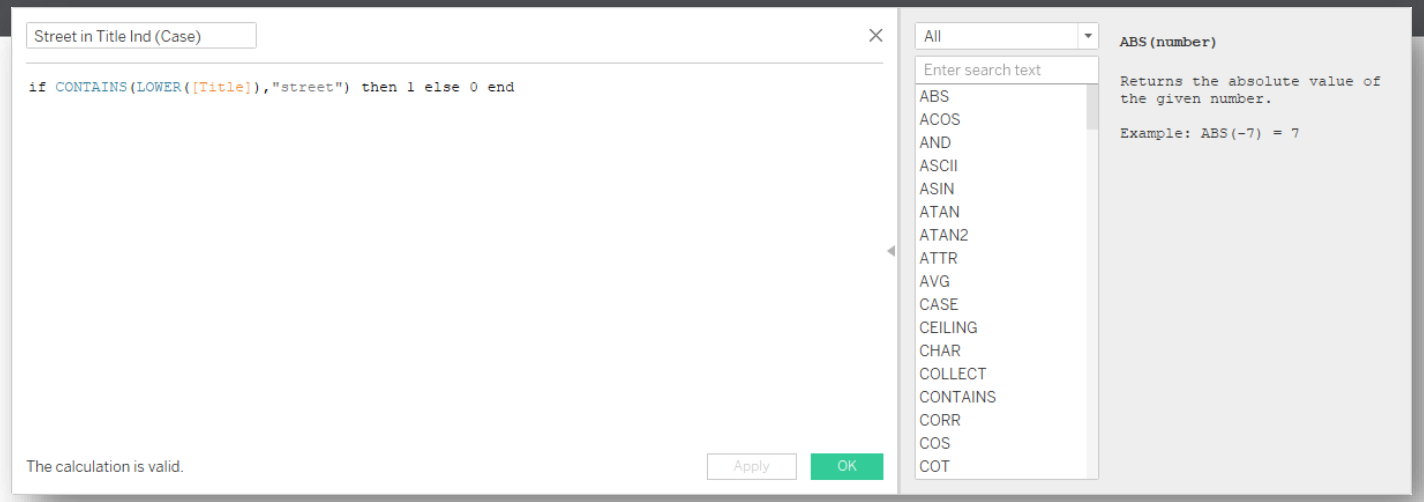
Then create a new calculated field named **Recreated Title** using Tableau's concatenation functions

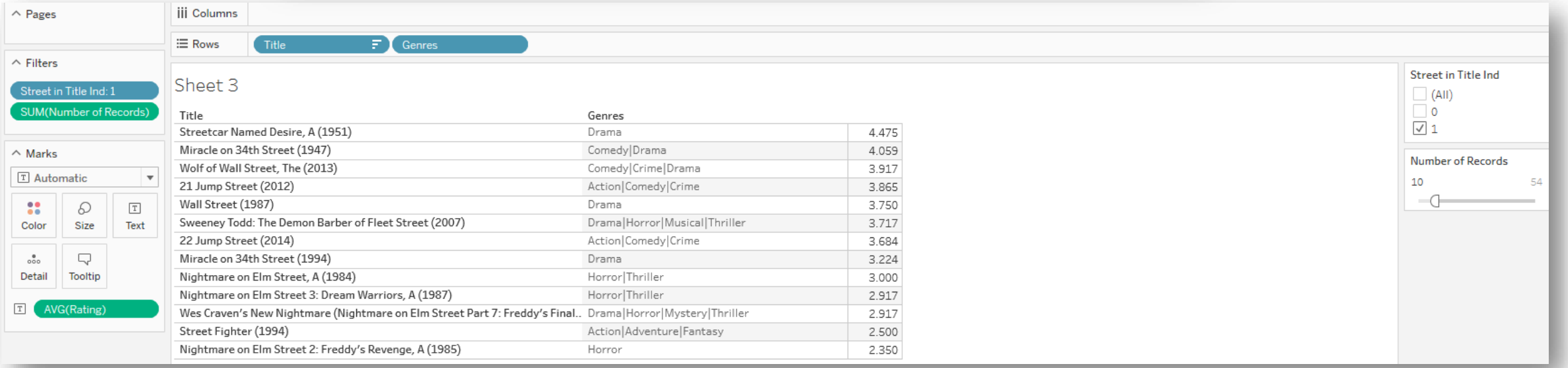


HANDS-ON EXERCISE

Create a new view with Title in the Rows shelf. Create a new calculated field named **Street in Title** that is equal to 1 if the movie title contains the word “street” and 0 otherwise. Filter the view based on this indicator. Sort the view descending by average movie rating. Add a filter to show only movies with at least 10 ratings. Name this view **Street in Title**.

