

Assignment-3

Name: G.Naga Sai Praneeth

Regd.No:19MIM10044

Classification of Term Deposit Buyers for Retail Banking

Dataset: bank.csv

Create the project and upload the dataset, after uploading the dataset create a data refinery.

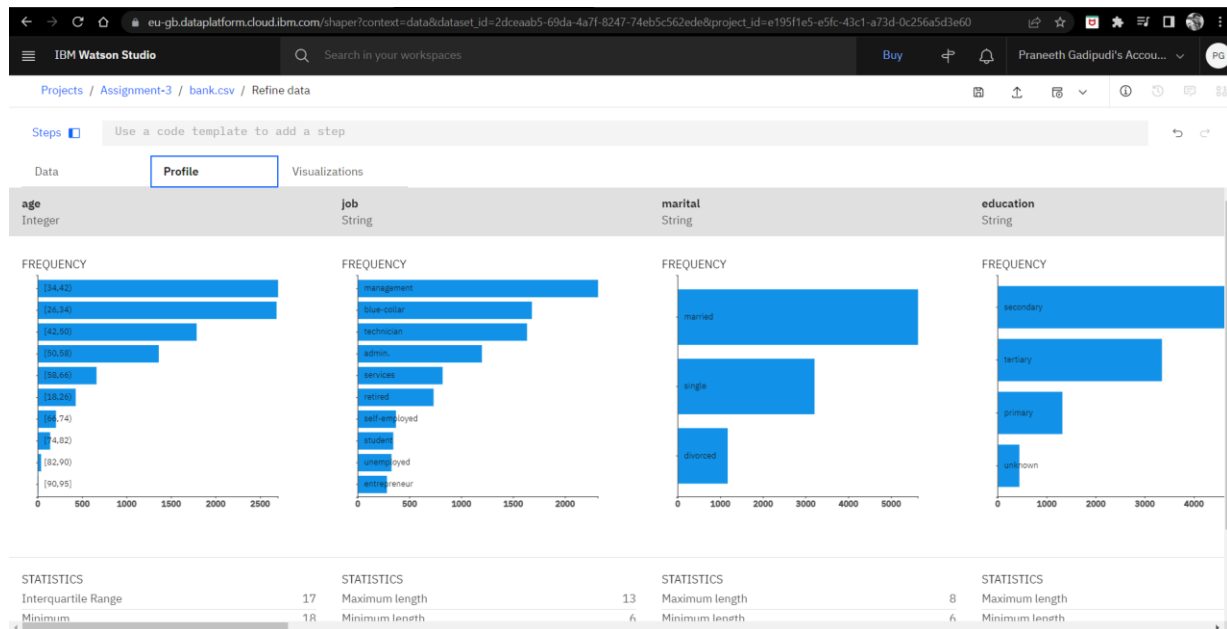
Data Refine:

The screenshot displays the IBM Watson Studio interface for a data refinery project. The top navigation bar shows the project path: **Projects / Assignment-3 / bank.csv / Refine data**. A search bar is available with the text "Search in your workspaces". The main content area is titled "Use a code template to add a step". Below this, a table displays the first 50 rows of the dataset, with columns: **age**, **job**, **marital**, **education**, **default**, **balance**, and **housing**. The table is paginated, showing rows 1 through 14. To the right of the table, a sidebar contains a "Details" tab and a "Help" tab. The "Details" tab shows the "LOCATION" as "Assignment-3", the "DATA REFINERY FLOW NAME" as "bank.csv_flow", and the "STEPS" as "0". The "DATA REFINERY FLOW OUTPUT" section shows a "Location" field.

	age	job	marital	education	default	balance	housing
1	59	admin.	married	secondary	no	2343	yes
2	56	admin.	married	secondary	no	45	no
3	41	technician	married	secondary	no	1270	yes
4	55	services	married	secondary	no	2476	yes
5	54	admin.	married	tertiary	no	184	no
6	42	management	single	tertiary	no	0	yes
7	56	management	married	tertiary	no	830	yes
8	60	retired	divorced	secondary	no	545	yes
9	37	technician	married	secondary	no	1	yes
10	28	services	single	secondary	no	5090	yes
11	38	admin.	single	secondary	no	100	yes
12	30	blue-collar	married	secondary	no	309	yes
13	29	management	married	tertiary	no	199	yes
14	46	blue-collar	single	tertiary	no	460	yes

