

Getting Started with SAS Visual Analytics (2024.09)

Access SAS Visual Analytics by clicking the icon for the applications menu in the upper left corner and selecting **Explore and Visualize** from the menu.

Access Data

1. From the welcome window, you can open an existing report or create a new report. Select **New report**.
2. Visual Analytics consists of three main areas: the left pane with icons for working with the overall report, a canvas in which you build the report, and a right pane with icons for working with specific objects. The canvas displays one or more page tabs across the top of the report.
3. Reports use data that is loaded into memory. In the Data pane, there are options to either add data or import data. Click **Import data > Local Files** and notice that you can load files from the local operating system, such as Microsoft Excel or CSV files. Click **Cancel** to close the file window, and then click **Cancel** to close the **Import data** window.
4. Click **Add data** to select from tables that have been loaded into memory. Enter **HOME** in the **Search** field and click the **Start search** icon. Select the **HOME_EQUITY** table. On the right side of the Choose Data window, select **Columns**. Browse the column list to verify that this is the right data and then click **Add**.
5. The Data pane displays all the columns, or data items, in the **HOME_EQUITY** table. Data items are organized by type. Character columns and numeric columns with a date format appear as categories. Distinct counts appear next to each category. Numeric variables appear as measures.

Prepare Data

1. The Data pane is used to modify or create new data items. We are interested in analyzing the **Loan Status** column. The value *0* represents a loan in good standing, and *1* is a loan that has defaulted. Later, we'll explore the relationship between **Loan Status** and the other columns associated with the loan application.
2. The values of **Loan Status** represent a binary response rather than a measure. Click the double arrow next to **Loan Status** to edit the data item properties. Change **Name** to **Loan Status Code** and **Classification** to **Category**. This data item now appears in the Category group and includes two distinct values.
3. A consumer of this report might not know what the *0* and *1* values represent. We can create a new data item that assigns more descriptive labels to the raw loan status codes. Right-click **Loan Status Code** and select **New Custom Category**.
4. Enter **Loan Status** in the **Name** field. The first group name will be **Current**. In the **Values** field, enter **0** and press the Enter key. Click **Add group** and enter **Default** in the **Name** field. To show an alternative method for assigning values, click the **Choose values** icon, select **1**, click **Add**, and then click **OK**. Expand **Remaining Values** and select **Show as missing**, and then click **OK**. Now we can choose to use either the raw loan status codes or descriptive labels in our report.

5. There are many other ways to alter or create new data items. For example, we will eventually view the number of loan applications for each month and year. In the Data pane, expand the properties of **Loan Application Date** and click the **Edit** button next to the **Format** field. The current format displays the month name, a two-digit day, and a four-digit year. Instead, select the **MMYYYY** format so that the displayed values are the three-letter month and four-digit year. Click **OK**.
6. Notice that there is a column named **State**. To map values based on the US state, expand the properties for **State** and change **Classification** to **Geography**. In the Edit Geography Item window, change **Name or code context** to **US State Names**, and 100% of the data values are mapped to the appropriate geographical region. Click **OK**.
7. Finally, hover the cursor over each of the data items in the Measure category. Notice that the default aggregation for all measures, with the exception of Frequency, is SUM. This means that as we build visualizations that include these measures, the summarized values will be totals. Because each row in the data is a separate loan, it makes more sense to look at the average for all measures. To change the aggregation method, select the check box for each measure. Or select at least one measure data item and select the **More** menu to the right of the New data item button. Choose **Select > Select measures with same aggregation**. Right-click over the selected data items and change **Aggregation** to **Average**.

Explore Data

1. To begin by examining the raw data, select the Objects pane on the left side and then drag **List Table** onto the canvas. On the right side of the canvas, select the Data Roles pane. Select **Add** next to **Columns** and then choose **Select all > Apply**. Drag **Loan Status** to the top of the list so that it's the first column.
2. You can scroll to view the different columns and values, or click on any of the columns, such as **State**, to sort the table in ascending or descending order.
3. If you right-click a column, there are several additional options. Right-click **Age of Oldest Credit Line** and select **Add cell visualization > Bar**. Right-click **Amount of Loan Request** and select **Add cell visualization > Heat map**. You can customize cell visualization in the Options pane.
4. Double-click **Page1** at the top of the report and enter **Home Equity Data** to name the report page.
5. Next, we'll create a collection of graphs to explore the data. At the top of the canvas, click the plus sign to create a new page. From the Data pane, drag **Loan Status** to the canvas. An autochart is generated, which means that Visual Analytics selects an appropriate graph based on the data type and values. In this example, the autochart is a bar chart.
6. The autochart can be easily changed to another visualization. Access the Object menu in the upper right corner of the bar chart and select **Change bar chart to > Pie chart**. In the Options pane, expand the Pie section. Under the Data Labels section, select **Percent of total values** and increase the **Text style** size to **12**. The chart indicates that approximately 20% of the loans default.
7. In the Data pane, select **Amount of Loan Request** and drag it to the right side of the canvas. The autochart is a histogram, which is a great way to view the distribution of the values.
8. Next, drag **Loan Application Date** below the histogram. This time, the autochart is a line plot that tracks the frequency of loan applications over time. This downward trend is expected because homeowners were more likely to refinance rather than apply for home equity loans during these years of declining interest rates.

