

Building Your Tableau Toolkit

Josh Mills, Data Science Leader (Sr. Consultant)

Enterprise Analytics Office

2021



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.

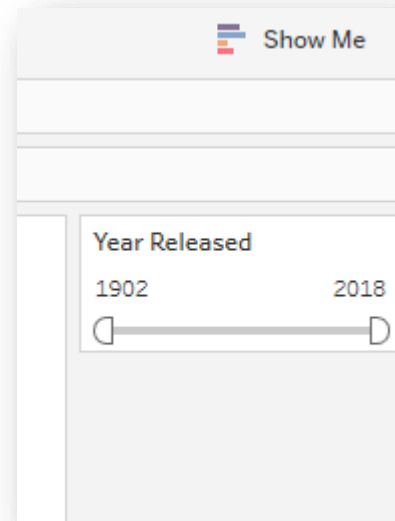
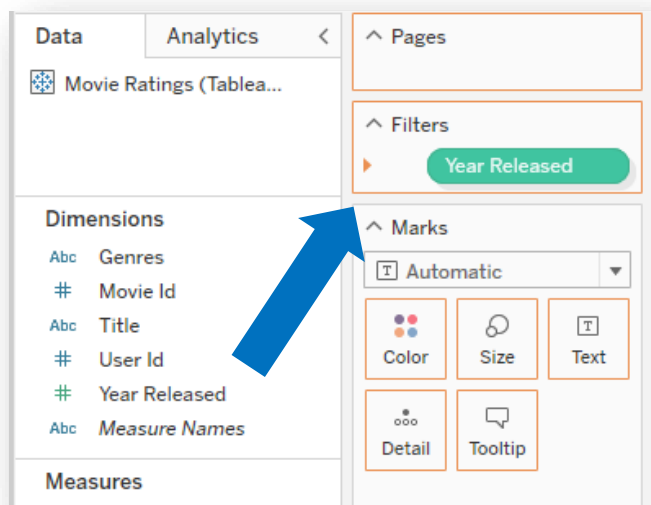
MovieLens Movie Ratings

- 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

Release Date

1/1/192210/23/1998

Date
Range

Release Date

Today

YearsQuartersMonthsWeeksDays

☐ Yesterday

☐ Last3days

☒ Today

☐ Next3days

☐ Tomorrow

10/13/2017 to 10/13/2017

Relative Date

Release Date

10/13/201710/13/2017

Start/End Date

Continuous Variable Filters - Examples

SUM(Rating)

12,541

Range of Values

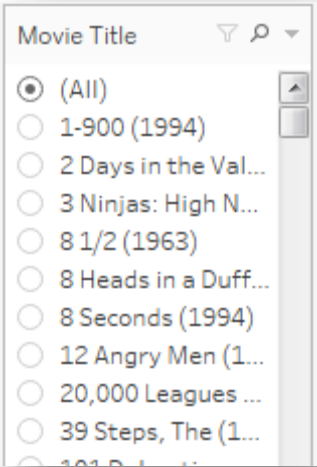
SUM(Rating)

12,541

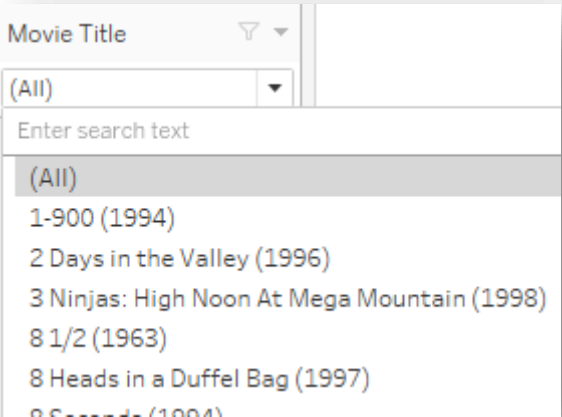
At Least/At Most

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

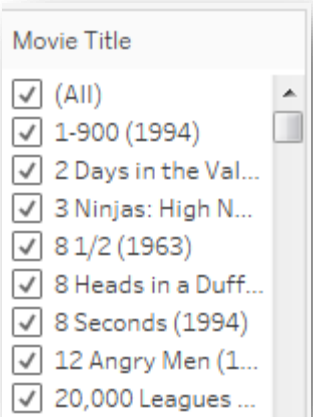
Discrete Variable Filters - Examples



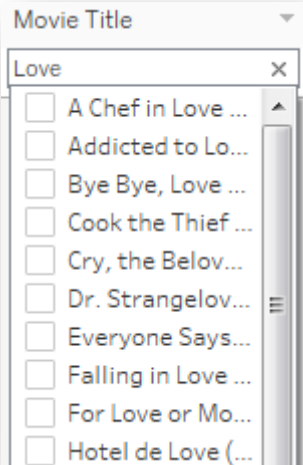
Single Value List



Single Value Dropdown

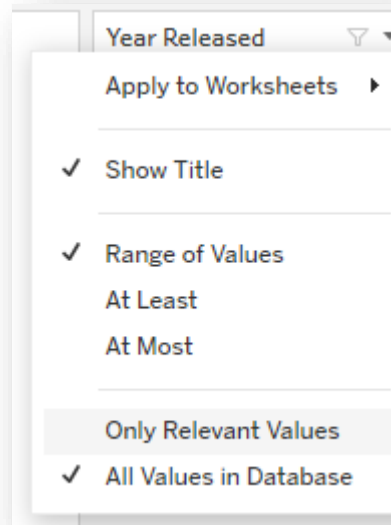


Multiple Value List/Dropdown



Custom List (Searchable)

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



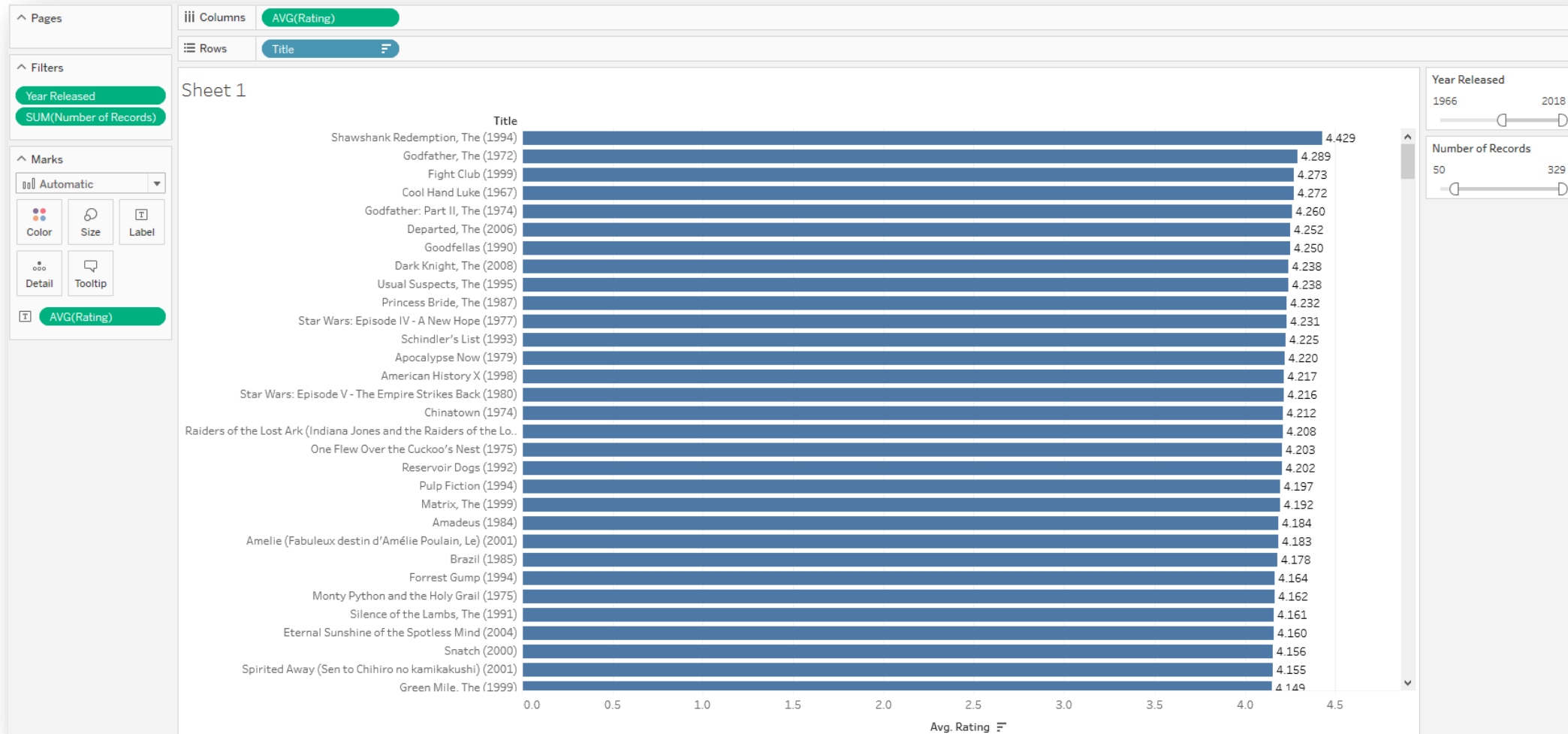
For example, you may want to show only relevant values in a filter rather than all possible values, particularly if the field you're using to filter has a high number of distinct values (> 500 typically)

