

# Exercise: Preparing Data Using SAS® Data Studio

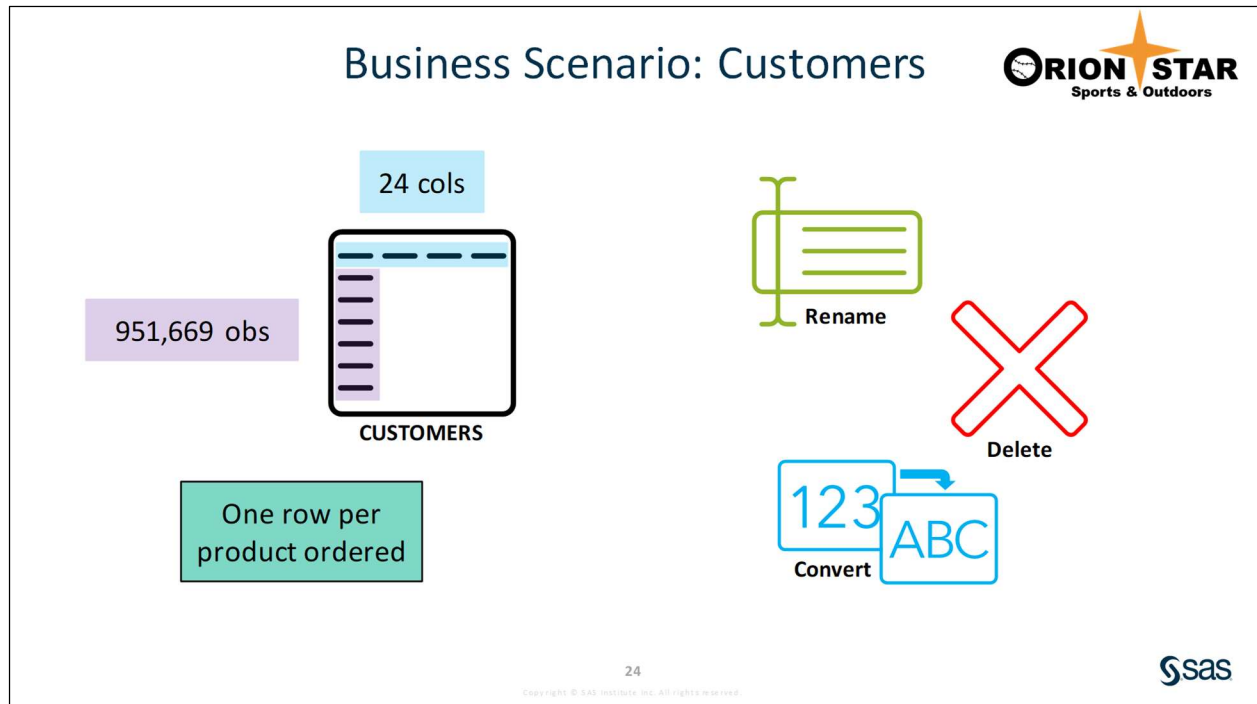
Login: : <https://vle.sas.com/vfl>

**2.2 Transforming Data Using SAS Data Studio .....**

**Preparing Data.....**

**Exercise 2B: Pages 19-20.....**

## 2.2 Transforming Data Using SAS Data Studio



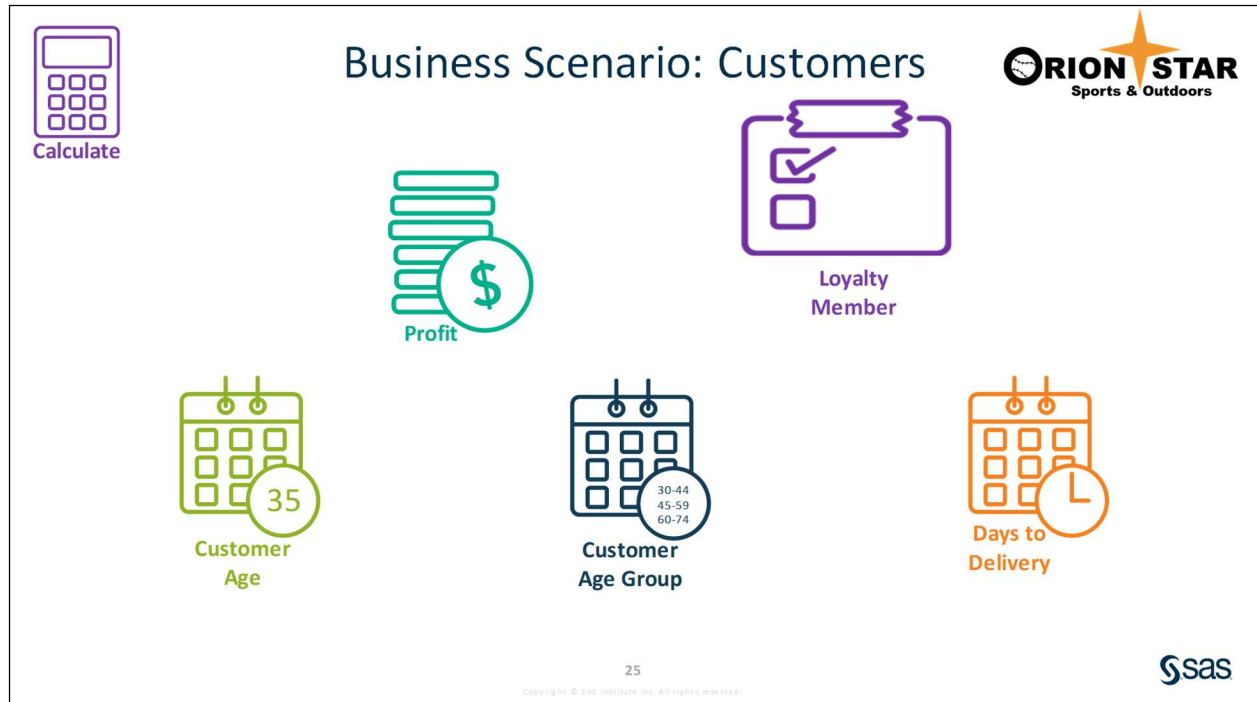
Here is a description of the categories in the **CUSTOMERS** table:

| Name                                     | Description                                   | Distinct Counts | Cleansing?                          |
|------------------------------------------|-----------------------------------------------|-----------------|-------------------------------------|
| <b>City Name</b>                         | City where customer resides                   | 10,507          |                                     |
| <b>Continent Name</b>                    | Continent where customer resides              | 5               |                                     |
| <b>Customer Birth Date</b>               | Date on which customer was born               | 4,368           |                                     |
| <b>Customer Country</b>                  | Country where customer resides                | 47              |                                     |
| <b>Customer Group Name</b>               | Loyalty member group                          | 3               |                                     |
| <b>Customer Name</b>                     | Name of customer                              | 67,793          |                                     |
| <b>Customer Type Name</b>                | Loyalty member level                          | 7               |                                     |
| <b>Customer ID</b>                       | Unique identifier for customer                | 68,300          | Change data type. Trim white space. |
| <b>Date Order was Delivered</b>          | Date on which order was delivered to customer | 1,840           |                                     |
| <b>Date Order was placed by Customer</b> | Date on which order was placed by customer    | 1,825           |                                     |
| <b>Order Type</b>                        | Method in which order was placed              | 3               |                                     |
| <b>Order ID</b>                          | Unique identifier for order                   | 747,953         | Change data type. Trim white space. |
| <b>Postal code</b>                       | Postal code where customer resides            | 19,340          |                                     |
| <b>State Name</b>                        | State or province where customer resides      | 272             |                                     |

**Note:** By default, all datetime variables have a format of Date with Month Name.

Here is a description of the measures in the **CUSTOMERS** table:

| Name                                                    | Description                               | Minimum | Maximum  | Average | Number Missing |
|---------------------------------------------------------|-------------------------------------------|---------|----------|---------|----------------|
| <b>Cost</b>                                             | Cost per unit                             | 0.40    | 1,583.60 | 77.76   | 0              |
| <b>Discount in percent of Normal Total Retail Price</b> | Discount (% of normal total retail price) | 0.30    | 0.60     | 0.38    | 942,517        |
| <b>Quantity Ordered</b>                                 | Quantity ordered                          | 1.00    | 10.00    | 1.68    | 0              |
| <b>Retail Price</b>                                     | Total revenue                             | 0.63    | 9,385.80 | 139.96  | 0              |



**Note:** The following data items are not needed for the analysis and will be removed: **Customer Name** (after the split), **Street Name**, **Street ID**, **Date Order was Delivered** (after calculation), **Employee ID**, **Product ID**, **xyContinentLat**, and **xyContinentLon**.

**Note:** **Profit** and **Days to Delivery** are calculated in SAS Data Studio. **Customer Age** is calculated in Visual Analytics using the Now operator, so the age updates every time that the report is opened. Both **Customer Age Group** and **Loyalty Member** can be calculated in SAS Data Studio using custom code. These data items can also be calculated in SAS Visual Analytics, which you see in a later lesson.

**Profit** is calculated as **Retail Price (Total Revenue) – Cost (Unit Cost) \* Quantity Ordered**.

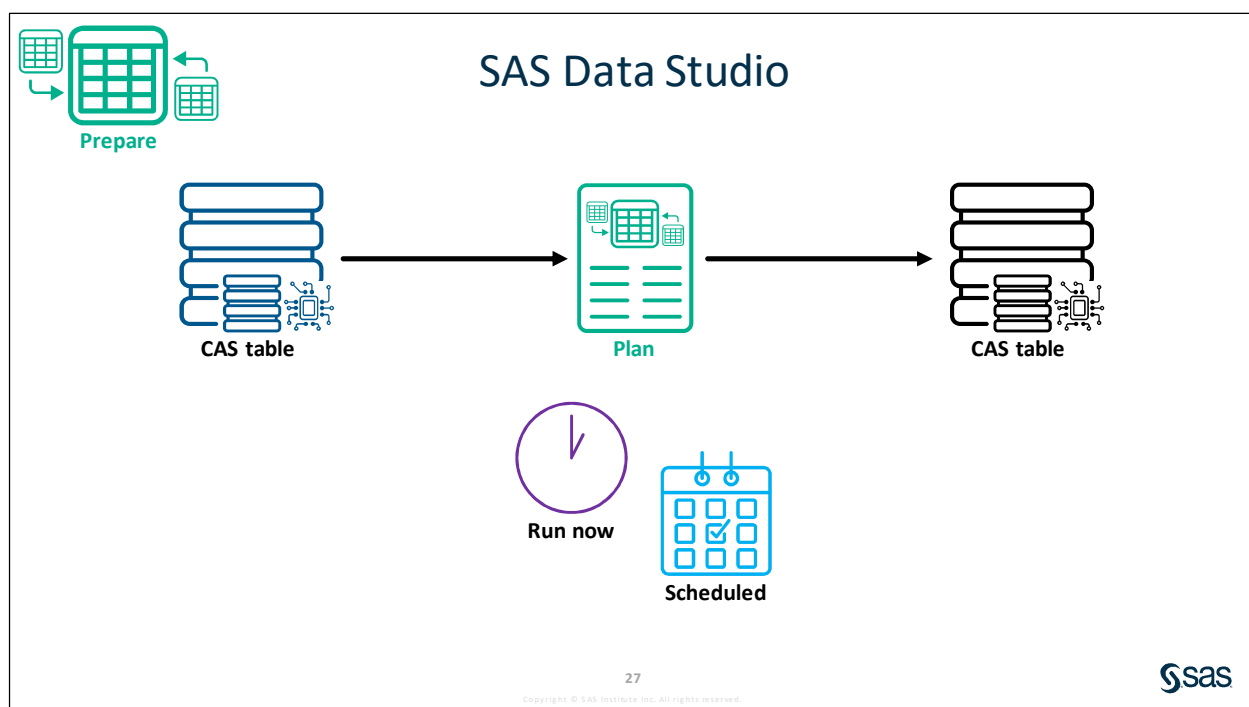
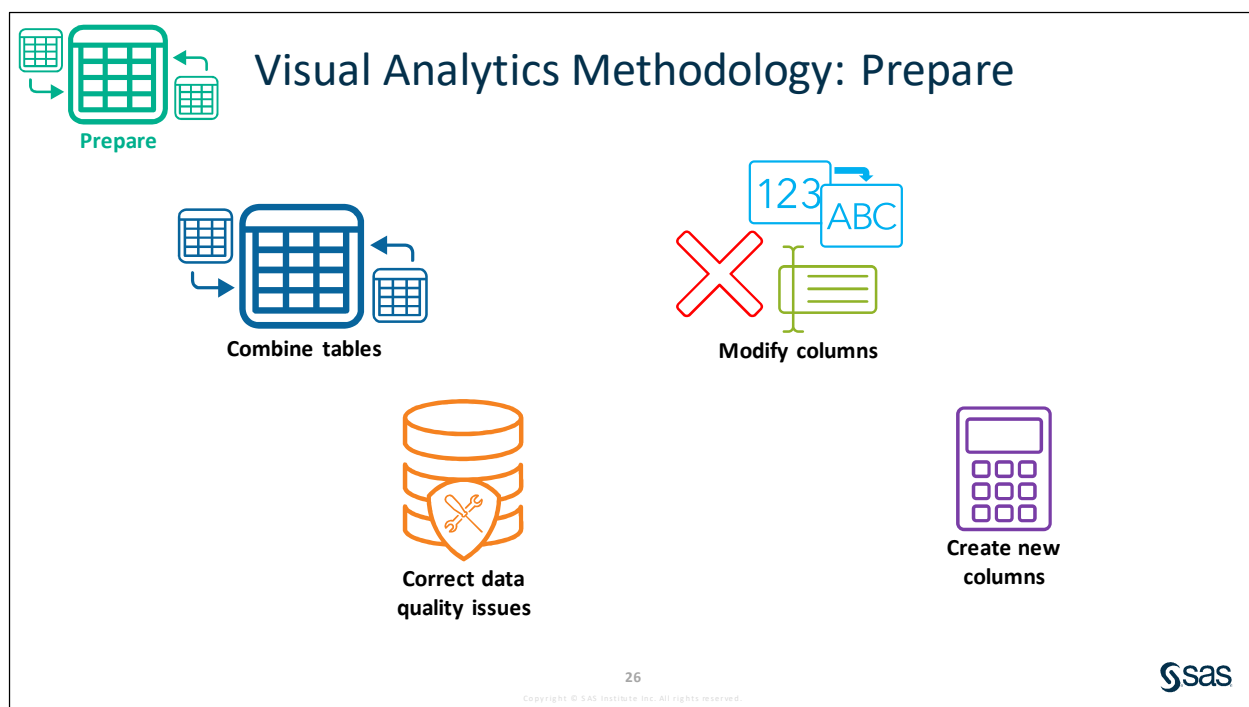
**Days to Delivery** is calculated as **Date Order was Delivered – Date Order** (that is, the date on which the order was placed by the customer).

