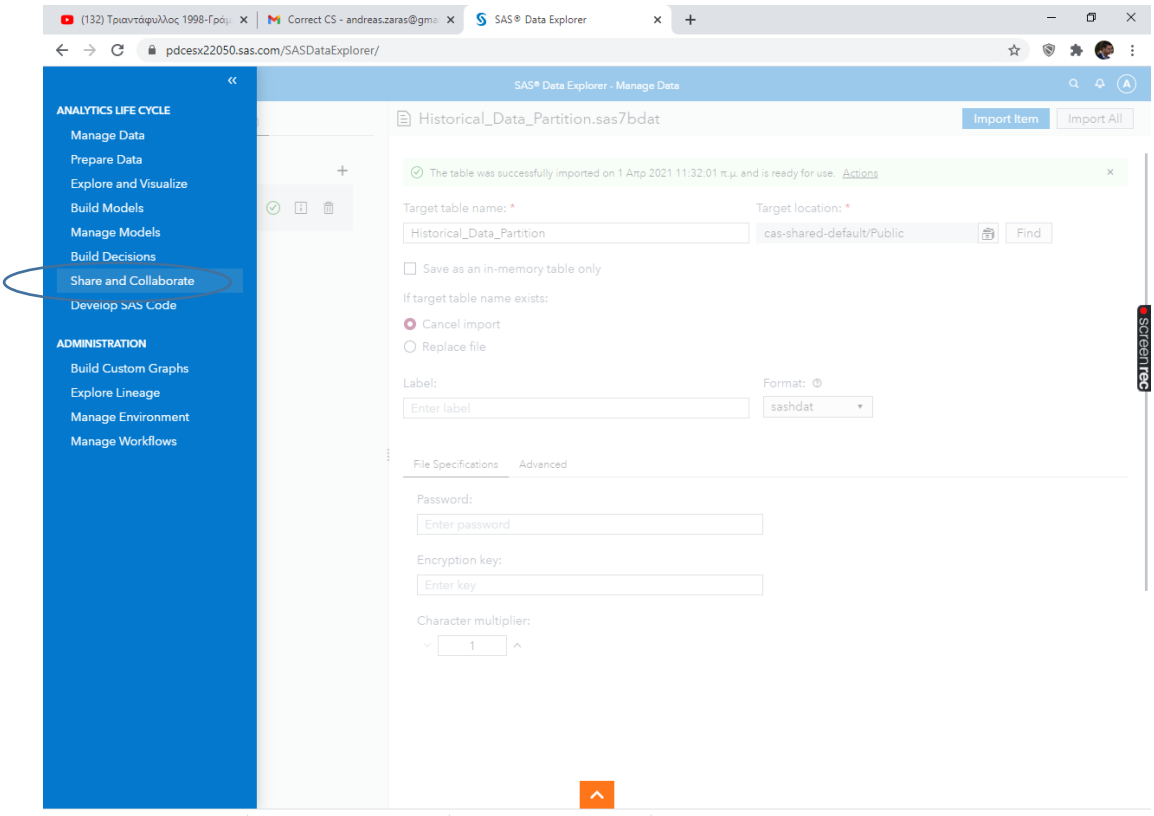
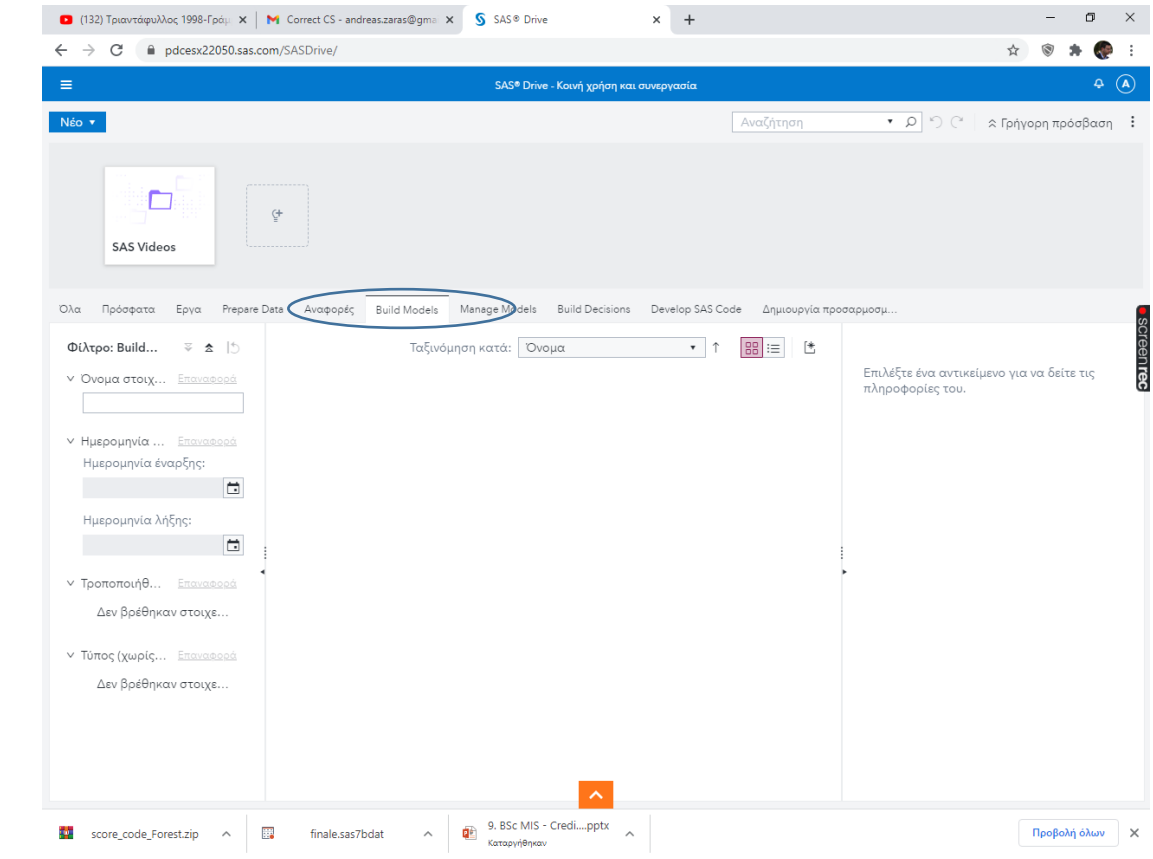
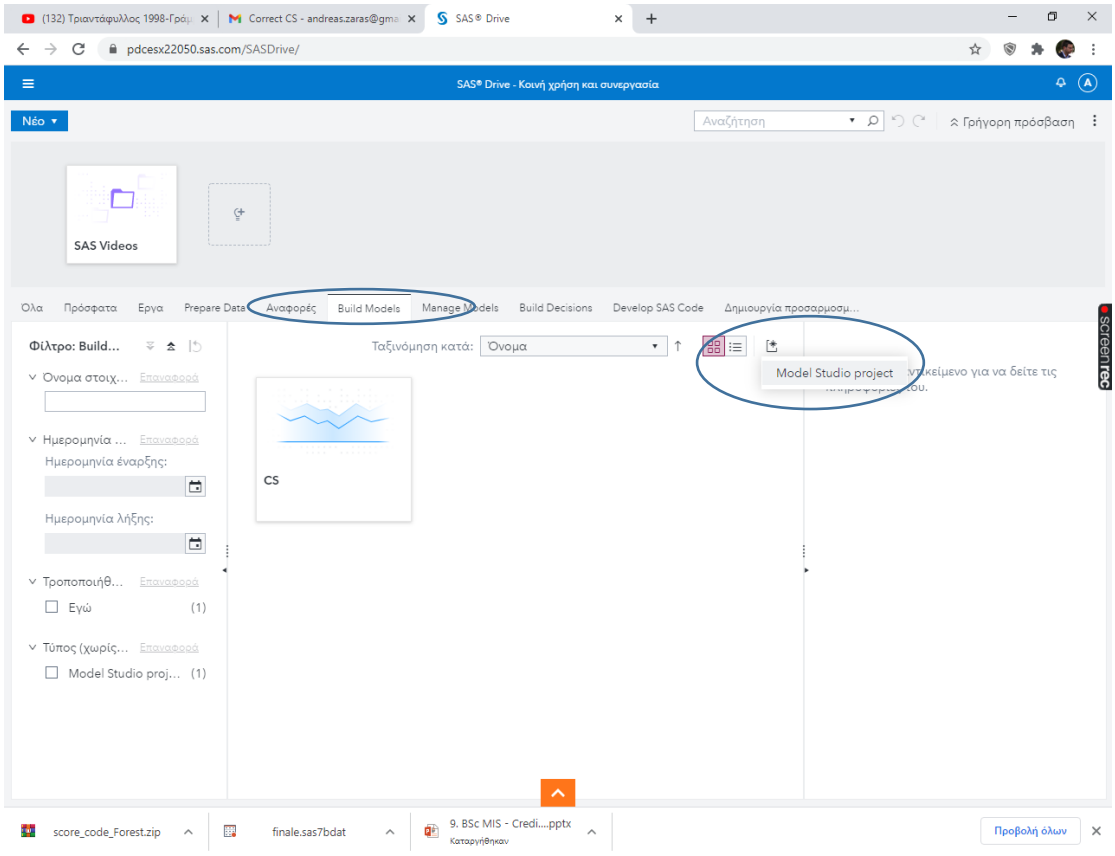
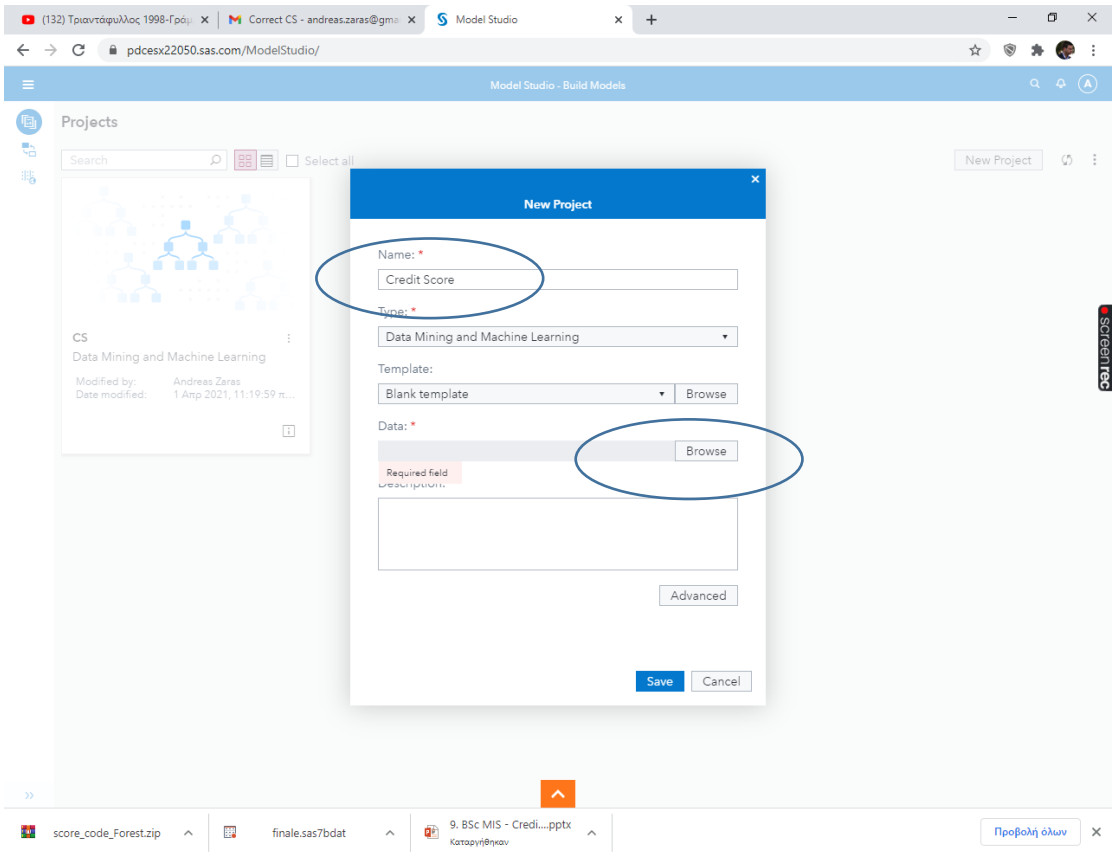
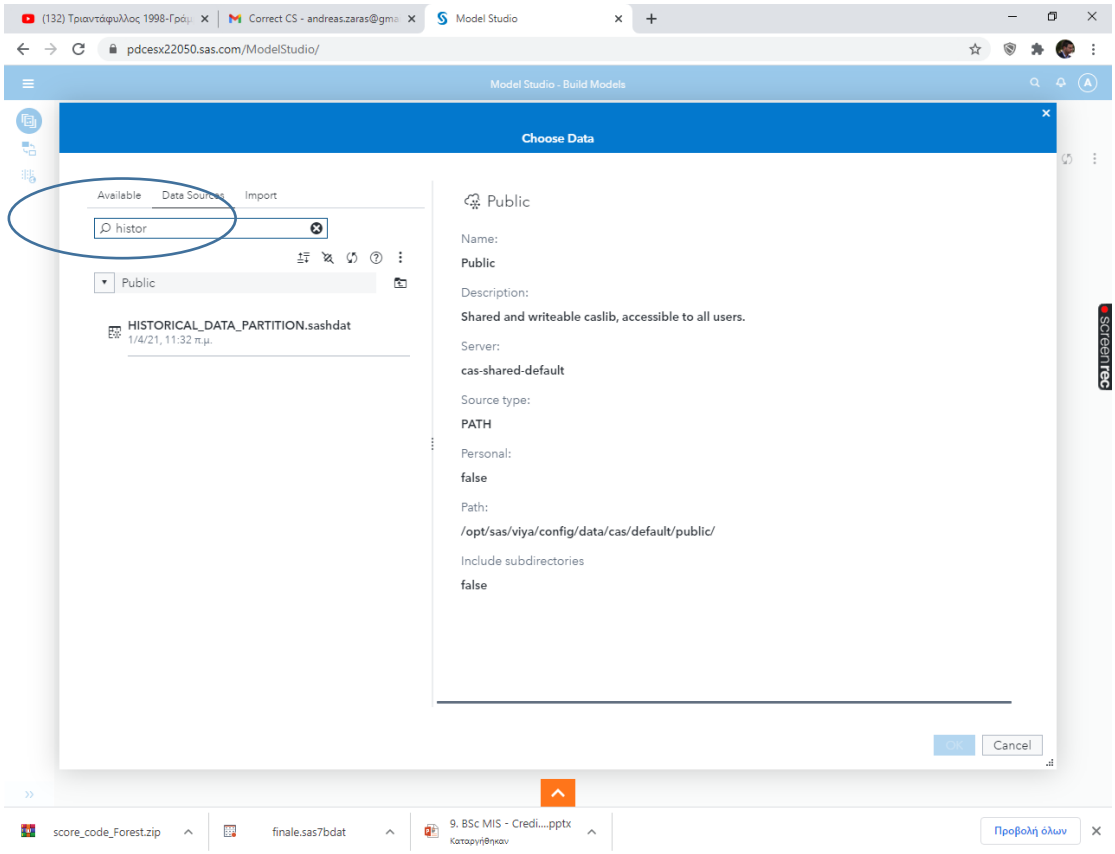
Import the Historical Data Partition set and the Credit Score Data Set by following the Upload Data to VFL pdf document. Then:

