Start SAS Enterprise Miner with CLAIM_DATA_V2_TRAIN

Didem B. Aykurt

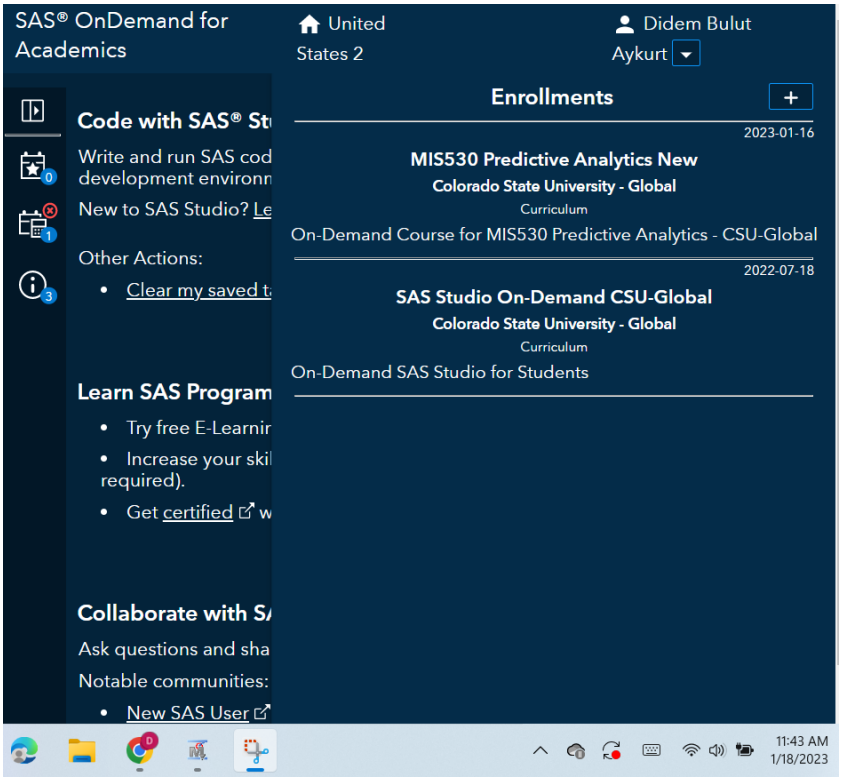
Colorado State University Global

MIS530; Predictive Analytics

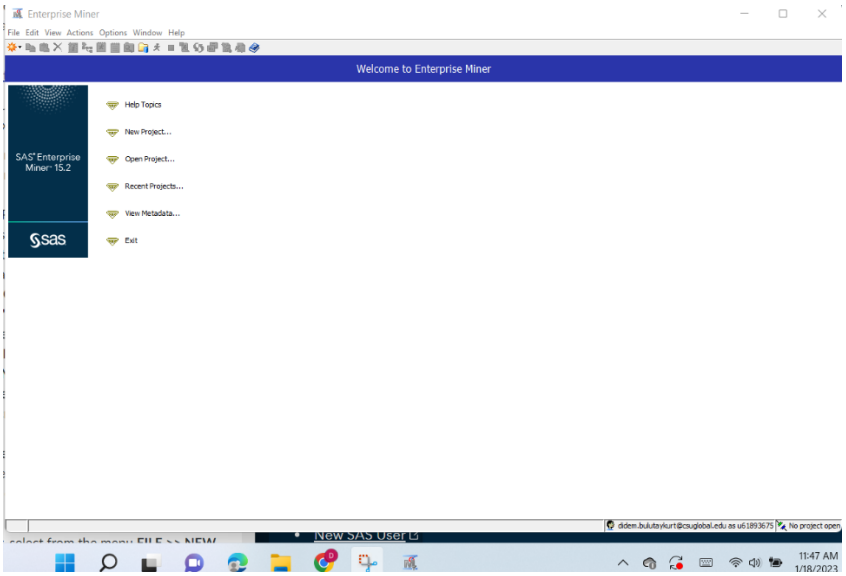
Dr.Jennifer Catalano

January 22, 2023

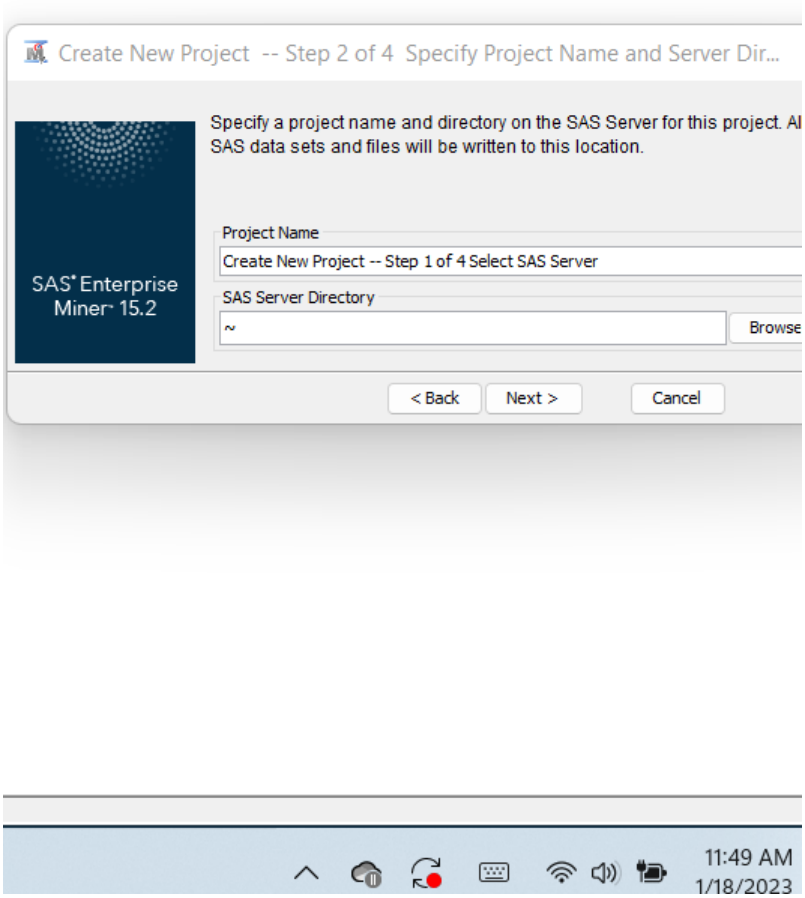
1-Course enrollment: I used the SAS OnDemand for Academic website to add a new course, left the site with the enrolment icon, and then clicked the plus icon to pop up the new tab to give system information.



