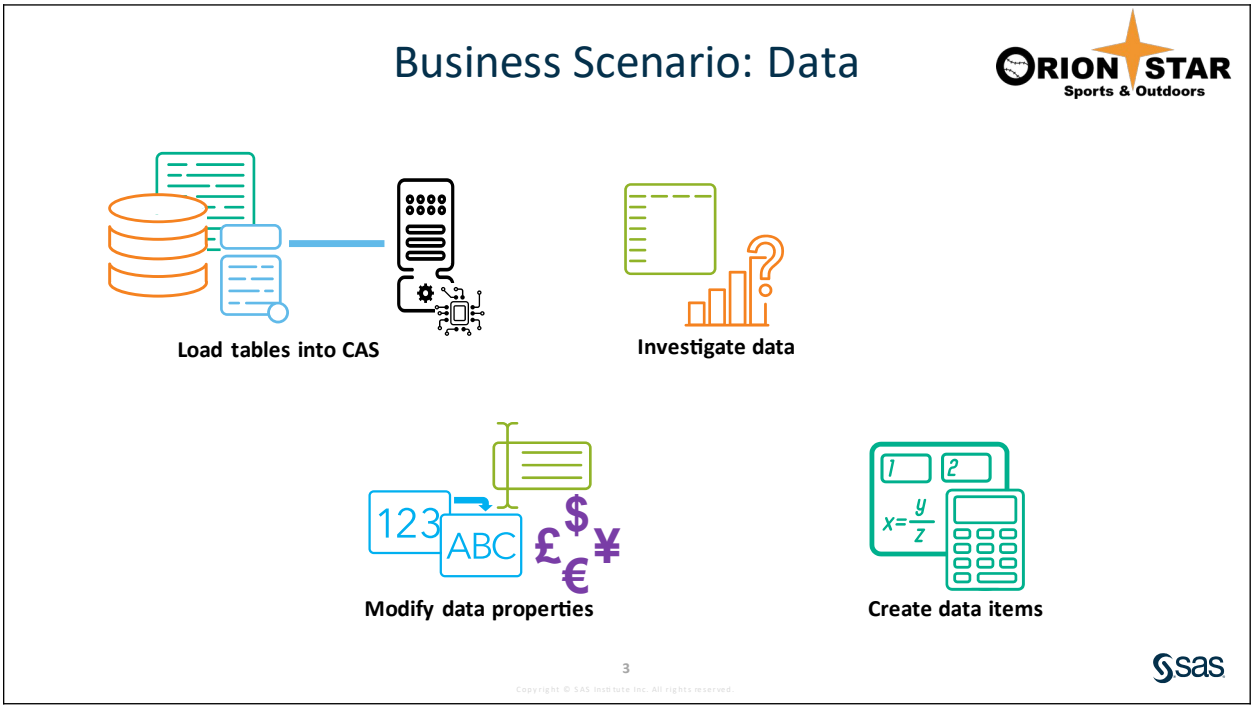
Lesson 2 Preparing Data Using SAS® Data Studio

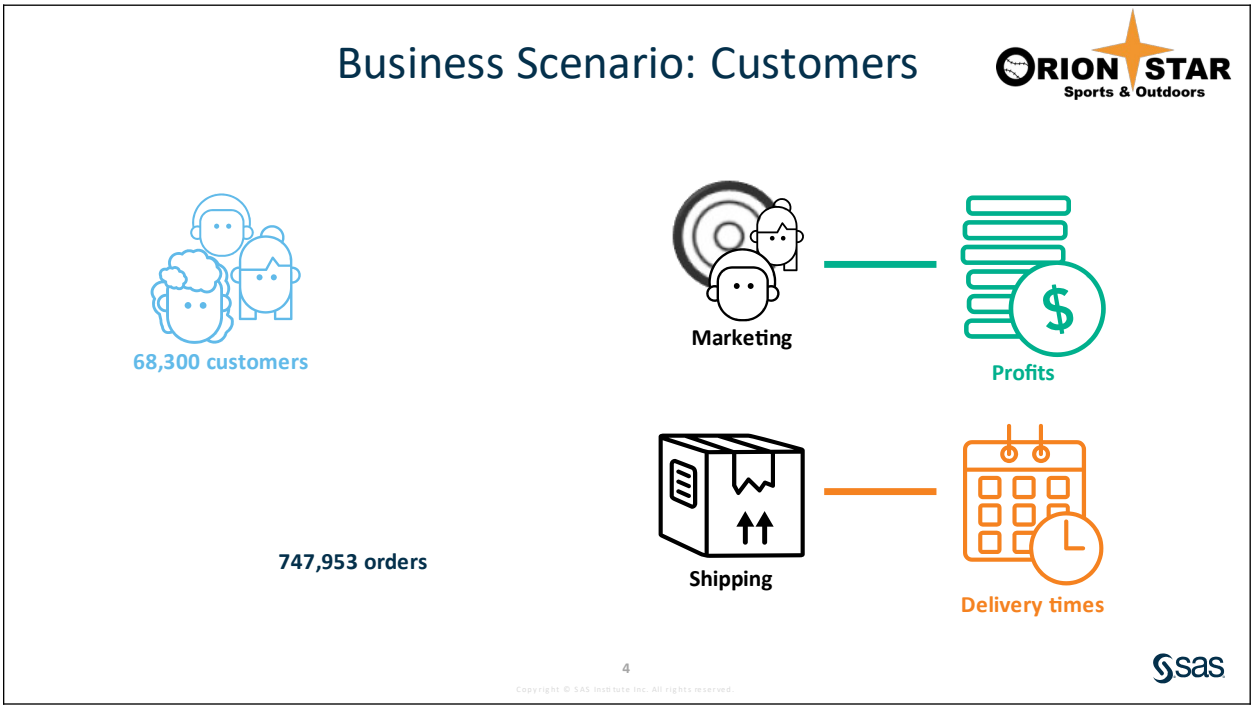
Submit: Open a NEW word document. Save it as A6XX (XX – your initials). Provide answers to Questions 1-8 (from Exercise 2 – pages 27-29). Support each answer with a relevant screenshot(s).

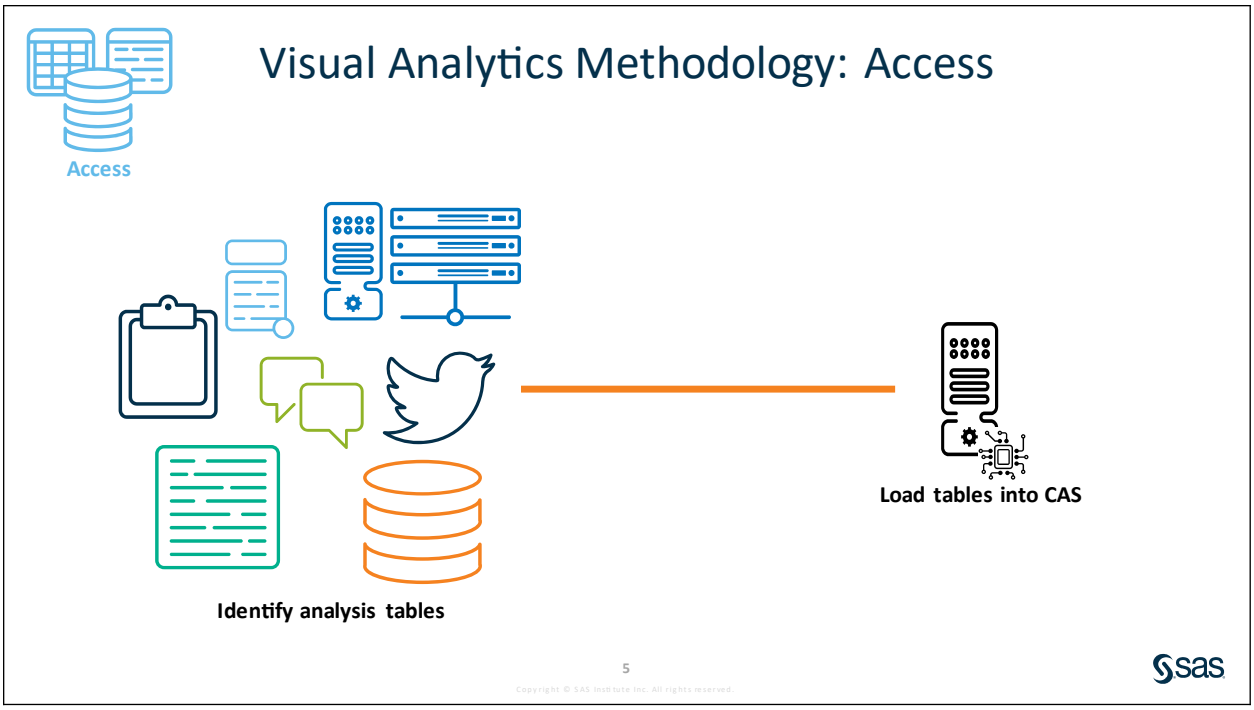
NOTE: DO NOT submit your work within this document provided to you by your instructor. You must submit a NEW word document created by you.

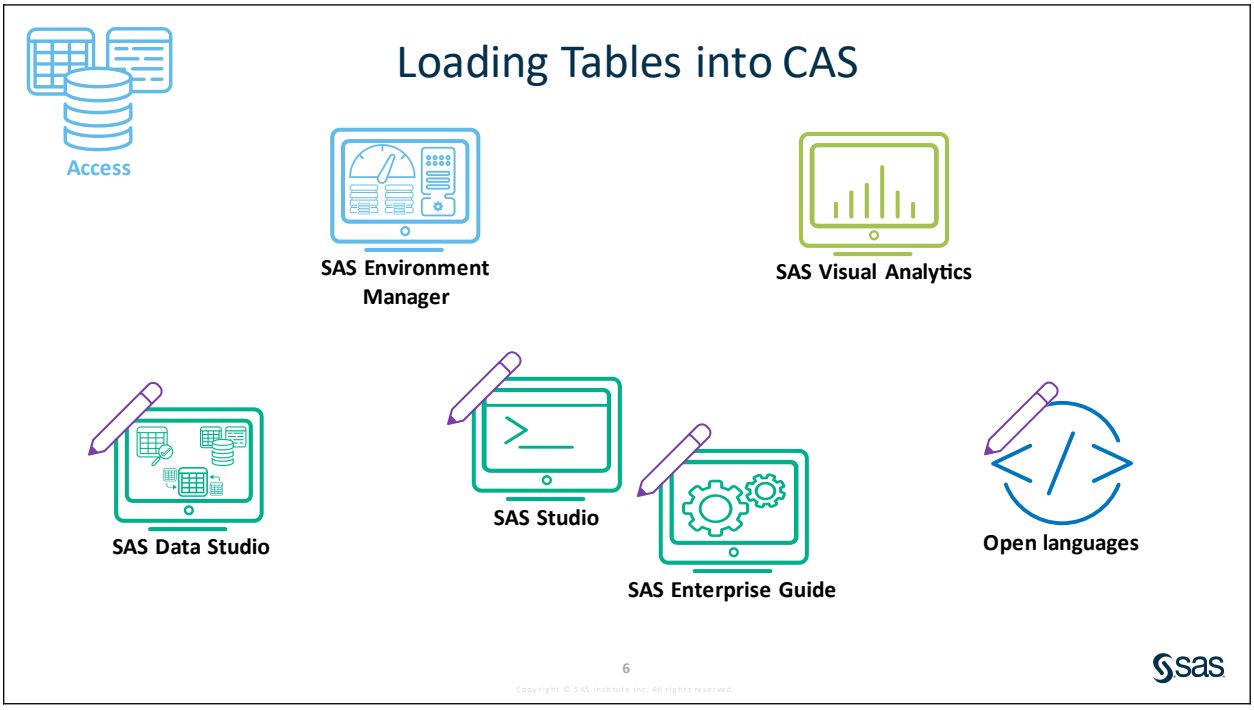
1. To Login to SAS: https://vle.sas.com/vfl