HANDS-ON EXERCISE

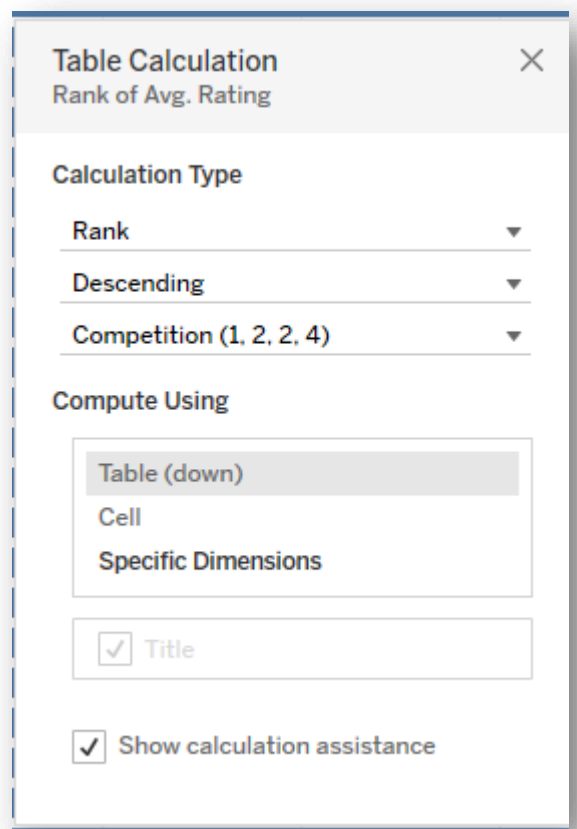
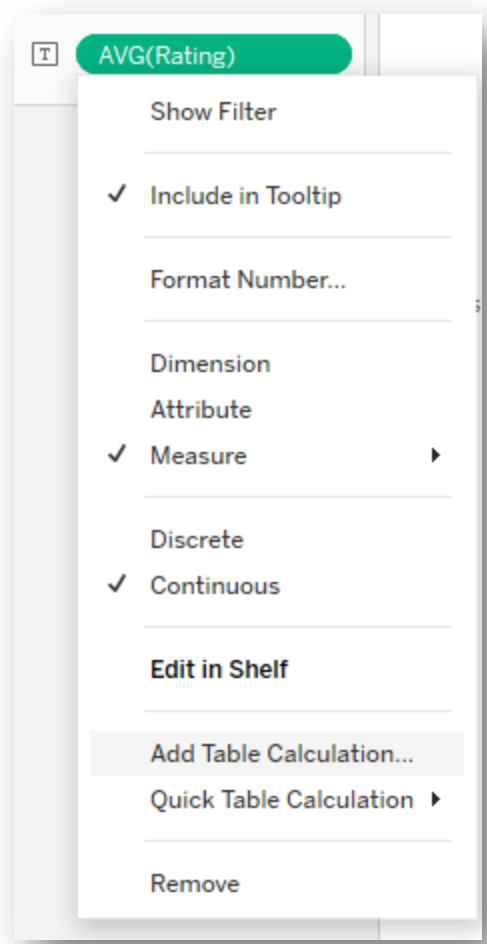
Create a new view (named **Movie Rankings**) that shows average movie rating by movie. Add filters for year released, and filter the view to show only movies released between 1966 and 2018. Then, add a filter to show only movies with at least 50 ratings. Sort the view by average movie rating, descending.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

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

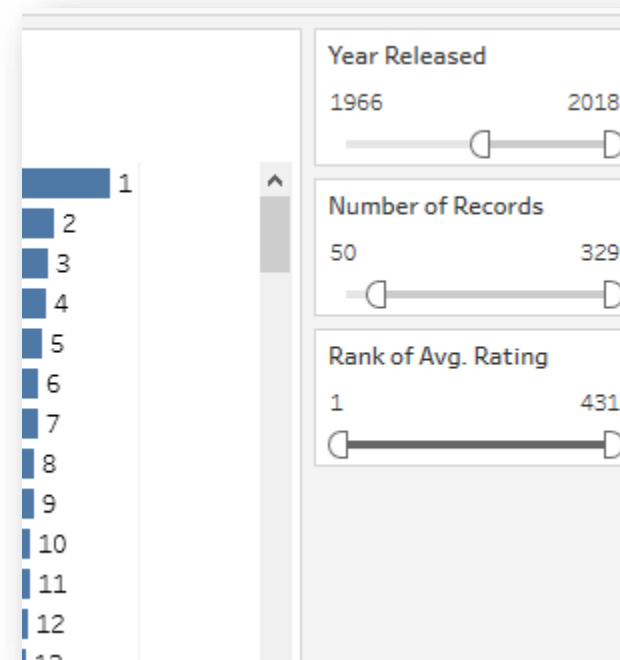
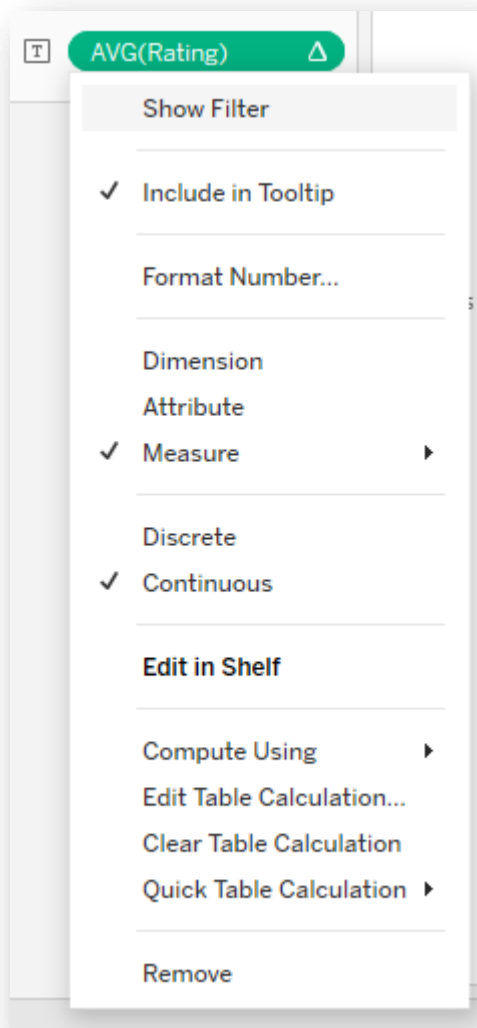
Solution



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

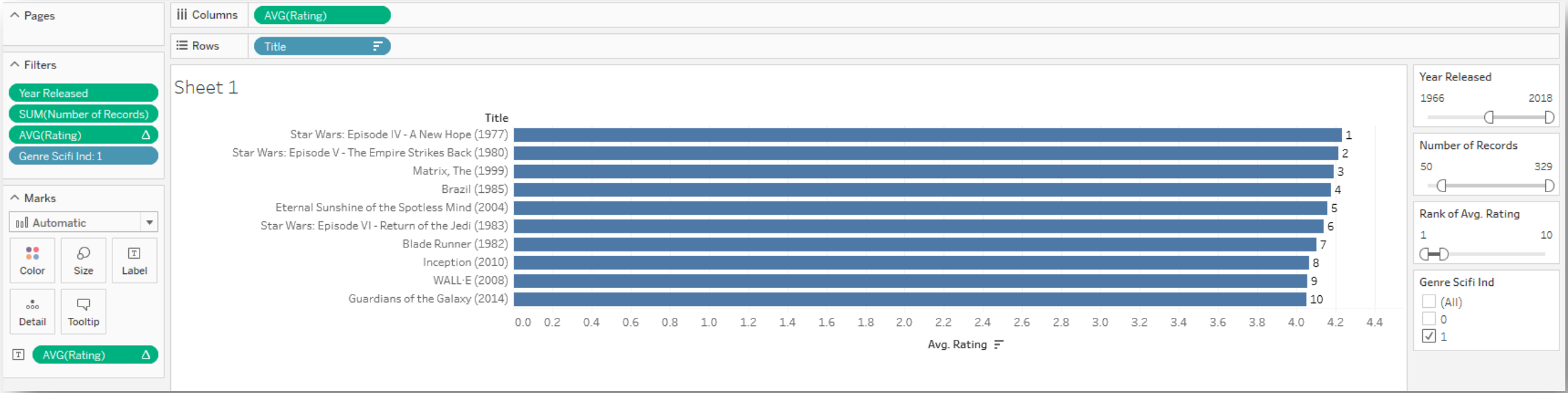


HANDS-ON EXERCISE

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

HINT: The Genre indicators can be converted to dimensions.

