Chapter 2: Preparing Data Using SAS® Data Studio

2.1 Investigating Data in SAS Visual Analytics

2.2 Transforming Data Using SAS Data Studio

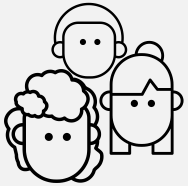
Objectives

- Describe the data used in the demonstrations and exercises.
- Discuss the Access phase of the SAS Visual Analytics methodology.
- Discuss the types of files that can be loaded into CAS using self-service import.
- Discuss the Investigate phase of the Visual Analytics methodology.
- Describe the SAS Visual Analytics interface.
- Discuss when to use list tables and crosstabs in Visual Analytics.
- Describe how the automatic chart changes based on the selected data items.

Business Scenario: Data

ORION STAR
Sports & Outdoors

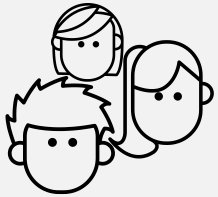
Demonstrations



68,300 customers

747,953 orders

Exercises



648 employees

Orion Star has many SAS data sets that contain information for the different divisions. In order to use this data in Visual Analytics, the following actions need to be performed:

- Tables need to be loaded into CAS.
- Data quality issues need to be corrected.
- Some data items might need to be created for the analysis.

Other things might be discovered along the way.

sas

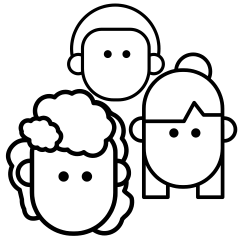
Business Scenario: Customers

You have been hired as an analyst and report designer for the Marketing Division of Orion Star.

Here is the information for your first assignment:

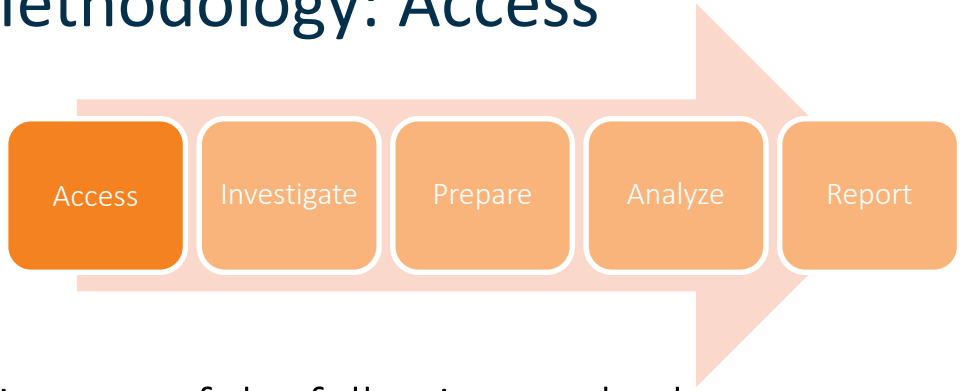
- The Marketing team has asked for an analysis of profits.
- The Shipping team has asked for an analysis of delivery times.

You need to access and investigate the data to determine whether it is ready to be used by the analysts.