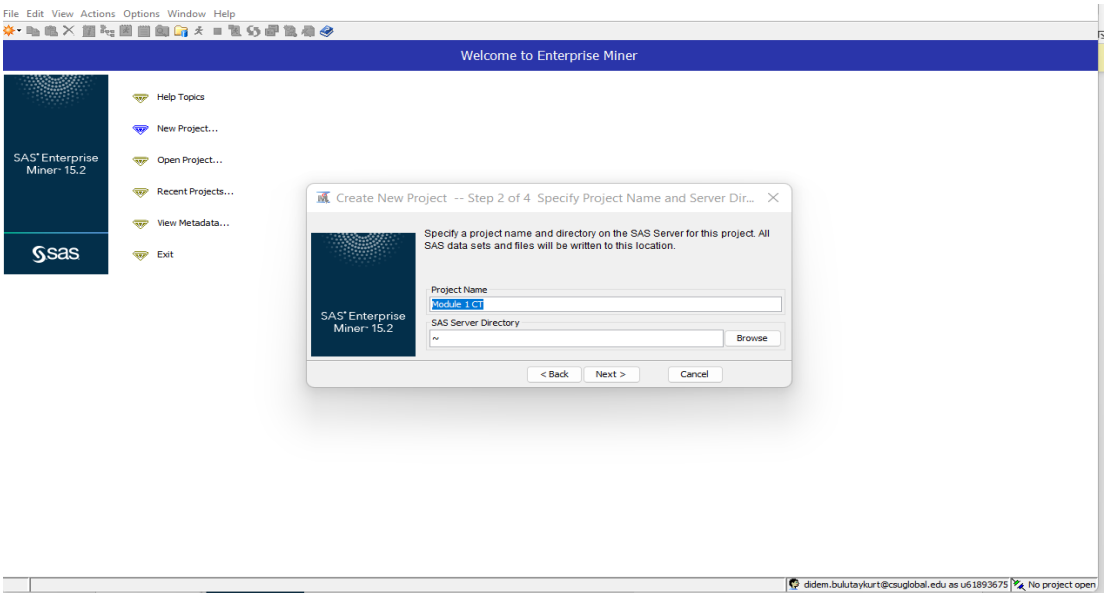
2-Running SAS Enterprise Miner: The left side of the SAS OnDemand for Academics page has a hammer icon that shows how to download and the URL to connect to SAS Enterprise Miner. This part was a little strange because when I gave an email and password to log on, it said, “Public access denied.” How I solved this: I uninstalled the program, cleared data on Google’s history page, and deleted all SAS links. Then I opened the SAS OnDemand Academic page to do all the steps without any open page and web. I also didn’t close the SAS Academic page; I got a URL link from that page and solved the problem.



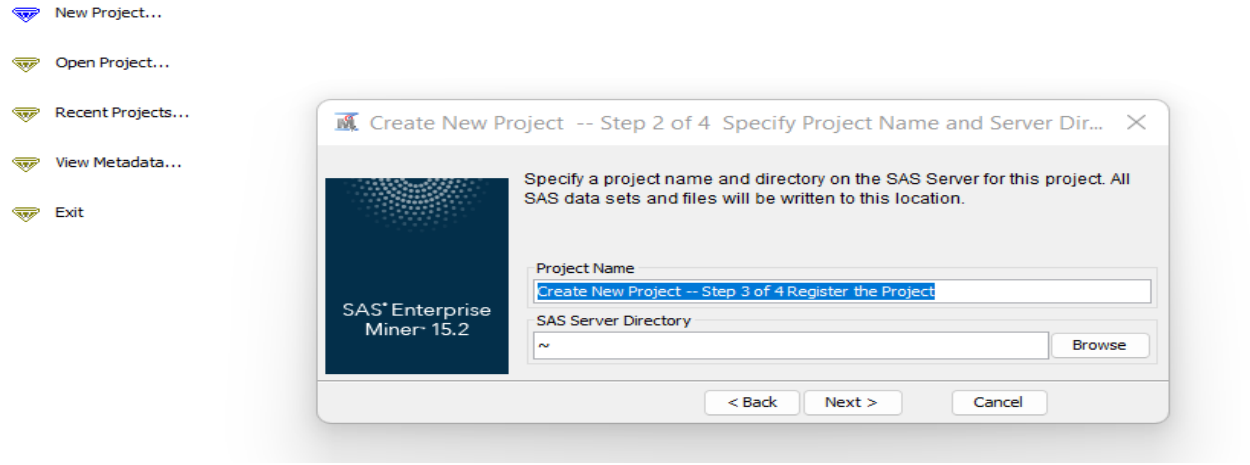
3- New project; click Create a new project on the list, named “Create New Project – Step 1 of 4 Select SAS Server”.



