MODUL 8

SELF-SERVICE DATA PREPARATION IN SAS® VIYA®

Introduction to SAS Data Studio and Plans

THEME DESCRIPTION

Students are able to do transform data into the form required for analytics, in accordance with the SAS Data Preparation methodology address transforming the data: Prepare and Cleanse. In order to standardize your data to ensure consistency, parse data, or group data to create one best record for Analytics.

WEEKLY LEARNING OUTCOMES (SUB-LESSONS)

CLO-3-Sub-CLO-8, Able to effectively implement the process and role of data curation in data management through the use of SAS® Viya cloud technology– C6.

Through the following learning steps:

1. SAS Data Explorer: Explore and load the CLIENT_INFO table
2. SAS Data Studio: Using Suggestions to Prepare Data
3. Add and configure the Remove duplicates suggestion
4. Add the Parsing CUSTOMER_NAME suggestion
5. SAS Drive: Saved Plan overview

PRACTICUM SUPPORT

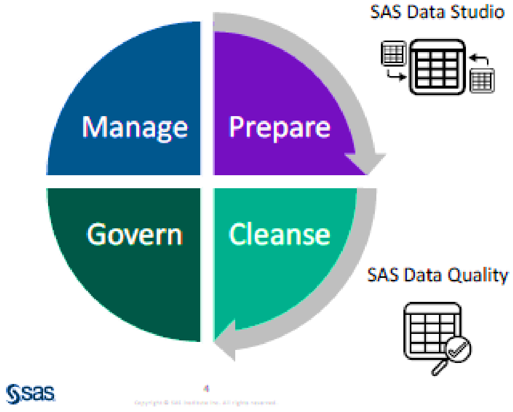
c. Windows Operating System

d. (any) Browser Application

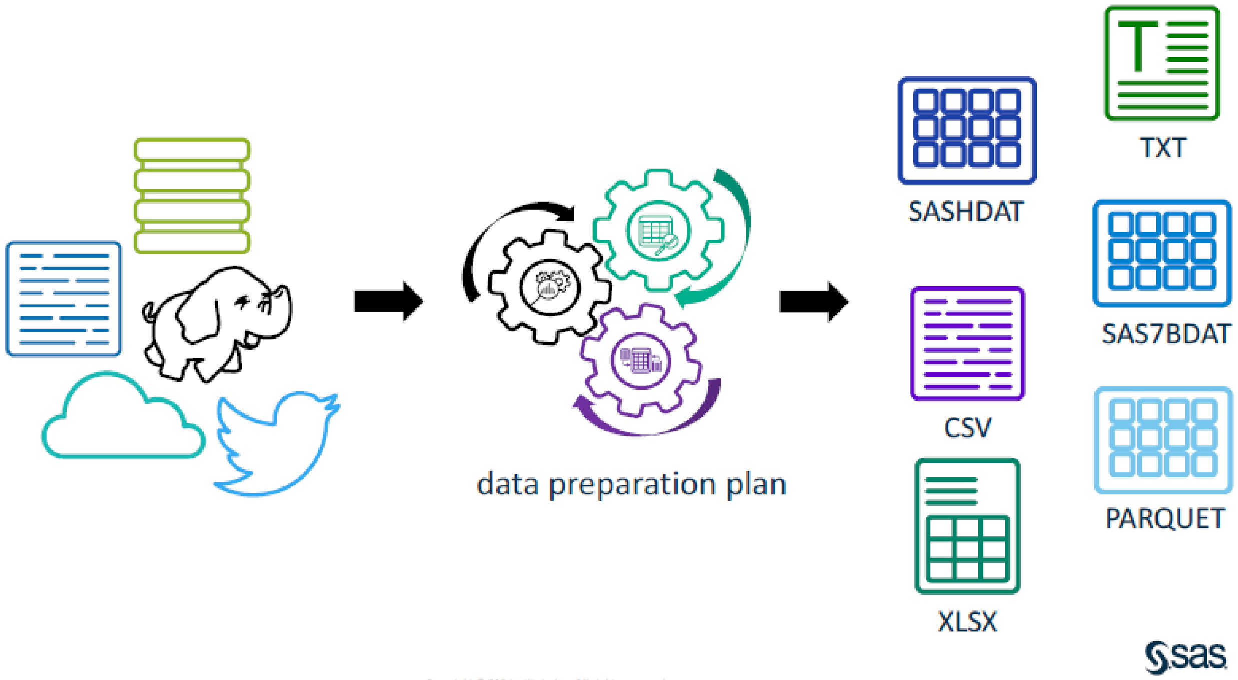
PRACTICUM GUIDELINES

Exploring the SAS Data Preparation Applications

- Our IS-429 BDA practicum class today illustrates exploring the SAS Self-Service Data Preparation applications in the SAS Viya environment.



- The activities, and practice are about investigating the environment and SAS Data Preparation applications:
  - SAS Drive
  - SAS Data Explorer
  - SAS Data Studio
  - SAS Data Quality
- A **data preparation plan** is a collection of transforms that are performed on one or more tables that **prepare data for analysis and reporting**.

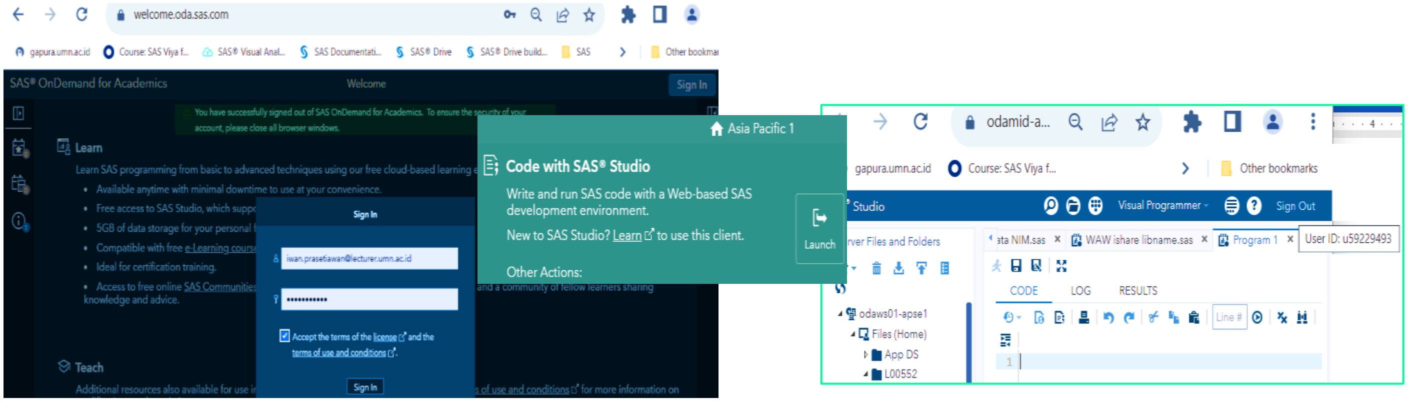


- You **create plans in SAS Data Studio**, and you can save them to a SAS folder location where other users can also access them.

PRACTICUM STEPS

Login to SAS Studio

Start, https://welcome.oda.sas.com



You have successfully signed out of SAS OnDemand for Academics. To ensure the security of your account, please close all browser windows.

Learn SAS programming from basic to advanced techniques using our free cloud-based learning e

- Available anytime with minimal downtime to use at your convenience.
- Free access to SAS Studio, which suppc
- 5GB of data storage for your personal f
- Compatible with free e-Learning cours
- Ideal for certification training.
- Access to free online SAS Communities knowledge and advice.

and a community of fellow learners sharing

Additional resources also available for use i

s of use and conditions for more information on

Sign In

iwan.prasetiawan@lecturer.umn.ac.id

••••••••••••

☑ Accept the terms of the license and the terms of use and conditions.

Sign In

Code with SAS® Studio

Write and run SAS code with a Web-based SAS development environment.

New to SAS Studio? Learn to use this client.

Other Actions:

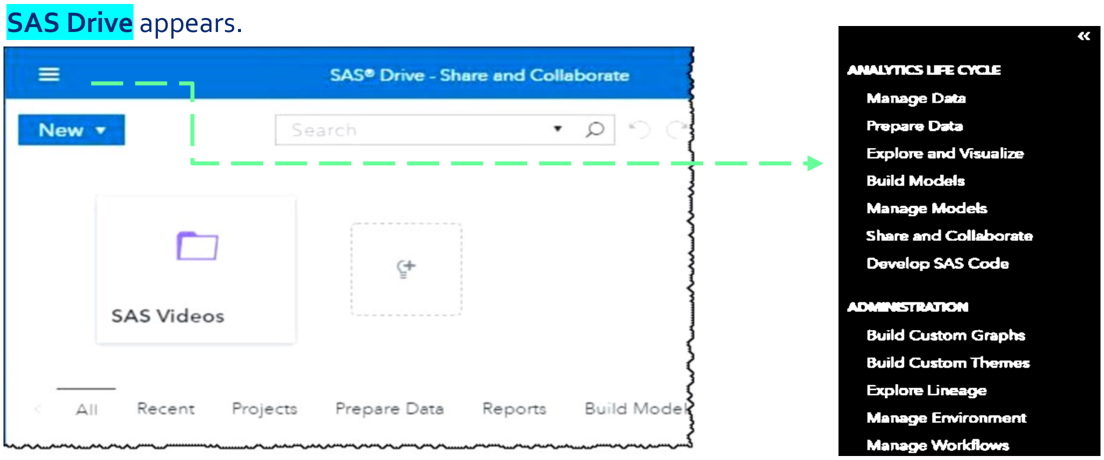
User ID: u59229493

Open SAS Drive Explorer

1) When you have signed in to SAS OnDemand for Academics, then open the SAS Studio application and open another tab of your browser and click SAS® Drive to go to SAS Drive page.

- *If you have problems, you can try using this url:* https://v4e041.vfe.sas.com/SASDrive/ **or go to** https://vle.sas.com/course/view.php?id=7715 then click "Launch SAS Viya for Learners3.5".

2) SAS Drive appears.