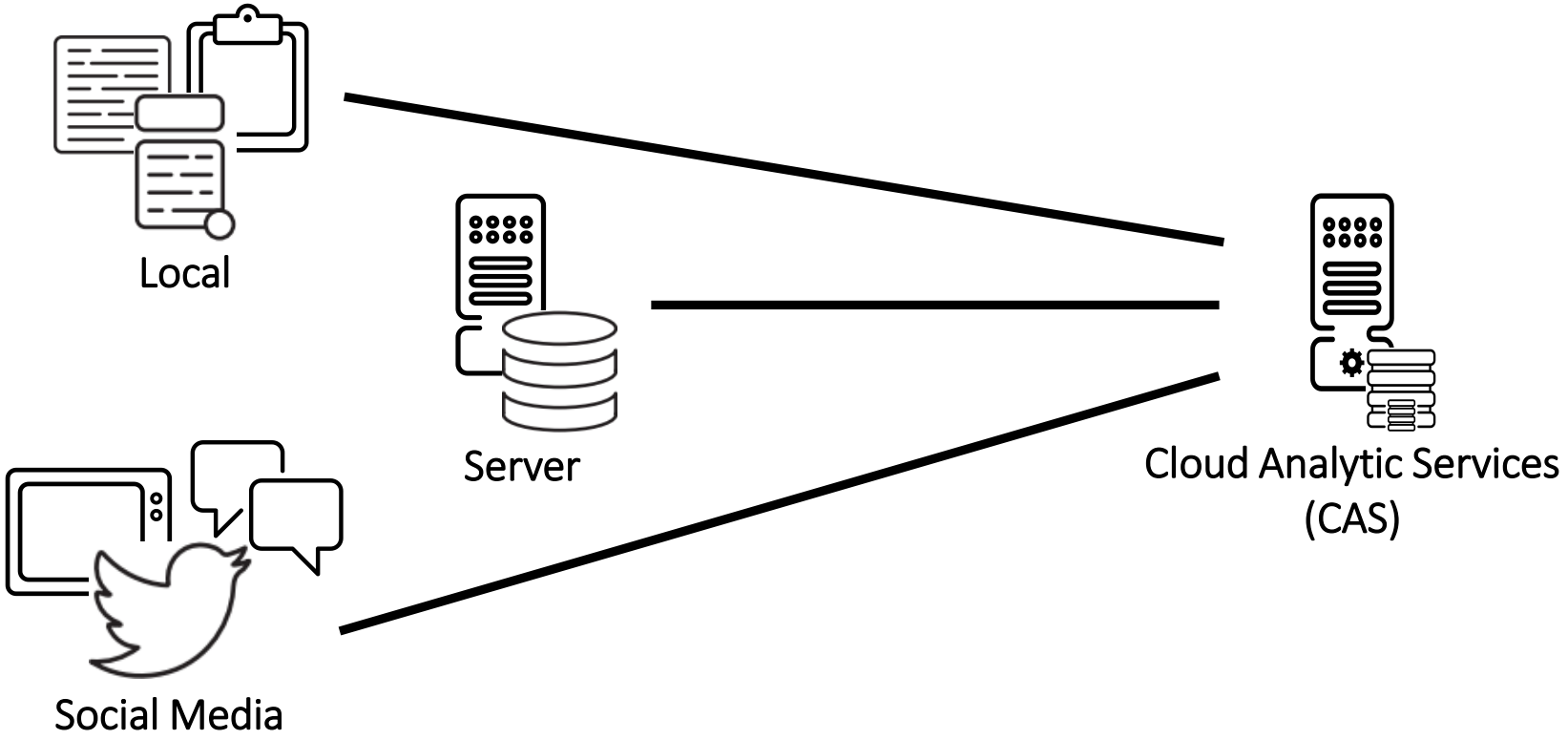
Visual Analytics Methodology: Access

In the ***Access*** phase, you might need to complete the following steps:



- Identify or create analysis tables (or do both).
- Load the analysis tables into CAS using one of the following methods:
  - importing data using SAS Environment Manager
  - creating plans in SAS Data Studio
  - uploading data using the task in SAS Enterprise Guide
  - executing SAS code (using SAS Studio or Enterprise Guide)
  - using self-service import
  - executing other supported open languages: Python, Lua, Java