**Customer Age** is calculated as **(Today's Date – Customer Birth Date)/365.25**.

**Customer Age Group** uses **Customer Age** to create ranges of ages.

**Loyalty Member** is Yes if last two digits are 99, and No if the last two digits are 00.

**Note:** The actual calculations are more complex and are discussed in more detail in later sections.



In SAS Data Studio, you can view metrics about your tables and columns and perform data transformations on your data by creating plans. The data source must be loaded to CAS before data can be prepared in Data Studio. If you open a table that is not loaded, it is automatically loaded to the CAS server.

## SAS Data Studio: Table Details

Table Profile Metadata

Date profiled: Dec 17, 2019 09:18 AM

| Column                   | Unique         | Null     |
|--------------------------|----------------|----------|
| City                     | 1.10% (10,507) |          |
| Continent                | <0.01% (5)     |          |
| Cost                     | 0.20% (1,883)  | 77.76    |
| CustomerCount<br>ryLabel | <0.01% (47)    |          |
| Customer_Birth<br>Date   | 0.46% (4,368)  | 3,867.29 |
| Customer_Grou<br>p       | <0.01% (3)     |          |

Table Profile Metadata

CUSTOMERS

The session table is not current to the plan. Run the plan to update the table.

Result rows: 100

| City       | Con...  | Post... | Stat...    | Stre...     | xyC...   | xyC...  | Emp...  | Stre...  |
|------------|---------|---------|------------|-------------|----------|---------|---------|----------|
| Leinster   | Oceania | 6437    | Western... | Bembo...    | -18.3... | 138.... | 9999... | 16001... |
| Berowra    | Oceania | 2081    | New So...  | Circuit ... | -18.3... | 138.... | 9999... | 16001... |
| Berowra    | Oceania | 2081    | New So...  | Circuit ... | -18.3... | 138.... | 9999... | 16001... |
| Northbr... | Oceania | 2063    | New So...  | Exhibiti... | -18.3... | 138.... | 9999... | 16001... |

Table Profile Metadata

Filter

| # | Name           | Label          | Type   | Raw Length |
|---|----------------|----------------|--------|------------|
| 1 | City           | City Name      | char   | 45         |
| 2 | Continent      | Continent Name | char   | 45         |
| 3 | Postal_Code    | Postal code    | char   | 15         |
| 4 | State_Province | State Name     | char   | 38         |
| 5 | Street_Name    | Name of Street | char   | 68         |
| 6 | xyContinentLat | xyContinentLat | double | 8          |
| 7 | xyContinentLon | xyContinentLon | double | 8          |
| 8 | Employee_ID    | Employee ID    | double | 8          |

28


Copyright © SAS Institute Inc. All rights reserved.

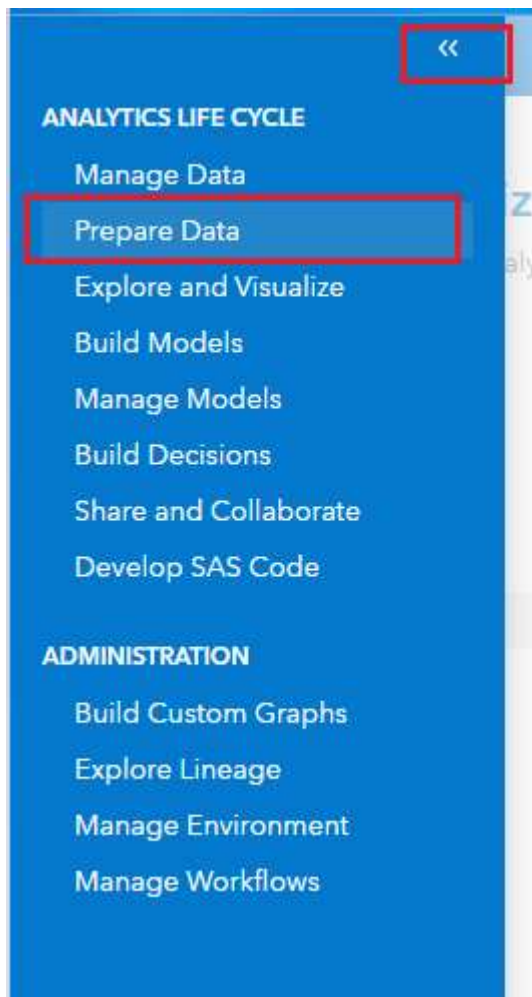
**Note:** Only the first 300 columns are displayed in the workspace. However, this does not affect your ability to work with all the data in the table. Any changes that are made apply to the entire table, not just the columns that are displayed.



