UNSW Business School/
Information Systems and Technology Management

SAS Viya for Learners – SAS Visual Analytics Workbook 1

Analyzing Data Using SAS Visual Analytics

SAS Visual Analytics Workbook:

Compiled/Modified By	Date	SAS Visual Analytics
Jacky Mo	Sep. 2021	SAS Viya for Learners

All the SAS Visual Analytics Workbooks will help the students to learn and gain experience and skills in data preparation; data exploration; creating reports; and constructing dashboard.

Reference:

This learning material is extracted from SAS® Academic Hub (LWYVA185) with the permission from SAS Australia to use and publish for teaching purpose at the University of New South Wales.

File Name:

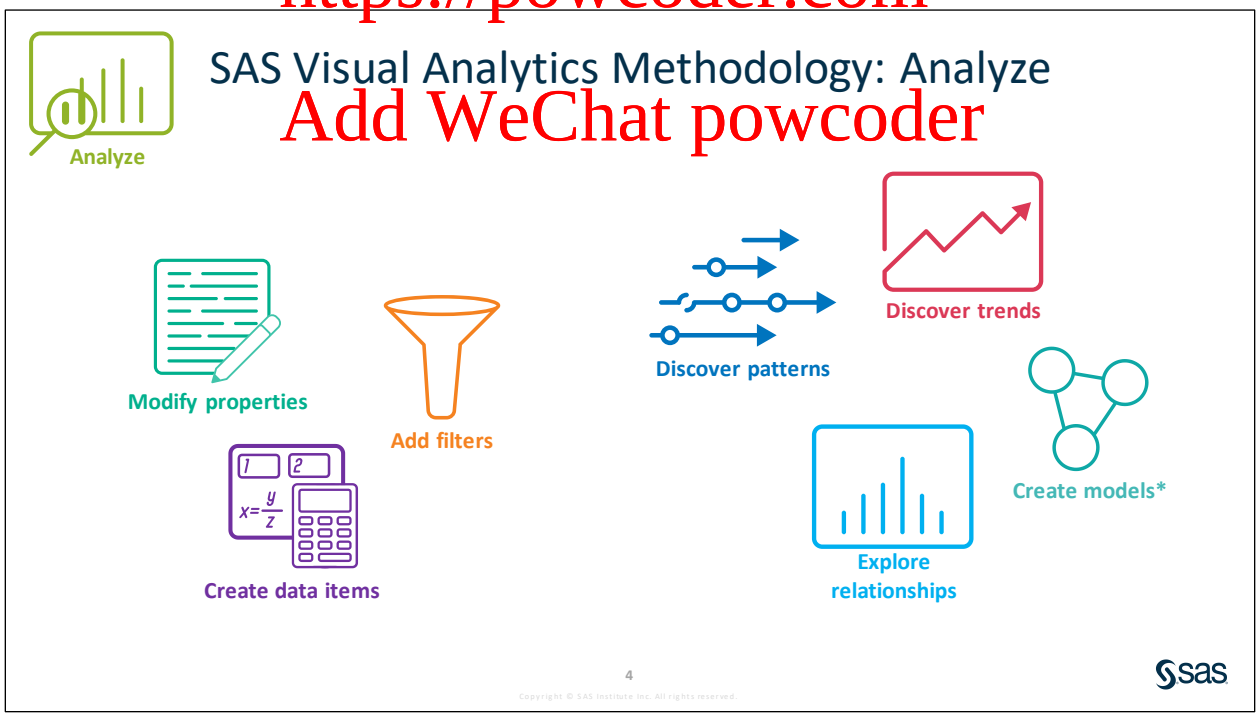
SAS Viya for Learners – SAS Visual Analytics Workbook 1

Copyright:

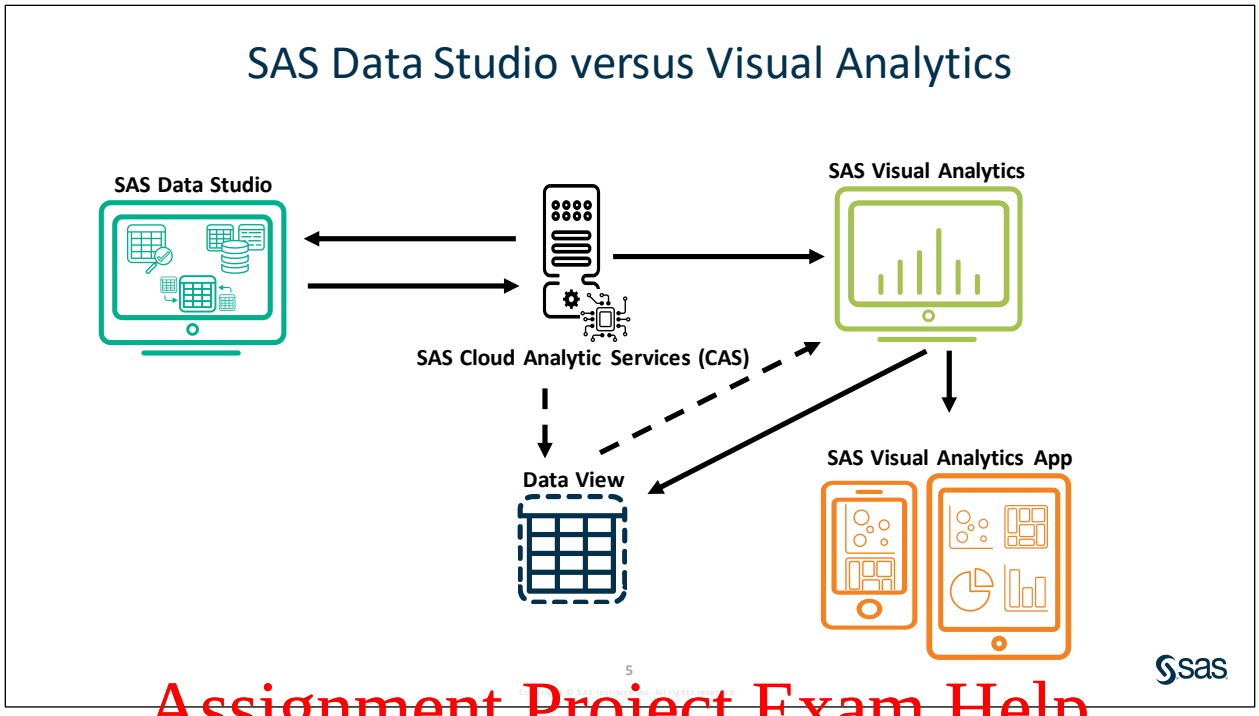
This learning material is created for Business Analytics courses offered by the School of Information Systems and Technology Management, the University of New South Wales, Sydney, Australia.

Lesson 3 Analyzing Data Using SAS® Visual Analytics

3.1 Working with Data Items



* Creating, testing, and comparing models can be accomplished with SAS Visual Statistics and SAS Visual Data Mining and Machine Learning.



SAS Visual Data Studio uses a CAS table as input and creates a CAS table as output.

SAS Visual Analytics uses a CAS table as input and creates a report that can be viewed in Visual Analytics or the SAS Visual Analytics app. Any changes to data made in Visual Analytics apply to the report only and do not affect the CAS table.

Beginning with Visual Analytics 8.3, report data views can be created to save and apply settings for a data source. A data view acts as a template for any settings that are modified, including data property changes, data source filters, hierarchies, geography data items, calculated items, and more. A data view does not update the CAS table. If the view is updated, your reports are not automatically updated with the new settings.

Data views are saved separately from your reports. If you create a data view in one report, you can apply it to other reports that use the same data source.

Data views can be shared by an application administrator so that other users can apply them to the data source.

A data source can have a default view as set by an application administrator. You can also set the default view for yourself. A default data view is automatically applied anytime that you add the data source to a report.

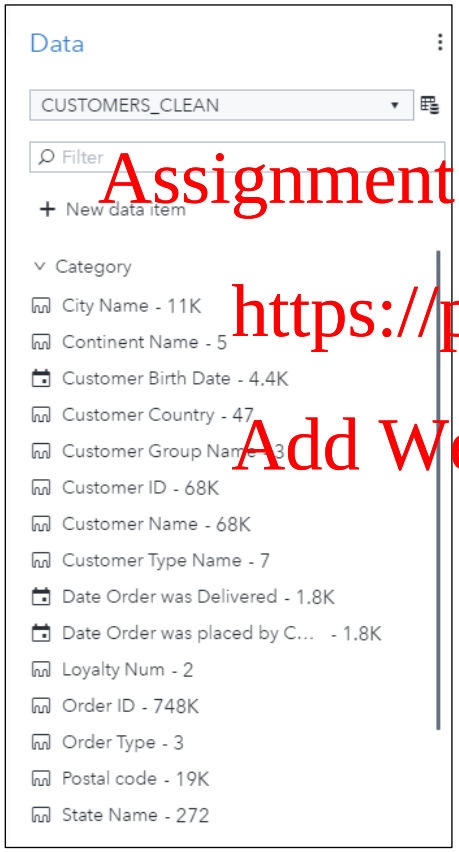
For more information about data views, see “Working with Data Views in Reports” in the *SAS Visual Analytics: Working with Report Data* documentation.

Working with Data Items

This demonstration illustrates how to modify data item properties (name, format, aggregation) in Visual Analytics.

1. From the browser window, sign in to SAS Viya.
2. In the upper left corner, click ≡ (**Show list of applications**) and select **Explore and Visualize**.

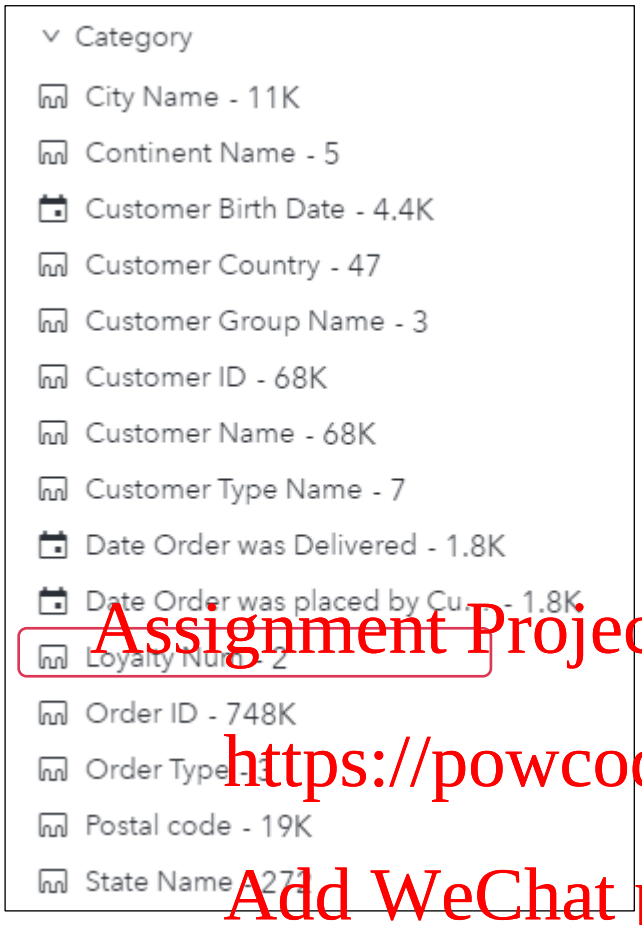
   SAS Visual Analytics appears.
3. Click **All Reports**.
   a. Navigate to the **Courses/YVA185/Basics/Demos (Marketing)** folder.
   b. Double-click the **VA1- Demo3.1** report to open it.
4. In the left pane, click **Data**.



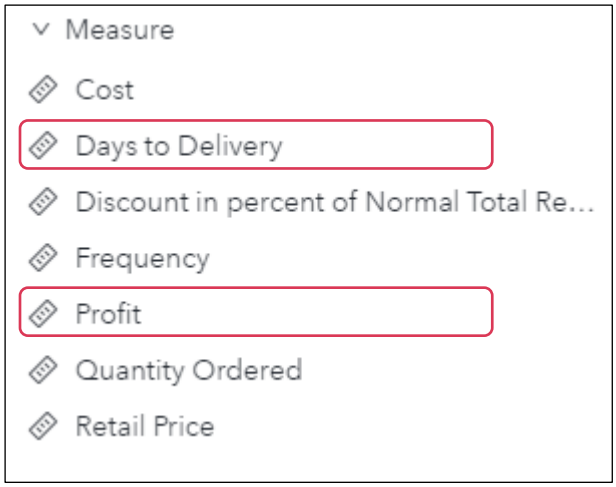
   The Data pane contains a list of data items from the **CUSTOMERS_CLEAN** table.
5. Verify that **Customer ID** and **Order ID** appear in the Category group, because the data type was changed to character in SAS Data Studio.

   Note: Character and datetime data items appear as categories in Visual Analytics.

6. Verify that the new column created in SAS Data Studio (**Loyalty Num**) appears in the Category group.



7. Verify that the new columns created in SAS Data Studio (**Days to Delivery** and **Profit**) appear in the Measure group.

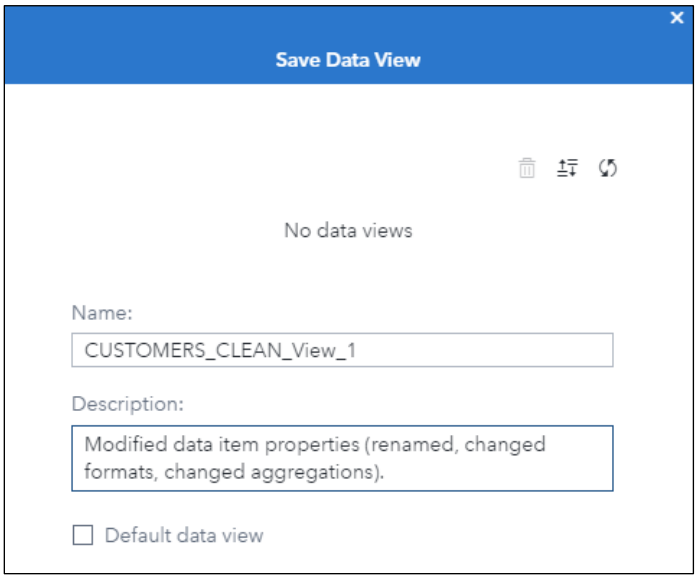


Note: Numeric (double) data items appear as measures in Visual Analytics.

Note: **Cost** and **Retail Price** were renamed in SAS Data Studio to **Unit Cost** and **Total Revenue**, respectively. Those new names are not reflected because Visual Analytics displays labels, not data source names.

8. Modify properties for a data item, **Date Order was Delivered**.
   a. In the Category group, right-click **Date Order was Delivered**.
   b. Select **Format** ⇨ **MMMYYYY (MONYY7)**.
   c. Next to **Date Order was Delivered**, click (**Edit properties**).
   d. In the **Name** field, enter **Delivery Date** and press the Enter key.
9. Modify properties for a data item, **Discount in percent of Normal Total Retail Price**.
   a. In the Measure group, next to **Discount in percent of Normal Total Retail Price**, click (**Edit properties**).
   b. For the **Aggregation** field, select **Average**.
   c. In the **Name** field, enter **Discount** and press Enter.
10. Modify the aggregation for a data item, **Days to Delivery**.
    a. In the Measure group, next to **Days to Delivery**, click (**Edit properties**).
    b. For the **Aggregation** field, select **Average**.
    c. In the **Name** field, enter **Average Days to Delivery** and press Enter.
11. Rename data items.
    a. In the Category group, next to **Date Order was placed by Customer**, click (**Edit properties**).
    b. In the **Name** field, enter **Order Date** and press Enter.
    c. In the Measure group, next to **Cost**, click (**Edit properties**).
    d. In the **Name** field, enter **Unit Cost** and press Enter.
    e. In the Measure group, next to **Quantity Ordered**, click (**Edit properties**).
    f. In the **Name** field, enter **Quantity** and press Enter.
    g. In the Measure group, next to **Retail Price**, click (**Edit properties**).
    h. In the **Name** field, enter **Total Revenue** and press Enter.
12. Create a data view.
    a. At the top of the Data pane, next to the table name, click (**Actions**) and select **Save data view**.
    b. For the **Name** field, verify that **CUSTOMER_CLEAN_View_1** is specified.

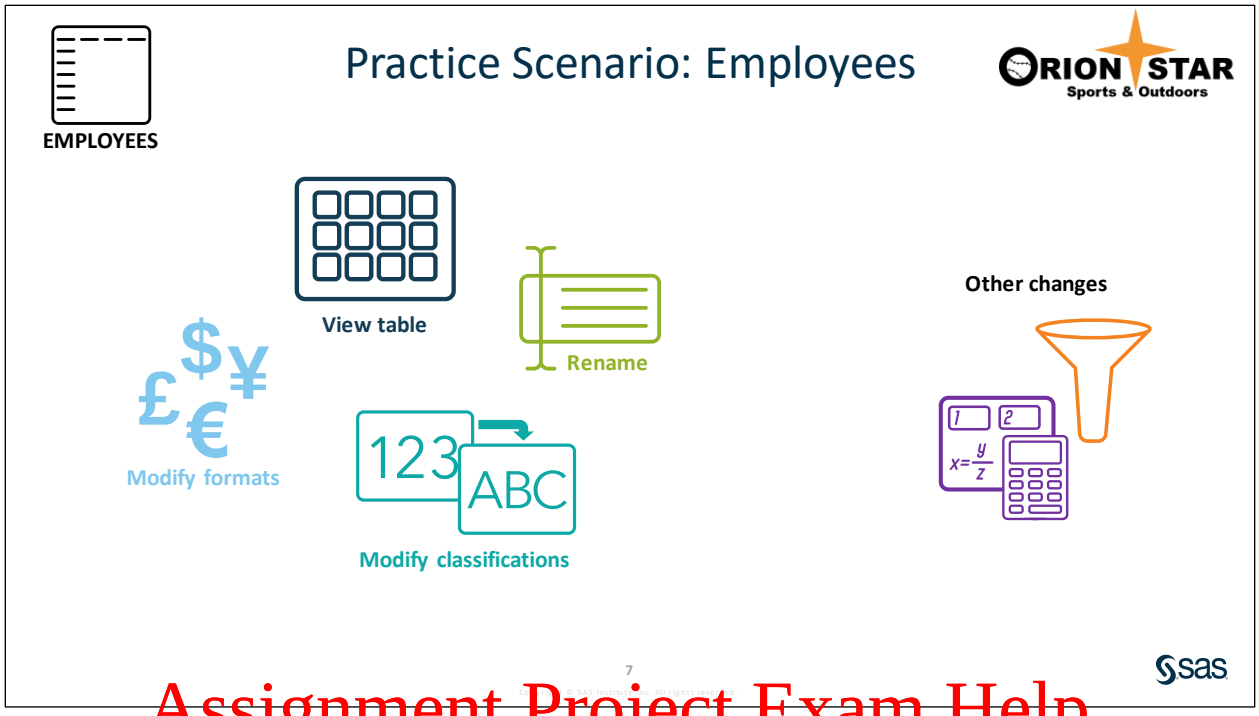
c. In the **Description** field, enter **Modified data item properties (renamed, changed formats, changed aggregations).**.



d. Click **Save**.

13. Save the report.

End of Demonstration



Practice

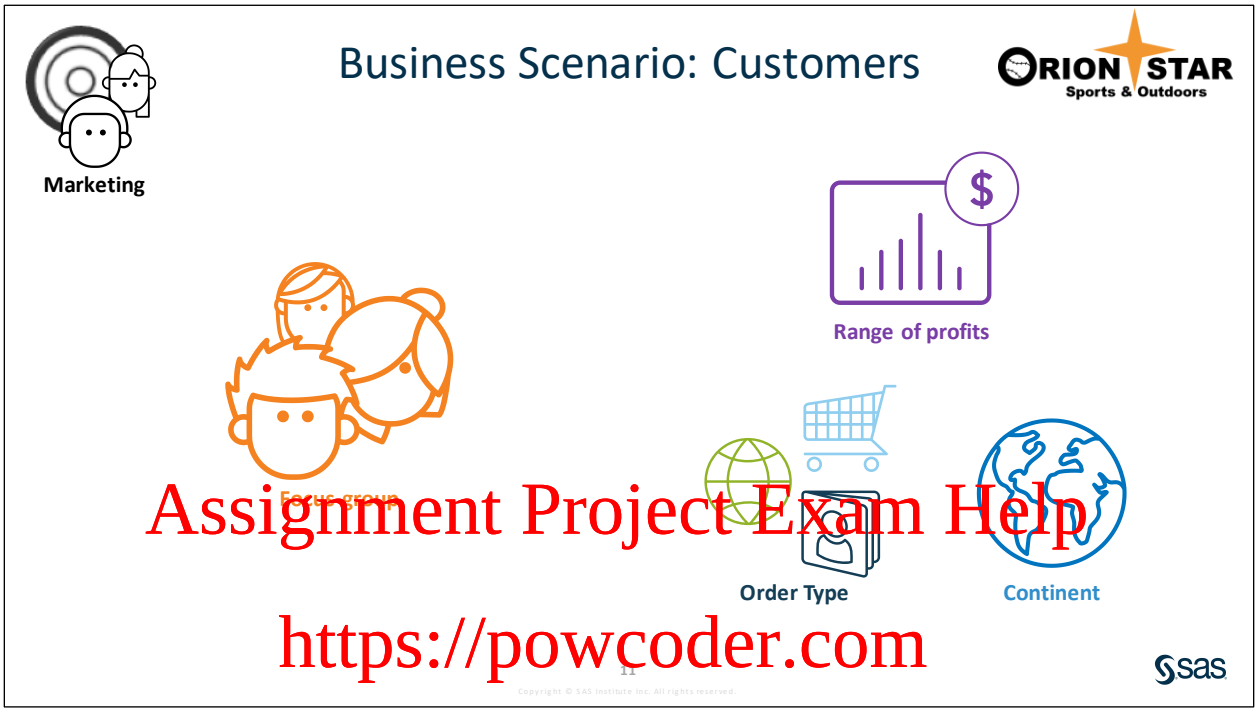
1. **Working with Data Items**
   a. Open the browser and sign in to SAS Viya.
   b. Open the **VA1- Practice3.1** report from the **Courses/YVA185/Basics/Practices (HR)** folder.
   c. View the data items in the Data pane and answer the following questions:
      What is the classification of **Employee ID**? **Manager at 1. level**?
      Answer: _____
      What does the **Frequency** data item represent?
      Answer: _____
   d. Change the classification for **Manager at 1. level** to **Category**.
   e. Change the format for **Annual Salary** to **Dollar13.2**.
   f. Rename the following data items:

Old name	New name
Employee ID	ID
Employee Name	Name
Manager at 1. level	Manager ID
Frequency	Number of Employees

   Note: Click (Actions) and select **Refresh EMPLOYEES_CLEAN** at the top of the Data pane to collapse the data item properties.
   g. Save the report.

End of Practices

3.2 Exploring Data with Charts and Graphs



3.01 Activity

Sign in to SAS Viya. Open the **VA1- Activity3.01** report (in the **/Courses/YVA185/Basics** folder).

What is the average of **Days to Delivery**?

Which factor is the most related to **Days to Delivery**?

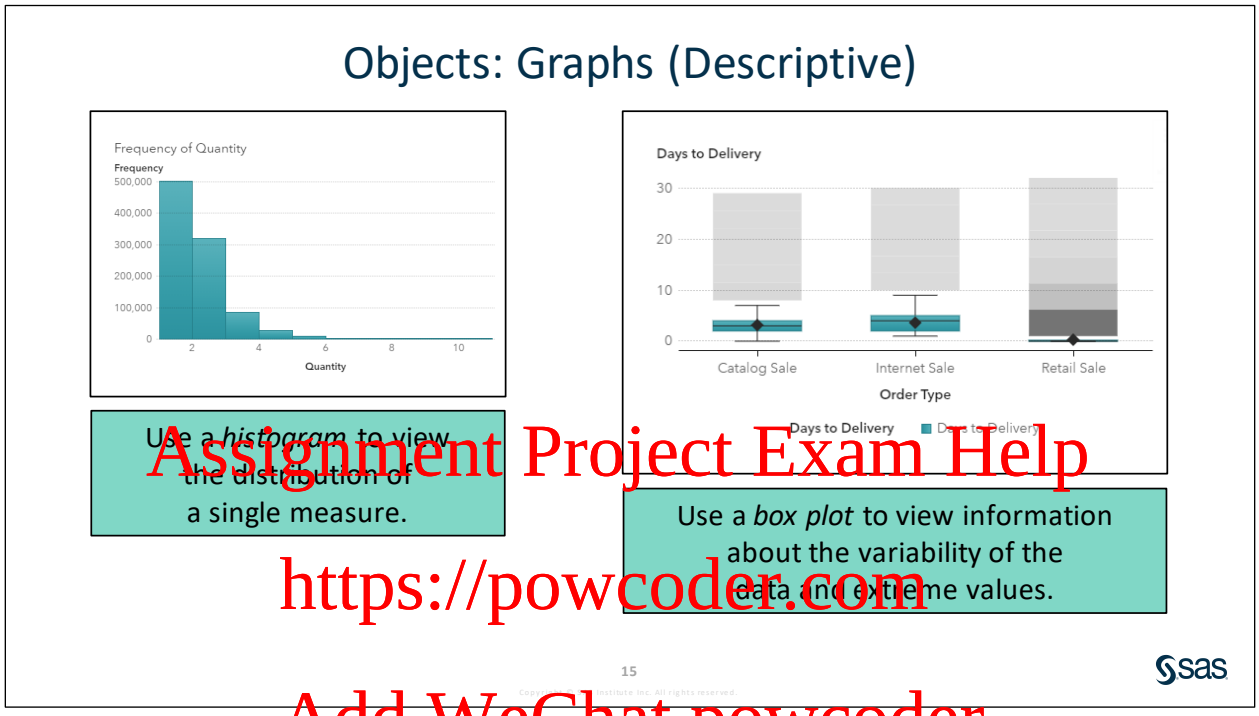
Which country has the highest average for **Days to Delivery**?

12

This report uses an automated explanation object to explore **Days to Delivery**. The automated explanation object determines the most important underlying factors for a specific response variable.

For more information about the automated explanation object, see “Working with Automated Explanation Objects” in the *SAS Visual Analytics: Working with Report Content* documentation.

Note: The automated explanation object is discussed in more detail in the SAS Visual Analytics 2 for SAS Viya: Advanced course.

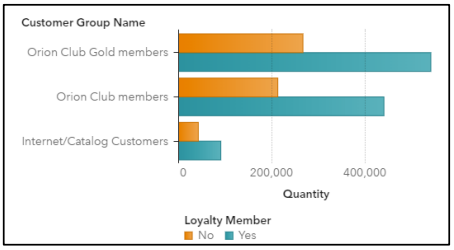


Histogram	The histogram contains a series of bars that represent the number of observations (or percentage of all observations) for a measure that fit in a specified value range (or bin). The shape of the distribution can be affected by the number of bins specified for the histogram. **Note:** If you use the default number of bins, then the minimum and maximum values on the histogram might not match your actual data values. However, if you specify the number of histogram bins, then the minimum and maximum values on the histogram match your actual data values exactly.
Box plot	The size and location of the box indicate the range of values between the 25th and 75th percentile (or the interquartile range). The diamond marker inside the box indicates the mean value, and the line inside the box indicates the median value. You can modify options to display outliers in the plot. Outliers are data points whose distance from the interquartile range are more than 1.5 times the size of the interquartile range. The whiskers (lines protruding from the box) can indicate either minimum and maximum values of the plot or the range of values outside of the interquartile range but close enough not to be considered outliers. If there are a large number of outliers, the range of outlier values is represented by a bar colored to represent the number of values inside the outlier range (as seen above).

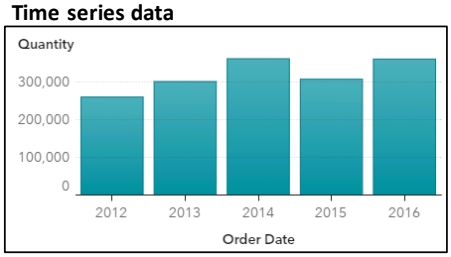
Objects: Graphs (Descriptive)

Use a *bar chart* to compare summarized data for the following:

Nominal values

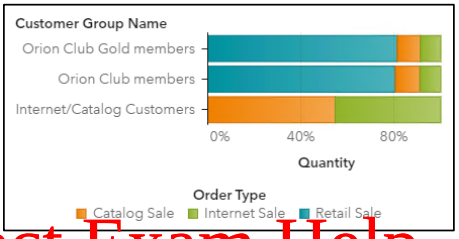


Time series data



Rankings

Parts of a whole



16

Bar chart

A bar chart displays data aggregated by the distinct values of a category. By default, the bars are sorted by descending order of the value of the first measure. For ranked bars, the data is sorted based on the values of the rank. Stacked bar charts enable you to compare totals for each category, as well as totals for all categories. However, comparing segments is difficult, and when there are many segments in the chart, it is difficult to read. To see relative differences (parts of a whole) in a bar chart, select **Normalize groups to 100%** for the **Group scale** option.

Note: Nominal values are categories whose data has no particular order.