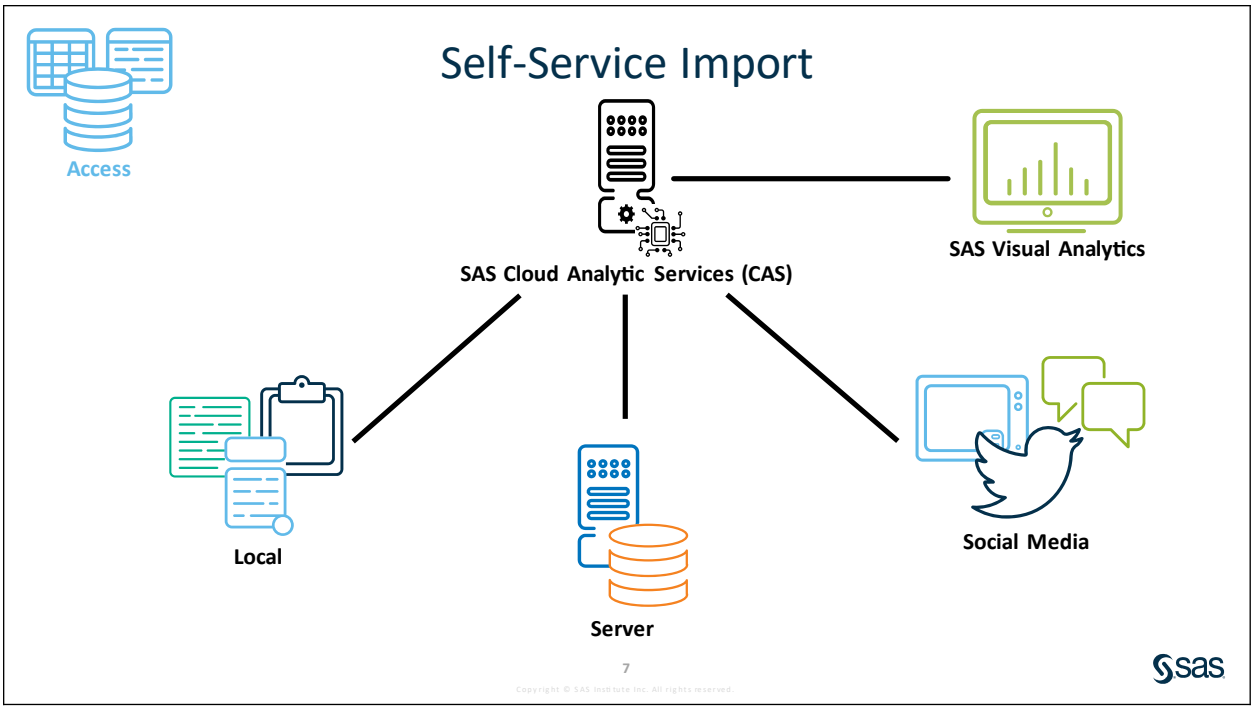
Investigating Data in SAS Visual Analytics







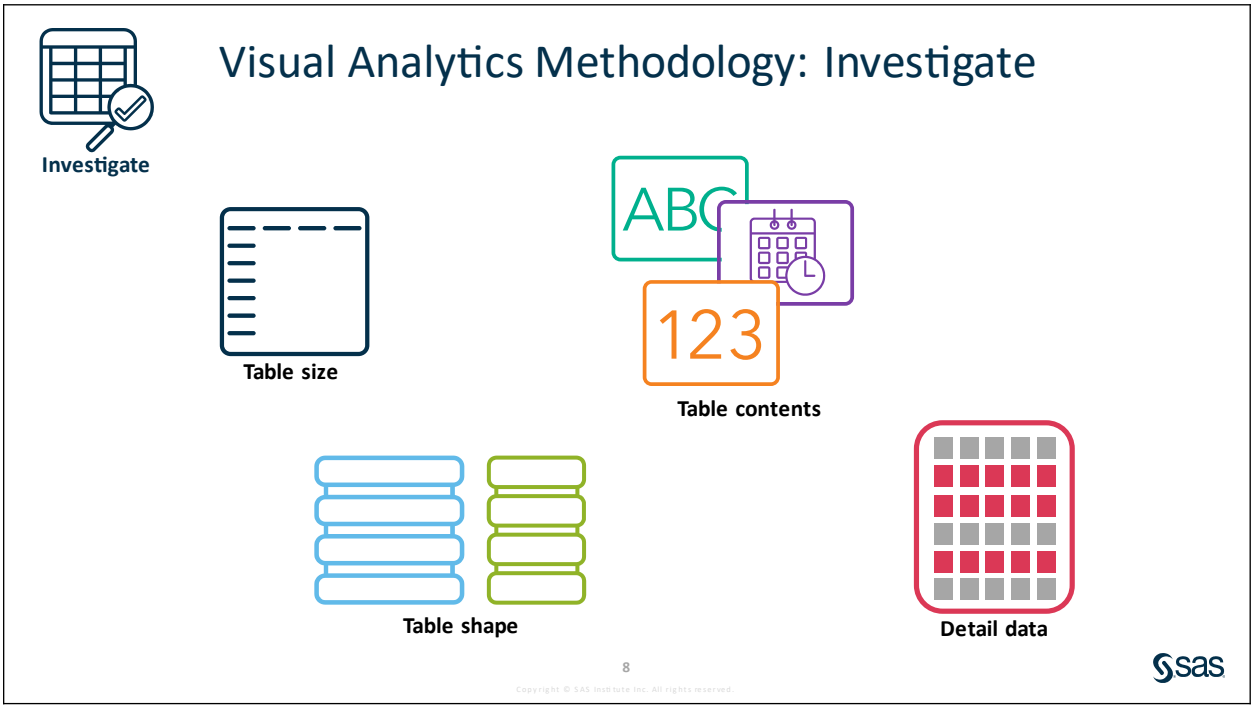


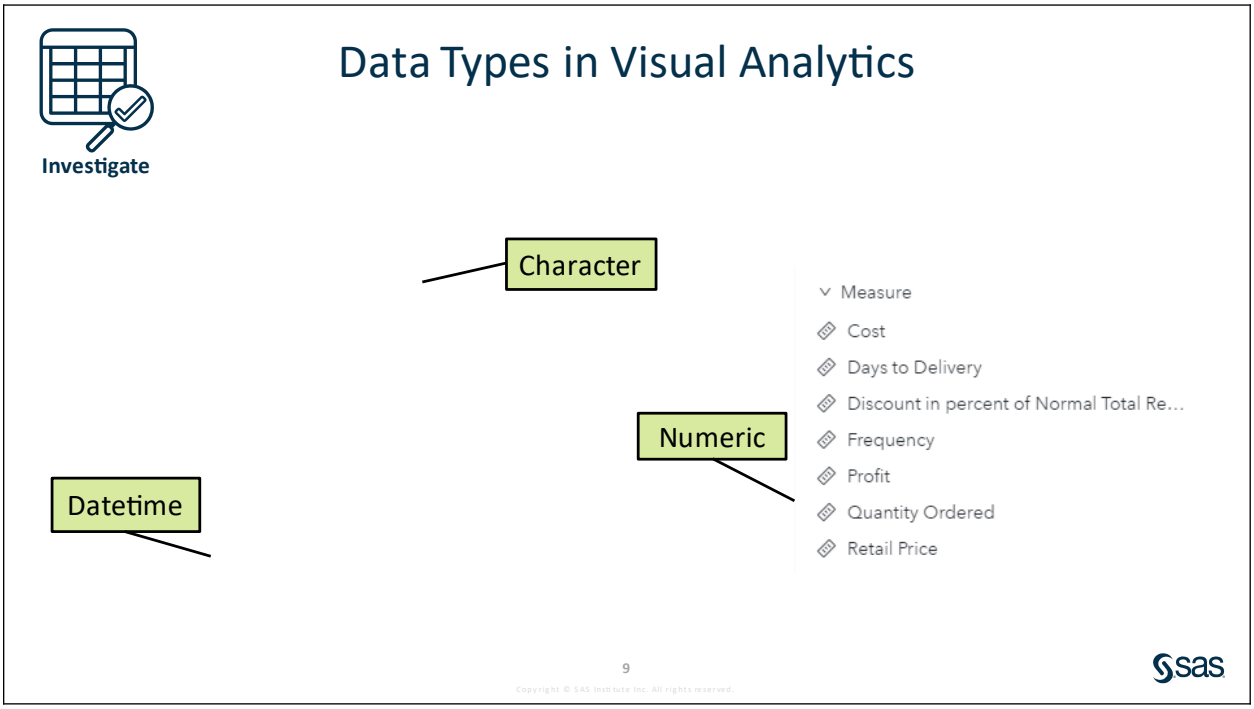


Here are the types of data that can be imported to CAS using self-service import:

Documents Directory	You can extract text and metadata from a collection of documents in a caslib and write this information to a table that can be analyzed using SAS Visual Text Analytics. For more information, see “Working with Data in CAS” in the *SAS Visual Analytics: SAS Data Explorer* documentation.
Local	You can import data from a Microsoft Excel spreadsheet (XLS or XLSX), a text file (CSV or TXT), or a SAS data set (SASHDAT or SAS7BDAT).
Server	After providing connection information, you can import a table into CAS from a database (Teradata, Oracle, Hadoop, and so on) or from the LASR Analytic Server.
Social Media	After authenticating with Facebook, Google Analytics, Twitter, YouTube, or Google Drive and providing search criteria (where applicable), you can import data to the CAS server. **Note:** Your access to, and use of, social media data through a social media provider’s public APIs is subject to the social media provider’s applicable license terms, terms of use, and other usage terms and policies.
Esri	The GeoEnrichment Service from Environmental Systems Research Institute (Esri) provides a large collection of geographic data sets, including population, income, housing, consumer behavior, and the natural environment. You can use this service to combine Esri enrichment data with a table in CAS. The resulting output table contains new columns of Esri data that are associated with geographic location codes in the source table. Note: In order to add geo-enrichment data to CAS tables, you must belong to the custom group (Esri Users), have credentials for Esri GeoEnrichment Service, accept Esri ArcGIS Online Services terms and conditions, and enable Esri premium services.

Folders	You can import text files, SAS data sets, and Microsoft Excel files that were saved to a SAS folder. For more information, see “Making Data Available to CAS” in the *SAS Visual Analytics: SAS Data Explorer* documentation.





In Visual Analytics, character and datetime data items are treated as *categories*—that is, data items whose distinct values can be used to group and aggregate measures. In the Data pane, distinct counts are displayed for each category data item. Numeric data items are treated as *measures*—that is, data items whose values can be used in computations.

In the Data pane, an icon next to each data item indicates the type of that item. Here are the data items that are available:

Category		A data item whose distinct values are used to group and aggregate measures. There are five types of categories: alphanumeric, date, datetime, time, and numeric.
Date and Time		A category data item whose distinct values are used to group and aggregate measures. There are three types of date categories: date, datetime, and time.
Custom Category		A data item that can be created based on either a category or numeric data item. A custom category is always a category data item with alphanumeric values.
Calculated (category)		A data item that is calculated from existing data items using an expression and returns an alphanumeric value.
Calculated (datetime)		A data item that is calculated from existing data items using an expression and returns a datetime value. Calculated dates and times are treated as categories with distinct values being governed by the chosen date or time format.
Geography		A category data item whose values are mapped to geographical locations or regions. These data items can be used to show data on a geographic map.
Hierarchy		A data item with a predefined arrangement of category data items, typically whose values are arranged with more general information at the top and more specific information at the bottom. The first level of the hierarchy is known as the *root* level.
Geographic Hierarchy		A hierarchy whose members are all geographic data items.
Interaction Effect		A user-created data role that can be used when there is a nonadditive relationship between two variables (the effect of one variable on a model changes as another variable changes). SAS Visual Statistics must be licensed for you to create and use an interaction effect.
Measure		A data item whose values can be used in computations. These values are numeric. By default, almost all measures have a default aggregation of Sum, but the aggregation can be modified.
Calculated (measure)		A data item that is calculated from existing data items using an expression and returns a numeric value. Numeric data items are treated as measures (with an aggregation type of Sum), or they can be changed to category data items.

Frequency		A measure data item whose value represents the number of observations in the selected data source. This data item is automatically added to the Data pane under the Measure group. You cannot change the classification for this data item. This data item is automatically assigned to some report objects when no measure is assigned.
Frequency Percent		A measure data item whose value represents the percentage of observations in the selected data source. This data item is automatically added to the Data pane under the Aggregated Measure group. You cannot change the classification for this data item.
Aggregated Measure or Time Period Calculation		A data item that represents special predefined operations, like distinct count, percentage of totals, percentage of subtotals, or frequency percent. Users can also create their own aggregated measure calculations. Aggregated measures cannot be used in all report objects, filters, controls, spark lines, or time series graphs. Some aggregated measures cannot be used in a detail rank. Percentage of subtotal items can be used only in a crosstab.
Parameter		A variable whose values can be changed and that can be referenced by other objects. You can use parameters in calculations, display rules, filters, and ranks.
Spline Effect		A spline function is a piecewise polynomial function in which the individual polynomials have the same degree and connect smoothly at certain points. Spline functions are used to fit smooth curves to a wide variety of data. SAS Visual Statistics must be licensed to create and use a spline effect.