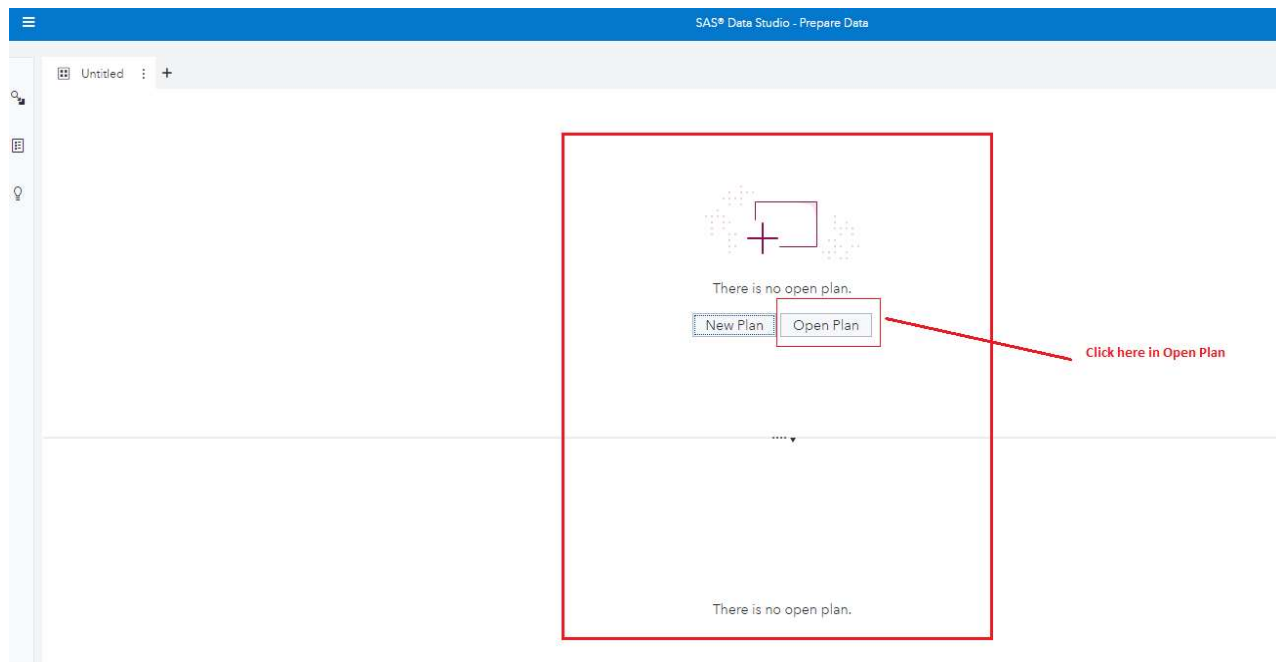
## Preparing Data

This demonstration illustrates how to view table and column profile information and view plan actions in SAS Data Studio.

1. From the browser window, Go to URL: <https://vle.sas.com/vfi>
2. Sign in.
1. Click  in the top left corner of SAS Drive. select **Prepare Data** as below:



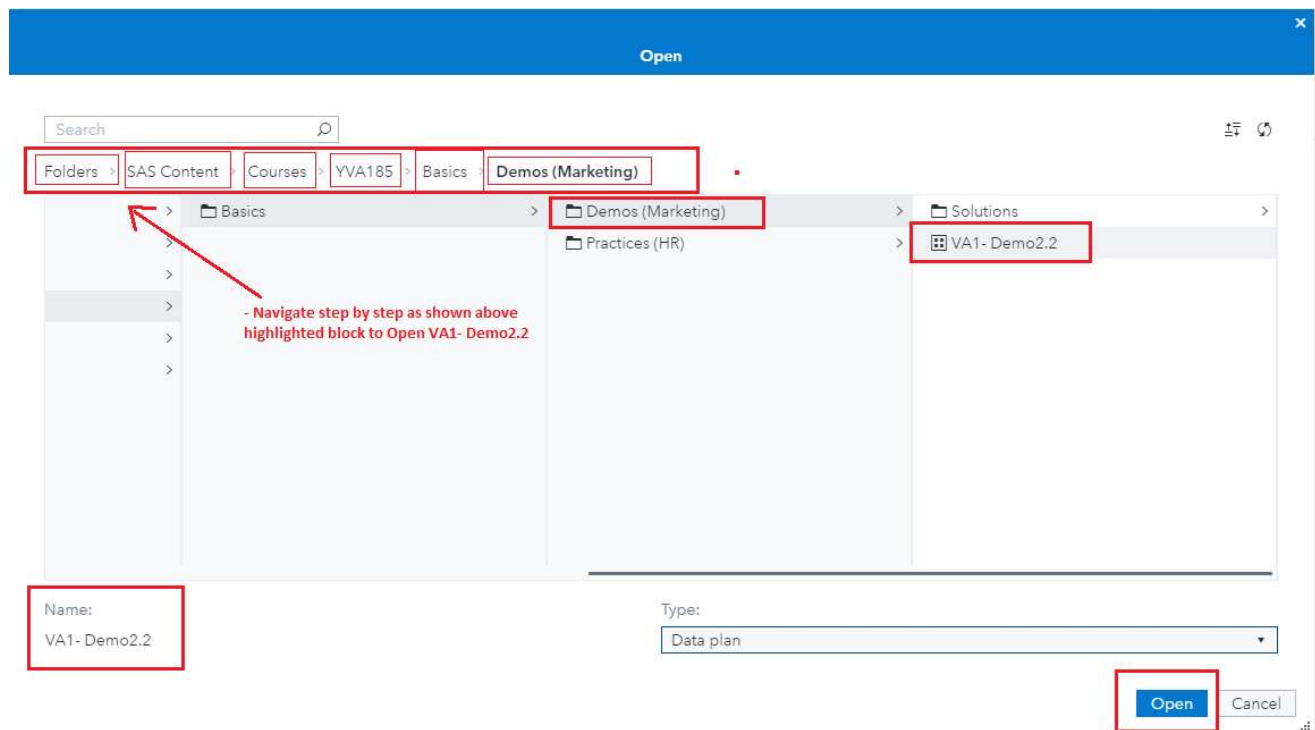
SAS Data Studio appears. You can open a table by creating a new plan or by opening an existing plan.