3.02 Multiple Choice Question

Which graph would help you determine whether a measure is normally distributed?

a. distribution plot
b. box plot
c. histogram
d. normality plot

17

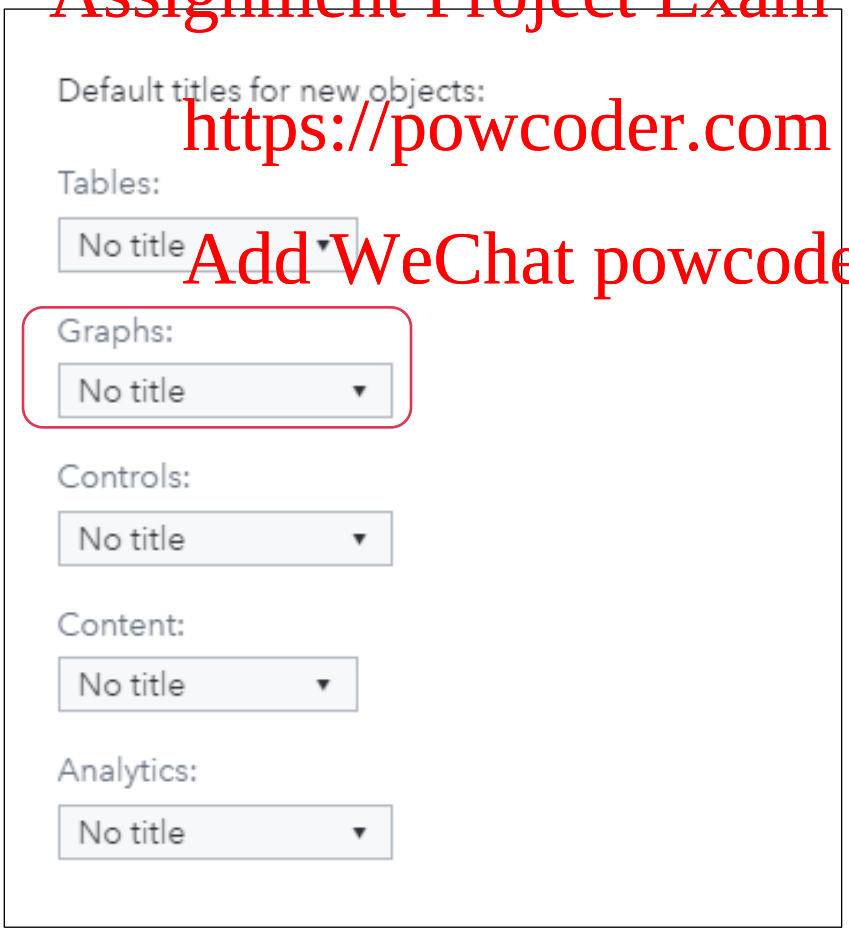
sas

Exploring Data: Part 1

This demonstration illustrates how to use the automatic chart to explore data and modify roles and options for charts and graphs in Visual Analytics.

1. From the browser window, sign in to SAS Viya.
2. In the upper left corner, click ≡ (**Show list of applications**) and select **Explore and Visualize**.

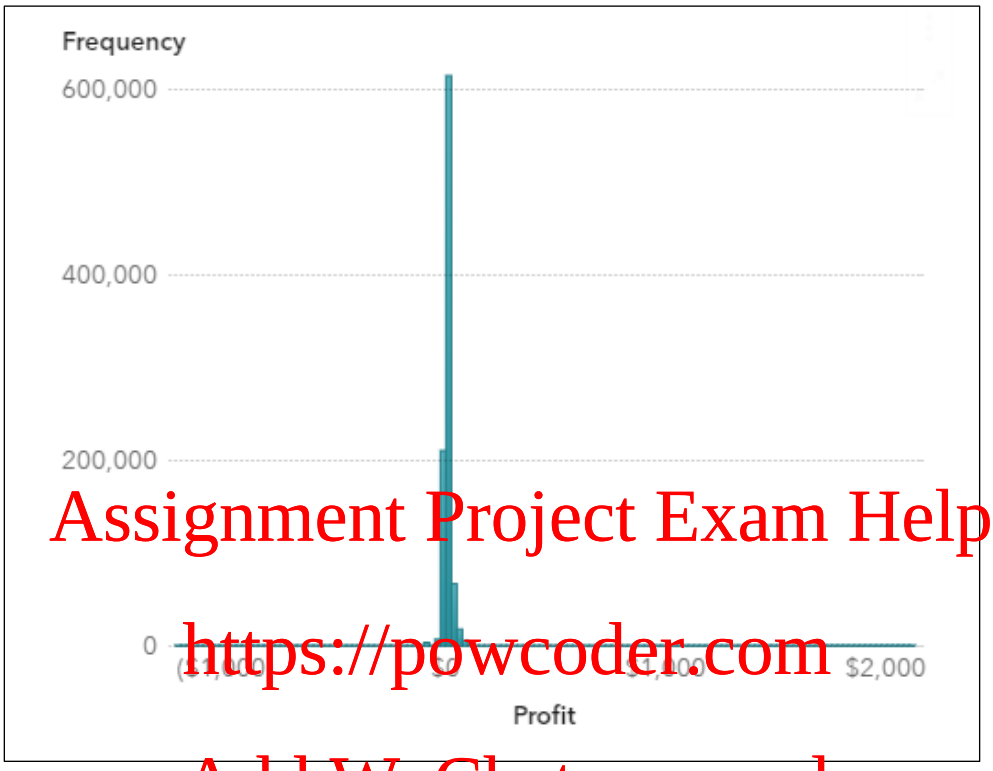
   SAS Visual Analytics appears.
3. Click **All Reports**.
   a. Navigate to the **Courses/YVA185/Basics/Demos (Marketing)** folder.
   b. Double-click the **VA1- Demo3.2a** report to open it.
4. Turn off automatic graph titles.
   a. In the upper right corner, select ***<user name>*** ⇨ **Settings**.
   b. On the left side of the window, select **General** under **SAS Visual Analytics**.
   c. Scroll down to **Default titles for new objects**.
   d. For **Graphs**, change **Automatic title** to **No title**.



   e. Click **Close**.

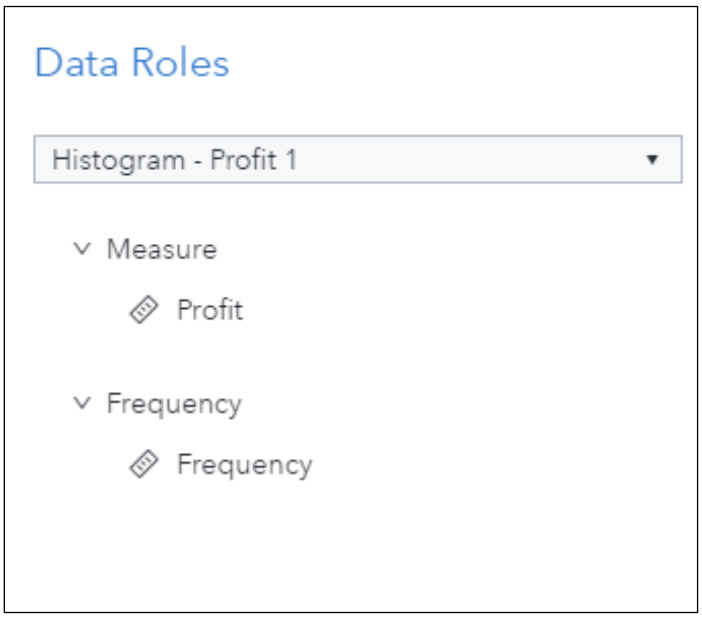
5. Create an automatic chart.
   a. In the left pane, click **Data**.
   b. Drag **Profit** from the Data pane to the canvas.

      The automatic chart functionality determines the best way to display the selected data.



      A histogram is used to display the distribution of profits.

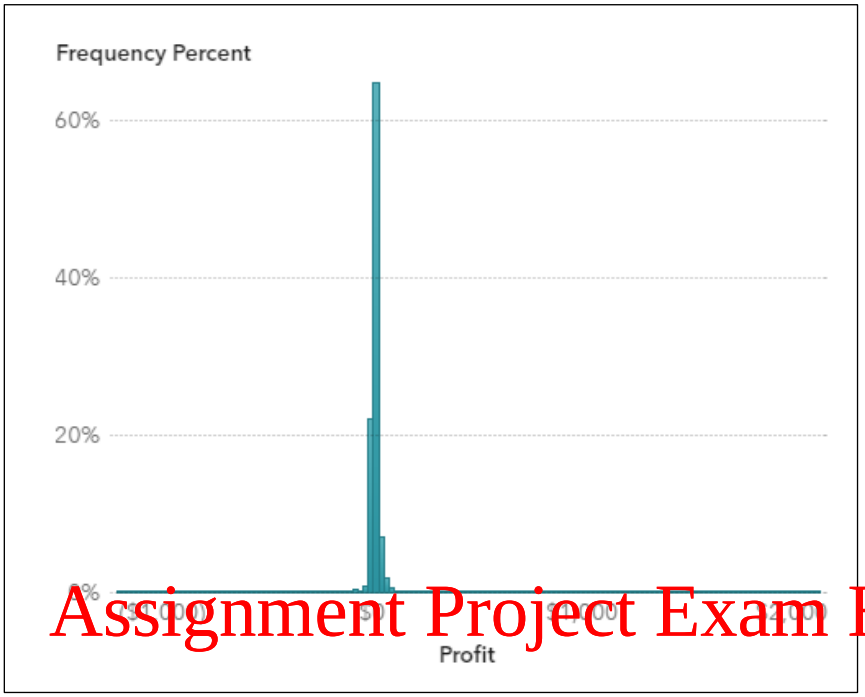
   c. In the right pane, click **Roles**.



      A histogram accepts two roles, Measure and Frequency.

d. For the **Frequency** role, select **Frequency** ⇨ **Frequency Percent**.

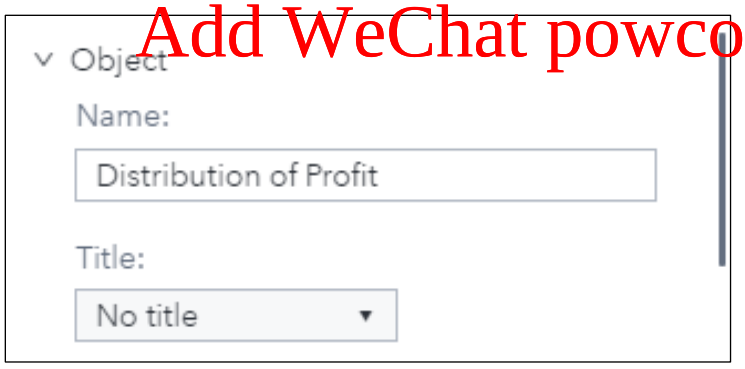
The histogram is updated to use frequency percent for the Y axis.



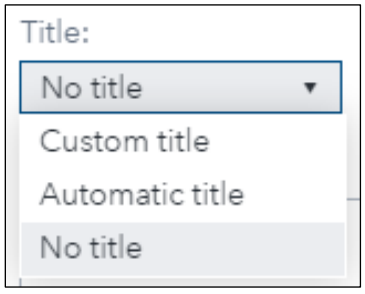
e. In the right pane, click **Options**.

1) Expand the **Object** group.

2) In the **Name** field, enter **Distribution of Profit**.

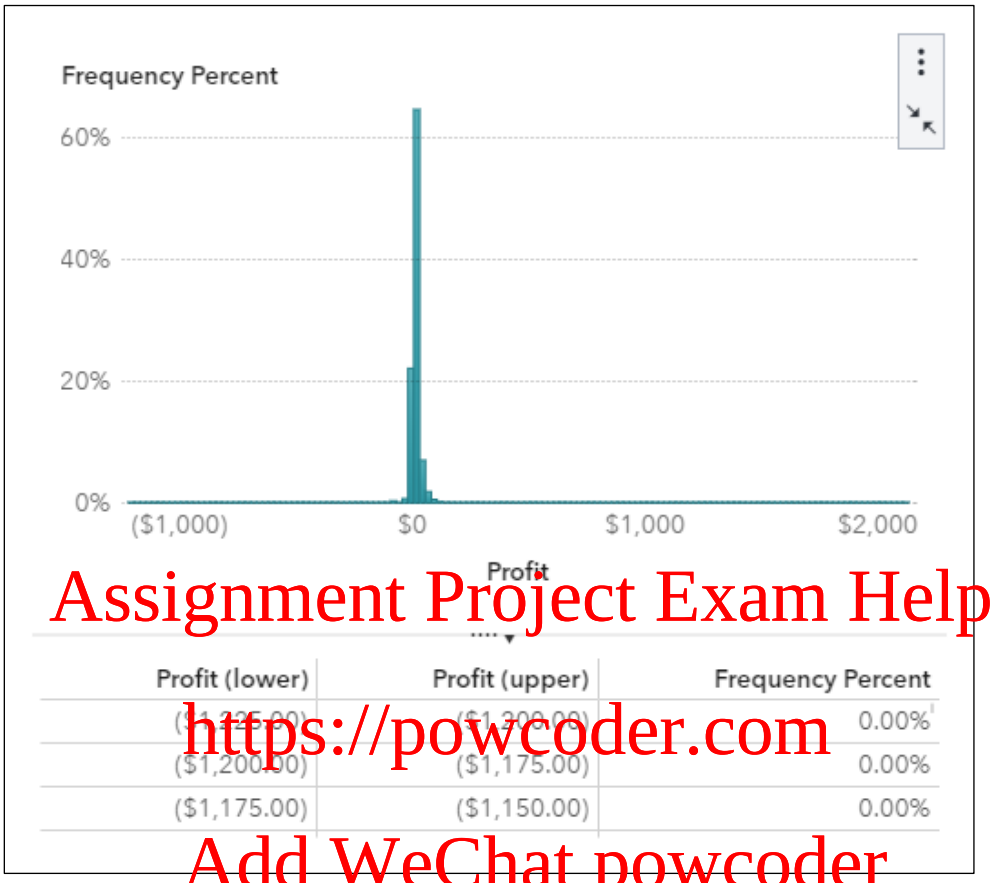


Note: The **Automatic title** setting was turned off for Graph objects in an earlier demo. You can turn it on for this graph by selecting **Automatic title**, or you can create a custom title by selecting **Custom title**.



f. In the upper right corner of the histogram, click (**Maximize**) to view additional details.

A table of data values is displayed at the bottom of the chart.



Profit (lower)	Profit (upper)	Frequency Percent
($1,225.00)	($1,200.00)	0.00%
($1,200.00)	($1,175.00)	0.00%
($1,175.00)	($1,150.00)	0.00%

g. Click the highest bar in the graph.

h. Scroll through the table to find the highlighted row.

Profit (lower)	Profit (upper)	Frequency Percent
($50.00)	($25.00)	0.88%
($25.00)	$0.00	22.23%
$0.00	$25.00	64.65%
$25.00	$50.00	7.16%

A majority of the products ordered are low-profit items, in the $0 to $25 range. Also notice that more than 20% of items result in a loss. Why is this problem occurring? Are these products ordered from a similar product area, geographical area, or order type? Could the costs be too high in these areas? What can we do to reduce costs?

i. In the upper right corner, click (**Restore**).

6. Create a crosstab.

a. In the left pane, click **Objects**.

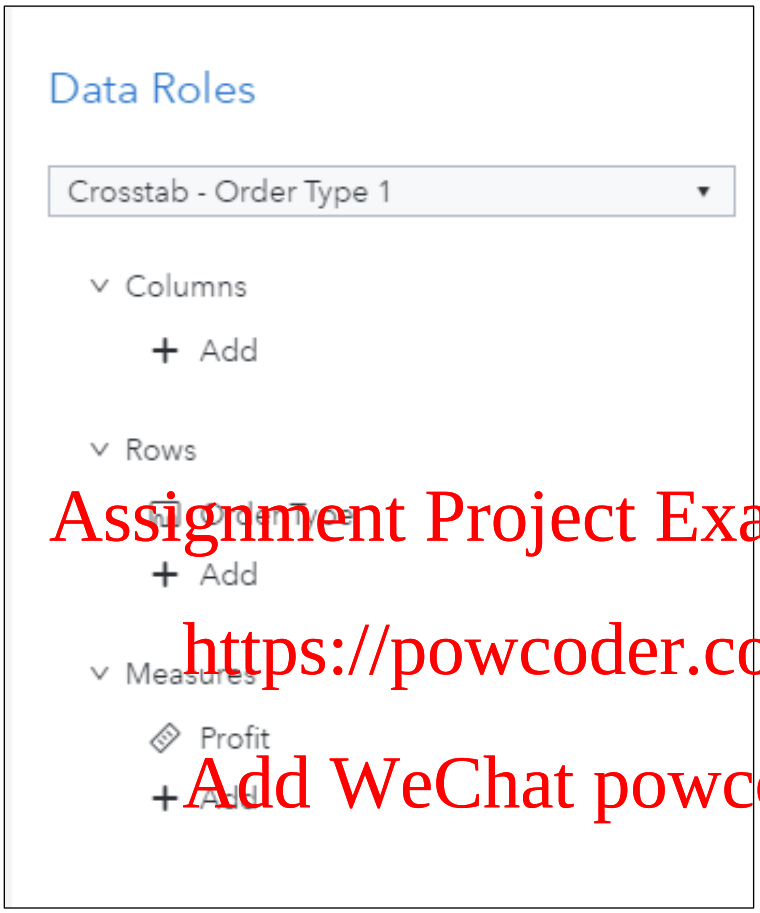
b. Drag the **Crosstab** object, from the Tables group, to the bottom of the canvas.

c. In the right pane, click **Roles**.

d. For the **Rows** role, select **Add** ⇨ **Order Type** and click **OK**.

e. For the **Measures** role, select **Frequency** ⇨ **Profit**.

The Roles pane should resemble the following:



Note: The Measures role is required for the crosstab object.

The crosstab should resemble the following:

Order Type ▲	Profit
Catalog Sale	$1,153,380.79
Internet Sale	$981,170.49
Retail Sale	$6,124,855.53

Profits are much lower in the internet and catalog channels. A company-wide policy mandates that we need to try to improve profits for orders through these channels.

f. On the Roles tab, for the **Columns** role, select **Add** ⇨ **Continent Name** and click **OK**.

The updated crosstab should resemble the following:

Continent Name ▲	Africa	Asia	Europe	North America	Oceania
Order Type ▲	Profit	Profit	Profit	Profit	Profit
Catalog Sale	$730.56	$7,564.99	$670,252.82	$423,428.89	$51,403.52
Internet Sale	($858.24)	$7,938.71	$559,663.83	$370,621.44	$43,804.75
Retail Sale	——	——	$4,429,533.94	$1,327,595.24	$367,726.36

g. In the right pane, click **Options**.

h. Expand the **Totals and Subtotals** group.

i. Select the **Totals** check box.

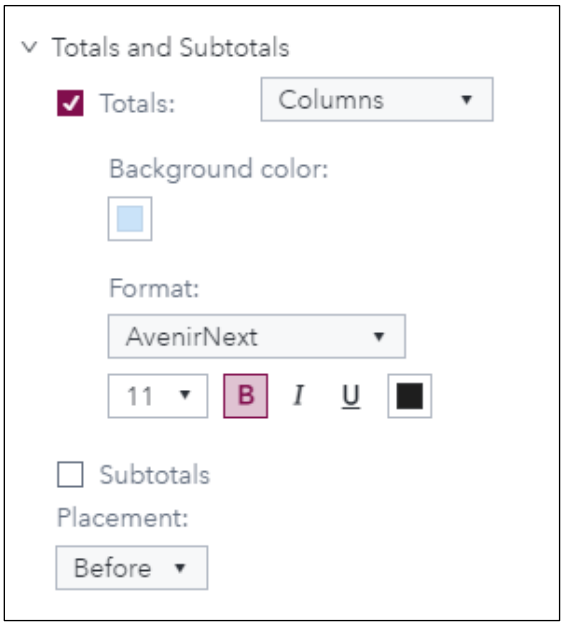
By default, totals are added to rows and columns.

j. Next to the **Totals** field, select **Columns**.

k. For the **Background color** field, click (**Select a color**).

l. Select **Pale blue**.

m. For the **Format** field, verify that B (**Bold**) is selected.



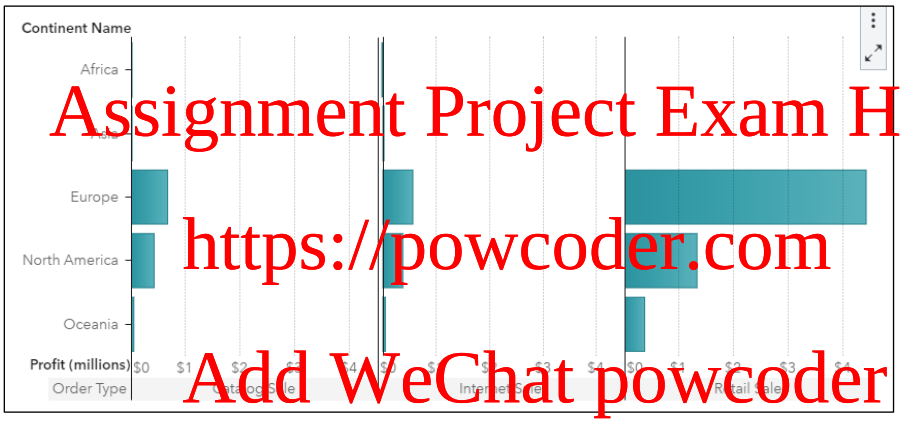
The updated crosstab should resemble the following:

Continent Name ▲	Africa	Asia	Europe	North America	Oceania
Order Type ▲	**Profit**	**Profit**	**Profit**	**Profit**	**Profit**
Total	**($127.68)**	**$15,503.70**	**$5,659,450.59**	**$2,121,645.57**	**$462,934.63**
Catalog Sale	$730.56	$7,564.99	$670,252.82	$423,428.89	$51,403.52
Internet Sale	($858.24)	$7,938.71	$559,663.83	$370,621.44	$43,804.75
Retail Sale	——	——	$4,429,533.94	$1,327,595.24	$367,726.36

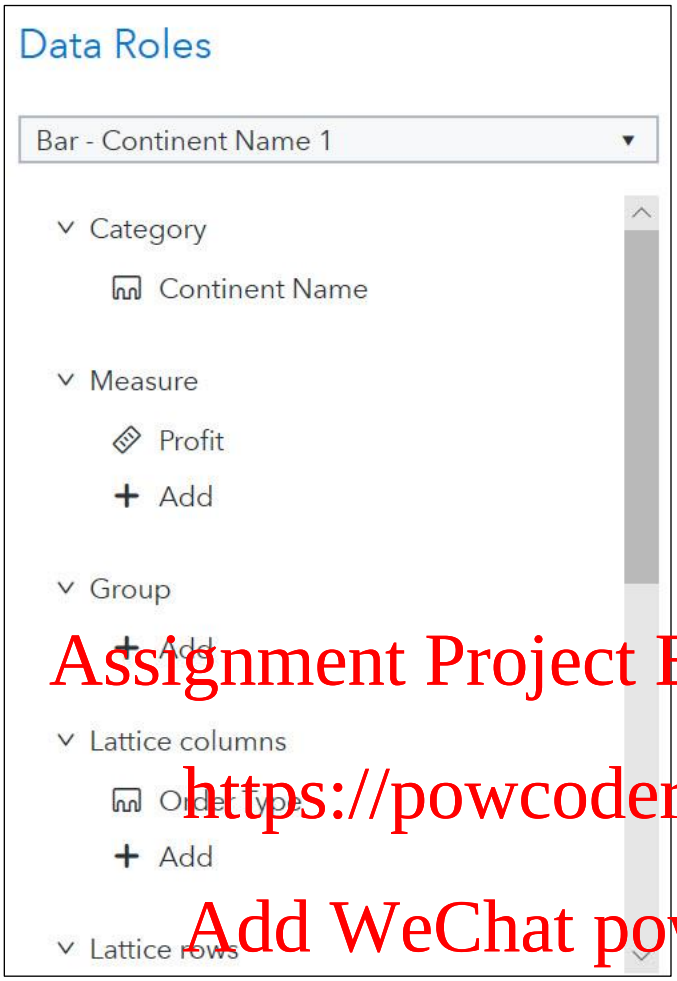
Profits are much lower in North America than in Europe. Because our corporate office is located in North America, we would expect higher profits. Also notice the loss in Africa for internet sales. Why is this loss occurring? Is this due to start-up operations (for example, building distribution facilities in Africa)? Are the losses consistent over time or has this changed over time?

7. Change the crosstab to a bar chart.
   a. Right-click the crosstab and select **Change Crosstab to** ⇨ **Bar chart**.

      The bar chart should resemble the following:



b. In the right pane, click **Roles**.

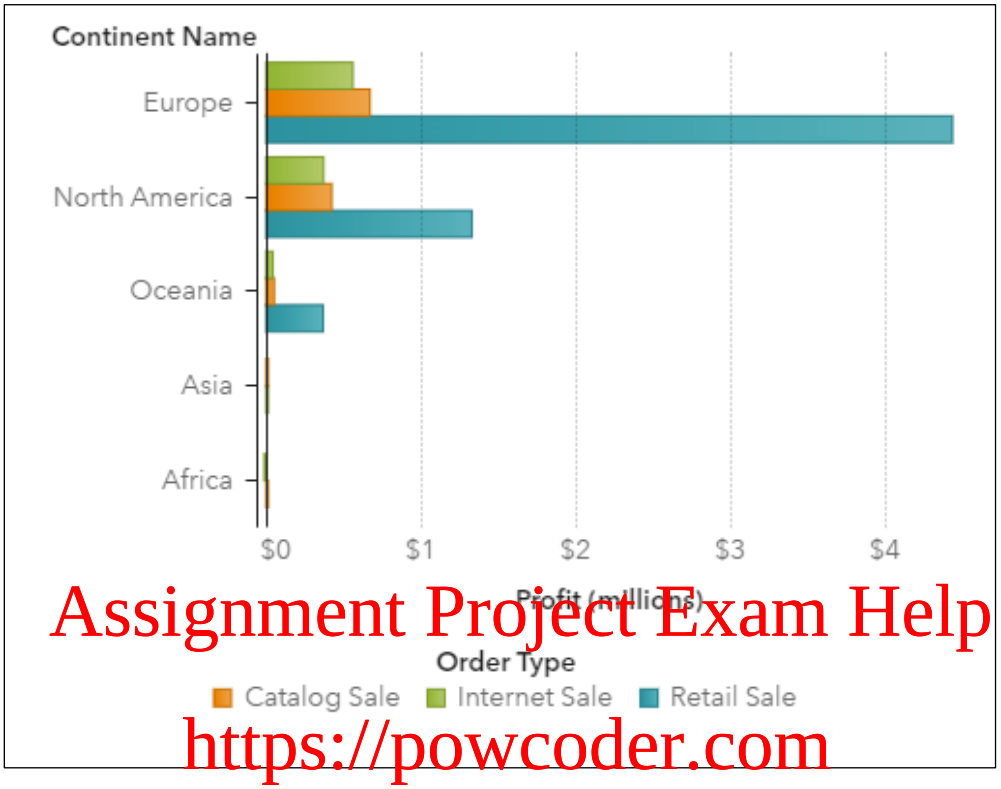


The bar chart has many more roles available.

- Category data items can be added to the Group role to show additional bars for each category, or to the Lattice columns and Lattice rows roles to add additional bar charts for each distinct category.
- Category and Measure data items can be added to the data tip values role to show additional information when a bar is selected.
- Datetime data items can be added to the Animation role to animate the bar chart.
- Category or date data items can be added to the Hidden role for mapping data sources, adding color-mapped display rules, or adding external links.

c. Drag **Order Type**, from the **Lattice columns** role, to the **Group** role.

The bar chart should resemble the following:



d. In the right pane, click **Options**.

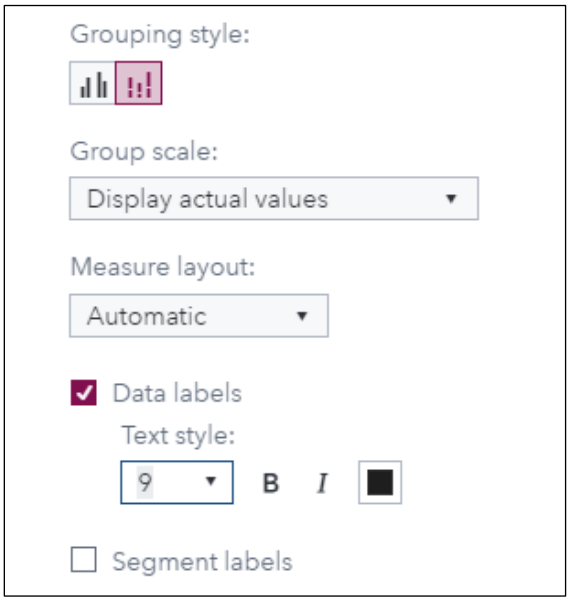
e. In the Object group, for the **Name** field, enter **Profit by Continent and Order Type**.

f. In the Bar group, for the **Grouping style** field, click (**Stacked**).

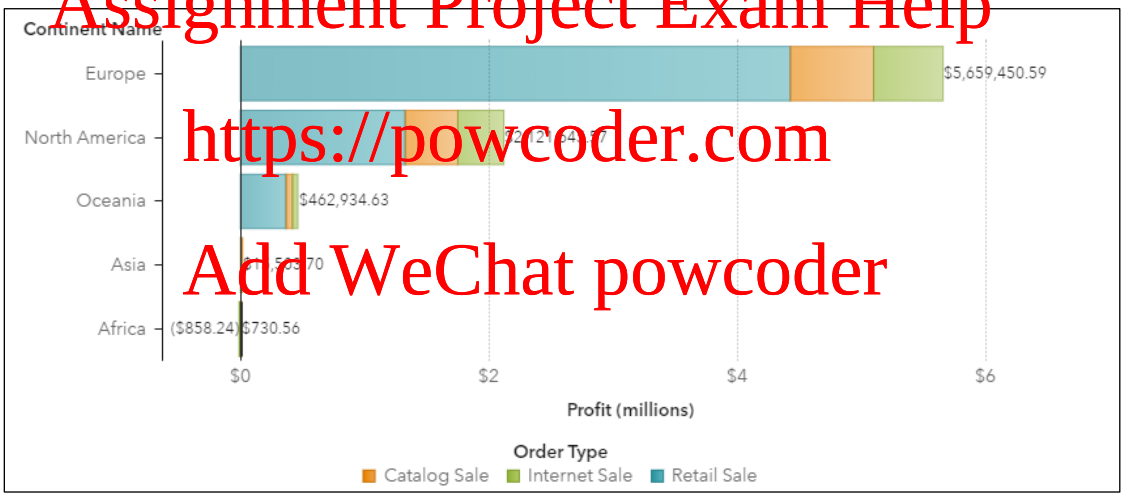
g. Select **Data labels**.

h. For the **Text style** field, select **9**.