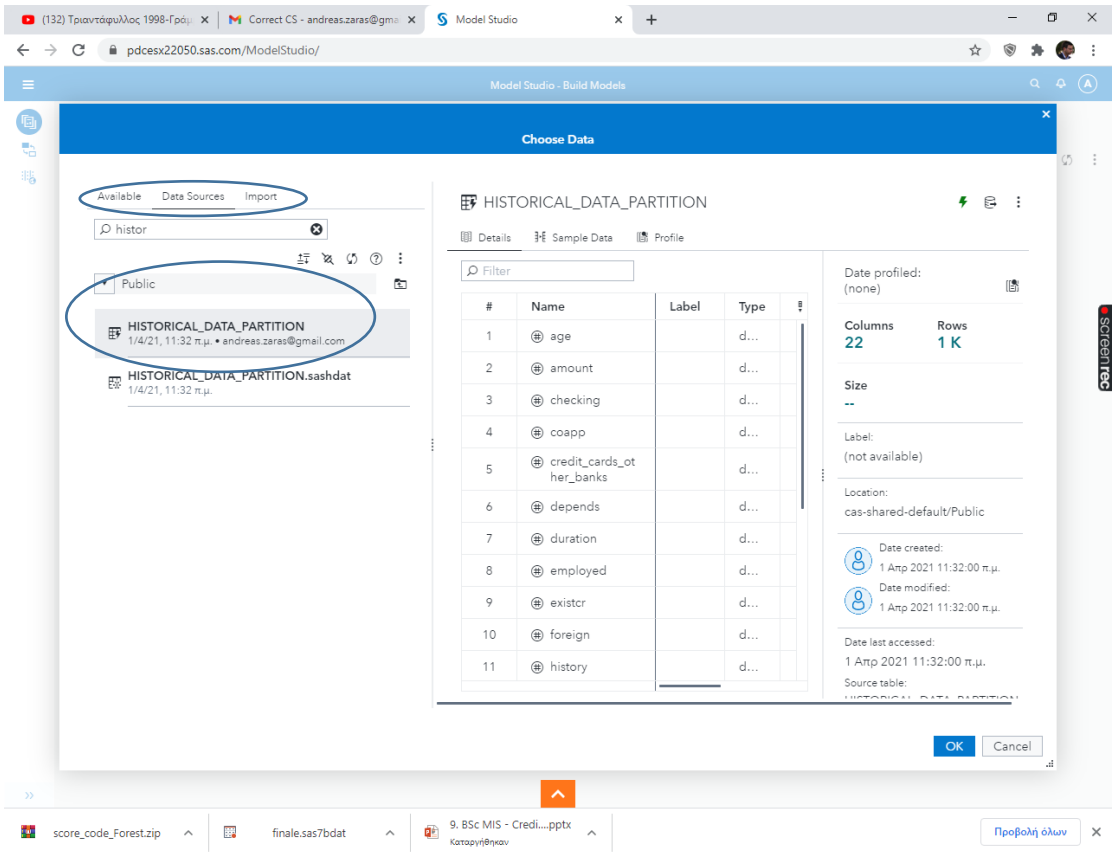


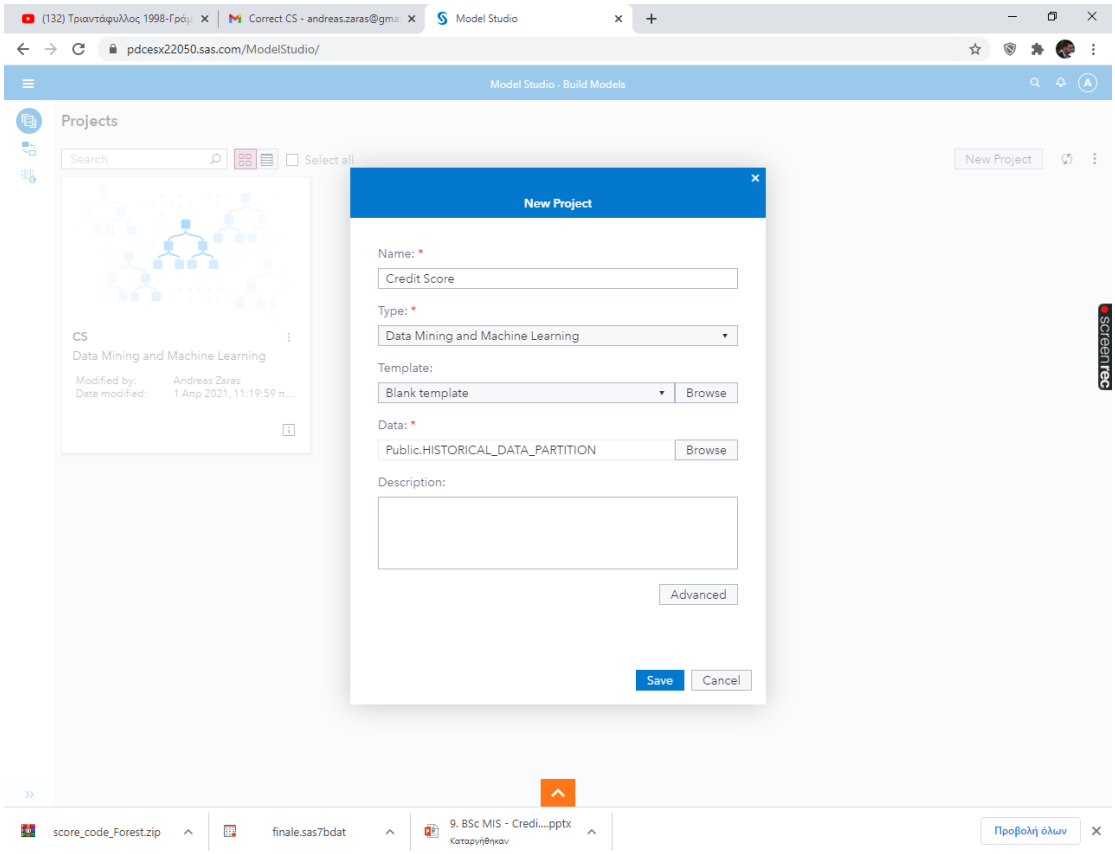



SAS® Drive
pdcesx22050.sas.com/SASDrive/
SAS® Drive - Κοινή χρήση και συνεργασία
Νέο
Αναζήτηση
Γρήγορη πρόσβαση
SAS Videos
Όλα
Πρόσφατα
Εργα
Prepare Data
Αναφορές
Build Models
Manage Models
Build Decisions
Develop SAS Code
Δημιουργία προσαρμοσμ...
Φίλτρο: Build...
Ταξινόμηση κατά:
Όνομα
Model Studio project
CS
score_code_Forest.zip
finale.sas7bdat
9. BSc MIS - Credi....pptx
Προβολή όλων