2. Click **Open Plan**.

- In the Open window, navigate to the **Folders/SAS Content/Courses/YVA185/Basics/Demos(Marketing)/ VA1-Demo2.2**
- Then click **Open** below to open the plan.

The steps of the plan are executed when the plan is opened.



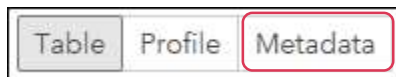


c. Once you click Open, your screen should resemble like below:

The screenshot shows the SAS Data Studio - Prepare Data interface. On the left is a sidebar with 'Add Transform' options. The main workspace displays the 'CUSTOMERS' table with columns: City, Continent, Postal Code, State/Province, Street Name, xyContinentLat, Employee ID, Street Address, Order Date, Delivery Date, and Days to Delivery. Below the table, a transformation step '8. Calculated Column' is shown, with the formula 'Delivery Date - Order Date' and the option to 'Create new column' named 'Days to Delivery'.

3. View metadata, profile, and table information for the source table.

a. In the top pane, click **Metadata**.



The Metadata view shows a list of columns (and their types) for the source table.

The screenshot shows the 'Metadata' view for the 'CUSTOMERS' table. It displays a table with columns: #, Name, Label, Type, Raw Length, Formatted Length, and Format. Below this, the '8. Calculated Column' transformation is visible, showing the formula 'Delivery Date - Order Date' and the option to 'Create new column' named 'Days to Delivery'.

| # | Name           | Label          | Type   | Raw Length | Formatted Length | Format |
|---|----------------|----------------|--------|------------|------------------|--------|
| 1 | City           | City Name      | char   | 90         | 90               |        |
| 2 | Continent      | Continent Name | char   | 90         | 90               |        |
| 3 | Postal_Code    | Postal code    | char   | 30         | 30               |        |
| 4 | State_Province | State Name     | char   | 76         | 76               |        |
| 5 | Street_Name    | Name of Street | char   | 136        | 136              |        |
| 6 | xyContinentLat | xyContinentLat | double | 8          | 15               | F      |

b. In the top pane, click **Profile** (beside **Metadata Tab**). Your screen will resemble like below:

| Column            | Unique          | Null         | Blank | Pattern Count | Mean      | Median    | Mode      | Standa... ↑ | Standa... | Minimum       | Maximum       | Data Type |
|-------------------|-----------------|--------------|-------|---------------|-----------|-----------|-----------|-------------|-----------|---------------|---------------|-----------|
| Discount          | 0.04% (4)       | 99.04% (942) |       |               | 0.38      | 0.40      | 0.30      | 0.08        | 0.00      | 0.30          | 0.60          | double    |
| Quantity          | <0.01% (10)     |              |       |               | 1.68      | 1.00      | 1.00      | 0.90        | 0.00      | 1.00          | 10.00         | double    |
| xyContinentLat    | <0.01% (5)      |              |       |               | 43.73     | 48.69     | 48.69     | 16.24       | 0.02      | -18.31        | 48.69         | double    |
| xyContinentLon    | <0.01% (5)      |              |       |               | -9.71     | 9.14      | 9.14      | 60.78       | 0.06      | -100.55       | 138.52        | double    |
| Cost              | 0.20% (1,883)   |              |       |               | 77.76     | 50.50     | 45.60     | 85.28       | 0.09      | 0.40          | 1,583.60      | double    |
| RetailPrice       | 0.79% (7,515)   |              |       |               | 139.96    | 84.10     | 55.40     | 185.73      | 0.19      | 0.63          | 9,385.80      | double    |
| Delivery_Date     | 0.19% (1,840)   |              |       |               | 19,95...  | 19,95...  | 19,78...  | 519.79      | 0.53      | 18,994.00     | 20,838.00     | double    |
| Order_Date        | 0.19% (1,825)   |              |       |               | 19,95...  | 19,95...  | 19,78...  | 519.80      | 0.53      | 18,993.00     | 20,818.00     | double    |
| Customer_Birth... | 0.46% (4,368)   |              |       |               | 3,867.29  | 4,812.00  | 1,277.00  | 5,977.96    | 6.13      | -8,035.00     | 12,050.00     | double    |
| Customer_ID       | 7.18% (68,300)  |              |       |               | 4,544,... | 4,414,... | 9,117,... | 2,721,4...  | 2,789.66  | 100.00        | 9,425,499.00  | double    |
| Order_ID          | 78.59% (747,... |              |       |               | 1,236,... | 1,236,... |           | 4,273,2...  | 4,380.37  | 1,230,000,... | 1,244,337,... | double    |
| Employee_ID       | 0.07% (648)     |              |       |               | 24,85...  | 120,9...  | 99,99...  | 43,113,...  | 44,19...  | 120,121.00    | 99,999,99...  | double    |
| Street_ID         | 3.90% (37,144)  |              |       |               | 6,272,... | 6,300,... | 8,300,... | 2,762,7...  | 2,831,... | 1,400,100,... | 9,700,100,... | double    |
| Product_ID        | 0.33% (3,151)   |              |       |               | 229,0...  | 230,1...  | 240,1...  | 10,647,...  | 10,91...  | 210,200,1...  | 240,800,2...  | double    |

c. If necessary, click **Run Profile** to execute the profile.

| Column          | Unique         | Null        | Blank | Pattern Count | Mean     | Median   | Mo...         |
|-----------------|----------------|-------------|-------|---------------|----------|----------|---------------|
| City            | 1.10% (10,506) | 0.01% (118) |       | 731           |          |          | London        |
| Continent       | <0.01% (5)     |             |       | 4             |          |          | Europe        |
| Cost            | 0.20% (1,883)  |             |       |               | 77.76    | 50.50    | 45.60         |
| CustomerCou...  | <0.01% (47)    |             |       | 15            |          |          | United States |
| Customer_Bir... | 0.46% (4,368)  |             |       |               | 3,867.29 | 4,812.00 | 1,277.00      |
| Customer_Gr...  | <0.01% (3)     |             |       | 3             |          |          | Orion         |

Basic profile metrics (Unique, Mean, Standard Deviation, Standard Error, Minimum, Maximum, Data Type, and Data Length) appear for all the columns in the **CUSTOMERS** data source.

**Note:** Advanced profile metrics (Null, Blank, Pattern Count, Median, Mode, Actual Type, Minimum Length, Maximum Length, Ordinal Position, Primary Key Candidate, and Non-null Count) appear if SAS Data Preparation is licensed at your site.

- d. In the top pane, click **Table**. You will see a message displaying under **CUSTOMERS**, saying session table is not current to the plan. Therefore, we will run the plan to update the table.
- e. To do that click on **Run** button on bottom right to update the table.

