



MTH782P - SAS for Business Intelligence
(Semester B) - 2023/24

Dr. Hong Qi

Analysing and Creating Business Intelligence Reports

SAS Report

Name: Mr. Vatsal Vipul Doshi

Student ID: 230504415

PART I (40 Points): Orion Star Product Report Analysis using SAS Viya

Q.1) Before starting your analysis, you should access and investigate the data set, including understanding its structure, data types, variables, dimensions, and possible data issues. Provide clear instructions on how you access and explore the data set and write a detailed description of the data set. **(12 points)**

Solution:

Throughout this analysis, we are using “**PRODUCTS_CLEAN**” data set from SAS Viya.

The following actions must be taken in order to open this data set:

- *Step 1:* Select your preferred browser and open SAS Viya.
- *Step 2:* Go to the SAS Content/Courses/YVA185/Basics route in SAS Viya's SAS Drive - Share and Collaborate page, browse the files there, and find and open the “**PRODUCT REPORT**” for Orion Star Sports & Outdoors.
- *Step 3:* Navigate to the applications list from the button in the upper-left corner, then select the Manage Data button to launch SAS Data Explorer.
- *Step 4:* From the list of available data sets, choose the “**PRODUCTS_CLEAN**” data set now.
- *Step 5:* To view the metadata for the data in the Report, select the Details tab.

Data set's description:

To begin our investigation, I initially searched the various tables or datasets to see which ones were relevant to our project objectives. Following that, I imported the selected datasets into CAS (Cloud Analytics Services) for further analysis. Upon first examination, it was found that the dataset contains around 951.7k records over 30 columns or features.

The structure of a dataset tells us how it's organized and what's inside. Each part of the dataset has its own special characteristics.



#	Name	Label	Type	Raw Length	Formatted Length	Format	Tags
1	City	City Name	char	90	90		
2	Continent	Continent Name	char	90	90		
3	CountryLabel	Country	char	60	60		
4	Postal_Code	Postal code	char	30	30		
5	State_Province	State Name	char	76	76		
6	Street_Name	Name of Street	char	136	136		
7	xyContinentLat	xyContinentLat	double	8	15	F	
8	xyContinentLon	xyContinentLon	double	8	15	F	
9	Order_Date	Date Order was placed by Cu...	double	8	9	DATE	
10	Delivery_Date	Date Order was Delivered	double	8	9	DATE	
11	Quantity	Quantity Ordered	double	8	12	COMMA	
12	RetailPrice	Retail Price	double	8	13	DOLLAR	
13	Cost	Cost	double	8	13	DOLLAR	
14	Discount	Discount in percent of Norma...	double	8	6	PERCENT	
15	OrderTypeLabel	Order Type	char	46	46		
16	Product_Category	Product Category	char	76	76		
17	Product_Group	Product Group	char	76	76		
18	Product_Line	Product Line	char	60	60		
19	Product_Name	Product Name	char	136	136		
20	SupplierContinent	Supplier Continent	char	120	120		
21	SupplierCountryLabel	Supplier Country	char	60	60		
22	Supplier_Name	Supplier Name	char	90	90		
23	xyContLat	xyContLat	double	8	12		

Variable: One important part is the variable, which is like a category that holds different kinds of information. For example, if we're talking about a dataset of customers, one variable might be "Customer Name," another might be "Order Date," and so on.

Label: Every variable has its own name, like a label. This name helps us know what the variable is about. For example, "Customer Name" tells us the variable is about the names of customers. Labels make it easier for us to understand what the data represents.

Datatype: Variables also have a type, which tells us what kind of information they hold. Some variables might hold numbers, others might hold words, and some might hold dates or times. This helps us know how to work with the data.

Length: When we look at a variable, we can also see its length. This tells us how much space it takes up in the dataset. There are two types of lengths we might see: "raw length," which is the maximum length of the data before any special formatting, and "formatted length," which is how long the data looks after it's been formatted nicely.

Format: Variables can also have a format, which tells us how the data is displayed or stored. For example, a date might be shown as "mm/dd/yyyy" or "dd-mm-yyyy," depending on the format chosen. This helps keep the data consistent and easy to understand.

Tags: Lastly, variables might have tags, which are like extra labels that give us more information about them. Tags help us organize and find specific variables when working with large datasets. They add extra context to the variables, making them even more useful for analysis.

So, the structure of a dataset includes all these things about its variables: names, labels, types, lengths, formats, and tags. Understanding these helps us make sense of the data and use it effectively for analysis and decision-making.

Dimensions: Dimensions in a dataset refer to the different categories or aspects of the data that provide context and meaning to the information contained within. They serve as the backbone for organizing and understanding the dataset's contents, allowing us to analyse and interpret the data effectively.