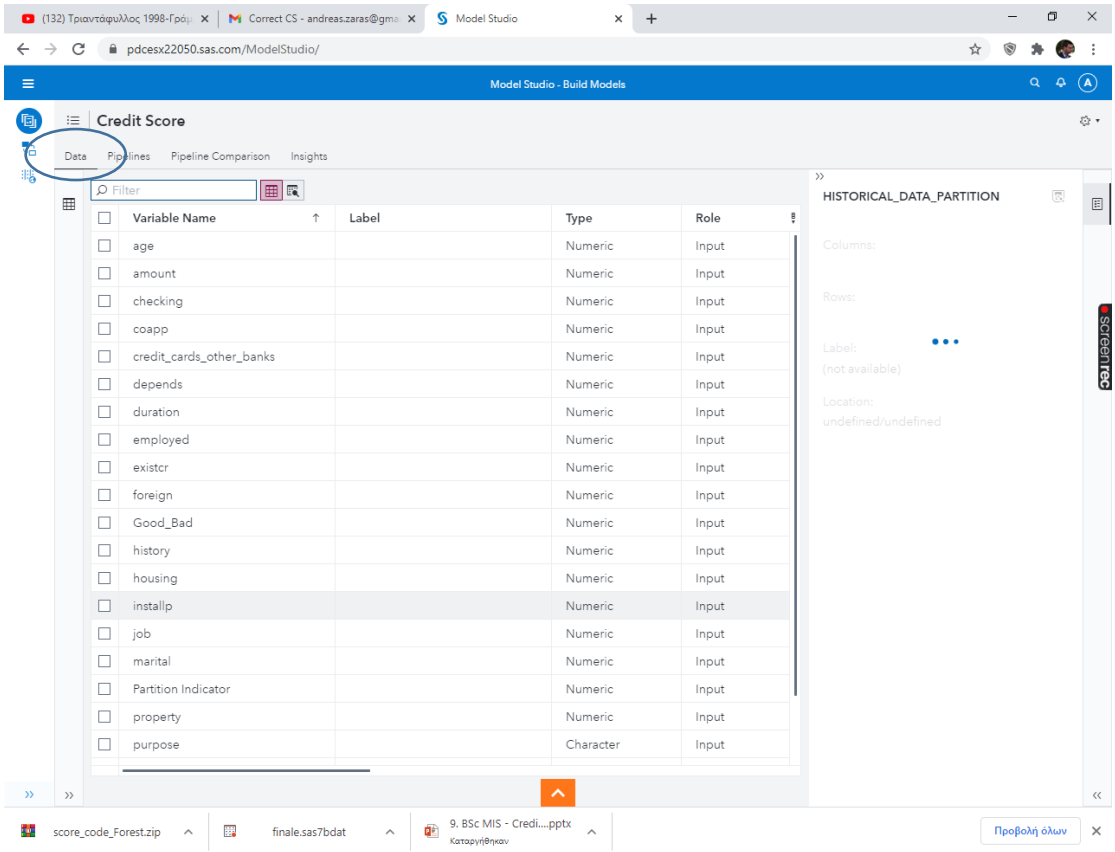

Model Studio
pdcesx22050.sas.com/ModelStudio/
Model Studio - Build Models
Projects
New Project
Name: *
Credit Score
Type: *
Data Mining and Machine Learning
Template:
Blank template
Browse
Data: *
Browse
Required field
Advanced
Save
Cancel
CS
Data Mining and Machine Learning
score_code_Forest.zip
finale.sas7bdat
9. BSc MIS - Credi....pptx
Προβολή όλων


Model Studio - Build Models
Choose Data
Available
Data Sources
Import
histor
Public
HISTORICAL_DATA_PARTITION.sashdat
1/4/21, 11:32 π.μ.
Public
Name:
Public
Description:
Shared and writeable caslib, accessible to all users.
Server:
cas-shared-default
Source type:
PATH
Personal:
false
Path:
/opt/sas/viya/config/data/cas/default/public/
Include subdirectories
false
OK
Cancel


Model Studio - Build Models
Choose Data
Available
Data Sources
Import
histor
Public
HISTORICAL_DATA_PARTITION
1/4/21, 11:32 π.μ. • andreas.zaras@gmail.com
HISTORICAL_DATA_PARTITION.sashdat
1/4/21, 11:32 π.μ.
HISTORICAL_DATA_PARTITION
Details
Sample Data
Profile
Filter
Name Label Type
1 age
2 amount
3 checking
4 coapp
5 credit_cards_other_banks
6 depends
7 duration
8 employed
9 existcr
10 foreign
11 history
Date profiled:
(none)
Columns
22
Rows
1 K
Size
--
Label:
(not available)
Location:
cas-shared-default/Public
Date created:
1 Απρ 2021 11:32:00 π.μ.
Date modified:
1 Απρ 2021 11:32:00 π.μ.
Date last accessed:
1 Απρ 2021 11:32:00 π.μ.
Source table:
OK
Cancel