Accessing Data: Self-Service Import

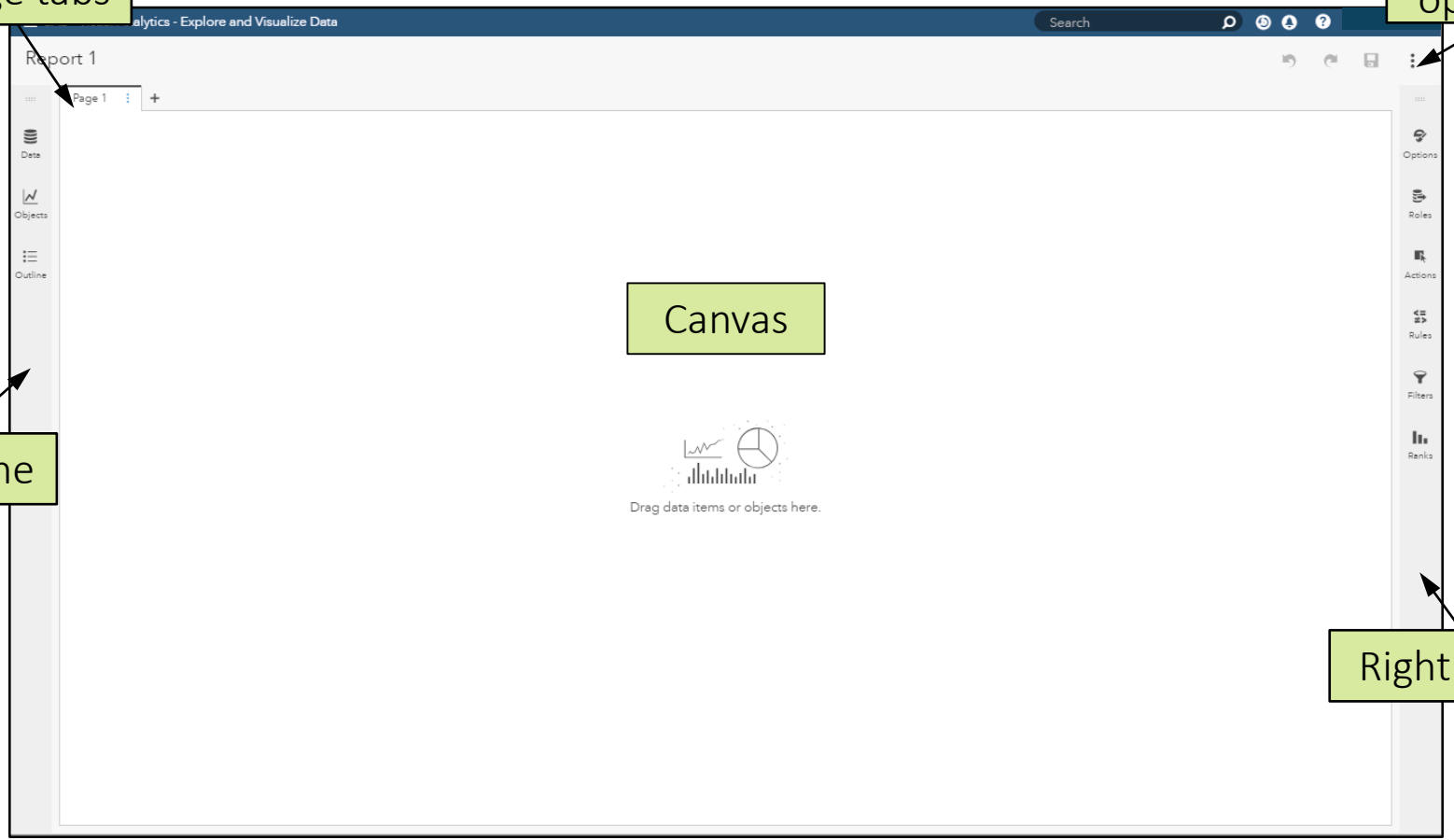


Visual Analytics Methodology: Investigate

In the *Investigate* phase, you need to inspect the tables and answer questions such as the following:

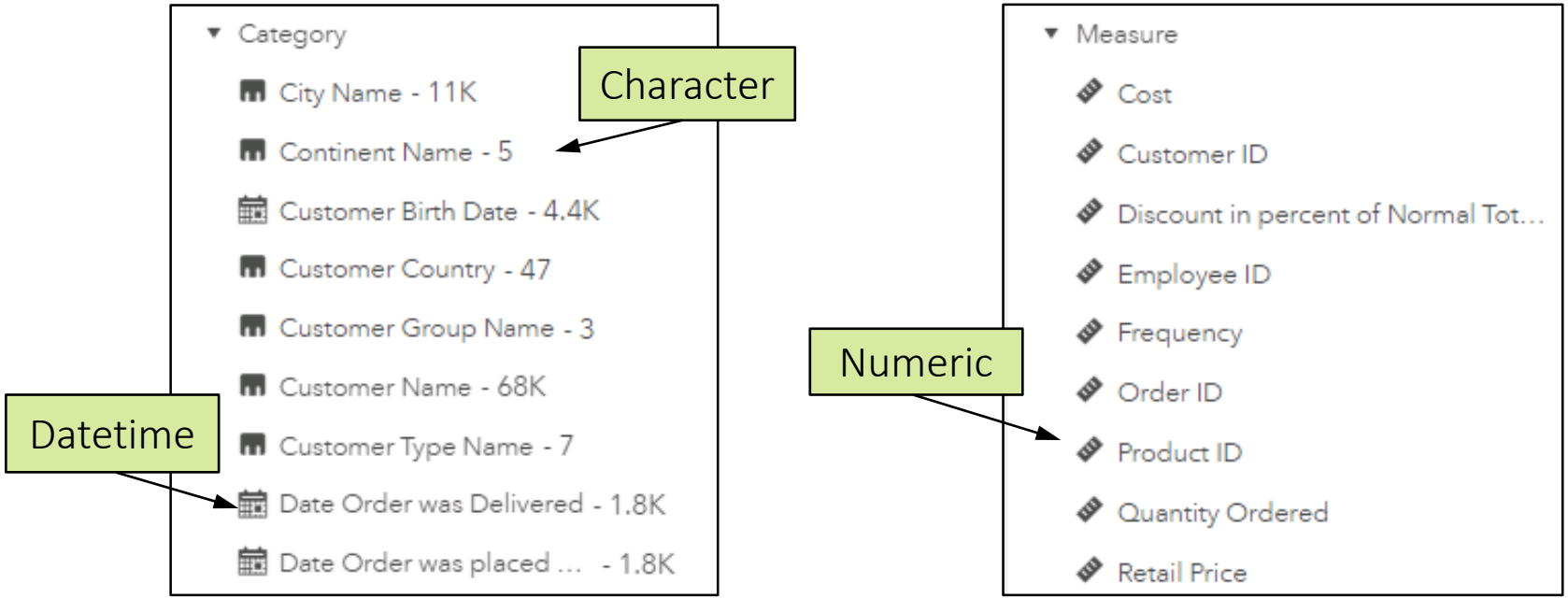
- How big is the data?
- How is the data shaped?
- Are there any data quality issues? Missing values?
- Are there any data items that need to be calculated for the analysis?