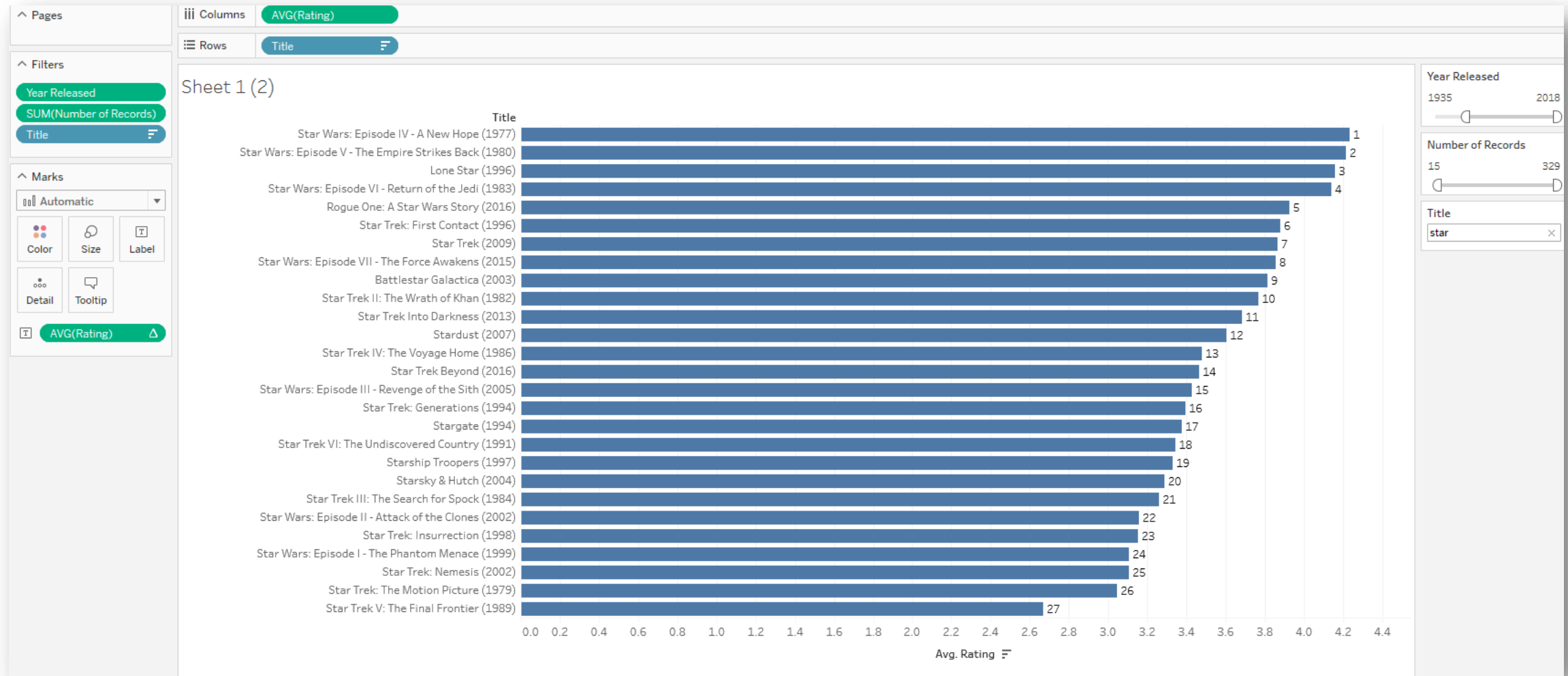
Solution



HANDS-ON EXERCISE

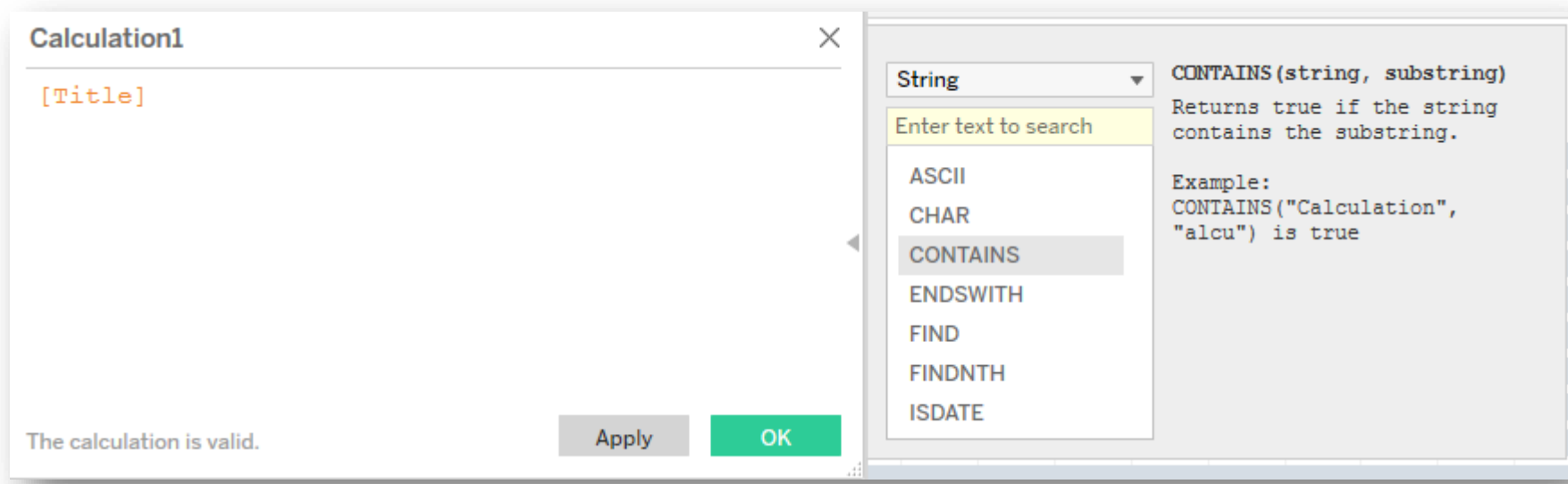
Create a copy of the view and add a filter for movie title. Remove the top X filter and the sci-fi filter. Reduce the minimum number of ratings to 15. Change the filter type to a wildcard match. Search for movies with the word “star” in the title.

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

First 4

MID([Movie Title],1,4)|

Pages

Columns

Rows

Movie Title

Filters

Marks

Automatic

Color

Size

Text

Detail

Tooltip

First 4

Sheet 1

Movie Title

Bridges of Madison Count..	Brid
Bringing Up Baby (1938)	Brin
Broken Arrow (1996)	Brok
Broken English (1996)	Brok
Bronx Tale, A (1993)	Bron
Brother Minister: The Ass..	Brot
Brothers in Trouble (1995)	Brot
Brothers Kiss, A (1997)	Brot
Brothers McMullen, The (..	Brot
Browning Version, The (1..	Brow
Buddy (1997)	Budd
Bulletproof (1996)	Bull
Bullets Over Broadway (1..	Bull
Burnt By the Sun (1994)	Burn

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!

Columns	
Rows	Movie Title
Sheet 1	
Movie Title	
Bridges of Madison Count..	False
Bringing Up Baby (1938)	False
Broken Arrow (1996)	False
Broken English (1996)	False
Bronx Tale, A (1993)	False
Brother Minister: The Ass..	False
Brothers in Trouble (1995)	False
Brothers Kiss, A (1997)	False
Brothers McMullen, The (.	False
Browning Version, The (1..	False
Buddy (1997)	False
Bulletproof (1996)	False
Bullets Over Broadway (1..	False
Burnt By the Sun (1994)	False
Burnt Offerings (1976)	False
Bushwhacked (1995)	False
Butch Cassidy and the Su..	False
Butcher Boy, The (1998)	False
Butterfly Kiss (1995)	False
Bye Bye, Love (1995)	False
Cabin Boy (1994)	False
Cable Guy, The (1996)	False
Calendar Girl (1993)	False
Canadian Bacon (1994)	False
Candidate, The (1972)	False
Candyman (1992)	True
Candyman: Farewell to th..	True
Cape Fear (1962)	False
Cape Fear (1991)	False
Captives (1994)	False

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

Year String

`RIGHT([Movie Title],6)`

Pages

Columns

Rows

Movie Title

Filters

Sheet 1

Movie Title

Table

Marks

Automatic

Color

Size

Text

Detail

Tooltip

Year String

Bridges of Madison Count..	(1995)
Bringing Up Baby	(1938)
Broken Arrow	(1996)
Broken English	(1996)
Bronx Tale, A	(1993)
Brother Minister: The Ass..	(1994)
Brothers in Trouble	(1995)
Brothers Kiss, A	(1997)
Brothers McMullen, The (..	(1995)
Browning Version, The (1..	(1994)
Buddy	(1997)
Bulletproof	(1996)
Bullets Over Broadway (1..	(1994)
B... ..	(1994)

Concatenating Strings

- Strings can be combined simply by using the + operator

City-State-Zip

Restaurant Inspections

X

[City] + ", " + [State] + " " + [ZIP]

The calculation is valid.

