Assignment 4 – Calculated Fields & Parameters

Professor John Sokol | Due 2/22

Background Information:

Calculated fields are custom data fields that are created within the Tableau environment. Calculated fields are usually created from existing data fields, but calculated fields can also be solely text and numbers.

Here are a few applications of calculated fields:

- To segment data
- To convert data types, such as converting a string to a date.
- To aggregate data
- To filter results
- To calculate ratios

Aggregation:

Before we take low level (detailed) dive into calculated fields, we need to understand the concept of aggregation, which is applicable to Tableau measures. Aggregation defines how data fields are expressed within Tableau. You have probably noticed when you drag a measure data field to the columns or rows tab, Tableau automatically aggregates the data field as a sum of all the numbers in the data field, like so:
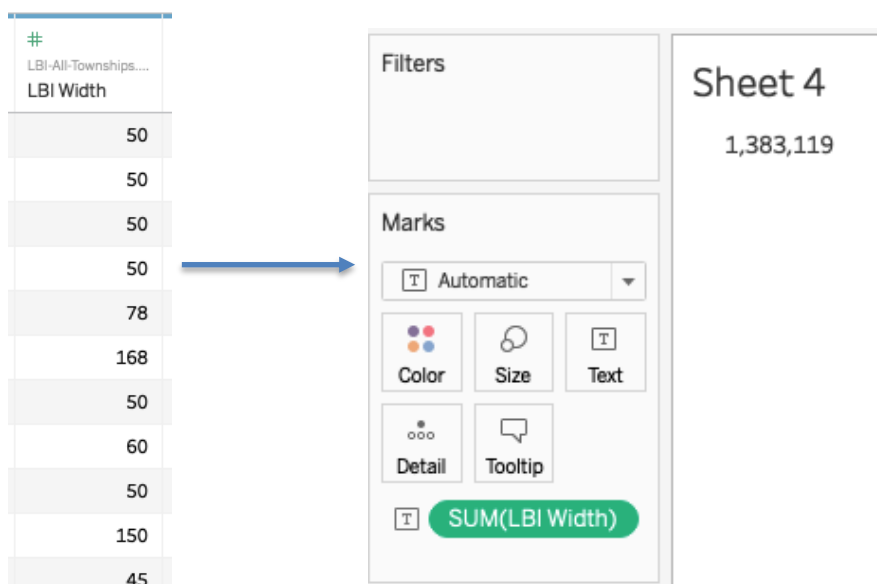

The LBI Width data field has about 19,000 rows, so the screenshot to the left is only a fraction of all the rows contained in the data field. Tableau defaults the aggregation to SUM( ) upon