9. You don't have to always start with an autochart to create a report item. From the Objects pane, you can select from a long list of options. Drag **Bar Chart** below the pie chart. You can use either the Assign Data button or the Data Roles pane to build the graph. Add **Job Category** to the Category role. Under Measure, remove **Frequency** and then add **Amount Due on Existing Mortgage** and **Value of Current Property**. The length of each bar represents the average measure because we previously changed the aggregation.
10. You can further customize the graph by selecting the Options pane. Expand the **Bar** category and change the direction to **vertical**.
11. These static images are informative, but it would be helpful to see how they differ by region. In the Data pane, right-click **Region** and select **Add as a page control**. Now you can click each Region button to see how the graphs change.
12. This page is a good first step toward understanding our home equity data. Double-click the **Page 2** label and enter **Loan Exploration**.
13. Let's dive a bit deeper into the geography data. Create another new page and name it **Geo Analysis**. Select **State** in the Data pane and drag it to the canvas. A map is created. Select the Roles pane on the right and add **Frequency** to the Size role and **Amount of Loan Request** to the Color role.
14. Return to the Data pane and select **New data item > Hierarchy**. A hierarchy enables you to drill into categories within your data. Add the largest grouping first, which is **Region** with four values. Then add **Division**, followed by **State**. Notice that the **Name** field automatically includes the hierarchy values. Click **OK**.
15. Drag the **Region – Division – State** hierarchy to the canvas, and Visual Analytics creates a bar chart. You can click on each bar label to drill into **Region** and **Division** to see the corresponding frequency of loans. You can also select **Actions** to link the map and the bar chart. In the Actions pane, select the check box for **Geo coordinate – State 1**. Each time that you drill into a particular bar on the Region chart, the map updates based on the filtered data.

Explain and Analyze Data

1. We've explored this historical loan data, but the next question is: How does the data help to predict **Loan Status**? Return to the Data pane, and then right-click **Loan Status** and select **Explain > Explain on new page**. Visual Analytics generates an automated explanation that indicates that the three factors most related to **Loan Status** are **Debt to Income Ratio**, **Number of Delinquent Credit Lines**, and **Value of Current Property**. Name this page **Predict Loan Status**.
2. Visual Analytics enables you to create a preliminary model to predict **Loan Status**. Access the Object menu in the upper right corner of the page and select **Duplicate as > Gradient boosting**. This model enables us to continue in the analytics life cycle. The Options pane enables you to customize and fine-tune the model.

Note: The Object menu for gradient boosting provides a Create Pipeline option. This hands the model off to Model Studio for deeper analysis and model comparison.
3. Another great feature about this model is that it can derive predicted values and probabilities. From the Object menu for gradient boosting, select **Derive Predicted** and then click **OK** to add Predicted and Probability data items.

4. The new data items are available in the Data pane and can be included in the Home Equity Data list report. Click the **Home Equity Data** page and select the List table. In the Data Roles pane, click **Add** next to **Columns**. Click **Select all** to choose both the Predicted and Probability data items and then click **Apply** to include the columns in the List table. If a new row is added to the data without a determined loan outcome, the model predicts the result.

Save and View the Report

1. Click the **Save** icon and select **My Folder**. Name the report **Home Equity Report** and click **Save**. The report is currently in Edit mode, so click the **View report** button in the upper left corner. Select the Loan Exploration page and confirm that all the charts update as you select each region. Select the Geo Analysis page and verify that the map updates as you select a bar representing a particular geographic area.
2. In the upper right corner of Visual Analytics, select **Opened reports** and click **Close all reports**.
3. We've packed a lot into this report, but we've just scratched the surface of what is possible in SAS Visual Analytics. If you would like to view other completed sample reports with even more objects and actions, navigate to **SAS Content > Products > SAS Visual Analytics > Samples**.
4. To learn more, select the **Help** menu and visit the Learning Center. There are several resources available for you to expand your Visual Analytics expertise.