The Bar group should resemble the following:

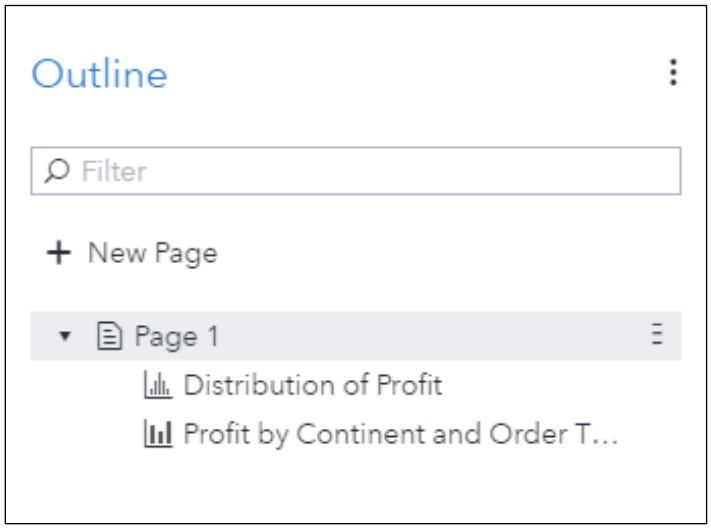


The updated bar chart should resemble the following:



Profits in North America are less than half of total profits in Europe. We need to understand why this discrepancy exists and try to improve profits in non-European countries.

8. In the left pane, click **Outline**.



The Outline pane displays a list of all pages and objects in the report.

9. Save the report.

End of Demonstration

Practice

2. Exploring Data: Part 1

a. Open the browser and sign in to SAS Viya.

b. Open the **VA1- Practice3.2a** report from the **Courses/YVA185/Basics/Practices (HR)** folder.

c. Create an automatic chart using the following data items:

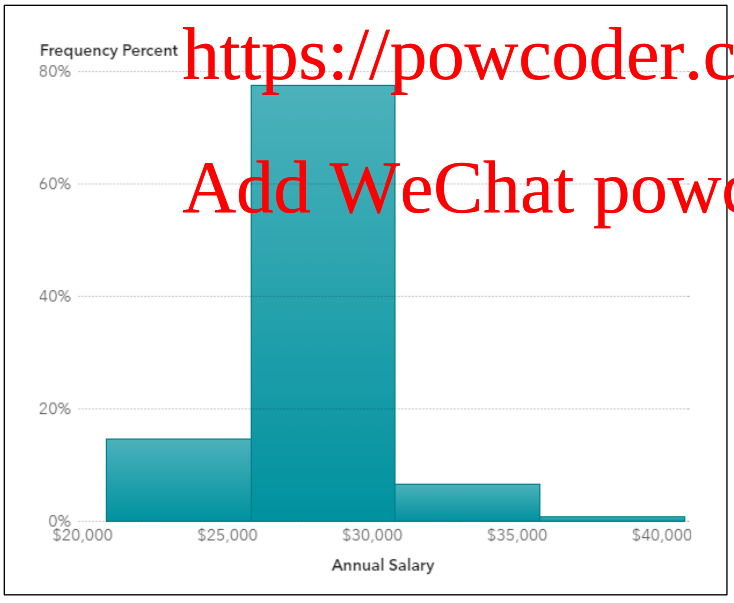
Annual Salary

Frequency Percent

d. Modify the following options for the automatic chart:

Name	Distribution of Salary
Bin range	Measure values
Set a fixed bin count	<selected>
Bin count	4

The automatic chart should resemble the following:



e. Maximize the histogram to answer the following question:

Into which range do the majority of salaries fall?

Answer: _____

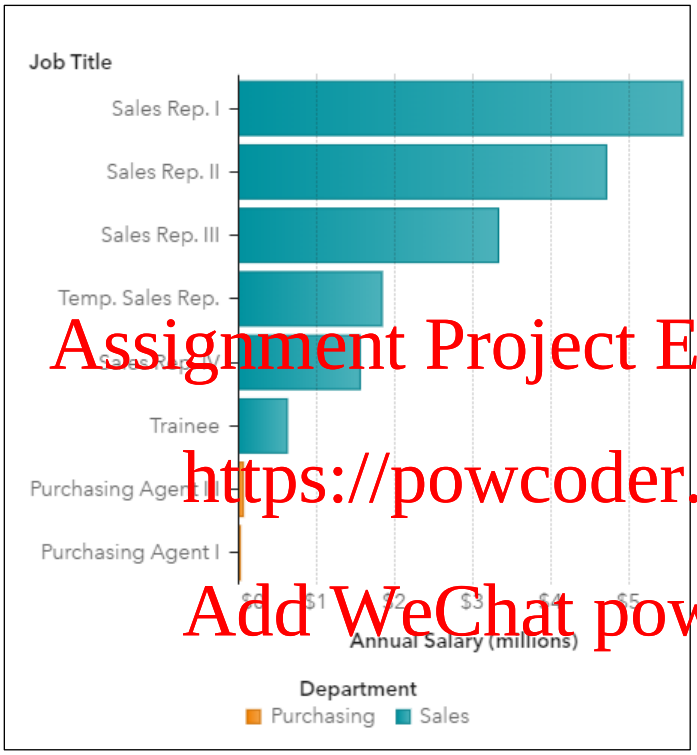
Hint: After answering the question, click (**Restore**) in the upper right corner.

f. Add a bar chart on the right of the automatic chart by assigning the following data items to the specified roles:

Category	**Job Title**
Measure	**Annual Salary**
Group	**Department**

g. Specify **Total Salary by Job and Department** as the name of the bar chart.

The bar chart should resemble the following:



h. Answer the following questions:

In which department are a majority of our salary costs spent? For which job title?

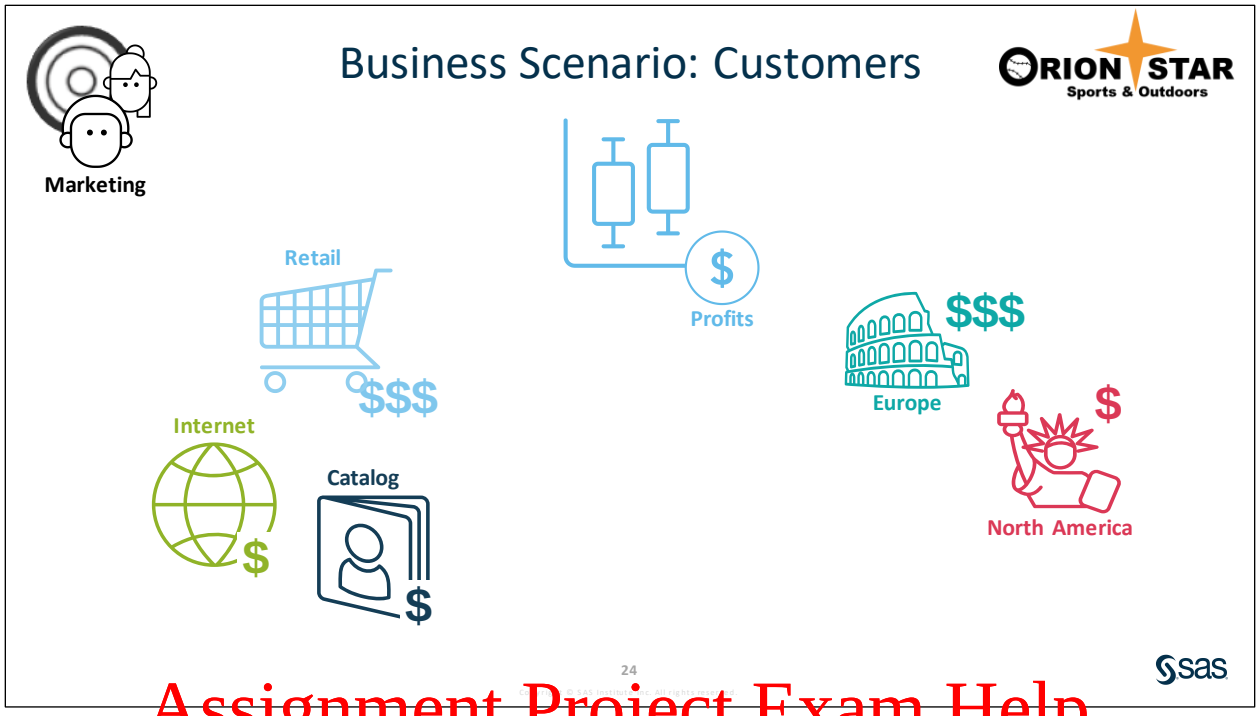
Answer: _____

What could be some reasons why salary costs are so much higher for this group?

Answer: _____

i. Save the report.

End of Practices

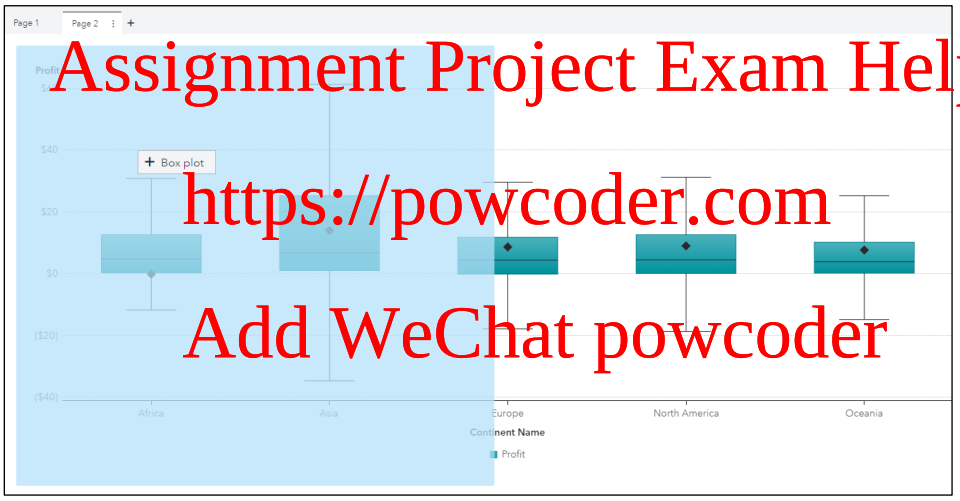


Exploring Data: Part 2

This demonstration illustrates how to use box plots to explore data in Visual Analytics.

1. From the browser window, sign in to SAS Viya.
2. In the upper left corner, click ☰ (**Show list of applications**) and select **Explore and Visualize**.

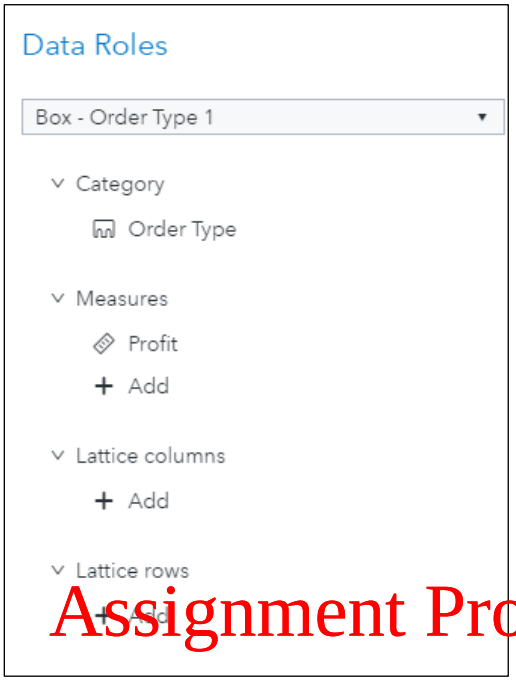
   SAS Visual Analytics appears.
3. Click **All Reports**.
   a. Navigate to the **Courses/YVA185/Basics/Demos (Marketing)** folder.
   b. Double-click the **VA1- Demo3.2b** report to open it.
4. In the upper left corner of the report, click the **Page 2** tab.
5. Create a box plot.
   a. In the left pane, click **Objects**.
   b. Drag the **Box plot** object, from the Graphs group, to the left side of the canvas.



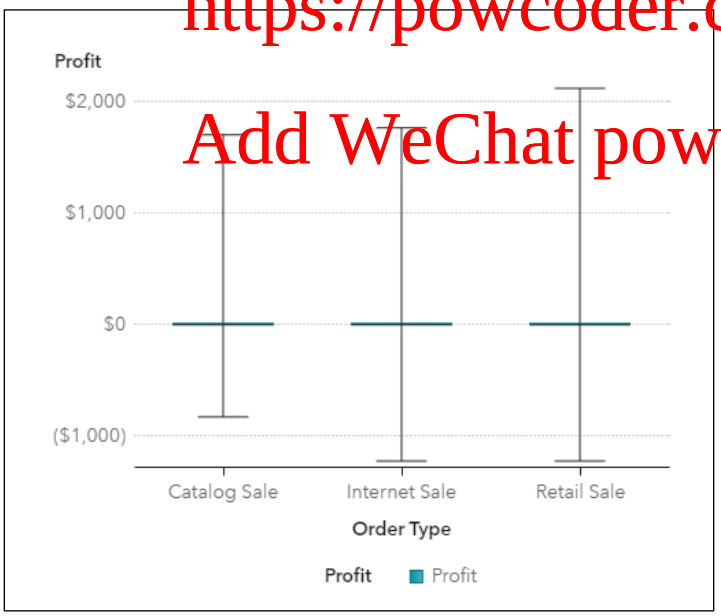
   c. In the right pane, click **Roles**.
   d. For the **Category** role, select **Add** ⇨ **Order Type**.

e. For the **Measures** role, select **Add** ⇨ **Profit** and click **OK**.

The Roles pane should resemble the following:



The box plot should resemble the following:

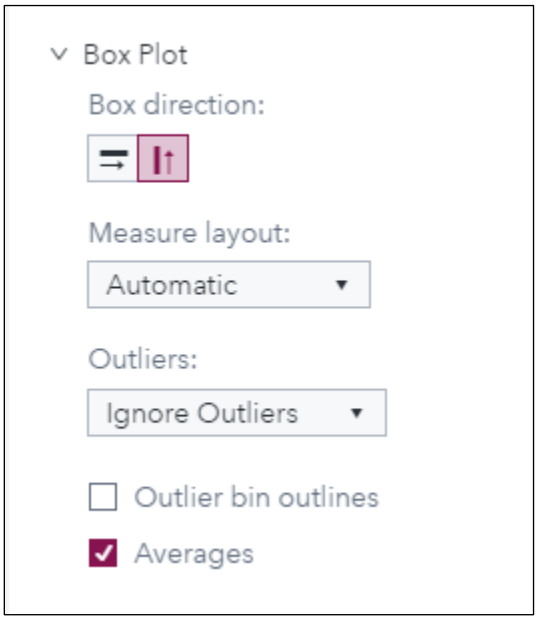


f. In the right pane, click **Options**.

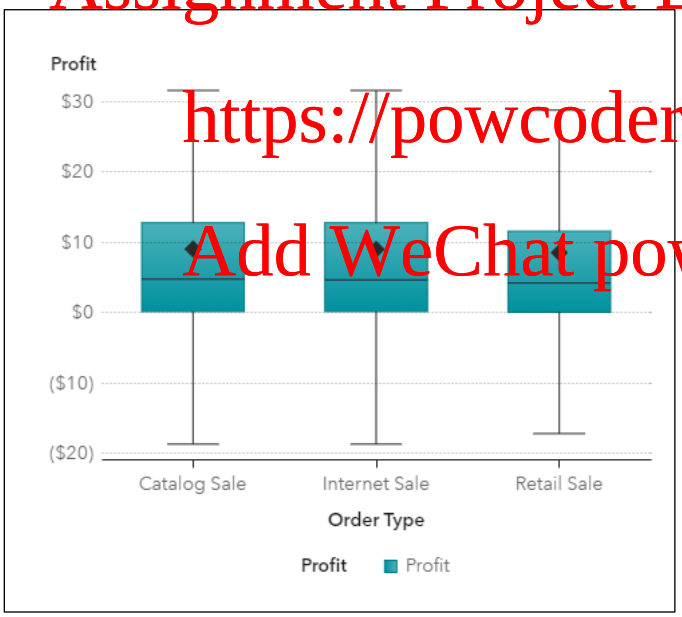
g. In the Object group, for the **Name** field, enter **Profit by Order Type**.

h. In the Box Plot group, for the **Outliers** field, select **Ignore Outliers**.

i. Select the check box for **Averages**.



The box plot should resemble the following:



j. In the upper right corner of the box plot, click (**Maximize**) to view additional details.

The detail table displays descriptive statistics for **Profit** for each order type.

Order Type	Minimum	Lower Whisker	First Quartile	Average	Median	Third Quartile
Catalog Sale	($826.26)	($18.63)	$0.20	$9.07	$4.80	$12.80
Internet Sale	($1,222.48)	($18.63)	$0.20	$9.04	$4.70	$12.80
Retail Sale	($1,222.48)	($17.13)	$0.10	$8.55	$4.25	$11.60

Even though total profits are highest for the retail sales channel, averages across all channels are very similar, but are a bit higher for catalog and internet sales. This reinforces our company-wide policy to try to increase profits in these channels. Total profits might be higher in retail because there are more customers or more orders for that channel.

k. In the upper right corner, click (**Restore**).

l. In the upper right corner of the **Profit by Continent** box plot, click (**Maximize**) to view additional details.

The detail table displays descriptive statistics for **Profit** for each continent.

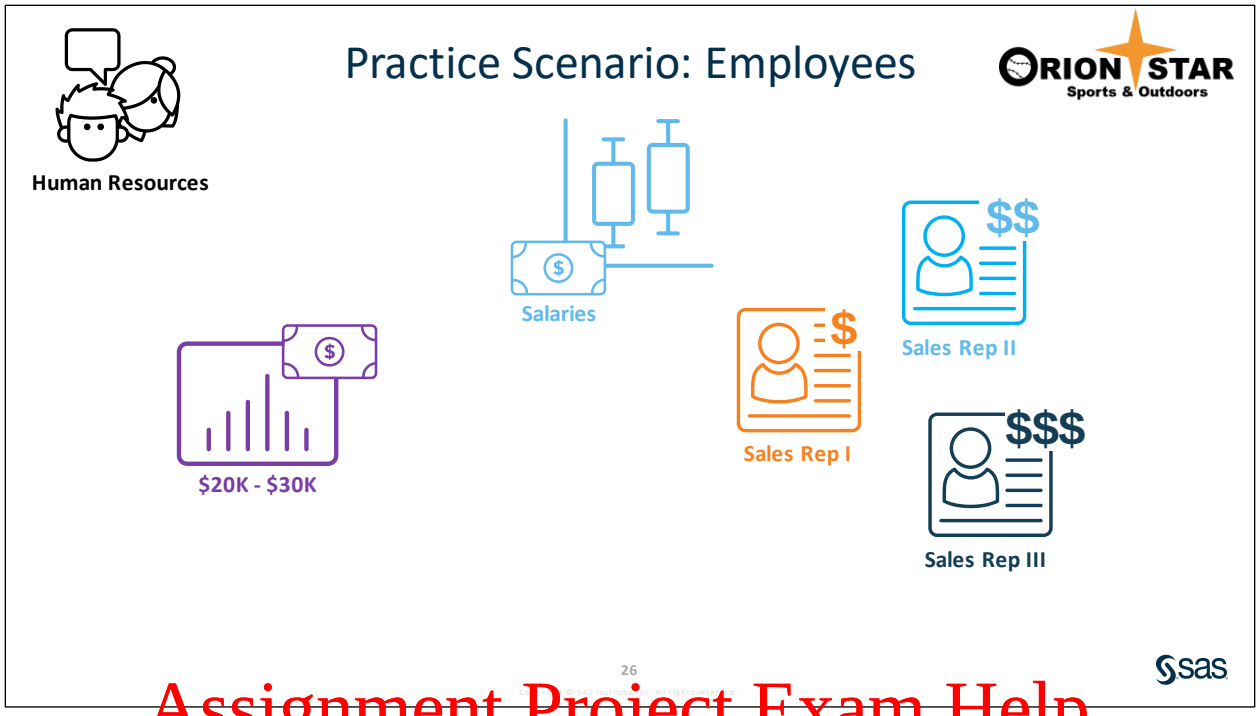
Continent Name	Minimum	Lower Whisker	First Quartile	Average	Median	Third Quartile	U
Africa	($374.42)	($11.70)	$0.30	($0.17)	$4.80	$12.60	
Asia	($258.84)	($34.62)	$1.00	$13.97	$6.80	$25.20	
Europe	[illegible]	[illegible]	[illegible]	[illegible]	$4.40	$11.80	
North America	($1,222.48)	($18.63)	$0.10	$9.00	$4.50	$12.60	
Oceania	($646.40)	($14.80)	$0.20	$7.66	$3.90	$10.20	

Even though total profits are highest for Europe, averages are higher in North America and Asia. Because our corporate office is located in North America, we will start by focusing on increasing profits in North America. Total profits might be higher in Europe because there are more customers or more orders for that continent. Also, note the negative average profits in Africa. Why is this occurring? What can we do to increase profits for that continent?

m. In the upper right corner, click (**Restore**).

6. Save the report.

End of Demonstration



Practice

3. Exploring Data: Part 2

a. Open the browser and sign in to SAS Viya.

b. Open the **VA1- Practice3.2b** report from the **Courses/YVA185/Basics/Practices (HR)** folder.

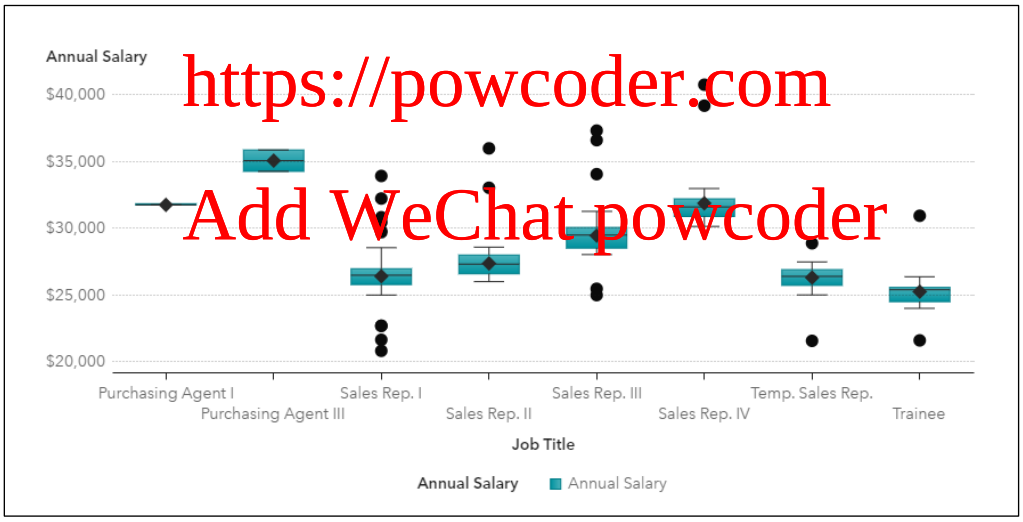
c. On Page 2, create a box plot by assigning the following data items to the specified roles:

Category	**Job Title**
Measures	**Annual Salary**

d. Modify the following options for the box plot:

Name	Salary Analysis by Job Title
Outliers	Show Outliers
Averages	<selected>

The box plot should resemble the following:



e. Maximize the box plot to answer the following questions:

Which job title has the highest average salary? The lowest?

Answer: _____

Which job title has the largest number of outliers?

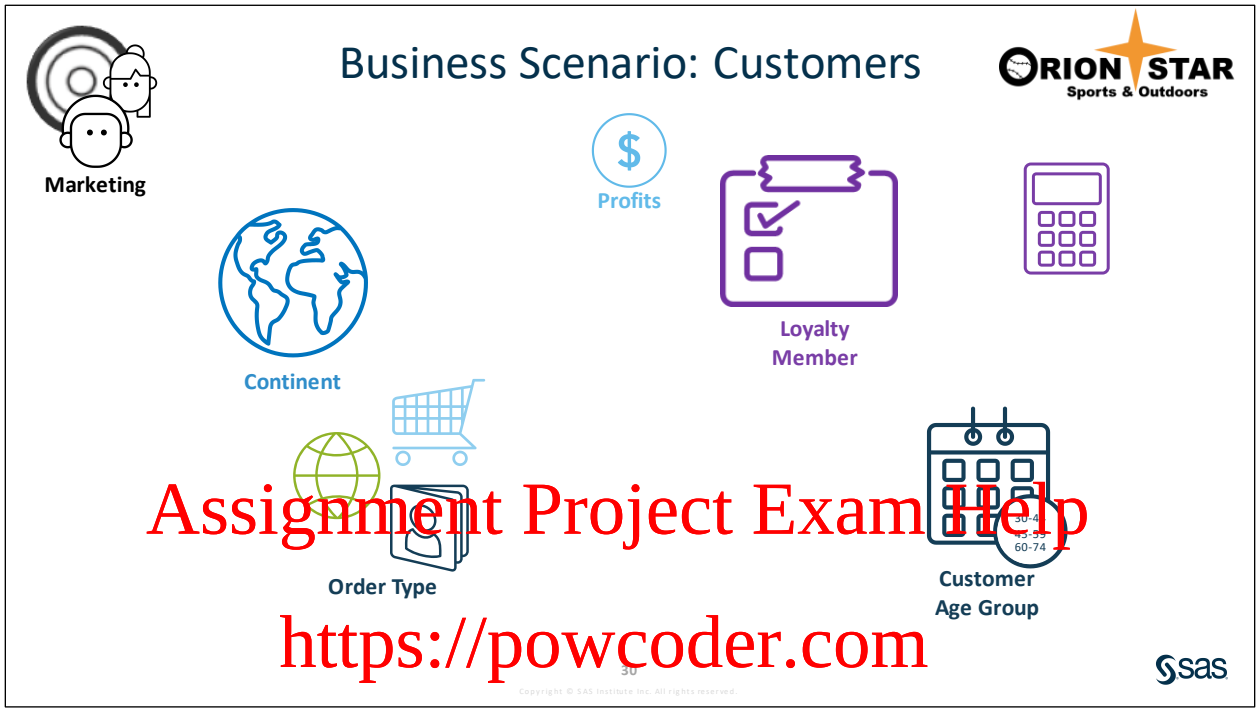
Answer: _____

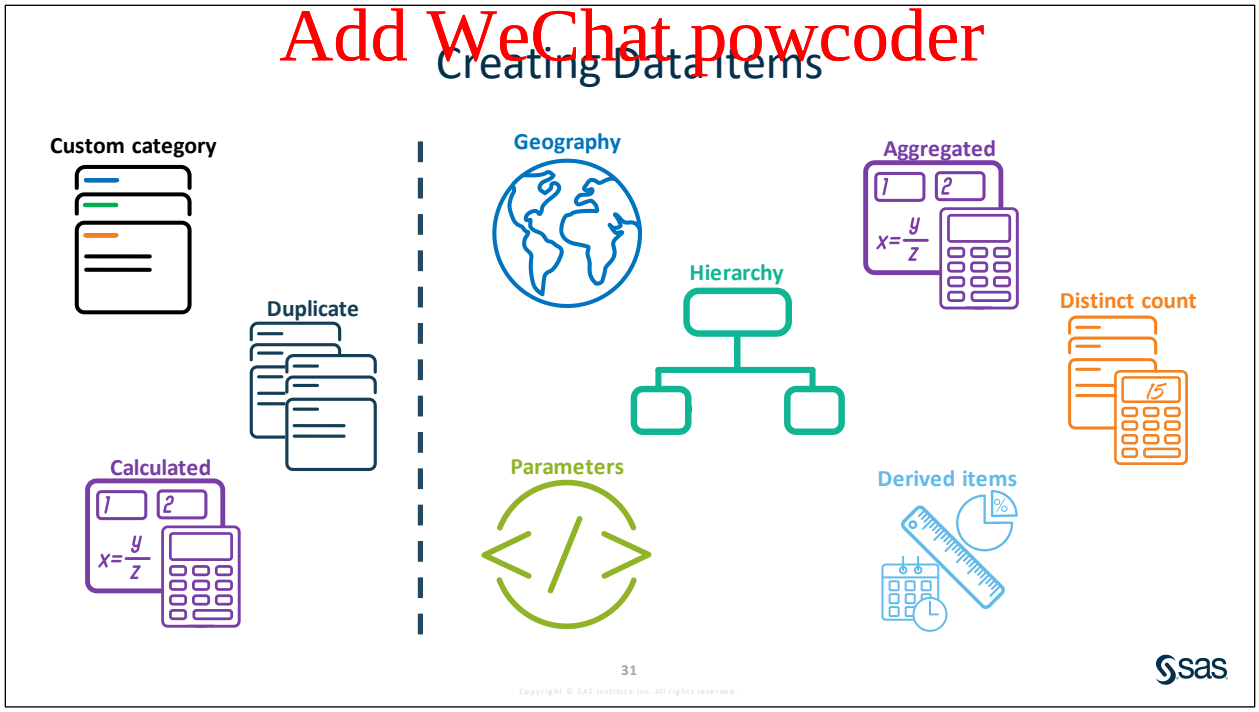
Hint: After answering the question, click (**Restore**) in the upper right corner.

f. Save the report.

End of Practices

3.3 Creating Data Items and Applying Filters





The following types of data items can be created in SAS Visual Analytics, using code, or in SAS Data Studio or SAS Enterprise Guide:

Custom category	A custom category creates labels for groups of values of category or measure data items. When you create a custom category from a measure data item, you can use intervals or distinct values to group the data. For more information about custom categories, see “Working with Custom Categories in a Report” in the *SAS Visual Analytics: Working with Report Data* documentation.
Duplicate	Both measures and categories can be duplicated (copied) in Visual Analytics. Duplicating measures enables you to compare the data using different aggregations in a table or graph or change the classification to a category for grouping other values in tables or graphs. Duplicating datetime values enables you to apply different formats to the values for use in tables or graphs. Duplicating calculated items enables you to make variations to a calculation. For more information about duplicating data items, see “Working with Data Items in a Report” in the *SAS Visual Analytics: Working with Report Data* documentation.
Calculated item	Calculated items are created by performing mathematical calculations on numeric values, or by performing operations on datetime data items or categories. All calculations are performed on unaggregated data. That is, the expression is evaluated for each row in the data source. For more information about creating calculated data items, see “Working with Calculated Items in a Report” in the *SAS Visual Analytics: Working with Report Data* documentation. For more information about operators, see “Reference: Operators for Data Expressions” in the *SAS Visual Analytics: Working with Report Data* documentation.

The following types of data items need to be created in Visual Analytics:

Geography	A geography data item is a category whose values are mapped to geographical locations or regions. Geography data items can be used with geo maps and other report objects. Geography data items can be created using predefined roles (for example, country names), by associating latitude and longitude coordinates with the values (custom), or by associating polygon data from a separate data source with map regions (custom). For more information about creating geography data items, see “Working with Geography Data Items” in the *SAS Visual Analytics: Working with Report Data* documentation.
Aggregated measure	Aggregated measures enable you to calculate new data items using aggregated values. This means that the calculation changes depending on the other data items available in the graph. For example, you can see the profit margin for each region or by each store. For more information about creating calculated data items, see “Working with Calculated Items in a Report” in the *SAS Visual Analytics: Working with Report Data* documentation.

