Project 10

Visualize Benford's Law

Benford's Law is a mathematical law that states that the leading, or left-most, digit in many real-life data sources is distributed in a very specific manner. Specifically, the number 1 occurs as the leading digit about 30% of the time, and as numbers get larger they occur less frequently, with the number 9 occurring less than 5% of the time. When fraudsters are fabricating data, they may not know to create fake data that conforms to Benford's law, and in some cases this makes it possible to detect fake data or at least to create doubt about its veracity.

This article describes how to apply Benford's Law to Sales data, using the **Sample - Superstore** data source provided with Tableau Desktop.

The process requires you to do the following:

1. Create calculated fields to use in your view.
2. Set up the view.

The following sections break these procedures down into specific instructions.

Create calculated fields to use in your view

1. In the **Analysis** menu, select **Create Calculated Field** to open the calculation editor. Name the calculation **Leftmost Integer** and type or paste the following in the formula area:
   ```
   LEFT(STR([Sales]),1)
   ```
2. Create a second calculated field and name it **Benfords Law**. Type or paste the following in the formula area:
   ```
   LOG(INT([Leftmost Integer])+1)-LOG(INT([Leftmost Integer]))
   ```

Set up the view

1. From the **Data** pane, drag **Leftmost Integer** to **Columns**, and then drag **Orders(Count)** to **Rows**.
2. Click **CNT(Orders)** on Rows and choose **Quick Table Calculation** > **Percent of Total**.
   Your view now shows the distribution of first digits, and the size of the bars (decreasing from left to right) suggests that the data in this case conforms to Benford's law. But we can do more to frame the data by adding reference distributions.
3. From the **Data** pane, drag **Benfords Law** to Detail on the Marks card. Click **Benfords Law** on the Marks card and select **Measure** > **Minimum**.
4. Switch from the **Data** pane to the **Analytics** pane and drag **Distribution Band** into the view. Drop it on **Cell**.
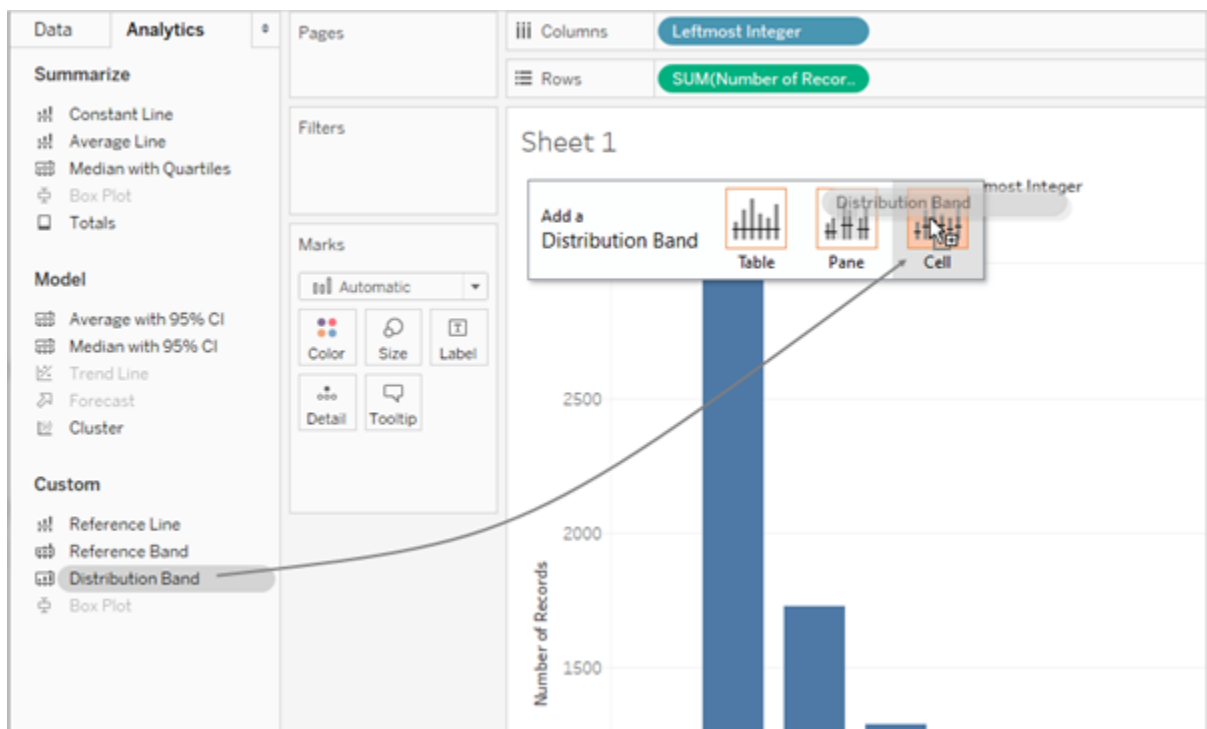
   Note: Distribution Bands are supported on web platforms starting with Tableau 10.2.
5. In the Edit Reference Line, Band, or Box dialog box, do the following:

a. Click in the Value field to view an additional set of options:
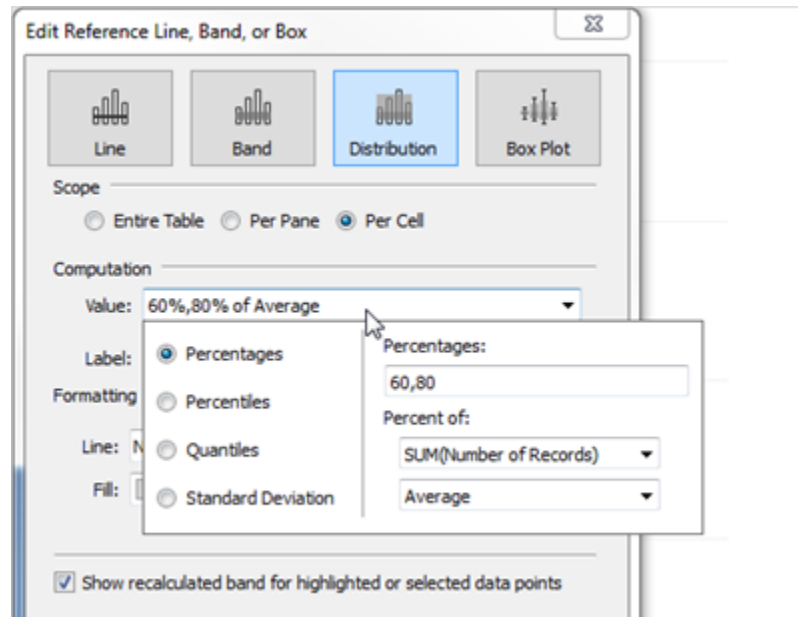
b. In the Percentages area, type `80,100,120`.
    This specifies that you want bands spanning from 80 to 100 percent, and from 100 to 120 percent. Next you will specify what value the percentages are referencing.

c. In the **Percent of** field, choose **MIN(Benfords Law)**.
    The Value field should now read `80%,100%,120% of Average Min. Benfords Law`.
    The remaining steps configure the appearance of the reference bands.

d. Set **Label** to **None**.

e. Set **Line** to the thinnest available line.

f. Choose **Fill Below**.

g. From **Fill**, select **Stoplight**.

h. Click **OK** to exit the Edit Reference Line, Band, or Box dialog box.

6. Click the toolbar button to display mark labels:

T

The finished view should look like this:

Sheet 1

Leftmost Integer
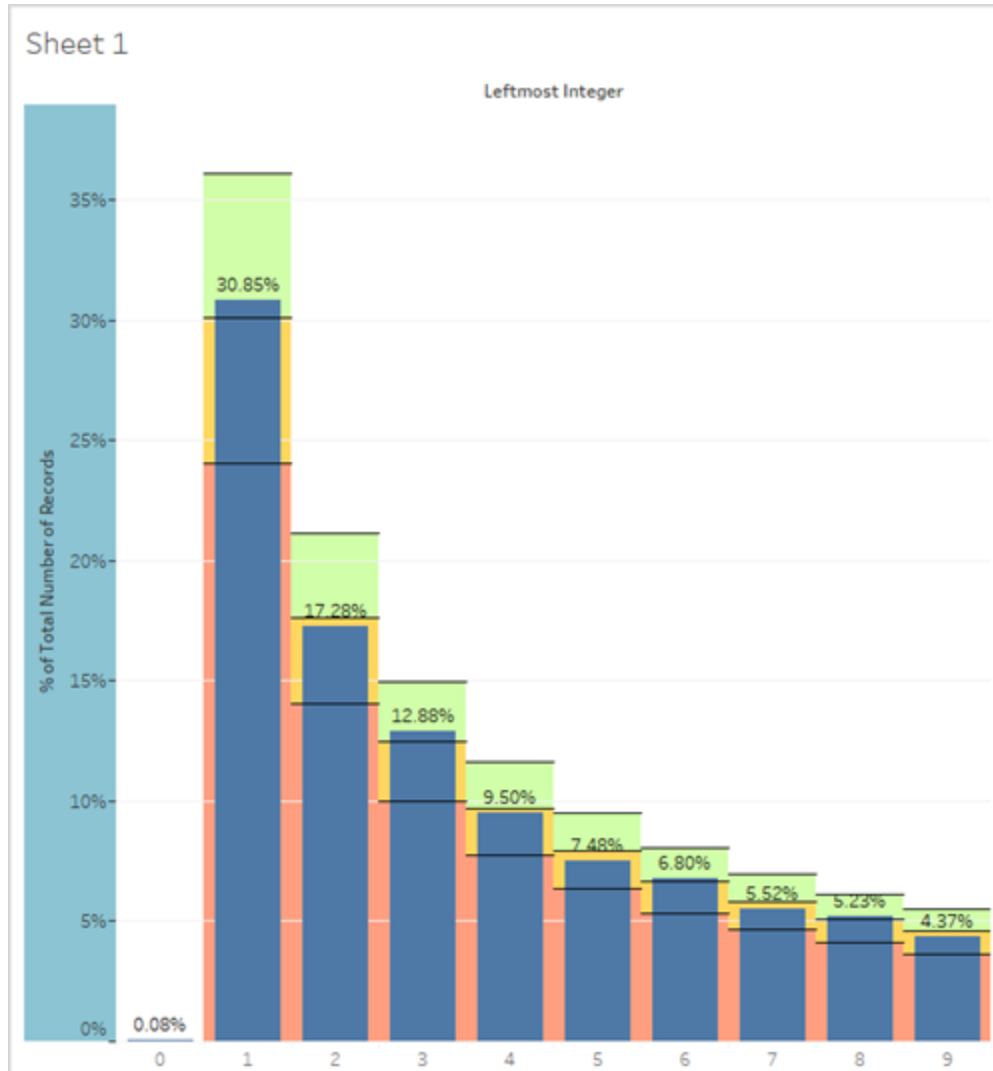
Even though Superstore is demo data, it's realistic as far as conforming to Benford's law. The blue bars that indicate actual percentages of initial digits align very well with the 100% value (that is, the line that separates the green zone from the yellow zone in the distribution bands) that shows expected Benford values in the view.