Objects: Tables

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

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

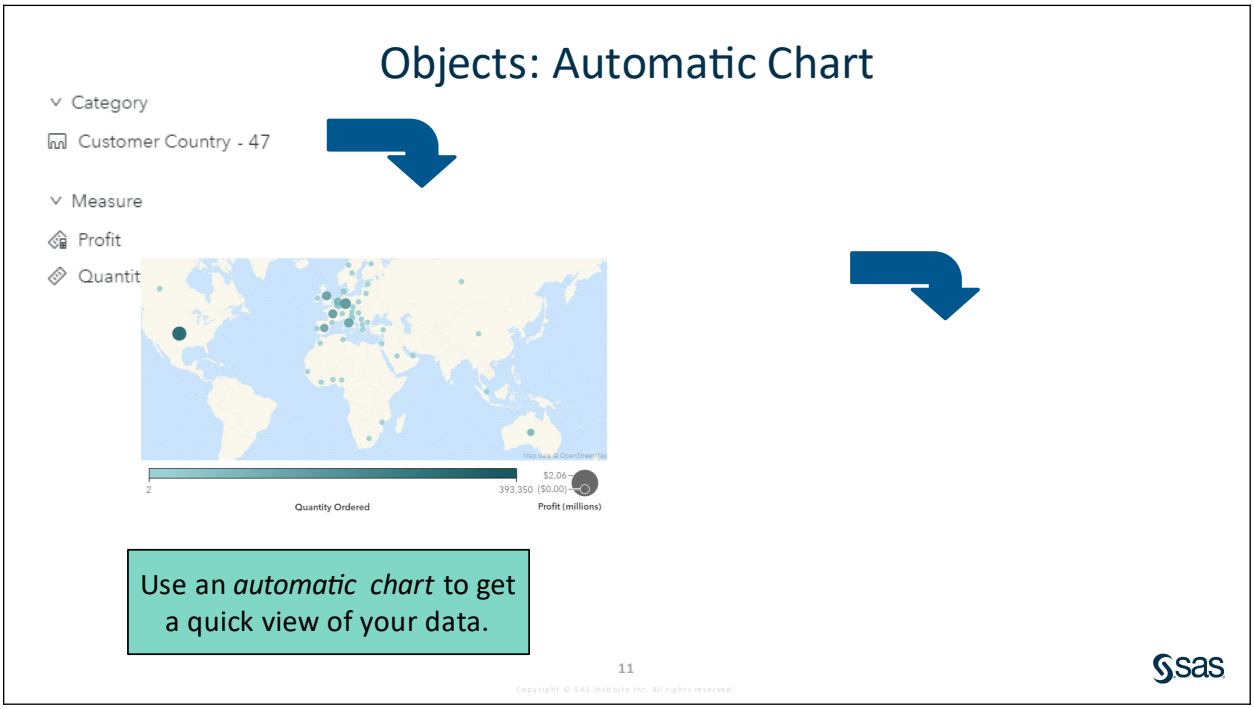
Order Type ▲	Catalog Sale	Internet Sale	Retail Sale
Continent Name ▲	Quantity Ordered	Quantity Ordered	Quantity Ordered
Africa	548	793	—
Asia	845	1,073	—
Europe	142,511	120,384	836,473
North America	63,480	55,688	280,652
Oceania	14,811	12,551	67,508

10

sas

List table	By default, the list table contains aggregated data with one row for each distinct combination of category values. If the **Detail data** option has been selected, then

	every row of the data source is displayed. **Note:** By default, the list table is sorted in ascending order by the first column. **Note:** To change the sorting, click on the heading for the column on which you want to sort. An arrow appears in the column heading to indicate the sorting. If the arrow points up, then the sort is ascending. If the arrow points down, then the sort is descending. **Note:** To sort on multiple columns, hold down the Ctrl key and click the columns to sort by, in order. **Note:** If detailed data is displayed in the list table, it cannot be the source of an action or link.
Crosstab	Each cell of the crosstab contains the aggregated measure values for a specific intersection of category values. You should consider placing lower cardinality (fewer distinct values) categories in the Columns role and higher cardinality (more distinct values) categories in the Rows role.



Data Items	Chart Type
One measure	Histogram
One category and any number of measures	Bar chart
One datetime category and any number of measures	Time series plot
One date or datetime category and one or more categories	List table
One geography and up to two measures	Geo map
One geography and three or more measures	Bar chart

One hierarchy and any number of measures	Bar chart
One hierarchy, one or more categories, and any number of other data items	Crosstab
Two or more hierarchies and any number of other data items	Crosstab
Two or more categories and any number of measures	List table
Two or three measures	Scatter plot or heat map*
Four or more measures	Scatter plot matrix or correlation matrix*

* The actual chart type depends on the cardinality of the data.

Suggested objects are generated in the following ways:

- **Correlated measures** - A correlation query runs against the data source. The suggested object has the lowest cardinality data item and the two most correlated measures. This can generate a butterfly chart, a dual axis bar chart, or a dual axis bar-line chart.
- **Lowest cardinality** - The lowest cardinality data item, a category with at least six distinct values, in the data source is used to create the suggested object using frequency. This can generate a bar chart, a dot plot, a line chart or needle plot, a pie chart, a treemap, or a word cloud.
- **Custom data items** - Custom data items are detected in the following order: hierarchies, custom categories, calculated items. Suggested objects are generated for each custom data item. The generated objects are the same as those listed for lowest cardinality.
- **Date data items** - If the data source contains a date or time data item, a suggested object (typically a time series plot) is generated using a random measure.
- **Single measures** - A histogram or key value object is generated for a single measure.

Note: The Suggestions pane is not available if the **Bypass retrieving cardinality values for the Data pane** or **Bypass retrieving correlation values for the Data pane** settings are selected.

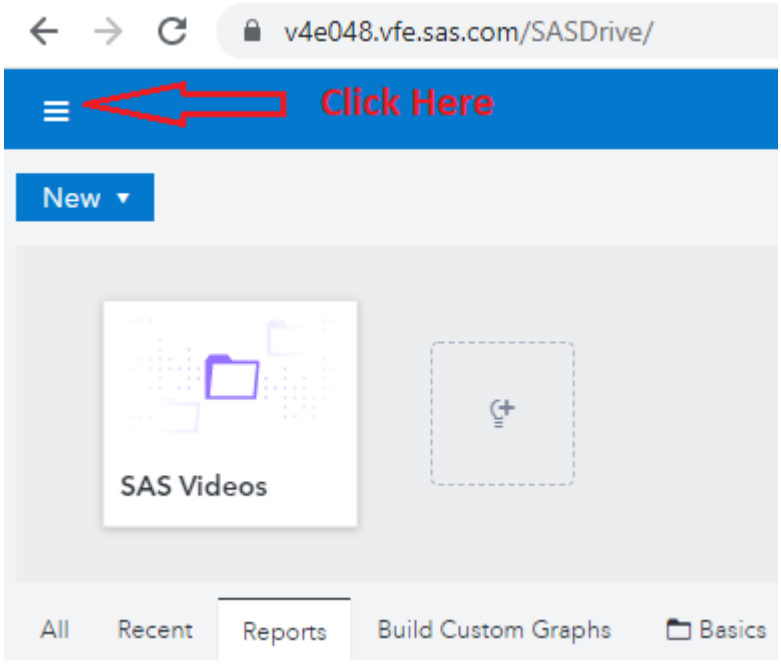
For more information about the suggested objects, see “About the Suggestions Pane” in the *SAS Visual Analytics: Working with Report Content* documentation.

Accessing and Investigating Data

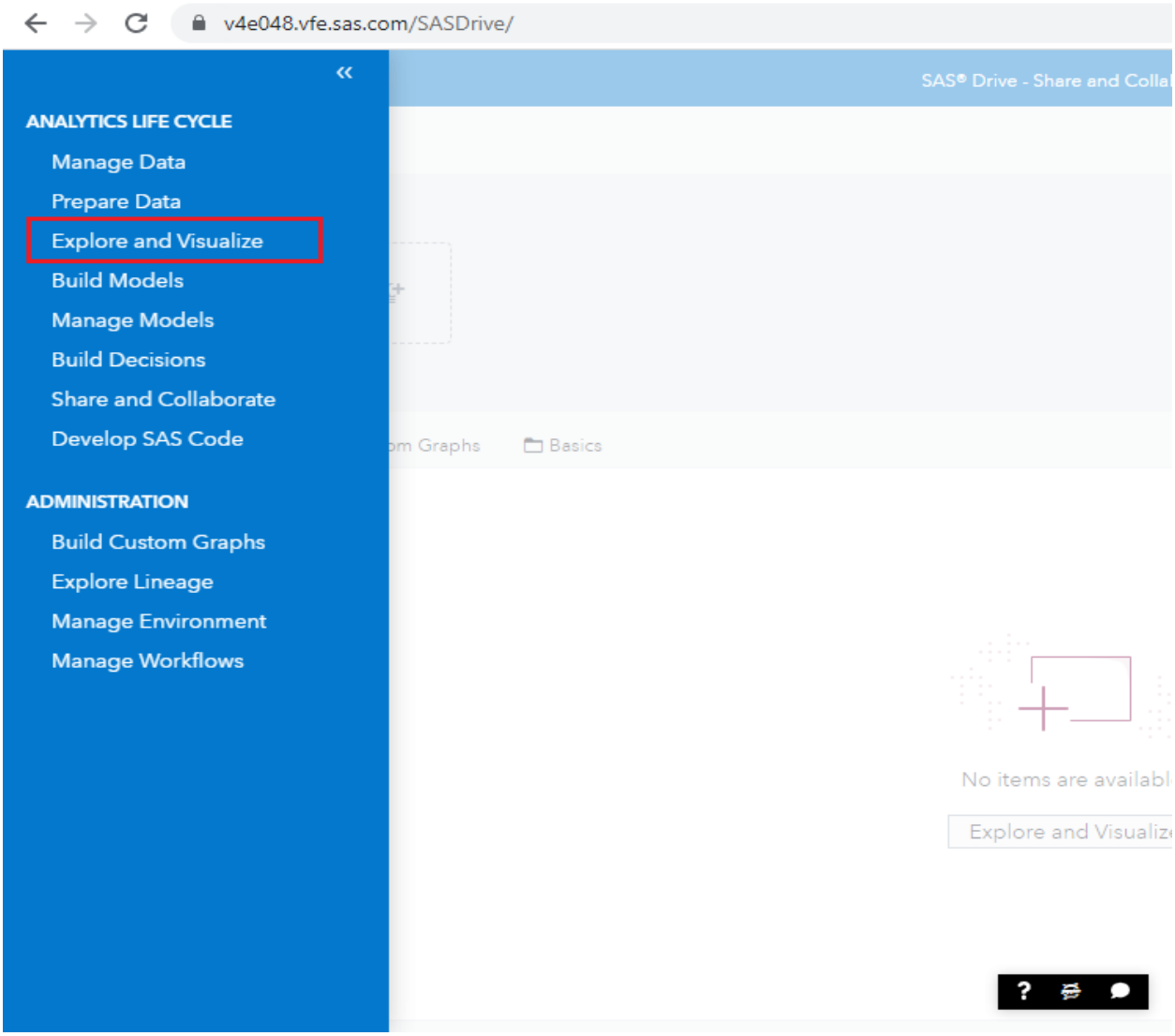
This demonstration illustrates how to access data in Visual Analytics through self-service import and how to investigate data for the business scenario.

2. Open a Chrome browser. Go to URL: https://vle.sas.com/vfl
3. Log in.

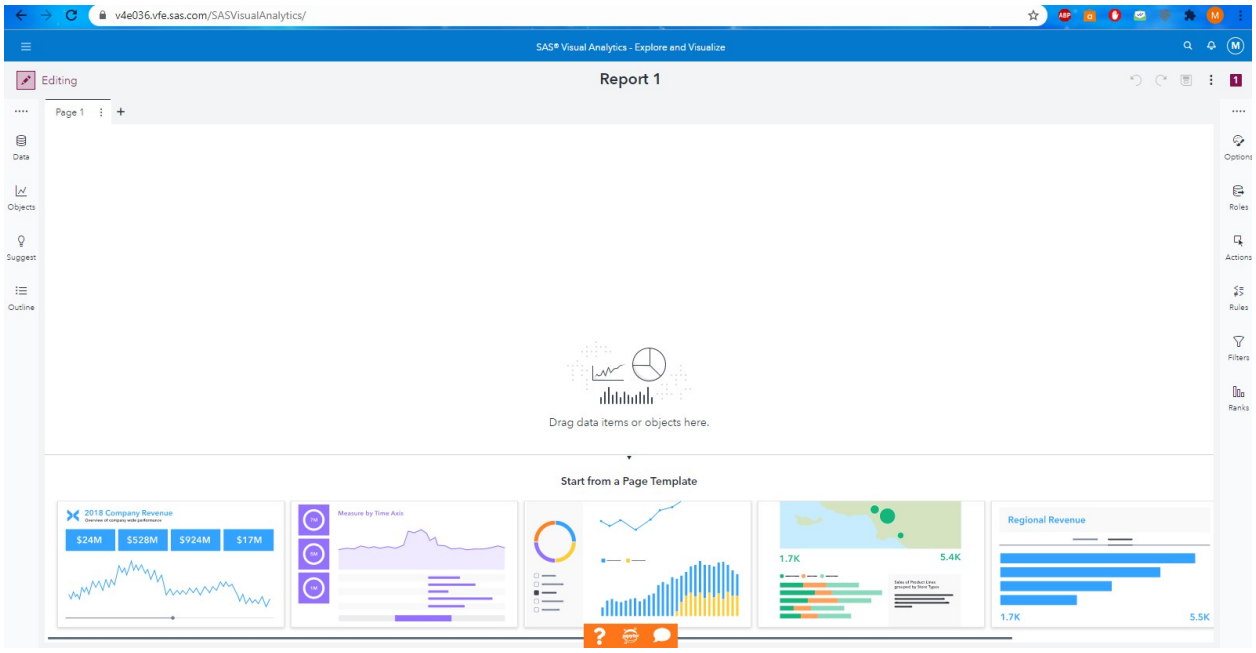
1. To select **New Report** in the top left corner of SAS Drive, do as below indicated in screenshot.