<table>
<tr><td>Hierarchy</td><td>A hierarchy is a defined arrangement of category data items based on a parent-child relationship. In many cases, the levels of the hierarchy are arranged with the more general information at the top (for example, year) and the more specific information at the bottom (for example, month). Hierarchies enable you to add drill-down functionality to graphs and tables. Hierarchies that consist of all geographic data items are considered geographic hierarchies and can be used in geo maps.

Note: You can create a date hierarchy from a date data item. The date hierarchy, by default, has levels for year, quarter, month, and day. A date hierarchy created from a datetime data item has levels, by default, for year, quarter, month, day, hour, minute, and second.

For more information about hierarchies, see “Working with Hierarchies in a Report” in the SAS Visual Analytics: Working with Report Data documentation.</td></tr>
<tr><td>Distinct count</td><td>A distinct count counts the number of distinct values of a category data item as an aggregated measure. This means that the calculation changes depending on the other data items available in the graph. For example, you can see the number of orders placed for each age group or the number of orders placed for each country by creating a distinct count from Order ID. For more information about creating distinct counts, see “Working with Data Items in a Report” in the SAS Visual Analytics: Working with Report Data documentation.

Note: If the category contains missing values, the distinct count is increased by one. A configuration setting can modify this behavior.</td></tr>
<tr><td>Parameter</td><td>A parameter is a variable whose value can be changed and that can be referenced by other report objects. Parameters can be used in control objects in Visual Analytics. When the value of the control changes, the parameter is updated with that value, and any report objects that reference that parameter are updated as well. Parameters can be used in calculations, display rules, filters, ranks, URLs, and text objects. For more information about parameters, see “Working with Parameters in Reports” in the SAS Visual Analytics: Working with Report Data documentation.</td></tr>
<tr><td>Derived item</td><td>Derived data items are aggregated measures that display values for the measure and the formula type on which the derived item is based.

The following types of derived items can be created from category data items:

Distinct count — Displays the number of distinct values for the selected category. For more information, see the distinct count row above.

Count — Displays the number of nonmissing values for the selected category.

Number missing — Displays the number of missing values for the selected category.

The following types of derived data items can be created from measure data items:

Cumulative total — Displays a running total of all the values for the measure on which it is based.</td></tr>
</table>

	Data suppression	Obscures aggregated data if individual data values could easily be inferred. Data suppression replaces all values for the measure on which it is based with an asterisk (*) unless a value represents the aggregation of a specified minimum number of values. For more information, see “Reference: Operators for Data Expressions” in the *SAS Visual Analytics 8.3: Working with Report Data* documentation.
	Difference from previous period	Displays the difference between the value for the current time period and the value for the previous time period.
	Difference from previous parallel period	Displays the difference between the value for the current time period and the value for the previous parallel time period within a longer time interval.
	Moving average	Displays a moving average (rolling average) for the measure on which it is based. The moving average calculates the average for each value with the specified number of preceding values.
	Percent difference from previous period	Displays the percentage difference between the value for the current time period and the value for the previous time period.
	Percent difference from previous parallel period	Displays the percentage difference between the value for the current time period and the value for the previous parallel time period within a longer time interval.
	Percent of subtotals	Displays the percentage of the subtotal value for the measure on which it is based. You can create a percentage of subtotal only when the source data item has an aggregation of Sum or Count. **Note:** The Percent of subtotals derived item is available only for use in crosstabs. **Note:** The Percent of subtotals derived item is relative to the subset of data that is selected by your filters and ranks.
	Percent of total - sum	Displays the percentage of the total value for the measure on which it is based. You can create a percentage of total only when the source data item has an aggregation of Sum or Count. **Note:** The Percent of total – sum derived item is relative to the subset of data that is selected by your filters and ranks.
	Period to date	Displays the aggregated value for the current time period and all of the previous time periods within a larger time interval.

	Year to date	Displays the aggregated value for the current time period and all of the previous time periods within the year. The year-to-date calculation subsets the data for each year using today's date (where today is evaluated each time you view the report).
	Year to date growth	Displays the percentage difference between the year-to-date value for the current time period and the year-to-date value for the same time period of the previous year. The year-to-date calculation subsets the data for each year using today's date (where today is evaluated each time you view the report).
	Year over year growth	Displays the percentage difference between the current time period and an equivalent time period from the previous year. The year-over-year calculation subsets the data for each year using today's date (where today is evaluated each time you view the report).
	For more information about derived items, see "Working with Data Items in a Report" in the *SAS Visual Analytics: Working with Report Data* documentation.	

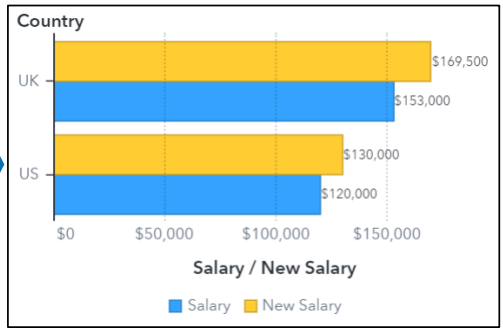
Note: Creating calculated items and aggregated measures is discussed in more detail in the SAS Visual Analytics 2 for SAS Viva: Advanced course.

Calculated Item: Example

Calculated items are created by performing operations on unaggregated data.

(Salary * Increase)

Country	Salary	Increase	New Salary
US	40,000	1.05	42,000
UK	65,000	1.10	71,500
UK	32,000	1.05	33,600
US	80,000	1.10	88,000
UK	56,000	1.15	64,400



32

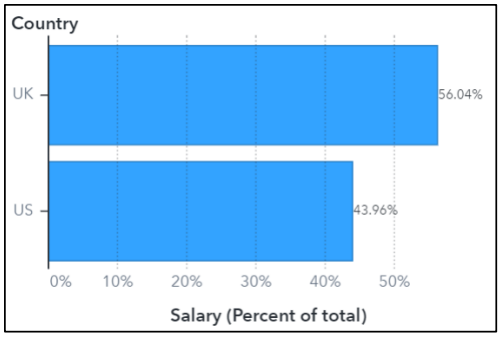
Aggregated Measure: Example

Aggregated measures are created by aggregating first and then performing the operation.

(Sum _ByGroup_ (Salary) / Sum _ForAll_ (Salary))

Country	Salary
US	40,000
UK	65,000
UK	32,000
US	80,000
UK	56,000

Country	Salary
US	120,000
UK	153,000
TOTAL	273,000



33

Note: Distinct counts and derived data items are special types of aggregated measure.

3.03 Activity

Match each new data item with the type of calculation.

___ Gross Profit Margin (Total Profit / Total Revenue)

___ Date (from month, day, year)

___ Hemisphere (from continents)

___ GDP Growth (year-over-year)

___ Number of Employees (distinct count)

___ State Abbreviations (uppercase)

A. calculated item

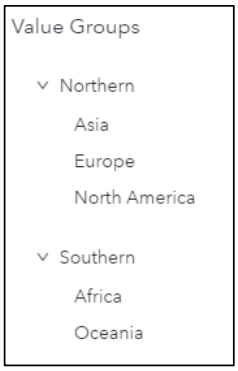
B. aggregated measure

34

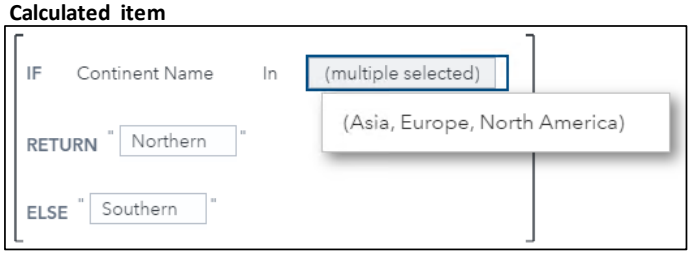
Custom Category: Example

Custom categories create labels for groups of category or measure data items.

Custom category



Calculated item



This calculated item and custom category produce equivalent results.

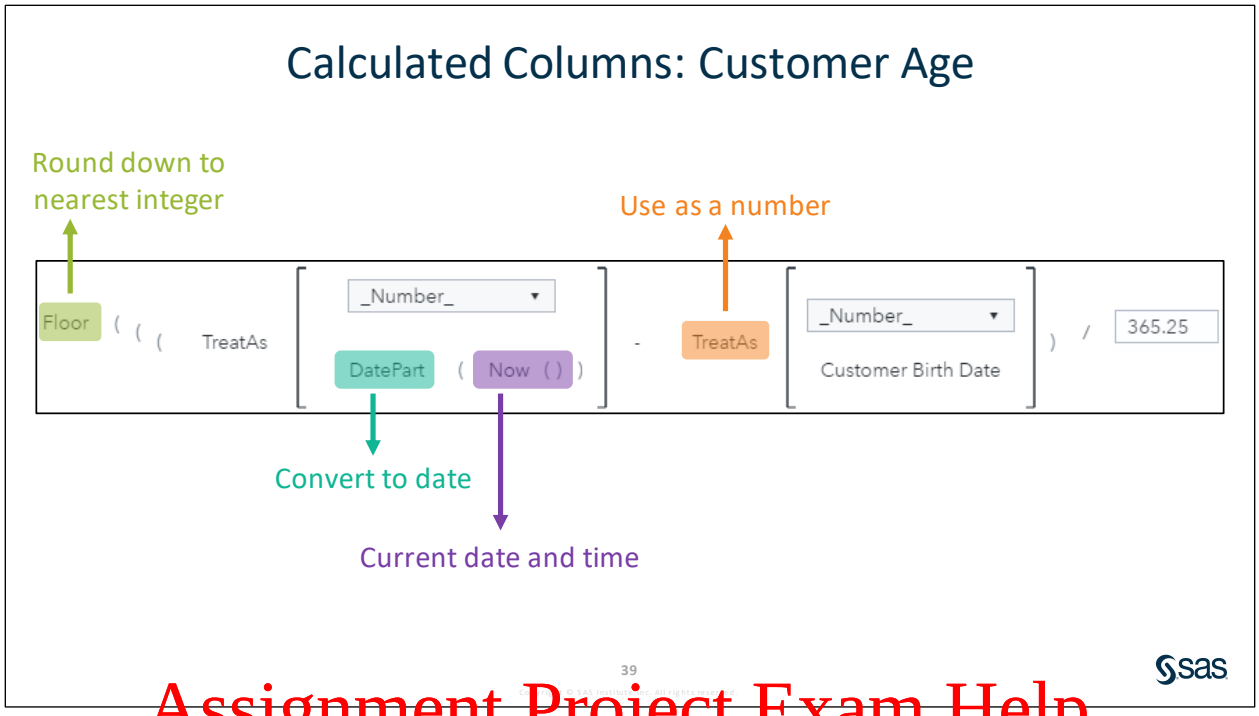
36

3.04 Activity

Given the values of **Customer Birth Date** and today's date, how would you calculate **Customer Age**?

Customer Birth Date
02Jan1983
28May1975
08May2008
19Sep2010
10Oct2017

37



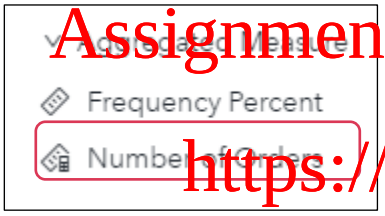
Note: SAS Visual Analytics treats datetime values as character data. To use numeric operators with datetime values, the TreatAs operator is required.

Creating Data Items

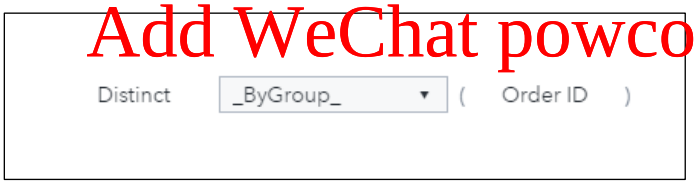
This demonstration illustrates how to create new data items (distinct counts, custom categories) in Visual Analytics.

1. From the browser window, sign in to SAS Viya.
2. In the upper left corner, click ☰ (**Show list of applications**) and select **Explore and Visualize**.

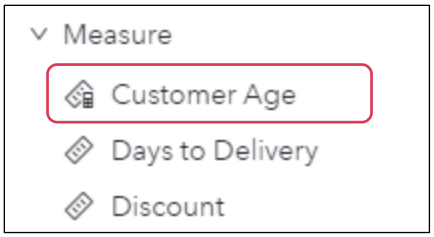
   SAS Visual Analytics appears.
3. Click **All Reports**.
   a. Navigate to the **Courses/YVA185/Basics/Demos (Marketing)** folder.
   b. Double-click the **VA1- Demo3.3a** report to open it.
4. In the upper left corner of the report, click the **Page 3** tab.
5. View new calculated items (**Number of Orders**, **Customer Age**, and **Customer Age Group**).
   a. In the left pane, click **Data**.
   b. In the Aggregated Measure group, view **Number of Orders** (new derived data item).



Note: You can view the calculation by right-clicking the calculated item and selecting **Edit**.

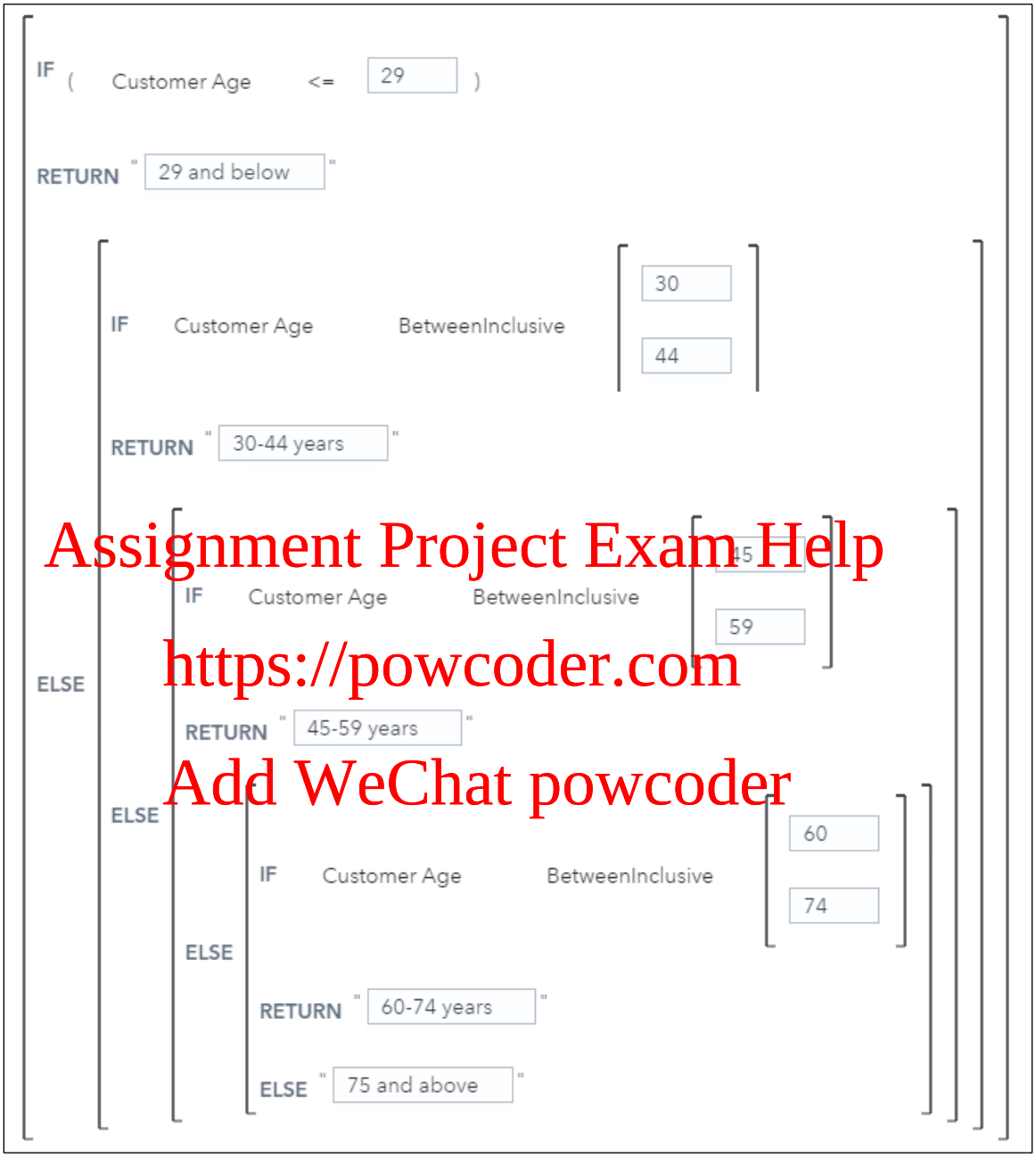


   c. In the Measure group, view **Customer Age** (new calculated data item).



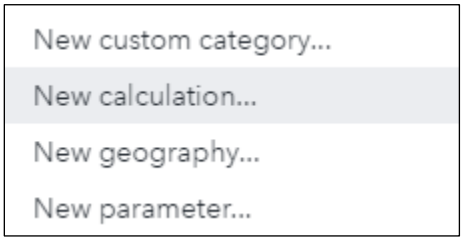
Note: You can view the calculation by right-clicking the calculated item and selecting **Edit**.

d. In the Category group, right-click **Customer Age Group** and select **Edit**.
The expression should resemble the following:

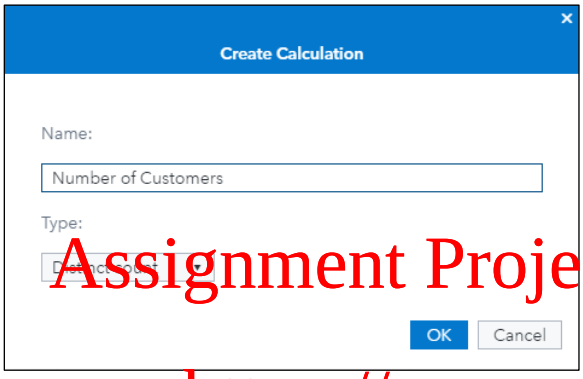


e. Click **Cancel** to close the Edit Calculated Item window.

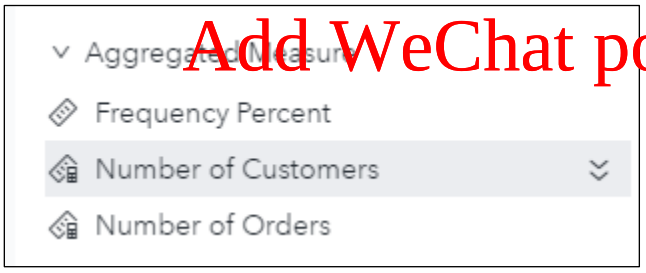
6. Create new distinct count data items.
   a. In the Category group, right-click **Customer ID** and select **New calculation**.



   b. In the **Name** field, enter **Number of Customers**.
   c. For the **Type** field, verify that **Distinct count** is selected.

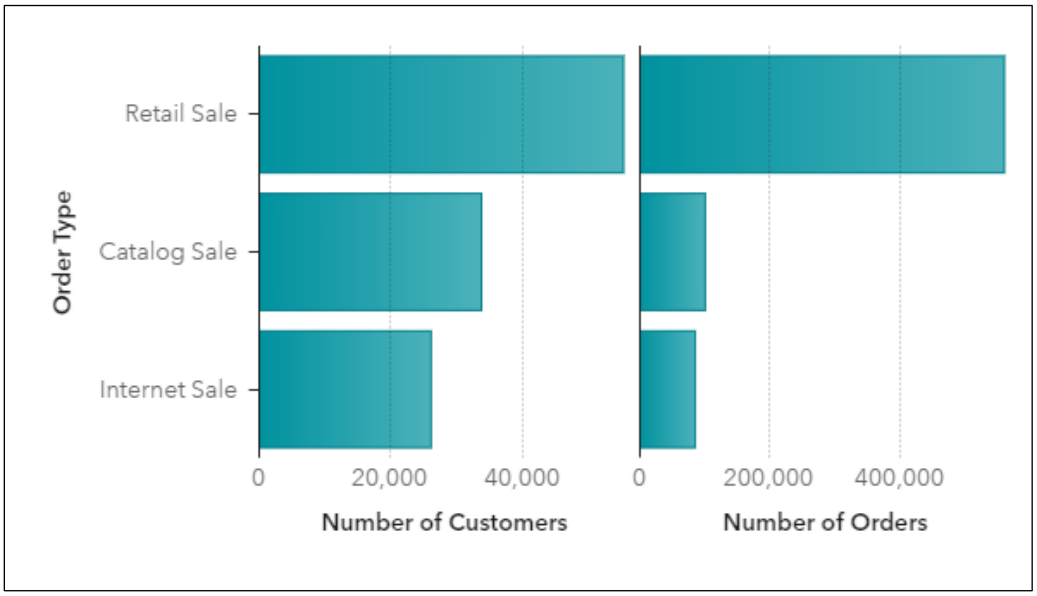


   d. Click **OK**.
   The new data item, **Number of Customers**, is added to the Aggregated Measure group.



7. Create an automatic chart.
   a. In the Data pane, select the following data items:
   Number of Orders
   Order Type
   Note: **Number of Customers** should already be selected.
   b. Drag the columns to the left side of the canvas.

The automatic chart functionality determines the best way to display the selected data.



Total profit is lower in the internet and catalog channels because there are fewer customers that place orders through those channels. There are also significantly fewer orders placed through those channels.

c. In the right pane, click **Options**.

d. In the Object group, for the **Name** field, enter **Customers and Orders by Order Type**.

8. Duplicate a data item and modify data item properties.

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

b. In the Measure group, right-click **Profit** and select **Duplicate**.

c. Next to the new data item, **Profit (1)**, click (**Edit properties**).

d. For the **Aggregation** field, select **Average**.

e. In the **Name** field, enter **Average Profit** and press Enter.

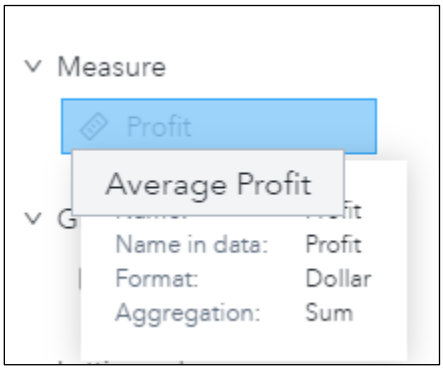
9. Modify the **Average Profits by Order Type and Continent** bar chart.

a. In the canvas, click the **Average Profit by Order Type and Continent** bar chart to make it active.

b. In the right pane, click **Roles**.

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

d. Drag **Average Profit** on top of **Profit** to replace the measure in the Roles pane.



The bar chart should resemble the following:

Ideally, we would want to increase orders placed for existing customers that produce the highest average profit. In this example, that would be Asian customers who order through the catalog. However, because corporate headquarters is located in North America, management has decided that the initial marketing strategy should focus on increasing sales among North American customers who order through the catalog and internet. Then, if the marketing strategy is successful, it is implemented in other locations.

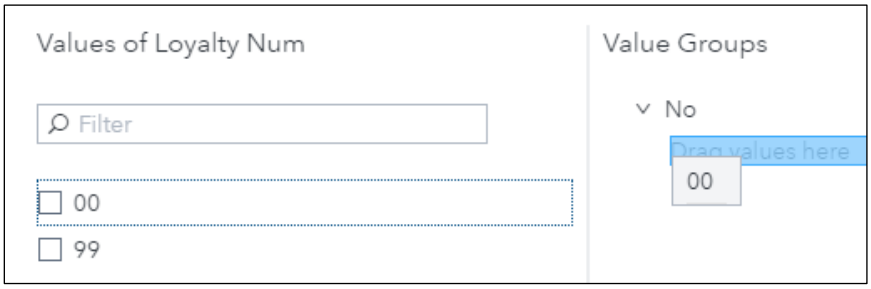
10. Create a new custom category, **Loyalty Member**.
    a. In the Data pane, select **New data item** ⇨ **Custom category**.
    b. In the New Custom Category window, in the **Name** field, enter **Loyalty Member**.

c. For the **Based on** field, select **Loyalty Num**.

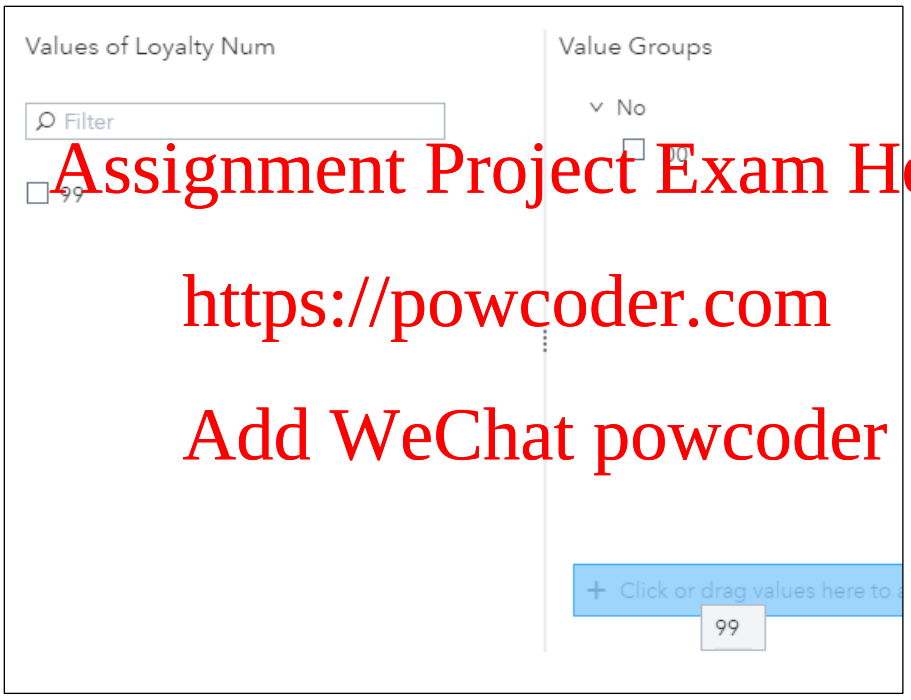
d. Select **Value Group 1** to edit the group name.

1) Type **No** and press Enter.

2) In the left pane, click **00** and drag to the **Drag values here** area on the right.



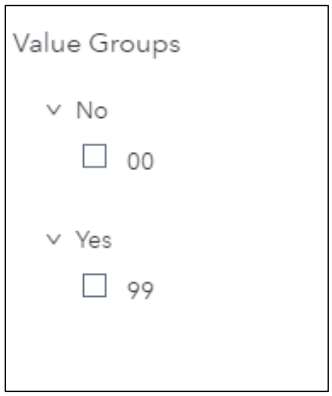
e. Drag **99** to the **Click or drag values here to add a value group** area.



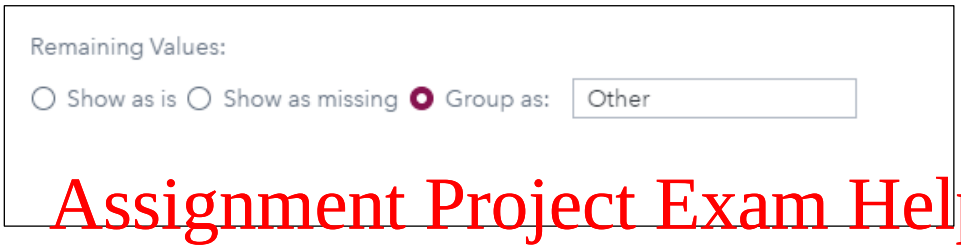
1) Select **Value Group 1**.

2) Type **Yes** and press Enter.

The Value Groups area should resemble the following:

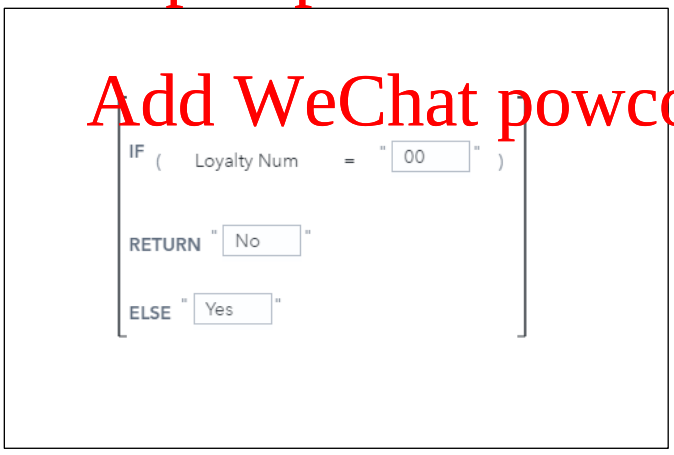


f. In the Remaining Values area, for the **Group as** field, verify that **Other** is specified.



g. Click **OK** to create the new custom category.

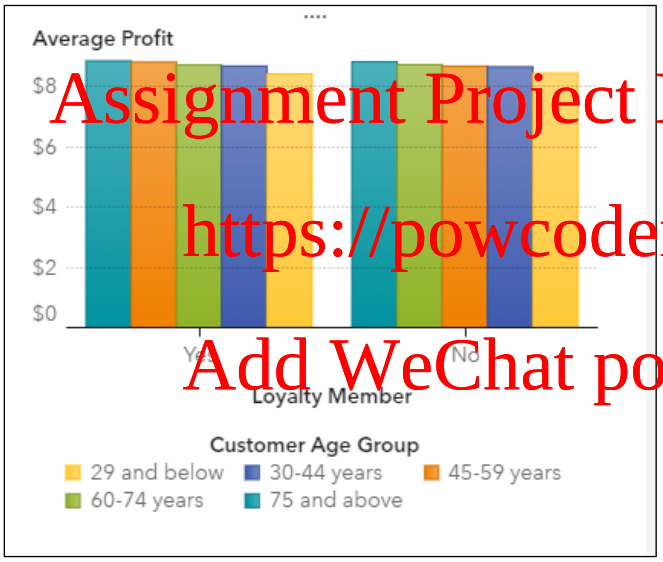
Note: As an alternative, you can also create a calculated data item with the following expression:



The new data item, **Loyalty Member**, should appear in the Category group.

