

Using Tableau Analytics

Tableau contains a wide range of analytical tools ranging from simple trend and reference lines through to data mining techniques such as cluster analysis and forecasting through time series analysis. It is also possible, given an appropriately-structured data source, to perform other data mining tasks such as association analysis.

Using Reference Lines

Reference lines highlight specific values or ranges on an axis based on either constant or computed values, e.g. for comparing monthly sales to an annual average. Most of these functions are calculated locally on the data that is already in the view. Reference line scope can be set to Entire Table, Per Pane, or Per Cell (Figure 1).



FIGURE 1: AVERAGE LINE WITH SCOPE OF ENTIRE TABLE (L), PER PANE (C), AND PER CELL (R)

Add a Reference Line

1. Right-click on the axis to be used for the reference line and choose **Add Reference Line**.
2. On the Add Reference Line box (Figure 2), choose **Line**.

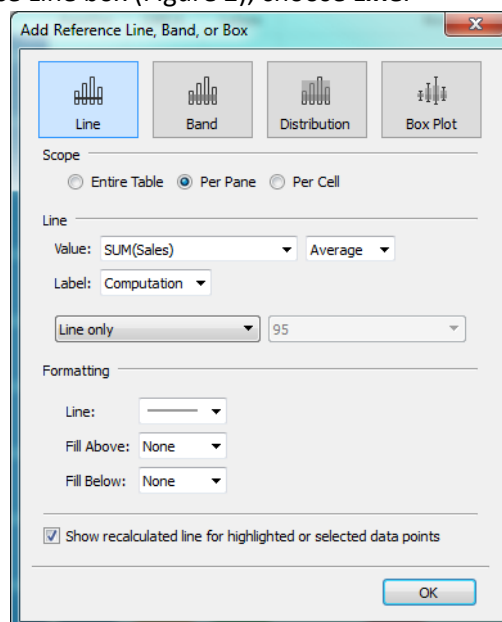


FIGURE 2: ADD REFERENCE LINE DIALOG BOX

3. Under **Scope**, specify **Entire Table**, **Per Pane**, or **Per Cell**.

- Under **Line**, set the **Value** computation to mark on the axis. You can select from the following options:

Option	Description
Average	Places a line at the average value along the axis
Constant	Places a line at the specified value on the axis
Maximum	Places a line at the maximum value
Median	Places a line at the median value
Minimum	Places a line at the minimum value
Sum	Places a line at the SUM of all values in either the cell, pane, or entire view, depending upon the scope selected
Total	Places a line at the aggregate of all of the values in either the cell, pane, or entire view, depending upon the scope selected

- Select how you want to label the line:

Option	Description
None	Does not include a label for the reference line
Value	Includes a label that corresponds to value on the axis
Computation	Displays and automatic label based on the computation and the measure selected
Custom	Adds a custom label to the text box (Use the menu to the right of the text box to insert values such as the computation or the value.)

- Format the line using the options shown on the dialog box, and click **OK**.

Reference bands

Reference bands show data that falls within a certain window value by displaying a shaded area between two constant or computed values (Figure 3). These values use the same calculations as reference lines, and most of the functions are calculated after the data is returned.

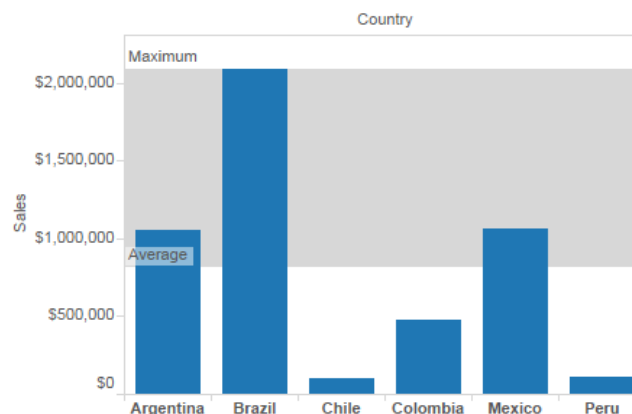


FIGURE 3: REFERENCE BAND WITH AVERAGE AND MAXIMUM VALUES

Add a Reference Band

- Right-click on the axis to be used for the reference line and choose **Add Reference Line**.
- On the Add Reference Line box choose **Band**.
- Under **Scope**, specify **Entire Table**, **Per Pane**, or **Per Cell**.
- Under **Band From**, set the **Value** computation to mark on the axis. You can select from the same options as were available for a reference line.
- Under **Band To**, set the **Value** computation to mark the other end of the band. You can select from the same options as were available for a reference line.