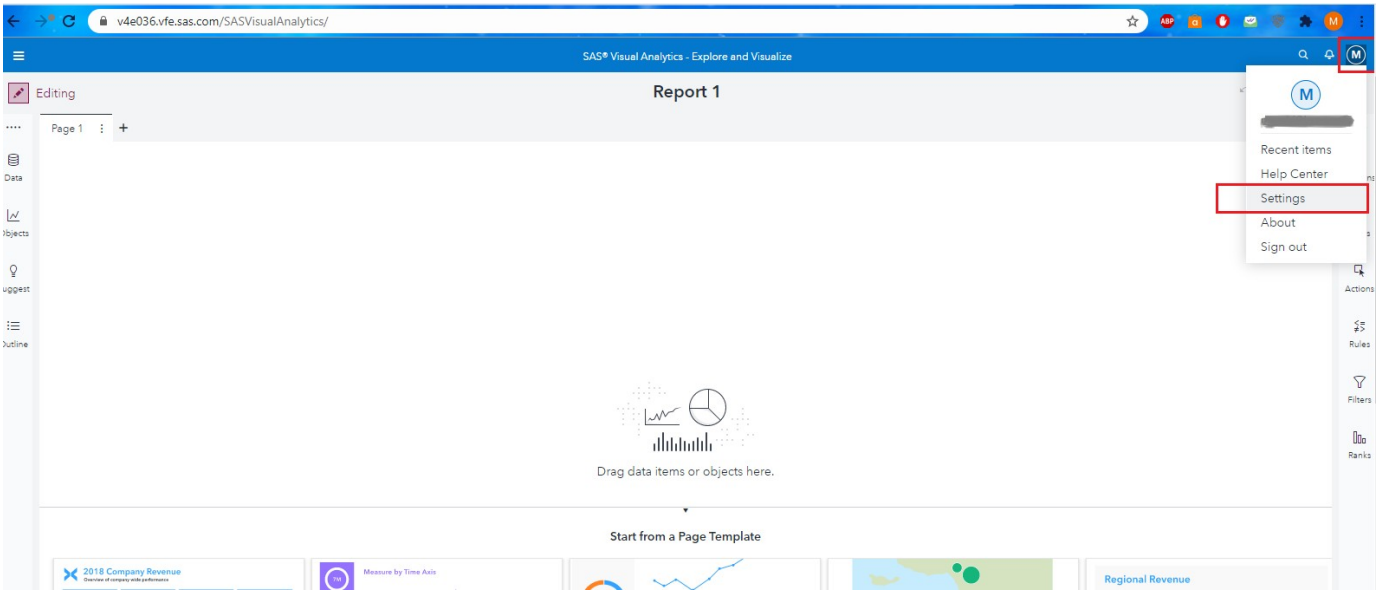
2. Then click on **Explore and Visualize** as below**.**



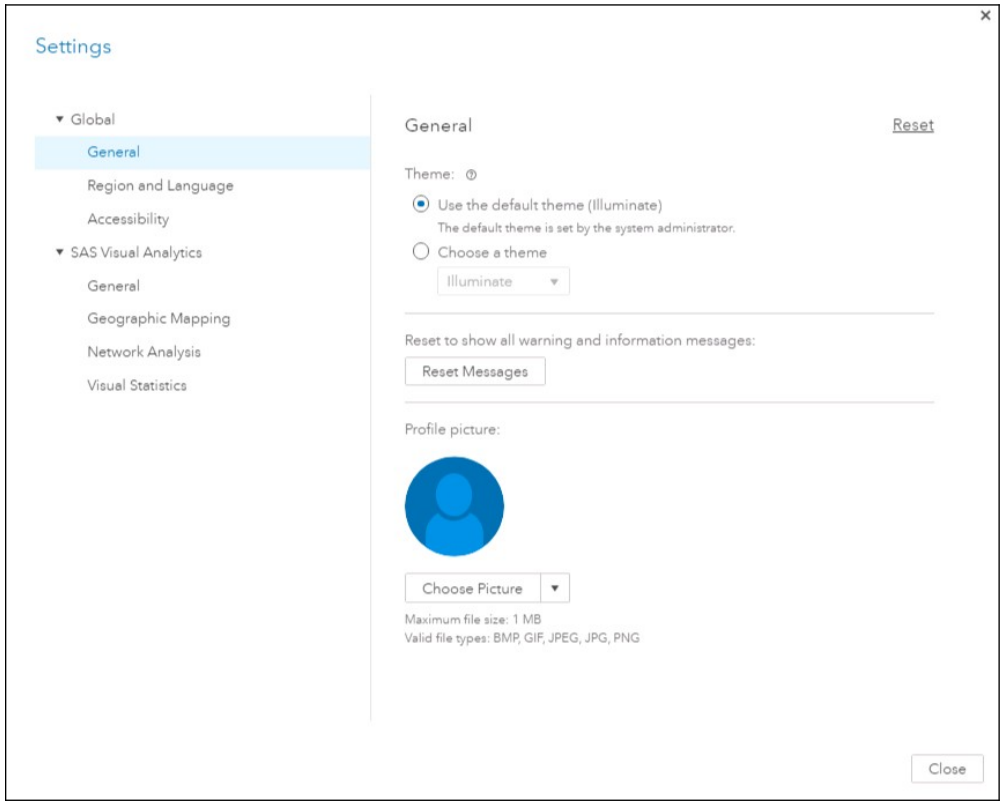
The Welcome to SAS Visual Analytics window appears as below.



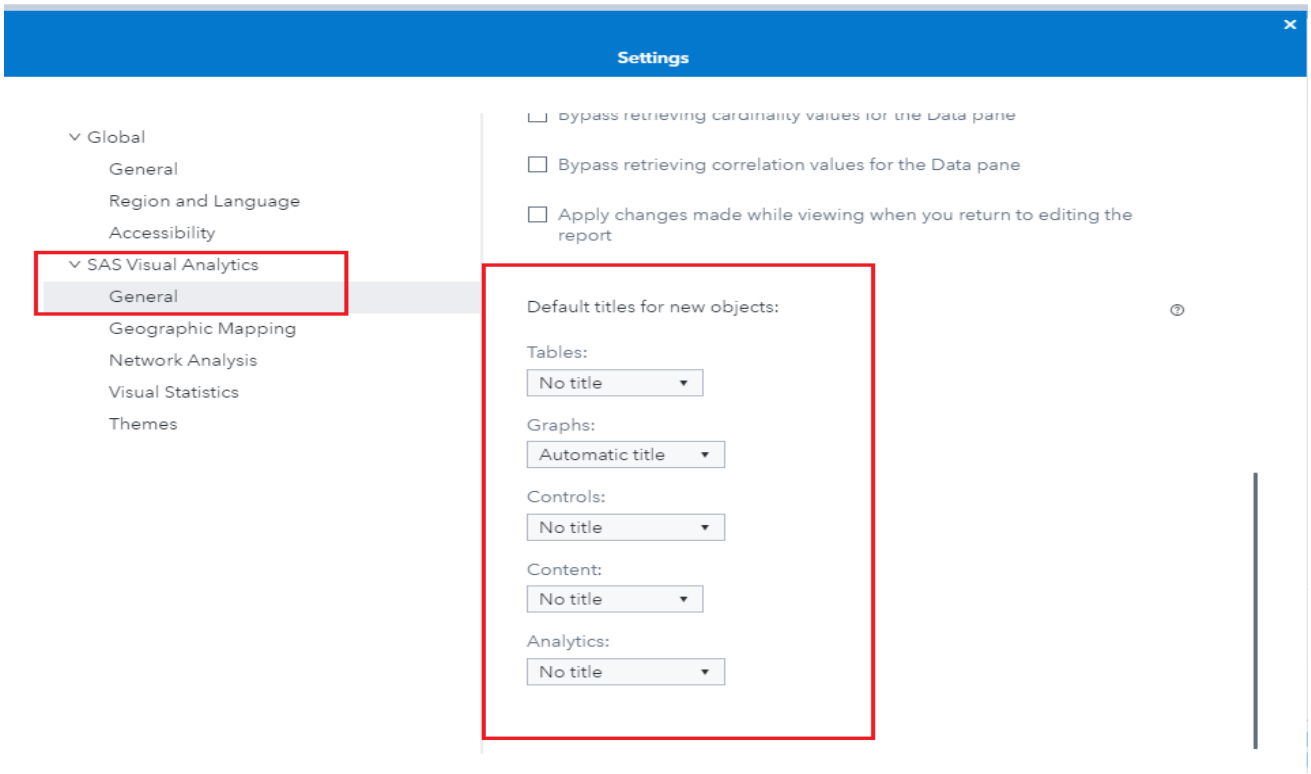
3. Check on your account profile here indicated as **M**, for SAS Visual Analytics as indicated below, then click **Settings.**



The Settings window appears as below.

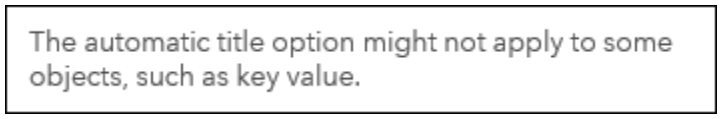


a. Select **General** under **SAS Visual Analytics** as below

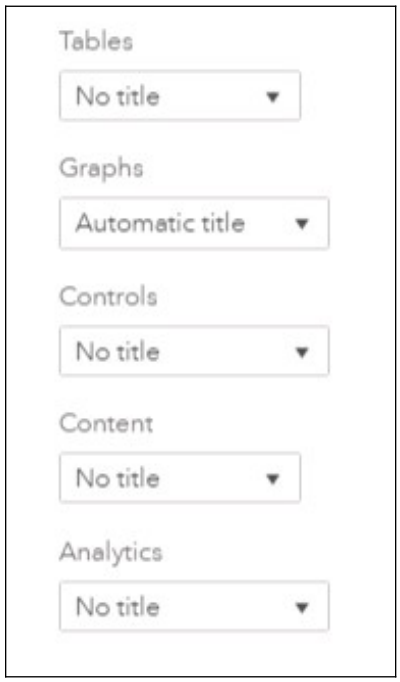


b. Scroll down to the right pane Settings section titled “Default titles for new object”

c. Click (**More information about this option**).



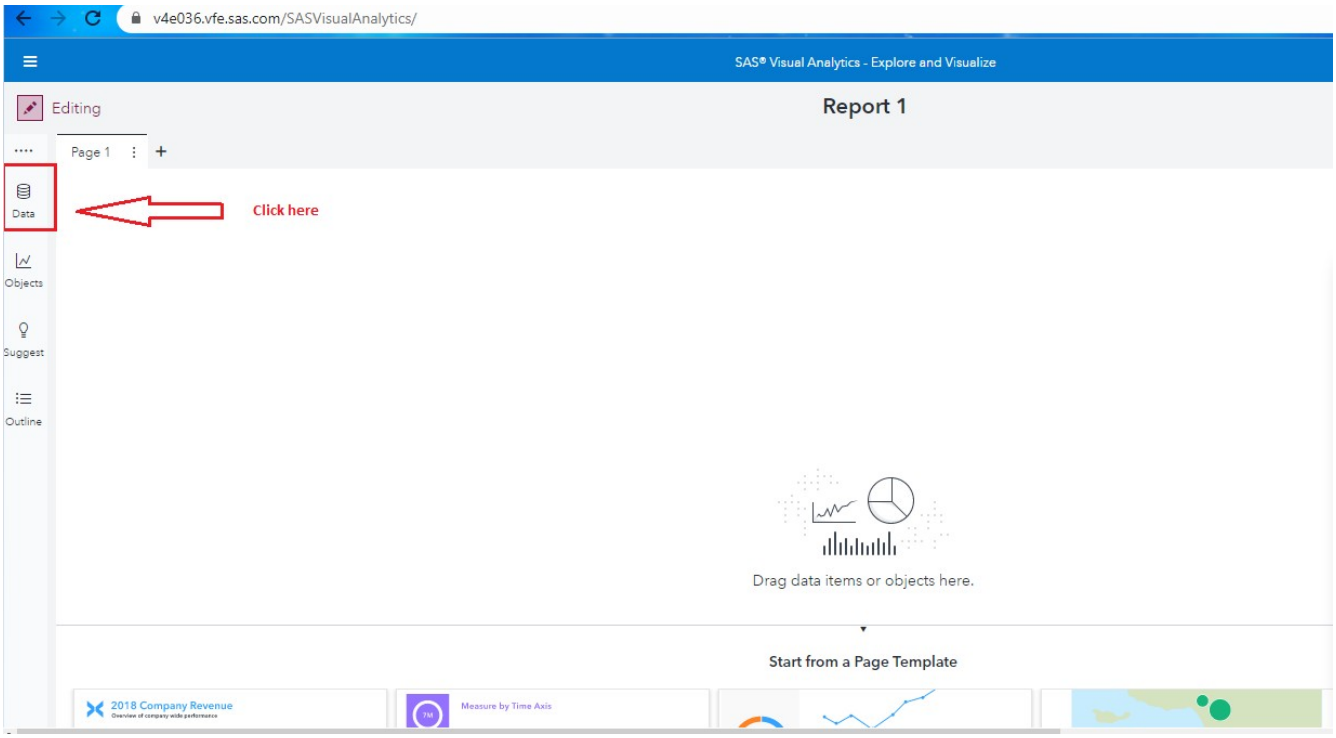
These options enable you to add automatic titles to specific objects in your report. Notice that automatic titles will be added to all Graph objects. **If ‘Automatic titles’ is not selected by default, select it from the drop down menu of “Graphs”**