dragging the data field to text in the marks card of a worksheet. The result is over one million feet of width. However, we can select other aggregate functions that meet our analytical needs:
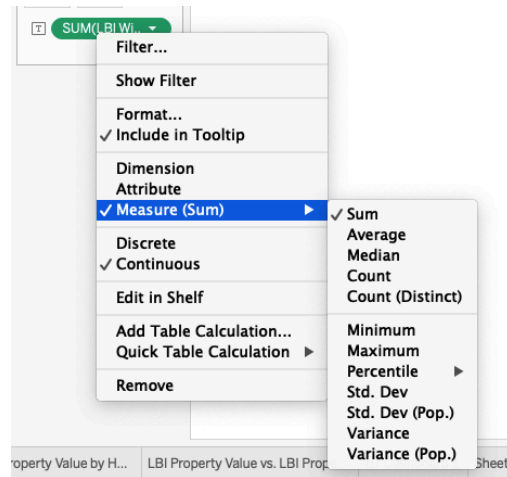
The two most common aggregates used are Sum( ) and Average( ). Median( ) is also useful when the average is significantly influenced by outliers, such as in the LBI dataset:
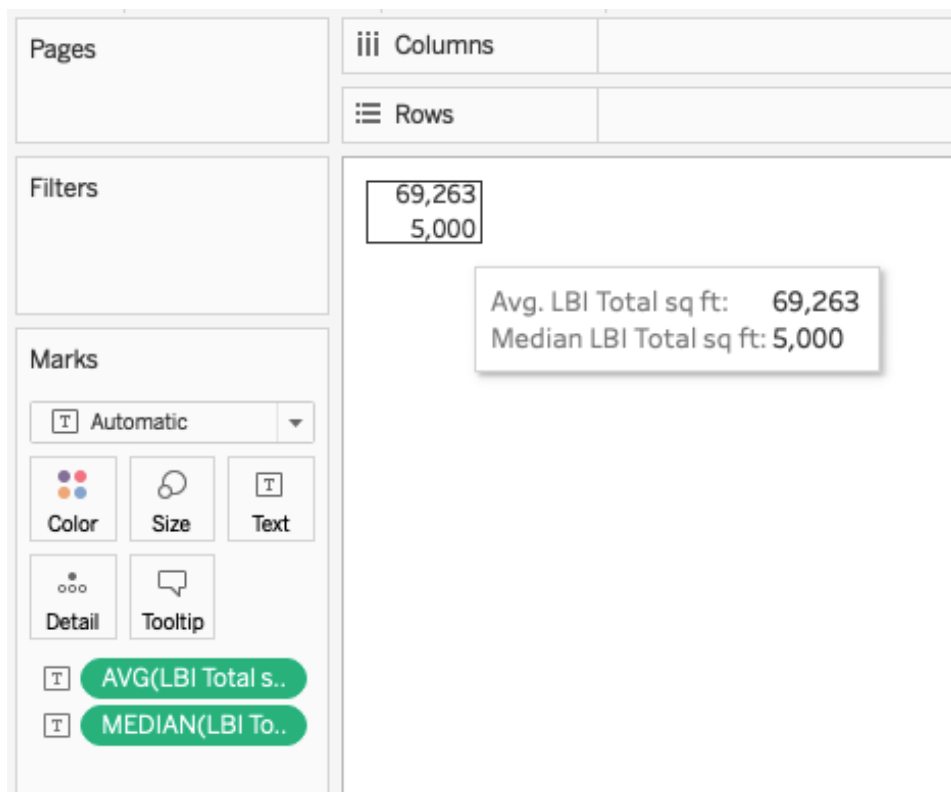
I created a calculated field of the total square feet of each property (the land, not the inside of the home) with [LBI Width] * [LBI Length]. Look at the average compared to the median values. It is obvious that the average is greatly influenced by municipal buildings that have more land

than a regular homeowner lot. Therefore, the median is a better representation of how much land a typical homeowner on LBI has. This is a great exercise to explore your data. I highly recommend looking at the average, median, and standard deviation of a data field before visualization to become more familiar with your data from a summary statistics standpoint and how the data is distributed.

Calculated Field Example:

Consider the LBI dataset that has the data fields 'LBI Town' and 'LBI State':

What if we want to create a new data field that combines the two into one data field? Like so:

We can do this within Tableau with the following calculated field:

Short and descriptive title

LBI Town + LBI State

[LBI Town] + ', ' + [LBI State]

LBI Town Data Field

LBI State Data Field

Comma (type string) to separate data fields. Notice the space after the comma

Tableau confirmation that syntax is correct

The calculation is valid.

Apply   OK

Calculated Field Elements:
Calculated Fields are defined by entering a formula into Tableau's calculated field editing dialog box:

Data   Analytics

LBI-All-Townships

Dimensions
Create Calculated Field...
Create Parameter...

Group by Folder
✓ Group by Data Source Table

✓ Sort by Name
Sort by Data Source Order

Hide All Unused Fields
Show Hidden Fields

Abc LBI Town
⊕ LBI Zip
Abc Updated
Abc Measure Names

Pages

Filters

Marks

T Automatic

Color   Size   Text

Detail   Tooltip

T SUM(LBI Width)

Descriptive calculated field name

Apply   OK

Elements of a calculated field:
- Functions
  - MIN( ), MAX( ), AVG( ), DATEDIFF( )...
- Existing Data field(s)
  - [Name], [Age], [LBI Town]...
- Operators

- o Mathematical operators and comparative operators for strings, dates, text, etc
- Miscellaneous
  - o Parameters for creating variables that are accessible to the Tableau user
  - o Comments for documenting notes within the formula box; use the double forward slash to create a comment: //, the comment will be greyed out.