11. Duplicate the **Average Profits by Order Type and Continent** bar chart.
    a. In the canvas area, right-click the **Average Profits by Order Type and Continent** bar chart and select **Duplicate** to copy the bar chart.
    b. Click ▪▪▪▪ above the new bar chart and drag to the drop zone to the bottom of the **Average Profits by Order Type and Continent** bar chart.
    c. In the right pane, click **Roles**.
    d. For the **Category** role, select **Order Type** ⇨ **Loyalty Member**.
    e. For the **Group** role, select **Continent Name** ⇨ **Customer Age Group**.
    f. In the right pane, click **Options**.
    g. In the Object group, for the **Name** field, enter **Average Profits by Loyalty Membership and Age Group**.
    h. In the Bar group, for the **Direction** field, click (**Vertical**).

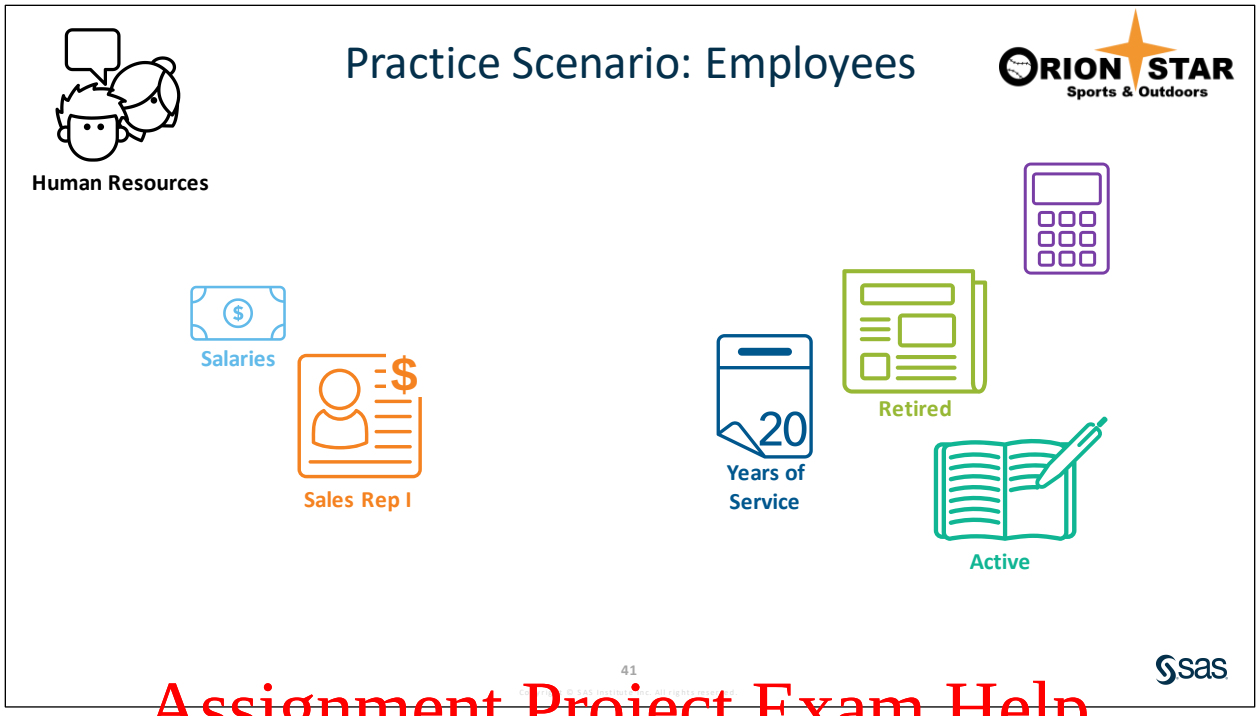
    The bar chart should resemble the following:



    Average profits are similar across loyalty members and non-loyalty members. Average profits are slightly higher for loyalty members in the 75 and above age group.

12. Save the report.

End of Demonstration



3.05 Activity

Given the values of **Employee Hire Date** and **Employee Termination Date**, how would you calculate **Years of Service**?

Employee Hire Date	Employee Termination Date
01Dec2004	28Feb2007
01Nov2005	.
25Jan2005	.
01Mar2005	28Feb2010
31May2005	31May2012
11Dec2005	.
01Sep2002	.

42

Practice

4. Creating Data Items

a. Open the browser and sign in to SAS Viya.

b. Open the **VA1- Practice3.3a** report from the **Courses/YVA185/Basics/Practices (HR)** folder.

c. Create a new data item, **Employee Status**, by assigning the following labels to the values:

Employee Status (label)	Employee Termination Date (value)
Active	.
Retired	<all remaining values>

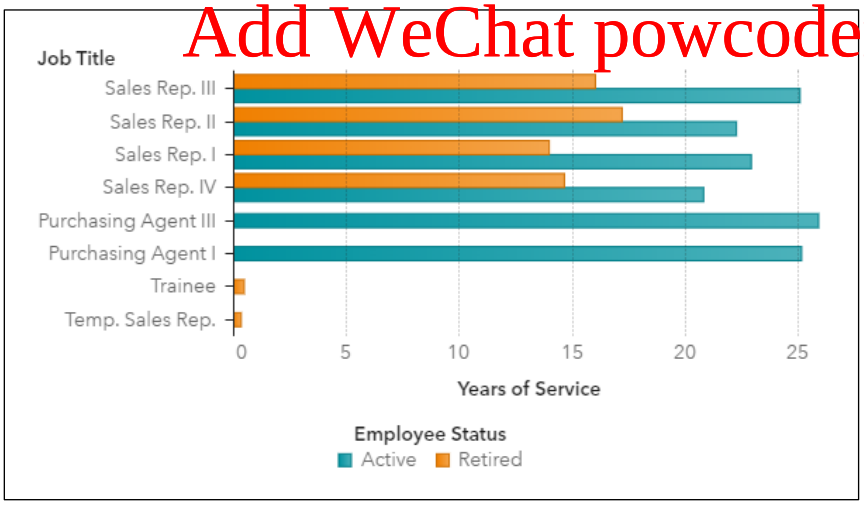
d. On Page 3, create a bar chart by assigning the following data items to the specified roles:

Category	**Job Title**
Measure	**Years of Service**
Group	**Employee Status**

e. Specify **Years of Service by Job Title and Status** as the name of the bar chart.

f. Change the aggregation for **Years of Service** to **Average**.

The bar chart should resemble the following:



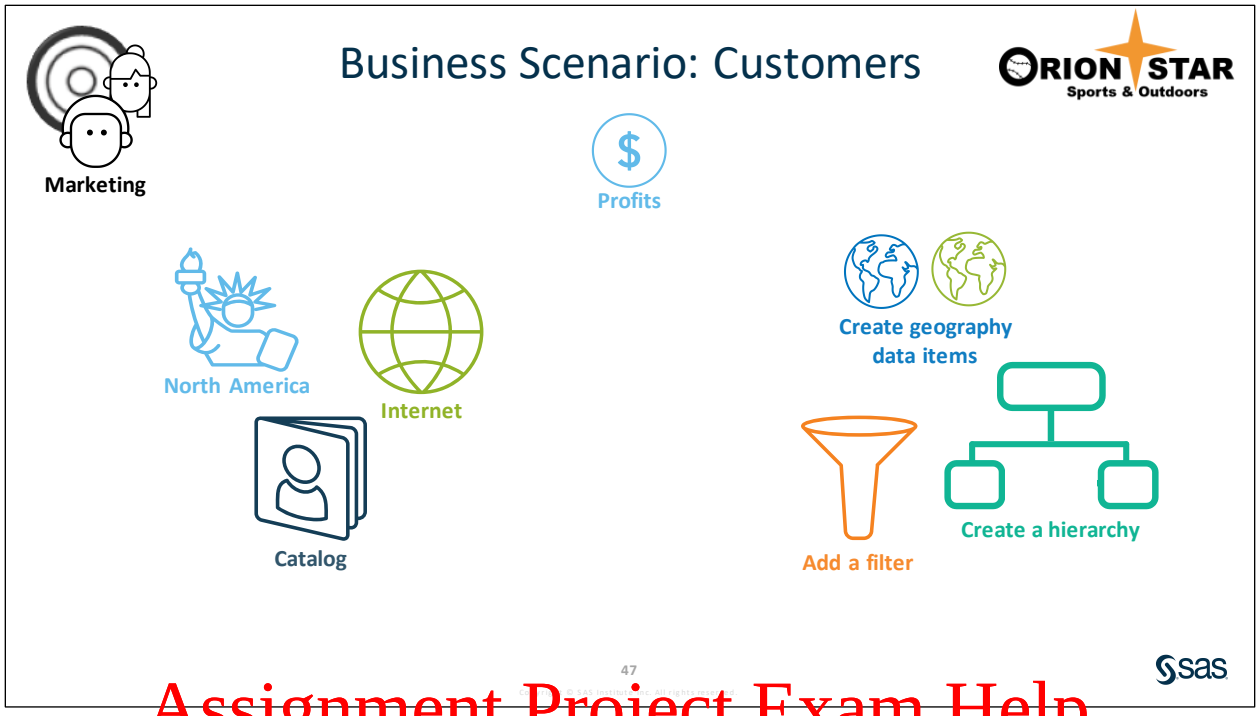
g. Answer the following questions:

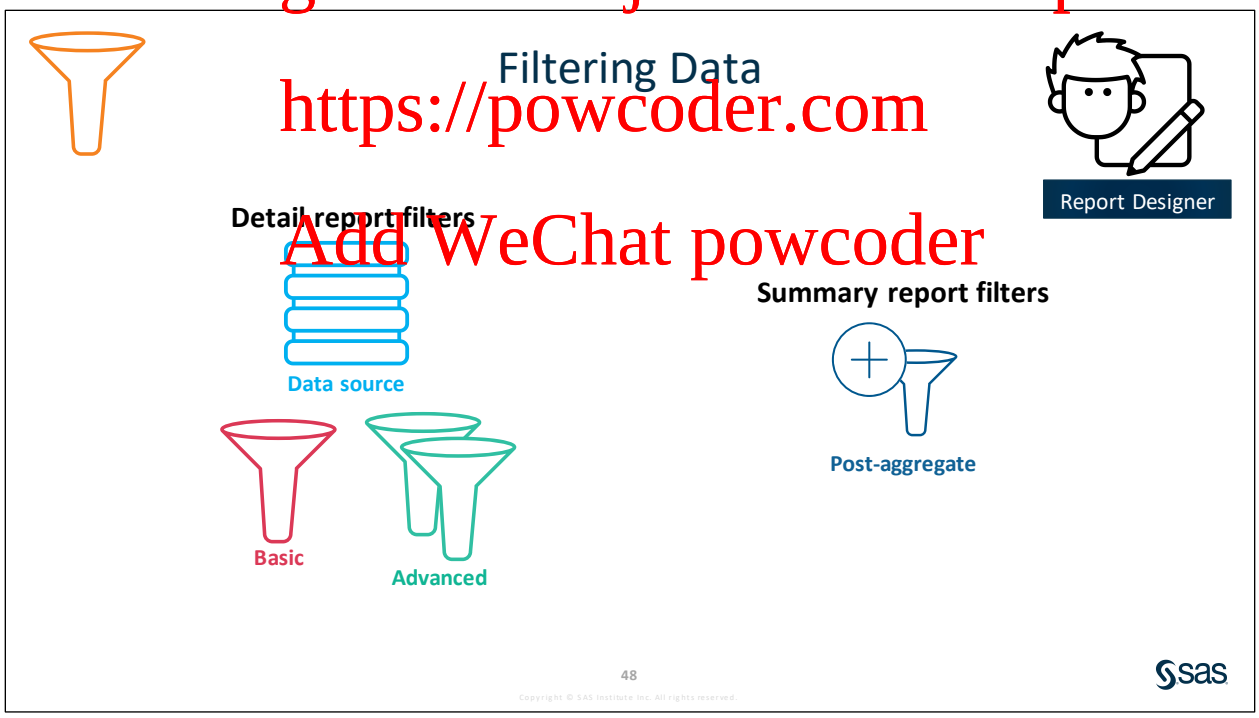
Which job title has the highest average years of service among active employees? Among retired employees?

Answer: _____

h. Save the report.

End of Practices





The following types of filters can be created and modified only by the report designer:

Data source filter	Subsets the data for the entire report and is applied to every report object that uses that data source. The data source filter acts as a pre-filter, by filtering the data before it is brought into Visual Analytics. This can be seen by the updated cardinality values in the Data pane after the filter has been applied.
Basic report filter	Subsets the data for individual report objects by using a single data item.
Advanced report filter	Subsets the data for individual report objects by using any number of data items and operators in the same expression.
Post-aggregate report filter	Subsets the data for individual report objects by using aggregated values, not summarized values. Post-aggregate report filters are available only for measure data items.

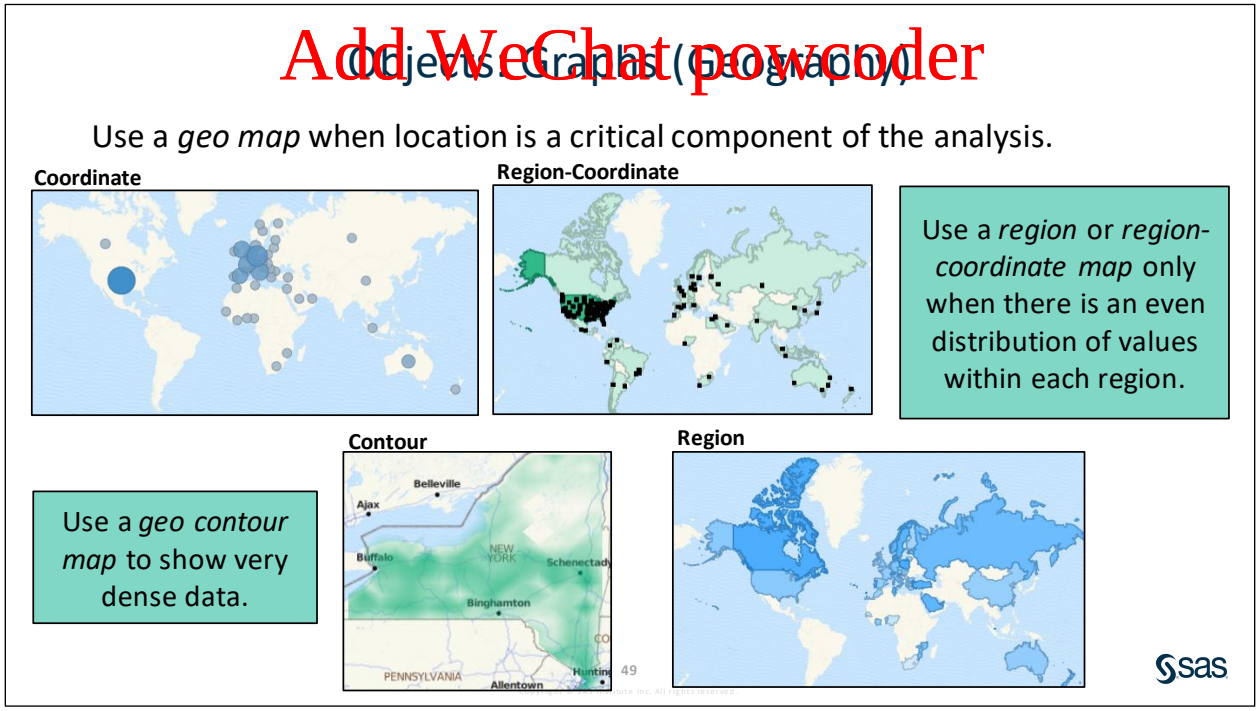
For more information about filters that can be created and modified by the report designer, see "Working with Report Filters" in the *SAS Visual Analytics: Working with Report Data* documentation.

Filters that can be modified by report viewers are discussed in more detail in a later section.

Filters are applied in the following order:

- data source filter/ cr filters
- basic or advanced report filter/ post-aggregate report filter
- prompts and actions

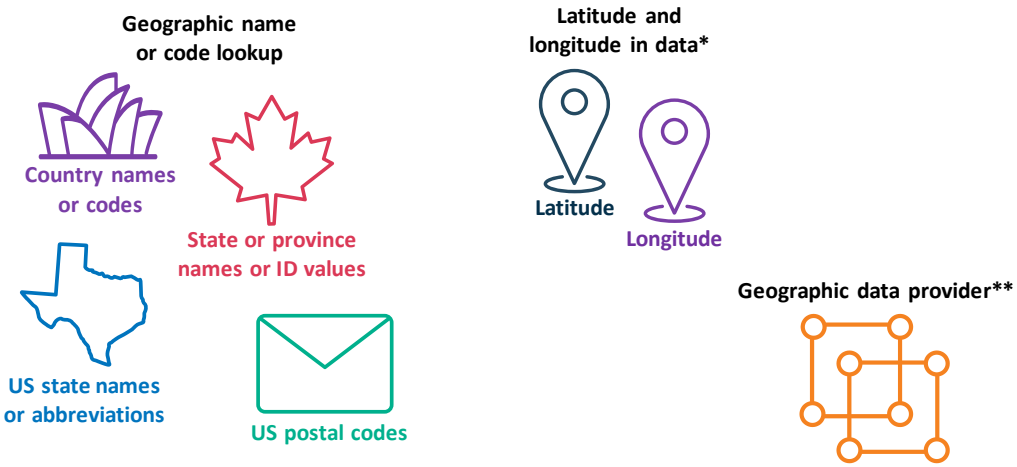
Note: More advanced filtering techniques are discussed in the SAS Visual Analytics 2 for SAS Viya: Advanced course.



Geo map	A geo map overlays data on a geographic map. Data can be displayed using colored regions, coordinates, or regions and coordinates, as a contour plot, or as a network. In order to display data on a geo map, at least one category data item must have values that are mapped to geographical locations or regions.	
	Region	A regions geo map (also known as a *choropleth map*) uses colors to show variations by location. However, larger regions appear more emphasized than smaller ones, which can affect perceptions of colors.
	Coordinate	A coordinates geo map (also known as a *dot distribution map* or a *dot density map*) displays a map with either a scatter plot or a bubble plot of coordinates. This type of map helps with detecting spatial patterns and understanding the distribution of data over a geographical region, which can help reveal patterns using clustered points. For a bubble plot, the bubble size helps with comparing proportions over regions without the size of the region causing distortions, but the size of the bubble can overlap with other bubbles and regions making the chart difficult to read.
	Region-Coordinate	A region-coordinate geo map displays a map using both colored regions and either a scatter plot or bubble plot of coordinates. This type of map is great for comparing two levels of data with the region colors representing more general information (like countries) and the coordinates representing more specific information (like customer locations).
	Contour	A contour geo map displays shaded regions over a geographical region. Contour maps are best used to show very dense data.
	Network	A network geo map displays a network diagram overlaid on a map. Network maps are helpful for understanding how location affects the relationships in the network. Network geo maps are discussed in more detail in a later lesson.

For more information about creating geography data items, see “Working with Geography Data Items” in the *SAS Visual Analytics: Working with Report Data* documentation.

Geography Data Items



50

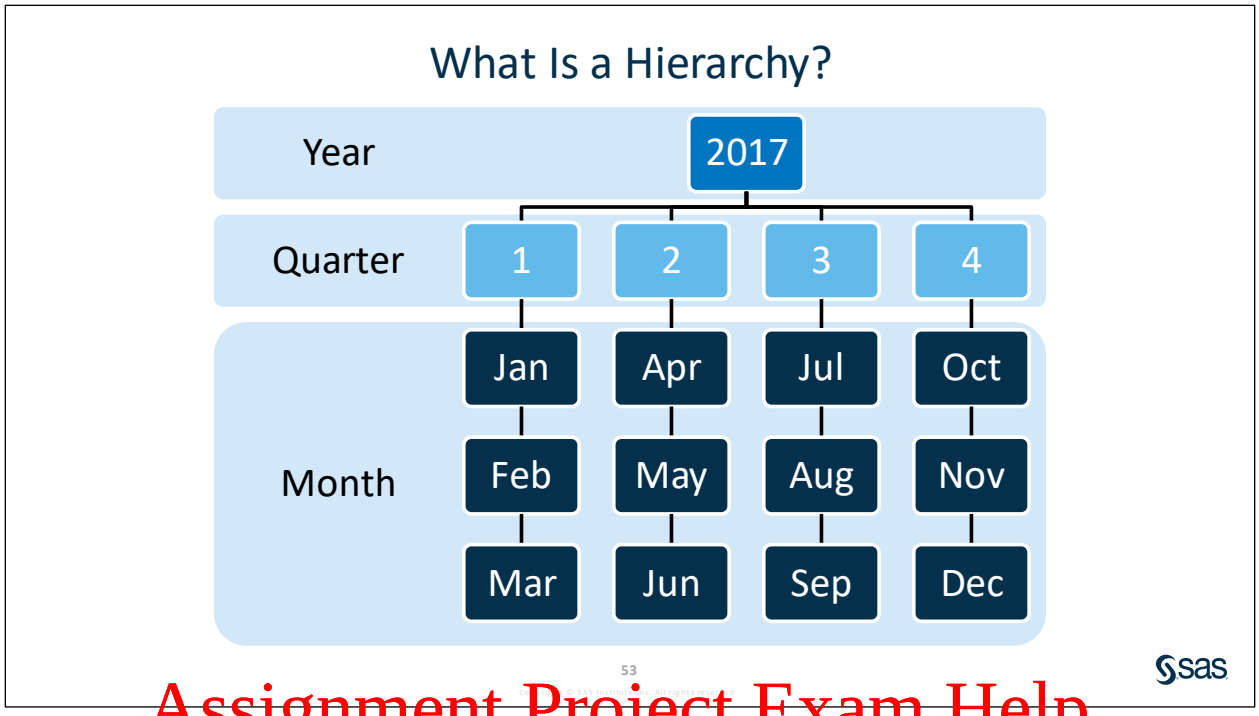
Note: By default, Visual Analytics supports country- and state-level polygons for regional overlays in geo maps. An administrator can define a custom polygon provider to create regional overlays for other types of data. For more information about how to define a custom polygon provider, see "Loading Geographic Polygon Data as a CAS Table" in *SAS Visual Analytics: Administration Guide*.

3.06 Multiple-Answer Question

Which object can use a data item that has a classification type of geography?

a. crosstab
b. geo map
c. table
d. bar chart

51



A *hierarchy* is a defined arrangement of categorical data items based on parent-child relationships.

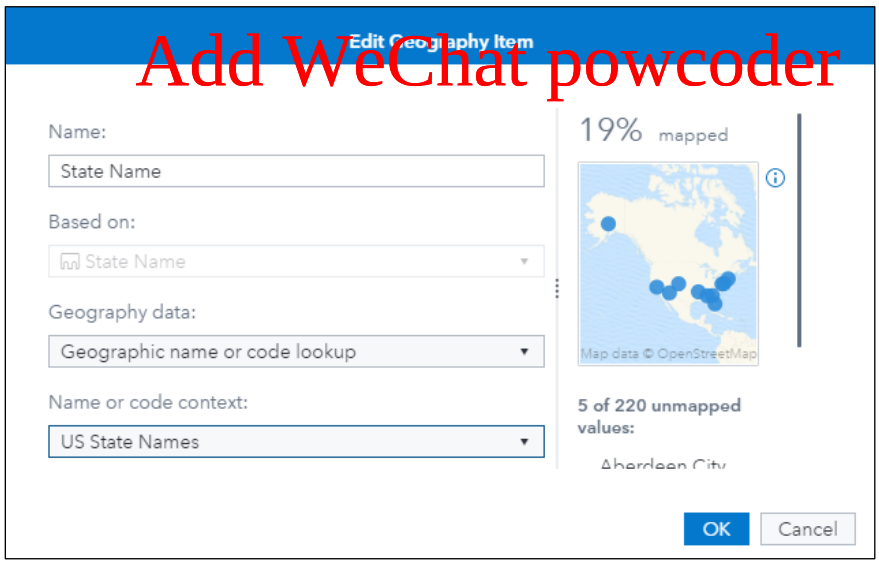
Applying Filters

This demonstration illustrates how to create new data items (geographic data items, hierarchies) and apply filters in Visual Analytics.

1. From the browser window, sign in to SAS Viya.
2. In the upper left corner, click (**Show list of applications**) and select **Explore and Visualize**.

   SAS Visual Analytics appears.
3. Click **All Reports**.
   a. Navigate to the **Courses/YVA185/Basics/Demos (Marketing)** folder.
   b. Double-click the **VA1- Demo3.3b** report to open it.
4. In the upper left corner of the report, click the **Page 4** tab.
5. Create new data items.
   a. In the left pane, click **Data**.
   b. In the Category group, next to **State Name**, click (**Edit properties**).
   c. For the **Classification** field, select **Geography**.
      1) For the **Geography data** field, verify that **Geographic name or code lookup** is selected.
      2) For the **Name or code context** field, select **US State Names**.

      The map on the right shows that 19% of state names are mapped.

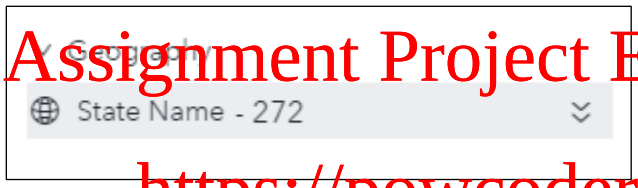


3) View the list of unmapped values.

These values represent states and provinces in other countries. Later, we add a data source filter to focus on the United States.

4) Click **OK**.

A new group, **Geography**, is added to the Data pane.



d. In the Data pane, next to **Postal code**, click (**Edit properties**).

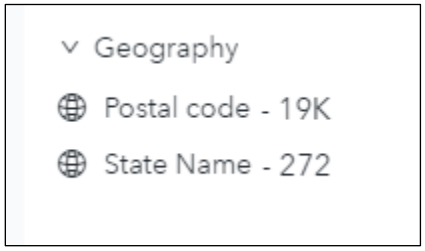
e. For the **Classification** field, select **Geography**.

1) For the **Geography data** field, verify that **Geographic names or code lookup** is selected.

2) For the **Name or code context** field, select **US ZIP Codes**.

3) Click **OK**.

The Geography group should resemble the following:



f. In the Data pane, select **New data item** ⇨ **Hierarchy**.

1) In the New Hierarchy window, for the **Name** field, enter **US Hierarchy**.

2) Double-click the following data items in the Available items list, in the specified order, to move them to the Selected items list:

State Name

Postal code

The New Hierarchy window should resemble the following:

3) Click **OK**.

A new group, **Hierarchy**, is added to the Data pane.

6. Add a data source filter.

a. In the Data pane, click (**Actions**) and select **Apply data filter**.

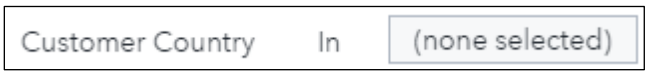
Note: Because the new geography data items cover only the United States, a data source filter is added to include only the data for products ordered in the United States.

b. On the left, verify that **Data Items** is selected.

c. Expand the **Character** group.

d. Select **Customer Country**.

e. In the Conditions area, double-click **Customer Country In (x)** to add it to the expression area.



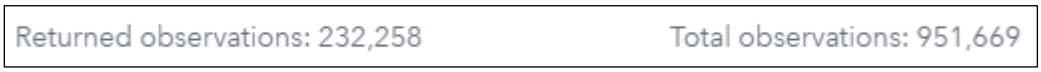
f. In the expression area, click **(none selected)**.

g. In the Select Data Values window, double-click **United States** to move it from the Available items list to the Selected items list.

h. Click **OK**.

The expression should resemble the following:

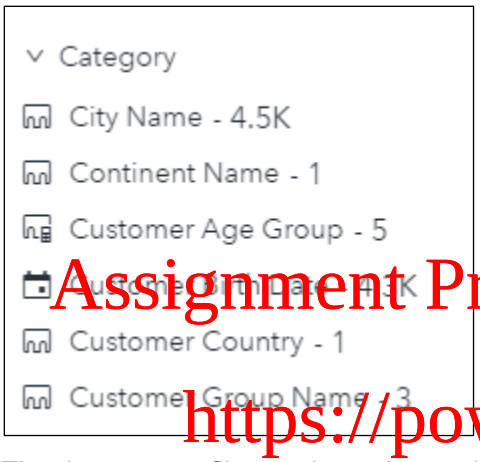
The bottom of the Apply Data Filter window should resemble the following:



Note: 232,258 observations have a value of *United States* for **Customer Country**.

i. Click **OK** to apply the data source filter.