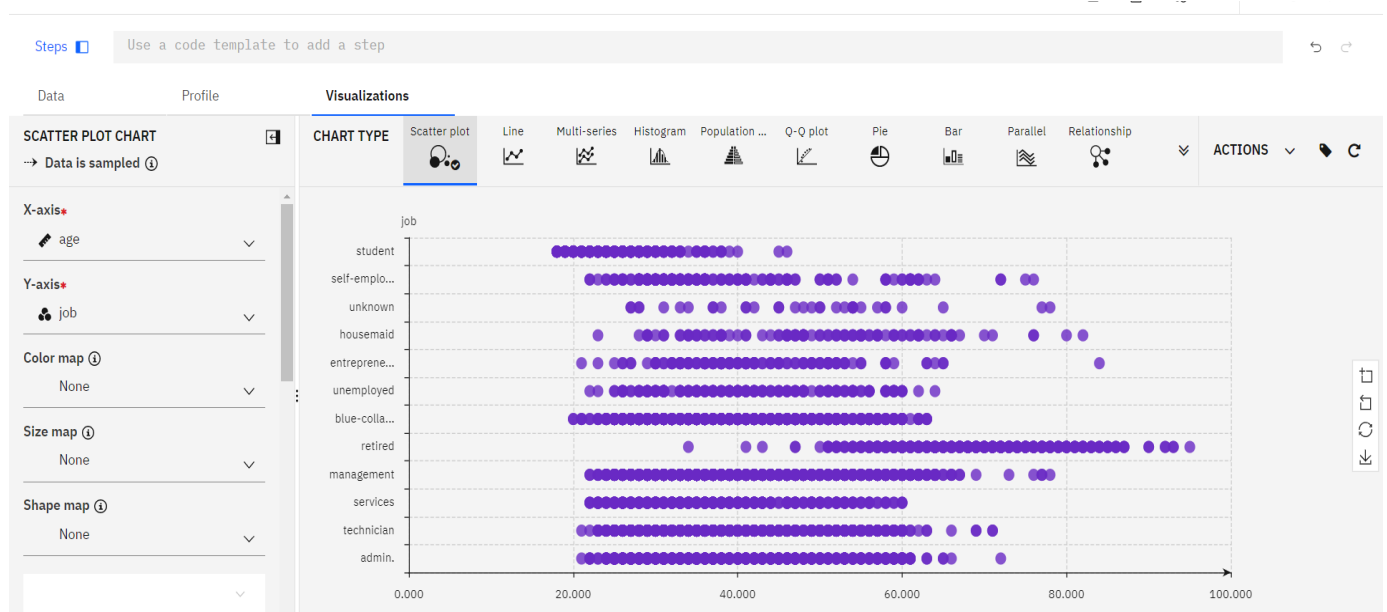
SOURCE FILE: bank.csv SAMPLE SIZE: First 50 rows

Profile:

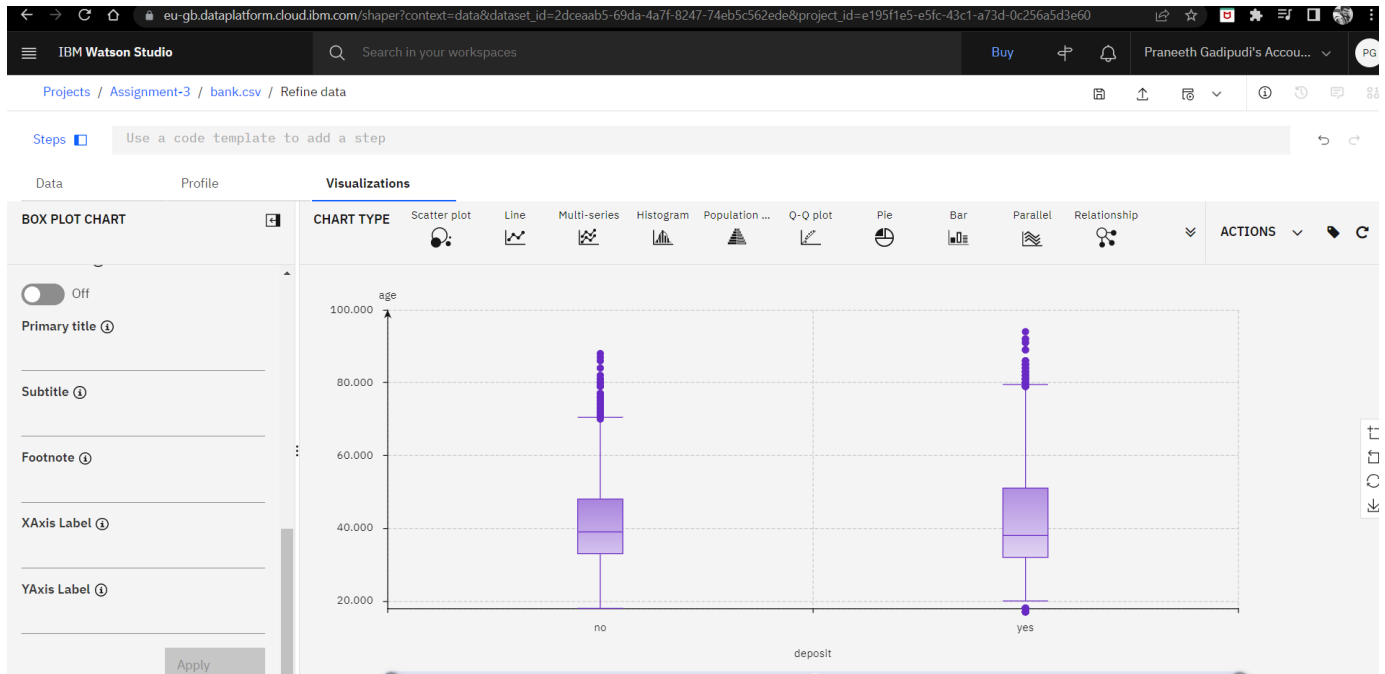


Visualization:

Scatter Plot age vs job



Box Plot:



Create a SPSS Modeler and dataset:

The screenshot shows the IBM Watson Studio interface for creating a SPSS Modeler dataset. The top navigation bar is consistent with the previous image. The breadcrumb trail indicates the current location: Projects / Assignment-3 / SPSS Assignment-3. Below the breadcrumb, there's a 'Find palette nodes' search bar and a list of available nodes: Import, Data Asset, User Input, Sim Gen, Extension Import, Record Operations, Field Operations, Modeling, Text Analytics, Graphs, Outputs, and Export. The 'Data Asset' node is selected, and its configuration panel is displayed on the right. The panel includes a 'Data Asset' title, a 'Data' section with 'Change data asset' and 'Refresh' buttons, a 'Source location' field set to 'bank.csv', a 'File format properties' section with 'File format' set to 'CSV', 'Encoding' set to 'UTF-8', and a checked 'First line is header' option. The 'Invalid data handling' is set to 'Fail'. At the bottom, there are 'Cancel' and 'Save' buttons.

Add Data Audit:

The screenshot shows the IBM Watson Studio interface. The top navigation bar includes the IBM Watson Studio logo, a search bar, and user information. The main workspace displays a 'Data Asset' icon connected to a '17 Fields' icon. The 'Run selection' button is highlighted in blue. The 'Outputs' panel on the right shows a list of outputs, including 'All results' and '17 Fields'.

Find palette nodes

Run selection

Import

Record Operations

Field Operations

Modeling

Text Analytics

Graphs

Outputs

Table

Data Asset

17 Fields

Outputs

Models

Outputs are available during your SPSS Modeler session. After your session ends, the outputs are no longer available.

All results

17 Fields Just now

Data Audit Output:

The screenshot shows the 'View Output: 17 Fields' window. The table displays the following data:

	Field	Sample Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
1	age		Continuous	18	95	41.232	11.913	0.863	--	11162
2	job		Categorical	--	--	--	--	--	12	11162
3	marital		Categorical	--	--	--	--	--	3	11162
4	education		Categorical	--	--	--	--	--	4	11162
5	default		Categorical	--	--	--	--	--	2	11162
6	balance		Continuous	-6847	81204	1528.539	3225.413	8.225	--	11162

Type Node:

IBM Watson Studio interface showing the 'Type' node configuration. The 'Data Asset' node is connected to the 'Type' node. The 'Type' node configuration panel is open, showing a table of fields with their measures, roles, and value modes. The 'Default mode' is set to 'Read metadata'.