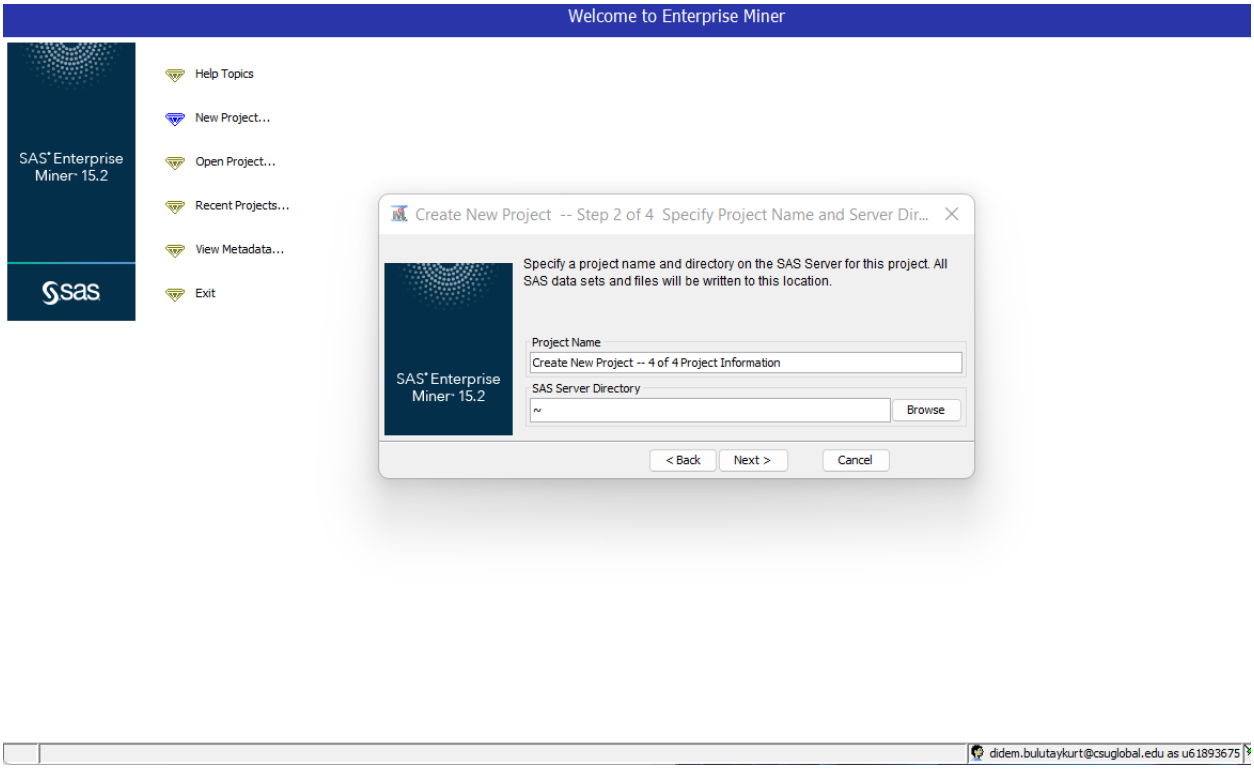
4- New project named “Module 1 CT. “



5- A new project named “Create New Project – Step 3 of 4 Register the Project.”

