**ASSIGNMENT**

**SAP ID:** 60004190013 **Name:** Aryan Parekh

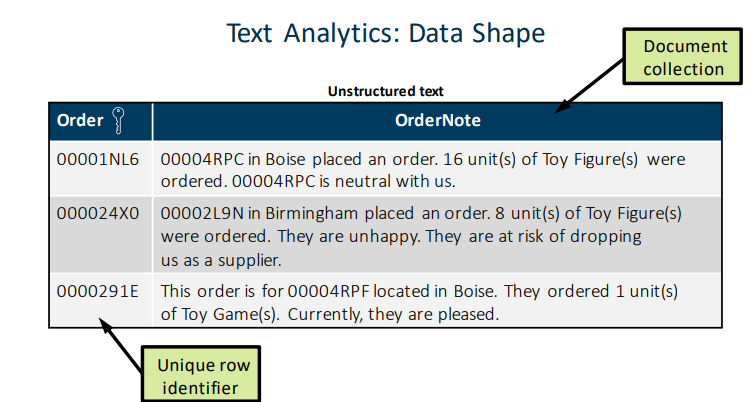
**Division:** A **Batch:** A1

**Subject:** Business Analytics **Semester:** VI

**Date of Submission:** 16th May, 2022

Q1). Discuss how a Text data source is analyzed in SAS.

First, Text is converted into a particular data shape for processing. It needs to be converted into a tabular format with 2 columns. The first column is called as a **Unique Row Identifier**, and the second column is called as **Document Collection**.



If your data source does not contain a unique row identifier, you can create one in Visual Data Builder by adding a generated key column to your query. After structuring the data, we pass it through various steps:

STEP 1: PARSING INTO INDIVIDUAL TERMS

Text from each row is taken and parsed into individual terms from a sentence or a paragraph.

STEP 2: COMPARISON WITH STOP WORD LIST (If available)

Next, the list obtained from step 1, containing list of words is compared with a stop list. A stop list is a table of common words that are ignored in text analytics.

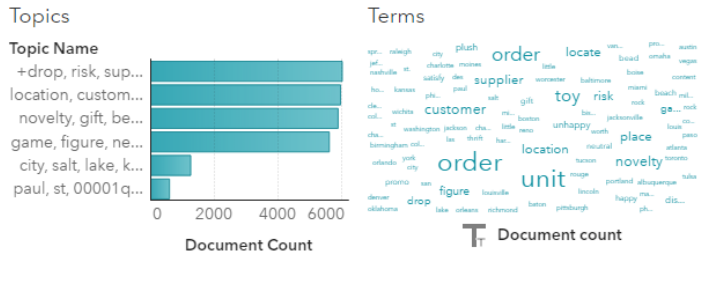
STEP 3: REMOVAL OF STOP WORDS FROM MAIN LIST:

If words from the main list are found in stop list, then they are removed at this stage and not considered for further analysis.

STEP 4: ANALYSIS

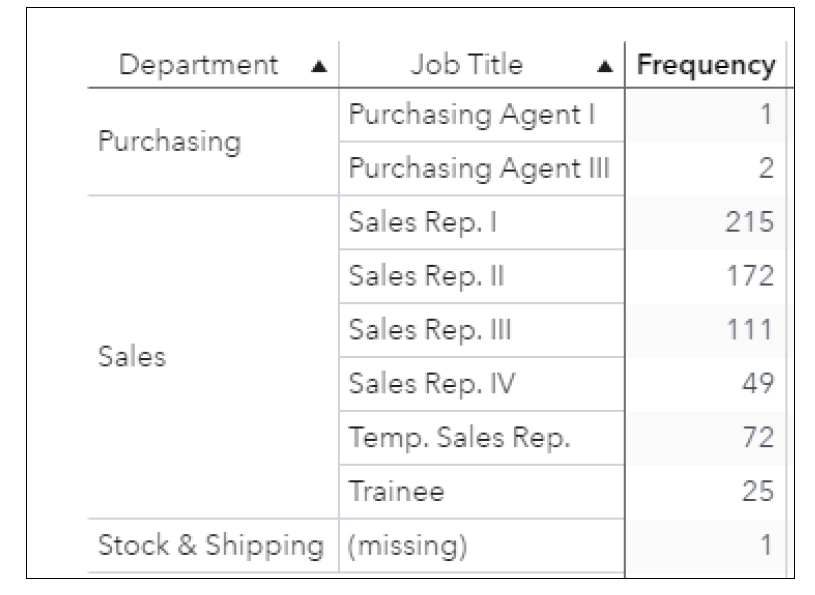
Now, the remaining words are taken so that analysis can be performed on it to determine topics.

At the end, the output of text analysis for document count looks like:



Q2).

1. View the crosstab of **Department** and **Job Title** and answer the question: Which **Department** contains the **missing** **Job Title**?