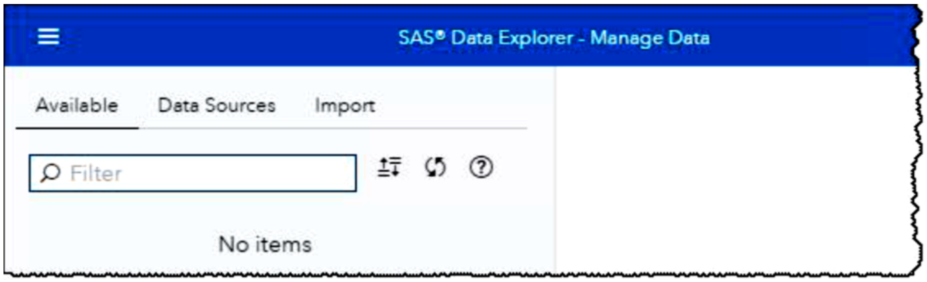
- With SAS Drive, you can manage all your content in one convenient location.
- You can search for an object that was created in a SAS application. You can share, tag, and preview your SAS content. Sharing takes only one click on the object.
- You decide who you want to share it with and whether they can read, edit, or share it.
- Tagging content can help you organize your work. Previewing files can help you find the right one.
- In addition to controlling your SAS content, you can drag-and-drop most content types from your local machine, and SAS Drive will upload it for you.
- For quick access to what you care most about, SAS Drive includes a collapsible area where you can place any content you choose. This paper includes screen shots and illustrative examples showing you how to manage all your SAS content in one place

3) Click ≡ (Show list of applications) to access SAS Viya applications.

4) The applications that are shown are controlled by the SAS products that are licensed and the permissions assigned to the user.

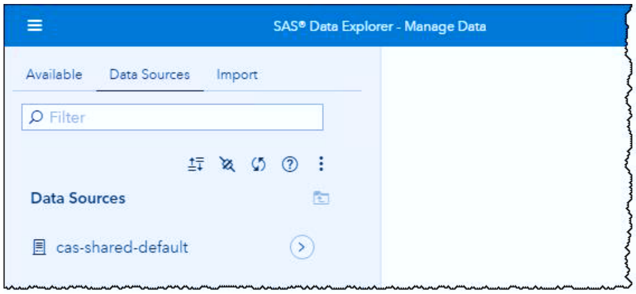
5) The list of actions in the list relate to a specific application that will be used to perform that action. For example, **selecting Manage Data opens SAS Data Explorer** to enable you to work with **data connections and data**.

6) Select **Manage Data** to **open SAS** Data **Explorer**.
SAS Data Explorer appears



1. Explore and load the CLIENT_INFO table.

a. Click the Data Sources tab



SAS® Data Explorer - Manage Data

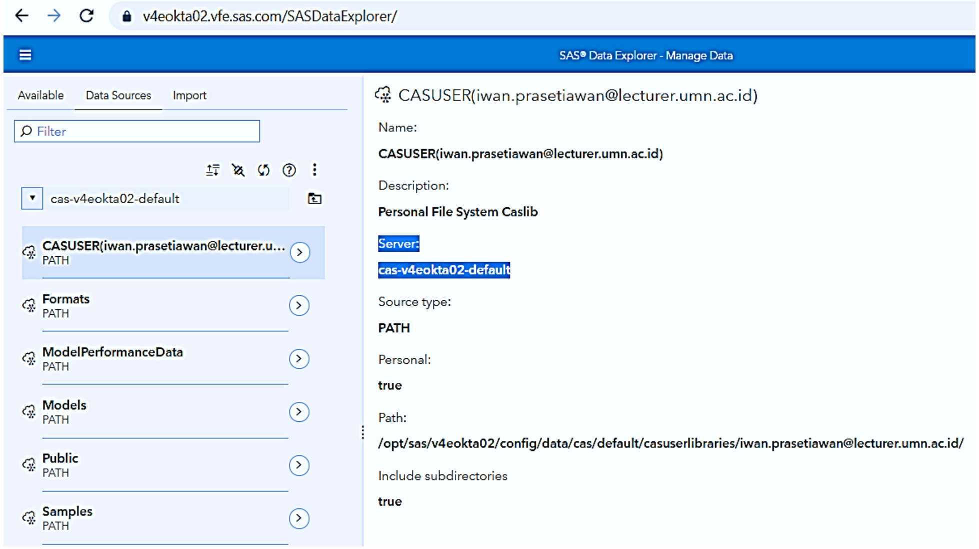
Available | Data Sources | Import

Filter

Data Sources

cas-shared-default

b. Only one CAS server is configured. Click (Down one level) for the **cas-shared-default server** entry to view the list of data sources that are configured on the **CAS server**.
Note: By default, users have their own CASUSER library and the Formats, Public, and Samples caslibs. Additional caslibs might be visible, depending on the SAS products that are licensed.



v4eokta02.vfe.sas.com/SASDataExplorer/

SAS® Data Explorer - Manage Data

Available | Data Sources | Import

Filter

cas-v4eokta02-default

- CASUSER(iwan.prasetiawan@lecturer.u... PATH
- Formats PATH
- ModelPerformanceData PATH
- Models PATH
- Public PATH
- Samples PATH

CASUSER(iwan.prasetiawan@lecturer.umn.ac.id)

Name:
CASUSER(iwan.prasetiawan@lecturer.umn.ac.id)

Description:
Personal File System Caslib

Server:
cas-v4eokta02-default

Source type:
PATH

Personal:
true

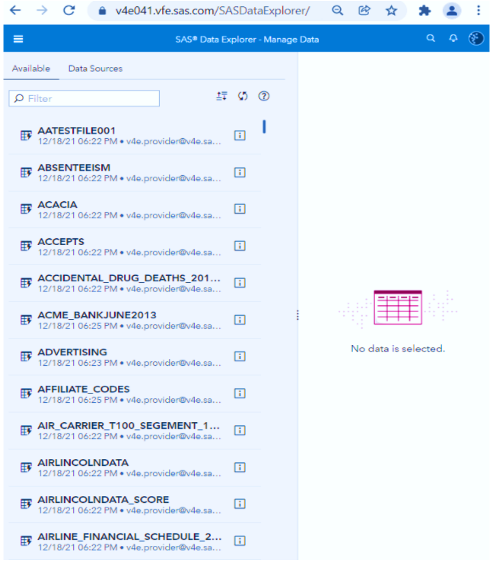
Path:
/opt/sas/v4eokta02/config/data/cas/default/casuserlibraries/iwan.prasetiawan@lecturer.umn.ac.id/

Include subdirectories
true

c. Double-click Public to open the connection and view the data sources. Click CLIENT_INFO.sashdat to view specifics of the table.

- *If you don't find the CLIENT_INFO table in Data Sources then move to Available tab.*

d. Click the Available tab.



v4e041.vfe.sas.com/SASDataExplorer/

SAS® Data Explorer - Manage Data

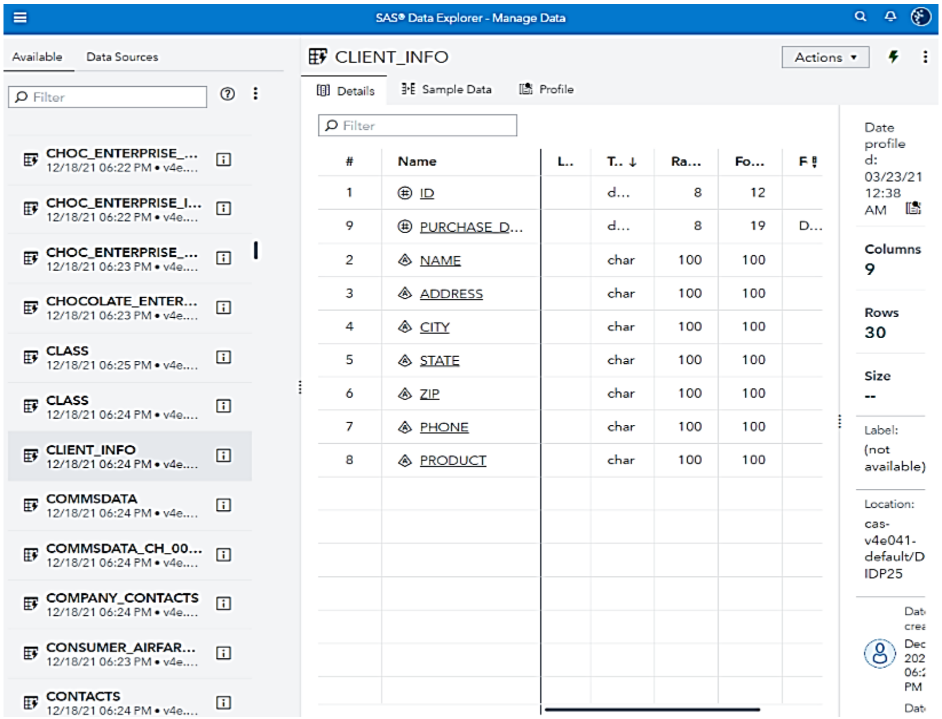
Available | Data Sources

Filter

- AATESTFILE001 12/18/21 06:22 PM • v4e.provider@v4e.sa...
- ABSENTEEISM 12/18/21 06:22 PM • v4e.provider@v4e.sa...
- ACACIA 12/18/21 06:22 PM • v4e.provider@v4e.sa...
- ACCEPTS 12/18/21 06:22 PM • v4e.provider@v4e.sa...
- ACCIDENTAL_DRUG_DEATHS_201... 12/18/21 06:22 PM • v4e.provider@v4e.sa...
- ACME_BANKJUNE2013 12/18/21 06:25 PM • v4e.provider@v4e.sa...
- ADVERTISING 12/18/21 06:23 PM • v4e.provider@v4e.sa...
- AFFILIATE_CODES 12/18/21 06:25 PM • v4e.provider@v4e.sa...
- AIR_CARRIER_T100_SEGEMENT_1... 12/18/21 06:22 PM • v4e.provider@v4e.sa...
- AIRLINCOLNDATA 12/18/21 06:22 PM • v4e.provider@v4e.sa...
- AIRLINCOLNDATA_SCORE 12/18/21 06:22 PM • v4e.provider@v4e.sa...
- AIRLINE_FINANCIAL_SCHEDULE_2... 12/18/21 06:22 PM • v4e.provider@v4e.sa...

No data is selected.