6. Select how you want to label the lines. You can select from the same options as were available for a reference line.
7. Formatting is available for the band, and for the lines that mark the two edges of the band.

Using the Analytics Pane

Common data analysis tools can be accessed from the Analytics Pane. This is a tab beside the Data Pane on the left of the workspace (Figure 4). This allows you to add any of these tools to your visualisation quickly by dragging and dropping the tool to the view. When you do this, Tableau will automatically open up the appropriate dialog box for you to add any necessary scope or related parameters. N.b. Not all analytical tools can be applied to all scopes, e.g. a constant reference line can only be applied to a whole table.

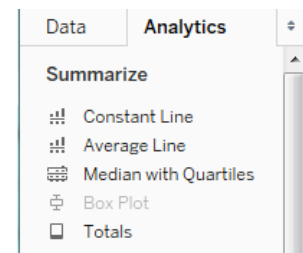


FIGURE 4: ANALYTICS PANE

If you add a reference line or other analytics to your view then if you select marks in the view, the analytics will automatically recalculate based upon the selected marks.

Tasks

1. Open the packaged workbook **Workshop 11.1 – Reference Line Starter.twbx**
2. Create a view (Figure 5) that does the following:
 - Includes a red line displayed with a value of \$500,000 across the view.
 - Has a label indicating that the line shows the quota.
 - Includes the value of the quota in the label.

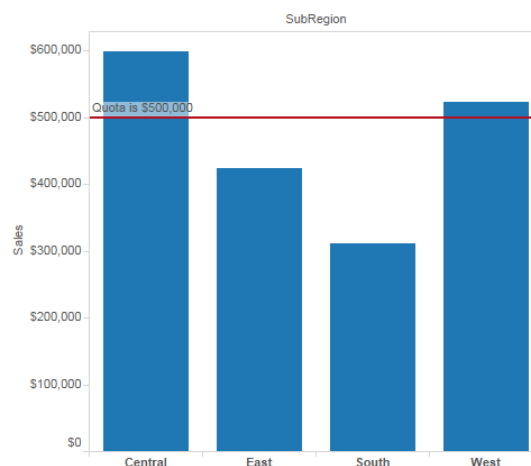


FIGURE 5: SALES BY SUBREGION WITH QUOTA LINE

3. Create a view (Figure 6) that does the following:
 - Shows Sales for each Customer Segment by SubRegion
 - Includes Average Sales by Segment as a reference line.

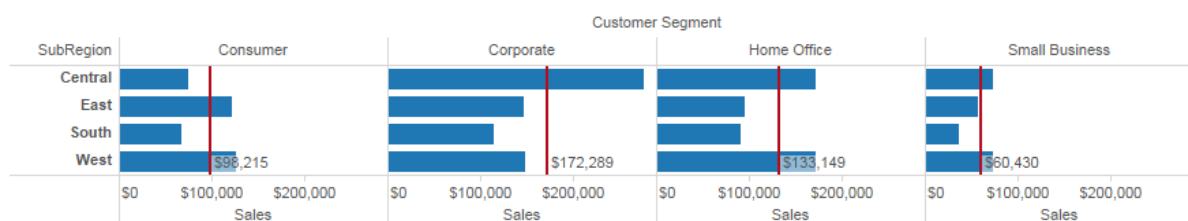


FIGURE 6: SALES BY SEGMENT AND SUBREGION

Trend Lines

Tableau's trend line feature allows you to construct interactive models that can be used to make predictions about your data, e.g. how profits will fare over the next financial period. To add trend lines to a view, both axes must contain a field that can be interpreted as a number. This includes date/time fields in both Discrete and Continuous forms, although generally trends are best calculated over long periods of time so Continuous date/time fields are better than Discrete ones.