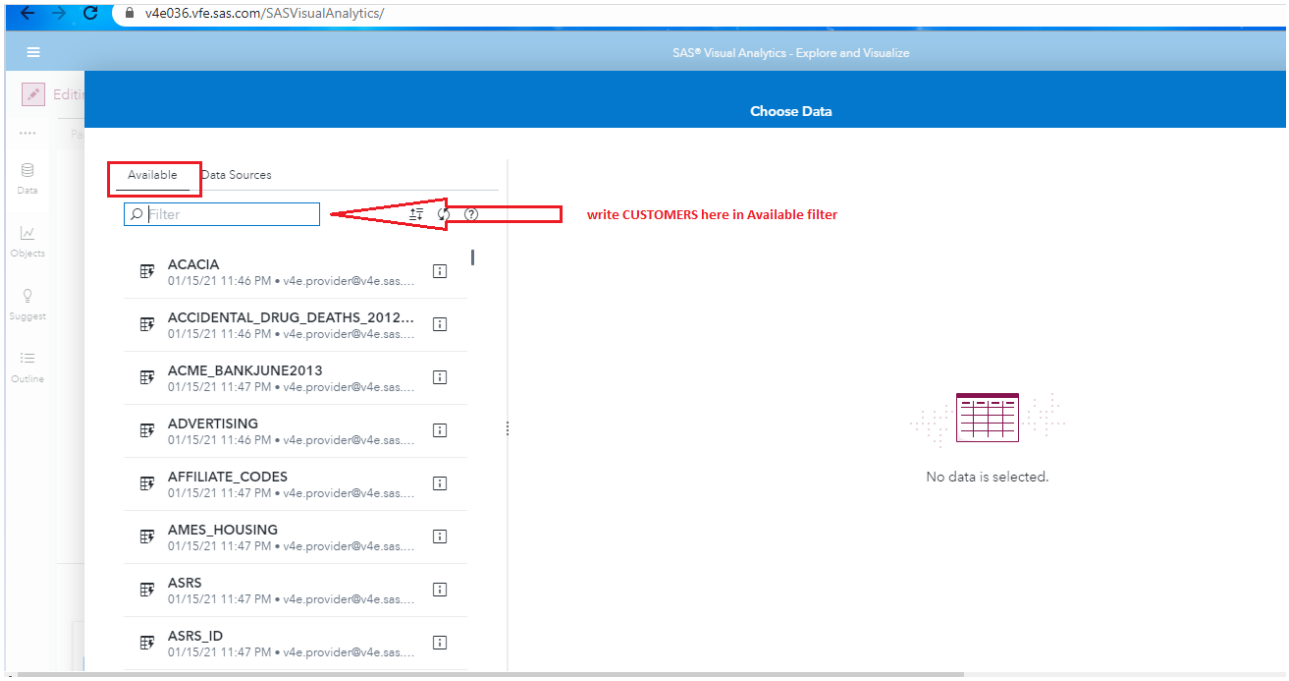
d. Click **Close** to close the Settings window.

4. Import a SAS data set into CAS.

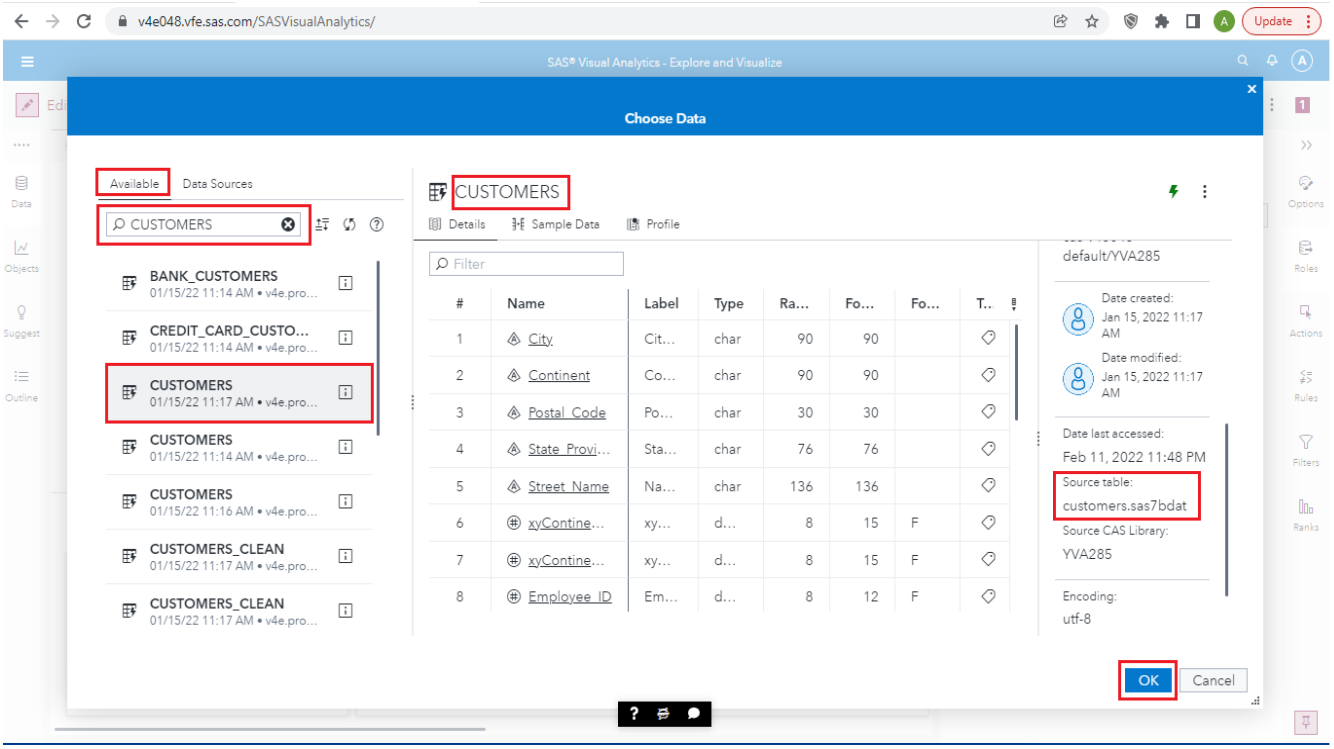
a. Click the **Data** icon in the left pane.



b. In the Choose Data window, click on **Available**, then write **CUSTOMERS** in the filter box as indicated below:

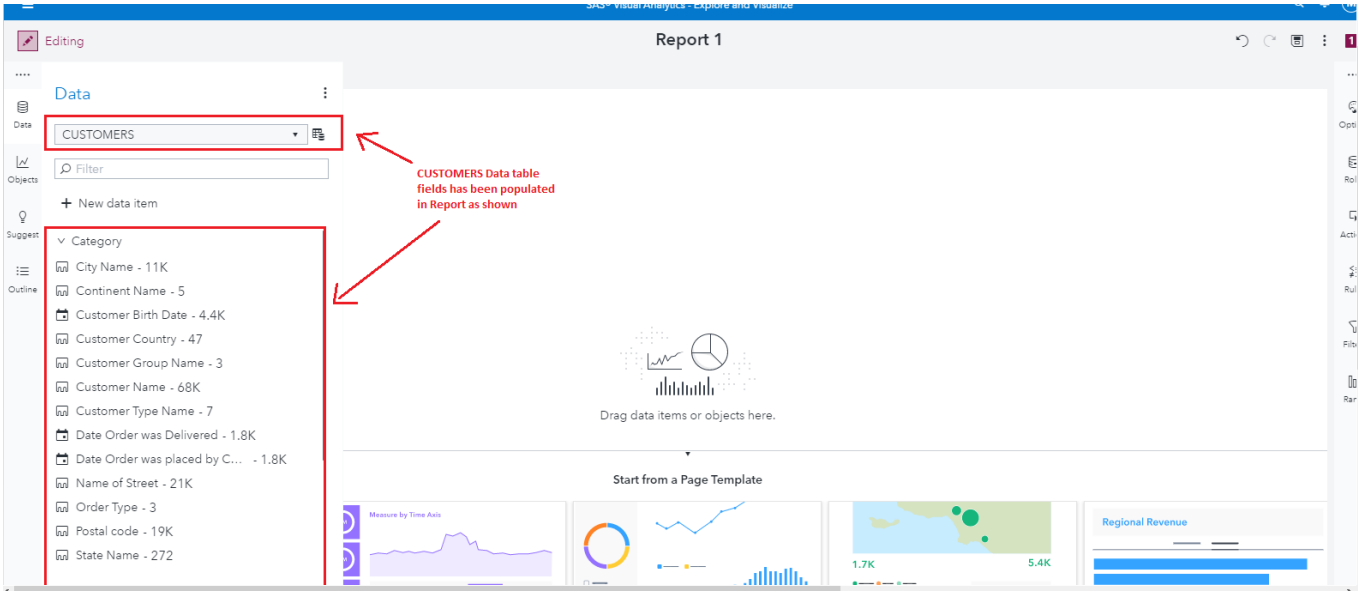


c. Once you filter with **CUSTOMERS,** below list will populate in the current window, then Click on CUSTOMERS table as indicated with **most recent updated date** from the list.

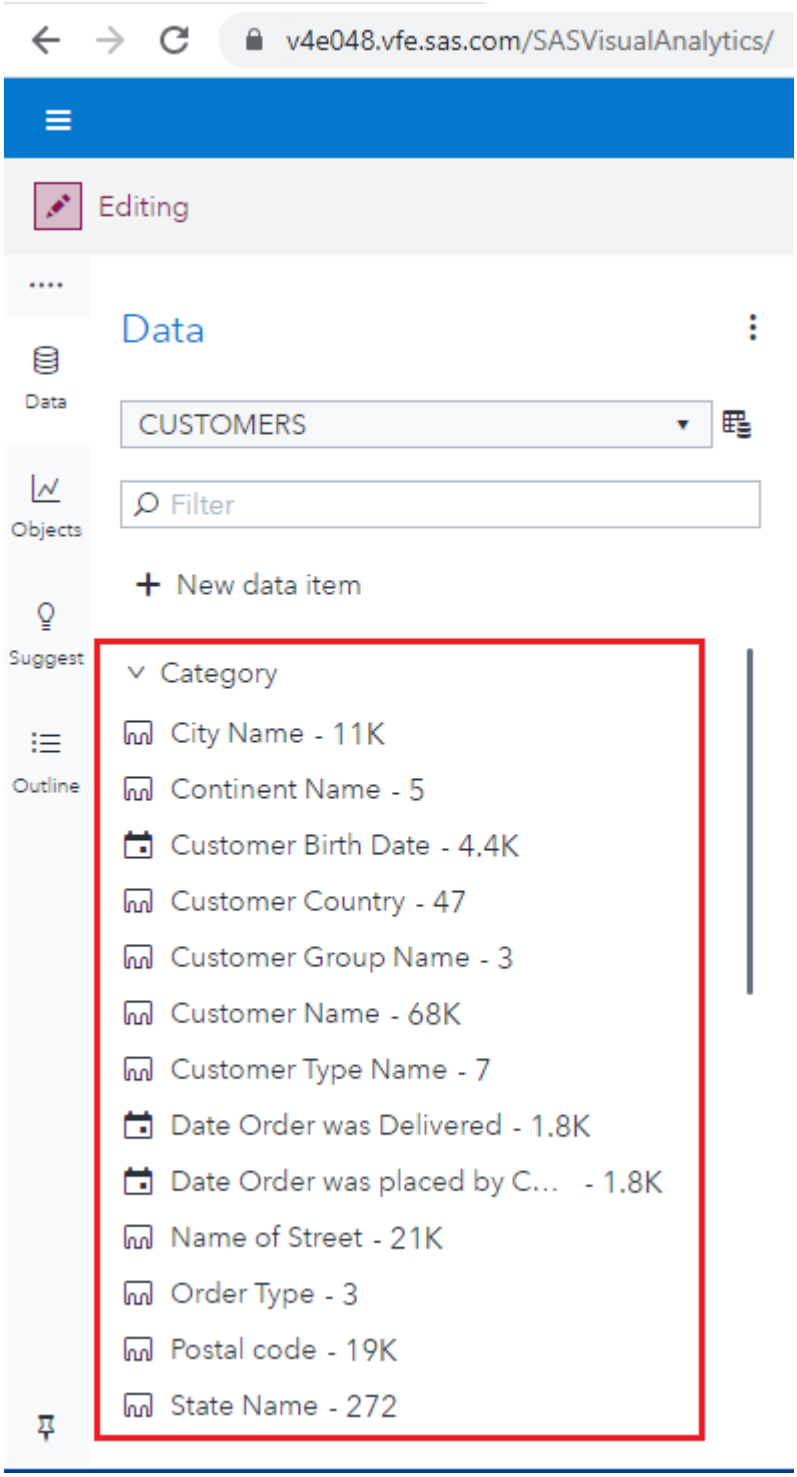


d. You will also see the source table as **customers.sas7bdat** table and then click **OK.**

d. The table is imported and is available to use with Visual Analytics. The Data pane is displayed and lists the data items from the **CUSTOMERS** table.