Solution



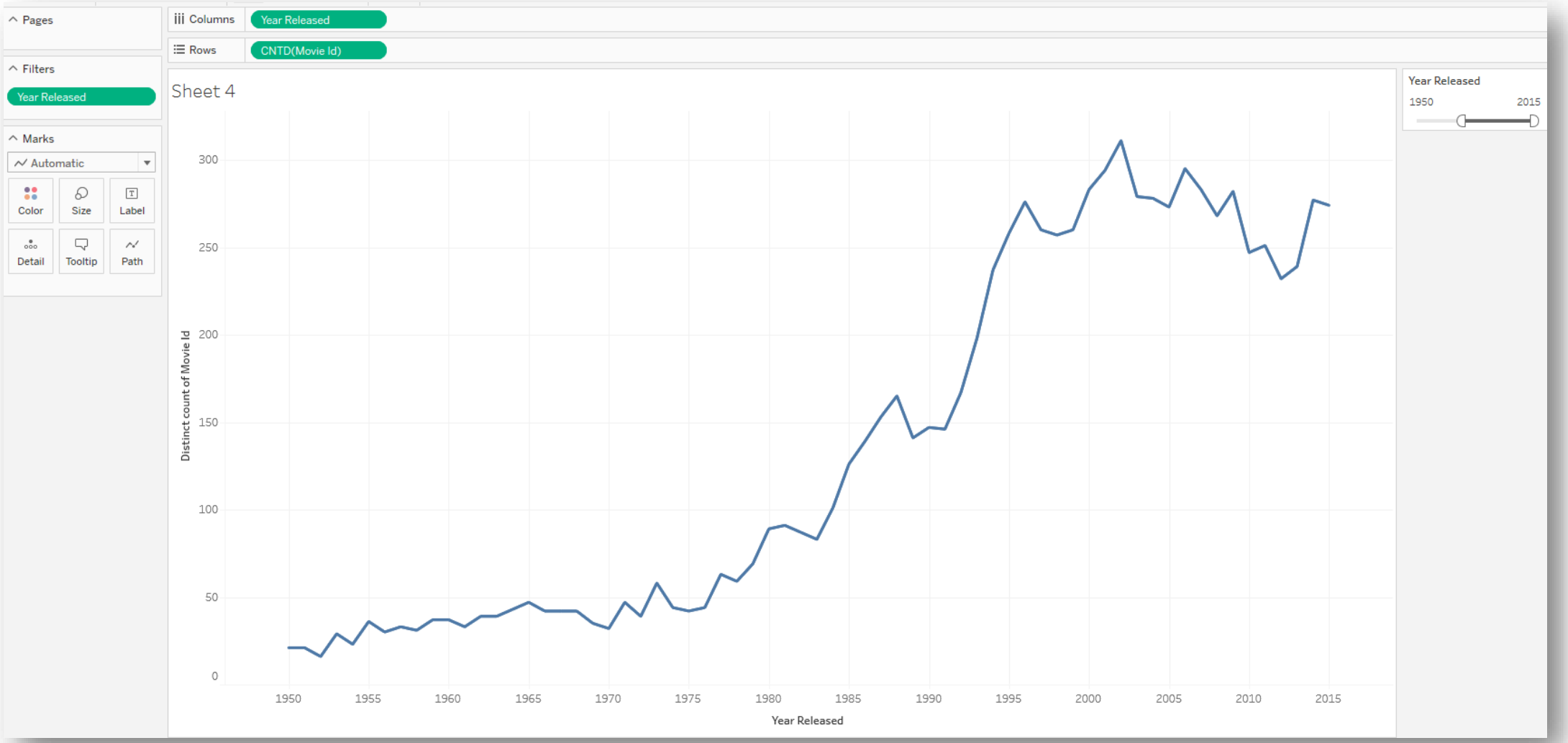


Title	Genres	
Streetcar Named Desire, A (1951)	Drama	4.475
Miracle on 34th Street (1947)	Comedy|Drama	4.059
Wolf of Wall Street, The (2013)	Comedy|Crime|Drama	3.917
21 Jump Street (2012)	Action|Comedy|Crime	3.865
Wall Street (1987)	Drama	3.750
Sweeney Todd: The Demon Barber of Fleet Street (2007)	Drama|Horror|Musical|Thriller	3.717
22 Jump Street (2014)	Action|Comedy|Crime	3.684
Miracle on 34th Street (1994)	Drama	3.224
Nightmare on Elm Street, A (1984)	Horror|Thriller	3.000
Nightmare on Elm Street 3: Dream Warriors, A (1987)	Horror|Thriller	2.917
Wes Craven's New Nightmare (Nightmare on Elm Street Part 7: Freddy's Final..	Drama|Horror|Mystery|Thriller	2.917
Street Fighter (1994)	Action|Adventure|Fantasy	2.500
Nightmare on Elm Street 2: Freddy's Revenge, A (1985)	Horror	2.350

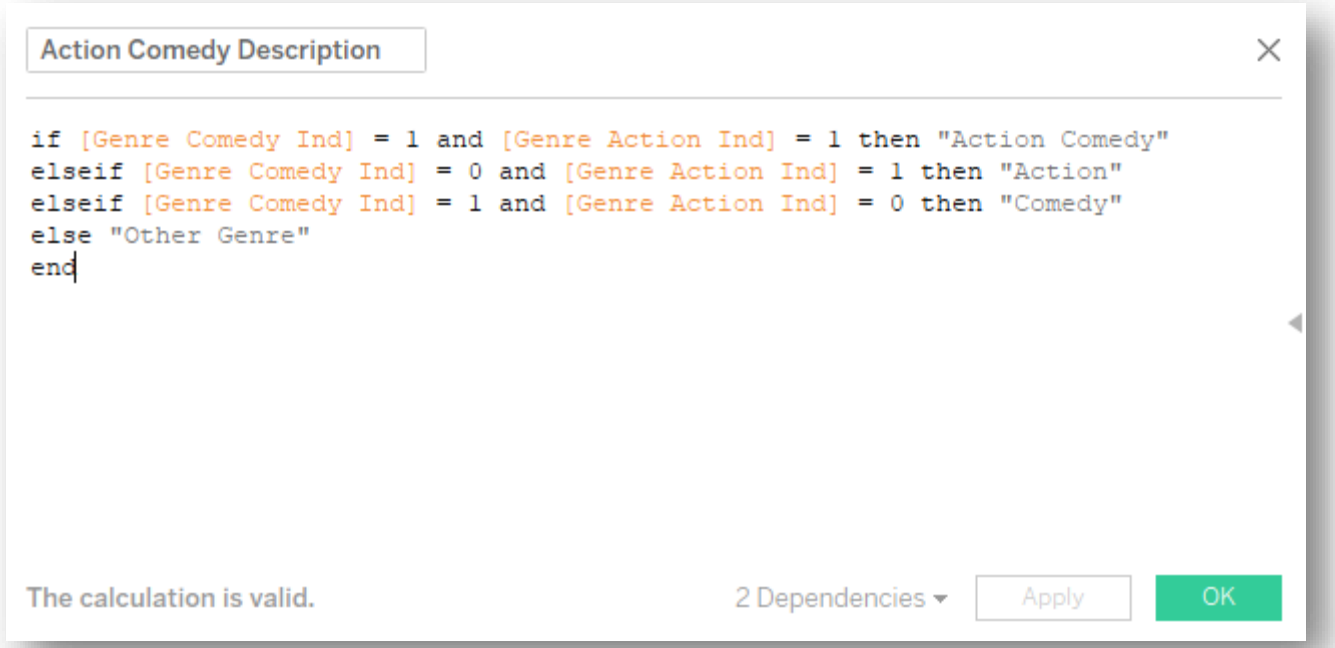
HANDS-ON EXERCISE

Create a new view (named **Count Over Time**) that shows number of movies released each year. Add a filter for year released. Limit the view to show only movies released between 1950 and 2015. Create a new calculated field named **Action Comedy Description** that uses the Genre indicators for Action and Comedy to identify movies that are: Action only, Comedy only, Action-Comedies, or other. Modify the view to add a color legend using **Action Comedy Description**.