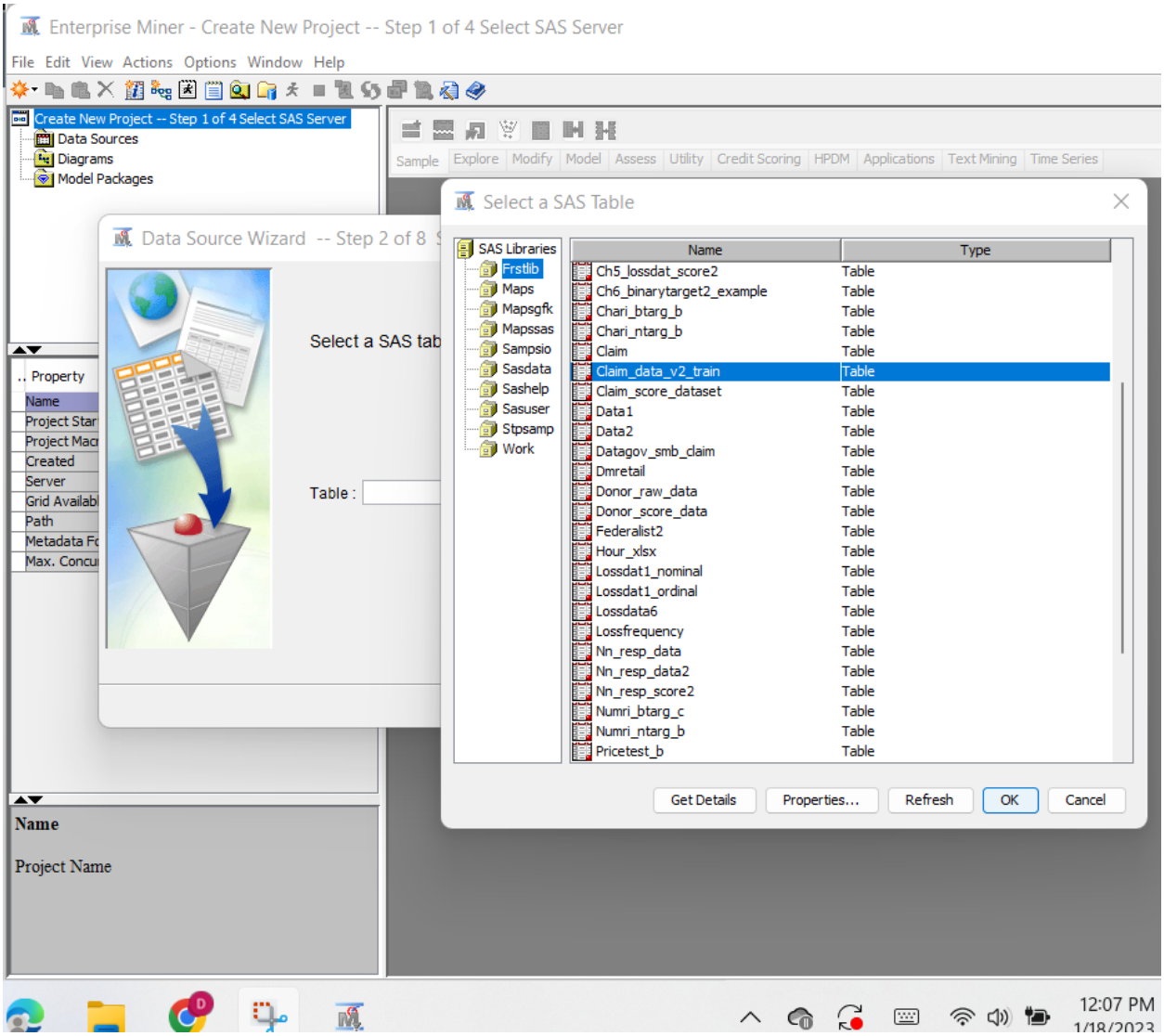


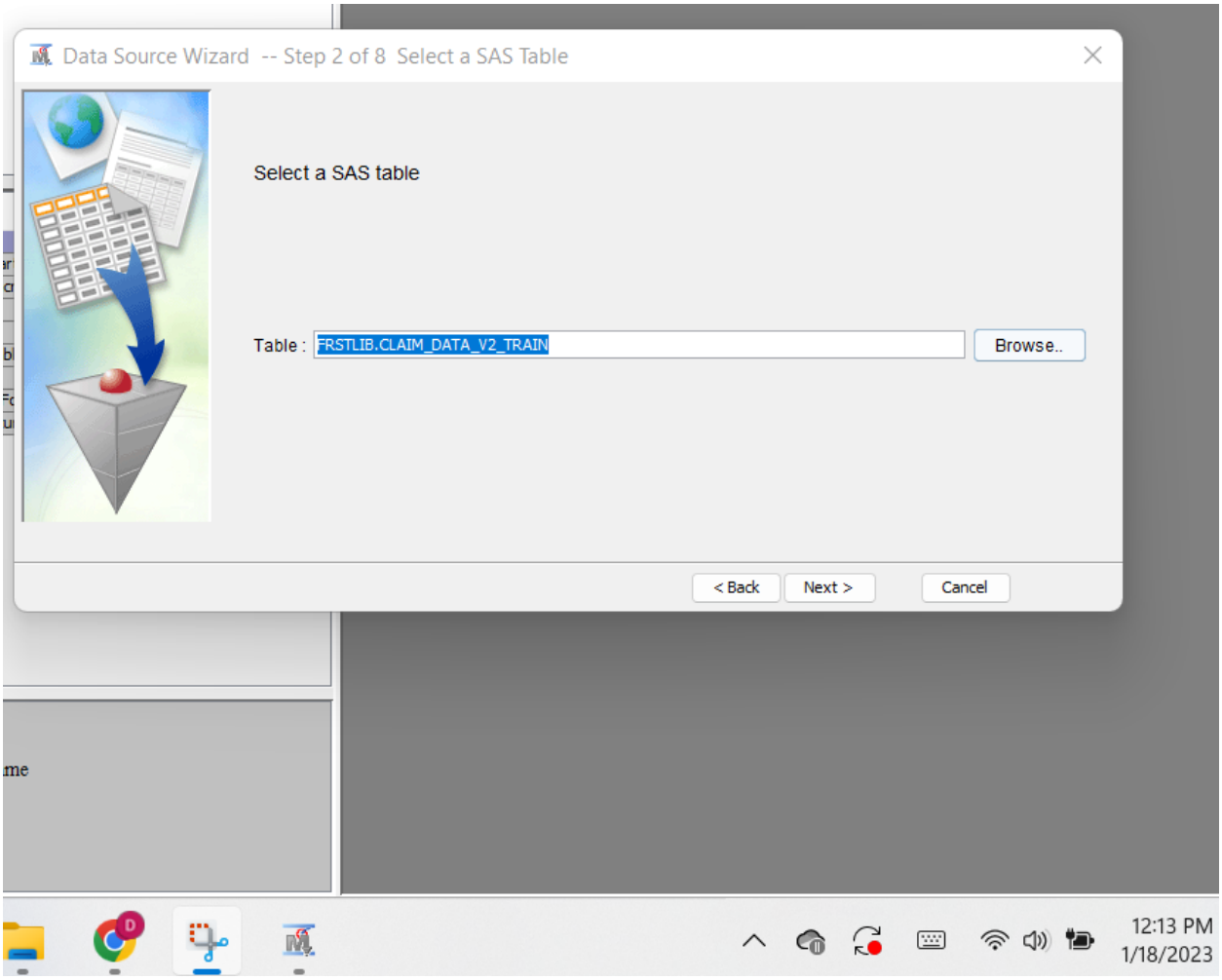
6- A new project named “Create New Project – Step 4 of 4 Project Information.”