Field	Measure	Role	Value mode	Values
# day	Continuous	None	Read	
# month	Categorical	None	Read	
# duration	Continuous	Input	Read	
# campaign	Continuous	Input	Read	
# pdays	Continuous	Input	Read	
# previous	Continuous	Input	Read	
# poutcome	Categorical	Input	Read	
# deposit	Categorical	Target	Read	

Output:

View Output: 17 Fields

Field	Sample Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
1 age		Continuous	18	95	41.232	11.913	0.863	--	11162
2 job		Categorical	--	--	--	--	--	12	11162
3 marital		Categorical	--	--	--	--	--	3	11162
4 education		Categorical	--	--	--	--	--	4	11162
5 default		Categorical	--	--	--	--	--	2	11162
6 balance		Continuous	-6847	81204	1528.539	3225.413	8.225	--	11162

Change the outliers and extremes to 0 and run the model again

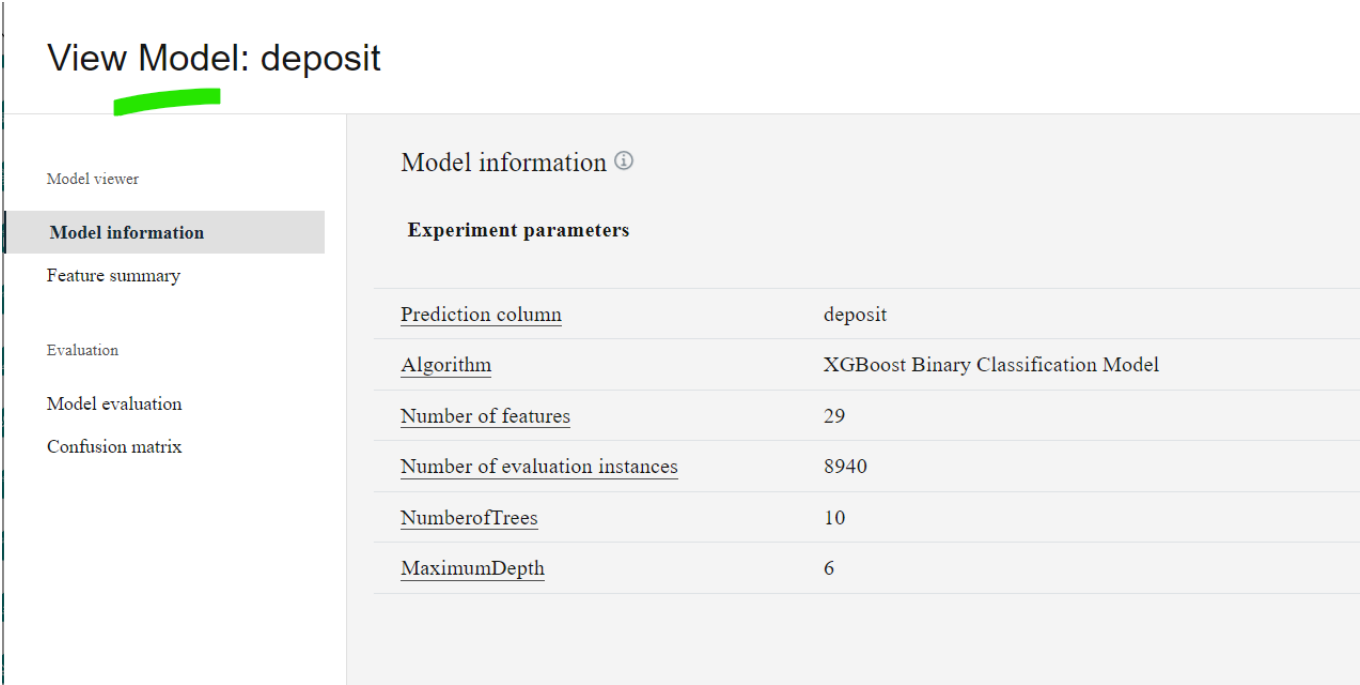
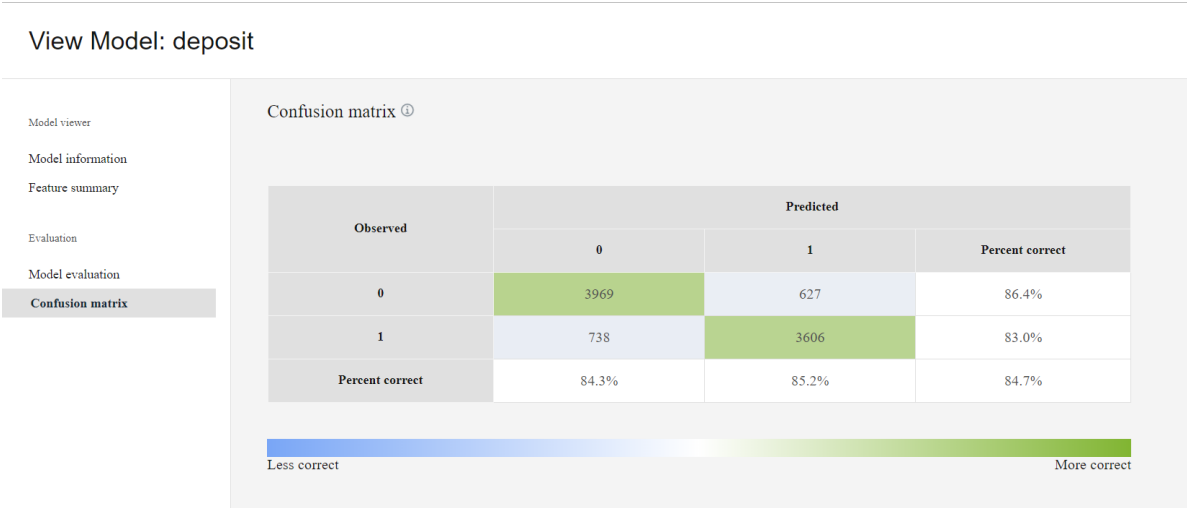
Partition Node:

The screenshot shows the IBM Watson Studio interface. On the left, a palette of nodes is visible, including 'Auto Data Prep', 'Type', 'Filter', 'Derive', 'Filler', 'Reclassify', 'Binning', 'RFM Analysis', 'Ensemble', 'Partition', 'SetToFlag', 'Restructure', 'Transpose', and 'Field Partitioner'. The main workspace displays a workflow: 'Data Asset' (purple icon) points to 'Type' (blue icon), which points to 'Partition' (blue icon). Below 'Data Asset', a green icon labeled '17 Fields' is shown. The 'Partition' node is highlighted with a dashed border. On the right, the 'Partition' node's settings panel is open. It includes a 'Derived Field Name' field with the value 'Partition', a 'Training Partition(%)' field with the value '80', and a 'Testing Partition(%)' field with the value '20'. There are checkboxes for 'Create validation partition' (unchecked), 'Repeatable partition assignment' (checked), and 'Use unique field to assign partitions' (unchecked). A 'Seed' field contains the value '1234567'. At the bottom of the settings panel are 'Cancel' and 'Save' buttons.

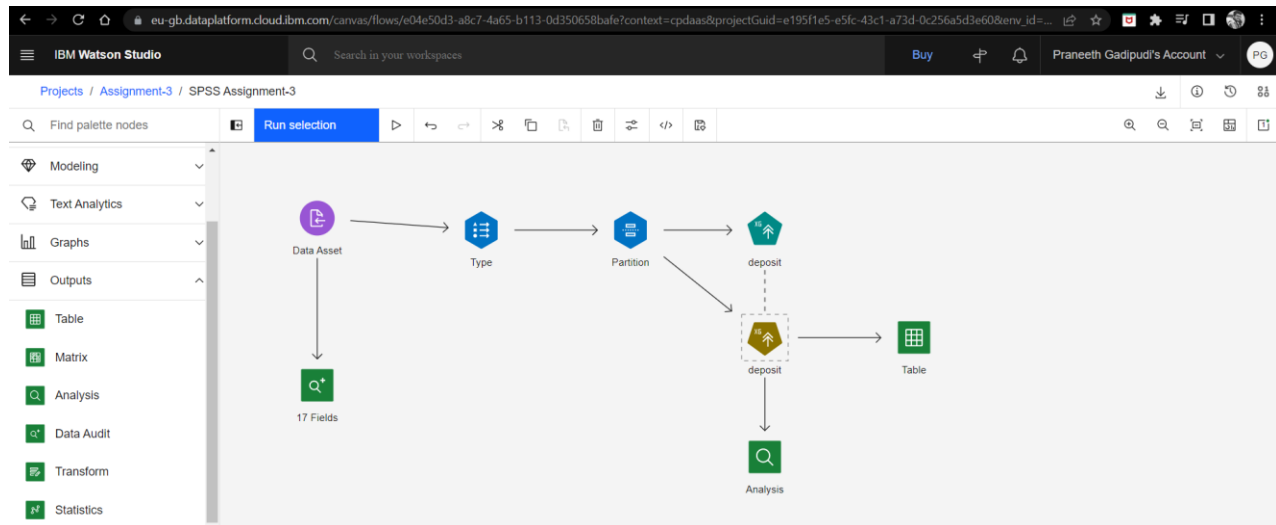
Creating XG Boost Tree Model:

The screenshot shows the IBM Watson Studio interface with a workflow for creating an XGBoost Tree model. The main workspace displays a workflow: 'Data Asset' (purple icon) points to 'Type' (blue icon), which points to 'Partition' (blue icon). Below 'Data Asset', a green icon labeled '17 Fields' is shown. From the 'Partition' node, two arrows point to 'deposit' nodes (green icons). On the right, the 'Outputs' panel is open, showing a list of outputs. The first output is 'All results' (checkbox unchecked). The second output is '17 Fields' (checkbox unchecked), with a status of 'Just now'. The 'Models' tab is also visible, showing a list of models.

Output:



Creating Table and Analysis Nodes:



Analysis Output:

The screenshot shows the 'View Output: Analysis of [deposit]' window in IBM Watson Studio. The output is titled 'Results for output field deposit' and includes a section 'Comparing \$XGT-deposit with deposit'. The data is presented in a table with columns for 'Partition', '1_Training', and '2_Testing'. The table is divided into 'Correct' and 'Wrong' rows, and a 'Total' row. The 'Correct' row shows 7,575 correct predictions (84.73%) for training and 1,787 correct predictions (80.42%) for testing. The 'Wrong' row shows 1,365 wrong predictions (15.27%) for training and 435 wrong predictions (19.58%) for testing. The 'Total' row shows 8,940 total predictions for training and 2,222 total predictions for testing.

'Partition'	1_Training	2_Testing
Correct	7,575 84.73%	1,787 80.42%
Wrong	1,365 15.27%	435 19.58%
Total	8,940	2,222

Output Table:

View Output: Table (20 fields, 11,162 records) [Compare](#)

	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	
1	59	admin.	married	secondary	no	2343	yes	no	unknown	5	may	1042	1	-1	0	unknown	y
2	56	admin.	married	secondary	no	45	no	no	unknown	5	may	1467	1	-1	0	unknown	y
3	41	technician	married	secondary	no	1270	yes	no	unknown	5	may	1389	1	-1	0	unknown	y
4	55	services	married	secondary	no	2476	yes	no	unknown	5	may	579	1	-1	0	unknown	y
5	54	admin.	married	tertiary	no	184	no	no	unknown	5	may	673	2	-1	0	unknown	y
6	42	management	single	tertiary	no	0	yes	yes	unknown	5	may	562	2	-1	0	unknown	y
7	56	management	married	tertiary	no	830	yes	yes	unknown	6	may	1201	1	-1	0	unknown	y
8	60	retired	divorced	secondary	no	545	yes	no	unknown	6	may	1030	1	-1	0	unknown	y
9	37	technician	married	secondary	no	1	yes	no	unknown	6	may	608	1	-1	0	unknown	y
10	28	services	single	secondary	no	5090	yes	no	unknown	6	may	1297	3	-1	0	unknown	y
11	38	admin.	single	secondary	no	100	yes	no	unknown	7	may	786	1	-1	0	unknown	y
12	30	blue-collar	married	secondary	no	309	yes	no	unknown	7	may	1574	2	-1	0	unknown	y
13	29	management	married	tertiary	no	199	yes	yes	unknown	7	may	1689	4	-1	0	unknown	y
14	46	blue-collar	single	tertiary	no	460	yes	no	unknown	7	may	1102	2	-1	0	unknown	y
15	31	technician	single	tertiary	no	703	yes	no	unknown	8	may	943	2	-1	0	unknown	y

Using Custom Field Roles and then select 4 attributes:

eu-gb.dataplatform.cloud.ibm.com/canvas/flows/e04e50d3-a8c7-4a65-b113-0d350658baf?context=cpdaas&projectGuid=e195f1e5-e5fc-43c1-a73d-0c256a5d3e60&env_id=...