**Stock & Shipping** Department contains the missing Job Title.

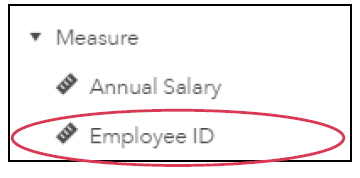
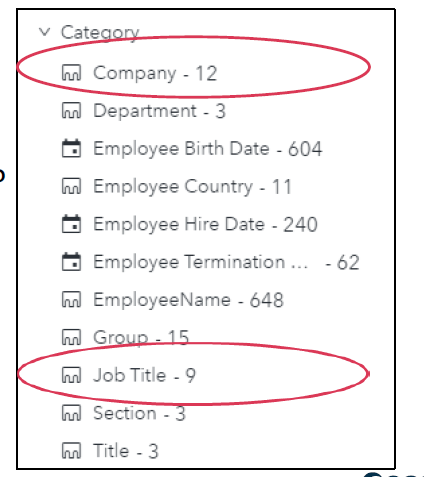
1. How will you filter the table to include only employees in the **Purchasing** and **Sales** **Departments**?

1) In the left pane, click the Data icon.

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

1. Click **Operators** on the left.
2. Expand **Boolean**.
3. Double-click **OR** to add it to the expression.
4. On the left, click **Data Items**.
5. In the Character group, select Department.
6. In the Conditions area, double-click **Department = ‘x’** to add it to the first condition in the expression area.
7. Enter **Purchasing** as the string for the first condition.
8. In the Character group, select Department.
9. In the Conditions area, double-click **Department = ‘x’** to add it to the second condition in the expression area.
10. Enter **Sales** as the string for the second condition.

Q3). Answer the following for the given data:



1. How many unique values does **Company** have? **Job Title**?

**Company** has 12 unique values. **Job Title** has 9 unique values.

1. What is the type (or classification) of **Employee ID**?

Classification of **Employee ID**: Measure

Q4). Justify which graph would help you determine whether a measure is normally distributed?

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.

The shape of data distribution in histogram can give us an approximate idea about skewness of data.

1. Normal Distribution: If histogram appears somewhat like this, then we can say that data may be normally distributed.



1. Skewed Distribution:
   1. Right Skewed Distribution: If histogram appears somewhat like this, then we can say that data may be right skewed.



* 1. Left Skewed Distribution: If histogram appears somewhat like this, then we can say that data may be left skewed.



Q5). Explain the difference between calculated item and aggregated measure with relevant examples.

|  |  |
| --- | --- |
| **Calculated Item** | **Aggregated Measure** |
| 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. | 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. |

Q6).

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



IF Employee Termination Date = “.”:

Years of Service = (Today – Employee Hire Date)/365.25

ELSE:

Years of Service = (Employee Termination Date – Employee Hire Date)/365.25

1. Justify which data item type will the expression below return



The given expression returns string “Active” or “Passive” depending on value of Employee Termination Date. Since the value returned in either case is string, the given data will be classified and put into **Category** group.

Q7). Explain the following with relevant examples.

1. Custom Category

A data item that can be created based on either a category or numeric data item. A custom category is always a category data item with alphanumeric values. 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 ranges or distinct values to group the data. E.g.: We have 5 regions: Africa, Asia, Europe, Oceania and North America. We can create two custom categories called Northern and Southern such that:



1. Derived items

Derived data items are aggregated measures that display values for the measure and the formula type on which the derived item is based.

Examples of derived items can be created from category data items:

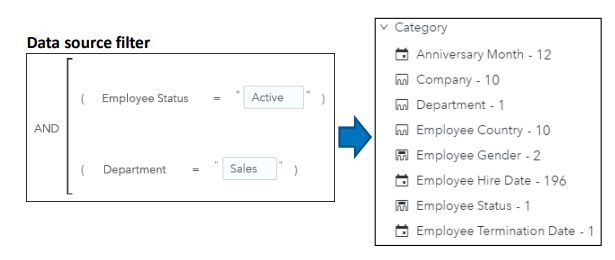
* Distinct Count
* Count
* Number missing

Examples of derived items can be created from measure data items:

* Difference from previous period
* Percent difference from previous period
* Year to Date

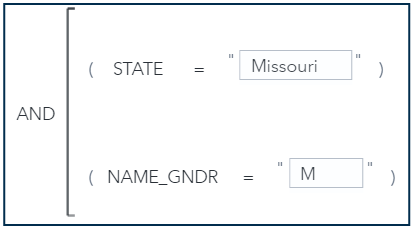
1. Data source filter

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 Report Builder. This can be seen by the updated cardinality values in the Data pane after the filter has been applied. E.g.:



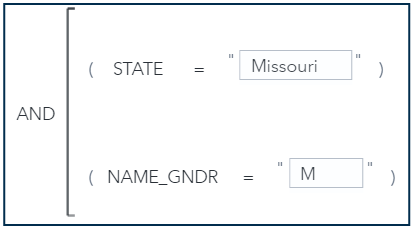
1. Basic report filter

Subsets the data for individual report objects by using a single data item and an equality condition. E.g.:



1. Advanced report filter

Subsets the data for individual report objects by using any number of data items and operators in the same expression. E.g.:

-

1. 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.