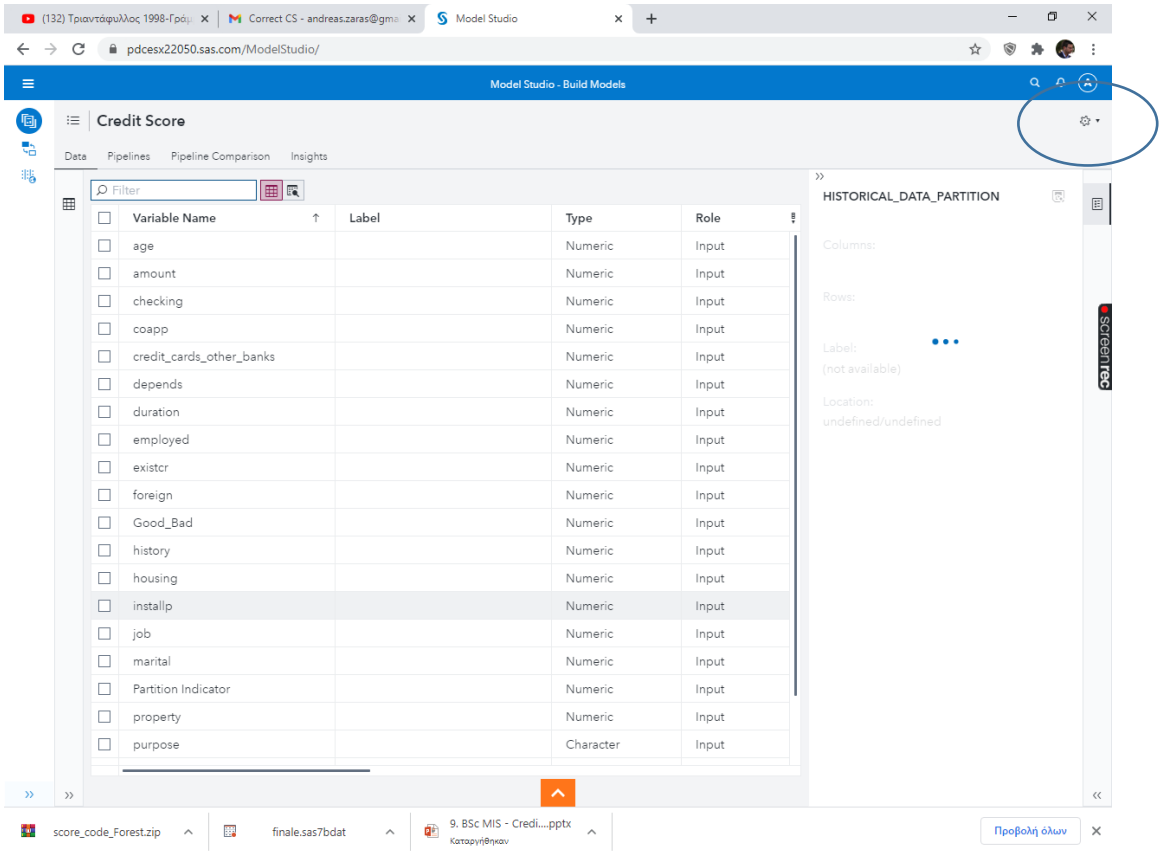


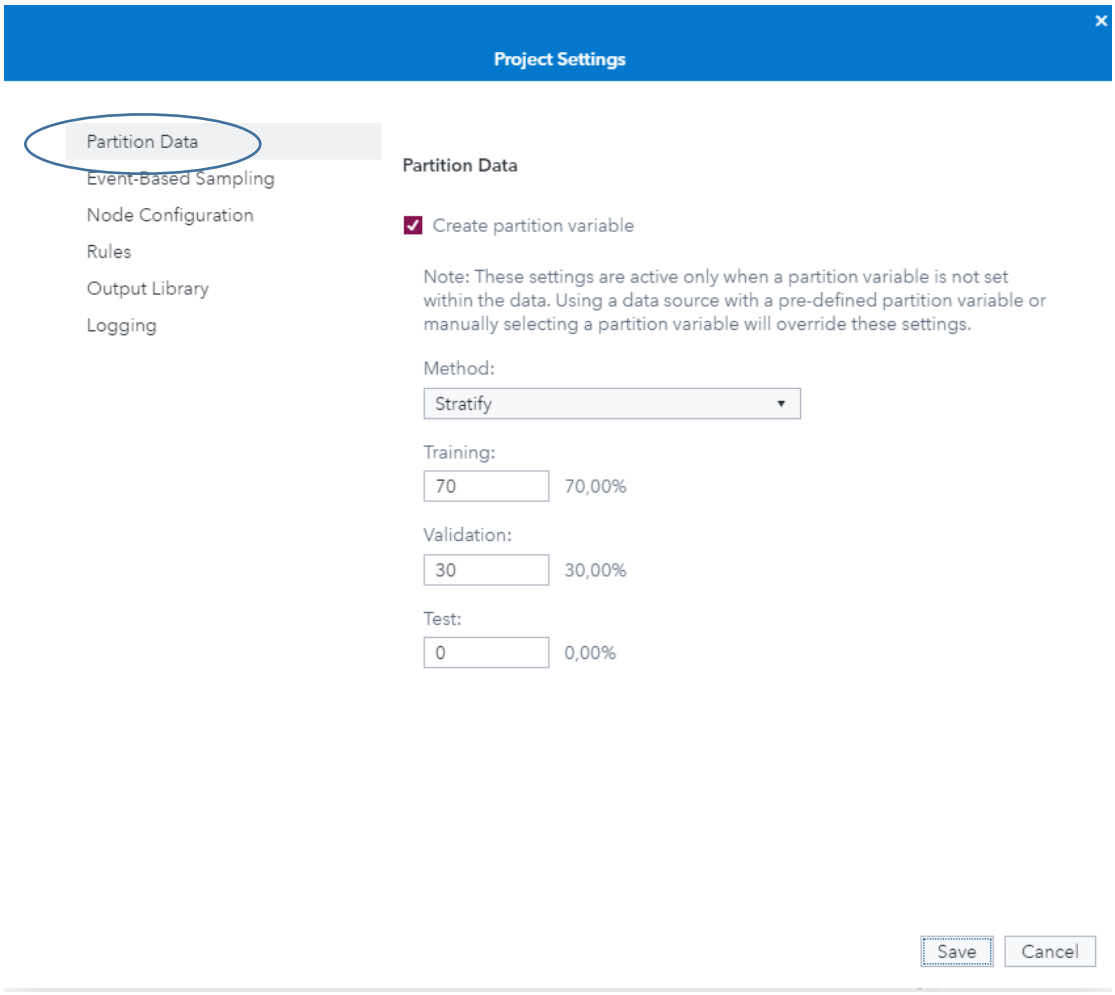
Set the roles and the measurement scales of the variables accoridn to the table in the next page.

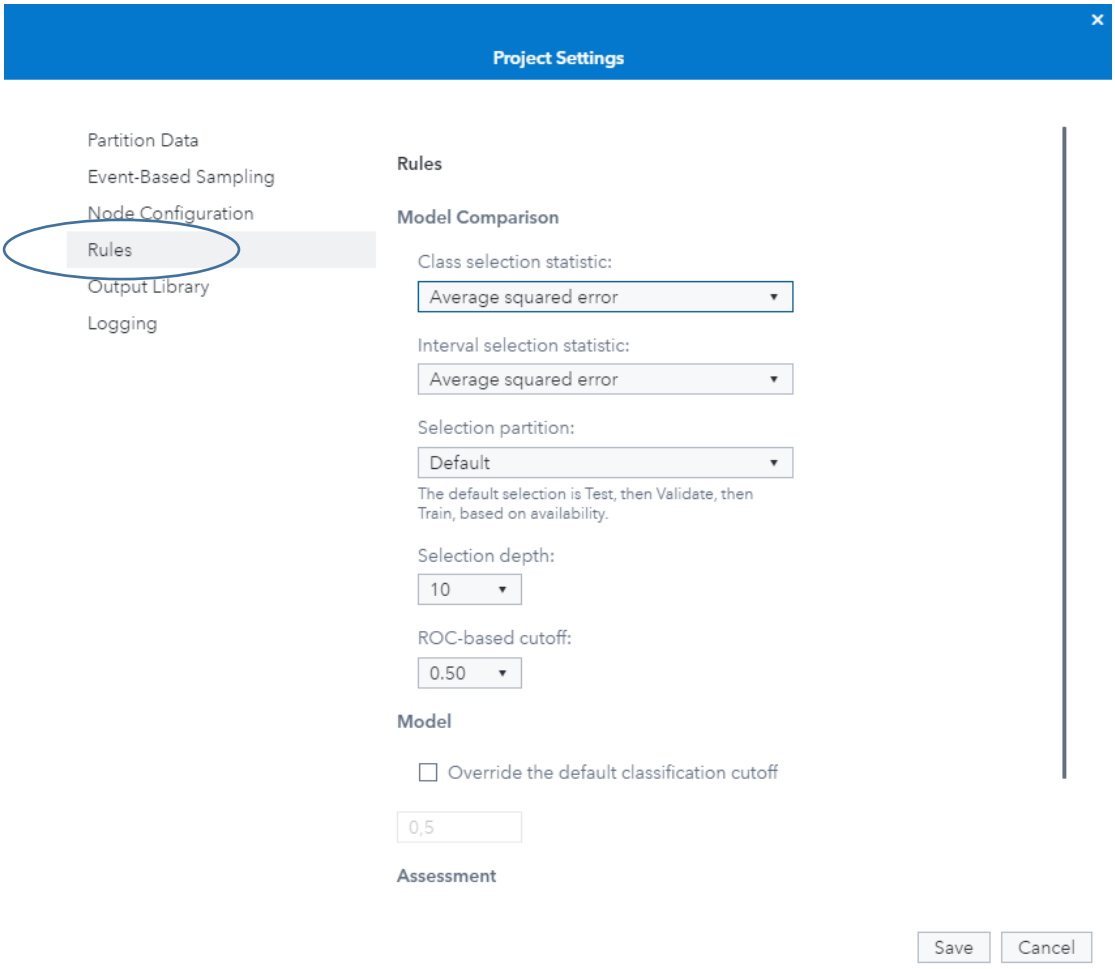


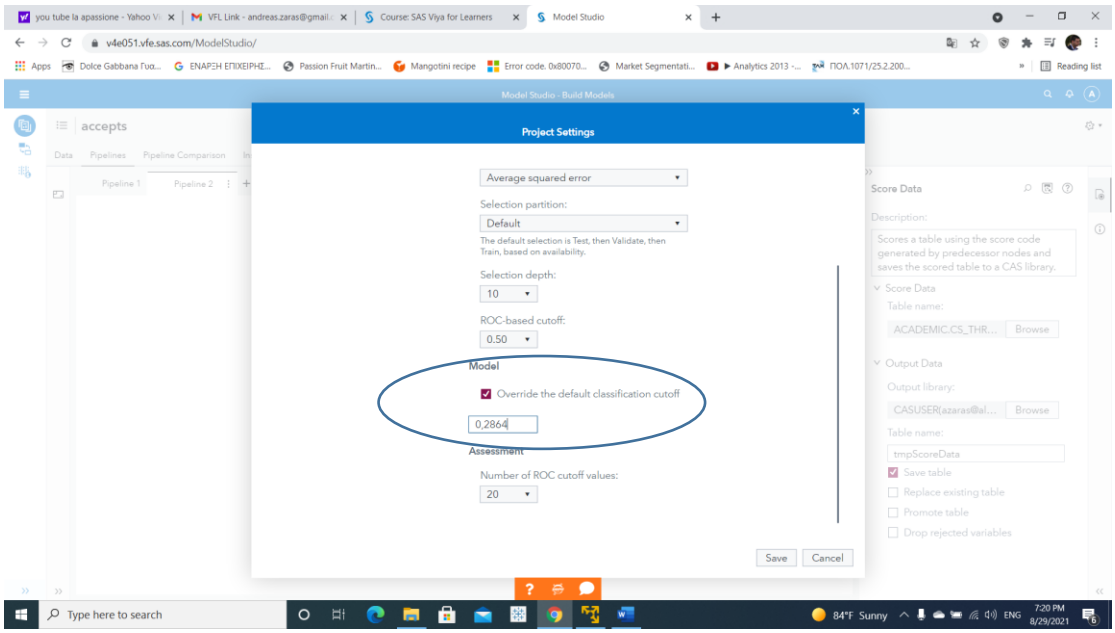
Variable Name	Label	Type	Role
age		Numeric	Input
amount		Numeric	Input
checking		Numeric	Input
coapp		Numeric	Input
credit_cards_other_banks		Numeric	Input
depends		Numeric	Input
duration		Numeric	Input
employed		Numeric	Input
existcr		Numeric	Input
foreign		Numeric	Input
Good_Bad		Numeric	Input
history		Numeric	Input
housing		Numeric	Input
installp		Numeric	Input
job		Numeric	Input
marital		Numeric	Input
Partition Indicator		Numeric	Input
property		Numeric	Input
purpose		Character	Input