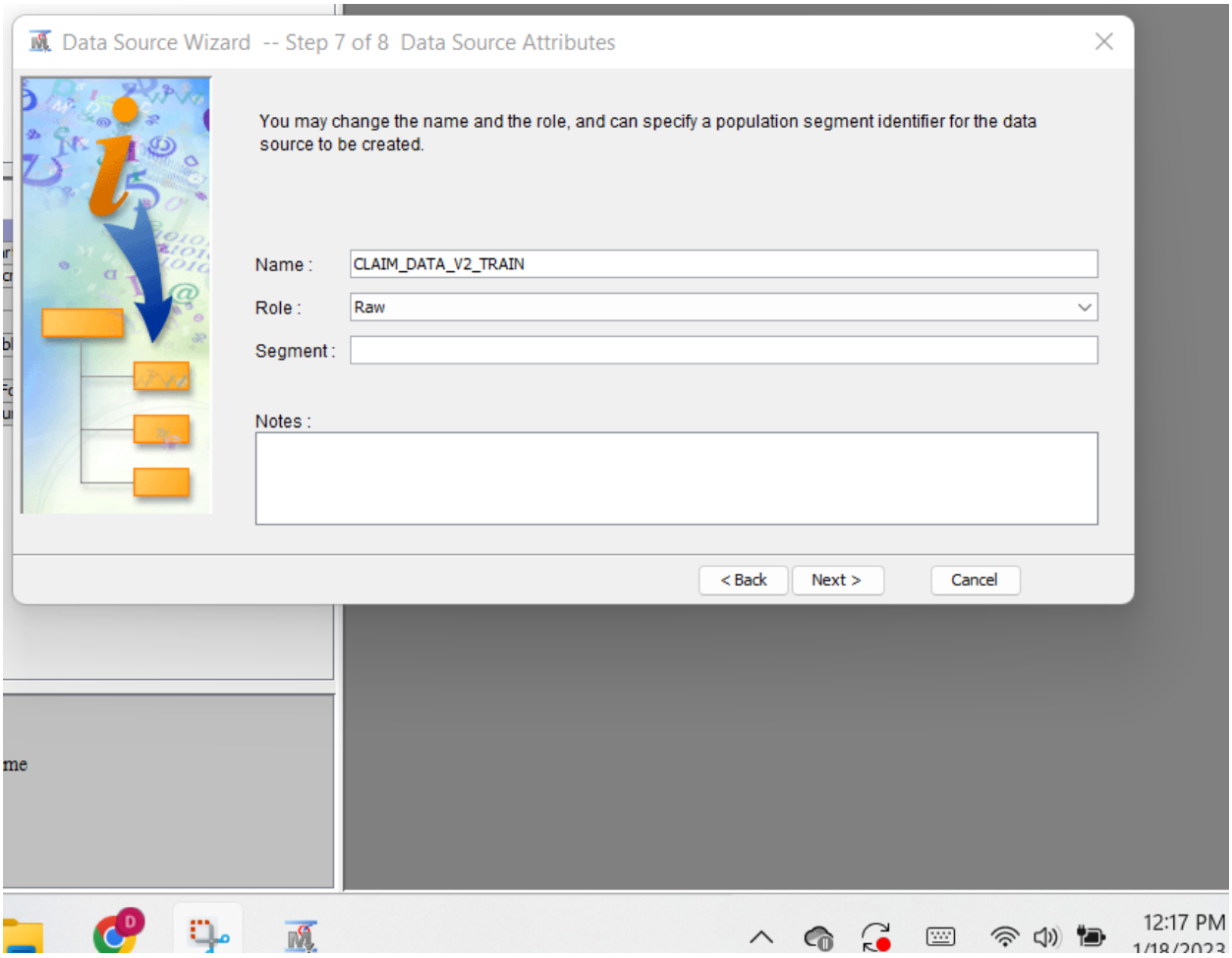


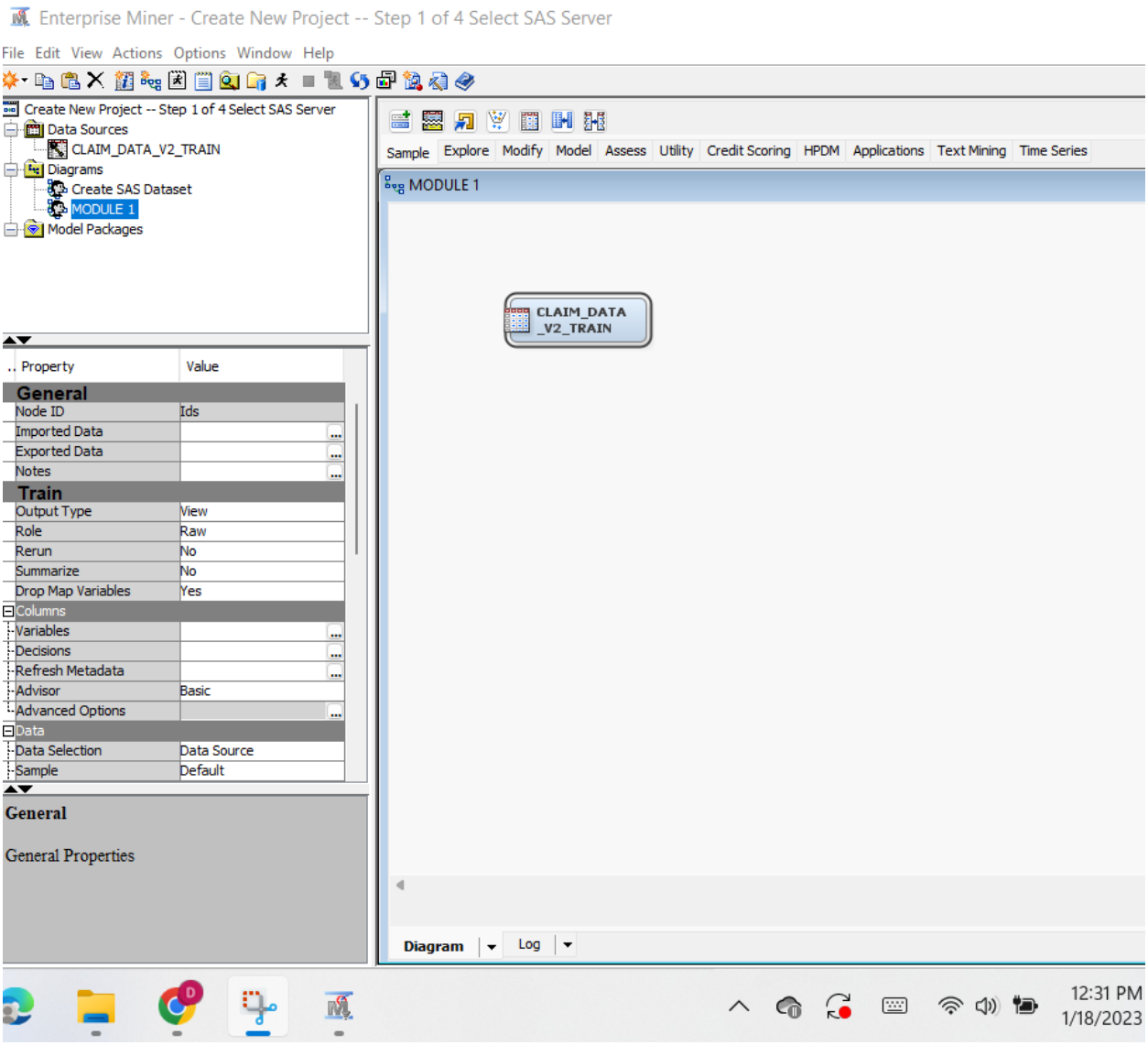
7- Created Diagram; right-click on Diagrams and name “MODULE 1”. That opens a page to work on it with the dataset. Then, I created a Library named “FRSTLIB” and chose the table CLAIM_DATA_V2_TRAIN. Drag the data source from the project panel to the Diagram located in the workspace. Right-click on the data and run. Click the Sample tab, drag the Data Partition node onto the process flow, and then connect the Data Partition node to the CLAIM_DATA_TRAIN node. The sample data set is divided into 60% train, and 40% will be validation.

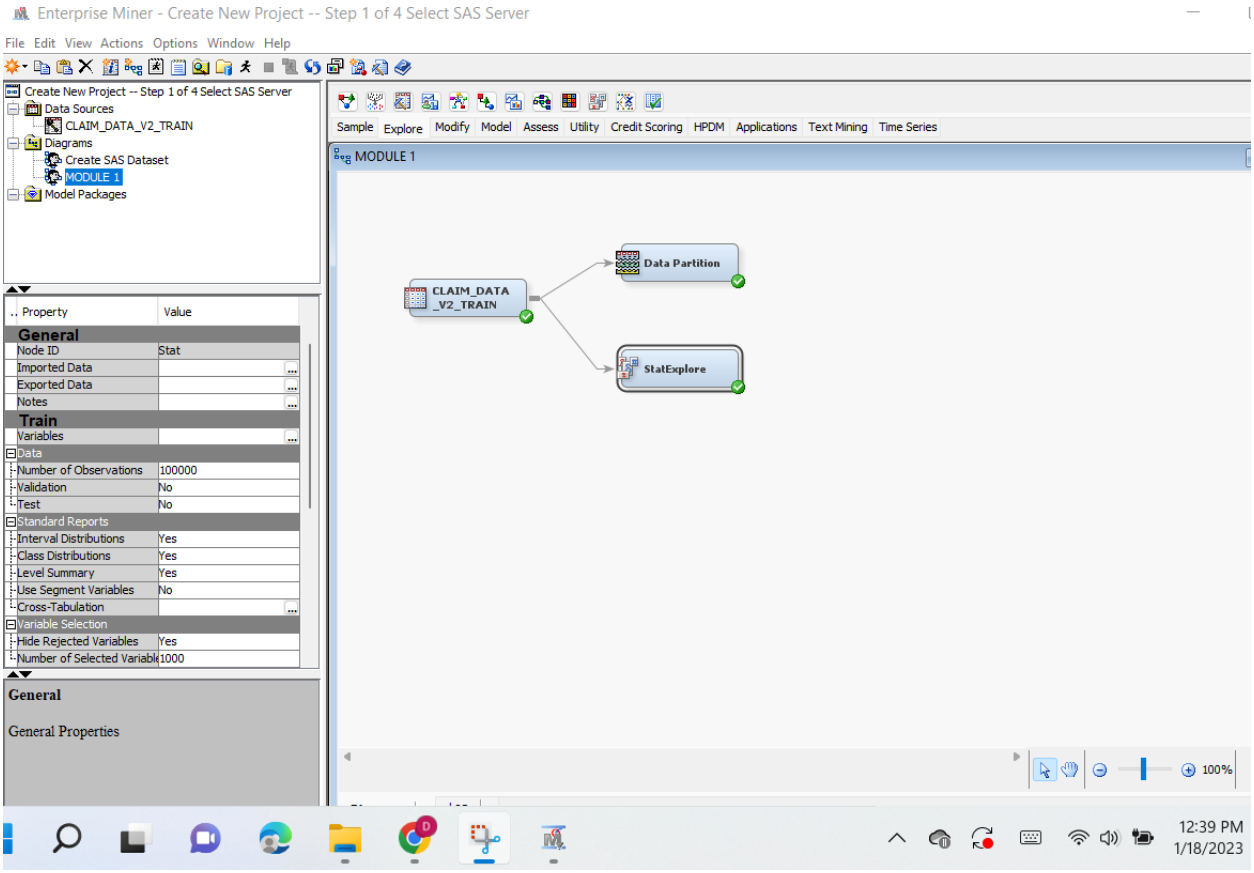
Data Exploration is the data preparation process. One of the tabs is the StatExplore node, which allows descriptive statistics to summarize the raw data like sums, average, and percent changes and describe the past. Also, prepare the data for future use in prescriptive analytics.