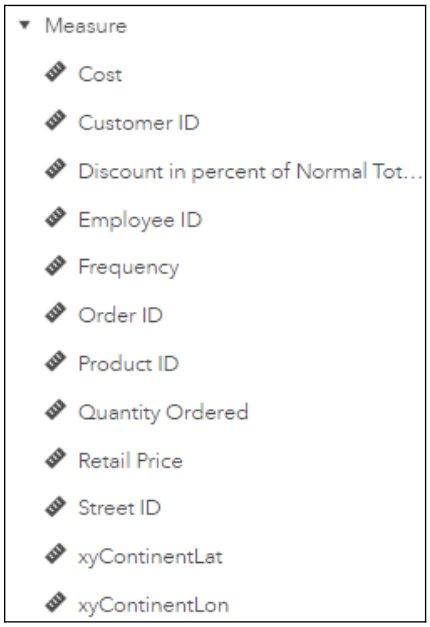


5. View the list of data items in the Category group.

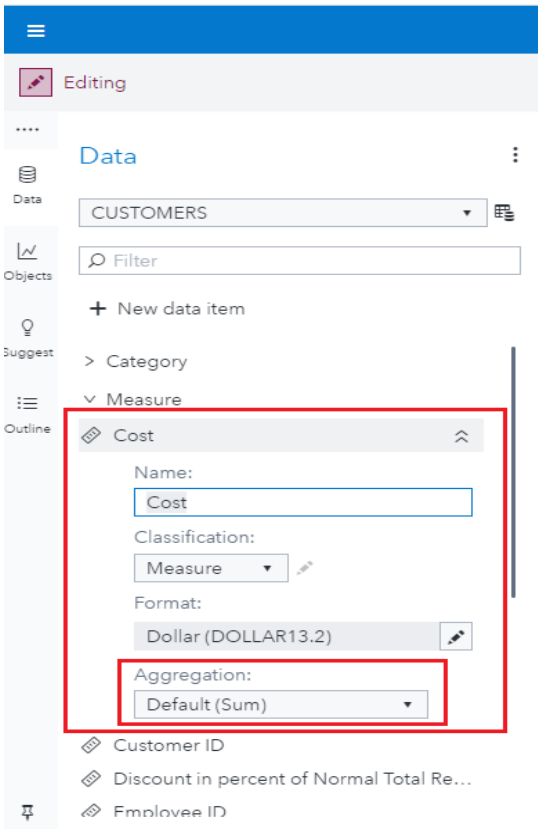


Character variables and numeric variables with a date format associated with them appear as categories in Visual Analytics. Distinct counts appear next to each category.

6. Scroll down in the **Data pane** to view the list of data items in the Measure group.



Numeric variables appear as measures in Visual Analytics. By default, all measures have an aggregation of Sum. You can check the same by clicking the down arrow as indicated below:

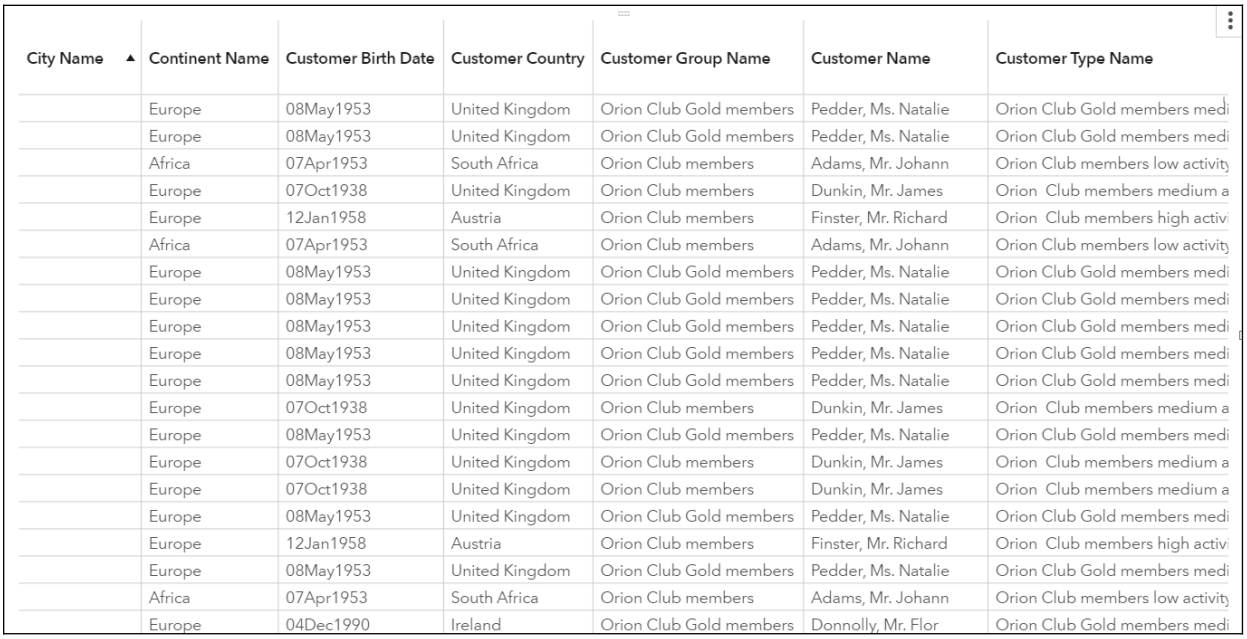


7. Use the list table object to view the imported table.
   a. In the left pane, click the **Objects** icon.
   b. Drag the **List Table** object, from the Tables group, to the canvas.
   c. In the right pane, click the **Roles** icon.
   d. For the **Columns** role, click **Add**.
   e. Select all data items ***except*** **Frequency** and **Frequency Percent**.

      Note: To make multiple selections, click the first column, hold down the Shift key, and click the last column.

   f. Click **OK**.

      The list table should resemble the following:



City Name	Continent Name	Customer Birth Date	Customer Country	Customer Group Name	Customer Name	Customer Type Name
	Europe	08May1953	United Kingdom	Orion Club Gold members	Pedder, Ms. Natalie	Orion Club Gold members medi
	Europe	08May1953	United Kingdom	Orion Club Gold members	Pedder, Ms. Natalie	Orion Club Gold members medi
	Africa	07Apr1953	South Africa	Orion Club members	Adams, Mr. Johann	Orion Club members low activity
	Europe	07Oct1938	United Kingdom	Orion Club members	Dunkin, Mr. James	Orion Club members medium a
	Europe	12Jan1958	Austria	Orion Club members	Finster, Mr. Richard	Orion Club members high activi
	Africa	07Apr1953	South Africa	Orion Club members	Adams, Mr. Johann	Orion Club members low activity
	Europe	08May1953	United Kingdom	Orion Club Gold members	Pedder, Ms. Natalie	Orion Club Gold members medi
	Europe	08May1953	United Kingdom	Orion Club Gold members	Pedder, Ms. Natalie	Orion Club Gold members medi
	Europe	08May1953	United Kingdom	Orion Club Gold members	Pedder, Ms. Natalie	Orion Club Gold members medi
	Europe	08May1953	United Kingdom	Orion Club Gold members	Pedder, Ms. Natalie	Orion Club Gold members medi
	Europe	08May1953	United Kingdom	Orion Club Gold members	Pedder, Ms. Natalie	Orion Club Gold members medi
	Europe	07Oct1938	United Kingdom	Orion Club members	Dunkin, Mr. James	Orion Club members medium a
	Europe	08May1953	United Kingdom	Orion Club Gold members	Pedder, Ms. Natalie	Orion Club Gold members medi
	Europe	07Oct1938	United Kingdom	Orion Club members	Dunkin, Mr. James	Orion Club members medium a
	Europe	07Oct1938	United Kingdom	Orion Club members	Dunkin, Mr. James	Orion Club members medium a
	Europe	08May1953	United Kingdom	Orion Club Gold members	Pedder, Ms. Natalie	Orion Club Gold members medi
	Europe	12Jan1958	Austria	Orion Club members	Finster, Mr. Richard	Orion Club members high activi
	Europe	08May1953	United Kingdom	Orion Club Gold members	Pedder, Ms. Natalie	Orion Club Gold members medi
	Africa	07Apr1953	South Africa	Orion Club members	Adams, Mr. Johann	Orion Club members low activity
	Europe	04Dec1990	Ireland	Orion Club Gold members	Donnolly, Mr. Flor	Orion Club Gold members medi

   g. Scroll through the columns to view the data.

      The Marketing team has asked for customer data to analyze the following:

      - profits by age group
      - profits by gender

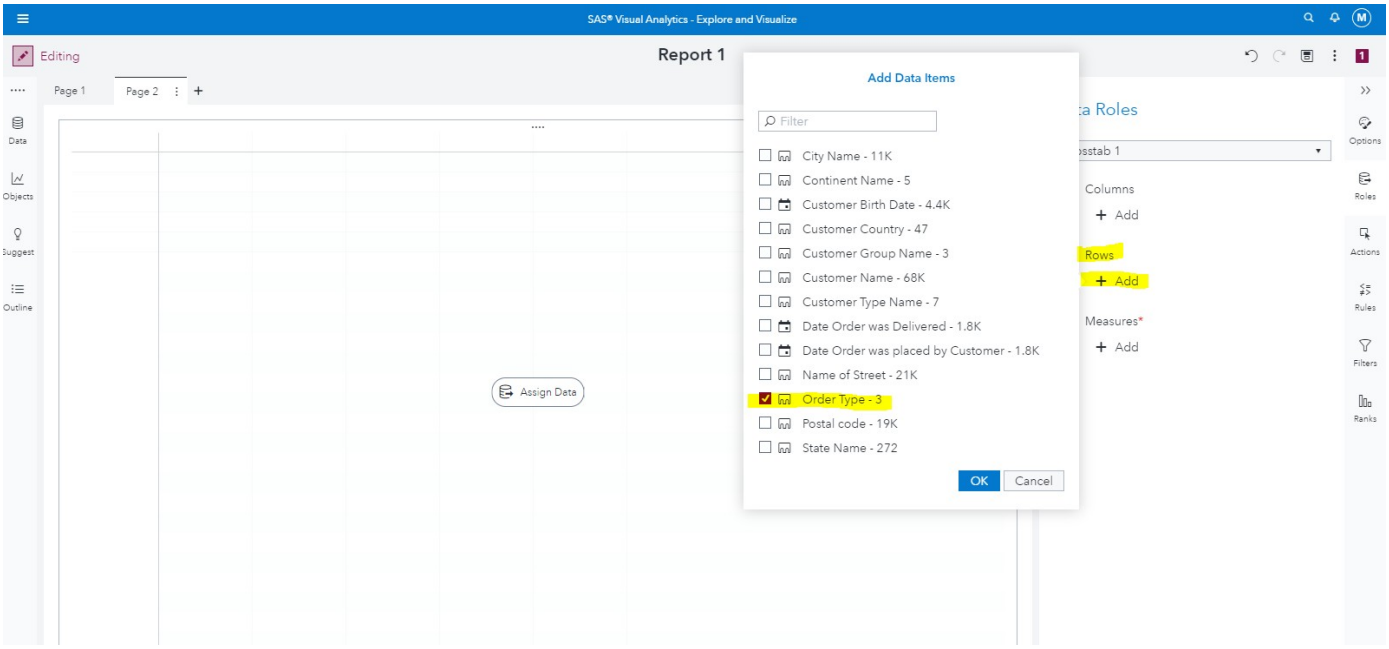
      The Shipping team has requested information about delivery times.

      Some data items (**Profit**, **Age Group**, **Gender**, and **Days to Delivery**) are not in the table but are needed for the analysis. The following existing data items can be used to create this information:

New data item	Contributing data items
Profit	**Cost** (unit cost), **Quantity Ordered**, **Retail Price** (total revenue)
Age Group	**Customer Birth Date**
Gender	**Customer Name** (*Pedder, Ms. Natalie* or *Finster, Mr. Richard*)

Days to Delivery	Date Order was Delivered, Date Order was placed by Customer

8. Click + (**New page**) in the upper left corner next to **Page 1**.
9. Use the crosstab object to view distinct values for order type.
   a. In the left pane, click the **Objects** icon.
   b. Drag the **Crosstab** object, from the Tables group, to the canvas.
   c. If necessary, click the **Roles** icon in the right pane.
   d. For the **Rows** role, select **Add** ⇨ **Order Type**. Below highlighted fields shows this formation:



   e. Click **OK**.

   The crosstab should resemble the following:

Order Type ▲	Frequency
Catalog Sale	127,129
Internet Sale	108,570
Retail Sale	715,970

   Order Type contains the method in which the order was placed: catalog, internet, or retail.