Message displaying to update the session table to the current plan.

**CUSTOMERS**  
The session table is not current to the plan. Run the plan to update the table.

| City       | Con...     | Post... | Stat...    | Stre...     | xyC...   | xyC...  | Emp...  | Stre...  | Ord...    | Deli...   | Ord...   | Prod...   | Qua... | Reta... |
|------------|------------|---------|------------|-------------|----------|---------|---------|----------|-----------|-----------|----------|-----------|--------|---------|
| Leinster   | Oceania    | 6437    | Western... | Bembo...    | -18.3... | 138...  | 9999... | 16001... | 01JAN2... | 07JAN2... | 12300... | 220101... | 3      | \$28... |
| Berowra    | Oceania    | 2081    | New So...  | Circuit...  | -18.3... | 138...  | 9999... | 16001... | 01JAN2... | 04JAN2... | 12300... | 220100... | 2      | \$11... |
| Berowra    | Oceania    | 2081    | New So...  | Circuit...  | -18.3... | 138...  | 9999... | 16001... | 01JAN2... | 04JAN2... | 12300... | 220101... | 2      | \$41... |
| Northbr... | Oceania    | 2063    | New So...  | Exhibiti... | -18.3... | 138...  | 9999... | 16001... | 01JAN2... | 03JAN2... | 12300... | 240100... | 1      | \$35... |
| Montréal   | North A... |         | Quebec     | rue Bea...  | 46.0...  | -100... | 9999... | 26001... | 01JAN2... | 04JAN2... | 12300... | 240200... | 1      | \$24... |

8. Calculated Column

1 Delivery\_Date - Order\_Date

Options for new columns

Run

- f. After you run the CUSTOMERS table you will see the below updated table to the current plan.

After running the table, Customer table is updated to current plan and showing session in brackets.

**CUSTOMERS (session)**  
The session table is current to the plan.

| Profit  | Days to Delivery | Total Revenue | Unit Cost | Order ID   | Customer... | Loyalty ... | City   | Con... | Post... | Stat...   | Ord...    |
|---------|------------------|---------------|-----------|------------|-------------|-------------|--------|--------|---------|-----------|-----------|
| \$9.20  | 0                | \$290.80      | \$70.40   | 1239725727 | 74386       | 99          | Roma   | Europe | 00165   | Lazio     | 20JUL2... |
| \$10.80 | 0                | \$249.40      | \$119.30  | 1239725727 | 74386       | 99          | Roma   | Europe | 00165   | Lazio     | 20JUL2... |
| \$4.60  | 0                | \$145.40      | \$70.40   | 1239725735 | 75610       | 99          | Milano | Europe | 20122   | Lombar... | 20JUL2... |
| \$10.00 | 4                | \$113.60      | \$51.80   | 1239725752 | 76245       | 99          | Verona | Europe | 37100   | Veneto    | 20JUL2... |
| \$10.40 | 4                | \$117.80      | \$53.70   | 1239725752 | 76245       | 99          | Verona | Europe | 37100   | Veneto    | 20JUL2... |

8. Calculated Column

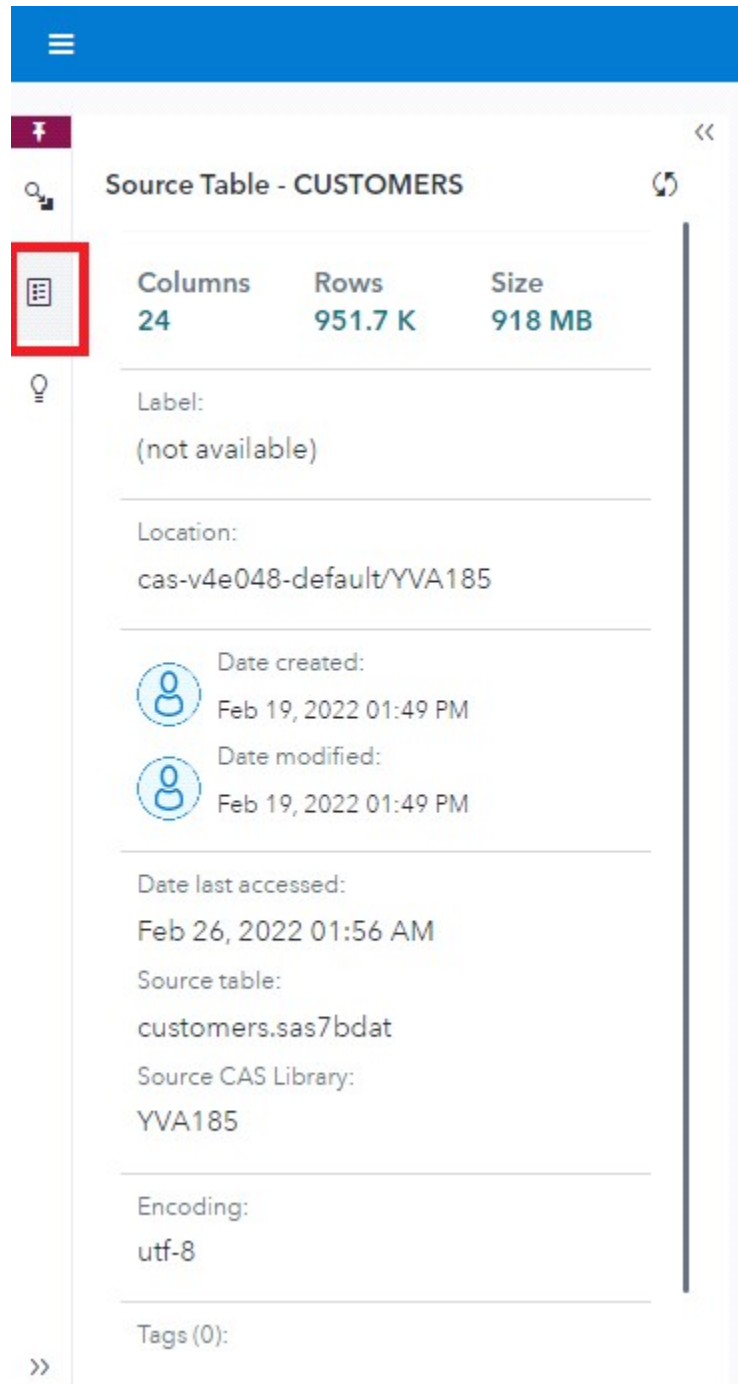
1 Delivery\_Date - Order\_Date

Options for new columns

Run

A sample of rows from the **CUSTOMERS** data source is displayed. These results might vary.