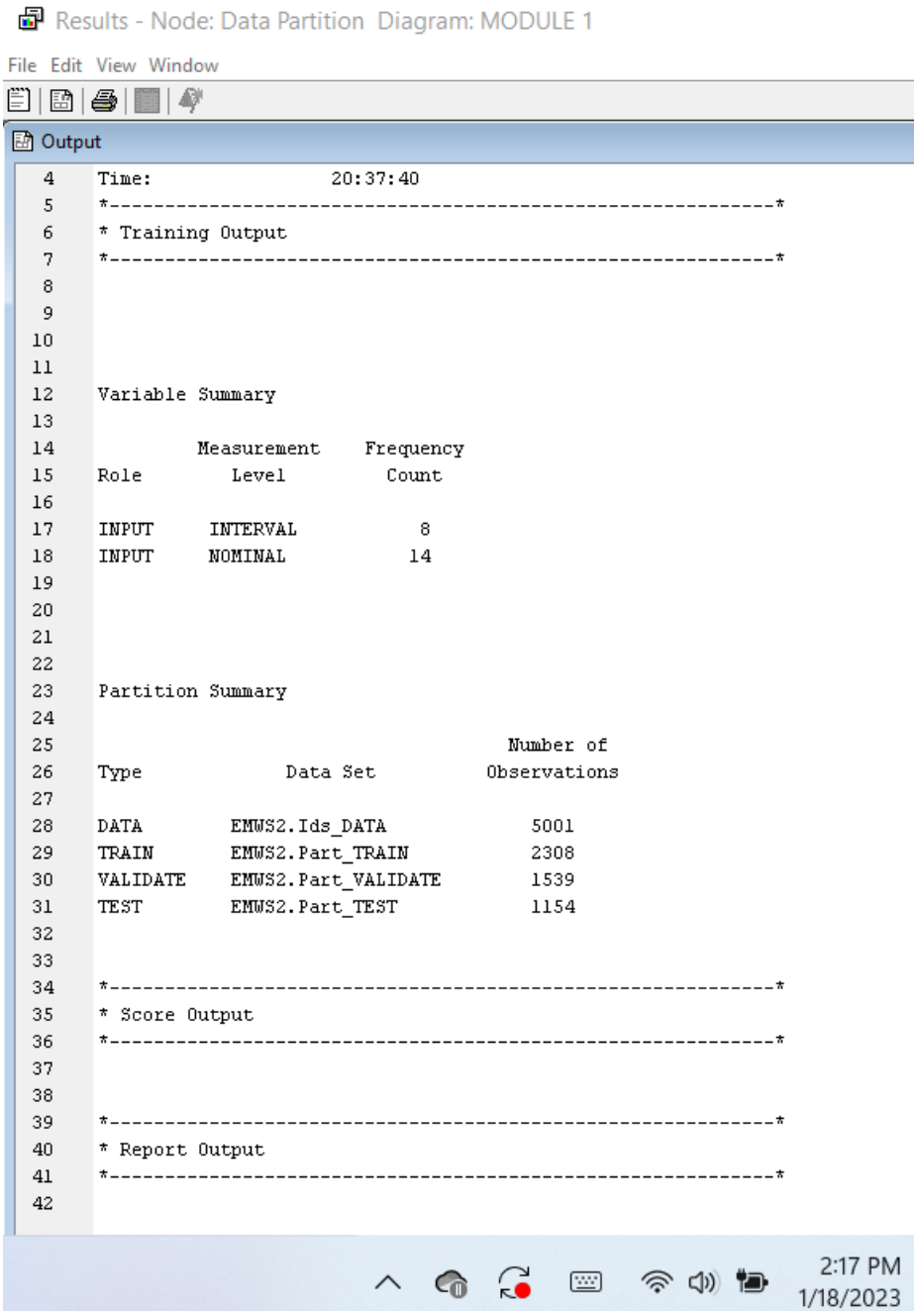

Enterprise Miner - Create New Project -- Step 1 of 4 Select SAS Server
File Edit View Actions Options Window Help
Create New Project -- Step 1 of 4 Select SAS Server
Data Sources
Diagrams
Model Packages
Sample Explore Modify Model Assess Utility Credit Scoring HPDM Applications Text Mining Time Series
Data Source Wizard -- Step 2 of 8
Select a SAS tab
Table :
Select a SAS Table
SAS Libraries
Frstlib
Maps
Mapsgfk
Mapssas
Sampsio
Sasdata
Sashelp
Sasuser
Stpsamp
Work
Name
Type
Ch5_lossdat_score2 Table
Ch6_binarytarget2_example Table
Chari_btarg_b Table
Chari_ntarg_b Table
Claim Table
Claim_data_v2_train Table
Claim_score_dataset Table
Data1 Table
Data2 Table
Datagov_smb_claim Table
Dmretail Table
Donor_raw_data Table
Donor_score_data Table
Federalist2 Table
Hour_xlsx Table
Lossdat1_nominal Table
Lossdat1_ordinal Table
Lossdata6 Table
Lossfrequency Table
Nn_resp_data Table
Nn_resp_data2 Table
Nn_resp_score2 Table
Numri_btarg_c Table
Numri_ntarg_b Table
Pricetest_b Table
Get Details
Properties...
Refresh
OK
Cancel
Property
Name
Project Star
Project Mac
Created
Server
Grid Availab
Path
Metadata Fc
Max. Concu
Name
Project Name
12:07 PM
1/18/2023