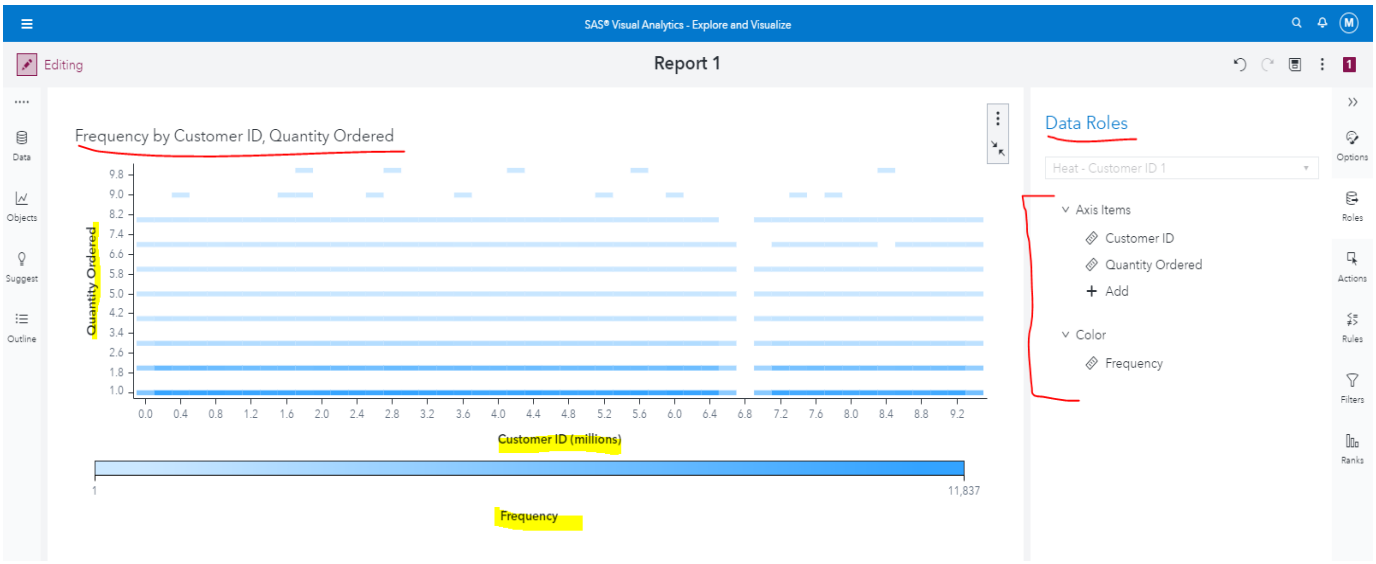
10. Use the automatic chart to view quantity ordered by customer.
    a. In the left pane, click the **Data** icon.
    b. Select the following data items, in the Measures group (in the order specified):

Customer ID

Quantity Ordered

c. Drag the selected items to the right of the crosstab.

The automatic chart should resemble the following:

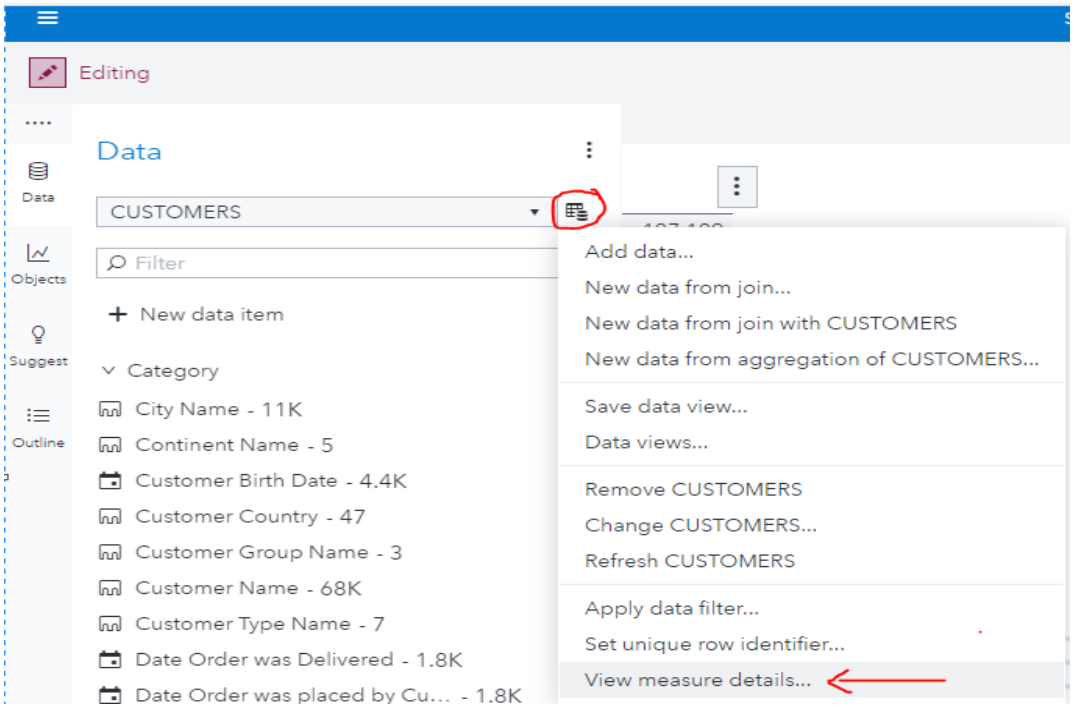


In Visual Analytics, all ID data items are classified as measures by default. An automatic chart of **Customer ID** and **Quantity Ordered** (two measures) yields a heat map that displays the relationship between the data items.

11. View descriptive information for the measure data items.

a. In the left pane, click the **Data** icon.

b. Click  (**Actions**) and select **View measure details**.



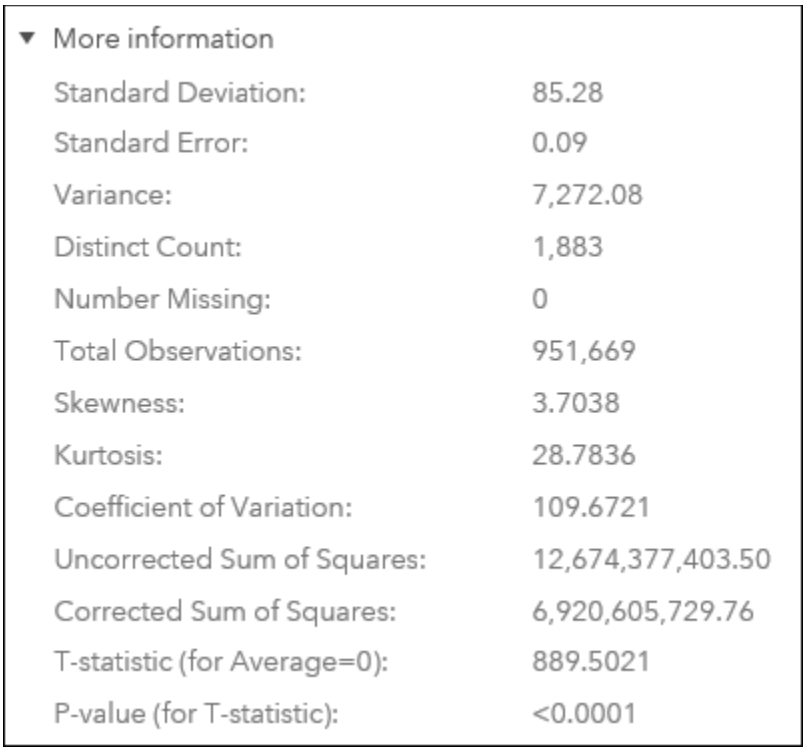
The **Measure Details window** displays descriptive statistics for each measure.

Measure Details

Name	Minimum	Maximum	Average	Sum
Cost	0.40	1,583.60	77.76	73,997,879.26
Customer ID	1.00	94,254.00	45,440.60	43,244,412,915.00
Discount in percent of Normal Total Retail Price	0.30	0.60	0.38	3,503.90
Employee ID	120,121.00	99,999,999.00	24,857,697.64	23,656,300,254,...

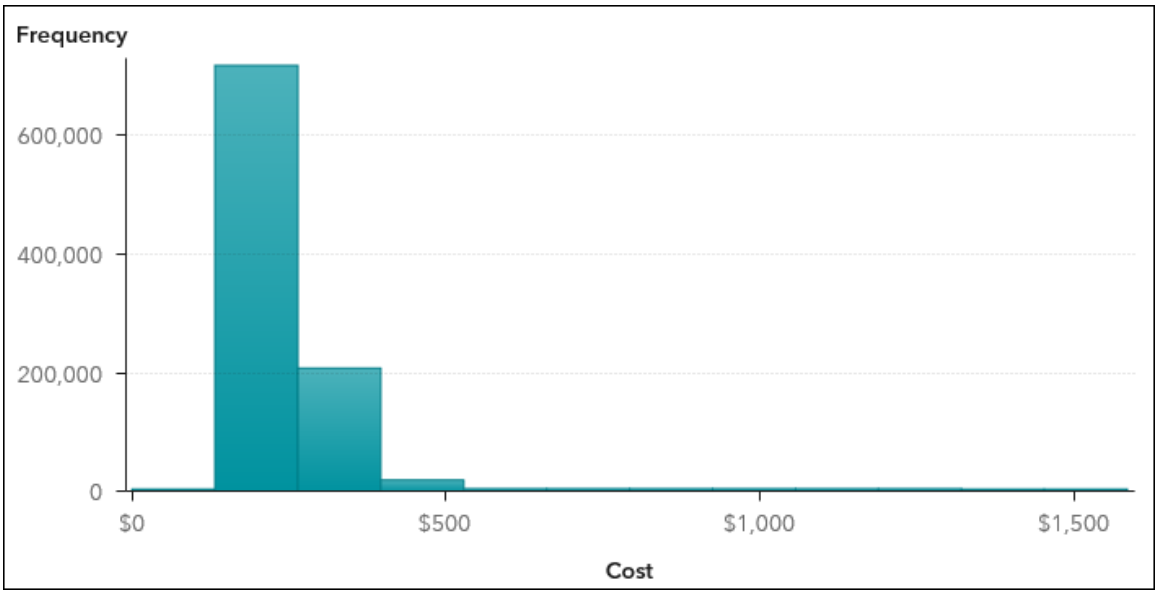
Note: **Customer ID**, **Employee ID**, and **Order ID** are numeric values and are classified as measures by default. They should be classified as categories because they should not be used in calculations. The results of summing or averaging these data items returns meaningless information.

c. With **Cost** selected in the **Measure Details window**, view the More information area.



Note: The number of rows (total observations) in the **CUSTOMERS** table appears in this list along with additional descriptive statistics for **Cost**.

d. View the graph on the right inside a canvas.

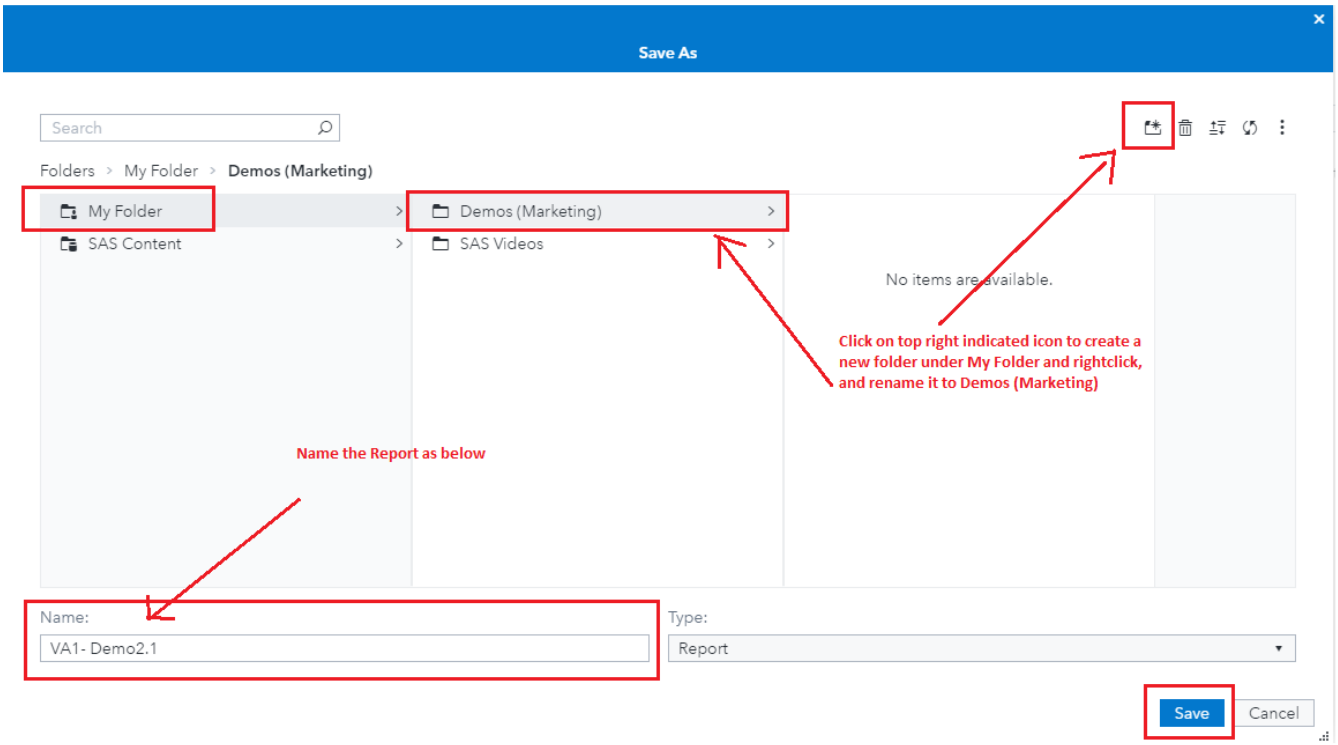


Note: The histogram displays the distribution of the **Cost** values.