- The Available tab enables you to see all tables that have been loaded into memory from any CAS server to which you have access. Clicked twice the CLIENT_INFO table to selected.



SAS® Data Explorer - Manage Data

Available | Data Sources

- CHOC_ENTERPRISE_... 12/18/21 06:22 PM • v4e....
- CHOC_ENTERPRISE_I... 12/18/21 06:22 PM • v4e....
- CHOC_ENTERPRISE_... 12/18/21 06:23 PM • v4e....
- CHOCOLATE_ENTER... 12/18/21 06:23 PM • v4e....
- CLASS 12/18/21 06:25 PM • v4e....
- CLASS 12/18/21 06:24 PM • v4e....
- CLIENT_INFO 12/18/21 06:24 PM • v4e....
- COMMSDATA 12/18/21 06:24 PM • v4e....
- COMMSDATA_CH_00... 12/18/21 06:24 PM • v4e....
- COMPANY_CONTACTS 12/18/21 06:24 PM • v4e....
- CONSUMER_AIRFAR... 12/18/21 06:23 PM • v4e....
- CONTACTS 12/18/21 06:24 PM • v4e....

CLIENT_INFO

Details | Sample Data | Profile

#	Name	L...	T.. ↓	Ra...	Fo...	F
1	ID		d...	8	12	
9	PURCHASE_D...		d...	8	19	D...
2	NAME		char	100	100	
3	ADDRESS		char	100	100	
4	CITY		char	100	100	
5	STATE		char	100	100	
6	ZIP		char	100	100	
7	PHONE		char	100	100	
8	PRODUCT		char	100	100	

Date profiled: 03/23/21 12:38 AM

Columns 9

Rows 30

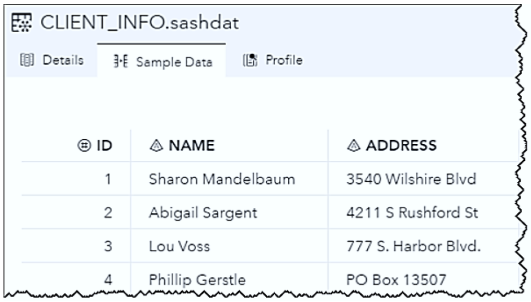
Size --

Label: (not available)

Location: cas-v4e041-default/DIDP25

The Details tab shows metadata information about the table such as the column properties, table size, and number of columns.

e. Click the Sample Data tab.



CLIENT_INFO.sashdat

Details | Sample Data | Profile

ID	NAME	ADDRESS
1	Sharon Mandelbaum	3540 Wilshire Blvd
2	Abigail Sargent	4211 S Rushford St
3	Lou Voss	777 S. Harbor Blvd.
4	Phillip Gerstle	PO Box 13507

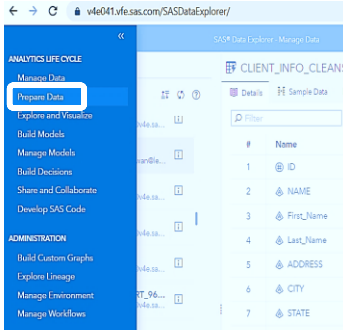
f. A sample data is shown displayed completely for 100 records data sample.

g. **Screenshot your page of the Client_Info sample data as in the output figure A.**

2. Using Suggestions to Prepare Data.

- This demonstration creates a simple plan using the Suggestions feature to quickly add and configure Data Quality transforms.

a. Clicking "**prepare data**" will take you to the **SAS Data Studio tools** for further data preparation activities.



v4e041.vfe.sas.com/SASDataExplorer/

SAS® Data Explorer - Manage Data

ANALYTICS LIFE CYCLE
- Manage Data
- Prepare Data
- Explore and Visualize
- Build Models
- Manage Models
- Build Decisions
- Share and Collaborate
- Develop SAS Code

ADMINISTRATION
- Build Custom Graphs
- Explore Lineage
- Manage Environment
- Manage Workflows

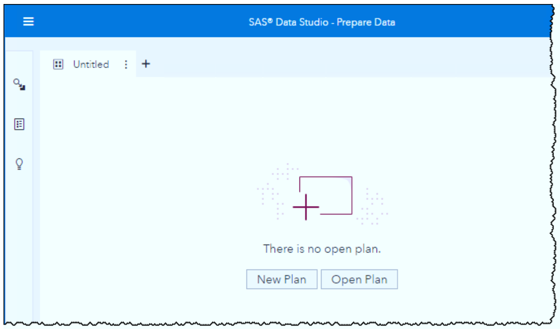
CLIENT_INFO_CLEAN

Details | Sample Data

#	Name
1	ID
2	NAME
3	First_Name
4	Last_Name
5	ADDRESS
6	CITY
7	STATE

b. Click ☰ (Show list of applications) and select Prepare Data to open **SAS Data Studio**.

h. Click New Plan. **SAS Data Studio** appears



There is no open plan.

The Choose Data window appears.

c. Select the **BANK_CUSTOMERS** table **as the data source**. Click the Data Sources tab.

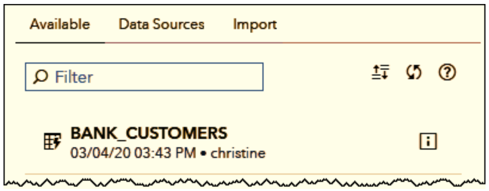
d. Click (Down one level) for the cas-shared-default server.

e. Click (Down one level) for the DIDP25 caslib.

f. Right-click BANK_CUSTOMERS.sashdat and select Load to load the table into memory.

- Ignore it, if the table is already available in CAS memory.
- This step shows what if you prepare a table for loading into CAS memory and will be available at available tab.

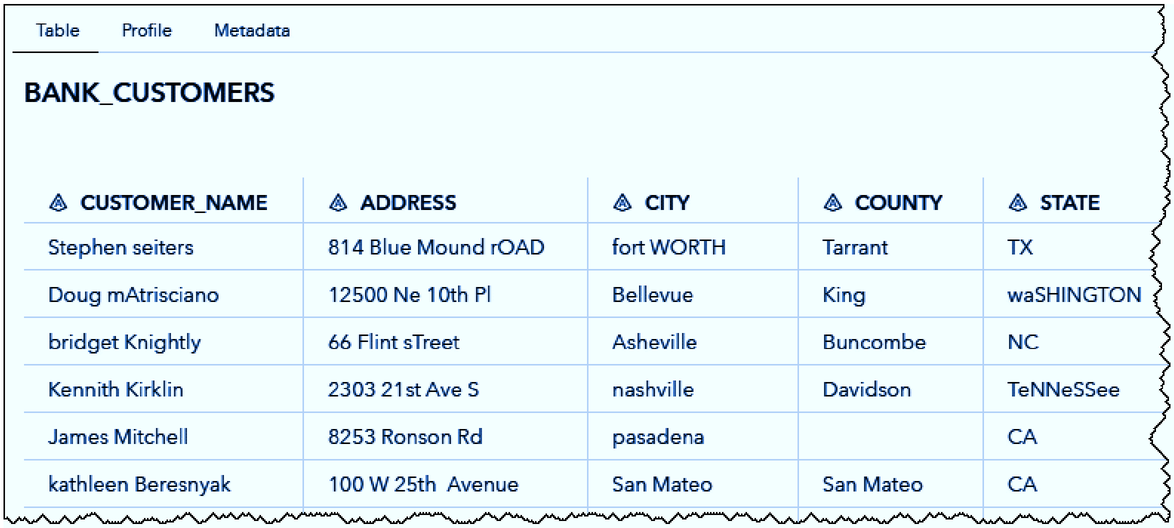
g. Click the Available tab. The BANK_CUSTOMERS table is displayed as an in-memory table.



BANK_CUSTOMERS
03/04/20 03:43 PM • christine

h. Double click the BANK_CUSTOMERS table. You have them now at your SAS Studio workspace.

i. Scroll to the right and expand the CUSTOMER_NAME, ADDRESS, CITY, and STATE columns to review the data.

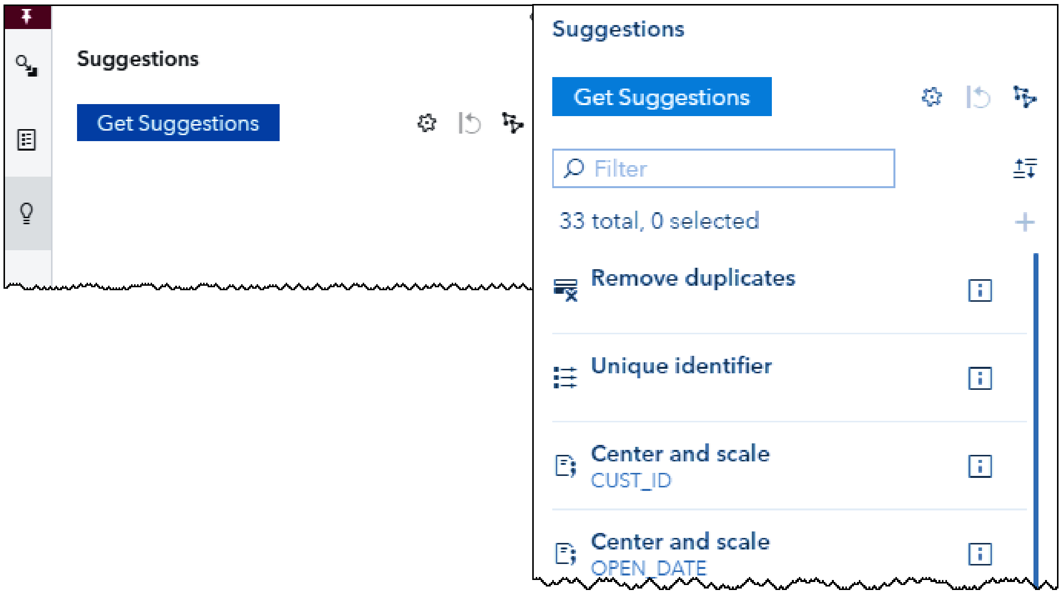


BANK_CUSTOMERS

CUSTOMER_NAME	ADDRESS	CITY	COUNTY	STATE
Stephen seiters	814 Blue Mound rOAD	fort WORTH	Tarrant	TX
Doug mAtrisciano	12500 Ne 10th Pl	Bellevue	King	waSHINGTON
bridget Knightly	66 Flint sTreet	Asheville	Buncombe	NC
Kennith Kirklin	2303 21st Ave S	nashville	Davidson	TeNNeSSee
James Mitchell	8253 Ronson Rd	pasadena		CA
kathleen Beresnyak	100 W 25th Avenue	San Mateo	San Mateo	CA

Note: The data in the *CUSTOMER_NAME*, *ADDRESS*, *CITY*, and *STATE* columns is **mixed case** and **not in a standard format**.