Sheets Affected

Pages

Columns

Rows

City

State

ZIP

Filters

Sheet 7

City

State

ZIP

Blue Diamond

Nevada

89004

Blue Diamond, Nevada 89004

Boulder City

Nevada

89005

Boulder City, Nevada 89005

89005-1401

Boulder City, Nevada 89005-1401

89005-1805

Boulder City, Nevada 89005-1805

89005-1808

Boulder City, Nevada 89005-1808

89005-1809

Boulder City, Nevada 89005-1809

89005-1825

Boulder City, Nevada 89005-1825

89005-1827

Boulder City, Nevada 89005-1827

89005-1841

Boulder City, Nevada 89005-1841

89005-1867

Boulder City, Nevada 89005-1867

89005-1903

Boulder City, Nevada 89005-1903

89005-1904

Boulder City, Nevada 89005-1904

89005-1908

Boulder City, Nevada 89005-1908

89005-1909

Boulder City, Nevada 89005-1909

89005-1914

Boulder City, Nevada 89005-1914

Marks

Automatic

Color

Size

Text

Detail

Tooltip

City-State-Zip

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.

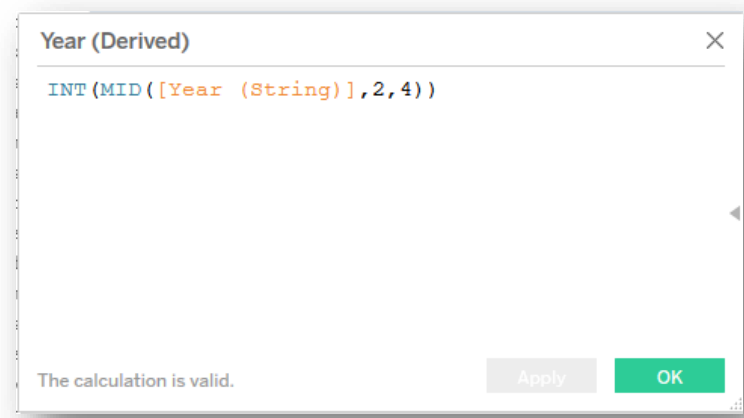
BONUS: Create a new calculated field that only contains the movie title without the year, named **Title Only**. Then, recreate the original movie title field using Tableau's concatenation functions, named **Recreated Title**. You will need to use a combination of the LEFT function (which performs a similar 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



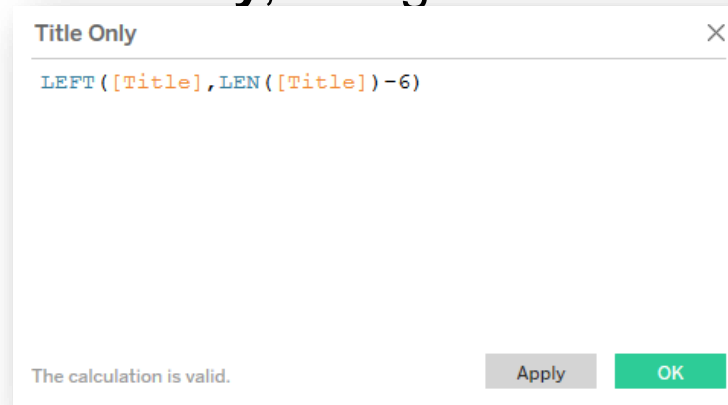
Then create a new calculated field named **Year (Derived)** using the MID and INT functions



**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

Street in Title Ind (Case)

```
if CONTAINS(LOWER([Title]),"street") then 1 else 0 end
```

The calculation is valid.

Apply

OK

All

Enter search text

ABS
ACOS
AND
ASCII
ASIN
ATAN
ATAN2
ATTR
AVG
CASE
CEILING
CHAR
COLLECT
CONTAINS
CORR
COS
COT

ABS (number)

Returns the absolute value of the given number.

Example: ABS(-7) = 7

Pages

Filters

Street in Title Ind: 1

SUM(Number of Records)

Marks

Automatic

Color

Size

Text

Detail

Tooltip

AVG(Rating)

Columns

Rows

Title Genres

Sheet 3

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

Street in Title Ind

☐ (All)
☐ 0
☒ 1

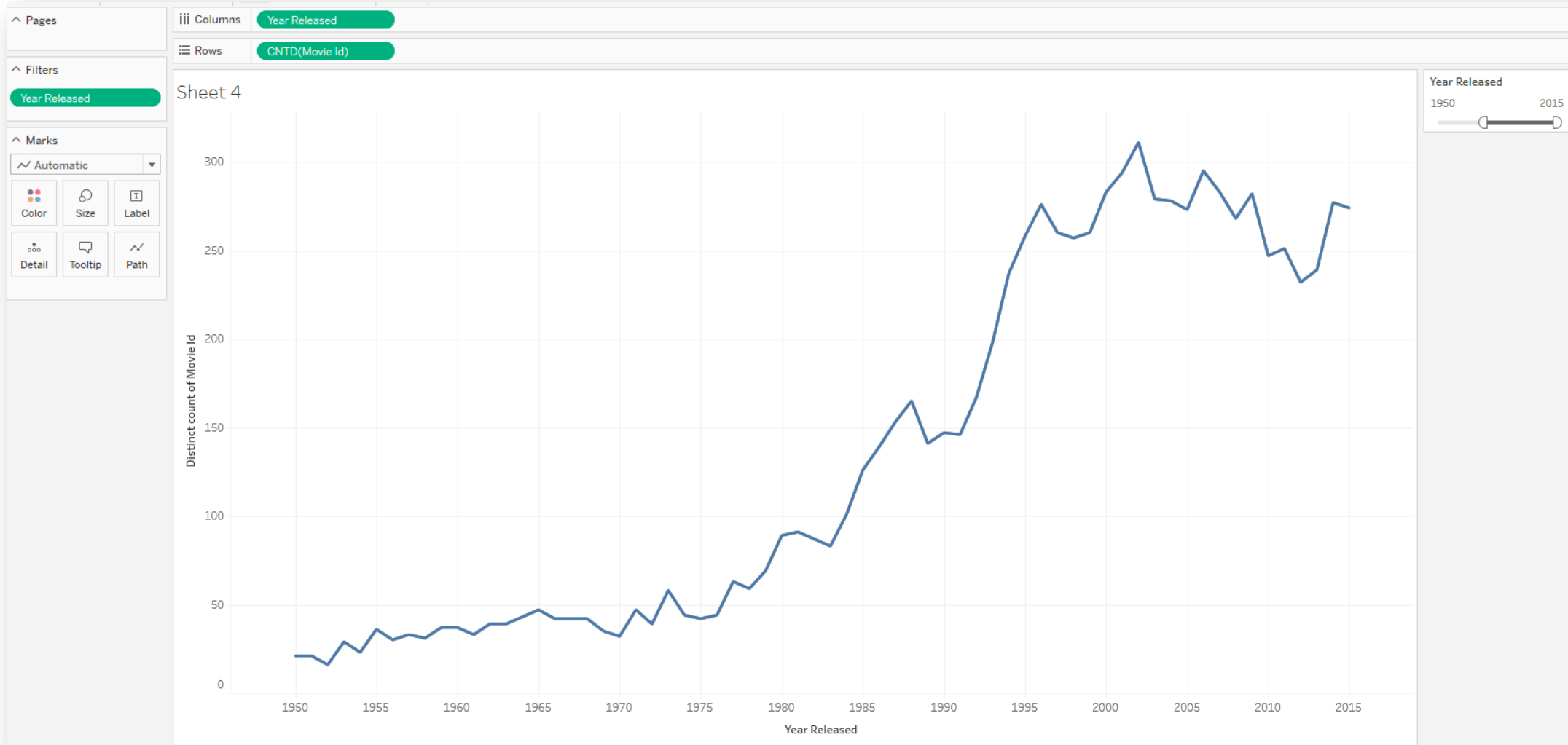
Number of Records

10 54

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

Action Comedy Description

×

```
if [Genre Comedy Ind] = 1 and [Genre Action Ind] = 1 then "Action Comedy"
elseif [Genre Comedy Ind] = 0 and [Genre Action Ind] = 1 then "Action"
elseif [Genre Comedy Ind] = 1 and [Genre Action Ind] = 0 then "Comedy"
else "Other Genre"
end
```

The calculation is valid.