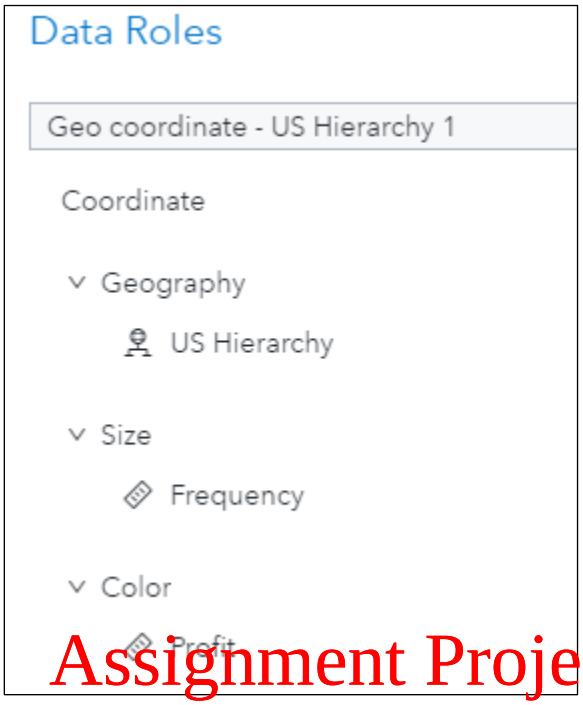
The Data pane should resemble the following:



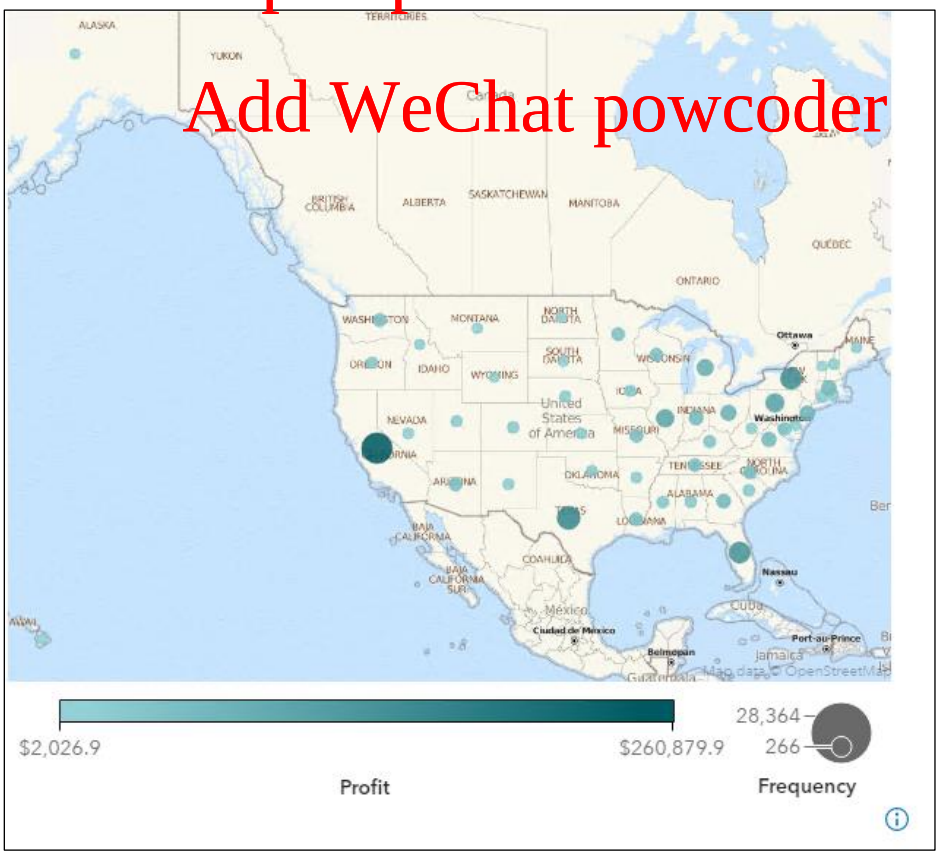
The data source filter updates the cardinality values that appear in the Data pane and is applied to every report object that uses this data source.

7. Create a geo map.
   a. In the left pane, click **Objects**.
   b. Drag the **Geo coordinate** object, from the Geographic group, to the canvas.
   c. In the right pane, click **Roles**.
   d. For the **Geography** role, select **Add** ⇨ **US Hierarchy**.
   e. For the **Size** role, **Add** ⇨ **Frequency**.
   f. For the **Color** role, select **Add** ⇨ **Profit**.

The Roles pane should resemble the following:



The geo coordinate map requires a geography data item for the Geography role. A measure data item can be added to the Color role to color the coordinates based on the measure. The geo map should resemble the following:



g. Place your cursor over ⓘ in the lower right corner of the geo map to view the warning.

No matches were found for supplied geography data items: PR
Some features may not be displayed on the map because of missing location information in the data.

Note: *PR* is not found in the US State Names predefined geographic role. You can filter this value out if you do not want to see the warning.

h. In the right pane, click **Options**.

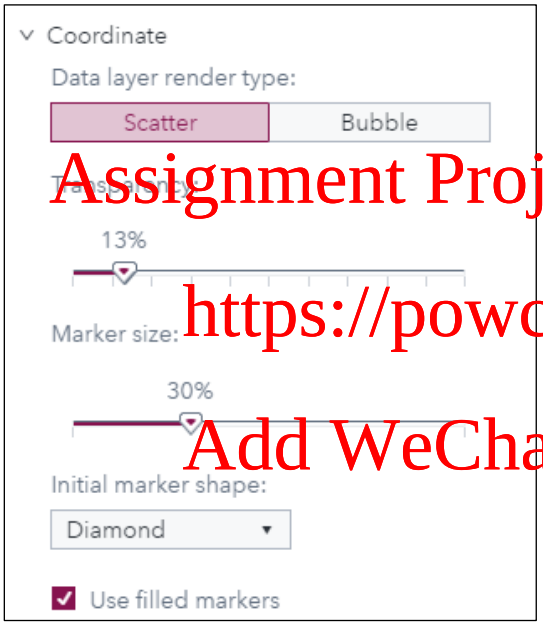
i. In the Object group, for the **Name** field, enter **Profit by Location**.

j. In the Coordinate group, verify that **Scatter** is selected for the **Data layer render type** field.

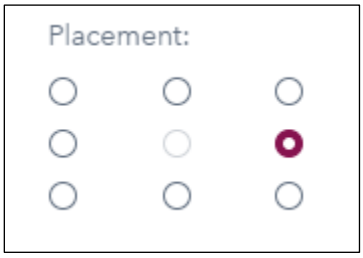
k. For the **Initial marker shape** field, select **Diamond**.

l. For the **Marker size** field, select **30**.

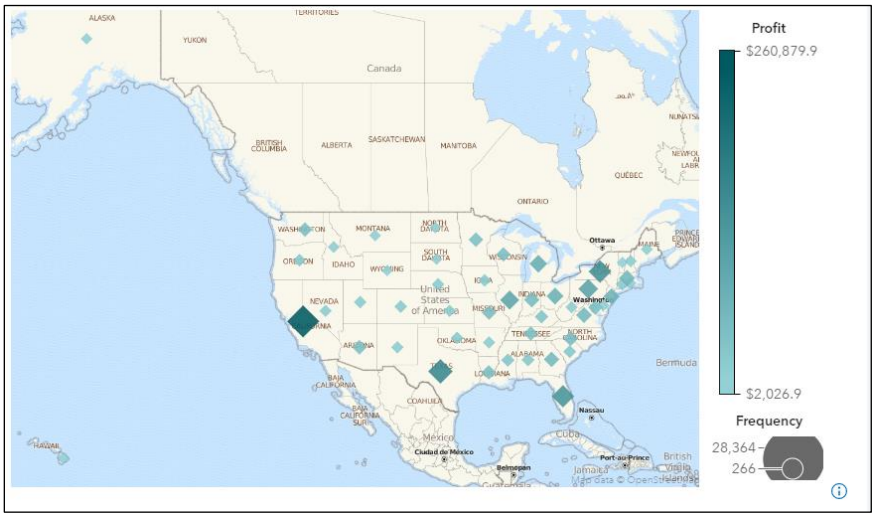
The Options pane should resemble the following:



m. In the Legend group, for the **Placement** field, choose the middle on the right side.



The updated geo map should resemble the following:



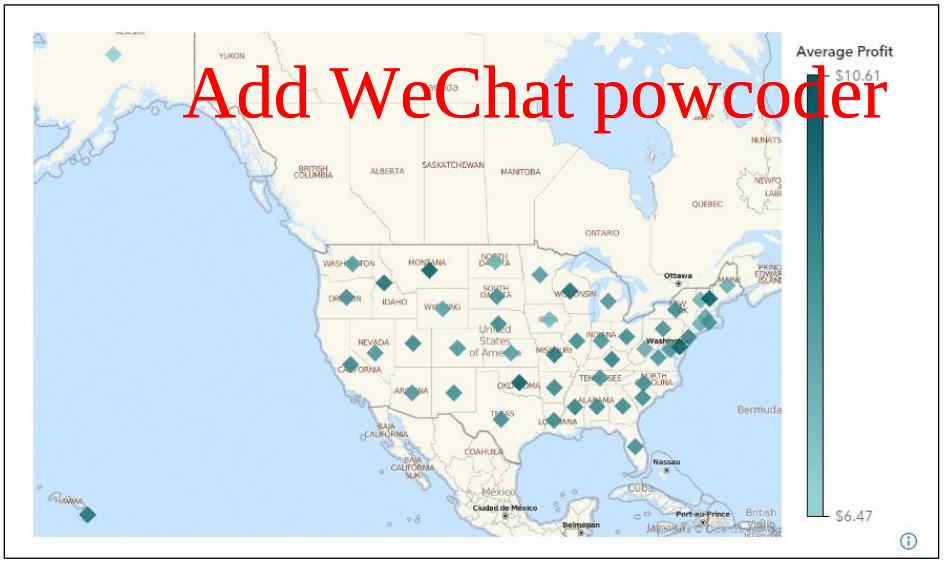
Highest total profits seem to be in larger states (California, Texas, and Florida), most likely because there are more customers and more orders placed in those states. Looking at average profits by location can give greater insight into orders placed in the United States.

n. Right-click the **Geo coordinate map** object and select **Remove data** ⇨ **Frequency**

o. Right-click the **Geo coordinate map** object and select **Replace data** ⇨ **Profit**.

p. Select **Average Profit**.

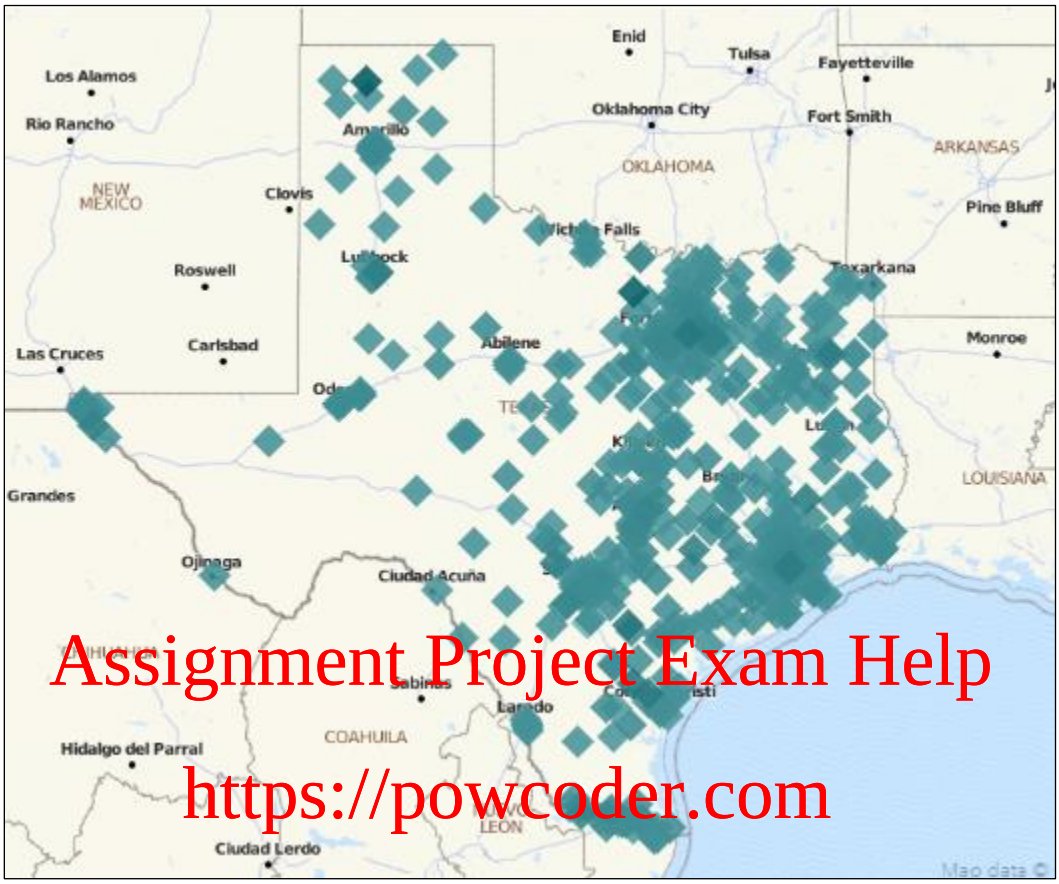
The updated geo map should resemble the following:



When looking at averages, there does not seem to be any clusters of higher average profits in any one location in the United States. High average profits seem to be evenly distributed across the United States.

q. Double-click the marker for **Texas**.

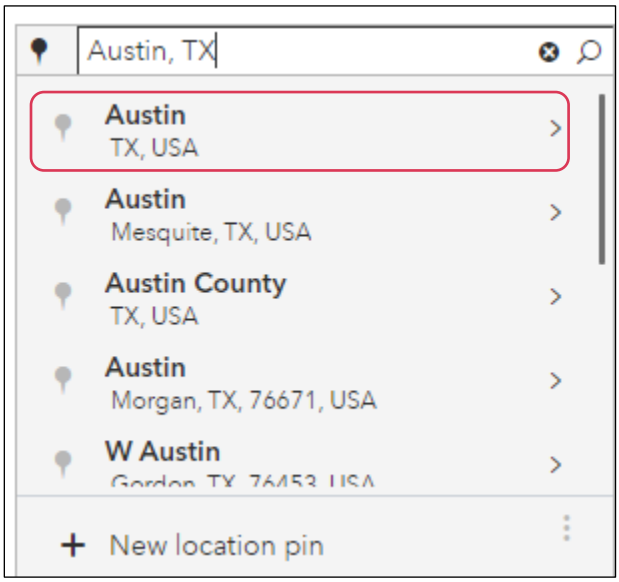
The geo map displays markers for all postal codes in Texas where products were ordered.



r. In the upper left corner, click (**Location**).

s. In the **Search** field, enter **Austin, TX**.

t. Double-click the first value in the search list, **Austin TX, USA**.



All locations containing combinations of **Austin, TX** are listed in the search. The location of Austin, Texas is marked on the geo map with a 1.

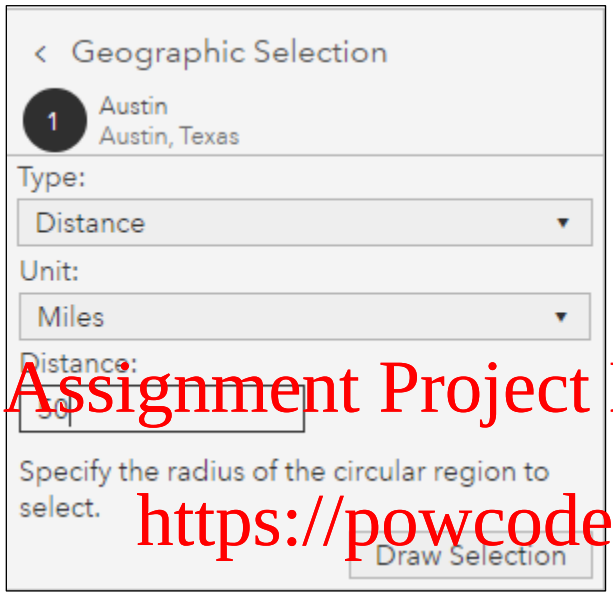
u. Select **Geographic selection**.

1) For the **Type** field, verify that **Distance** is selected.

2) For the **Unit** field, verify that **Miles** is selected.

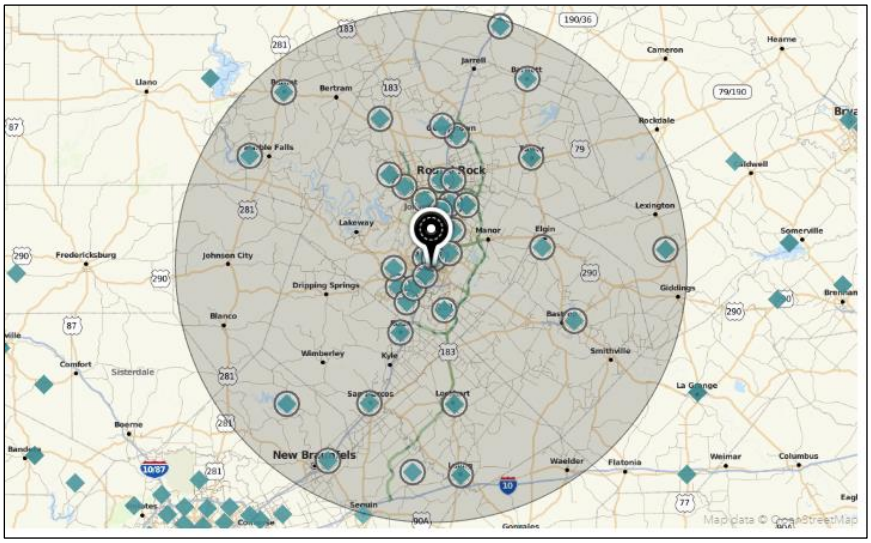
3) For the **Distance** field, enter **50**.

The Geographic Selection window should resemble the following:



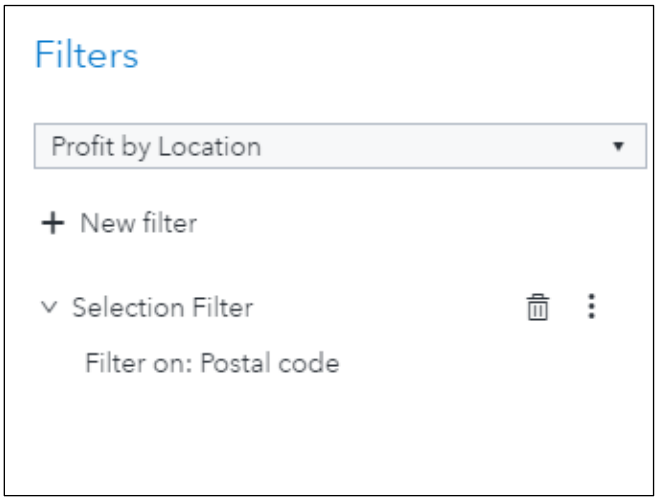
4) Click **Draw Selection**.

All customers within a 50-mile radius of Austin, TX, are highlighted.



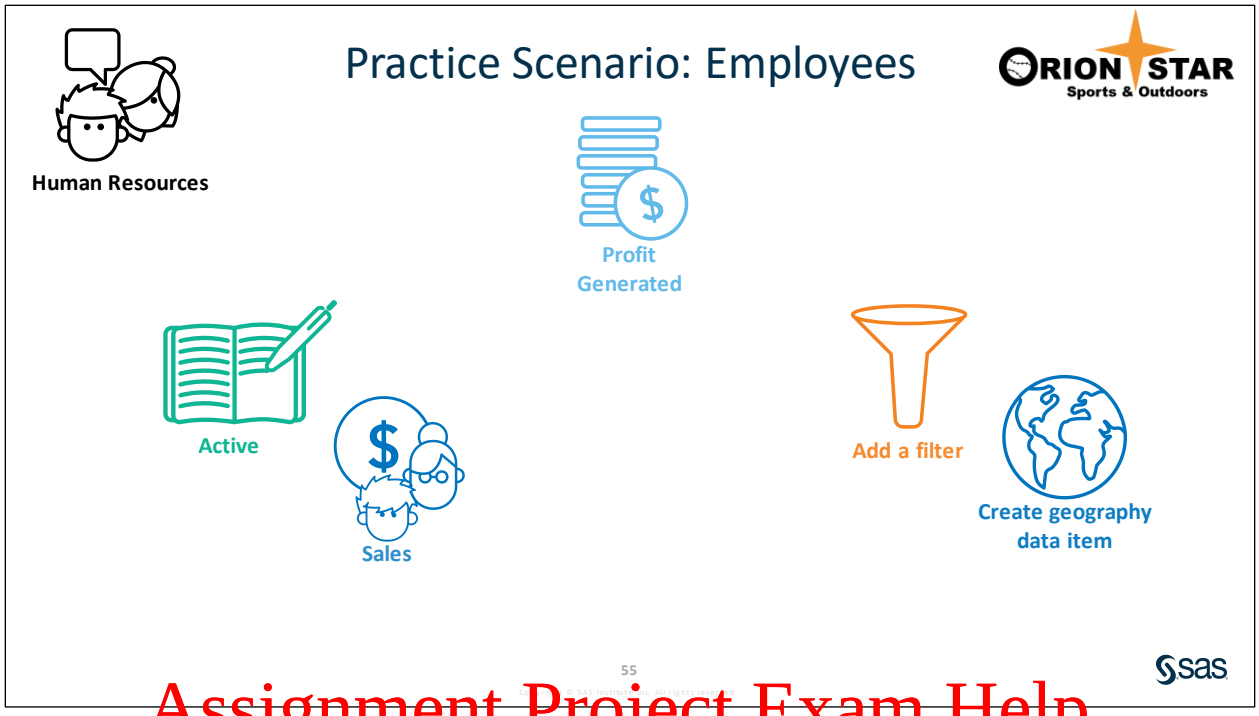
v. Right-click the Geo coordinate map and select **New filter from selection** ⇨ **Include only selection**.

w. In the right pane, click **Filters** to show the applied filter.



8. Save the report.

End of Demonstration



Practice

5. Applying Filters

a. Open the browser and sign in to SAS Viya.

b. Open the **VA1- Practice3.3b** report from the **Courses/YVA185/Basics/Practices (HR)** folder.

c. Add a data source filter to filter for active employees in the Sales Department.

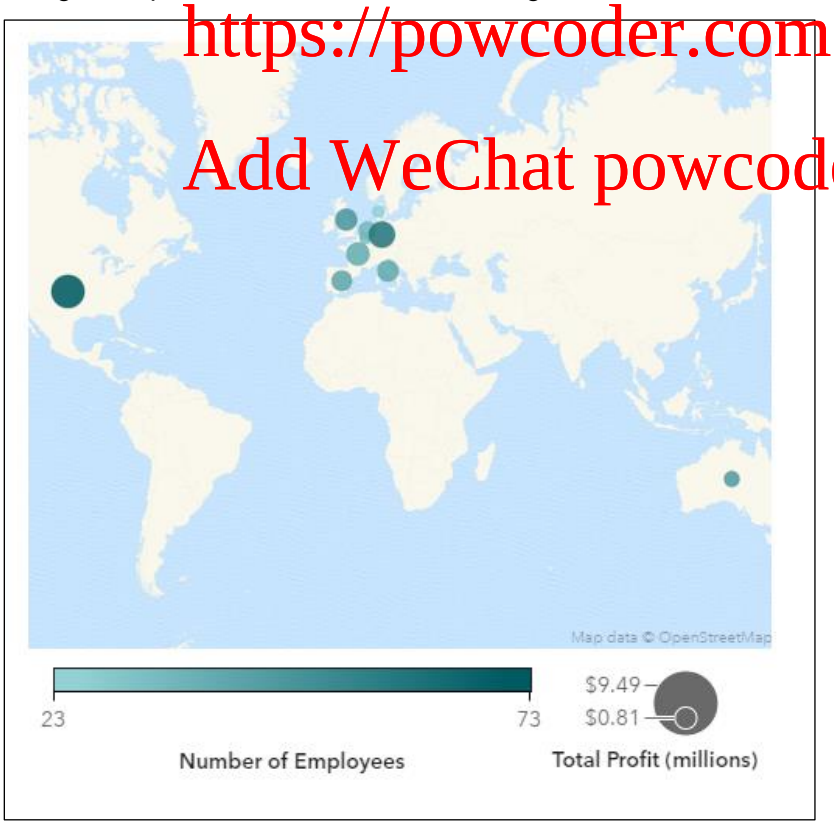
Note: Use the AND operator (in the Boolean group) to filter for multiple conditions. After the data source filter is applied, 429 observations should be returned.

d. Change the classification for **Employee Country** to **Geography ⇨ Country or Region ISO 2-Letter Codes**.

e. On Page 4, create a geo coordinate map by assigning the following data items to the specified roles:

Category	**Employee Country**
Size	**Total Profit**
Color	**Number of Employees**

The geo map should resemble the following:



f. Maximize the geo map to answer the following questions:

Management has decided that one possible criterion for promotion is profit generated. Which two countries generate the highest profit? Why do they have such high profits?

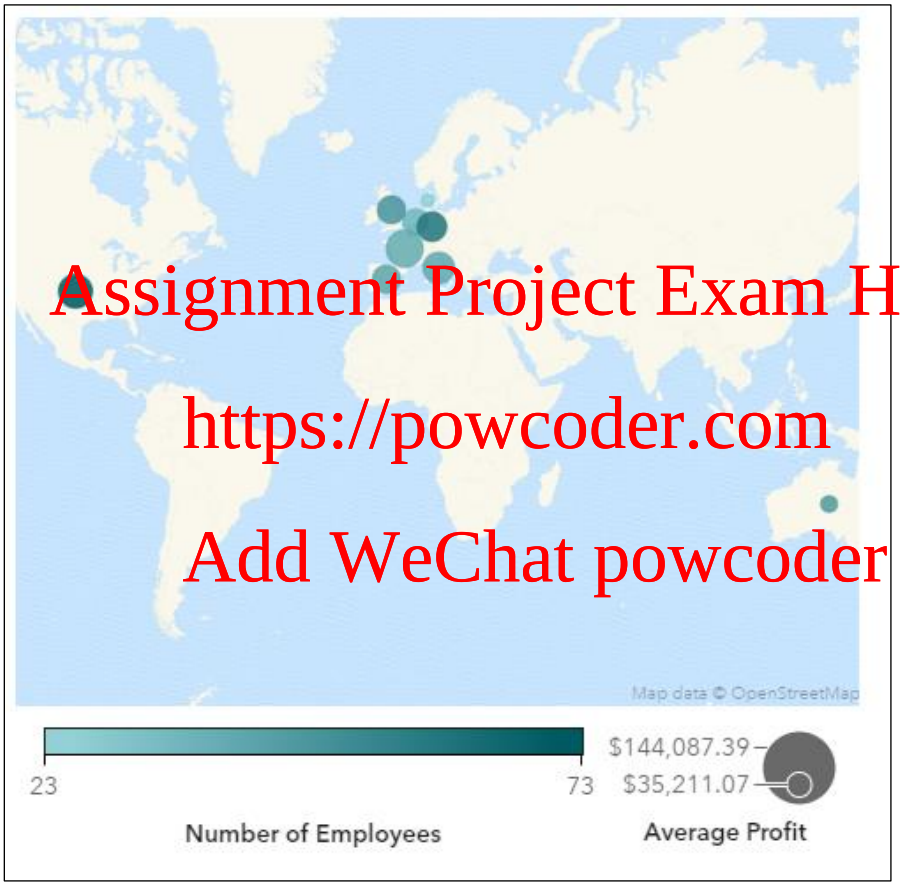
Answer: _____

Hint: After answering the questions, click (**Restore**) in the upper right corner.

g. In the geo map, specify **Average Profit** for the **Size** role.

h. Specify **Average Profit and Number of Employees by Country** as the name of the geo map.

The updated geo map should resemble the following:



i. Maximize the geo map to answer the following question:

Which country has the highest average profit? Highest number of employees?

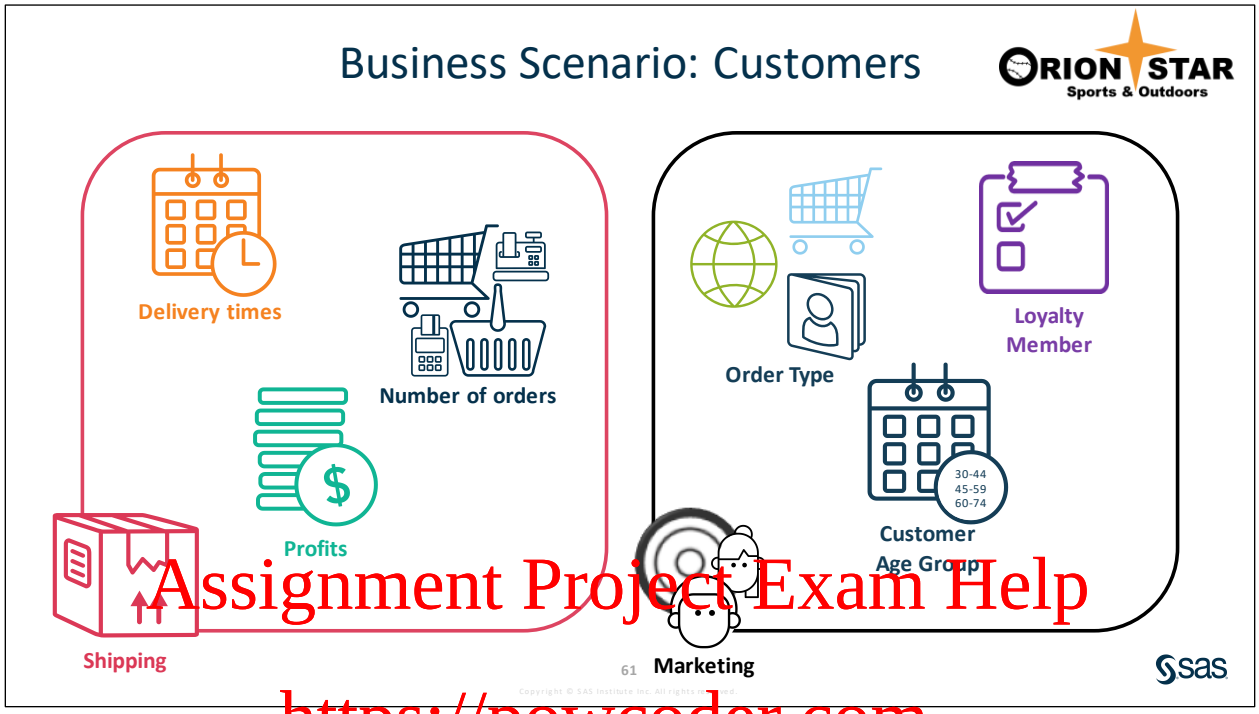
Answer: _____

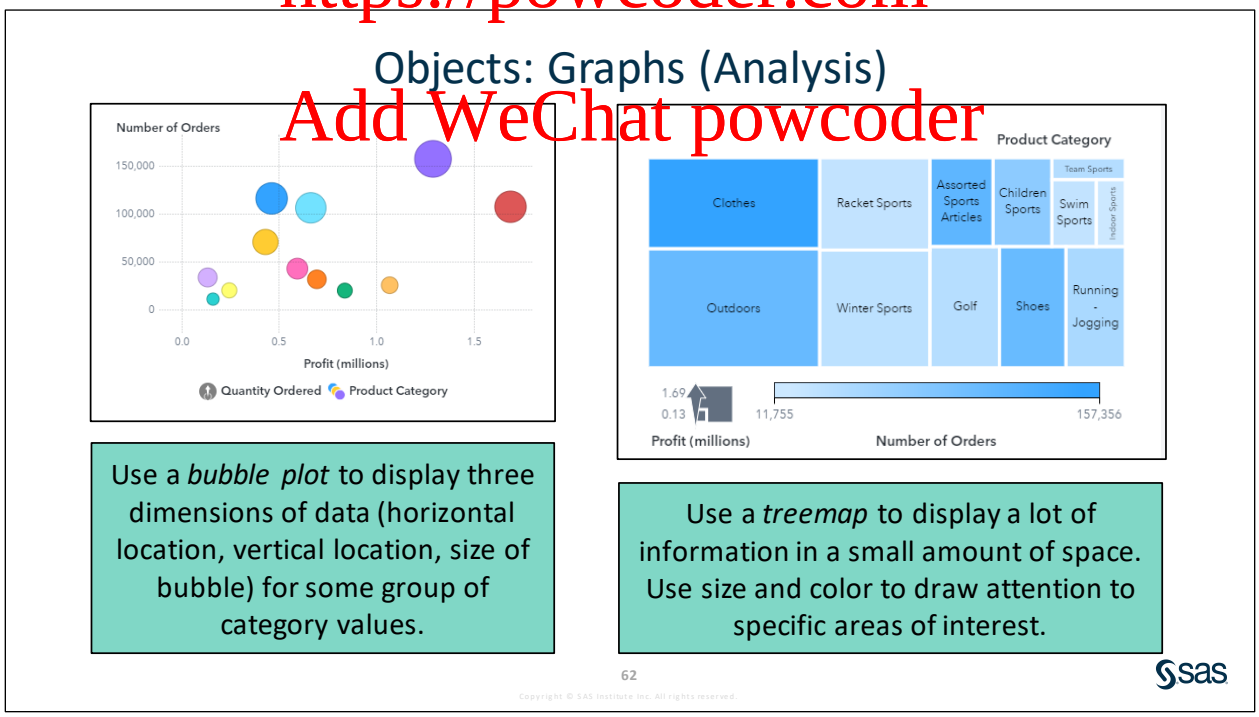
Hint: After answering the question, click (**Restore**) in the upper right corner.

j. Save the report.