4. In the left pane (the leftmost pan), click  (**Properties for the source table**) to show details about the source table.



The screenshot shows the SAS Data Studio interface. The left sidebar contains a red box around the 'Properties' icon (a document with a list). The main pane displays the 'Source Table - CUSTOMERS' details.

| Columns | Rows    | Size   |
|---------|---------|--------|
| 24      | 951.7 K | 918 MB |

Label:  
(not available)

Location:  
cas-v4e048-default/YVA185

Date created:  
Feb 19, 2022 01:49 PM

Date modified:  
Feb 19, 2022 01:49 PM

Date last accessed:  
Feb 26, 2022 01:56 AM

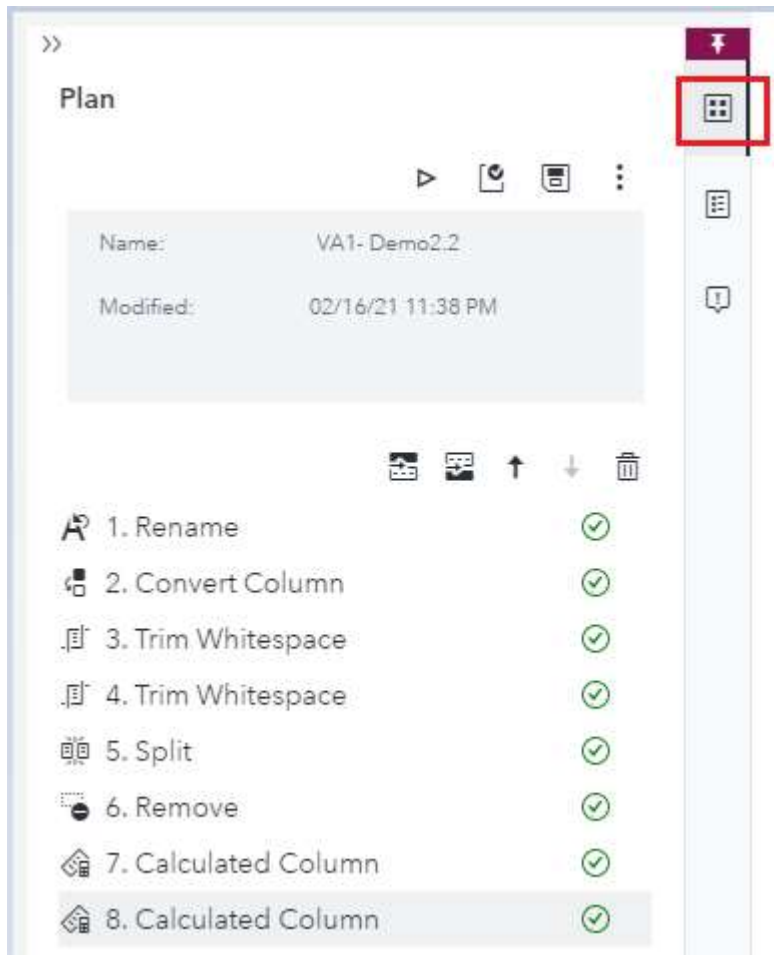
Source table:  
customers.sas7bdat

Source CAS Library:  
YVA185

Encoding:  
utf-8

Tags (0):

5. In the right pane (rightmost pane), click  (**Plan**) to view details about the steps performed in SAS Data Studio.



6. With step 8 selected (see the above picture), **Calculated Column**, view the calculation.

**8. Calculated Column** Run

|   |                            |
|---|----------------------------|
| 1 | Delivery_Date - Order_Date |
|---|----------------------------|

☐ Replace existing column  
☒ Create new column Days to Delivery  
[Options for new columns](#)

7. Select step 7, **Calculated Column** to view the calculation.

**7. Calculated Column** Run

|   |                                                   |
|---|---------------------------------------------------|
| 1 | ("Total Revenue"n - ("Unit Cost"n * "Quantity"n)) |
|---|---------------------------------------------------|

☐ Replace existing column  
☒ Create new column Profit  
[Options for new columns](#)

8. Click **Options for new columns** (Look at the bottom).

**Options for New Columns**

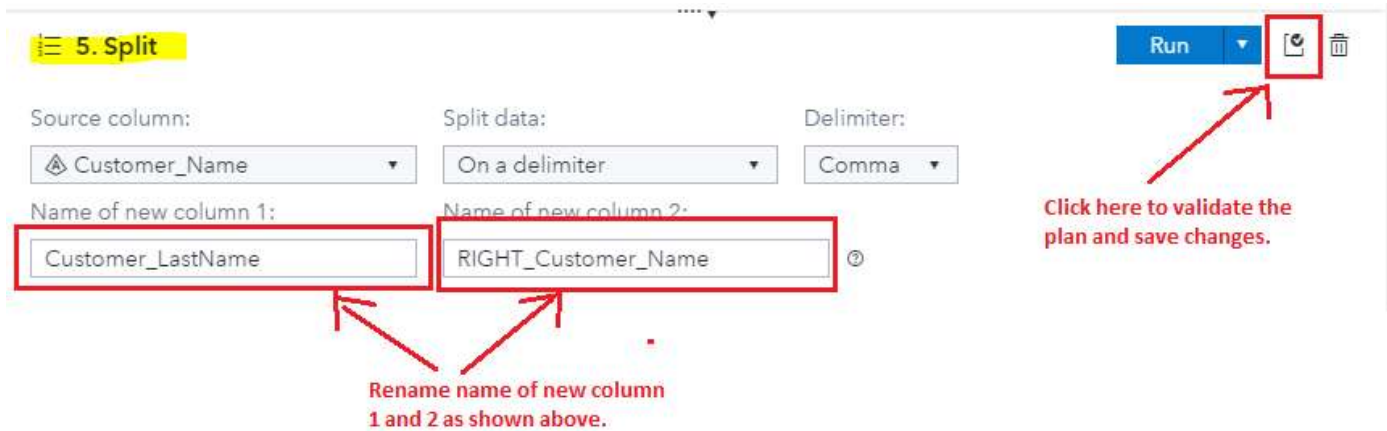
| Name of new column: | Type:  | Length: | Label: | Format:    |
|---------------------|--------|---------|--------|------------|
| Profit              | Double | 8       | Profit | DOLLAR12.2 |

OK Cancel

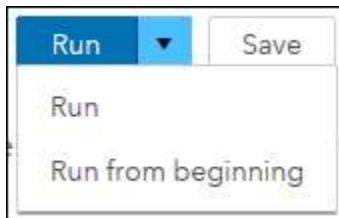
9. Click **OK**.