IBM Watson Studio

Search in your workspaces

Buy

Praneeth Gadipudi's Account

PG

Projects / Assignment-3 / SPSS Assignment-3

Find palette nodes

Run selection

Modeling

Text Analytics

Graphs

Outputs

Table

Matrix

Analysis

Data Audit

Transform

Statistics

Means

Report

Set Globals

Data Asset

Type

Partition

deposit

deposit

17 Fields

XGBoost Tree

deposit

Fields

Use custom field roles

Target

deposit

Predictors

Field name

housing

loan

contact

duration

Basic

Cancel

Save

Build the Model again:

eu-gb.dataplatform.cloud.ibm.com/canvas/flows/e04e50d3-a8c7-4a65-b113-0d350658baf?context=cpdaas&projectGuid=e195f1e5-e5fc-43c1-a73d-0c256a5d3e60&env_id=...

IBM Watson Studio

Search in your workspaces

Buy

Praneeth Gadipudi's Account

PG

Projects / Assignment-3 / SPSS Assignment-3

Find palette nodes

Run selection

Modeling

Text Analytics

Graphs

Outputs

Table

Matrix

Analysis

Data Audit

Transform

Data Asset

Type

Partition

deposit

deposit

Table

Analysis

17 Fields

Outputs

Models

Outputs are available during your SPSS Modeler session. After your session ends, the outputs are no longer available.

All results

17 Fields

1 minute ago

Table (20 fields, 11,162 re...)

6 minutes ago

Analysis of [deposit]

6 minutes ago

Analysis Output:

eu-gb.dataplatform.cloud.ibm.com/canvas/flows/e04e50d3-a8c7-4a65-b113-0d350658bafefcontext=cpdaas&projectGuid=e195f1e5-e5fc-43c1-a73d-0c256a5d3e60&env_id=...

IBM Watson Studio Search in your workspaces Buy Praneeth Gadipudi's Account

Projects / Assignment-3 / SPSS Assignment-3

View Output: Analysis of [deposit] #1

Compare

Collapse All

Results for output field deposit

Comparing \$XGT-deposit with deposit

'Partition'	1_Training		2_Testing	
Correct	7,575	84.73%	1,787	80.42%
Wrong	1,365	15.27%	435	19.58%
Total	8,940		2,222	

Table Output:

View Output: Table (20 fields, 11,162 records) #1																			
	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome			
1	59	admin.	married	secondary	no	2343	yes	no	unknown	5	may	1042	1	-1	0	unknown	y		
2	56	admin.	married	secondary	no	45	no	no	unknown	5	may	1467	1	-1	0	unknown	y		
3	41	technician	married	secondary	no	1270	yes	no	unknown	5	may	1389	1	-1	0	unknown	y		
4	55	services	married	secondary	no	2476	yes	no	unknown	5	may	579	1	-1	0	unknown	y		
5	54	admin.	married	tertiary	no	184	no	no	unknown	5	may	673	2	-1	0	unknown	y		
6	42	management	single	tertiary	no	0	yes	yes	unknown	5	may	562	2	-1	0	unknown	y		
7	56	management	married	tertiary	no	830	yes	yes	unknown	6	may	1201	1	-1	0	unknown	y		
8	60	retired	divorced	secondary	no	545	yes	no	unknown	6	may	1030	1	-1	0	unknown	y		
9	37	technician	married	secondary	no	1	yes	no	unknown	6	may	608	1	-1	0	unknown	y		
10	28	services	single	secondary	no	5090	yes	no	unknown	6	may	1297	3	-1	0	unknown	y		
11	38	admin.	single	secondary	no	100	yes	no	unknown	7	may	786	1	-1	0	unknown	y		
12	30	blue-collar	married	secondary	no	309	yes	no	unknown	7	may	1574	2	-1	0	unknown	y		
13	29	management	married	tertiary	no	199	yes	yes	unknown	7	may	1689	4	-1	0	unknown	y		
14	46	blue-collar	single	tertiary	no	460	yes	no	unknown	7	may	1102	2	-1	0	unknown	y		
15	31	technician	single	tertiary	no	703	yes	no	unknown	8	may	943	2	-1	0	unknown	y		