Based on the provided variables, we can categorize them into 6 different dimensions. Here's a breakdown:

1. Customer Information:

- Customer ID
- Customer Name
- Name of Street
- Postal code
- City Name
- State Name
- Country
- Quantity
- Retail Price
- Discount in percent
- Cost
- Profit
- Order Month
- Order Year

2. Order Information:

- Order ID
- Order Type
- Order Date
- Delivery Date

3. Employee Information:

- Employee ID

4. Product Information:

- Product ID
- Product Name
- Product Line

- Product Category
- Product Group

- Name of Street
- Postal code
- City Name
- State Name

5. Location Information:

- Continent Name (Customer Continent)
- Supplier Continent
- Supplier Country
- Supplier Name
- Supplier ID
- Supplier Country

6. Geospatial Information:

- xyContinentLat
- xyContinentLon
- xySupContLat
- xySupContLong

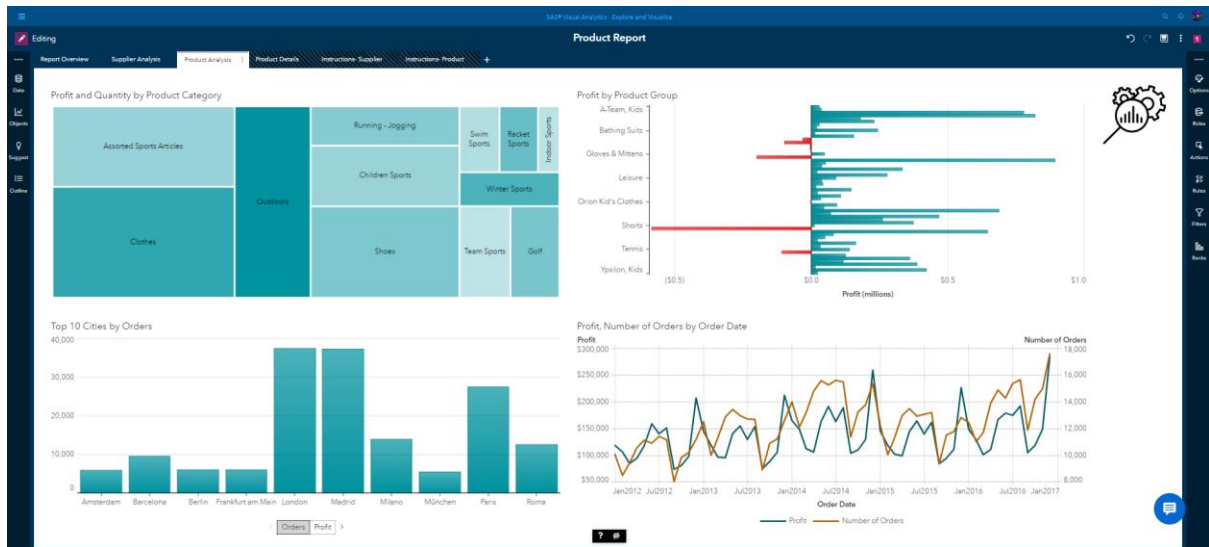
Each variable is categorized based on its relevance and relationship to other variables within its respective section. This categorization helps organize the variables into logical groups for analysis and understanding of the dataset.

Data issue:

PRODUCTS_CLEAN						
Report "03/31/24 01:14 PM"						
Column	Unique	Null ↓	Blank	Pattern Count	Mean	Media
⊕ Discount	0.04% (4)	99.04% (942,517)			0.38	0.4
△ State_Province	0.03% (271)	12.31% (117,192)		93		
△ Postal_Code	2.04% (19,339)	0.39% (3,716)		42		
△ City	1.10% (10,506)	0.01% (118)		731		
△ Continent	<0.01% (5)			4		
⊕ Cost	0.20% (1,883)				77.76	50.5
△ Counted_label	<0.01% (47)			15		

- Several variables in the dataset under examination have noticeable instances of null values, which could cause problems with the data when a report is created.
- More specifically, 0.01% (118) of the total records show a slight occurrence of null values for the **'City'** attribute.
- In comparison, the variable **'Discount'** displays a notable occurrence of null values, making up 99.04% (942,517) of the dataset.
- The **'Postal Code'** field also has 0.39% (3,716) null values, however the **'State_province'** variable has a 12.31% (117,192) null value occurrence.
- These cases of missing data may compromise the report's correctness and completeness, which may have an effect on the reliability of any analytical conclusions drawn from the dataset.

Q.2) From the Report Overview, Supplier Analysis, and Product Analysis pages, choose one page to analyse. Provide a detailed explanation of the objects presented on this page, including how and why they are created. **(16 points)**