10. Select step 5, **Split**, to view the transformation.


- a. If not already selected, please select “**Customer\_Name**” from the **Source column** Drop Down Menu and “**On a delimiter**” from **Split data** Drop Down Menu. Then rename the “**Name of new column 1**” as **Customer\_LastName** and “**Name of new column 2**” as **RIGHT\_Customer\_Name**. Then click on highlighted **icon** next to Run button to validate the plan and save changes. While renaming the two columns, make sure there is no space in between or after the words. Otherwise you will get an error message



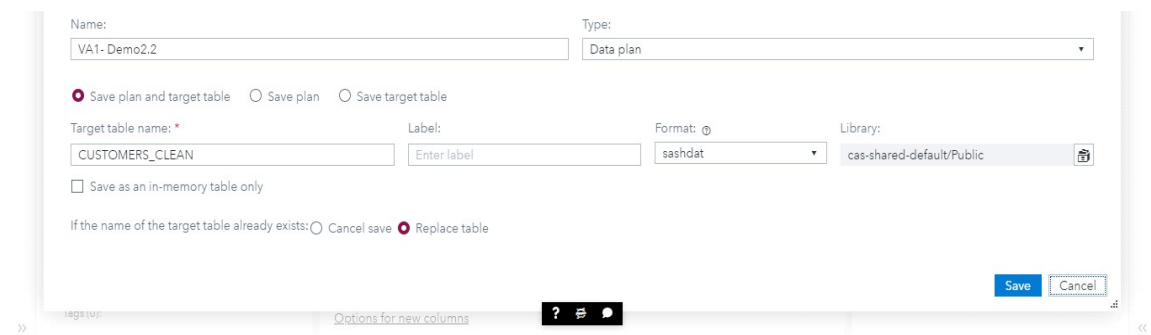
11. Click  and select **Run from beginning** to run the plan.



12. Save the plan and the result table.


- In the upper right corner of the plan, click  (**More**) and select **Save As**.
- Navigate to the **Shared Data/Basic/Demos (Marketing)** folder.
- Verify that **VA1-Demo2.2** is entered in the **Name** field.
- Verify that **Save plan and target table** is selected at the bottom of the window.
- Verify that **CUSTOMERS\_CLEAN** is listed in the **Table name** field.
- Verify that **cas-shared-default/Public** is specified for the **Library** field.
- Verify that **Replace table** is selected for the **If the name of the target table already exists** option.

The bottom portion of the Save As window should resemble the following:



- Click **Cancel**.

**Note:** When a plan is saved, the result table is automatically loaded to the CAS server.


**Note:** You can open the result table for the plan in another application by clicking  (**More**) and selecting **Actions** ⇒ **Saved table** ⇒ **Explore and Visualize Data**.

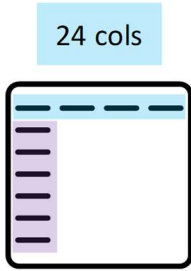
13. Select **your user profile** ⇒ **Sign Out** in the upper right corner to sign out of SAS Data Studio.

**End of Demonstration**



## Practice Scenario: Employees

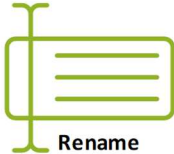





24 cols

648 obs


EMPLOYEES



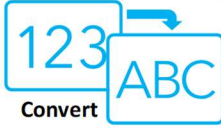
Rename




Filter



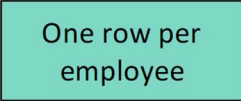
Delete



Convert




Change case



One row per employee

34



Copyright © SAS Institute Inc. All rights reserved.

## Practice Scenario: Employees





Calculate



Employee Group



Employee Tenure



Anniversary Month



Employee Type

35



Copyright © SAS Institute Inc. All rights reserved.

**Note:** The following data items are not needed for the analysis and will be removed: **Employee Birth Date, Section, Total Customers, Total Products Ordered, Total Quantity Ordered, Levels of Management, Manager at 2. level, Manager at 3. level, Manager at 4. level, Manager at 5. level, and Manager at 6. level.**

**Note:** **Anniversary Month** is calculated in SAS Data Studio. **Employee Tenure** is calculated in Visual Analytics using the Now operator, so the years of service value updates every time the report is opened. **Employee Type** can be calculated in SAS Data Studio using custom code. These data items can also be calculated in SAS Visual Analytics, which you see in a later lesson.

**Anniversary Month** is calculated as the name of the month in which the employee was hired.

**Employee Tenure** is calculated as  $(\text{Employee Termination Date} - \text{Employee Hire Date})/365.25$  for retired employees and as  $(\text{Today's Date} - \text{Employee Hire Date})/365.25$  for active employees.



**Employee Type** is *Retired* if the termination date is not missing and *Active* if the termination date is missing.

**Note:** The actual calculations are more complex and are discussed in more detail in later sections.



## Exercise 2B: Provide answers to Questions 1-5 (from Exercise 3 – pages 19-20). Support each answer with a relevant screenshot (s).


### 1. Preparing Data

- a. Open the browser and sign in to SAS Viya. SAS Drive is displayed by default.
- b. Open and run the **VA1- Practice2.2** plan in the **Courses/YVA185/Basics/Practices (HR)** folder.
  - 1) Click  (**Show list of applications**) and select **Prepare Data**.
  - 2) Click **Open Plan**.
  - 3) Navigate to the **Folders/SAS Content/Courses/YVA185/Basics/Practices (HR)** folder.
  - 4) Double-click **VA1- Practice2.2** to open the plan.
  - 5) Select **Run** to execute/update the plan.
- c. In the right pane, click  (**Properties for the result table**) and then answer the question.

- 1) How many rows are in the **EMPLOYEES** table after the actions of the plan are applied?

**Answer 1:**

**SS1:**

- d. In the right pane, click  (**Plan**) and then answer the questions.
  - 2) How many convert column actions were performed? On which column (or columns)?
    - Hint: **Select the Convert Column step.**

**Answer 2:**

**SS2:**

- 3) Which column was changed to uppercase?
  - Hint: **Select the Change Case step.**

**Answer 3:**

**SS3:**



4) What filter was applied to the table?

- Hint: **Select the Filter step**

**Answer 4:**

**SS4:**

5) What is the name of the new output table created from the plan?

- Hint: Click  (**Plan**).
- Click  (**Options**) and select **Save as**.

**Answer 5:**

**SS5:**