SAS Visual Analytics Interface



Data Types in Visual Analytics

There are two main data classifications in Visual Analytics:



Objects: Tables

Click to sort

Customer Name	Quantity Ordered
Zyryi, Mr. Christoher	5
Zwilling, Mr. Timothy	58
Zwikker, Ms. M.E.	34
Zwikker, Mr. Jan	96
Zwikker, Mr. F.W.A.	11
Zwietering, Ms. T.W.A	17
Zwier, Mr. Frank	17

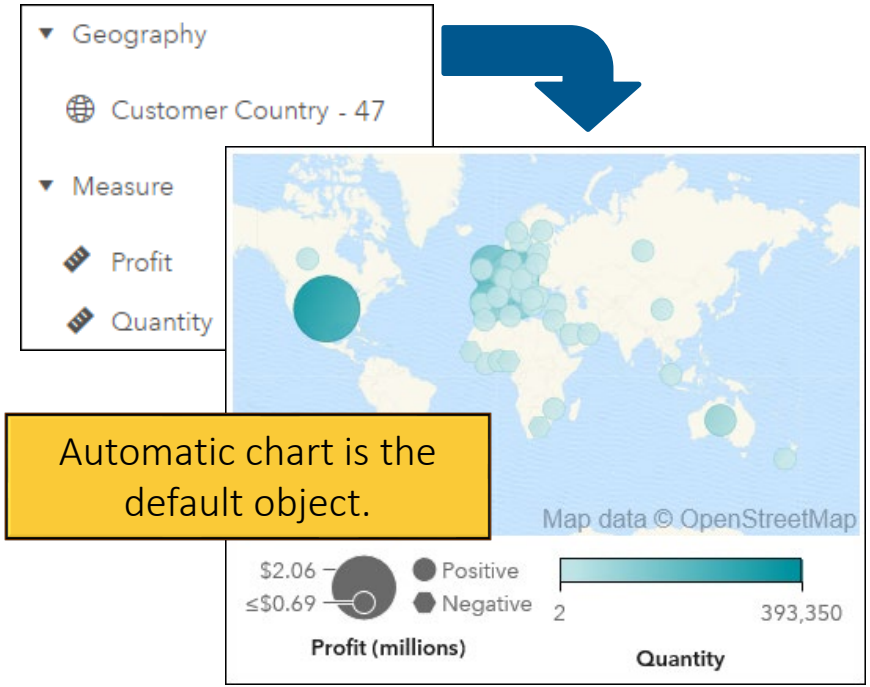
Use a *list table* to view summary or detail data about your data source.

Use a *crosstab* to view summary information for multiple categories.