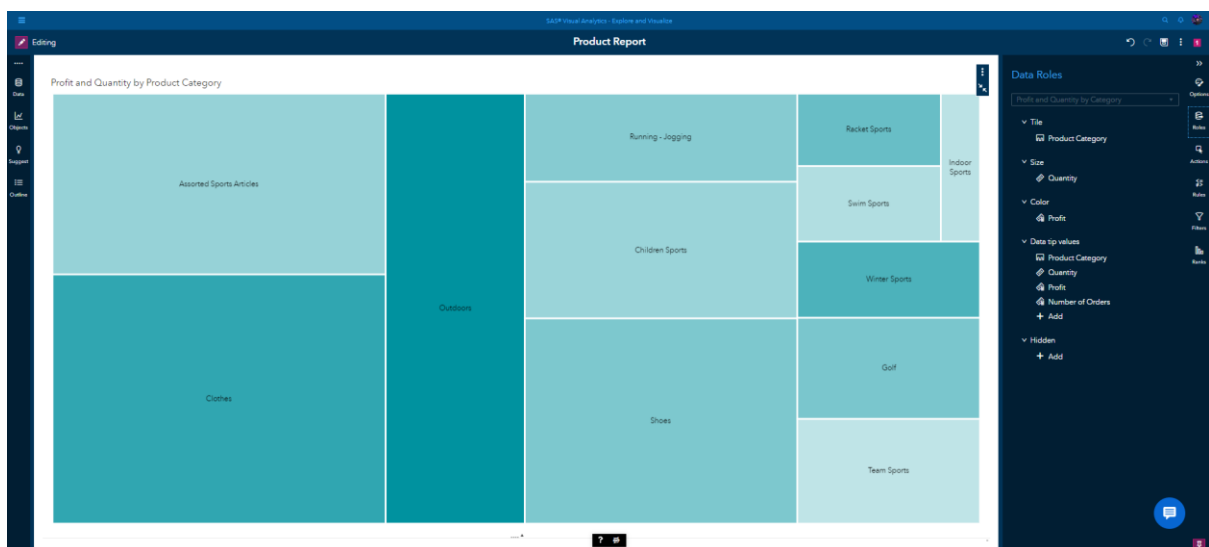


Solution:

We will be analysing the "Product Analysis" page from the Orion Star report. There are 5 objects on the page when we navigate to it.

1. Tree Map
2. Horizontal Bar Chart
3. Stack Container (consisting of 2 bar charts)
4. Dual Axis Time Series Plot
5. Information Tab

Tree Map



- A data visualization method called a Tree Map uses nested rectangles to show hierarchical data, with each rectangle's size denoting a numerical value while preserving the hierarchical structure.

- This map illustrates how "Product Category" relates to "Quantity," "Profit," and "Number of Orders."
- Every tile belongs to a particular product category.
- The size of the tiles is determined by the quantity of orders.
- The tile's hue indicates the size of the profit margin for that specific product category. Higher profit margins correspond to darker shades.
- Any tile will demonstrate details about the "Product Category," "Quantity," "Profit," and "Number of Orders" when your cursor is over it.

How is it created?

- Step 1: In the report pane, create a page called "Product Analysis."
- Step 2: From the panel on the left side of the page, select the Objects tab and then select Tree Map from the list of objects.
- Step 3: We will now select the Roles tab from the panel on the right in order to assign the roles to the map. We did this by dragging the data columns from the left-side Data tab.
- Step 4: Now, we configure the map's "Title," "Size," "Colour," And "Data Tip Values."
- Step 5: To modify the "Object Name," "Style," "Layout," "Graph Frame," "Tree Map," "Legend," and "Data Options," select the Options tab located on the right-side panel.

Horizontal Bar Chart



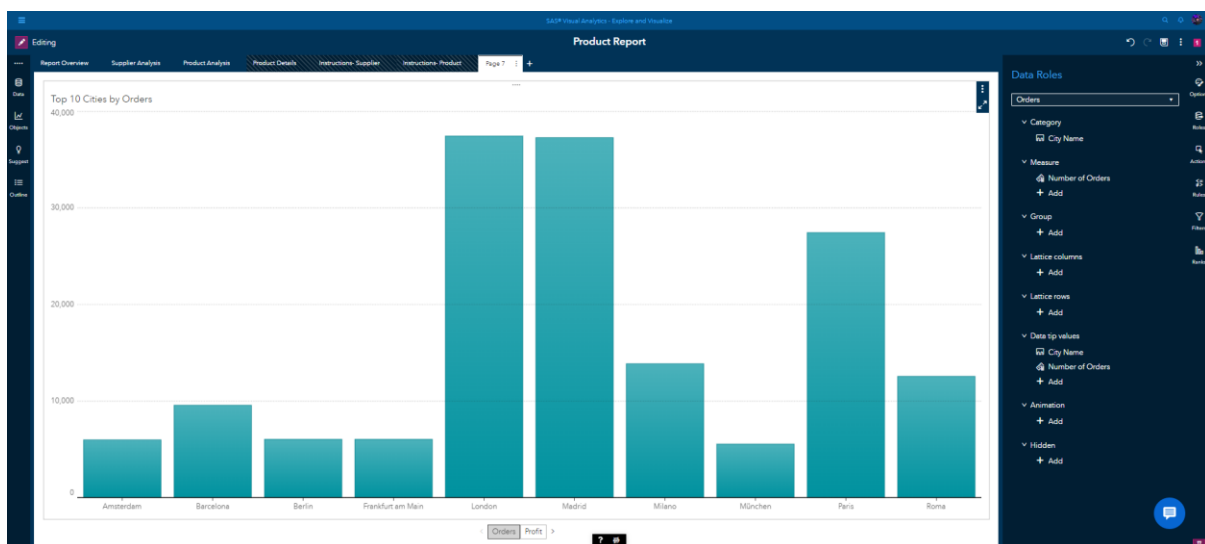
- A horizontal bar chart is a type of graphical data representation where categories are shown along the horizontal axis and each bar's length indicates a particular data value.
- This graph displays the relationship between the "Product Group" and "Profit", with the category on the y-axis and profit on the x-axis.
- Each bar belongs to a particular product group.
- The length of the bar shows the amount of profit each group of products has made.
- The bars are coloured red to indicate loss and green to indicate profit, respectively.