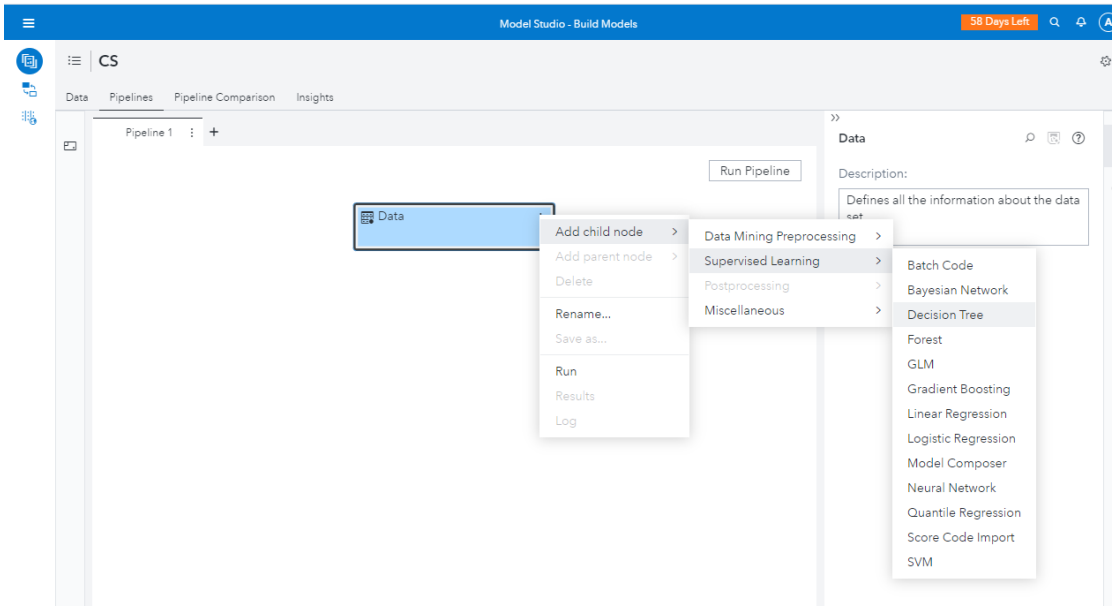
Variable Name	Variable Description	Variable Scale	Variable Role
Age	Age in years	Interval	Input
Amount	Amount of loan	Interval	Input
Checking	Status of existing checking account: 1: No Checking Account, 2: <$0, 3: 0 - <$200, 4: >=$200	Ordinal	Input
Coapp	Other applicants/ guarantors: 1: none, 2: co-applicant, 3: guarantor	Nominal	Input
Depends	Number of dependents	Interval	Input
Duration	Duration in months	Interval	Input
Employed	Presently employed since: 1: unemployed, 2: <1year, 3: 1 to <4 years, 4: 4 to <7 years, 5: >=7 years	Ordinal	Input
Existcr	Number of existing credits at this bank	Interval	Input
Foreign	Foreign worker: 1: yes, 2: no	Binary	Input
Housing	1: rent, 2: own, 3: for free	Nominal	Input
Installp	Installment rate in percentage of disposable income	Interval	Input
Job	1: Unemployed/ Unskilled - non resident, 2: Unskilled - resident, 3: Skilled employee/ official, 4: Manager/ Shelf Employed/ Highly Qualified employee officer	Nominal	Input
Marital	Marital Status: 1: Male: Divorced/ Separated, 2: Female: Divorced/ Separated/ Married 3: Male: Single, 4: Male married/ windowed, 5: Female: Single	Nominal	Input
Purpose	Purpose of loan: 0: Car (new), 1: Car (used), 2: Furniture/ Equipment, 3: Radio/ TV, 4: Domestic appliances, 5: House Repairs, 6: Education, 7: Vacation, 8: Retraining	Nominal	Input
Other Loan Obligations	1: Bank, 2: Stores, 3: None	Nominal	Input
Savings	Savings Account: 1: <$100, 2: 100-<$500, 3: $500-<$1000, 4: >= $1000,	Ordinal	Input
Telephone	1: None, 2: Yes, registered under the customer's name	Binary	Input
Credit Cards in Other Banks	1: No Credit Cards, 2: 1 Credit Card, 3: 2 Credit Cards, 4: >= 2 Credit Cards	Ordinal	Input
Good/ Bad	Good/ Bad payer	Binary	Target
Property	1: Real Estate, 2: Car, 3: Life Insurance, 4: No Property	Nominal	Input
Resident	Years beginning permanent residence	Interval	Input
History	0: No Credits Taken/ All Credits Paid Back Dully; 1: All Credits at this Bank Paid Back Dully; 2: Existing Credits Paid Back Dully Until Now; 3: Delay in Paying Off in the Past; 4: Critical Account/ Other Credits Existing (Not at this Bank)	Nominal	Input
Partition_ID	Partition Variable	Binary	Partition

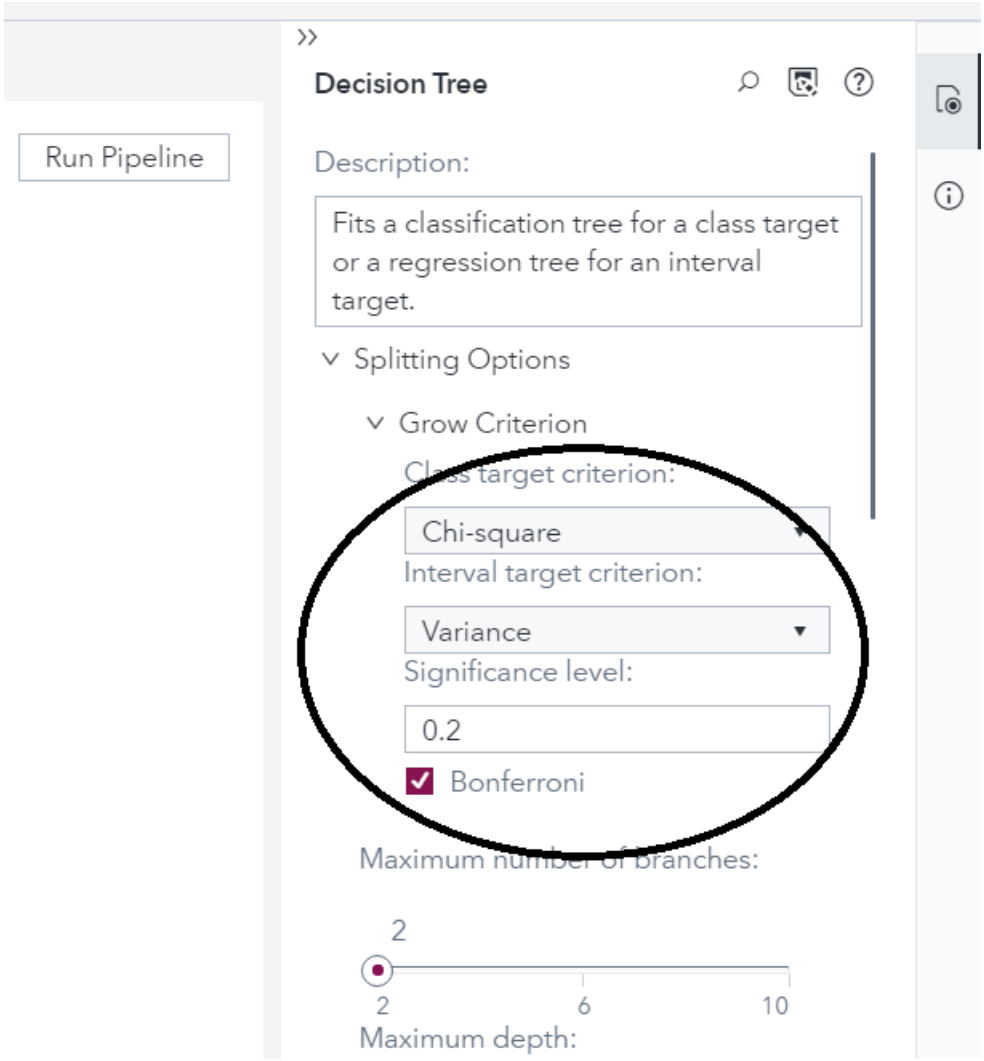

Model Studio - Build Models
Credit Score
Data
Pipelines
Pipeline Comparison
Insights
Filter
Variable Name
Label
Type
Role
age Numeric Input
amount Numeric Input
checking Numeric Input
coapp Numeric Input
credit_cards_other_banks Numeric Input
depends Numeric Input
duration Numeric Input
employed Numeric Input
existcr Numeric Input
foreign Numeric Input
Good_Bad Numeric Input
history Numeric Input
housing Numeric Input
installp Numeric Input
job Numeric Input
marital Numeric Input
Partition Indicator Numeric Input
property Numeric Input
purpose Character Input
HISTORICAL_DATA_PARTITION
Columns:
Rows:
Label:
(not available)
Location:
undefined/undefined
score_code_Forest.zip
finale.sas7bdat
9. BSc MIS - Credi....pptx
Προβολή όλων

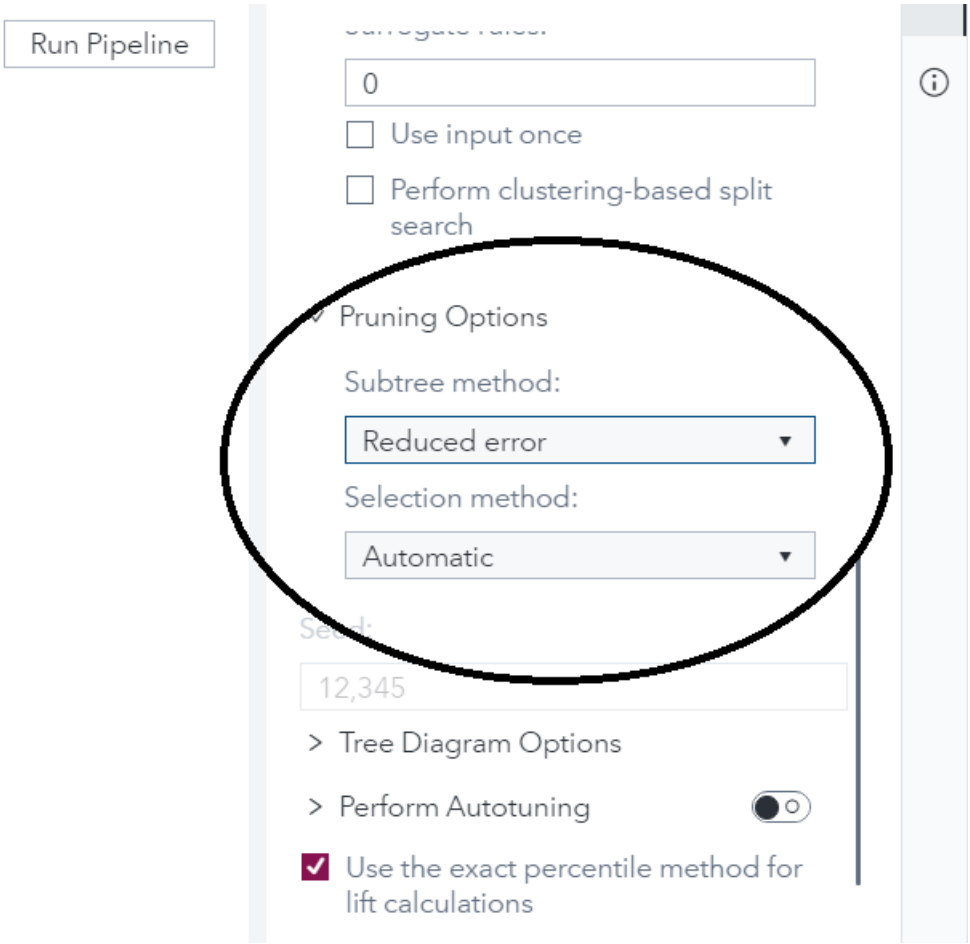

Project Settings
Partition Data
Event-Based Sampling
Node Configuration
Rules
Output Library
Logging
Partition Data
Create partition variable
Note: These settings are active only when a partition variable is not set within the data. Using a data source with a pre-defined partition variable or manually selecting a partition variable will override these settings.
Method:
Stratify
Training:
70
70,00%
Validation:
30
30,00%
Test:
0
0,00%
Save
Cancel