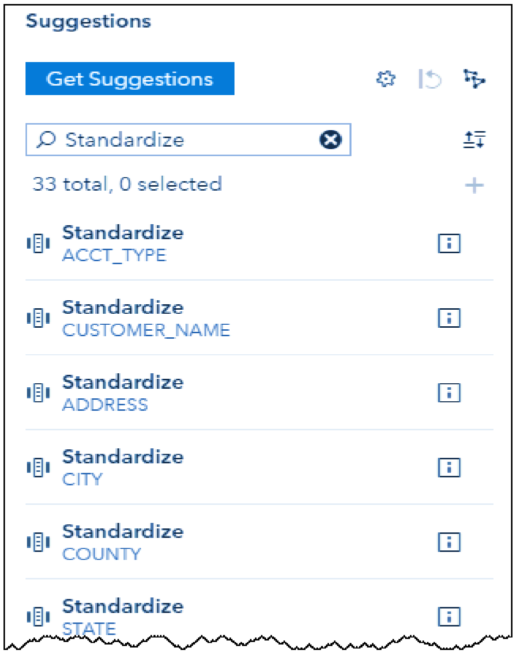
j. Click (Suggestions) in the left-hand pane. Click Get Suggestions.



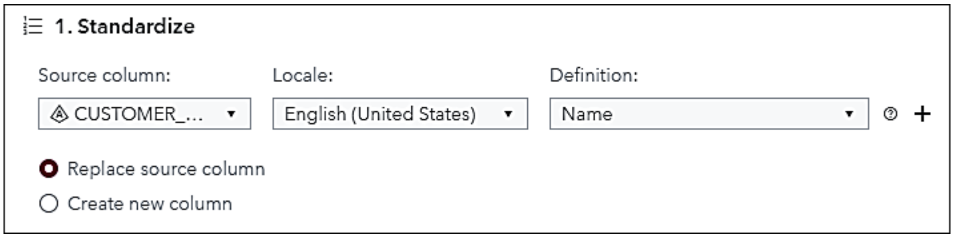
The system has provided some of 32 suggestions for the selected table for you to use.

k. **Standardize the Column**. Enter "Standardize: in the Filter box.

- **Twelve standardizations are suggested** for this table. These columns have standardization suggestions because they were the right type of data, character, and they seemed like they contained data that corresponds to one of our standardization algorithms.
- Sometimes the suggestions make a mistake. For example, ACCT_TYPE is probably unique to this bank data and cannot be standardized by one of our default standardization algorithms. It might look enough like a phone number or address that the sort of data in the column was misclassified. However, most of the suggestions are correct. Name, Address, City, County, and State can be standardized.
- Standardization put every value in the column in a consistent format.
  For example, it applies consistent casing, abbreviations, and ordering to address values.



l. Double-click the **Standardize CUSTOMER_NAME suggestion** to add it to the data preparation plan.



1. Standardize

Source column: CUSTOMER_...

Locale: English (United States)

Definition: Name

● Replace source column

○ Create new column

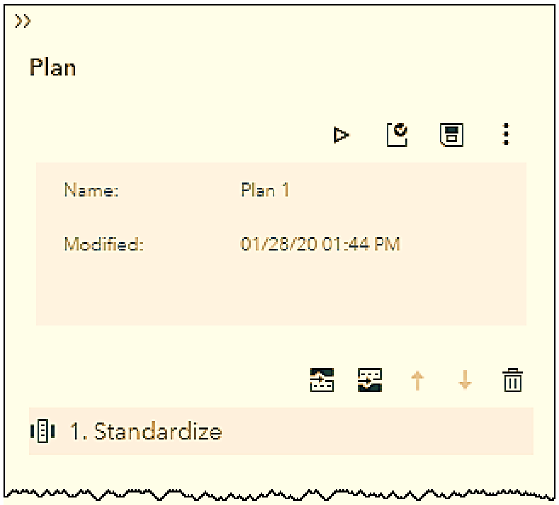
m. The Standardize transform is added to the plan.

- Verify that the **CUSTOMER_NAME column** is selected as the source column.
- Verify that the selected locale is **English (United States)**.
- Verify that the selected definition is **Name**

The suggestion gives us default settings, but we can change them.

n. Click **Replace source** column

o. This is the first transform in the plan. Transforms can be deleted from anywhere in the plan or moved forward or backward in the processing order.



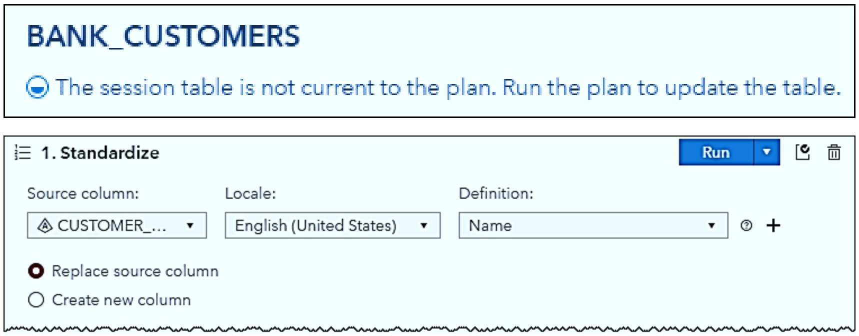
Plan

Name: Plan 1

Modified: 01/28/20 01:44 PM

1. Standardize

p. The **session table is an in-memory table**. Right now, it reflects the source table BANK_CUSTOMERS. It is not current to the plan because we have not run the plan yet. When we run the plan, the session table is updated to reflect the transforms used. Click Run.



BANK_CUSTOMERS

The session table is not current to the plan. Run the plan to update the table.

1. Standardize

Source column: CUSTOMER_...

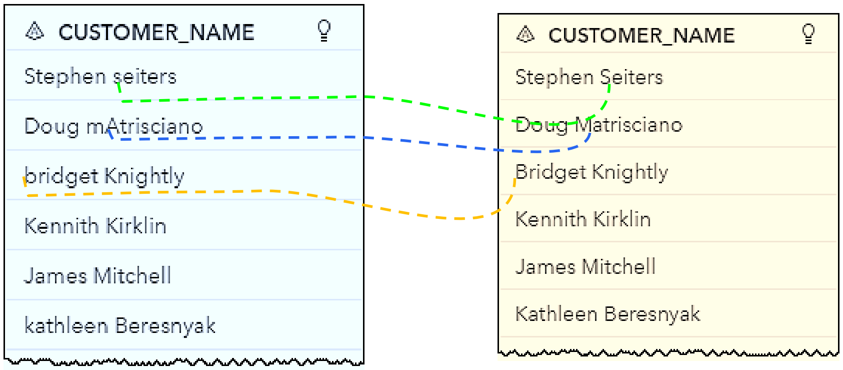
Locale: English (United States)

Definition: Name

● Replace source column

○ Create new column

Partial List of the Original Data   Partial List of the Standardized Data



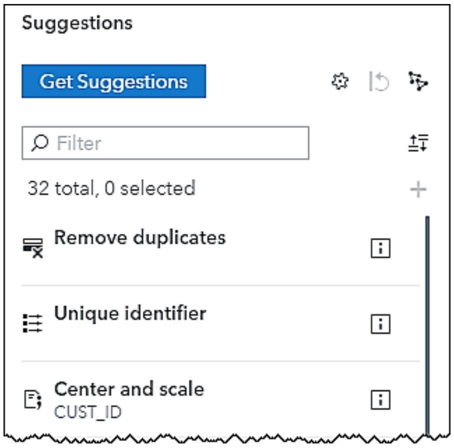
CUSTOMER_NAME
Stephen seiters
Doug mAtrisciano
bridget Knightly
Kennith Kirklin
James Mitchell
kathleen Beresnyak

CUSTOMER_NAME
Stephen Seiters
Doug Matrisciano
Bridget Knightly
Kennith Kirklin
James Mitchell
Kathleen Beresnyak

The casing for the CUSTOMER_NAME column is now consistent. Other standardizations such as applying consistent abbreviations to name suffixes might have been applied, but we can see that the data in the first six rows are now consistently in proper case.

q. Click the X to remove the Standardize filter.

r. Click Get Suggestions in the Suggestions pane to regenerate the suggestions.



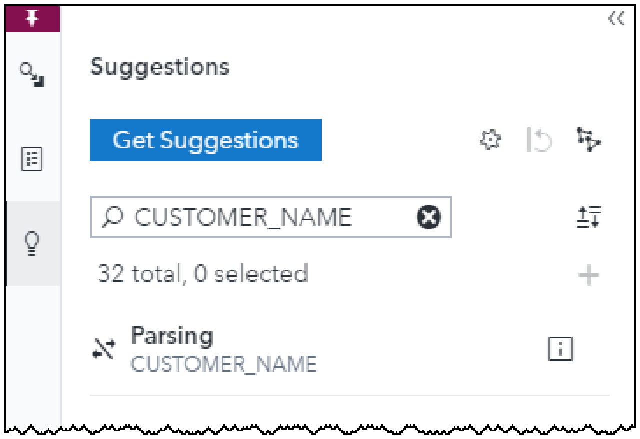
Suggestions

32 total, 0 selected

Remove duplicates

Unique identifier

Center and scale
CUST_ID



Suggestions

CUSTOMER_NAME

32 total, 0 selected

Parsing
CUSTOMER_NAME

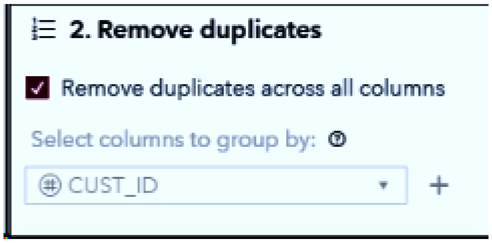
s. Enter CUSTOMER_NAME in the Filter box

- The Standardize CUSTOMER_NAME suggestion has been removed from the list of suggestions because it was already applied to the session table.
- The suggestions evaluate the session table as we update it. If we create more columns, suggestions might be generated for those columns as well.

t. **Screenshot your page of Bank_Customers updated as in the output figure B**.