Solution – Initial View with Filter



Solution – Action Comedy Description Derivation



Solution – Final View with Color Legend

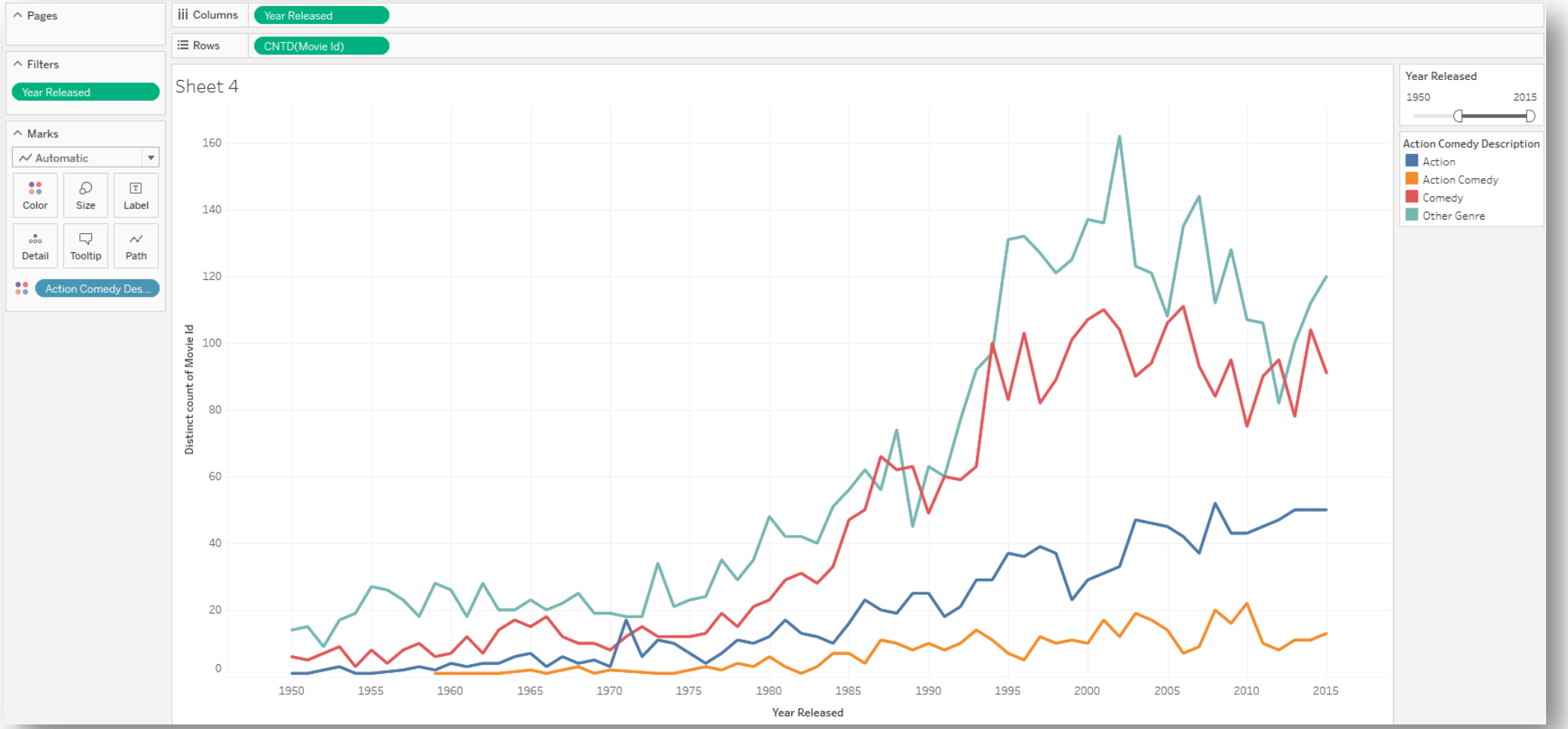
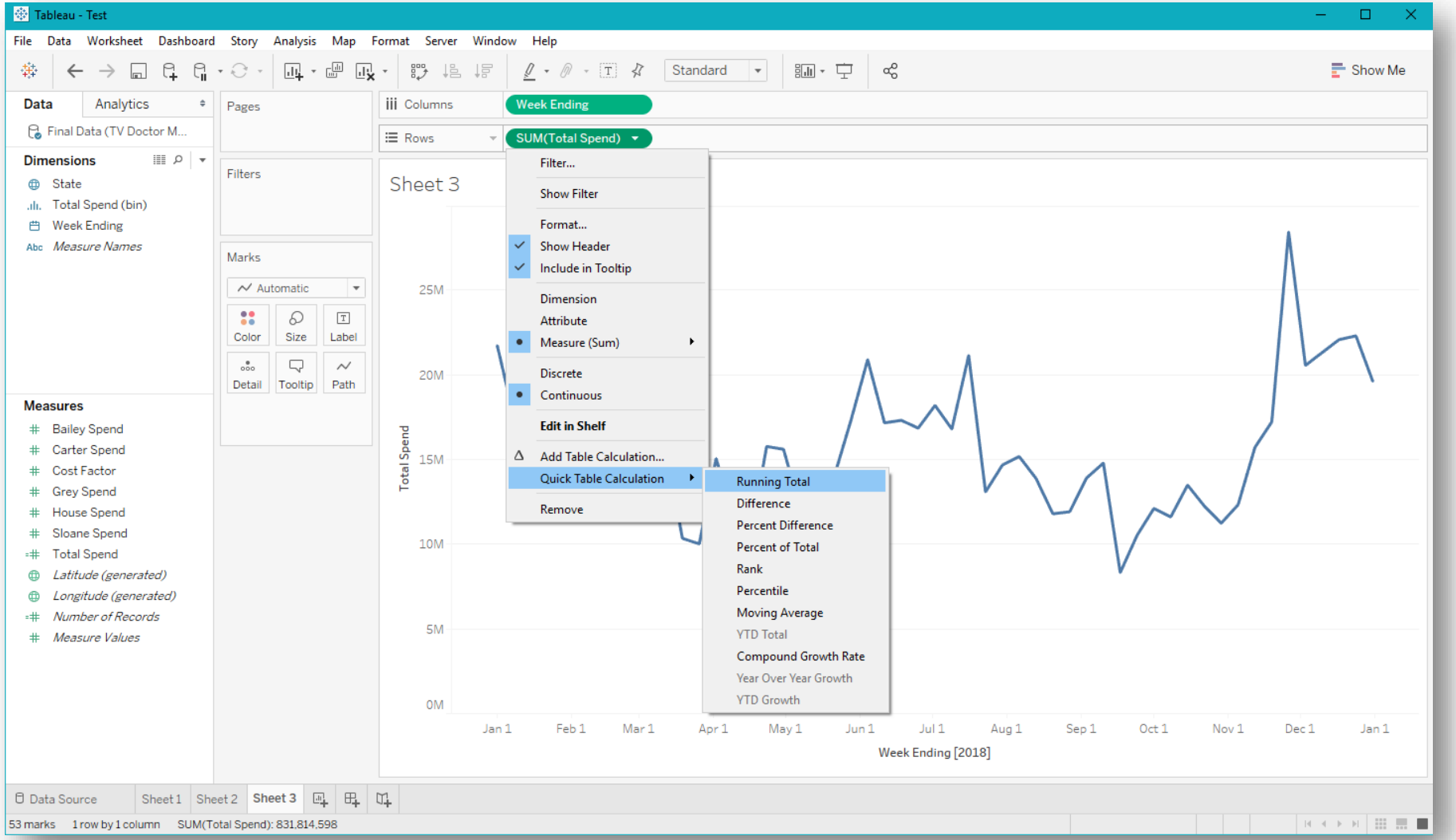