Functions:
The complete Tableau documentation for functions can be found here. The following are notable functions used frequently when creating calculated fields:
Number functions:
- MAX(number, number): Returns the maximum of the two arguments
  - o MAX(5, 7) = 7
- MIN(number, number): Returns the minimum of the two arguments
  - o MIN(5, 7) = 5
- ZN(expression): Returns the expression if it is NOT null, otherwise returns zero. Use this function to use zero values instead of null values. Only works with numbers/measures

String functions:
- LEFT(string, number): Returns the left-most number of characters in the string
  - o LEFT("Matador", 4) = "Mata"
- LEN(string): Returns the length of the string
  - o LEN("Matador") = 7
- TRIM(string): Returns the string with leading and trailing spaces removed
  - o TRIM("   Data Visualization   ") = "Data Visualization"

Date functions:
- DATEDIFF(date_units, start_date, end_date): Returns the difference between start_date and end_date expressed in units of date_units.
  - o DATEDIFF('month', 1/1/2016, 5/31/2016) = 5
- NOW( ): Return the current date and time
  - o NOW( ) = 2018-10-12 8:41:00 PM
- TODAY( ): Returns the current date
  - o TODAY( ) = 2018-10-12

Type conversion: Converting a data field from one data type to another data type.
- DATE( )
  - o Returns a date given a number, string, or date expression

- FLOAT( )
  - Returns a float; converts the argument (the data field you place inside the parenthesis) to data type float.

Logical Functions:
- IF Statement: IF <expr> THEN <then> [ELSEIF <expr2> THEN <then2>…] [ELSE <else>] END:
  - Tests a series of expressions returning the <then> value for the first true <expr>.
  - IF [Profit] > 0 THEN 'Profitable' ELSEIF [Profit] = 0 THEN 'Breakeven' ELSE 'Loss' END
- ISNULL(expression): Returns true if the expression does not contain valid data (Null).

Although calculated fields usually incorporate a function, they are not required. Calculated fields can also be used to create a new data field of just numbers or text:

In this example, I use a calculated field to create a title, which is of data type string, for the median property value.

Operators:
- General
  - + (addition): Adds two numbers; Concatenates two strings; Adds days to date.
    - 7 + 3 = 10; 'abc' + 'def' = 'abcdef'; April 15th 2018 + 15 = April 30th 2018.
  - − (subtraction): Subtracts two numbers. Subtracts days from date.
    - 7 − 3 = 4; April 15th 2018 - 15 = April 1st 2018.

- Arithmetic
  - * (multiplication): Numeric multiplication.

- 23 * 2 = 46.
    - / (division): Numeric division.
        - 45 / 2 = 22.5.
    - % (modulus): Remainder of numeric division.
        - 13 % 2 = 1; 2 goes into 13, 6 times with 1 as the remainder.
    - ^ (power): Numeric exponent.
        - 2^3 = 8.

- Comparison
    - = or == (equal to):
        - Compares two numbers or two strings or two dates to be equal. Returns the Boolean value TRUE if they are, else returns false.
    - ! = or <> (not equal to):
        - Compares two numbers or two strings or two dates to be unequal. Returns the Boolean value TRUE if they are, else returns false.
    - > (greater than):
        - Compares two numbers or two strings or two dates where the first argument is greater than second. Returns the boolean value TRUE if it is the case, else returns false.
    - < (less than):
        - Compares two numbers or two strings or two dates where the first argument is smaller than second. Returns the boolean value TRUE if it is the case, else returns false.

- Logical
    - AND
        - If the expressions or Boolean values present on both sides of AND operator is evaluated to be TRUE, then the result is TRUE. Else the result is FALSE.
    - OR
        - If any one or both of the expressions or Boolean values present on both sides of AND operator is evaluated to be TRUE, then the result is TRUE. Else the result is FALSE.
    - NOT
        - This operator negates the Boolean value of the expression present after it.