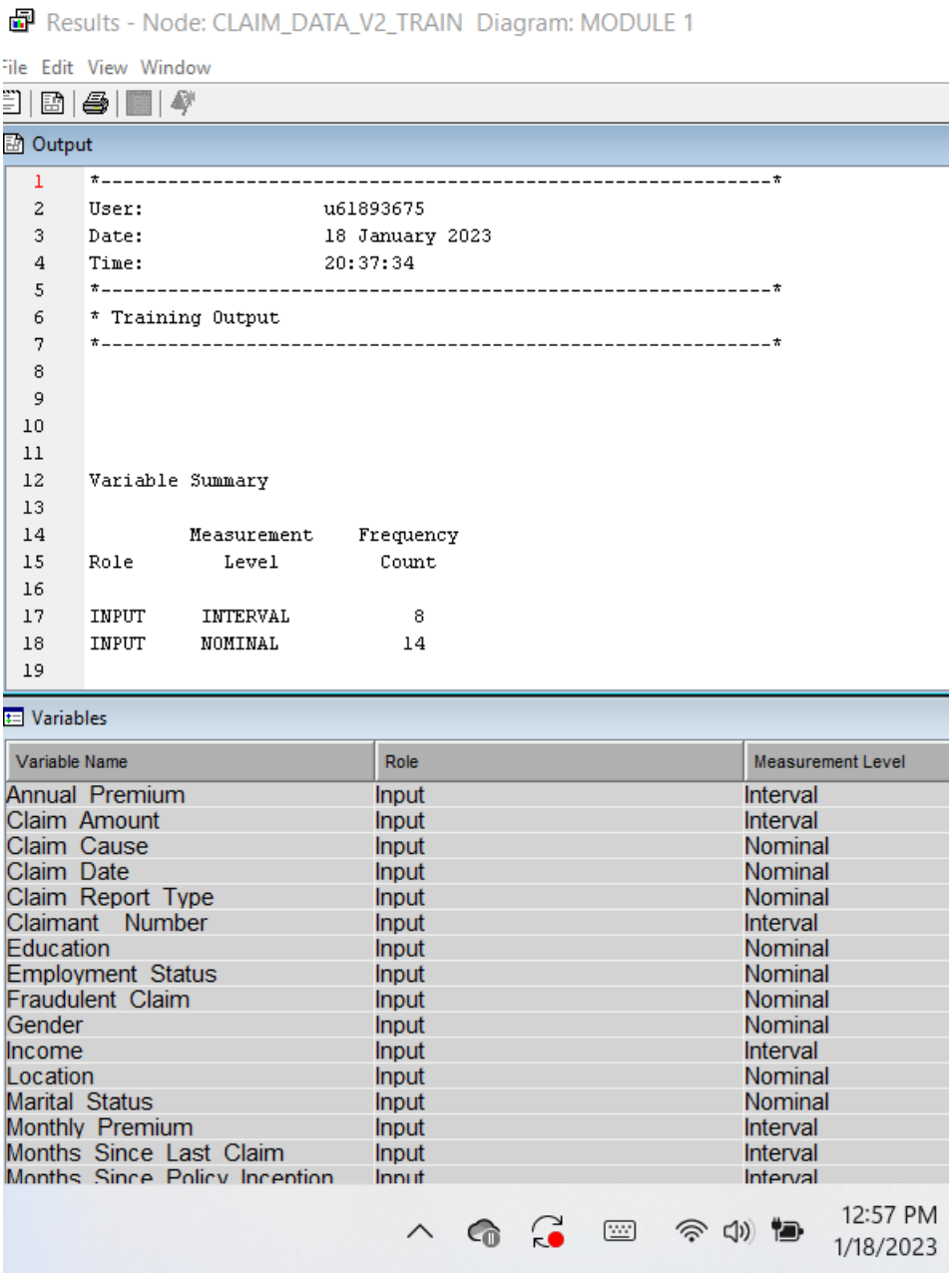

Data Source Wizard -- Step 2 of 8 Select a SAS Table
Select a SAS table
Table :
FRSTLIB.CLAIM_DATA_V2_TRAIN
Browse..
< Back
Next >
Cancel
12:13 PM
1/18/2023


Data Source Wizard -- Step 7 of 8 Data Source Attributes
You may change the name and the role, and can specify a population segment identifier for the data source to be created.
Name :
CLAIM_DATA_V2_TRAIN
Role :
Raw
Segment :
Notes :
< Back
Next >
Cancel
12:17 PM


Enterprise Miner - Create New Project -- Step 1 of 4 Select SAS Server
File Edit View Actions Options Window Help
Create New Project -- Step 1 of 4 Select SAS Server
Data Sources
CLAIM_DATA_V2_TRAIN
Diagrams
Create SAS Dataset
MODULE 1
Model Packages
Property
Value
General
Node ID
Ids
Imported Data
Exported Data
Notes
Train
Output Type
View
Role
Raw
Rerun
No
Summarize
No
Drop Map Variables
Yes
Columns
Variables
Decisions
Refresh Metadata
Advisor
Basic
Advanced Options
Data
Data Selection
Data Source
Sample
Default
General
General Properties
Sample
Explore
Modify
Model
Assess
Utility
Credit Scoring
HPDM
Applications
Text Mining
Time Series
MODULE 1
CLAIM_DATA _V2_TRAIN
Diagram
Log
12:31 PM
1/18/2023