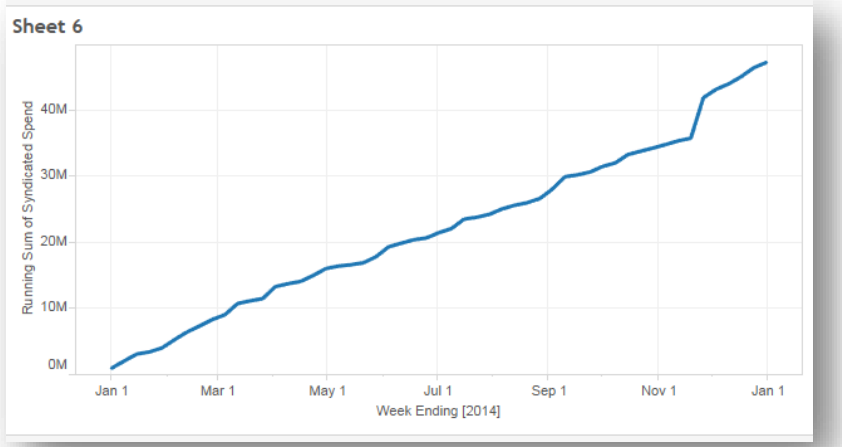
Table Calculations

Table calculations can be used to transform existing measures based on the dimensions in your views

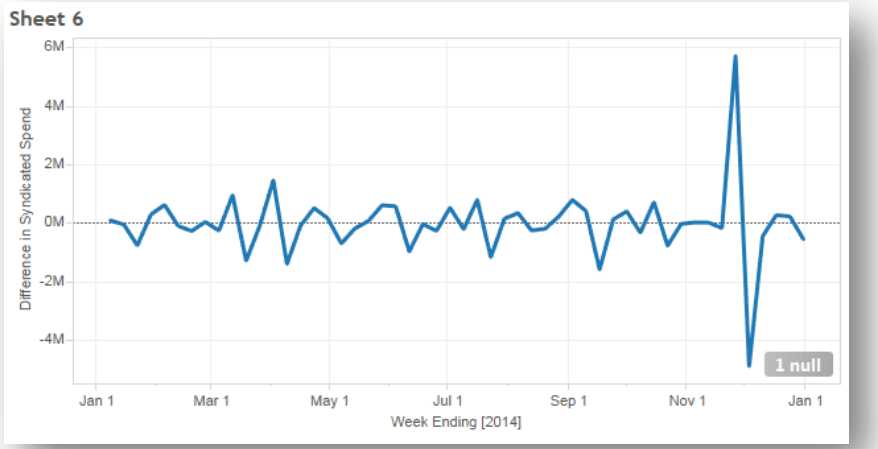


Examples include cumulative sums, % of total, absolute/percent differences, moving average, and rank