Parameters:

For the previous assignment, we saw how parameters can be used as a static value for a bar graph reference line. But the real power of parameters lies with their ability to be changed by the Tableau user. Consider this visualization of annual income percentiles by Ryan Sleeper:

https://www.ryansleeper.com/household-income-rank/

Ryan uses a parameter as the household income value. As you change the annual household income selection, the comment next to the line graph changes as well. This is in contrast to calculated fields, as the data within calculated fields are NOT changeable by the dashboard viewer.

Watch my video tutorial on parameters to see how I modify the target reference line value in the IT tickets by location dataset.

Instructions:

Calculated fields:
- Watch the four calculated fields Davis videos
- Watch my video tutorial on calculated fields involving the Walmart dataset

Parameters:
- Watch my video tutorial on parameters involving the IT Tickets by location dataset

You should have read the content above, so all of these learning resources will prepare you to complete the assignment.

Deliverables:

1. Submit a Tableau packaged workbook that contains:
    - The Walmart dataset imported into the packaged workbook
    - The calculated fields I created in my video tutorial and in lecture
        - Order to Ship Time (Days)
        - Max Profit
        - Min Profit
        - Order ID new
        - Profit Categories

In the same Tableau packaged workbook, import the Titanic dataset and create the following calculated fields:

2. A calculated field that converts the numerical values of the 'Pclass' field to string categories
    a. 3 = 'Upper class'
    b. 2 = 'Middle class'
    c. 1 = 'Lower class'
3. Consider the yearBirth field in the titanic dataset. Create a calculated field of how many years it has been from each passenger's year of birth to present day. As an example, 200 years have passed for a passenger born in 1818.
4. Create a calculated field of a new passenger ID from these existing data fields, of the structure: Embarkment + Pclass + YearBirth
    a. Location of embarkment
    b. Pclass
    c. YearBirth

5. Consider the cabin data field. Create a calculated field where if the cabin is specified, leave it unchanged. However, if the cabin is not specified, the Cabin defaults to 0 instead of null. Use the ISNULL( ) function.

6. Create a calculated field of the maximum fare

7. Create a calculated field of the minimum fare

8. Create a calculcated field to determine if a passenger was alone in their trip on the Titanic. Consider the 'Sib Sp' and 'Parch' data fields:
   a. 'Sib Sp' shows numerically how many family members (siblings and spouse) are aboard with the passenger.
   b. 'Parch' shows numerically how many parents/children are aboard with the passenger.
   c. Therefore, a passenger was alone if BOTH of these data fields are 0. The logic is the following:
      i. If 'Sib Sp' = 0 and 'Parch' = 0, then 'Alone', else 'Not alone'

9. Consider the 'Name' data field. Notice how the first and last names of each passenger are together in one field, separated by a comma. Create two calculated fields derived from the 'Name' data field to break up the first and last names of the passengers into their own separate data fields, as shown below:

Abc	=Abc	=Abc
titanic_clean_.csv	Calculation	Calculation
Name	Last Name	First Name
Braund, Mr. Owen Harris	Braund	Mr. Owen Harris
Cumings, Mrs. John Bradley (Florence Briggs Thayer)	Cumings	Mrs. John Bradley (Florence Briggs Thayer)
Heikkinen, Miss. Laina	Heikkinen	Miss. Laina
Futrelle, Mrs. Jacques Heath (Lily May Peel)	Futrelle	Mrs. Jacques Heath (Lily May Peel)
Allen, Mr. William Henry	Allen	Mr. William Henry
Moran, Mr. James	Moran	Mr. James
McCarthy, Mr. Timothy J	McCarthy	Mr. Timothy J
Palsson, Master. Gosta Leonard	Palsson	Master. Gosta Leonard
Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	Johnson	Mrs. Oscar W (Elisabeth Vilhelmina Berg)
Nasser, Mrs. Nicholas (Adele Achem)	Nasser	Mrs. Nicholas (Adele Achem)

Look at the Tableau string function documentation to determine the best function for the task.

10. In the same Tableau packaged workbook, create a bar graph from existing data or one of your newly created calculated fields.
     a. Make sure to add color to the bar graph based on an arbitrary threshold value
     b. Create a parameter target reference line
     c. Insert this parameter into the color calculated field
     d. Place this bar graph in a dashboard view
     e. Show parameter controls.