

SAS Studio and Flows in SAS Viya

Cross Topics network meeting in Kristiansand
The 7 of March

Pia Skare Rønnevik, Customer Success Manager, FANS
Email: pia.roennevik@sas.com



SAS Studio

For everyone

Who:

Data engineers/ETL developers
Analysts
Business Intelligence Analysts
Customer Intelligence Analysts
Data scientists

.....

Modes of work:

- No code
- Low code
- Own code

Some benefits of No code/Low code approach:

- Fast build – save time to insights (time is money)
- Easy to use – tasks and flow steps
- Standardization – mostly autogenerated code

SAS Viya

SAS Platform

SAS Viya

- an open, cloud-enabled, analytic run-time environment

SAS Cloud Analytic Services (CAS)



in-memory
engine



fast
processing



data of
any size

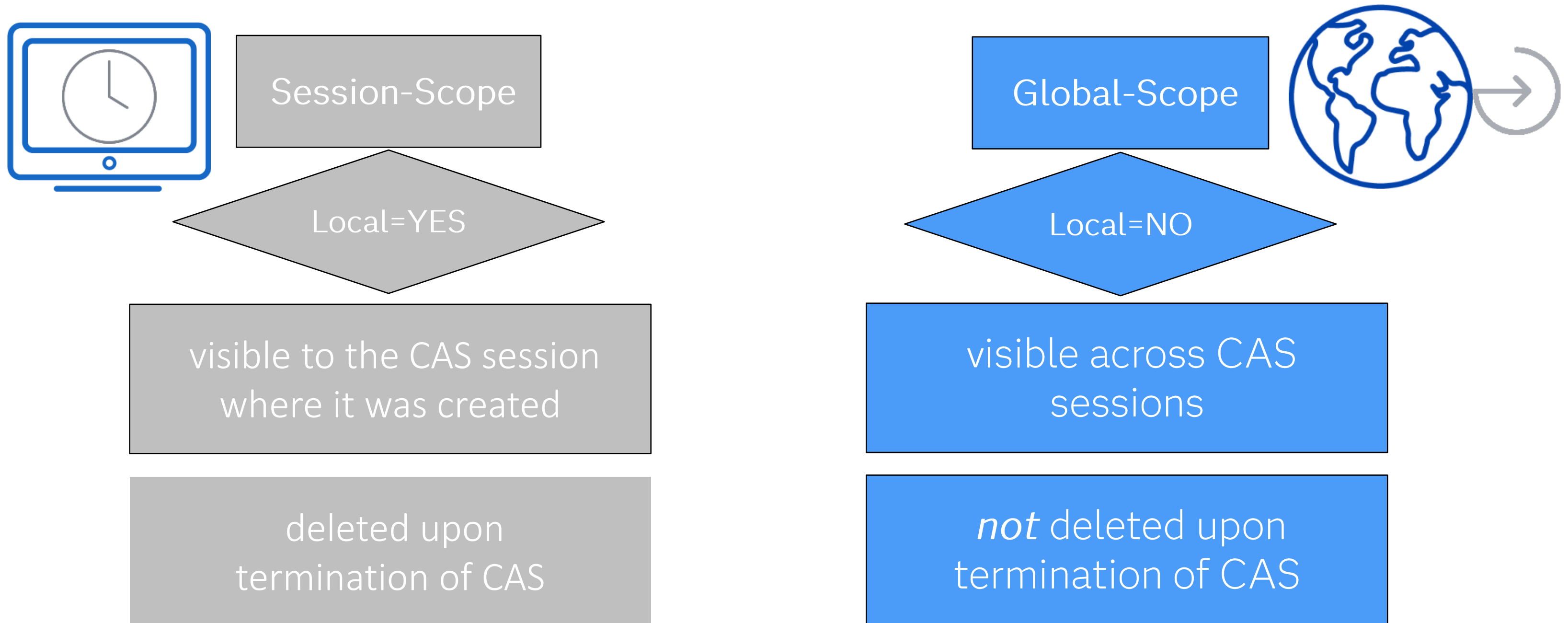
SAS vs CAS?

- You can run your existing SAS 9 programs on SAS Viya platform without changes!
- You won't lose anything, but you won't win either!



- You can change your existing SAS 9 programs so that they can run in CAS and not in SAS.
- You take advantage of multi-threading and parallel processing => You can win a lot!

Session versus Global



Session versus Global



Session-Scope

default

Promote=NO

visible only to the CAS session where it was created

visible only to the user who created the table

dropped from memory upon termination of CAS



Global-Scope

Promote=YES



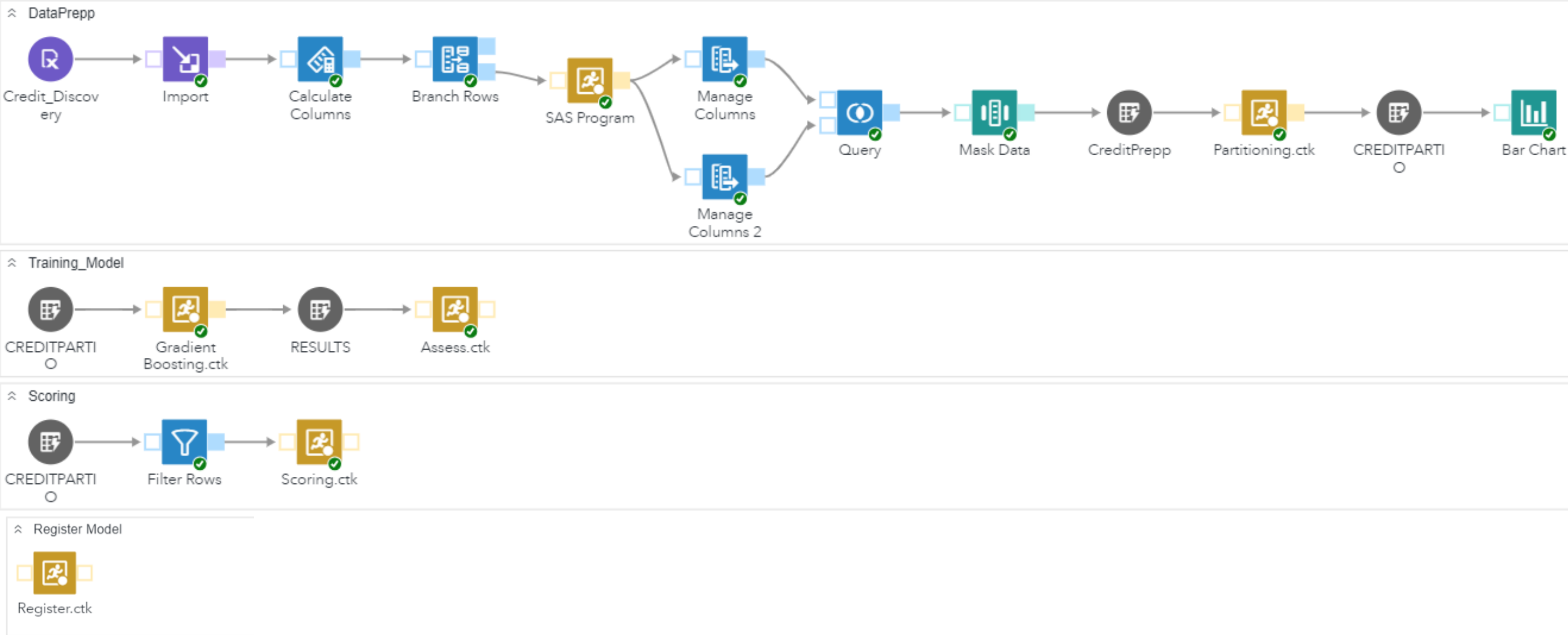
promoted table

visible across CAS sessions

visible to any user who can access the global-scope caslib

not dropped from memory upon termination of CAS

SAS Studio - flows



Create a flow

SAS® Studio - Develop Code and Flows

New Options View Open Save All

SAS program
Python program
Flow
Query
Custom step
Quick import
Job
Task
More file types

Develop
Python Program
SAS Program
Transform Data
Branch Rows
Calculate Columns
Filter Rows
Insert Rows
Manage Columns
Mask Data
Query
Rank Data
Remove Duplicates
Select Random Sample
Sort
Split Columns
Stack Columns
Transpose Data
Union Rows
Integrate
Execute Decisions
Implement SCD
Load Table
Merge Table
Statistical Process Control
Analysis of Means

Start Page x +

GET STARTED

Program in SAS
Build a flow
Import data
Query data

NEW Explore new features in SAS Studio

LEARN

Learn SAS Studio - videos, tutorials, and training

Learn SAS programming

STAY CONNECTED

Join the community

Request a feature

RECENTS

Flow-Demo.flw /Users/norpro/My Folder	Feb 28, 2024, 4:22:36 PM
Program 7 CAS Action.sas /Users/norpro/My Folder/CAS Actions	Feb 22, 2024, 3:50:24 PM
DISTINCTCARS CASUSER	Feb 22, 2024, 3:40:40 PM
ORIGINBYMAKE CASUSER	Feb 22, 2024, 3:38:12 PM
CARS CASUSER	Feb 22, 2024, 2:31:59 PM
Program 5 CAS Action.sas /Users/norpro/My Folder/CAS Actions	Feb 22, 2024, 3:38:56 PM
CARS_FREQ CASUSER	Feb 22, 2024, 3:35:40 PM
Program 6 CAS Action.sas /Users/norpro/My Folder/CAS Actions	Feb 22, 2024, 3:18:49 PM
Program 4 CAS Action.sas /Users/norpro/My Folder/CAS Actions	Feb 22, 2024, 3:17:15 PM
Program 3 CAS Action.sas /Users/norpro/My Folder/CAS Actions	Feb 22, 2024, 3:16:29 PM
Program 2 CAS Action.sas /Users/norpro/My Folder/CAS Actions	Feb 22, 2024, 2:37:25 PM
Program 1 CAS Action.sas /Users/norpro/My Folder/CAS Actions	Feb 22, 2024, 1:59:58 PM

Create flow from a SAS program

The screenshot displays the SAS Studio interface with the following components:

- Steps Panel (Left):** A sidebar containing various step categories such as Data (Input and Output), Develop, Transform Data, Integrate, Statistical Process Control, and Visualize Data. The 'SAS Program' step is highlighted under the 'Develop' category.
- Code Editor (Center):** A window titled 'SAS Program.sas' containing SAS code for processing earthquake data. The code includes a data step to create a new variable 'Depth_Cat' based on 'Depth' and a frequency procedure to analyze the distribution of 'Depth_Cat'.

```
1 data earthquakes;
2 set sashelp.quakes;
3 length Depth_Cat $12;
4 Magnitude=round(Magnitude,.1);
5 if Depth<70 then Depth_Cat="Shallow";
6 else if Depth<300 then Depth_Cat="Intermediate";
7 else if Depth<=700 then Depth_Cat="Deep";
8 run;
9
10 proc freq data=earthquakes nlevels;
11 tables Depth_Cat 'Type'n /nocum;
12 run;
13
14
```
- Log Panel (Right):** A panel showing the execution log with the message "There are no messages." It includes status indicators for Errors (0), Warnings (0), and Notes (0).
- Context Menu (Far Right):** A menu is open over the Log panel, listing actions such as "Schedule as a job", "Create flow from program", "Add to My Favorites", "Open in a browser tab", "E-Mail", "Print", "Download", "Tab layout", and "Refresh".
- Annotations:** Two dark blue callout boxes with white text and arrows. The first box, pointing to the 'Code to Flow' button in the toolbar, contains the text "If a flow is available, add the SAS program here". The second box, pointing to the "Create flow from program" menu item, contains the text "Create flow from a SAS program".

Flow steps

Flow file name

Flow tab section

Flow properties

Preview tab section

Add Notes

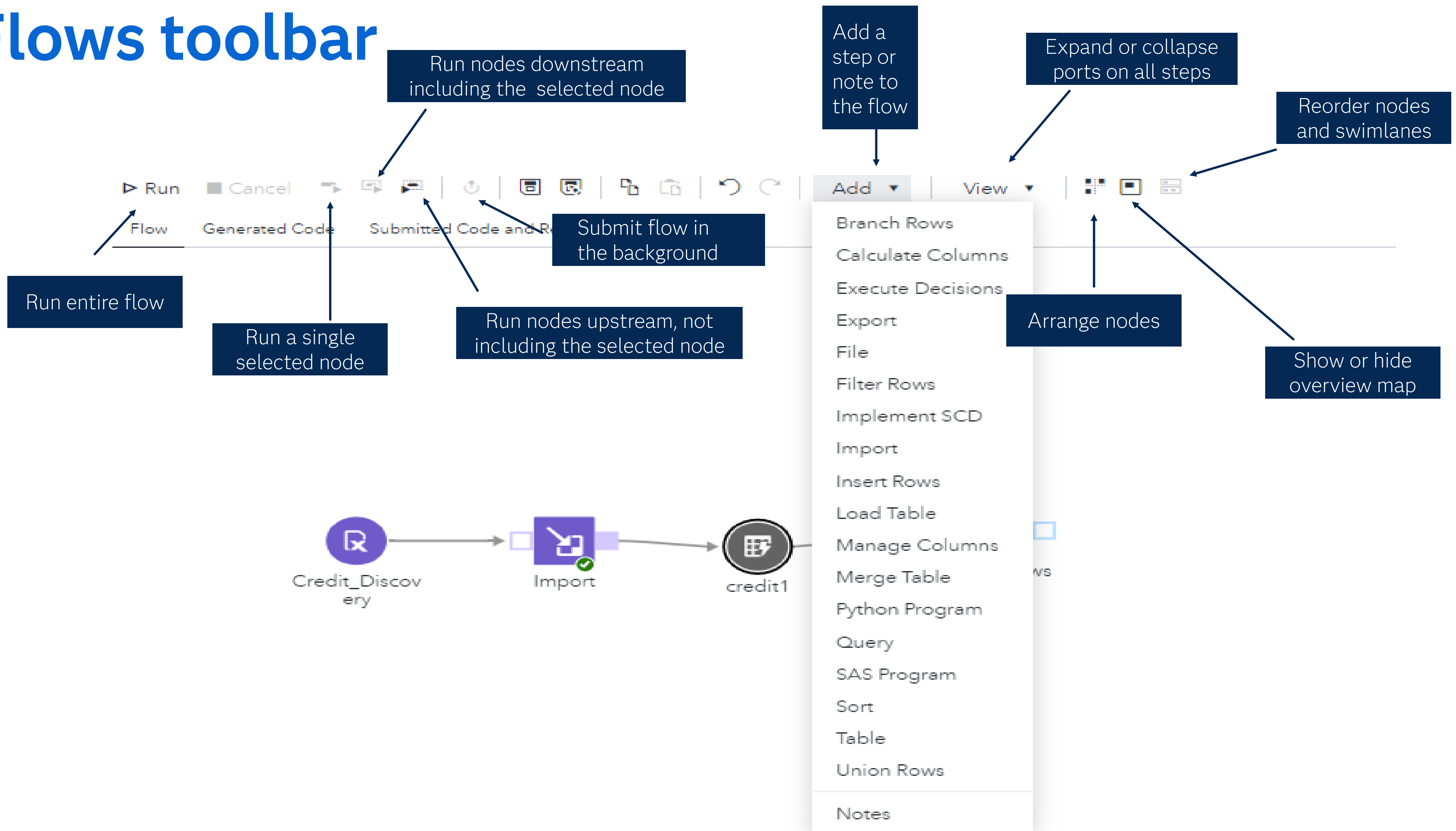
Submission order

The screenshot displays the SAS Studio interface for developing code and flows. The main workspace shows a flow diagram with three swimlanes: 'Flow', 'Training_Model', and 'Scoring'. The 'Flow' swimlane contains steps: Credit_Discovery, Import, Calculate Columns, Branch Rows, SAS Program, Manage Columns, Query, Mask Data, CreditPrepp, Partitioning.ctl, CREDITPARTI, and Bar Chart. The 'Training_Model' swimlane contains: CREDITPARTI, Gradient Boosting.ctl, RESULTS, and Assess.ctl. The 'Scoring' swimlane contains: CREDITPARTI, Filter Rows, and another Assess.ctl step. A 'Steps' panel on the left lists various SAS actions like 'Data (Input and Output)', 'Develop', 'Transform Data', 'Integrate', 'Statistical Process Control', and 'Visualize Data'. A 'Flow properties' panel on the right shows the 'Submission Order' table. A 'Preview tab section' at the bottom displays a table of columns for a 'Query' node.

Order	Swimlane Name
1	DataPrepp
2	Training_Model
3	Scoring

Table	Source	Name	Label	Type	Length	Format	Informat
Calculated	RevenuePercent	RevenuePercent		Character			
t1	acctcd	acctcd	acctcd	Character	1	\$1.	\$1.
t1	address	address	address	Character	33	\$33.	\$33.
t1	age	age	age	Numeric	8	BEST.	
t1	AGE_RNG	AGE_RNG	AGE_RNG	Character	7	\$7.	\$7.
t1	atmuser	atmuser	atmuser	Character	1	\$1.	\$1.
t1	behave_scr	behave_scr	behave_scr	Numeric	8	BEST.	
t1	CARD_TYPE	CARD_TYPE	CARD_TYPE	Character	4	\$4.	\$4.
t1	City	City	City	Character	13	\$13.	\$13.

Flows toolbar

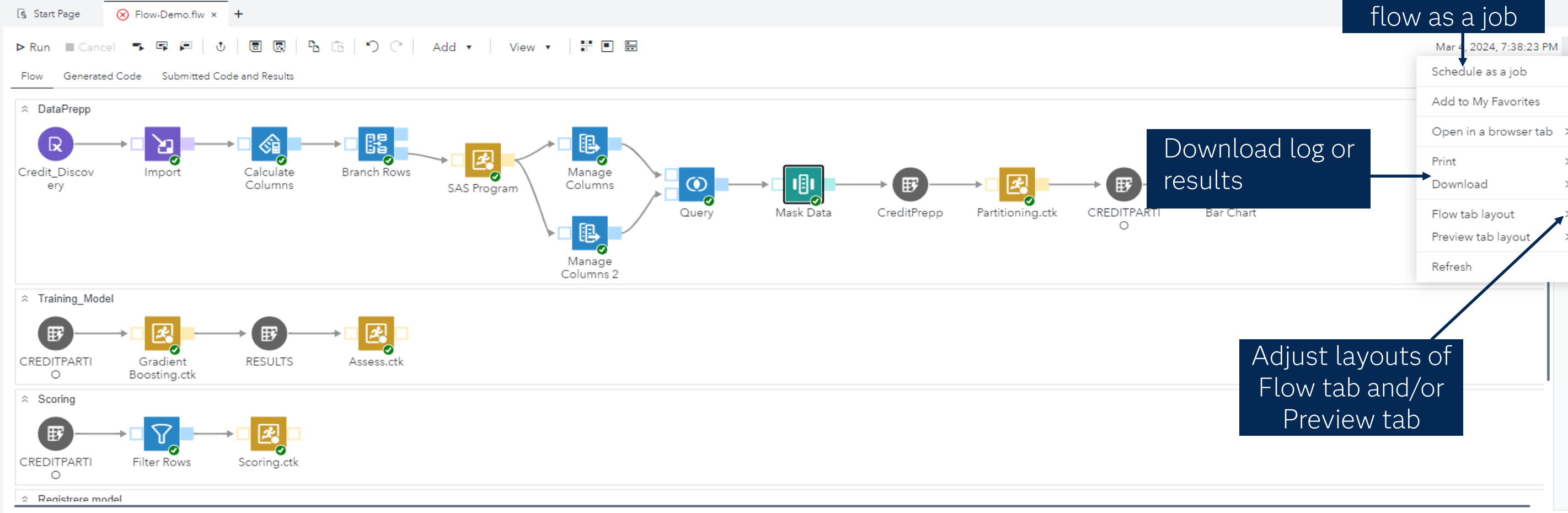


Steps

Type to filter list

SAS Steps : Shared

- Data (Input and Output)
 - Export
 - File
 - Import
 - Table
- Develop
 - Python Program
 - SAS Program
- Transform Data
 - Branch Rows
 - Calculate Columns
 - Filter Rows
 - Insert Rows
 - Manage Columns
 - Mask Data
 - Query
 - Rank Data
 - Remove Duplicates
 - Select Random Sample
 - Sort
 - Split Columns
 - Stack Columns
 - Transpose Data
 - Union Rows
- Integrate
 - Execute Decisions
 - Implement SCD
 - Load Table
 - Merge Table
- Statistical Process Control
 - Analysis of Means
 - Control Charts
 - Pareto Analysis
- Visualize Data
 - Bar Chart
 - Bar-Line Chart
 - Box Plot
 - Bubble Map
 - Bubble Plot
 - Choropleth Map
 - Heat Map
 - Histogram



Schedule the flow as a job

Mar 4, 2024, 7:38:23 PM

- Schedule as a job
- Add to My Favorites
- Open in a browser tab
- Print
- Download
- Flow tab layout
- Preview tab layout
- Refresh

Download log or results

Adjust layouts of Flow tab and/or Preview tab

Mask Data

Data Obfuscation Options Output Node Notes

Obfuscation method:

- Masking (QKB locale is required.)
- Hashing
- Substitution

Data masking definition: *

Mask All Characters

Mask column: *

City

Output Options

By default, the masked column replaces the input column. If a new column is created, the default name for the masked column is in the format ColumnName_Masked.

Column options:

- Replace existing column
- Create new column

Additional actions, the availability varies by node

Refresh

View tabs

SAS steps to put in a flow – just drag steps into the flow tab section

Steps



Type to filter list

SAS Steps : Shared

Data (Input and Output)

- Export
- File
- Import
- Table

Develop

- Python Program
- SAS Program

Transform Data

- Branch Rows
- Calculate Columns
- Filter Rows
- Insert Rows
- Manage Columns
- Mask Data
- Query
- Rank Data
- Remove Duplicates
- Select Random Sample
- Sort
- Split Columns
- Stack Columns
- Transpose Data
- Union Rows

Integrate

- Execute Decisions
- Implement SCD
- Load Table
- Merge Table

Statistical Process Control

- Analysis of Means
- Control Charts
- Pareto Analysis

Visualize Data

- Bar Chart
- Bar-Line Chart
- Box Plot
- Bubble Map
- Bubble Plot
- Choropleth Map
- Heat Map
- Histogram
- Line Chart
- Pie Chart
- Scatter Map
- Scatter Plot
- Series Plot
- Text Map

Machine Learning

- Moving Window Principal Component Analysis

Statistics

- Coin Toss Simulation
- Combinations
- Correlation Analysis
- Dice Roll Simulation
- Distribution Analysis
- One-Way Frequencies
- Permutations
- Poker Hand Probability
- Same Birthday Probability
- Summary Statistics
- t Tests
- Table Analysis

Data Quality

- Clean Data
- Match Codes
- Parse Data

Enrichment

- Geocode Data
- Verify & Geocode Addresses - Loqate
- Verify Email Addresses - Loqate
- Verify Phone Numbers - Loqate

Examine Data

- Characterize Data
- Describe Missing Data
- List Data
- List Table Attributes

Manage Models

- Register Python Model
- Register SAS Model



SAS Snippets to put in a flow – just right click and add

Snippets



Type to filter list

SAS Snippets : My Snippets

Standard

Catalogs

- Edit a SOURCE Entry
- List Catalog Entries
- List Catalogs
- Print GRSEG Entry

Data

- DS2 Code
- DS2 Package
- DS2 Thread
- Generate CSV File
- Generate PowerPoint Slide
- Generate XML File
- Import CSV File
- Import XLSX File
- Simulate Linear Regression Data
- Simulate One-Way ANOVA Data

Data Quality

- Apply Schemes
- Calculate Matchcodes
- Case Data
- Cluster (Entity Resolution) - Proc DQMATCH
- Extract Data
- Guess Locale
- Identify Gender
- Match Action Clustering
- OPTNET Clustering
- Parse Data
- Perform Identification Analysis
- RTENG Clustering
- Run Data Profiling - Proc DATAMETRICS
- Standardize Data
- Survivorship

Descriptive

- Custom ODS Output
- PROC SQL

Graph

- Bar Panel
- Box Panel
- Comparative Scatter Plot
- Dot Plot
- Fit Plot
- HBar Plot
- HighLow Plot
- Histogram Plot
- Scatter Plot Matrix
- VBox Plot

IML

Macro

Viya Foundation

- Cloud Analytic Services
 - Create CAS Connection
 - Delete caslib
 - Delete Table or File from caslib
 - Disconnect CAS Session
 - Generate SAS librefs for caslibs
 - List CAS Session Options
 - List CAS Sessions for SAS Client
 - List CAS Sessions for User ID
 - Load Data to caslib
 - New CAS Session
 - New caslib for Path
 - Reconnect CAS Session
 - Save Table to caslib
 - Terminate CAS Session

Image Processing

- Convert Color
- Display Image
- Load Images
- Mutate Images
- Rescale Images
- Resize Images
- Save Images

Machine Learning

- Compare Several ML Algorithms
- Compare Two ML Algorithms
- Generalized Linear Models
- Load Data
- Prepare and Explore Data
- Supervised Learning
- Unsupervised Learning

Task to put in a flow – fill in information before adding tasks to a flow

Tasks



Type to filter list

SAS Tasks : My Tasks

- Standard
 - Econometrics
 - Aggregate Loss Models
 - Causal Models
 - Cross-sectional Data Models
 - Multivariate Time Series Analysis
 - Panel Data Models
 - Severity Models
 - Spatial Regression Models
 - Univariate Time Series Analysis
 - Forecasting
 - Modeling and Forecasting
 - Time Series Data Preparation
 - Time Series Exploration
 - Optimization and Network Analysis
 - Biconnected Components
 - Connected Components
 - Cycle Detection
 - Linear Assignment
 - Maximal Cliques
 - Minimum Cost Network Flow
 - Minimum Cut
 - Minimum Spanning Tree
 - Shortest Paths
 - Transitive Closure
 - Traveling Salesman Problem
 - Prepare Data
 - Examine Data
 - Transform Data
 - Statistical Process Control
 - Analysis of Means
 - Capability Analysis
 - Control Charts
 - Pareto Analysis
 - Statistics
 - Cluster Analysis
 - Combinatorics and Probability
 - Descriptive
 - High-Performance Statistics
 - Linear Models
 - Multivariate Analysis
 - Power and Sample Size
 - Survival Analysis
 - Visualize Data
 - Viya Foundation
 - Cloud Analytic Services
 - CAS Operations
 - Connect to CAS
 - Econometrics
 - Aggregate Loss Models
 - Cross-sectional Data Models
 - Hidden Markov Models
 - Panel Data Models
 - Severity Models
 - Spatial Regression Models
 - Evaluate and Implement Models
 - Assess
 - Register
 - Scoring
 - Forecasting
 - Time Series Mining
 - Time Series Modeling
 - Machine Learning
 - Automated Machine Learning
 - Automated Feature Engineering
 - Computer Vision
 - Load Images
 - Semi-supervised Learning
 - Semi-supervised Learning
 - Supervised Learning
 - Unsupervised Learning
 - Optimization and Network Analysis
 - Biconnected Components
 - Centrality Metrics
 - Community Detection
 - Connected Components
 - Core Decomposition
 - Cycle Detection
 - Linear Assignment
 - Maximal Cliques
 - Minimum Cost Network Flow
 - Minimum Cut
 - Minimum Spanning Tree
 - Reach Network
 - Shortest Paths
 - Summary Statistics
 - Transitive Closure
 - Traveling Salesman Problem
 - Prepare and Explore Data
 - Binning
 - Imputation
 - Partitioning
 - Sampling
 - Summary
 - Transform Data
 - Variable Selection
 - Statistics
 - Clustering
 - Decision Tree
 - Generalized Linear Models
 - Linear Regression
 - Logistic Regression
 - Partial Least Squares Regression
 - Principal Component Analysis
 - Quantile Regression
 - Text Analytics
 - Boolean Rules
 - Segmentation
 - Text Parsing and Topic Discovery
 - Text Scoring
 - Text Summarization

The demo-data

- A data set from a credit company, where over 40% of customer accounts can be written off.
- If the company does nothing, they could lose over \$240 million.
- The dataset consists of 44 columns and 10,000 rows.