2 Dependencies ▾

Apply

OK

Solution – Final View with Color Legend

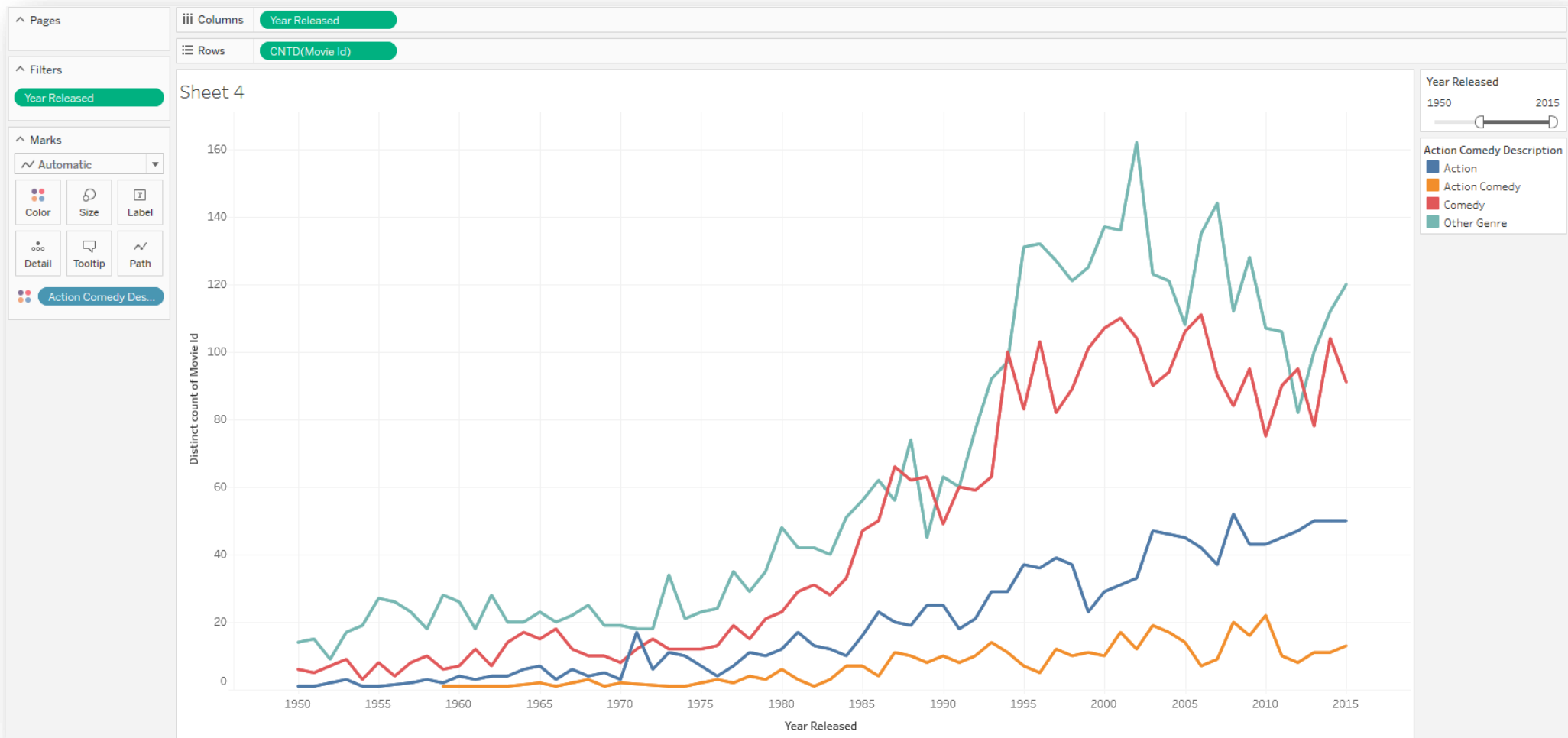
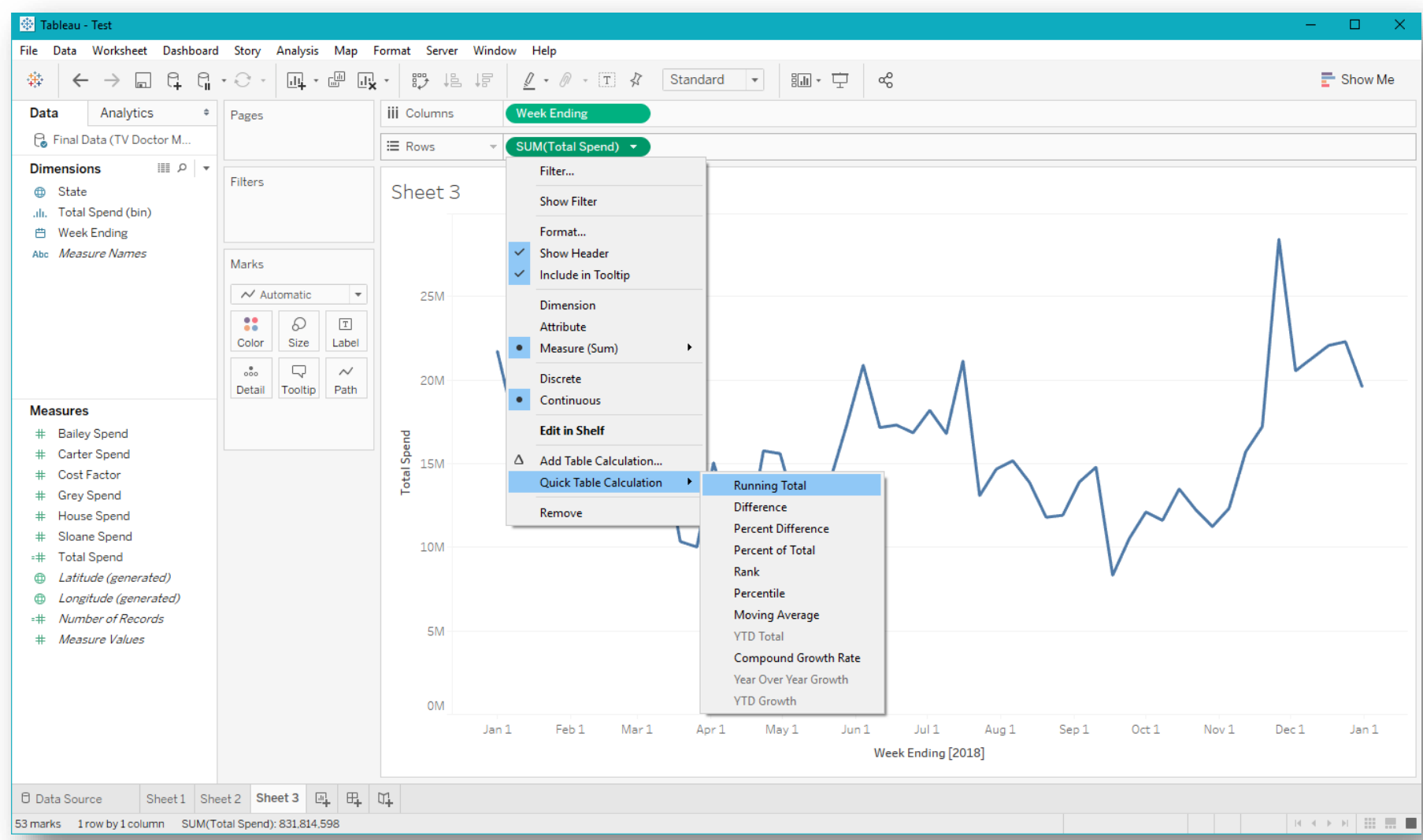


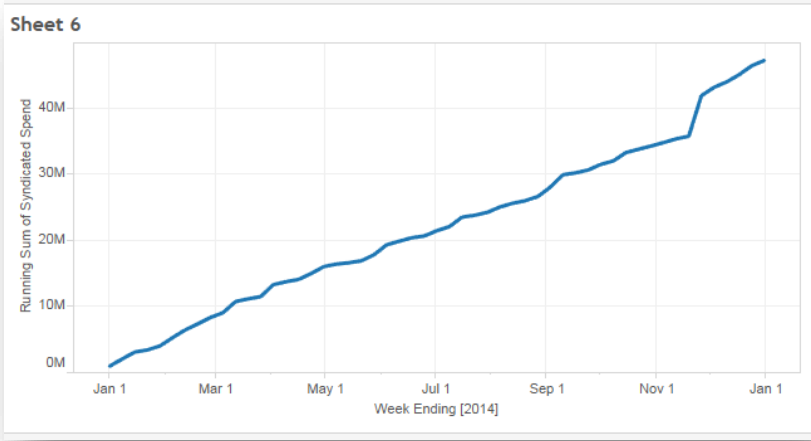
Table Calculations

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

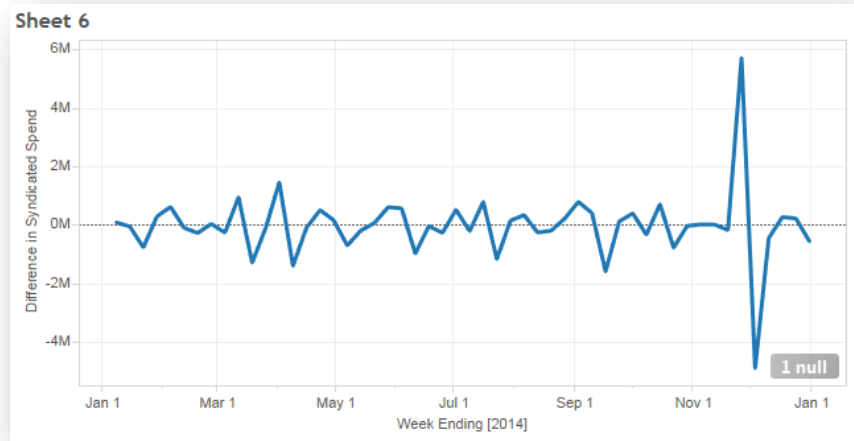


Examples include cumulative sums, % of total, absolute/percent differences, moving average, and rank