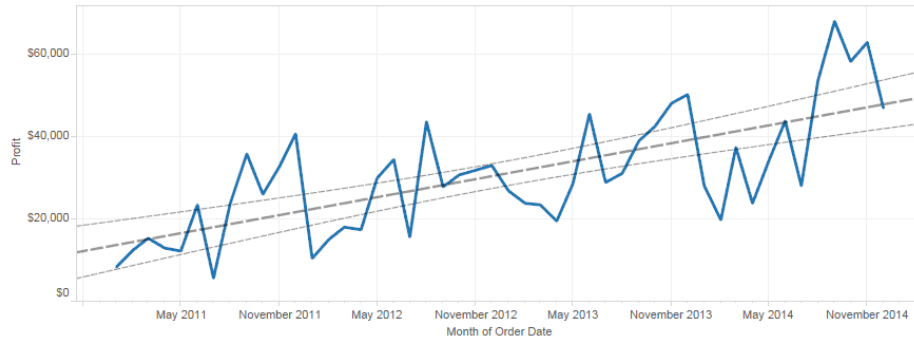


FIGURE 7: PROFIT VARIATION OVER TIME. LINEAR TREND WITH 95% CONFIDENCE LEVELS

By default, Tableau includes 95% confidence bands with its trend lines. These can be turned off from the Trend Lines Options box, accessed by right-clicking on the trend line and choosing **Edit Trend Lines**.

Add a trend line

1. Create a view with two numerical fields.
2. Right-click in the view and choose **Trend Lines**. This adds a linear trend line and adds a separate trend line for each colour if there is a discrete field on the **Colour** shelf.
3. Right-click the view again and select **Edit Trend Lines**.
4. In the Edit Trend Lines dialogue box, specify one of the following options:



Trend lines can also be accessed from the **Analytics** pane, under **Model**.

Tasks

1. Open the packaged workbook **Workshop 11.2 – Trend Lines with Small Multiples Starter.twbx**
2. Create a view (Figure 8) that does the following:
 - Shows a scatter plot with Sales and Profit
 - Filters data for United States only.
 - Disaggregates data to show all transactions

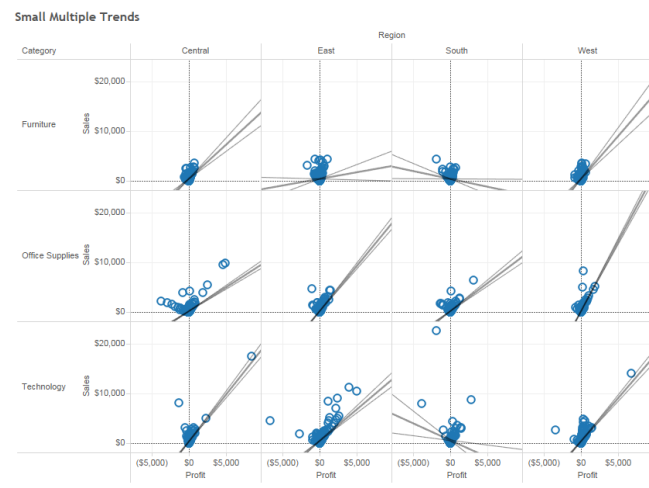


FIGURE 8: SMALL MULTIPLE TRENDS

Forecasting

Forecasting is a way to estimate future values based on continuous historical data. Tableau provides built-in statistical models to forecast your data including models that account for seasonality and trends. Forecasting can be turned on when viewing data over time.

Turn on Forecasting

Forecasting can be turned in two ways:

- Right-click in the view, select **Forecast**, and choose **Show Forecast**.
- From the **Analytics** pane, under **Model**, drag and drop **Forecast** into the view.

This will take the existing line graph and add an extension to it that covers the next major accounting period, showing the forecast values in a paler shade of the line colour (Figure 9). The forecast will assume the same basic shape as the annual shape of the source data but will take into account the trends identified within that source data. Forecast parameters can be changed by right-clicking on the view, selecting **Forecast**, and choosing **Forecast Options**.