- When your cursor is over any bar, information about the "Product Group," "Profit," and "Number of Products" will appear.

How is it created?

- Step 1: From the list of objects on the Objects tab, create a new object named Bar Chart in the "Product Analysis" page.
- Step 2: To add the roles to the chart, we will now choose the Roles tab from the panel on the right. To accomplish this, we moved the data columns from the Data tab on the left, just like the tree map.
- Step 3: Now, we configure the map's "title," "size," "colour," and "data tip values."
- Step 4: To modify the "Object Name," "Style," "Layout," "Graph Frame," "Bar," "X Axis Options," "Y Axis Options," "Reference Lines," and "Legend," select the Options tab located on the right-side panel.
- Step 5: One thing to keep in mind is that the bar chart is by default always vertical. We adjust the orientation from the Bar section in the options tab on the right to convert it to a horizontal chart.

Stack Container



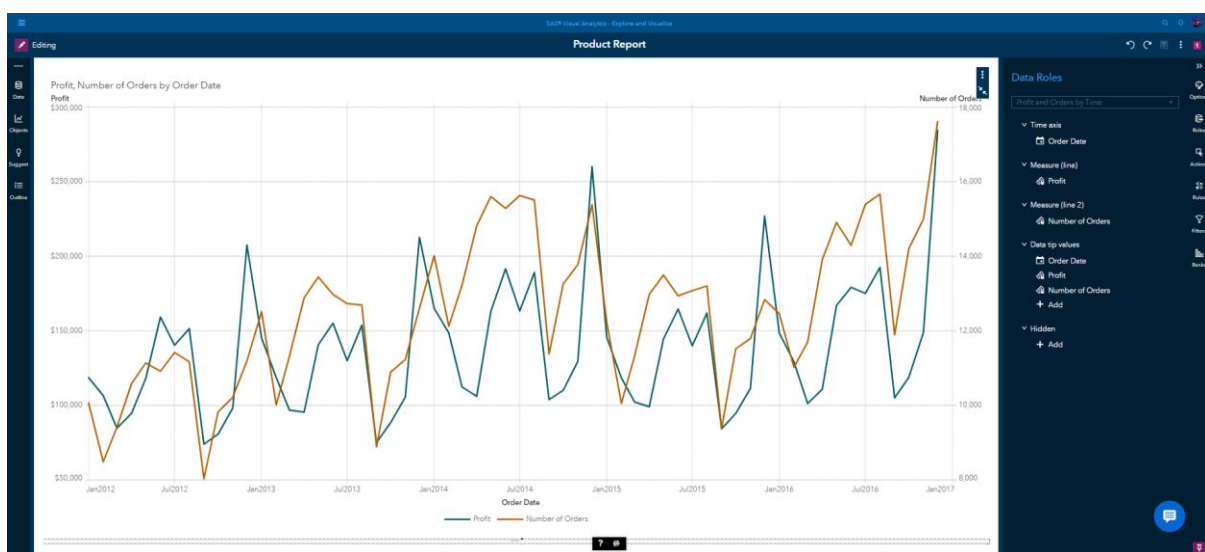
- For better report presentation and organisation, a stack container is a layout element that groups and arranges several visualisations or objects vertically inside a single container.
- We have grouped two bar charts in this container, which display the total orders and profits of the top ten cities in our data set.
- In the two charts, the bar heights represent the number of orders and profits, respectively.
- You will see information about the "City Name," "Orders," and "Profits," correspondingly, when your cursor is over any bar.

How is it created?

- Step 1: Find and create a stacking container by searching the objects list.

- Step 2: Go to the layout area of the Options tab, choose "stacking" under container type, and arrange the chart buttons according to your preference to stack the two charts on top of each other.
- Step 3: Except for altering this chart to horizontal, the rest of the processes remain the same, although depending on the chart we desire for this section.
- Step 4: We proceed to the ranks section and construct a new rank by picking the city name data in order to display the top ten cities.
- Step 5: Next, we'll configure "top count," which counts the highest-ranking cities in the data. We specify the "number of orders" and "profit" by which it will count, and we set the count to 10 to display the top 10 cities.