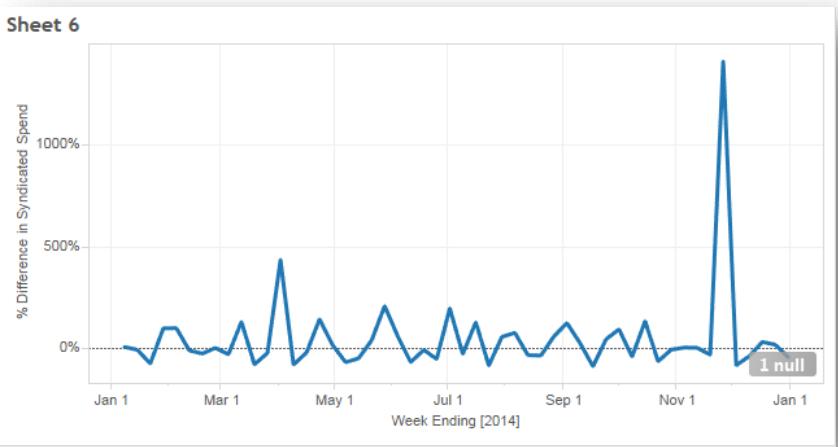
Running Total/Cumulative Sum



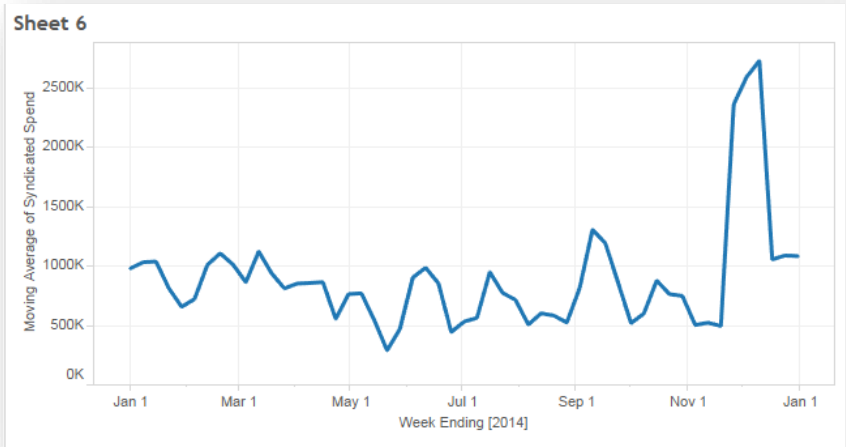
Absolute Difference



% Difference

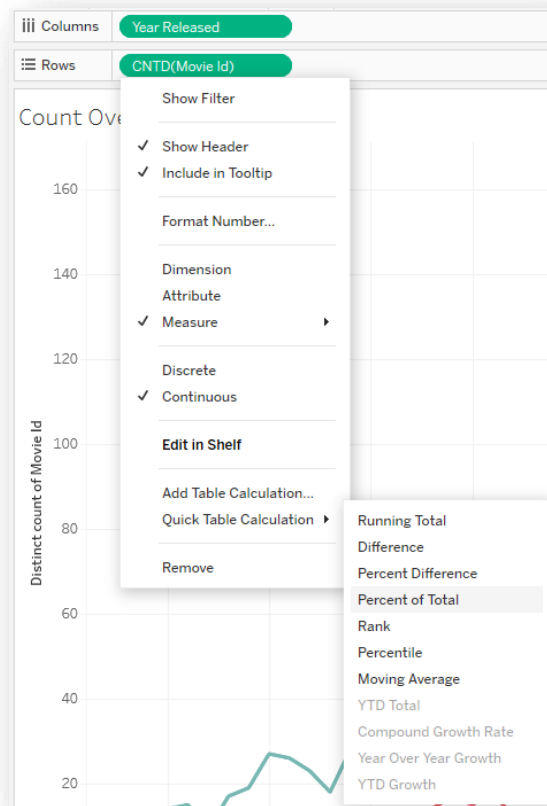


Moving Average



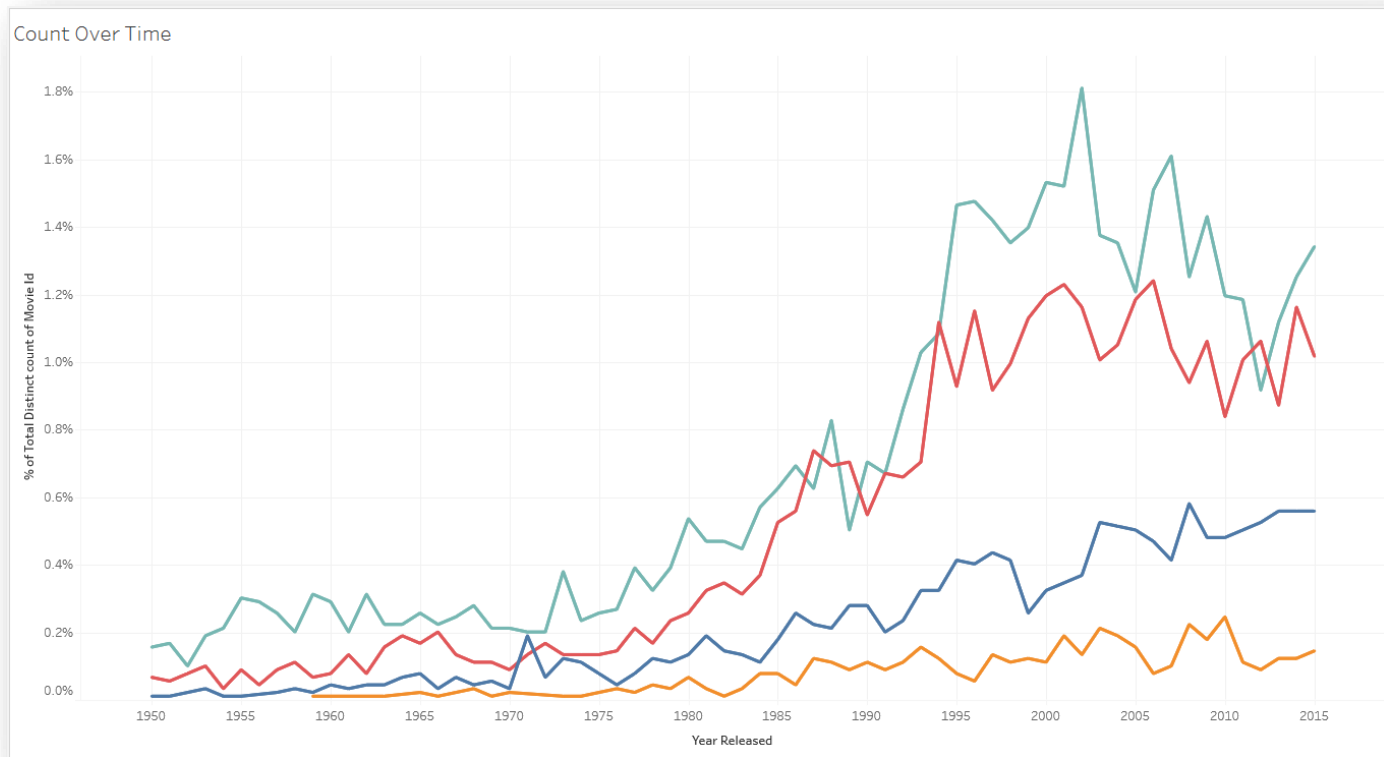
Example – Movie Count by Genre and By Year

- **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



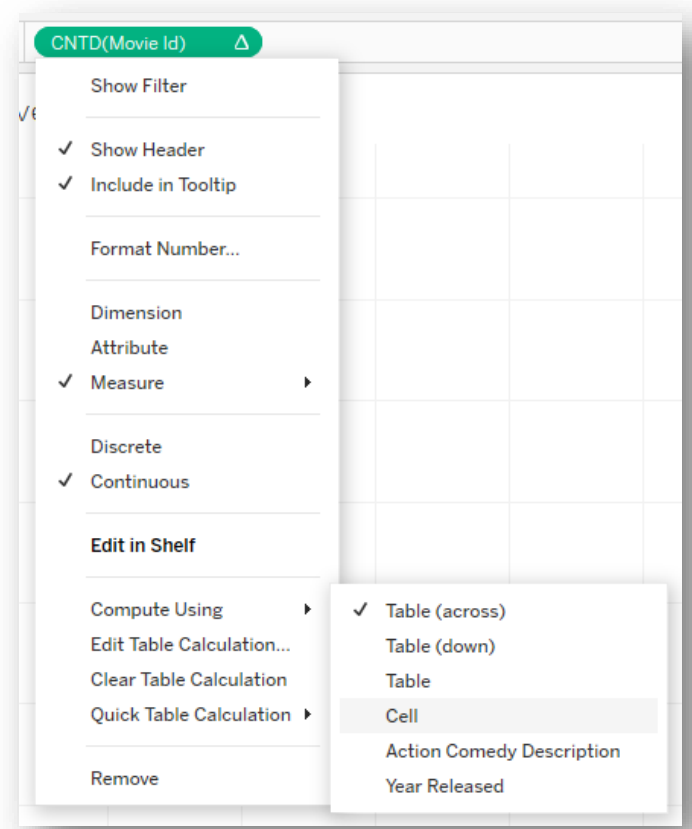
The resulting view is a bit misleading, and the view will need modifications to answer the question being posed

- 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.