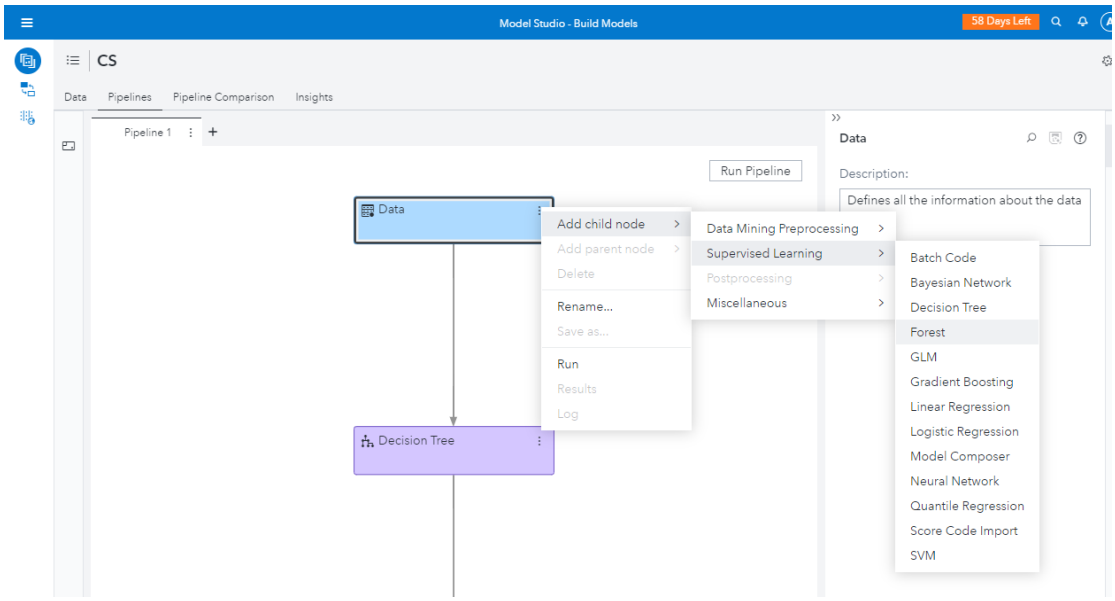

Project Settings
Partition Data
Event-Based Sampling
Node Configuration
Rules
Output Library
Logging
Rules
Model Comparison
Class selection statistic:
Average squared error
Interval selection statistic:
Average squared error
Selection partition:
Default
The default selection is Test, then Validate, then Train, based on availability.
Selection depth:
10
ROC-based cutoff:
0.50
Model
Override the default classification cutoff
0,5
Assessment
Save
Cancel

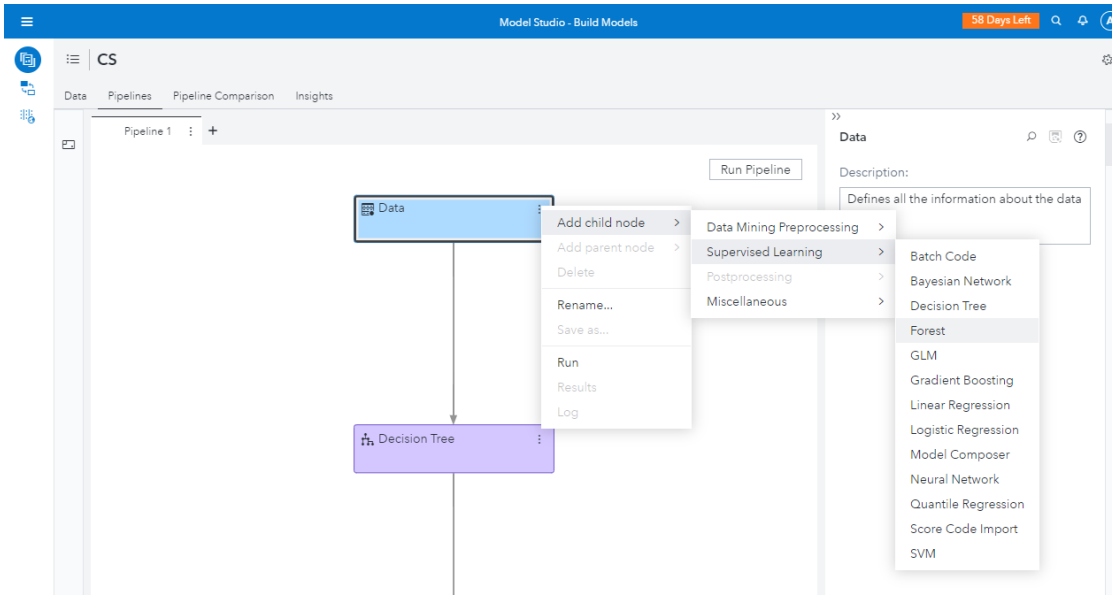

Project Settings
Average squared error
Selection partition:
Default
Selection depth:
10
ROC-based cutoff:
0.50
Model
Override the default classification cutoff
0,2864
Assessment
Number of ROC cutoff values:
20
Save
Cancel

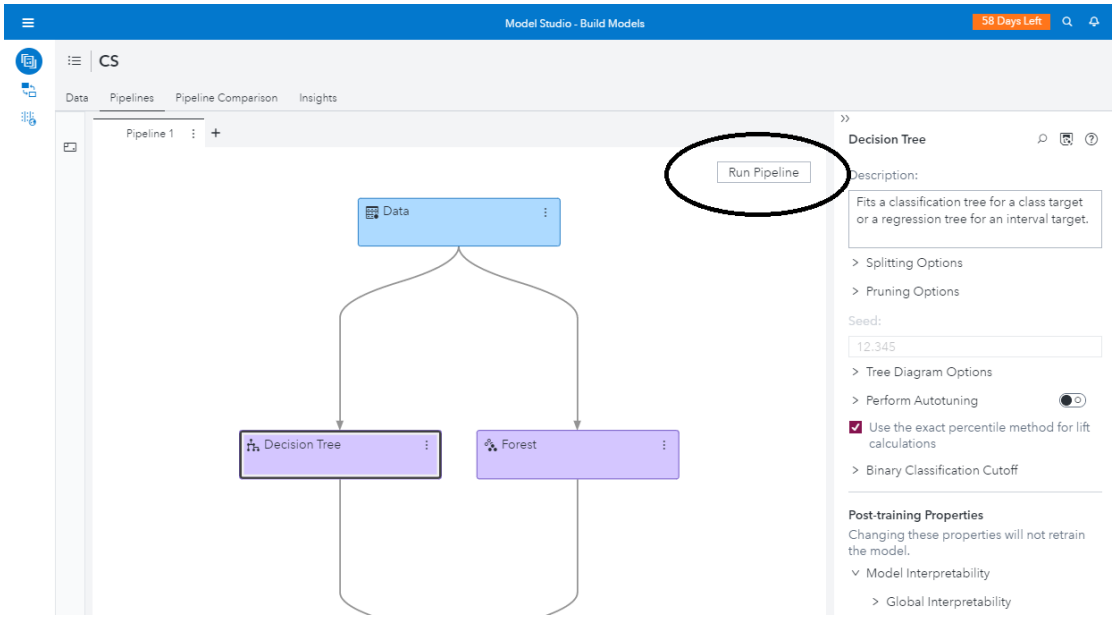

Model Studio - Build Models
58 Days Left
CS
Data
Pipelines
Pipeline Comparison
Insights
Pipeline 1
Run Pipeline
Data
Add child node
Add parent node
Delete
Rename...
Save as...
Run
Results
Log
Data Mining Preprocessing
Supervised Learning
Postprocessing
Miscellaneous
Batch Code
Bayesian Network
Decision Tree
Forest
GLM
Gradient Boosting
Linear Regression
Logistic Regression
Model Composer
Neural Network
Quantile Regression
Score Code Import
SVM
Data
Description:
Defines all the information about the data

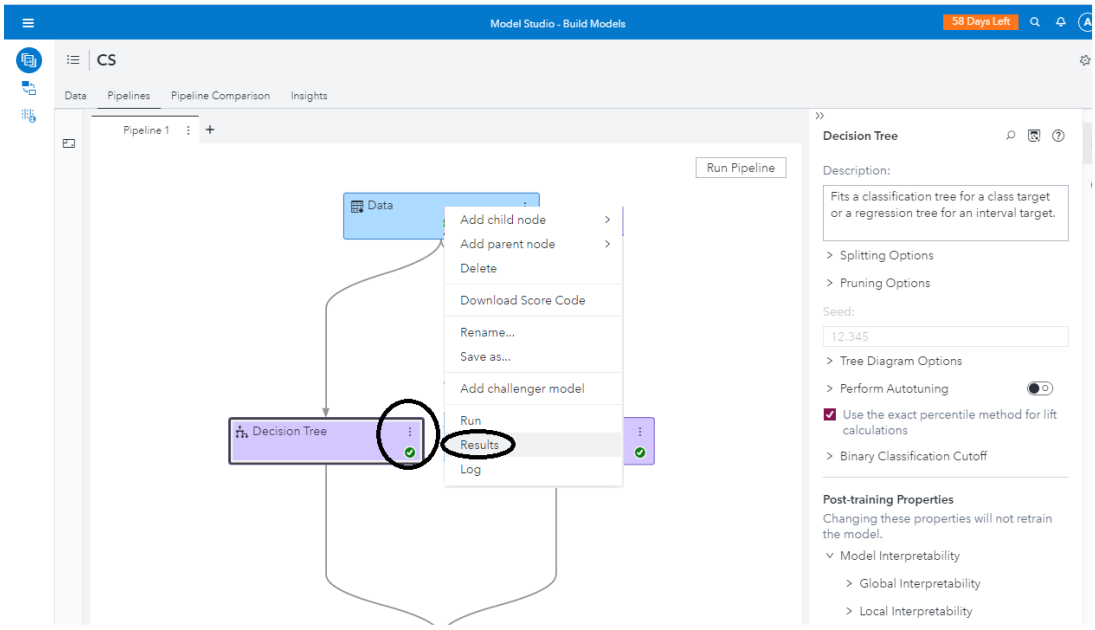

Run Pipeline
Decision Tree
Description:
Fits a classification tree for a class target or a regression tree for an interval target.
Splitting Options
Grow Criterion
Class target criterion:
Chi-square
Interval target criterion:
Variance
Significance level:
0.2
Bonferroni
Maximum number of branches:
2
2
6
10
Maximum depth:

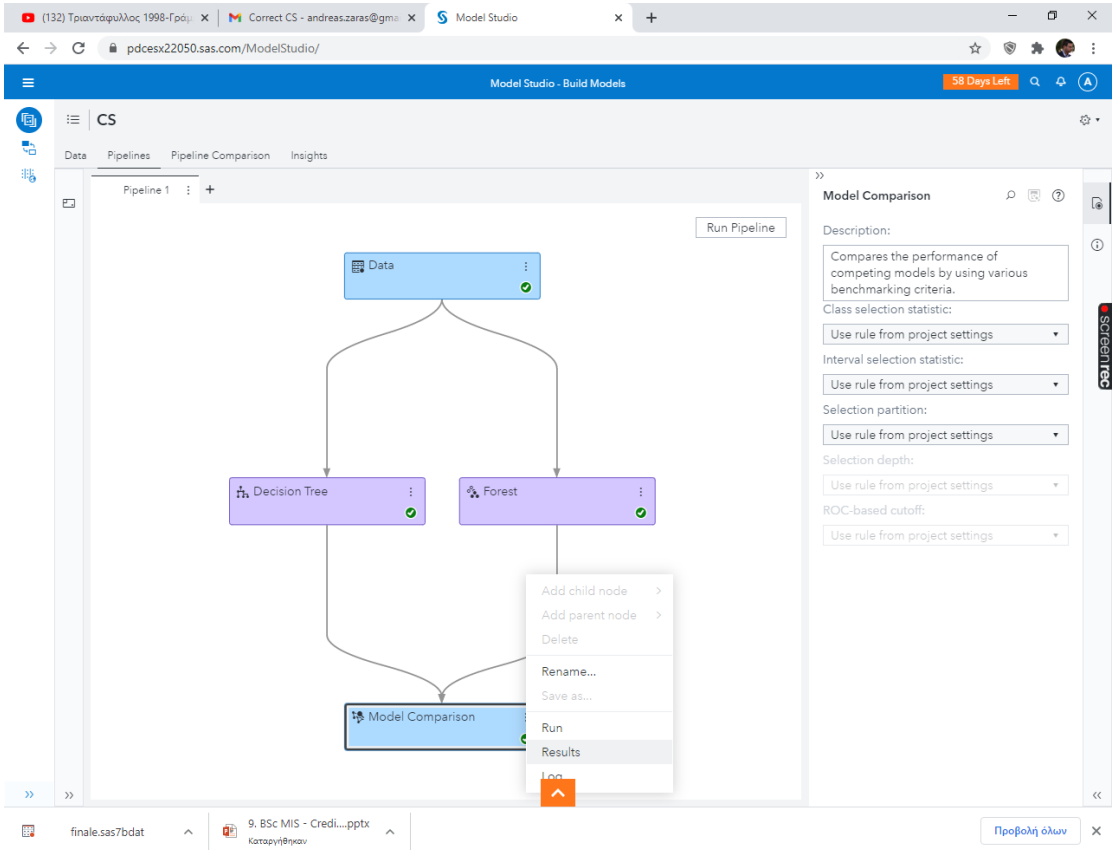

Run Pipeline
0
Use input once
Perform clustering-based split search
Pruning Options
Subtree method:
Reduced error
Selection method:
Automatic
12,345
Tree Diagram Options
Perform Autotuning
Use the exact percentile method for lift calculations

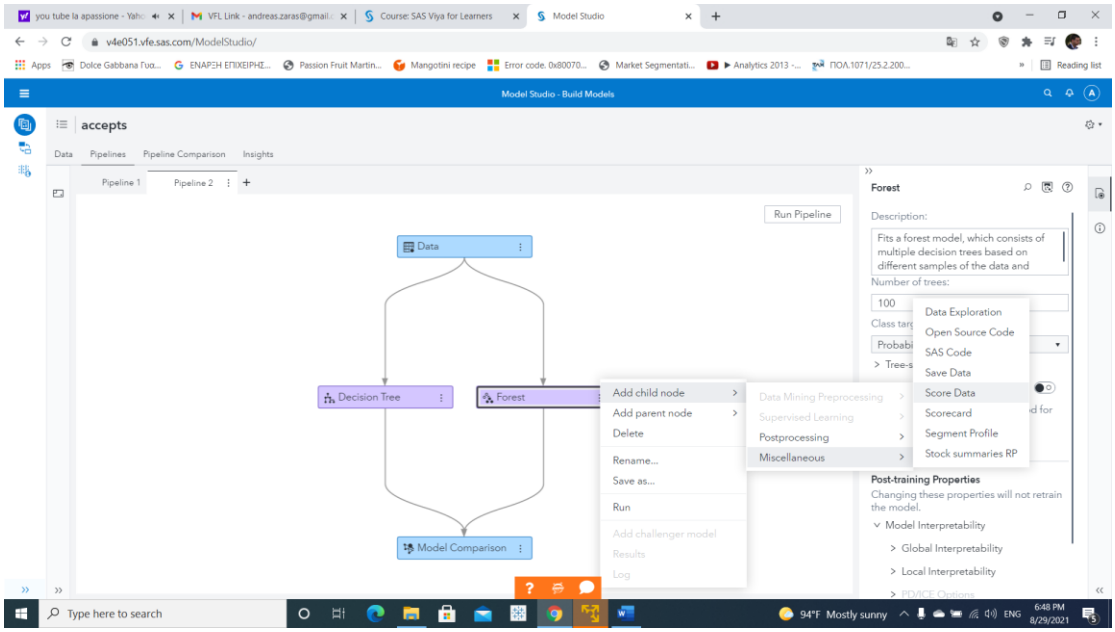

Model Studio - Build Models
58 Days Left
CS
Data
Pipelines
Pipeline Comparison
Insights
Pipeline 1
Run Pipeline
Data
Add child node
Add parent node
Delete
Rename...
Save as...
Run
Results
Log
Data Mining Preprocessing
Supervised Learning
Postprocessing
Miscellaneous
Batch Code
Bayesian Network
Decision Tree
Forest
GLM
Gradient Boosting
Linear Regression
Logistic Regression
Model Composer
Neural Network
Quantile Regression
Score Code Import
SVM
Decision Tree
Data
Description:
Defines all the information about the data

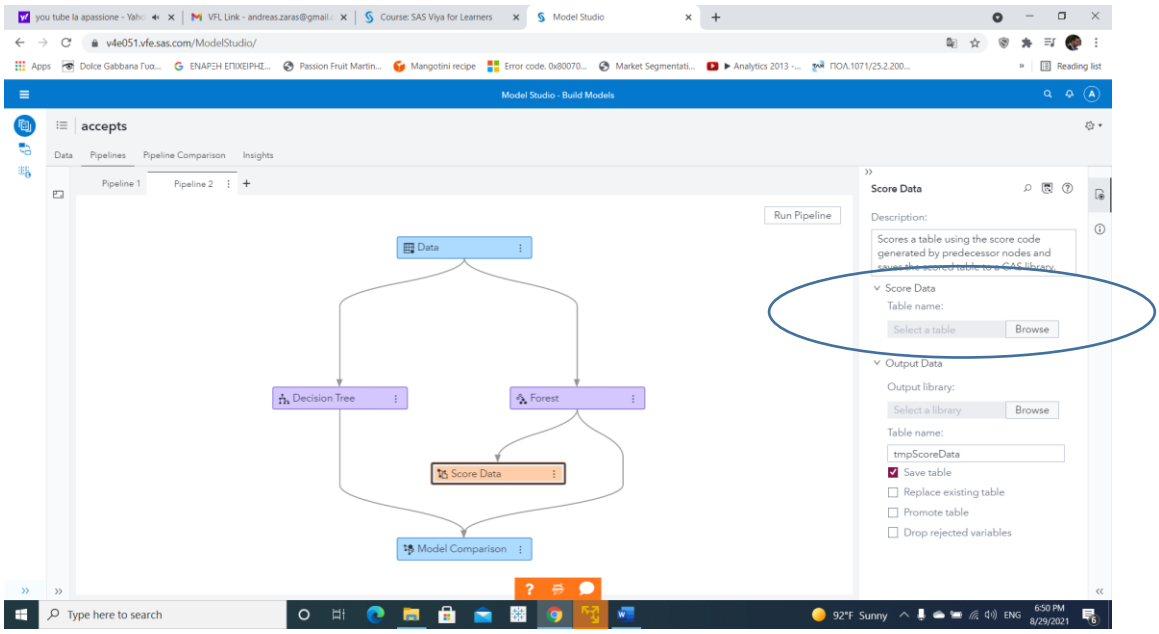

Model Studio - Build Models
58 Days Left
CS
Data
Pipelines
Pipeline Comparison
Insights
Pipeline 1
Run Pipeline
Data
Add child node
Add parent node
Delete
Rename...
Save as...
Run
Results
Log
Data Mining Preprocessing
Supervised Learning
Postprocessing
Miscellaneous
Batch Code
Bayesian Network
Decision Tree
Forest
GLM
Gradient Boosting
Linear Regression
Logistic Regression
Model Composer
Neural Network
Quantile Regression
Score Code Import
SVM
Decision Tree
Data
Description:
Defines all the information about the data


Model Studio - Build Models
58 Days Left
CS
Data
Pipelines
Pipeline Comparison
Insights
Pipeline 1
Run Pipeline
Data
Decision Tree
Forest
Decision Tree
Description:
Fits a classification tree for a class target or a regression tree for an interval target.
Splitting Options
Pruning Options
Seed:
12345
Tree Diagram Options
Perform Autotuning
Use the exact percentile method for lift calculations
Binary Classification Cutoff
Post-training Properties
Changing these properties will not retrain the model.
Model Interpretability
Global Interpretability