Dual Axis Time Series Plot

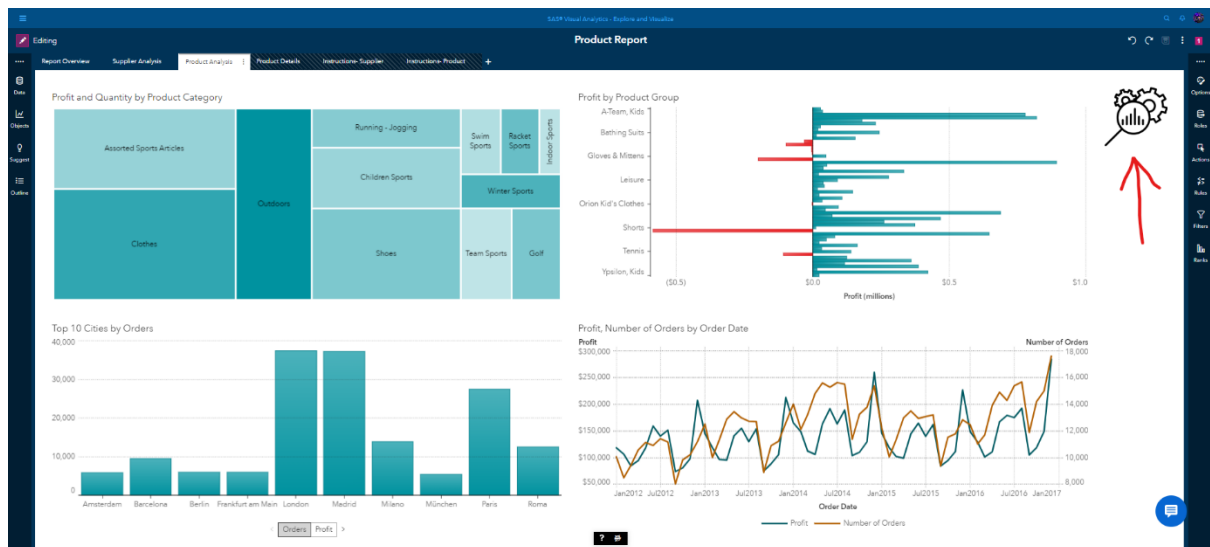


- A Dual Axis Time Series Plot is a type of data visualisation approach that shows two independent time series with different y-axes on the same chart. This makes it possible to compare patterns and relationships between various variables throughout time.
- This object compares the trends in "Profit" and "Number of orders" between January 2012 and January 2017.
- 'Number of orders' is on the right Y-axis, while 'Profit' is on the left. The 5-year period starting in January is also displayed on the X-axis.
- In contrast to other objects, this one doesn't display any data when we move the pointer over any trend lines.

How is it created?

- Step 1: From the list of items, choose the Dual Axis Time Series Plot.
- Step 2: In the time axis role, set the date period variable.
- Step 3: In measure line 1, enter "profit," and in measure line 2, enter "number of orders." this will divide them into two separate axes.
- Step 4: We use distinct, eye-catching colours from the style section of the options tab on the right to distinguish these axes.