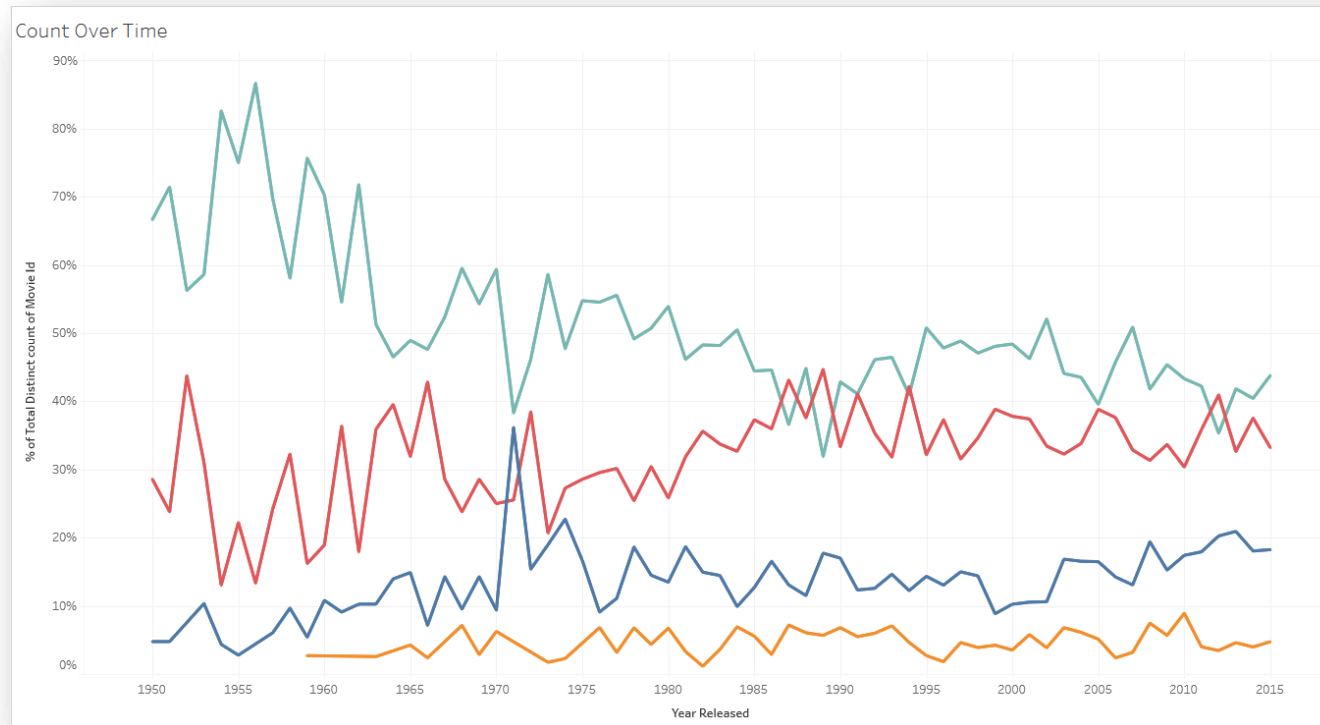
Modify the view by choosing the drop-down menu for CNTD(Movie Id) and choosing Compute Using → Cell.

- 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

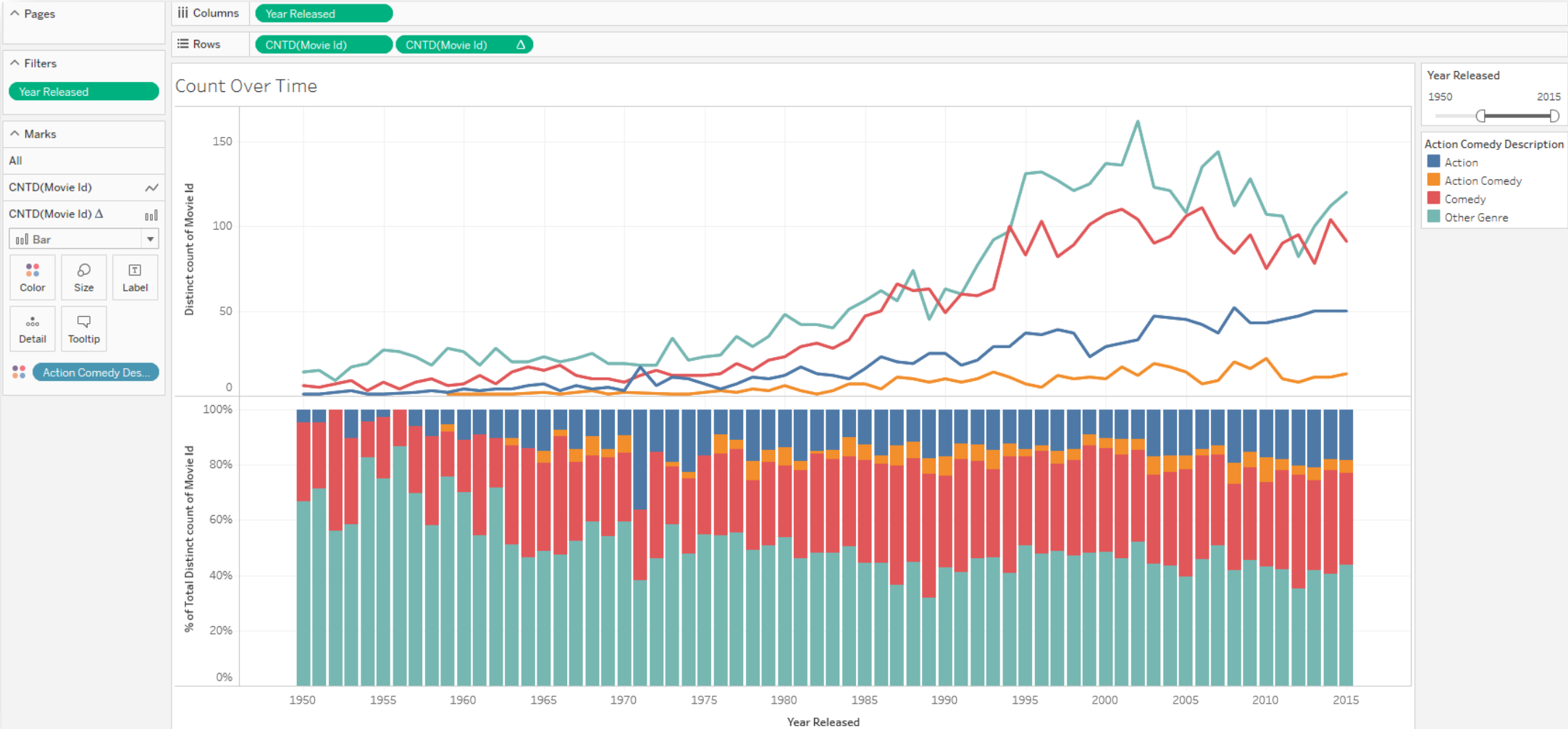


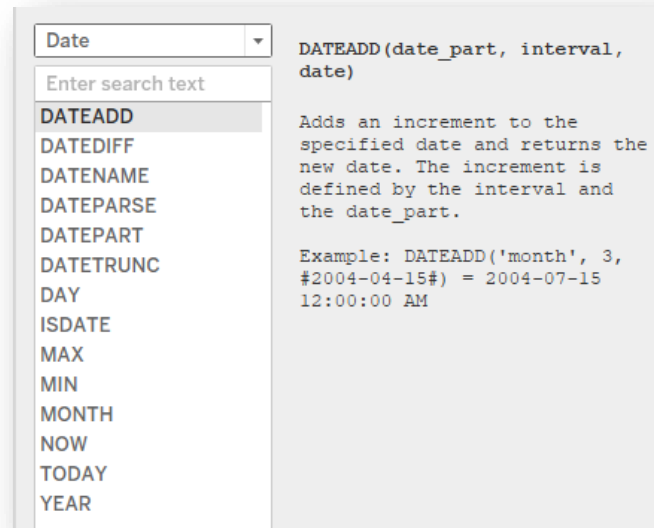
Table Calculations – Additional Resources

- [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

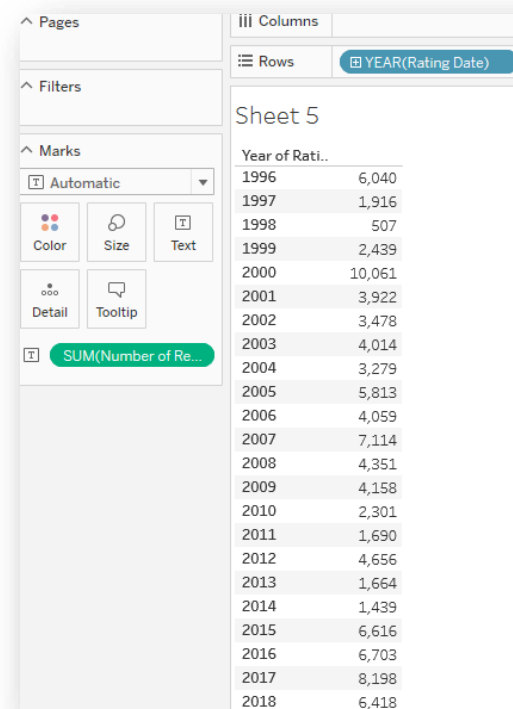
Tableau can work with both dates and timestamps, and a wide variety of functions are available

- 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



Columns: YEAR(Rating Date)

Rows: SUM(Number of Re...)