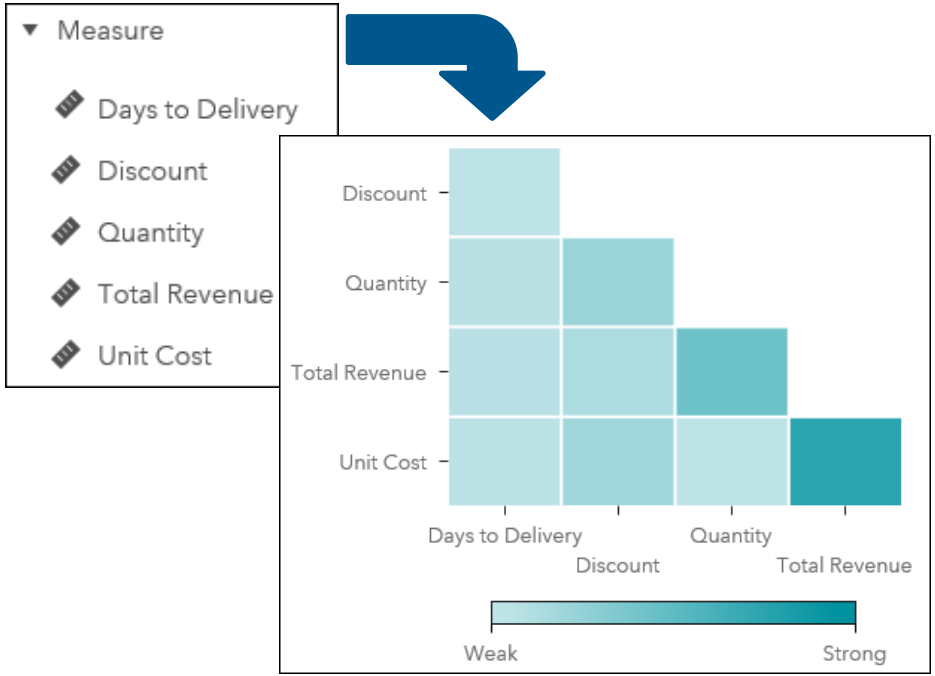
Order Type ▲	Catalog Sale	Internet Sale	Retail Sale
Continent Name ▲	**Quantity Ordered**	**Quantity Ordered**	**Quantity Ordered**
Africa	548	793	.
Asia	845	1,073	.
Europe	142,511	120,384	836,473
North America	63,480	55,688	280,652
Oceania	14,811	12,551	67,508

Objects: Automatic Chart

An *automatic chart* selects the chart type based on the assigned data. Automatic charts provide a quick view of the data.



Automatic chart is the default object.



Chapter 2: Preparing Data Using SAS® Data Studio

2.1 Investigating Data in SAS Visual Analytics

2.2 Transforming Data Using SAS Data Studio

Objectives

- Discuss the Prepare phase of the Visual Analytics methodology.
- Describe the Data Studio interface.
- Discuss the information displayed in the Table Profile window.
- Discuss the information displayed in the Column Profile window.
- Apply data transformations (rename, modify classification, remove white space, change case, filter, remove) to columns in SAS Data Studio.
- Create new columns (splitting, calculated) in SAS Data Studio.

Business Scenario: Customers

The **CUSTOMERS** table contains a total of 951,669 observations and 24 columns. Each row represents a product ordered by a customer, so there are multiple rows for each order and multiple rows for each customer.

The following data cleansing operations need to be performed:

Rename
Retail Price ▸ Total Revenue
Cost ▸ Unit Cost

Convert column
Customer_ID
Order_ID


Business Scenario: Customers

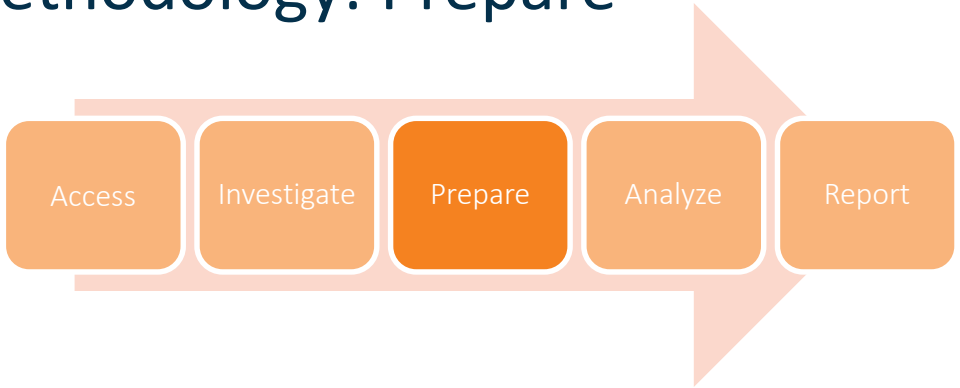
The Marketing team has asked you for an analysis of profits, and the Shipping team has asked for an analysis of delivery times.