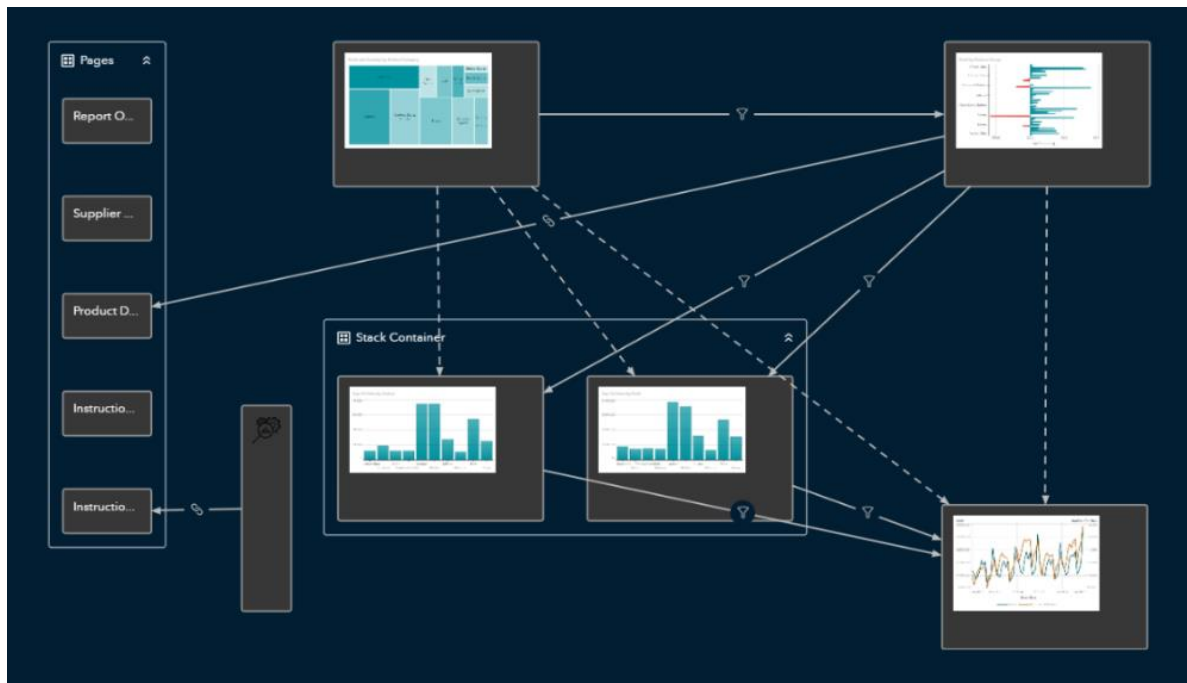
Information Section



- This section doesn't accomplish anything particularly special, unlike other objects. All it does is send us to a different page that summarises the contents of our "Product Analysis" page.

Q.3) The report creator of the Report Overview, Product Analysis, and Supplier Analysis pages might have used actions, filters, and links. Select one page from the three, and provide a clear explanation of any used actions, filters, and links on your chosen page, and step-by-step instructions on how to set them up in SAS Viya. **(12 points)**

Solution:



We will discuss the "Product analysis" page in this question, just as we did in the previous one.

We will use the graphic that appears in the upper-right corner of the Action tab to assist us understand any used actions, filters, and links.

Actions

- A drill-down action is initiated when a user clicks on a particular product category on the tree map, displaying detailed information about that category's items in the remaining objects.
- When a user selects the "Shoes" category, for example, the dual axis time series plot, the stack container below, and the horizontal bar charts all drill down and display only the detailed data that is available for that specific category.
- As we can see, the plots and charts are updated in accordance with the category guidelines.

Filters

- With filters, users can improve or dynamically subset the data shown in a visualisation according to defined criteria.
- Filters are responsible for these drill down operations.
- Every object is filtered sequentially. That is, the tree map filters the horizontal bar chart, the chart filters the stack container, which contains two bar charts, and lastly these charts filter the dual axis time series plot.

Links

- Links enable consumers to easily move between related topics by connecting various pages inside a report or visualisation.
- The horizontal bar charts on this page are linked to the "Product Details" page, whereas the "Information Section" is linked to a separate page called "Instructions- Product."

PART II (60 Points): Visual Story Line Development for Global Toys Corporation

Q.1) Explore the data and produce a report in SAS Viya for your company's senior management and the business intelligence department. For time consideration of the coursework assignment, this SAS report only needs to contain one or two pages. The business intelligence report should cover key facts about the toy company's performance on a global and regional level. These facts should include financial, marketing related data, and/or efficient use of resources. **(30 points)**

Solution:



We have established a product analysis page for the company, where we obtain insights into each product line in the sector of orders.

This page consists of four different objects:

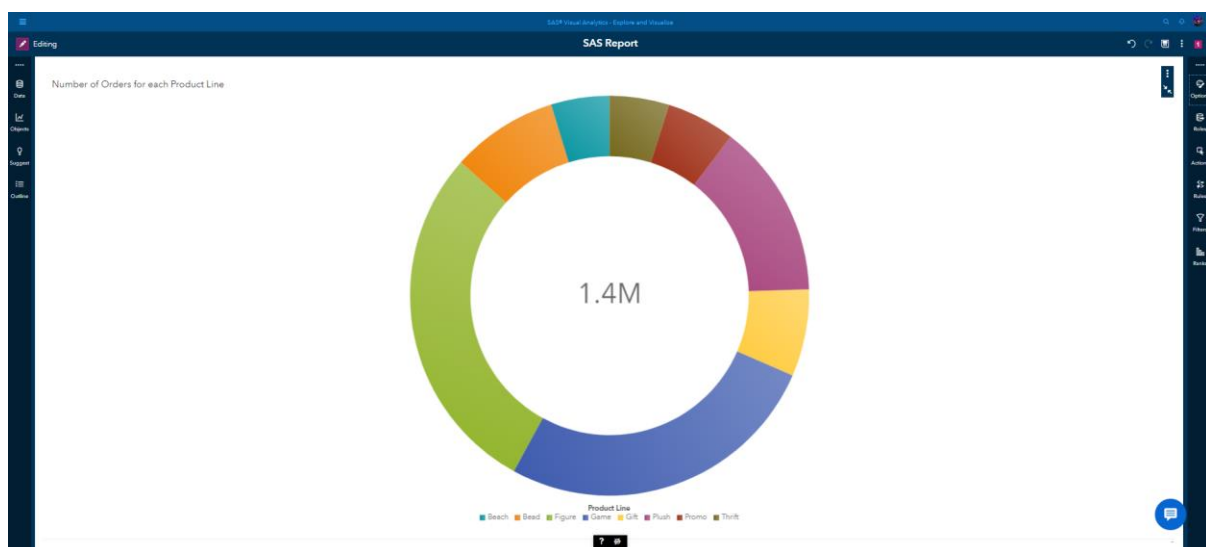
1. Slider
2. Pie Chart
3. Geo Coordinate
4. Line Chart