Q8). Which object can use a data item that has a classification type of geography? List the essential steps for creating a geo-map.

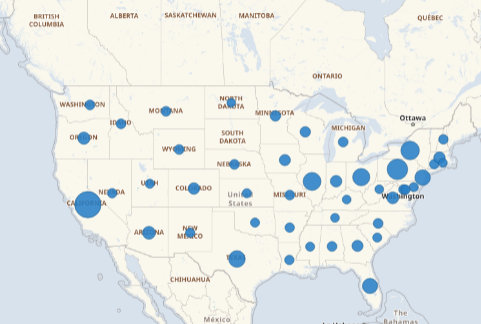
Geo map is the object which can be used for data item with classification type of geography up to two measures. If there are three or more measures long with one geography, then a Bar Chart is recommended. Even crosstabs and tables can use classification type of geography. But, out of all these, Geo map is the only one that won’t give any output without a geography type classification.

Steps to create a geo-map:

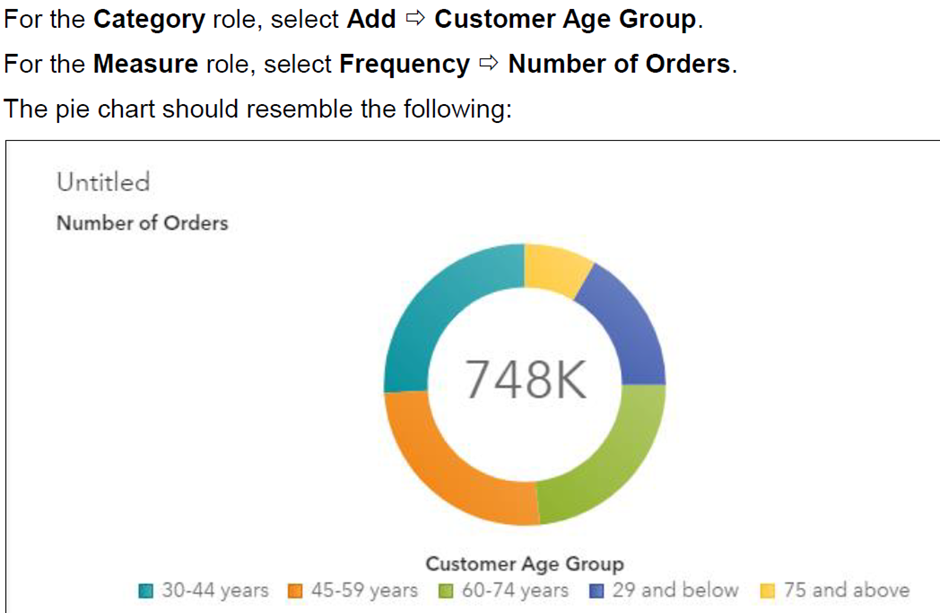
1. In the left pane, click the Data icon.
2. Click (Edit properties) next to State Name.
3. Select Geography for the Classification field.
4. Verify that Predefined geographic names and codes is selected for the Geography data type field.
5. Select US State Names for the Geography field.
6. Click OK.

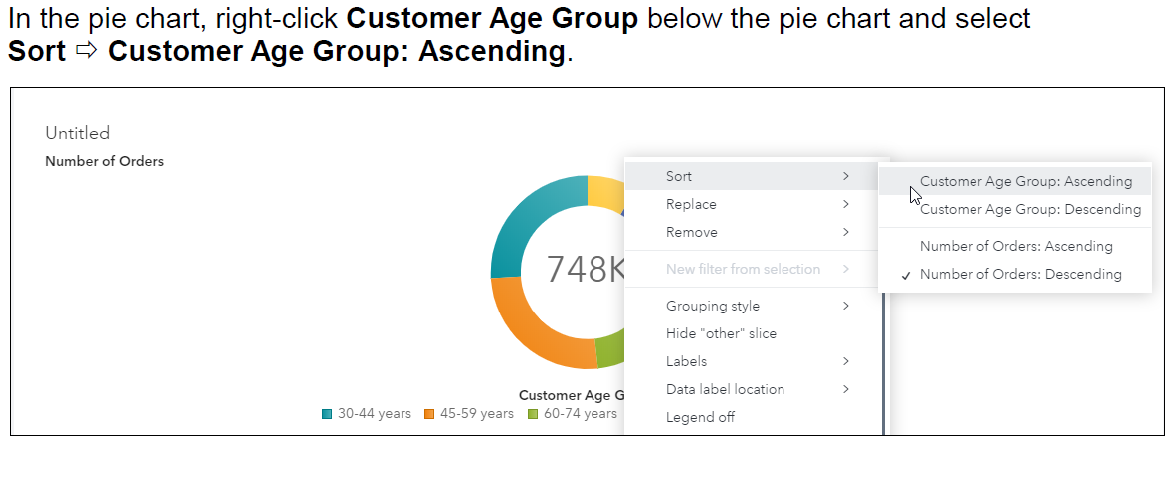
A new group, Geography, is added to the Data pane.