In order to perform this analysis, the following data items need to be calculated:

- **Customer_LastName**, **Title**, and **Customer_FirstName**
- **Profit**
- **Days to Delivery**
- Customer Age and Customer Age Group
- Customer Gender

Visual Analytics Methodology: Prepare

In the ***Prepare*** phase, you need to complete the following tasks:



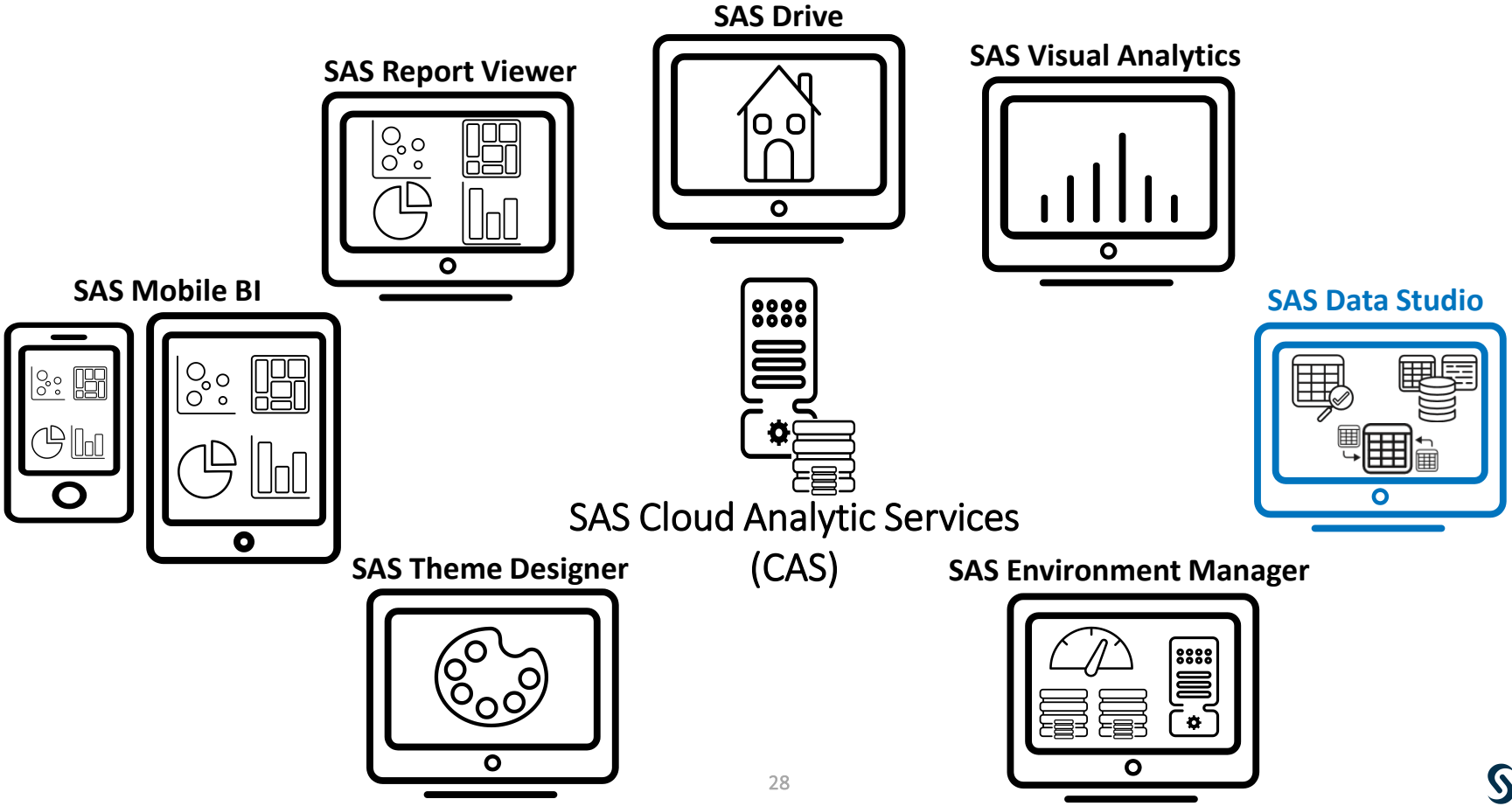
Correct any data issues discovered.