Slider



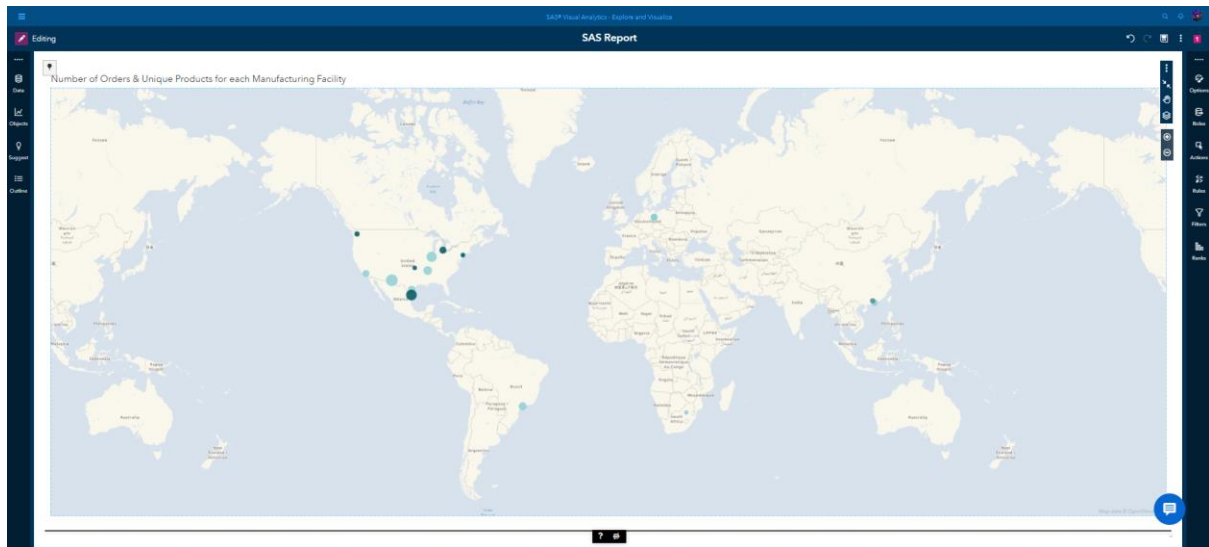
- Sliders are commonly used in SAS Viya when creating interactive dashboards or reports to allow users to filter or alter data based on a continuous range of values. They are especially useful in situations when users need to dynamically filter data using numeric or date/time variables.
- For this report, we are utilising it to pick the time period by selecting the year of data we want to see.

Pie Chart



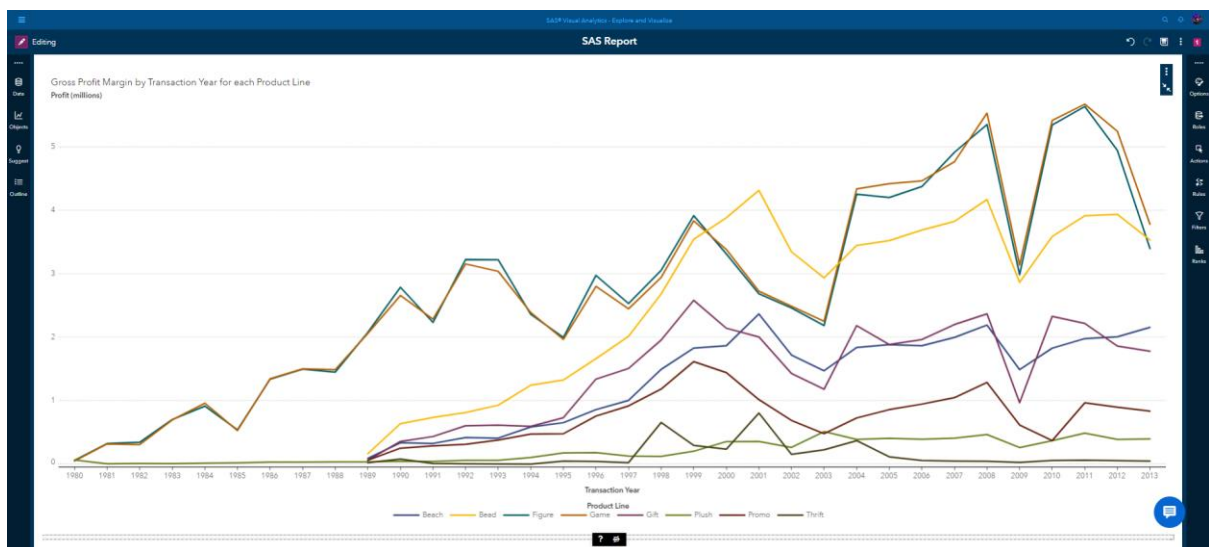
- A pie chart object in SAS visually shows the proportionate distribution of categorical data by splitting a circle into slices, with each slice representing a category and its relative size showing its proportion in the overall dataset.
- In this report, the pie chart splits the total number of orders in eight different product lines based on their metrics.