3. Add and configure the Remove duplicates suggestion

a. Double-click the **Remove duplicates suggestion**



2. Remove duplicates

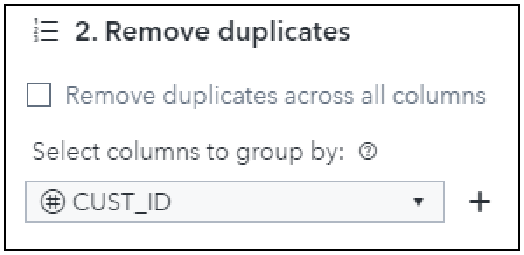
☑ Remove duplicates across all columns

Select columns to group by:

CUST_ID

We can choose to remove duplicates based on the values of one or more specific columns instead of using all columns. As an ID value for the table, we want **CUST_ID to be unique**, so let's remove duplicates by the CUST_ID column.

b. Clear the Remove duplicates across all columns option. Select the CUST_ID column as the column to group by.



2. Remove duplicates

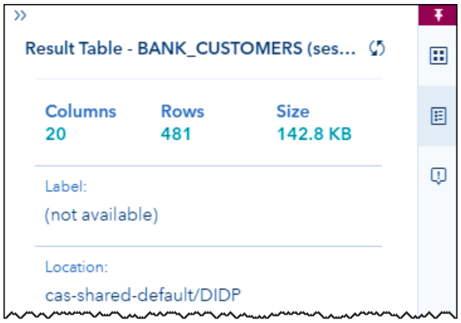
☐ Remove duplicates across all columns

Select columns to group by:

CUST_ID

c. Click (**Properties for the Result table**) in the right pane. There are **currently 483 rows** in the table.

d. Click Run. The **new row count is 481**. Two rows with duplicate CUST_ID values were removed



Result Table - BANK_CUSTOMERS (ses...

Columns	Rows	Size
20	481	142.8 KB

Label:
(not available)

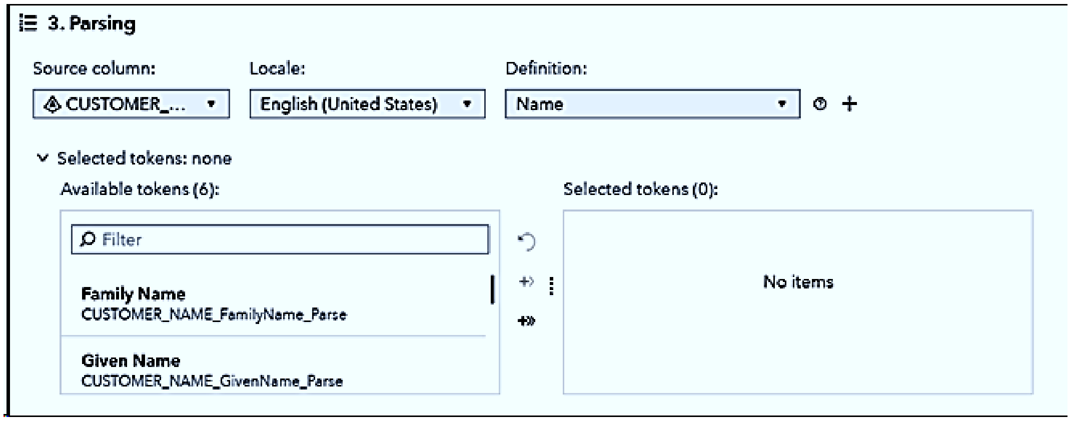
Location:
cas-shared-default/DIDP

e. **Screenshot your page of removing duplicates Cust_Id as in the output figure C**.

4. **Add the Parsing CUSTOMER_NAME suggestion**.

a. Double-click the Parsing CUSTOMER_NAME suggestion.

- The **Parsing** transform enables us to **break up a string into smaller semantic** pieces called **tokens**.
- For example, a Name string might have smaller pieces like a first name, middle name, or last name, which all have meaning separate from the larger value.



3. Parsing

Source column: CUSTOMER_...

Locale: English (United States)

Definition: Name

Selected tokens: none

Available tokens (6):

Filter

Family Name
CUSTOMER_NAME_FamilyName_Parse

Given Name
CUSTOMER_NAME_GivenName_Parse

Selected tokens (0):

No items

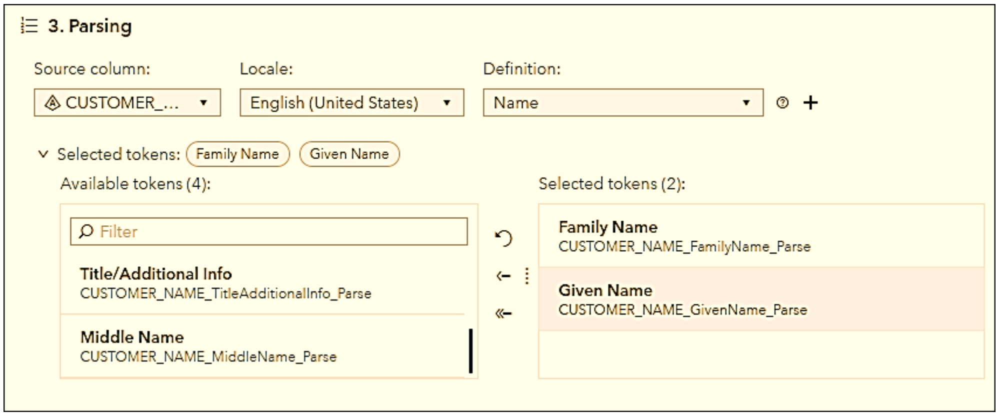
b. Verify that the CUSTOMER_NAME column is selected as the source column.

c. Verify that the selected locale is English (United States).

d. Verify that the selected definition is Name.

e. Double-click the Family Name token to add it to the Selected tokens box.

f. Double-click the Given Name token to add it to the Selected tokens box



3. Parsing

Source column: CUSTOMER_...

Locale: English (United States)

Definition: Name

Selected tokens: Family Name, Given Name

Available tokens (4):

Filter

Title/Additional Info
CUSTOMER_NAME_TitleAdditionalInfo_Parse

Middle Name
CUSTOMER_NAME_MiddleName_Parse

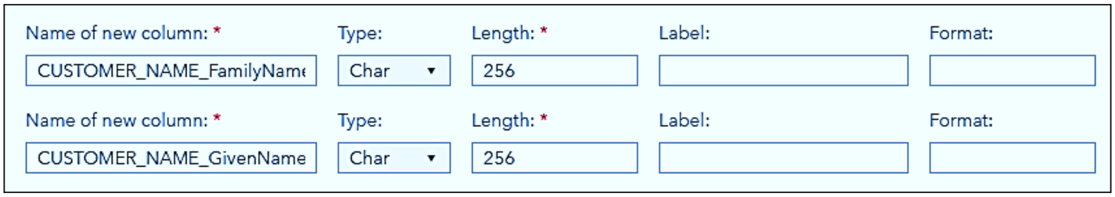
Selected tokens (2):

Family Name
CUSTOMER_NAME_FamilyName_Parse

Given Name
CUSTOMER_NAME_GivenName_Parse

Notice the long column name under the token name. We will change these to simpler column names.

g. Click “Options for new columns”.



Name of new column: *	Type:	Length: *	Label:	Format:
CUSTOMER_NAME_FamilyName	Char	256		
CUSTOMER_NAME_GivenName	Char	256		

In this window, you can change column properties.

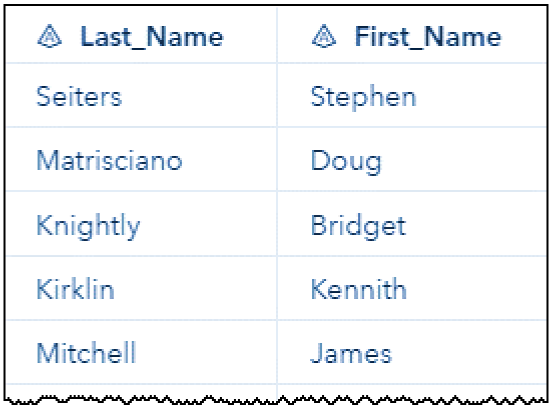
h. Rename CUSTOMER_NAME_FamilyName_Parse to Last_Name.

i. Rename CUSTOMER_NAME_GivenName_Parse to First_Name.

j. Changes Length for First and Last Name to 64

k. Click OK to save the column properties.

l. Click Run to run the data preparation plan.

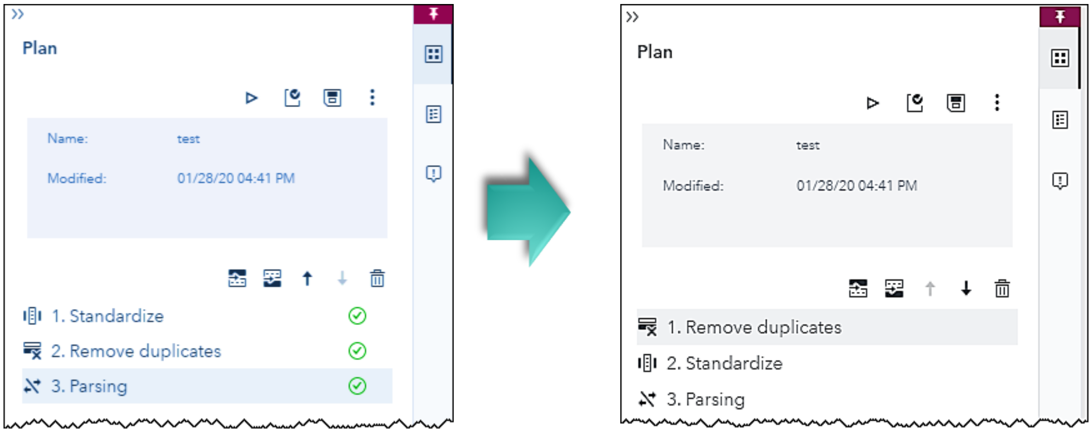


Last_Name	First_Name
Seiters	Stephen
Matrisciano	Doug
Knightly	Bridget
Kirklin	Kennith
Mitchell	James

There are now two new columns for First_Name and Last_Name.

m. Screenshot your parsing page of Customer_name suggestion as in the output figure D

5. **Change the order of the transforms.**
   a. Change the order of the transforms.