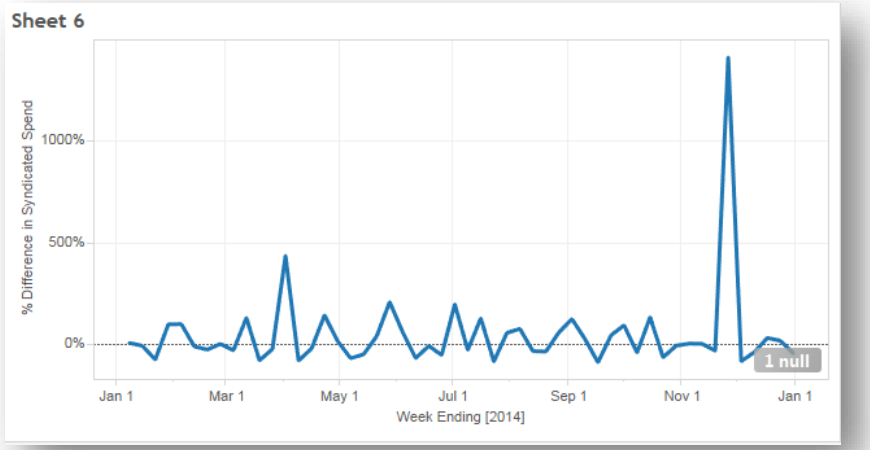
Running Total/Cumulative Sum



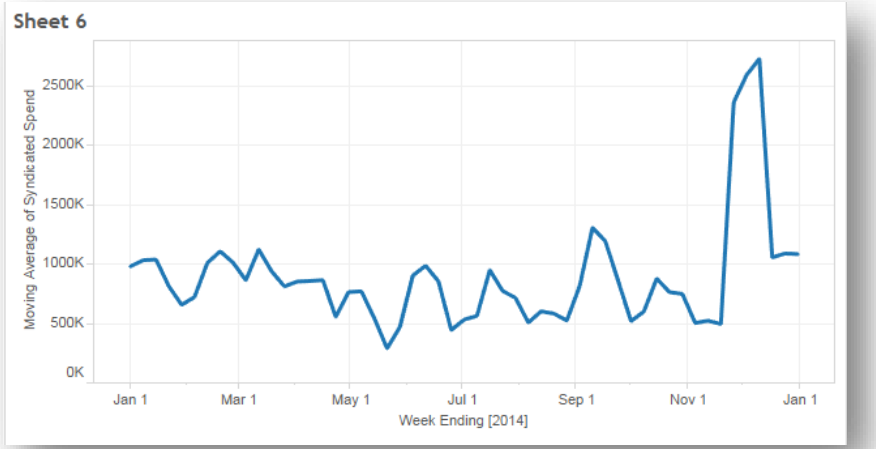
Absolute Difference



% Difference

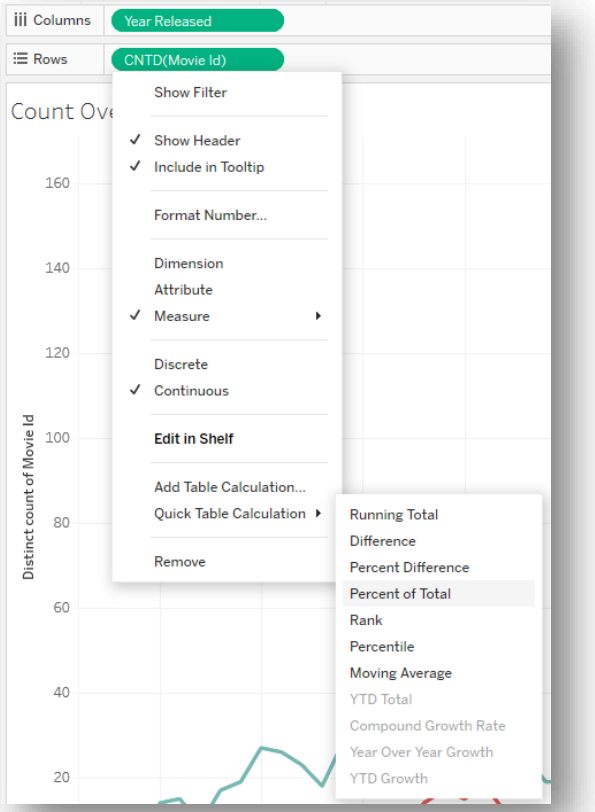


Moving Average



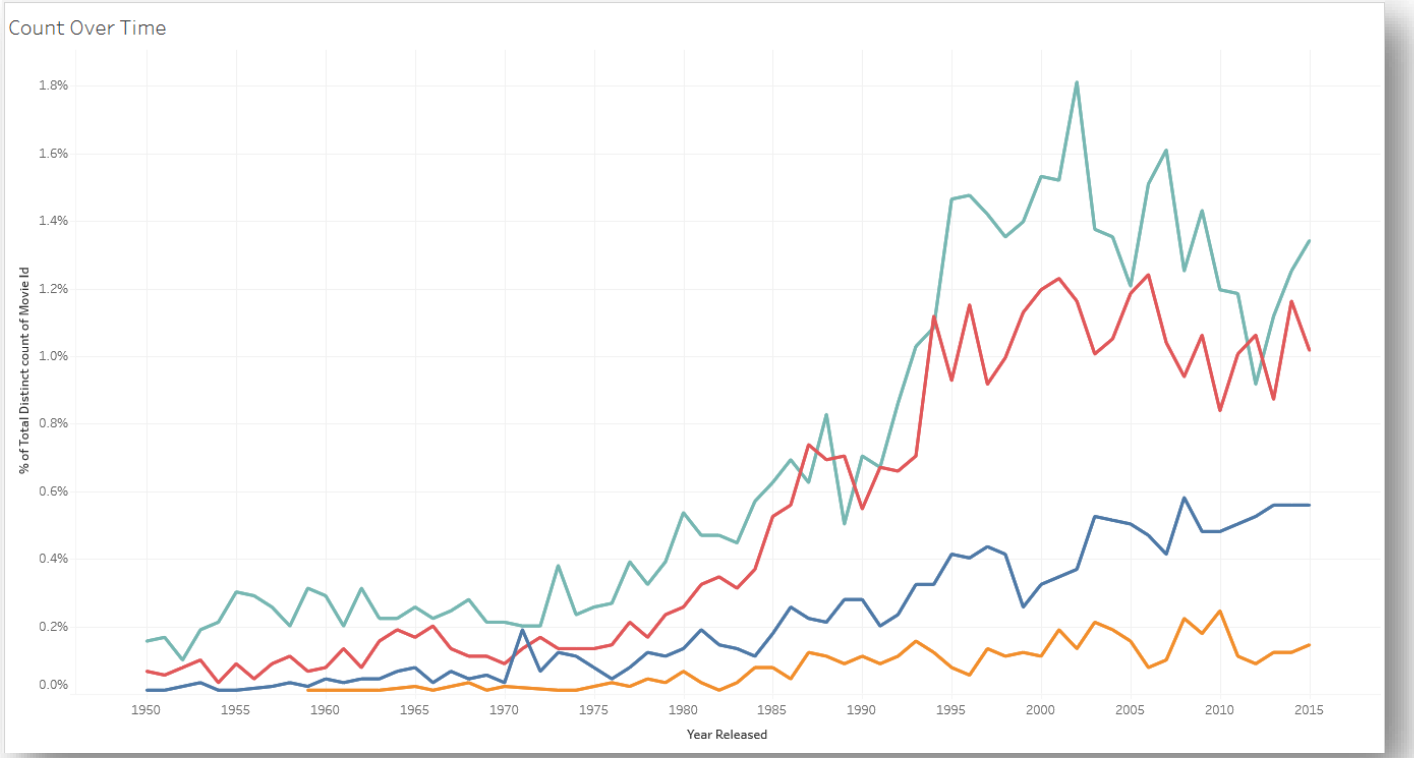
Example – Movie Count by Genre and By Year

- **Suppose we want to see the percentage of movies that are action-comedies**, by year. To start, click on the drop-down menu for CNTD(Movie Id) and choose Quick Table Calculation → Percent of Total



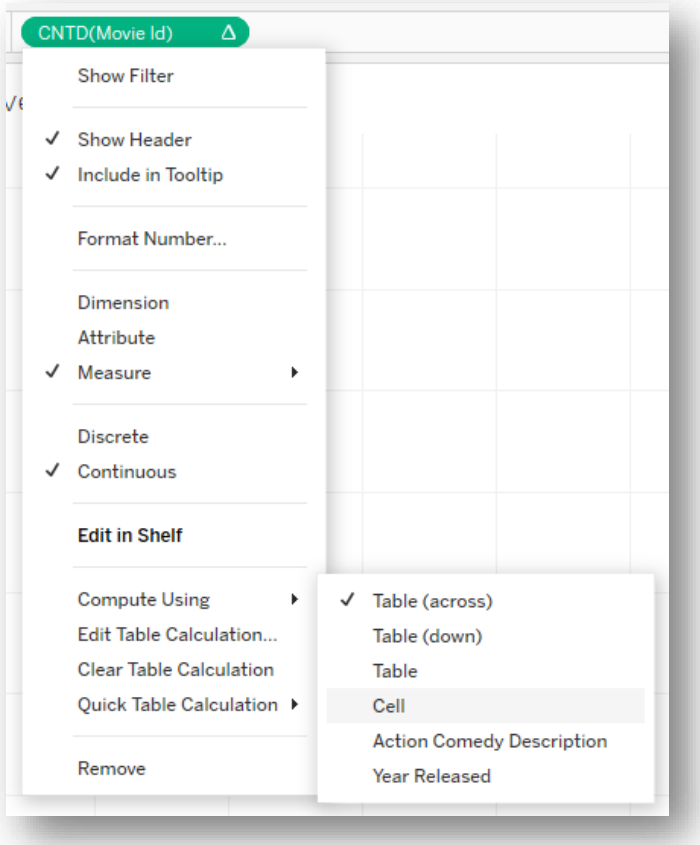
The resulting view is a bit misleading, and the view will need modifications to answer the question being posed

- The current view shows the cumulative percentage of movies released for each particular category, by year. Adding the percentage across all years and across all series would sum to 100%. This does not address the question of what percentage of movies released each year that are action-comedies.