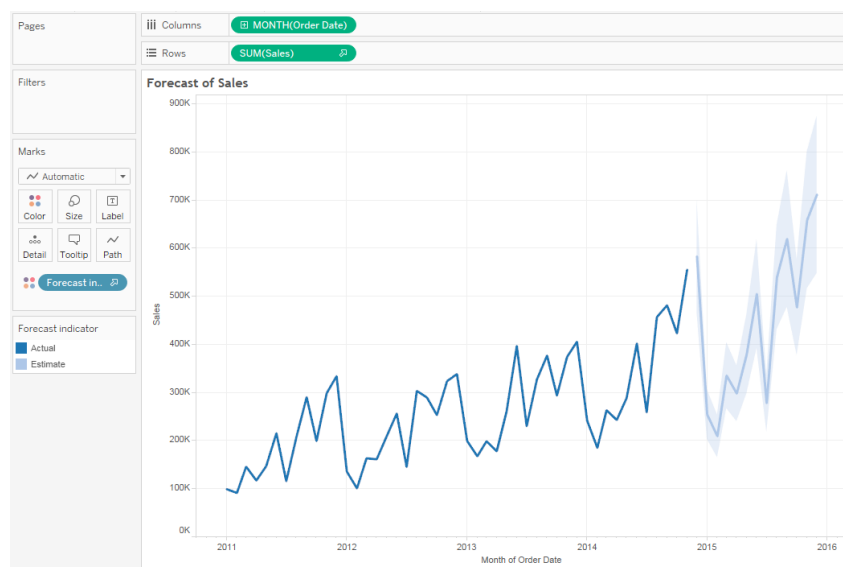


FIGURE 9: FORECAST OF SALES

Tasks

1. Open the packaged workbook **Workshop 11.3 – Trend Lines with Forecast Starter.twbx**
2. Create a view (Figure 10) that does the following:
 - Shows Continuous “MMM YYYY” on Columns and Profit on Rows.
 - Has a single-value, drop-down Quick Filter using Category.
 - Shows Forecast data for 2 years.
 - The marks are formatted with the Colour Blind 10 palette, using label options and mark size as shown.

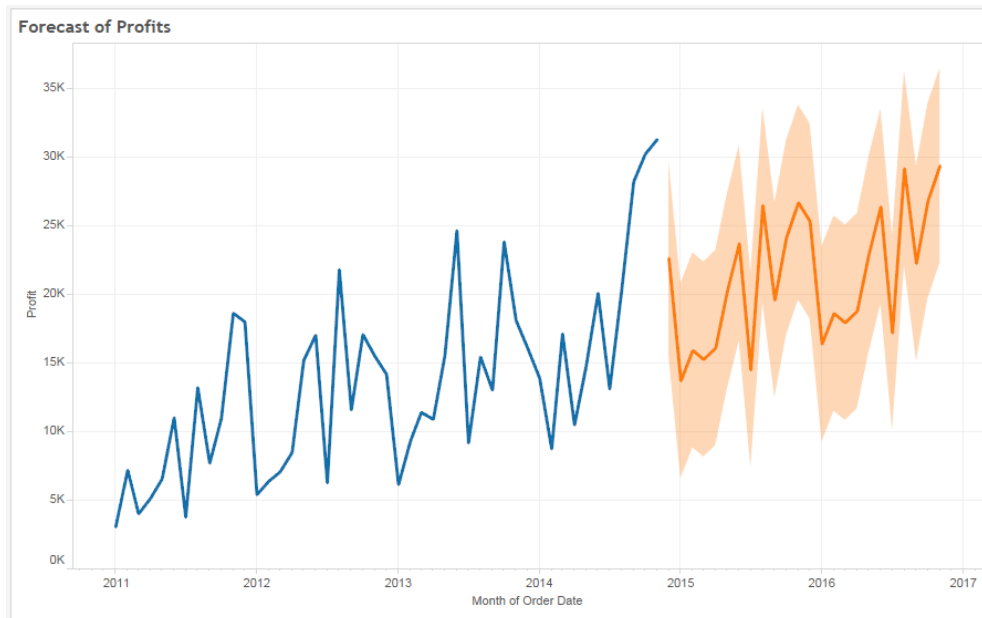


FIGURE 10: FORECAST OF PROFITS

Market Basket Analysis

Market basket analysis is a data mining technique, also known as association analysis, which allows you to look for relationships such as “How many people bought both A and B?” and “Which pairings of products are most commonly purchased?”

In order to do this, the data source must be appropriately structured to support this kind of data. In particular, the data source must be self-joined to itself on a unique identifier. We also need to define a second condition to identify the data field that contains the things we want to compare. Finally, we need to remove duplicate matches because we don’t want to include customers who only bought items from one category.

The following Market Basket Analysis is based on the data in the file **Workshop 11.4 – Market Basket Analysis Starter.twbx**. The aim is to identify which SubCategories are bought together.

Create a Market Basket Heat Map

1. In the **Data Source**, drag a second copy of the **Orders** table to the canvas.
2. In the Join dialog box, do the following:
 - Make sure that the **Inner** join type is selected.
 - From the drop down list under **Data Source**, select the joining attribute (**Customer Name**).
 - From the drop down list for the joined table (**Orders (copy)**), select the same joining attribute (**Customer Name**).
 - Make sure that the equal sign = is selected between them.