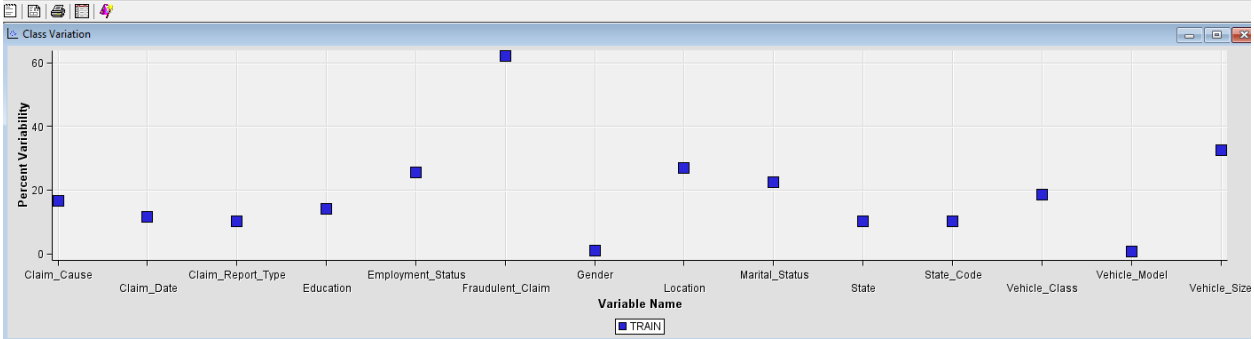

Enterprise Miner - Create New Project -- Step 1 of 4 Select SAS Server
File Edit View Actions Options Window Help
Create New Project -- Step 1 of 4 Select SAS Server
Data Sources
CLAIM_DATA_V2_TRAIN
Diagrams
Create SAS Dataset
MODULE 1
Model Packages
Sample Explore Modify Model Assess Utility Credit Scoring HPDM Applications Text Mining Time Series
MODULE 1
CLAIM_DATA _V2_TRAIN
Data Partition
StatExplore
Property
Value
General
Node ID
Stat
Imported Data
Exported Data
Notes
Train
Variables
Data
Number of Observations
100000
Validation
No
Test
No
Standard Reports
Interval Distributions
Yes
Class Distributions
Yes
Level Summary
Yes
Use Segment Variables
No
Cross-Tabulation
Variable Selection
Hide Rejected Variables
Yes
Number of Selected Variable
1000
General
General Properties
100%
12:39 PM
1/18/2023


Results - Node: Data Partition Diagram: MODULE 1
File Edit View Window
Output
Time: 20:37:40
* Training Output
Variable Summary
Measurement Frequency
Role Level Count
INPUT INTERVAL 8
INPUT NOMINAL 14
Partition Summary
Number of
Type Data Set Observations
DATA EMWS2.Ids_DATA 5001
TRAIN EMWS2.Part_TRAIN 2308
VALIDATE EMWS2.Part_VALIDATE 1539
TEST EMWS2.Part_TEST 1154
* Score Output
* Report Output
2:17 PM
1/18/2023


Results - Node: CLAIM_DATA_V2_TRAIN Diagram: MODULE 1
File Edit View Window
Output
--
User: u61893675
Date: 18 January 2023
Time: 20:37:34
--
* Training Output
--
Variable Summary
Measurement Frequency
Role Level Count
INPUT INTERVAL 8
INPUT NOMINAL 14
Variables
Variable Name
Role
Measurement Level
Annual Premium Input Interval
Claim Amount Input Interval
Claim Cause Input Nominal
Claim Date Input Nominal
Claim Report Type Input Nominal
Claimant Number Input Interval
Education Input Nominal
Employment Status Input Nominal
Fraudulent Claim Input Nominal
Gender Input Nominal
Income Input Interval
Location Input Nominal
Marital Status Input Nominal
Monthly Premium Input Interval
Months Since Last Claim Input Interval
Months Since Policy Inception Input Interval
12:57 PM
1/18/2023

Results - Node: StatExplore Diagram: MODULE 1

File Edit View Window



Output

```
*------------------------------------------------------------*
User:                  u61893675
Date:                  18 January 2023
Time:                  20:38:17
*------------------------------------------------------------*
* Training Output
*------------------------------------------------------------*



Variable Summary

         Measurement    Frequency
Role        Level         Count

INPUT     INTERVAL           8
INPUT     NOMINAL           14
```

Output

Class Variable Summary Statistics
(maximum 500 observations printed)

Data Role=TRAIN

Data Role	Variable Name	Role	Number of Levels	Missing	Mode	Mode Percentage	Mode2	Mode2 Percentage
TRAIN	Claim_Cause	INPUT	5	0	Collision	41.53	Hail	31.55
TRAIN	Claim_Date	INPUT	3	0	12/01/2018	39.99	12/15/2018	39.99
TRAIN	Claim_Report_Type	INPUT	4	0	Agent	37.77	Branch	28.21
TRAIN	Education	INPUT	6	11	College	29.83	Bachelor	29.81
TRAIN	Employment_Status	INPUT	5	0	Employed	62.51	Unemployed	25.39
TRAIN	Fraudulent_Claim	INPUT	2	0	N	93.86	Y	6.14
TRAIN	Gender	INPUT	2	0	M	50.61	F	49.39
TRAIN	Location	INPUT	4	3	Suburban	63.11	Rural	19.62
TRAIN	Marital_Status	INPUT	3	0	Married	58.59	Single	26.41
TRAIN	State	INPUT	5	0	Iowa	30.91	Missouri	28.91

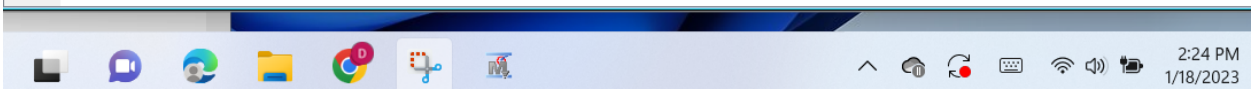
Output

(maximum 500 observations printed)

Data Role=TRAIN

Variable	Role	Mean	Standard Deviation	Non Missing	Missing	Minimum	Median	Maximum	Skewness	Kurtosis
Annual_Premium	INPUT	1134.368	313.4503	5001	0	600	1140	1680	0.016055	-1.17369
Claim_Amount	INPUT	787.7633	655.9633	5001	0	189.8684	577.3521	7422.852	2.922369	12.62942
Claimant__Number	INPUT	3501	1443.809	5001	0	1001	3500	6001	0	-1.2
Income	INPUT	41310.45	227690.4	5001	0	0	34621	15967801	68.48667	4790.542
Monthly_Premium	INPUT	94.53069	26.12086	5001	0	50	95	140	0.016055	-1.17369
Months_Since_Last_Claim	INPUT	15.0042	11.13965	5001	0	0	13	60	0.666671	0.052198
Months_Since_Policy_Inception	INPUT	48.23495	28.09665	5001	0	0	48	99	0.051865	-1.14433
Outstanding_Balance	INPUT	23728.15	13827.4	4992	9	4	23988	47996	0.005256	-1.20029

--
* Score Output



Reference

Richard V. McCarthy; Mary M. McCarthy; Wendy Ceccucci, 2022. *Applying Predictive Analytics Finding Value in Data*. Second edition.