- split columns
- change case
- change data type or format (or both)
- remove white space
- delete or rename columns
- sort columns

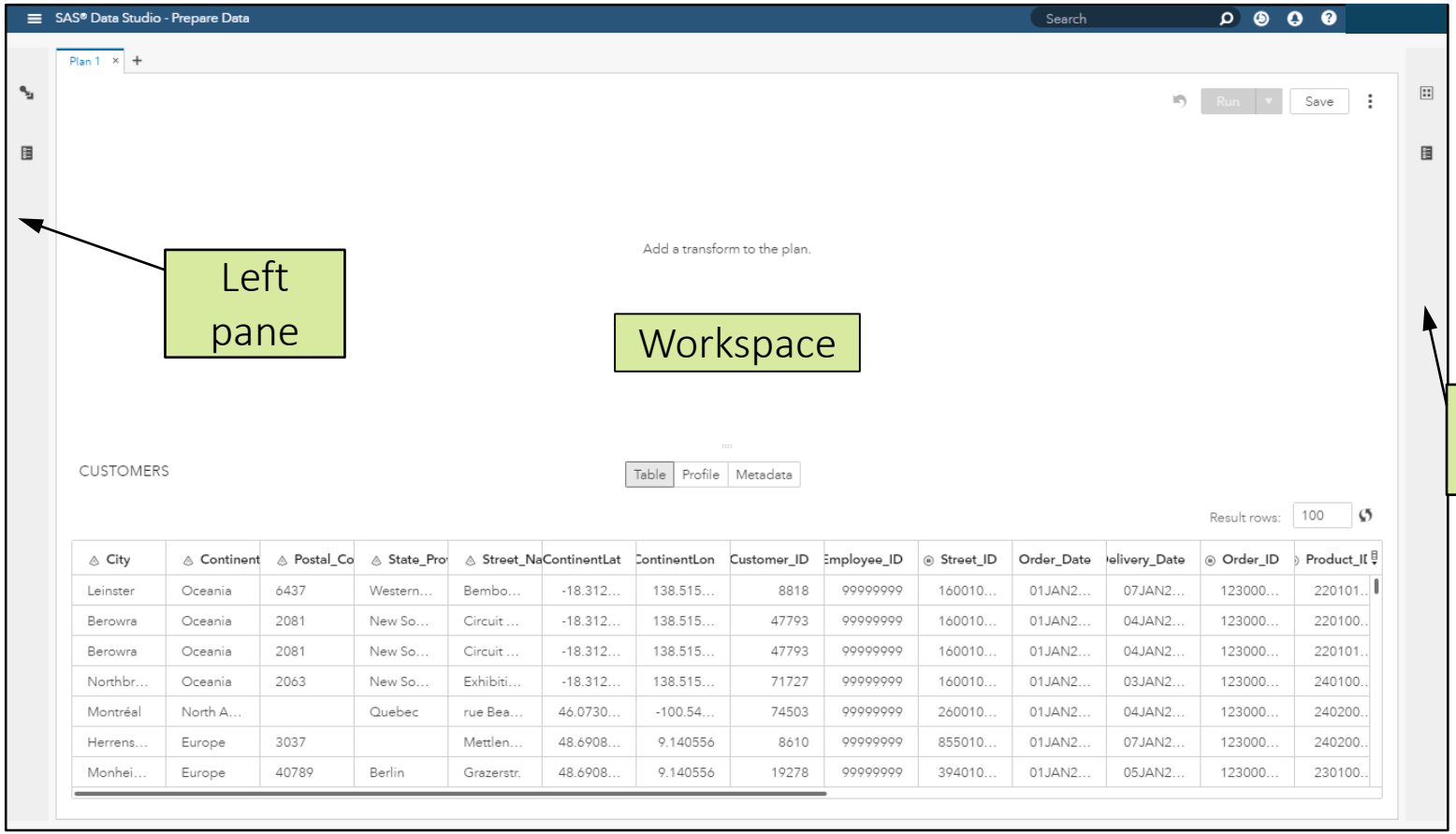
Create new calculated items needed for analysis.

Note: These transformations and calculations are saved in a plan.