Sheet 5

Year of Rati..	
1996	6,040
1997	1,916
1998	507
1999	2,439
2000	10,061
2001	3,922
2002	3,478
2003	4,014
2004	3,279
2005	5,813
2006	4,059
2007	7,114
2008	4,351
2009	4,158
2010	2,301
2011	1,690
2012	4,656
2013	1,664
2014	1,439
2015	6,616
2016	6,703
2017	8,198
2018	6,418

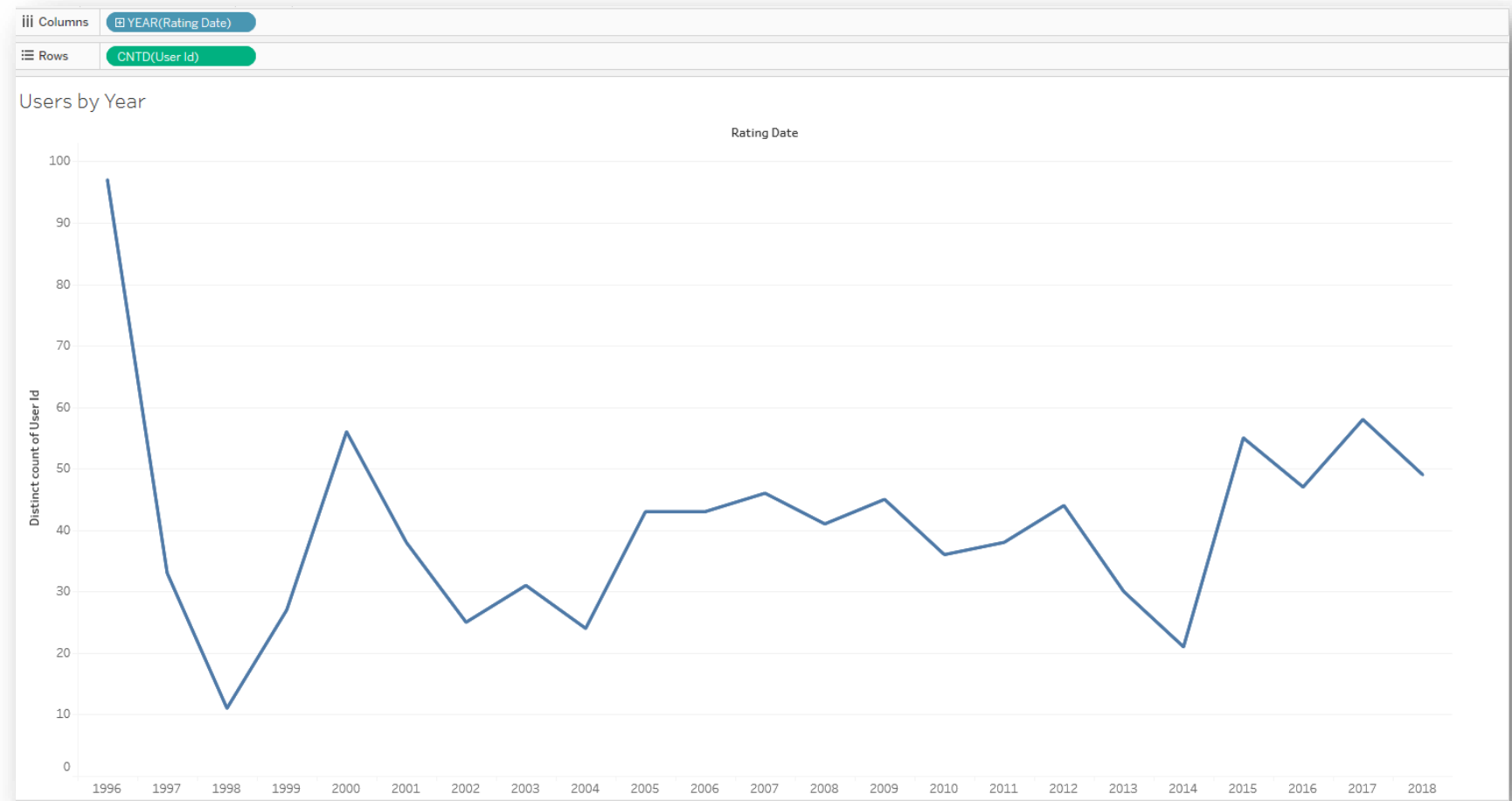
HANDS-ON EXERCISE

Build views to answer the following questions:

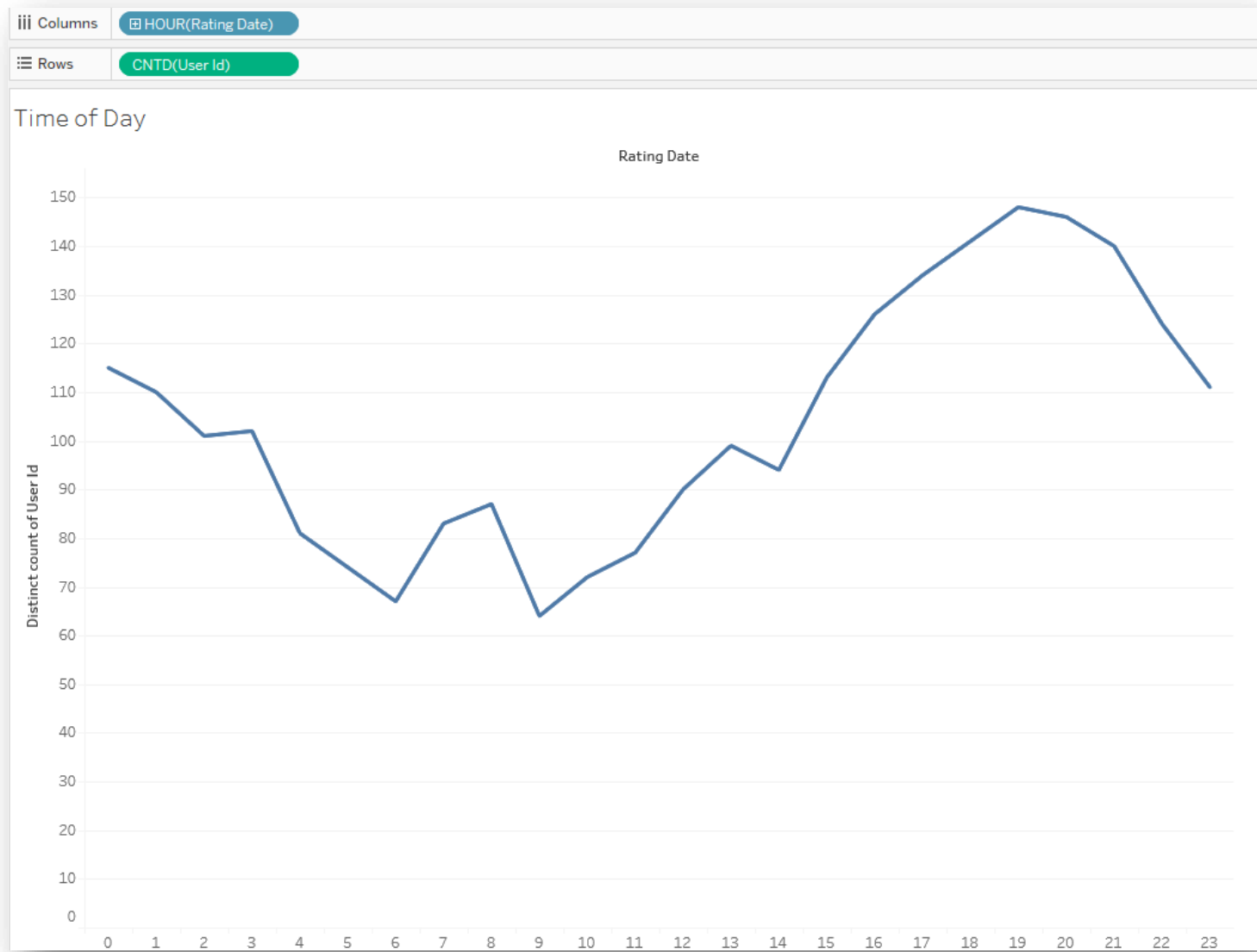
- What is the count of distinct users, by year?
- What time of day were users most likely to rate movies?

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