Name: test

Modified: 01/28/20 04:41 PM

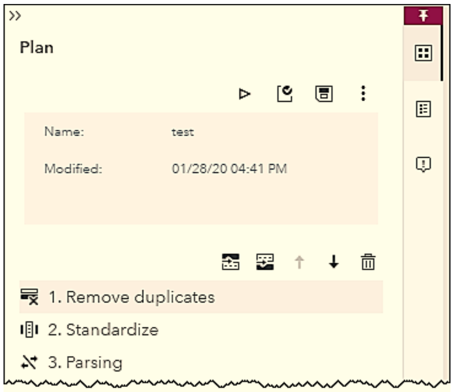
1. Standardize
2. Remove duplicates
3. Parsing

Name: test

Modified: 01/28/20 04:41 PM

1. Remove duplicates
2. Standardize
3. Parsing

   b. If our data had a lot of duplicates, it would be more efficient to remove duplicates before doing other processing. Let's make that transform the first one to run.
   c. If necessary, click ⊞ (Plan).
   d. Click the Remove duplicates transform.
   e. Click ↑(up arrow) to move it above Standardize



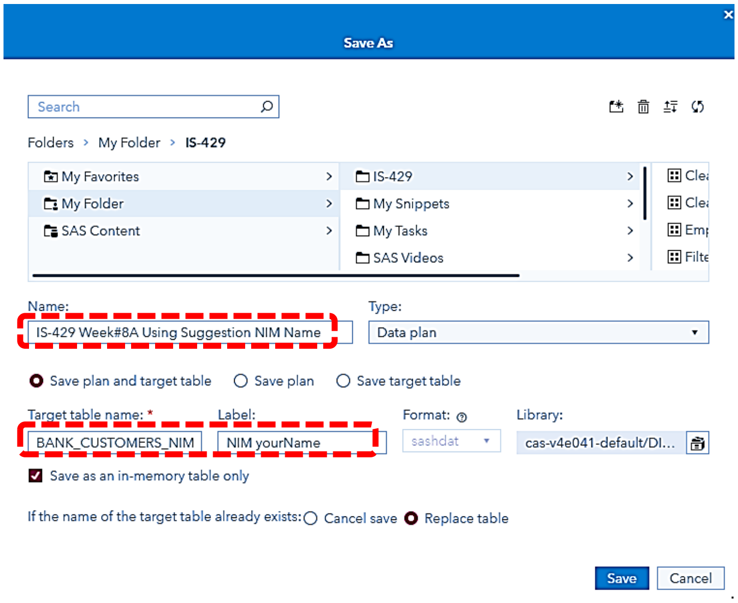
Name: test

Modified: 01/28/20 04:41 PM

1. Remove duplicates
2. Standardize
3. Parsing

Remove duplicates is now the first transform in the plan.

   f. **Save the plan**. Click ⋮ (Options) in the upper right corner to see the plan options.
   g. Click Save As to save the plan as follows:



Name: IS-429 Week#8A Using Suggestion NIM Name

Type: Data plan

● Save plan and target table ○ Save plan ○ Save target table

Target table name: * BANK_CUSTOMERS_NIM

Label: NIM yourName

Format: sashdat

Library: cas-v4e041-default/DI...

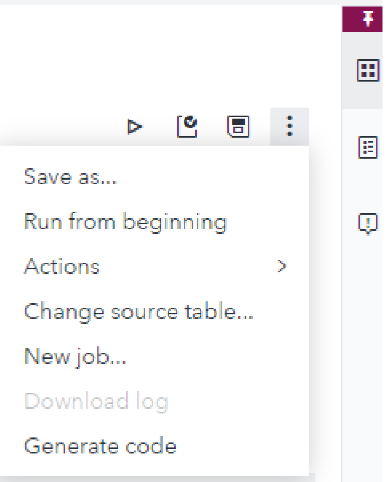
☑ Save as an in-memory table only

If the name of the target table already exists: ○ Cancel save ● Replace table

h. Verify that Save plan and target table is selected.
   - We can open and change or run saved plans later.
   - If we save the target table, we can open and use it in different applications or as a source for a new plan.

i. Verify that sashdat is selected as the format.
   - Target tables can be saved in many formats depending on the target caslib.
   - For PATH caslibs, the target formats include SASHDAT, TXT, CSV, SAS7BDAT, XLSX, XLS, PARQUET, and ORC. If you do not need to save the data on disk, the Save as in-memory table only check box enables you to load the table to memory without saving it to disk.

j. Click Save to your folder.. Click ⋮ Options) and **Generate code for your Plan**:



   - The system will automatically generate the code in SAS format and place it directly into the "download" folder in your workspace with a .txt file format.
   - **Rename your plan code generated as "IS-429 Lab Week#8 Plan NIM Name.sas"**
   - Select Close to close the plan.

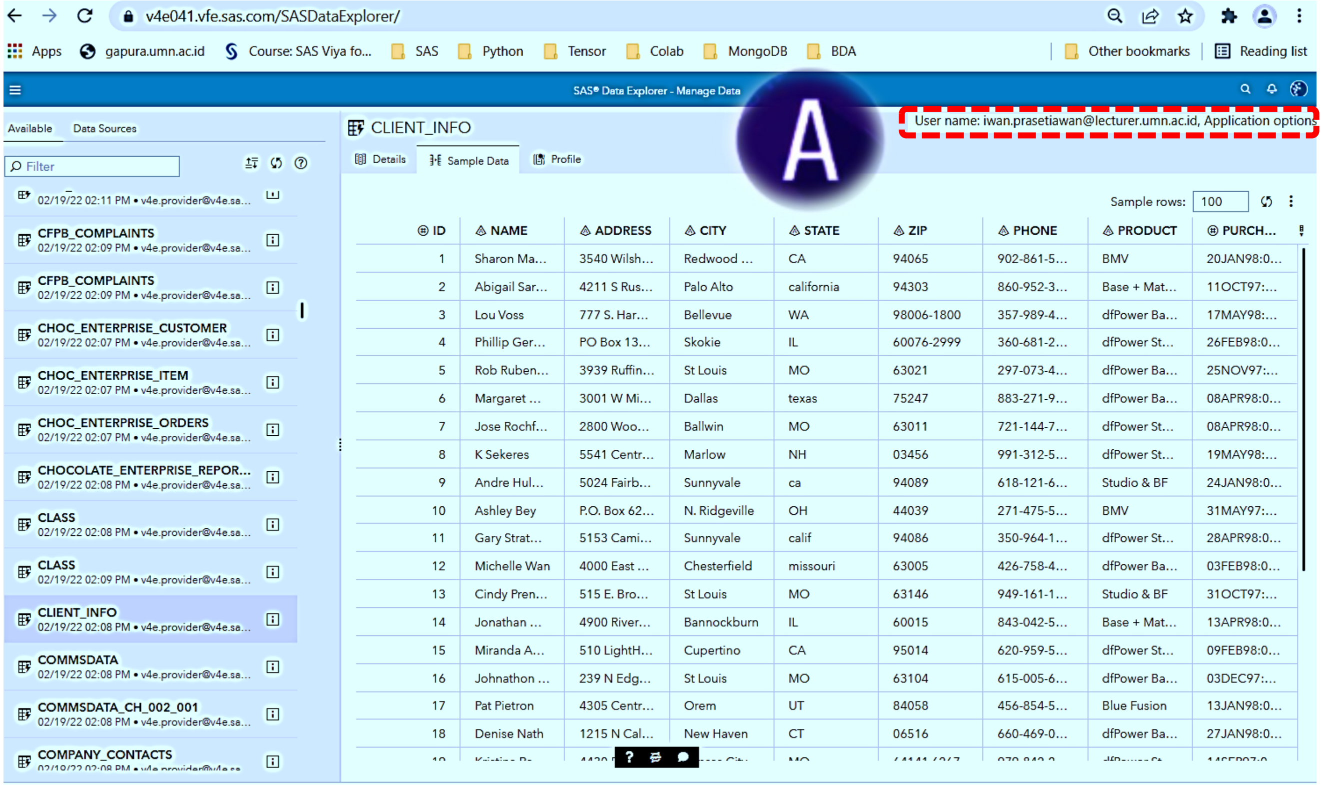
k. Return to SAS [SAS DataExplorer]

n. **Screenhot your Saved Plan page as in the output figure E**.