1. In the left pane, click the Objects icon.
2. Drag the Geo Map object, from the Graphs group, to the canvas.
3. For the Category role, select Add, then select State Name.
4. Verify that Frequency is specified for the Size role.
5. Geo Map is ready:

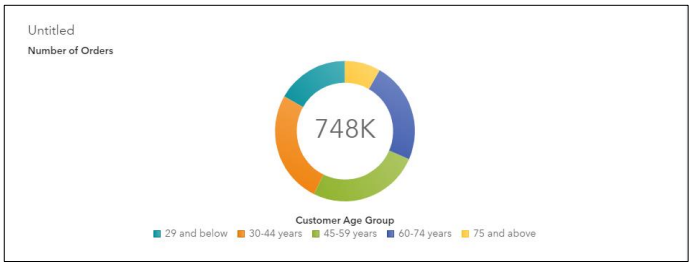


Q9).





How would the updated pie-chart look like?



Q10). Explain with relevant examples

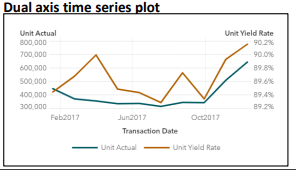
1. Dual axis bar chart

A dual axis bar chart displays two bar charts with a shared category axis and separate response axes. Use a dual axis bar chart when the value for both measures does not depend on the prior value. E.g.:



1. Dual axis time series plot

A dual axis time series plot displays two time series with a common time axis on separate response axes.



1. Graph-Level Display Rules

Graph Level Display Rules are the rules that can be applied on a particular graph. They provide a flexible structure to specify conditions. They are mainly used to highlight the specifications that user wants to highlight. Display rules for graphs can be added using multiple ways:

* an expression
* color mapped values

1. Table-Level Display Rules

Table Level Display Rules are the rules that can be applied on a particular table. They provide a flexible structure to specify conditions. They are mainly used to highlight the specifications that user wants to highlight. Display rules for tables in a list table can be added using:

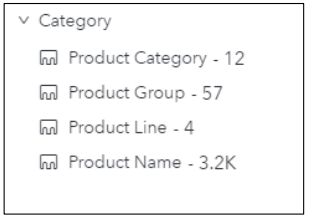
* an expression
* a gauge
* color mapped values

Table level display rules can also be added to a crosstab.

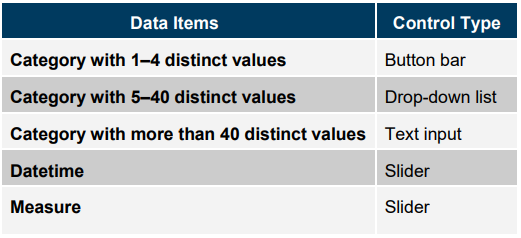
1. Report-Level Display Rules

Report Level Display Rules are the rules that are applicable to all the objects in a report. They provide a flexible structure to specify conditions. They are mainly used to highlight the specifications that user wants to highlight. Mainly, Report level display rules are added using color mapped values as they provide a uniform color to a specific object throughout the report.

Q11). Given the distinct values, which control object would you use to filter for each category displayed below? Justify your answer.



We know that:



Therefore, suggested control types for each category according to their distinct value count and data type:

|  |  |
| --- | --- |
| **Category** | **Control Type** |
| Product Category | Drop-down list |
| Product Group | Text input |
| Product Line | Button Bar |
| Product Name | Text Input |

Q12). What are the tips for designing reports? Justify what type of chart would you use to show profit information by continent?

Tips for Designing Reports:

* Understand the audience.
* Make accessible to all.
* Tell a single data story.
* Use visually appealing, easy to understand objects.
* Use the simplest graph.
* Use consistent fonts.
* Limit the number of objects.
* Limit the number of pages.

Profit information for a continent can be negative as well as positive. Since there is a chance of it being negative, we can’t use pie charts or tree maps as they can’t display negative values.

Bubble plots can’t be used either as we don’t have three different measures.

Bar Charts, on the other hand can be used to display all types of values, and will fit well for a 2D plot with continent on X-axis and Profit Information on Y-axis.