End of Practices

3.4 Performing Data Analysis





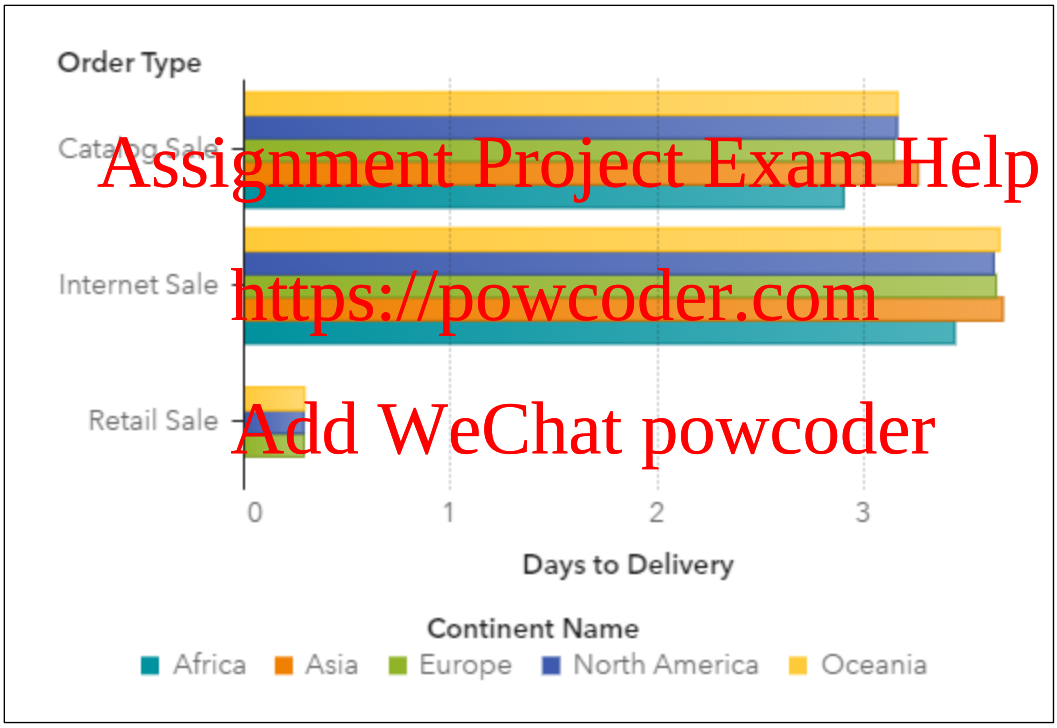
Bubble plot	A bubble plot displays the values of at least three measures by using plot markers (bubbles) of varying sizes in a scatter plot. The values of two measures determine the location of the bubble in the plot, and the value of the third measure determines the size of the bubble. Bubble plots can be animated to show changes in data over time. **Note:** A bubble's size is scaled relative to the minimum and maximum values of the size variable.
Treemap	A treemap displays a hierarchy or category as a set of rectangular tiles. The value of a category or hierarchy node is represented by tiles, and measures can be added to both size and color the tiles. Typically, the size and color are used to draw attention to areas of interest (for example, top contributors). The measures used to size and color the tiles should mean something when compared. Do not use the same measure for both the size and color as this violates the law of redundancy. The measure used to size the tiles cannot be below zero and must have an aggregation of sum. **Note:** The layout of the tiles in the treemap is dependent on the size of the display area because it uses a space-filling algorithm to lay the tiles out. This means that the same treemap might appear slightly different while editing a report than it does while viewing a report or in the Visual Analytics app.

Analyzing Data

This demonstration illustrates how to analyze data with graphs in Visual Analytics.

1. From the browser window, sign in to SAS Viya.
2. In the upper left corner, click ☰ (**Show list of applications**) and select **Explore and Visualize**.

   SAS Visual Analytics appears.
3. Click **All Reports**.
   a. Navigate to the **Courses/YVA185/Basics/Demos (Marketing)** folder.
   b. Double-click the **VA1- Demo3.4a** report to open it.
4. In the upper left corner of the report, click the **Page 4** tab.
5. View the **Delivery Days by Order Type and Continent** bar chart.



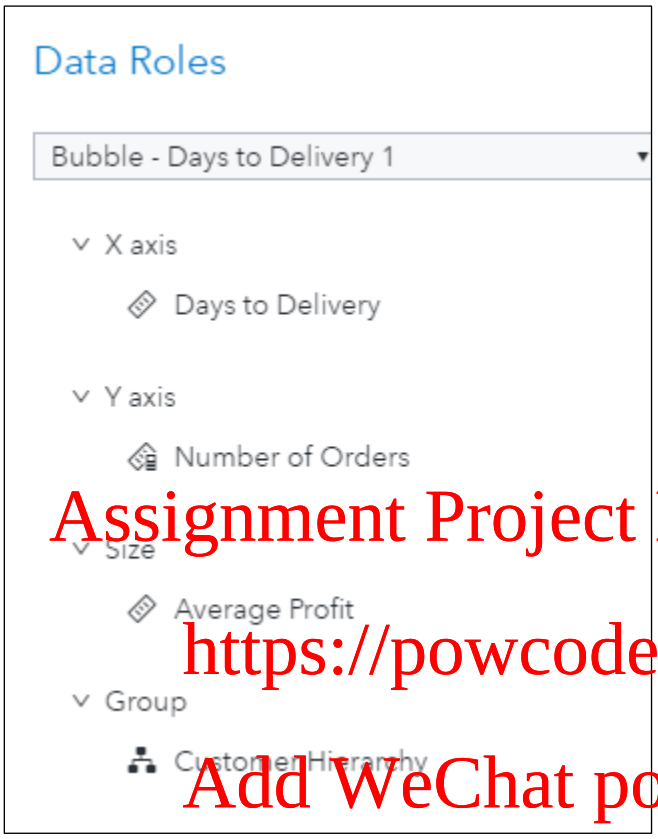
   In general, catalog sales take slightly less time to be delivered than internet sales. We might need to look at our internet process to try to minimize the difference. For most continents, the average days to delivery are the same, except that Africa has lower delivery times than other continents. This could be because there are no retail stores in Africa, but that does not explain why Asia has higher delivery times. We might need to look at our distribution facilities in Africa and Asia to determine the discrepancy.
6. Create a bubble plot.
   a. In the left pane, click **Objects**.
   b. Drag the **Bubble plot** object, from the Graphs group, to the right side of the canvas.
   c. In the right pane, click **Roles**.
   d. For the **Group** role, select **Add** ⇨ **Customer Hierarchy**.

e. For the **X axis** role, select **Add** ⇨ **Days to Delivery**.

f. For the **Y axis** role, select **Add** ⇨ **Number of Orders**.

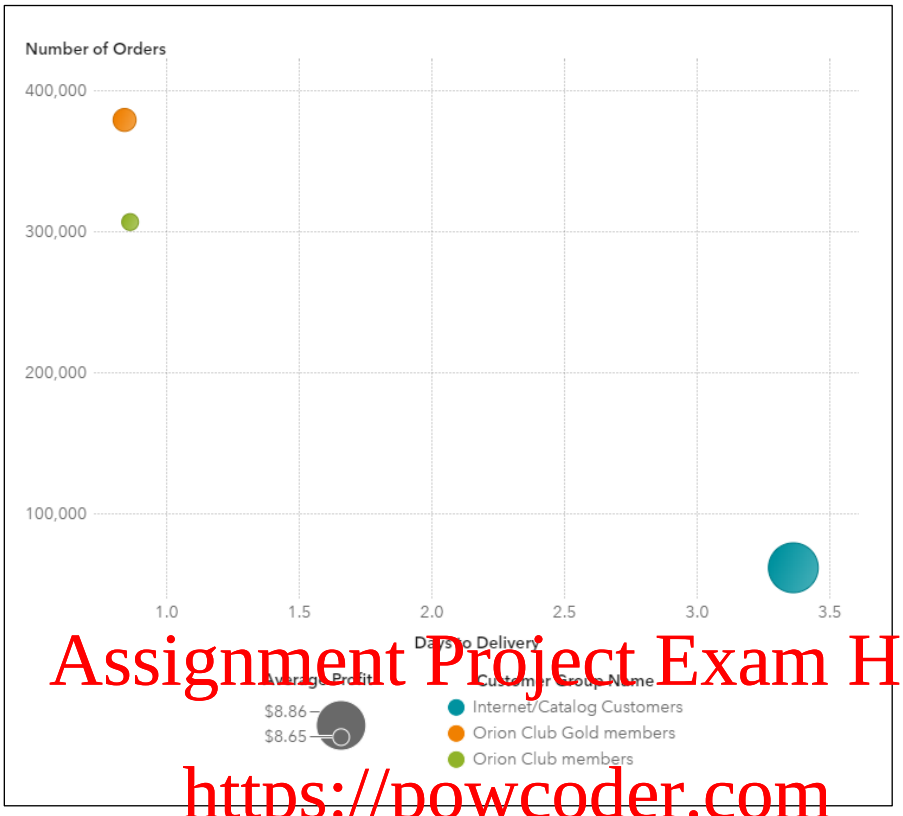
g. For the **Size** role, select **Add** ⇨ **Average Profit**.

The Roles pane should resemble the following:



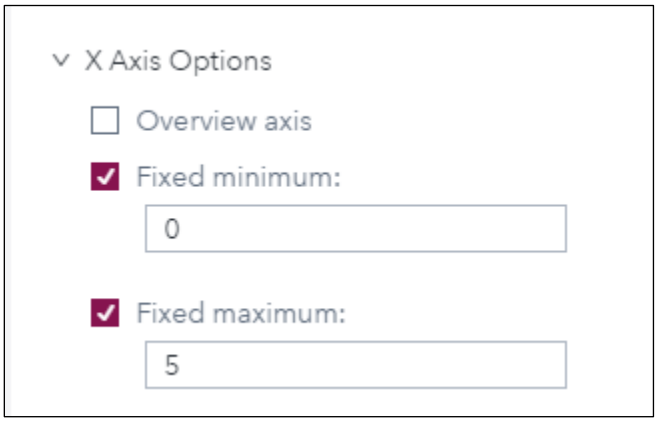
Measure data items can be added to the X axis and Y axis roles to determine the placement of the bubble. A measure data item can be added to the Size role to determine the size of the bubble.

The bubble plot should resemble the following:



h. In the right pane, click **Options**.

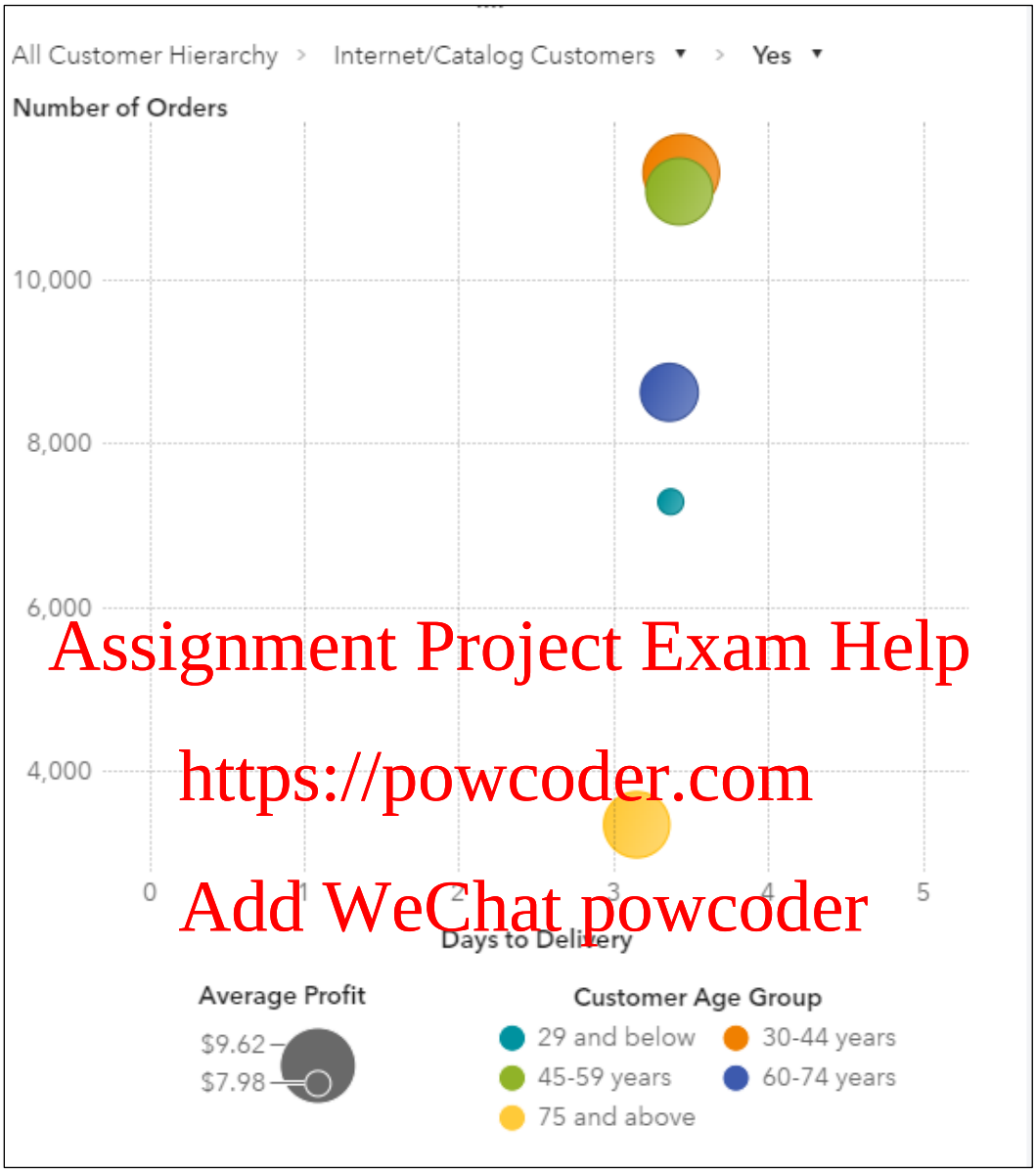
1) In the **X Axis Options** group, select **Fixed minimum**.
2) In the **Fixed minimum** field, enter **0**.
3) Select **Fixed maximum**.
4) In the **Fixed maximum** field, enter **5**.



i. In the bubble plot, double-click the **Internet/Catalog Customers** bubble.

j. Double-click the **Yes** bubble.

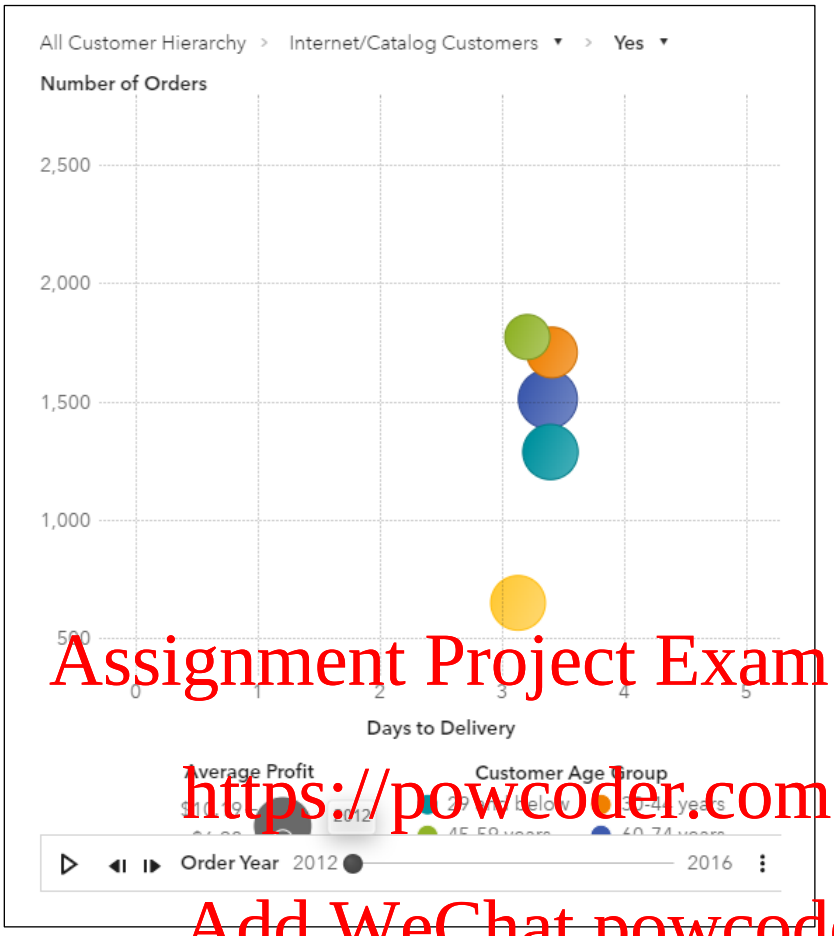
The bubble plot should resemble the following:



Next, we can analyze customers to determine which groups our marketing strategy can focus on. For internet/catalog orders among female customers, it seems the older age groups (60-74 years and 75 and above) place the fewest orders, but the oldest age group (75 and above) has the highest average profit. We should create marketing materials specifically for these groups to try to increase the number of orders.

7. Animate the bubble plot.
   a. In the right pane, click **Roles**.
   b. For the **Animation** role, select **Add** ⇨ **Order Year**.

The updated bubble plot should resemble the following:

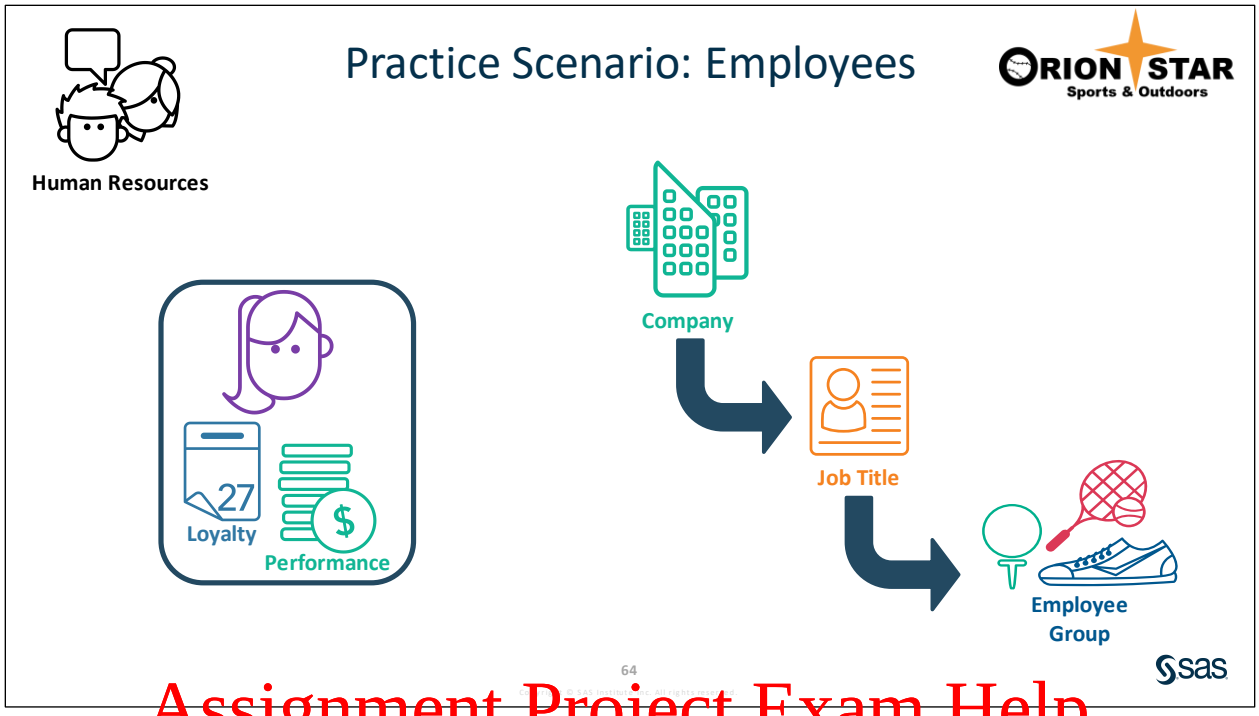


c. In the lower left corner, click ▷ to play the animation.

   For loyalty customers who have placed internet/catalog orders, the days to delivery remain nearly constant over the years. However, the number of orders has a marked increase in 2014 for customers in the 30–44 age group and a slight drop in 2015, and then seems to remain constant. For the older age groups (60–74 years and 75 and above), the number of orders remains fairly constant, but average profit decreases over time.

8. Save the report.

End of Demonstration



Practice

6. Analyzing Data

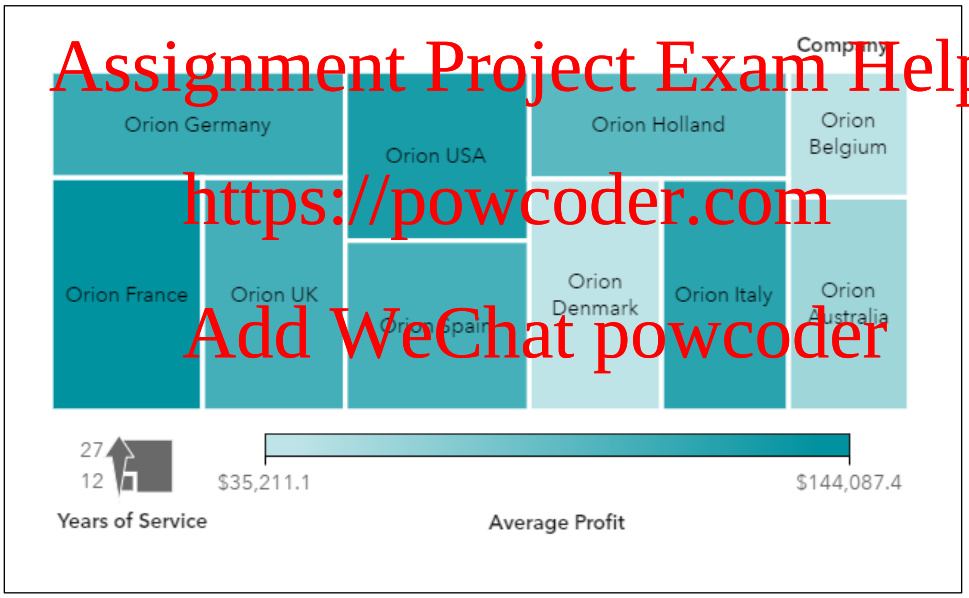
a. Open the browser and sign in to SAS Viya.

b. Open the **VA1- Practice3.4a** report from the **Courses/YVA185/Basics/Practices (HR)** folder.

c. On Page 5, create a treemap by assigning the following data items to the specified roles:

Tile	**Company**
Size	**Years of Service**
Color	**Average Profit**
Data tip values	**Add Number of Employees**

The treemap should resemble the following:



d. Create a new hierarchy (**Employee Hierarchy**) that contains the following categories:

Company

Job Title

Group

e. In the treemap, specify **Employee Hierarchy** for the **Tile** role and navigate through the hierarchy to answer the following questions:

Which two companies have the highest average profit generated (one possible criterion for promotion)?

Answer: _____

For these two companies, which job titles have the highest average years of service and average profit generated?

Answer: _____

f. Save the report.

End of Practices

Business Scenario: Customers

Marketing

Discounts

Delivery times

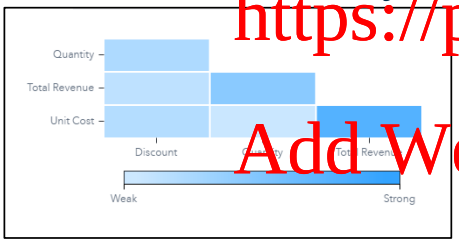
Unit costs

Total revenue

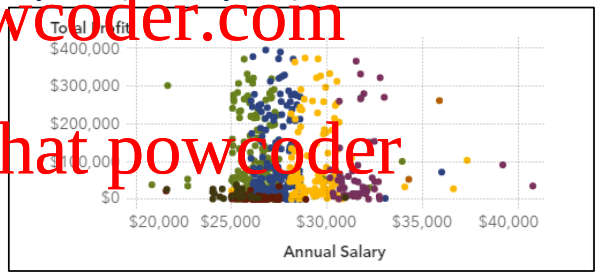
Predict profit and number of orders

68

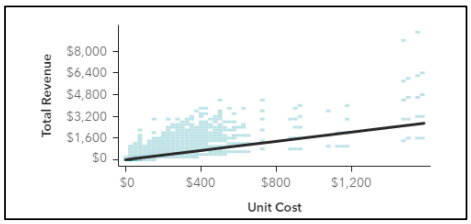
Objects: Graphs (Analysis)



Use a *correlation matrix* to evaluate the linear relationship between measures.



Use a *scatter plot* to evaluate the relationship between two measures.



Use a *heat map* to evaluate the relationship between two high-cardinality measures, between two categories, or between a category and a measure.

69

Scatter plot	A scatter plot displays the values of two measures using markers. When more than two measures are added to the graph, a scatter plot matrix is displayed, which shows a series of scatter plots for every possible pairing of the measures applied to the graph. Scatter plots can be used to visualize trends between measures and to pinpoint any possible outliers. **Note:** Scatter plots do not use aggregated data. Because of this, you get an error message if you attempt to create a scatter plot using more than 40,000 rows of data. For more information about data limits, see "High-Cardinality Thresholds for Objects" in the *SAS Visual Analytics: Reference* documentation.
Heat map	A heat map displays the distribution of values for two data items by using a table with colored cells. When more than two data items are added to the graph, a heat map matrix is displayed, which shows a series of heat maps for every possible pairing of the data items applied the graph. Heat maps can be used to visualize trends between high-cardinality measures and to pinpoint any possible outliers. If multiple measures are added to a heat map, the relationship between the measures can be visualized by adding a fit line.
Correlation matrix	A correlation matrix displays the degree of correlation between multiple measures as a matrix of rectangular cells, where each cell represents the intersection of two measures and the color of the cell indicates the degree of correlation between those two measures. The correlation values are calculated by using Pearson's correlation coefficient and are identified as weak (if the absolute value of the correlation is 0.3 or lower), moderate (if the absolute value of the correlation is greater than 0.3 and less than or equal to 0.6), or strong (if the absolute value of the correlation is greater than 0.6). Positive correlation values indicate that as one measure increases, the other measure increases as well, whereas negative correlation values indicate that as one measure increases, the other measure decreases.

3.07 Activity

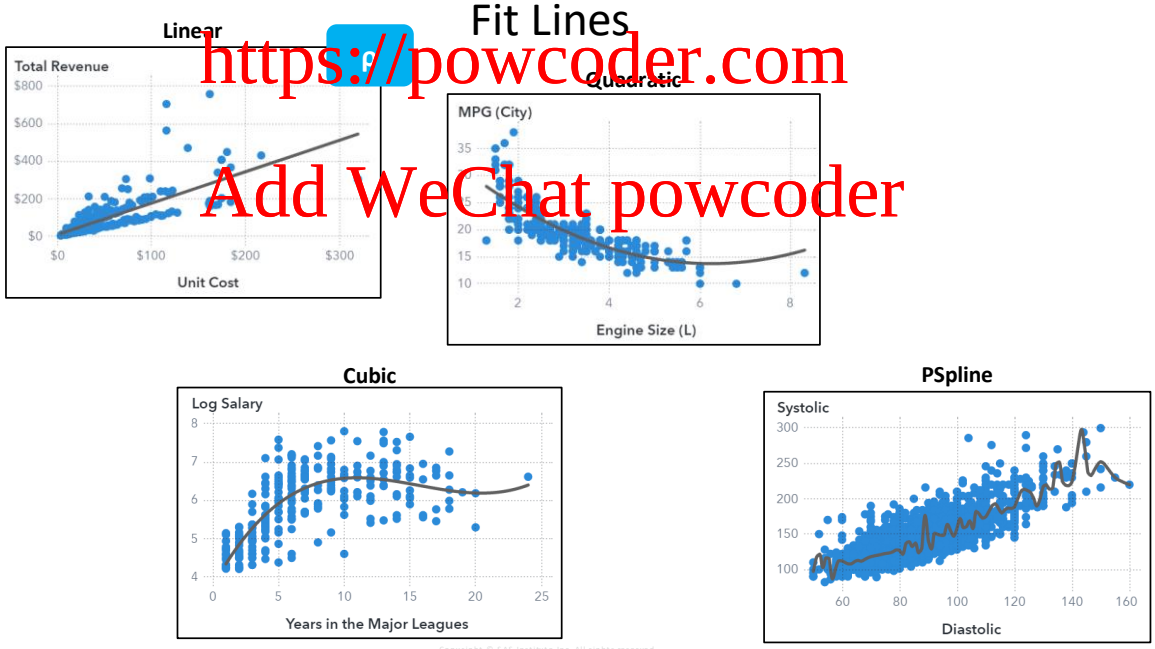
Each report object has a threshold for how much data it can visually display. Many report objects will not display high-cardinality data items with a large number of unique values.