➲ Finally, today's practicum is over, **collect your doc/docx and sas file into IS-429 Week#8 NIM yourName.zip** and submit immediately today to e-Learning IS-429 Practicum Week#8

RESULTS/ OUTPUT

A. SAS Drive Explorer: Explore and load the CLIENT_INFO table



User name: iwan.prasetiawan@lecturer.umn.ac.id, Application options

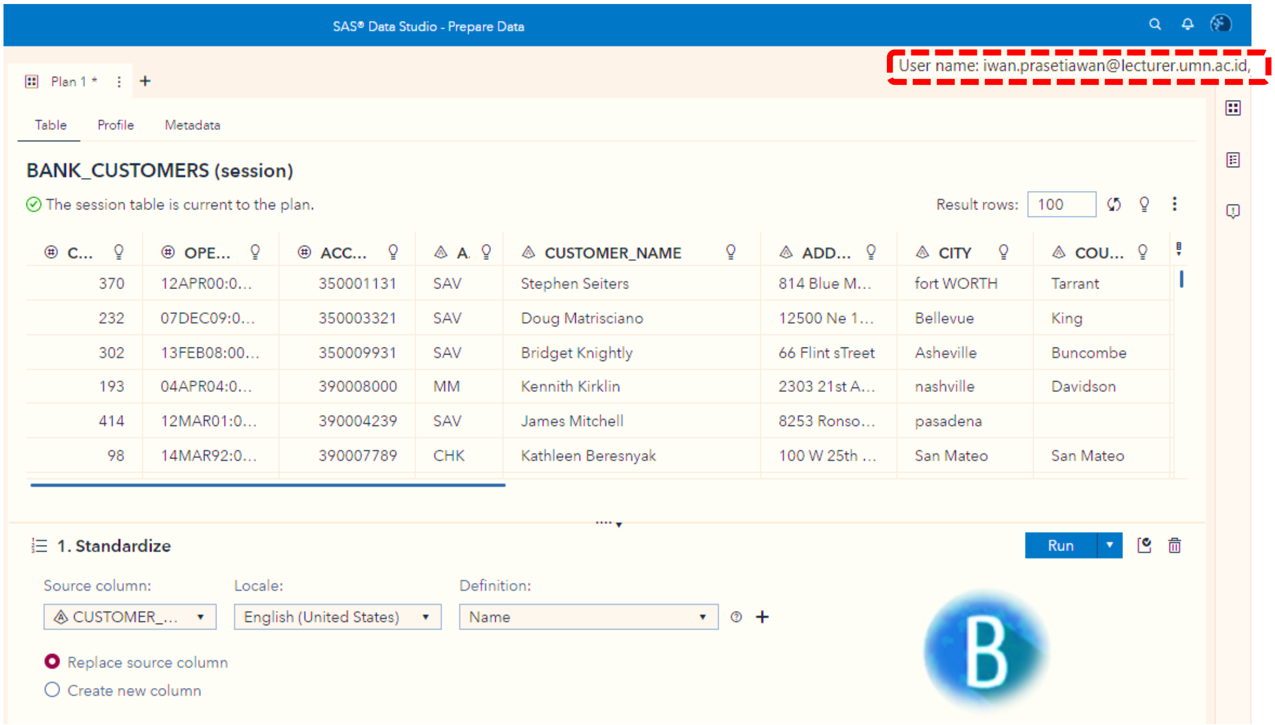
	02/19/22 02:11 PM • v4e.provider@v4e.sa...
CFPB_COMPLAINTS	02/19/22 02:09 PM • v4e.provider@v4e.sa...
CFPB_COMPLAINTS	02/19/22 02:09 PM • v4e.provider@v4e.sa...
CHOC_ENTERPRISE_CUSTOMER	02/19/22 02:07 PM • v4e.provider@v4e.sa...
CHOC_ENTERPRISE_ITEM	02/19/22 02:07 PM • v4e.provider@v4e.sa...
CHOC_ENTERPRISE_ORDERS	02/19/22 02:07 PM • v4e.provider@v4e.sa...
CHOCOLATE_ENTERPRISE_REPOR...	02/19/22 02:08 PM • v4e.provider@v4e.sa...
CLASS	02/19/22 02:08 PM • v4e.provider@v4e.sa...
CLASS	02/19/22 02:09 PM • v4e.provider@v4e.sa...
CLIENT_INFO	02/19/22 02:08 PM • v4e.provider@v4e.sa...
COMMSDATA	02/19/22 02:08 PM • v4e.provider@v4e.sa...
COMMSDATA_CH_002_001	02/19/22 02:08 PM • v4e.provider@v4e.sa...
COMPANY_CONTACTS	

CLIENT_INFO

Sample rows: 100

ID	NAME	ADDRESS	CITY	STATE	ZIP	PHONE	PRODUCT	PURCH...
1	Sharon Ma...	3540 Wilsh...	Redwood ...	CA	94065	902-861-5...	BMV	20JAN98:0...
2	Abigail Sar...	4211 S Rus...	Palo Alto	california	94303	860-952-3...	Base + Mat...	11OCT97:...
3	Lou Voss	777 S. Har...	Bellevue	WA	98006-1800	357-989-4...	dfPower Ba...	17MAY98:...
4	Phillip Ger...	PO Box 13...	Skokie	IL	60076-2999	360-681-2...	dfPower St...	26FEB98:0...
5	Rob Ruben...	3939 Ruffin...	St Louis	MO	63021	297-073-4...	dfPower Ba...	25NOV97:...
6	Margaret ...	3001 W Mi...	Dallas	texas	75247	883-271-9...	dfPower Ba...	08APR98:0...
7	Jose Rochf...	2800 Woo...	Ballwin	MO	63011	721-144-7...	dfPower St...	08APR98:0...
8	K Sekeres	5541 Centr...	Marlow	NH	03456	991-312-5...	dfPower St...	19MAY98:...
9	Andre Hul...	5024 Fairb...	Sunnyvale	ca	94089	618-121-6...	Studio & BF	24JAN98:0...
10	Ashley Bey	P.O. Box 62...	N. Ridgeville	OH	44039	271-475-5...	BMV	31MAY97:...
11	Gary Strat...	5153 Cami...	Sunnyvale	calif	94086	350-964-1...	dfPower St...	28APR98:0...
12	Michelle Wan	4000 East ...	Chesterfield	missouri	63005	426-758-4...	dfPower Ba...	03FEB98:0...
13	Cindy Pren...	515 E. Bro...	St Louis	MO	63146	949-161-1...	Studio & BF	31OCT97:...
14	Jonathan ...	4900 River...	Bannockburn	IL	60015	843-042-5...	Base + Mat...	13APR98:0...
15	Miranda A...	510 LightH...	Cupertino	CA	95014	620-959-5...	dfPower St...	09FEB98:0...
16	Johnathon ...	239 N Edg...	St Louis	MO	63104	615-005-6...	dfPower Ba...	03DEC97:...
17	Pat Pietron	4305 Centr...	Orem	UT	84058	456-854-5...	Blue Fusion	13JAN98:0...
18	Denise Nath	1215 N Cal...	New Haven	CT	06516	660-469-0...	dfPower Ba...	27JAN98:0...

B. SAS Data Studio, Using Suggestions to Prepare Data: Standardize Customer_Name



User name: iwan.prasetiawan@lecturer.umn.ac.id,

BANK_CUSTOMERS (session)

The session table is current to the plan.

Result rows: 100

C...	OPE...	ACC...	A	CUSTOMER_NAME	ADD...	CITY	COU...
370	12APR00:0...	350001131	SAV	Stephen Seiters	814 Blue M...	fort WORTH	Tarrant
232	07DEC09:0...	350003321	SAV	Doug Matrisciano	12500 Ne 1...	Bellevue	King
302	13FEB08:00...	350009931	SAV	Bridget Knightly	66 Flint sTreet	Asheville	Buncombe
193	04APR04:0...	390008000	MM	Kennith Kirklin	2303 21st A...	nashville	Davidson
414	12MAR01:0...	390004239	SAV	James Mitchell	8253 Ronso...	pasadena	
98	14MAR92:0...	390007789	CHK	Kathleen Beresnyak	100 W 25th ...	San Mateo	San Mateo

1. Standardize

Source column: CUSTOMER_...

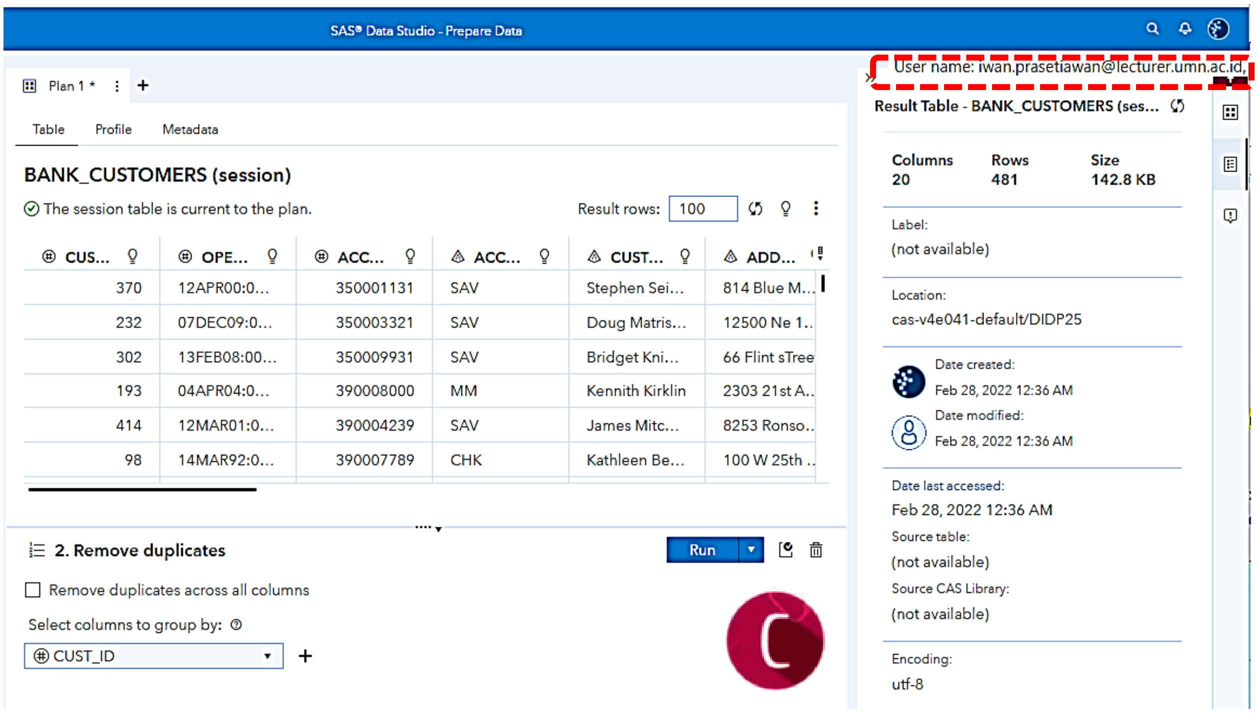
Locale: English (United States)

Definition: Name

● Replace source column

○ Create new column

C. Add and configure the Remove duplicates suggestion



SAS® Data Studio - Prepare Data

User name: iwan.prasetiawan@lecturer.umn.ac.id,

Plan 1 *

Table | Profile | Metadata

BANK_CUSTOMERS (session)

The session table is current to the plan.