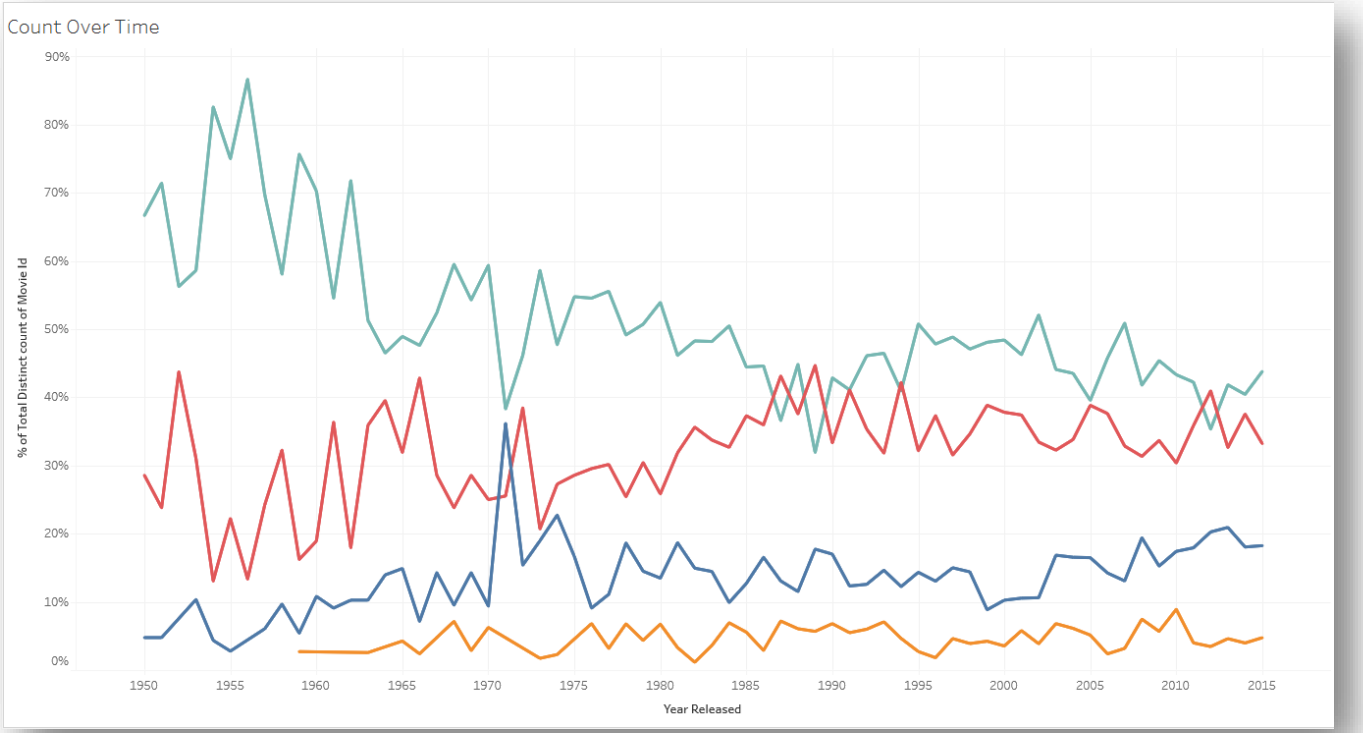
Modify the view by choosing the drop-down menu for CNTD(Movie Id) and choosing Compute Using → Cell.

- This will calculate the percentage of movies released for each genre for each year



The resulting view provides a clearer view into what percentage of all movies released are action-comedies

- From here, a few observations can be made:
  - Action comedies were not released until around the mid-1960s
  - Action comedies made up a larger share of movies in the mid-2000s before a sustained decrease occurred after 2010



HANDS-ON EXERCISE

Experiment with different ways of representing the percentage of movies released by type and by genre. Modify the **Count Over Time** view include the original view showing count of movies released by genre.

HINT: The Show Me function can be useful for experimenting with different views.