SAS Viya Applications

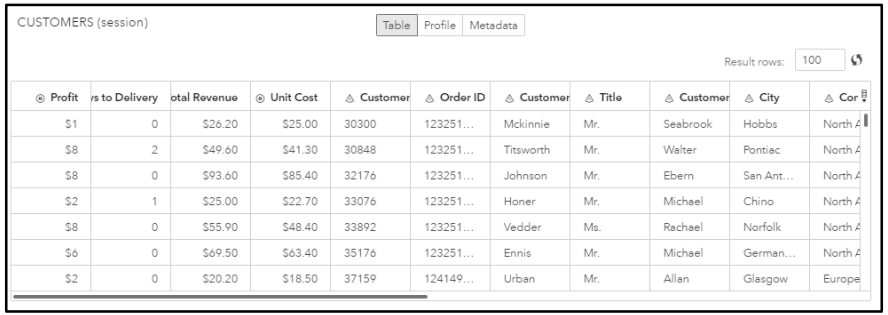


SAS Data Studio

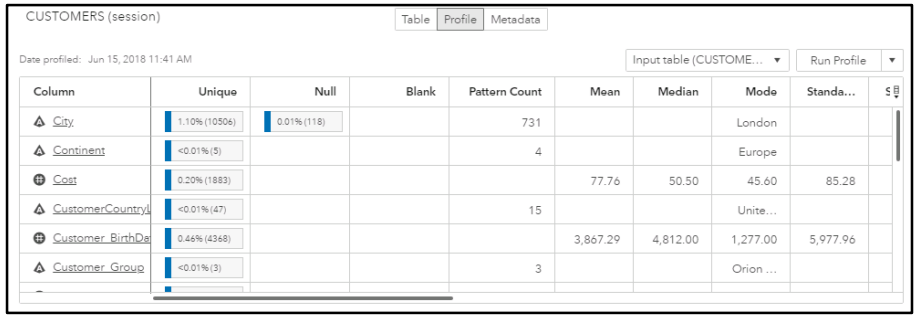


Table, Profile, and Metadata Information

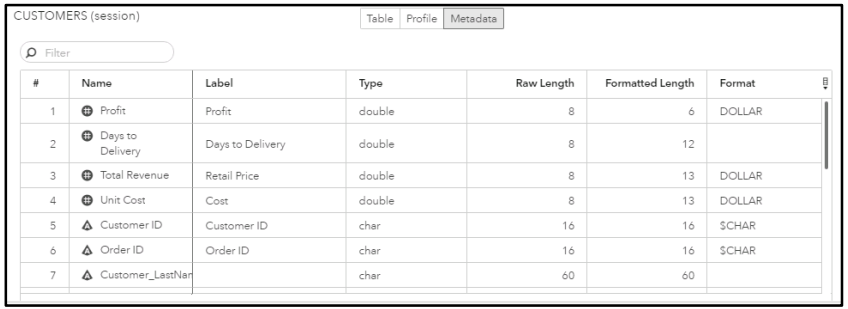
At the bottom of the workspace, you can view the Table, Profile, and Metadata information about the table.



Profit	s to Delivery	otal Revenue	Unit Cost	Customer	Order ID	Customer	Title	Customer	City	Cor
$1	0	$26.20	$25.00	30300	123251…	Mckinnie	Mr.	Seabrook	Hobbs	North A
$8	2	$49.60	$41.30	30848	123251…	Titsworth	Mr.	Walter	Pontiac	North A
$8	0	$93.60	$85.40	32176	123251…	Johnson	Mr.	Ebern	San Ant…	North A
$2	1	$25.00	$22.70	33076	123251…	Honer	Mr.	Michael	Chino	North A
$8	0	$55.90	$48.40	33892	123251…	Vedder	Ms.	Rachael	Norfolk	North A
$6	0	$69.50	$63.40	35176	123251…	Ennis	Mr.	Michael	German…	North A
$2	0	$20.20	$18.50	37159	124149…	Urban	Mr.	Allan	Glasgow	Europe



Column	Unique	Null	Blank	Pattern Count	Mean	Median	Mode	Standa…
City	1.10% (10506)	0.01% (118)		731			London	
Continent	<0.01% (5)			4			Europe	
Cost	0.20% (1883)				77.76	50.50	45.60	85.28
CustomerCountryL	<0.01% (47)			15			Unite…	
Customer_BirthDa	0.46% (4368)				3,867.29	4,812.00	1,277.00	5,977.96
Customer_Group	<0.01% (3)			3			Orion …	



#	Name	Label	Type	Raw Length	Formatted Length	Format
1	Profit	Profit	double	8	6	DOLLAR
2	Days to Delivery	Days to Delivery	double	8	12	
3	Total Revenue	Retail Price	double	8	13	DOLLAR
4	Unit Cost	Cost	double	8	13	DOLLAR
5	Customer ID	Customer ID	char	16	16	$CHAR
6	Order ID	Order ID	char	16	16	$CHAR
7	Customer_LastNar		char	60	60	

Data Transformations

You can perform data transforms such as changing data types, splitting columns, creating calculated columns, joining tables, and filtering data.