Geo Coordinate



- In SAS, a geo coordinate object allows you to represent and manipulate geographic coordinates (latitude and longitude) for spatial analysis and visualisation.
- This coordinate indicates the location of the manufacturing facilities, the total number of orders, and how many distinct items are produced.

Line Chart



- In SAS, a line chart object illustrates trends or correlations over a continuous axis by connecting data points with lines.
- From 1980 until 2013, each product line is represented by eight different lines.

These objects are linked with actions, filters, and links. That is, by picking a time period from the slider, we will receive updated data on each and every object.

Q.2) Provide clear explanations on how you choose your story line, what alternative story lines you have considered, and why you decided not to include them in your report. **(15 points)**

Solution:

Title: Unveiling Insights: A Revolutionary Strategy for Insight Toys Corporation

Introduction:

In the ever-changing world of toy production, staying ahead of trends and consumer preferences is critical. As Insight Toys Corporation's recently appointed Marketing Director, I am excited to deliver a detailed analysis that will transform our approach to product strategy and market positioning. We methodically developed a visual story line using powerful business intelligence technologies, delving deep into our business performance, and providing actionable insights to catapult us to unprecedented success.

Storyline:

Setting the stage:

- Emphasise the importance of data-driven decision-making in today's competitive marketplace.
- Introduce Insight Toys Corporation as a world leader in toy manufacture.
- Emphasise the Marketing Director's critical role in creating the company's future strategy.

Unveiling the Product Analysis Page

- Introduce the product analysis page, a key component of our strategic strategy.
- Emphasise the significance of understanding each product line's success in terms of growth and innovation.

The Advantages of Interactive Filtering:

- Showcase the slider feature, which allows you to dynamically select time ranges for examination.
- Explain how this interactive functionality allows users to personalise findings to their unique requirements and goals.

Piecing Together the Puzzle:

- Dive into the pie chart depiction, which shows how orders are distributed across product lines.
- Show how this visualisation helps discover top-performing product segments and areas for development.

Mapping our footprint:

- Explore the geo coordinate object, which depicts Insight Toys' manufacturing sites and their impact on production and distribution.
- Discuss the role of geographic knowledge in improving supply chain efficiency and market penetration.

Tracking Trends Over Time:

- Introduce a line chart to show historical patterns and correlations in product performance across decades.
- Show how this visualisation allows us to identify long-term growth trends and forecast market movements.

Conclusion:

In conclusion, Insight Toys Corporation's strategic imperative is to harness the power of data to drive innovation and stay ahead of the curve. We are ready to reinvent our product strategy, boost market positioning, and solidify our role as a toy industry innovator by harnessing the insights gained from our thorough investigation. With a clear goal and a data-driven approach, the opportunities for development and success are endless.

Q.3) During your report design and creation process, you must decide on the appropriate visualization tools/types to use based on the data you choose and the information you intend to portray. How did you choose the objects in your report (charts, tables, etc.)? How will the charts be perceived by a non-technical user? What questions would they ask and what answers could they derive based on your report? **(15 points)**

Solution:

When selecting visualisation tools/types for the report design and development process, the decision was made based on the specific data available and the information to be displayed. Here's how the objects in the report were selected:

Slider:

Purpose: Allow users to dynamically filter data depending on time periods (years).

Reasoning: The slider allows users to interactively select the time period they want to analyse, giving them more freedom and control while analysing historical data.

Pie Chart:

Purpose: To visualize the distribution of orders across different product lines.

Reasoning: Pie charts provide a simple and understandable representation of proportions, allowing non-technical consumers to quickly assess the relative success of each product line.

Geo Coordinate:

Purpose: To illustrate the geographic distribution of Insight Toys' manufacturing facilities and their impact on production and distribution.

Reasoning: Geo coordinates provide a spatial perspective, allowing users to understand the geographic spread of manufacturing facilities and the associated order volumes. This helps in assessing the global footprint of the company and optimizing supply chain operations.

Line Chart:

Purpose: To visualize trends and correlations in product performance over time.

Reasoning: Line charts are effective in showing trends and patterns over continuous axes, making them suitable for showcasing historical data trends. By plotting product performance over decades, users can identify long-term growth trajectories and potential market shifts.

Perception by Non-Technical Users:

Non-technical users are likely to find the report visually appealing and easy to understand due to the use of intuitive visualization tools.

They may ask questions such as:

- "What is the trend in sales over the years?"
- "Which product lines contribute the most to our overall sales?"

- "Where are our manufacturing facilities located, and how does it impact our production and distribution?"

The report provides answers to these questions:

- The line chart shows the trend in sales over decades, allowing users to identify patterns and seasonality.
- The pie chart illustrates the distribution of orders across product lines, helping users identify top-performing and underperforming product segments.
- The geo coordinate map displays the geographic spread of manufacturing facilities, providing insights into their location and impact on production and distribution.

Overall, the report empowers non-technical users to derive actionable insights from the data, enabling informed decision-making and strategic planning for Insight Toys Corporation.