

Project Based Learner Series

Exploring Data using Visualizations



Cristina Anton
Cristina.Anton@sas.com



Rachel McLawhon
Rachel.McLawhon@sas.com



Who We Are

Global Academic Programs

Meet the Instructors



Cristina
Anton



Rachel
McLawhon



Jackie
Johnson



Tom
Grant

Global Academic Programs

Our Mission: Empowering Analytics for Academics



www.sas.com/academics

We support the next generation by providing **free software and resources** for academia.

Our “customers” are students, educators, independent learners, and academic researchers.

Resources for Teaching and Learning



SOFTWARE

- SAS® Viya® for Learners
- SAS® OnDemand for Academics
- SAS® Viya® Workbench for Learners



TRAINING

- e-Learning
- How-to videos
- Onsite & virtual training



EDUCATIONAL RESOURCES

- Teaching materials
- Curriculum consulting
- Academic specializations
- SAS Educator Portal
- SAS Skill Builder for Students

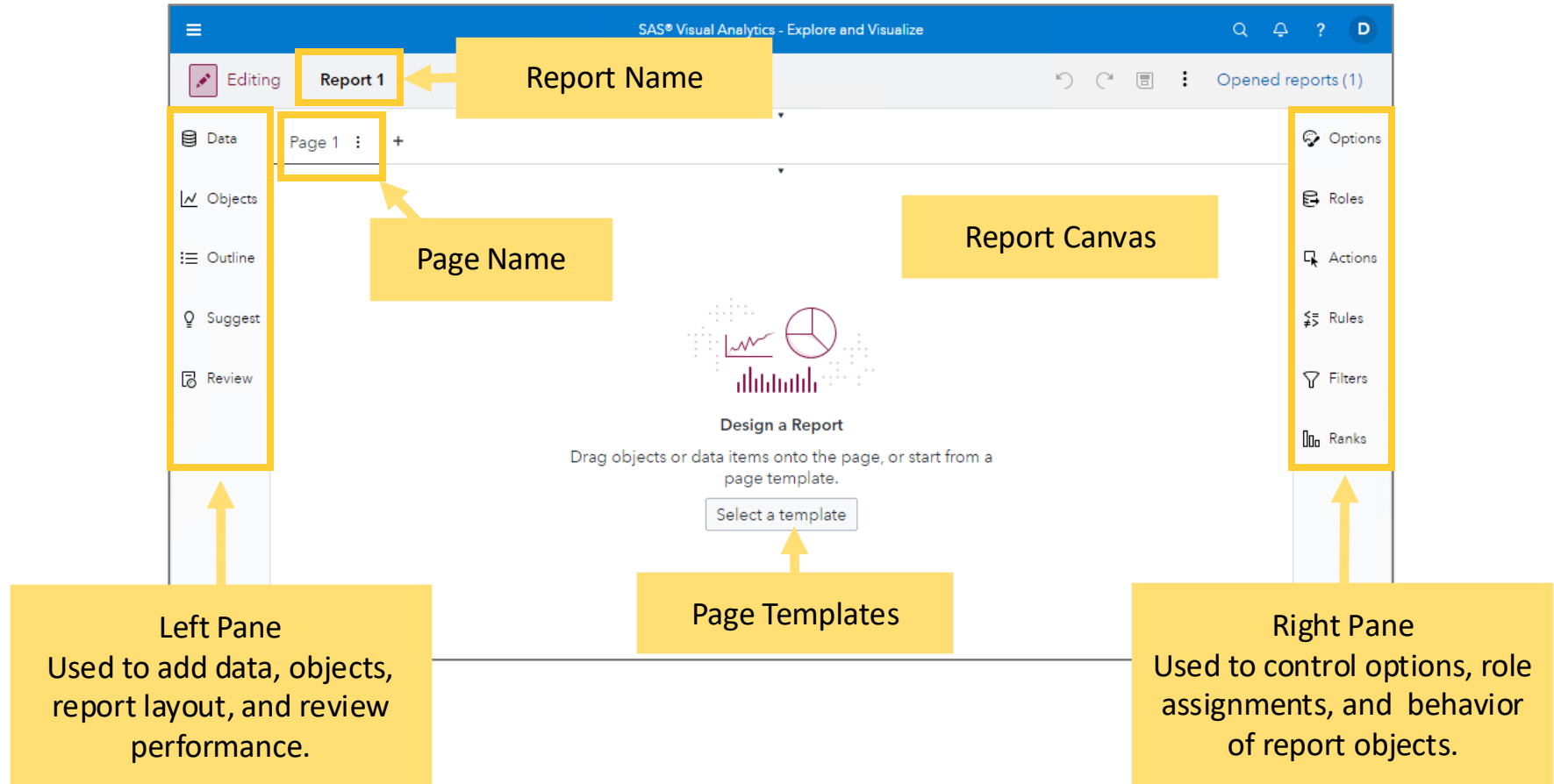
Starting our journey...



What Is SAS Visual Analytics?

SAS Visual Analytics is a web-based product that leverages SAS High-Performance Analytics technologies to empower organizations to explore huge volumes of data very quickly to identify patterns, trends, and opportunities for further analysis.

Working with the Reporting Interface



About the Data Source

- Able to add multiple data sources to a report.
- Cardinality is only displayed for category, date, time, and datetime data items.

Frequency and Frequency Percent are automatically added for each data source.

The screenshot shows a software interface for configuring data sources. The main panel is titled 'Data' and contains a dropdown menu currently set to 'PRDSALE'. Below this is a 'Filter' input field, a '+ New data item' button, and a 'Category' section with a list of items: Country - 3, Division - 2, Month - 12, Product - 5, Product type - 2, Region - 2, and Year - 2. The 'Country - 3' item is highlighted with a yellow box. Below the 'Category' section is a 'Measure' section with a list of items: Actual Sales, Frequency, Predicted Sales, and Aggregated Measure. The 'Frequency' item is highlighted with a yellow box. Below the 'Measure' section is an 'Aggregated Measure' section with a list of items: Frequency Percent, which is also highlighted with a yellow box. Annotations with yellow boxes and arrows point to these elements: 'Active data source' points to the 'PRDSALE' dropdown; 'Add a data source' points to the 'Add data' button; 'Cardinality' points to the 'Country - 3' item; and the text 'Frequency and Frequency Percent are automatically added for each data source.' points to the 'Frequency' and 'Frequency Percent' items.

Active data source

PRDSALE

Add data

Add a data source

Cardinality

Country - 3

Frequency

Frequency Percent



It is recommended to have no more than six data sources per report.

Data Item Tooltips

Hover over a data item to display its tooltips.

Displays data items used in

- data mappings
- hierarchies
- objects
- filters
- calculated items
- rules
- ranks.

The screenshot shows a BI tool interface with a sidebar on the left containing a 'Data' section. The 'Actual Sales' item is highlighted with a yellow box. A tooltip is displayed over this item, showing its details. The main area of the interface shows a dashboard titled 'Actual Sales Summary' with a chart and a table.

Data Item Details:

- Name: Actual Sales
- Name in data: ACTUAL
- Format: Dollar
- Aggregation: Sum

Used by:

- Difference expression
- Key value - Actual Sales 1 role
- Time - Year Month 1 role
- List table - Product type 1 role
- Treemap - Product 1 role
- Targeted bar - Region 1 role

Dashboard Data:

Actual Sales: \$369K

Filters: No selections

Buttons: FURNITURE, OFFICE

Chart: Predicted Sales by Year

Actual Sales	Predicted Sales	Difference
\$69,463	\$70,923	\$1,460
\$79,432	\$76,478	-\$2,954



Are There Outliers Values of Cylinders?

There are 5 outlier values of Cylinders. These outliers do not change the overall sum, average, or median by more than 5%.



Box plot of outlier values

What are the Details of these Outliers?

	Cylinders	Engine Size (L)	MPG (City)	Horsepower	Weight (LBS)	MSRP
	12	6	12	420	5399	\$75,000
	12	5.5	13	493	4473	\$128,420
	12	5.5	13	493	4429	\$126,670
	10	6.8	10	310	7190	\$41,475
	10	8.3	12	500	3410	\$81,795

Detailed list table of outlier values and related measures

What Is the Effect of Outliers on Cylinders?

Metric	Including Outliers	Excluding Outliers	Outlier Impact	Difference
Sum	2474	2418	2.26%	56
Average	5.8075117371	5.7434679335	1.10%	0.0640438036
Median	6	6	0.00%	0

Impact assessment that quantifies the effect of outlier values

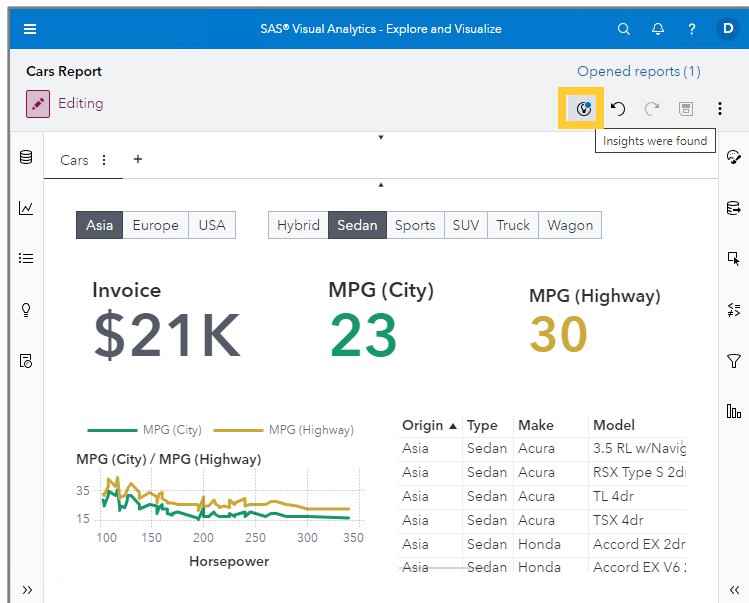
Close

Insights

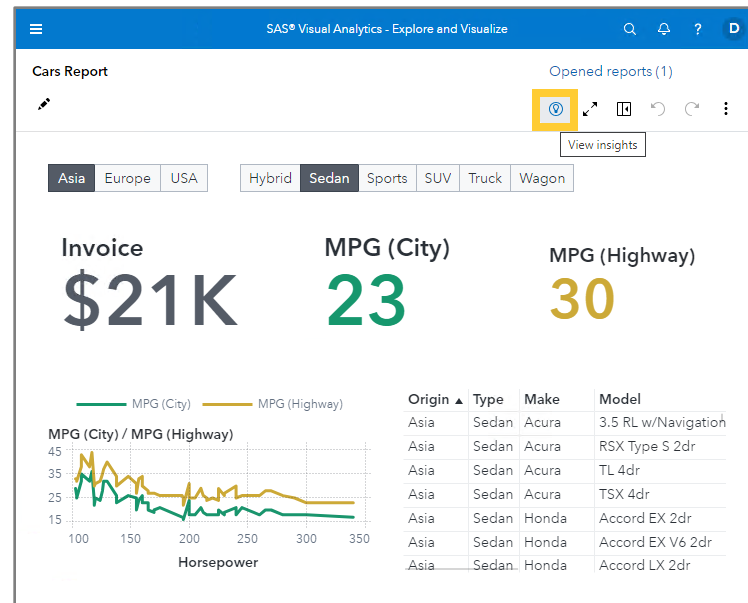
- Indicators alert users that insights have been automatically detected.
- Only data items used in the report content are analyzed.



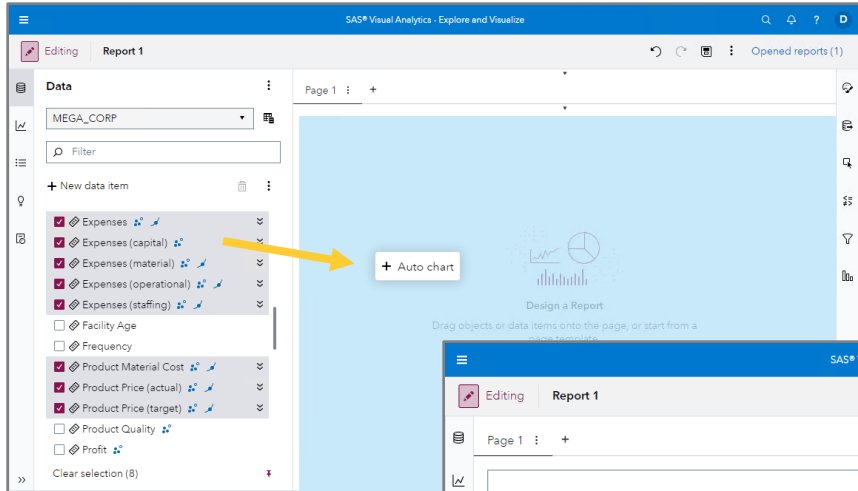
Edit Mode: Insights found



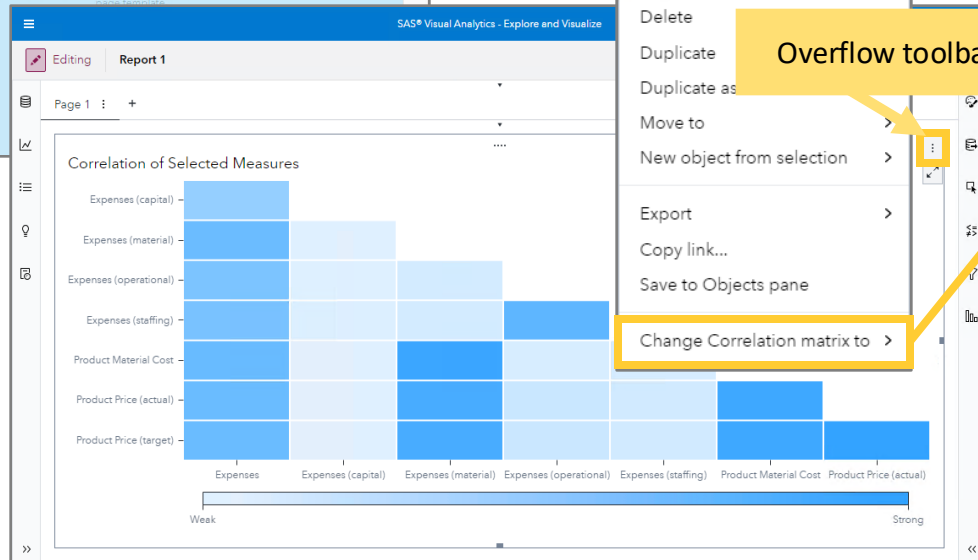
View Mode: View Insights



Automatic Chart and Change Object Type



- Quick view of data
- Selects best chart type



Overflow toolbar

- Box plot
- Crosstab
- Heat map
- List table
- Numeric series plot
- Parallel coordinates plot
- Scatter plot
- Automated explanation
- Bar chart
- Bubble change plot
- Bubble plot
- Butterfly chart
- Cluster
- Comparative time series plot
- Decision tree
- Dot plot
- Dual axis bar chart



List table

View summary or detail

Product Brand ▲	Product Line	Product Sale	Product Quality	Number of Products
Novelty	Bead	\$453,780.79	87%	13,371
Novelty	Gift	\$483,820.91	87%	15,906
Novelty	Kiosk	\$460,017.64	86%	80,841
Novelty	Promo	\$456,098.01	87%	10,146
Novelty	Store	\$451,327.08	87%	30,061
Toy	Figurine	\$1,406,497.10	86%	47,425
Toy	Game	\$1,340,276.72	87%	24,962
Toy	Plush	\$1,232,834.23	86%	29,364

- View distinct values
- Identify range of values
- Spot data quality issues



Values are aggregated by default.



Place high-cardinality categories in rows.



Crosstab

- Spot data quality issues
- Broader insight

View distinct categories and summary of measures

Product Brand ▲		Novelty		Toy	
Facility Continent ▲	Facility Country ▲	Product Quality	Product Sale	Product Quality	Product Sale
Europe	Denmark	87%	\$17,534.44	88%	\$16,911.57
	France	88%	\$16,759.51	89%	\$26,913.14
	Germany	89%	\$58,661.07	89%	\$69,943.17
	Italy	89%	\$30,528.83	89%	\$37,815.75
	Norway	89%	\$37,560.40	90%	\$47,500.83
	Spain	90%	\$136,785.75	90%	\$200,808.93
	Sweden	90%	\$57,266.41	90%	\$80,312.61
	United Kingdom	90%	\$141,040.65	90%	\$211,459.07
North America	Canada	87%	\$49,619.02	84%	\$56,983.75
	Mexico	77%	\$12,659.72	77%	\$28,663.89
	United States	86%	\$1,680,58...	84%	\$1,889,59...
Oceania	Australia	86%	\$54,385.88	87%	\$70,250.15
	New Zealand	86%	\$11,659.91	86%	\$20,542.58



Place low-cardinality categories in columns.

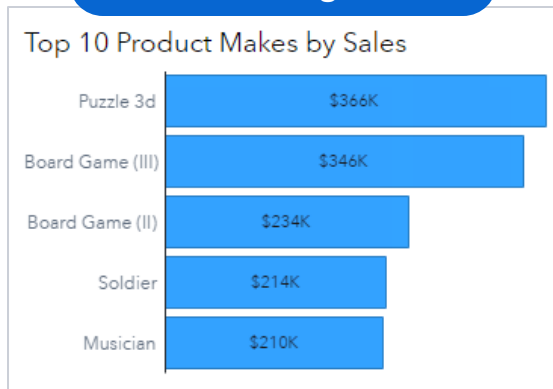


Bar chart

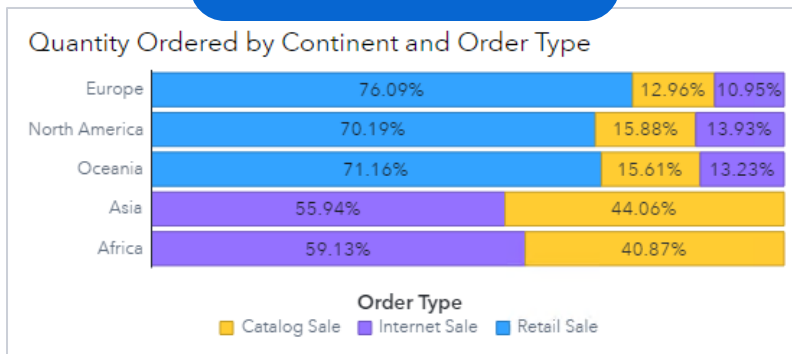
General comparisons



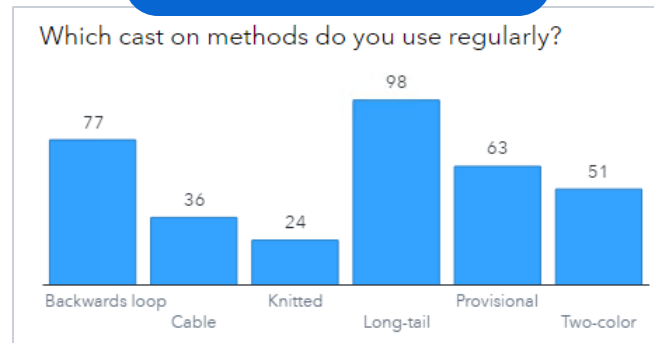
Rankings



Part-to-whole



Questionnaire results

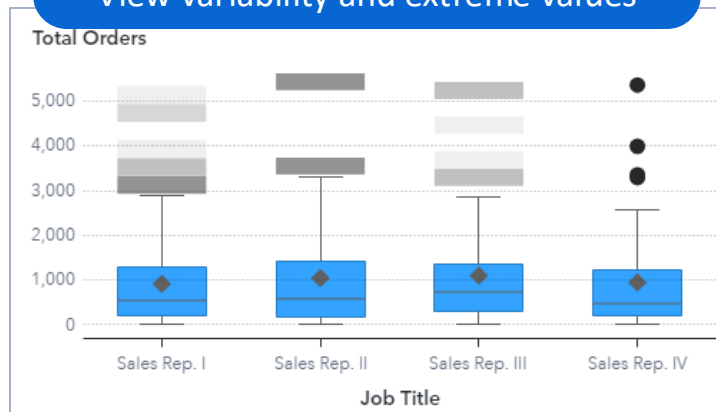


Use a zero baseline

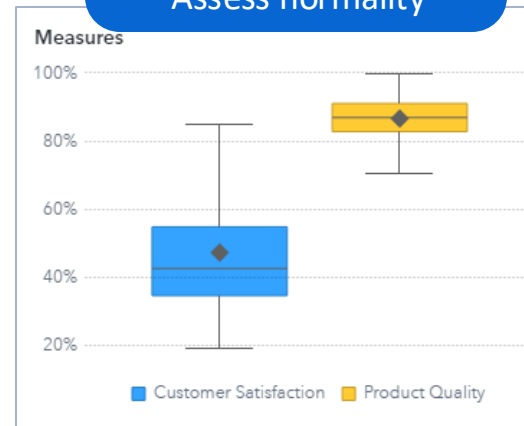


Box plot

View variability and extreme values

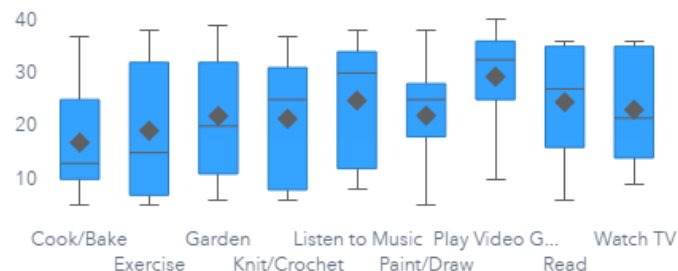


Assess normality



Show survey or questionnaire results

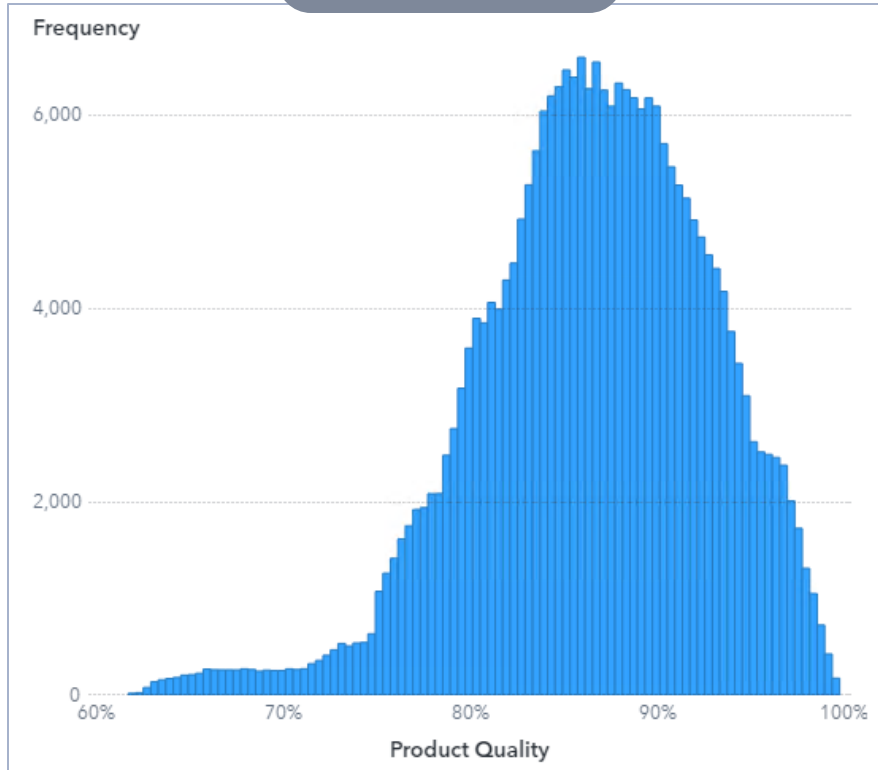
How many hours a week do you ...?





Histogram

Distribution



- Gauge overall range
- Identify values with high number of observations
- Determine normality



Bubble plot

Three dimensions

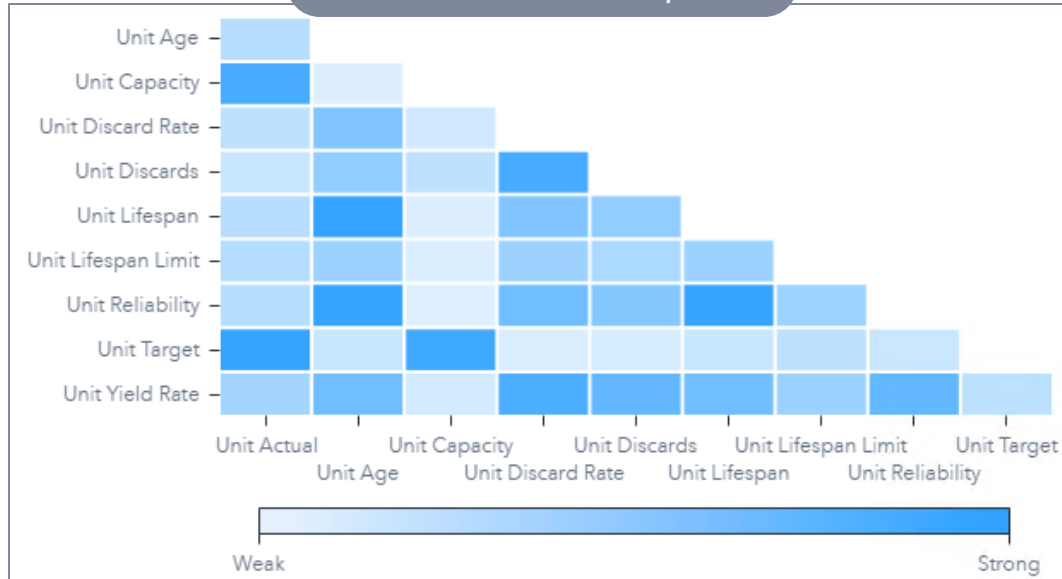


- Two measures = location
- Third measure = size
- Size scaled to relative minimum and maximum values



Correlation matrix

Linear relationship



- Cell represents intersection of two measures
- Color indicates degree of correlation
 - Light = weak correlation
 - Dark = strong correlation



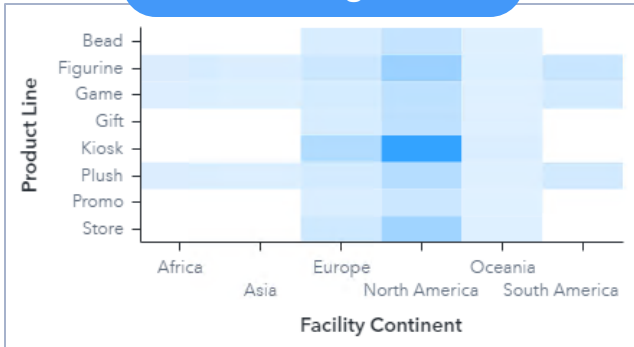
Heat map

Two high-cardinality measures

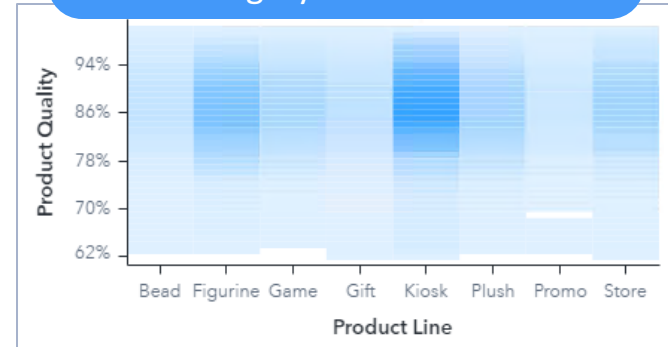


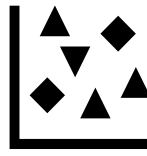
- Visualize trends between measures
- Pinpoint possible outliers
- Identify nonlinear relationships

Two categories



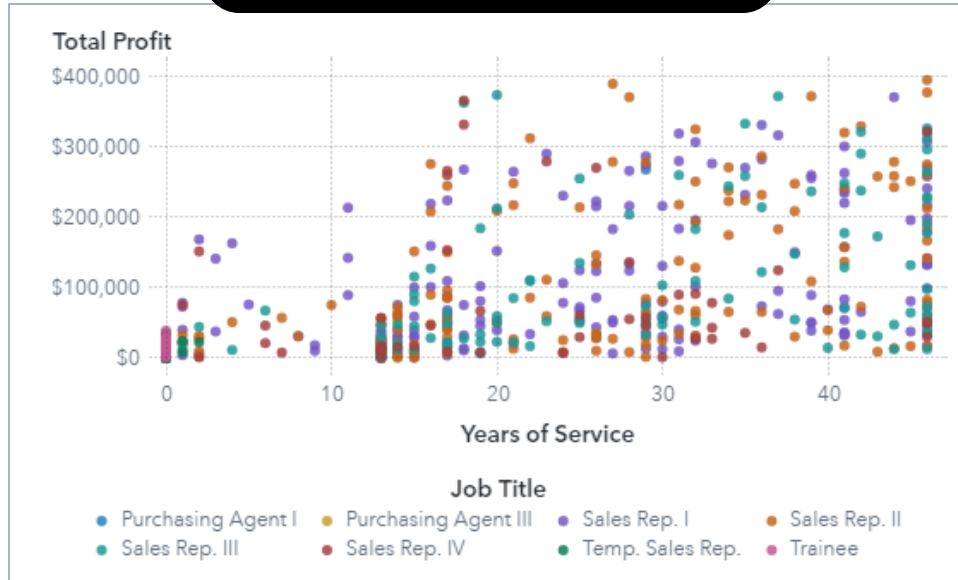
Category and measure





Scatter plot

Two low-cardinality measures



- Visualize trends
- Pinpoint possible outliers
- Identify non-linear relationships

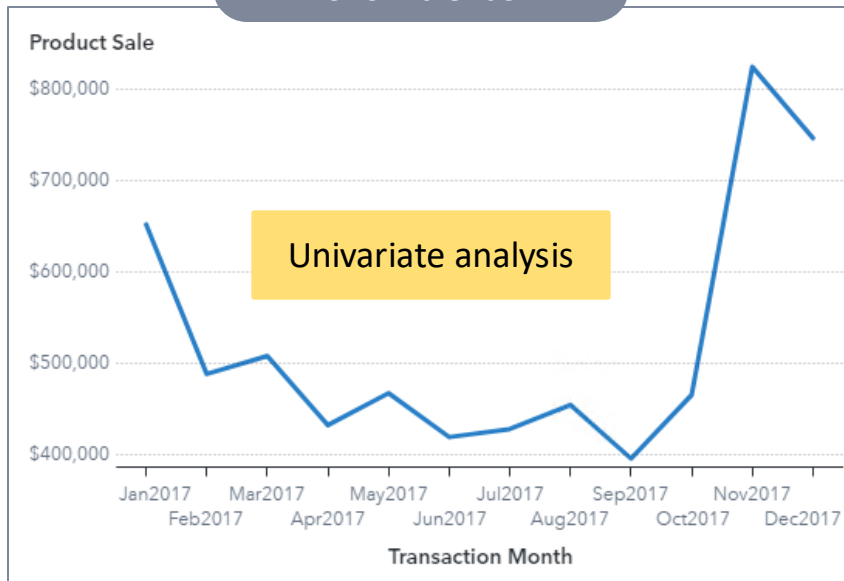


Display independent variable on horizontal axis

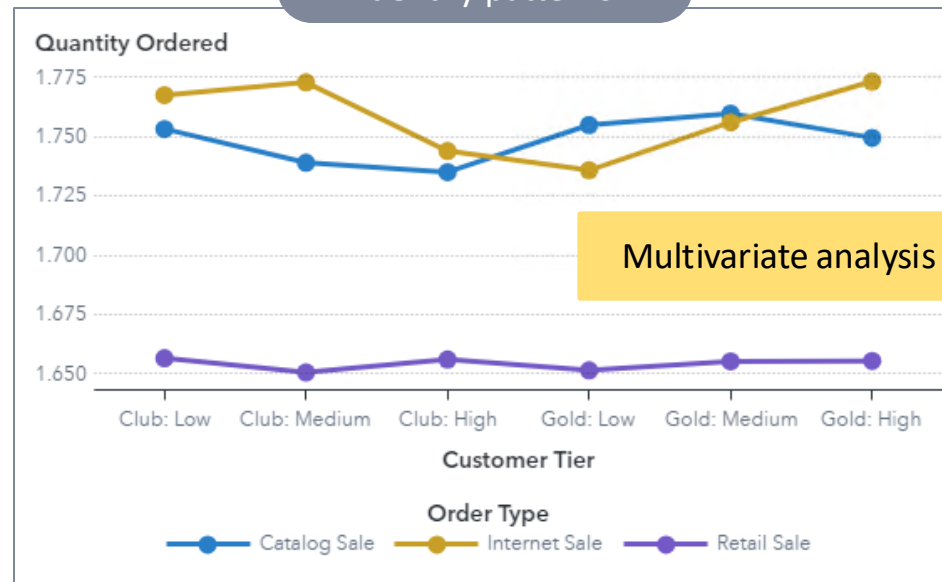


Line chart

Show trends



Identify patterns

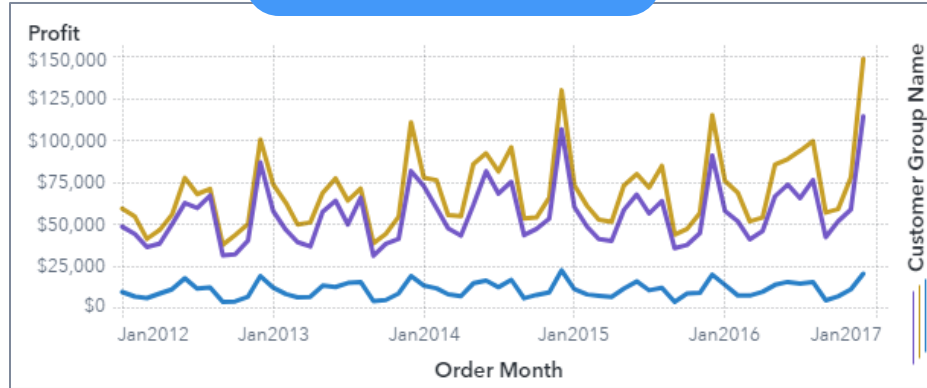


Set a zero baseline using the Options pane

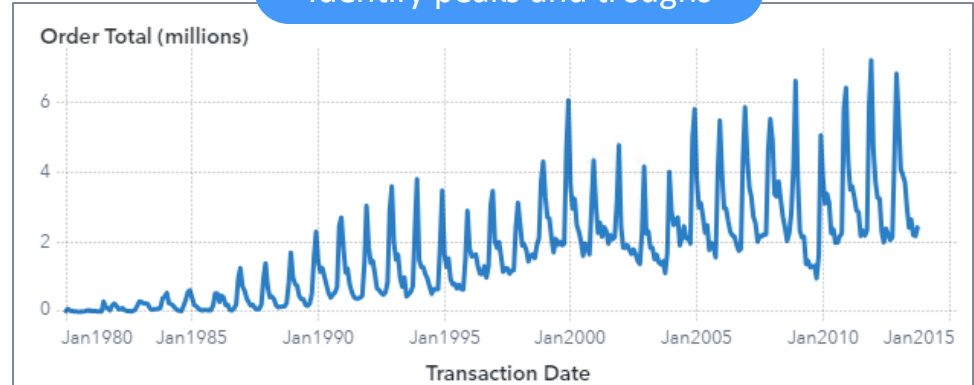


Time series plot

Show trends



Identify peaks and troughs





Automated Explanation

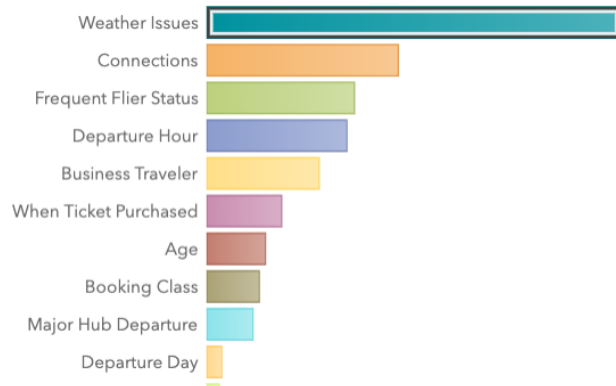
Advanced Analysis

What are the characteristics of No Show?

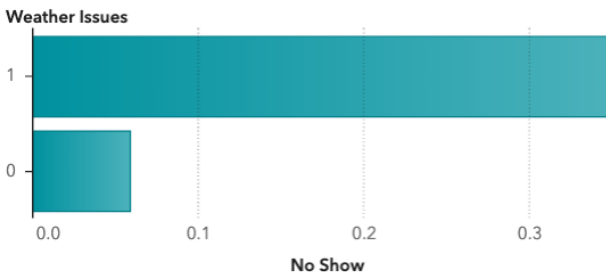
No Show ranges from 0 to 1. Average No Show is 0.1. No factor best differentiates the highest (top 10%) and the lowest (bottom 10%) No Show cases. The three most related factors are Weather Issues, Connections, and Frequent Flier Status.

There are 998 cases that might be outliers, with No Show greater than or equal to 1.

What factors are most related to No Show?



What is the relationship between No Show and Weather Issues?



When Weather Issues is 1, the average of No Show is 0.35. When Weather Issues is 0, the average of No Show is 0.06. The most common Weather Issues value is 0.

Determine the most important data items that explain the values of the selected data item.

Portals and e-Learning Access

Project Based Learner Series

Course Page

[Live web](#)[Overview](#)[Course Materials](#)[Instructor Material](#)

Course Materials

Session 1: Getting Started with Data and SAS

- 🔗 Create a SAS Profile with your university affiliated email address
- 🔗 Register for Viya for Learners Software
- 🔗 Launch Viya for Learners Software
- 📄 For Students: Accessing SAS Skill Builder for Students
- 📄 For Educators: Accessing the SAS Educator Portal

Session 2: Exploring Data Using Visualizations

- 📄 Activity: Exploring Tennis Players' Earnings in SAS Visual Analytics

Session 3: Data Wrangling and Descriptive Statistics

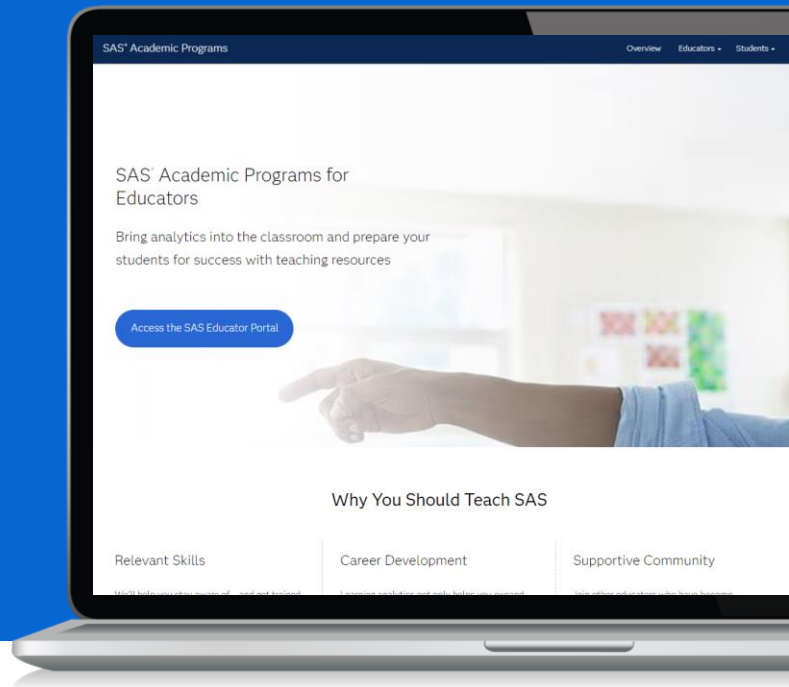
- 🔗 Course: SAS Viya Overview
- 🔗 Getting Started with SAS Studio

<https://learn.sas.com/course/view.php?name=PBL5>

Educator Portal

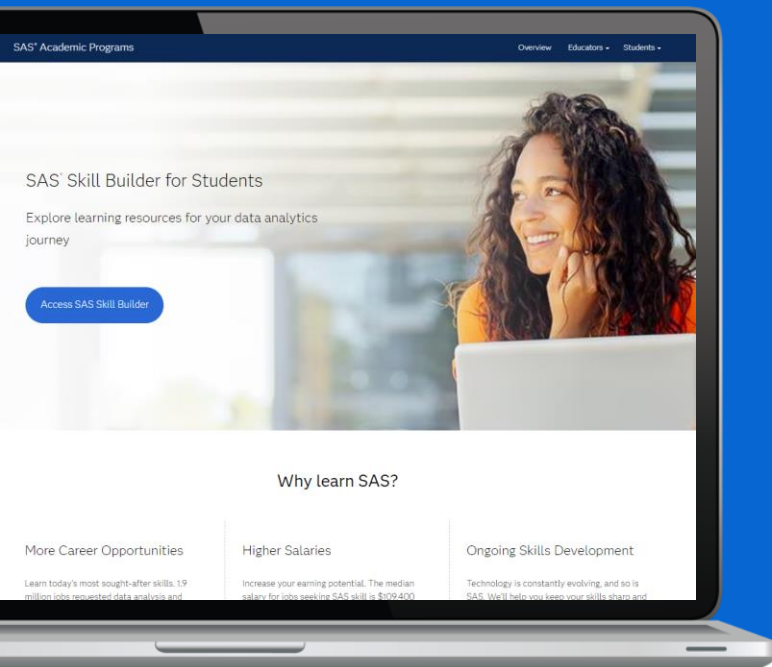
All-in-One Resource Center

- E-learning & Certification Paths
- Teaching Materials
- Certification Preparation Materials
- Educator Workshops & Training
- Discounts & Promotions



SAS Skill Builder for Students

Your future starts here



- Learn on your time
 - Free e-learning courses
 - Certification prep materials
- Free software
- Career resources
 - Discover why data analytics is a rewarding career
 - Learn about what it means to be a data analytics professional
 - Connect with employers
- Cross-industry access
 - Analytical skills for any discipline

www.sas.com/skillbuilder

Explore Helpful Resources



[SAS Visual Analytics Gallery](#)

A place to share and chat about visualizations.



[Ask the Expert](#)

View other user webinars that provide insights into using SAS products to make your job easier.



[SAS Support Communities](#)

Ask questions, get answers and share insights with SAS users.



[SAS Users YouTube Channel](#)

A plethora of videos on hundreds of topics, just for SAS users.