Result rows: 100

CUS…	OPE…	ACC…	ACC…	CUST…	ADD…
370	12APR00:0…	350001131	SAV	Stephen Sei…	814 Blue M…
232	07DEC09:0…	350003321	SAV	Doug Matris…	12500 Ne 1…
302	13FEB08:00…	350009931	SAV	Bridget Kni…	66 Flint sTree
193	04APR04:0…	390008000	MM	Kennith Kirklin	2303 21st A…
414	12MAR01:0…	390004239	SAV	James Mitc…	8253 Ronso…
98	14MAR92:0…	390007789	CHK	Kathleen Be…	100 W 25th …

2. Remove duplicates

Run

□ Remove duplicates across all columns

Select columns to group by:

CUST_ID

Result Table - BANK_CUSTOMERS (ses…

Columns	Rows	Size
20	481	142.8 KB

Label:
(not available)

Location:
cas-v4e041-default/DIDP25

Date created:
Feb 28, 2022 12:36 AM

Date modified:
Feb 28, 2022 12:36 AM

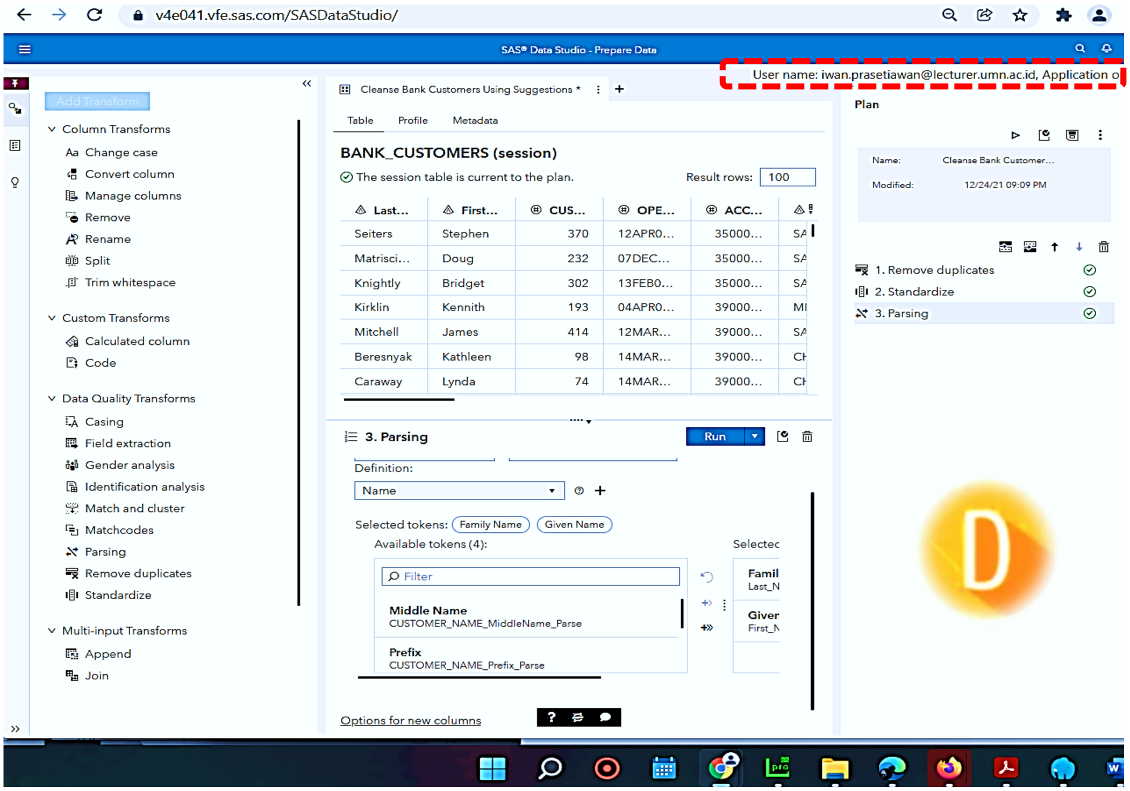
Date last accessed:
Feb 28, 2022 12:36 AM

Source table:
(not available)

Source CAS Library:
(not available)

Encoding:
utf-8

D. Add the Parsing CUSTOMER_NAME suggestion



v4e041.vfe.sas.com/SASDataStudio/

SAS® Data Studio - Prepare Data

User name: iwan.prasetiawan@lecturer.umn.ac.id, Application o

Cleanse Bank Customers Using Suggestions *

Table | Profile | Metadata

BANK_CUSTOMERS (session)

The session table is current to the plan.

Result rows: 100

Last…	First…	CUS…	OPE…	ACC…	
Seiters	Stephen	370	12APR0…	35000…	SA
Matrisci…	Doug	232	07DEC…	35000…	SA
Knightly	Bridget	302	13FEB0…	35000…	SA
Kirklin	Kennith	193	04APR0…	39000…	MI
Mitchell	James	414	12MAR…	39000…	SA
Beresnyak	Kathleen	98	14MAR…	39000…	CH
Caraway	Lynda	74	14MAR…	39000…	CH

3. Parsing

Run

Definition:
Name

Selected tokens: Family Name | Given Name

Available tokens (4):

Filter

Middle Name
CUSTOMER_NAME_MiddleName_Parse

Prefix
CUSTOMER_NAME_Prefix_Parse

Selectec

Famil
Last_N

Giver
First_N

Options for new columns

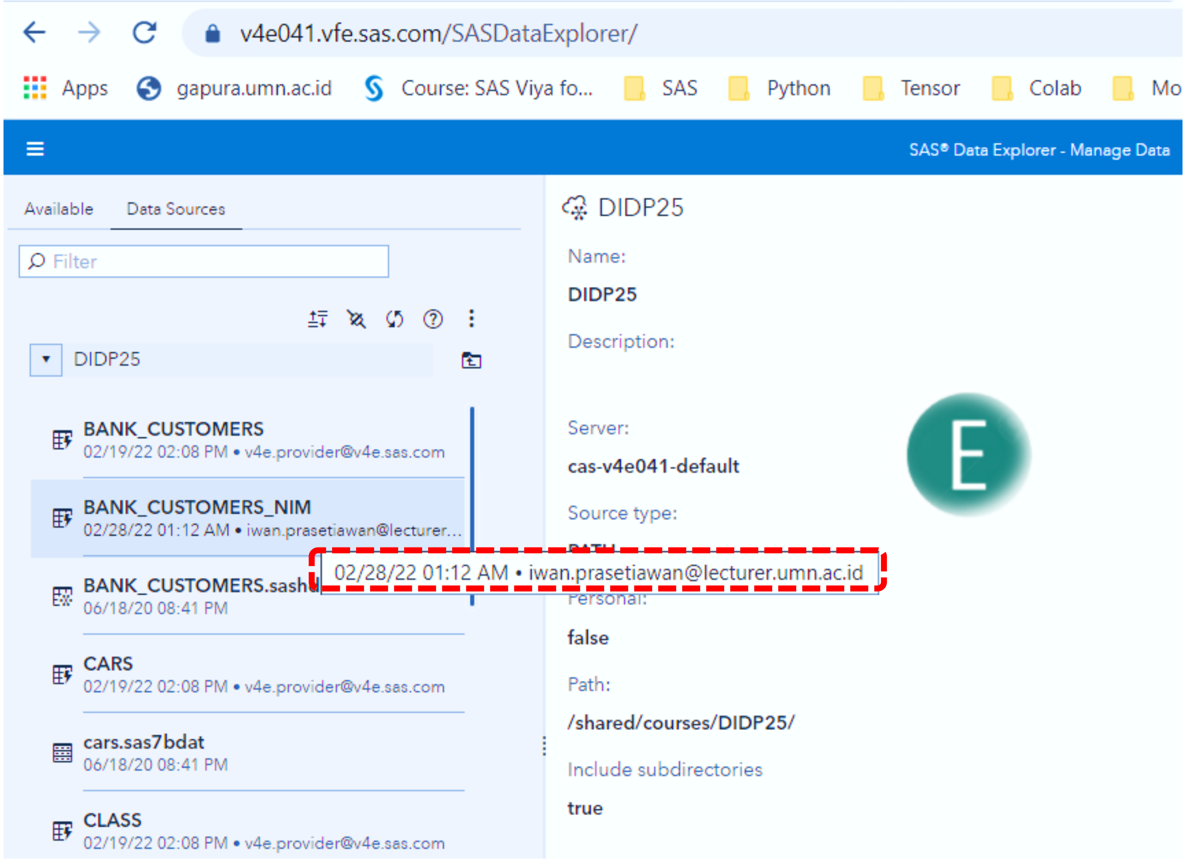
Plan

Name: Cleanse Bank Customer…

Modified: 12/24/21 09:09 PM

1. Remove duplicates ✓
2. Standardize ✓
3. Parsing ✓

E. SAS Data Explorer: New Bank Customer of yours adding the DIDP25 CAS table.



DIDP25

BANK_CUSTOMERS
02/19/22 02:08 PM • v4e.provider@v4e.sas.com

BANK_CUSTOMERS_NIM
02/28/22 01:12 AM • iwan.prasetiawan@lecturer...

02/28/22 01:12 AM • iwan.prasetiawan@lecturer.umn.ac.id

BANK_CUSTOMERS.sash
06/18/20 08:41 PM

CARS
02/19/22 02:08 PM • v4e.provider@v4e.sas.com

cars.sas7bdat
06/18/20 08:41 PM

CLASS
02/19/22 02:08 PM • v4e.provider@v4e.sas.com

DIDP25

Name:
DIDP25

Description:

Server:
cas-v4e041-default

Source type:
PATH

Personal:
false

Path:
/shared/courses/DIDP25/

Include subdirectories
true

The End

REFERENCE

1. Anna Yarbrough. 2020. Introduction to Data Curation for SAS® Data Scientists Course Notes. SAS Institute Inc. Cary, NC, USA.
2. SAS Institute Inc. 2020. SAS® Viya® Programming: Getting Started. SAS Institute Inc. Cary, NC, USA.
3. Johnny Starling, Erin Winters, and Anna Yarbrough. 2020.Self-Service Data Preparation in SAS® Viya® Course Notes. . SAS Institute Inc. Cary, NC, USA
4. SAS® Support | Documentation
5. Other additional references are excerpts from various Online Learning/websites.