What are some examples of high-cardinality data items?

What are some examples of low-cardinality data items?

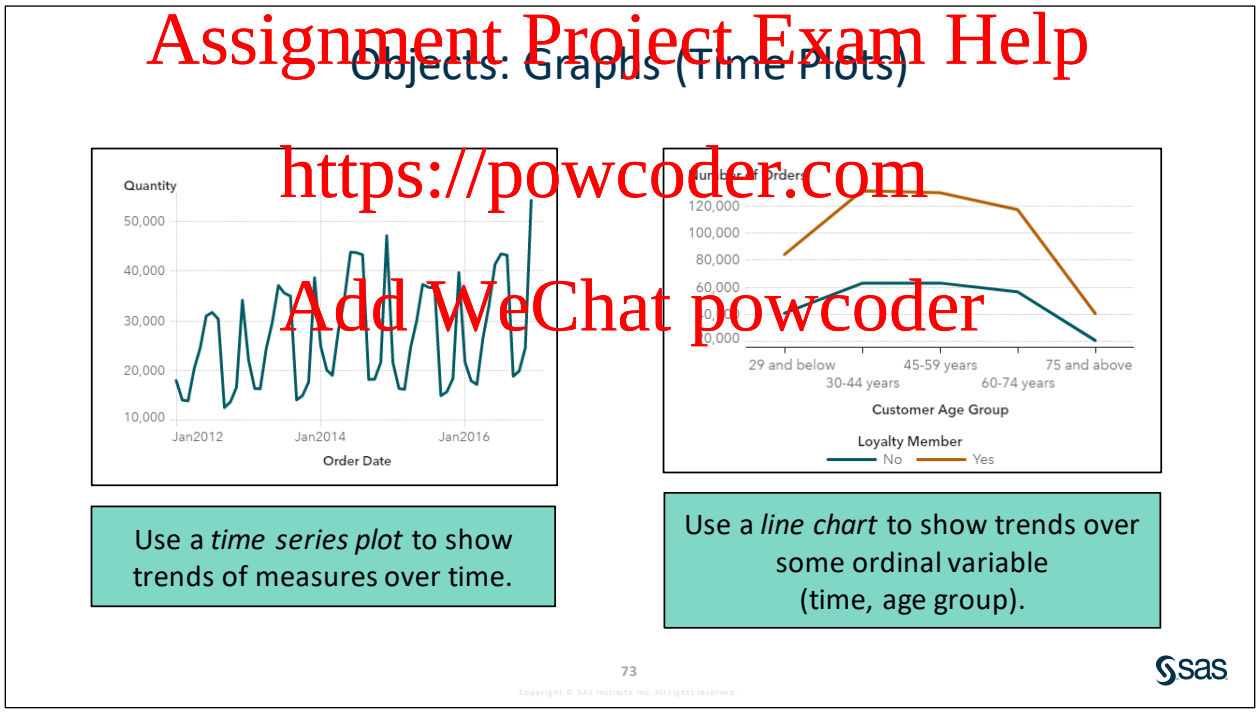
70

Fit Lines



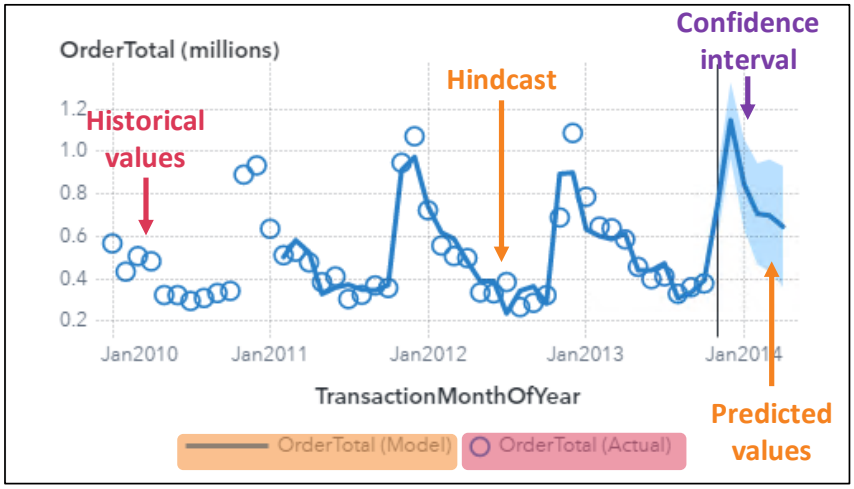
Fit lines can be added to scatter plots and heat maps to plot the relationship between variables. The following types of fit lines are available:

Best Fit	Selects the most appropriate model (linear, quadratic, or cubic) for your data. This method uses backward variable selection to select the highest-order model that is significant.
Linear	Creates a linear fit line (a straight line that best represents the relationship between measures) using a linear regression algorithm. For this method, correlation information is automatically added to the plot.
Quadratic	Creates a quadratic fit line (a line with a single curve that best represents the relationship between measures). This method produces a line with the shape of a parabola.
Cubic	Creates a cubic fit line (a line with two curves that best represents the relationship between measures). This method often produces a line with an S shape.
PSpline	Creates a penalized B-spline, which is a smoothing spline that closely fits the data. This method can display a complex line with many changes in its curvature.



Line chart	A line chart displays data by using a line that connects data values across some interval, such as time or a series of ordinal ranges. Stacked line charts enable you to compare totals for each category, as well as totals for all categories. However, comparing segments is difficult, and it is often hard to tell the difference between segments. To create a stacked line chart, select **Stack Filled** for the **Grouping style** option.
Time series plot	A time series plot displays data over time by using a line that connects the data values.

Objects: Analytics (Forecasting)



Use a *forecasting* object to show estimates of future values based on historical trends in the data.

74

Forecasting

A forecasting object uses the statistical trends in your data to predict future values. The forecast displays a line with predicted values and a colored band that represents the confidence interval. By default, the next six periods are forecast and the 95% confidence interval is displayed. Historical values for the forecasting model are displayed as markers only (without a line). Historical predicted values (hindcast) are displayed as part of the forecast line. SAS Visual Analytics automatically tests the following forecasting models against your data and selects the best model:

- ARIMA
- Damped-trend exponential smoothing
- Linear exponential smoothing
- Seasonal exponential smoothing
- Simple exponential smoothing
- Winters method (additive)
- Winters method (multiplicative)

Note: Forecasting accounts for cyclical patterns by using standard intervals of time (for example, 60 minutes in an hour, 24 hours in a day, and so on). If your data uses nonstandard values (for example, 48 thirty-minute cycles per day), then cyclical patterns are not considered in the forecast.

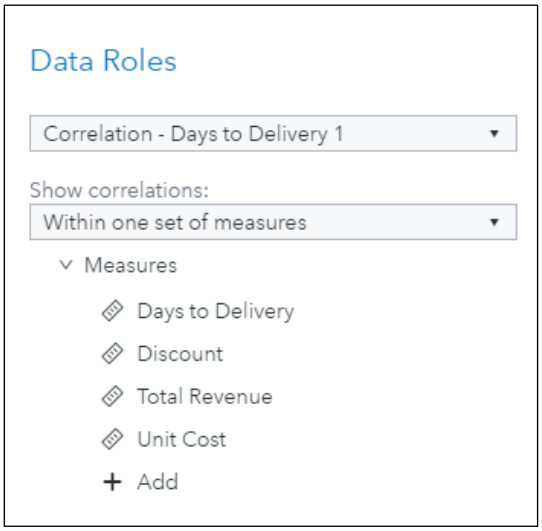
Note: If SAS Visual Statistics and SAS Visual Data Mining and Machine Learning are licensed at your site, you can create models instead of relying on the model automatically selected for forecasting.

Note: If SAS Visual Forecasting is licensed at your site, you can automatically produce large-scale time series analyses and hierarchical forecasts.

Adding Data Analysis

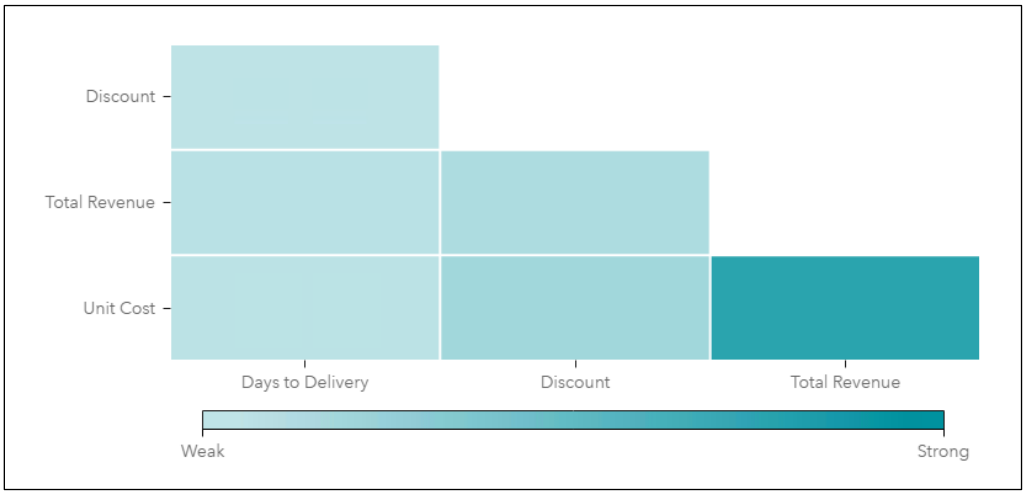
This demonstration illustrates how to add data analysis to graphs in Visual Analytics.

1. From the browser window, sign in to SAS Viya.
2. In the upper left corner, click ☰ (**Show list of applications**) and select **Explore and Visualize**. SAS Visual Analytics appears.
3. Click **All Reports**.
   a. Navigate to the **Courses/YVA185/Basics/Demos (Marketing)** folder.
   b. Double-click the **VA1- Demo3.4b** report to open it.
4. In the upper left corner of the report, click the **Page 5** tab.
5. Create a correlation matrix.
   a. In the left pane, click **Objects**.
   b. Drag the **Correlation matrix** object, from the Graphs group, to the top of the canvas.
   c. In the right pane, click **Roles**.
   d. For the **Measures** role, click **Add**.
   e. In the Add Data Items window, select the following measures:
      Days to Delivery
      Discount
      Total Revenue
      Unit Cost
   f. Click **OK**.
      The Roles pane should resemble the following:



Only measure data items can be used for the correlation matrix.

The correlation matrix should resemble the following:



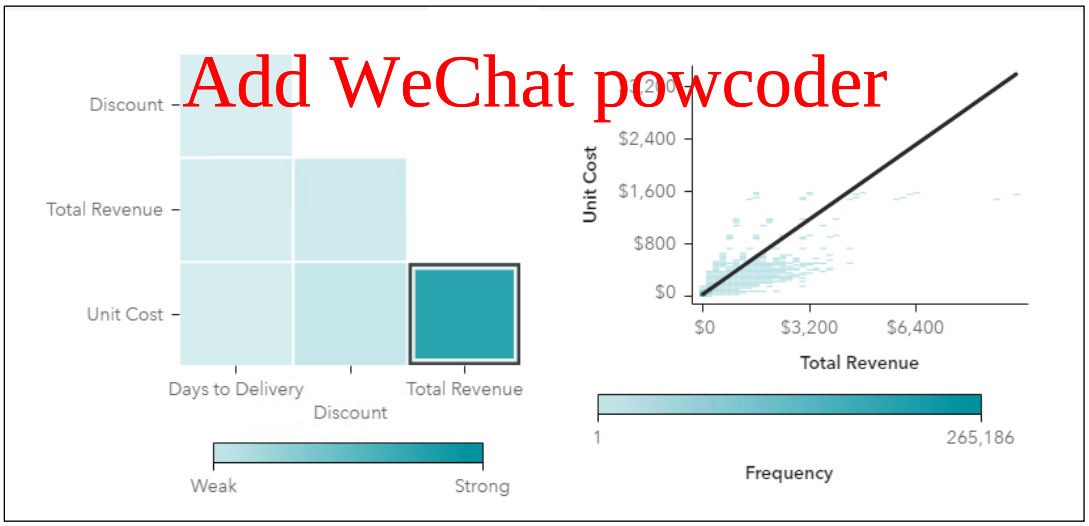
There is a strong relationship between **Unit Cost** and **Total Revenue**. Placing your cursor over the cell shows a correlation of 0.7790, meaning that as **Unit Cost** increases, so does **Total Revenue**. We should examine these two measures more closely to better understand the relationship.

g. Select the cell for **Unit Cost** and **Total Revenue**.

h. Right-click the correlation matrix and select **New object from selection** ⇨ **Heat map**.

A heat map is created below the correlation matrix.

i. Drag the heat map to the right of the correlation matrix.

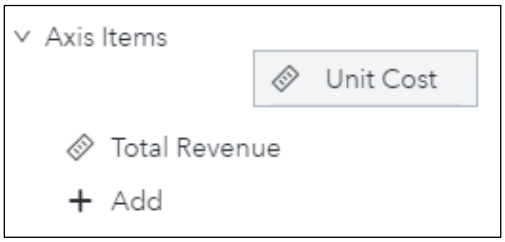


We want to move **Unit Cost** to the horizontal axis because we want to see how a unit change in cost affects **Total Revenue**.

j. If necessary, click the heat map to select it.

k. In the right pane, click **Roles**.

l. Select **Unit Cost** and drag it above **Total Revenue**.



m. In the upper right corner of the heat map, click (**Maximize**) to view additional details.

n. In the detail table below the chart, click **Unit Cost, Total Revenue analysis**.

Unit Cost, Total Revenue | Unit Cost, Total Revenue analysis

Property	Value
Model type	Linear
Model description	The linear fit is the straight line that best represents the relationship betwe...
R-square value	0.6068
Correlation	A correlation of 0.78 suggests there is a strong linear relationship between...
Correlation help	A positive correlation value means that as one variable increases the seco...
Slope	1.6966
Function	f(x)=8.0391 + 1.6966x
Average x	77.76
Average y	139.96
Standard deviation x	85.2765
Standard deviation y	185.7319
Observations	951,65

The linear fit line between unit cost and total revenue indicates that a dollar increase in costs increases revenues by $1.69.

o. In the upper right corner, click (**Restore**).

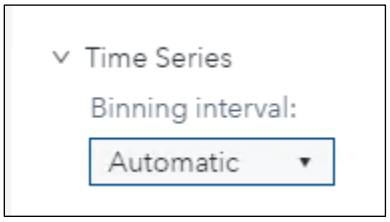
6. Modify the time series plot.

a. Click the time series plot to select it.

b. Right-click the time series plot and select **Change Time series plot to** ⇨ **Forecasting**.

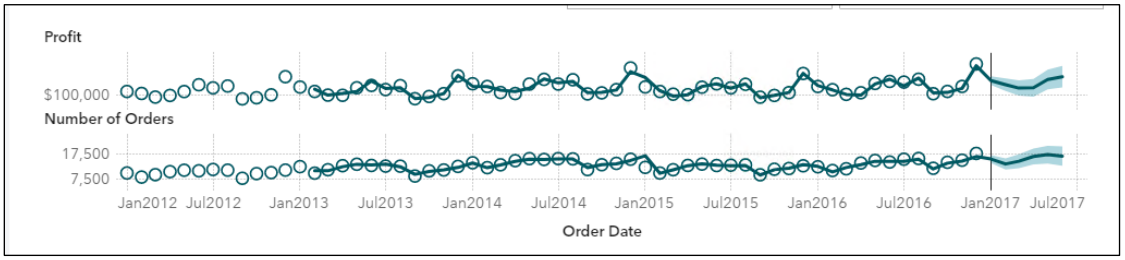
c. In the right pane, click **Options**.

d. In the Time Series group, for the **Binning interval** field, select **Automatic**.



e. Increase the size of the Forecast object to see the forecasted values.

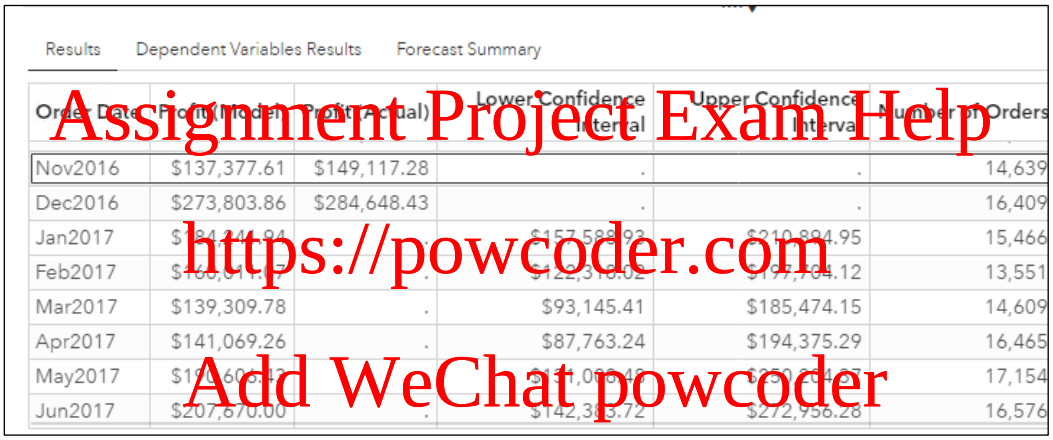
The forecast plot should resemble the following:



We can see that **Profit** and **Number of Orders** are closely related. When the number of orders rise, so do profits. The forecast shows that this trend is expected to continue in the near future.

f. In the upper right corner of the forecast plot, click (**Maximize**) to view additional details.

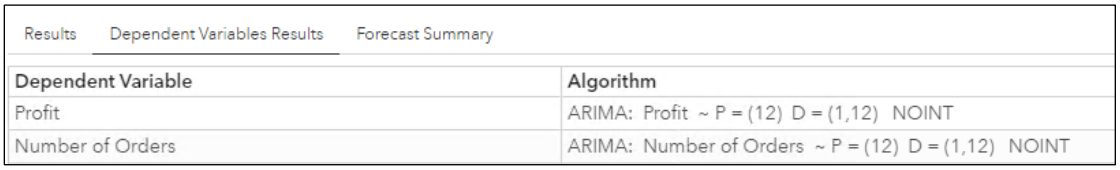
g. Scroll to the bottom of details table below the chart.


Results | Dependent Variables Results | Forecast Summary

Order Date	Profit (Model)	Profit (Actual)	Lower Confidence Interval	Upper Confidence Interval	Number of Orders
Nov2016	$137,377.61	$149,117.28	.	.	14,639
Dec2016	$273,803.86	$284,648.43	.	.	16,409
Jan2017	$184,244.94		$157,588.93	$210,894.95	15,466
Feb2017	$166,011.07	.	$122,310.02	$197,764.12	13,551
Mar2017	$139,309.78	.	$93,145.41	$185,474.15	14,609
Apr2017	$141,069.26	.	$87,763.24	$194,375.29	16,465
May2017	$196,602.32	.	$141,089.48	$252,115.17	17,154
Jun2017	$207,670.00	.	$142,383.72	$272,956.28	16,576

The forecasted values for profit and number of orders, along with values for the lower and upper confidence intervals, are displayed.

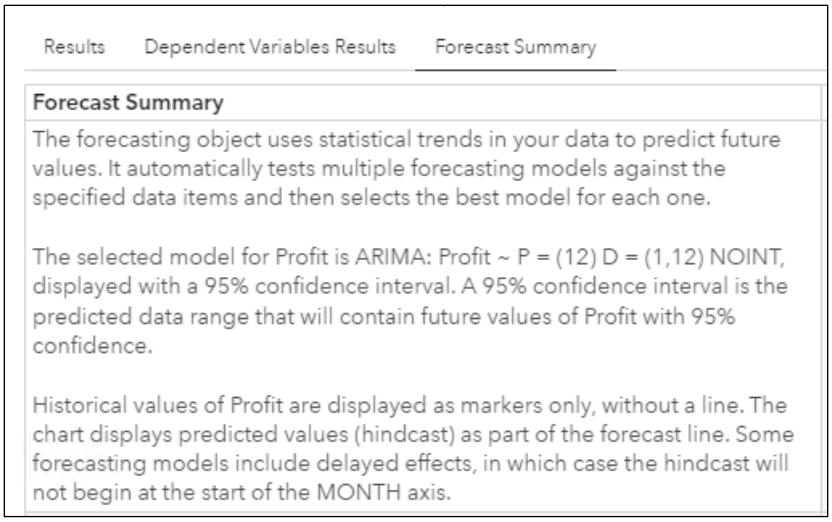
h. Click **Dependent Variables Results** in the table of data values below the chart.


Results | Dependent Variables Results | Forecast Summary

Dependent Variable	Algorithm
Profit	ARIMA: Profit ~ P = (12) D = (1,12) NOINT
Number of Orders	ARIMA: Number of Orders ~ P = (12) D = (1,12) NOINT

Visual Analytics has determined that the ARIMA algorithm best forecasts profit and number of orders. This algorithm cannot be changed.

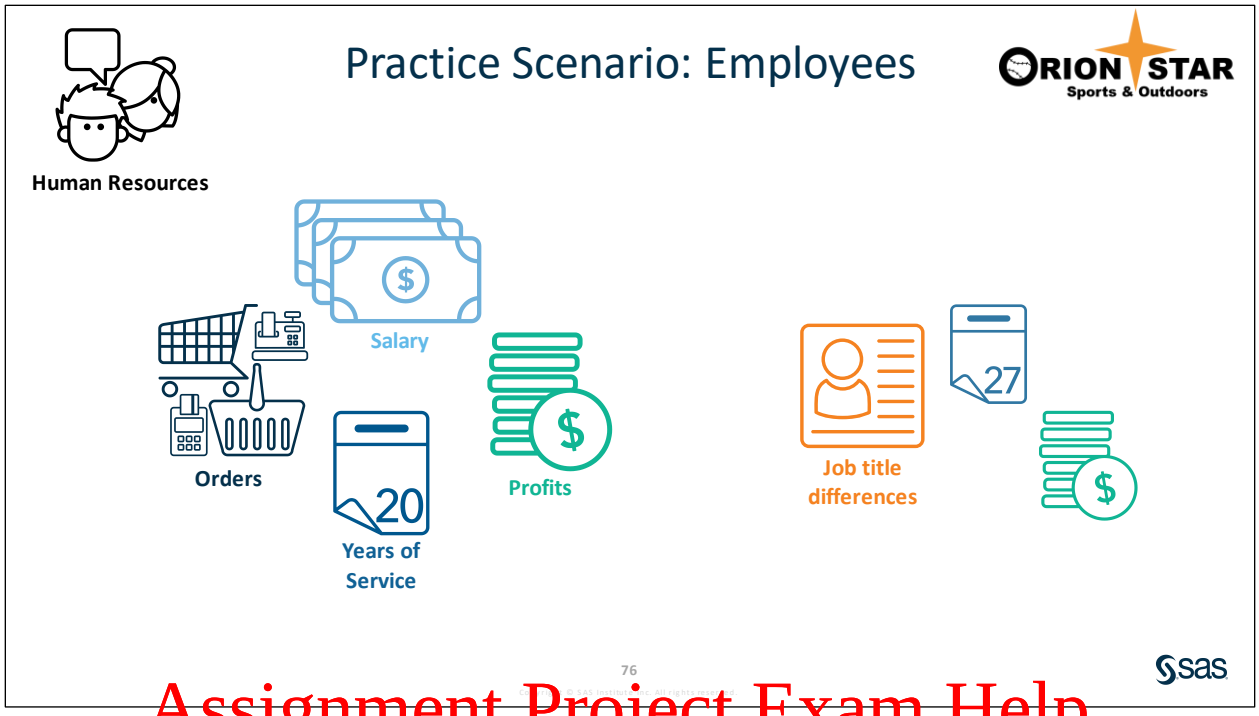
i. Click **Forecast Summary** in the table of data values below the chart to view the Forecast Summary.



j. In the upper left corner, click (**Restore**).

7. Save the report.

End of Demonstration



Practice

7. Adding Data Analysis

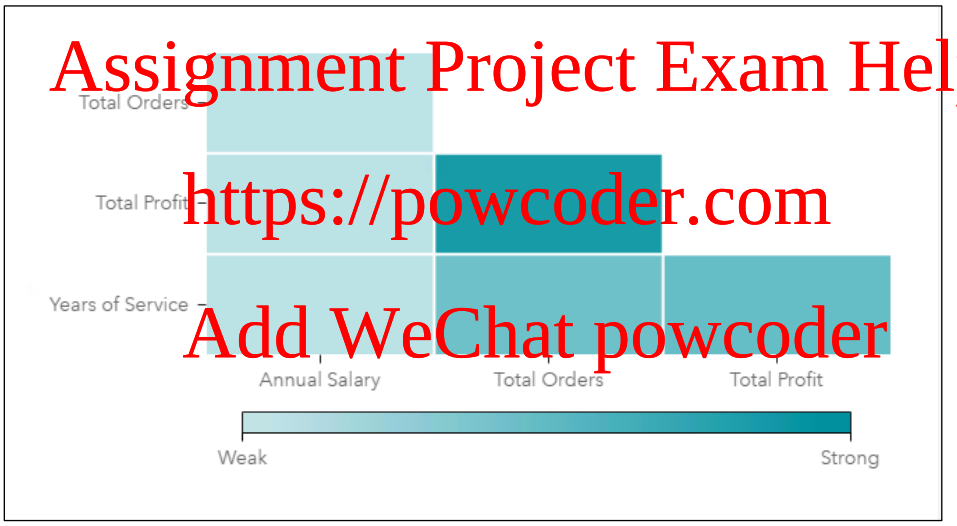
a. Open the browser and sign in to SAS Viya.

b. Open the **VA1- Practice3.4b** report from the **Courses/YVA185/Basics/Practices (HR)** folder.

c. On Page 6, create a correlation matrix by assigning the following data items to the specified roles:

Measures	**Annual Salary**
	Total Orders
	Total Profit
	Years of Service

The correlation matrix should resemble the following:



d. Answer the following question:

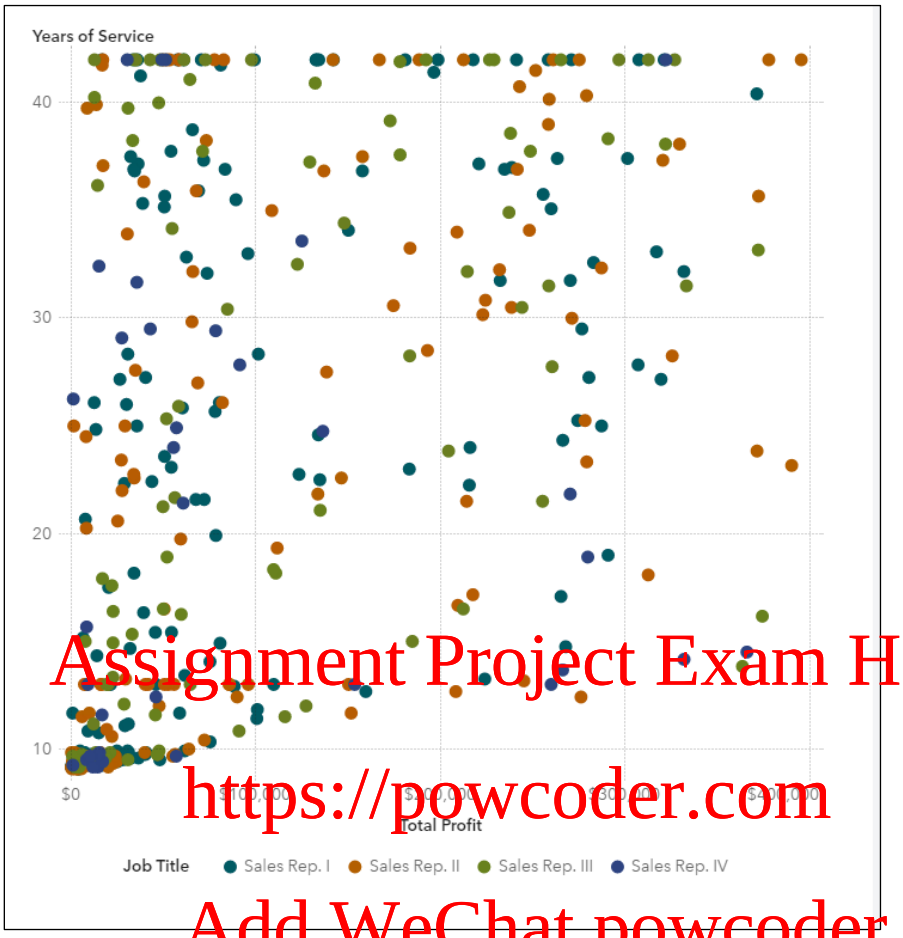
What is the degree of correlation between **Total Orders** and **Total Profit**?

Answer: _____

e. Create a scatter plot, on the right of the correlation matrix, by assigning the following data items to the specified roles:

Measures	**Total Profit**
	Years of Service
Color	**Job Title**

The scatter plot should resemble the following:



f. Answer the following question:

Using years of service and profit generated as promotion criteria, do you notice any differences between job titles?

Answer: _____

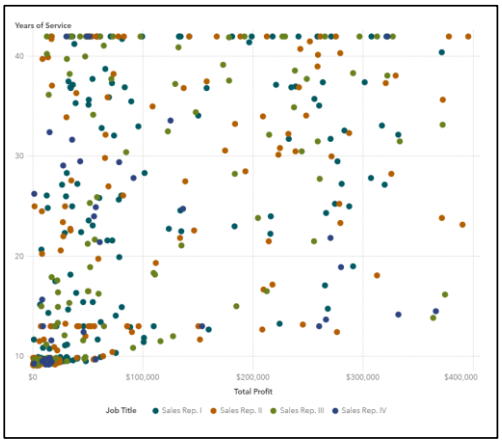
g. Save the report.

End of Practices

3.7 Adding Data Analysis – Solution

Using years of service and profit generated as promotion criteria, do you notice any differences between job titles?

We want to focus on employees in the upper right quadrant of the scatter plot. In that area, there seems to be an equal representation of Sales Rep. I, Sales Rep. II, and Sales Rep. III.



79