Possible Solution

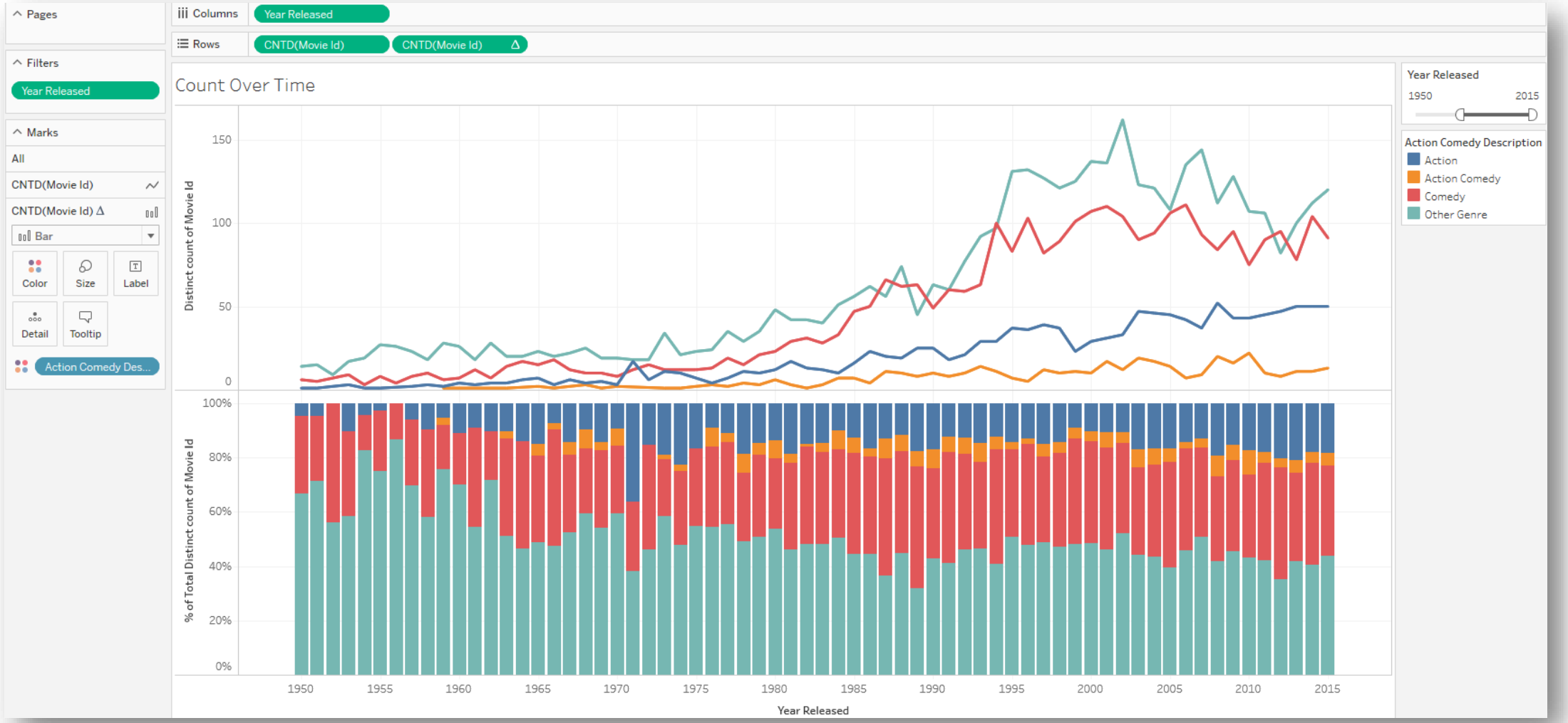


Table Calculations – Additional Resources

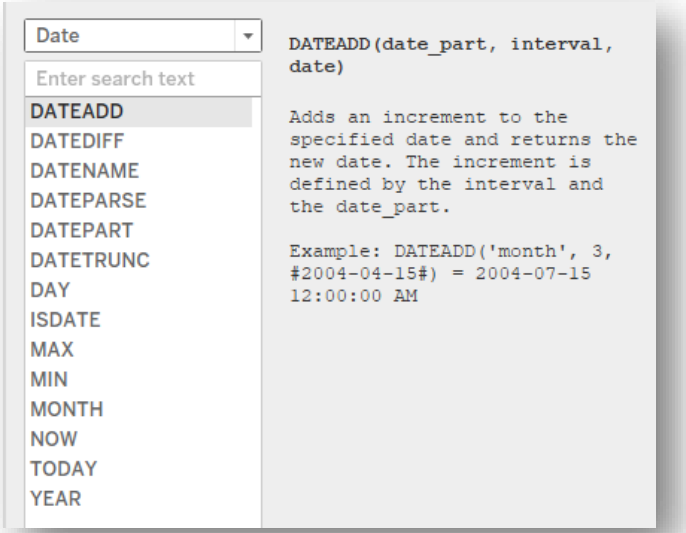
- This article covers multiple use cases for table calculations and provides reproducible examples
- This article provides an overview of Tableau's advanced table calculation settings as well as an instructional graphic
- This article provides examples for creating your own table calculations "from scratch" using calculated fields

Working with Dates and Timestamps

Tableau can work with both dates and timestamps, and a wide variety of functions are available

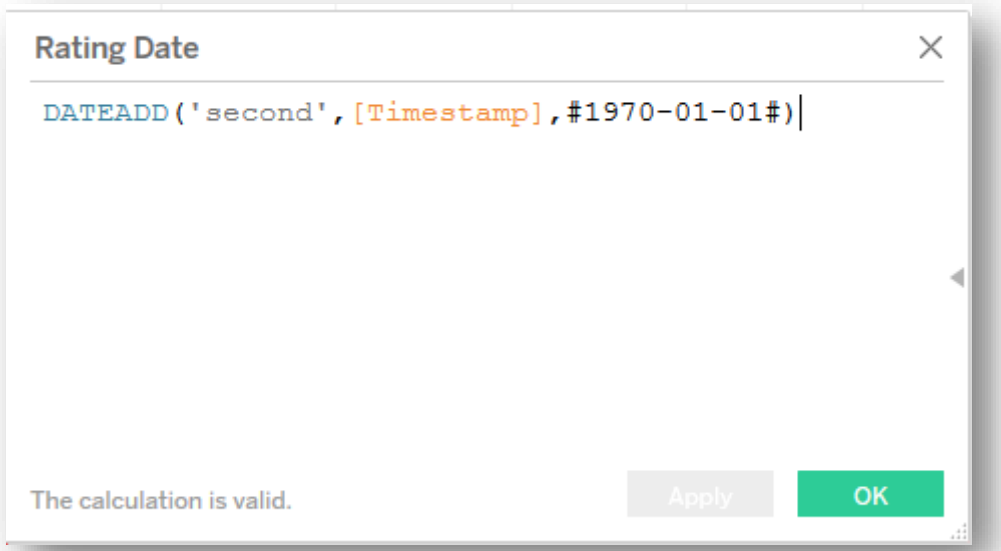
- In addition to these functions, three other functions are available to create both date fields and datetime fields:
  - MAKEDATE
  - MAKEDATETIME

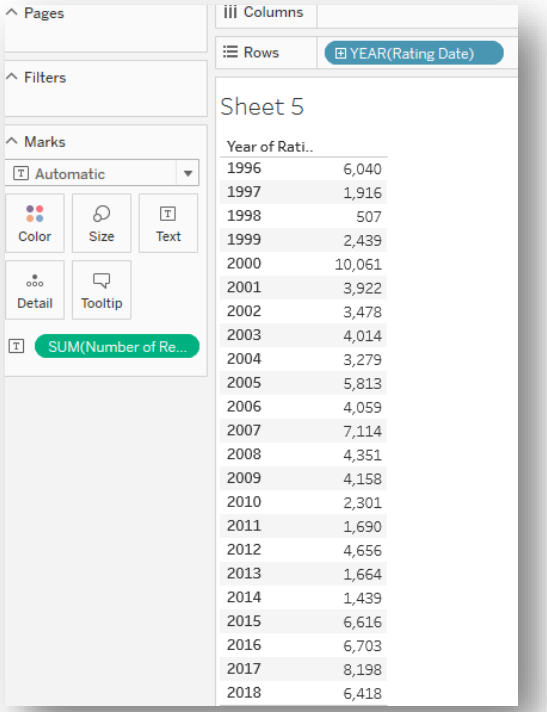
- This link provides additional information about using date functions in Tableau



DATEADD can be used to add any date or time component (termed date_part in Tableau) to a date

- Valid `date_part` values include: year, quarter, month, dayofyear, weekday, week, hour, minute, and second
- In the below example, the Timestamp variable will be used to derive the date each rating was generated