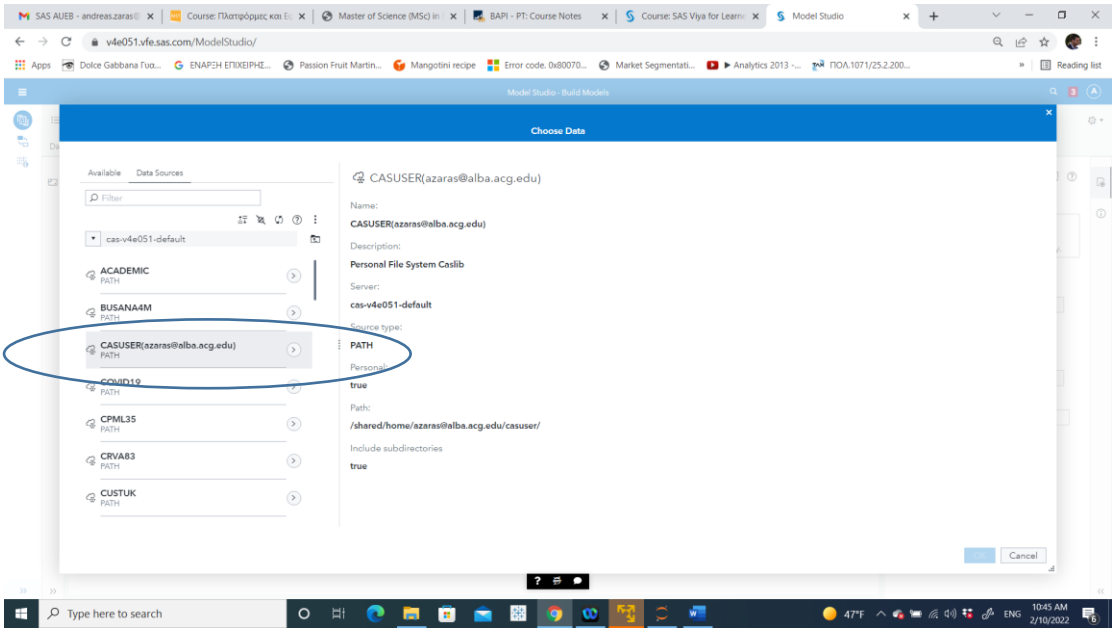

Model Studio - Build Models
58 Days Left
CS
Data
Pipelines
Pipeline Comparison
Insights
Pipeline 1
Run Pipeline
Data
Add child node
Add parent node
Delete
Download Score Code
Rename...
Save as...
Add challenger model
Run
Results
Log
Decision Tree
Decision Tree
Description:
Fits a classification tree for a class target or a regression tree for an interval target.
Splitting Options
Pruning Options
Seed:
12345
Tree Diagram Options
Perform Autotuning
Use the exact percentile method for lift calculations
Binary Classification Cutoff
Post-training Properties
Changing these properties will not retrain the model.
Model Interpretability
Global Interpretability
Local Interpretability

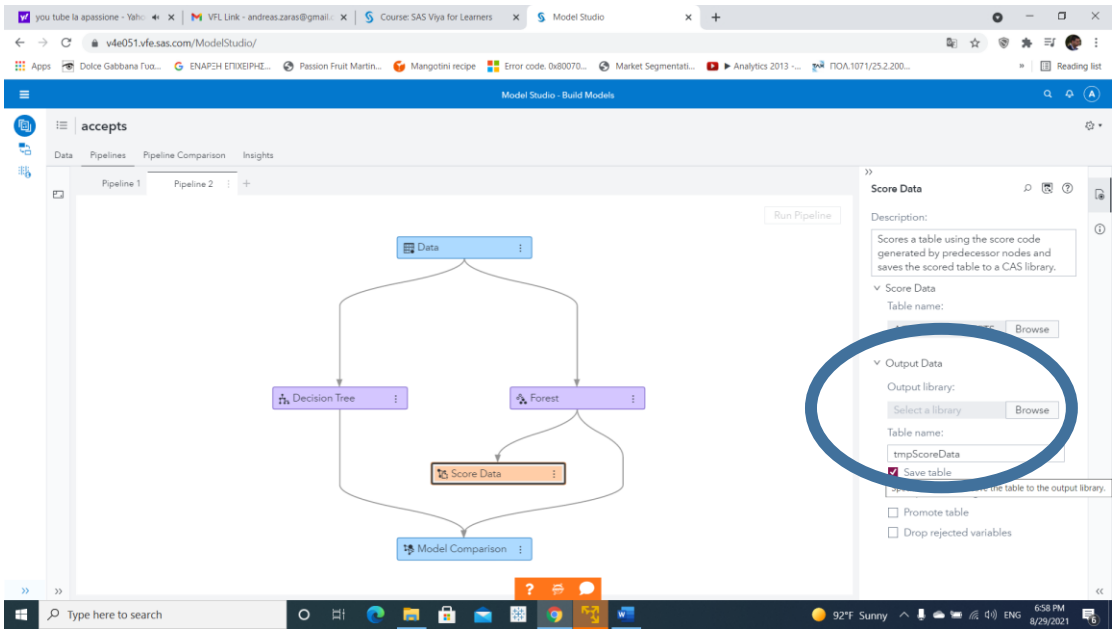

Model Studio
pdcesx22050.sas.com/ModelStudio/
Model Studio - Build Models
58 Days Left
CS
Data
Pipelines
Pipeline Comparison
Insights
Pipeline 1
Run Pipeline
Data
Decision Tree
Forest
Model Comparison
Add child node
Add parent node
Delete
Rename...
Save as...
Run
Results
Log
Model Comparison
Description:
Compares the performance of competing models by using various benchmarking criteria.
Class selection statistic:
Use rule from project settings
Interval selection statistic:
Use rule from project settings
Selection partition:
Use rule from project settings
Selection depth:
Use rule from project settings
ROC-based cutoff:
Use rule from project settings
finale.sas7bdat
9. BSc MIS - Credi....pptx

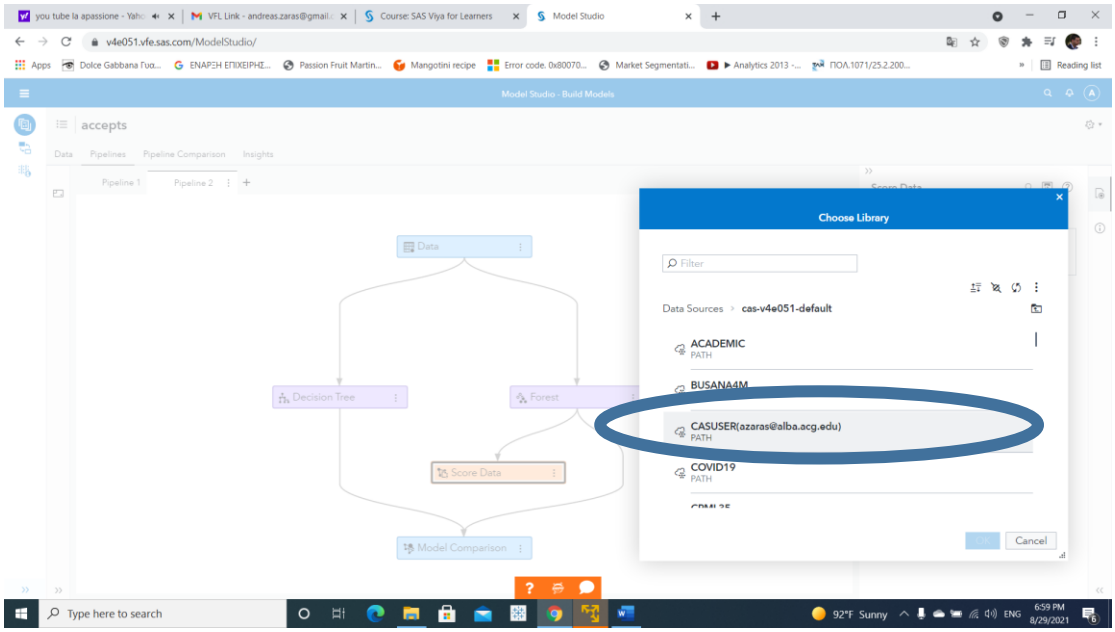


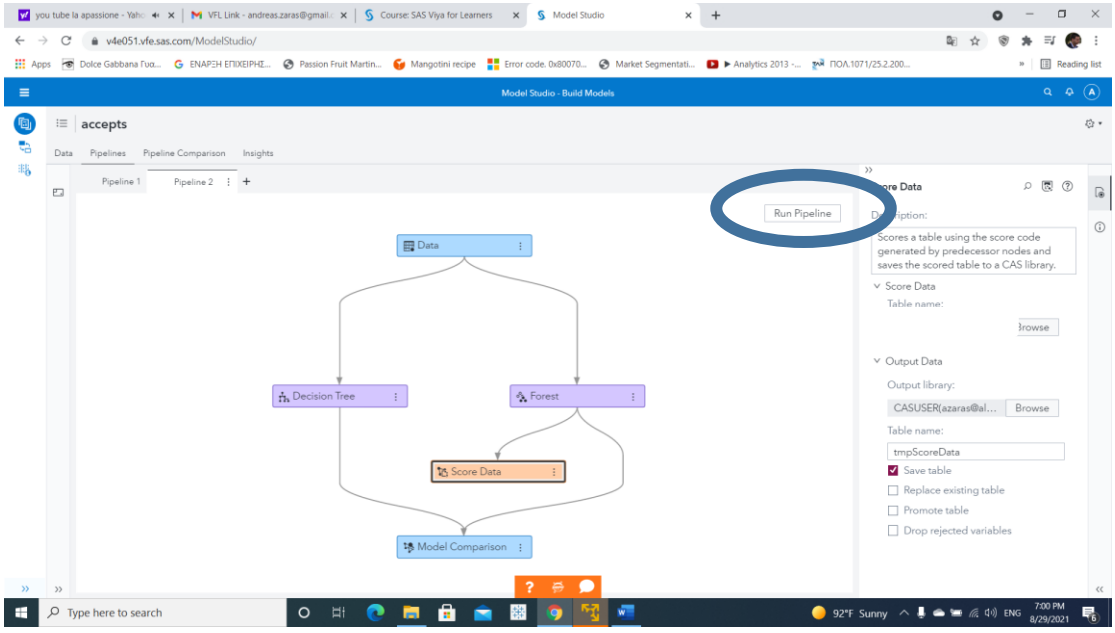


Select the credit_score data set from CASUSER library


Choose Data
CASUSER(azaras@alba.acg.edu)
Name:
CASUSER(azaras@alba.acg.edu)
Description:
Personal File System Caslib
Server:
cas-v4e051-default
Source type:
PATH
true
Path:
/shared/home/azaras@alba.acg.edu/casuser/
Include subdirectories
true
Cancel


accepts
Data
Decision Tree
Forest
Score Data
Model Comparison
Score Data
Description:
Scores a table using the score code generated by predecessor nodes and saves the scored table to a CAS library.
Table name:
Output Data
Output library:
Table name:
tmpScoreData
Save table
Promote table
Drop rejected variables





You can change the name of tmpScoredata to Scored_Data.

Run the pipeline, left click in the three dots in the Score Data node and select Results. Then select Output Data and then View Output Data