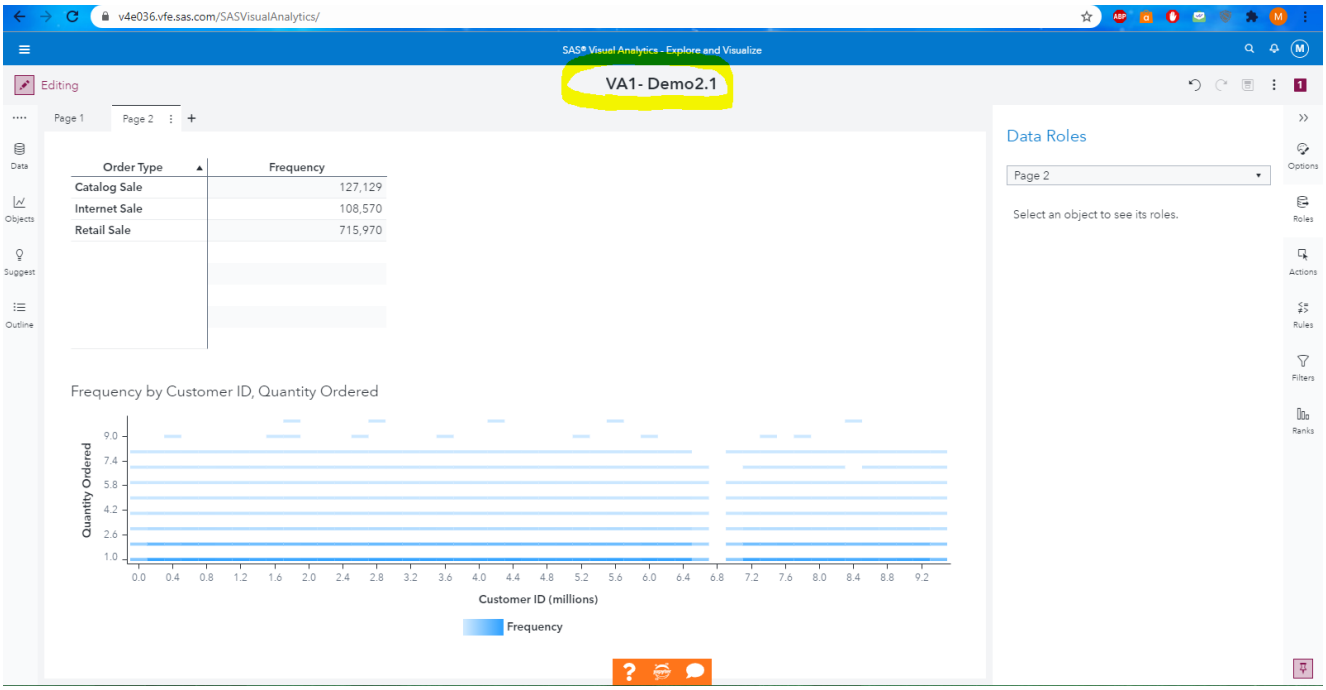
e. Click **Close** to close the Measure Details window.

12. In the upper right corner, click ⋮ (**Menu**) and select **Save as**.

13. Once Save As window pops up select **My Folder** → create a new folder by clicking on top rightmost first icon as indicated below and rename folder by **Right-Click** as name as **Demos (Marketing).**

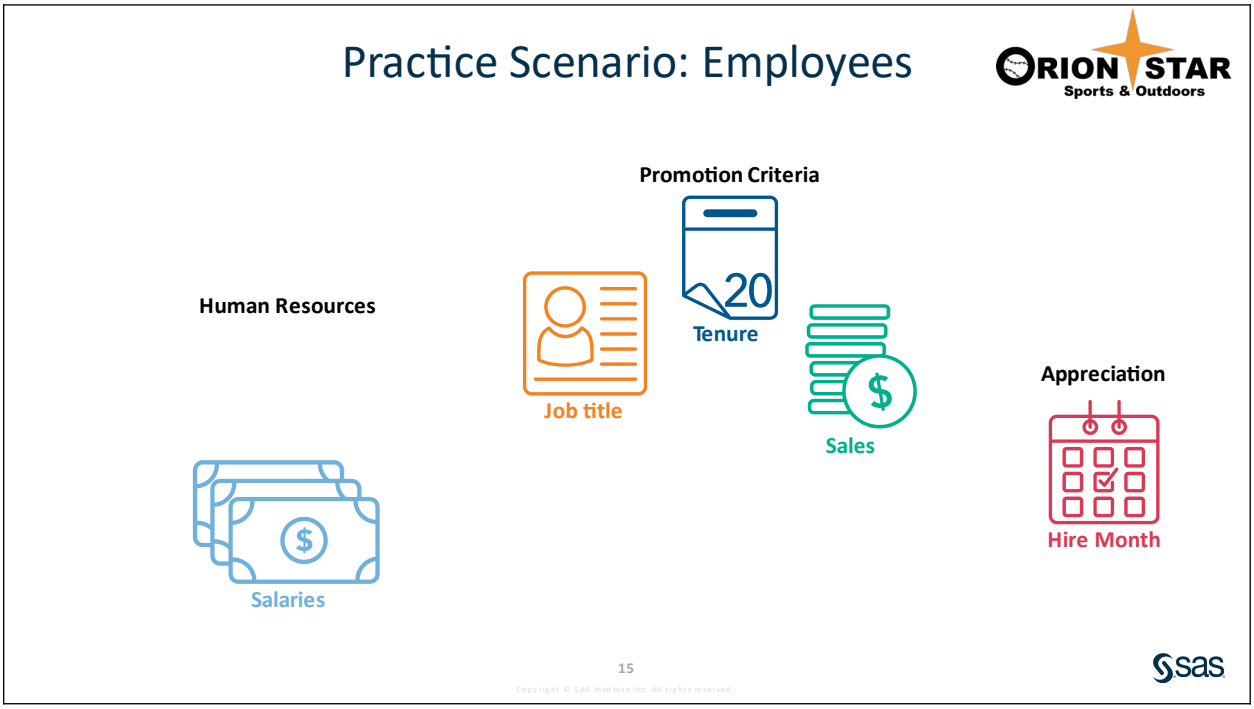


14. Enter **VA1- Demo2.1** in the **Name** field as indicated above screenshot.
15. Click **Save**.
16. Once you Save your screen will resemble as below:



17. Select **Your profile** ⇨ **Sign Out** in the upper right corner to sign out of SAS Visual Analytics.

End of Demonstration



Exercise 2: Provide answers to Questions 1-8. Support each answer with a relevant screenshot(s)

1. **Accessing and Investigating Data**
   - **a.** Open the browser and sign in to SAS Viya. SAS Drive is displayed by default.
   - **b.** In the upper left corner, click ☰ (**Show list of applications**) and select **Explore and Visualize**. SAS Visual Analytics appears.
   - **c.** Open the **VA1- Practice2.1** report in the **Courses/YVA185/Basics/Practices (HR)** folder.
     1) Click **All Reports**.
     2) Navigate to the **Courses**/**YVA185/Basics/Practices (HR)** folder.
     3) Double-click the **VA1- Practice2.1** report to open it.
   - **d.** View the Data pane and answer the questions.
     - Verify that you are editing the report.
     - In the left pane, click **Data**.

     Answer the following questions:

1. How many unique values does **Company** have? **Job Title**?
   - Hint: View the list of Category data items on the Data pane

   Answer 1:

   SS1:
2. What is the type (or classification) of **Employee ID**?
   - Hint: View the list of measure data items on the Data pane.

   Answer 2:

   SS2:

   - **e.** View the list table of all data items on Page 1 and answer the questions.
     - If necessary, click the **Page 1** tab at the top of the canvas.

The list table should resemble the following:

Company ▲	Department	Employee Birth Date	Employee Country	Employee Hire Date
Logistics	Stock & Shipping	.		.
Orion Australia	Sales	06Jun1953	au	01Jan1978
Orion Australia	Sales	22Apr1990	au	01Oct2010
Orion Australia	Sales	20Jul1948	au	01Jan1978
Orion Australia	Sales	18Aug1990	au	01Sep2010
Orion Australia	Sales	09Nov1990	au	01Nov2010
Orion Australia	Sales	21Feb1990	au	01Dec2010
Orion Australia	Sales	21Nov1978	au	01Jan1997
Orion Australia	Sales	07May1983	au	01Jan2002

- Scroll through the columns and answer the following questions:

3. What is the case of **Employee Country (Upper case /Lower case)**?

Answer 3:

SS3:

4.Which employee represents sales over the internet or through the catalog?

Answer 4:

SS4:

5.Which data item can be used to determine whether an employee is active (currently employed) or retired (formerly employed)?

Answer 5:

SS5:

View the crosstab of **Department** and **Job Title** on Page 2 and answer the question.

- Click the **Page 2** tab at the top of the canvas.
- View the crosstab and answer the following question:

6.Which department contains the missing job title?

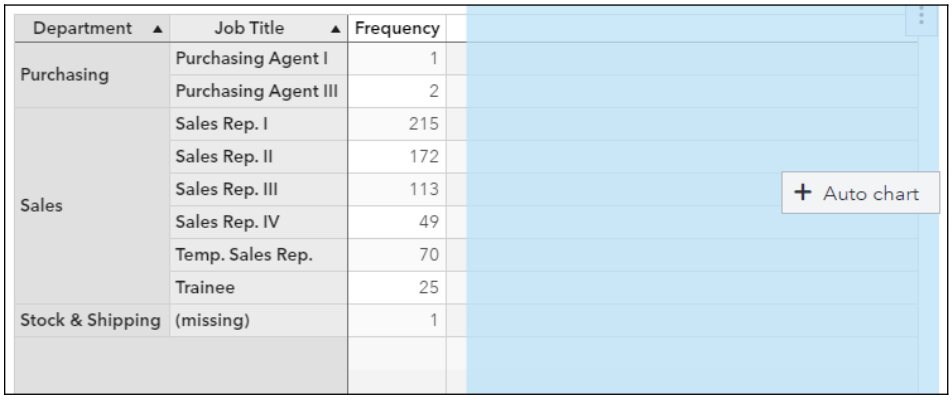
Answer 6:

SS6:

Create an auto chart of **Company** (on the right side of the crosstab) and answer the questions.

- In the left pane, click **Data**.

- Drag **Company** from the Data pane to the right side of the canvas.



7. What is the largest company, based on the number of employees? The smallest?
   - Hint: **Place your cursor over the bars to see the frequency.**

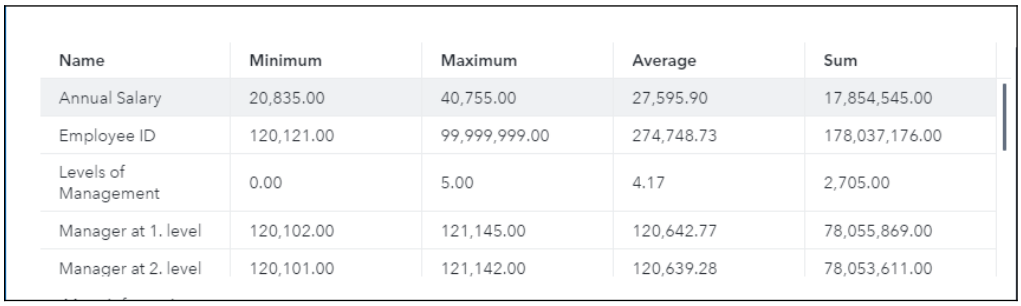
Answer 7:

SS7:

View the measure details (from the Data pane) and answer the questions.

- In the left pane, click **Data**.
- Click  (**Actions**) and select **View measure details**.

  The Measure Details table shows the minimum, maximum, average, and sum for each measure.



Name	Minimum	Maximum	Average	Sum
Annual Salary	20,835.00	40,755.00	27,595.90	17,854,545.00
Employee ID	120,121.00	99,999,999.00	274,748.73	178,037,176.00
Levels of Management	0.00	5.00	4.17	2,705.00
Manager at 1. level	120,102.00	121,145.00	120,642.77	78,055,869.00
Manager at 2. level	120,101.00	121,142.00	120,639.28	78,053,611.00

8. What is the minimum total profit generated by an employee? The maximum? The average? The total profit generated by all employees?

Answer 8:

SS8: