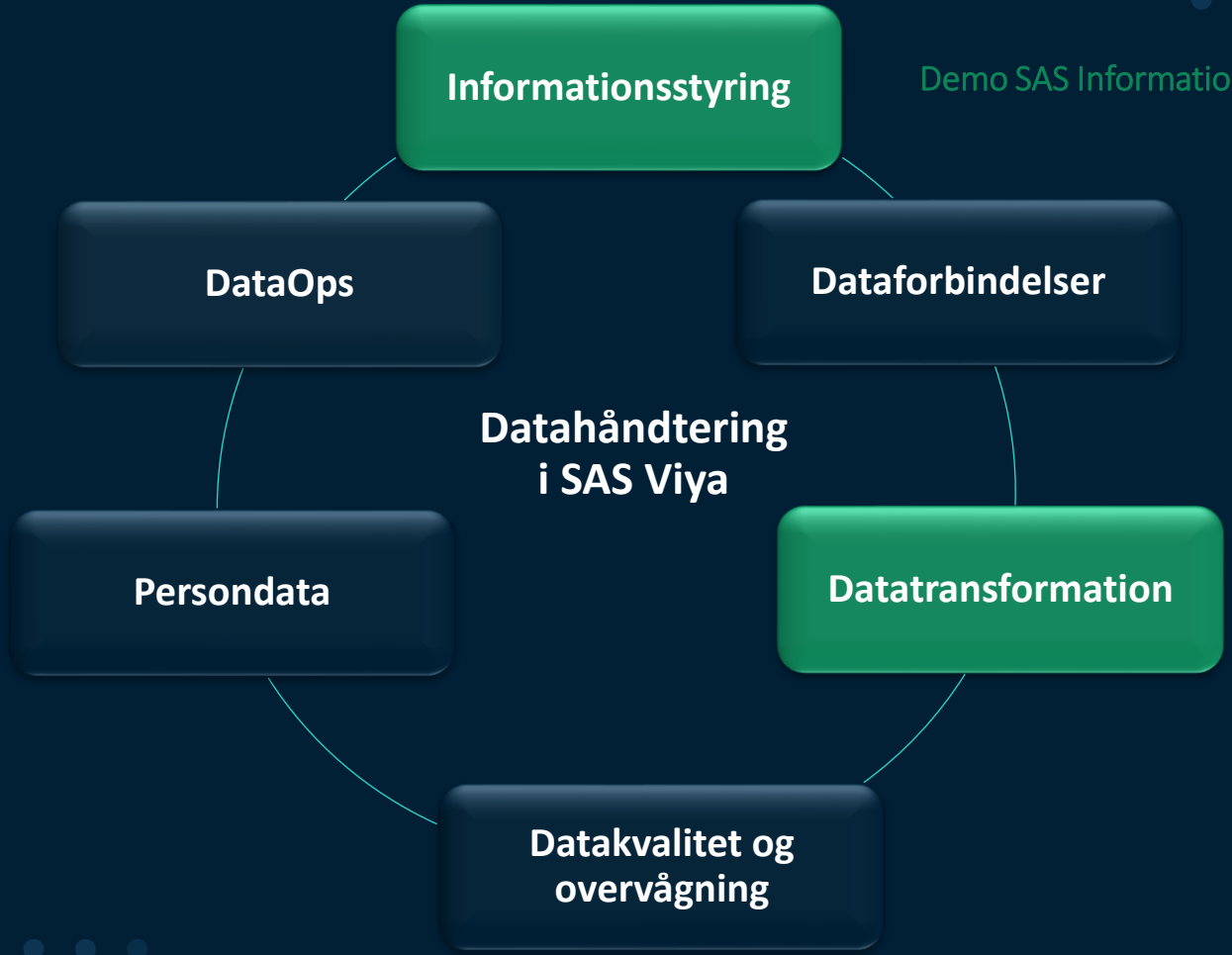
SAS FANS
Data Management

Cecily Hoffritz

Data engineering advisory, SAS Nordics

2. december 2021

30 years employed @ sas

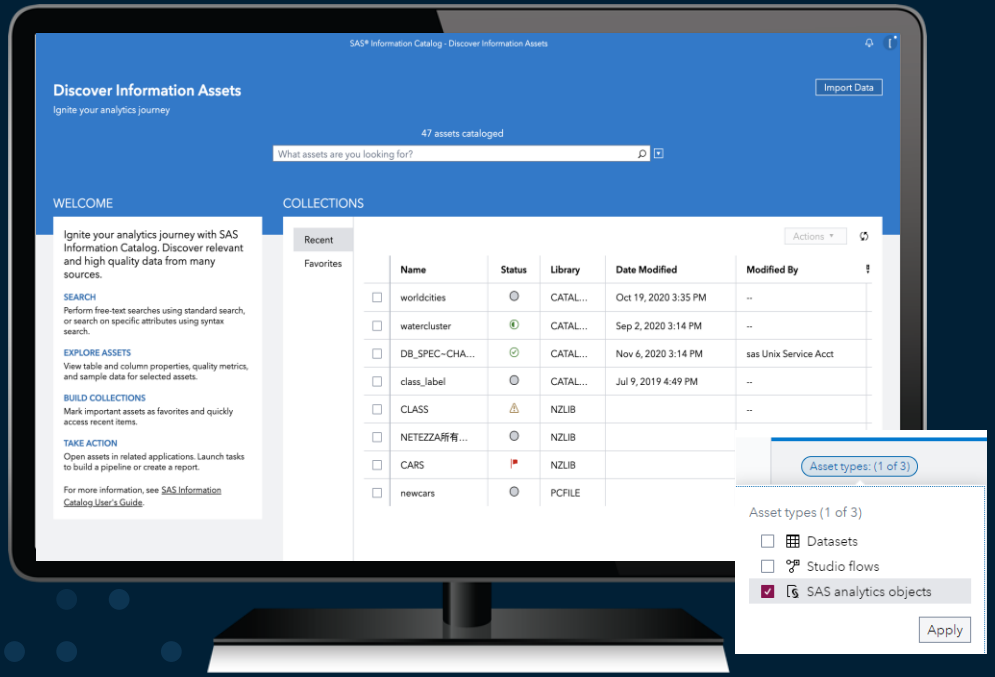


Demo SAS Information Catalog applikation

Demo SAS Studio &
Python

SAS Information Governance

Styring over et broget datalandskab med et informationskatalog



Et informationskatalog gør det muligt

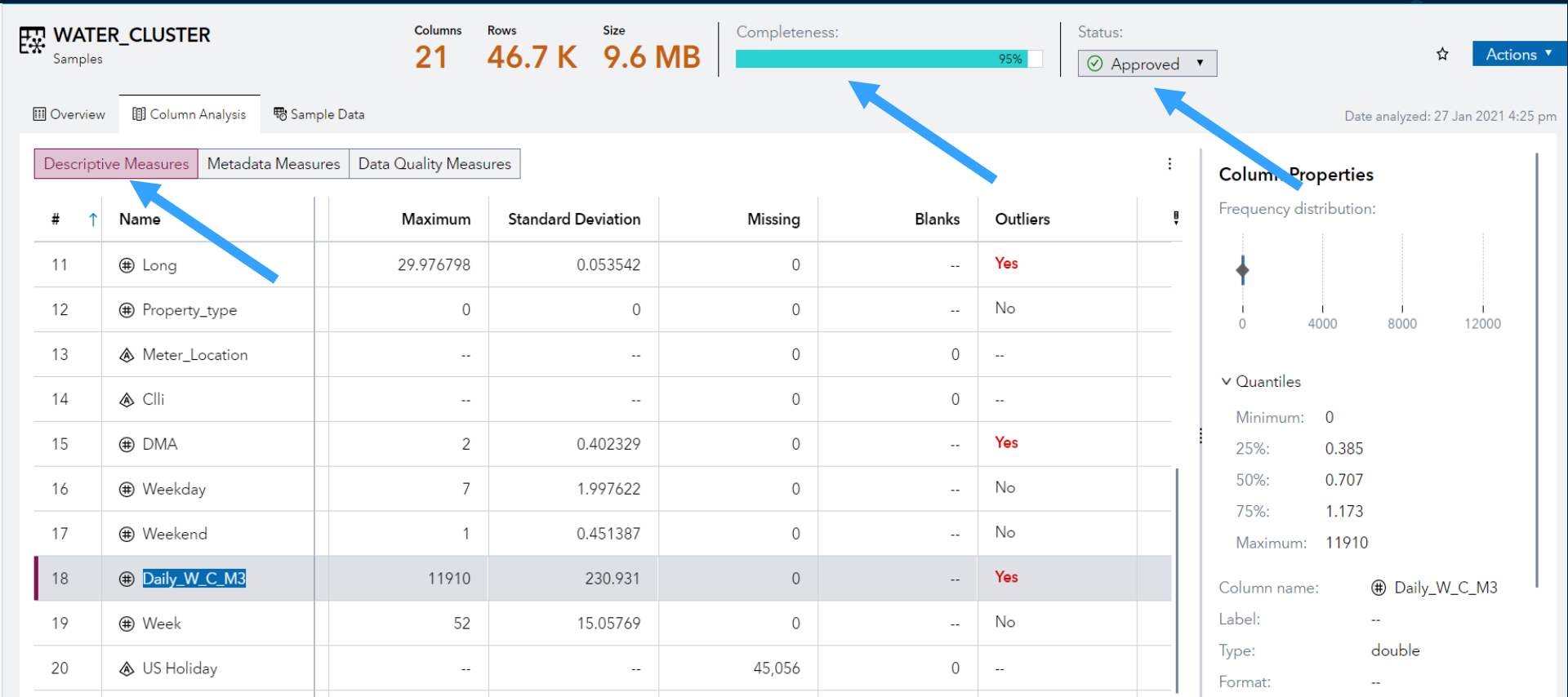
- At søge efter data og undersøge deres analyseparathed.
- At søge efter analyseaktiver med henblik på genbrug og læring

Hvorfor et informationskatalog?

- Tidsbesparende - afkorter processen fra data til insigt
- Øget datakvalitet og forbedret dataanalyse
- Øget samarbejde mellem teams – bedre data er et fællesanliggende
- Bedre styring på data- og analyseaktiverne
- Øget tillid til resultaterne – man kan stole på data

Informationsstyring med SAS Information Governance

Er data analyseparate?



Informationsstyring med SAS Information Governance

Er data analyseparate?

WATER_CLUSTER
Samples

Columns	Rows	Size
21	46.7 K	9.6 MB

Completeness: 95%

Status: Approved

Actions

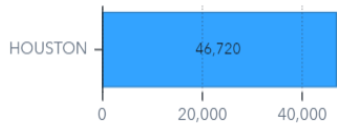
Overview | Column Analysis | Sample Data

Date analyzed: 27 Jan 2021 4:25 pm

Descriptive Measures | Metadata Measures | Data Quality Measures

#	Name	Label	Type	Actual Type	Logical Type	Format	Le
2	Month	Month	double	--	Nominal		
3	Day		double	--	Interval		
4	Date		double	--	Interval	MMDDYY	
5	Serial		double	--	Interval	BEST	
6	Property		double	--	Interval	BEST	
7	Address		char	String	Nominal	$CHAR	
8	City		char	String	Unary	$CHAR	
9	Zip		double	--	Interval	BEST	
10	Lat		double	--	Interval	BEST	
11	Long		double	--	Interval	BEST	

Frequency distribution:



Column name:	City
Label:	--
Type:	char
Actual type:	String
Format:	$CHAR
Length:	7
Minimum length:	7
Maximum length:	7
Primary key candidate:	No
Logical type:	Unary
Semantic type:	CITY
Information privacy:	Candidate

Informationsstyring med SAS Information Governance

Er data analyseparate?

WATER_CLUSTER
Samples

Columns	Rows	Size
21	46.7 K	9.6 MB

Completeness: 95%

Status: Approved

Actions

Overview | Column Analysis | Sample Data

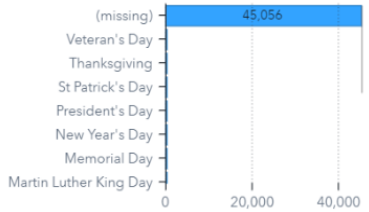
Date analyzed: 27 Jan 2021 4:25 pm

Descriptive Measures | Metadata Measures | Data Quality Measures

#	Name	Completeness	Uniqueness	Most Common Value	Least Common Value	Pattern Count	Semantic
9	Zip	100%	0%	77094	77020 (2 more)	--	POSTAL C
10	Lat	100%	0%	-95.408872 (1 more)	-95.71878 (19+ more)	--	LATITUDE
11	Long	100%	0%	29.785747 (1 more)	29.651167 (19+ more)	--	LONGITU
12	Property_type	100%	0%	0	0	--	GENERIC
13	Meter_Location	100%	0%	external	internal	1	
14	Clli	100%	0%	HSTNTXJA	HSTNTXHO	1	
15	DMA	100%	0%	2	1	--	
16	Weekday	100%	0%	4 (1 more)	1	--	
17	Weekend	100%	0%	0	1	--	
18	Daily_W_C_M3	100%	11%	0	0.023 (19+ more)	--	
19	Week	100%	0%	51 (19+ more)	52	--	DATE
20	US Holiday	4%	0%	Veteran's Day (11 m...	Independence Day	13	
21	CLUSTER	100%	0%	4	1	--	

Column Properties

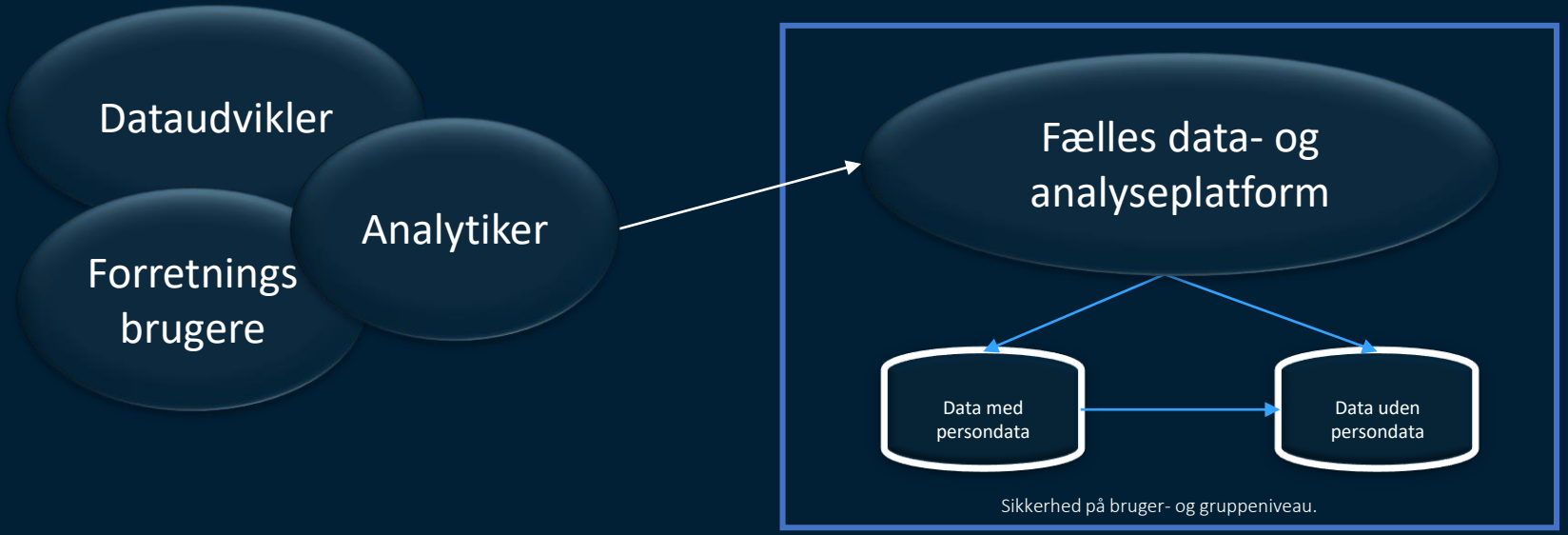
Frequency distribution:



Column name:	US Holiday
Label:	--
Type:	char
Actual type:	String
Format:	$CHAR
Length:	27
Minimum length:	8
Maximum length:	27
Primary key candidate:	No
Logical type:	Nominal
Semantic type:	--
Information privacy:	--

Informationsstyring med SAS Information Governance

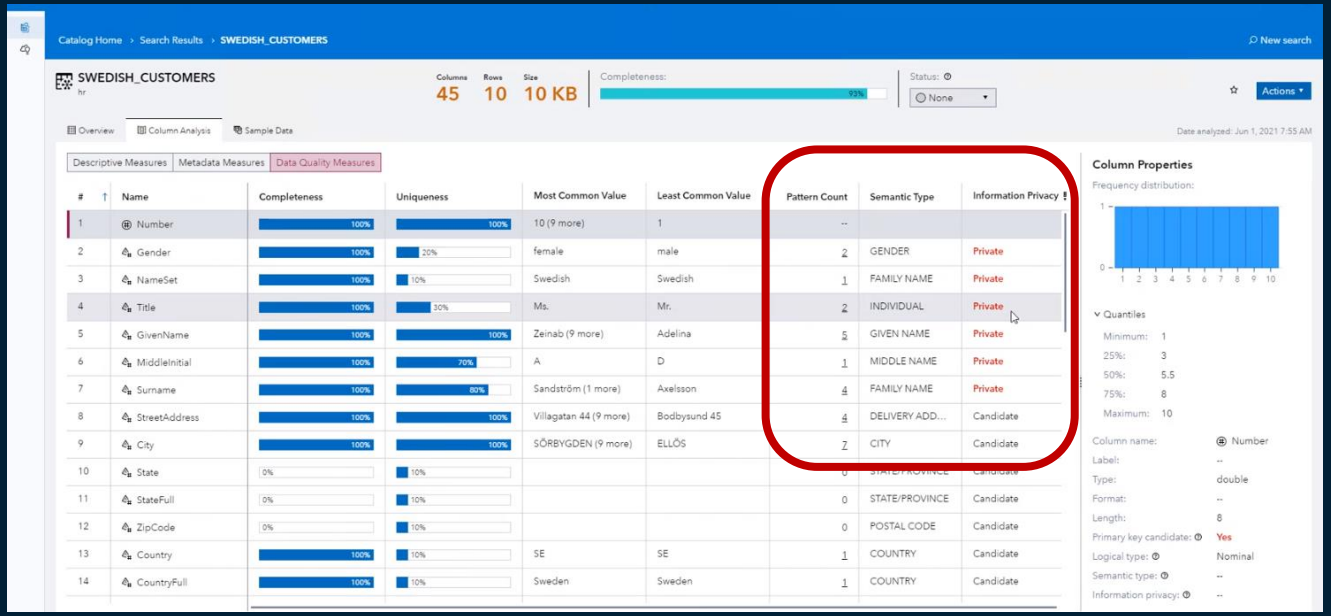
Use case: håndtering af persondata



- Sporbarhed gennem hele processen, fra data til analyse.
- Logning af aktivitet på data - hvilke data kigger man specifikt på og hvornår.

Informationsstyring med SAS Information Governance

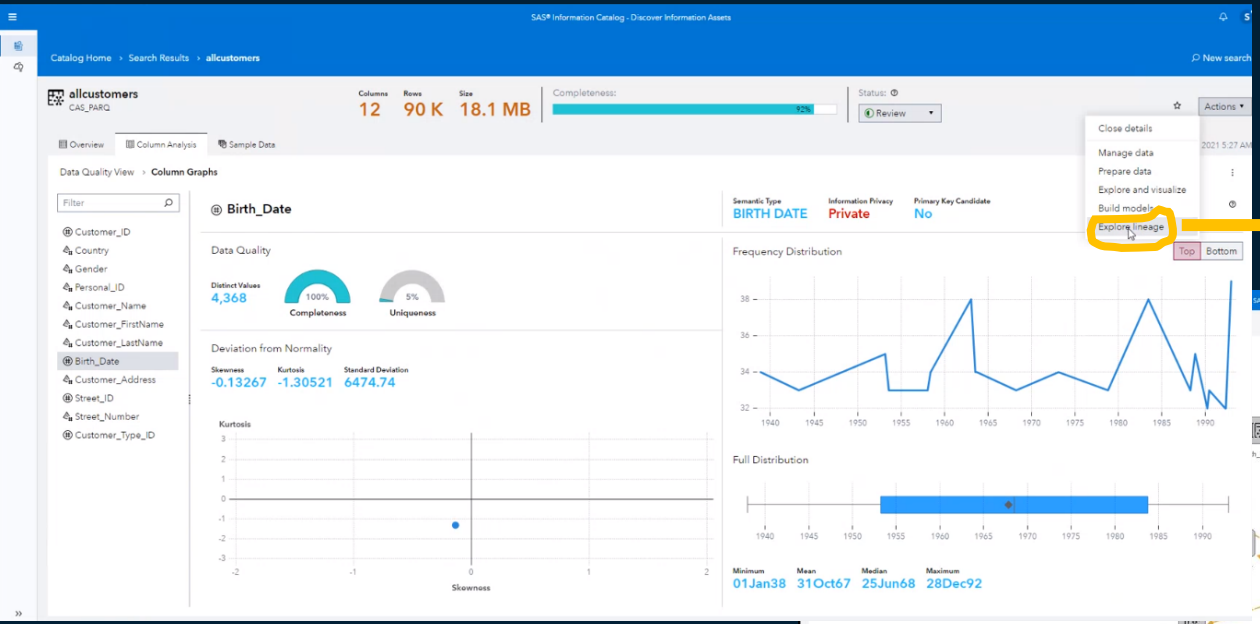
Håndtering af persondata med SAS



Trin 1- Overblik:
Få et straks overblik over evt. persondata gennem den automatiske persondata-scanning

Informationsstyring med SAS Information Governance

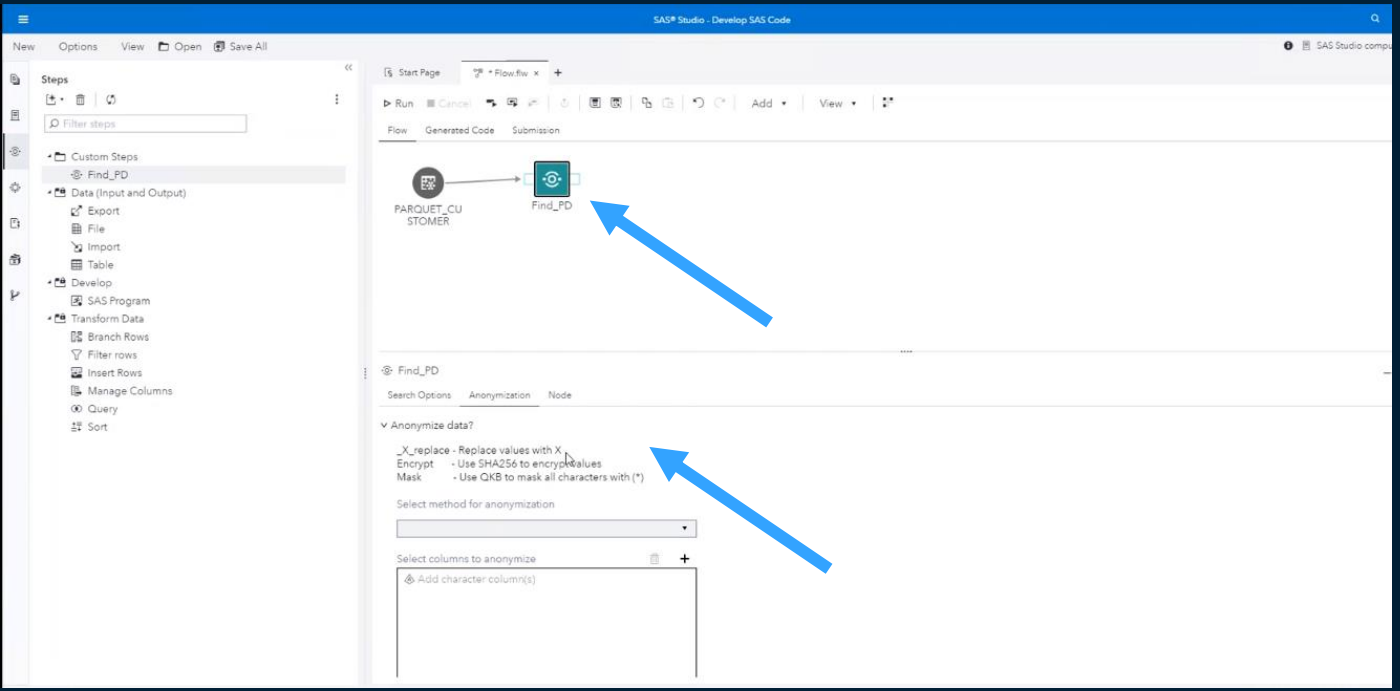
Håndtering af persondata med SAS



Trin 2 - Sporbarhed:
Skab overblik over de sammenhænge, hvor data indgår.

Informationsstyring med SAS Information Governance

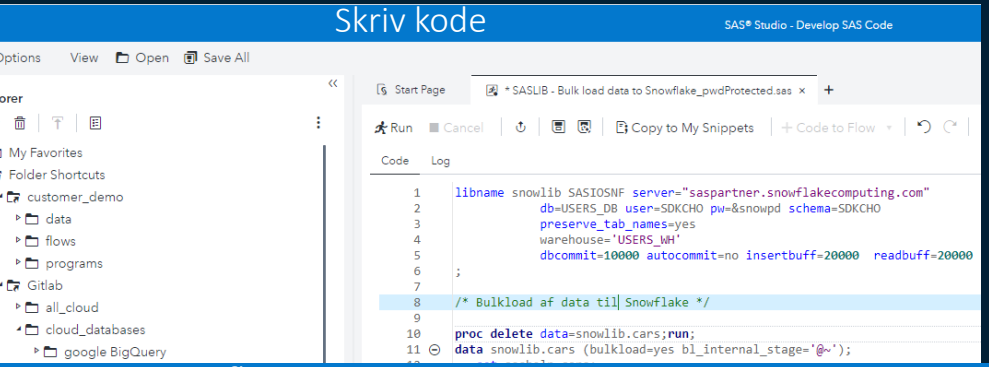
Håndtering af persondata med SAS

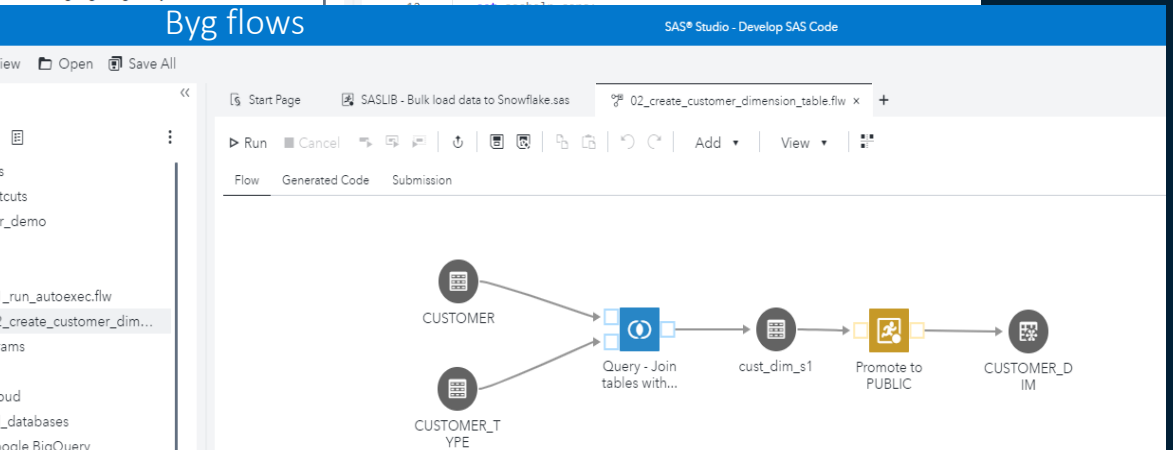


Trin 3 - Beskyttelse:
Skab anonymiserede data til analysen.

Datatransformation

Gør data klar til analyse med SAS Studio



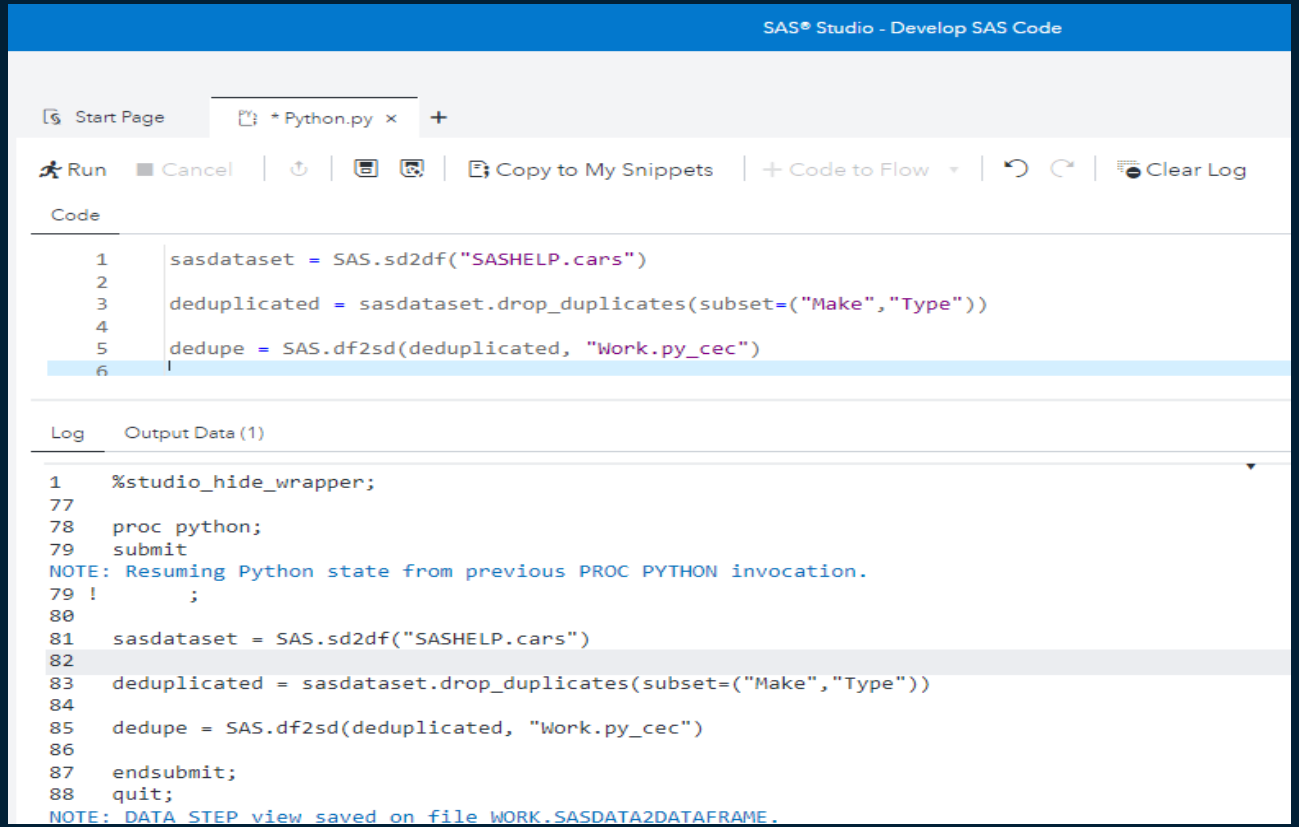


- SAS Studio er én applikation med et væld af funktionalitet til datatransformation, dataintegration og analyse/statistik for alle typer af brugere

Agilitet
Samarbejde
Integration

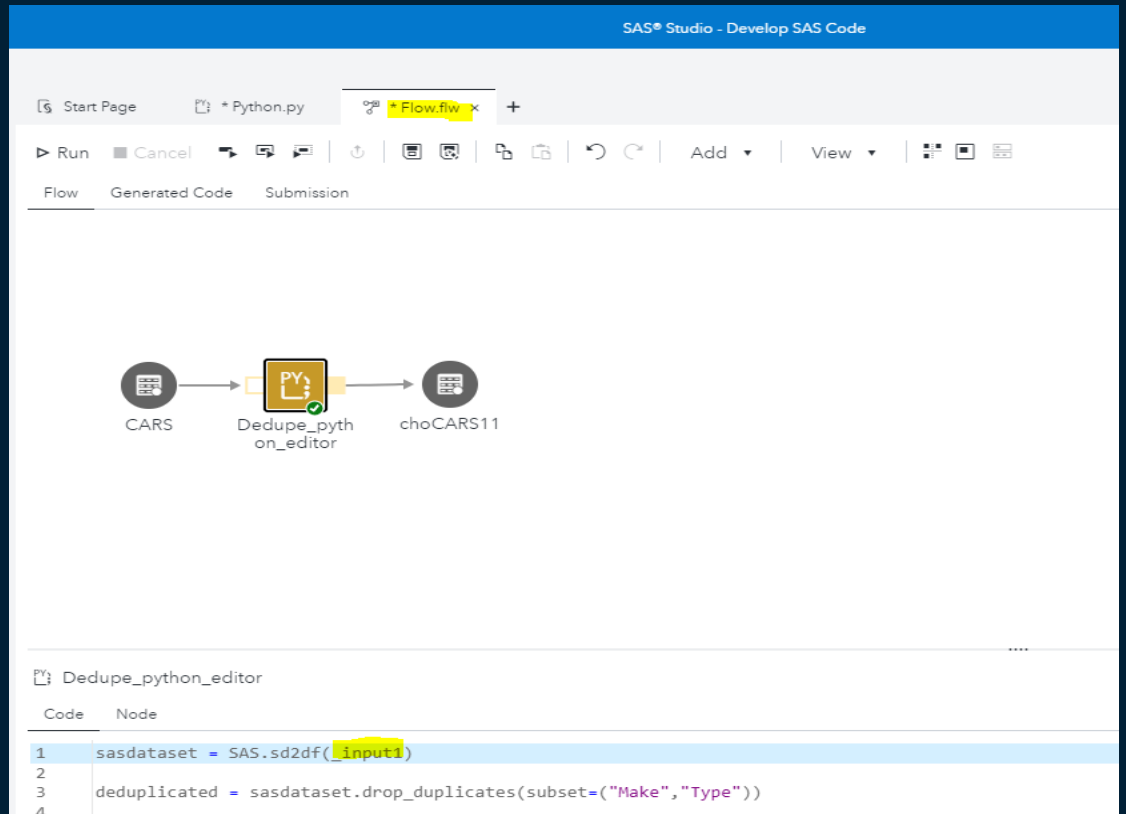
SAS Studio & Python

Kør Python kode i Python.py editor i SAS Studio



SAS Studio & Python

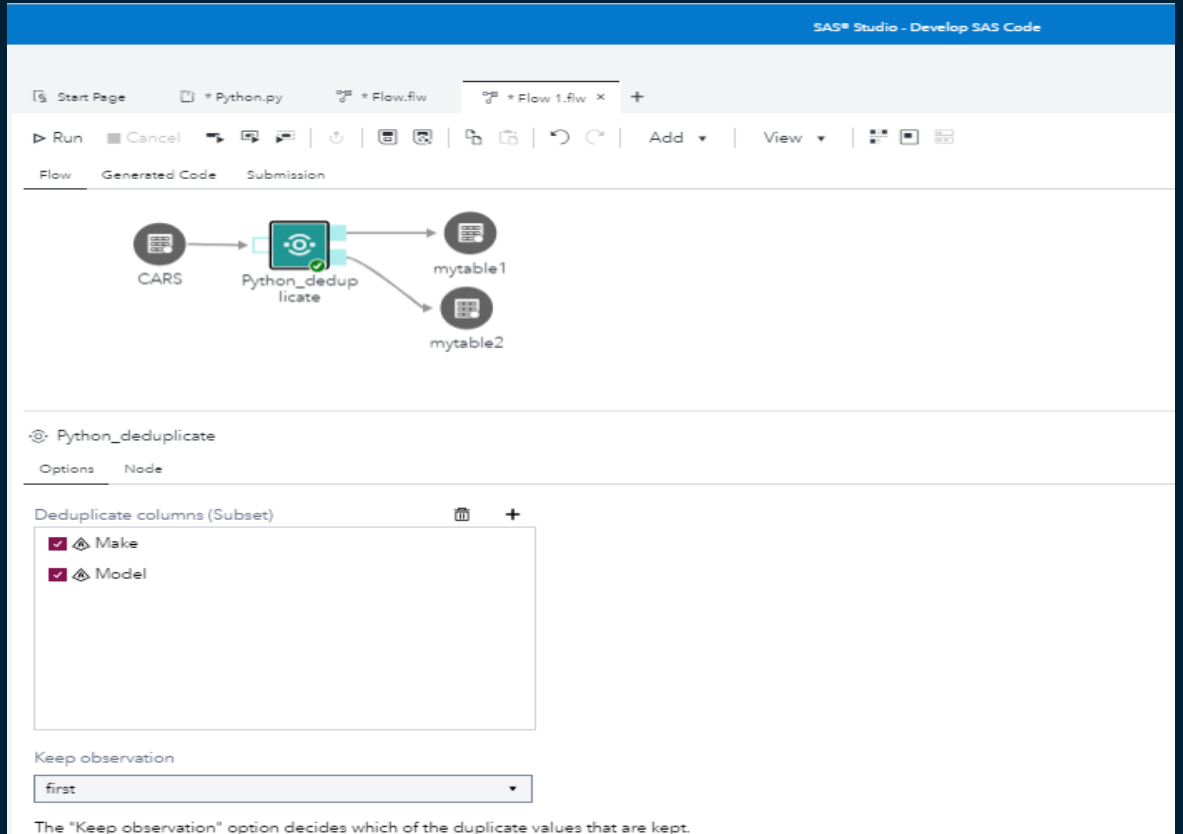
Integrér Python i en SAS Studio data pipeline



SAS Studio & Python

Gør Python super nemt at anvende for SAS Studio brugere

Kodefrit



Migrering af data til UTF-8 for SAS Viya

- Migrér til UTF-8 encoding links:
  - Determine the Encoding of a Data Set.
  - Migrating Data from WLATIN1 to UTF-8.
  - Determine Storage Size Requirements.
  - Use CEDA to Read Data.
  - Determine Whether the CVP Engine Is Needed to Read Your Data without Truncation.
  - Convert Indexes and Integrity Constraints to UTF-8.
  - Convert Format Catalogs to UTF-8.
  - Read External Files.

Læs mere om SAS og UTF8

Læs mere om session encoding og dataafkortning

Tak

Cecily Hoffritz | LinkedIn