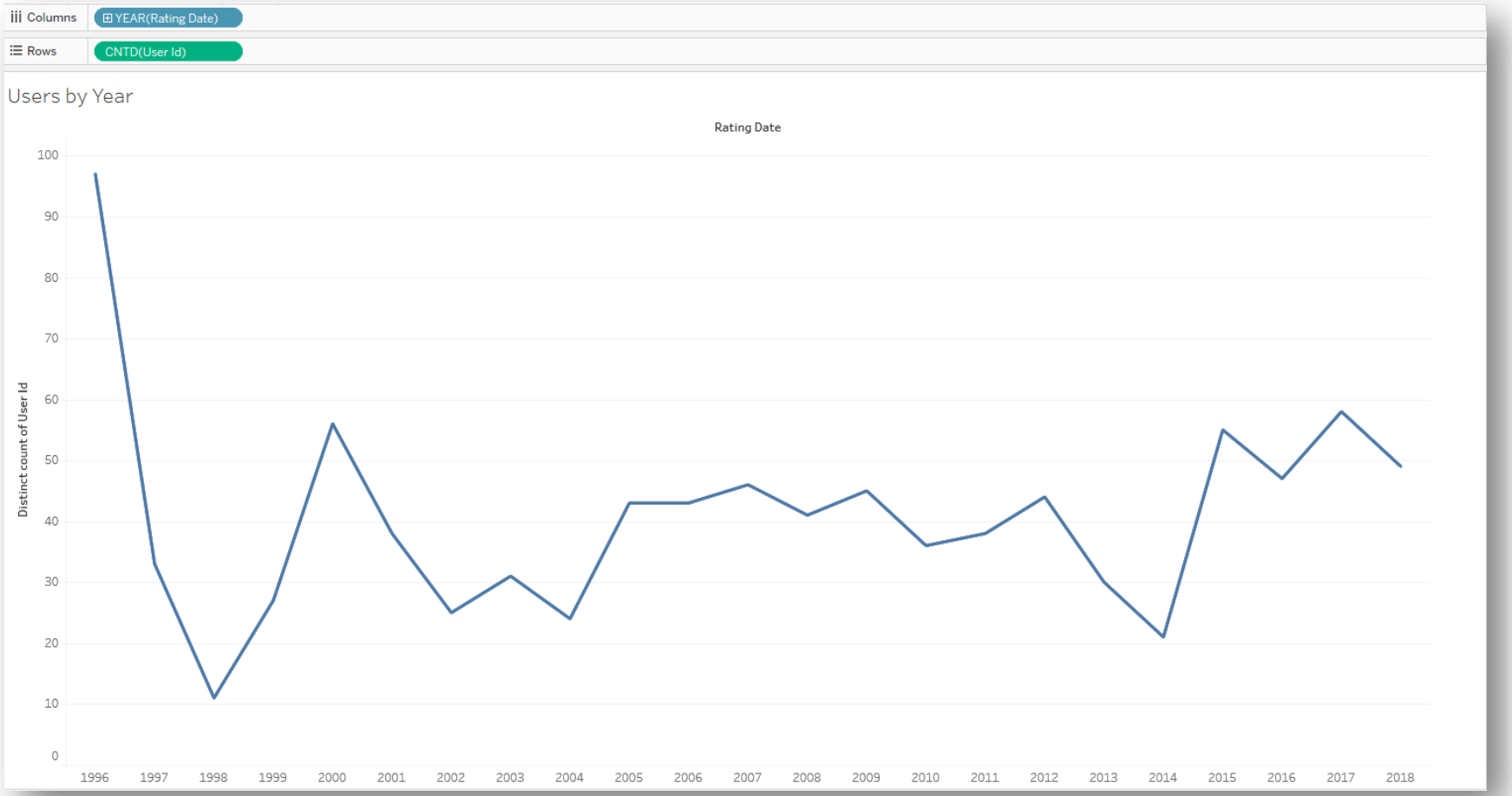
HANDS-ON EXERCISE

Build views to answer the following questions:

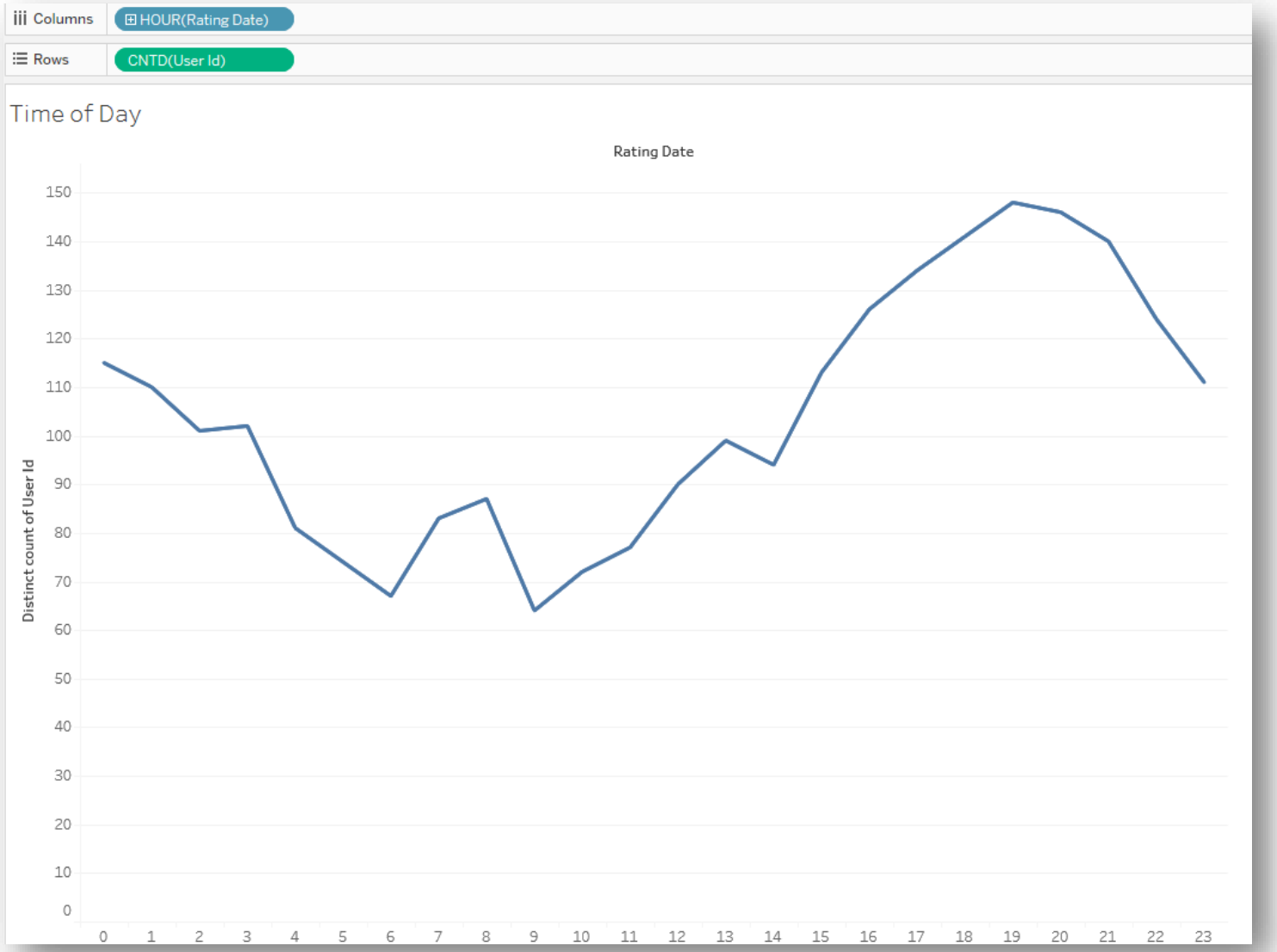
- What is the count of distinct users, by year?
- What time of day were users most likely to rate movies?

HINT: The **User ID** variable can be used to count users.

Solution – Count of Users by Year



Solution – Time of Day



EXTENDED HANDS-ON EXERCISE

- Derive a new calculated field named **Genres with Commas** that replace the pipe (|) character with commas.
- Recall that the Genre variable can identify one or more genres for a movie. Create a new view named **Genres** to answer the following question: What are the most common genre combinations?
- Create a dashboard that allows the user to filter the **Movie Rankings** view based on selections made in the **Genres** view.

HINT: The REPLACE function can be used to replace characters or substrings.

Example Solution - Dashboard