- Click the sheet tab to begin building your view. The data source will now contain records that show two sub-categories: the sub-category of an item bought and the sub-category of the other item bought with it.
- It is possible that a customer only bought items that were in the same category. We are not interested in this, so we need to add a filter to remove these. The filter will be based on a calculated field.
 - Right-click in the **Data** pane and select **Create > Calculated Field**
 - In the Calculated Field dialog box name the calculated field **Sub-Category Filter** and create the following calculation:

Sub-Category Filter

[Sub-Category] <> [Sub-Category (Orders)]

The calculation is valid. Sheets Affected ▼ Apply OK

This creates a calculated field shown with a T|F (Boolean) data type.

- From the **Data** pane, drag the new **Sub-Category Filter** to the **Filters** pane.
 - Show the filter, and set it to **True**.
- With the data source and filter now prepared, we can create the visualisation. We are interested in the number of records that include each pair of sub-categories.
 - Drag **Sub-Category** to the **Columns** shelf.
 - Drag **Sub-Category (Orders)** to the **Rows** shelf.
 - Drag **Number of Records** to the **Text** field on the **Marks** card.
 - Drag another copy of **Number of Records** to the **Color** field on the **Marks** card.
 - Set the drop-down on the **Marks** pane to **Square**.
 - Change the colour scheme so that there is a clear distinction between low and high values.

Figure 11 (below) shows the heat map generated by this method. The data is duplicated across the diagonal as there is no difference between the sub-category pairs of, for instance, Art/Labels and Labels/Art.

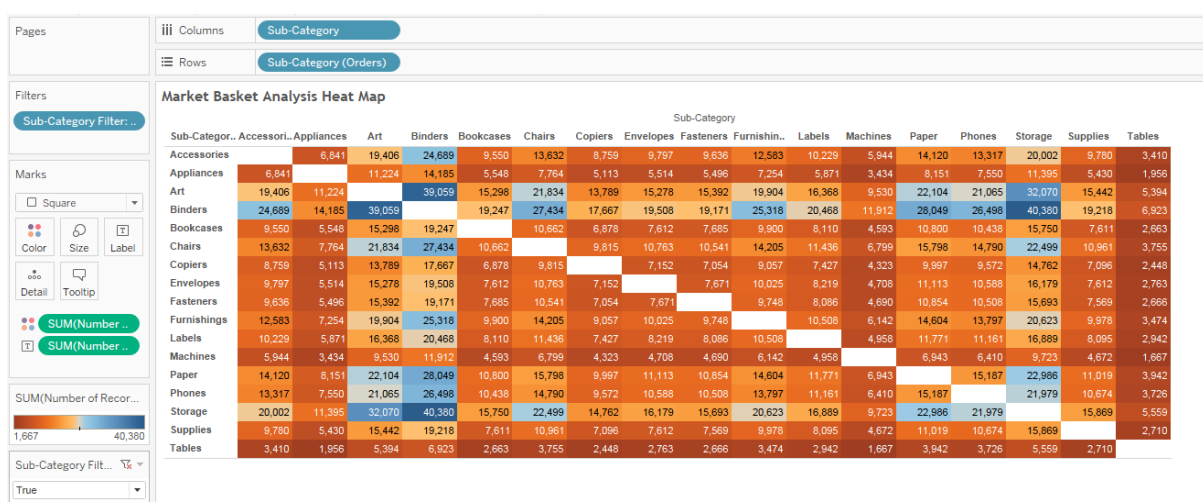


FIGURE 11: HEAT MAP SHOWING MARKET BASKET ANALYSIS OF SUB-CATEGORIES

The central diagonal is blank, as those sales where two items of the same sub-category were bought have been removed by the filter (and can be viewed by setting the filter to False). The colouring clearly shows that there are some combinations that appear together frequently: Art/Binders and Binders/Storage are the most obvious, but all of the blue squares show combinations that were frequently bought together.

Tasks

1. Using the file **Workshop 11.4 – Market Basket Analysis Starter.twbx**, work through the process of creating the heat map shown in Figure 11.
2. Create a view that shows how often items from the same sub-category were bought together (Figure 12).



FIGURE 12: SAME-SUB-CATEGORY HEAT MAP